# Biologically Inspired Bottom-Up Visual Attention Model

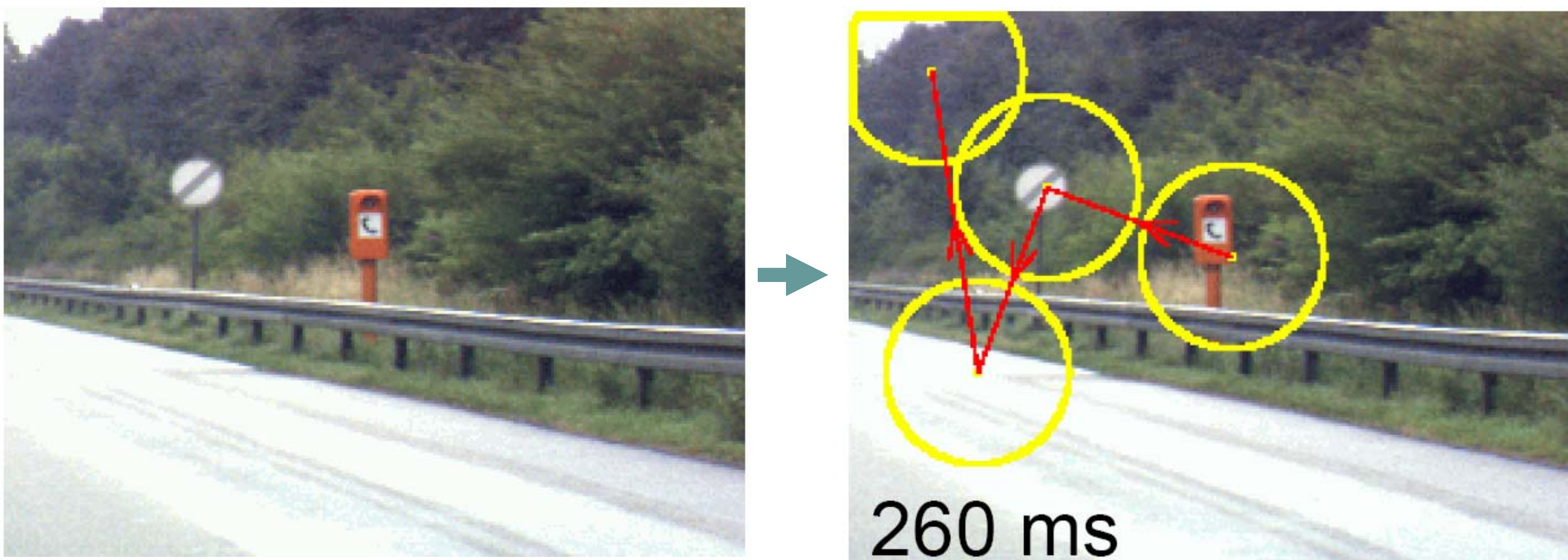
Laurent Itti, et al, 1998

Mike Onorato

COS 598B

# Overview

- Task: Detect most salient regions.
- Decrease dimensionality

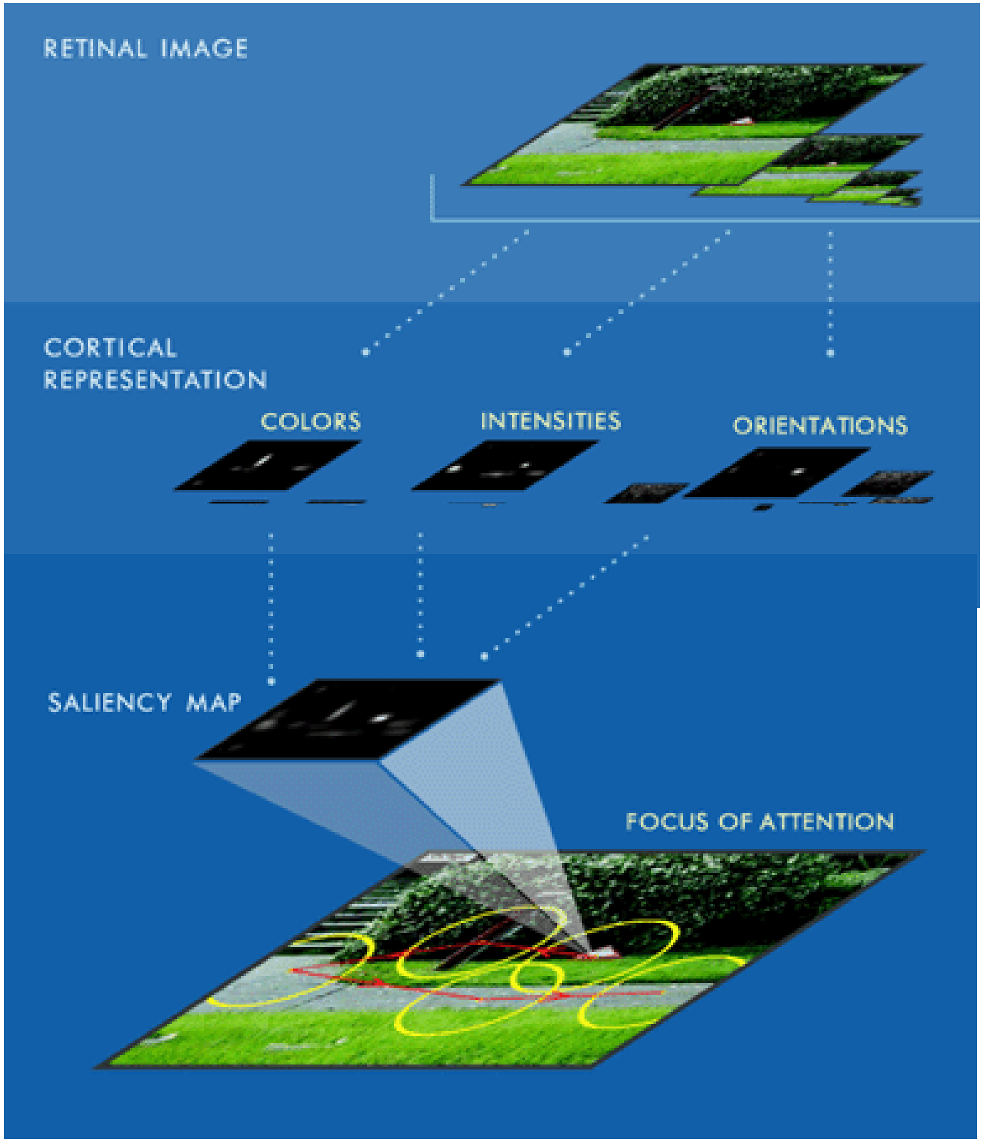



Source: [Laurent Itti, USC iLab: http://ilab.usc.edu/bu/]  (Same for all images unless otherwise specified.)

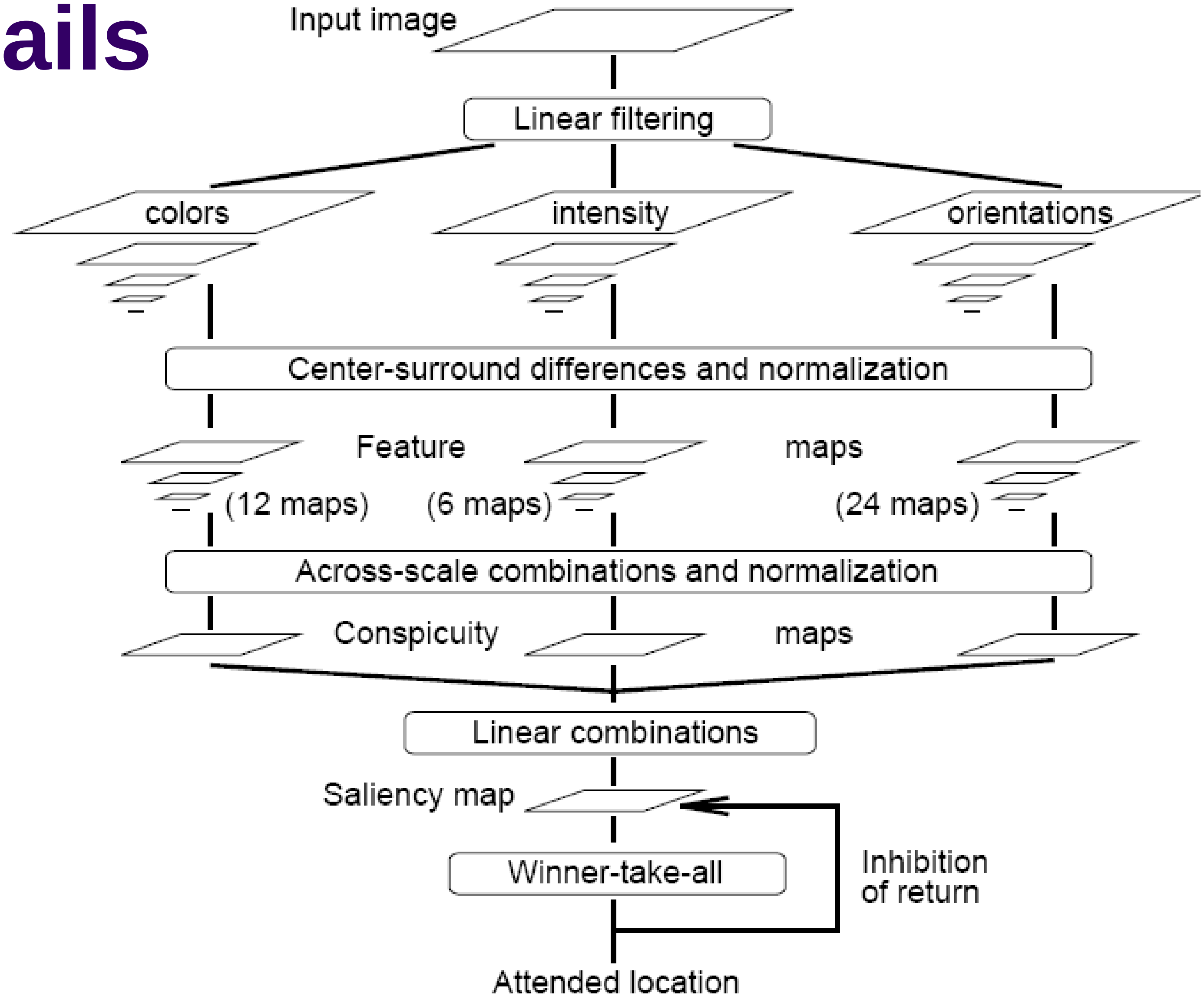
# Part 1: Model

# Overview

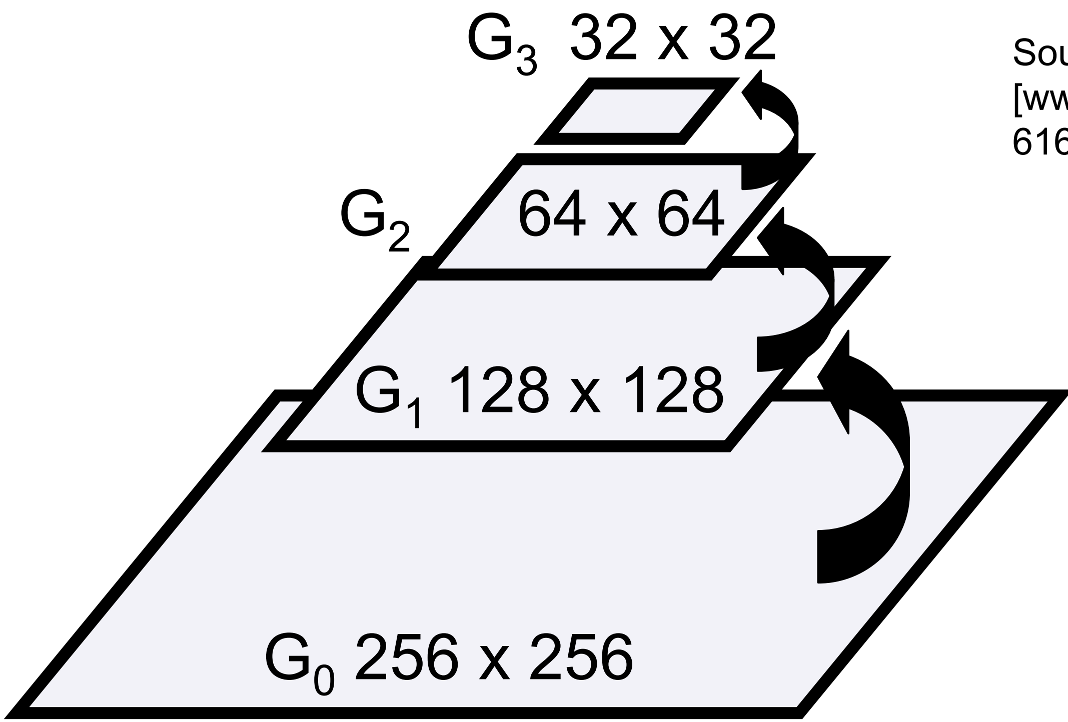
- Bottom-Up Method



# Details


Input image
Linear filtering
colors
intensity
orientations
Center-surround differences and normalization
Feature
maps
(12 maps)
(6 maps)
(24 maps)
Across-scale combinations and normalization
Conspicuity
maps
Linear combinations
Saliency map
Winner-take-all
Inhibition of return
Attended location

