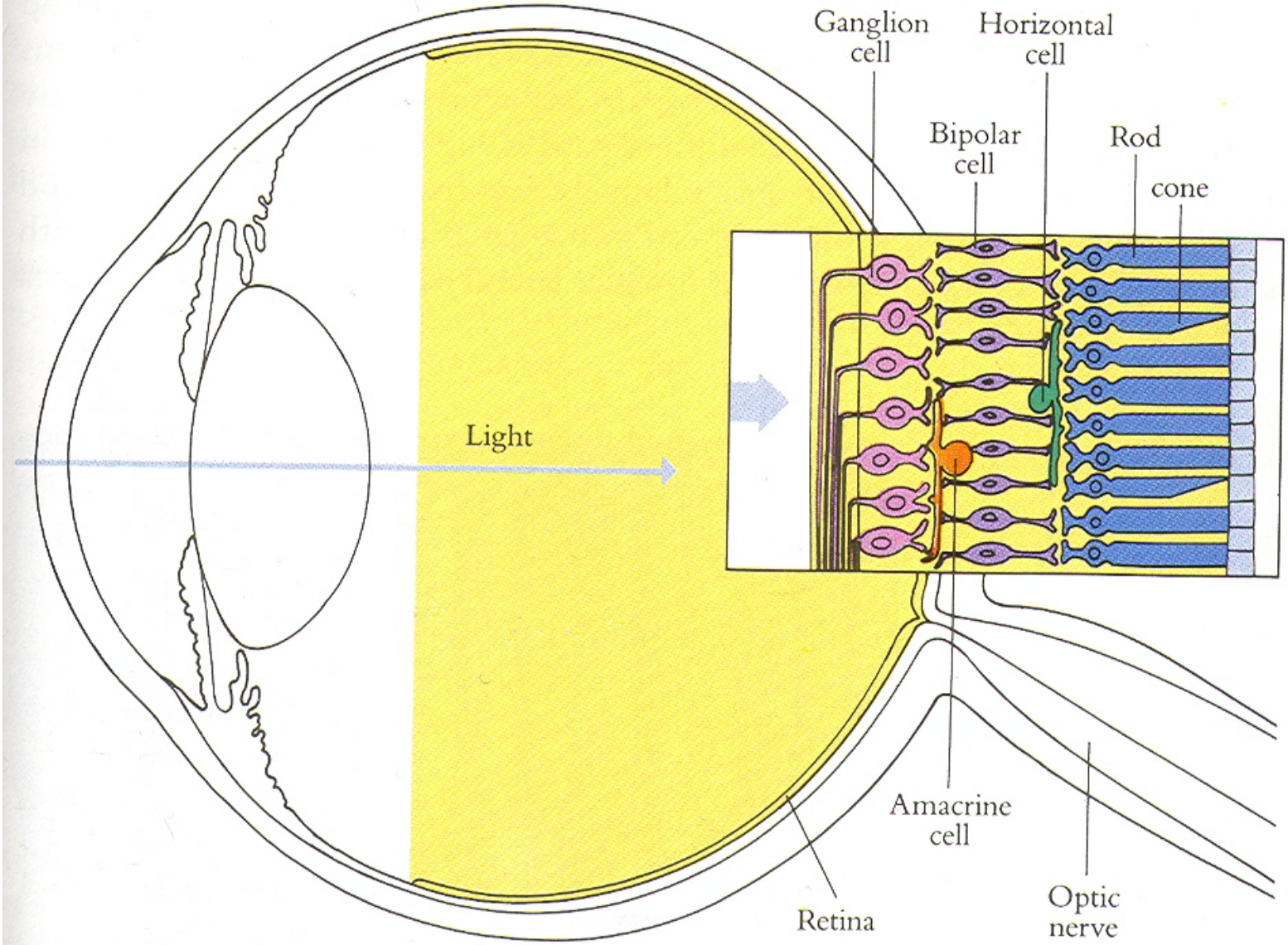
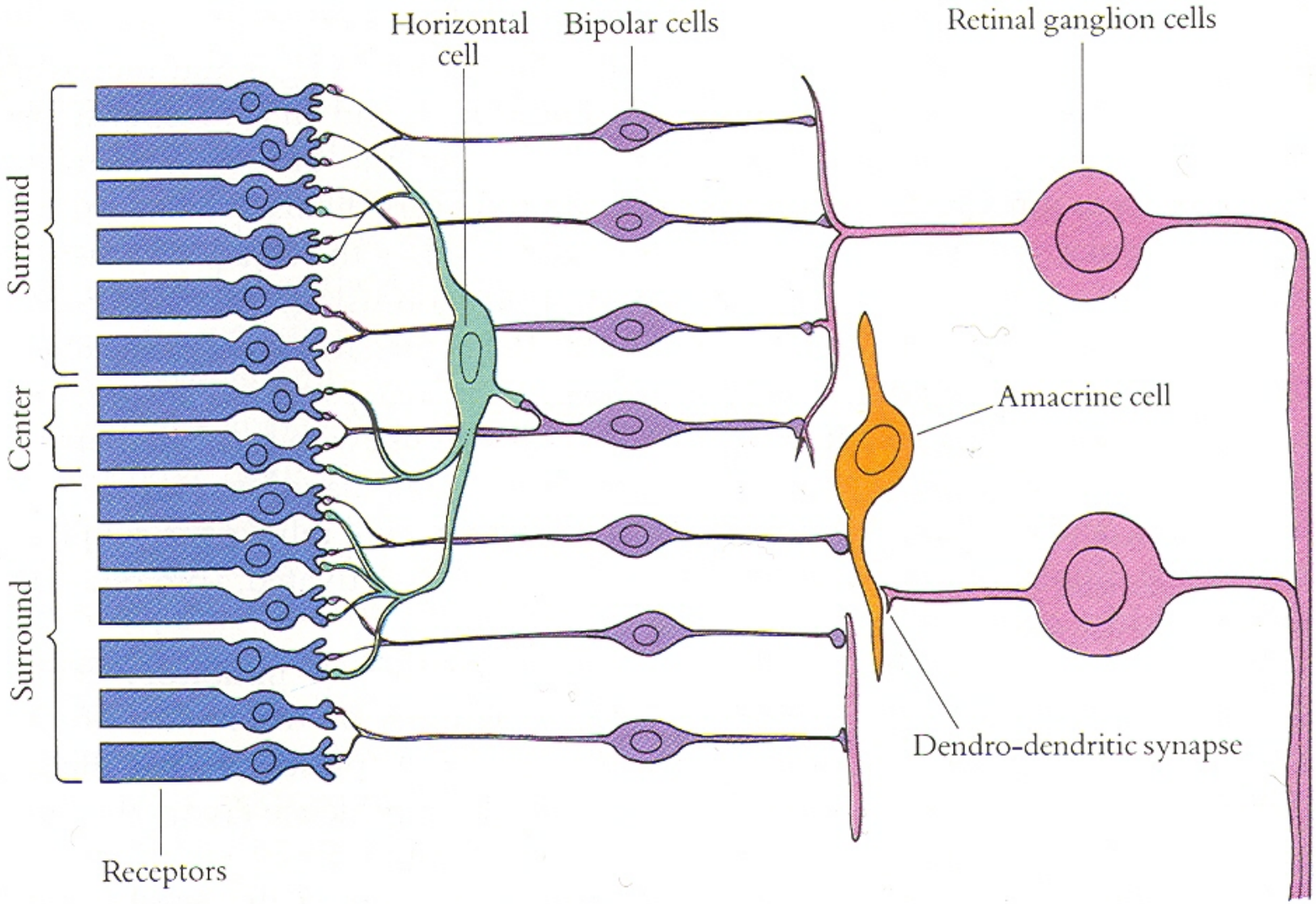
Inspiration from Biological Vision

© Quint Buchholz

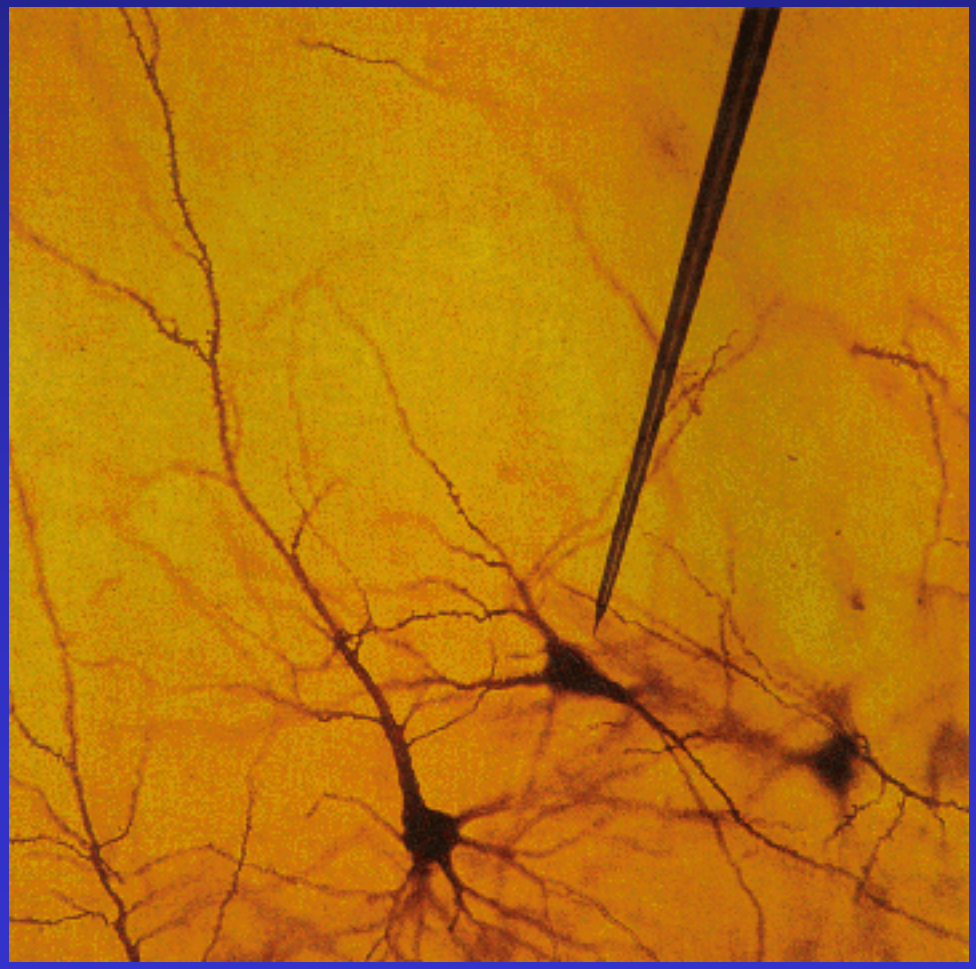
Ways to understand human vision

- Psychophysics
  - Present humans/animals with stimuli
  - Quantitatively measure perceptual response
  - E.g. Julesz’s Visual Search experiments
- Electrophysiology
  - Stick probes into brains
  - Present stimuli
  - Quantitatively measure electrical response
- Imaging (e.g. fMRI)


Ganglion cell
Horizontal cell
Bipolar cell
Rod
cone
Light
Amacrine cell
Retina
Optic nerve


Horizontal cell
Bipolar cells
Retinal ganglion cells
Surround
Center
Surround
Amacrine cell
Dendro-dendritic synapse
Receptors

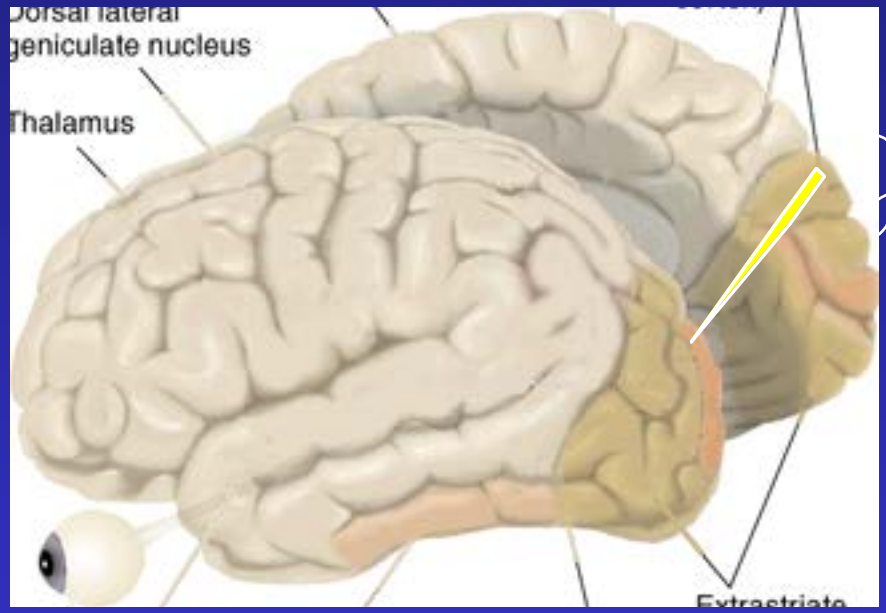
Single Cell Recording



© Stephen E. Palmer, 2002

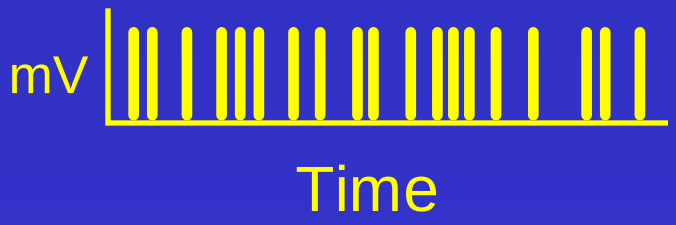
Single Cell Recording

Microelectrode



Amplifier

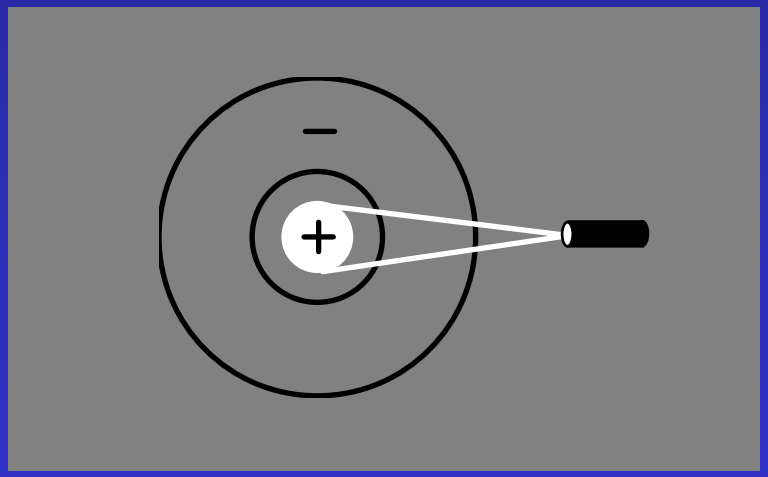
Electrical response
(action potentials)

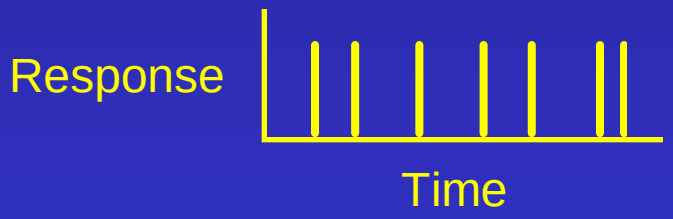


© Stephen E. Palmer, 2002

Retinal Receptive Fields

Receptive field structure in ganglion cells:
On-center Off-surround





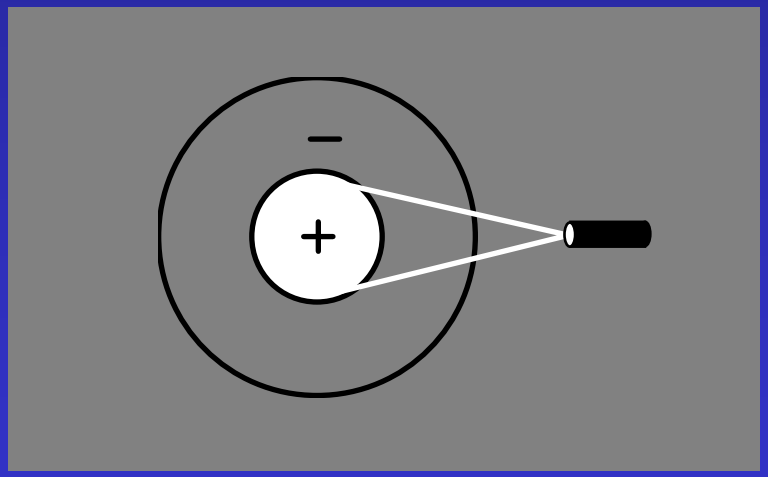
Stimulus condition

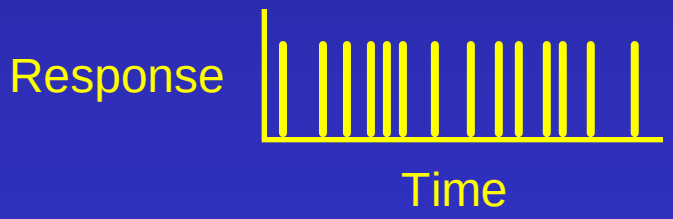
Electrical response

© Stephen E. Palmer, 2002

Retinal Receptive Fields

Receptive field structure in ganglion cells:
On-center Off-surround





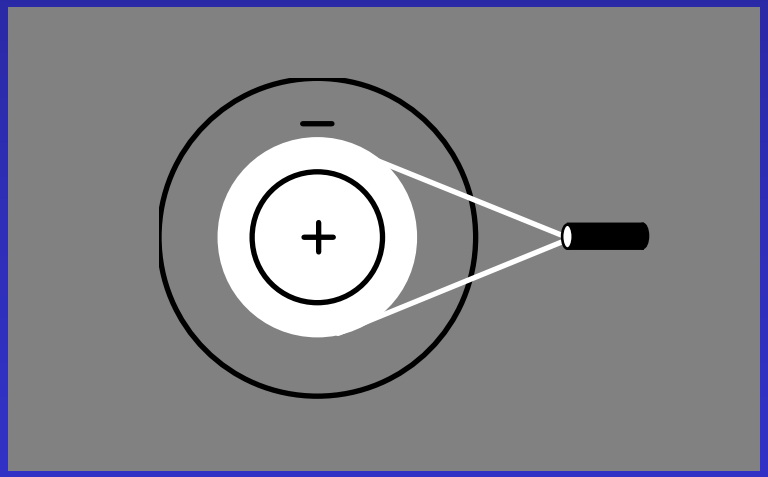
Stimulus condition

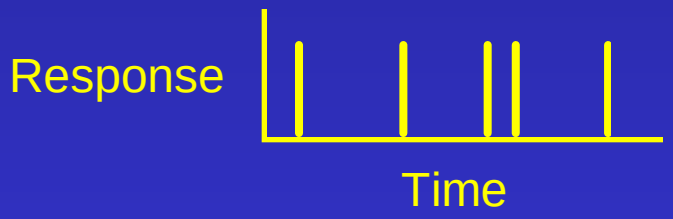
Electrical response

© Stephen E. Palmer, 2002

Retinal Receptive Fields

Receptive field structure in ganglion cells:
On-center Off-surround





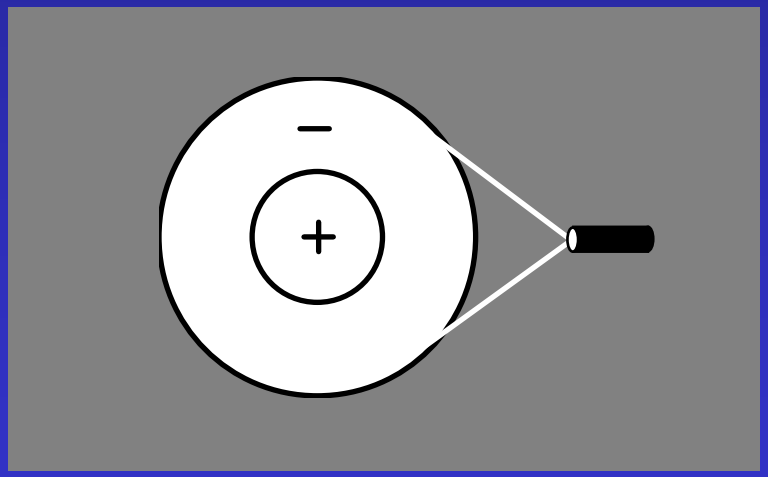
Stimulus condition

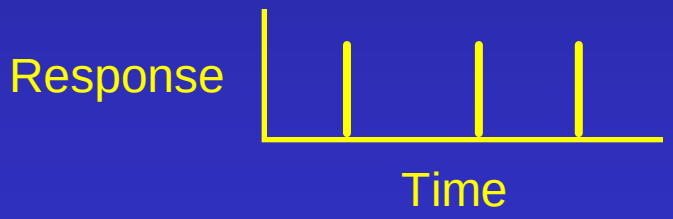
Electrical response

© Stephen E. Palmer, 2002

Retinal Receptive Fields

Receptive field structure in ganglion cells:
On-center Off-surround





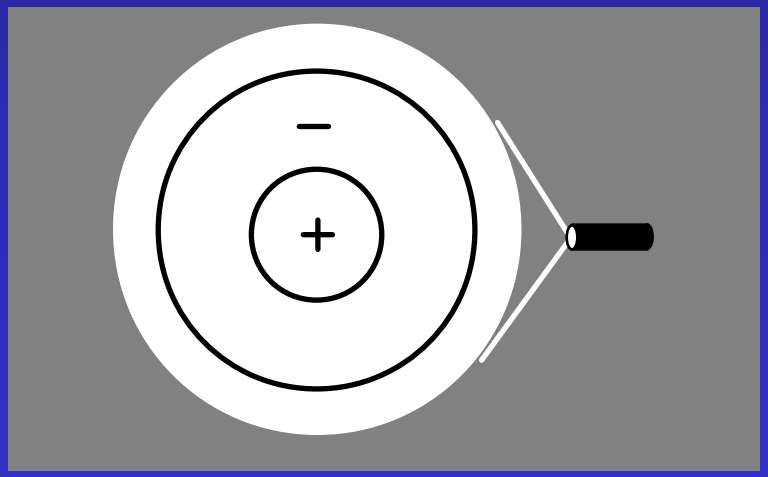
Stimulus condition

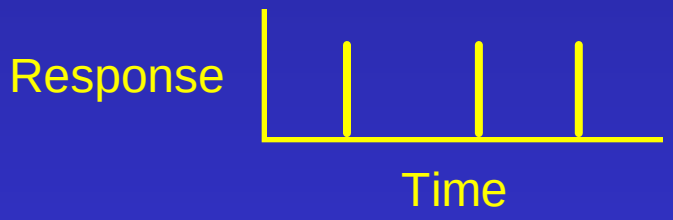
Electrical response

© Stephen E. Palmer, 2002

Retinal Receptive Fields

Receptive field structure in ganglion cells:
On-center Off-surround





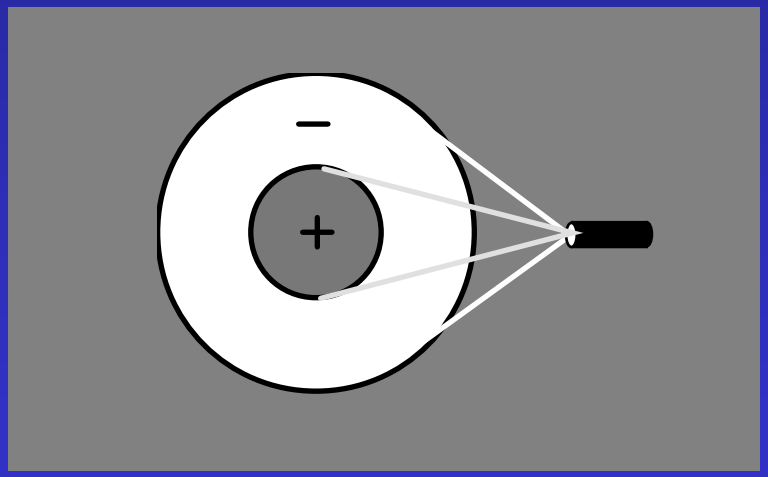
Stimulus condition

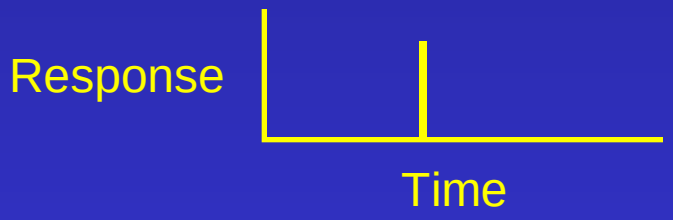
Electrical response

© Stephen E. Palmer, 2002

Retinal Receptive Fields

Receptive field structure in ganglion cells:
On-center Off-surround





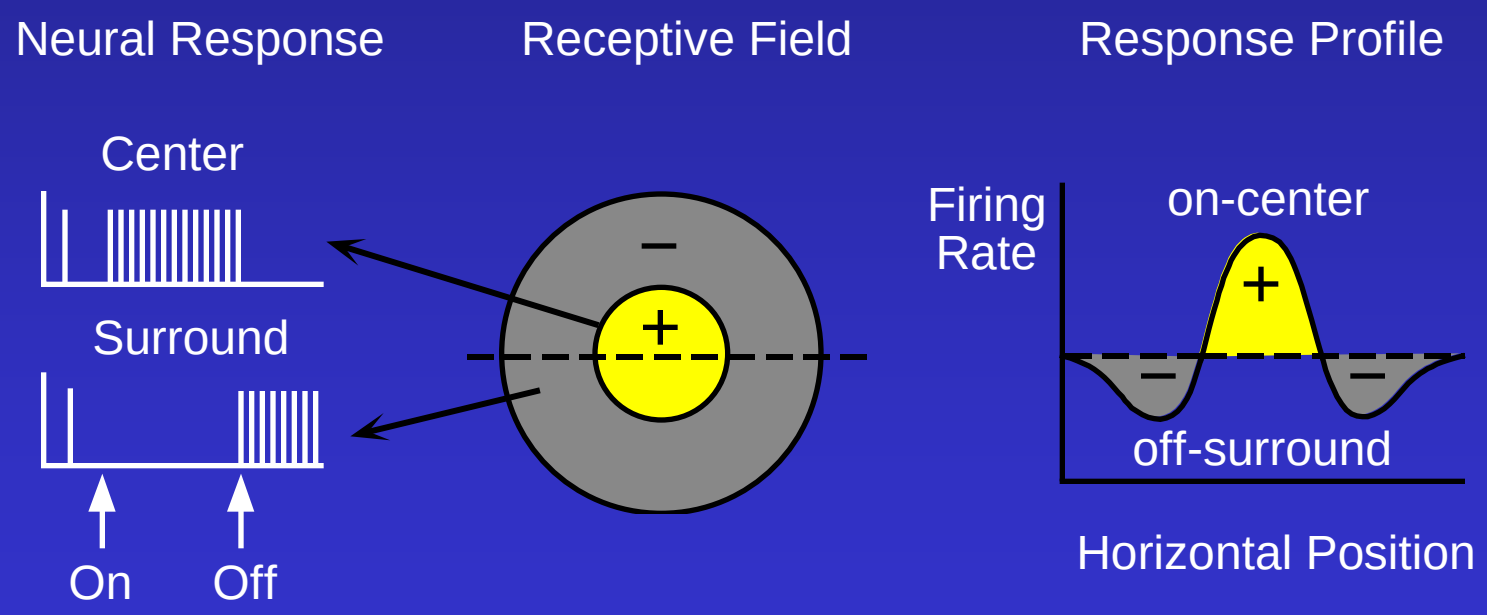
Stimulus condition

Electrical response

© Stephen E. Palmer, 2002

Retinal Receptive Fields

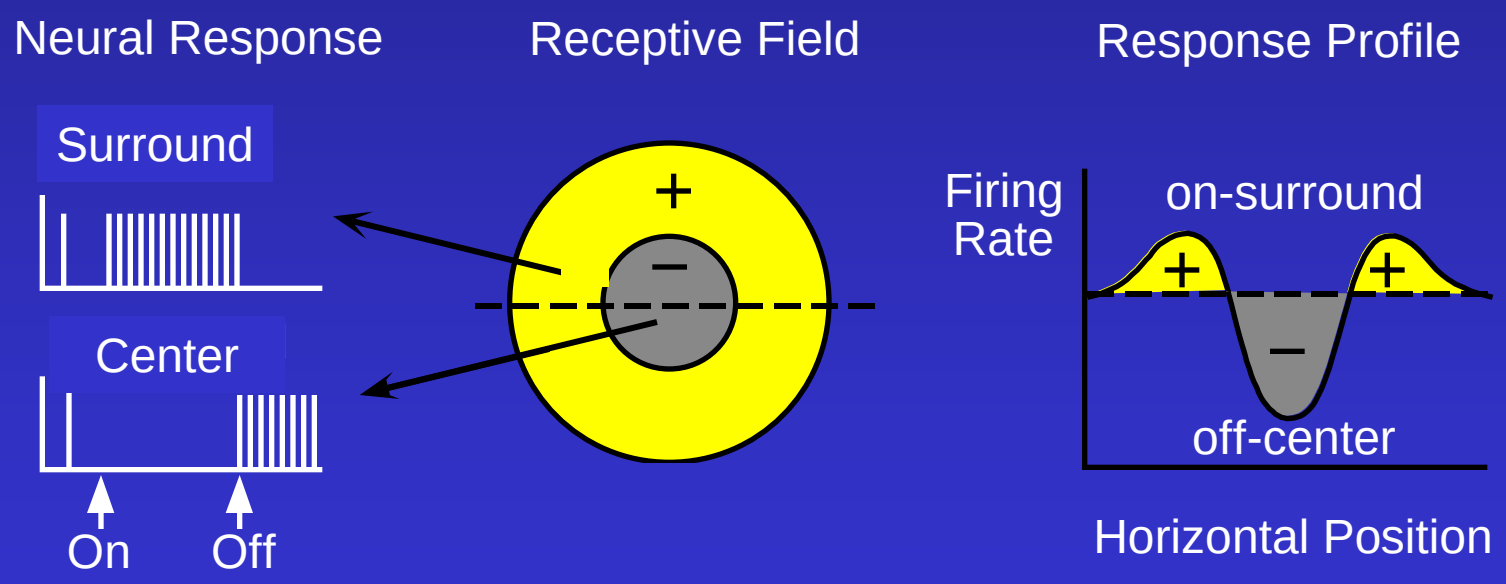
RF of On-center Off-surround cells



© Stephen E. Palmer, 2002

Retinal Receptive Fields

RF of Off-center On-surround cells



© Stephen E. Palmer, 2002

Receptive Fields

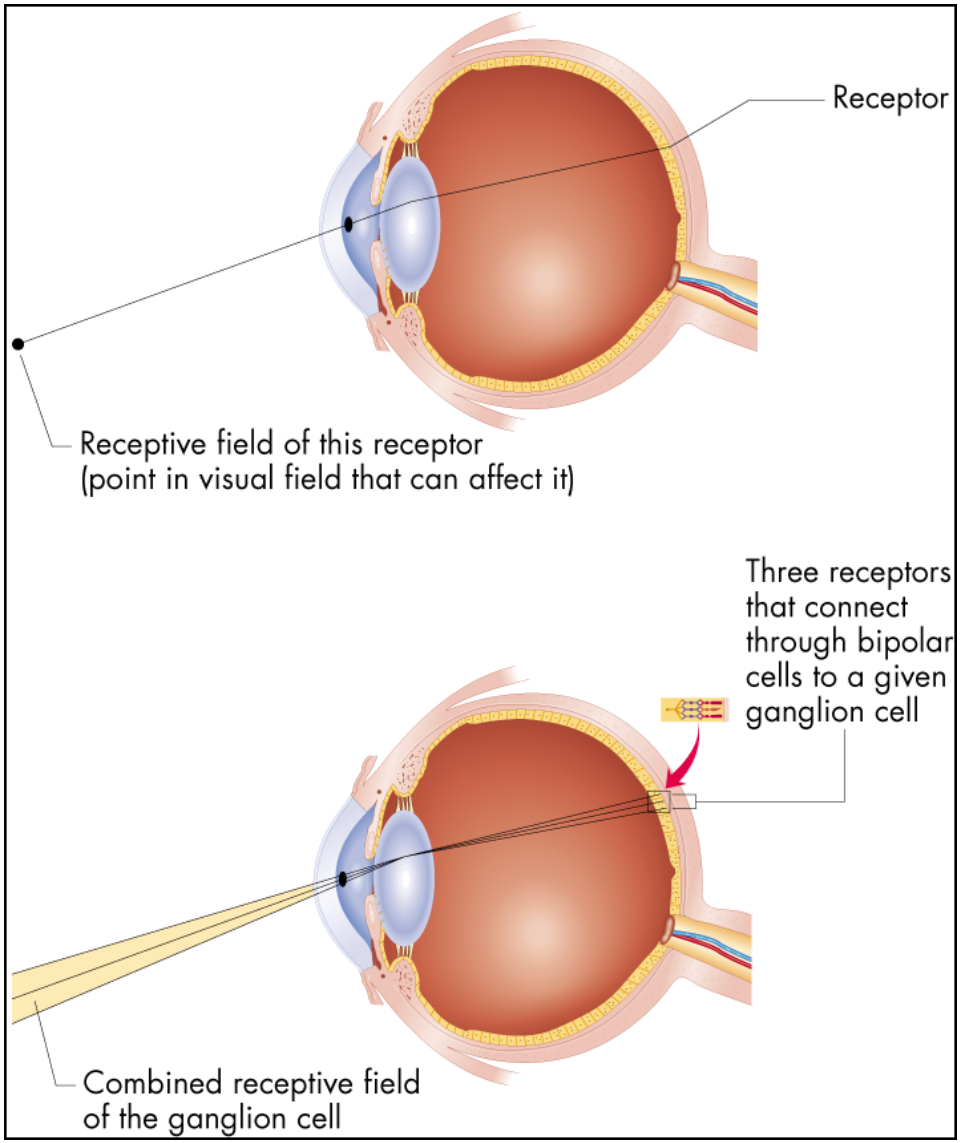
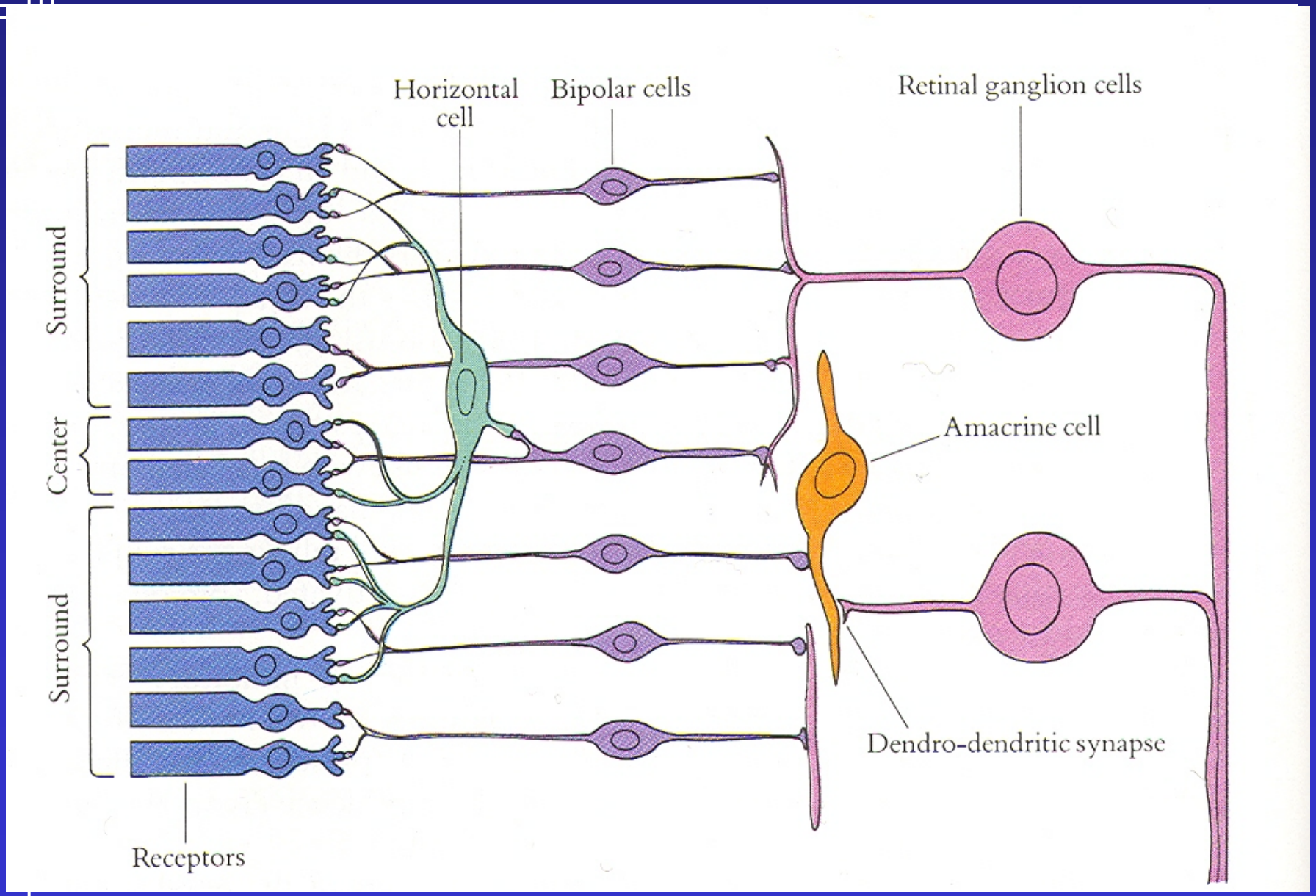


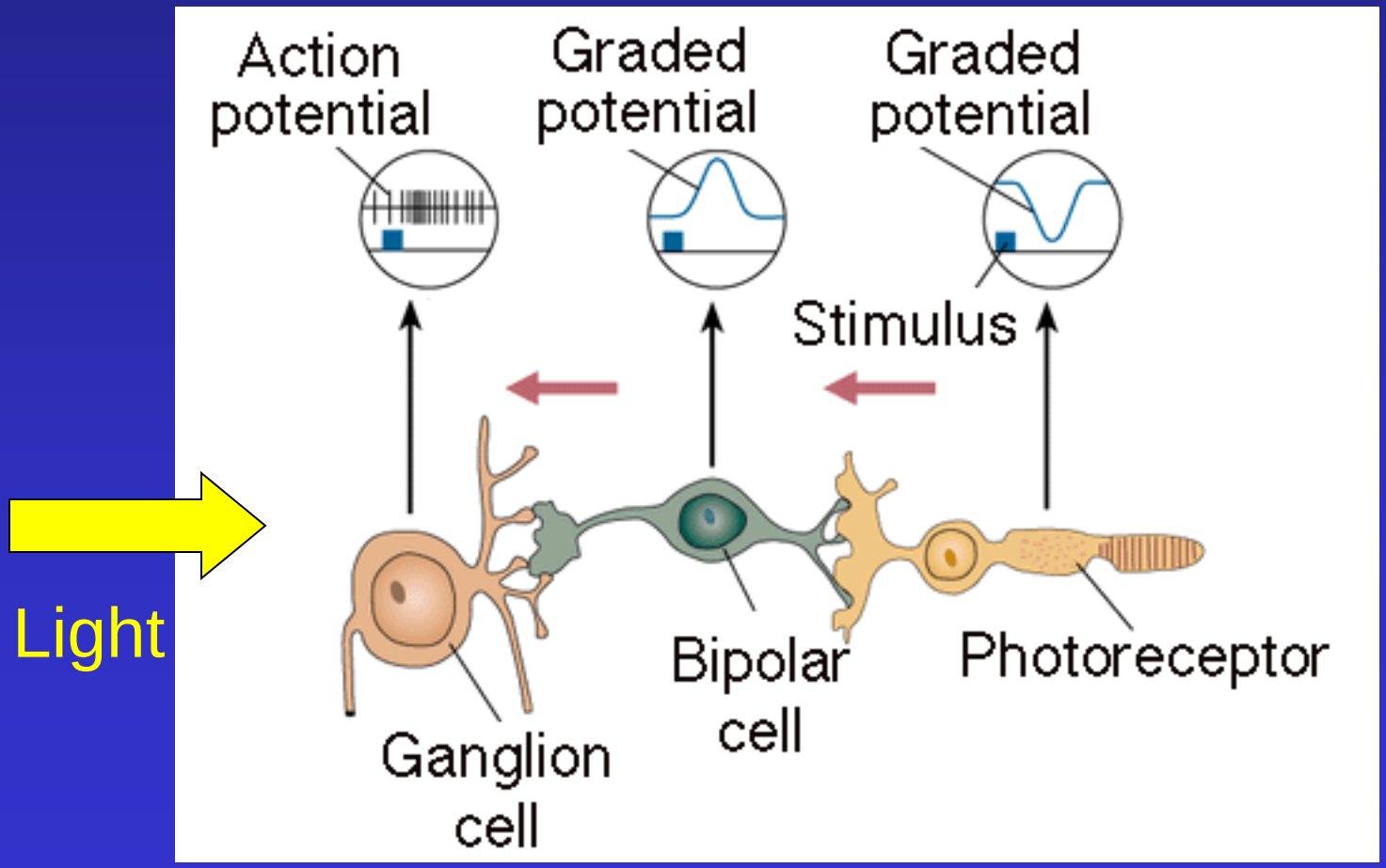
Figure 6.16 Receptive fields
The receptive field of a receptor is simply the area of the visual field from which light strikes that receptor. For any other cell in the visual system, the receptive field is determined by which receptors connect to the cell in question.

