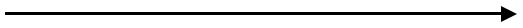
Active Vision

Key points:

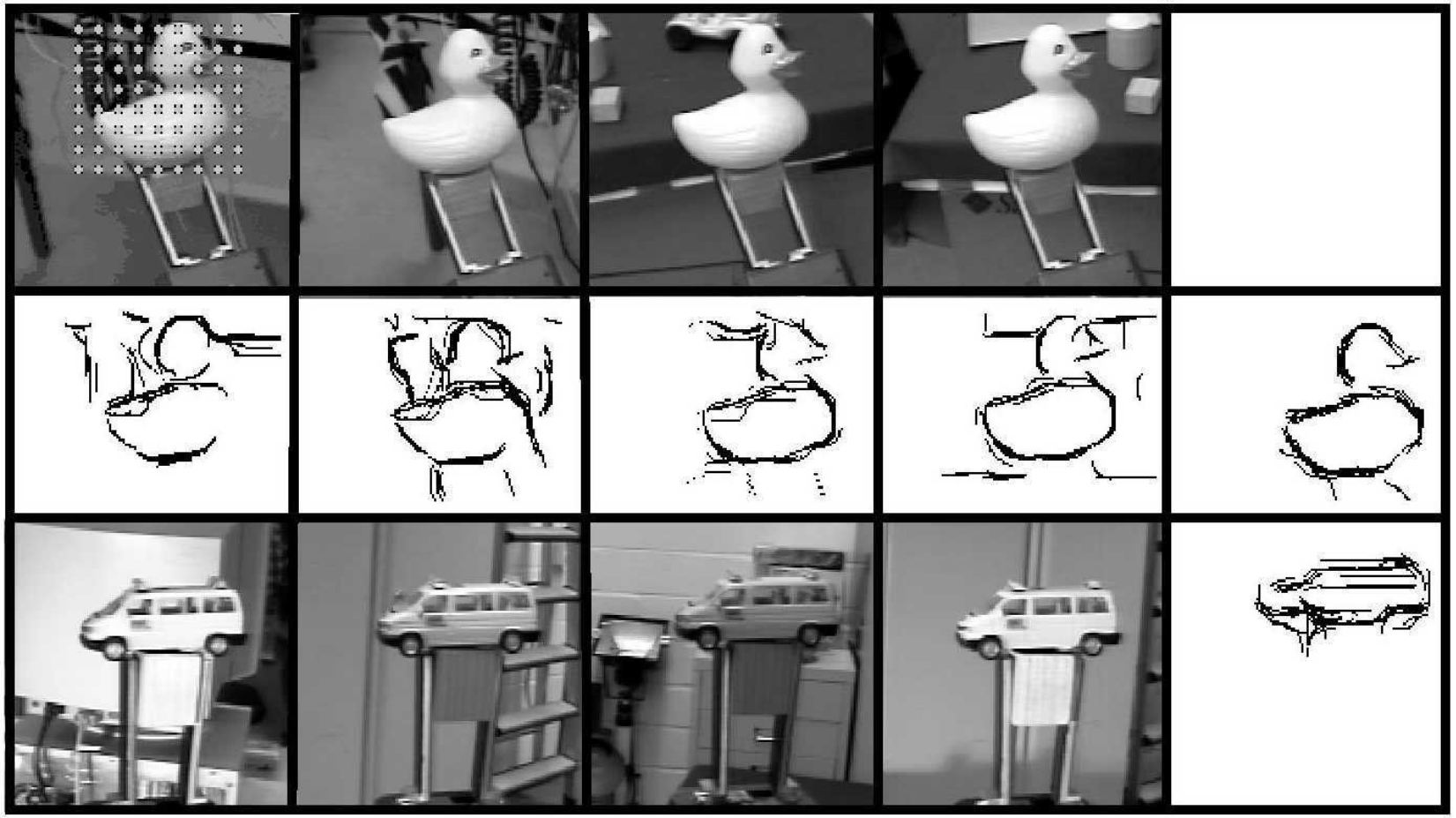
- Acting to obtain information
- Eye movements
- Depth from motion parallax
- Extracting motion information from a spatio-temporal pattern
- Obtaining structure from motion

The importance of activity

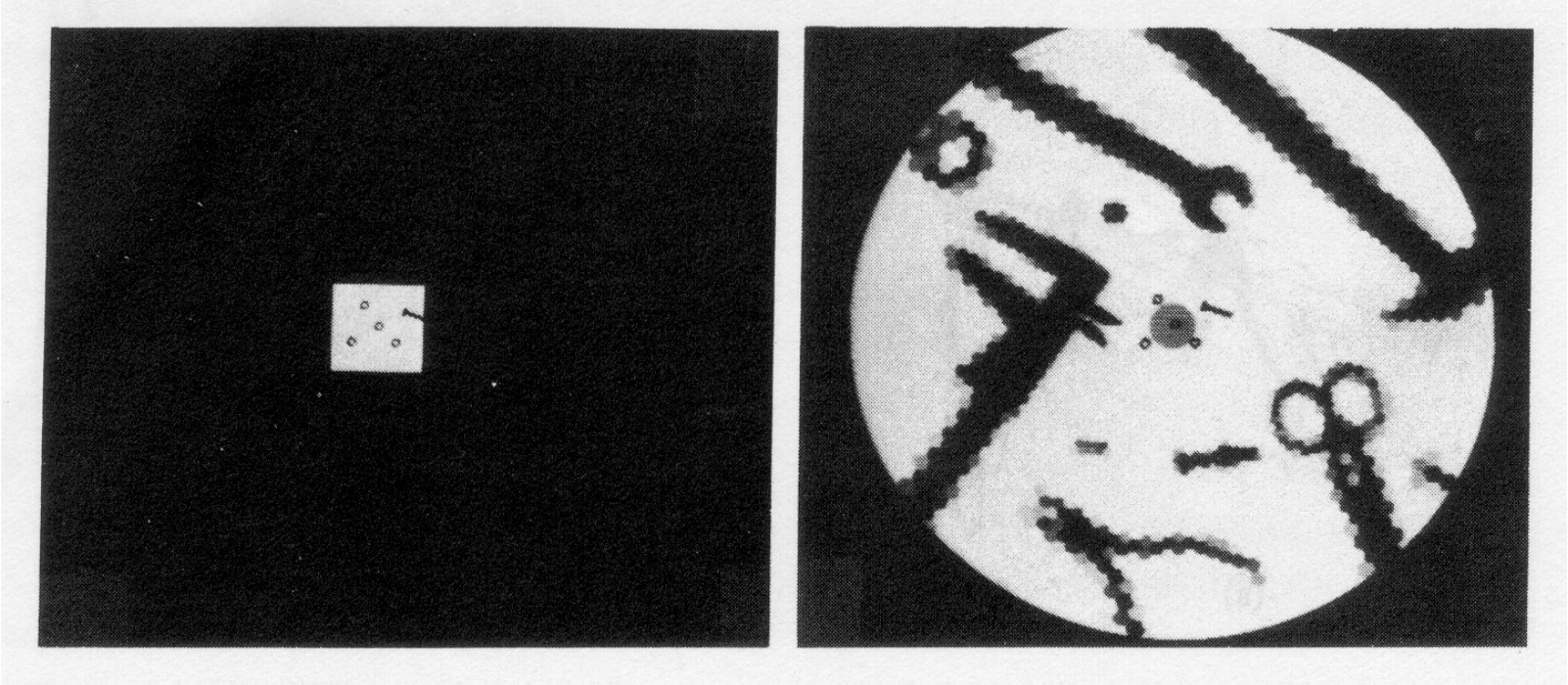
- Computer vision is often approached as a problem of passive extraction of information from single images
- But most natural vision systems try to actively sample the visual scene
- Can solve some problems with active system that are hard with passive system



