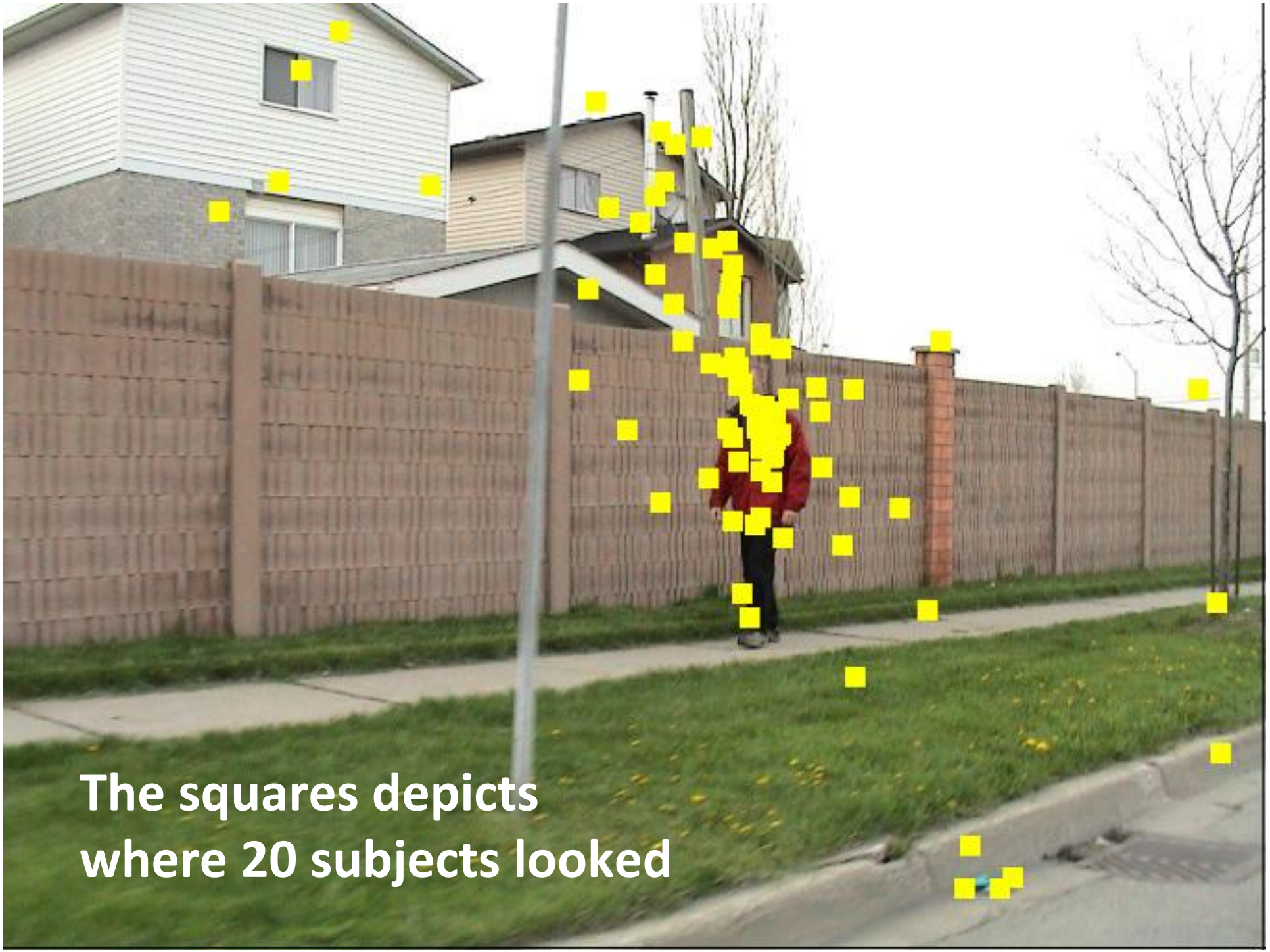

Where do you look on this image?