Retinal Receptive Fields



Retinal Receptive Fields

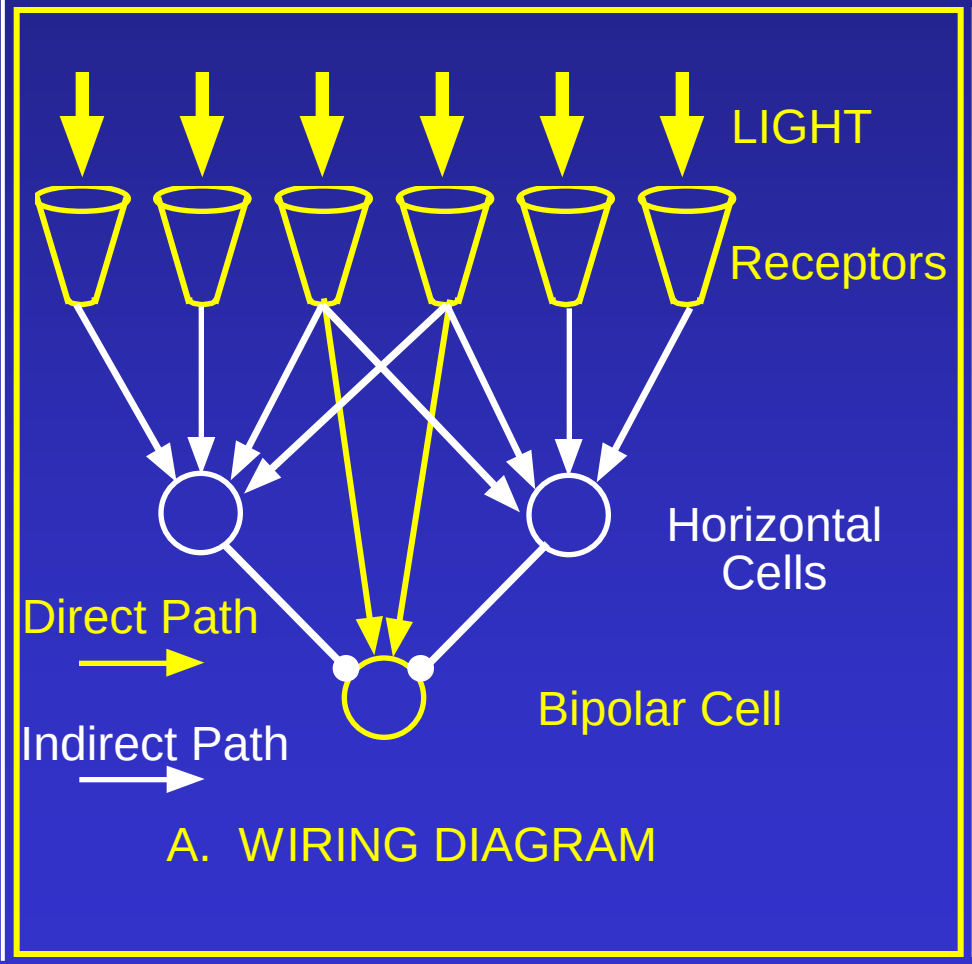
Receptive field structure in bipolar cells

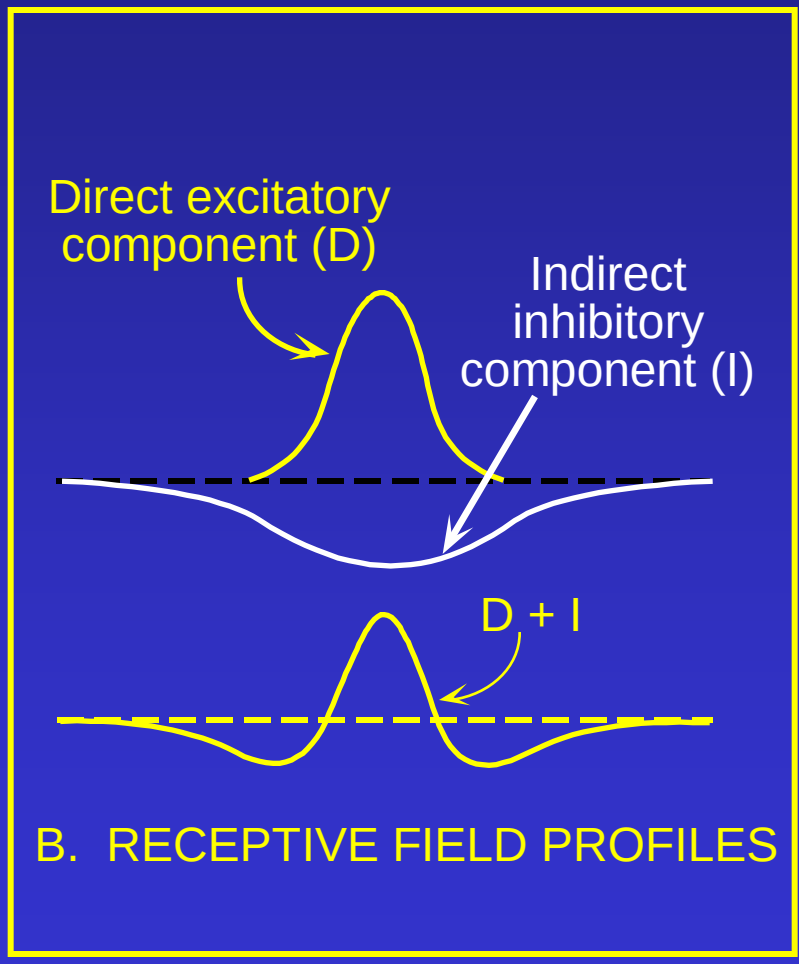


© Stephen E. Palmer, 2002

Retinal Receptive Fields

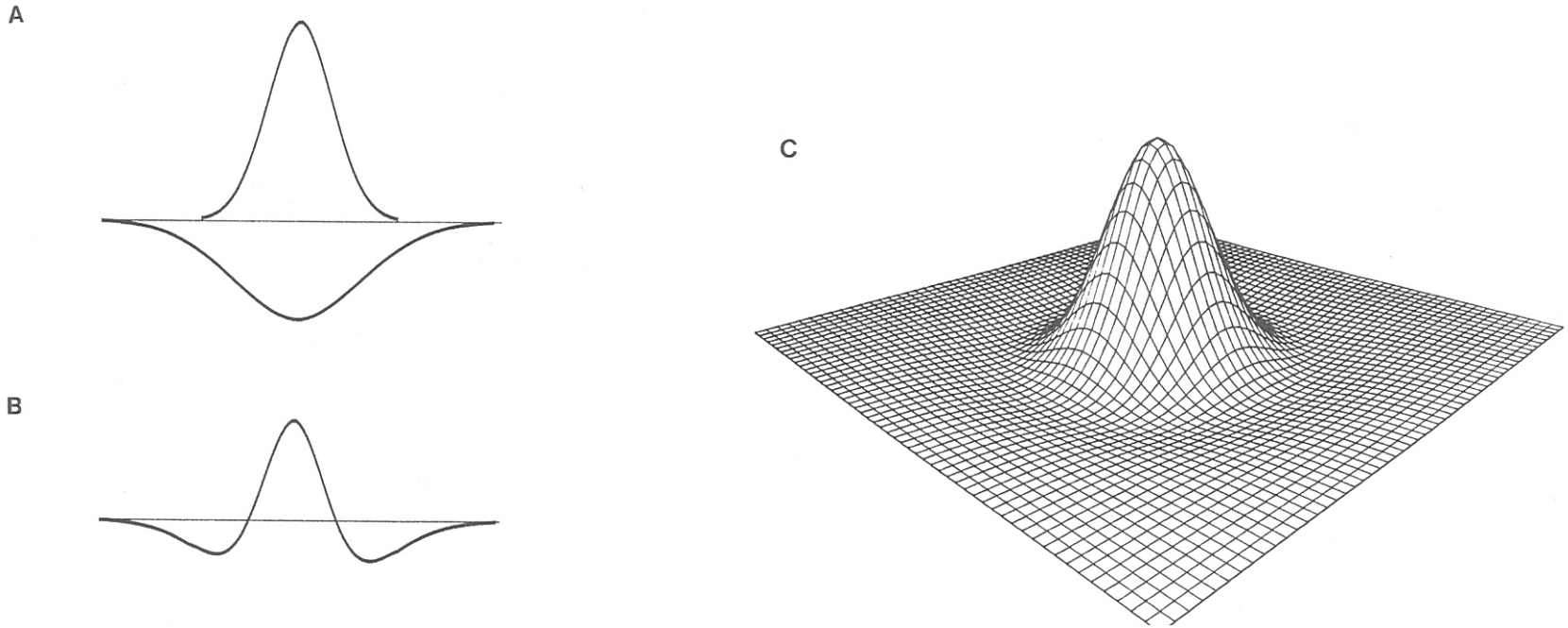
Receptive field structure in bipolar cells





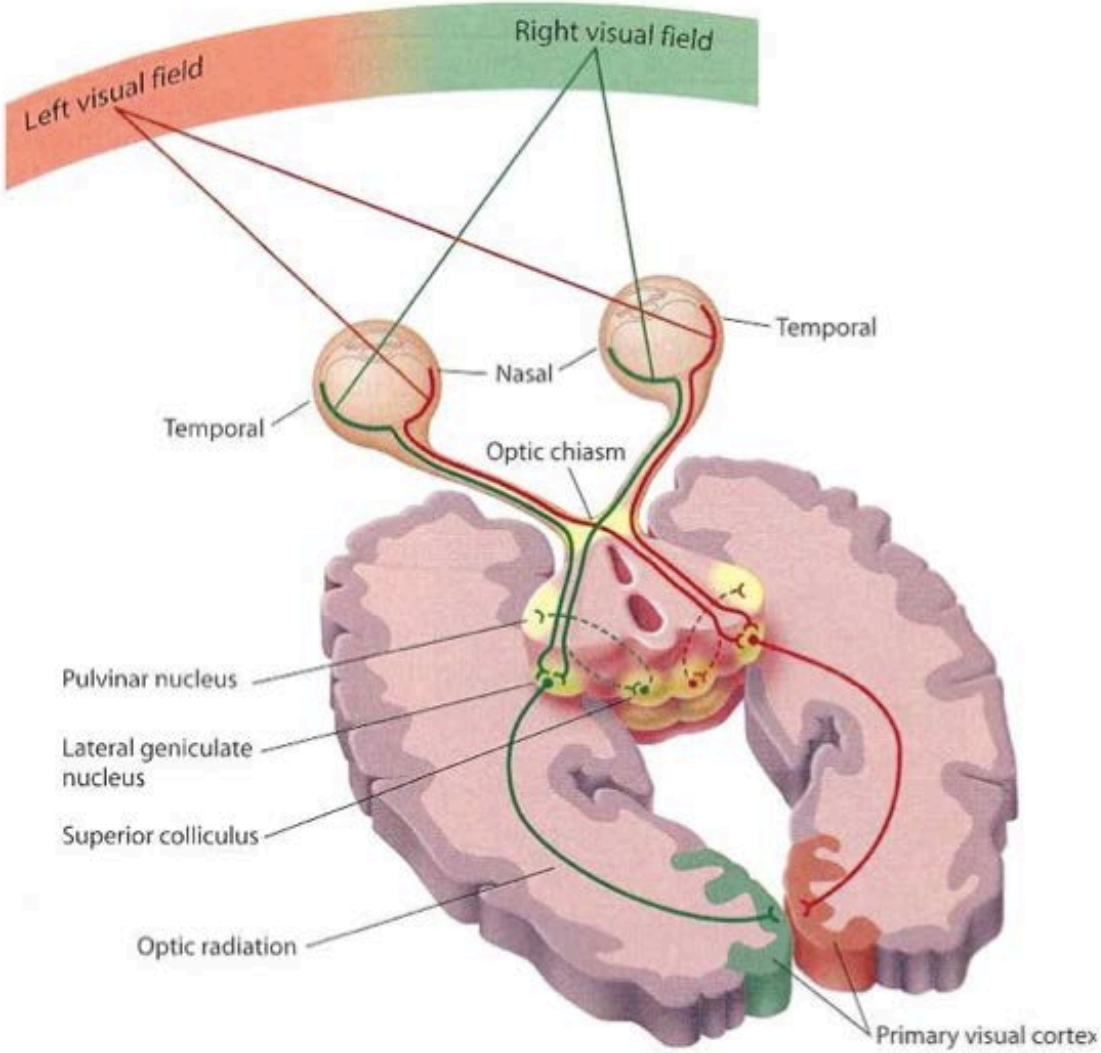
© Stephen E. Palmer, 2002

The receptive field of a retinal ganglion cell can be modeled as a “Difference of Gaussians”



$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

Anatomy of Pathway to Visual Cortex

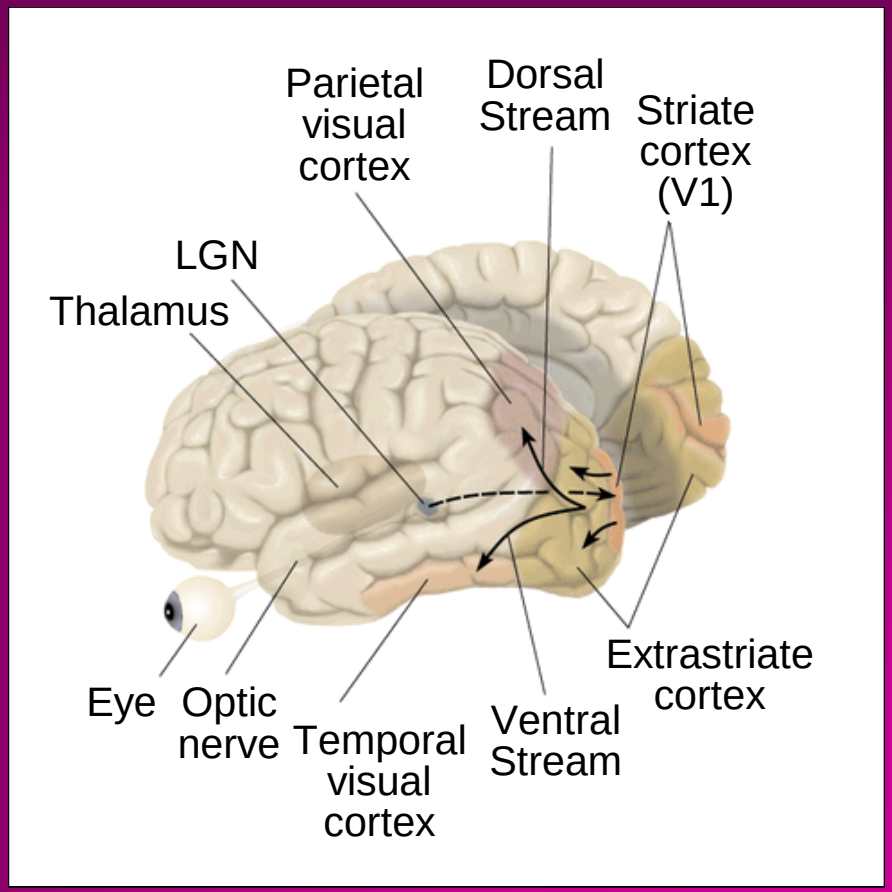


© Stephen E. Palmer, 2002

Visual Cortex

Cortical Area V1

aka:
Primary visual cortex
Striate cortex
Brodman's area 17



© Stephen E. Palmer, 2002

Cortical Receptive Fields

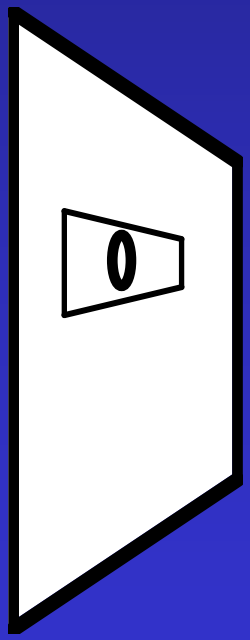
Single-cell recording from visual cortex

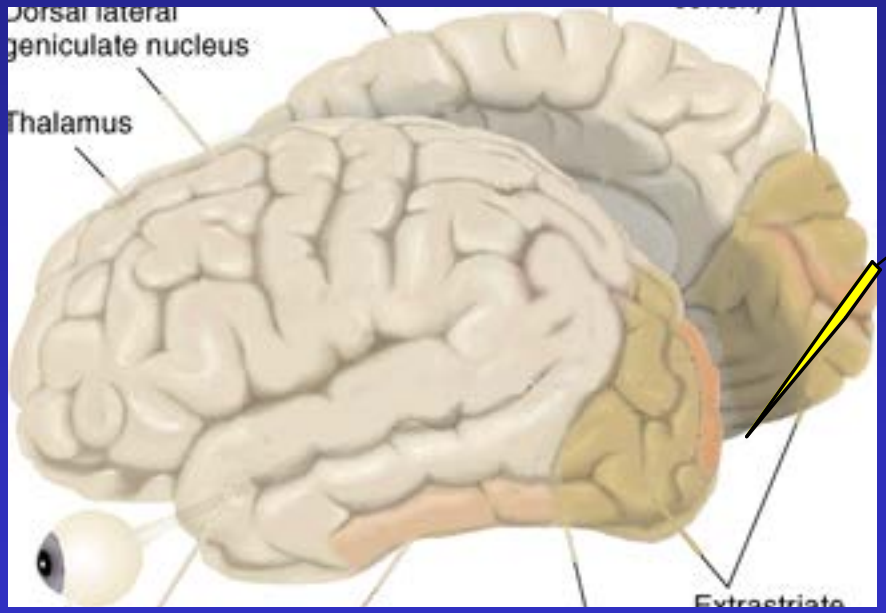
David Hubel & Thorston Wiesel

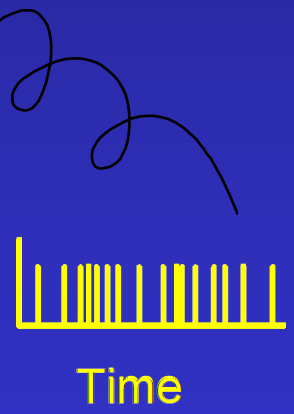
© Stephen E. Palmer, 2002

Cortical Receptive Fields

Single-cell recording from visual cortex







https://www.youtube.com/watch?v=IOHayh06LJ4

© Stephen E. Palmer, 2002

Cortical Receptive Fields

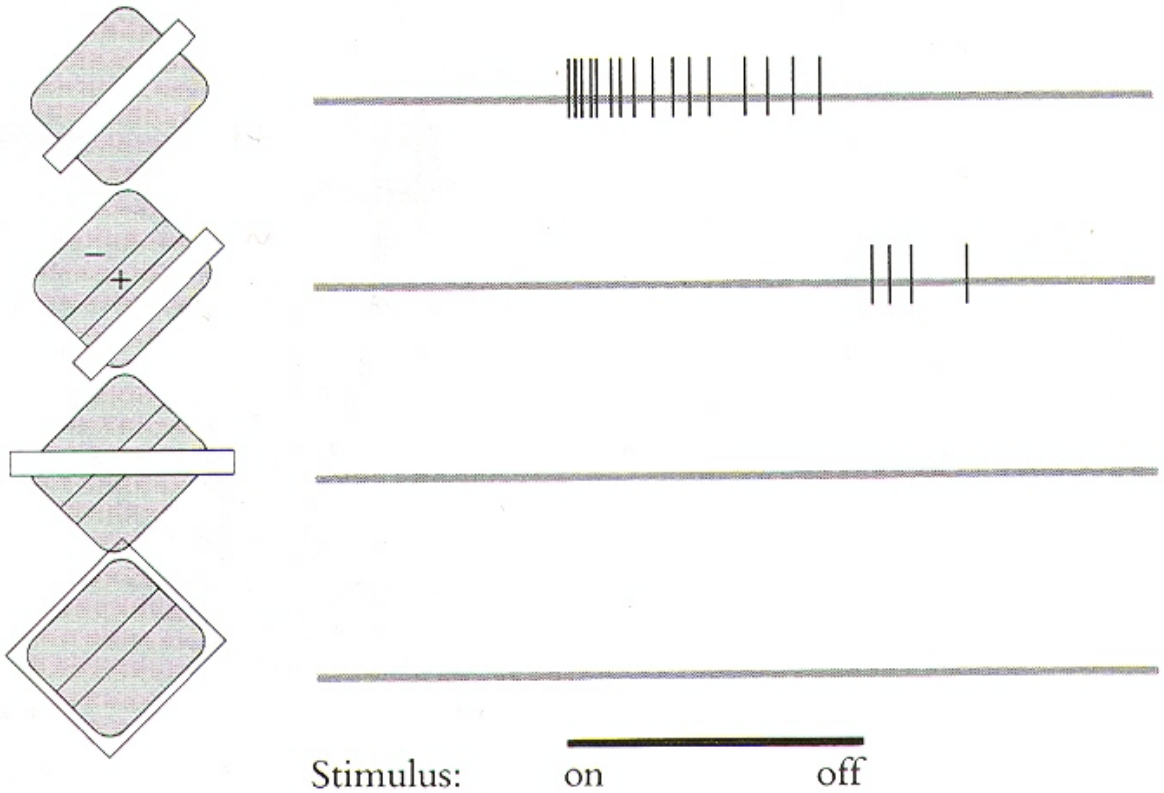
Three classes of cells in V1

Simple cells

Complex cells

Hypercomplex cells

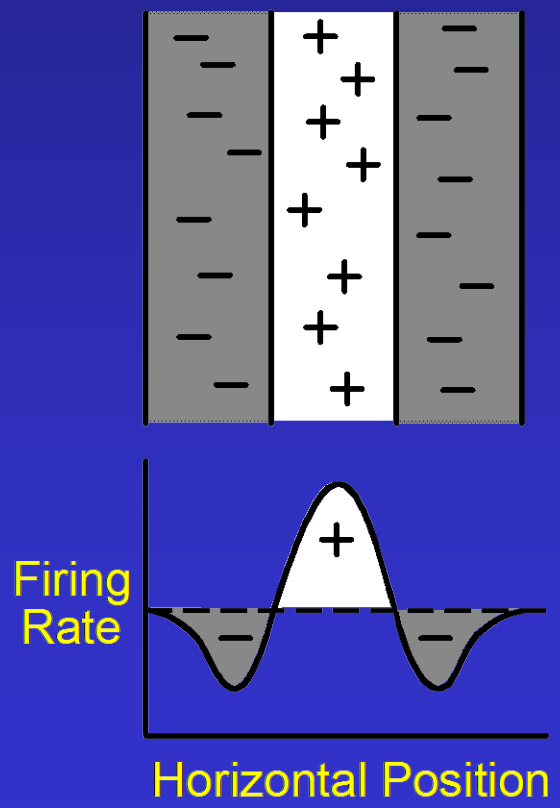
© Stephen E. Palmer, 2002


−
+
Stimulus:
on
off

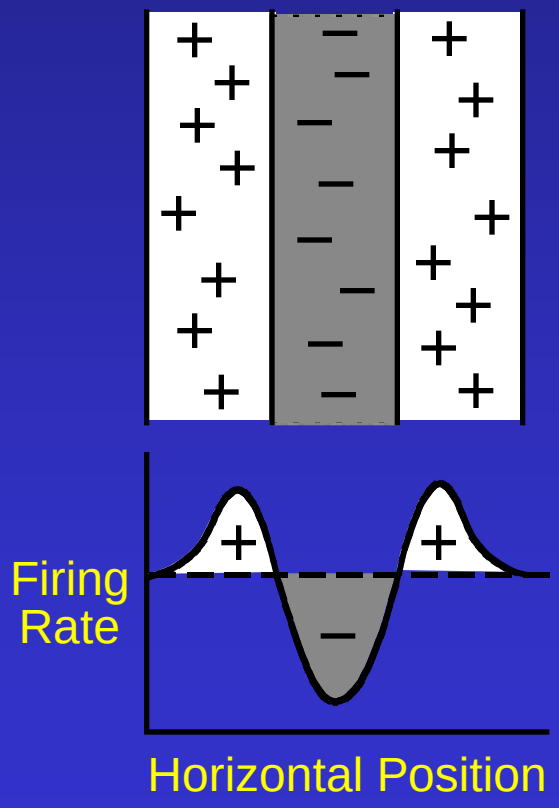
Cortical Receptive Fields

Simple Cells: “Line Detectors”

A. Light Line Detector



B. Dark Line Detector

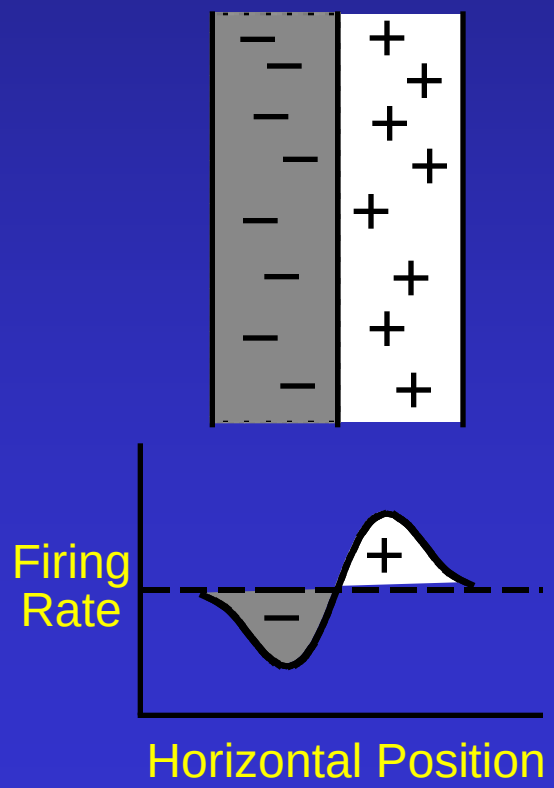


© Stephen E. Palmer, 2002

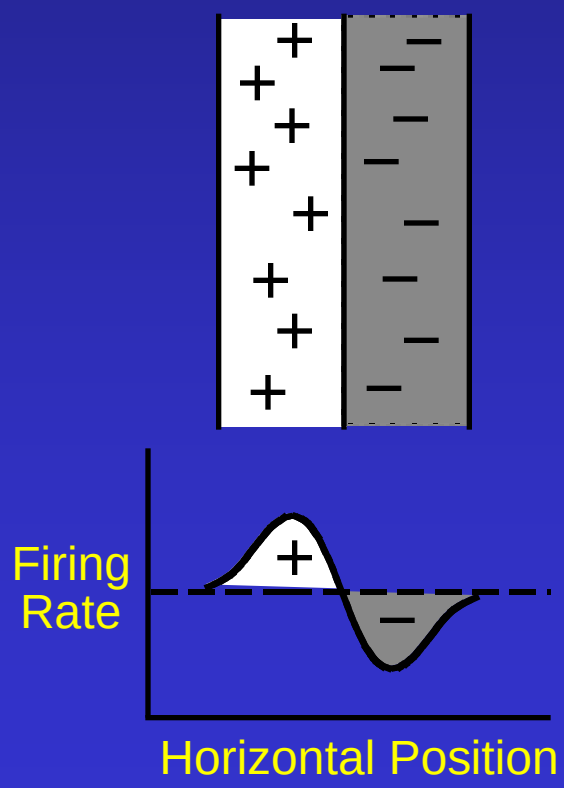
Cortical Receptive Fields

Simple Cells: “Edge Detectors”

C. Dark-to-light Edge Detector



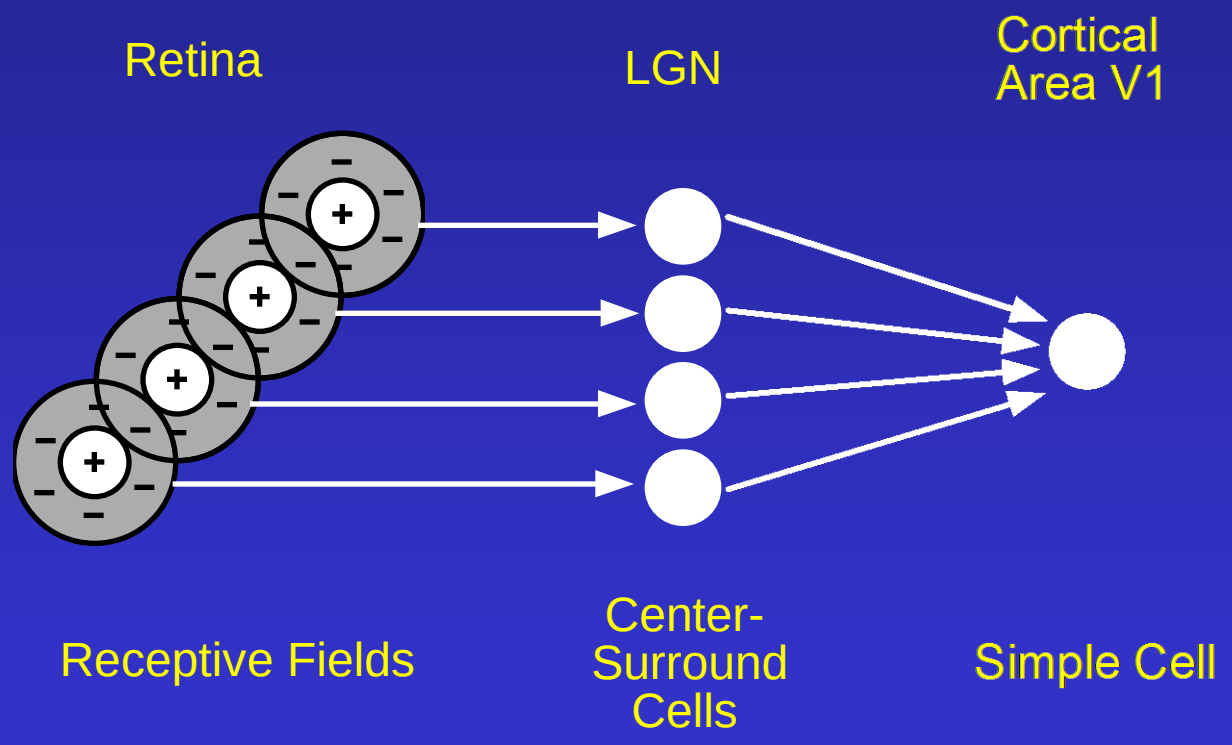
D. Light-to-dark Edge Detector



© Stephen E. Palmer, 2002

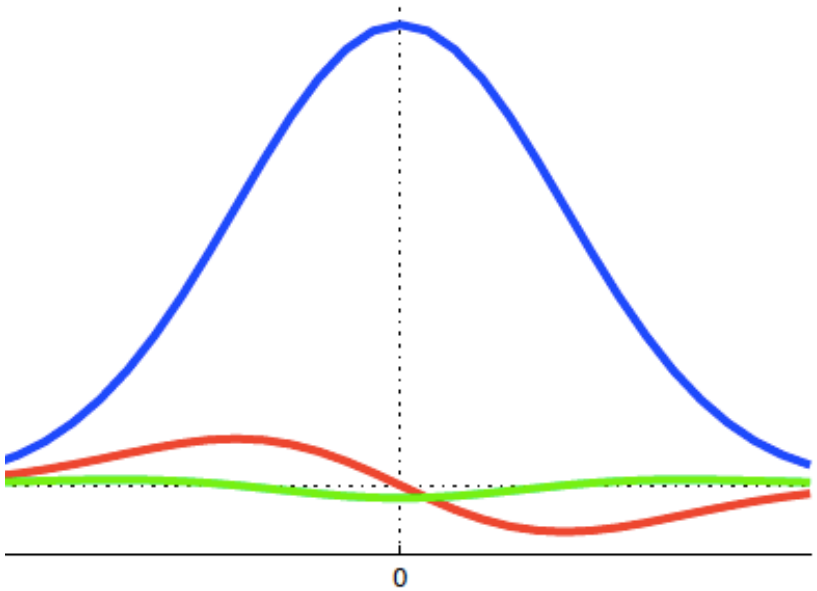
Cortical Receptive Fields

Constructing a line detector



© Stephen E. Palmer, 2002

The 1D Gaussian and its derivatives



$$G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

$$G'_\sigma(x) = \frac{d}{dx} G_\sigma(x) = -\frac{1}{\sigma}\left(\frac{x}{\sigma}\right) G_\sigma(x)$$

$$G''_\sigma(x) = \frac{d^2}{dx^2} G_\sigma(x) = \frac{1}{\sigma^2}\left(\frac{x^2}{\sigma^2} - 1\right) G_\sigma(x)$$

$G'_\sigma(x)$'s maxima/minima occur at $G''_\sigma(x)$'s zeros. And, we can see that $G'_\sigma(x)$ is an odd symmetric function and $G''_\sigma(x)$ is an even symmetric function.

Oriented Gaussian Derivatives in 2D

$$f_1(x,y) = G'_{\sigma_1}(x)G_{\sigma_2}(y) \tag{10.4}$$
$$f_2(x,y) = G''_{\sigma_1}(x)G_{\sigma_2}(y) \tag{10.5}$$

We also consider rotated versions of these Gaussian derivative functions.

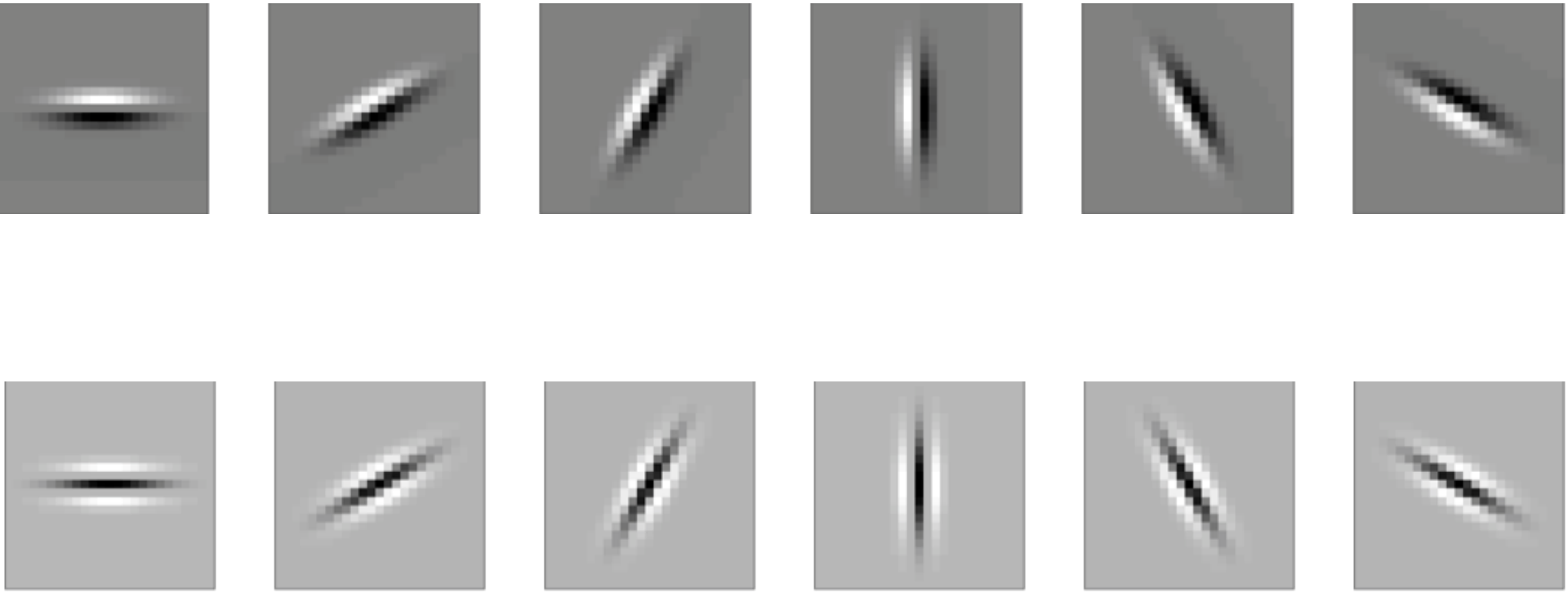
$$Rot_\theta f_1 = G'_{\sigma_1}(u)G_{\sigma_2}(v) \tag{10.6}$$
$$Rot_\theta f_2 = G''_{\sigma_1}(u)G_{\sigma_2}(v) \tag{10.7}$$

where we set

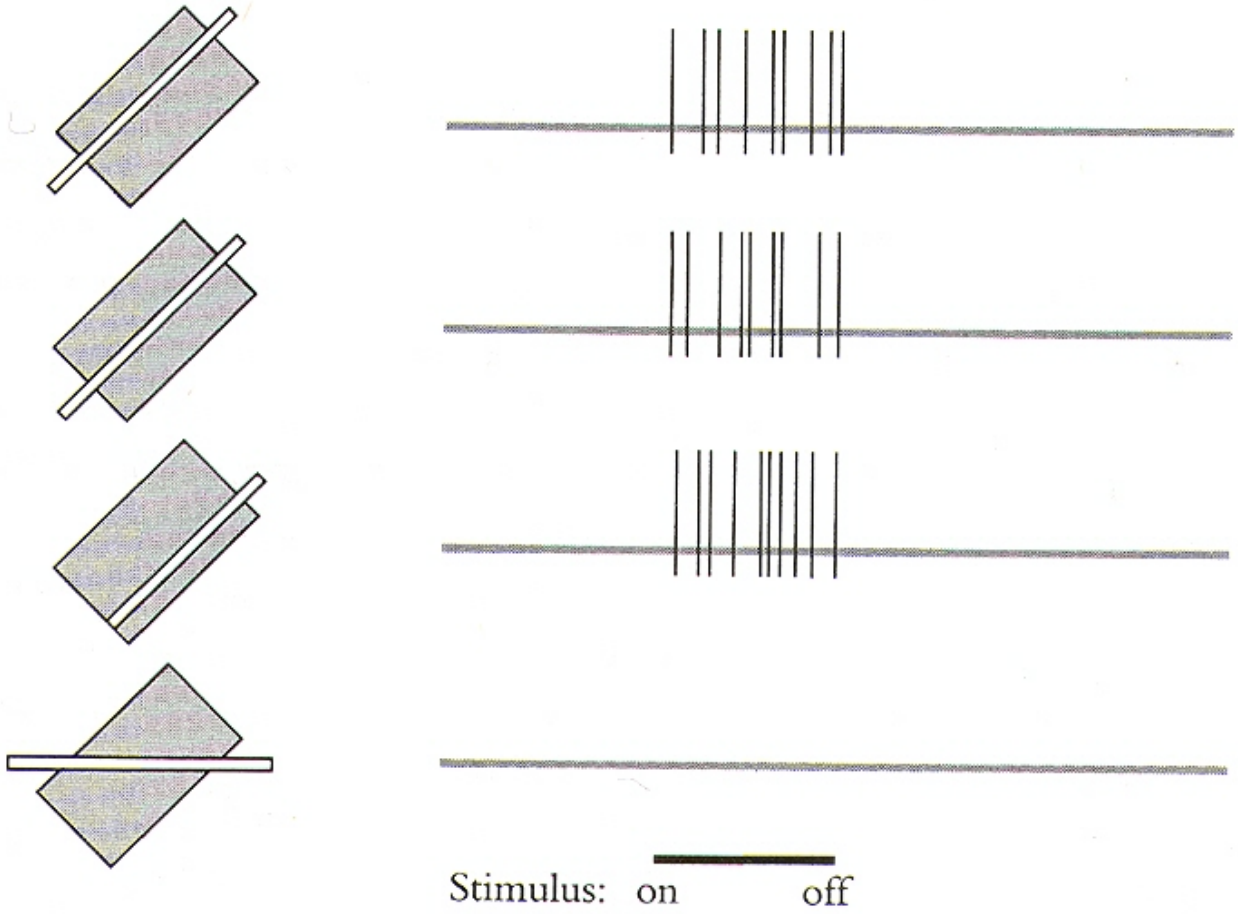
$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

These are useful when we convolve with 2D images, e.g. to detect edges at different orientations.

