Quantifying the relative influence of photographer bias and viewing strategy on scene viewing

Ali Borji, Dicky N. Sihite, and Laurent Itti
University of Southern California, Computer Science, Neuroscience

http://ilab.usc.edu/publications/doc/Borji_etal11vss.pdf

Supported by the National Science Foundation, the General Motors Corporation, and the Army Research Office

(1) Introduction & Motivation

Saccade distributions while observers freely view natural scenes and videos are often found to be highly biased toward the image center (**center-bias effect**) [1].

Two main sources of center-bias are: 1) **photographer bias** (natural tendency of photographers to place objects of interest near the center) and 2) **viewing strategy** (tendency of subjects to look at the center to extract more information) [2].

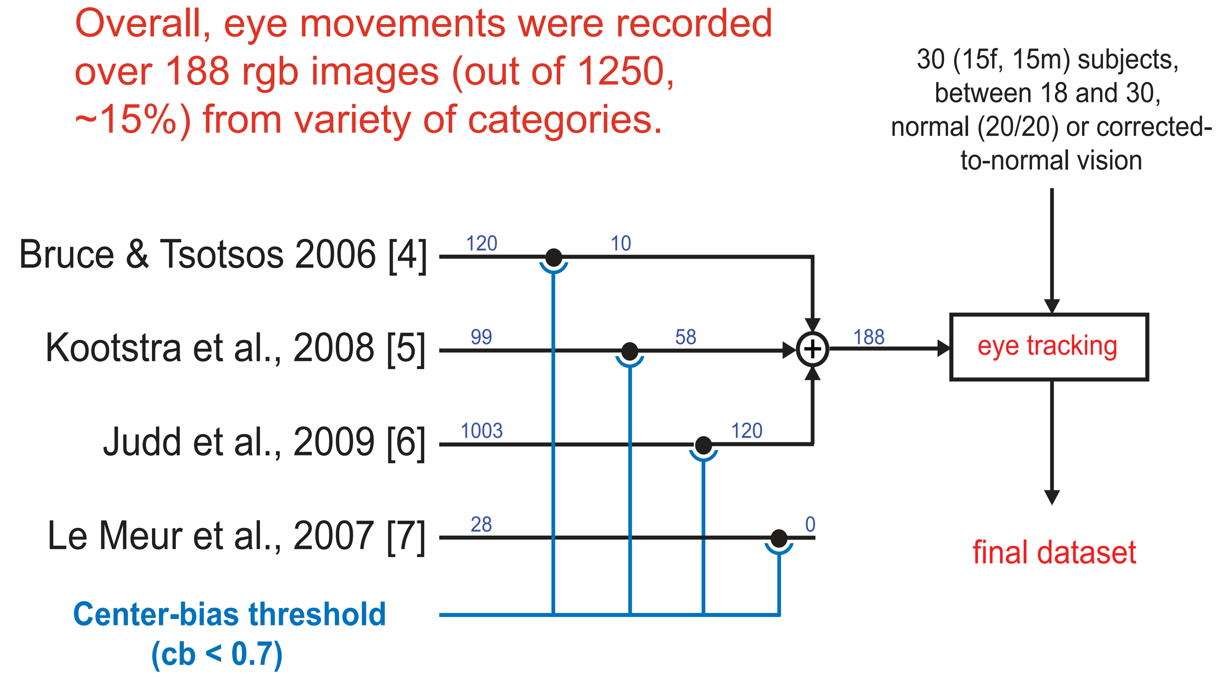
Our quantitative comparison of 30 saliency models over three standard datasets of still images (Bruce & Tsotsos 2006 [4], Kootstra et al., 2008 [5] and Judd et al., 2009 [6]), shows that model rankings do not agree [3]. Interestingly, a trivial central Gaussian blob saliency model outperforms many models in regard to predicting where humans look.

Here, we measure the relative influence of center-bias causes and introduce a less center-biased dataset as a benchmark for fair evaluation of saliency models.

(2) Experimental Procedure

Method, Stimuli & Subjects: Stimuli were selected from four datasets and were shown to new subjects to remove variability.

