CE6902 Computer Vision

Visual Attention

Overview

- Visual Attention Basics
- Bottom-Up Visual Attention - Saliency
- Top-Down Visual Attention – Visual Search

Visual Attention Basics

- *What's Attention*
- *Bottom-Up Controls*
- *Top-Down Controls*

What's Attention

Everyone knows what attention is. It is the taking possession of the mind, in clear and vivid form, of one out of what seem several simultaneous possible objects or trains of thought. Focalization, concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrain state ....

– William James (1890)

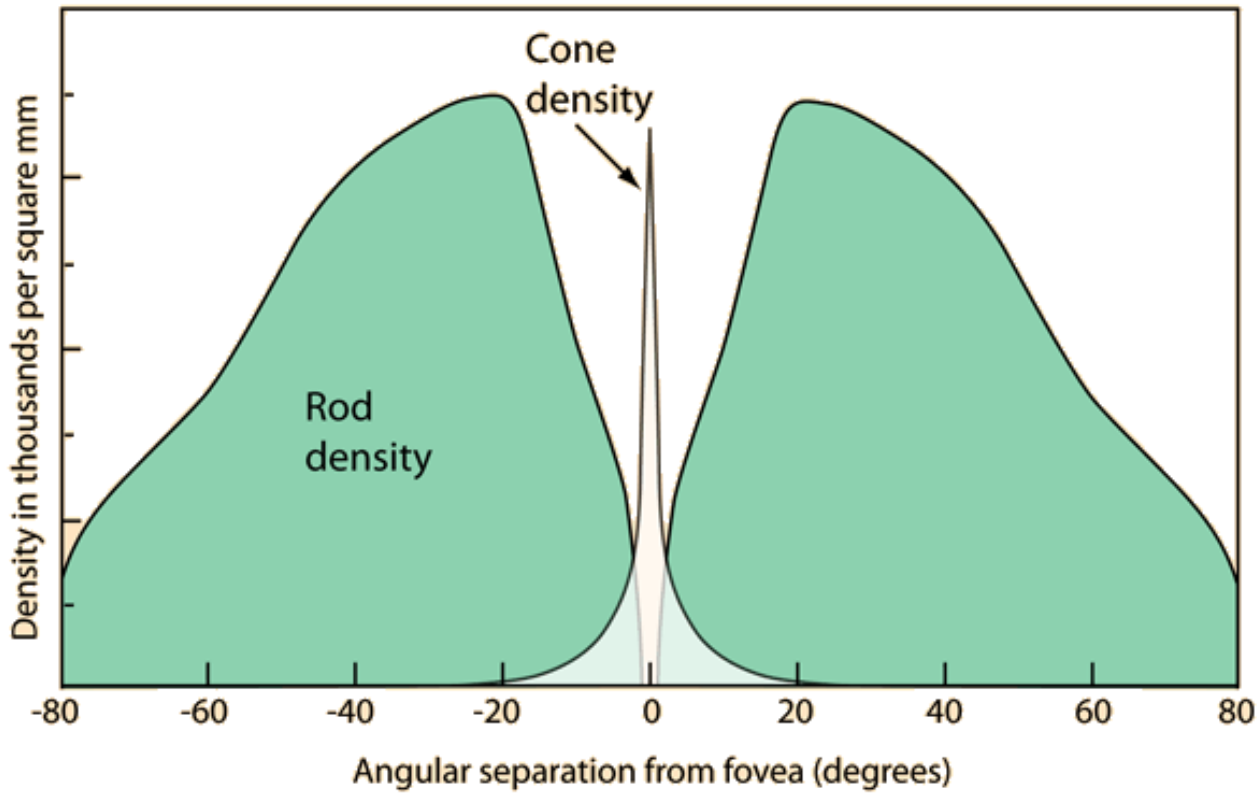
What's Visual Attention

Why do we need visual attention?

What's Visual Attention

Why do we need visual attention?

- *120 million rods (intensity/motion)*
- *7 million cones (color)*
- *Fovea: 2 degrees of cones*



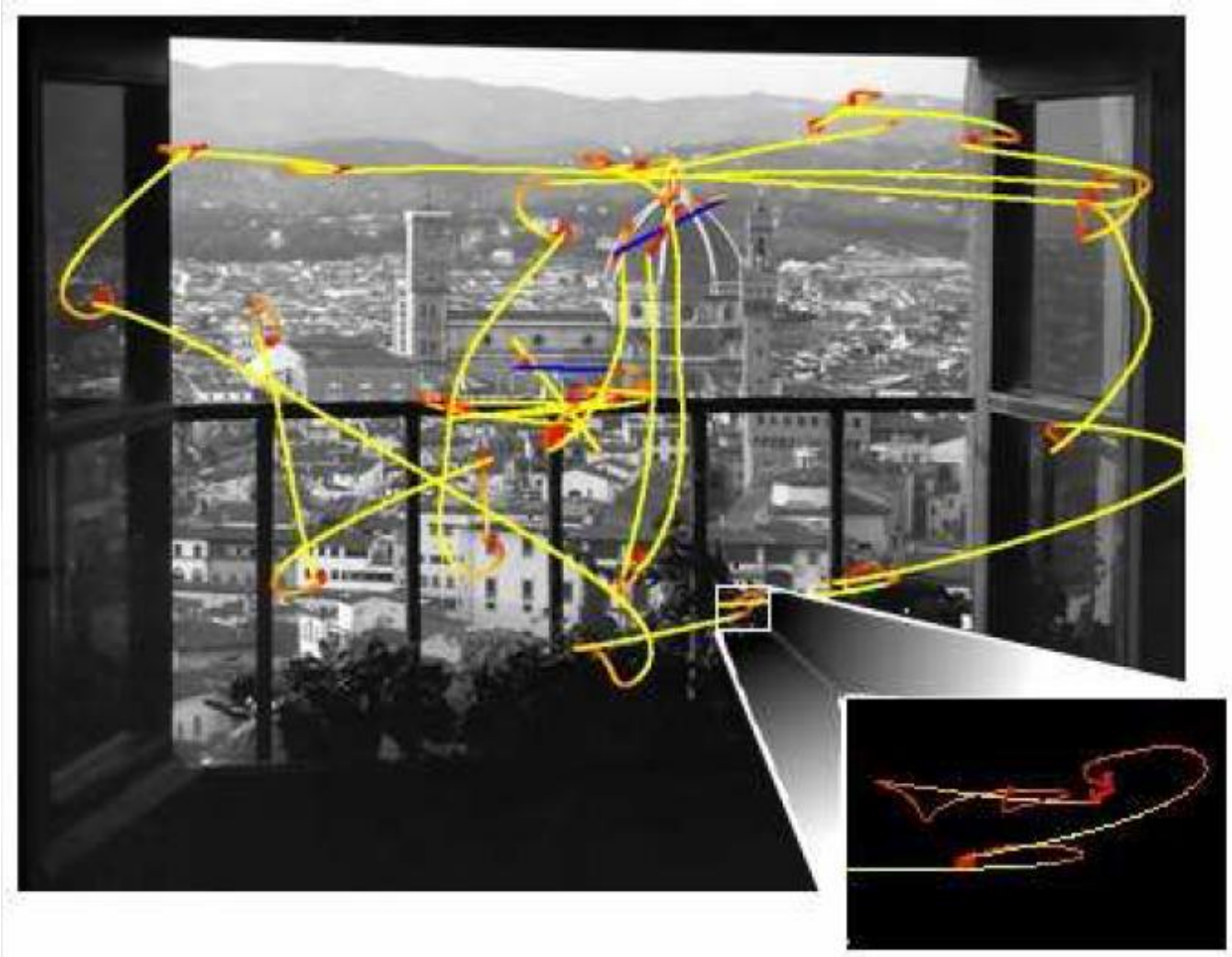
Bottom-Up Controls

What determines where we look? One control is bottom-up as determined by characteristics of the scene.

- *Stimulus salience - areas of stimuli that attract attention due to their properties*
  - ✓ *Color, contrast, and orientation are relevant properties*
  - ✓ *Saliency maps show that fixations are related to such properties in the initial scanning process*
- *Very fast (up to 20 shifts/s)*
- *Involuntary / automatic*

Bottom-Up Controls

***Saccades**: quick eye movements from one fixation to another.*



Bottom-Up Controls

Not all parts of a scene are sampled equally.

Top-Down Controls

What determines where we look? Another control is top-down which is determined by tasks/goals.

- *Where to attend (spatial attention)*
- *What features to attend to (feature-based attention)*

