ACM RESEARCH

Salient Features in VR Viewports Prediction

Sunny Guan, Matthew Vu, David Wank, Natasha Trayers, Agustin Diocares

UT DALLAS The University of Texas at Dallas

Introduction

Abstract

The accuracy of viewport prediction is critical for adapting 360-degree video streaming. The ability to accurately predict viewports for longer horizons would allow for selective buffering which helps reduce data bandwidth and provide smoother video playback.

For short prediction horizon, viewports are close to the previous sample but for longer horizons, viewport prediction accuracy drops. A longer buffer is desirable for streaming under unstable network conditions. This research investigates the viewport prediction problem for longer horizons by analyzing user interactions with "salient features" - particular areas of interest - in a 360-degree video.

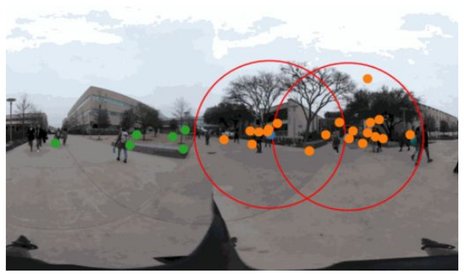

Fig 1: Visual correlation between salient features

Hypothesis

Through the salient feature correlation of VR Viewport videos and the usage of a machine learning model on moving central objects, we can significantly reduce data bandwidth with viewport prediction based on the features users are more likely to look at.

We tested this hypothesis using a previous set of VR videos with complete data regarding salient feature locations and viewport traces from 60 different users watching the videos[1].

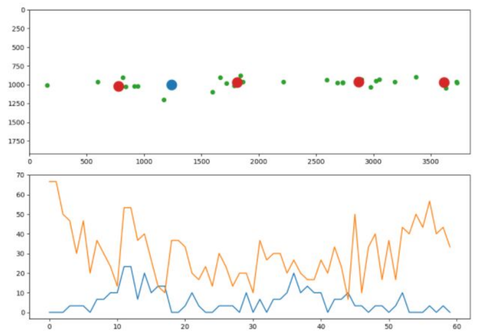

Fig 2: Upper Bound of K-Means Clusters

Methods, Analysis, and Process

Salient Feature Correlations

We determined a "correlation ratio" to measure the accuracy of salient features in viewport prediction. Our correlation ratio was determined by the percentage of user viewport traces the salient features captured divided by the percentage of the screen that the salient features covered.

Our results showed a moderate to significant correlation between salient features and user traces, on a video-to-video basis. In some cases, the correlation ratio determined was moderately above a "normal" sample- as was the case of video 18, with a ratio of 3.19- and in some cases it was immensely significant, as was the case with video 29 and its ratio of 11.67 compared to normal.

We then examined the relationship between salient features and user traces from the perspective of individual types of users. In every video examined, we found no significant difference between correlation ratios for the whole video and ratios for these "predicates"- the difference between these ratios would rarely surpass 5%.

Results

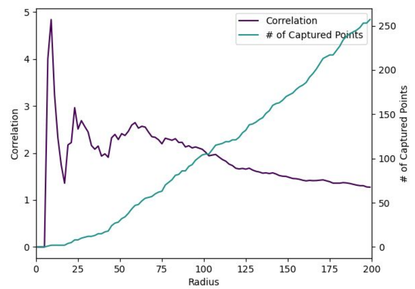

Fig 3: Correlation Ratio vs Salient Feature Radius

This is an example of the correlation results for one video for different radii. We also included the number of captured points for each radius. As we can see, a radius around 50-75 pixels for this particular video is optimal for the correlation ratio. The correlation trends towards 1 for larger radii since four times as many points are required for twice the radius. The initial peak around r=5 is due to low captured points and low area percentage, which is discarded in favor of the second and third local maxima.

Video ID	Base Correlation Ratio	Diff. for Male Subjects	Diff. for Inexp. Subjects
Video 3	0.86	-0.56%	-15.56%
Video 6	14.16	6.62%	-0.74%
Video 12	6.95	-0.74%	4.54%
Video 18	3.19	3.01%	1.44%
Video 23	8.26	-2.14%	5.08%
Video 24	5.28	-2.56%	-6.38%
Video 29	11.67	-7.27%	0.35%

Table 1

Conclusion & Future Directions

Conclusion

Our work provides a better understanding of how salient features can be used in VR viewport prediction for accurate long-horizon predictions.

Prediction Model

By training a prediction model on user viewport traces and salient feature data, we were able to achieve an average error of 10.96%, which is measured by calculating the mean-squared error between each salient feature and their corresponding cluster centroid.

Future research could be aimed at improving the model through better image processing and error calculation. By utilizing an effective model, videos can be preprocessed to accurately predict user viewport and reduce network bandwidth.

Another relevant direction our research could have gone is using edge detection algorithms like ORB or Harris Corner Detection. These could automatically preprocess feature traces more accurately than with the current model.

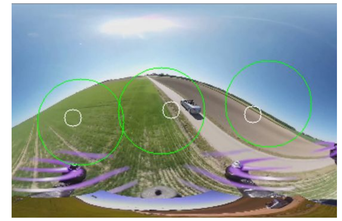

Fig 4: Prediction Example (Green is predicted, white is K-Means)

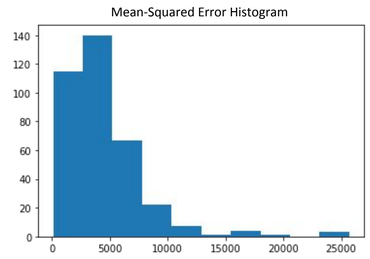

Fig 5: MSE Distribution of Prediction Centroids

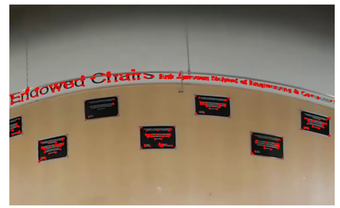

Fig 6: Harris Corner Detection

References

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.
2. Feng Qian, Bo Han, Qingyang Xiao, Vijay Gopalakrishnan (2018). Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices
3. Tilke Judd, Krista Ehinger, Frédo Durand, & Antonio Torralba (2009). Learning to Predict Where Humans Look. In IEEE International Conference on Computer Vision (ICCV).