Visual Attention:

From Theory to Efficient Implementation

Francesco Rea

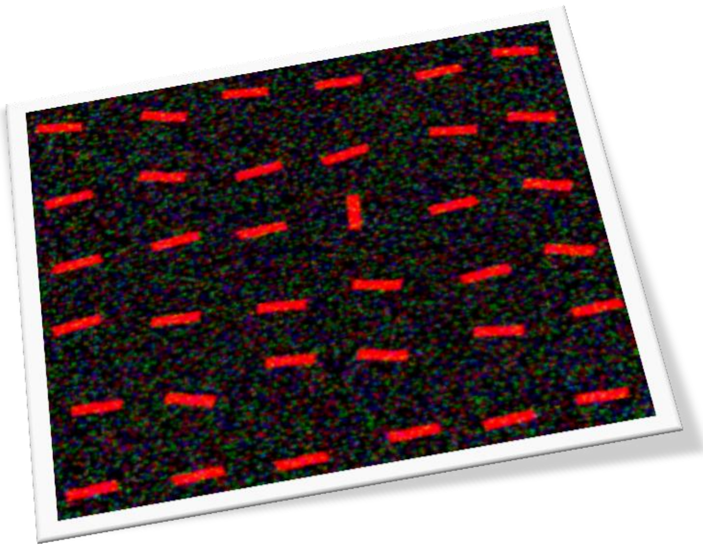
Outline

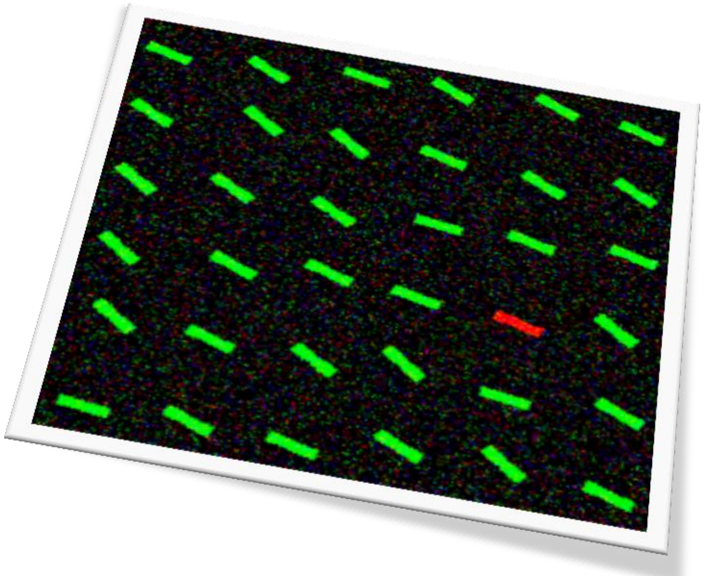
- 1. Evolutionary Reasons of Visual Attention Development
  - 1.1 Reduction in of processing demand in vision
  - 1.2 Binding mechanism and association through visual attention
  - 1.3 Digression : Interaction with spatial variant vision
  - 1.4 Applications
- 2. Classification of visual attention
  - 2.1 region-based and feature-based
  - 2.2 top-down and bottom-up
  - 2.3 object-based visual attention
- 3. Biological processes of visual attention
  - 3.1 Retina and Ganglion Cells
  - 3.2 Lateral Geniculate Nucleus
  - 3.3 Magnocellular and Parvocellular Pathways
  - 3.4 Simple Cells and Complex Cells in primary visual cortex
  - 3.5Visual cortex: color and orientation

Outline

Relevant Advantages of Visual Attention

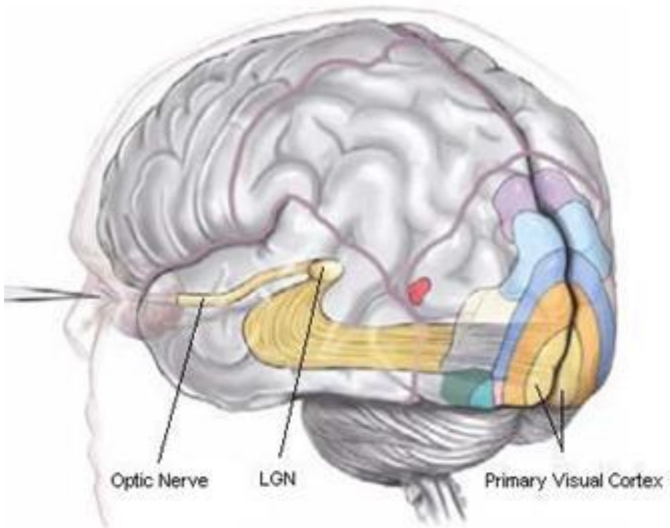
- Reduction of the processing demand in vision
- Binding mechanism and association through visual attention
- Digression : Interaction with spatial variant vision
- Applications



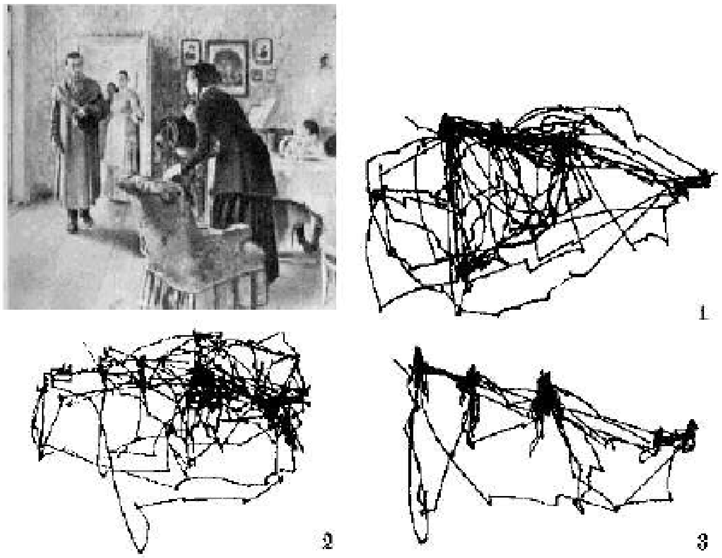


Relevant Advantages of Visual Attention

- During the phylogenetic evolution, visual attention has been valid solution for fast reaction time despite the bottleneck in the visual processing
  - limit in processing estimated in the order of $10^7 - 10^8$ bits per second at the optical nerve



- Reduction of the processing demand in vision



" The visual attention is a mechanism that reduces the effect of processing bottleneck in visual cortex, limiting the computation to the smaller subregion highlighted by the attentional focus "(Koch, The Quest for Consciousness, 2004).

Relevant Advantages of Visual Attention

- Digression : Interaction with Foveated vision
  - Result of the phylogenetic evolution
  - Model: log-polar mapping

$$\begin{cases} R = \lambda^{\rho} \cdot r_0 \\ \omega = \frac{2\pi\theta}{\Theta} \end{cases}$$

R is the distance between the center of the receptive field (RF) and the c the mapping.
ω is the counter-clockwise angular distance of the center of the RF from horizontal axis.
λ is the ratio between the widths of two subsequent rings.
r_0 is the radius of the ring with index ρ = 0.
Θ is the number of RF's per ring

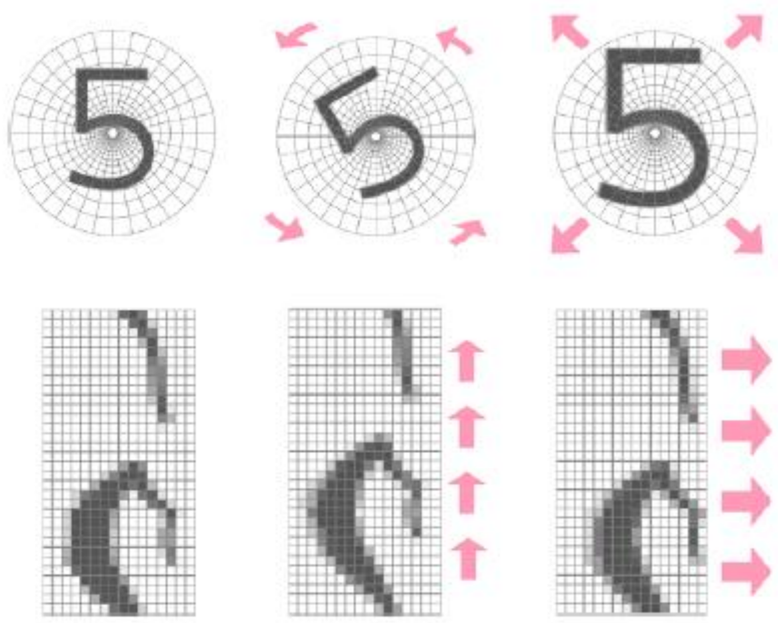
$$\begin{cases} x = R \cdot \cos(\omega) \\ y = R \cdot \sin(\omega) \end{cases} \quad \begin{cases} x = \lambda^{\rho} \cdot r_0 \cdot \cos(\frac{2\pi\theta}{\Theta}) \\ y = \lambda^{\rho} \cdot r_0 \cdot \sin(\frac{2\pi\theta}{\Theta}) \end{cases} \quad \begin{cases} \rho = \log_{\lambda} \frac{R}{r_0} = \log_{\lambda} \frac{\sqrt{x^2+y^2}}{r_0} \\ \theta = \frac{\Theta\omega}{2\pi} = \frac{\Theta}{2\pi} \arctan \frac{y}{x} \end{cases}$$

fovea: $\begin{cases} R = \rho + \varepsilon \\ \omega = \frac{2\pi\theta}{\Theta} \end{cases}$

otherwise: $\begin{cases} R = \lambda^{\rho} \cdot r_0 \\ \omega = \frac{2\pi\theta}{\Theta} \end{cases}$

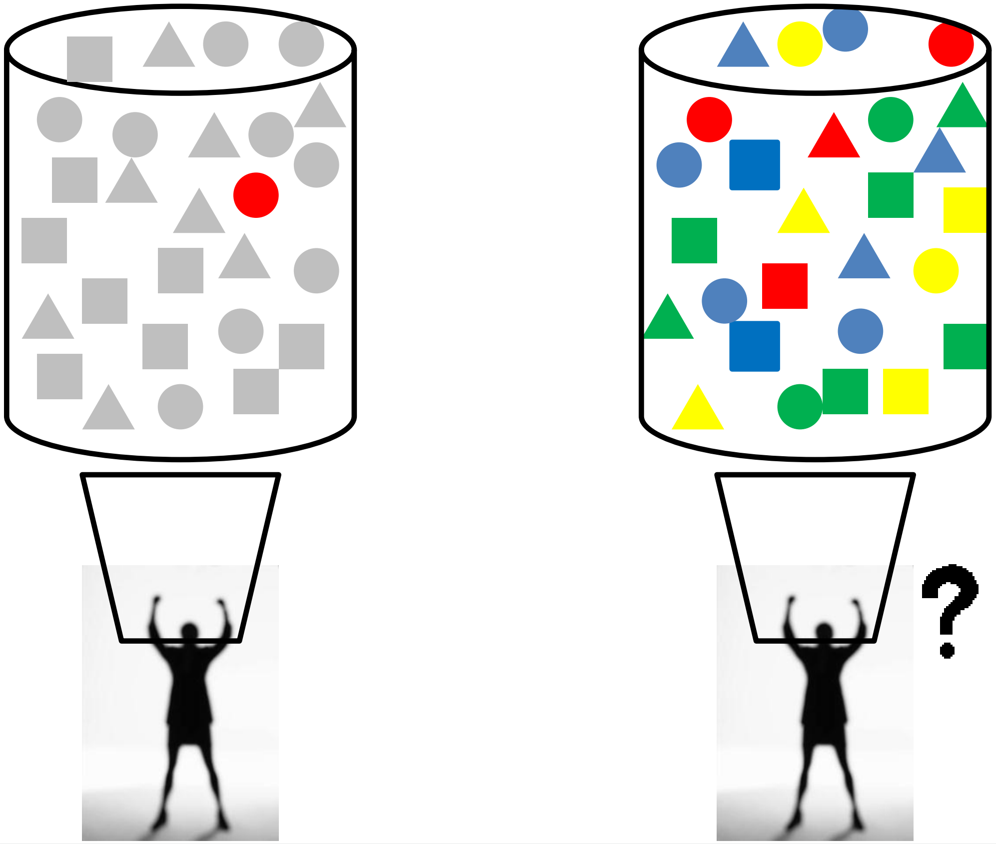
Relevant Advantages of Visual Attention

- Digression : Interaction with Foveated vision
  - Advantages:
    - Elegant trade of between wide field of view, high resolution and little data process
    - Biological Plausibility
    - Rotation and scale invariance



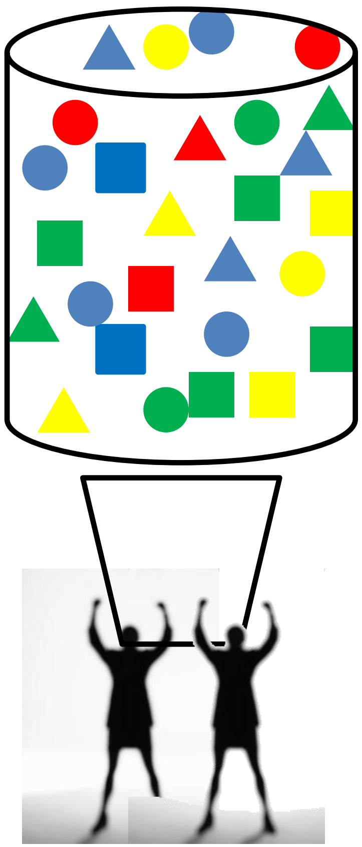
Relevant Advantages of Visual Attention

- Binding mechanism and association through visual attention

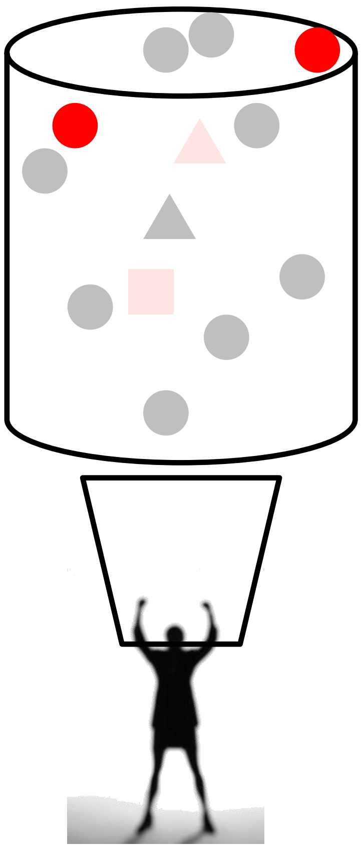


Relevant Advantages of Visual Attention

- Binding mechanism and association through visual attention



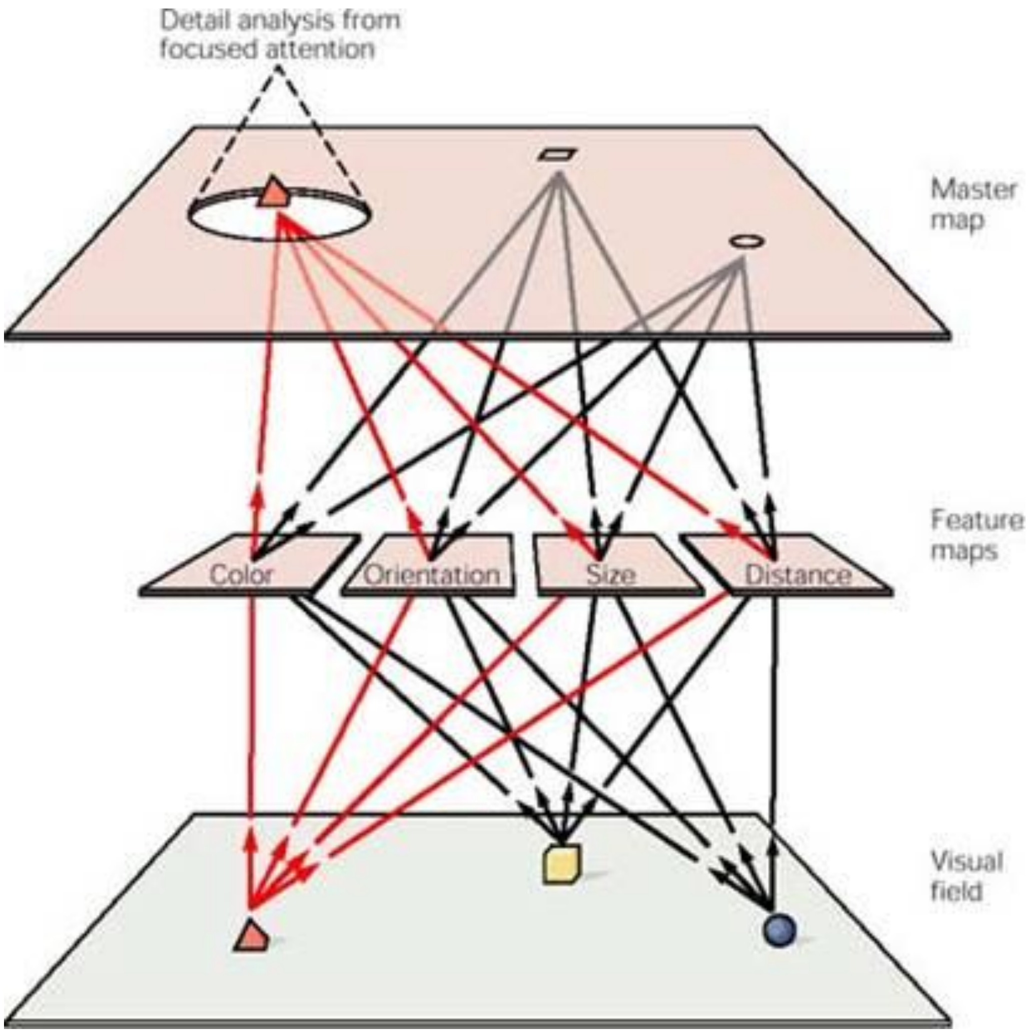
FORM COLOR



COLOR: red, FORM: circle

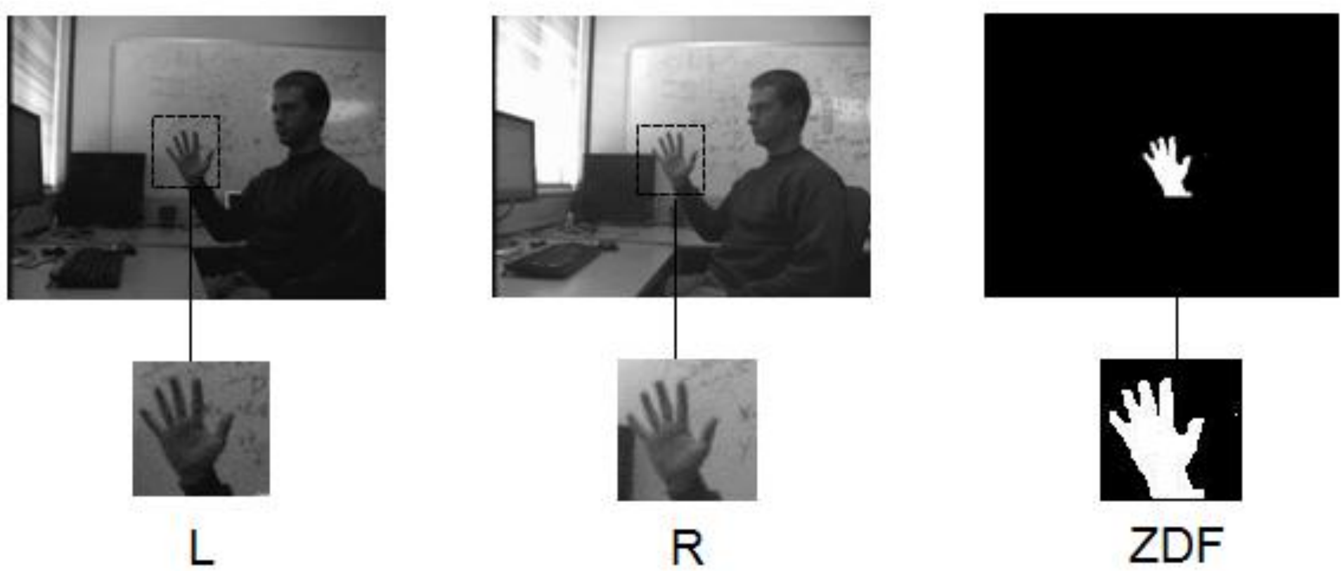
Classification of visual attention models

