

Visual Texture (in human and machine)



Somewhere in Cinque Terre, May 2005

CS194: Intro to Computer Vision and Comp. Photo
Alexei Efros, UC Berkeley, Fall 2022

What is Texture?

- Texture depicts spatially repeating patterns
- Many natural phenomena are textures



radishes



rocks



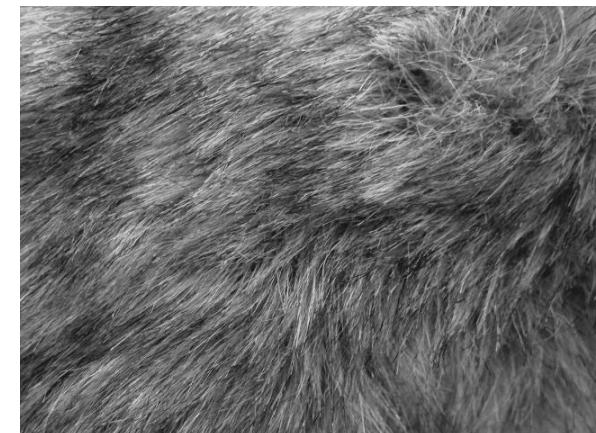
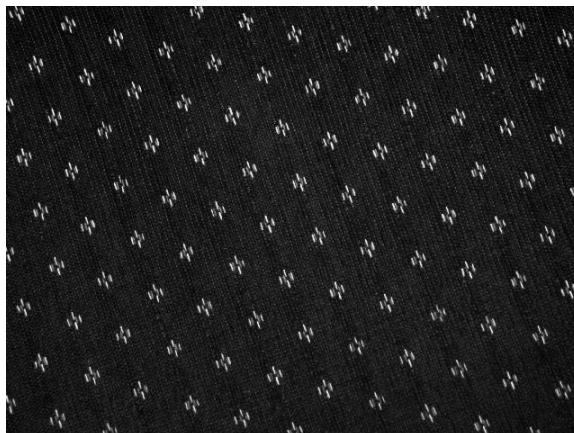
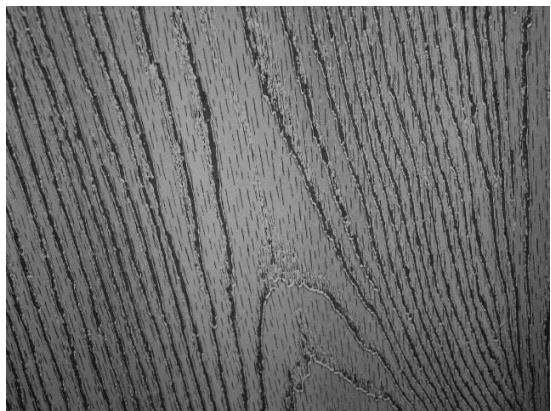
yogurt

Texture as “stuff”

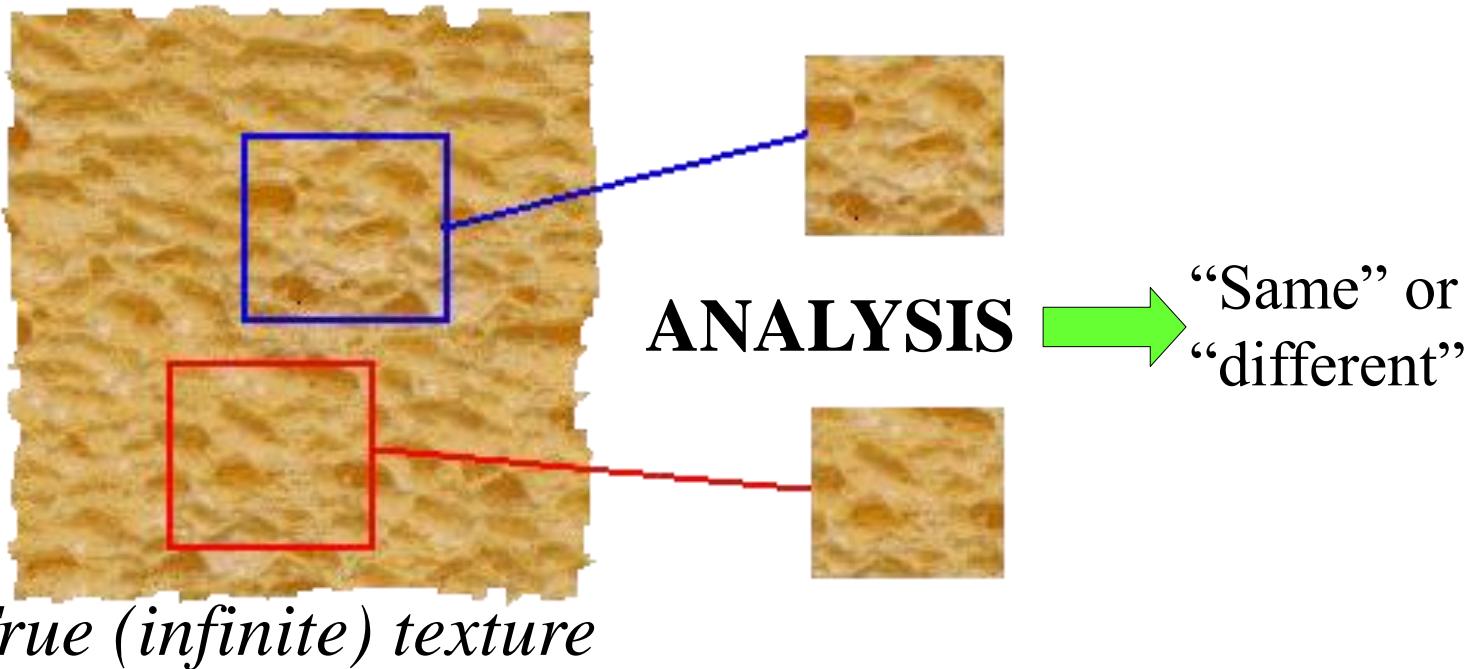


Source: Forsyth

Texture and Material



Texture Analysis

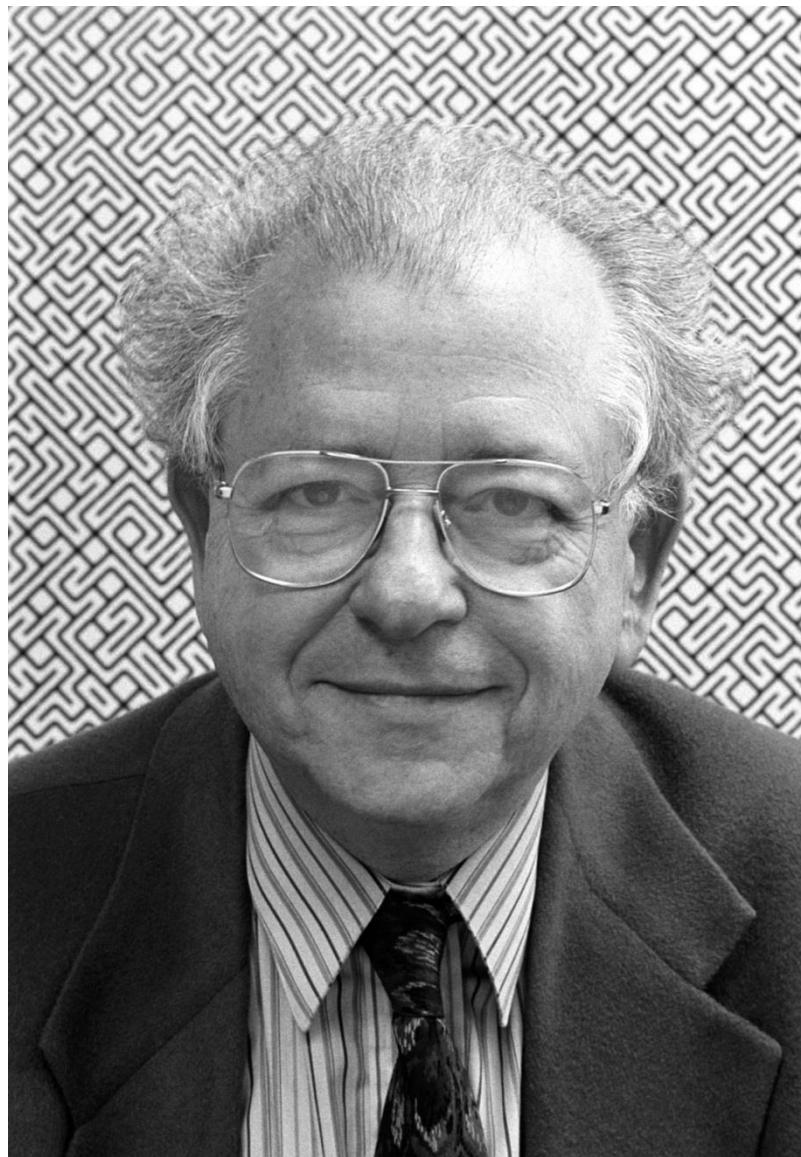


Compare textures and decide if they're made of the same “stuff”.

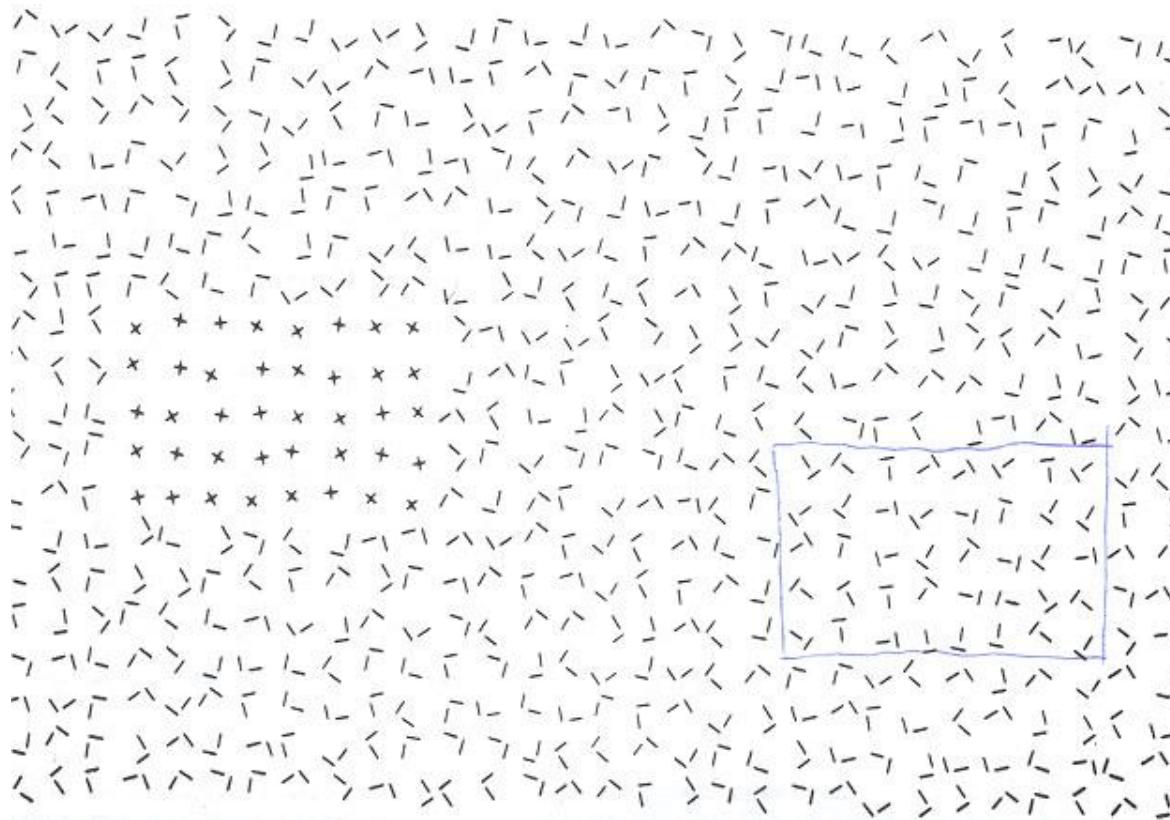
When are two textures similar?



Béla Julesz, father of texture

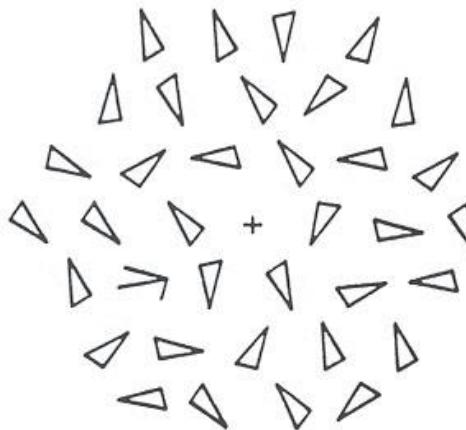
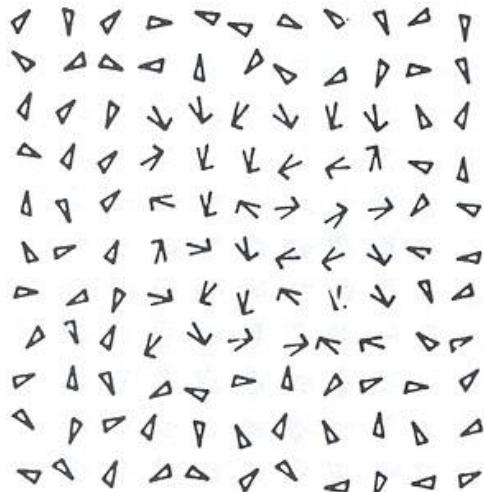
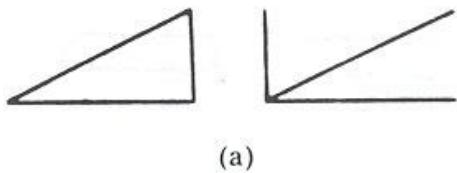


Texton Discrimination (Julesz)



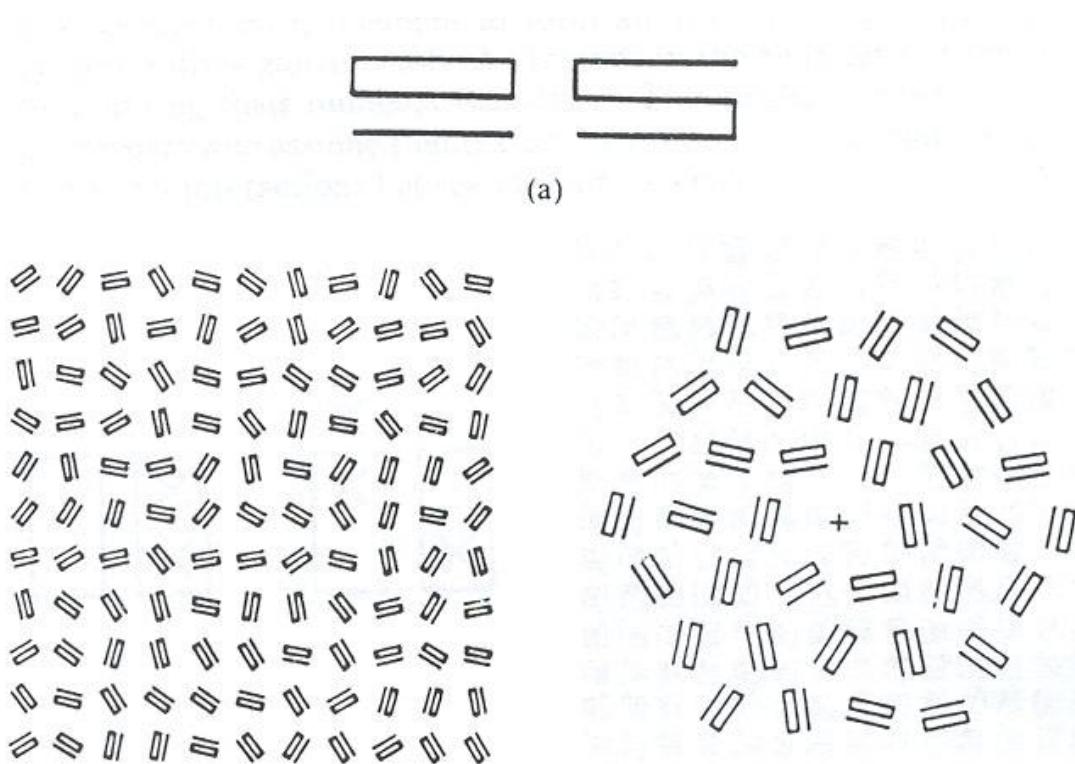
Human vision is sensitive to the difference of some types of elements and appears to be “numb” on other types of differences.

Search Experiment I



The subject is told to detect a target element in a number of background elements.
In this example, the detection time is independent of the number of background elements.

Search Experiment II



In this example, the detection time is proportional to the number of background elements,
And thus suggests that the subject is doing element-by-element scrutiny.

Preattentive vs Attentive Vision (Julesz)

Human vision operates in two distinct modes:

1. Preattentive vision

parallel, instantaneous (~100--200ms), without scrutiny, independent of the number of patterns, covering a large visual field.

2. Attentive vision

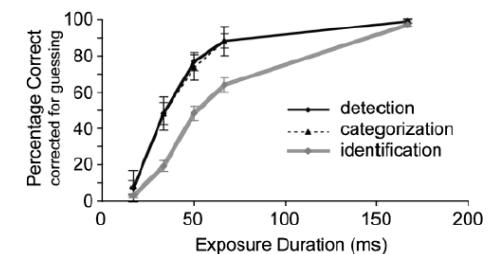
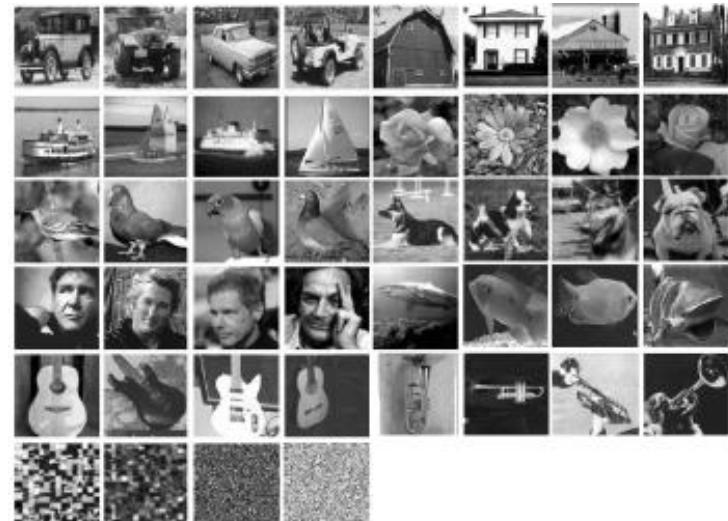
serial search by focal attention in 50ms steps limited to small aperture.

Evidence for Pre-attentive Recognition (Thorpe)

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**

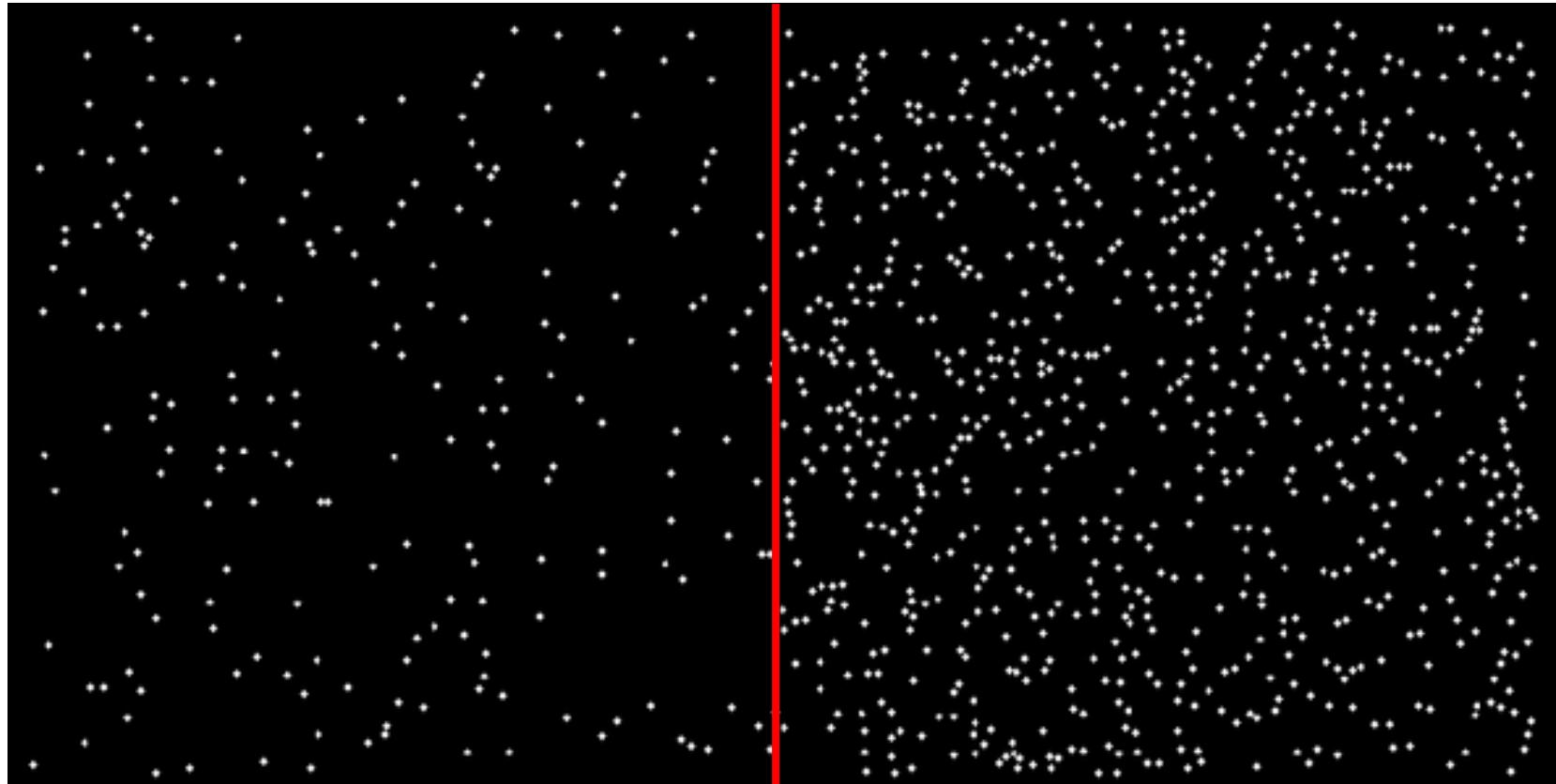
Detection and categorization are practically simultaneous (Grill-Spector & Kanwisher, 2005)



Julesz Conjecture

*Textures cannot be spontaneously discriminated if they have the **same first-order and second-order statistics** of texture features (textons) and differ only in their third-order or higher-order statistics.*

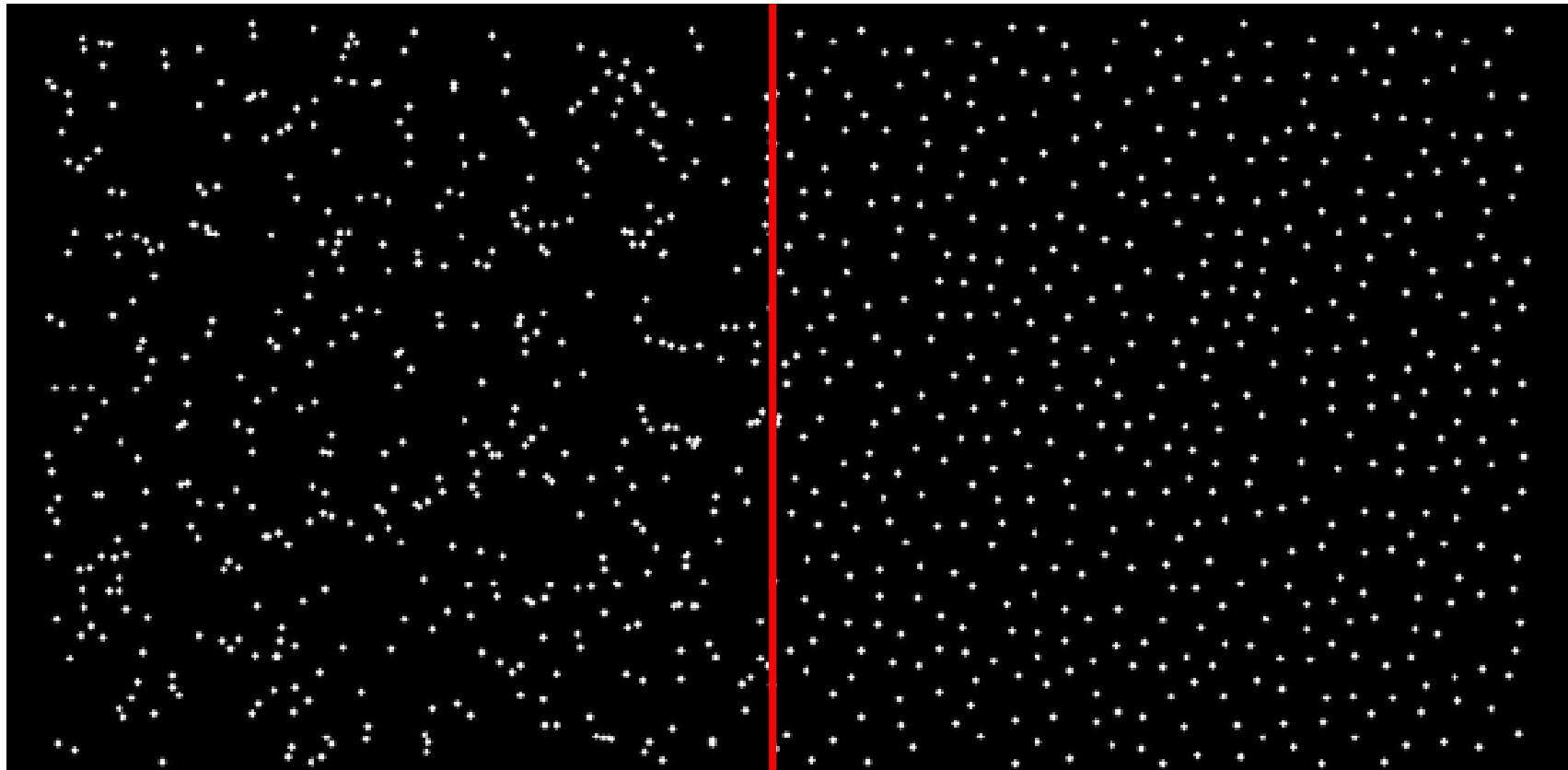
1st Order Statistics



5% white

20% white

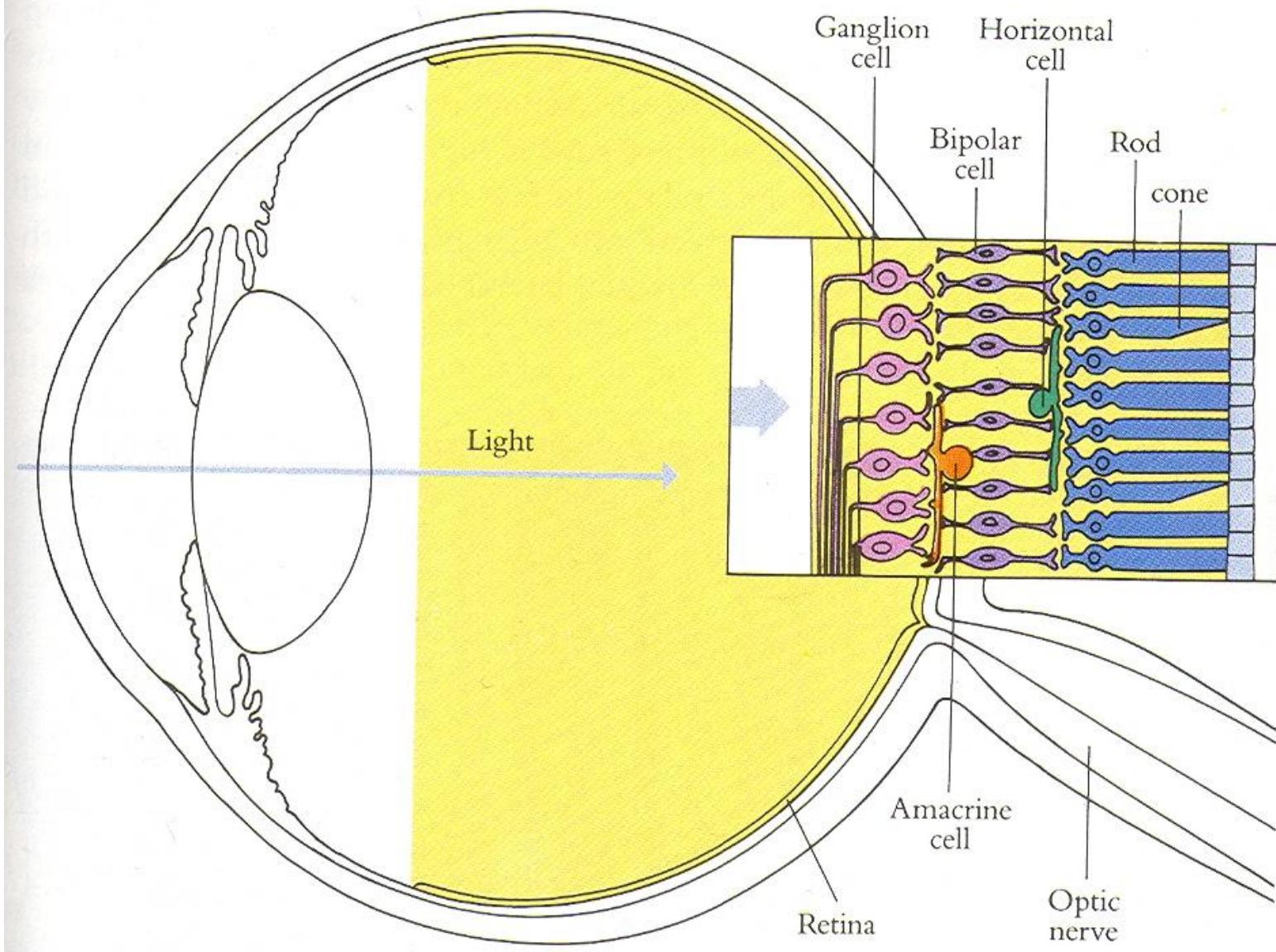
2nd Order Statistics

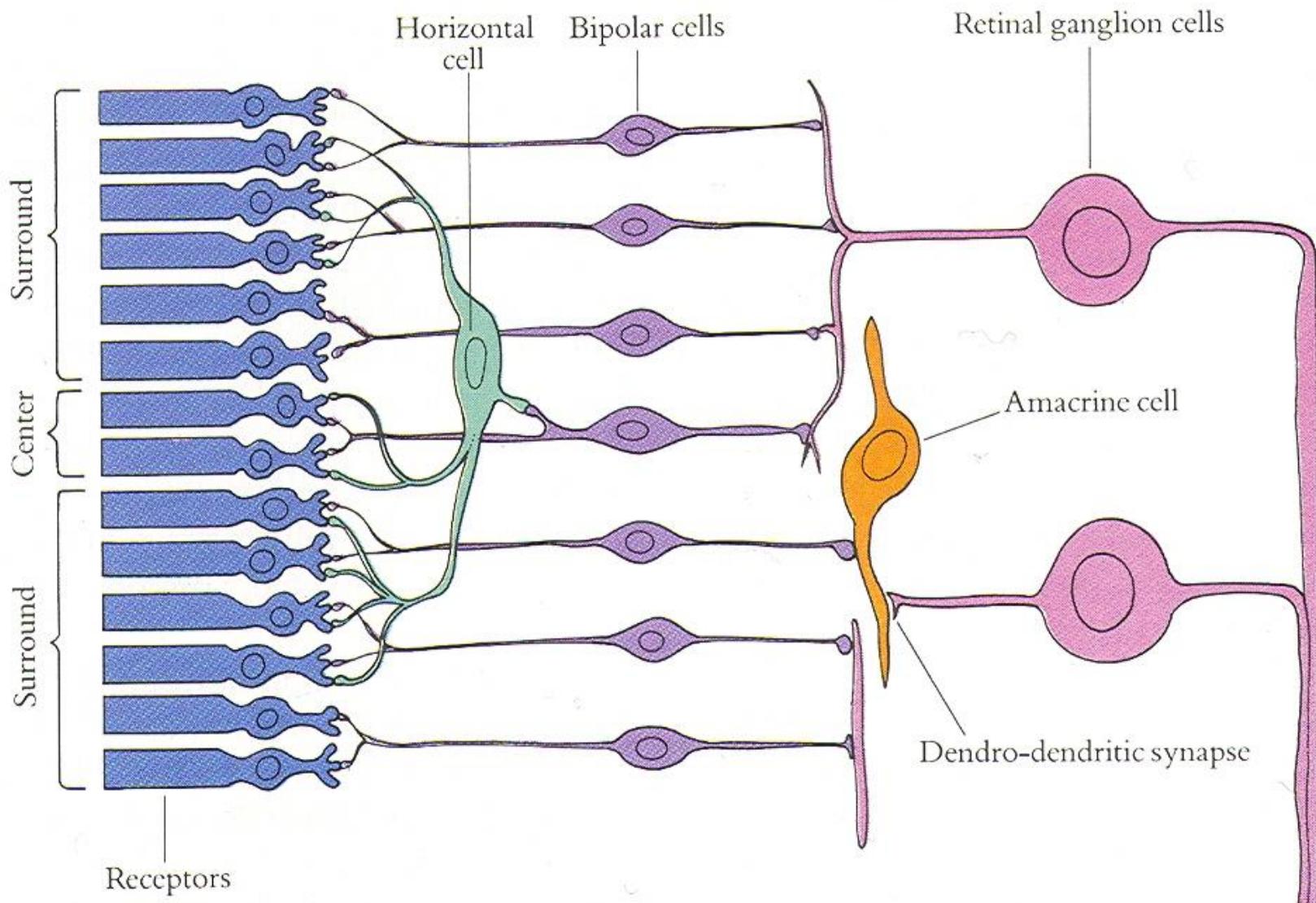


10% white

Big Question

What is the statistical unit (texton) of texture in real images?

