Connecting Gaze, Scene, and Attention: Generalized Attention Estimation via Joint modeling of Gaze and Scene Saliency

Eunji Chong, Nataniel Ruiz, Yongxin Wang, Yun Zhang, Agata Rozga, and James M. Rehg.

Our Approach

- We propose the new problem of “generalized attention estimation” and design a system that can model the visual attention of subjects in unconstrained scenarios which works across most natural scenarios.
- We exploit three public datasets that have been originally collected for different tasks to solve this problem.

Method

- Input = full scene image, a person’s face location whose visual attention we want to predict, and the close-up face image.
- Scene and face images go through separate convolutional layers in such a way that (a) (b) and (c) contribute to saliency, and (b) and (d) contribute to gaze angle prediction. In the last layer, the final feature vectors for two tasks are combined to estimate how likely the person is actually fixating at a gaze target in the frame.
- Loss = Cross Entropy + Euclidean + Project-and-Compare.

References