Combining bottom-up and top-down attentional influences

Vidhya Navalpakkam
Department of Computer Science
USC, Los Angeles
navalpak@usc.edu

Laurent Itti
Department of Computer Science
USC, Los Angeles
itti@usc.edu

Visual attention to salient and relevant scene regions is crucial for an animal's survival in the natural world. It is guided by a complex interplay of at least two factors – image-driven, bottom-up salience [1] and knowledge-driven, top-down guidance [2, 3]. For instance, a ripe red fruit among green leaves captures visual attention due to its bottom-up salience, while a non-salient camouflaged predator is detected through top-down guidance to known predator locations and features. Although both bottom-up and top-down factors are important for guiding visual attention, most existing models and theories are either purely top-down [4] or bottom-up [5, 6]. Here, we present a combined model of bottom-up and top-down visual attention.

Our proposed model first computes the naive, bottom-up salience of every scene location for different local visual features (e.g., different colors, orientations and intensities) at multiple spatial scales in a manner described in [6]. Next, the top-down component uses learnt statistical knowledge of the local features of the target and distracting clutter, to optimize the relative weights of the bottom-up maps such that the overall salience of the target is maximized relative to the surrounding clutter. Such optimization renders the target more salient than the distractors, thereby maximizing target detection speed [7].

Finding the optimal top-down weights that maximize the target's salience relative to

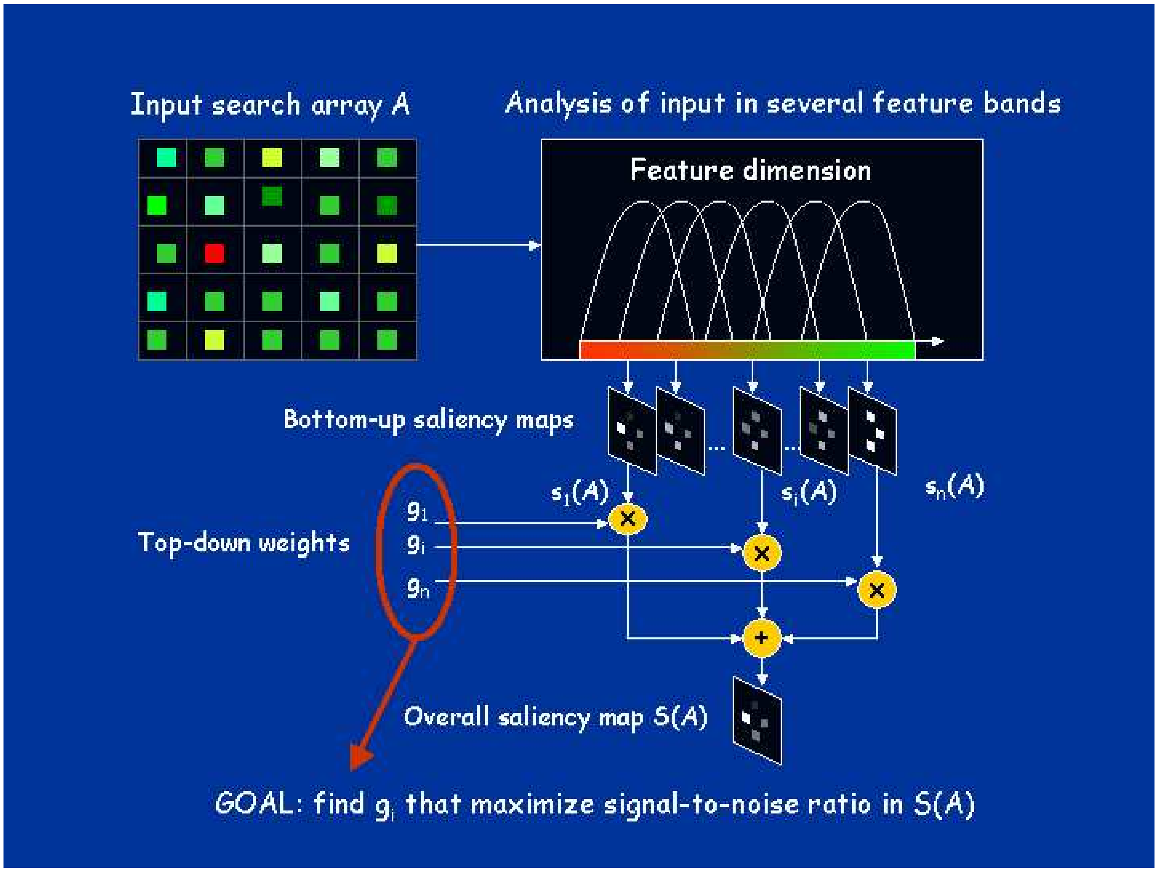

Figure 1: Overview of our model. The incoming visual scene A is analyzed in several feature dimensions such as color, orientation, texture. For simplicity, only color is shown here. We assume that each feature dimension is encoded by a population of neurons with broad and overlapping bell-shaped tuning curves. Within each dimension, bottom-up saliency maps ($s_1(A)...s_n(A)$) are computed for different feature values and combined in a weighted linear manner to form the overall saliency map ($S(A)$) for that dimension. Given this model, our theory suggests the optimal set of top-down weights ($g_1...g_n$) on bottom-up saliency maps such that the target's salience relative to the background is maximized, i.e., the signal-to-noise ratio is maximized

the distracting background is challenging as salience computations involve complex, non-linear spatial interactions and depend on several random variables such as the statistical distribution of target and distractor features, the spatial configuration or arrangement of target and distractors in the scene, photoreceptor or neural noise in response to the visual stimulus. In an earlier paper [8], we formalized the different bottom-up and top-down factors influencing salience, and derived a theory of optimal top-down biasing (i.e., weighting) of bottom-up ups. Our theoretical result is simple and intuitive, suggesting that each bottom-up map must be weighted according to the $\mathcal{SNR}$ contained in it. Mathematically,

$$g_i = \frac{\mathcal{SNR}_i}{\frac{\sum_i \mathcal{SNR}_i}{n}} \tag{1}$$

where g_i is the top-down weight on the i^{th} bottom-up saliency map ($i \in \{1...n\}$), $\mathcal{SNR}$ is the signal-to-noise ratio in the overall saliency map for that dimension, while $\mathcal{SNR}_i$ represents the signal-to-noise ratio contained in the i^{th} saliency map within that dimension.

