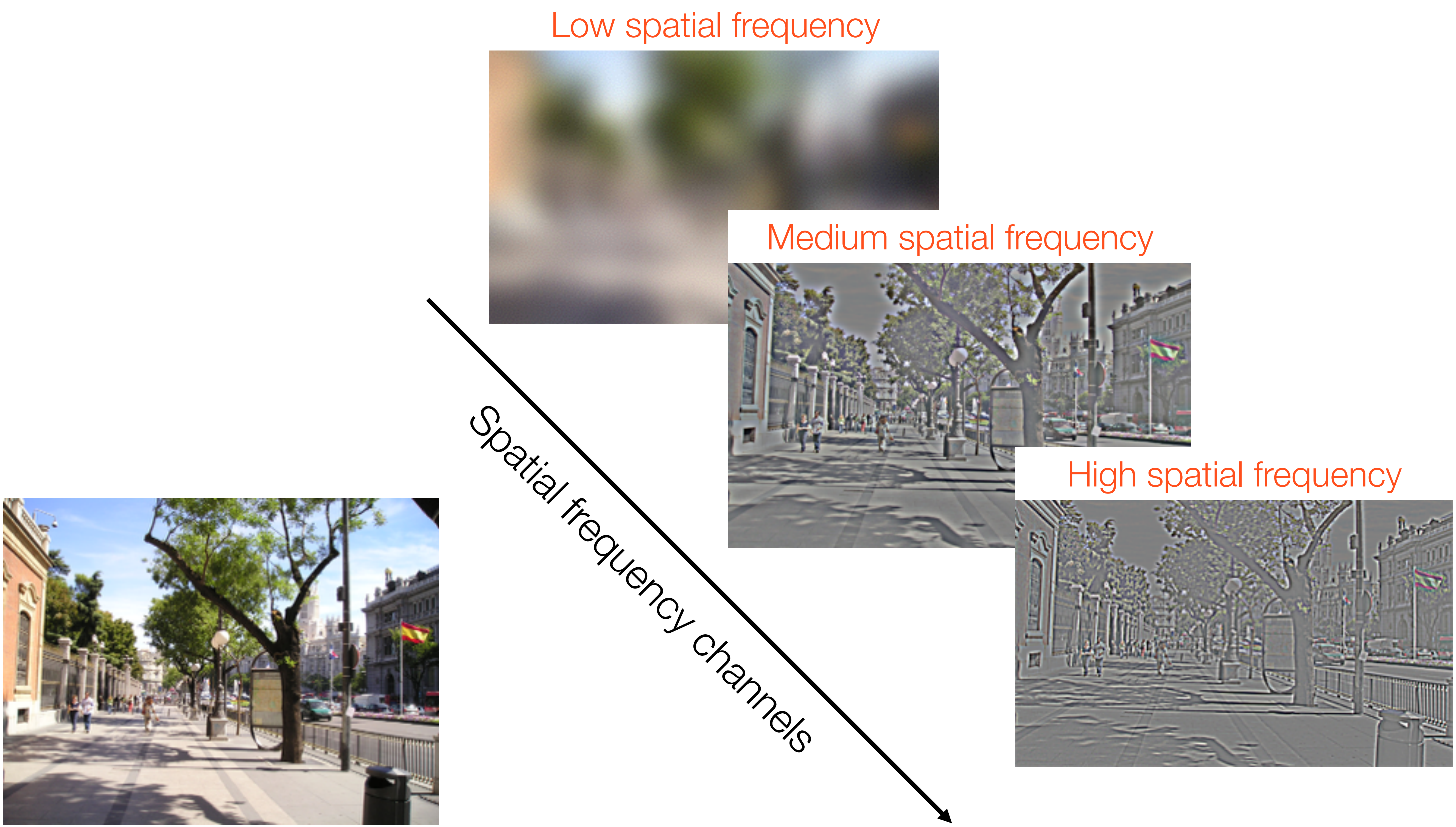
Computational Vision

Primary visual cortex

- Orientation selectivity
- Spatial frequency
- Normalization
- Color opponency

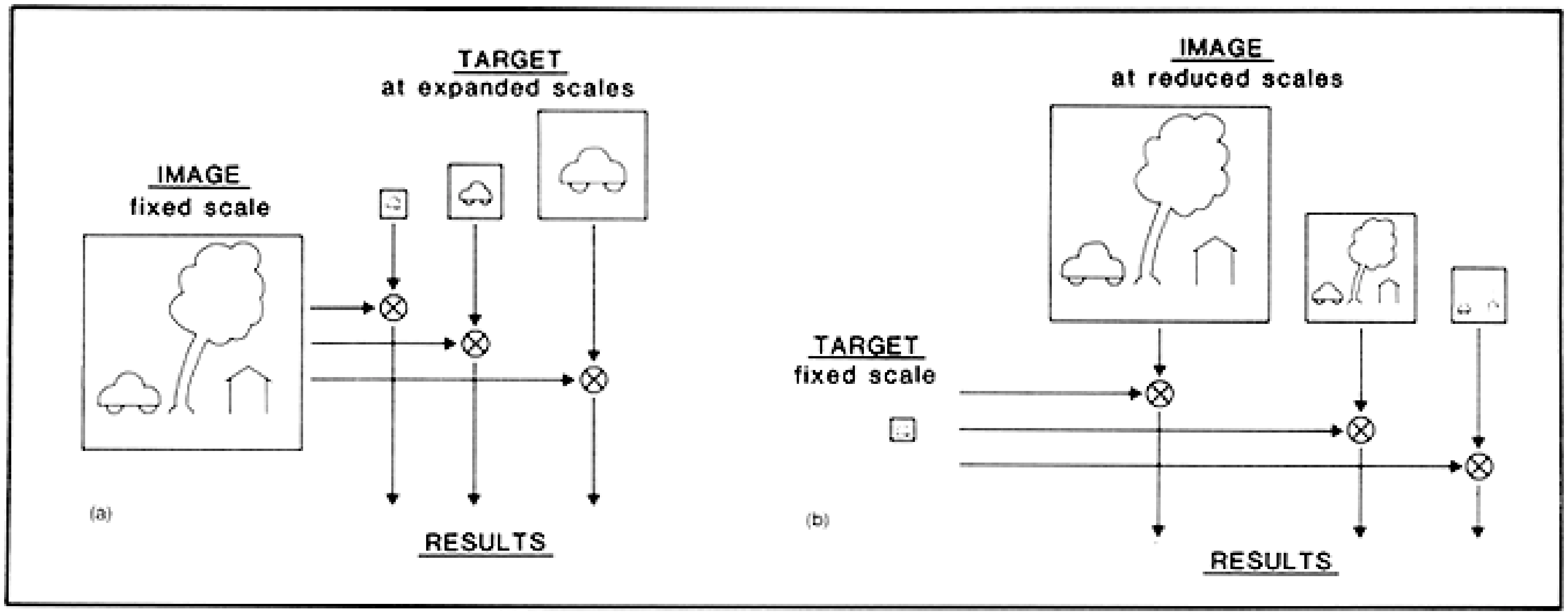
Frequency channels



Hybrid Images

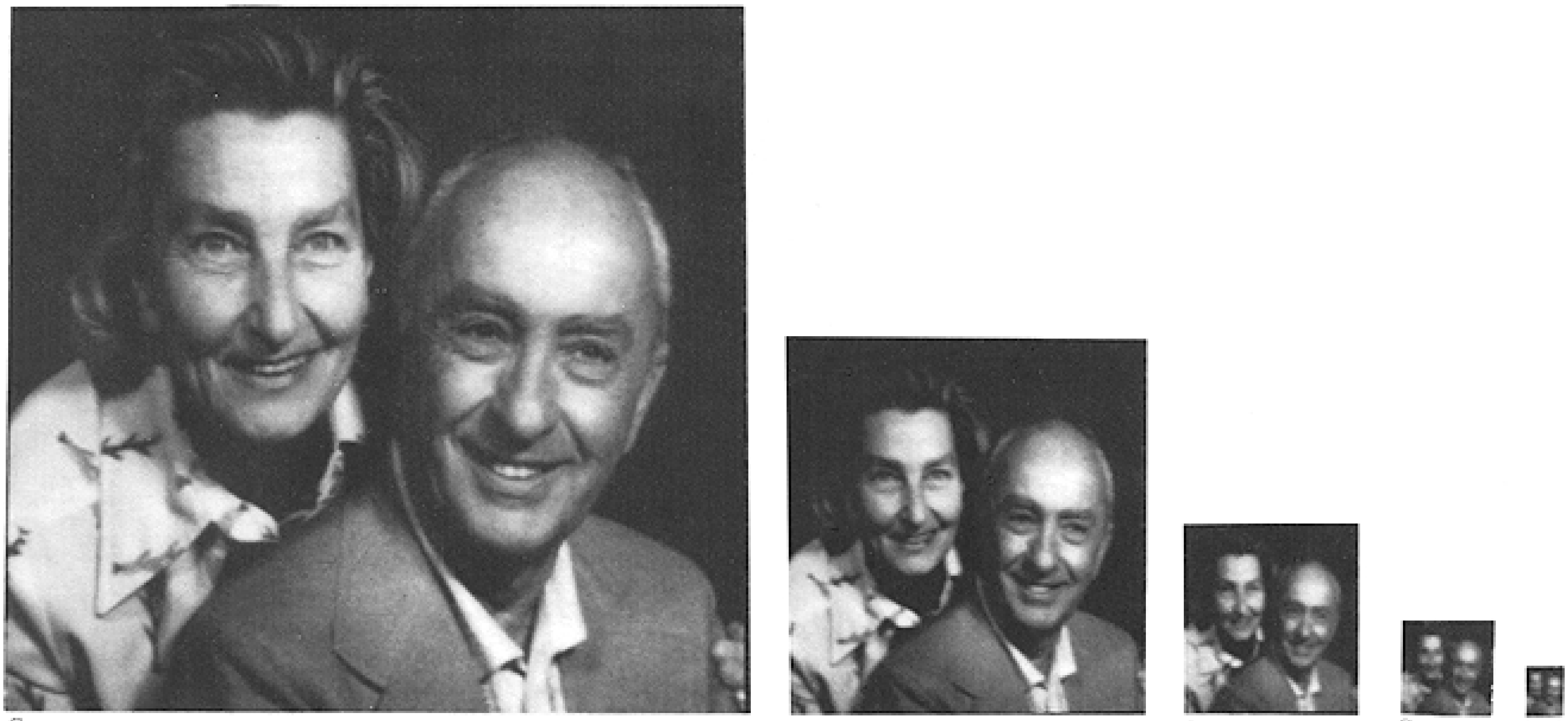
Aude Oliva (MIT)

