A Biological Plausible Algorithm for Video Saliency Detection

Zhaoji Zhang[1,2], Ya-tang Li[1,*]

[1] *Chinese Institute for Brain Research, Beijing (CIBR), Beijing, 102206, China*
[2] *College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China*

*Corresponding author
E-mail: yatangli@cibr.ac.cn

Abstract: **Objective** Visual saliency plays an important role in scene-driven, bottom-up attention. However, it remains unclear how the brain encodes visual saliency. Decades ago, Itti et al. proposed a brain-inspired visual attention model that generates saliency maps from static images. Yet, real-world visual perception relies on continuous streaming of visual input, where motion-based saliency is essential for guiding visual attention. Here, we propose a model inspired by this earlier work to generate continuous saliency maps from video input. **Methods** We first implemented the original model with improved saliency detection to generate static saliency map. We then introduced a retina-inspired motion detector to extract motion information and generate motion saliency maps. These static and motion saliency maps were then fused to create master saliency maps. Finally, focuses of attention (FOAs) were determined by a greedy multiple Gaussian-peak fitting algorithm that balance winner-take-all and maximal information selection. **Results** Our model effectively detected salient regions in video streams with millisecond-level latency. Performance was assessed by comparing the generated FOAs with human attention data from the DHF1K dataset. The model outperformed all traditional models and even a few deep-learning network. In videos where attention is primarily driven by bottom-up cues, our model matched the average performance of top deep-learning models, while offering significantly faster processing and requiring far fewer parameters. **Conclusion** Our biologically plausible computer vision model offers an efficient solution for video saliency detection, shedding light on the neural encoding of bottom-up saliency in the brain.

Keywords: video saliency; bottom-up attention; feature extraction, computer vision; visual search