6.869 Advances in Computer Vision Spring 2010

http://people.csail.mit.edu/torralba/courses/6.869/6.869. computervision.htm

Lecture 22
Miscellaneous

Project presentations

May 5 1pm – 2:30pm

2:30pm – 4pm

Presentations

5 Min + 2 min questions

- **Jon Brookchire**
- **Daniel Jonathan McDuff**
- **Matt Miller**
- **Julian Hernandez Munoz***
- **Nick Loomis**
- **Geoff Gorman ***
- **Roarke Horstmeyer**
- **Kevin Chiu**
- **Michael Rubinstein**
- Haitham Hassanieh
- **Michael Kuo**
- **Derya Akkaynay**
- **Nicholas Edelman***
- **Adam Kraft**
- **Matt Hirsch**
- **Jianxiong Xiao**
- Huan Liu*
- **Jenny Liu**
- Haitham Hassanieh
- **Guy-Richard Kayombya**
- **Lawson Wong***
- **Hao-Yu Wu**
- **Phillip Isola**
- **Jeff Kaeli**
- **Gershon Dublon****

How to give a talk

http://www.cs.berkeley.edu/~messer/Bad_talk.html

http://www-psych.stanford.edu/~lera/talk.html

First, some bad news

The more you work on a talk, the better it gets: if you work on it for 3 hours, the talk you give will be better than if you had only worked on it for 2 hours. If you work on it for 5 hours, it will be better still. 7 hours, better yet…

All talks are important

There are no unimportant talks.

There are no big or small audiences.

Prepare each talk with the same enthusiasm.

How to give a talk

Delivering:

Look at the audience! Try not to talk to your laptop or to the screen. Instead, look at the other humans in the room.

You have to believe in what you present, be confident… even if it only lasts for the time of your presentation.

Do not be afraid to acknowledge limitations of whatever you are presenting. Limitations are good. They leave job for the people to come. Trying to hide the problems in your work will make the preparation of the talk a lot harder and your self confidence will be hurt.

Let the audience see your personality

- They want to see you enjoy yourself.
- They want to see what you love about the work.
- People really respond to the human parts of a talk. Those parts help the audience with their difficult task of listening to an hour-long talk on a technical subject. What was easy, what was fun, what was hard about the work?
- Don’t be afraid to be yourself and to be quirky.

The different kinds of talks you'll have to give as a researcher

- 2-5 minute talks
- 20 -30 minute conference presentations
- 30-60 minute colloquia

How to give a talk

Talk organization: here there are as many theories as there are talks. Here there are some extreme advices:

1. Go into details / only big picture
2. Go in depth on a single topic / cover as many things as you can
3. Be serious (never make jokes, maybe only one) / be funny (it is just another form of theater)

Corollary: ask people for advice, but at the end, if will be just you and the audience. Chose what fits best your style.

What everybody agree on is that you have to practice in advance (the less your experience, the more you have to practice). Do it with an audience or without, but practice.

The best advice I got came from Yair Weiss while preparing my job talk:

"just give a good talk"

How to give the project class talk

Initial conditions:

- I started with a great idea
- It did not work
- The day before the presentation I found 40 papers that already did this work
- Then I also realized that the idea was not so great

How do I present?

- Just give a good talk

Sources on writing technical papers

- How to Get Your SIGGRAPH Paper Rejected, Jim Kajiya, SIGGRAPH 1993 Papers Chair, http://www.siggraph.org/publications/instructions/rejected.html
- Ted Adelson's Informal guidelines for writing a paper, 1991. http://www.ai.mit.edu/courses/6.899/papers/ted.htm
- Notes on technical writing, Don Knuth, 1989.

http://www.ai.mit.edu/courses/6.899/papers/knuthAll.pdf

- What's wrong with these equations, David Mermin, Physics Today, Oct., 1989. http://www.ai.mit.edu/courses/6.899/papers/mermin.pdf
- Ten Simple Rules for Mathematical Writing, Dimitri P. Bertsekas http://www.mit.edu:8001/people/dimitrib/Ten_Rules.html

Knuth

24. The opening paragraph should be your best paragraph, and its first sentence should be your best sentence. If a paper starts badly, the reader will wince and be resigned to a difficult job of fighting with your prose. Conversely, if the beginning flows smoothly, the reader will be hooked and won't notice occasional lapses in the later parts.

    Probably the worst way to start is with a sentence of the form "An x is y." For example,

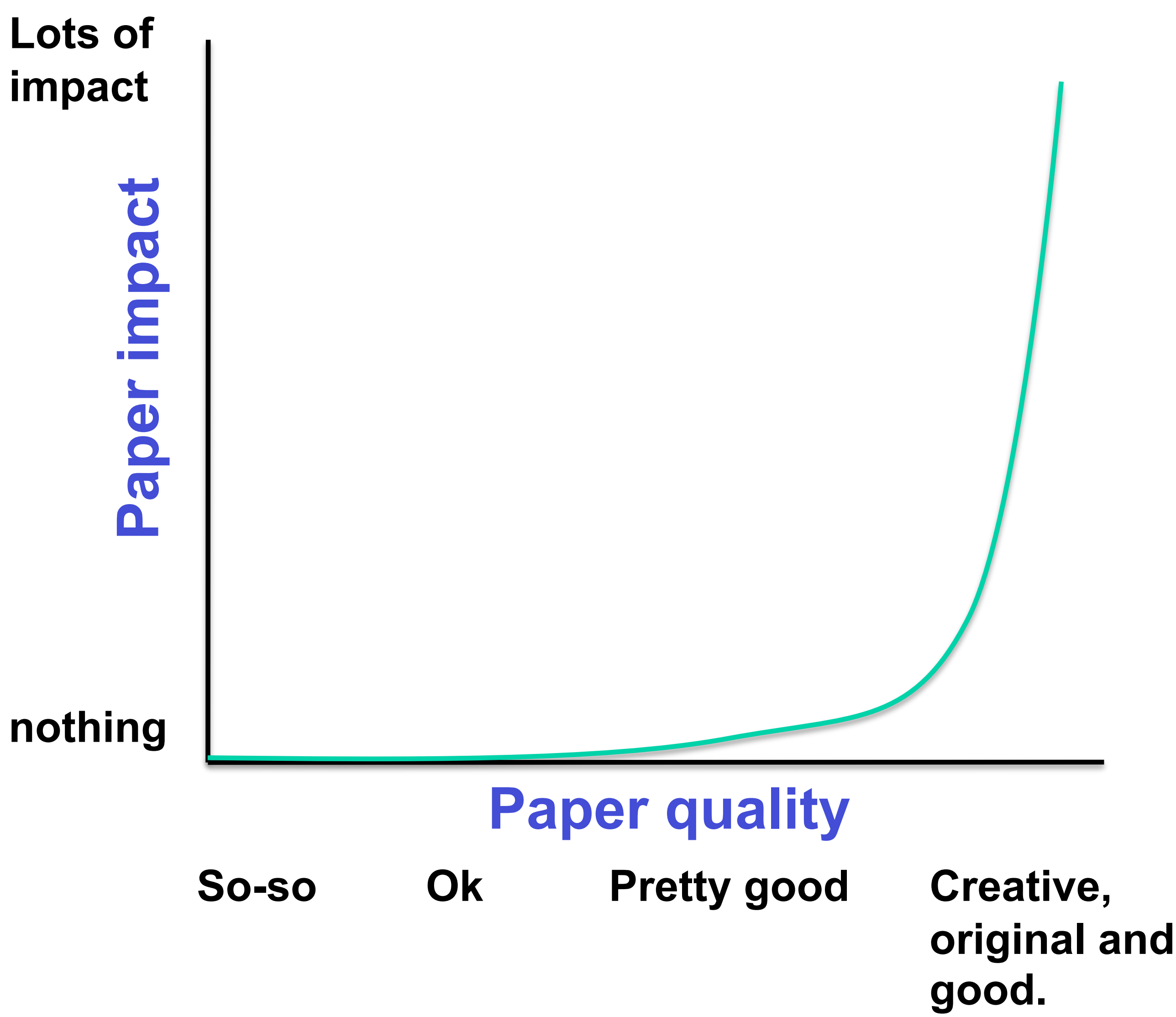
    Bad: An important method for internal sorting is quicksort.
    Good: Quicksort is an important method for internal sorting, because ...

    Bad: A commonly used data structure is the priority queue.
    Good: Priority queues are significant components of the data structures needed for many different applications.

Knuth on equations

13. Many readers will skim over formulas on their first reading of your exposition. Therefore, your sentences should flow smoothly when all but the simplest formulas are replaced by "blah" or some other grunting noise.

The paper impact curve



Tracking

Overview

For a nice overview, check David Fleet's page:

http://www.cs.toronto.edu/~fleet/research/Talks/humaneva07.pdf

Also this tutorial:

http://www.cs.toronto.edu/~ls/iccv2009tutorial/

Jepson, Fleet, and El-Maraghi tracker

IEEE Conference on Computer Vision and and Pattern Recognition, Kauai, 2001, Vol. I, pp. 415–422

Robust Online Appearance Models for Visual Tracking

Allan D. Jepson[*] David J. Fleet[†] Thomas F. El-Maraghi[*†]

[*] Department of Computer Science, University of Toronto, Toronto, M5S 1A4
[†] Xerox Palo Alto Research Center, 3333 Coyote Hill Rd, Palo Alto, CA 94304

Wandering, Stable, and Lost appearance model

- Introduce 3 competing models to explain the appearance of the tracked region:
  - A stable model—Gaussian with some mean and covariance.
  - A 2-frame motion tracker appearance model, to rebuild the stable model when it gets lost
  - An outlier model—uniform probability over all appearances.
- Introduce 3 competing processes to explain the observations

The motion tracker

- Motion prior prefers slow velocities and small accelerations.
- The WSL appearance model gives a likelihood for each possible new position, orientation, and scale of the tracked region.
- They combine that with the motion prior to find the most probable position, orientation, and scale of the tracked region in the next frame.
- Gives state-of-the-art tracking results.

Jepson, Fleet, and El-Maraghi tracker

Figure 4. The adaptation of the model during tracking. (top) The target region in selected frames 200, 300, 480. (bottom) The stable component's mixing probability (left) and mean (right) for the selected frames.

Jepson, Fleet, and El-Maraghi tracker

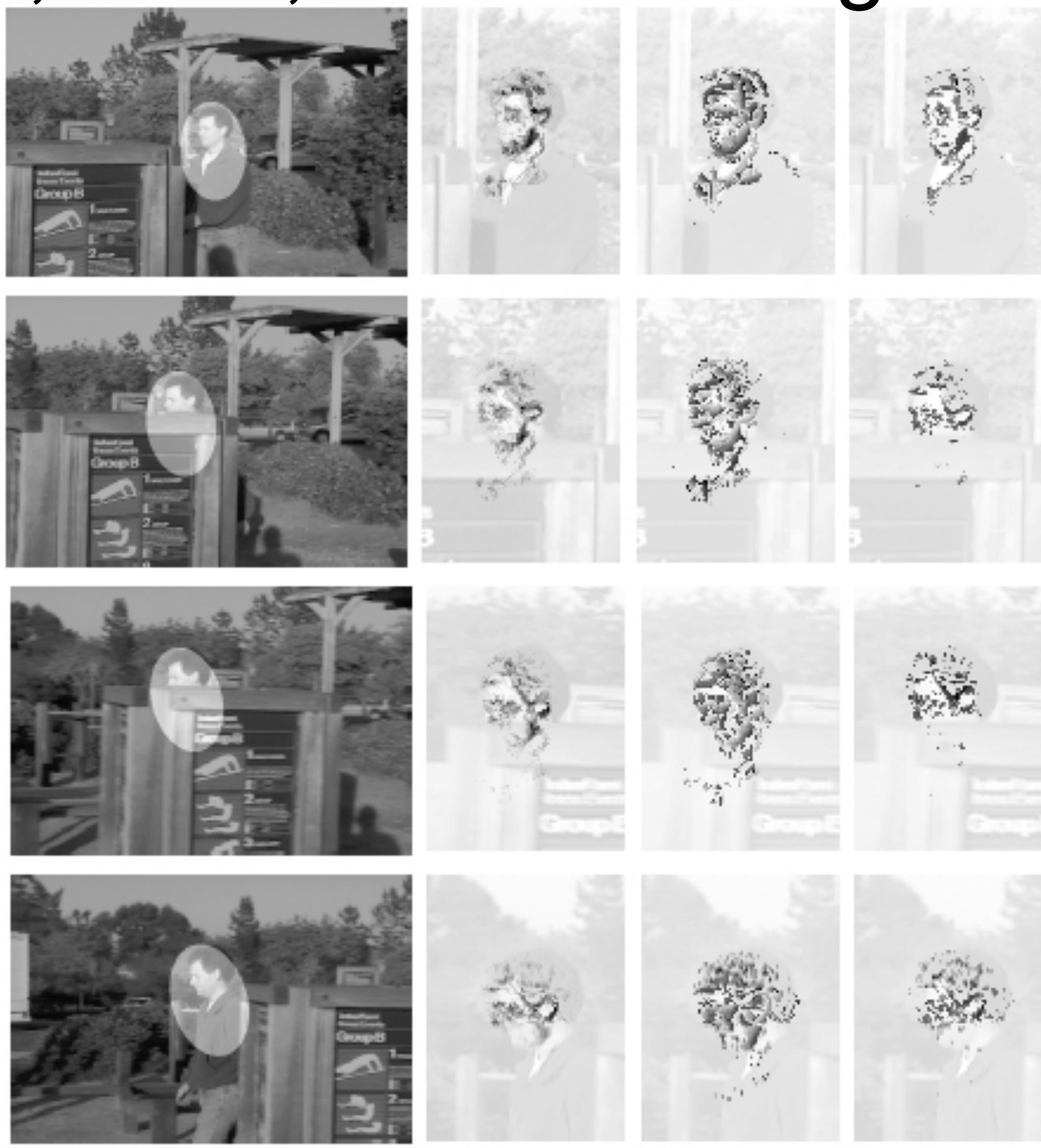


Figure 3. Each row shows, from left to right, the tracking region, the stable component's mixing probability $m_s(\mathbf{x}, t)$, mean $\mu_s(\mathbf{x}, t)$, and ownership probability $o_s(\mathbf{x}, t)$. The rows correspond to frames 244, 259, 274, and 289, top to bottom. Note the model persistence and the drop in data ownership within the occluded region.

Contour tracking by stochastic propagation of conditional density - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://research.microsoft.com/users/misard/abstracts/eccv96.isard.html Go

Getting Started Latest Headlines

Contour tracking by stochastic propagation of conditional density

Michael Isard and Andrew Blake

Proc. European Conference on Computer Vision, vol. 1, pp. 343--356, Cambridge UK, (1996).

Abstract

The problem of tracking curves in dense visual clutter is a challenging one. Trackers based on Kalman filters are of limited use; because they are based on Gaussian densities which are unimodal, they cannot represent simultaneous alternative hypotheses. Extensions to the Kalman filter to handle multiple data associations work satisfactorily in the simple case of point targets, but do not extend naturally to continuous curves. A new, stochastic algorithm is proposed here, the **Condensation** algorithm --- Conditional Density Propagation over time. It uses `factored sampling', a method previously applied to interpretation of static images, in which the distribution of possible interpretations is represented by a randomly generated set of representatives. The **Condensation** algorithm combines factored sampling with learned dynamical models to propagate an entire probability distribution for object position and shape, over time. The result is highly robust tracking of agile motion in clutter, markedly superior to what has previously been attainable from Kalman filtering. Notwithstanding the use of stochastic methods, the algorithm runs in near real-time.

Click here for a compressed postscript version

Back to

Michael Isard's home page

Done

File Edit View History Bookmarks Tools Help

http://www.robots.ox.ac.uk/~misard/condensation.html

Getting Started Latest Headlines

Google west concord, ma 5&10 Search PageRank Check AutoLink AutoFill Subscribe Options west

Gmail - old conden... | Mozilla Firefox Sta... | west concord, ma ... | Olive Garden Italia... | Approve new CSA... | leafmv.mpg (video... | The Condens...

The Condensation Algorithm

Background

Tracking objects through highly cluttered scenes is difficult. We believe that for tracking to be robust when following agile moving objects, in the presence of dense background clutter, probabilistic algorithms are essential. Previous algorithms, for example the Kalman filter, have been limited in the range of probability distributions they represent. We have developed a new algorithm, the **Condensation** algorithm (**Con**ditional **Dens**ity Propag**ation**) which allows quite general representations of probability. Experimental results show that this increased generality does indeed lead to a marked improvement in tracking performance. In addition to permitting high-quality tracking in clutter, the simplicity of the **Condensation** algorithm also allows the use of non-linear motion models more complex than those commonly used in Kalman filters. We have implemented a mixed discrete/continuous tracker in the **Condensation** framework which switches between multiple continuous Auto-Regressive Process motion models according to a discrete transition matrix. Also, by using the statistical technique of *importance sampling* it is possible to build a **Condensation** tracker which runs in real time, and we have implemented a real-time hand-tracker on a low-end SGI workstation. My D.Phil. thesis gives a thorough description of the algorithm and some applications.

Sample Code

Download source code of a simple implementation of the **Condensation** algorithm.