The squares depicts
where 20 subjects looked

Saliency and Visual Attention


Adapted from T. Judd

Adapted from T. Judd


Adapted from T. Judd

Adapted from T. Judd


Adapted from T. Judd

Adapted from T. Judd

Adapted from T .Judd

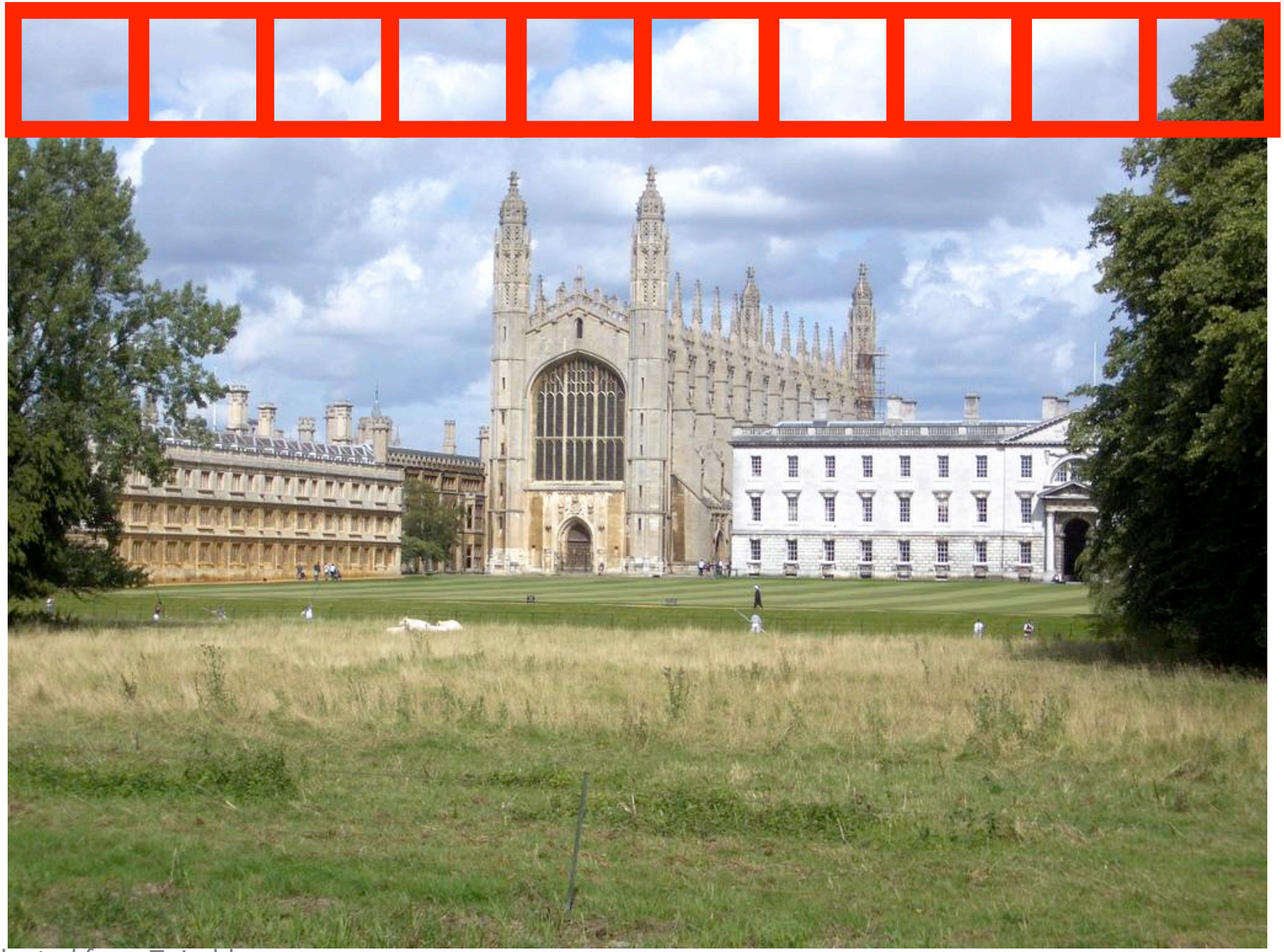
Adapted from T. Judd

Adapted from T. Judd

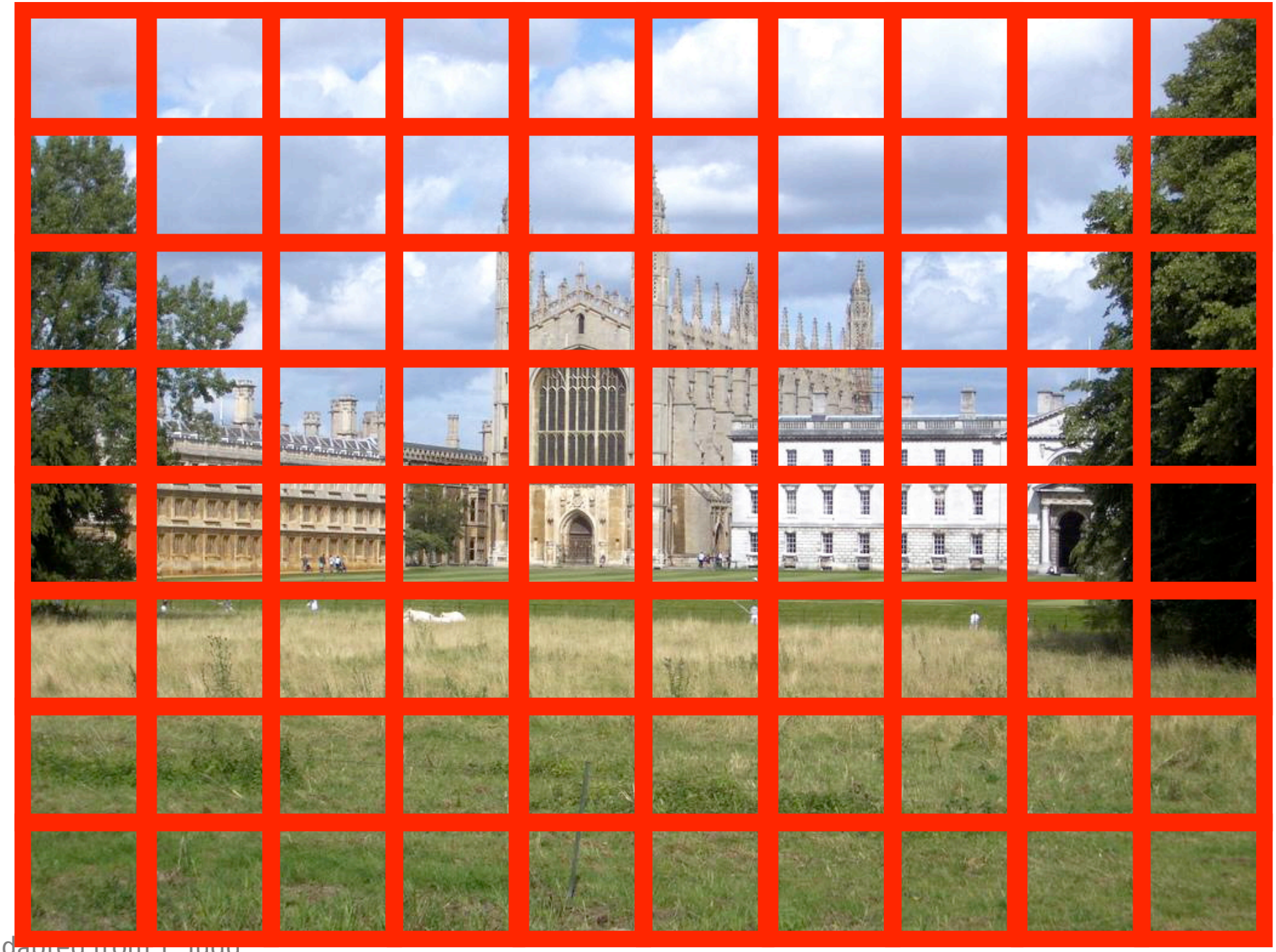
Adapted from T. Judd

Adapted from T. Judd

Adapted from T. Judd

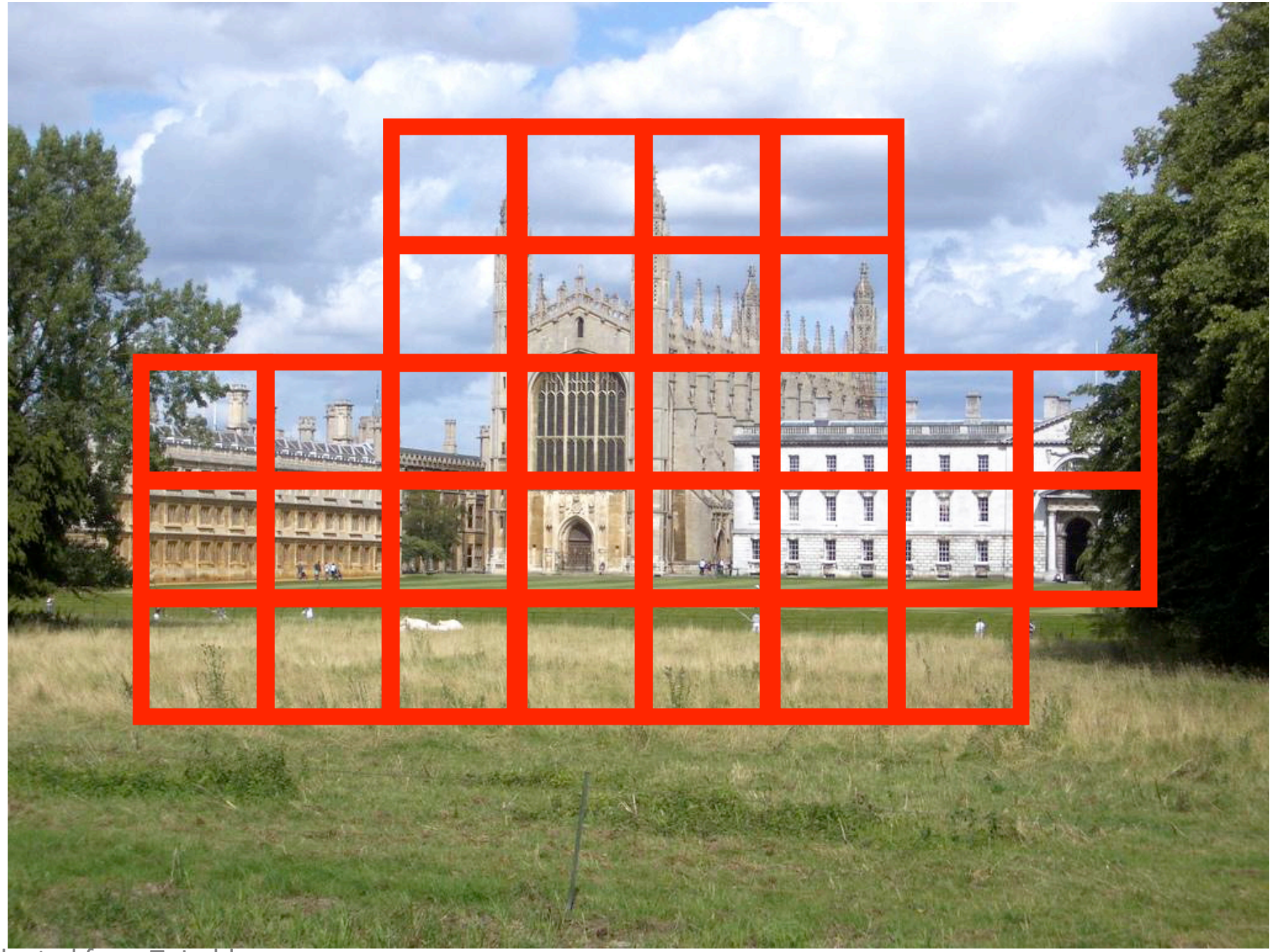


Adapted from T. Judd



Adapted from T. Judd

Adapted from T. Judd

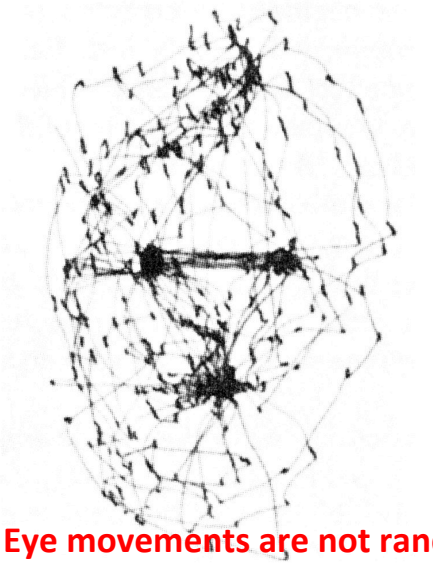


Adapted from T. Judd

Adapted from T. Judd

Scene Analysis and Eye Movements

- Our visual system processes an enormous amount of data coming from the retina. ~10^9 bits/sec (Itti, PhD Thesis, 2000)
- We have developed some **selection mechanisms** to find the most relevant parts of a scene.
  - Warning (noticing predators, sudden motion, etc.)
  - Exploration (finding preys, locating objects, etc.)



Eye movements are not random

Adapted from T. Judd, M. S. Lewicki

Attention and Scene Analysis

- **Attention** is a complex set of interrelated processes.
  - selection of information (bottom-up)
  - integration of that information with existing knowledge (top-down)
- Why do perceptual systems **need attention**?
  - limited resources
  - Even though we have 10^{12} neurons, the brain is still not sufficient to process all the information coming out of the retina
  - simplifies the problem computationally by selecting information
  - perceptual constancy by separating the "foreground" from the "background"

Adapted from M. S. Lewicki

Terminology

- **Gaze** refers to coordinated eye-head movements during shifts in visual attention.
- **Fixation** is the maintaining of the visual gaze on a single location.
- **Saliency** describes the distinctive nature of regions/ objects about how they stand out in relation to their surroundings (grab our attention).

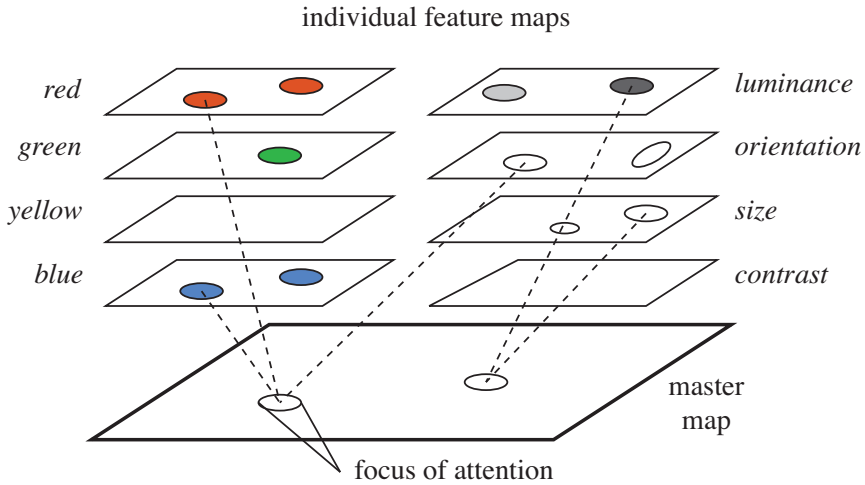
What is attention?

- *"Everyone knows what attention is."*

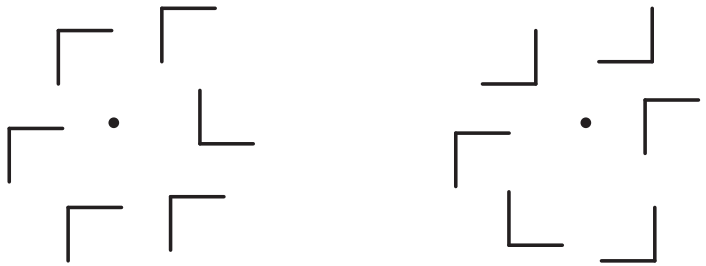
  William James

- Theories of attention
  - Feature-Integration Theory [Treisman and Gelade 1980]
  - Textons [Julezs 1981]
  - Guided Search Theory [Wolfe 1989]
  - Similarity [Duncan and Humphreys 1987]

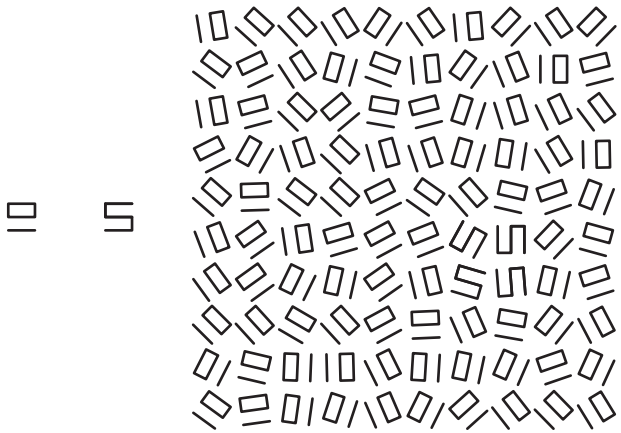
Theories of Visual Attention



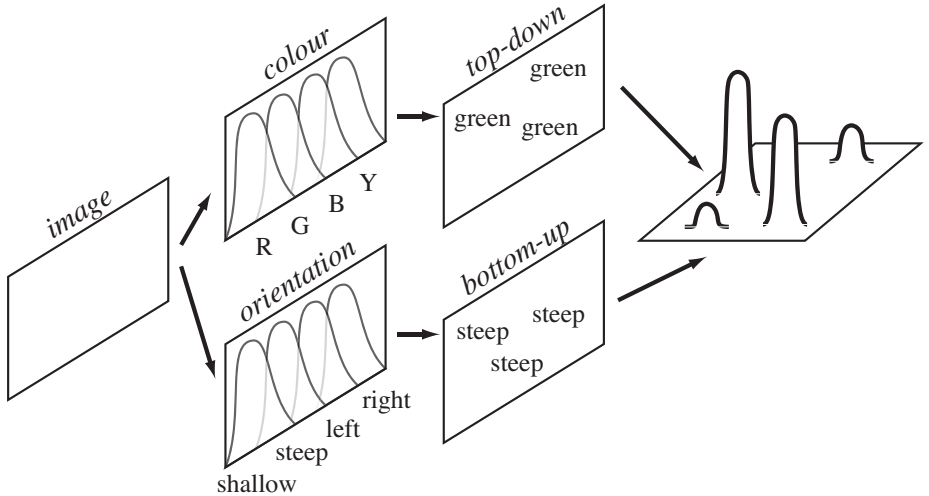
- Feature-Integration Theory [Treisman and Gelade 1980]



- Similarity [Duncan and Humphreys 1987]



- Textons [Julezs 1981]



- Guided Search Theory [Wolfe 1989]

[Healey and Enns 2011]

Attention Mechanisms

- **Attention** is a complex set of interrelated processes
  - selection of information (bottom-up)
  - integration of that information with existing knowledge (top-down)
- Bottom-up attention
  - very rapid, task-independent
- Top-down attention
  - slower, task dependent

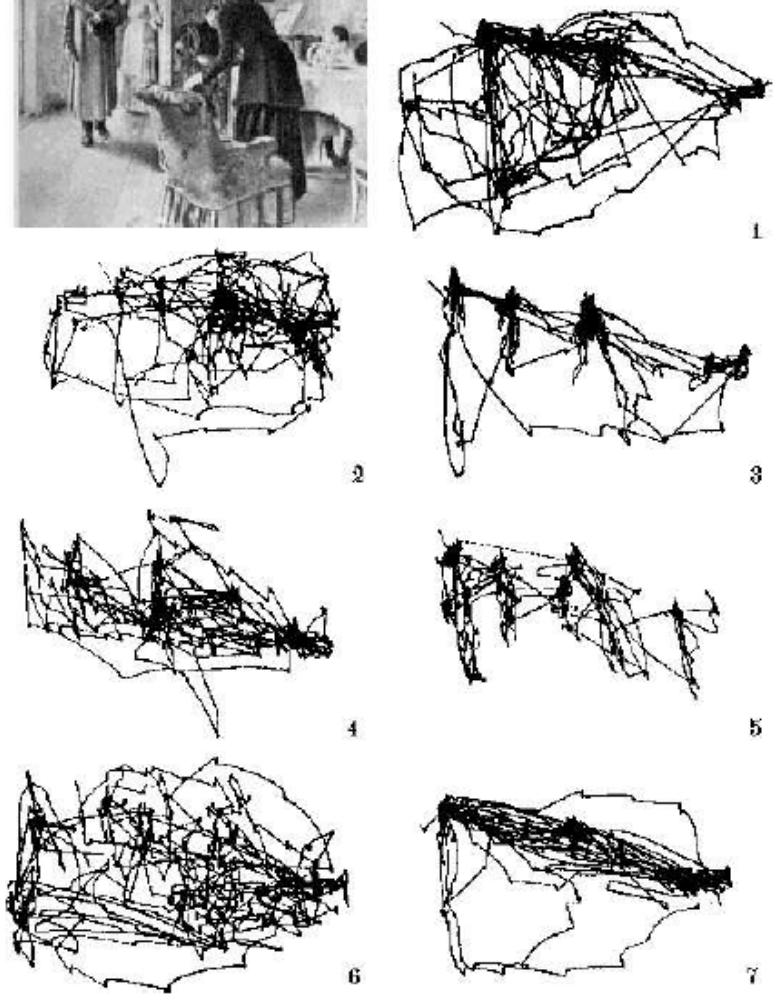
Top-down Attention

"They did not expect him" by Repin

Yarbus (1967) was the first to show that *task* influences fixation locations

Adapted from T. Judd

Saccades and Fixations



Free Examine

What are the material circumstances of the family?

What are their ages?

What were they doing before arrival?

Remember the clothes

Remember object and person positions

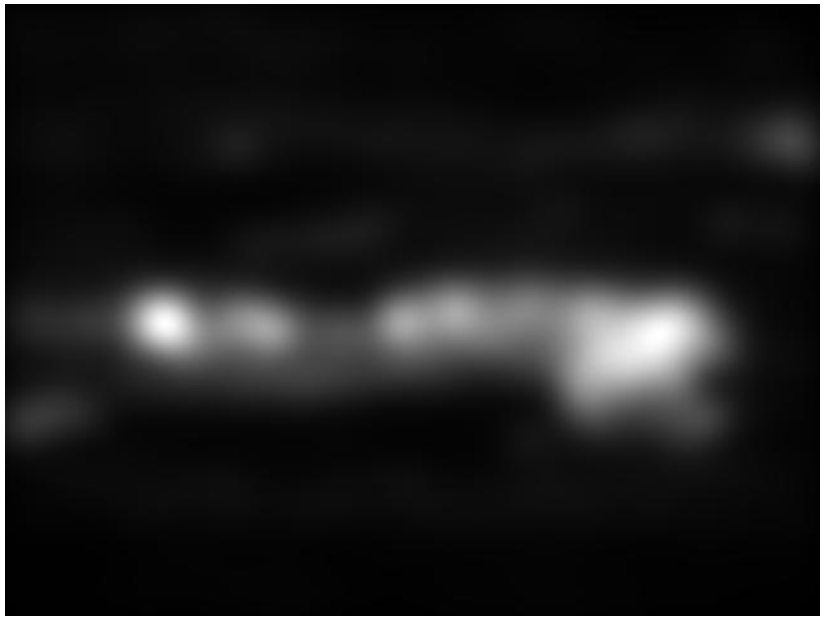
How long has the unexpected visitor been away?

