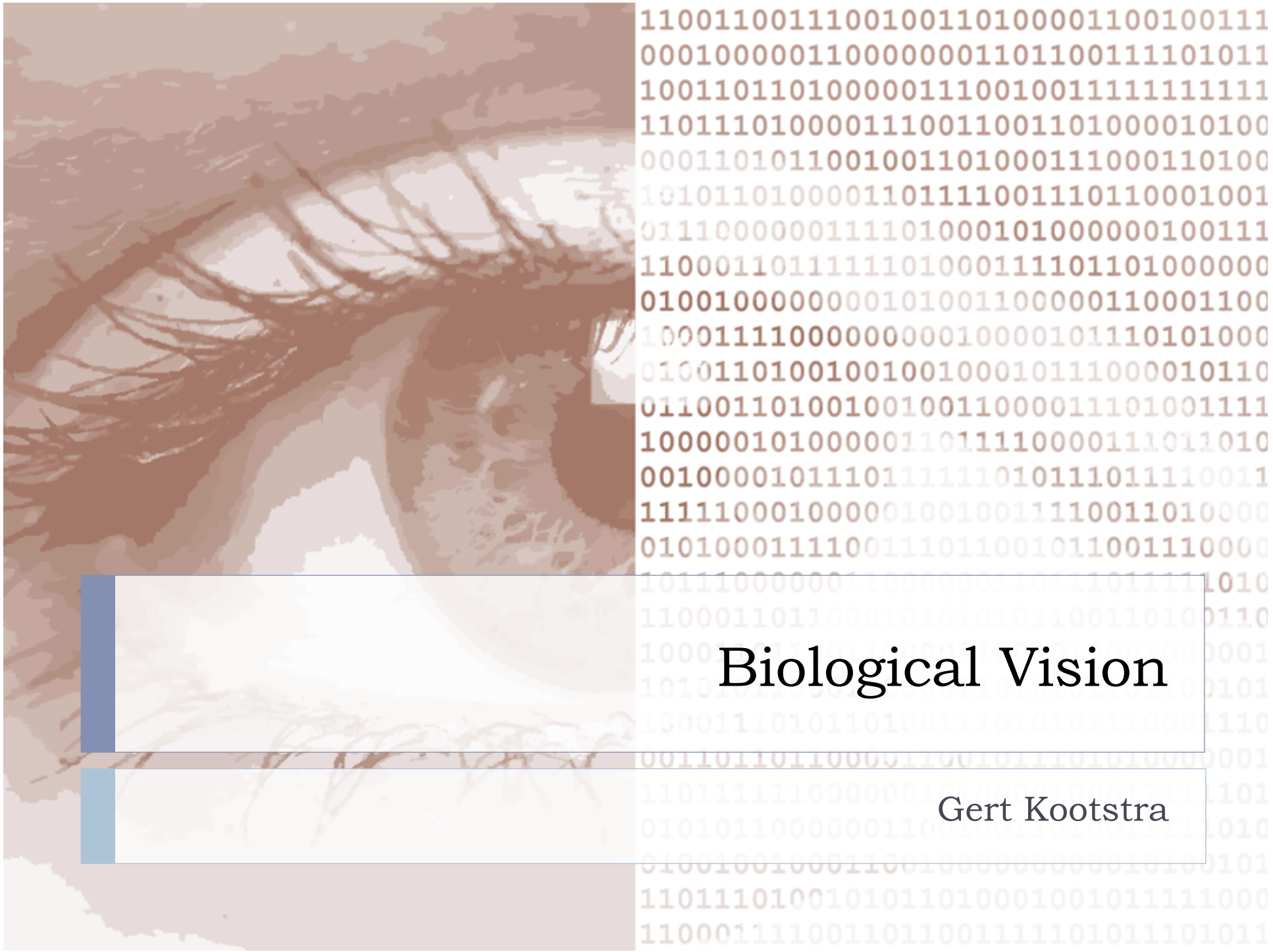


Biological Vision

Gert Kootstra

This lecture

- Overview of vision in biological systems
  - The human eye
  - Neural foundations
  - Some visual illusions (habituation, context)
  - Visual attention
  - Active vision

- *What we can learn from biological vision for machine vision?*

Biological vision

Biological Vision

- The purpose of vision is for the survival and reproduction of the species
  - The visual system is optimized for this
- Example
  - We are good in recognizing other human beings, or estimating the speed of a ball to catch it
  - We are bad in reading QR codes

Biological vs Machine Vision

- Biological vision
  - Billions of years of evolution
  - Very well adapted to our needs
- Machine vision
  - 3-4 decades of research
  - Many more challenges than solutions
- Insights in biological vision can give inspiration
  - Neurophysiology (Studying the brain on neural level)
  - Psychophysiology (Linking behavior to internal proc.)

Sensation and perception

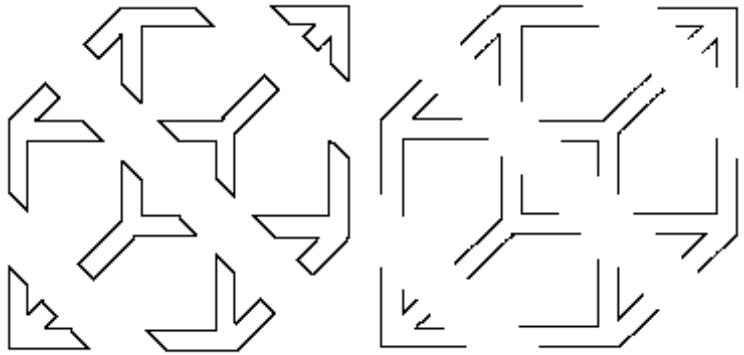
- Sensation
  - The measurement of the outside world
- Perception
  - Interpretation of this information
  - Direct perception
  - "Intelligent" perception

Direct perception

- All the information to make the percept is directly available in the stimulus
- Invariant properties of the stimulus
  - Perceiving entities despite transformations like changes in light, pose, etc
- Affordance
  - Actions that can be performed on the percepts

Intelligent perception

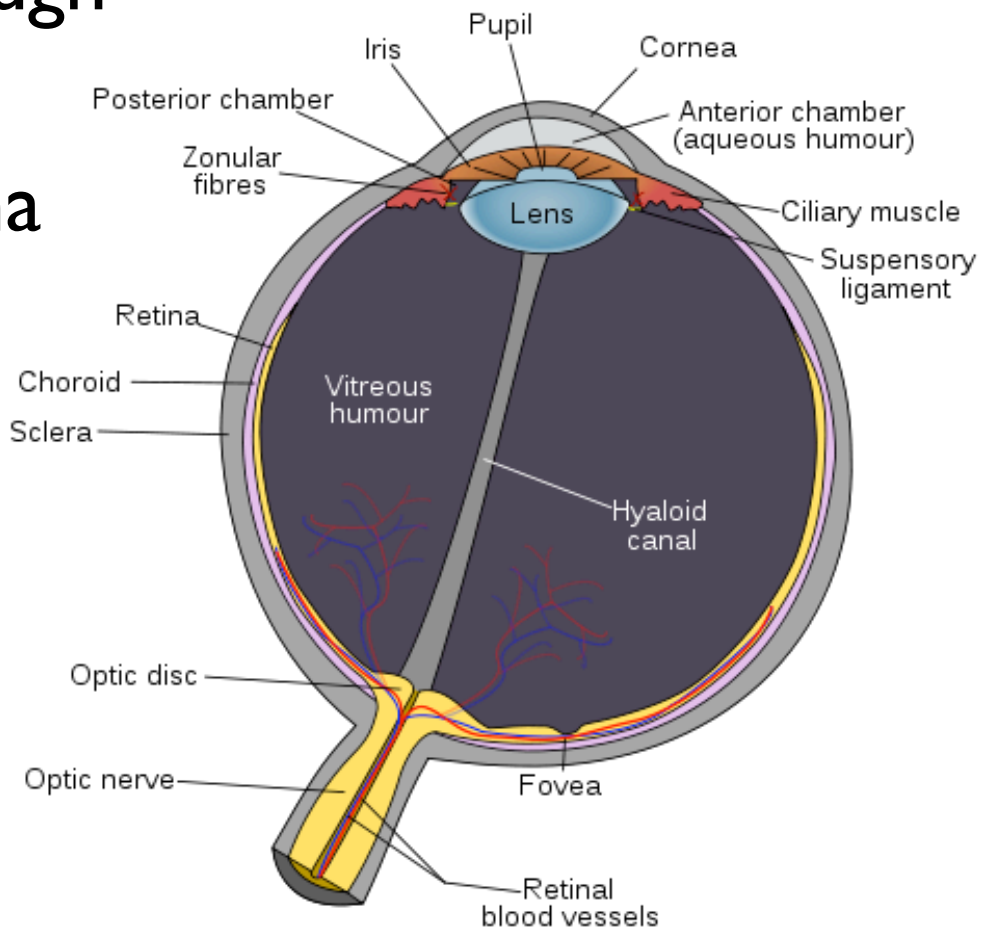
- Perception as reasoning
- The percept might depend on
  - Knowledge, past experiences, expectations
- Direct vs "Intelligent" perception



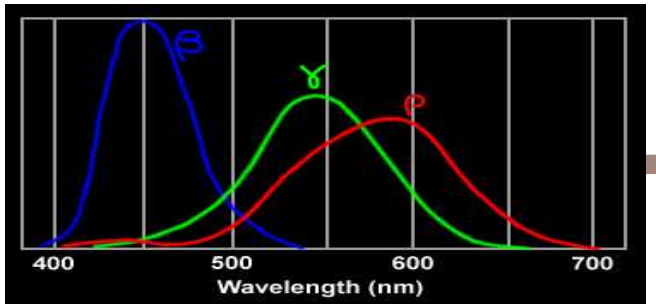
  - Although the stimulus is almost the same in b), we see a cube due to past experiences

The human eye

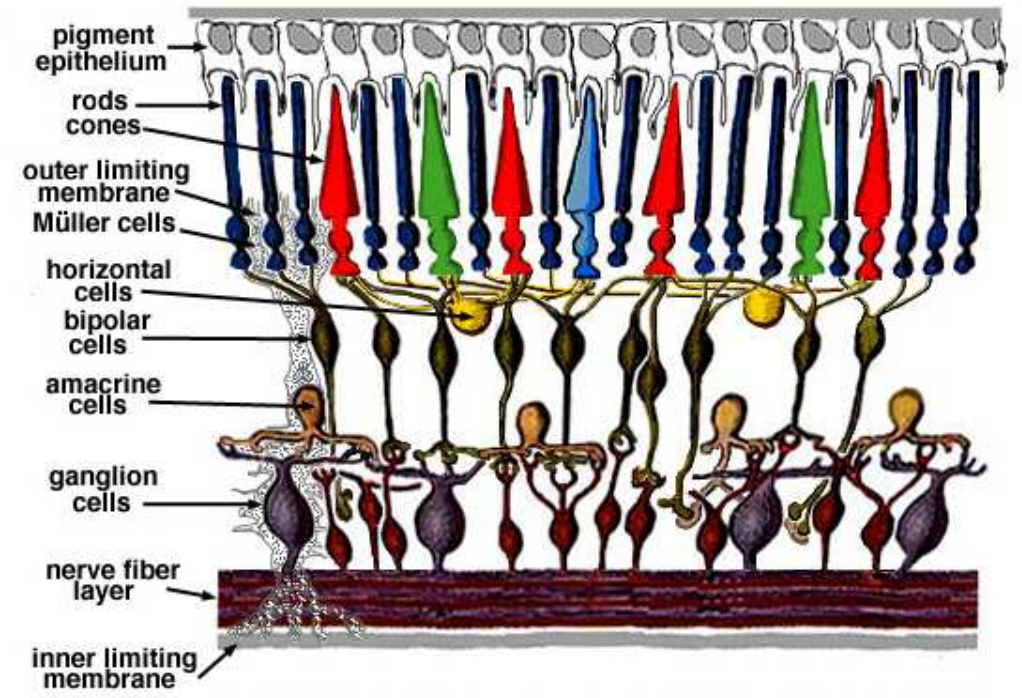
- Light comes in through the pupil
- Is focus on the retina by the lens
- Retina contains photoreceptors to transform photons into electric potentials



Photoreceptors

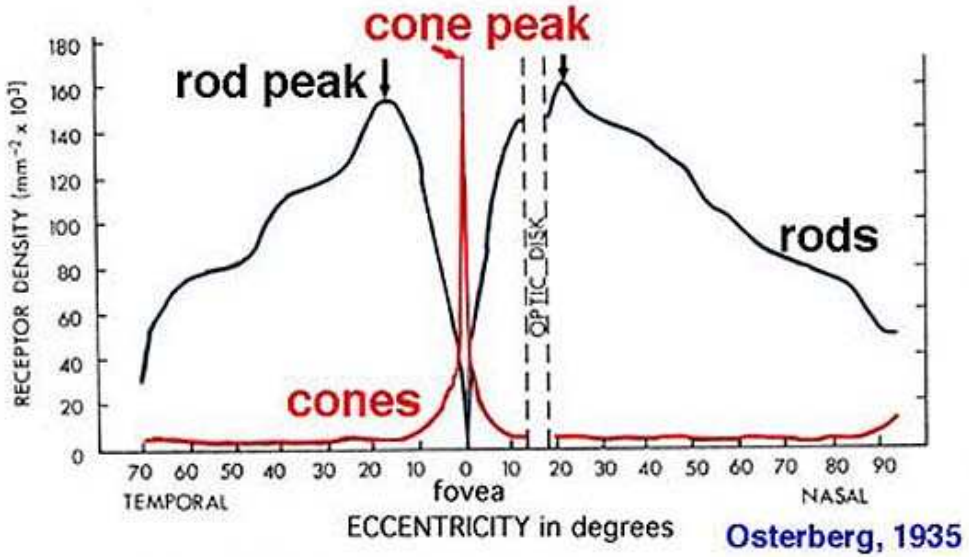


- Cones
  - Come in three different wave lengths: 'red', 'green', 'blue' (visual spectrum: 400 – 700 nm)
  - For color perception
- Rods
  - Very sensitive
  - For brightness and motion perception
  - 'Dark' vision



The retina

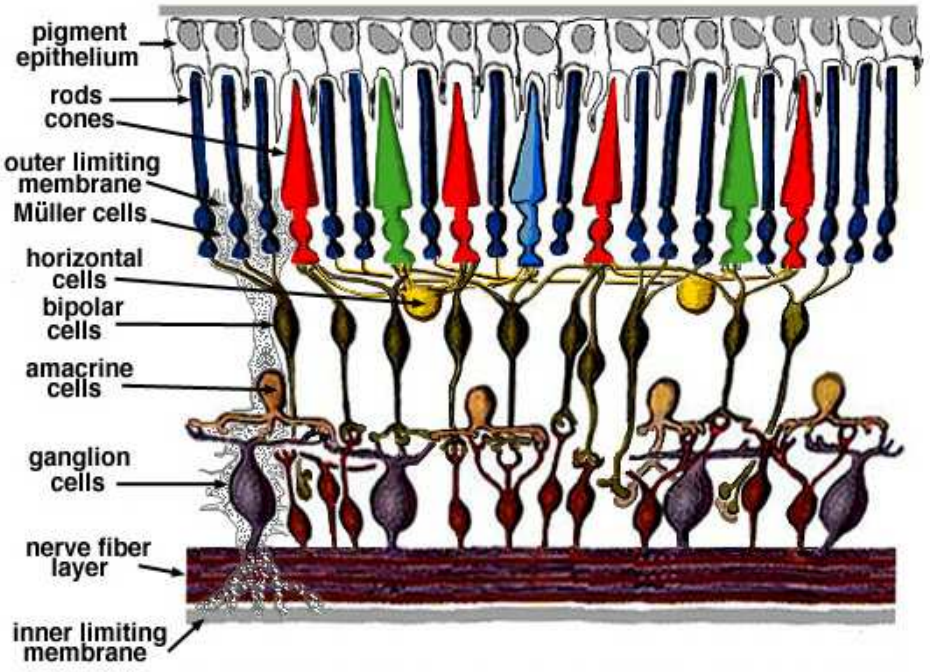
- The distribution of receptors on the retina
- Cones in the fovea (2°)
- Rods mainly in the parafovea and periphery



- Fovea is used for high-acuity vision
- Periphery for motion detection

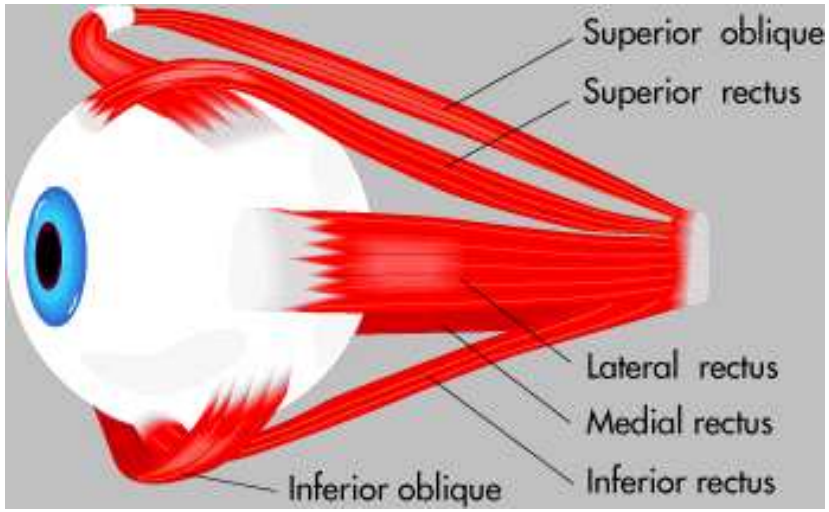
Retinal cells in the eye

- Horizontal cells
  - Lateral inhibition
- Bipolar cells
  - Center-surround
- Ganglion cells
  - Parvo: color contrast
  - Magno: brightness cont., for low-contrast stimuli, fast => motion
  - Konio: involved in color vision
  - Connection to the rest of the brain



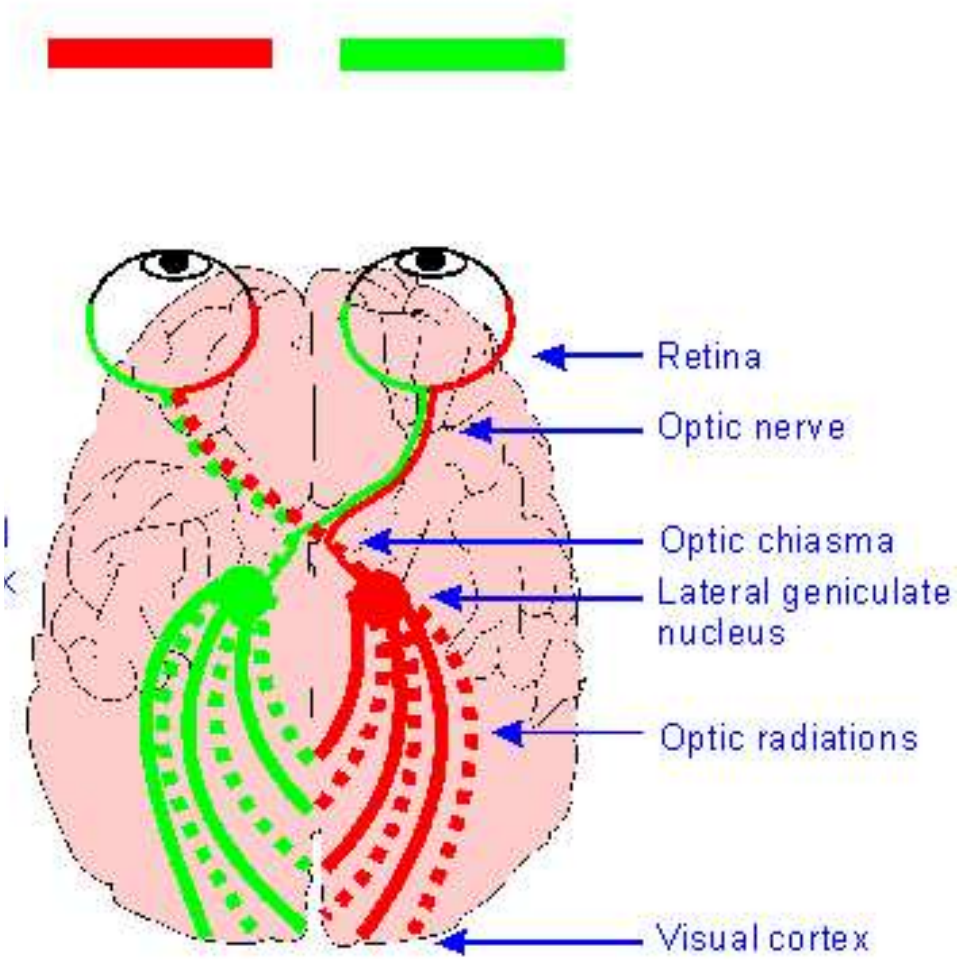
The eye is an active sensor

- Six extrinsic muscles
  - Eye movements
- Pupil diameter
- Lens curvature to focus
- Eye movements
  - Saccades: rapid gaze shifts (800 degrees/sec)
  - Fixation: brief stable focus
  - 4-5 eye movements per second



