Motion Integration Cosmo 2012

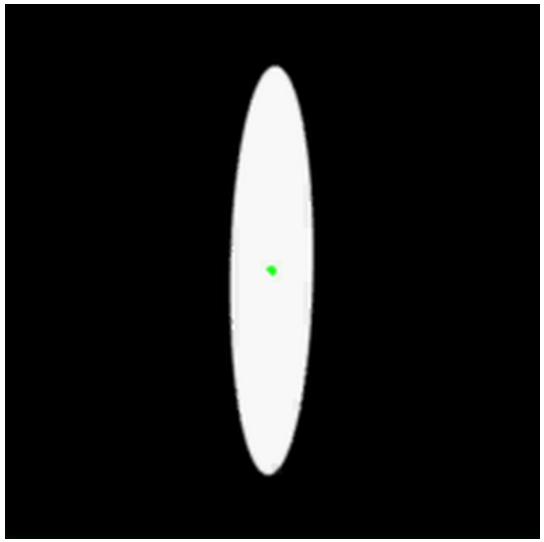
Paul Schrater

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.

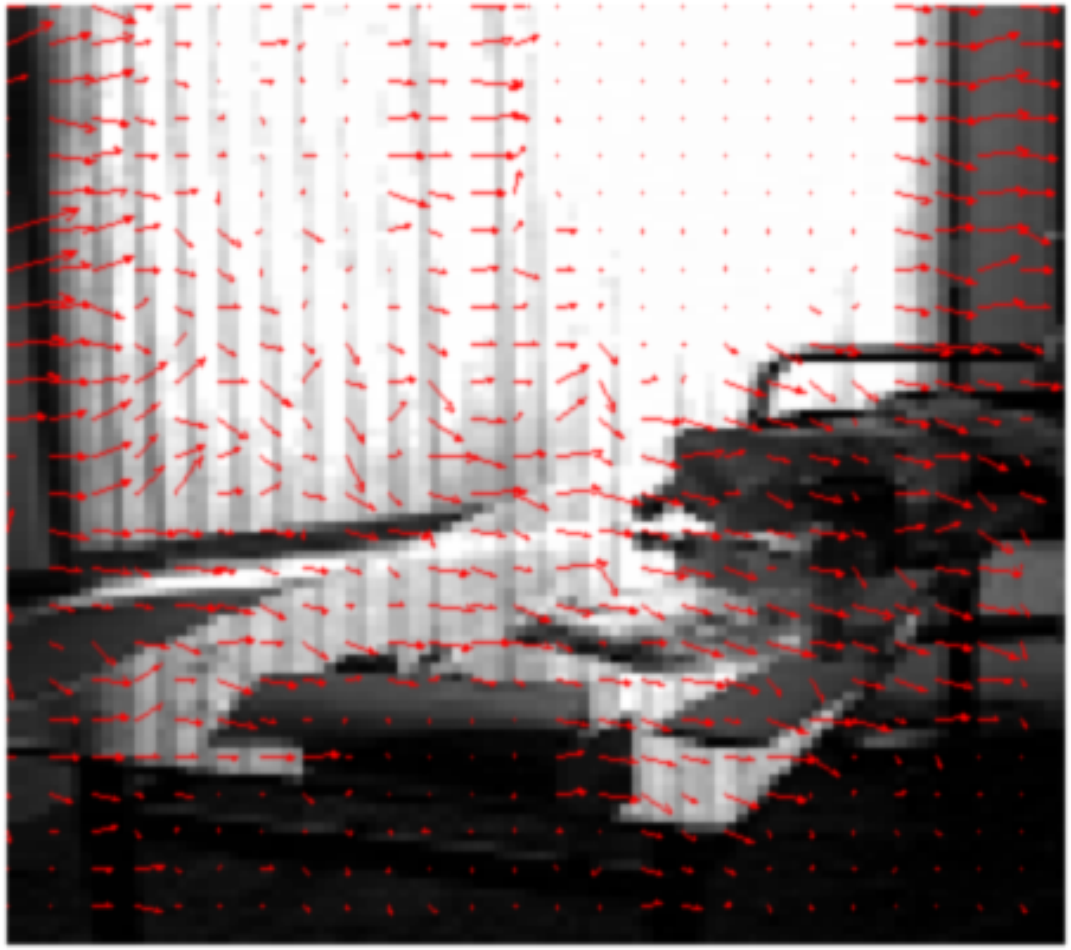
  $I(x,y,t)$ $\quad$ $v_{OBJ}(x,y,z,t)$

Motion Perception as Cue Integration

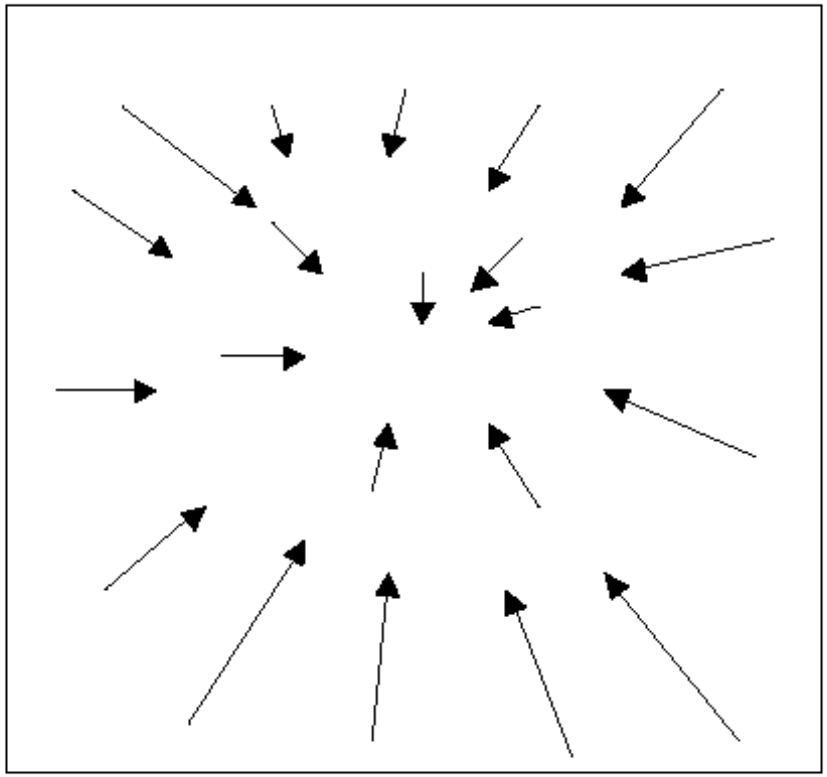
Local Translations

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

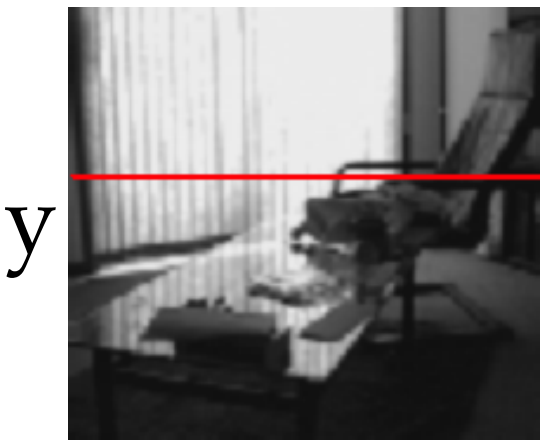
Time ~100msec
Space ~1-10deg

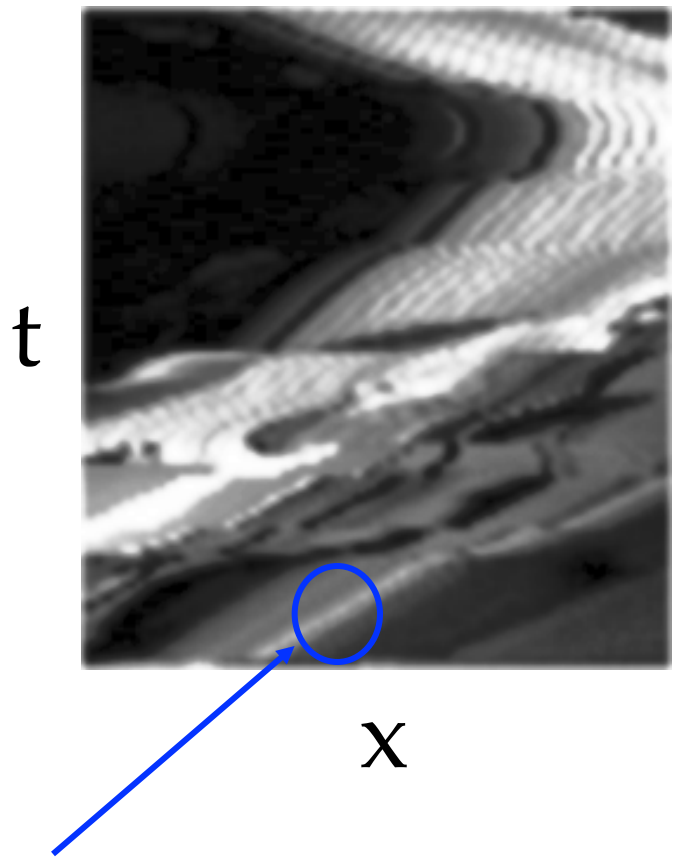


Larger goal is to estimate self/object motion

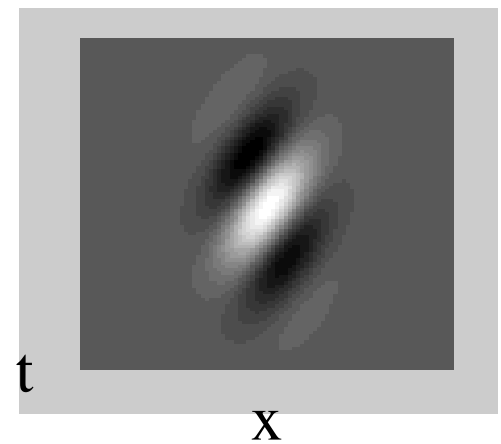


X-T Slice of Translating Camera





Simple signals for V1



Differential approach: Optical flow constraint equation

Brightness should stay constant as you track motion

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

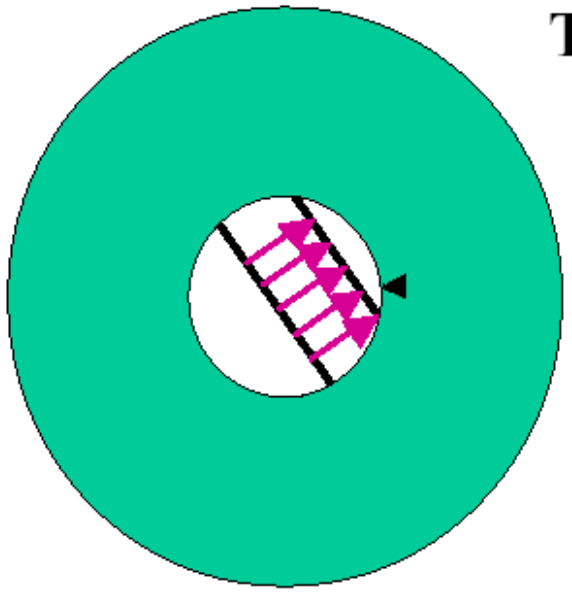
1st order Taylor series, valid for small δt

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

Constraint equation

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

"BCCE" - Brightness Change Constraint Equation

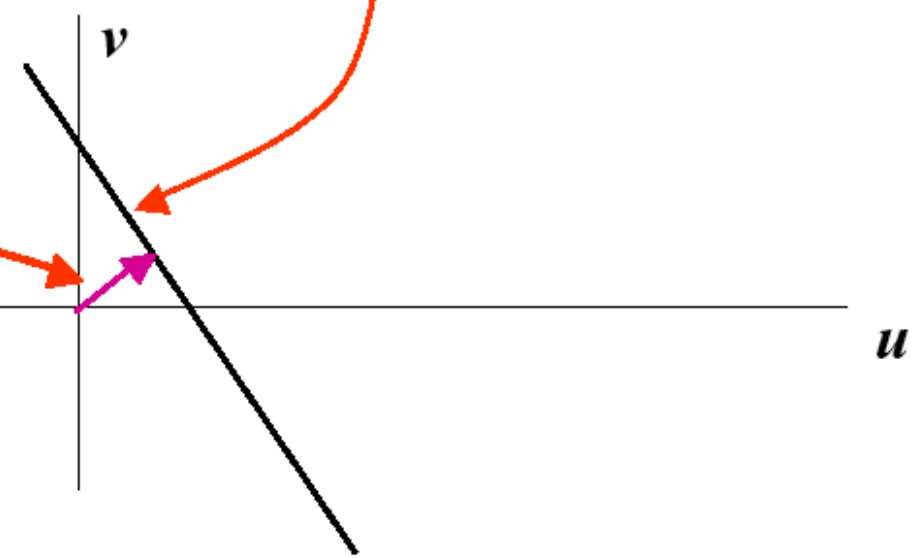


The gradient constraint:

$$I_x u + I_y v + I_t = 0$$

$$\nabla I \bullet \vec{U} = 0$$

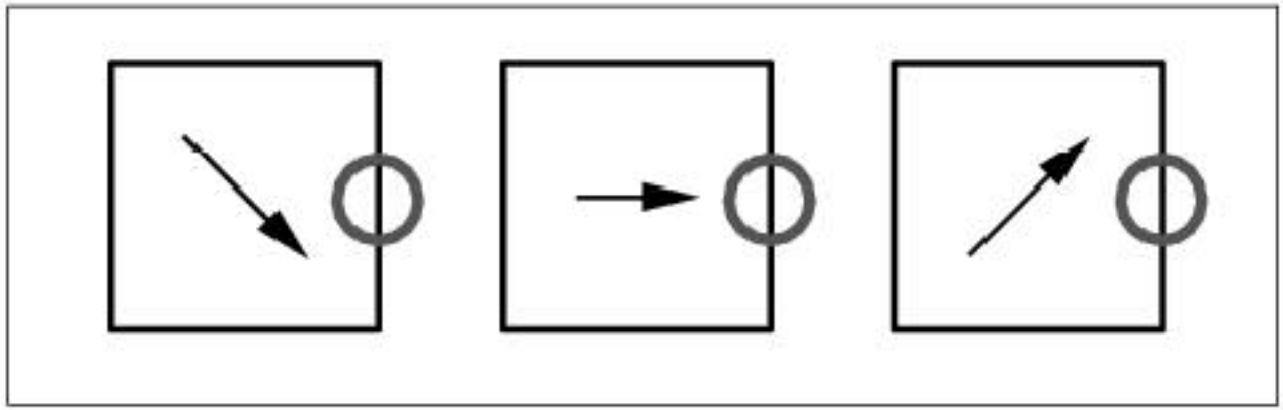
Defines a line in the *(u,v)* space



Normal Flow:

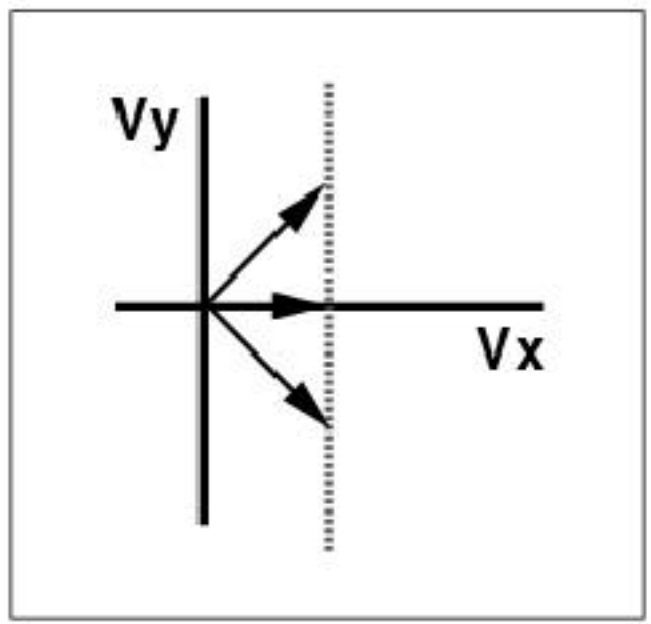
$$u_{\perp} = -\frac{I_t}{|\nabla I|}\frac{\nabla I}{|\nabla I|}$$

Object motion via Edge integration



a

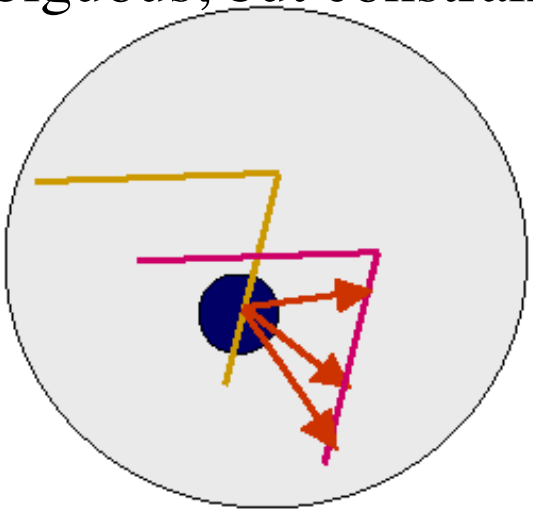
Normal flow:
Motion component in the direction of the edge