Slide credit: D. Hoeim

[Yarbus 1967]

Computational Models of Visual Attention

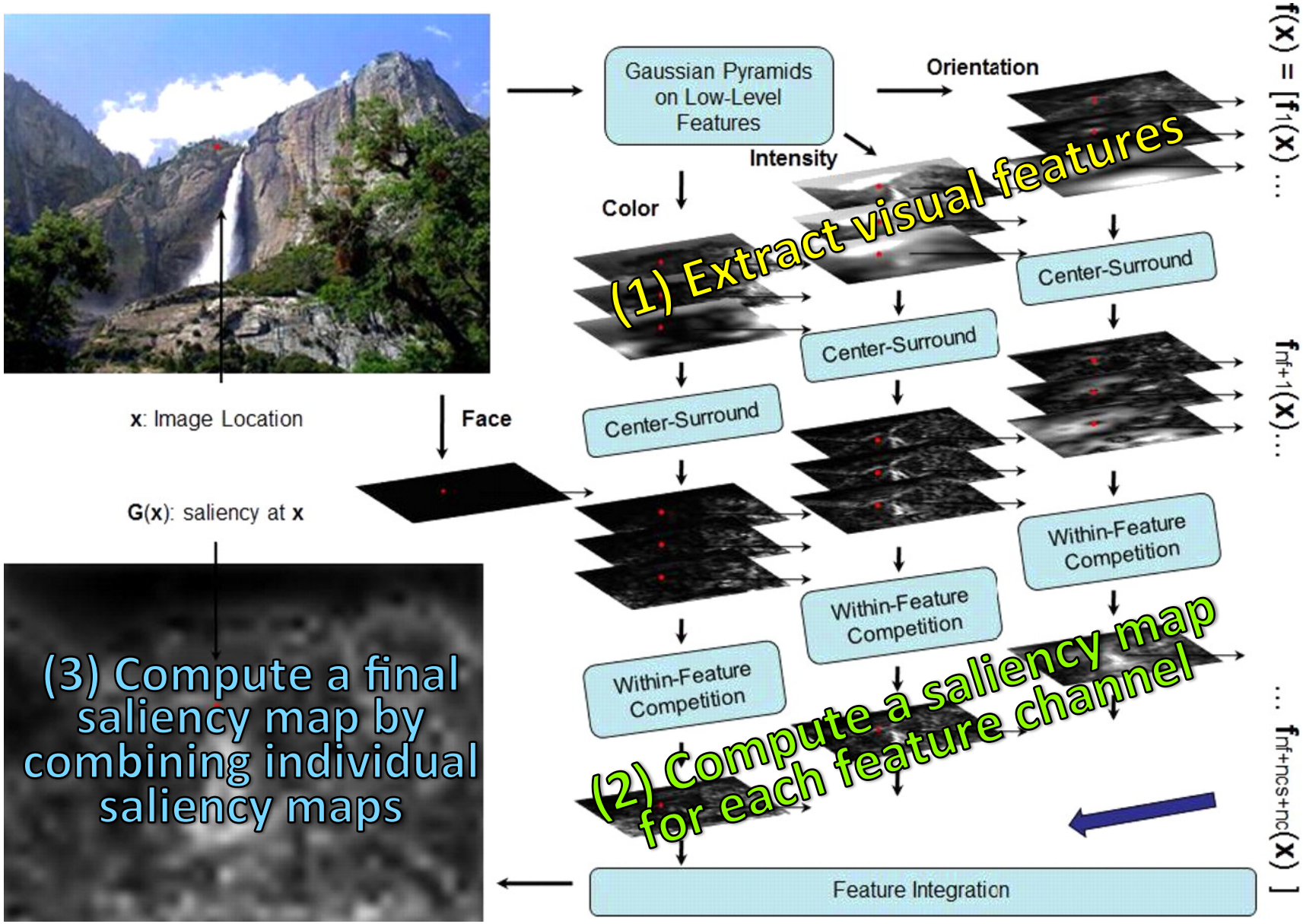
- Can machines predict where the humans look at a given image?



Itti-Koch Model (1998)

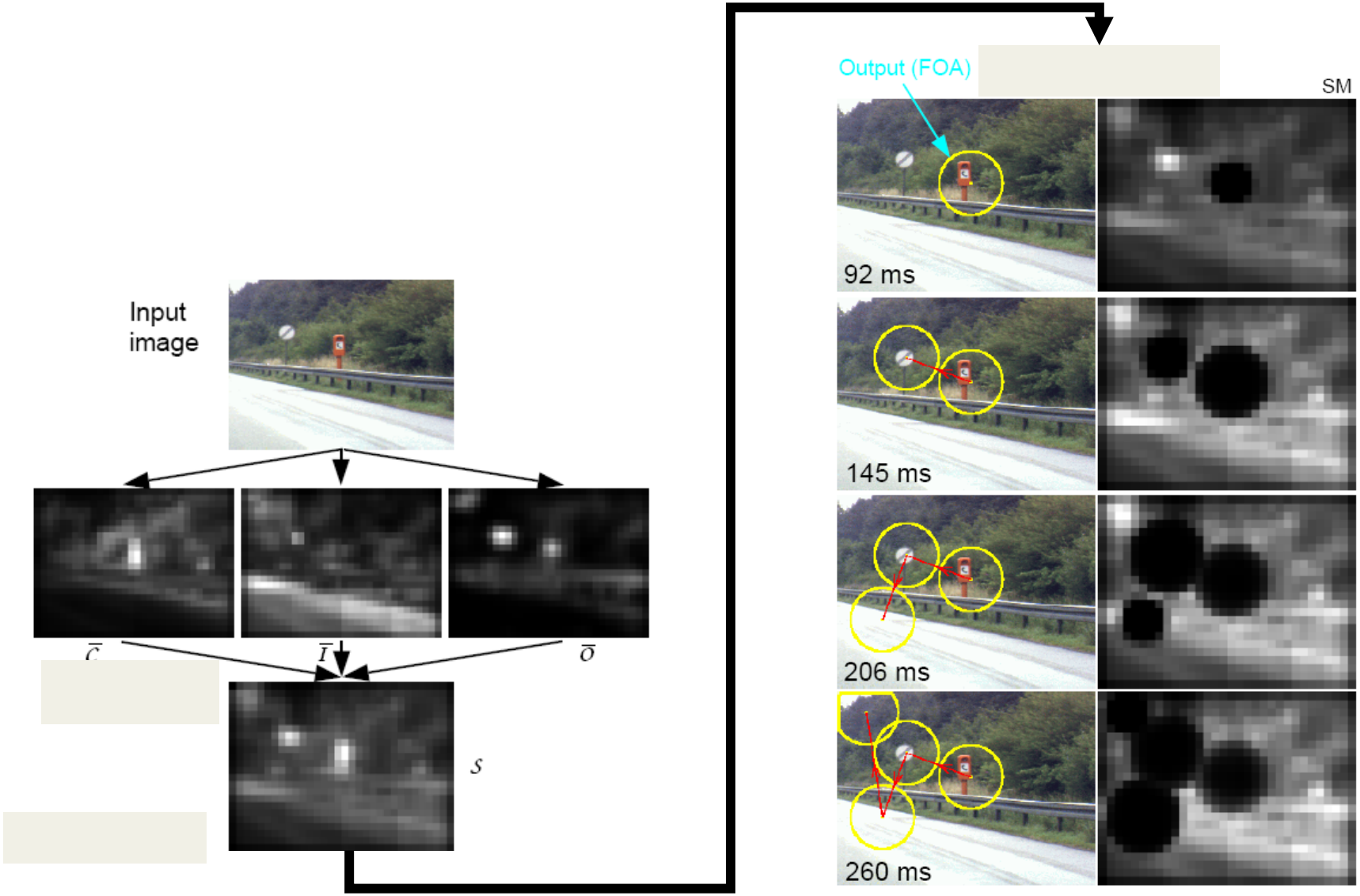
- *First computational model of visual attention to* predict where people look
- An implementation of Koch & Ullman, 1985
- employs a multi-scale center-surround mechanism which imitates the workings of the retinal receptive field

Bottom-up Saliency Models



Zhao Q , Koch C J Vis 2012;12:22

Predicted Fixations



Slide credit: D. Hoeim

[Itti Koch Niebur 1998]

Many models of saliency have been introduced

Biologically Inspired

Itti and Koch (1998)

Cerf et al. (2007)

Hou and Zhang (2007)

Rosenholtz (1999)

Itti and Baldi (2006)

Le Meur et al. (2006)

Seo & Milanfar (2009)

Zhang & Cottrell (2008)
SUN model

Goferman et al. (2009)

Achanta (2010)

Mathematically Inspired

Heral et al. (2007)
Graphical Model

Avraham and Lindenbaum (2009)
Esaliency

Bruce and Tsotsos (2009)
Information theoretic approach

Kienzle et al., (2007)

Gao and Vasconulos (2005)

Itti and Baldi (2006)
"Surprise" model

Navalpakkam and Itti (2005)

Elazary and Itti (2010)

Add top-down features

Ehinger et al., (2009)
(search task)

Oliva et al. (2003)

Torralba et al. (2006)

Zhang et al. (2008)

Kanan et al. (2009)

Adapted from T. Judd

Which one is the best?

Biologically Inspired

Itti and Koch (1998)

Cerf et al. (2007)

Hou and Zhang (2007)

Rosenholtz (1999)

Itti and Baldi (2006)

Le Meur et al. (2006)

Seo & Milanfar (2009)

Zhang & Cottrell (2008)
SUN model

Goferman et al. (2009)

Achanta (2010)

Mathematically Inspired

Heral et al. (2007)
Graphical Model

Avraham and Lindenbaum (2009)
Esaliency

Bruce and Tsotsos (2009)
Information theoretic approach

Kienzle et al., (2007)

Gao and Vasconulos (2005)

Itti and Baldi (2006)
"Surprise" model

Navalpakkam and Itti (2005)

Elazary and Itti (2010)

Add top-down features

Ehinger et al., (2009)
(search task)

Oliva et al. (2003)

Torralba et al. (2006)

Zhang et al. (2008)

Kanan et al. (2009)

Adapted from T. Judd

Which one is the best?

Biologically Inspired

Itti and Koch (1998)

Cerf et al. (2007)

Hou and Zhang (2007)

Rosenholtz (1999)

Itti and Baldi (2005)

Le Meur et al. (2006)

Seo & Milanfar (2009)

Zhang & Cottrell (2008)
SUN model

Goferman et al. (2009)

Achanta (2010)

Mathematically Inspired

Heral et al. (2007)
Graphical Model

Avraham and Lindenbaum (2009)
Esaliency

Bruce and Tsotsos (2009)
Information theoretic approach

Kienzle et al., (2007)

Gao and Vasconulos (2005)

Itti and Baldi (2006)
"Surprise" model

Navalpakkam and Itti (2005)

Elazary and Itti (2010)

Add top-down features

Ehinger et al., (2009)
(search task)

Oliva et al. (2003)

Torralba et al. (2006)

Zhang et al. (2008)

Kanan et al. (2009)

Benchmark needed!

Adapted from T. Judd

How benchmark data sets are formed?

screen resolution 1280x1024

each image shown for 3 seconds

eye tracker measures location of eye fixation several times a second.

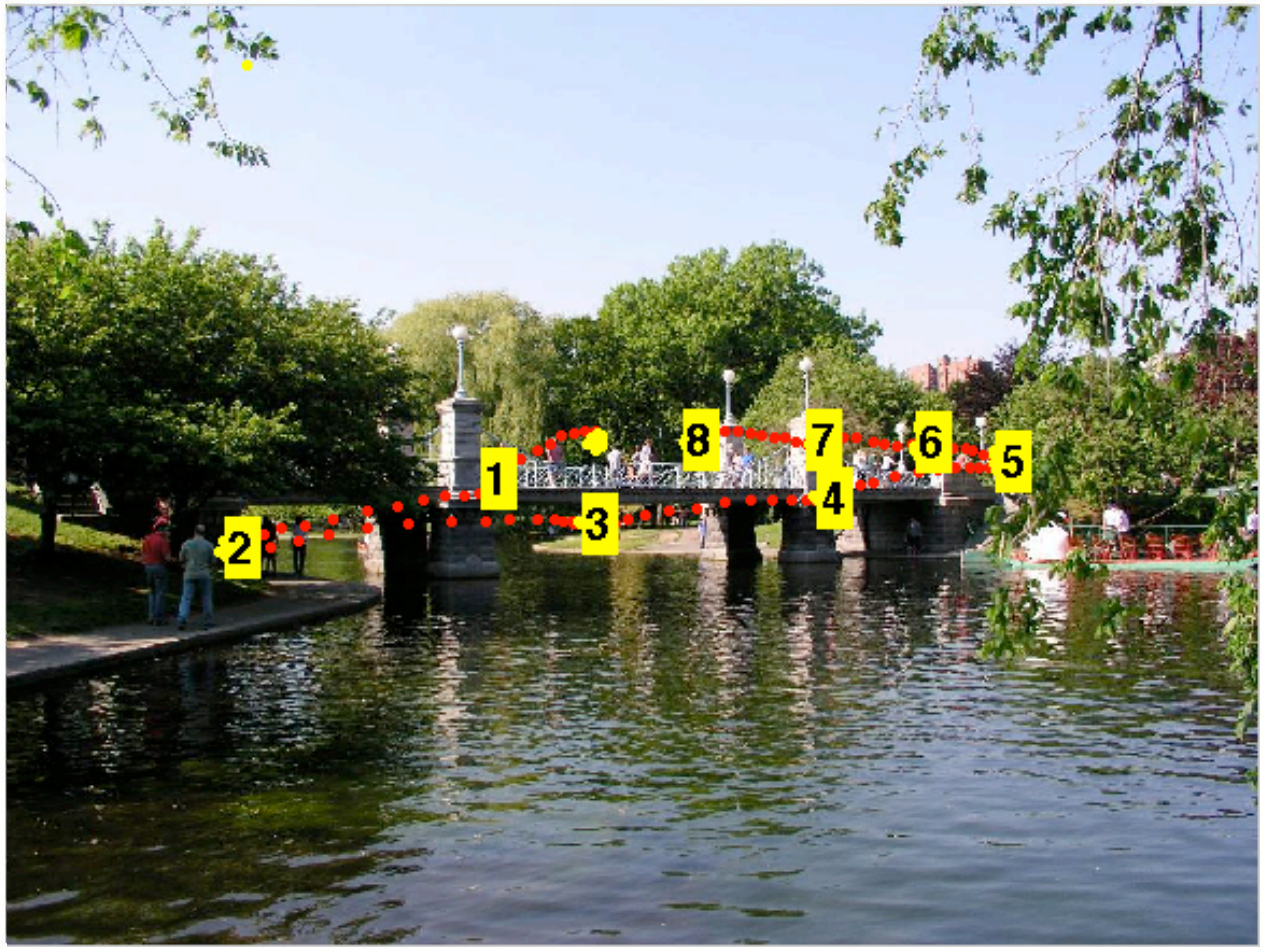
user rests head in chin rest

[Photo Credit: Jason Dorfman CSAIL website]

