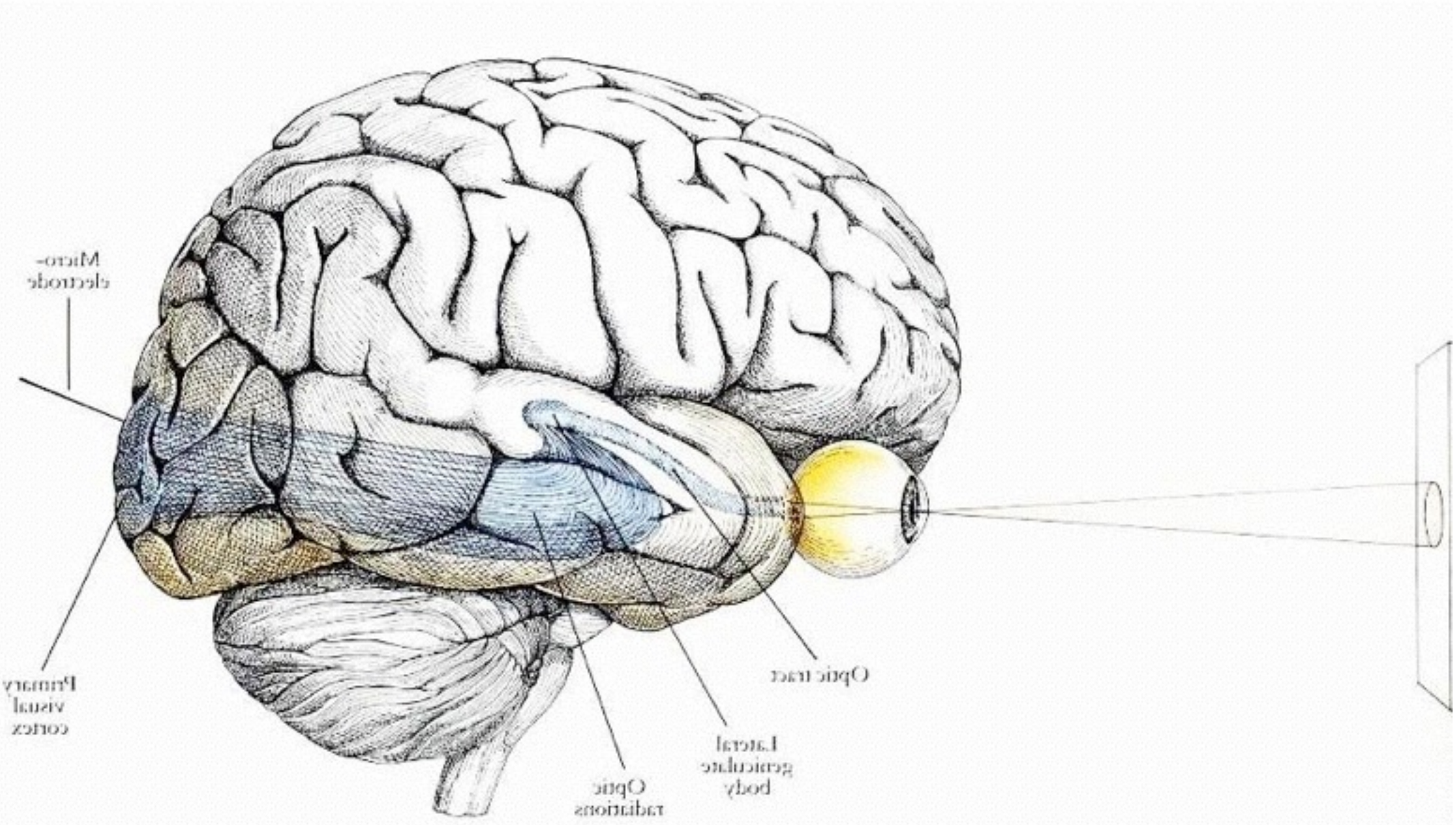
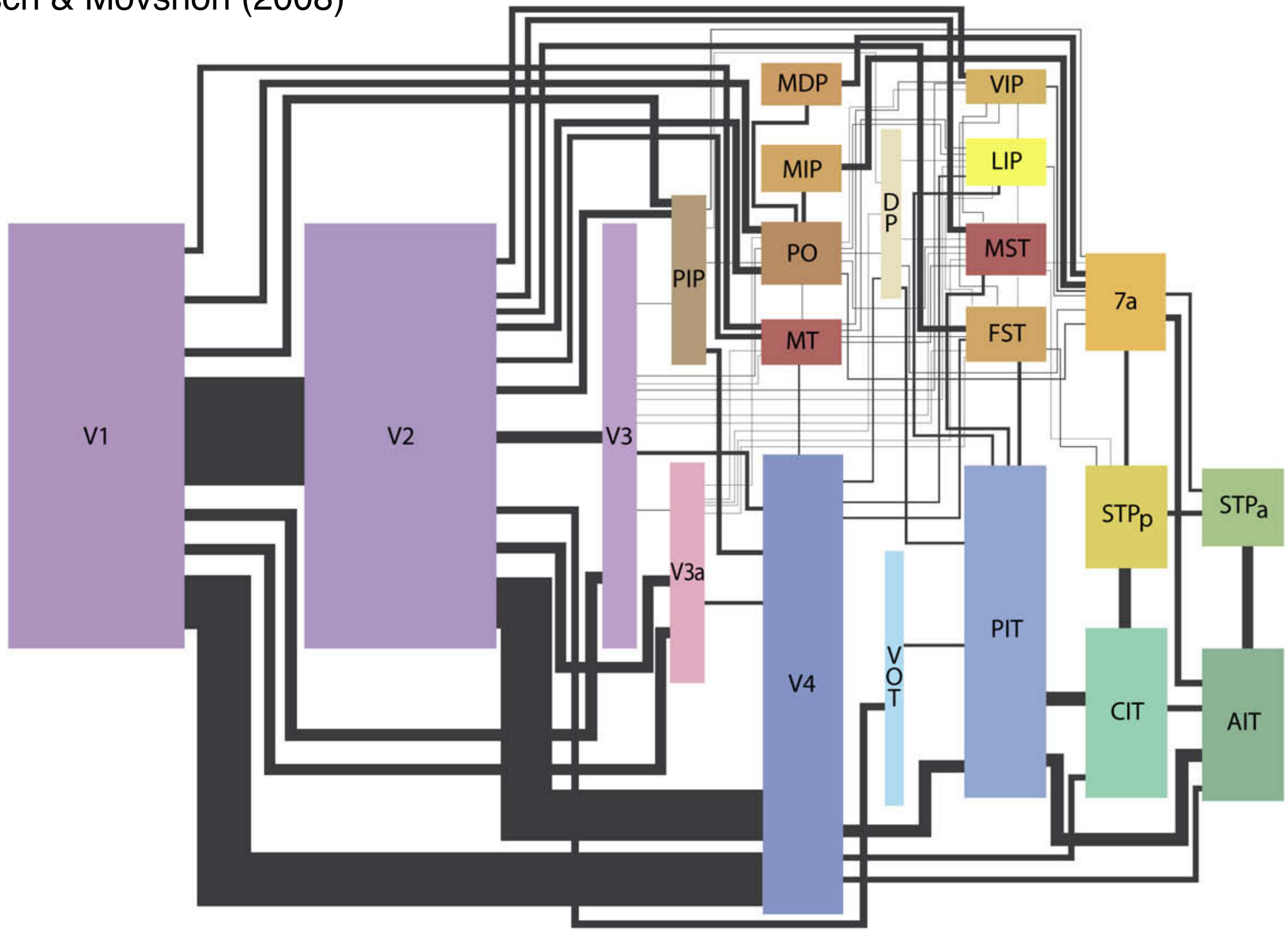
Perception as an Inference Problem

Bruno A. Olshausen
Helen Wills Neuroscience Institute, School of Optometry
and Redwood Center for Theoretical Neuroscience
UC Berkeley

What are the principles governing information processing in this system?



Wallisch & Movshon (2008)



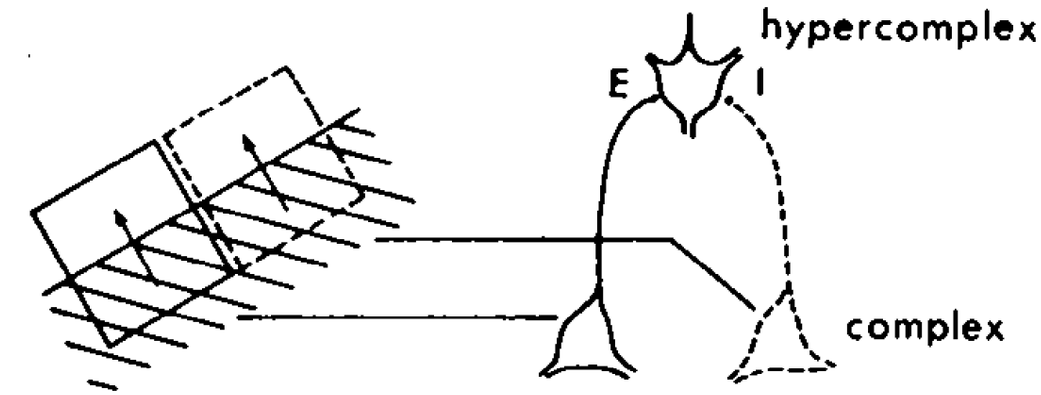
‘Gabor filters’ . . ? . . objects . faces

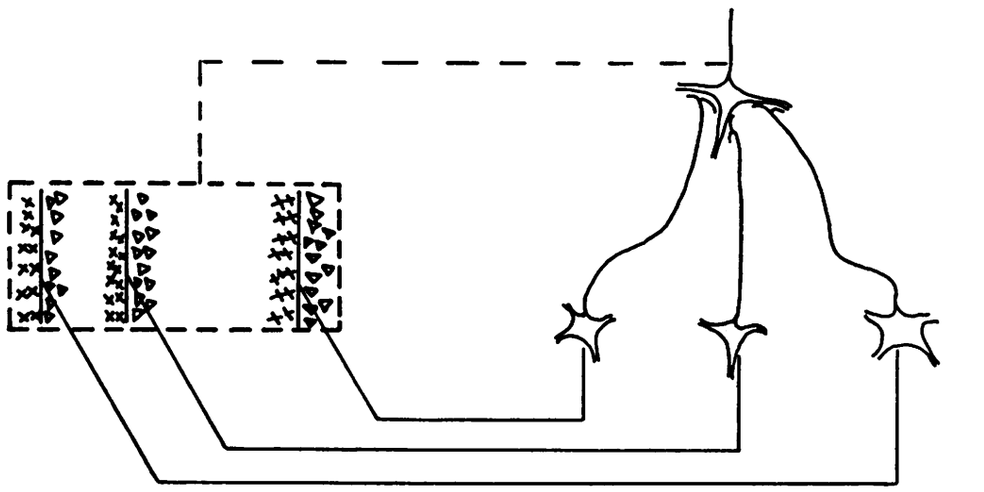
Two views of visual system function

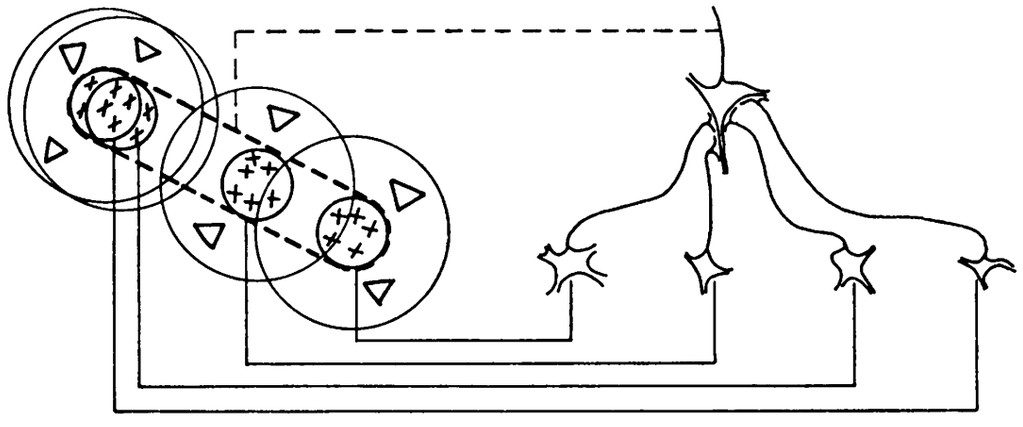
- *Deduction*
  - feature extraction, classification
  - (Hubel & Wiesel; Fukushima; 'deep learning')

- *Inference*
  - generative models, recurrent computation
  - (Helmholtz; Nakayama; Kersten & Yuille;
    Geman; Lee & Mumford)

