Biological Vision and Applications
Module 05-02: Visual attention: Cognitive model

Hiranmay Ghosh

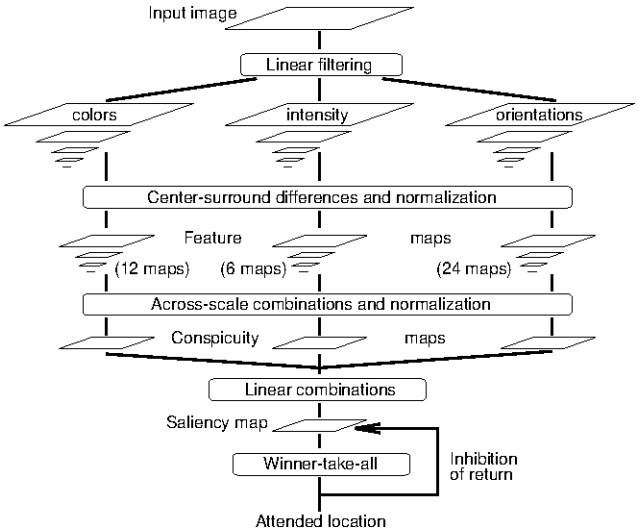
Cognitive Models

Based on Feature Integration Theory

- Motivated by the observations
  - Higher acuity at central vision, lower at peripheral
  - Early vision can distinguish local contrasts
    - Intensity contrast (Dark vs. Bright)
    - Color contrast (Red vs. Green and Blue vs. Yellow)
    - Edge Orientation
  - Features are subsequently integrated
    - Treismann's Feature Integration Theory

Itti's model (1998)

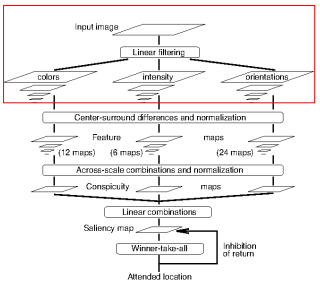
Overview

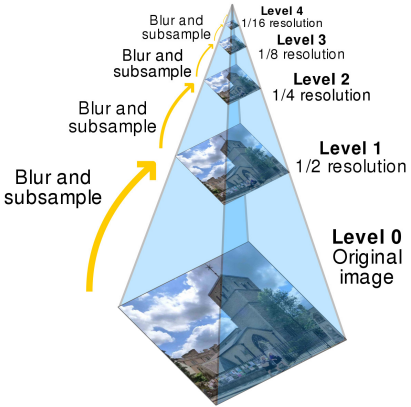


Itti's model: Stage 1

Multi-resolution image analysis

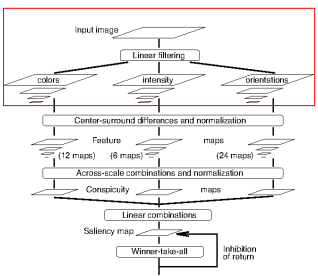
- Multi-resolution analysis of input image
  - Using Gaussian pyramids (9 scales)





Itti's model: Stage 1

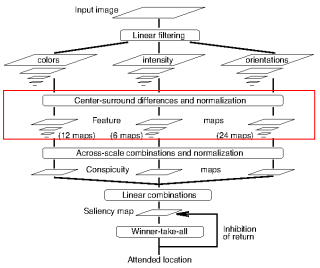
Feature extraction



- For images at each resolution level, 3 features are extracted
  - Color (C): R-G and B-Y contrasts
  - Intensity (I): B-W contrast
  - Edge Orientations (O): 0, 45, 90, 135 degrees
- $2 + 1 + 4 = 7$ features extracted for each resolution level

Itti's model: Stage 2

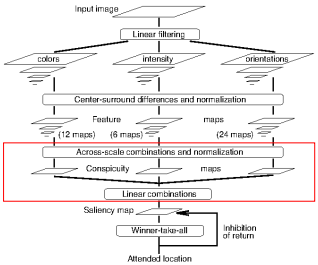
Center-surround operations: Multi-scale feature maps