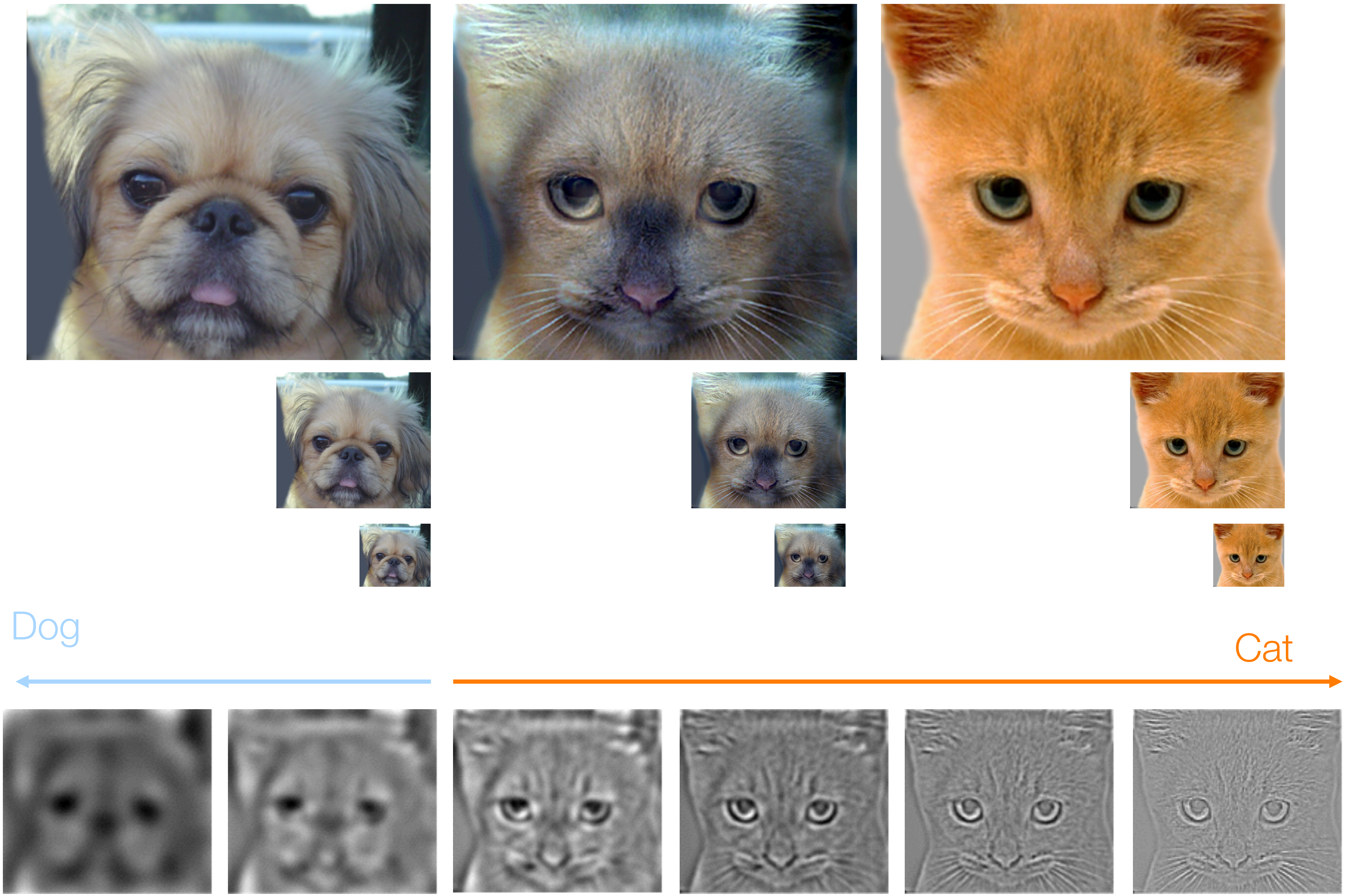
Image pyramids vs. frequency channels



brains computers

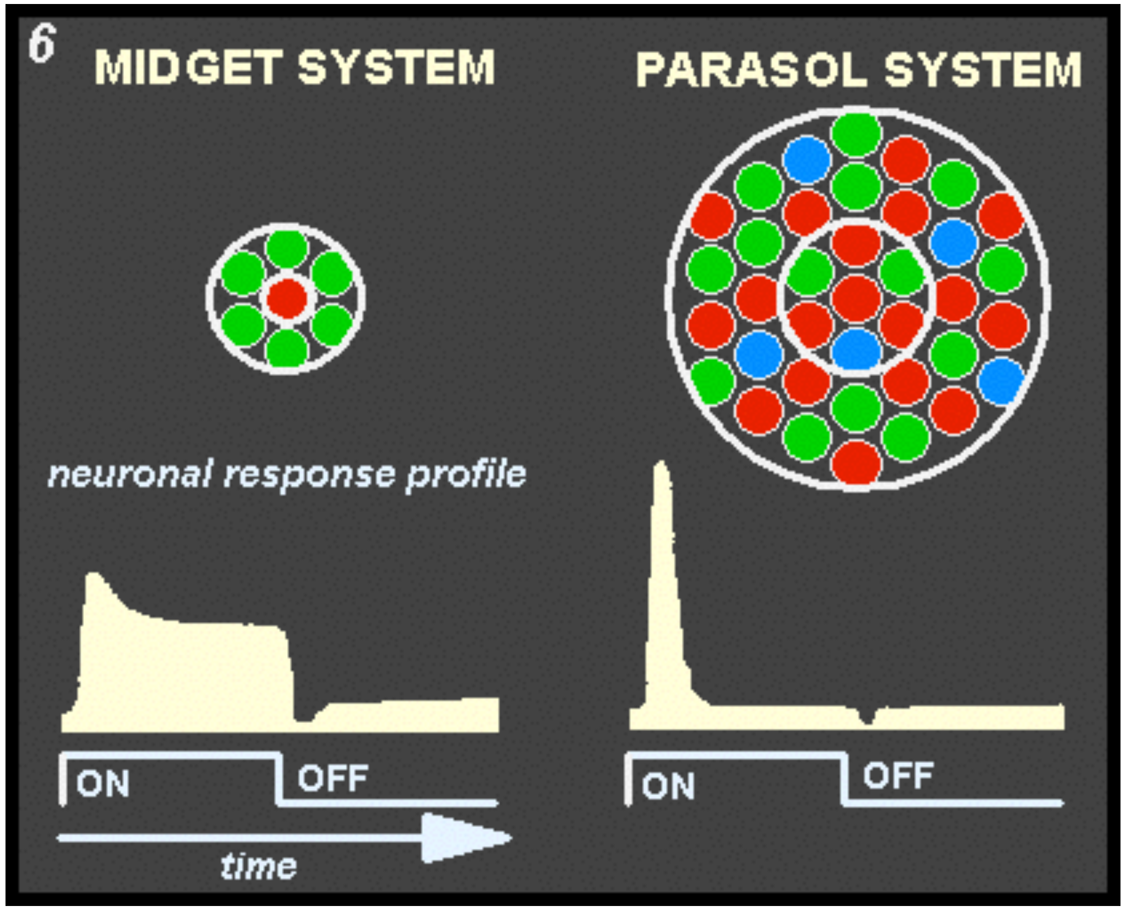
Image pyramids vs. frequency channels



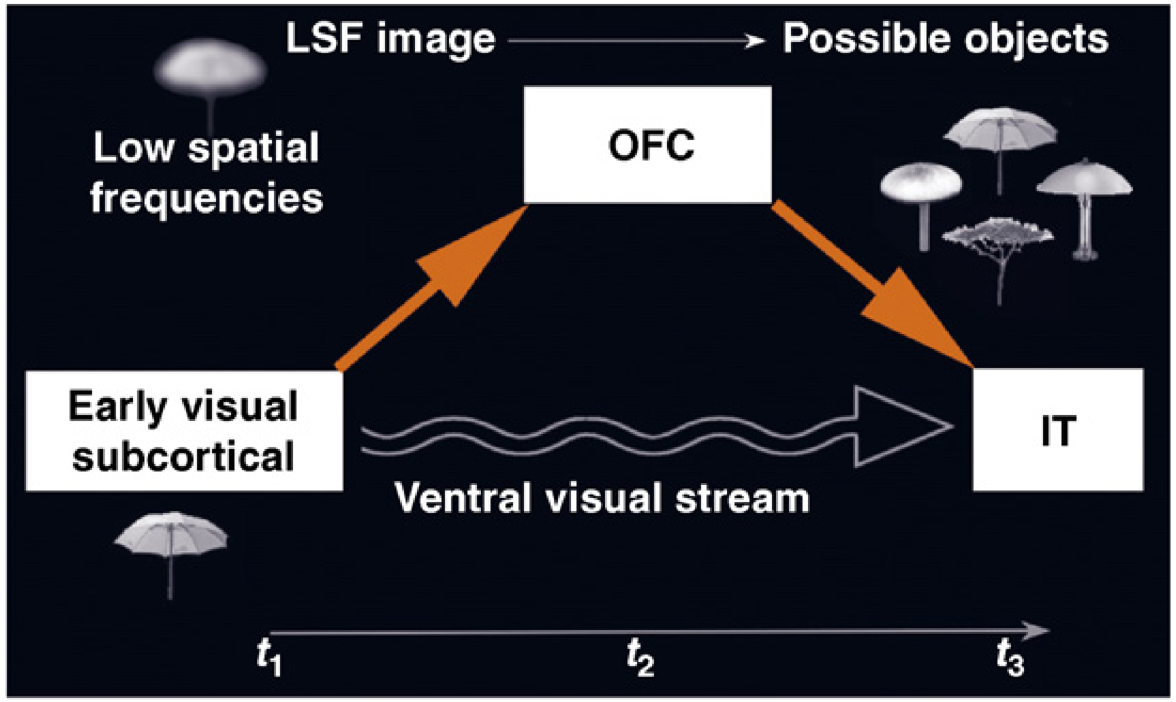

Dog
Cat

Theories of object recognition

high SF (slow) low SF (fast)



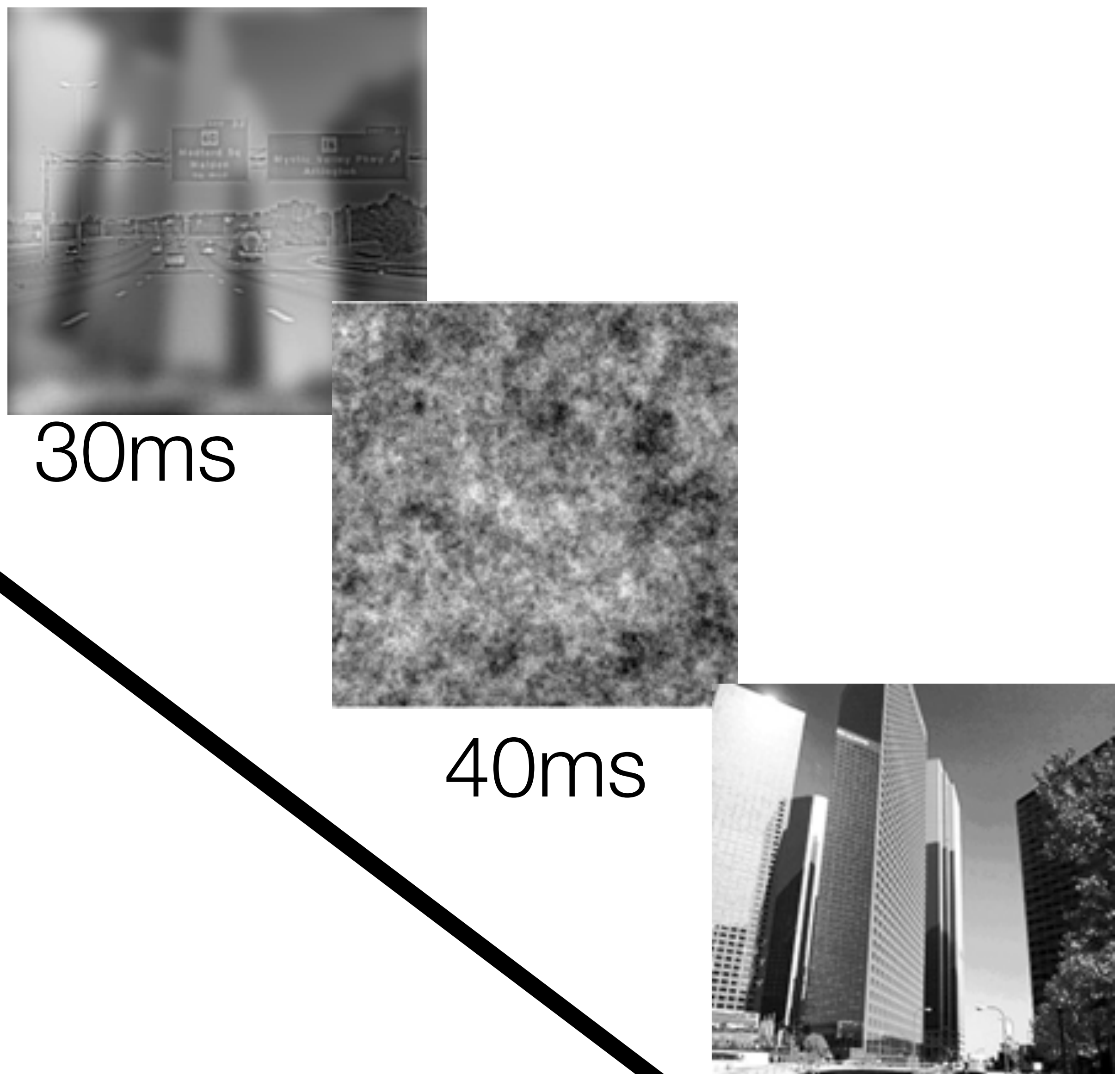
source: Peter Schiller



Bar '07

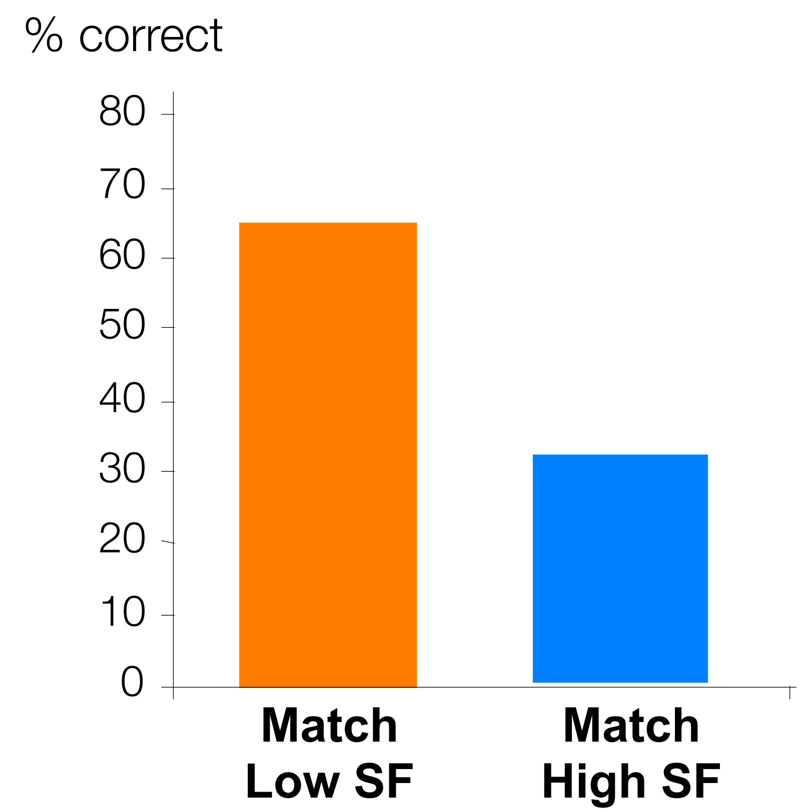
Coarse-to-fine processing

Duration: 30 ms



30ms

40ms



The second image can be:
- New image
- Match to Low SF (city)
- Match to High SF (highway)