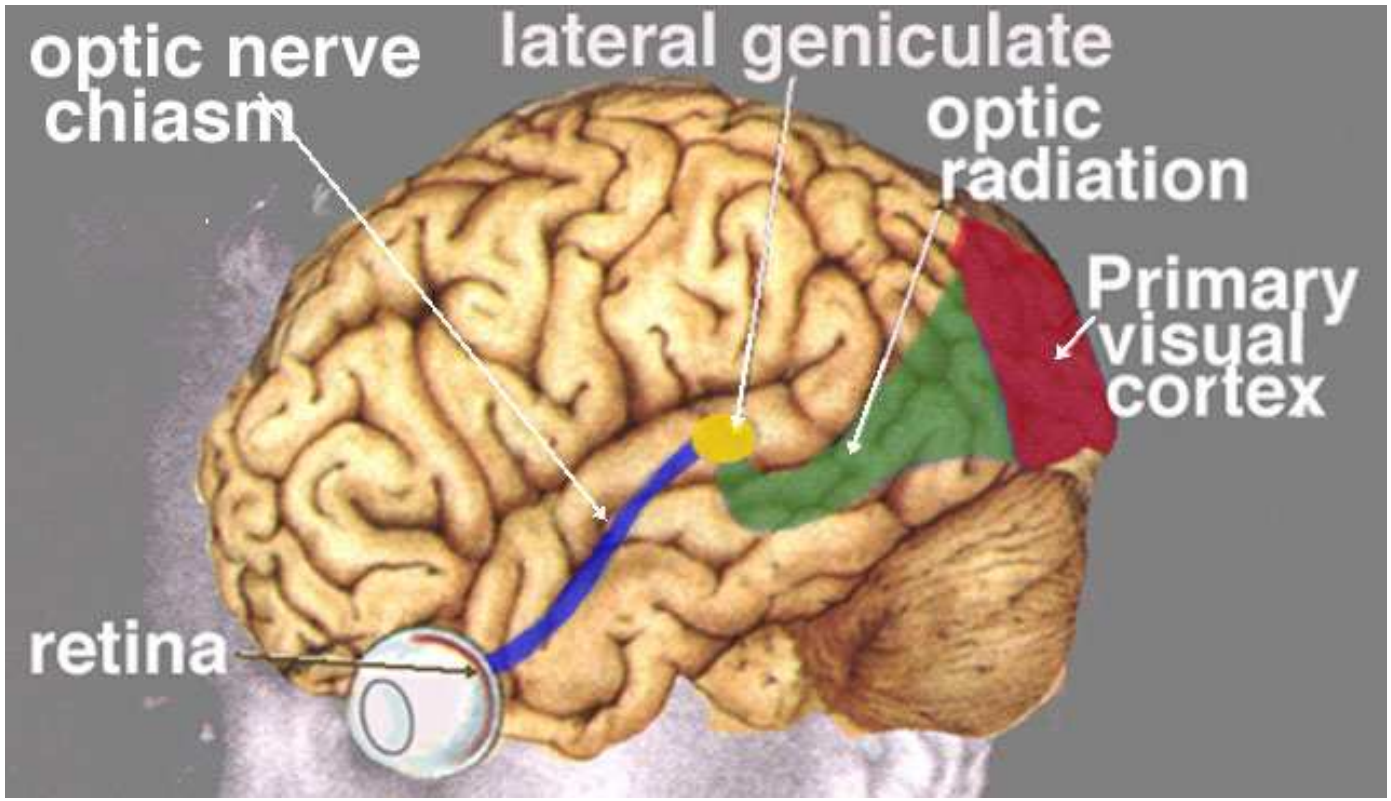
Visual Path Ways

- Optical chiasm
  - Left visual field to right hemisphere
  - Vise versa

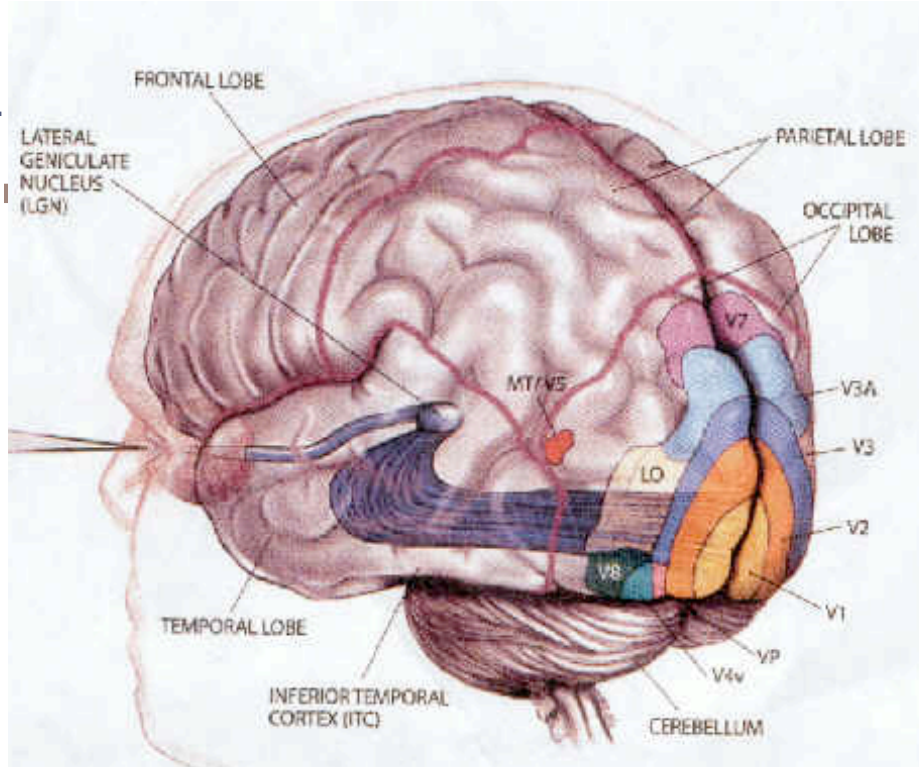


Lateral Geniculate Nucleus

- LGN passes information from the eye to the visual cortex

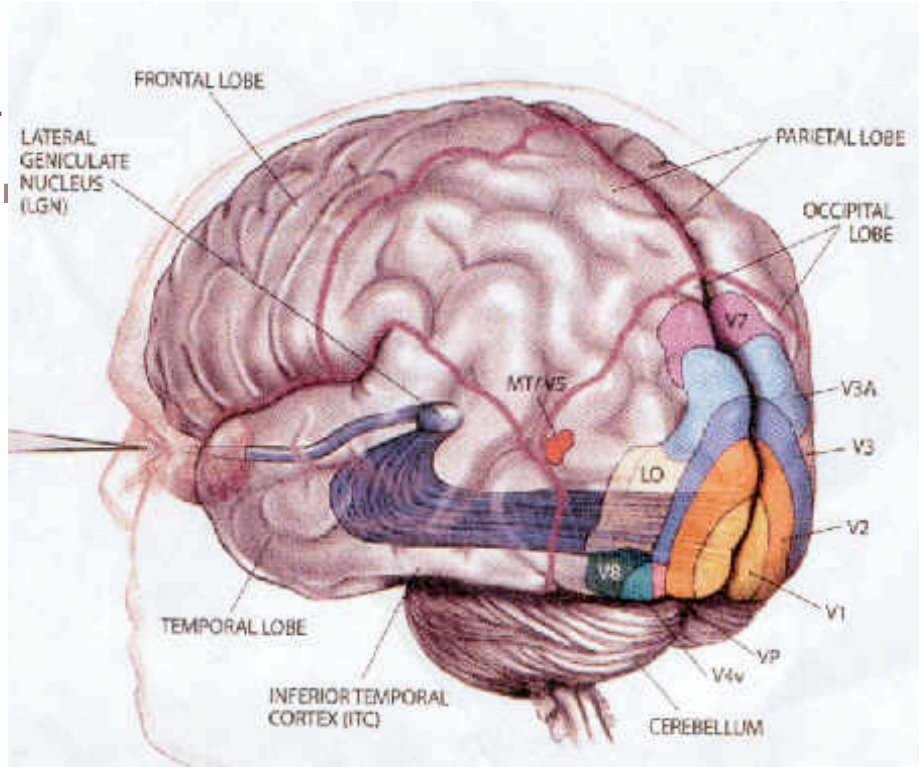


The visual cortex



- V1
  - Retinotopic mapping
  - Cortical magnification of foveal information
  - Cells sense orientation, direction, speed.
- V2
  - Tuned to orientation, spatial frequency, and color
  - Illusionary contours, figure-ground

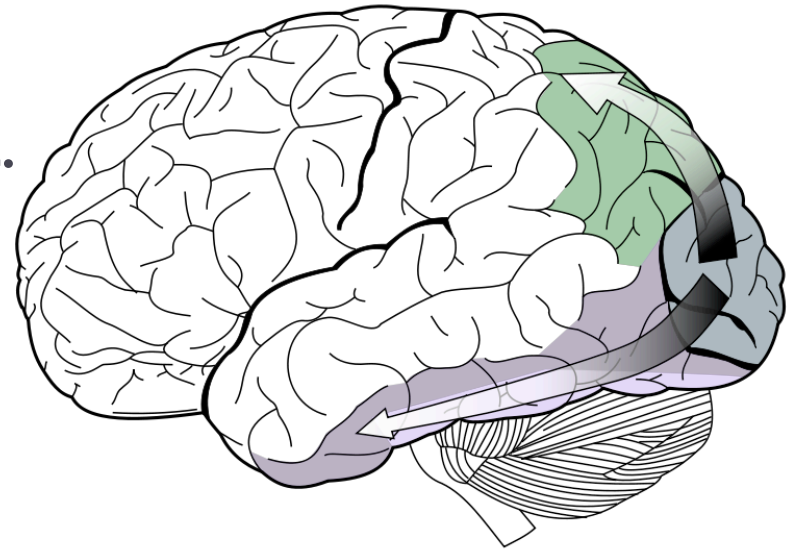
The visual cortex



- V3/V3A
  - Global motion
- V4
  - Simple shapes
- V5/MT
  - Motion perception of more complex objects
  - Control of eye movements
- Inferotemporal cortec
  - Complex objects: faces

Ventral and dorsal stream

- Dorsal stream (green)
  - V1, V2, MT, post. pariatal cort.
  - "Where" pathway
  - Motion and object location
  - Guidance of eyes and arms

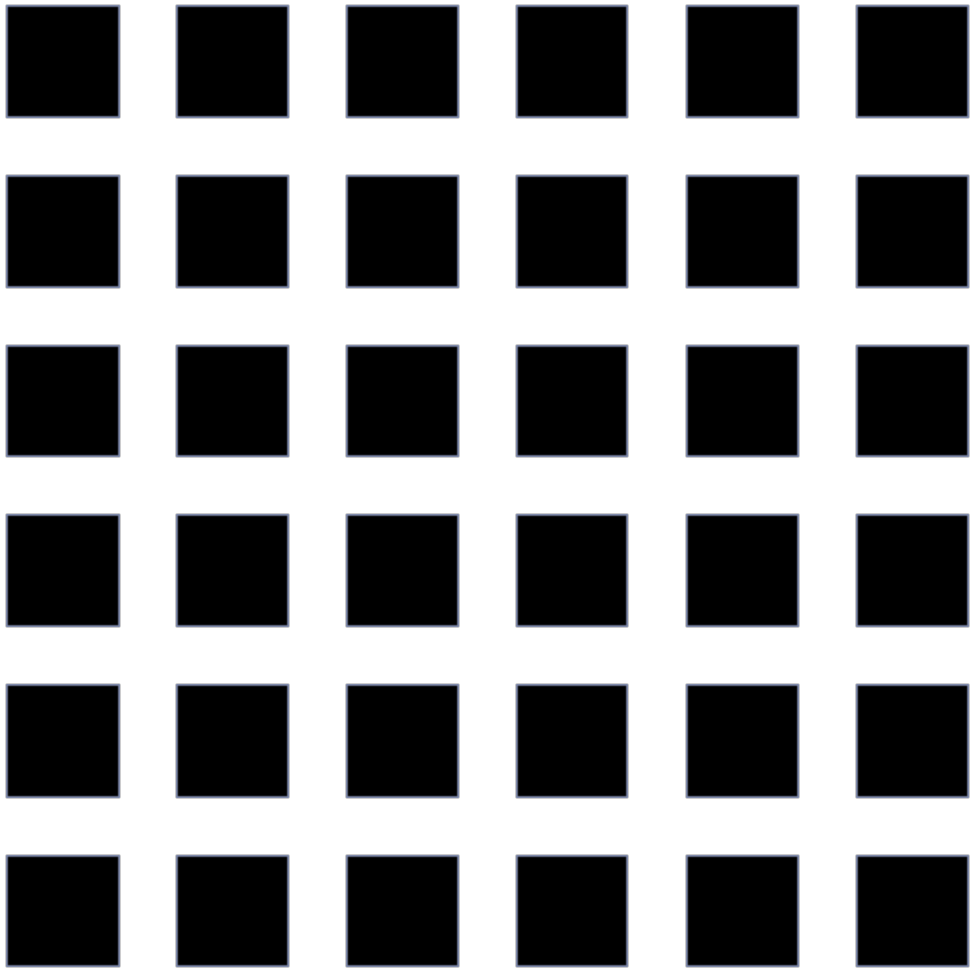


- Ventral stream (purple)
  - V1, V2, V4, inferior temporal cortex
  - "What" pathway
  - Form/object recognition, long term memory

Insights from a visual illusion

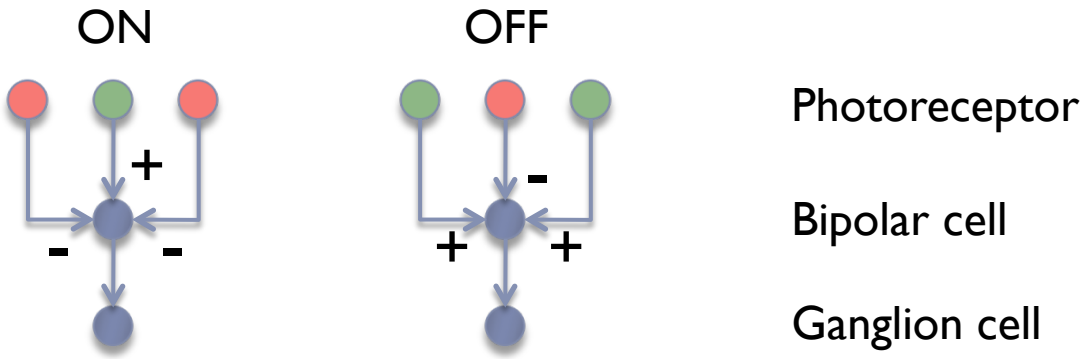
- We can gain insights about the visual system from visual illusions
  - The Hermann-grid illusion

The Hermann-grid illusion



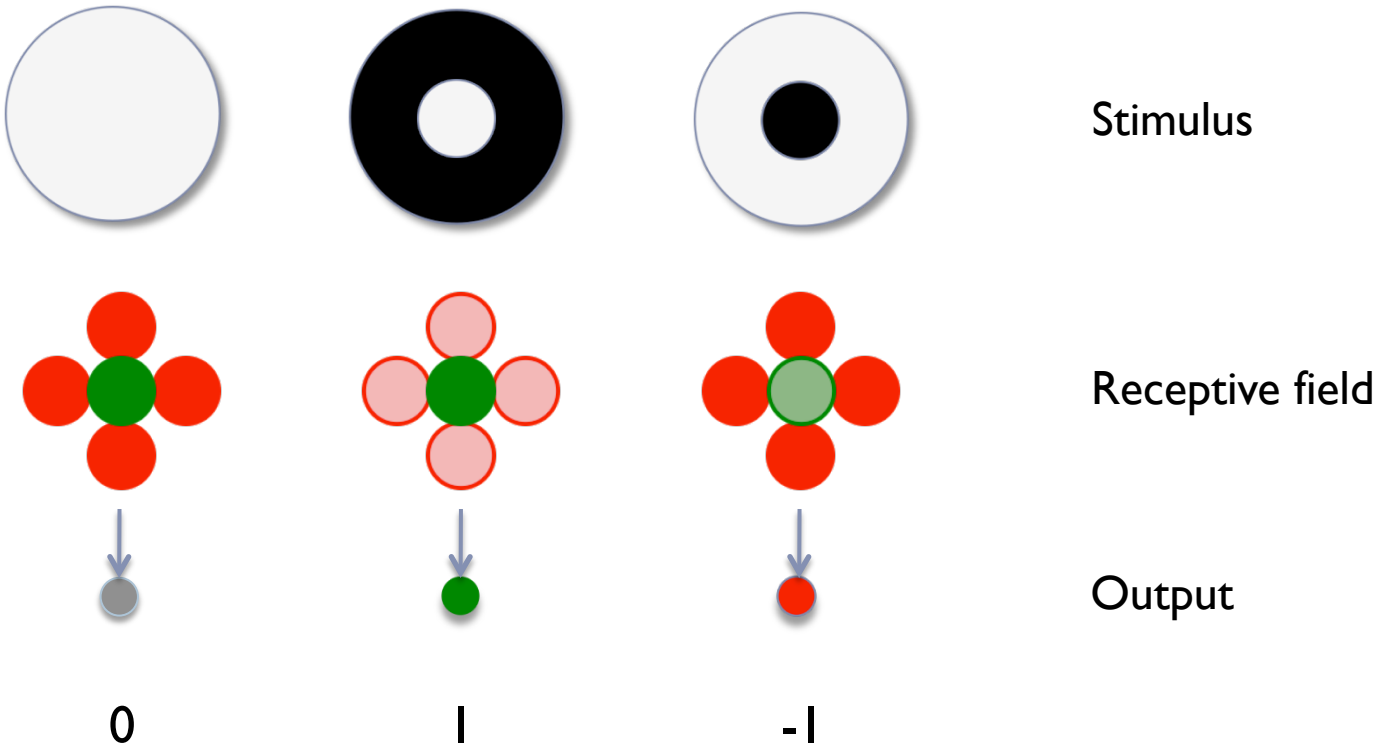
Explanation: lateral inhibition

- Hermann grid illusion
  - Black spots at the corners in the parafoveal view
  - Disappear when focused on
- This is a result from lateral inhibition
  - Neighboring photoreceptors give inhibition
  - Center-surround cells (ON/OFF)



Lateral inhibition

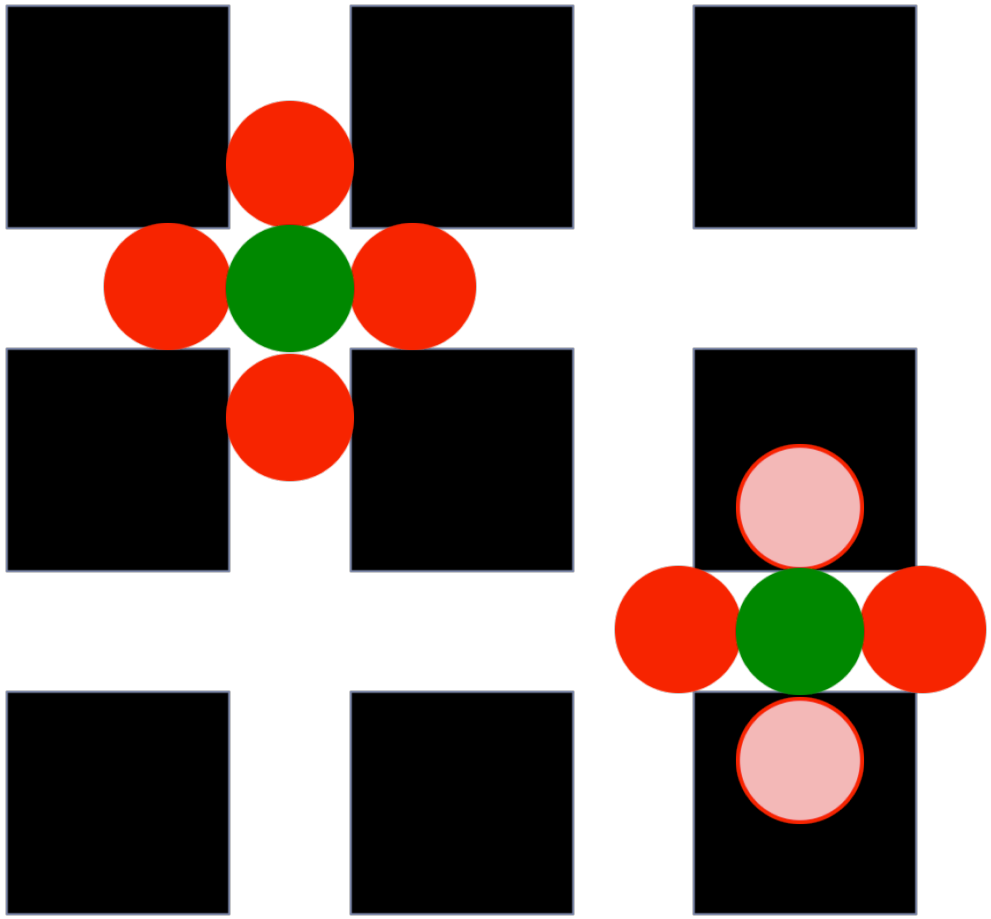
- The effect of a ON center-surround cell



Explanation of Hermann grid

Perception at the periphery

Smaller response => black dot

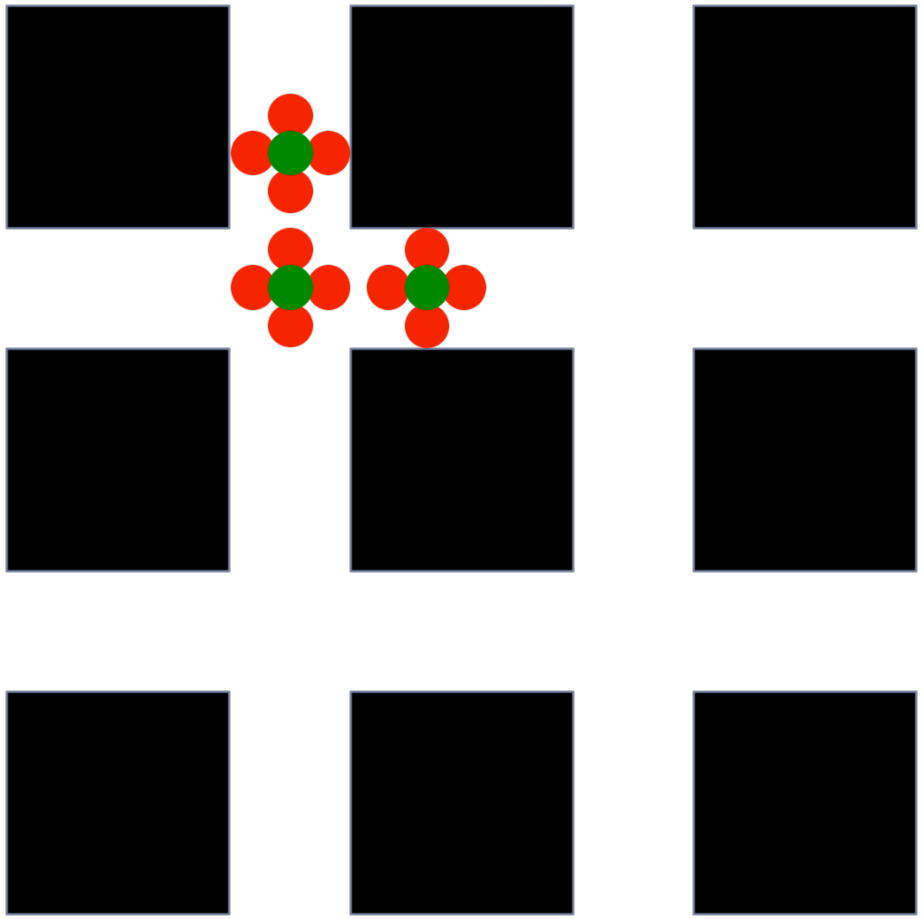


Larger response

Explanation of Hermann grid

Perception at the fovea

Smaller receptive fields



Same response for all cells

Center-surround organization

- Through out the visual system
  - Ganglion cells
  - Cells in Lateral Geniculate Nucleus (LGN)
  - Visual cortex
- Larger and larger receptive fields
- For different features