Oriented Gaussian First and Second Derivatives

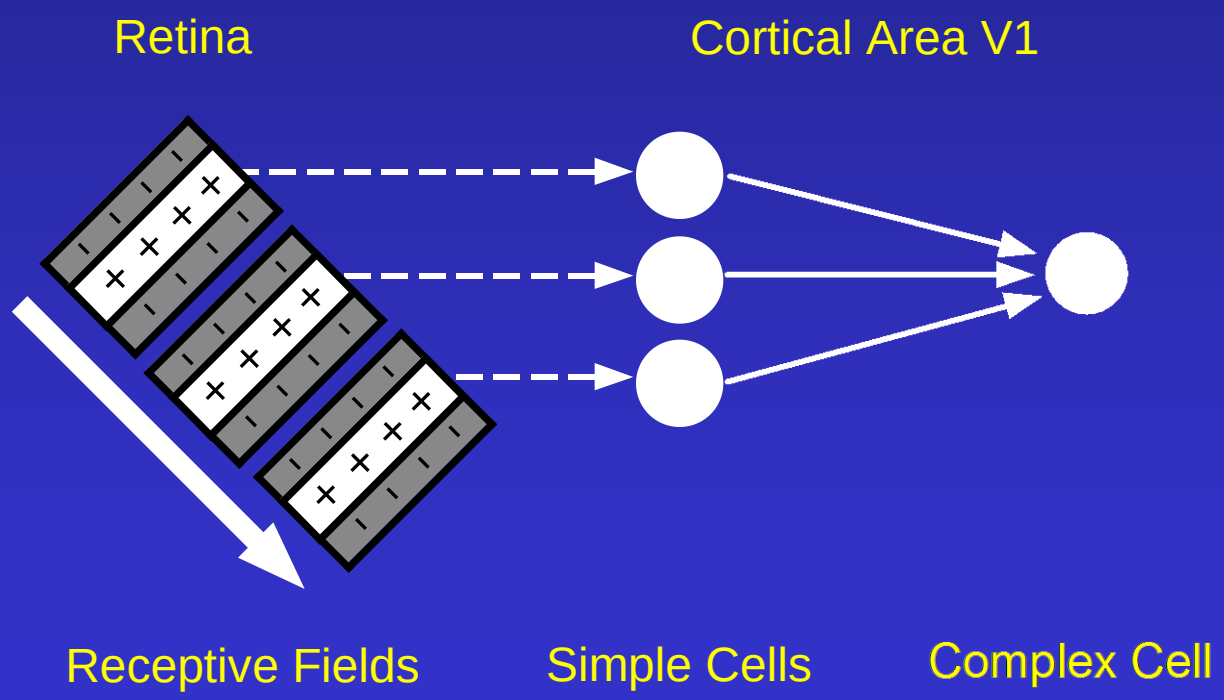


Receptive fields of complex cells



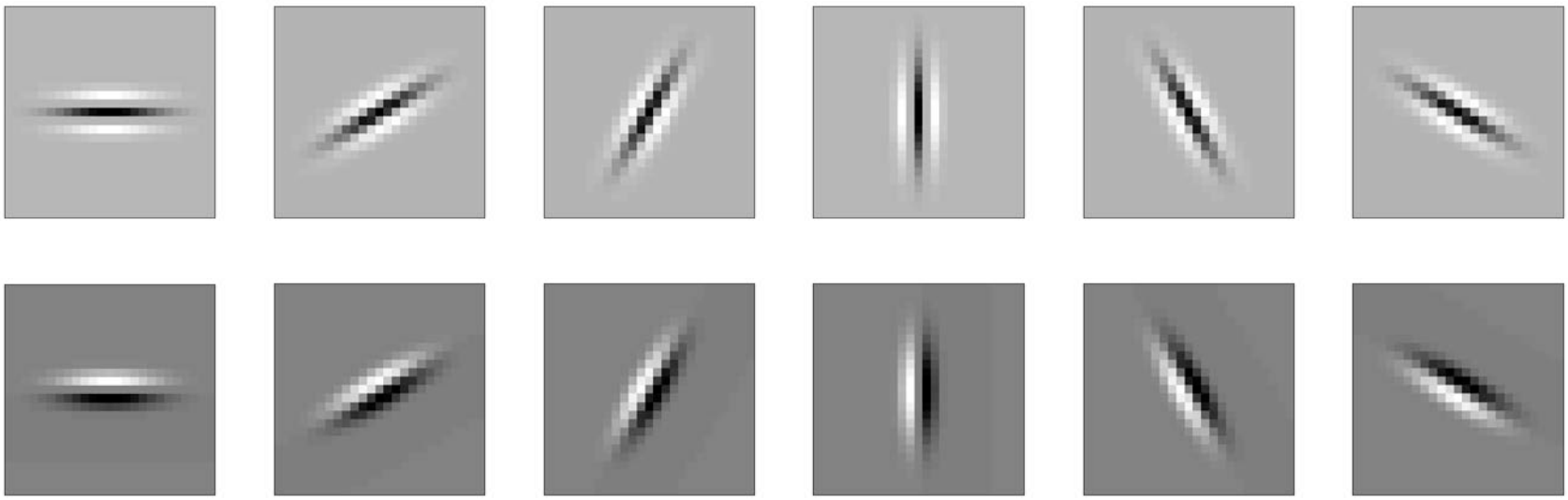
Cortical Receptive Fields

Constructing a Complex Cell



© Stephen E. Palmer, 2002

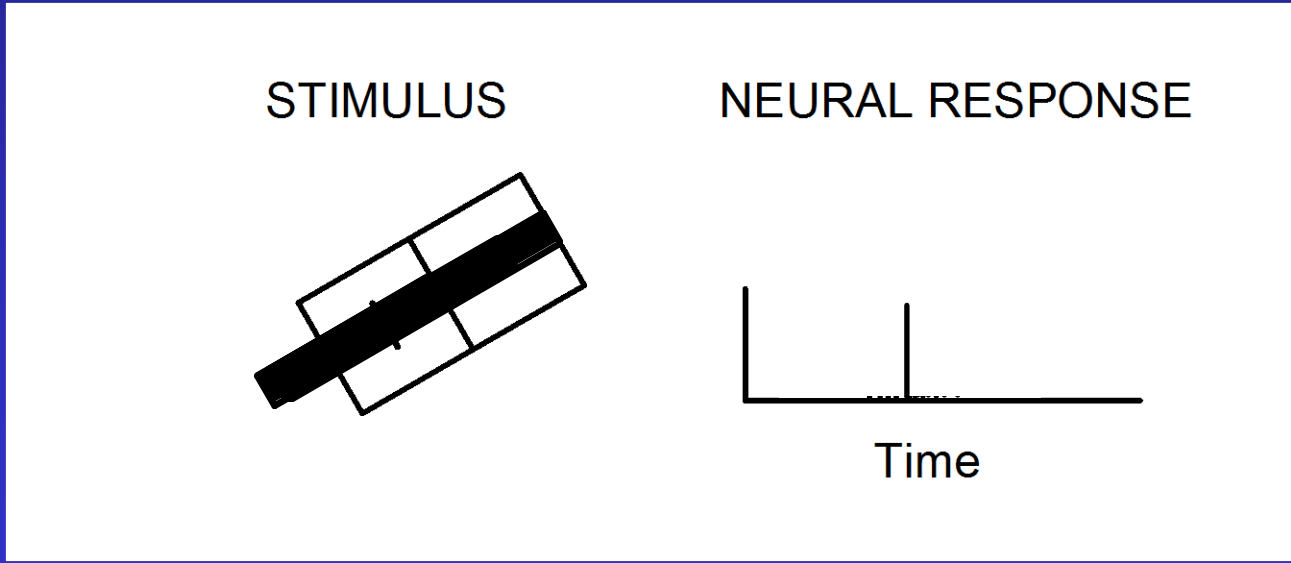
Orientation Energy



- $OE = (I * f_{odd})^2 + (I * f_{even})^2$
- Can be used to model complex cells, as this is insensitive to phase
- Multiple scales

Cortical Receptive Fields

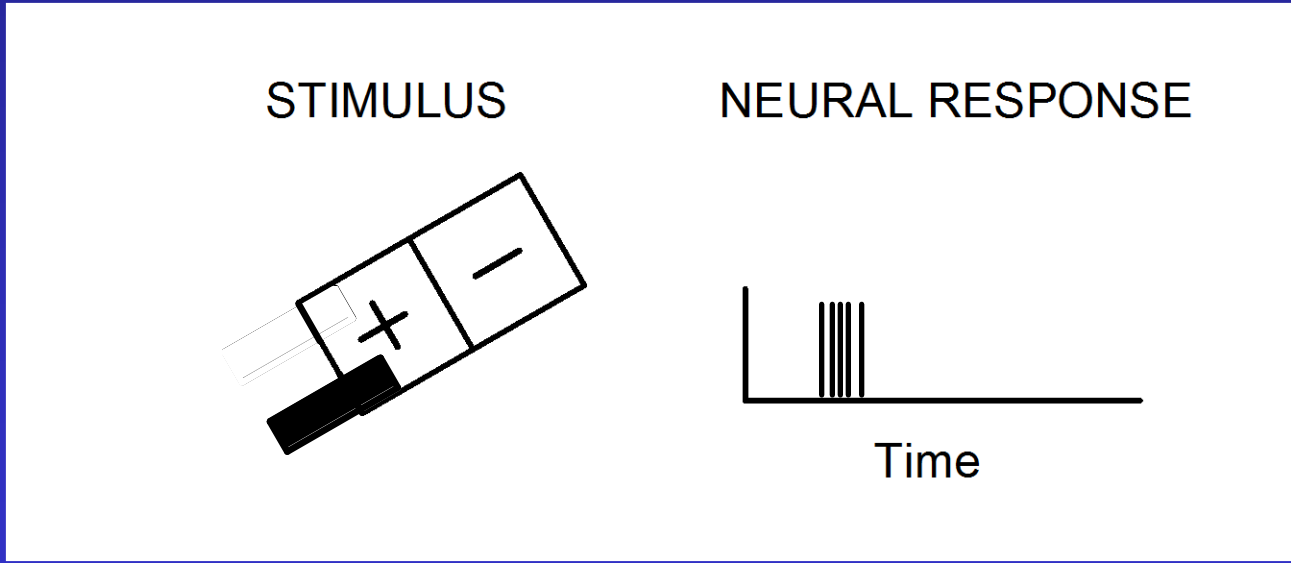
“End-stopped” Simple Cells



© Stephen E. Palmer, 2002

Cortical Receptive Fields

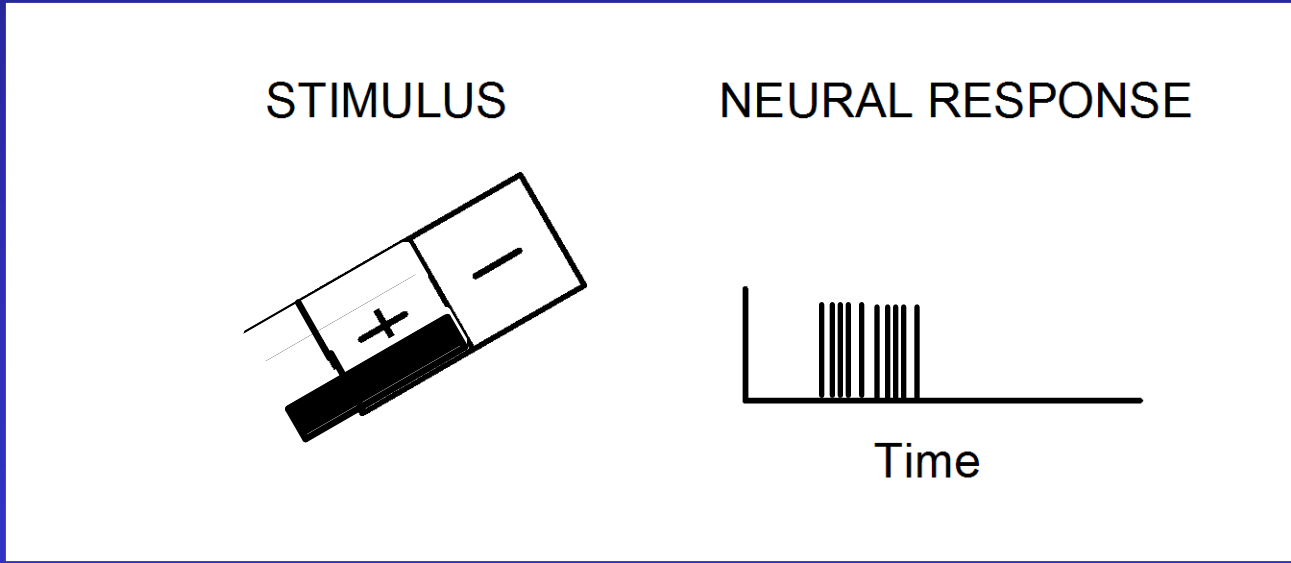
Hypercomplex Cells



© Stephen E. Palmer, 2002

Cortical Receptive Fields

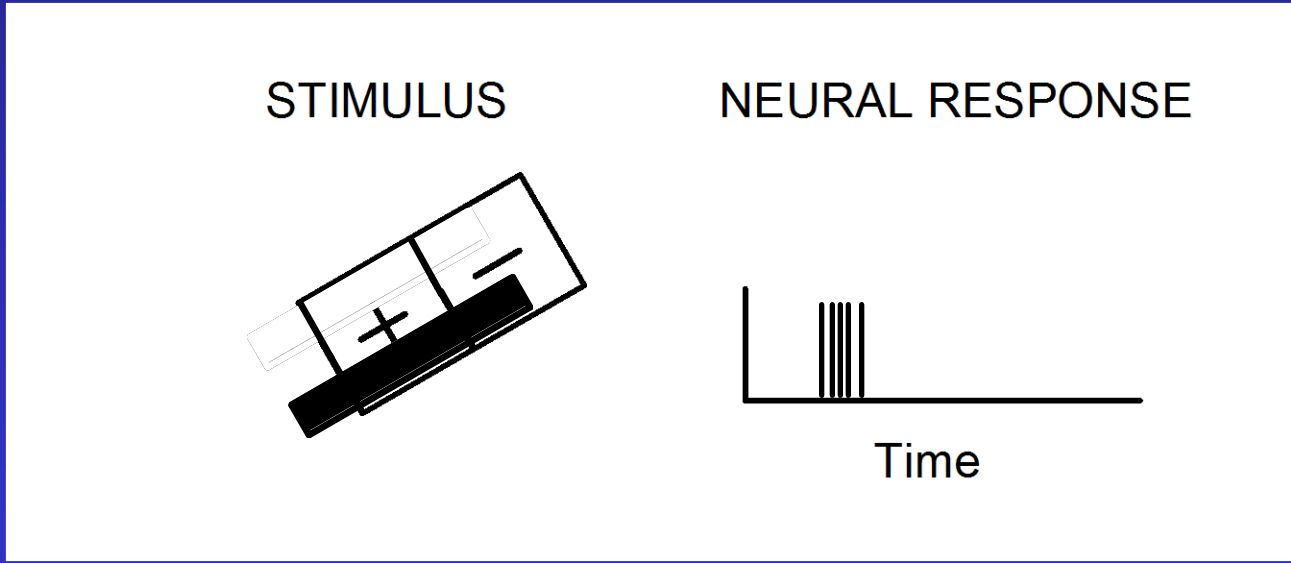
Hypercomplex Cells



© Stephen E. Palmer, 2002

Cortical Receptive Fields

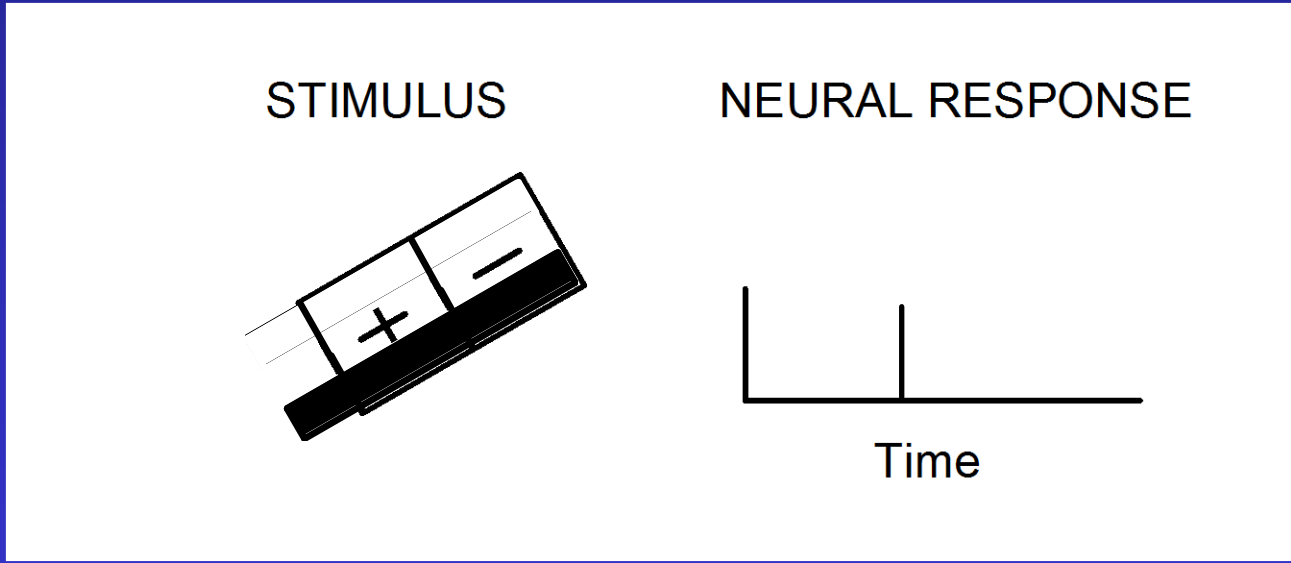
Hypercomplex Cells



© Stephen E. Palmer, 2002

Cortical Receptive Fields

Hypercomplex Cells

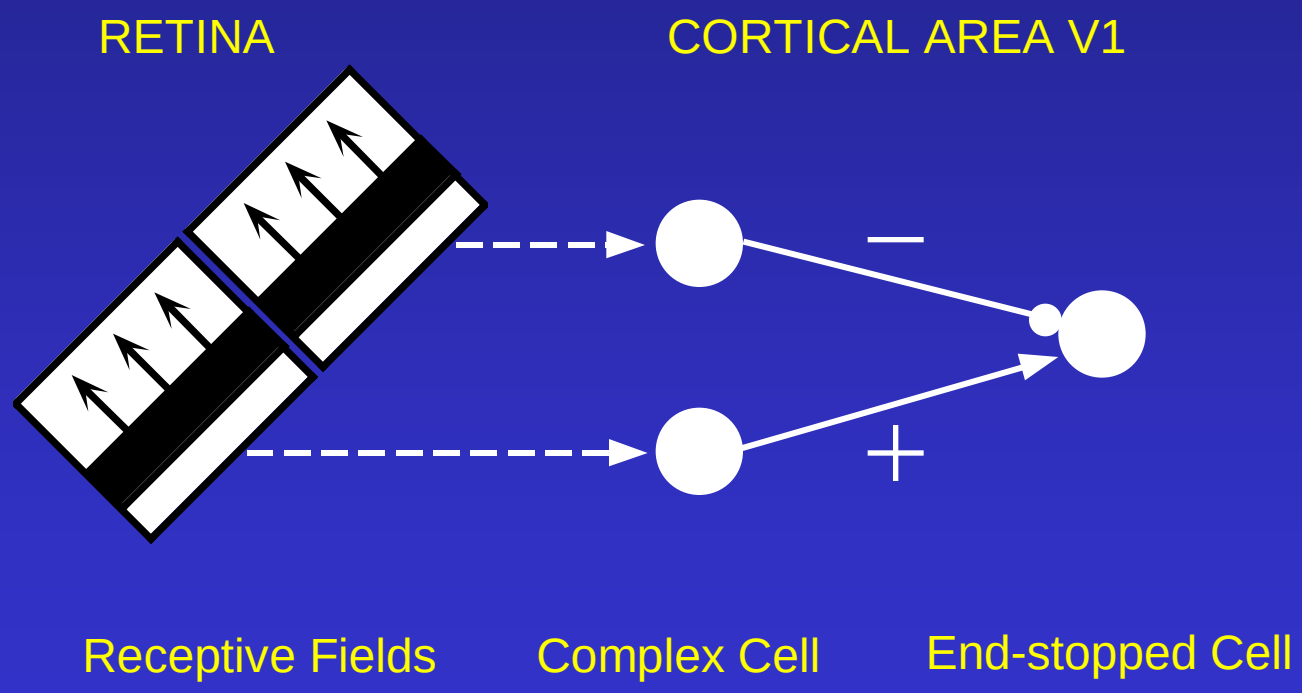


“End-stopped” Cells

© Stephen E. Palmer, 2002

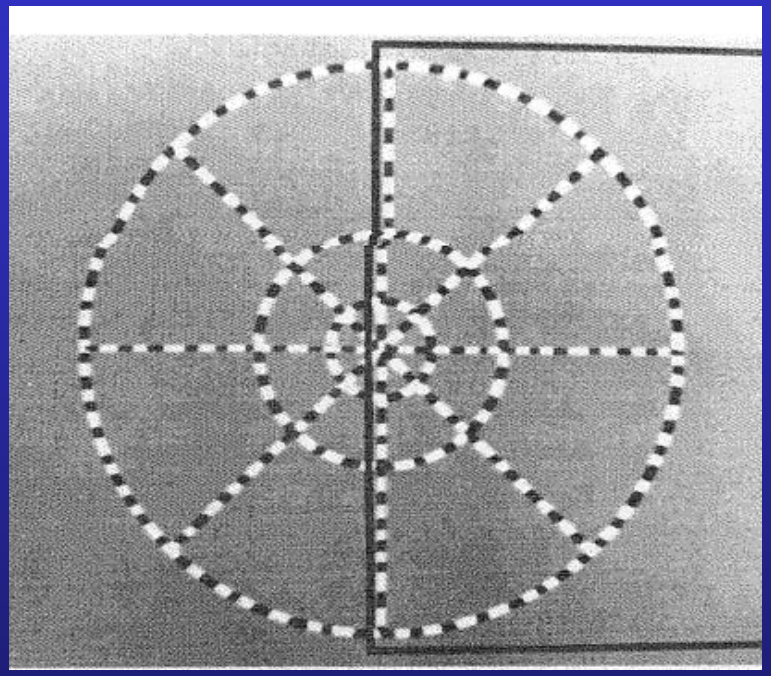
Cortical Receptive Fields

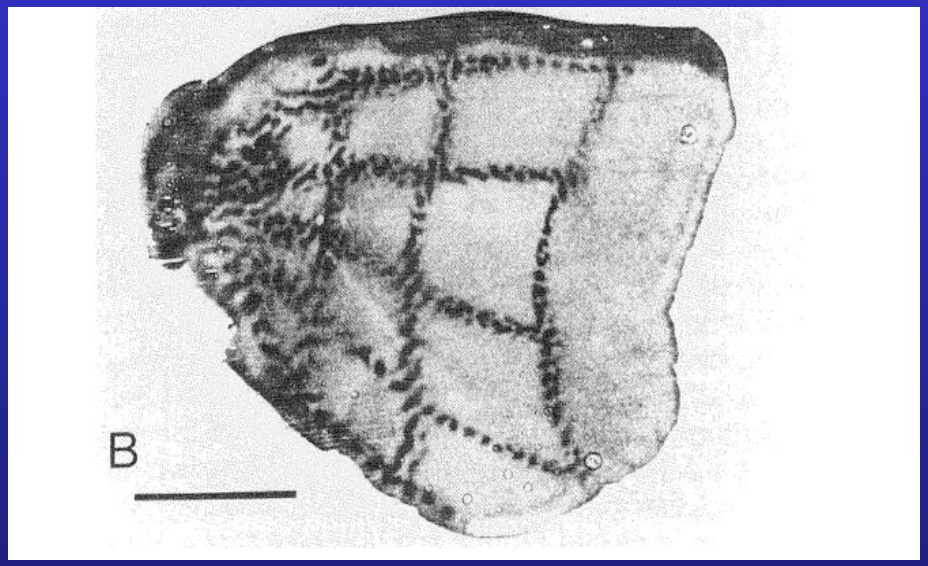
Constructing a Hypercomplex Cell



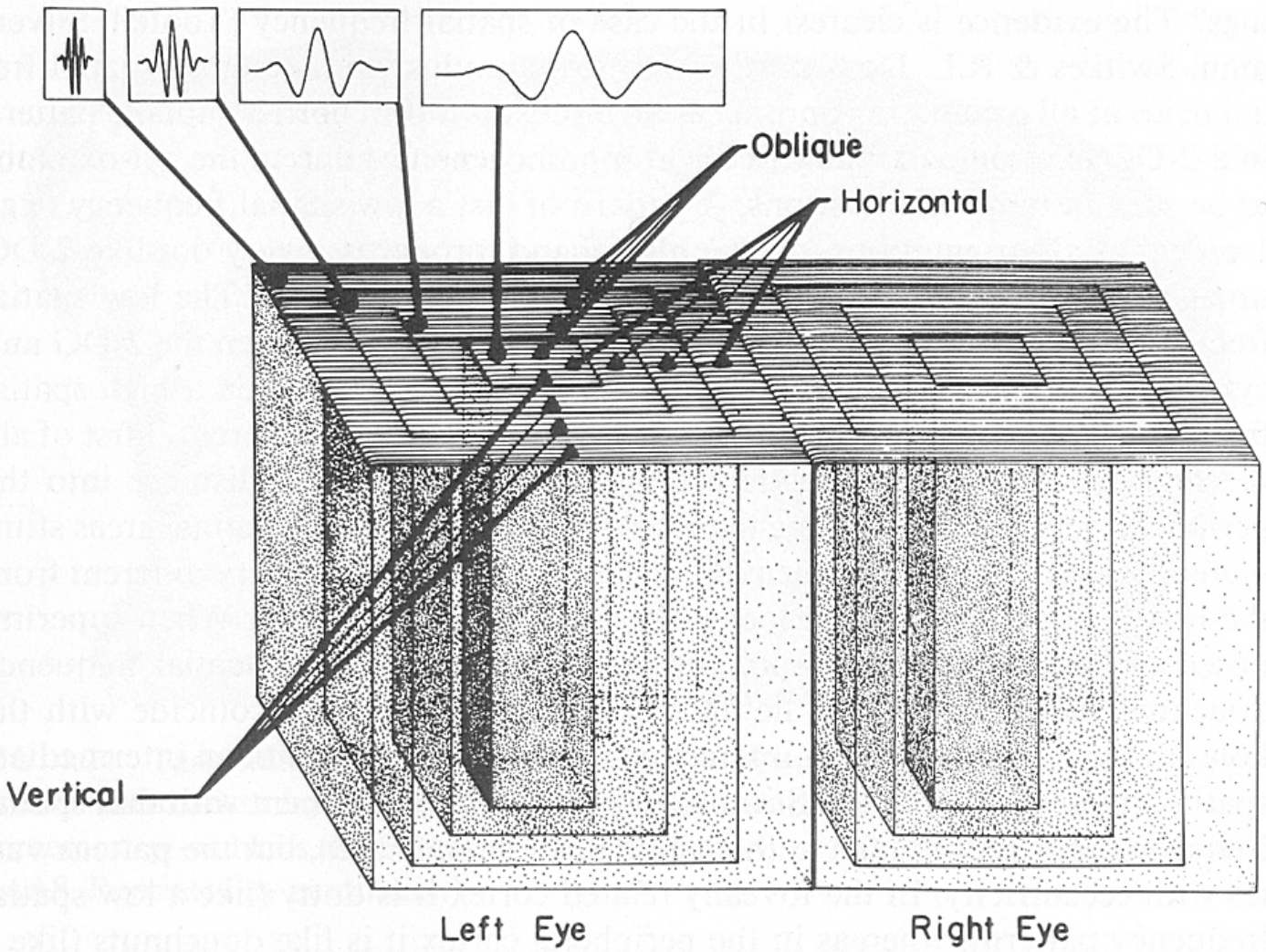
© Stephen E. Palmer, 2002

Mapping from Retina to V1





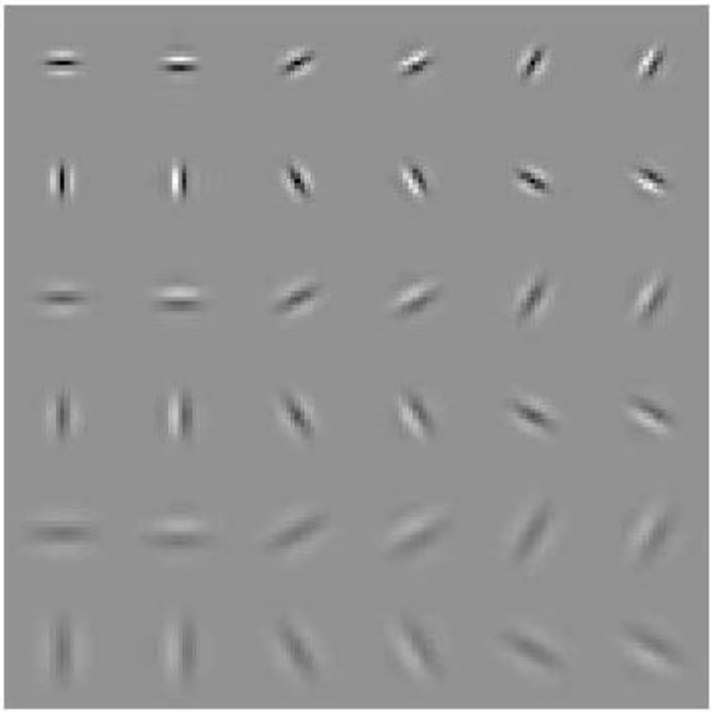
Hypercolumns in visual cortex



Model of Striate Module in Monkeys

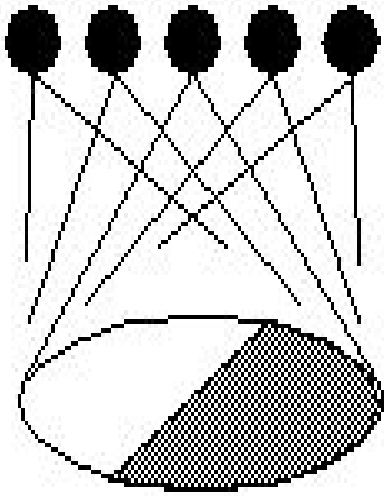
Modeling hypercolumns

- Elongated directional Gaussian derivatives
- Gabor filters could be used instead
- Multiple orientations, scales

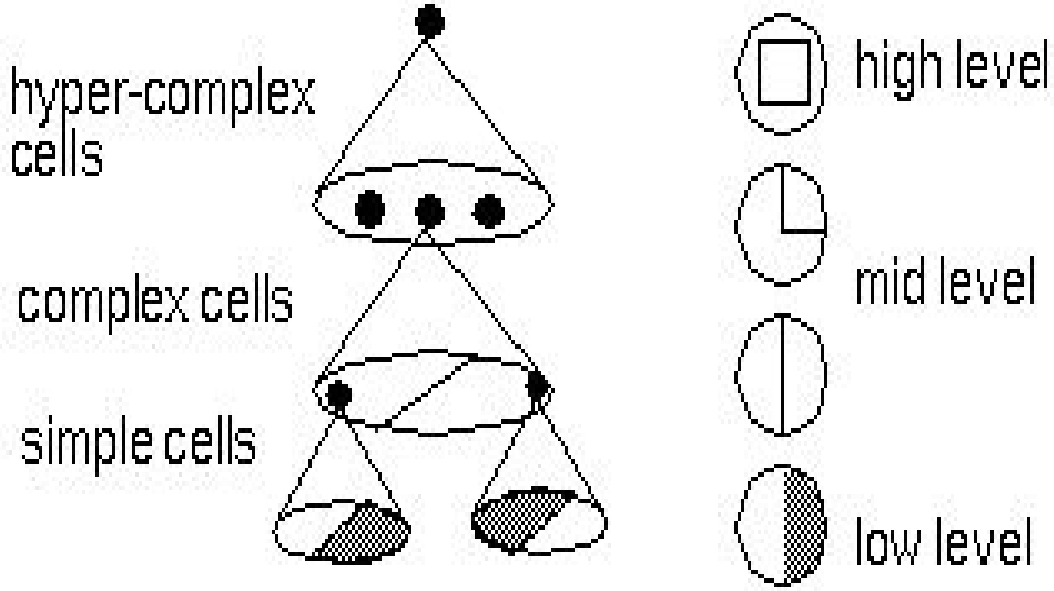


Hubel & Weisel

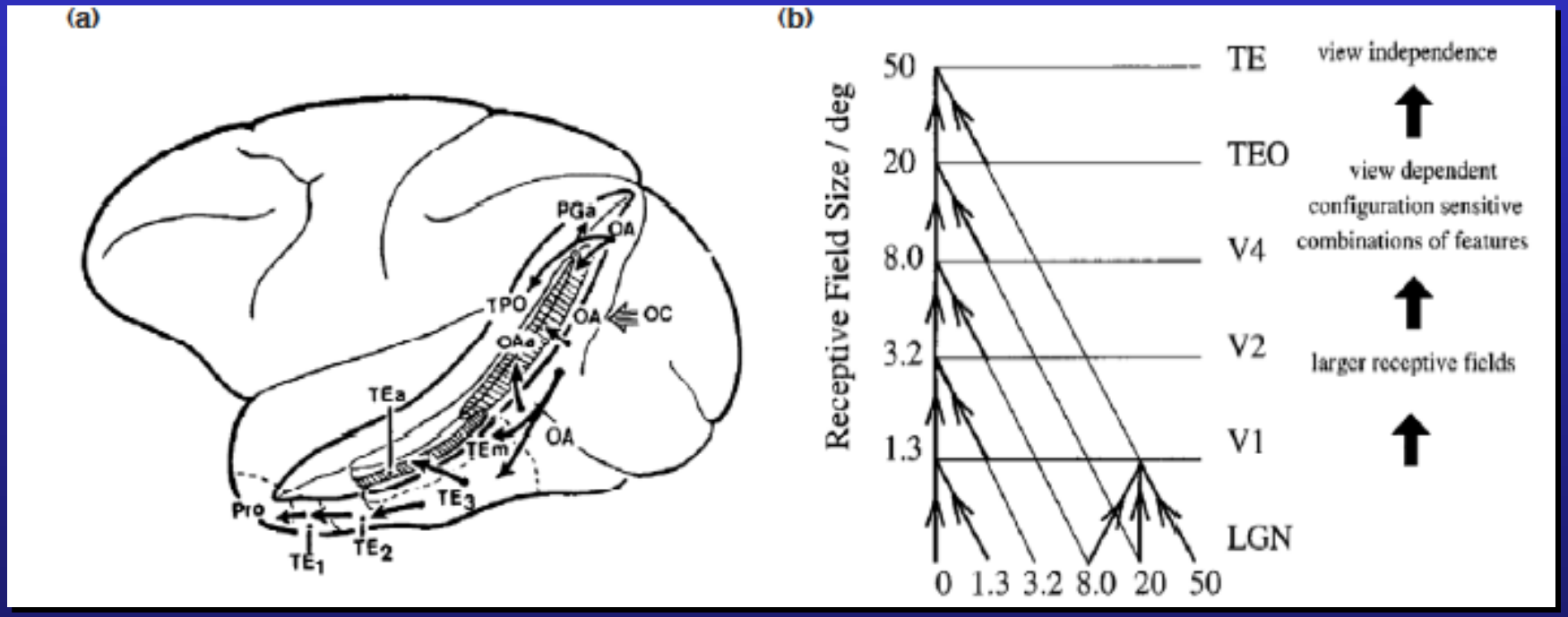
topographical mapping



featural hierarchy

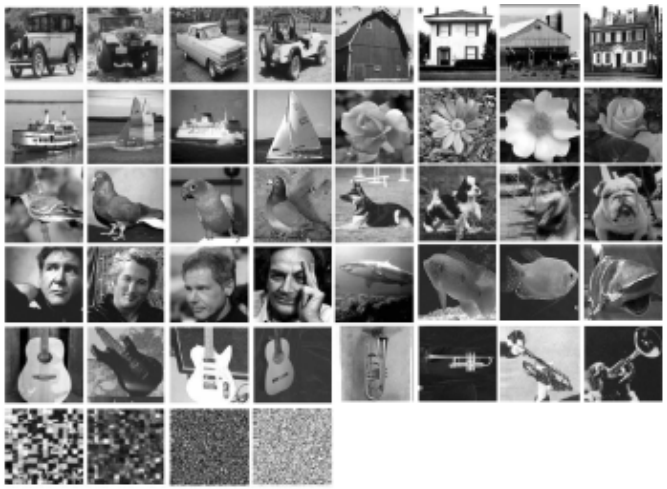


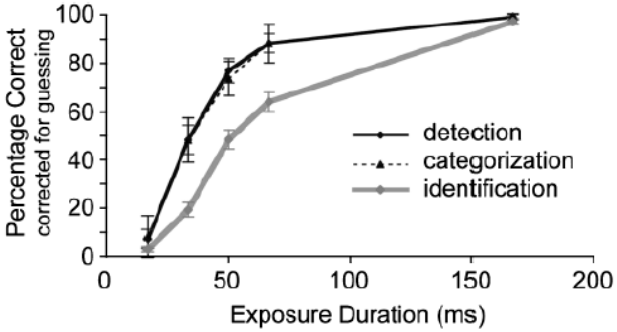
Rolls et al (2000) model of ventral stream



Object Detection can be very fast

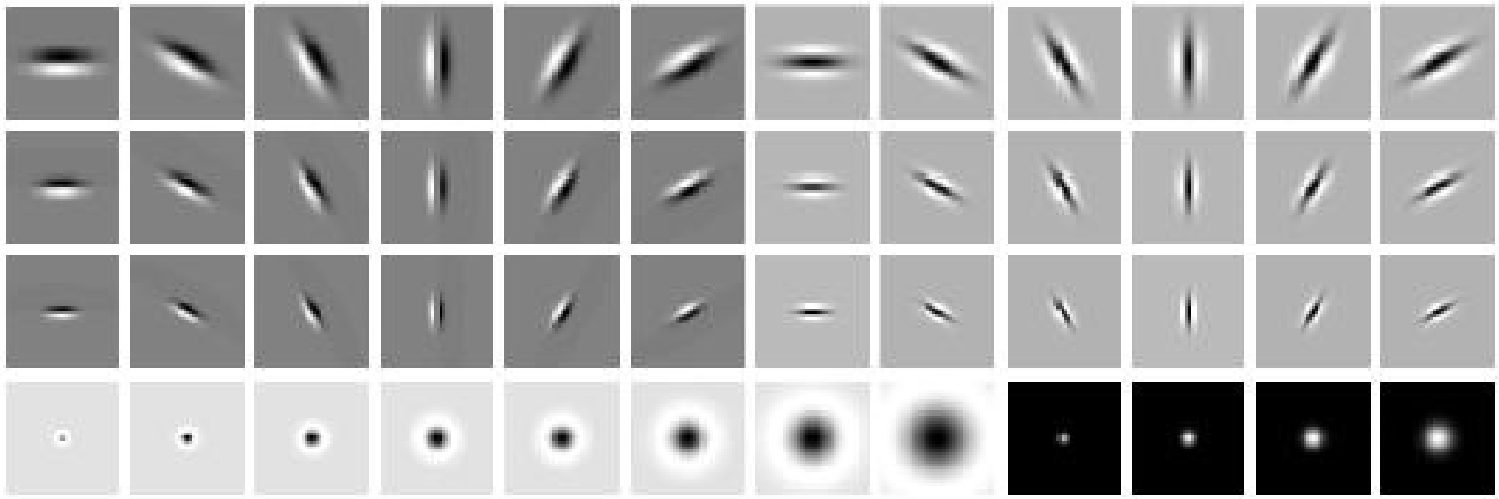
- On a task of judging animal vs no animal, humans can make mostly correct saccades in 150 ms (Kirchner & Thorpe, 2006)
  - Comparable to synaptic delay in the retina, LGN, V1, V2, V4, IT pathway.
  - Doesn't rule out feed back but shows feed forward only is very powerful
- Detection and categorization are practically simultaneous (Grill-Spector & Kanwisher, 2005)





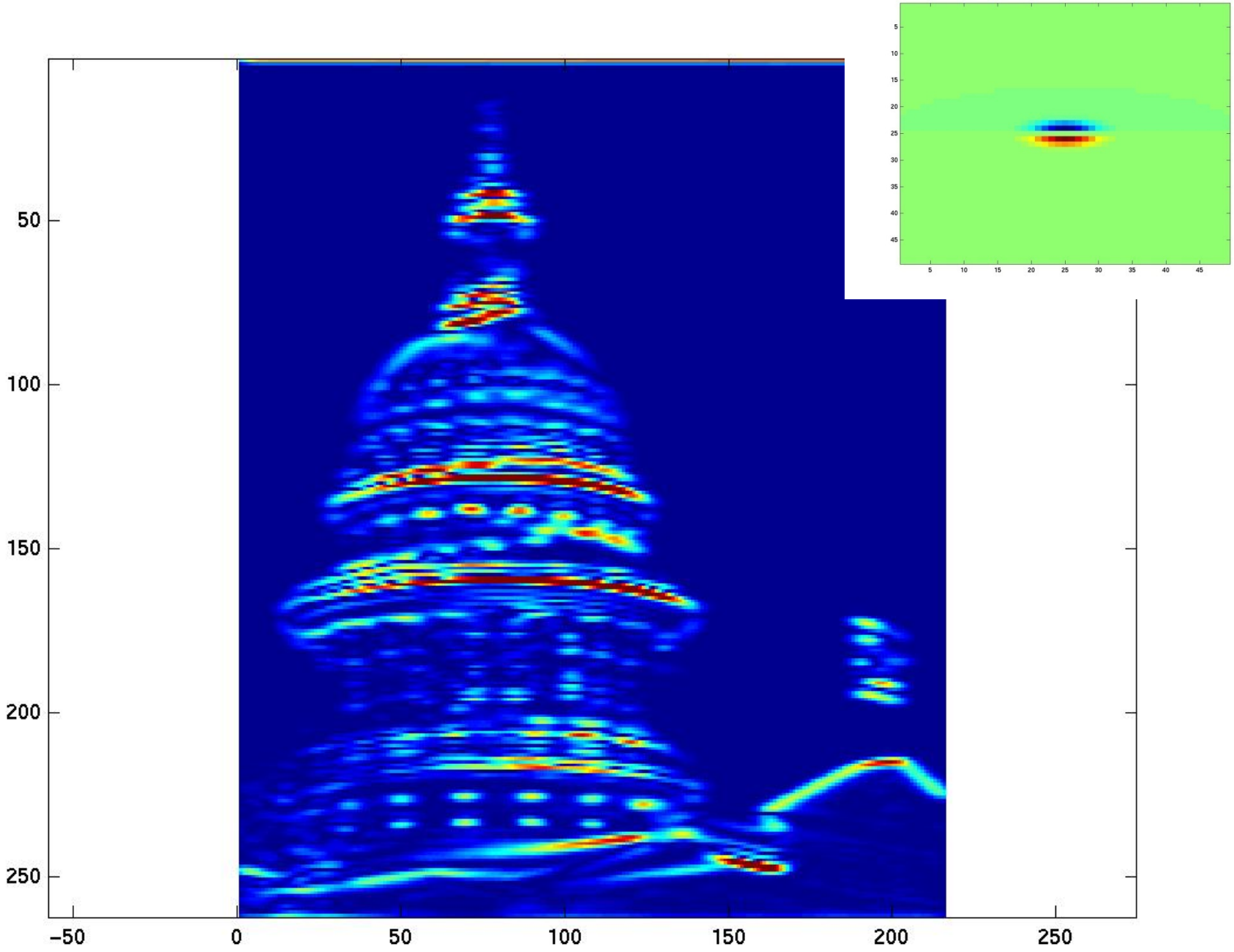
Overcomplete representation: filter banks

LM Filter Bank

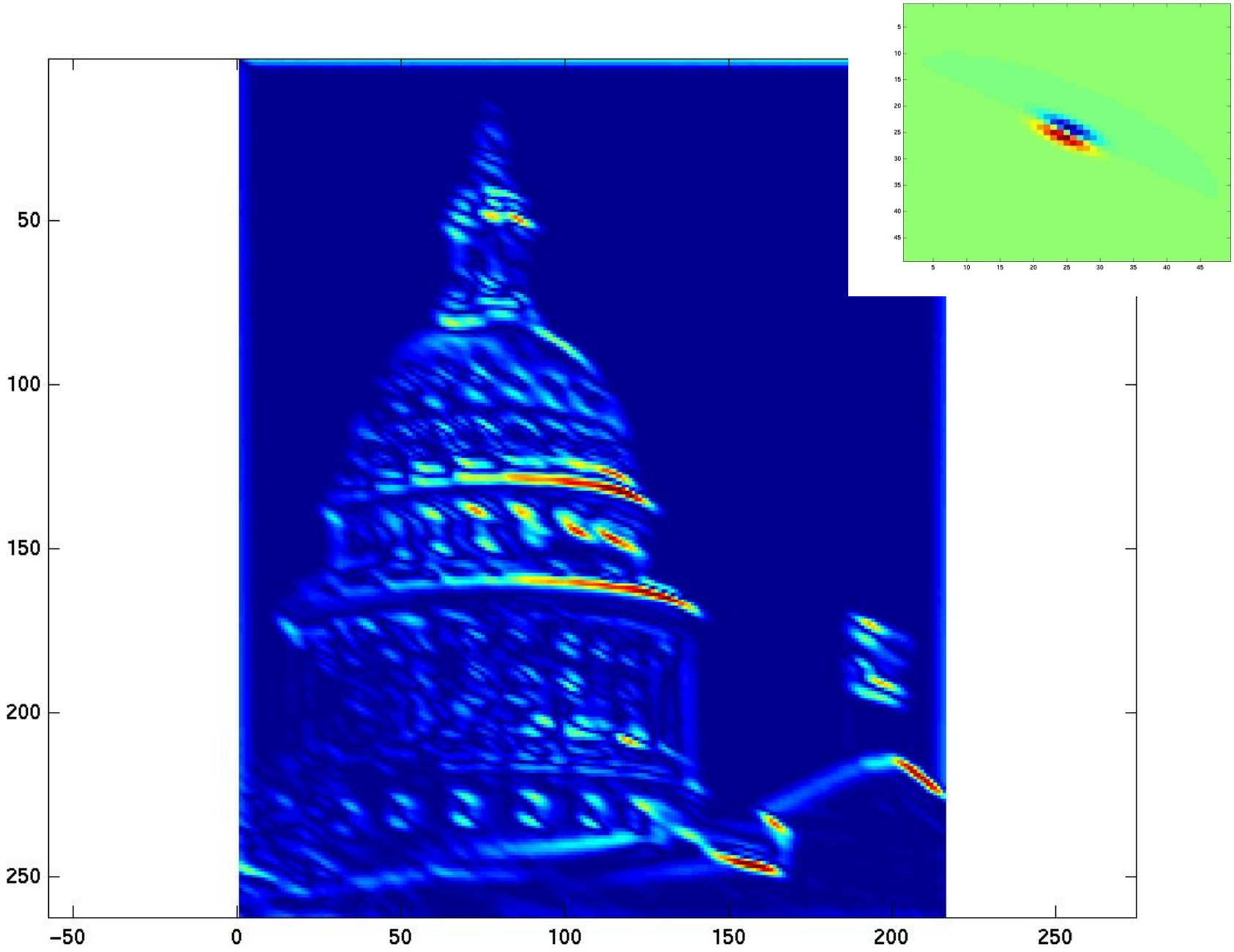


Code for filter banks: *www.robots.ox.ac.uk/~vgg/research/texclass/filters.html*

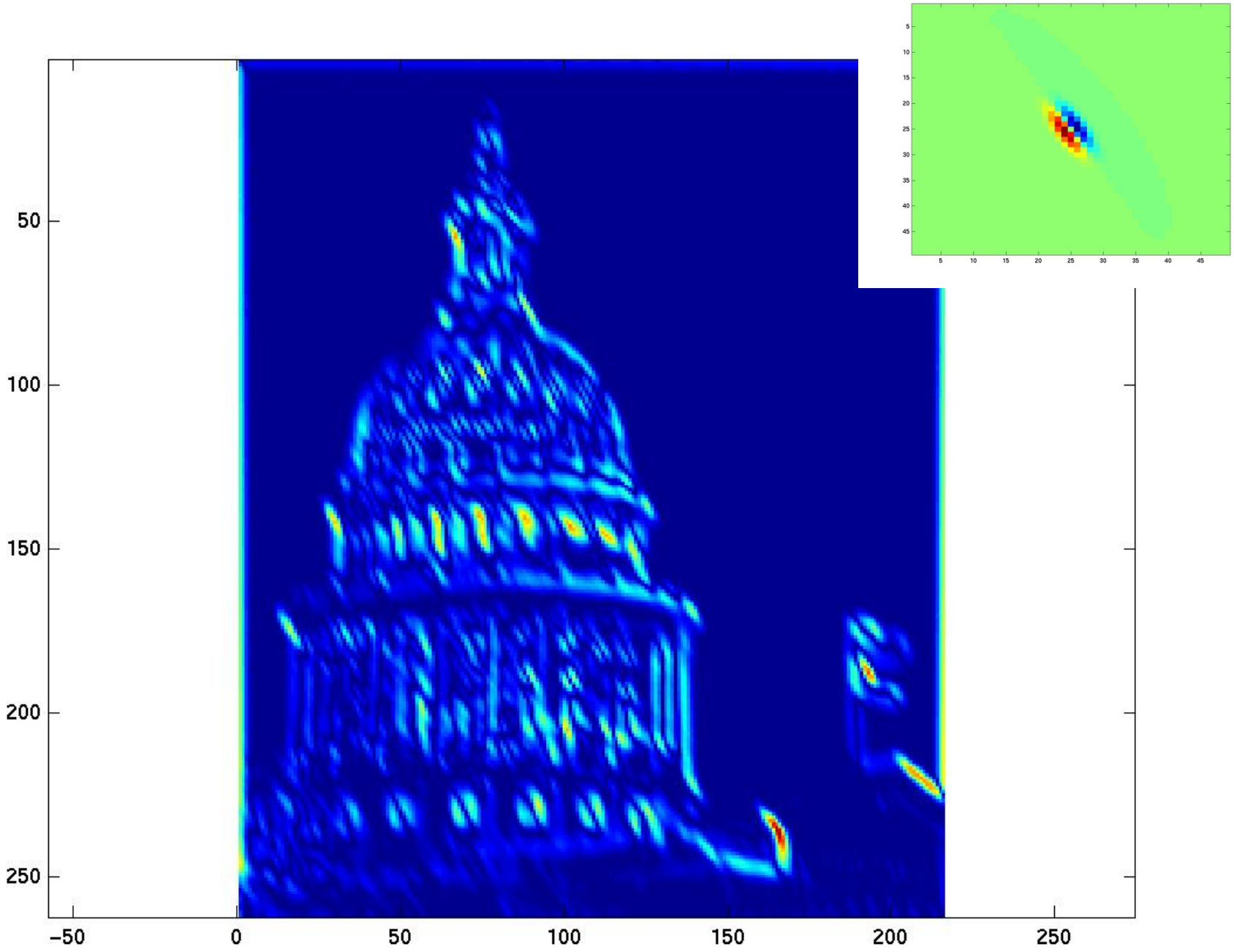
Image from http://www.texasexplorer.com/austincap2.jpg



