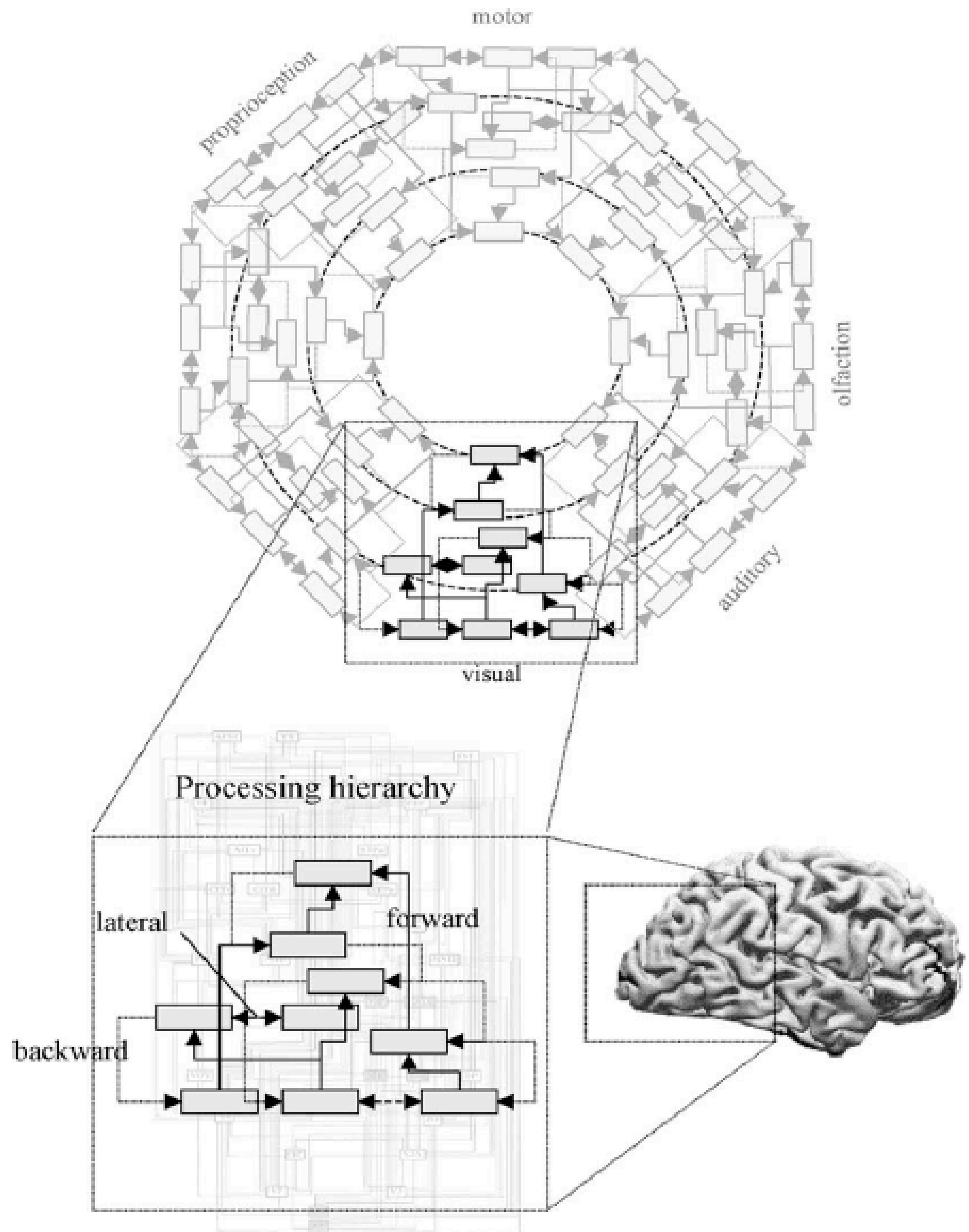
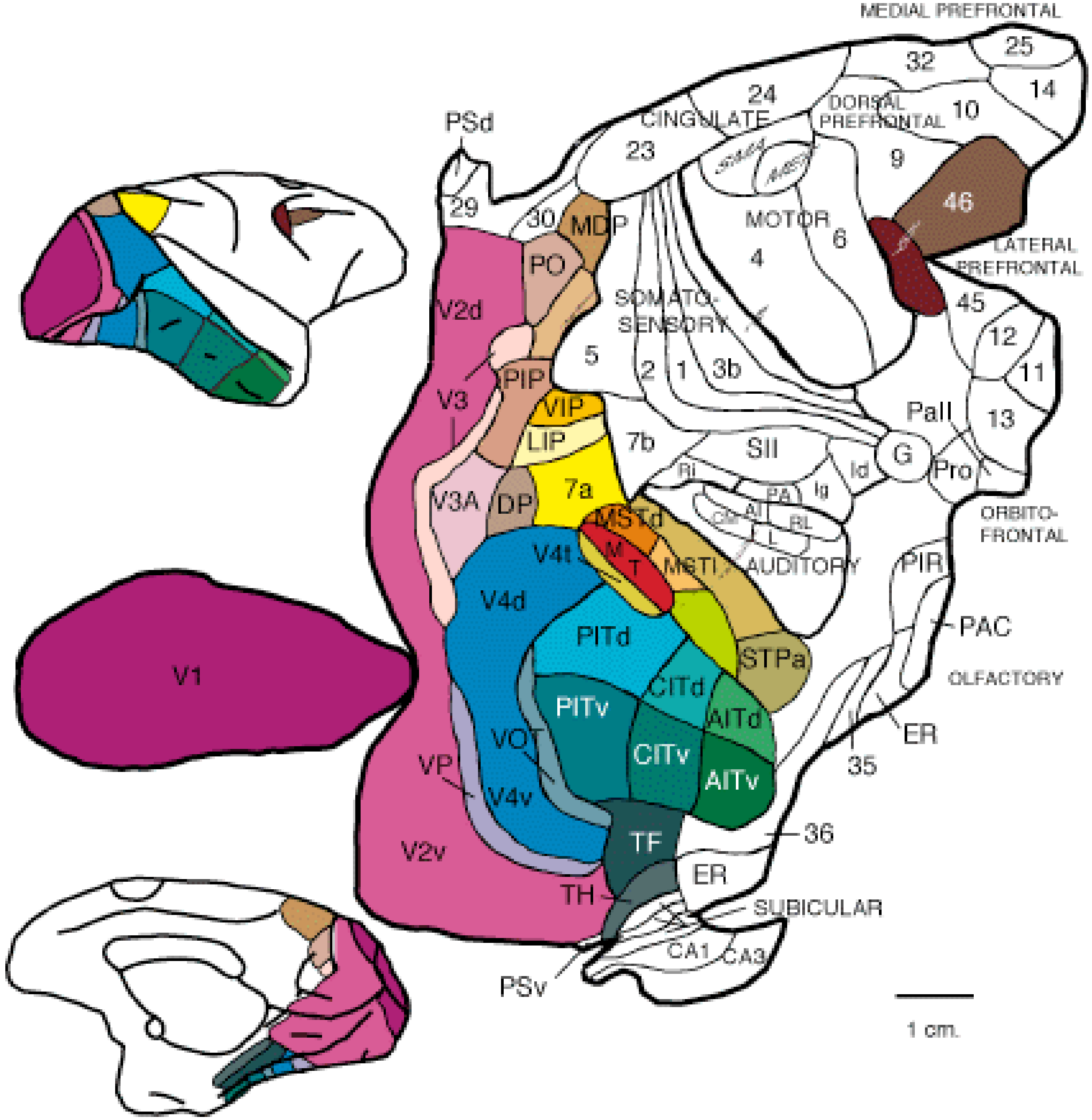
Vision as Optimal Inference

- The problem of visual *processing* can be thought of as computing a belief distribution
- Conscious perception better thought as a *decision* based on both beliefs and the utility of the choice.

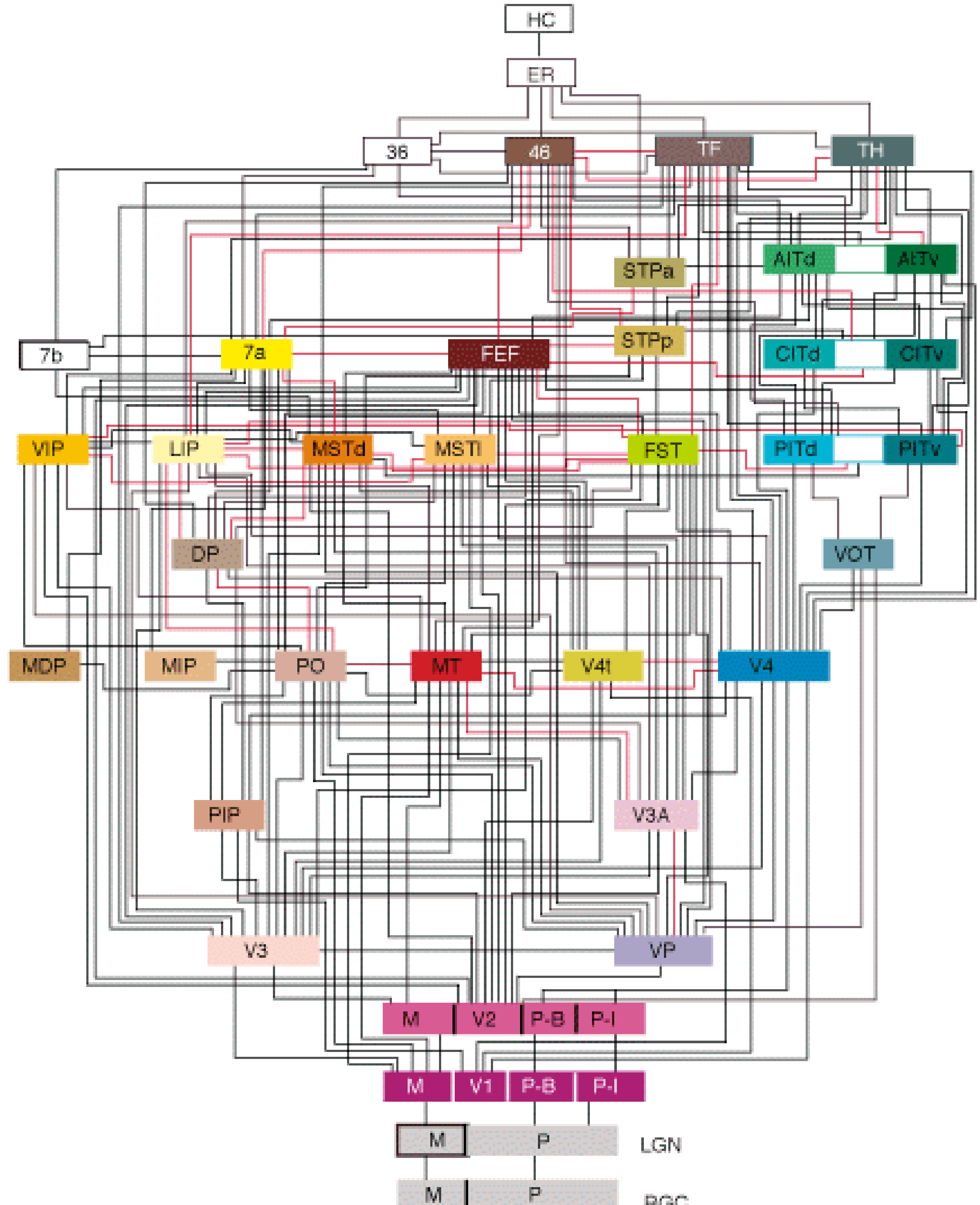
Hierarchical Organization of Visual Processing



Visual Areas



Circuit Diagram of Visual Cortex



Motion Perception as Optimal Estimation

Local Translations

OpticFlow:
(Gibson,1950)
Assigns local image velocities v(x,y,t)

Time ~100msec
Space ~1-10deg

Measuring Local Image Velocity

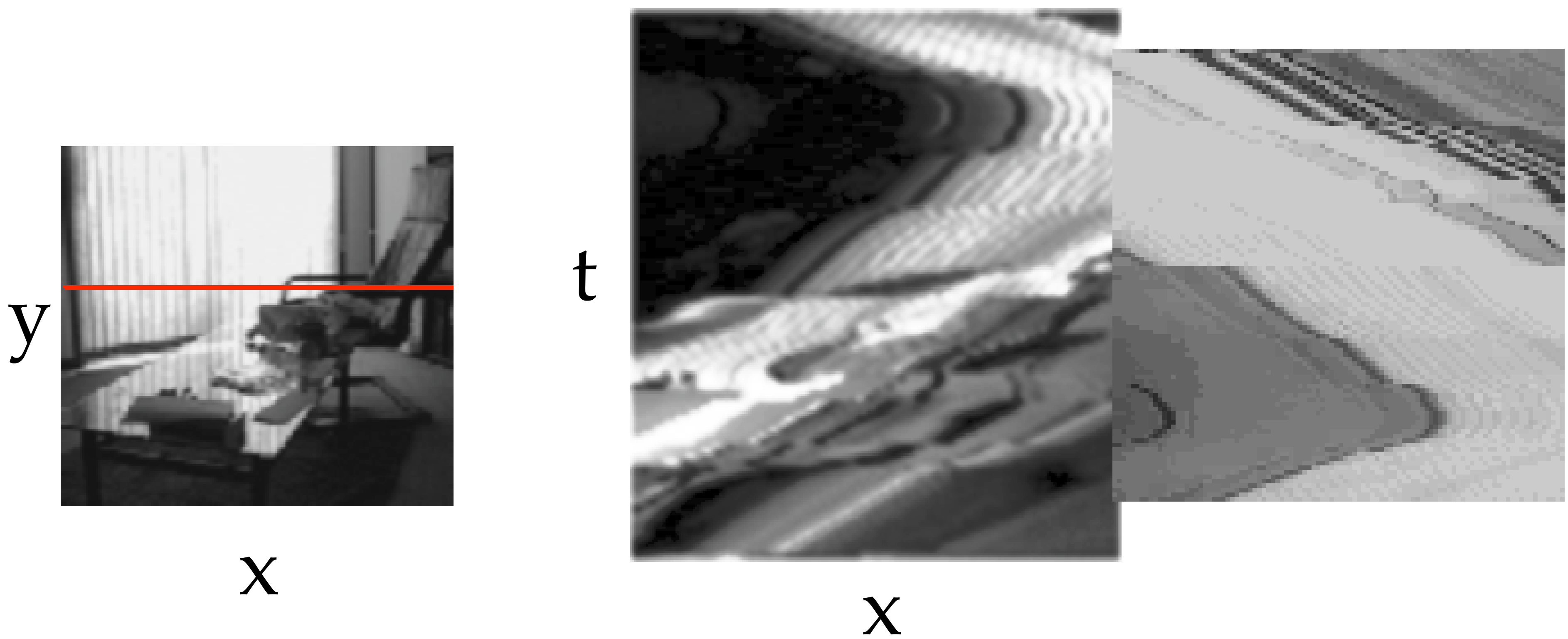
Reasons for Measurement

- Optic Flow useful:
  - Heading direction and speed, structure from motion,etc.
- Efficient:
  - Efficient code for visual input due to self motion (Eckert & Watson, 1993)

How to measure?