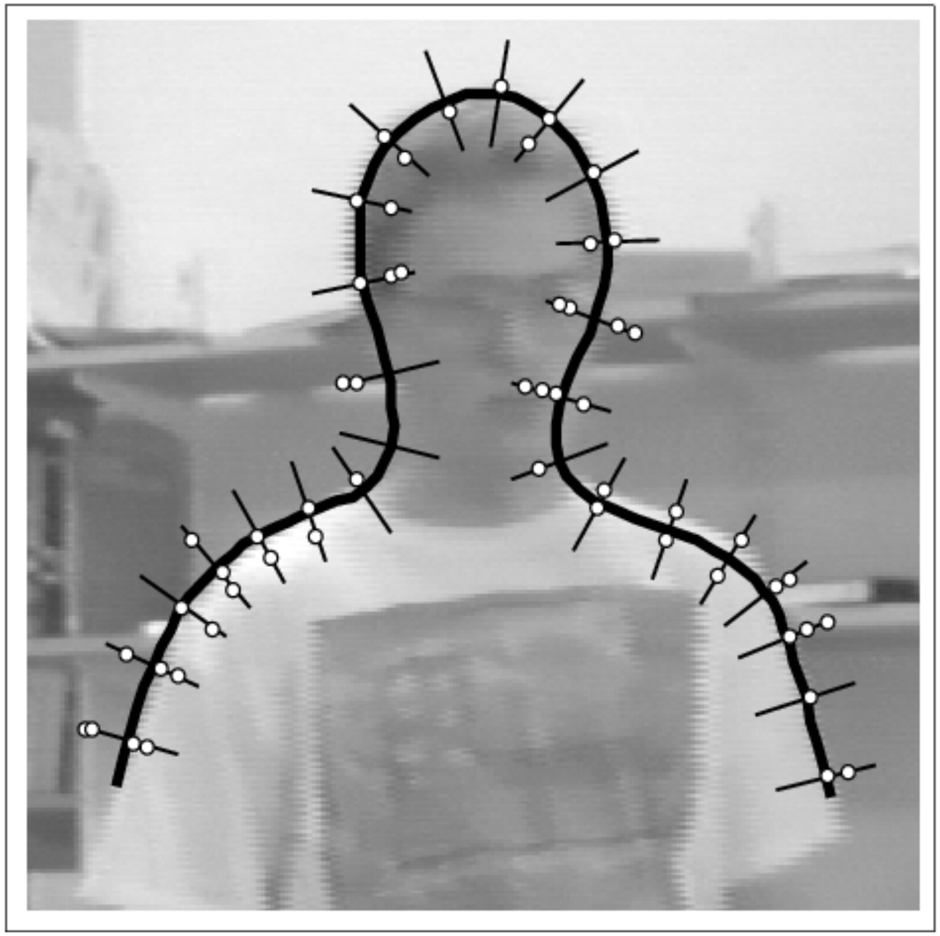
Results

Here is an MPEG (2.3Mb) showing the **Condensation** algorithm tracking a leaf blowing in the wind, against a background of

Find: Next Previous Highlight all Match case

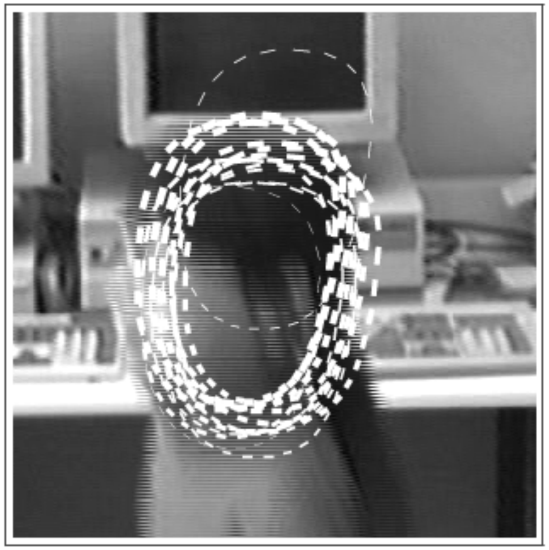
Done

Contour tracking



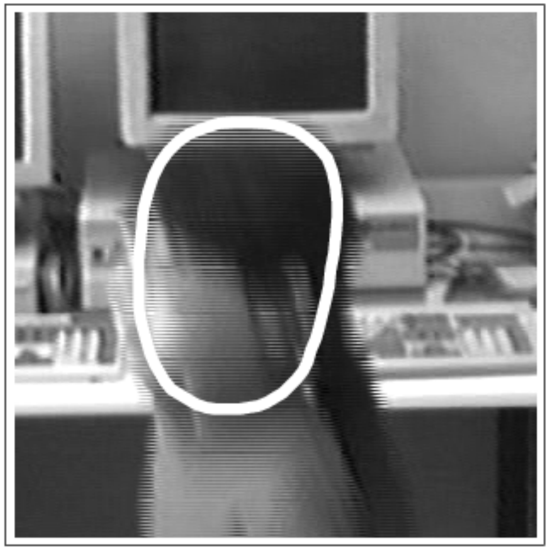
[Isard 1998]

Head tracking



(a)

Picture of the states represented by the top weighted particles



(b)

The mean state

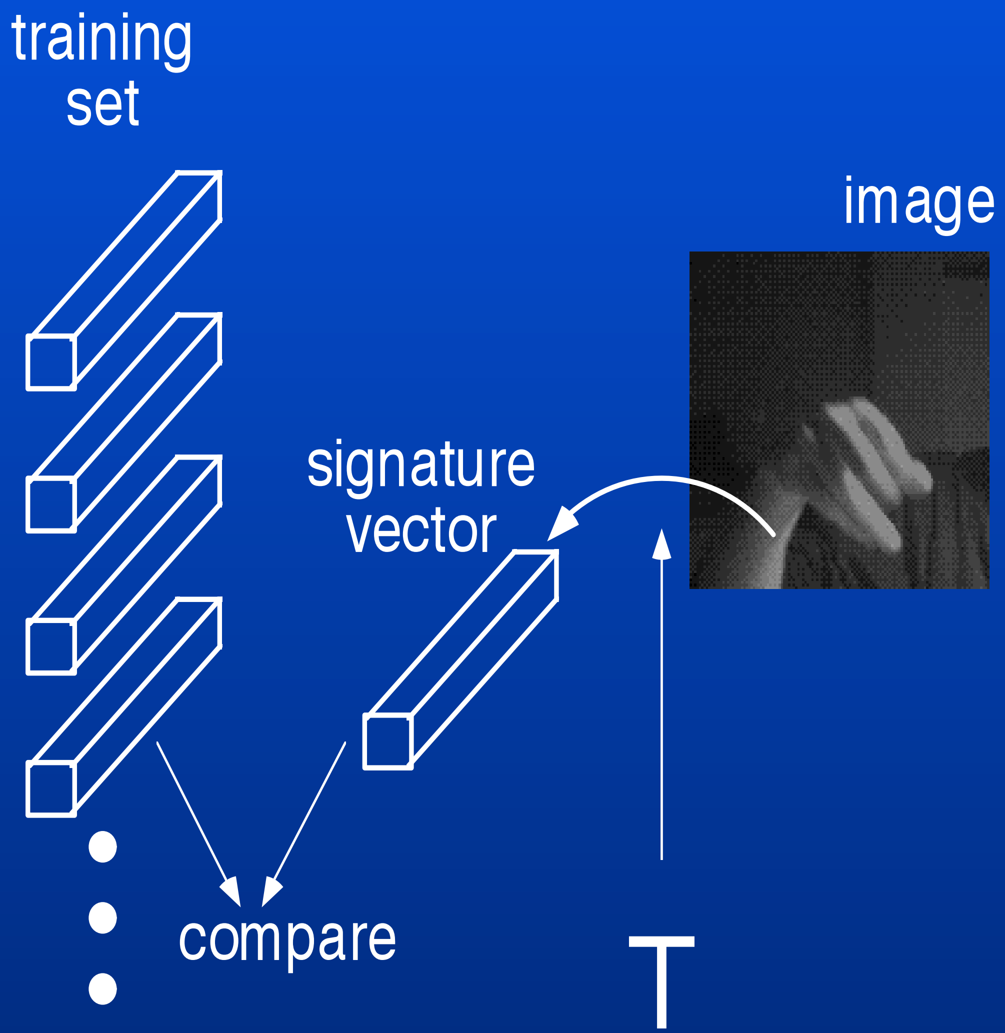
[Isard 1998]

Leaf tracking

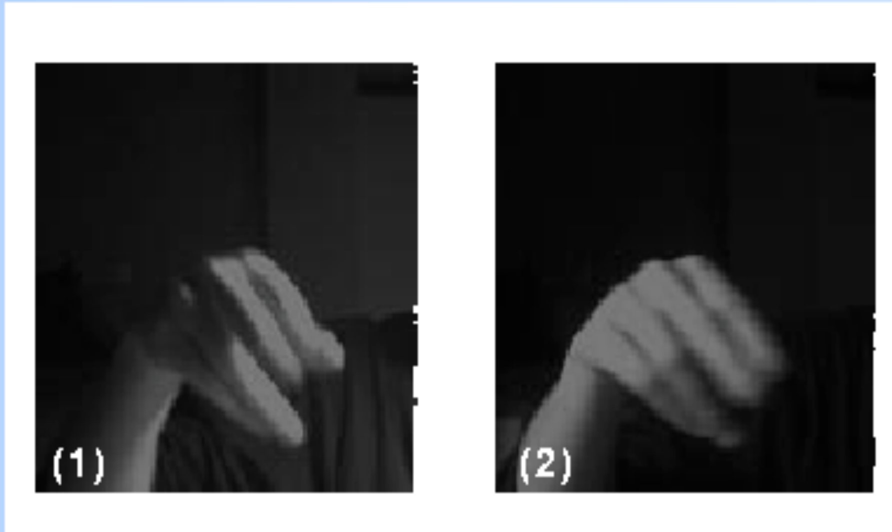
[Isard 1998]

Gesture recognition

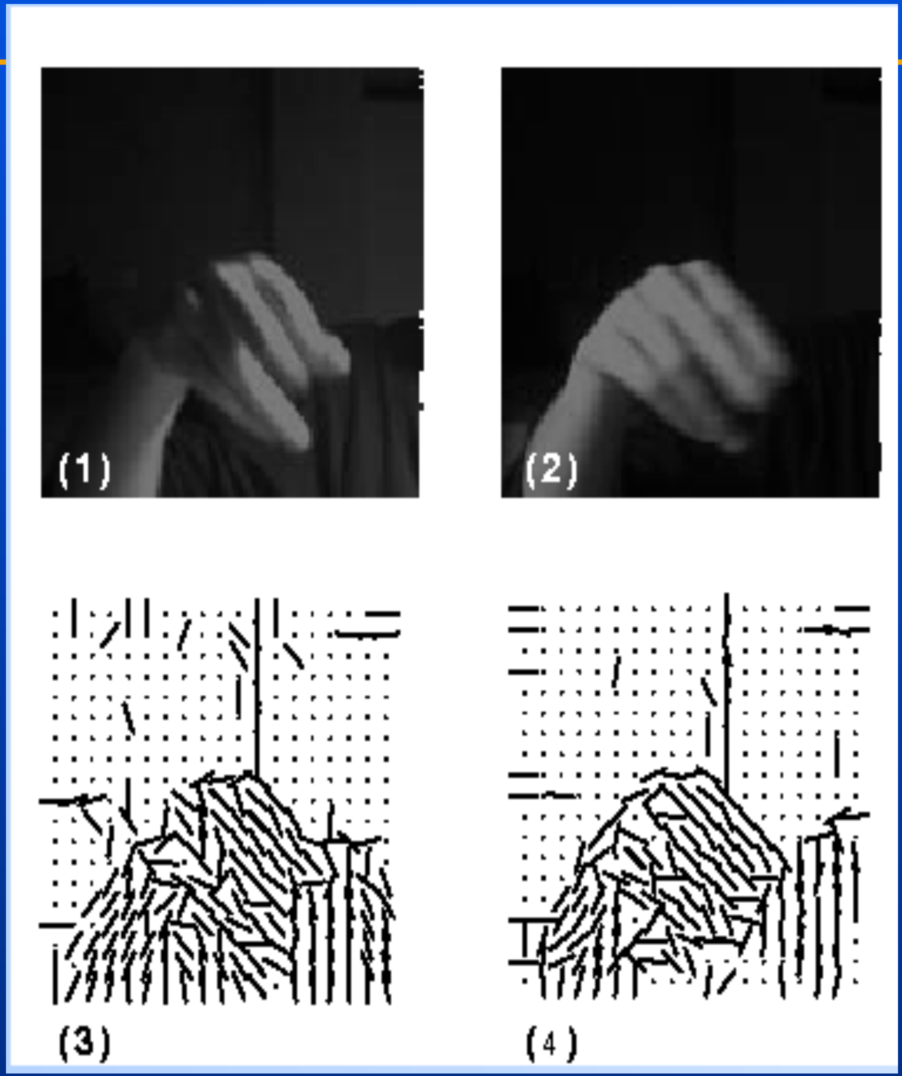
Real-time hand gesture recognition by orientation histograms



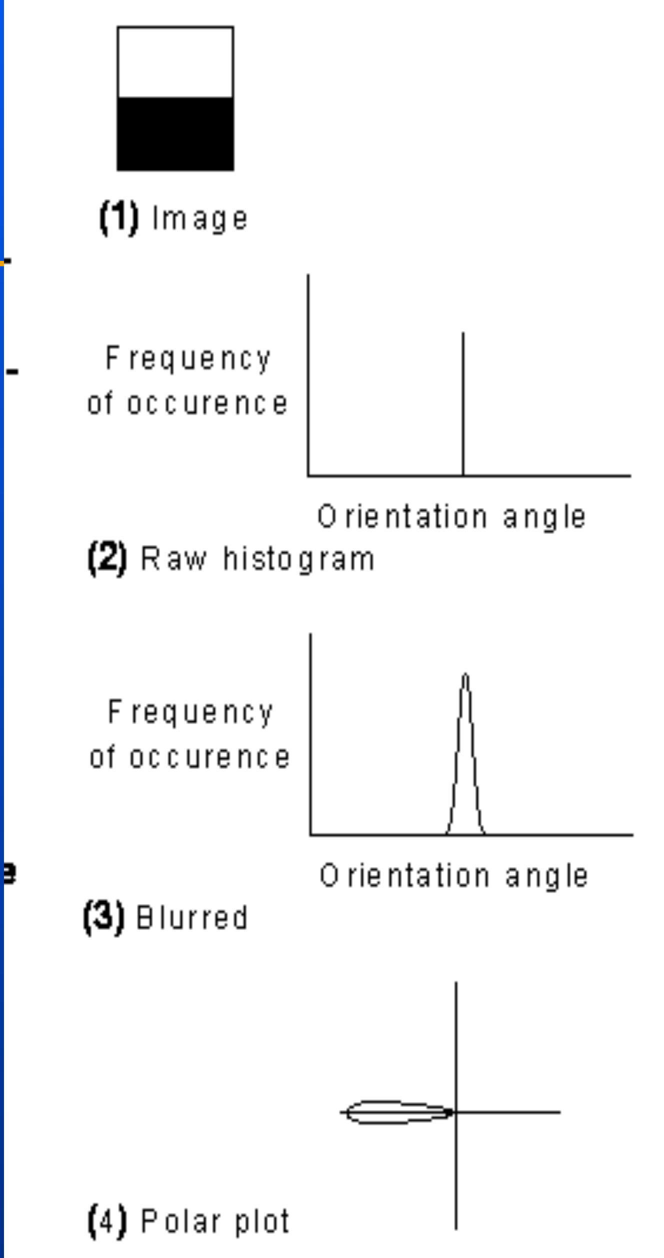
Orientation measurements (bottom) are more robust to lighting changes than are pixel intensities (top)



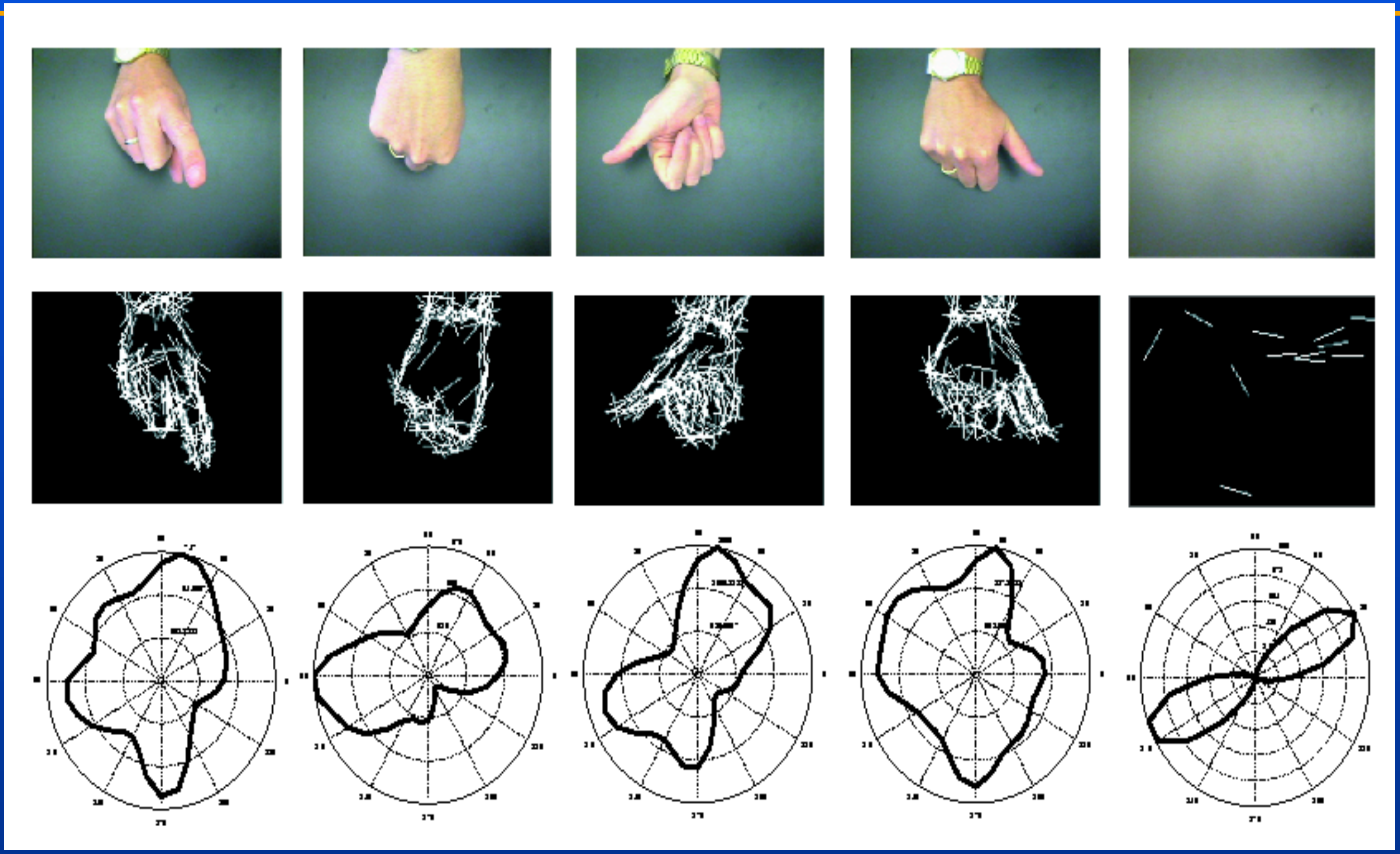
Orientation measurements (bottom) are more robust to lighting changes than are pixel intensities (top)



C Simple illustration of an orientation histogram. (1) An image of a horizontal edge has only one orientation at a sufficiently high contrast. (2) Thus the raw orientation histogram has counts at only one orientation value. (3) To allow neighboring orientations to sense each other, we blurred the raw histogram. (4) The same information, plotted in polar coordinates. We define the orientation to be the direction of the intensity gradient, plus 90 degrees.



Images, orientation images, and orientation histograms for training set



Test image, and distances from each of the training set orientation histograms (categorized correctly).

