Computational Saliency Models

Cheston Tan, Sharat Chikkerur
{cheston,sharat}@mit.edu

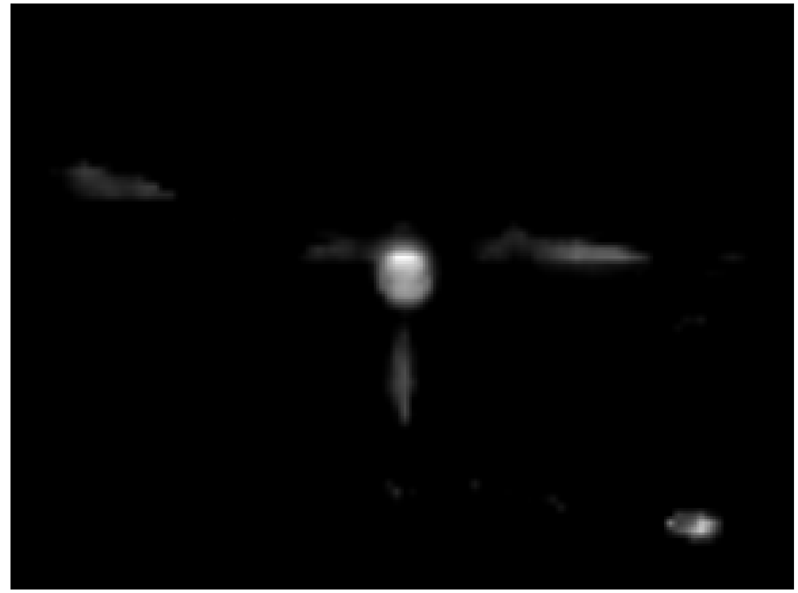
Outline

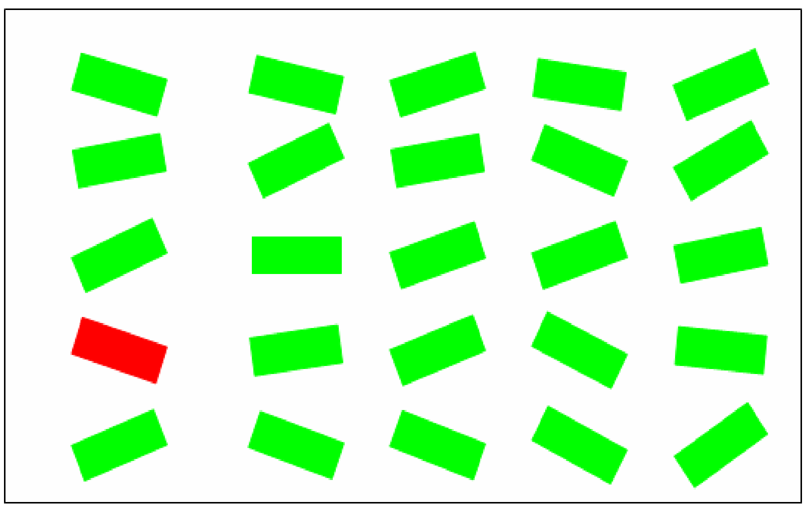
- **Saliency 101**
- **Bottom up Saliency Model**
  - Itti, Koch and Neibur, “A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, 20(11), ‘98
  - Itti and Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention”, Vision Research, ‘00
- **Contextual Guidance Model : (Bottom up + Top Down)**
  - A. Torralba, A. Oliva, M. Castelhano and J. M. Henderson, “Contextual guidance of attention and eye movements in real-world scenes: the role of global features in object search”. Psychological Review, ’06
  - A. Torralba,“Modeling global scene factors in attention", JOSA, 20(7), 03
  - A. Torralba, "Contextual Priming for Object Detection", IJCV, 53(2), 03
- **Summary**
- **Demo**
  - Comparison of bottom up saliency models

Saliency 101

- **What is saliency?**
  - But first, what is "Attention"?
  - Biological visual system process complex scenes 'serially' (despite parallel computation)
  - Specific parts of the scene are "attended" by covert or overt attention (eye movements).
- **What drives attention?**
  - Saliency!
    - Bottom up saliency
      - Driven by scene features, Fast!
    - Top down saliency
      - Driven by volitional control, Slow
- **Biological Evidence**
  - Believed to be located in posterior pareital cortex ,V4
  - Spike modulation observed in V1,V2,V4 (Luck et al., '97; Reynolds et al '00)

Bottom up saliency

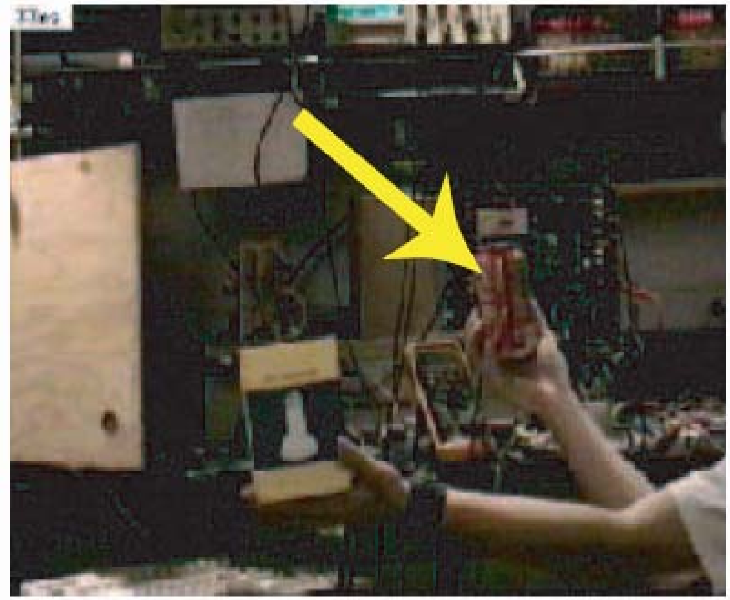


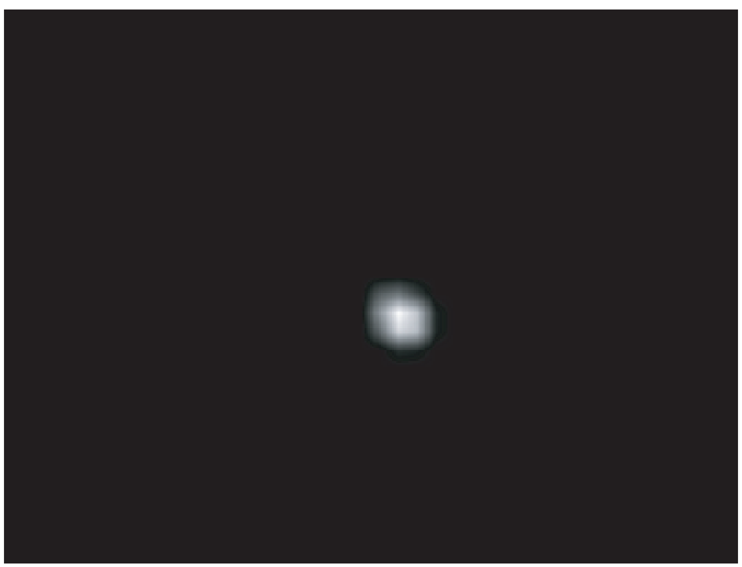


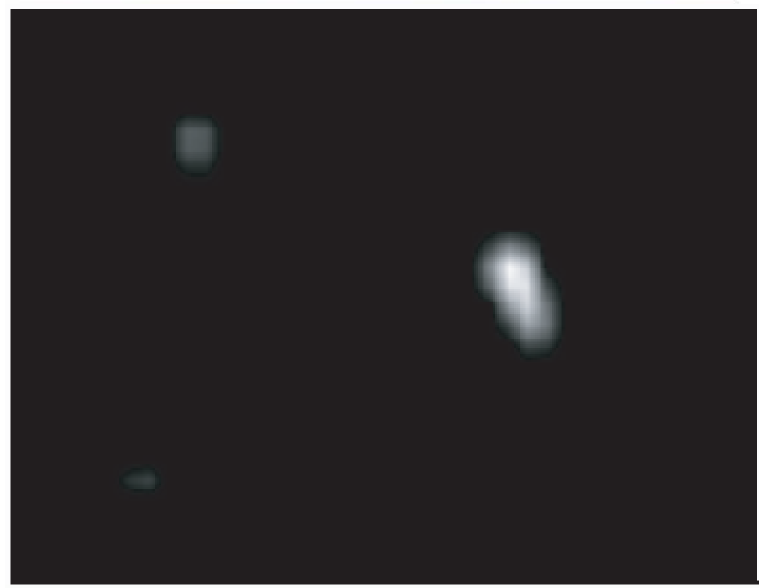
Top down saliency- Spatial Modulation

(Torralba)

Top down saliency-Feature modulation







(Navalpakkam and Itti)

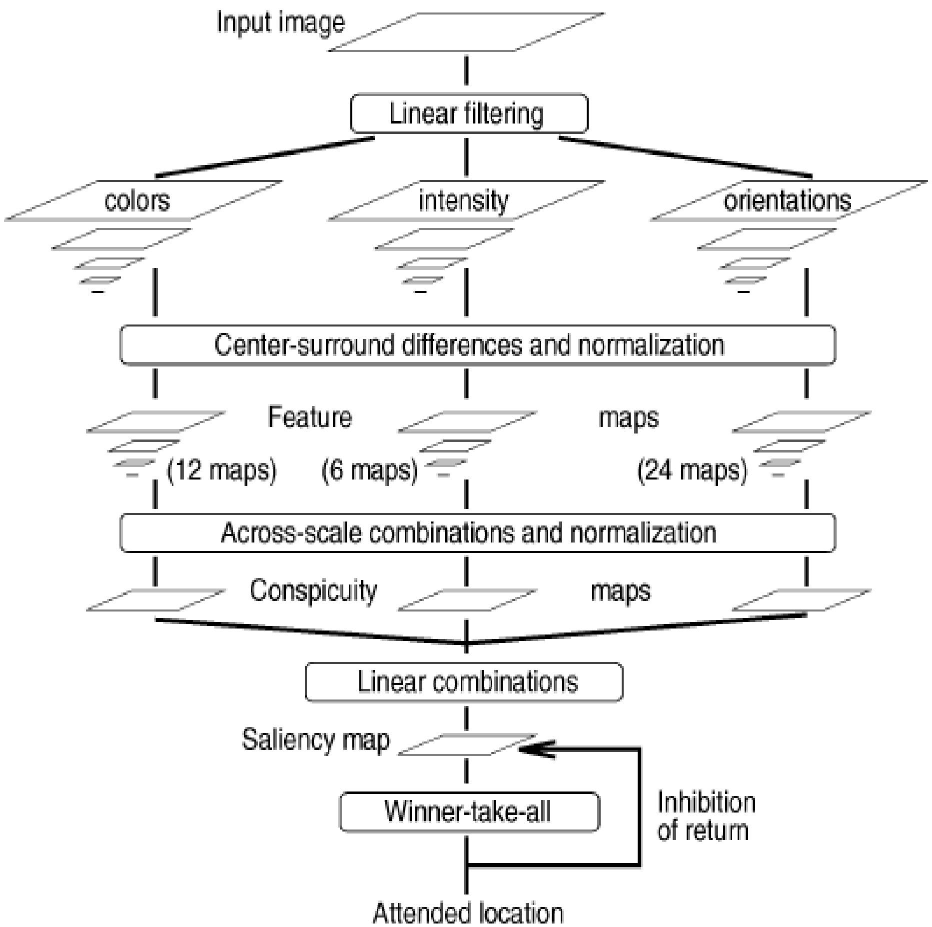
Computational Saliency Model

- **Bottom up saliency**
  - Intuition: Unusual/Salient items should draw our attention and be easy to search for.
  - Unusual targets? :A target whose features are *outliers* to the local distribution of features.
  - How do we detect outliers?
    - Explicit statistical Model, Ruth Rosenholtz et al., Torralba et al.
    - Approximate estimation with center surround filters, Itti et al.
- **Top down saliency**
  - Intuition: Searching is task oriented
  - Task priors change relevance of locations and features
  - How are the priors manifested?
    - Modulation : additive boosting
    - Gain control: multiplicative boosting/supression (Luck et. al, 97')

Outline

- **Saliency 101**
- **Bottom up**
  - Itti, Koch and Neibur, “A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, 20(11), ‘98
  - Itti and Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention”, Vision Research, ‘00
- **Top Down**
  - A. Torralba, A. Oliva, M. Castelhano and J. M. Henderson, “Contextual guidance of attention and eye movements in real-world scenes: the role of global features in object search”. Psychological Review, ’06
  - A. Torralba,“Modeling global scene factors in attention", JOSA, 20(7), 03
  - A. Torralba, "Contextual Priming for Object Detection", IJCV, 53(2), 03
- **Summary**
- **Demo**
  - Comparison of bottom up saliency models

Itti and Koch Algorithm



- **Feature maps**
  - Compute strength of individual features
- **Conspicuity maps**
  - Compute saliency of individual features through center surround
- **Saliency maps**
  - Combines saliency from different features
- **Inhibition of return**
  - Models covert attention

(Itti et al.)