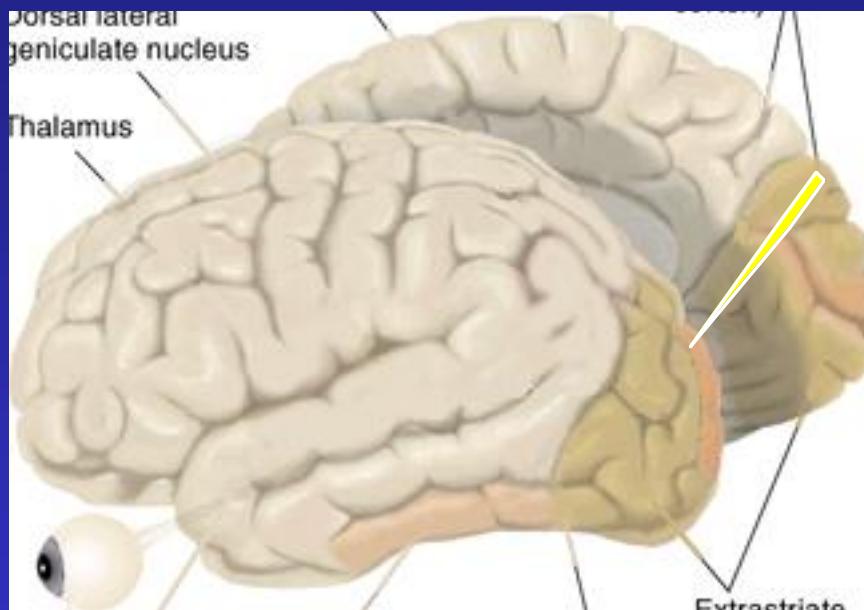
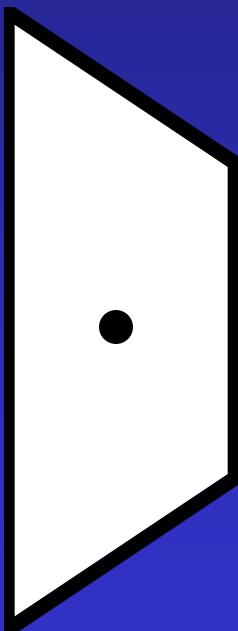




Single Cell Recording



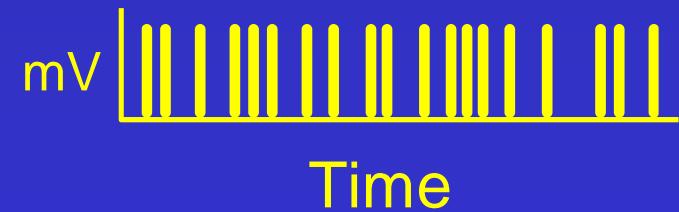
Single Cell Recording



Microelectrode

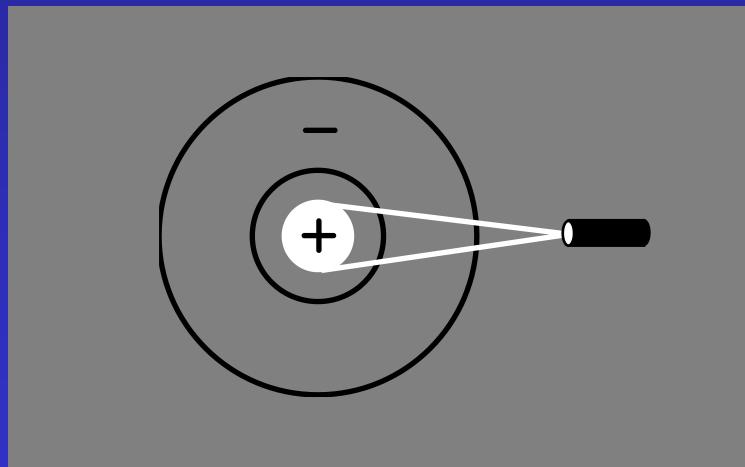
Amplifier

Electrical response
(action potentials)

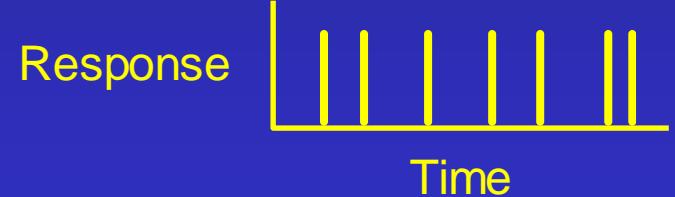


Retinal Receptive Fields

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



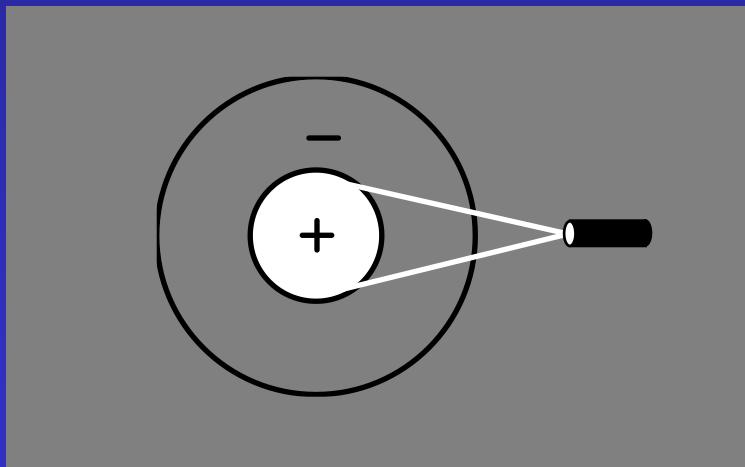
Stimulus condition



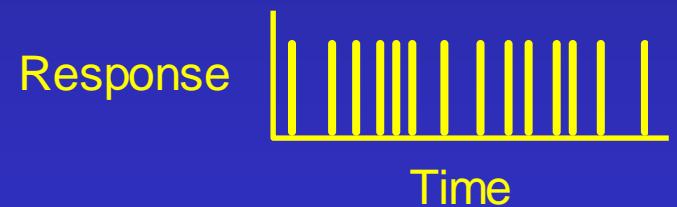
Electrical response

Retinal Receptive Fields

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



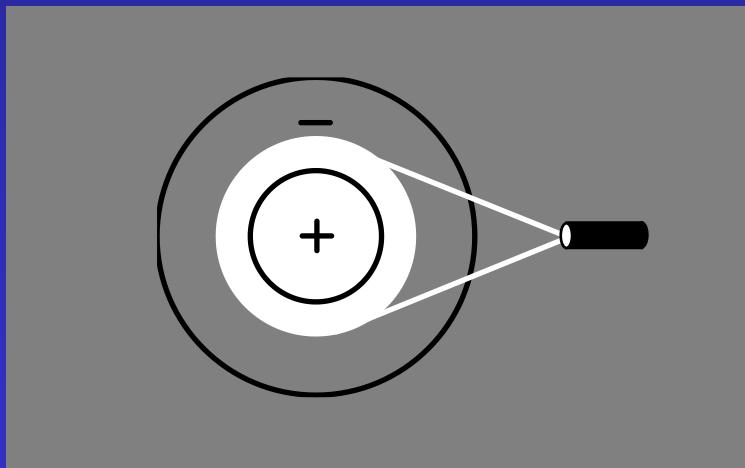
Stimulus condition



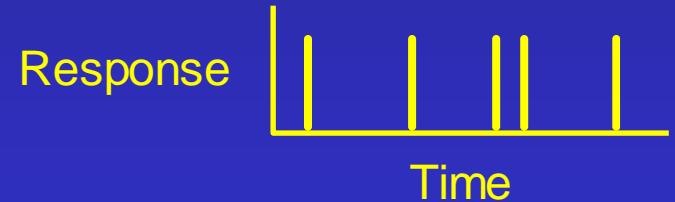
Electrical response

Retinal Receptive Fields

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



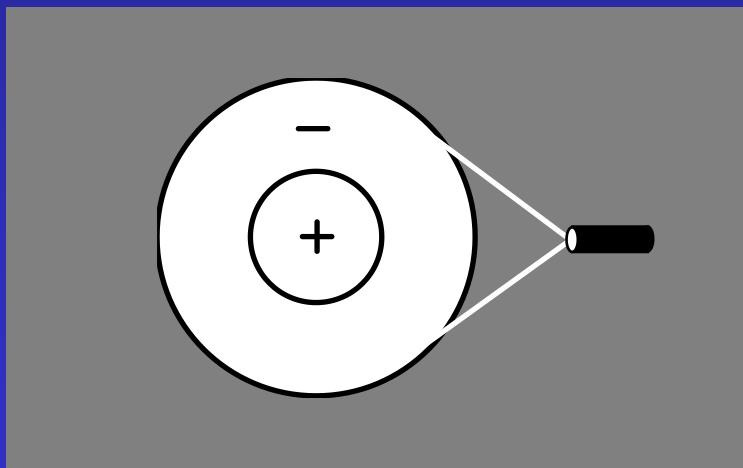
Stimulus condition



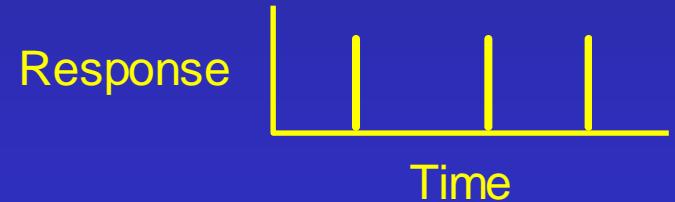
Electrical response

Retinal Receptive Fields

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



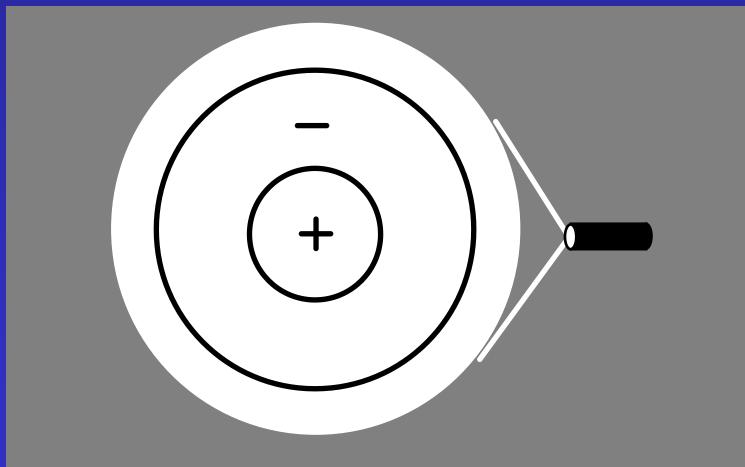
Stimulus condition



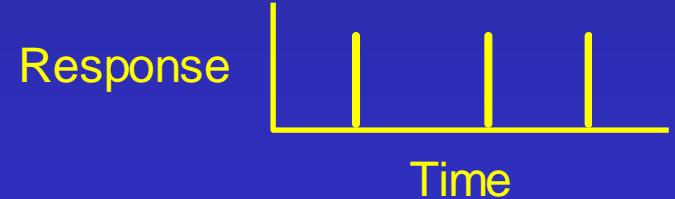
Electrical response

Retinal Receptive Fields

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



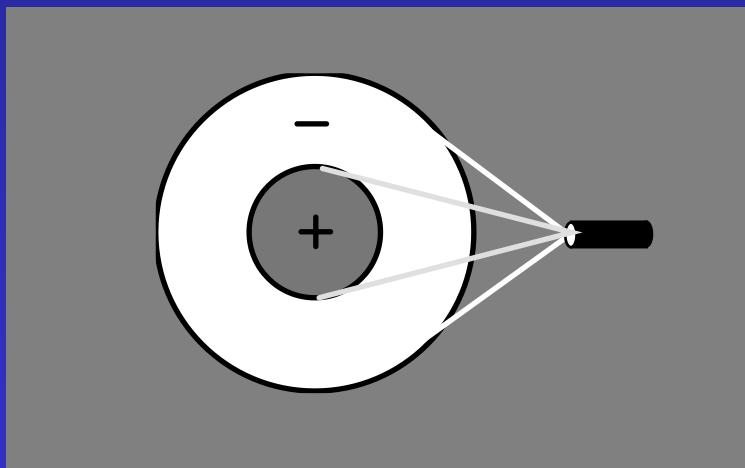
Stimulus condition



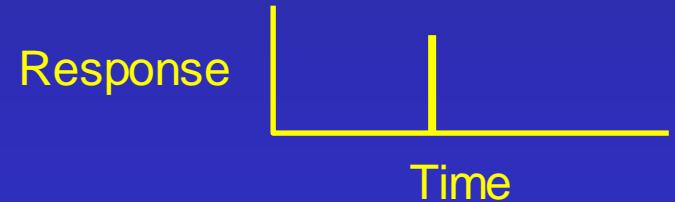
Electrical response

Retinal Receptive Fields

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



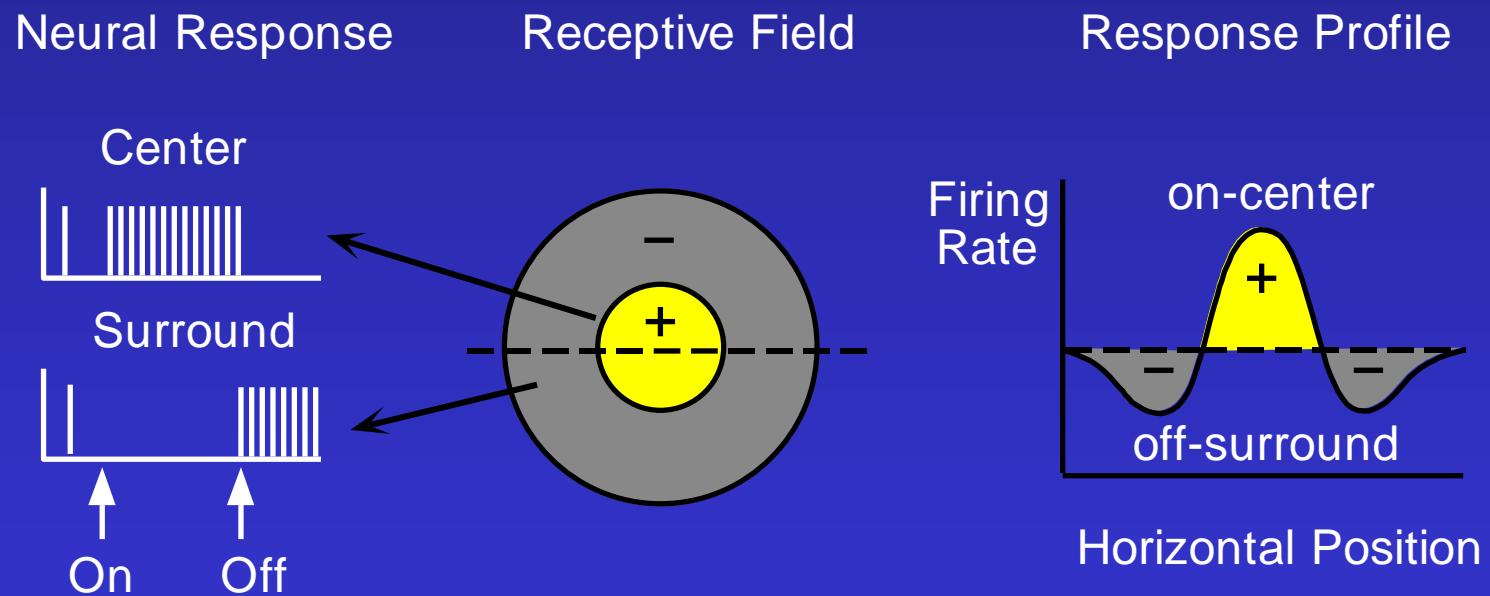
Stimulus condition



Electrical response

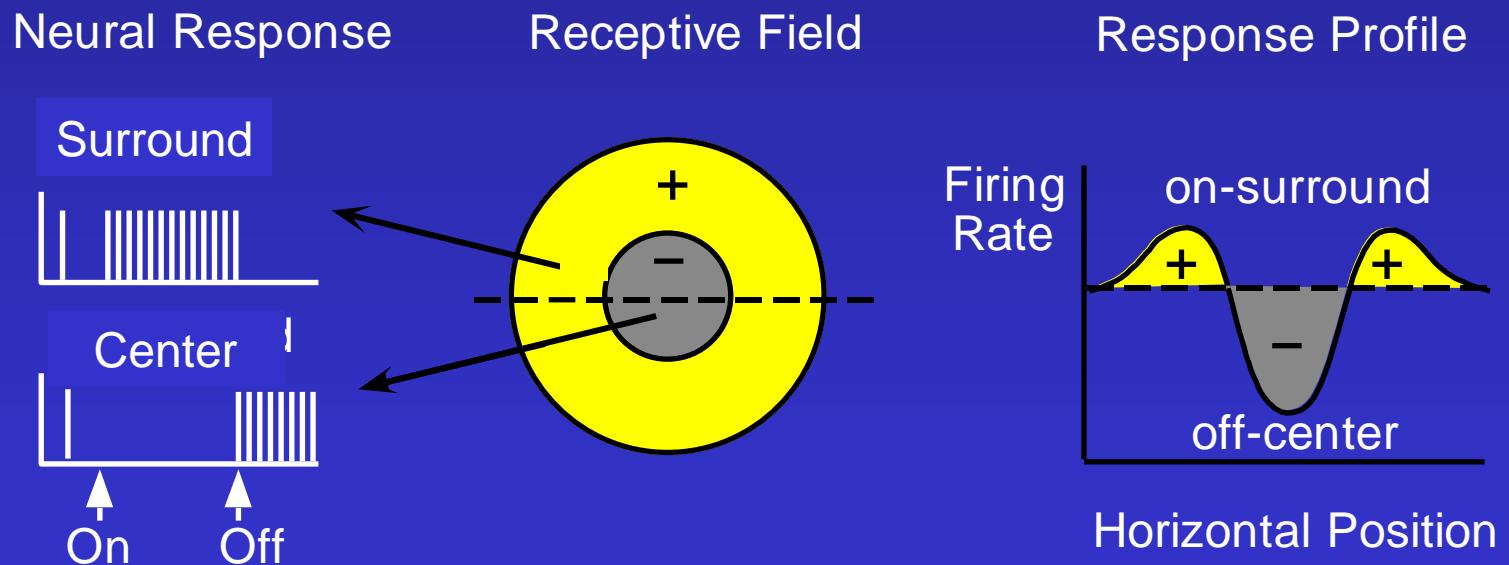
Retinal Receptive Fields

RF of On-center Off-surround cells

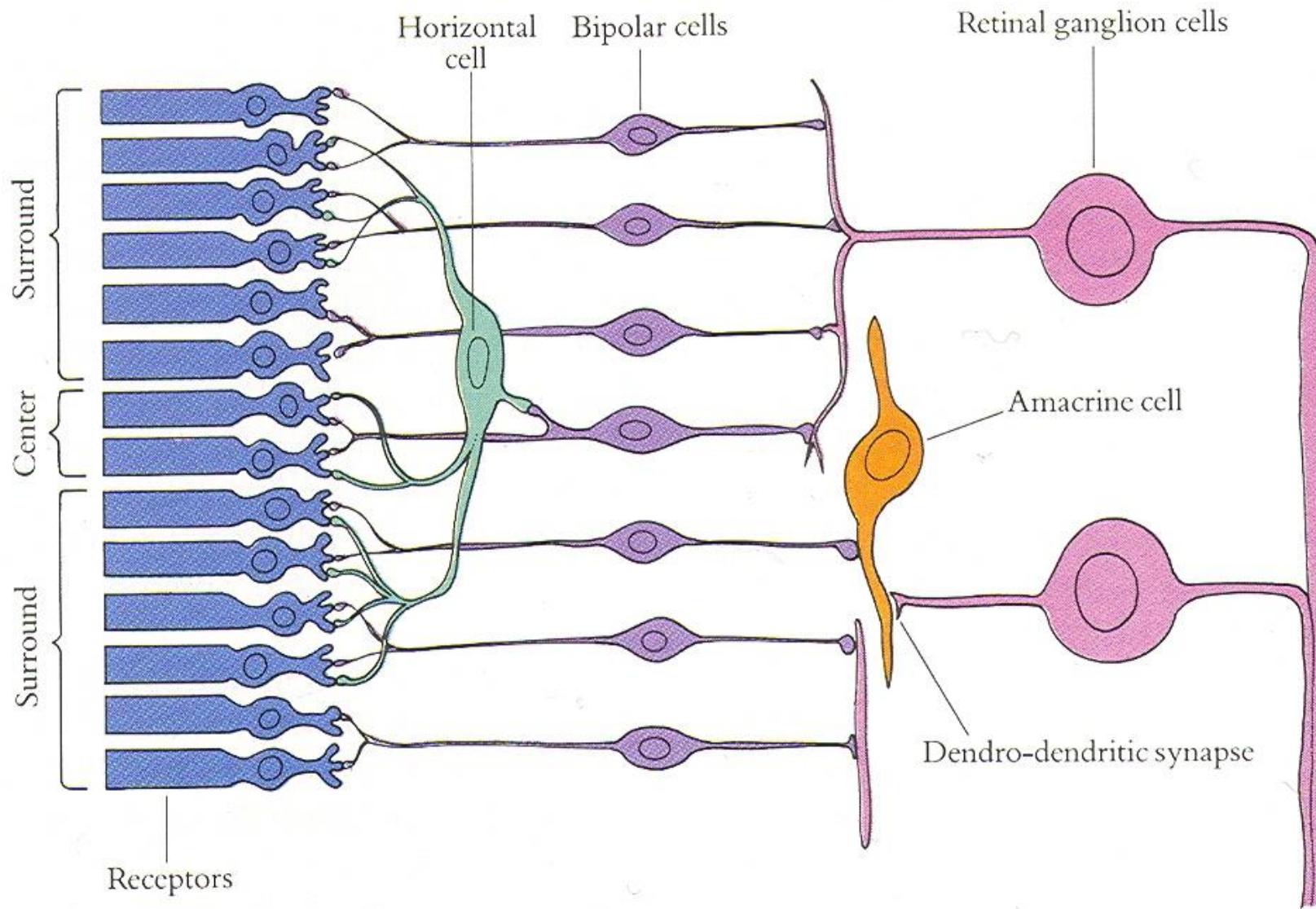


Retinal Receptive Fields

RF of Off-center On-surround cells

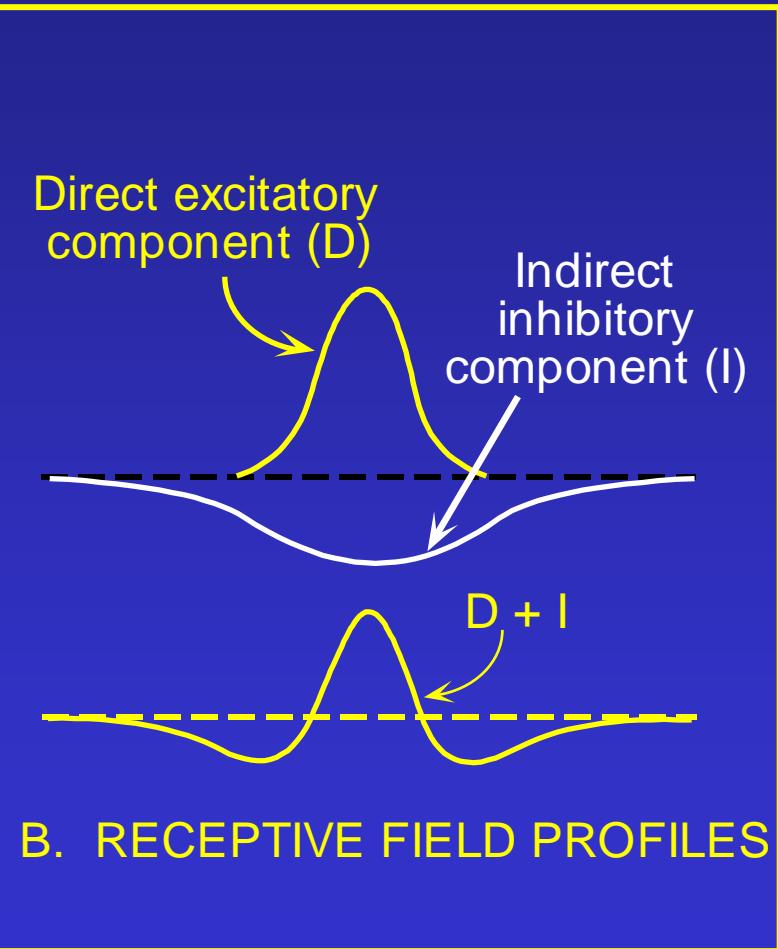
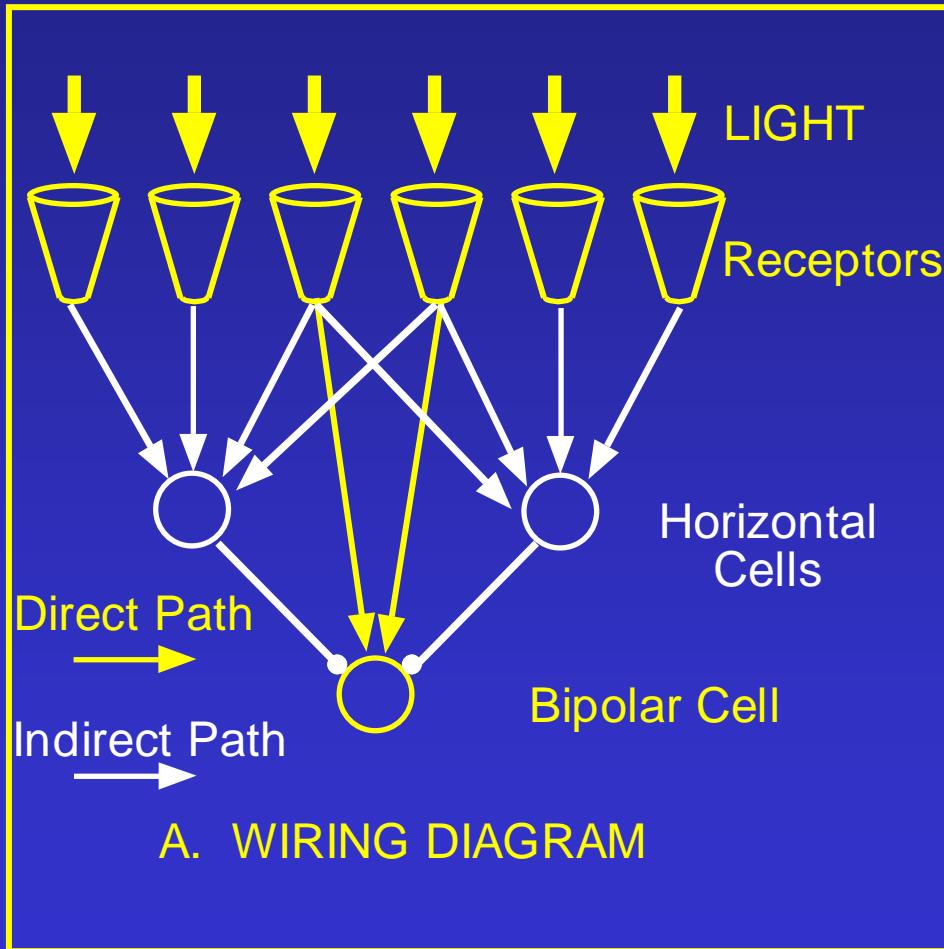