Hubel & Wiesel (1962, 1965)



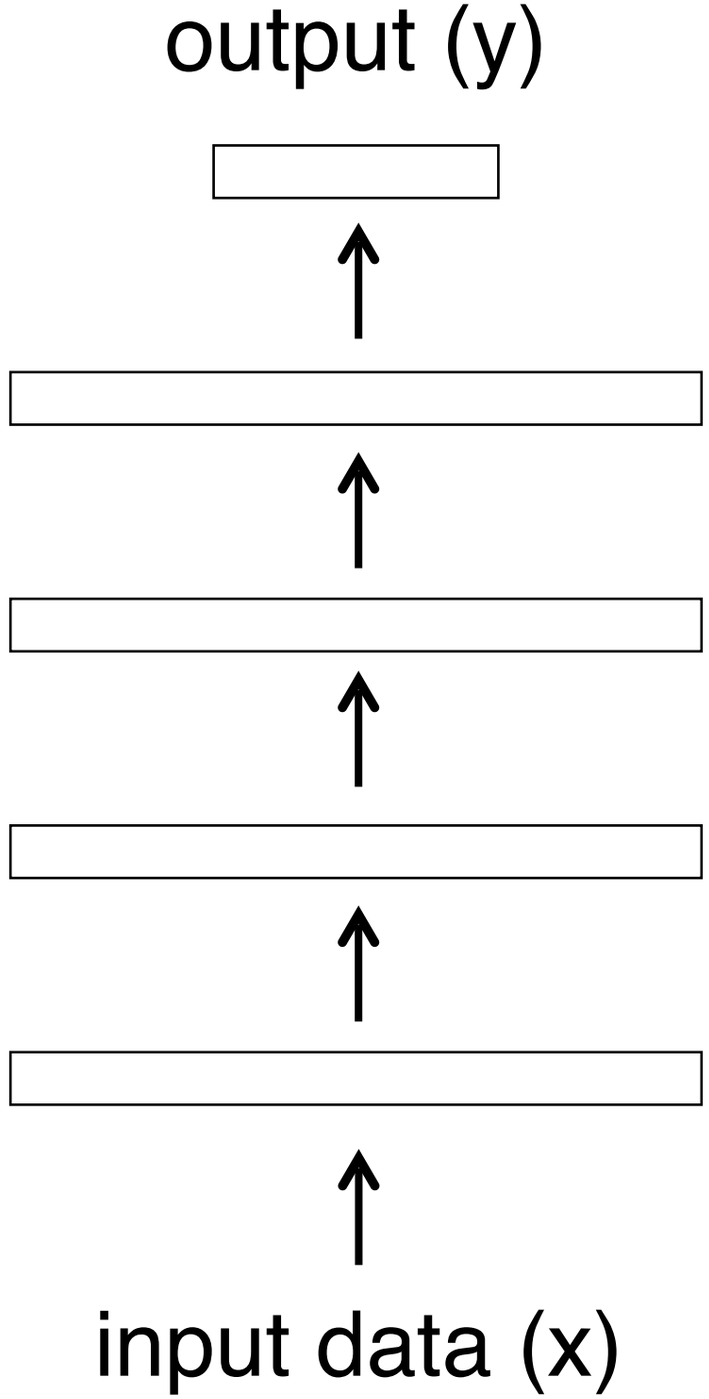




Hypercomplex

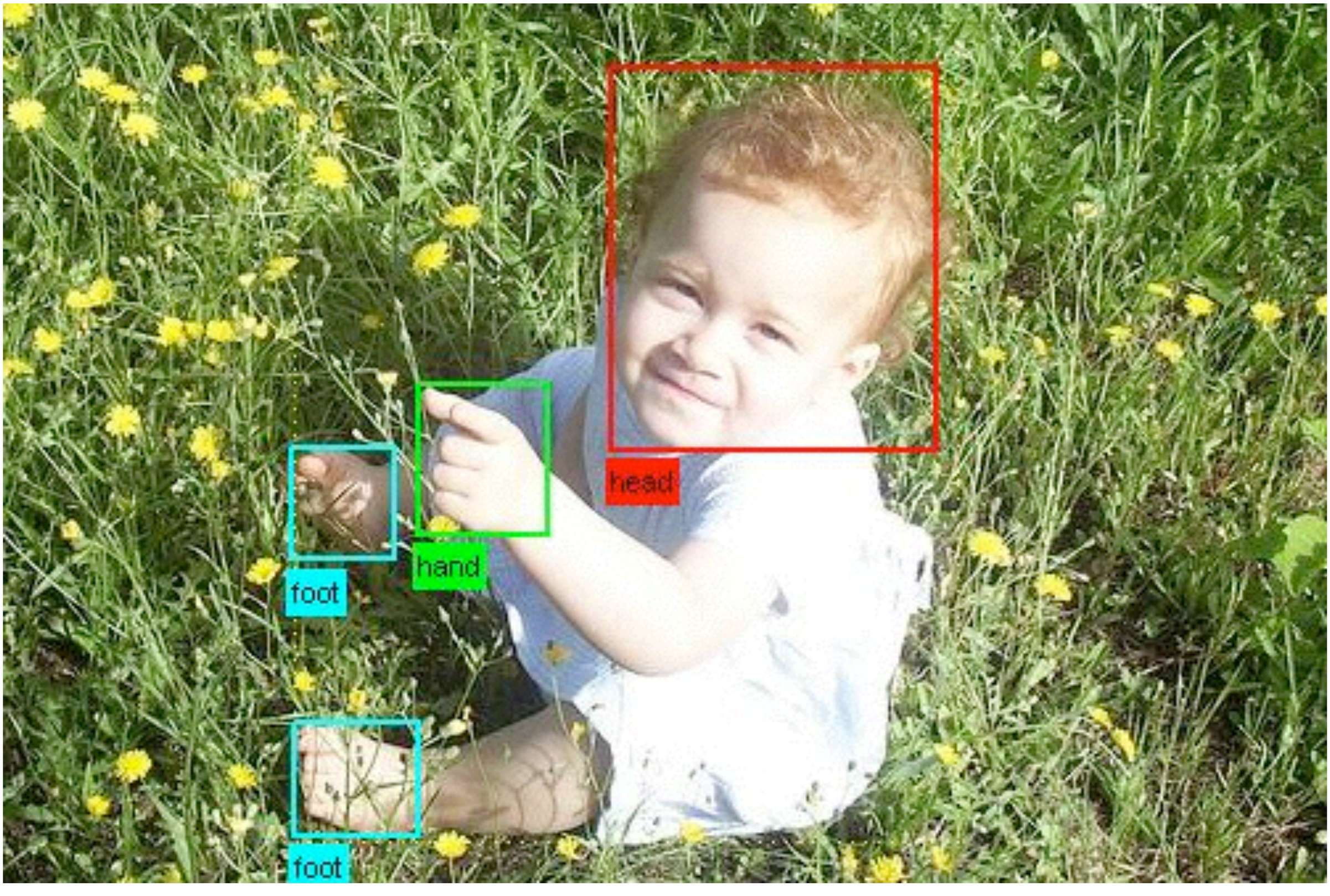
Complex

Simple



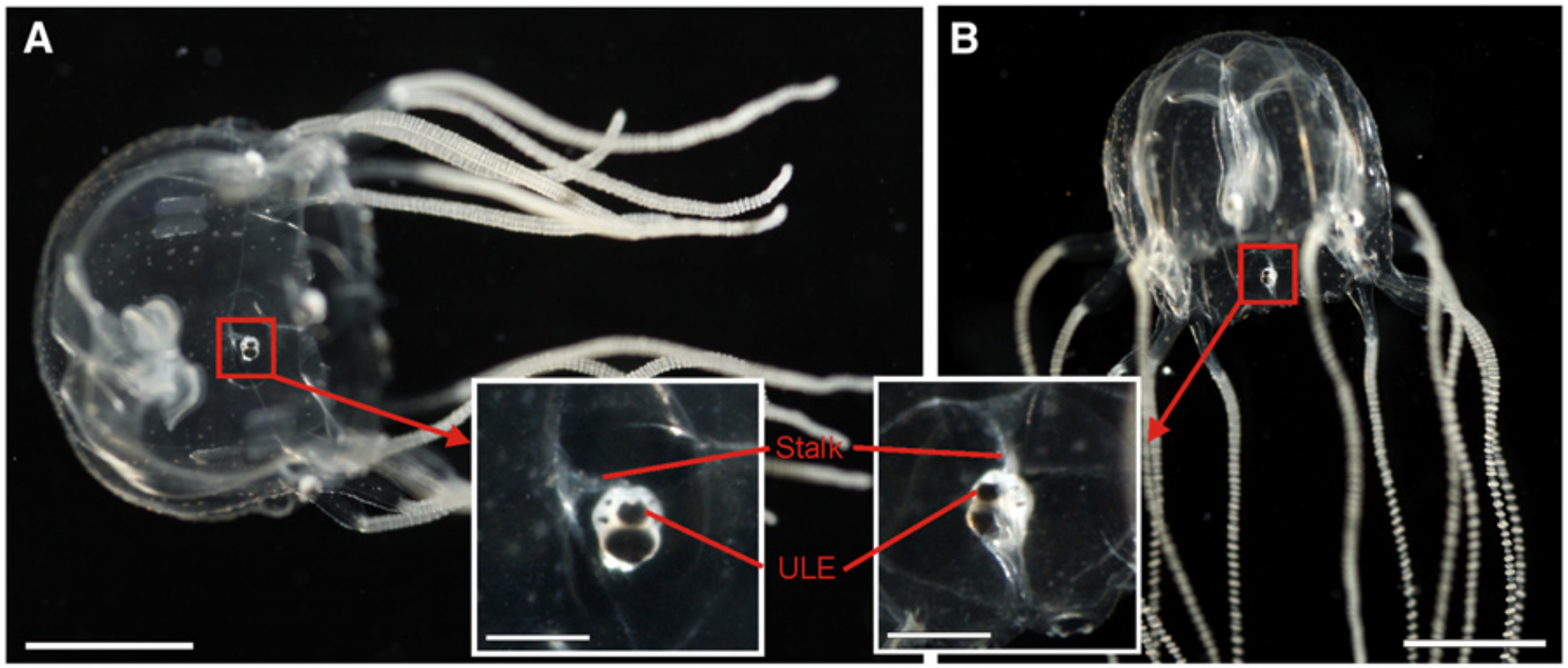
$$y = f(x; w)$$

Is this the goal of vision?