Example

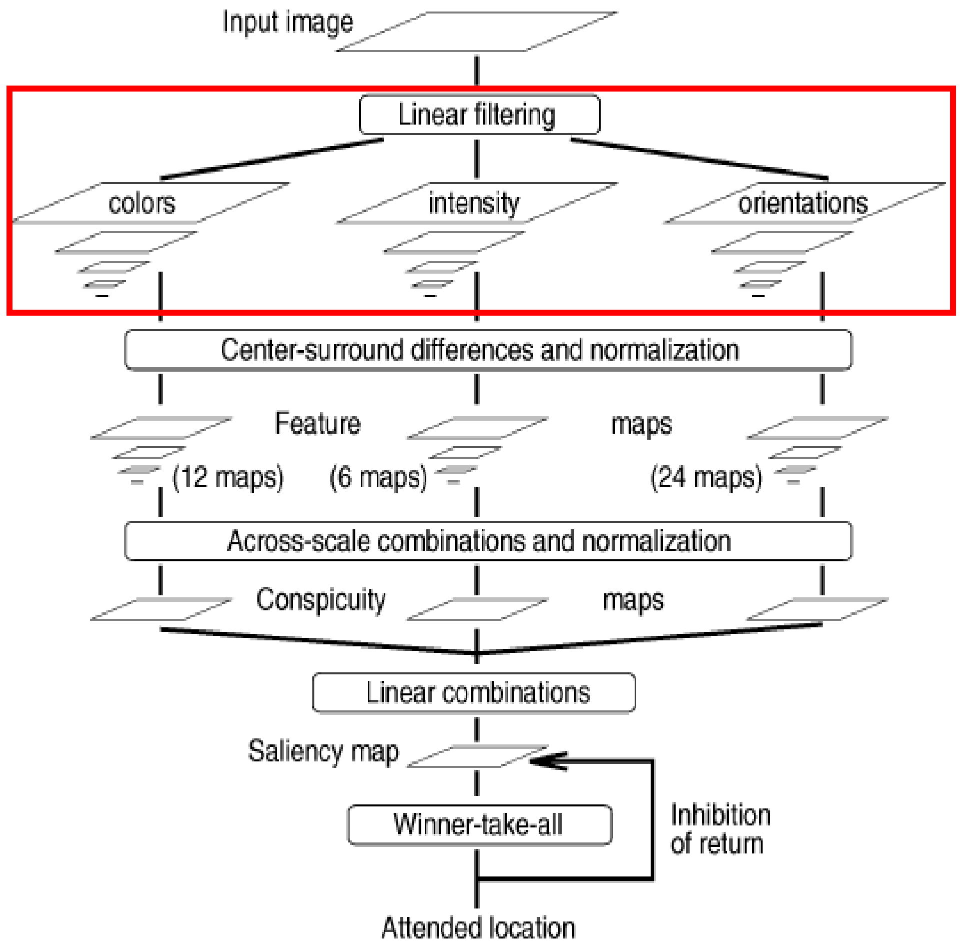
(Itti et al.)

Feature maps

$$I=\frac{r+g+b}{3}, I(\sigma)=(I)_\sigma$$
$$(I)_0=I$$
$$(I)_1=\left(I_0 * G\right)\downarrow_2, (I)_\sigma=\left(I_{\sigma-1} * G\right)\downarrow_2$$

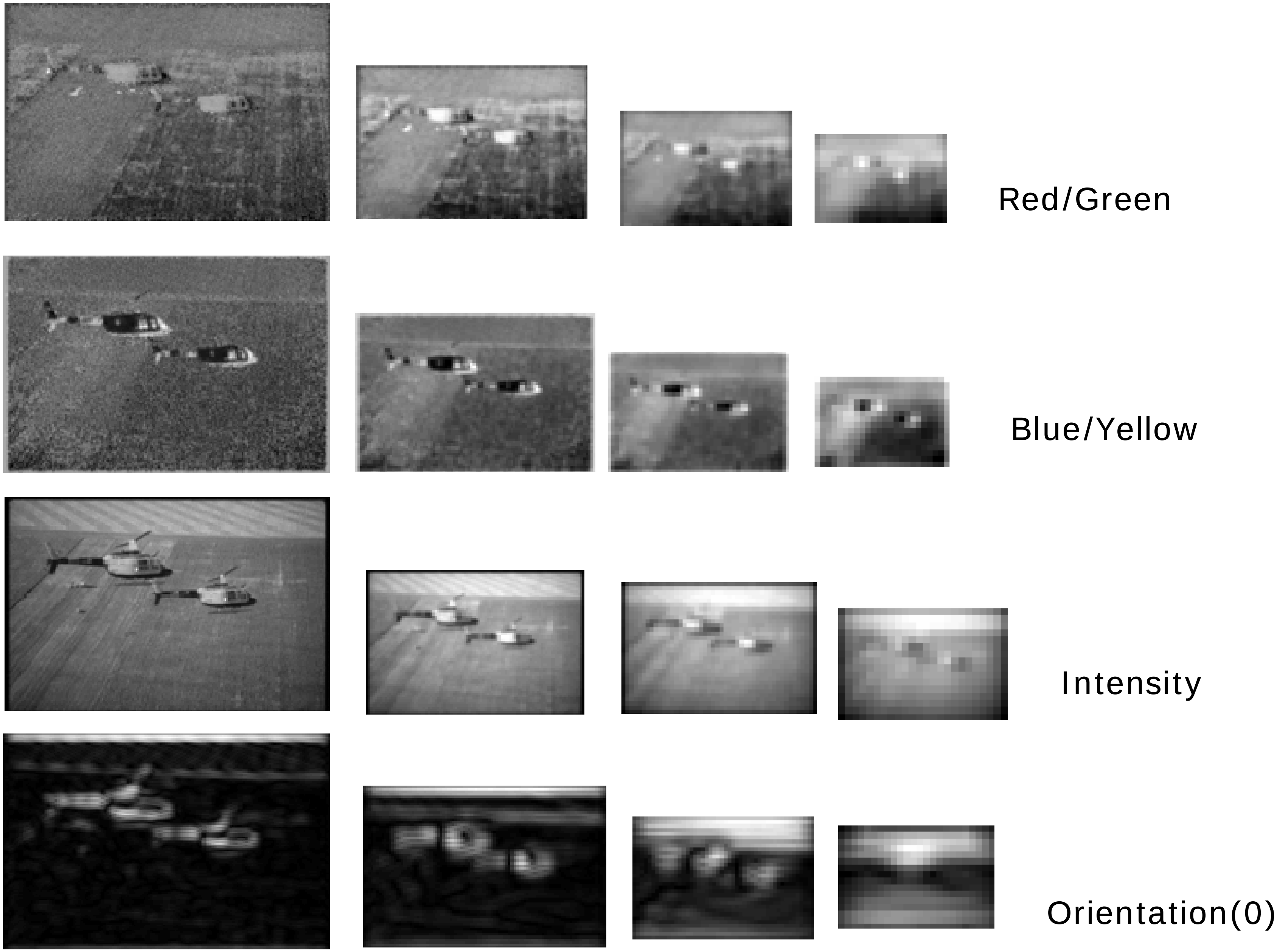
$$O(\sigma,\theta)=\left(I * Gabor(\theta)\right)_\sigma$$

$$R=r-\frac{(g+b)}{2}, G=g-\frac{(r+b)}{2}$$
$$B=b-\frac{(r+g)}{2}, Y=\frac{(r+g)}{2}-\left|\frac{r-g}{2}\right|-b$$
$$R(\sigma), G(\sigma), B(\sigma), Y(\sigma)$$



(Itti et al.)

Example Features



Center surround and normalization

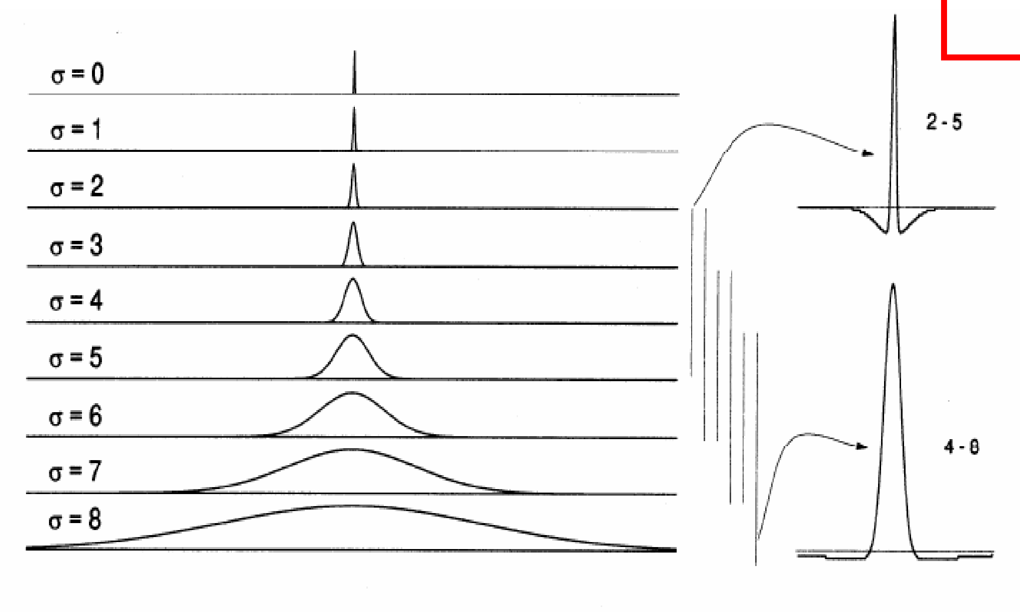
$$I(c,s) = |I(c) - I(s)|$$

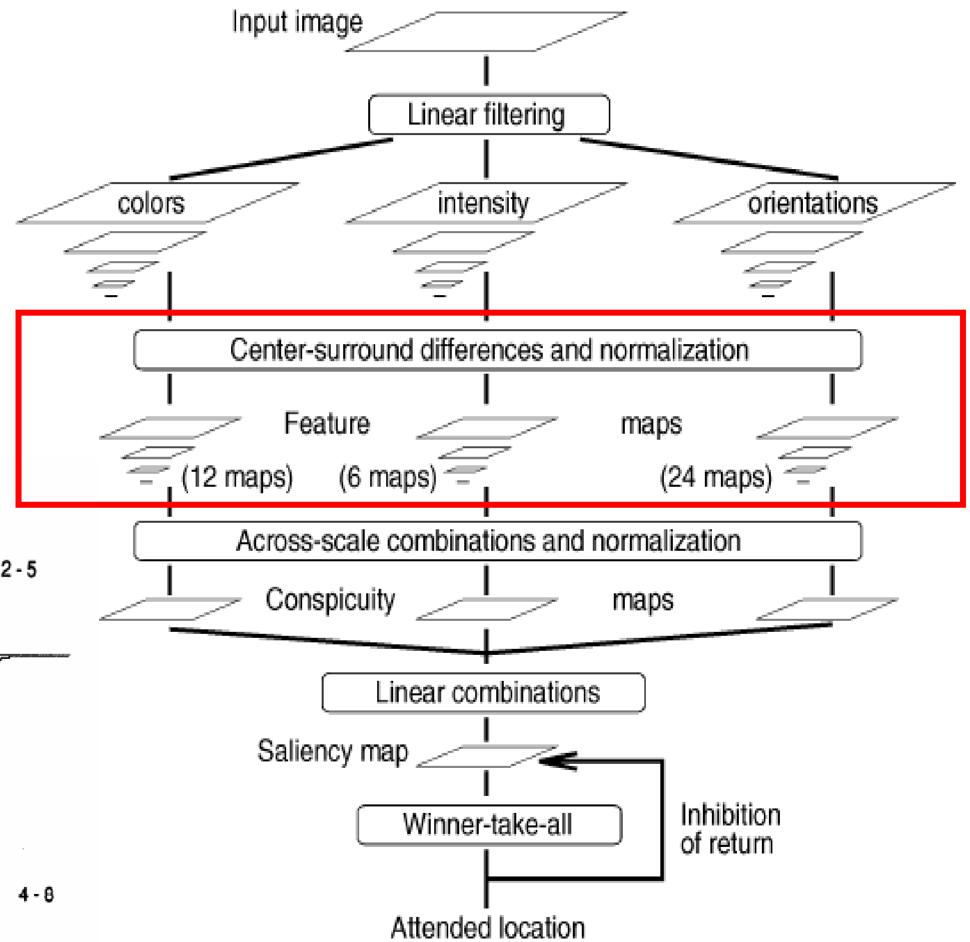
$$RG(c,s) = |(R(c) - G(c)) - (G(s) - R(s))|$$

$$BY(c,s) = |(B(c) - Y(c)) - (Y(s) - B(s))|$$

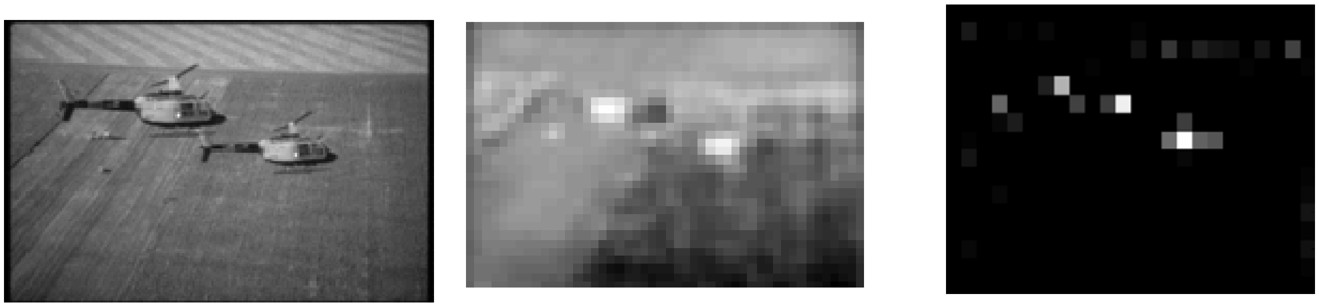
$$O(c,s,\theta) = | O(c,\theta) - O(s,\theta) |$$