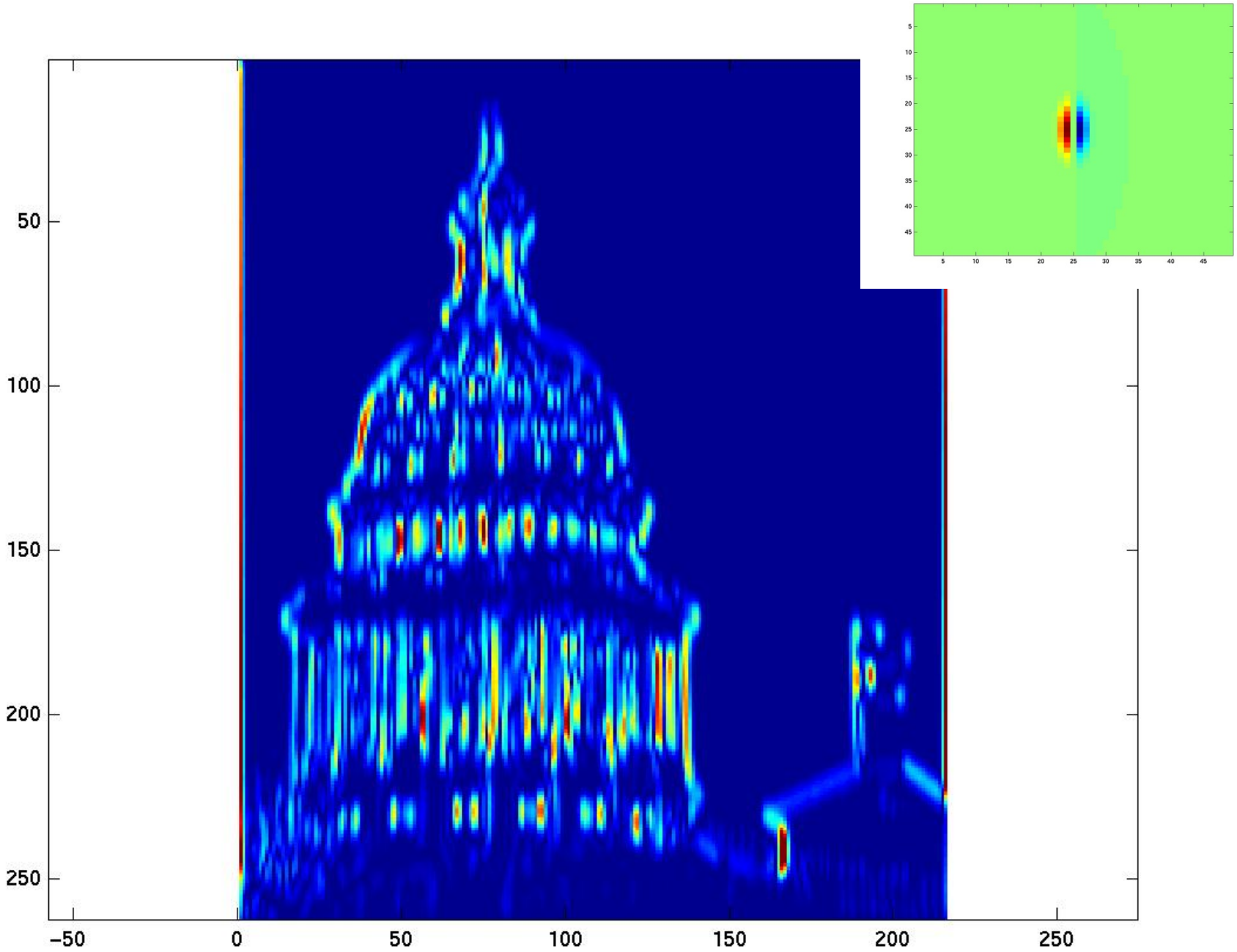
Kristen Grauman



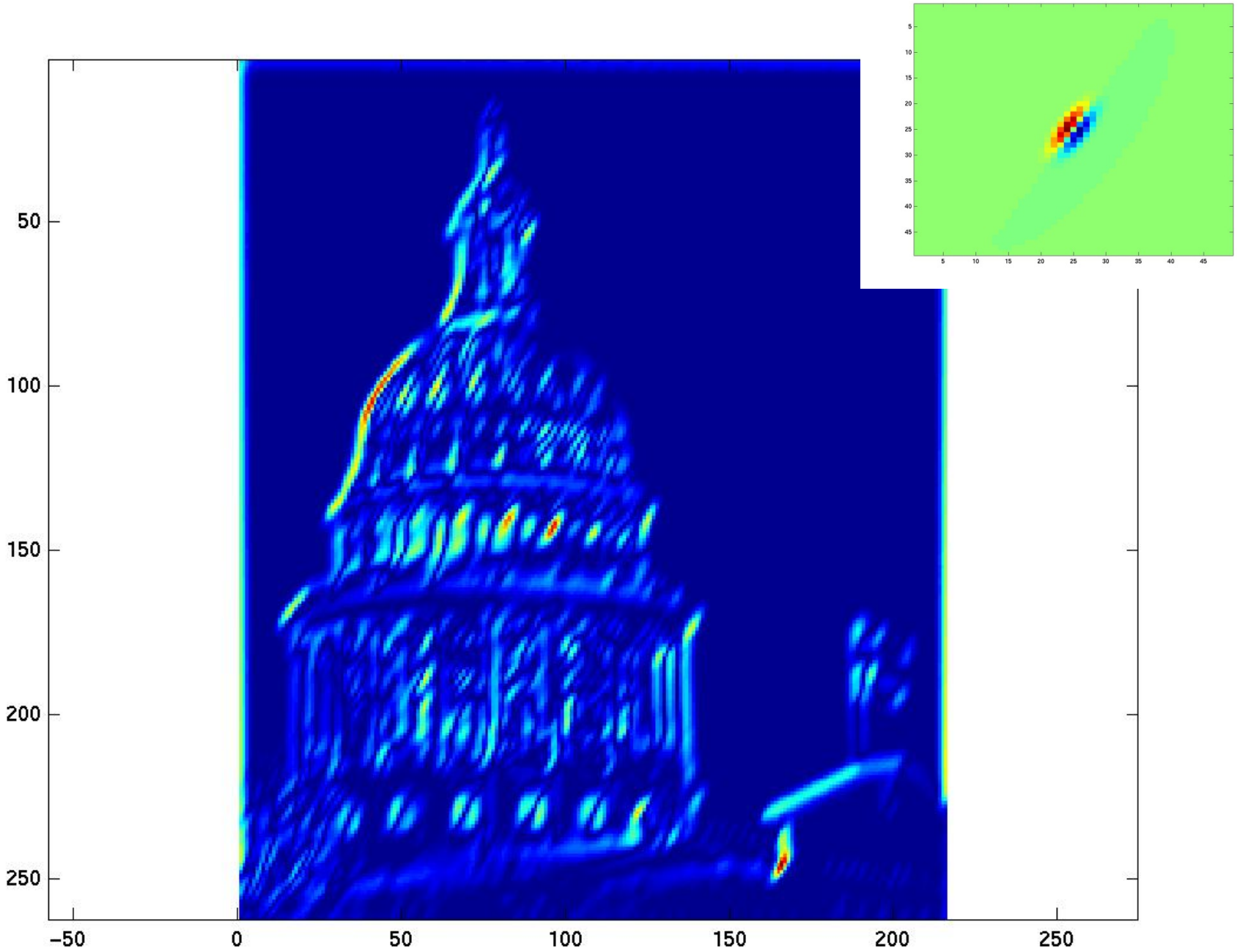
Kristen Grauman



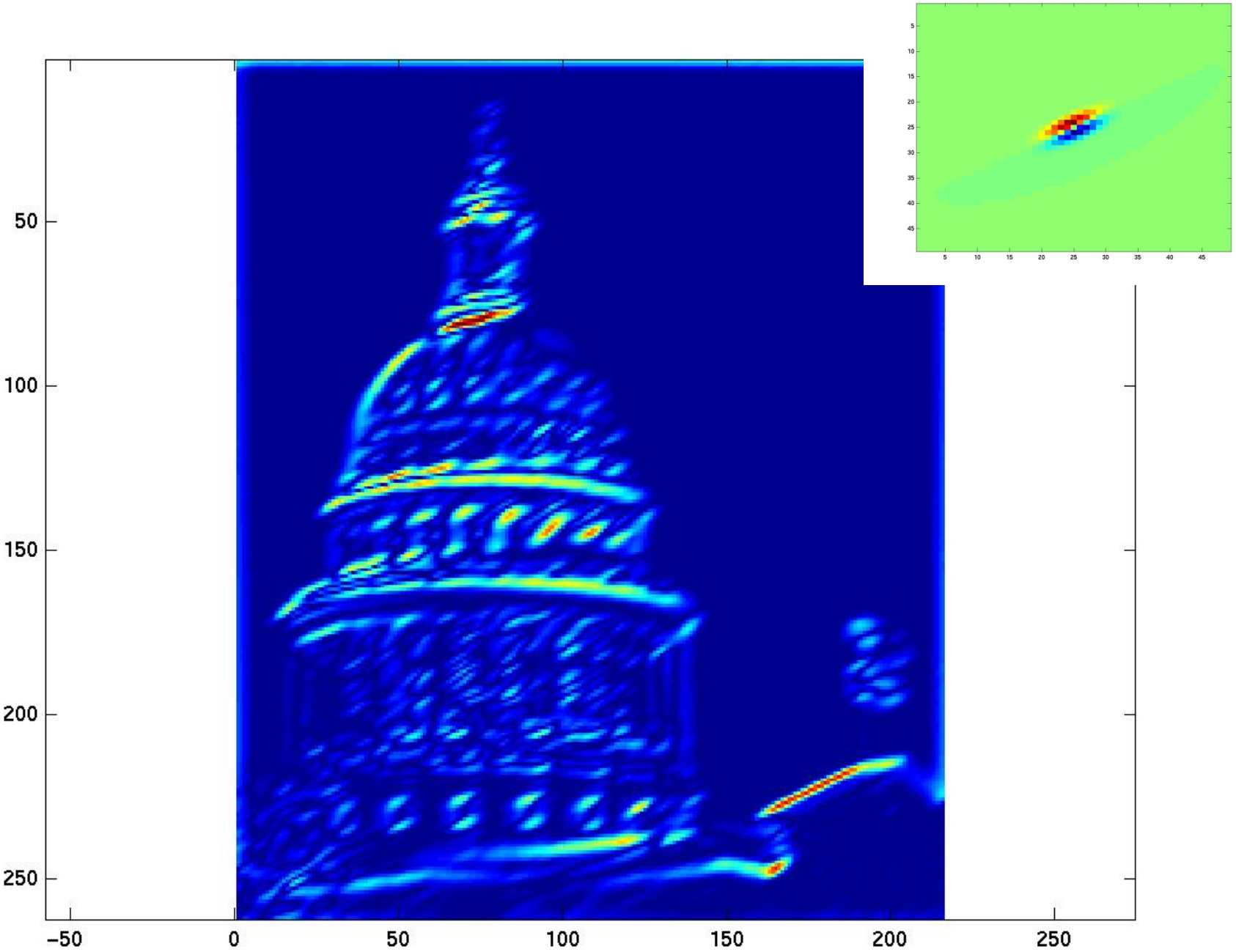
Kristen Grauman



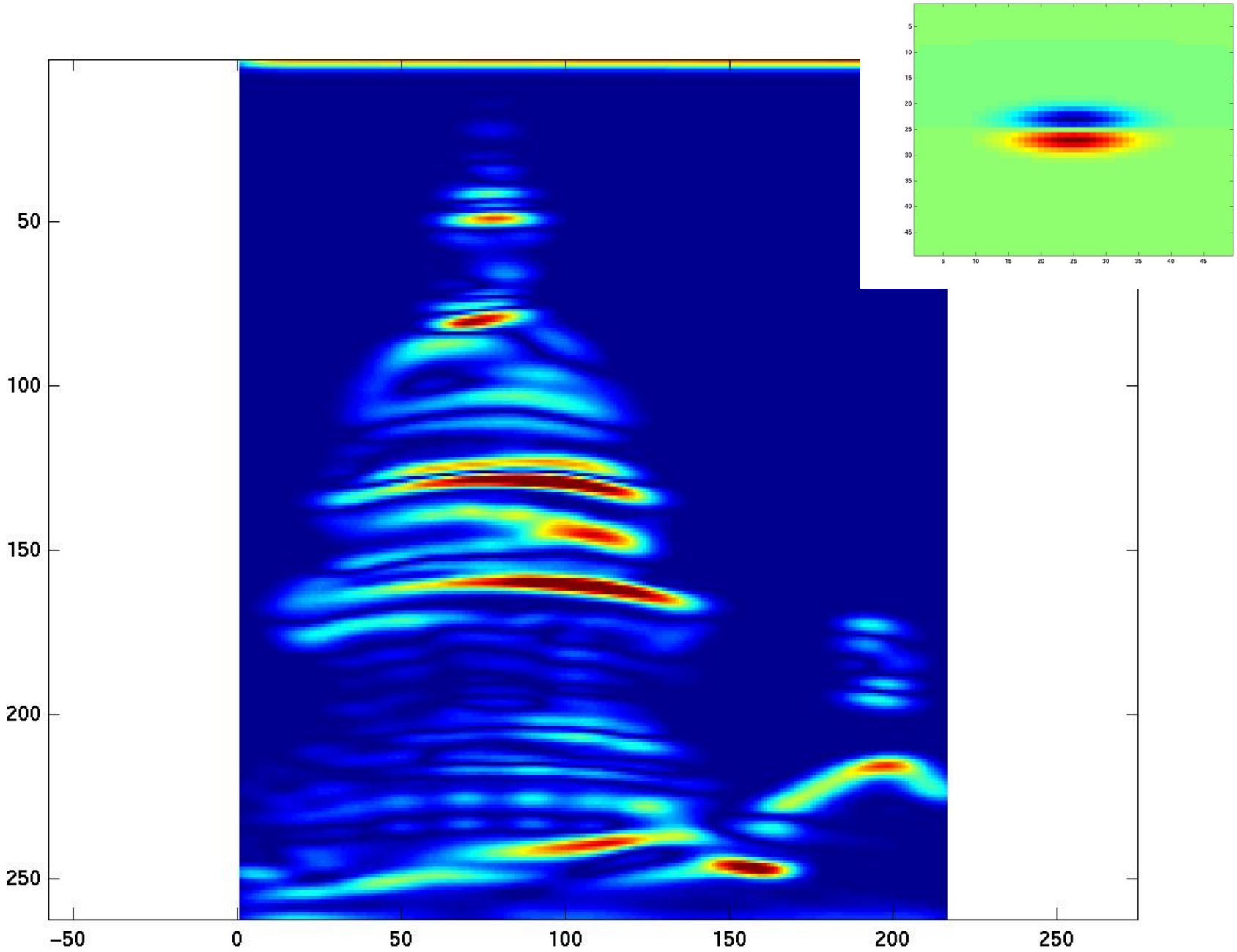
Kristen Grauman



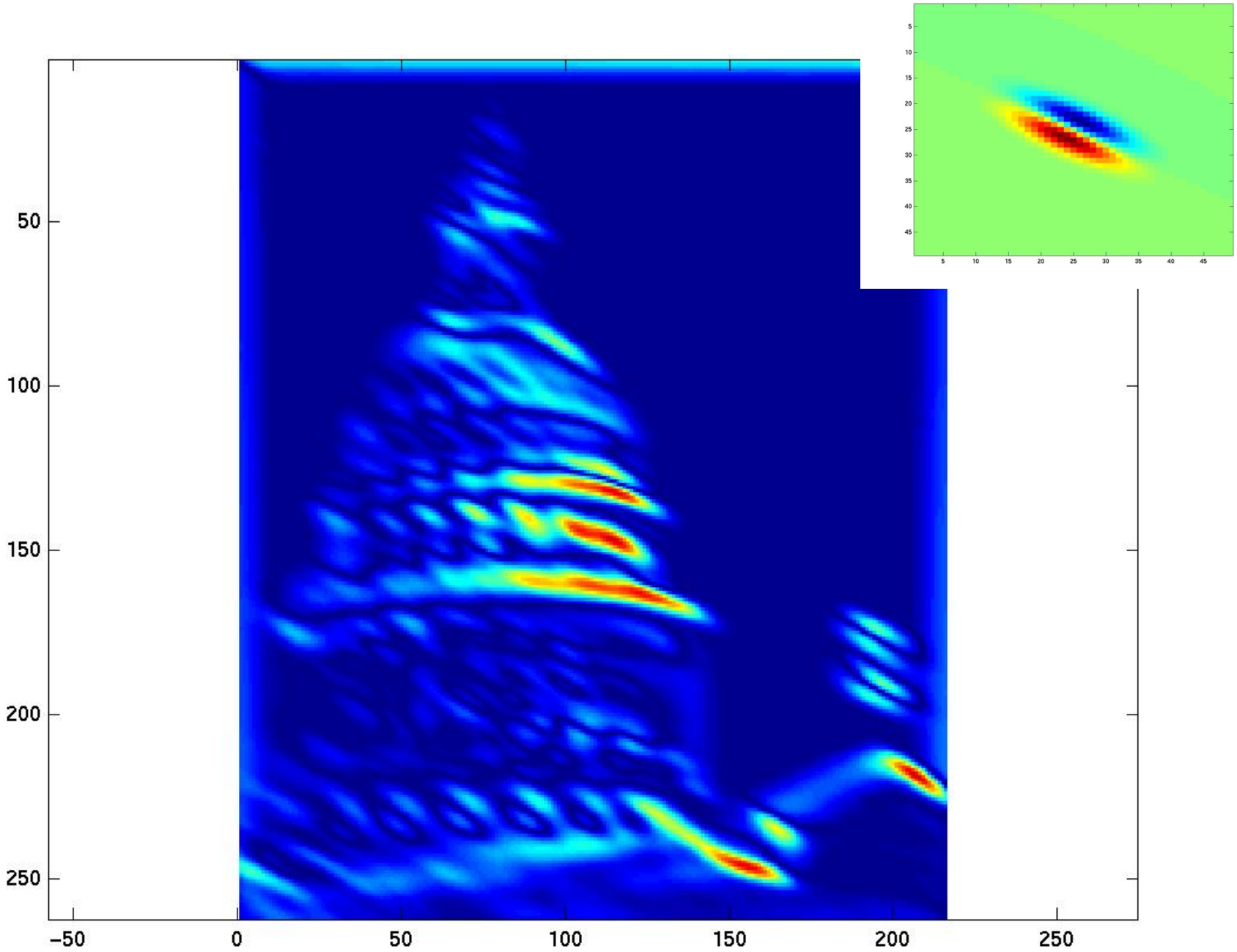
Kristen Grauman



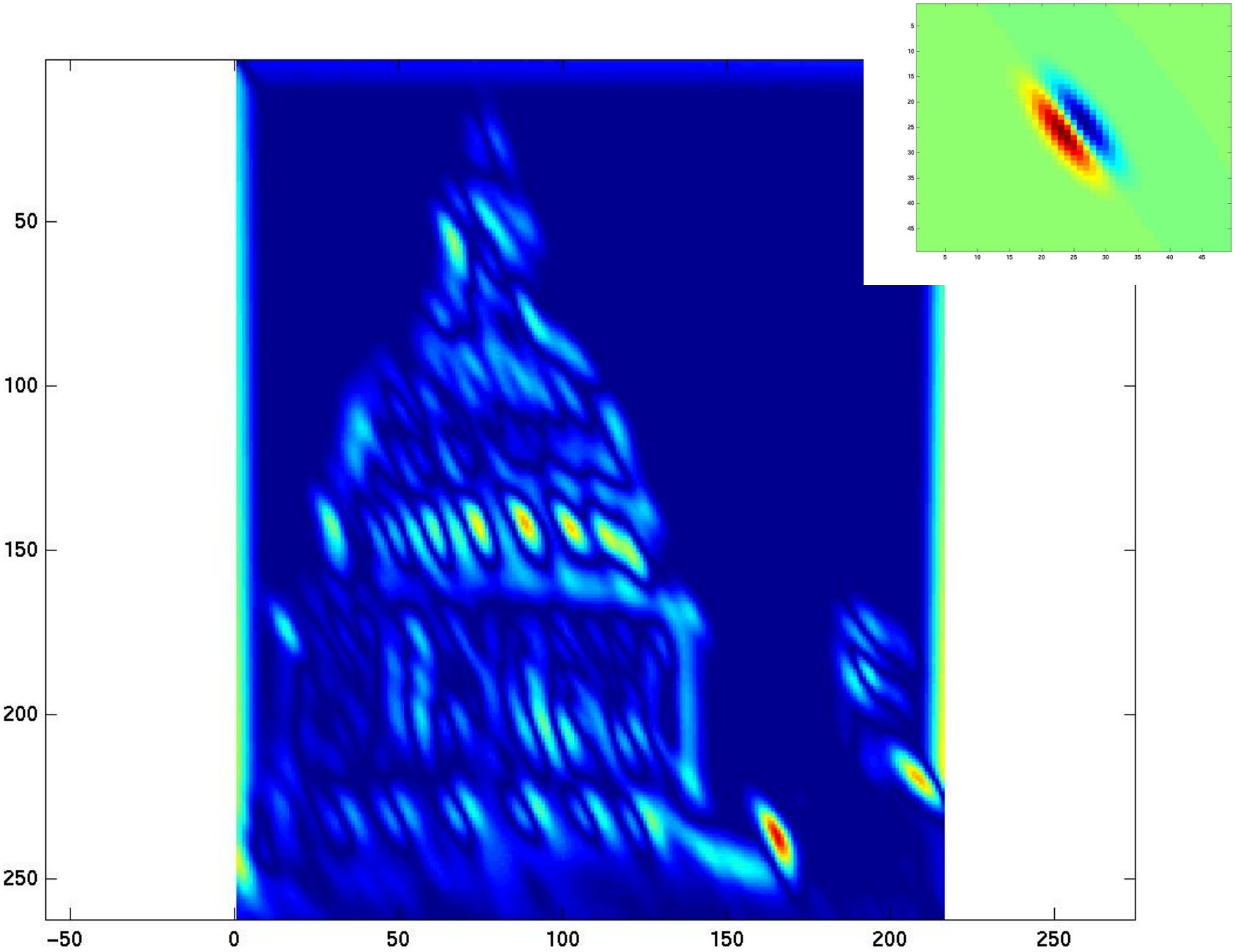
Kristen Grauman



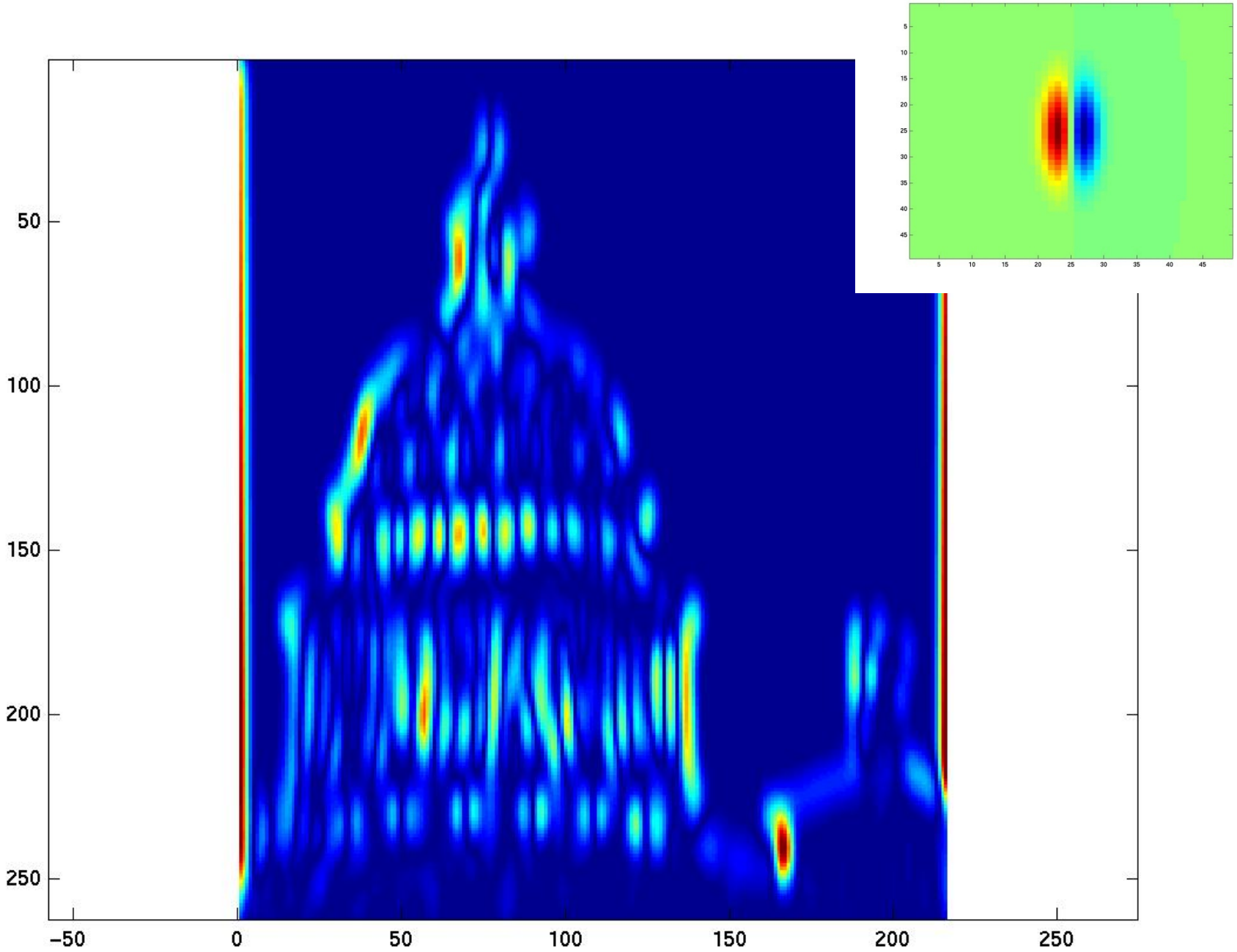
Kristen Grauman



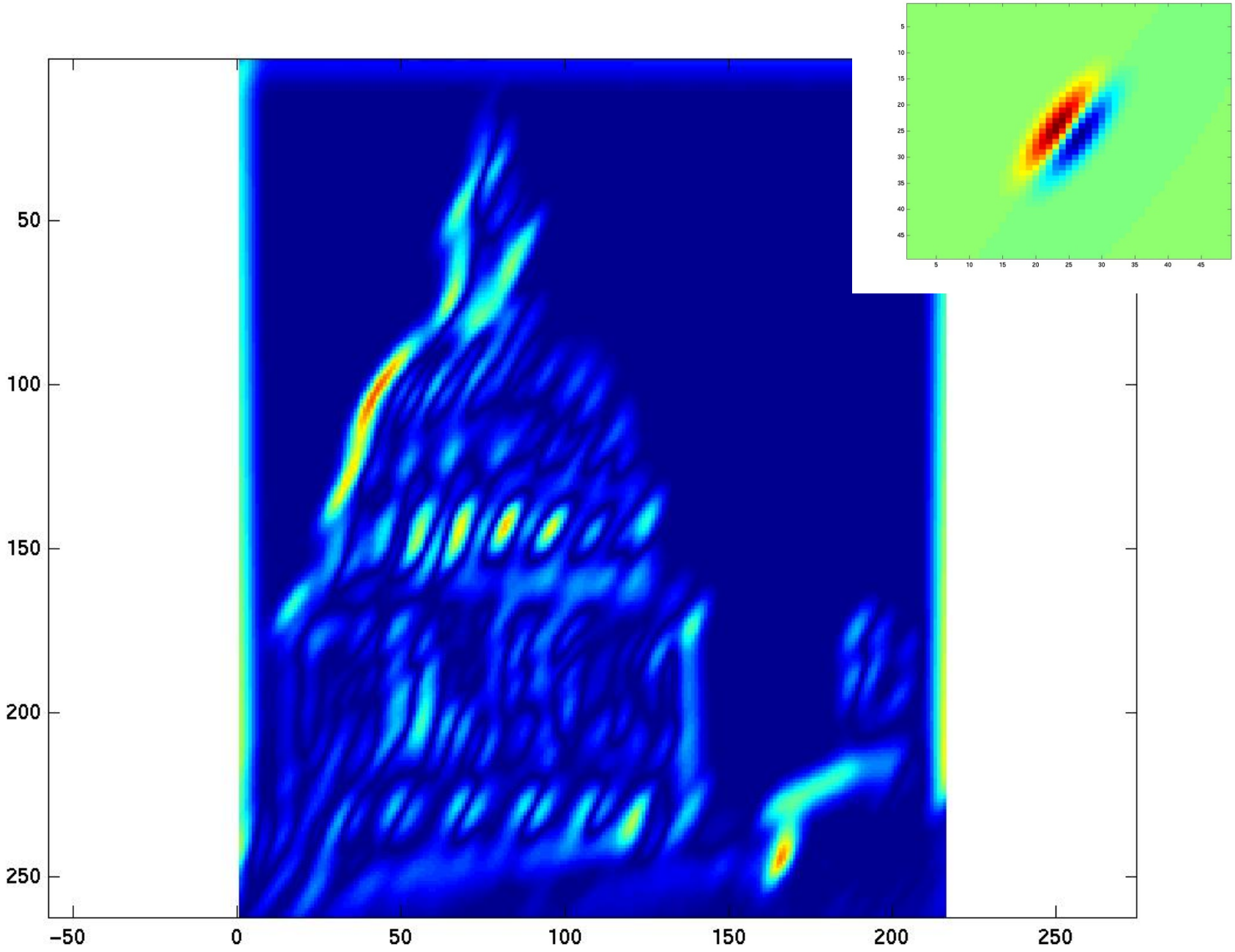
Kristen Grauman



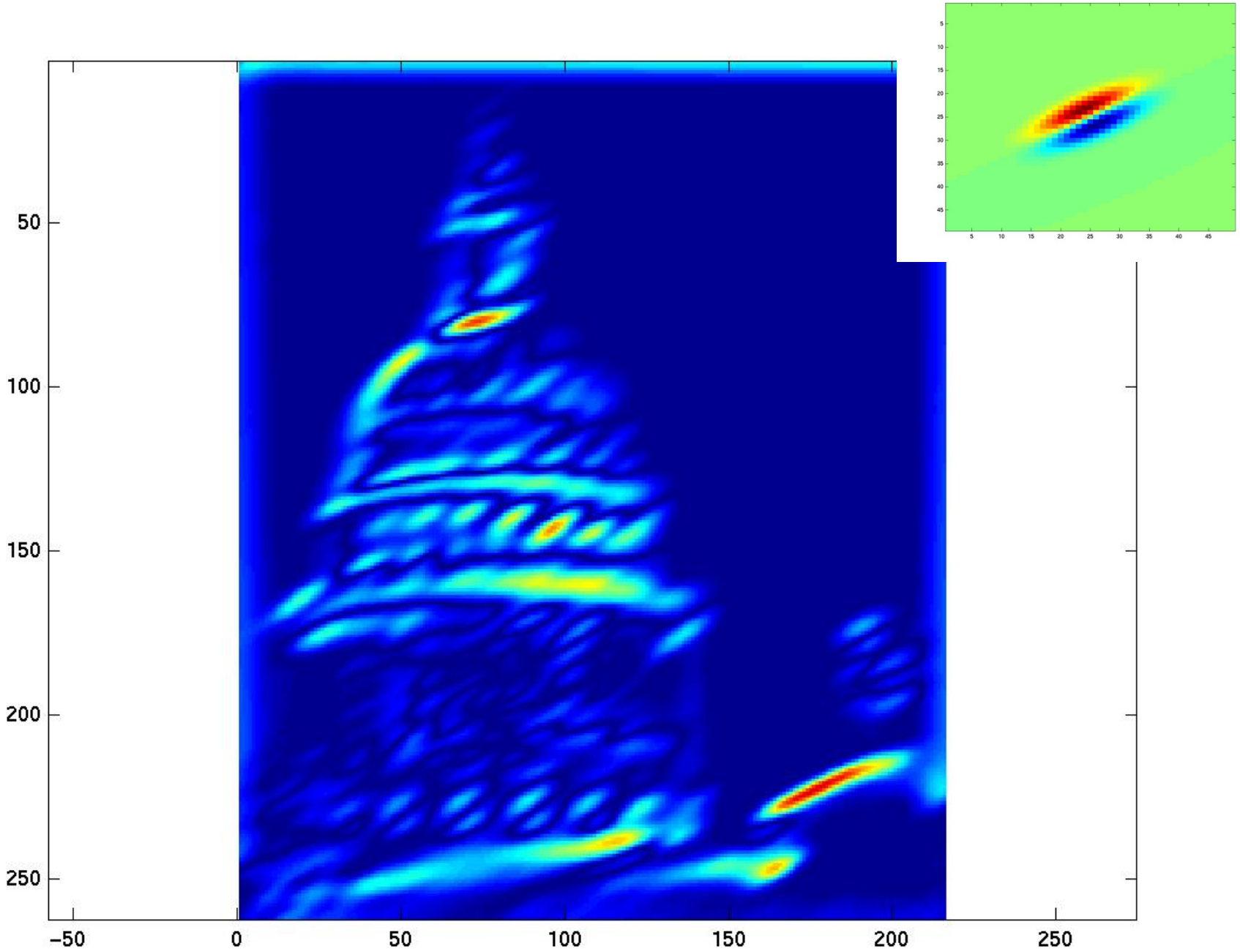
Kristen Grauman



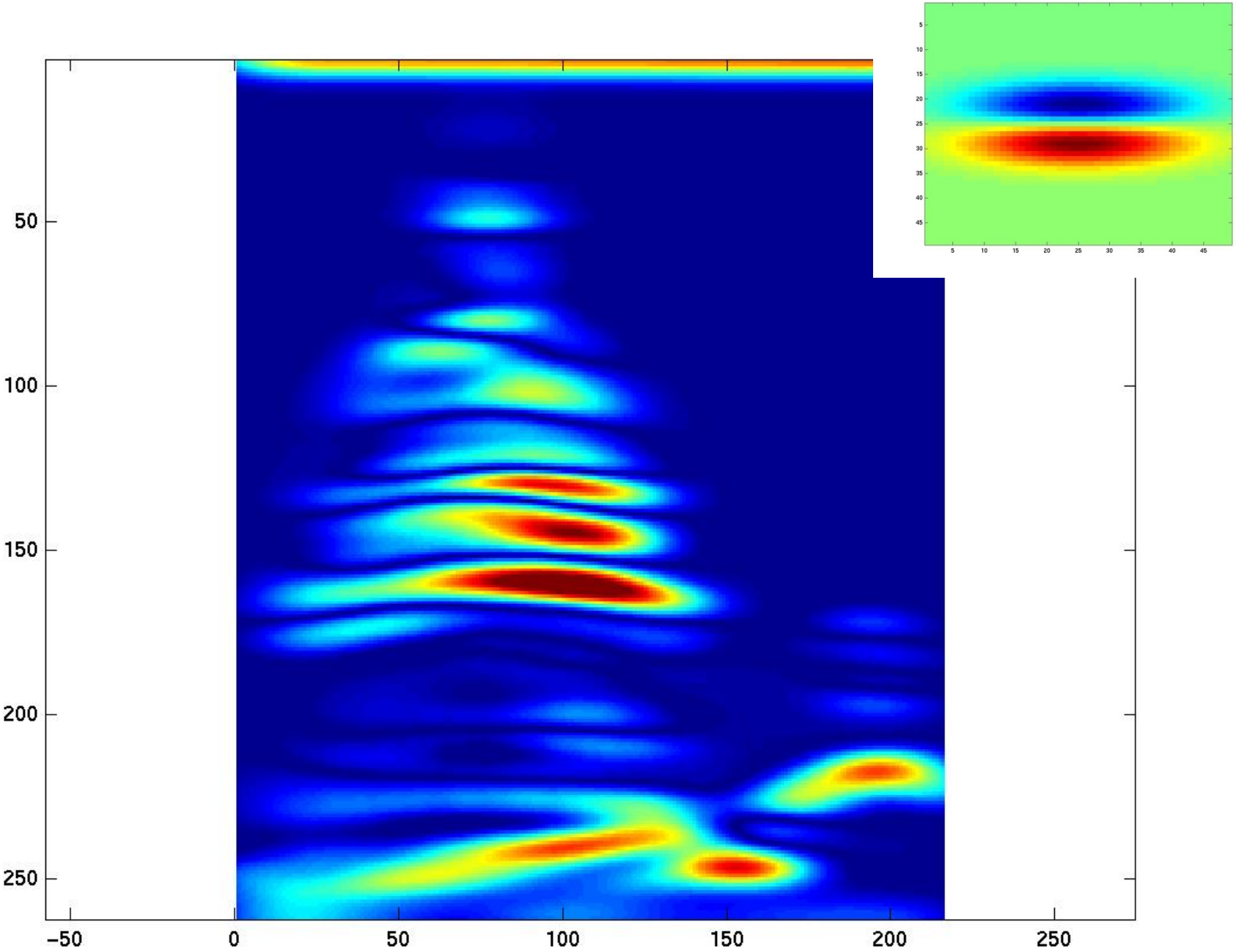
Kristen Grauman



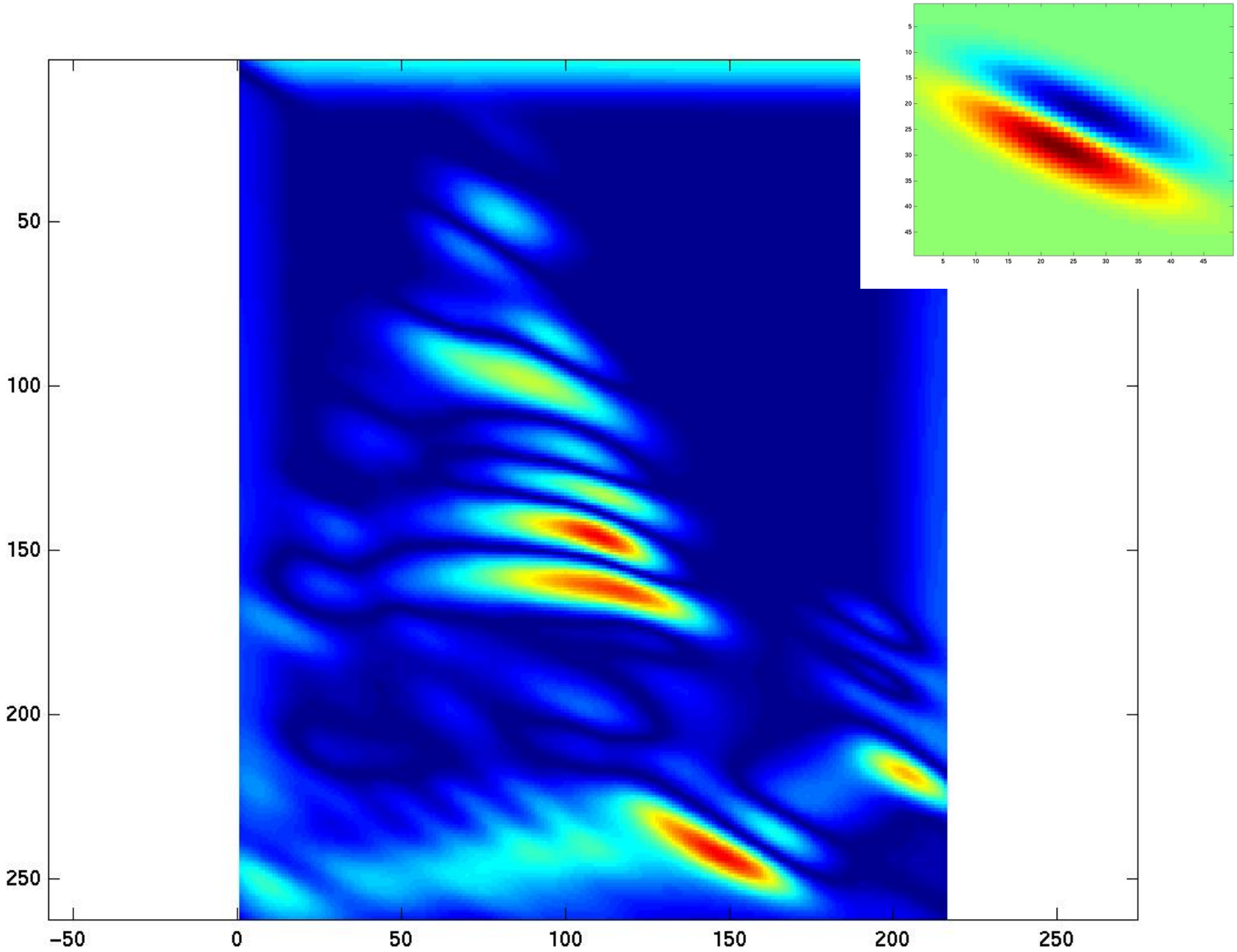
Kristen Grauman



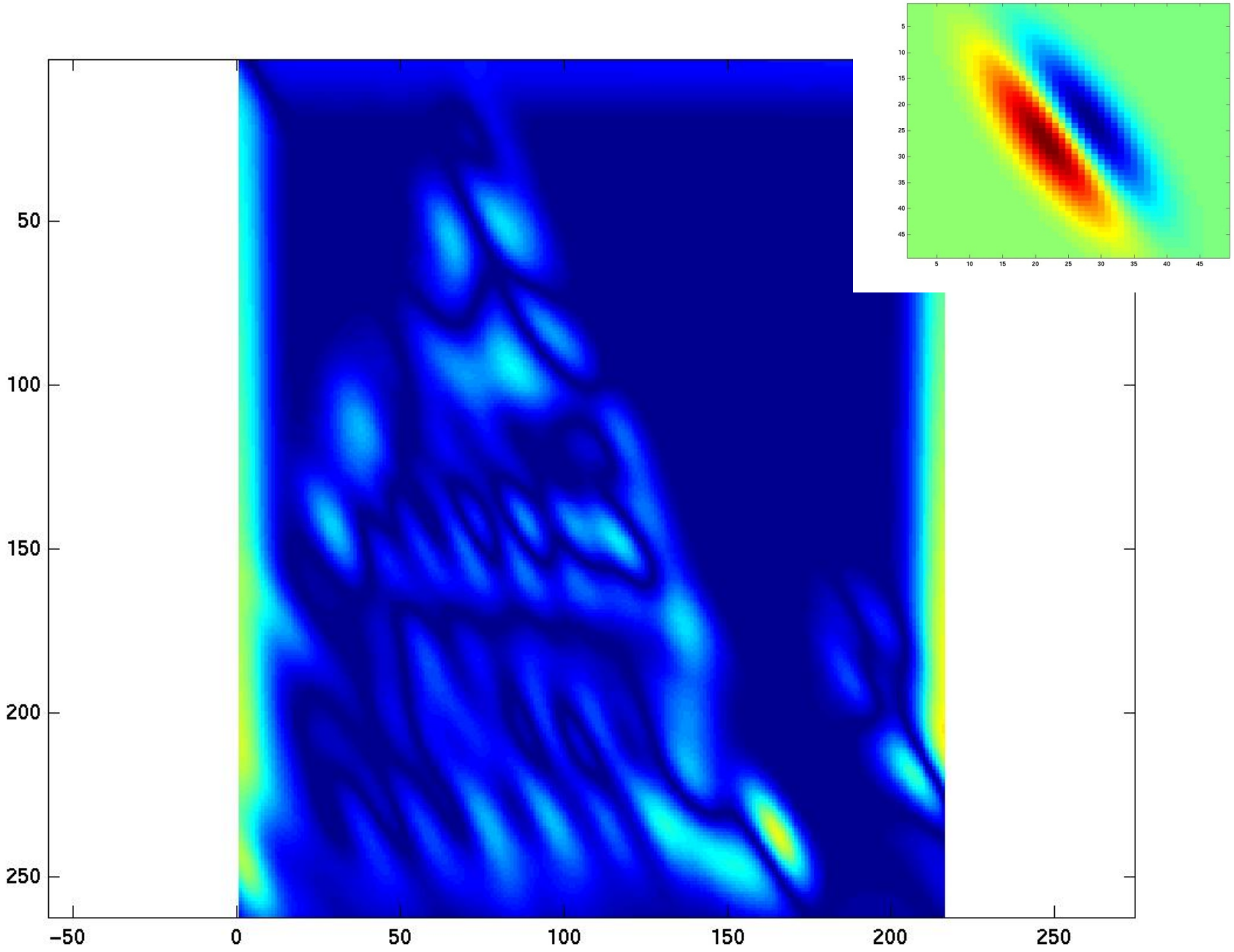
Kristen Grauman



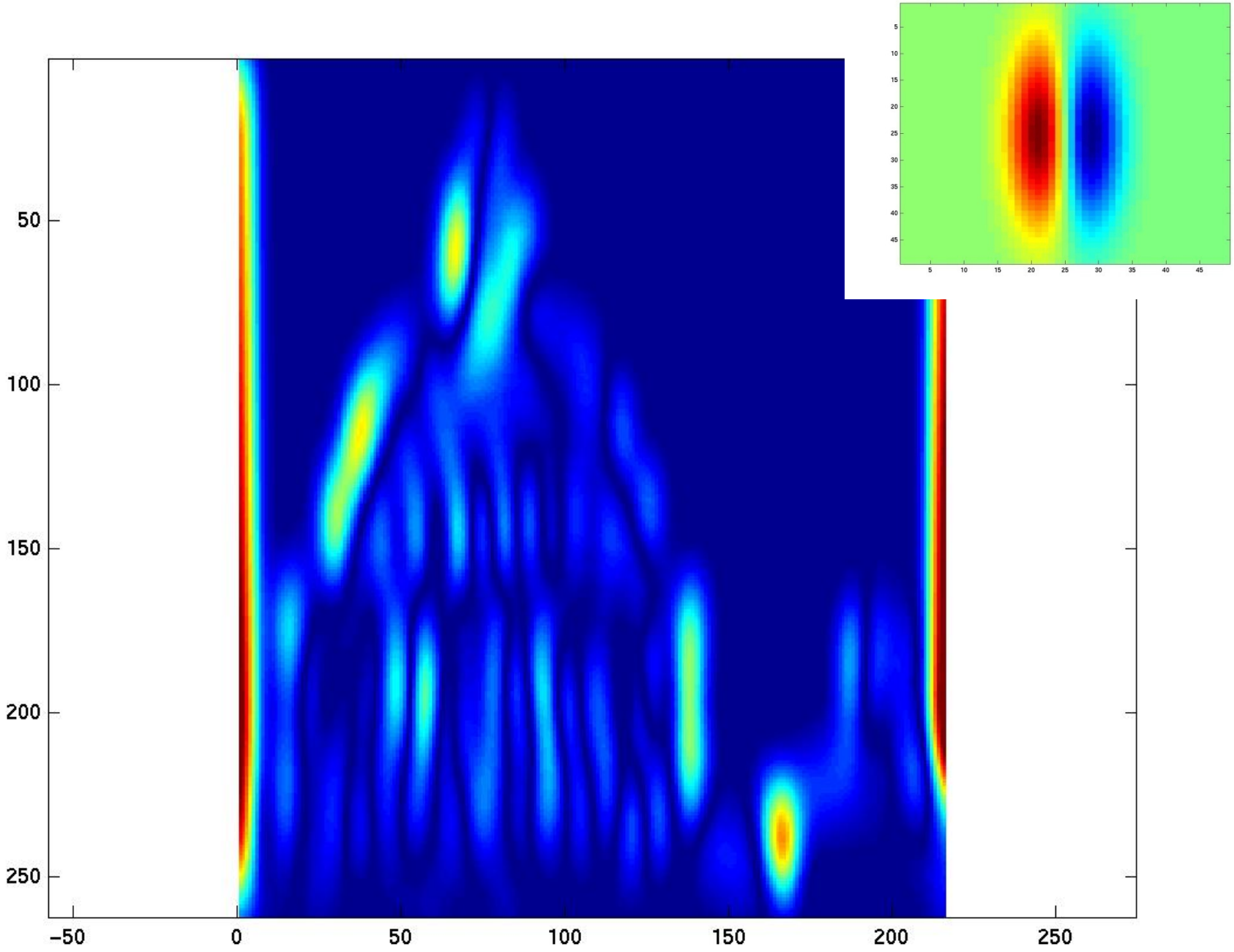
Kristen Grauman



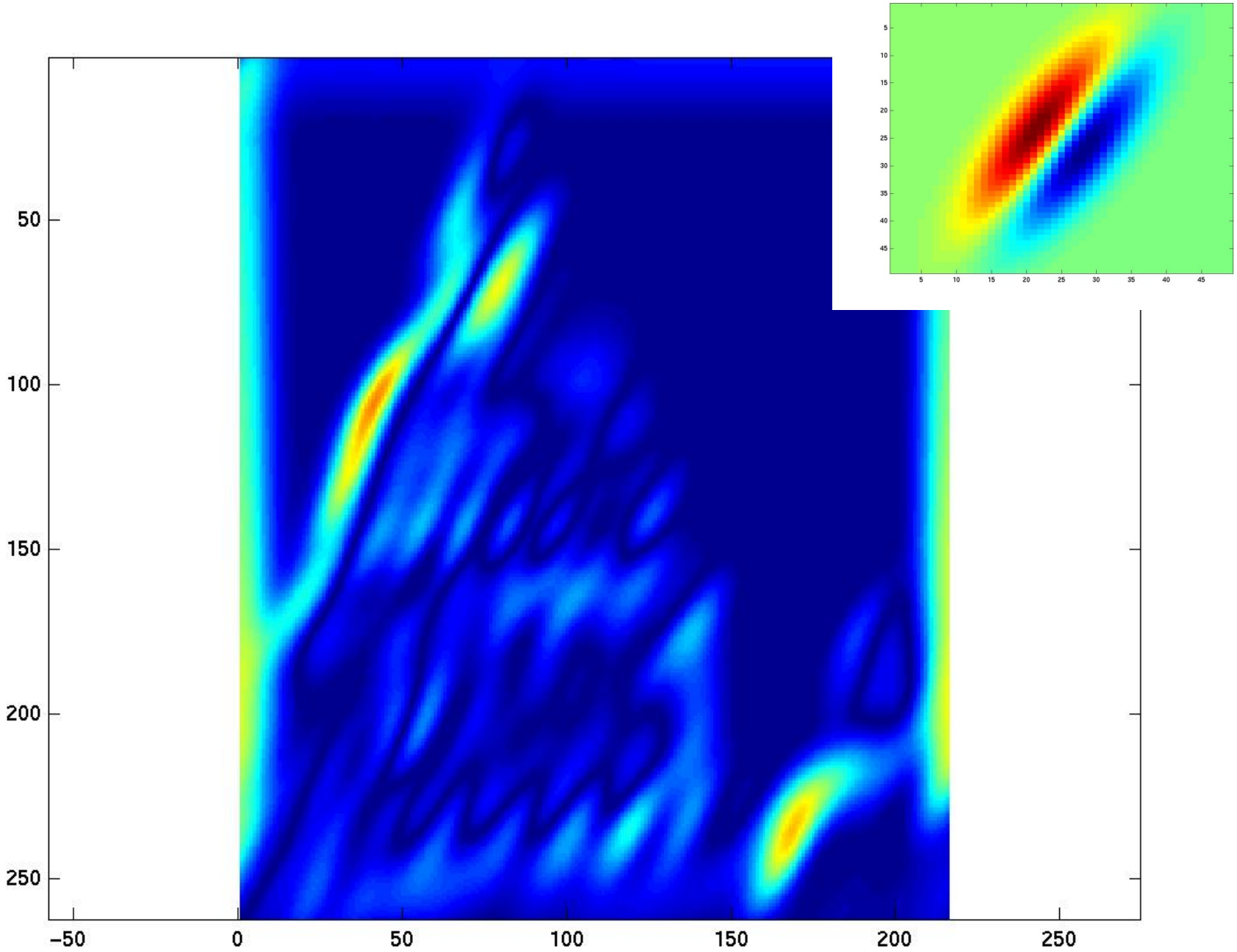
Kristen Grauman



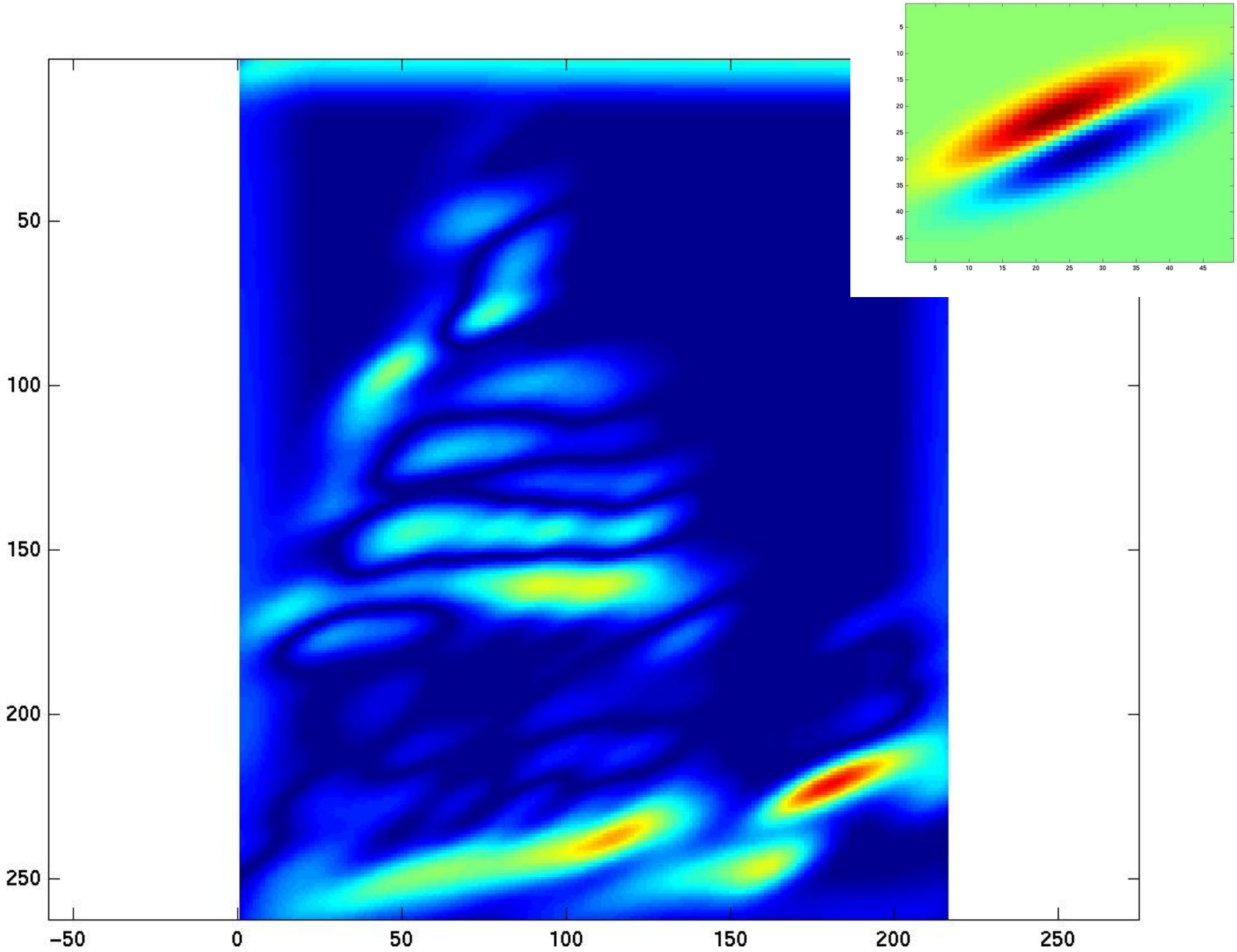
Kristen Grauman



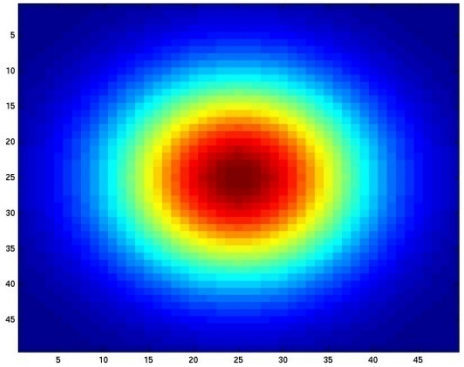
Kristen Grauman



Kristen Grauman



Kristen Grauman



Kristen Grauman

What does it capture?

v = F * Patch (where F is filter matrix)

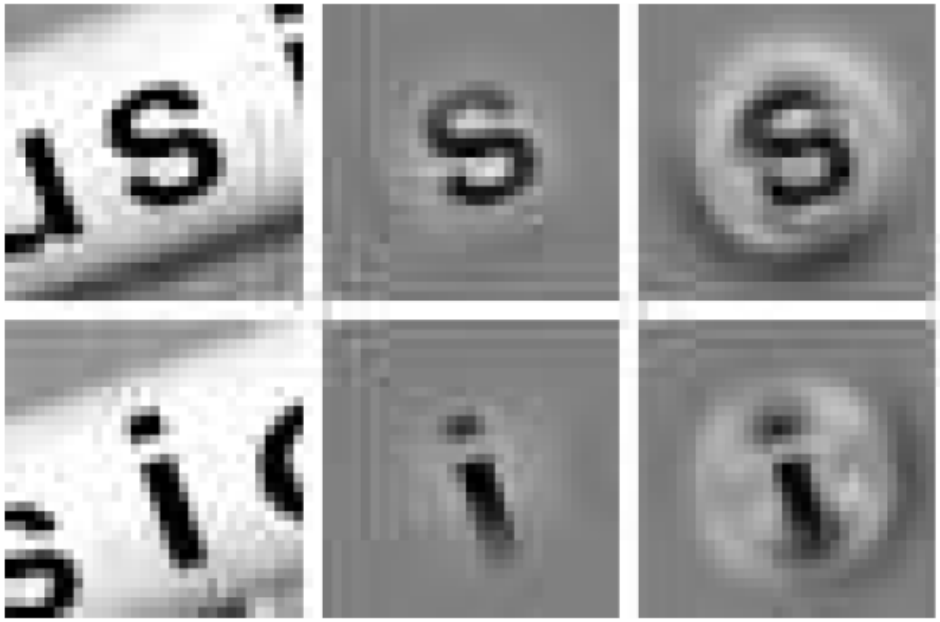


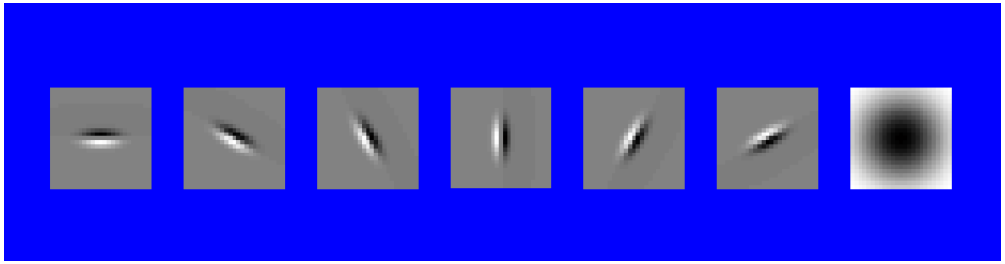
Fig. 3. Image reconstruction. Two example image patches (*left*), were reconstructed (*right*) from spatial filter responses at their center. Original image patches masked by a Gaussian (*middle*) are shown for comparison.

How can we represent visual information?

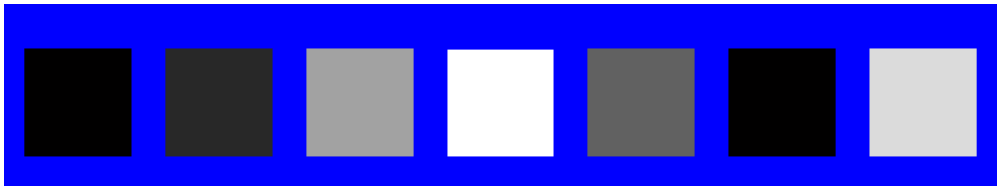
- Idea 1: Record simple statistics (e.g., mean, std.) of absolute filter responses