- Stimuli are perceived relative to the context

Gradient Illusion

- Brightness perceived relative to the context

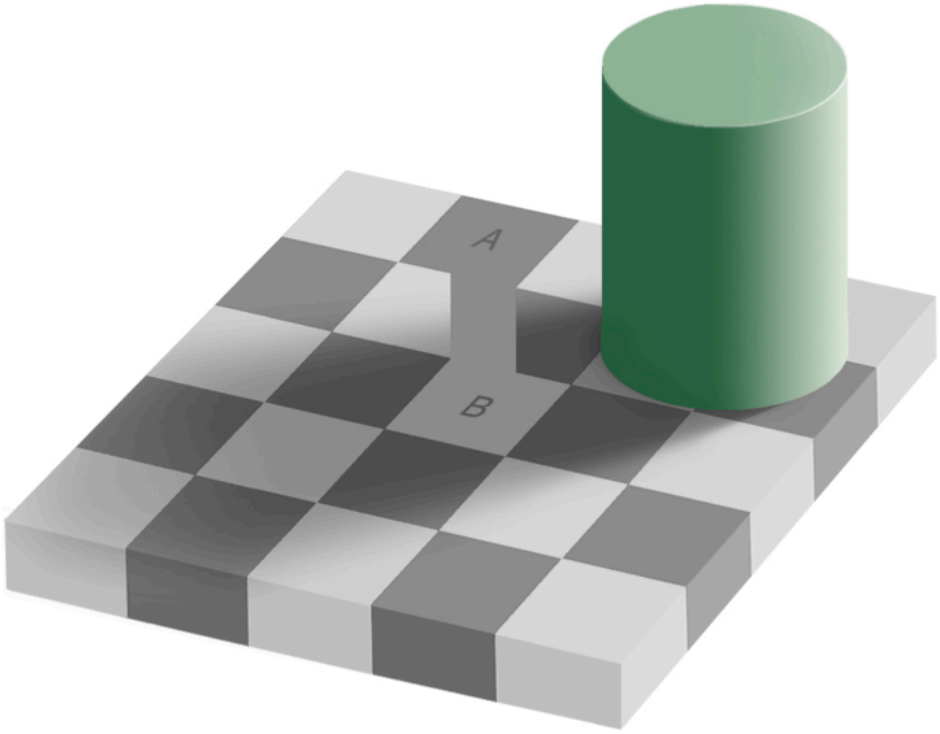
The contrast effect

- The human visual system is not a very good physical light meter.
- Brightness perception enhances or diminishes based on the contrast with the surroundings

- Is this a failure or a feature of the system?

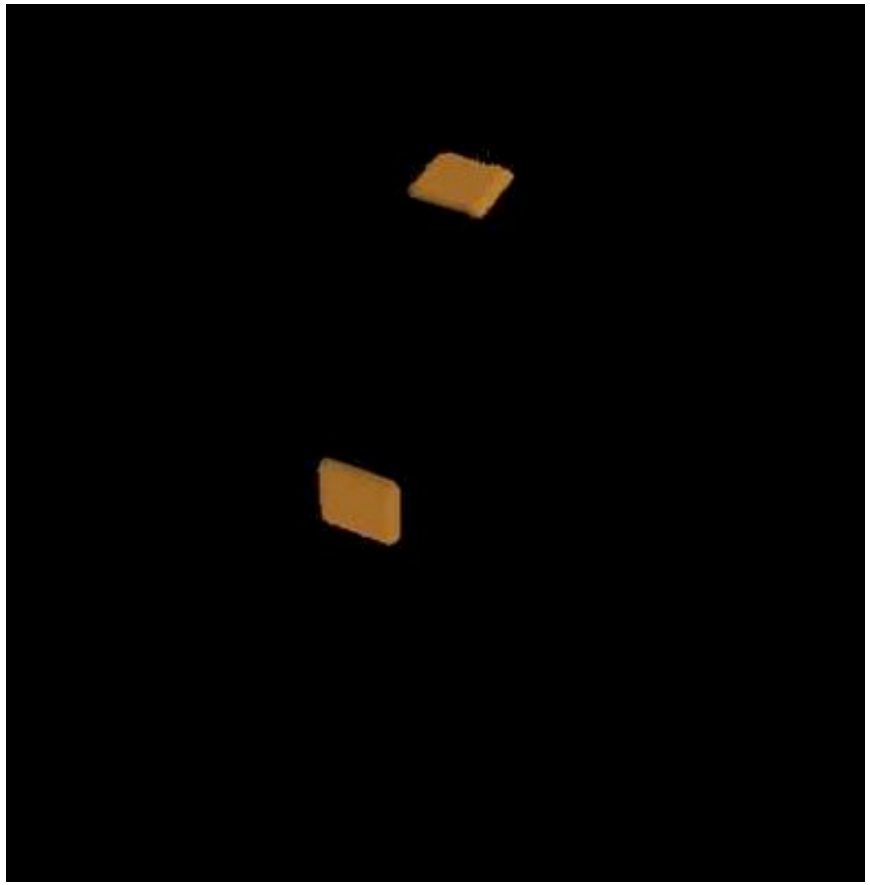
Adelson's checker-shadow illusion

- Which square is brighter, A or B?



Color tile illusion

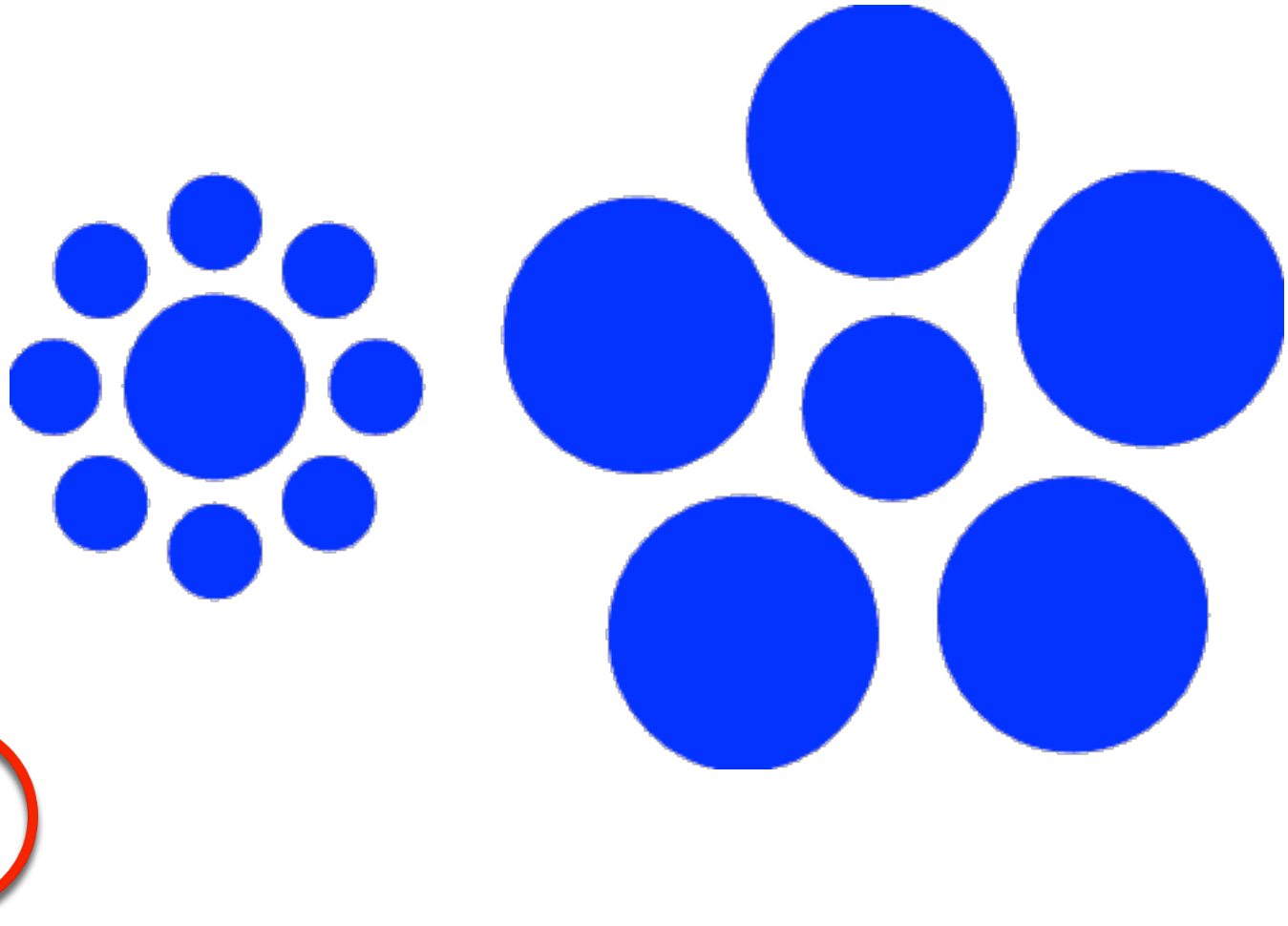
- Name the colors on the cube



Color/Brightness Constancy

- The contrast effect is a feature, not a failure
- The human visual system solves the brightness constancy problem.
  - It is about what is black and what white, not about the absolute brightness values.
- And color constancy
  - Identifying colors in different light conditions and with shadows
- This is important for machine vision

Contrast in shape



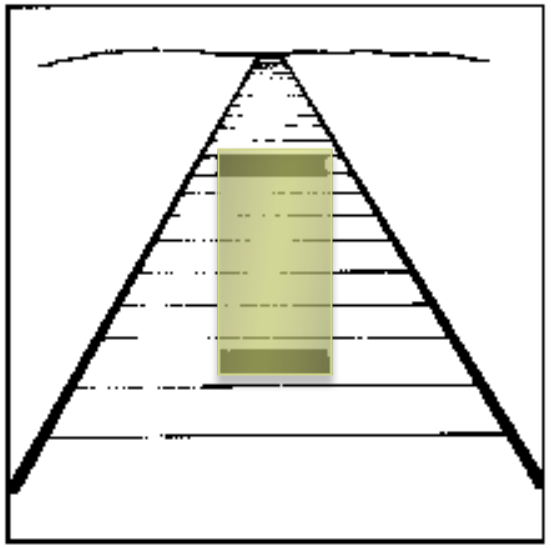
Contrast in beauty

- Which Penelope is more attractive?

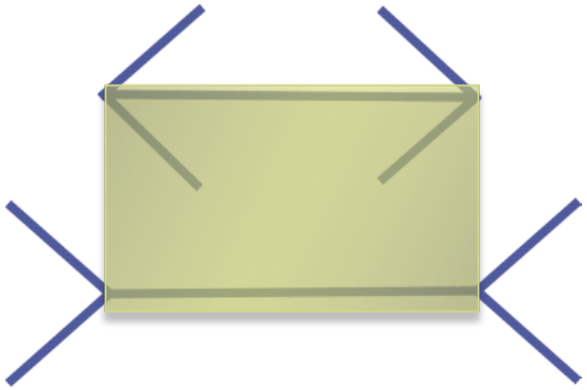



Use of contextual information

- Context:
  - Ponzo illusion

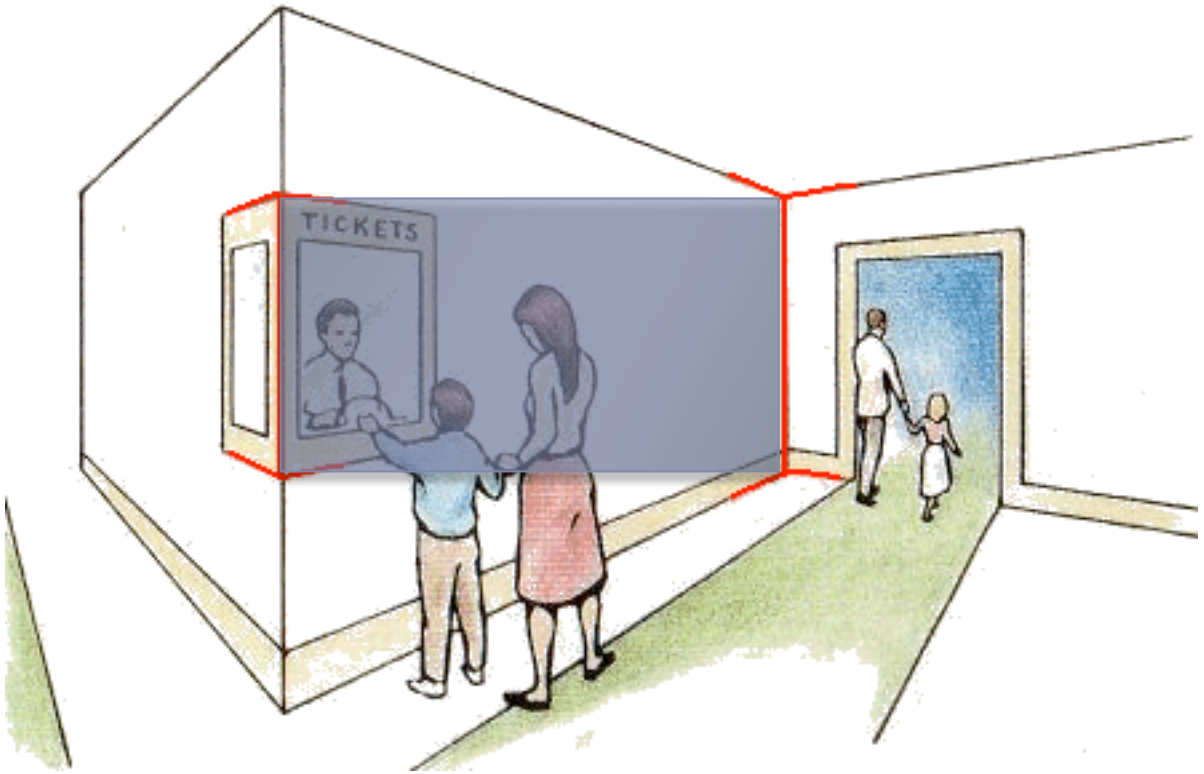


  - Müller-Lyer illusion



We perceive things in 3D

- Müller-Lyer illusion: world knowledge



Let machines perceive things in 3D

- Promotes the use of 3D in machine vision
- However, 3D scene understanding is a difficult problem
  - Having stereo information is not the same as understanding the 3D structures
  - World knowledge

Visual attention

Visual attention

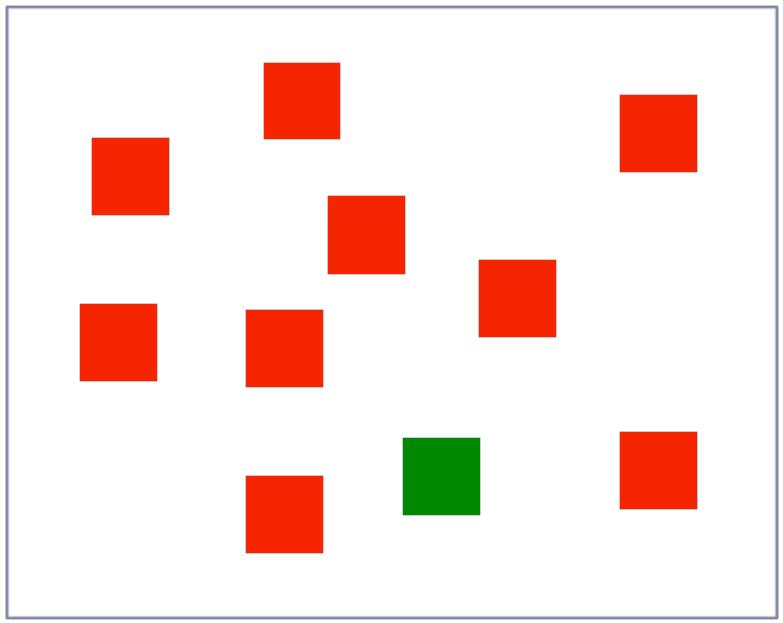
- Visual attention
  - Focusing of attention to small, but interesting parts of the visual field
- Overt and covert visual attention
  - Overt: Making eye/head/body movements to focus
  - Covert: Mentally focusing attention
- Advantage of (visual) attention
  - only process interesting/relevant information

Eye movements

- Periphery => determine what is interesting
- Focus the gaze on that part of the visual field
- Viewed on high resolution with fovea

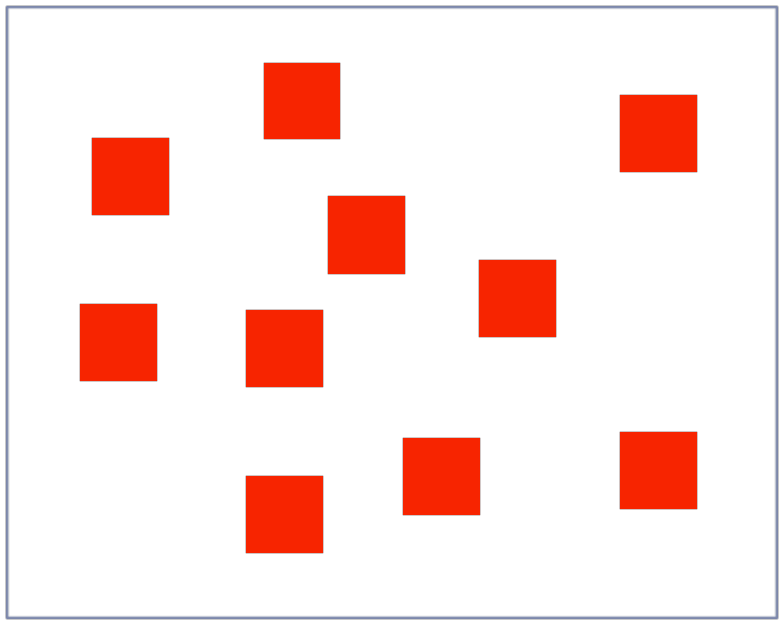
Visual search

- Raise your right hand when you see the odd figure

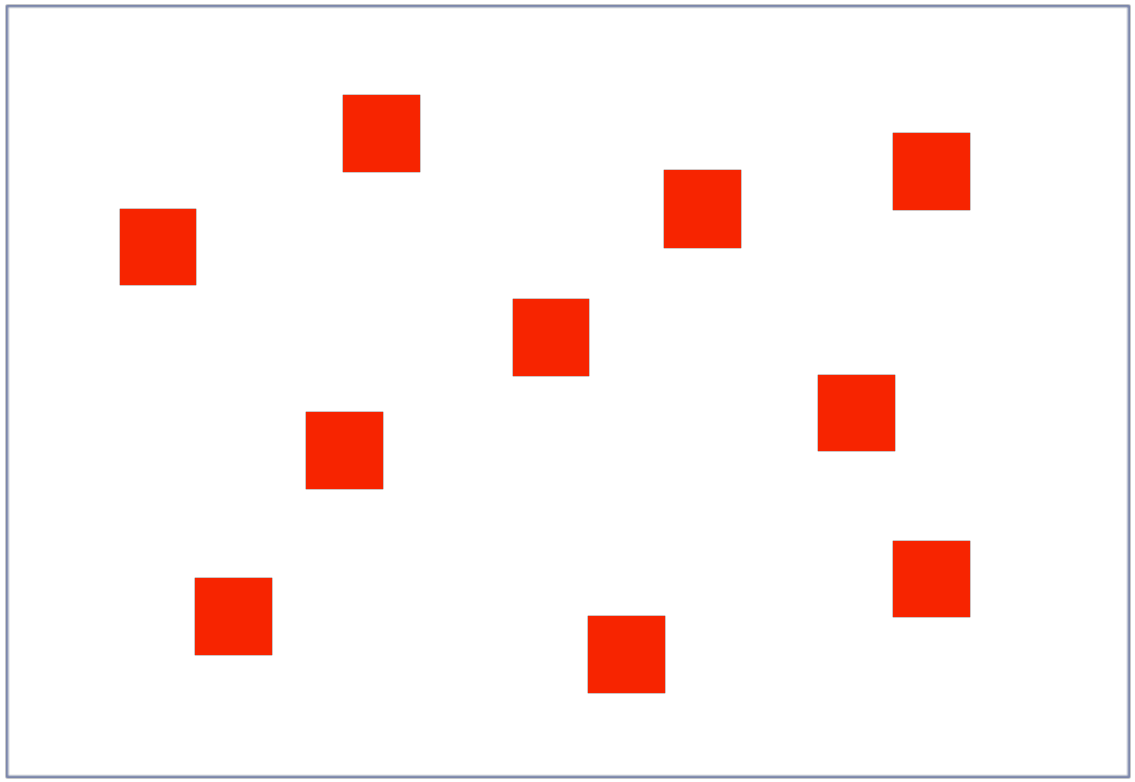


Visual search

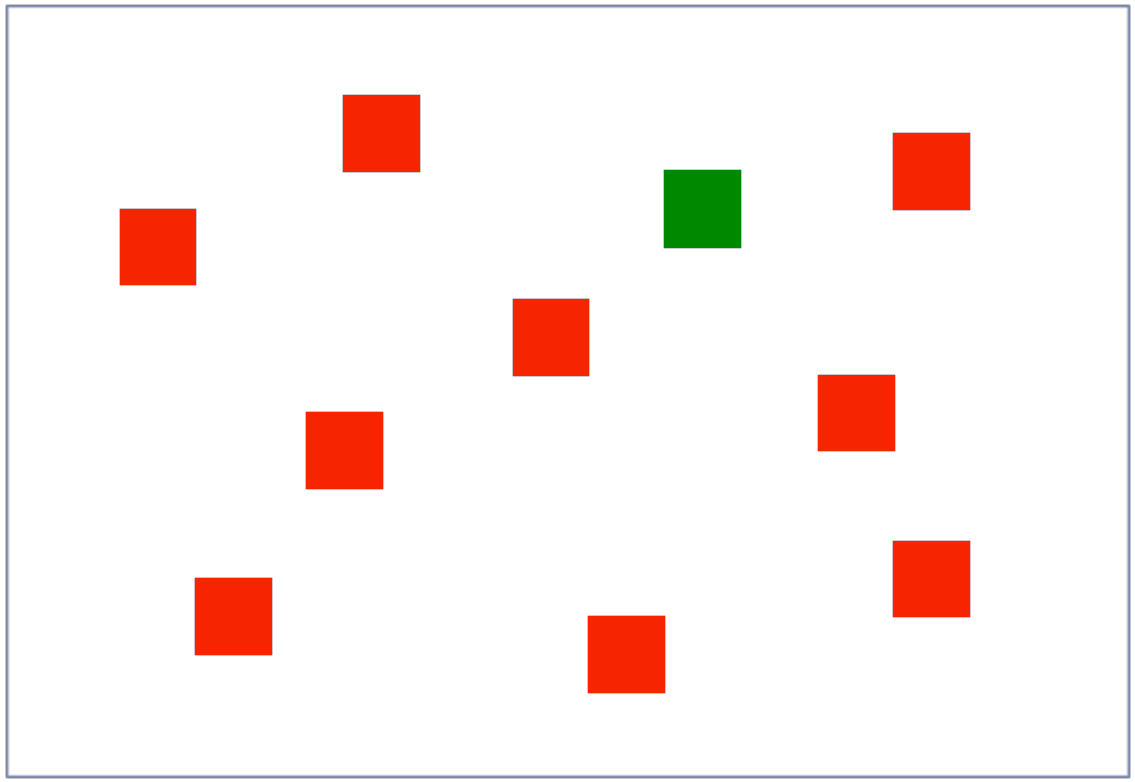
- Raise your left hand when there is no odd figure