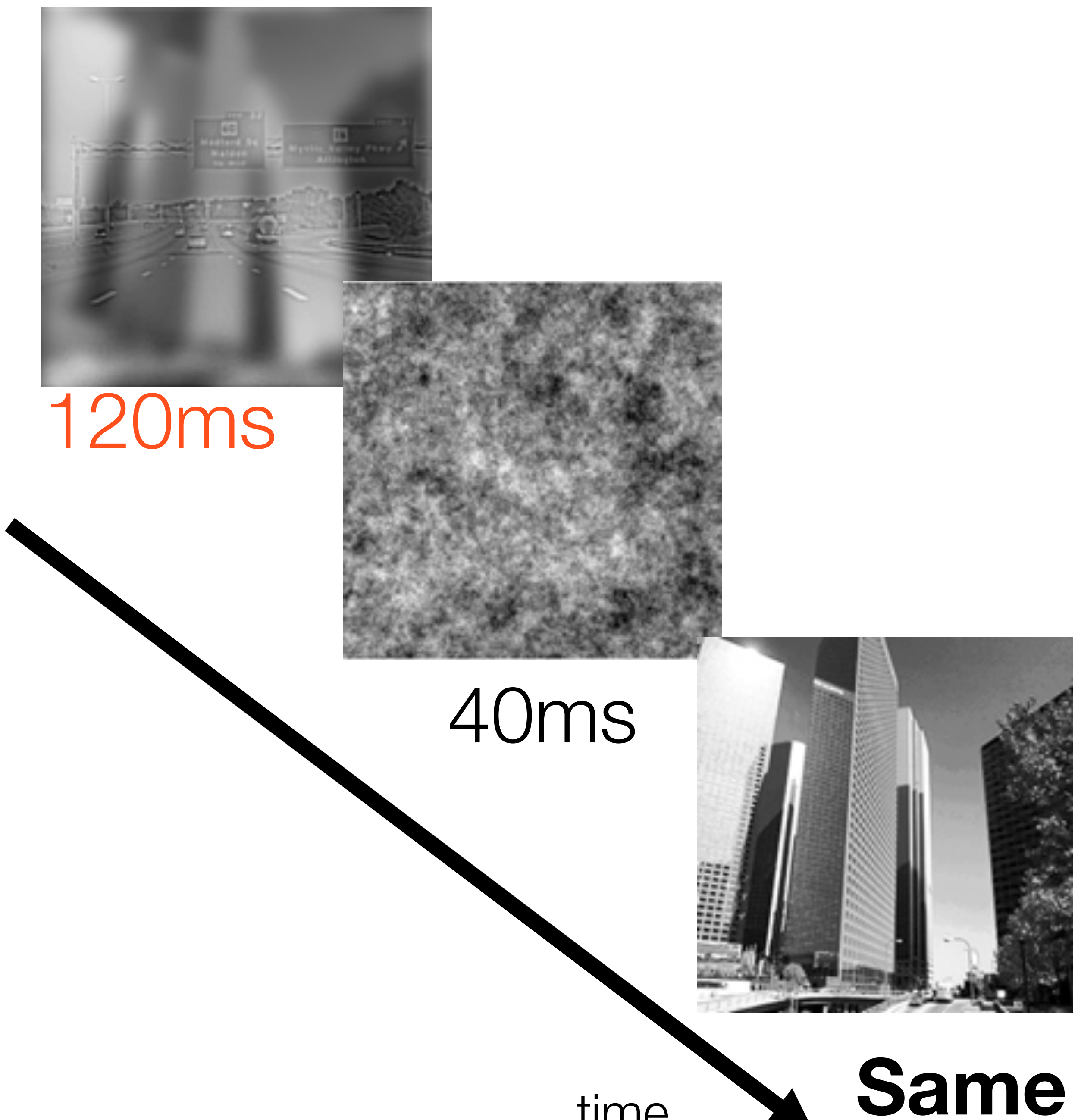
time

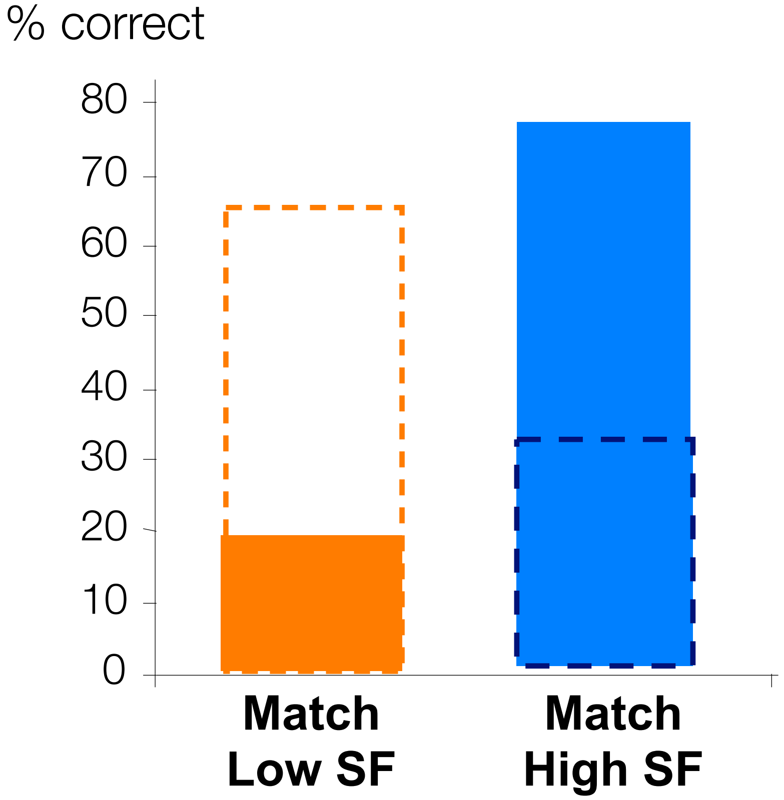
Same or different ?

Schyns & Oliva '94

Coarse-to-fine processing

Duration: 120 ms





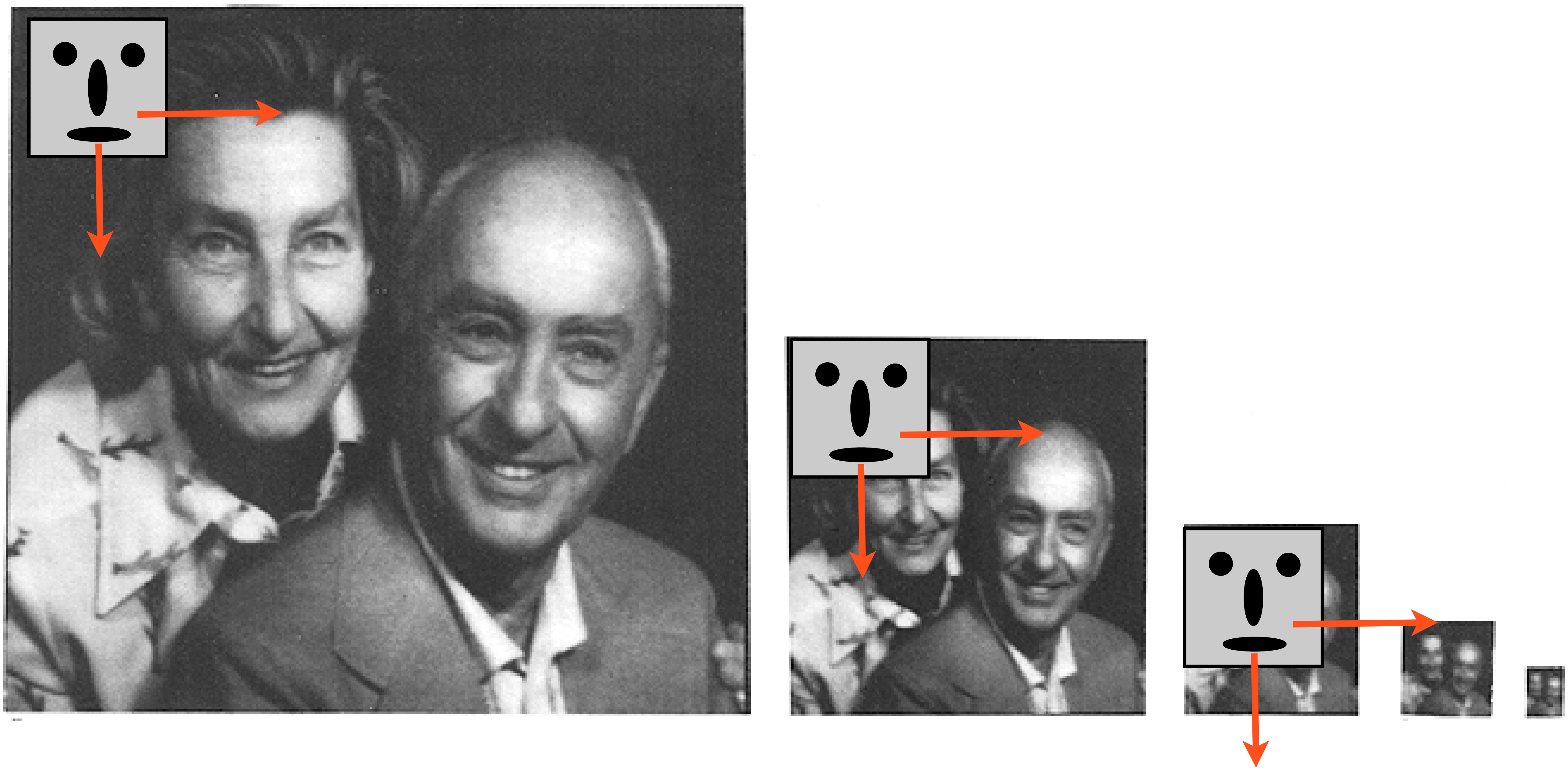
The second image can be:
- New image
- Match to Low SF (city)
- Match to High SF (highway)

Same or different ?