Through eye tracking experiments

Adapted from T. Judd

Fixations for one observer

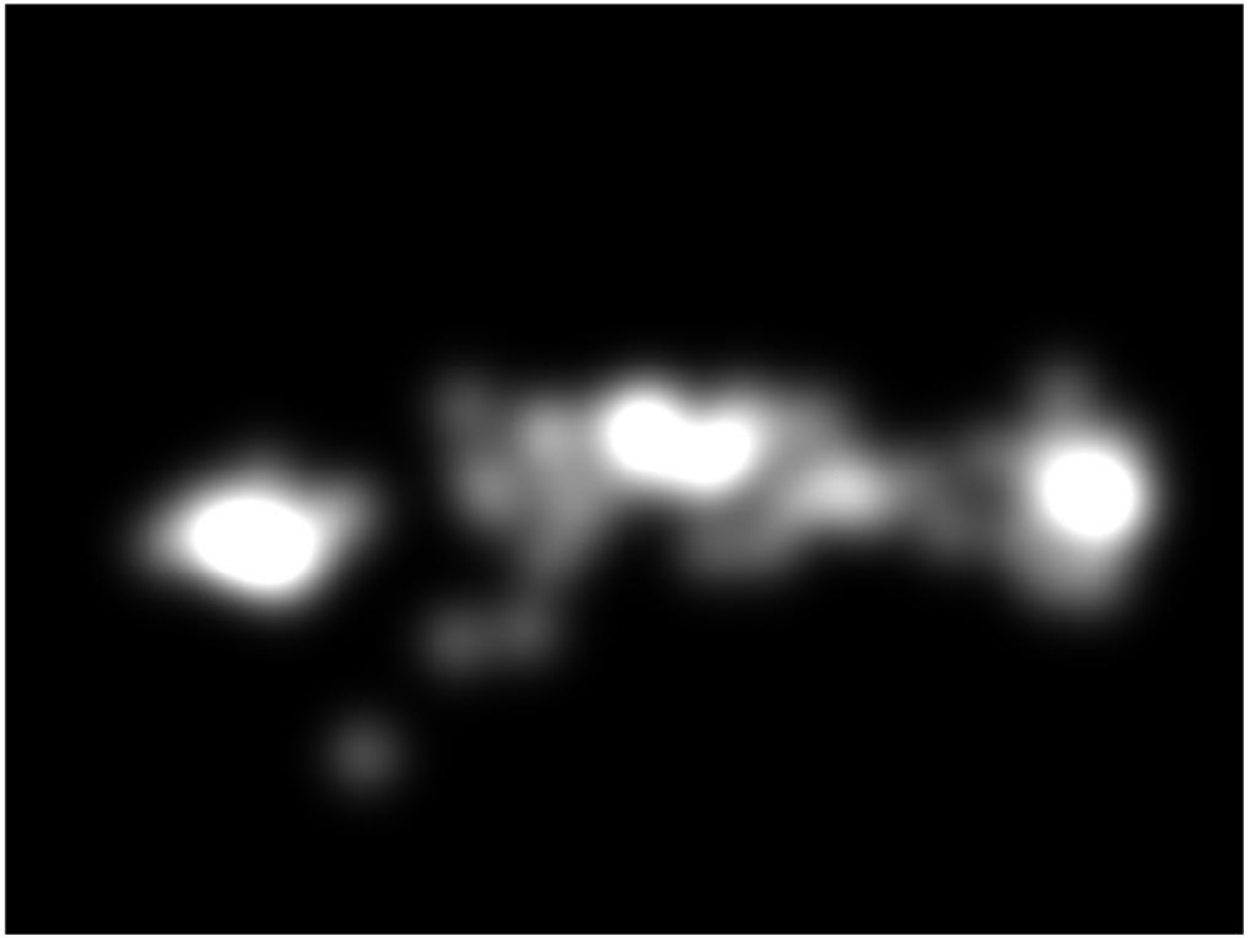


Adapted from T. Judd

Fixations from 15 observers

Adapted from T. Judd

Fixation map created from gaussian convolution over fixations



Fixation map

Adapted from T. Judd

Human Eye Fixations Data Sets

- Bruce data set [Bruce and Tsotsos 2006]
  - 120 color images of indoor and outdoor scenes
  - Eye movement data from 20 subjects
- MIT 1003 data set [Judd et al. 2009]
  - 1003 natural color images (779 landscape + 228 portrait)
  - Eye movement data from 15 subjects
- MIT 300 data set [Judd et al. 2012]
  - 300 natural images (223 landscape + 77 portrait)
  - Eye movement data from 39 subjects

Sample Images

Bruce data set

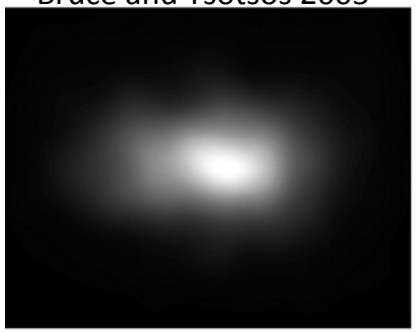
MIT 1003 data set

MIT 300 data set

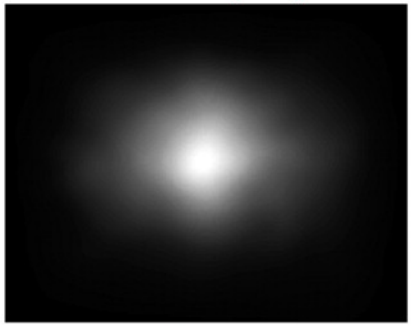
Why are fixations center biased?

- photographer bias
- viewing strategy

Bruce and Tsotsos 2005



Judd 2009



Judd 2011

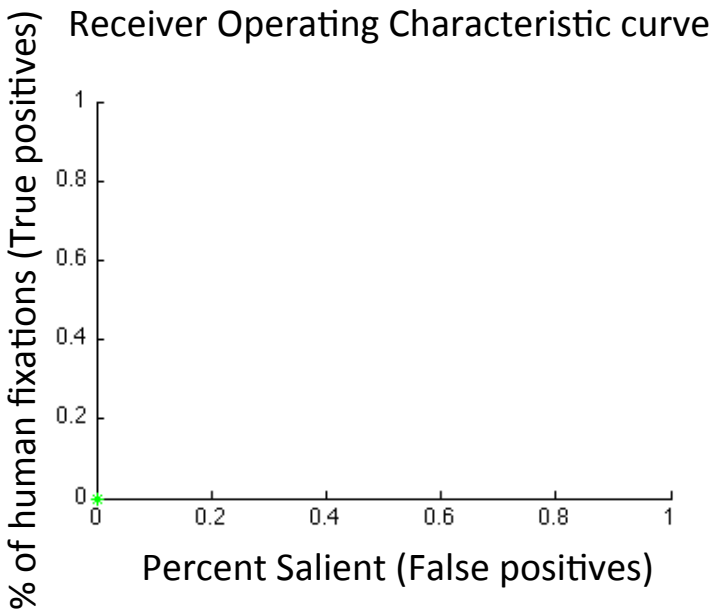
Adapted from T. Judd

Evaluation Metrics

- Area under ROC curve (AUC)
- Earth Mover's Distance (EMD)
- Normalized Scanpath Saliency (NSS)
- Similarity Score
- …

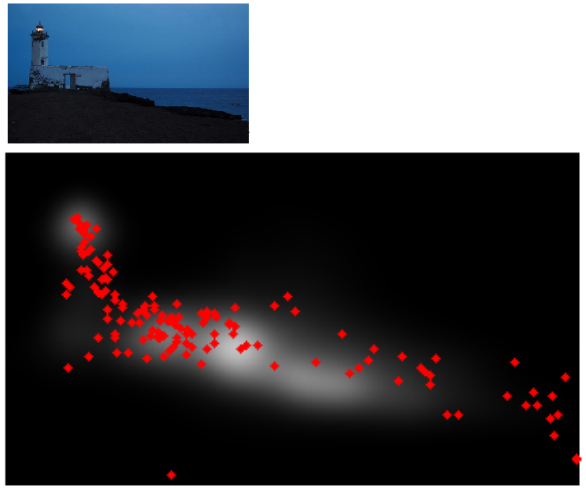
Area under ROC curve (AUC) as a performance measure