May target inconspicuous locations in visual scenes

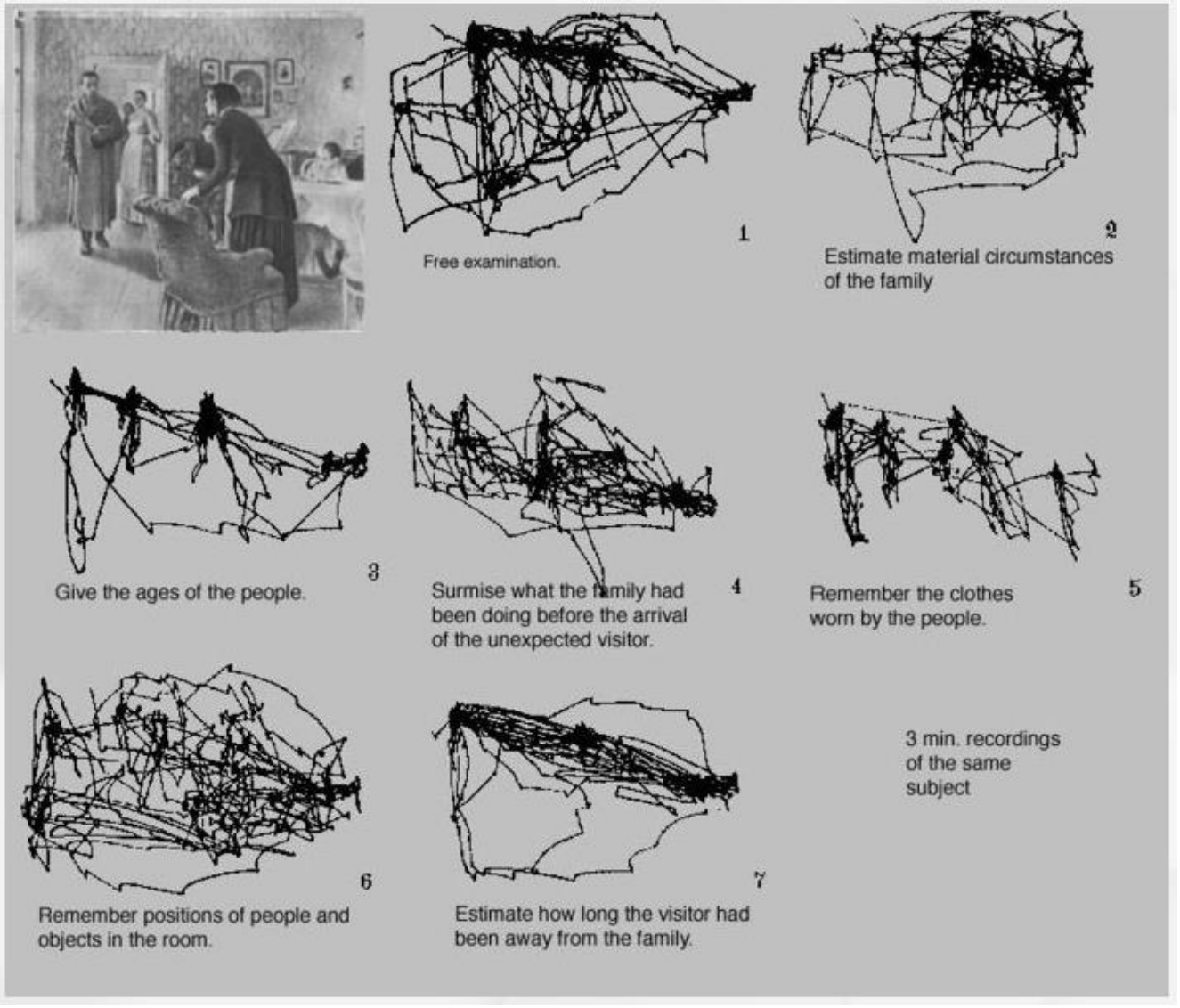
Slower (5 shifts/s or fewer; like eye movements)

Top-Down Controls

The unexpected visitor

Top-Down Controls

Top-down tasks strongly affect where people will look at



Bottom-Up Attention – Saliency

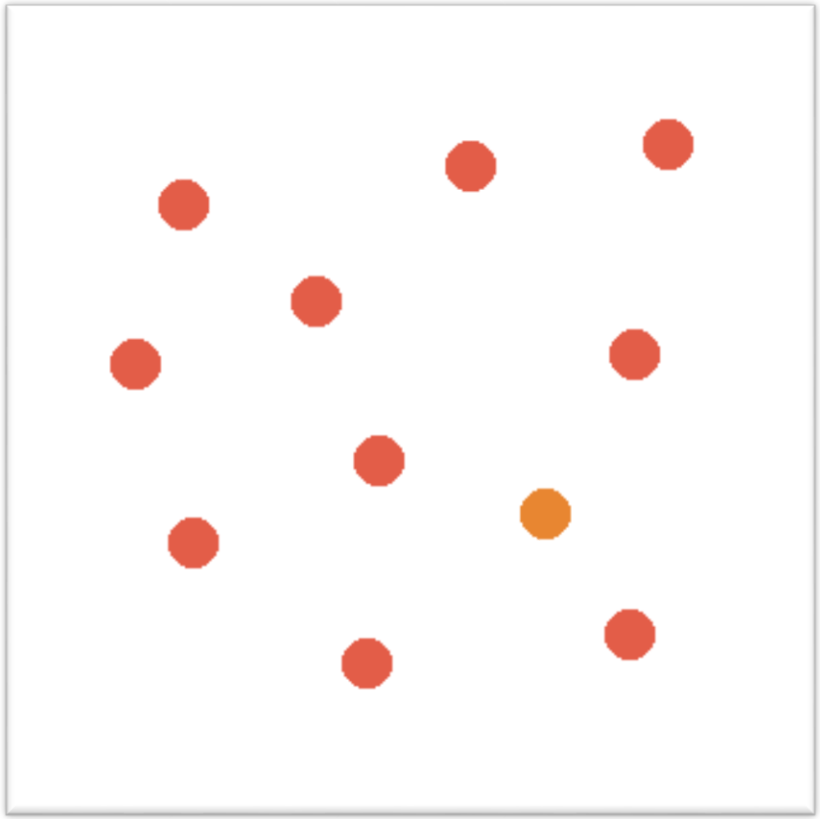
- *What's Visual Saliency*
- *Visual Saliency Modeling*
- *A Visual Saliency Model*
- *Visual Saliency Applications*

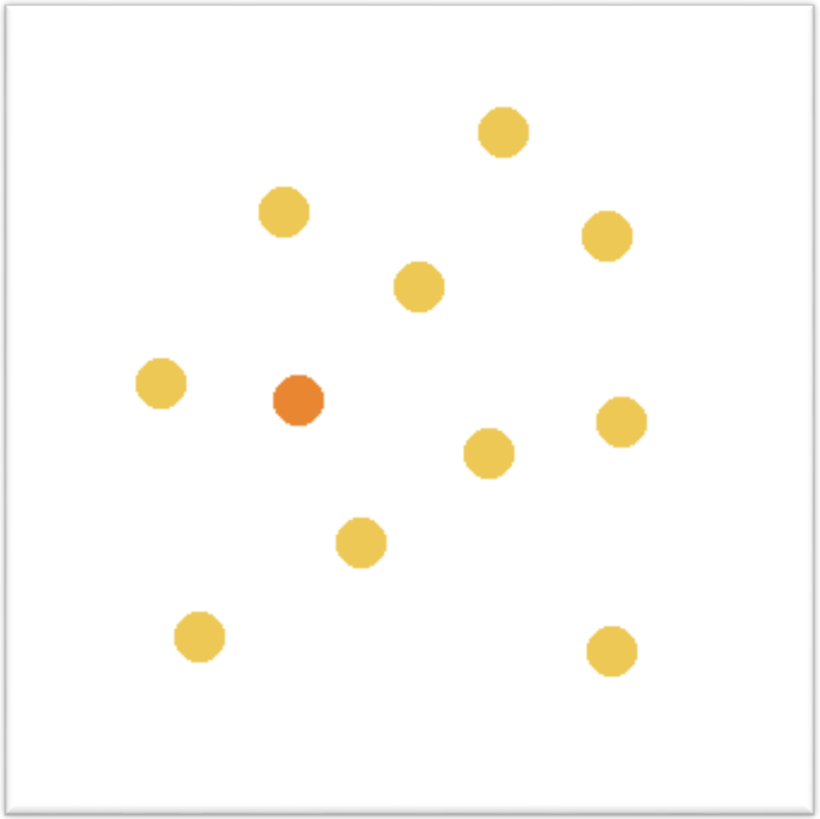
What's Visual Saliency

- *It is "Distinct subjective perceptual quality which makes some items in the world stand out from their neighbors and immediately grab our attention"*
- *It characterizes the bottom-up attentive process*
- *It could be modulated or even overridden by top-down factors*

What's Visual Saliency

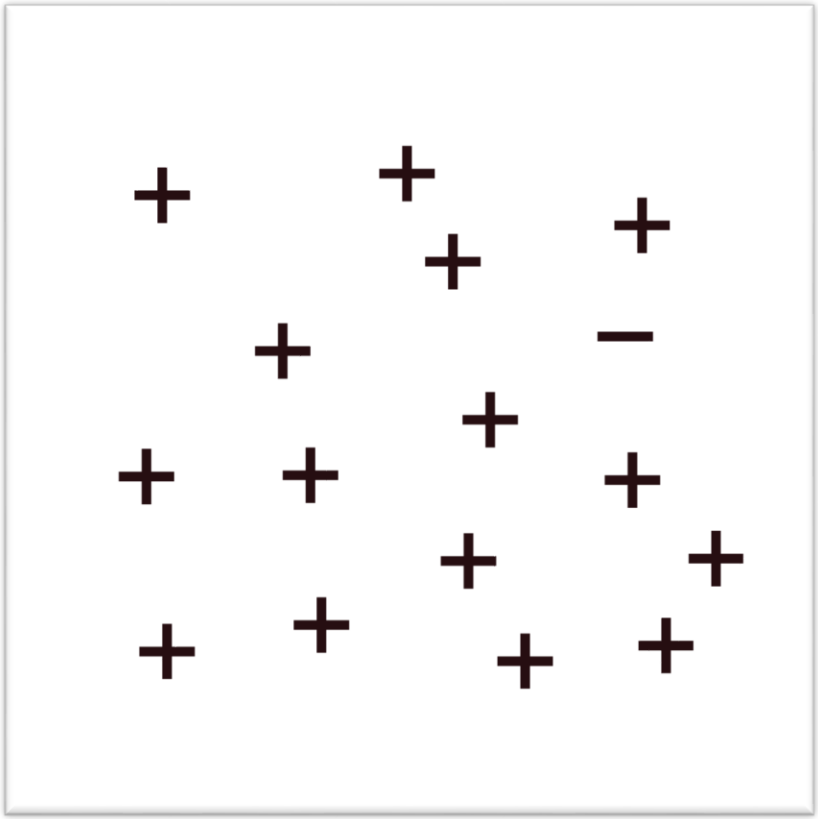
*Visual saliency is characterized by unusual/uncommon low-level visual features which could be **colors**:*