Can you match the texture to the response?

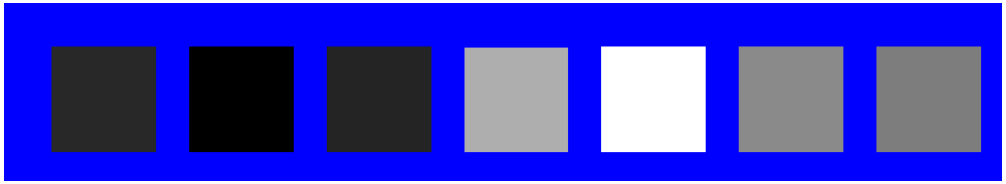
Filters



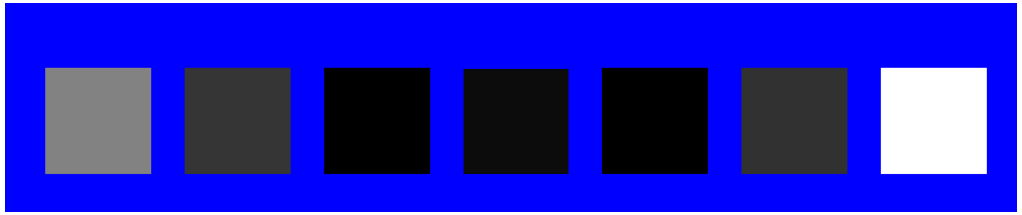
1



2

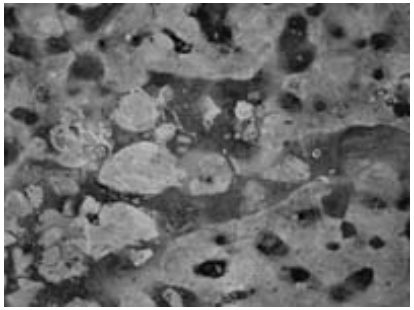


3



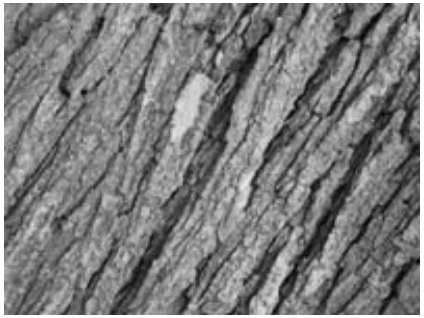
Mean abs responses

A



B

C



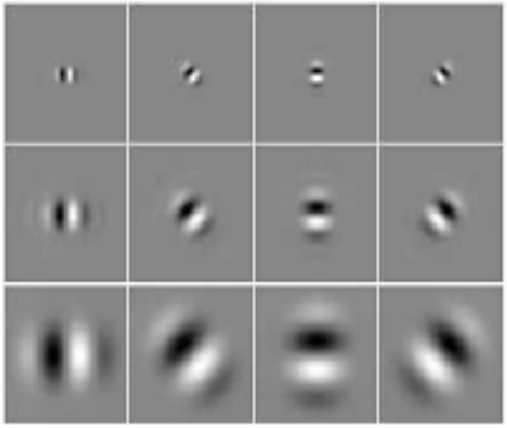
How can we represent visual information?

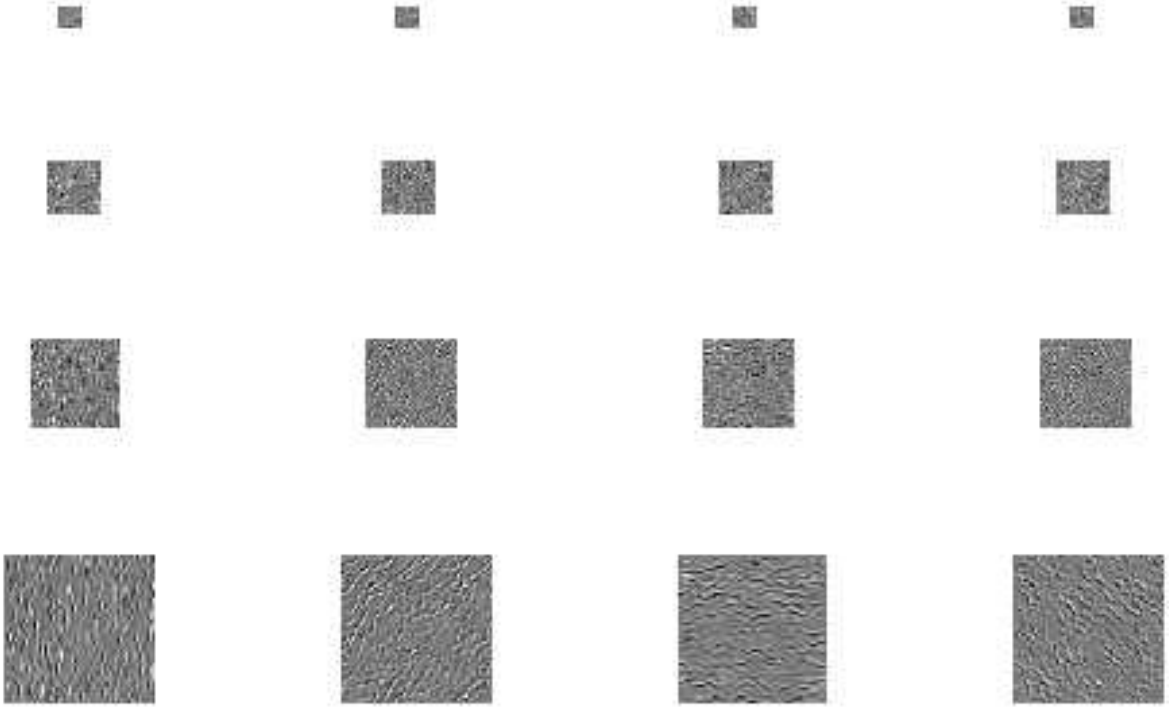
- Can be thought of as an single “orientation histogram”

- <u>Idea 2</u>: Marginal histograms of filter responses
  - one histogram per filter

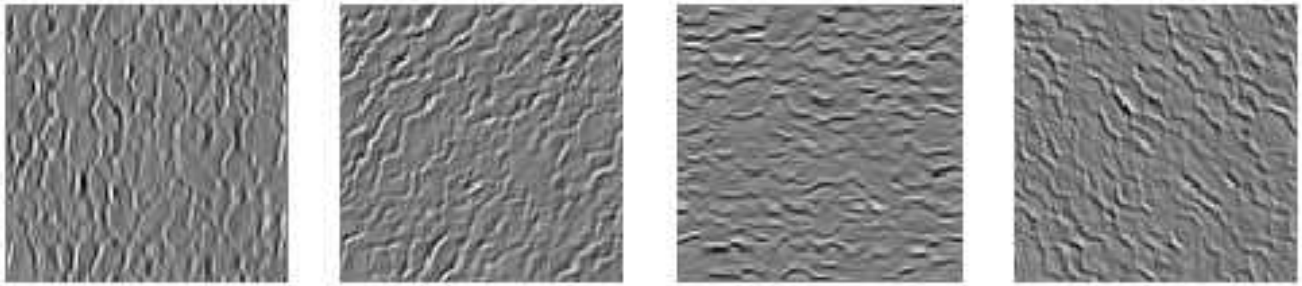
Multi-scale filter decomposition

Filter bank

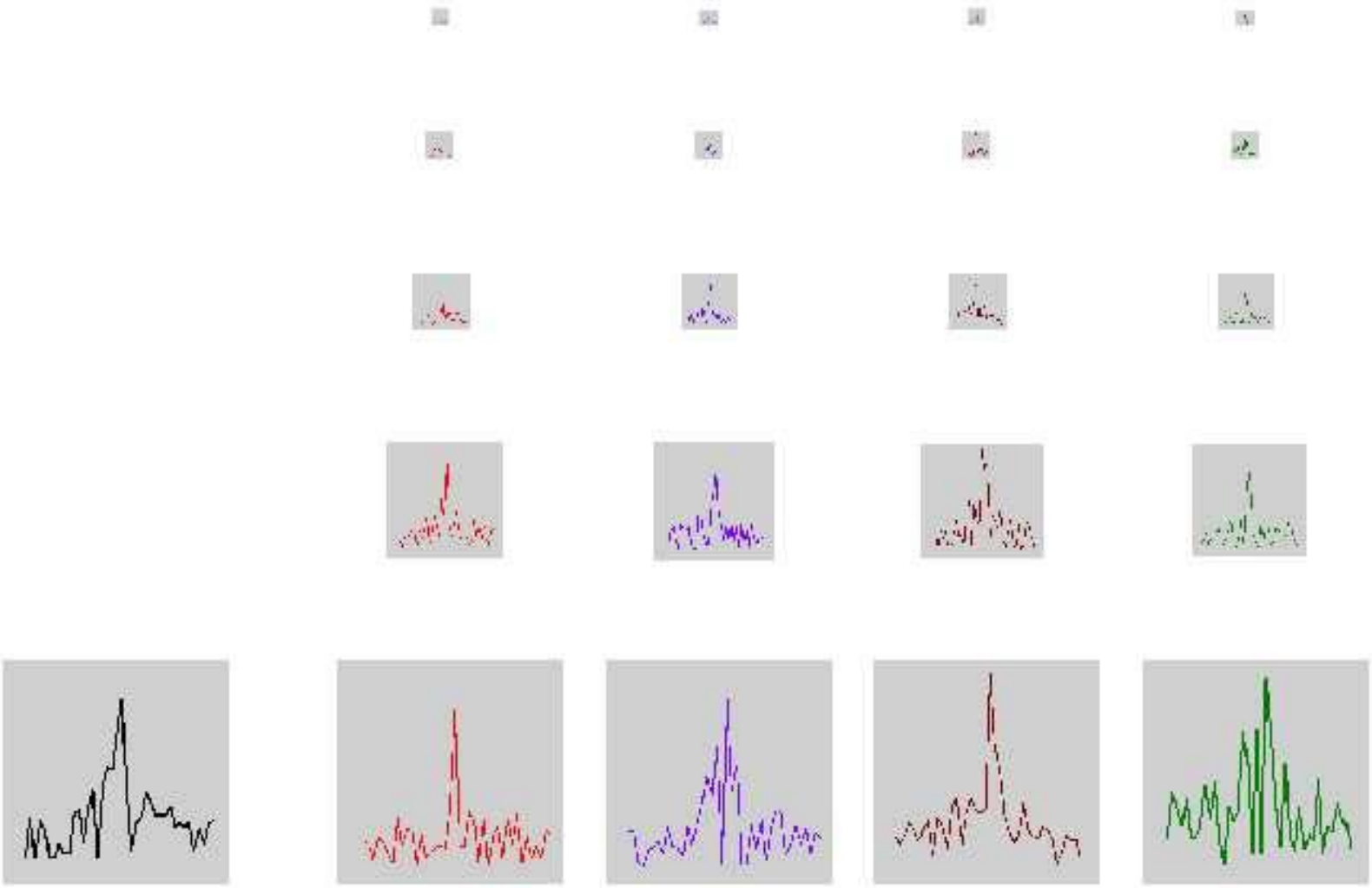




Input image



Filter response histograms

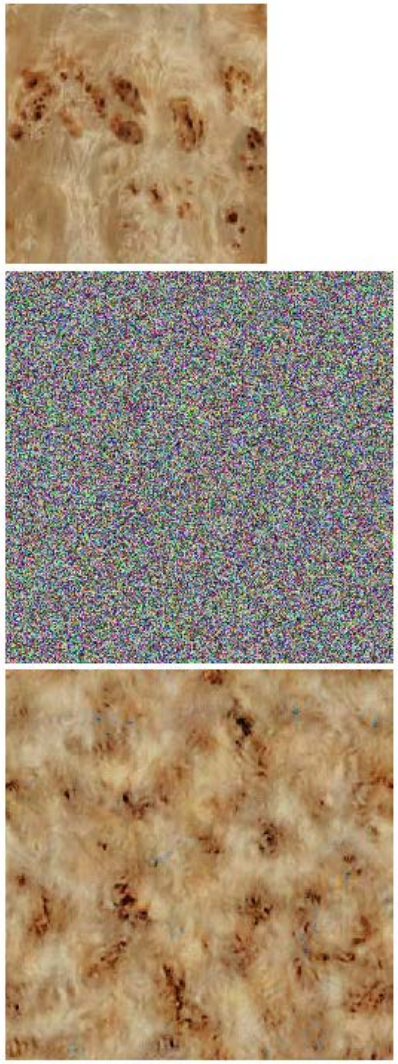


Heeger & Bergen, SIGGRAPH‘95

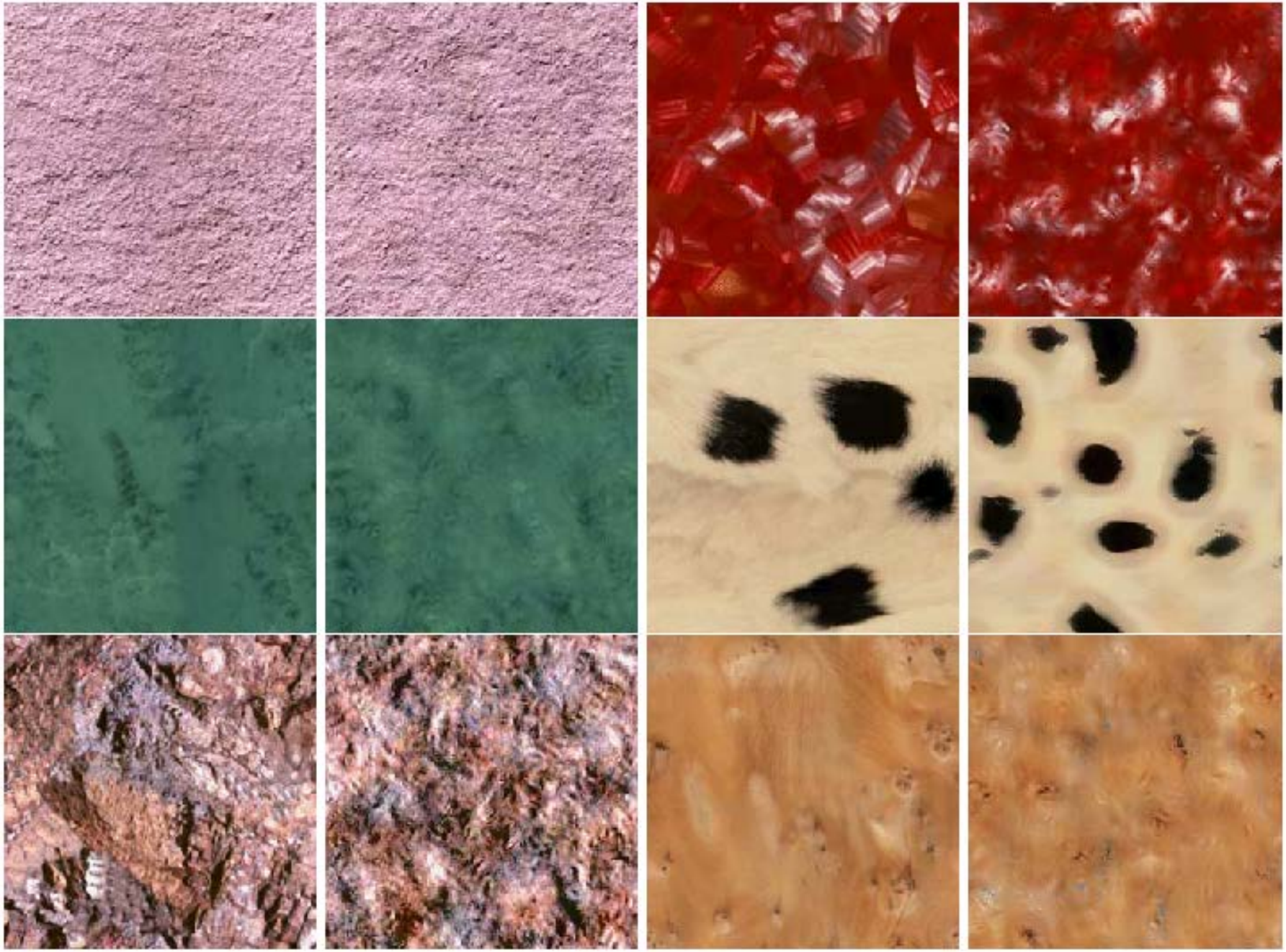
Start with a noise image as output

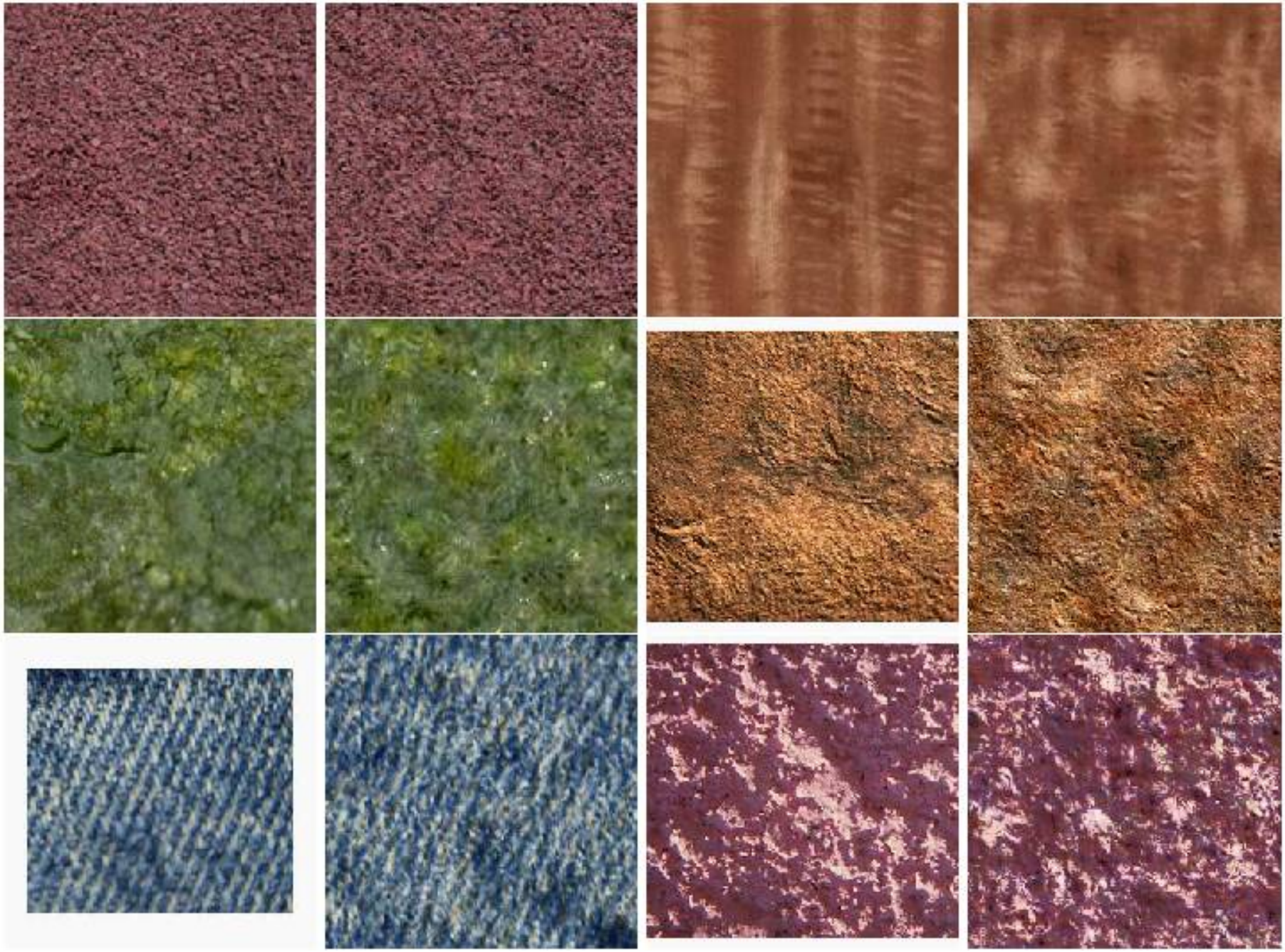
Main loop:

- Match pixel histogram of output image to input
- Decompose input and output images using multi-scale filter bank (Steerable Pyramid)
- Match subband histograms of input and output pyramids
- Reconstruct input and output images (collapse the pyramids)



Heeger, Bergen, Pyramid-based texture analysis/synthesis, SIGGRAPH 1995





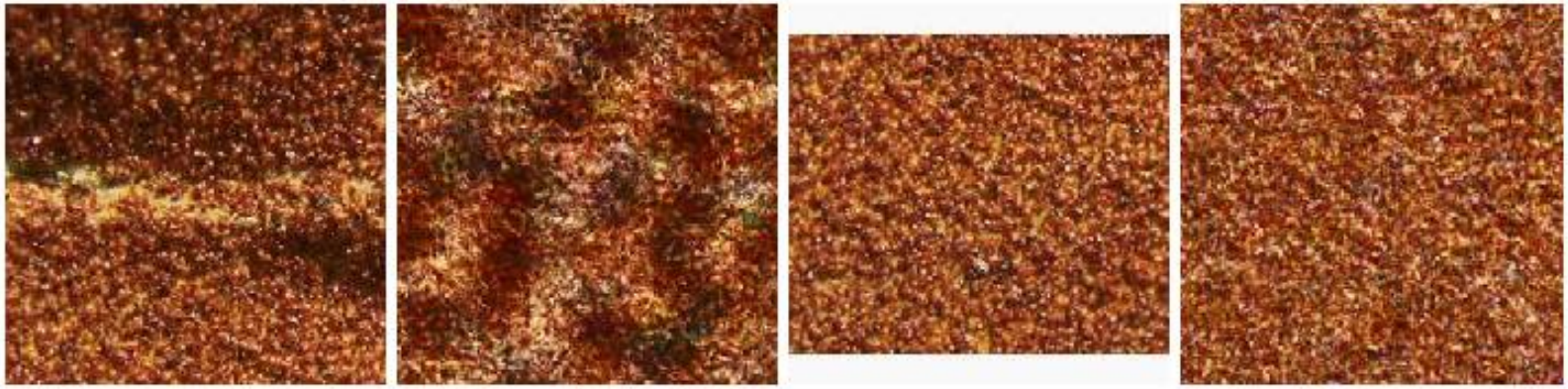


Figure 7: (Left pair) Inhomogeneous input texture produces blotchy synthetic texture. (Right pair) Homogenous input.

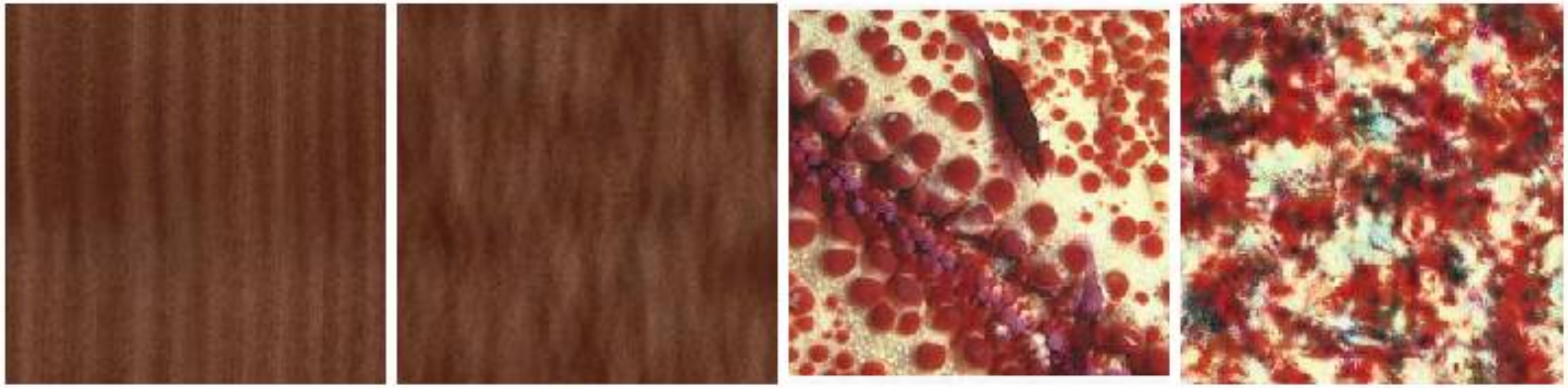


Figure 8: Examples of failures: wood grain and red coral.