Saliency Map

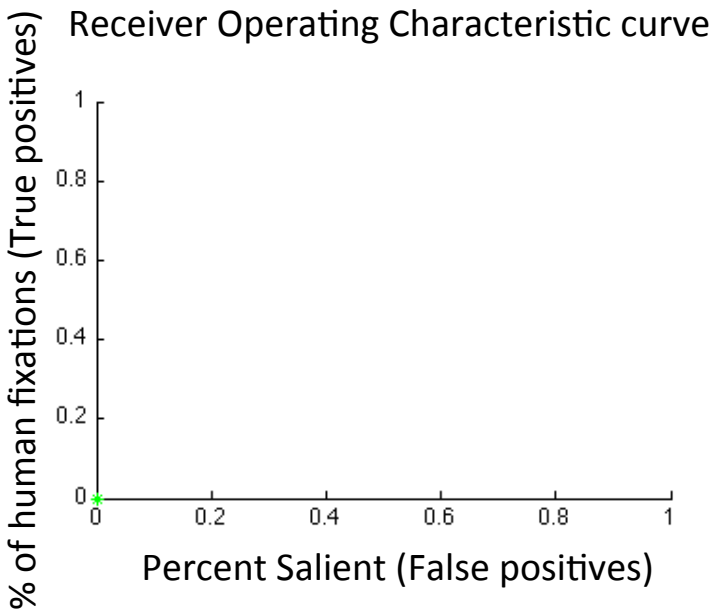


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

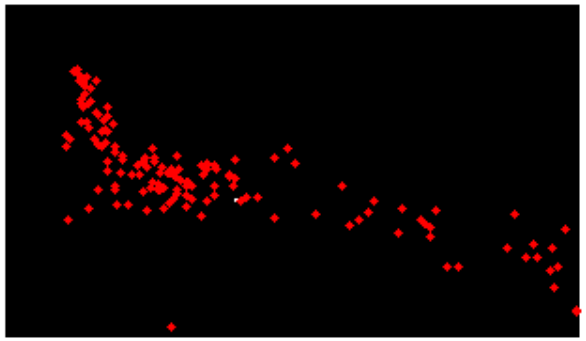


Human Fixations

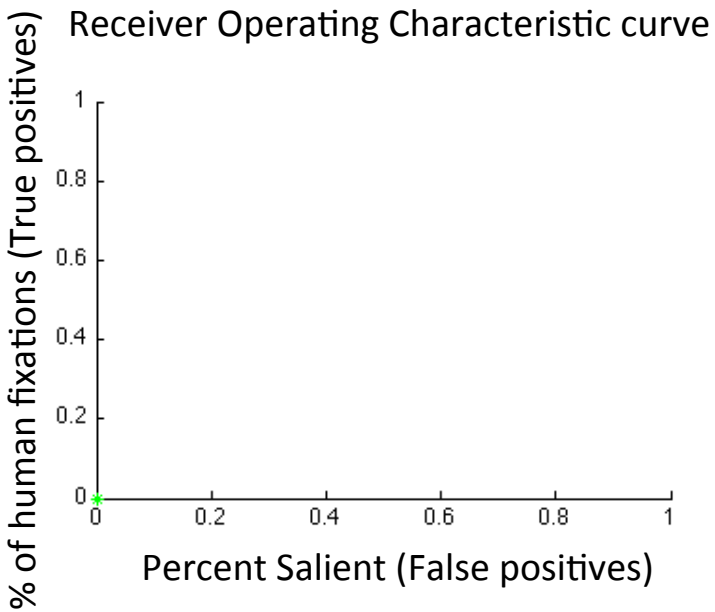


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

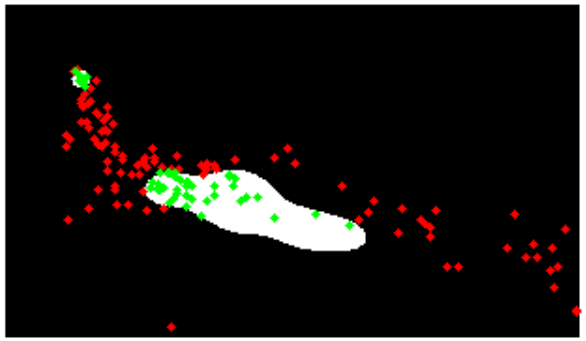


Thresholded Saliency Map

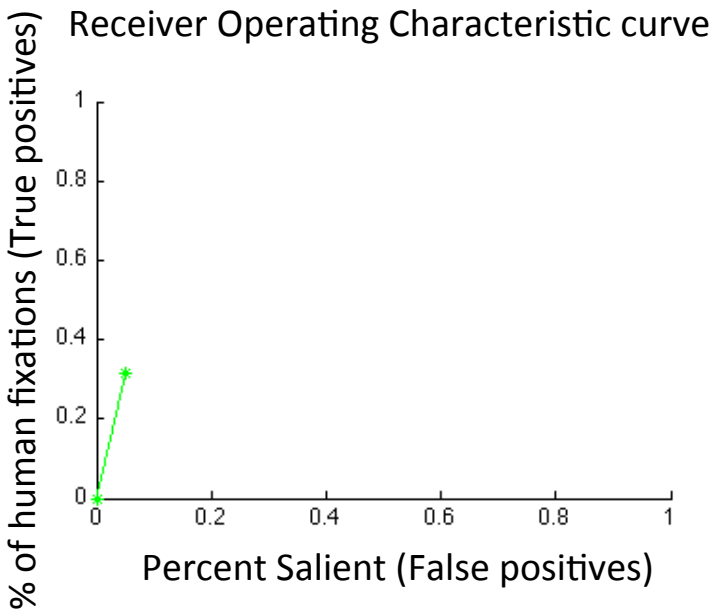


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

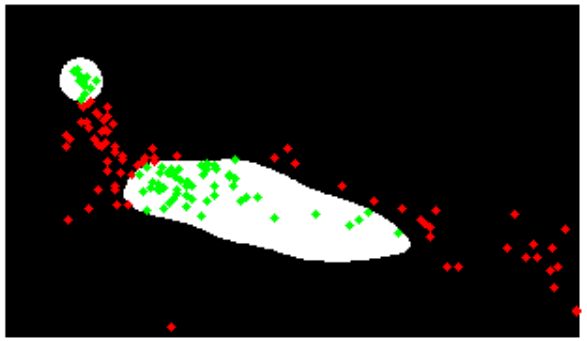


Thresholded Saliency Map

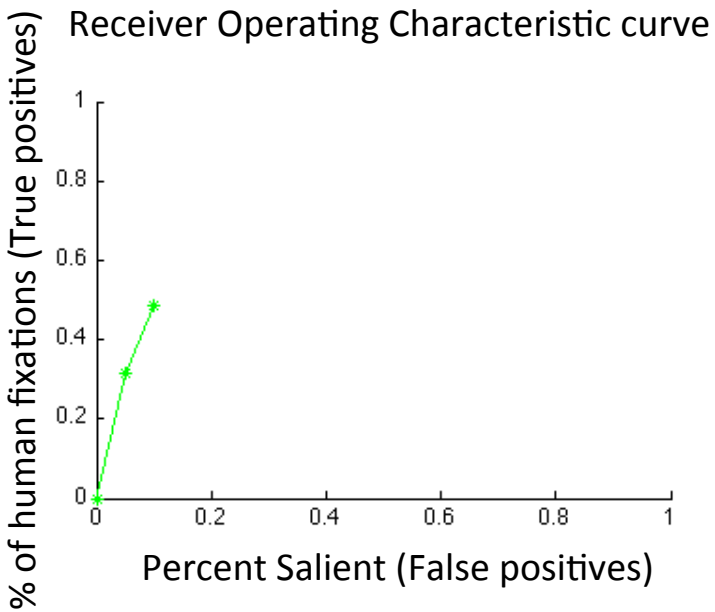


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

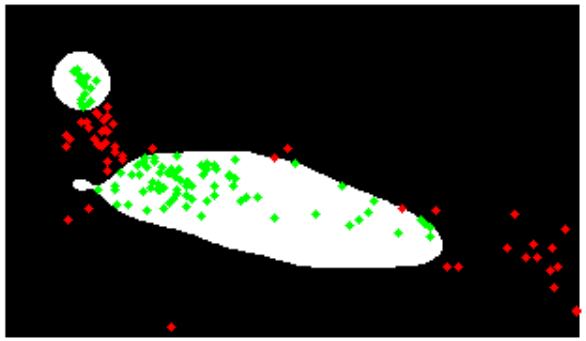


Thresholded Saliency Map

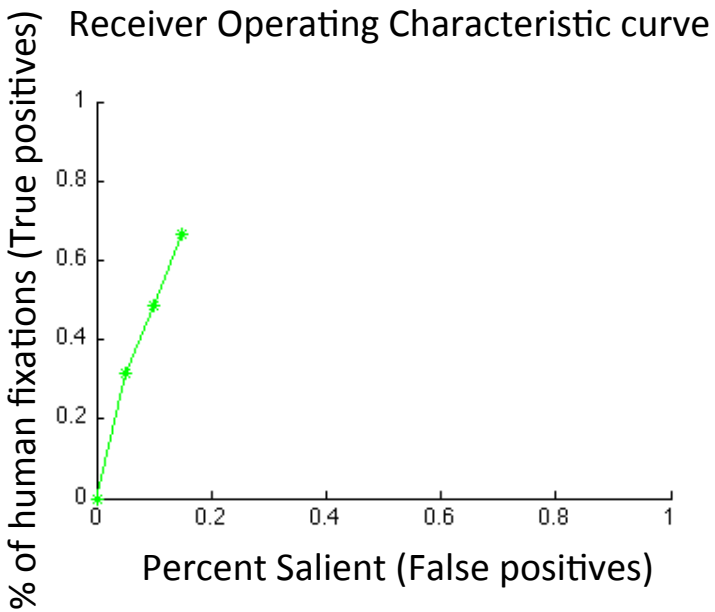


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure



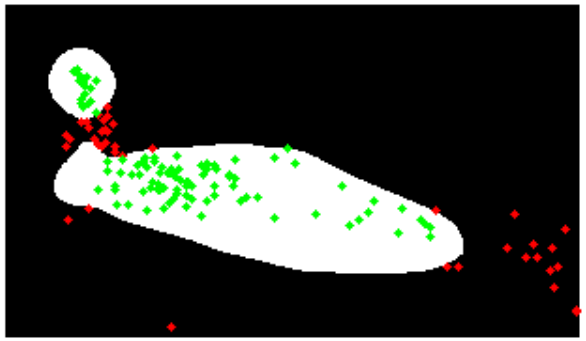
Thresholded Saliency Map



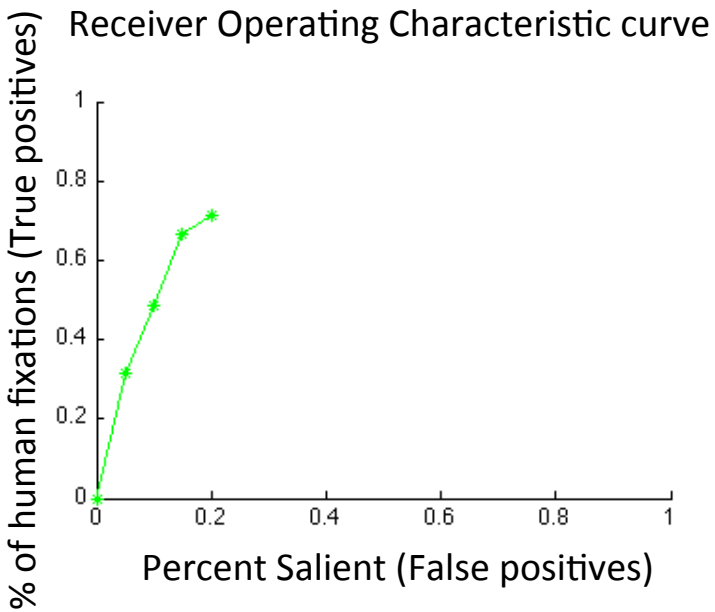
Percentage of fixations are calculated that lie within the salient portion of the map

Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

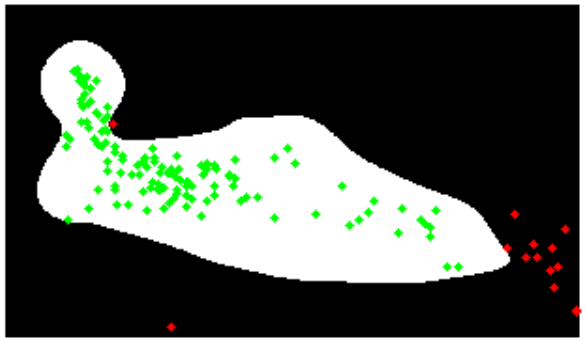


Thresholded Saliency Map

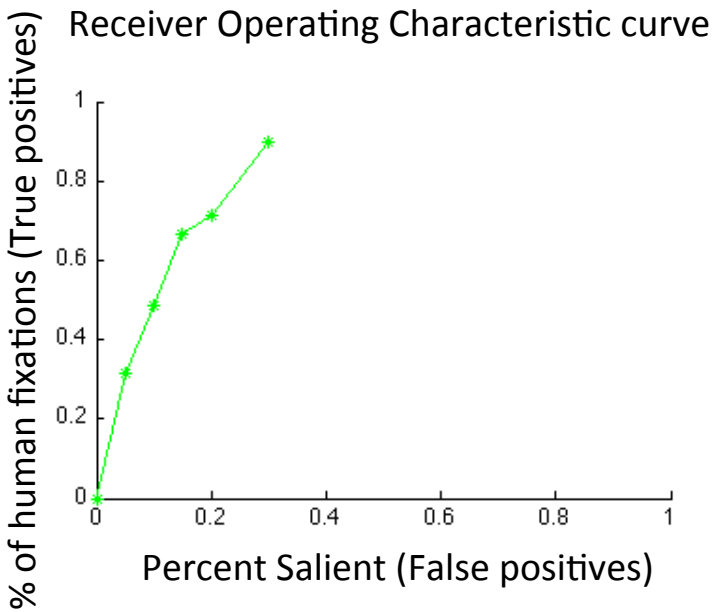


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

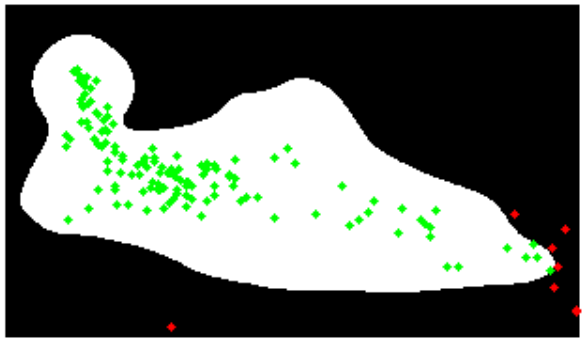


Thresholded Saliency Map

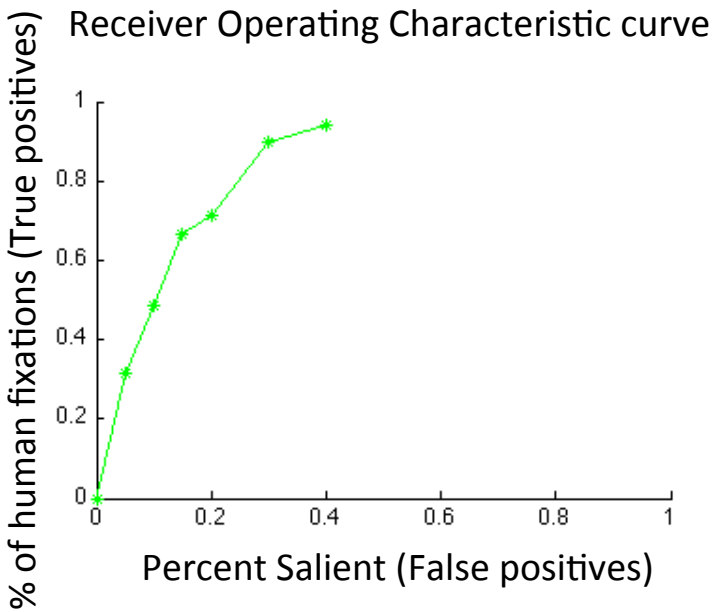


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

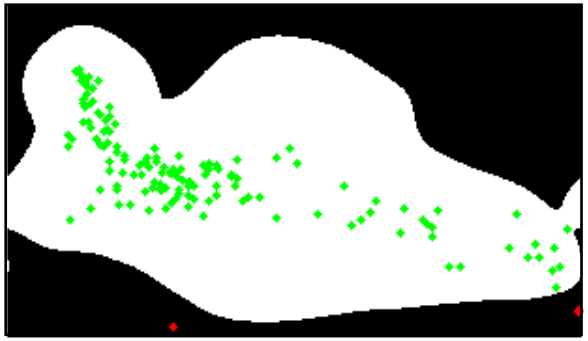


Thresholded Saliency Map

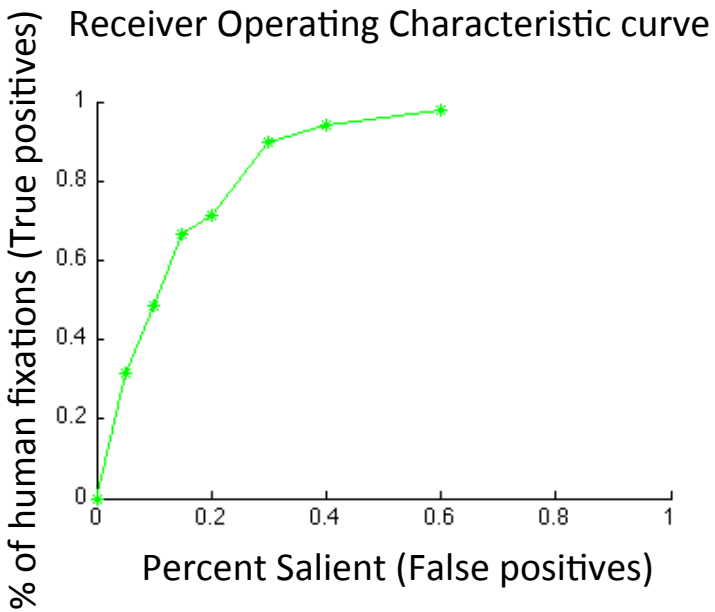


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

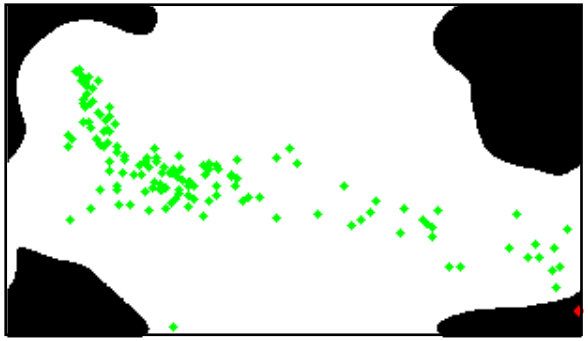


Thresholded Saliency Map

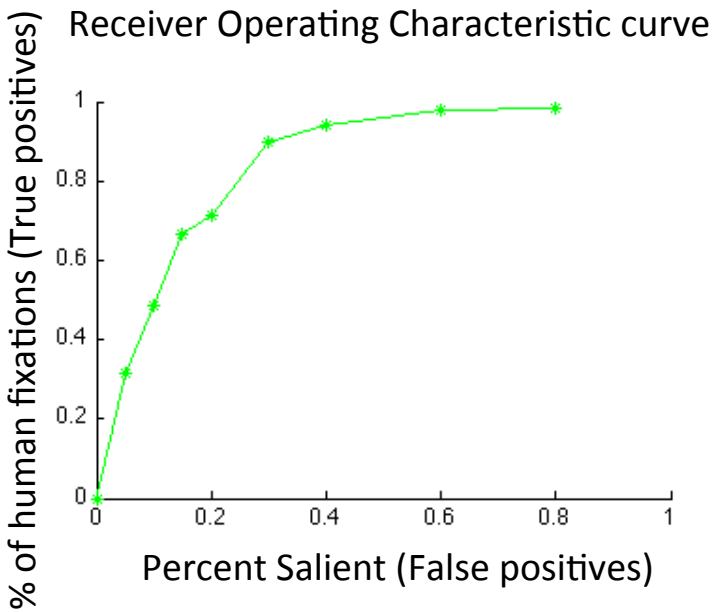


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

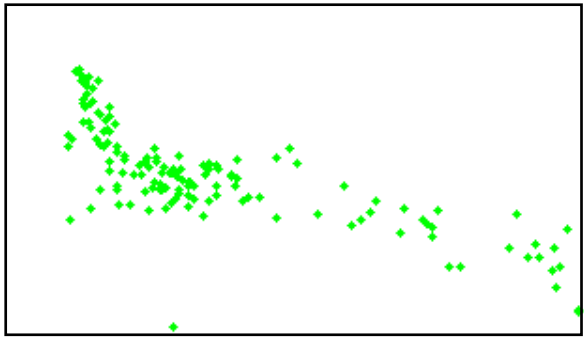


Thresholded Saliency Map

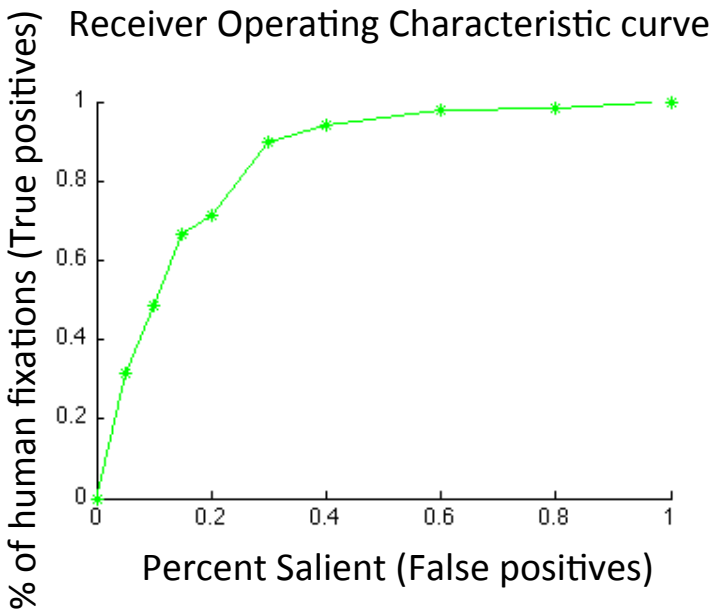


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure



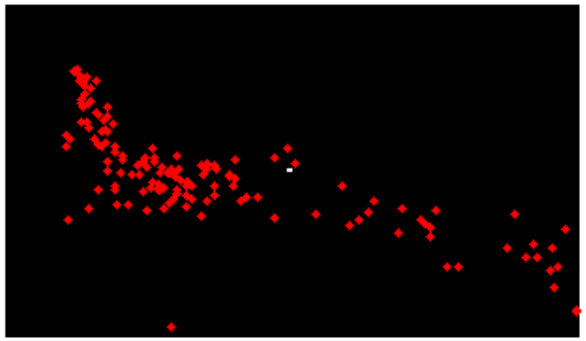
Thresholded Saliency Map



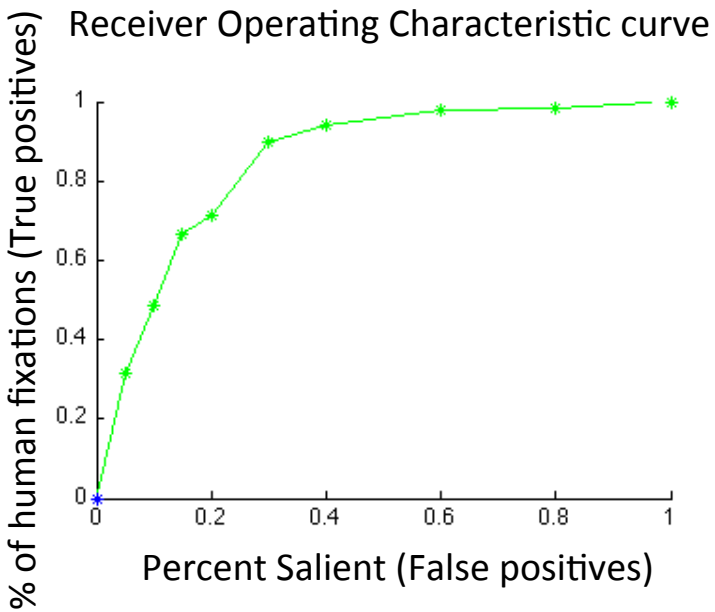
ROC curve always starts at 0 ends at 1

Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

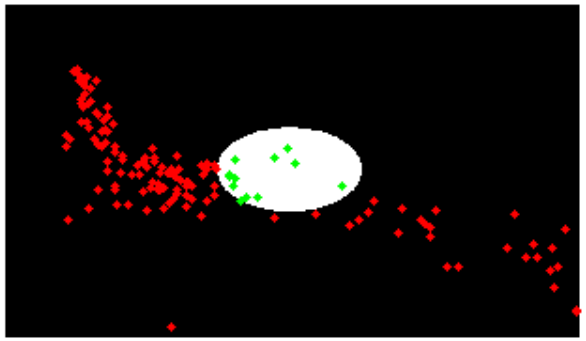


Thresholded Center Map

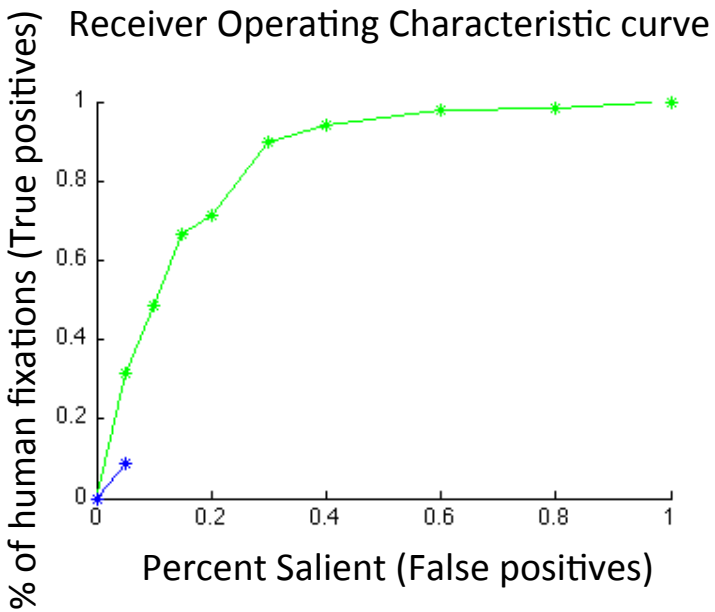


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

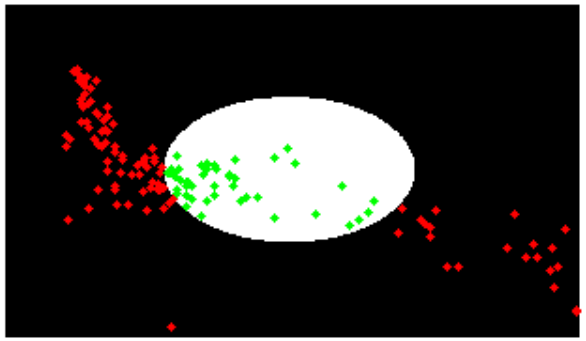


Thresholded Center Map

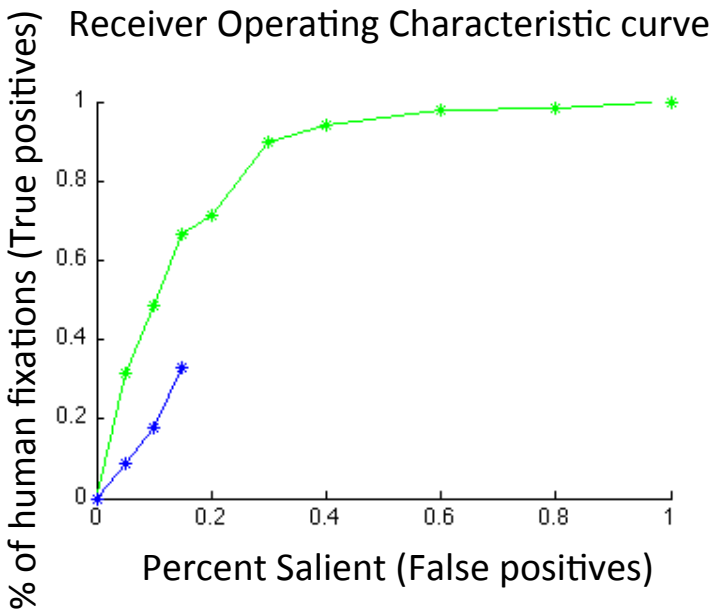


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

