Video 360 Content Navigation for Mobile HMD Devices

Jounsup Park*, Mingyuan Wu, Eric Lee
Klara Nahrstedt
University of Illinois at Urbana-Champaign

Yash Shah*, Arielle Rosenthal, John Murray,
Kevin Spiteri, Michael Zink, Ramesh Sitaraman
University of Massachusetts, Amherst

ABSTRACT

We demonstrate a video 360 navigation and streaming system for Mobile HMD devices. The Navigation Graph (NG) concept is used to predict future views that use a graph model that captures both temporal and spatial viewing behavior of prior viewers. Visualization of video 360 content navigation and view prediction algorithms is used for assessment of Quality of Experience (QoE) and evaluation of the accuracy of the NG-based view prediction algorithm.

KEYWORDS

video 360, View prediction, Rate adaptation, QoE

1 INTRODUCTION

Mobile Head Mounted Displays (HMD) allow viewers to enjoy Virtual Reality (VR) contents including video 360 anywhere. Mobile HMDs receive video streams from media server and render the video to show a 360-degree view based on the viewer's head movement. However, video 360 requires very high bandwidth to be streamed over the mobile network, because video 360 movies have very high spatial resolution. Therefore, it is necessary to cut the video into smaller resolution videos, so called tiles, and stream the tiles which are necessary to cover the viewport. Since there is a latency between the media server and the mobile HMD, the mobile HMD should request future tiles in advance from the server to continuously show high quality video to the viewer.

Even though viewers have freedom to see anywhere in the video 360 content, there are physical restrictions that limit the viewers' behavior. For example, a viewer cannot suddenly turn their head more than 180-degree and the viewer cannot watch two separate regions at the same time. Moreover, there are common head movement patterns that many viewers show while they are watching the same video. These facts help modeling how viewers navigate through the video 360 content. The Navigation Graph (NG) concept [7] is used to build the viewing and navigation behavior model of video 360 content. This model gathers view information from many prior viewers for each video 360 movie to generate a NG for each video 360 and this information is provided to the mobile HMDs when they request video segments. Mobile HMDs predict the viewer's future view using the NG information to request appropriate qualities of future tiles.

*Both authors contributed equally on this paper and can be contacted at following: jounsup@illinois.edu, ynshah@umass.edu

MM '20, October 12–16, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-7988-5/20/10.
https://doi.org/10.1145/3394171.3414389

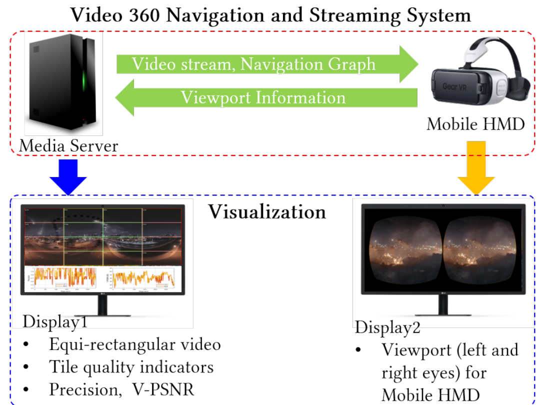

Figure 1: Video 360 Content Navigation System

Our approach of using the NG-concept is different from other view prediction algorithms using Linear Regression (LR) [1, 6] or classification of viewers into groups [5] to use different probability maps. LR is useful for capturing temporal variations of viewer's behavior and classification of viewers is helpful to find spatial regions of viewers' interest. The NG represents a systematic viewing model that gathers multiple viewers' information and records both temporal and spatial navigation behaviors of viewers. Moreover, the NG is compatible with Dynamic Adaptive Streaming over HTTP (DASH) because mobile HMDs can perform rate adaptation heuristic using NG-information that can be found in the Media Presentation Description (MPD) file.