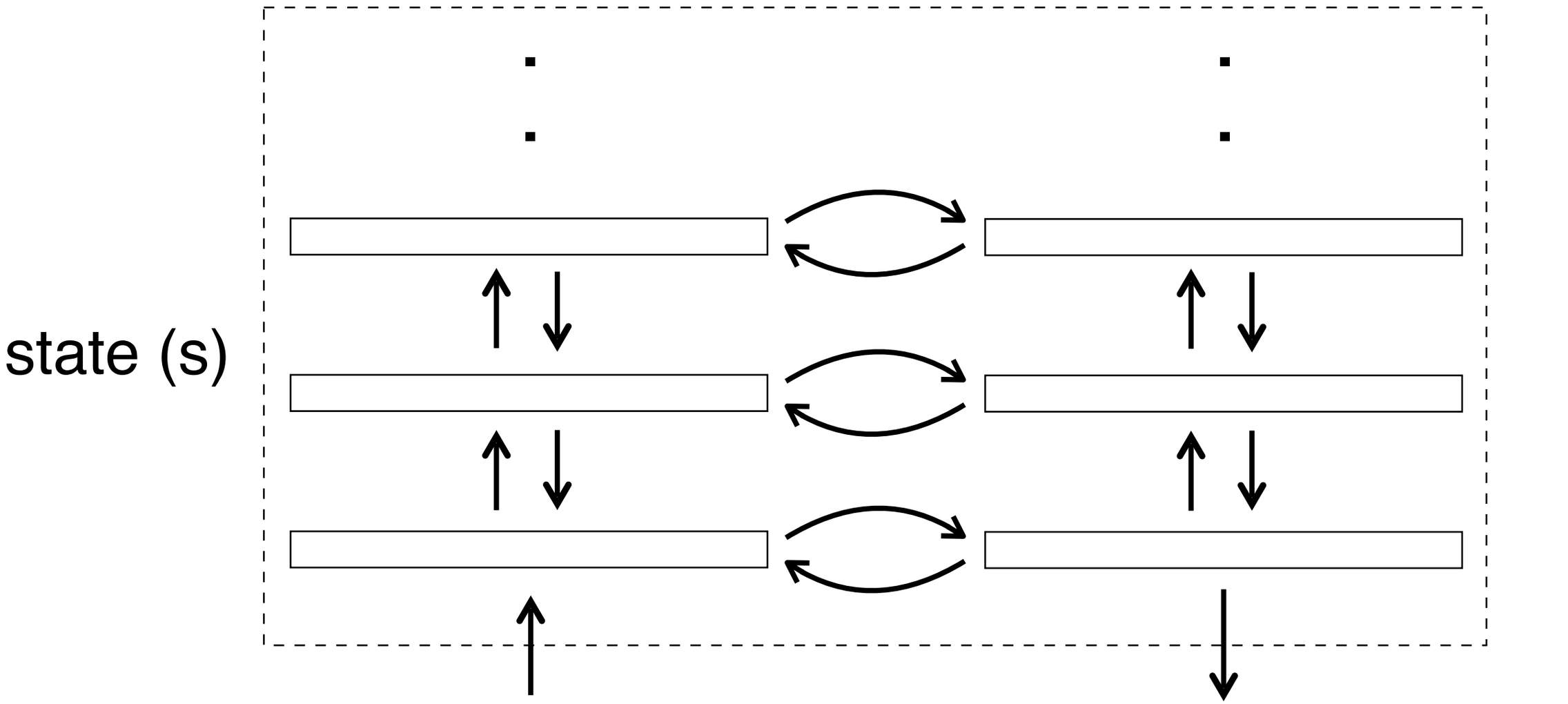


jumping spider

sand wasp



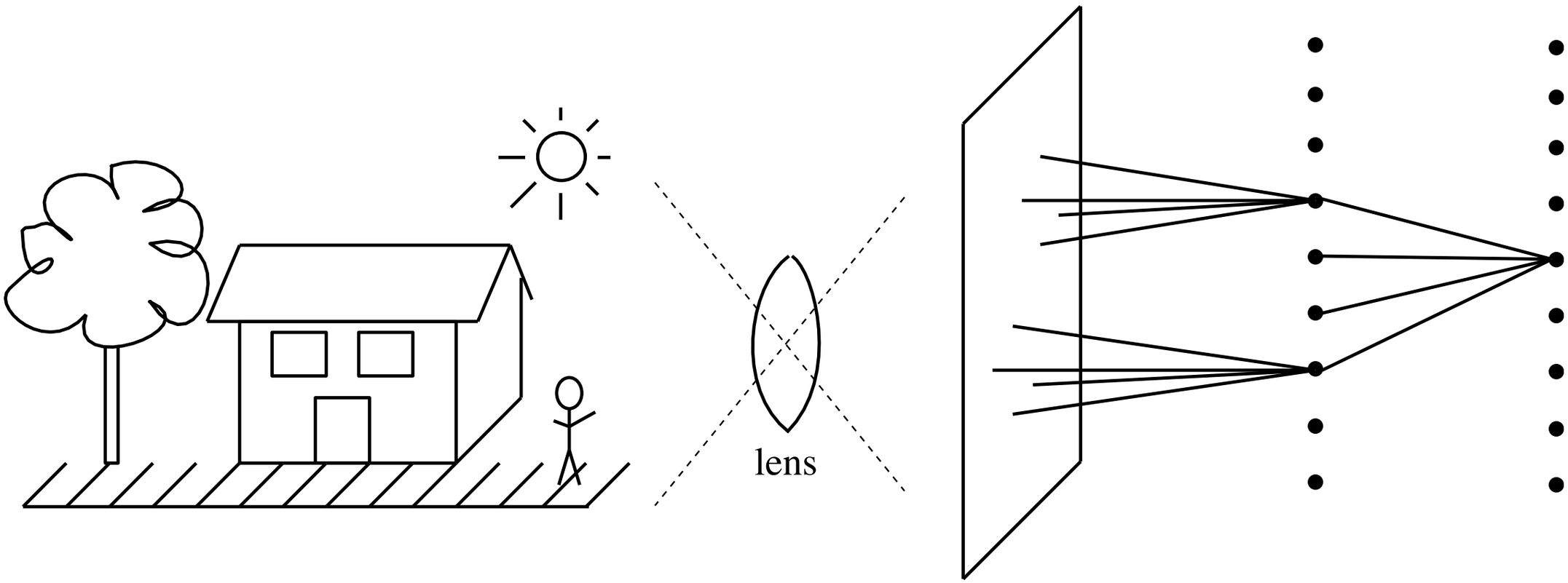
box jellyfish



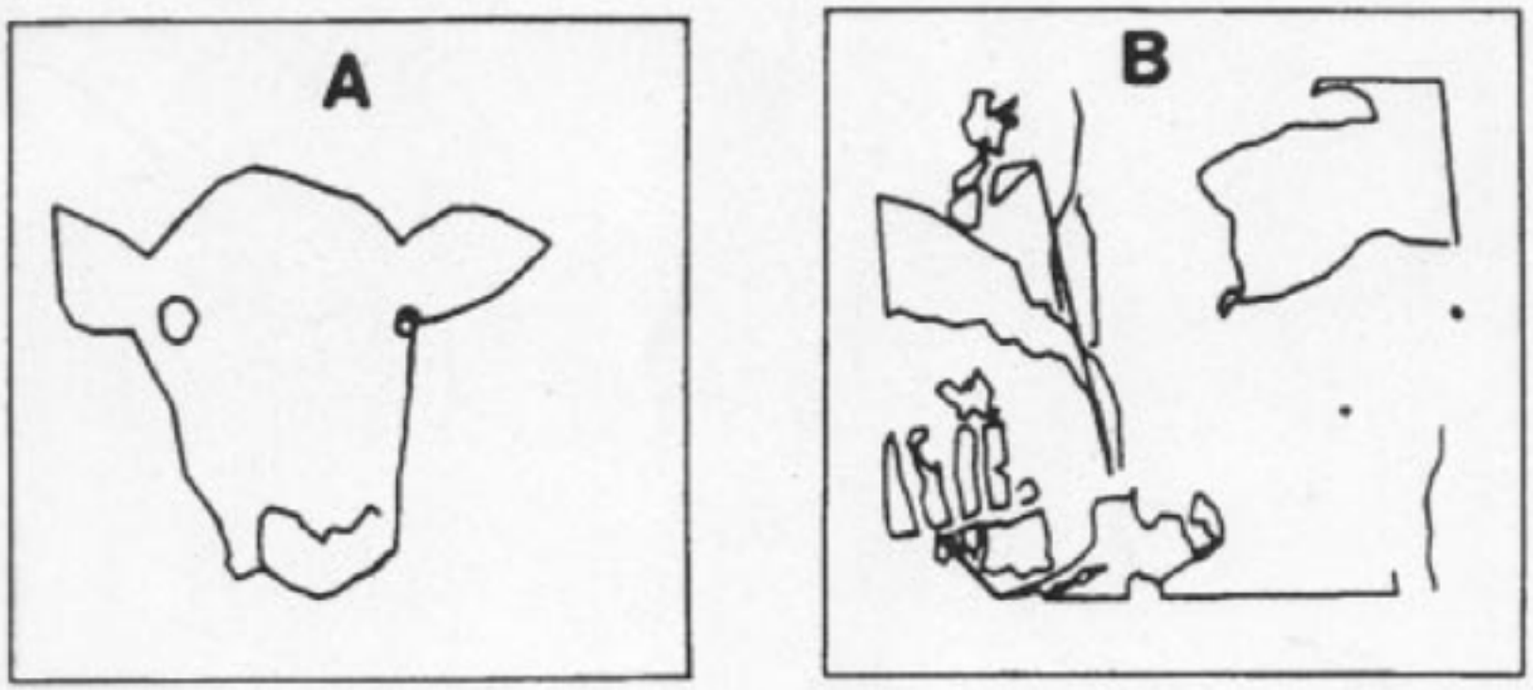
$$\tau \dot{s} + s = g(s, x, a; w)$$

$$a = f(s)$$

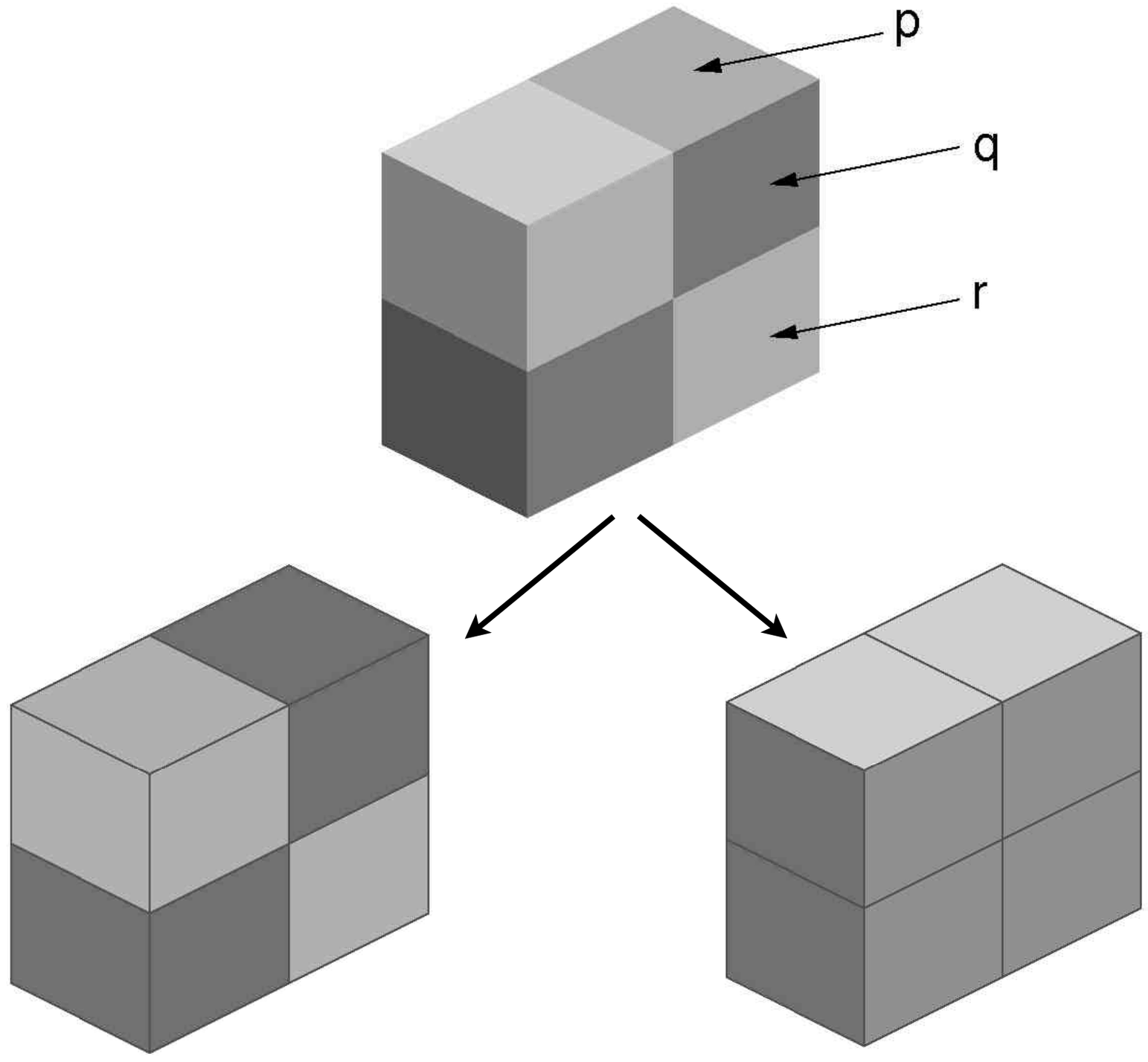
Vision as inference



b



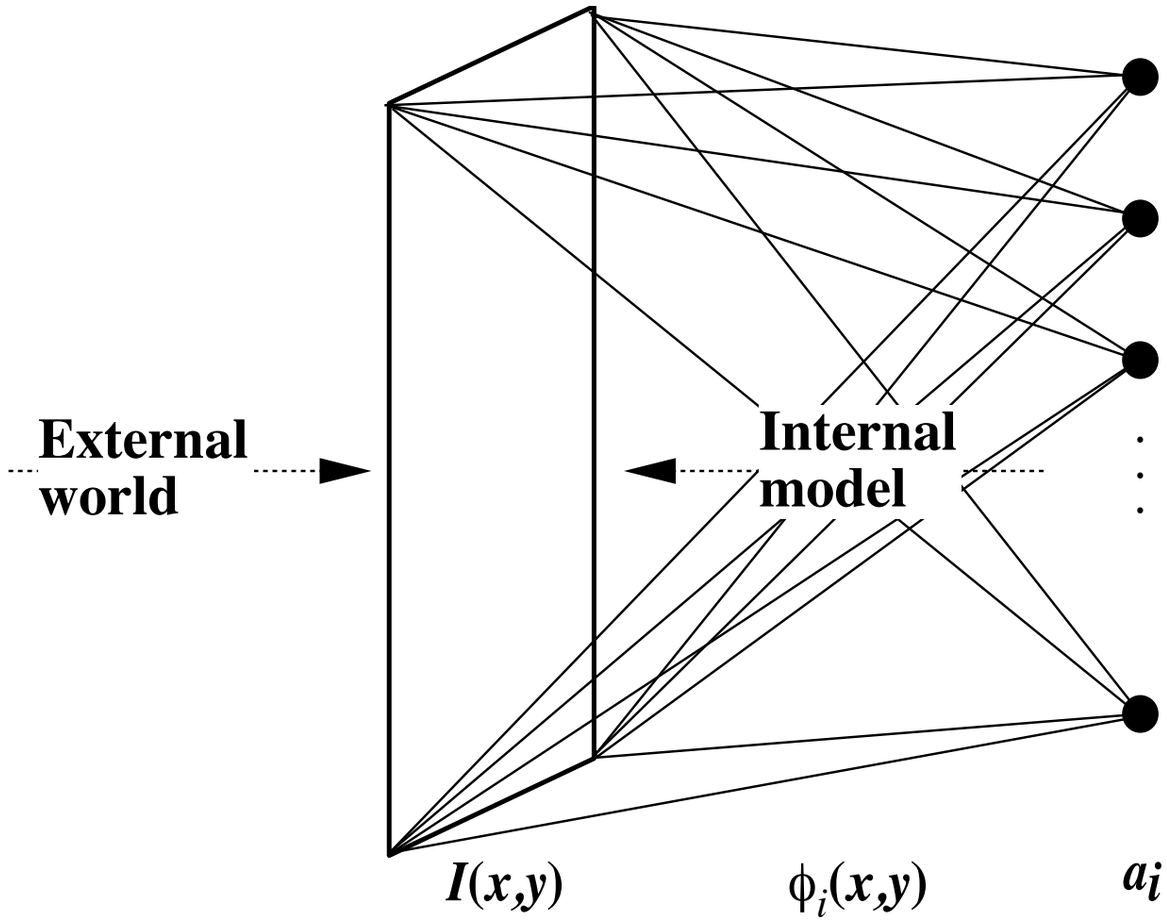
Separation of shape and reflectance



(Adelson, 2000)

Possible neural circuits for inferential computation in V1

1. Sparse coding
2. Separating form and motion from time-varying images



Sparse coding image model

(Olshausen & Field, 1996;
Chen, Donoho & Saunders 1995)

$$I(\vec{x}) = \sum_{i=1}^{M} a_i \, \phi_i(\vec{x}) + \epsilon(\vec{x})$$

image

