

Saliency Maps for Eye Movement Prediction on Dynamic Scenes

Martin Böhme, Michael Dorr, Thomas Martinetz, and Erhardt Barth

Institute for Neuro- and Bioinformatics
University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

Eye movements have been shown to be affected both by top-down factors (observer and task-dependent strategies) and bottom-up factors (static or dynamic characteristics of the scene). The bottom-up component of eye movement control is often modelled using so-called saliency maps that assign a certain amount of saliency (i.e. probability of being chosen as a fixation target) to each point in the scene. Various approaches have been proposed for constructing saliency maps from image data, both for static and dynamic scenes (e.g. [1, 2]).

We investigate how well eye movements on dynamic scenes can be predicted using saliency maps based on the concept of intrinsic dimensionality [3], i.e. the number of locally non-constant dimensions in the time-varying image signal $f(x, y, t)$; for static images, intrinsic dimensionality has been shown to be a plausible model for biological vision processes [4].

We also introduce the concept of the *empirical saliency map*, which is computed from the actual eye movements of test subjects and gives us an idea of what the best results are that we can expect from predictions based on saliency maps.

To assess how well the saliency maps predict eye movements, we extracted a certain number of candidate locations in each frame and compared these locations with the eye movements made by test subjects on 18 high-resolution real-world video sequences.

Our results show that it is unrealistic to expect accurate predictions from a single candidate location, even if that location is based on observers' actual eye movements (i.e. an empirical saliency map). However, if five to ten candidate locations are used, there is a high probability that one of them will be close to the attended location. This holds for both the empirical and analytical saliency maps, though the performance of the latter still lags behind the ideal embodied by the former. We conjecture that top-down processes in the brain select one of the candidate locations chosen using bottom-up image properties.

References

- [1] E Barth, J Drewes, and T Martinetz. Dynamic predictions of tracked gaze. In *Seventh International Symposium on Signal Processing and its Applications*, Paris, 2003. Special Session on Foveated Vision in Image and Video Processing.
- [2] L Itti. Models of bottom-up attention and saliency. In L Itti, G Rees, and J K Tsotsos, editors, *Neurobiology of Attention*, pages 576–582. Elsevier, San Diego, CA, 2005.
- [3] C Zetsche and E Barth. Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research*, 30:1111–1117, 1990.
- [4] E Barth and A B Watson. A geometric framework for nonlinear visual coding. *Optics Express*, 7:155–185, 2000.