$$c \in \{2,3,4\}, s = c + \delta, \delta \in \{3,4\}$$

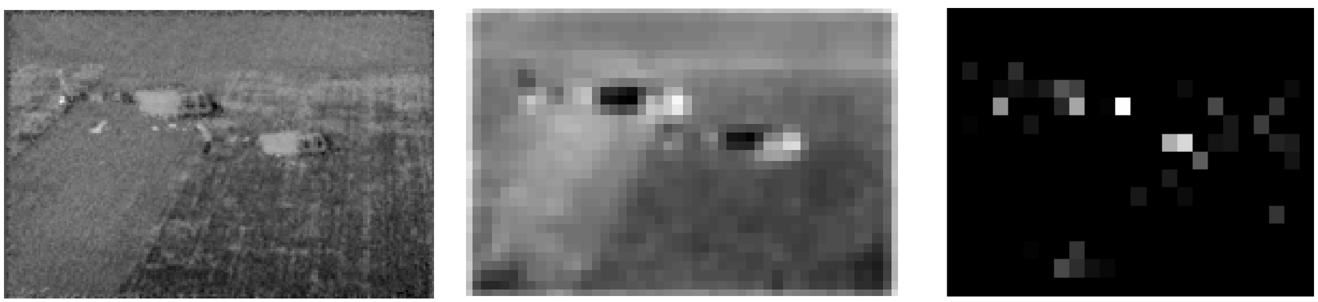




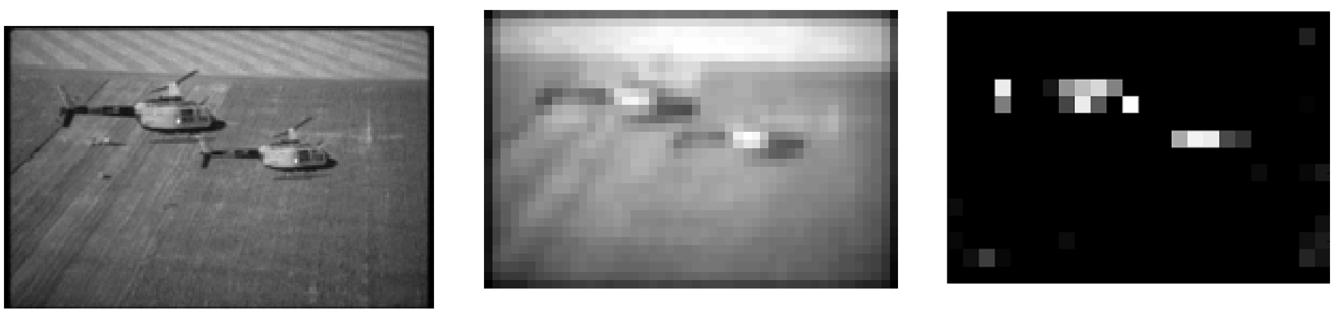
Example: Center surround



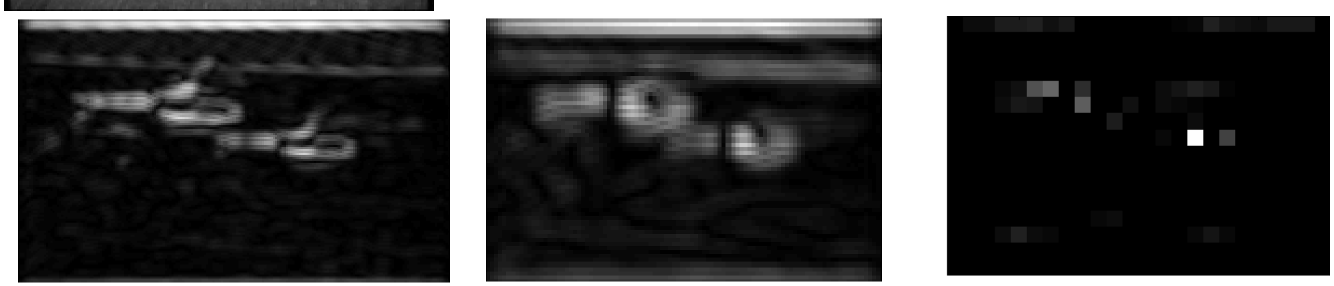
Red/Green



Blue/Yellow



Intensity



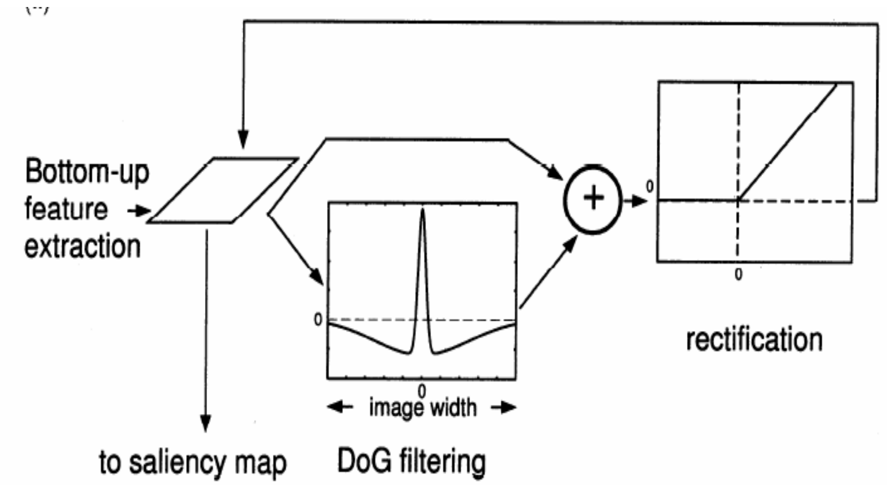
Orientation(0)

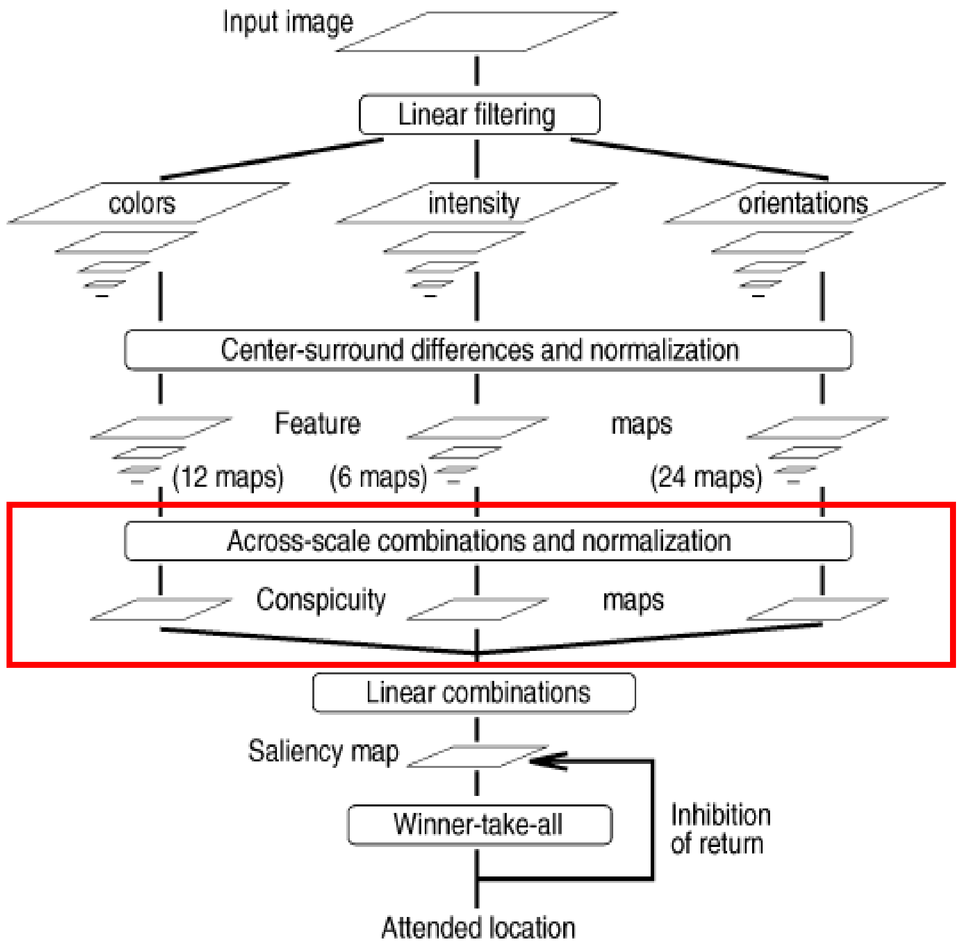
Example: Conspicuity maps

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

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

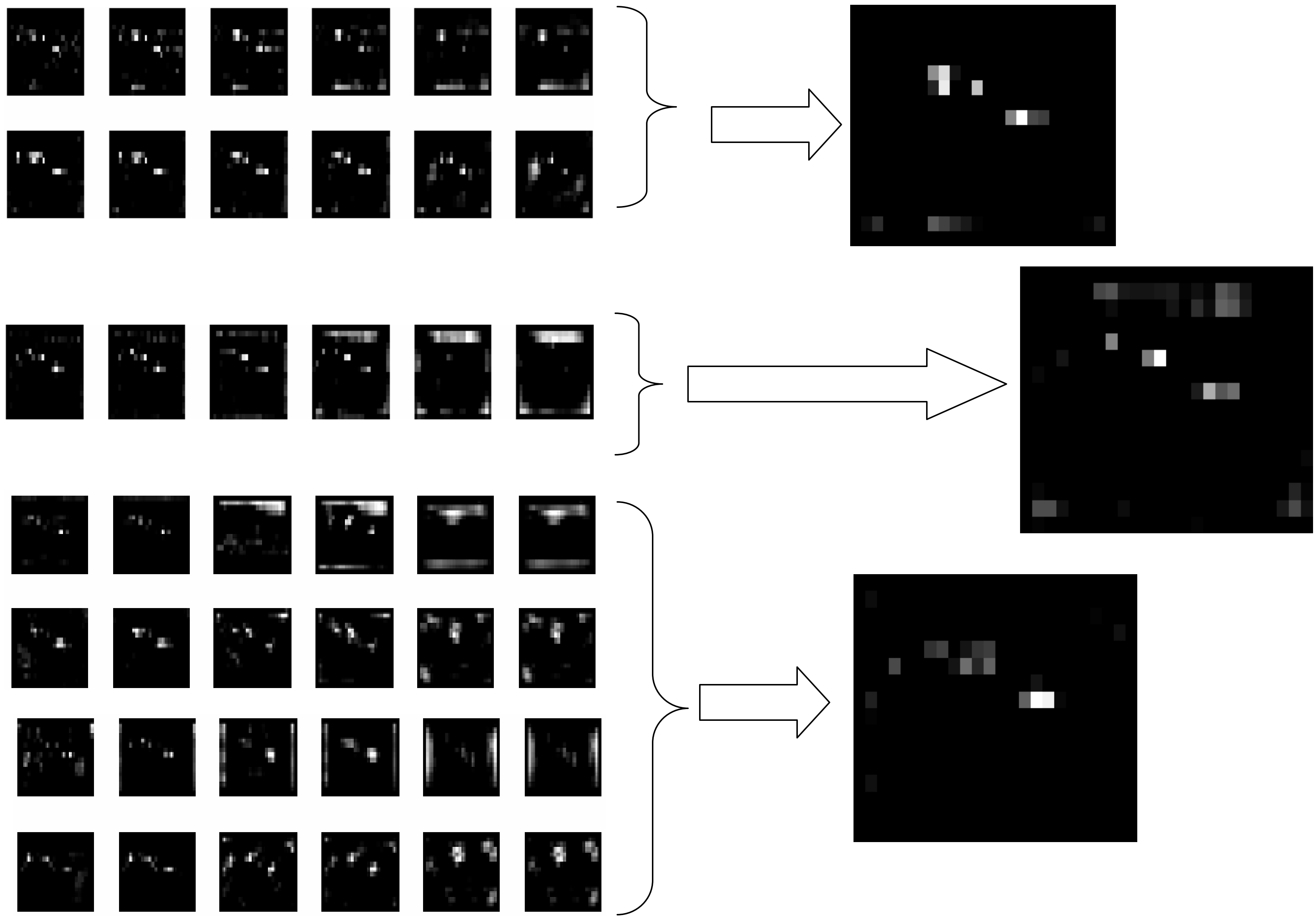
$$\overline{O} = \sum_{\theta} \mathbf{N}\left(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathbf{N}(O(c,s,\theta)) \right)$$





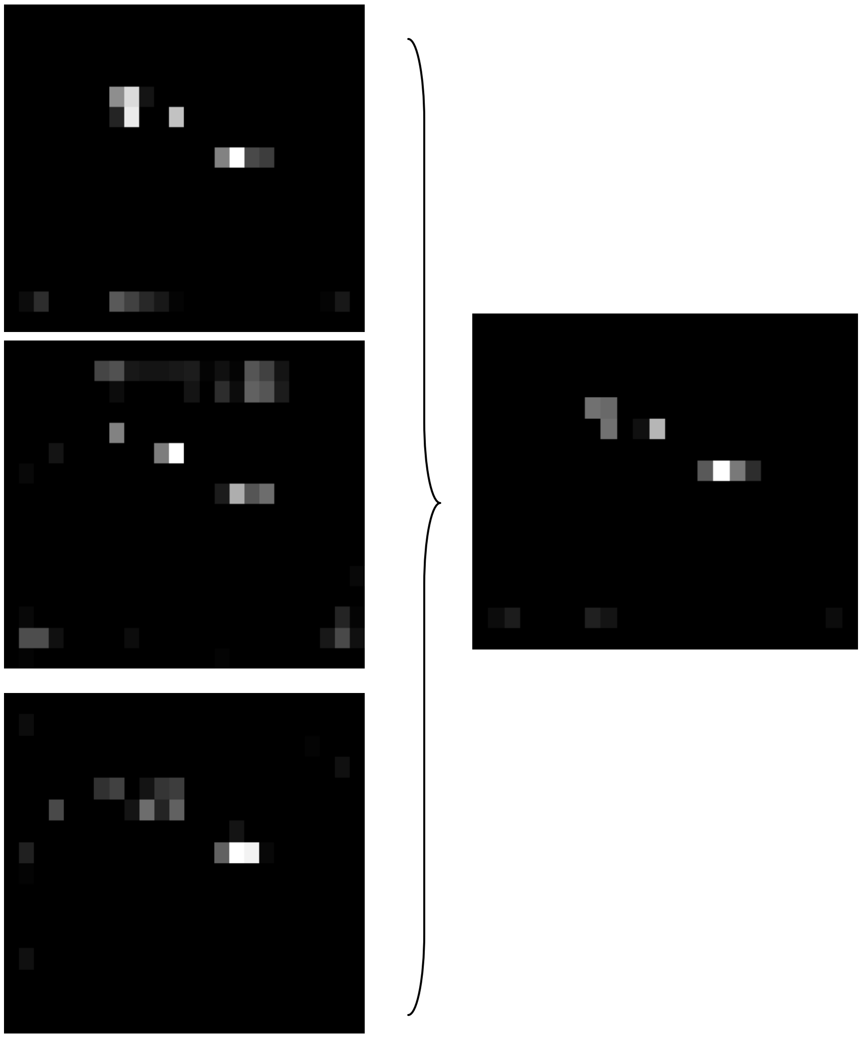
(Itti et al.)

