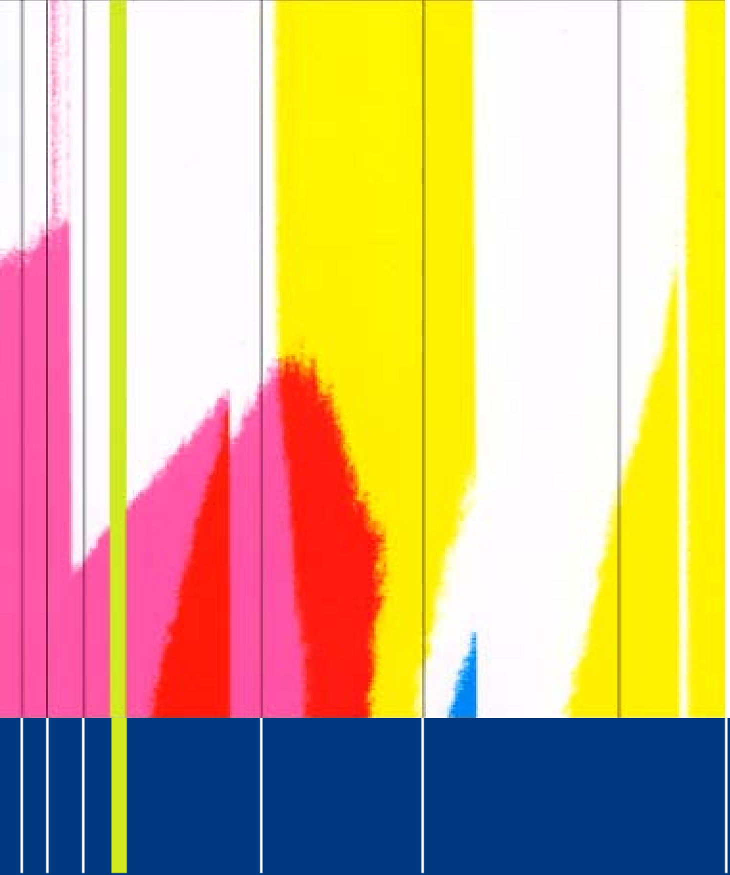


On the role of context in probabilistic models of visual saliency

Date

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

Neil Bruce, Pierre Kornprobst

NeuroMathComp Project Team,
INRIA Sophia Antipolis, ENS Paris, UNSA, LJAD

Overview

What is saliency?

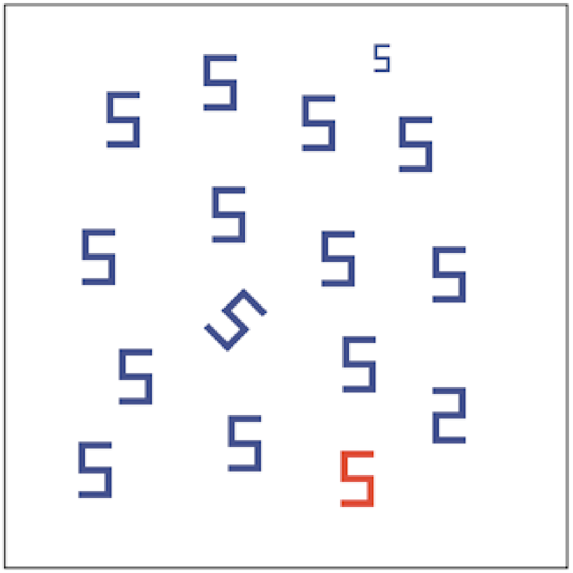
Modeling saliency

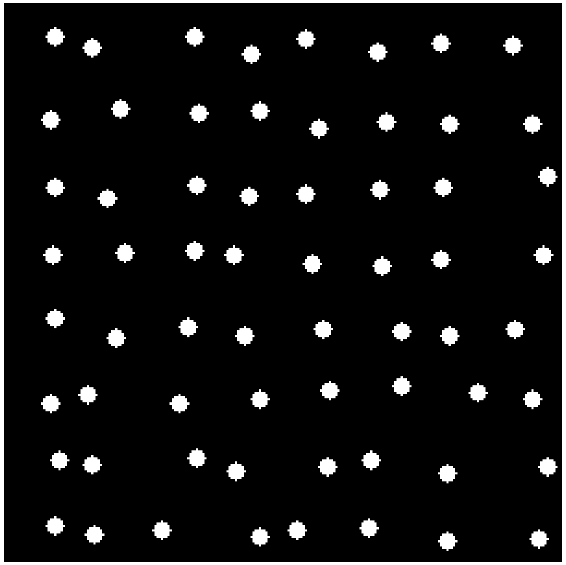
Probabilistic Models

Definitions of the support region

Future directions and conclusions

What is saliency?





What is saliency?

What is saliency?

- Visual content that is conspicuous, seemingly causing the automatic and immediate deployment of attention independent of task
- In computer vision, ROI selection allows focused processing on a subset of visual input overcoming complexity of visual search
- Saliency is one element that dictates where attention is to be focused (for people and machines)

Models of saliency

Models of saliency

- Purpose

Two categories:

i. To indicate what is of interest in an image / video

ii. To describe how saliency is implemented in humans

- Models

Three categories:

i. Inspired by observation (quantitative or qualitative) of how saliency is achieved in primates (behavioral or structural)

ii. Derived: Generally resulting in an expression with probabilistic terms

iii. Based on Image/signal processing principles

Models of saliency

- Purpose

Two categories:

1. i. To indicate what is of interest in an image / video
2. ii. To describe how saliency is implemented in humans

- Models

Three categories:

i. Inspired by observation (quantitative or qualitative) of how saliency is achieved in primates (behavioral or structural)

ii. Derived: Generally resulting in an expression with probabilistic terms

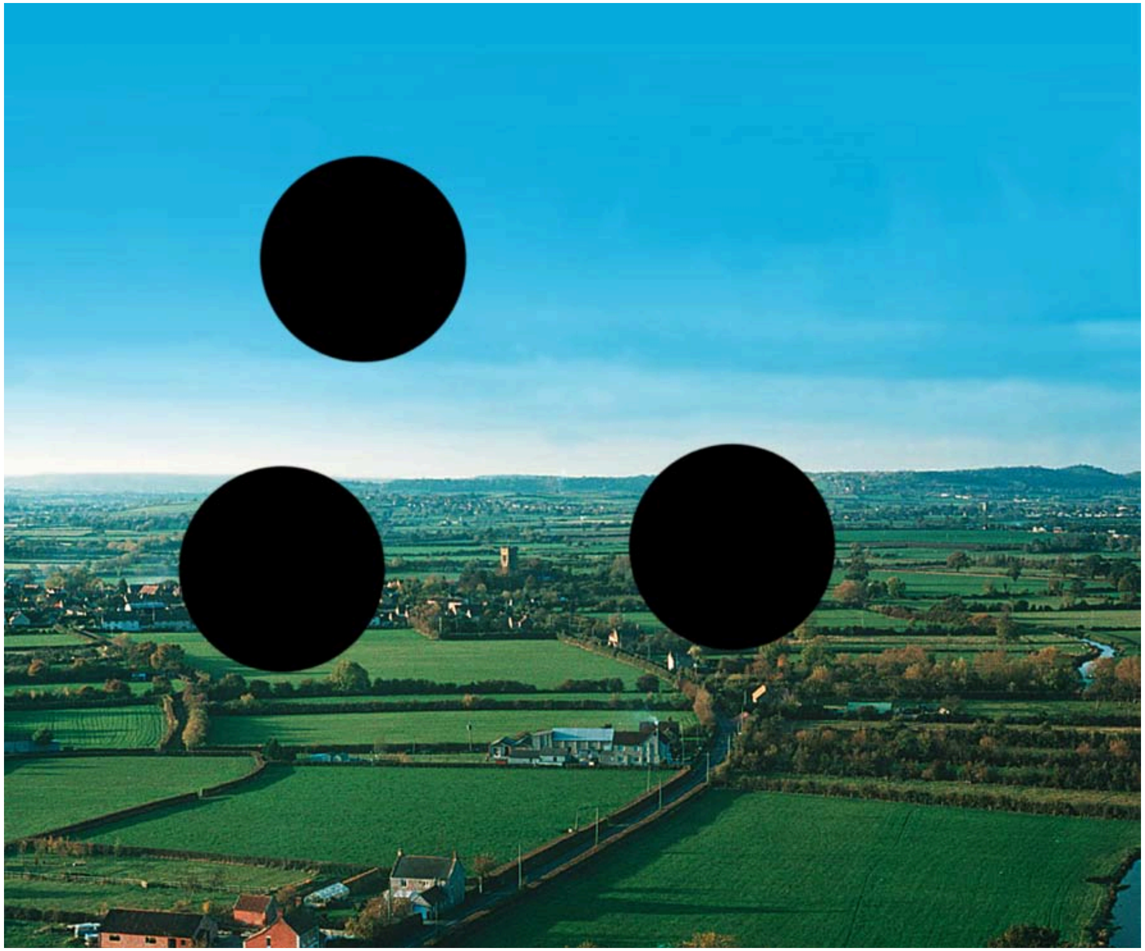
iii. Based on Image/signal processing principles