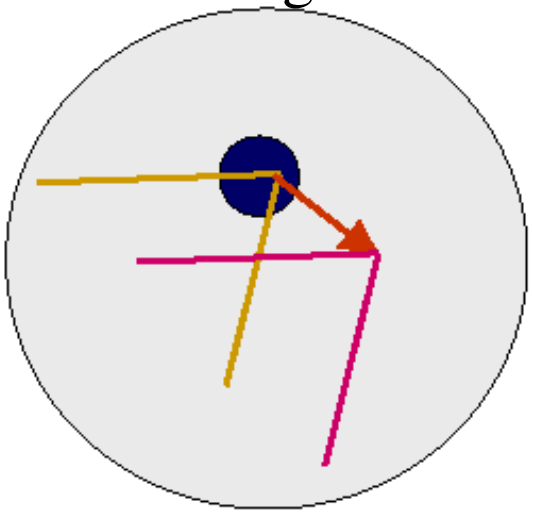
b

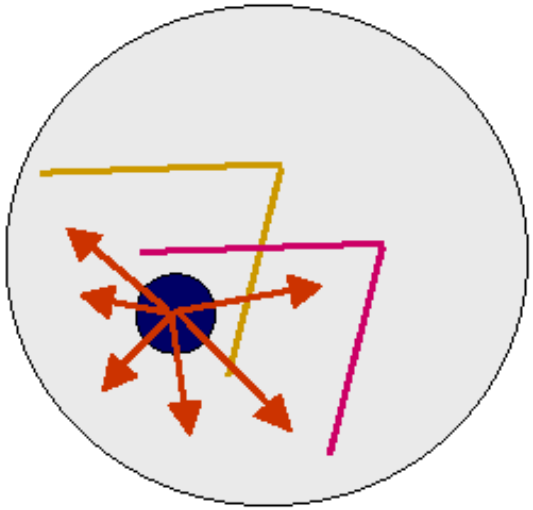
Local Patch Analysis

Ambiguous, but constrained



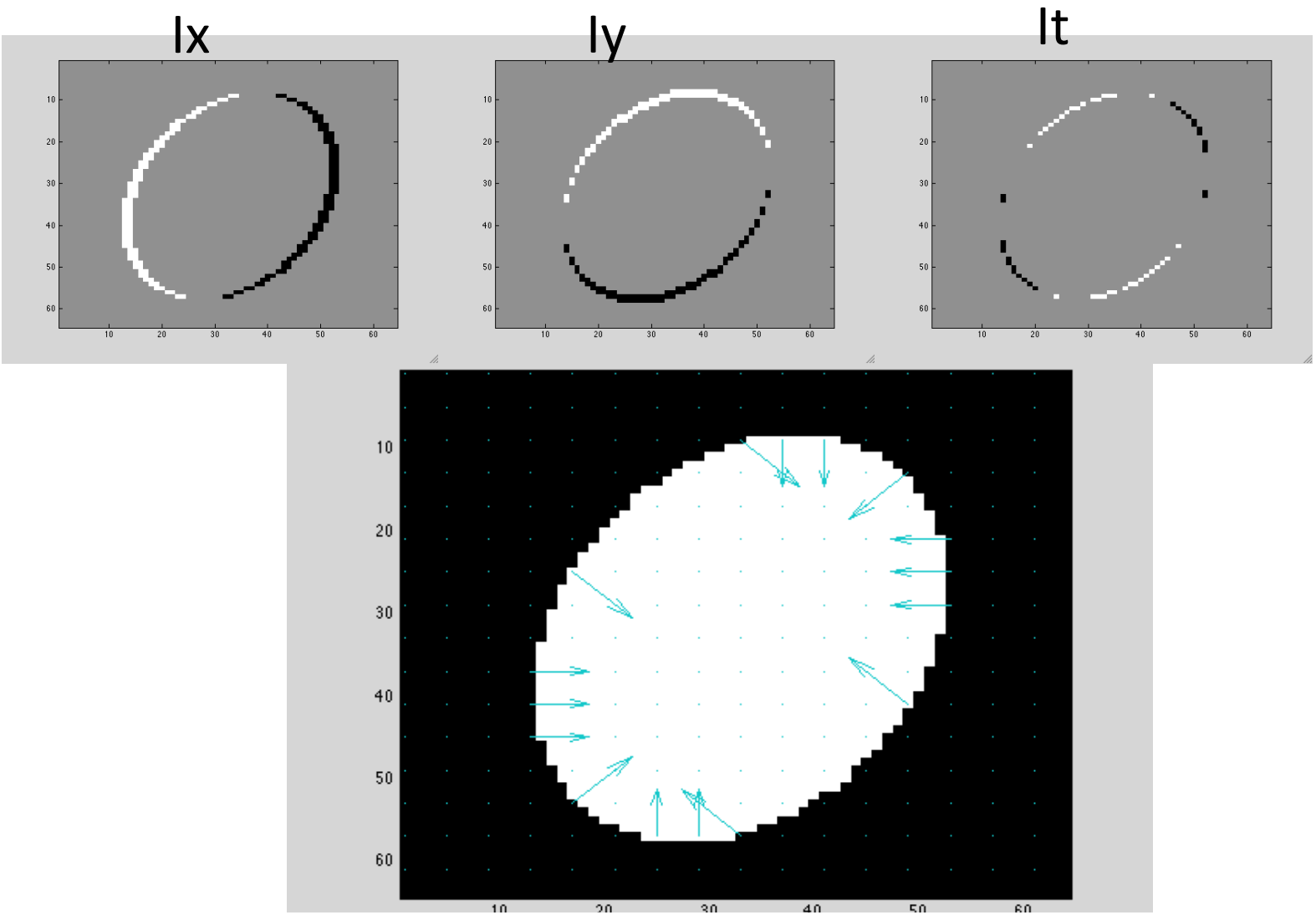
Unambiguous





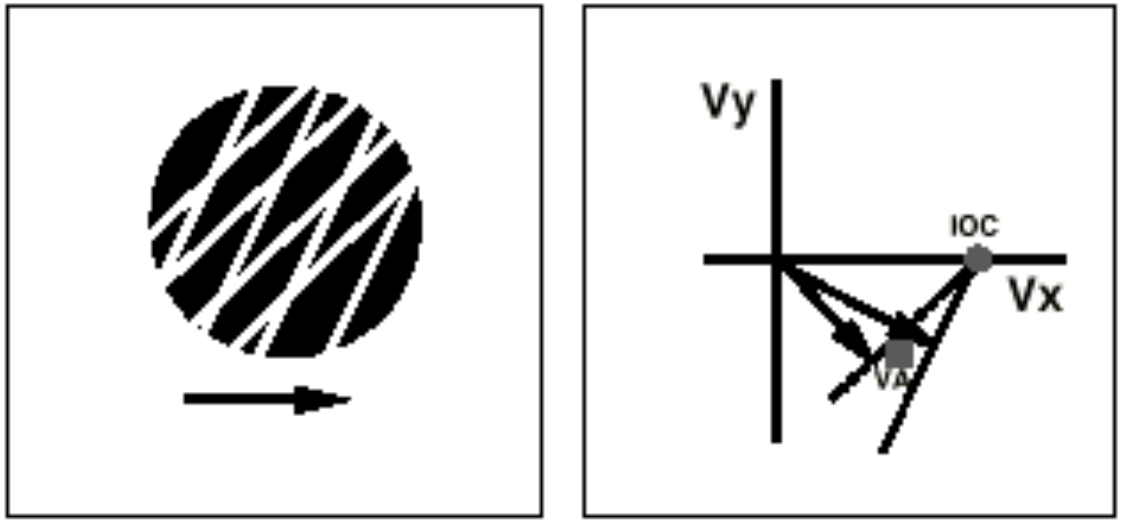
Ambiguous

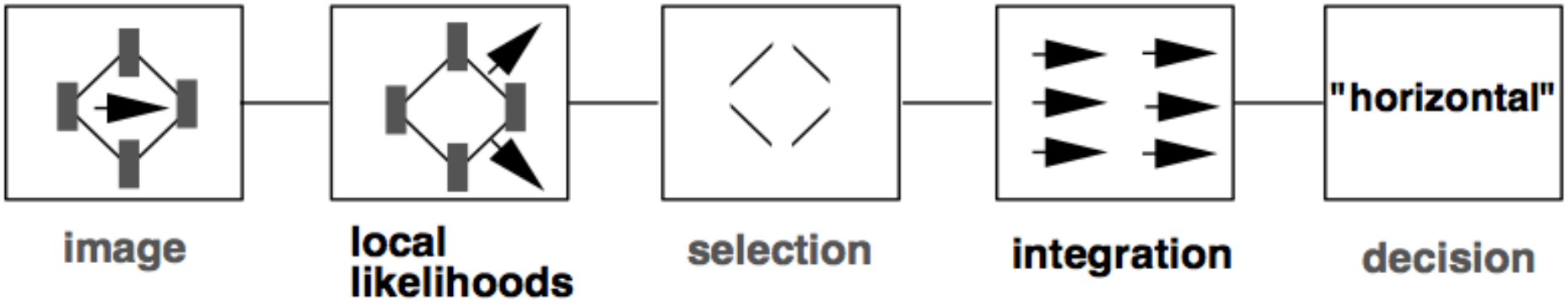
Visualizing Derivatives



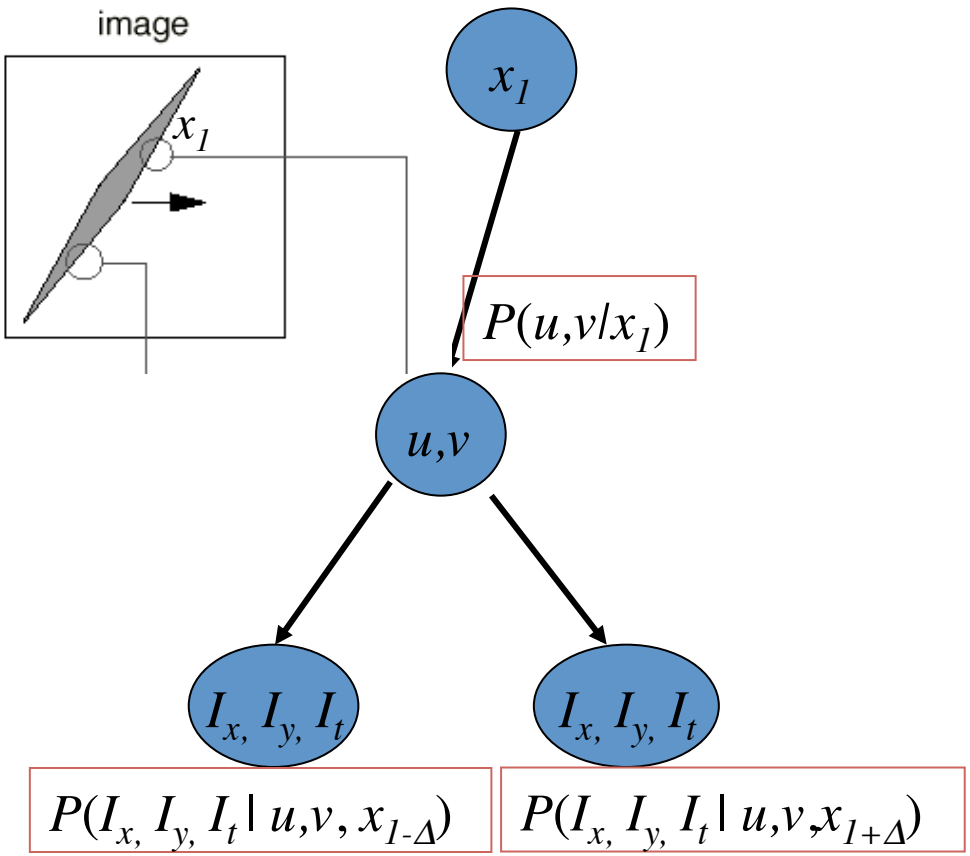
Motion processing as cue combination

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



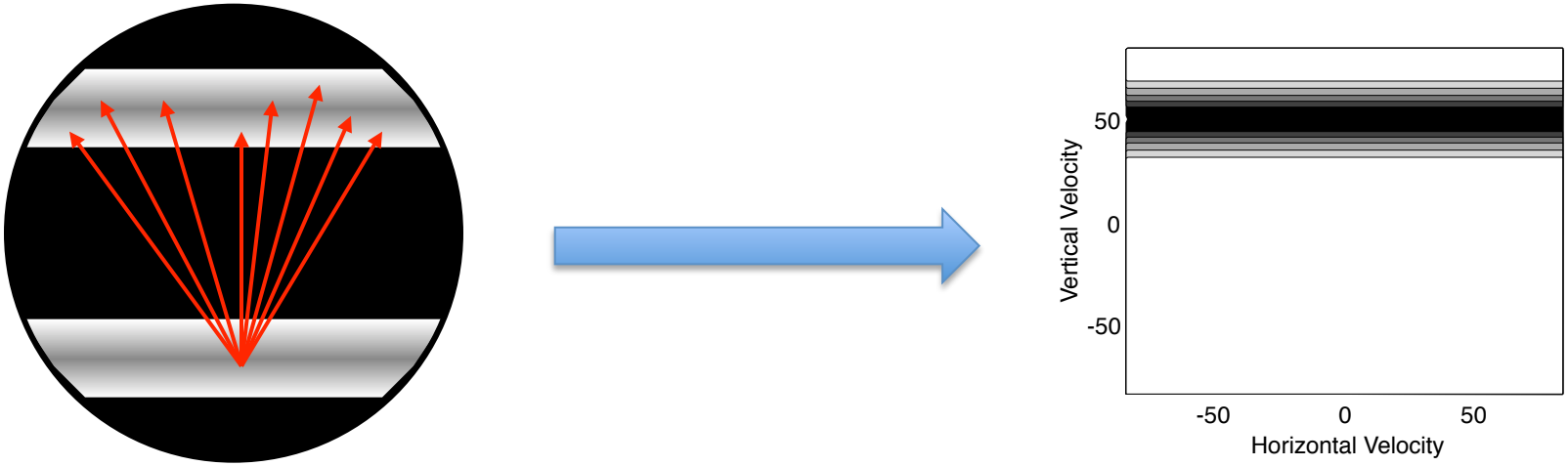


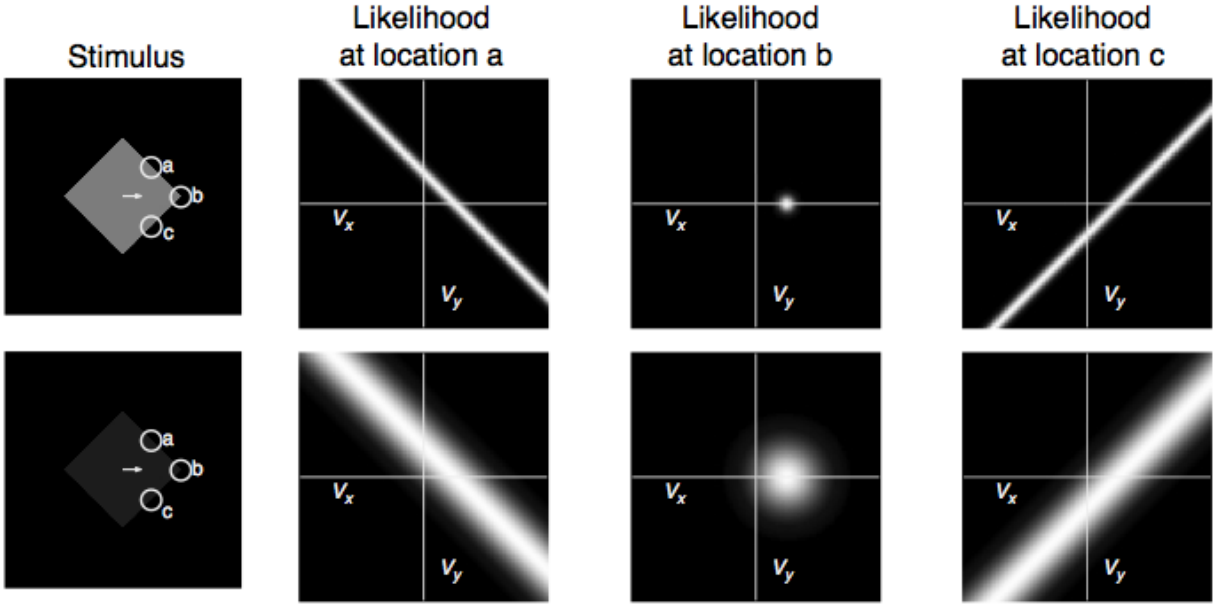
As a Graphical Model: Local



Integrate intensities
locally around x_1 and across time

Likelihoods

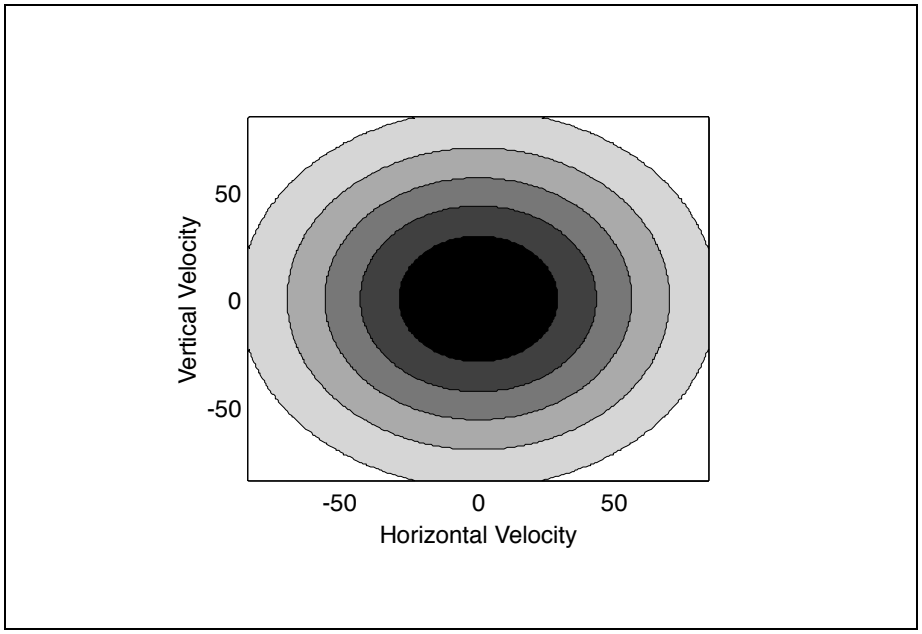




Likelihood functions for three local patches of a horizontally translating diamond stimulus, computed using equation (4). Intensity corr
ɔability. Top, high-contrast sequence. Bottom, low-contrast sequence, with the same parameter σ. At edges, the local likelihood is a 'fu
line; at corners, the local likelihood peaks around the veridical velocity. The sharpness of the likelihood decreases with decreasing cont

Assume a single prior on velocity for all

- Human observers favor slow motions



Likelihood

$$I(x,y,t) = I(x - v_x\Delta t, y - v_y\Delta t, t - \Delta t) + \eta$$

$$I(x,y,t) - I(x - \mathbf{v}_x\Delta t, y - \mathbf{v}_y\Delta t, t - \Delta t) = \eta$$

$$P(I(x_i,y_i,t)\,|\,\mathbf{v}) \propto \exp\left(-\frac{\left(I(x_i,y_i,t) - I(x - \mathbf{v}_x\Delta t, y - \mathbf{v}_y\Delta t, t - \Delta t)\right)^2}{2\sigma^2}\right)$$

$$I(x_i - \mathbf{v}_x\Delta t, y_i - \mathbf{v}_y\Delta t, t - \Delta t) \approx I(x_i,y_i,t) - I_x\mathbf{v}_x\Delta t - I_y\mathbf{v}_y\Delta t - I_t\Delta t$$

$$\begin{aligned} I(x_i,y_i,t) - I(x_i - \mathbf{v}_x\Delta t, y_i - \mathbf{v}_y\Delta t, t - \Delta t) &= I_x\mathbf{v}_x\Delta t + I_y\mathbf{v}_y\Delta t + I_t\Delta t \\ &= (I_x\mathbf{v}_x + I_y\mathbf{v}_y + I_t)\Delta t \end{aligned}$$

First-order Taylor series expansion

$$P(I(x_i,y_i,t)\,|\,\mathbf{v}) \propto \exp\left(-\frac{(I_x\mathbf{v}_x + I_y\mathbf{v}_y + I_t)^2}{2\sigma^2}\right)$$

Likelihood

$$P\left(I\left(x,y,t\right)\mid \mathbf{v}\right)=\prod_i P\left(I\left(x_i,y_i,t\right)\mid \mathbf{v}\right)$$

$$\propto \prod_i \exp\left(-\frac{1}{2\sigma^2}\left(I_x\mathbf{v}_x+I_y\mathbf{v}_y+I_t\right)^2\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\int_{x,y} w\left(x,y\right)\left(I_x\mathbf{v}_x+I_y\mathbf{v}_y+I_t\right)^2 dxdy\right)$$

Gaussian window
(local patches or mixture of local patches)

Posterior

$$\log\left(P\left(\mathbf{v} \mid I\left(x,y,t\right)\right)\right) = \log\left(P\left(I\left(x,y,t\right) \mid \mathbf{v}\right)\right) + \log\left(P\left(\mathbf{v}\right)\right)$$

$$\propto \left(\sum_{x,y} -\frac{1}{2\sigma^2}\left(I_x\mathbf{v}_x + I_y\mathbf{v}_y + I_t\right)^2\right) - \frac{1}{2\sigma_p^2}\left(\mathbf{v}_x^2 + \mathbf{v}_y^2\right)$$

Optimal Percepts

- Perceived motion corresponds to the MAP estimate

Posterior is Gaussian → MAP is mean

$$\mathbf{v}^* = \arg\max_{\mathbf{v}} P(\mathbf{v} \mid I), P(\mathbf{v} \mid I) = P(\mathbf{v}) \prod_i P\left(I\left(x_i, y_i\right) \mid \mathbf{v}\right)$$

$$\begin{pmatrix} \sum I_x^2 + \frac{\sigma^2}{\sigma_p^2} & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{pmatrix} \mathbf{v}^* = -\begin{pmatrix} \sum I_x I_t \\ \sum I_y I_t \end{pmatrix}$$

Only one free parameter

Simple Theory is simply Bayesian Linear Regression

- Assume observations from a deterministic function with added Gaussian noise:

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon \qquad \text{where} \qquad p(\epsilon|\beta) = \mathcal{N}(\epsilon|0, \beta^{-1})$$

- which is the same as saying,

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1}).$$

- Given observed inputs, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, and targets, $\mathbf{t} = [t_1, \ldots, t_N]^{\mathrm{T}}$, we obtain the likelihood function

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1}).$$

- Taking the logarithm, we get

$$\begin{aligned}\ln p(\mathbf{t}|\mathbf{w},\beta) &= \sum_{n=1}^{N} \ln \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n),\beta^{-1}) \\ &= \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) - \beta E_D(\mathbf{w})\end{aligned}$$

- where

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2$$

- is the sum-of-squares error.

General Case

- Computing the gradient and setting it to zero yields

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}, \beta) = \beta \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} = \mathbf{0}.$$

- Solving for w, we get

$$\mathbf{w}_{\mathrm{ML}} = \overbrace{\left(\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^{\mathrm{T}}}^{\text{The Moore-Penrose pseudo-inverse, } \boldsymbol{\Phi}^{\dagger}.} \mathbf{t}$$

- where

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}.$$

Bayesian Linear Regression

Define a conjugate prior over w

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0).$$

Combining this with the likelihood function and using results for marginal and conditional Gaussian distributions, gives the posterior

$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$$

where

$$\begin{aligned} \mathbf{m}_N &= \mathbf{S}_N \left(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta \mathbf{\Phi}^{\mathrm{T}}\mathbf{t} \right) \\ \mathbf{S}_N^{-1} &= \mathbf{S}_0^{-1} + \beta \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}. \end{aligned}$$

Bayesian Linear Regression (2)

- A common choice for the prior is

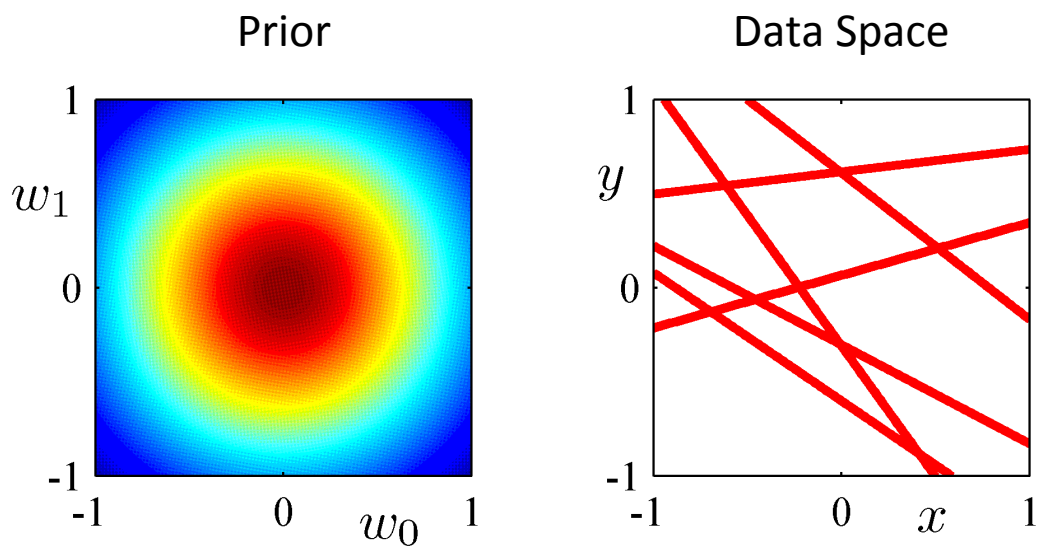
$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

- for which

$$\begin{aligned} \mathbf{m}_N &= \beta \mathbf{S}_N \mathbf{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_N^{-1} &= \alpha \mathbf{I} + \beta \mathbf{\Phi}^{\mathrm{T}} \mathbf{\Phi}. \end{aligned}$$