- Region-based vs. Feature-based
  - *Reference : Treisman 1986*



Relevant Advantages of Visual Attention

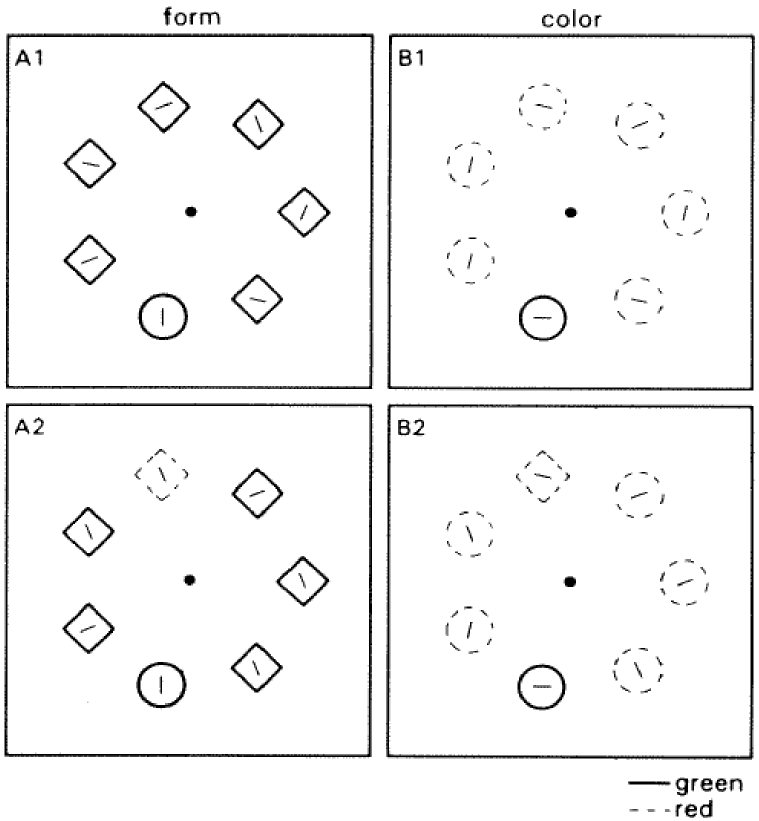
- Applications:
  - Object Recognition: starting from redeployment of fixation point
    - Zero disparity Filter : ZDF



  - Tracking: attention locus as prior

Classification of visual attention models

- Region-based vs. Feature-based
  - In literature also know as *singleton search mode* vs *feature search mode*
  - *Reference : Teeuwes 1992, Bacon & Egeth 1994, Teeuwes 2004 , Leber & Egeth 2006*
  - Theeuwes's (2004) model, there exist two essential criteria for examining stimulus-driven capture.
    1. parallel search , display size independent
    2. salience of the irrelevant distractor must not be compromised by characteristics of the search display



Classification of visual attention models

- Top-down vs. bottom-up visual attention
  - Theory and implementation
  - Experiment : Where is Waldo?

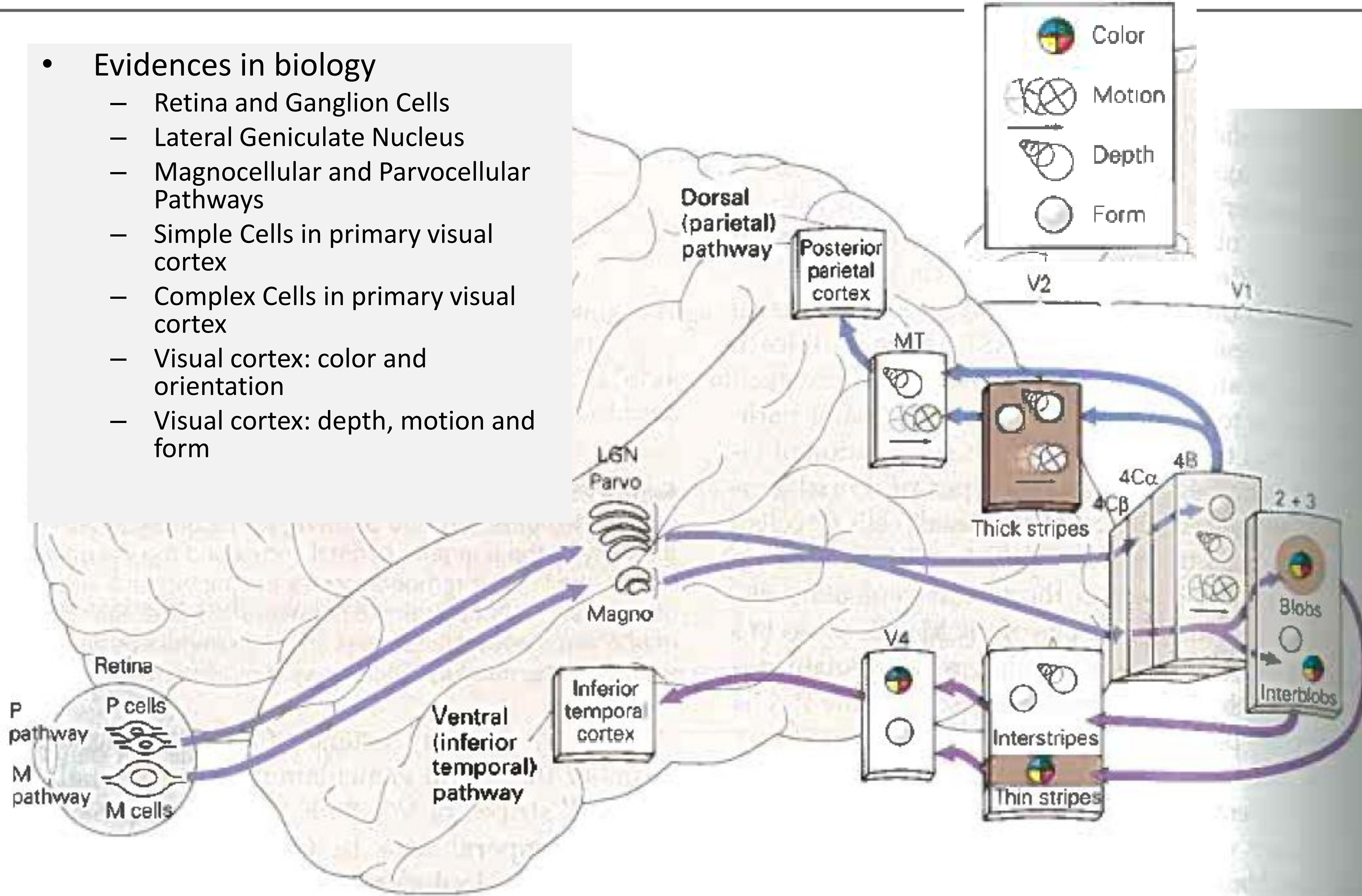
Classification of visual attention models

- Reference

  - Bacon, W.F. and Egeth, H.E. (1994). Overriding stimulus-driven attentional capture. Percept. Psychophys. *55, 485-496.*
  - Burr, D. C., Morrone, M. C. & Ross, J. Selective suppression of the magnocellular visual pathway during saccadic eye movements.*Nature* 371, 511-513 (1994).
  - Theeuwes, J. (1992). Perceptual selectivity for color and form. Percept. Psychophys. *51, 599-606.*
  - Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. Psychon. Bull. Rev. *11, 65-70*

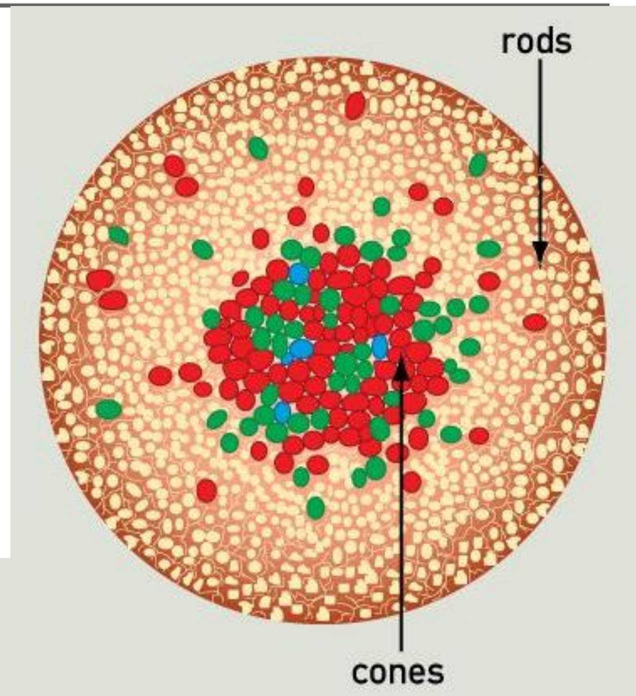
Biological process in visual attention

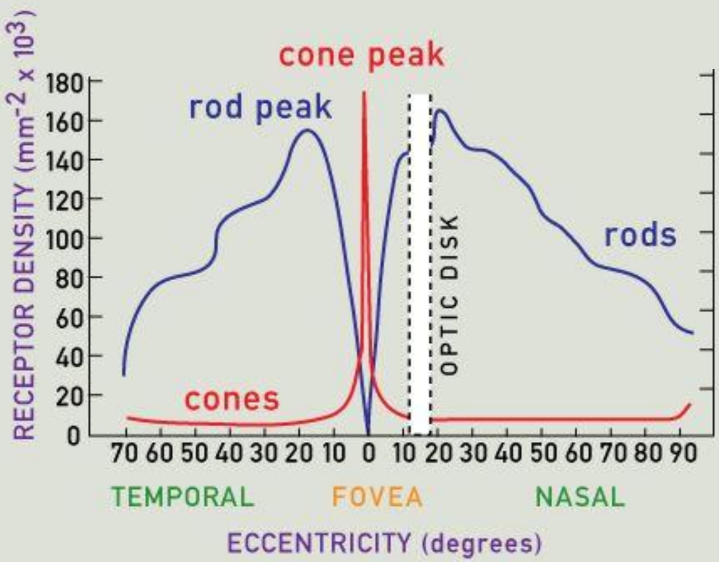
- Evidences in biology
  - Retina and Ganglion Cells
  - Lateral Geniculate Nucleus
  - Magnocellular and Parvocellular Pathways
  - Simple Cells in primary visual cortex
  - Complex Cells in primary visual cortex
  - Visual cortex: color and orientation
  - Visual cortex: depth, motion and form