Probabilistic models of saliency

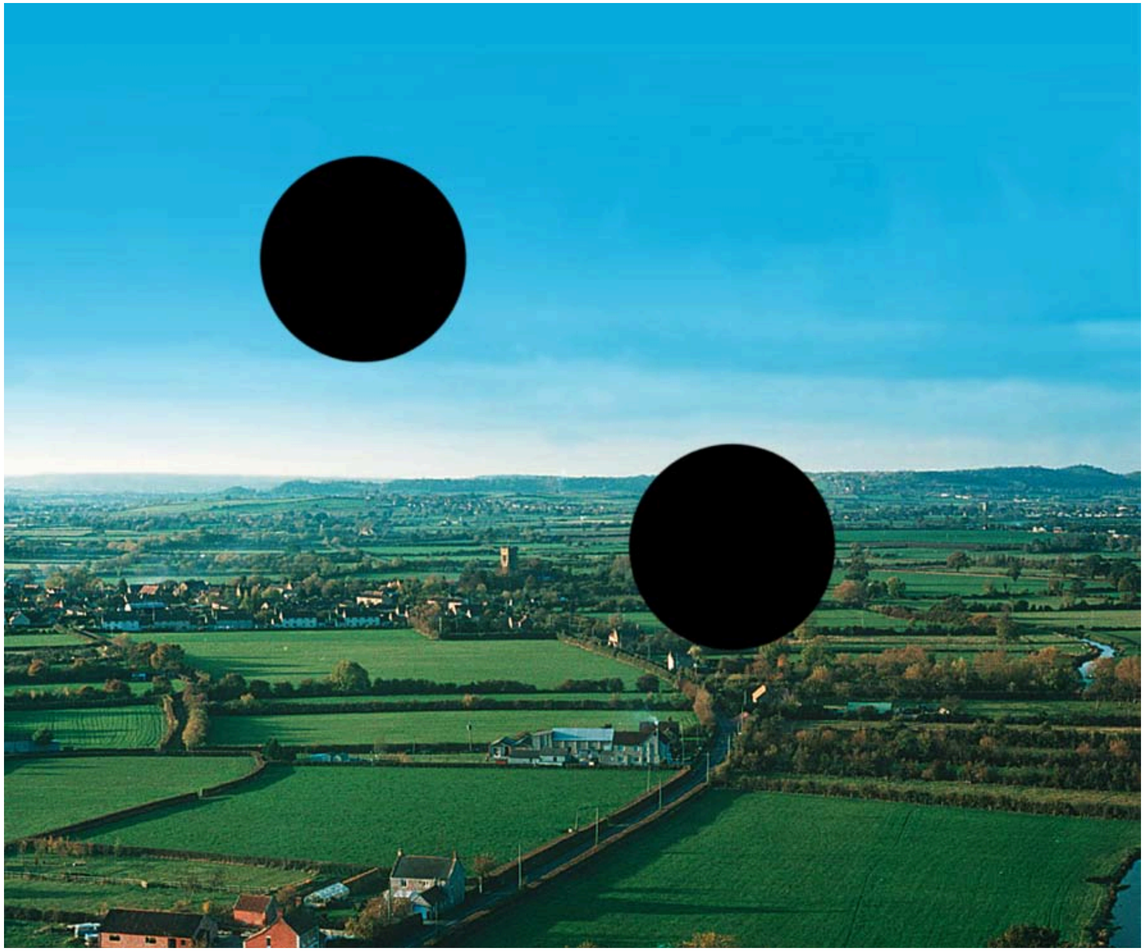
INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