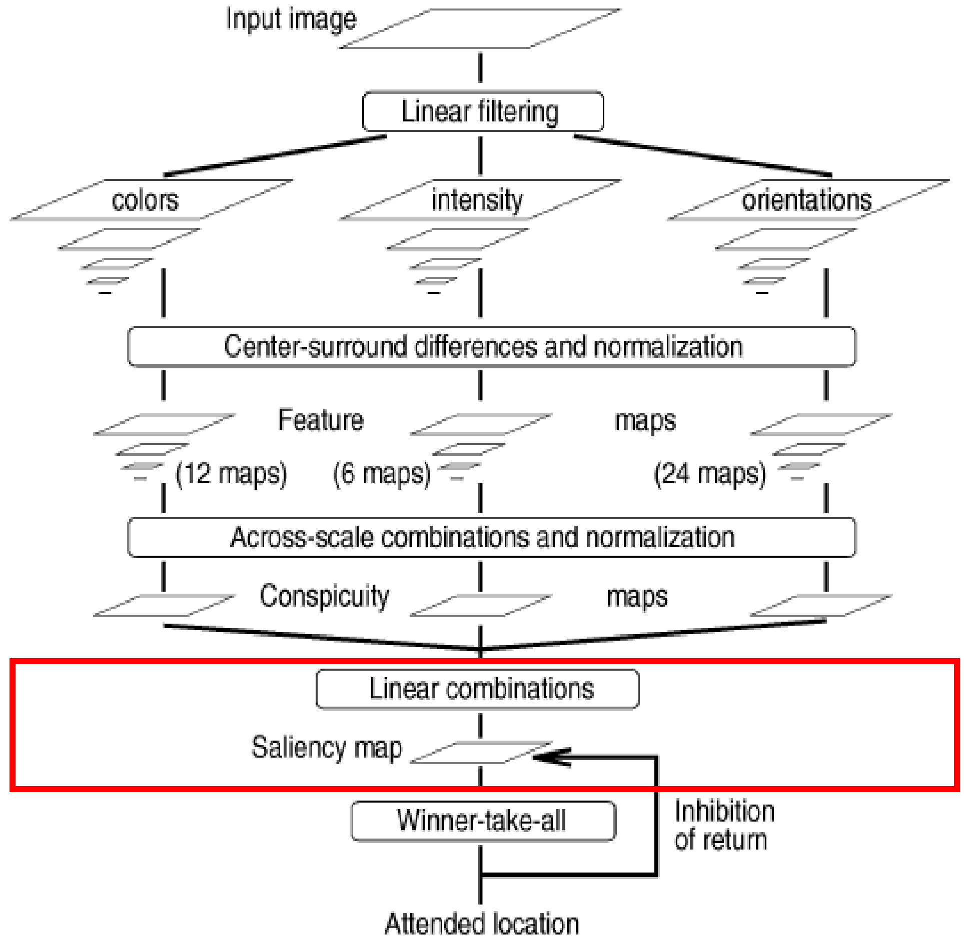
Example: Conspicuity maps



Saliency maps

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

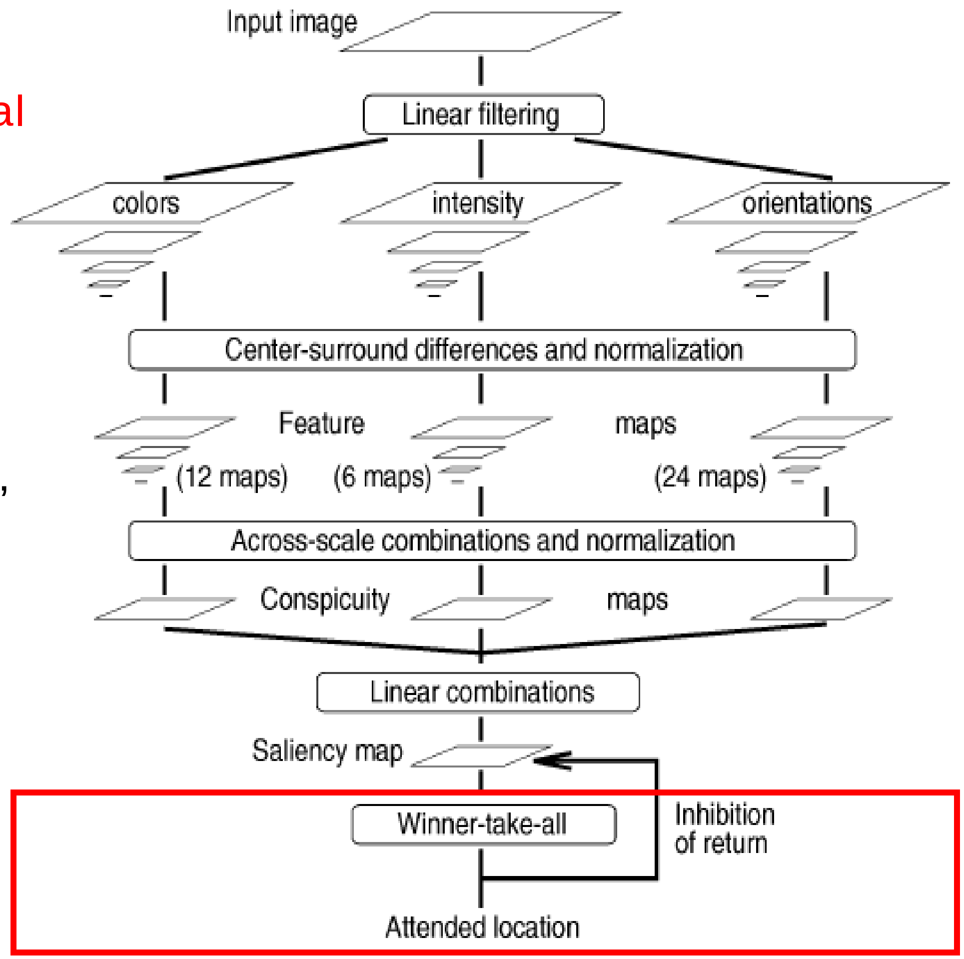




(Itti et al.)

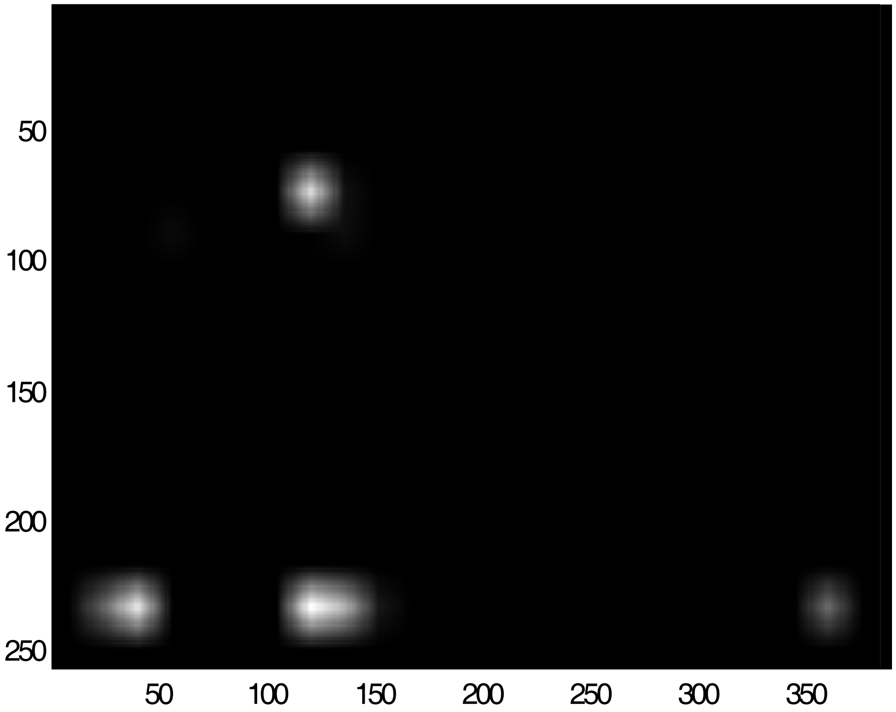
IOR: Inhibition of return

- Attention shifts are modeled using IAF neurons
- Salience map feeds into a WTA neural network
- Attention is first shifted to the most salient location
- The region is consequently suppressed, and attention is shifted to the next most salient location
- The FOA is shifted in 'simulated time' to model human attention mechanism

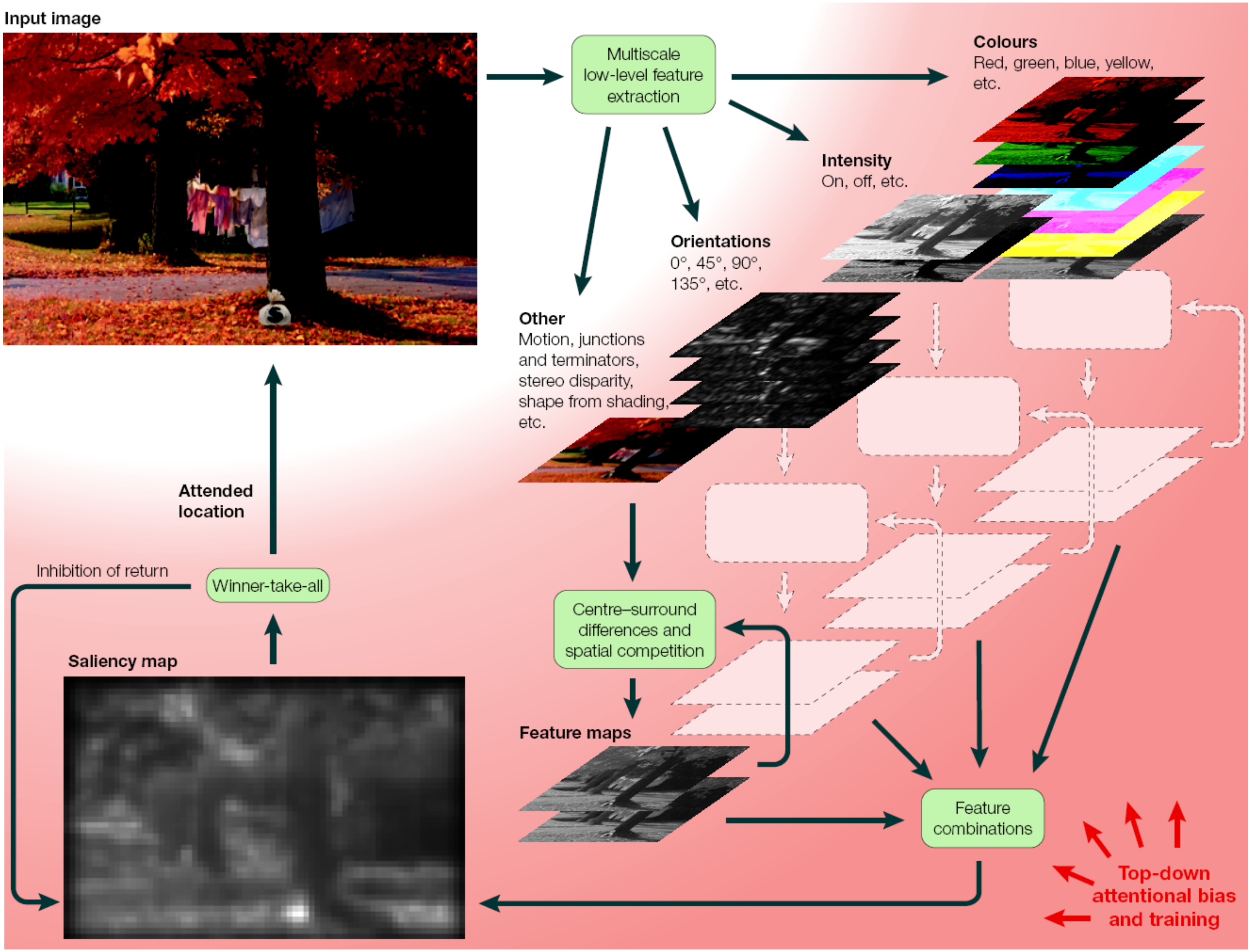


(Itti et al.)

