VR User Behavior Analysis

Varin Sikand, Ryan Aspenleiter, Rishi Villa, Shreyon Roy, Dr. Prakash

Introduction

Abstract

Streaming a 4k virtual reality (VR) video requires about 600 Mbit/s internet speed connection. This makes the option of viewing VR videos in high resolution not feasible as the average user does not have access to bandwidths of such magnitude.

A potential solution to this problem is to adaptively render each frame such that we only fully render what is being actively looked at (Fig. 1). To accomplish this, we developed a heat map-based approach that makes use of previous user data to determine what's being looked at.

Hypothesis

With data detailing where the user viewports are centered within the frame, we will be able to first discard any areas that are not able to be seen and render the remaining area based on heat magnitude. All of this can be accomplished with negligible loss of user experience and will result in considerable file size savings.

This hypothesis was tested using a selection of 6 VR videos, each with a separate collection of viewport traces per frame from 60 different users[1].

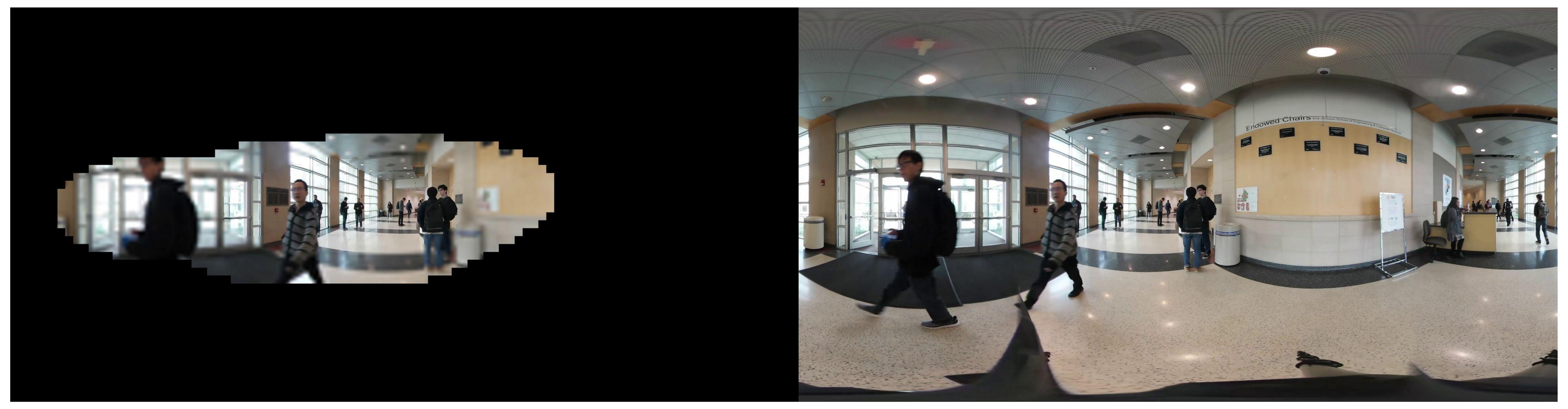

Fig. 1: Compressed (left) vs Original (right)

Process

Heat Maps

First, we developed a heat map by splitting the frame into square regions and then counting the number of viewport centers within each region. Predictably, this generated a very simple heat map that could define the most popular general region but not how attention is distributed throughout the frame, at least not continuously.

To improve on this approach, we developed a heatmap that stacked viewport sized ellipses of varying heat for each user's center of vision. (Fig. 2). These circles relied on a "voting" system that assigned regions close to the center higher values while the values moving away from the center decayed according to one of our four decay functions. Consequently, the areas that expand outward from the center are given a lower resolution, while the rate at which this decrease in resolution quality occurred depended on the decay function selected.