Segmenting (Kruger, 1998)



Improving resolution



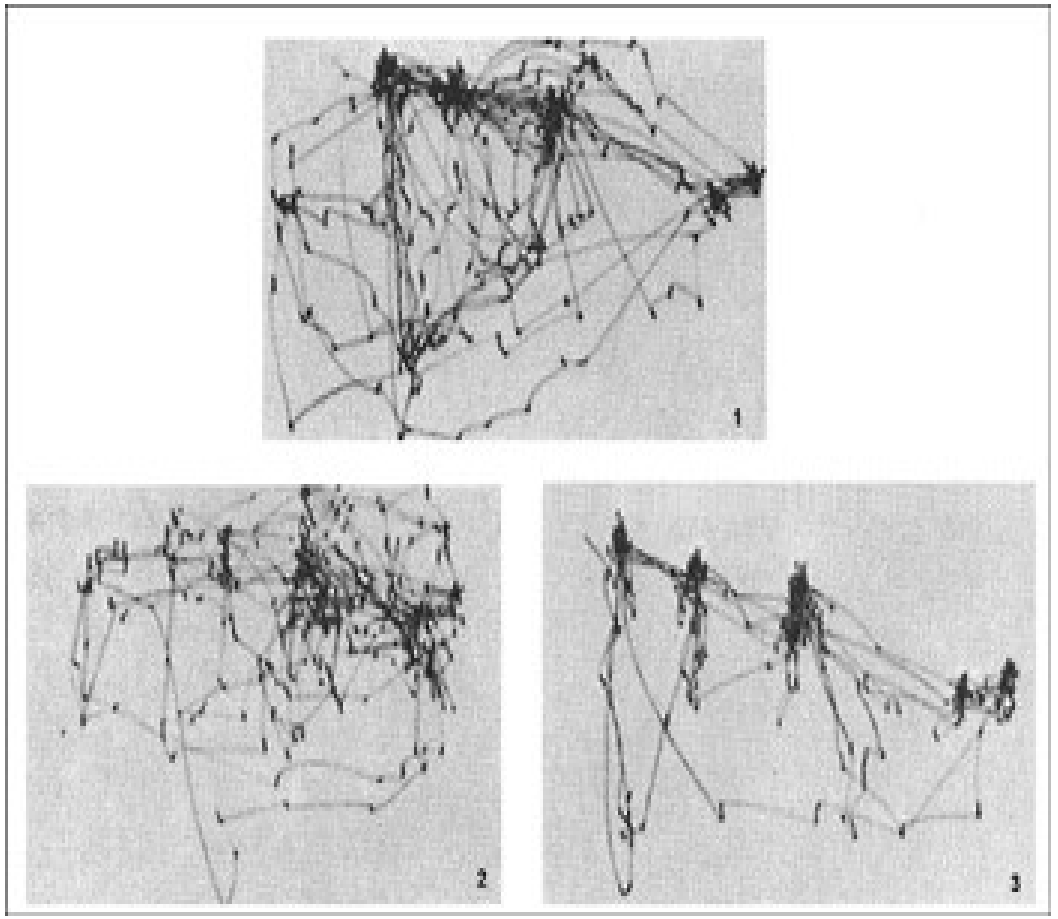
Making best use of fixed number of receptors e.g. 100x100 pixels (Ballard, 1991)

Eye movements

- Increase the effective resolution by *saccade* movements of high resolution area (fovea)
- Creates impression that see complete detailed scene, but this is illusory

Eye movement patterns indicate attention and task

1. Describe room.
2. What was happening before?
3. People's ages.



Eye movements

- Increase the effective resolution by *saccade* movements of high resolution area (fovea)
- Creates impression that see detailed scene, but this is illusory
- Task dependent, indicates attention

Mid-lecture Question

The human eye/brain system?