2 VIDEO 360 NAVIGATION AND STREAMING SYSTEM

In this technical demonstration, we show a video 360 content navigation and streaming system between a media server and a mobile HMD (Figure 1) [1]. The media server stores video 360 content and NG information for each video 360 movie. Raw video 360 movies have very high resolution, so each movie is divided into tiles. Tiles are compressed using ffmpeg [2] to generate multiple copies of the tiles with different qualities. We divided the equi-rectangular video into 4×4 tiles and stored in a media server. Each tile is encoded in three different qualities. The information about the adaptation sets of tiles are stored in MPD file. The NG is updated whenever the server receives view information from the mobile HMDs. We have built the NG based on 48-viewers' trace dataset [3]. The NG information is provided to the mobile HMD along with the MPD. Therefore, a viewer can watch high-quality video by predicting future views using the NG information.

[1]https://youtu.be/LirmYGTzEPI

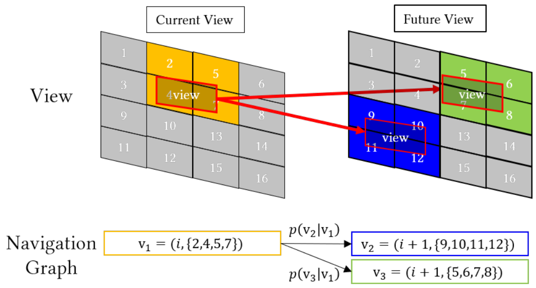

Figure 2: View Prediction using Navigation Graph (NG)

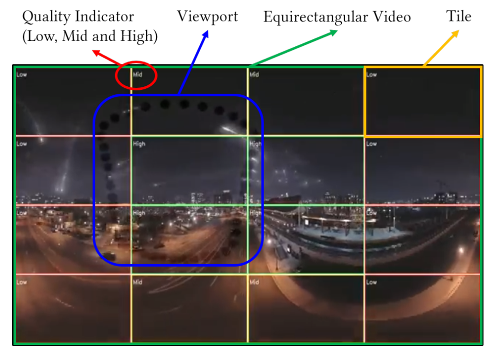

Figure 3: 4×4 **Tiled Equirectangular Video with Viewport**

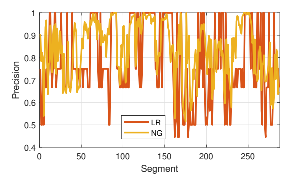

Figure 4: Precision

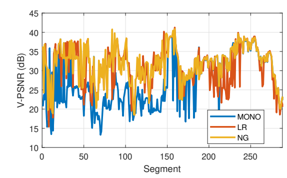

Figure 5: V-PSNR

We installed the ToucanVR [4] player on a Samsung Galaxy S8 phone to make the Android device work as a mobile HMD. ToucanVR can play the video either stored in the device or streamed from the media server. The ToucanVR player performs rate adaptation algorithm to request appropriate qualities of tiles. NG information is used to calculate the probability that the tiles will be shown in the next segment. The GearVR headset is attached to the Samsung Galaxy S8 phone to display viewport and it sends back the viewport information to the server. The server gathers viewport information from the mobile HMD and updates the NG.

Figure 2 shows the example of an NG that is used to predict future views given current view information. The NG has two possible future views $\mathbf{v}_2$ and $\mathbf{v}_3$ when the current view is $\mathbf{v}_1$, which is a set of visible tiles. The view transition probabilities, $p(\mathbf{v}_2|\mathbf{v}_1)$ and $p(\mathbf{v}_3|\mathbf{v}_1)$, are derived from the statistics for how many times other viewers move from the current view to the subsequent views. The mobile HMD requests tiles that have higher probability to be seen first. Therefore, we can show high quality video in the viewer's viewport even when the network condition is bad.

3 VISUALIZATION OF SYSTEM PERFORMANCE

We visualize the performance of the navigation and streaming system using two displays (Display 1 and Display 2 in Figure 1) to evaluate Quality of Experience (QoE) and an accuracy of the view prediction algorithm while streaming video 360. There are several performance metrics introduced to measure the performance of video 360 streaming systems. However, none of the metrics can solely represent the performance of the systems because multimedia experience is difficult to express via numbers. Therefore, we implement the visualization of real viewing experience of the viewers along with performance metrics to show actual performance of the systems. It shows the actual quality of video 360 streamed over the mobile network and corresponding performance metrics, such as Viewport-PSNR (V-PSNR) and Precision (prediction accuracy).