Overall, eye movements were recorded over 188 rgb images (out of 1250, ~15%) from variety of categories.



Eye Tracking:

Stimuli were presented on a monitor (60 Hz refresh rate), full HD 1024 x 1280 pixels (corresponding to 43.6º x 28º field of view). Participants' heads were stabilized with a chin rest. Eye position was tracked by an EyeLink camera (SR Research), with 1 kHz and gaze position accuracy of 0.5º average from participants' right eye. Eye tracker was calibrated with 5 point calibration at the beginning of the experiment. An image was presented to the subject for 3 seconds followed by a 2 second blank screen in between. Stimuli were rgb images with resolution of 1024 x 1280. Mean screen luminance was 30 cd/m2. Overall, 46429 saccades were recorded.

(3) Evaluated Data Sets

Bruce & Tsotsos [4]

Stimuli: 120 (indoor and outdoor)

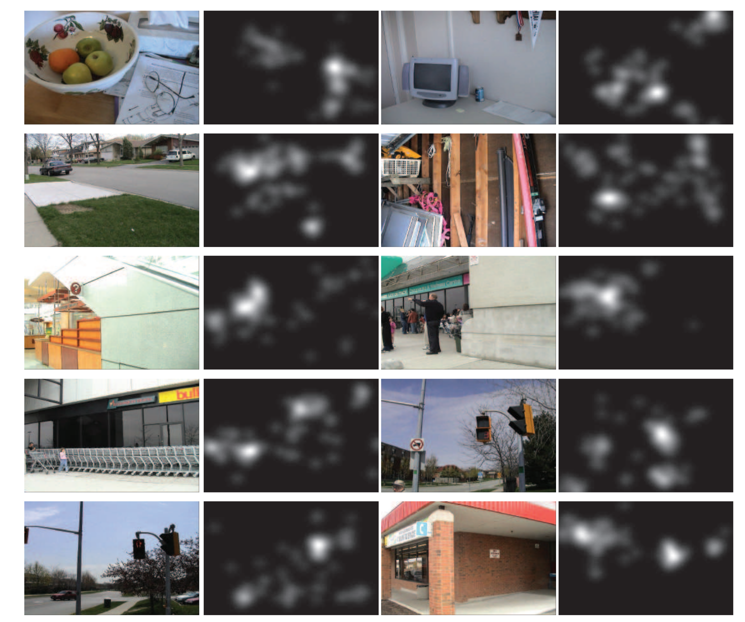
Subjects: 20

Resolution: 681 x 511

Viewing distance: 75 cm

Presentation time: 4 s

www-sop.inria.fr/members/Neil.Bruce



Kootstra et al. [5]

Stimuli: 99 (12 Animals, 12 Automan, 16 Buildings, 20 flowers, 41 natural scenes)

Subjects: 31 (15f, 16m)

Resolution: 1024 x 768

Viewing distance: 70 cm

http://www.csc.kth.se/~kootstra/

Judd et el. [6]

Stimuli: 1003 (779 landscape images and 228 portrait images)