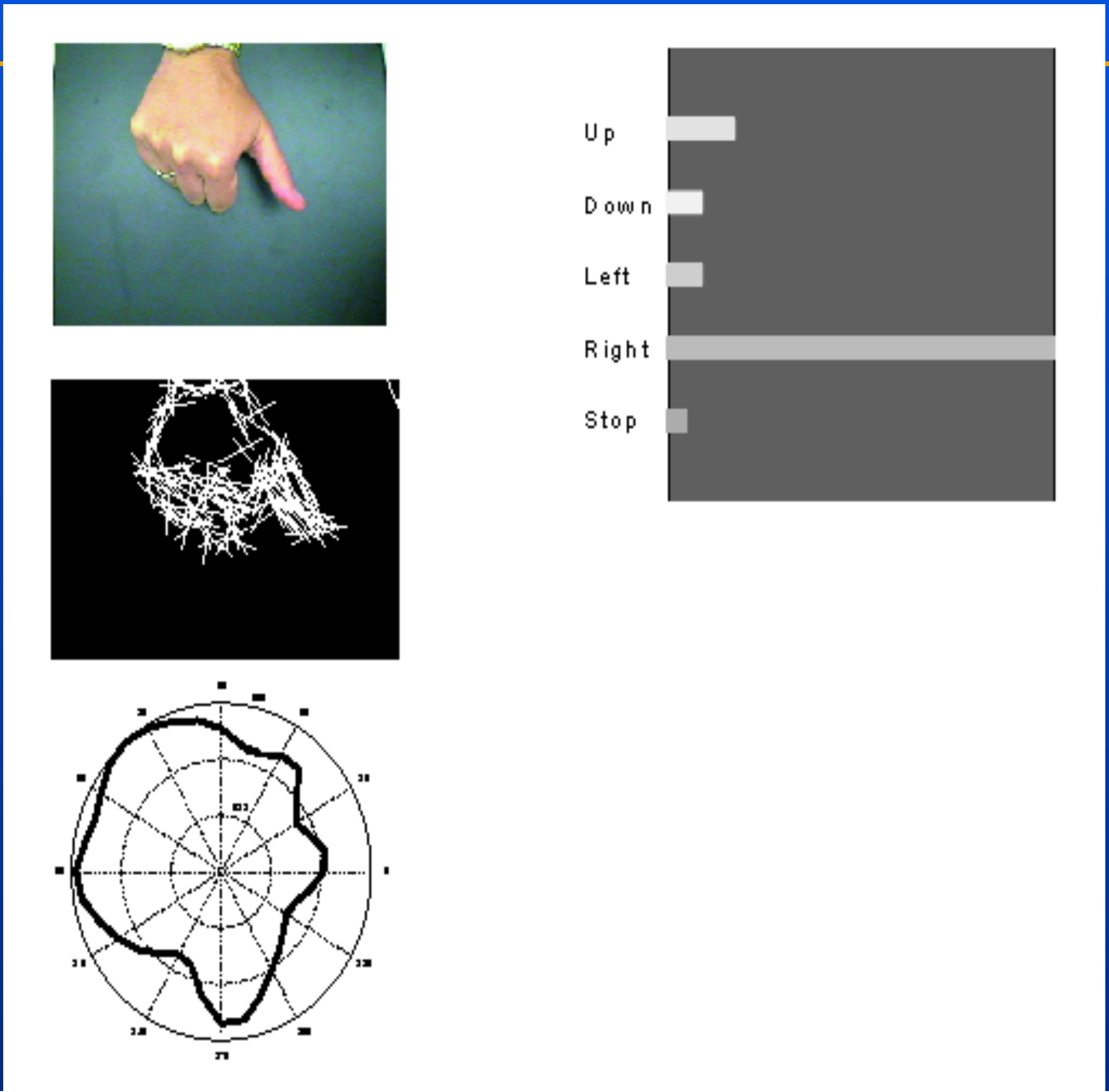


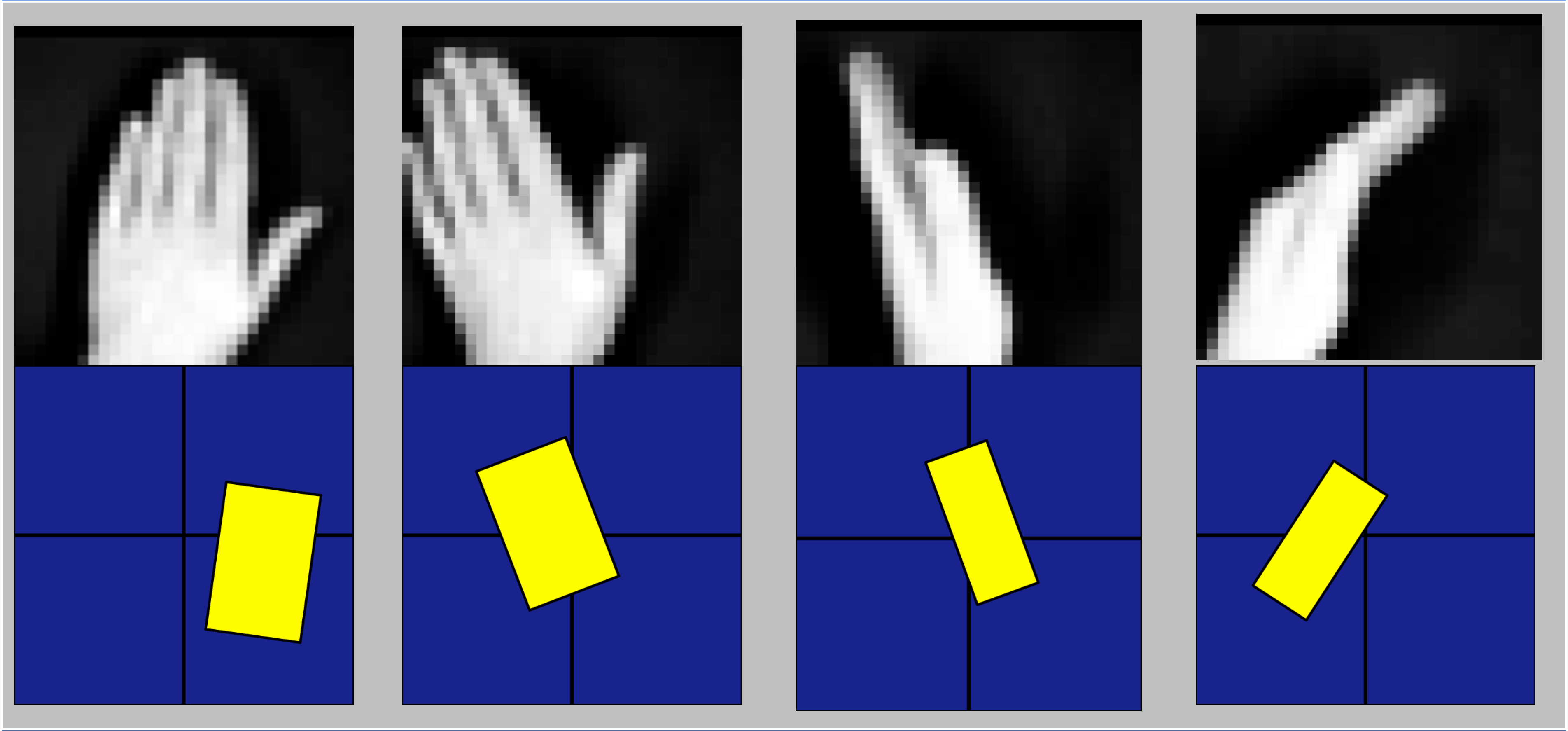
Image moments give a very coarse image summary.

$$M_{00} = \sum_x \sum_y I(x,y) \qquad M_{10} = \sum_x \sum_y x\, I(x,y)$$

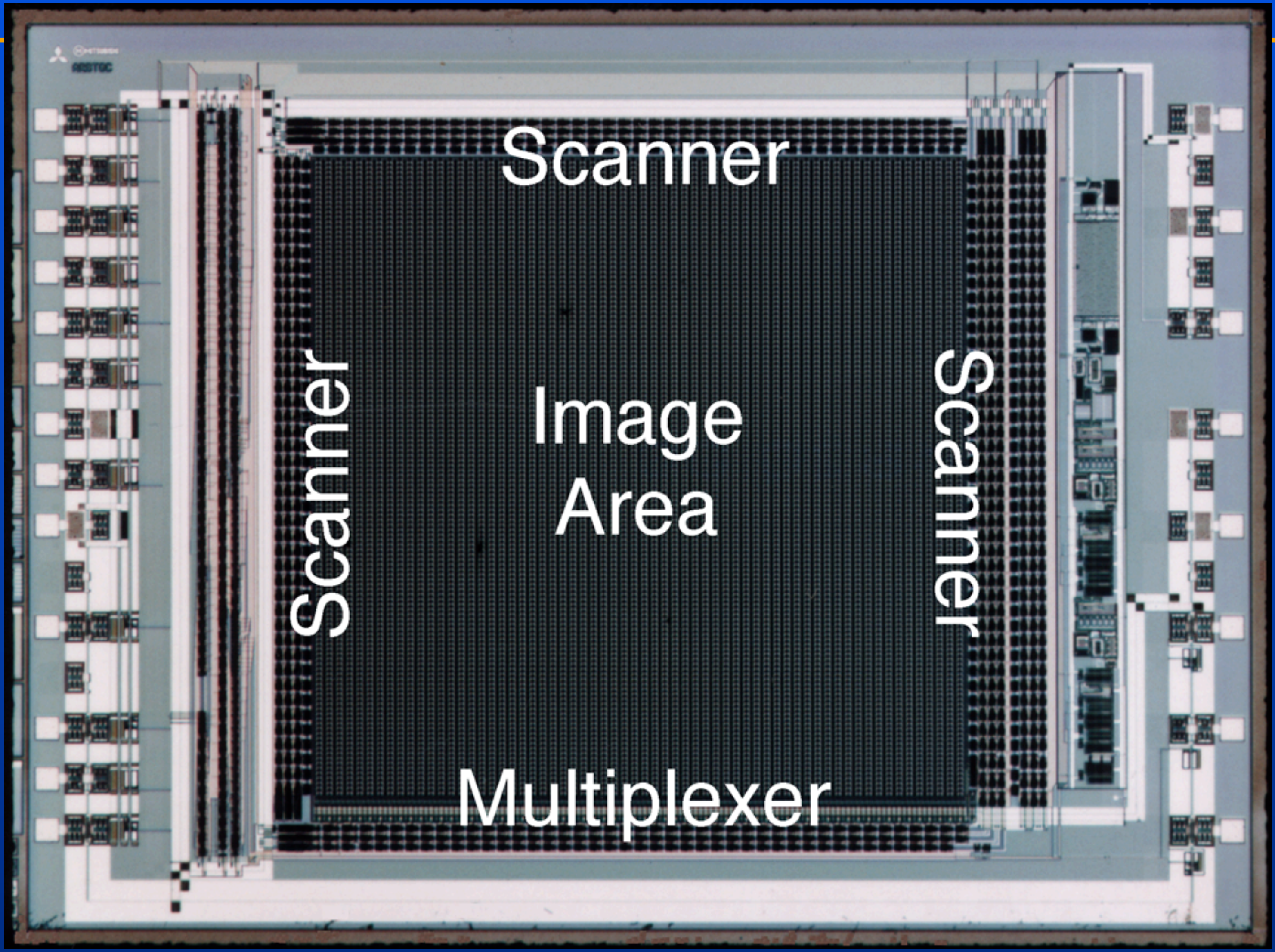
$$M_{01} = \sum_x \sum_y y\, I(x,y) \qquad M_{20} = \sum_x \sum_y x^2\, I(x,y)$$

$$M_{11} = \sum_x \sum_y xy\, I(x,y) \qquad M_{02} = \sum_x \sum_y y^2\, I(x,y)$$

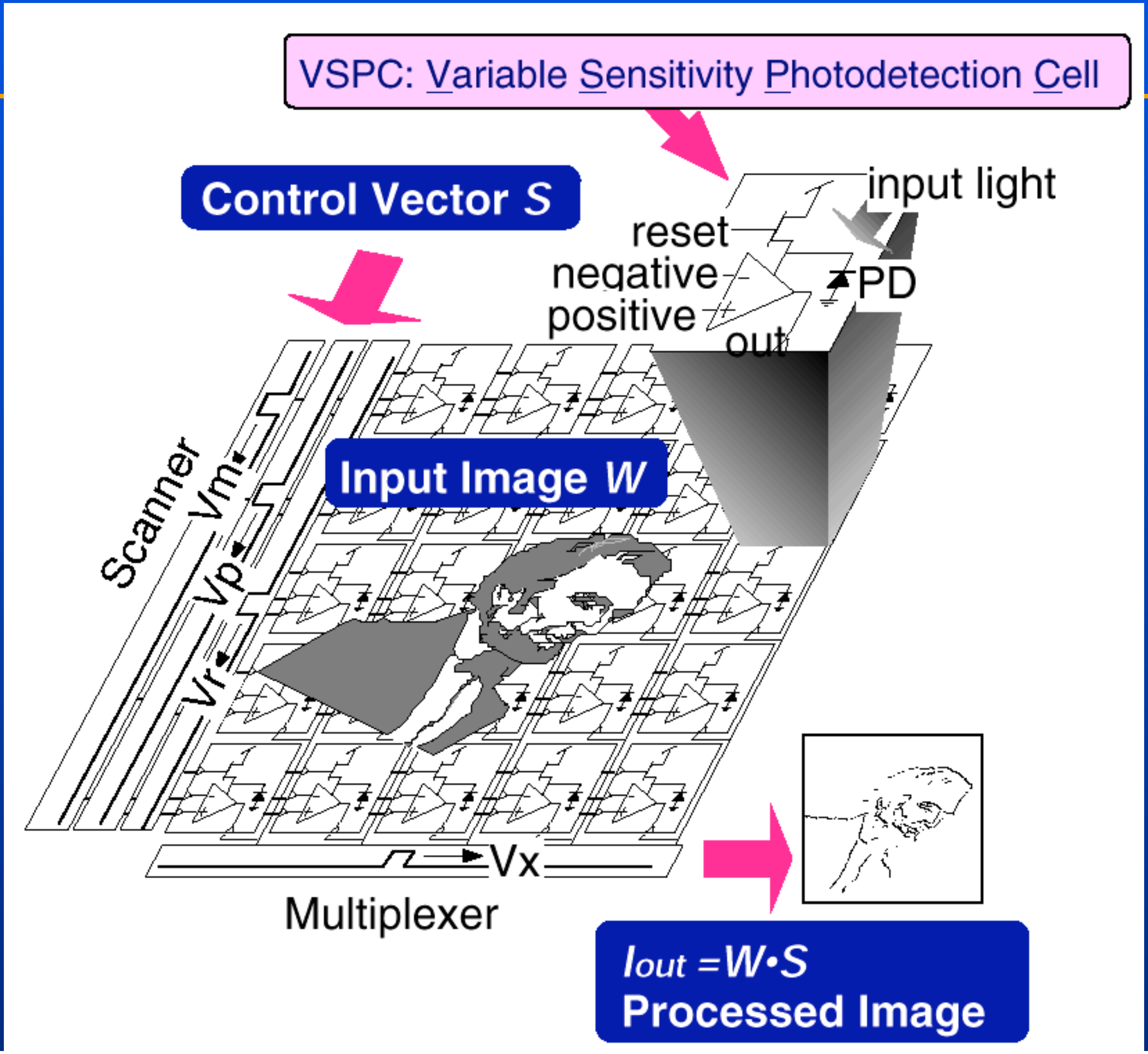
Hand images and equivalent rectangles having the same image moments



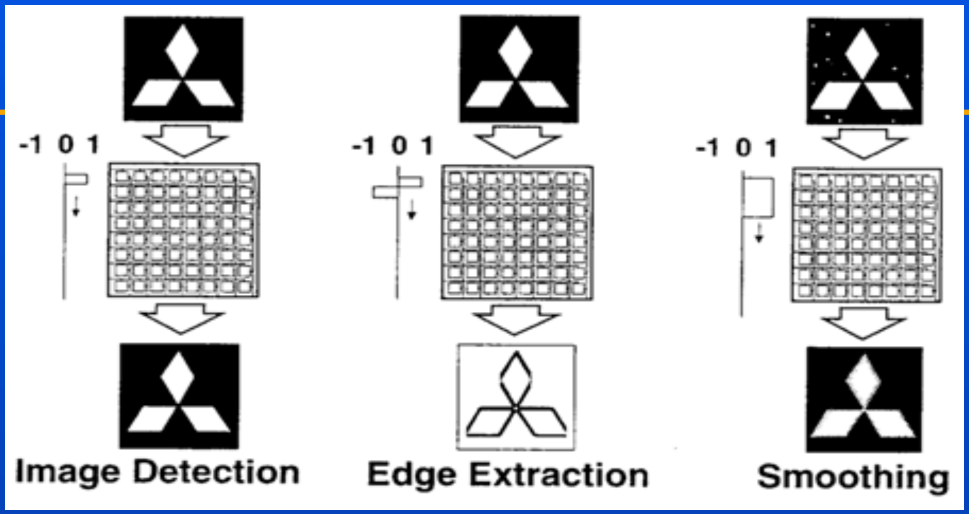
Artificial Retina chip for detection and low-level image processing.

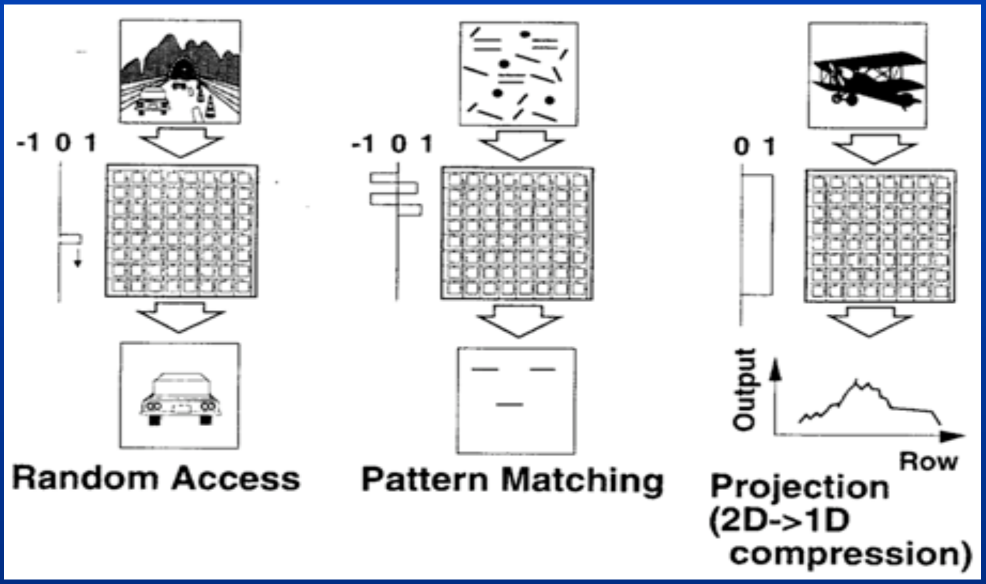


Artificial Retina chip



Artificial Retina functions



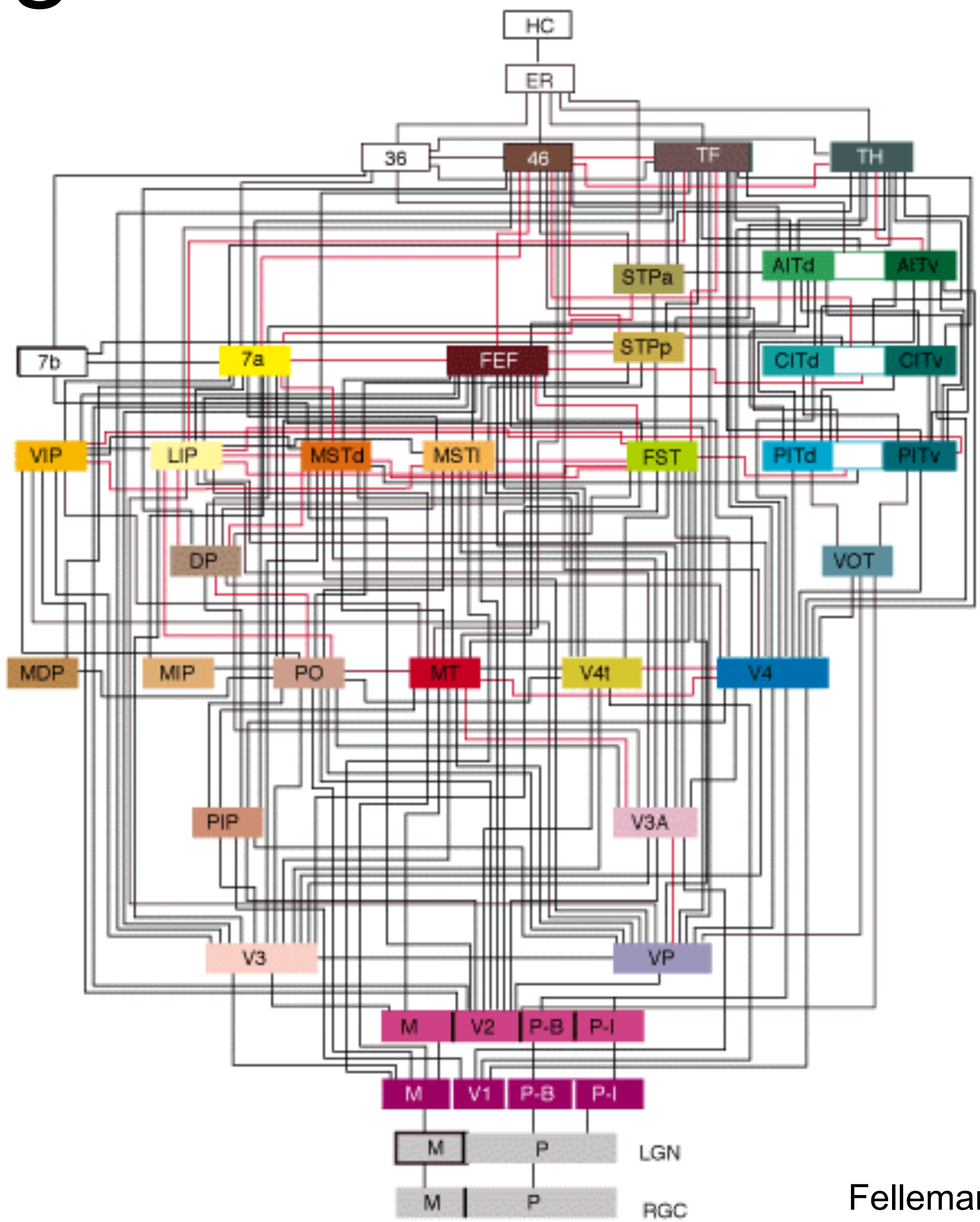


Model-based hand tracking with texture, shading and self-occlusions

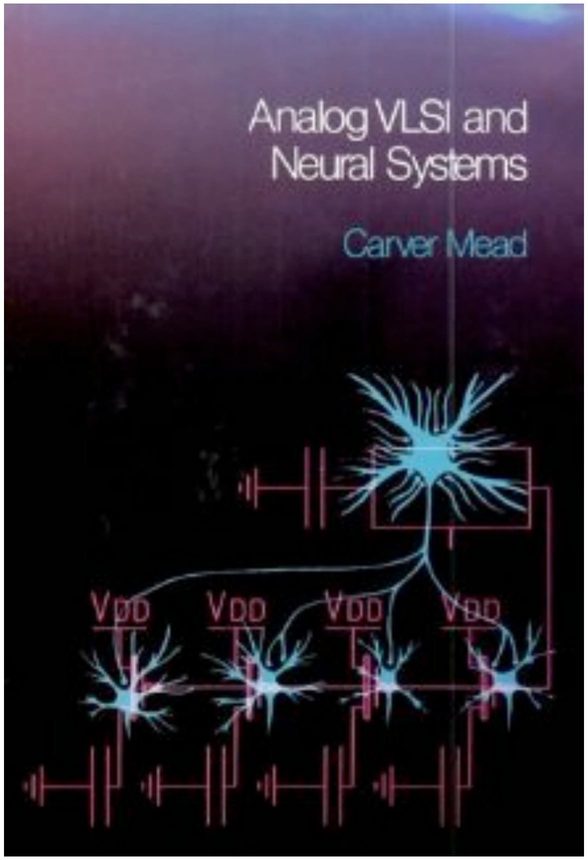
De la Gorce, M., Paragos, N. and Fleet, D.J.

