

Lecture 9. Motion Perception

Reading Assignments:

Chapter 10.

Announcements

- no lecture next week (3/13)

- white papers due the following week (3/20):

in one page (no more, no less), describe:

introduction about why your topic is important

brief summary of state of the art in the field

plans for the project

Moving objects

should not be confused with...

- spatio-temporal noise
- changes in lighting and shadows
- appearance or disappearance of image elements (e.g., flickering light)

What defines motion, then?

changes in the image with time, which do not fall under the categories of the previous slide.

and keep in mind that:

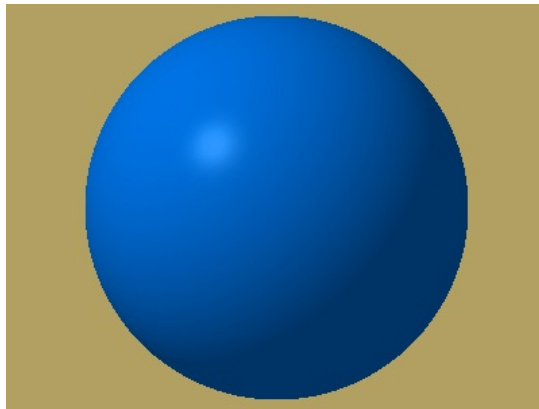
- “objects” may not be contiguous but may consist of several disjoint parts (e.g., flock of birds)
- motion may be non-rigid and involve deformations (e.g, person walking)
- moving objects may be partially occluded (e.g., behind a picket fence)
- observer and eye motion cause the entire scene to appear to move

Motion Field vs. Optical Flow

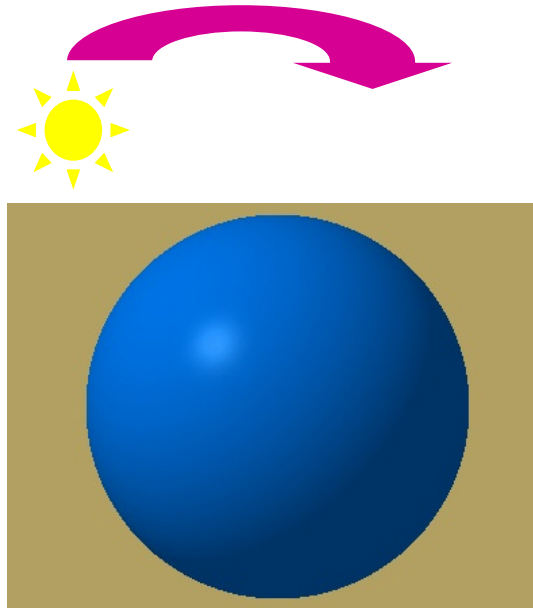
Motion field: assigns a motion vector to every point in the image – this is the “true” motion of objects.

Optical Flow: apparent motion of the brightness pattern of the image – this is what we can measure.

are they identical? ideally, yes, but...



Rotating sphere: motion field
but no optical flow



Moving sun: optical flow but
no motion field

The Computational Challenge

Infer **motion field** from the apparent **optical flow**.

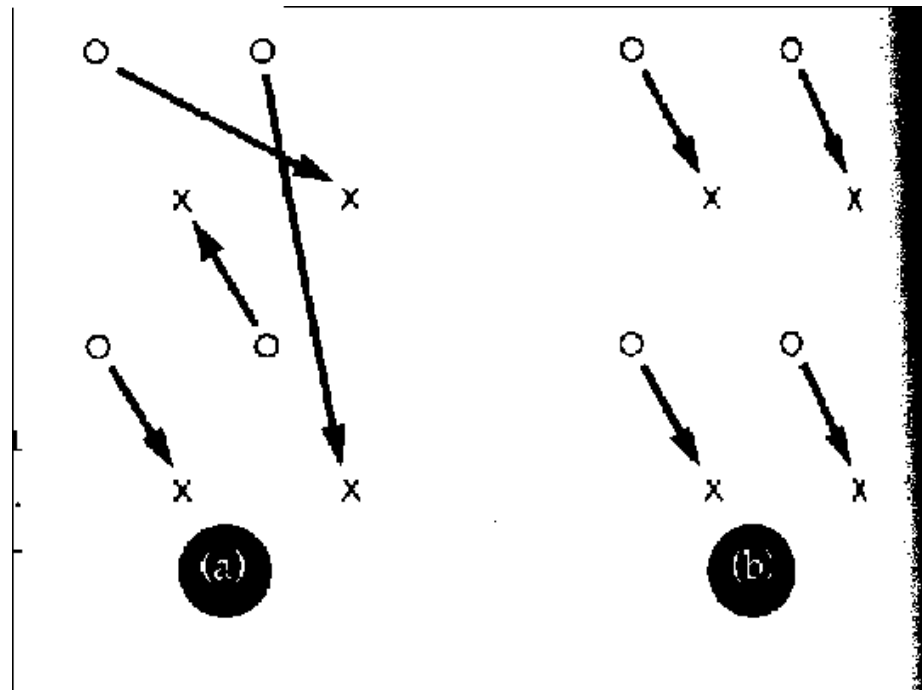
This is particularly difficult and requires high-level knowledge in many ambiguous situations (see previous slide).

Just computing the optical flow is a very complicated problem!

- correspondence problem: given series of frames, how do we pair features in one frame to corresponding features in the next frame, such as to infer the motion of that feature?
- aperture problem: if we don't see the whole object, recovering its motion is ambiguous.