Retinal Receptive Fields



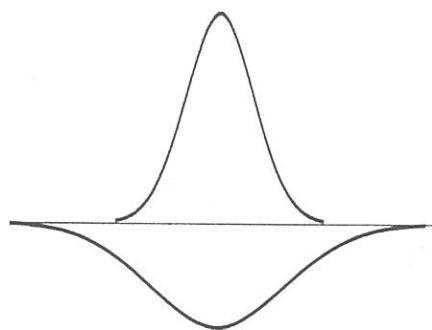
Retinal Receptive Fields

Receptive field structure in bipolar cells

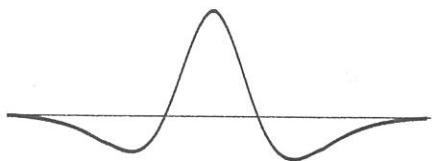


The receptive field of a retinal ganglion cell can be modeled as a “Difference of Gaussians”

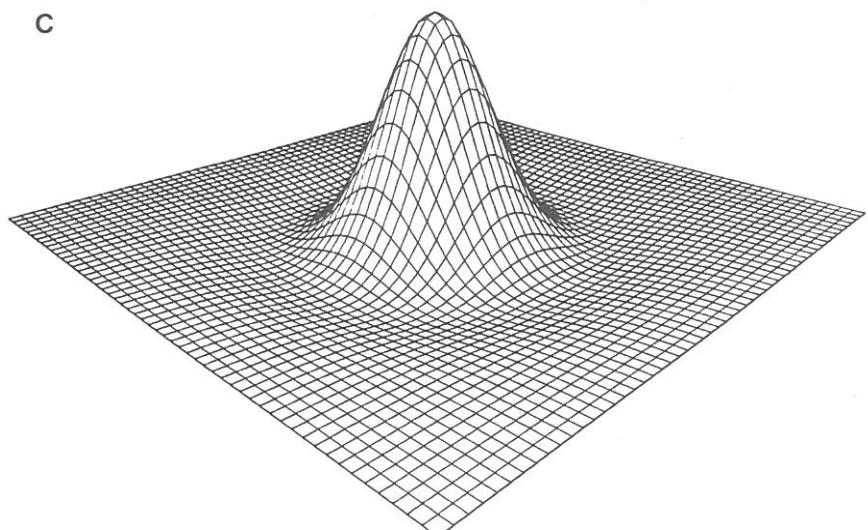
A



B



C



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

Receptive Fields

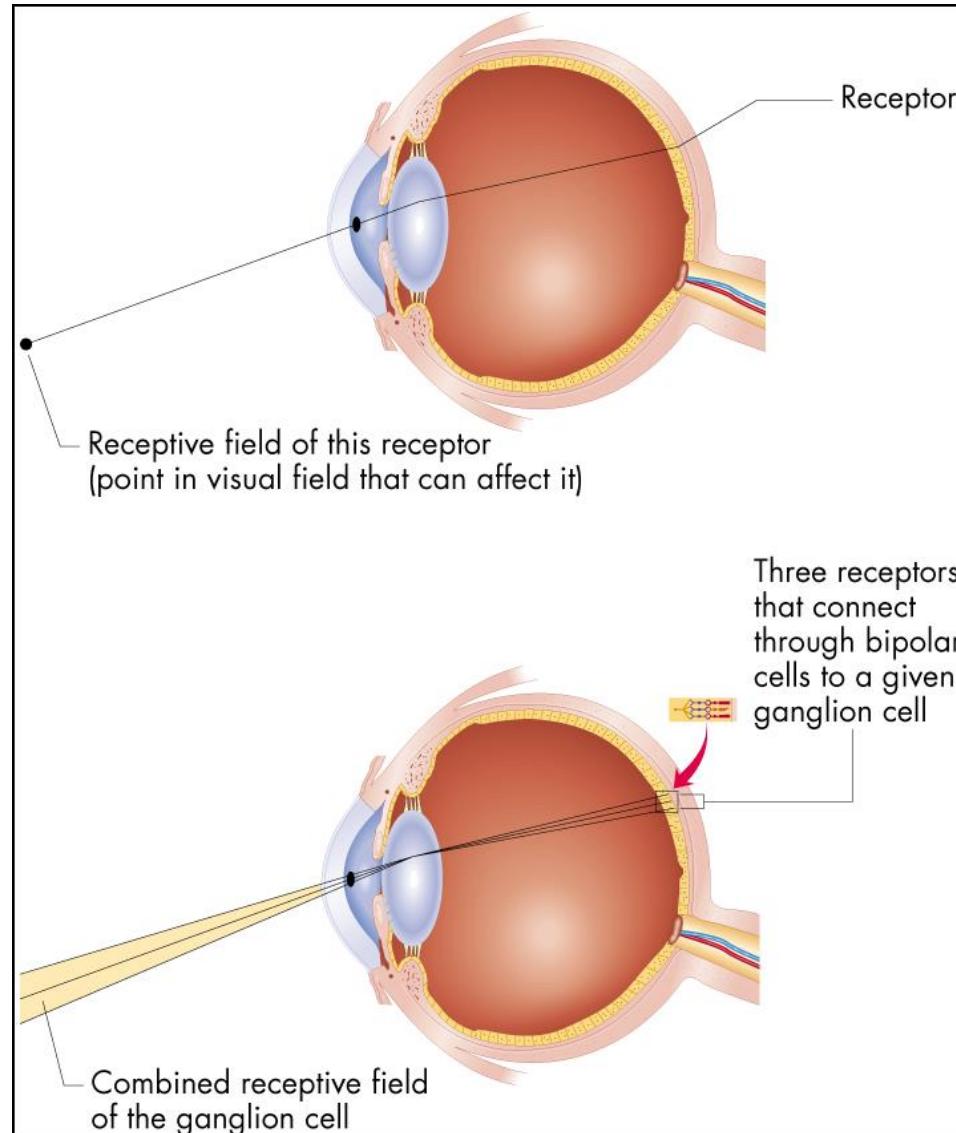
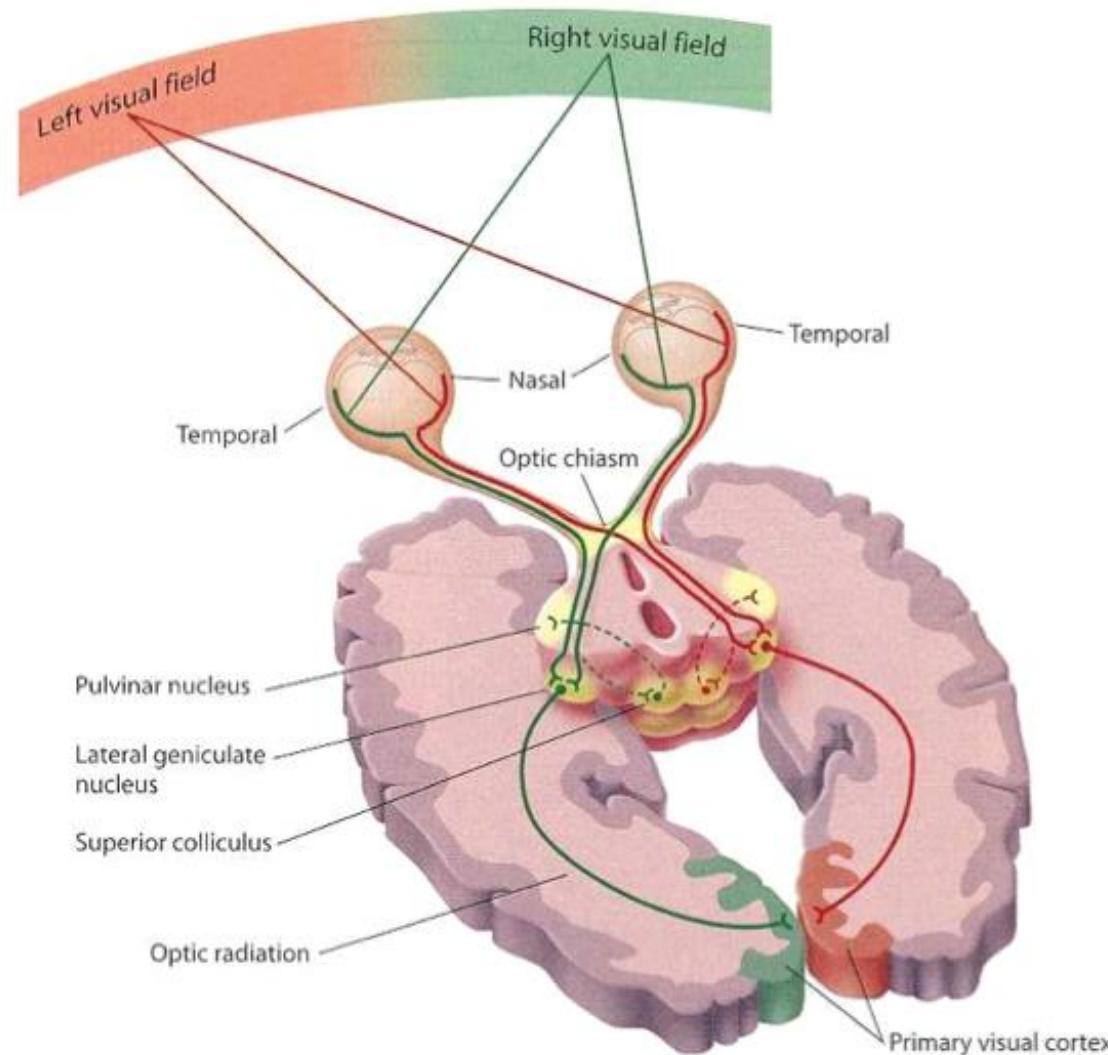


Figure 6.16 Receptive fields

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.

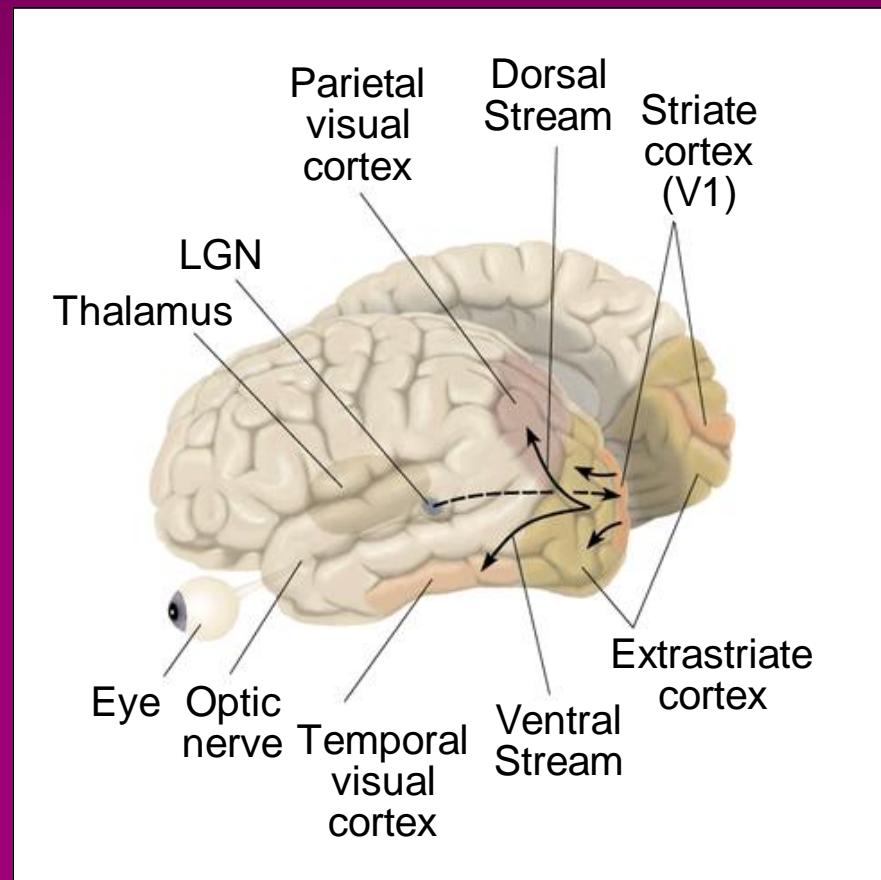
Anatomy of Pathway to Visual Cortex



Visual Cortex

Cortical Area V1

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



Cortical Receptive Fields

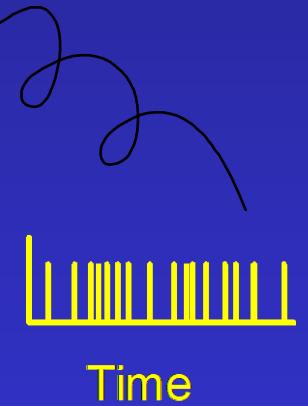
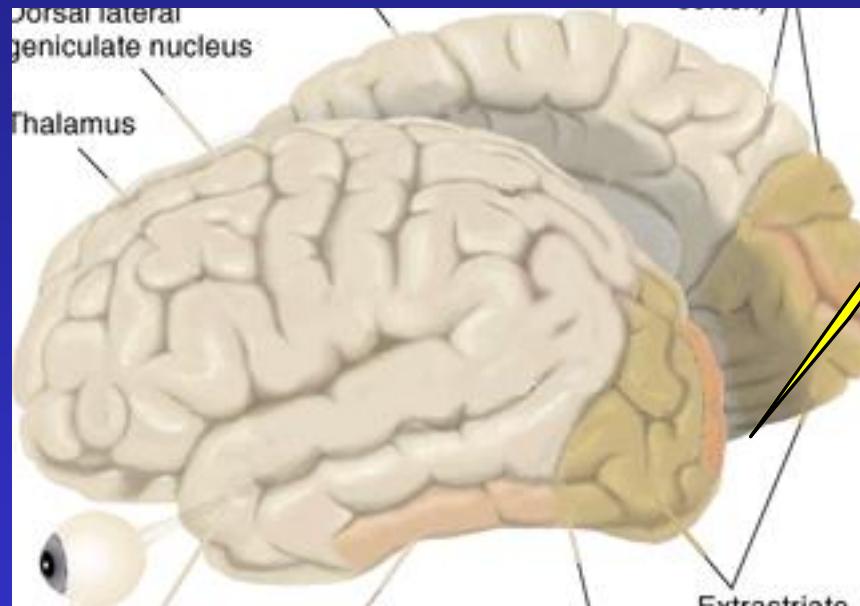
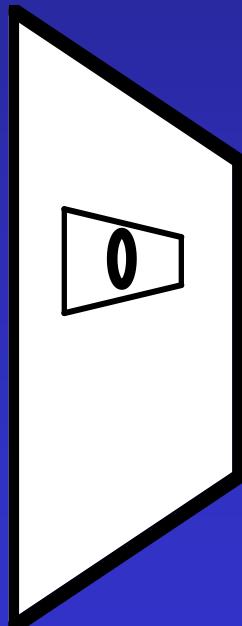
Single-cell recording from visual cortex



David Hubel & Thorston Wiesel

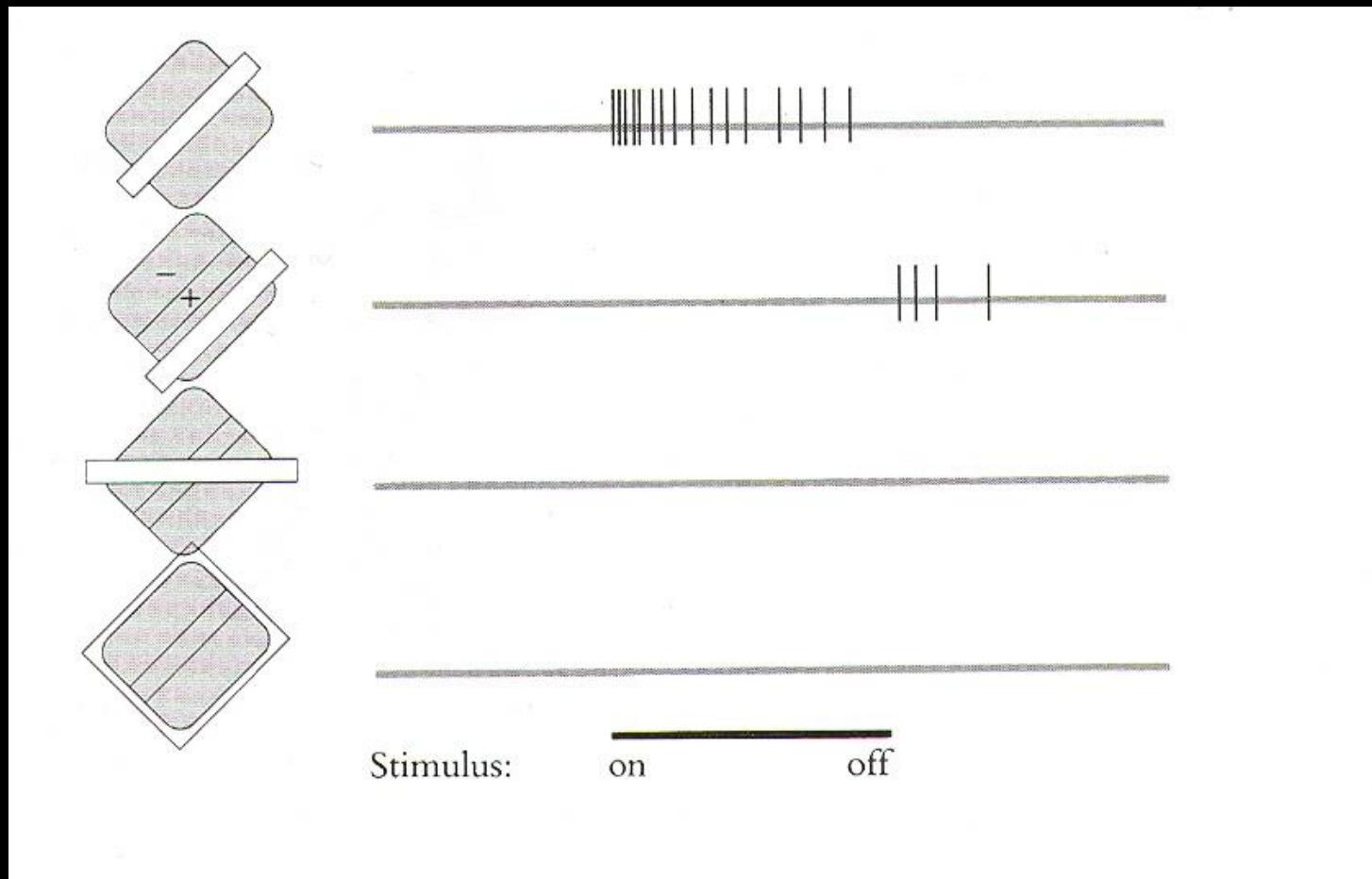
Cortical Receptive Fields

Single-cell recording from visual cortex





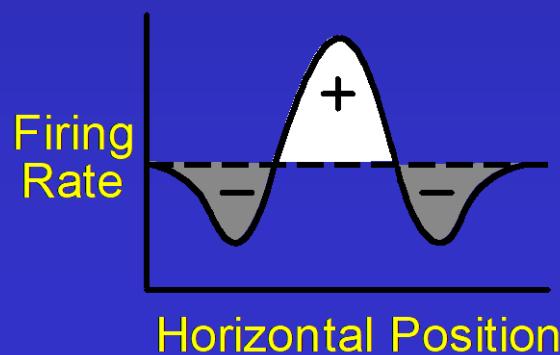
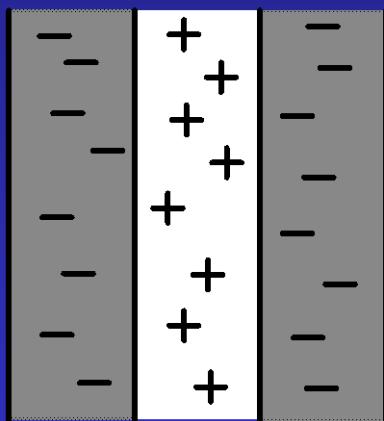
<https://www.youtube.com/watch?v=IOHayh06LJ4>



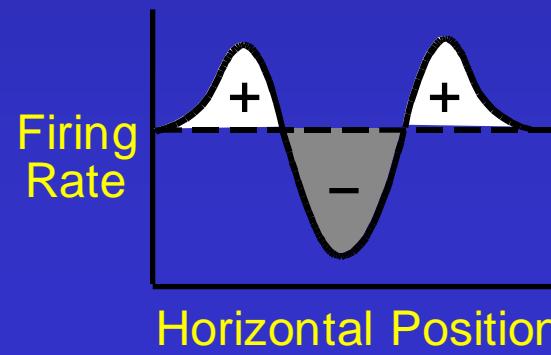
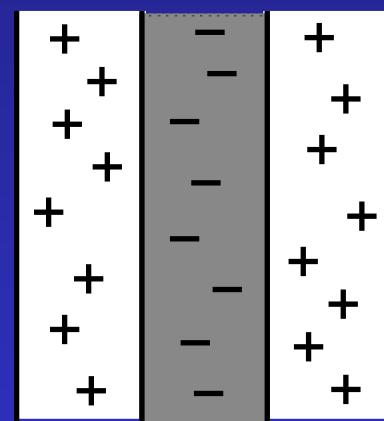
Cortical Receptive Fields

Simple Cells: “Line Detectors”

A. Light Line Detector



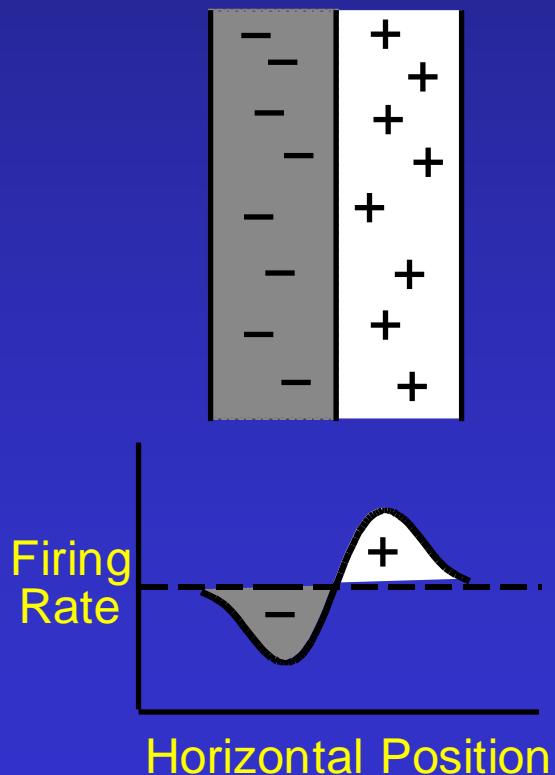
B. Dark Line Detector



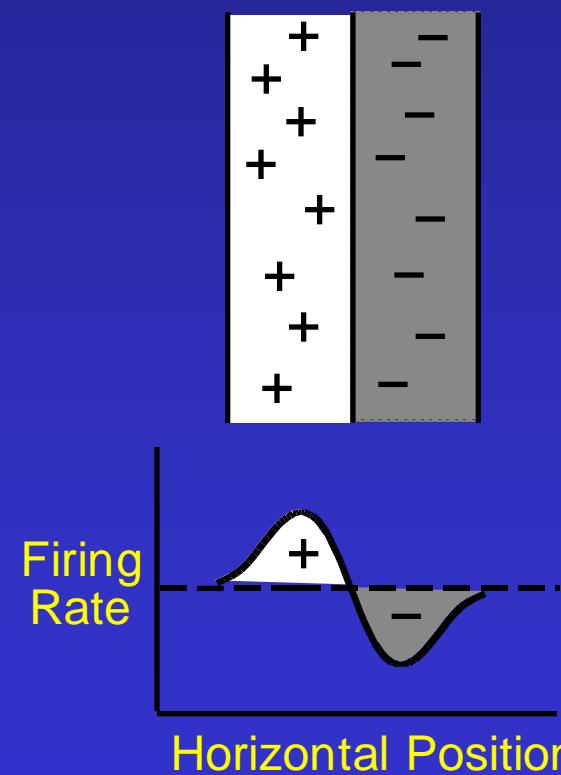
Cortical Receptive Fields

Simple Cells: “Edge Detectors”

C. Dark-to-light Edge Detecto

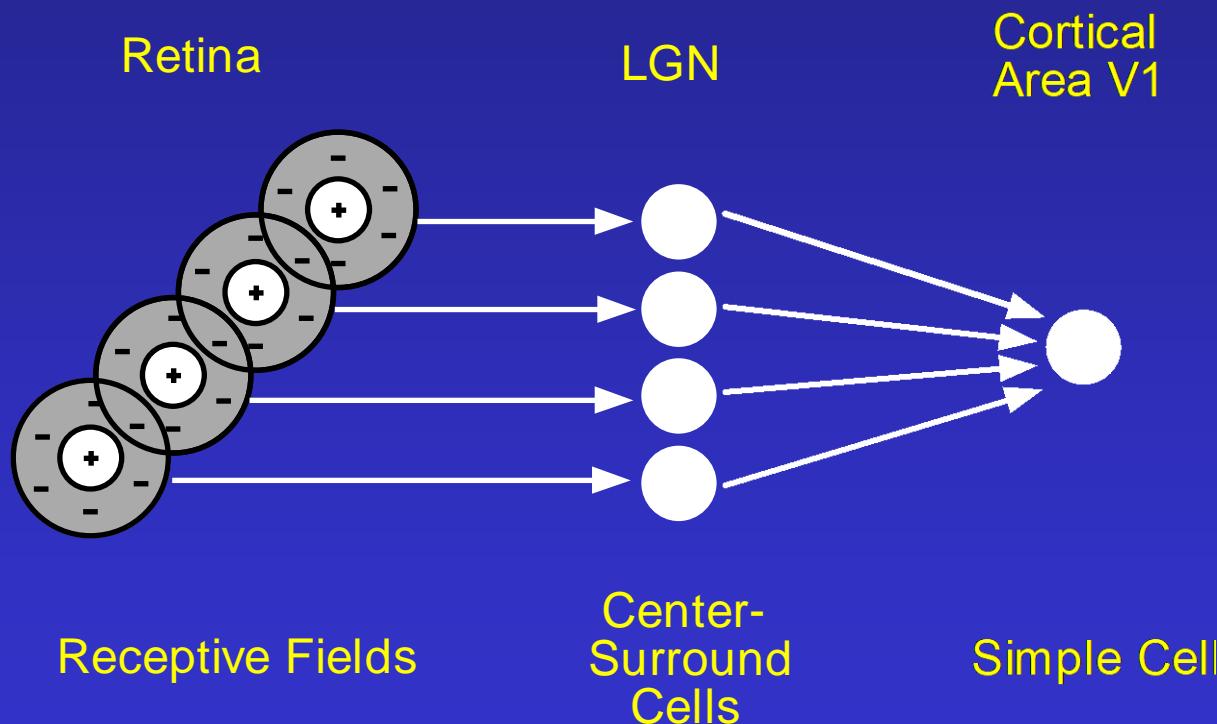


D. Light-to-dark Edge Detecto

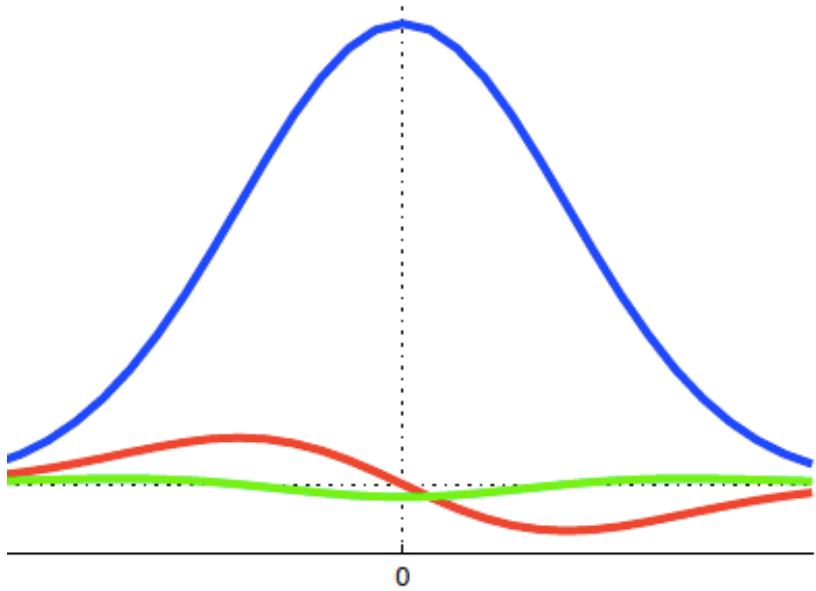


Cortical Receptive Fields

Constructing a line detector



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

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

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

We also consider rotated versions of these Gaussian derivative functions.