1 - sees the whole scene

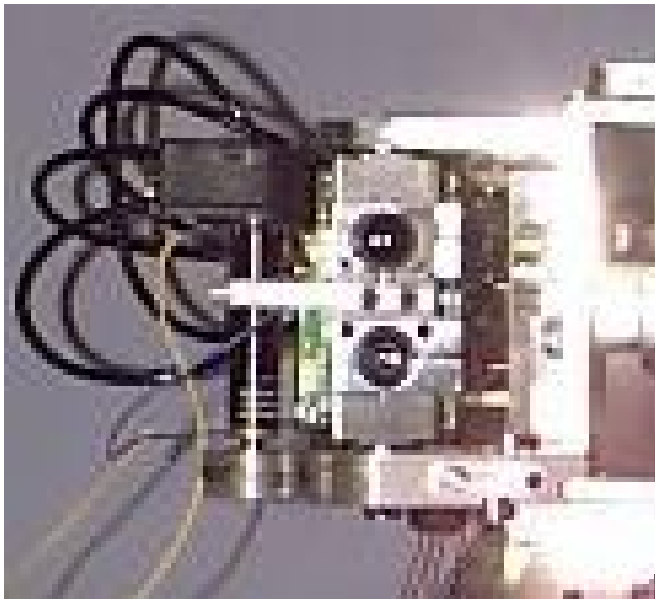
2 - has higher resolution in the fovea

3 - saccades

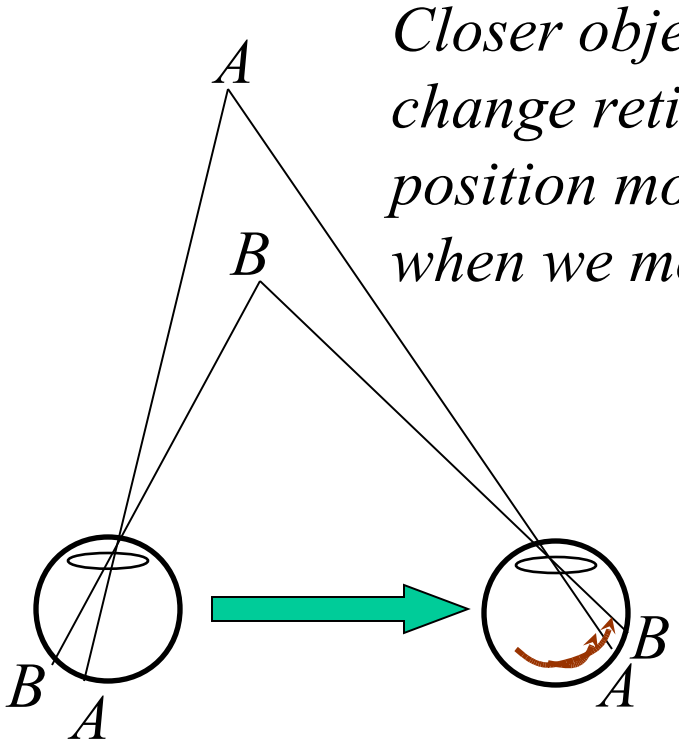
4 - is directed by an attention process

Eye movements and localisation

- Knowing where the eye/camera is pointing tells us the direction of objects of interest (requires proprioception to know relative angles)
- Can also extract depth through motion parallax