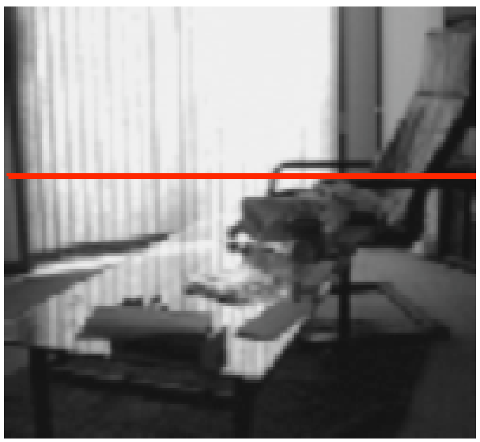
- Look at the characteristics of the signal

X-T Slice of Translating Camera



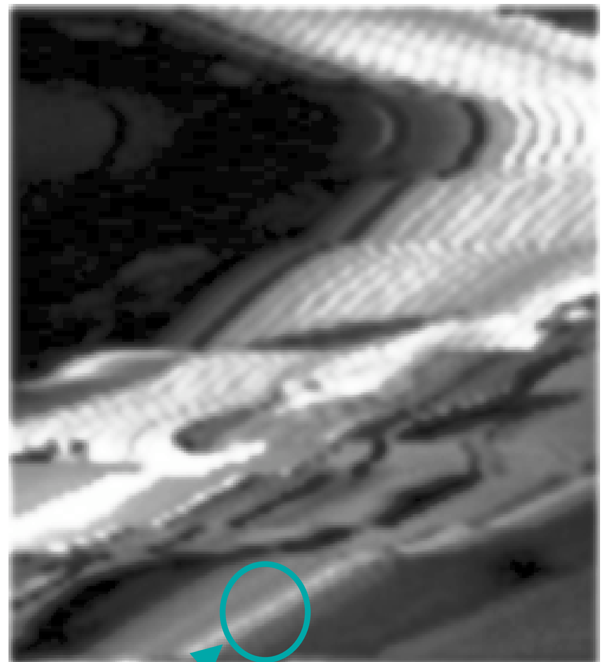
X-T Slice of Translating Camera

y



x

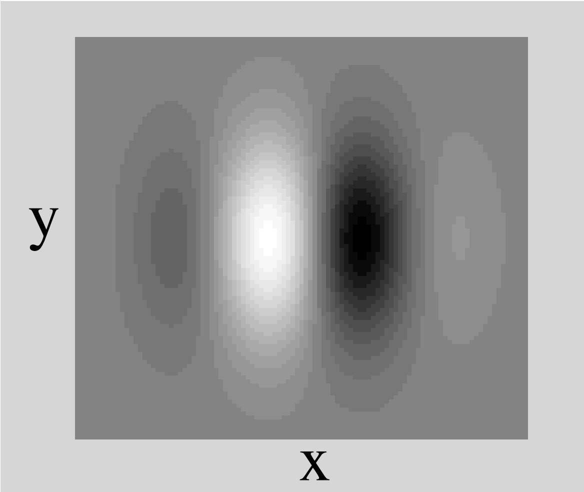
t

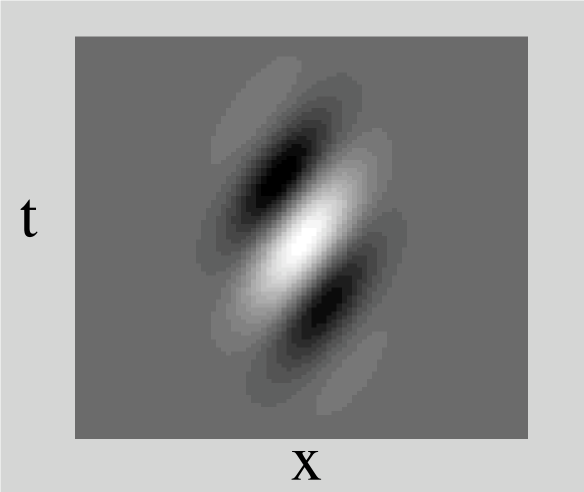


x

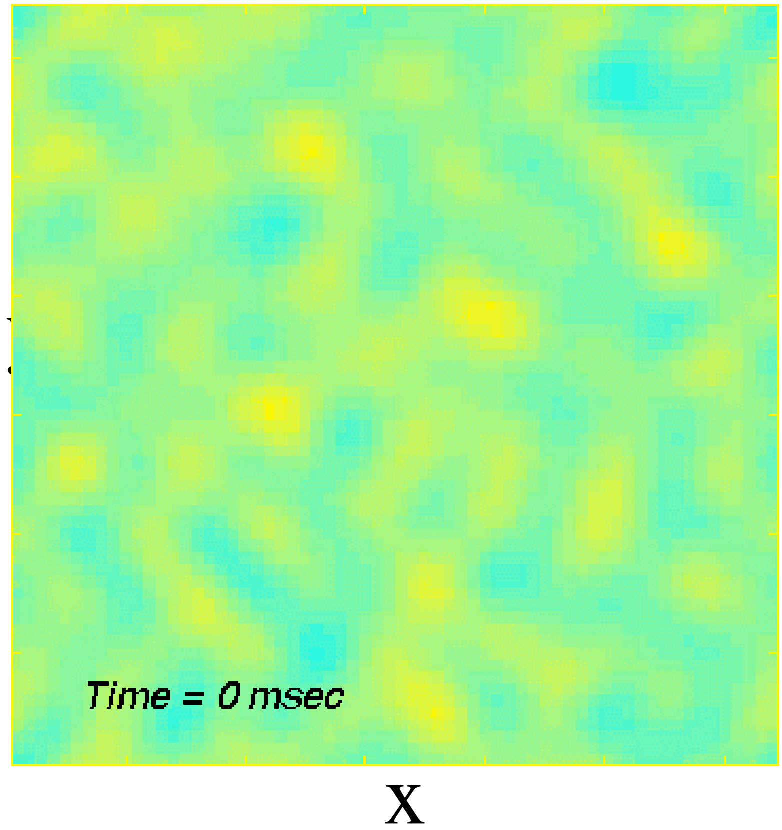
Local translation

Early Visual Neurons (V1)





Ringach et al (1997)

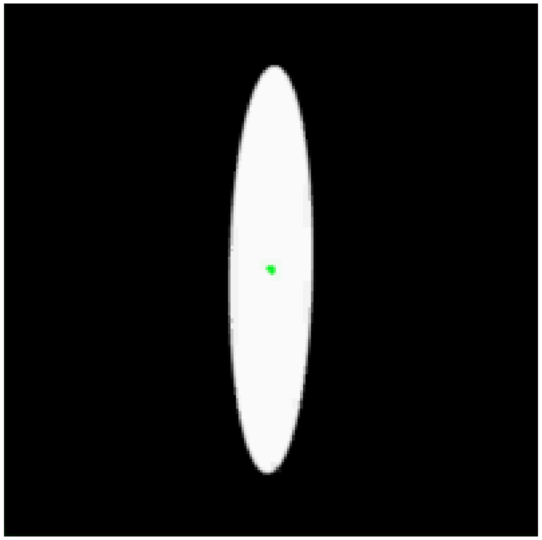


What is Motion?

As Visual Input:

- Change in the spatial distribution of light on the sensors.

Minimally, $dI(x,y,t)/dt \neq 0$



As Perception:

- Inference about causes of intensity change, e.g.

$I(x,y,t) \longrightarrow v_{OBJ}(x,y,z,t)$

Motion Field: Movement of Projected points

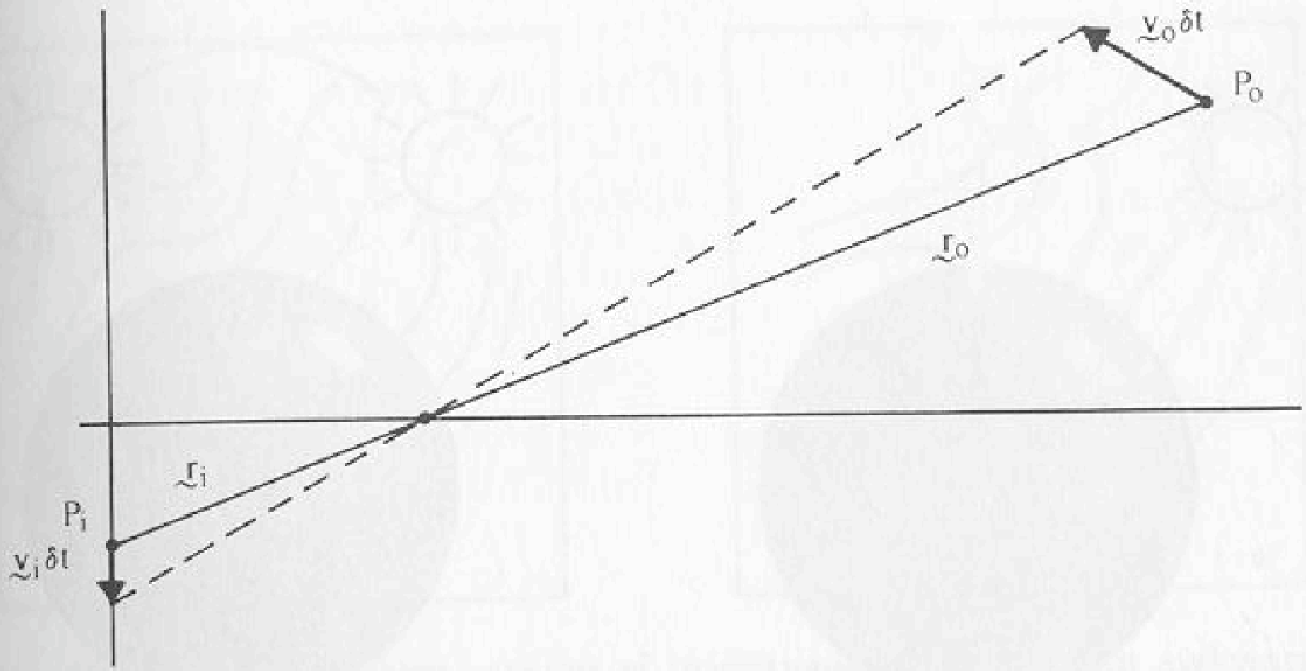
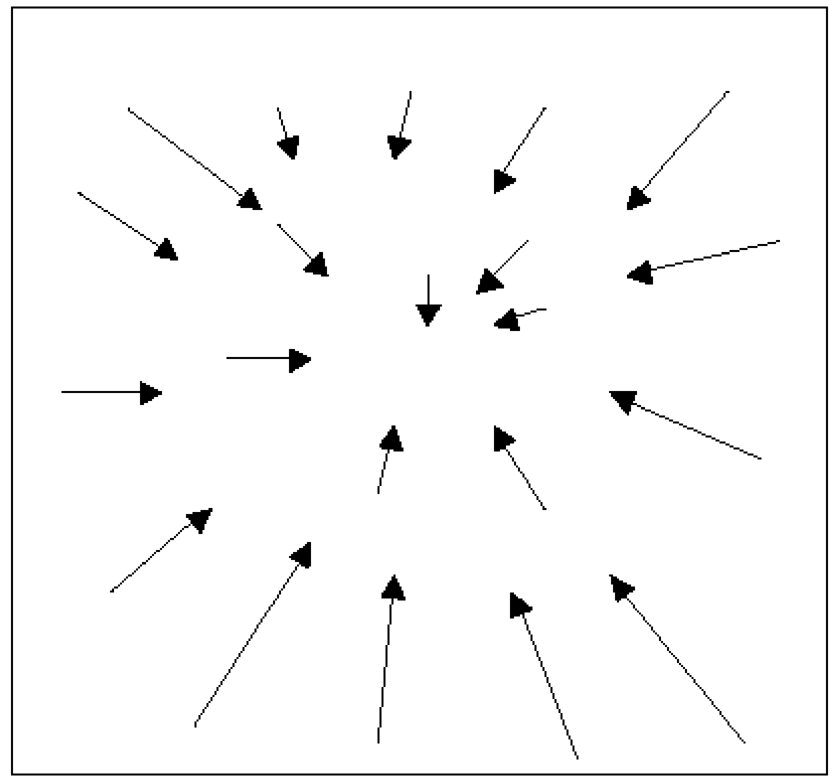


Figure 12-1. Displacement of a point in the environment causes a displacement of the corresponding image point. The relationship between the velocities can be found by differentiating the perspective projection equation.