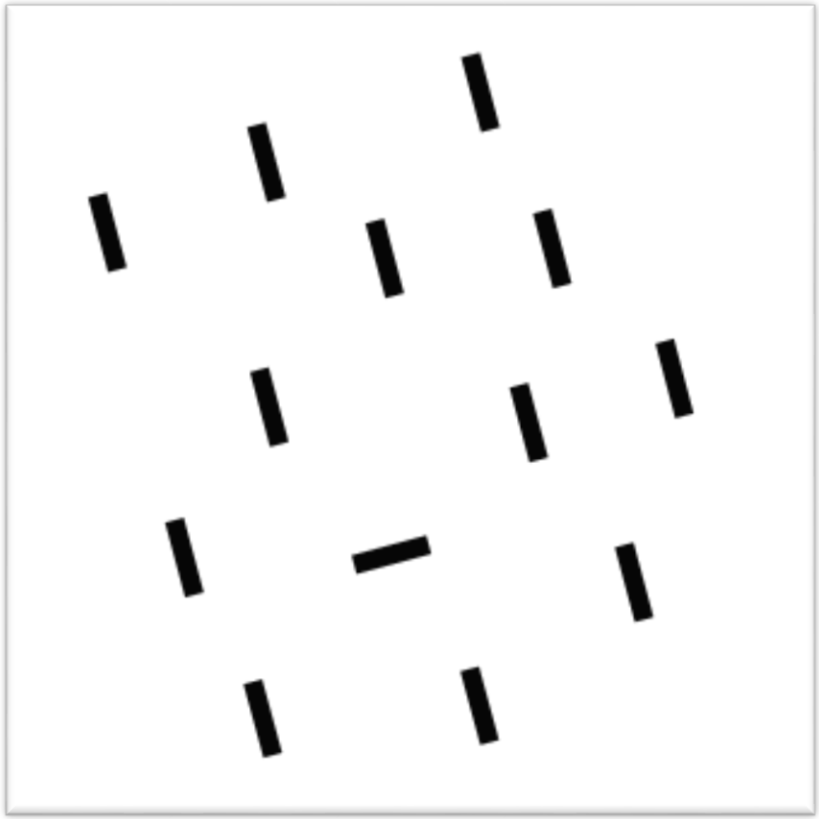




What's Visual Saliency

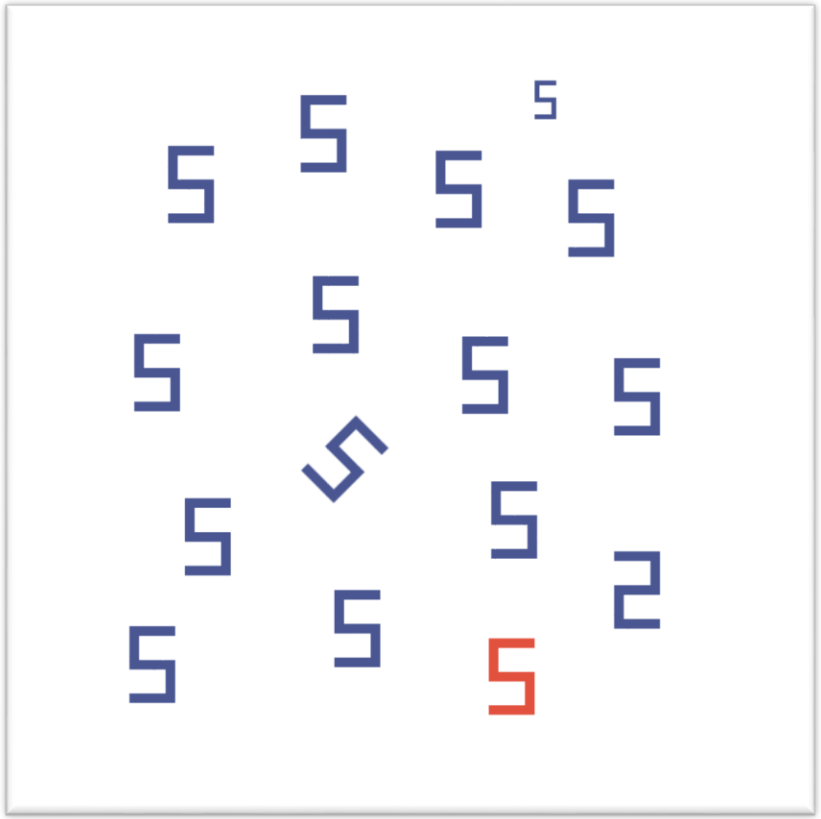
*Object **shapes** and **orientations**:*

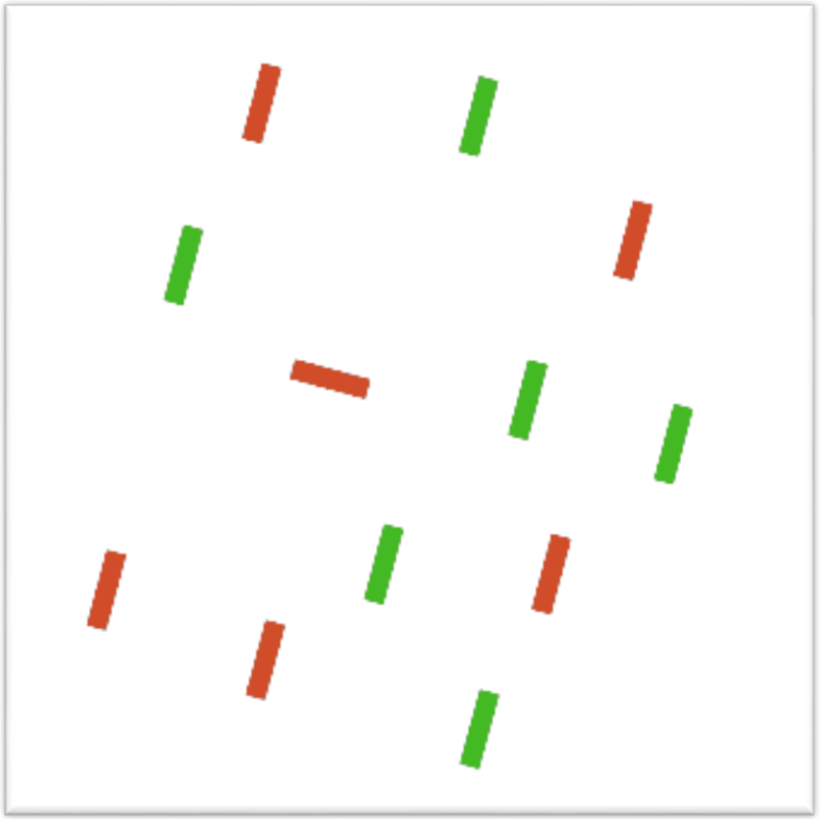




What's Visual Saliency

Or a mixture of them:



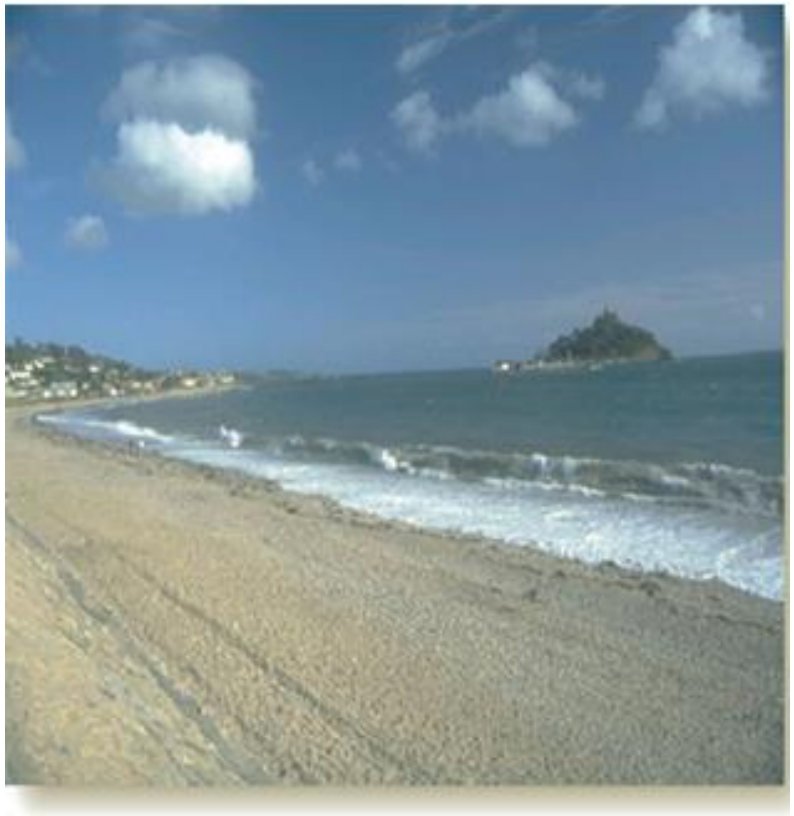


What's Visual Saliency

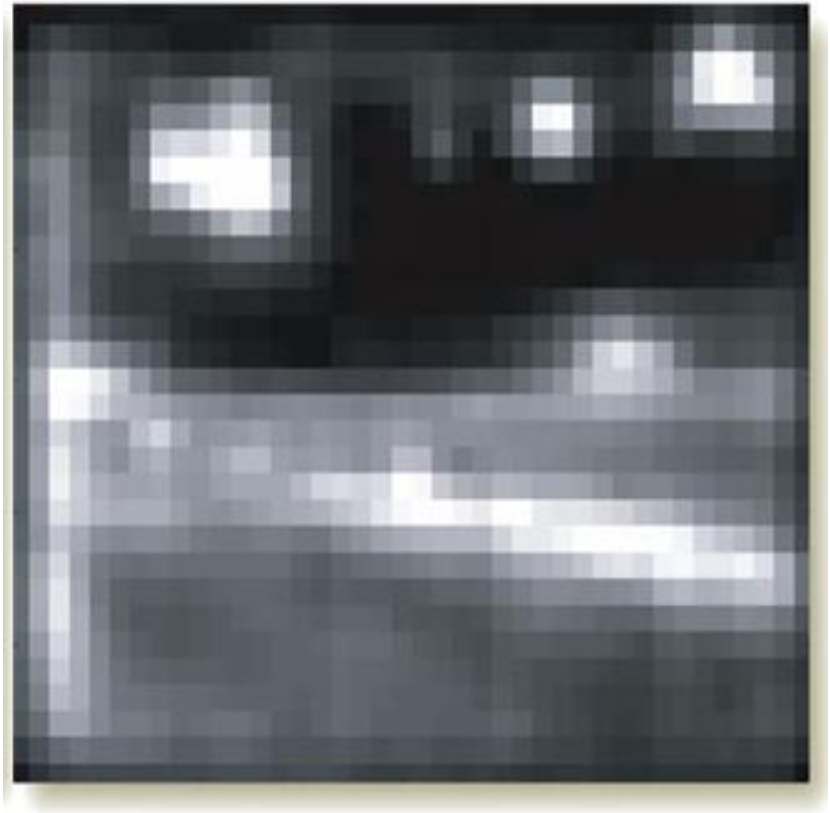
Visual saliency is often formed by a mixture of different saliency features in complex scenes.