First, view prediction results are represented by the selected qualities of tiles (Figure 3). We color-code the outline of the tiles and label the tiles as Low, Medium or High to represents the qualities of the tiles which will be shown to the viewer. The viewport is overlaid with the video to indicate the pixels that a viewer is actually seeing. The outline of the actual viewport is represented by black dots. In ideal case, tiles that have overlap with the viewport should have high quality. Performance metrics, such as precision and V-PSNR, are also plotted on the display. The prediction accuracy is measured as precision, which is measured as $\sum_{t=1}^{T} \min(w_t, g_t)$ where g_t is a normalized ground truth that has 0 for non-visible tiles and 1/(number of visible tiles) for visible tiles, and w_t is a predicted probability that tile t will be shown in the future segment. Figure 4 shows the precision results using NG-based prediction and Baseline method (Linear Regression [6]). V-PSNR represents how similar the received video is with the original videos. Figure 5 shows the V-PSNR performance when NG-based prediction and Baseline methods (LR and MONO) are used for the streaming system. MONO represents the V-PSNR when every tile has the same probability to be shown in the future segment assuming that the streaming system has no prediction system. Second, we show the real viewport in our system. The display of the mobile HMD is mirrored to the monitor, so it can be observed what viewers watch. Mobile HMD shows two videos for left and right eyes. Therefore, there are two similar videos shown in the mirrored display (Display 2 in Figure 1).

4 CONCLUSION

We implement video 360 navigation and streaming system for mobile HMD. It streams the video 360 from the server to the mobile HMD over the mobile network. It efficiently utilizes the limited bandwidth of mobile network by predicting future view using Navigation Graph. We also implement the visualization of the system to efficiently evaluate the system performance, including view prediction and Quality of Experience (QoE).

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. CNS-1900875 and No. CNS-1901137.

REFERENCES

[1] Y. Bao, H. Wu, T. Zhang, A. A. Ramli, and X. Liu. 2016. Shooting a moving target: Motion-prediction-based transmission for 360-degree videos. In *2016 IEEE International Conference on Big Data (Big Data)*. 1161–1170. https://doi.org/10.1109/BigData.2016.7840720

[2] Fabrice Bellard. https://www.ffmpeg.org. (FFmpeg https://www.ffmpeg.org).

[3] W. Chenglei, T. Zhihao, W. Zhi, and Y. Shiqiang. 2017. A Dataset for Exploring User Behaviors in VR Spherical Video Streaming. In *ACM Multimedia Systems Conference (MMSys)*.

[4] Savino Dambra, Giuseppe Samela, Lucile Sassatelli, Romaric Pighetti, Ramon Aparicio-Pardo, and Anne-Marie Pinna-Déry. 2018. Film Editing: New Levers to Improve VR Streaming. In *Proceedings of the 9th ACM Multimedia Systems Conference (MMSys '18)*. Association for Computing Machinery, New York, NY, USA, 27–39. https://doi.org/10.1145/3204949.3204962

[5] X. Lan, Z. Xinggong, and G. Zongming. 2018. CLS: A Cross-user Learning Based System for Improving QoE in 360-degree Video Adaptive Streaming. In *Proceedings of the 26th ACM International Conference on Multimedia (MM '18)*. ACM, New York, NY, USA, 564–572. https://doi.org/10.1145/3240508.3240556

[6] X. Lan, X. Zhimin, B. Yixuan, Z. Xinggong, and G. Zongming. 2017. 360Prob-DASH: Improving QoE of 360 Video Streaming Using Tile-based HTTP Adaptive Streaming. In *ACM International Conference on Multimedia (MM)*.

[7] J. Park and K. Nahrstedt. 2019. Navigation Graph for Tiled Media Streaming. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*. Association for Computing Machinery, New York, NY, USA, 447–455. https://doi.org/10.1145/3343031.3351021