Basic Idea

- 1) Estimate point motions
- 2) use point motions to estimate camera/object motion
- Problem: Motion of projected points not directly measurable.
- -Movement of projected points creates displacements of image patches -- Infer point motion from image patch motion
  - Matching across frames
  - Differential approach
  - Fourier/filtering methods

Differential approach: Optical flow constraint equation

Brightness should stay constant as you track motion

$$I(x+u\delta t, y+v\delta t, t+\delta t) = I(x,y,t)$$

1 st order Taylor series, valid for small δt

$$I(x,y,t) + u\delta t I_x + v\delta t I_y + \delta t I_t = I(x,y,t)$$

Constraint equation

$$\boxed{uI_x + vI_y + I_t = 0}$$

"BCCE" - Brightness Change Constraint Equation

Brightness constancy constraint line

$uI_x + vI_y + I_t = 0$:
one equation,
two unknowns!

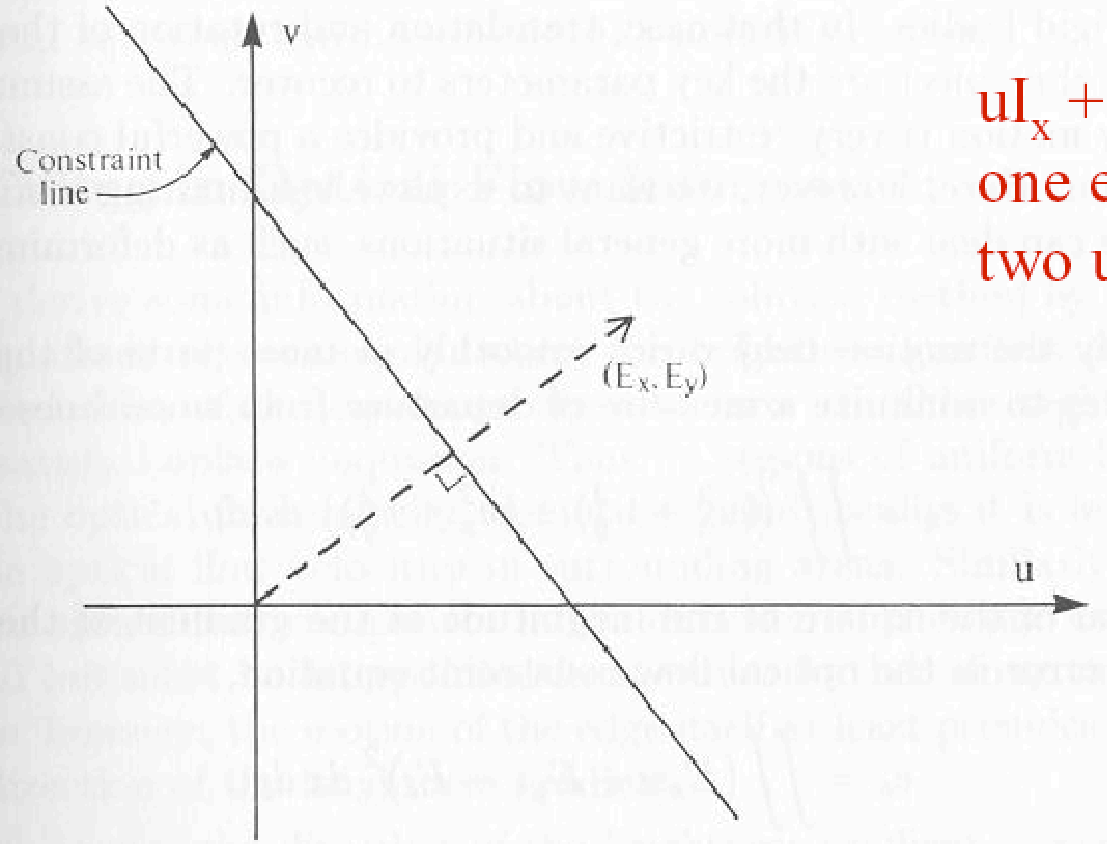
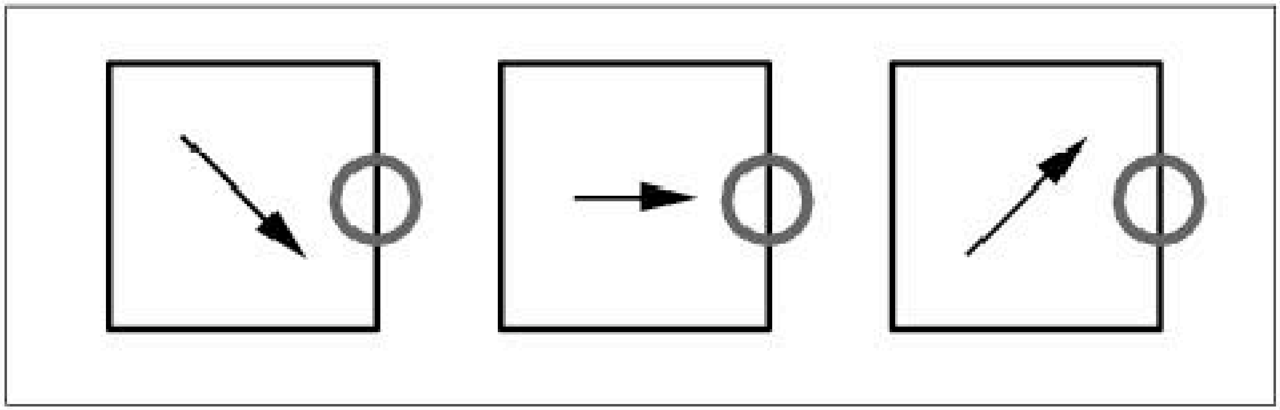


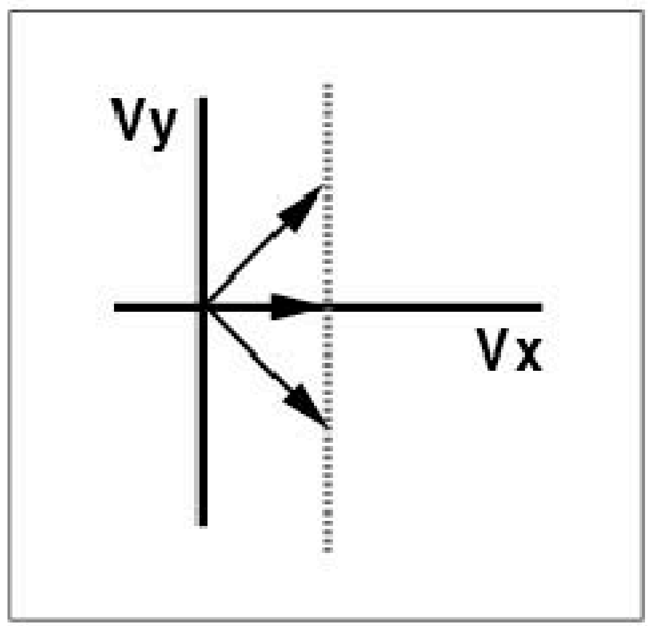
Figure 12-4. Local information on the brightness gradient and the rate of change of brightness with time provides only one constraint on the components of the optical flow vector. The flow velocity has to lie along a straight line perpendicular to the direction of the brightness gradient. We can only determine the component in the direction of the brightness gradient. Nothing is known about the flow component in the direction at right angles.

Problem: Images contain many edges-- Aperture problem

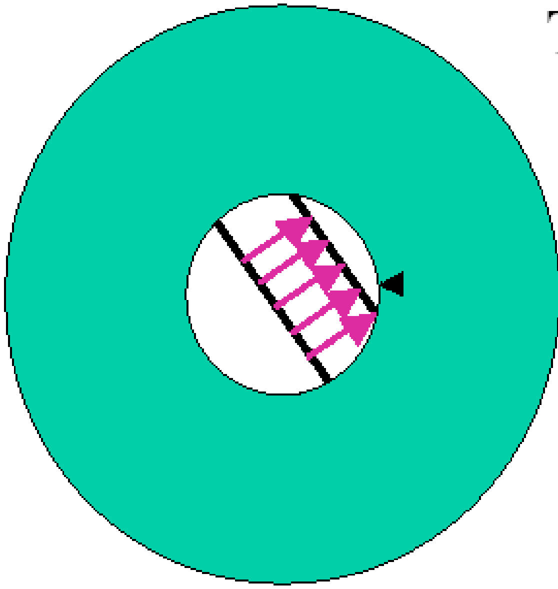


a

Normal flow:
Motion component in the direction of the edge



b



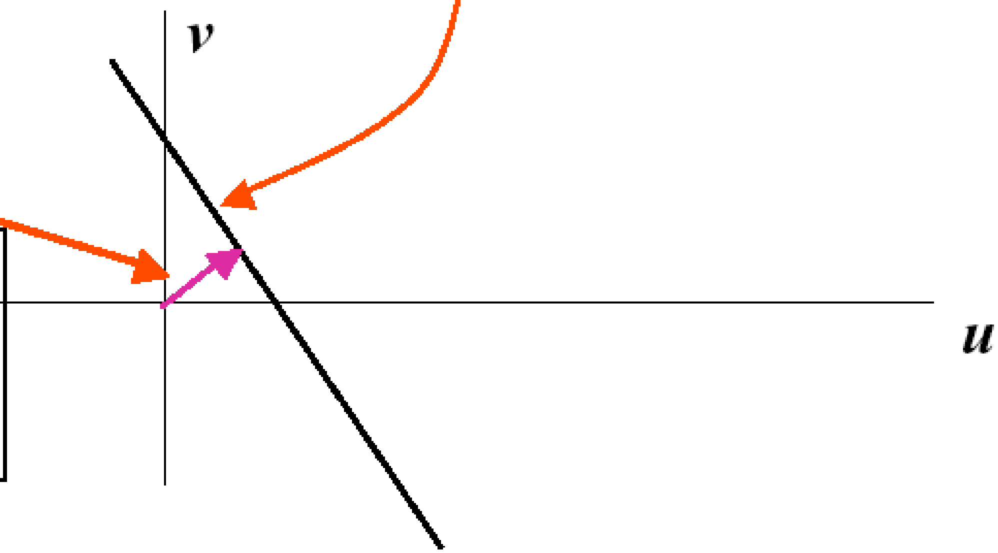
The gradient constraint:

$$\boxed{\begin{aligned} I_x u + I_y v + I_t &= 0 \\ \nabla I \bullet \vec{U} &= 0 \end{aligned}}$$