Subjects: 15

Resolution: 1280 × 1024

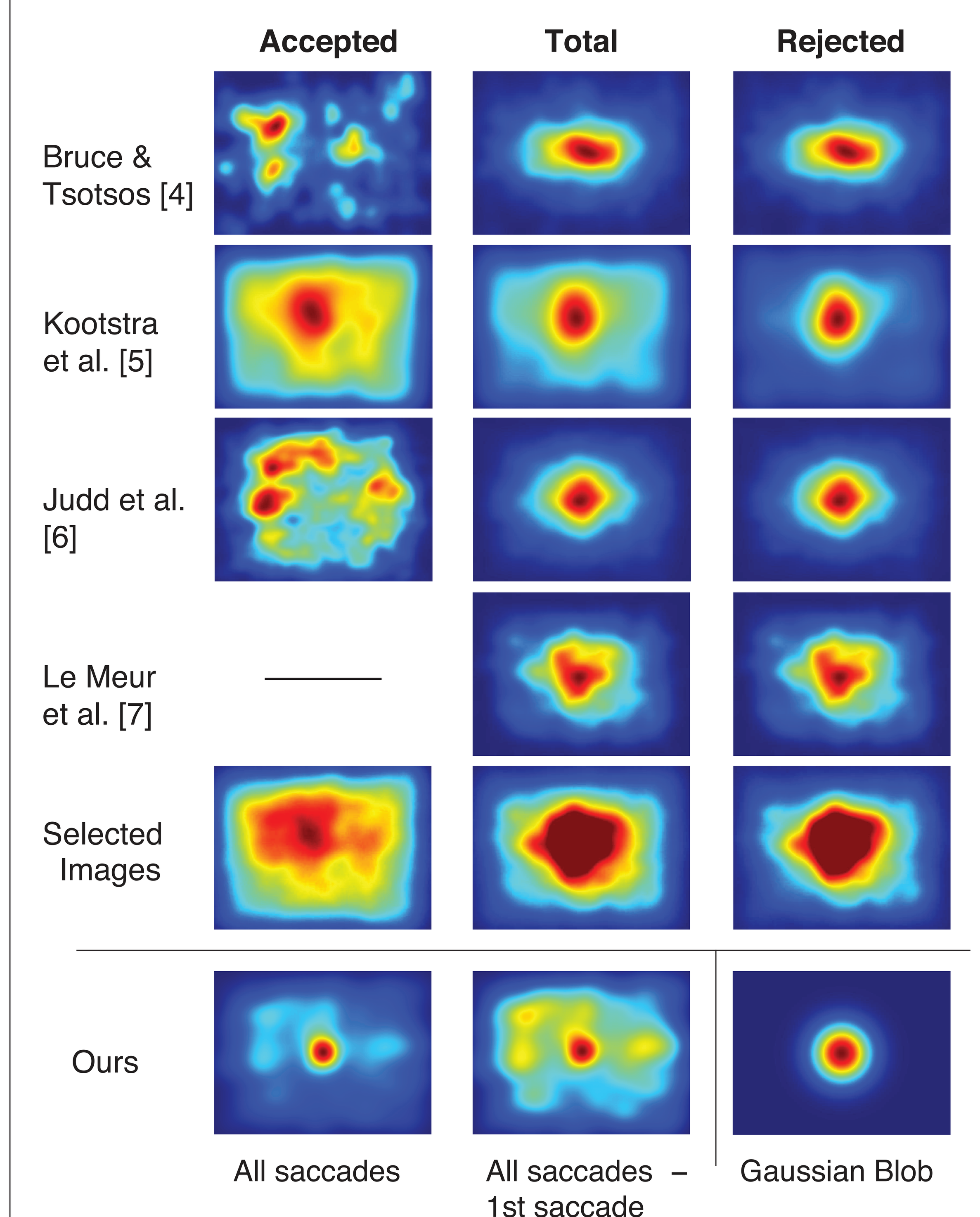
Viewing distance: 61 cm

Presentation time: 3 s

http://people.csail.mit.edu/tjudd

(4) Average Heat Maps

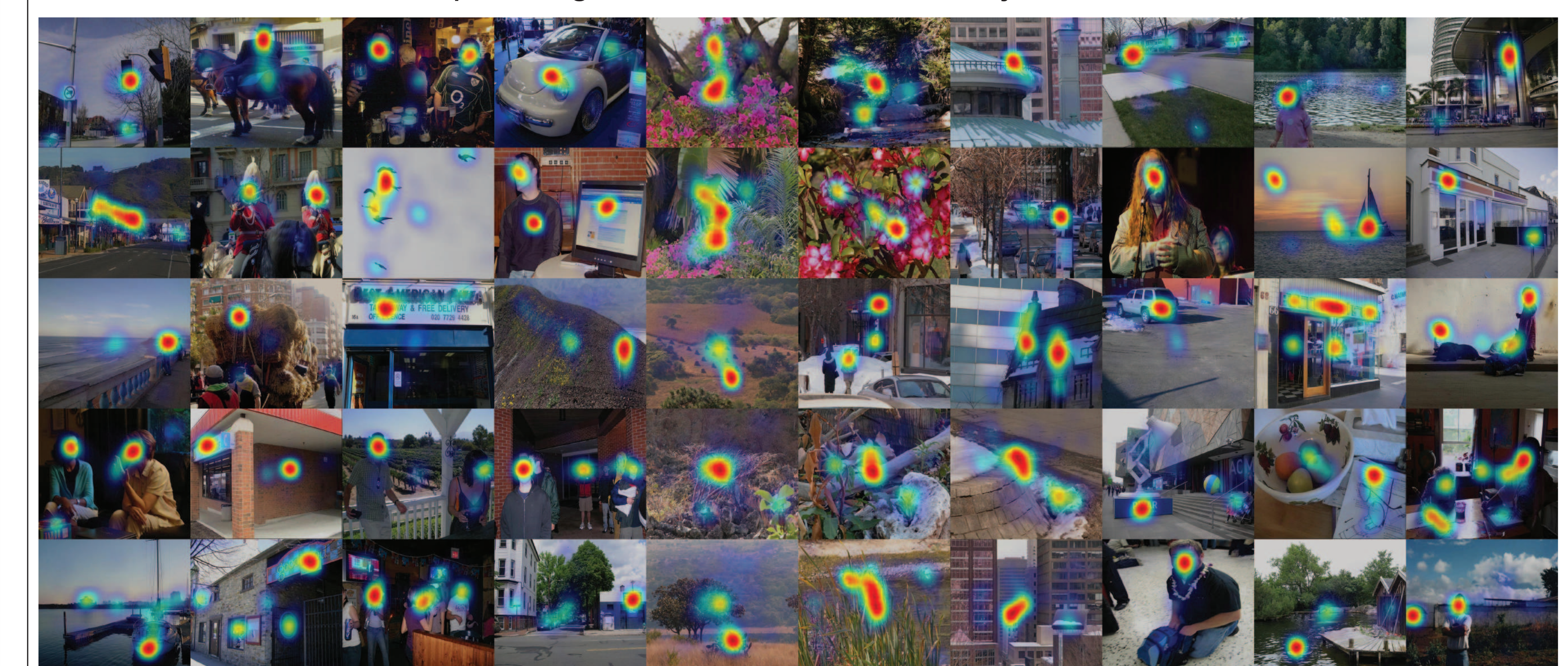
Average heat map of accepted, total and rejected images. The density map of accepted images has a wider spread. Rejected images have a very high center-bias.



(6) Final Selected Stimuli

The center-bias index for **recorded eye movements** over **selected images** and the **Judd** dataset were 0.73 and 0.88, respectively. After removing the first saccade, these values dropped to 0.66 and to 0.84.

Sample images with fixations of all subjects overlaid



Conclusions & References

1) Selected images have less image content/interesting objects at the center. This caused a significant drop in center-bias index. Removing the first saccade further reduces the cb index. Analyzing the images where fixations were at the center, we noticed that either there was an object at the center or there was not much difference between image center and the rest of the scene (e.g. some scenes from nature). Overall, based on our results we conclude that photographer bias is more important than viewing strategy.
2) Widely used datasets are center-biased, while our collected data shows less bias which makes it a suitable benchmark dataset for eye fixation prediction evaluation.

[1] B. W. Tatler, The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. J. Vision, 2007.
[2] P. Tseng, R. Carmi, I. G. M. Cameron, D.P. Munoz, L. Itti, Quantifying center bias of observers in free viewing of dynamic natural scenes, J. Vision, 2009.
[3] A. Borji and L. Itti, submitted.
[4] N.D.B. Bruce, J.K. Tsotsos, Saliency based on information maximization, NIPS, 2005.
[5] G., Kootstra, A. Nederveen, and B. de Boer, Paying attention to symmetry, BMVC, 2008.
[6] T. Judd and K. Ehinger and F. Durand and A. Torralba, Learning to predict where humans look, ICCV, 2009.
[7] O. Le Meur, P. Le Callet and D. Barba, Predicting visual fixations on video based on low-level visual features, Vision Research, 2007.
[8] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE PAMI, 1998.