Schyns & Oliva ‘94

Application to machine vision

e.g, multi-scale object detection

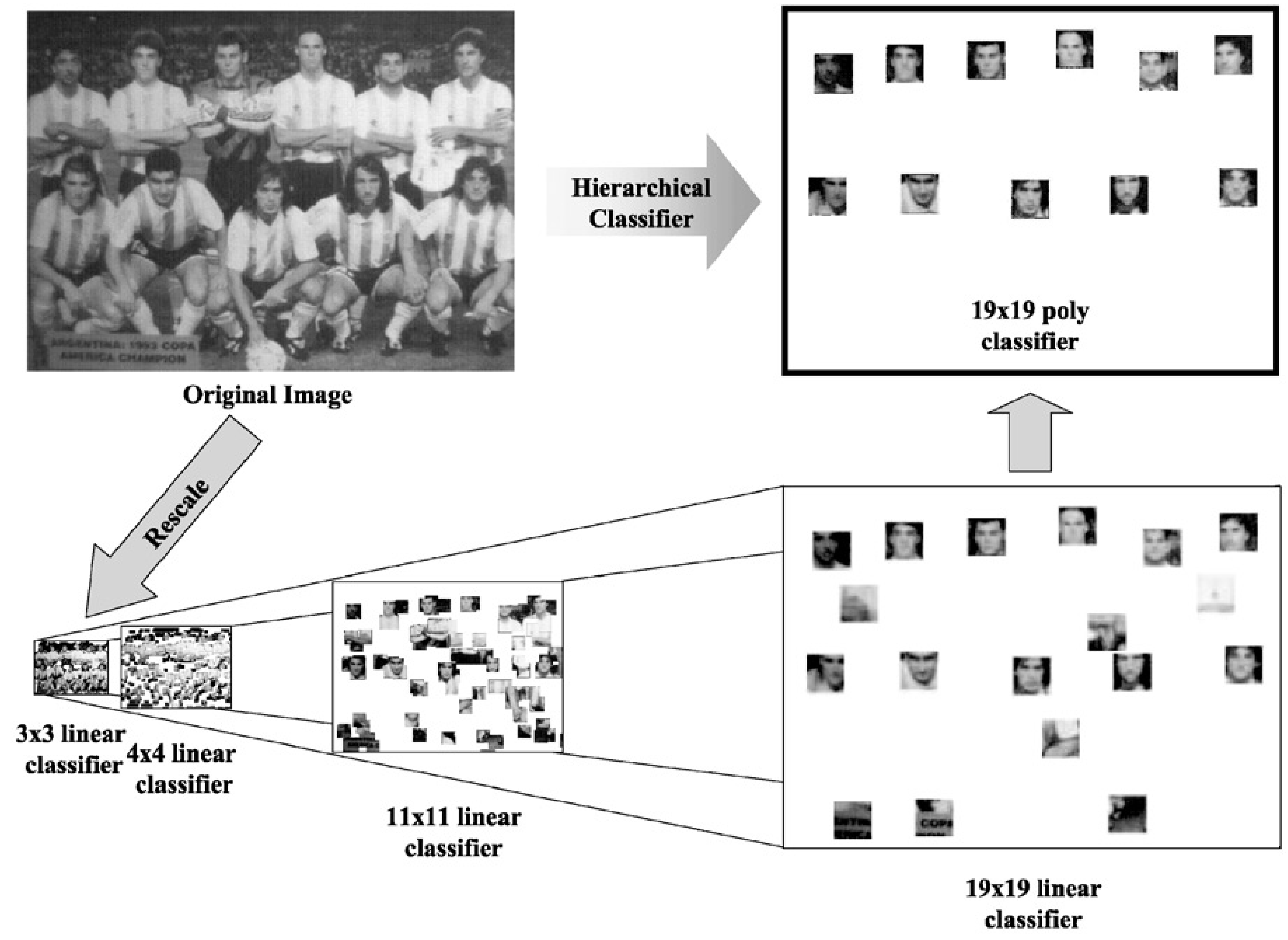


Application to machine vision

e.g, coarse-to-fine object strategies

speed-up: over 300 times

Heisele Serre & Poggio ’01

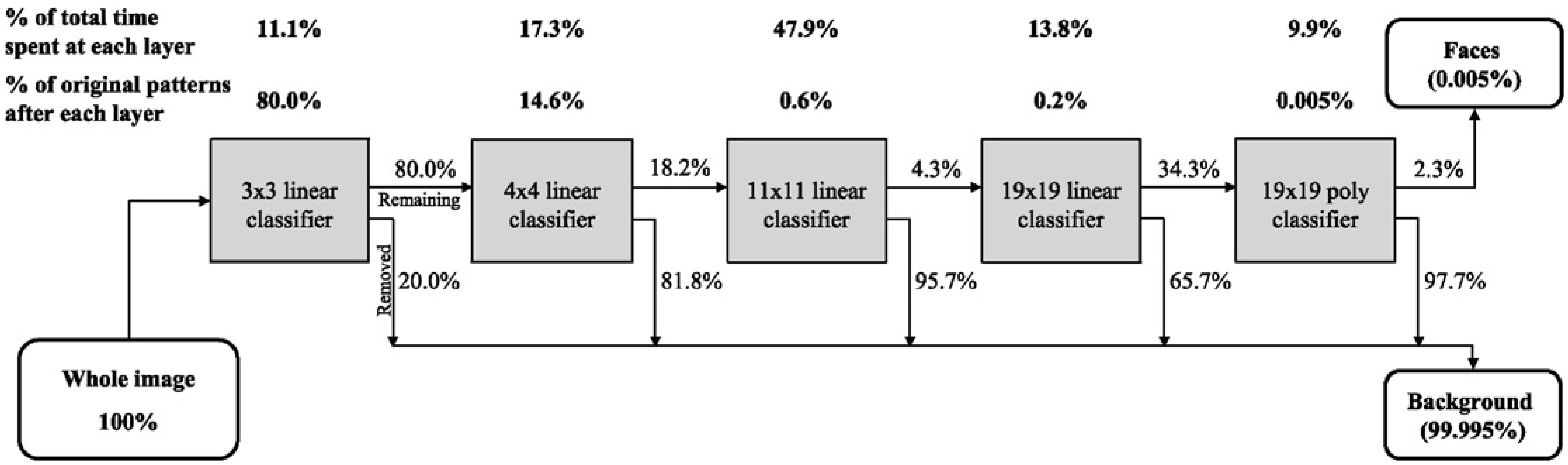


see also Fleuret & Geman ’01

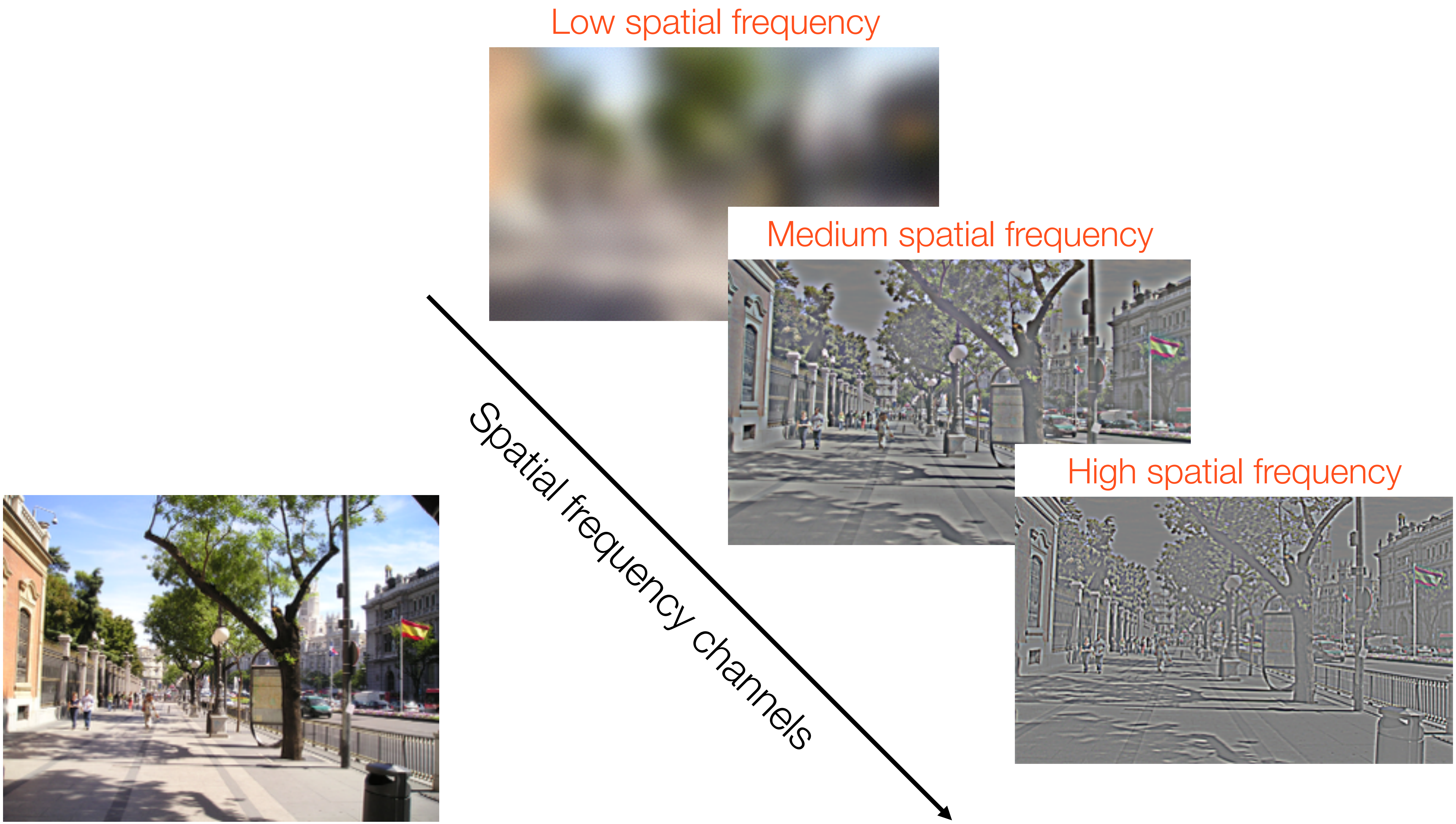
Application to machine vision

e.g, coarse-to-fine object strategies

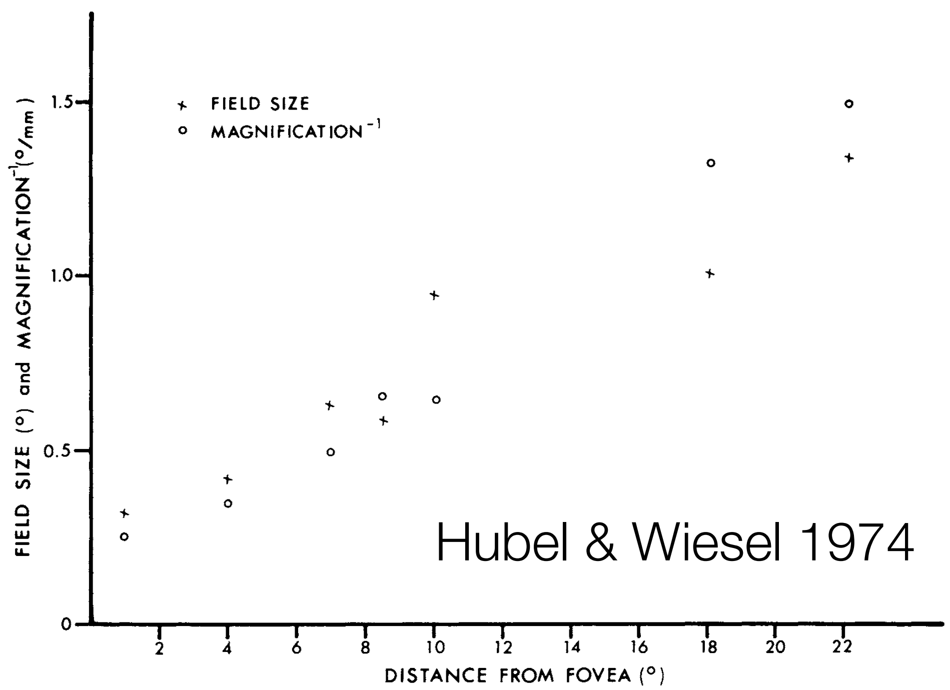
Heisele Serre & Poggio '01



Frequency channels

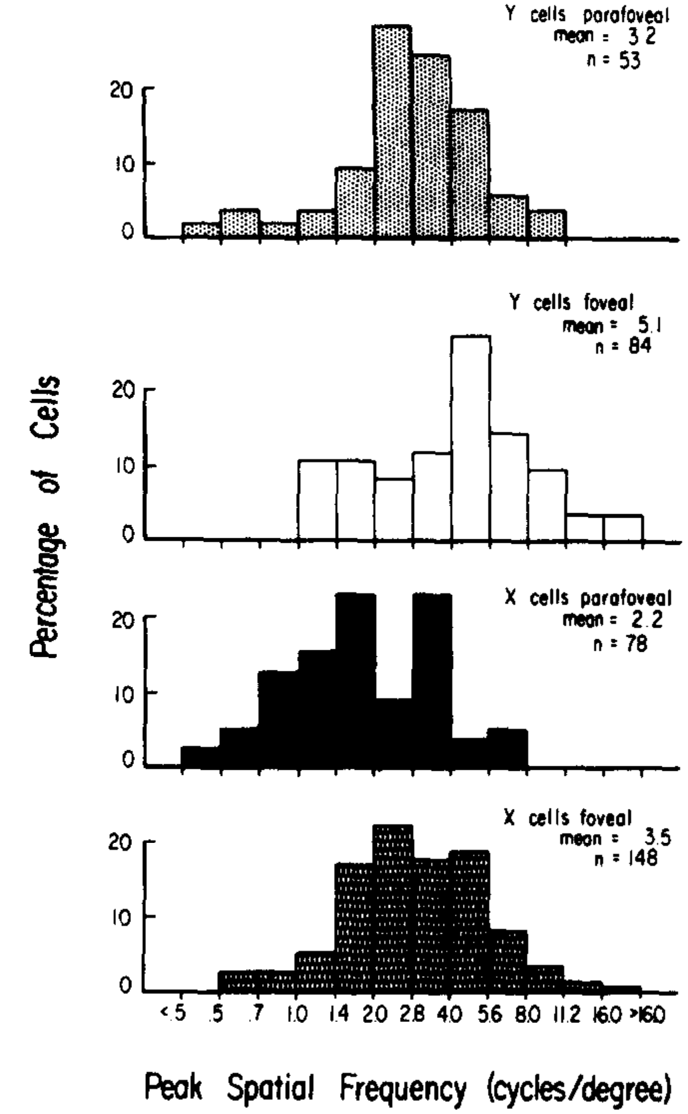