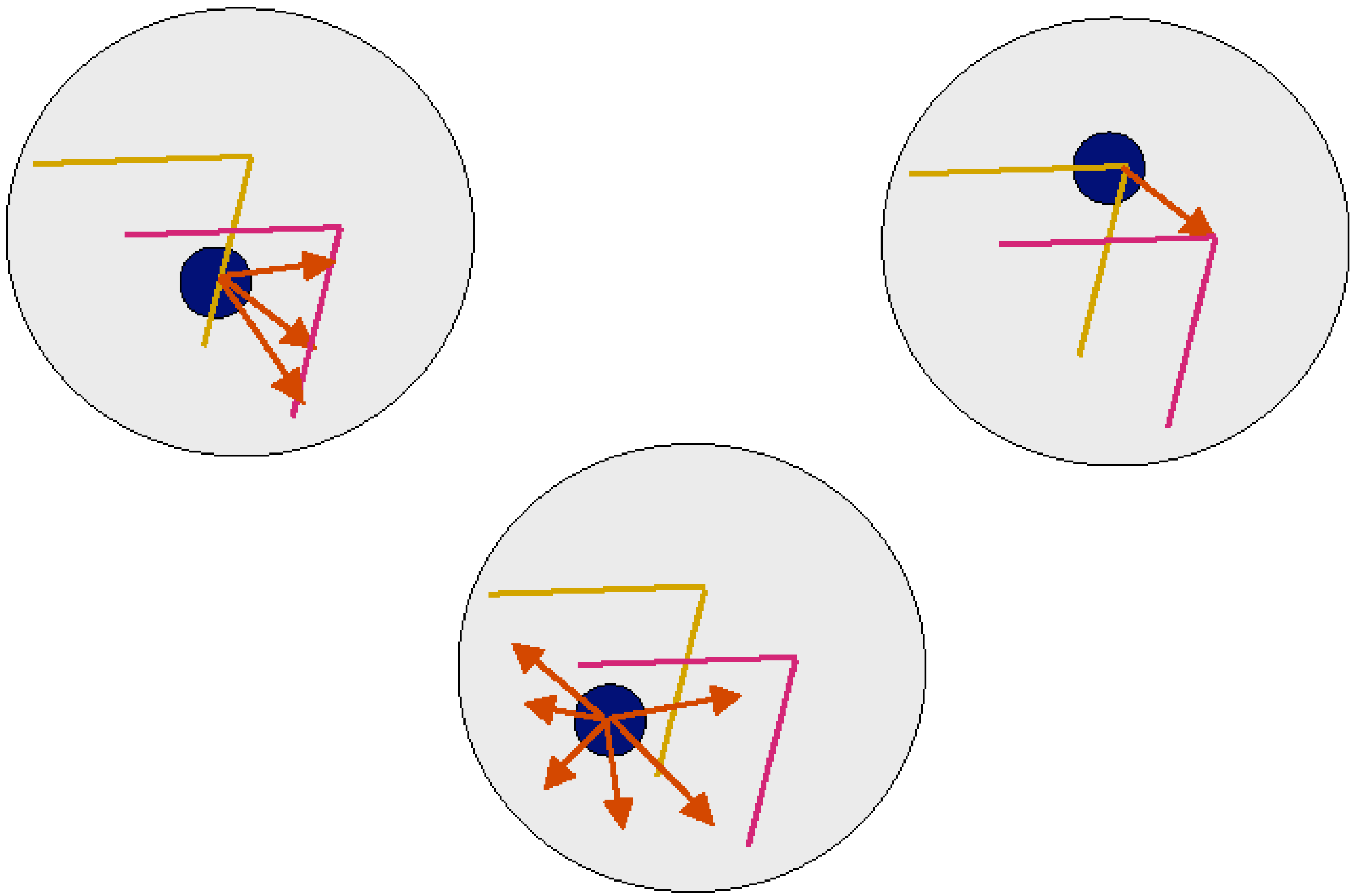
Defines a line in the *(u,v)* space

Normal Flow:

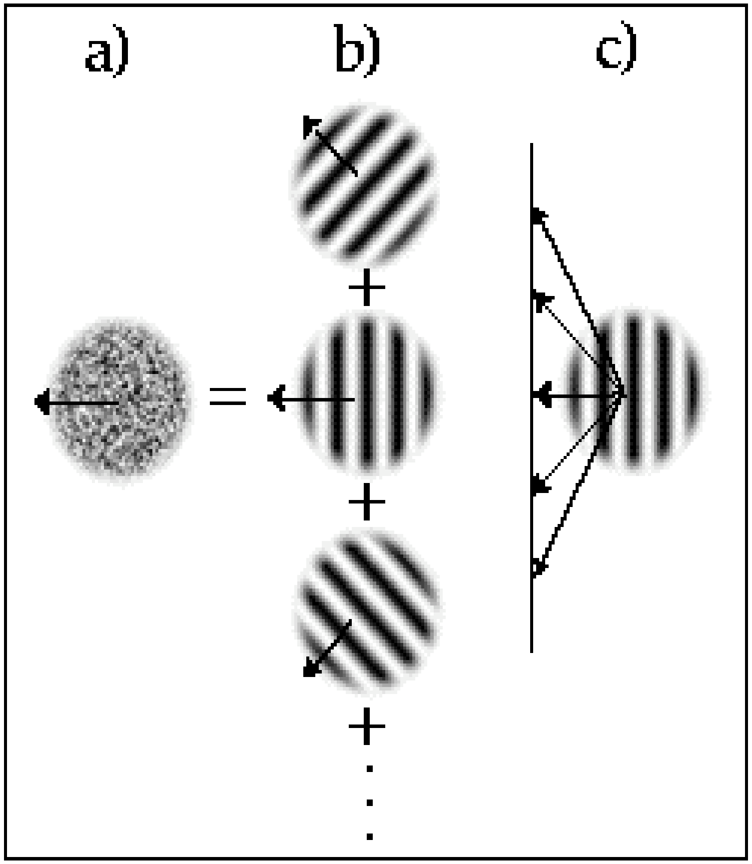
$$\boxed{u_{\perp} = -\frac{I_t}{|\nabla I|}\frac{\nabla I}{|\nabla I|}}$$



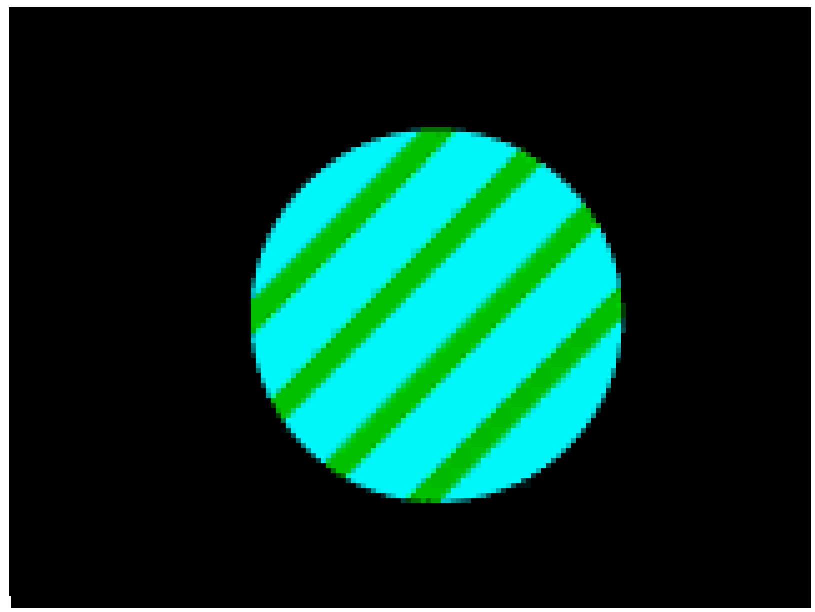
Local Patch Analysis



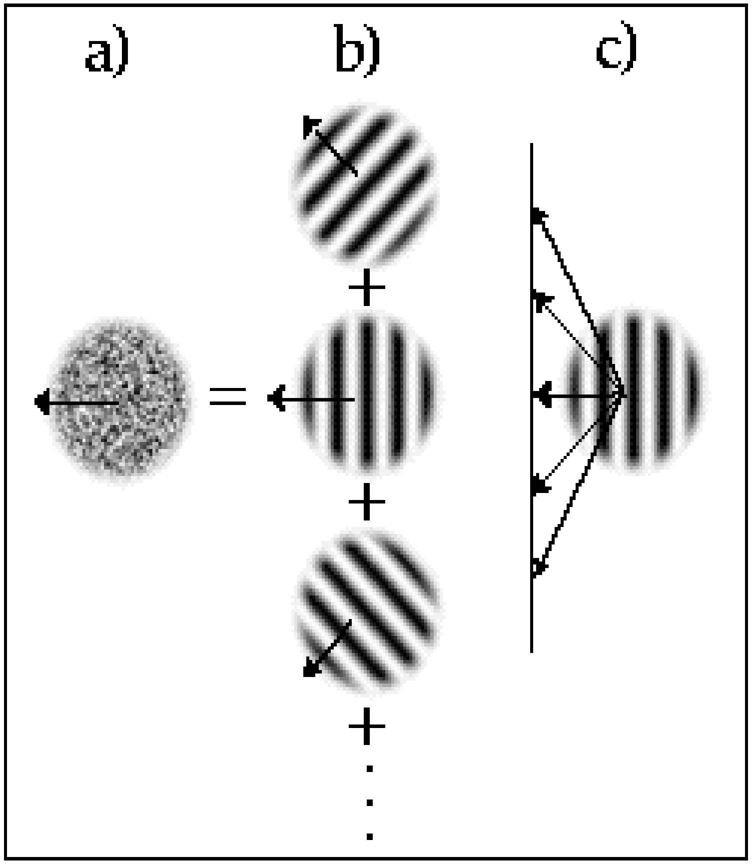
Aperture Problem (Motion/Form Ambiguity)



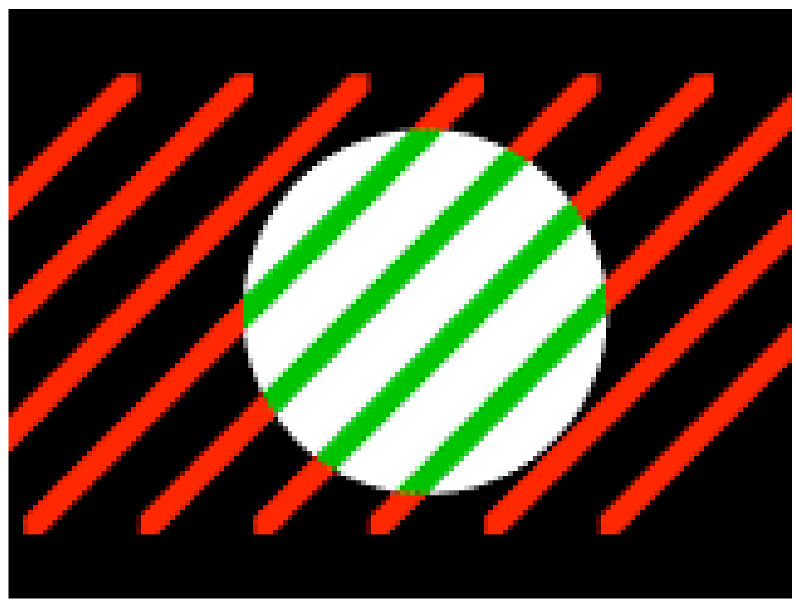
Result: Early visual measurements are ambiguous w.r.t. motion.



Aperture Problem (Motion/Form Ambiguity)



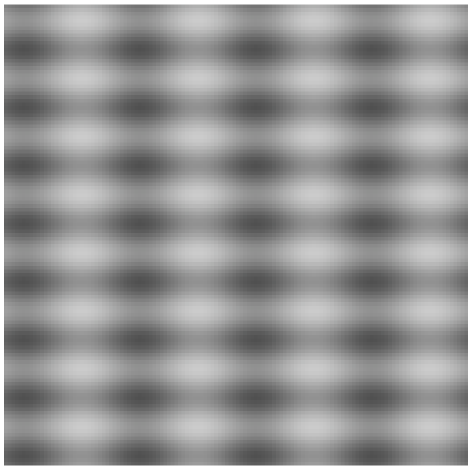
However, both the motion and the form of the pattern are implicitly encoded across the *population* of V1 neurons.



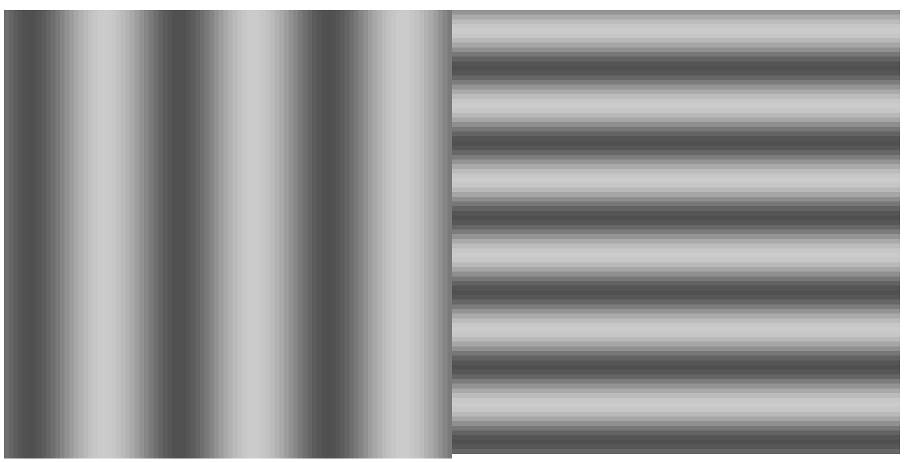
Actual motion

Plaids

Rigid motion

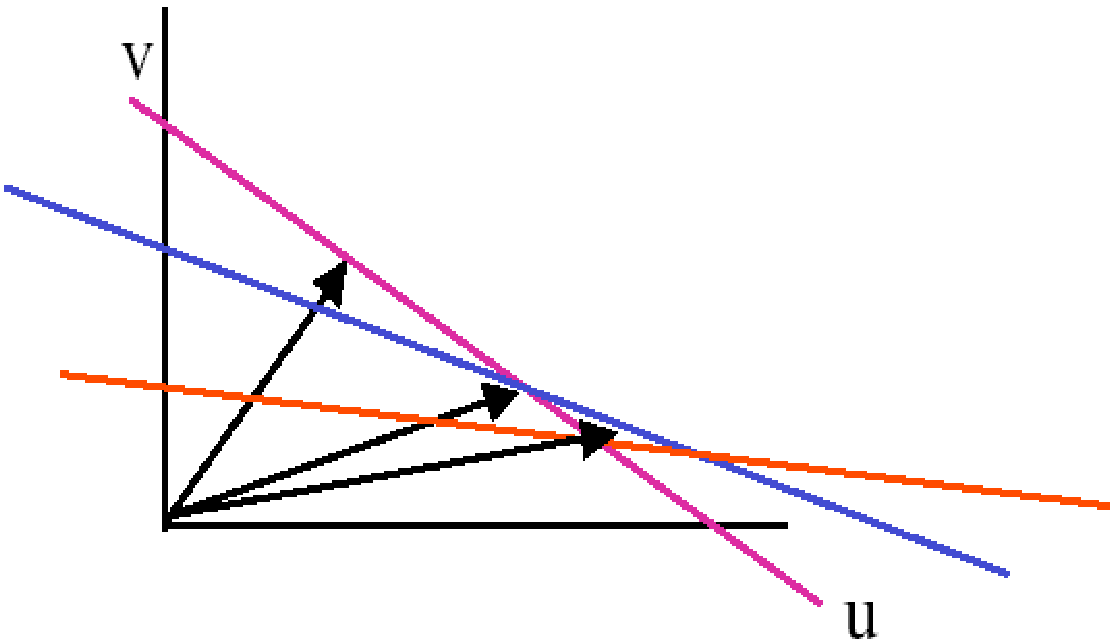


This pattern was created by super-imposing two drifting gratings, one moving downwards and the other moving leftwards.