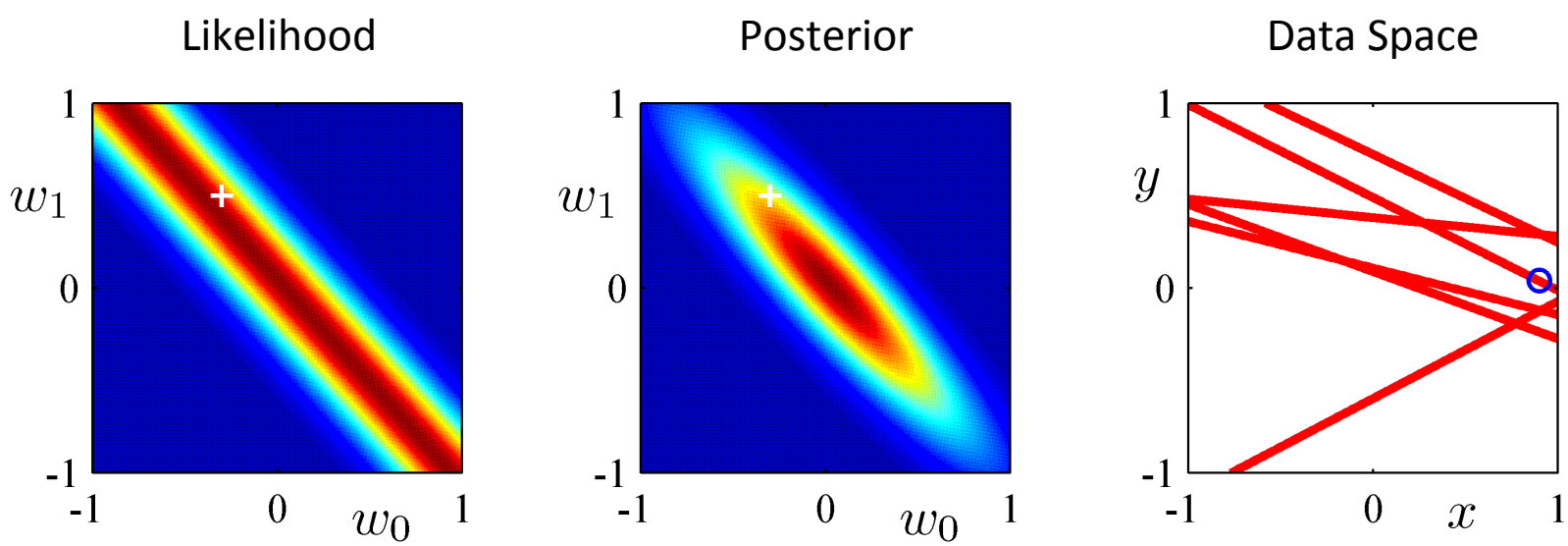
Bayesian Linear Regression (3)

0 data points observed



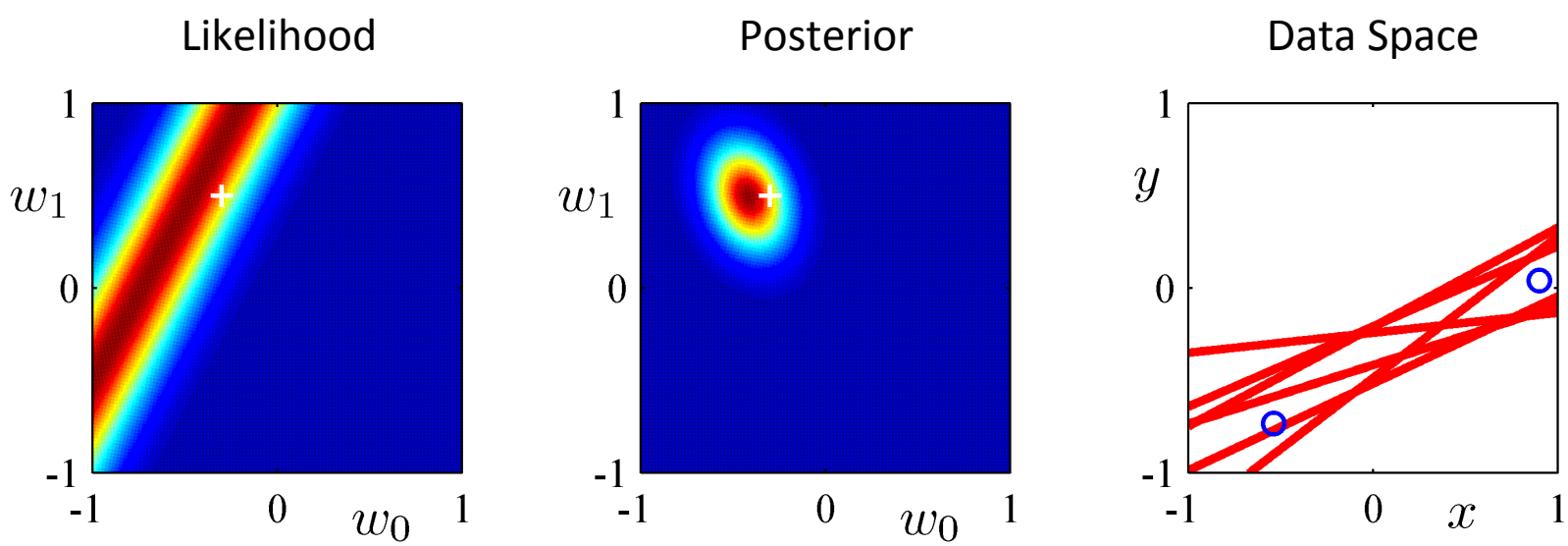
Bayesian Linear Regression

1 data point observed



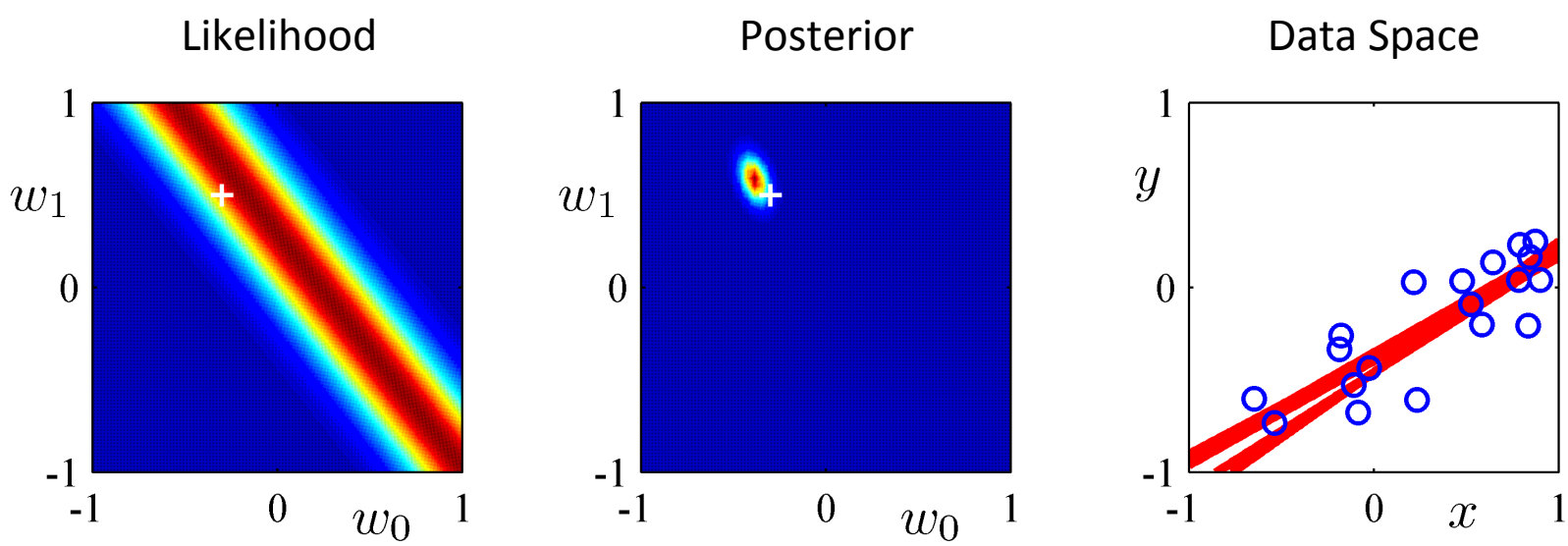
Bayesian Linear Regression

2 data points observed



Bayesian Linear Regression (6)

20 data points observed



For motion theory, points are regions with non-zero image derivatives

Likelihood

$$-L(\mathbf{v}) = \frac{1}{2\sigma^2}\sum_{x,y}\left(I_x\mathbf{v}_x + I_y\mathbf{v}_y + I_t\right)^2 + \frac{1}{2\sigma_p^2}\sum_{x,y}\left(\mathbf{v}_x^2 + \mathbf{v}_y^2\right)$$

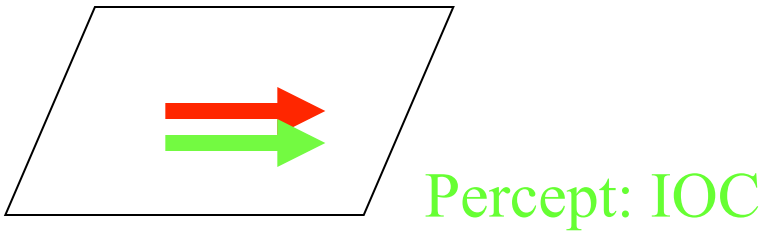
$$\frac{\partial L(\mathbf{v})}{\partial \mathbf{v}} = \frac{1}{2\sigma^2}\begin{pmatrix} 2\sum_{x,y}\left(I_x\mathbf{v}_x + I_y\mathbf{v}_y + I_t\right)I_x \\ 2\sum_{x,y}\left(I_x\mathbf{v}_x + I_y\mathbf{v}_y + I_t\right)I_y \end{pmatrix} + \frac{1}{2\sigma_p^2}\begin{pmatrix} 2\sum_{x,y}\mathbf{v}_x \\ 2\sum_{x,y}\mathbf{v}_y \end{pmatrix}$$

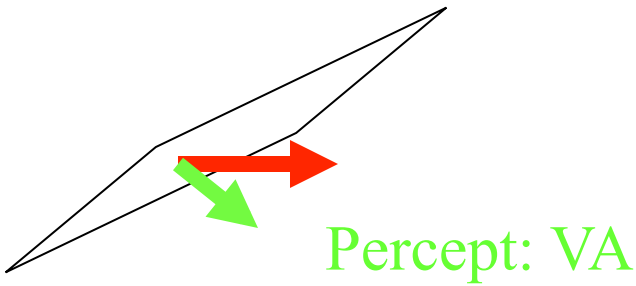
$$= -\frac{1}{\sigma^2}\left(\begin{pmatrix} \sum_{x,y} I_x^2 + \frac{\sigma^2}{\sigma_p^2} & \sum_{x,y} I_y I_x \\ \sum_{x,y} I_y I_x & \sum_{x,y} I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{pmatrix}\mathbf{v} + \begin{pmatrix} \sum_{x,y} I_t I_x \\ \sum_{x,y} I_t I_y \end{pmatrix}\right)$$

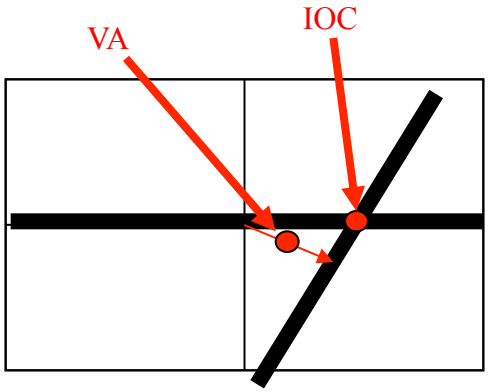
$$= 0$$

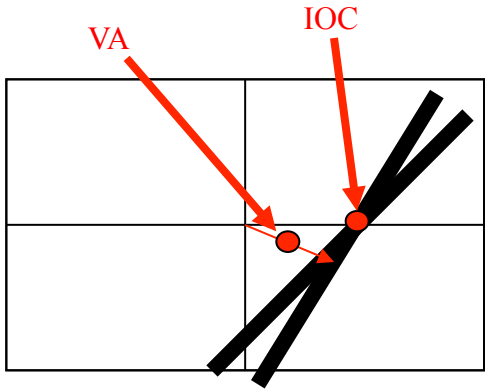
Simple idea accounted for many observations left out of previous theories