Figure 9: More failures: hay and marble.

Simoncelli & Portilla ’98+

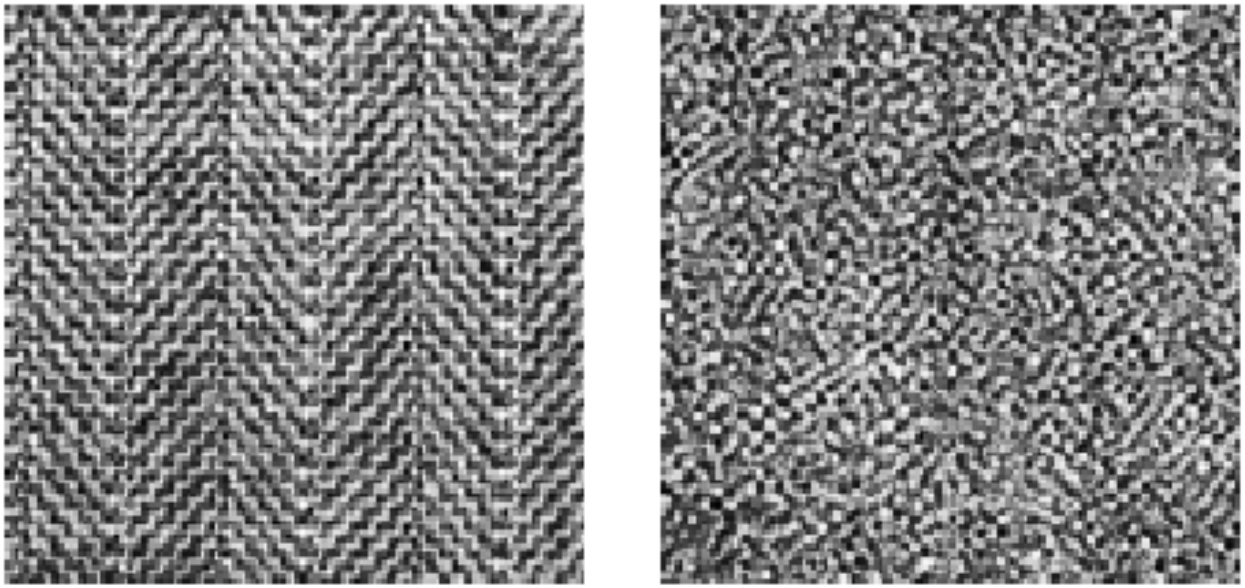


Figure 1. Textures with matching marginal statistics.

Marginal statistics are not enough

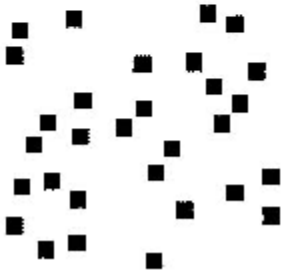
Neighboring filter responses are highly correlated

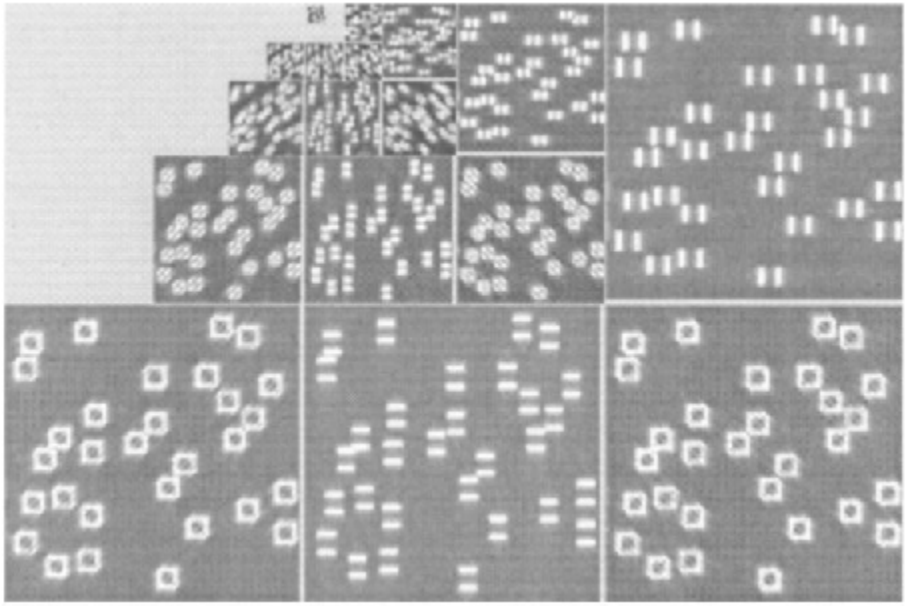
- an edge at low-res will cause an edge at high-res

Let’s match 2nd order statistics too!

J Portilla and E P Simoncelli. *A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients.* Int'l Journal of Computer Vision. 40(1):49-71, October, 2000.

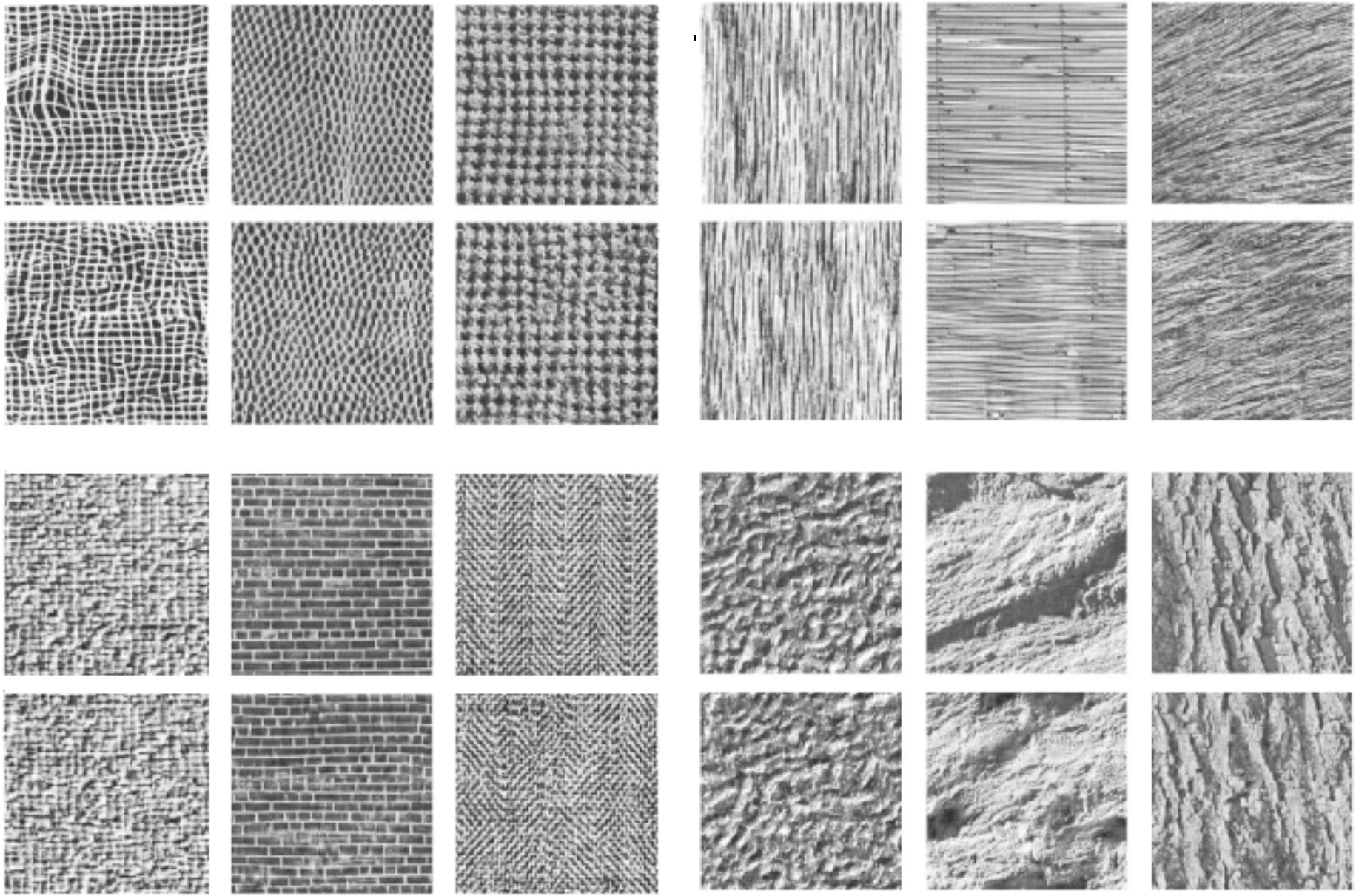
Simoncelli & Portilla ’98+





Match joint histograms of pairs of filter responses at adjacent spatial locations, orientations, and scales.

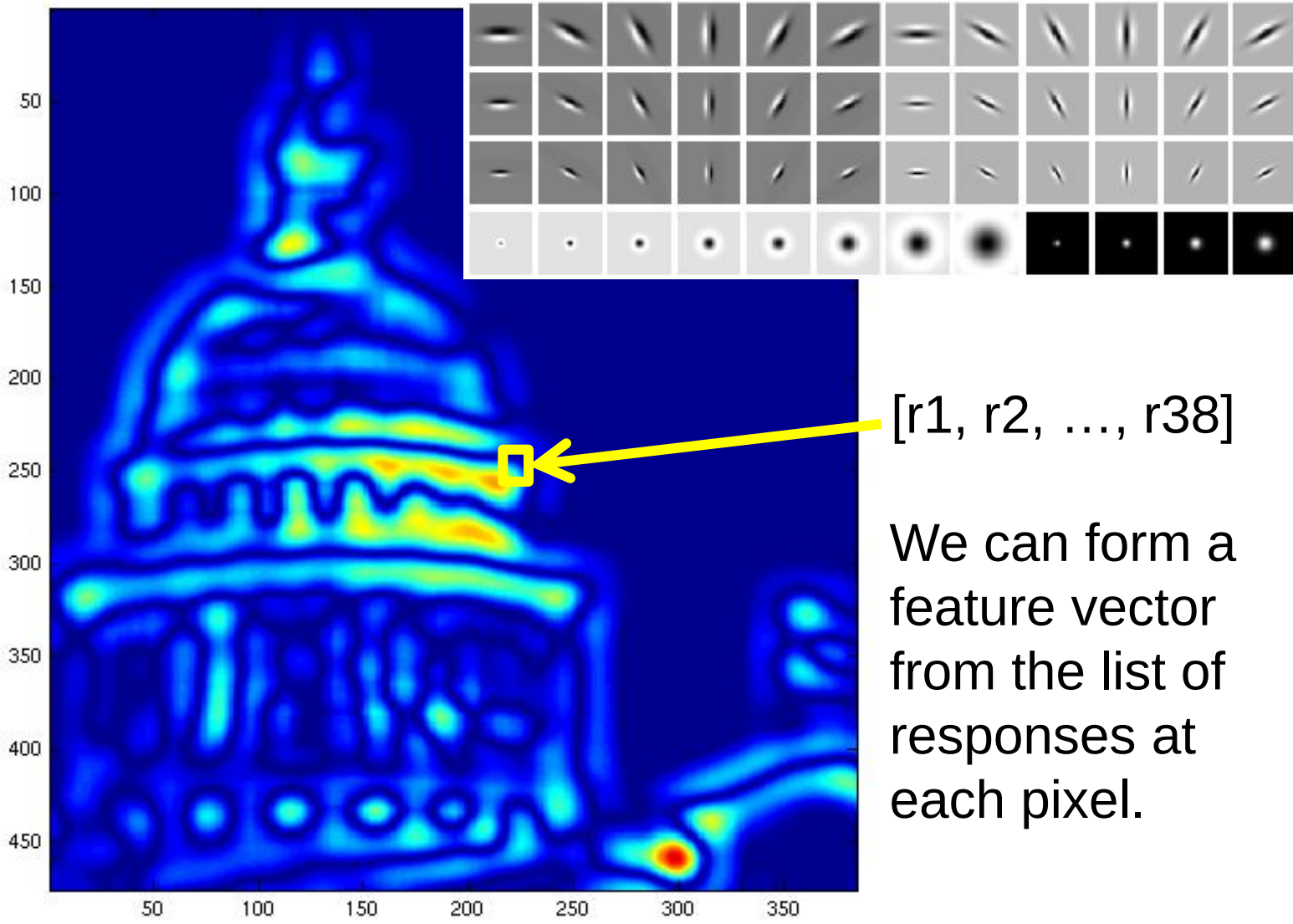
Optimize using repeated projections onto statistical constraint sufraces



How can we represent visual informaiton?

- Marginal filter response histograms don't talk to each other (in a direct way)
- <u>Idea 3</u>: Histograms of joint responses (textons)

Filter Response

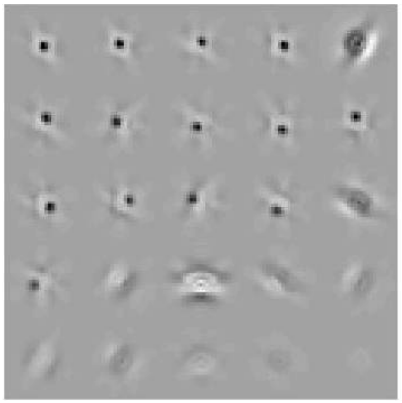


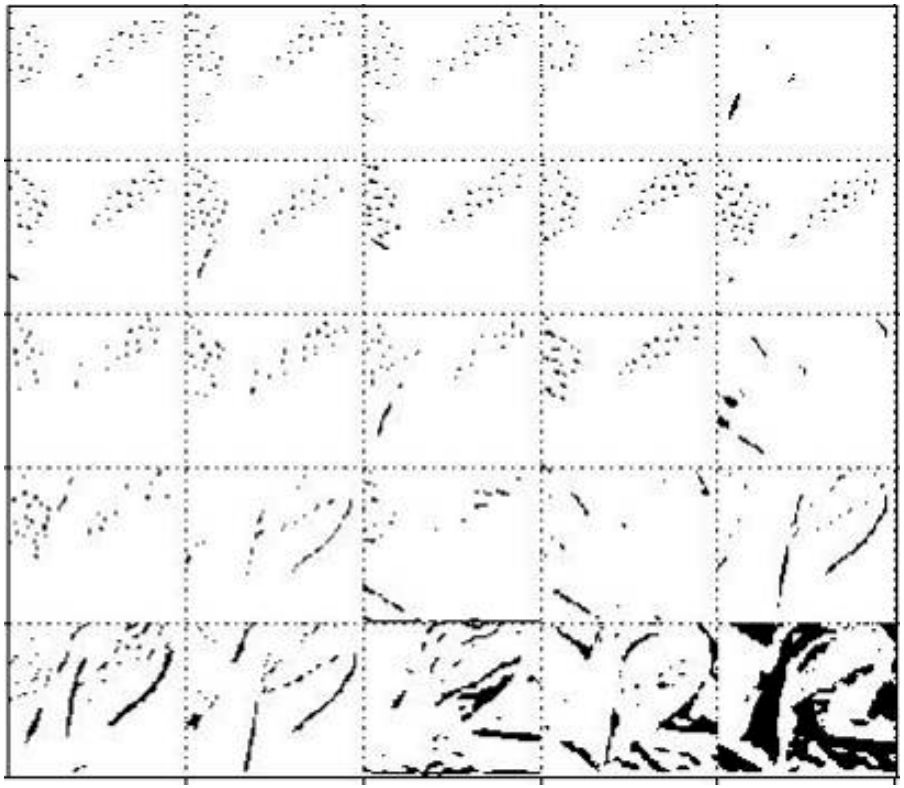
[r1, r2, …, r38]

We can form a feature vector from the list of responses at each pixel.

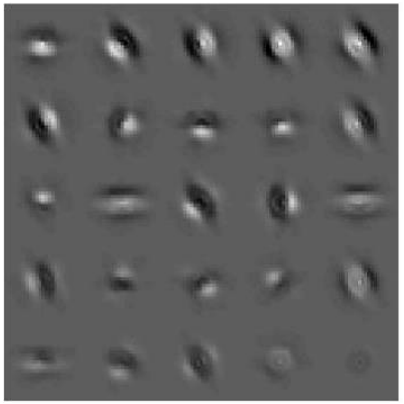
Textons (Malik et al, IJCV 2001)

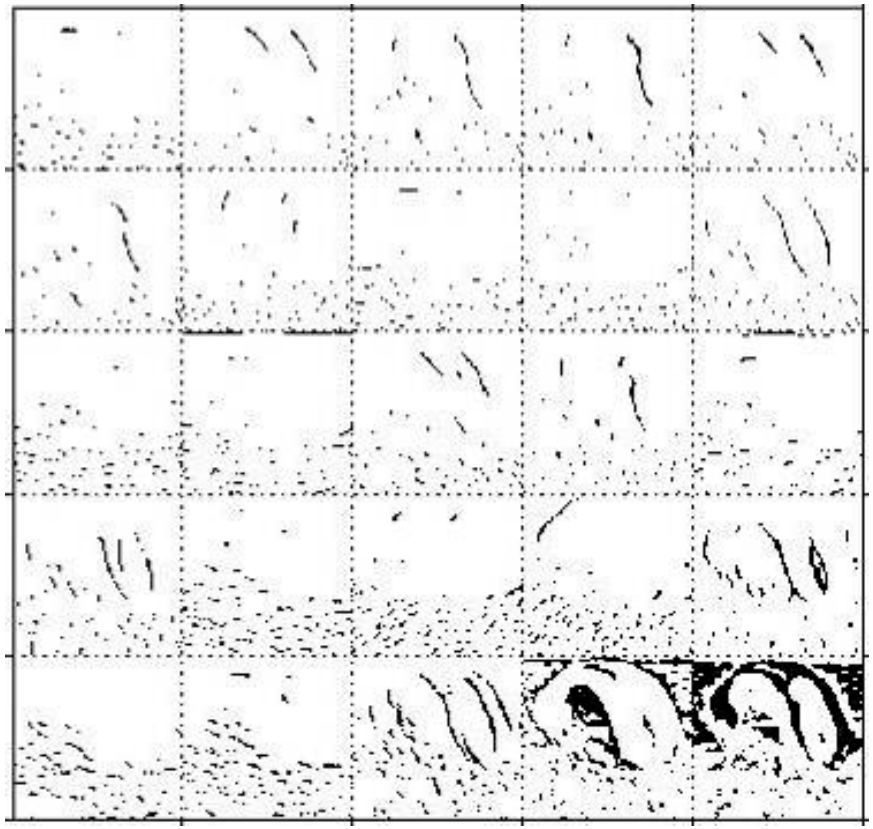
- Cluster vectors of filter responses





Textons (cont.)





Object → **Bag of ‘words’**

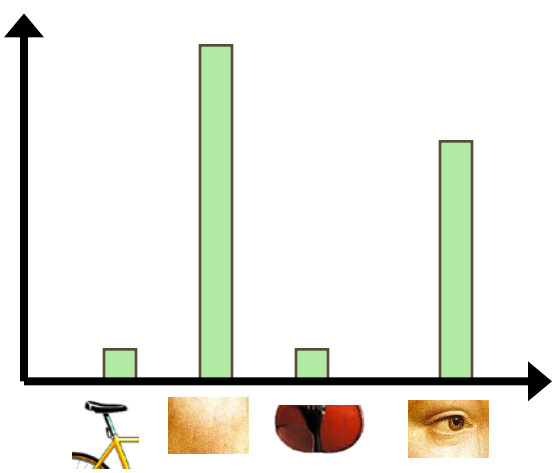
Analogy to documents

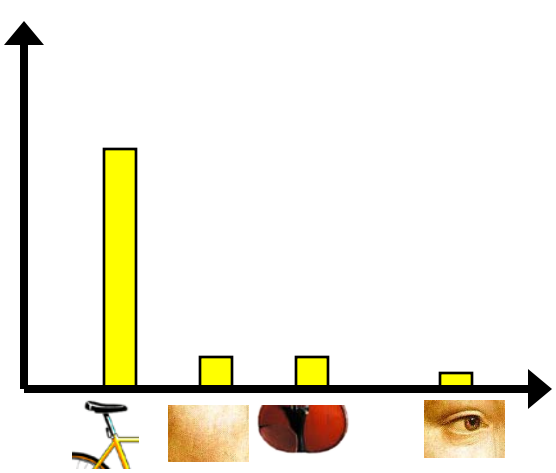
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that r[illegible] our eyes. For a long ti[illegible] retinal image wa[illegible] isual centers i[illegible] s a movie s[illegible] image [illegible] discove[illegible] know th[illegible] perceptio[illegible] more com[illegible] following the [illegible] ath to the various c[illegible] ortex, Hubel and Wiesel ha[illegible] demonstrate that the *message abou[illegible] image falling on the retina undergoes [illegible]-wise analysis in a system of nerve cell[illegible] stored in columns. In this system each c[illegible] has its specific function and is responsible [illegible] a specific detail in the pattern of the retinal image.*

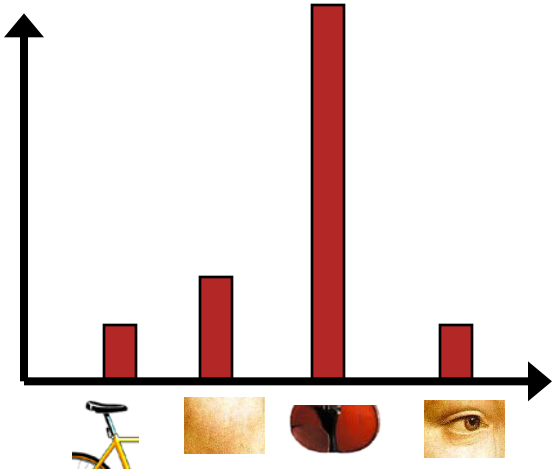
sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel

China is forecasting a trade surplus of $90bn (£51bn) to $100bn this year, a threefold increase on 2004's $32bn. The Commerce Ministry said the surplus would be created by a predicted 30% [illegible] $750bn, compared w[illegible] $660bn. T[illegible] annoy th[illegible] China's [illegible] deliber[illegible] agrees [illegible] yuan is [illegible] governo[illegible] also need[illegible] demand so [illegible] country. China [illegible] e yuan against the do[illegible] nd permitted it to trade within a narrow [illegible] but the US wants the yuan to be allowed [illegible] de freely. However, Beijing has made it c[illegible] t it will take its time and tread carefully be[illegible] allowing the yuan to rise further in value.

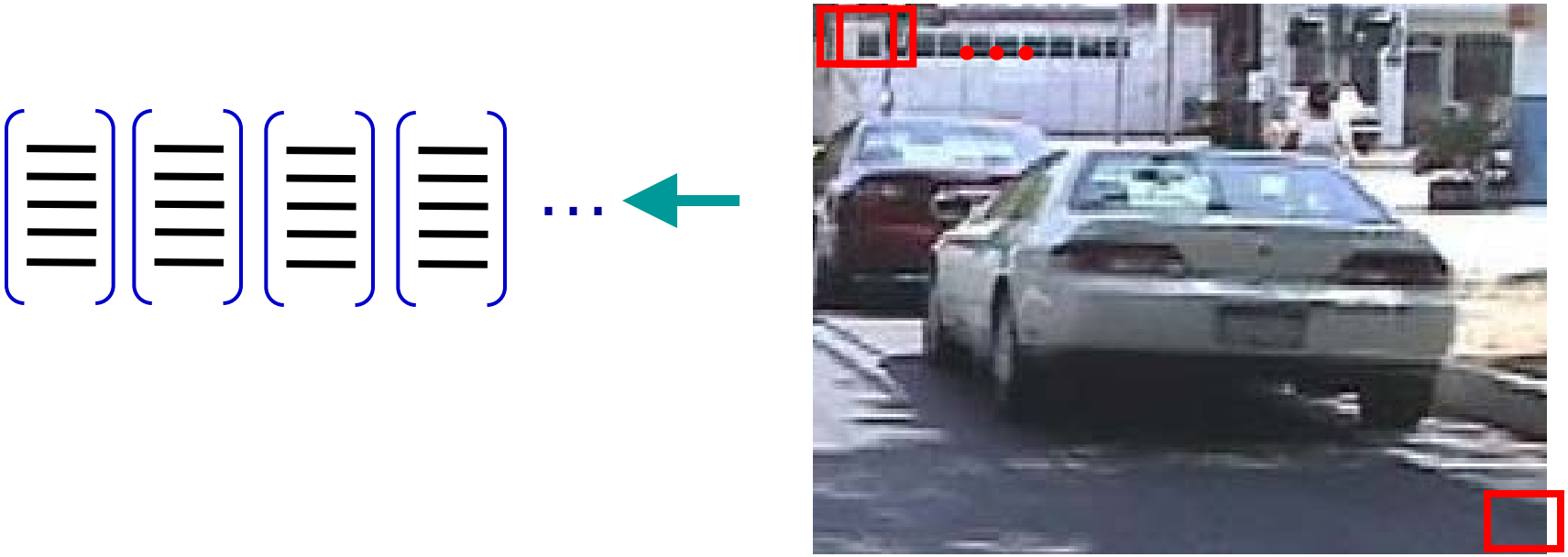
China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value



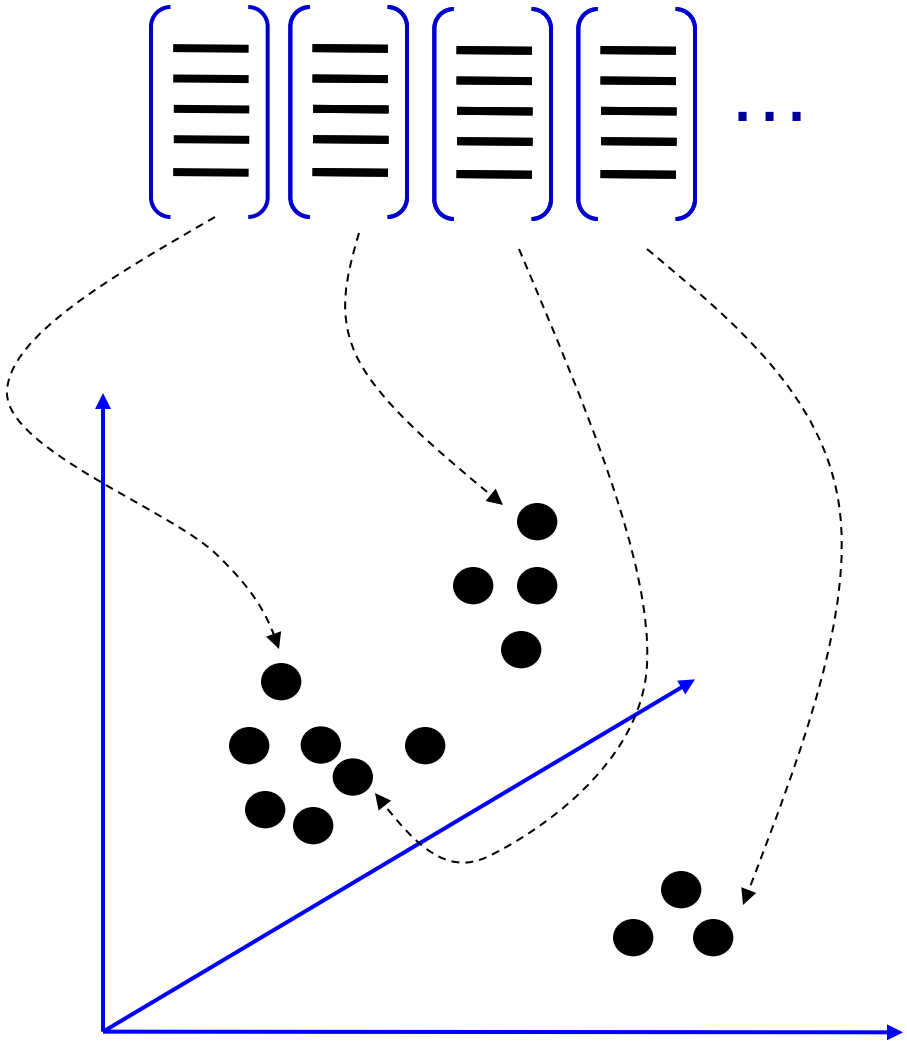




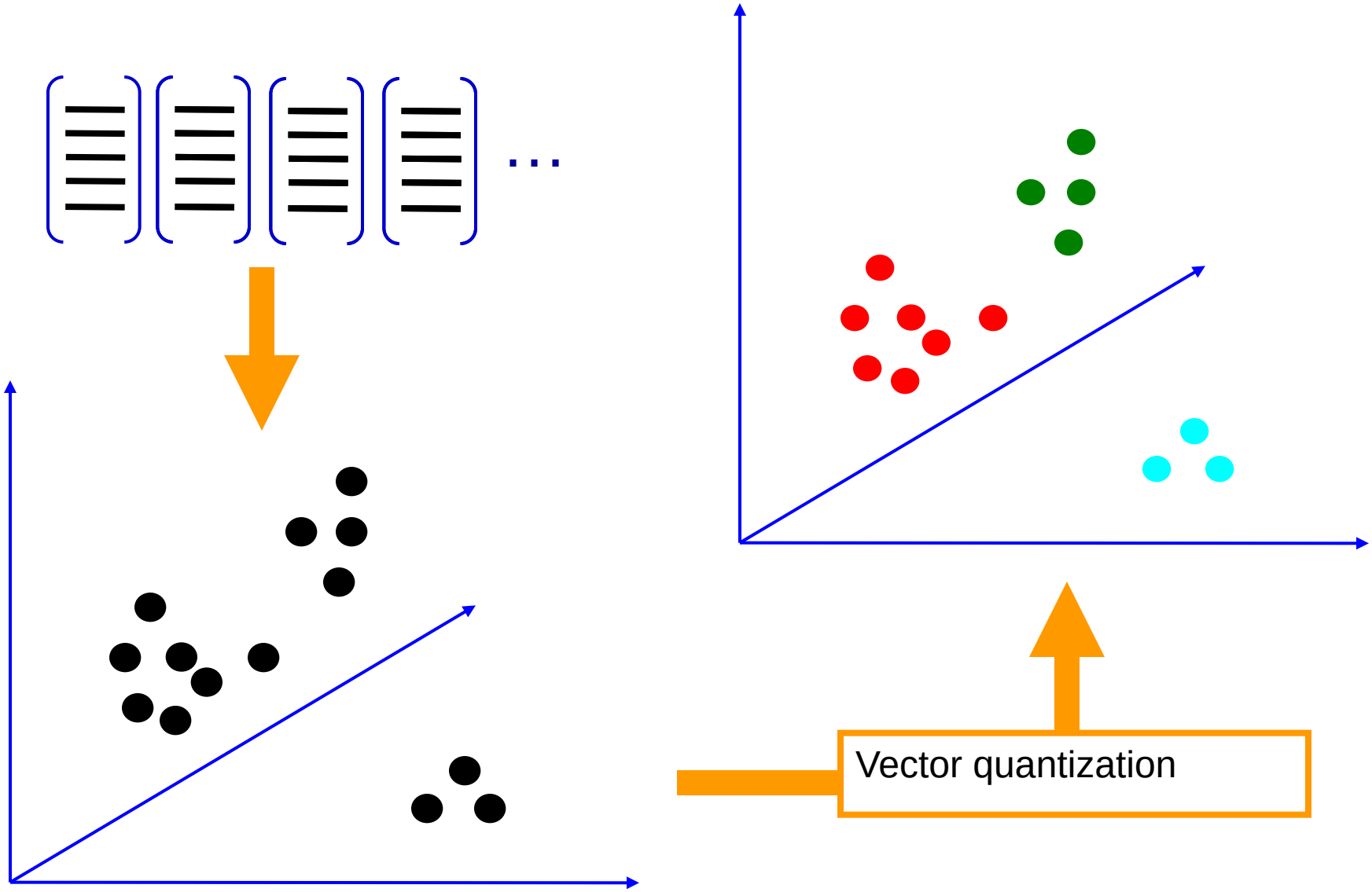
Patch Features



dictionary formation



Clustering (usually k-means)



Slide credit: Josef Sivic

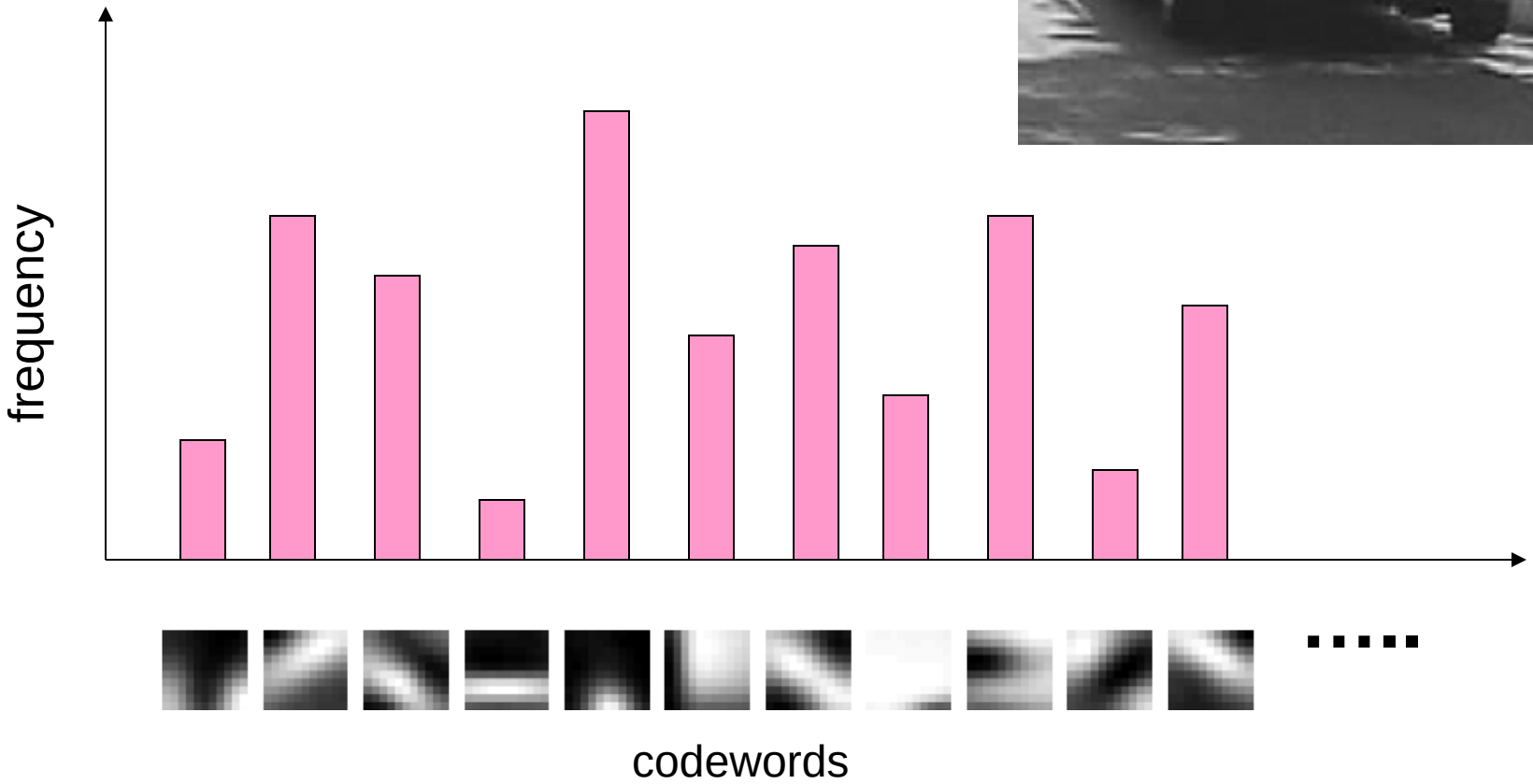
Feature Representation

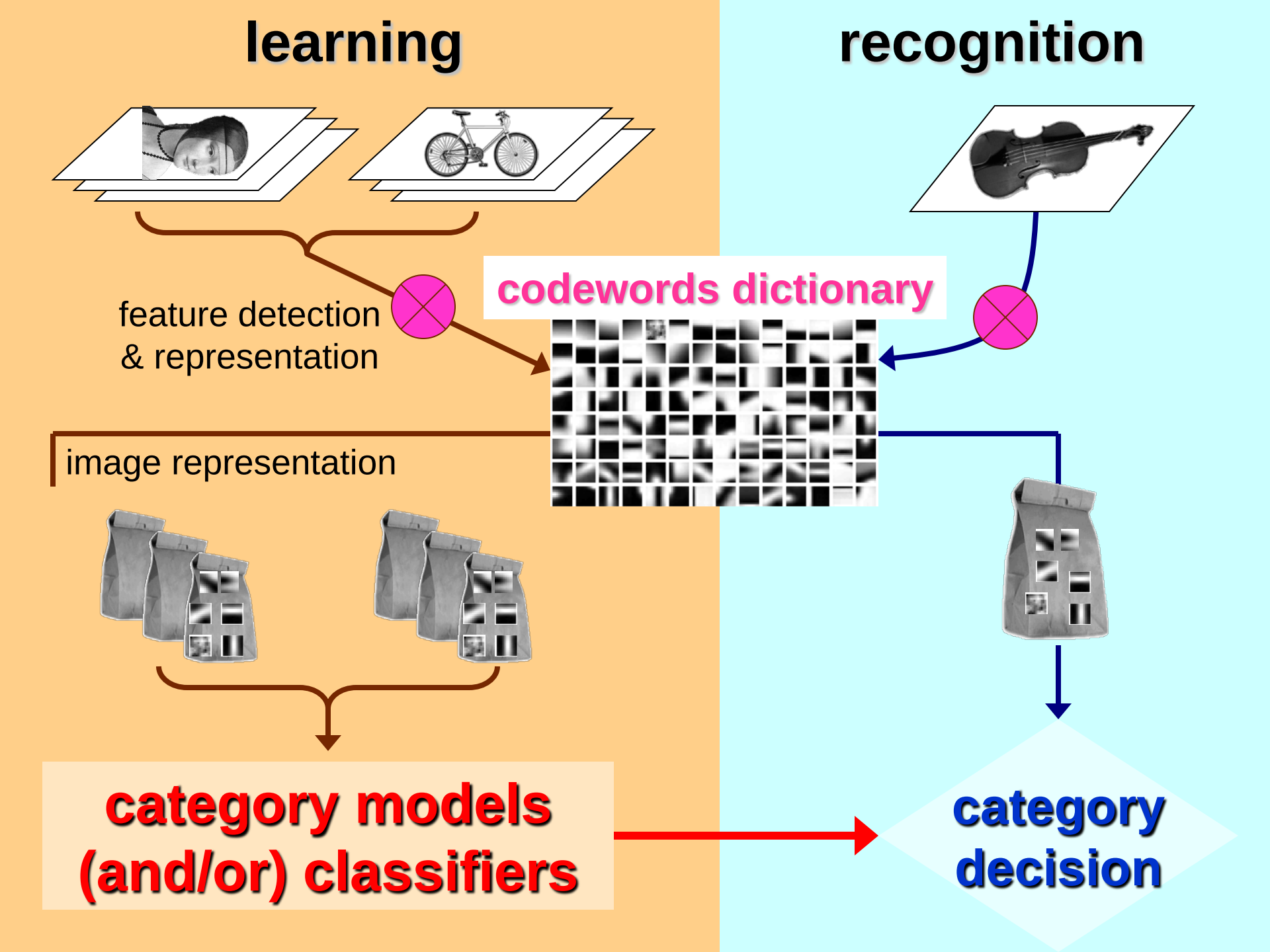
Visual words, aka textons, aka keypoints:

K-means clustered pieces of the image

- Various Representations:
  - Filter bank responses (textons)
  - Image Patches
  - SIFT descriptors
- Either image-specific or “universal” dictionary

