Focus of Attention

CS 510

Lecture #15

April 8th, 2013

Colorado State University

What is attention?

- "The selective aspect of processing" – Kosslyn[1]
- "processes that enable an observer to recruit resources for processing selected aspects of the retinal image more fully than nonselected aspects" – Palmer[2]

Overt vs. Covert Attention

- Overt attention: observable movements of eyes, head & body to orient eyes
  - Foveas: 90% of receptors, ±2°
  - Allocation to 3D point in space
    - Vergence & focus
  - Average dwell time: ~300ms[4]
  - Saccadic movement
    - Very fast: ~30ms, up to 900°/sec
    - Suppression: no input during saccade
  - World appears as sequence of displaced, small, high resolution, stereo images with low resolution peripheries

Overt vs. Covert Attention (II)

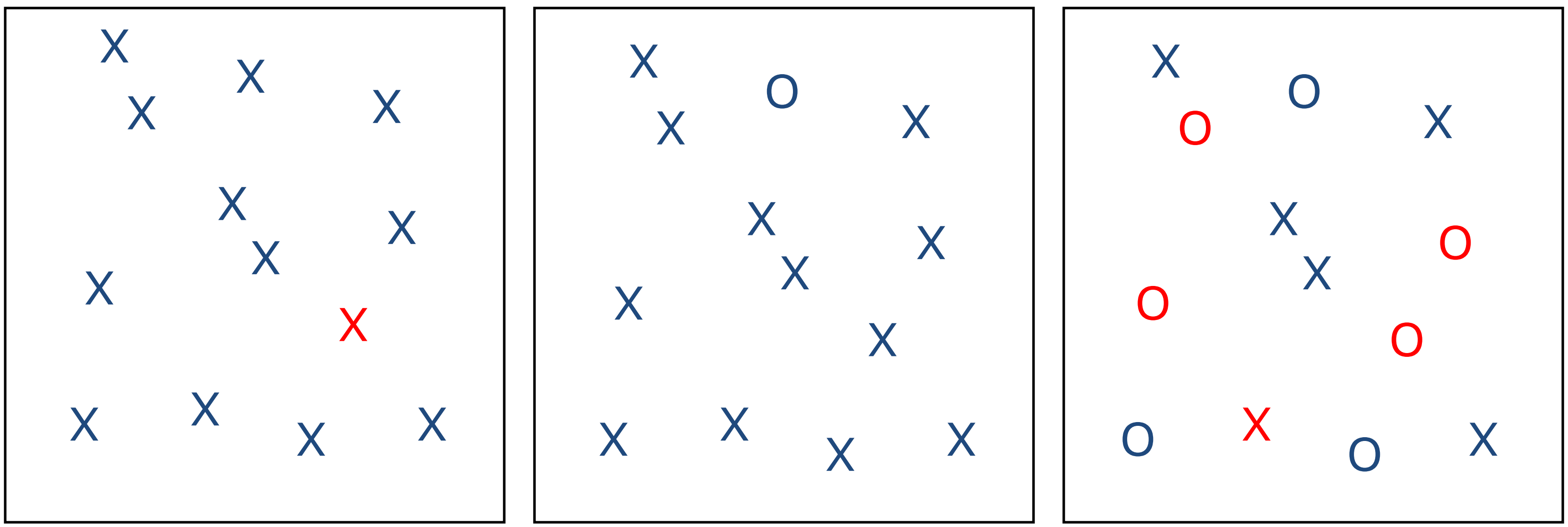
- You don’t process all the data in your foveal image
- Covert attention: selection of retinal data to process (“inner eye”)
  - Cannot be observed directly
  - Its existence is not in dispute
  - Its form is a matter of intense debate
  - Assumption: insufficient resources necessitate covert attention.
- Covert attention is the subject of this lecture

3 Models of Covert Attention

1. Feature Integration Theory
   - “Pre-attentive” low-level features computed in parallel across the image
     - E.g. color, edge orientations, motion
   - In visual search, attention can jump to locations based on pre-attentive features (“pop-out”)
   - Conjunctions of features or complex features require sequential search
   - Implicitly assumes attention is like a spotlight

Feature Integration Theory (II)

