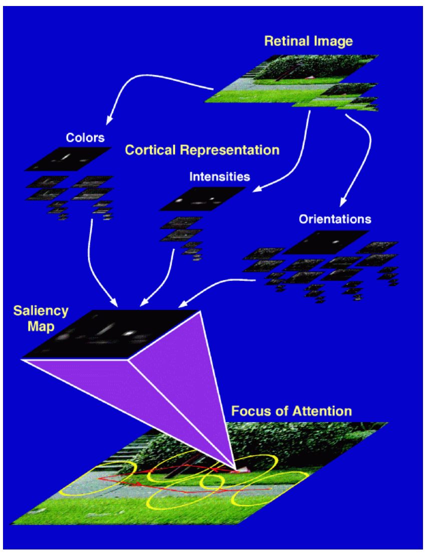
Attention

The Saliency-Map Model

Suggested reading:

- Itti, L., Koch, C., Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20:1254-1259.
- Itti, L., Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res., 40:1489-1506.

Attention: The Saliency Map Model

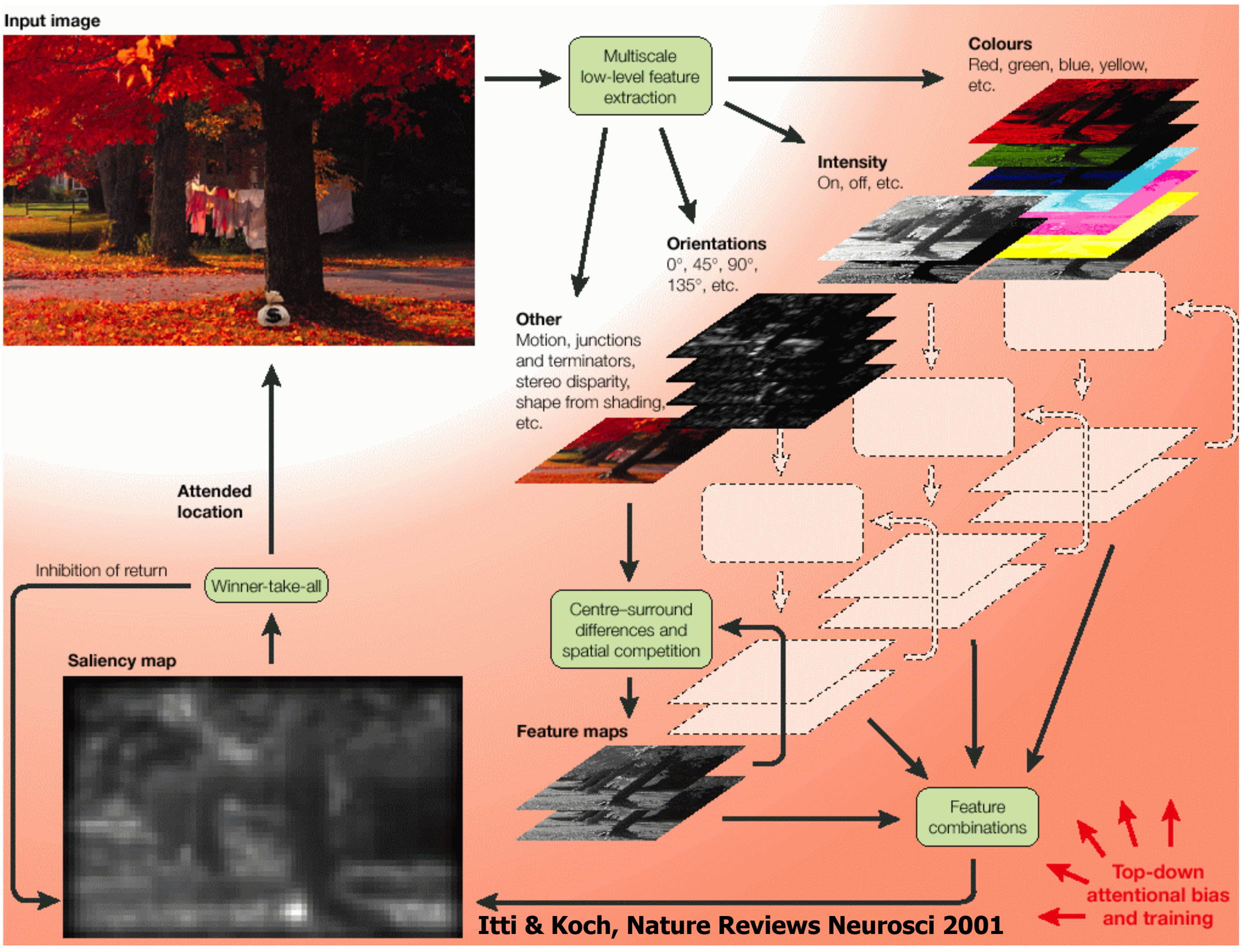


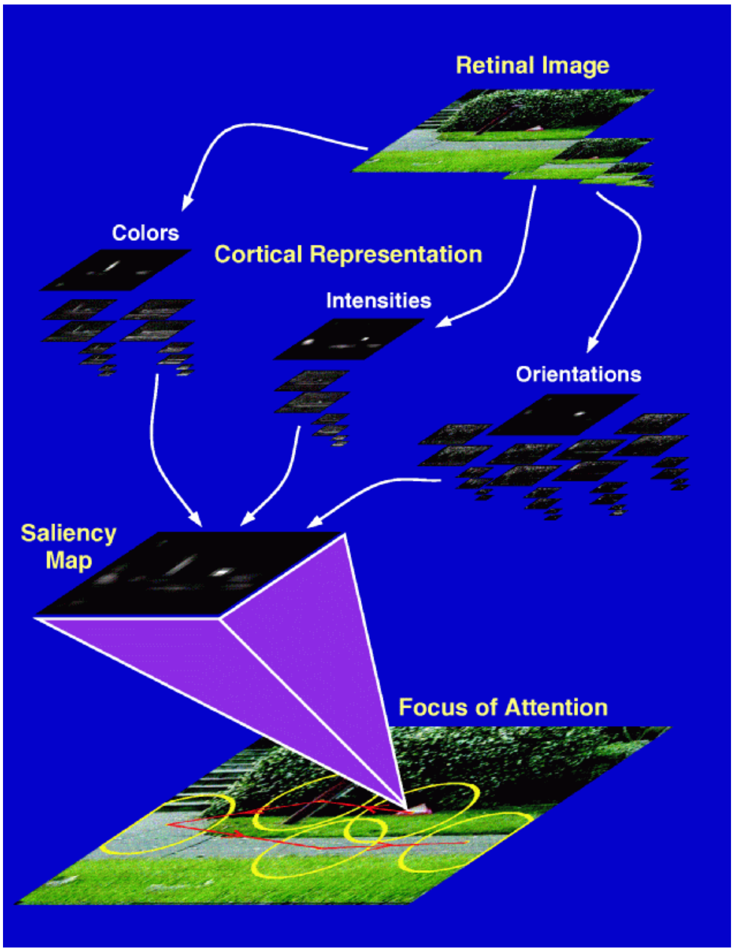
Contents:

- Saliency map Model
- Visual preprocessing
- Center surround differences
- Normalization
- Conspicuity maps
- Saliency map
- Example
- Visual Search

The Saliency-Map Model

- Localizes **salient** points in the visual field.
- Saliency is based on (bottom-up) scene-based properties
- Reduces computation by a selection on basis of preattentively computed simple features.
- Addresses some problems with the integration of different feature dimensions into a space-related map.