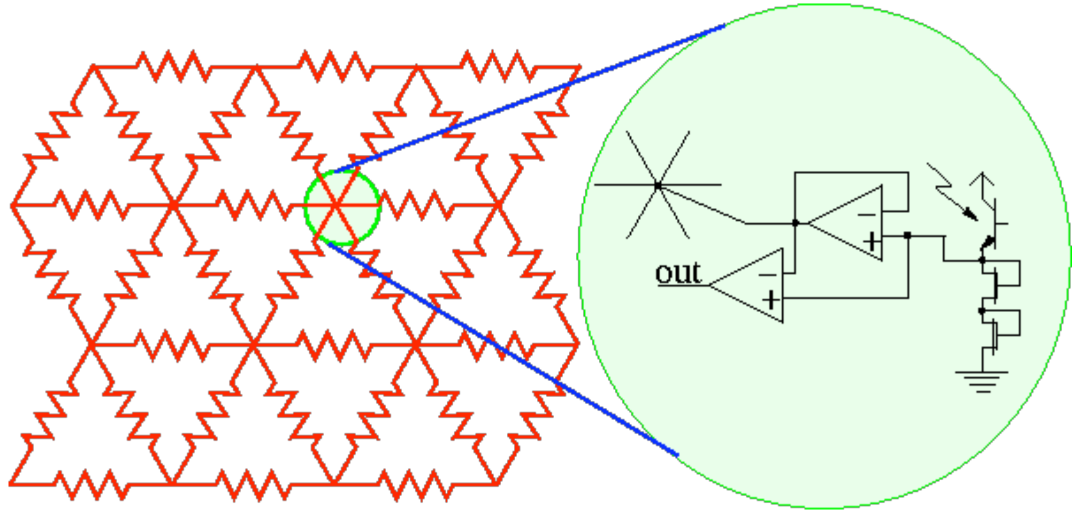
Biologically inspired computer vision

Diagram of the visual system



Felleman and Van Essen, 1991





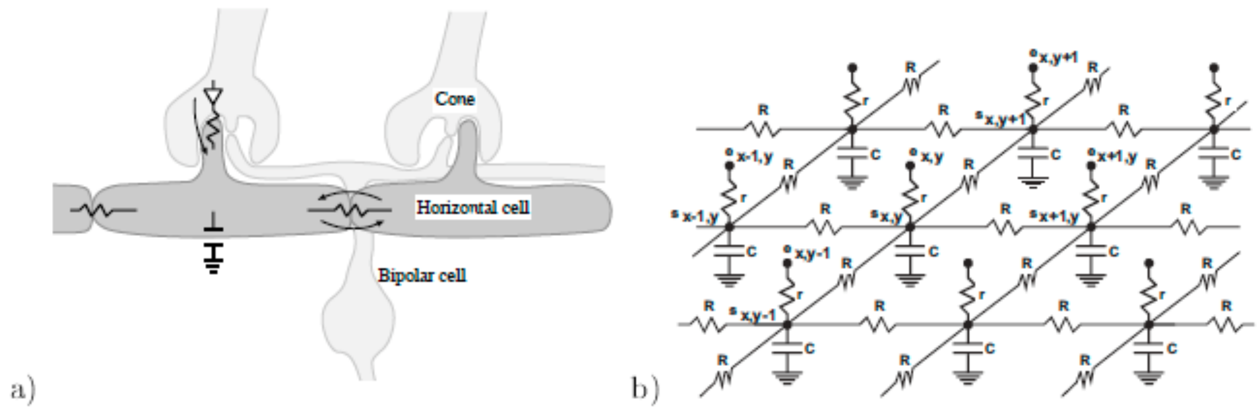
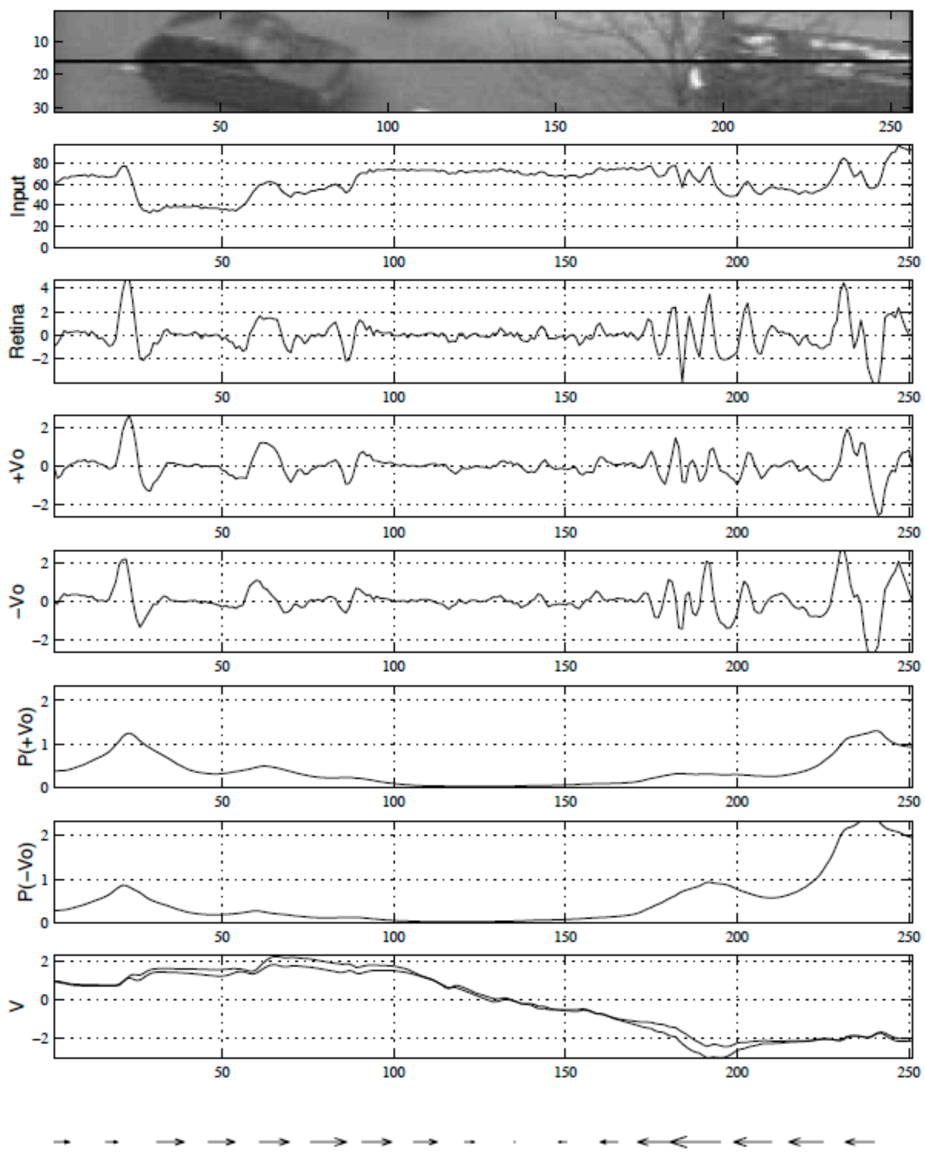
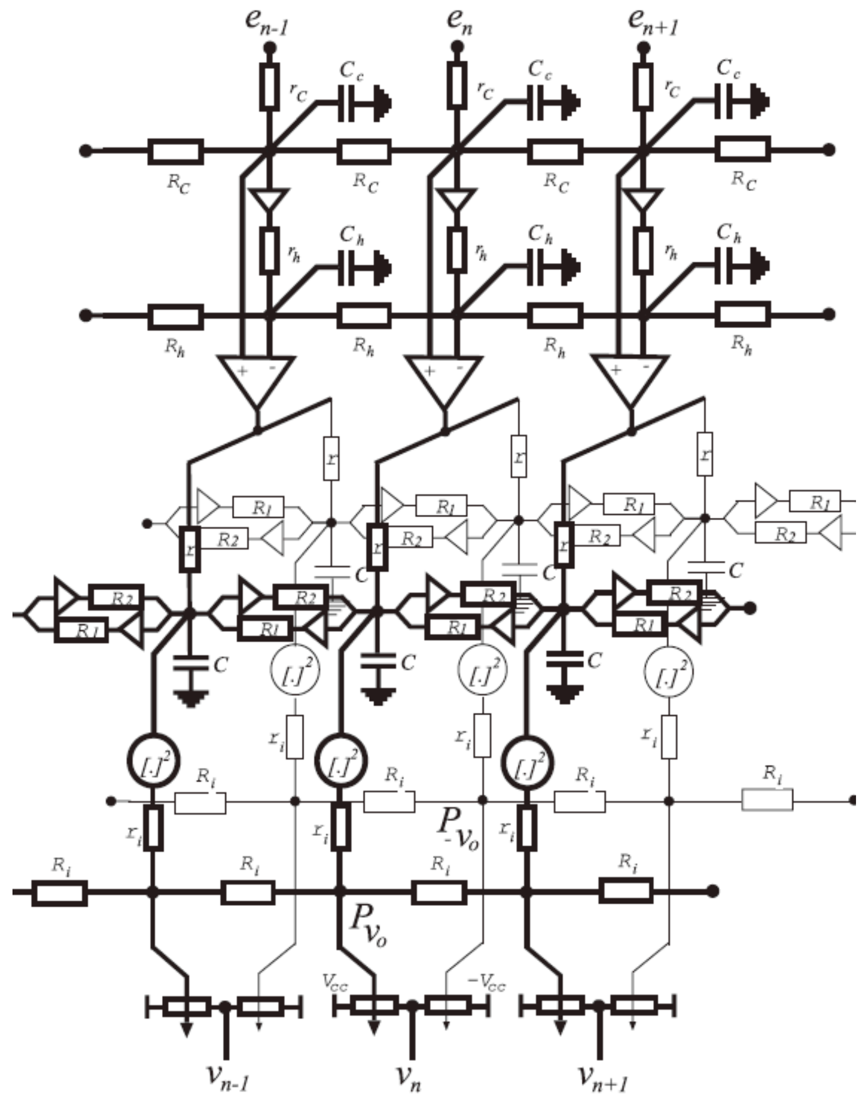


Figure 1.2: *a) Schema of the horizontal cell layer of the retina. b) RC analog network.*

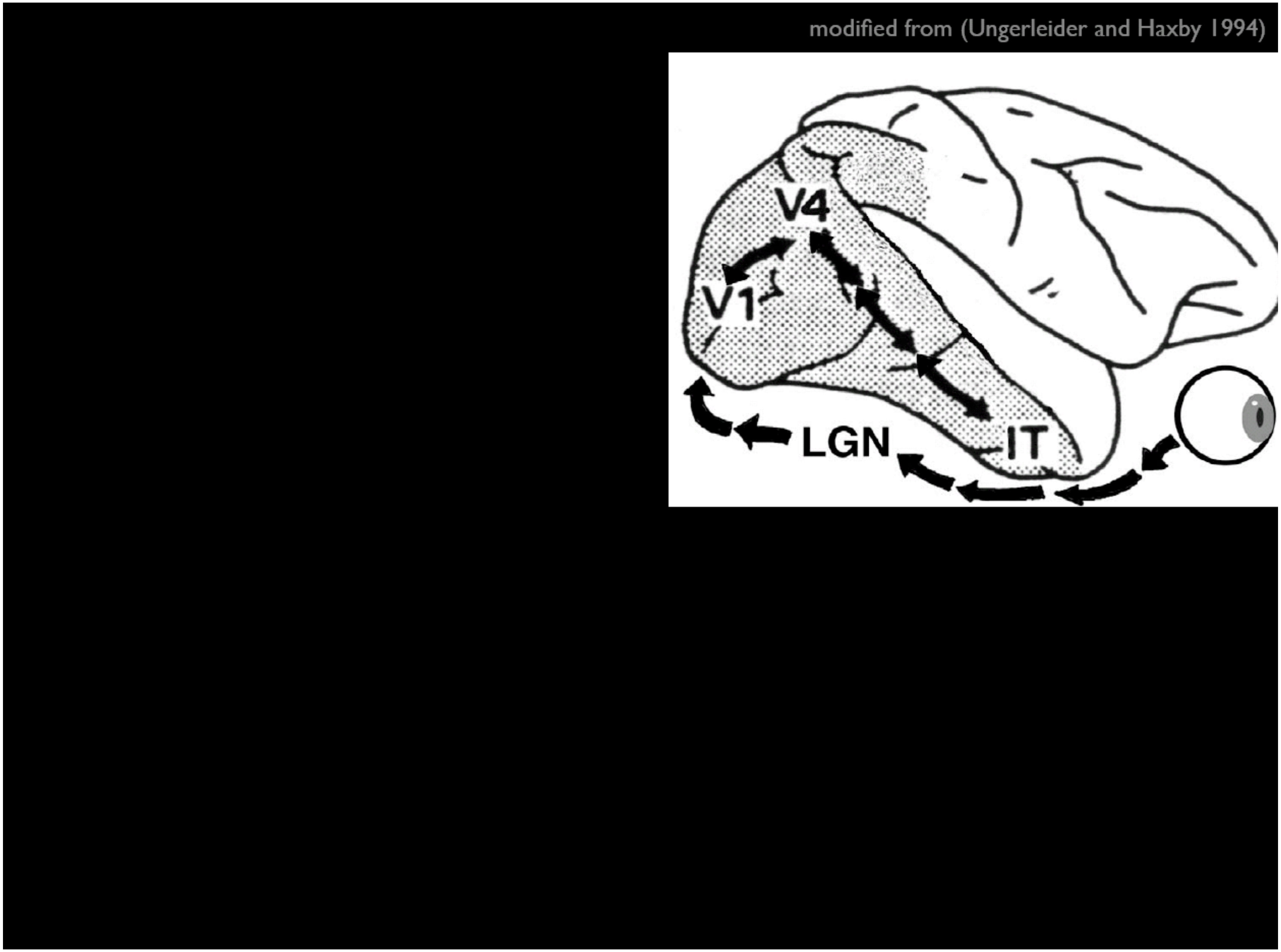
http://www.iee.et.tu-dresden.de/iee/analog/papers/mirror/visionchips/vision_chips/mahowald_mead.html