Example: IOR



Time line: Bottom-up attention



Koch, Ullman 1985

Itti, Koch, Niebur 1998

Itti and Koch 2001

V. Navalpakkam and Itti, 2005

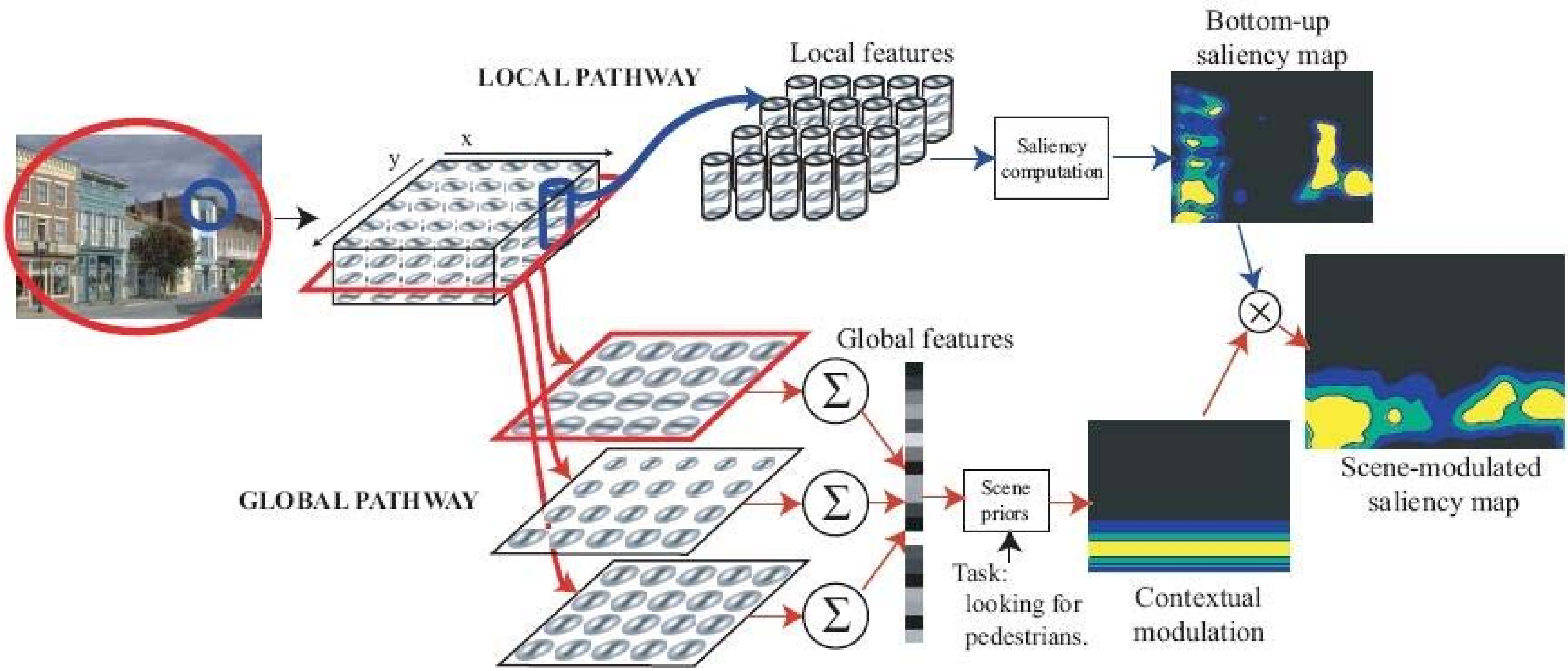
D.Walter And Koch,2006

Courtesy, D.Walter

Outline

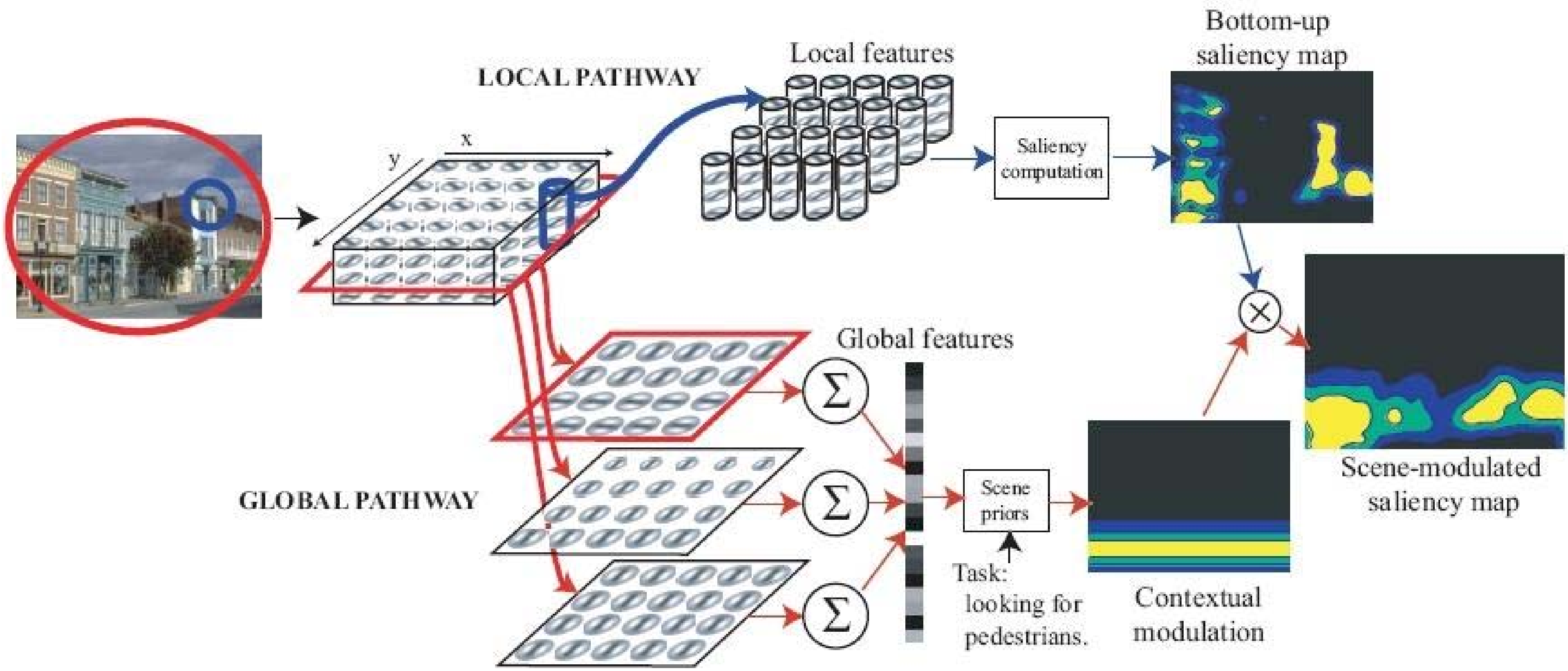
- **Saliency 101**
- **Bottom up**
  - Itti, Koch and Neibur, “A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, 20(11), ‘98
  - Itti and Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention”, Vision Research, ‘00
- **Top Down**
  - A. Torralba, A. Oliva, M. Castelhano and J. M. Henderson, “Contextual guidance of attention and eye movements in real-world scenes: the role of global features in object search”. Psychological Review, ’06
  - A. Torralba,“Modeling global scene factors in attention", JOSA, 20(7), 03
  - A. Torralba, "Contextual Priming for Object Detection", IJCV, 53(2), 03
- **Summary**
- **Demo**
  - Comparison of bottom up saliency models

Contextual Guidance model



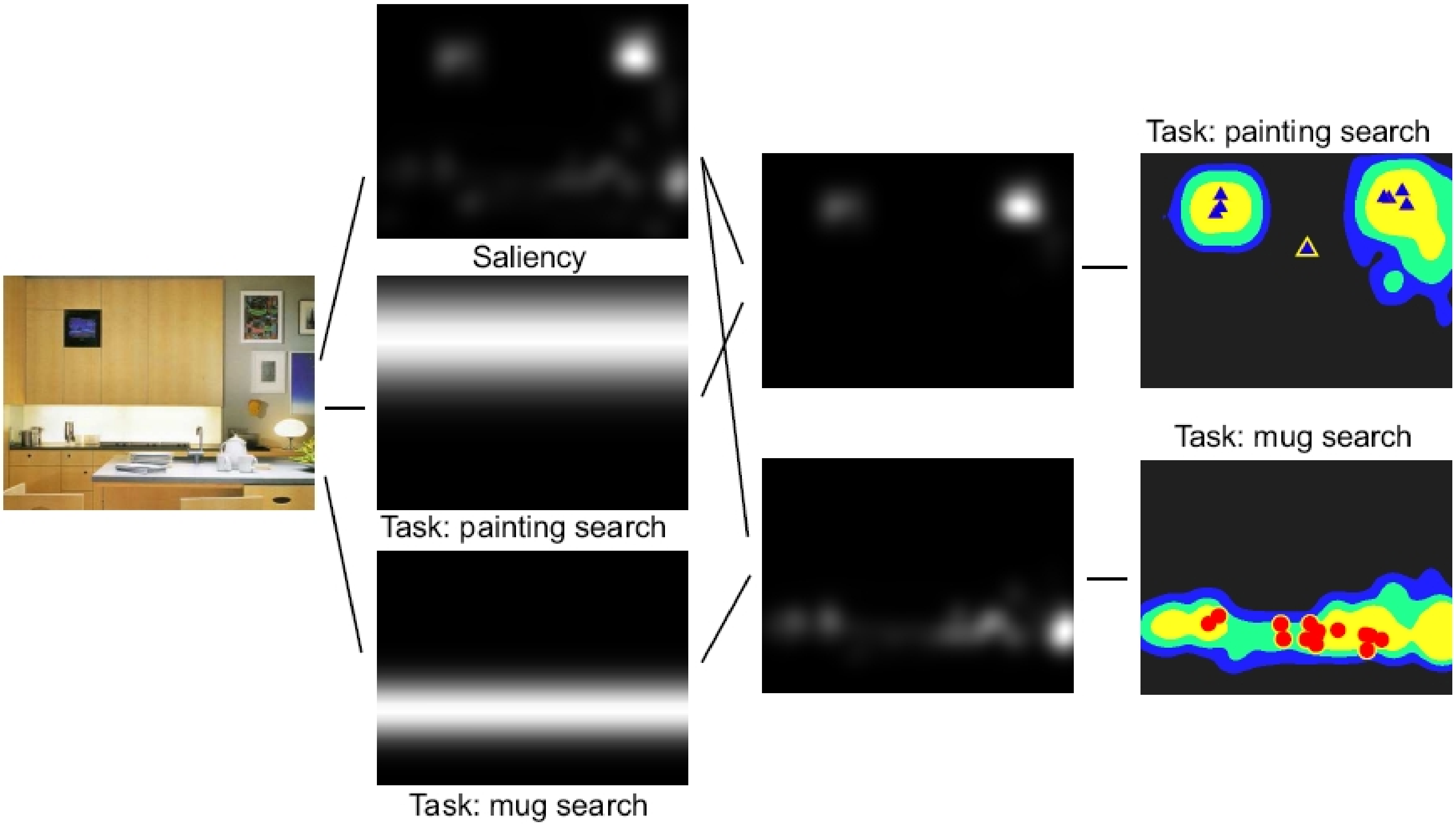
- Saliency and global-context features computed in parallel, feed-forward manner
- Search task exerts top-down control

Contextual Guidance model

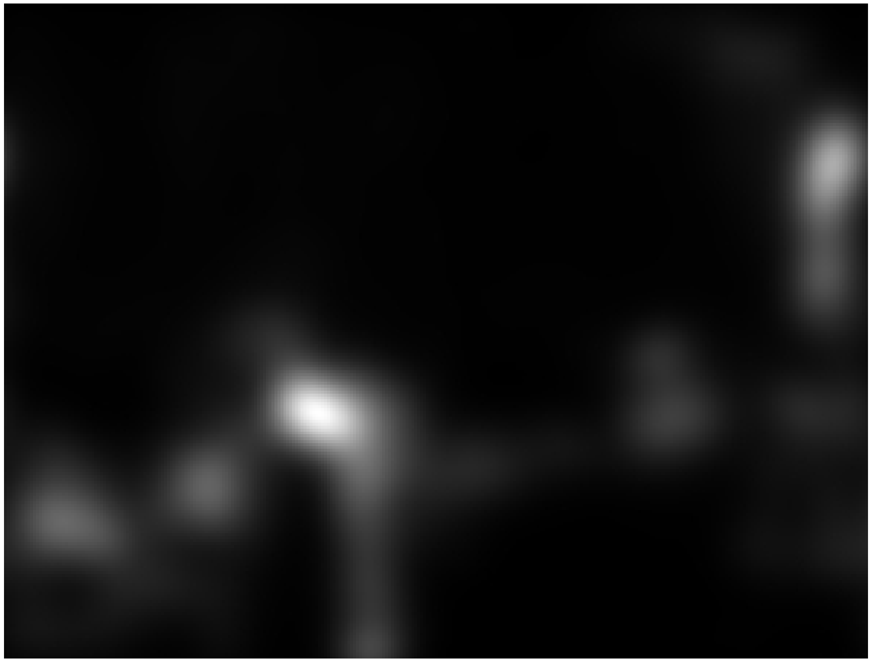


- Saliency map modulated by contextual information
- Probability of target presence by integration of task constraints and global and local image information

Contextual Guidance model



Statistical approach to saliency



- Statistically distinguishable from background
- Locations differing from neighboring regions more informative
- Rare image features more likely to be objects

Statistical approach to saliency

- Each color channel passed through bank of multiscale oriented filters (e.g. Steerable pyramid) to extract local features
- Model distribution of features using multivariate power-exponential distribution

- Normalization constant, k
- Exponent α
- Mean η
- Covariance Δ

$$\log p(L) = \log k - \frac{1}{2}\left[(L-\eta)^t \Delta^{-1} (L-\eta)\right]^\alpha$$

Statistical approach to saliency

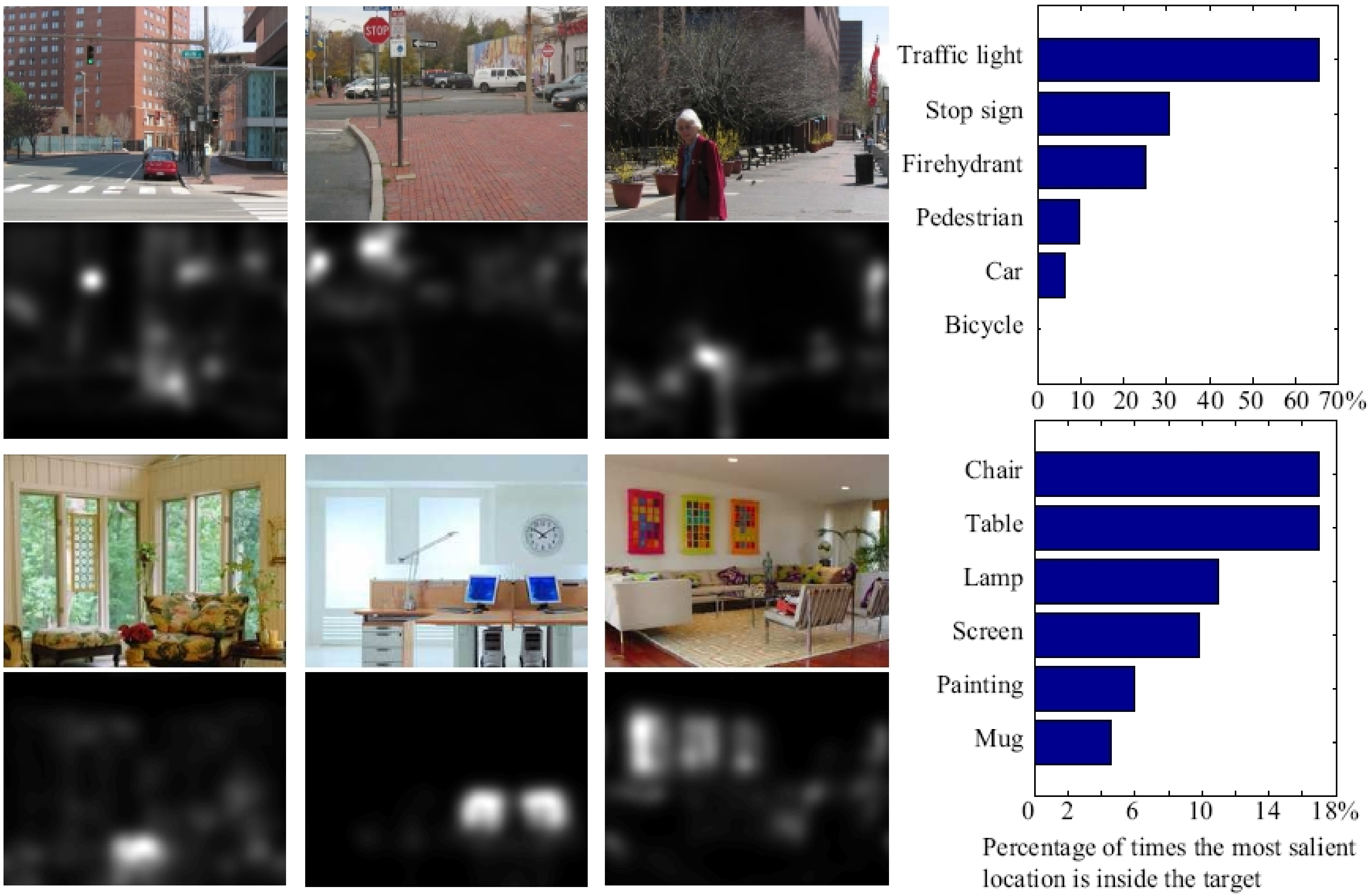
- Each color channel passed through bank of multiscale oriented filters (e.g. Steerable pyramid) to extract local features
- Model distribution of features using multivariate power-exponential distribution