- Example: Rhombus

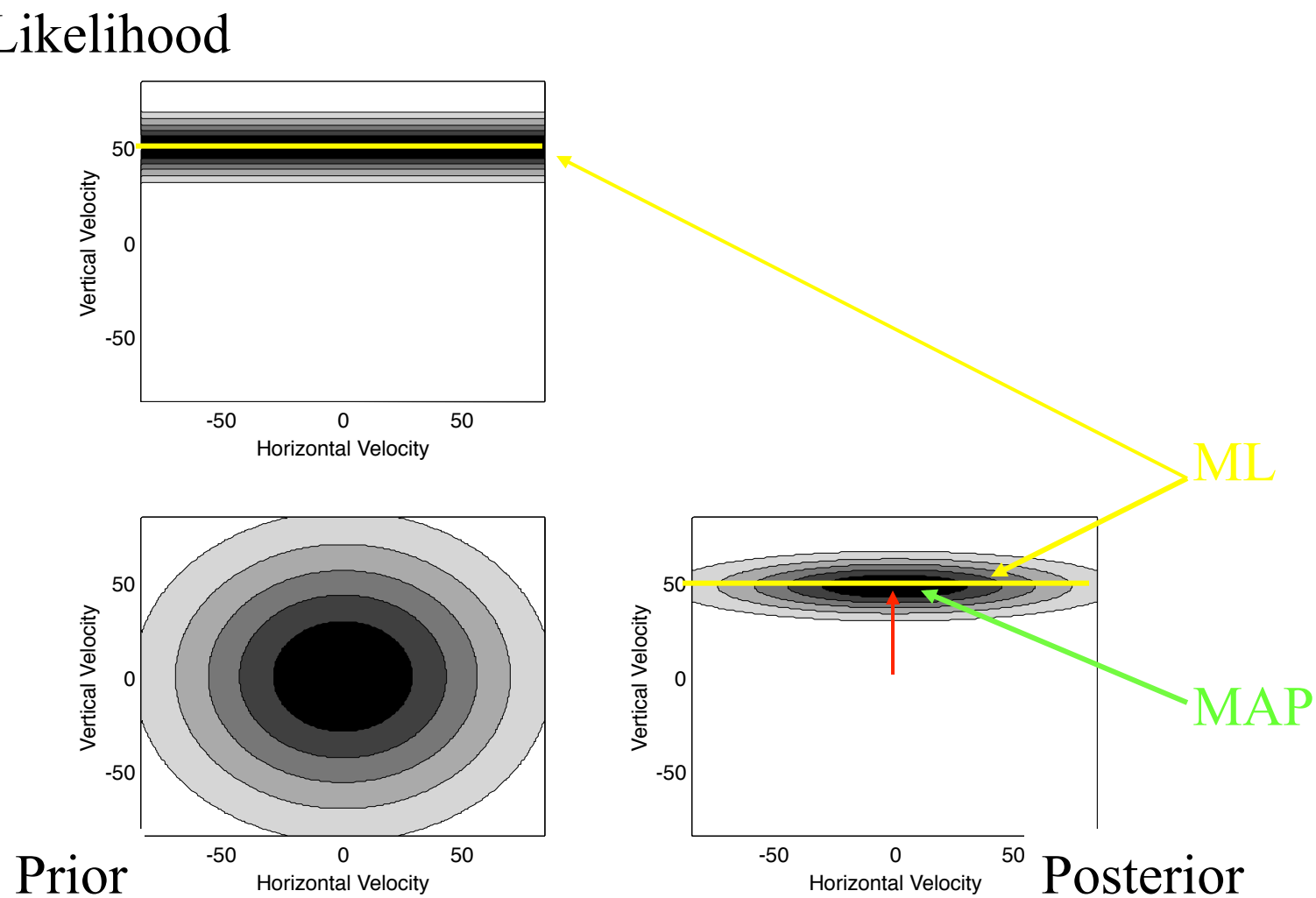








Motion Through An Aperture

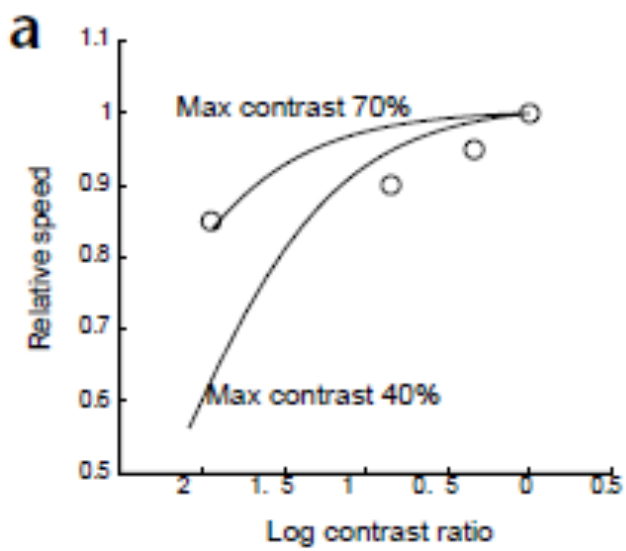


Motion And Contrast

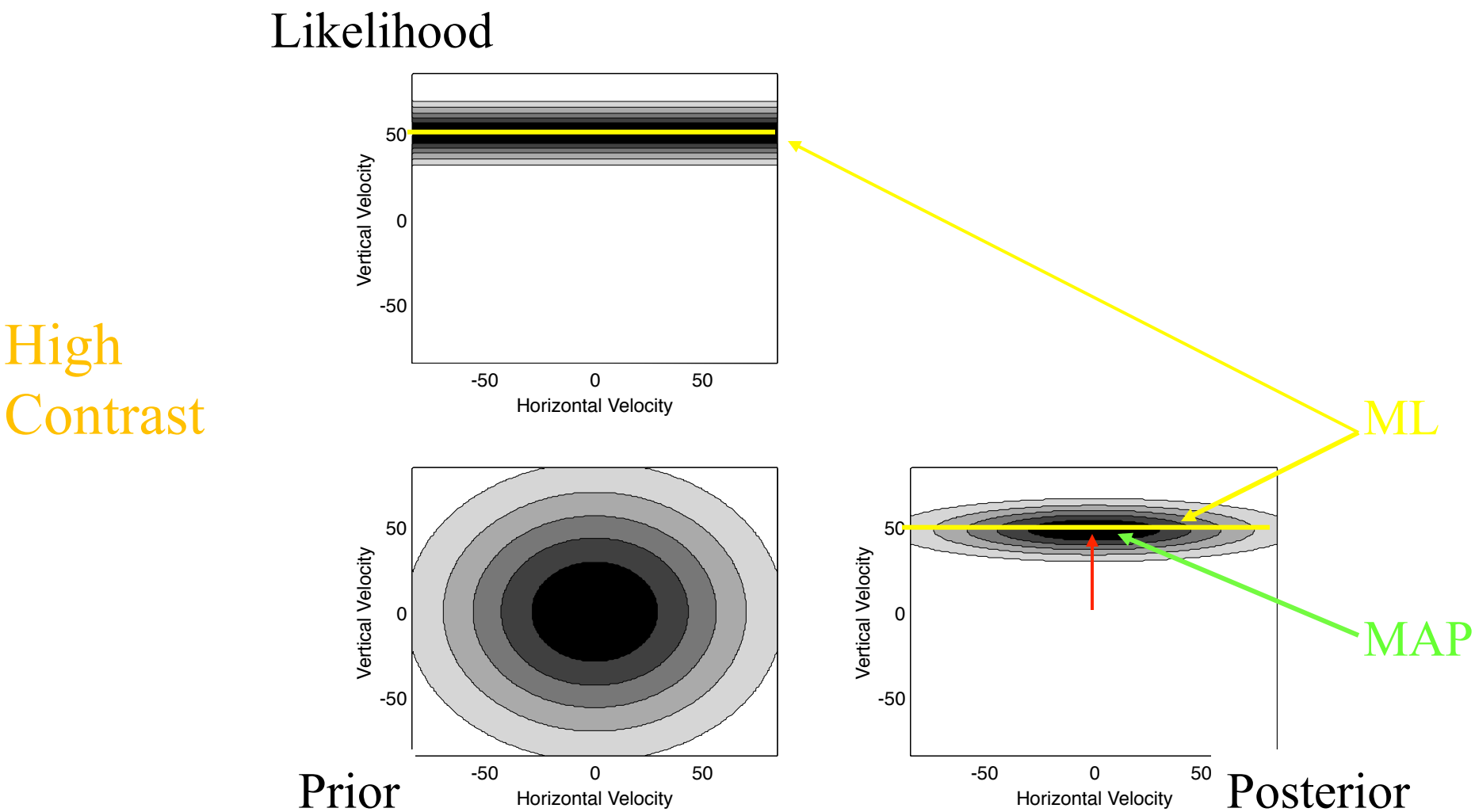
- Driving in the fog
  - Drivers tend to speed up
- In low contrast situations, the prior dominates
  - In fog, poor quality visual information about speed
  - Priors biased toward slow speeds

Motion And Constrast

- Individuals tend to underestimate velocity in low contrast situations

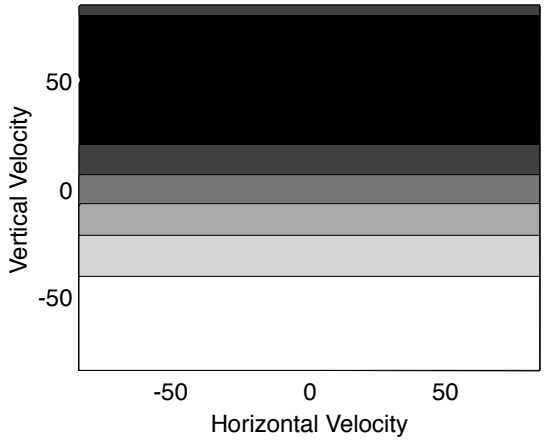


Motion And Contrast

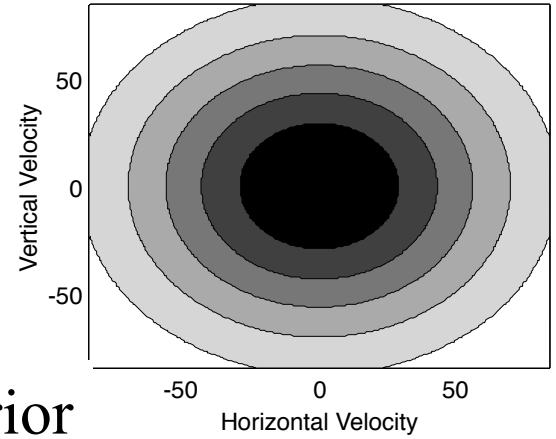


Motion And Contrast

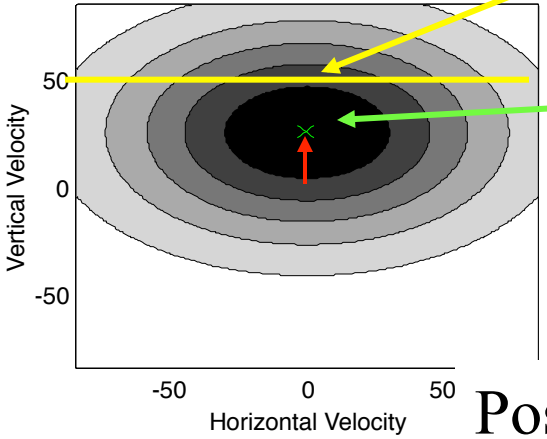
Likelihood



Low Contrast



Prior



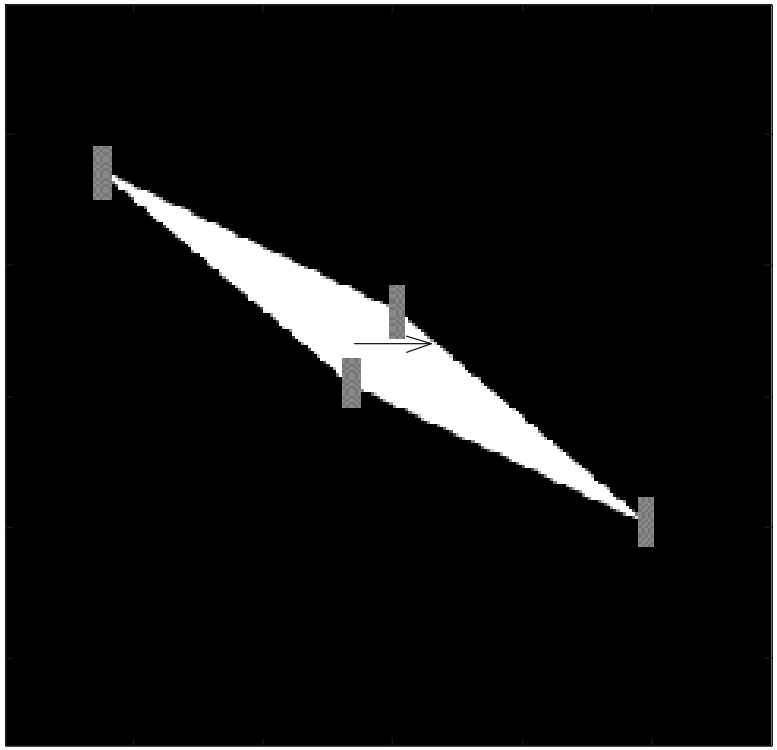
ML

MAP

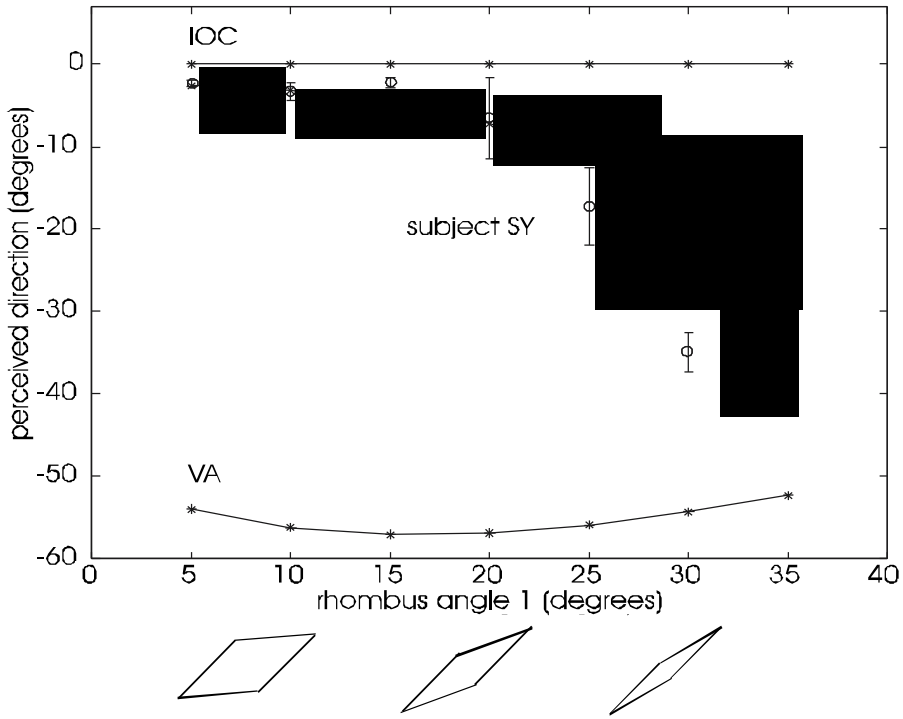
Posterior

Rhombus Demo

Moving Rhombus



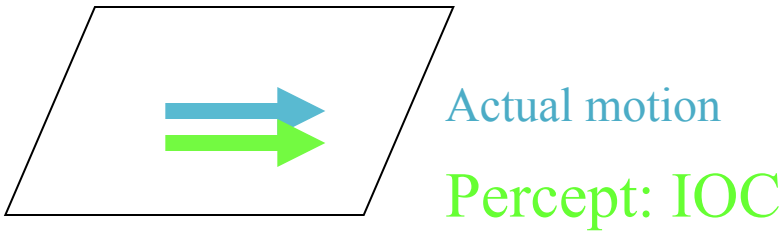
a.

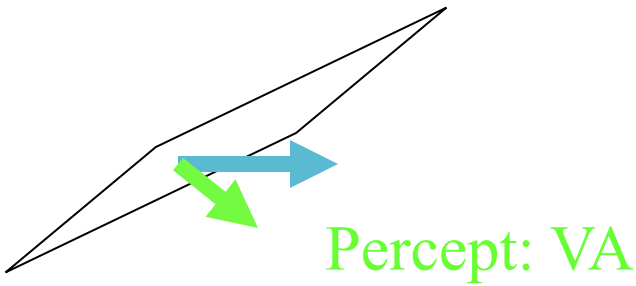


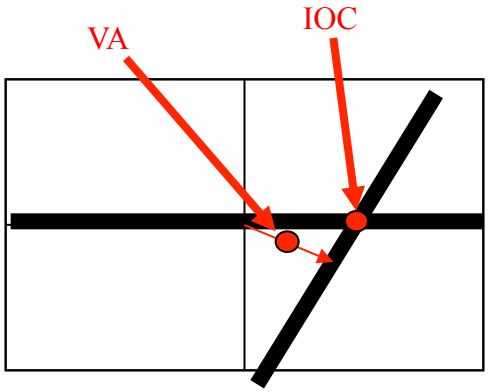
b.