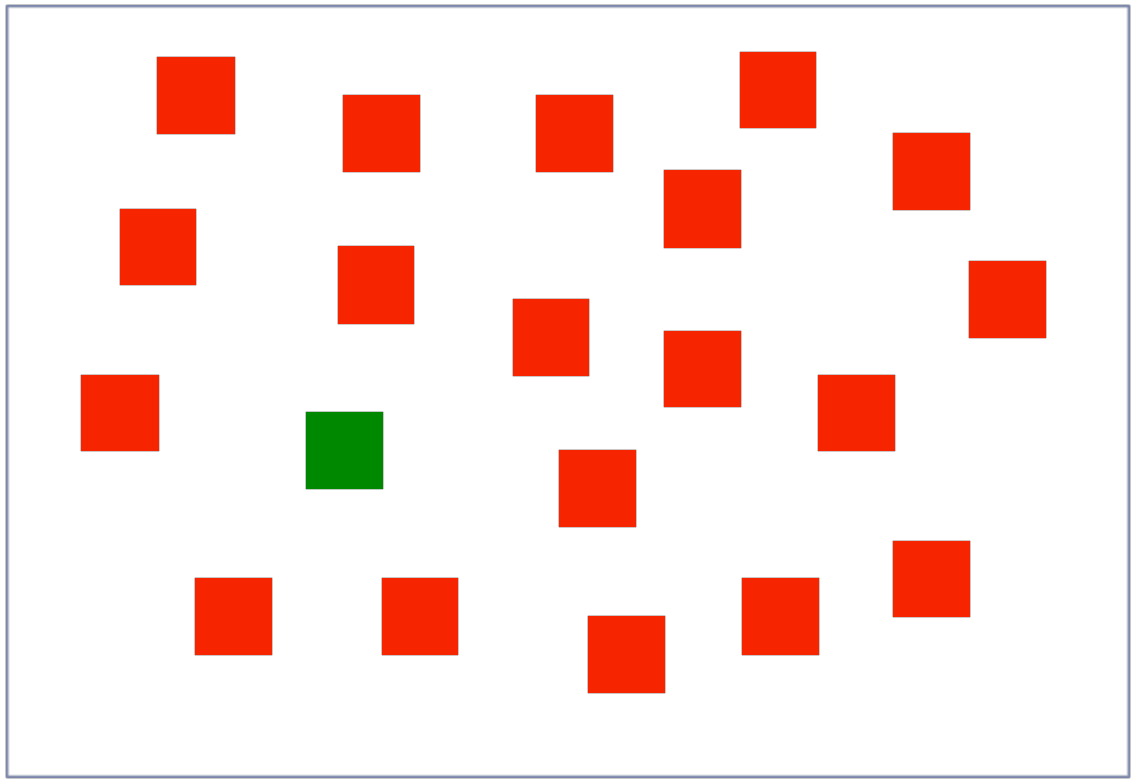
Experiment



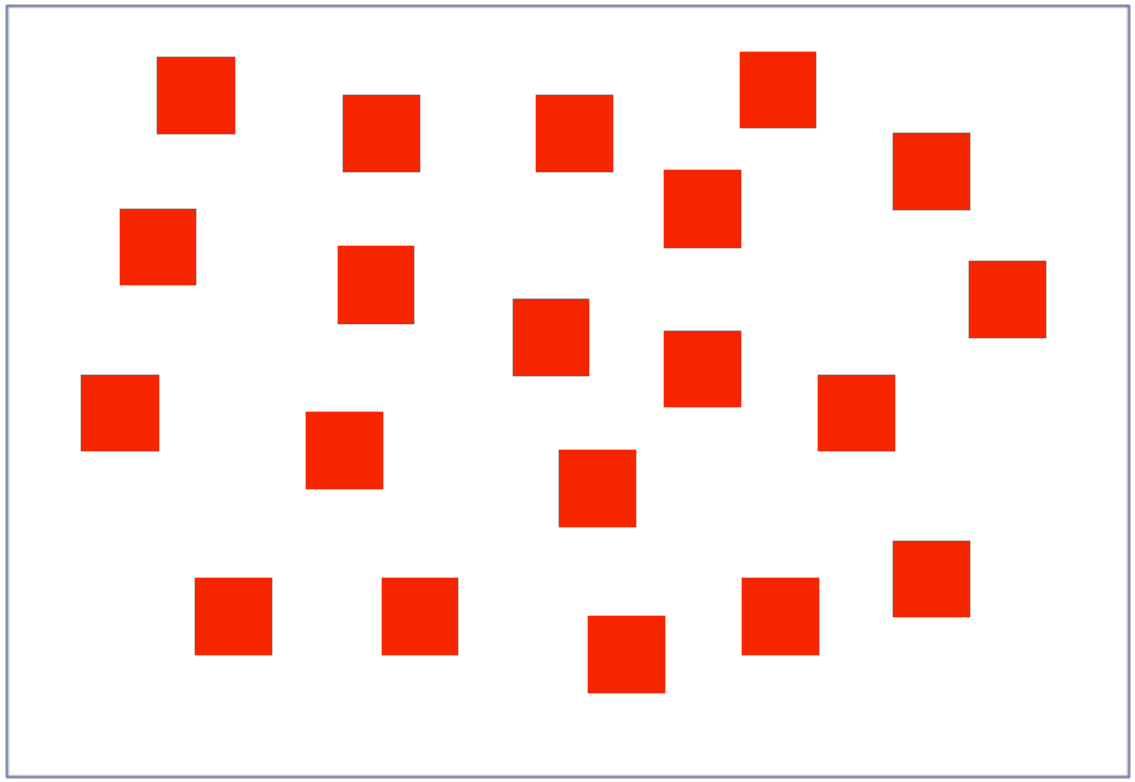
Experiment



Experiment

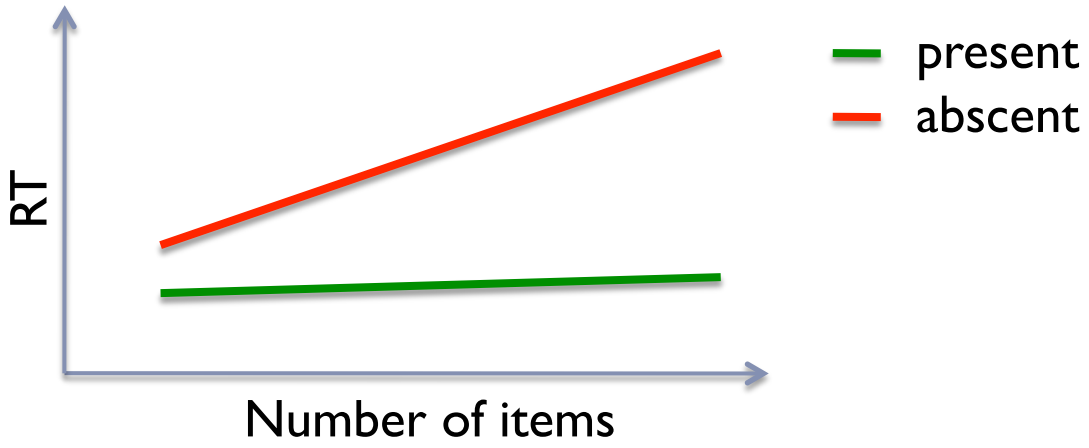


Experiment



The pop-out effect

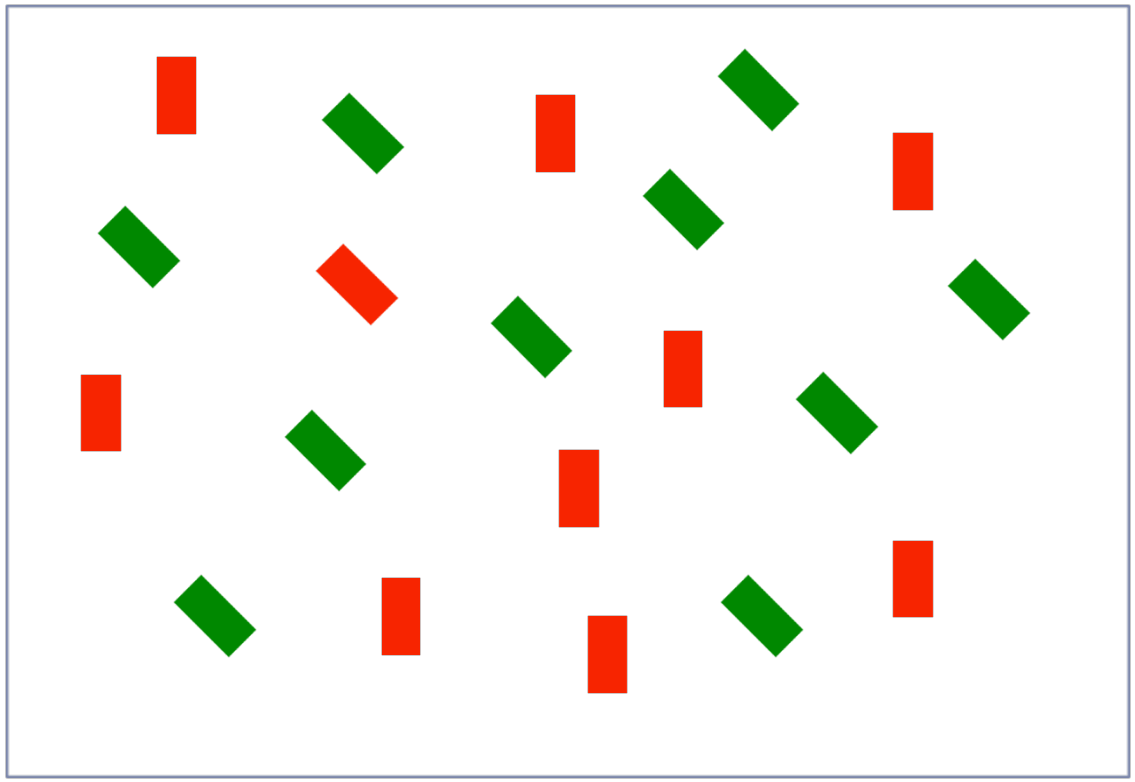
- This stimulus results in a pop-out effect



  - Efficient search for the target

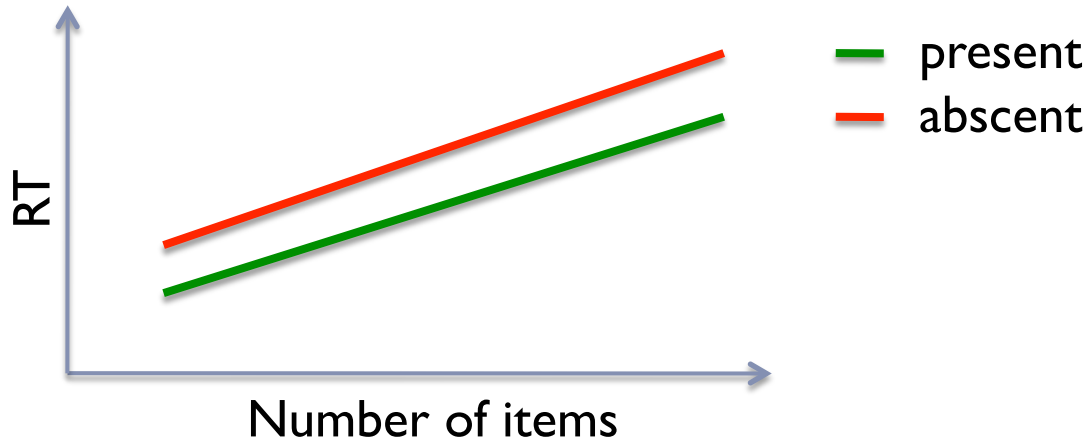
Conjunction search

- Now search for the tilted red bar 



The pop-out effect

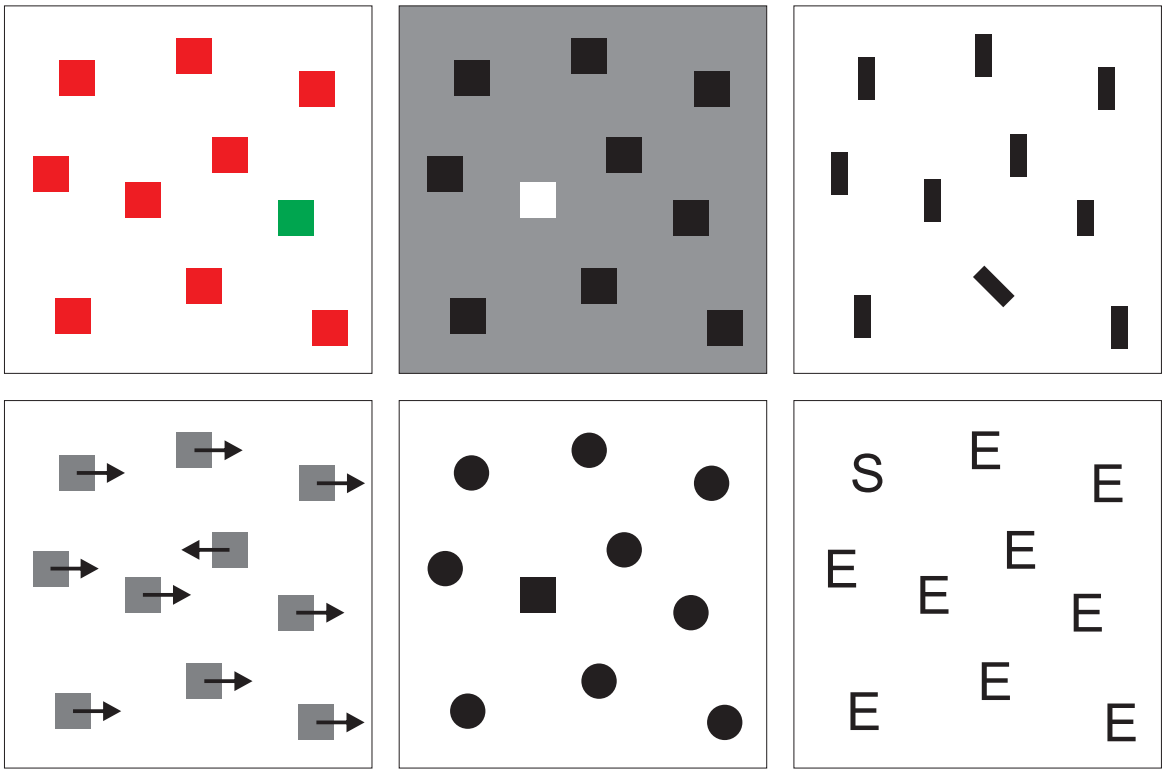
- Reaction times in conjunction search



  - Inefficient search for conjunction target

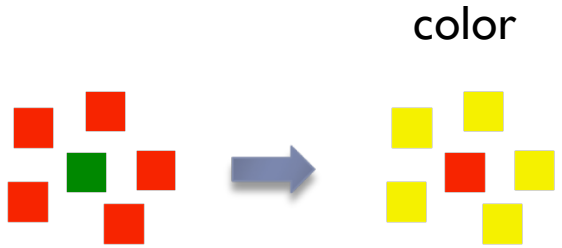
Basic features

- Efficient search for basic features
  - Color
  - Brightness
  - Orientation
  - Motion
  - Shape
  - …

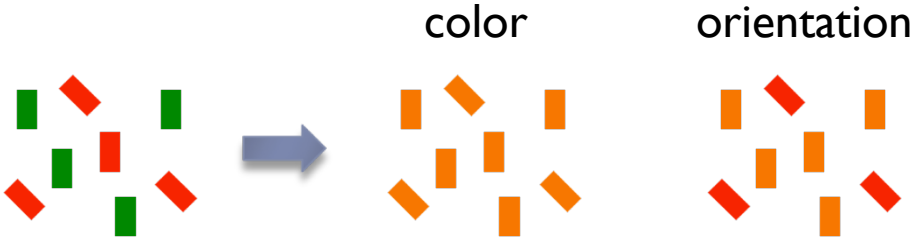


Visual-search theory

- Feature-integration theory (Treisman & Gelade)
  - Different basic features are processed in parallel
  - Center-surround contrast is represented in separate feature maps
  - Feature maps are integrated in overall saliency map
- Predictions
  - Single search:



  - Conjunction search:



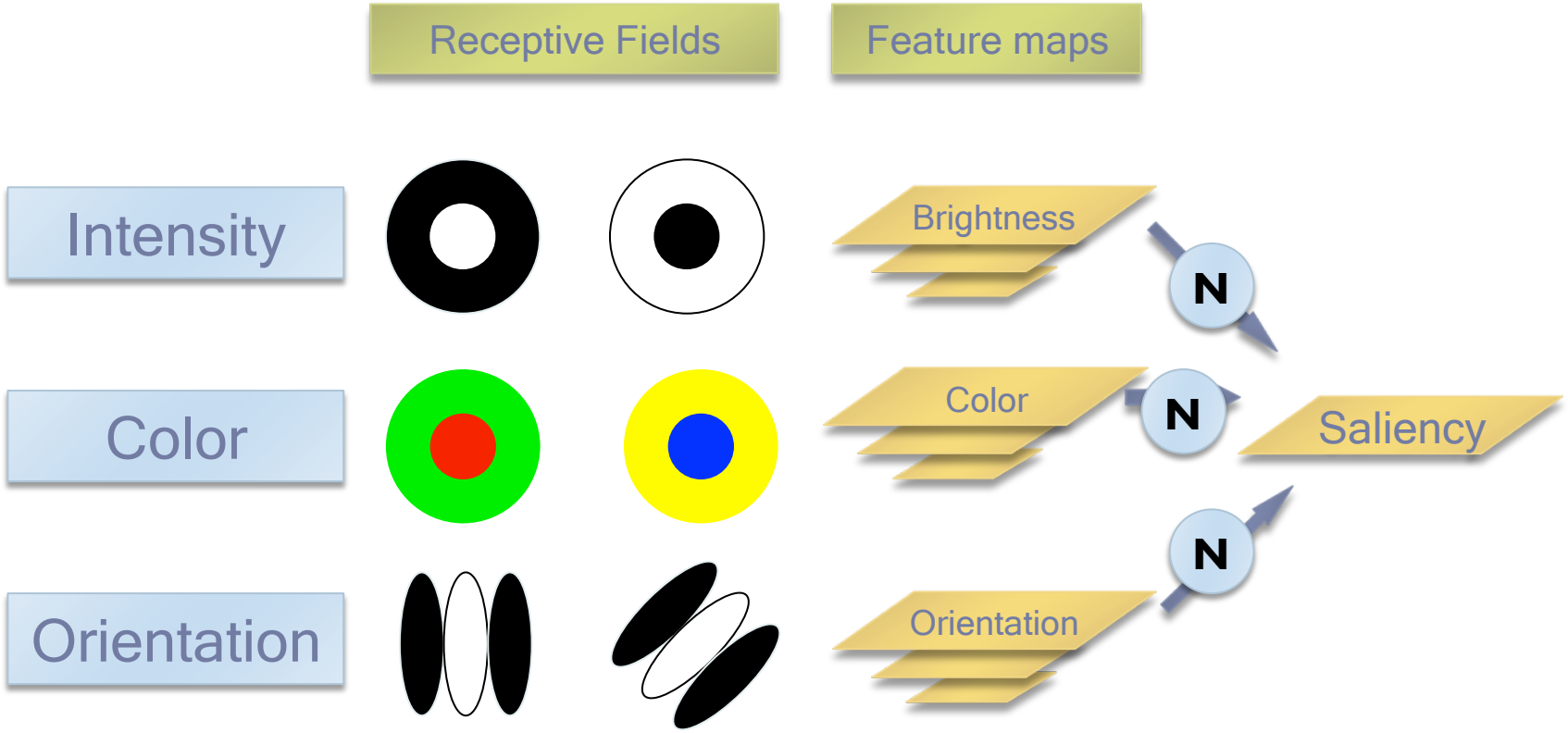
Contrast-saliency model
(Itti, Koch & Niebur, 1998)