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

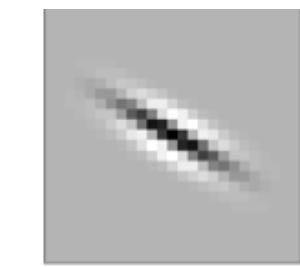
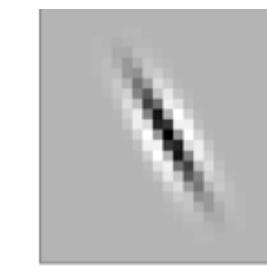
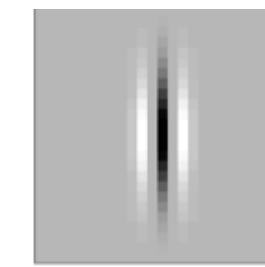
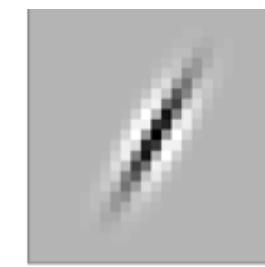
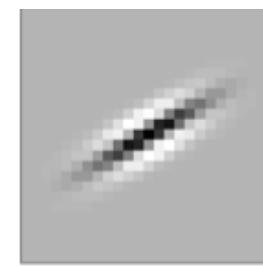
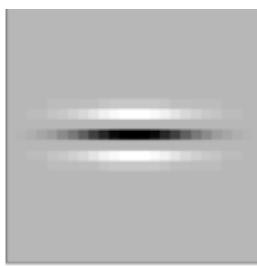
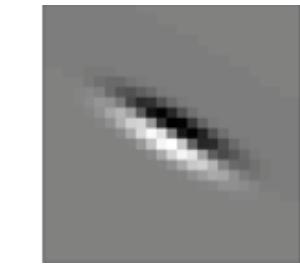
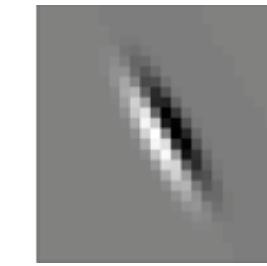
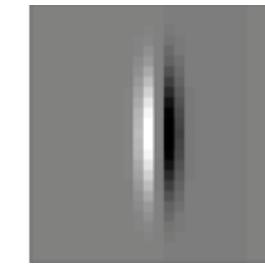
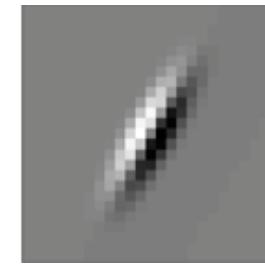
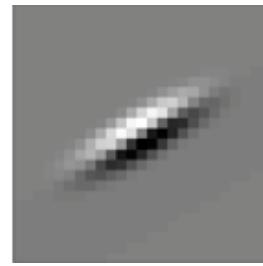
$$Rot_\theta f_2 = G''_{\sigma_1}(u)G_{\sigma_2}(v) \quad (10.7)$$

where we set

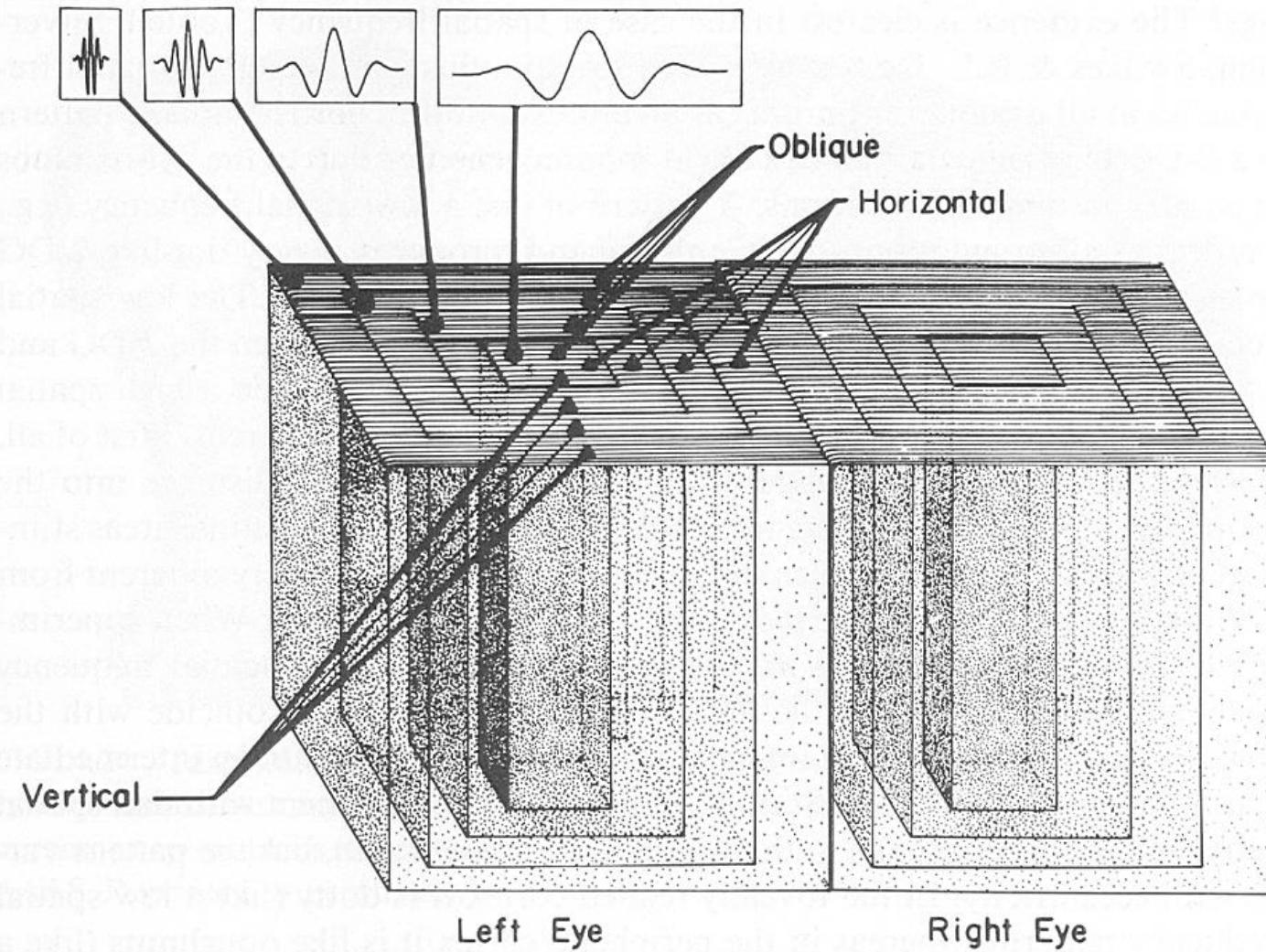
$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

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

Oriented Gaussian First and Second Derivatives



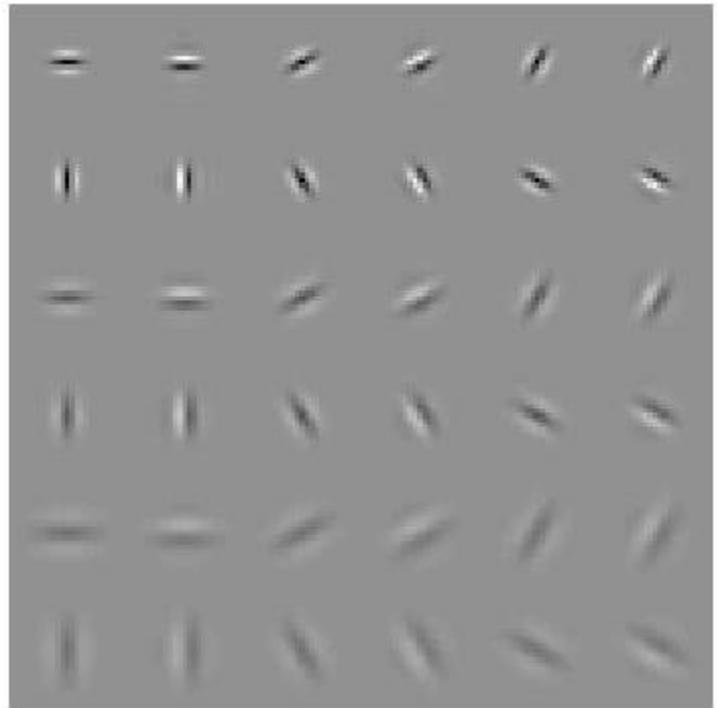
Hypercolumns in visual cortex



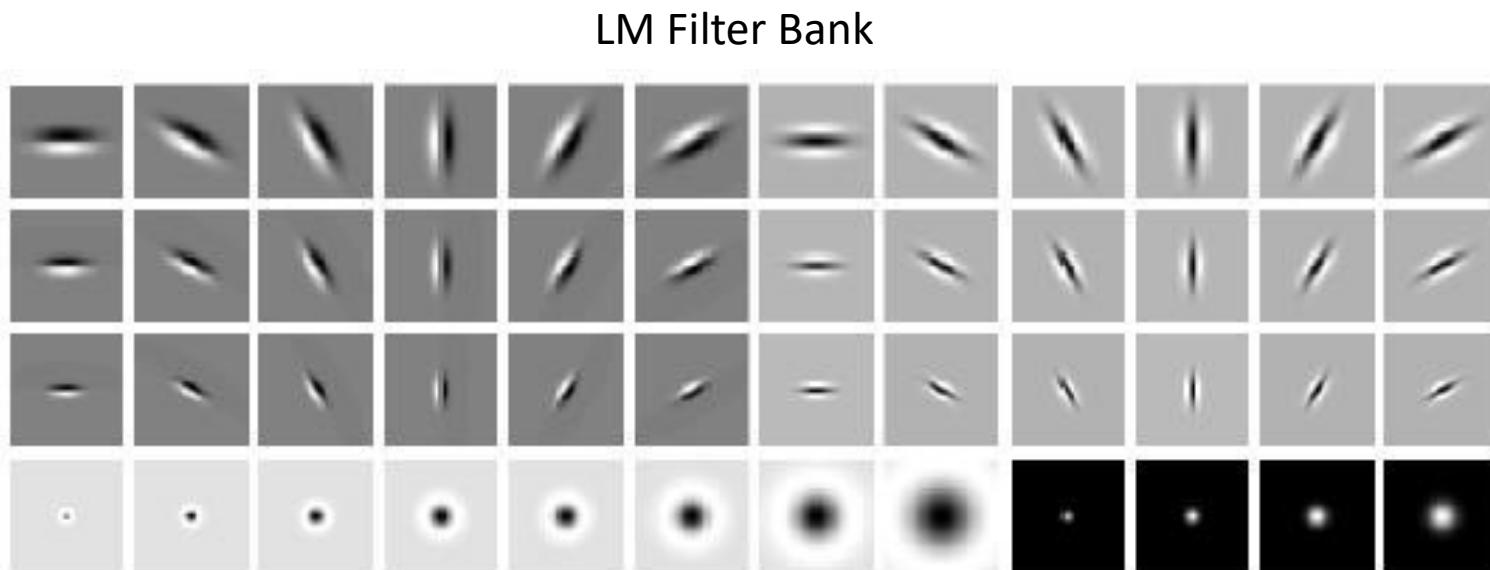
Model of Striate Module in Monkeys

Modeling hypercolumns

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



Overcomplete representation: filter banks

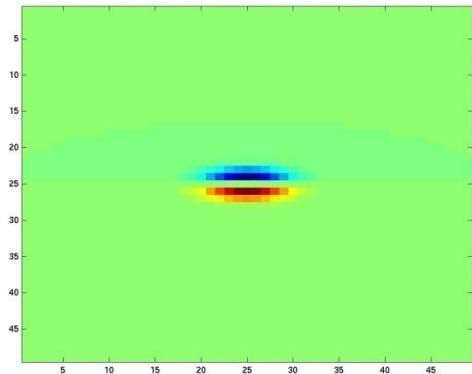
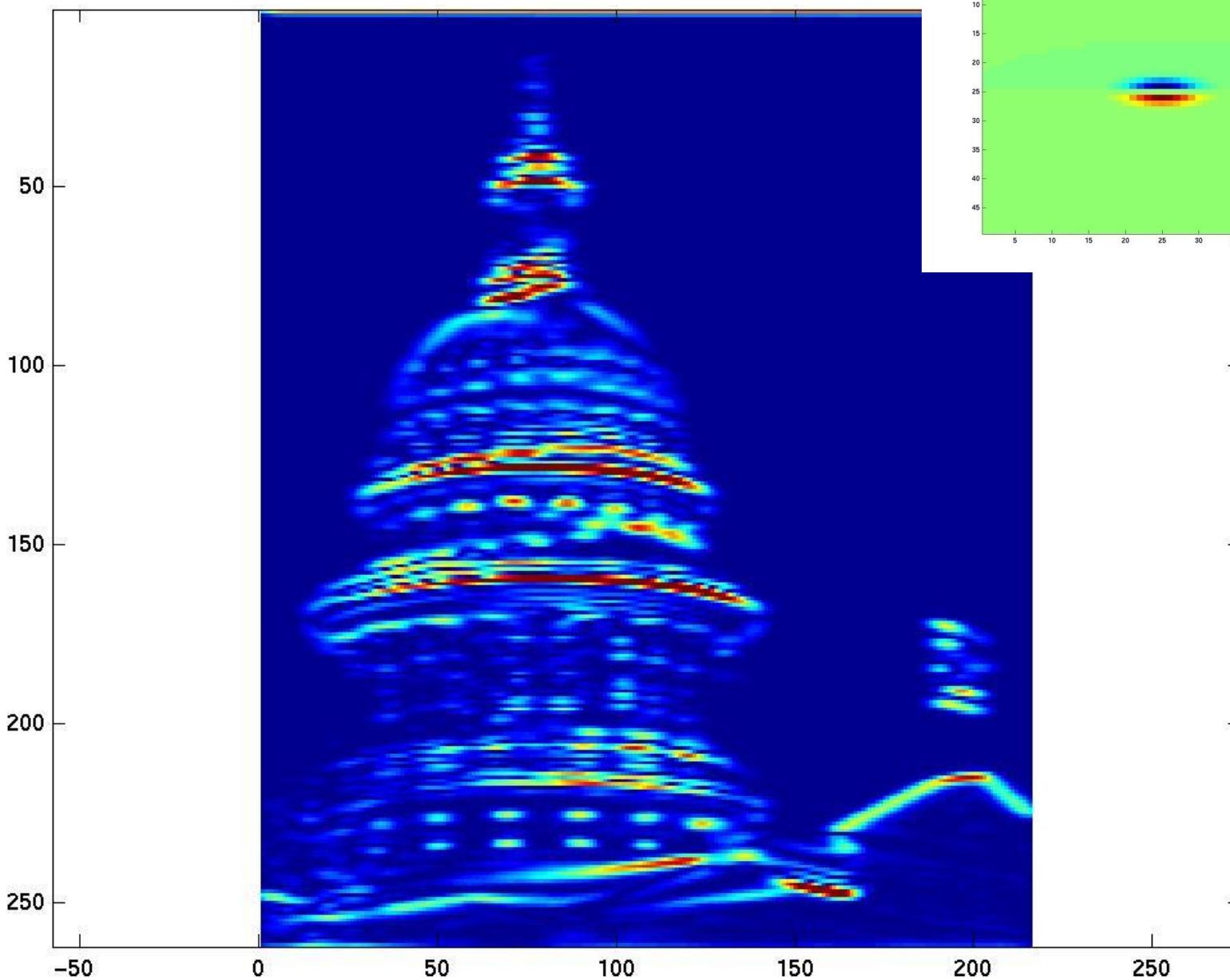


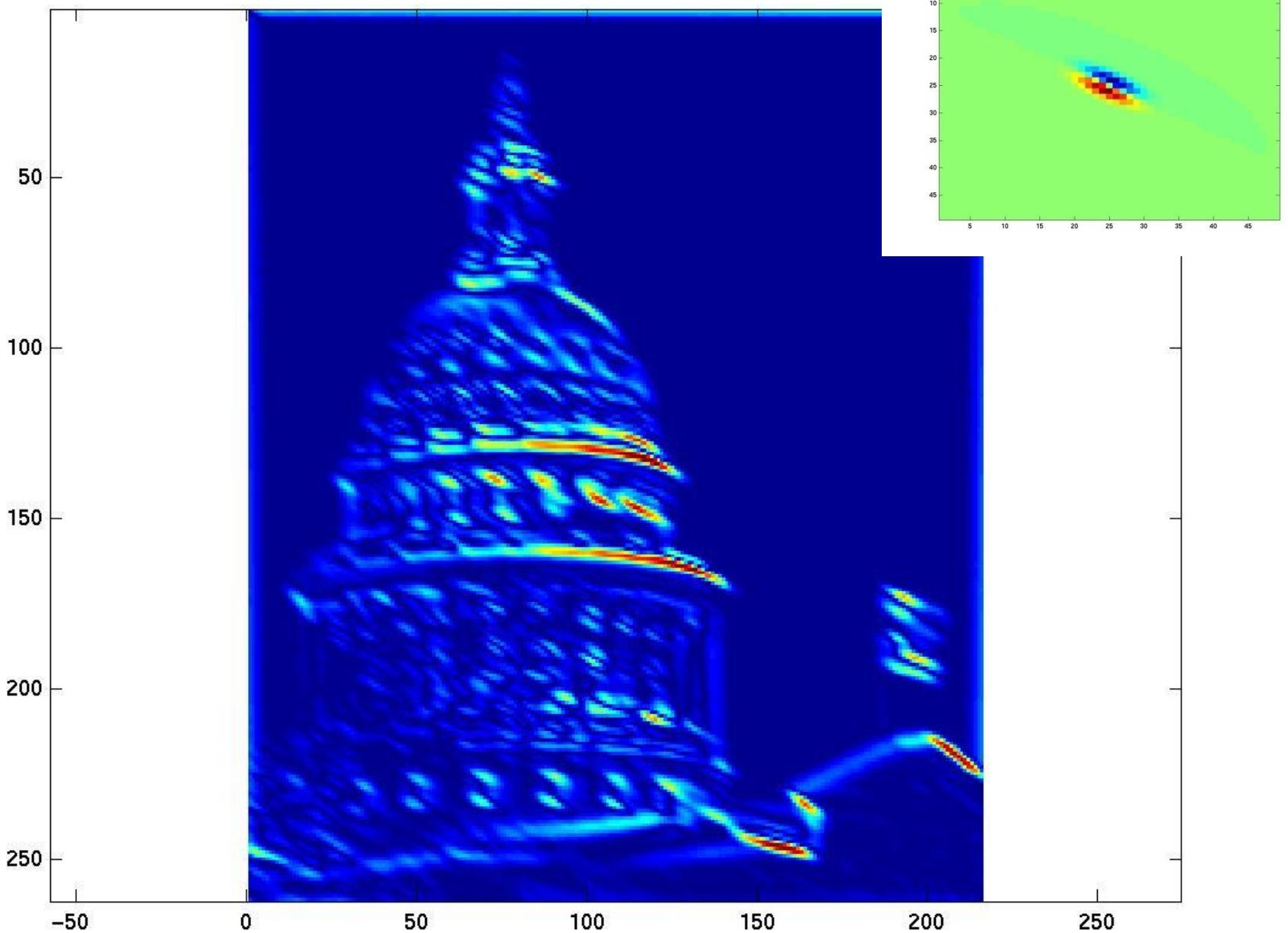
Code for filter banks: www.robots.ox.ac.uk/~vgg/research/texclass/filters.html

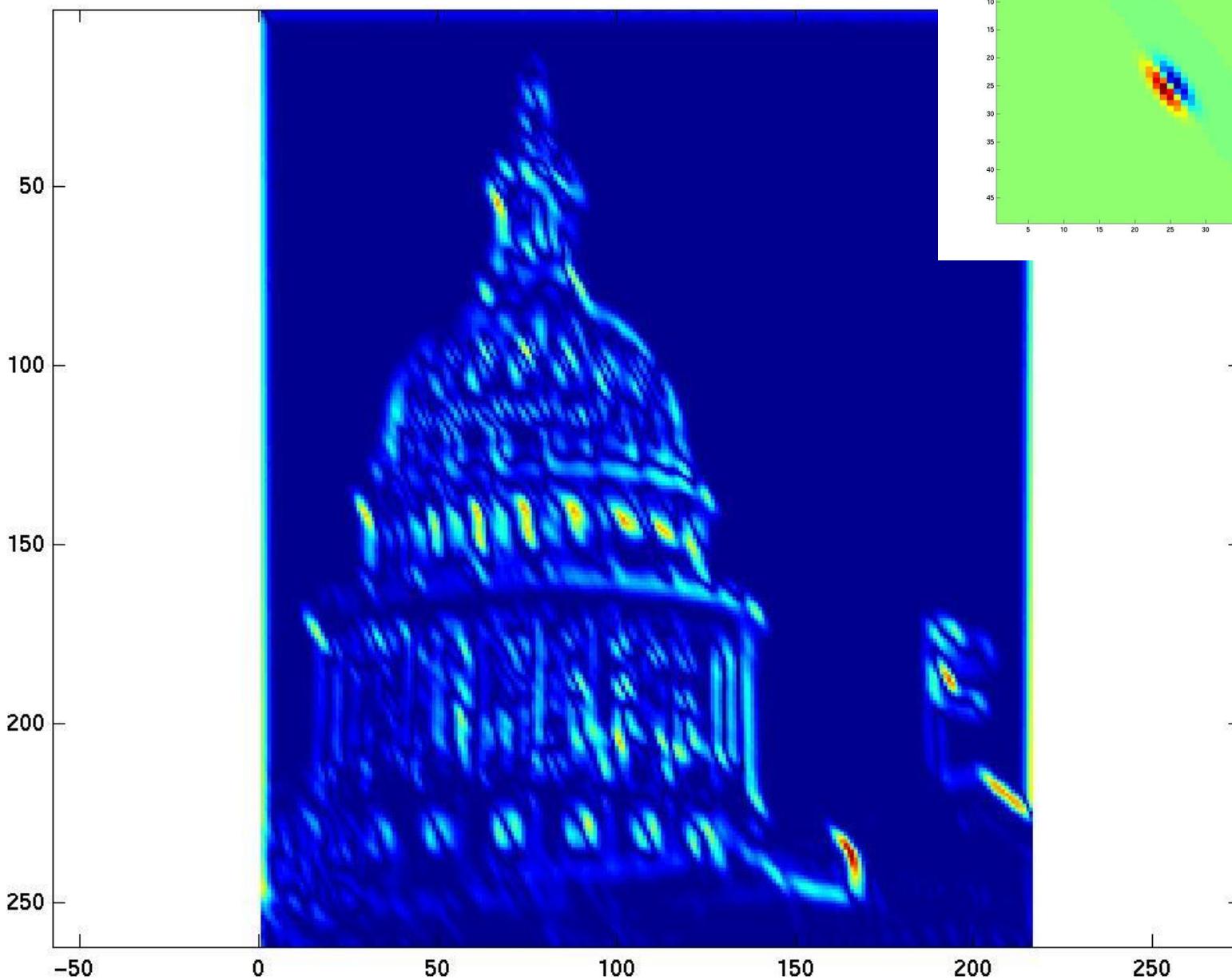
 Image from <http://www.texasexplorer.com/austincap2.jpg>

