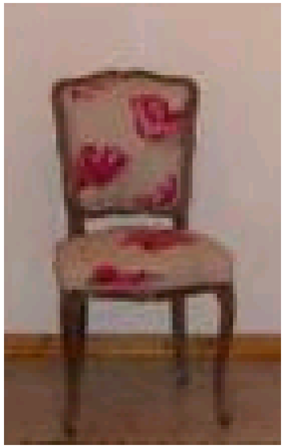
Texture: statistical models of vision

Somewhere in Cinque Terre, May 2005

CS180: Intro to Computer Vision and Comp. Photo
Alexei Efros, UC Berkeley, Fall 2025

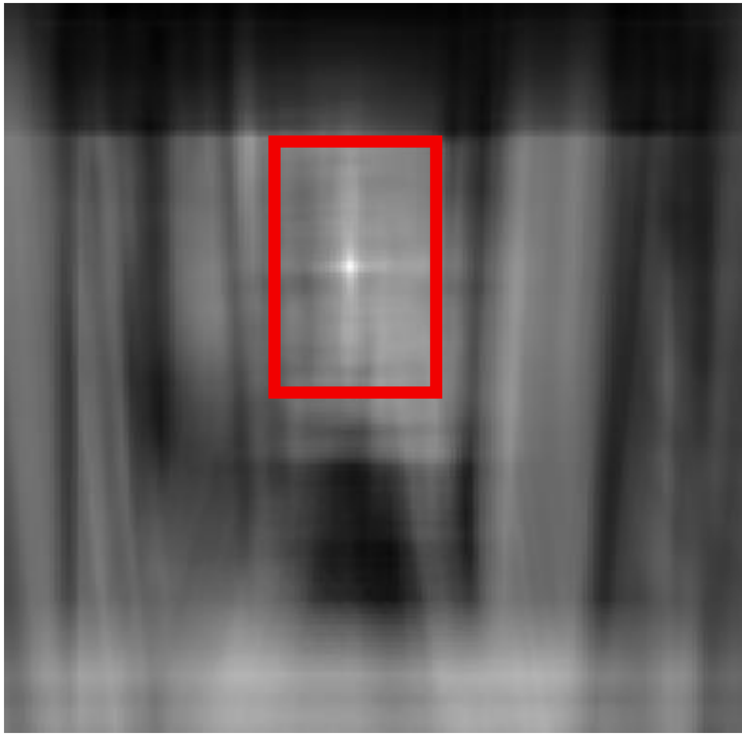
Object recognition
Is it really so hard?

This is a chair

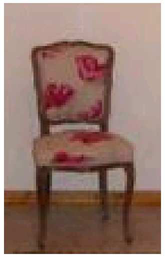


Find the chair in this image

Output of normalized correlation

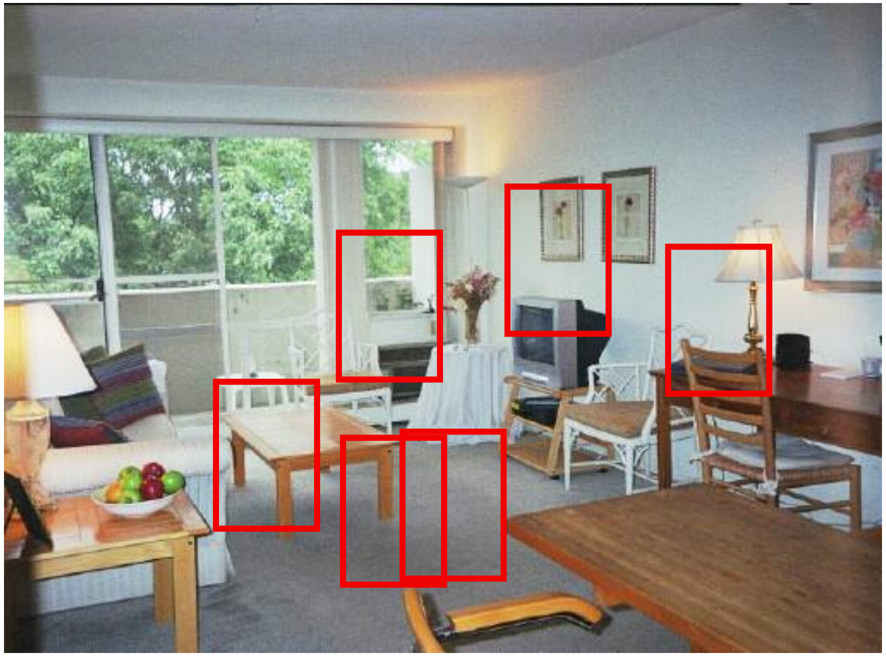


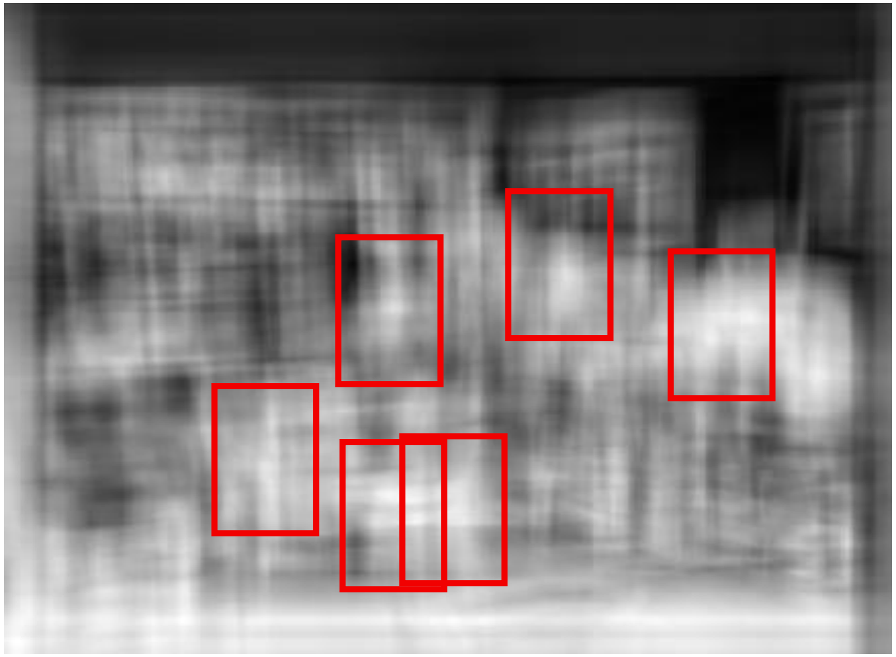
Slide by Antonio Torralba



Object recognition
Is it really so hard?

Find the chair in this image





Slide by Antonio Torralba

Recognition: Instance vs. Category

- Instance recognition:
  - “Find me this particular chair again”
  - Often simple template matching works OK
    - Even better with many small templates, a.k.a. feature descriptors
- Category recognition:
  - “find me all chairs”
  - Templates don’t work. Why?
  - Focus on things that might be <u>invariant</u> across the category
  - Relates to concept of “texture”

What is Texture?

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

radishes

rocks

yogurt

When are two textures similar?

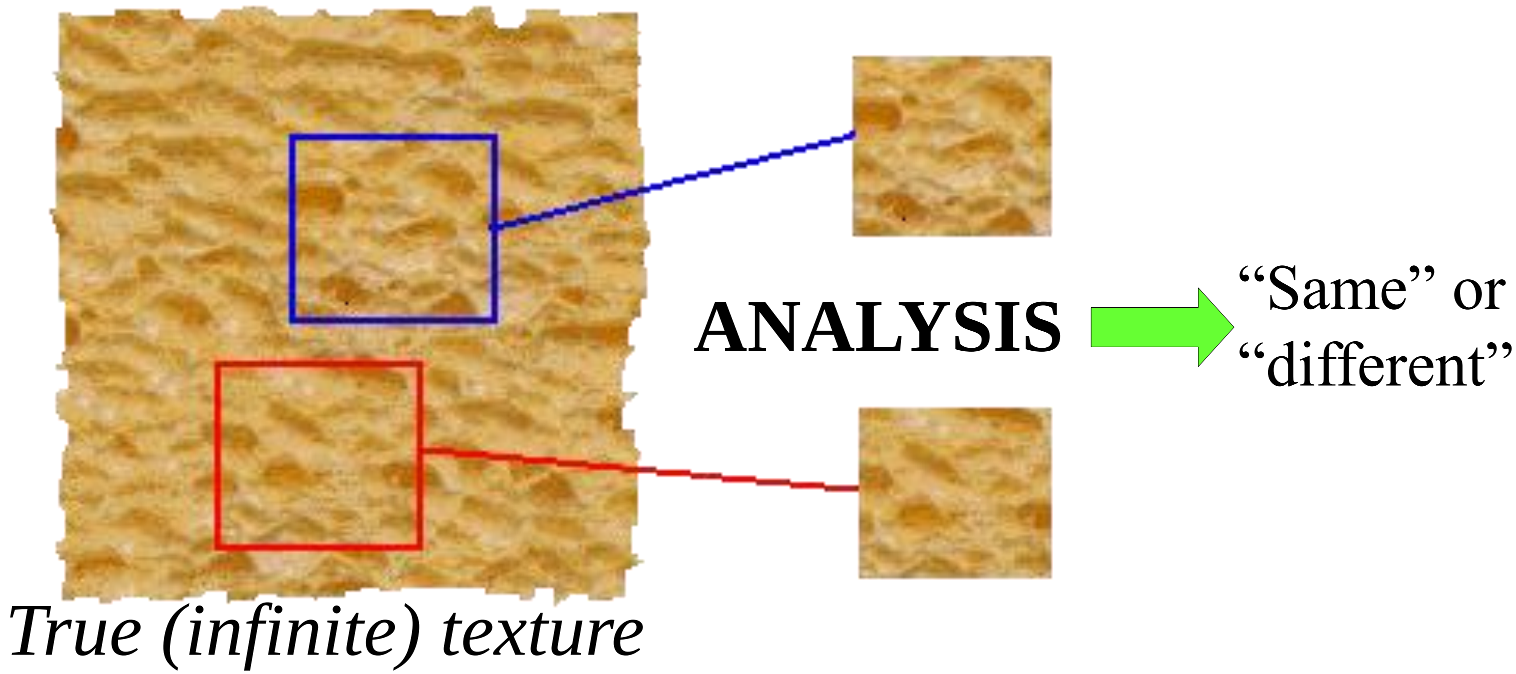




Texture as "stuff"

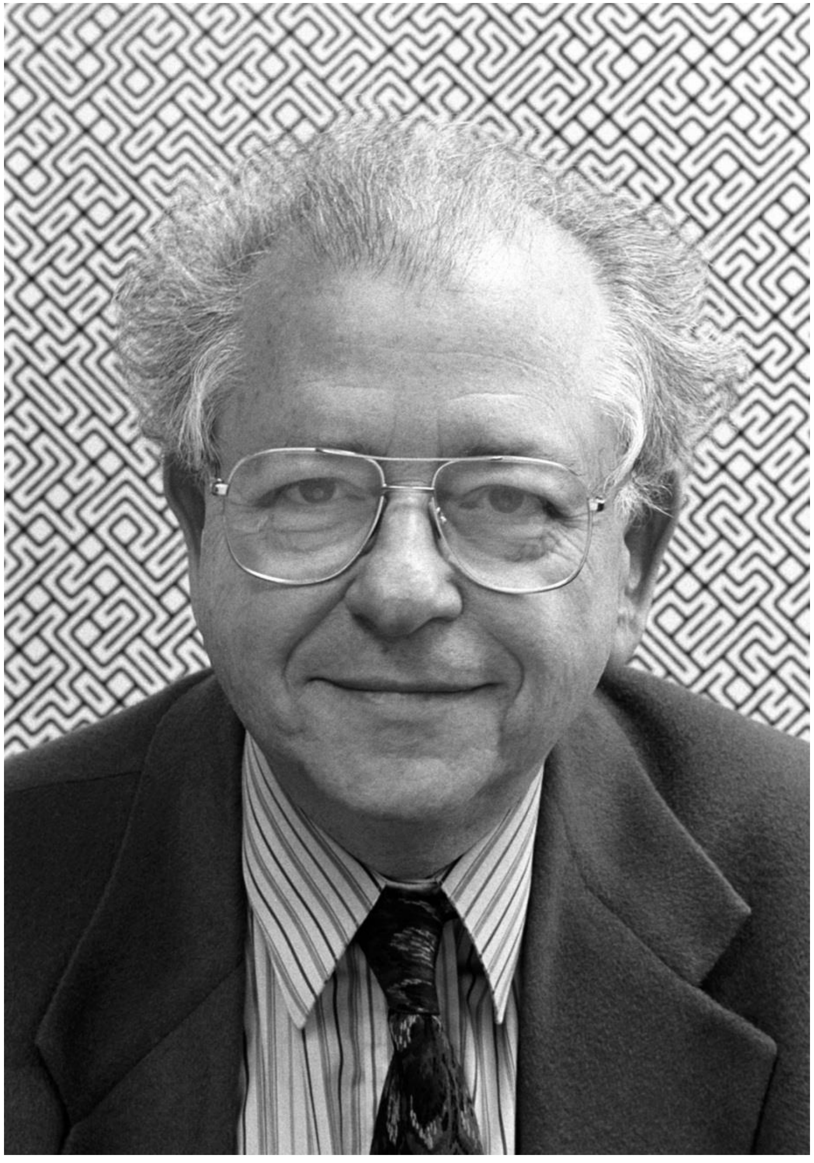
Source: Forsyth

Texture Analysis



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

Béla Julesz, father of texture

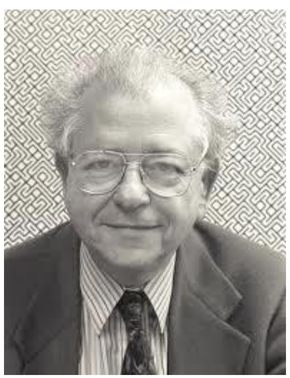


REVIEW ARTICLES

Textons, the elements of texture perception, and their interactions

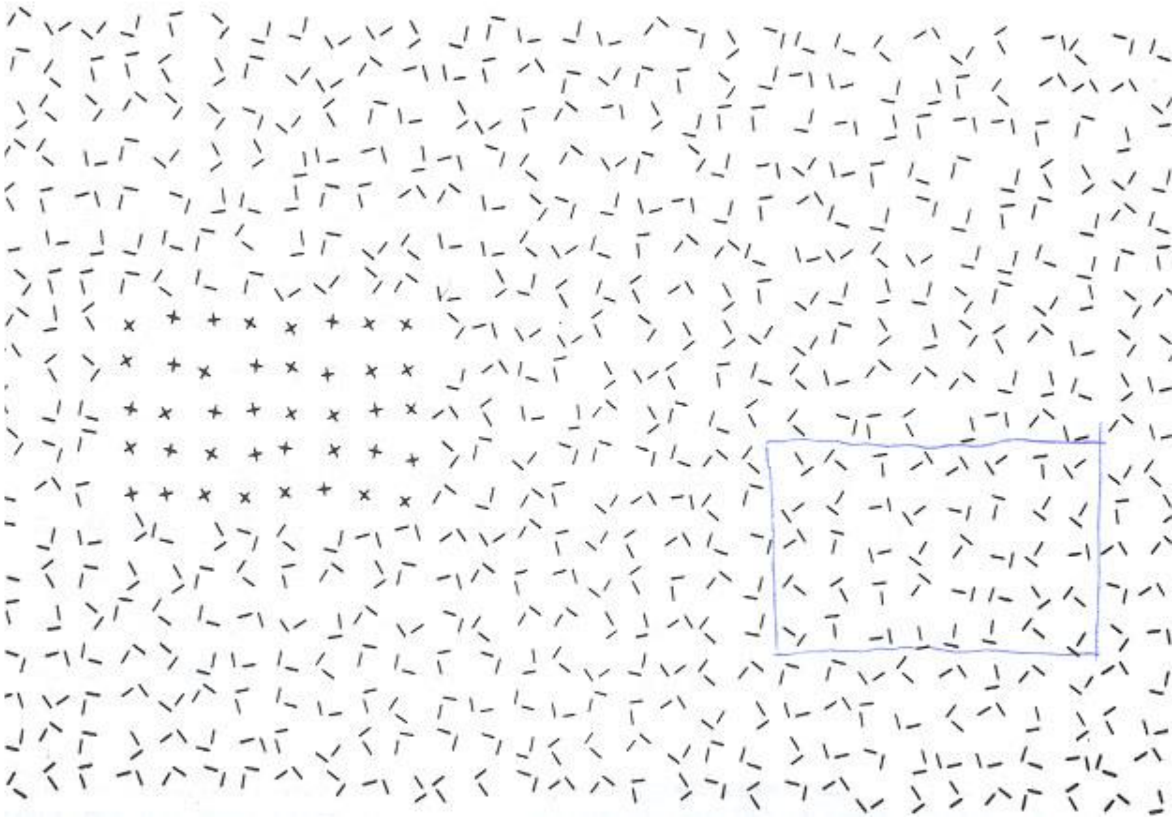
Bela Julesz
Bell Laboratories, Murray Hill, New Jersey 07974, USA

Research with texture pairs having identical second-order statistics has revealed that the pre-attentive texture discrimination system cannot globally process third- and higher-order statistics, and that discrimination is the result of a few local conspicuous features, called textons. It seems that only the first-order statistics of these textons have perceptual significance, and the relative phase between textons cannot be perceived without detailed scrutiny by focal attention.



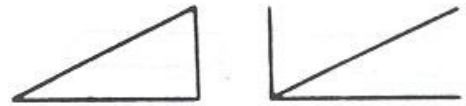
Bela Julesz, "Textons, the Elements of Texture Perception, and their Interactions". Nature 290: 91-97. March, 1981.

Texton Discrimination (Julesz)

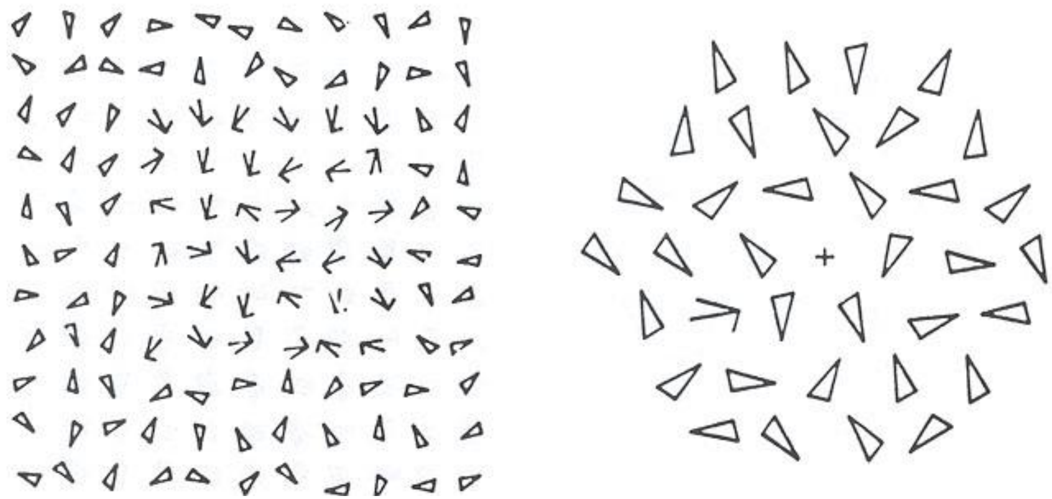


Human vision is sensitive to the difference of some types of elements and appears to be "numb" on other types of differences.

Search Experiment I

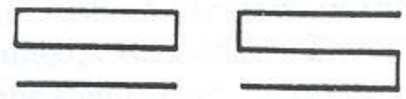


(a)

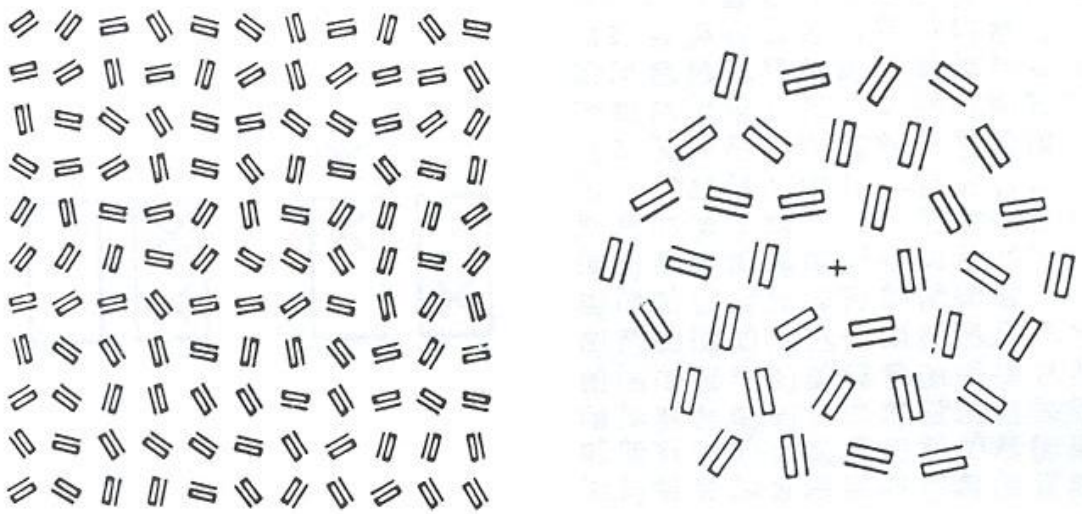


The subject is told to detect a target element in a number of background elements.
In this example, the detection time is independent of the number of background elements.