Here are the two components displayed side-by-side.

Combining Local Constraints



$$\nabla I^1 \bullet U = -I_t^1$$

$$\nabla I^2 \bullet U = -I_t^2$$

$$\nabla I^3 \bullet U = -I_t^3$$

etc.

Find Least squares solution for multiple patches.

Motion processing as optimal inference

- Slow & smooth: A Bayesian theory for the combination of local motion signals in human vision, Weiss & Adelson (1998)

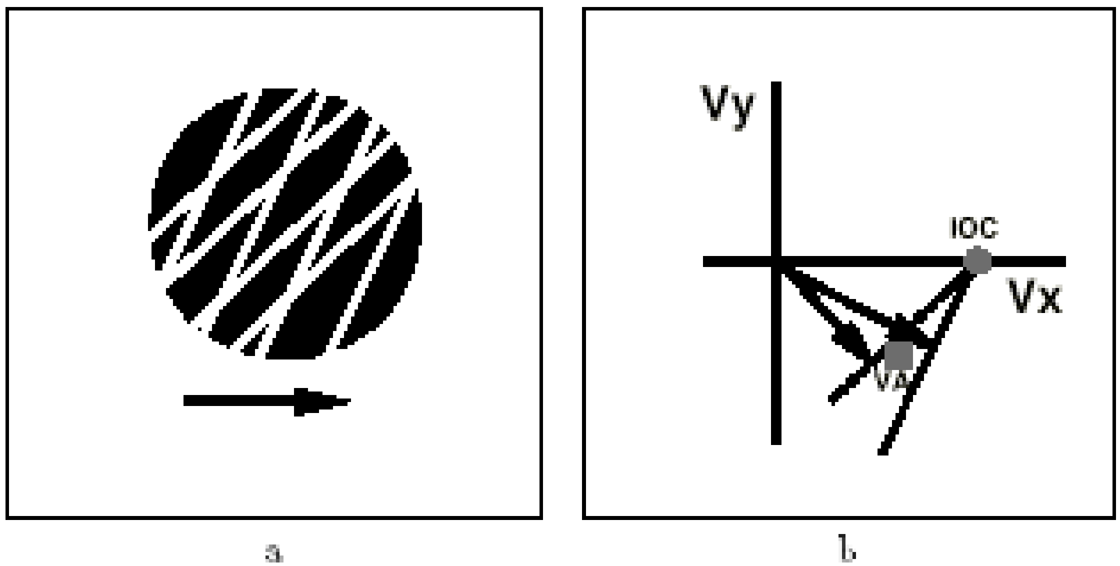
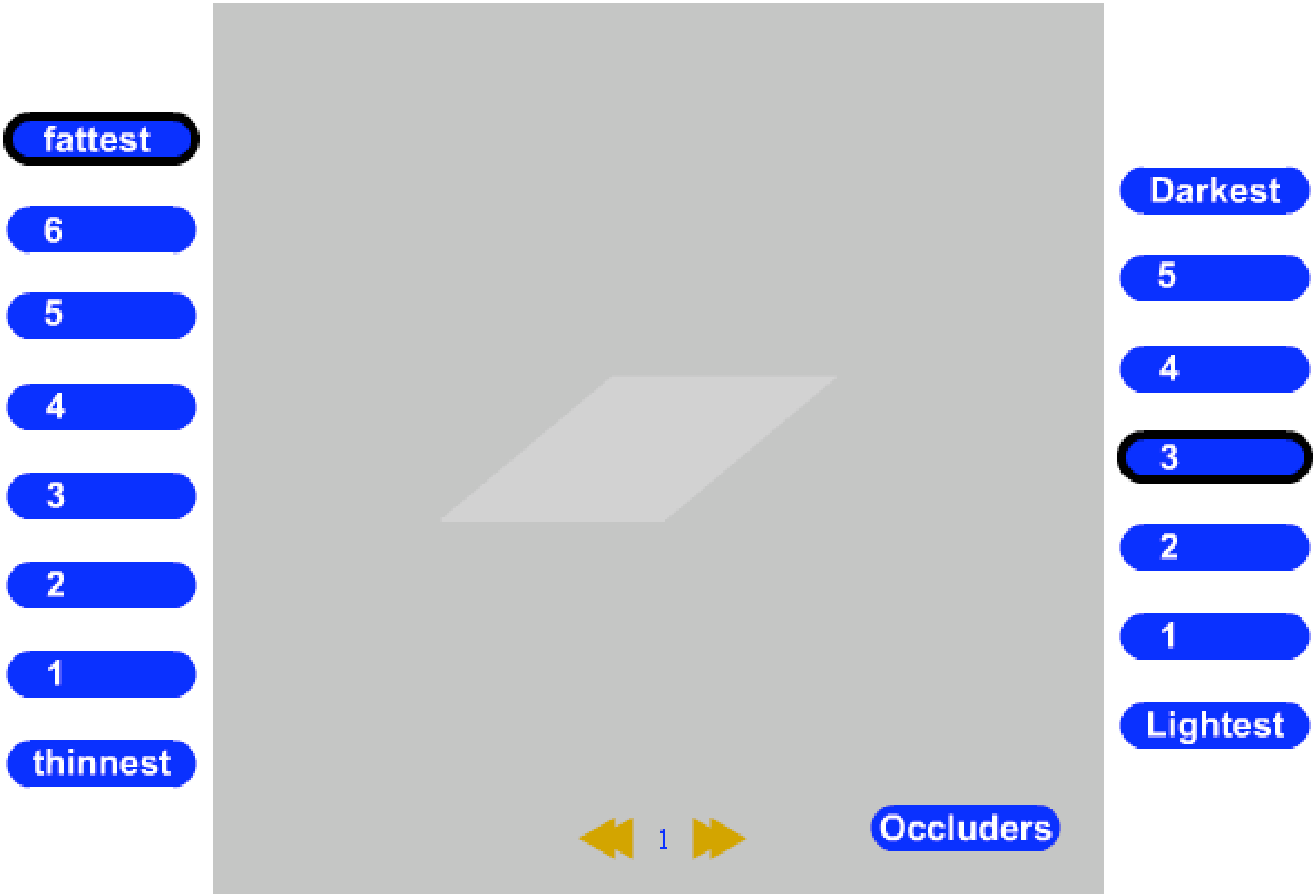


Figure from: Weiss & Adelson, 1998



Use the left blue buttons to change the size/shape of the moving figure; use the right blue buttons to change the color of the background. Use the yellow buttons to control the speed.

Modeling motion estimation

Local likelihood: $L(v) \propto e^{-\sum_r w(r)(I_x v_x + I_y v_y + I_t)^2 / 2\sigma^2}$

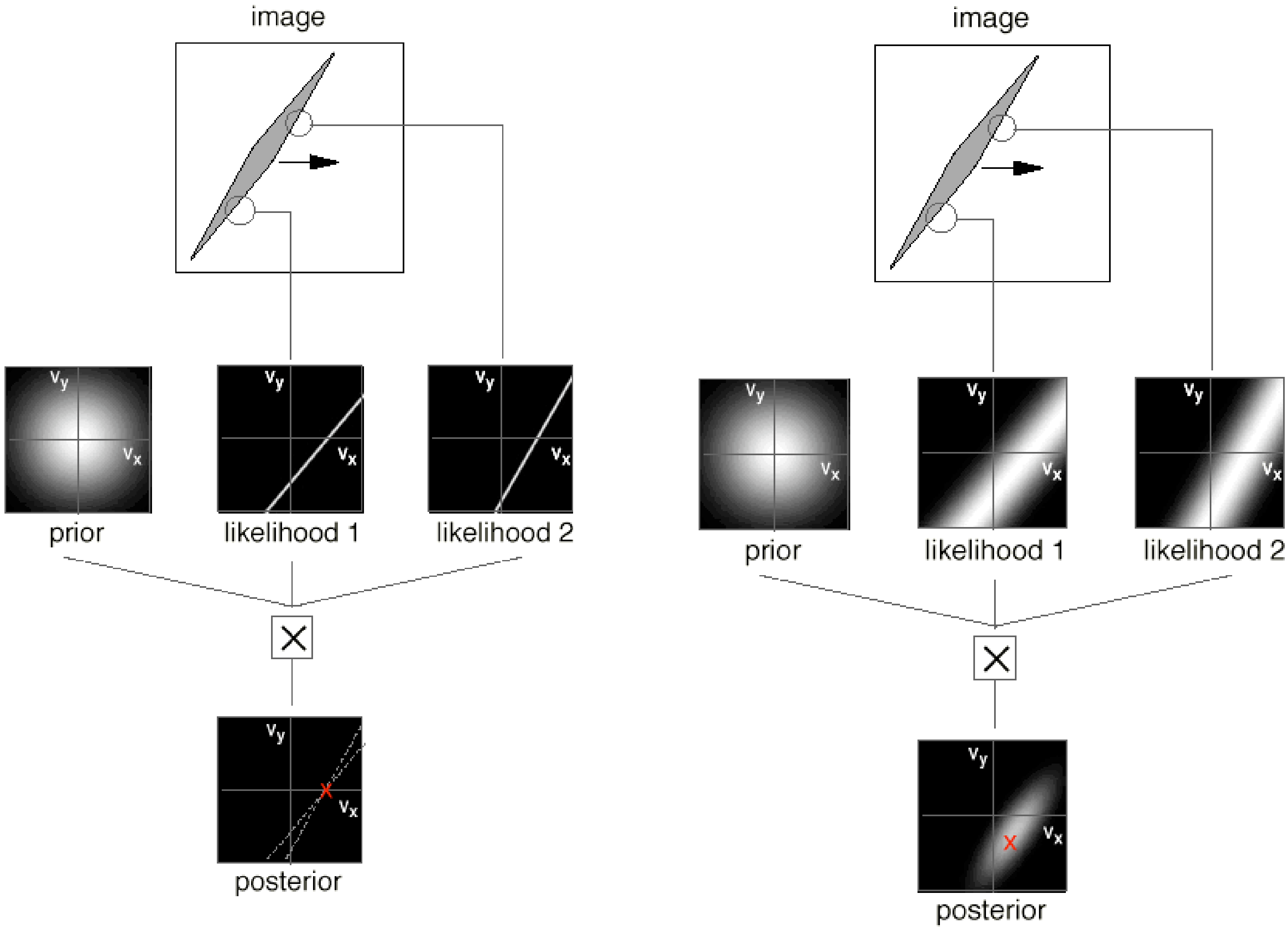
Global likelihood: $L_r(v) \rightarrow p(I \mid \theta) \propto \prod_r L_r(\theta)$