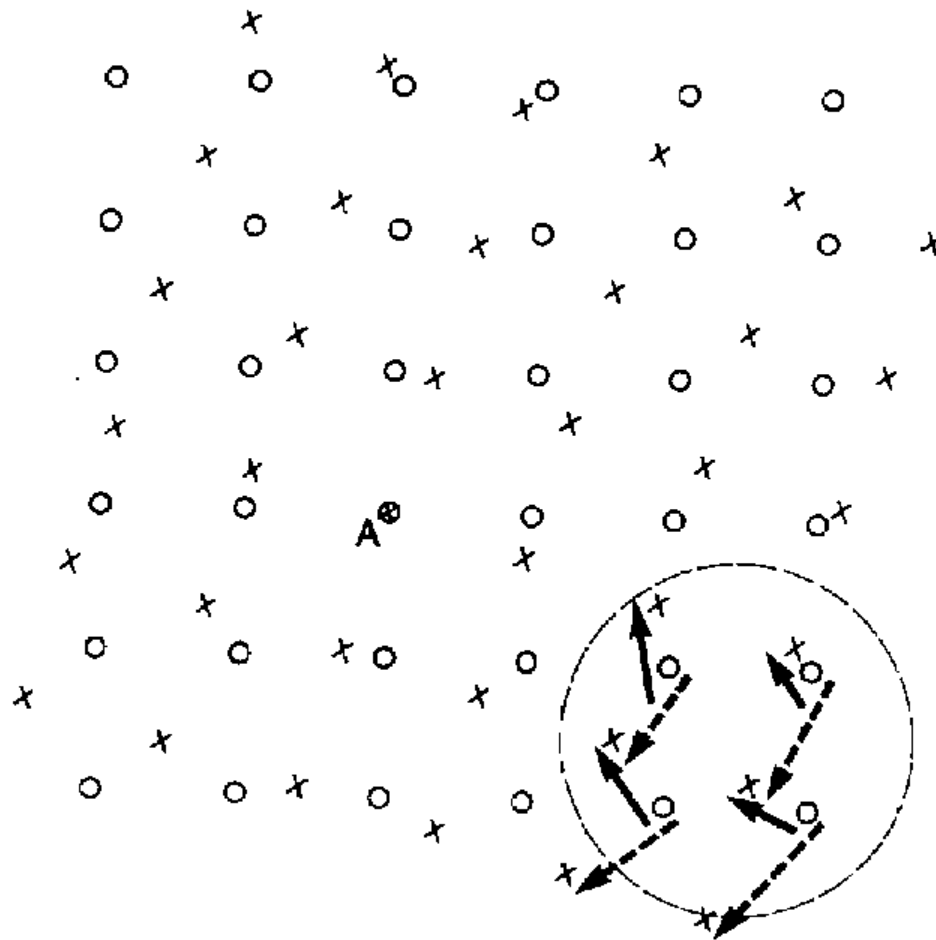
The Correspondence Problem

Many possible pairing between features extracted from successive frames may exist... identical to the correspondence problem in stereo vision, except that here we can use more than two frames to infer the pairing.

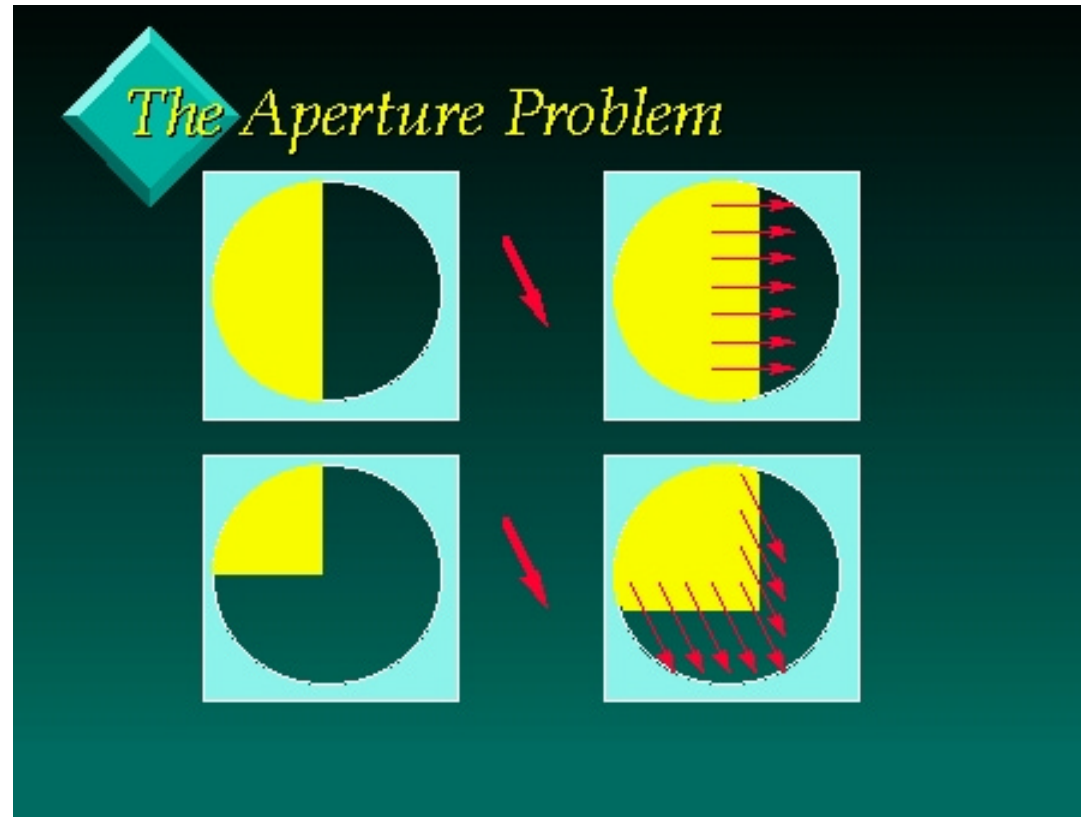


Note

Sometimes the “true” solution may not be the most obvious one!



The Aperture Problem



in the upper two images: even though the edge is moving down-right, we perceive its motion as being rightwards only...

For an edge viewed through an aperture, we can only recover the component of the motion field that is orthogonal to the edge.