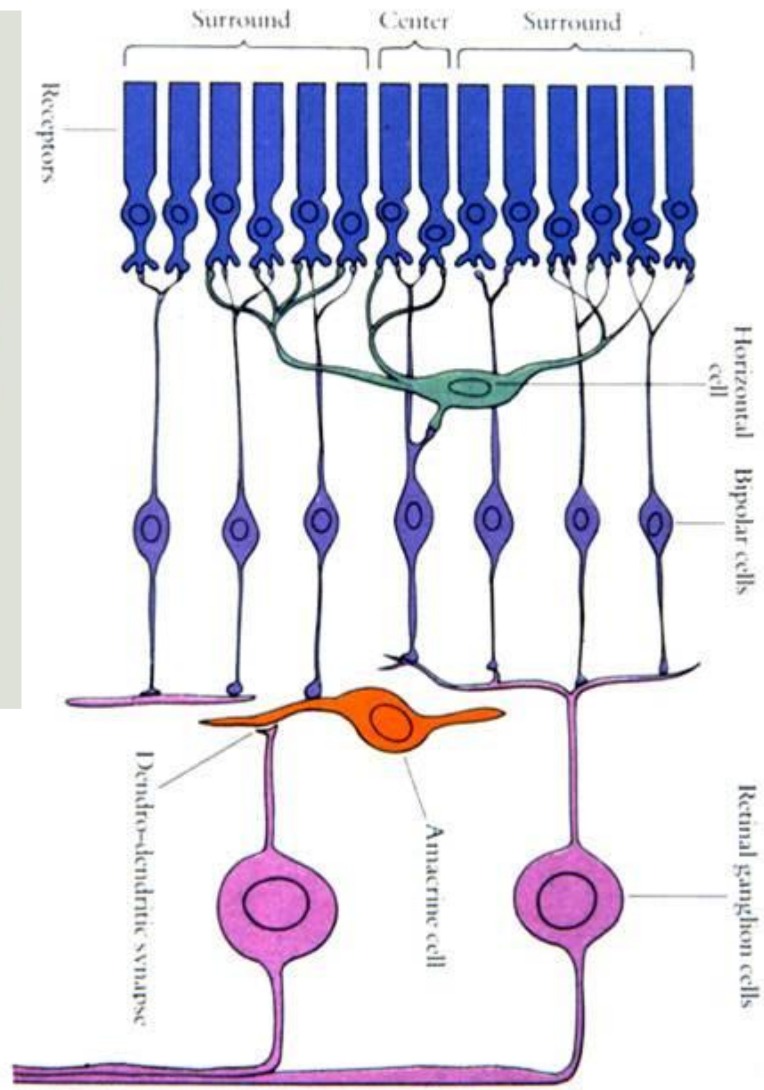


Biological process in visual attention

- Retina

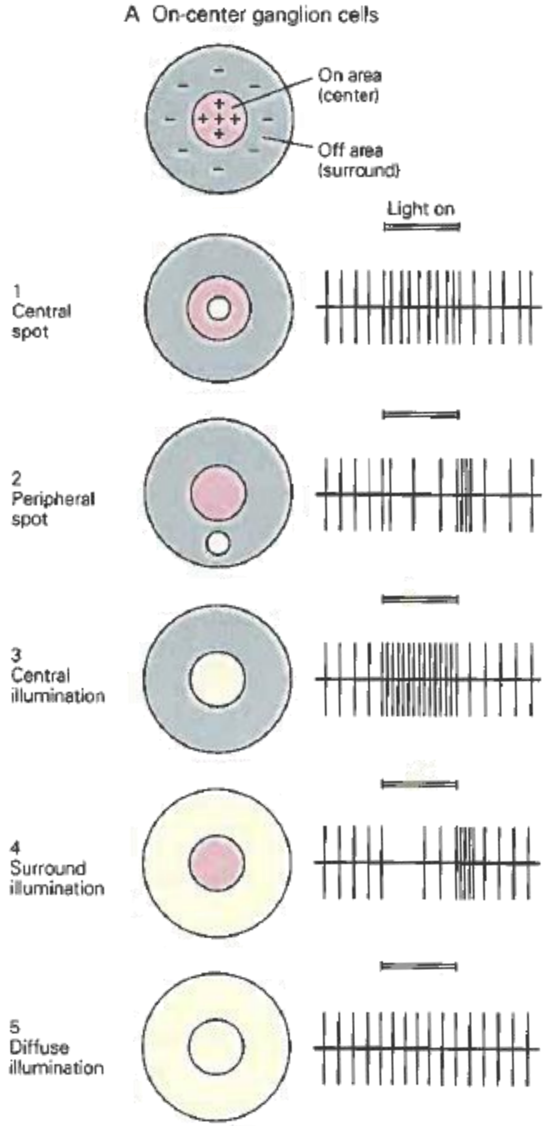






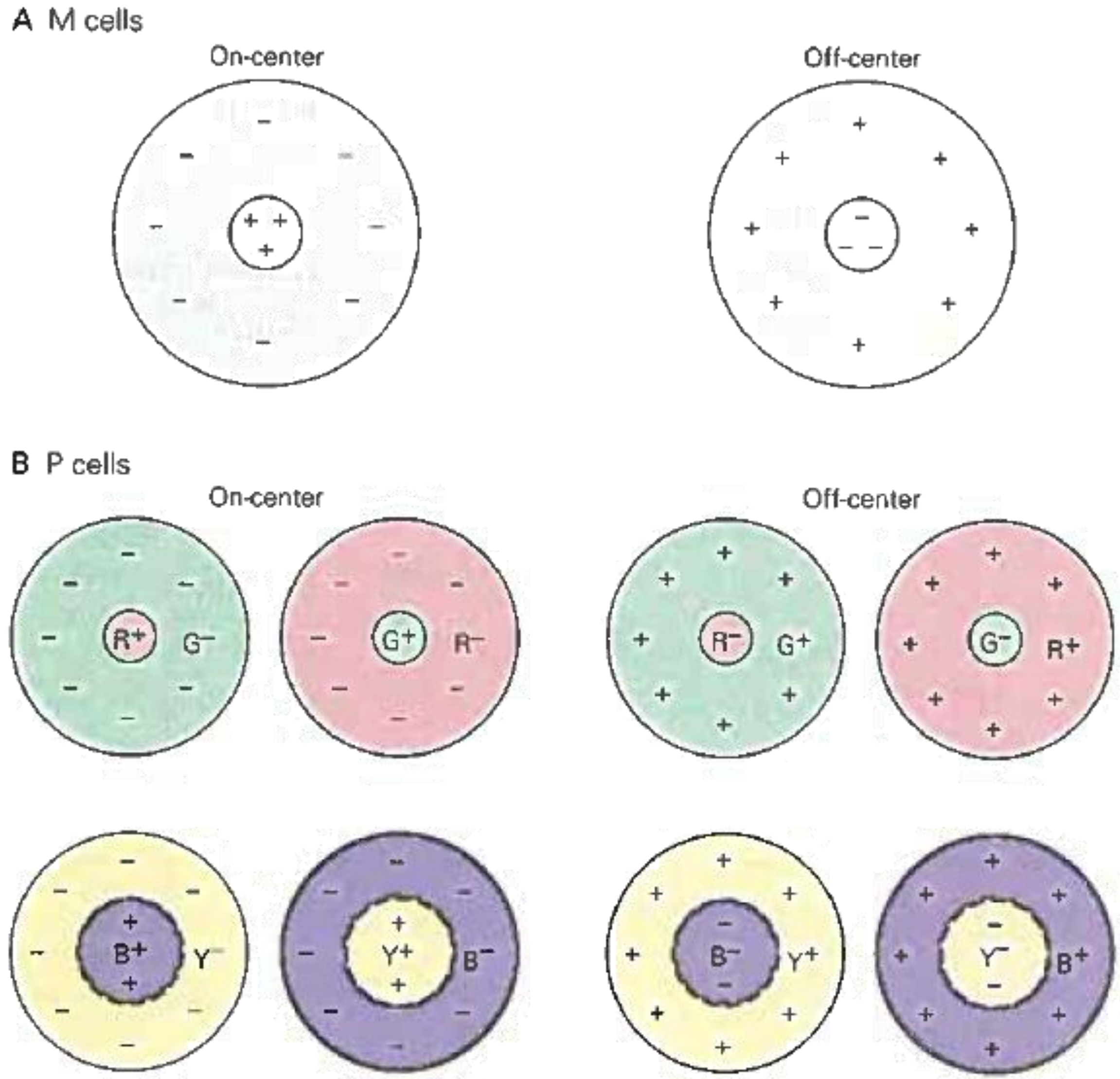
Biological process in visual attention

- Ganglion Cells
- Types of cell
  - M cells (magno), motion
  - P cells (parvo), color
- On center and Off center paradigm



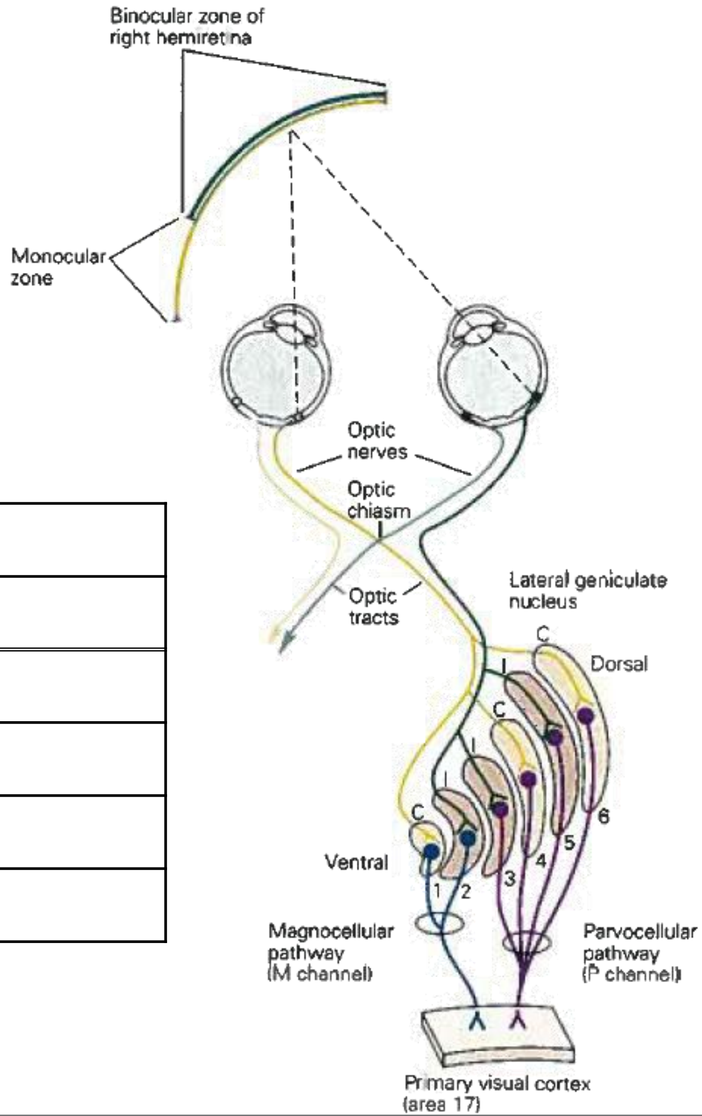
Biological process in visual attention

- M cells
  - Large receptive field
  - Respond to big objects and movement
- P cells
  - Numerous and small receptive fields
  - Involved in forms and colors



Biological process in visual attention

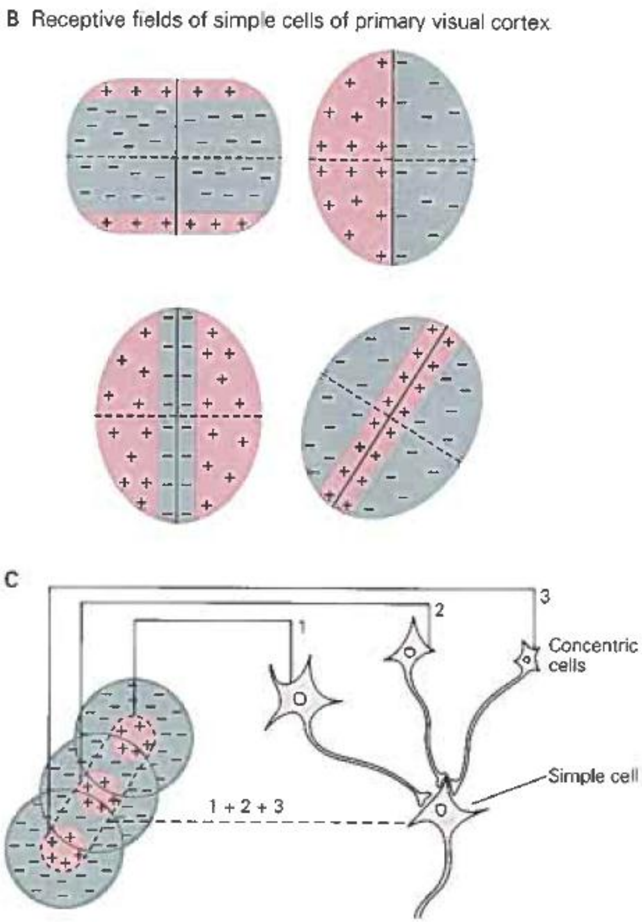
- Lateral Geniculate Nucleus LGN
  - 6 layers
  - Single Eye receptivity of layers
  - D.Huber & T.Wiesel 1960 found organization similar to retina

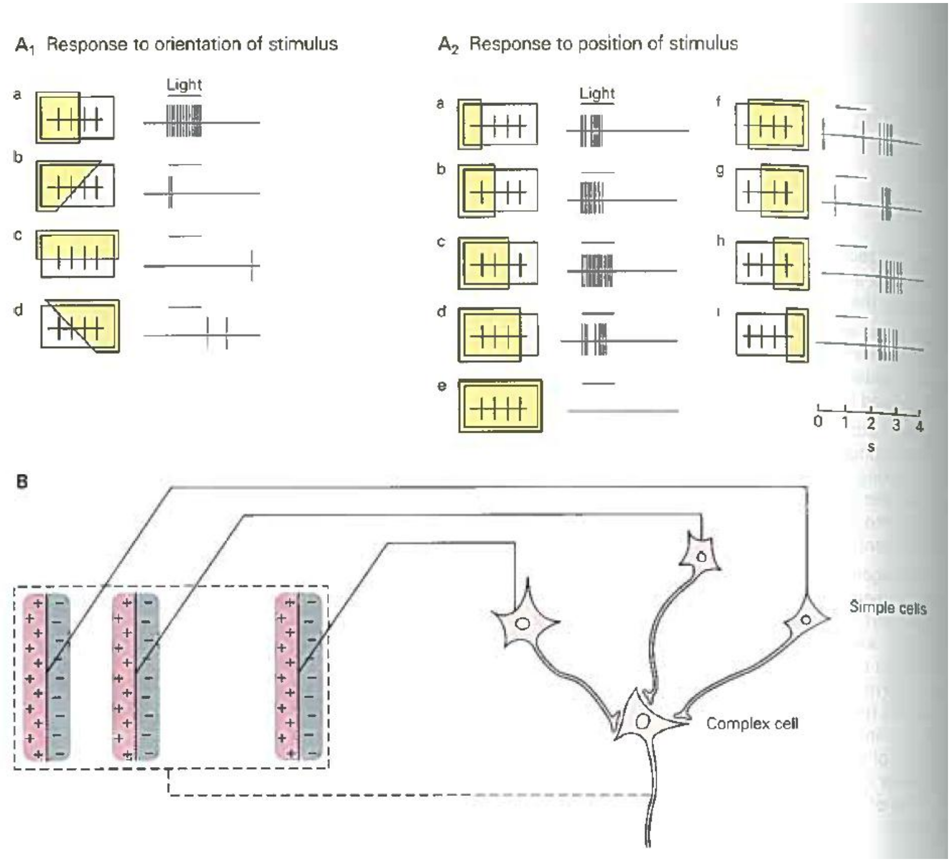


Stimulus Feature	Sensitivity	
	M Cells	P Cells
Color Contrast	No	Yes
Luminance Contrast	Higher	Lower
Spatial Frequency	Lower	Higher
Temporal Frequency	Higher	Lower

Biological process in visual attention

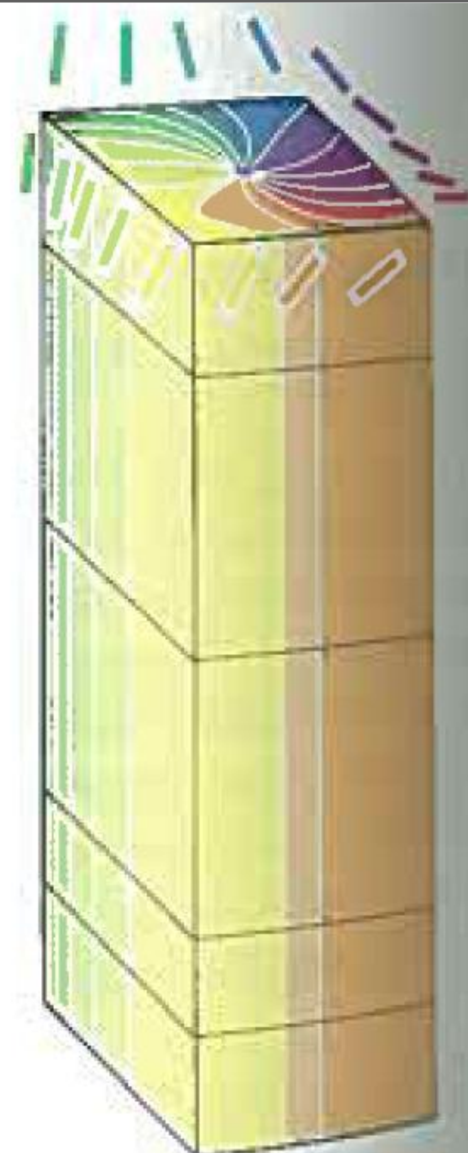
- Simple cells and Complex cells in primary visual cortex





Biological process in visual attention

- Visual cortex: color and orientation
  - *Hypercolumns*
    - *Orientation Columns*
    - *Blobs*
    - *Ocular Dominance*
  - Orientation column 1 x 1 x 2 (Hubel & Wiesel 1978)
  - Comparing local changes in reflectance, indicating activity
  - Blobs : color response



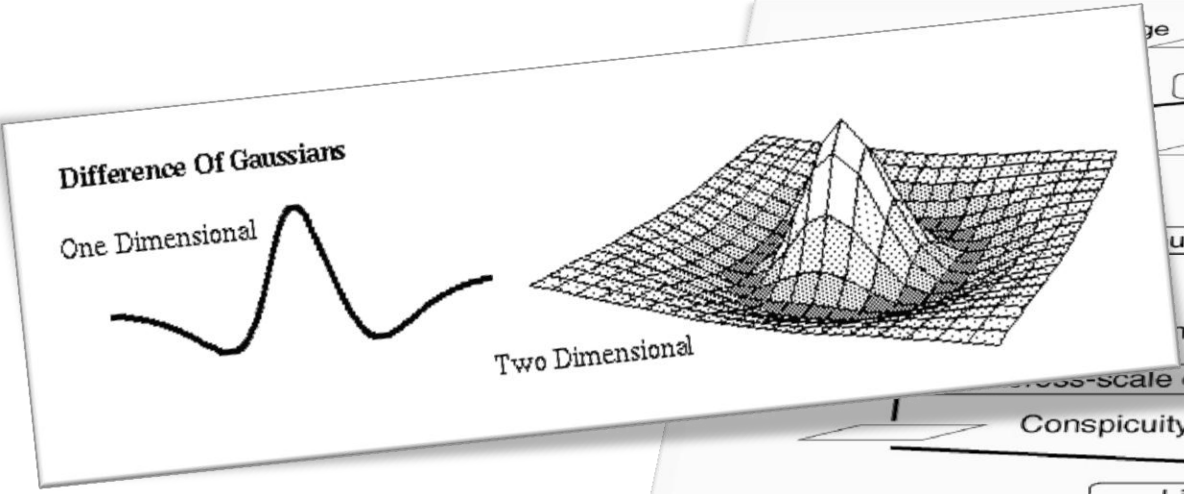
Courtesy of Gary Blasdel.