Search Experiment II



(a)



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

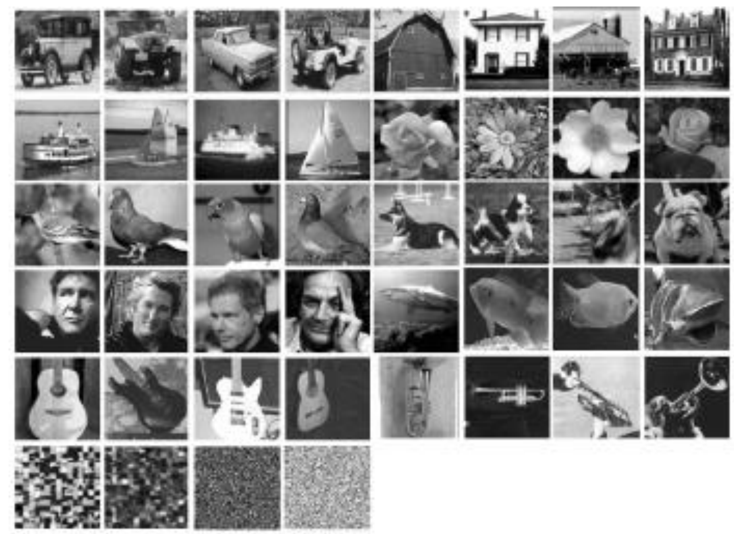
serial search by focal attention in 50ms steps limited to small aperture.

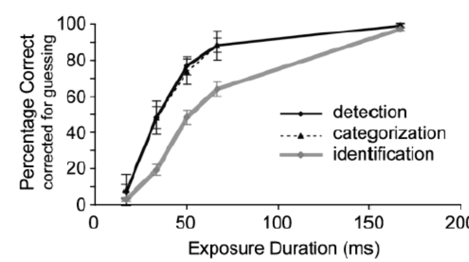
Evidence for Pre-attentive Recognition (Thorpe)

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)

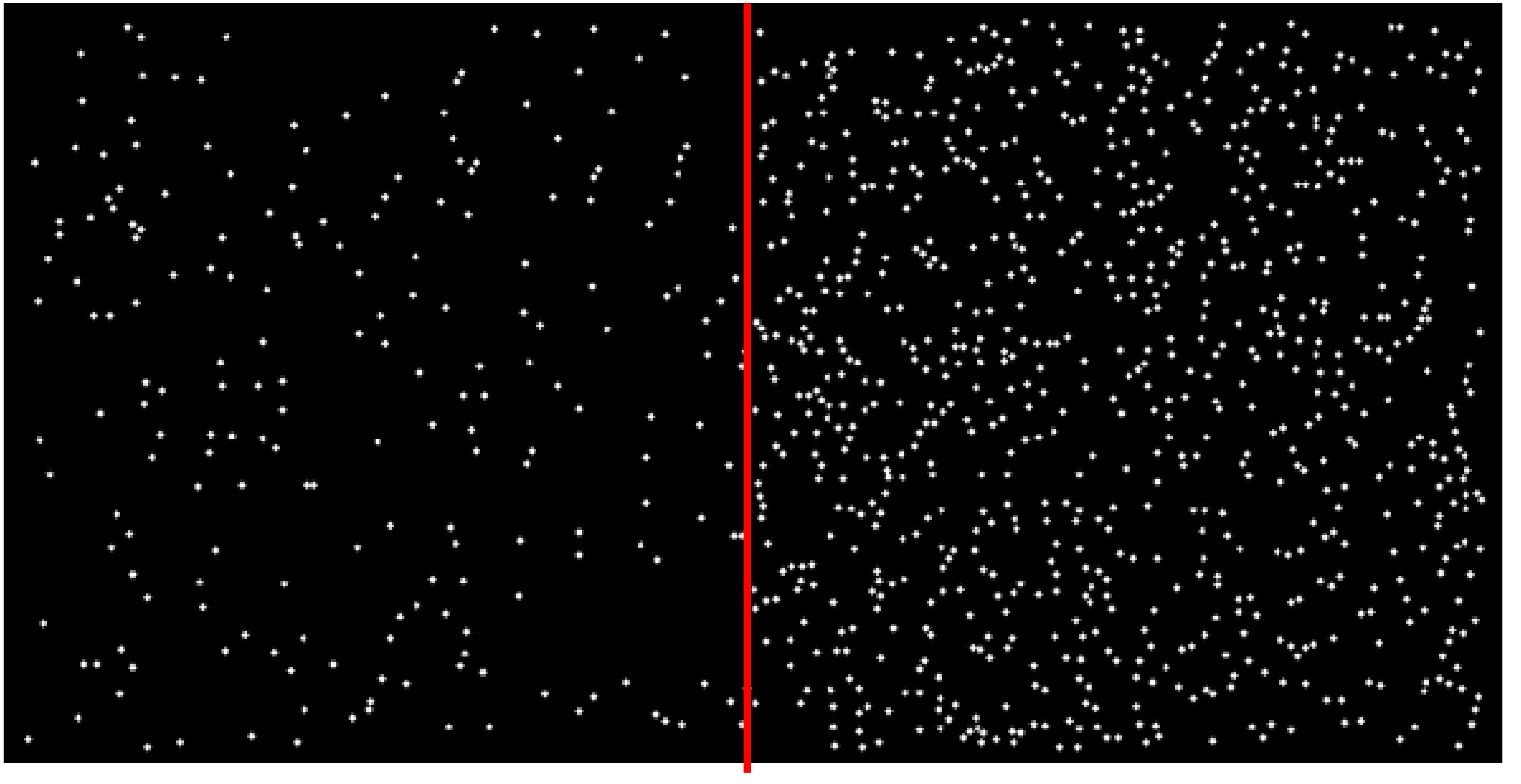




Julesz Conjecture

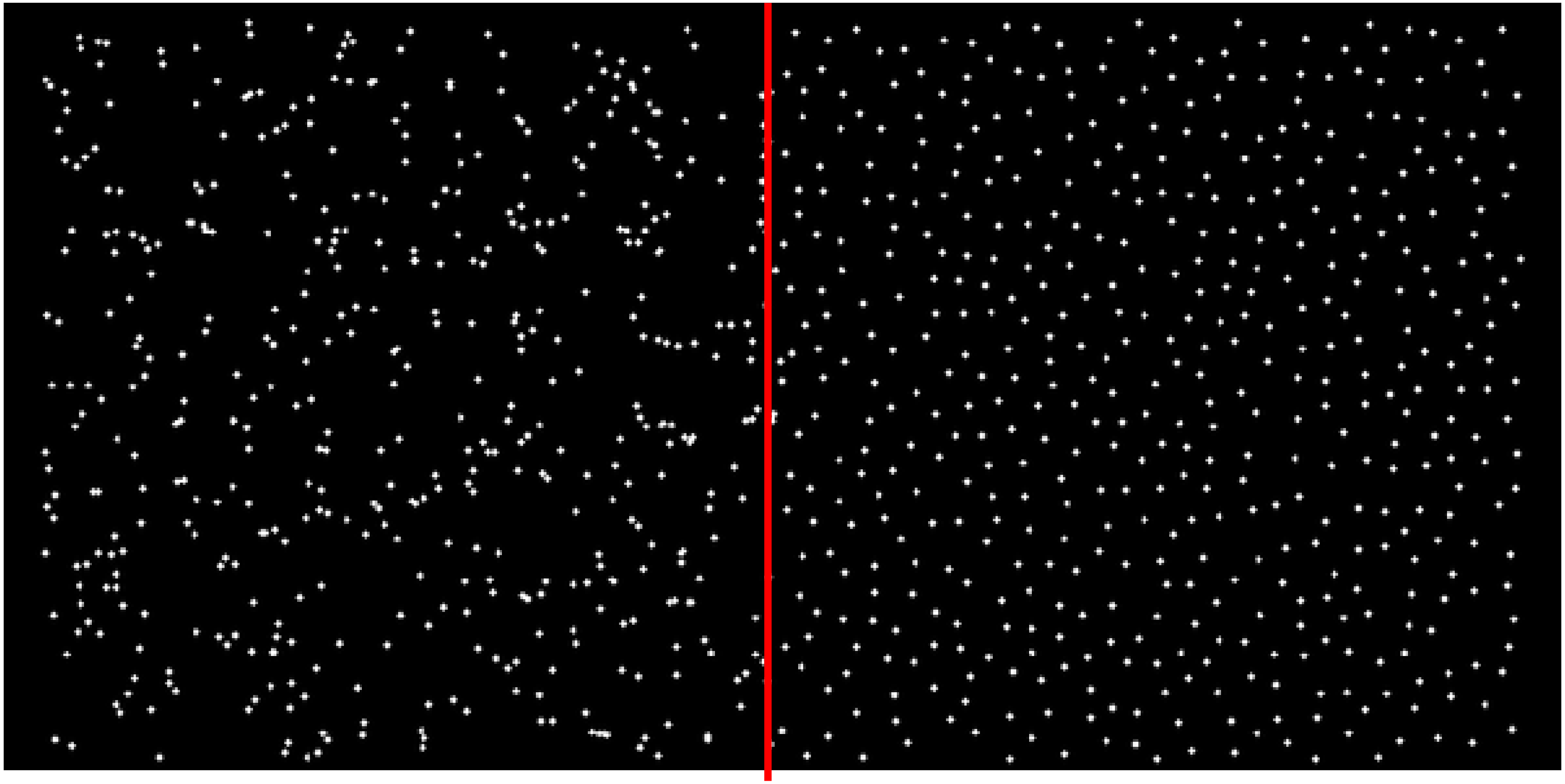
*Textures cannot be spontaneously discriminated if they have the **same first-order and second-order statistics** of texture features (textons) and differ only in their third-order or higher-order statistics.*

1st Order Statistics



5% white   20% white

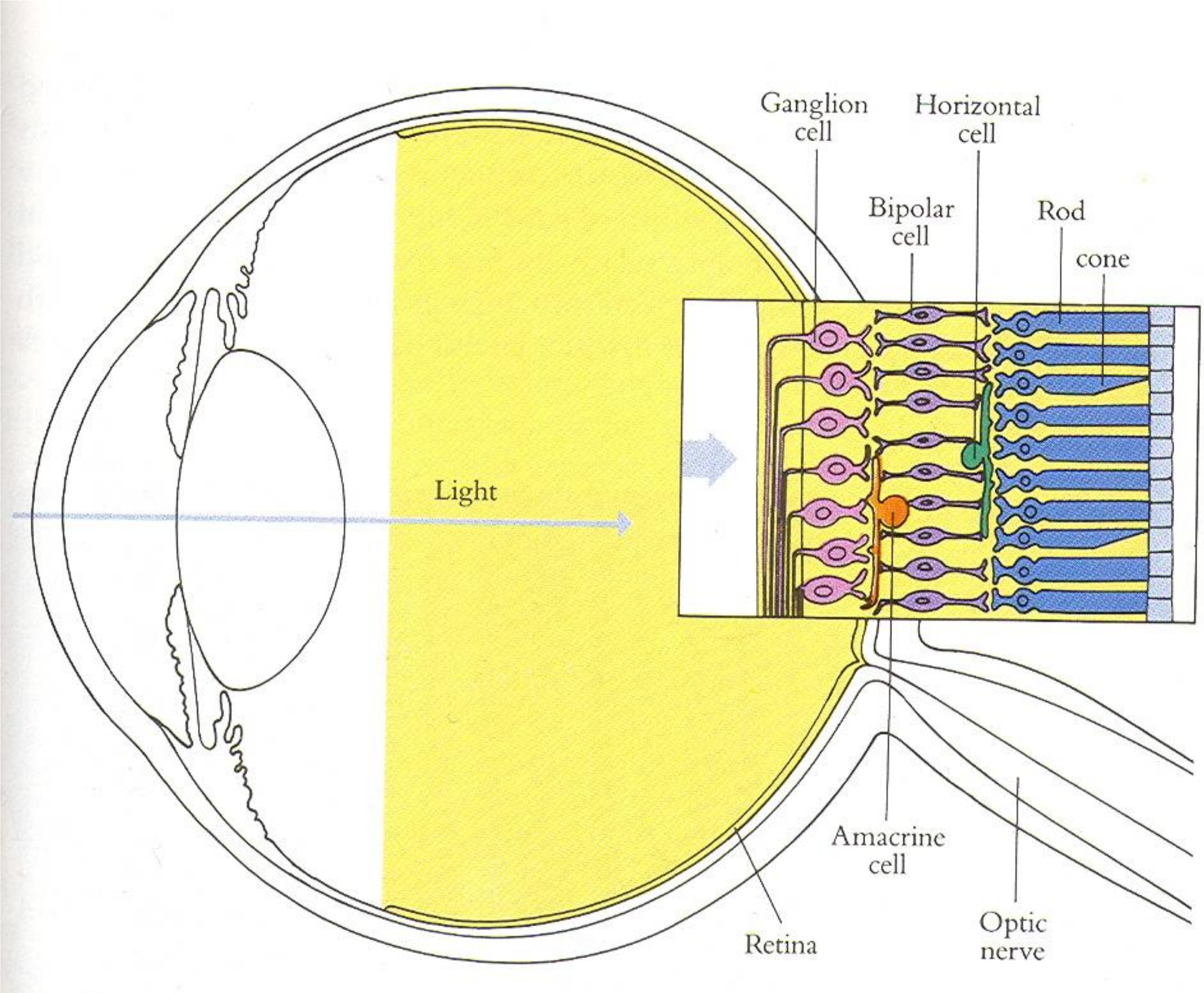
2nd Order Statistics

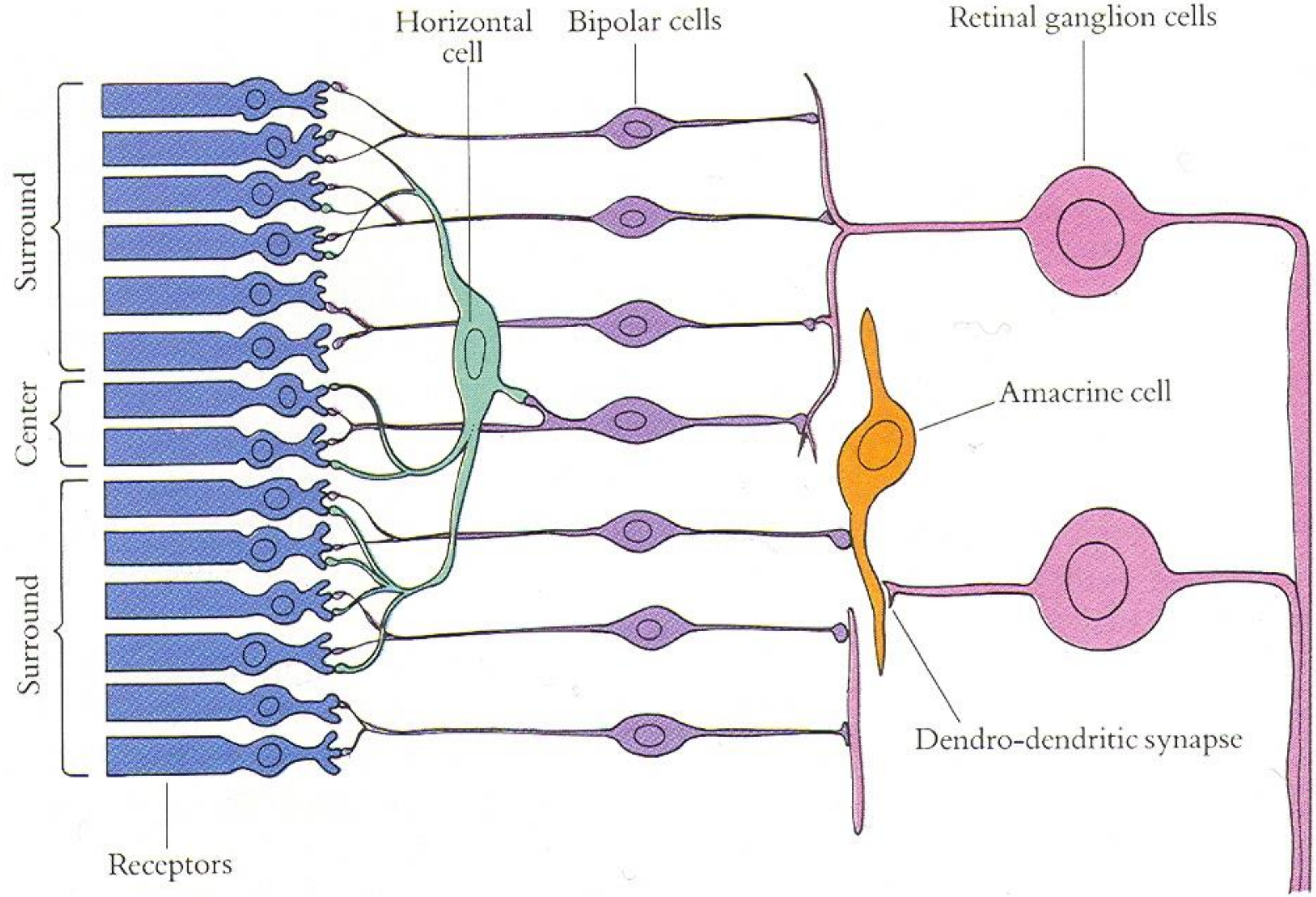


10% white

Big Question

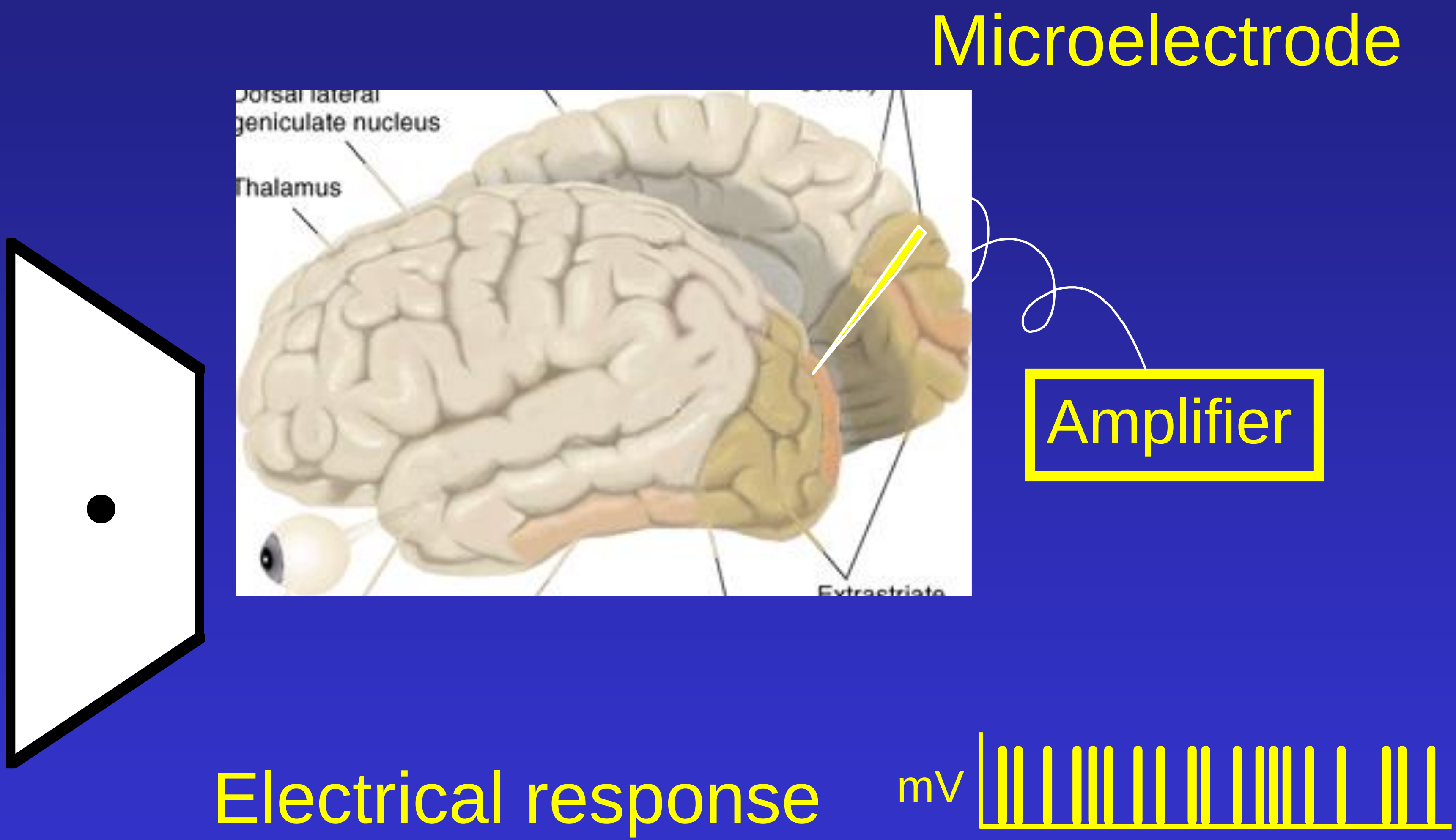
What is the statistical unit (texton) of texture in real images?