- Normalization constant, **k**
- Exponent **α**
- Mean **η**
- Covariance **Δ**

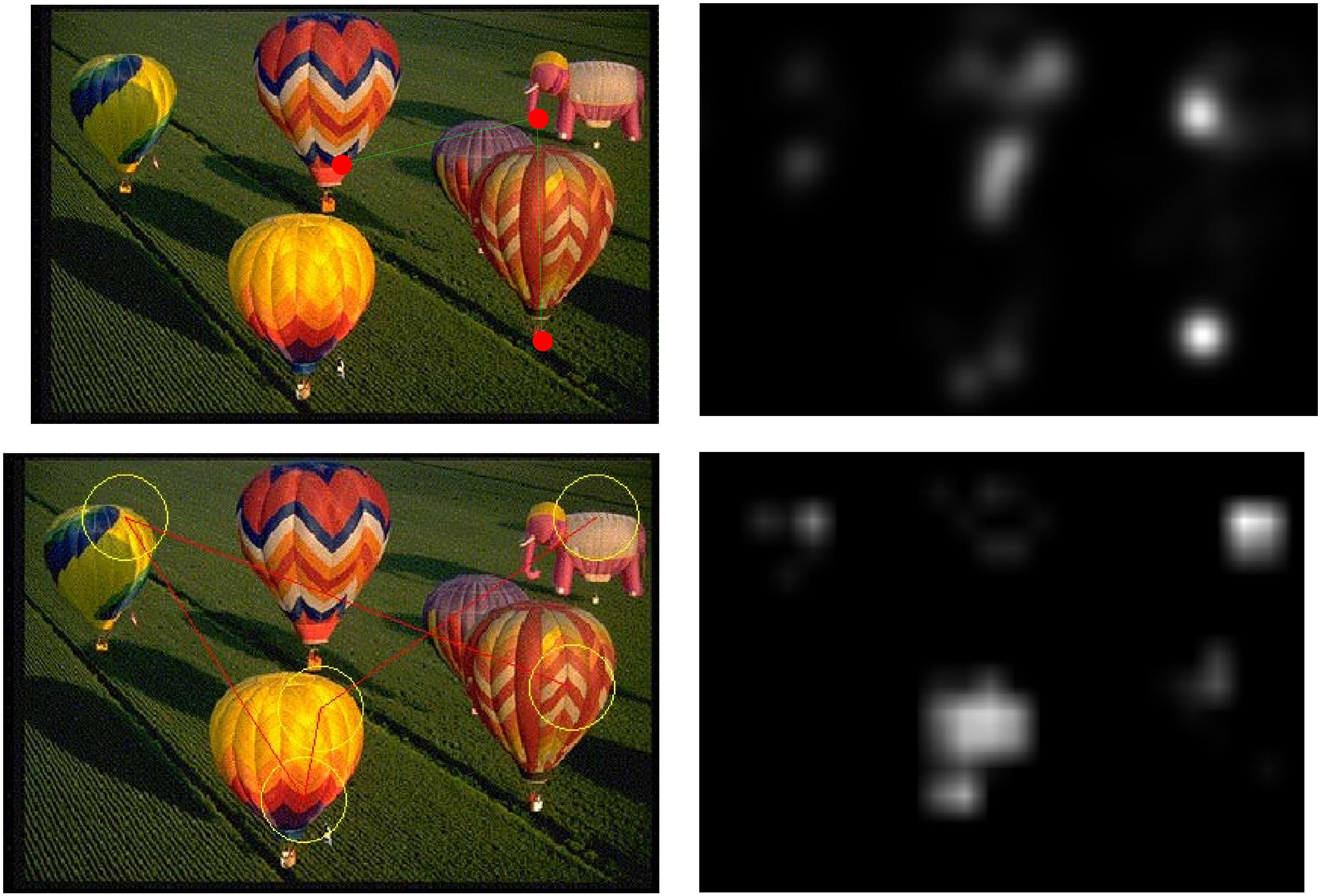
$$\log p(L) = \boxed{\log k - \frac{1}{2}\left[(L-\eta)^t \Delta^{-1} (L-\eta)\right]^{\alpha}}$$

$$\boxed{\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$ Gaussian

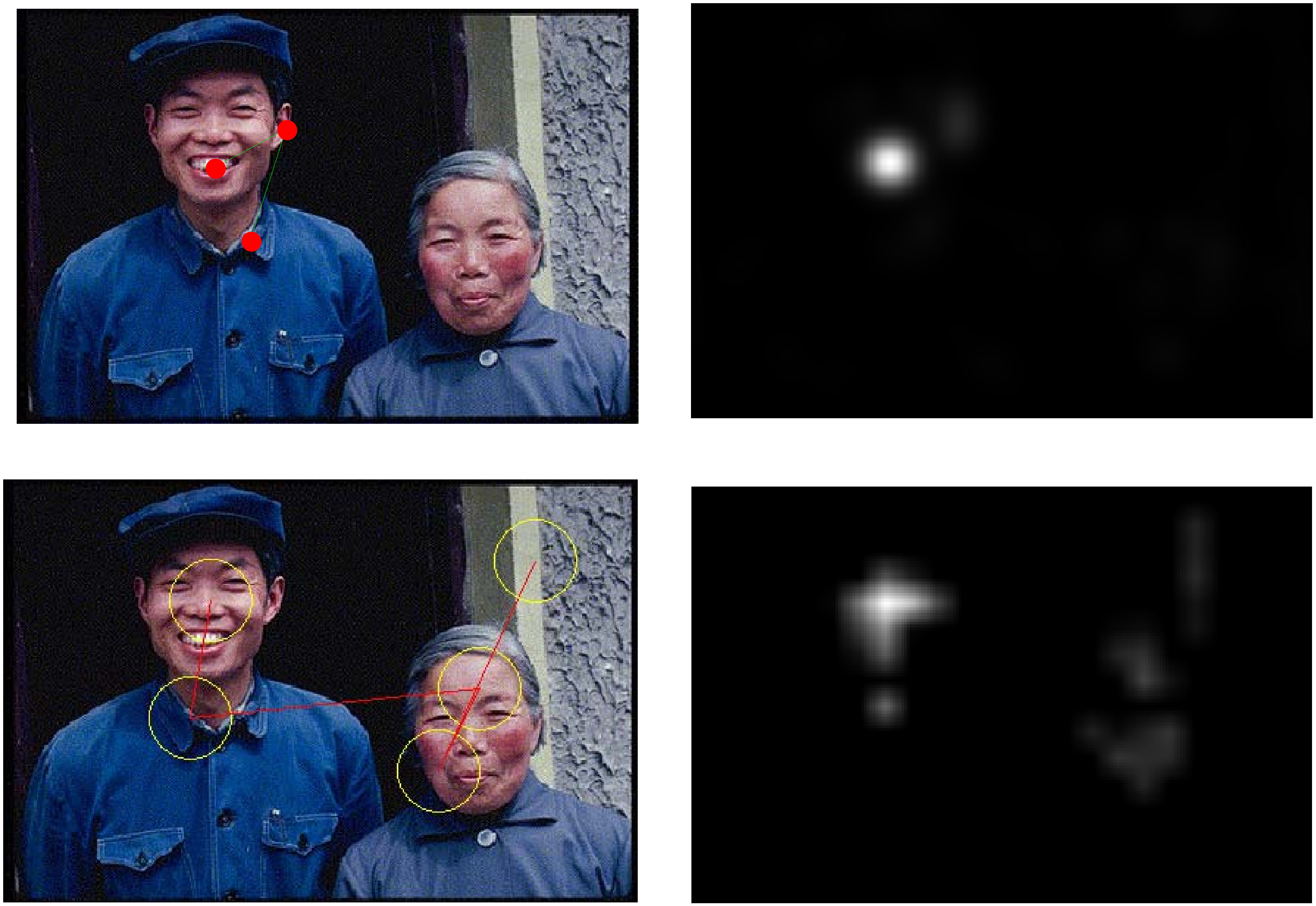
Statistical approach to saliency



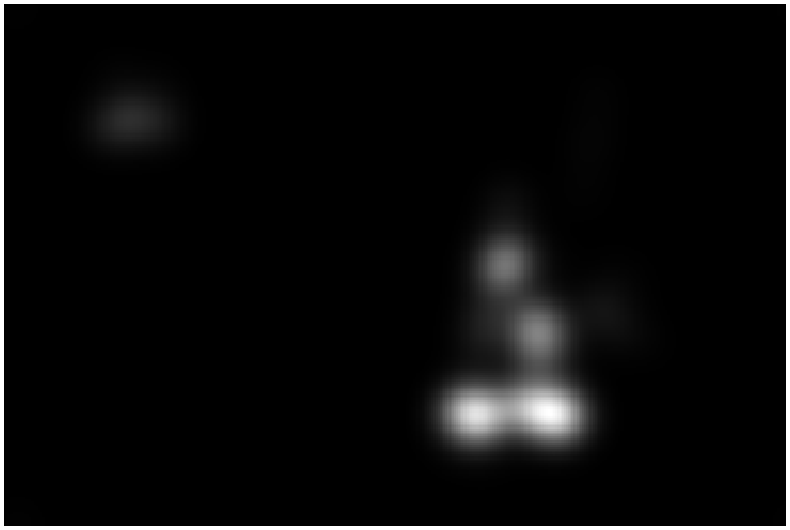
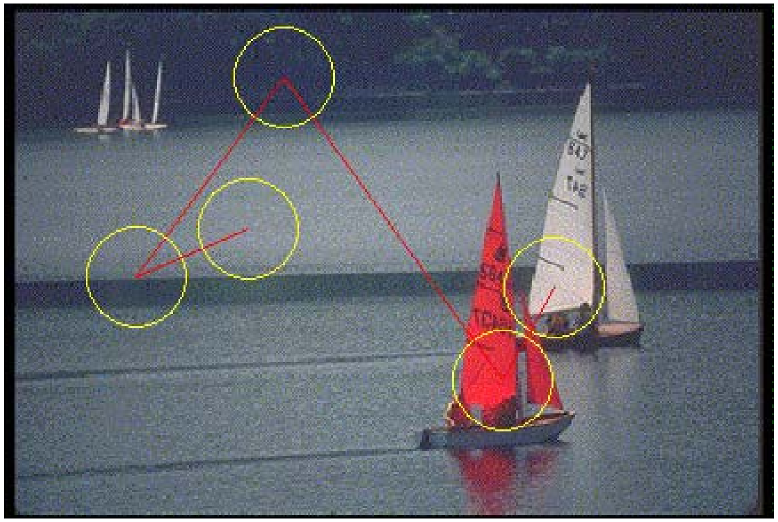
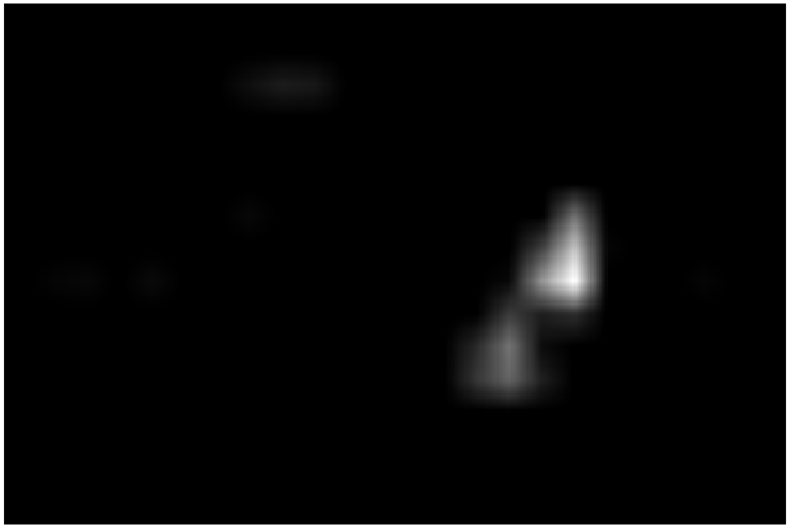
Comparison



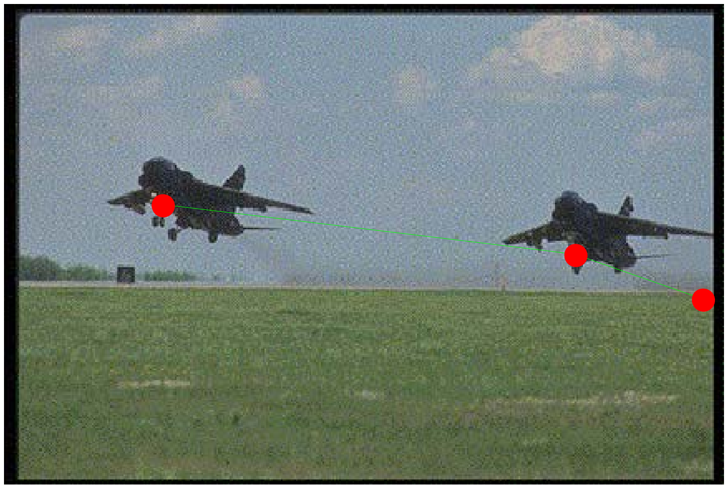
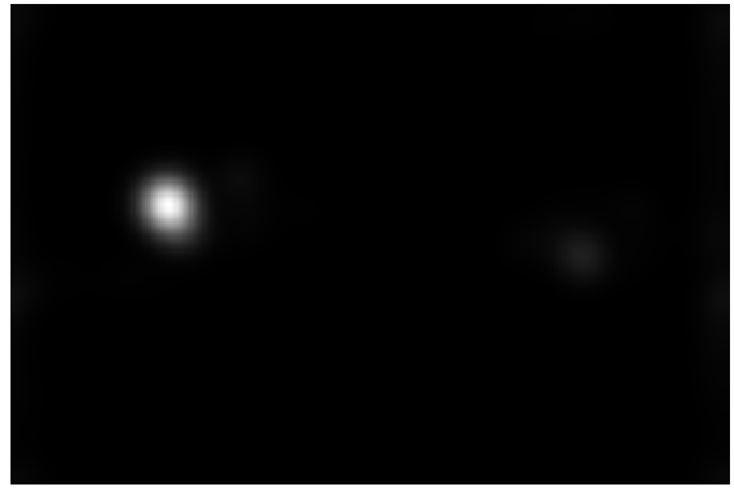
Comparison