Visual Saliency Modelling

Visual saliency can often be characterized by a saliency map.



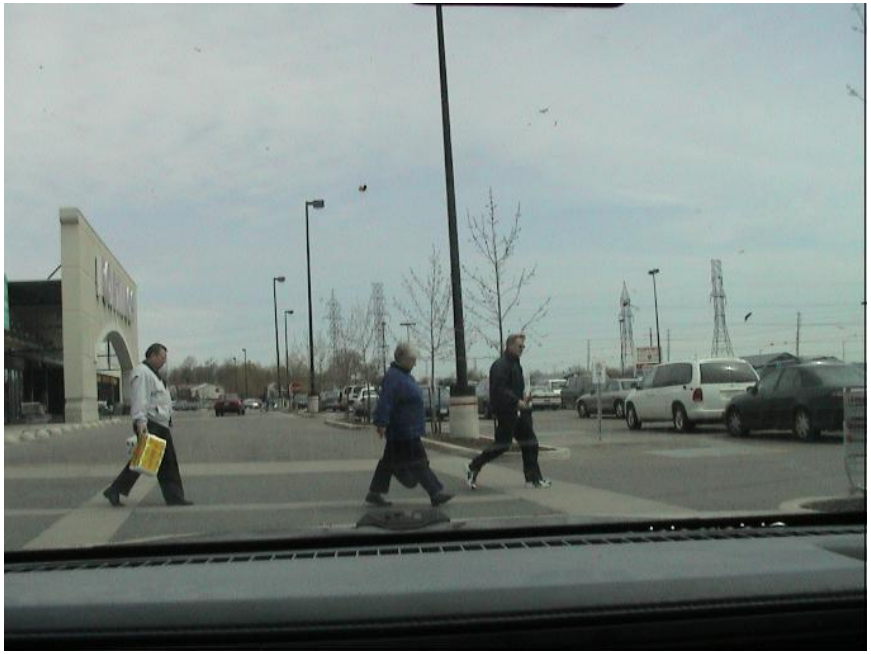
(a) Visual scene

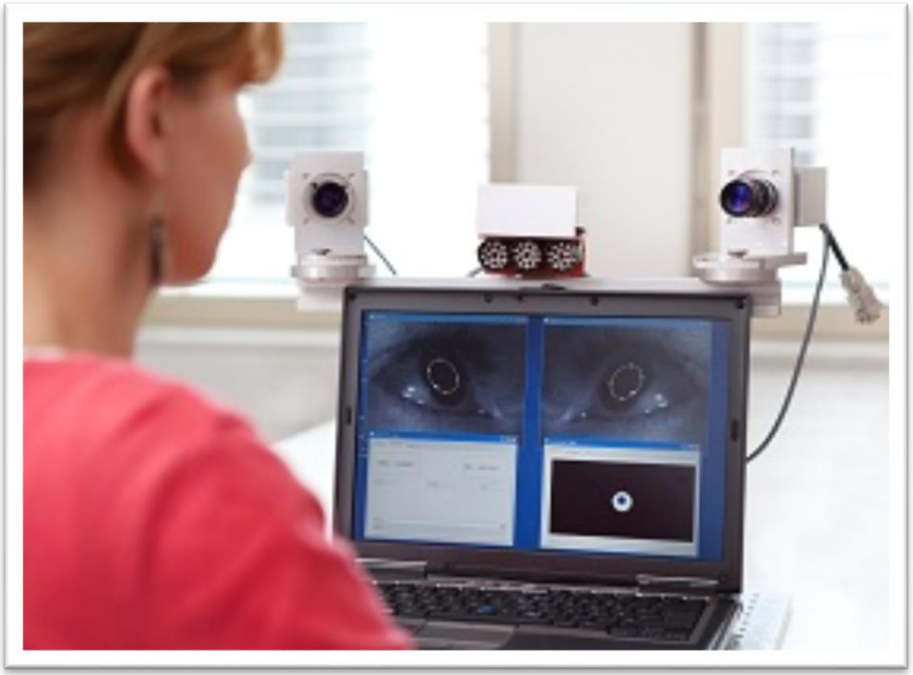


(b) Saliency map

Visual Saliency Modelling

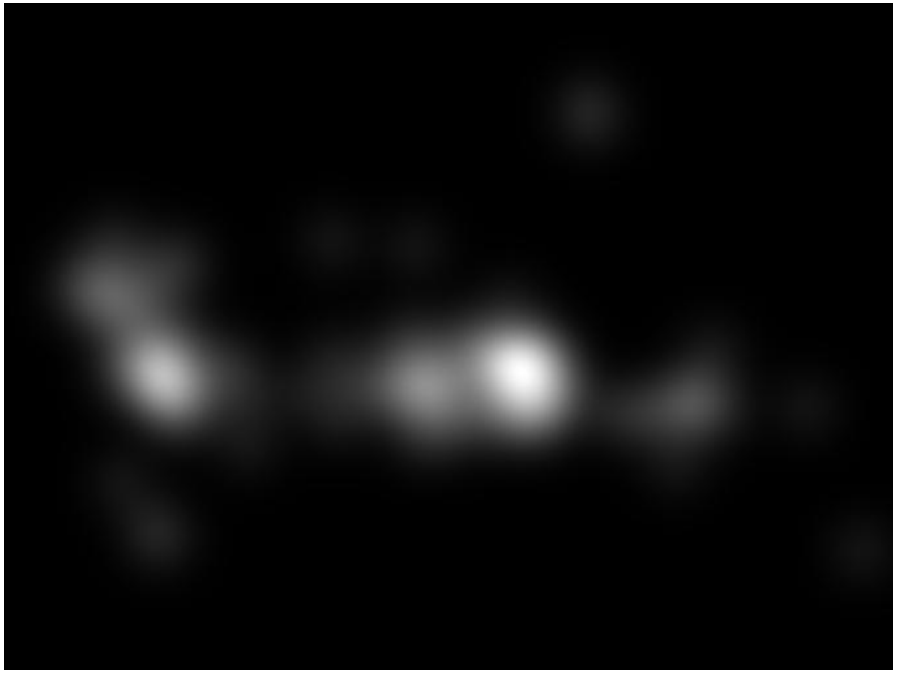
*How to get saliency maps? It can be derived by **eye trackers**.*





Visual Saliency Modelling

Eye trackers collect fixation points from a group of subjects, and saliency maps can be computed from the collected fixation points.



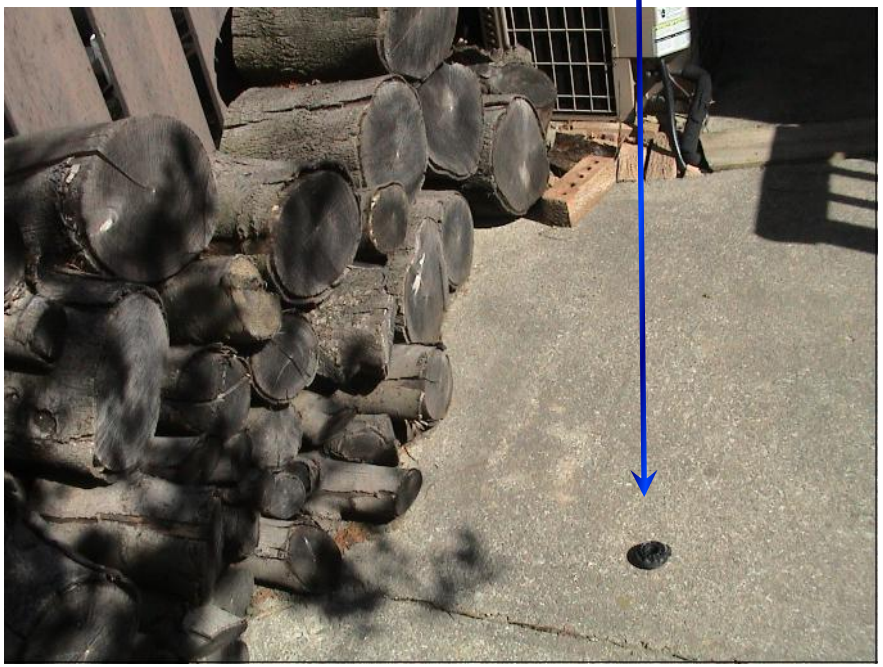
Visual Saliency Modelling

Saliency map can be derived by computational models without collecting fixation points. It can be computed by different approaches:

- *Spatial contrast based models that make use of a center-surround mechanism within spatial domain*
- *Frequency analysis based models that make use of frequency and amplitude within frequency domain*
- *Learning based models that learn image statistics or directly the eye fixations*

A Visual Saliency Model

Visual saliency characterizes "unusualness" of an object or an image region, which is often perceived by either global "uncommonness" or local "discontinuity".



**Robust and Efficient Saliency Modeling from Image Co-occurrence Histograms , S. Lu, C. Tan, J.H. Lim, TPAMI, 2014.*

A Visual Saliency Model

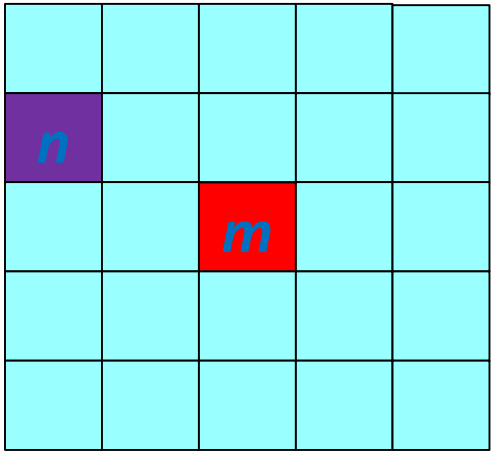
The global "uncommonness" corresponds to low occurrence frequency whereas the local "discontinuity" corresponds to image contrast across the object boundary which can be captured by co-occurrence of neighbouring pixels. We use Image co-occurrence histogram (ICH) , which captures both occurrence of image pixels and co-occurrence of neighbouring pixel pairs, for saliency computation.