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

Single Cell Recording

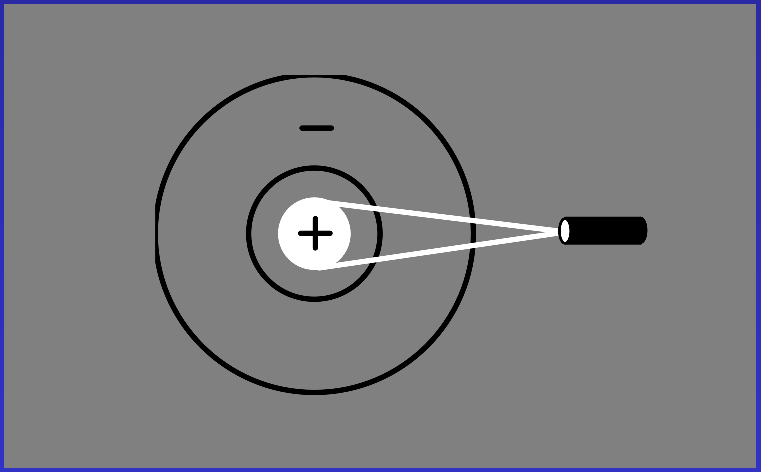


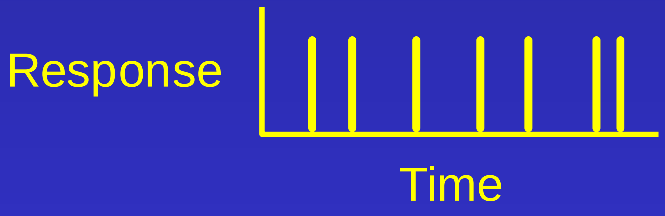
Electrical response
(action potentials)

Retinal Receptive Fields

Receptive field structure in ganglion cells:

On-center Off-surround





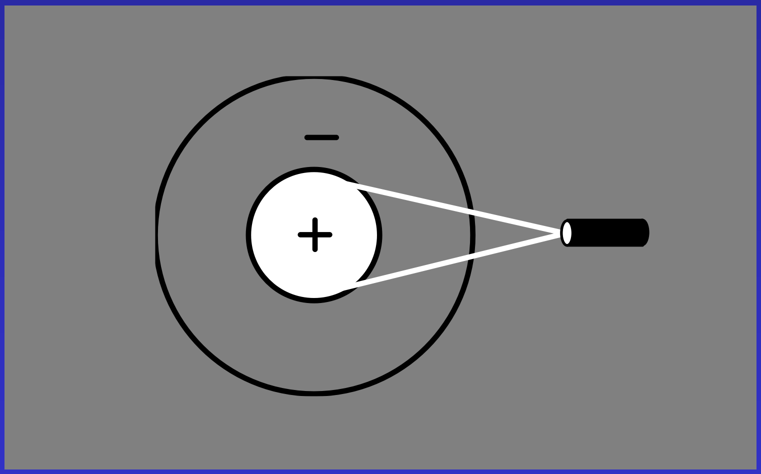
Stimulus condition

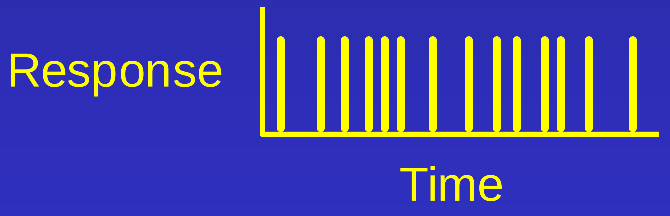
Electrical response

Retinal Receptive Fields

Receptive field structure in ganglion cells:

On-center Off-surround





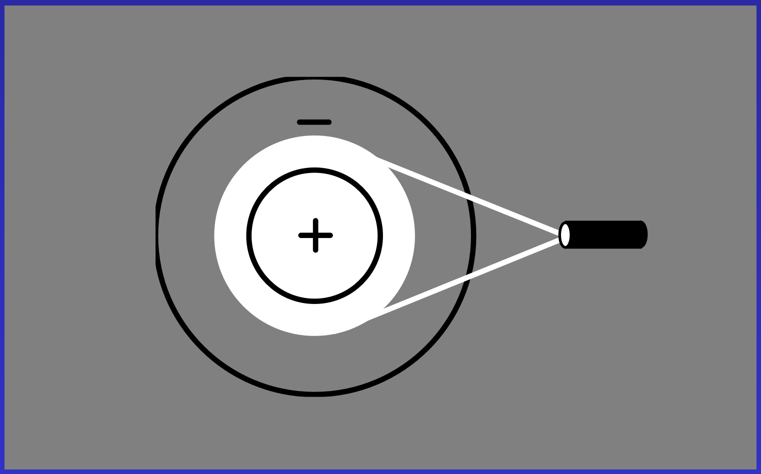
Stimulus condition

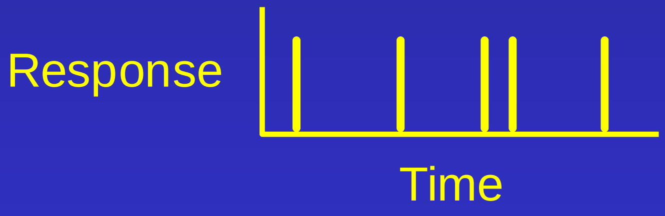
Electrical response

Retinal Receptive Fields

Receptive field structure in ganglion cells:

On-center Off-surround





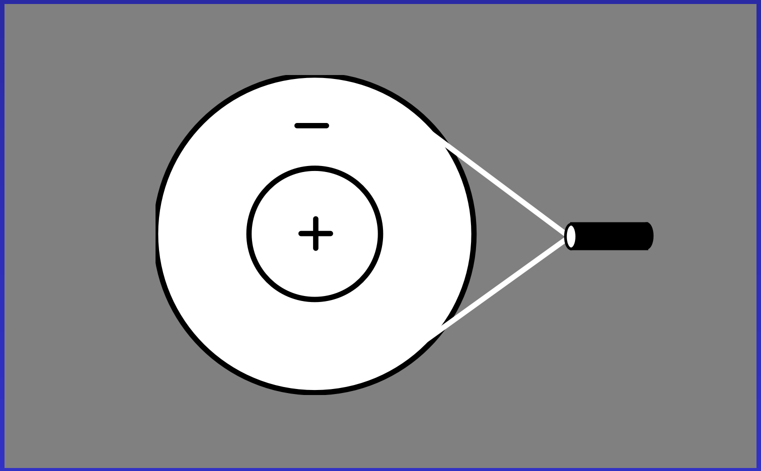
Stimulus condition

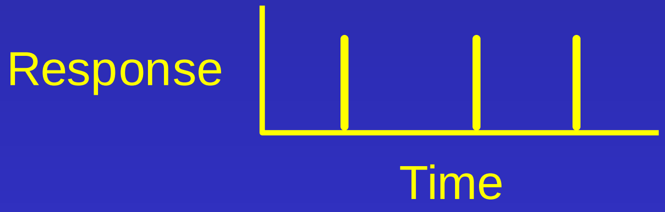
Electrical response

Retinal Receptive Fields

Receptive field structure in ganglion cells:

On-center Off-surround





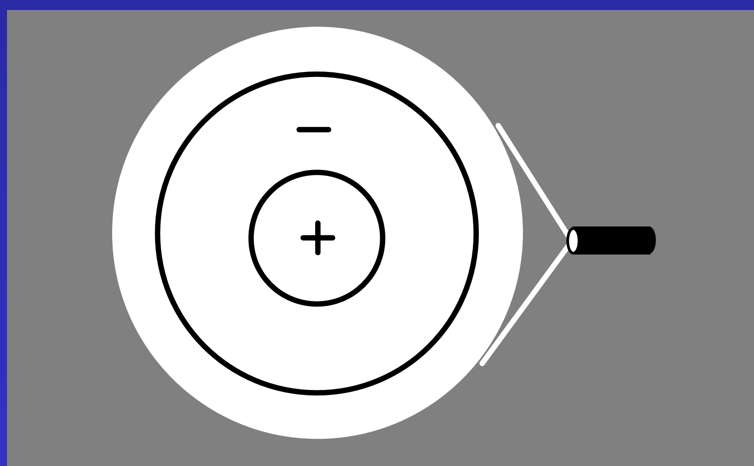
Stimulus condition

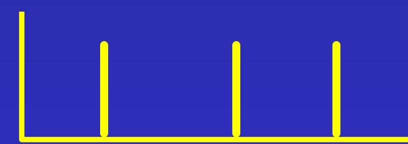
Electrical response

Retinal Receptive Fields

Receptive field structure in ganglion cells:

On-center Off-surround





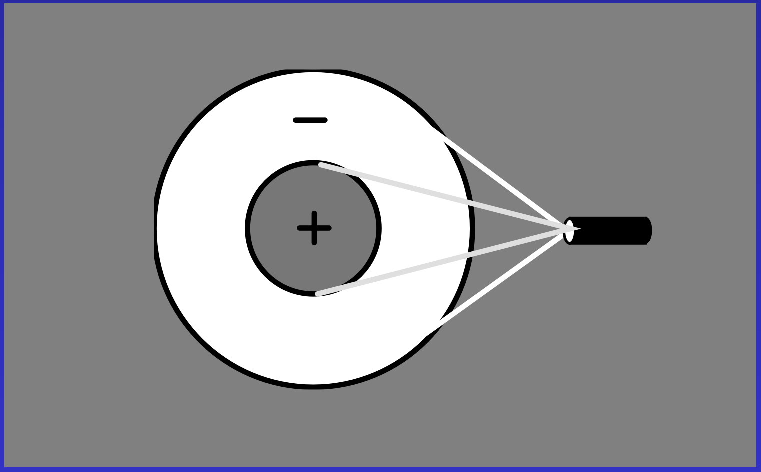
Stimulus condition

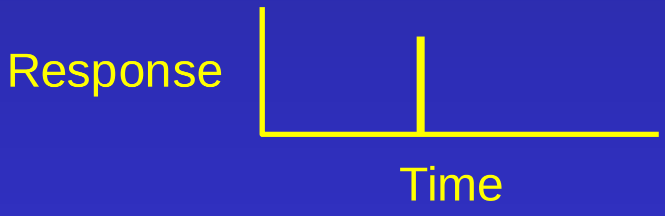
Electrical response

Retinal Receptive Fields

Receptive field structure in ganglion cells:

On-center Off-surround



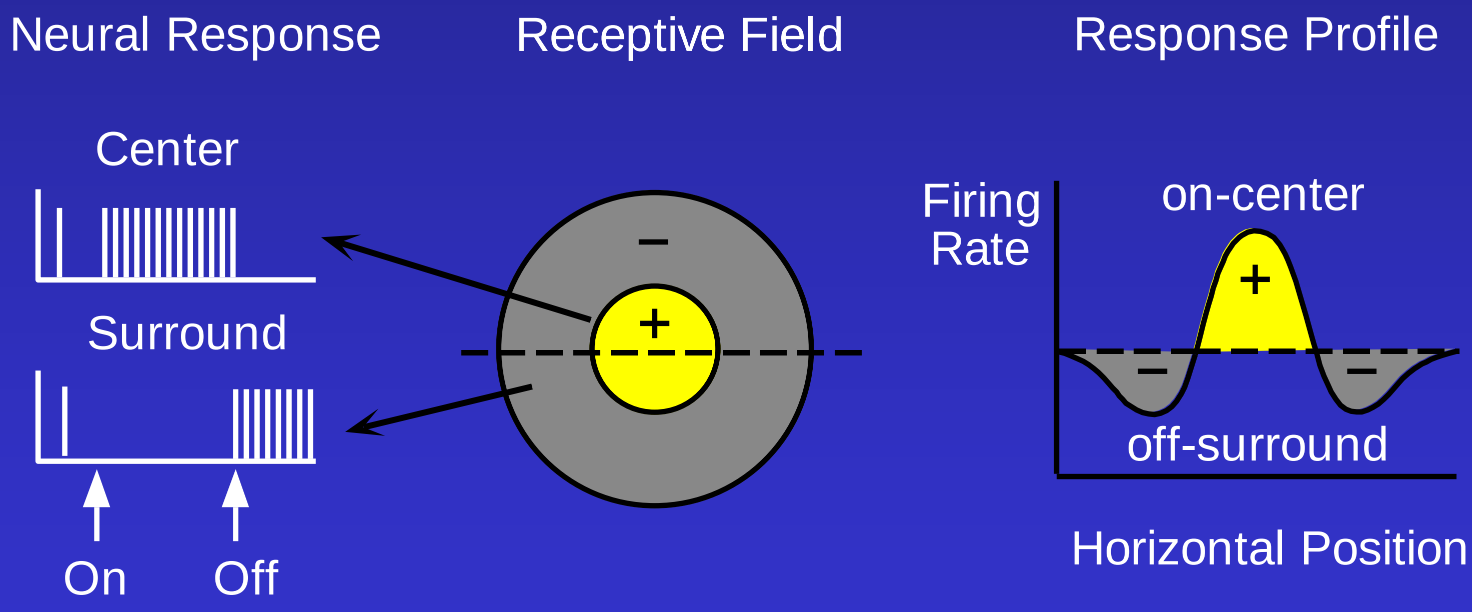


Stimulus condition

Electrical response

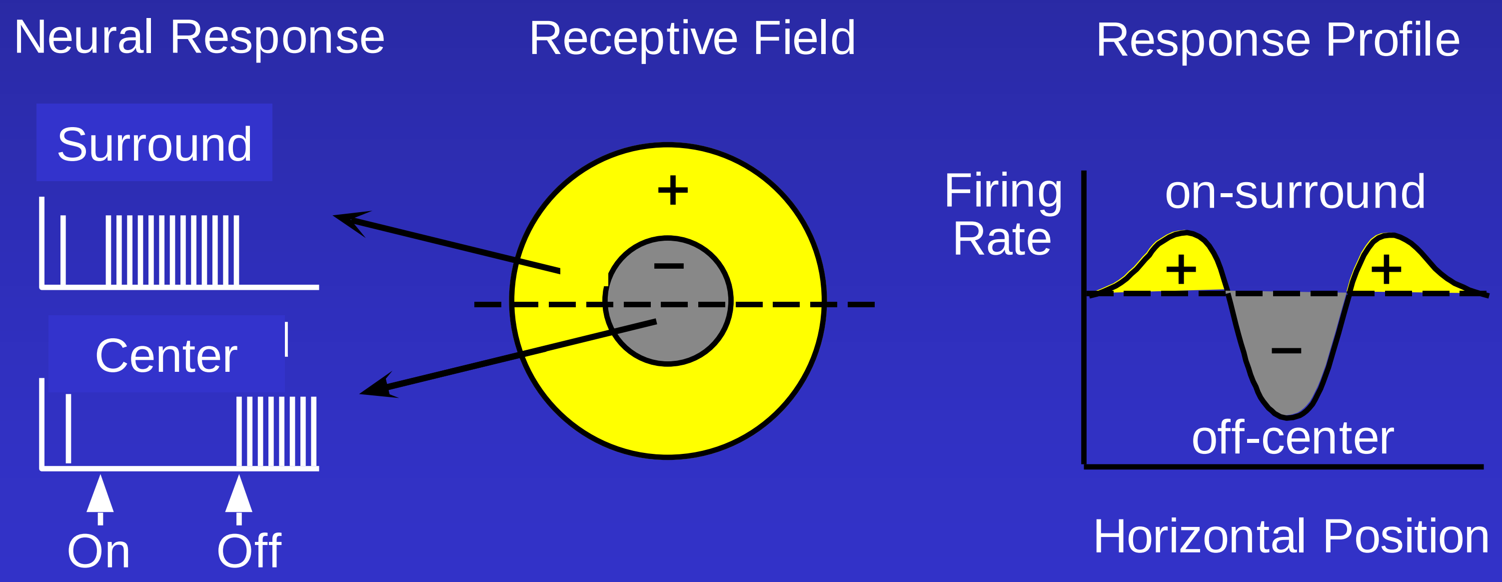
Retinal Receptive Fields

RF of On-center Off-surround cells

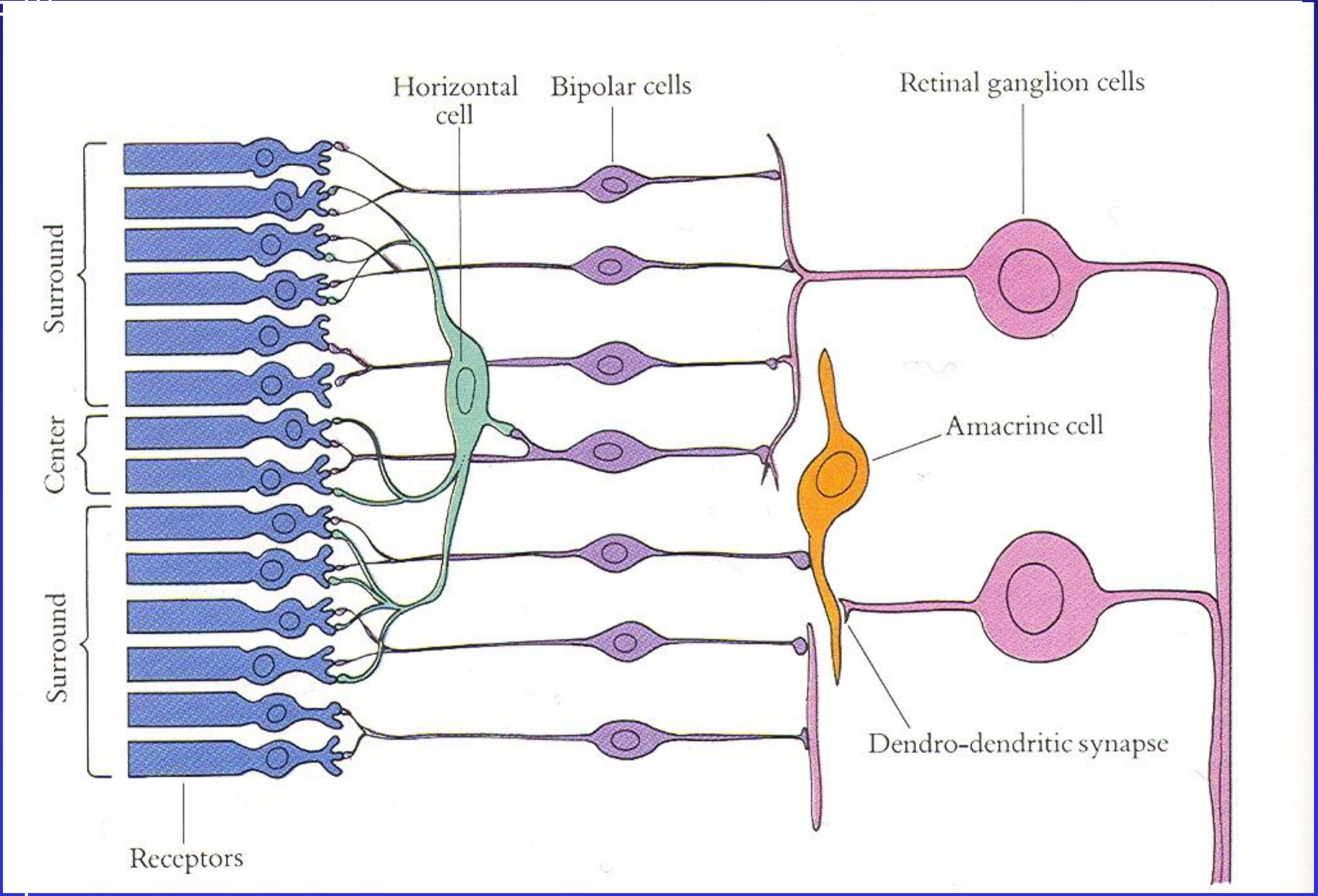


Retinal Receptive Fields

RF of Off-center On-surround cells

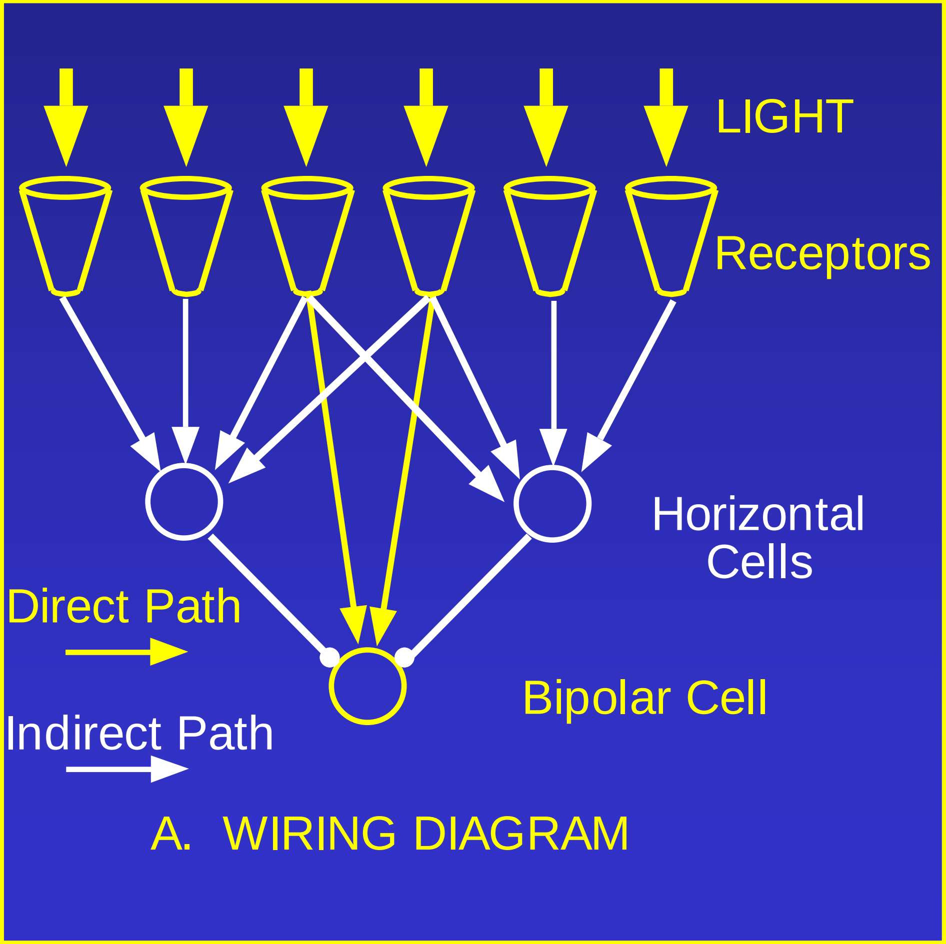


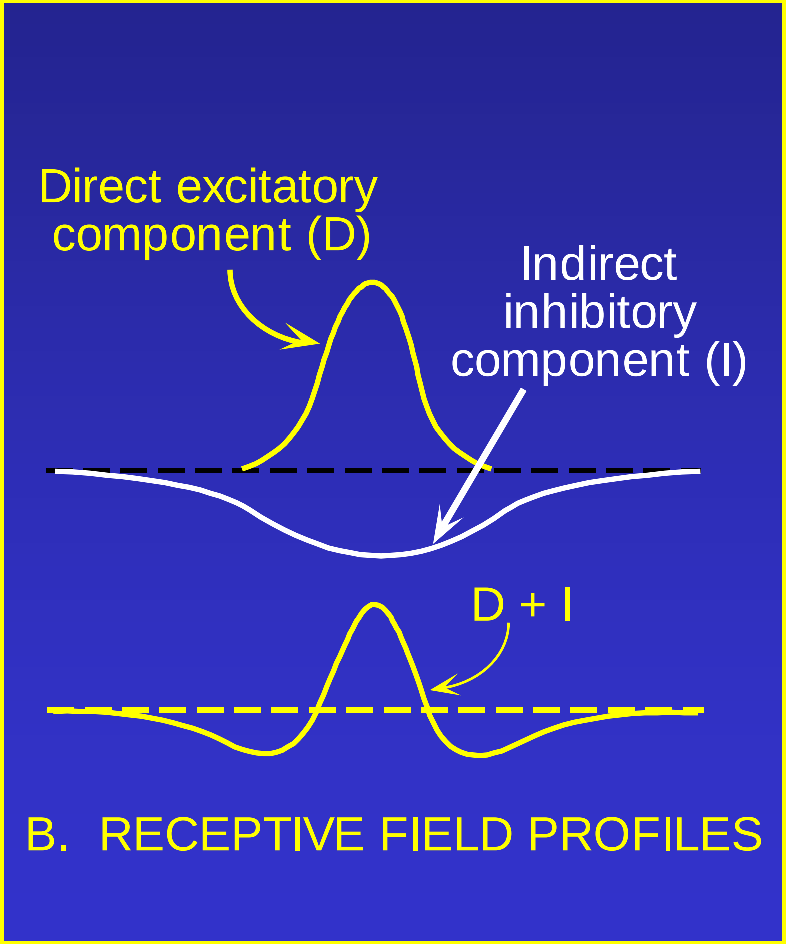
Retinal Receptive Fields



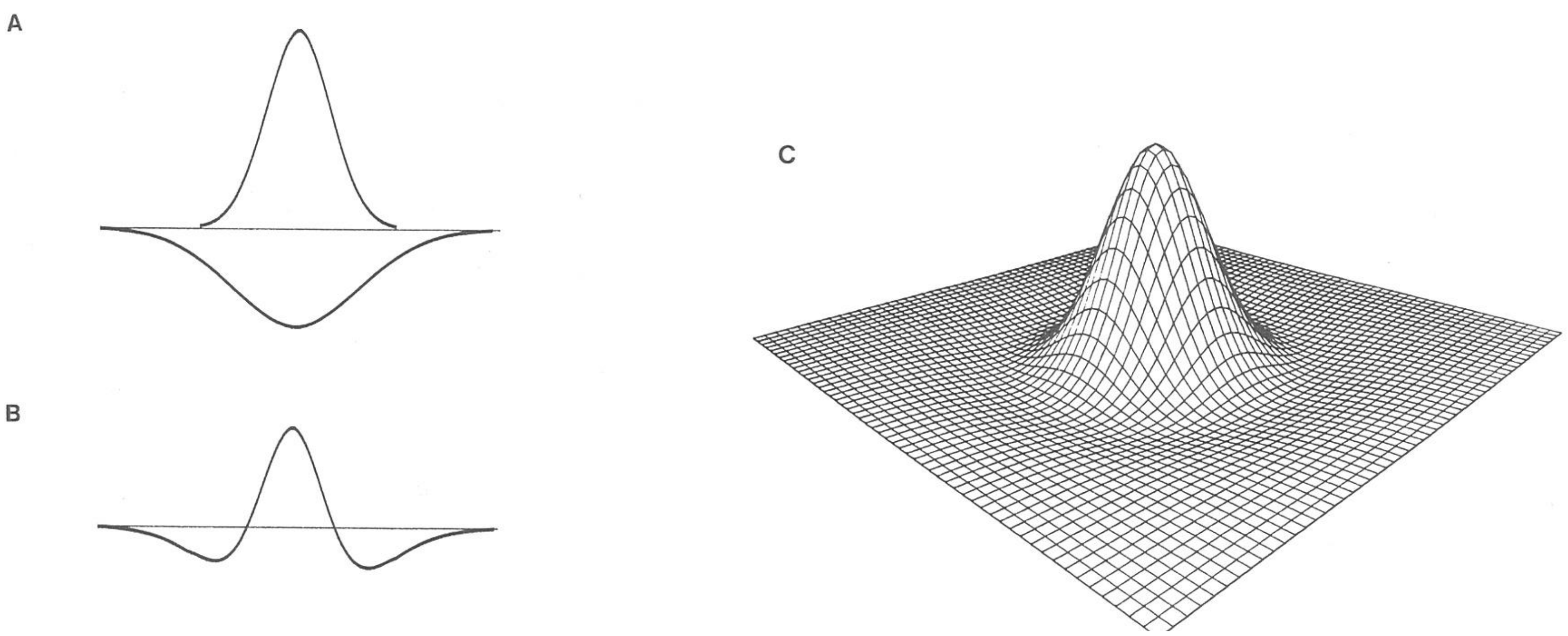
Retinal Receptive Fields

Receptive field structure in bipolar cells





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



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

Receptive Fields

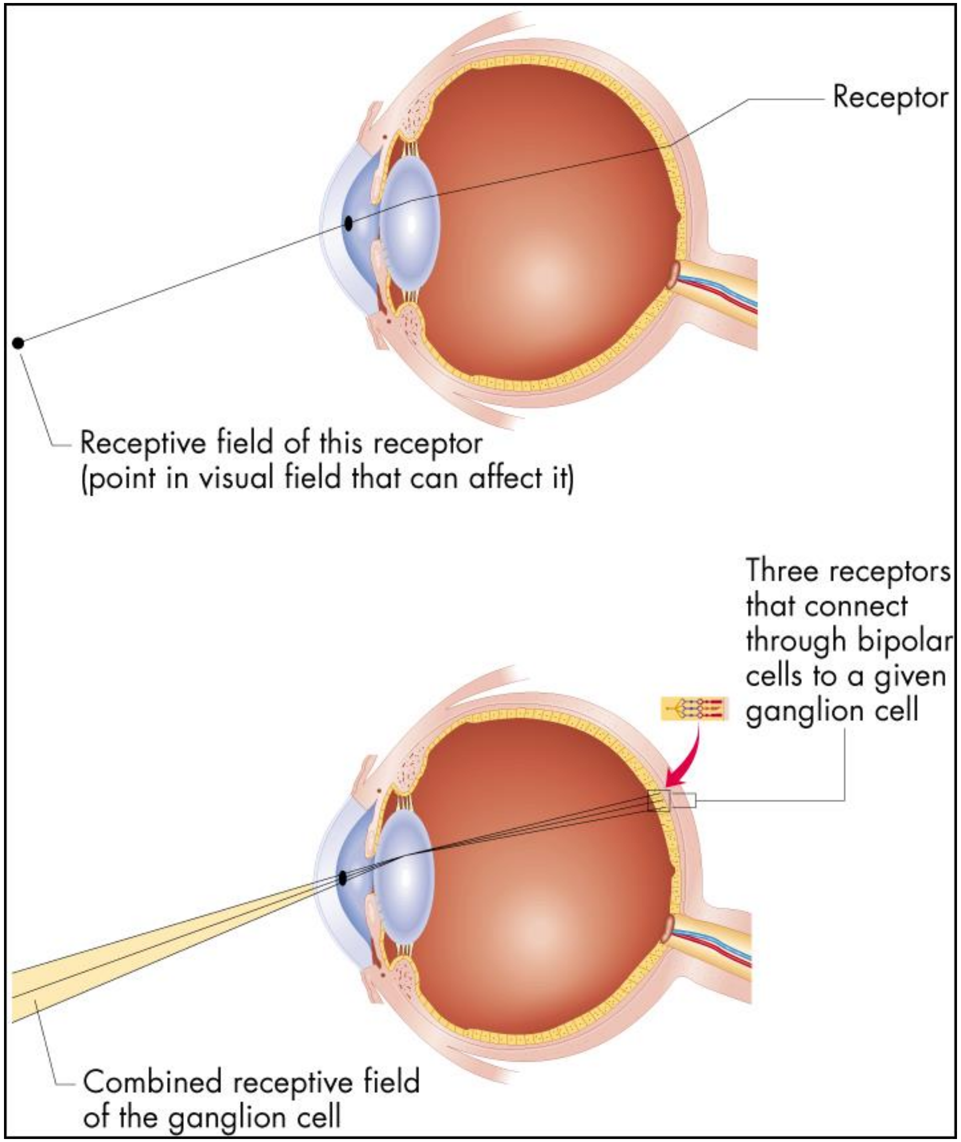
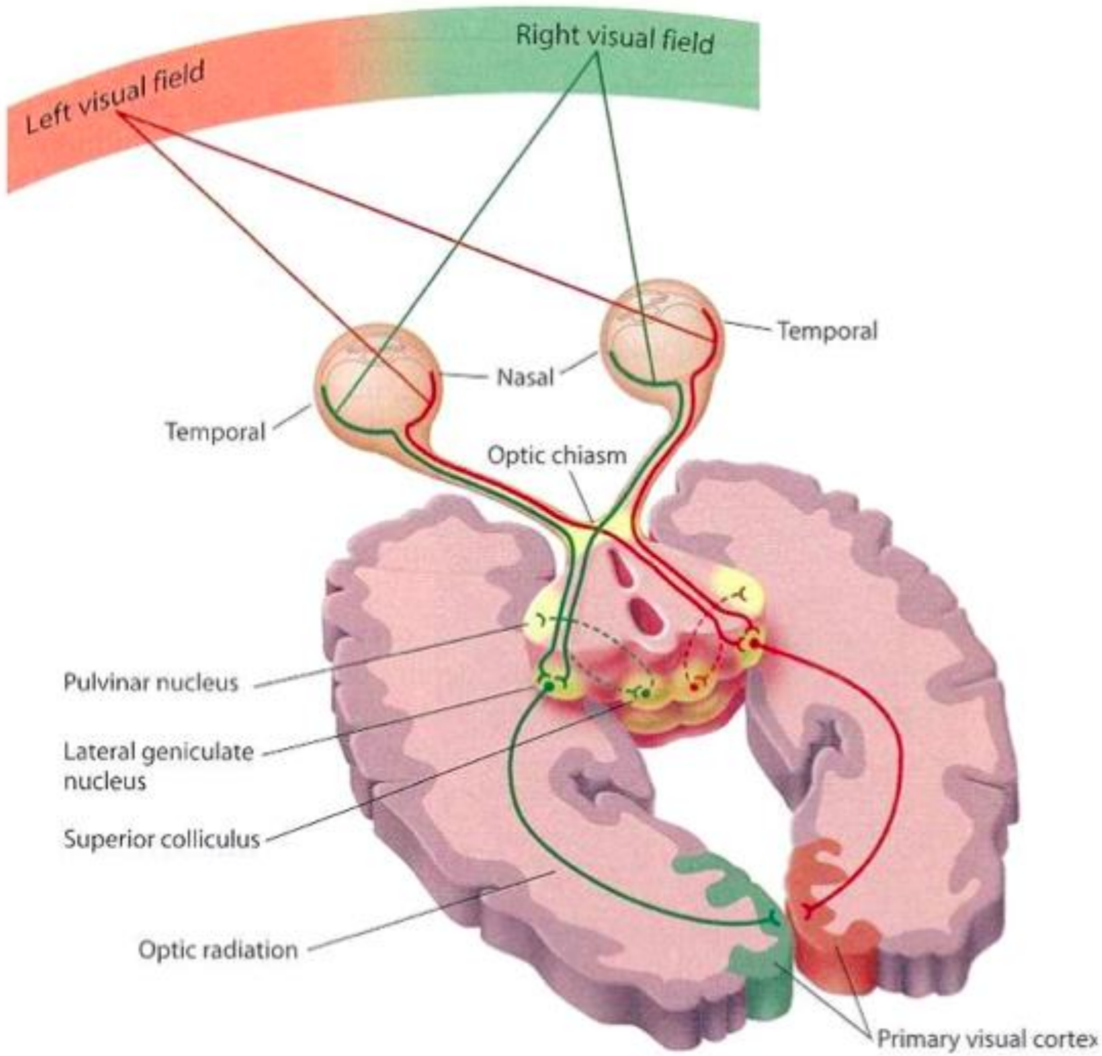


Figure 6.16 Receptive fields
The receptive field of a receptor is simply the area of the visual field from which light strikes that receptor. For any other cell in the visual system, the receptive field is determined by which receptors connect to the cell in question.

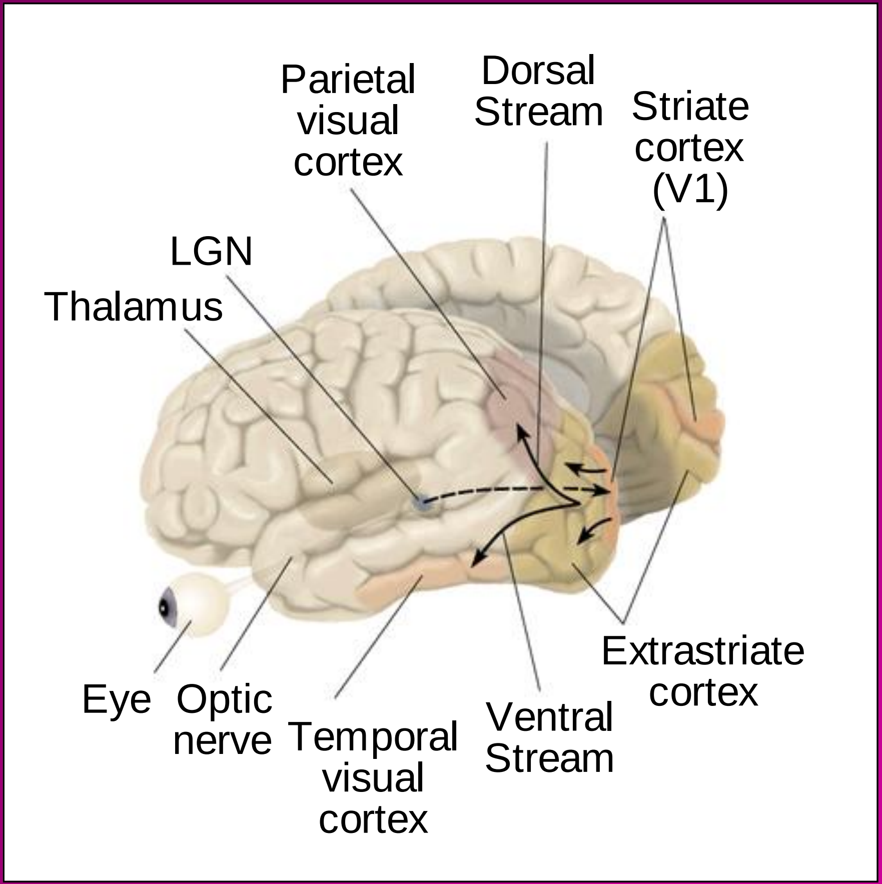
Anatomy of Pathway to Visual Cortex



Visual Cortex

Cortical Area V1

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



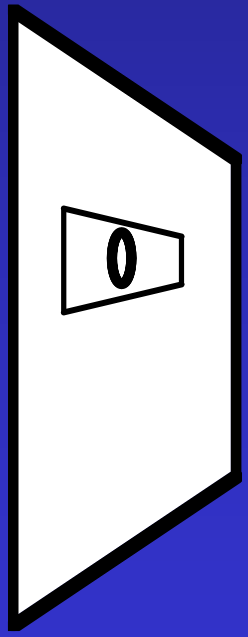
Cortical Receptive Fields

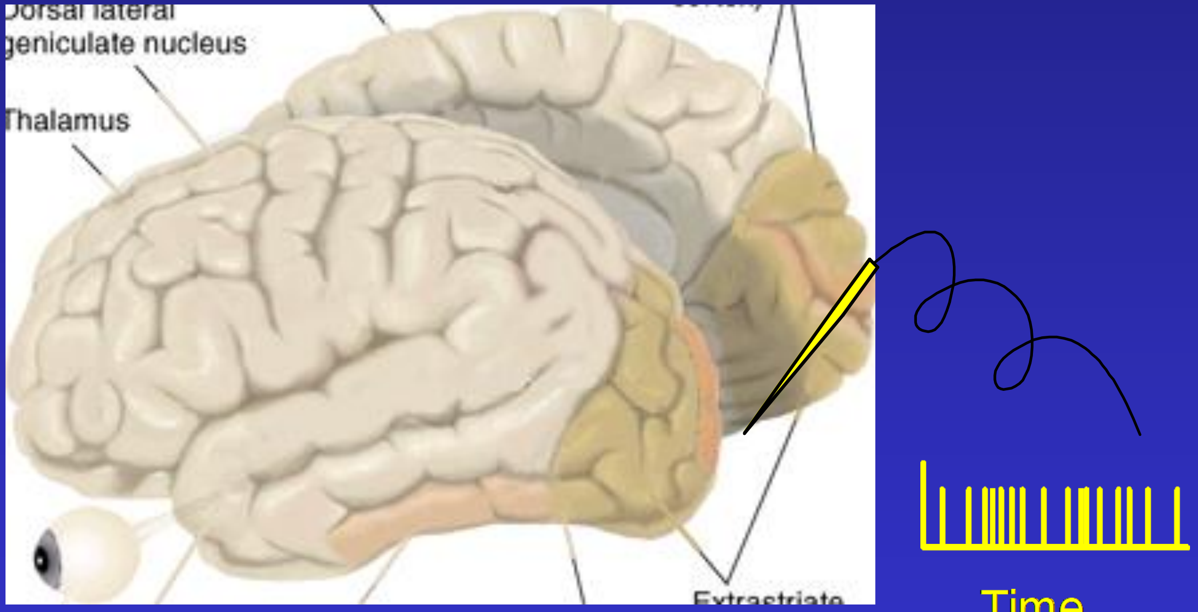
Single-cell recording from visual cortex