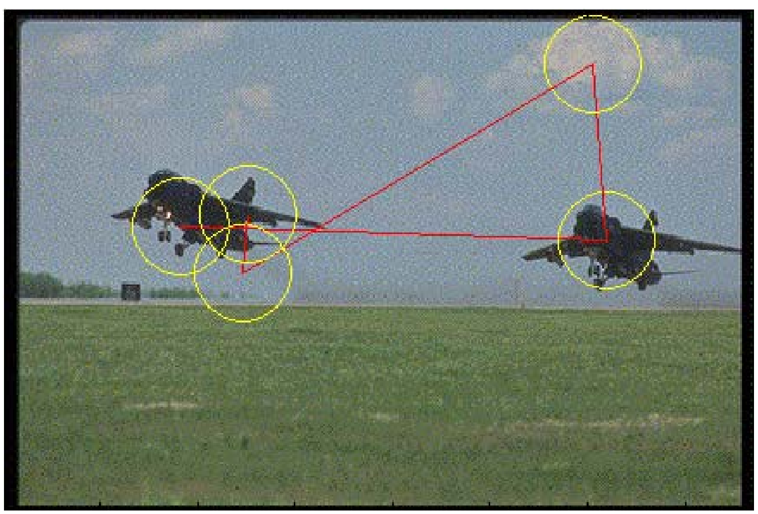
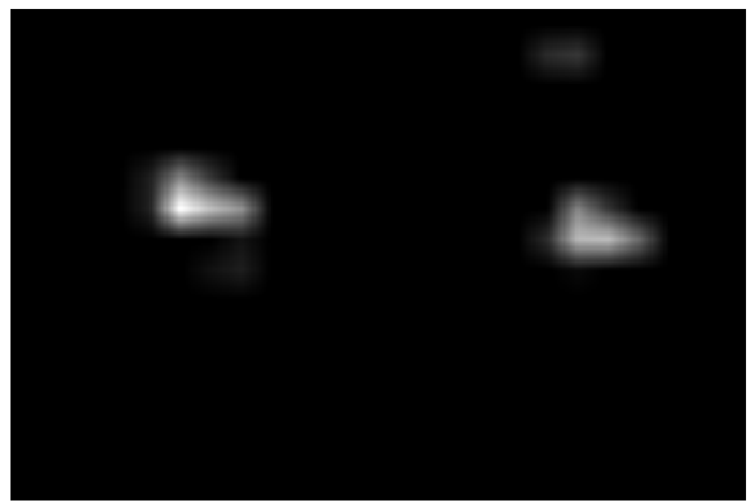


Comparison

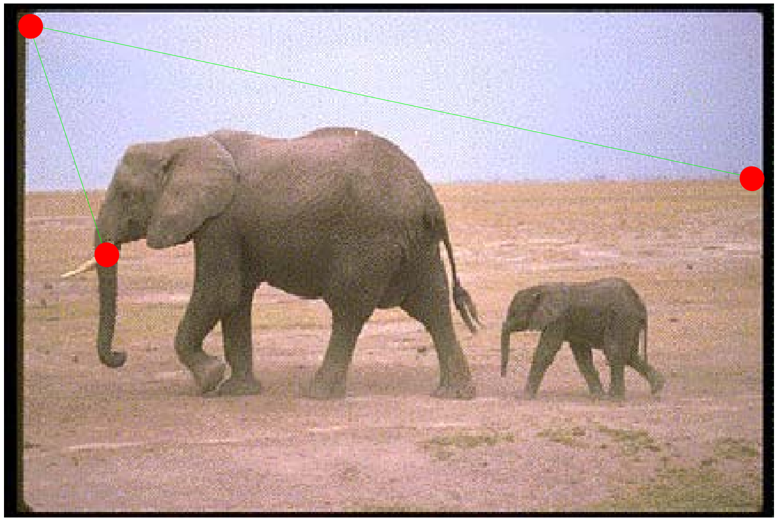





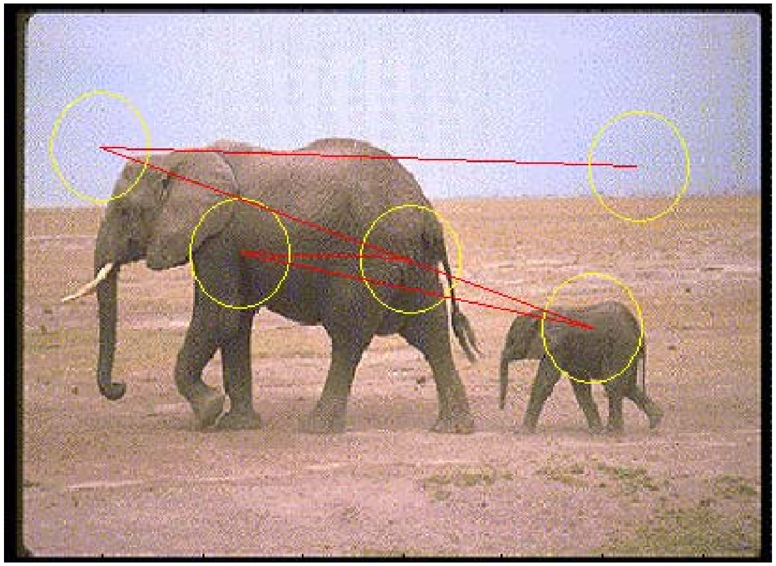
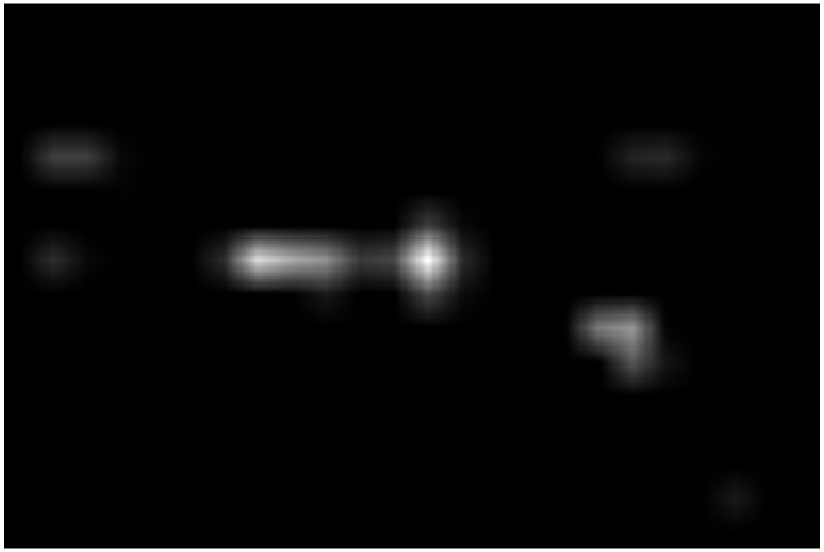
Comparison