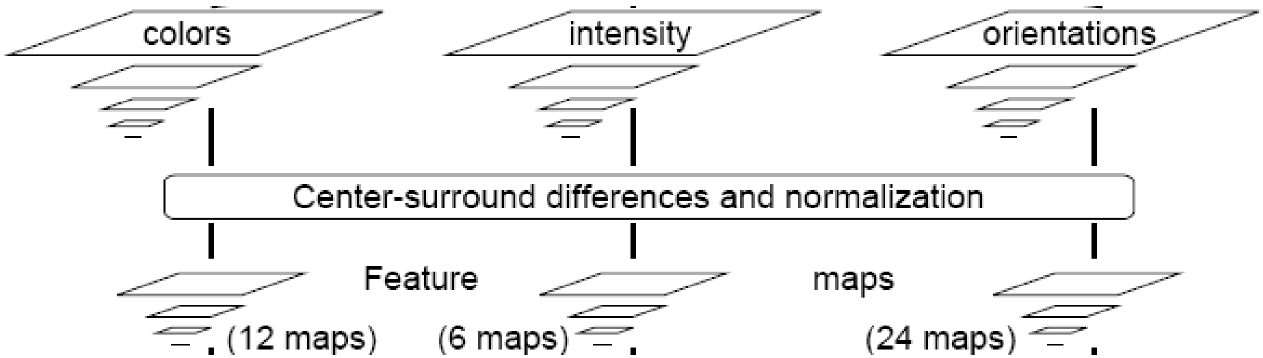
# Step 1: Linear Filtering

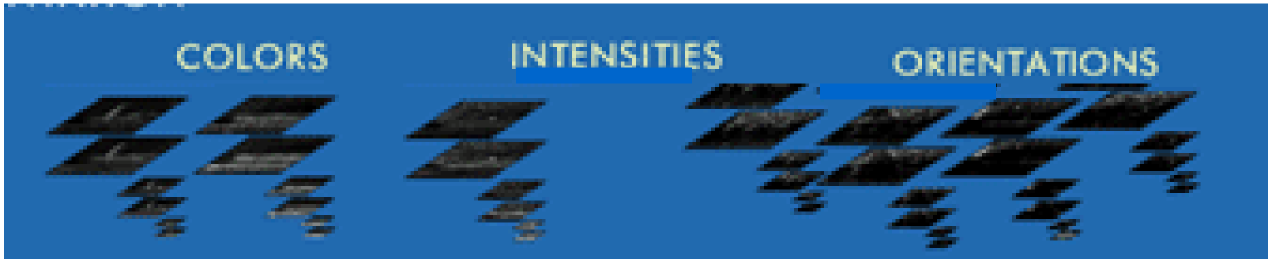



Source: [www.singularsys.com/research/courses/616/funk-project-pres.ppt]

- 9 scales from 1:1 to 1:256

# Step 2: Extract Feature Maps

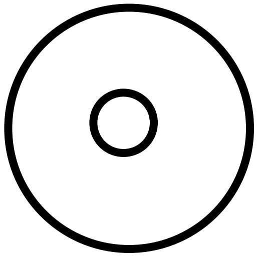


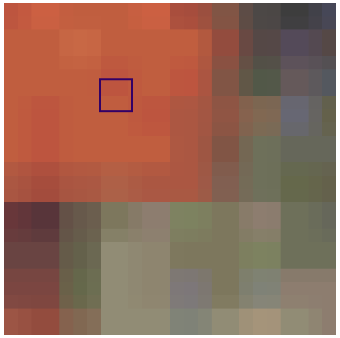




- Center-surround operations at multiple scales

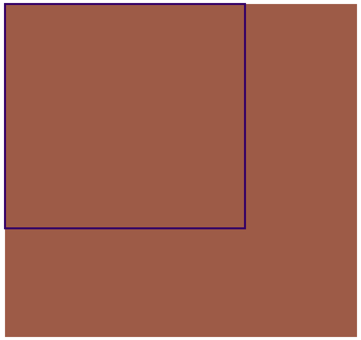
# Center-Surround Operations

- Difference between value of pixel at two different scales:
  - c: center pixel scale
  - s: surround pixel scale





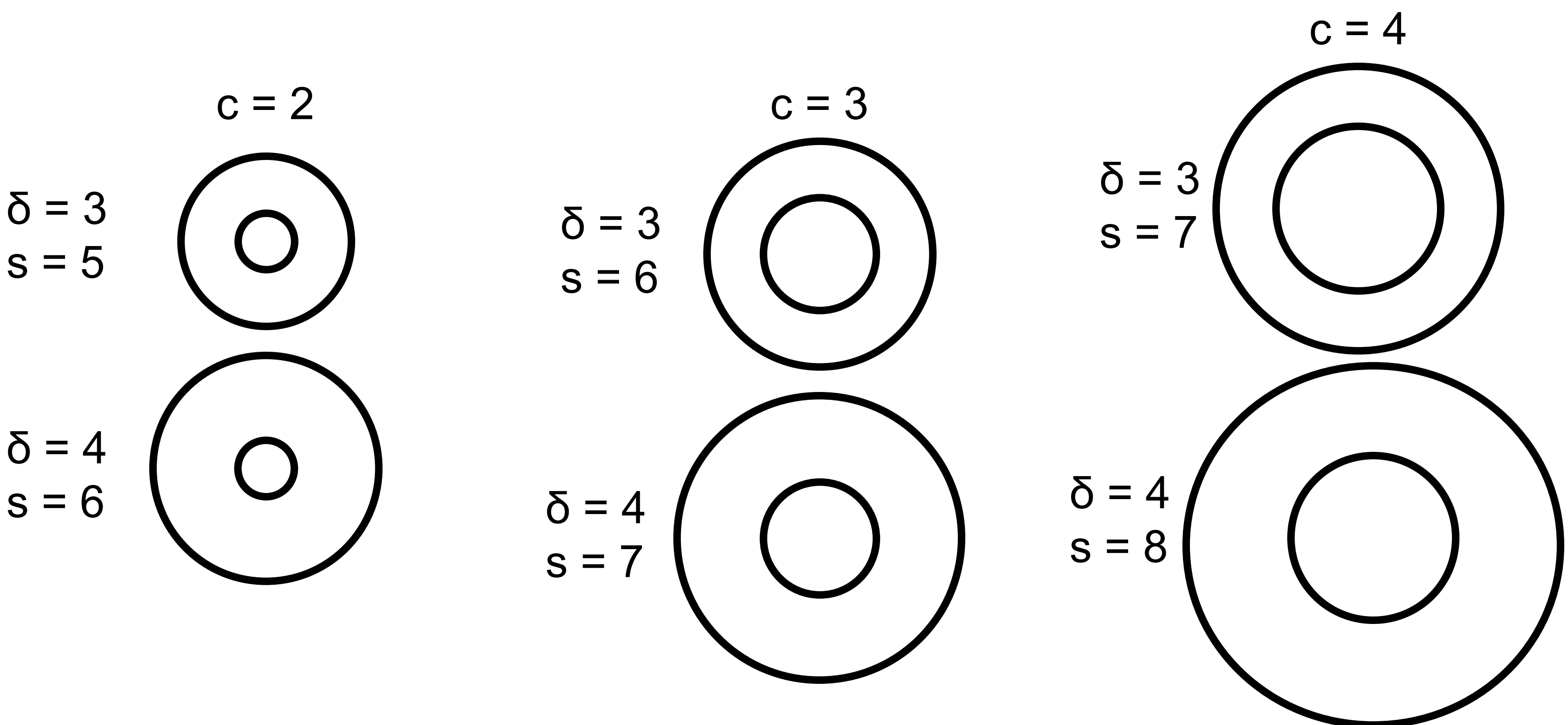
c



s

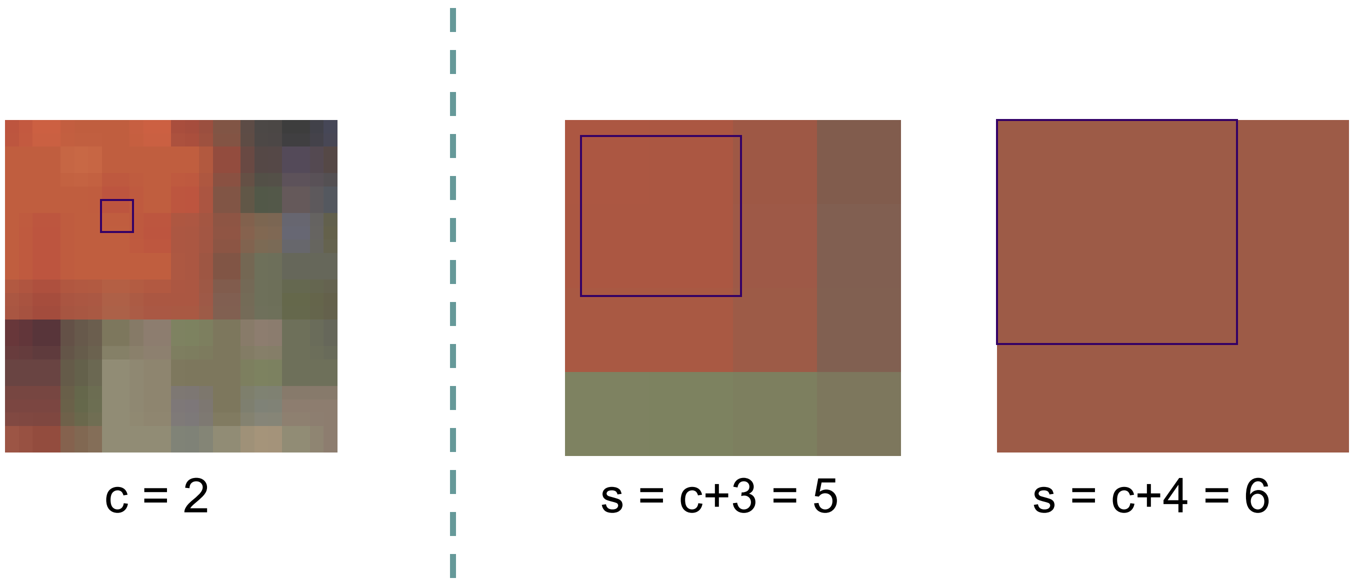
# Center-Surround Operations

- c = pixel at scale c Є {2,3,4}
- s = pixel at scale c+δ, where δ Є {3,4}
- 6 different scale combinations



# Center-Surround Operations

- c = pixel at scale c Є {2,3,4}
- s = pixel at scale c+δ, where δ Є {3,4}
- 6 different scale combinations