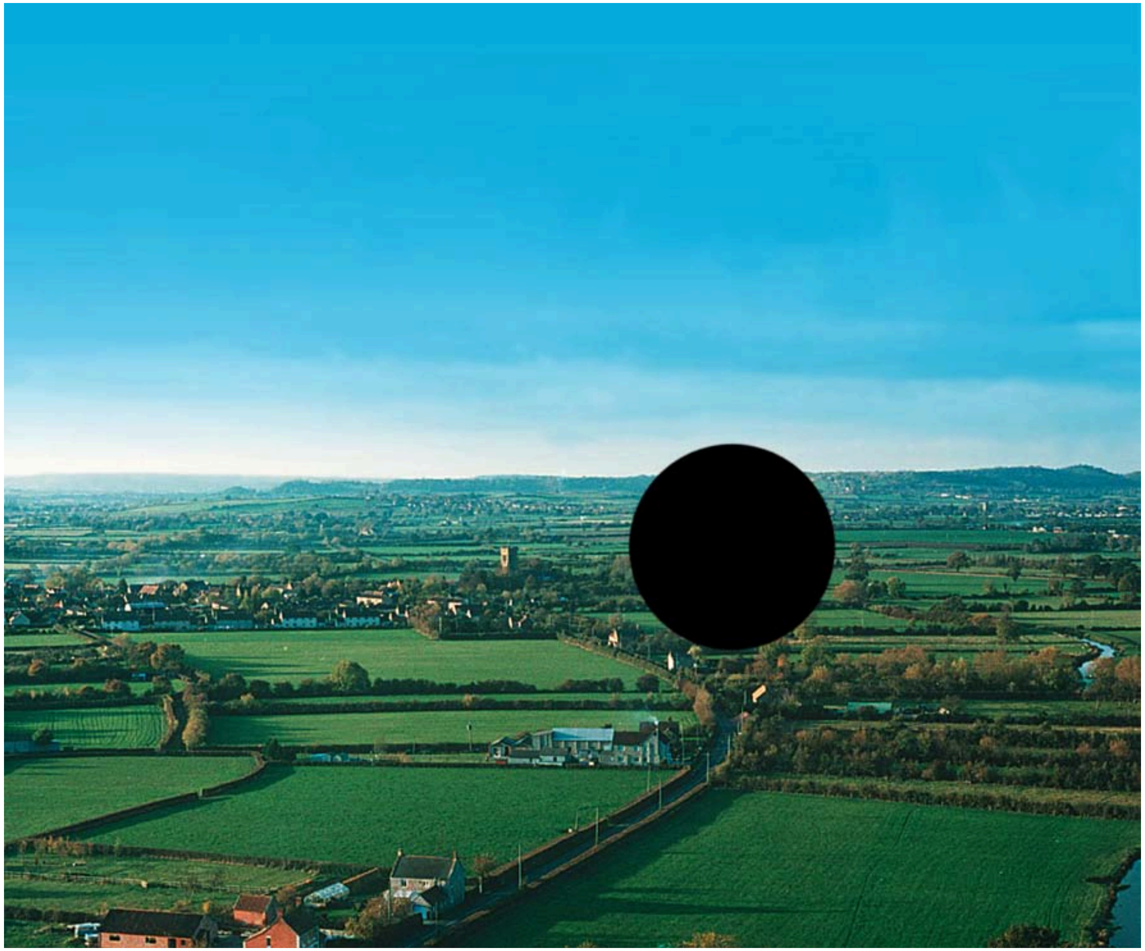
Visual content and expectation



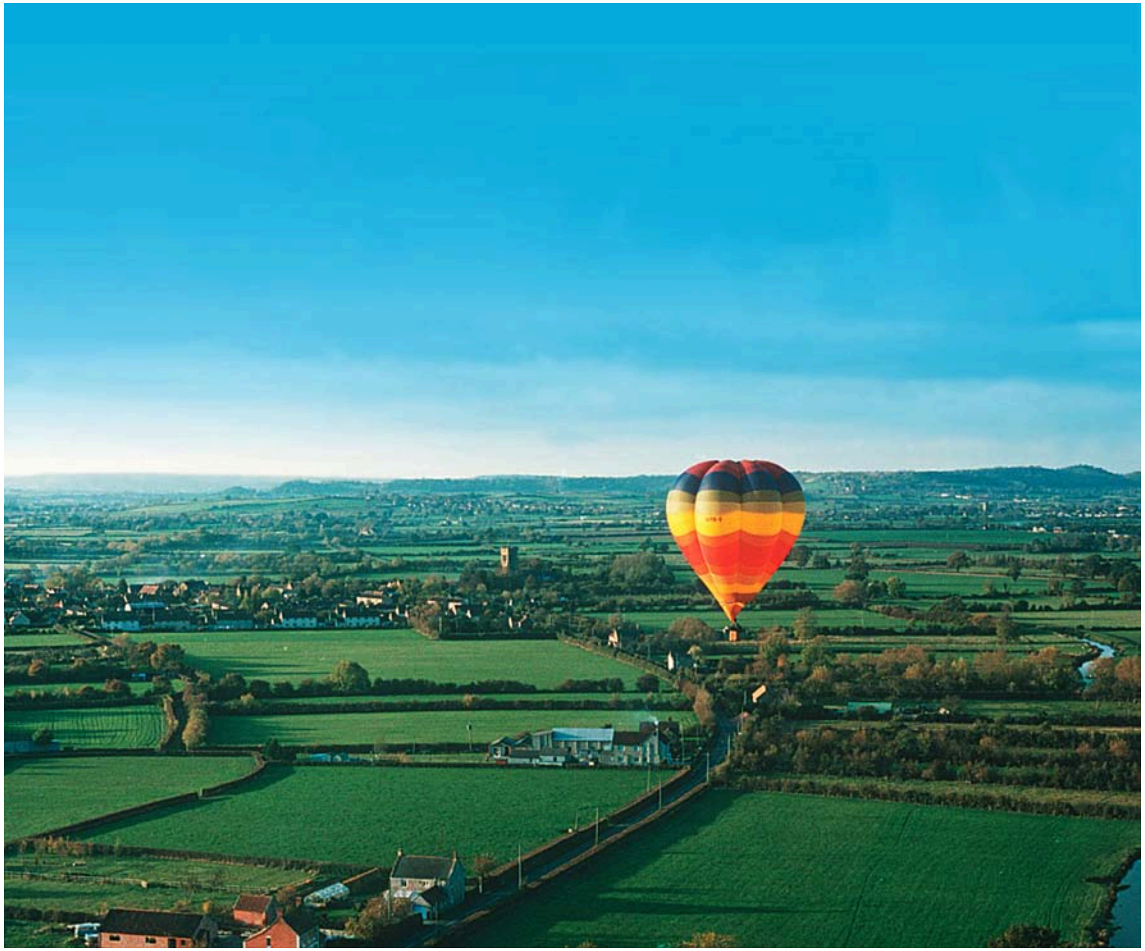
Visual content and expectation



Visual content and expectation

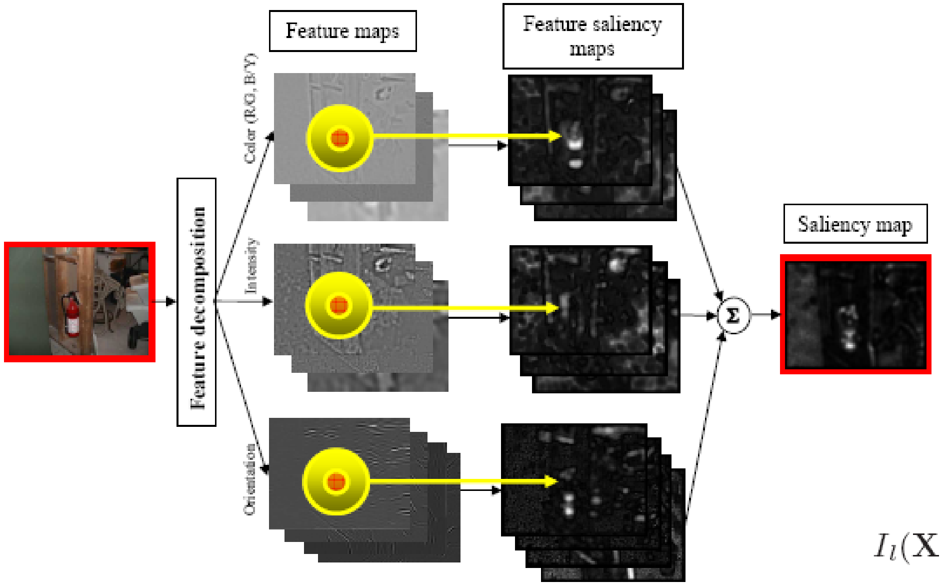


Visual content and expectation



Discriminant Saliency

- Mutual information between set of features X and class variable Y
- Y distinguishes centre from surround in bottom up case



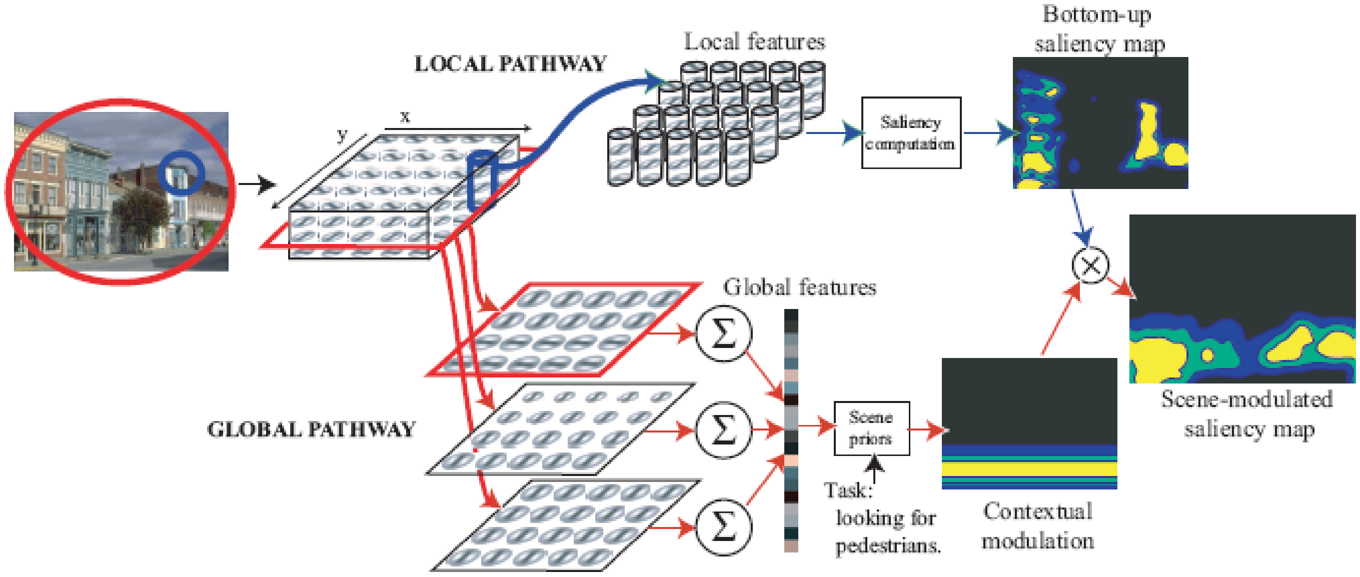
$$I_l(\mathbf{X};Y) = \sum_c \int p_{\mathbf{X}(l),Y(l)}(\mathbf{x},c) \log \frac{p_{\mathbf{X}(l),Y(l)}(\mathbf{x},c)}{p_{\mathbf{X}(l)}(\mathbf{x})p_{Y(l)}(c)} d\mathbf{x}$$

Gao and Vasconcelos, 2007

Bayesian Approaches

- Gist based processing

$$p(O,\ X|L,\ G) \qquad S(X) = \frac{1}{p(L|G)} p(X|O = 1,\ G)$$



Torralba et al., 2006

Bayesian Approaches

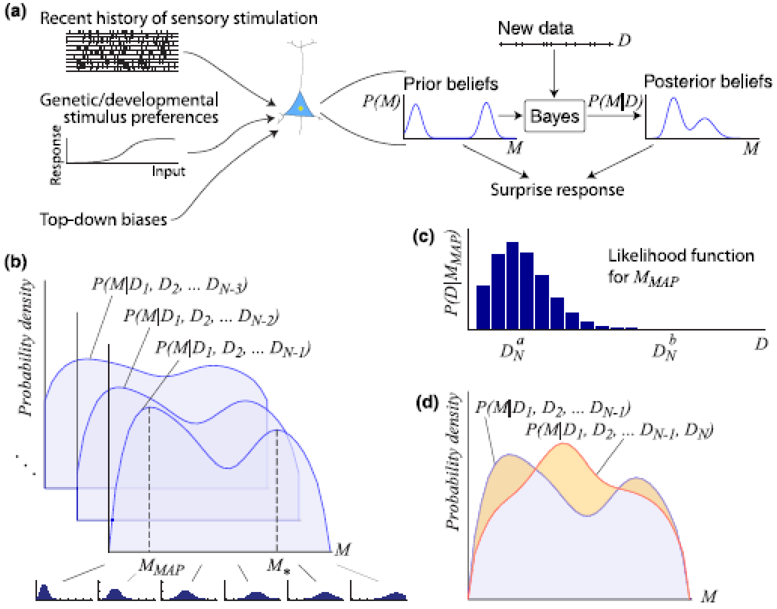
- SUN
- Saliency Using Natural Image Statistics (but see also Bruce 2004)