Spatial acuity and eccentricity



Hubel & Wiesel 1974

Hubel & Wiesel 1974

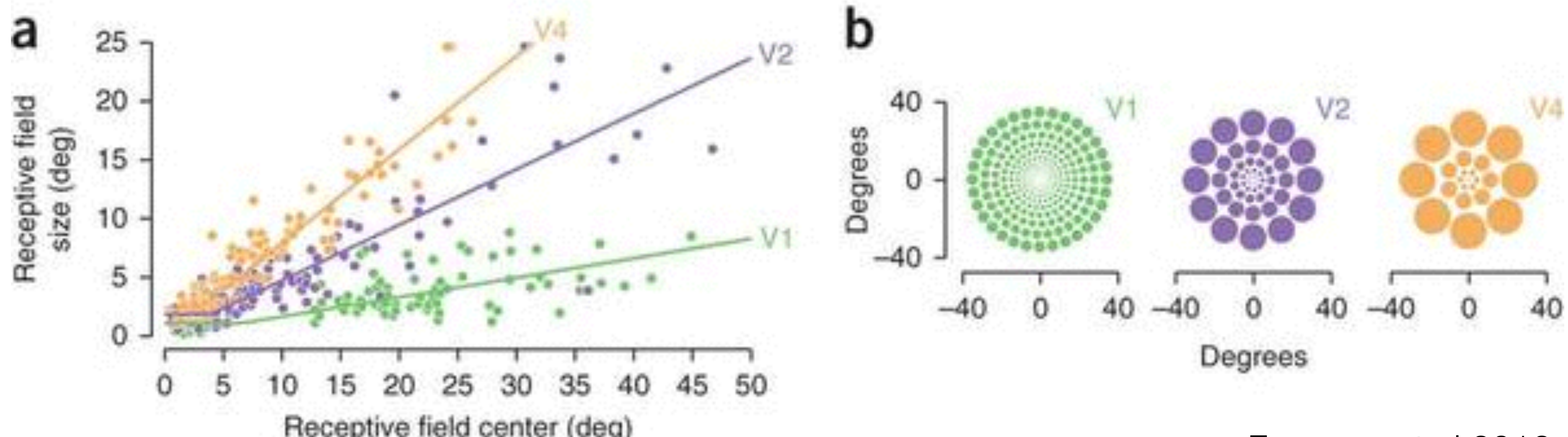


3º-5º

0º-1.5º

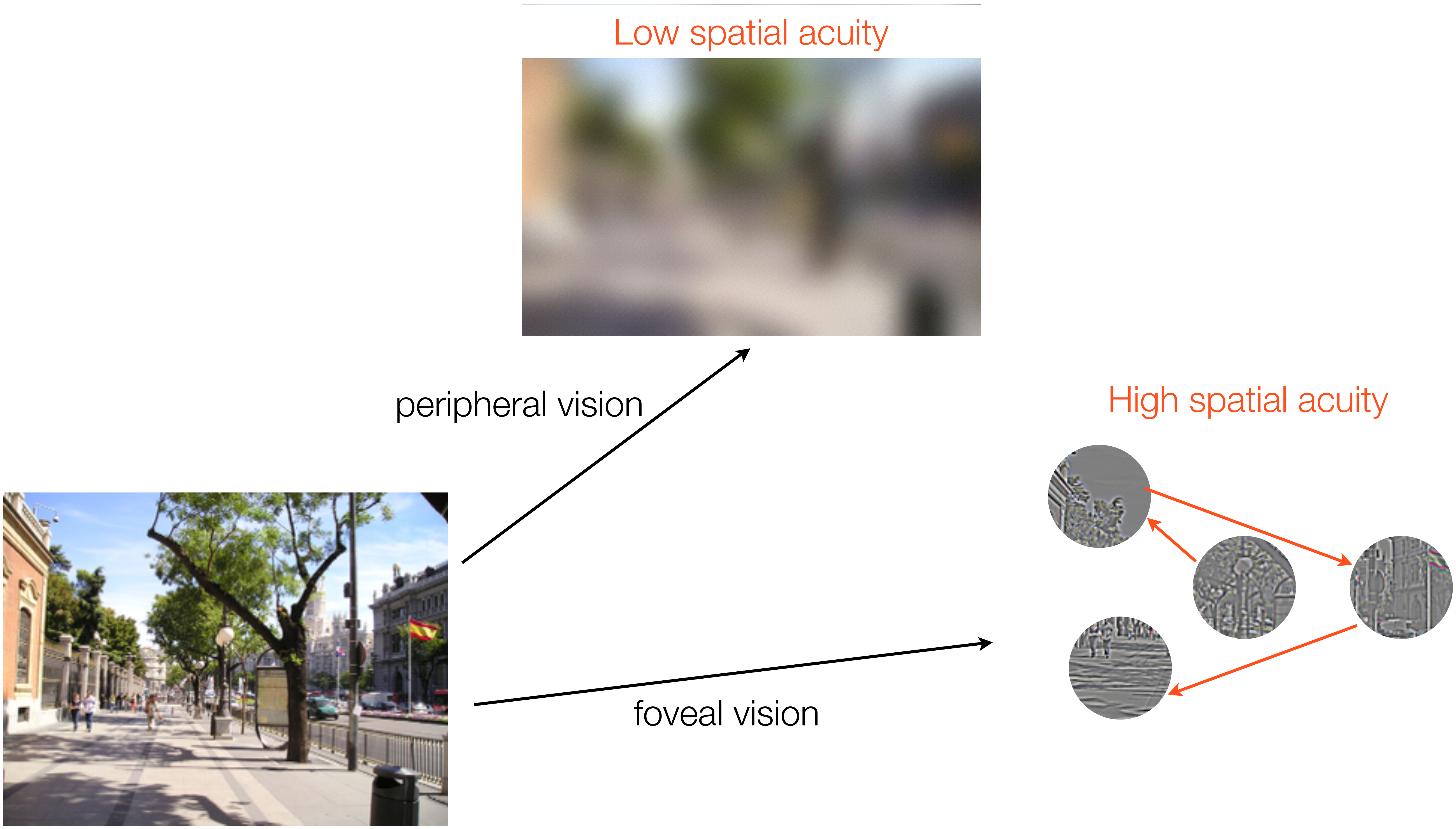
3º-5º

0º-1.5º

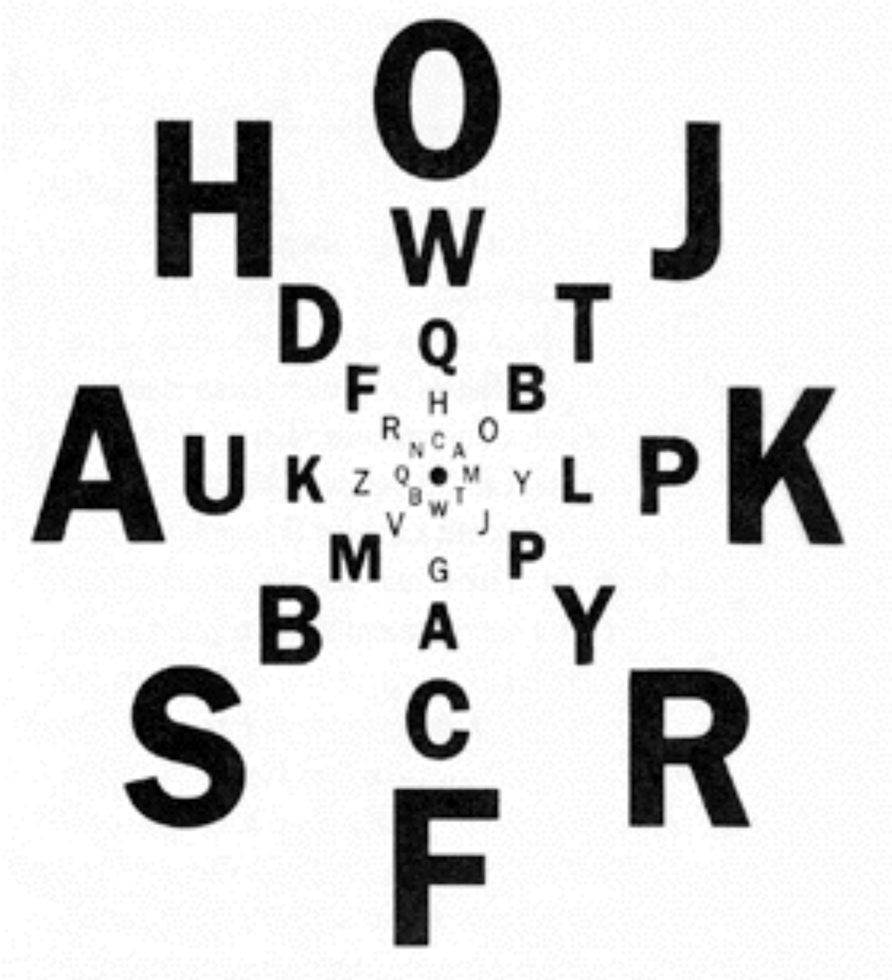


Freeman et al 2012

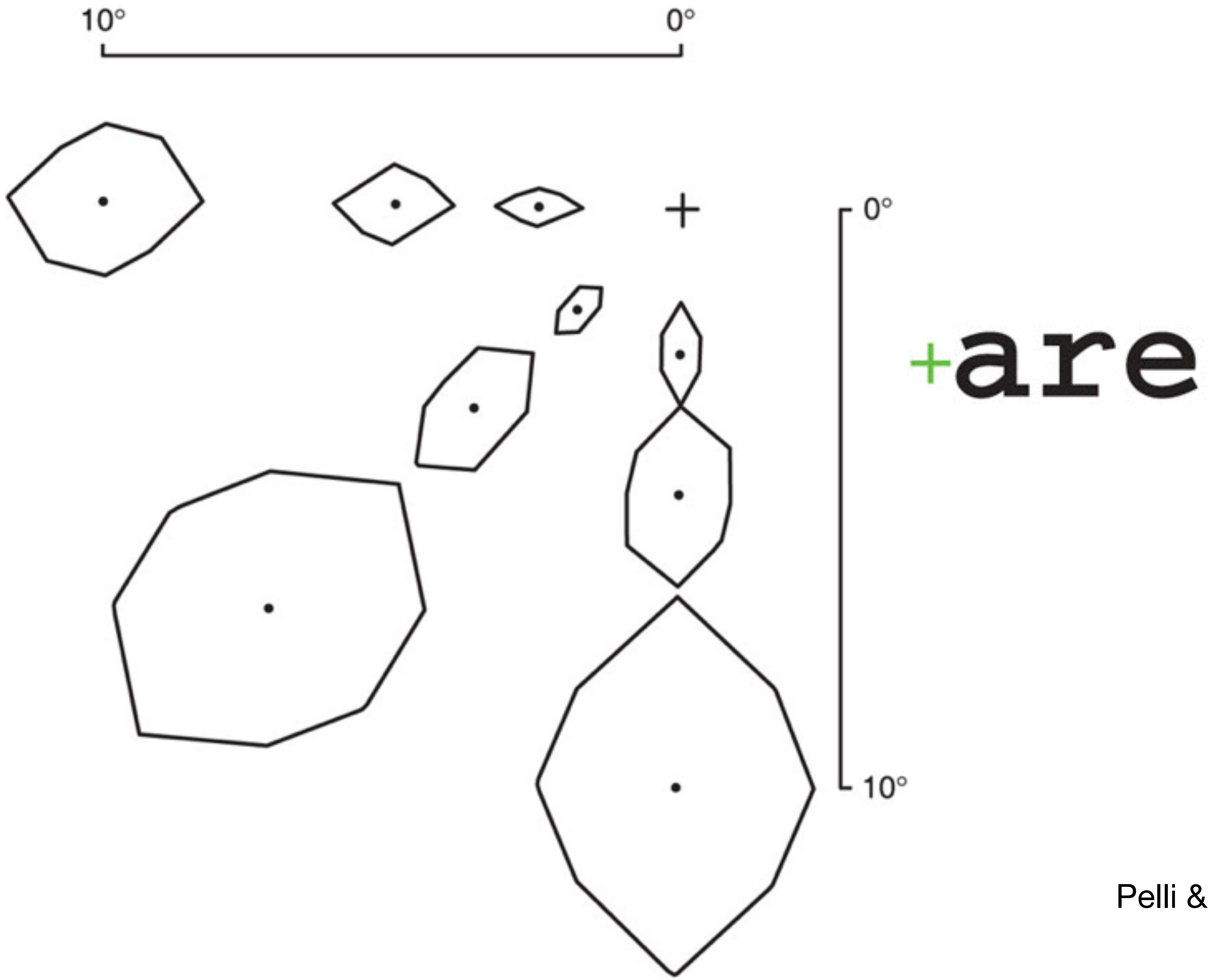
Spatial acuity and eccentricity



Spatial acuity and eccentricity



Perceptual crowding



Pelli & Tillman 2008

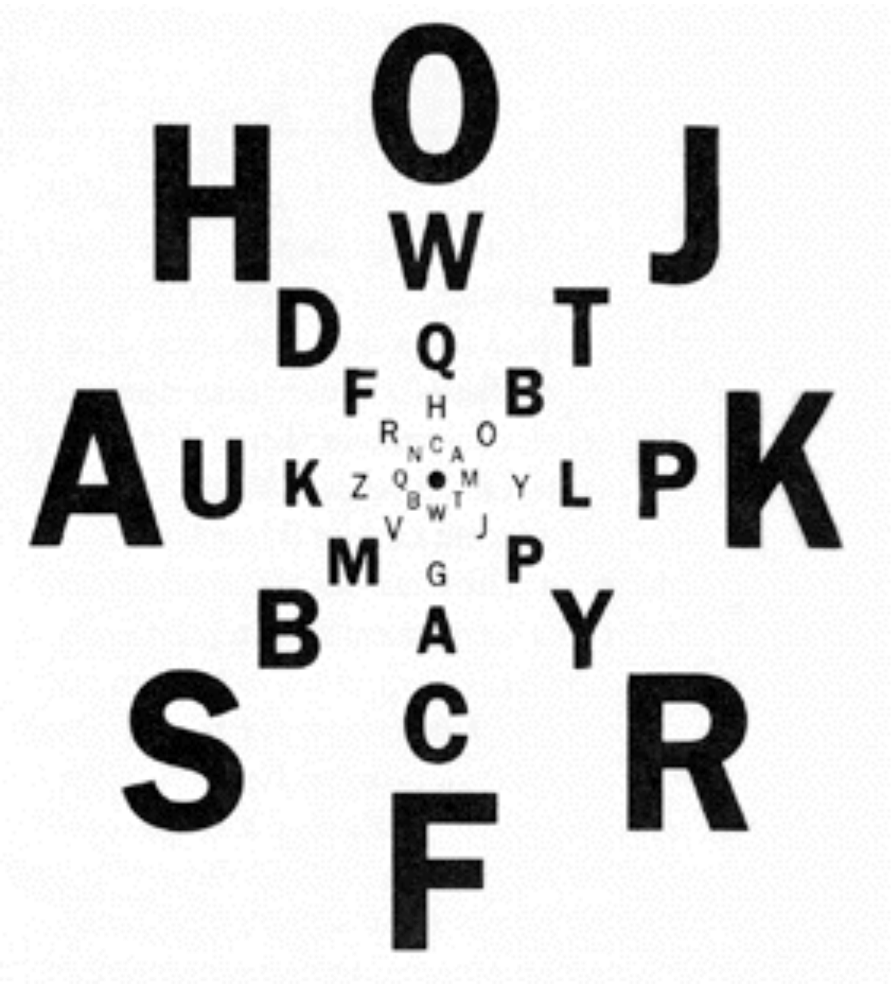
Spatial resolution and acuity

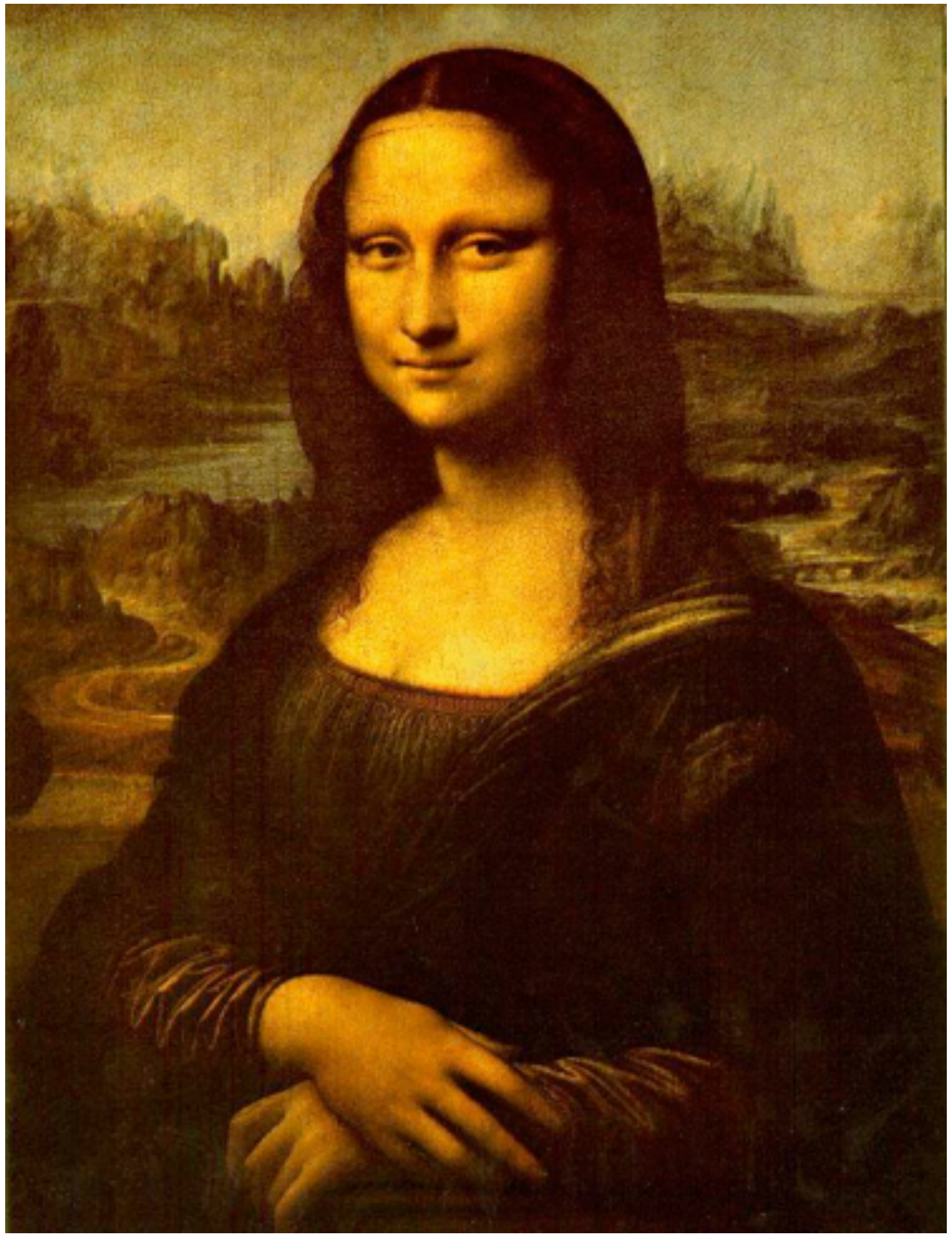