Basic Approach

Standard assumption: **brightness constraint equation**:

the brightness of an image patch does not change as the observer moves around the object.

$$\frac{dI(x, y, t)}{dt} = 0.$$

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = I_x u + I_y v + I_t = 0.$$

so,

$$(I_x, I_y) \cdot (u, v) = -I_t$$

Hence values (u, v) satisfying the constraint form a straight line, and all a local measurement can do is identify this line (aperture problem).

We can compute the component of the optical flow in the direction of the brightness gradient $(I_x, I_y)^T$, but not in the orthogonal direction.

Typical Algorithms

Most computer vision algorithms are based on adding a **regularization constraint** to the basic equation.

e.g.,

Regularizing constraint: the optical flow field should vary smoothly in most parts of the image

associated error measure:
$$e_s = \iint ((u_x^2 + u_y^2) + (v_x^2 + v_y^2)) dx dy$$

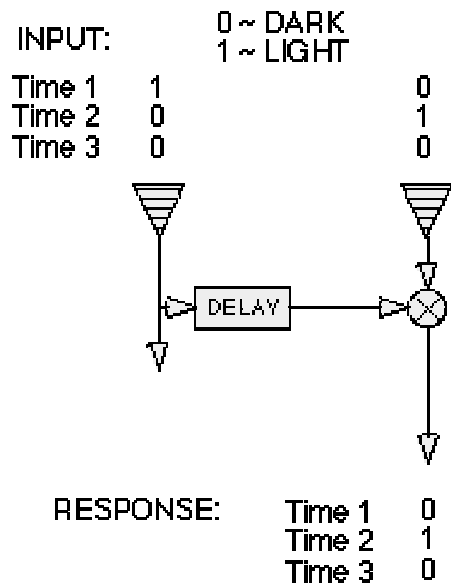
Error associated with the brightness constraint:

$$e_c = \iint (I_x u + I_y v + I_t)^2 dx dy$$

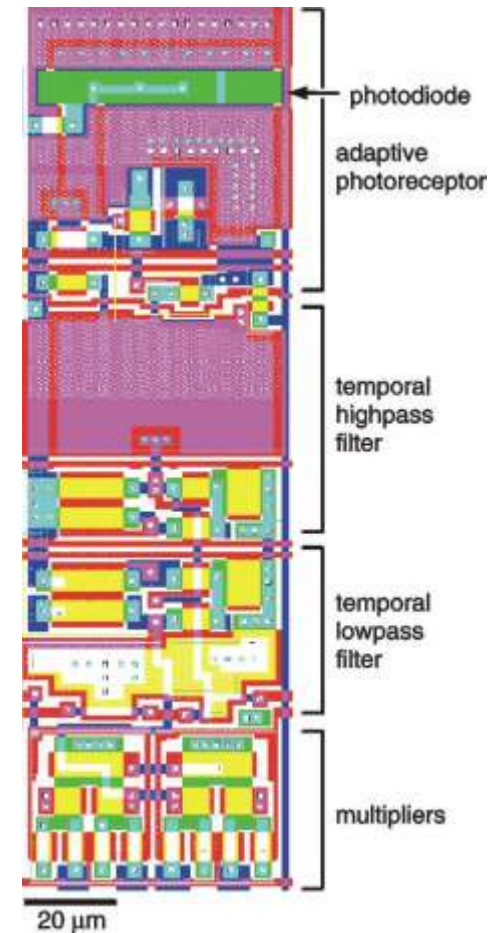
algorithm: find (u,v) that minimize $e_s + \lambda e_c$

Reichardt Motion Detector

Basic Reichardt detector: compares light intensity at two locations, with a time delay.



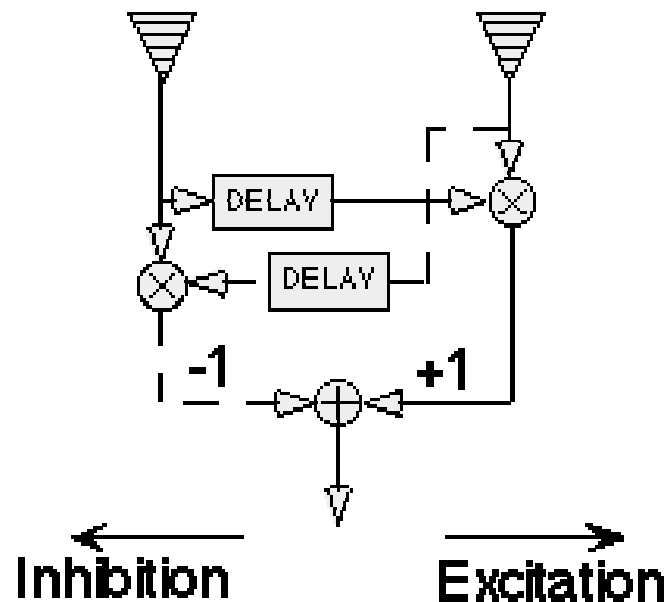
hence, it is tuned to a given velocity and direction of motion.



Analog VLSI implementation
(Harrison et al, 2000)

Full Reichardt Detector

Combines a detector in one direction with another one for the opposite direction.