Multi-resolution pyramid

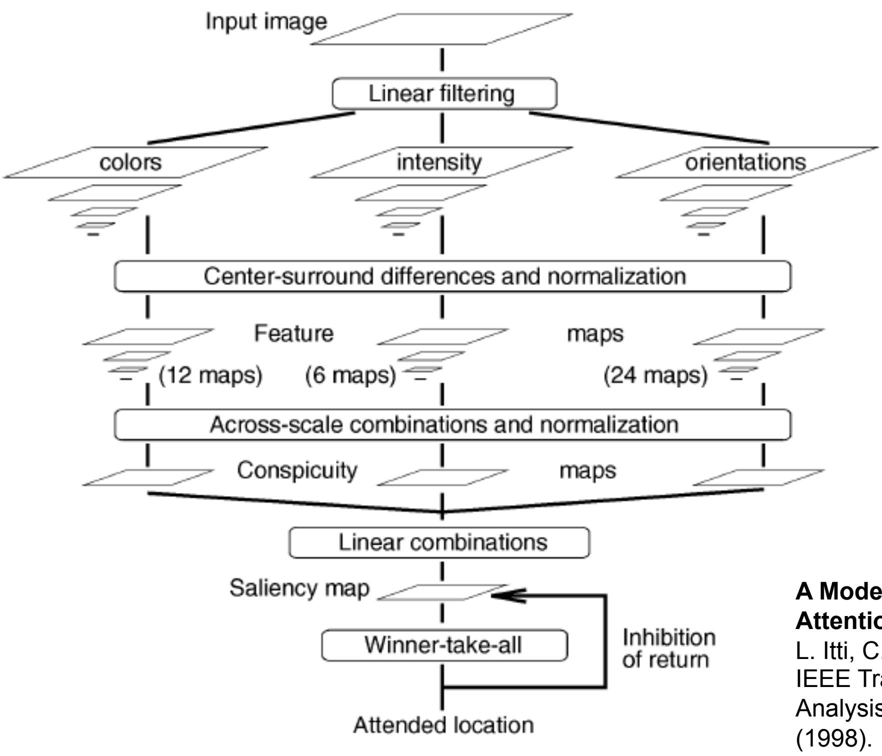
Computation of different channels
Feature filter for colors
intensity
orientations

Combination of the feature maps into a conspicuity map for each channel

Combination of the conspicuity maps into a saliency map

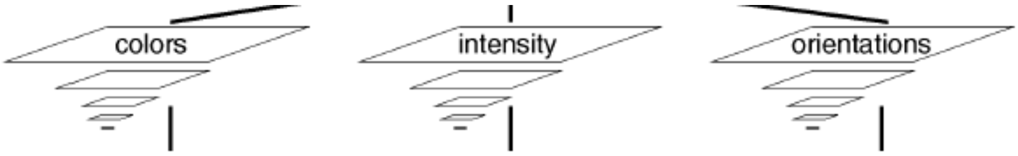
Serial selection of salient locations

Saliency-Model



A Model of Saliency-Based Visual Attention for Rapid Scene Analysis
L. Itti, C. Koch, and E. Niebur
IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (1998).

Visual Preprocessing



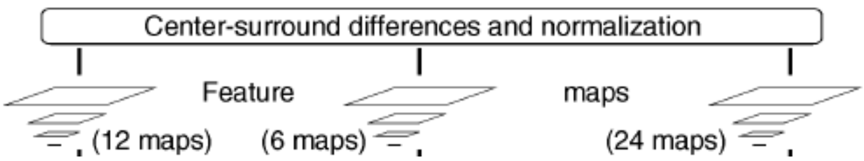
Starting from r, g, and b, the color values (red, green and blue) of the input image, an intensity image I=(r+g+b)/3 is obtained.

For each pixel in the pyramid, generate color channels R=r-(g+b)/2 for red, G=g-(r+b)/2 for green, B=b-(r+g)/2 for blue, and Y=(r+g)/2-|r-g|/2-b for yellow (negative values are set to zero).

Determine color opponency RG=R-G and BY=B-Y.

The detection of local orientation at each point in the image is achieved using overcomplete steerable filters O

Center-surround differences



Compute center-surround differences to determine contrast, by taking the difference between a fine (center) and a coarse scale (surround) for a given feature. This operation across spatial scales is done by interpolation to the fine scale and then point-by-point subtraction.