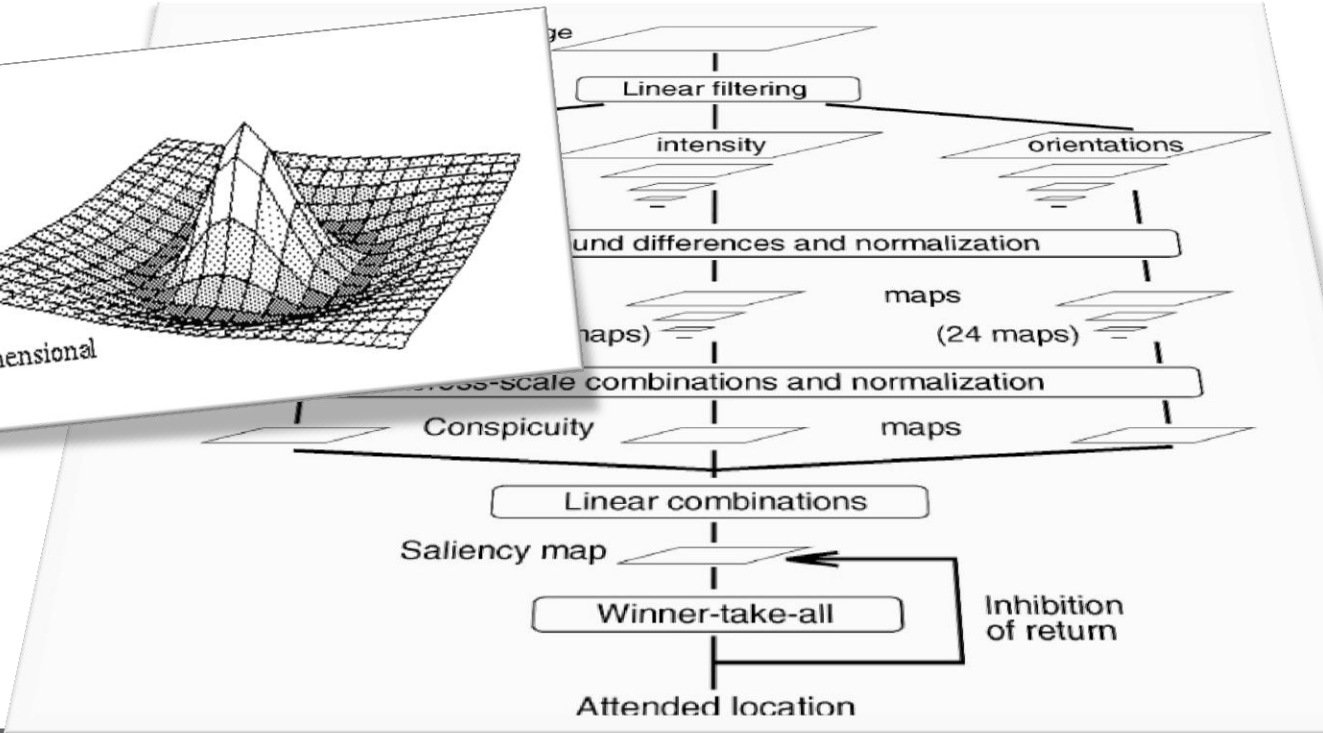
Biological process in visual attention

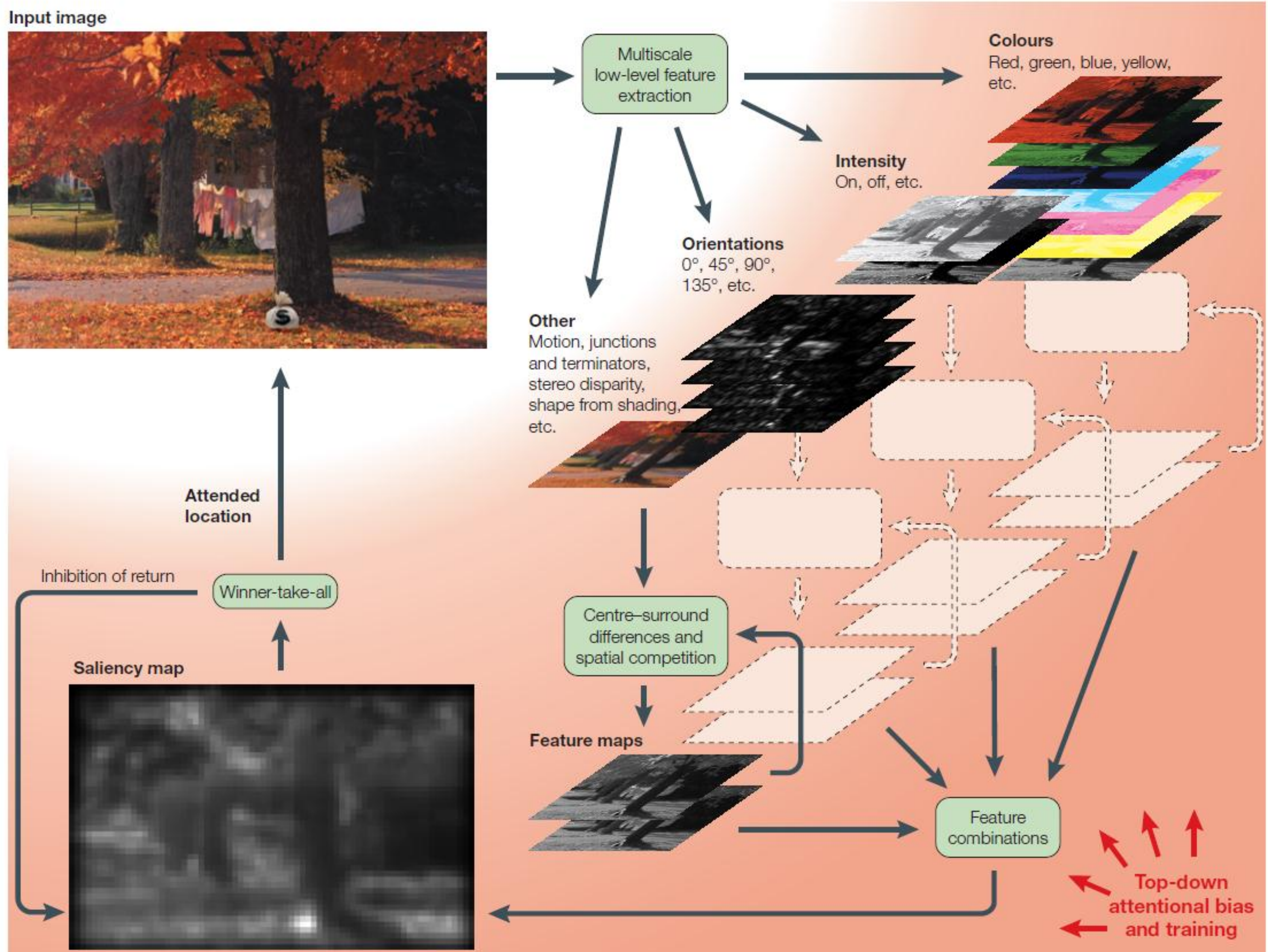
- Reference
  - Hubel D. & Wiesel T. 1965 Binocular interaction in striate cortex of kittens reared with artificial squint. J Neuropshysiol 28 : 1041-1059
  - Hubel D., Wiesel T., Stryker MP. 1978 Anatomical Demonstration of orientation columns in macaque monkey. J Comp Neurol 177:361-379

Preattentive System

4.1 Itti & Koch Model
4.2 Digression : Convolution
4.3 Digression : Difference of Gaussian
4.4 Digression : Pyramids of Convolutions
4.4 Feature Extraction in the Preattentive System







Input image
Multiscale low-level feature extraction
Colours
Red, green, blue, yellow, etc.
Intensity
On, off, etc.
Orientations
0°, 45°, 90°, 135°, etc.
Other
Motion, junctions and terminators, stereo disparity, shape from shading, etc.
Attended location
Inhibition of return
Winner-take-all
Centre–surround differences and spatial competition
Saliency map
Feature maps
Feature combinations
Top-down attentional bias and training

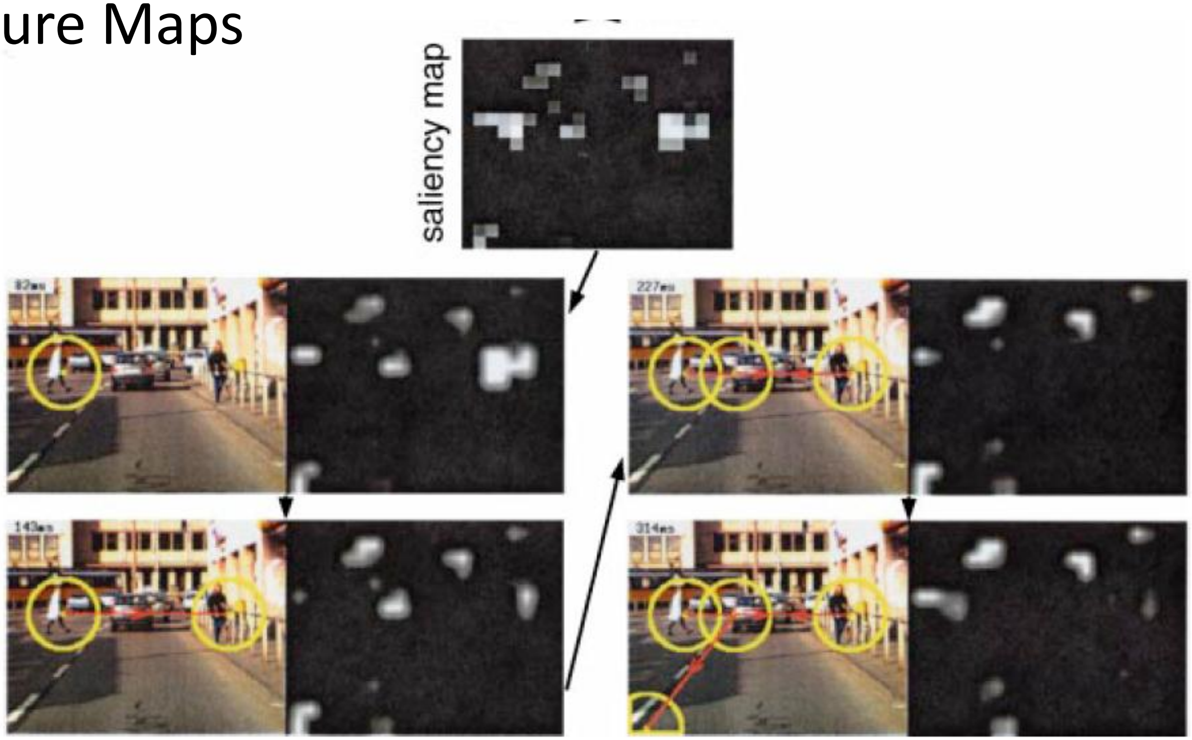
Preattentive System

- Itti & Koch Model
  - Koch & Ullman 1985
  - Itti & Koch 1998, 1999, 2001

•Multiple Multiscale Feature Maps from cartesian image
•Saliency Map

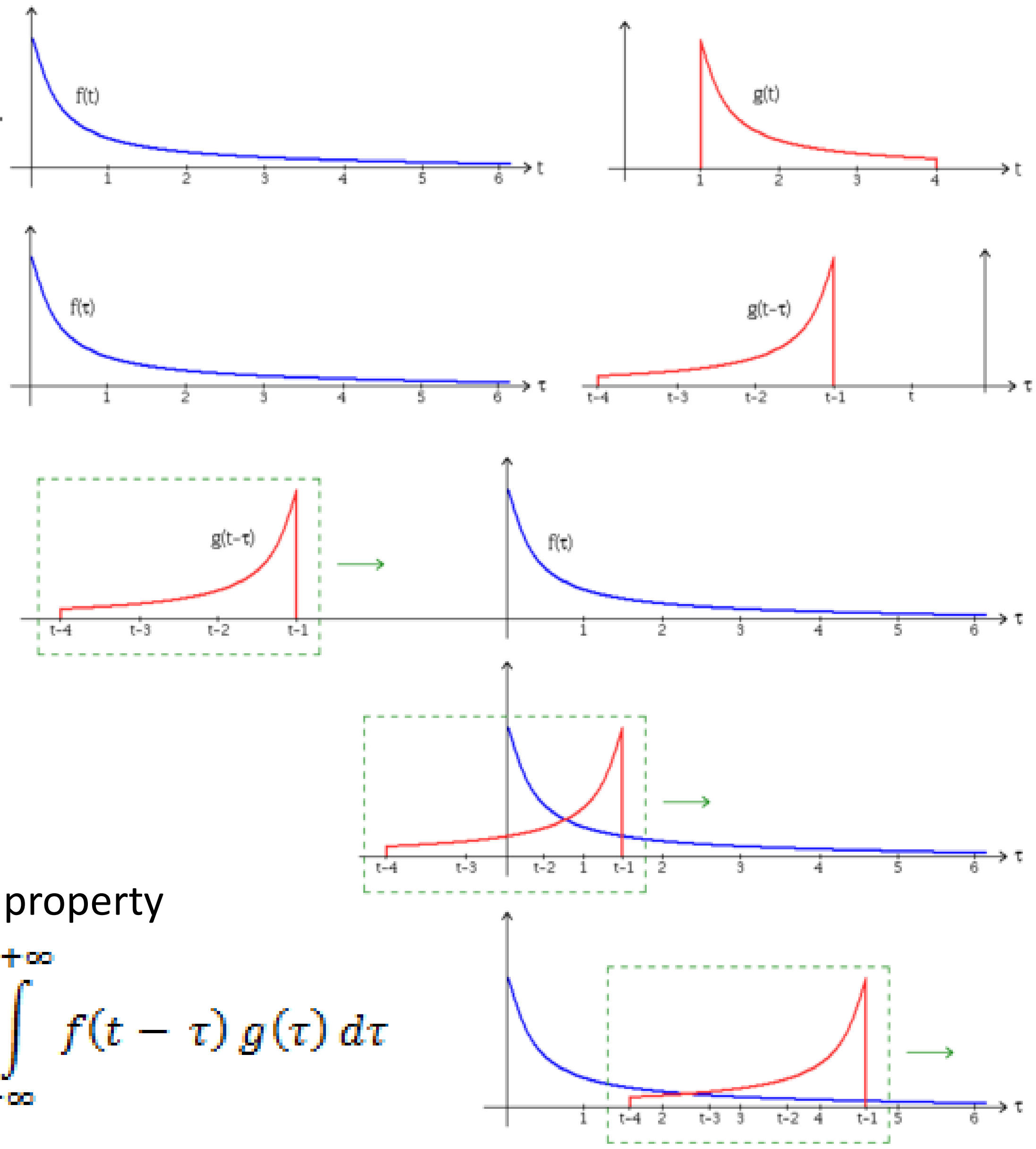
•WTA Winner Take All
•IOR Inhibition of Return

•Covert attention



Preattentive System

- Digression :
Convolution 1D

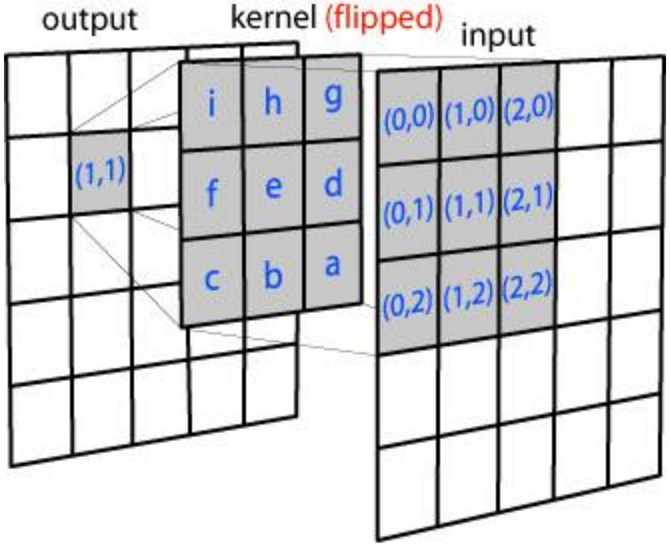


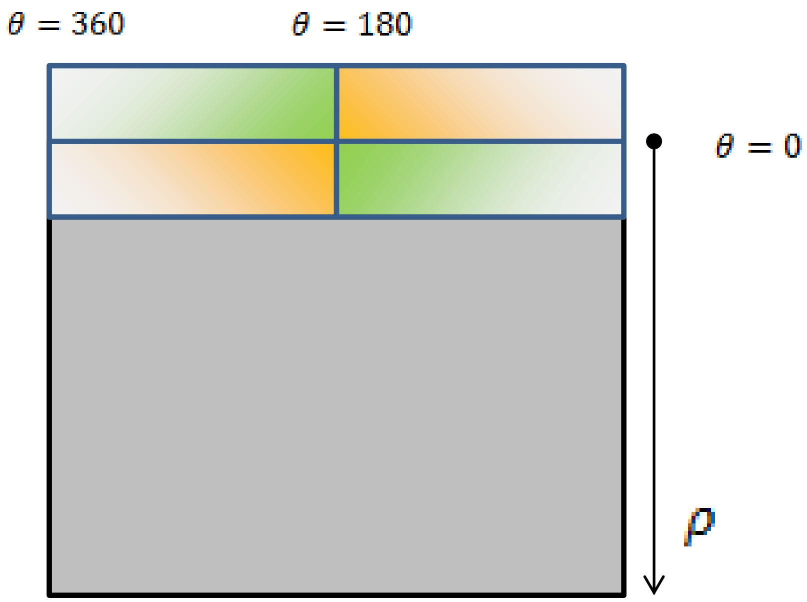
Commutative property

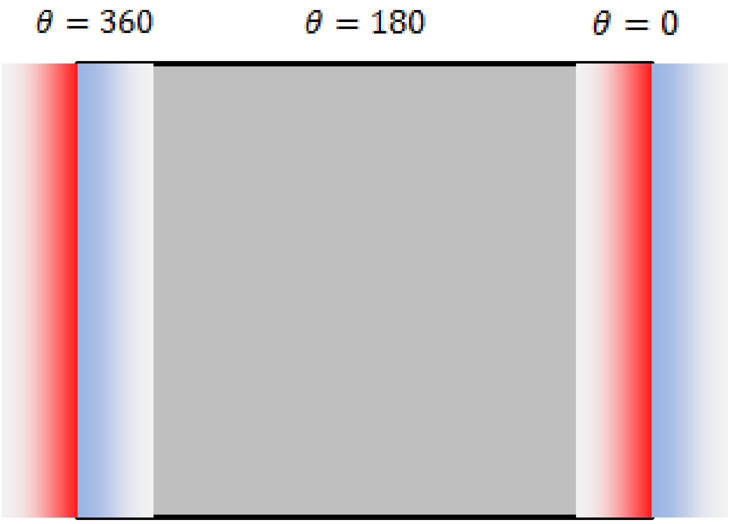
$$(f * g)(t) \stackrel{\mathrm{def}}{=} \int_{-\infty}^{+\infty} f(\tau)\, g(t-\tau)\, d\tau = \int_{-\infty}^{+\infty} f(t-\tau)\, g(\tau)\, d\tau$$

Preattentive System

- Digression : Convolution 2D
  - Windowing filter
  - Separable Kernels
  - Approach with log-polar image







Preattentive System

- Digression: Difference of Gaussian
  - “Mexican Hat”, simmetrical
  - On-Center Cell