Motion Parallax

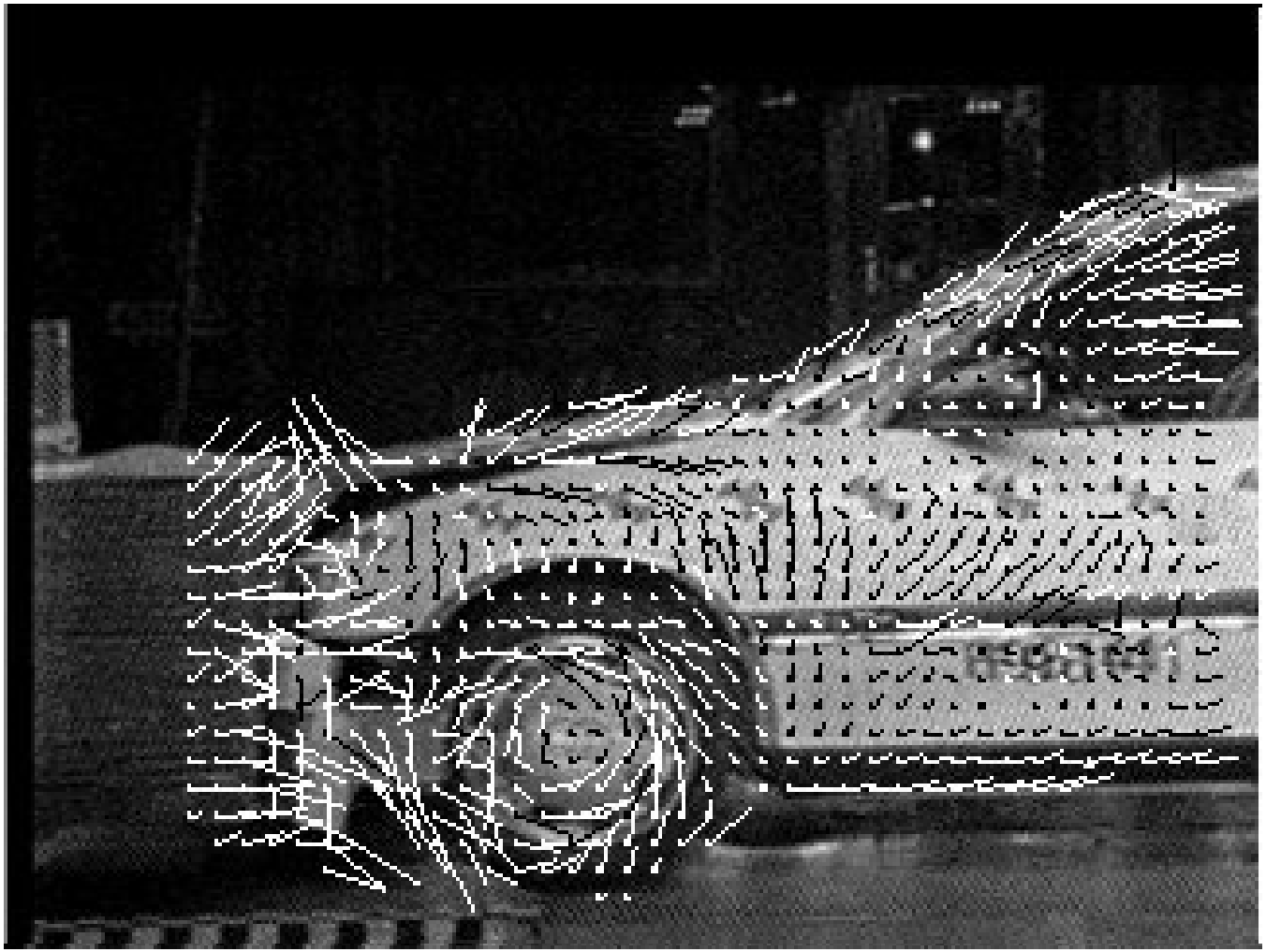


Closer objects change retinal position more when we move

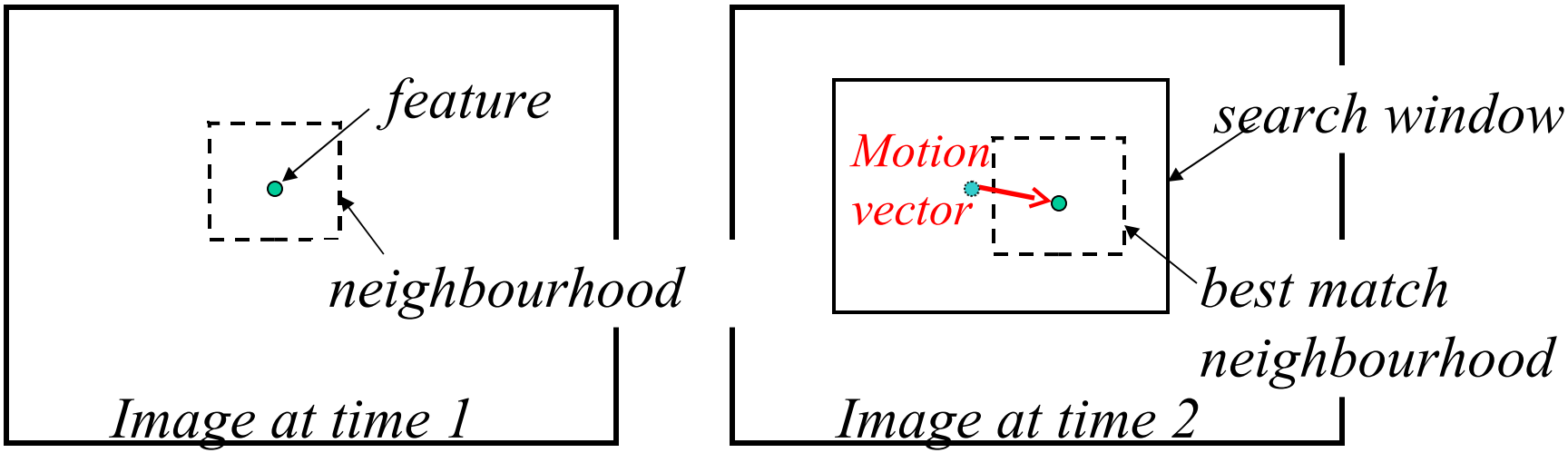
Geometrically equivalent to binocular stereopsis

Motion perception

- Like parallax, much important information about the world comes from sensing visual motion
- E.g. breaking camouflage, sensing self motion, seeing what is happening…
- ‘Active vision’ sometimes taken to mean vision based on sequences of images
- Aim is to extract the flow-field:



Method 1: feature tracking



- Assumes have first extracted features
- Assumes gradual change, and may need to deal with false matches or disappearing features
- In biological systems, motion perception is more fundamental than seeing objects change position

Motion as a perceptual primitive

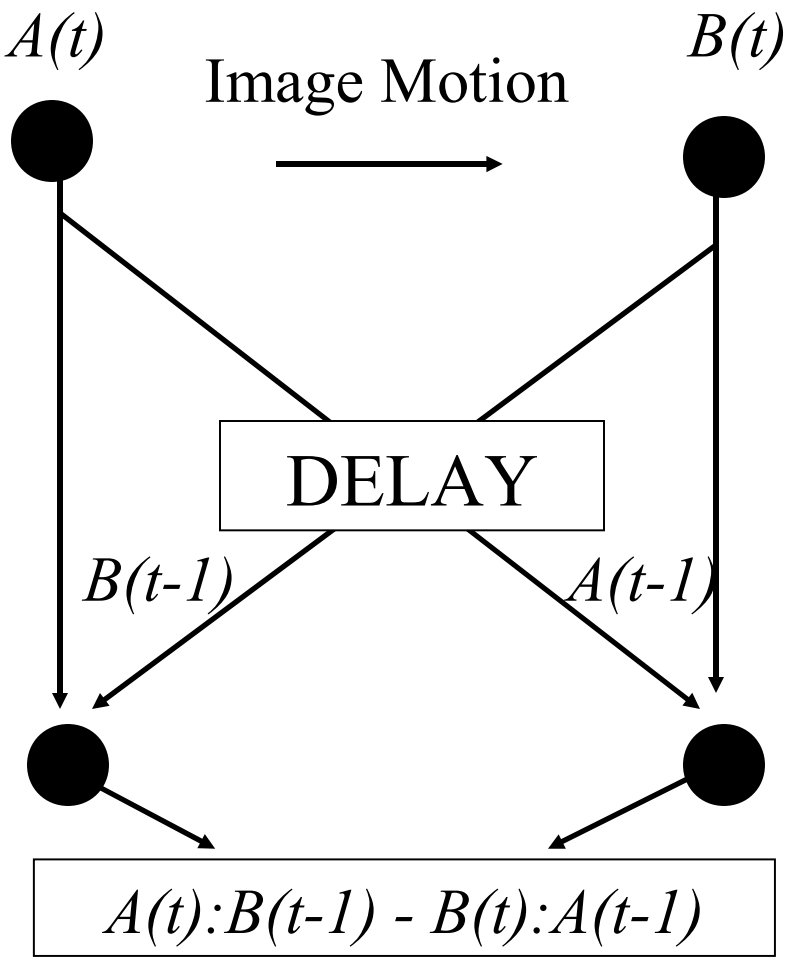
- Can detect motion at faster times and smaller distances than can detect motionless features

- Patients with motion-blindness (Zihl et al, 1983)