Saliency model of Itti & Koch

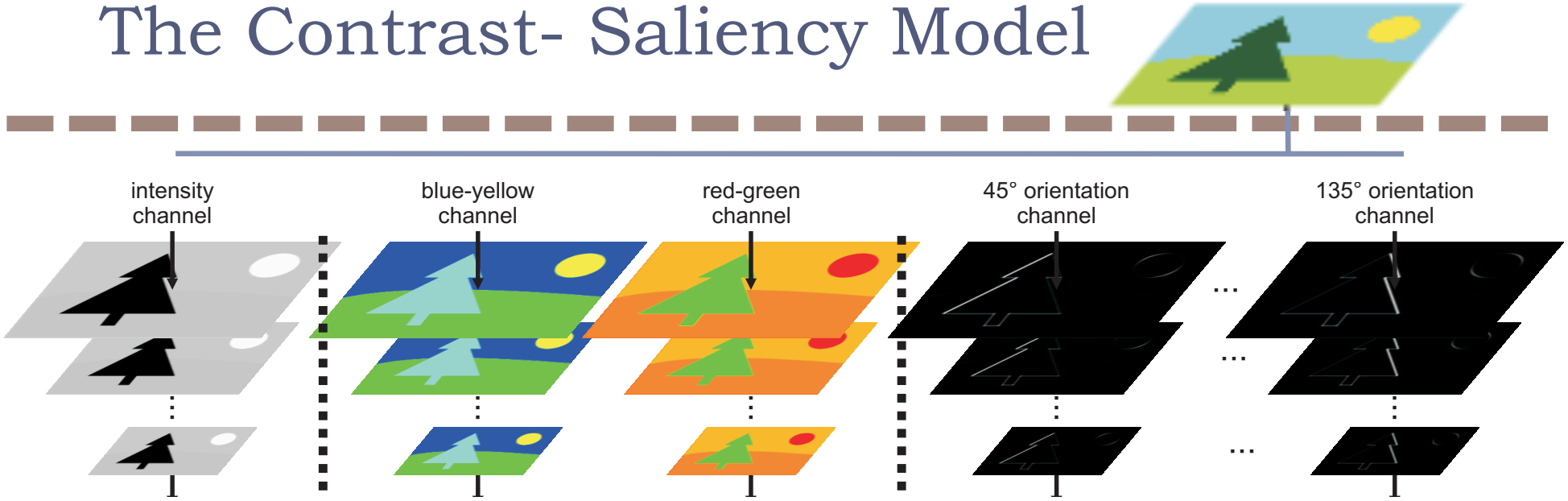
- Itti, Koch and Niebur, 1998
- Inspired by the Feature-integration theory
- Based on center-surround contrasts of basic feature
  - Brightness
  - Color
  - Orientation
- Used frequently in machine/robot vision

Saliency model of Itti & Koch

- Saliency model (Itti, Koch & Niebuhr 1998)

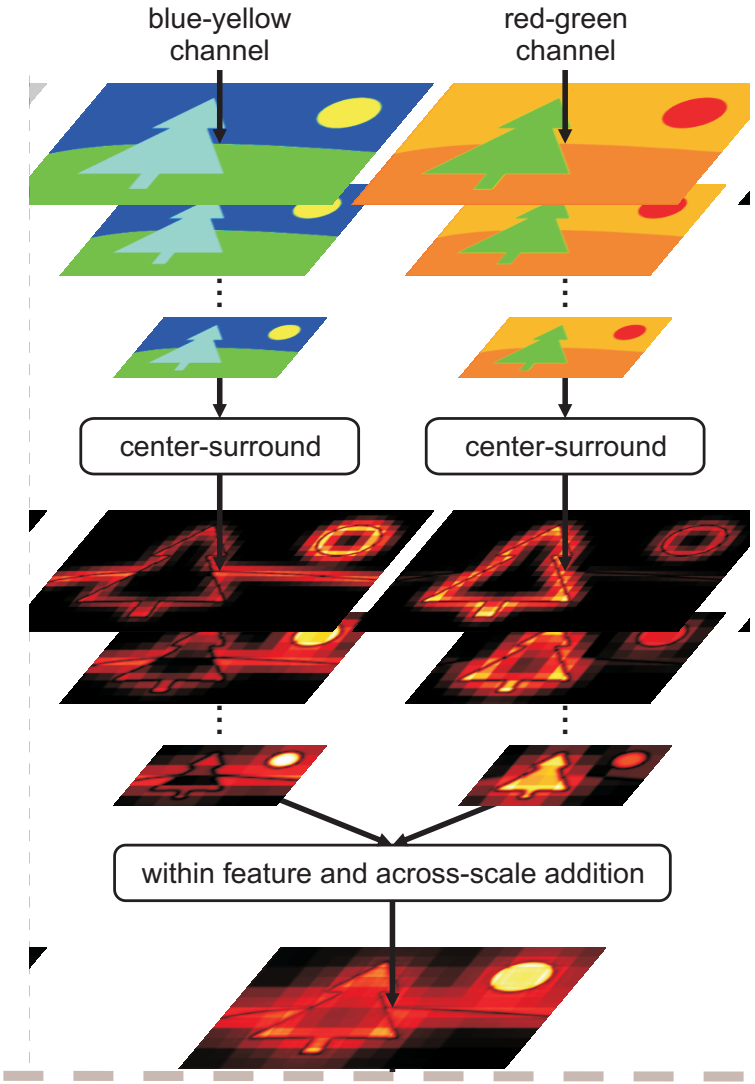


The Contrast- Saliency Model

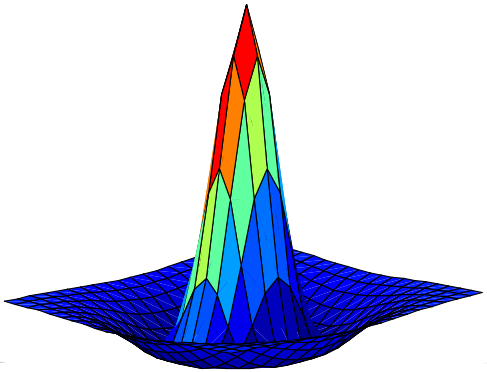


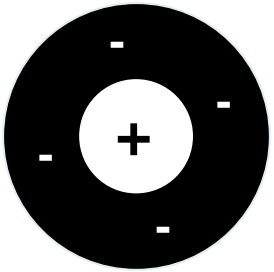
- Different feature channels
  - 1 brightness
  - 2 color channels (R-G, B-Y)
  - 4 orientation channels (0°, 45°, 90°, 135°)
- On seven different scales – image pyramid

The Contrast-Saliency Model



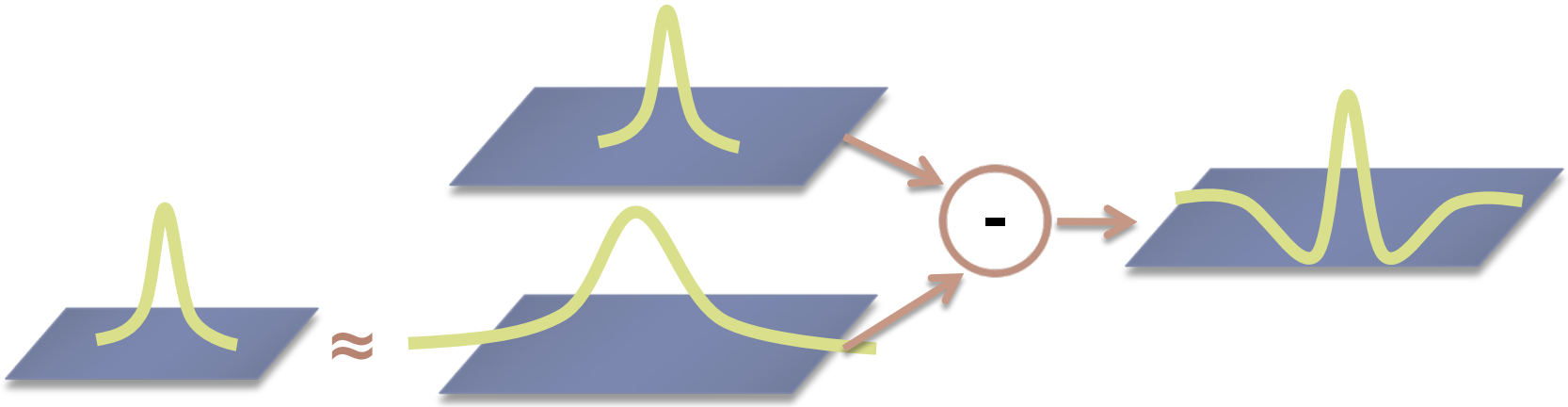
- Center-surround
  - Calculating the center-surround responses on the different scales
  - Convolution of images with Difference-of-Gaussians





Center-Surround Calculation

- An approximation of Difference of Gaussians
  - Smoothing of every scale with a Gaussian kernel
  - Subtraction of two different scales
  - $(I_c * G) - (I_{c+\delta} * G)$



Center-Surround Calculation

- A faster approximation with integral images

Making

Using

Original image

1	2	1	2
3	0	2	1
1	3	1	1
2	1	0	1

1	2	1	2
3	0	2	1
1	3	1	1
2	1	0	1

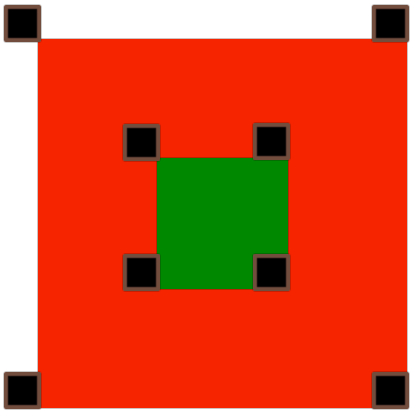
Integral image

1	3	4	6
4	6	9	12
5	10	14	18
7	13	17	22

1	3	4	6
4	6	9	12
5	10	14	18
7	13	17	22

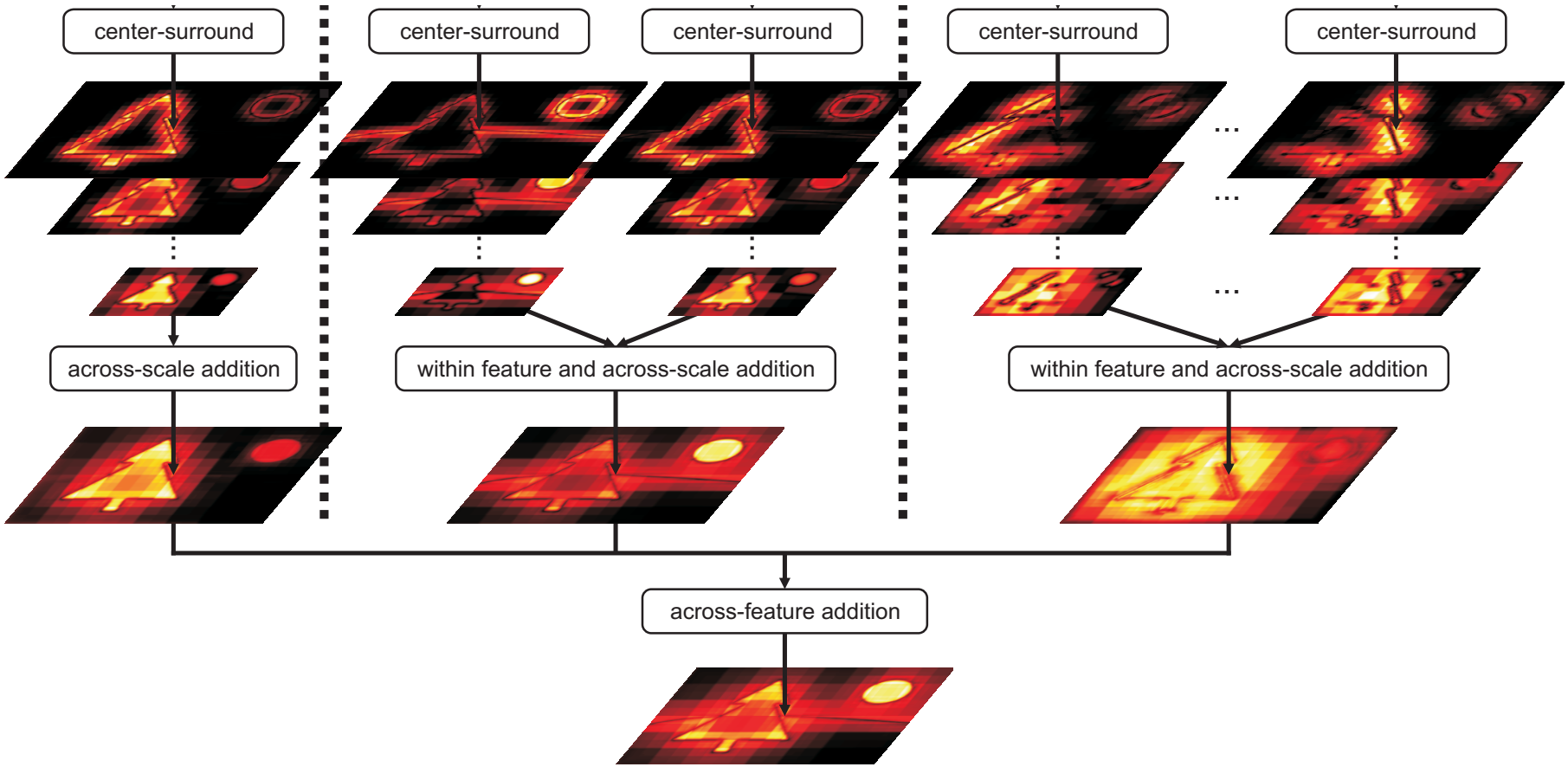
14 + 1 - 5 - 4 = 6

Difference Of Gaussian



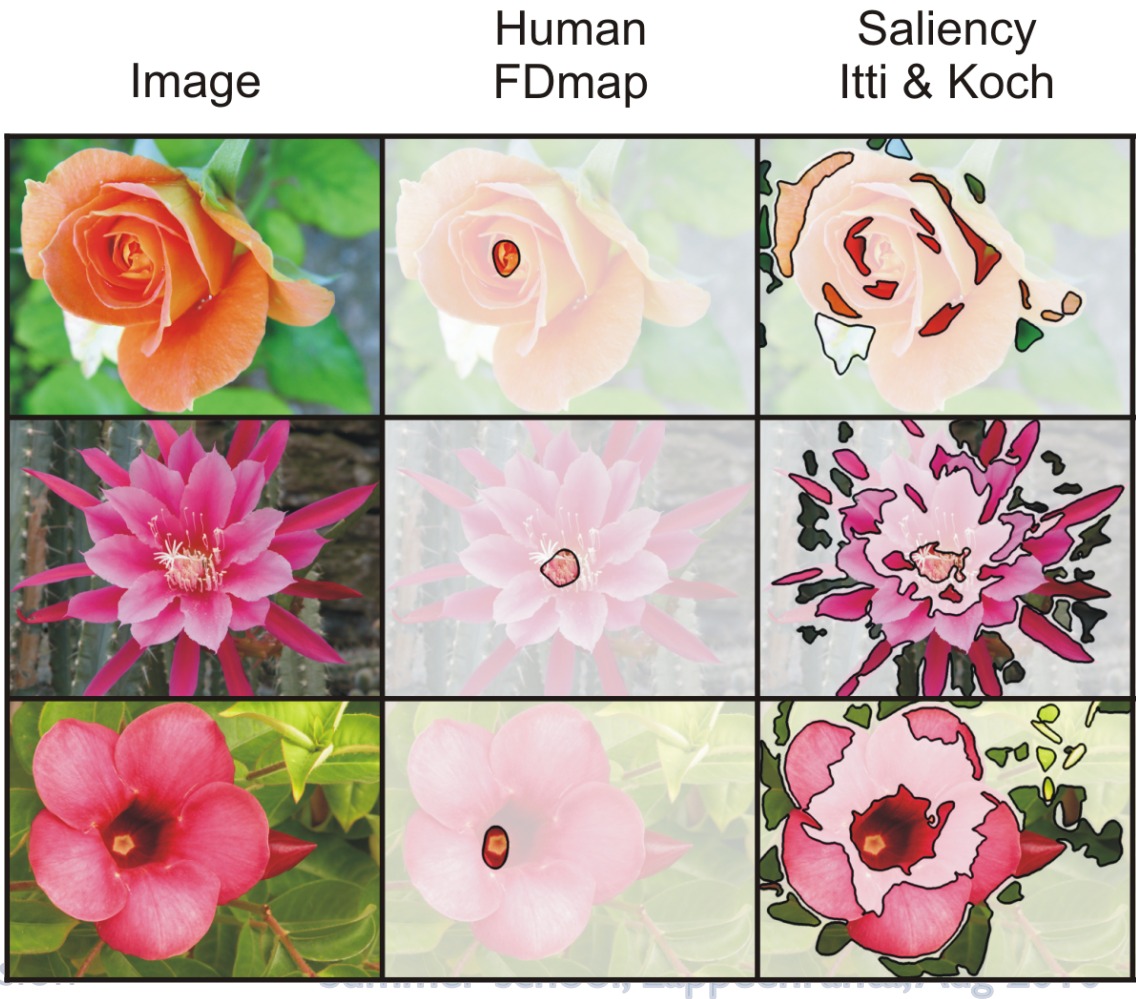
8 additions and subtraction

Across Scale Addition



Contrast-Saliency vs Eye Fixations

- This model predicts human eye fixations to some extend, but…
- … predicts saliency at borders and corners, not at the center of the object.



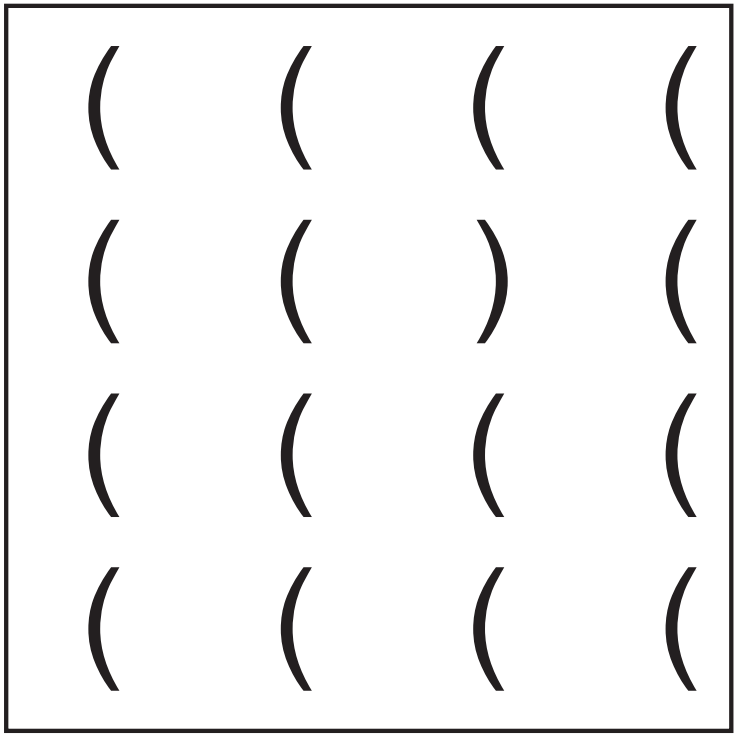
Basic Features

- The model of Itti and Koch is based on contrast in basic feature
- However, human visual attention entails much more than basic feature pop-out.
- Configurations of basic features, (proto) objects attract attention

Configural features

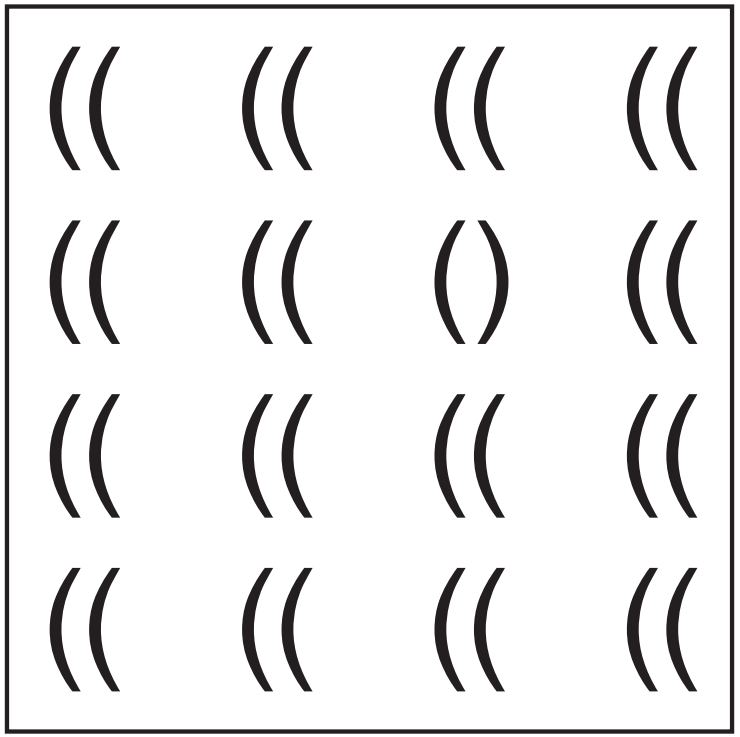
Configural superiority

- Find the) among (



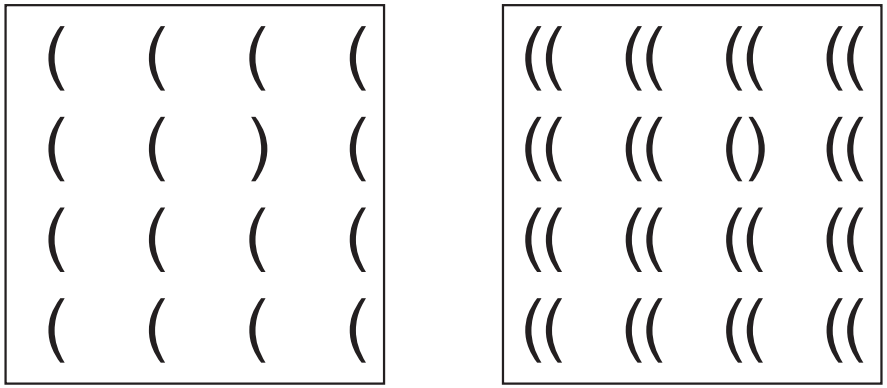
Configural superiority

- Now make it harder by adding more items



Configural superiority

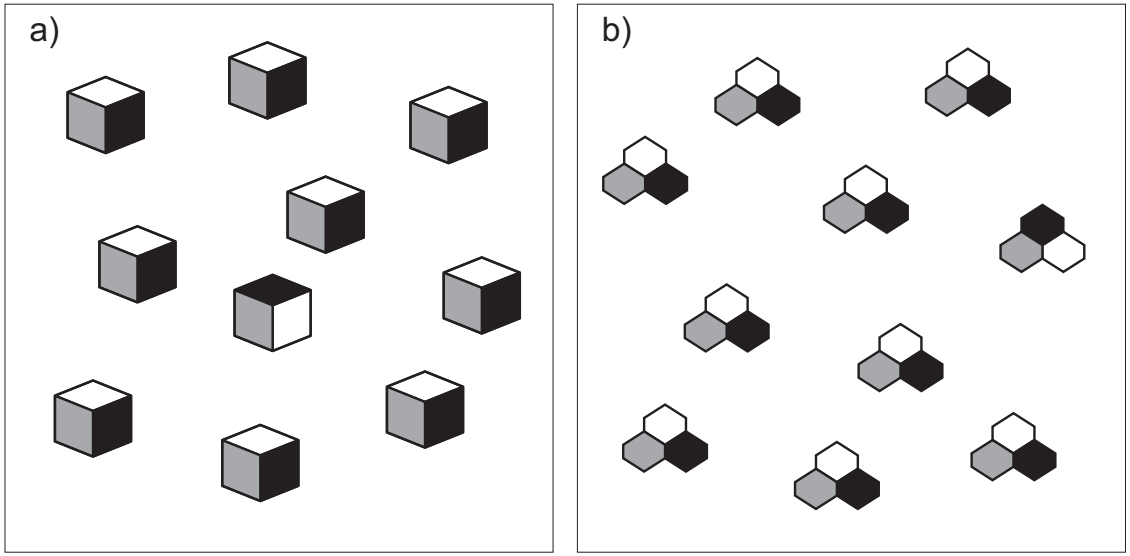
- Search becomes more efficient
- Configural superiority
  - Humans don't see 32 items, but 16 figures, () or ((
- Features