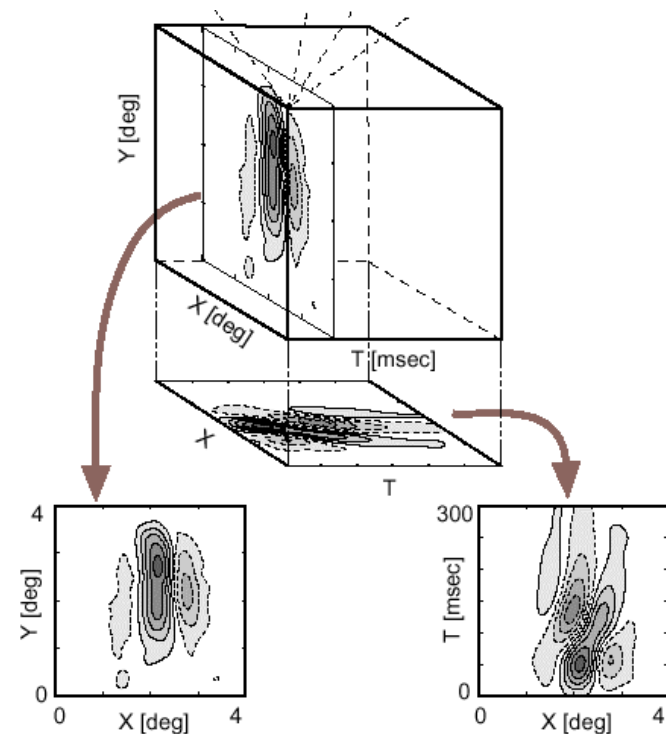


Spatio-Temporal Energy Model

Adelson & Bergen, 1984.

Remember the complex cells in V1: preferred stimulus is a bar with a given orientation and moving in a given direction (aperture problem yields that the preferred direction of motion is orthogonal to the preferred bar orientation).

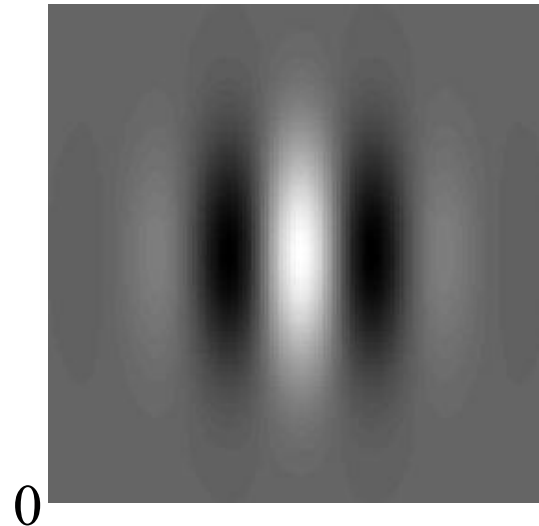
The spatio-temporal energy model is based on such cells with 3D (x,y,t) bandpass receptive fields.



Spatio-Temporal Energy Model

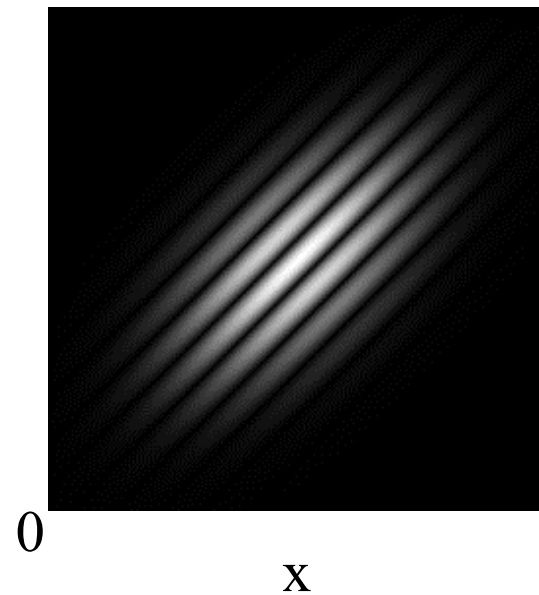
(x,y) bandpass receptive field
tunes 3D filter to oriented bars

y



(x,t) bandpass receptive field
tunes 3D filter to given velocity
in x direction.