- Retinal motion detector neurons

- Visual cortex motion features detectors

Method 1: correlation



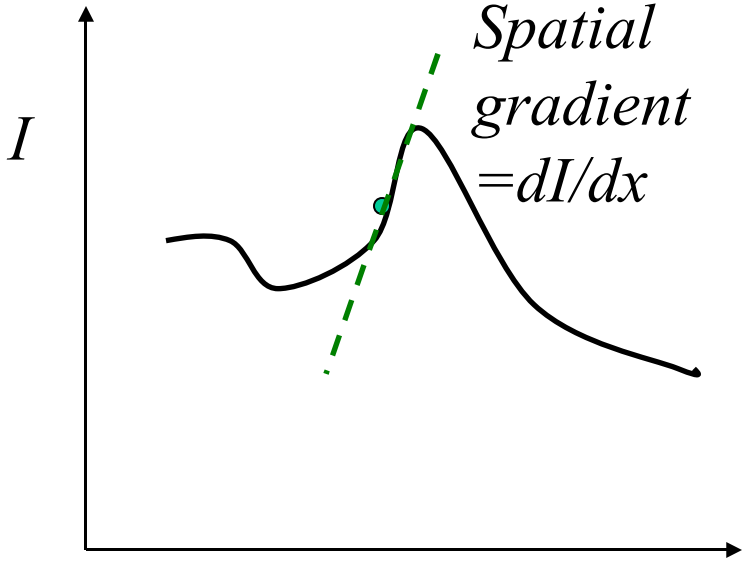
If image is moving from A to B, then A(t) ***will not*** *correlate with B(t-1)*

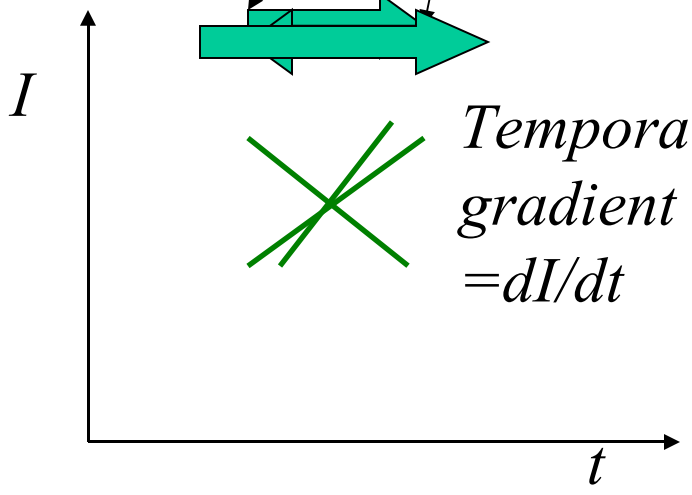
If image is moving from A to B, then A(t-1) ***will*** *correlate with B(t)*

Sign indicates direction of movement

Observed neural circuits in brain region V1

Method 2: gradients





$$V = \frac{dx}{dt} = -\frac{\frac{dI}{dt}}{\frac{dI}{dx}} \quad x$$

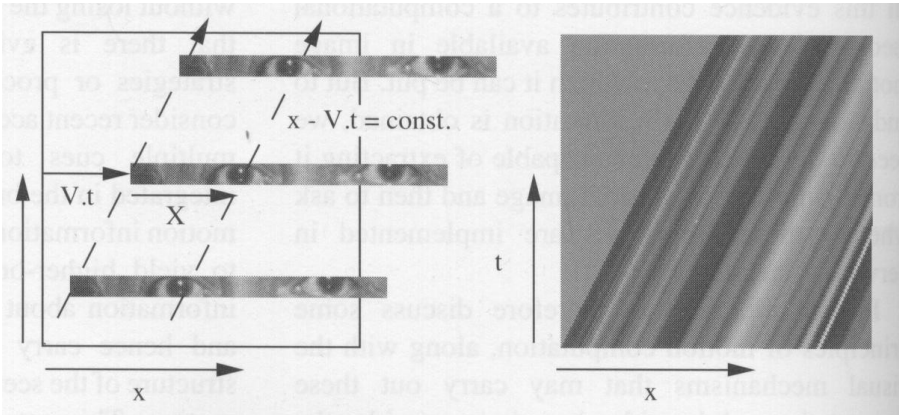
Optical Flow Constraint Equation:

$$V\frac{dI}{dx} + \frac{dI}{dt} = 0$$

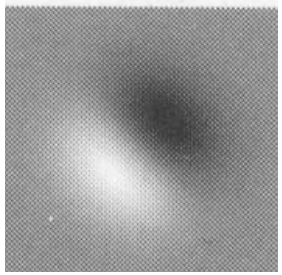
Method 4: energy models

Consider space-time 'image' I(x,y,t)

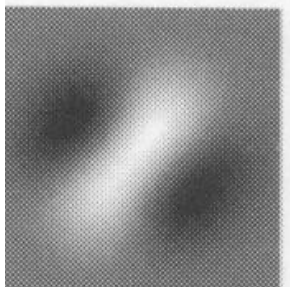
(pictures from Bruce et al 1996)



Create filters that detect intensity edges (or similar features) oriented in space-time

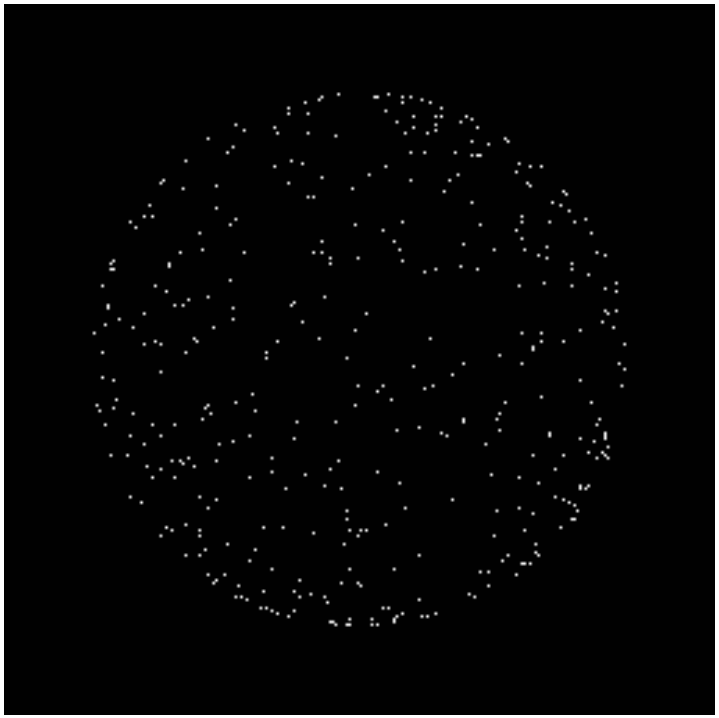


Summing and differencing their outputs can produce equivalent output to gradient or correlation models