  - Curvature
  - Emerging features: symmetry and closure

Attention for objects

- Objects attract more attention than basic features
  - Easier to find target when it can be interpreted as a (3D) object



Gestalt theory

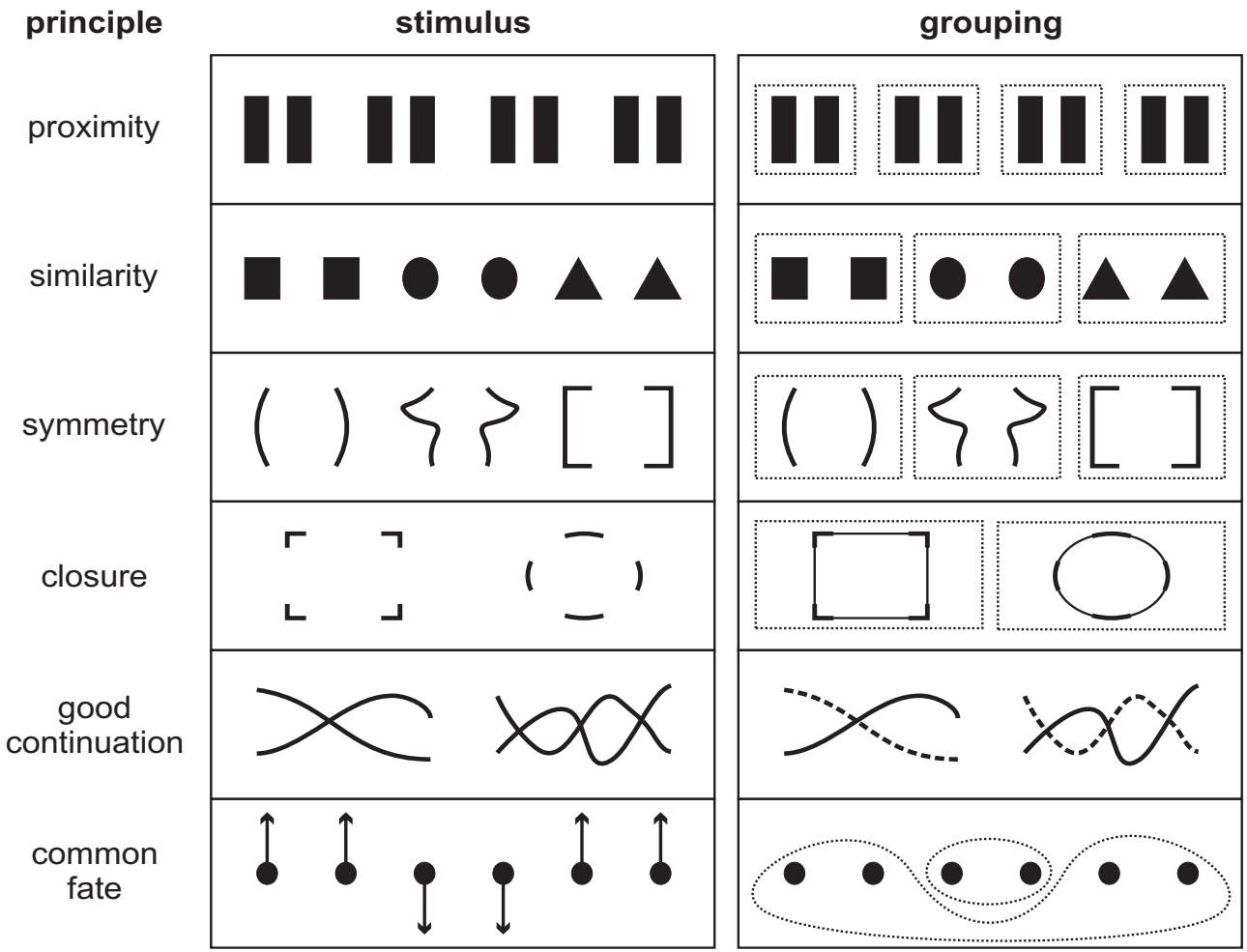
From parts to wholes

- The visual system has the tendency to group the parts into larger whole (objects/scenes).

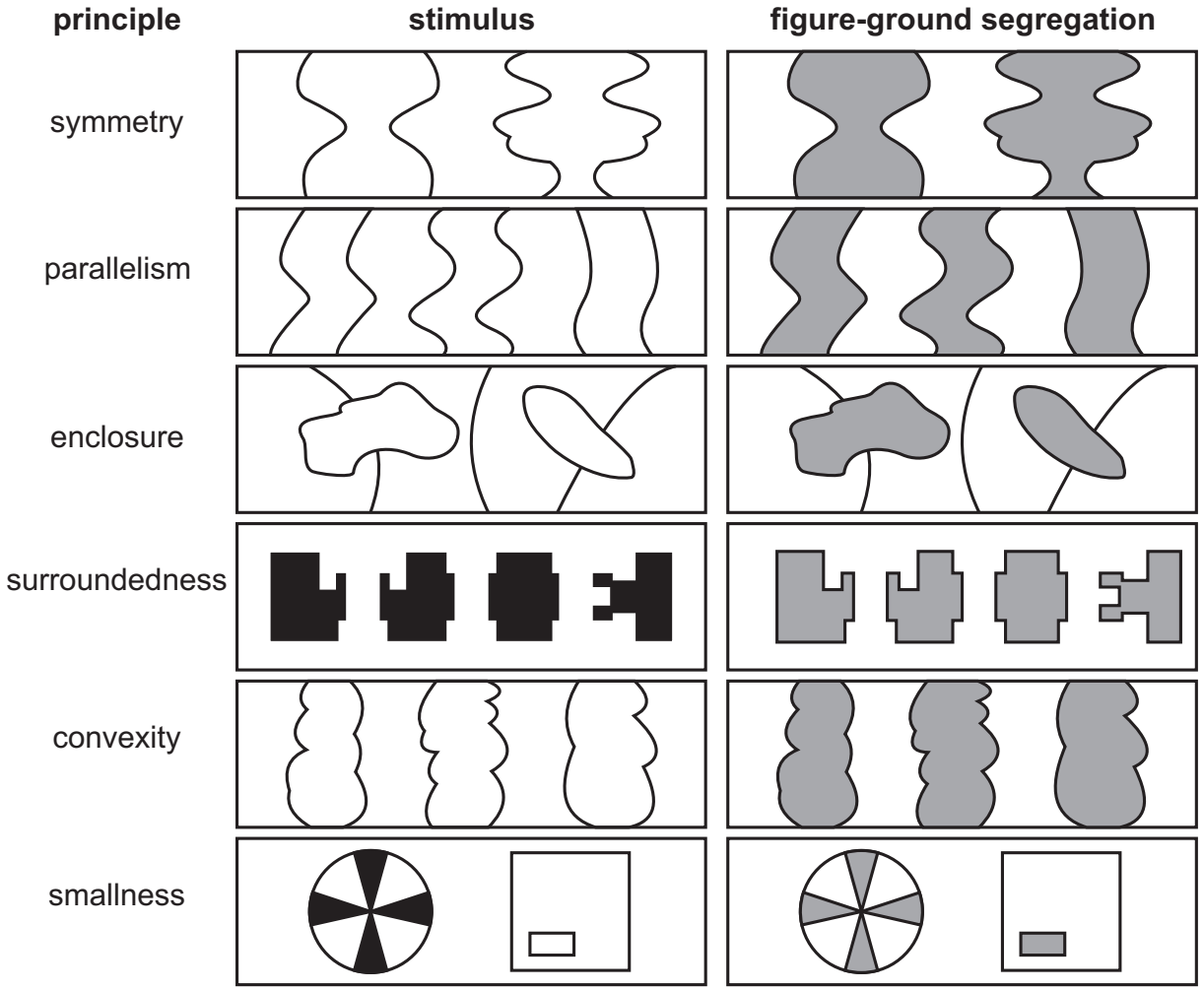
From parts to wholes

- How individual elements are grouped into wholes is studied in the Gestalt theory
- Gestalt principles
  - Grouping principles
    - How are things grouped?
  - Figure-ground segregation
    - How are object separated from the background?
- Correspond to the configural features

Gestalt: Grouping



Gestalt: Figure-ground segregation



Gestalt in Computer Vision

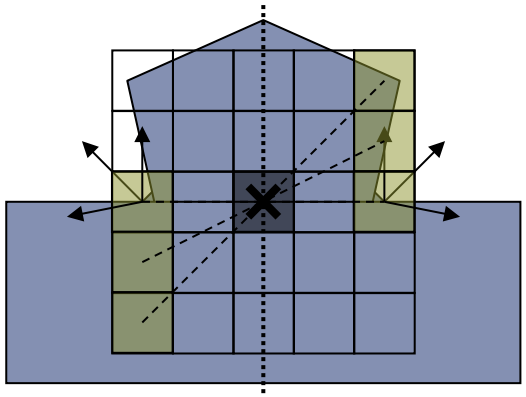
- Can we use this in computer vision?

- Yes, for instance symmetry:
  - In visual-attention models to predict human eye fixations (Kootstra '08/'09)
  - In robotic vision to find landmarks (Kootstra '09)
  - In object detection and segmentation (Kootstra '10)

Symmetry-saliency model
(Kootstra ’09)

Symmetry-saliency model

- Symmetry as a salient feature



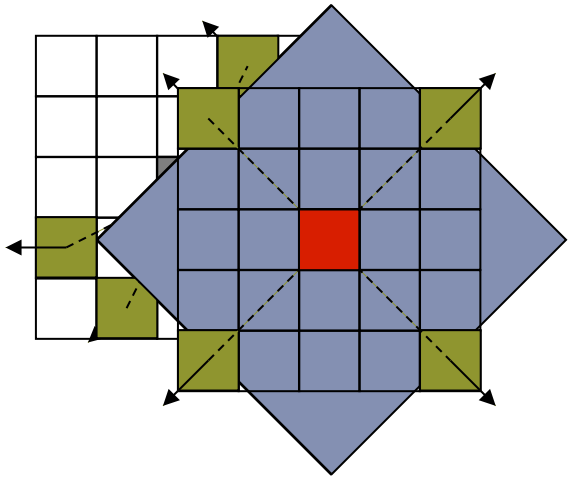
Do this for all pixel pairs in kernel

$$s_{ij} = \left(1 - \cos\left(\gamma_i + \gamma_j\right)\right) \cdot \left(1 - \cos\left(\gamma_i - \gamma_j\right)\right)$$

$$S(p) = \sum_{ij \in \Gamma} s_{ij} \cdot w_{ij}$$

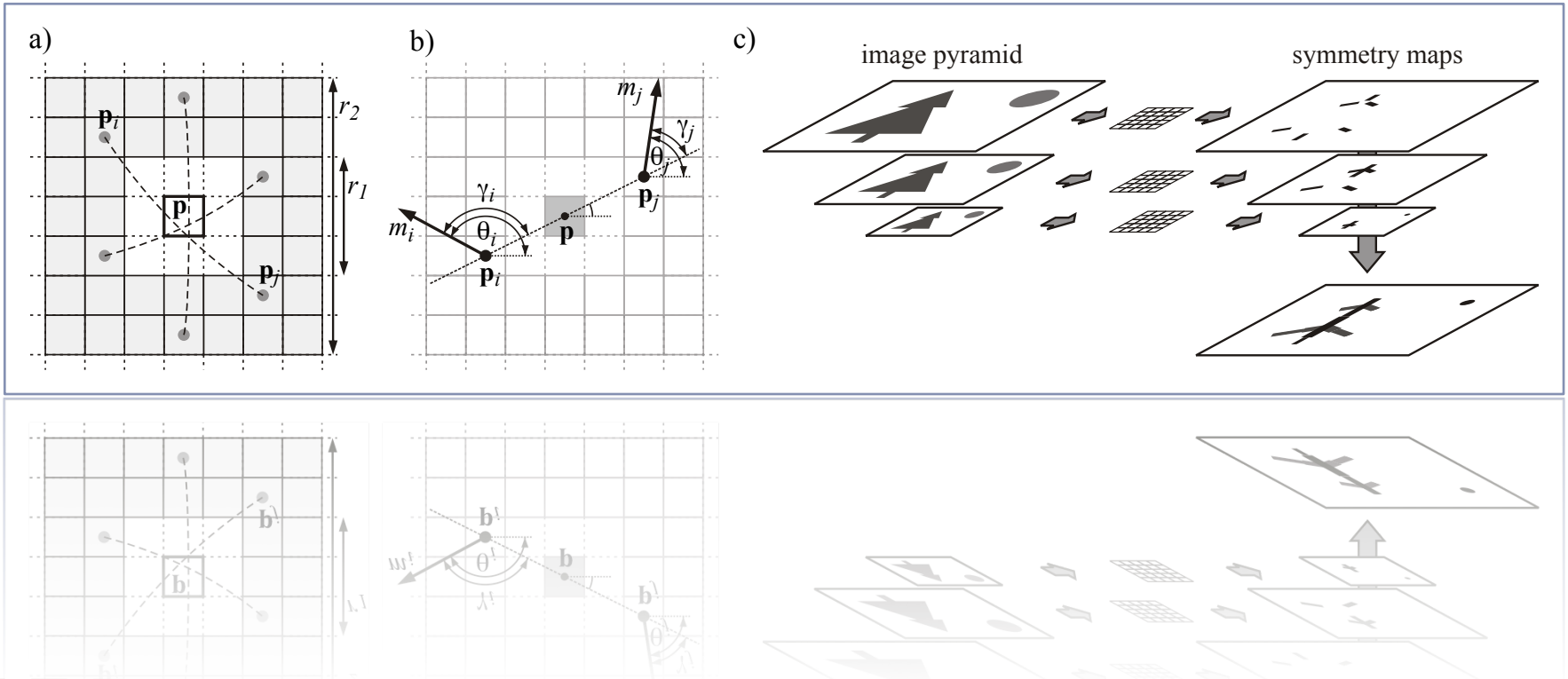
Symmetry-saliency model

- An example



Multi-scale symmetry model

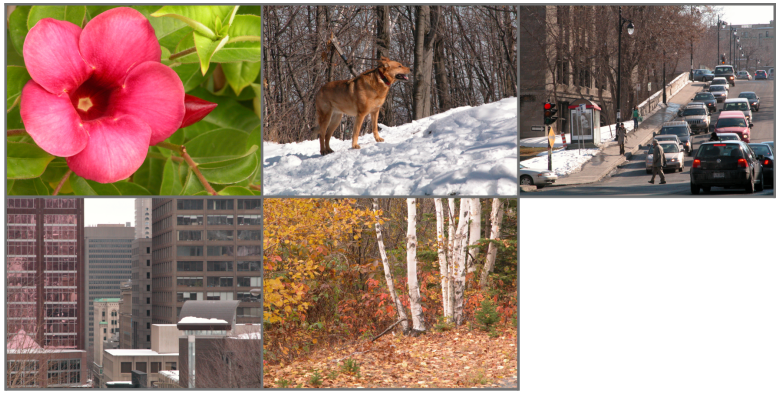
- The responses at different scales are summed up to obtain the symmetry-saliency map

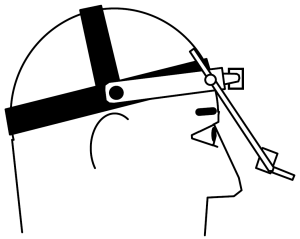


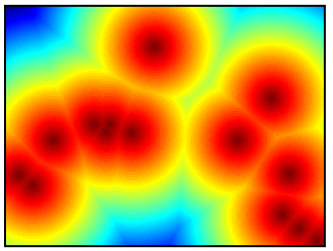
Prediction of eye movements using symmetry
(Kootstra ’09)

