Top Down Attentional Guidance During Visual Search

Ankit Awasthi
Keerti Choudhary

Guide: Prof. Amitabha Mukerjee

Top Down Vs Bottom Up

- Top Down influences

 -> Contexual Guidance

 -> Top Down Object Information

- Bottom Up influences

 -> Sharp Contrast

Top Down Object Information

- A dictionary of features is made using images of the specified object.
- For each positive sample,20 randomly cropped samples are chosen as negative examples.
- A number of weak classifiers(120) based on the features are used to score the image for the presence of that target
- The scores of these classifiers for multiple scales are combined(as in boosting) to give the final score.

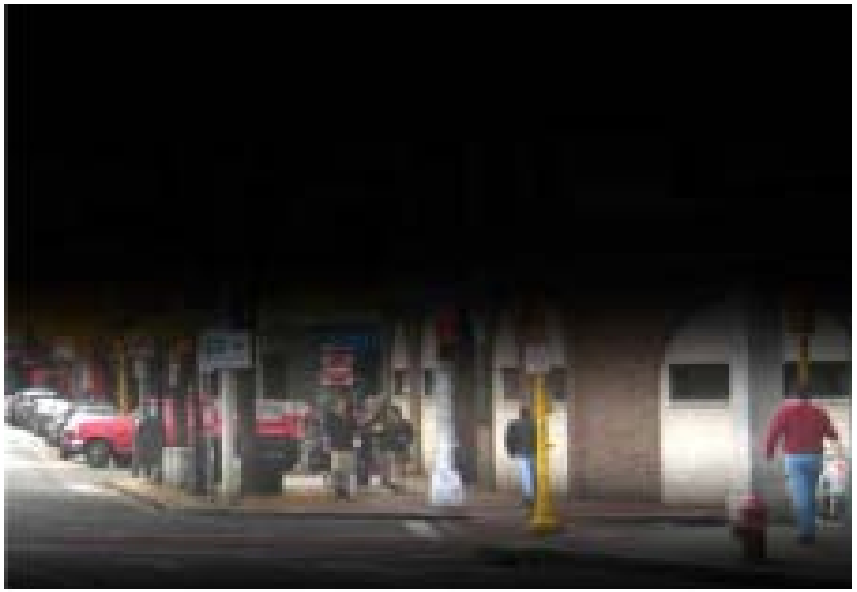
Contexual Guidance

Feature Extraction

- The image is filtered with Gabor filters at 4 scales and 6 orientations
- Each filtered image is down sampled into 4X4 blocks.
- The filtered response is averaged for each of the blocks.
- The resulting feature vector of 384 (16X24) is reduced to 100 dimensions using pca.

Training of Context Model

- For each image we have the global feature vector and corresponding target object location.
- A GMM for the joint probability of the global feature vector and the object location is fit to the training data.



Confidence in Context Model

•Results are pretty good if familiar scenes from familiar perspectives are taken
•Consider the following example(Torralba et al 2009)

Confidence in Context Model

- We have a Gaussian Mixture Model (3 gaussians) for the global feature vectors of the training images.
- If the query image is within some proximity (say 2σ) of any of the gaussian peaks ,we give it a certain weight and much lower weight otherwise.

$k = 0.25$ if $d < \sigma$

$k = 0.15$ if $d < 2\sigma$

$k = 0.1$ otherwise

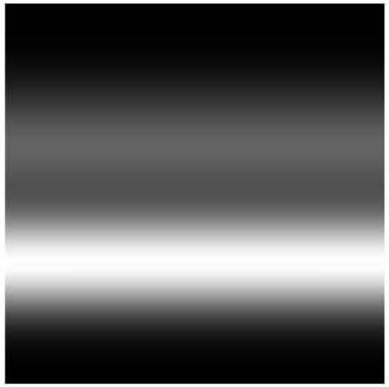
Combining Context and Target Information

- M_c : Context Map
- Mt : Target Based Map
- $M = M_c^{k} * M_t^{(1-k)}$

In [1] (Torralba et al, 2006) for saliency and context based maps k=0.2

Currently we are using k=0.15, confidence in context model yet to be tested.