Structure from motion

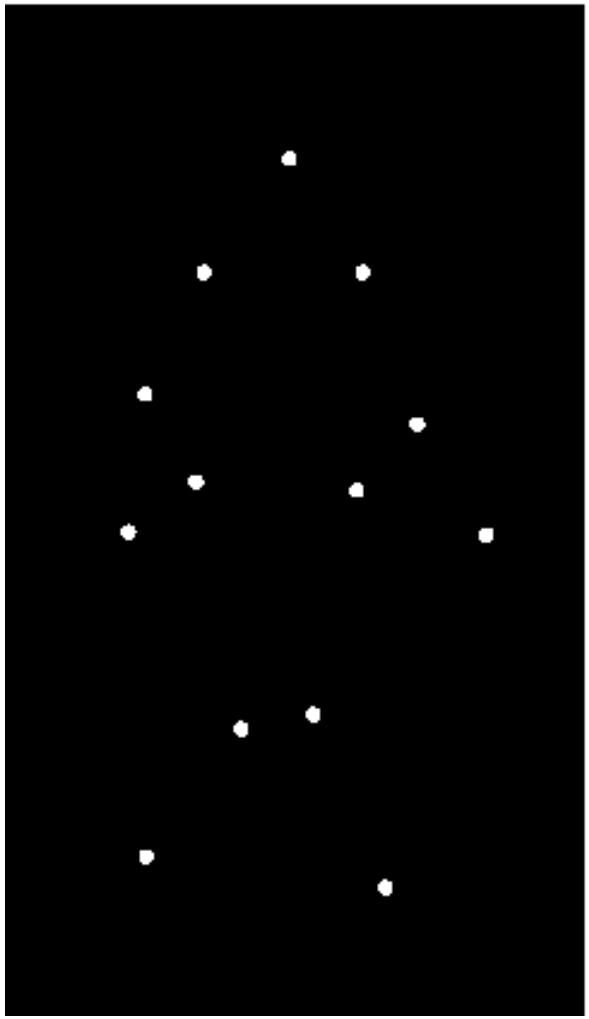
- Motion field contains information about the 3-d structure of objects (e.g. strong depth effect)



- If rigid body, and can track points, can geometrically recover structure of scene and movement of camera - active field in Computer Vision.

Structure from motion 2

What is this?



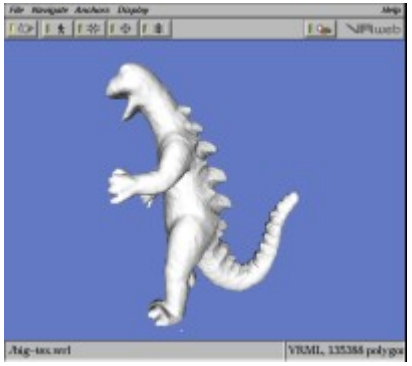
Structure from motion 3

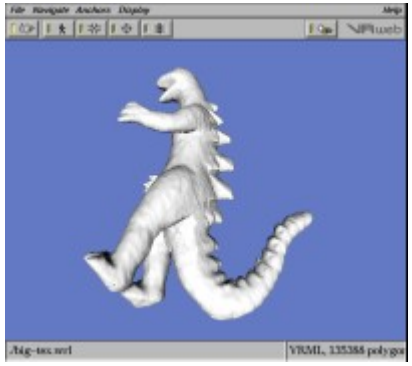
3D structure emerges from pattern of motion

Structure from motion

Source images

Reconstructed from tracked feature points:





Attention from motion

- Can use flow-field to determine where to redirect the eyes – moving stimuli are *salient*
- Mechanism to determine new eye position:
  - Calculate the flow field
  - Enhance changes to detect new stimuli
  - Smooth to offset noise
  - Implement ‘winner-take-all’ connection to choose most salient movement, and inhibit return to same location
- Note that then have to solve problem of mapping visual target onto correct motion of camera

Vijayakumar et al. 2001

Eye movements

- Increase the effective resolution by *saccade* movements of high resolution area (fovea)
- Creates impression that see detailed scene, but this is illusory
- Task dependent, indicates attention
- Why/how do we interpret the world as static when the image is constantly moving?

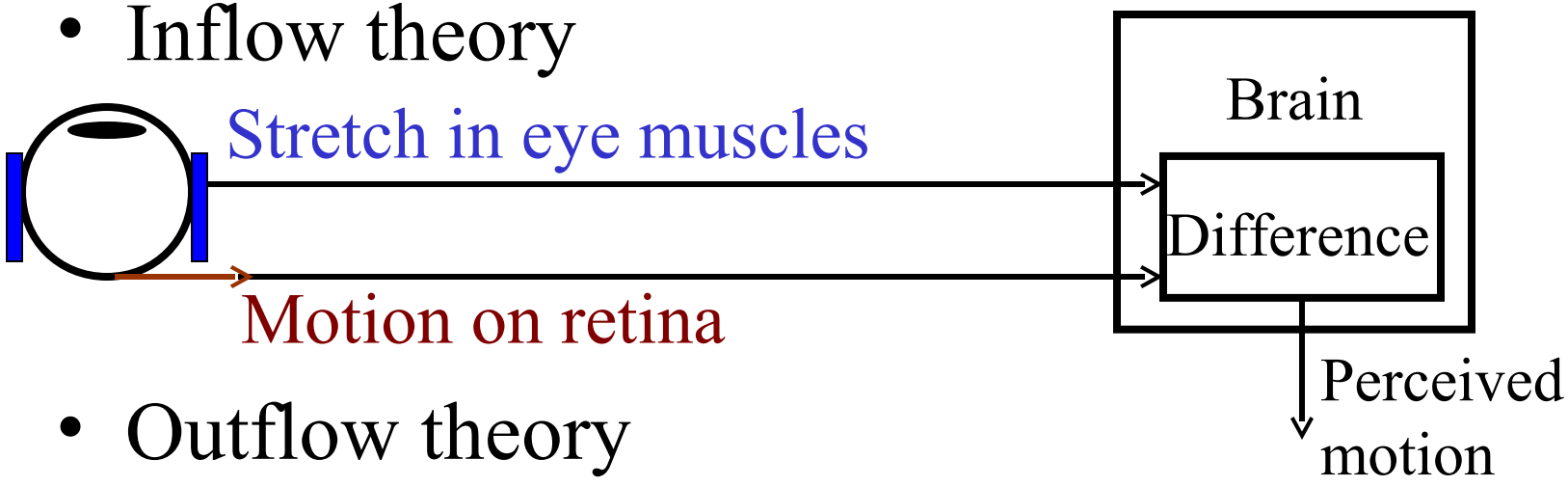
Explaining movement in retinal image?