Renoir's Madame Henriot (1876)

Degas' Woman ironing (1869)

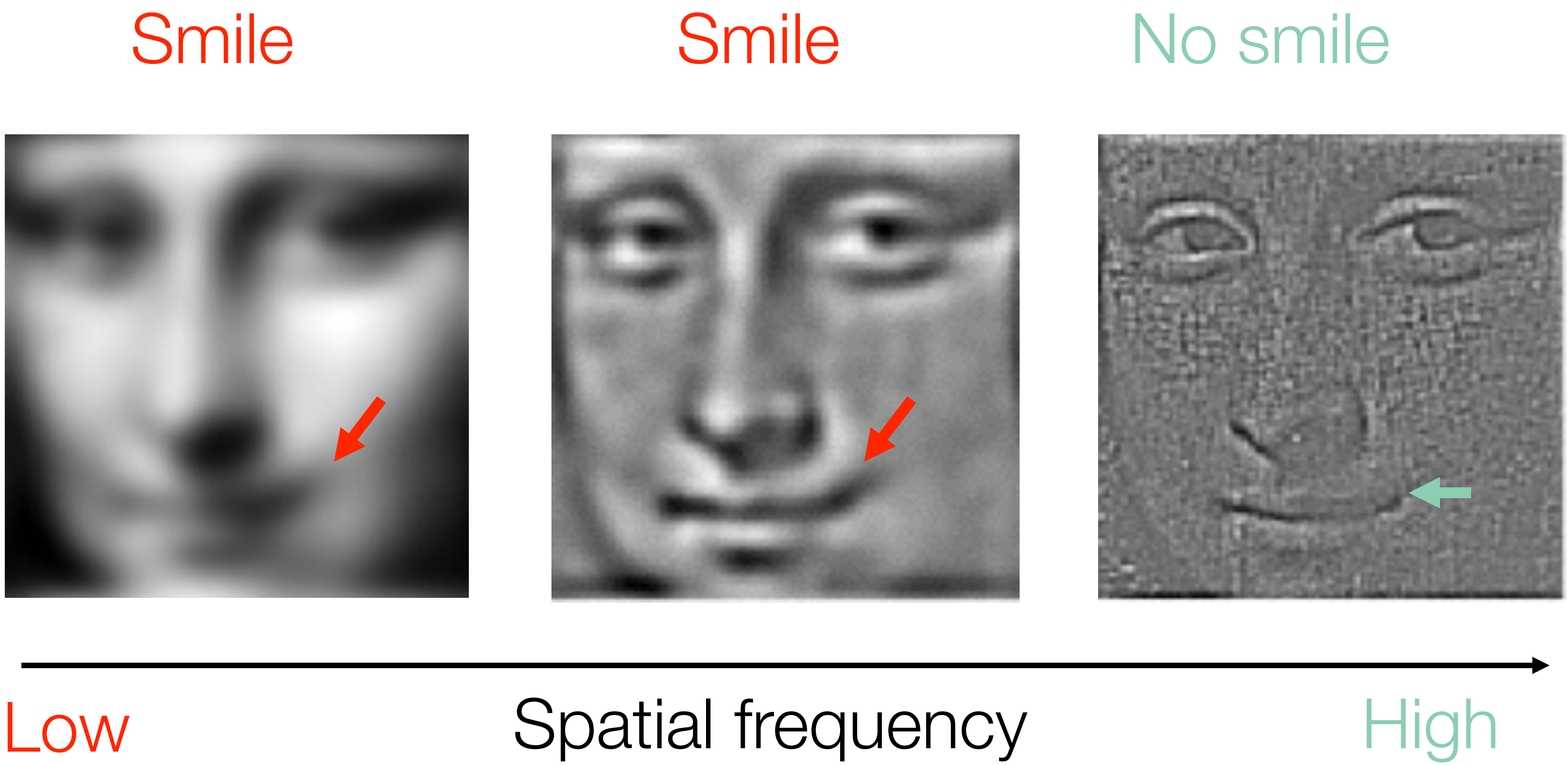
Ingres' Mrs Charles Badham (1816)

Spatial resolution and acuity





Subtle expression
Leonardo da Vinci’s Mona Lisa



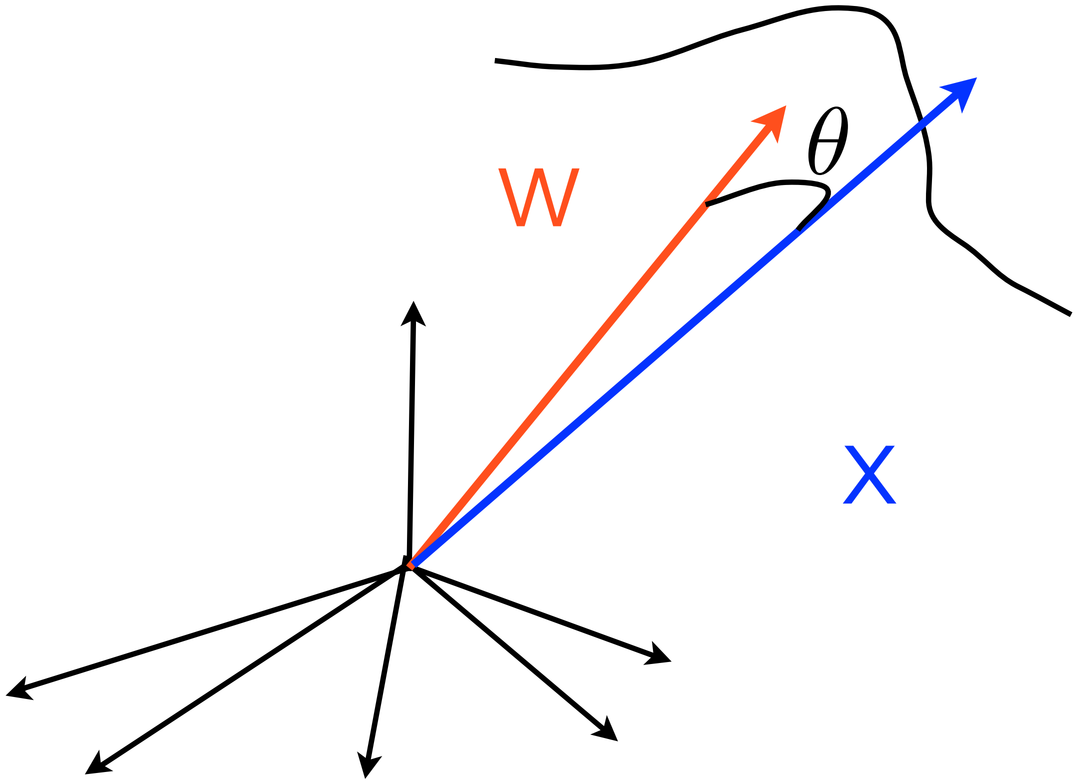
Computational Vision

Primary visual cortex

- Orientation selectivity
- Spatial frequency
- Normalization
- Color opponency

On the need for normalization circuits

$$
\begin{aligned}
y(x) &= \sum w_i x_i \\
&= \mathbf{w} \cdot \mathbf{x} \\
&= ||\mathbf{w}||\,||\mathbf{x}|| \cos(\theta)
\end{aligned}
$$