A Visual Saliency Model

*ICH can be built efficiently. Consider an integer image **I**. Let **K = {1, 2,…,k}** be a set of **k** possible image values within **I** (**k** is 256 for a 8-bit integer image). The ICH of **I** is defined as follows:*

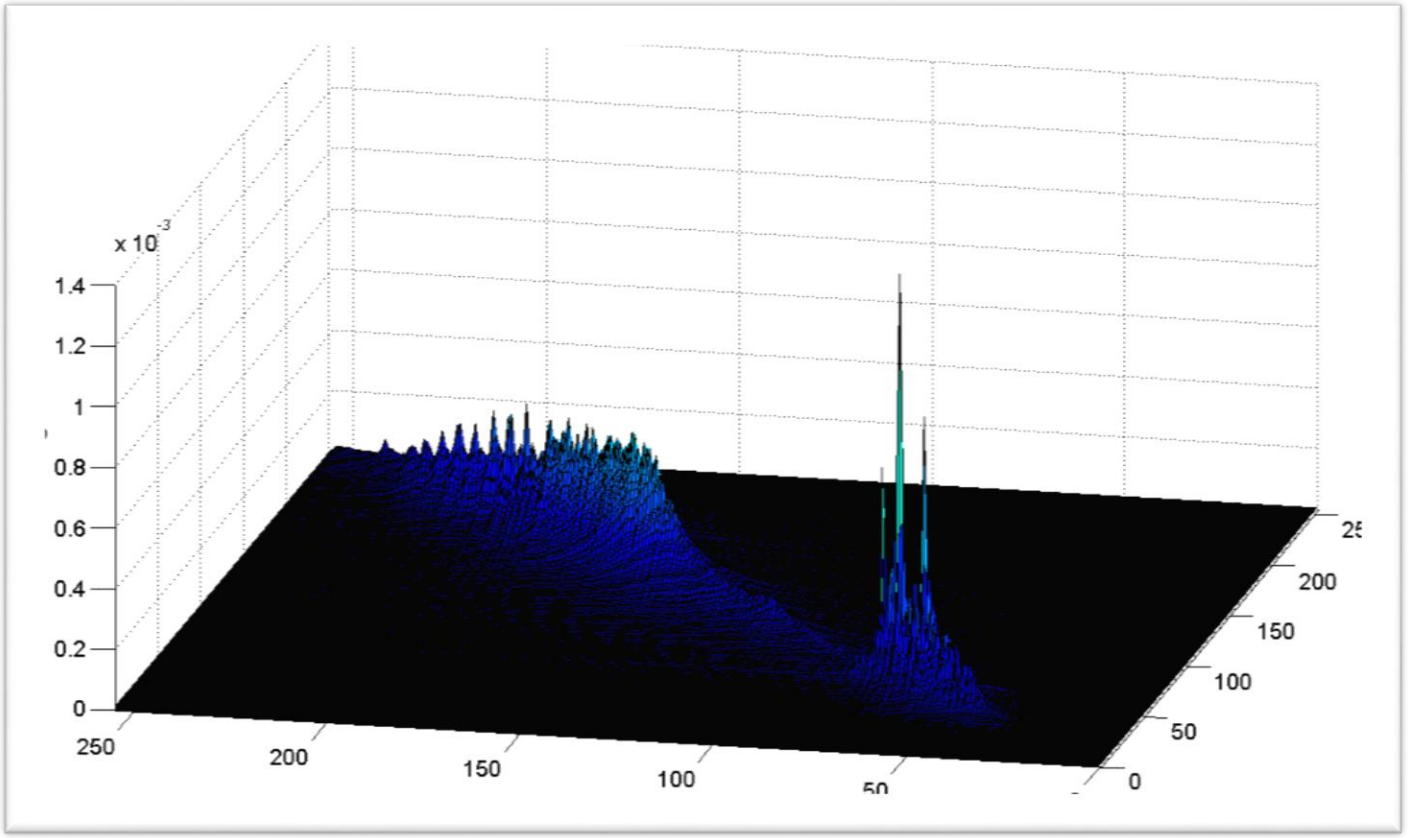
$$H = \left[h(m,n)\right],\ m,n \in K$$

*where **H** is a square symmetric matrix of size **k ✕k**. An ICH element **h(m, n)** is the co-occurrence frequency of image value **n** within the neighborhood of image value **m**.*



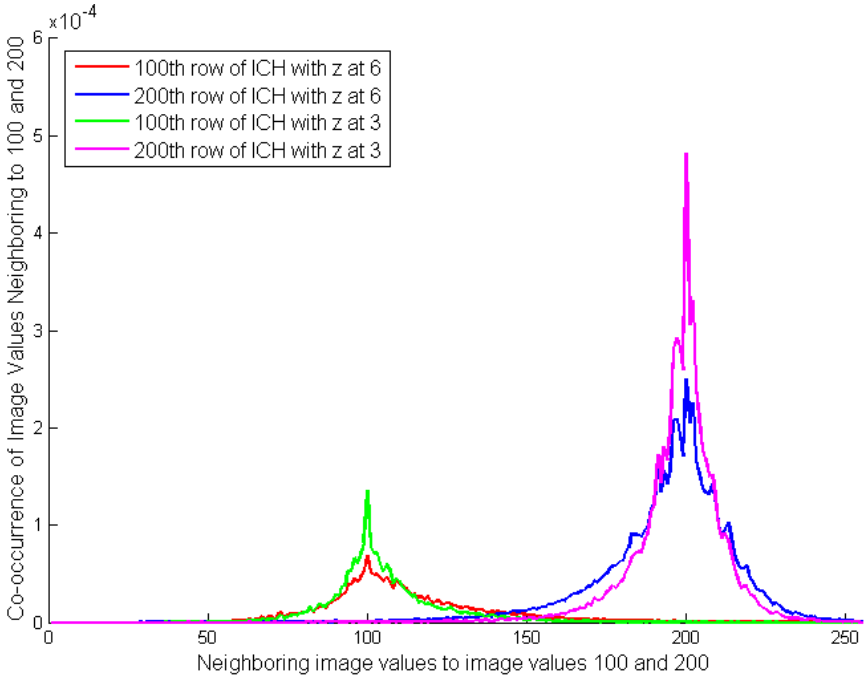
A Visual Saliency Model

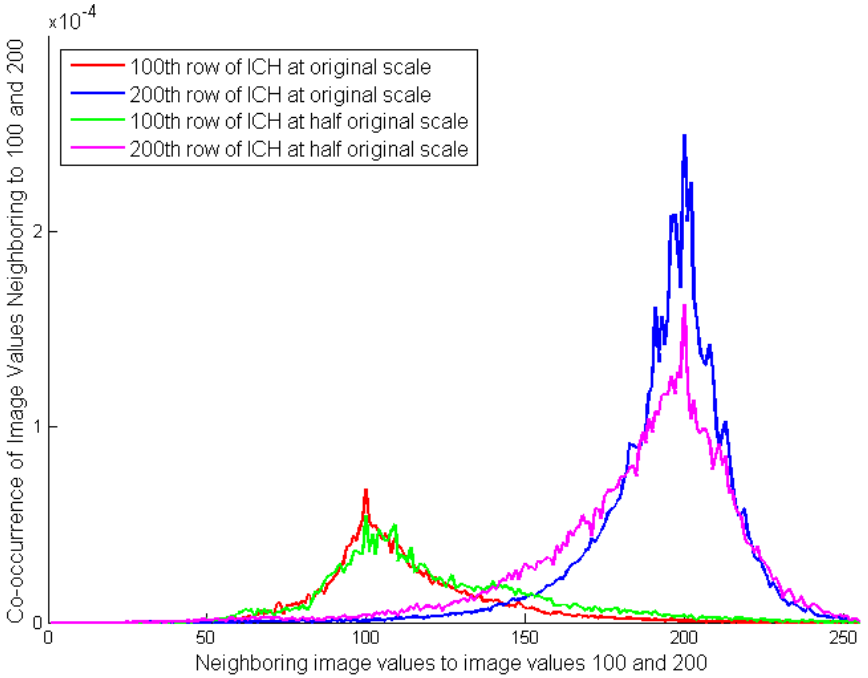
The graph below shows a sample ICH.



A Visual Saliency Model

The two graphs show good properties of ICH.





A Visual Saliency Model

*With the **H**, a probability mass function (PMF), **P,** is first computed:*

$$P = \frac{H}{\sum_{m=1}^{k}\sum_{n=1}^{k} h(m,n)}$$

As saliency is usually negatively correlated with occurrence/co-occurrence, an inverted PMF, , is computed as follows:

$$\overline{p}(m,n) = \begin{cases} 0 & if \;\; p(m,n) = 0 \\ 0 & if \;\; p(m,n) > U \\ U - p(m,n) & if \;\; p(m,n) \le U \end{cases}$$

*Where **U** is a uniform distribution whose value is defined by the average of **P**.*

A Visual Saliency Model

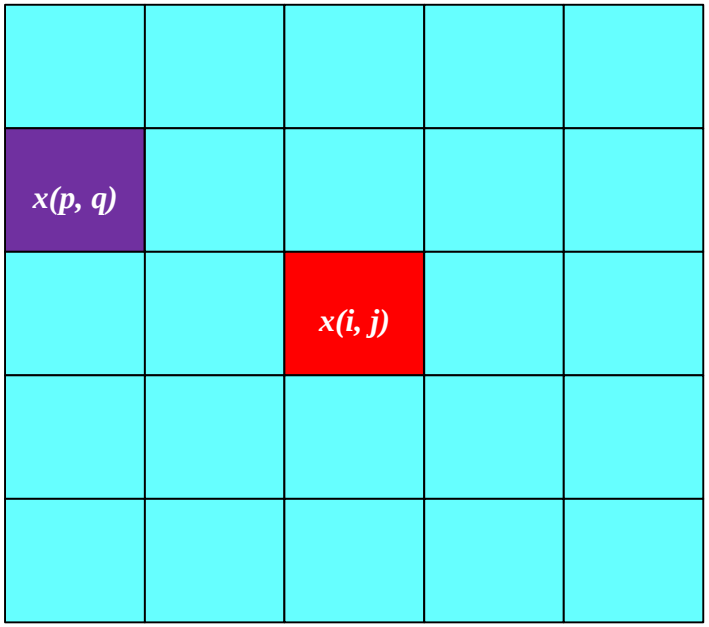
Saliency can be computed from the inverted PMF as follows:

$$S(i,j) = \sum_{p=i-z}^{i+z} \sum_{q=j-z}^{j+z} \overline{p}(x(i,j), x(p,q))$$

where z denotes the size of the neighborhood which is the same as used for the ICH construction.