Image representation




learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision

Scene Classification (Renninger & Malik)

beach

mountain

forest

city

street

farm

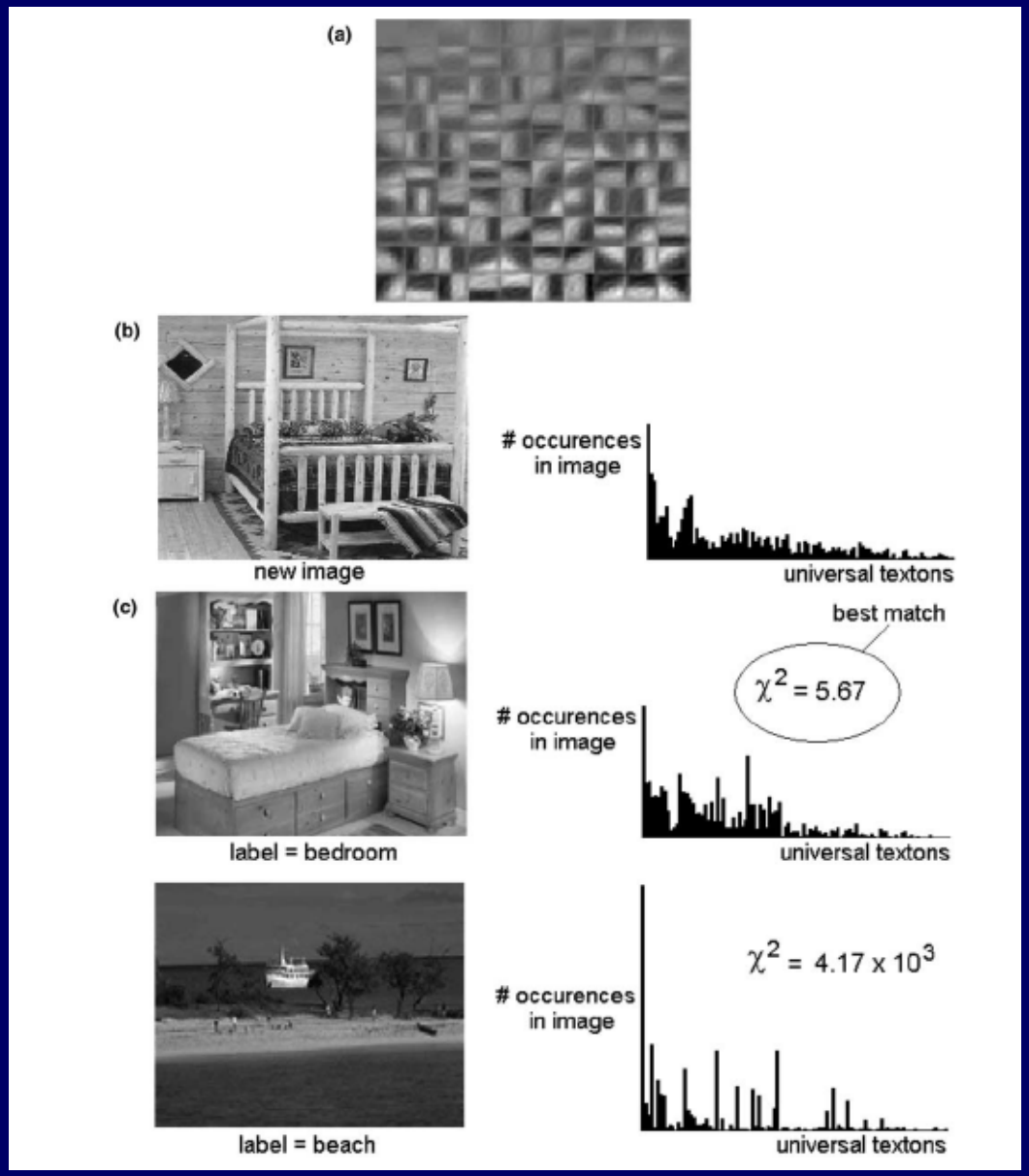
kitchen

livingroom

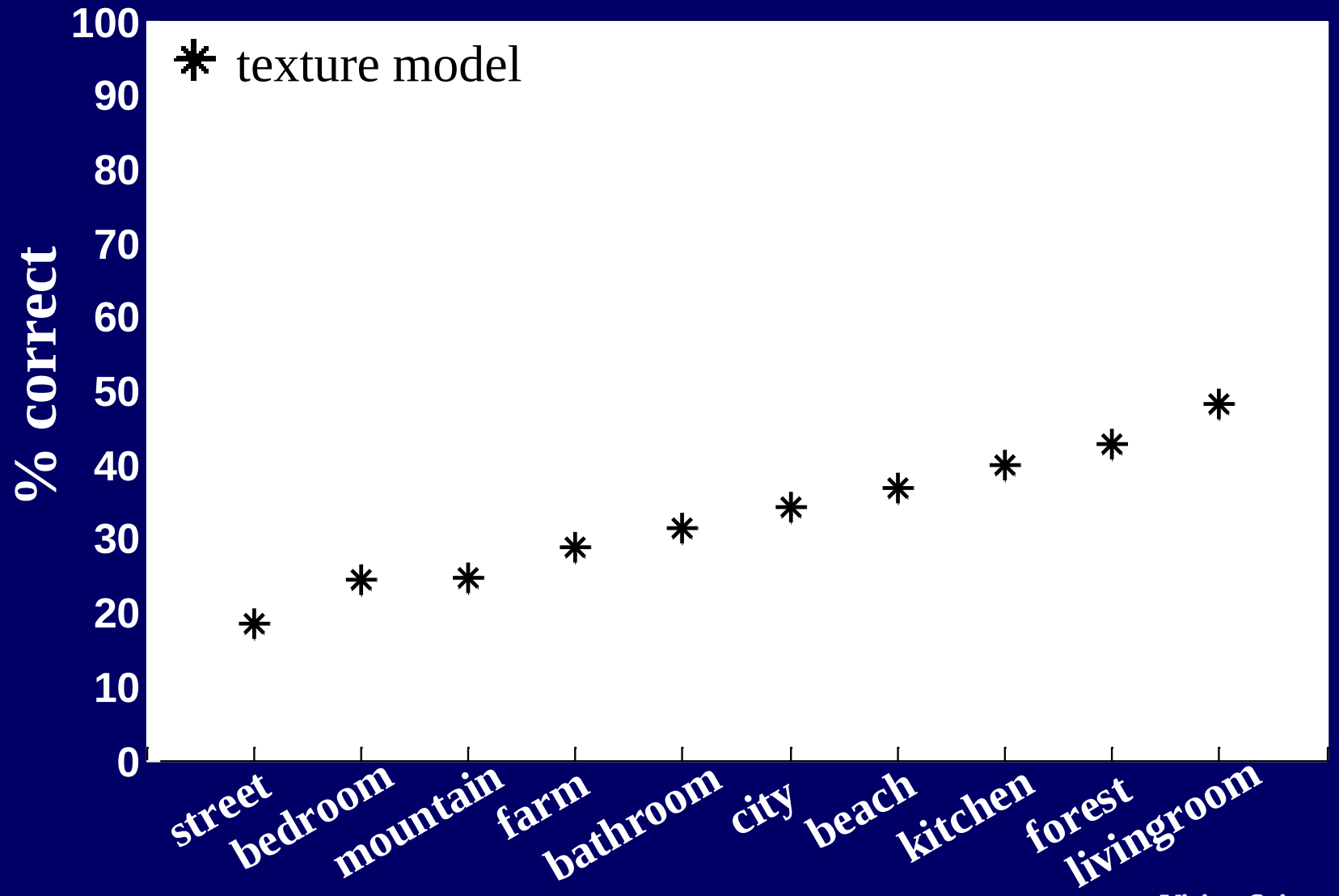
bedroom

bathroom

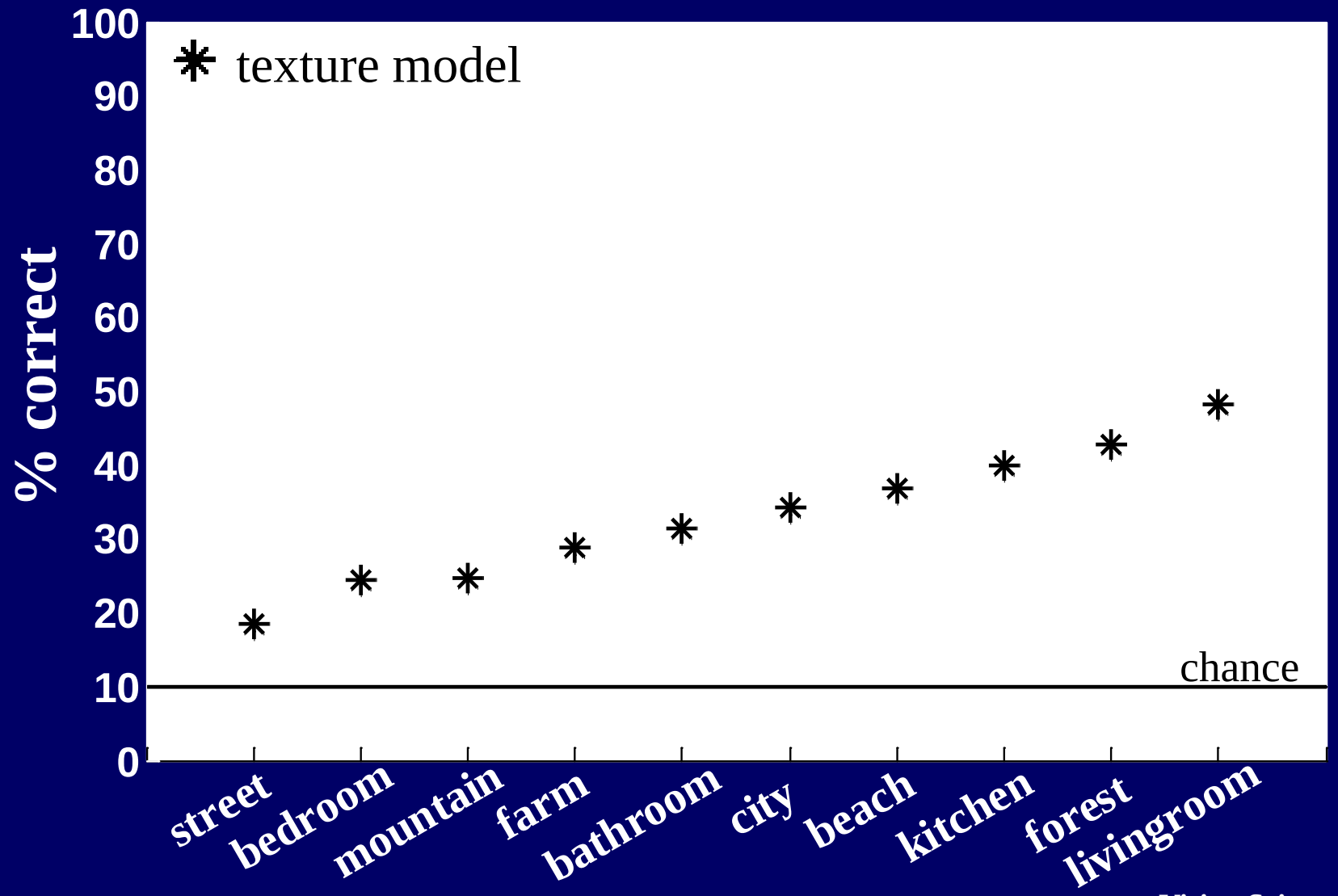
Texton Histogram Matching



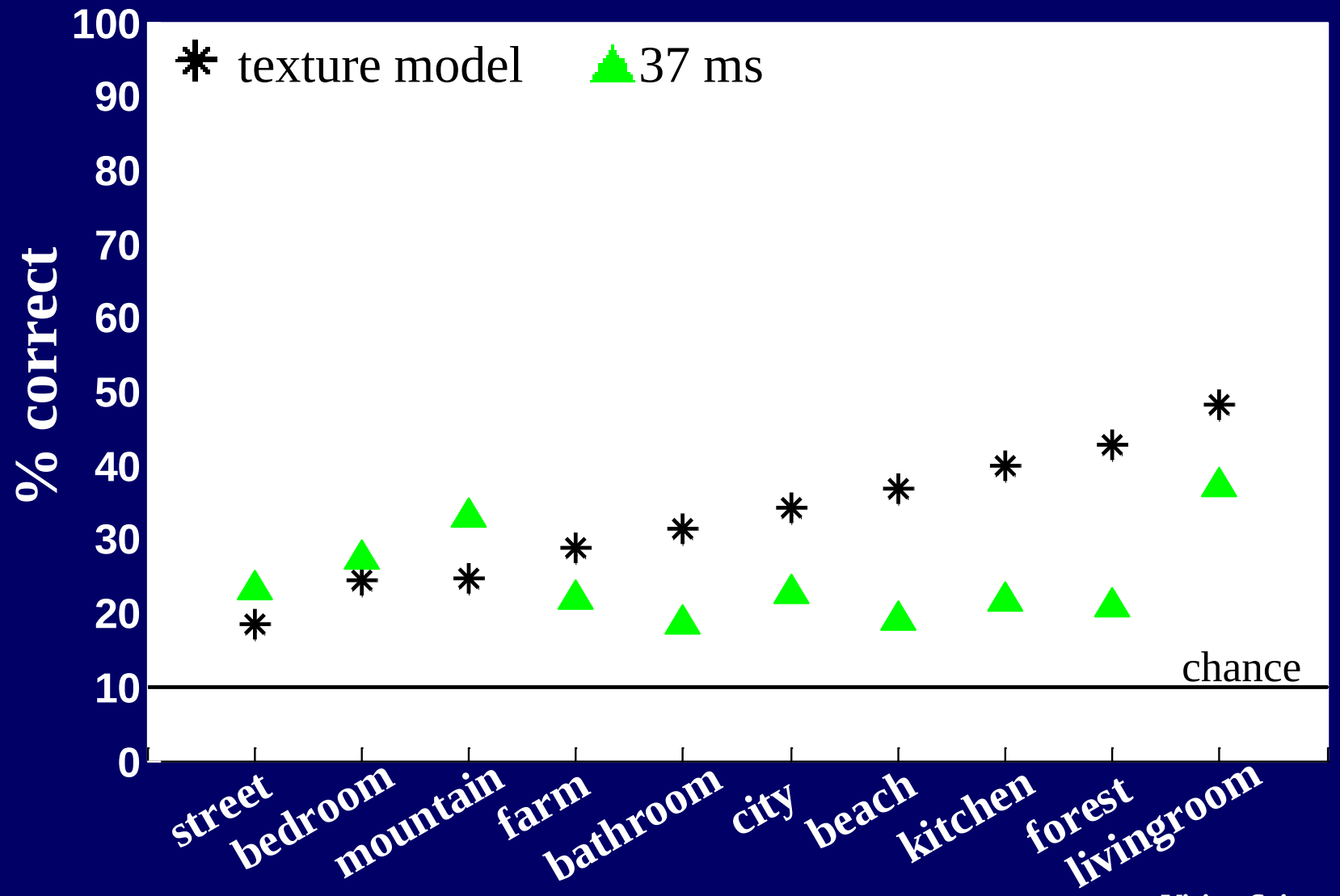
Discrimination of Basic Categories



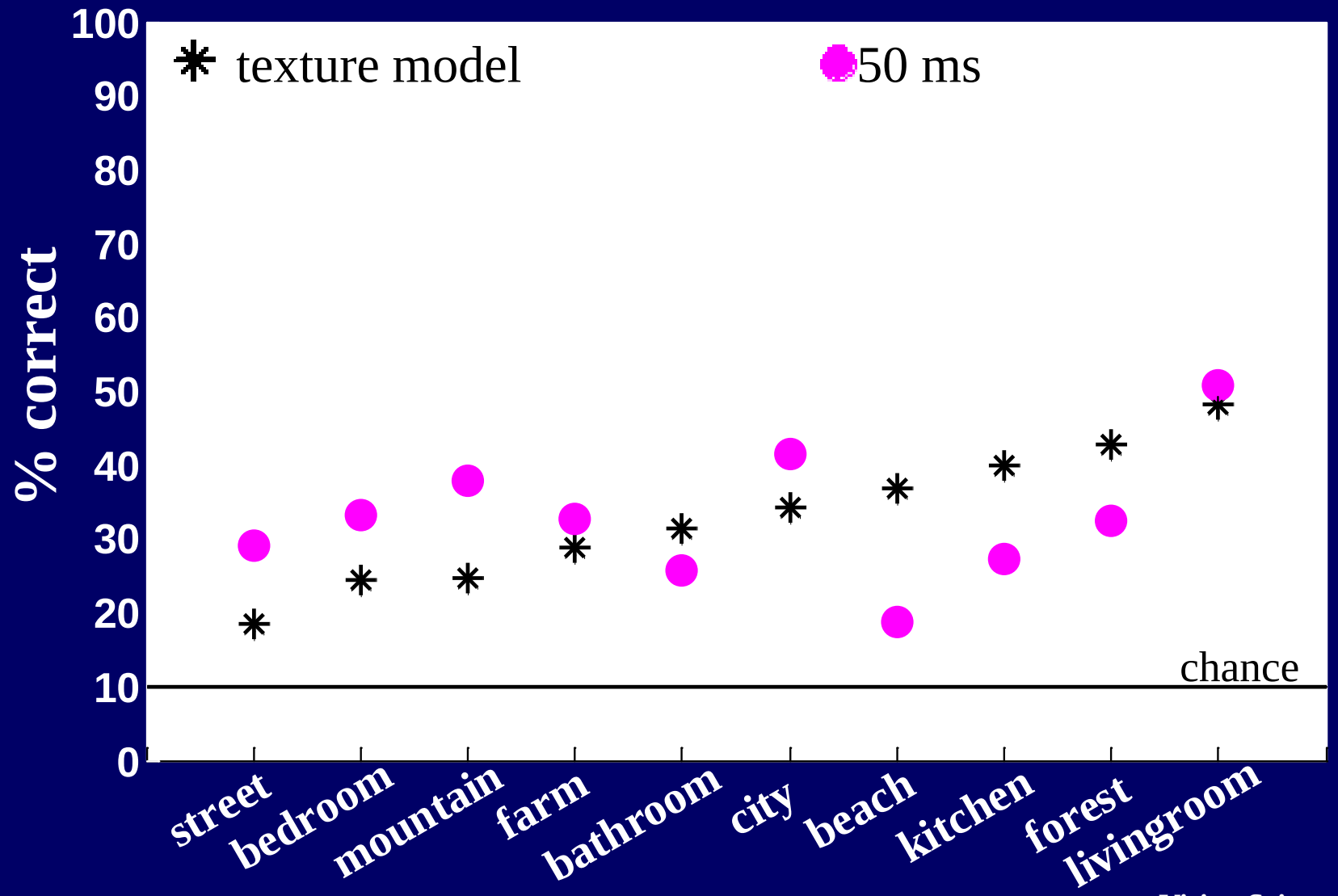
Discrimination of Basic Categories



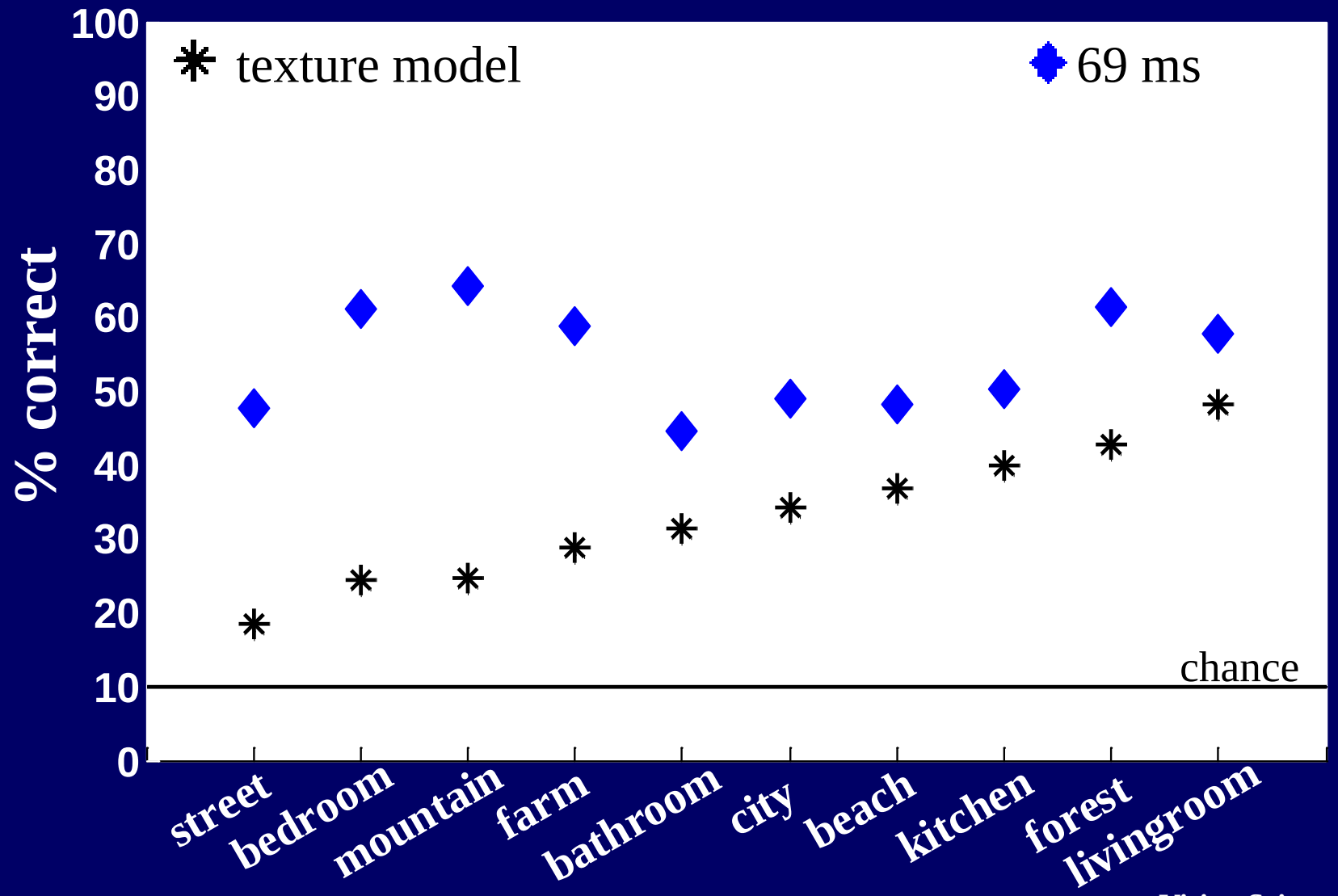
Discrimination of Basic Categories



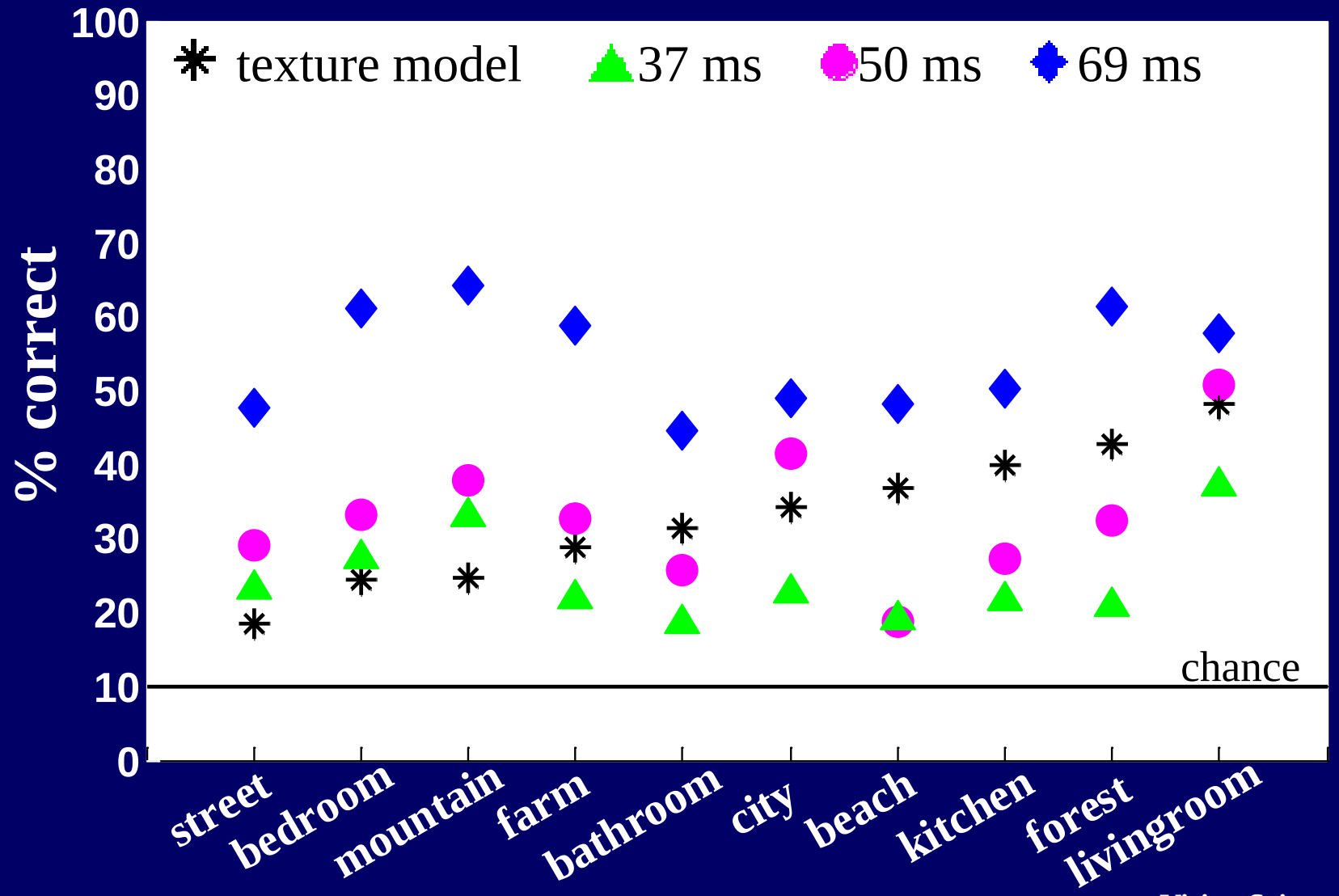
Discrimination of Basic Categories



Discrimination of Basic Categories



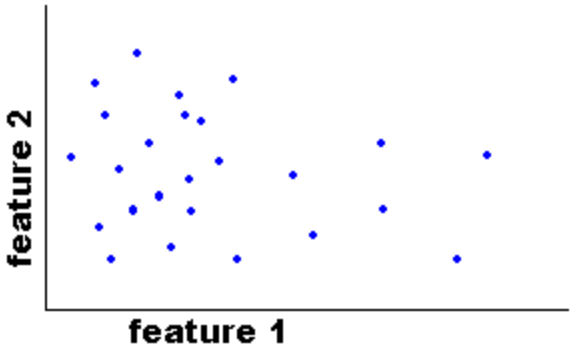
Discrimination of Basic Categories

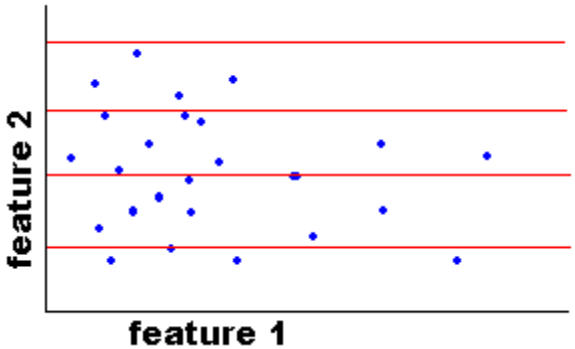


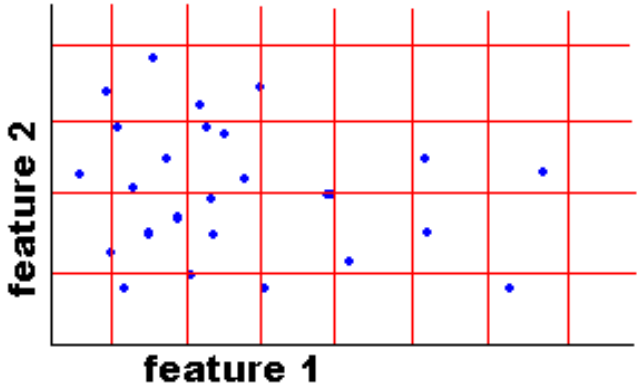
Scene Recognition using Texture

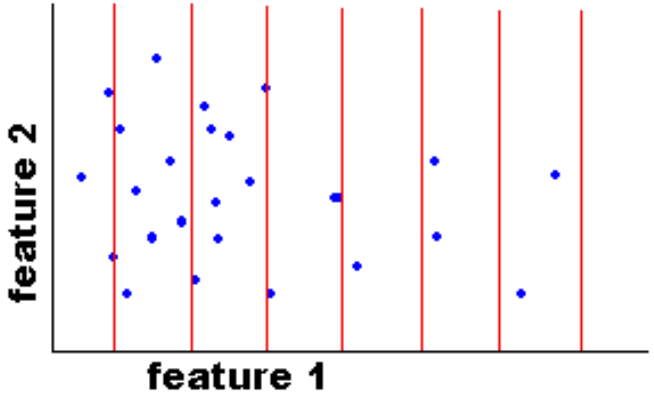
Image Representations: Histograms

Images from Dave Kauchak









Joint histogram

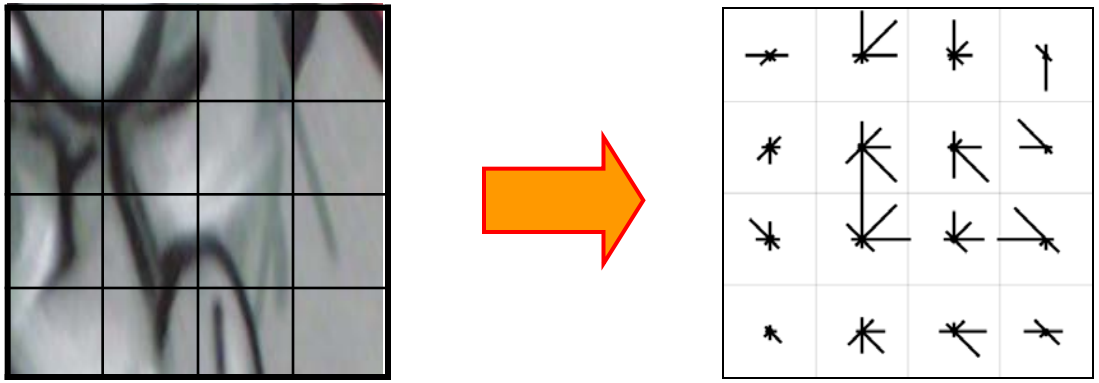
- Requires lots of data
- Loss of resolution to avoid empty bins

Marginal histogram

- Requires independent features
- More data/bin than joint histogram

Feature descriptors: SIFT

- Descriptor computation:
  - Divide patch into 4x4 sub-patches
  - Compute histogram of gradient orientations (8 reference angles) inside each sub-patch
  - Resulting descriptor: 4x4x8 = 128 dimensions



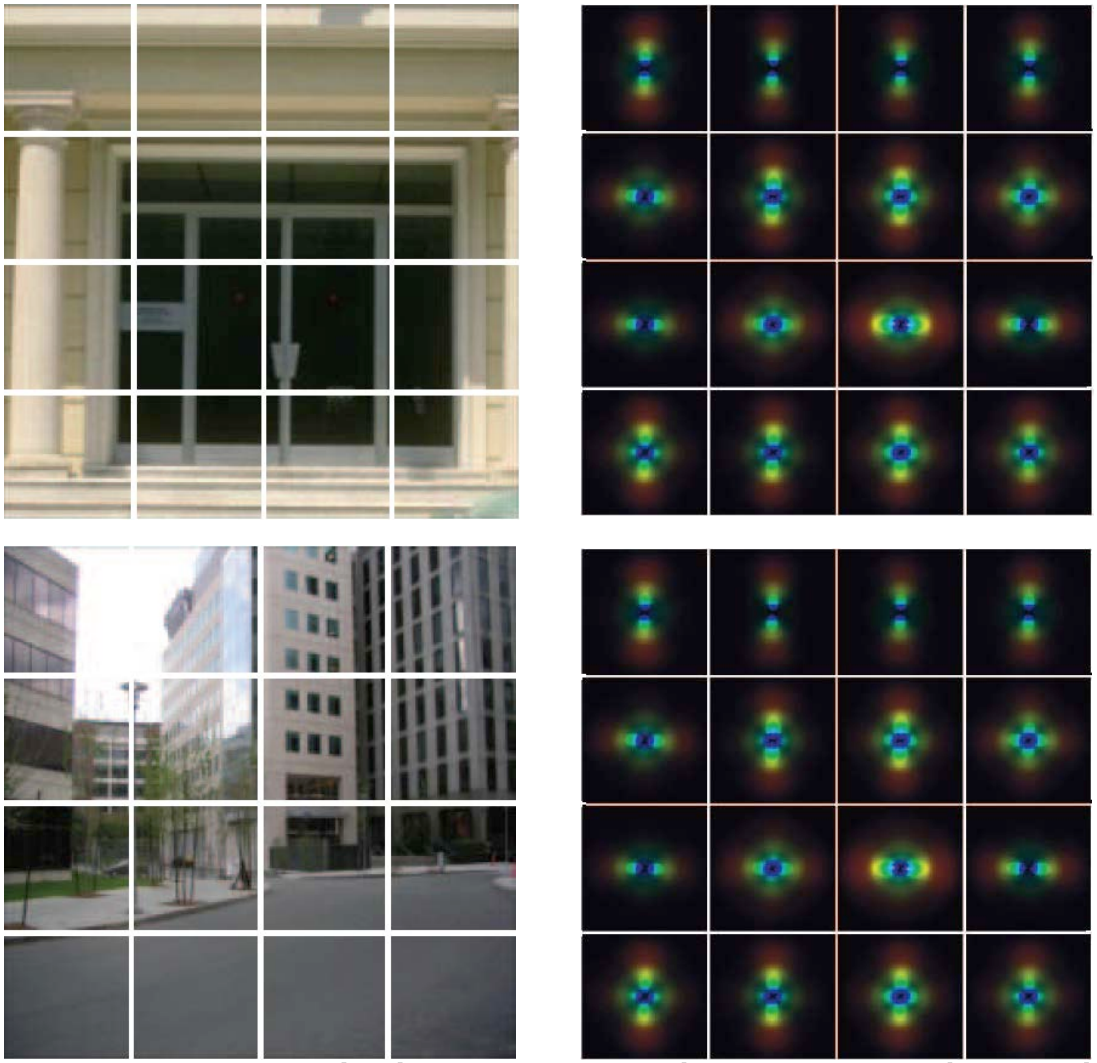
David G. Lowe. "Distinctive image features from scale-invariant keypoints." *IJCV* 60 (2), pp. 91-110, 2004.

SIFT captures some invariances

- Small deformations can affect the matching score a lot

Gist descriptor

Oliva and Torralba, 2001



- Apply oriented Gabor filters over different scales
- Average filter energy in each bin

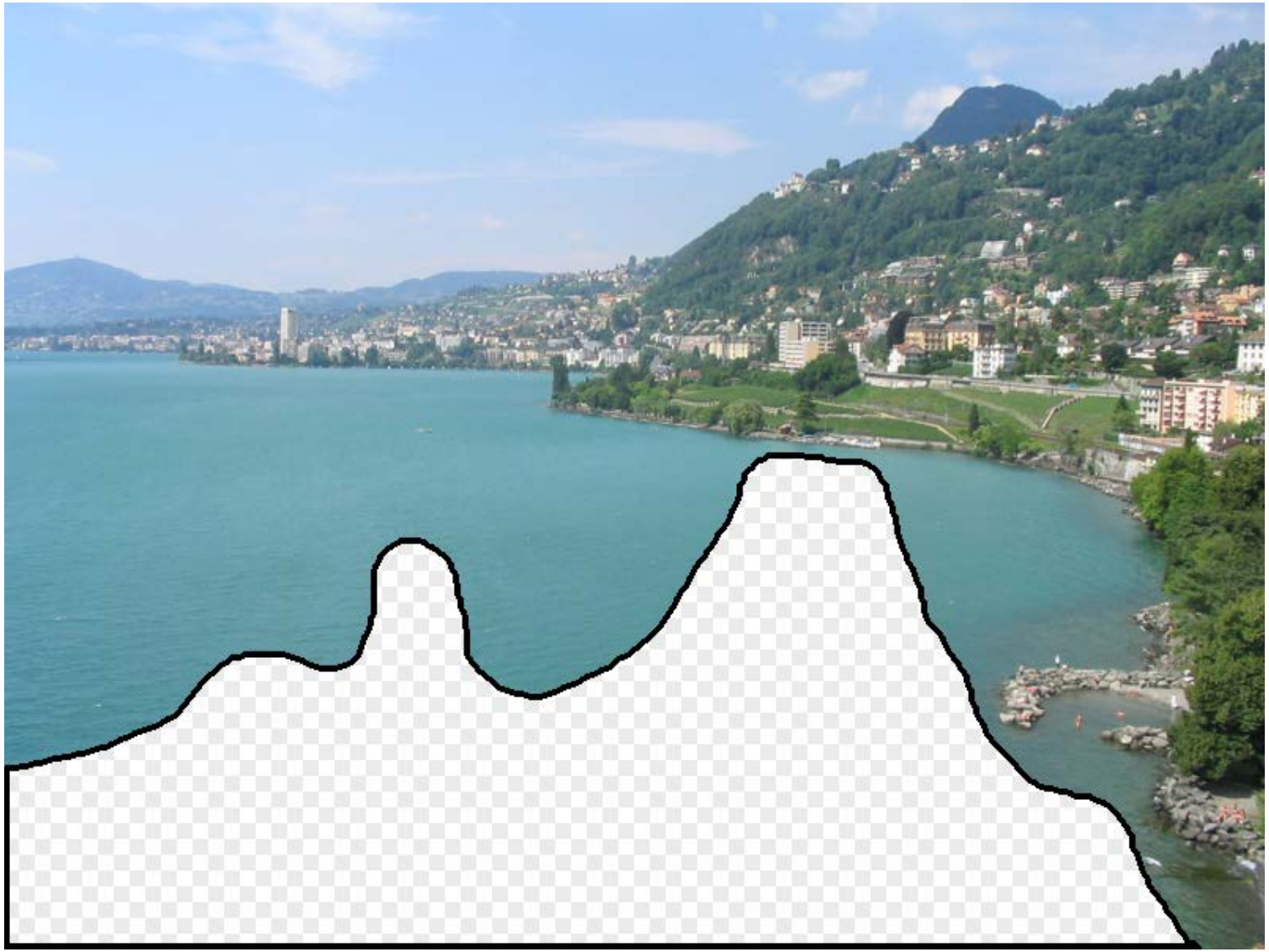
8 orientations
4 scales
x 16 bins
512 dimensions

Similar to SIFT (Lowe 1999) applied to the entire image

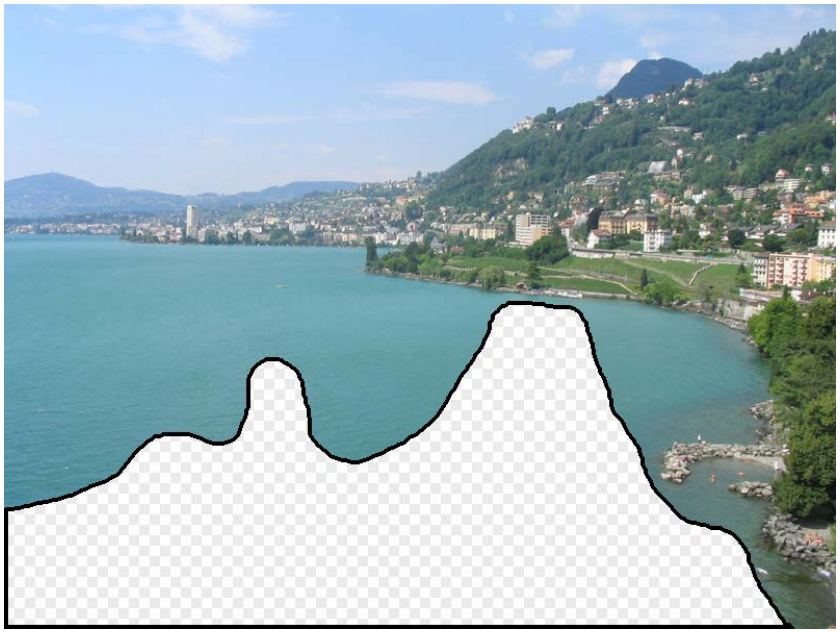
M. Gorkani, R. Picard, ICPR 1994; Walker, Malik. Vision Research 2004; Vogel et al. 2004; Fei-Fei and Perona, CVPR 2005; S. Lazebnik, et al, CVPR 2006; …

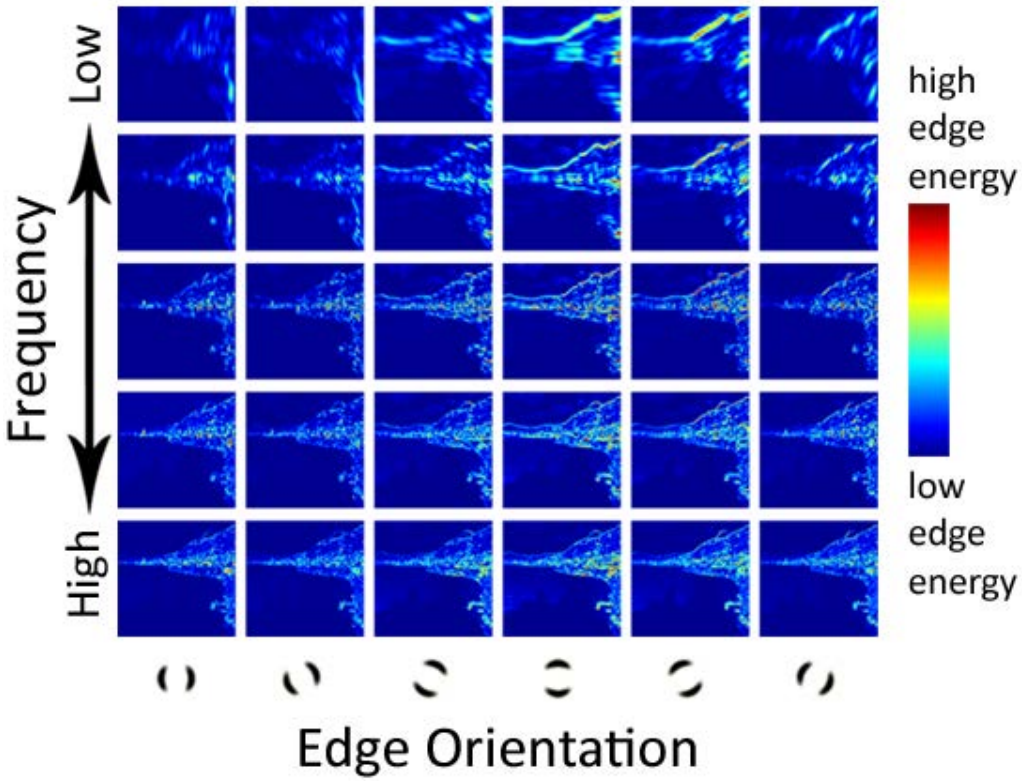
Playing with GIST

Scene Completion (Hays et al, SIGGRAPH 2007)