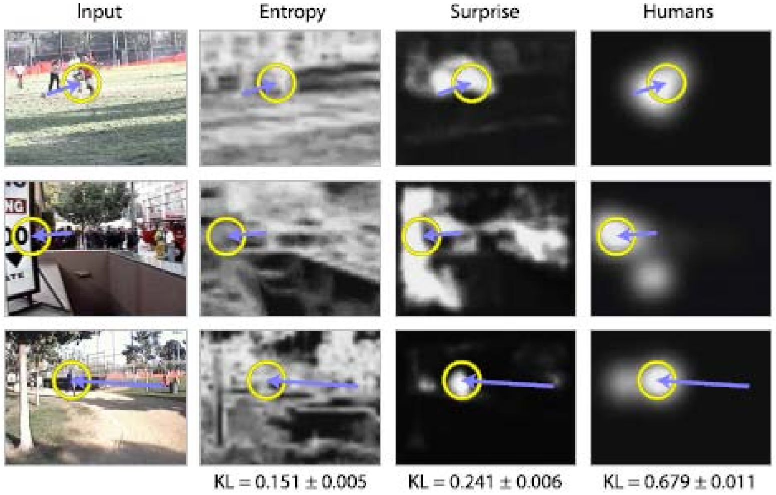
$$s_z = p(C = 1|F = f_z, L = l_z)$$

$$\underbrace{\frac{1}{p(F = f_z)}}_{\substack{\text{Independent}\\ \text{of target}\\ \text{(bottom-up saliency)}}} \underbrace{\underbrace{p(F = f_z|C = 1)}_{\text{Likelihood}}\,\underbrace{p(C = 1|L = l_z)}_{\text{Location prior}}}_{\substack{\text{Dependent on target}\\ \text{(top-down knowledge)}}}$$

Zhang et al., 2008

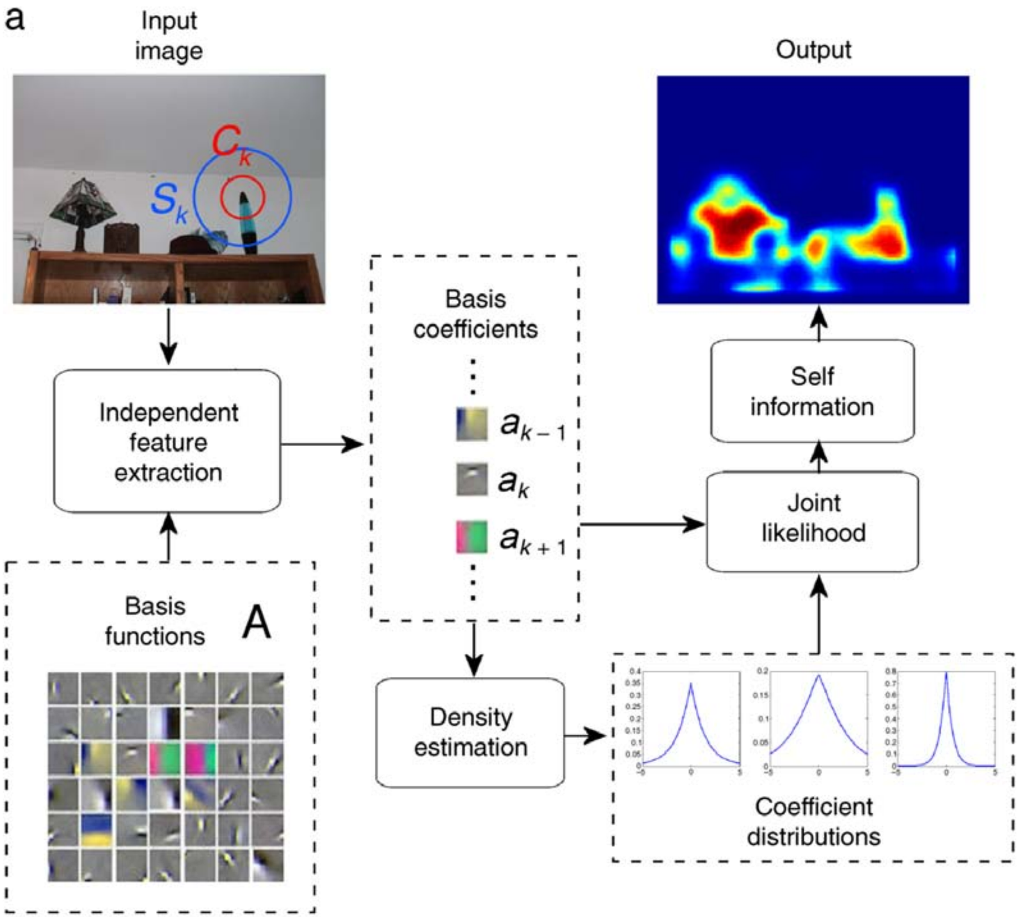
Surprise





Itti and Baldi, 2006

AIM: Attention by Information Maximization



Bruce and Tsotsos, 2006, 2009

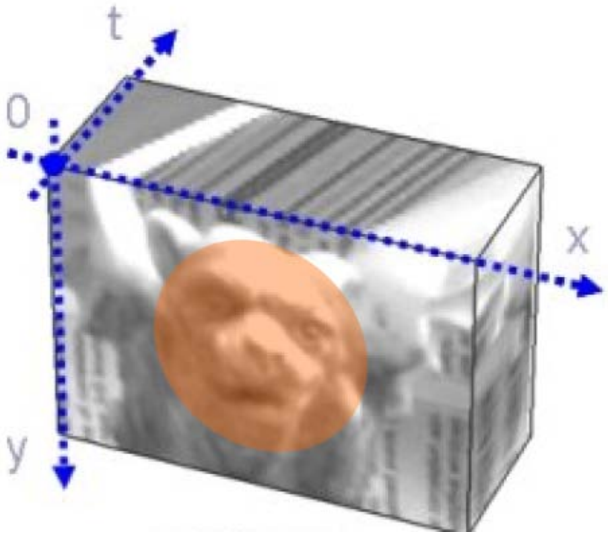
Definitions of the support region

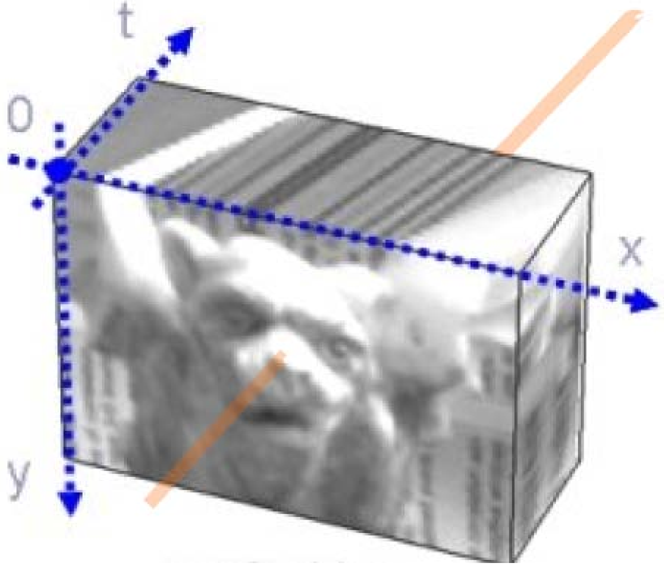
What is the estimate of local content based on?