neural activities (sparse)

features

other stuff

Energy function

$$E = \frac{1}{2}|\mathbf{I} - \Phi\,\mathbf{a}|^2 + \lambda \sum_i C(a_i)$$

preserve information

be sparse

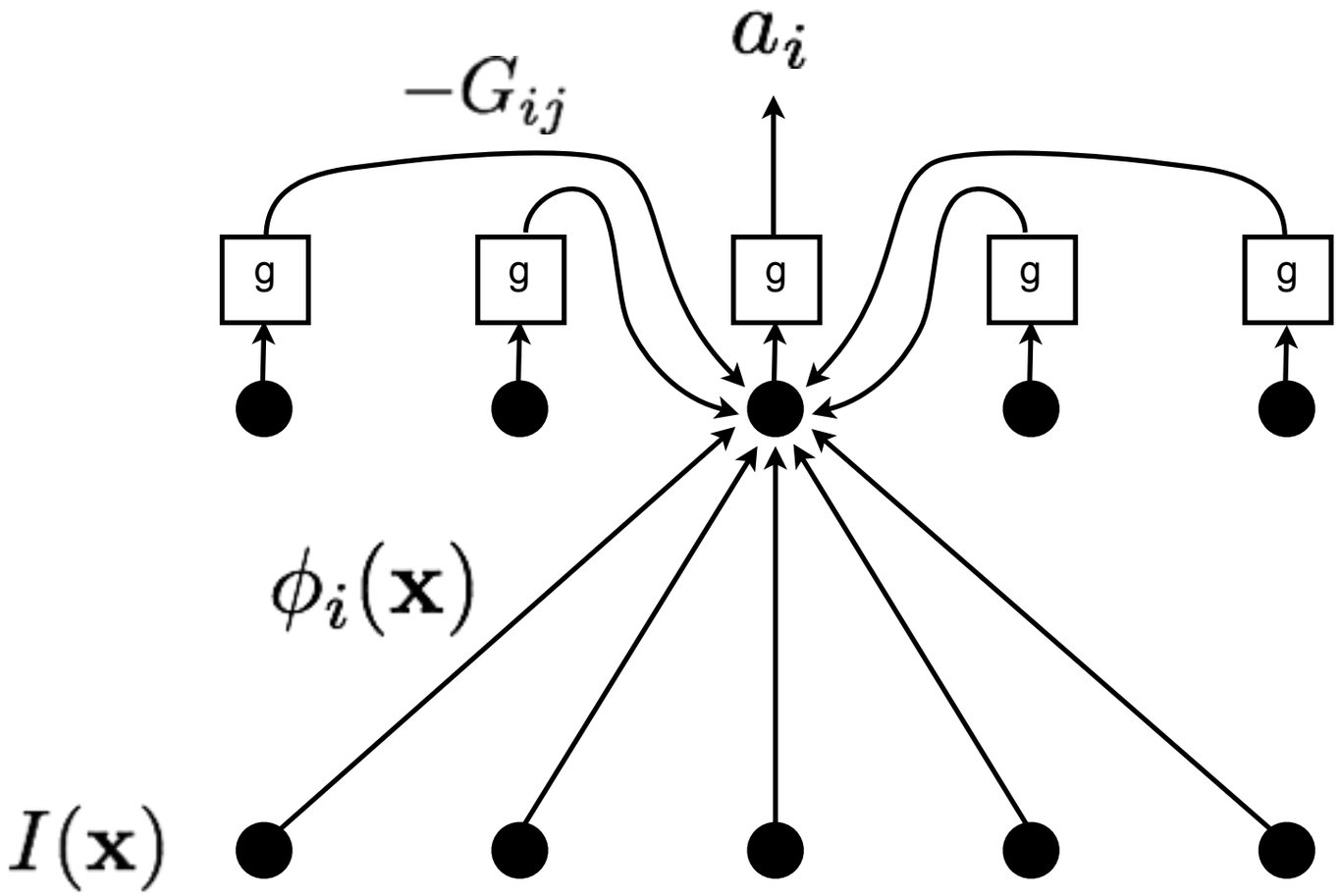
Energy function

$$E = \frac{1}{2}|\mathbf{I} - \Phi\,\mathbf{a}|^2 + \lambda \sum_i C(a_i)$$

-log **P(I | a)** **P(a)**

Coefficients a_i may be computed via thresholding and lateral inhibition

('LCA' - Rozell, Johnson, Baraniuk & Olshausen, 2008)



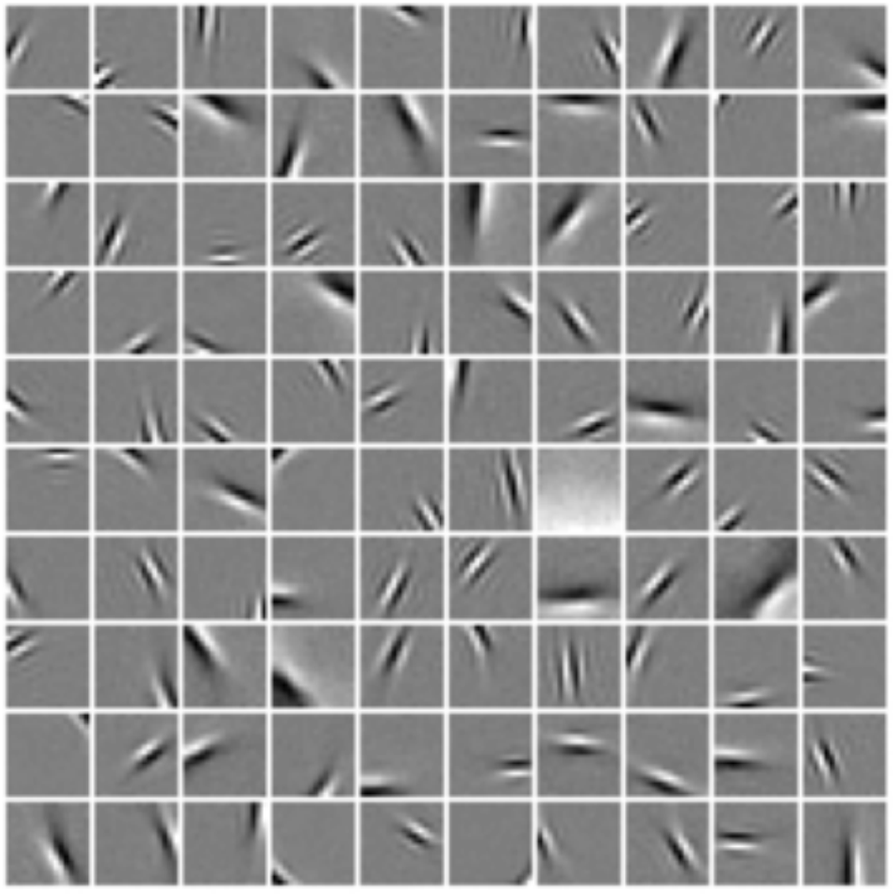
$$\tau\,\dot{u}_i + u_i \;=\; b_i - \sum_{j \neq i} G_{ij}\,a_j$$

$$a_i \;=\; g(u_i)$$

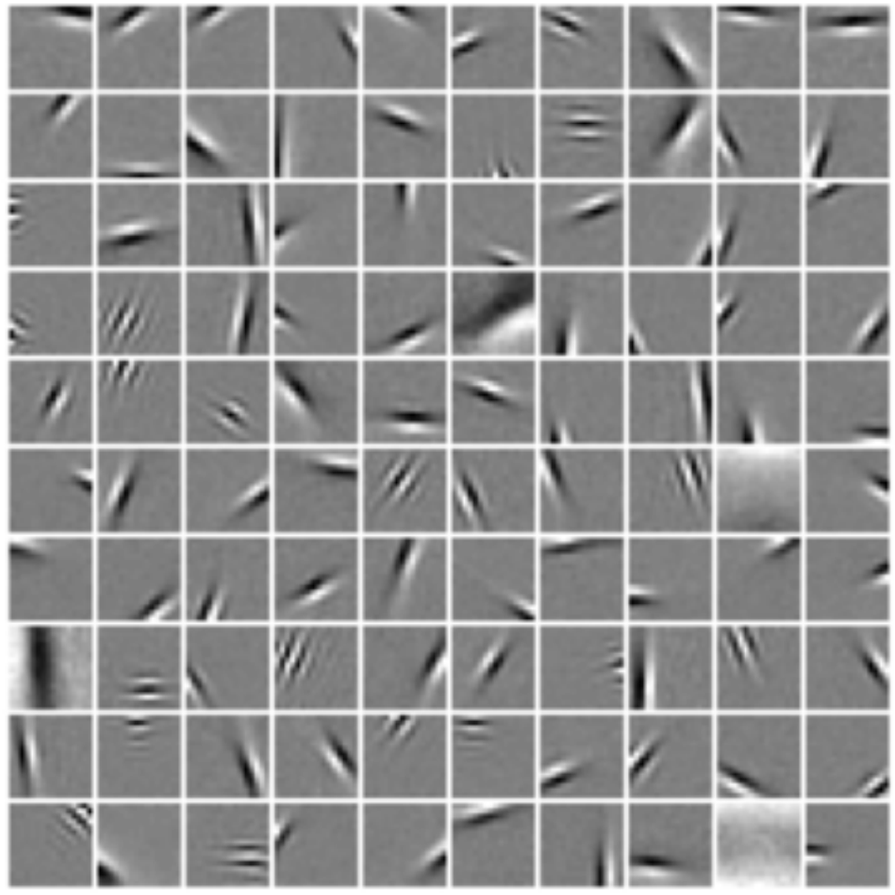
$$b_i = \sum_{\mathbf{x}} \phi_i(\mathbf{x})\, I(\mathbf{x})$$