Prior: $P(V) \propto e^{-\sum_r (Dv)^t(r)(Dv)(r)/2}$

$P(V) \rightarrow P(\theta)$

Posterior: $P(\theta \mid I) \propto P(I \mid \theta) P(\theta)$

From: Weiss & Adelson, 1998



Figures from: Weiss & Adelson, 1998

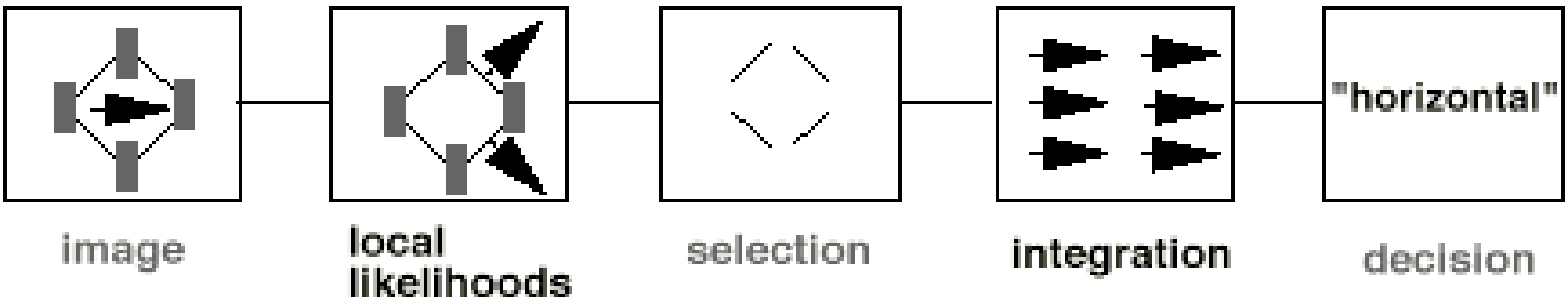
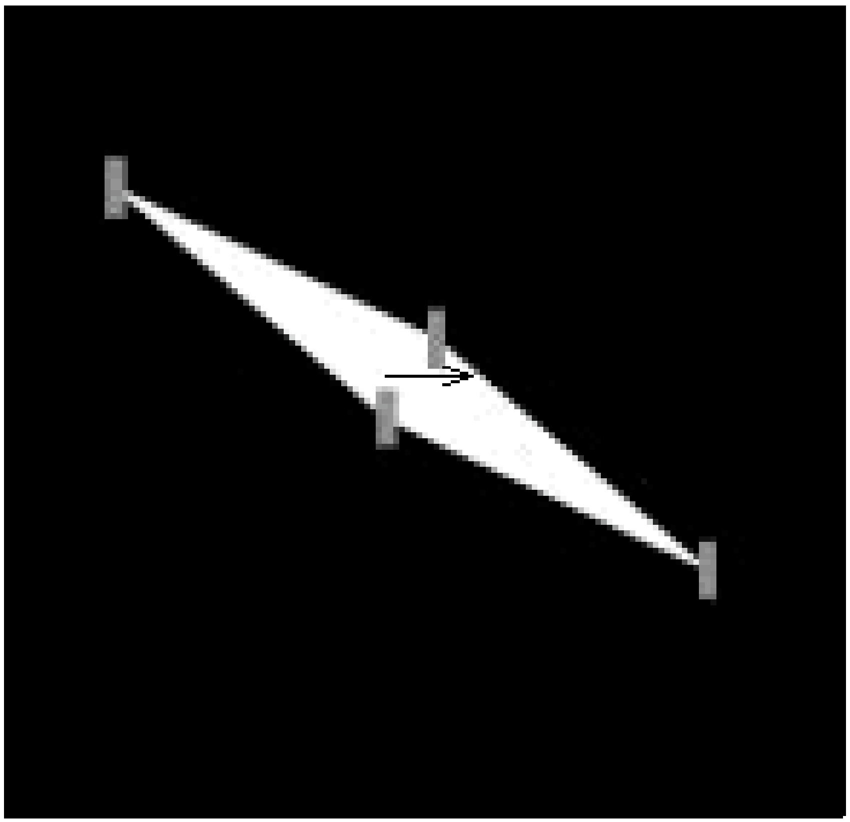
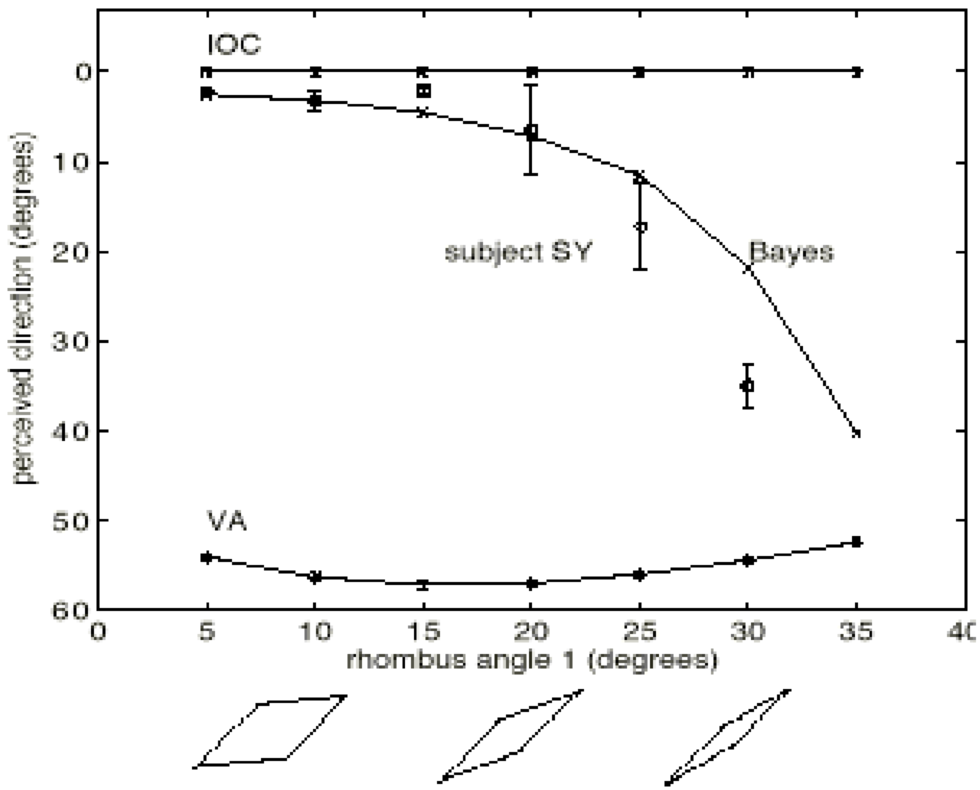


Figure from: Weiss & Adelson, 1998

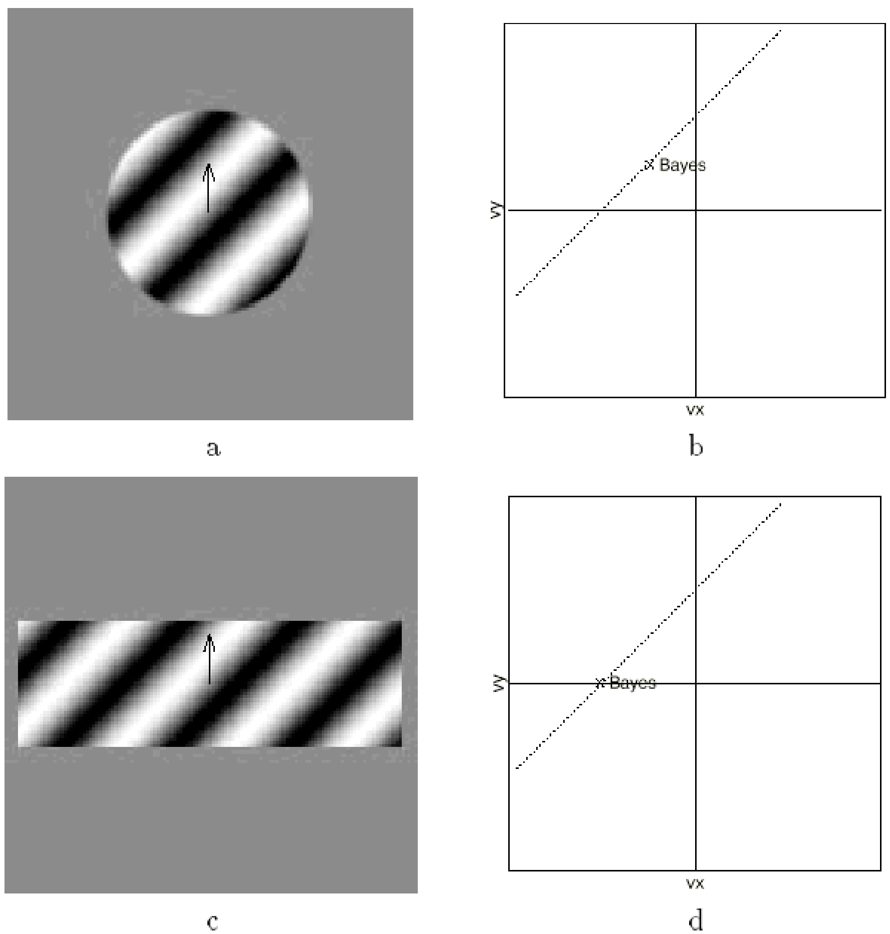


a



b

Figure from: Weiss & Adelson, 1998



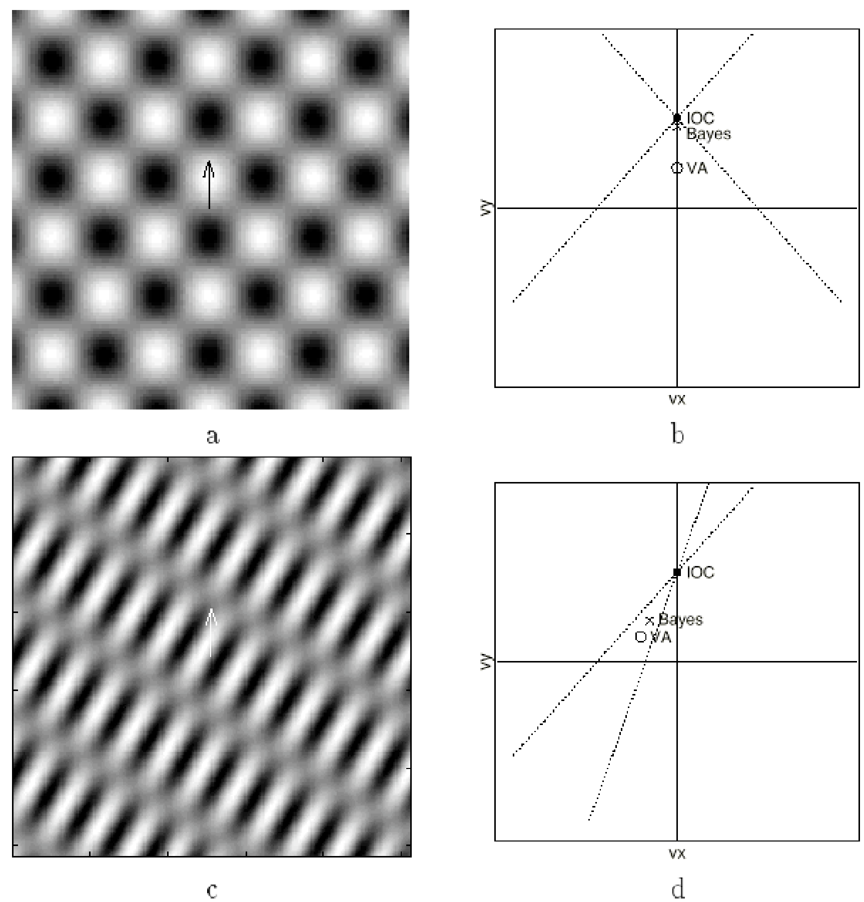
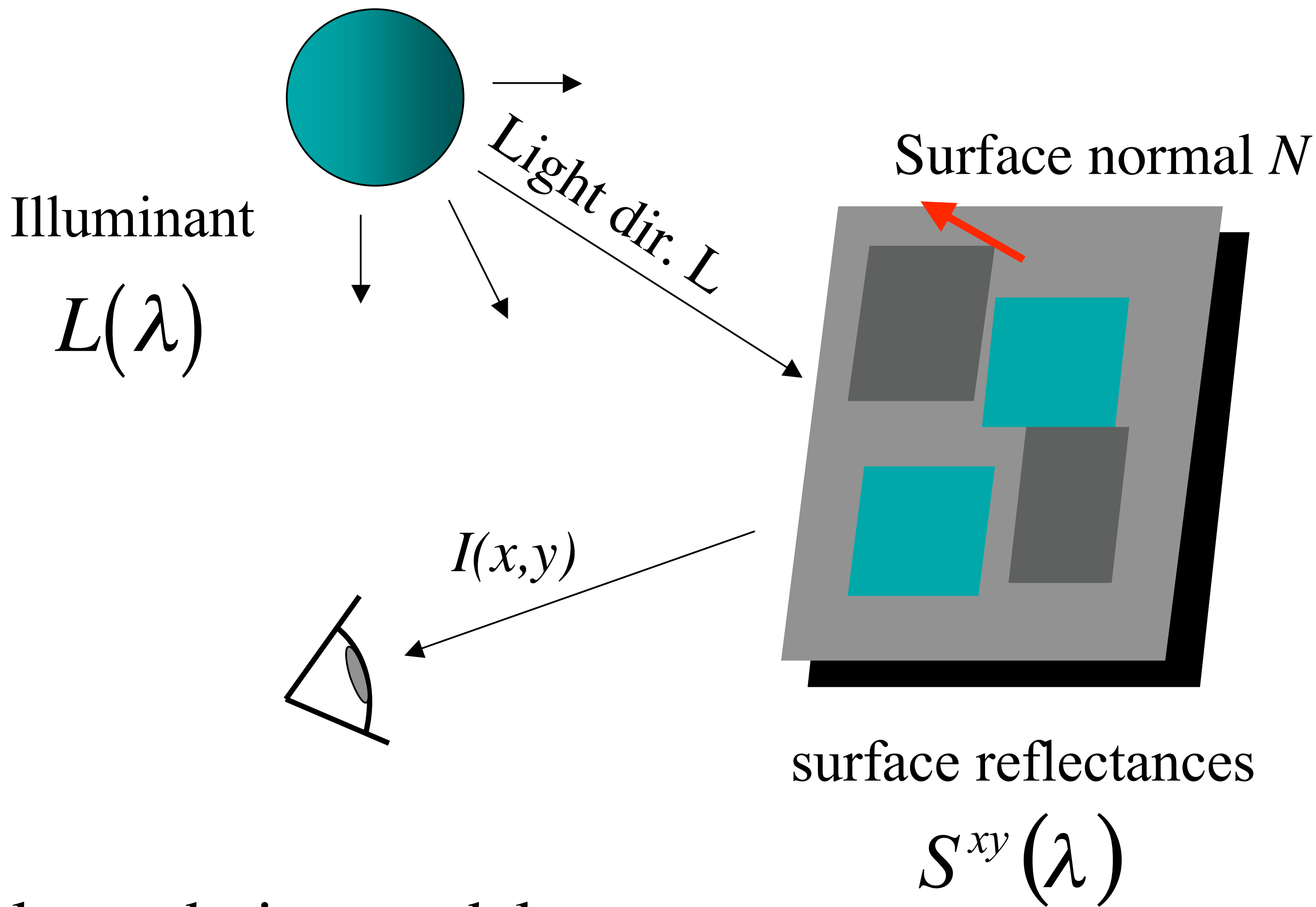