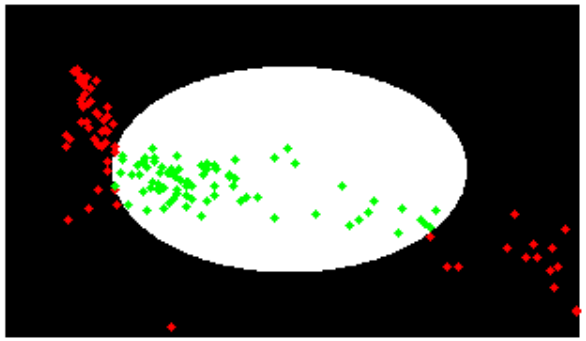


Thresholded Center Map

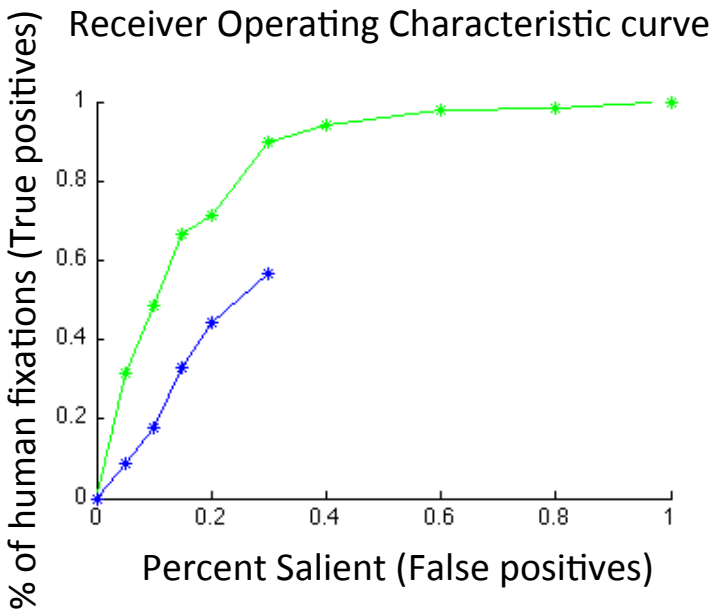


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

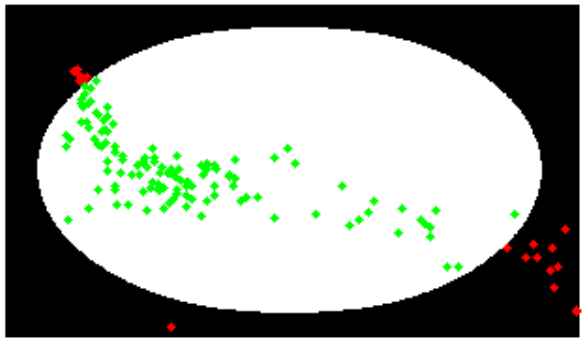


Thresholded Center Map

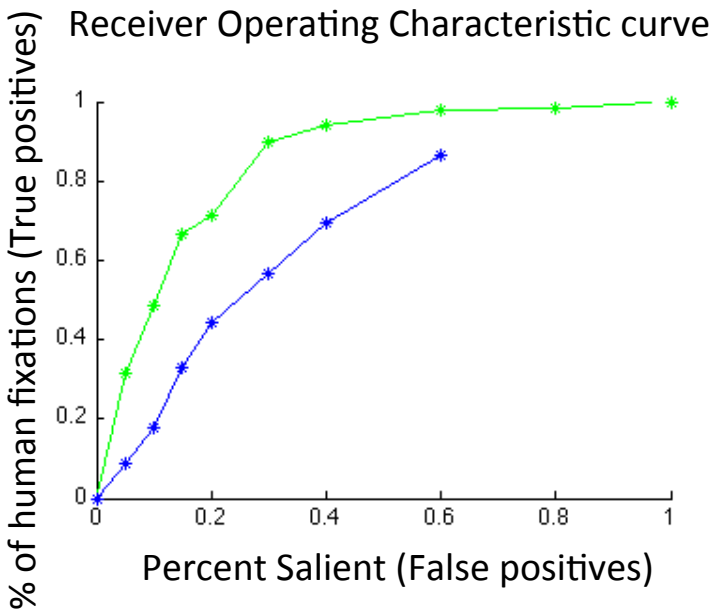


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

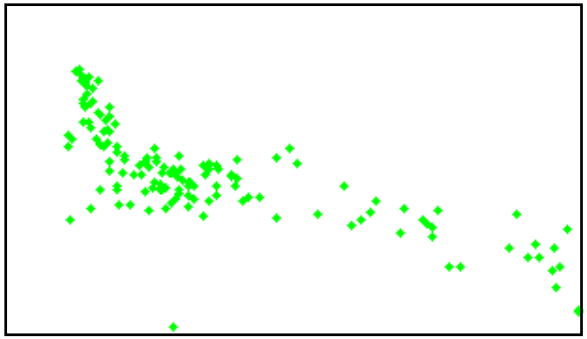


Thresholded Center Map

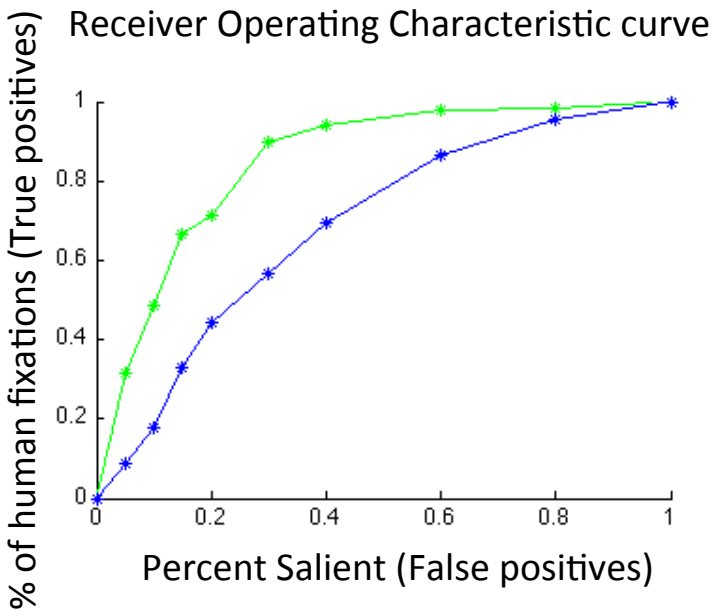


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

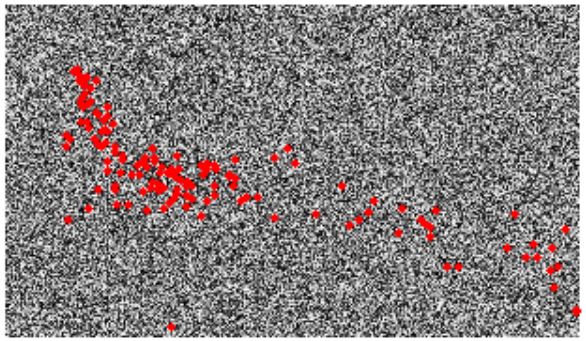


Thresholded Center Map

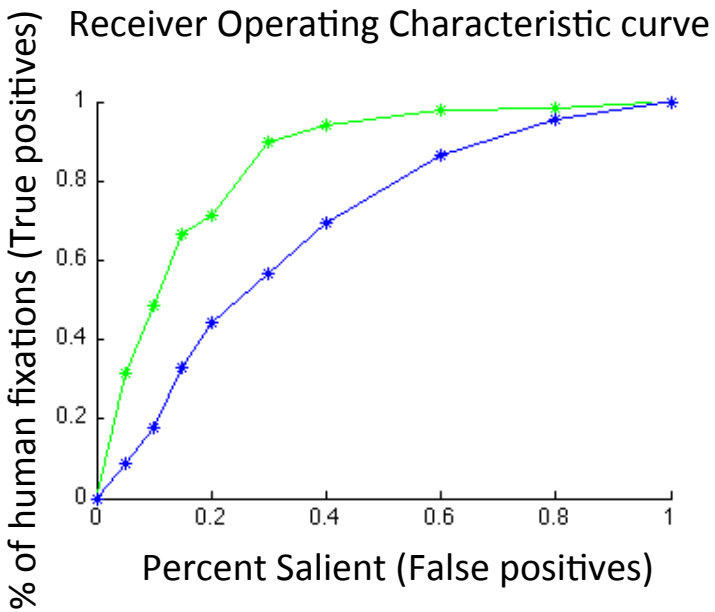


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

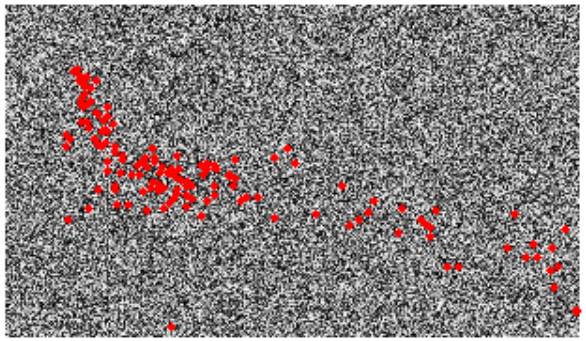


Random Noise map

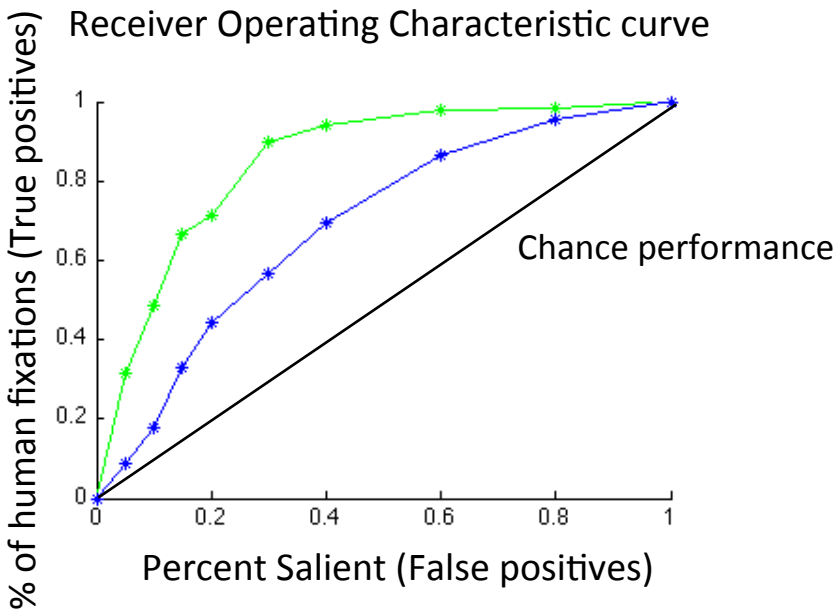


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

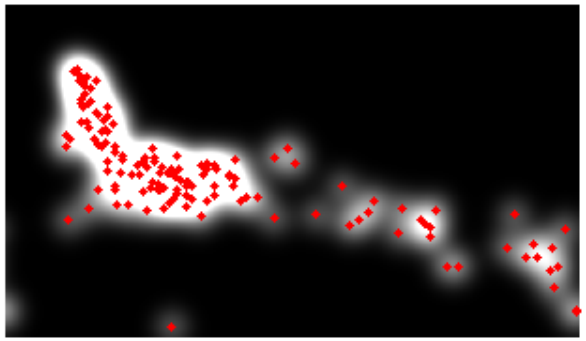


Random Noise map

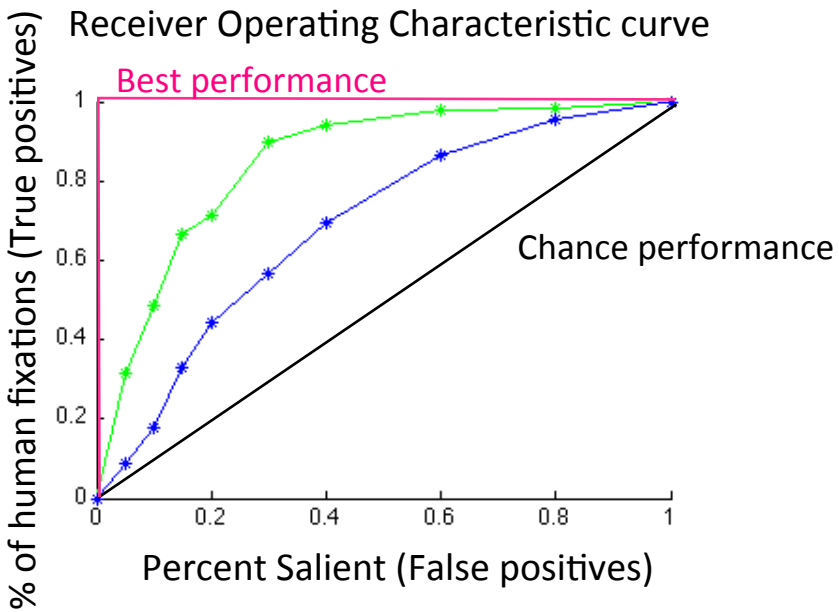


Adapted from T. Judd

Area under ROC curve (AUC) as a performance measure

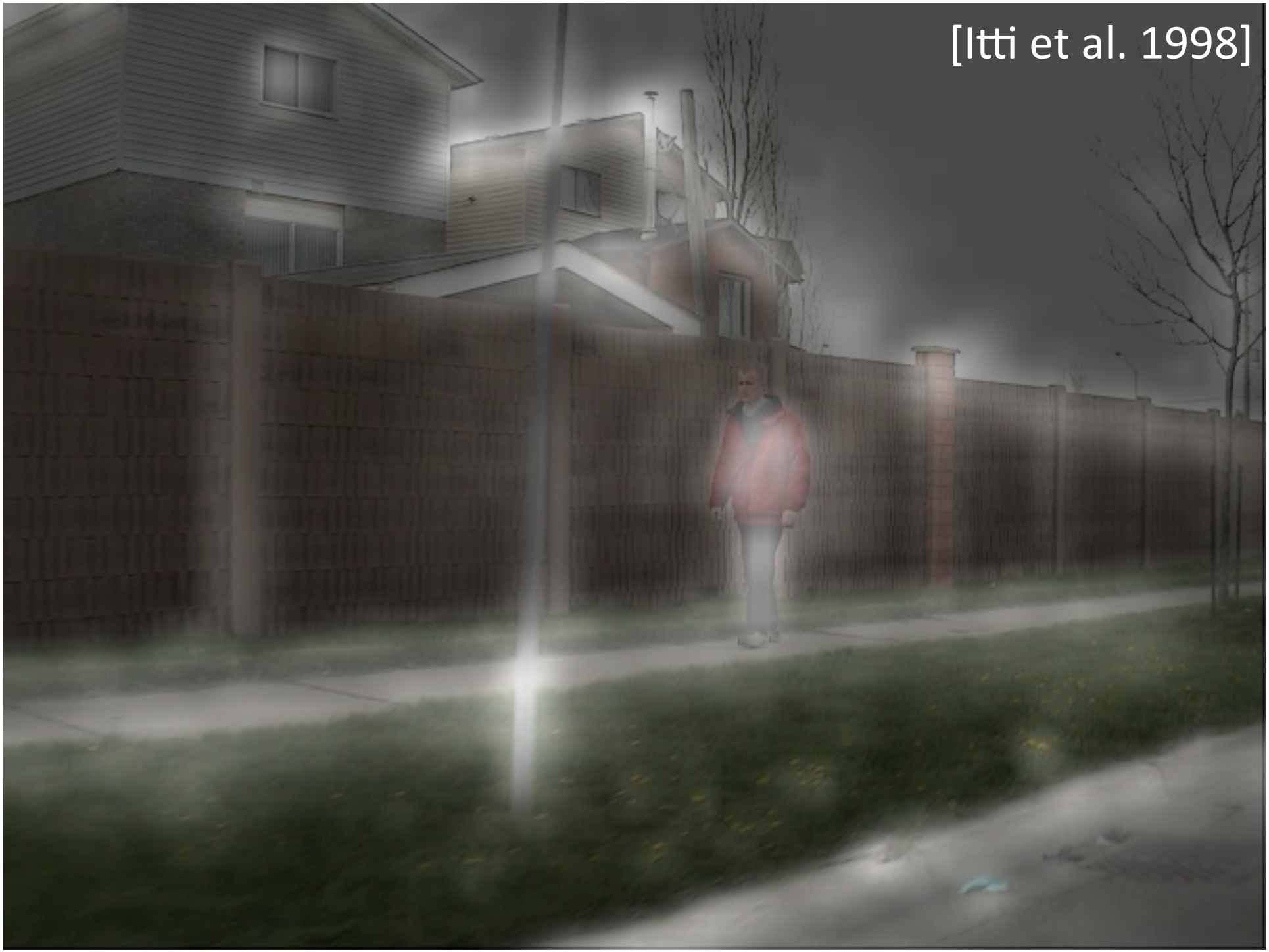


Perfect saliency map



Adapted from T. Judd


Density map from human fixation data


[Itti et al. 1998]


[Harel et al. 2007]


[Hou and Zhang 2009]


[Goferman et al. 2010]


Our Model


Density map from human fixation data


Density map from human fixation data


[Itti et al. 1998]

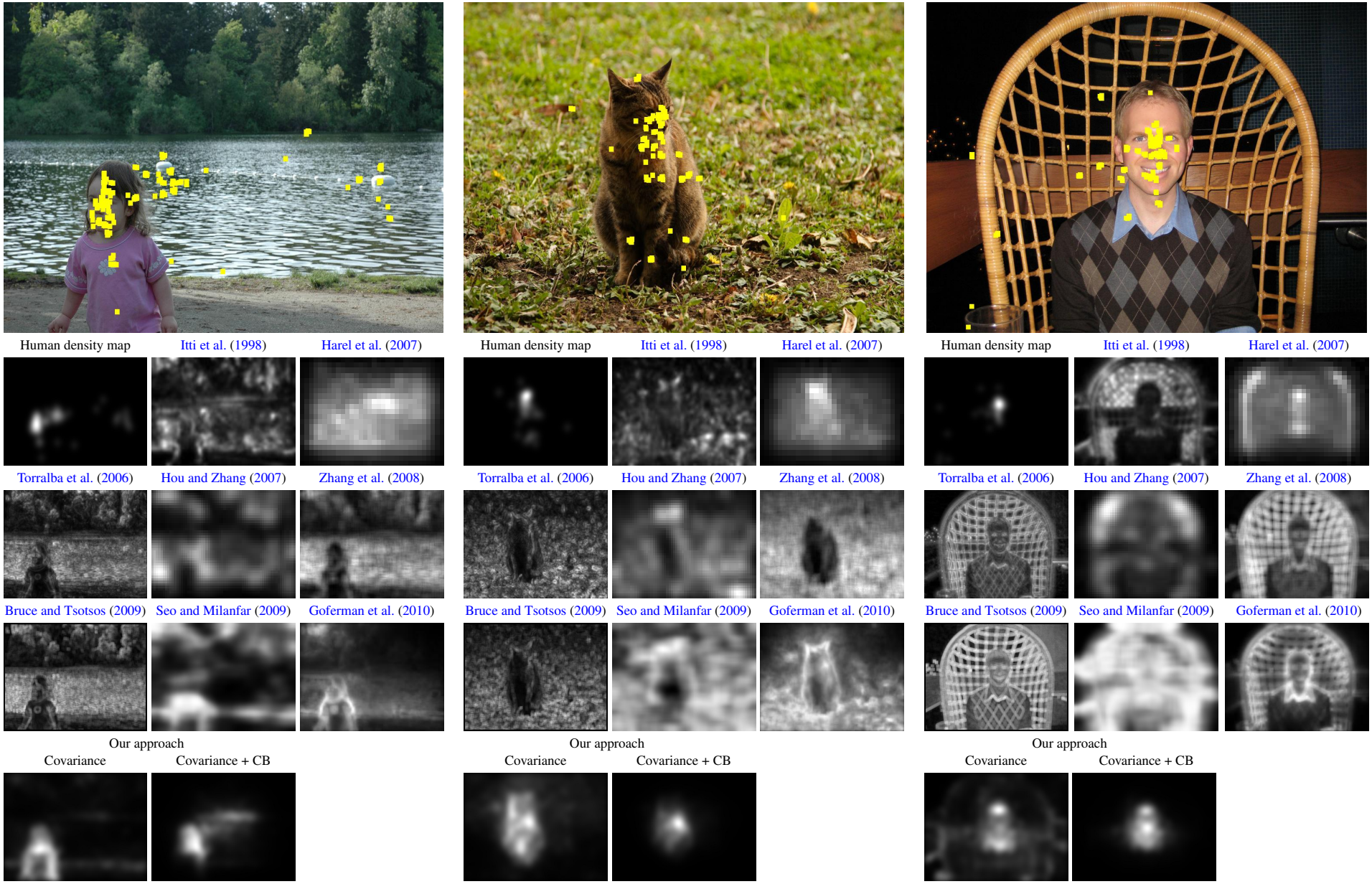

[Harel et al. 2007]