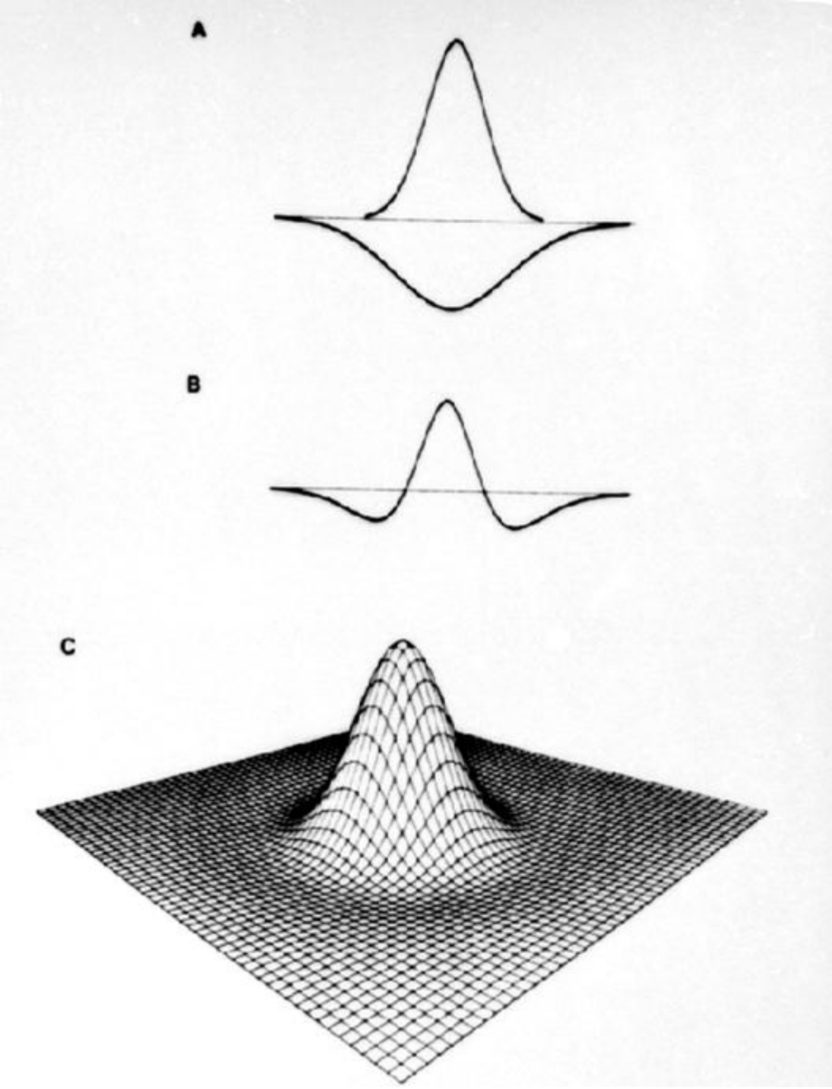
- Approach 1
  - Center: Gaussian variance 1 and Kernel 3x3
  - Surround: Gaussian variance 3 and Kernel 5x5
  - Unbalanced: the ratio β /α is chosen equal to 1.5 (Smirnakis et al. 1997)

e.g:

$$R^{+}G^{-}(x,y) = \alpha \cdot R * g_c - \beta \cdot G * g_s$$

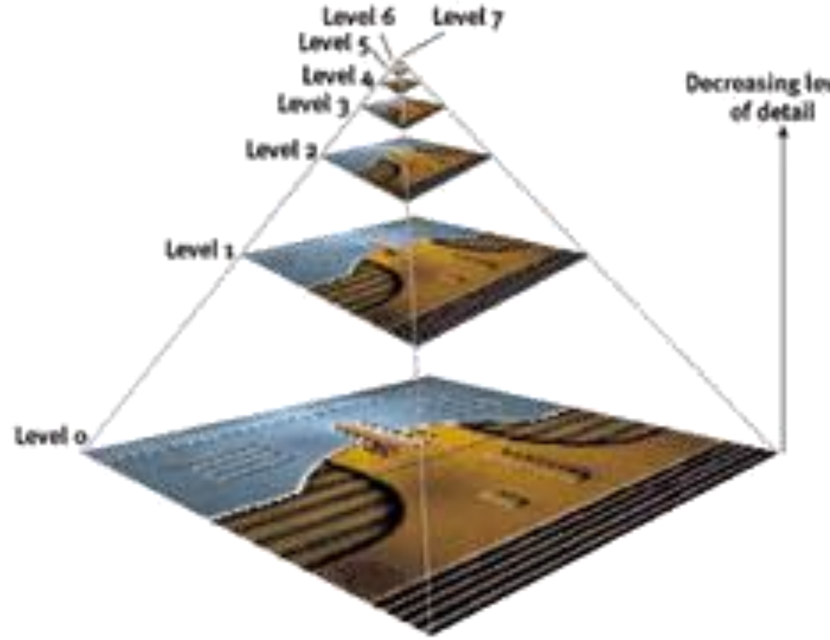
  - Reference: Hurvich & Jameson 1957 An opponent-process theory of color vision

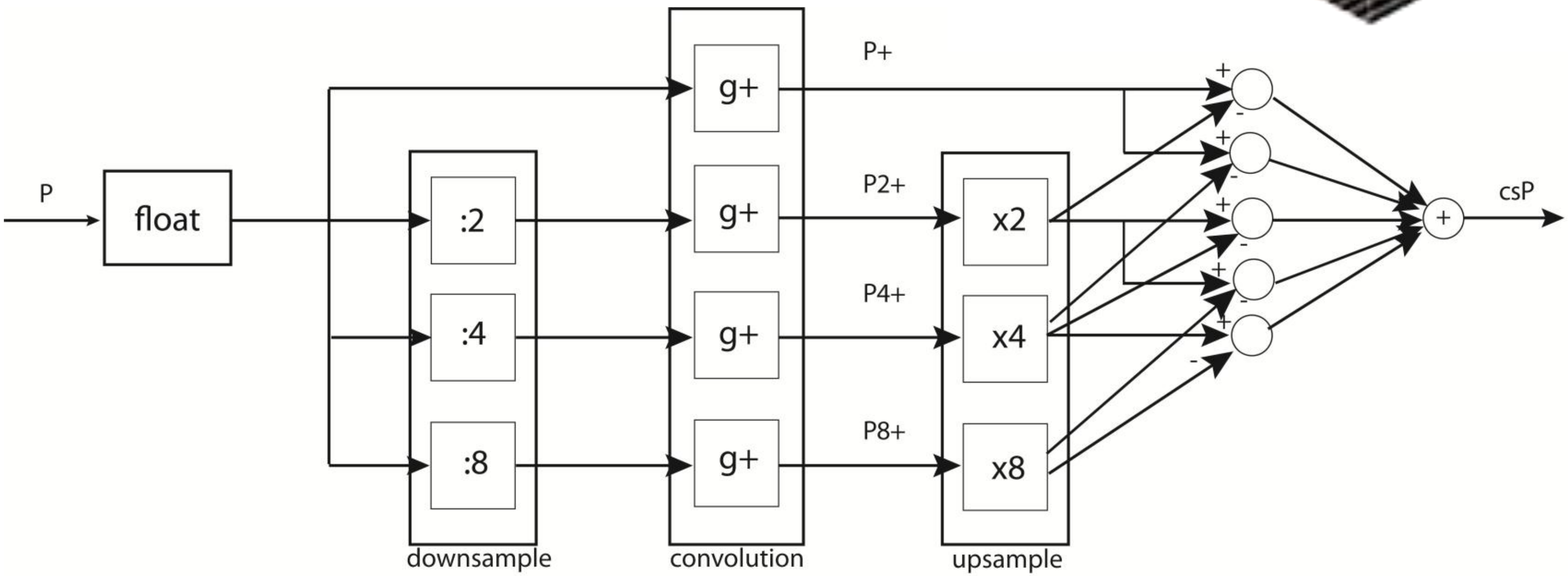
Reference: S. M. Smirnakis et al 1997, "Adaptation of retinal processing to image contrast and spatial scale



Preattentive System

- Approach 2: Pyramids of Convolutions

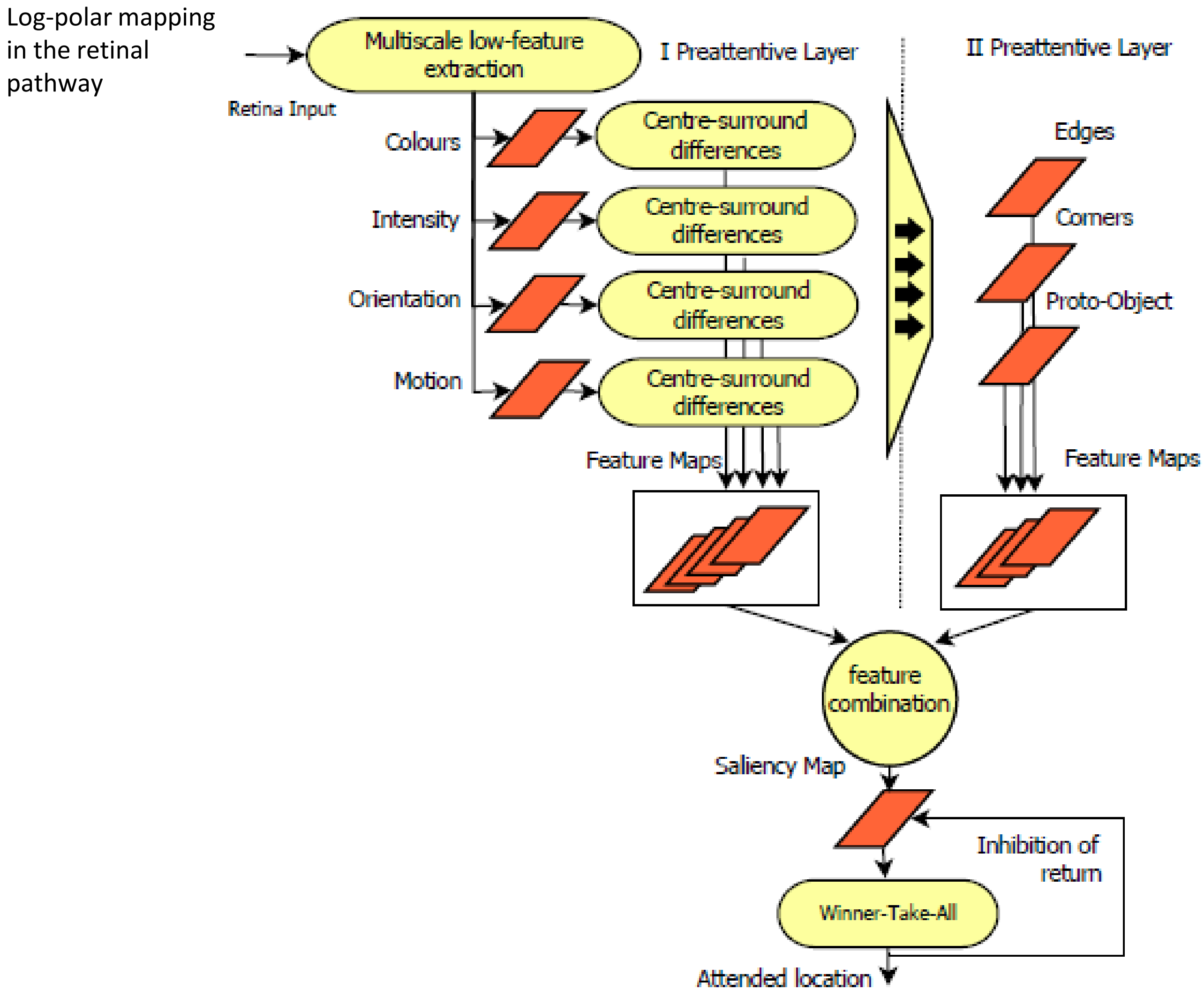




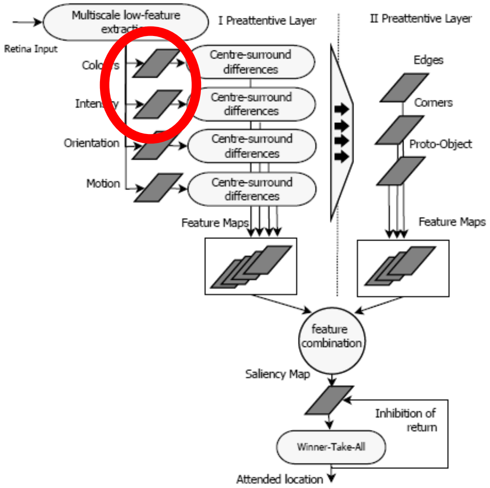
Reference : A. Ude et al, 2005

Preattentive System

- Feature Extraction in the Preattentive System



Preattentive System



– Intensity Uniqueness

- Gaussian pyramid from the intensity channel image I

– Chrominance Uniqueness

- Gaussian pyramids calculated over the both U and V channels
- After this, contributions of U and V are summed up
- process orthogonal U and V chrominance opponents

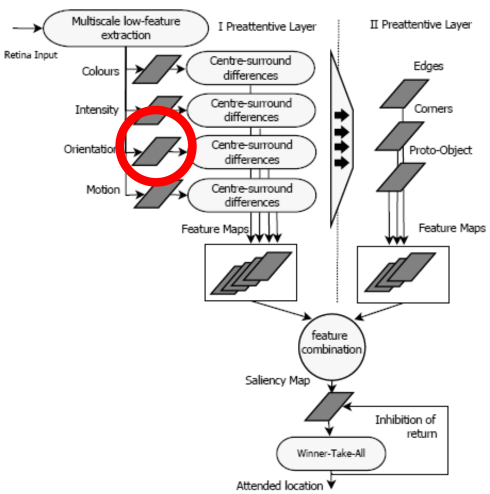
  U is approximately a yellow-magenta

  V approximates cyan-pink response

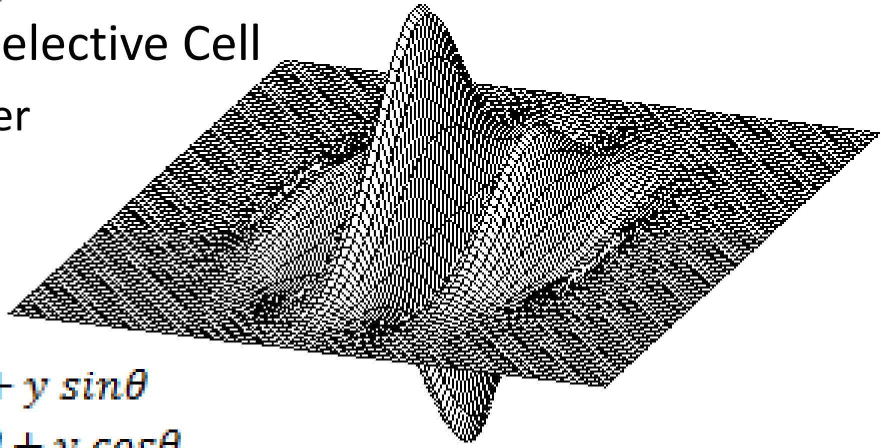
– Color Opponency Maps

- Derived using mapping from YUV to RGB (reduced computation cost)

Preattentive System



- Orientation Selective Cell
  - Gabor Filter



- Define:

$$\acute{x} = x\cos\theta + y\sin\theta$$
$$\acute{y} = -x\sin\theta + y\cos\theta$$

- Complex form

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{\acute{x}^2+\gamma^2\acute{y}^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{\acute{x}}{\lambda}+\psi\right)\right)$$

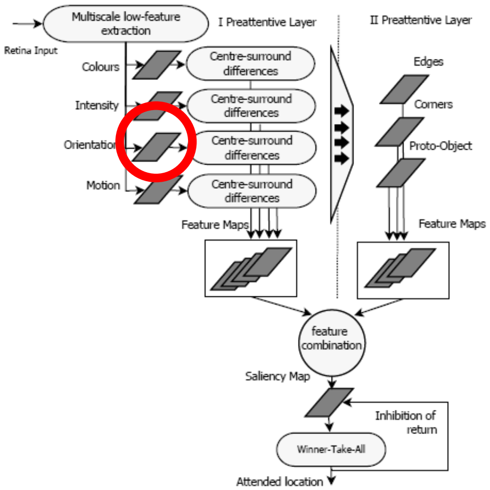
- Components

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{\acute{x}^2+\gamma^2\acute{y}^2}{2\sigma^2}\right)\cos\left(2\pi\frac{\acute{x}}{\lambda}+\psi\right)$$

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{\acute{x}^2+\gamma^2\acute{y}^2}{2\sigma^2}\right)\sin\left(2\pi\frac{\acute{x}}{\lambda}+\psi\right)$$

- λ : wavelength of the sinusoidal factor
- θ : orientation of Gabor function
- ψ : phase offset
- σ : sigma of the Gaussian envelope
- γ : spatial aspect ratio, ellipticity