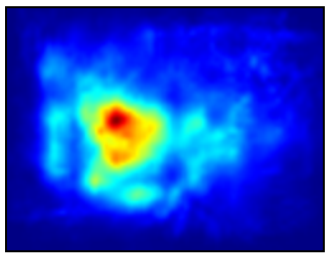
Compare to human eye fixations

- 31 participants
- 100 images
- Compare maps



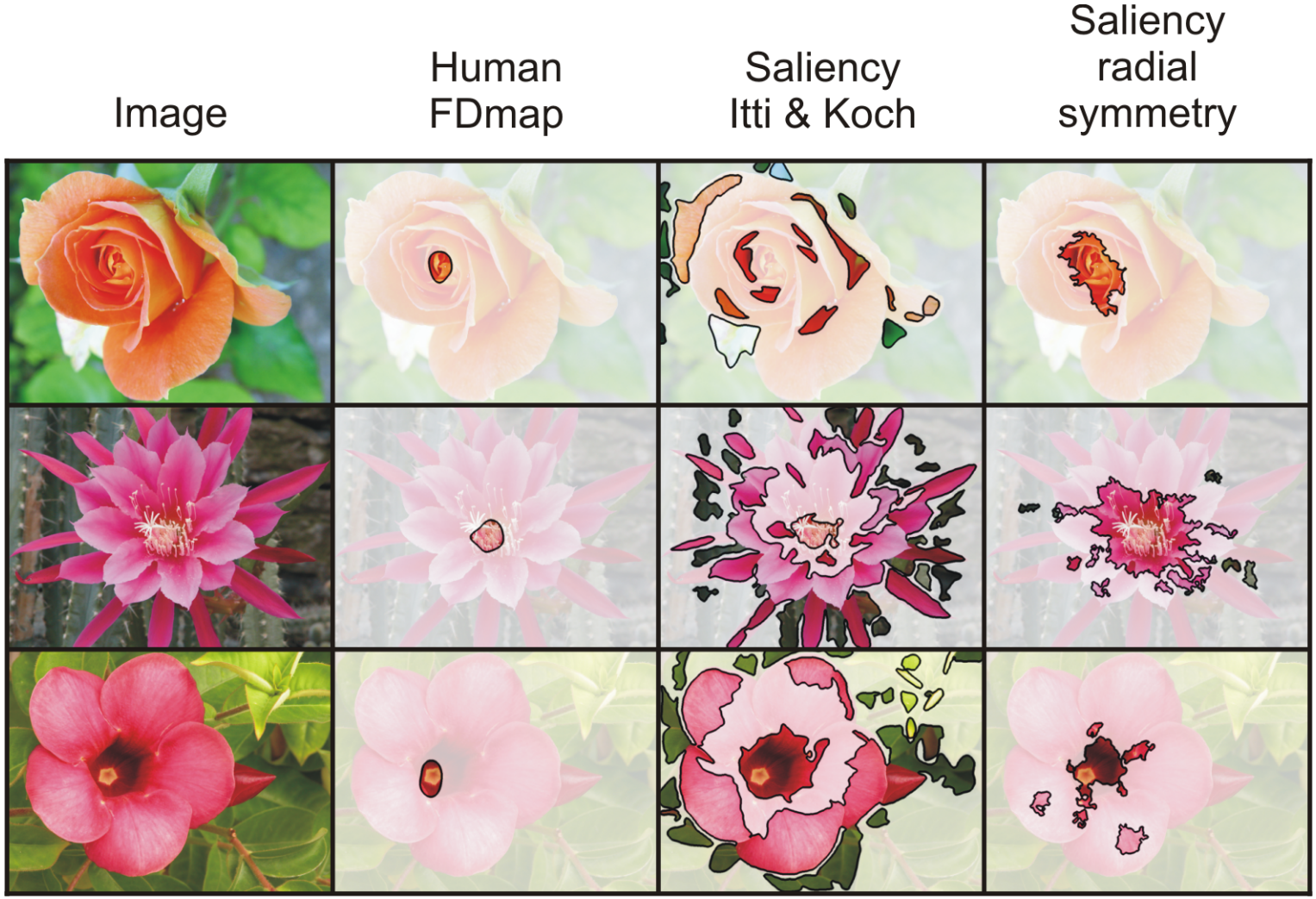






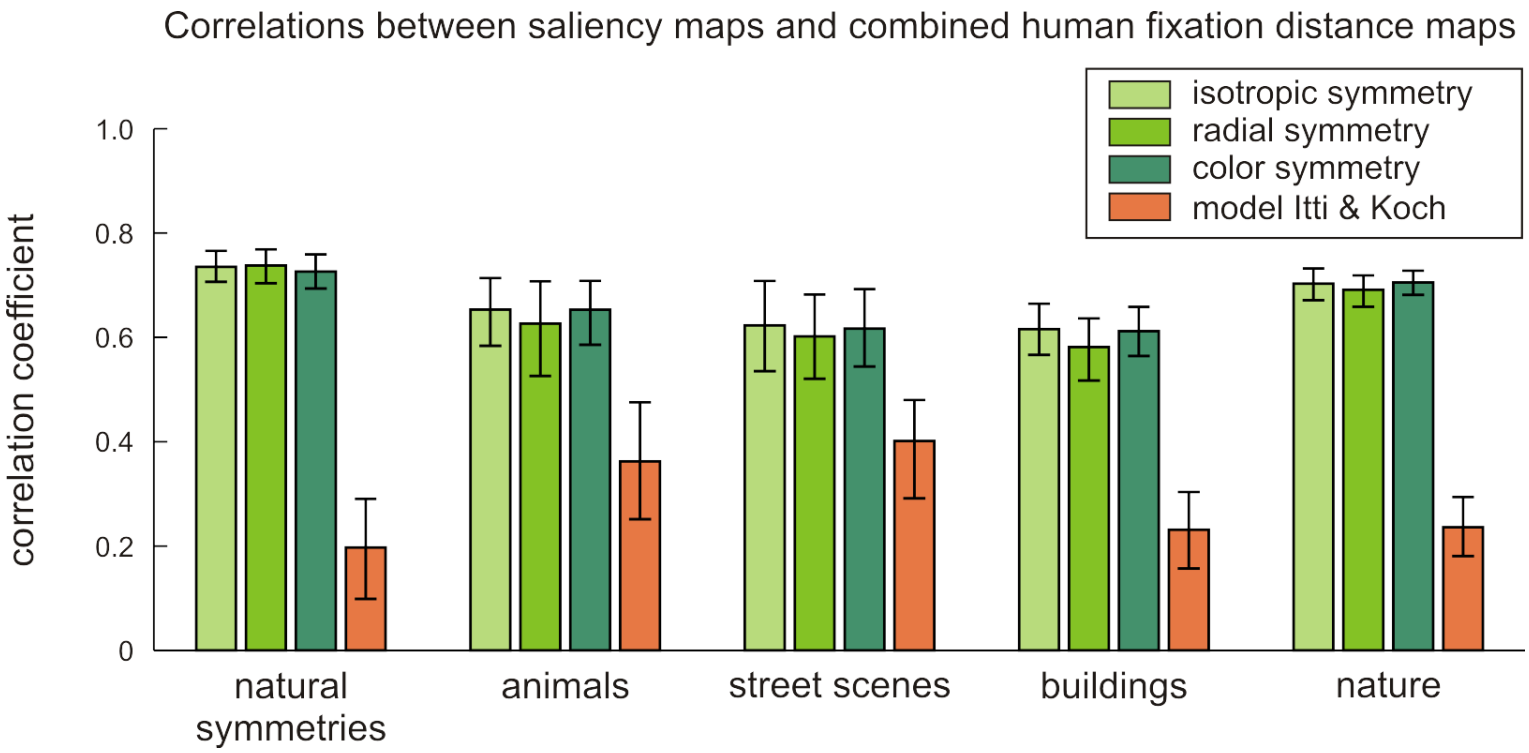
Correlation

Symmetry-saliency model: Results



Symmetry-saliency model: Results

- Human eye fixations are better predicted using symmetry than using center-surround contrast



Symmetry-saliency model

- What can we do with a computer model that predicts human eye fixations?
  - Human-machine interaction
  - Human assistance
  - In advertisement
  - To check if things attract attention (traffic signs)
  - …
- Same model used to guide robotic attention

Conclusion and Discussion

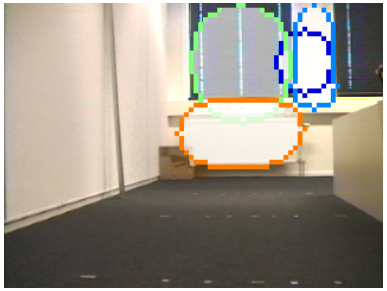
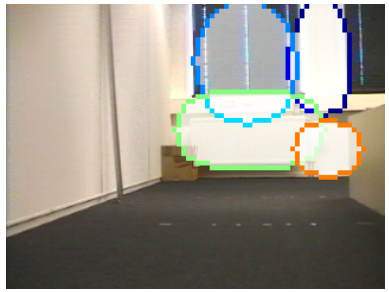
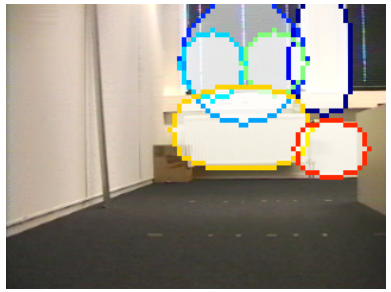
- Good correlation of symmetry saliency model with human data
- Symmetry correlates better than contrast
- Amount of symmetry at fixation especially high for early fixations
- Suggests that symmetry is a salient feature

- Symmetry as a bottom-up cue for object detection

Selection of landmarks for visual SLAM
(Kootstra ’09)

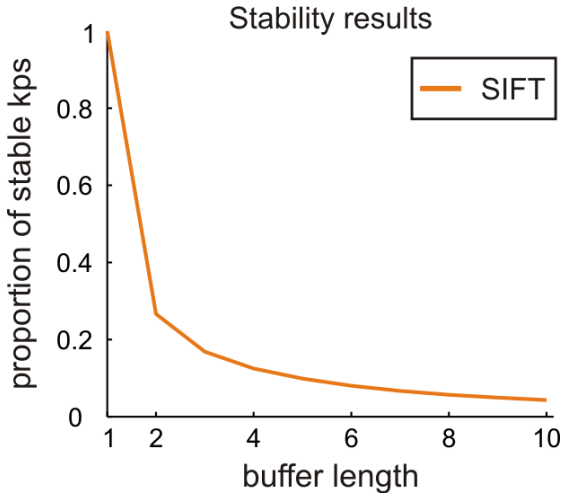
Objective

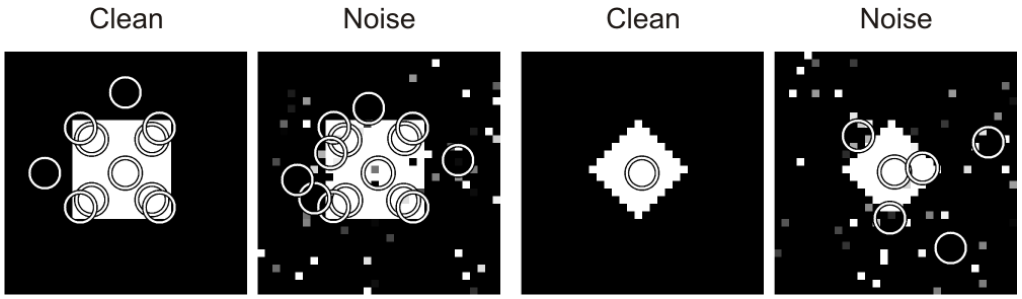
- Improving the quality of landmark detection for visual SLAM
  - Find stable visual landmarks robust to noise, changes in illumination, and changes in perspective

Common approach

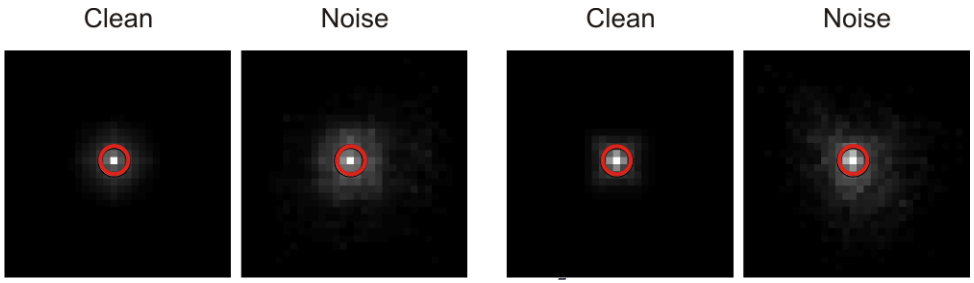
- Detection of interest points
  - Usually based on contrast: SIFT, SURF, etc.
- Problem SIFT for landmark selection
  - Many interest points, many unstable
  - Susceptible to noise





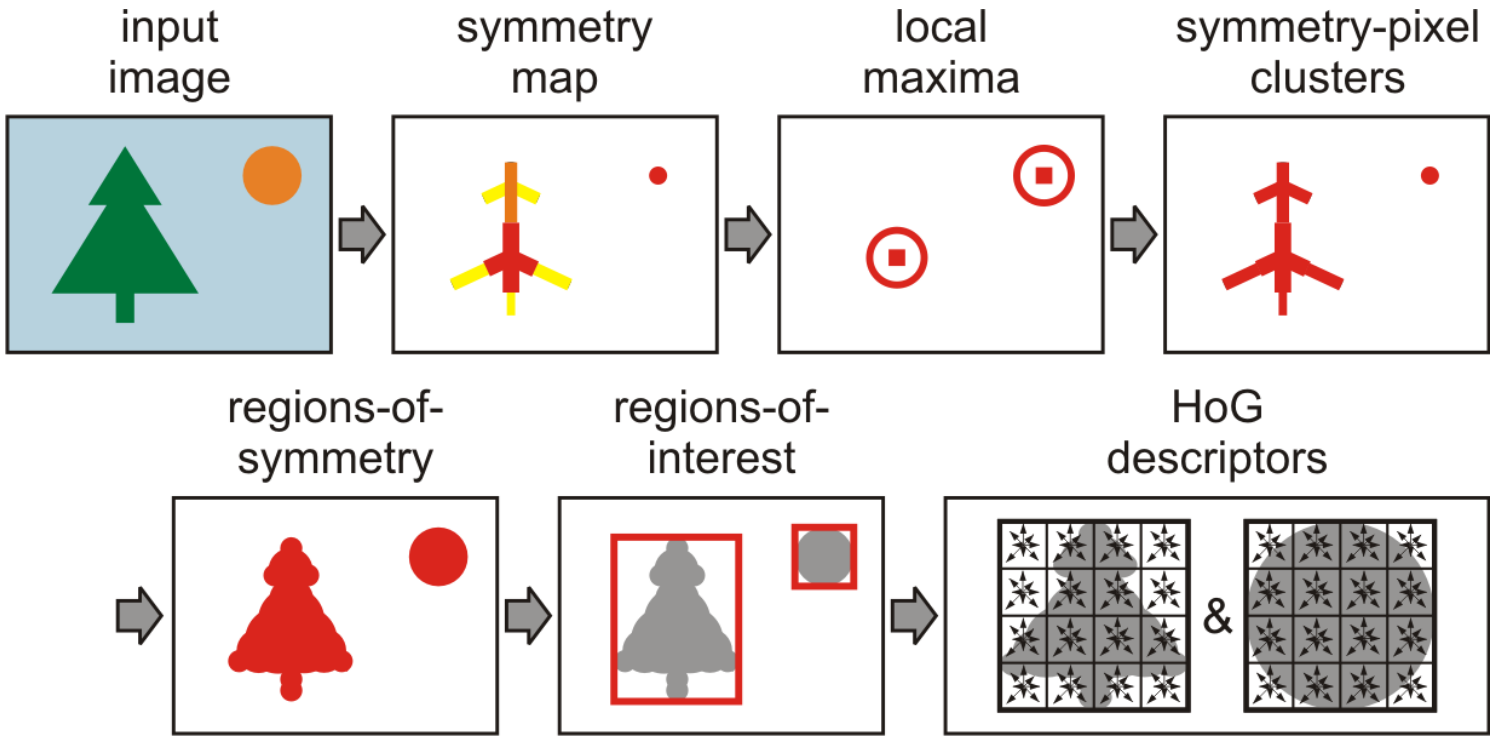
Our proposal

- Using local **symmetry** in the image to detect regions of interest



- Symmetry
  - More robust to noise sym. pattern
  - Symmetry cue for FG segregation
  - Indoor env: many sym. Patterns

SymRoID

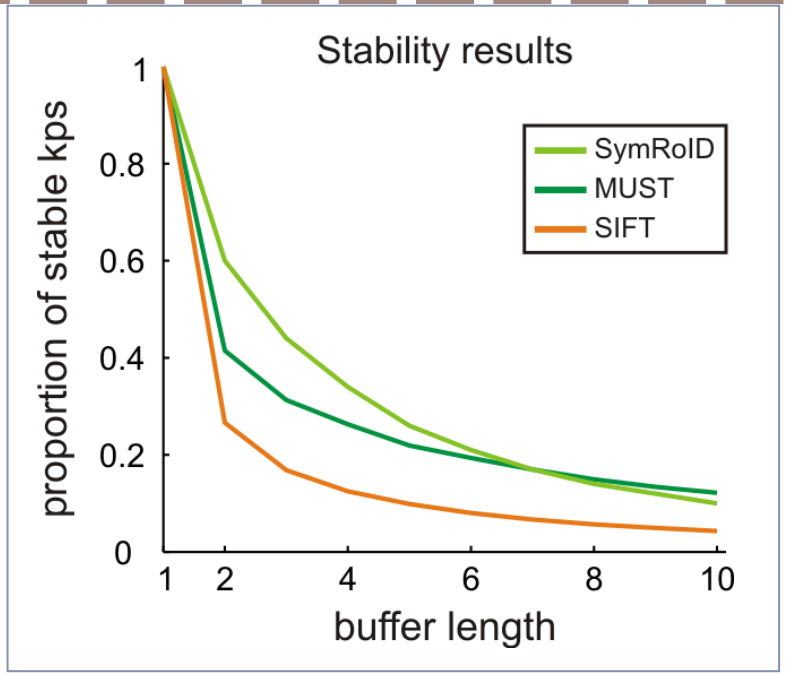


- From symmetry map to regions of interest
- Descriptor relative to size of region (more affine invariant)

Experiment 1: Stability

- Points/regions are tracked over a number of frames.
- Measure: the proportion of regions that are found in all frames

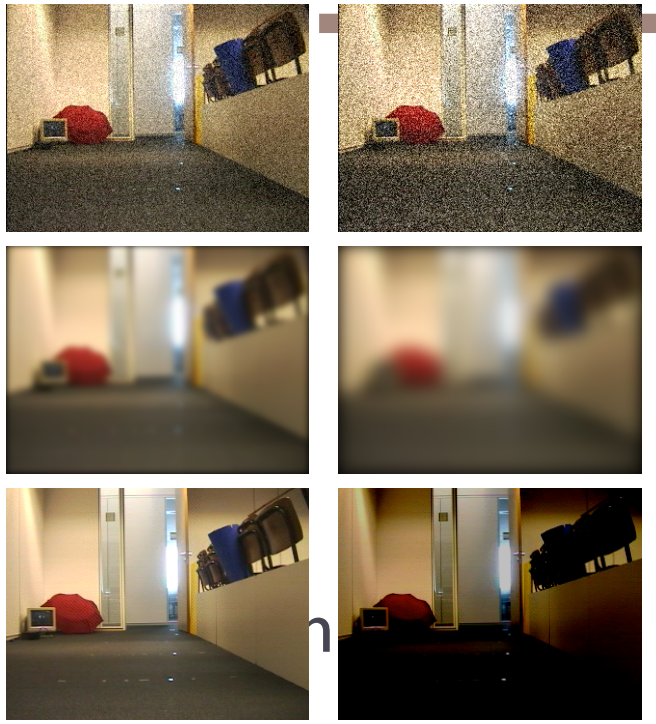
Results 1: Stability



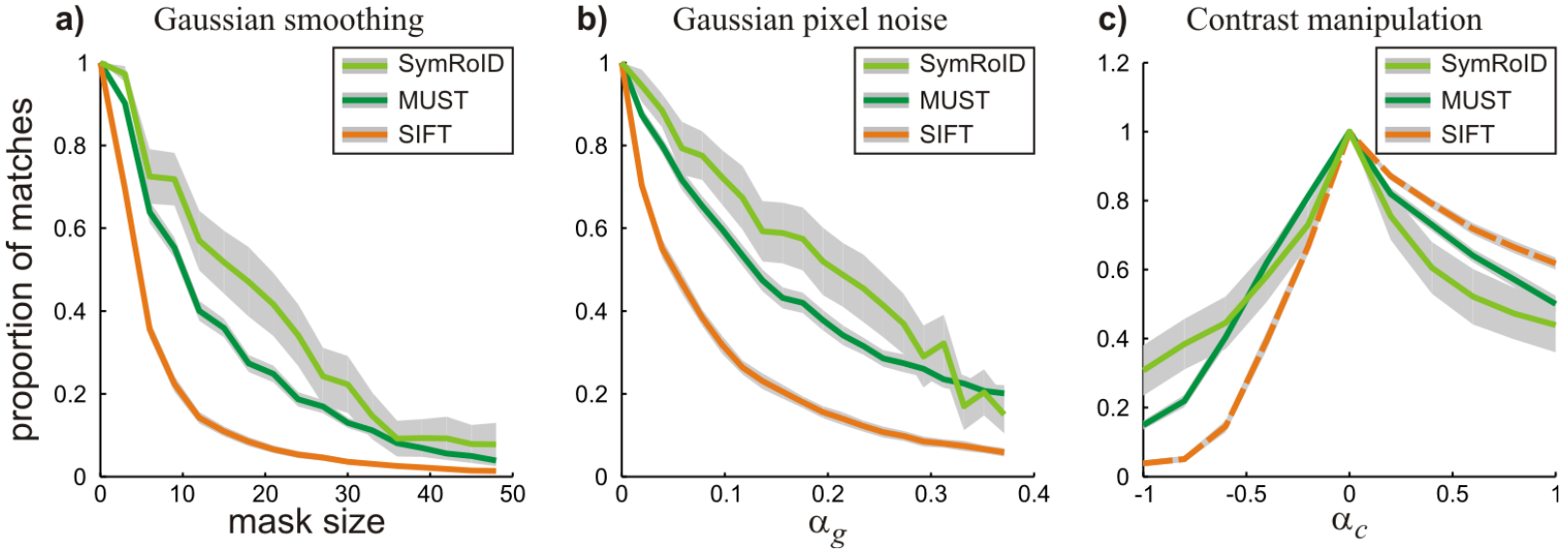
- Use of symmetry results in better stability over the sequence
- Regions are more stable than points
- Deals better with small changes in perspective

Experiment 2: Robustness

- Robustness
  - Noise
  - Light condition
  - Measure: proportion of points [illegible] original image



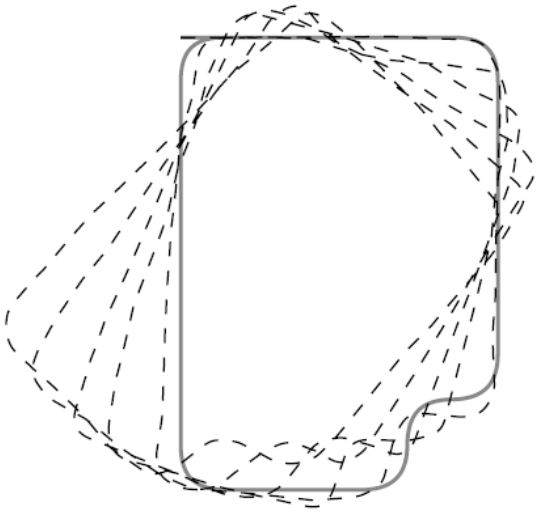
Results 2: Robustness



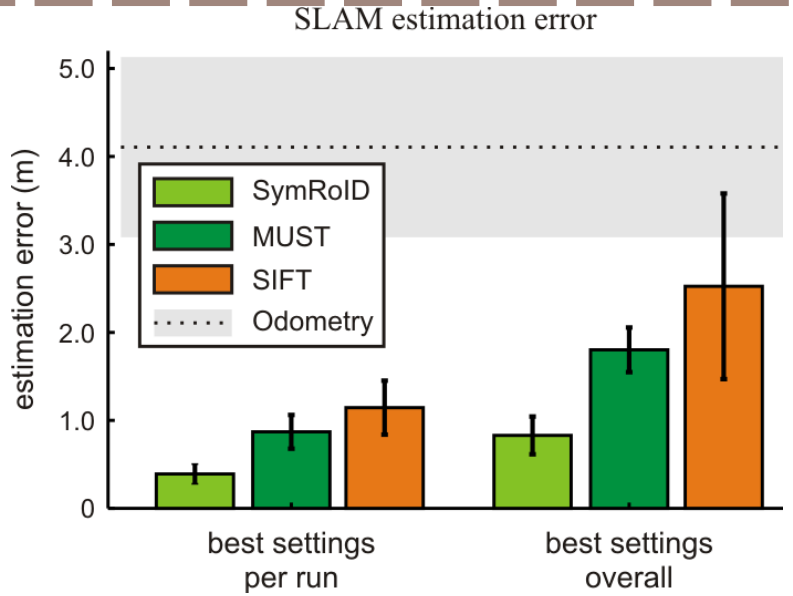
- Symmetry results in better noise robustness
- Better robust to a decrease contrast, not to an increase
- Regions are more robust than points

Experiment 3: SLAM performance

- Pioneer II robot
- 10 runs through an office environment
- Much odometric noise
- Quantitative analysis: hand labeled ground-truth
- Each run, 4 rounds
  - 3 to establish map with KF
  - 4th tot test the performance



Results 3: SLAM performance



- Buffer settings and matching parameters optimized

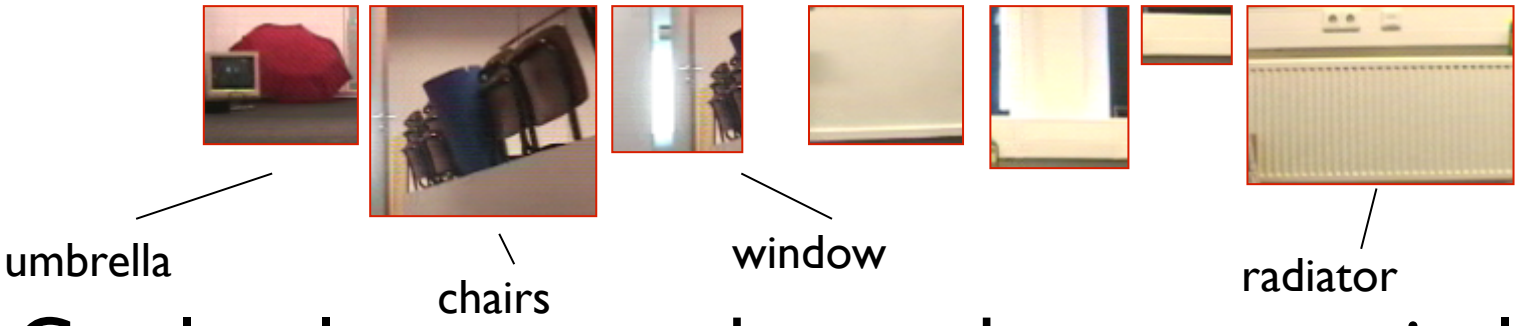
- Symmetry results in better SLAM performance
- The use of symmetry *regions* outperforms the rest

Conclusions / Discussion

- The use of local symmetry for SLAM:
  - Better stability, more invariant to perspective changes
  - Better noise robustness
  - Better landmarks ➔ better SLAM perform.
- Problem:
  - Contrast enhancement

Conclusion/Discussion

- SymRoID segmentation sometimes corresponds with semantic entities (Gestalt principle)

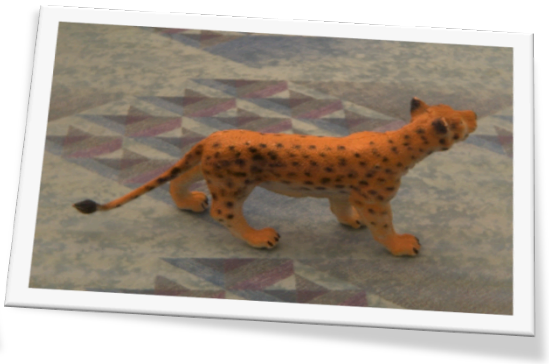


- Can local symmetry be used to automatically segment objects?