# Step 2a: Extract Intensity Maps

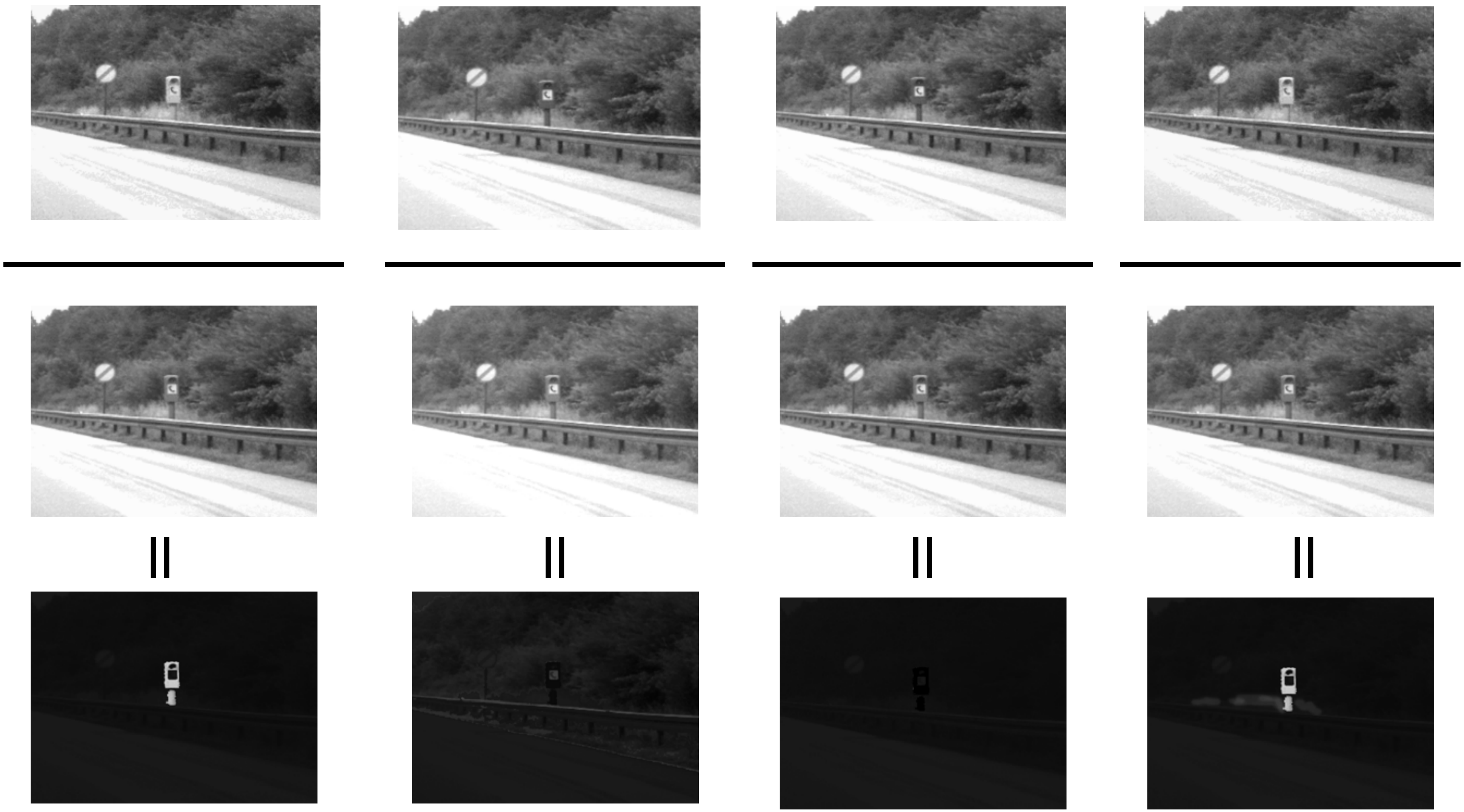
I(x): intensity map at scale x

θ: pixel difference operation

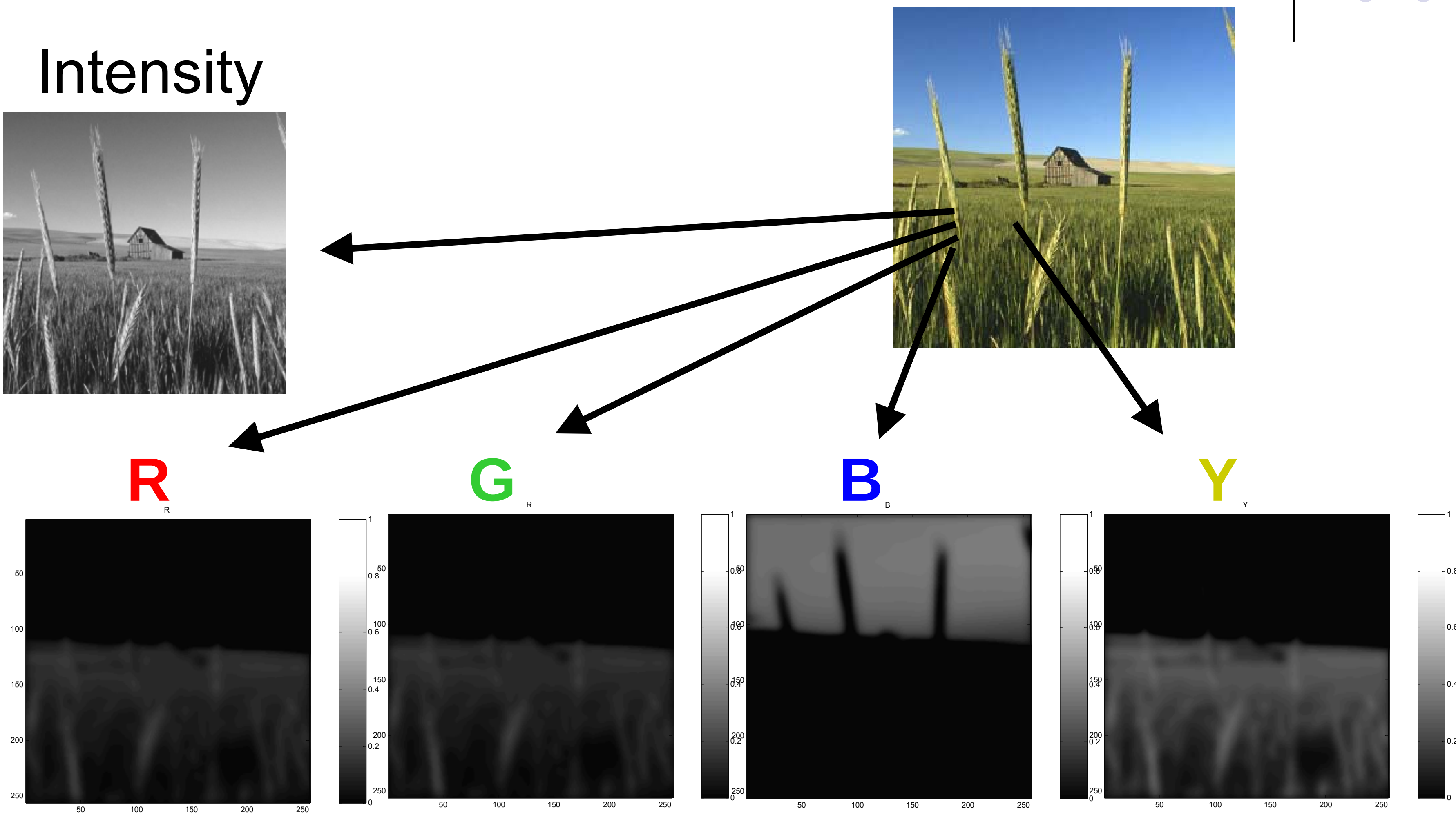
$\mathcal{I}(c,s) = |I(c)\ \theta\ I(s)|$

# Step 2b: Color Maps

Red: R(x) Green: G(x) Blue: B(x) Yellow: Y(x)



# Step 2b: Color Maps

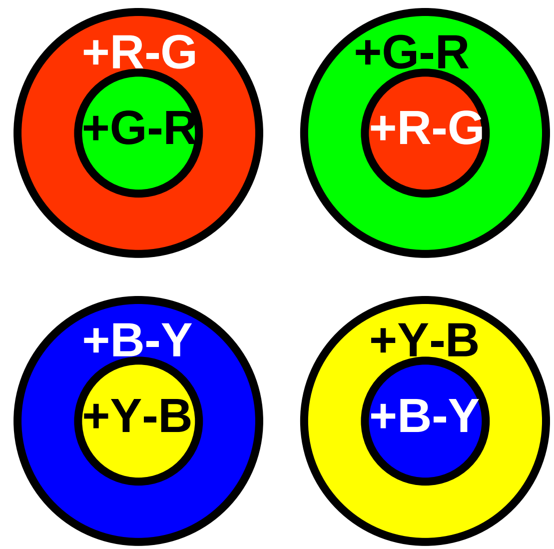



Source: [www.singularsys.com/research/courses/616/funk-project-pres.ppt]

# Step 2b: Extract Color Maps

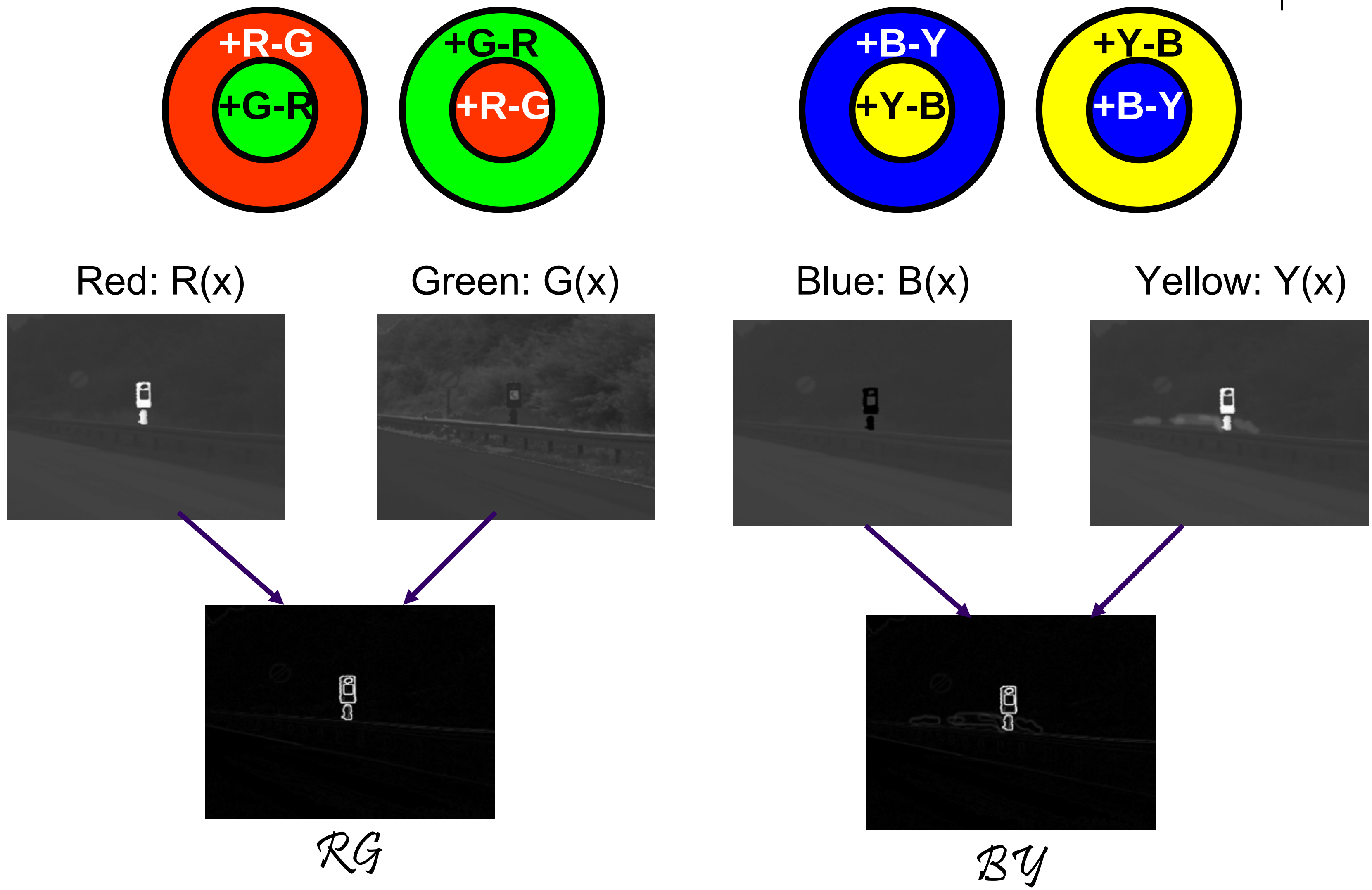
- Create a red-green and blue-yellow color map

$$\mathcal{RG}(c,s) = |( R(c) - G(c) ) \theta ( G(c) - R(c) )|$$
$$\mathcal{BY}(c,s) = |( B(c) - Y(c) ) \theta ( Y(c) - B(c) )|$$