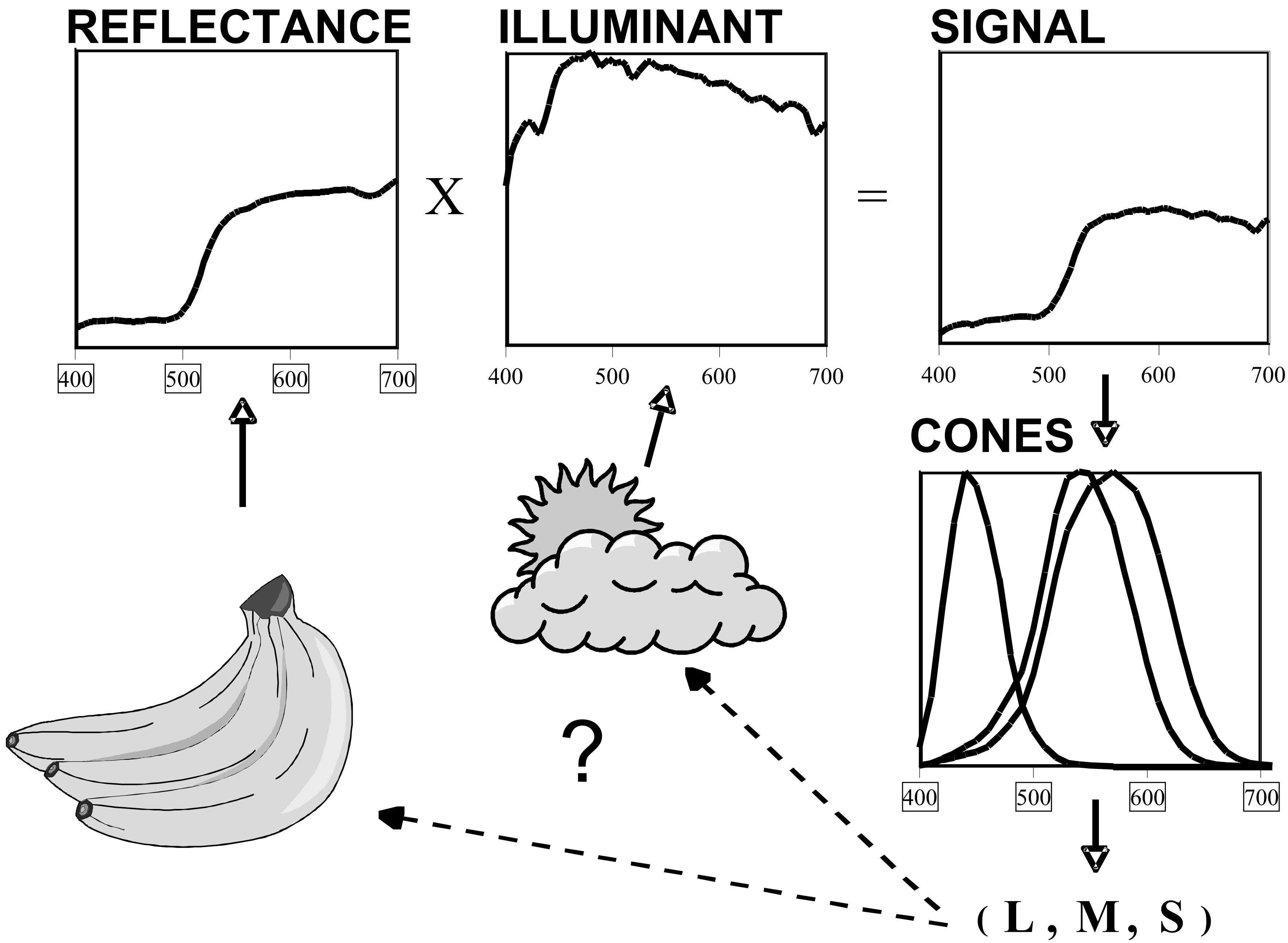
Figure from: Weiss & Adelson, 1998

Lightness perception as optimal inference

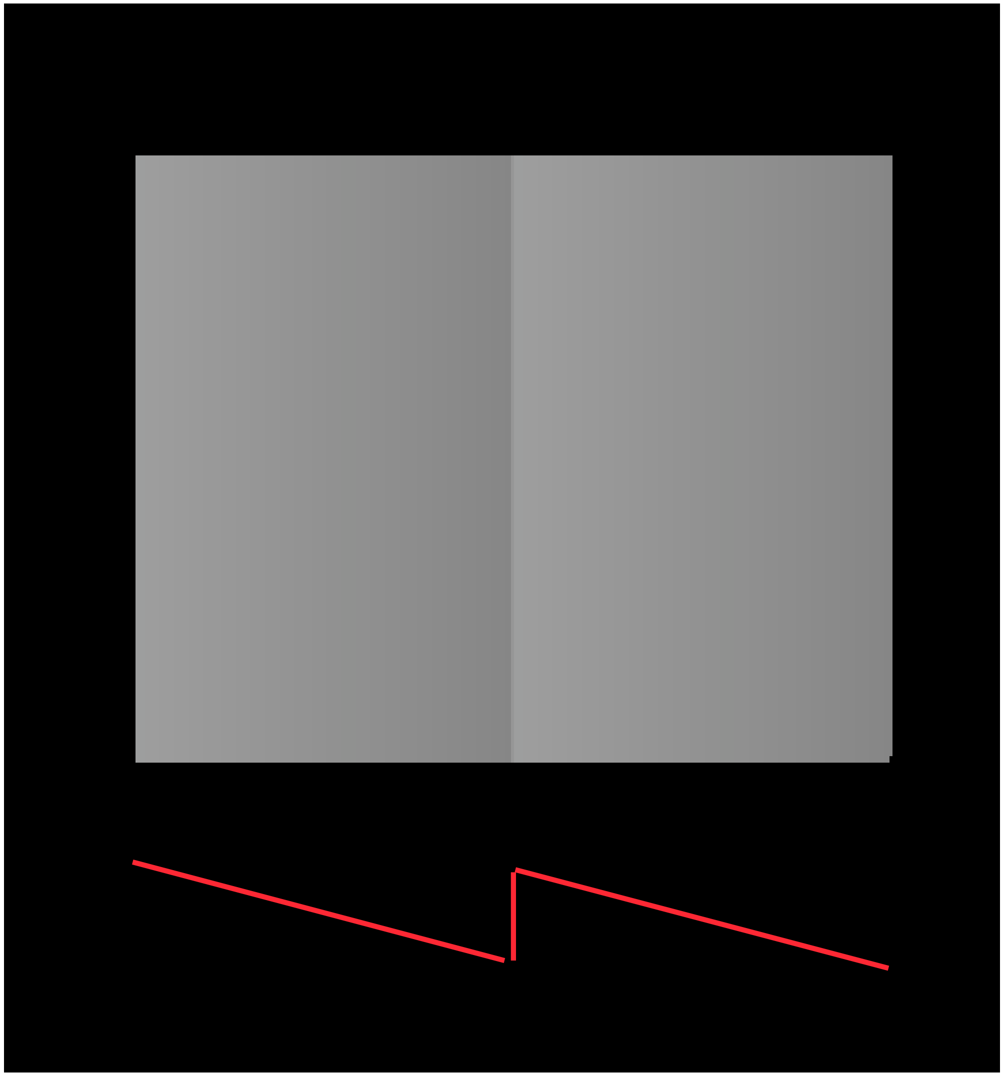


Simple rendering model

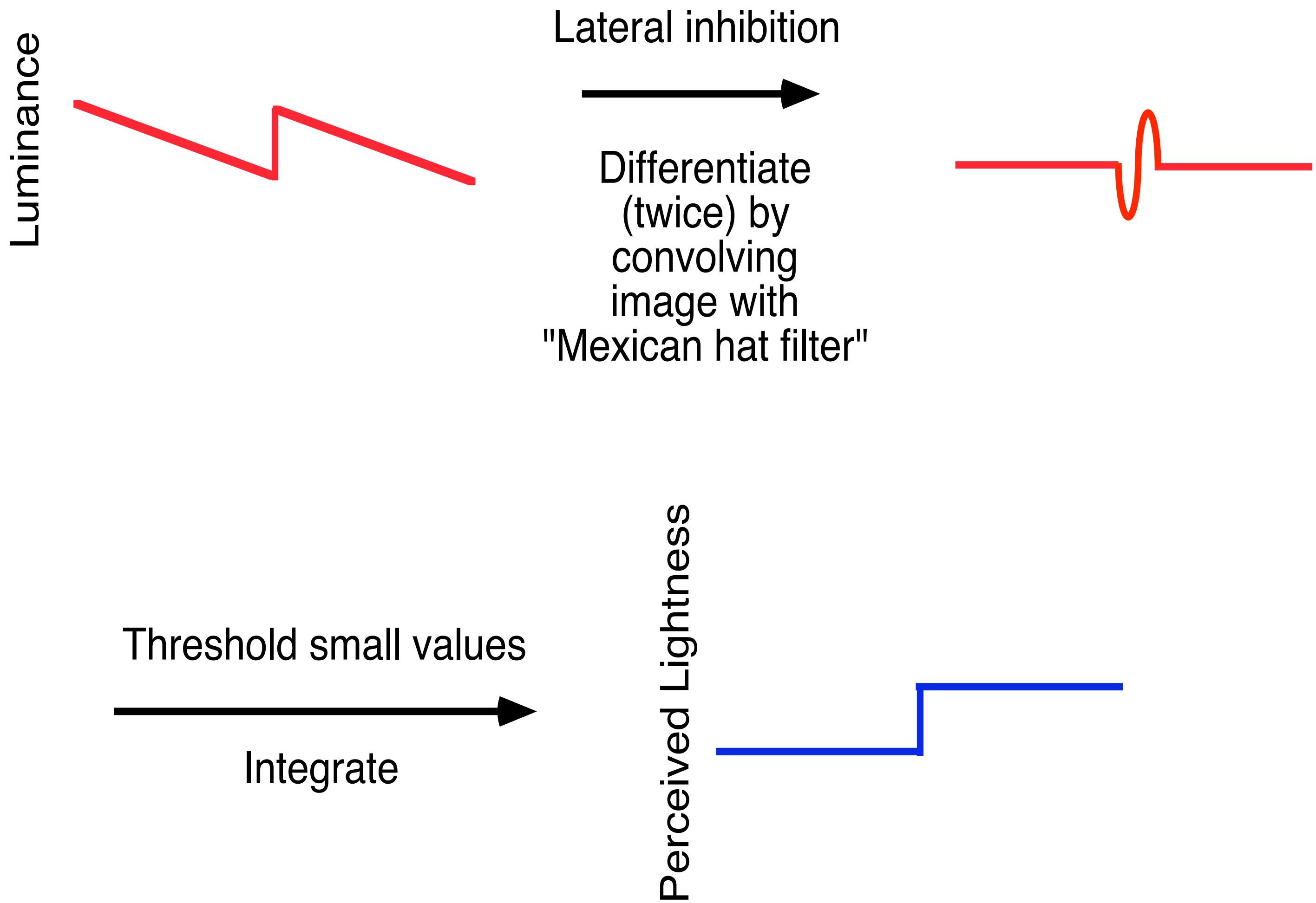
$$I(x,y) = S^{xy}(\lambda)\left[\vec{L}(\lambda)\cdot\vec{N}(x,y)\right]$$


REFLECTANCE
X
ILLUMINANT
=
SIGNAL
400
500
600
700
400
500
600
700
400
500
600
700
CONES
400
500
600
700
?
(L , M, S)