Results

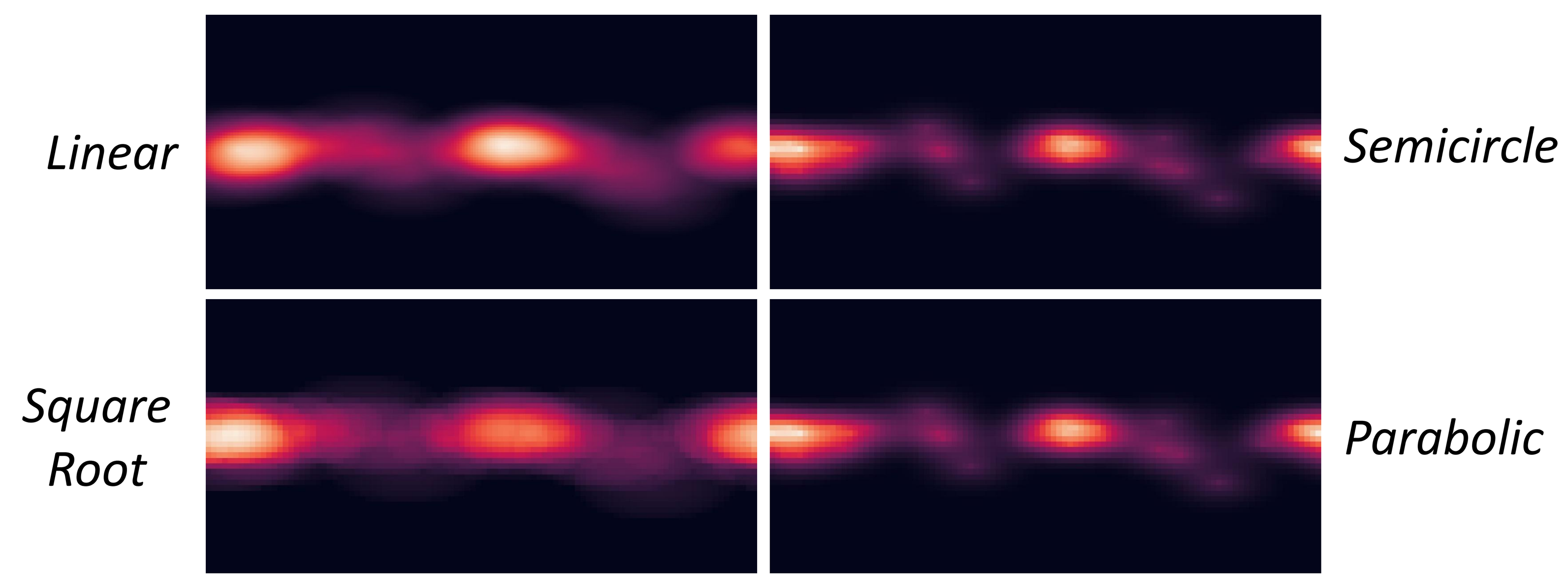

Fig. 2: Gradient Heatmaps per Decay Function

Results

After proceeding with the heat map discussed prior and choosing the two least extreme of our decay functions (semicircle decay and linear decay) to compare, we saw notable differences in how aggressively the image was compressed. We defined the storage statistic as the ratio of the compressed image's size as compared to the original. Taking the central 80% of a user's viewport as "important", we similarly defined the "user experience rating" as the percentage that "important" area that is rendered in full resolution. We found that in exchange for a 10% average increase in storage size we could achieve a 78.5% average user experience rating with the linear decay function as opposed to the 55% rating with our semicircle decay function. In both cases, the storage space is reduced by over 50% and every user is guaranteed to not have any discarded (black) space within their viewport.

Fig. 3: User Experience + Storage for Semicircle Decay (top) vs Linear Decay (bottom)

Conclusion & Future Directions

Conclusion

Our work gives a proof-of-concept which shows that our methodology can be used to dramatically reduce the amount of space a VR video requires without compromising on user experience.

Machine Learning

The methods used in this project rely on having a large amount of viewport data. Because this may not be the case in many applications, we tested the feasibility of using machine learning models to predict where users are looking based on a smaller amount of data.

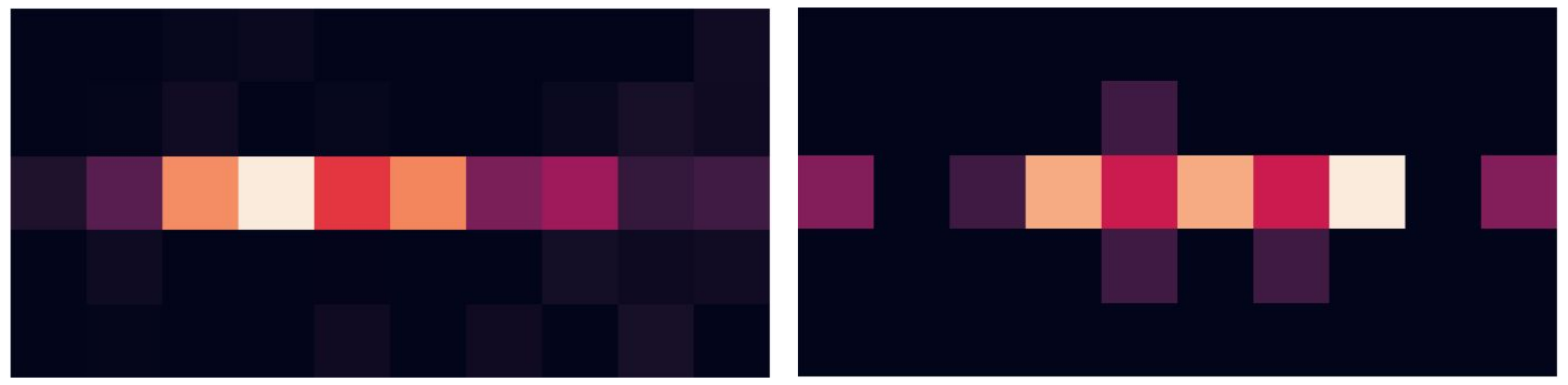

Fig. 4: Predicted Simplified Heatmap (left) vs Actual Simplified Heatmap (right)

Although the model we developed is not fully optimized, it is still accurate enough to indicate that using machine learning to accommodate for a lack of perfect data is viable (Fig. 4). These models could be used to train with new viewport data as it becomes available to continuously improve how videos are compressed.

Future Directions

- Including facial recognition and computer vision to ML Model
- Dividing users into clusters based on common viewing qualities
  - Allows for more accurate and specific heat maps
- Accounting for auditory influences to user viewing activity
- Collaborating with optometrists to standardize voting functions

References

[1] Afshin Taghavi Nasrabadi, Aliehsan Samiei, Anahita Mahzari, Ryan P. McMahan, Ravi Prakash (2019). A Taxonomy and Dataset for 360 Videos. In Proceedings of the 10th ACM Multimedia Systems Conference.

Acknowledgements

The VR User Behavior Analysis team would like to thank Dr. Ravi Prakash and his team for providing valuable viewport data for several VR videos.