- Done for each feature
- Center at hi-res, Surround at lo-res
- Scales used:
  - Center: $c = \{2, 3, 4\}$
  - Surround: $s = c + \delta$ $[\delta = \{3, 4\}]$
- Multi-scale differences
  - $\mathcal{F} = \mid F(c) \ominus F(s) \mid$
- 6 scales for each feature
- $7 \times 6 = 54$ "feature maps"
  - Each represents local contrast at a location based on a feature at a certain scale

Itti's model: Stage 3

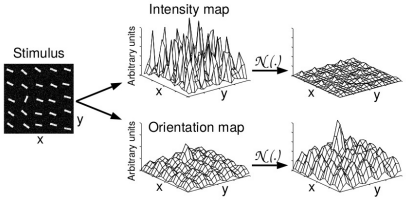
Combining the features: Conspicuity and Saliency Maps



- Feature maps are combined
- Equal weights – normalized $N()$
- Combined in two stages
  - Intra-feature-class, giving three *conspicuity maps*
    - $\bar{I} = \bigoplus_{c,s} N(I(c,s))$
    - $\bar{C} = \sum_{RG,BY} \bigoplus_{c,s} N(C(c,s))$
    - $\bar{O} = \sum_{\theta} \bigoplus_{c,s} O(c,s)$
  - Inter-feature-class, giving the final *saliency map*
    - $S = \bar{I} + \bar{C} + \bar{O}$

Itti's model: Stage 3

Normalization

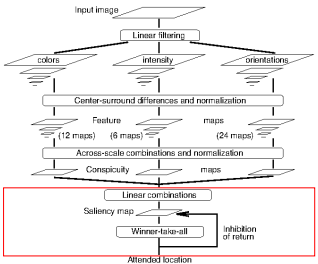


Maxima	**6.00**	**7.00**	**5.00**	**6.00**	**5.00**
Normalized	**0.04**	**0.05**	**0.03**	**0.04**	**0.03**
Maxima	**6.00**	**20.00**	**5.00**	**6.00**	**5.00**
Normalized	**0.16**	**0.53**	**0.13**	**0.16**	**0.13**

- Two reasons to normalize
  - Features are at arbitrary scale
  - Normalize to a fixed range $[0 \dots M]$
- Some feature may have many nearly equal peaks, indicating texture
  - Find the global maximum M
  - Compute the average of all other local maxima $\bar{m}$
  - Multiplying the map by $(M - \bar{m})^2$

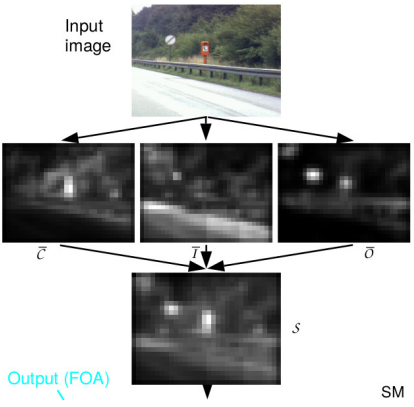
Itti's model: Stage 4

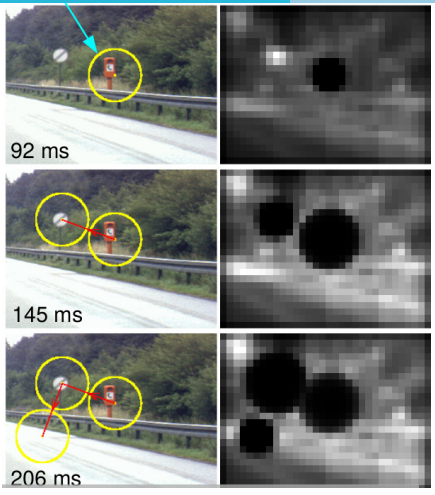
"Winner take it all" and "Return Inhibition" policies



- Winner-take-it-all policy
  - The image location with highest saliency attracts attention
  - All other locations are ignored
- Return Inhibition policy
  - Attention never returns to a location once attended
  - The neurons at the attended place tires out.
  - Attention moves to the location with next highest salience.

Results


Input
image
$\bar{C}$
$\bar{I}$
$\bar{O}$
$\mathcal{S}$
Output (FOA)
SM


92 ms
145 ms
206 ms

Quiz

Quiz 05-02

End of Module 05-02