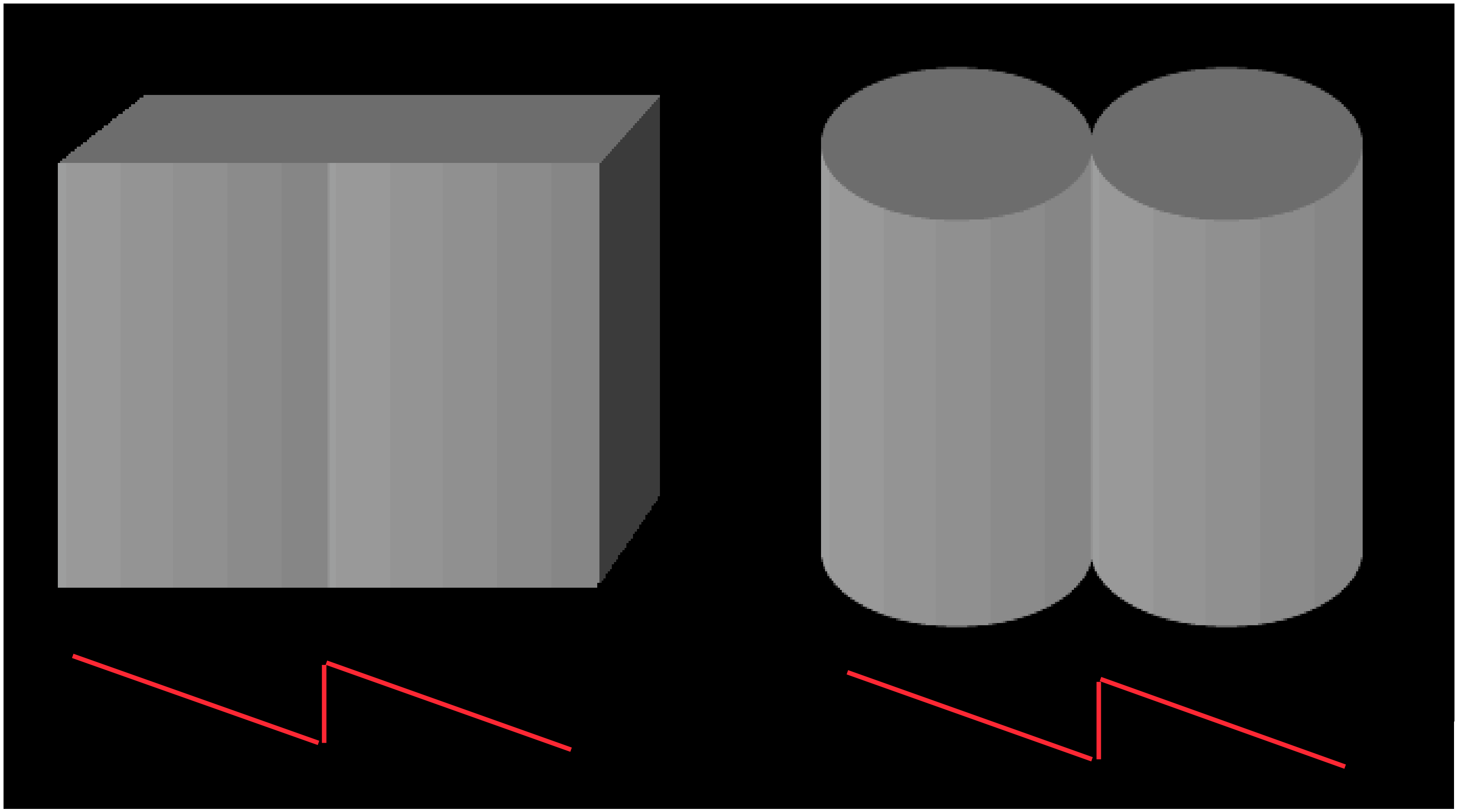
Land & McCann's lightness illusion



Neural network filter explanation



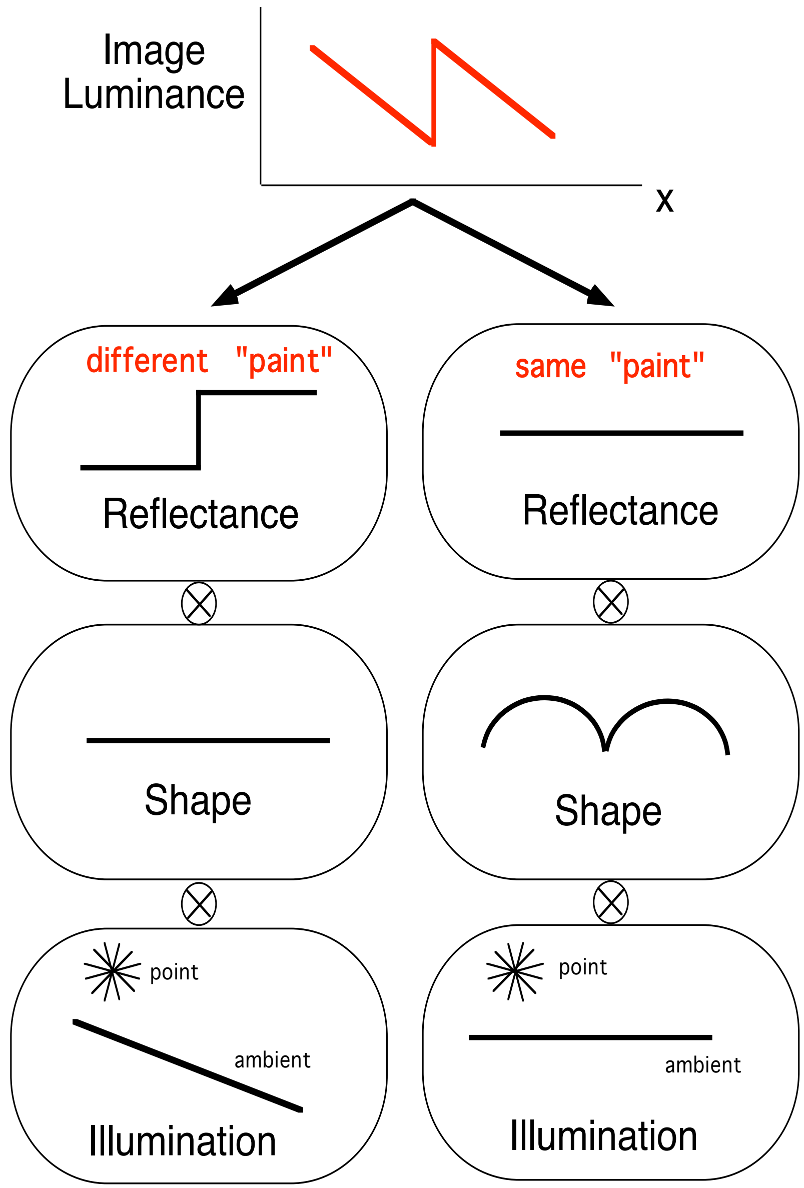
Apparent surface shape affects lightness perception



- Knill & Kersten (1991)

Inverse graphics solution

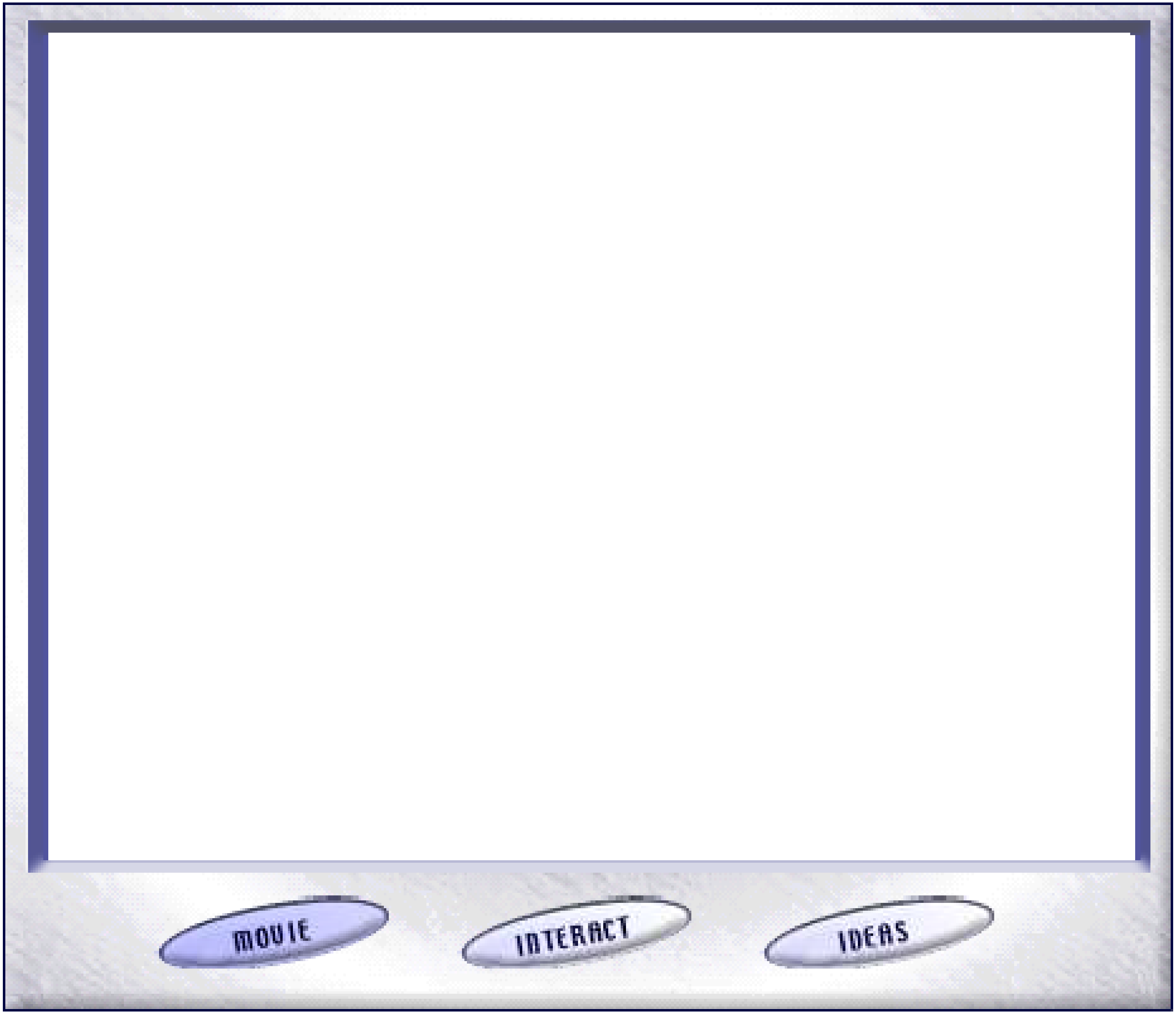
What model of material reflectances, shape, and lighting fit the image data?

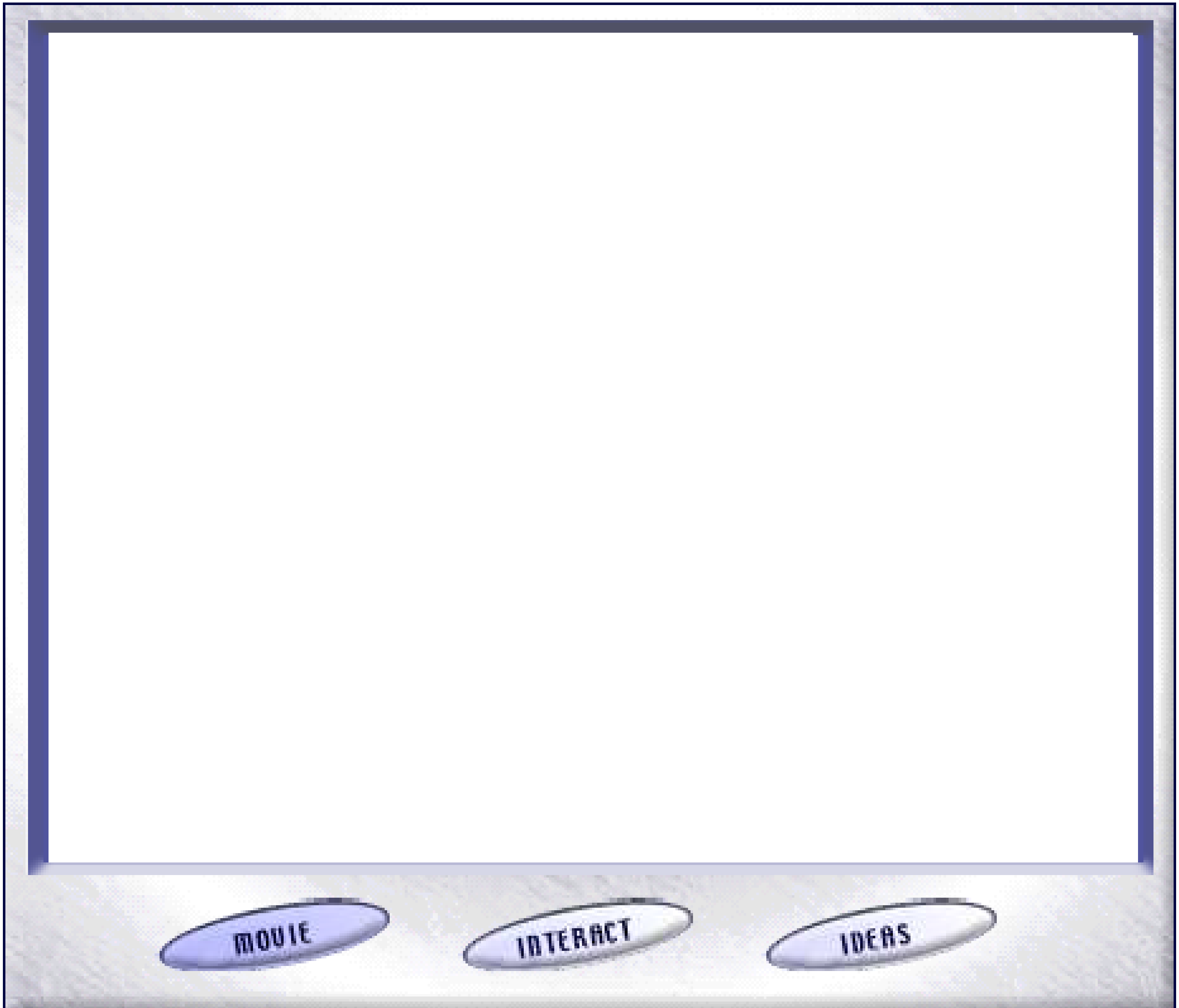


MOVIE

INTERACT

IDEAS


MOVIE
INTERACT
IDEAS


MOVIE
INTERACT
IDEAS