Find the red target | Find the round target | Find the red 'X' target



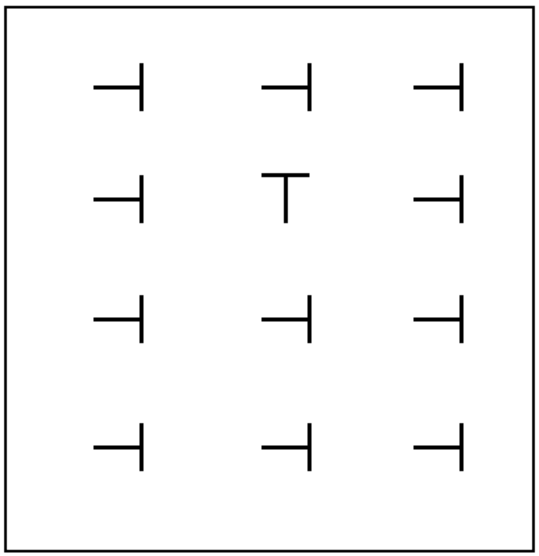
Colorado State University

3 Models (II)

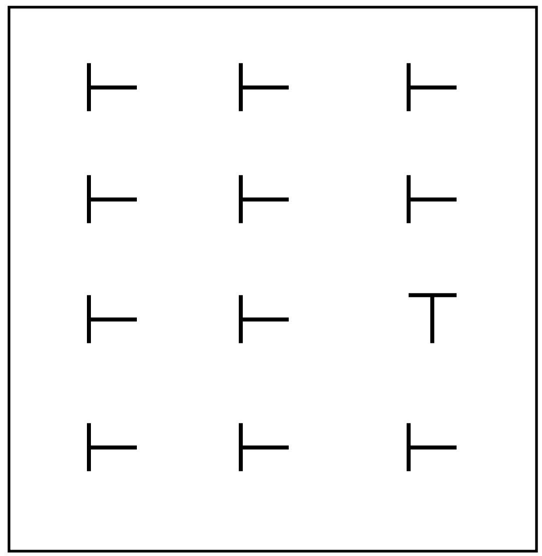
2. Integrated Competition Hypothesis
   - “Pop-out” effect depends on:
     - Homogeneity of distractors
     - Homogeneity of targets (seq. pres.)
   - Primary role of attention is *segmentation* (or grouping)
   - Low-level features important as the basis of segmentation

Integrated Competition Hypothesis (II)