	YES:	NO:
IF eyes not moving	See motion	Don't see motion
IF eyes are moving	Perceive stationary world during saccade	Perceive moving world (e.g. tracking, stabilised image)

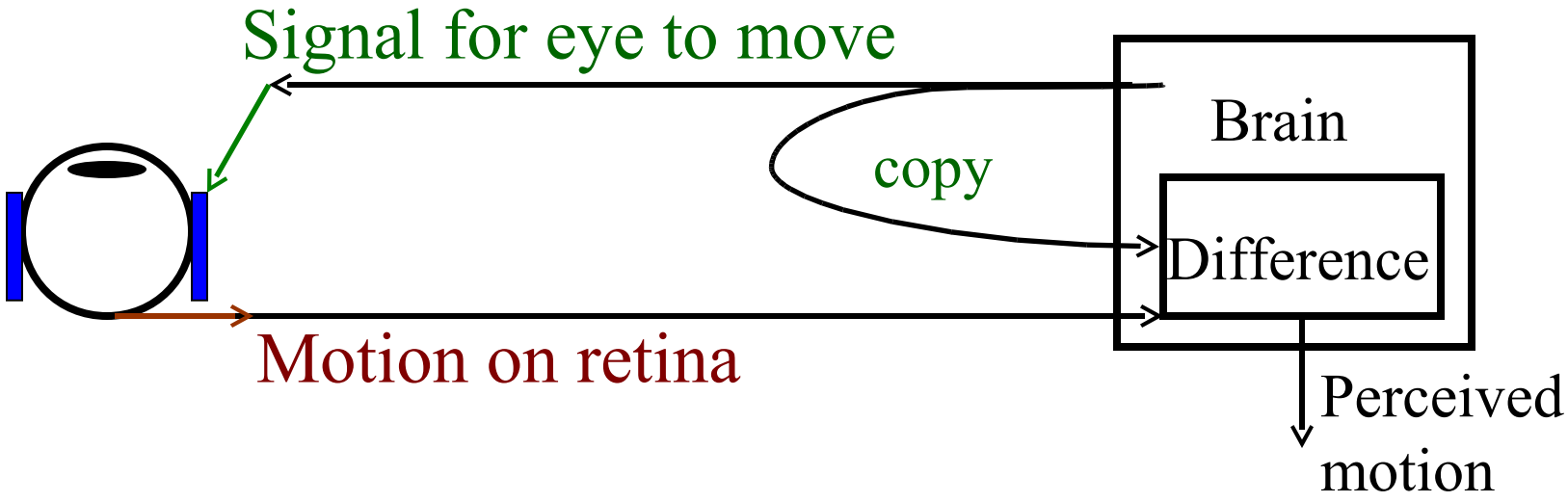
Visual system uses info to construct perception – but which info?

How does the visual system take eye movements into account?