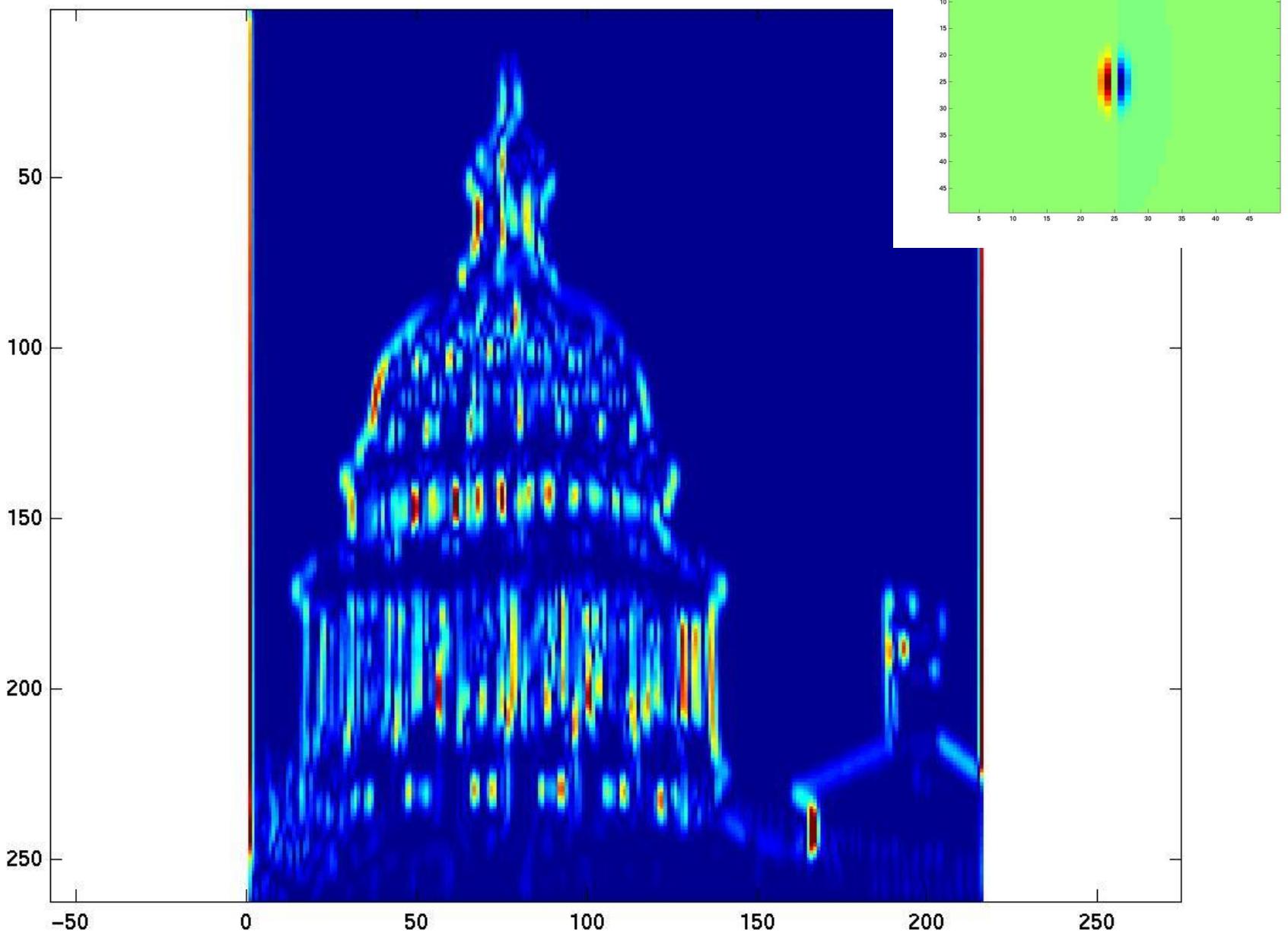


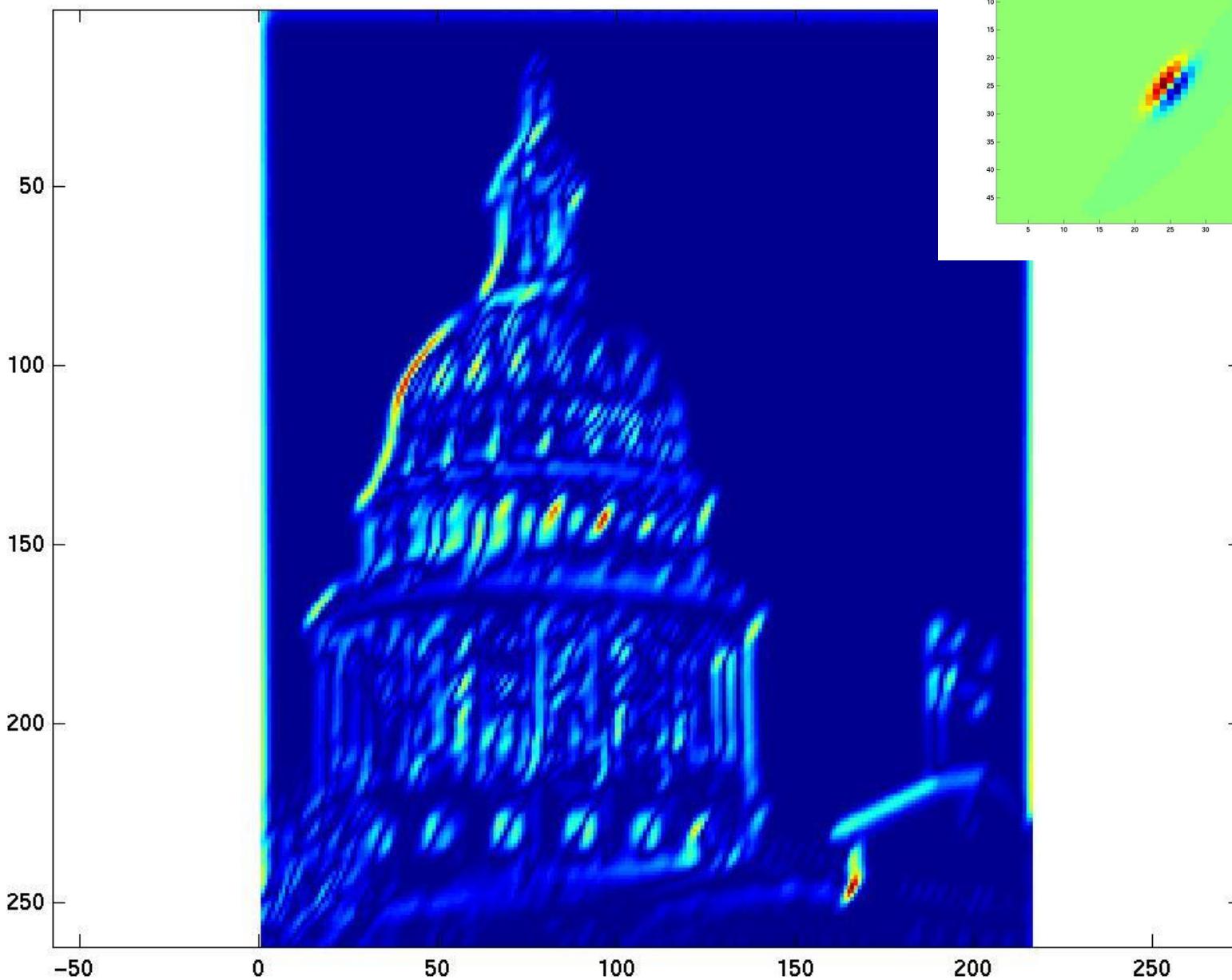
 Kristen Grauman

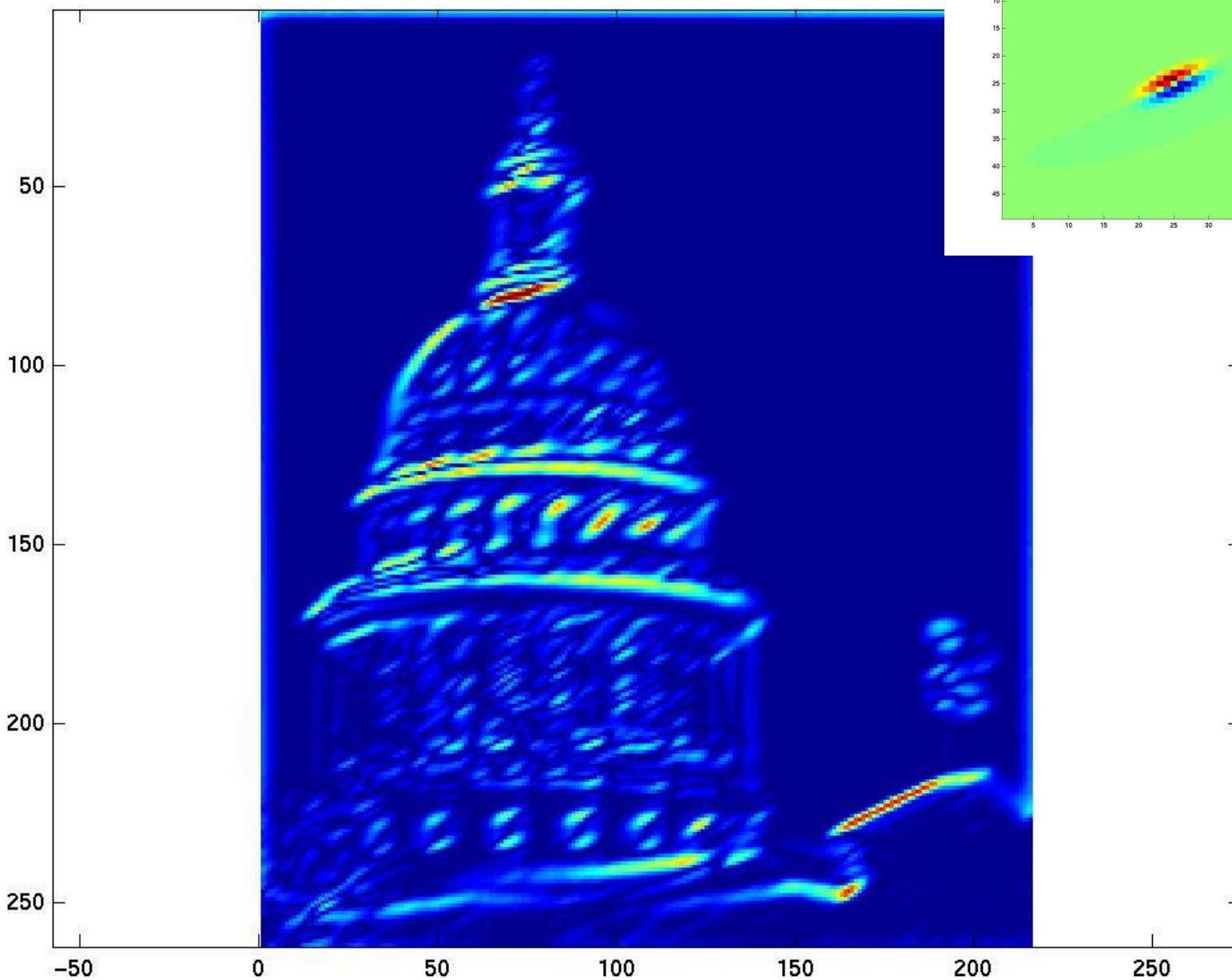


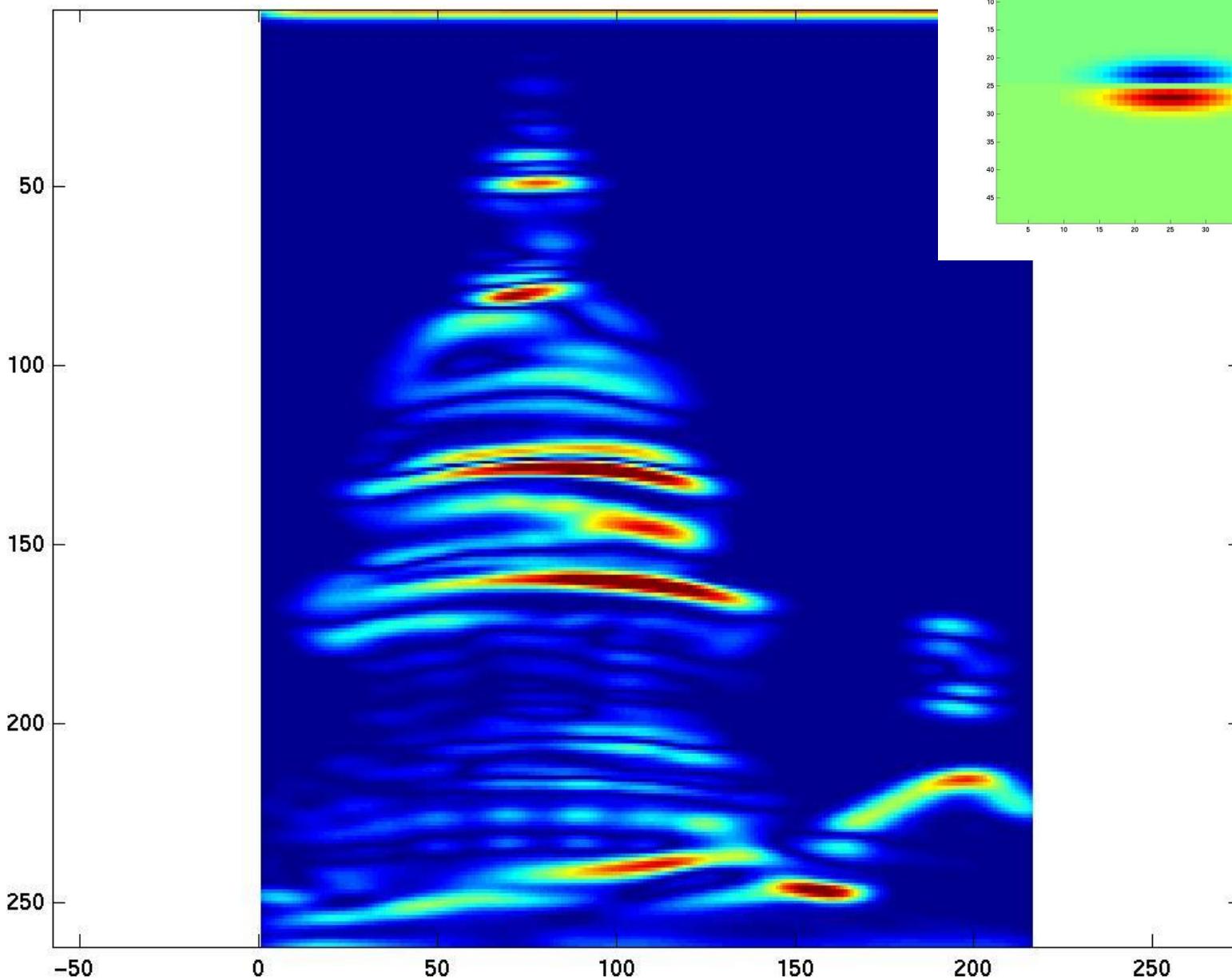


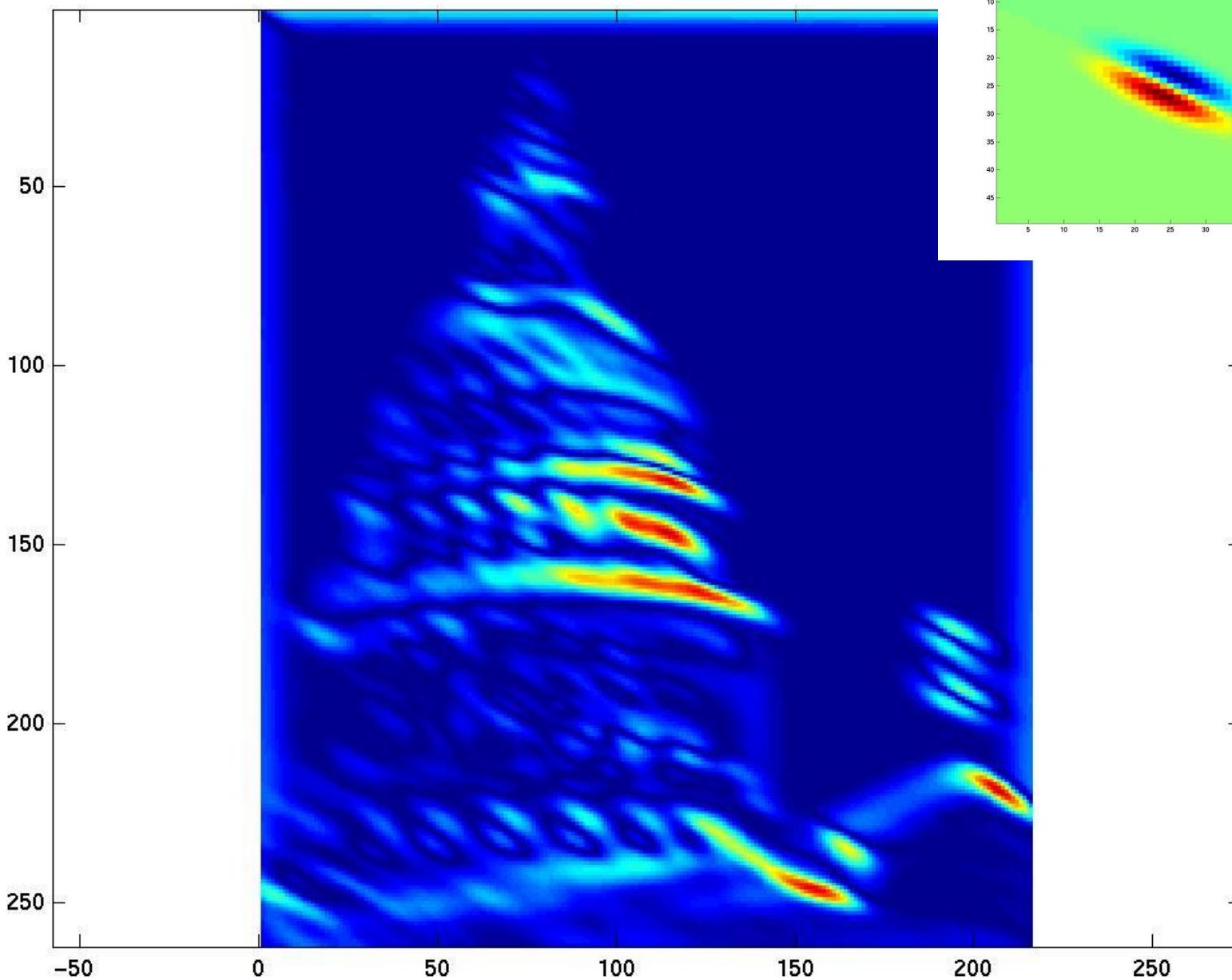


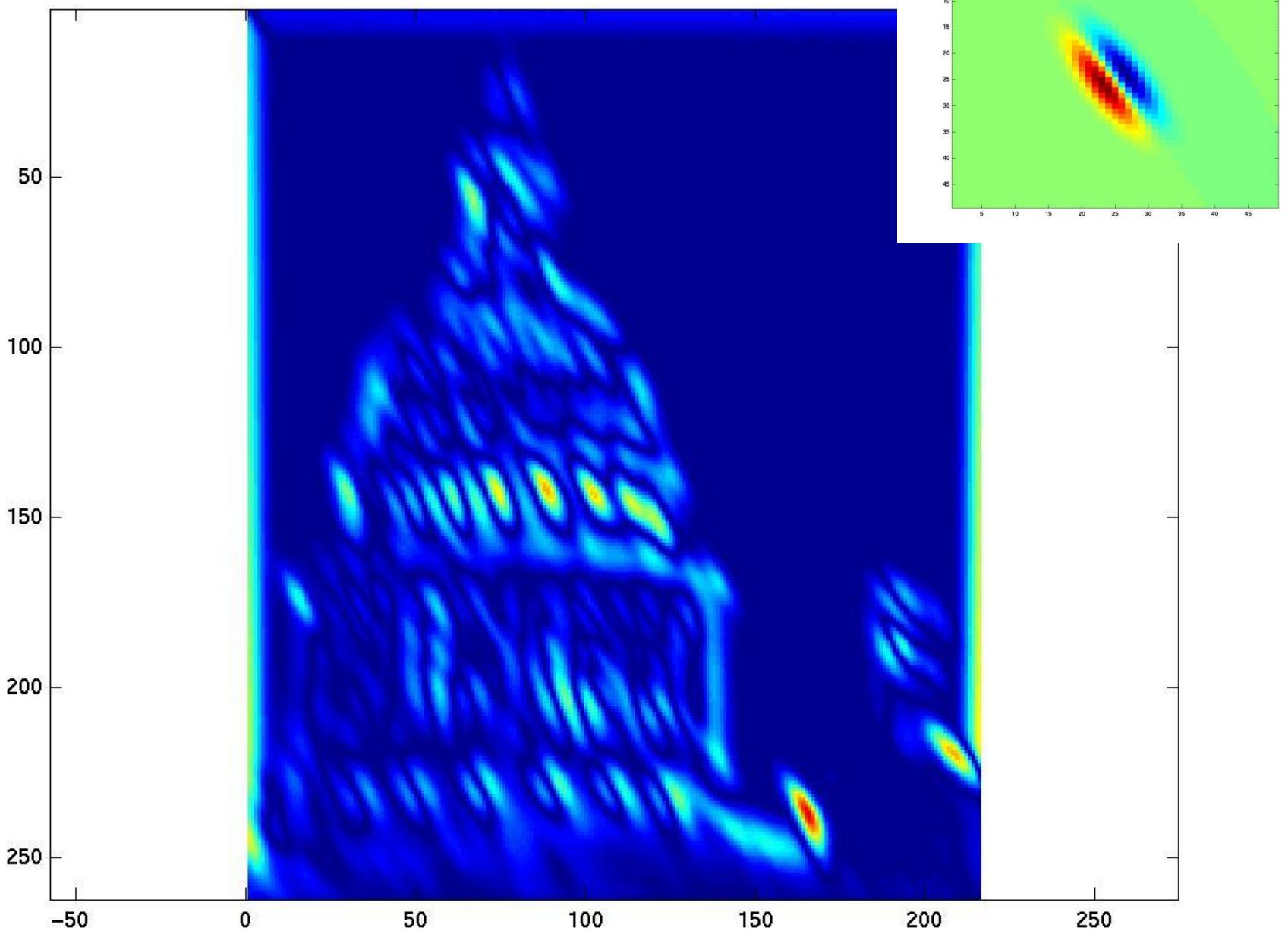


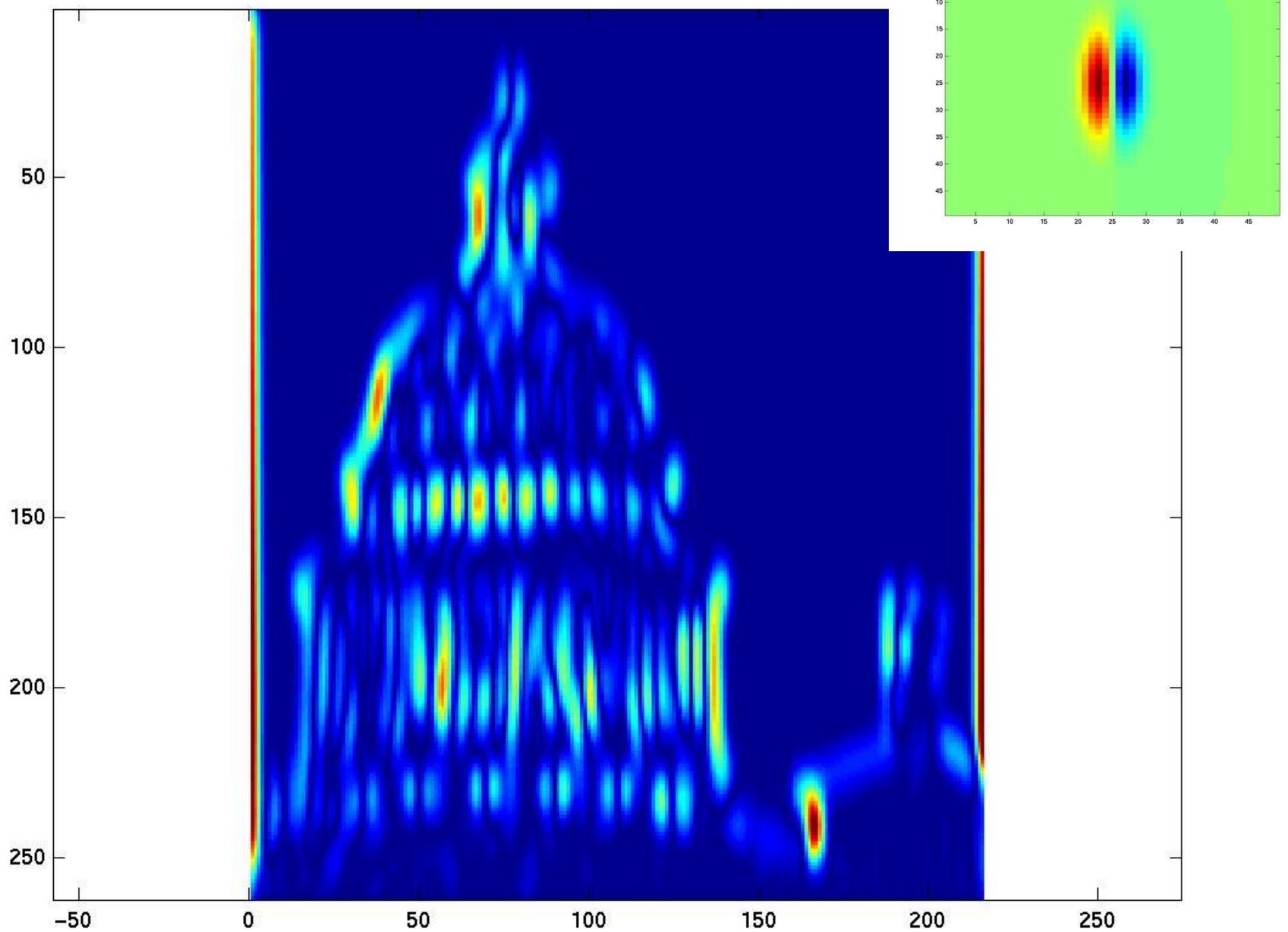


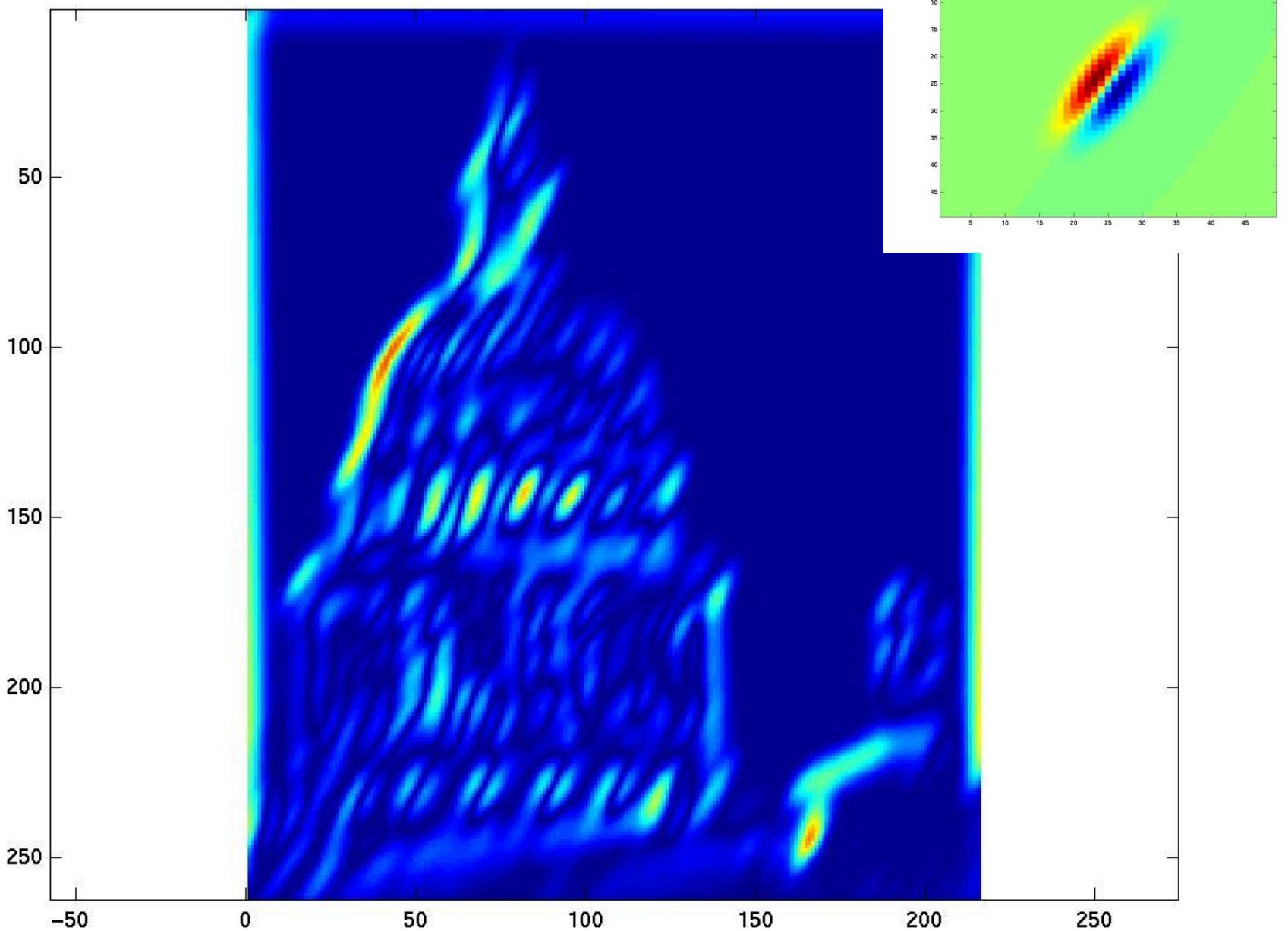


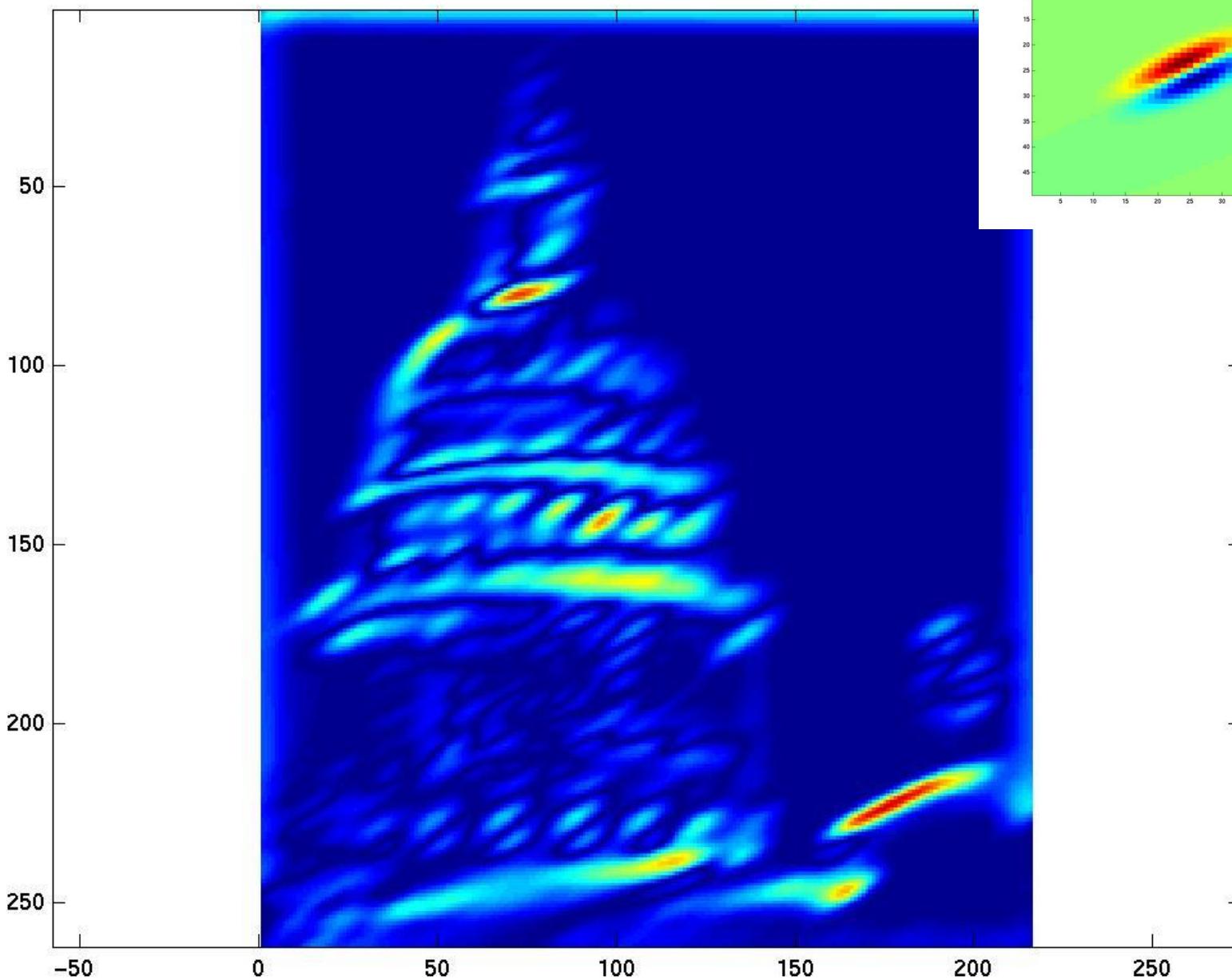


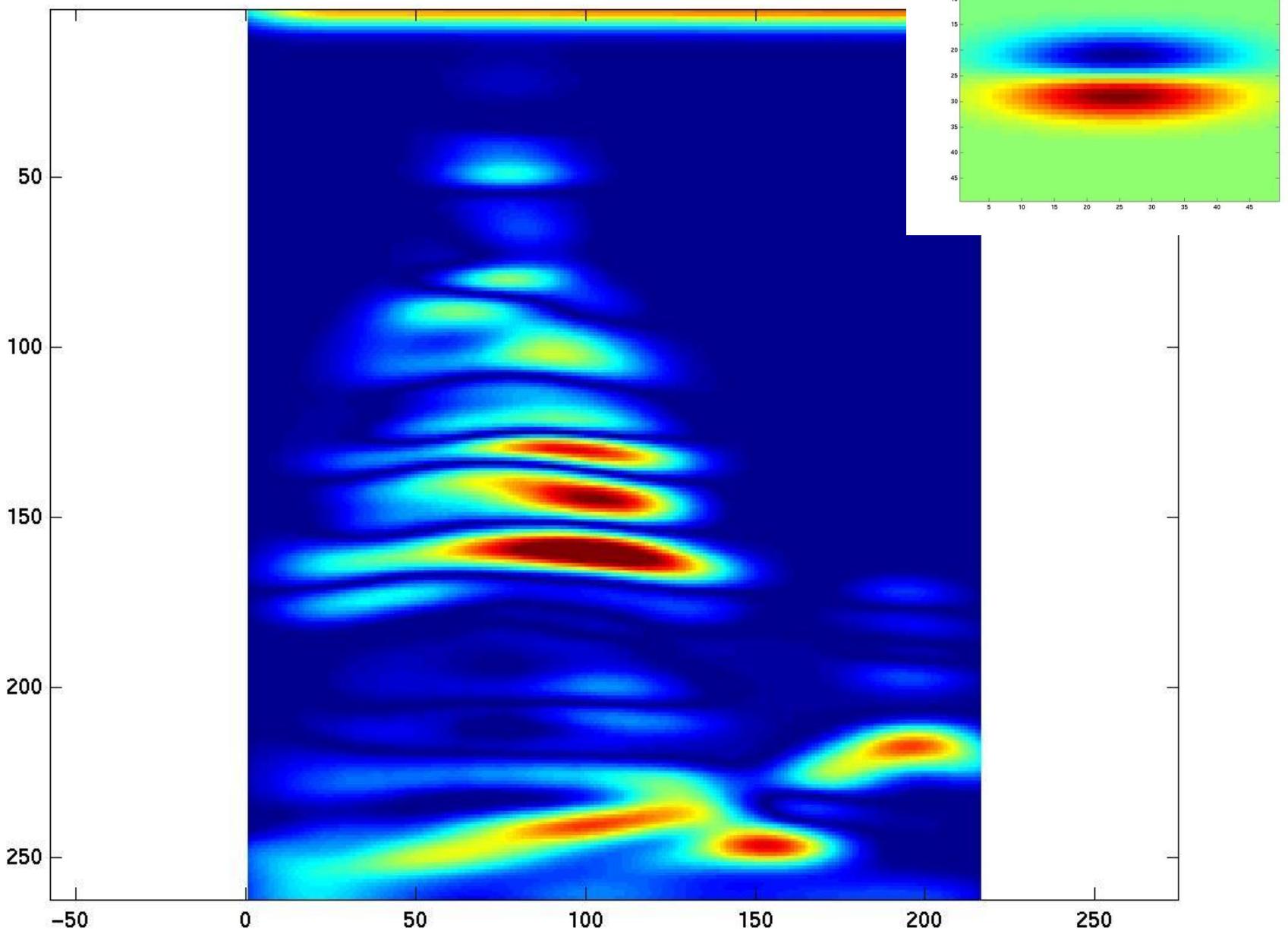


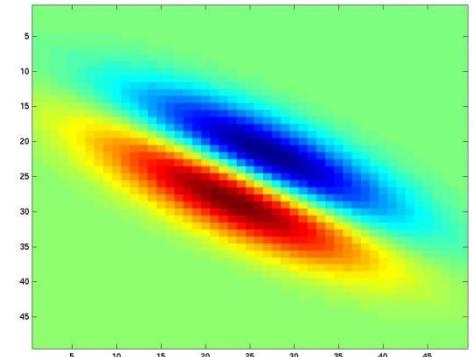
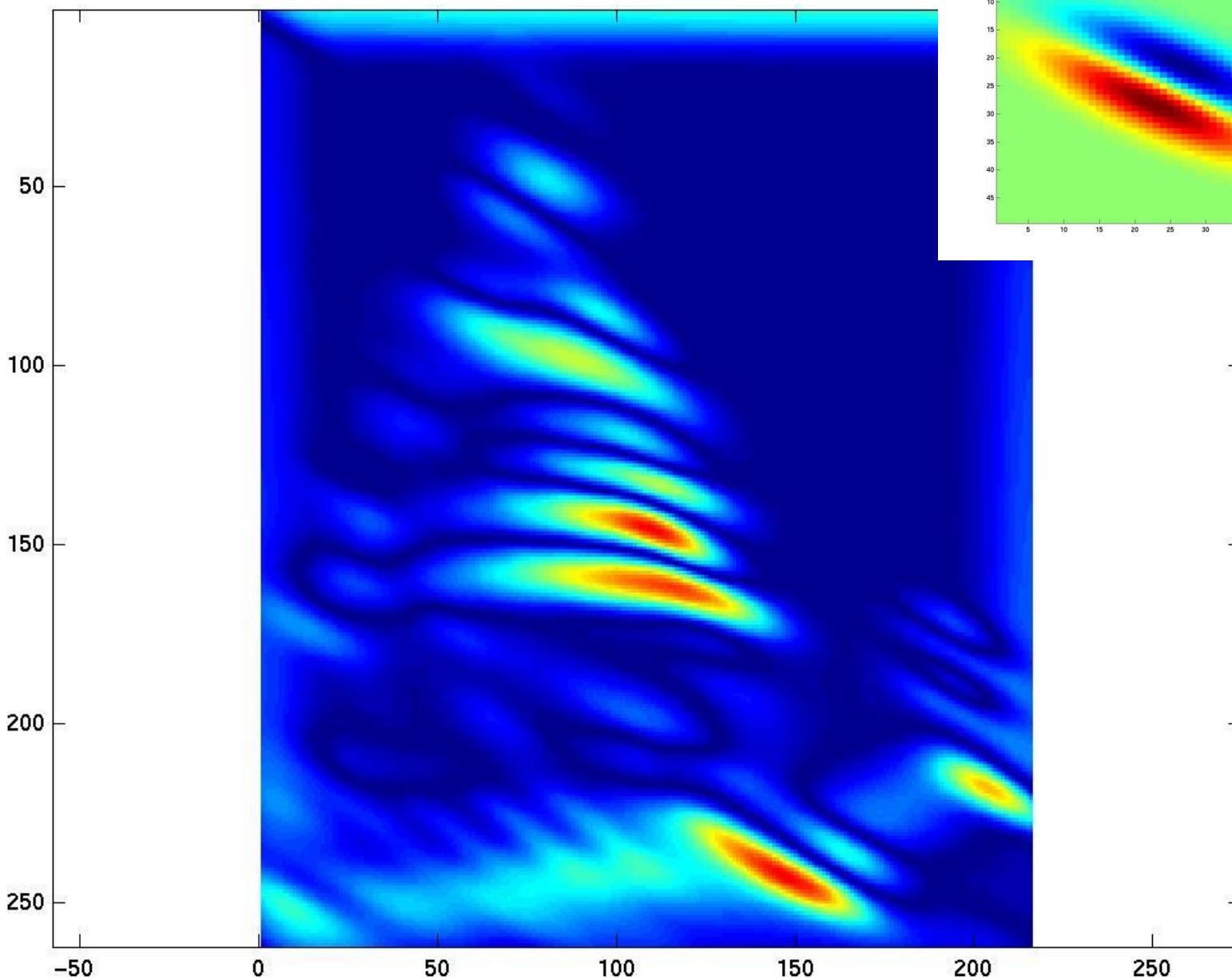