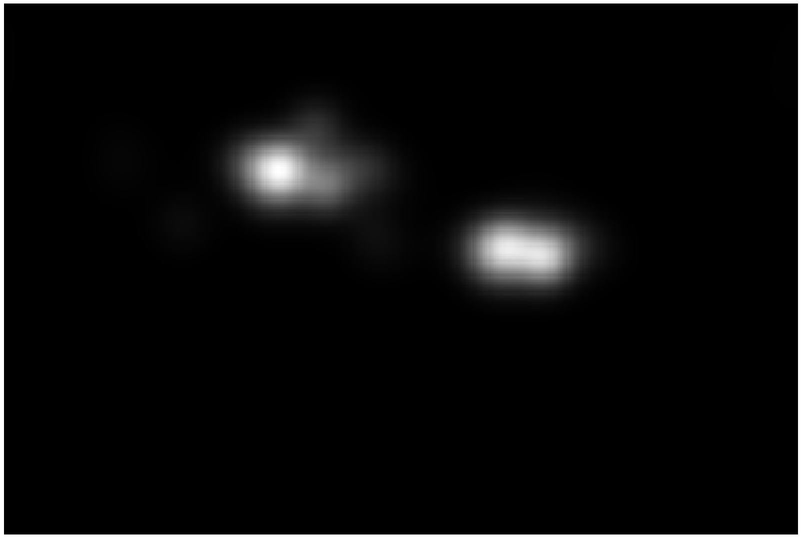



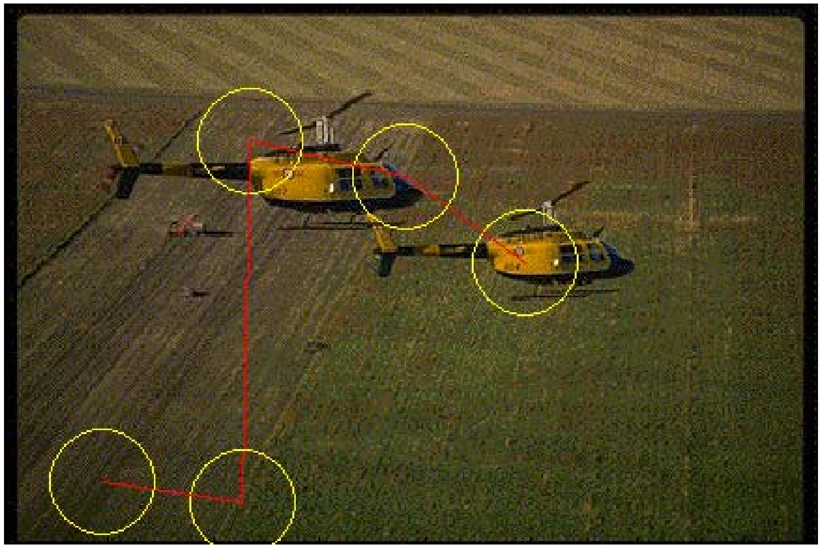
Comparison

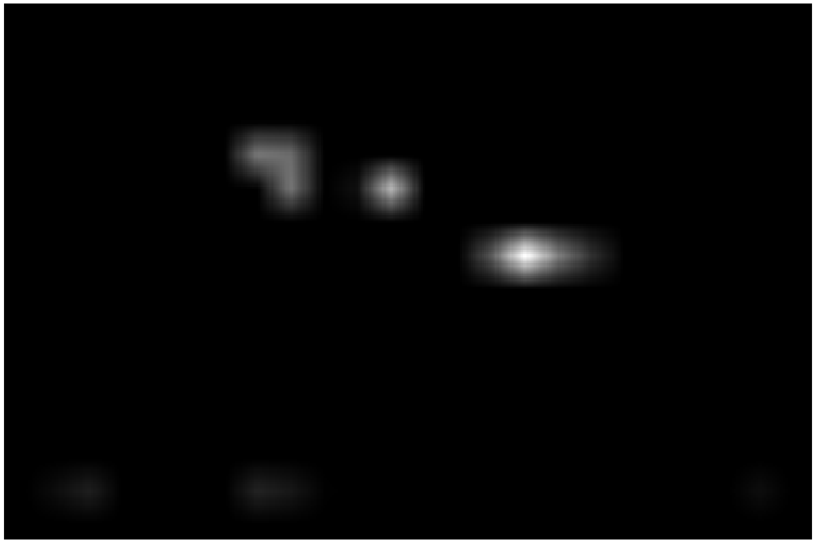


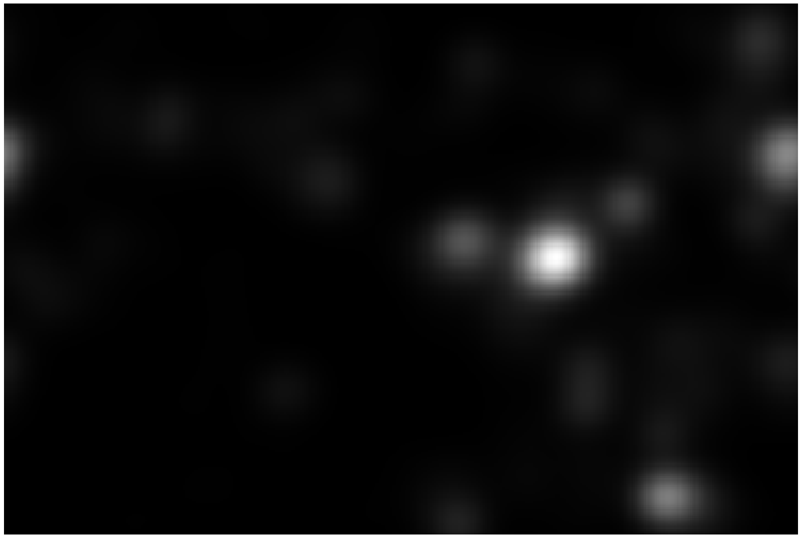
Comparison



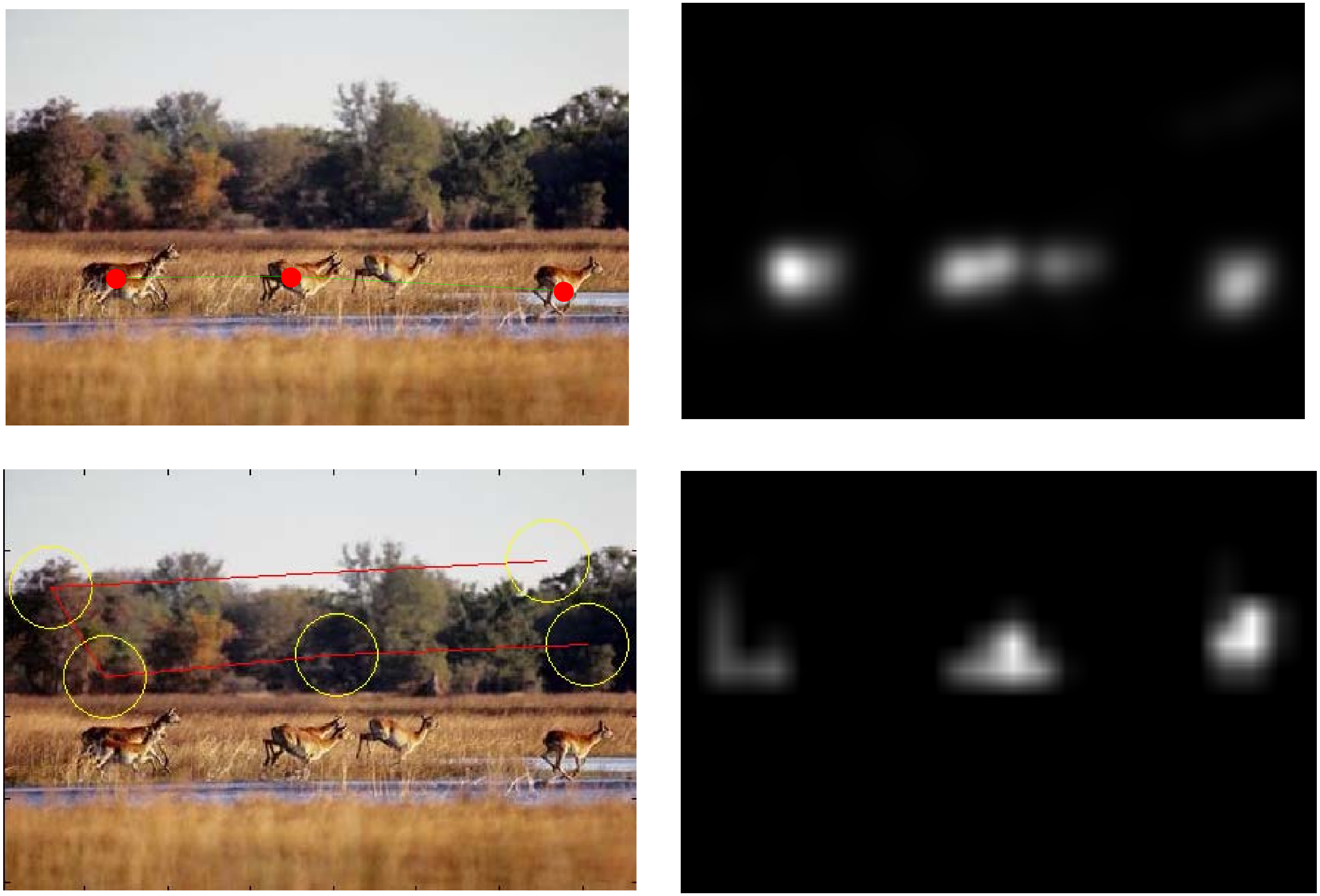




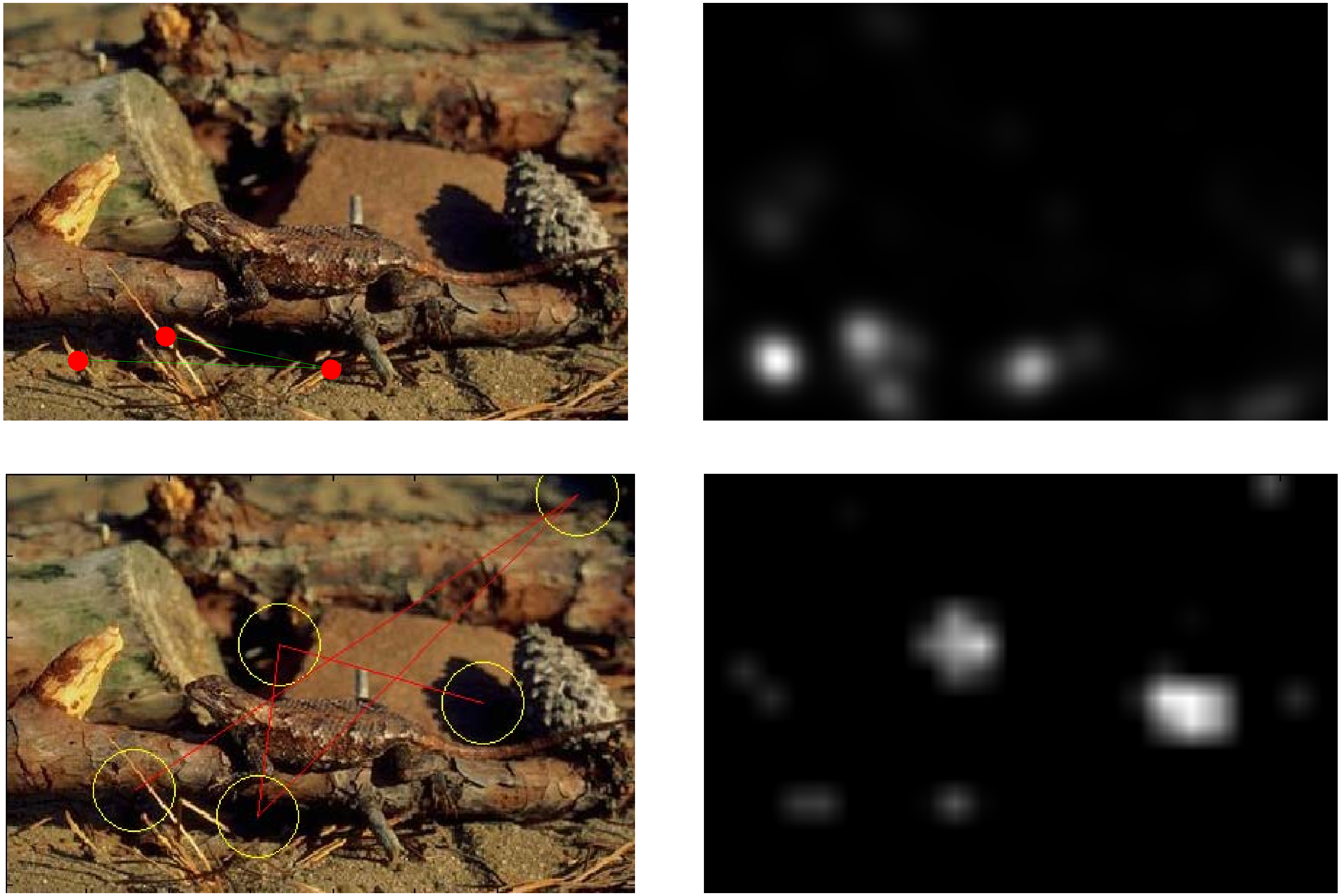
Comparison



Comparison

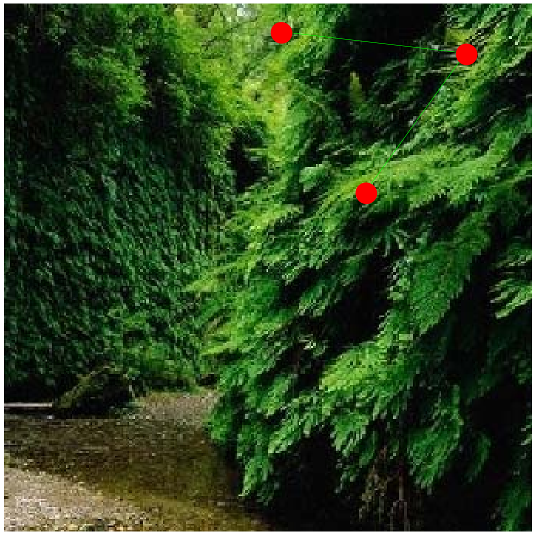
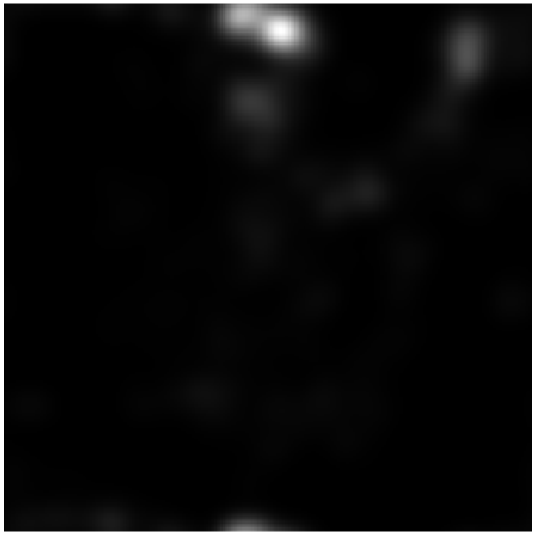


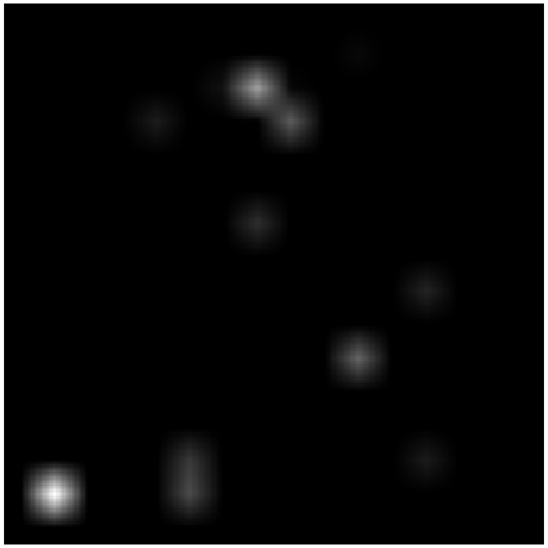
Comparison



Comparison

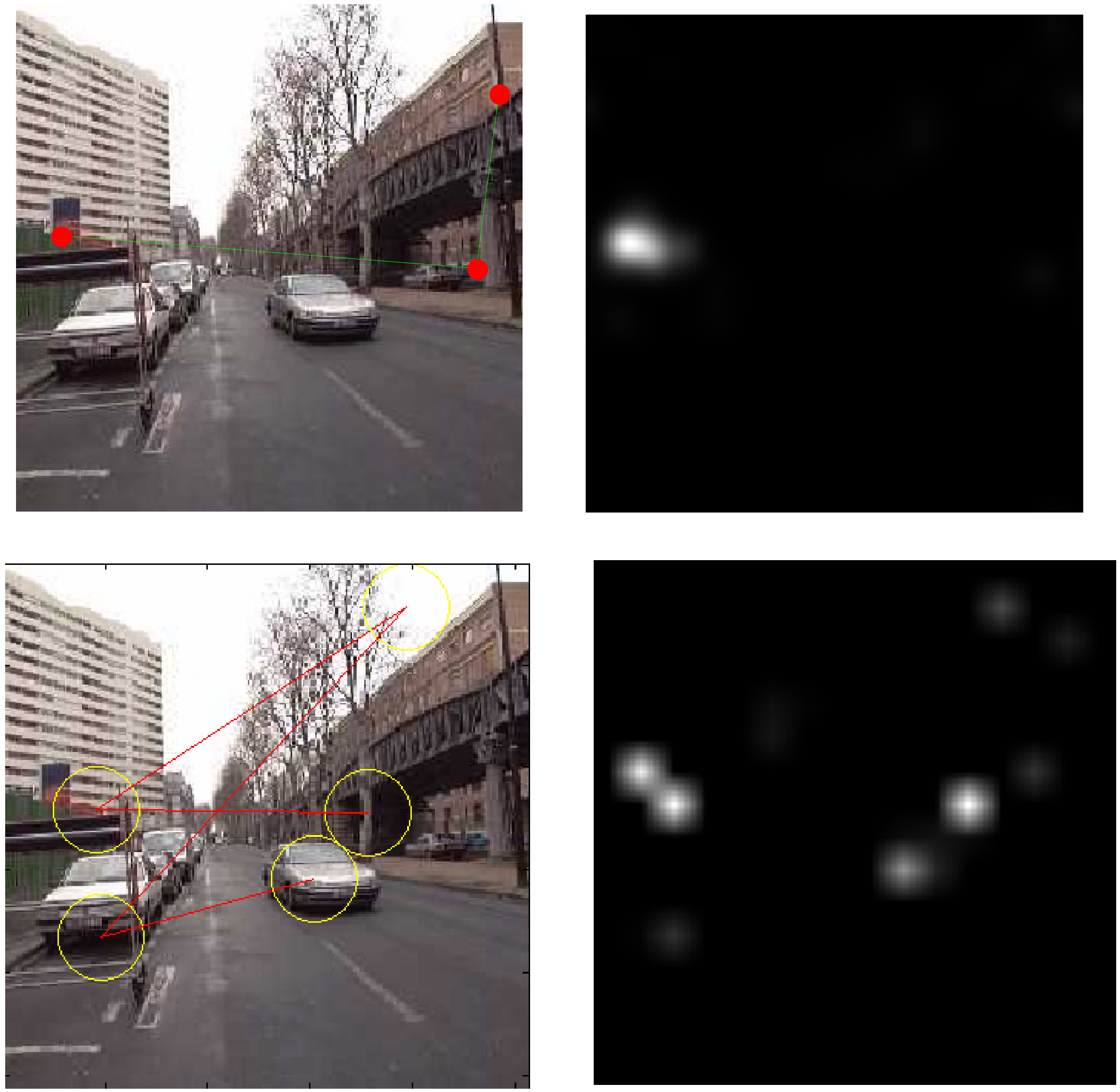
Comparison



Comparison

Comparison



Comparison

Summary

- **Saliency is the underlying mechanism that drives attention**
- **Saliency: bottom-up or top-down**
- **Bottom up: feature driven**
- **Top down: Task driven**
- **Computing bottom-up**
  - Detects outliers in feature space
  - Itti et al. algorithm- uses center surround filters
  - Torralba et al. – explicitly model statistics of features
  - Rosenholtz – gaussian modeling of features
- **Comparison**

Thank You

Biological Plausibility

- **Pop-out**
  - Search time/False positives is independent of the number of distractors
- **Search**
  - Search time increases linearly with the number of distractors
- **Performs better than humans!**