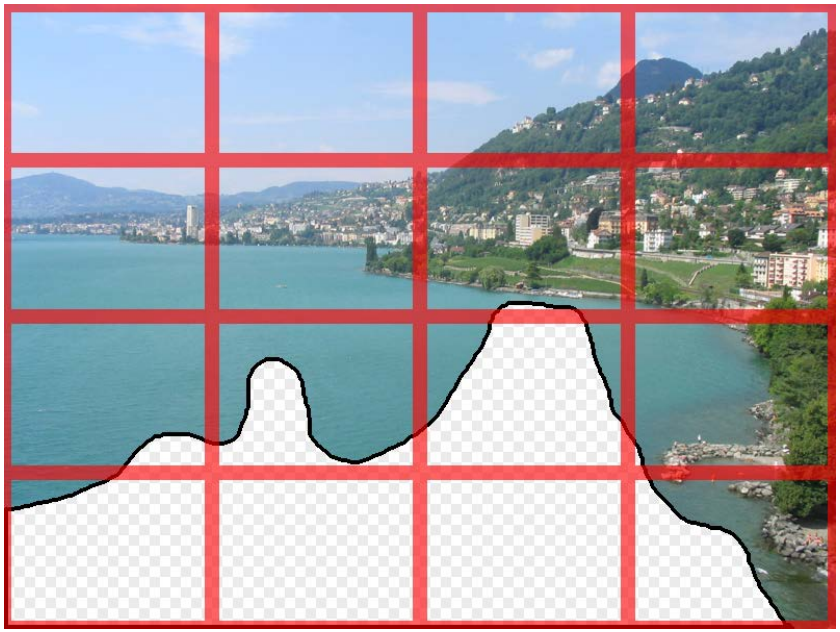


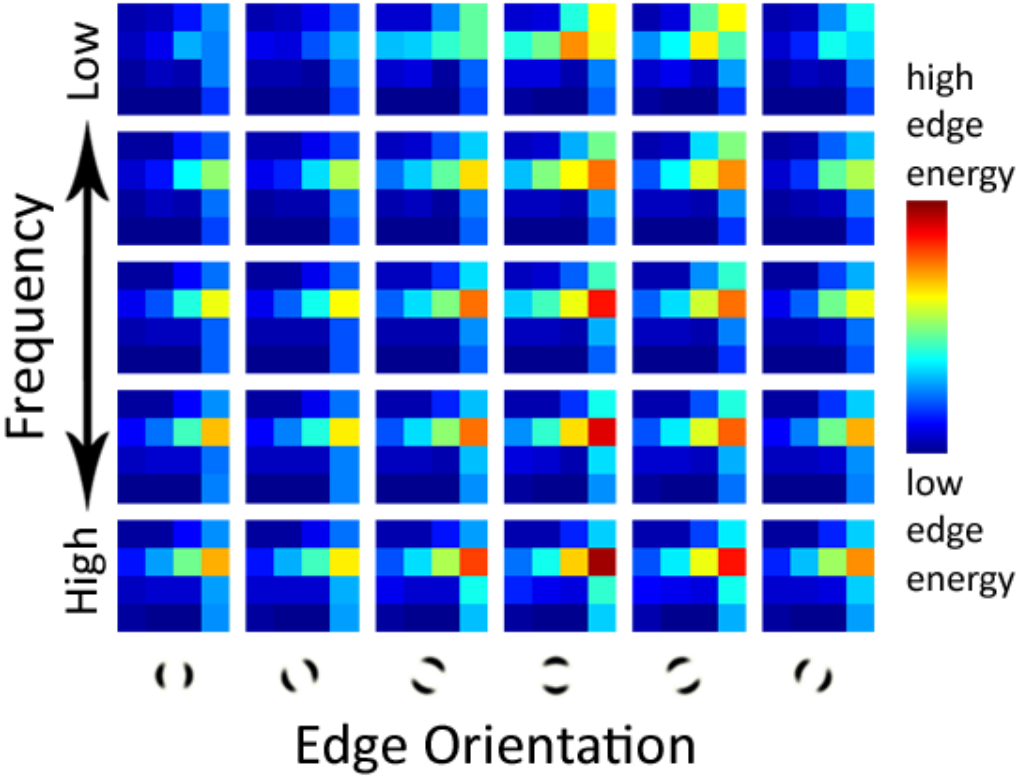
Scene Descriptor





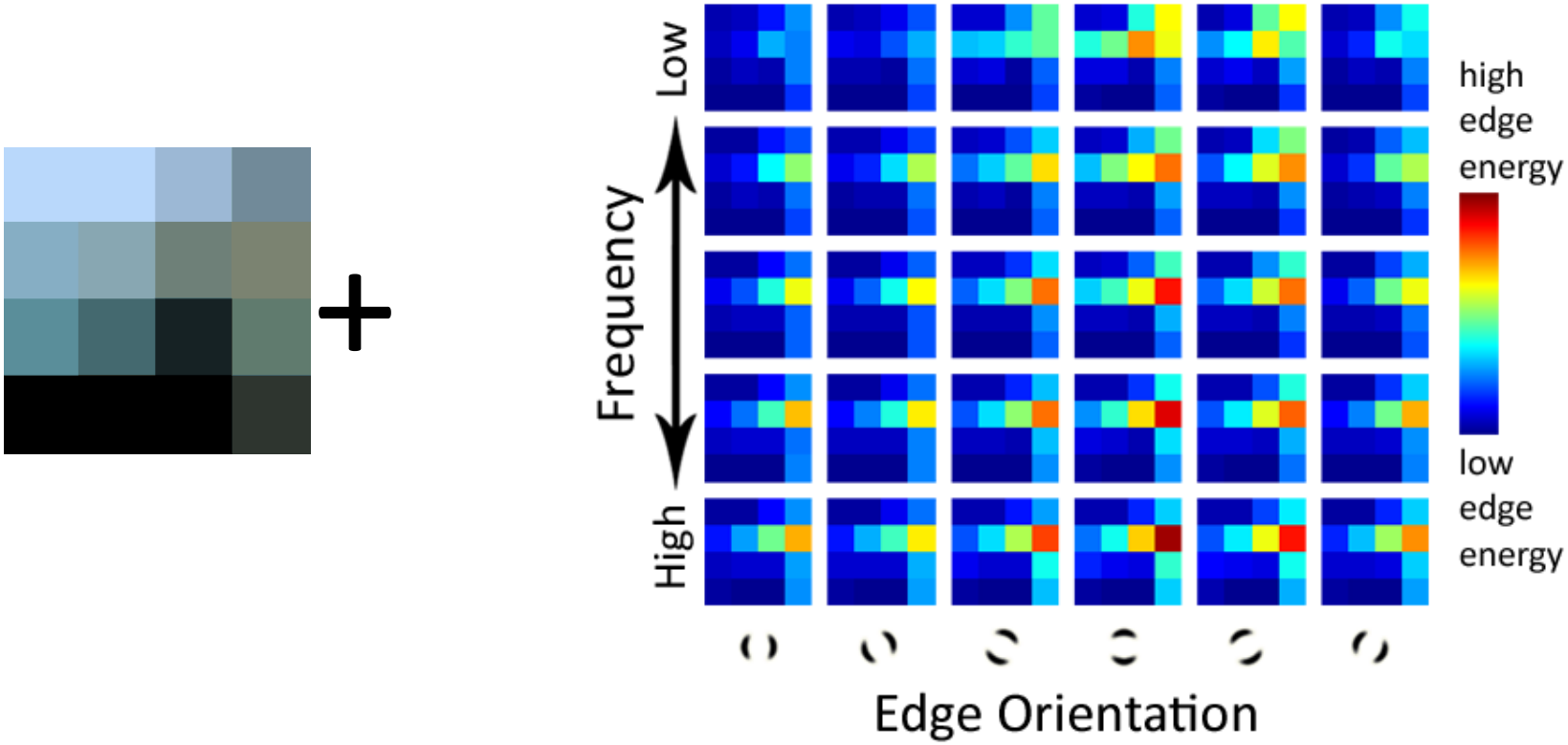
Scene Descriptor



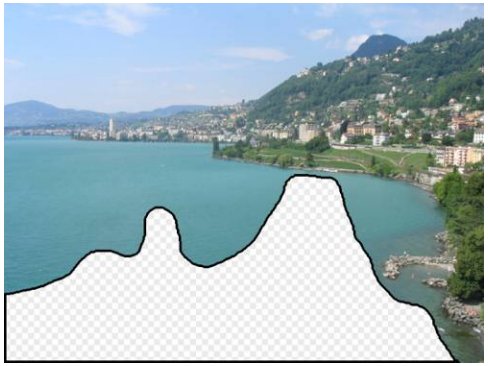


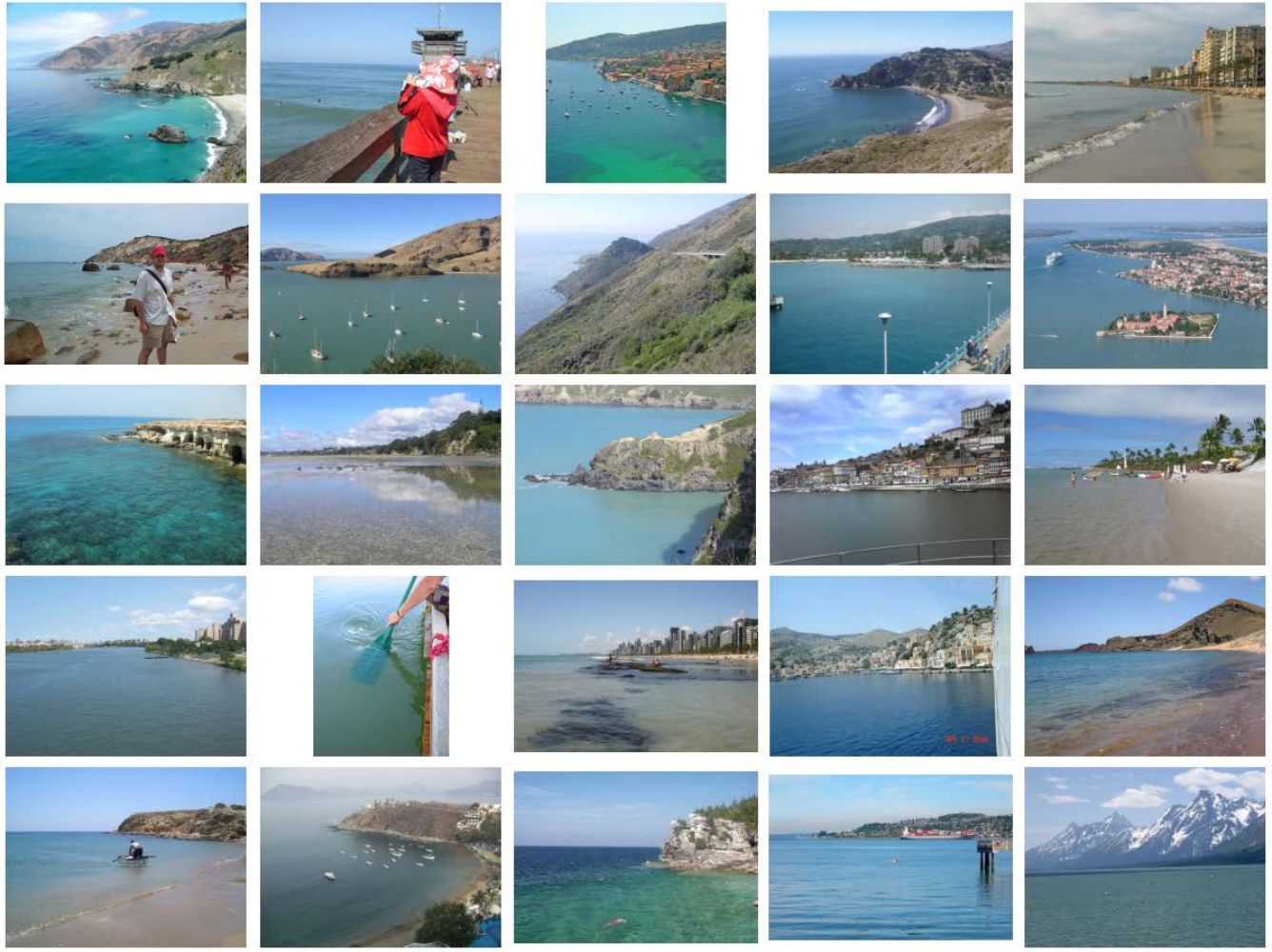
Scene Gist Descriptor
(Oliva and Torralba 2001)

Scene Descriptor



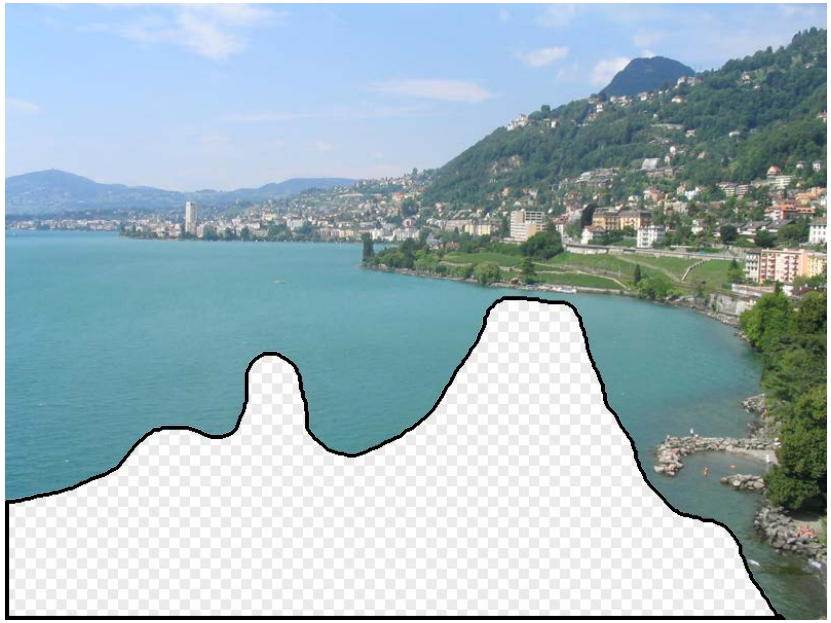
2 Million Flickr Images





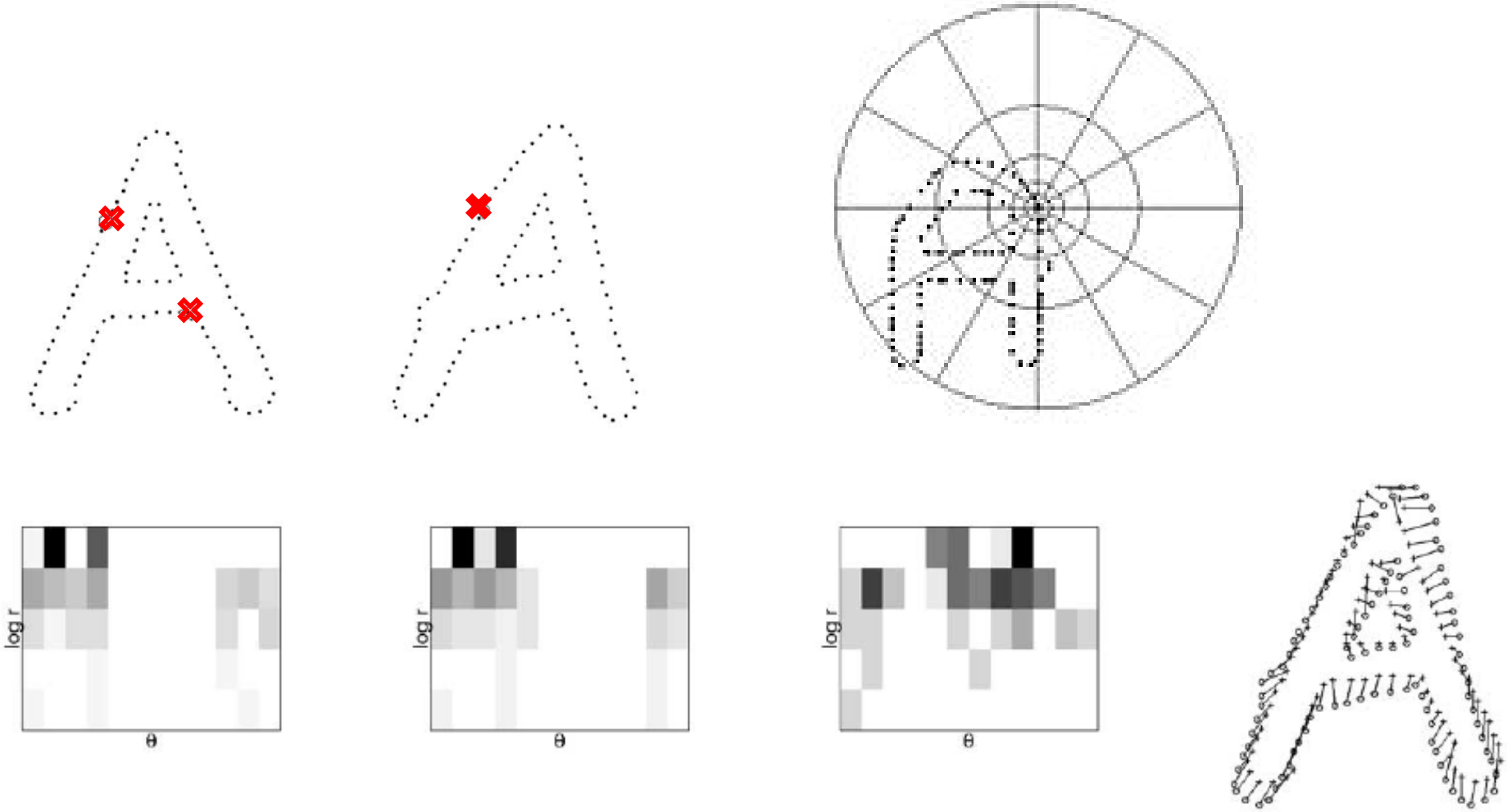
… 200 total

Context Matching



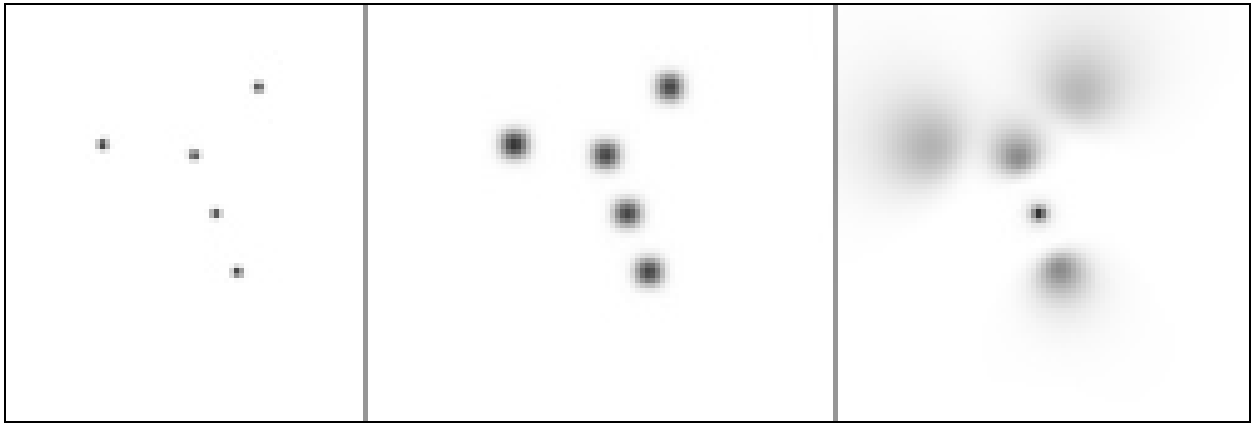
Graph cut + Poisson blending

Shape context

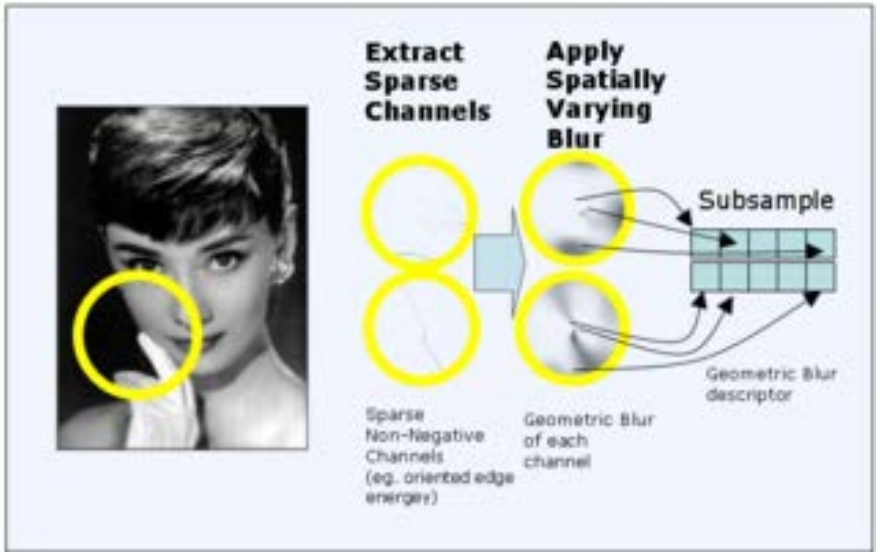


Belongie and J. Malik. "Shape Matching and Object Recognition Using Shape Contexts", IEEE Trans. PAMI April 2002.

Geometric blur

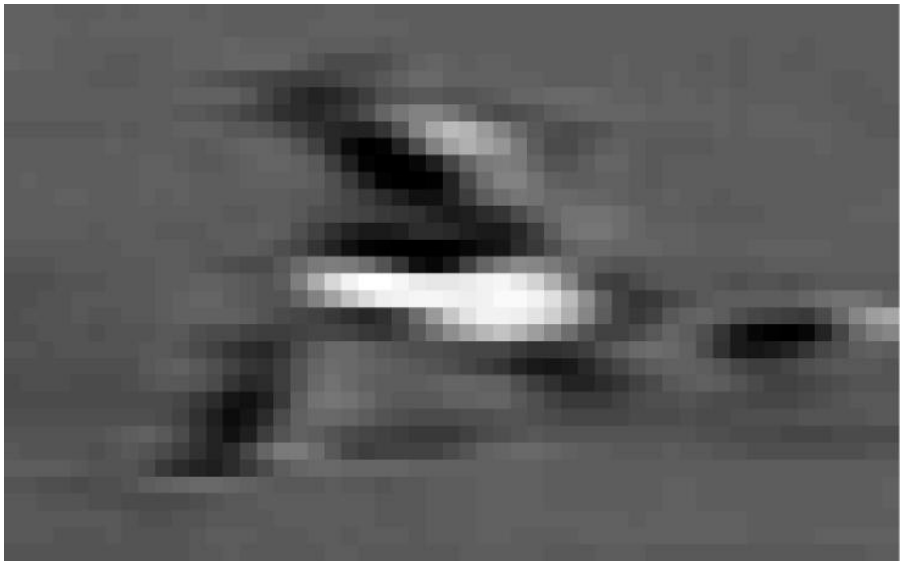


Discrete points Gaussian blur Variable blur

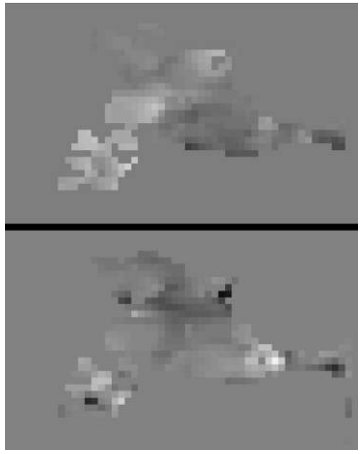


Alexander C. Berg, Jitendra Malik. Template Matching. Computer Vision and Pattern Recognition (CVPR) 2001

(Geometric) Blur + half-wave rectification

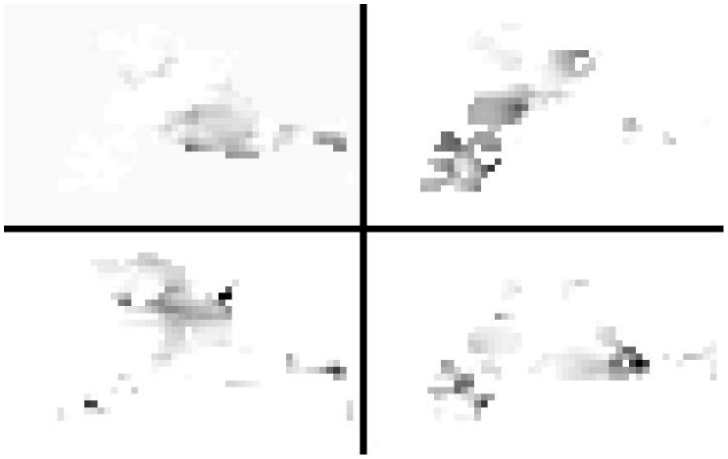


gradients



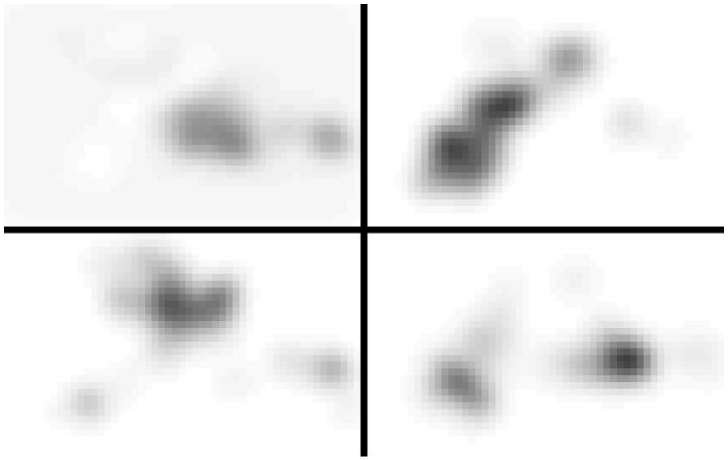
F_x, F_y

Half-wave rect.



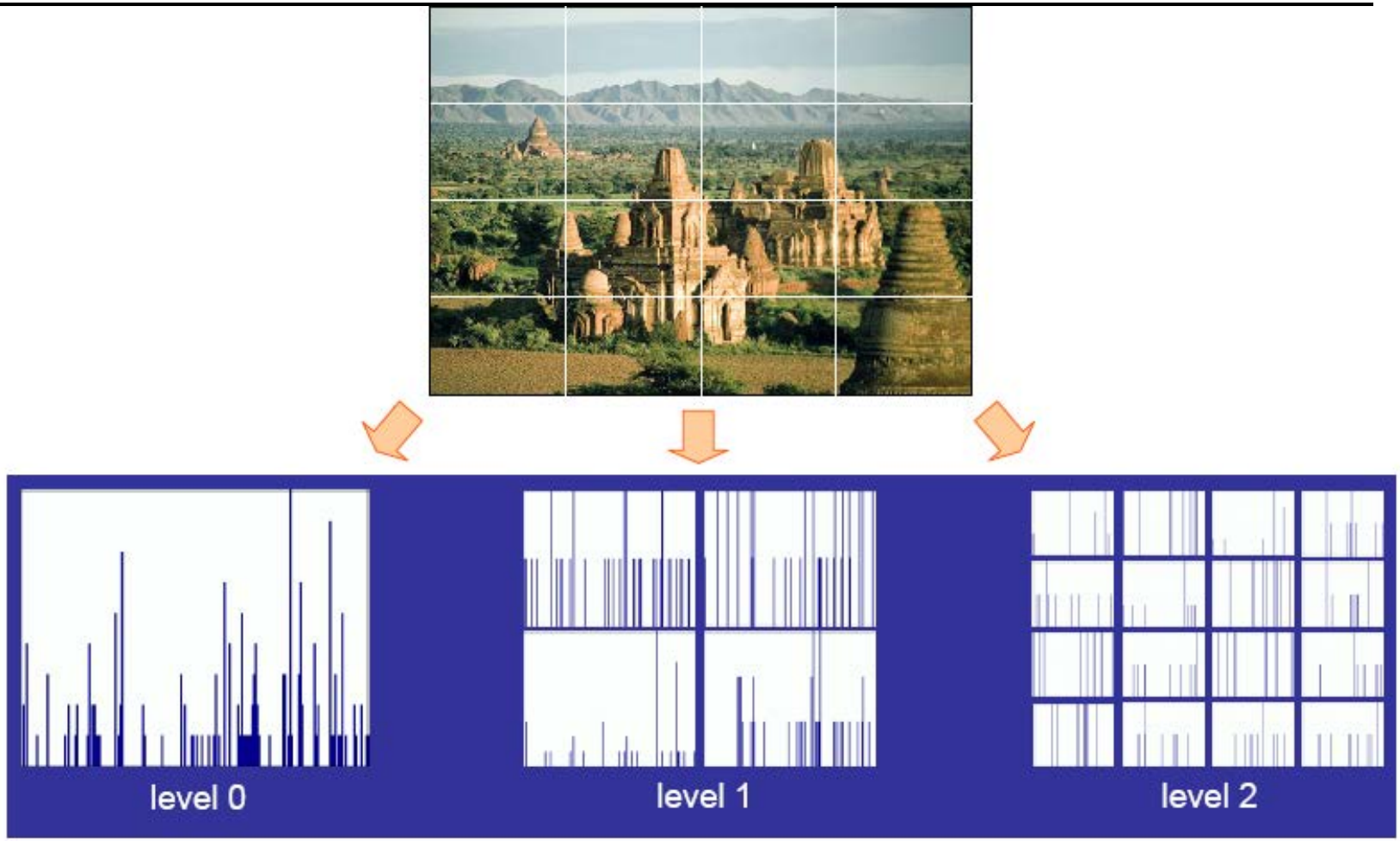
$F_x^-, F_x^+, F_y^-, F_y^+$

blur



blurred $F_x^-, F_x^+, F_y^-, F_y^+$

Spatial Pyramid Matching



Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories

S. Lazebnik, C. Schmid, and J. Ponce, *CVPR 2006*

Story so far…

1. Compute some low-level features
2. Pool/histogram/blur feature responses
3. Use as visual representation

But how do we decide on the features?

Why these filters?

Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images.

Bruno A. Olshausen[1] and David J. Field

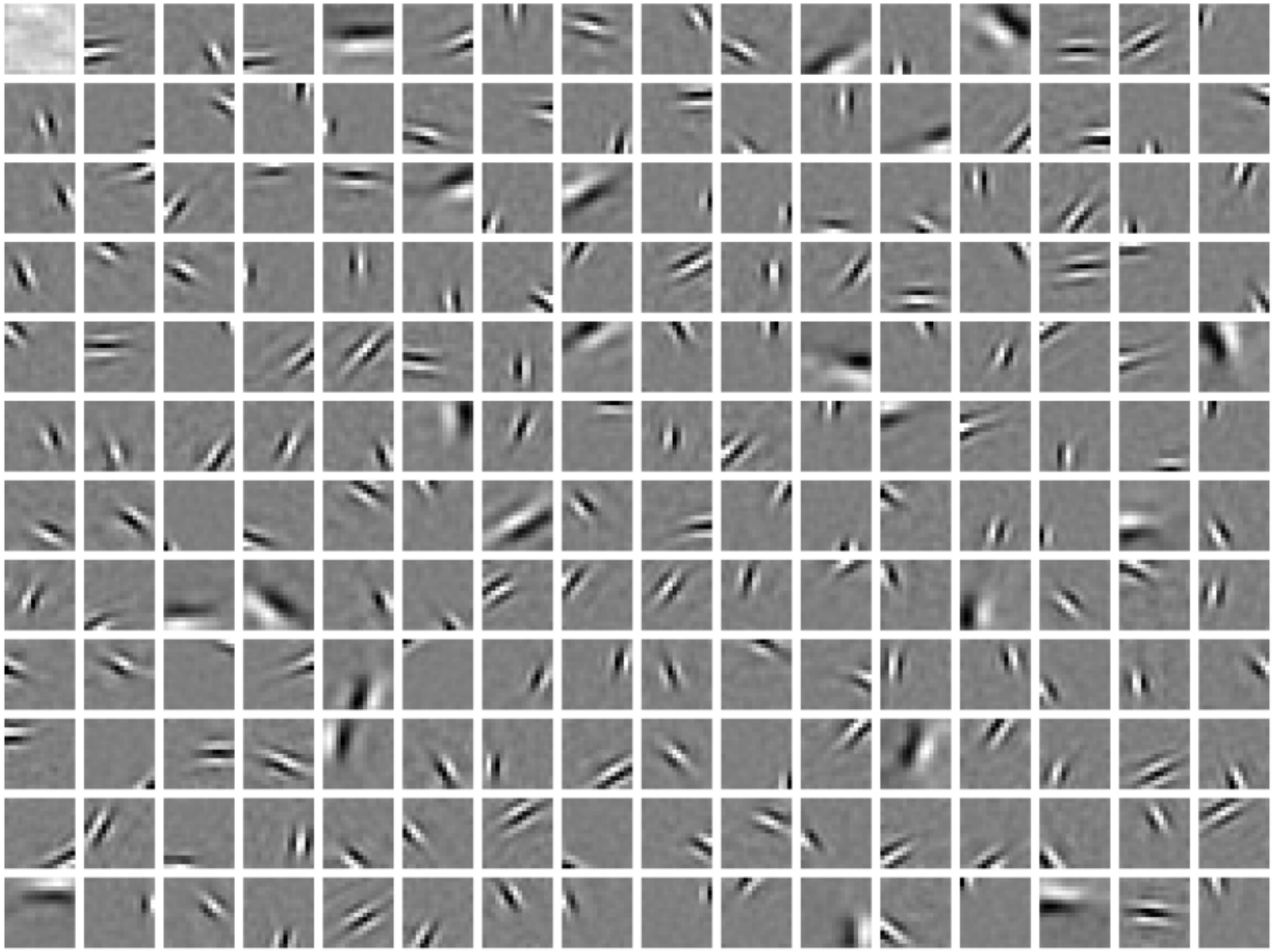
$$E = -[\text{preserve information}] - \lambda[\text{sparseness of } a_i]\,, \tag{2}$$

where λ is a positive constant that determines the importance of the second term relative to the first. The first term measures how well the code describes the image, and we choose this to be the mean square of the error between the actual image and the reconstructed image:

$$[\text{preserve information}] = -\sum_{x,y}\left[I(x,y) - \sum_i a_i\,\phi_i(x,y)\right]^2 . \tag{3}$$

Learned filters

a.



The story so far…

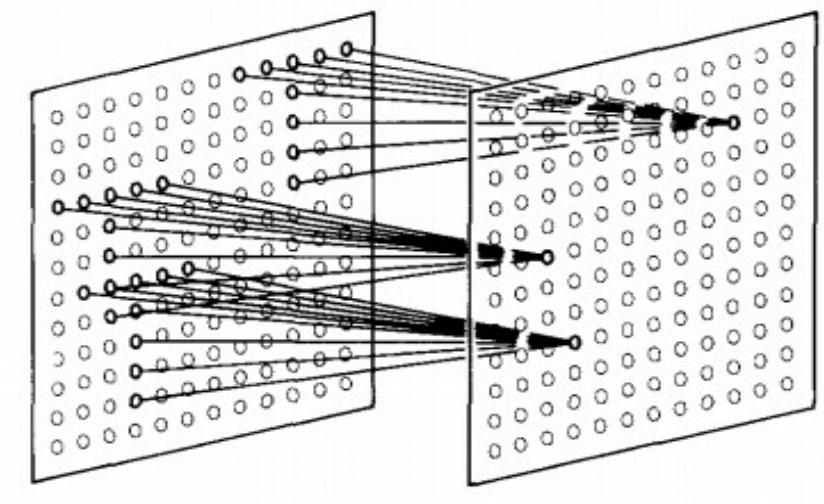
Neurons show increasing specificity higher in the visual pathway

V1 simple and complex cells are orientation-tuned

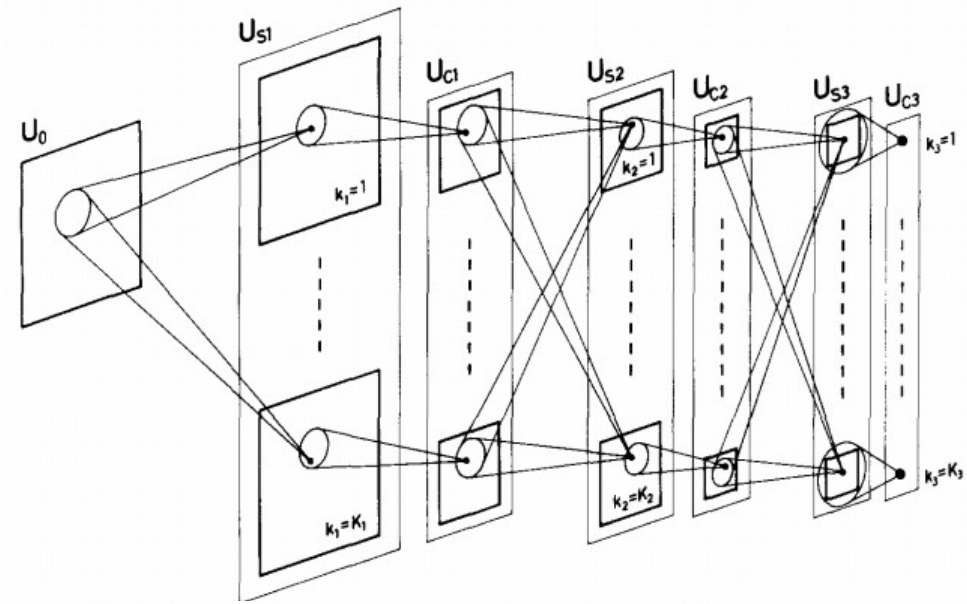
Convolution with a linear kernel followed by simple non-linearities is a good model for computation in retina, LGN and V1, but beyond that we do not have satisfactory computational models

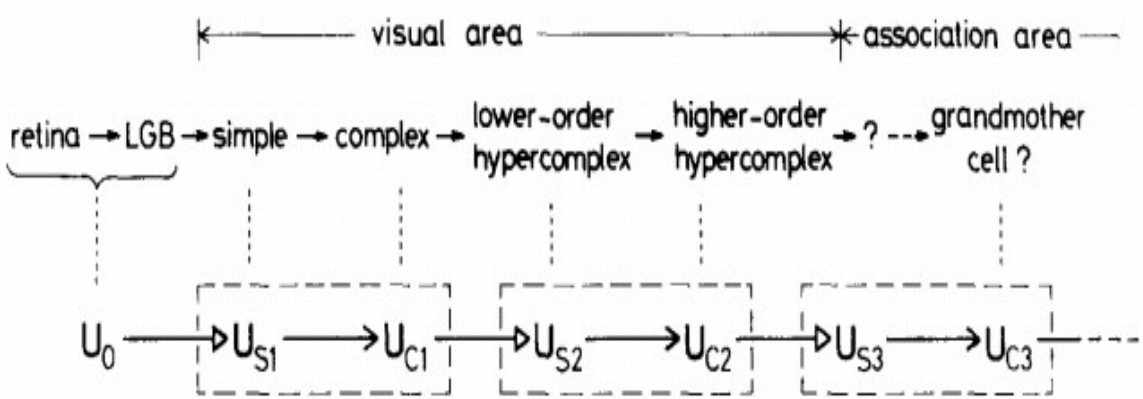
Good designs of visual systems are likely to be hierarchical and “mostly” feedforward

Neurocognitron *[Fukushima 1980]*

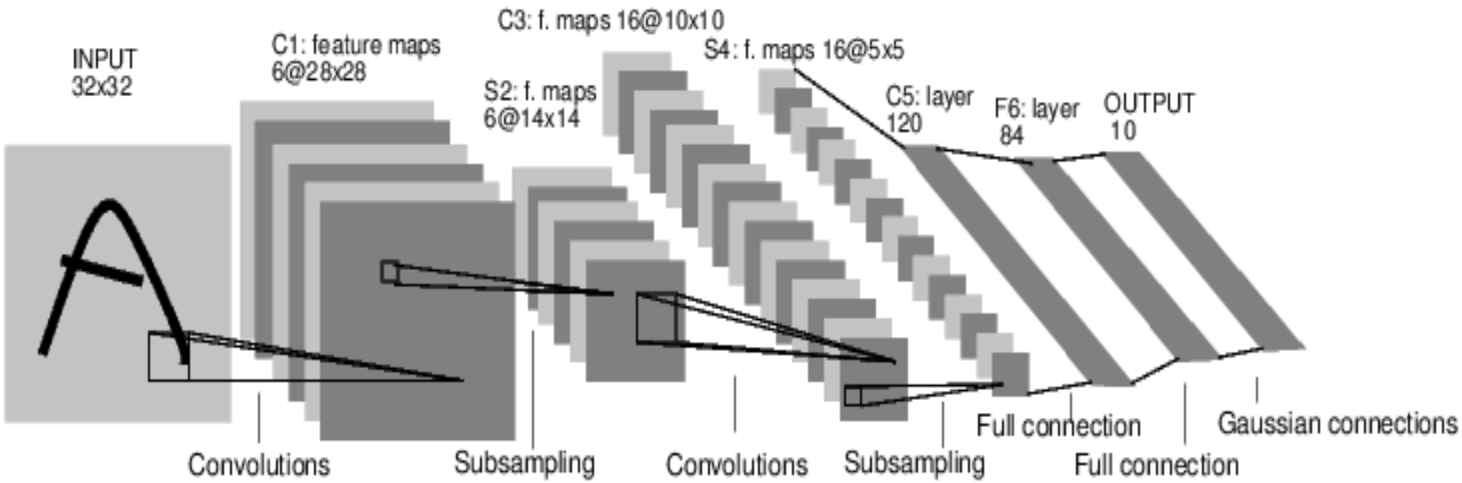


"sandwich" architecture (SCSCSC…)
simple cells: modifiable parameters
complex cells: perform pooling

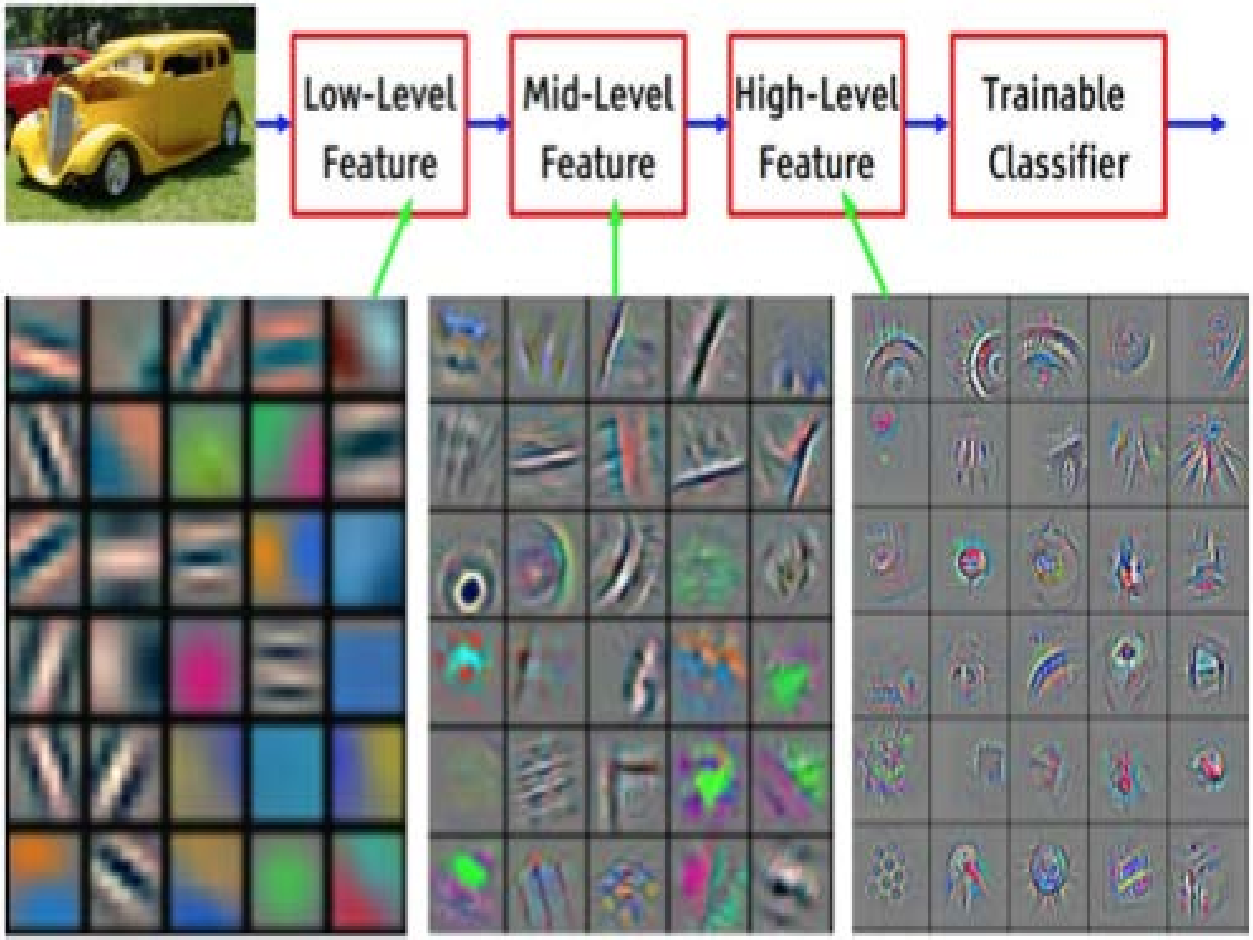




Convolutional Neural Nets (LeCun et al)



[LeNet-5, LeCun 1998]



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

[slide courtesy Yann LeCun]