Preattentive System



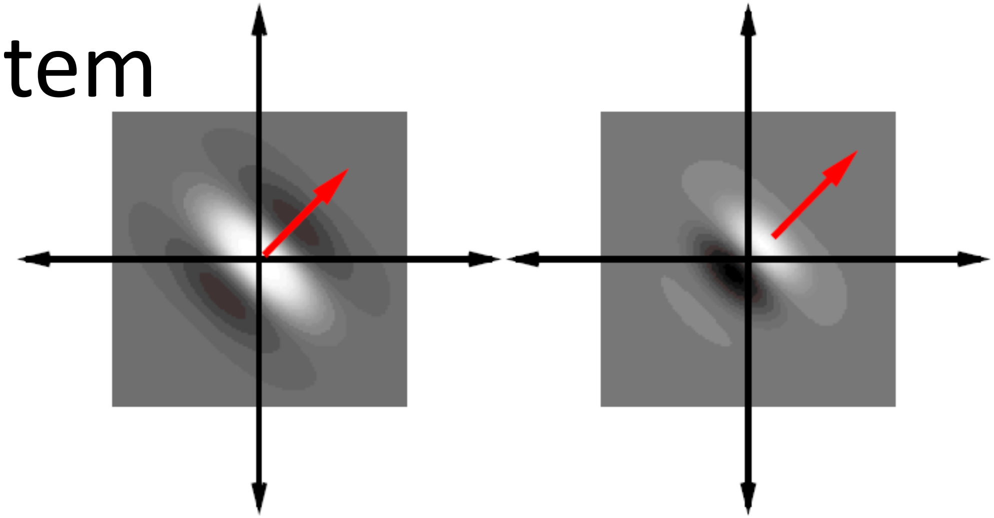
- We use

$$\gamma = 1$$

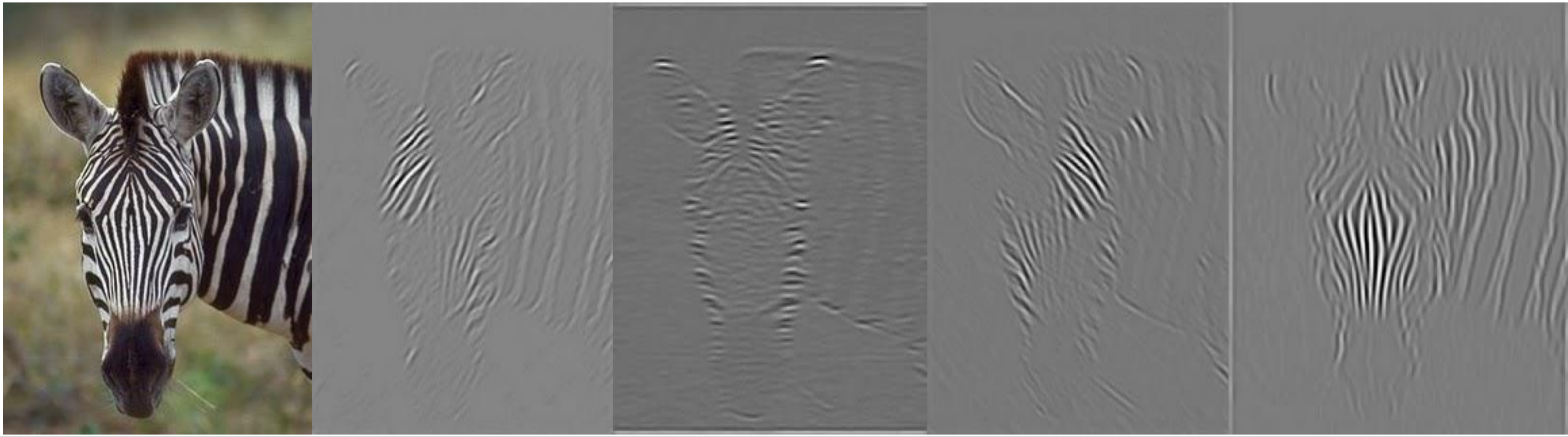
$$\delta = {}^{7}/_{3}$$

$$\lambda = 7$$

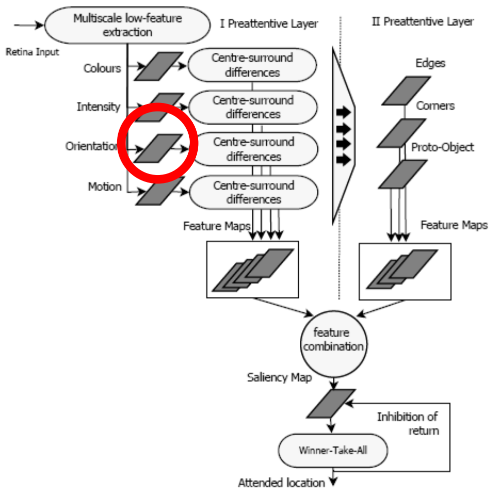
$$\psi \epsilon \left\{0, \frac{\pi}{2}\right\}$$



- With this rule we can obtain 5x5 asymmetrical kernel
- Behavior of Orientation Biased LGN cell

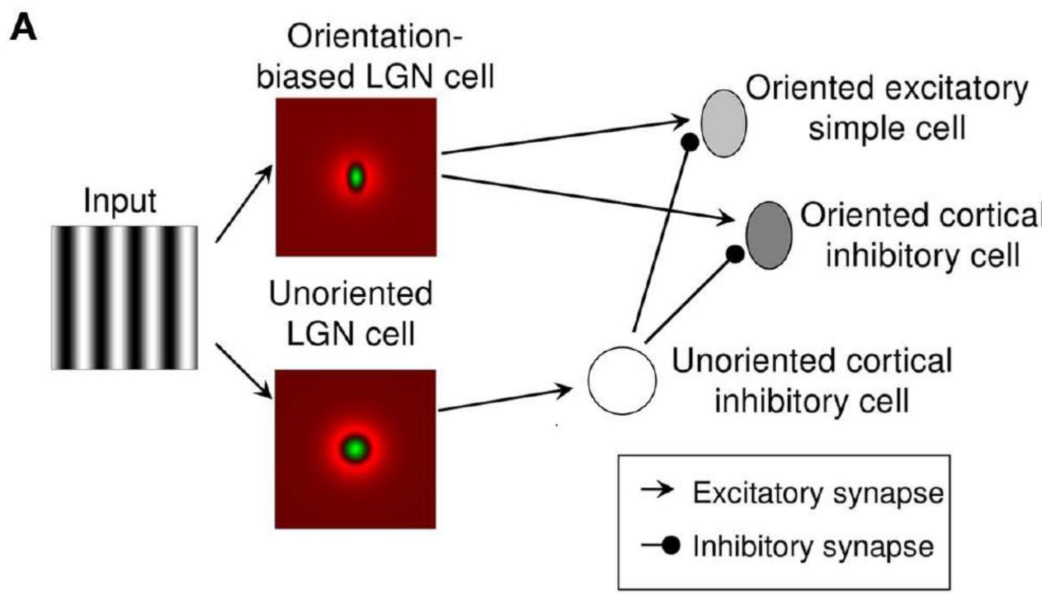


Preattentive System



– Behavior of oriented excitatory simple cell

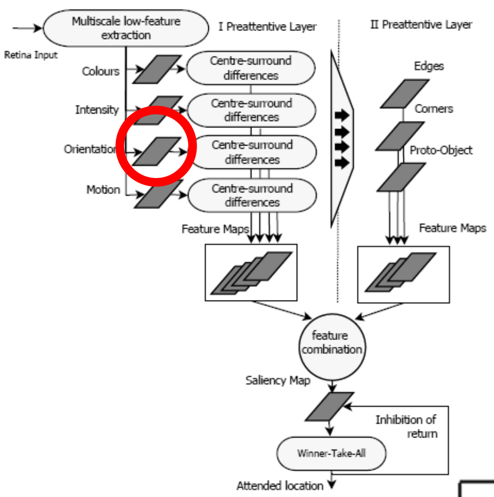
- Kuhlmann & Vidyasagar 2011

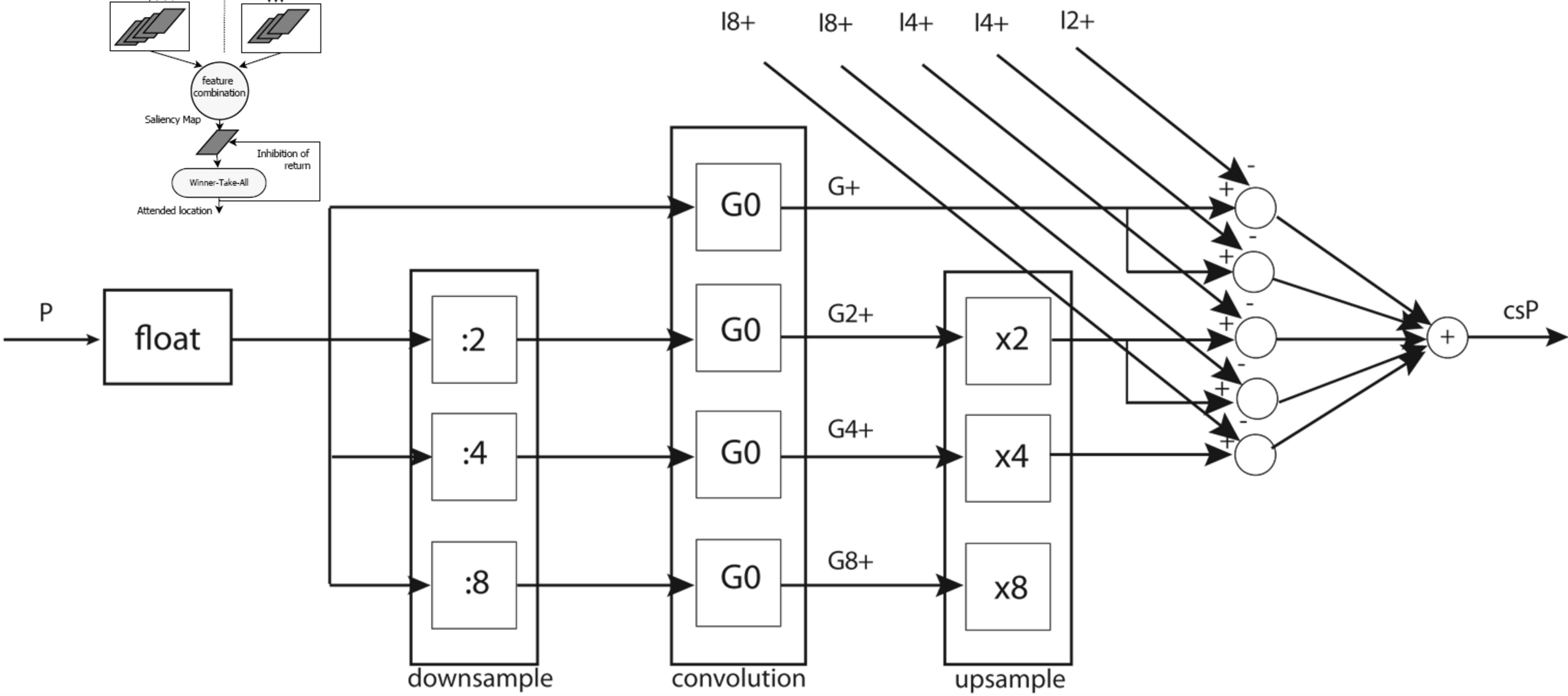


- anisotropic ON LGN exciting contribution and isotropic ON LGN inhibitory contribution are recurrently connected across orientation following Mexican hat weighting profile.

Preattentive System

– Implementation of the mechanism through pyramids of convolutions

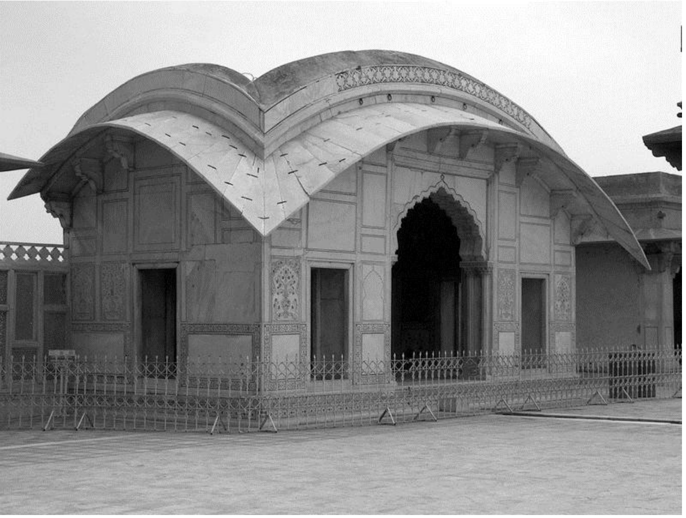


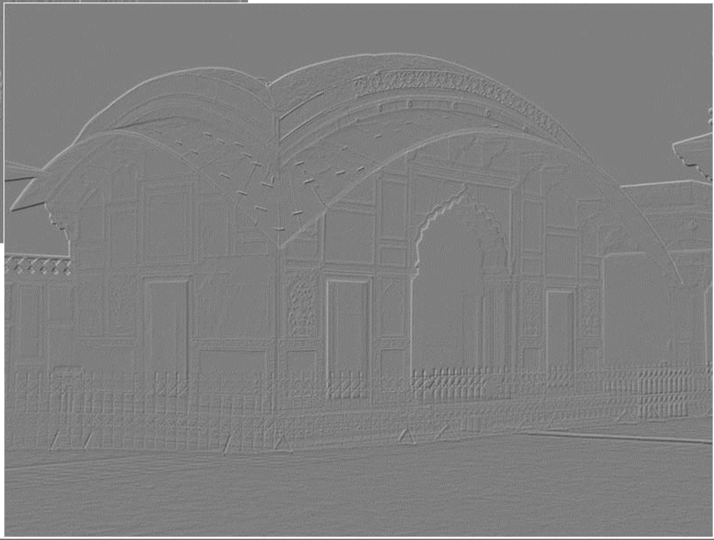


Preattentive System

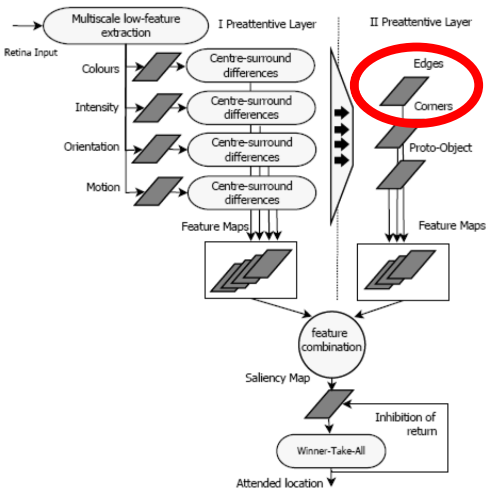
- Kirsch operator

$$\mathbf{g}^{(1)} = \begin{bmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, \mathbf{g}^{(2)} = \begin{bmatrix} +5 & +5 & -3 \\ +5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, \mathbf{g}^{(3)} = \begin{bmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{bmatrix}, \mathbf{g}^{(4)} = \begin{bmatrix} -3 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & +5 & -3 \end{bmatrix}$$





Complete Implementation of Visual Attention

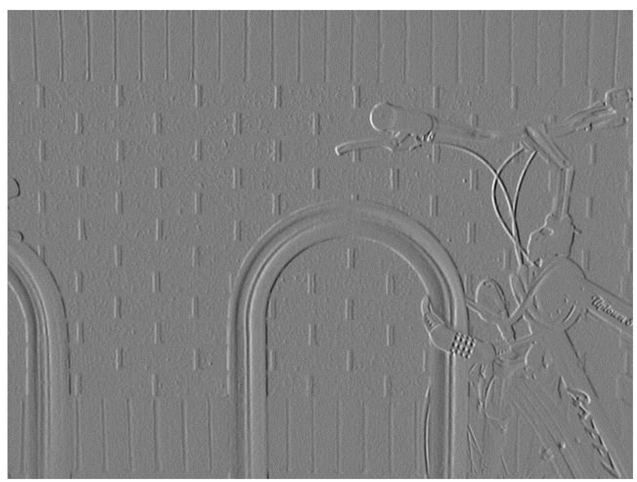


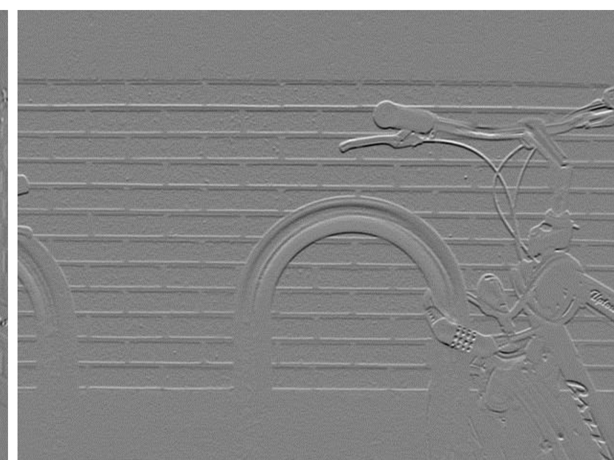
- Edges
  - Sobel operator , conolution in 2-dimesional space

$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \mathbf{A} \quad \text{and} \quad \mathbf{G}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * \mathbf{A}$$

  - Two filters are separable : Horizontal and Vertical

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$$





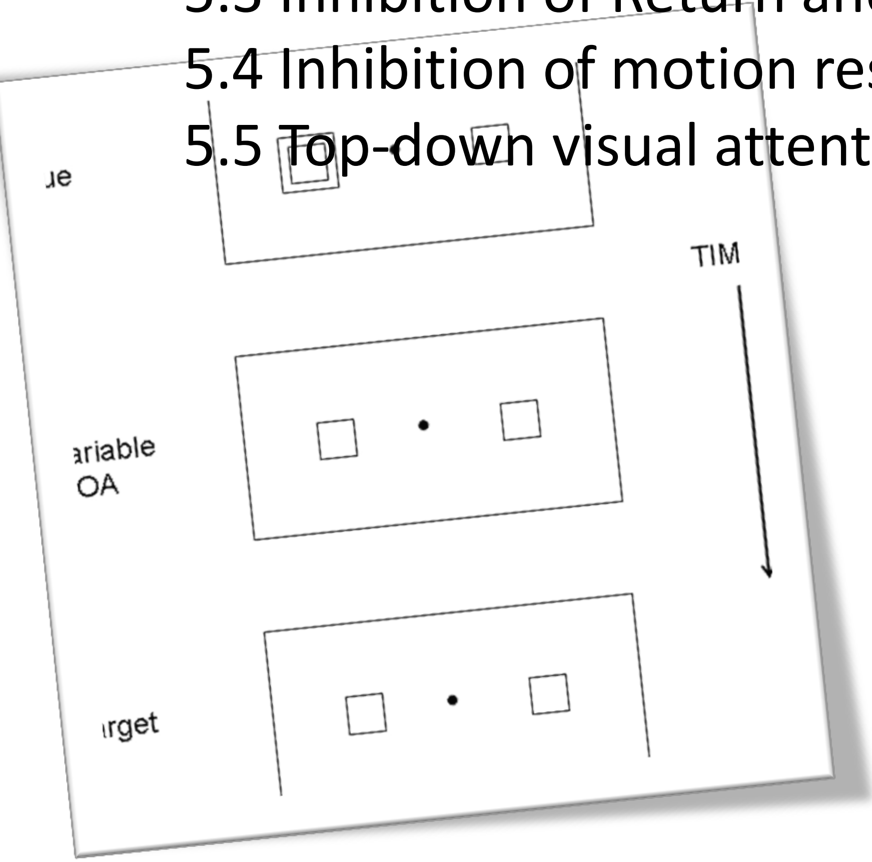
Preattentive System

- Reference:

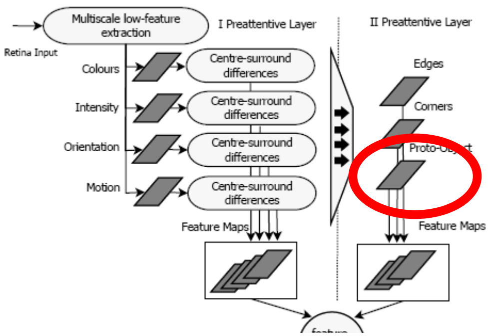
  - •Hurvich L.M. & Jameson D. 1957 An opponent-process theory of color vision
  - •Itti & Koch ,1998 A model of early visual processing
  - •Itti & Koch, 2001 Computational Modeling of Visual Attention
  - • Koch & Ullman, 1985 Shifts in selective visual attention towards the underlying neural circuitry
  - • Treisman, A., 1986. Features and objects in visual processing, Scientific American, 254, No. 11, 114-125
  - • Kuhlmann L and Vidyasagar TR (2011) A Computational study of how orientation bias in the lateral geniculate nucleus can give rise to orientation selectivity in primary visual cortex. *Front. Syst. Neurosci.* **5**:81
  - •A. Ude, V. W.-H. (2005). Distributed visual attention on a humanoid robot. *Proc. IEEE-RAS/RSJ Int. Conf. on Humanoid Robots*, (pp. 381-386). Tsukuba

Complete Implementation of Visual Attention

5.1 Proto objects
5.2 Saliency Map
5.3 Inhibition of Return and Habituation
5.4 Inhibition of motion response
5.5 Top-down visual attention

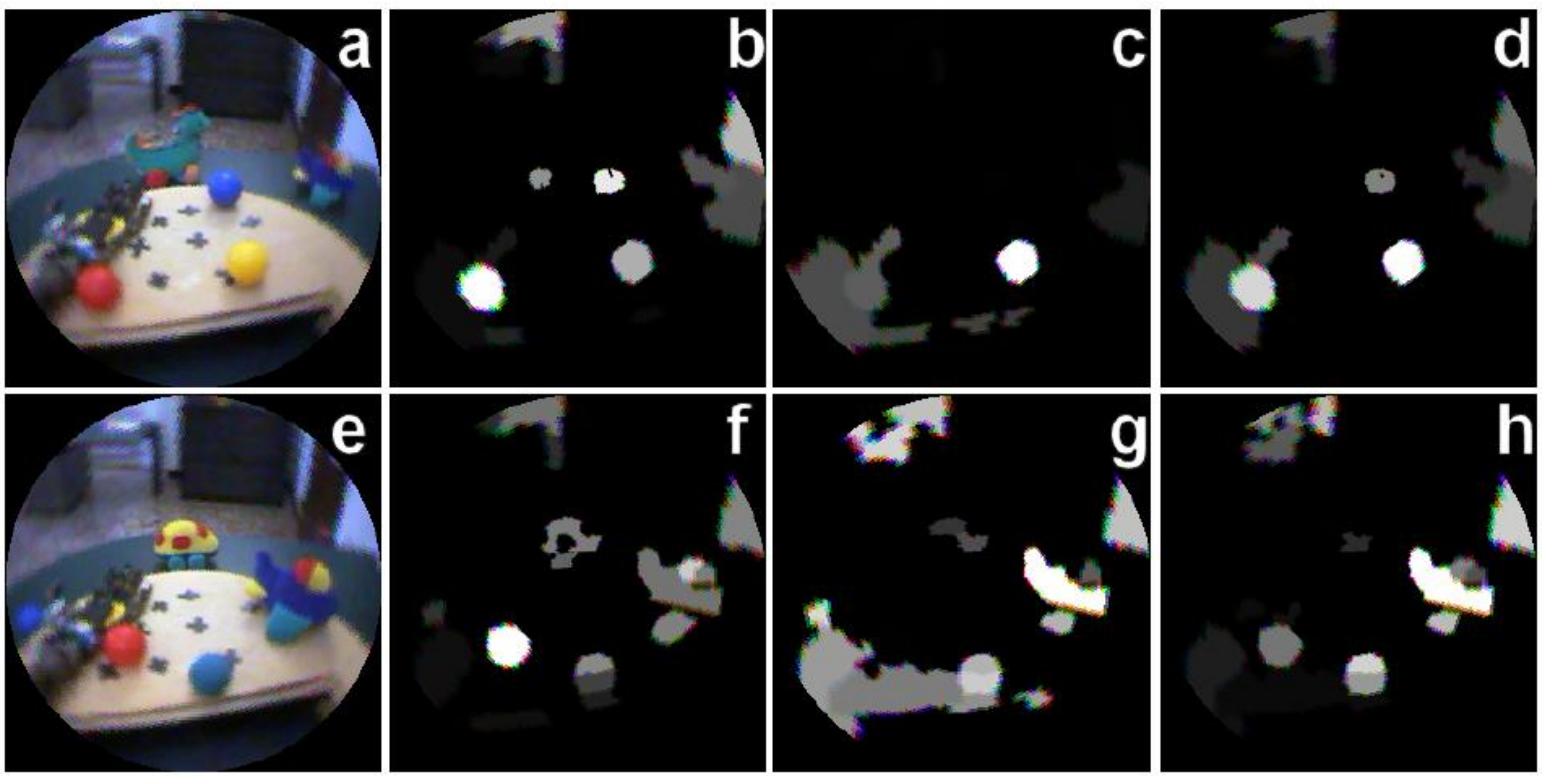


Complete Implementation of Visual Attention

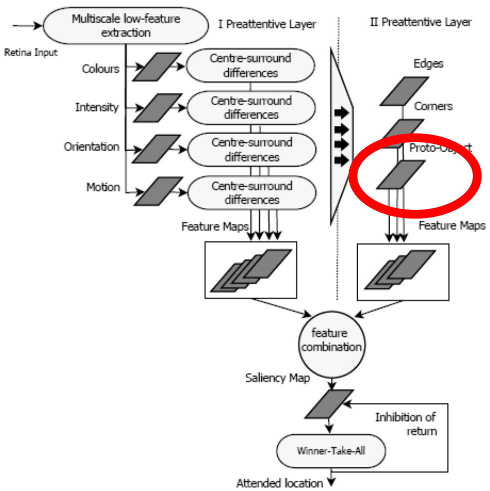


- Proto Objects
  - •Reference : Orabona et al 2005, Walter & Koch 2006, Rensink 2000

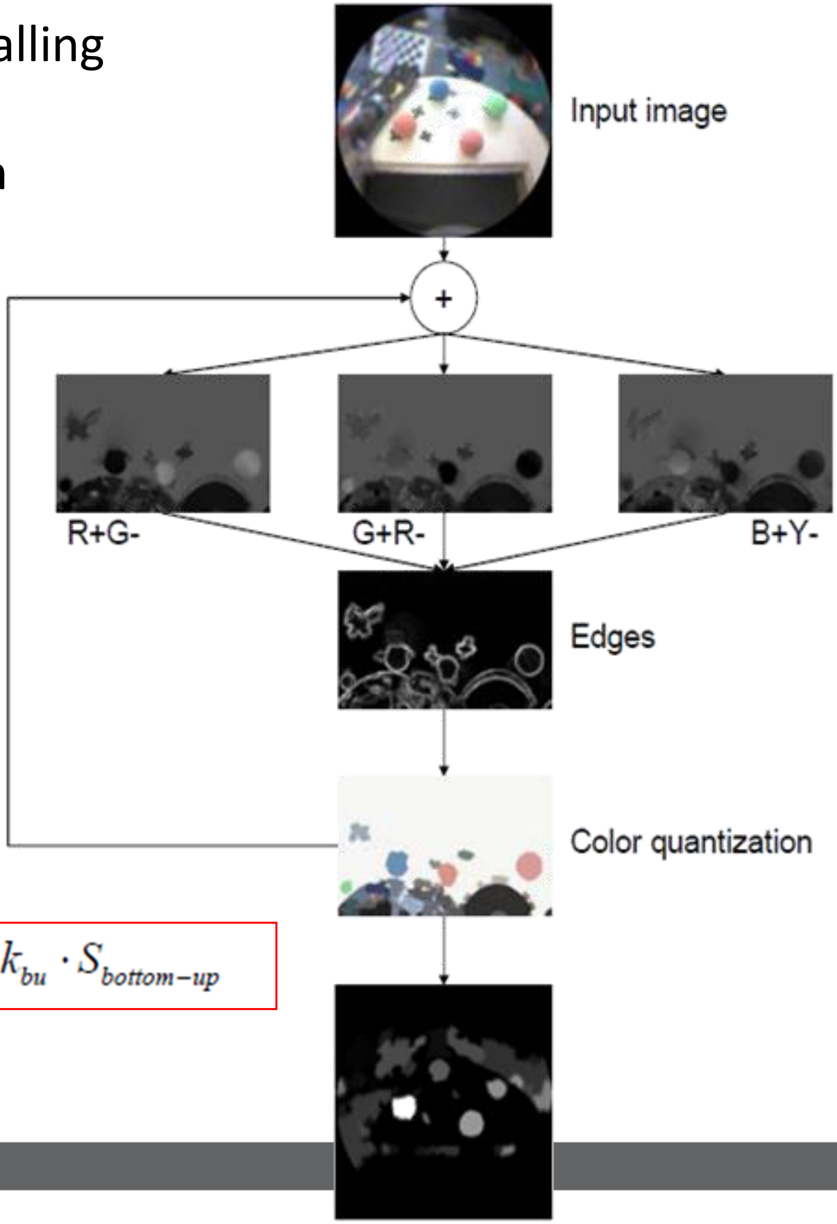
Proto-object: volatile units of visual information that can eventually be bound into coherent and stable objects.



Complete Implementation of Visual Attention



1. Watershed (rain falling algorithm)
2. Color Quantization
3. Feature Map



$$S_{BU} = \sqrt{\Delta RG^2 + \Delta GR^2 + \Delta BY^2}$$

$$\Delta RG = (R + G -)_{blob} - (R + G -)_{surround}$$

$$\Delta GR = (G + R -)_{blob} - (G + R -)_{surround}$$

$$\Delta BY = (B + Y -)_{blob} - (B + Y -)_{surround}$$

$$S_{TD} = \sqrt{\Delta RG^2 + \Delta GR^2 + \Delta BY^2}$$

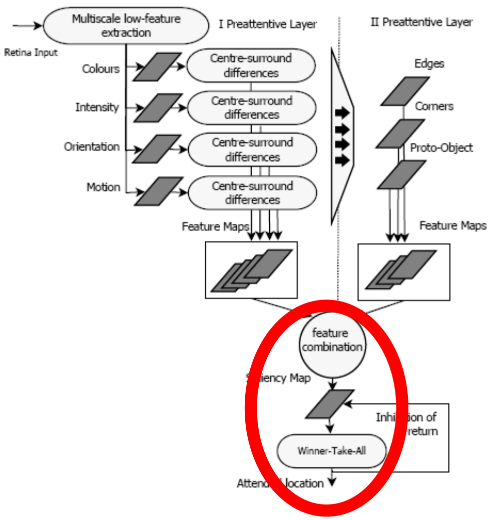
$$\Delta RG = (R + G -)_{blob} - (R + G -)_{target}$$

$$\Delta GR = (G + R -)_{blob} - (G + R -)_{target}$$

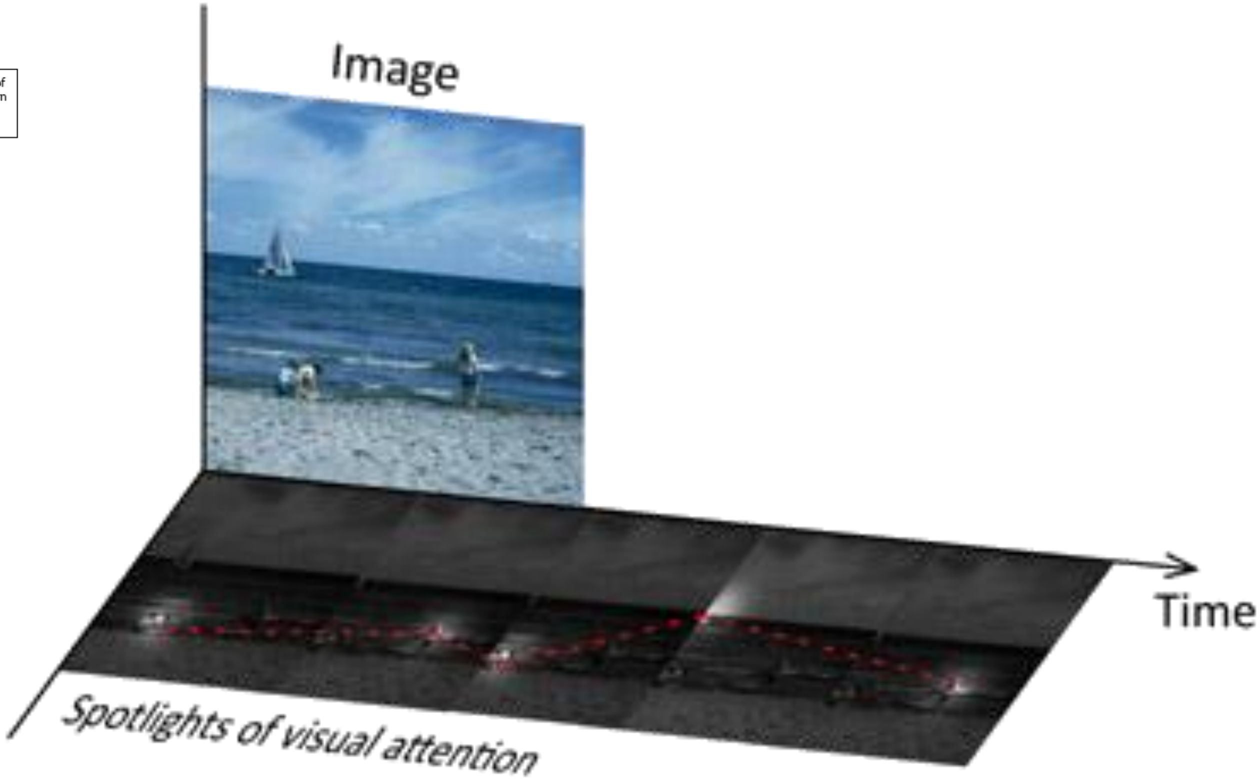
$$\Delta BY = (B + Y -)_{blob} - (B + Y -)_{target}$$

$$\boxed{S = k_{td} \cdot S_{top-down} + k_{bu} \cdot S_{bottom-up}}$$

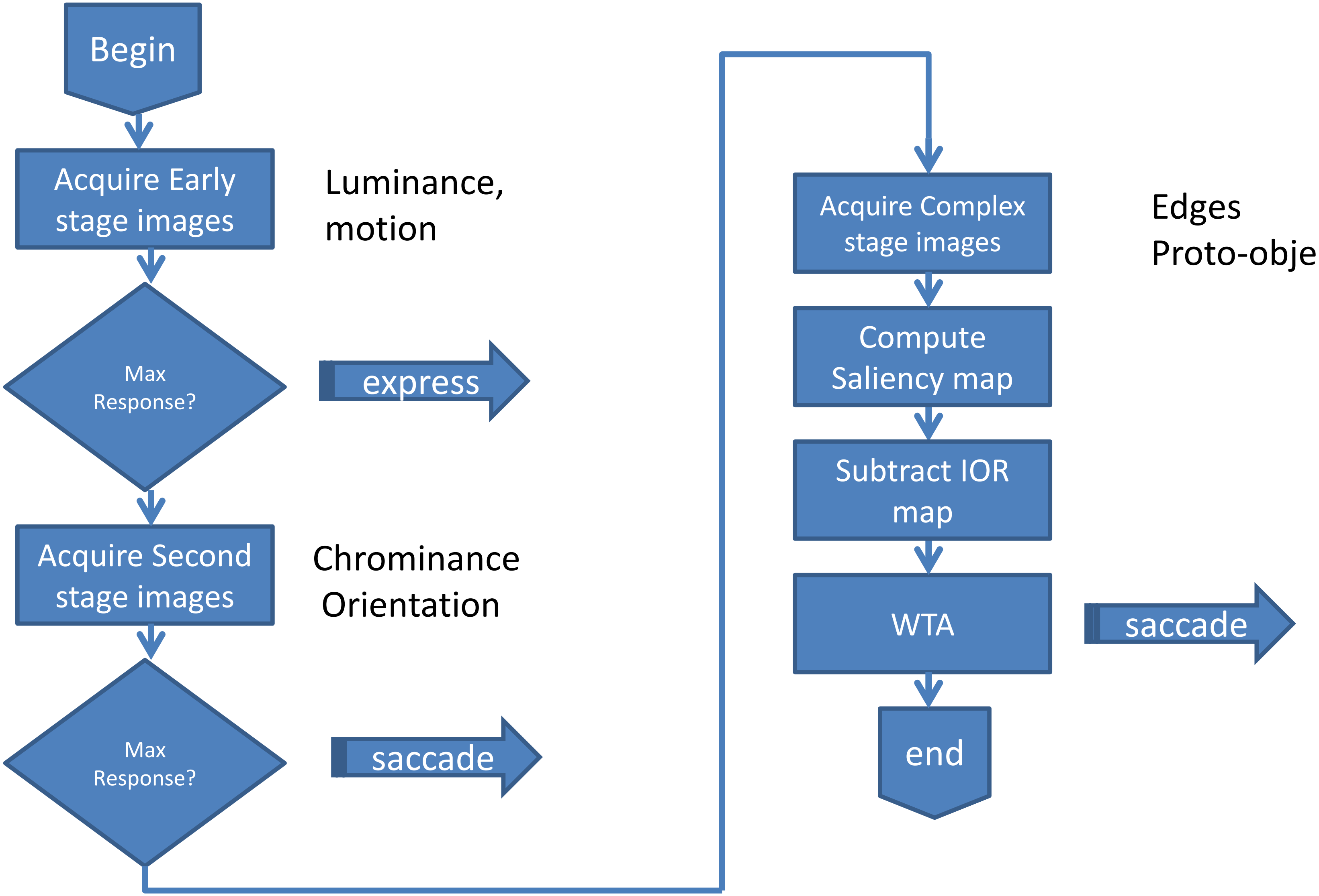
Complete Implementation of Visual Attention



Selective attention
- Singular Saliency Map
- WTA
- IOR



Complete Implementation of Visual Attention



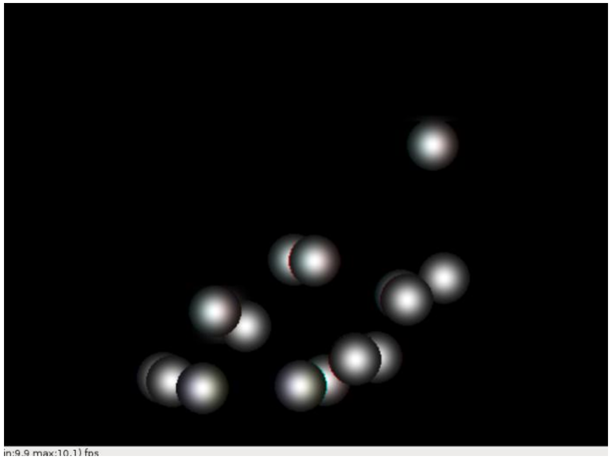
Complete Implementation of Visual Attention

- Inhibition of Return and Habituation
  - Reference: Posner et al 1985, Rafal et al 1989, Klein 2000.