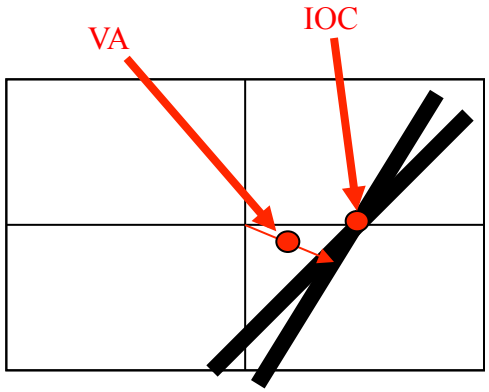
Moving Rhombus

- Example: Rhombus



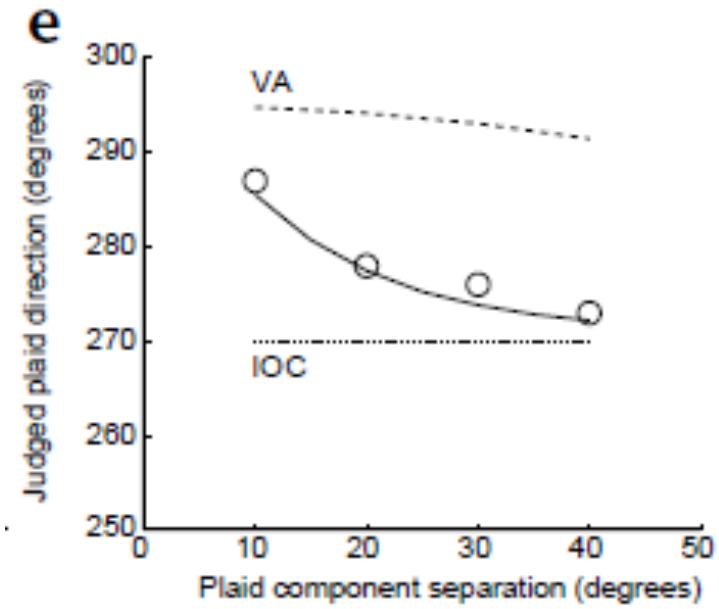




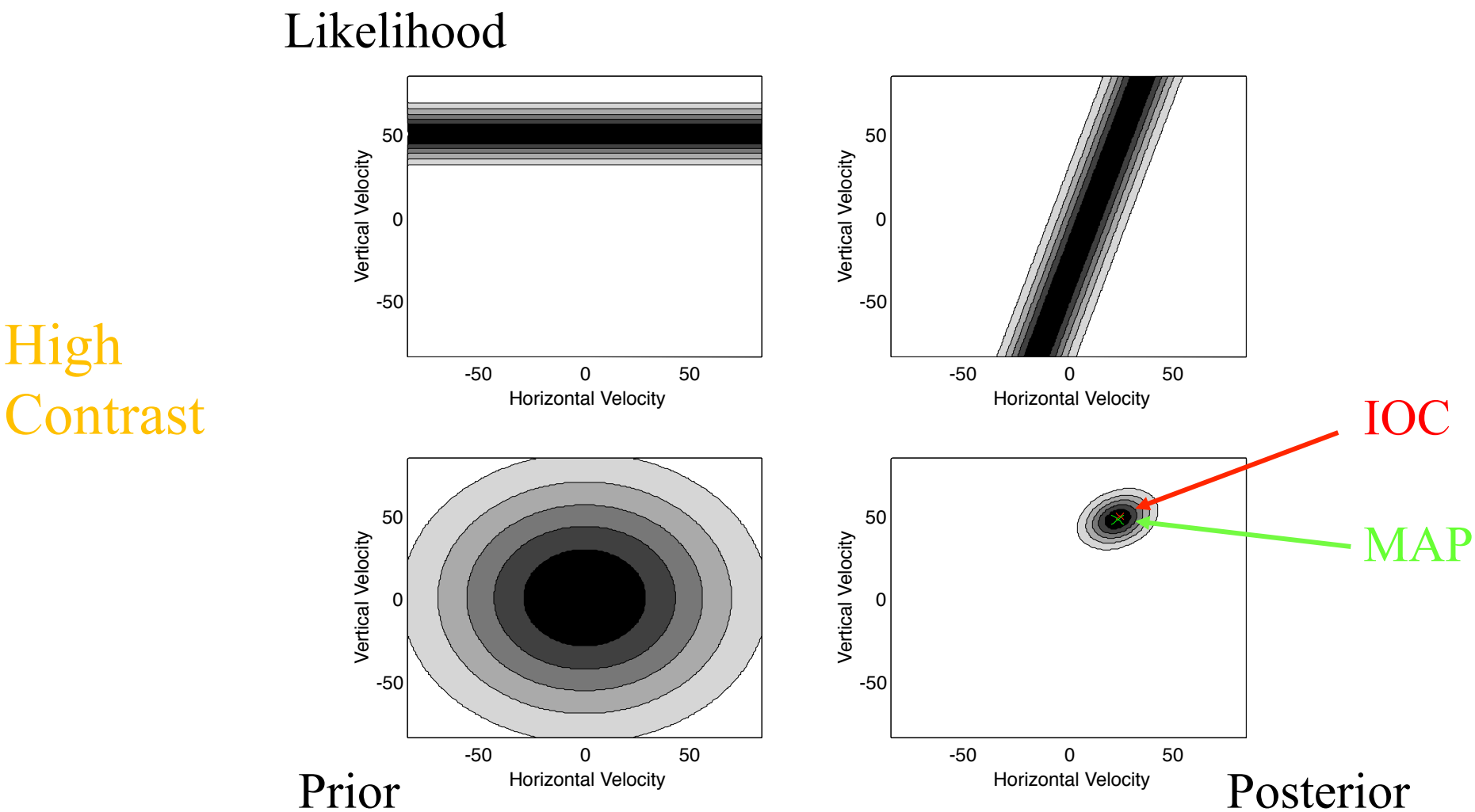


Relative Orientation of Plaids

- Type II plaids
- Same information as in rhomboid (two distinct orientations)



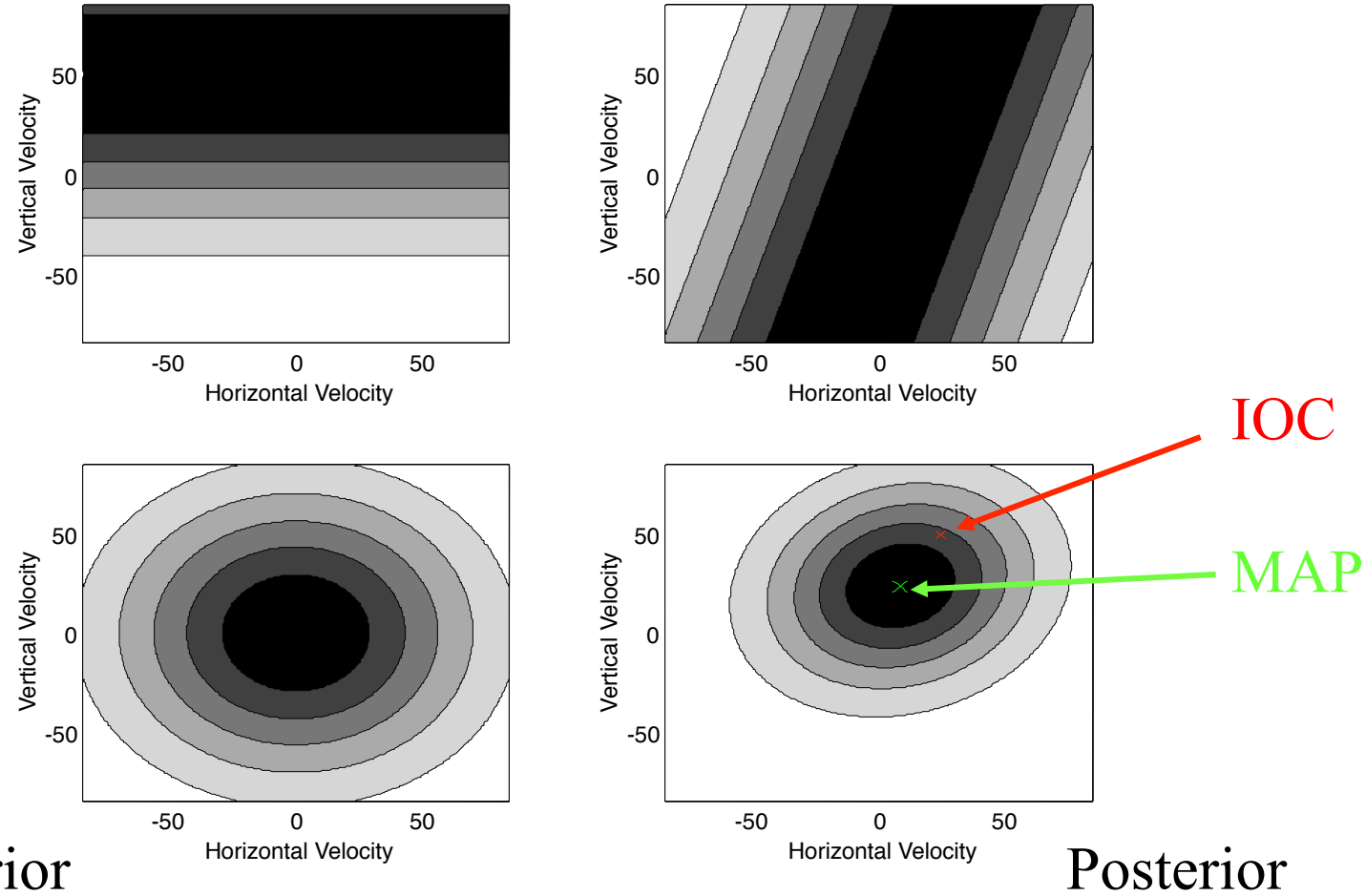
Moving Rhombus



Moving Rhombus

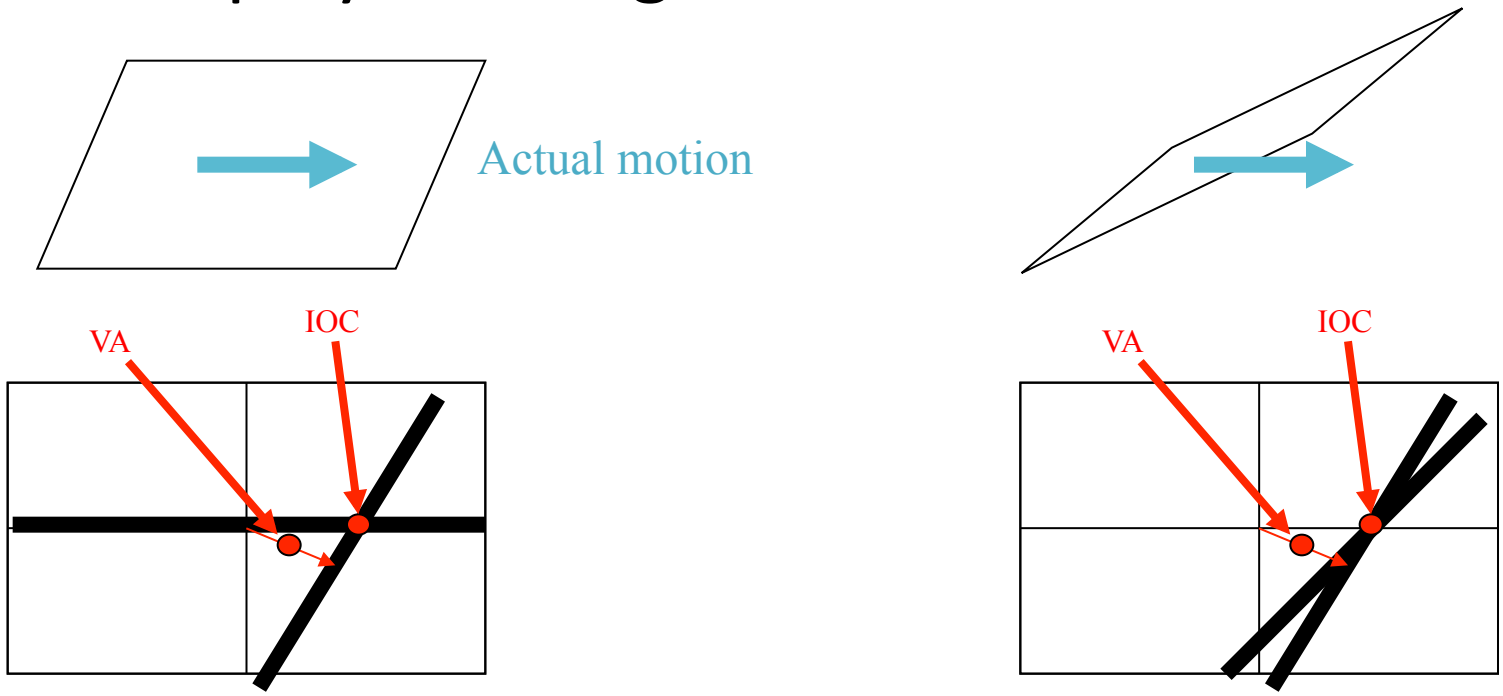
Likelihood

Low Contrast

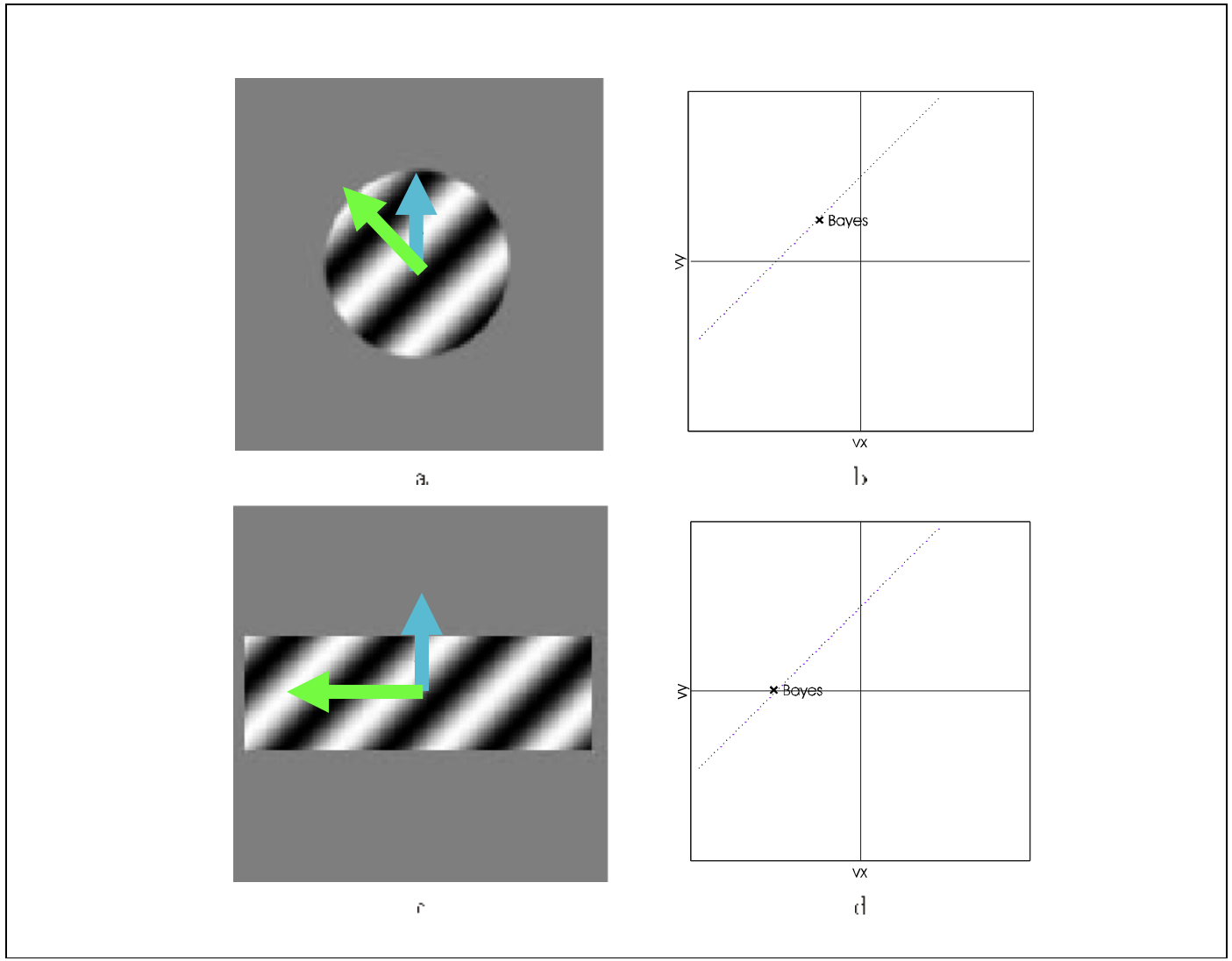


Why Does Rhombus Angle Interact With Contrast?