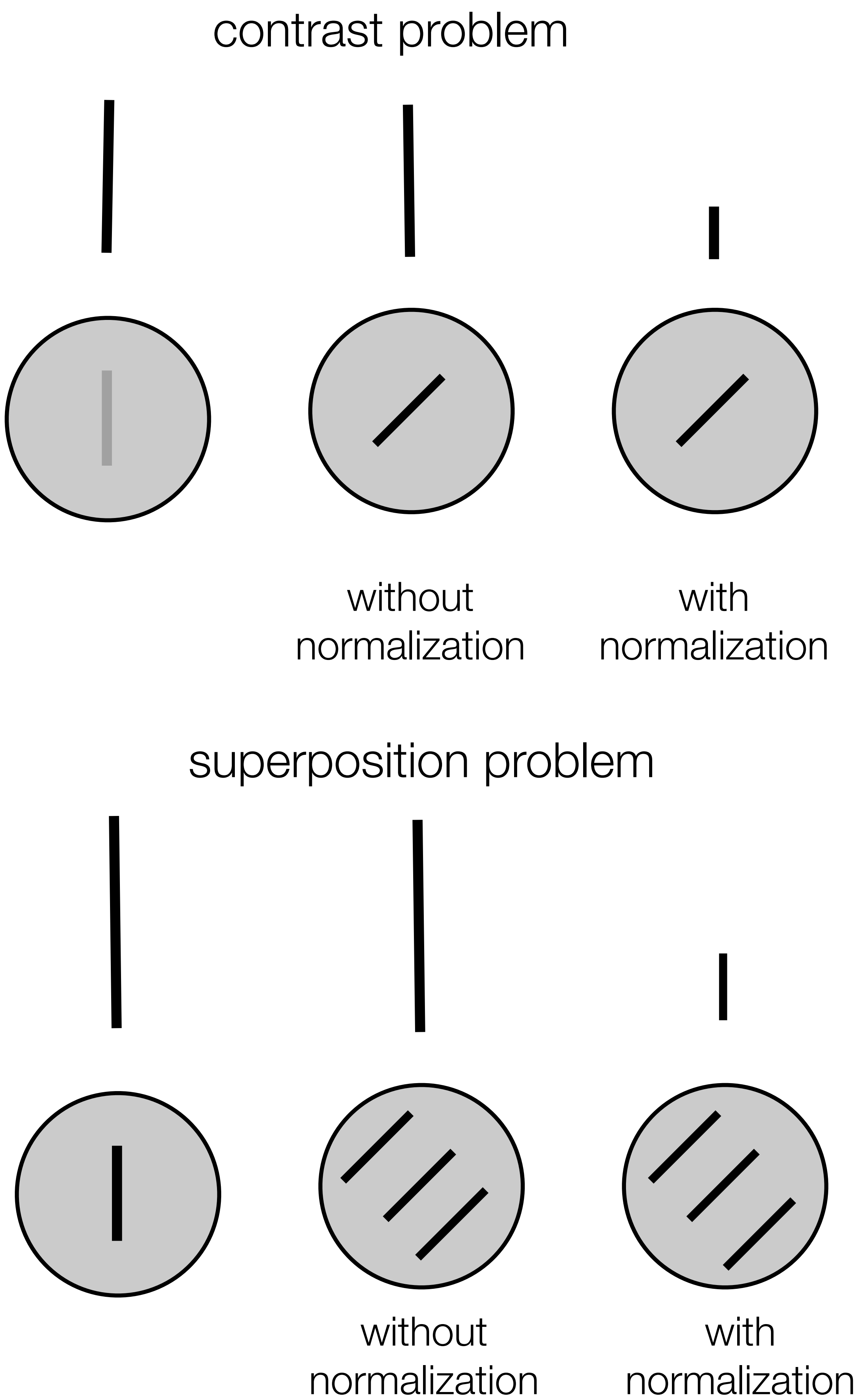


$$
\begin{aligned}
y(x) &= \sum w_i x_i \\
&= \frac{\mathbf{w} \cdot \mathbf{x}}{||\mathbf{w}||\,||\mathbf{x}||} \\
&= \cos(\theta)
\end{aligned}
$$

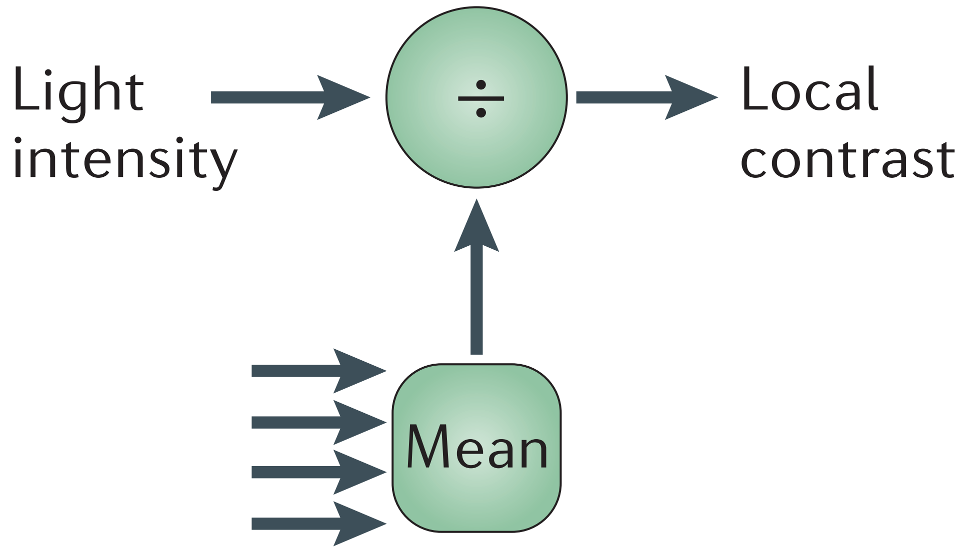
On the need for normalization circuits

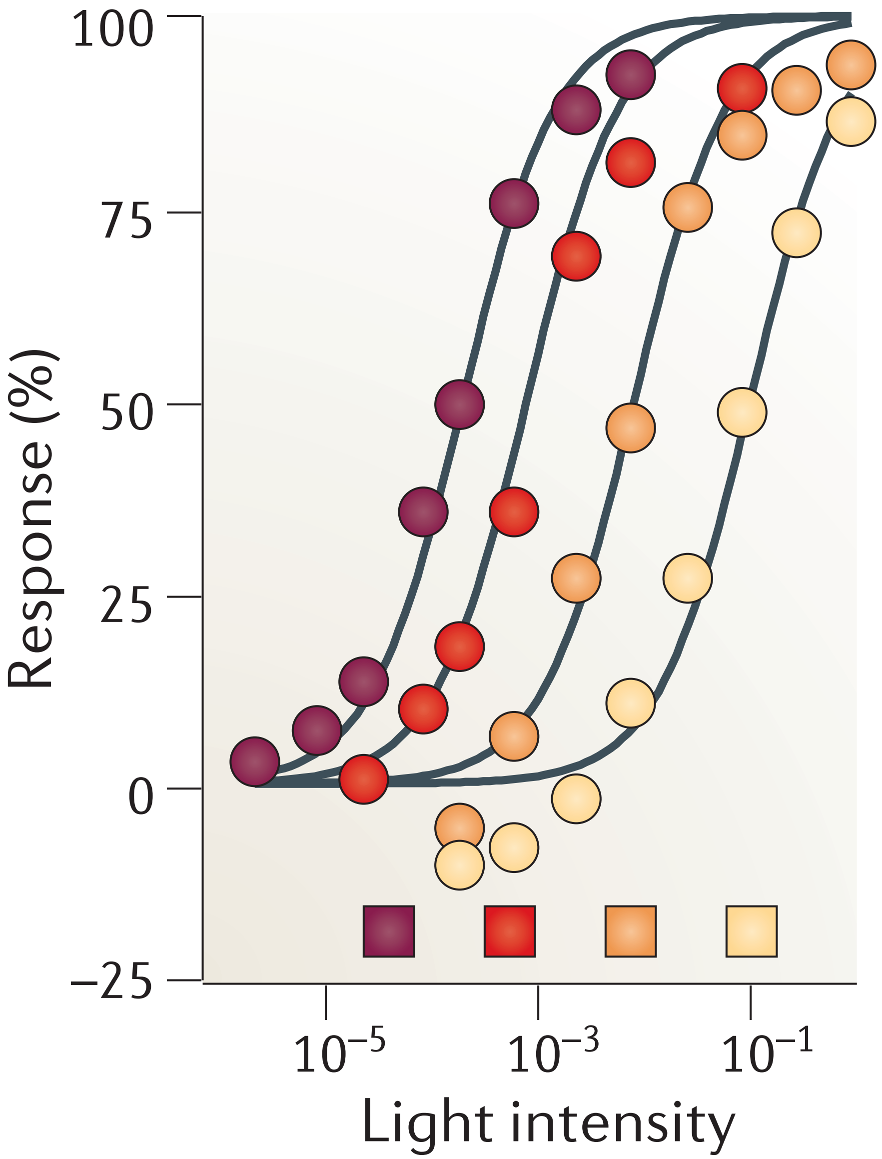
$$
\begin{aligned}
y(x) &= \sum w_i x_i \\
&= \mathbf{w} \cdot \mathbf{x} \\
&= ||\mathbf{w}|| \, ||\mathbf{x}|| \cos(\theta)
\end{aligned}
$$

$$
\begin{aligned}
y(x) &= \sum w_i x_i \\
&= \frac{\mathbf{w} \cdot \mathbf{x}}{||\mathbf{w}|| \, ||\mathbf{x}||} \\
&= \cos(\theta)
\end{aligned}
$$