$$G_{ij} = \sum_{\mathbf{x}} \phi_i(\mathbf{x})\, \phi_j(\mathbf{x})$$

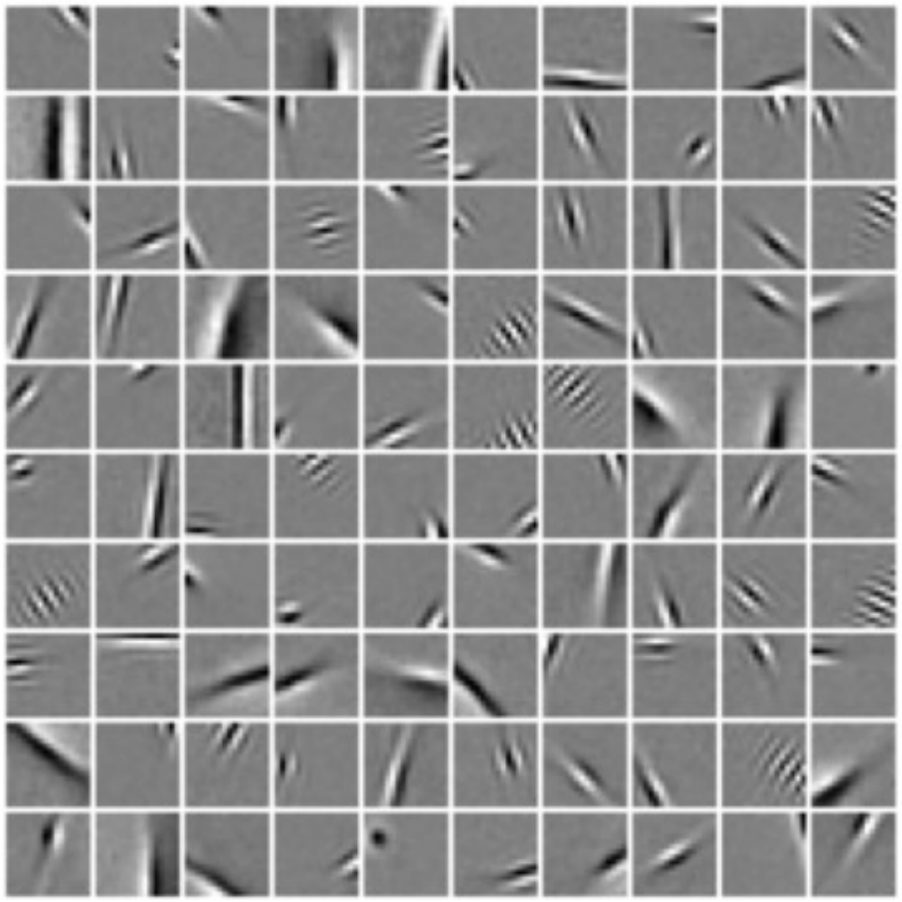
1.25x



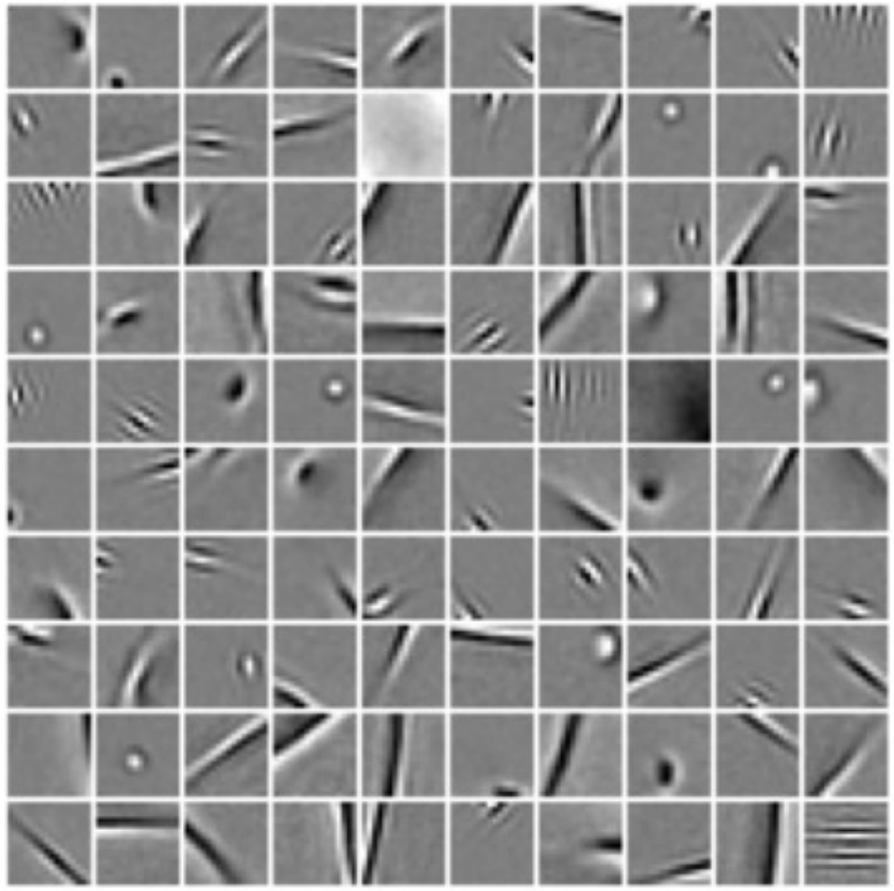
2.5x



5x



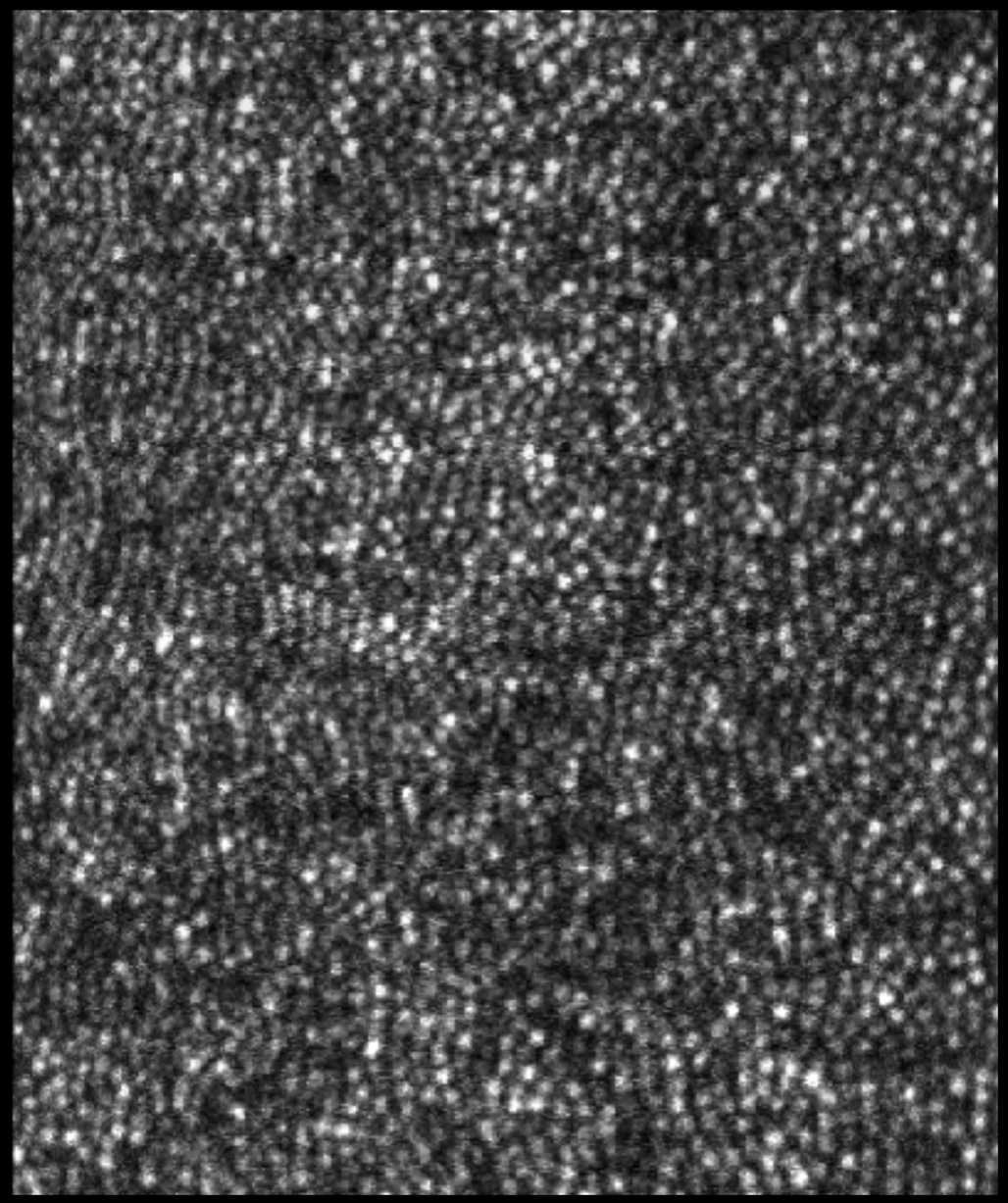
10x



Two examples

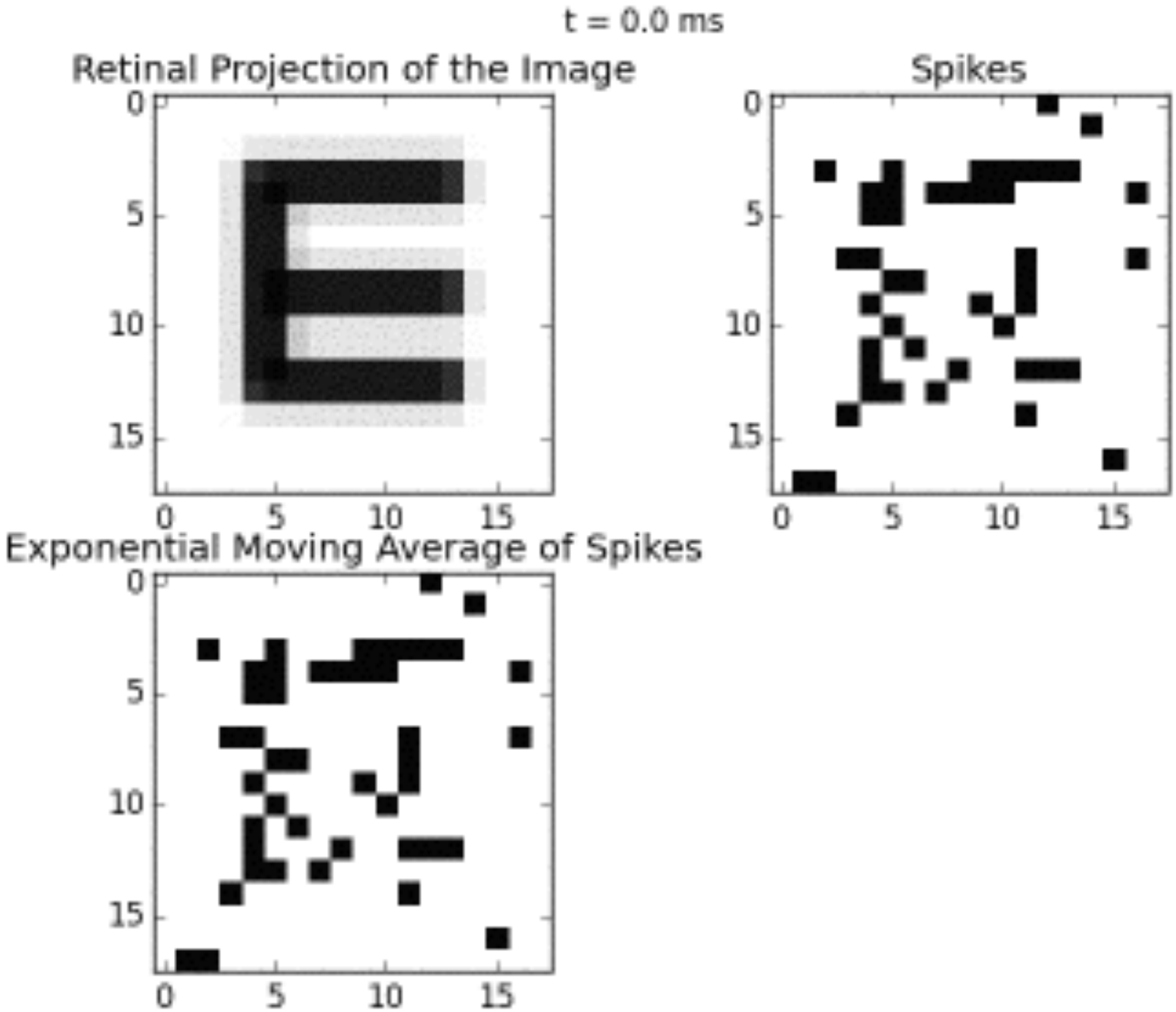
1. Sparse coding
2. **Separating form and motion from time-varying images**