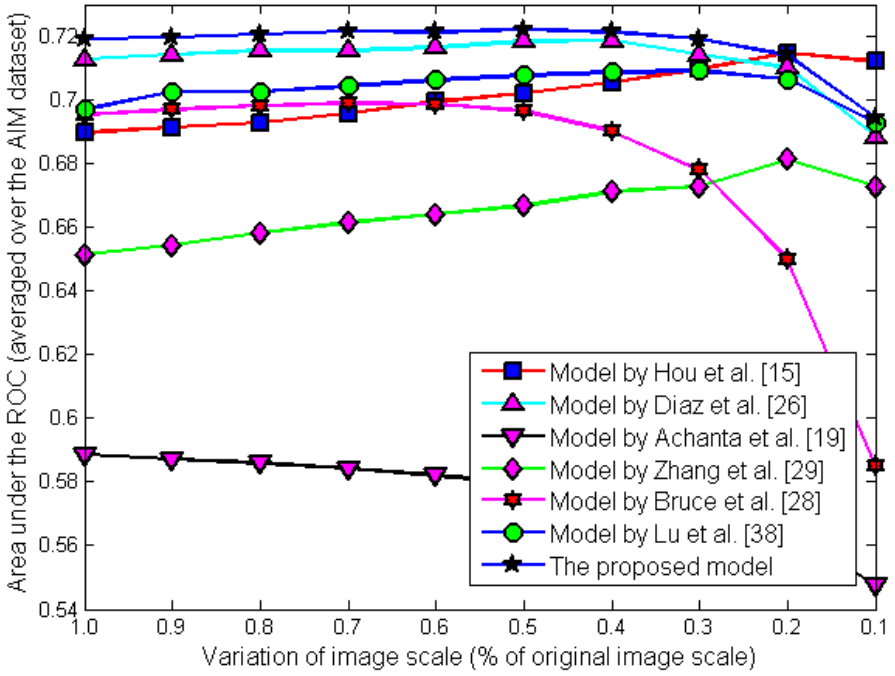
The saliency map can be computed by:

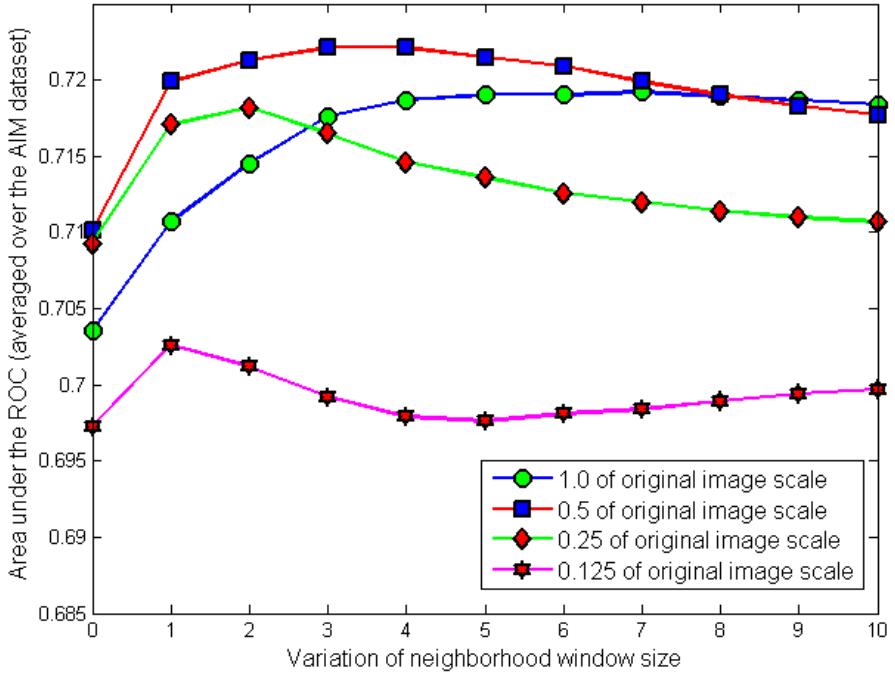
$$\mathbf{S} = G\left(\sum_{c=1}^{3} S_c \right)$$



A Visual Saliency Model

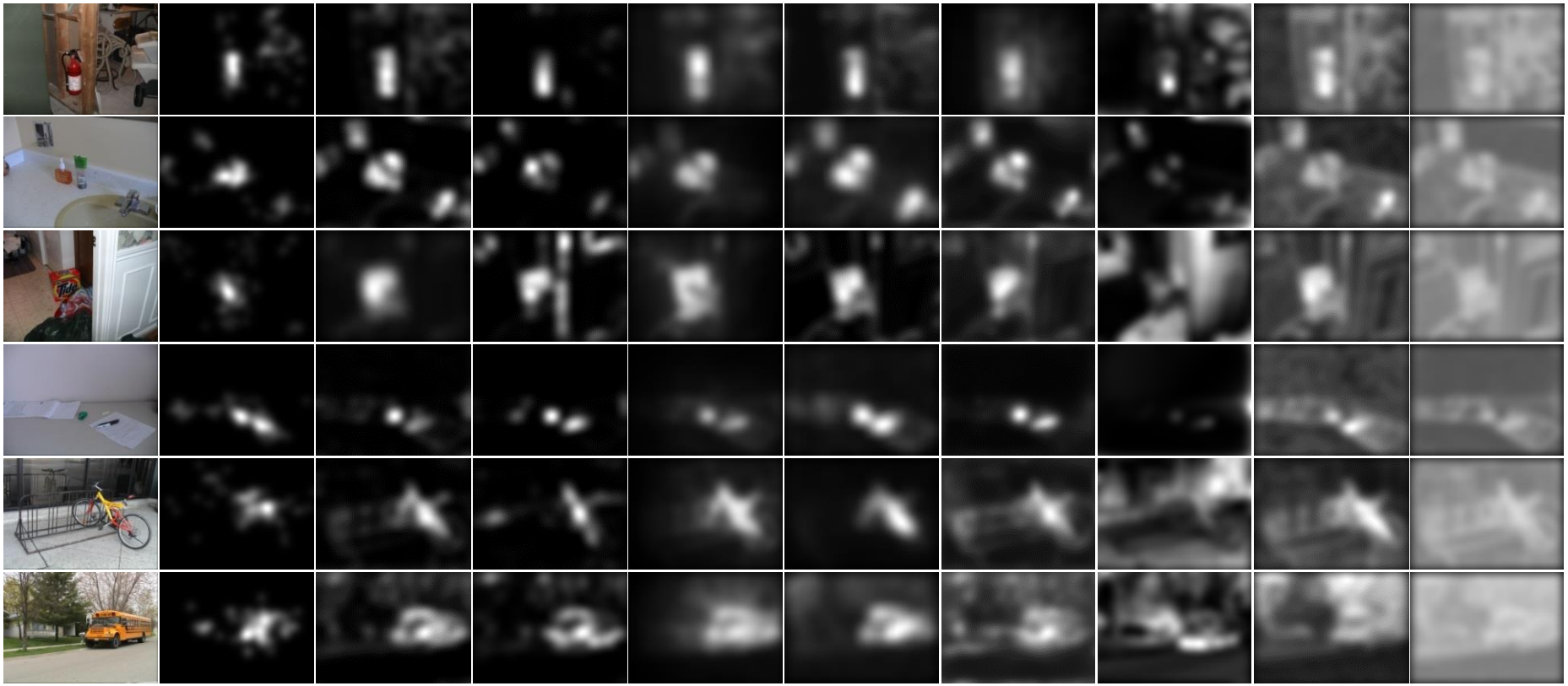
The first graph shows AUC of the proposed model and six state-of-the-art models when the image scale changes. The second shows AUC of the proposed model when the neighborhood size z changes.