(5) Results

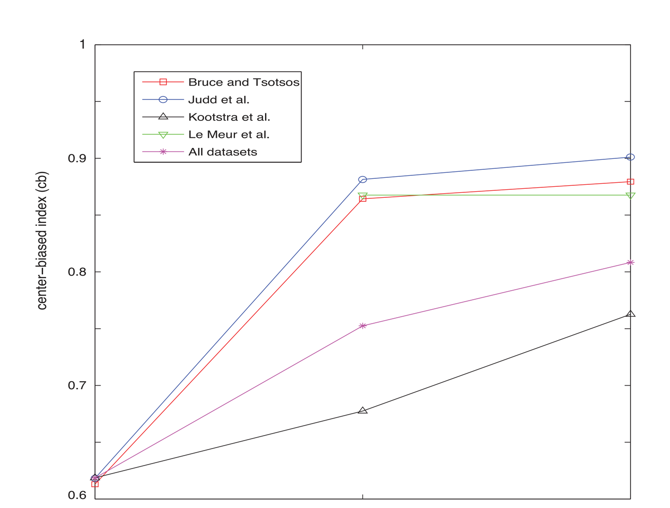
Center-bias (cb) index: The sum and count of heat map values are calculated in growing disks (5 pixel increments) from center to the image border. The cb index was the sum of mean map values of disks with radii smaller than or equal to a threshold divided by sum of all mean maps of all disks:

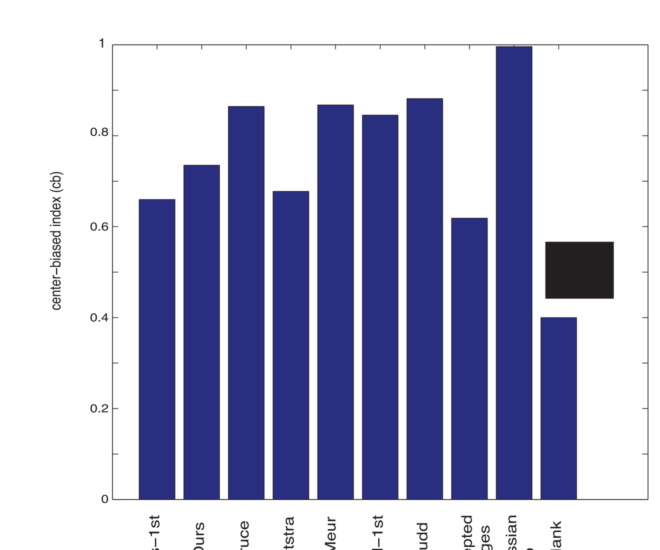
$$S_{R_i} = \sum_{x,y \in R_i} s_{xy}, C_{R_i} = |x,y \in R_i|, M_{R_i} = S_{R_i}/C_{R_i}$$

$$cb = \frac{\sum_{i=1}^{i=r} M_{R_i}}{\sum_i M_{R_i}}$$

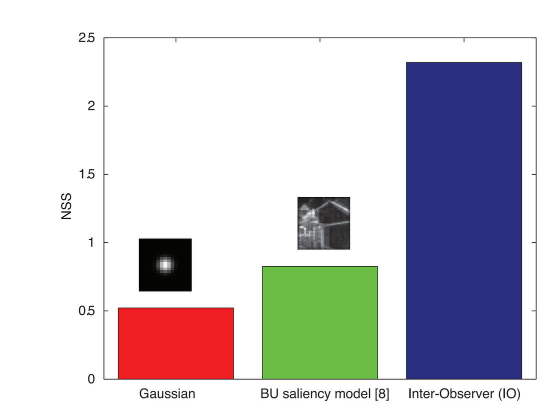
s_{xy} is the saliency of point (x,y) in the heat map of an image. |.| is the set size operator. R_i is the set of (x,y) coordinates that fall in a disk centered at image center with radius 5i. We chose r to be 40% of the distance from center to the corner of image.

The average center-bias index of four datasets, accepted images and recorded eye movements over accepted images.





Eye movement prediction power of **BU saliency model** [8], versus **Gaussian** and **Inter-Observer** models.



Average **entropy** of datasets, over rectangular regions with increasing size from center.