To verify whether the optimal theory can account for existing behavioral and physiological data in visual search literature, we tested the predictions of the optimal theory on simulated networks of neurons. The results of our simulation are consistent with several bottom-up effects such as pop-out vs. conjunction search [9], distractor heterogeneity [10], target-distractor discriminability [10–13] and linear separabilty [14, 15], as well as top-down effects such as uncertainty in target's features [3, 7, 16], role of priming [17, 18], target enhancement [19], distractor suppression [19, 20], linear separabilty effect [21]. Thus the theory successfully accounts for most reported effects in visual search literature.

We further tested the theory by evaluating its performance on natural images. Examples of saliency maps produced by the optimal theory are shown in comparison to those generated by the naive bottom-up saliency model [6].

To summarise, by combining top-down, knowledge-driven and bottom-up, image-driven approaches, we account for a large body of visual search literature. Systematic testing on natural images reveals that a model that combines both top-down and bottom-up effects performs significantly better than a model that is purely bottom-up. The promising results of our model suggest that the human visual system may guide attention by applying optimal top-down weights on bottom-up saliency maps, so that the desired target objects may be detected quickly in distracting backgrounds.

Acknowledgements

This research was supported by the National Geospatial-Intelligence Agency, the National Science Foundation, and the Human Frontier Science Program. The authors affirm that the views expressed herein are solely their own, and do not represent the views of the United States government or any agency thereof.

References

[1] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, Mar 2001.

[2] A. Yarbus. *Eye Movements and Vision.* Plenum Press, New York, 1967.

[3] Jeremy M Wolfe, Todd S Horowitz, Naomi Kenner, Megan Hyle, and Nina Vasan. How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Res*, 44(12):1411–1426, Jun 2004.

[4] R P Rao, G Zelinsky, M Hayhoe, and D H Ballard. Eye movements in iconic visual search. *Vision Research*, 42(11):1447–1463, Nov 2002.

[5] J K Tsotsos, S M Culhane, W Y K Wai, Y H Lai, N Davis, and F Nuflo. Modeling visual-attention via selective tuning. *Artificial Intelligence*, 78(1-2):507–45, 1995.

[6] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, May 2000.

[7] J. M. Wolfe, S. J. Butcher, and M. Hyle. Changing your mind: On the contributions

of top-down and bottom-up guidance in visual search for feature singletons. *J Exp Psychol Hum Percept Perform*, 29(2):483–502, 2003.

[8] V. Navalpakkam and L. Itti. Optimal cue selection strategy. In *Advances in Neural Information Processing Systems, Vol. 19 (NIPS*2005)*, pages 1–8, Cambridge, MA, 2006. MIT Press.

[9] A. Treisman and G. Gelade. A feature integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[10] J Duncan and G W Humphreys. Visual search and stimulus similarity. *Psychological Rev*, 96:433–458, 1989.

[11] H. Pashler. Target-distractor discriminability in visual search. *Percept Psychophys*, 41(4):385–392, Apr 1987.

[12] A. L. Nagy and R. R. Sanchez. Critical color differences determined with a visual search task. *Journal of the Optical Society of America A 7*, 7:1209–1217, 1990.

[13] A. Treisman. Search, similarity, and integration of features between and within dimensions. *J Exp Psychol Hum Percept Perform*, 17(3):652–676, Aug 1991.

[14] M. D'Zmura. Color in visual search. *Vision Research 31*, 6:951–966, 1991.

[15] B. Bauer, P. Jolicoeur, and W. B. Cowan. Visual search for colour targets that are or are not linearly-separable from distractors. *Vision Research 36*, 10:1439–1465, 1996.

[16] Timothy J Vickery, Li-Wei King, and Yuhong Jiang. Setting up the target template in visual search. *J Vis*, 5(1):81–92, Feb 2005.

[17] R. M. Shiffrin and W. Schneider. Controlled and automatic human information processing: Ii. perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84:127–190, 1977.

[18] V. Maljkovic and K. Nakayama. Priming of pop-out: I. role of features. *Mem Cognit*, 22(6):657–672, Nov 1994.

[19] N. P. Bichot and J. D. Schall. Priming in macaque frontal cortex during popout visual search: feature-based facilitation and location-based inhibition of return. *J Neurosci*, 22(11):4675–4685, Jun 2002.

[20] J. J. Braithwaite and G. W. Humphreys. Inhibition and anticipation in visual search: evidence from effects of color foreknowledge on preview search. *Percept Psychophys*, 65(2):213–237, Feb 2003.

[21] J. Hodsoll and G. W. Humphreys. Driving attention with the top down: the relative contribution of target templates to the linear separability effect in the size dimension. *Percept Psychophys*, 63(5):918–926, Jul 2001.

Figure 2: Comparison of saliency maps of the naive bottom-up model (second row) vs. our optimal model are shown during search for a phone on a desk (first column), a coke can in a cluttered scene (second column), and a pen in a distracting background (third column). Although the target is not bottom-up salient, prior knowledge of the target and the distracting background (acquired through training) helps in improving the $\mathcal{SNR}$, thereby rendering the target more salient and suppressing noisy activity due to the distractors.