- **What is the space that defines the estimate of local content?**
- Local spatial surround
  Gao and Vasconcelos 2007, Bruce and Tsotsos 2009
- Whole image
  Bruce and Tsotsos 2006, Torralba et al. 2006
- Natural image statistics
  Bruce 2004, Zhang et al. 2008
- Temporal history
  Itti and Baldi, 2006

Spatiotemporal Extent of Context

- Goal is to provide some examples of how context shapes model behavior
- This may be applied to any probabilistic definition

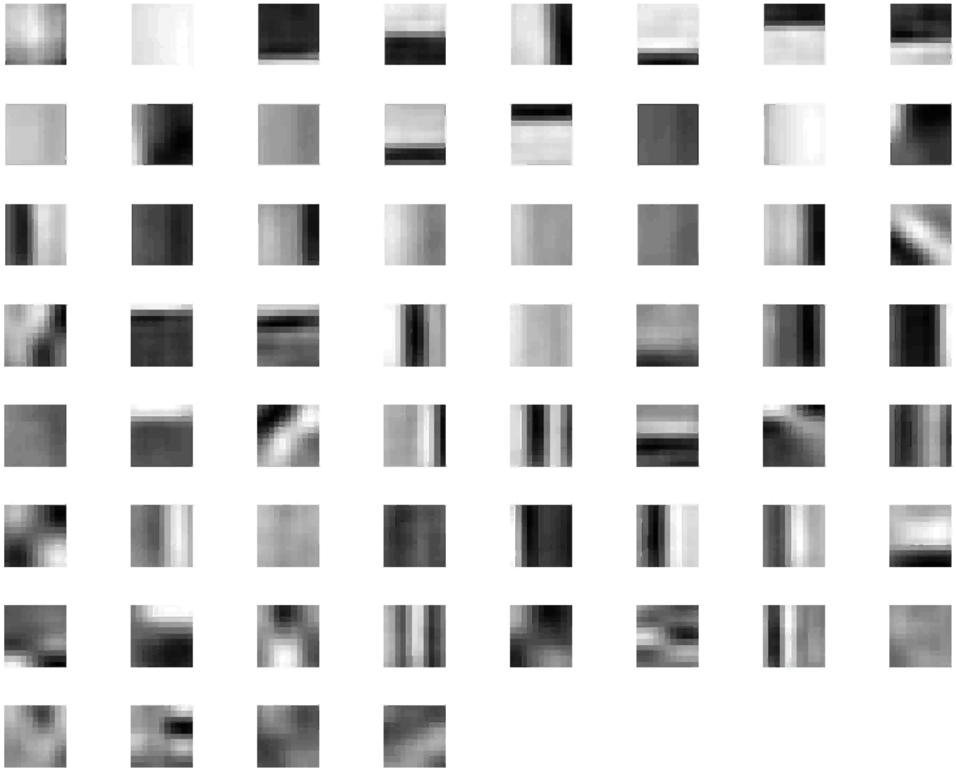




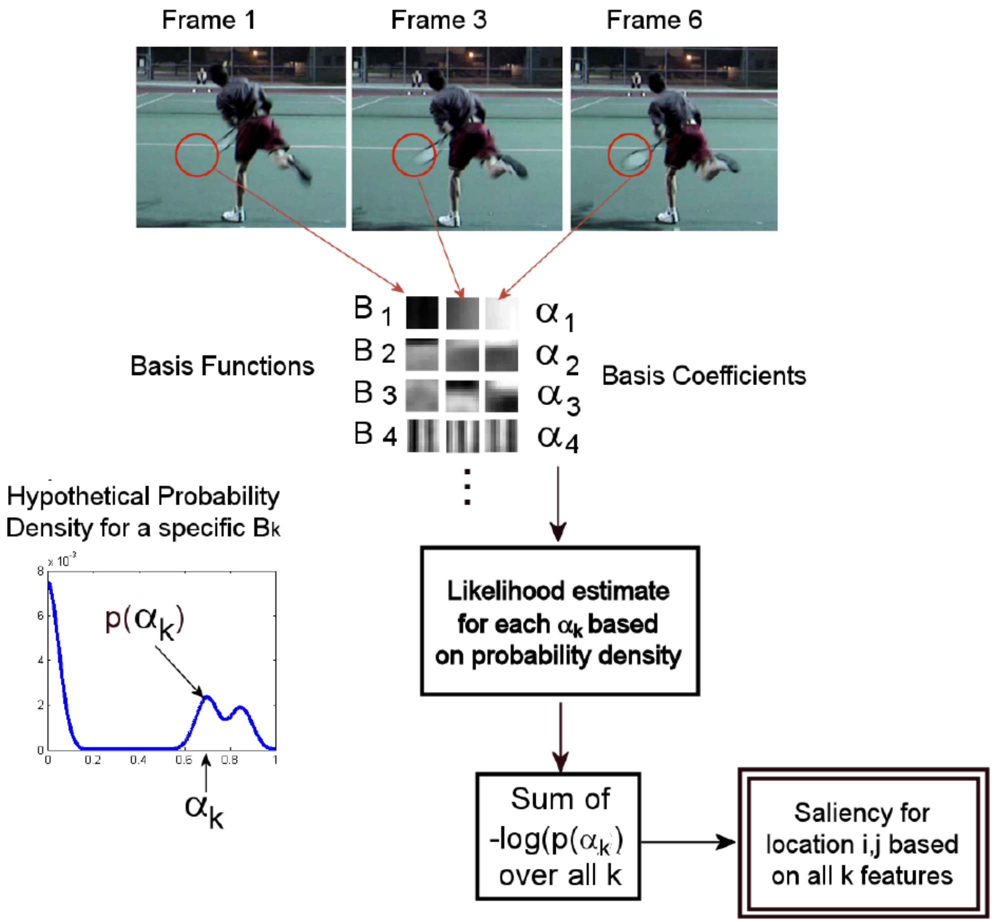
Representing P(L,F) jointly

- Similar to the case of Bruce & Tsotsos 2009, Gao & Vasconcelos, 2007 except support is not purely spatial
- Can do this explicitly, retaining a probability density function for each location
- Requires significant memory since the temporal extent may require storing a representation of a PDF for each pixel parameterized or quantized

Spatiotemporal Cells



Spatiotemporal Features

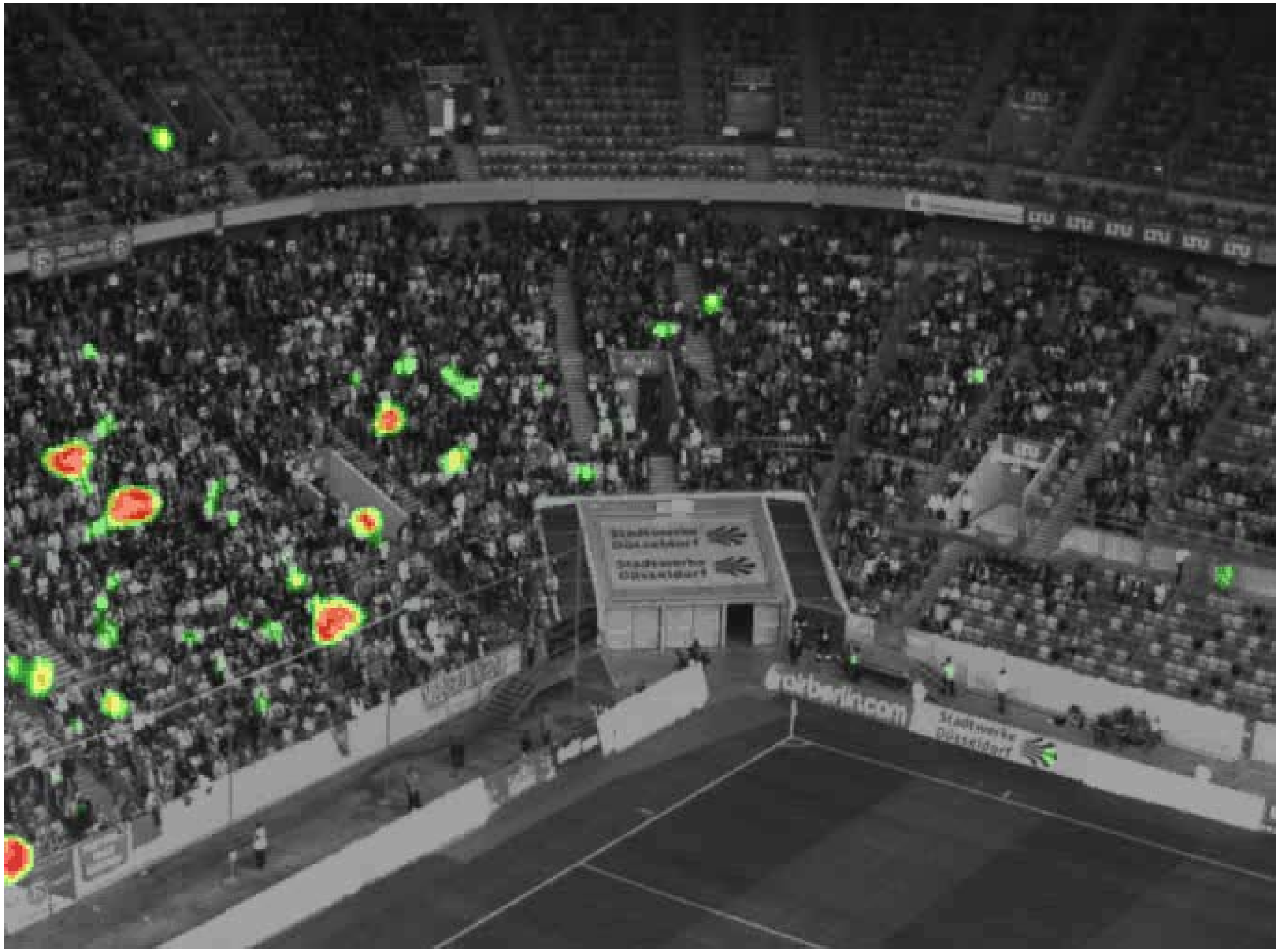


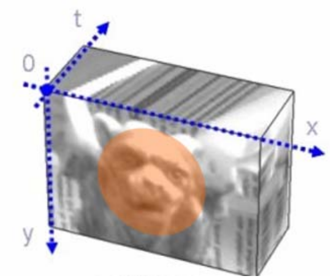
Likelihood estimate for each α_k based on probability density