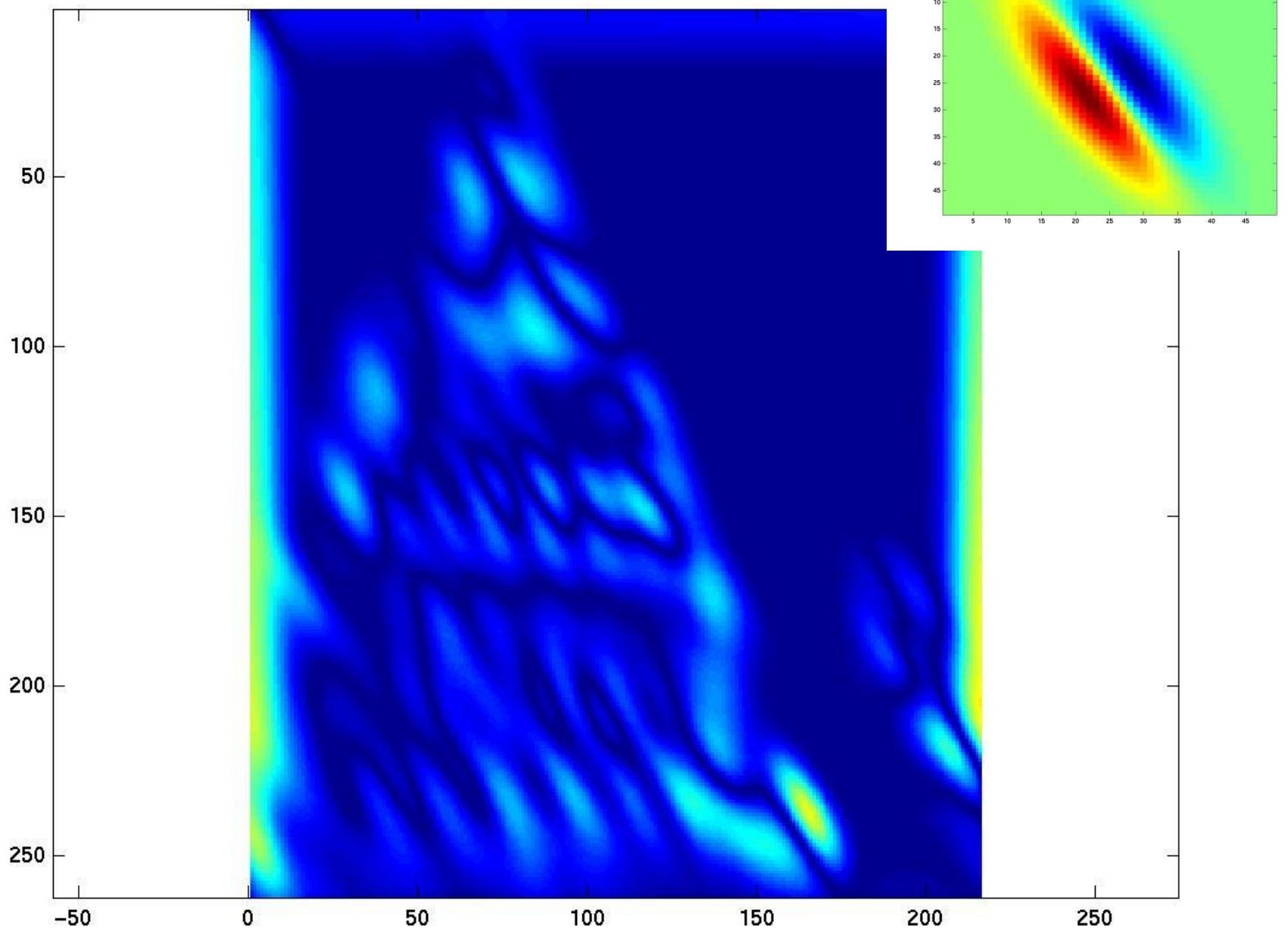


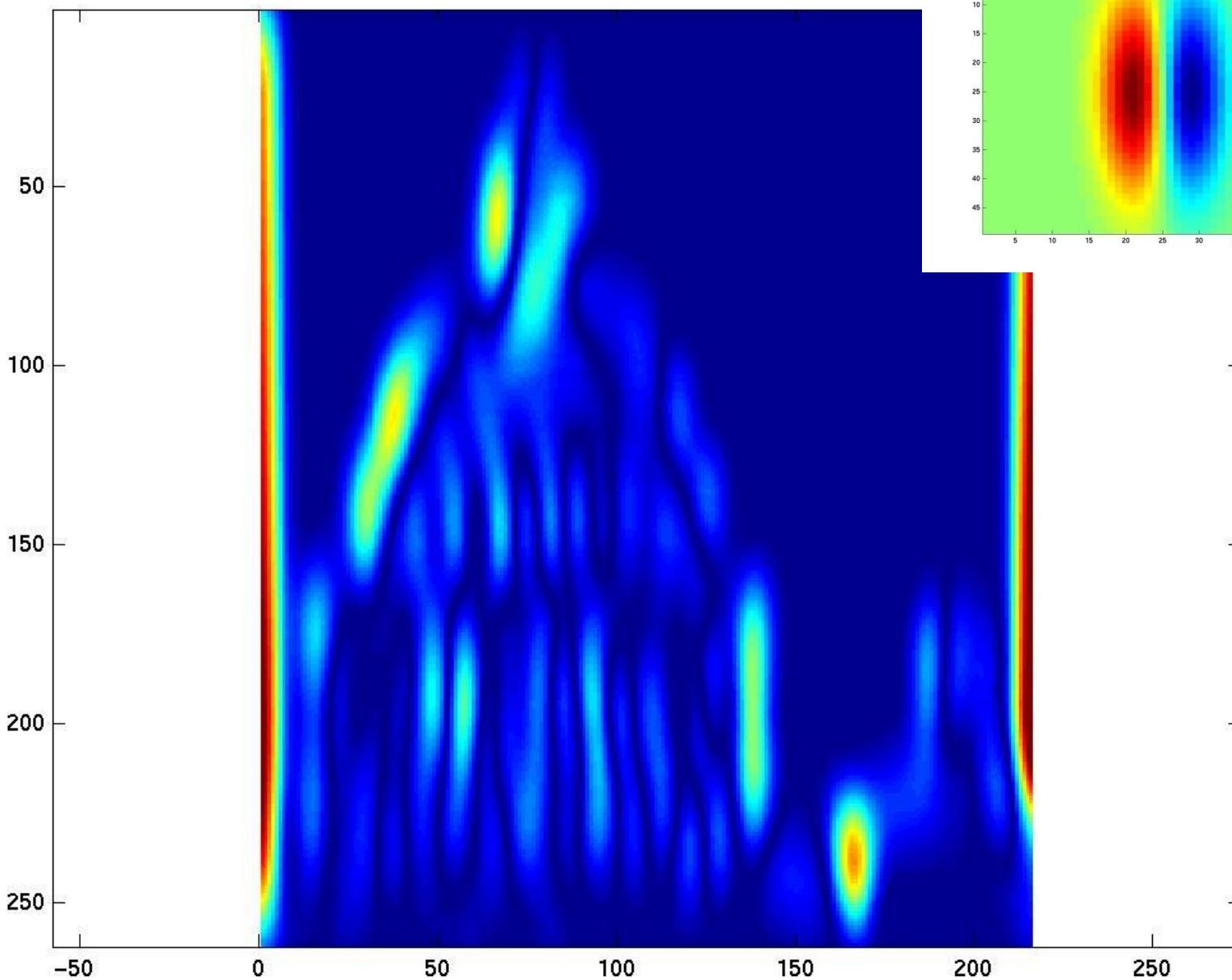


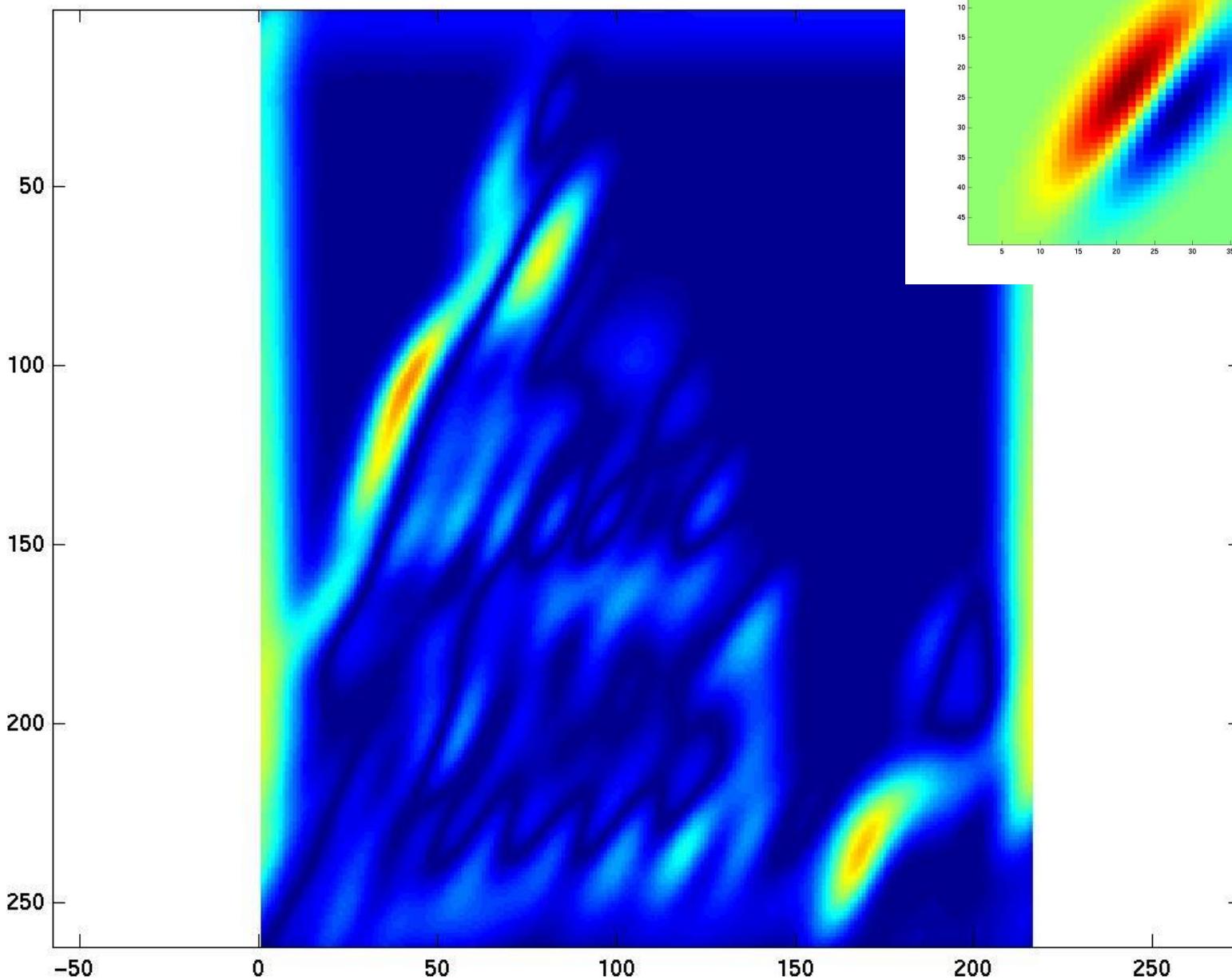


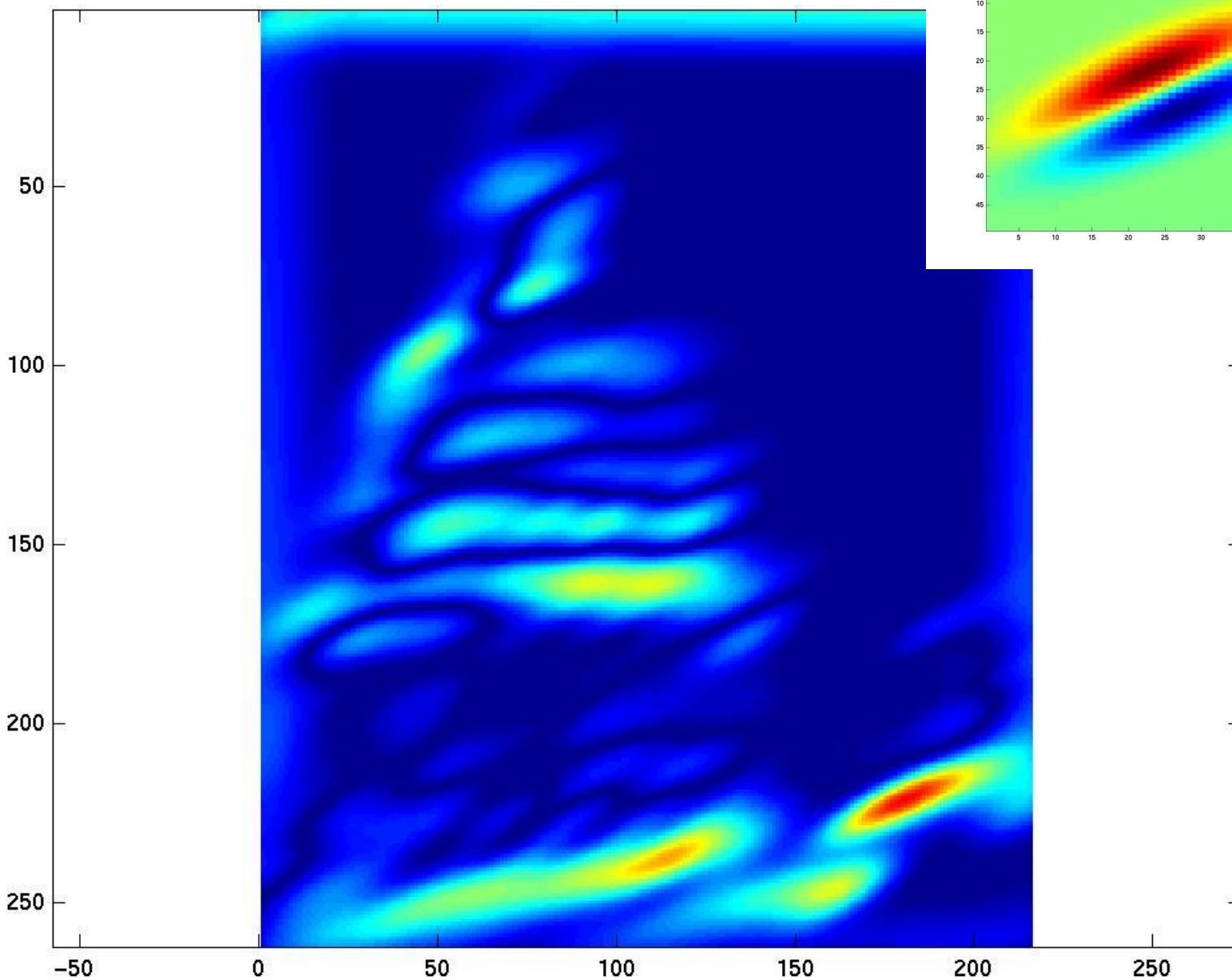


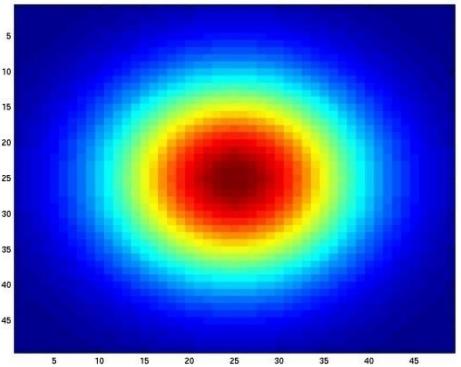
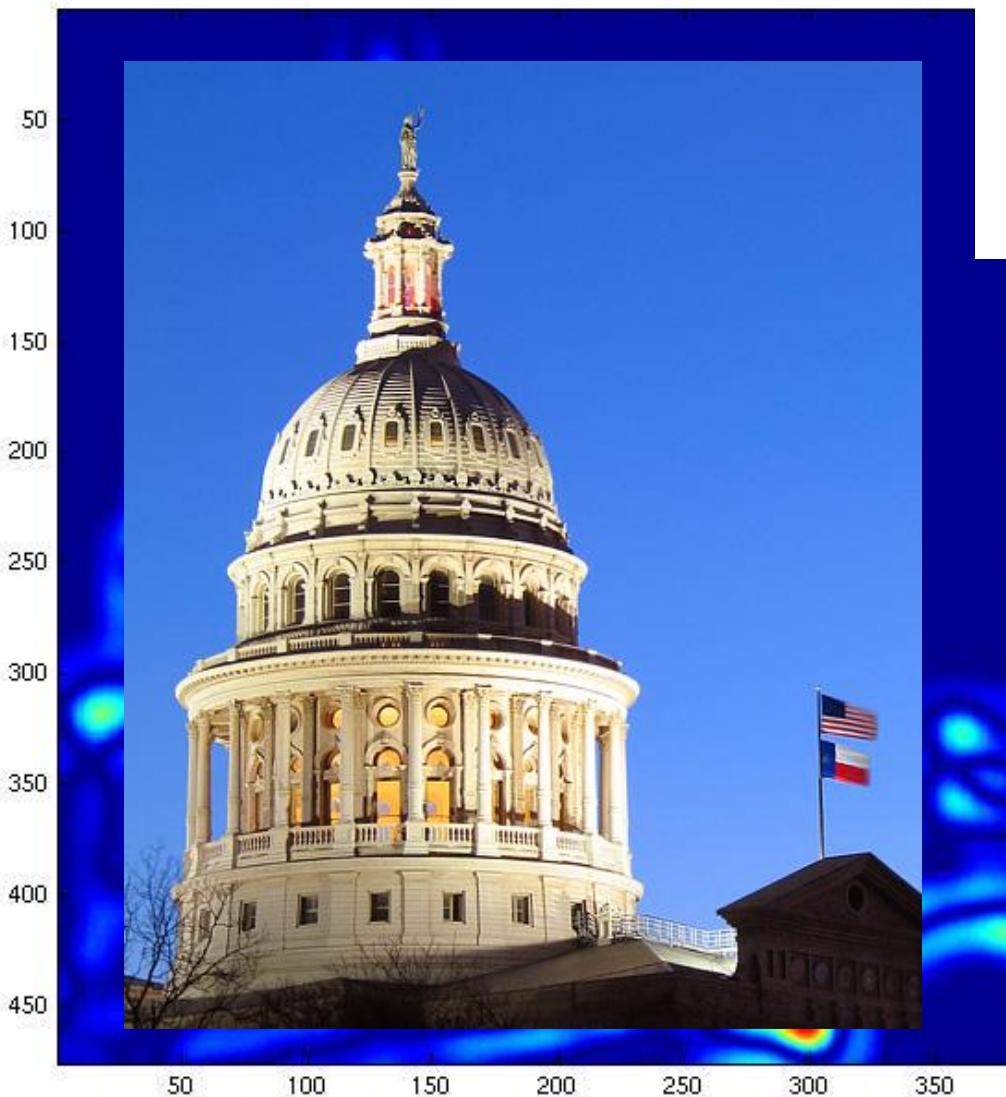








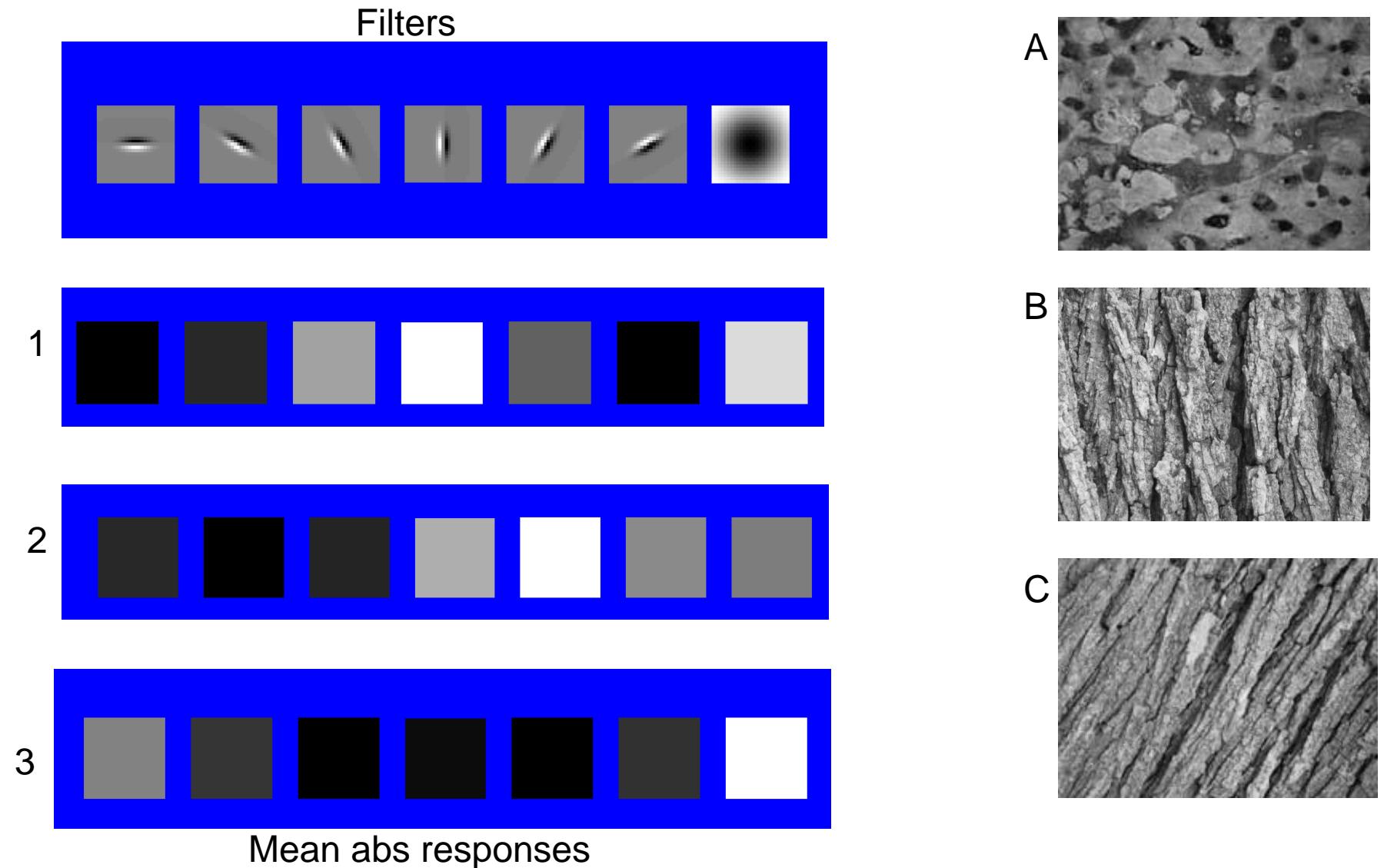




How can we represent texture?

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

Can you match the texture to the response?

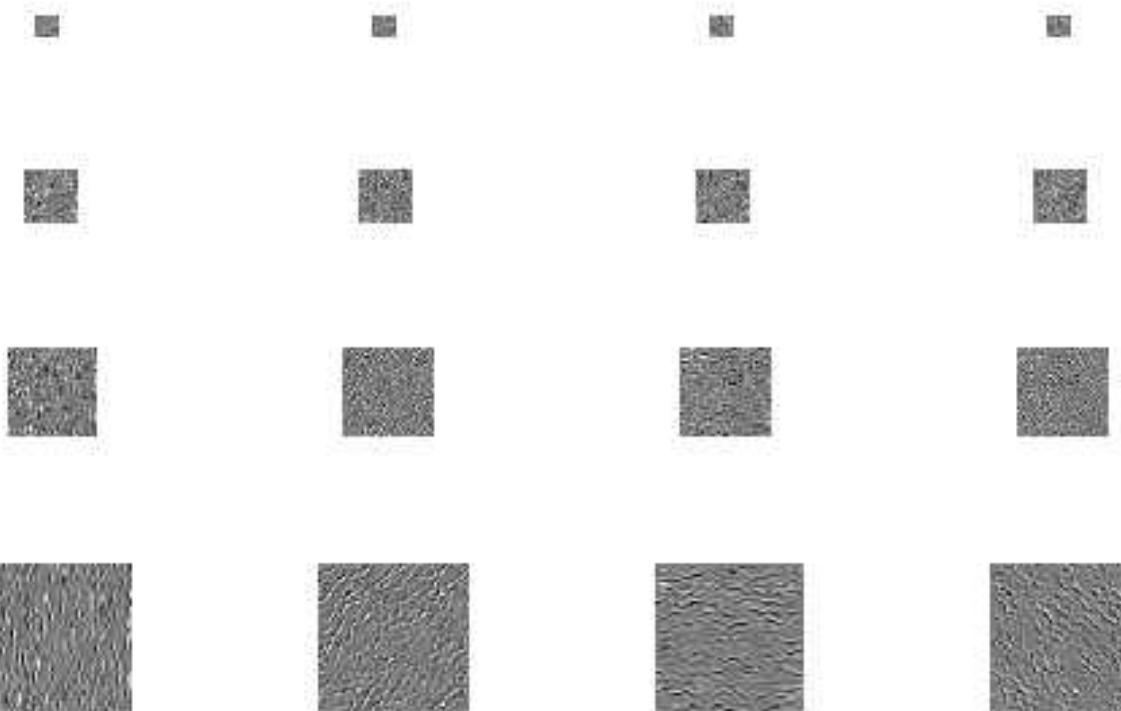
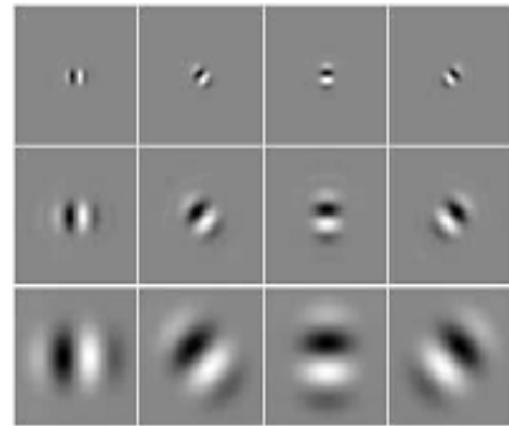


How can we represent texture?