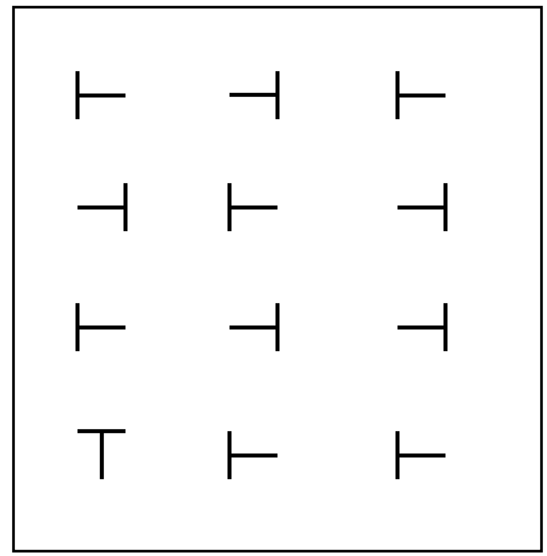
Find the upright T



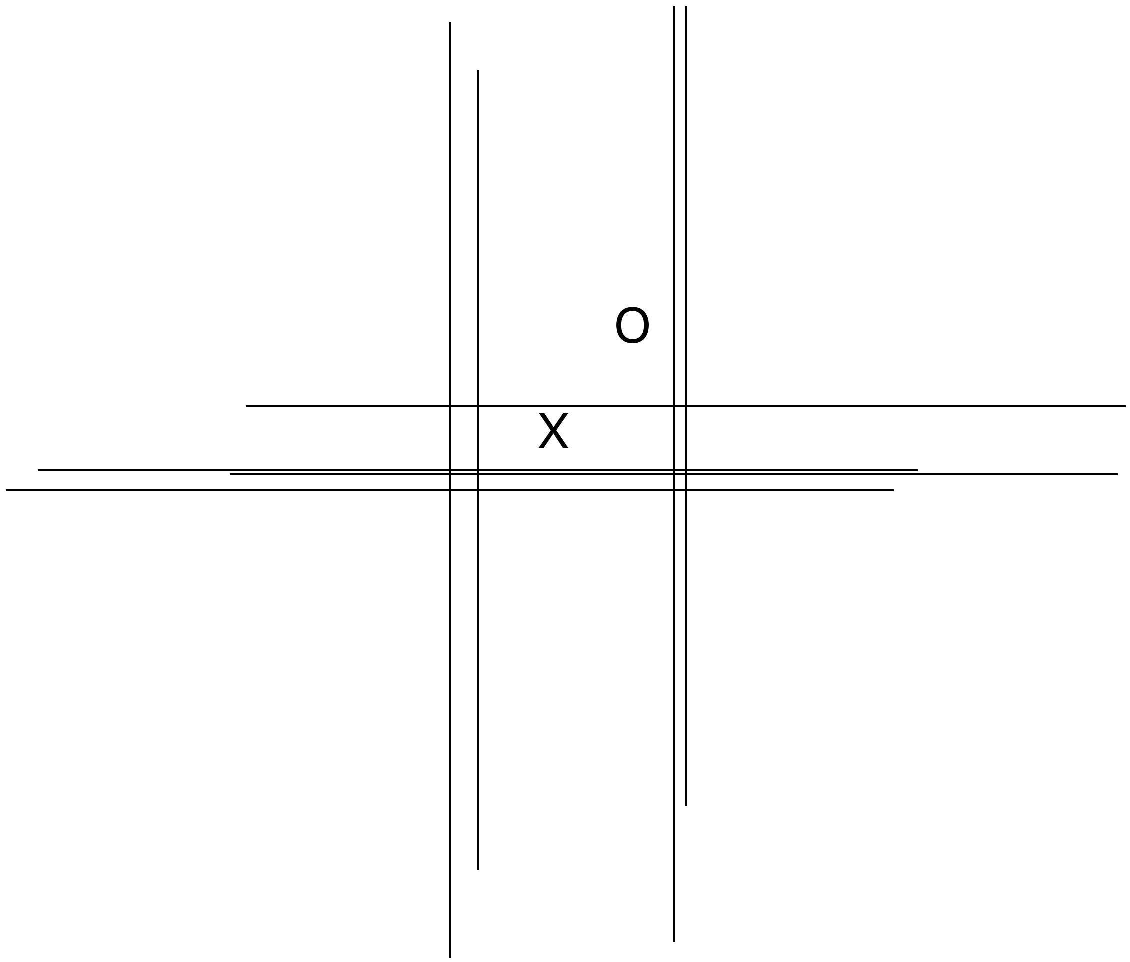
Find the upright T



Find the upright T



Task: Which Line is Longer?



Colorado State University

3 Models (III)

- Inattentional Blindness Theory
  - When concentrating on the task, most subjects will not see additional objects
    - Depends on semantics of additional object
  - Additional objects are interpreted
    - Cause priming effects
  - Hypothesis: all objects in visual field are interpreted
    - Attention is a late effect, caused by attentional bottleneck

How does this effect computer vision?

- Scale space theory
  - Image pyramids
- Difference of Gaussians (DoGs)
  - Impulse detection
  - Determines location & scale
- Refinements
  - Corners
  - Entropy

Resolution

- Definition:

  The *resolution* of an image is the inverse of the spatial area covered by each pixel. This depends on:

  1. The image size of the camera
  2. The field of view of the lens
  3. The distance to the target

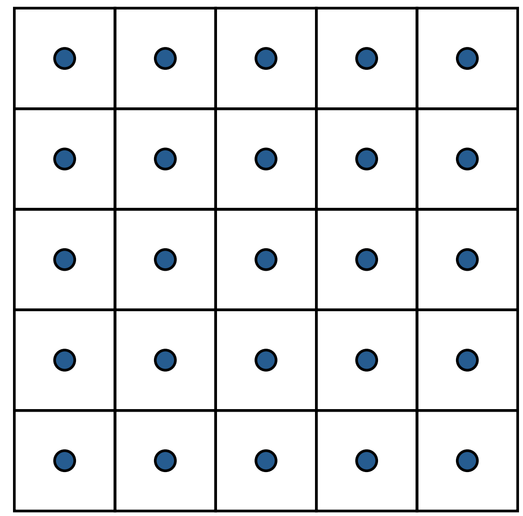
  Note that doubling the distance to the target halves the image resolution.

Scale space

- The appearance of an object is a function of the image resolution:
  - A checkerboard becomes a uniform gray surface as the resolution decreases.
  - A thin black bar goes from being a bar (with parallel lines) to a single line to nothing.
- The goal of scale space theory is to simulate what happens to the appearance of an object as resolution decreases.

Base Case: Raw Image

- We model pixels in raw images a point-wise intensity estimates, covering no area.
  - Not quite right: pixels sample over a small area
  - Areas don't overlap
    - Except for blooming
  - Gaps between areas
- Highly sensitive to micro-tranlsations
- Same model our ray tracers used…



Scale Space

- We want pixels to average information over an area, to improve stability and reduce aliasing
- We being by convolving the image with a Gaussian with $\sigma = 1$
  - Now most of the information comes from an area one pixel in size
- $\sigma = 1$ is the “base” resolution

Varying Resolution

- To cut the resolution in half, we convolve the original image with G(σ=2).
- To cut the resolution by a quarter, use σ = 4, etc.
- Problem: this is very expensive
  - To cover a lot of scales, the image gets convolved a lot of times
  - The convolution masks keep getting bigger!