- Less information with acute angle
- Less information with low contrast
- Priors play a stronger role

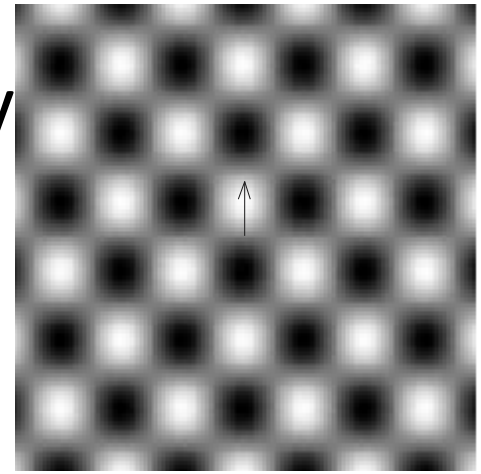


Barberpole Illusion

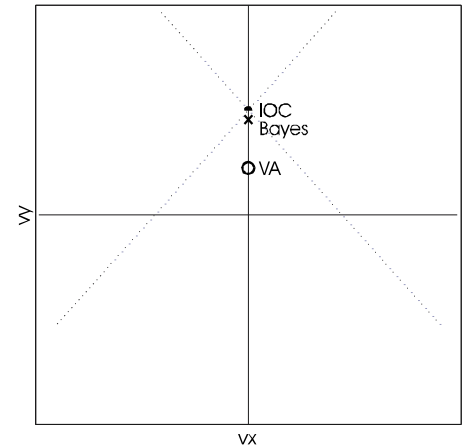


Plaid Motion: Type I and II

- Type I: true velocity lies between two normals

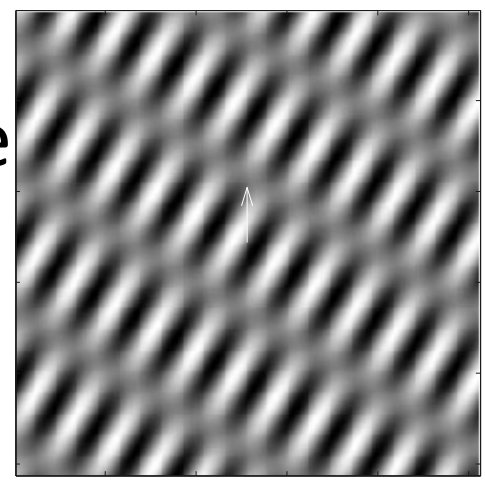


a.

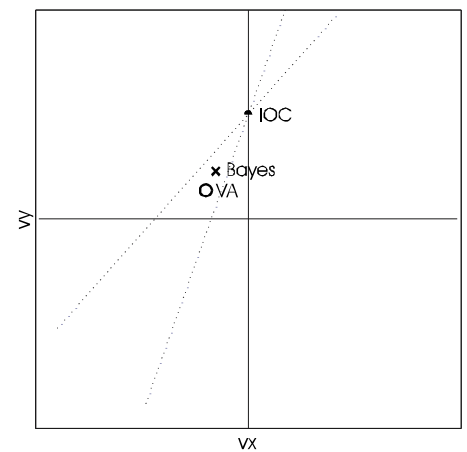


b

- Type II: true velocity lies outside two normals

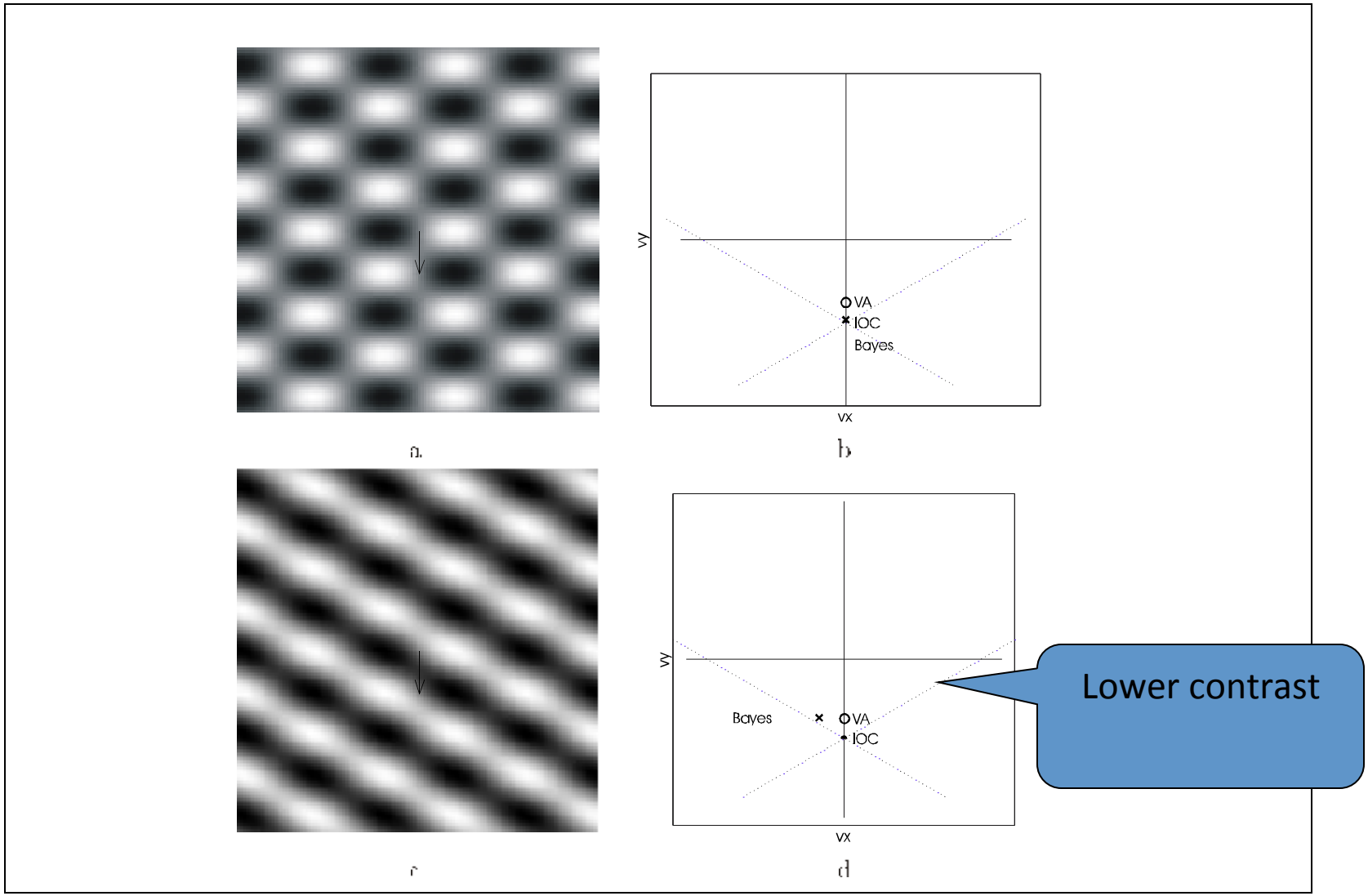


c



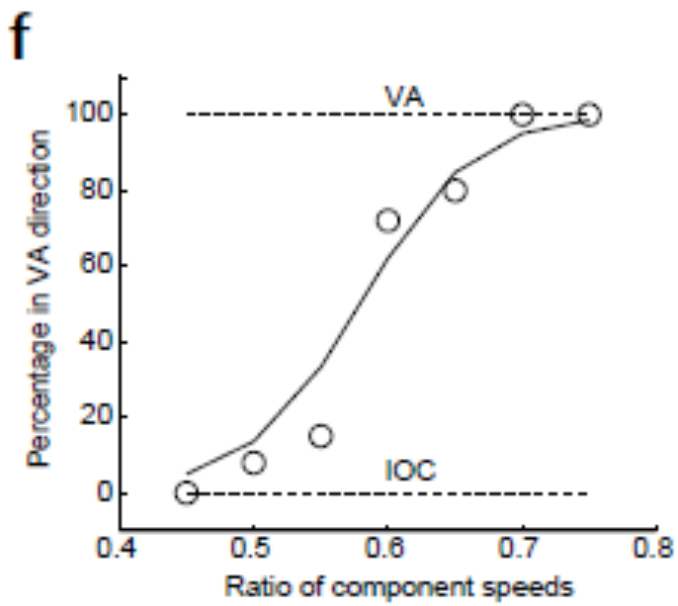
d

Plaids and Contrast



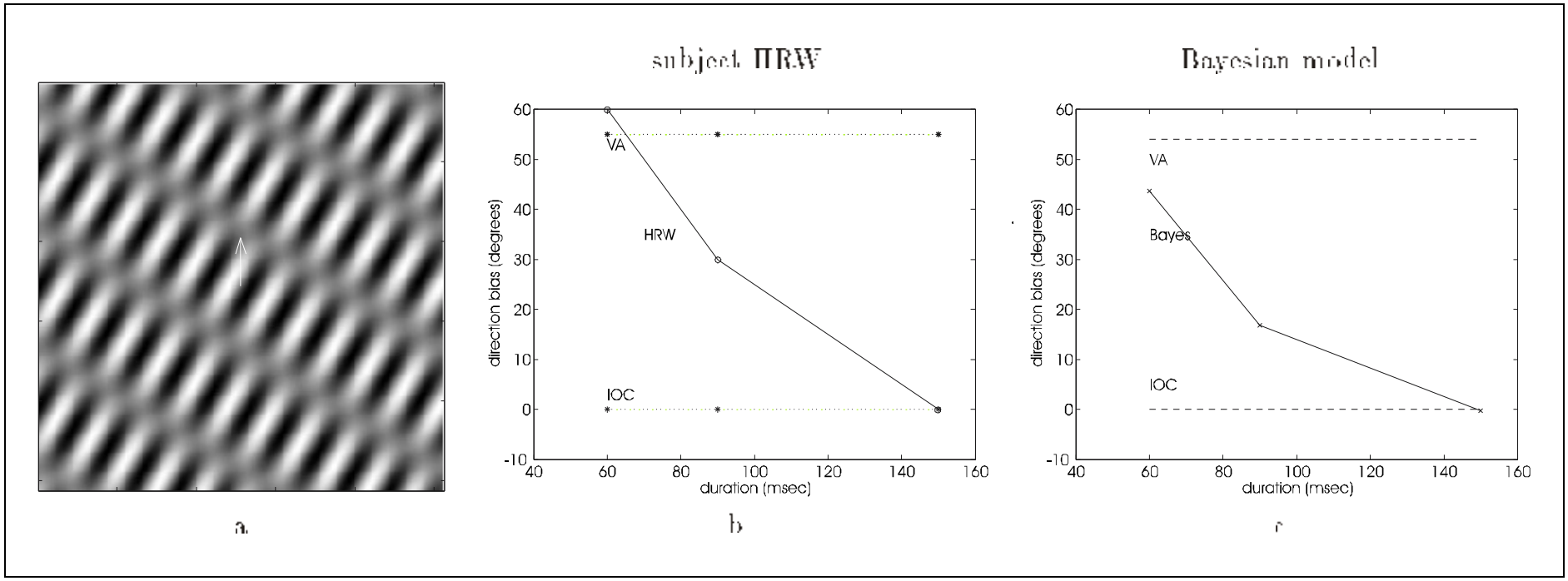
Plaids and Speed

- Perceived direction of type II plaids depends on relative speed of components



Plaids and Time

- Viewing time reduces uncertainty



Motion Illusions As Optimal Percepts

- Mistakes of perception are the result of a *rational* system designed to operate in the presence of uncertainty.
- A proper rational model incorporates actual statistics of the environment
  - Here, authors *assume* without direct evidence: (1) preference for slow speeds, (2) noisy local image measurements, (3) velocity estimate is the mean/ mode of posterior distribution
- "Optimal Bayesian estimator" or "ideal observer" is relative to these assumptions!