- Inflow theory

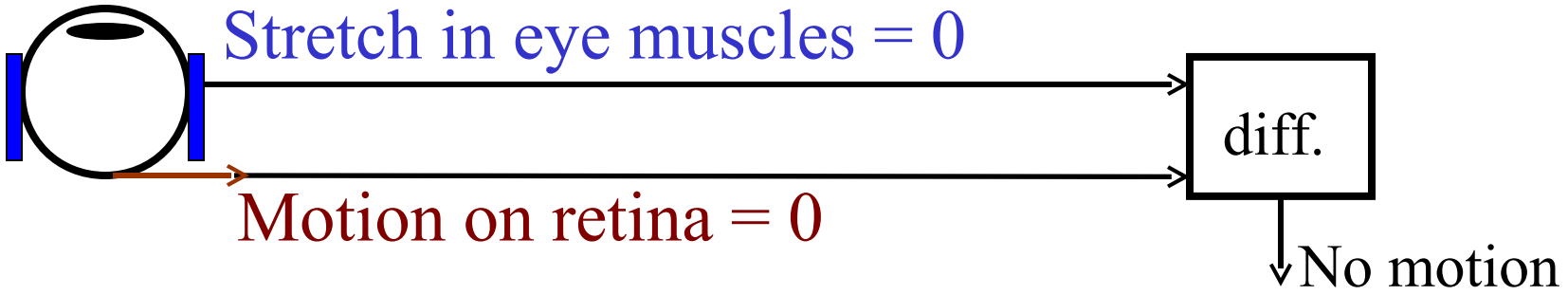


- Outflow theory

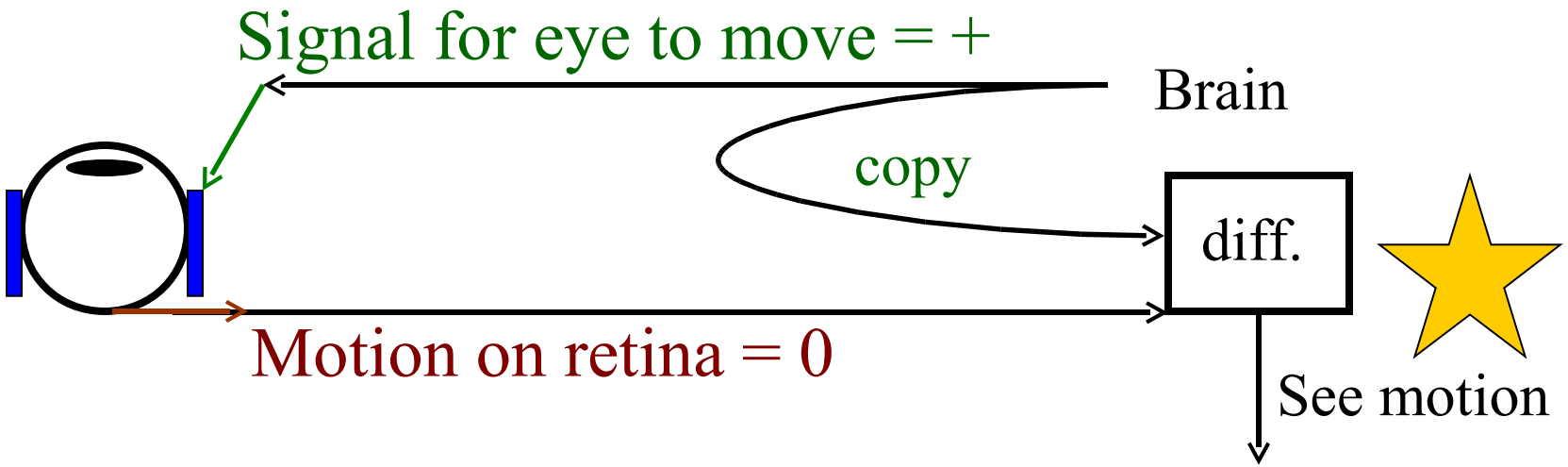


If immobolise the eyes and attempt to move them:

- Inflow theory predicts see no motion

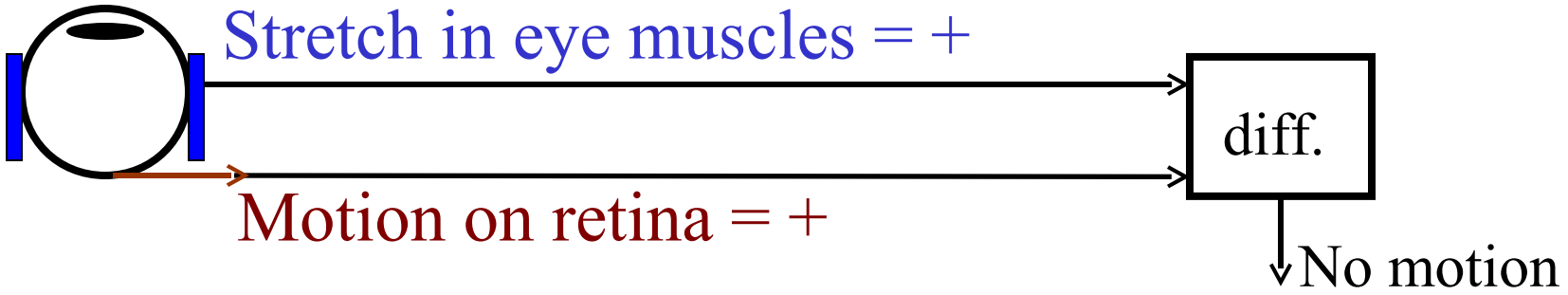


- Outflow theory predicts see motion

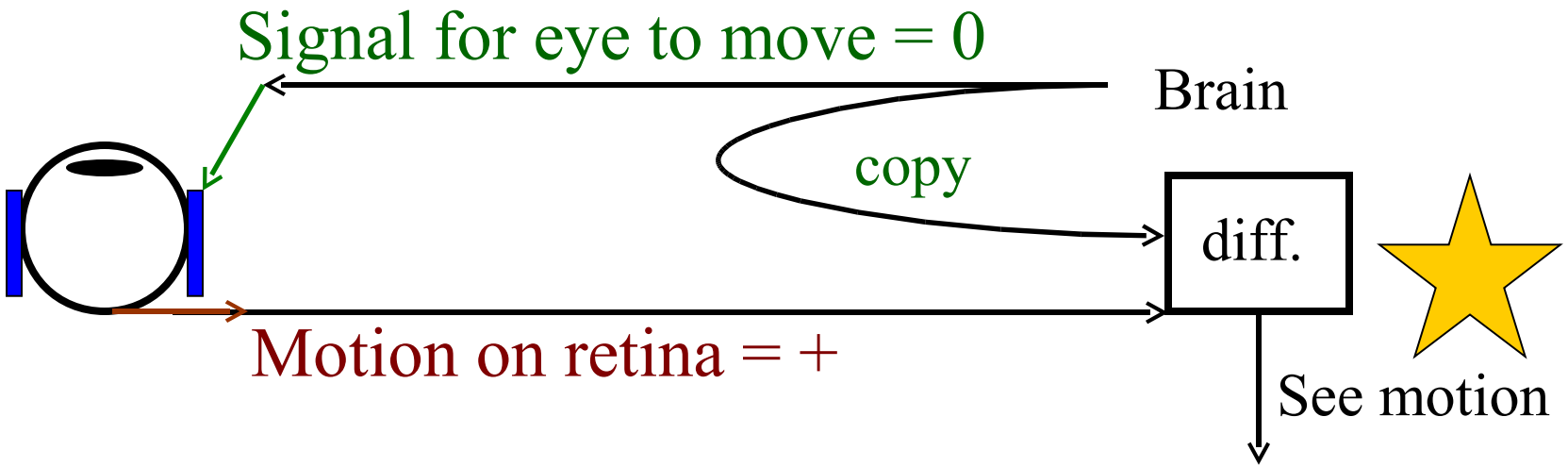


If eye is moved passively:

- Inflow theory predicts see no motion



- Outflow theory predicts see motion



Active Vision Summary

*Changing **Visual Field** gives info about:*

- Visual details
- Groupings
- Object depth and 3D structure
- Interesting Objects

Issues to consider:

- Self or scene movement
- Depth info from: features, area correlation, optical flow, spatio-temporal patterns