Sum of $-\log(p(\alpha_k))$ over all k

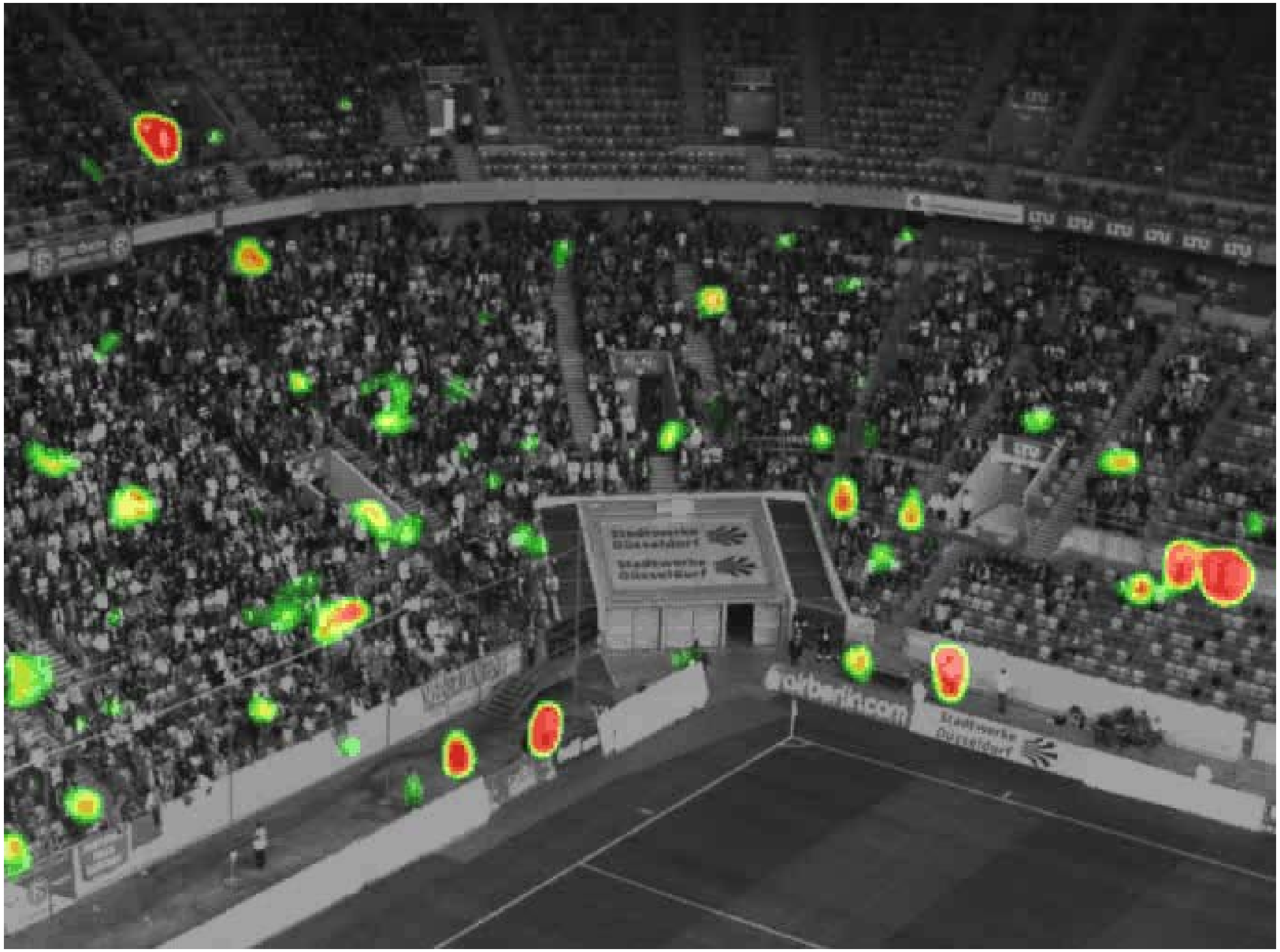
Saliency for location i,j based on all k features