David Hubel & Thorston Wiesel

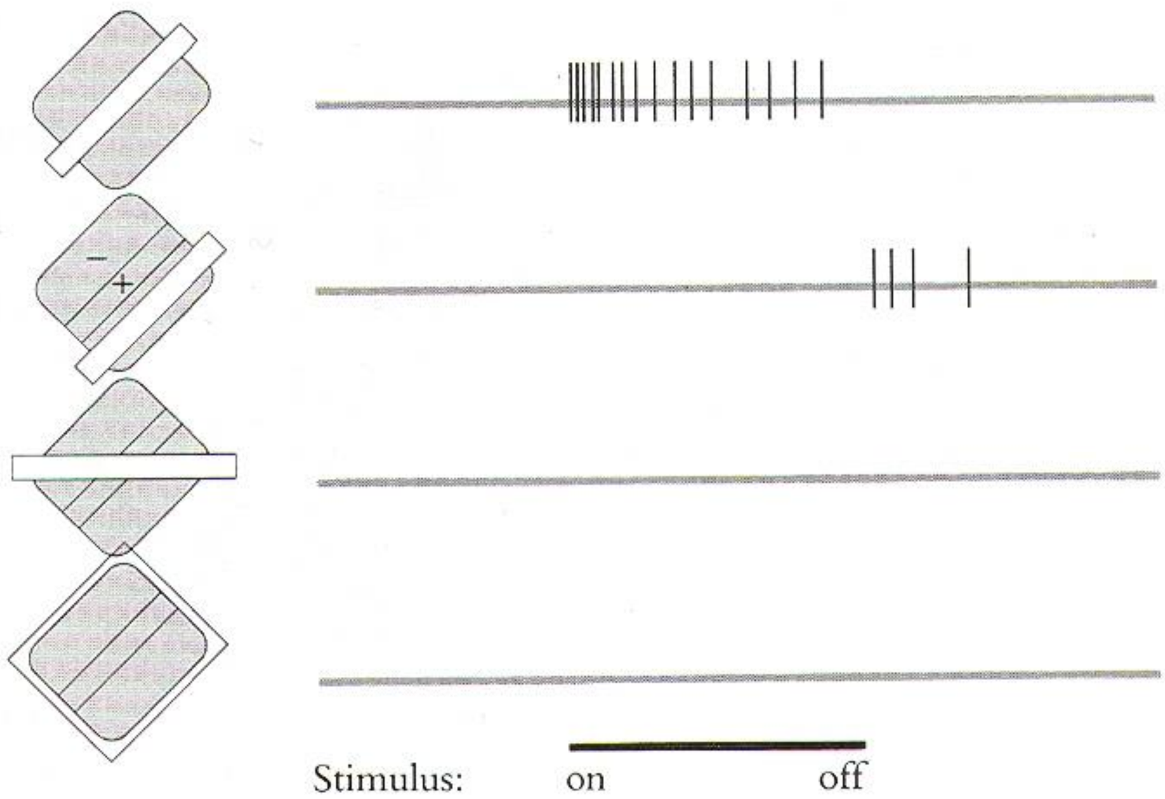
Cortical Receptive Fields

Single-cell recording from visual cortex



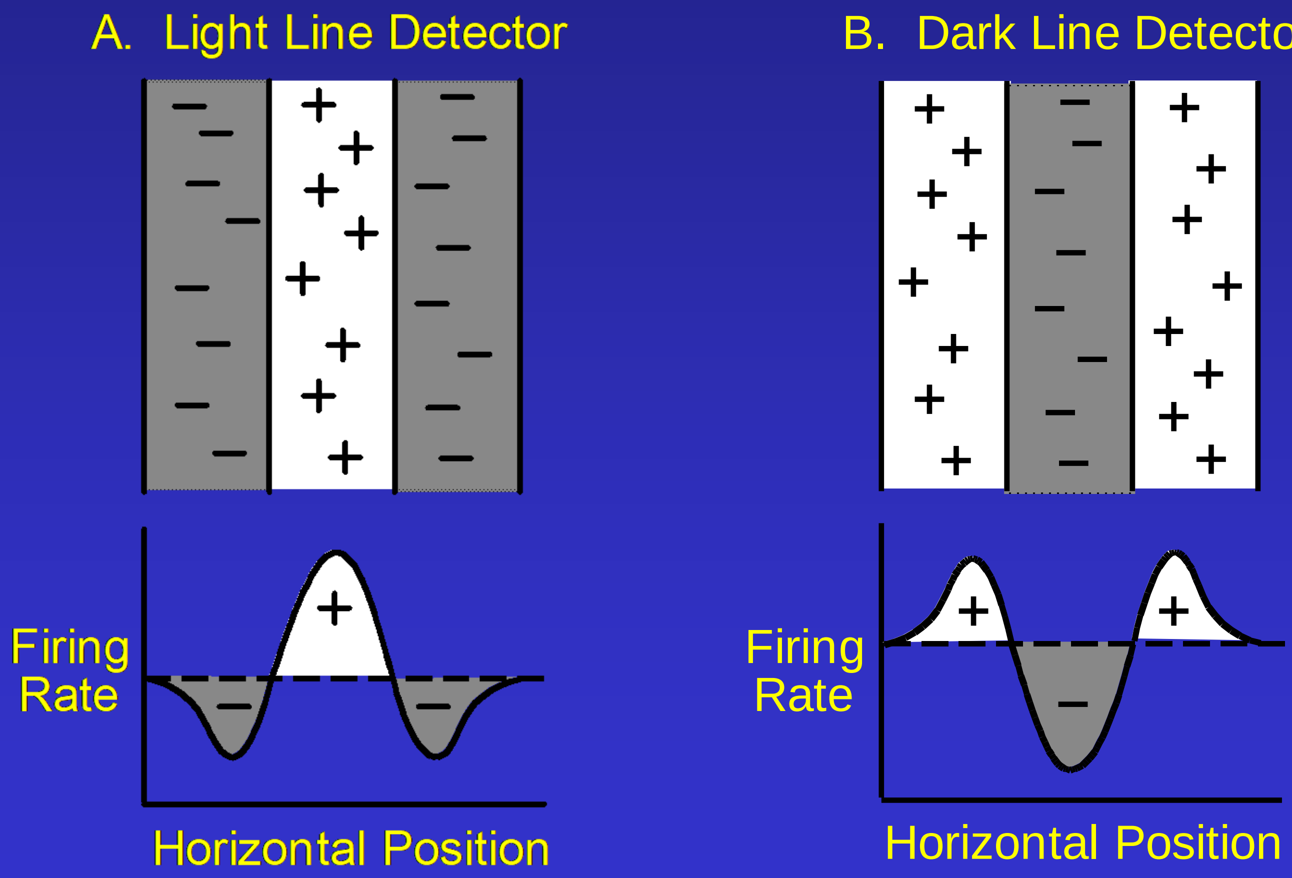


https://www.youtube.com/watch?v=IOHayh06LJ4


−
+
Stimulus:
on
off

Cortical Receptive Fields

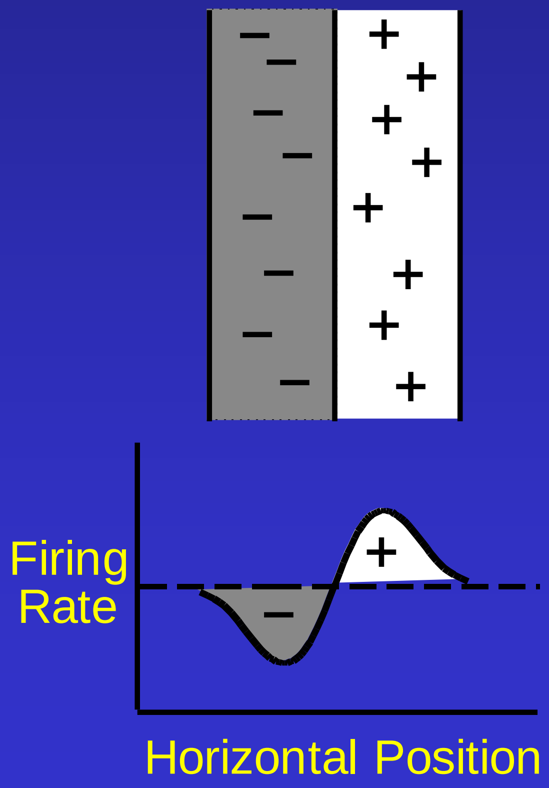
Simple Cells: "Line Detectors"



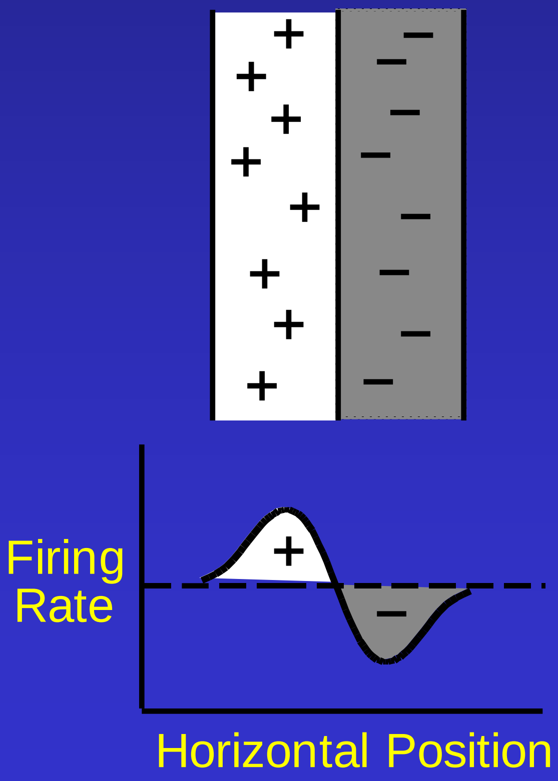
Cortical Receptive Fields

Simple Cells: "Edge Detectors"

C. Dark-to-light Edge Detector

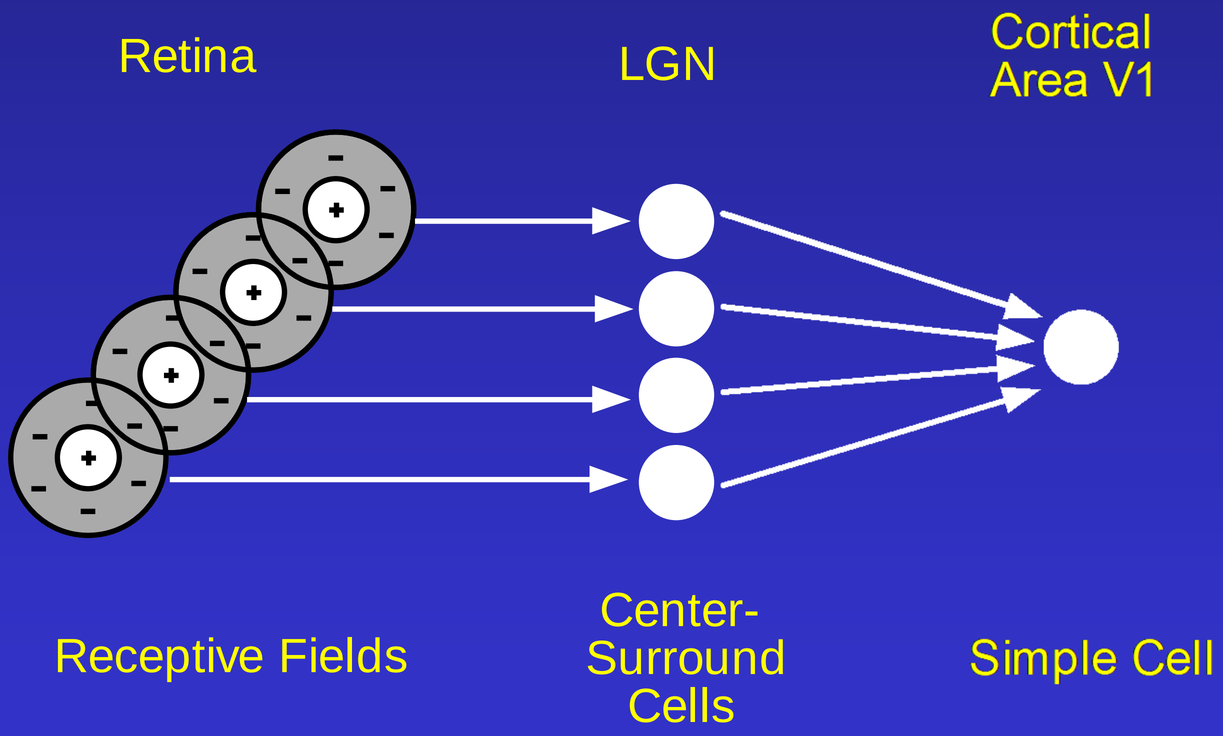


D. Light-to-dark Edge Detector

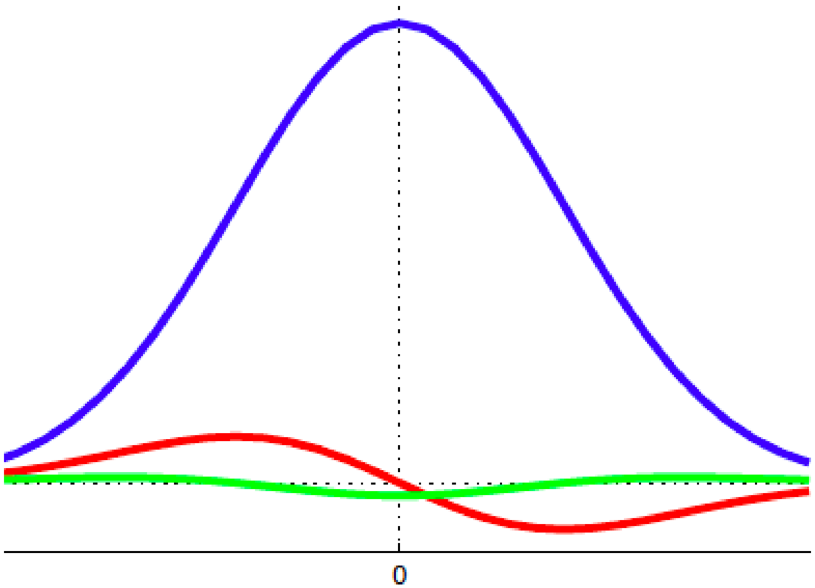


Cortical Receptive Fields

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

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

We also consider rotated versions of these Gaussian derivative functions.

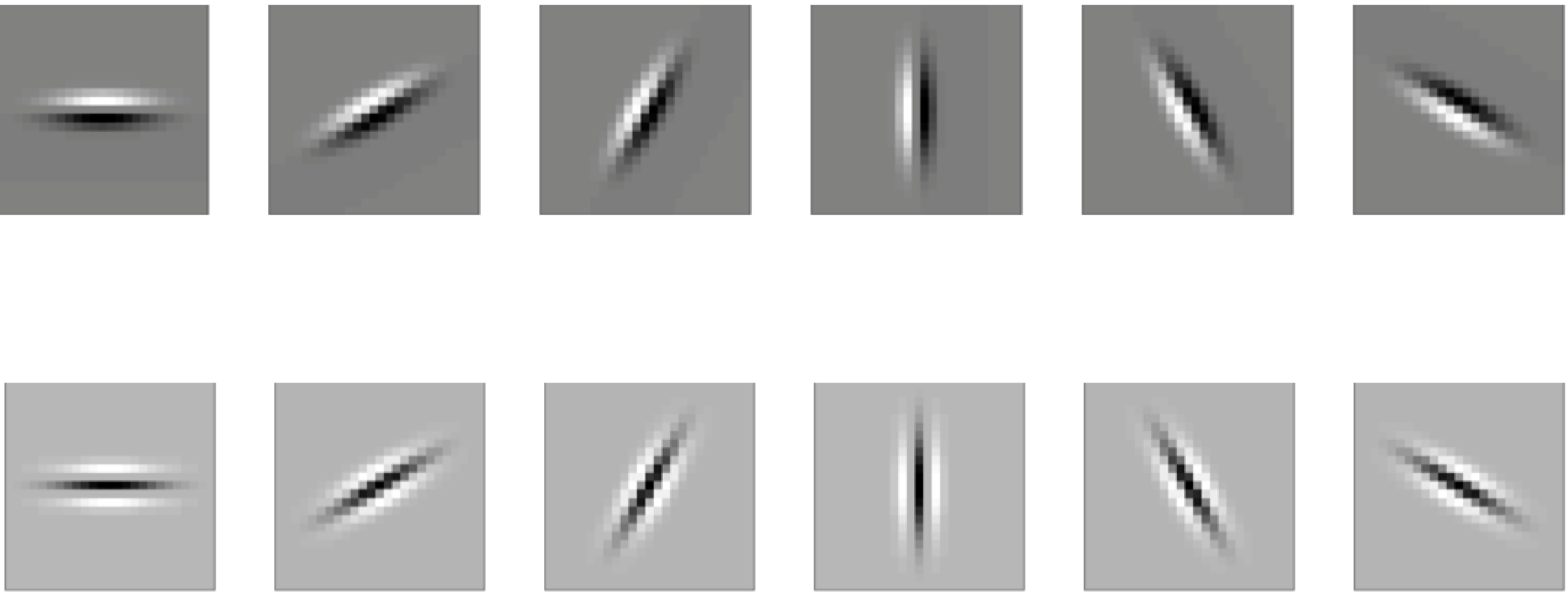
$$Rot_\theta f_1 = G'_{\sigma_1}(u) G_{\sigma_2}(v) \tag{10.6}$$
$$Rot_\theta f_2 = G''_{\sigma_1}(u) G_{\sigma_2}(v) \tag{10.7}$$

where we set

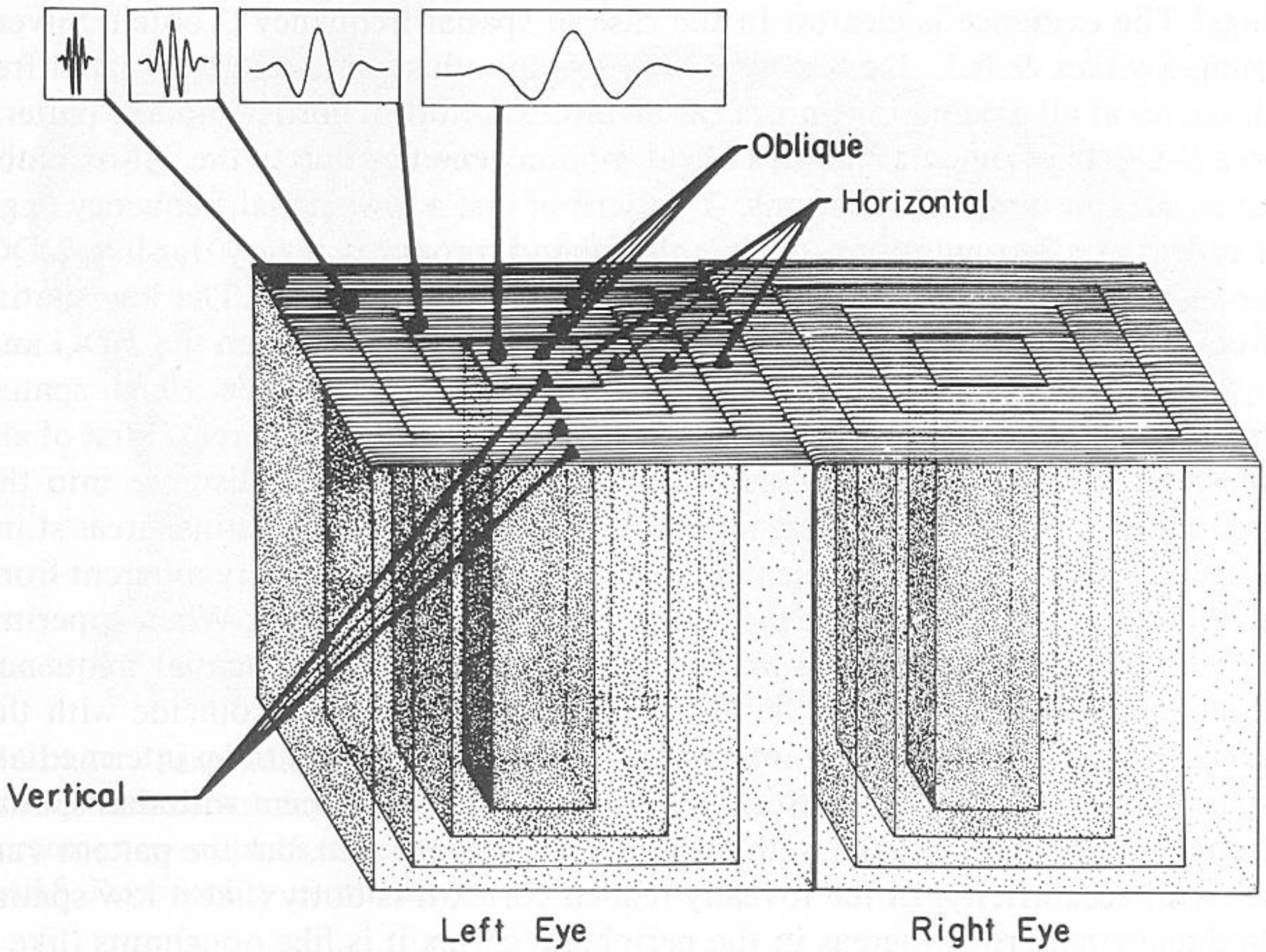
$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