Modeling motion estimation

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

Pattern Motion

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

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

Spatially Local Prior

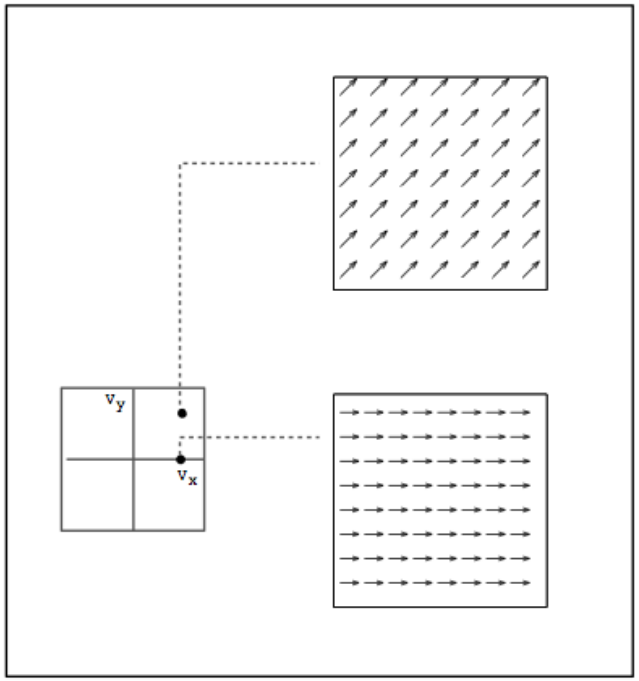
$P(\theta)$

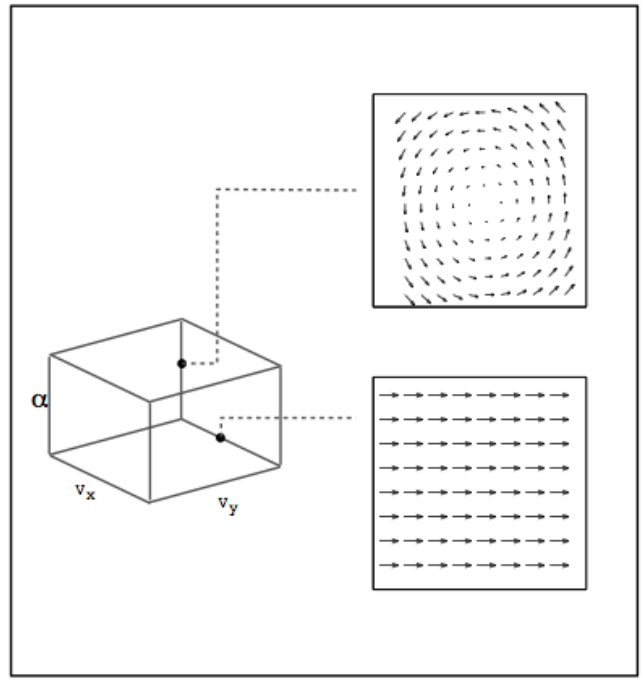
Pattern Prior

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

From: Weiss & Adelson, 1998

Hierarchical Prior for Pattern Motion



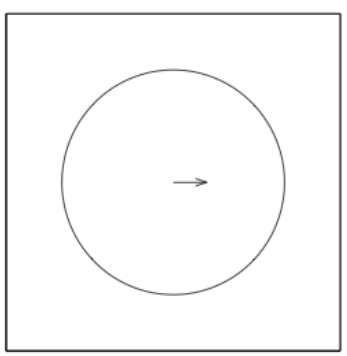


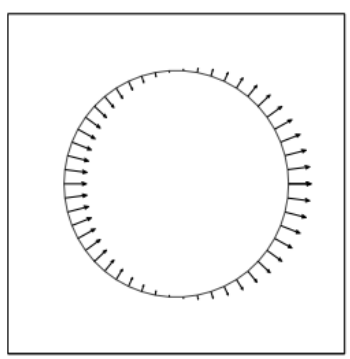
SMOOTHNESS

$$P(V) = \alpha e^{-J(V)}$$

$$J(V) = \sum_{xy} \|Dv(x,y)\|^2$$

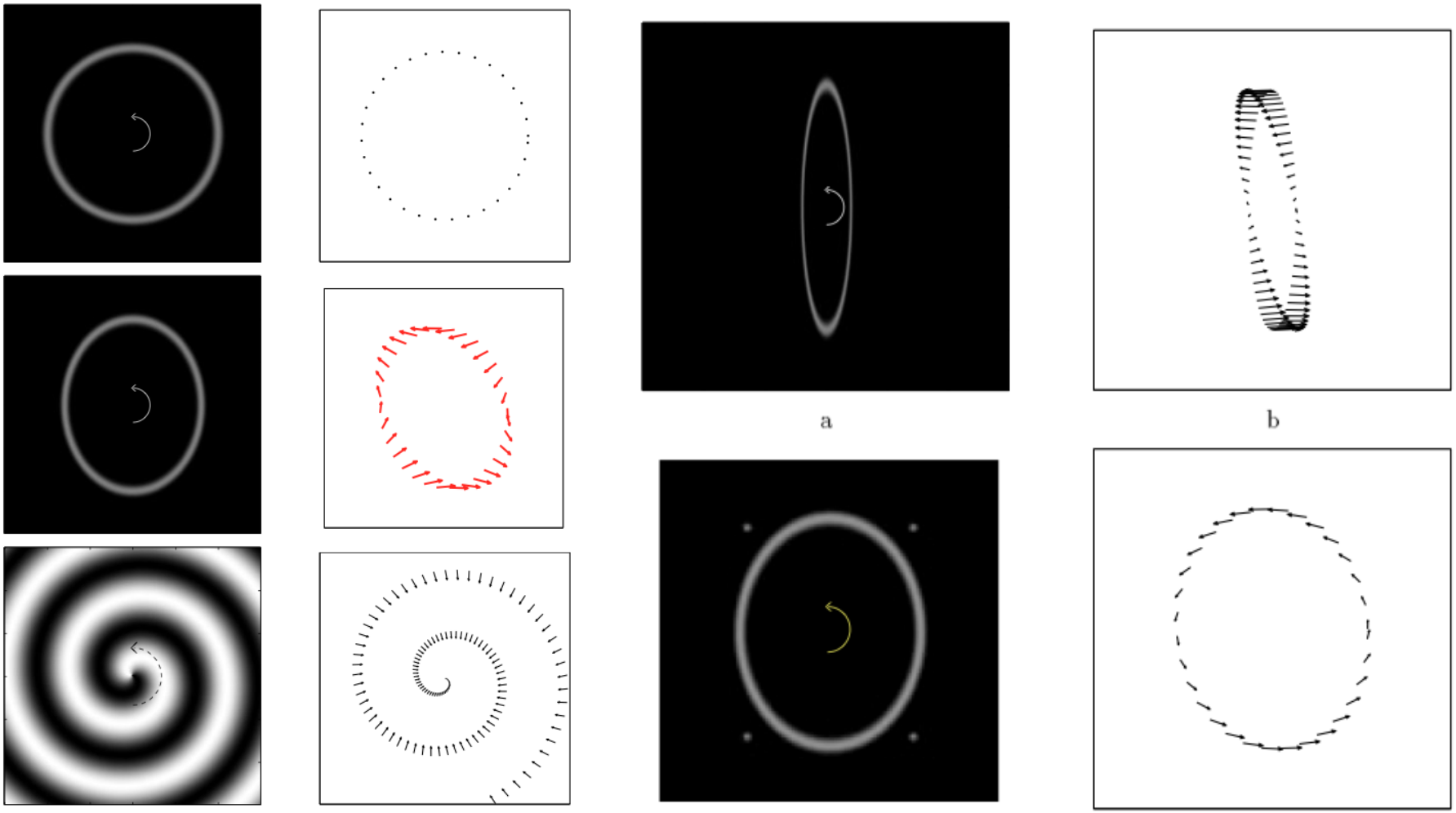
$$Dv = \sum_{n=0}^{\infty} a_n \frac{\partial^n}{\partial x} v$$



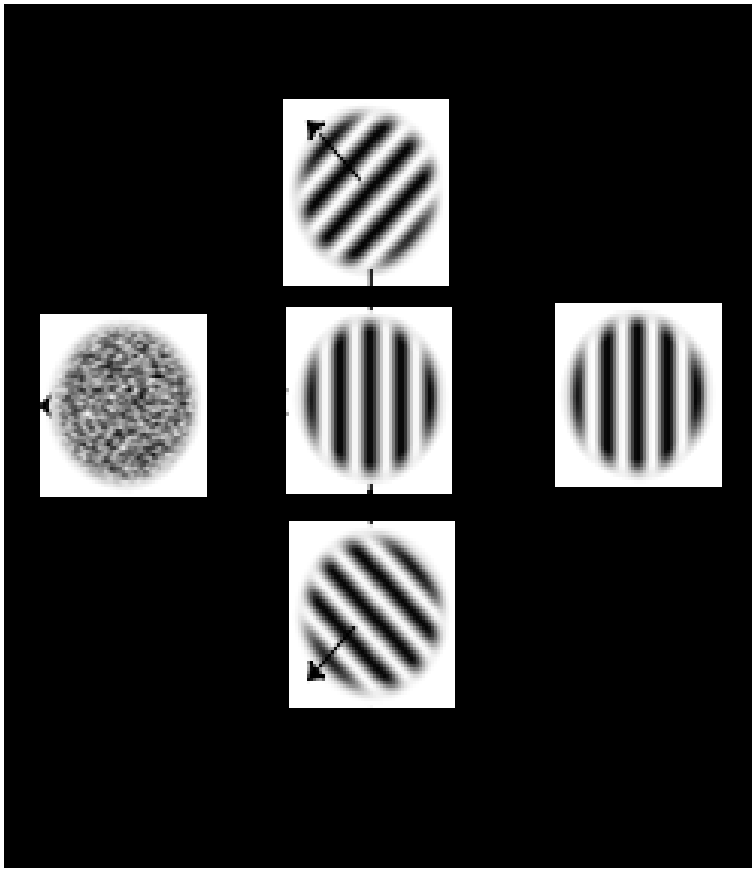


In modern ML terms, this is Gaussian Process Regression

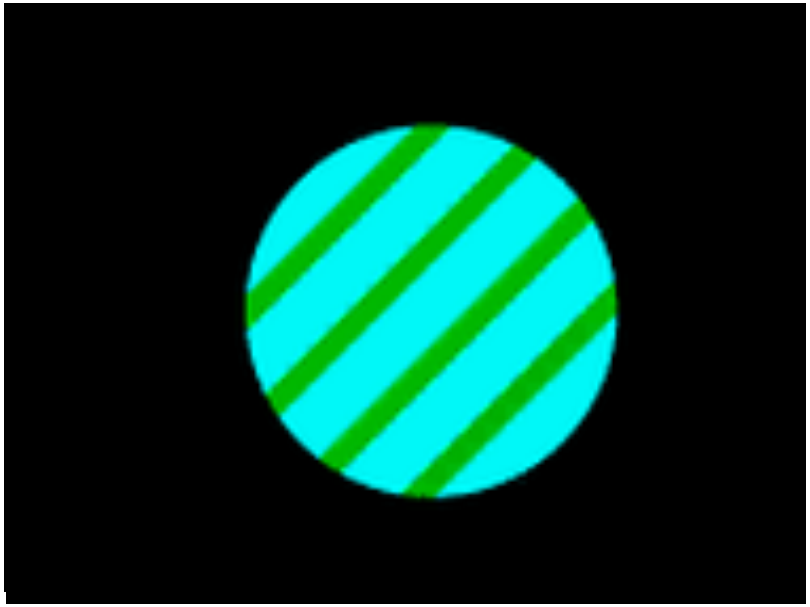
Smoothness captures many flow effects



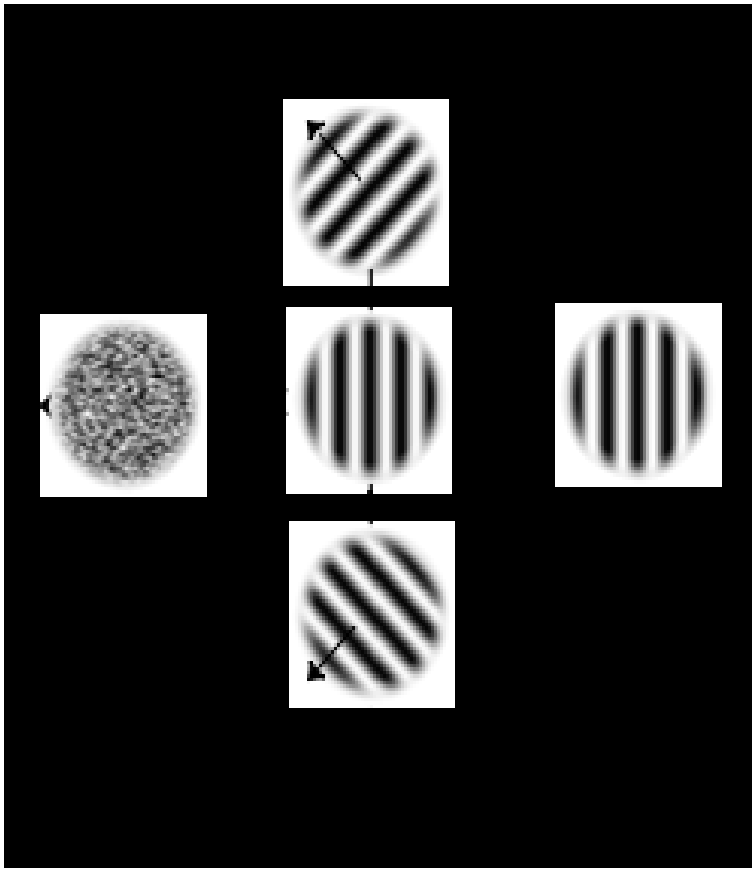
Aperture Problem (Motion/Form Ambiguity)



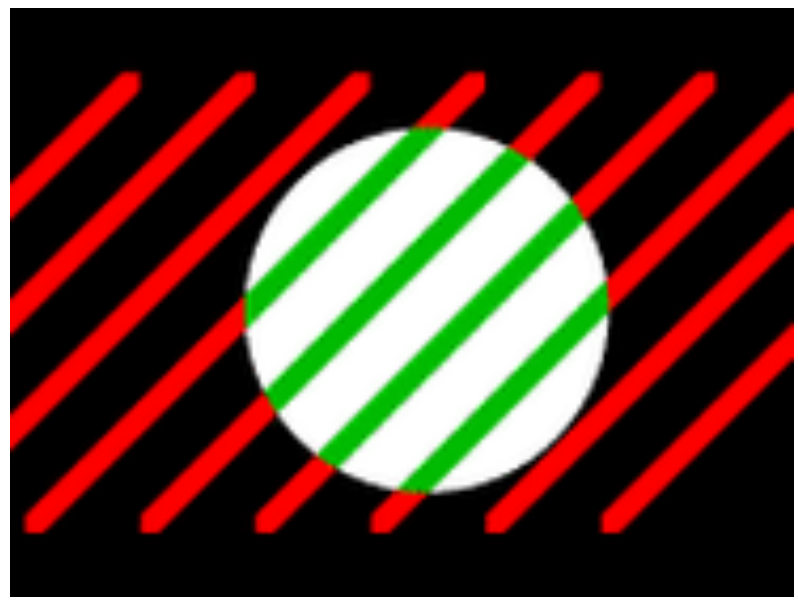
Result: Early visual measurements are ambiguous w.r.t. motion.



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 →