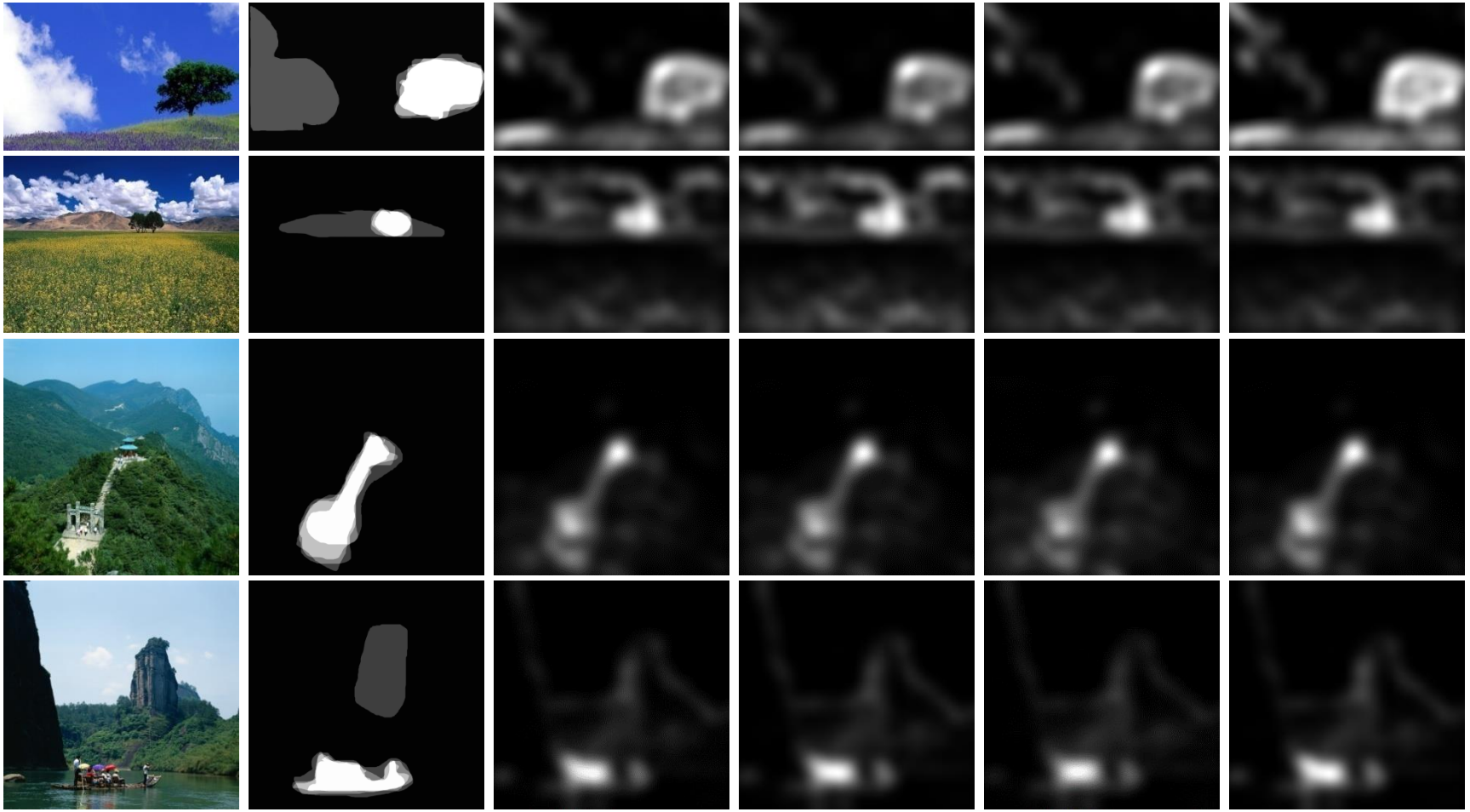
A Visual Saliency Model

The figure below illustrates experimental results based on sample images from the AIM dataset.



A Visual Saliency Model

*The ICH based model is tolerant to the only model parameter, the neighborhood size **z**. The graph shows saliency maps of images from the SR dataset when **z** is set at 2, 4, 6, and 8 pixels, respectively.*

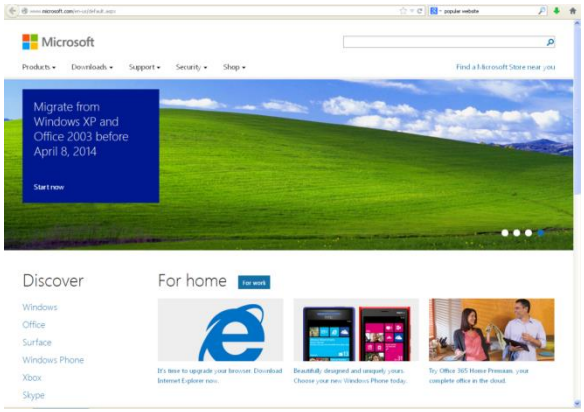


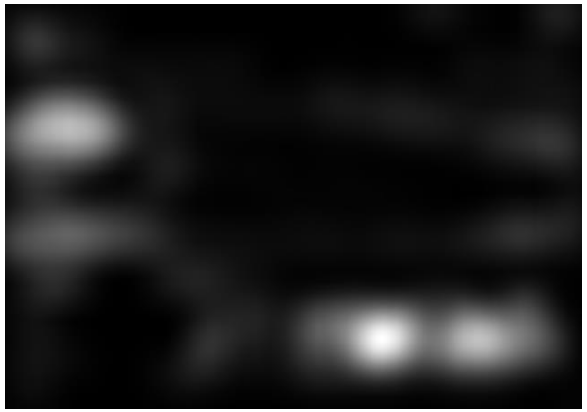
Visual Saliency Applications

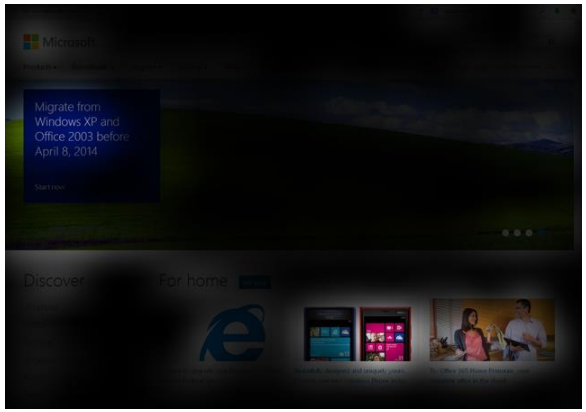
The proposed model can be applied for the extraction of salient objects through thresholding of the generated saliency maps.

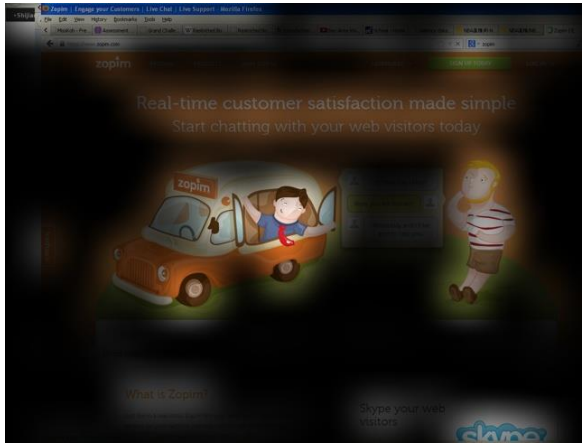
Visual Saliency Applications

The proposed saliency model can provide a great reference for evaluation of design of web pages, advertisement, posters, etc.









Visual Saliency Applications

The proposed saliency model can be applied for adaptive picture-in-picture placement.

Top-Down Attention – Visual Search

- *Human Visual Search Basics*
- *Human Visual Search Methanisms*
- *Visual Search Models*

Human Visual Search Basics

- ***Visual search*** *is the common task of looking for something in a cluttered visual environment. The item under search is termed the* ***target****, while non-target items are termed* ***distractors****.*
- *There are two typical types of visual search: Feature Search vs. Conjunction Search*

✓ *Feature search (Parallel)*

- *Differentiated by a single feature; pop-out; saliency.*

✓ *Conjunction search*

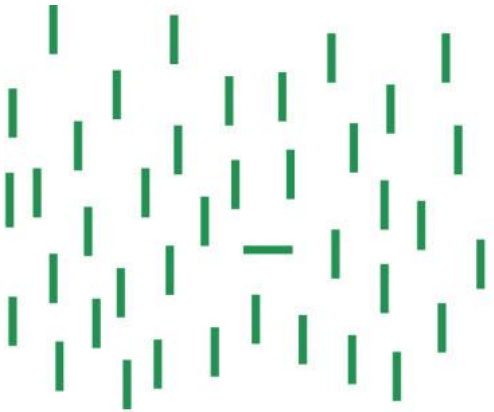
- *features need to be combined to find the target.*
- *need attention.*
- *Because you can attend only one item at a time, the conjunction search becomes more difficult when more items are in the stimulus frame.*

Human Visual Search Mechanisms

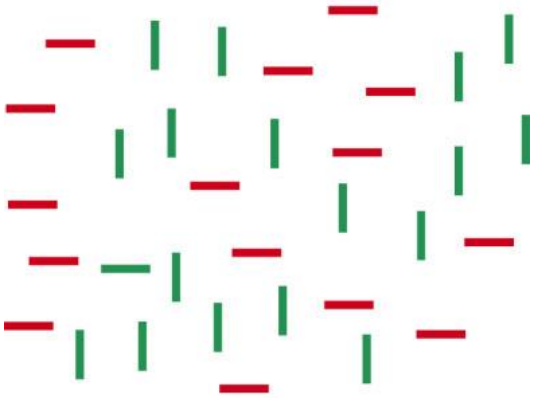
Feature search and conjunction search can be illustrated by a visual search task: Looking for the target —

- *Feature search*
  - *This is easy because you find the target by looking for a single feature.*
  - *→ you don't need attention*

- *Conjunction search*
  - *For this you need to **combine** two or more features (color and orientation)*
  - *→ you need attention*



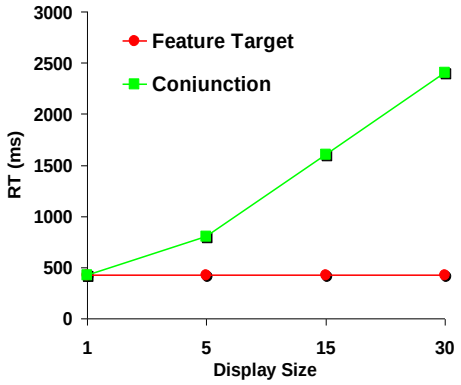
(a)



(b)