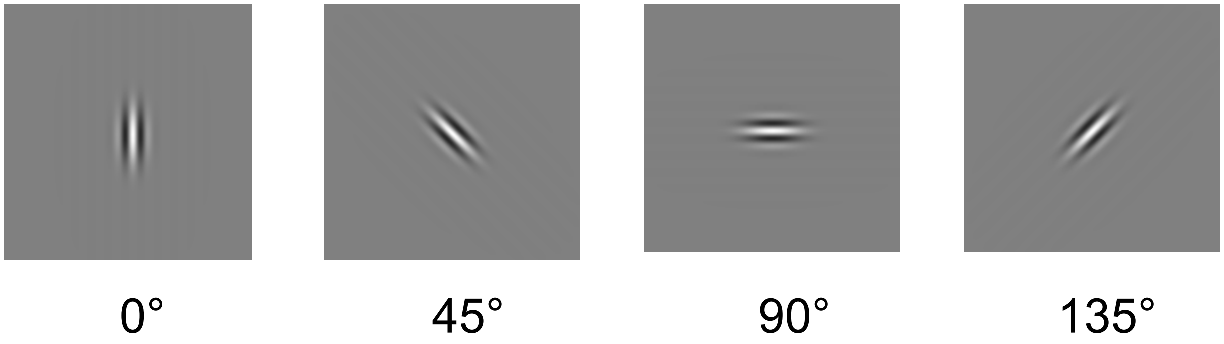

Source:
[www.singularsys.com/research/courses/616/funk-project-pres.ppt]

# Step 2b: Extract Color Maps



# Step 2c: Orientation Maps

- Gabor Filtering:
  - Difference between image and Gabor filter
  - Gabor filters for 4 different orientations




Source: [http://www.cs.rug.nl/~imaging/simplecell.html]

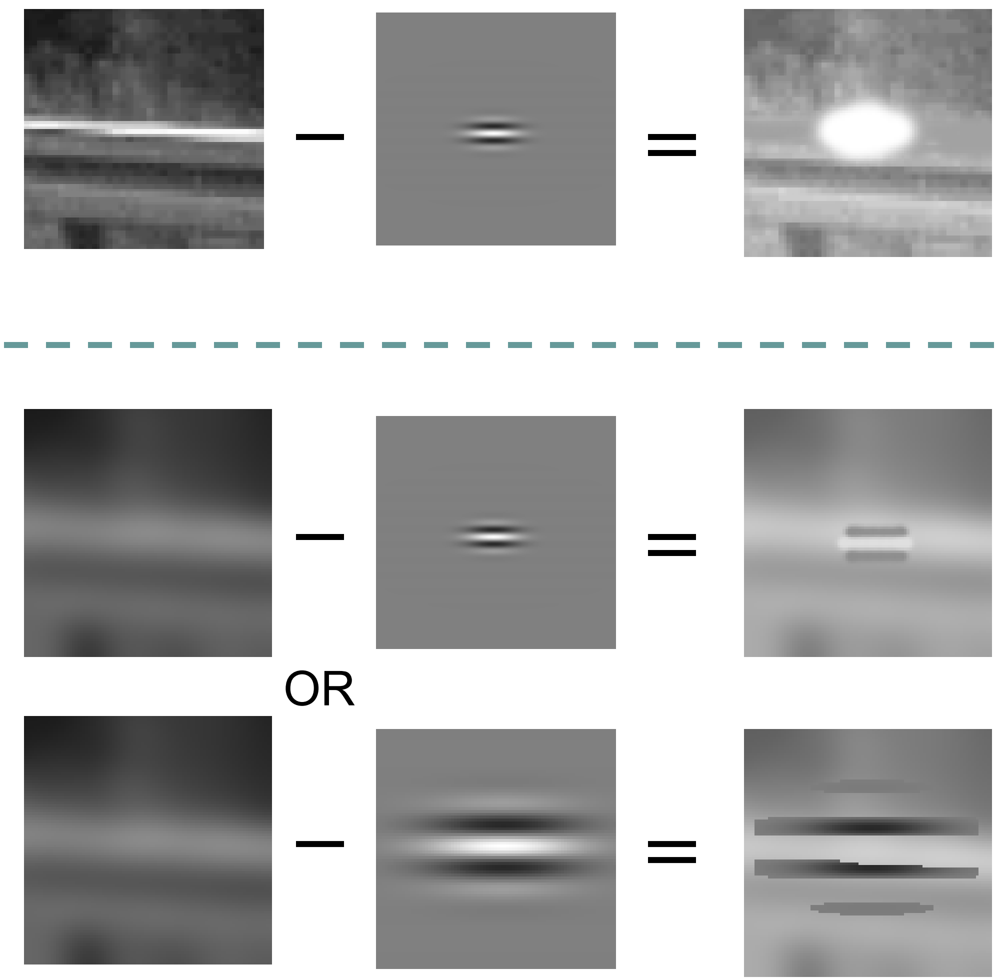
# Step 2c: Orientation Maps

O(x,θ):

θ = 90°

x = 2

x = 5



# Step 2c: Orientation Maps

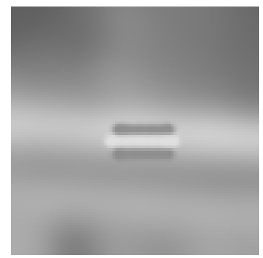
$$\mathcal{O}(c,s,\theta) = |O(c,\theta)\ \theta\ O(s,\theta)|$$

$\theta = 90°$

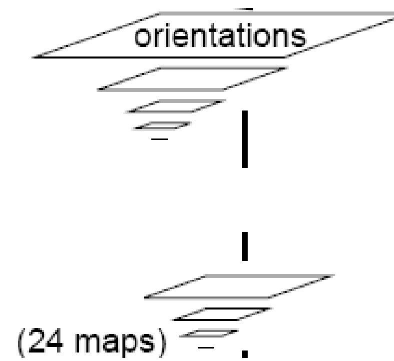
$s = 2$

$c = 2+3 = 5$

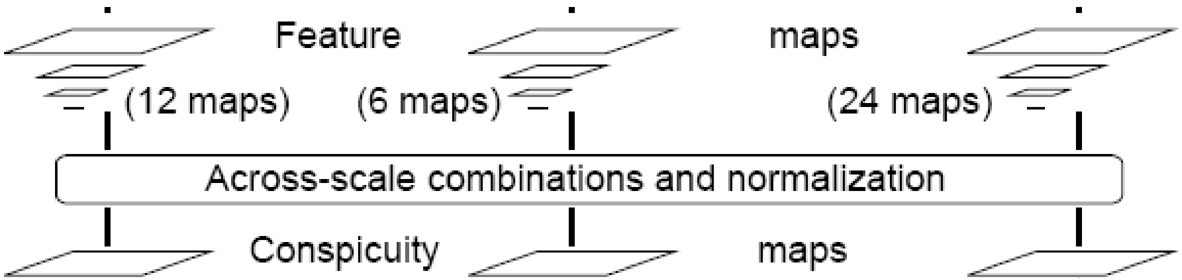
$O(c,\theta)$

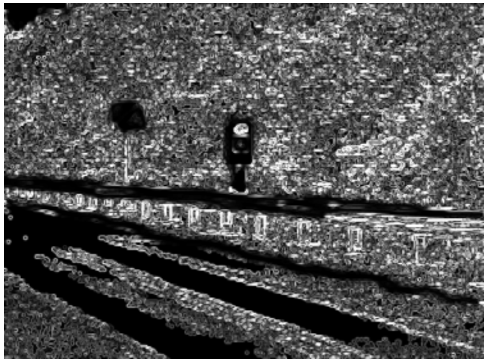


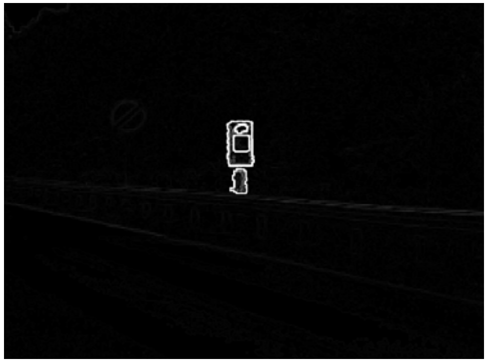
$O(s,\theta)$




# Step 3: Combine Feature Maps Into Conspicuity Maps

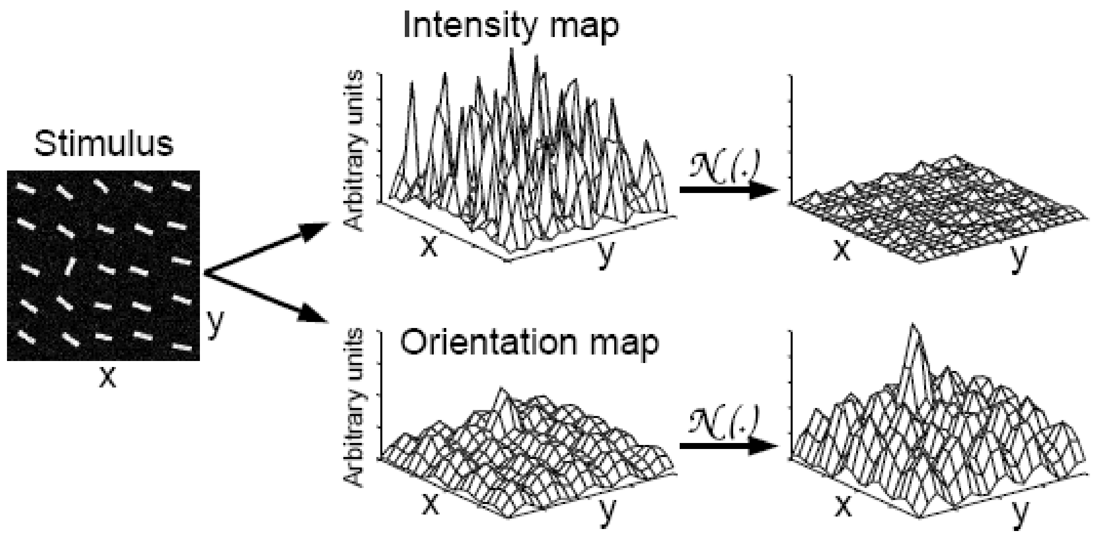


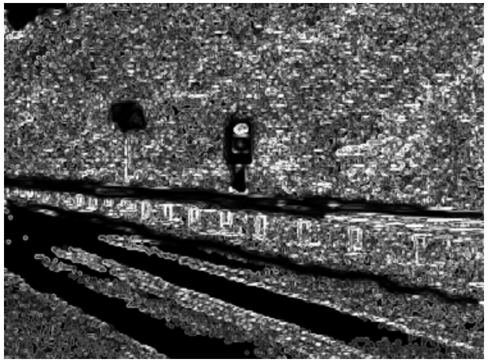



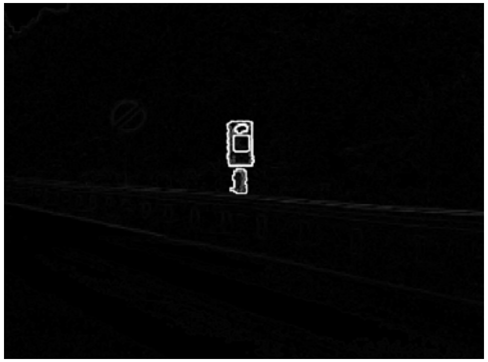


# Step 3: Combine Feature Maps Into Conspicuity Maps

- Normalize map values to range [0..1]
- **m** = avg local max
- $\mathcal{N}$ = (1-**m**)$^2$