t

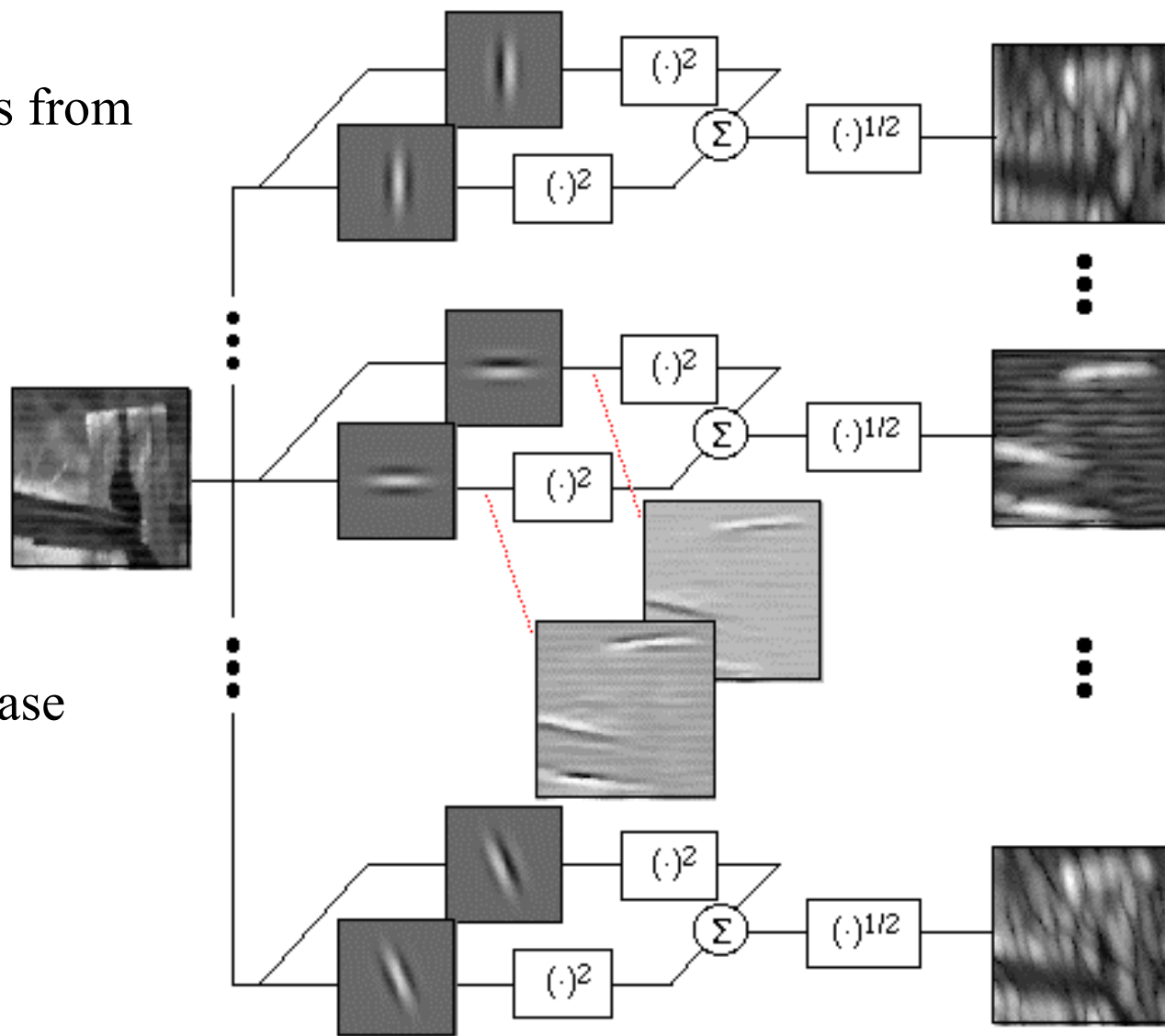


idem for (y,t)

Quadrature Pairs

Take sum of squared responses from a zero-phase and a 90deg-phase filters:

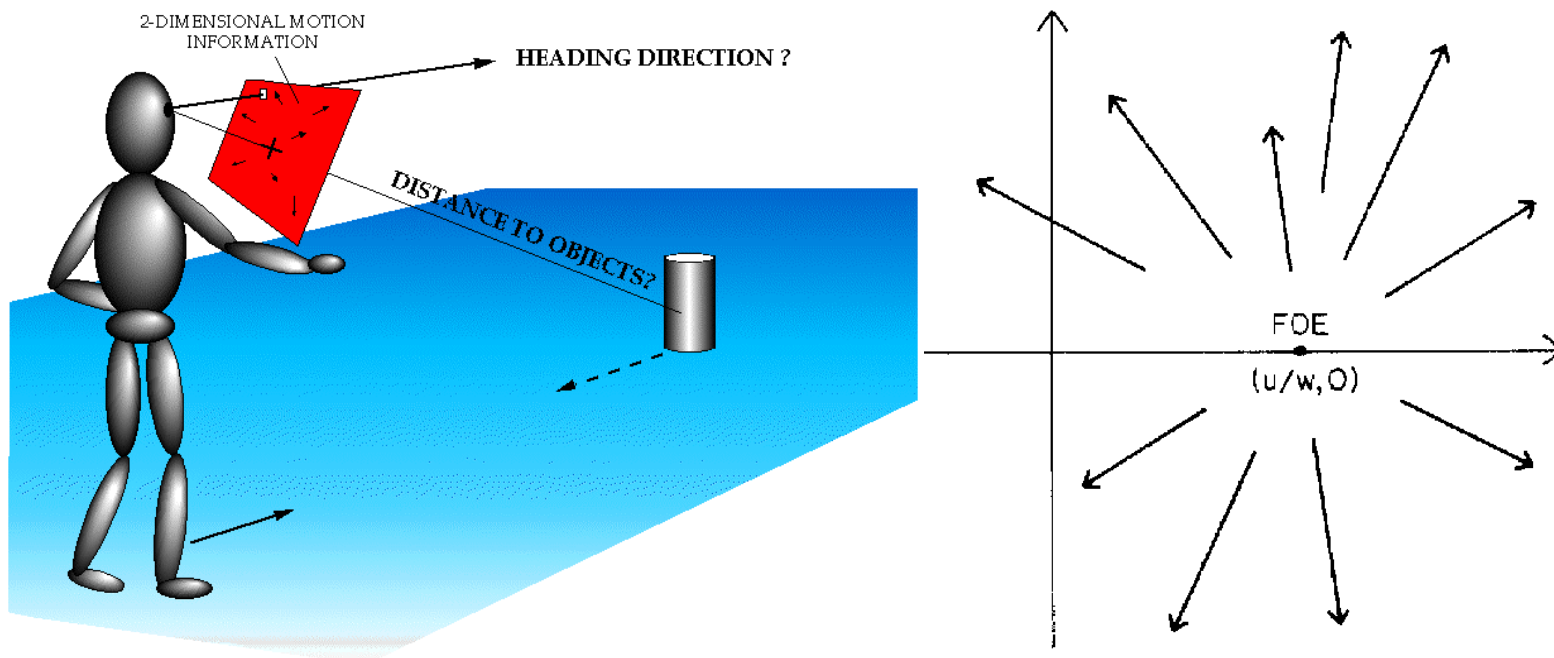
result is an “energy” which does not depend on phase of stimulus.



