Putting Saliency in its Place

John K. Tsotsos and Yulia Kotseruba
Dept. of Electrical Engineering and Computer Science
York University, Toronto, Ont. Canada

The development of a model of human vision is dependent on the choice of which experimental observations act as its constraints (Tsotsos 2014). Here, we examine the constraints that have gone into the current conceptualization of saliency models and their role within the visual process as a whole. The dominant belief, quite entrenched in both computer and biological vision communities, regarding visual processing is that much can be done within a single data-driven feedforward pass through the visual system. This is supported by broadly recognized experiments (Potter & Faulconer 1975; Thorpe et al. 1996), recently strengthened by Potter et al. (2014). Within this single feedforward pass, the original definition of saliency placed it in the role of an early gating mechanism (Koch & Ullman 1985), as a computational counterpart to selection from the master map of locations in Feature Integration Theory (Treisman & Gelade 1980). The most conspicuous location is selected and its image characteristics routed to higher levels for further analysis.

This experimental work is not disputed; however, when taken together and critically examined, problems arise. First, we took the actual image sets used by Potter and Thorpe, computed saliency maps using several leading algorithms, and found that the predicted first fixation or region of interest seem unrelated to the task posed to subjects. Second, we examined saliency algorithms with respect to their computational time requirements and found a mismatch between time available due to stimulus duration and time needed to support lateral computational mechanisms. Even when considered in light of the many incarnations of vision architectures that involve saliency, the inescapable conclusion is that at least one of the following is true: a) we just don't know how to compute a saliency map that provides guidance for categorization as part of a single feedforward pass; or, b) salience computation and subsequent early gating cannot be part of the fast categorization process. These conclusions are independent of other unresolved issues in saliency computation (Bruce et al. 2015). This is not to say that early selection plays no useful role in machine vision; it can help tame computational load problems. However, our results suggest that the case made to date for an early gating role in biological vision is extremely weak.

Towards the end of placing saliency in its proper place, the **S**elective **T**uning **A**ttentive **R**eference model (STAR) is previewed briefly (Tsotsos 2011, Tsotsos & Kruijne 2014, with saliency playing multiple roles. The first is the common stimulus-driven local feature conspicuity representation (*stimulus-based attentional push)*, but here, restricted to the visual periphery and participating in decisions to change fixation overtly for reasons of surprise, novelty and exploration. We base its computation on the AIM framework (Bruce & Tsotsos 2009). Another is an *object-centred conspicuity* to drive central visual field fixation changes. These changes may be covert or short-range overt, and are made to examine object components for purposes such as description, comparison or discrimination as well as pursuit. The central-peripheral distinction is not due solely to the anisotropy of retinal receptor distribution, but also required to solve the boundary problem present in any layered hierarchical representation (Tsotsos 2011). A third representation is that of *task-specific attentional pull*. This represents priority or urgency (conspicuity among unfinished components in the task domain) to attend a particular location, feature or object related to current task. Cognitive Programs lay out the temporal and causal sequence of operations comprising a task (Tsotsos & Kruijne 2014), working memory stores completed task elements, and the disparity at a given time between program and memory lead to task element priority. STAR integrates proposals for the executive controller, working and short-term memory components, attentional mechanisms, eye movements, the visual processing hierarchy and the communication among these that is hoped to enable flexible and generalizable visual task execution.

Acknowledgements We thank Molly Potter, Carl Hagmann, Simon Thorpe and Nadège Macé for sharing their stimulus sets with us. This research was supported by a grant from the Office for Naval Research (N00178-14-P-4312) and by the Canada Research Chairs Program. Comments from Neil Bruce, Calden Wloka and Brad Wyble greatly improved the presentation.

References

Bruce, N.D.B., Tsotsos, J.K. (2009). Saliency, Attention, and Visual Search: An Information Theoretic Approach, *J. of Vision*, 9:3, p1-24.
Bruce, N.D.B., Wloka, C., Frosst, N., Rahman, S., Tsotsos, J.K. (2015). On computational modeling of visual saliency: Examining what's right, and what's left. *Vision Research*.
Koch, C., Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry, *Human Neurobiology* 4, 219-227.
Potter, M.C., Faulconer, B.A. (1975). Time to understand pictures and words, *Nature* 253, 437 – 438.
Potter, M.C., Wyble, B., Hagmann, C.E., McCourt, E.S. (2014). Detecting meaning in RSVP at 13ms per picture, *Attention, Perception, & Psychophysics*, 76(2), 270-279.
Thorpe, S., Fize, D., Marlot, C. (1996). Speed of processing in the human visual system, *Nature* 381 (6582), 520 - 522.
Treisman, A., Gelade, G. (1980). A feature integration theory of attention, *Cognitive Psychology* 12, 97 – 136.
Tsotsos, J.K. (2011). **A Computational Perspective on Visual Attention**, The MIT Press.
Tsotsos, J.K., Kruijne W. (2014). Cognitive programs: Software for attention's executive, *Frontiers in Psychology: Cognition* 5:1260.
Tsotsos, J.K. (2014). It's All About the Constraints, *Current Biology* 24(18), pR854–R858, 22 September.