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

Oriented Gaussian First and Second Derivatives



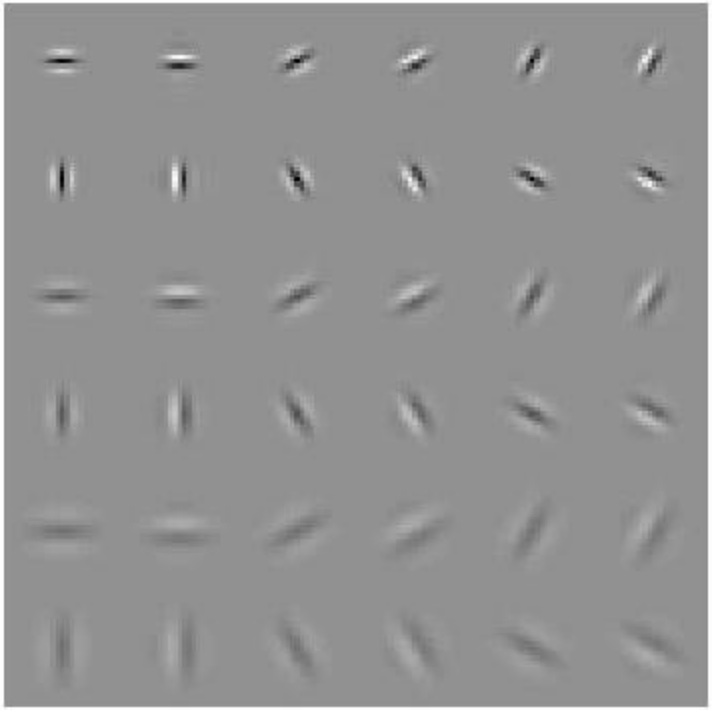
Hypercolumns in visual cortex



Model of Striate Module in Monkeys

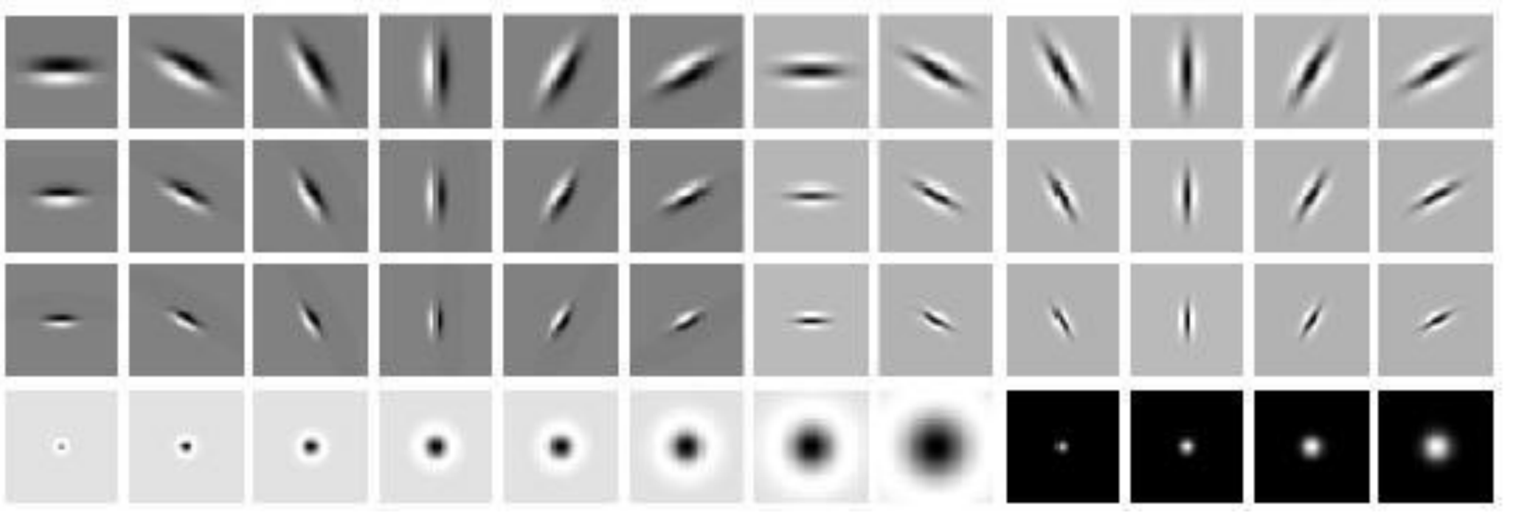
Modeling hypercolumns

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



Overcomplete representation: filter banks

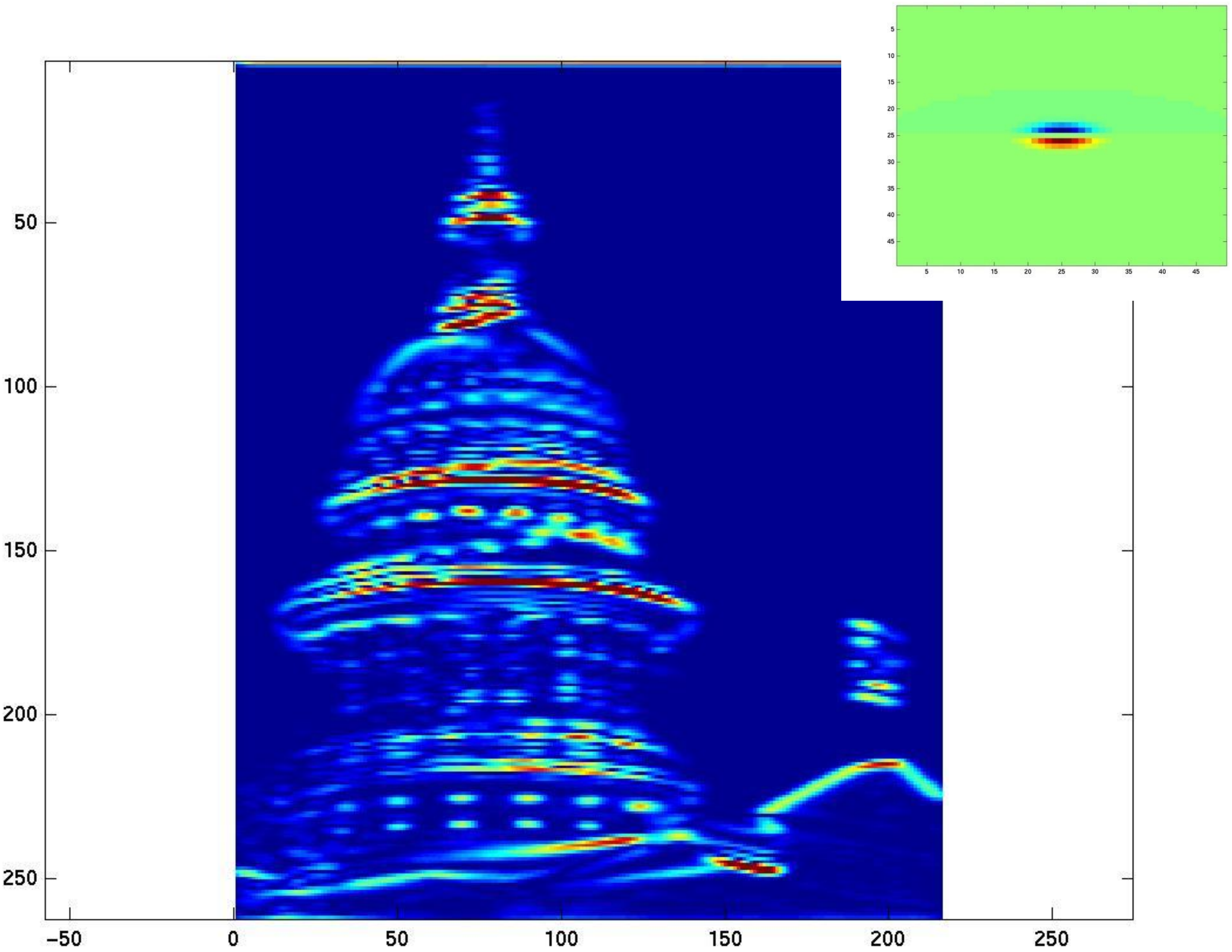
LM Filter Bank



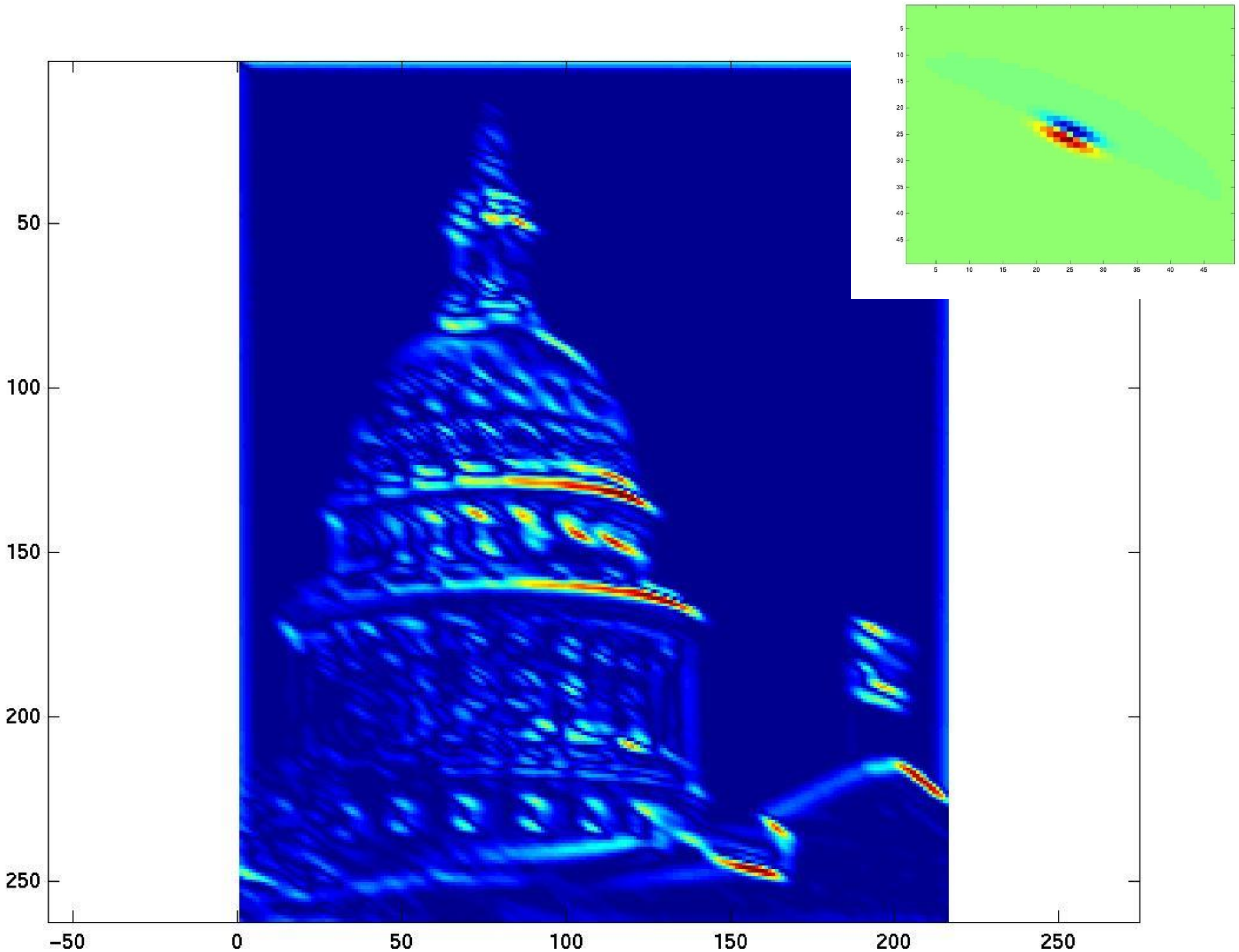
Code for filter banks: *www.robots.ox.ac.uk/~vgg/research/texclass/filters.html*