[Hou and Zhang 2009]


[Goferman et al. 2010]


Our Model

Some other comparative results



Understanding attention enables applications in computer graphics & vision, design

- image cropping / thumbnailing
- image and video compression
- non photorealistic rendering
- scene understanding
- advertising and package design
- web usability
- localization / recognition
- object detection
- navigational assistance
- robot active vision
- surveillance systems
- assistive technology for blind or low-vision people

Detecting Salient Objects

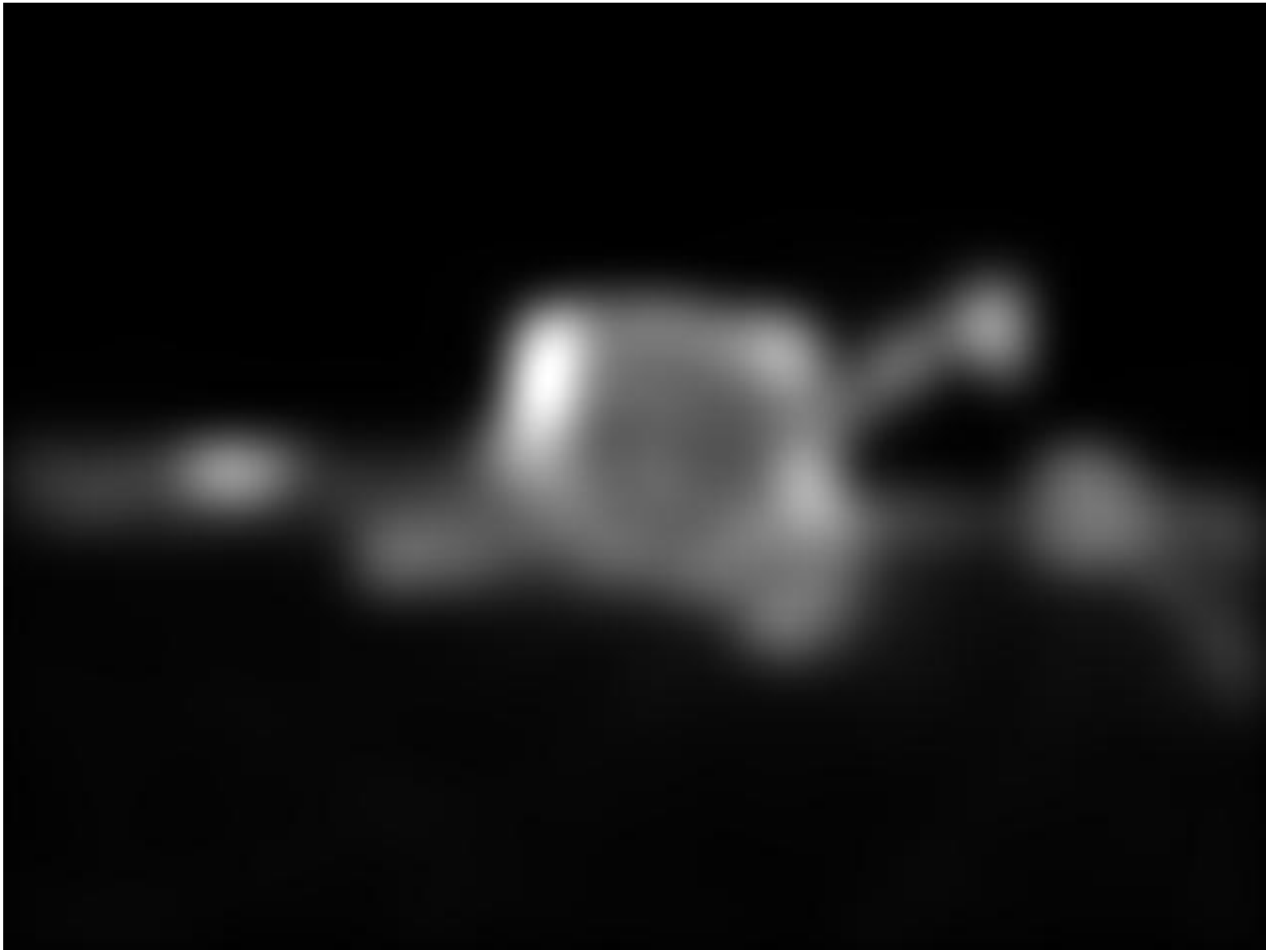
- The task is to detect foreground objects that attracts more attention in a given image

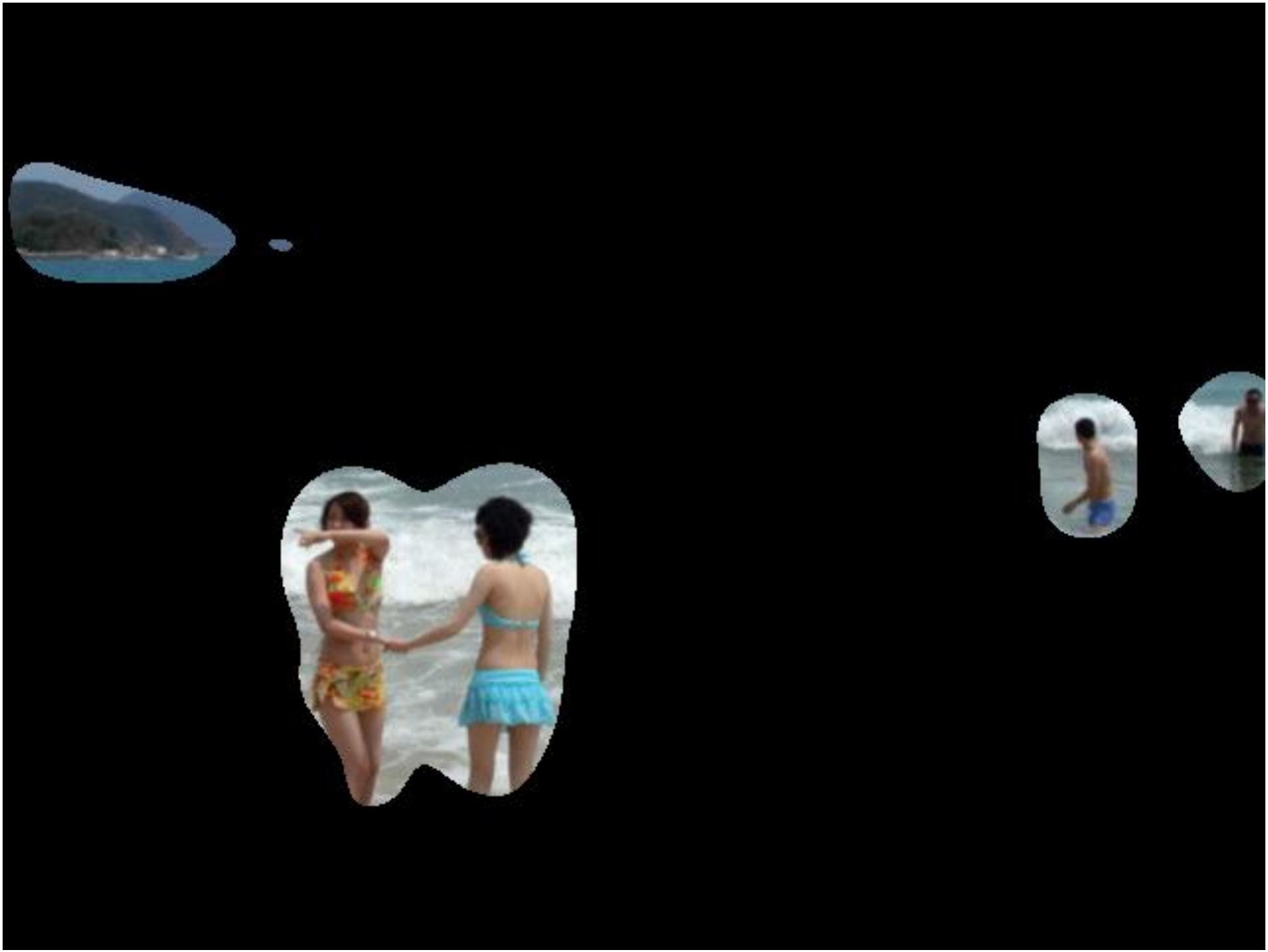
Detecting Salient Objects

- The task is to detect foreground objects that attracts more attention in a given image

Detecting Salient Objects

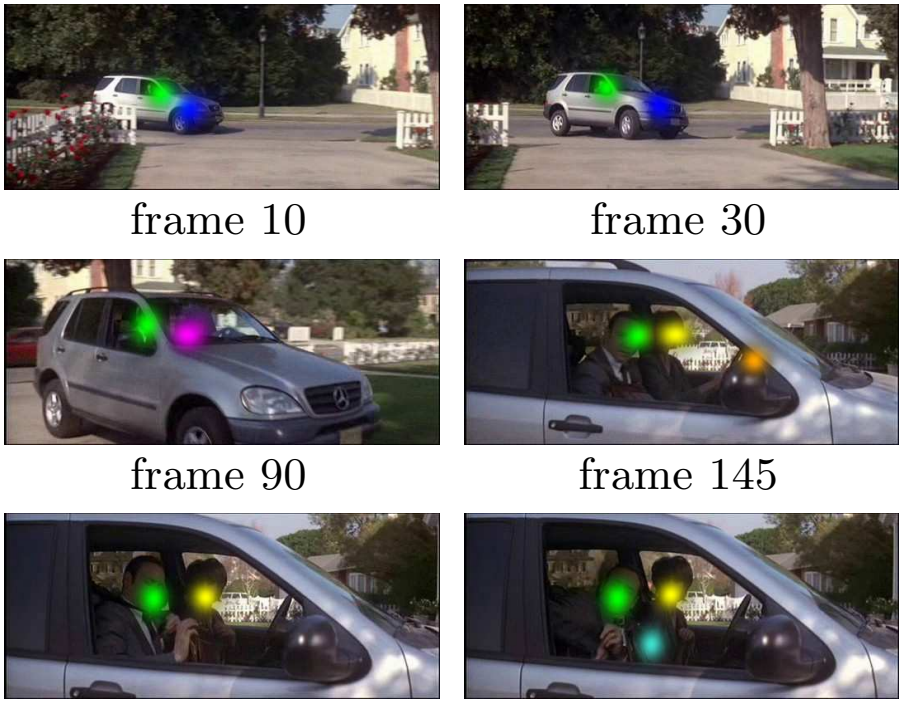
- ImgSal data set [Li, Levine, An, Xu & He, 2012]
- 235 natural images of size 480 x 640 pixels
- 19 subjects are asked to label the most salient objects in the images presented to them
- 6 different categories
  - large salient regions,
  - intermediate salient regions,
  - small salient regions,
  - cluttered backgrounds,
  - repeating distractors,
  - large and small salient regions)



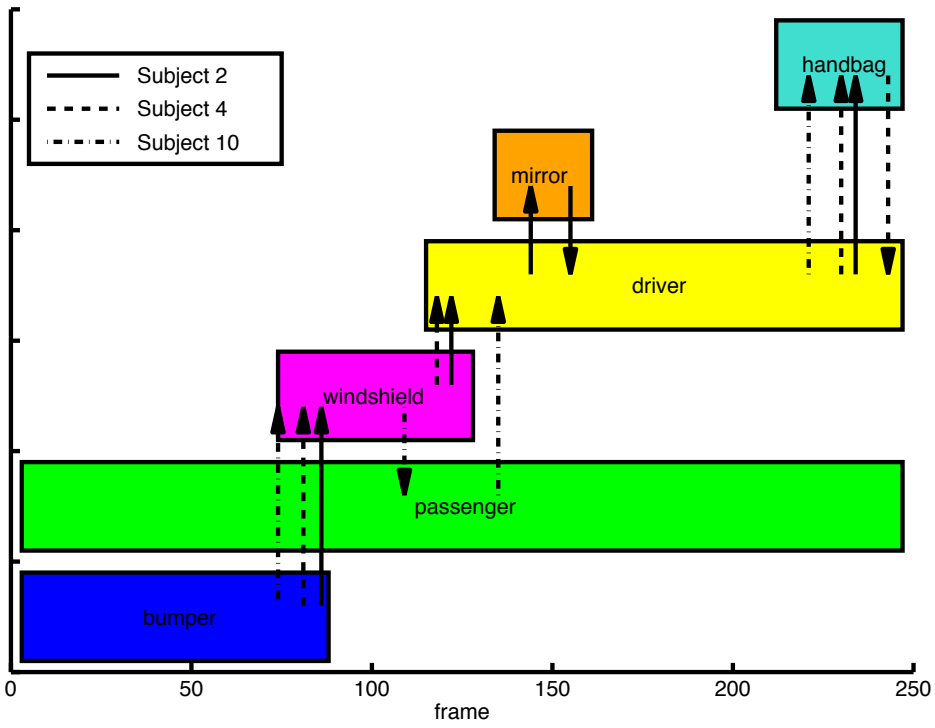


Action Recognition and Gaze

- Predicted human fixations can improve action recognition performances (Mathe and Sminchisescu, ECCV 2012)



fixations



Scan path

frame 25 frame 50 frame 90 frame 150

frame 15 frame 40 frame 60 frame 90

frame 10 frame 35 frame 70 frame 85

frame 3 frame 10 frame 18 frame 22

frame 4 frame 12 frame 25 frame 31

Summary

- Can machines predict where the humans look at a given image?
- State-of-the-art computational models of visual attention reach a certain point of success.
- Selection mechanisms provided by the computational saliency models could improve several machine perception applications.

Thanks for your attention!