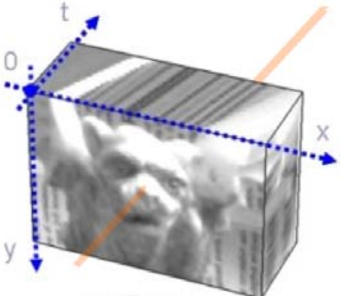
Representing P(L,X) jointly



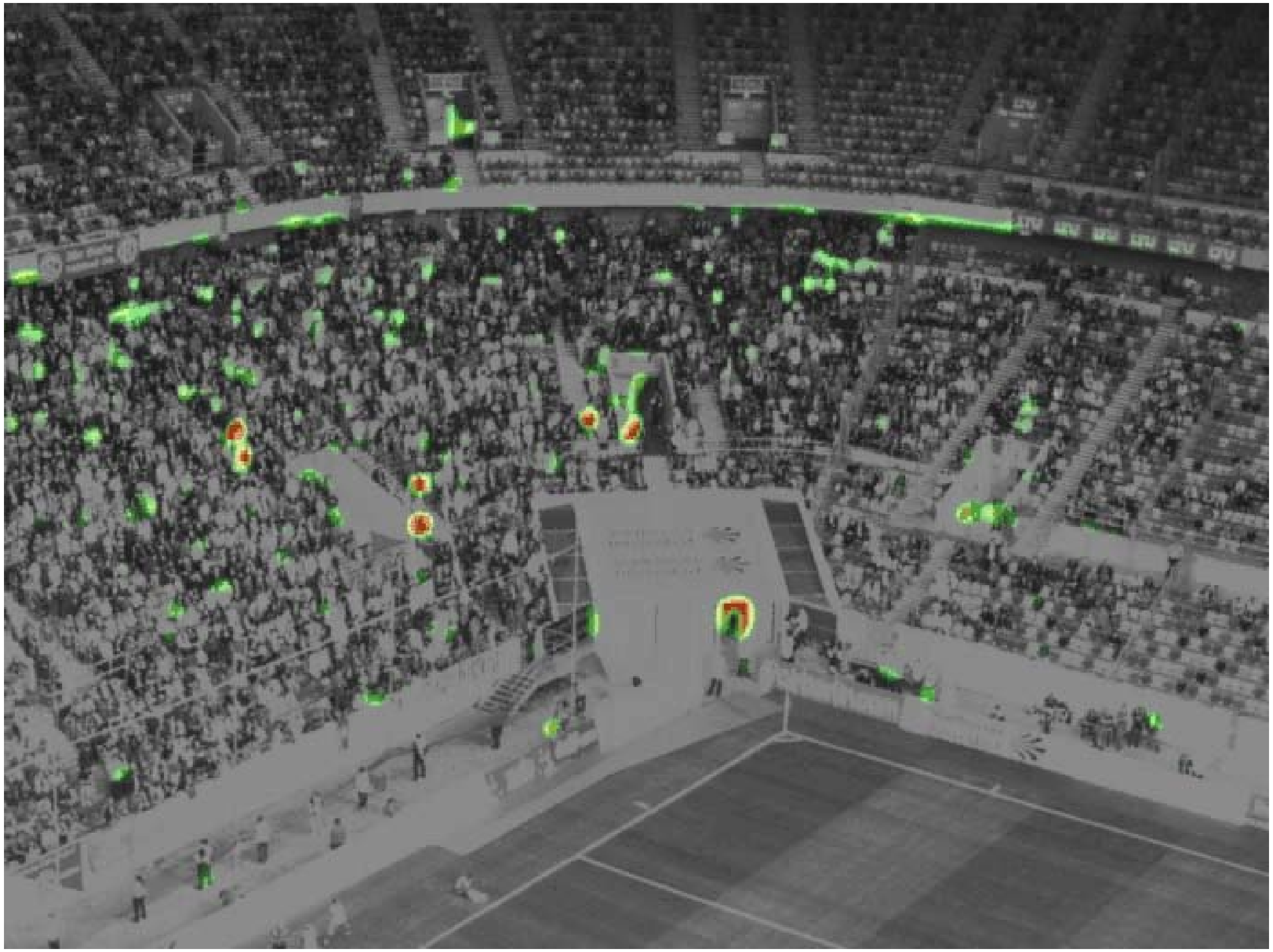


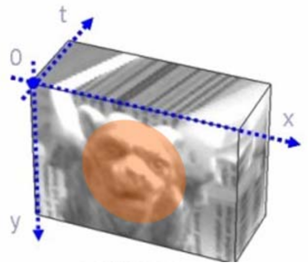
Representing P(L,X) jointly



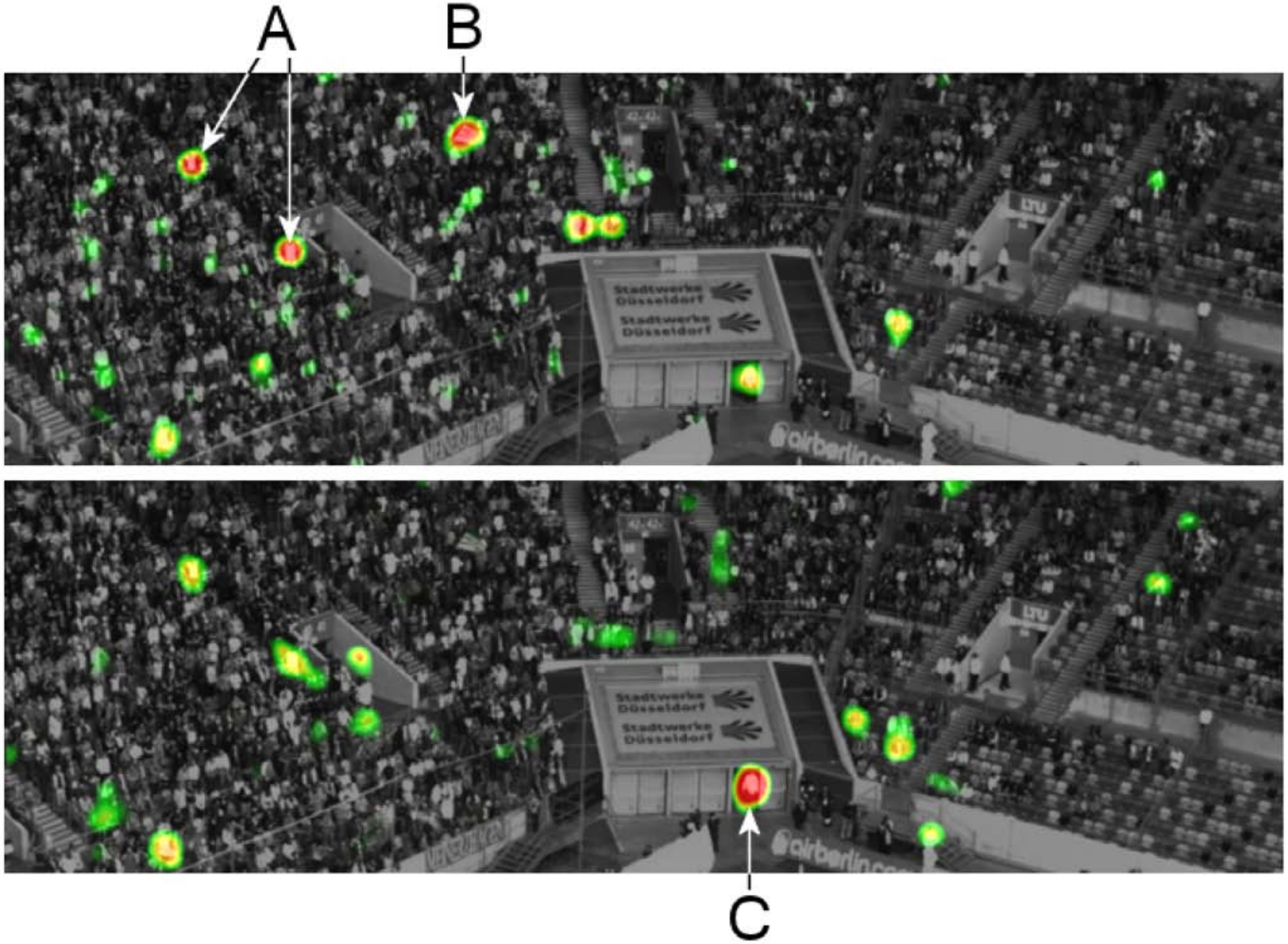


Representing P(L,X) jointly



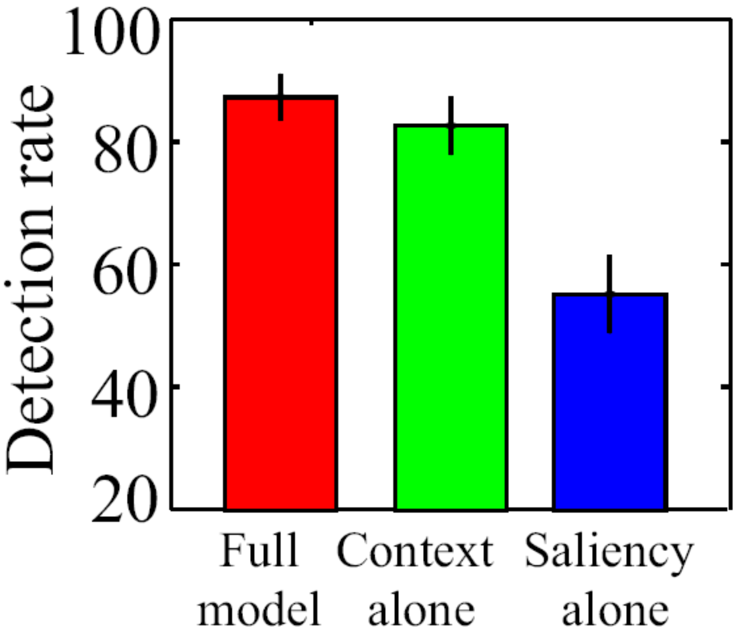


A single frame difference



Gist

- Global receptive fields that capture the “gist” of the scene



Torralba et al., 2006

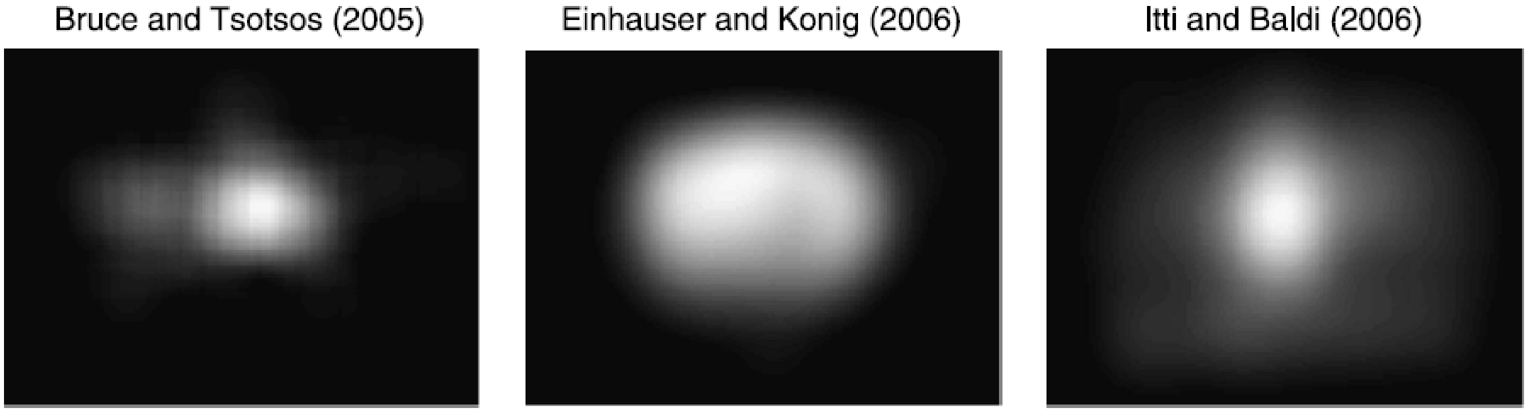
Separability of location and features

- Zhang et al., 2008 approach

$$\log s_z = \underbrace{-\log p(F = f_z)}_{\text{Self-information}} + \underbrace{\log p(C = 1 | L = l_z)}_{\text{Location prior}}.$$

- Allows influence of location to be modeled explicitly
  e.g. central bias – More on this later

Central bias



Bruce and Tsotsos (2005)