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

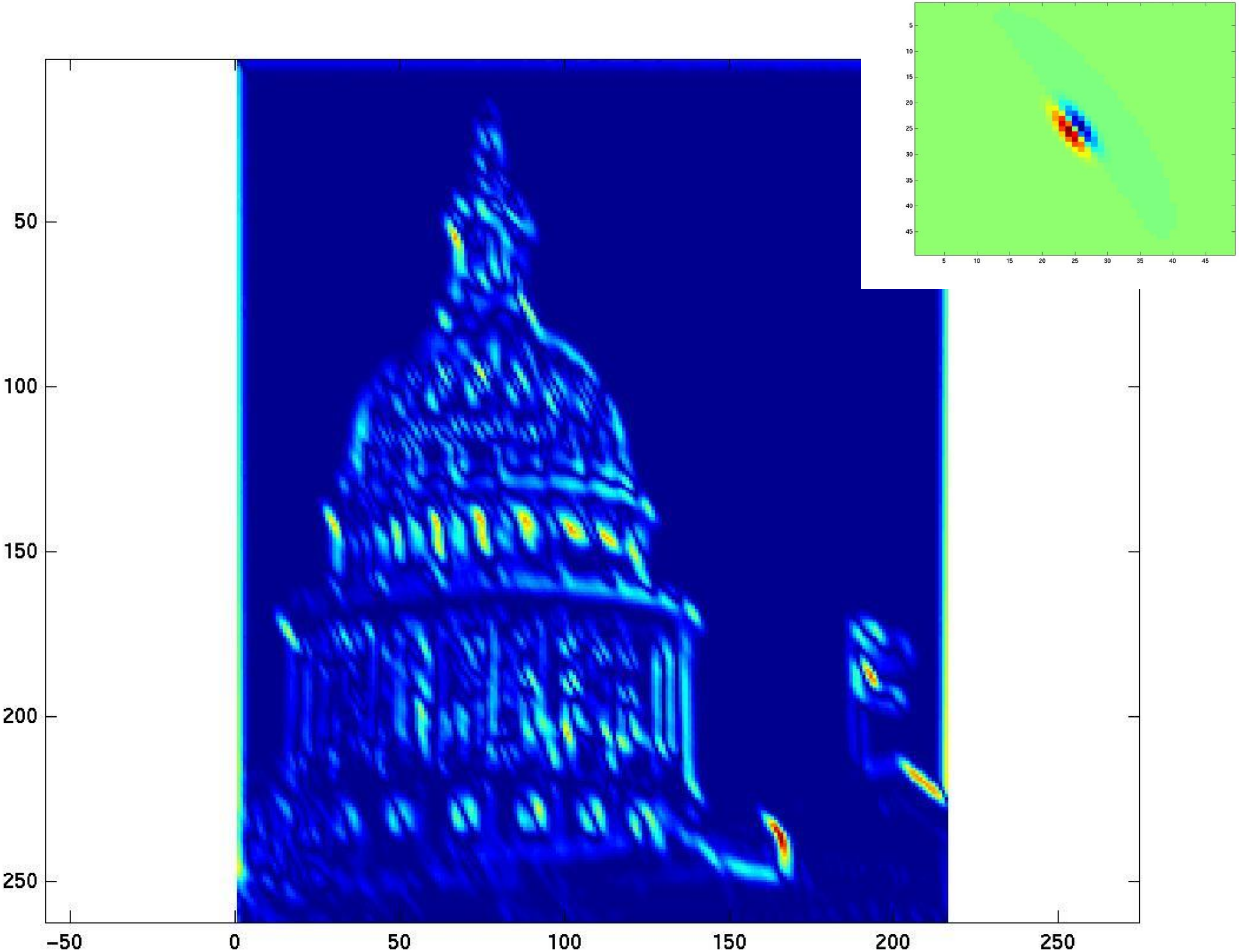
Kristen Grauman



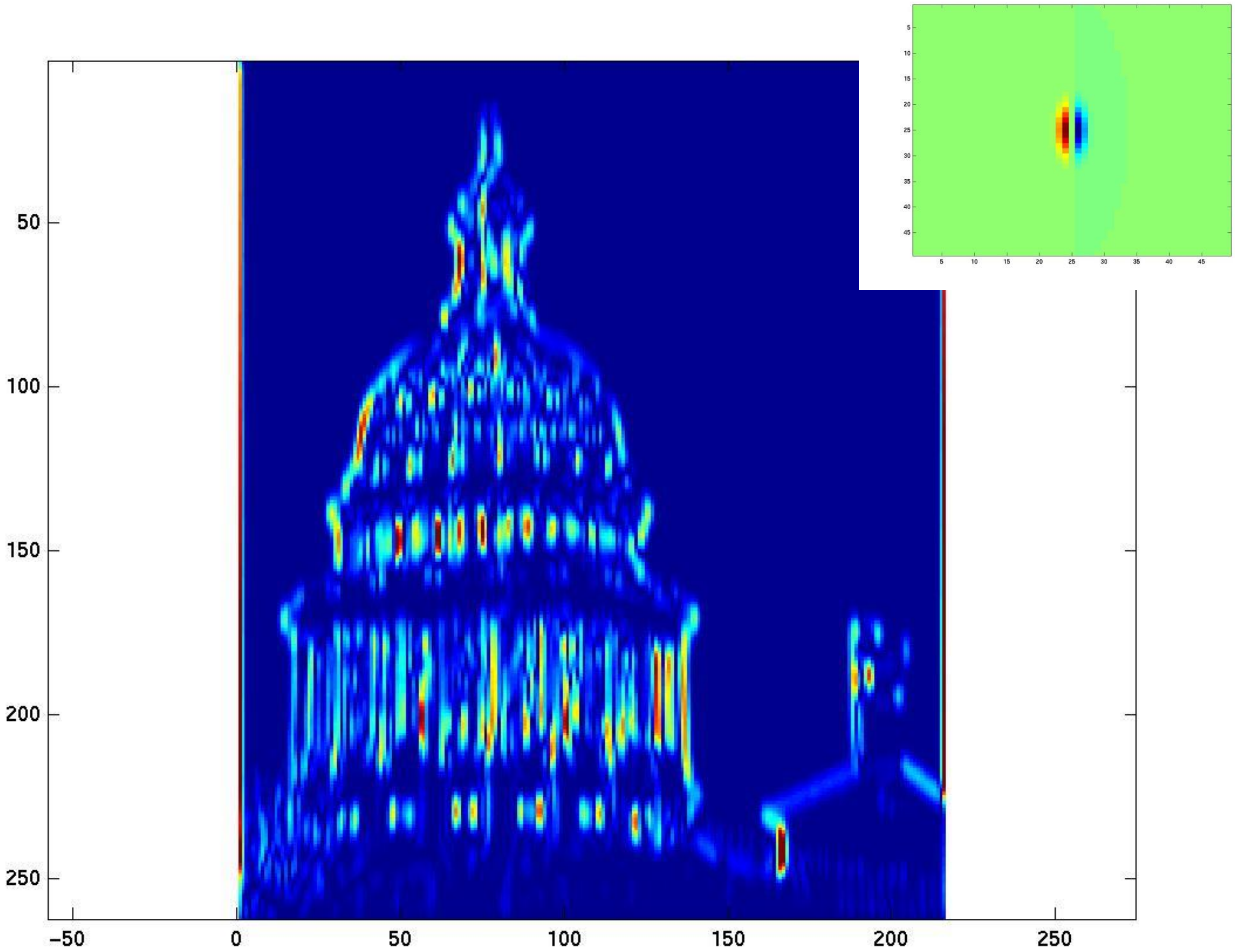
Kristen Grauman



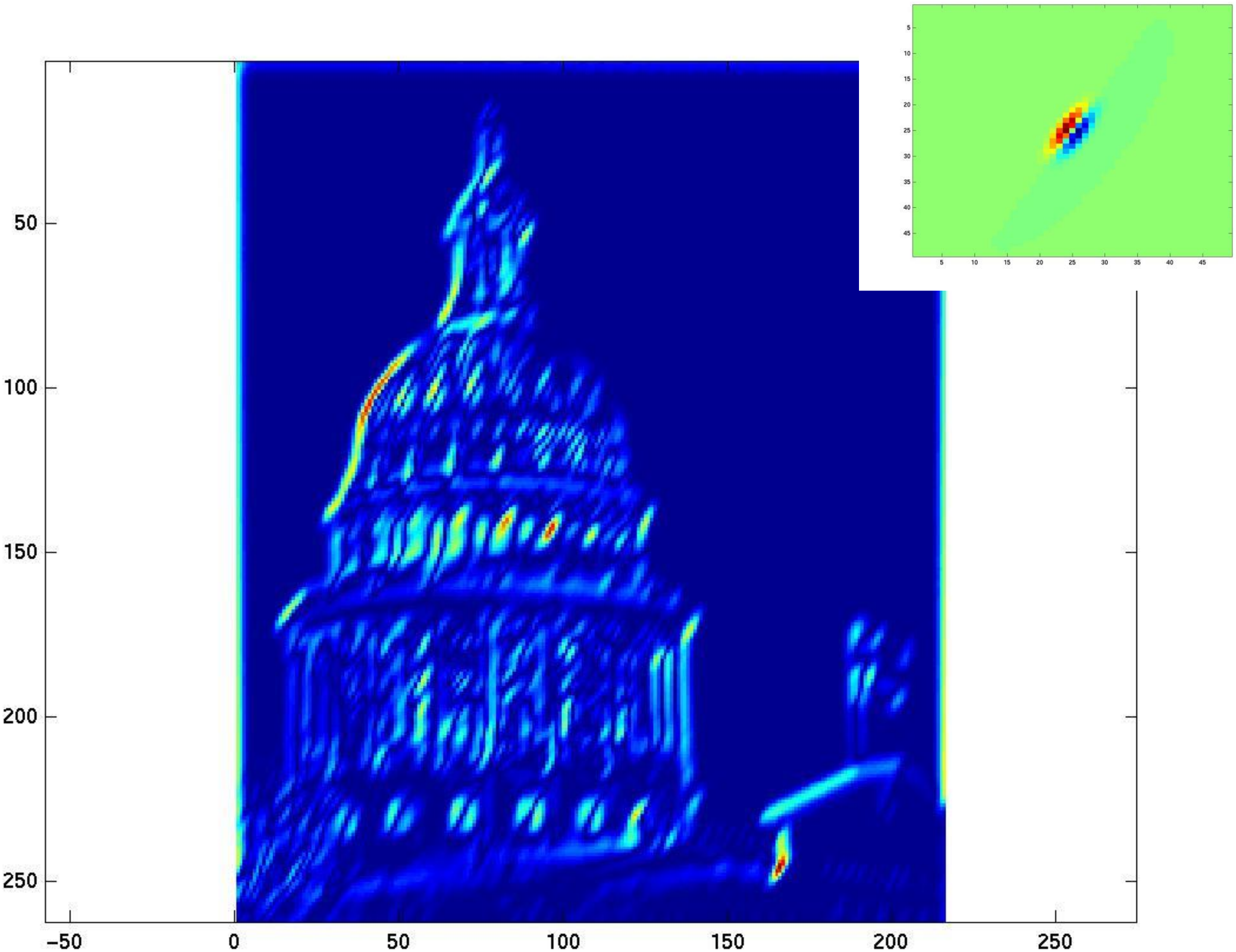
Kristen Grauman



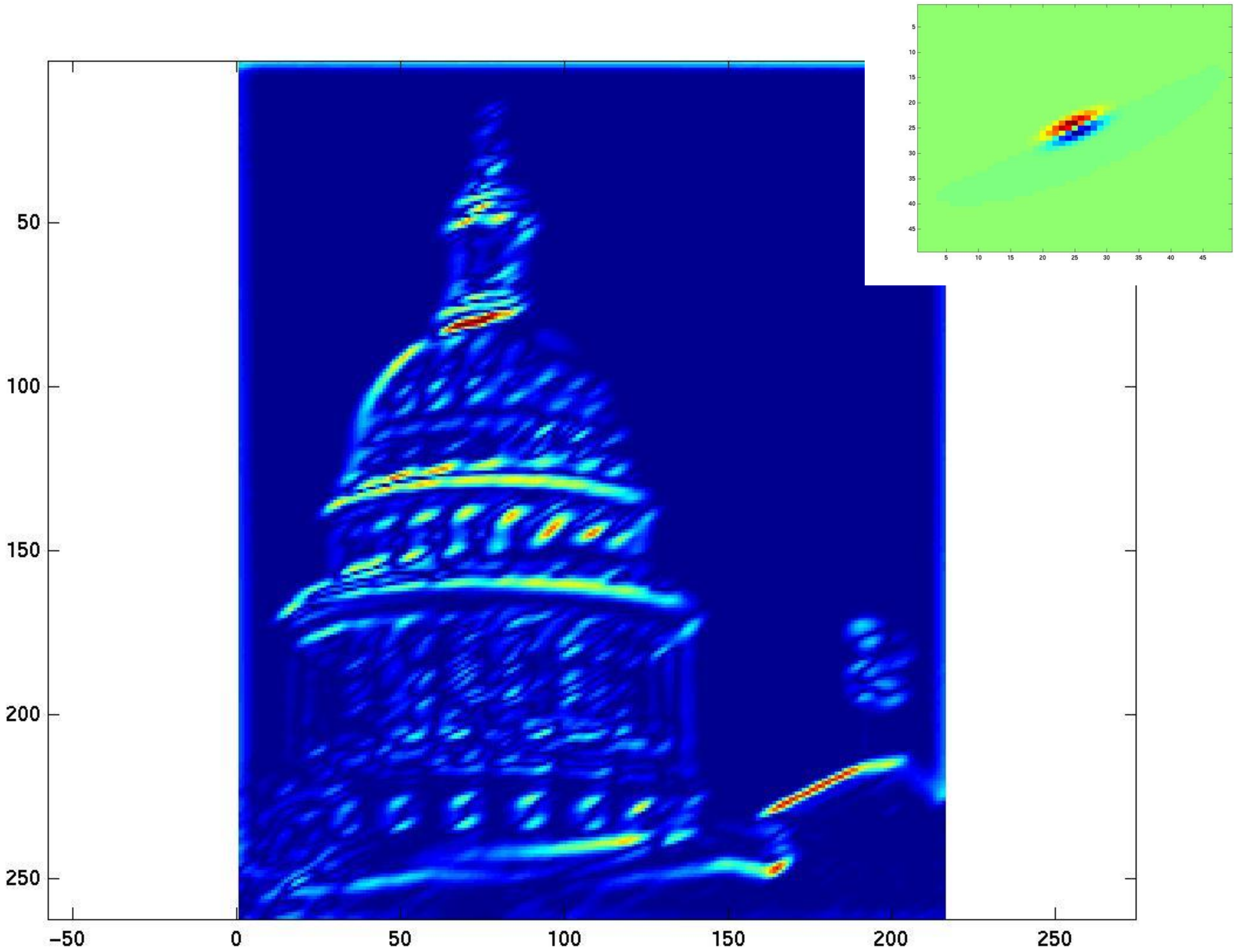
Kristen Grauman



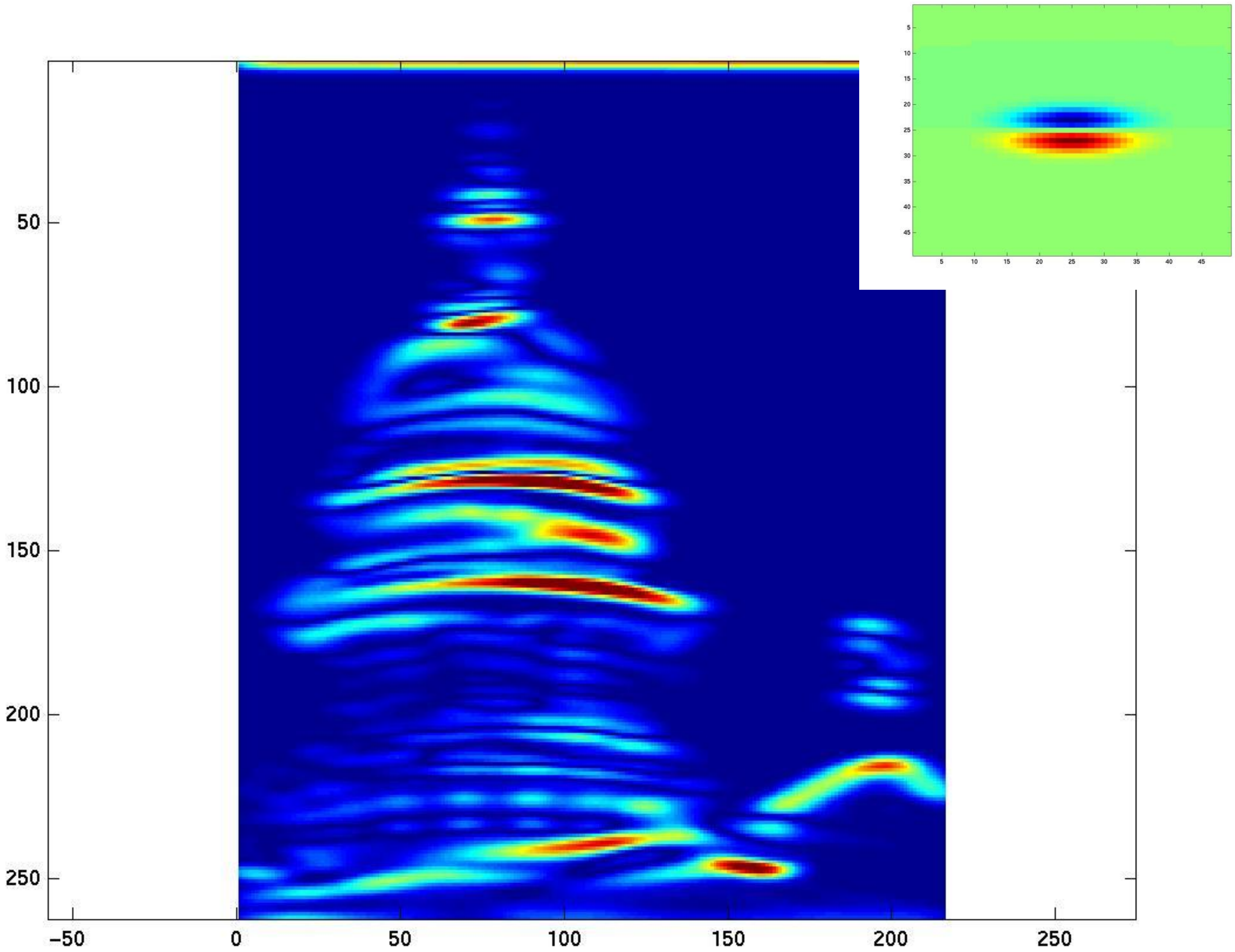
Kristen Grauman



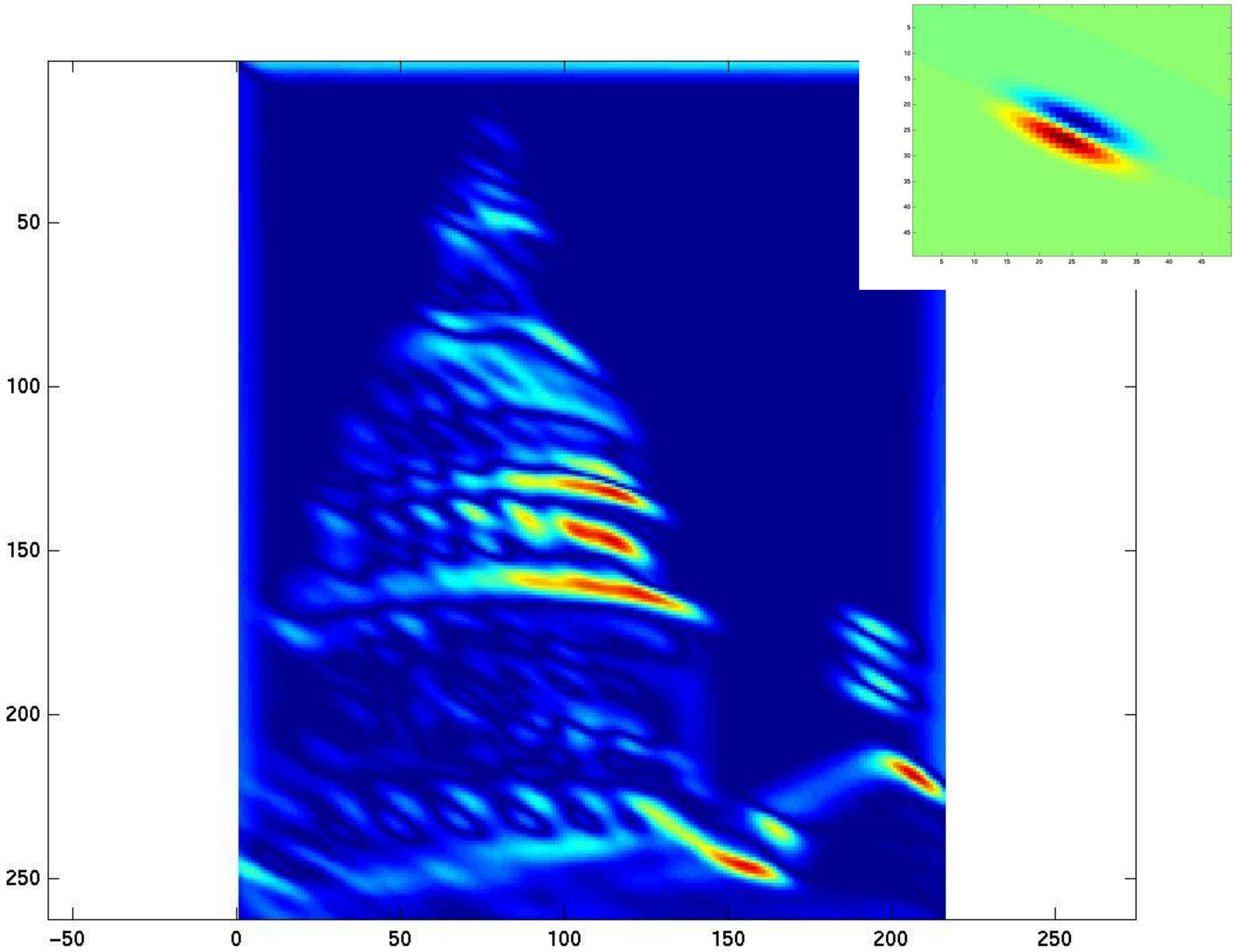
Kristen Grauman



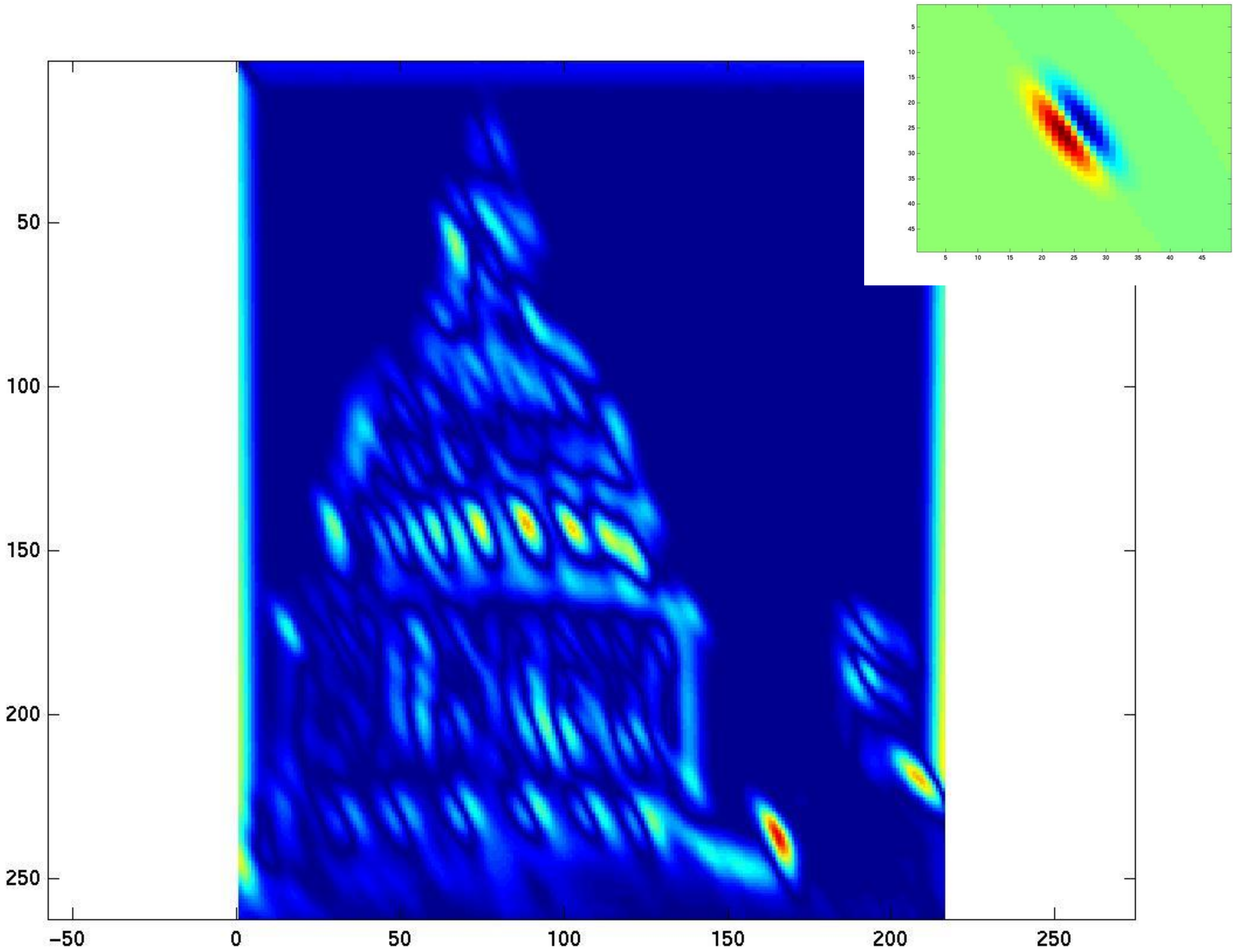
Kristen Grauman



