Variability of eye movements on natural videos

M Dorr, M Böhme, T Martinetz, K Gegenfurtner[†], and E Barth
Institute for Neuro- and Bioinformatics, University of Lübeck
[†]Abteilung Allgemeine Psychologie, Universität Giessen

We investigate the variability of eye movement patterns on high-resolution natural videos as well as on movie trailers. To compare the variability across subjects with the variability across multiple presentations of a video to the same subject, we performed two experiments. In the first experiment ("reference condition"), 56 subjects watched 6 high-resolution video clips of natural outdoor scenes (1280x720 pixels, 30 Hz, 20s duration). In the second experiment ("repetitive condition"), 10 subjects watched each of these movies 10 times (in two blocks of 5 on consecutive days).

For data analysis, we first discarded all trials where more than 5 percent of gaze samples were invalid, e.g. due to excessive blinks. We then adjusted the number of data sets so that for each movie, we had the same number of data sets (47-52) for both the reference and the repetitive condition.

For each frame, we employed a clustering algorithm on all fixations to obtain a variable number of circular regions of 2.4 degrees diameter that had a high fixation probability.

In frames with "hot spots", a single such region can account for up to 60% of all fixations. On average and depending on the video, 7 to 18 regions are required to predict 60% of fixations. In the repetitive condition, the variability is slightly lower (7-16 regions). This difference is significant for 4 out of the 6 video clips.

To explore how much variability can be reduced using conventional film-making techniques, we also conducted a similar experiment with professionally cut action movie trailers. Here, eye movements tend to cluster in only a very few, but larger clusters. Using a cluster diameter of 4.8 degrees, only 2-3 clusters can predict 60% of all fixations (as opposed to 5-10 such clusters for natural videos).

We can conclude that eye movements on dynamic natural scenes cluster in only a small number of regions with high saliency. The variability across multiple presentations is lower than across multiple subjects, albeit by only a small margin. Professional camera work can reduce such variability to a large extent.