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

$$\mathcal{RG}(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))|$$

$$\mathcal{BY}(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| .$$

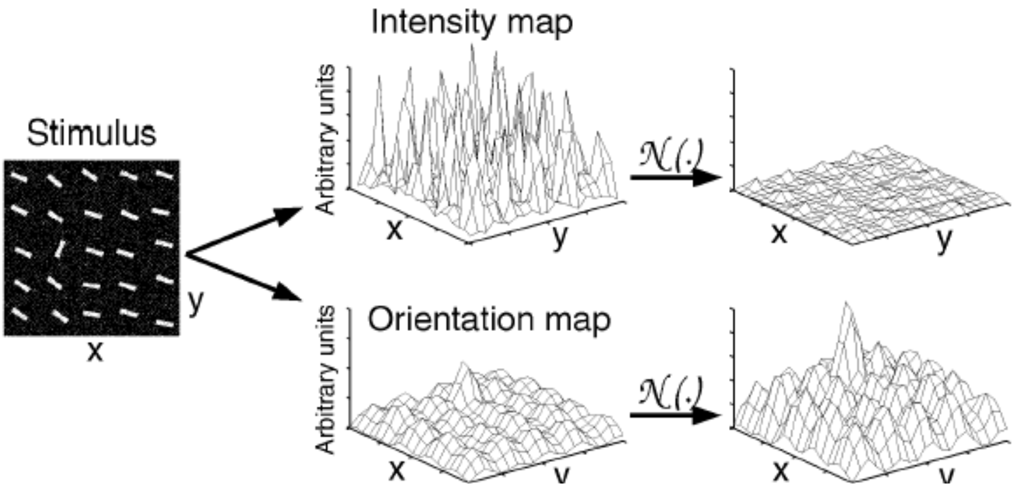
$$c \in \{2, 3, 4\}$$

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

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

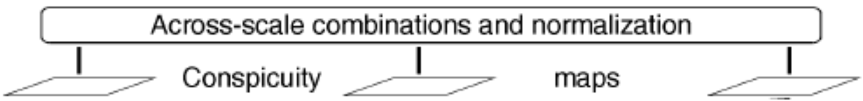
Normalization

$\mathcal{N}(.)$



1) normalizing the values in the map to a fixed range $[0..M]$, in order to eliminate modality-dependent amplitude differences;
2) finding the location of the map's global maximum M and computing the average $\overline{m}$ of all its other local maxima; and
3) globally multiplying the map by $(M - \overline{m})^2$.

Conspicuity maps



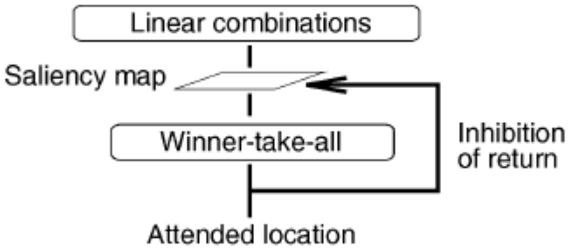
$$\overline{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c=4} \mathcal{N}\left(I(c,s)\right)$$

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

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

The feature maps are combined into three conspicuity maps at the scale 4. This is obtained through across-scale addition by reducing each map to the lowest resolution (scale 4) and point-by-point addition.

