Scan Path and Movie Trailers for Implicit Annotation of Videos

Pallavi Raiturkar *, Andrew Lee, and Eakta Jain

University of Florida

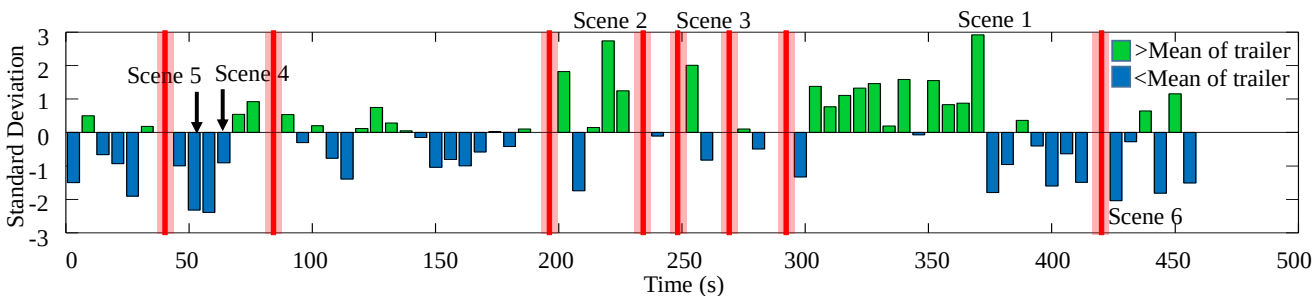

Figure 1: *Top: The red bounding boxes mark the excerpts from the video that were also in its official trailer. We compute the total scan path for seven subjects in the trailer excerpt. The mean and standard deviation across all subjects is considered to be a model of scan paths on the most interesting portions of a video. The green bars mark excerpts from the video that have a larger total scan path than the model average, and the blue bars mark excerpts that have lesser total scan path than the model average.*

Abstract

Affective annotation of videos is important for video understanding, ranking, retrieval, and summarization. We present an approach that uses excerpts that appeared in the official trailers of movies, as training data. Total scan path is computed as a metric for emotional arousal, based on previous eye tracking research. Arousal level on trailer excerpts is modeled as a Gaussian distribution, and signed distance from the mean of this distribution is used to separate out exemplars of high and low emotional arousal in movies.

Keywords: video understanding, eyetracking, scan path analysis

Concepts: •**Information systems** → *Multimedia and multimodal retrieval;* •**Computing methodologies** → *Perception;*

1 Introduction

Affective annotation of videos is an important problem, useful in video understanding and summarization. However, predicting regions of high emotional arousal is difficult because the components that go into creating an emotional response in a viewer are complex. Because self-reported emotional arousal by users is not reliable, researchers have started looking toward implicit annotation methods for videos. Previous work has shown that the total scan path is higher for emotionally arousing images when compared to neutral images [Bradley et al. 2011]. We hypothesize that eye tracking data together with official movie trailers could be used to create a training set for affective video understanding. Total saccade length is computed as a metric for emotional arousal. Arousal level on trailer excerpts is modeled as a Gaussian distribution, and signed distance from the mean of this distribution is used to separate out exemplars of high and low emotional arousal.

*e-mail:pallaviraiturkar@ufl.edu

SAP '16, July 22-23, 2016, Anaheim, CA, USA
ISBN: 978-1-4503-4383-1/16/07
DOI: http://dx.doi.org/10.1145/2931002.2948723

2 Method and Discussion

We selected approximately 15 minute clips from four movies as stimuli. We collected eyetracking data from 7 participants as they watched the videos, using a SensoMotoric eyetracker. Recorded gaze points were converted into fixations and saccades using the Velocity Threshold algorithm. We accumulated the timestamps of all frames that were common between the official trailers of the movies, and the clips shown to the participants. A 6 second window centered around these timestamps was denoted as a trailer excerpt. We divided the rest of the video clip into windows of 6 seconds (non-trailer excerpts). We calculated the total scan path, or, sum of all saccade lengths, for each of the trailer excerpts across all subjects. The mean and standard deviation of this distribution was computed. For each non-trailer excerpt, the signed distance of the total scan path from the trailer mean was calculated. We normalized the data by dividing it by the standard deviation.

Figure 1 shows the mean total scan path in these excerpts, and their distance from the trailer mean for one of our example videos, *Children of Men* (acquired from YouTube; movie id = 3QgOUWs-FeKg, trailer id = 2VT2apoX90o). We selected 6 excerpts, the 3 furthermost from the trailer mean in the positive direction (marked in green), and the 3 furthermost from the trailer mean in the negative direction (marked in blue). The excerpts were ranked in order of the distance of total scan path from the trailer mean. Scene 2 and Scene 3 show a huge group of people attacking passengers in the car, subsequently shooting a woman dead. The higher scan path in these exciting scenes is consistent with previous work on images. Similarly, Scene 4, 5 and 6 are relatively low arousing, with mean scan path lower than the trailer mean. These scenes involve introduction of characters, and there is nothing worth "standing up and taking notice". Future work could evaluate and improve the accuracy of this model, by using features apart from total scan path, such as average fixation duration, pupil dilation, etc. We could also analyse the use of cinematic tools such as camera movement, scene changes, close-ups and wide shots.

References

BRADLEY, M. M., HOUBOVA, P., MICCOLI, L., COSTA, V. D., AND LANG, P. J. 2011. Scan patterns when viewing natural scenes: Emotion, complexity, and repetition. *Psychophysiology 48*, 11, 1544–1553.