Colorado State University

Image pyramids

- Fortunately, the lower the resolution, the fewer pixels you need.
- $I \otimes G(\sigma=a) = (I \otimes G(\sigma=b)) \otimes G(\sigma=c)$, iff $a^2=b^2+c^2$
- Therefore, start with an image with $\sigma=1$.
  - Convolve it with $G(\sigma=\sqrt{3})$
  - This produces an image with $\sigma=2$
  - Now subsample every other pixel
    - Since you have halved the resolution
  - Repeat

Source: http://www.jaredjacobs.com/stanford/cs223b/results/girl-1.png

Image Pyramids (II)

- The result is an image pyramid
  - Every image ½ the width and height of its parent
  - Every image has σ =1
    - after subsampling
- Total cost of pyramid construction:
  - 1 convolution & downsample
  - + 1/4(convolution & downsample)
  - + 1/16(convolution &…)
- Total cost < 1.5*(convolution & downsample)
- Note: it is possible to have intermediate images within scales
  - But only downsample when σ=2

Focus of Attention

- The goal of focus of attention is to:
  1. Pick locations & scales in an image
  2. That convey information about the scene
  3. Would be identified again if the object occurs in another image
     - i.e. repeatable

Method: find impulses in $f(x, y, \sigma)$

Image source: http://micro.magnet.fsu.edu/primer/java/digitalimaging/processing/diffgaussians/diffgaussiansfigure1.jpg

Difference of Gaussians (DoG)

- A Difference-Of-Gaussians (DoG) function is an impulse filter, constructed by subtracting two Gaussians with different σ's.

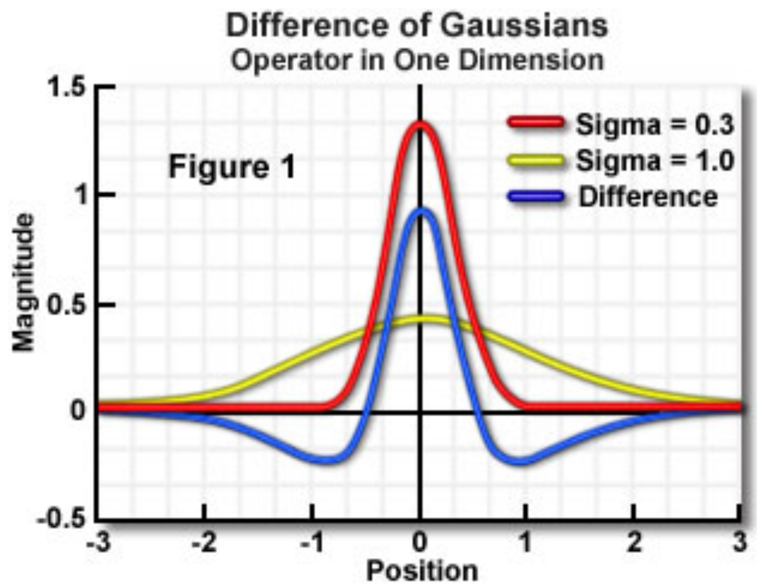
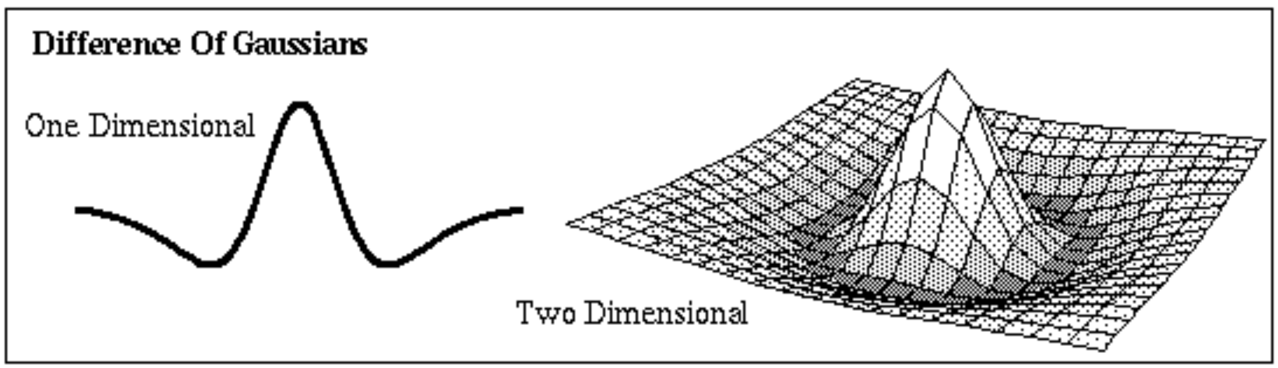