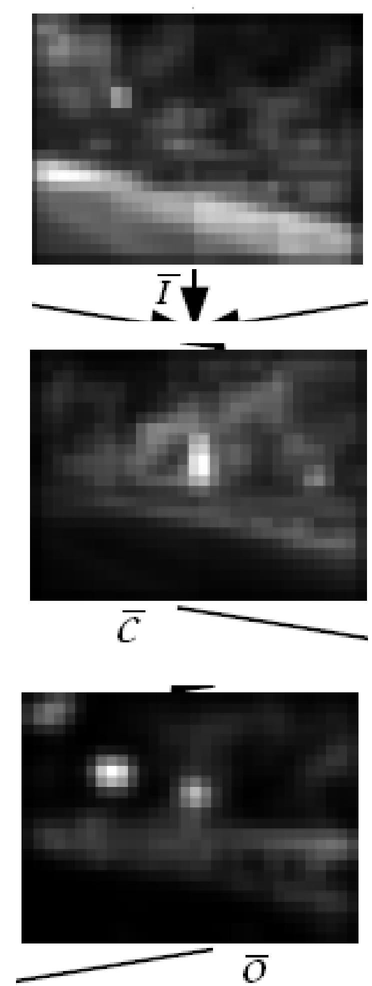




# Step 3: Combine Feature Maps Into Conspicuity Maps

- Conspicuity Maps:

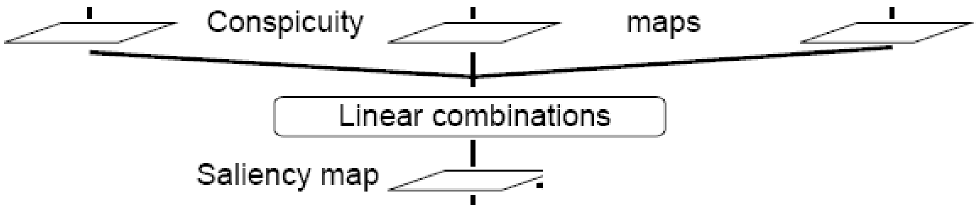



$$\overline{\mathcal{I}} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{I}(c,s))$$

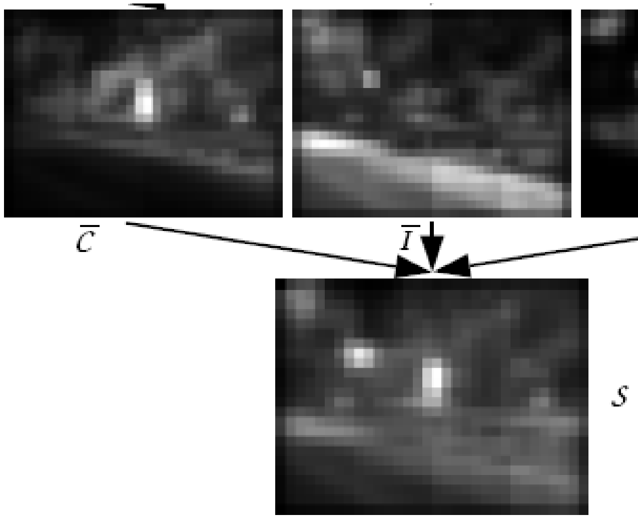
$$\overline{\mathcal{C}} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(\mathcal{RG}(c,s)) + \mathcal{N}(\mathcal{BY}(c,s))]$$

$$\overline{\mathcal{O}} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} \mathcal{N}\left( \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{O}(c,s,\theta)) \right)$$

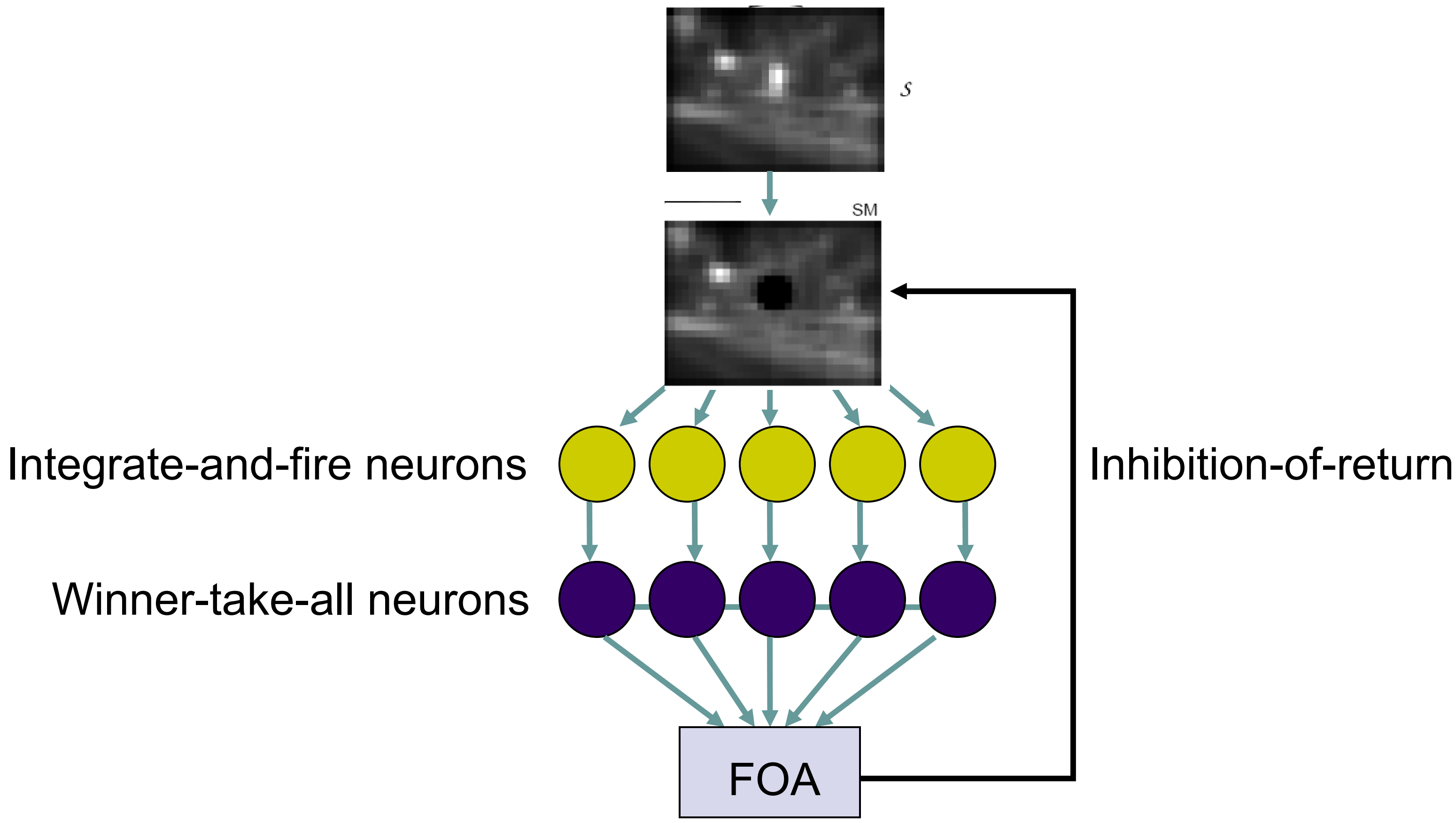
# Step 4: Combine Conspicuity Maps Into Saliency Map



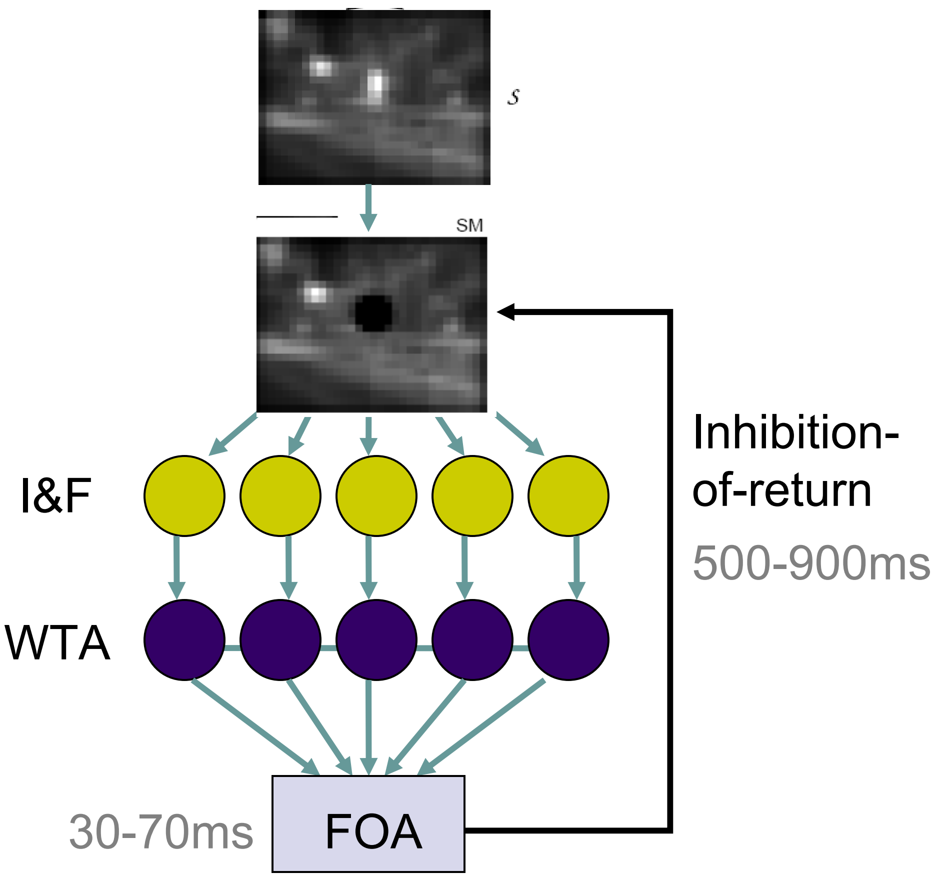

$$\mathcal{S} = \frac{1}{3}\left(\mathcal{N}(\overline{\mathcal{I}}) + \mathcal{N}(\overline{\mathcal{C}}) + \mathcal{N}(\overline{\mathcal{O}})\right)$$