Visual perception requires separation of form and motion from time-varying retinal images



(eye movement data from Austin Roorda, UC Berkeley)

Simple averaging is not sufficient



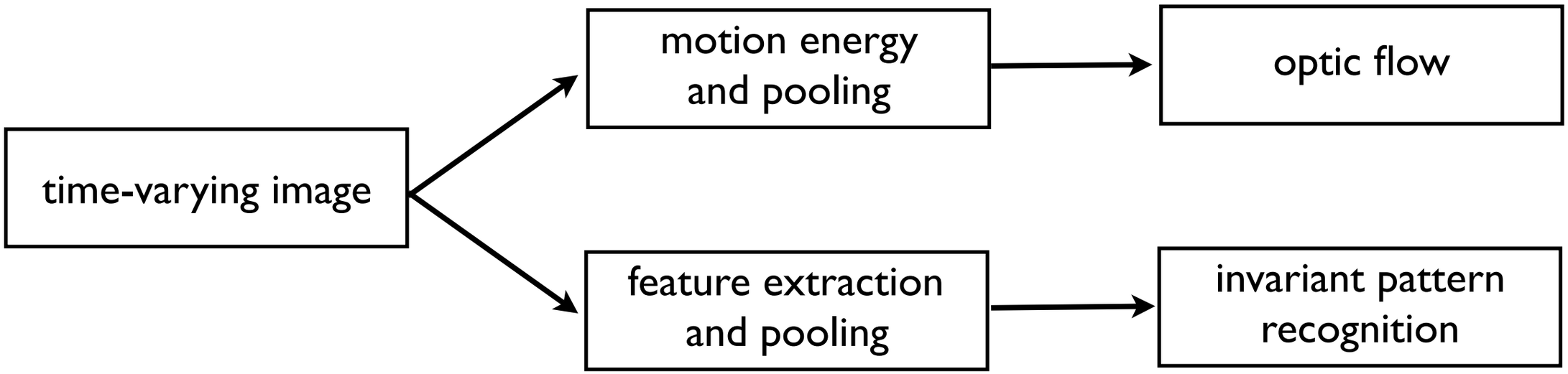
The problem

$$I(\vec{x}, t) = S(\vec{x} - \Delta\vec{x}(t)) + \epsilon(\vec{x}, t)$$

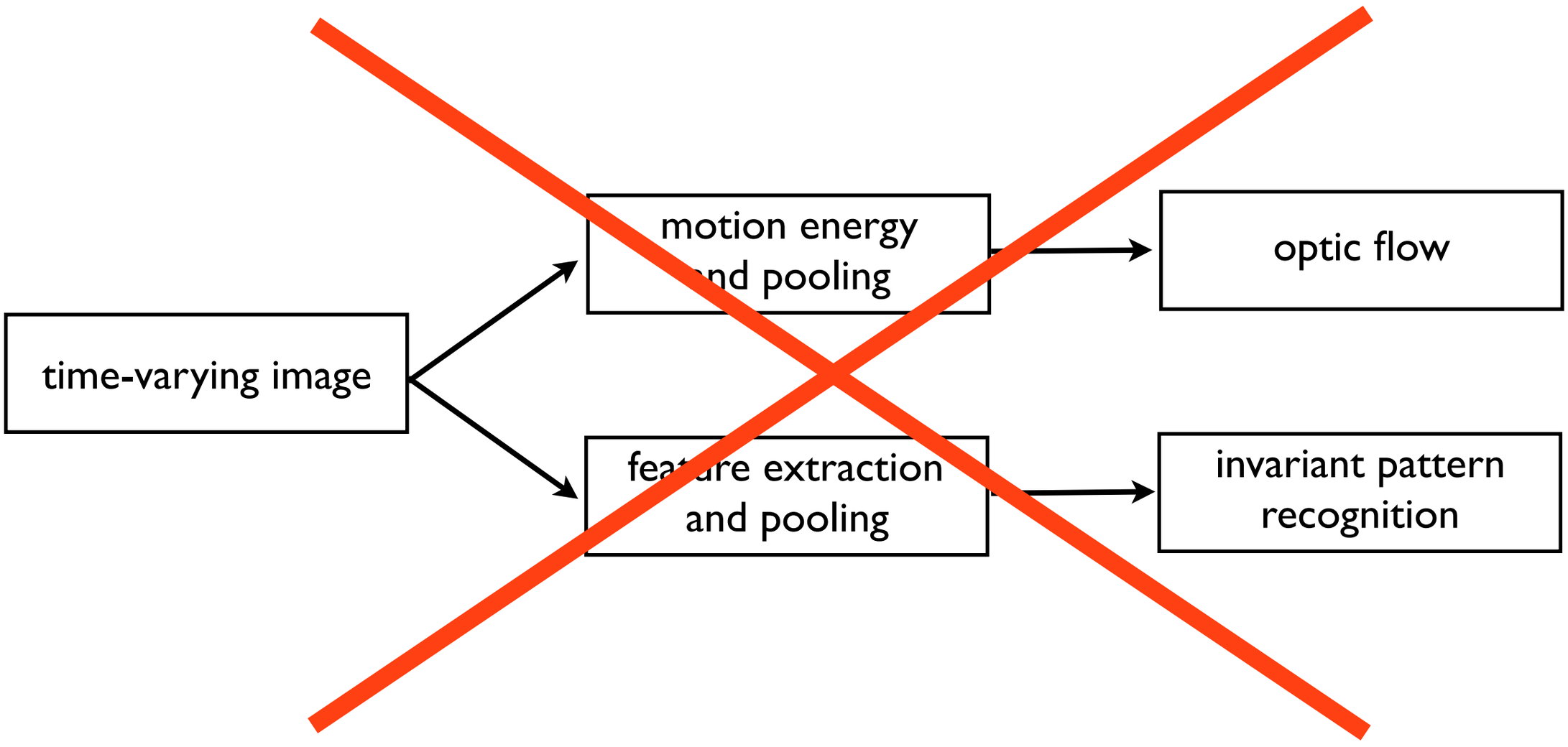
$$\hat{\Delta\vec{x}}(t) = \arg\min_{\Delta\vec{x}(t)} |I(\vec{x}, t) - S(\vec{x} - \Delta\vec{x}(t))|^2$$

$$\hat{S}(\vec{x}) = \int I(\vec{x} + \Delta\vec{x}(t))\, dt$$

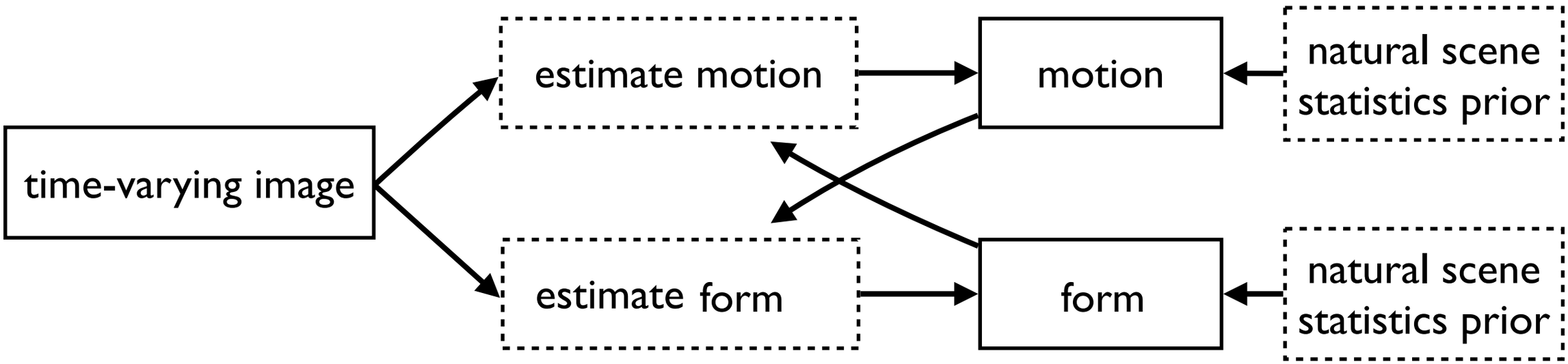
Traditional models compute motion and form independently



Traditional models compute motion and form independently

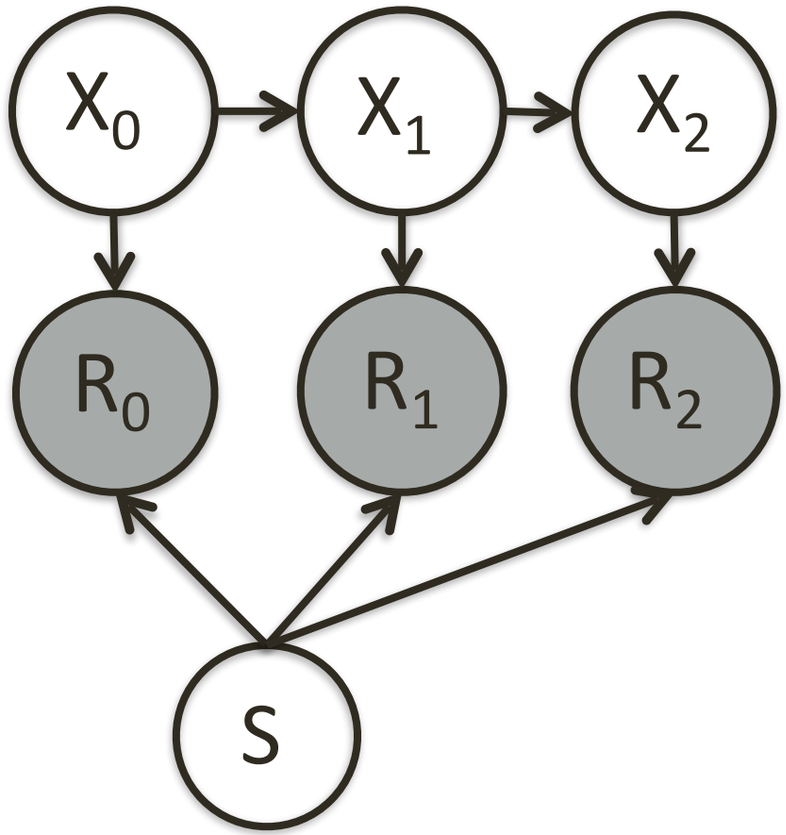


Motion *and* form must be estimated simultaneously



Graphical model for separating form and motion

(Alex Anderson, Ph.D. thesis)



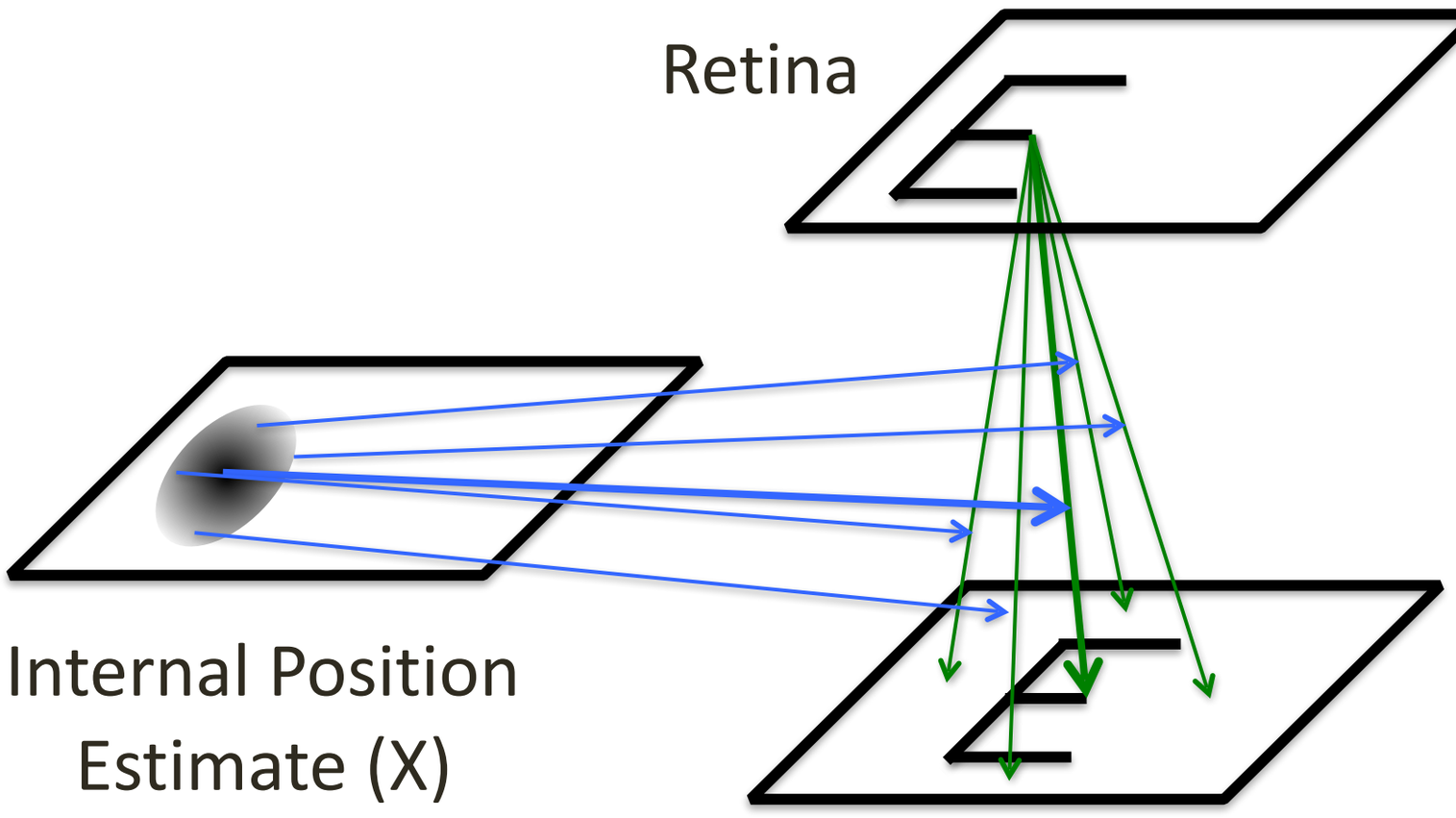
Eye position

Spikes
(from LGN afferents)

