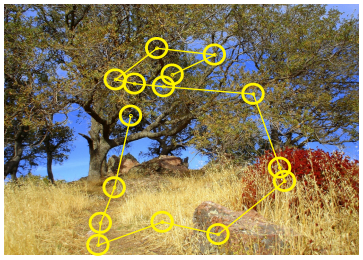
Prediction of Human Eye Fixations

Gert Kootstra, Arco Nederveen and Bart de Boer
Artificial Intelligence, University of Groningen, The Netherlands



Introduction

Humans view scenes with a series of eye fixations. The eye movements are controlled with bottom-up and top-down mechanisms. In this study, we investigate the role of the stimulus. In order to do this, we compare the results of human experiments with computational models of visual attention.

- Which visual features underlie the fixations?
- Can we predict the fixations with bottom-up models?

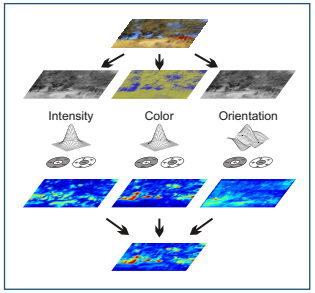
Methods

A. Analysis of fixation points

The saliency of the fixation points on an image is compared with the average saliency of that image.

Saliency measured:

- Intensity entropy
- Color entropy
- Saliency model of Itti *et al.* (1998)



Saliency Model (SM):

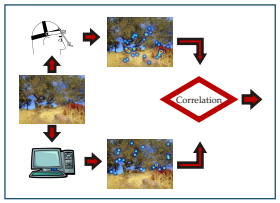
- Intensity, color and orientation
- Locally contrasting points
- Center-surround filters on different scales

B. Predicting human fixations

We compare the human fixations with predictions from region-of-interest detectors.

The region-of-interest detectors:

- Saliency model of Itti *et al.*
- Keypoint selection of SIFT model (Lowe,2004)
  - Consideres only intensity



Measuring the quality of prediction:

- Comparing human and predicted f.d.m. With correlation

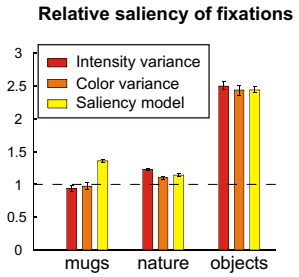
Eye track experiment

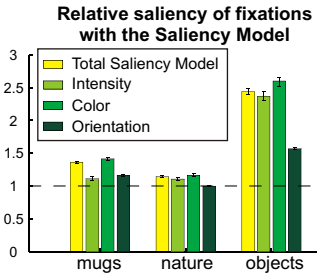
- 43 participants
- 3 sets images (see figure)
- 10 images per set
- Each image is viewed 4 times (including 2 mirrored versions) for 5 seconds
- Free-viewing, no specific task

Results

A. Analysis of fixation points

The plots show the average saliency of the human fixations per image category, relative to the average saliency in the images.



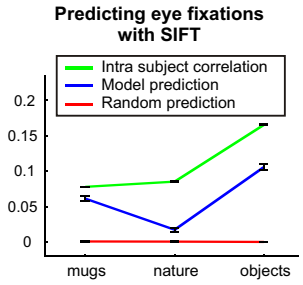


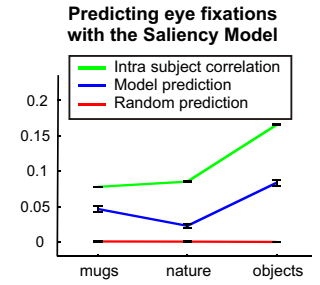
Some observations:

- Overall, the saliency measures of the human fixations is significantly higher than the average saliency.
- Mugs and nature images, low saliency. Reasons:
  - Nature: many salient parts all over the image
  - Mugs: too few salient parts. Fixations on low-salient parts.
- Within SM, color is most present in the human fixations, orientation least.

B. Predicting human fixations

The plots show the result of the prediction with SIFT and the Saliency Model. The random prediction serves as a bottom-line. The intra subject correlation as a top-line.





Some observations:

- Both model predict fixations significantly better than random, and significantly worse than the intra subject correlation.
- Difference between model prediction and intra subject correlation show that participants are interested in parts that cannot be explained by the saliency of the stimulus
- SIFT predicts better than SM (except for nature), and on all categories better than the intensity part of SM. (not shown)
- Similar to section A, fixations on mugs and nature images are harder to predict.

Discussion

- Participants fixate on more salient parts of the images
- Fixation strategies among categories cannot be compared, caused by differences in image structure
- Bottom-up models predict fixations better than random, but worst than the intra subject correlation:
  - Top-down control plays a role
- The mechanism for selecting interest points in SIFT performs better than that of SM

Future work: bottom-up and top-down control

- Hypotheses: fixations are more salient on images with little semantic content (Parkhusrt and Ernst, 2003)
- Experiment: semantic images and transformed versions with no meaning, but the same amount of saliency

Itti, L., Koch, C. & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE PAMI, 20*(20), 1254-1259.
Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Computer Vision 60*(2), 91-110
Parkhurst, D. J. & Ernst, N. (2003). Scene Content Selected by Active Vision. *Spatial Vision, 16*(2), 125-154.