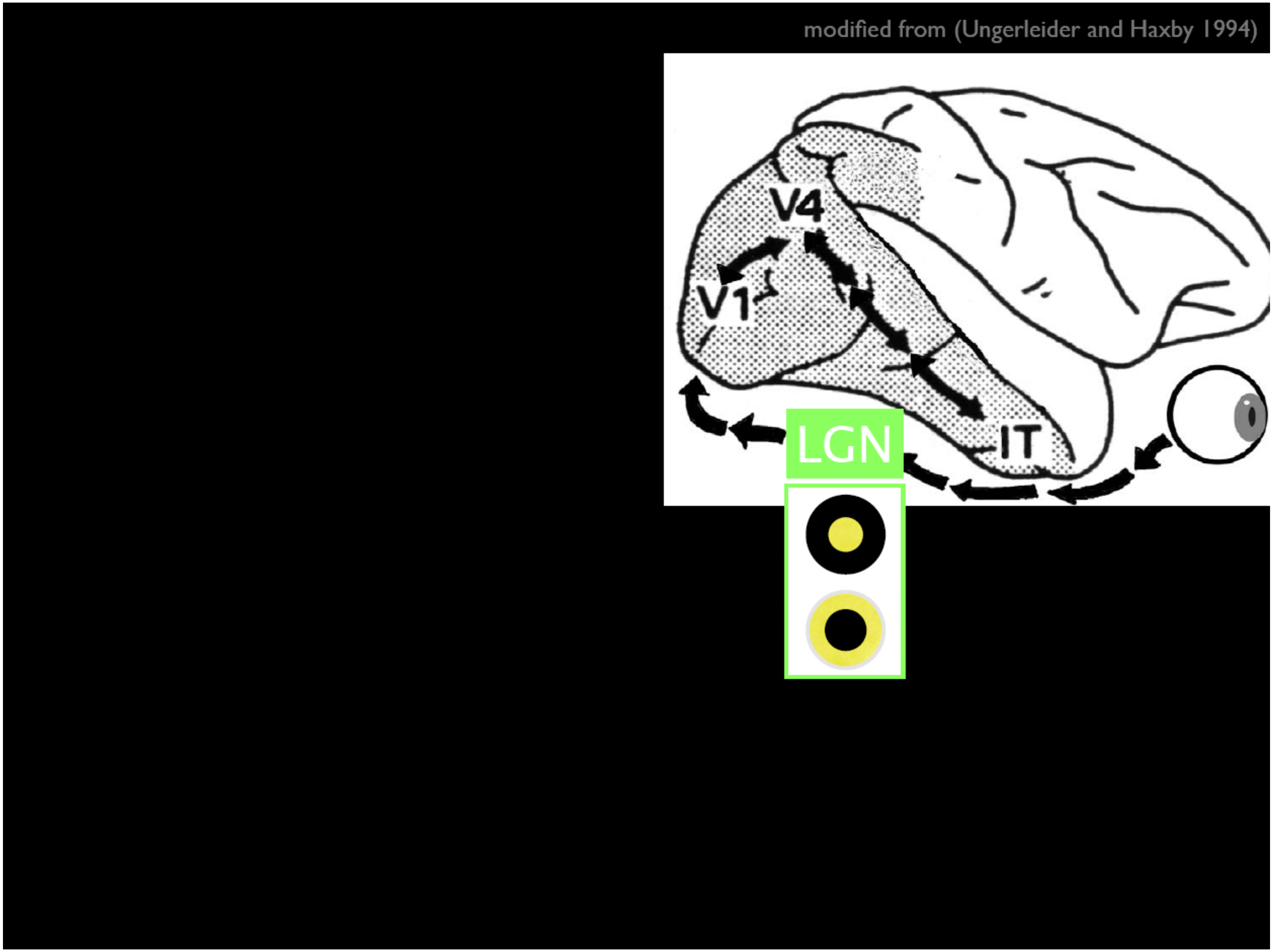
Input
Retina
+Vo
−Vo
P(+Vo)
P(−Vo)
V



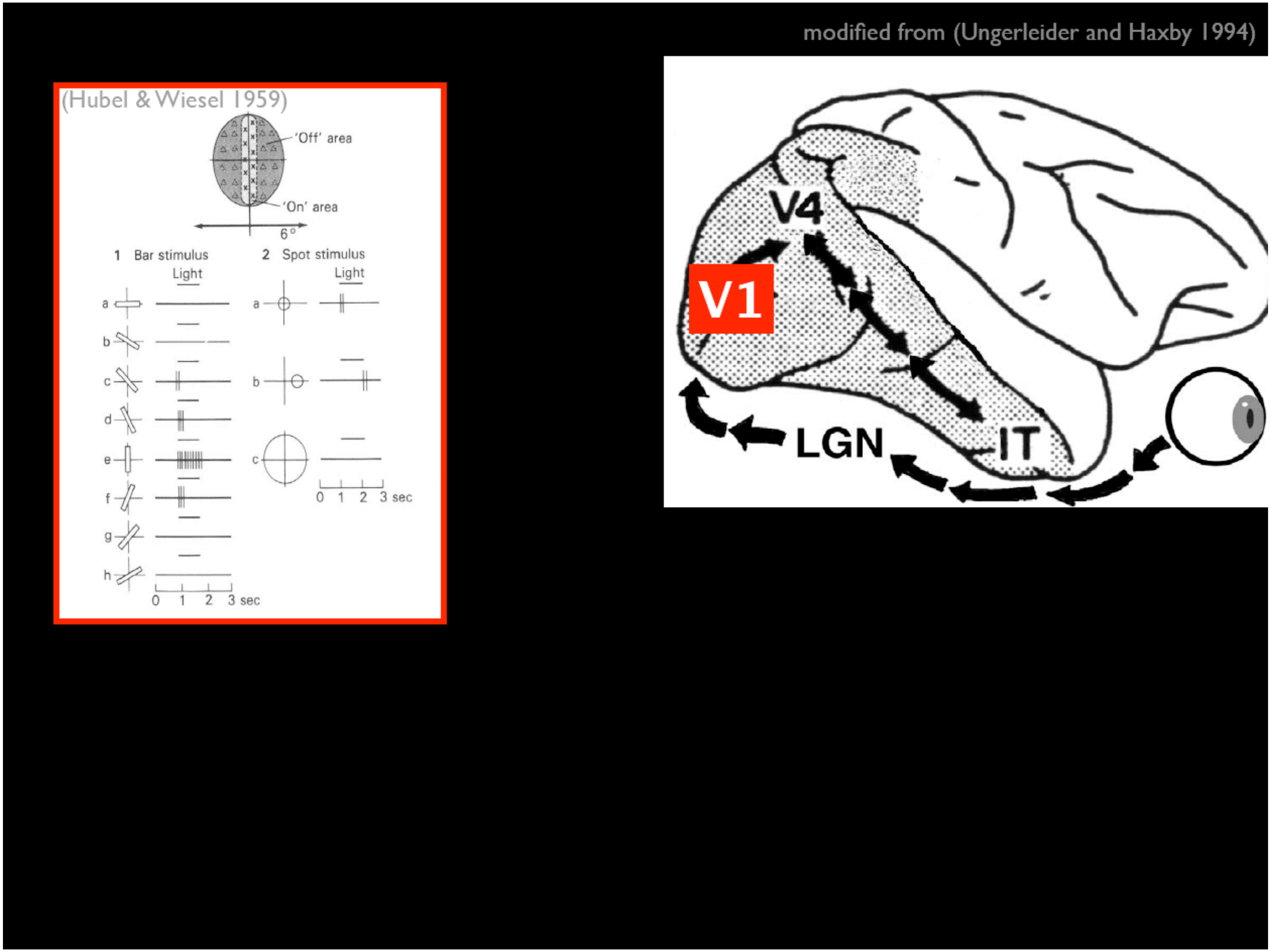
e_{n-1}
e_n
e_{n+1}
C_c
r_C
R_C
C_h
r_h
R_h
R_1
R_2
r
C
$[.]^2$
r_i
R_i
P_{-V_o}
P_{V_o}
V_{cc}
$-V_{cc}$
v_{n-1}
v_n
v_{n+1}



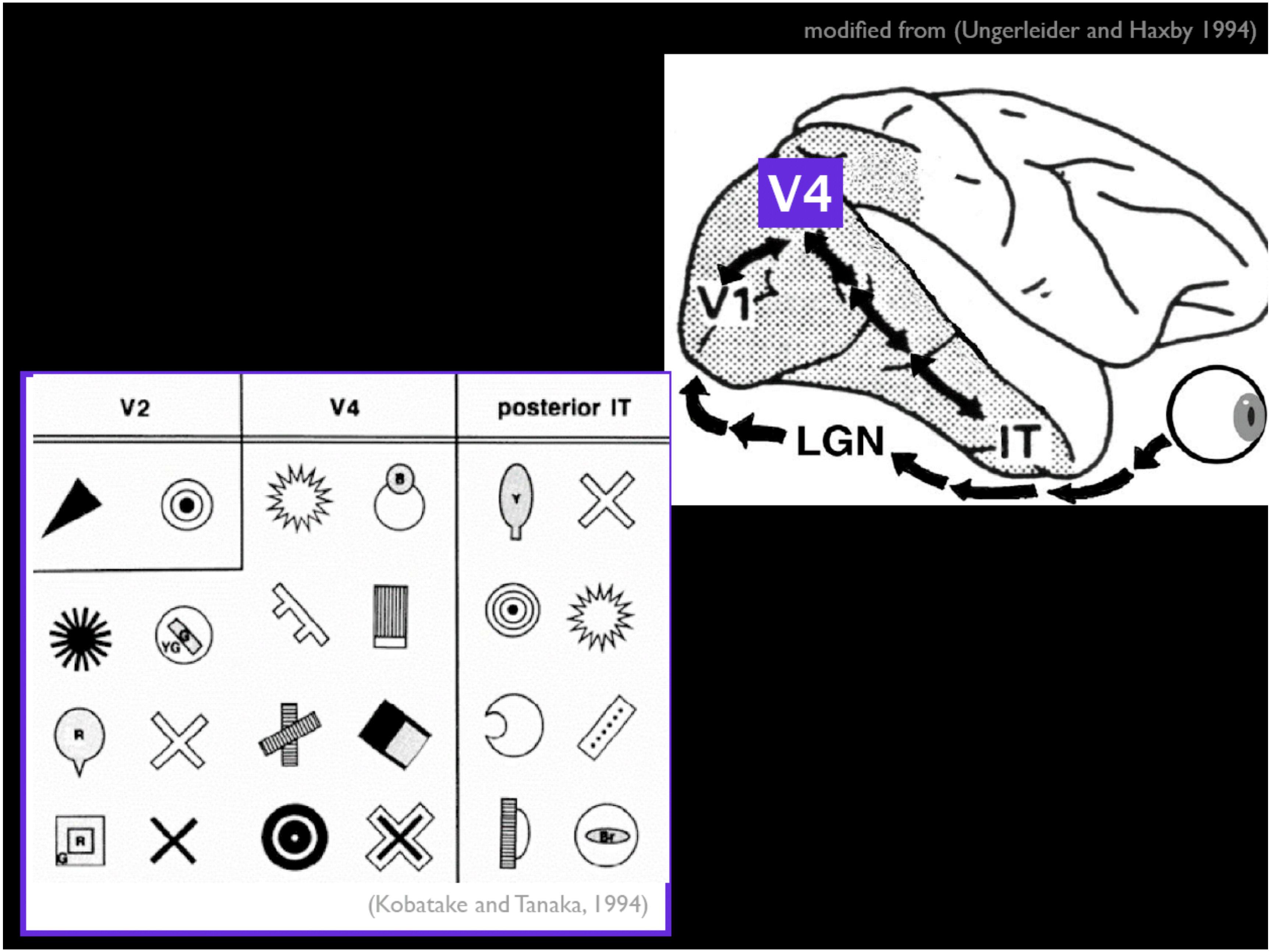
Modified by T. Serre from Ungerleider and Haxby, and then shamelessly copied by me.



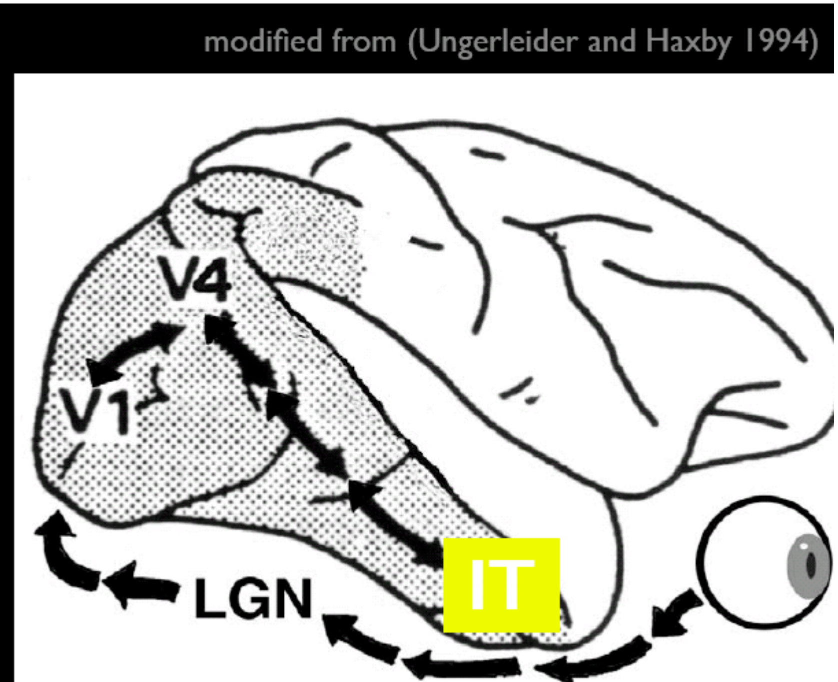
Modified by T. Serre from Ungerleider and Haxby, and then copied by me.

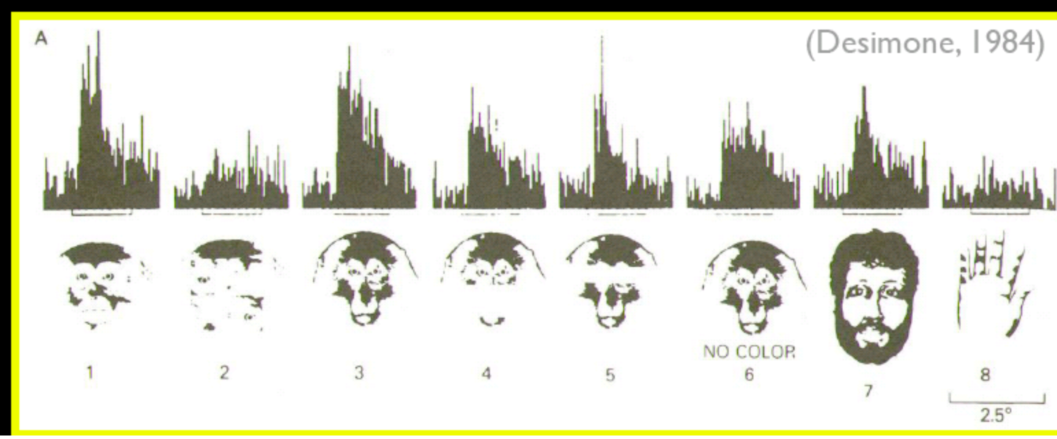


Modified by T. Serre from Ungerleider and Haxby, and then copied by me.



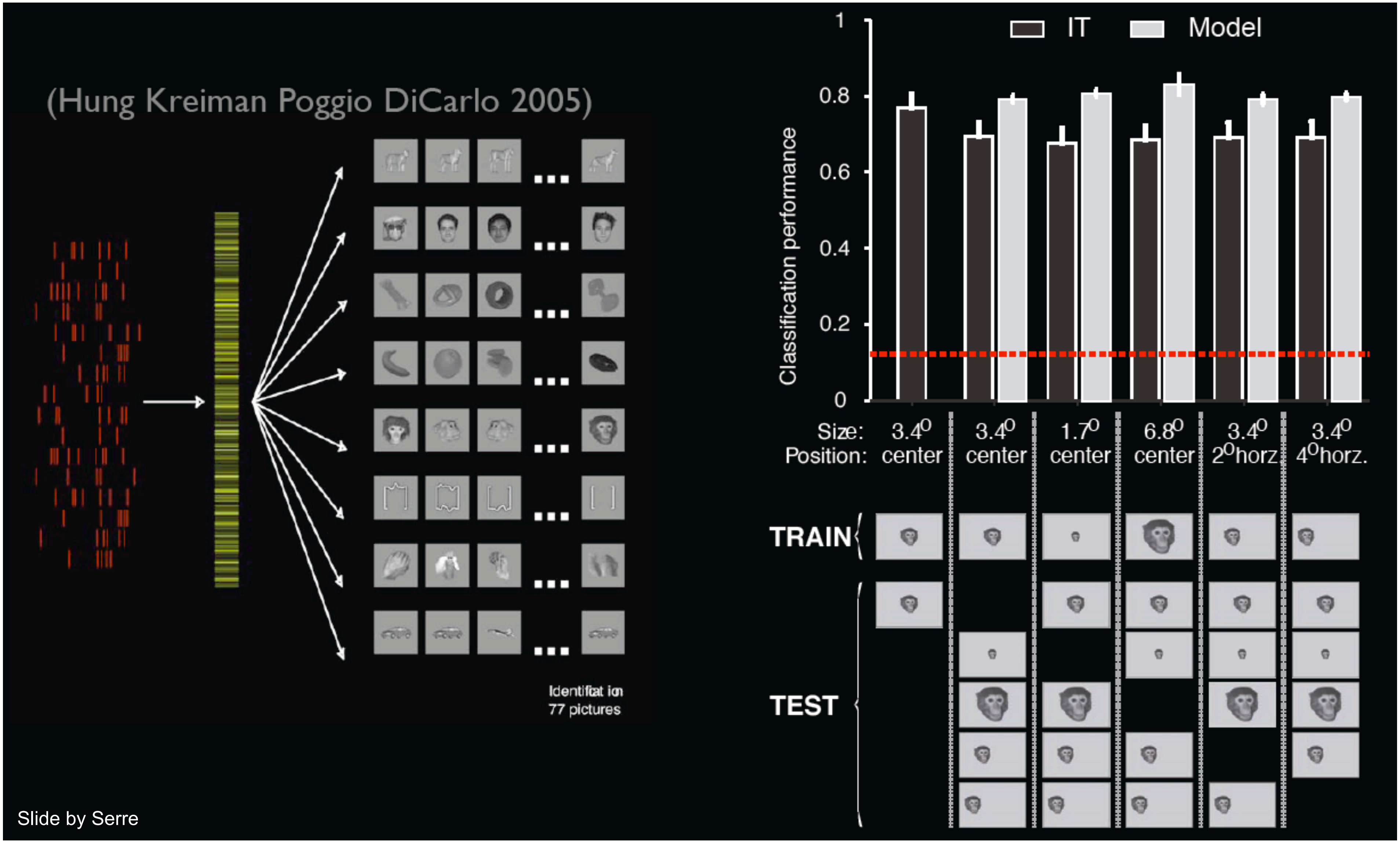
Modified by T. Serre from Ungerleider and Haxby, and then copied by me.





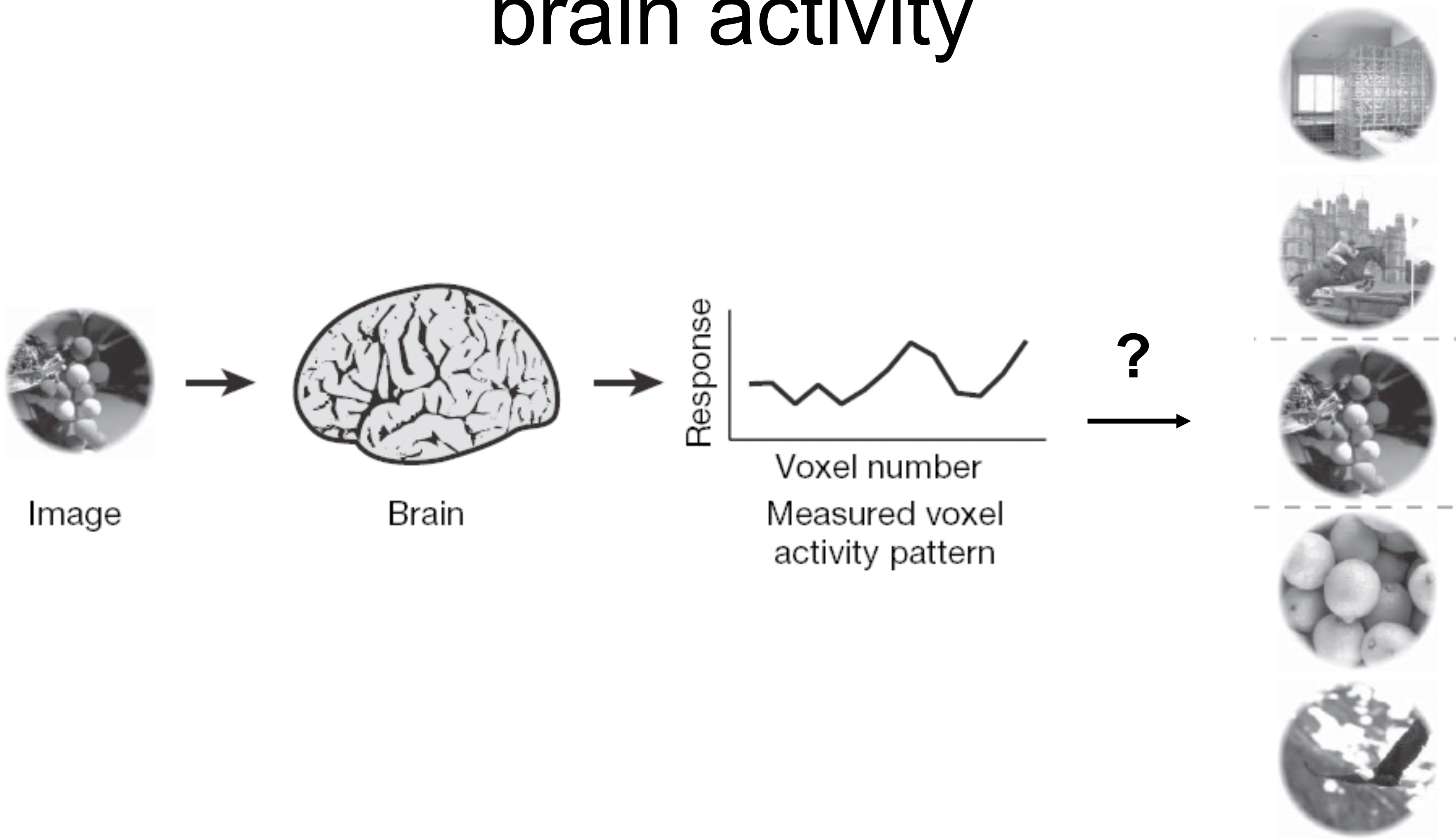
Modified by T. Serre from Ungerleider and Haxby, and then copied by me.