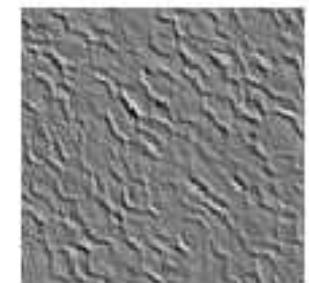
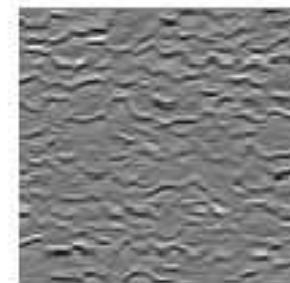
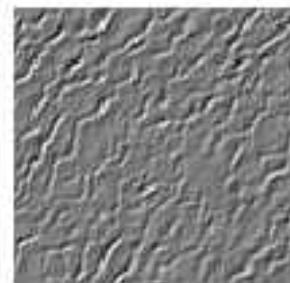
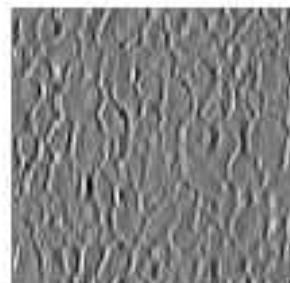
- Can be thought of as a single “orientation histogram”
- Idea 2: Marginal histograms of filter responses
 - one histogram per filter

Multi-scale filter decomposition

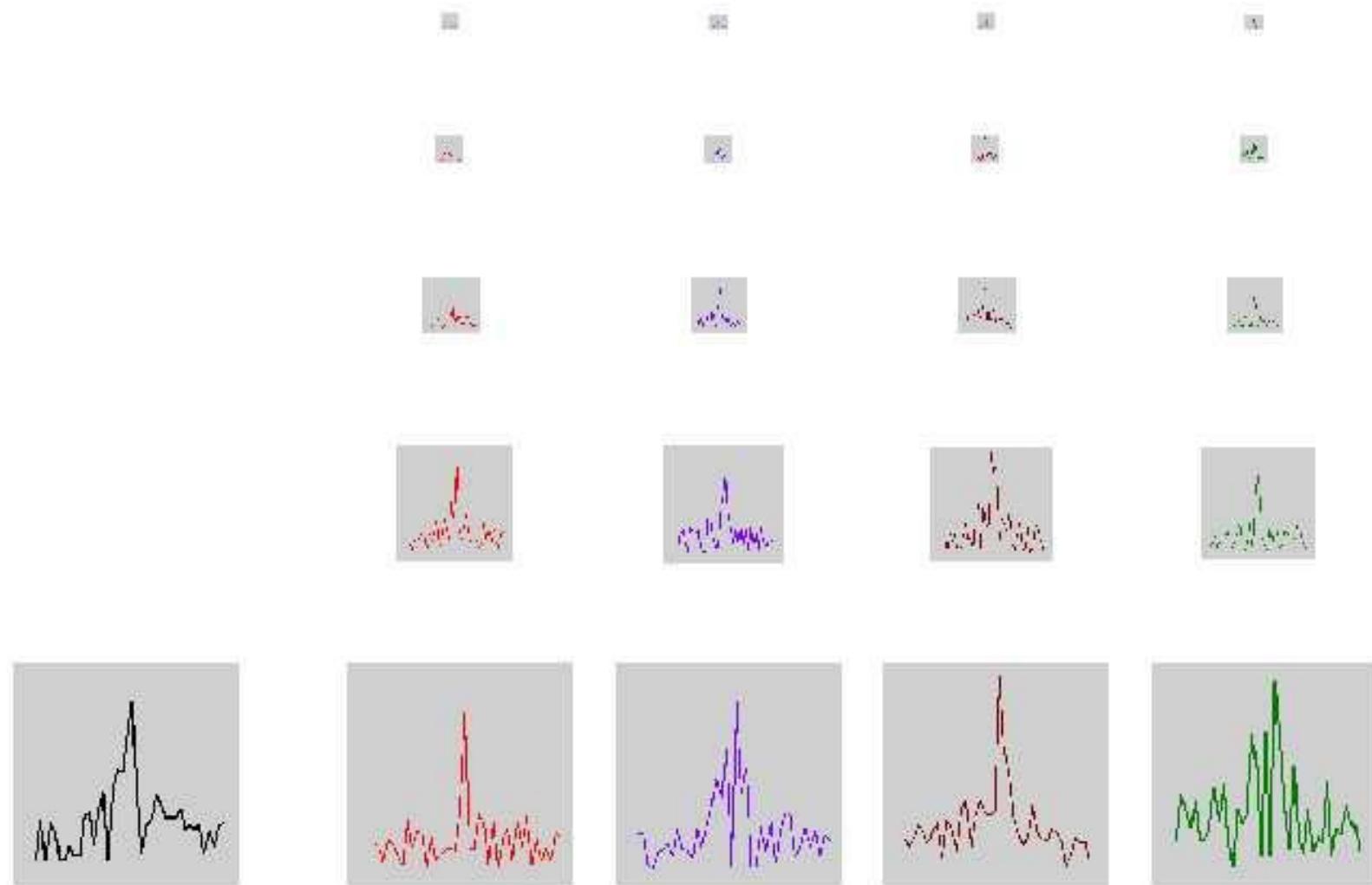
Filter bank



Input image



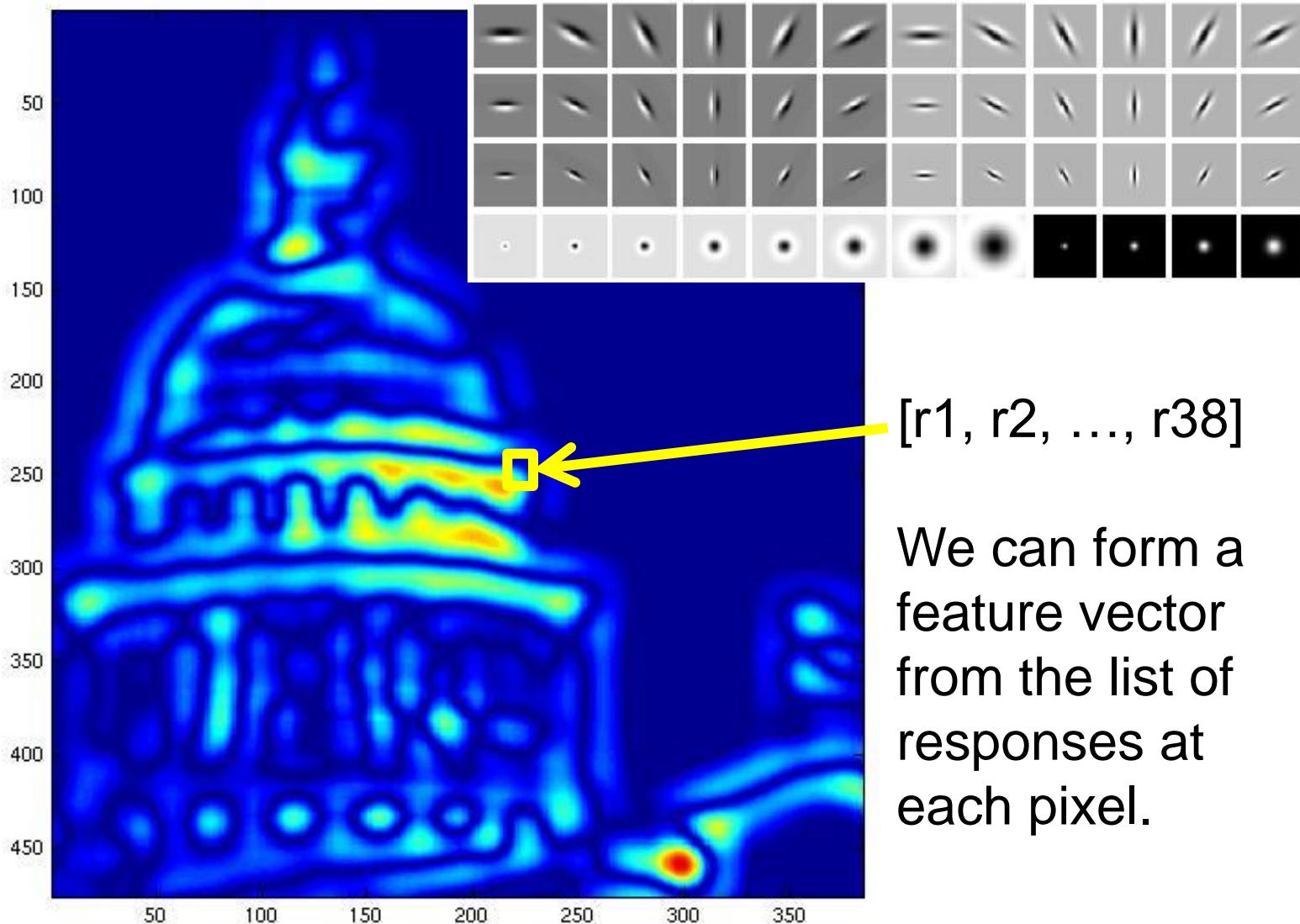
Filter response histograms



How can we represent texture?

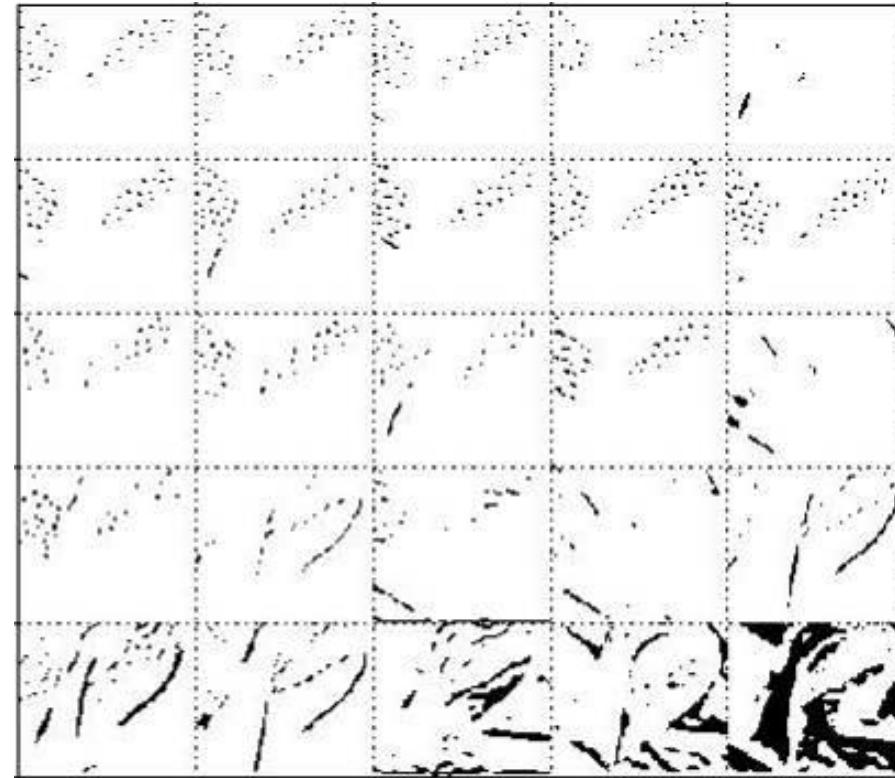
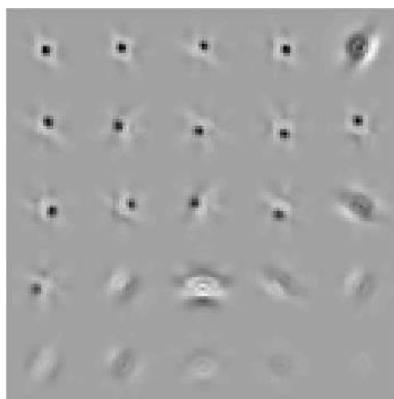
- Marginal filter response histograms don't talk to each other (in a direct way)
- Idea 3: Histograms of joint responses (textons)

Filter Response

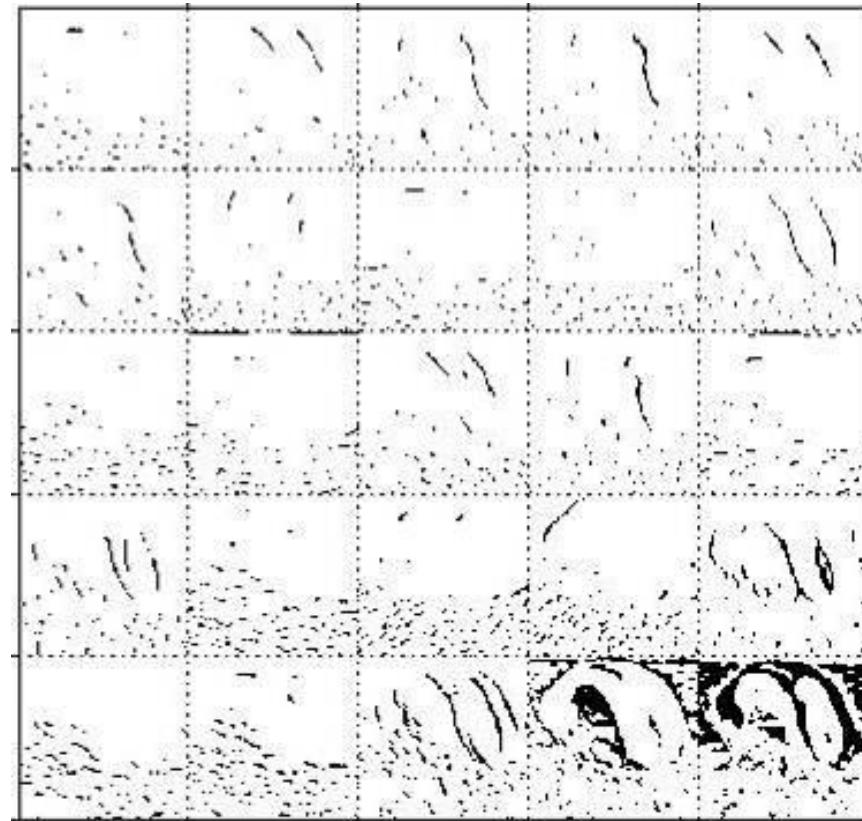
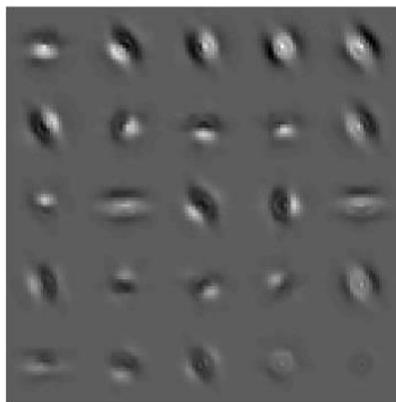


Textons (Malik et al, IJCV 2001)

- Cluster vectors of filter responses



Textons (cont.)



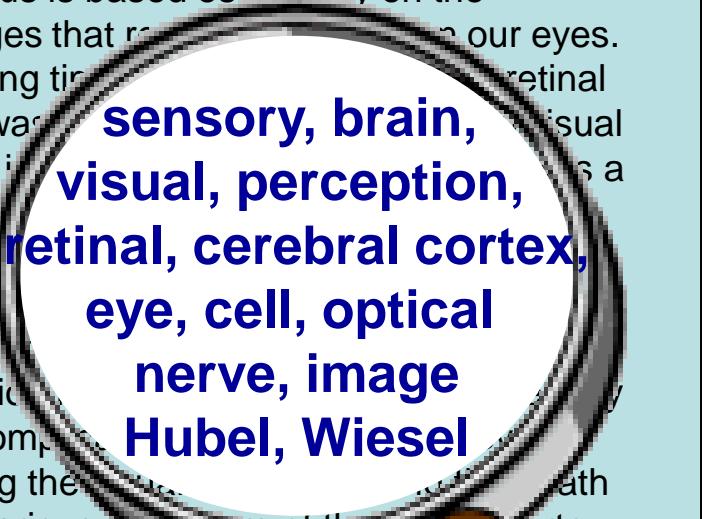
Object

Bag of ‘words’



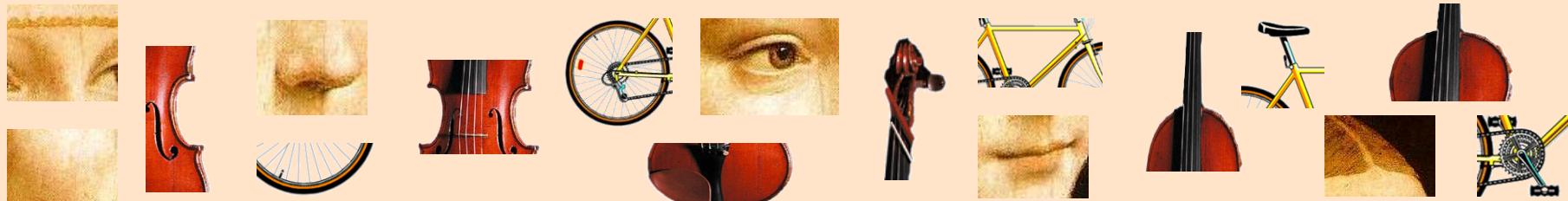
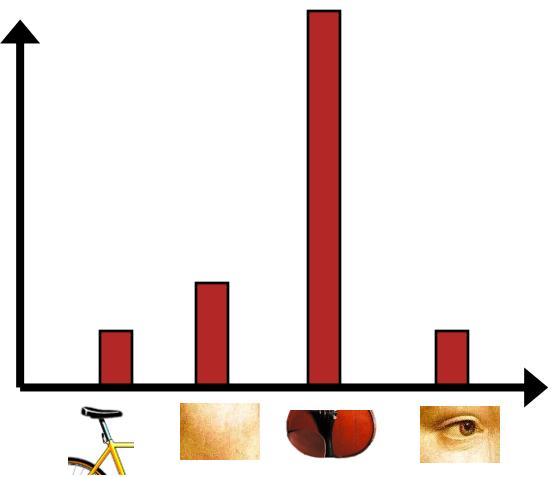
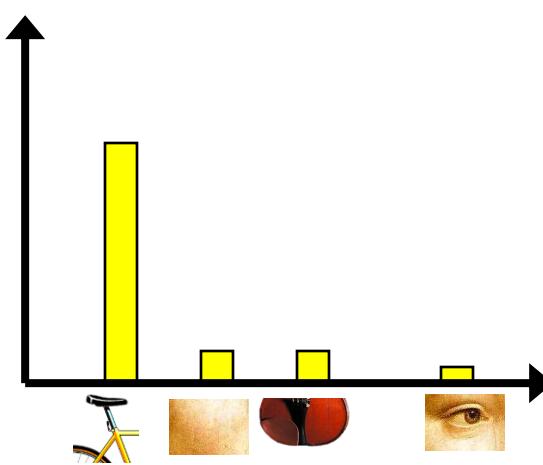
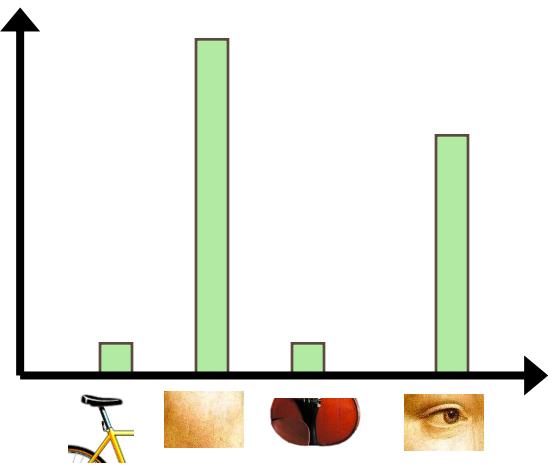
Analogy to documents

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 reach us through our eyes. For a long time it was believed that the retinal image was processed in the visual centers in the brain. In 1960, Hubel and Wiesel discovered that the visual system is a complex system of nerve fibers that follow a specific path to the various centers of the cerebral cortex. Hubel and Wiesel have demonstrated that the message about the image falling on the retina undergoes top-down analysis in a system of nerve cells stored in columns. In this system each column has its specific function and is responsible for a specific detail in the pattern of the retinal image.

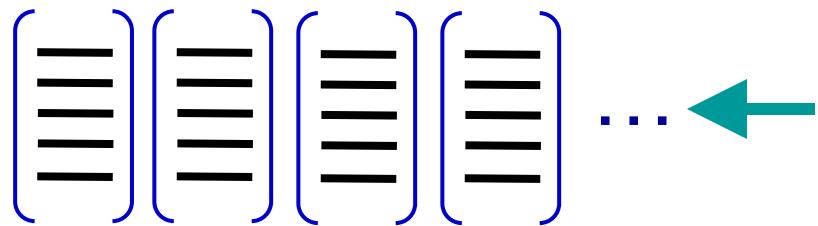


China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$660bn. The move will annoy the US, which wants China's central bank to deliberately devalue the yuan to ease its trade surplus. The Chinese government also needs to encourage domestic demand so that the country can buy more from the US. China has been allowed to let the yuan against the dollar rise, and permitted it to trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

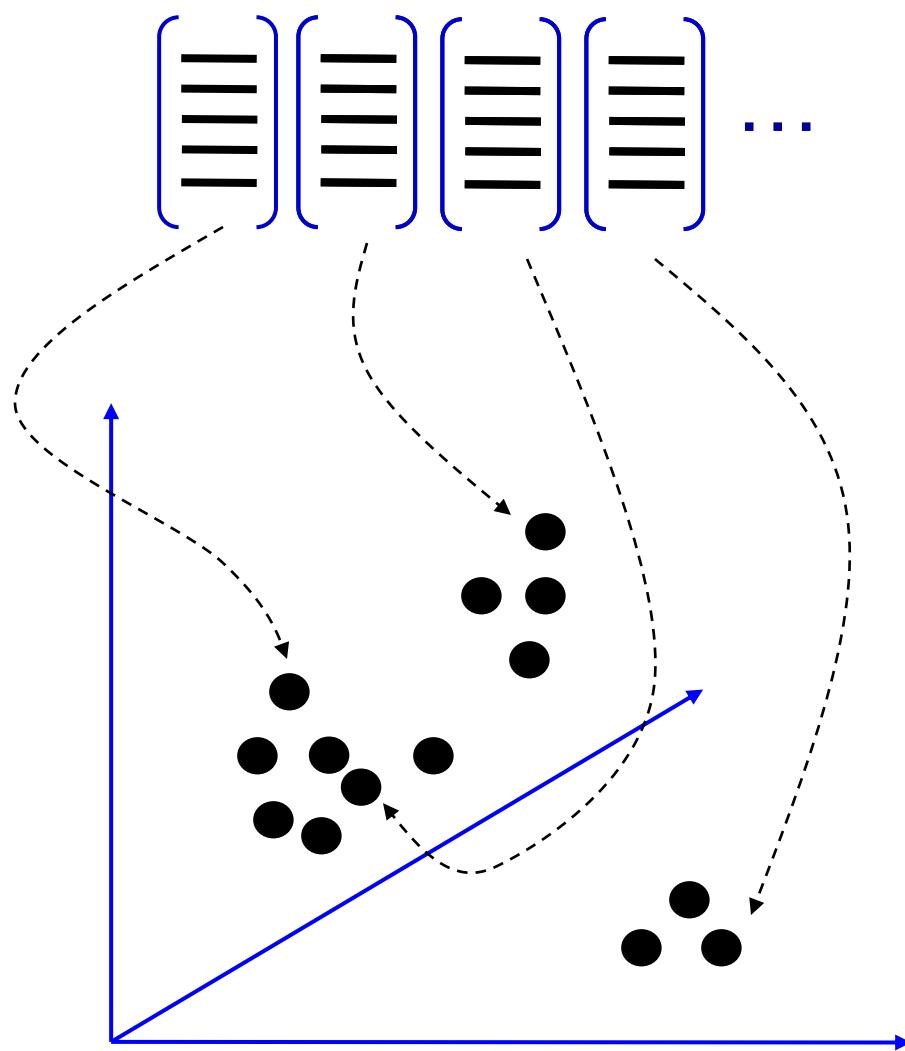




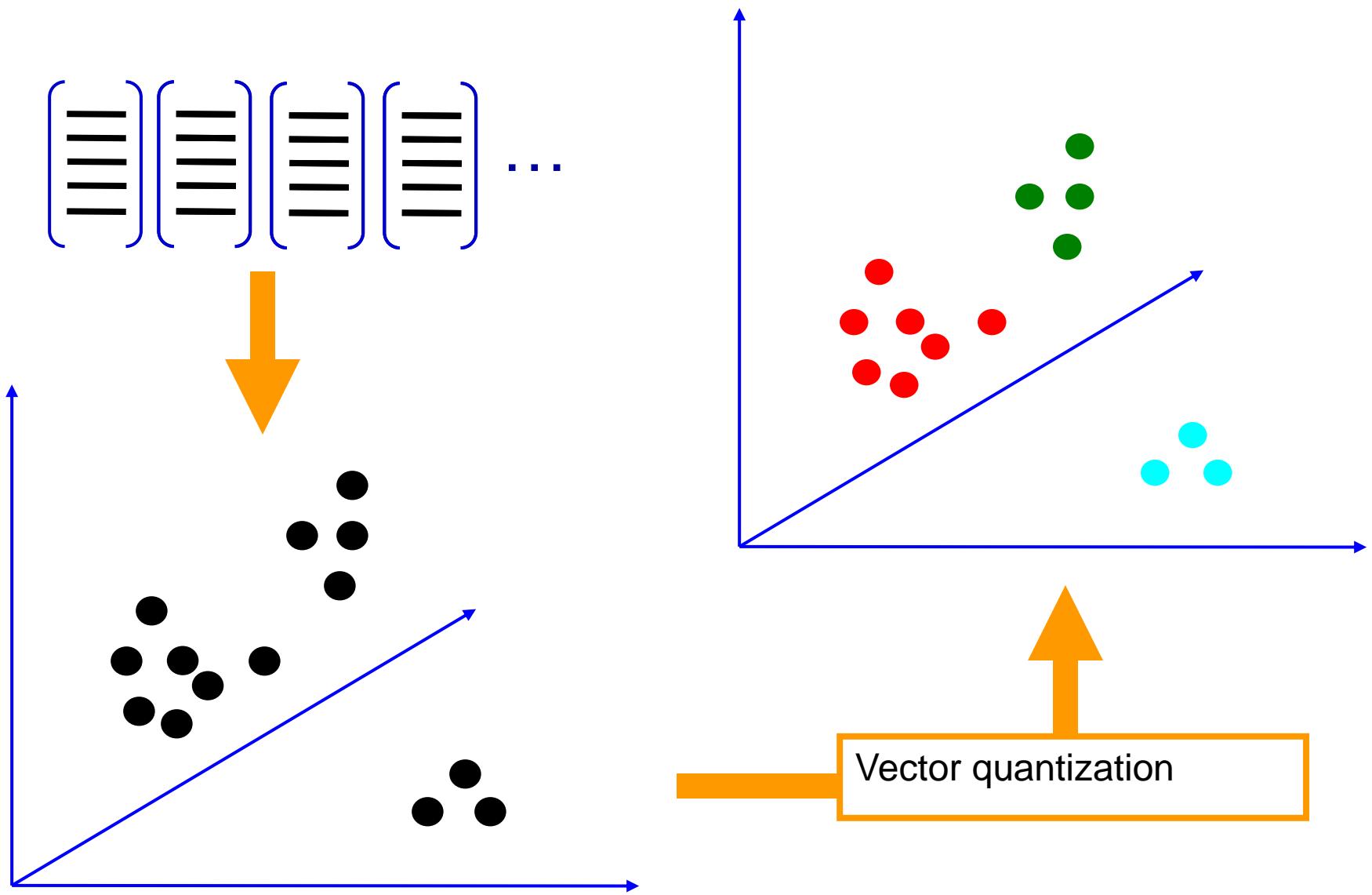
Patch Features



dictionary formation



Clustering (usually k-means)