  - Interpretation:
    - different processes, covert vs. overt attention
    - Inhibition of Return as reference eye movements

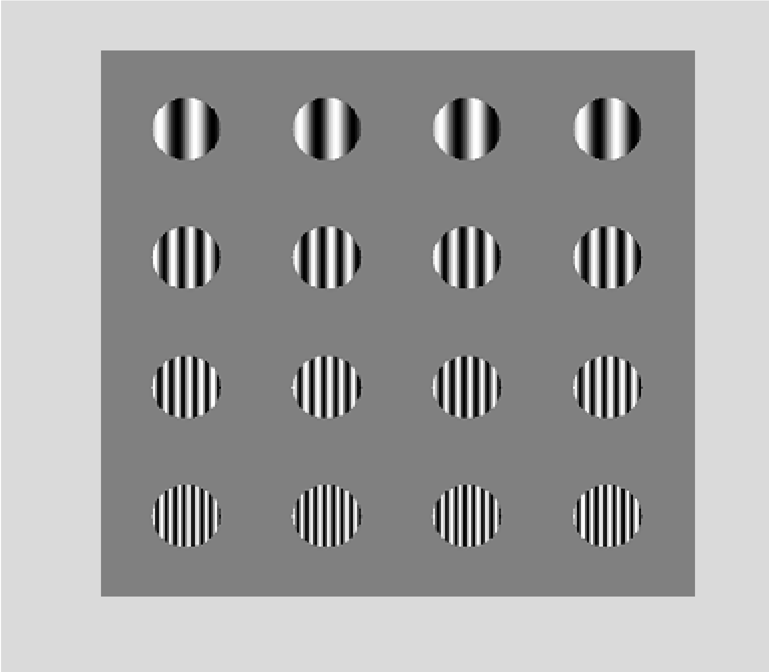
  - Implementation: Mosaic

  - Experiment



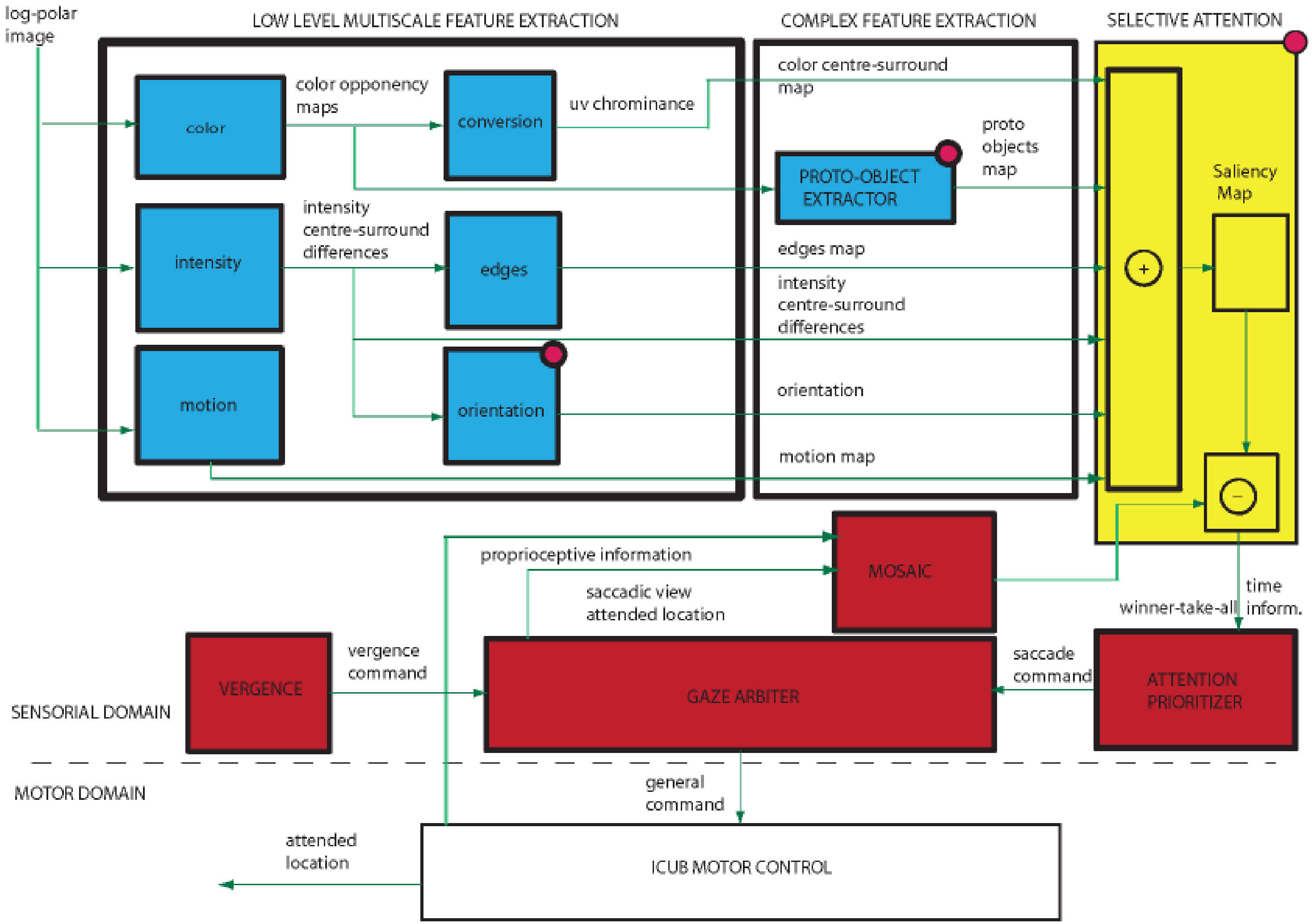
Complete Implementation of Visual Attention

- Inhibition of Motion Response
  - Reference: Burr et al 1994 Selective suppression of Magnocellular visual pathway during saccadic eye movements
  - Pattern of high spatial frequency(modulated in color): not suppression but enhancing
  - But only for color-blind magnocellular low spatial sense of motion

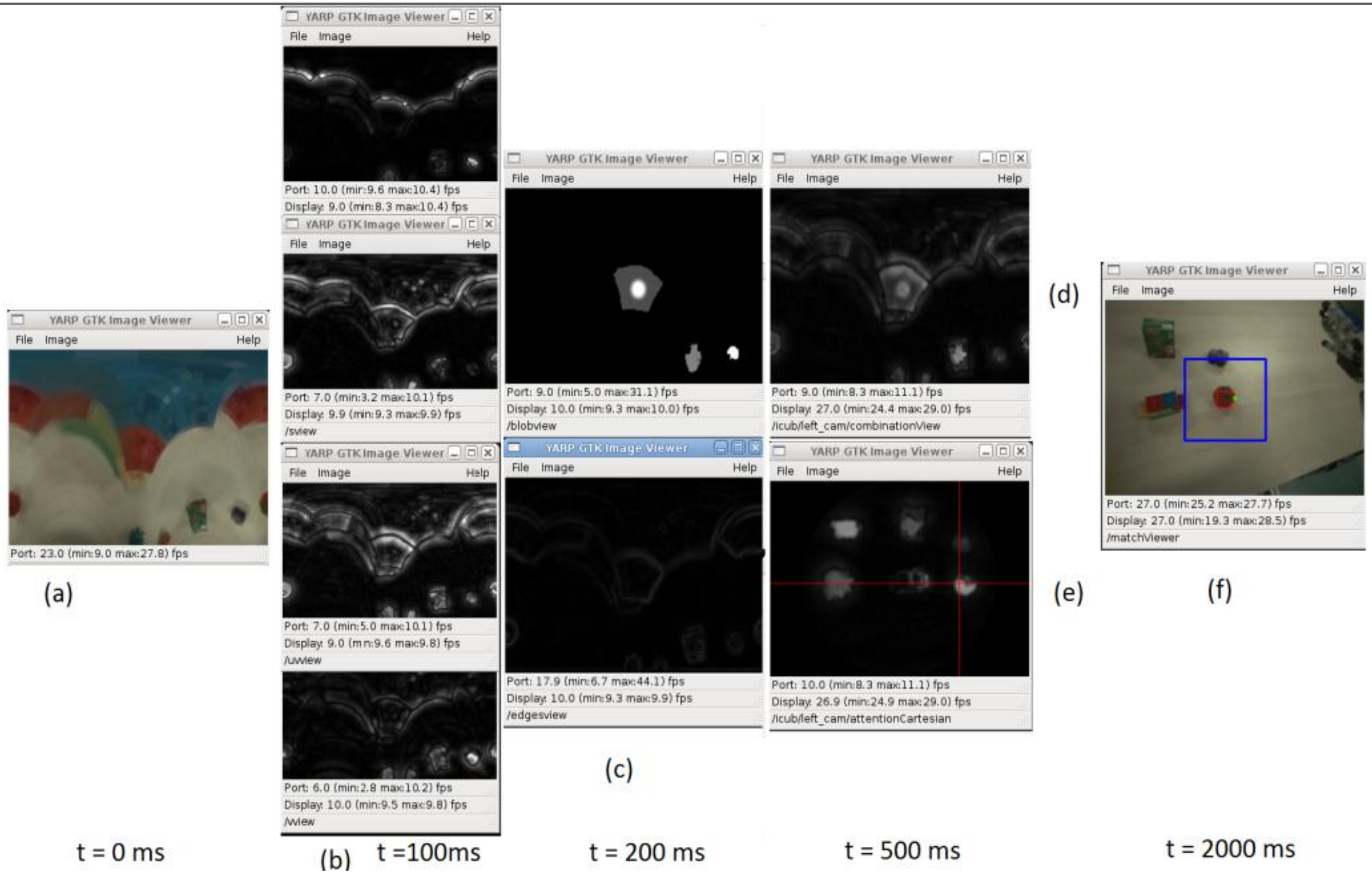


Complete Implementation of Visual Attention

- Top-Down visual attention



Complete Implementation of Visual Attention



Complete Implementation of Visual Attention

- Reference

  - Burr, D. C., Morrone, M. C. & Ross, J. Selective suppression of the magnocellular visual pathway during saccadic eye movements.*Nature* 371, 511-513 (1994).
  - Posner, M.I. *et al. (1985) Inhibition of return: neural basis and function. Cognit. Neuropsychol. 2, 211–228*
  - Rafal, R.D. *et al. (1989) Saccade preparation inhibits reorienting to* recently attended locations. *J. Exp. Psychol. Hum. Percept. Perform. 15,* 673–685
  - F.Orabona, G. G. (2005). Object-based visual attention: a model for a behaving robot. Washington, DC, USA: IEEE Computer Society
  - D. Walther, C. K. (2006). Modeling attention to salient proto-objects. *Neural Networks , 19*, pp. 1395-1407.
  - Resink, R. (2000). The Dynamic Representation of scenes. *Visual Cognition , 7* (1/2/3), 17-42.

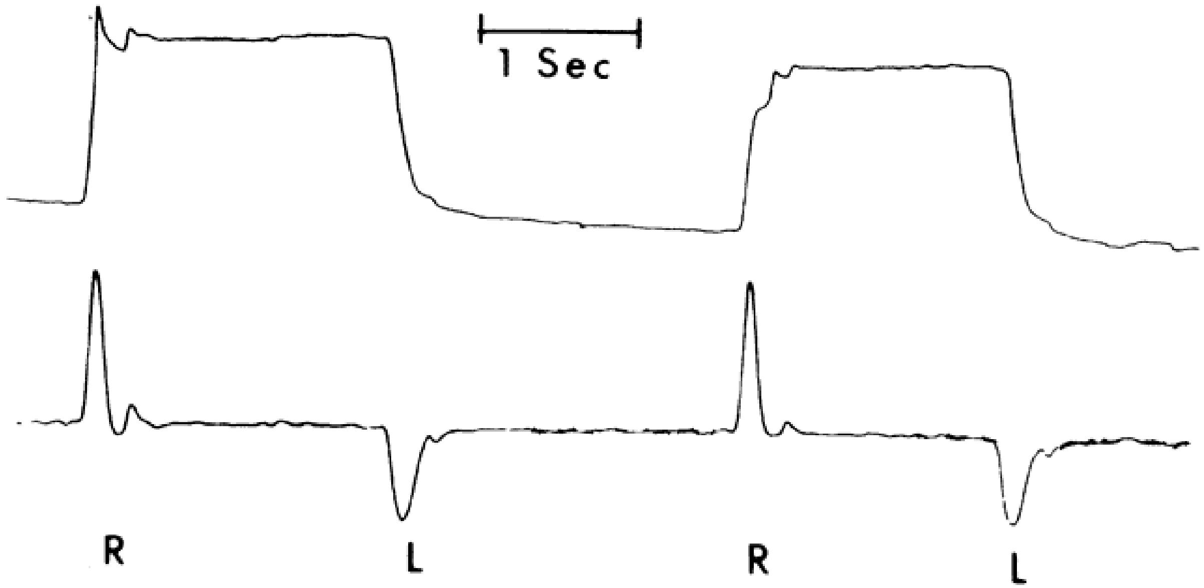
Appendix: Oculomotor commands

- Saccade
- Vergence
- Smooth Pursuit
- Microsaccade
- Tremor

Appendix: Oculomotor commands

- Saccade
  - Monocular

  left drive eye



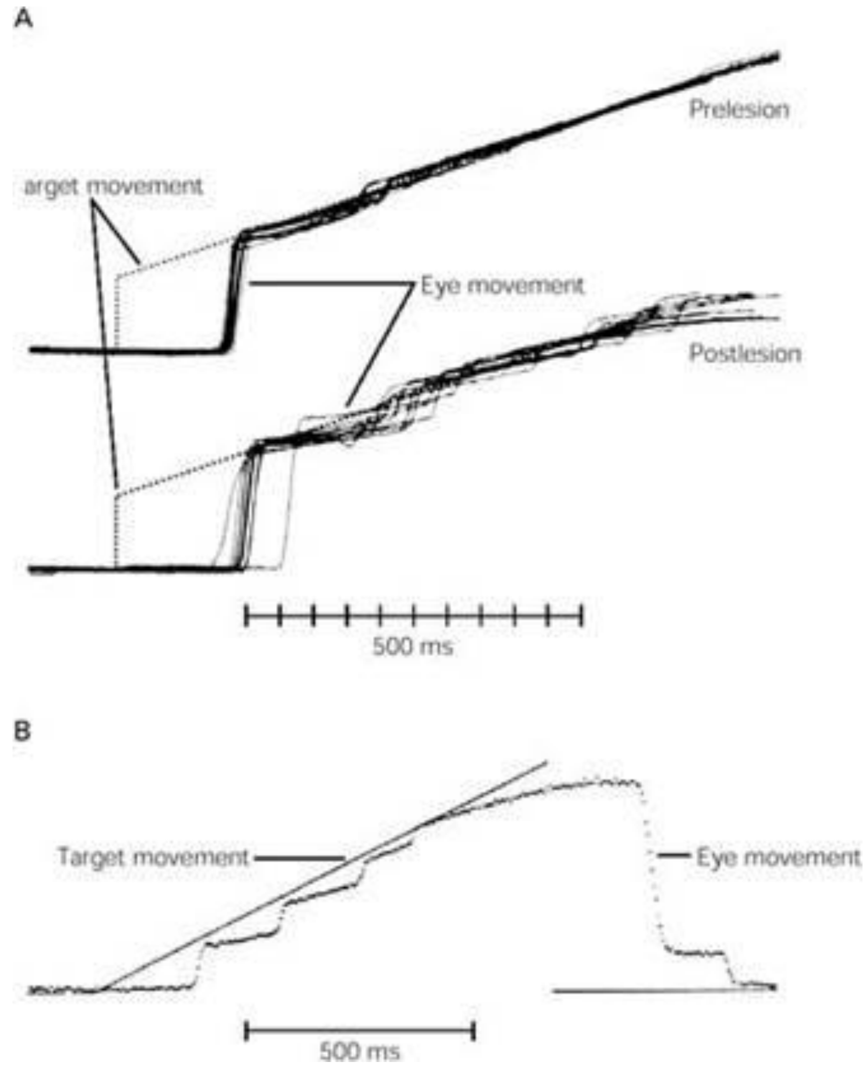
- Vergence
  - Stereo information
  - Correlation histogram

Appendix: Oculomotor commands

- Smooth Pursuit

Impaired monkey

Impaired patient



Appendix: Oculomotor commands

- Microsaccade & Tremor