IT readout

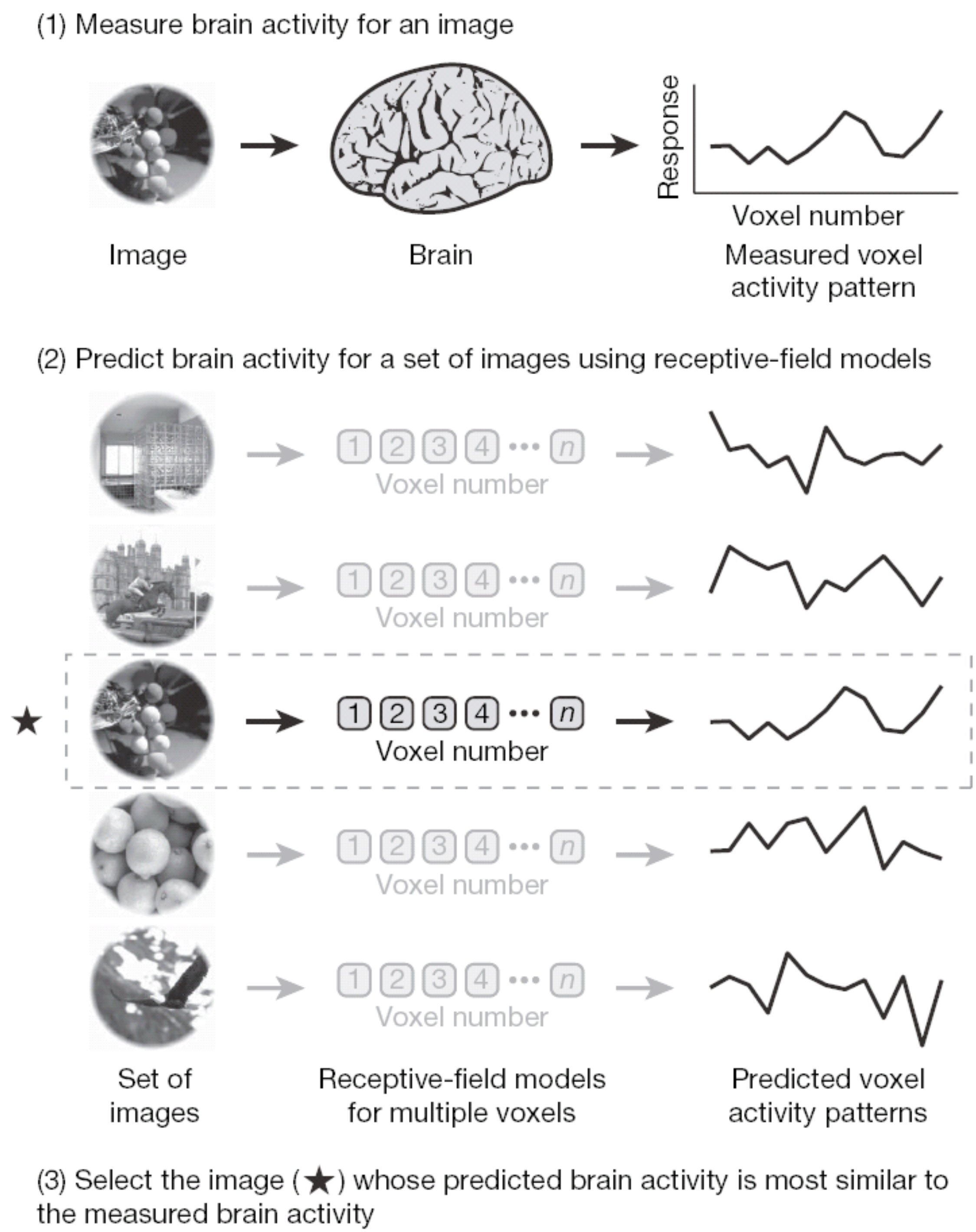


Slide by Serre

Identifying natural images from human brain activity



Kay, K.N., Naselaris, T., Prenger, R.J., & Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature, 452, 352-355.

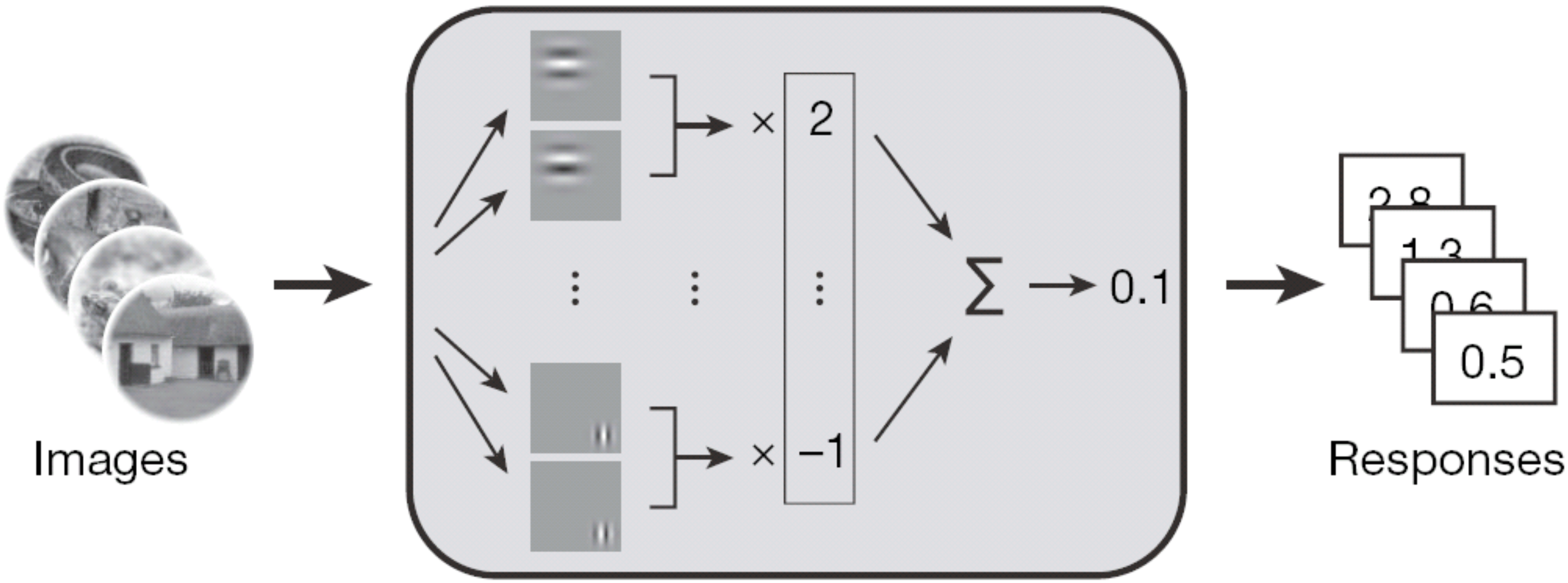


(1) Measure brain activity for an image
Response
Voxel number
Image
Brain
Measured voxel activity pattern
(2) Predict brain activity for a set of images using receptive-field models
1 2 3 4 ••• n
Voxel number
1 2 3 4 ••• n
Voxel number
★
1 2 3 4 ••• n
Voxel number
1 2 3 4 ••• n
Voxel number
1 2 3 4 ••• n
Voxel number
Set of images
Receptive-field models for multiple voxels
Predicted voxel activity patterns
(3) Select the image (★) whose predicted brain activity is most similar to the measured brain activity

Voxel Activity Model

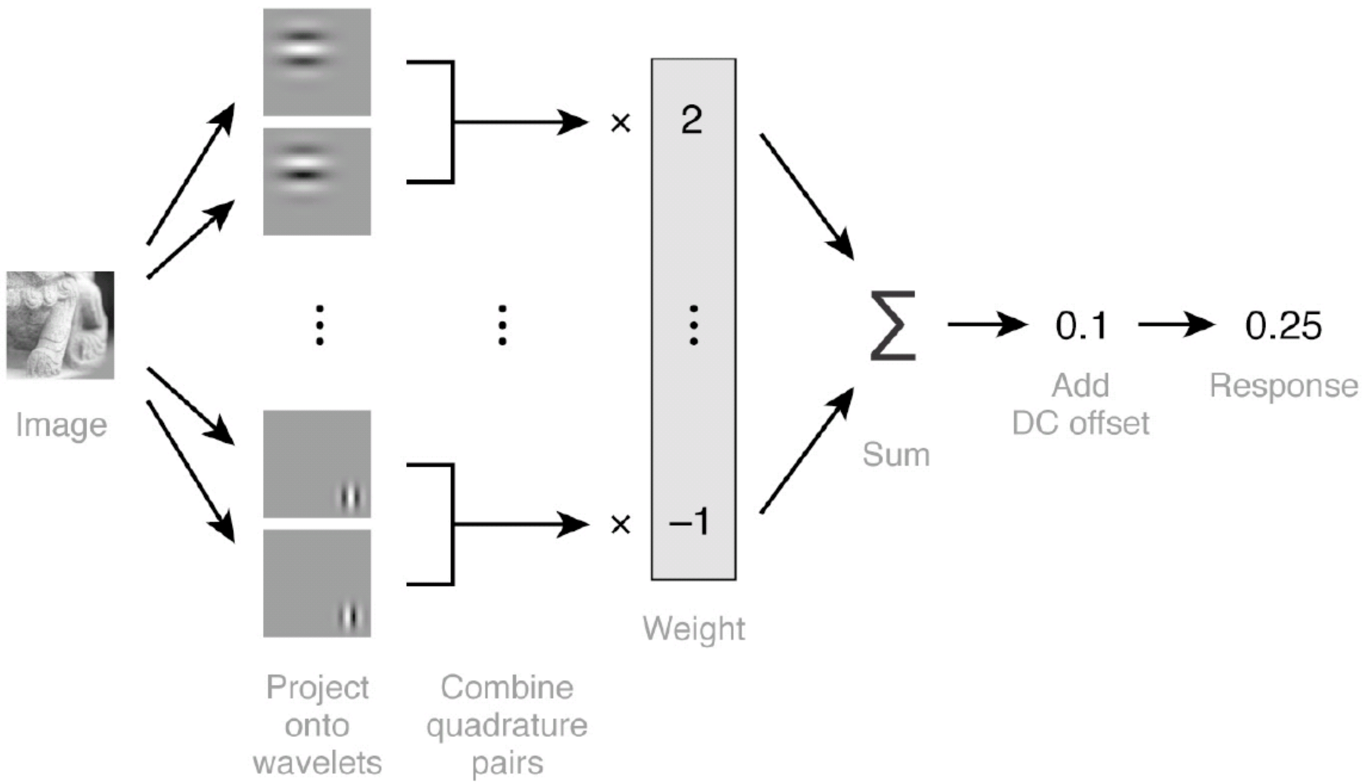
Goal: to predict the image seen by the observer out of a large collection of possible images. And to do this for new images: this requires predicting fMRI activity for unseen images.

Estimate a receptive-field model for each voxel



Receptive-field model for one voxel

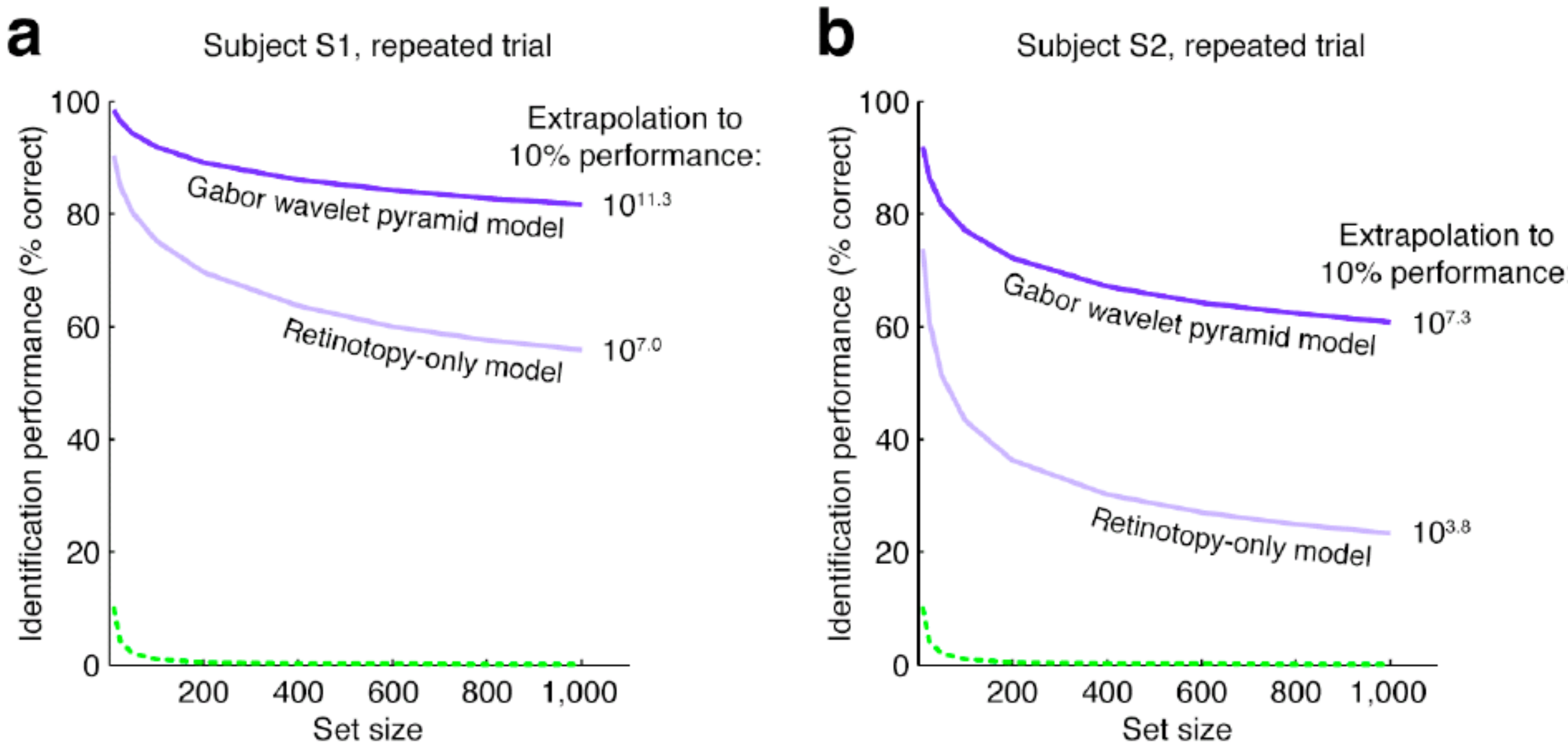
Kay, K.N., Naselaris, T., Prenger, R.J., & Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature, 452, 352-355.



Supplementary Figure 3. Gabor wavelet pyramid model. Each image is projected onto the individual Gabor wavelets comprising the Gabor wavelet pyramid (see Supplementary Fig. 2). The projections for each quadrature pair of wavelets are squared, summed, and square-rooted, yielding a measure of contrast energy. The contrast energies for different quadrature wavelet pairs are weighted and then summed. Finally, a DC offset is added. The weights are determined by gradient descent with early stopping (see Supplementary Methods 6).

Kay, K.N., Naselaris, T., Prenger, R.J., & Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature, 452, 352-355.

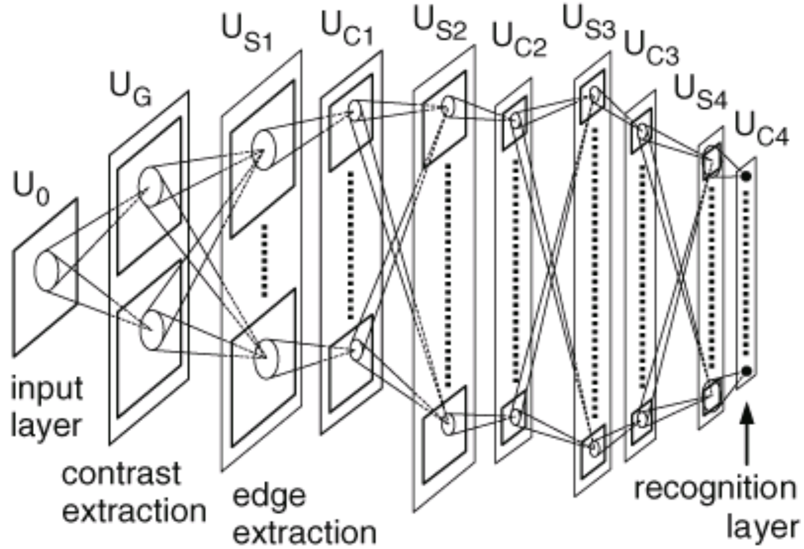
Performance



Kay, K.N., Naselaris, T., Prenger, R.J., & Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature, 452, 352-355.

Neocognitron

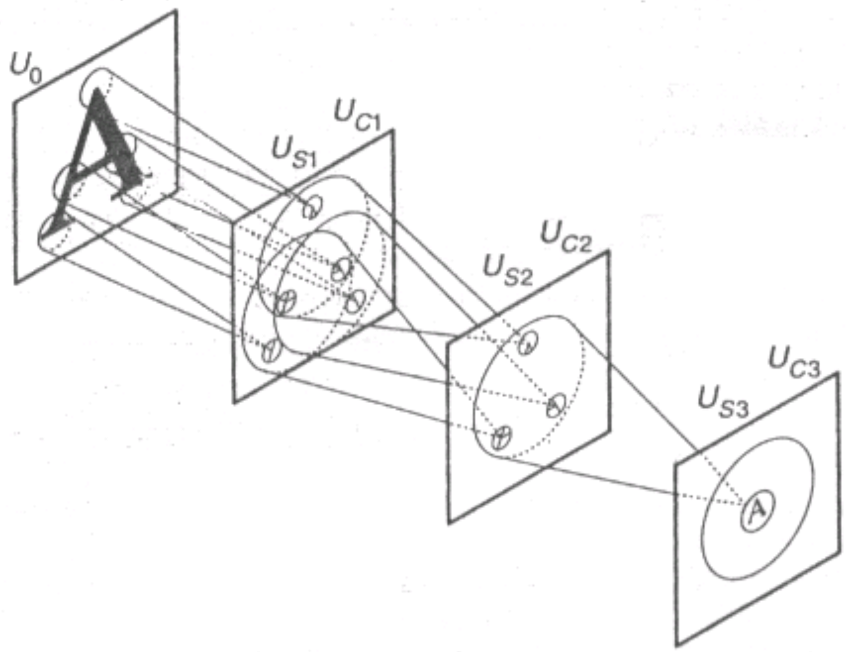
Fukushima (1980). Hierarchical multilayered neural network

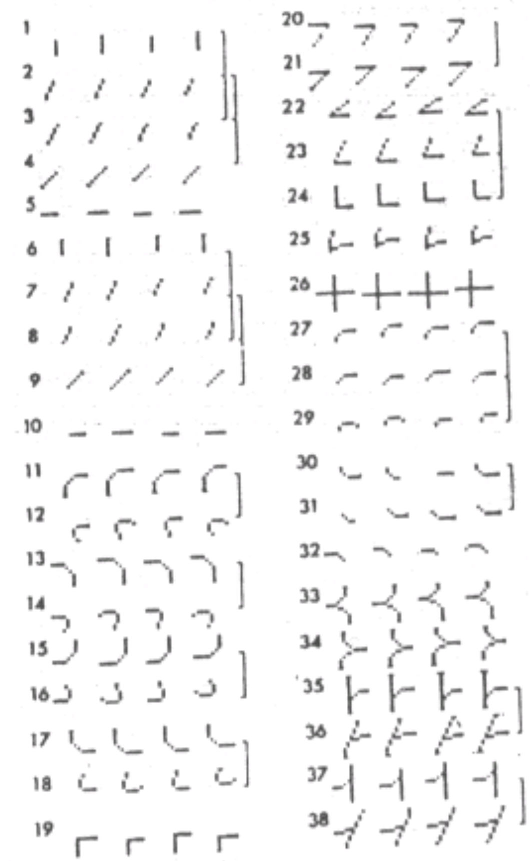


S-cells work as feature-extracting cells. They resemble simple cells of the primary visual cortex in their response.

C-cells, which resembles complex cells in the visual cortex, are inserted in the network to allow for positional errors in the features of the stimulus. The input connections of C-cells, which come from S-cells of the preceding layer, are fixed and invariable. Each C-cell receives excitatory input connections from a group of S-cells that extract the same feature, but from slightly different positions. The C-cell responds if at least one of these S-cells yield an output.