Human Visual Search Mechanisms

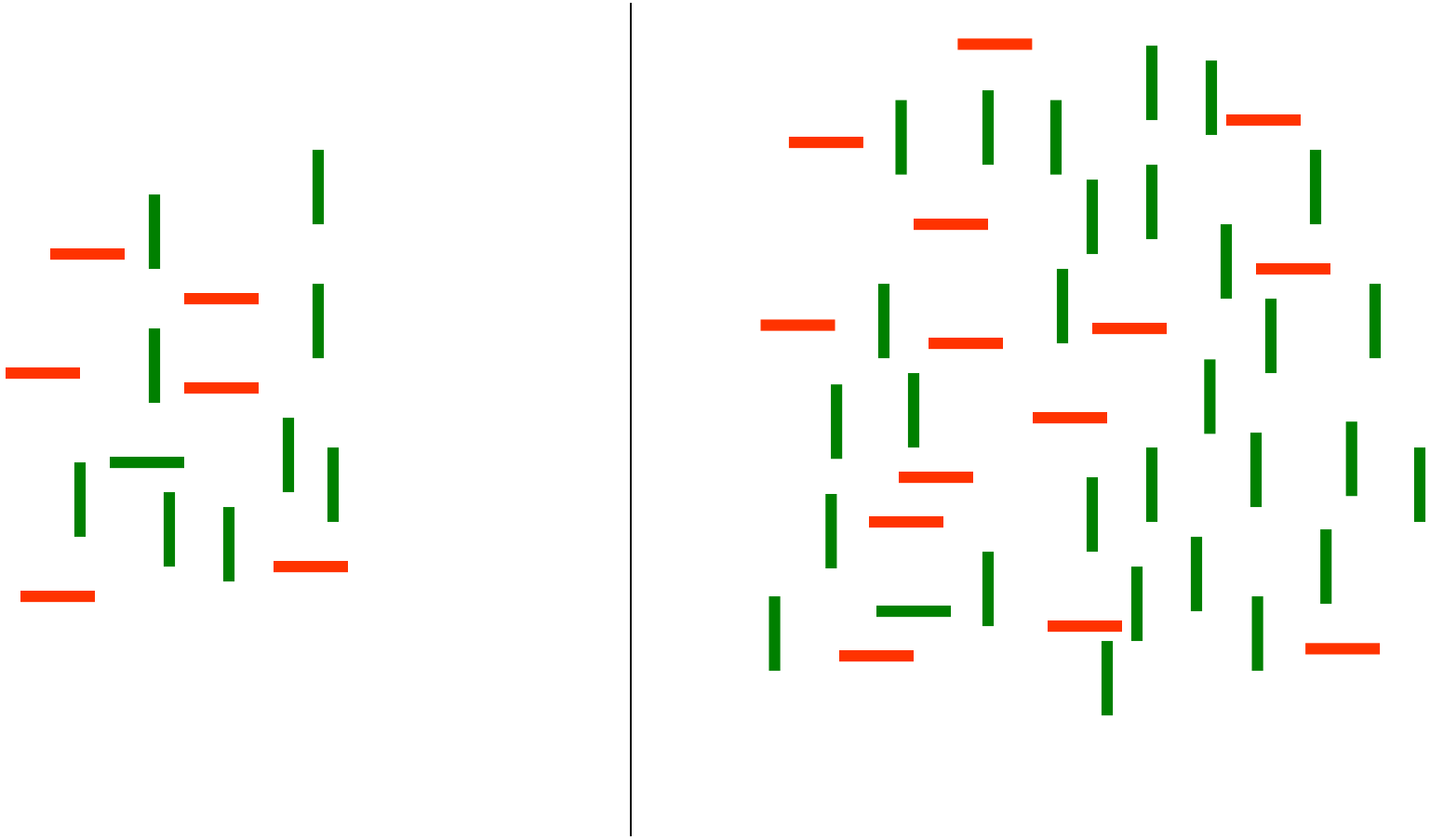
Human subject studies show that feature search is less affected by the display size as compared with conjunction search.



- *Feature targets pop out*
  - *flat display size function*
- *Conjunction targets demand serial search*
  - *non-zero slope*

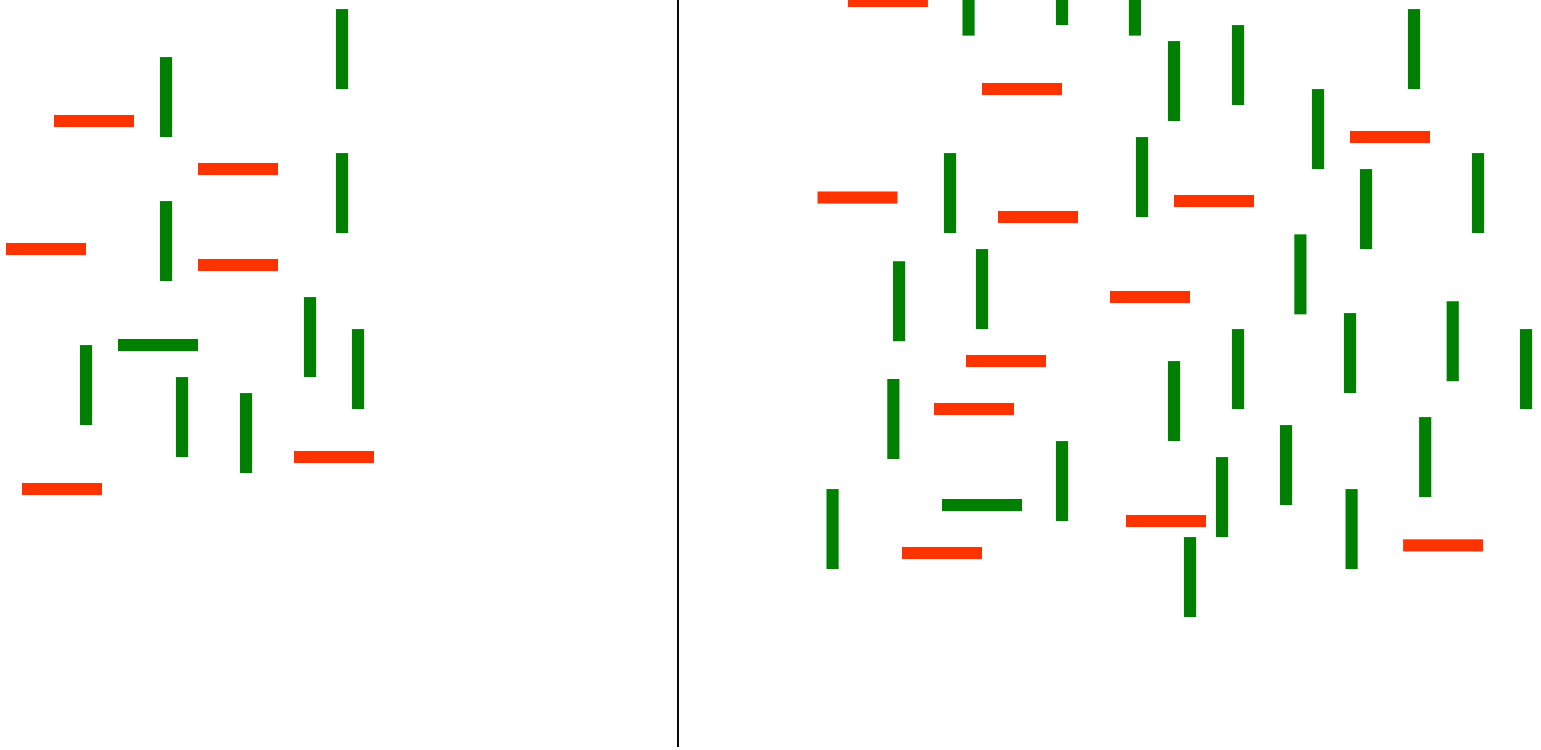
Human Visual Search Mechanisms

Find — *Which is more difficult?*



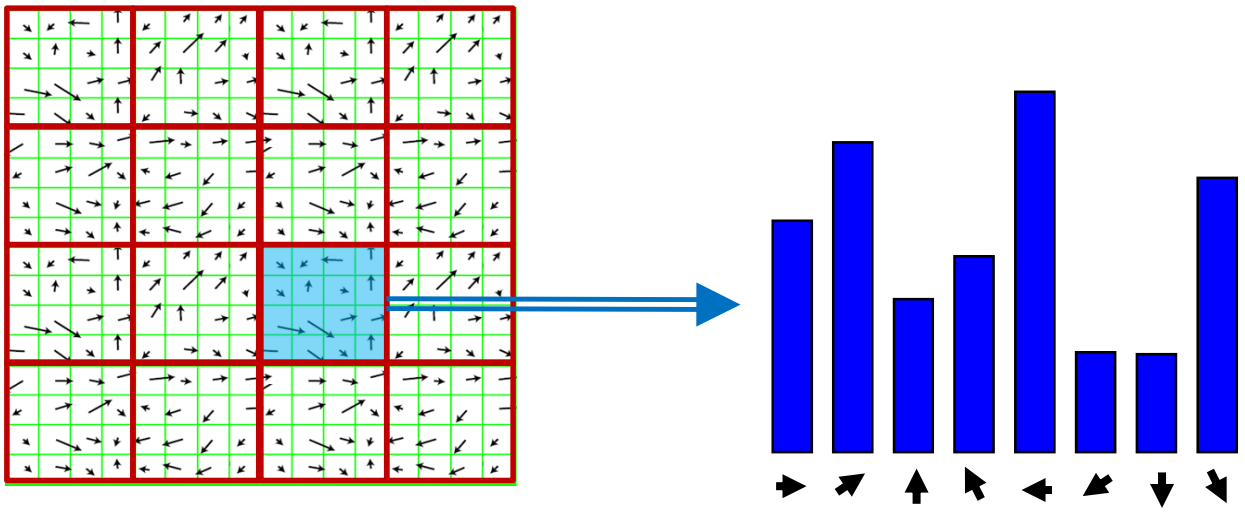
Visual Search Models

Find — *Which is more difficult?*

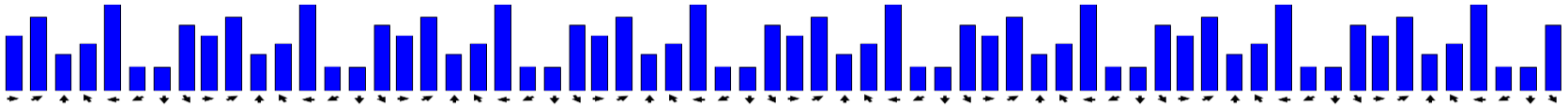


Visual Search Models

Visual search modelling is quite similar to object/target detection.



The result: 128 dimensions feature vector.



Summary

- *What are bottom-up and top-down attention*
- *What is visual saliency and how to model it*
- *How does human visual search work and how to model it*