Perceptual Video Compression with Combined Scene Analysis and Eye-Gaze Tracking

Javed Khan
Kent State University
javed@kent.edu

Oleg Komogortsev
Kent State University
okomogor@cs.kent.edu

1 Introduction

Live eye-tracker based perceptual compression is an active area of research. One of the challenges of the real-time eye-gaze based perceptual video compression is a delay found in any transcoding system. The delay might be large due to transmission/compression time. In this work we investigate the impact of the delay on the area of the image which requires high quality coding. We study the potential for performance improvement by combining eye-tracking with fast scene analysis to narrow down the high acuity area without eye-gaze containment loss.

2 Exposition

It is believed that both scene content and an eye movement patterns determine the precise area of human attention. Our combined scheme uses both these facts to calculate the exact area that requires high quality coding. The combined scheme is represented in *Perceptual Object Window* (W^{POW}), which is a combination of the two schemes *Reflex Window* (W^{RW}) and *Object Window* (W^{TW}). We evaluate the performance of each method through a set of performance factors*: gaze containment* – the amount of eye-gazes contained over some period of time by a given windowing scheme, and *perceptual coverage* – the fraction of the video frame covered by a windowing scheme. The perceptual coverage directly shows what percentage of the image requires the highest quality coding. The less perceptual coverage there is, the more potential for compression exists.

The objective of the Reflex Window is to contain eye fixations by estimating an eye velocity in the future. Given a set of past eye-positions while considering a current value for delay introduced by the encoding system, the W^{RW} denotes a zone where an eye will be at during a certain point in the future from its current position with certain likelihood [2002].

The Object Window represents a portion of the image that contains an object within the video frame. The position and size of such an area depends on the location and the dimensions of the object, its speed, etc. The objective of the W^{TW} construction mechanism is to identify and track the object in real time and provide the boundaries of the object's shape [2003]. W^{TW} tracks the object that the subject is currently looking at.

We propose several methods for *Perceptual Object Window* construction. Two of them are modifications of the Object Tracking Window and three of them are combinations of both W^{RW} and W^{TW}. The first two methods are based on the idea of approximating the W^{TW} by two simple shapes: rectangle and circle. Rectilinear approximation (W^{RTW}) and circular approximation (W^{CTW}) are instances of Perceptual Object Window (W^{POW}). Both of these methods shows some improvement in gaze containment compared to the original W^{TW}, however, they increase perceptual coverage and consequently decrease the potential for compression. It is noteworthy that W^{CTW} is a preferable alternative than W^{RTW}, as it is a circular shape and is therefore more suitable for perceptual compression. The combined method considers information from both W^{RW} and W^{CTW} for W^{POW} construction, taking into consideration such factors as overlapping, relative position to each other, and performing dynamic size control. The results for this method show significant improvement over the original W^{RW} and W^{TW}. In the event of a 155ms system delay, the combined method reduces perceptual coverage by 10-20%, and retains gaze containment at a level of 70-90%. In the event of a 1sec delay (the situation W^{RW} becomes large around 30% of the video frame), the combined method performs far better. It reduces perceptual coverage by 30-70%, while maintaining gaze containment at 70-90%.

3 Conclusion

Recent research in eye-gaze image and video compression mainly focuses on the acuity distribution function around the fovea, which expands to about 2 degrees. However, it seems the impact of transcoding loop delay can be much more important, as an eye can potentially move a long way during that period. The dynamic estimation and optimization of the acuity area introduced by the network delay becomes more important than a precise calculation of peripheral acuity degradation.

Considering the delay factor we can state that peripheral degradation approximation might work as well as the very precise one in a practical system. Unfortunately, the delay aspect of perceptual encoding has not been thoroughly investigated. We propose a solution to that problem by predicting future eye movements based on the previous eye behavior and we further improve our scheme by incorporating intelligent object tracking into our algorithm. The combined approach allows significant reduction of the area of the image requiring high quality coding. It provides a greater compression potential without bringing a considerable computational overhead, which some object detection schemes might introduce.

References

J. KHAN, O. KOMOGORTSEV. 2002. Dynamic Gaze Span Window-based Foveation for Perceptual Media Streaming. *Technical Report TR2002-11-01*, Kent State University, [available at URL http://bristi.facnet.mcs.kent.edu/medianet].

J KHAN and Z. GUO. 2003. Flock-of-Bird Algorithm for Fast Motion Based Object Tracking and Transcoding in Video Streaming. In *The 13th IEEE International Packet Video Workshop 2003*, Nantes, France.