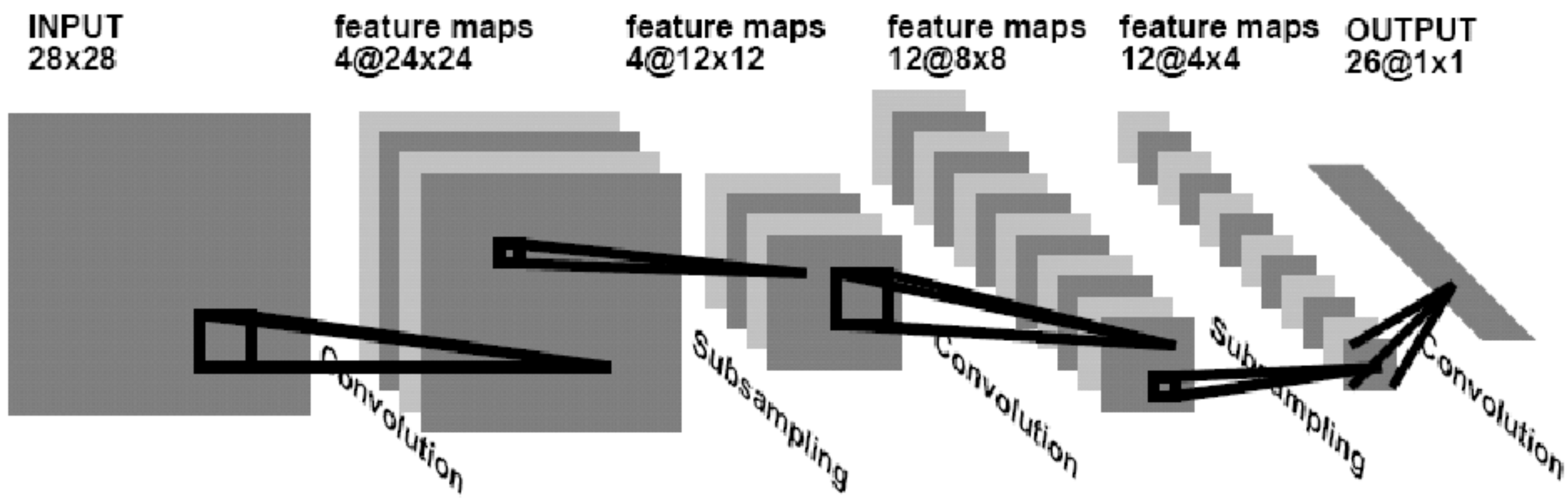
Neocognitron





Learning is done greedily for each layer

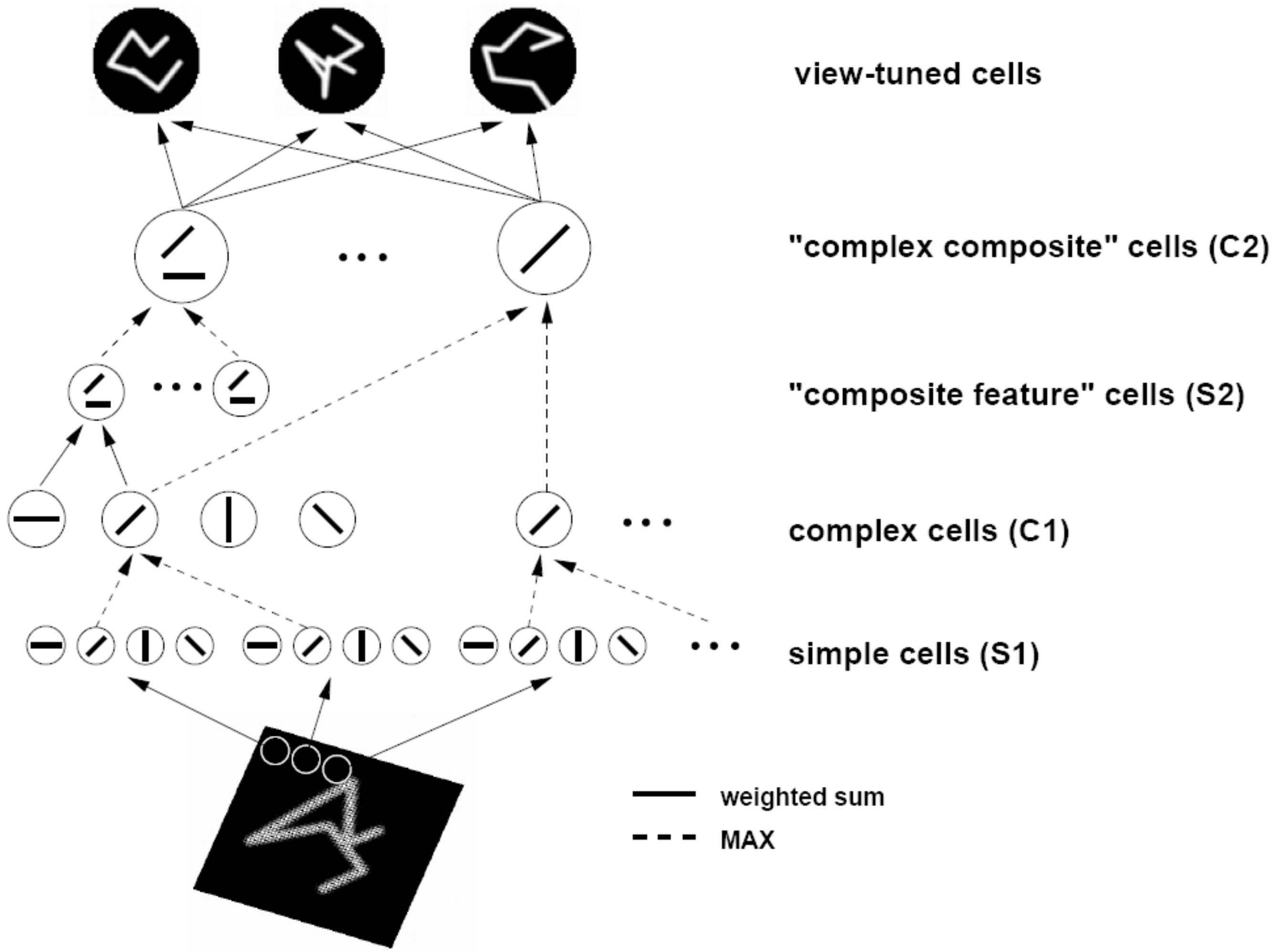
Convolutional Neural Network



Le Cun et al, 98

The output neurons share all the
intermediate levels

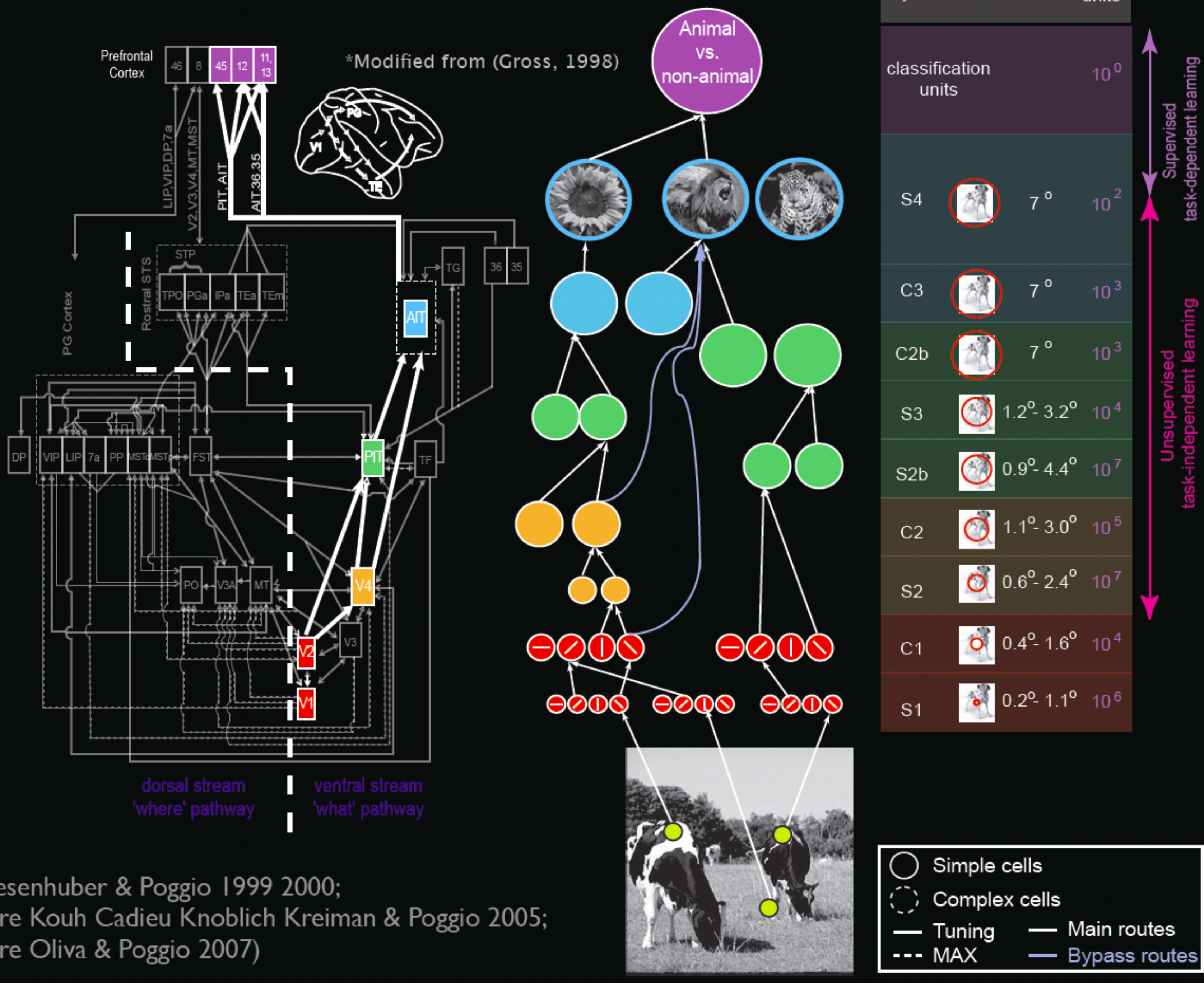
Hierarchical models of object recognition in cortex



Hierarchical extension of the classical paradigm of building complex cells from simple cells.
Uses same notation than Fukushima: “S” units performing template matching, solid lines and “C” units performing non-linear operations (“MAX” operation, dashed lines)

Riesenhuber, M. and Poggio, T. 99

Slide by T. Serre



(Riesenhuber & Poggio 1999 2000;
Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005;
Serre Oliva & Poggio 2007)

2 key learning stages:

- Task-specific circuits:
  - Supervised learning from ~100-1000 labeled examples
  - Linear classifier on top of VTUs (S4 units) [~RBF] (see Fredman Riesenhuber Poggio Miller, 2001, 2003)

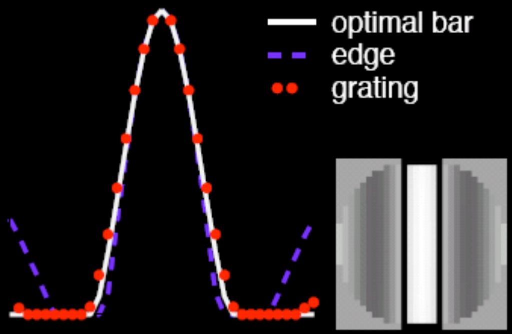
- Large dictionary of reusable features:
  - “unbound” features (Treisman & Gelade 1980; Wolfe & Bennett 1997; Schyns & Oliva 1994)
  - Different levels of invariance and complexity
  - Unsupervised learning from natural images ~developmental-like learning stage

Related to Edelman & Poggio (Edelman & Poggio 1990)

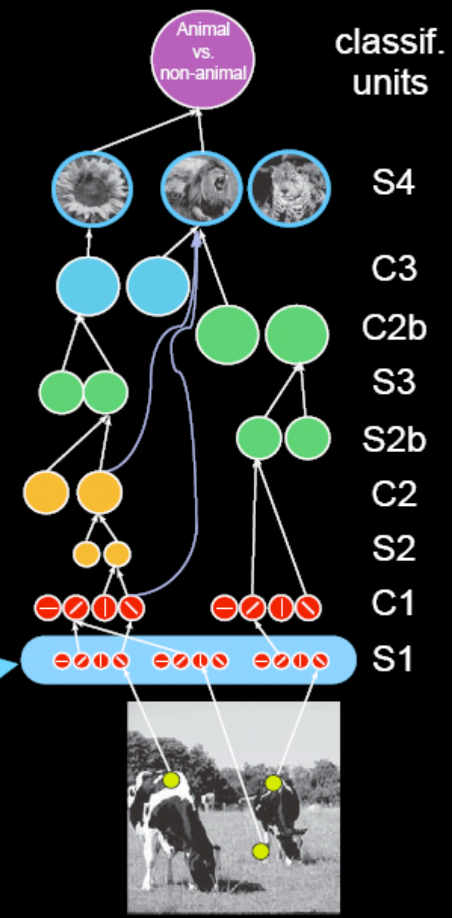
Related to Ullman’s visual features of intermediate complexity (Ullman et al 2002)

Gabor filters (Jones & Palmer 1987)

Slide by T. Serre



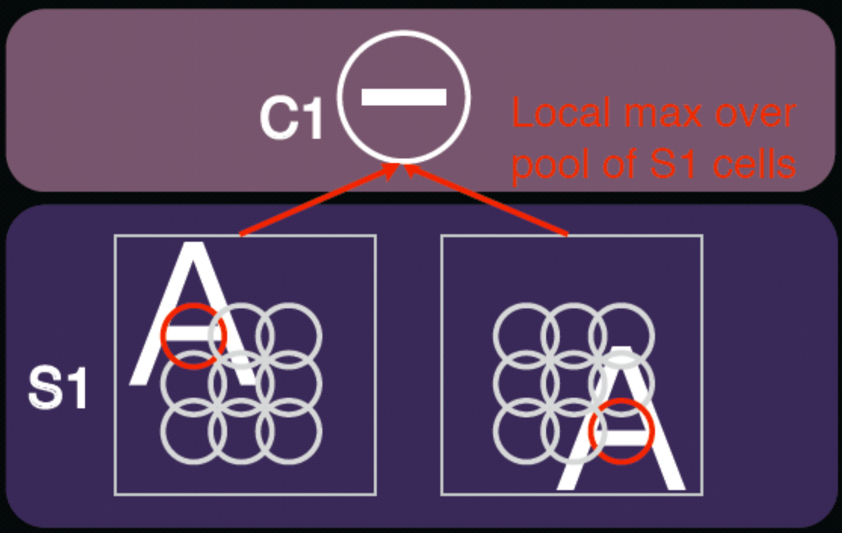
S1 units

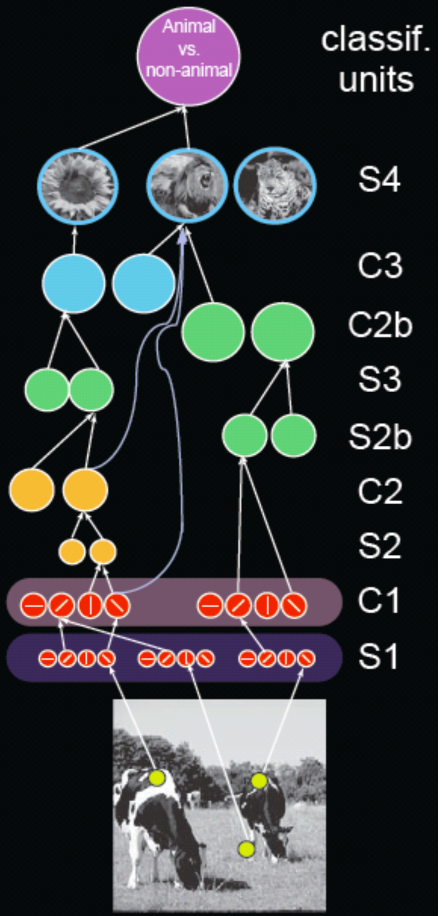


- ✦ Gabor filters
- ✦ Parameters fit to V1 data (Serre & Riesenhuber 2004)
  - 17 spatial frequencies (=scales)
  - 4 orientations

C1 units

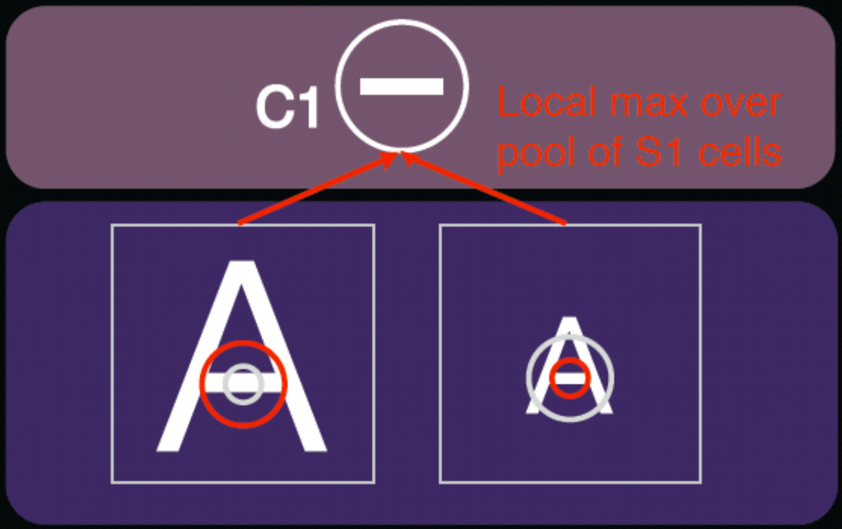
Increase in tolerance to **position** (and in RF size)

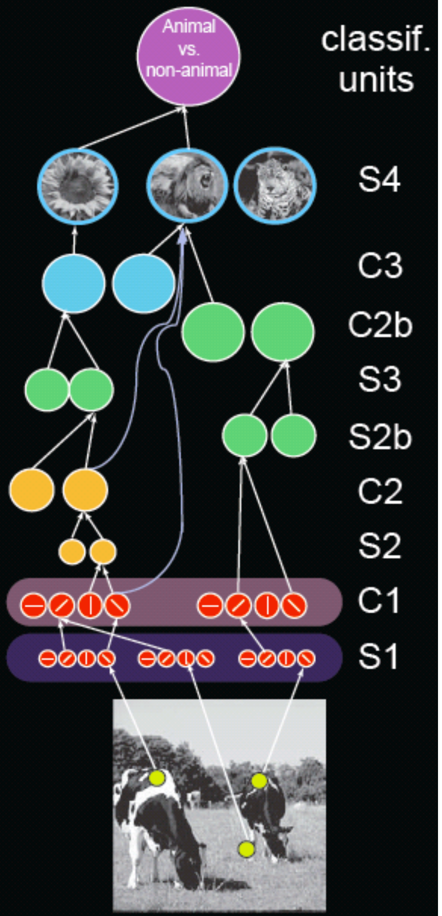




C1 units

Increase in tolerance to **scale**

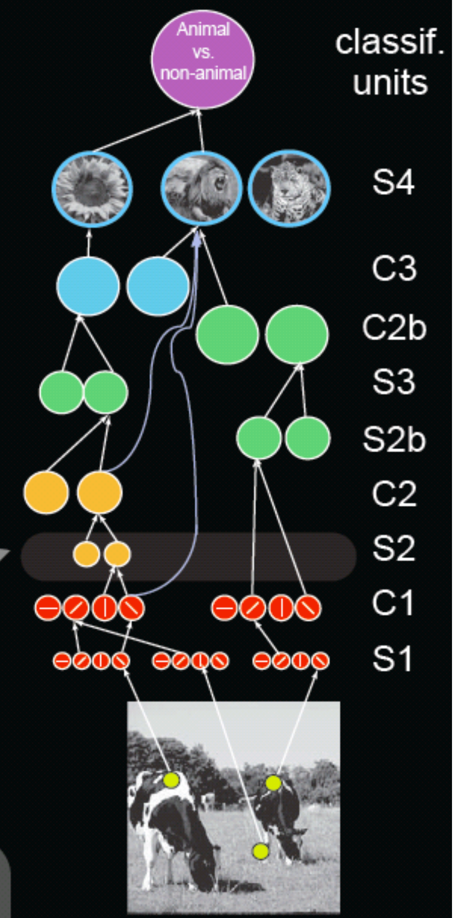


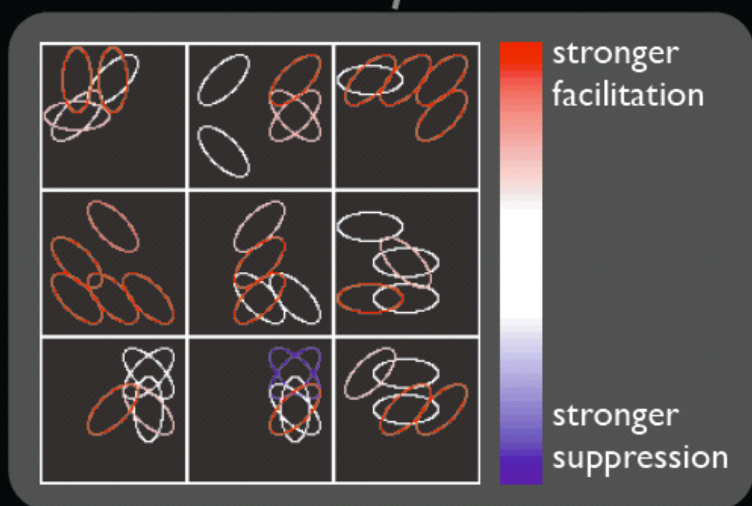


S2 units

- ✦ Features of moderate complexity (n~1,000 types)
- ✦ Combination of V1-like complex units at different orientations