Image source: http://www.liden.cc/Visionary/Images/DIFFERENCE_OF_GAUSSIANS.GIF

DoG (II)

- DoGs are also called the "Mexican Hat" filter when 2D
- Strong positive response: on-center, off-surround
- Strong negative response: off-center, on-surround

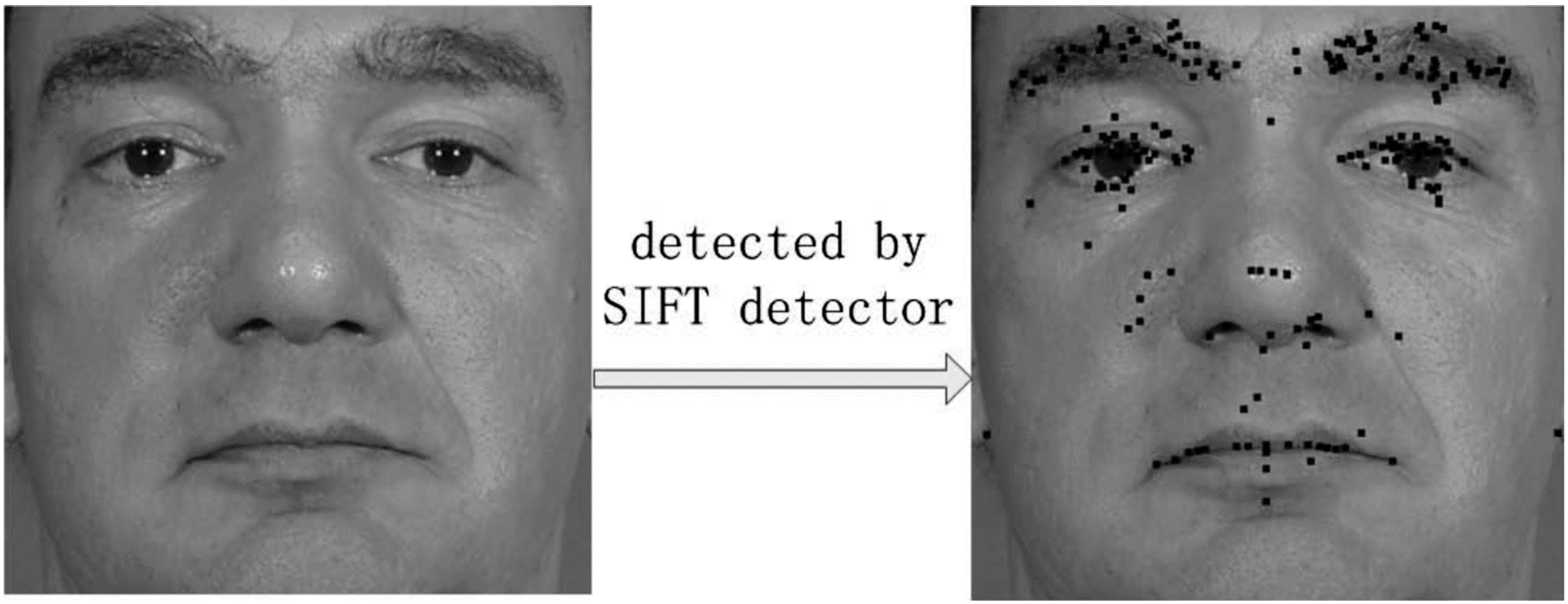


Focus of Attention

- Basic focus-of-attention strategy
  - Build an image pyramid
  - Subtract one layer from another
    - This creates images of DoG responses
    - $I \otimes (M-N) = (I \otimes M) - (I \otimes N)$
  - Find extrema in x, y, and σ of the DoG responses
    - Both positive and negative
- The image windows around the DoG extrema are fixed-size "focus of attention" windows.

Colorado State University

SIFT (DoG)
Interest Points Example



http://opticalengineering.spiedigitallibrary.org/article.aspx?articleid=1089401

Example #2

This image shows:

- Scale (circle size)
- Dominant Orientation
  - 1^{st} eigenvector of Harris operator

http://computervisionblog.wordpress.com/tag/sift-feature-point/