Locomotion

Several properties of the motion field can be exploited for locomotion:

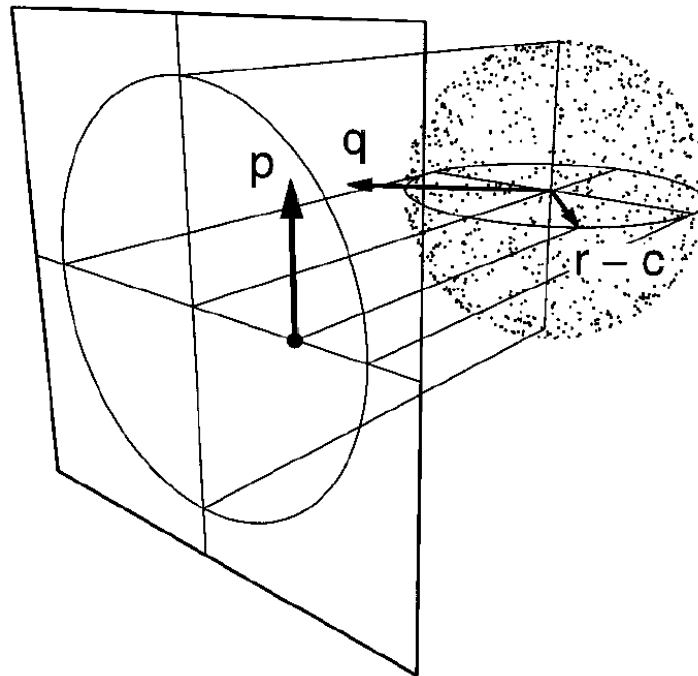
- pure translation in the direction of gaze yields an expanding flow field
- one point has zero motion: **the focus of expansion (FOE)**; we will reach it if we continue moving in the same direction forever.



Complex Motion

We can recognize objects from their motion with amazing facility (e.g., Johansson, 1973).

In addition to basic feature detection and regularization constraints, this must be thanks to additional “high-level” constraints on the shape of objects and their usual appearance when they move.



What Motion Perception Tells Us

depth and form

- motion parallax

- structure from motion

- biological motion

classification and identification

- living vs. non-living

- actions

- gender

- individuals (self and friends)

doesn't involve complex thought processes, quick to compute

What Motion Perception Tells Us

visual guidance

direction identification

judging time of arrival

retinal image size expands exponentially as we approach an object:

judging time to contact

a wide range of creatures respond to looming (expansion)

Case Study

Biological perception of looming motion and time-to-contact.

Looming Motion

time: t

distance: $D(t)$

angle subtended by object on retina: θ

absolute rate of image expansion: ρ

relative rate of expansion: τ

response of looming-sensitive neuron
in the locust: η

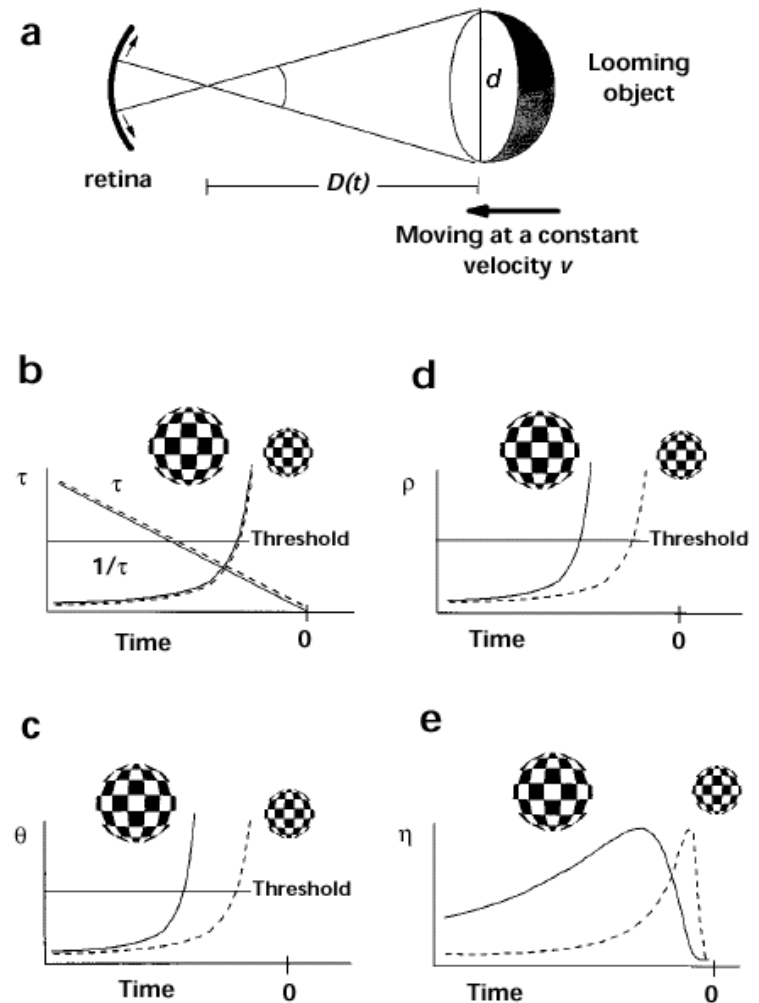


Fig. 1. A schematic diagram of a spherical object of diameter d directly approaching an animal's eye (a). When an object moves at constant velocity V , at time t , it is at distance $D(t)$ away from the eye, and subtends a visual angle of $\theta(t)$. Time course of four optical variables, τ (and $1/\tau$), θ , ρ , and η derived from the edges of two spherical objects of different sizes, moving at a constant speed directly toward the eye are shown in (b), (c), (d) and (e) respectively (solid line corresponds to larger sphere, dashed line to smaller sphere). Horizontal lines represent hypothetical threshold values for onset of neuronal firing. Time was plotted backwards from the time of collision, $t = 0$.

Sun & Frost, Nature Neuroscience, 1998

Looming

Neurons
in the pigeon's
nucleus rotundus.