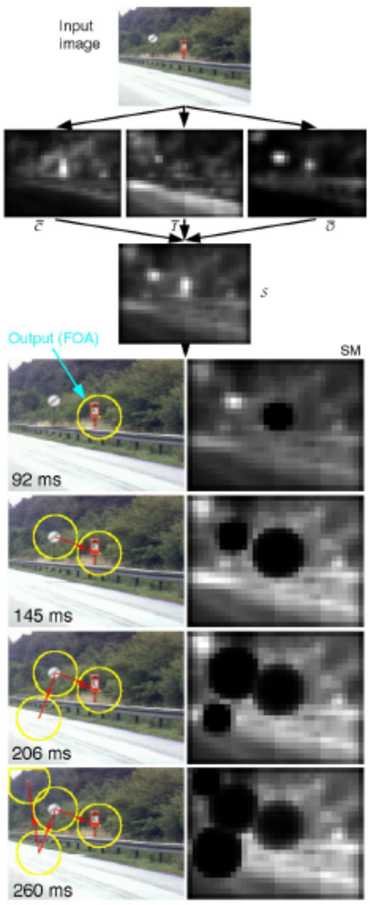
Saliency Map



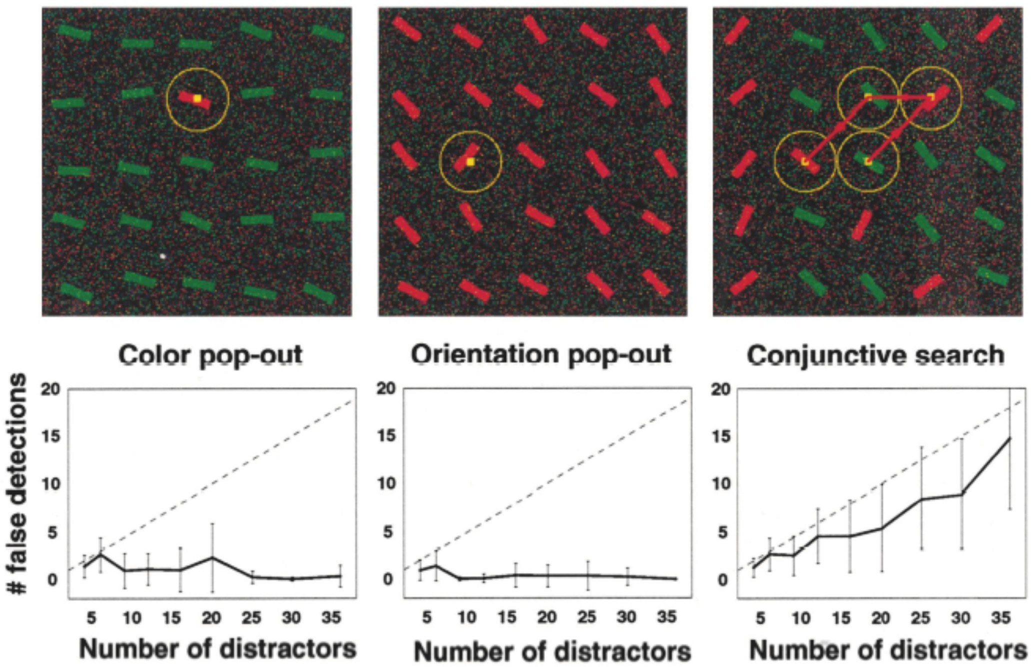
$$S = \frac{1}{3}\left(\mathcal{N}(\bar{I}) + \mathcal{N}(\bar{C}) + \mathcal{N}(\bar{O})\right)$$

The three conspicuity maps are normalized and summed into the final input *S* to the saliency map.

Example



Visual Search



The model suggests a serial search mechanism. Noise in the scene avoids local pop-out.

Discussion

- The saliency model provides a useful algorithm for guiding vision to potentially meaningful parts of a scene.

- It selects only a point in space, as compared to an object or region. Region selection has to be added by a separate mechanism.
- Saliency is restricted to simple features.
- Attention is defined solely as the selection in space (no, or only indirect feature-based selection).
- The advantage of this mechanism for object recognition is limited, since a selection in space does not necessarily promote object-recognition.

Additional reading:

- Peters, R. J., Iyer, A., Itti, L., Koch, C. (2005) Components of bottom-up gaze allocation in natural images, Vision Research, 45: 2397-2416.
- Einhäuser W., Spain, M. Perona, P. (2008) Objects predict fixations better than early saliency. Journal of Vision, 8(14):18,1-26.