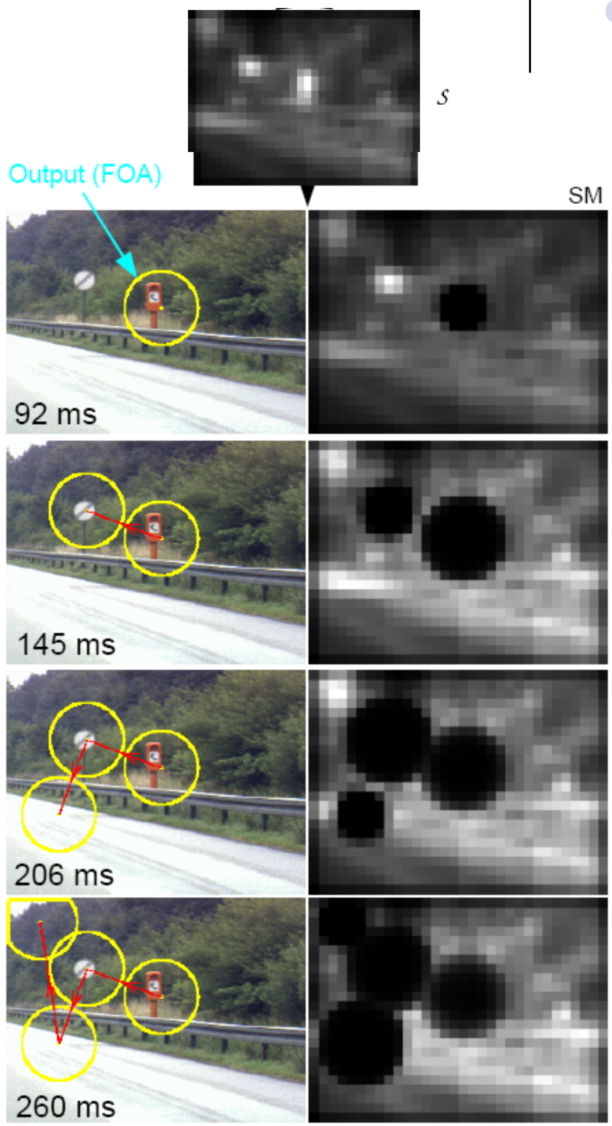


# Step 5: Process regions in order of saliency



# Step 5: Process regions in order of saliency

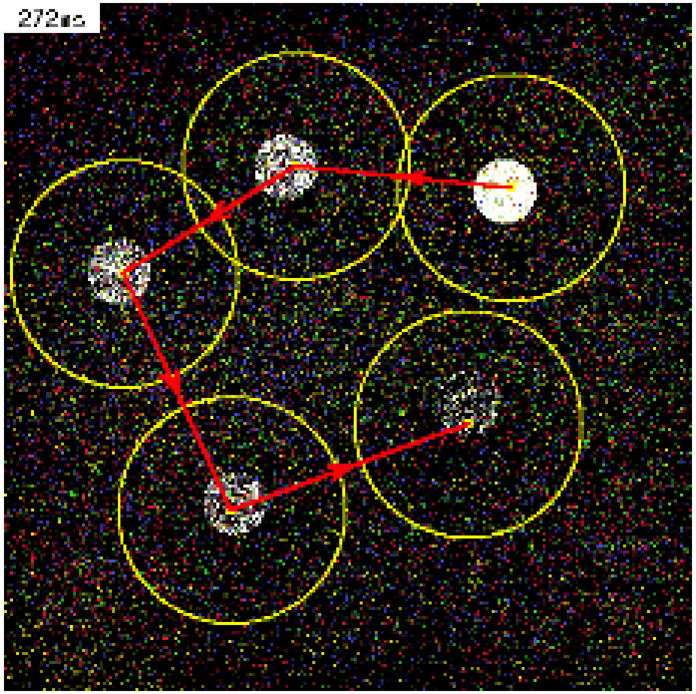


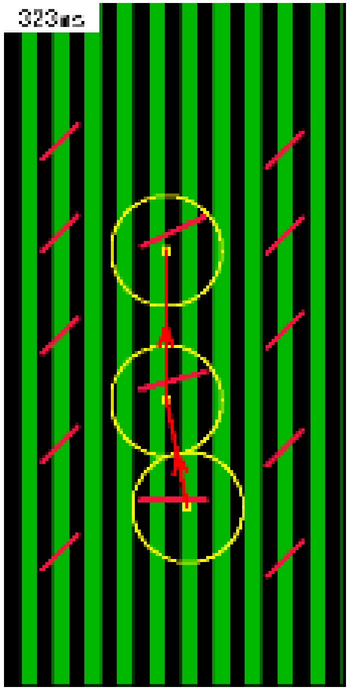



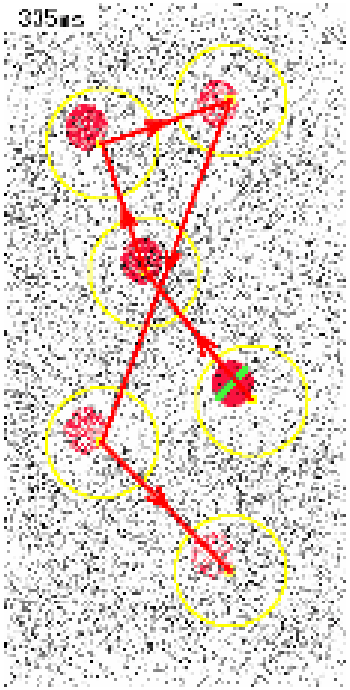
# Part 2: Results

# Experiments

- Same shape, different contrast, orientation or color

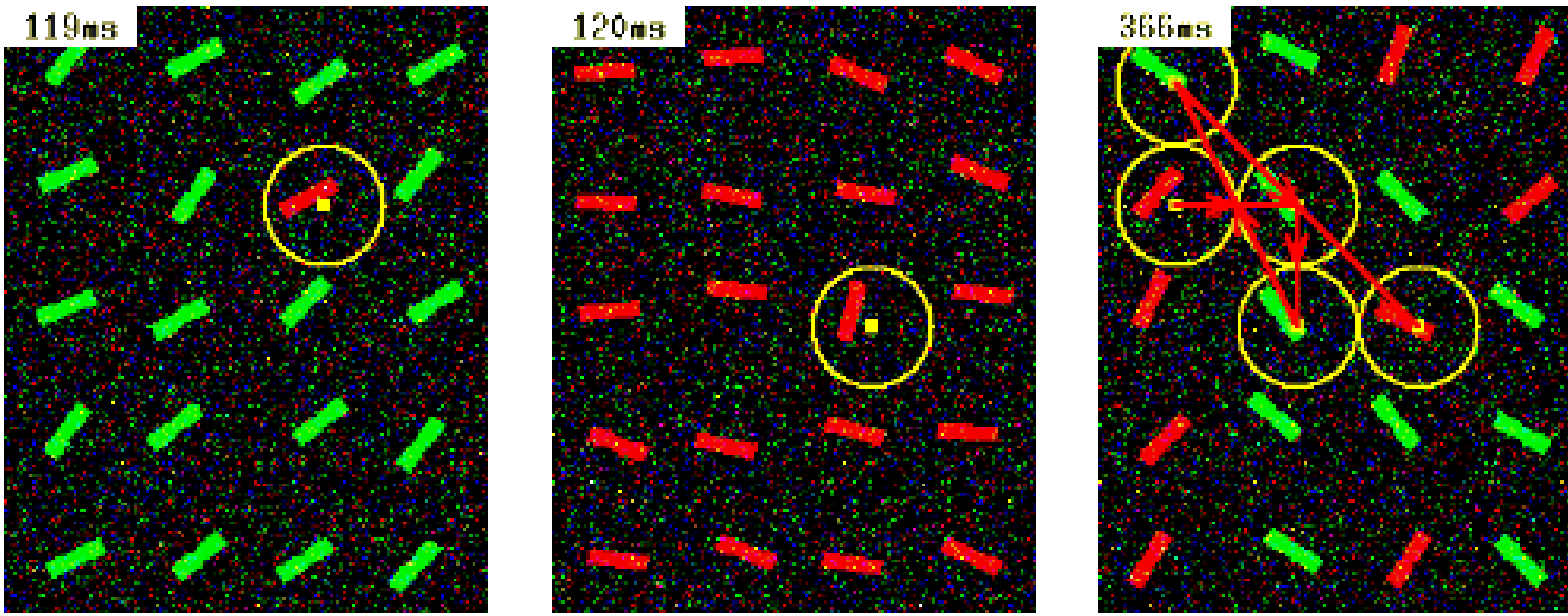


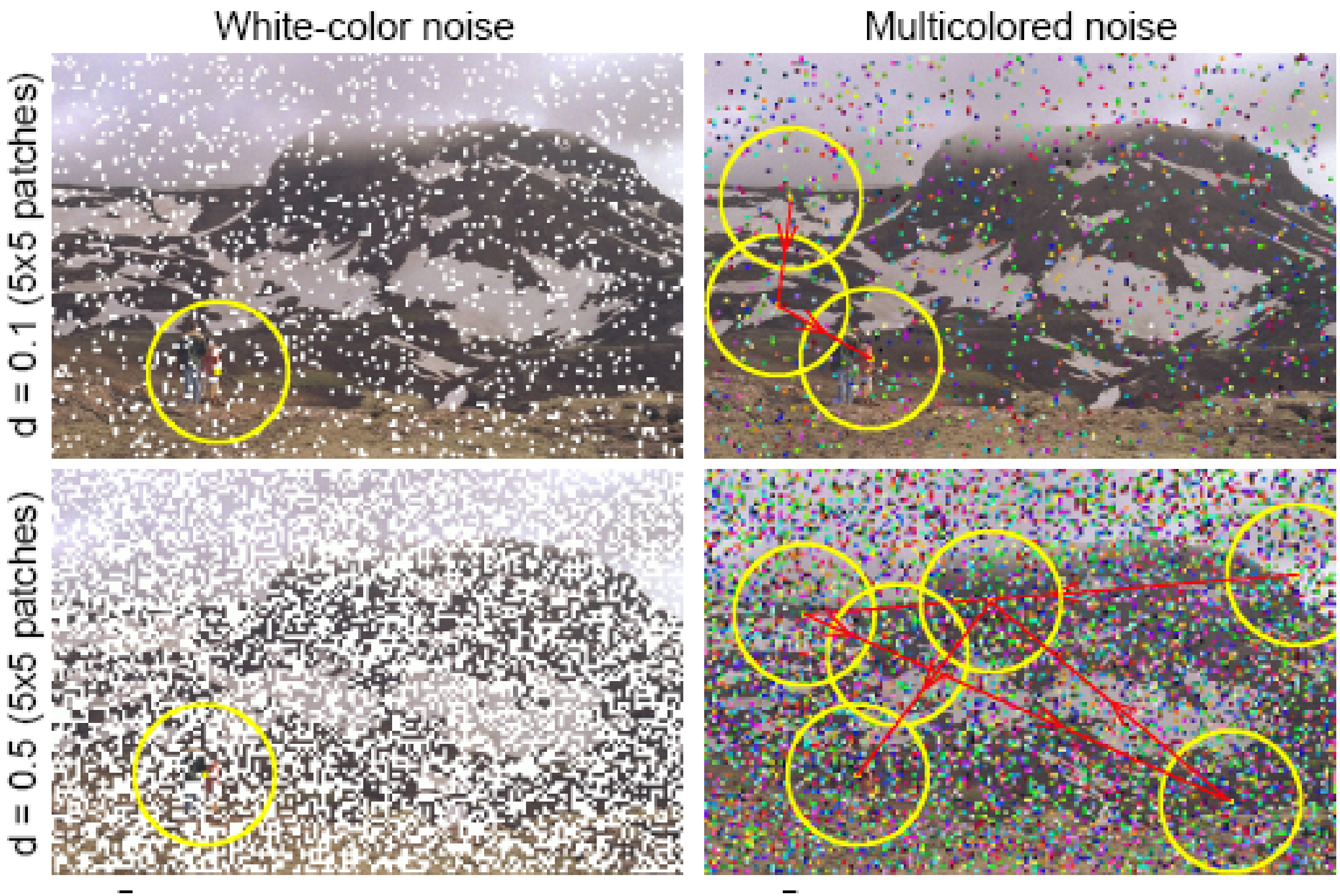



# Experiments

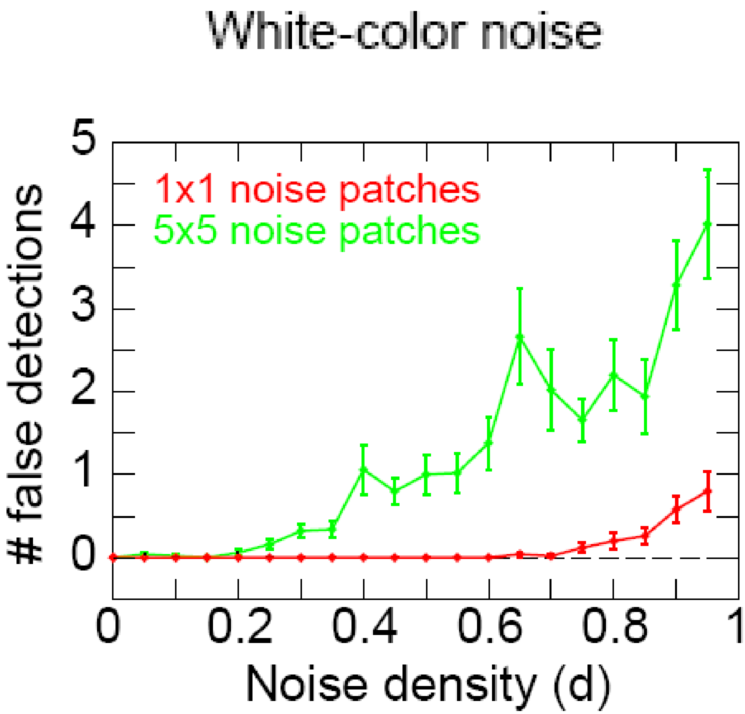
- Pop-out:

