Paper Summary and Critique

Robust Saliency-Driven Quality Adaptation for Mobile 360-Degree Video Streaming

Name	Student ID	Course	Lecture #
Sahil Pattni	40216177	COMP 691	3

Summary

Streaming high-definition 360-degree videos is bandwidth expensive and cannot be fully supported by the currently mass-implemented wireless networks' capacity. Consider a 360-video as a collection of *tiles*: The experienced quality of the video can be improved by prioritizing the quality of tiles that the user is most likely to look at / be focused on. Most of the research in this domain prior to this paper focused on *head-movement trajectory (HMT)*-based optimization methods. The authors in this paper propose an alternative method, which is an improvement upon their previous conference submission. The proposed system - named **RoSal360** - considers each video tile's probability of being focused on (i.e. its saliency) based on historical gaze data from previous viewers. The system can be segregated into two sub-models:

- A deep neural network model that accurately predicts the transmission time of a video tile (and considers the tile's size). This is used in conjunction with a video tile's assumed saliency to determine the tile's video quality.
- A reinforcement learning model to correct for saliency bias: Outlier viewers may focus on tiles in a video that have been marked un-salient based on historical aggregate viewer data, and therefore will experience a deterioration of their quality of experience.

The paper simulates the stream of their gaze-annotated 360-degree video against various historical network environment data (WiFi, 4G/LTE and 5G). The paper determines that the **RoSal360** algorithm:

- Reduced the re-buffering ratio by up to a factor of 4.11.
- Achieved up to a 4.57 dB improvement in gaze-driven Peak-Signal-to-Noise Ratio (PSNR).

Strengths

- The proposed solution is more accurate than previous state-of-the-art (SotA) solutions, and attempts to mitigate the effects of saliency-bias in order to maximize the quality of experience.
- Model inference can be run cheaply both server-side and client-side.
- The saliency-aware allocation method devised by the authors is a computationally effective approximation of the exhaustive search method, with a speed up factor of 87.

Weaknesses

The system does not perform as well for salient-uniform videos as opposed to salient-uneven videos. Since a salient-uniform video would have similar saliency scores for all regions of the video, there would be no meaningful insight to be derived that could aid in the prediction of user gaze. A caveat to be noted is that the authors found most of their videos used in the experiment to be salient-uneven, and therefore this potential loss of gains would not be accurately represented in their experimental data.

Applicability to practice

When collecting user-gaze data, the users were shown uniformly high-quality tiles. However the authors note that they observed an effect of spatially uneven video quality distribution: users were more likely to gaze at a high-quality region compared to a low-quality region, even if both are deemed to have a high level of saliency. Therefore, not only does the assumed saliency affect the quality, but the quality may also affect the actual saliency. In practice, it may make financial sense to optimize several regions of the video as opposed to a uniform quality stream.

Comments for improvement

The user-gaze collection was done on a relatively small sample size of 30 people, although the authors did try to maximize the diversity of the group.

Even so, with a larger amount of diversity, a larger sample size may need to be considered to determine its effect on the user gaze data.