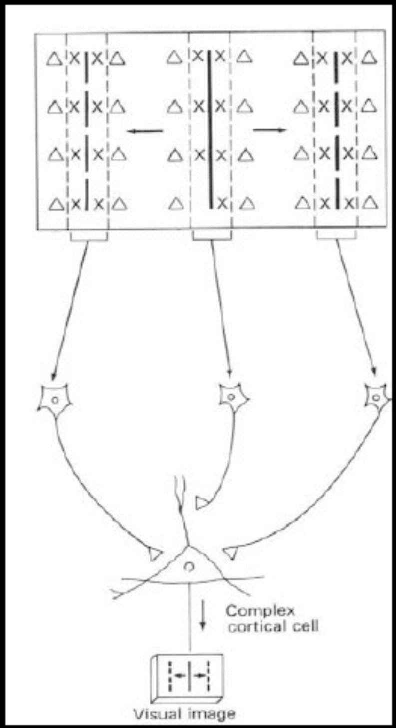
  - Synaptic weights ***w*** learned from natural images
  - 5-10 subunits chosen at random from all possible afferents (~100-1,000)



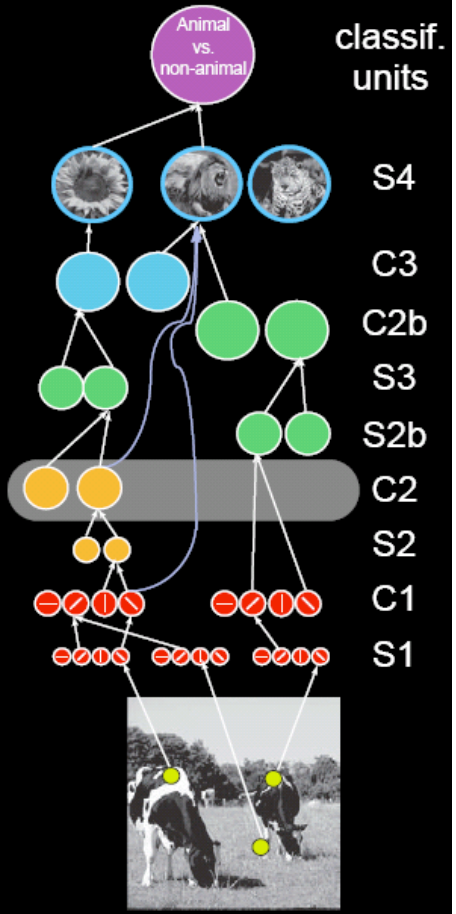


C2 units

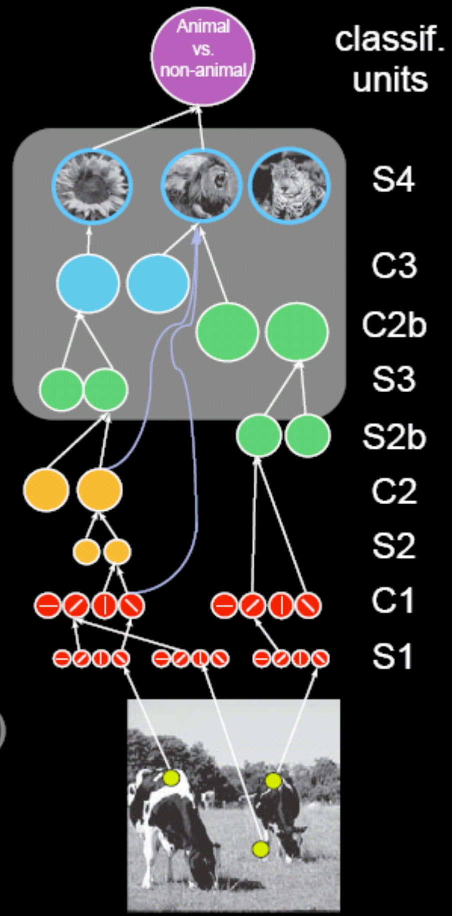
- ✦ Same selectivity as S2 units but increased tolerance to position and size of preferred stimulus
- ✦ Local pooling over S2 units with same selectivity but slightly different positions and scales
- ✦ S2 units in V2 and C2 in V4?



(Hubel & Wiesel 1959)

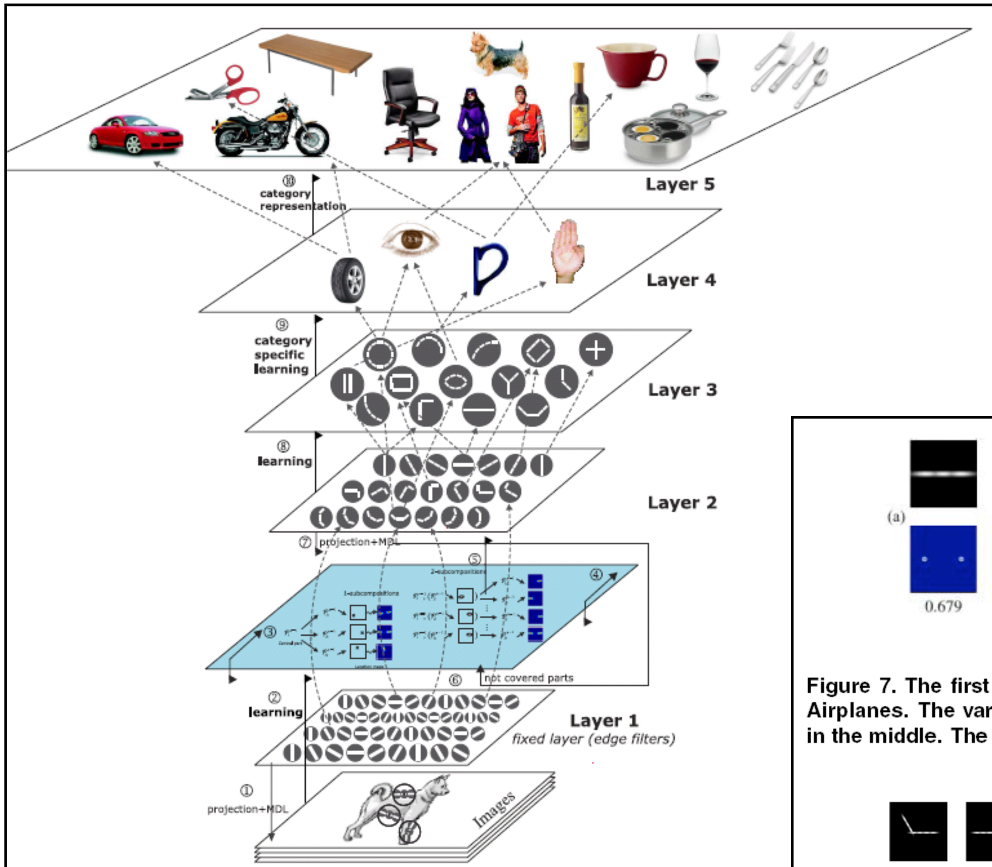


Beyond C2 units

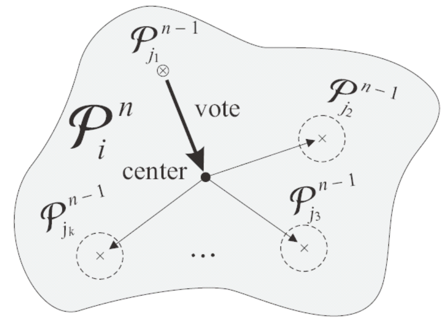


- ✦Units increasingly complex and invariant
- ✦ S3/C3 units:
  - Combination of V4-like units with different selectivities
  - Dictionary of ~1,000 features = num. columns in IT (Fujita 1992)
- ✦ S4 units:
  - View-tuned units (imprinted with part of the training set, e.g. animal and non-animal images but still unsupervised)
  - Tuning and invariance properties agrees with IT data (Logothetis, Pauls & Poggio 1995)

Learning a Compositional Hierarchy of Object Structure



The architecture



Parts model

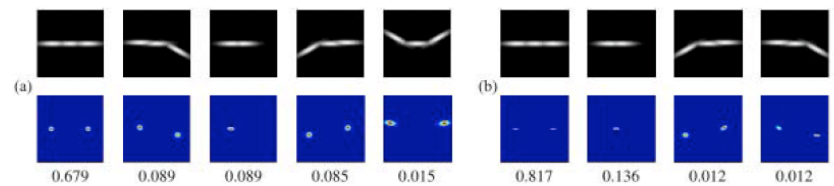


Figure 7. The first row depicts the final parts comprising Layer II obtained for (a) Cliparts and (b) Airplanes. The variances of position distributions of parts, relative to the central part, are depicted in the middle. The feature probabilities are listed in the last row.

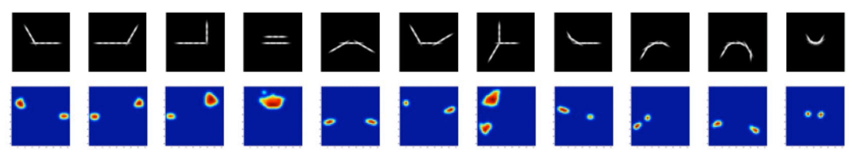


Figure 8. (a) Examples of Layer 3 parts, (b) variances of positions of the surrounding subparts

Learned parts

Fidler & Leonardis, CVPR'07; Fidler, Boben & Leonardis, CVPR 2008

Learning a Compositional Hierarchy of Object Structure

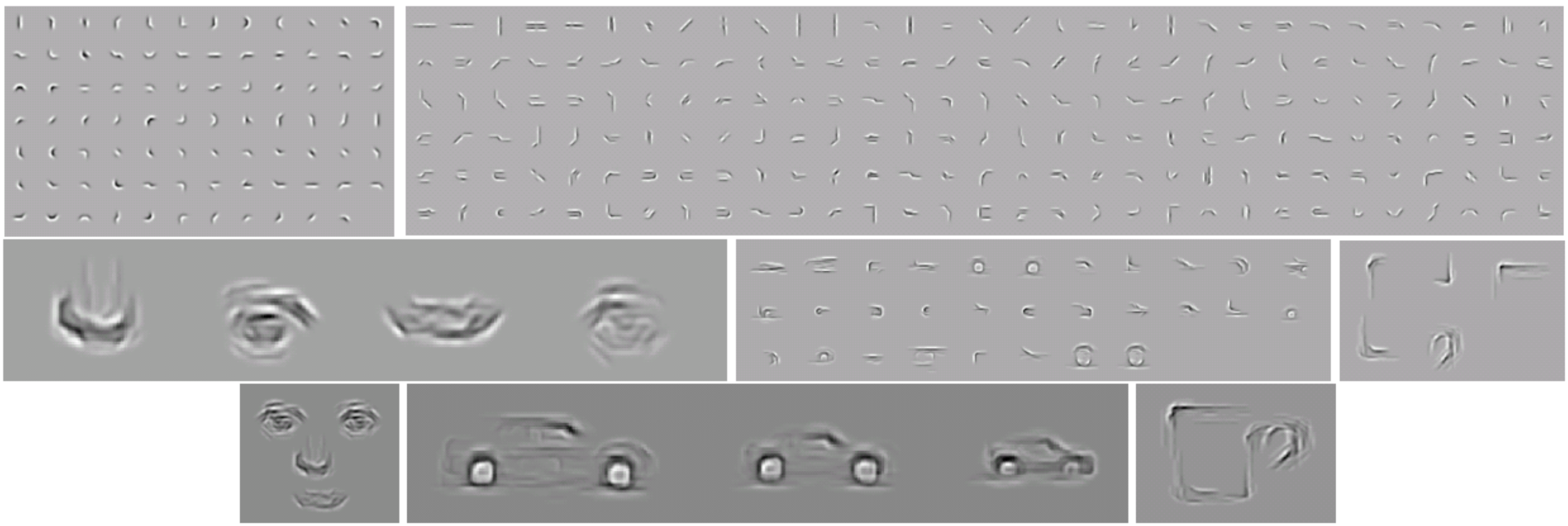
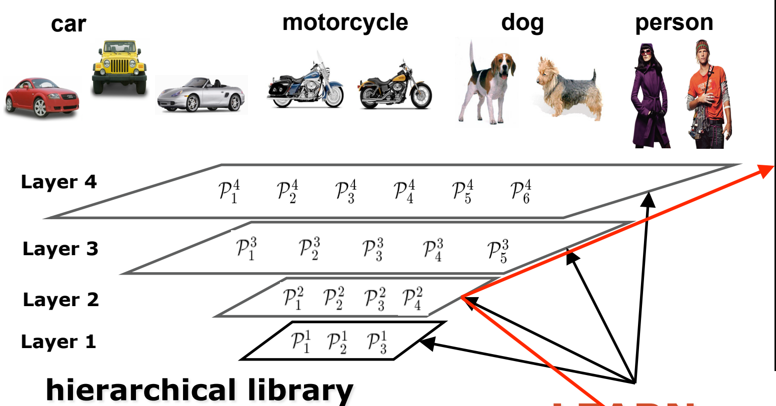


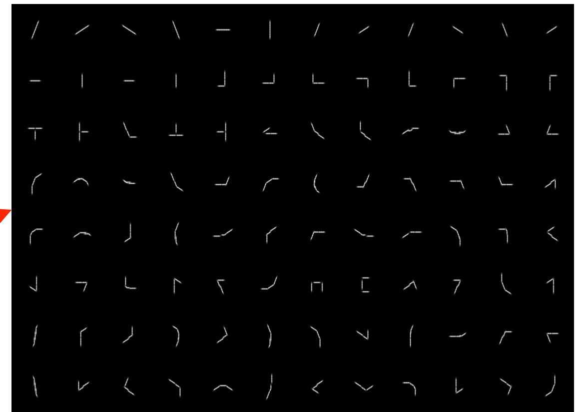
Figure 4. Mean reconstructions of the learned parts (spatial flexibility also modeled by the parts is omitted due to lack of space). **1st row**: $\mathcal{L}_2$, $\mathcal{L}_3$ (the first 186 of all 499 parts are shown), **2nd row**: $\mathcal{L}_4$ parts for faces, cars, and mugs, **3rd row**: $\mathcal{L}_5$ parts for faces, cars (obtained on 3 different scales), and mugs.

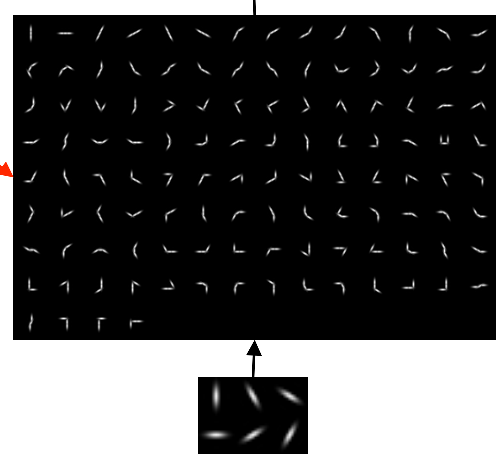
Fidler & Leonardis, CVPR'07; Fidler, Boben & Leonardis, CVPR 2008

Learning a Compositional Hierarchy of Object Structure

- Fidler & Leonardis, CVPR'07
- Fidler, Boben & Leonardis, CVPR 2008

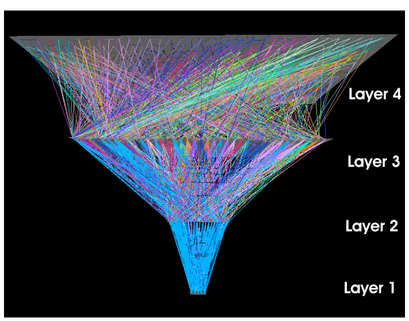


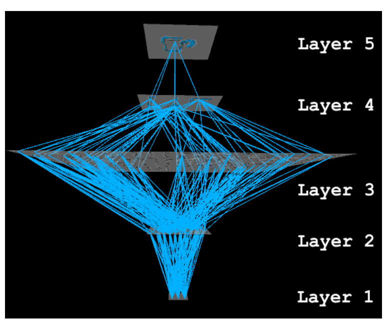




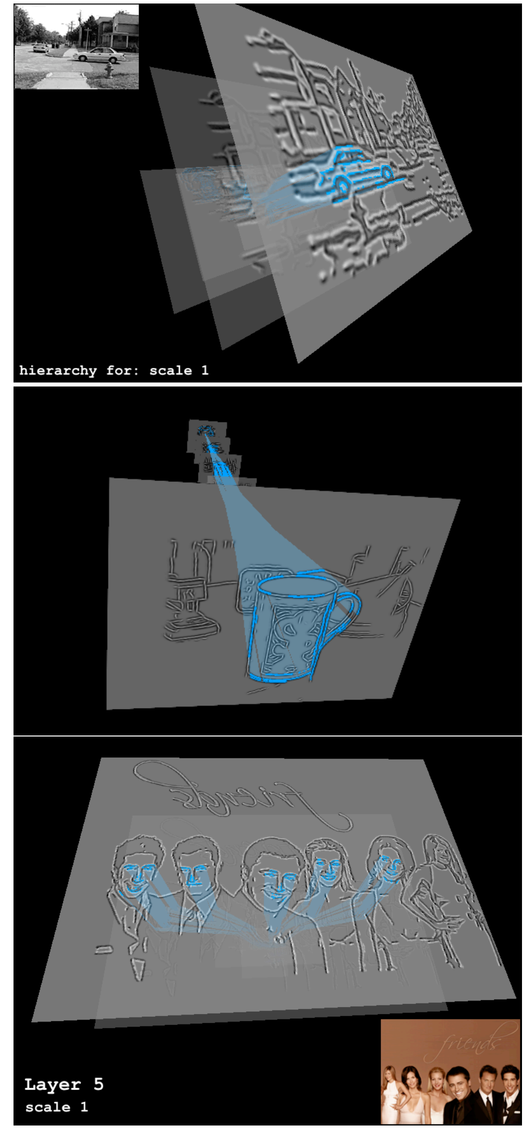
Learned L1 – L3

- Hierarchical compositional architecture
- Features are shared at each layer
- Learning is done on natural images
- Indexing and matching detection scheme





Learned hierarchical vocabulary



Detections