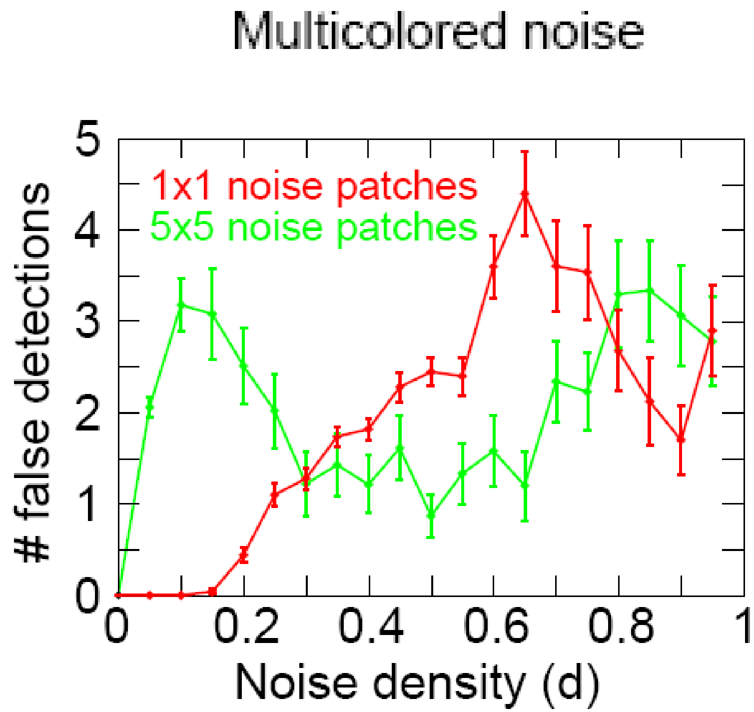


# Noise Sensitivity Experiment



# Noise Sensitivity Experiment

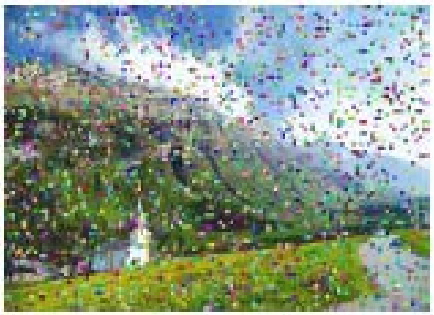


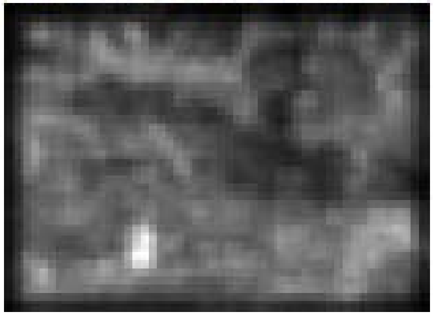




- Only one image
- # trials per density not stated

# Spatial Frequency Content Models

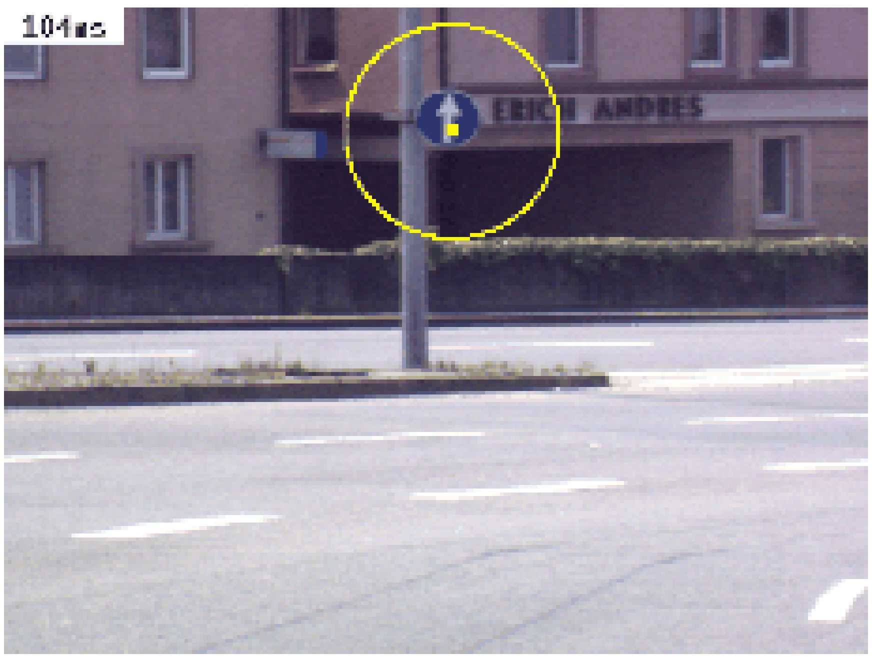
- Eye-tracking study shows certain locations are attended to more than others [Reinagel and Zador]
- Measured spatial frequency content (SFC) by:
  - At each image location, extract 16x16 patch of I(2), R(2), G(2), B(2), and Y(2)
  - Apply 2D Fast Fourier Transform

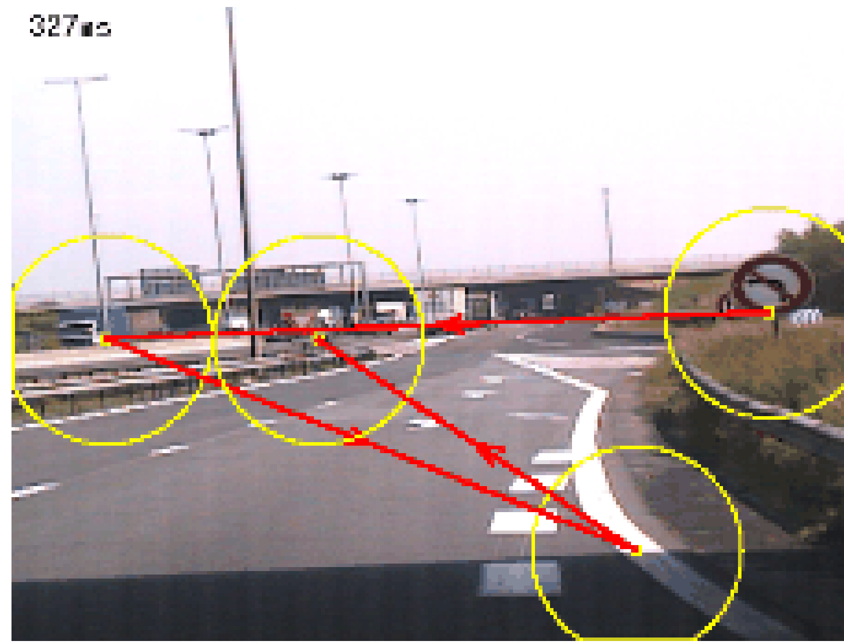




# SFC Comparison Experiment

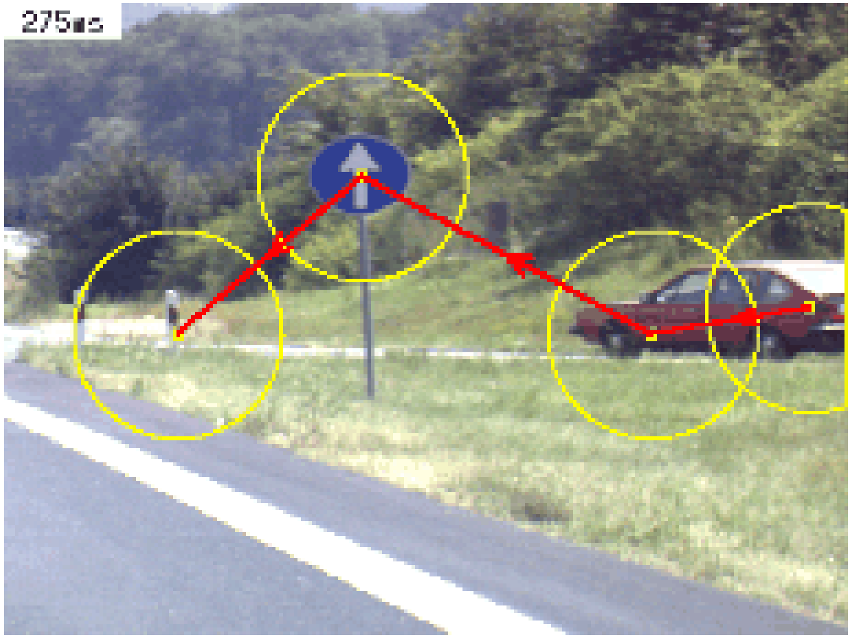
- Dataset:
  - Natural scenes with traffic signs (90 images)

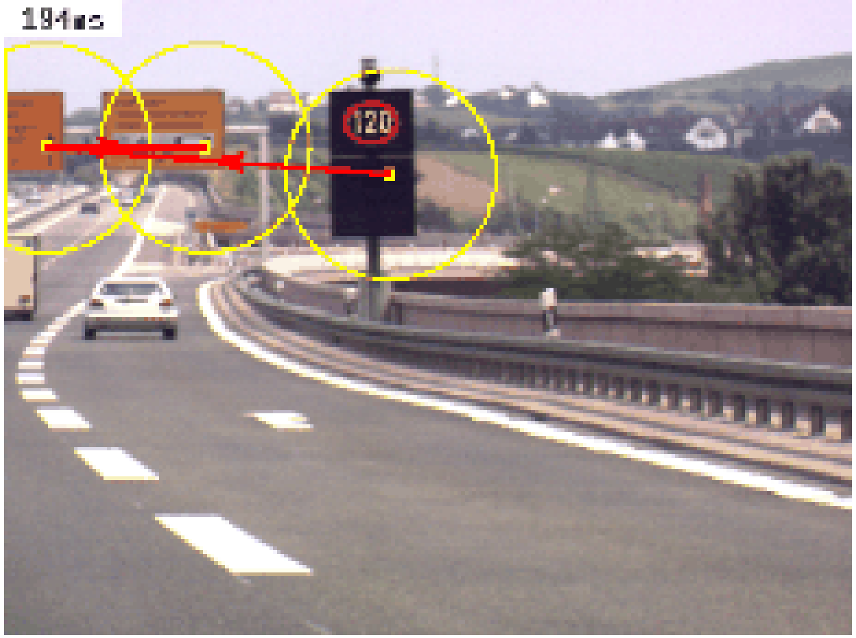


# SFC Comparison Experiment

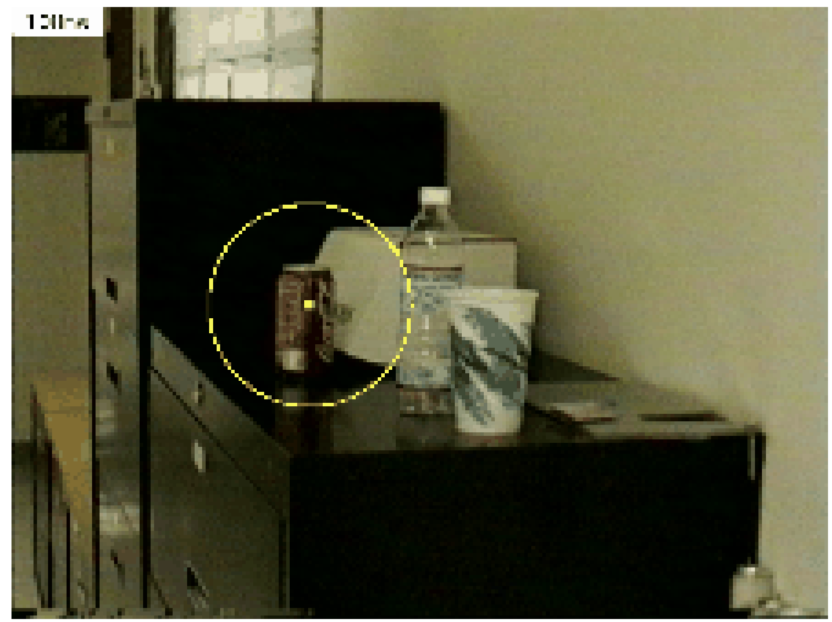
- Dataset:
  - Natural scenes with traffic signs (90 images)