Gain-control circuits in the (turtle) retina





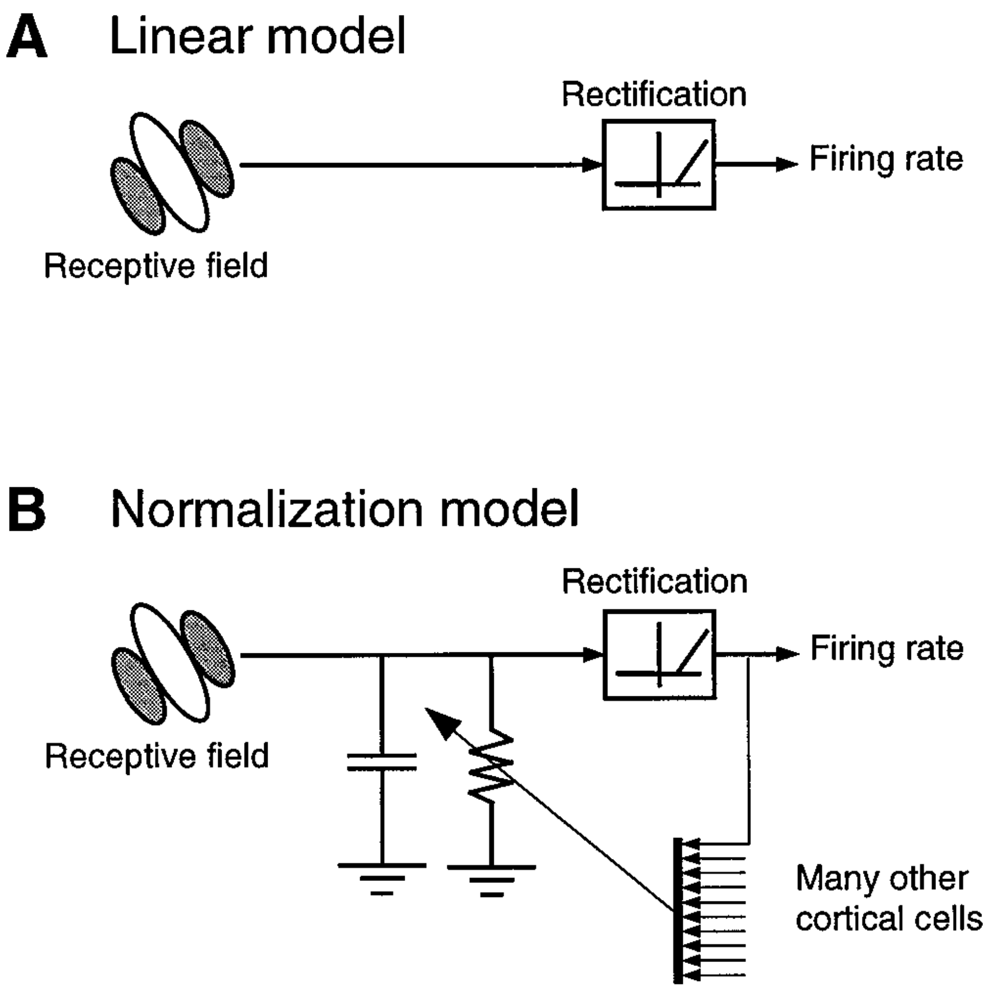
Carandini & Heeger 2011

Gain-control circuits

$$y(\mathbf{x}) = \frac{\mathbf{w}^* \cdot \mathbf{x}}{||\mathbf{x}|| + c}$$

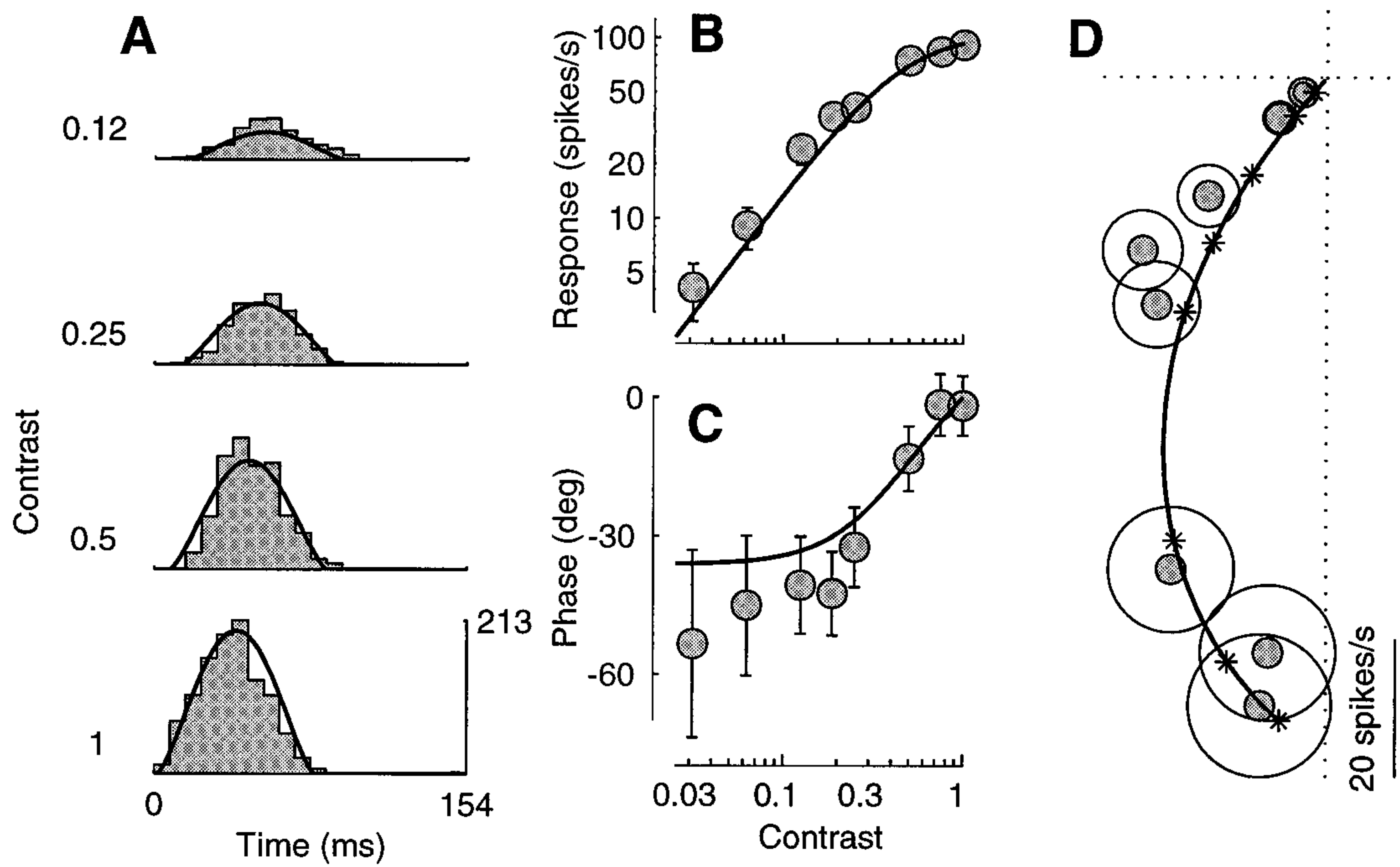
$$y(\alpha\mathbf{x}) = \frac{\mathbf{w}^* \cdot \alpha\mathbf{x}}{||\alpha\mathbf{x}|| + c}$$

$$= \frac{\alpha\mathbf{w}^* \cdot \mathbf{x}}{\alpha(||\mathbf{x}|| + c/\alpha)}$$



Carandini Heeger & Movshon '96

Gain-control circuits



Carandini et al 1997

Gain-control circuits

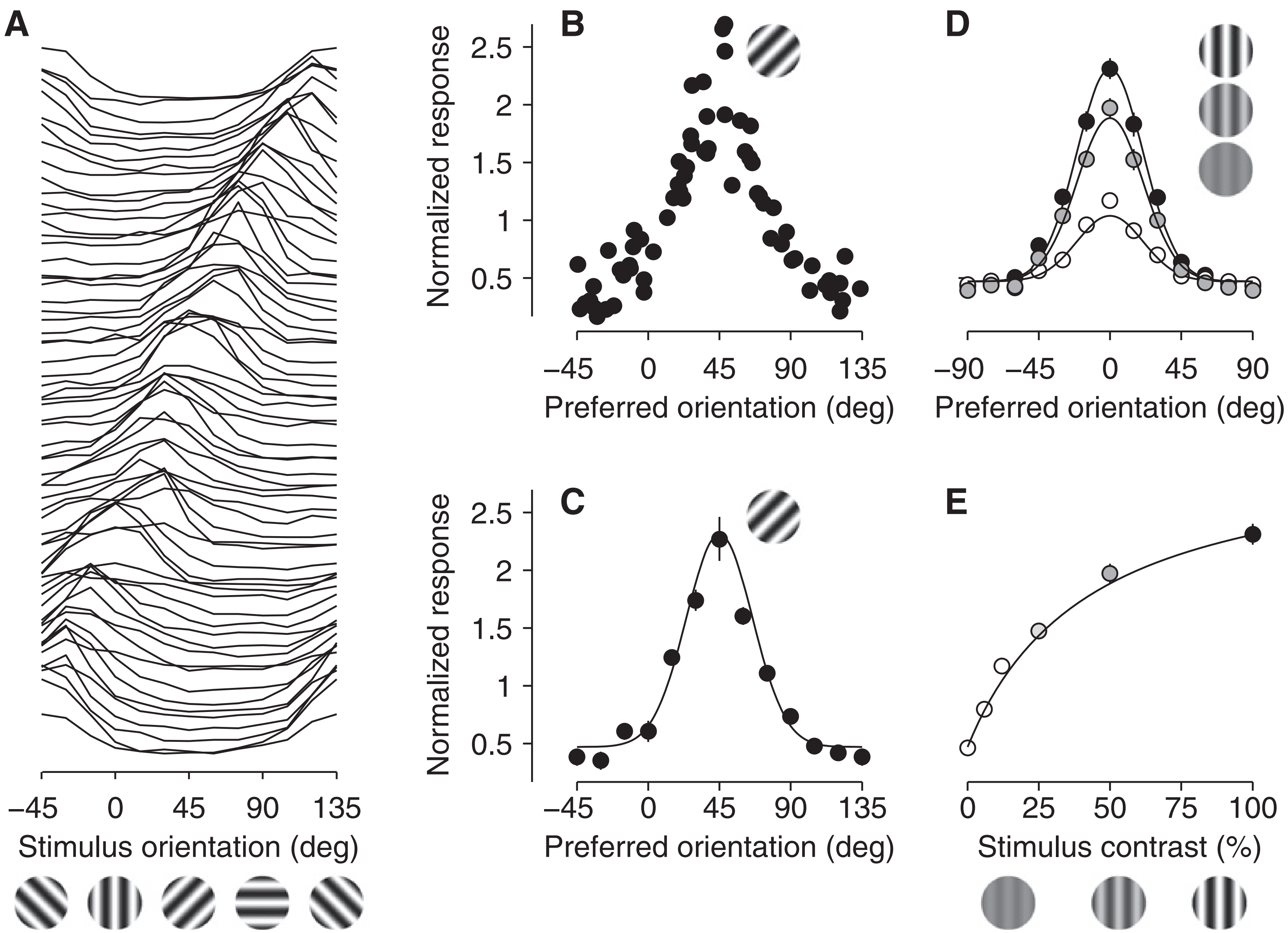


Figure 1. Tuning Curves and Population Responses to Single Orientation Stimuli

(A) Orientation tuning curves of all responsive sites (66 of 96), sorted according to preferred orientation. Each tuning curve is normalized by its mean across orientations.

(B) Population response to a 45° stimulus: responses of all sites (dots) as a function of preferred orientation of each site.

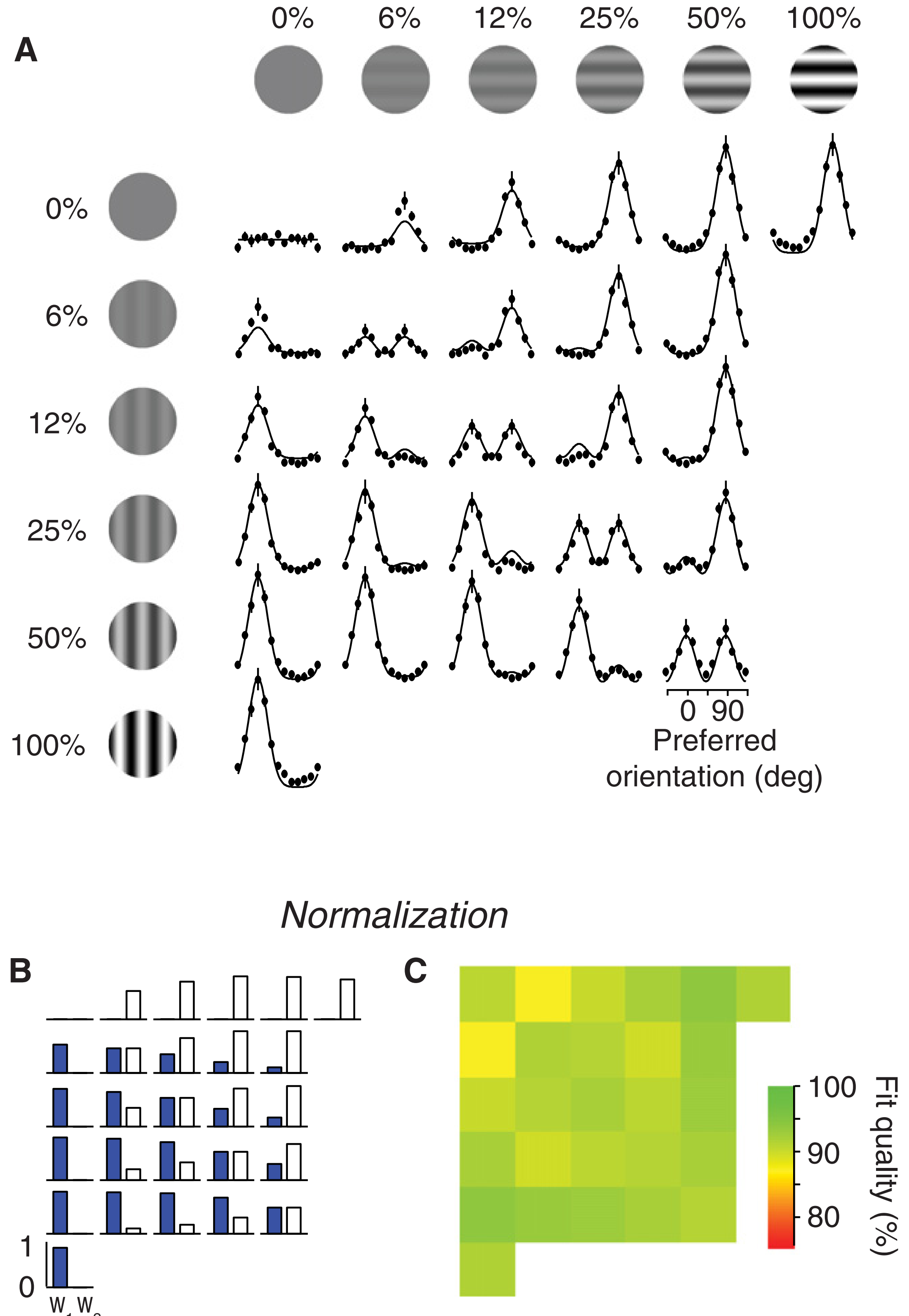
(C) The population response in (B) after binning sites with similar orientation preference (bin width: 15°). The curve is the best fitting circular Gaussian. Error bars indicate ±1 SE of responses across sites in each bin.

(D) Population response to a 0° stimulus for three contrasts: 12%, 50%, and 100%. The abscissa indicates preferred orientation relative to stimulus orientation. Data for stimuli of multiple orientations (0°, 30°, and 60°) are combined to obtain each population response. Error bars indicate ±1 SE of responses across sites in each bin. The curves fitting the data are circular Gaussians differing only in amplitude.

(E) Amplitude of the population responses as a function of stimulus contrast. The curve is the best-fitting hyperbolic ratio function (c_{50} = 42.1%, n = 1.0). All fits are given by Equation (1). Experiment 84-12-16.

Busse et al 2008

Gain-control circuits



Busse et al 2008

Gain control circuits in computer vision!

What is the Best Multi-Stage Architecture for Object Recognition?

Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ranzato and Yann LeCun
The Courant Institute of Mathematical Sciences
New York University, 715 Broadway, New York, NY 10003, USA
koray@cs.nyu.edu

- Normalization appears to be the most important component of a good computer vision system (over learning algorithms, architecture, etc)