Sun & Frost,
Nature Neuroscience, 1998

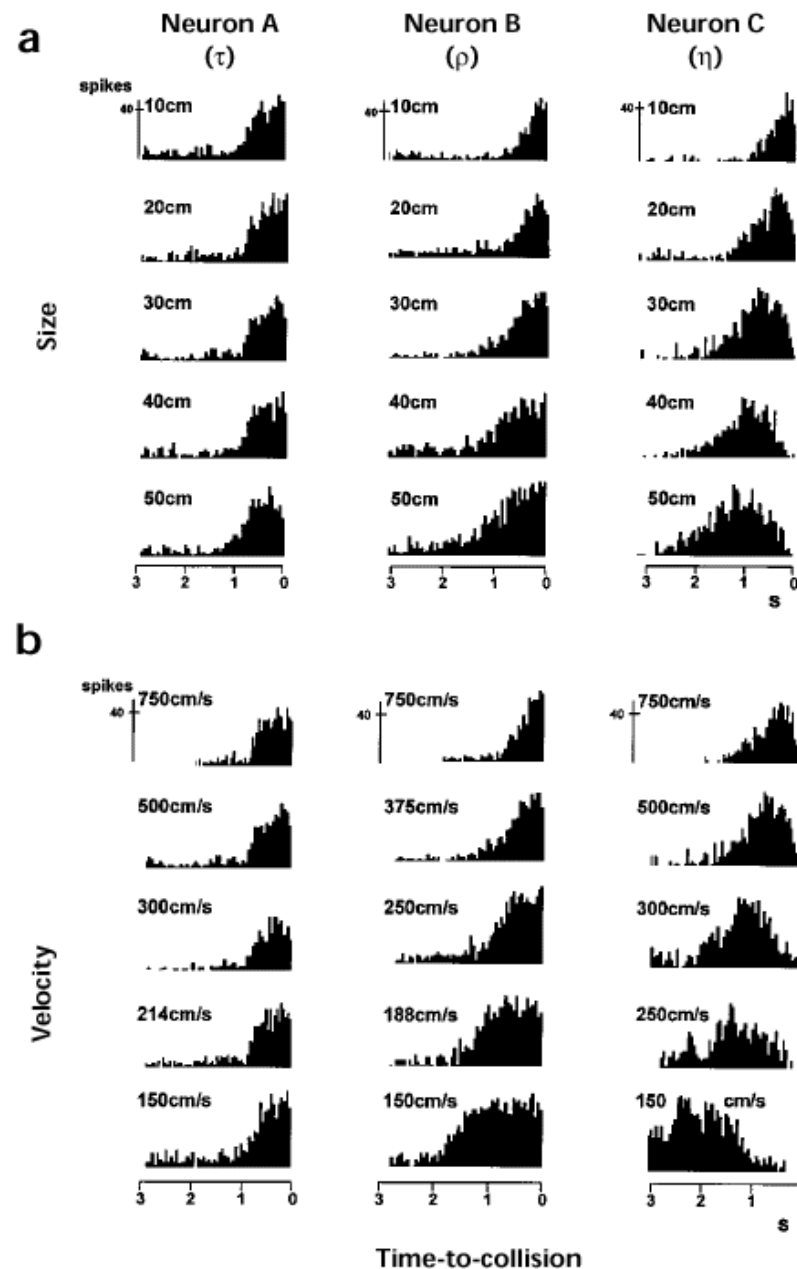


Fig. 2. Based on the differences in the time course of the neuronal responses relative to the moment of collision, the looming sensitive neurons in nucleus rotundus have been classified into three distinct classes. This figure shows the response pattern (post-stimulus time histograms) for a typical neuron in each of the three classes (neuron A, B, and C for τ , ρ , and η respectively) to a series of stimuli (a simulated moving sphere with a soccer-ball pattern) of varying sizes (**a**) or varying velocities (**b**), moving along the direct collision-course path toward the bird. Responses are the sum of five trials and are referenced to time zero, which is the time when the stimulus would have collided with the bird. The simulated path was 15 m in length. In (a), velocity for neuron B was 375 cm per s and for neurons A and C was 500 cm per s. In (b), object size was 30 cm for all three neurons. Note that for neuron A, the timing of the response remains invariant despite substantial changes in size and velocity, whereas for neuron B

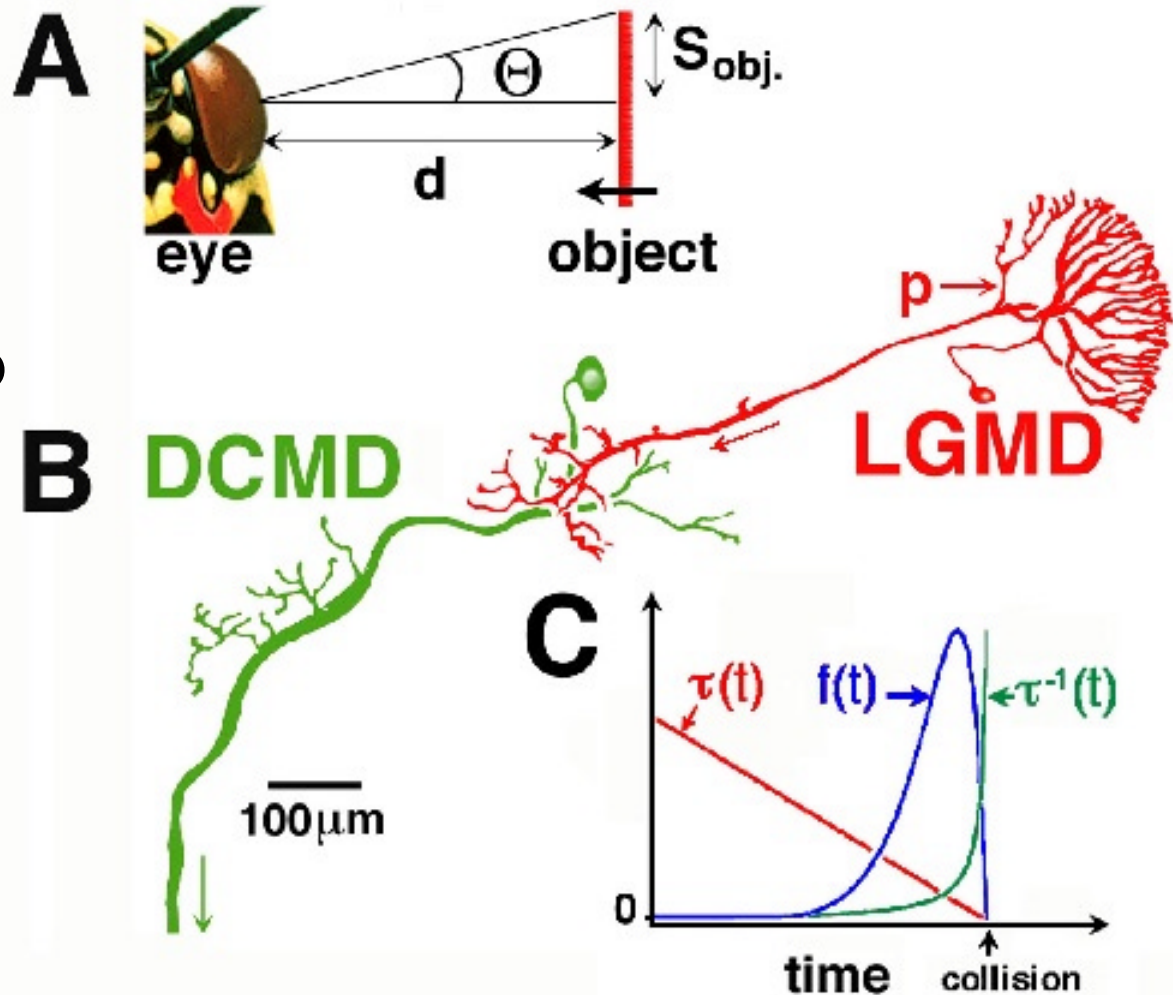
and neuron C, the timing depends on object size and velocity, with larger or slower objects evoking an earlier response. Further quantitative examination suggests that neurons A, B and C encode optical variables τ , ρ , and η respectively.