Bottom-up detection of objects
(Kootstra '10)

Objectives of study

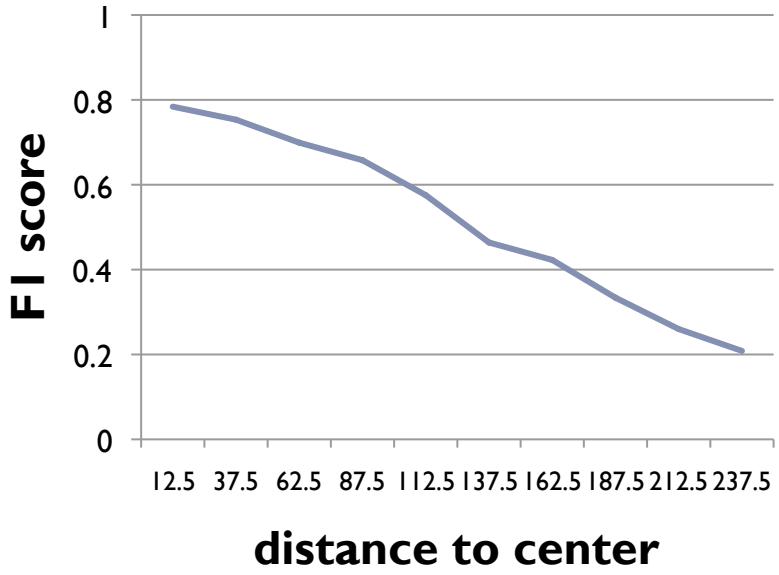
- Objectives
  - Detect unknown objects in the environment
  - To initialize and automate object segmentation
- Why symmetry
  - A bottom-up cue for the presence of an object
  - Focus on object's center



Symmetry for Fixation Points

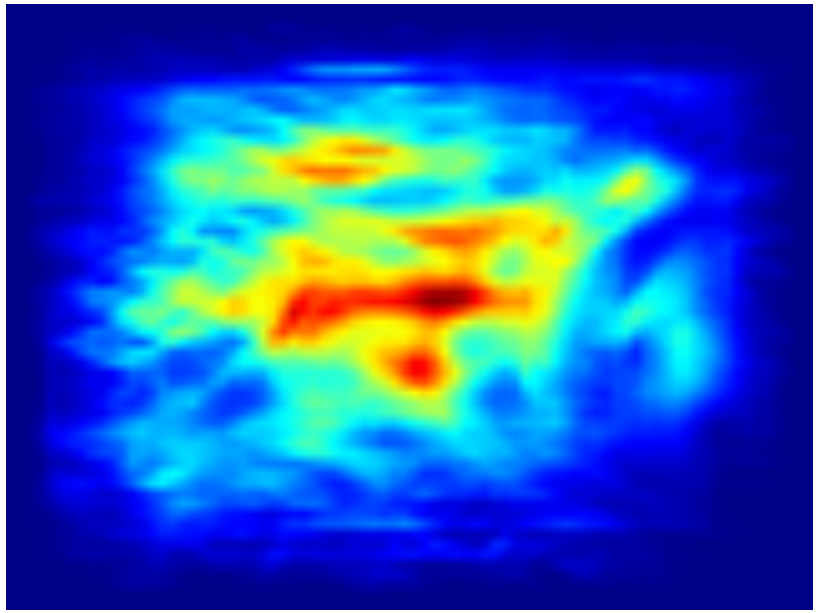
- Many object segmentation methods assume a fixation point to initiate the segmentation
- We automate the fixation-point selection
  (Kootstra, Bergström, Kragic, ICPR'10)

- The location of the fixation point is important
- Segmentation is best when points are close to center



Symmetry Saliency

- Symmetry-saliency map



From saliency map to fixation point

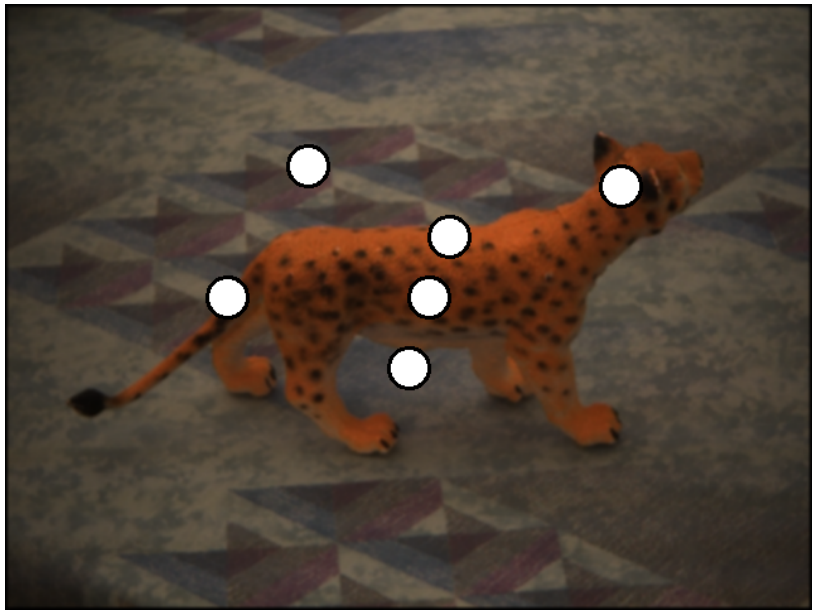
- An iterative procedure with inhibition-of-return (IOR) to generate fixation points

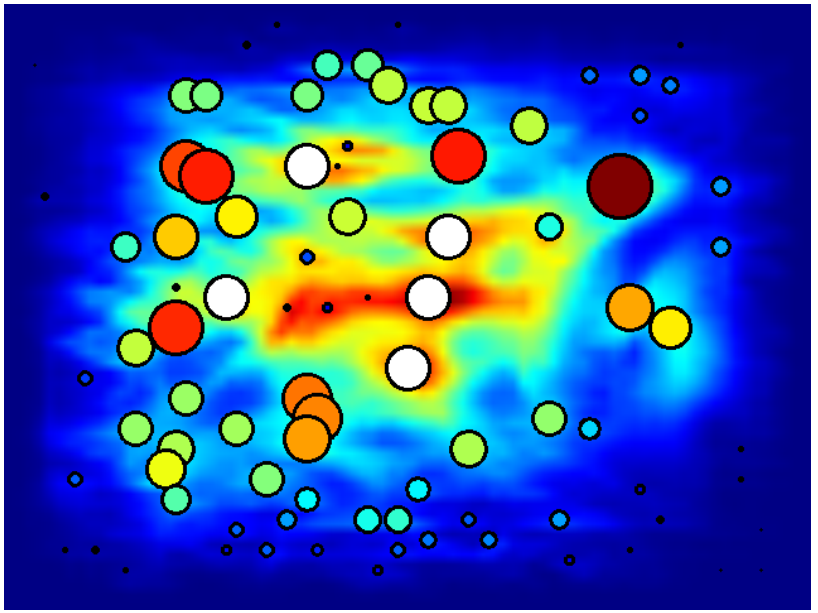
1. Get all local maxima
2. While fixation-point-needed
    1. Get highest local maximum, f_i
    2. Find salient area belonging to f_i
    3. Reduce all local maxima in that area (IOR)

Inhibition of Return

- Usually Gaussian suppression of saliency map
  - Produces new local optima on near boundaries of the Gaussian kernel
  - Might suppress fixation points on neighboring objects
- We suppress existing local maxima belonging to the same symmetric area

From saliency map to fixation point



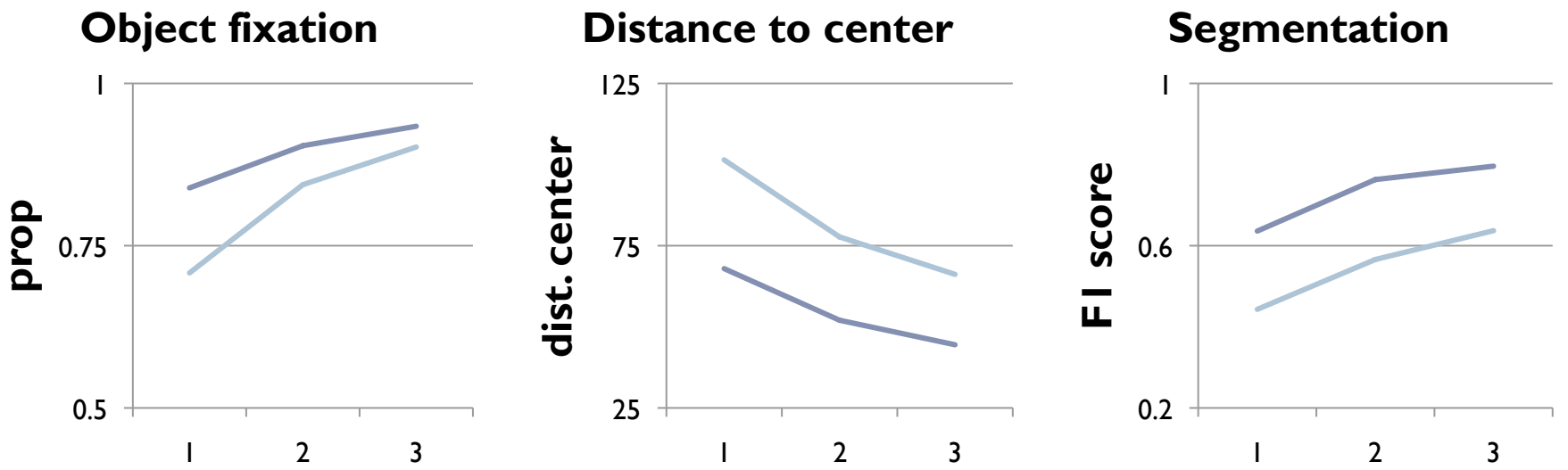


Symmetry to start segmentation

- Bottom-up segmentation methods need a fixation point to start the segmentation
- Fixation points selected using symmetry are good candidates for object segmentation

- Active segmentation method (Mishra et al 2009)
- Later our own method

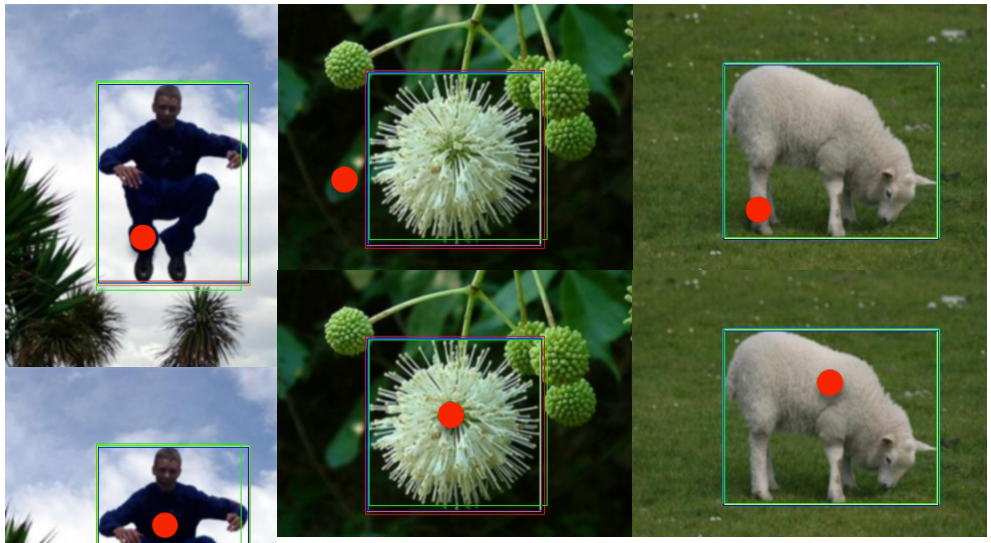
Object Detection Results



- Using symmetry:
  - Better salient object detection
  - Fixation points closer to the center
  - Results in better object segmentations

Symmetry
Contrast

Object Detection Results



Contrast

Symmetry

Symmetry

Contrast

Conclusions - segmentation

- Using local symmetry is more successful in detecting objects than contrast
- Fixation points are closer to the objects center
- Object segmentation is more successful

- The symmetry-saliency method is fast
  - 100-200 ms on a CPU
  - 5-10 ms on a GPU

Back to visual attention

From parts to wholes

- Hierarchy
  - Color, brightness
  - Lines, edges
  - Forms
  - Scenes
- Higher interpretations subsume lower ones

Change blindness

- What changes in this image?

- This
- basic
- Failu
  - Eng
  - Res

Count ball passes of white team

Gorilla in our Midst

Inattentional Blindness

- Very present stimuli can be unnoticed if the task puts attention on something else
- The visual system does not process all available information, but focuses attention
- Failure or feature?
  - Should we really process all information?

Visual Attention

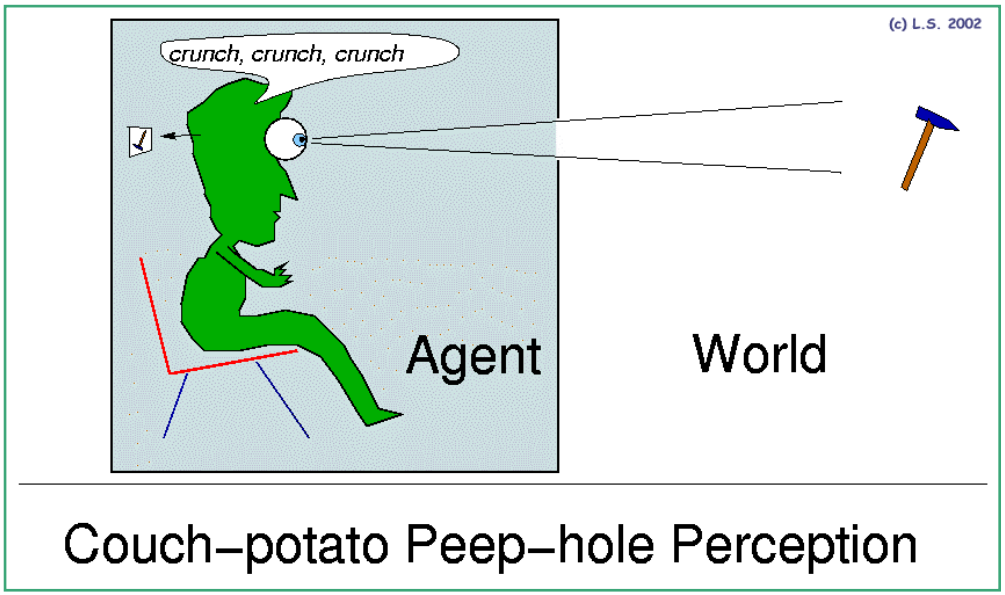
- Bottom-up attention
  - Basic features
  - Configural features
- Top-down attention
  - Task
  - Context
  - Experience

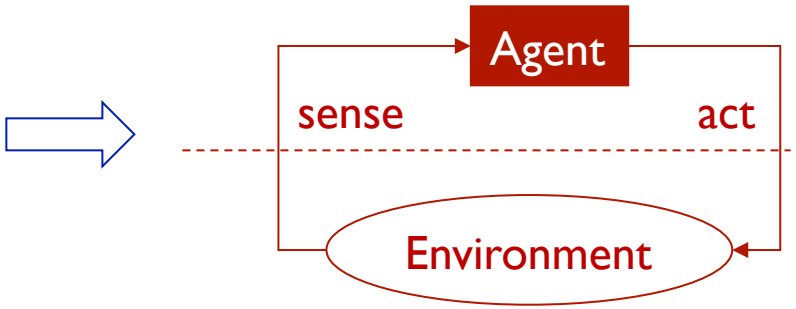
- A good understanding of visual attention is very helpful for machine vision
  - Faster search, detection, recognition, …

Active vision

Active Vision in Natural Systems

- Vision in biological systems is an active process
  - No couch-potato peep-hole perception, but
  - Interaction with the world





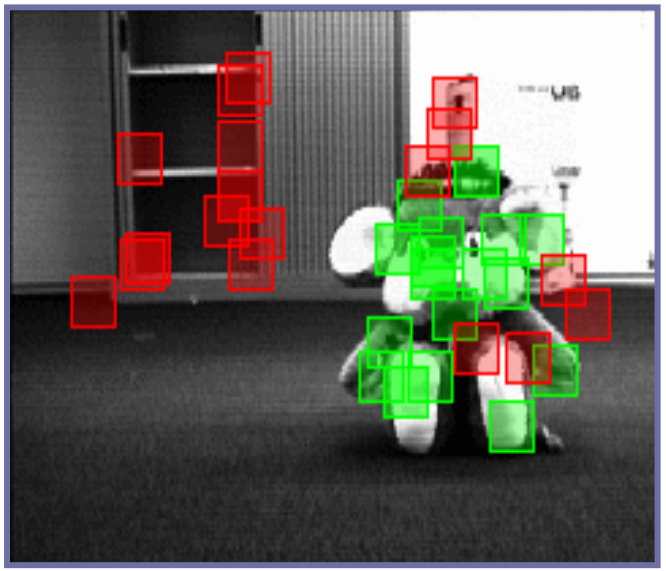
Active Vision in Natural Systems

- Active vision
  - Deliberately move the body to acquire new sensory information
- Two examples
  - Locust moves it head to estimate depth
  - My nephew explores new toys to get more info

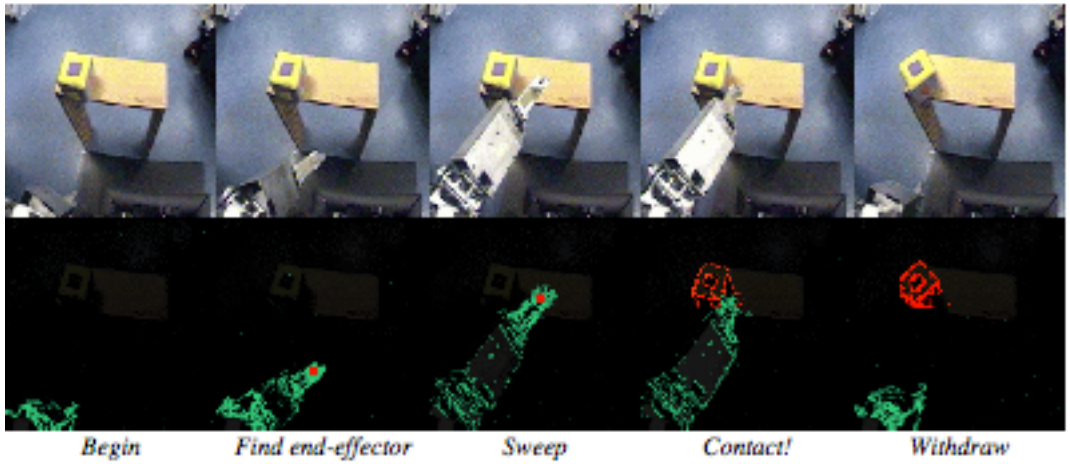
Active Vision in Machines

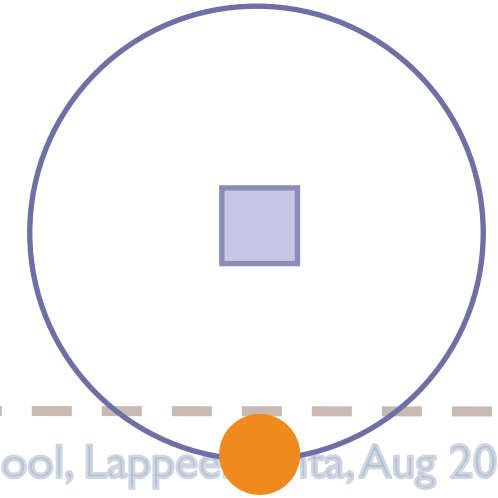
- Active object exploration
  - For segmentation and learning

Kootstra et al 2008



Fitzpatrick & Metta 2002





Biological and Machine Vision

Biological and Machine Vision

- There is still much unknown about biological vision, but…
- Multi-disciplinary research
  - Machine vision can get inspiration from biology
    - E.g., in visual attention
  - Biological theories can be tested in machines
    - E.g., use of symmetry in vision

What have we learnt?

- The human eye
- Visual pathways
- Center-surround organization
- Some visual illusions showing contrast effect and use of contextual information
- Visual attention
  - Bottom up
  - Top down
- Gestalt theory
- Vision is an active process

Take-home message

- Get inspired by biological systems