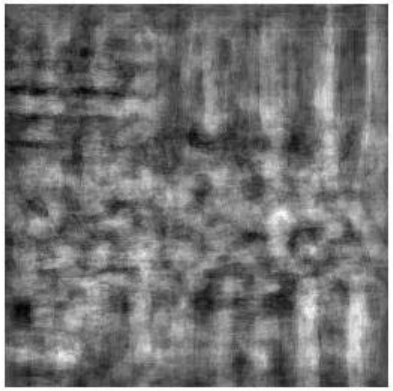
Context Map



Combined Map

Original Image

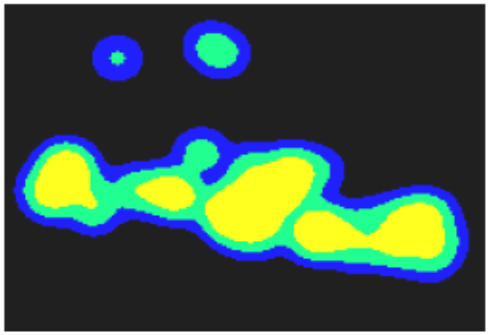
Target Features Map



Consistency Among Humans(Torralba et al, 2006)

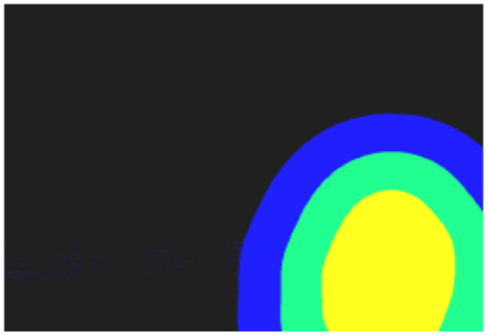
-A-

Region defined by 7 participants

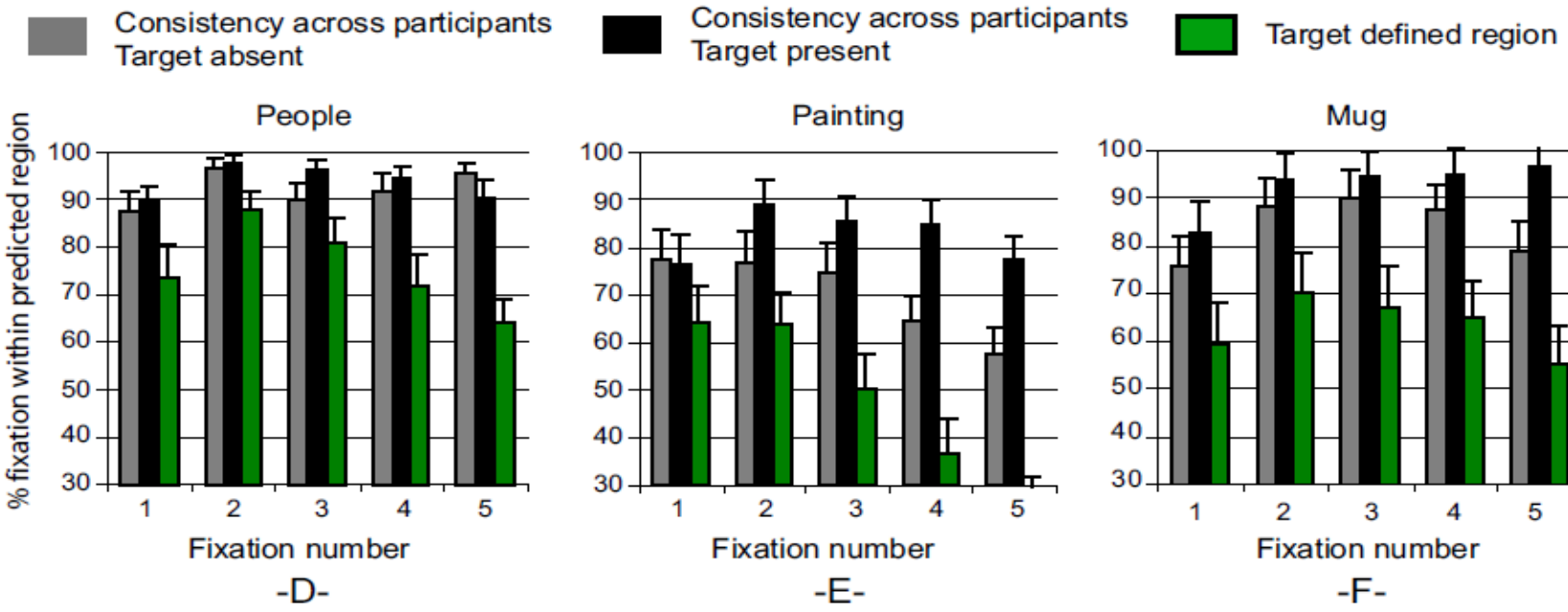


-B-

Region defined by the target



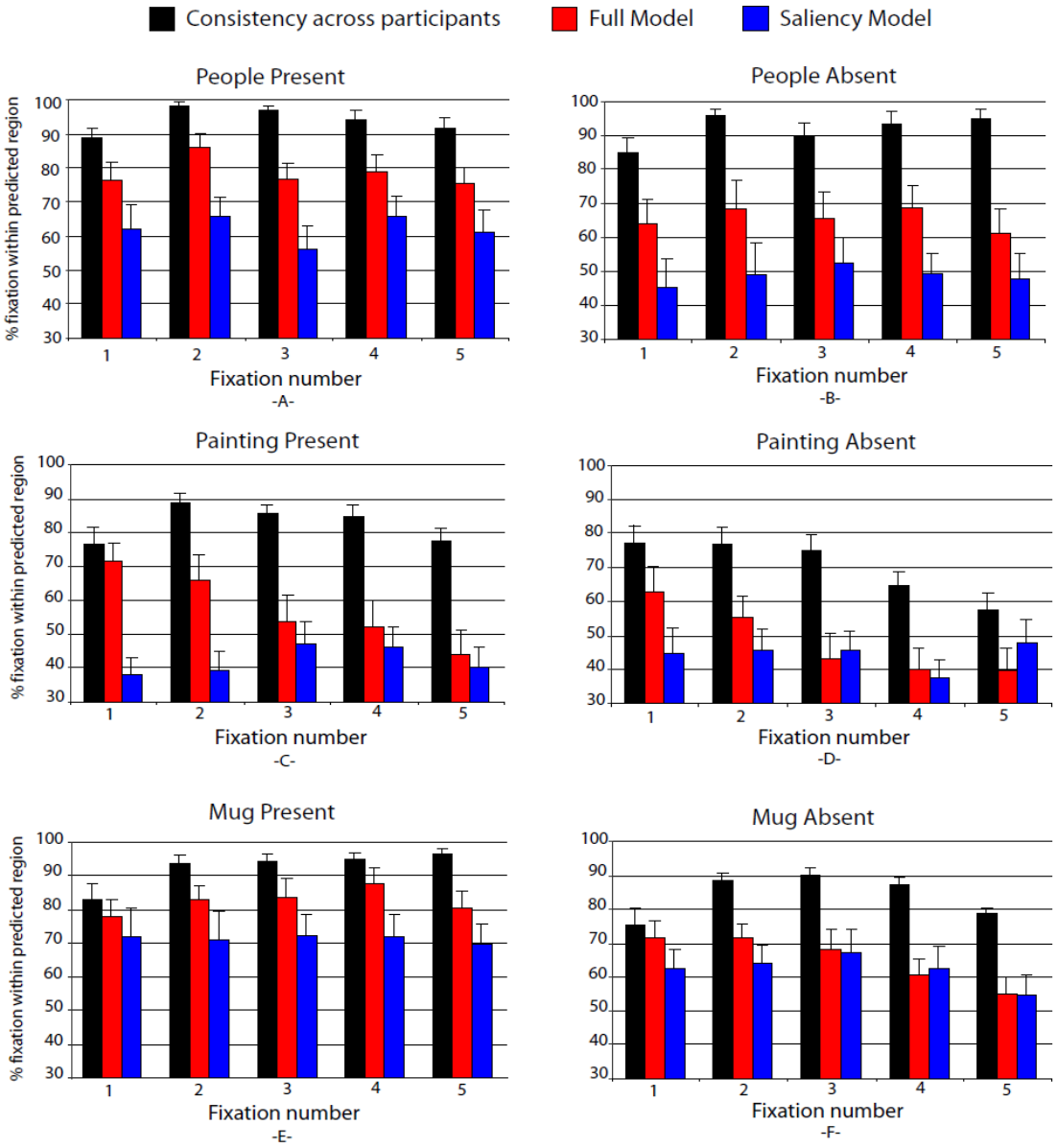
-C-



-D-

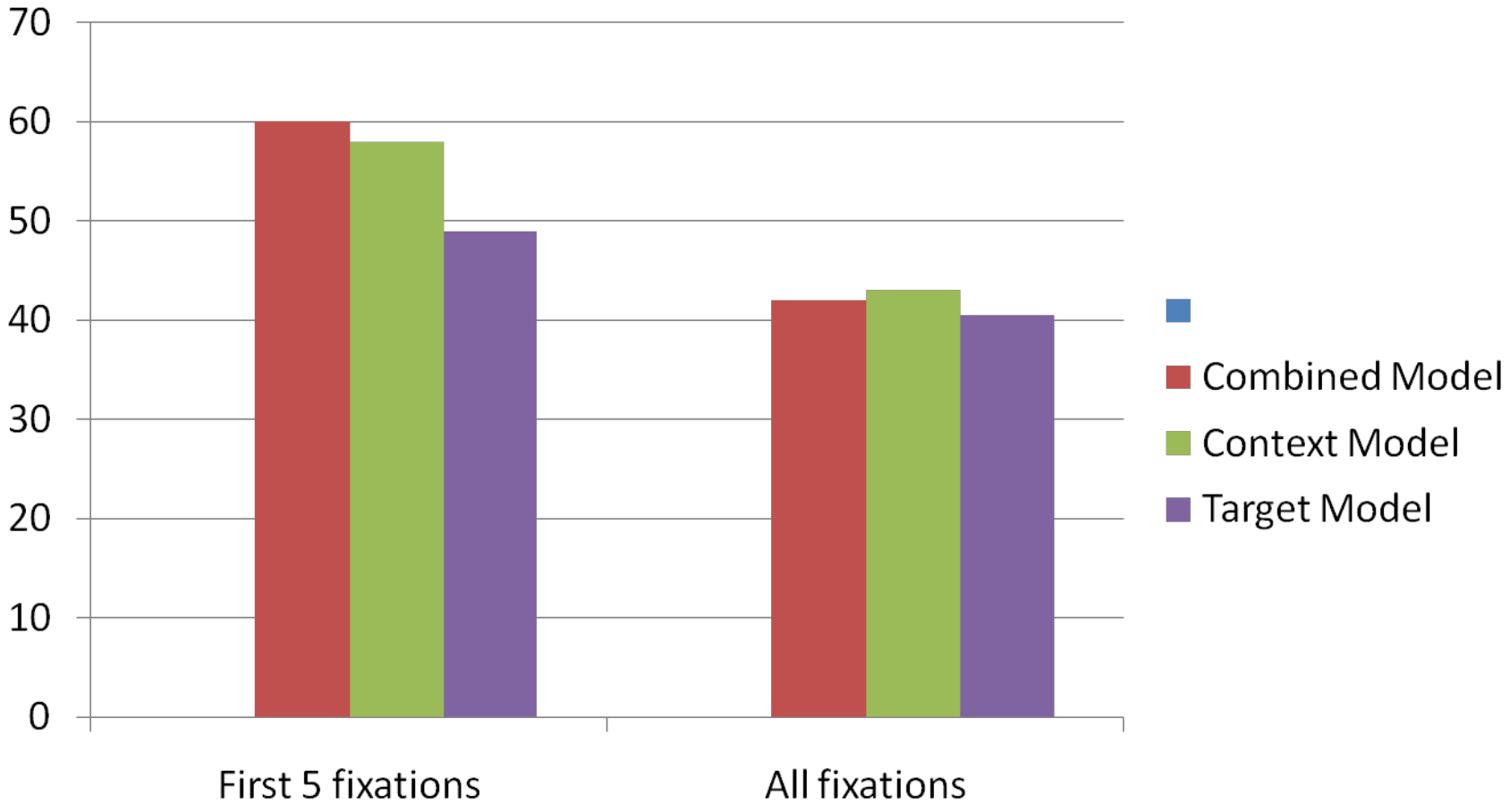
-E-

-F-



Torralba et al 2006

Results (search task : person + walking)



Work Left

- Testing the model on other search tasks.
- Compare the cases when target object is present and when the target object is absent.

Possible Extensions

- Incorporate Bottom Up saliency
- Better Context Models
- Issues like Center Bias in eye movements.

References

- B. C. Russell, A. Torralba, K. P. Murphy, W. T. Freeman, *LabelMe: a database and web-based tool for image annotation*. International Journal of Computer Vision, pages 157-173, Volume 77, Numbers 1-3, May, 2008.
- A Torralba, A. Oliva , M. S. Castellhano, J. M. Henderson , *Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search* , Psychological Review, , pages 766-786 , Volume 113 , Number 4, October 2006
- Kanan C., Tong M., Zhang L., Cottrell G. (2009). *SUN: Top-down saliency using natural statistics. Visual Cognition,* 17, 979–1003.
- John M Henderson & Antje Nuthman(2010)*, Object-based attentional selection in scene viewing,* Journal of Vision(2010), 10(8):20, 1-19
- A. Torralba, K. P. Murphy and W. T. Freeman. (2004). *Sharing features: efficient boosting procedures for multiclass object detection.* Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). pp 762- 769.