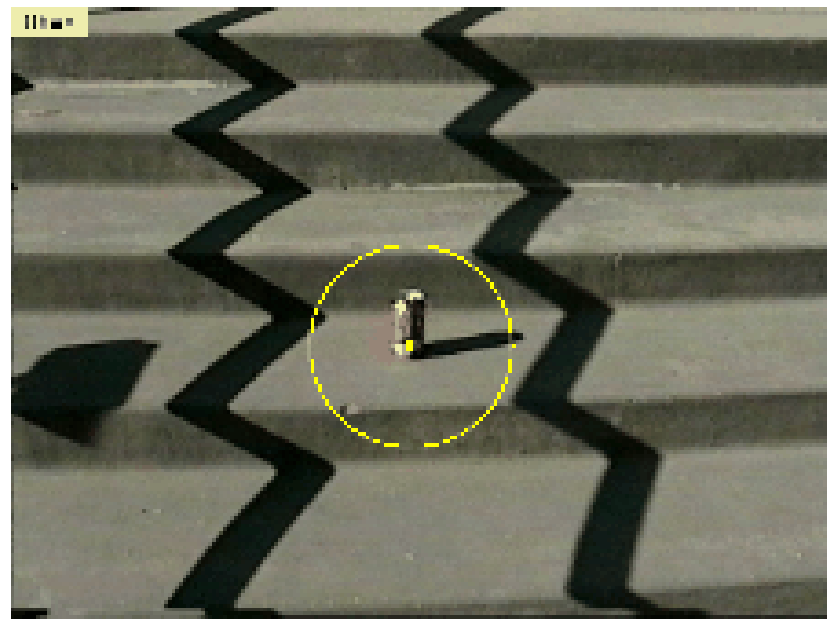
# SFC Comparison Experiment

- Dataset:
  - Natural scenes with traffic signs (90 images)
  - Red soda can (104 images)



# SFC Comparison Experiment

- Dataset:
  - Natural scenes with traffic signs (90 images)
  - Red soda can (104 images)



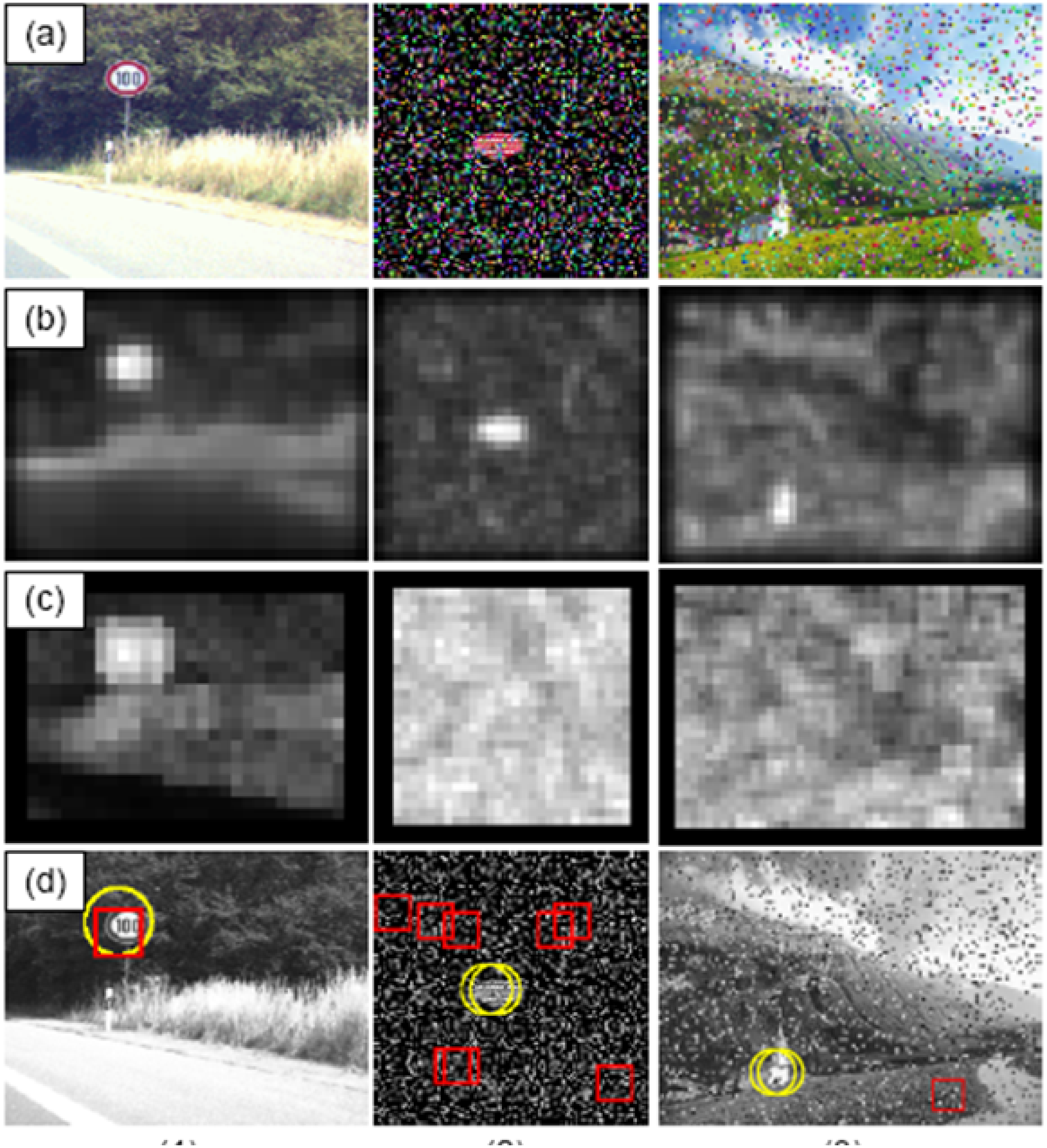
# SFC Comparison Experiment

- Dataset:
  - Natural scenes with traffic signs (90 images)
  - Red soda can (104 images)
  - Vehicle’s emergency triangle (64 images)

# Results

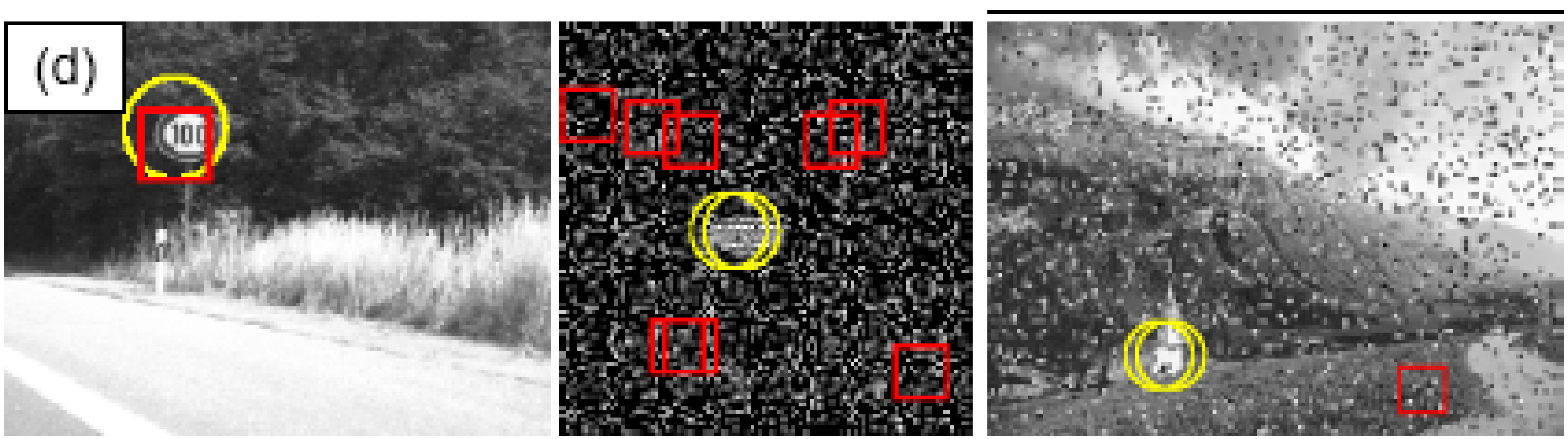
Spatial Frequency Content Maps (Red)

Saliency Maps (Yellow)



# SFC Comparison Experiment

- Results:
  - 1st location: SFC 2.5 ± 0.05 times the average SFC
  - …
  - 8th location: SFC 1.6 ± 0.05 times the average SFC



# Military Vehicle Experiment

- Time taken to attend to military vehicle
- Compare to 62 human observers

# Military Vehicle Experiment

- Time taken to attend to military vehicle
- Compare to 62 human observers

# Military Vehicle Experiment

- Time taken to attend to military vehicle
- Compare to 62 human observers

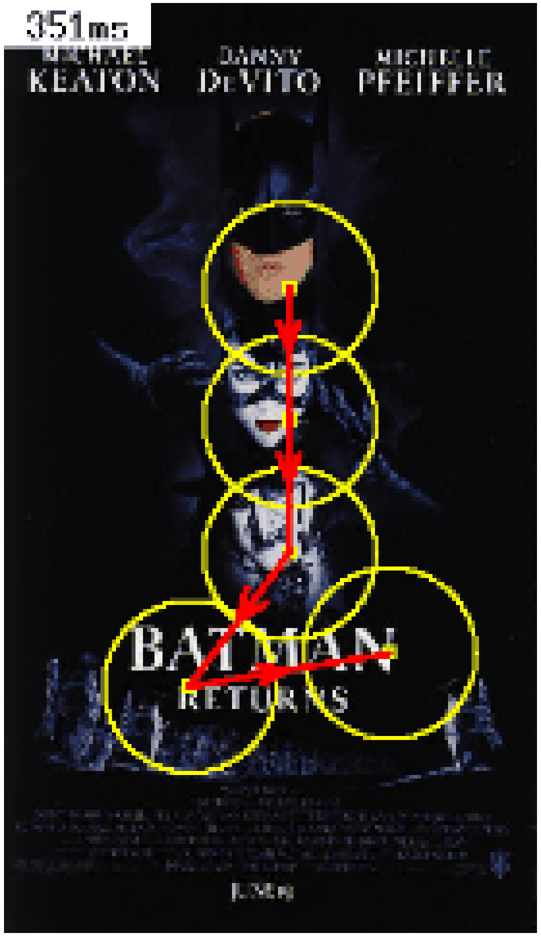
# Military Vehicle Experiment

- Time taken to attend to military vehicle
- Compare to 62 human observers

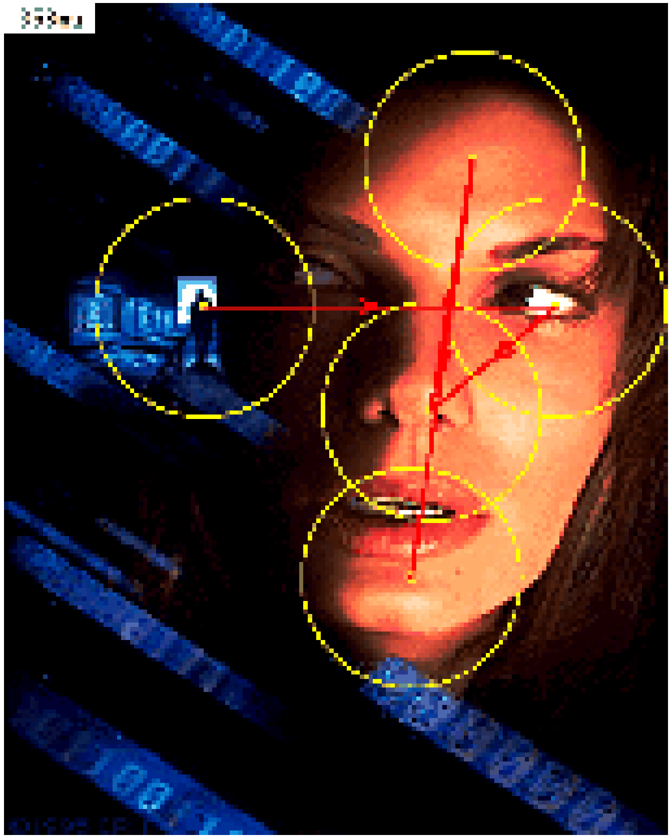
# Military Vehicle Experiment

- Results:
  - Itti's model finds target in fewer attentional shifts in 75% of trials

# Natural Images




# Natural Images



# Why the Model is Effective

- Fast
  - Parallel processing
  - No top-down knowledge
- Similar to primate visual system

# Why it Models the Primate Visual System Closely

- Parallel and bottom-up maps
- Maps of orientation, intensity and color
- Linear filtering
- Center-surround operations
- Winner-take-all
- Slow sequential attention shifting

# Criticisms of the Model

- Cannot detect junctions of features
- Cannot detect features other than color, intensity and orientation
- No content completion or closure
- Does not include magnocellular motion channel

# Criticisms of the Experiments

- Noise experiment was not thorough
- Running time data not given
- No quantitative results for pop-out experiments

# Additional Experiments

- Compare to data from human eye-tracking
- Extend this framework to do additional tasks and provide experimental results
  - Scene classification