Looming Detector in the Locust

Gabbiani et al, Science, 1996.

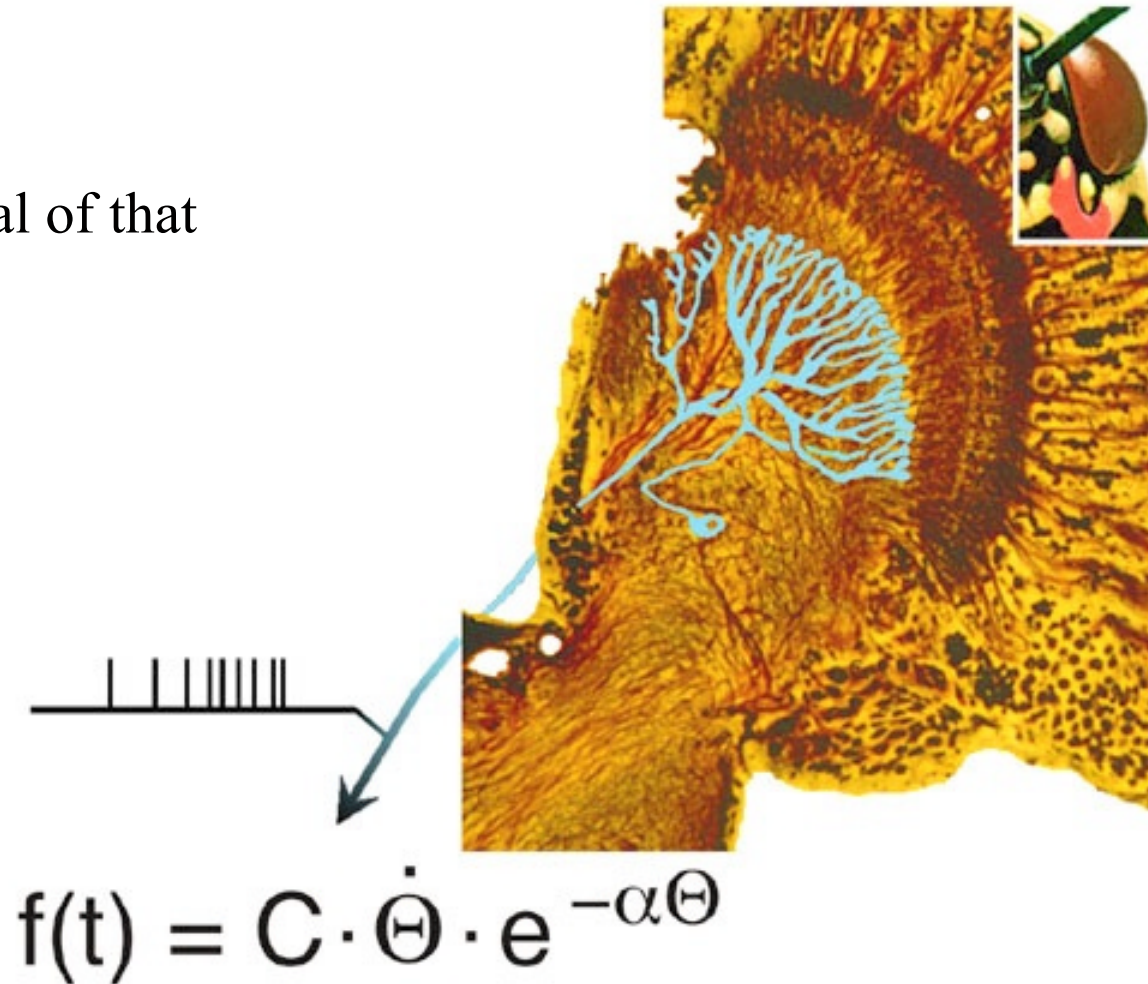
First stage (DCMD neurons): extract features

Second stage (LGMD neuron): compute looming



Neural response to looming

proportional to
time-derivative of
visual angle times
negative exponential of that
angle.



Detecting Looming Motion

Analog-VLSI
implementation

(Indiveri & Koch)