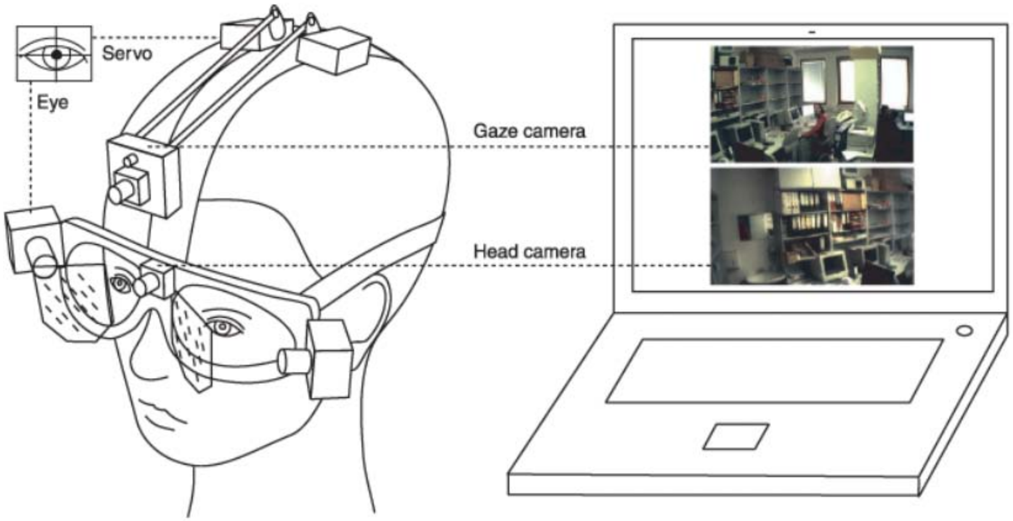
Einhauser and Konig (2006)

Itti and Baldi (2006)

Zhang and Cottrell, 2008

- Some authors have noted that fixation data tends to have a strong central bias (Le Meur et al. 2006, Zhang et al. 2008)
- When is it ok to assume a central bias?

Central bias



Schumann et al., 2009

- Eye in head movements tend to be upward biased
- … but, images do tend to be composed, and a centrally biased prior on saliency may be useful

Environmental Statistics

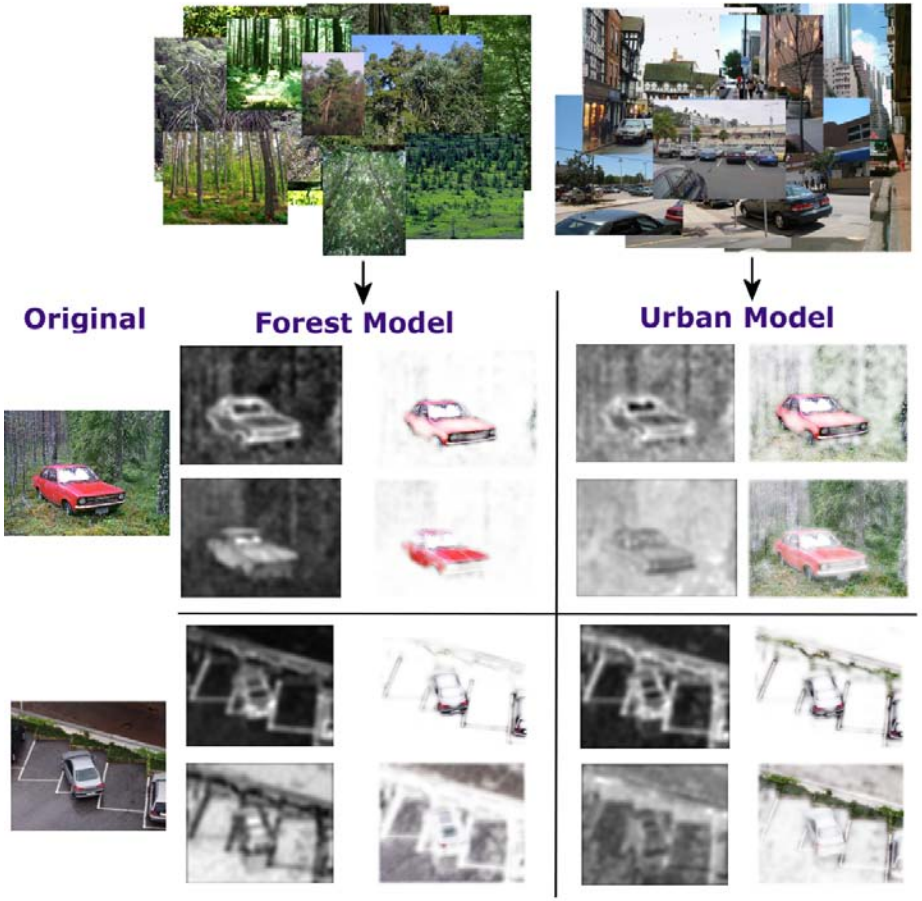
INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

Environmental Extent of Context

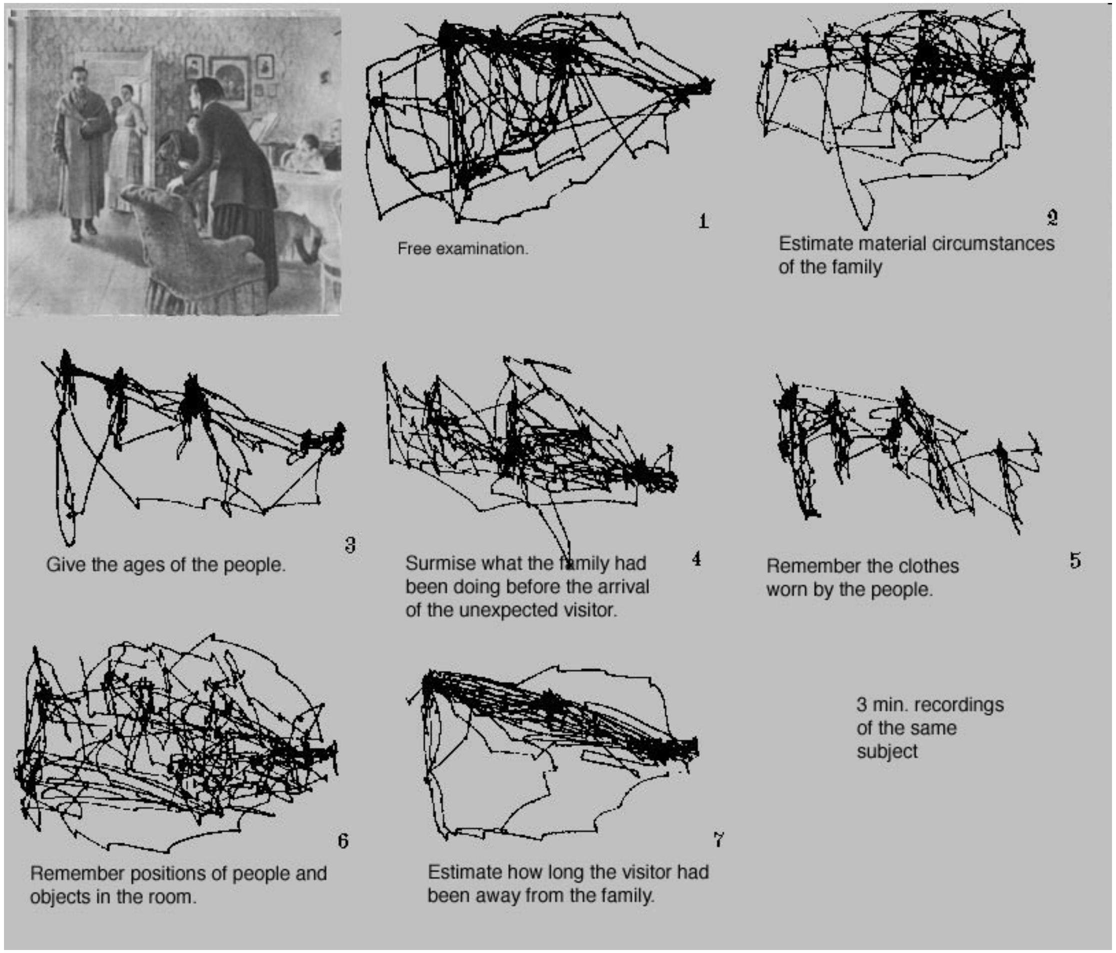
- Already mentioned various extents
  e.g. local surround, image, natural images
- As another example of how context may influence salience, can consider statistics of specific environments
- e.g. Forest, mountains, city, computer science building

Using Context



Future directions and Conclusions

Future Directions



Free examination.

Estimate material circumstances of the family

Give the ages of the people.

Surmise what the family had been doing before the arrival of the unexpected visitor.

Remember the clothes worn by the people.

Remember positions of people and objects in the room.

Estimate how long the visitor had been away from the family.

3 min. recordings of the same subject

Summary and Conclusions

- Adaptability in the contextual model (e.g. task) – At least from a general vision perspective
- Context includes more than just saliency
- Machine vision in general appears to be paying more attention to this problem (semantic labeling, image grammars)
- Saliency is one element that dictates where attention is to be focused (for people and machines) but not the only one