Components of Bottom-Up Gaze Allocation in Natural Scenes

Robert J. Peters (1,2), Asha Iyer (2), Christof Koch (2), Laurent Itti (1)
(1) University of Southern California, Computer Science, Neuroscience
(2) California Institute of Technology, Computation and Neural Systems

http://ilab.usc.edu/rjpeters/pubs/2005_VSS_Peters.pdf

Supported by HHMI, NGA, DCI,
and the ERC program of the NSF

introduction

Eye movements provide an overt measure of the internal allocation of spatial attention.

Here, we use human eye movement recordings to test ***computational models*** of ***bottom-up spatial visual attention***.

method

Eye movement psychophysics:
- 3 image categories, 110 images per category
- 4 subjects, 3 sec. free-viewing per image

Computational models:
- lower bound given by **pixel-shuffled control**
- and **image-shuffled control**
- upper bound given by **inter-observer control**
- saccade location predictions made by
- **baseline salience model**
- **enhanced salience model** includes more
- detailed physiological mechanisms

Analysis metrics:
- quantify the statistical correspondence between
- human saccade targets and the predictions of
- salience models and control models

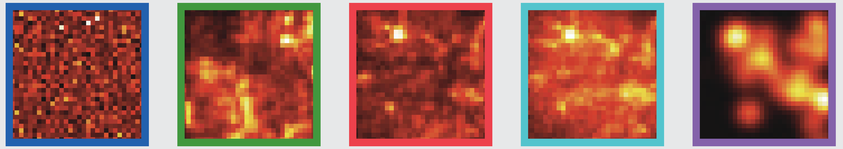
results

- **baseline salience model** predictions are far better than chance levels (**pixel-shuffled** and **image-shuffled** controls)
- physiological interactions among orientation-tuned units (**enhanced salience model**) play a significant role in gaze allocation
- model predictions reach 50-90% of upper bound derived from **inter-observer control**
- strength of bottom-up influences on eye movements decays over time
- "center bias" varies with image category

images

models — *trying to predict human eye movements*

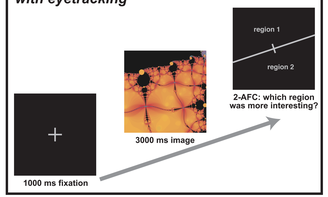
pixel-shuffled control | image-shuffled control | baseline salience | enhanced salience | inter-observer control



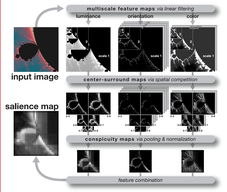
pixel-shuffled control for image A =
- randomly scramble pixels from
- image A's salience map

image-shuffled control for image A =
- take salience map from
- a different image B

free-viewing psychophysics task with eyetracking



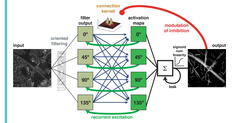
baseline salience model



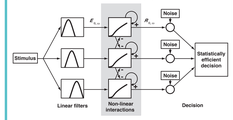
- multiple image features extracted in parallel channels
- features are subject to coarse global competition
- all features combine to give a single salience map

enhanced salience model

long-range orientation interactions for contour facilitation

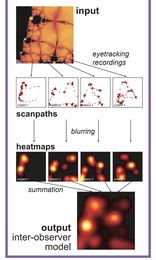


short-range orientation interactions (cross-orientation suppression)

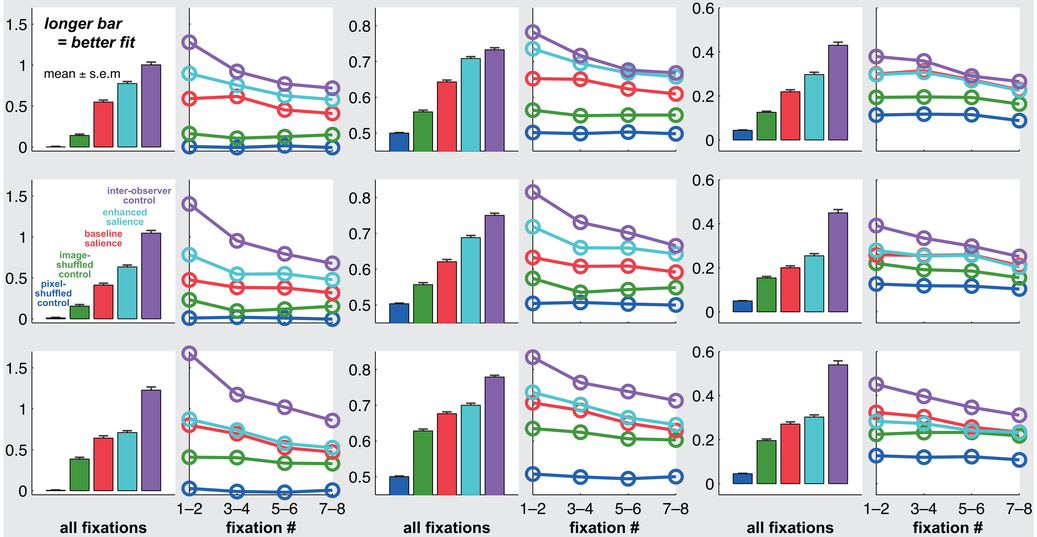


inter-observer control

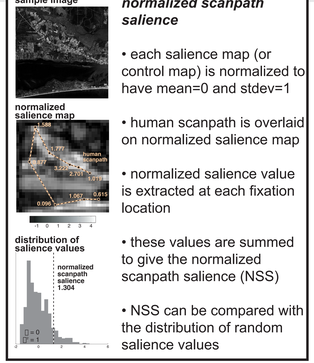
try to predict each subject's saccade locations from the average of the other subjects



metrics — *measuring the performance of the models*



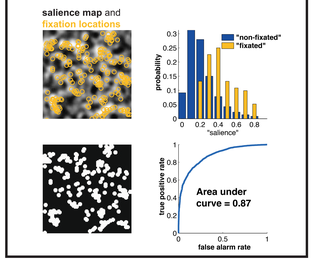
normalized scanpath salience



- each salience map (or control map) is normalized to have mean=0 and stdev=1
- human scanpath is overlaid on normalized salience map
- normalized salience value is extracted at each fixation location
- these values are summed to give the normalized scanpath salience (NSS)
- NSS can be compared with the distribution of random salience values

receiver operating characteristic
- *assuming a simple threshold:*

how well can fixated and non-fixated locations be distinguished?



Kullback-Leibler distance
- *with no assumptions about decision criteria:*

how well can fixated and non-fixated locations be distinguished?

$$KL = 0.5 \cdot \sum_i p_i(\mathbf{f}) \cdot log\left(p_i(\mathbf{f}) / p_i(\mathbf{nf})\right) + 0.5 \cdot \sum_i p_i(\mathbf{nf}) \cdot log\left(p_i(\mathbf{nf}) / p_i(\mathbf{f})\right)$$