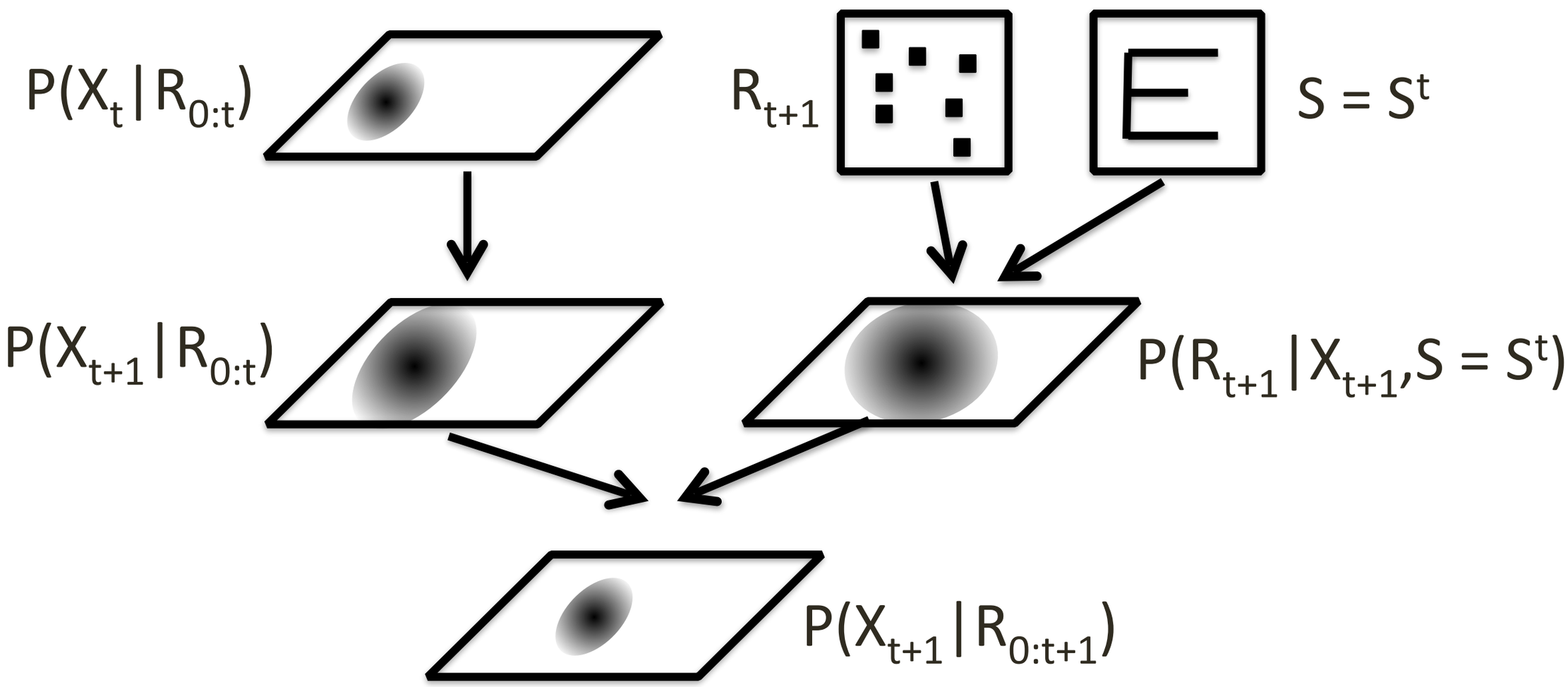
Pattern

$$\hat{S} = \arg\max_{S} \ \log P(R|S)$$

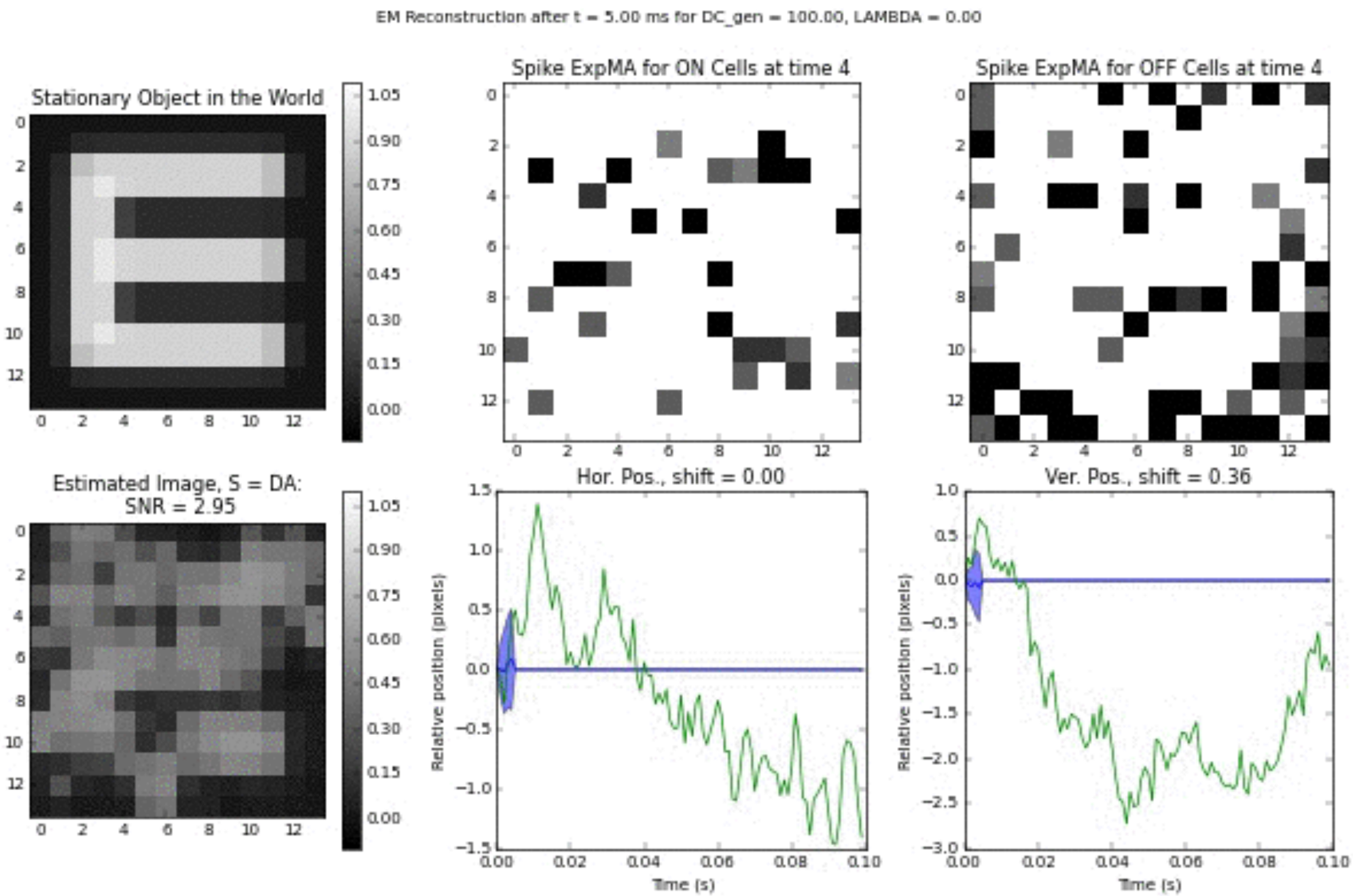
Given current estimate of position (X), update S



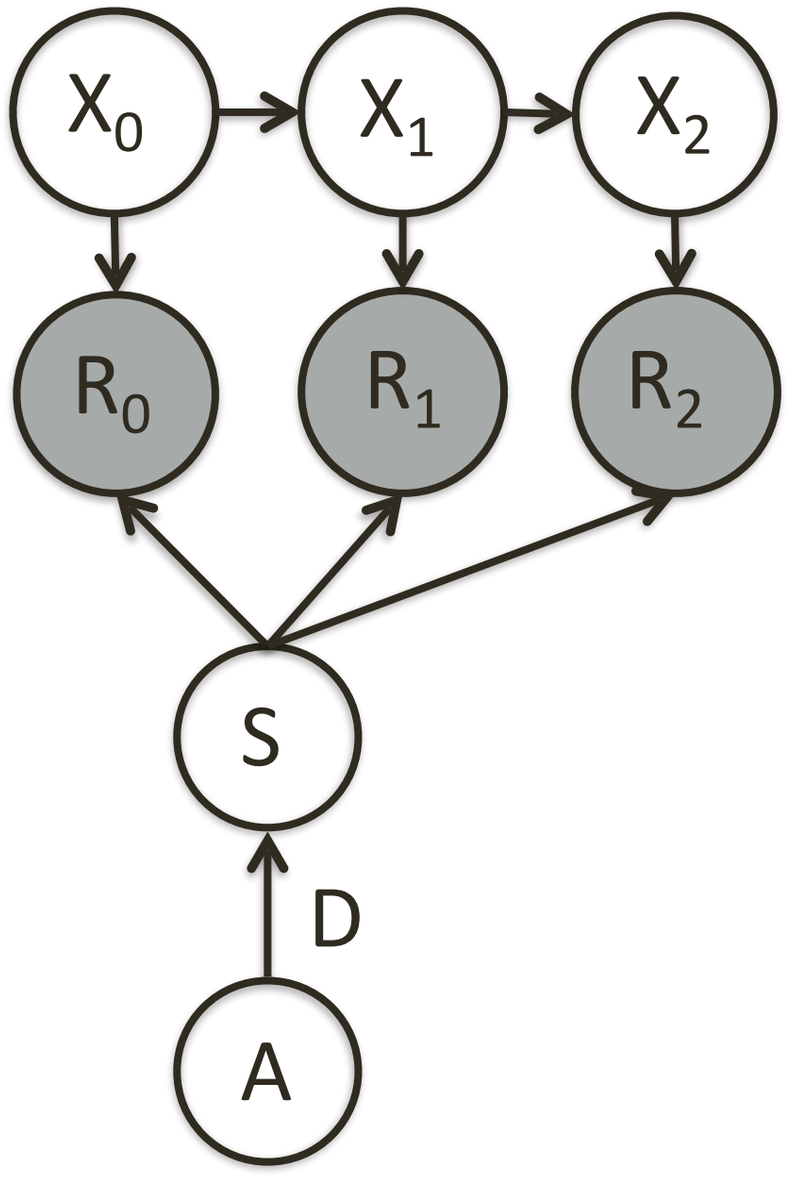
Given current estimate of form (S), update X



Joint estimation of form (S) and position (X)



Including a prior over form (S)



Eye position

Spikes
(from LGN afferents)

