Novelty-based Spatiotemporal Saliency Detection for Prediction of Gaze in Egocentric Video

The automated analysis of video captured from a first-person perspective has gained increased interest since the advent of marketed miniaturized wearable cameras. With this a person is taking visual measurements about the world in a sequence of fixations which contain relevant information about the most salient parts of the environment and the goals of the actor. We present a novel model for gaze prediction in egocentric video based on the spatiotemporal visual information captured from the wearer's camera, specifically extended using a subjective function of surprise by means of motion memory, referring to the human aspect of visual attention. Spatiotemporal saliency detection is computed in a bioinspired framework using a superposition of superpixel- and contrast based conspicuity maps as well as an optical flow based motion saliency map. Motion is further processed into a motion novelty map that is constructed by a comparison between most recent motion information with an exponentially decreasing memory of motion information. The innovative motion novelty map is experienced to be able to provide a significant increase in the performance of gaze prediction. Experimental results are gained from egocentric videos using eye-tracking glasses in a natural shopping task and prove a 6.48% increase in the mean saliency at a fixation in terms of a measure of mimicking human attention.