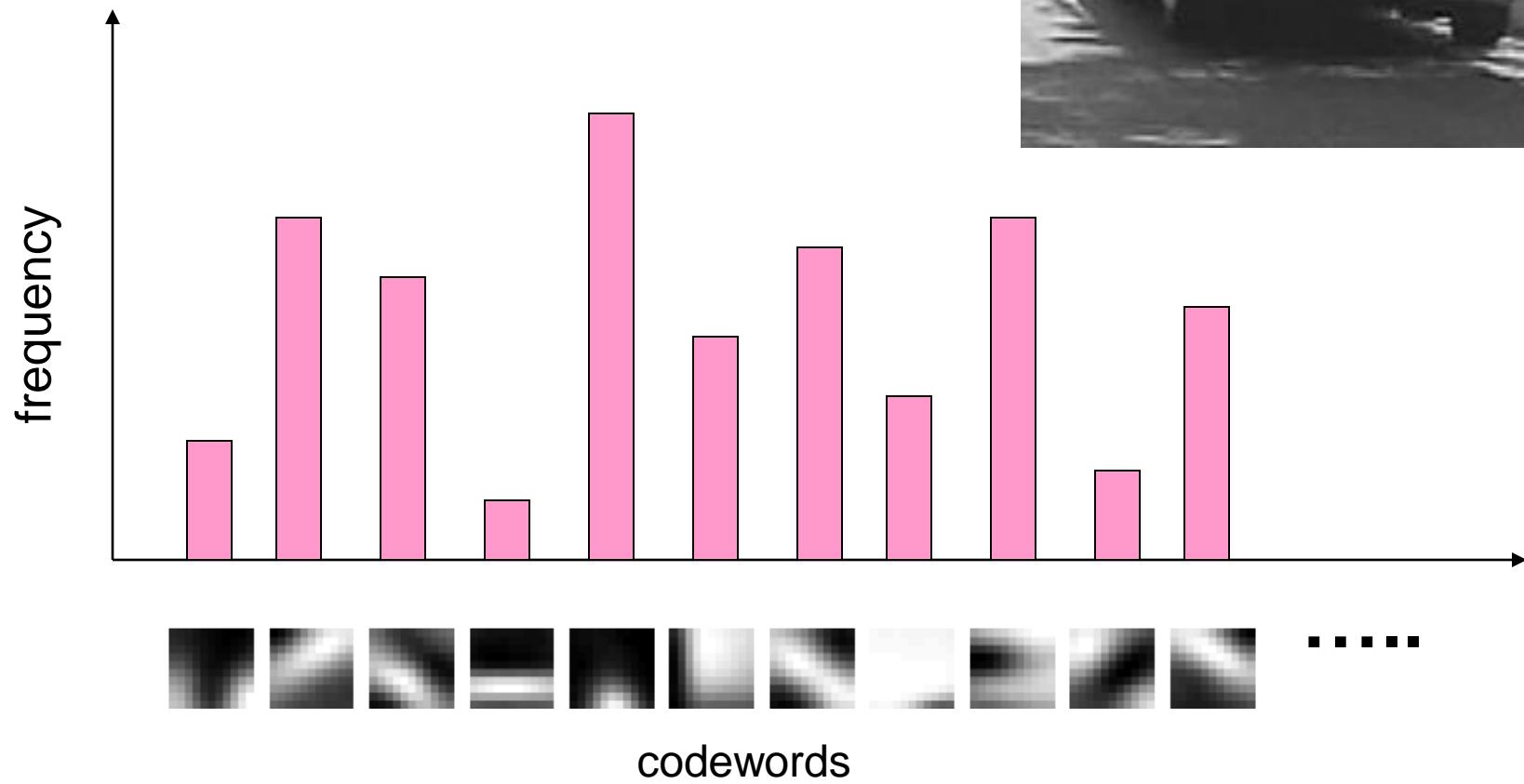


Feature Representation

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



Scene Classification (Renninger & Malik)

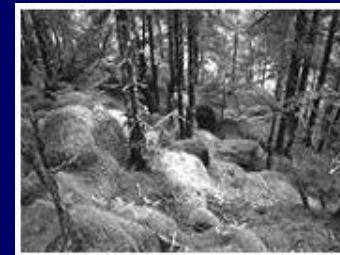
beach



mountain



forest



city



street



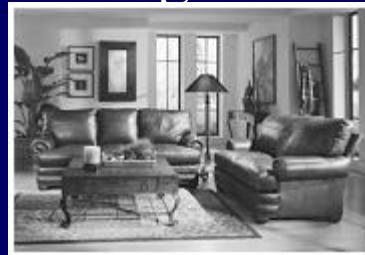
farm



kitchen



livingroom



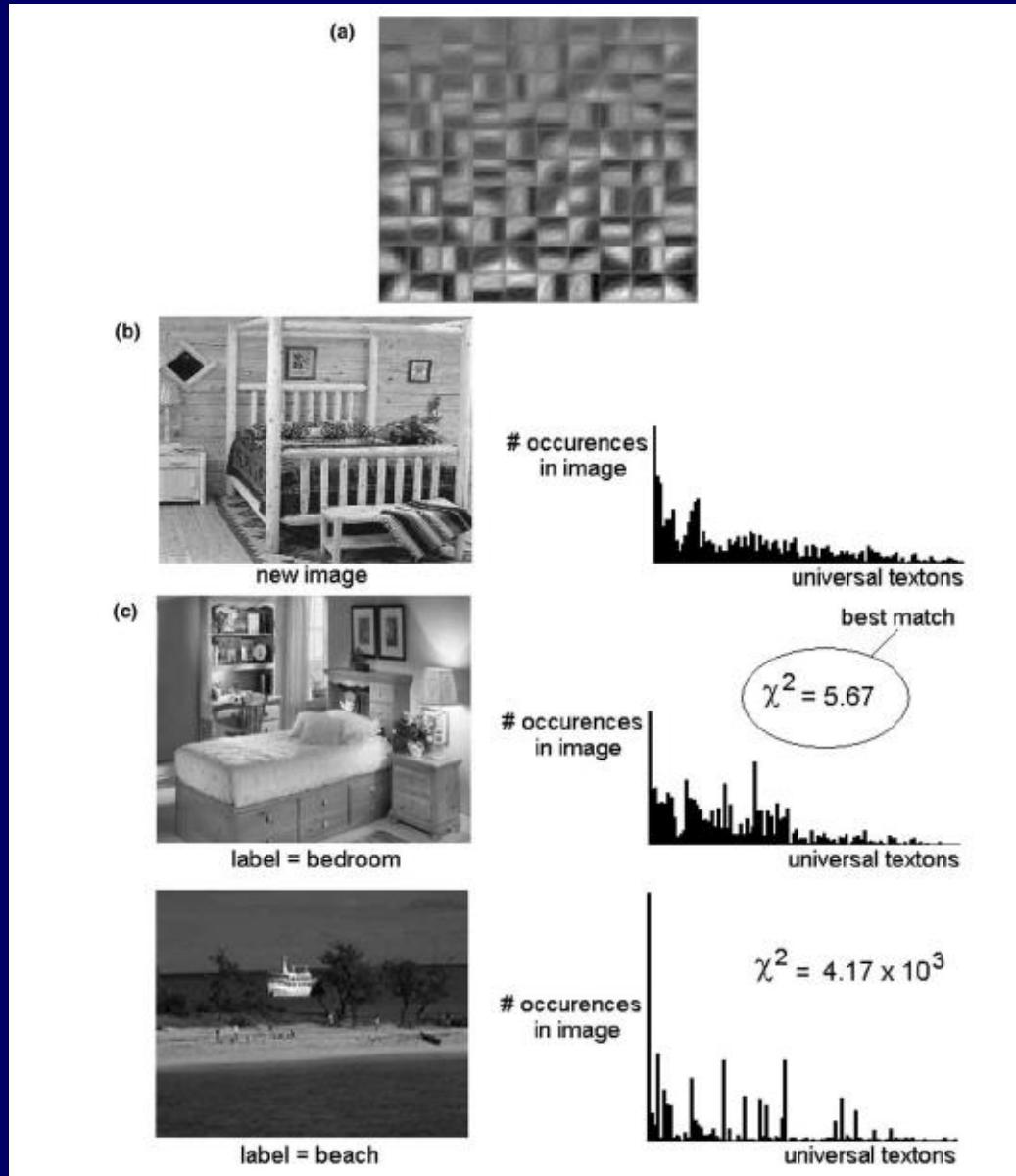
bedroom



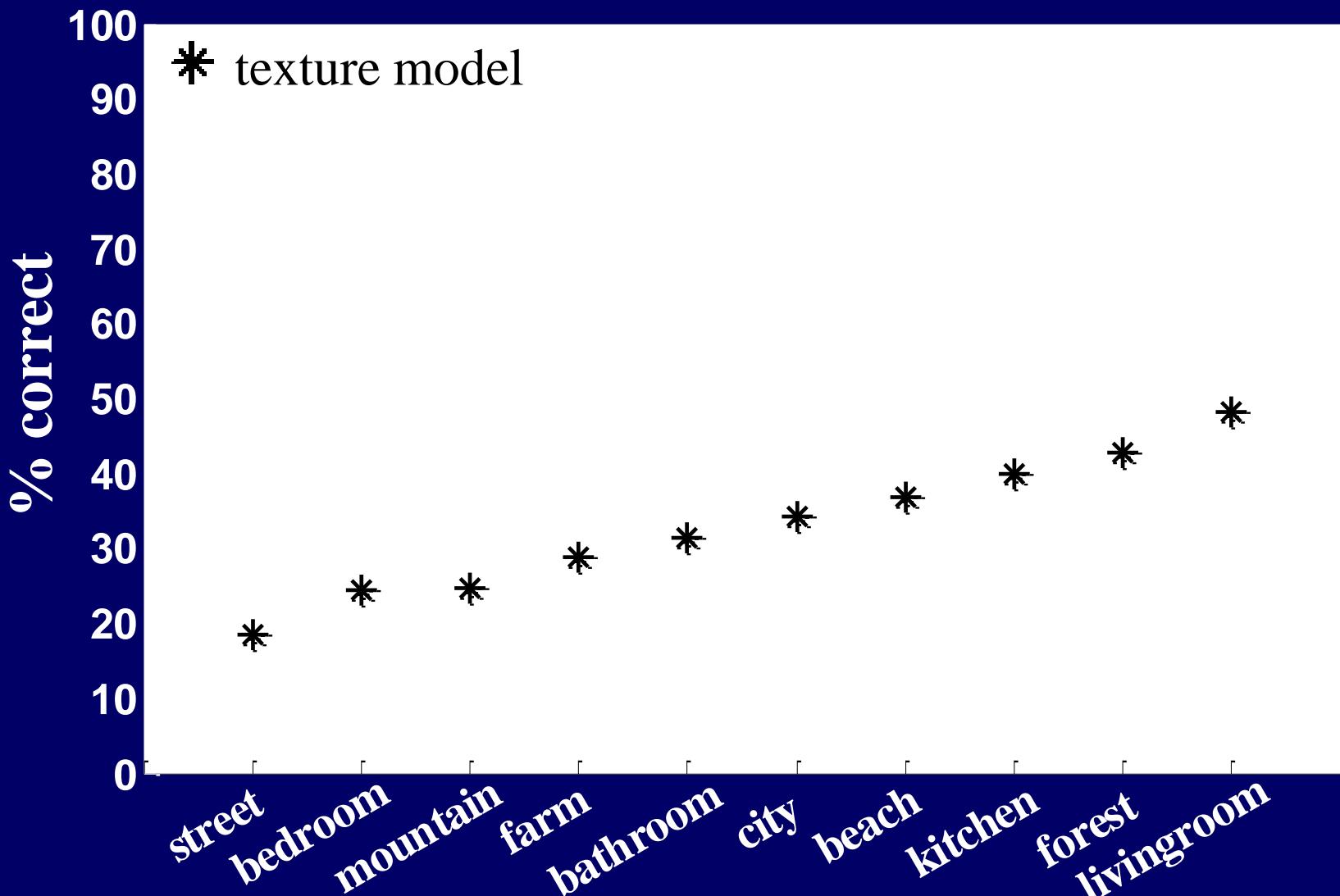
bathroom



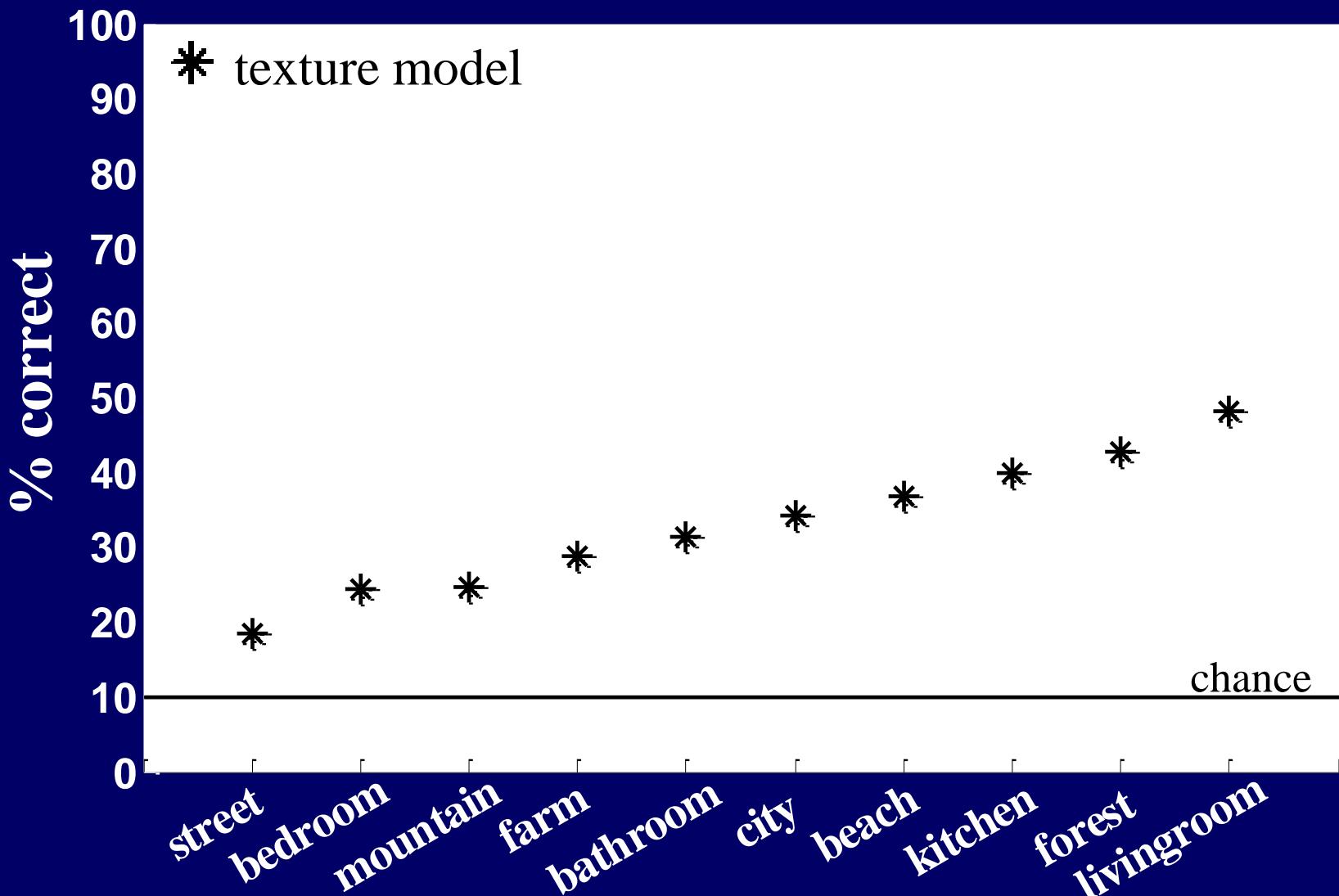
Texton Histogram Matching



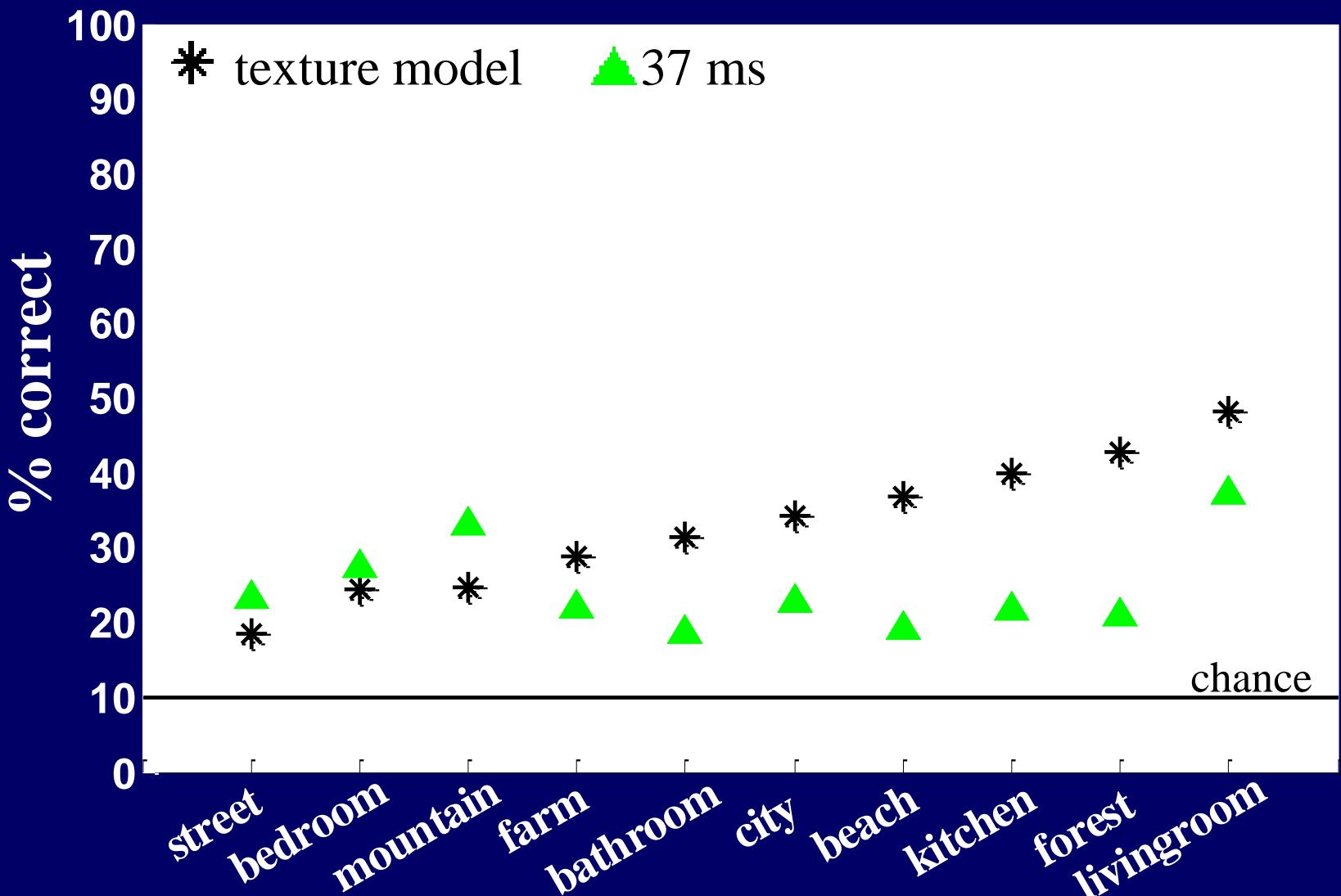
Discrimination of Basic Categories



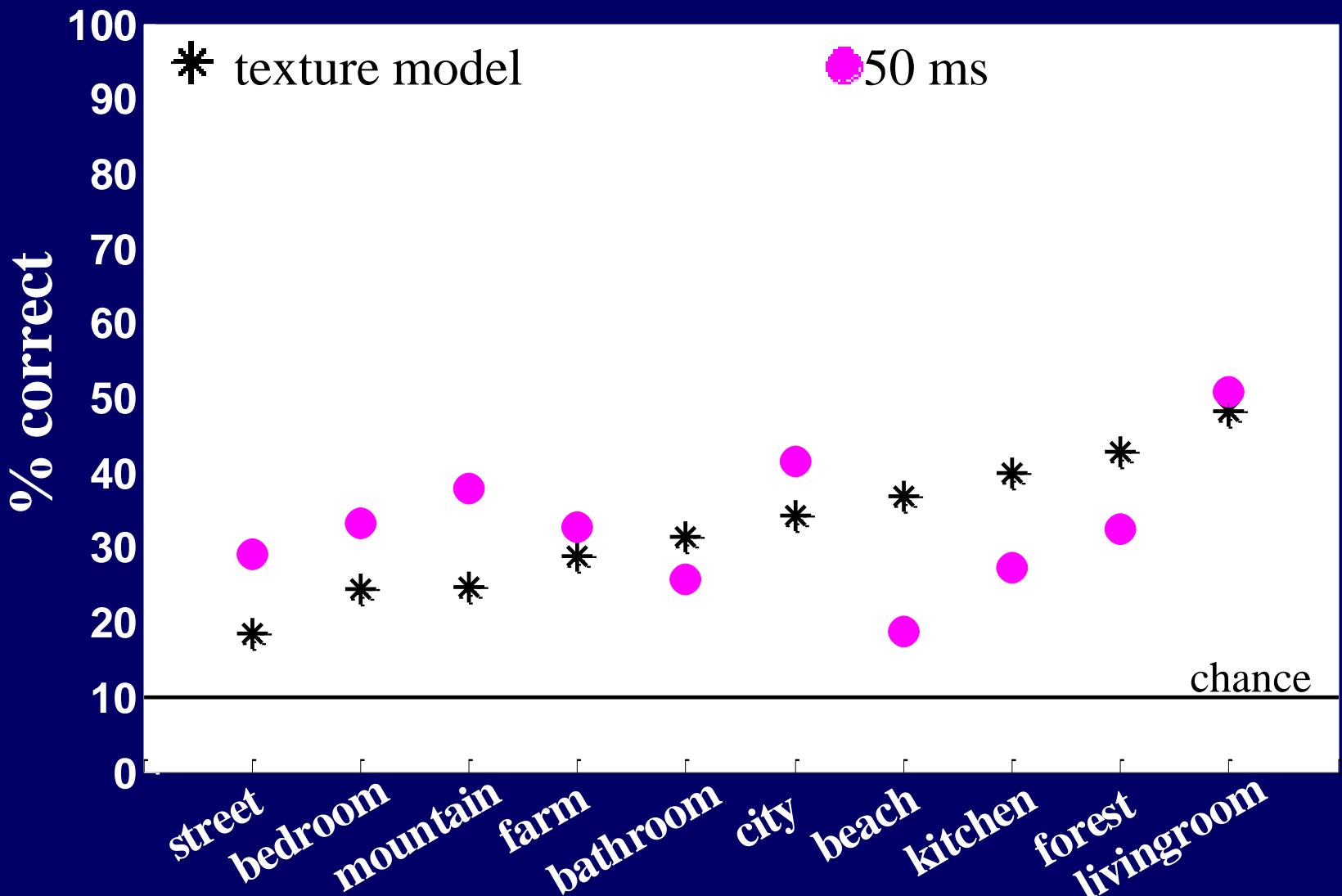
Discrimination of Basic Categories



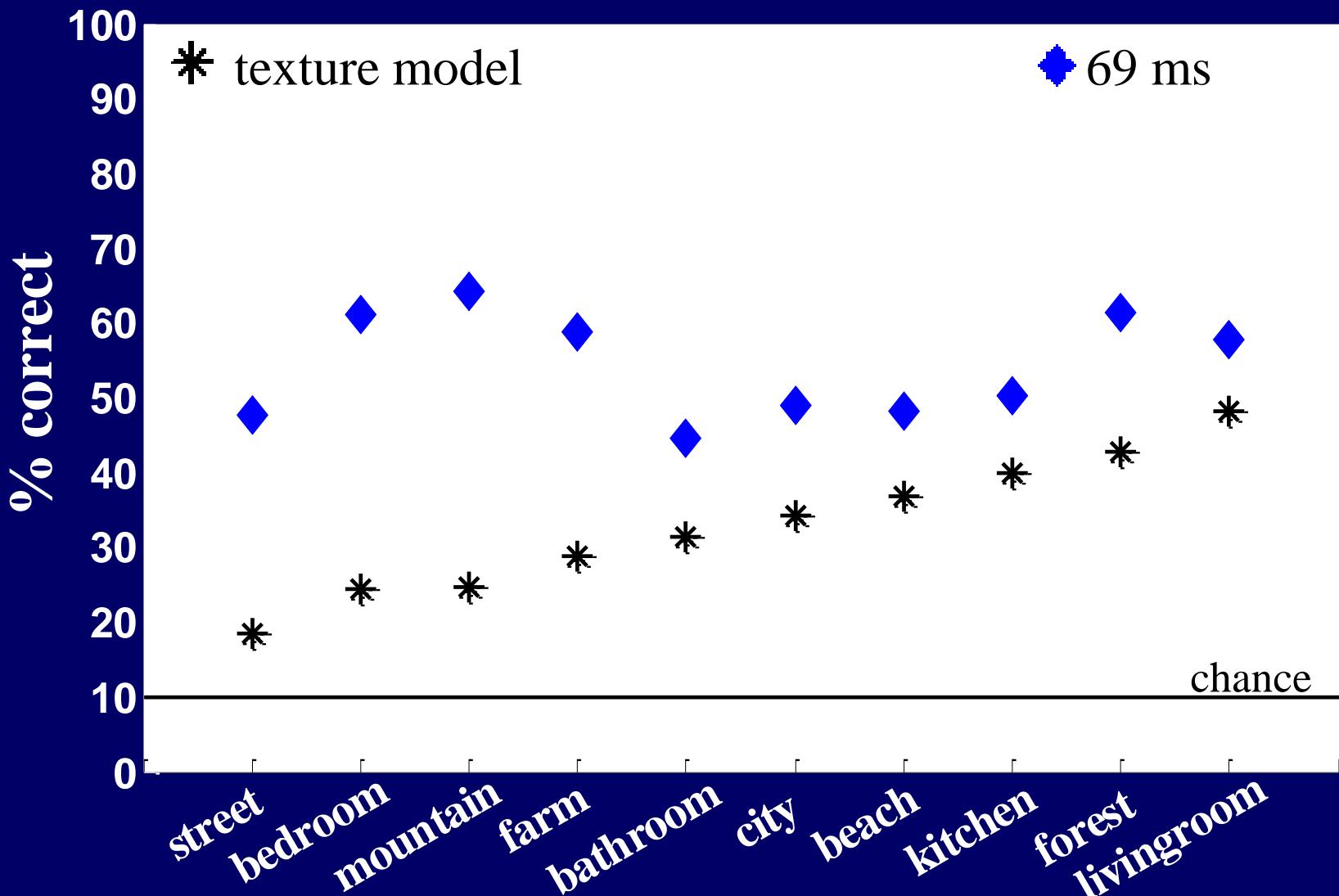
Discrimination of Basic Categories



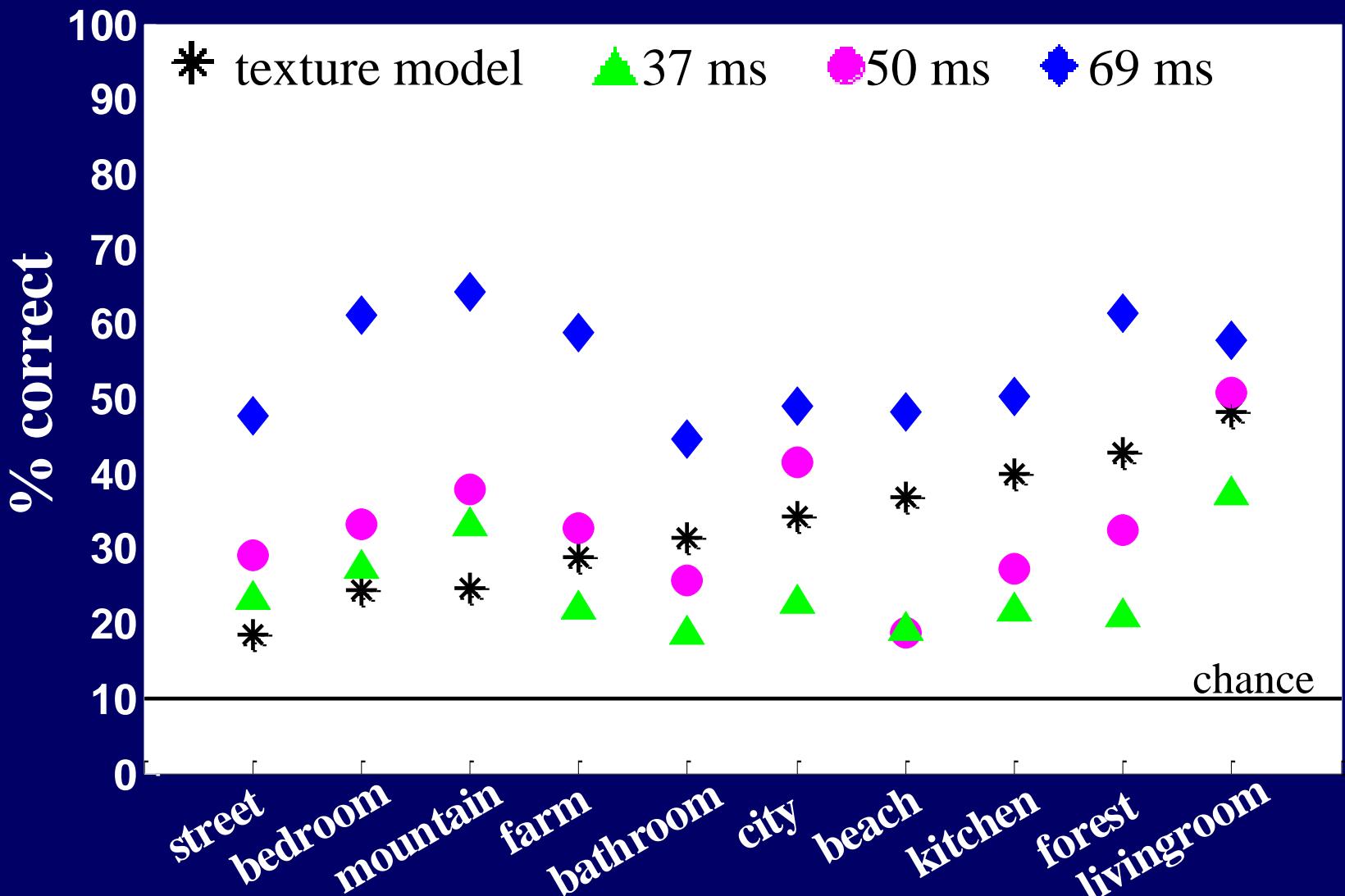
Discrimination of Basic Categories



Discrimination of Basic Categories



Discrimination of Basic Categories



Scene Recognition using Texture



Why these filters?

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

Bruno A. Olshausen¹ and David J. Field

$$E = -[\text{preserve information}] - \lambda[\text{sparserness of } a_i], \quad (2)$$

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:

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

Learned filters

a.