Pattern

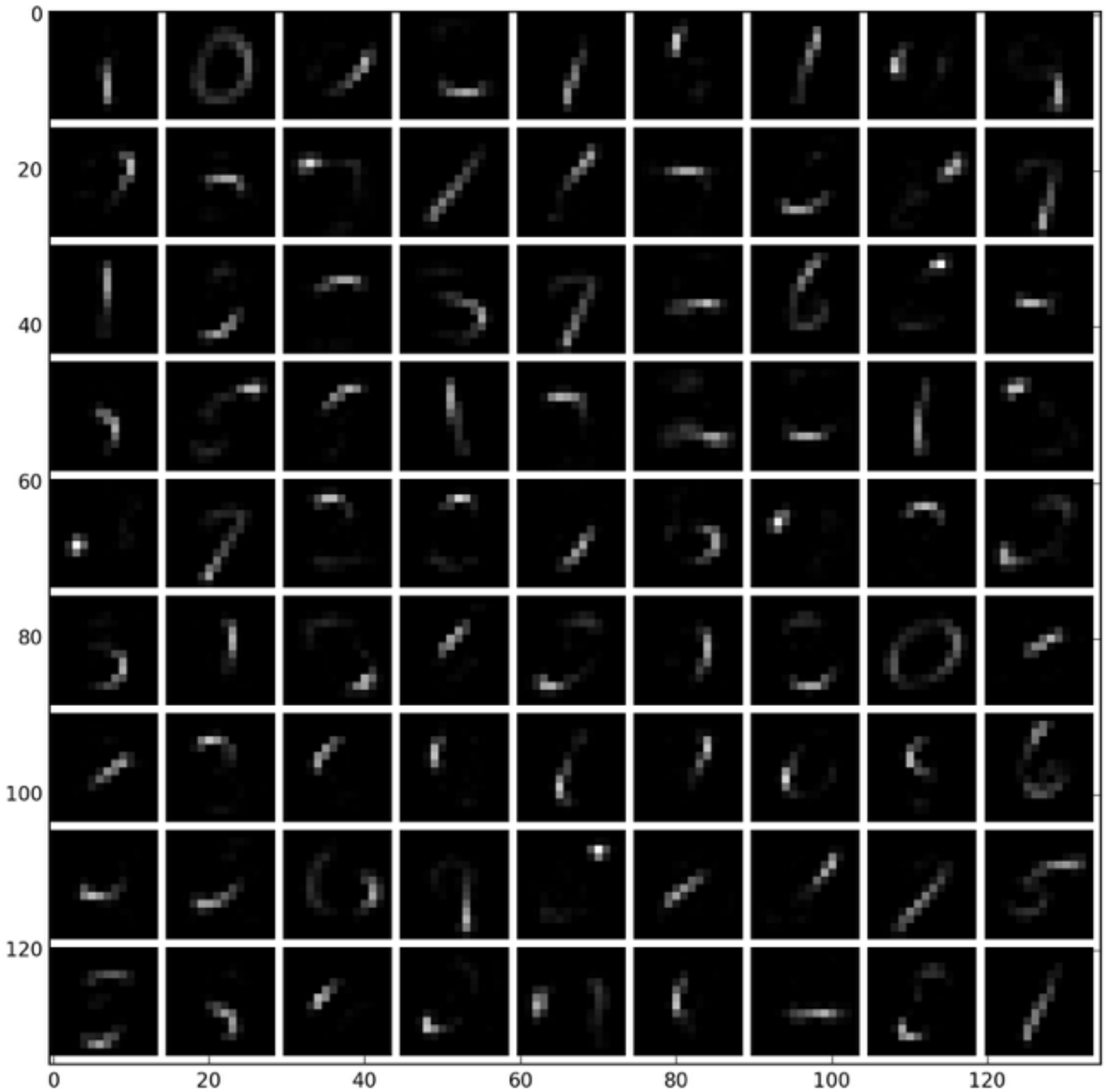
Dictionary

Sparse representation

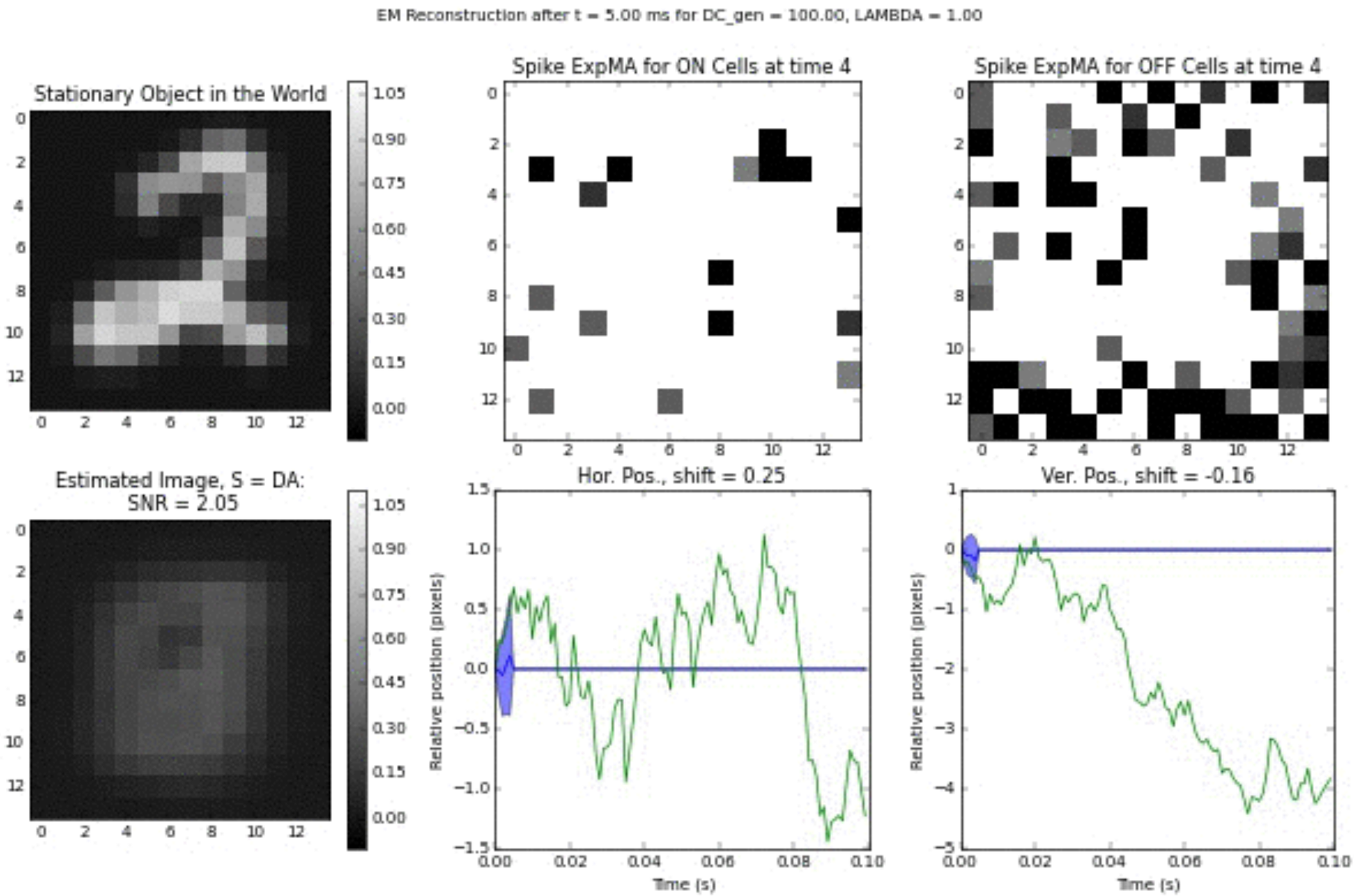
$$\hat{A} = \arg\max_{A} \; \log P(R|A) + \log P(A)$$

sparse

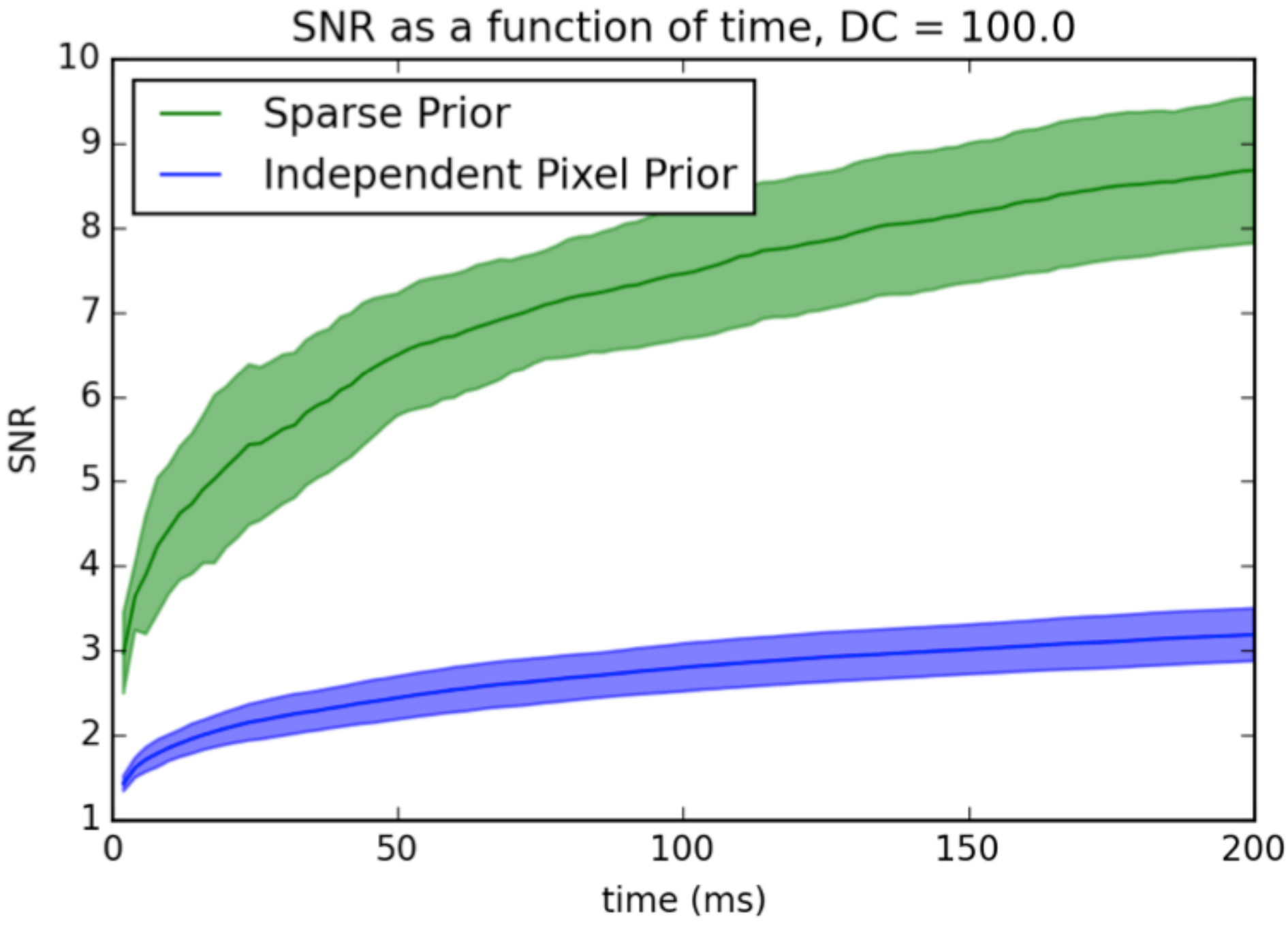
Learned dictionary *D*



Prior over form (S) improves inference



Form prior improves inference



Main points

- Perception seems better described as an inference problem that attempts to disentangle underlying causes from image data.

- Inference involves bidirectional information flow both within and between levels of representation.

- This moves us away from thinking of ‘receptive fields’ and instead toward how populations of neurons interact to perform collective computations.

Papers

Olshausen BA (2014) Perception as an Inference Problem. In: *The Cognitive Neurosciences V*. M. Gazzaniga, R. Mangun, Eds. MIT Press.
http://redwood.berkeley.edu/bruno/papers/perception-as-inference.pdf

Rozell CJ, Johnson DH, Baraniuk RG, Olshausen BA (2008). Sparse Coding via Thresholding and Local Competition in Neural Circuits. *Neural Computation, 20*, 2526-2563.
http://redwood.berkeley.edu/bruno/papers/rozell-sparse-coding-nc08.pdf

Olshausen BA (2013) Highly overcomplete sparse coding. In: *SPIE Proceedings vol. 8651: Human Vision and Electronic Imaging XVIII*, (B.E. Rogowitz, T.N. Pappas, H. de Ridder, Eds.), Feb. 4-7, 2013, San Francisco, California.
http://redwood.berkeley.edu/bruno/papers/highly-overcomplete-SPIE.pdf