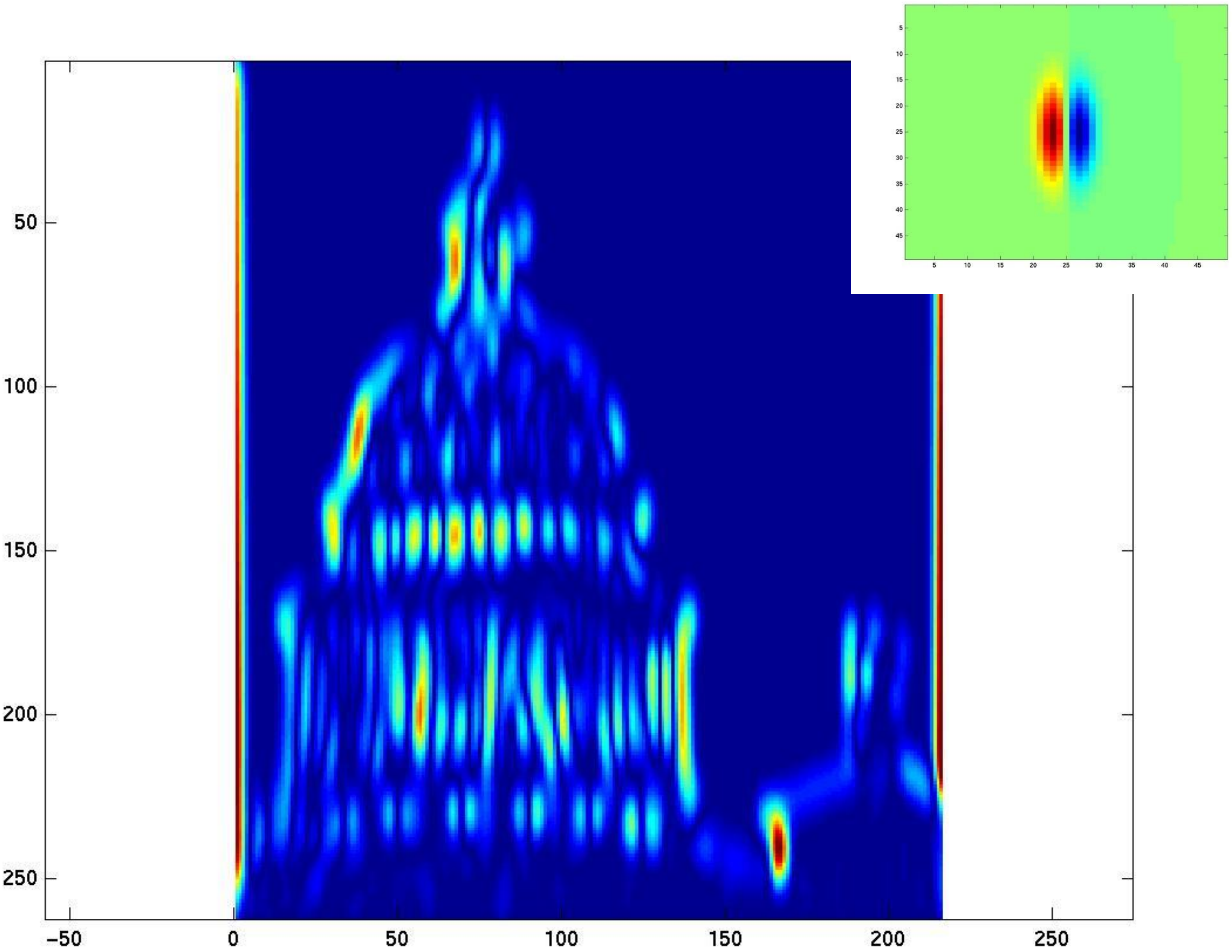
Kristen Grauman



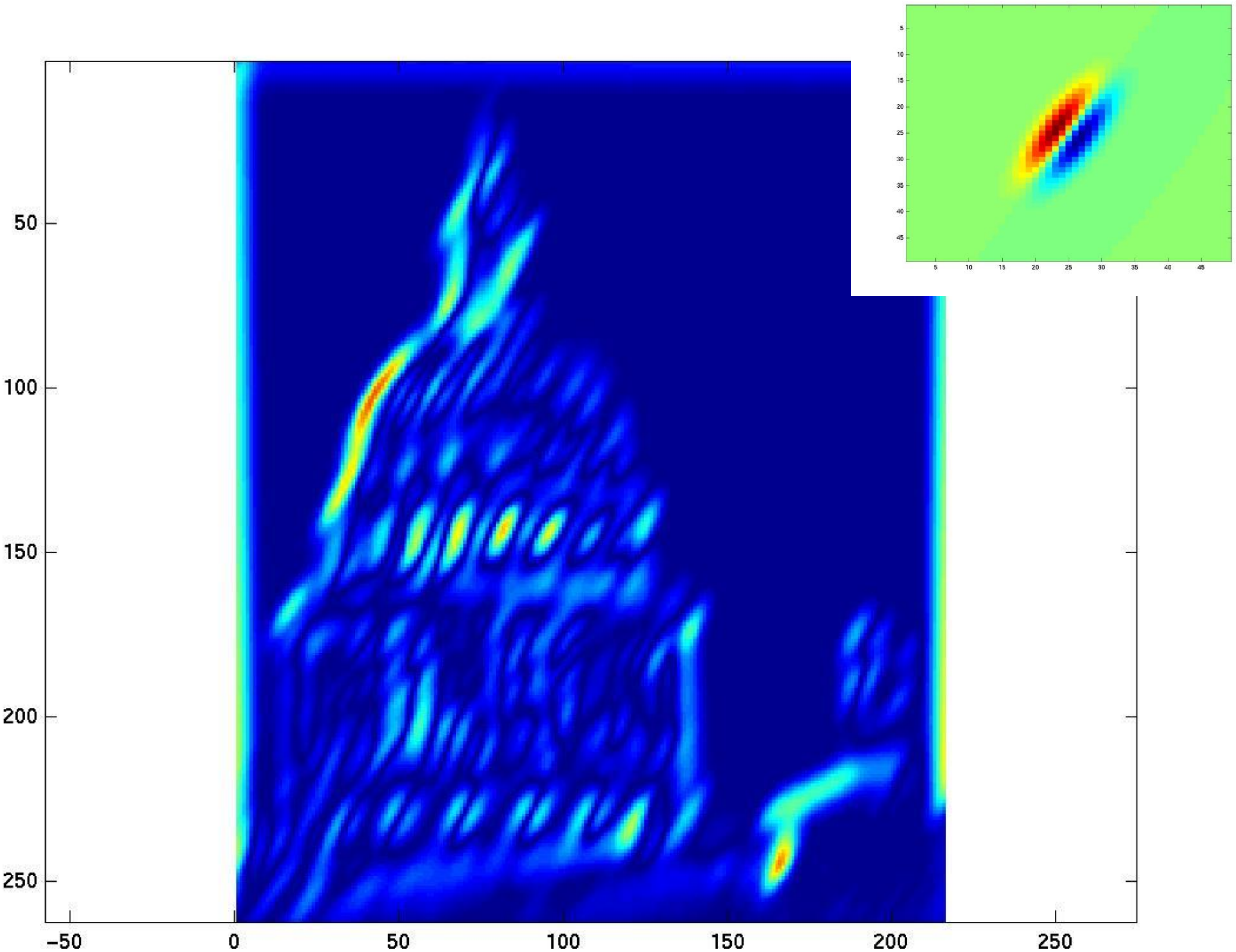
Kristen Grauman



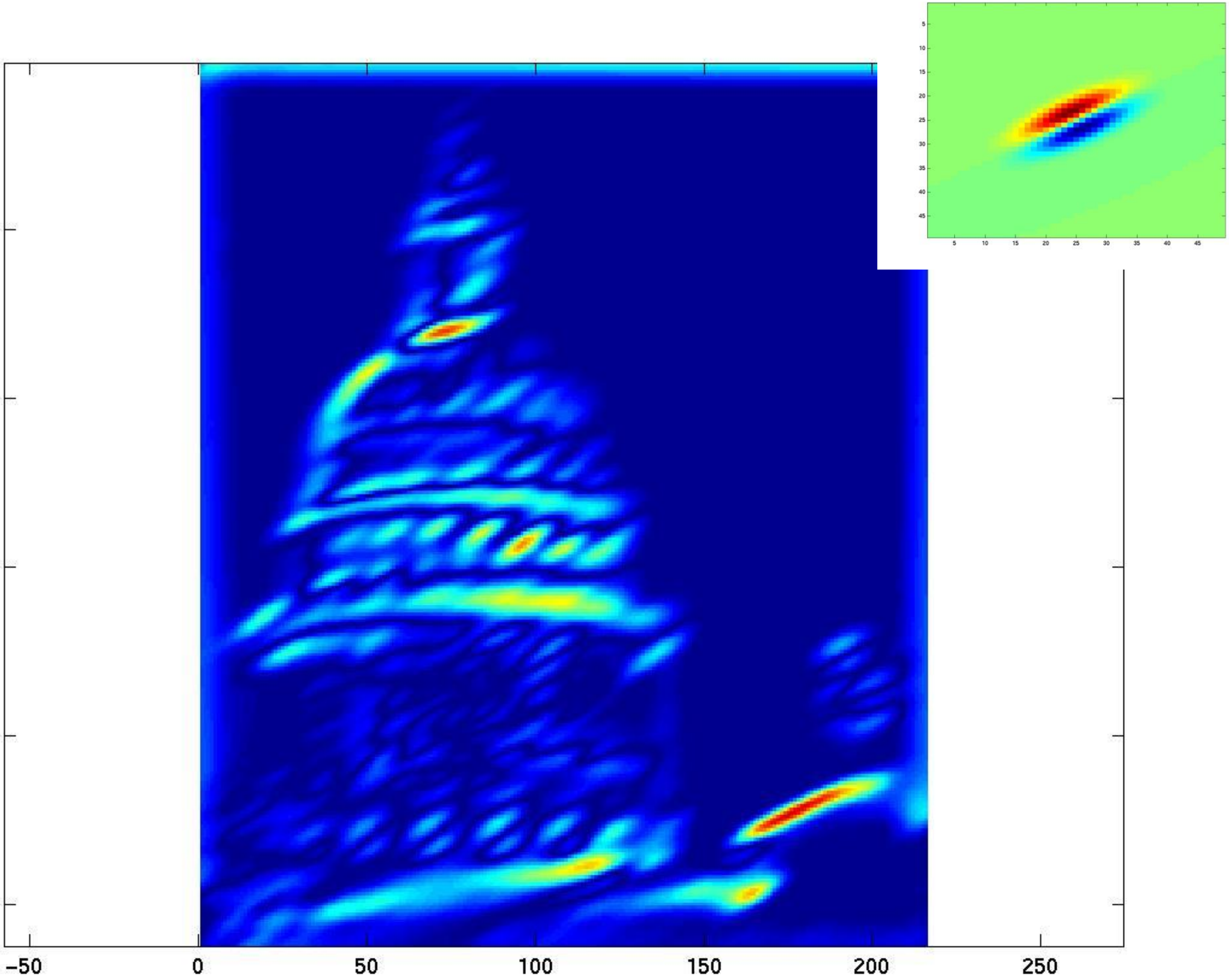
Kristen Grauman



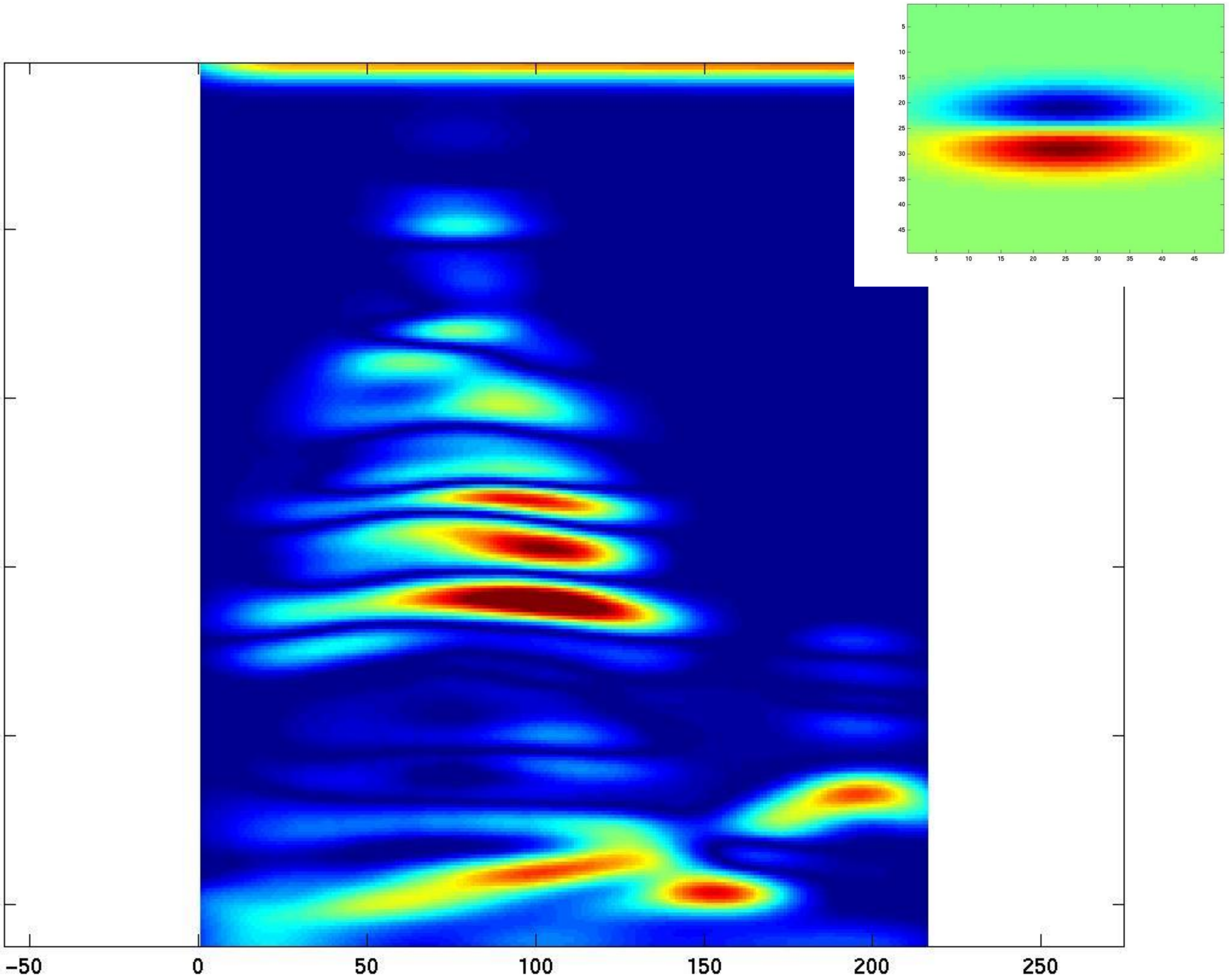
Kristen Grauman



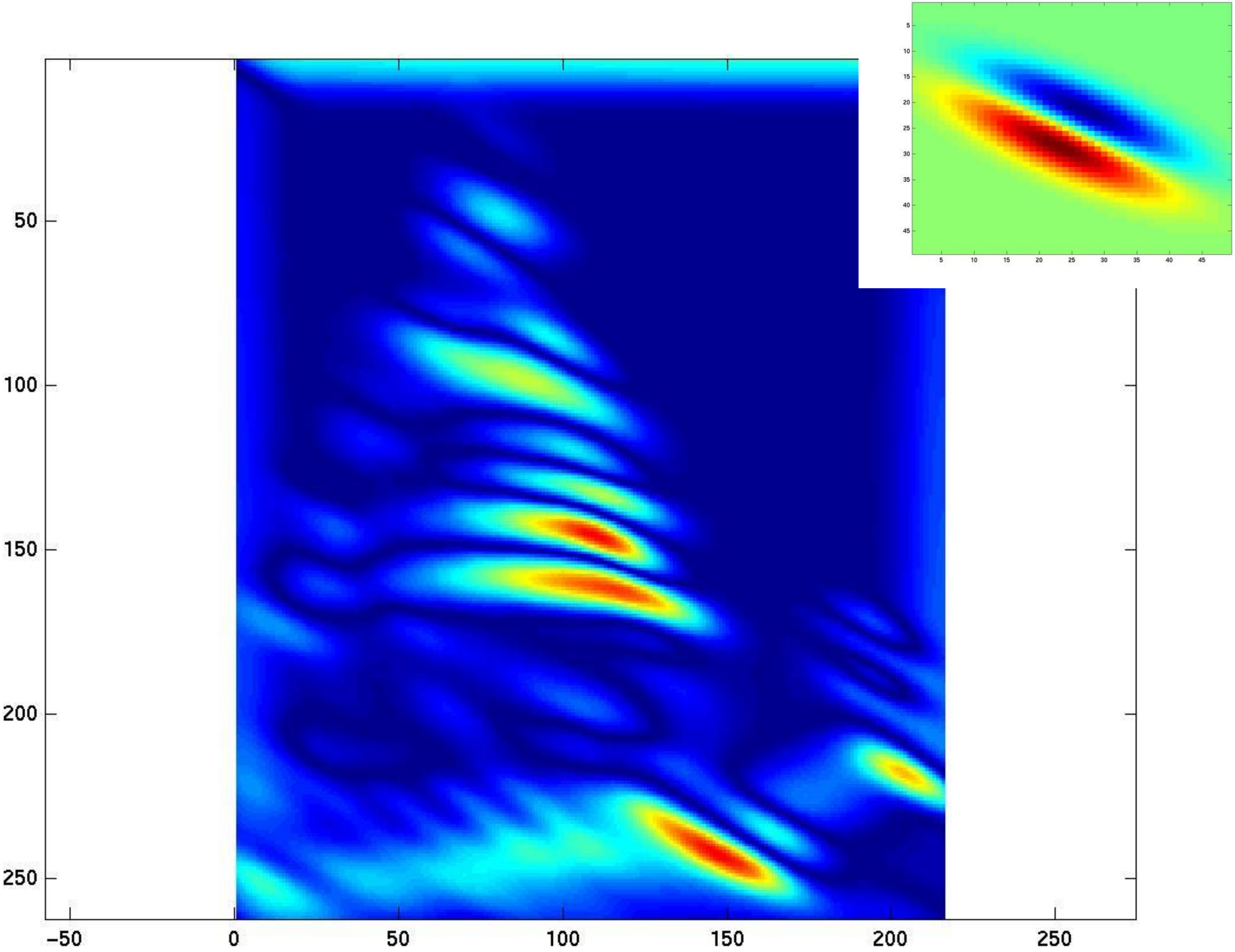
Kristen Grauman



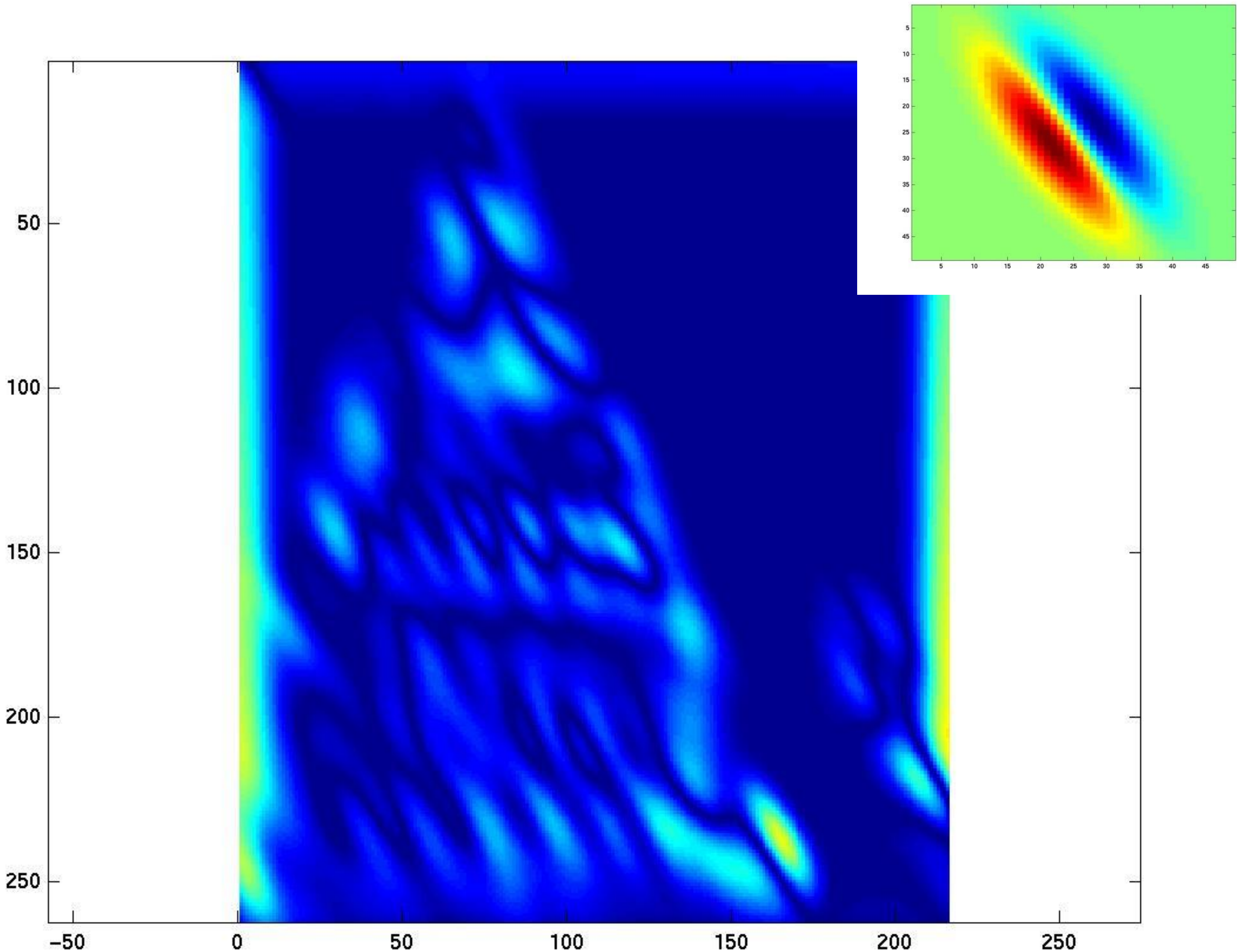
Kristen Grauman



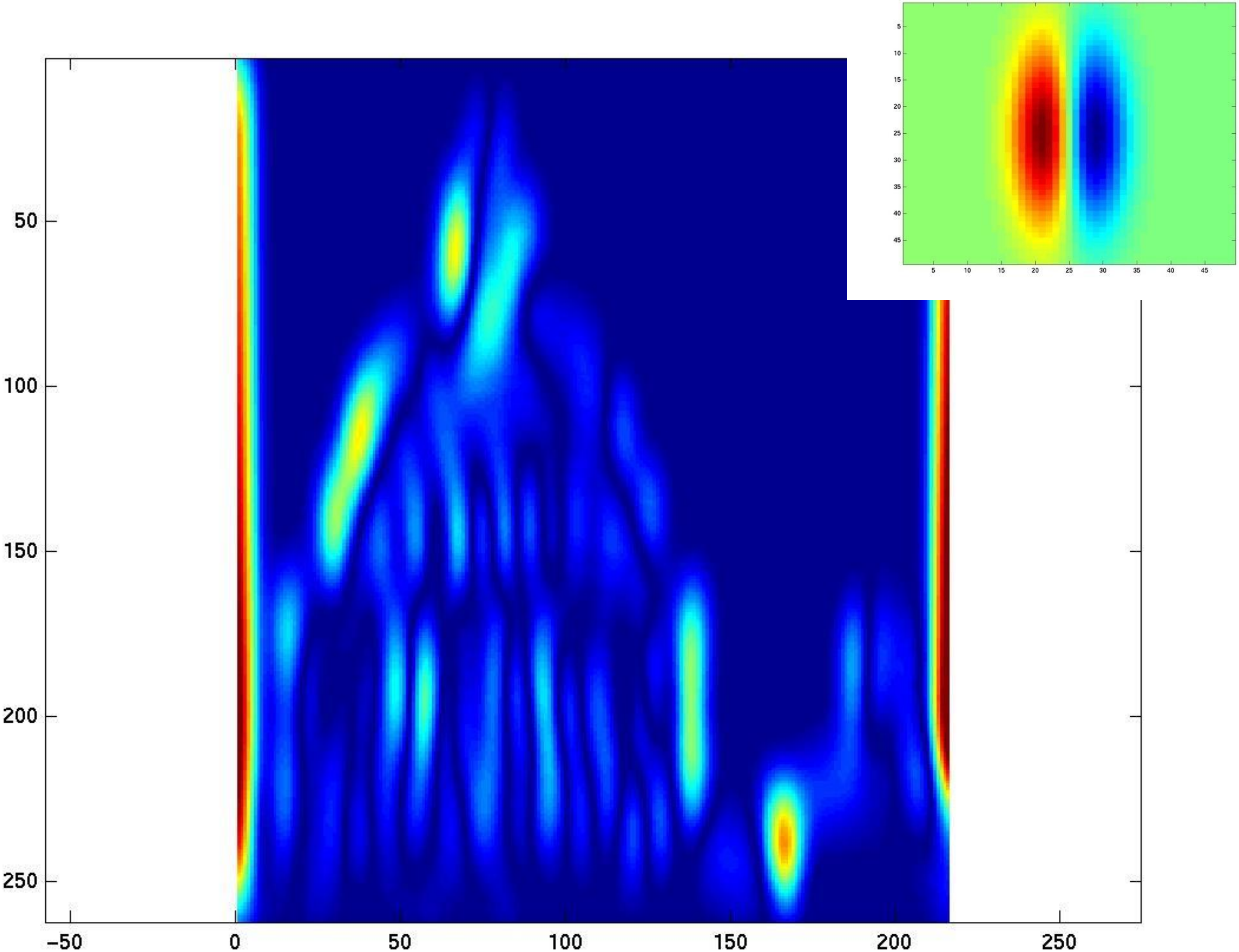
Kristen Grauman



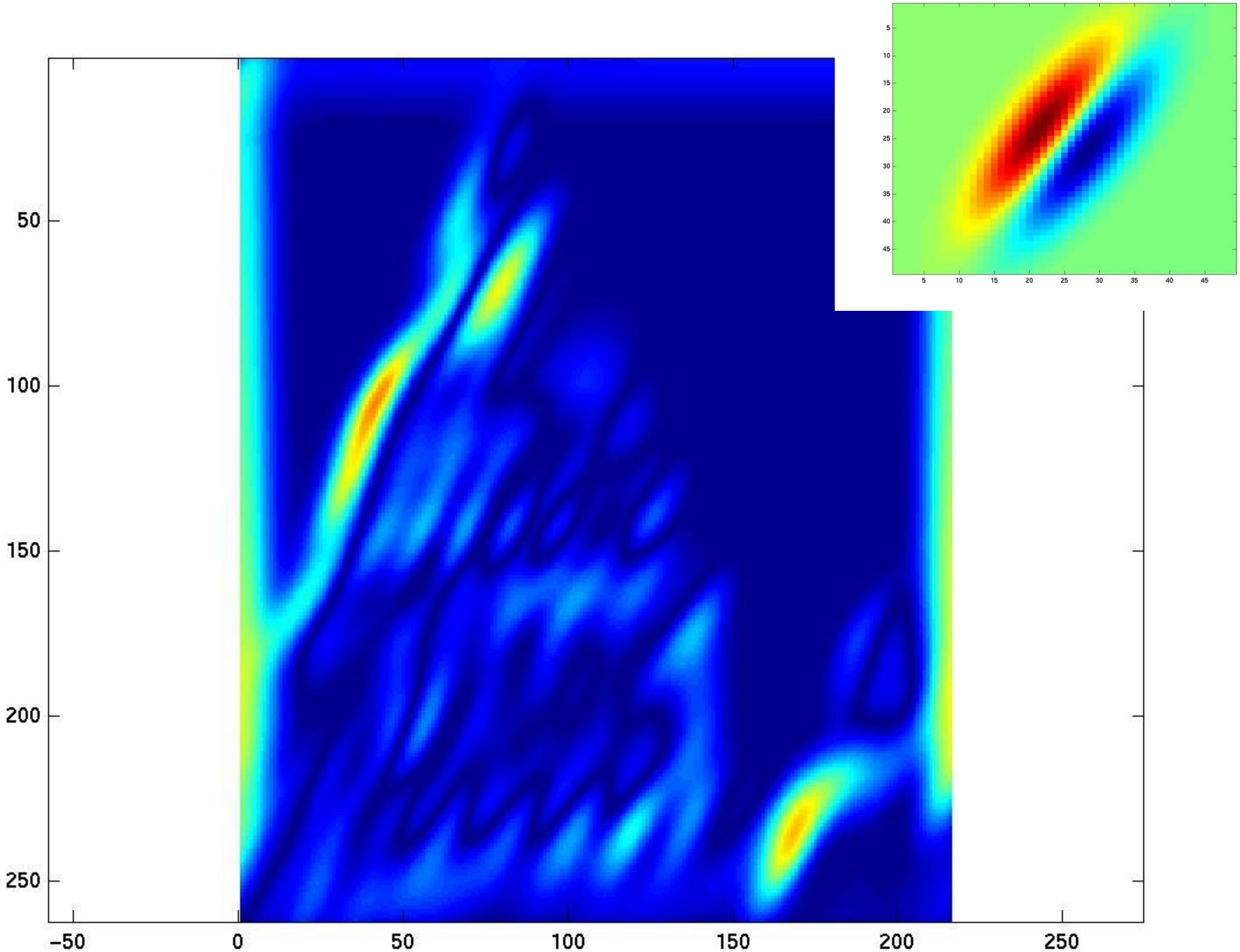
Kristen Grauman



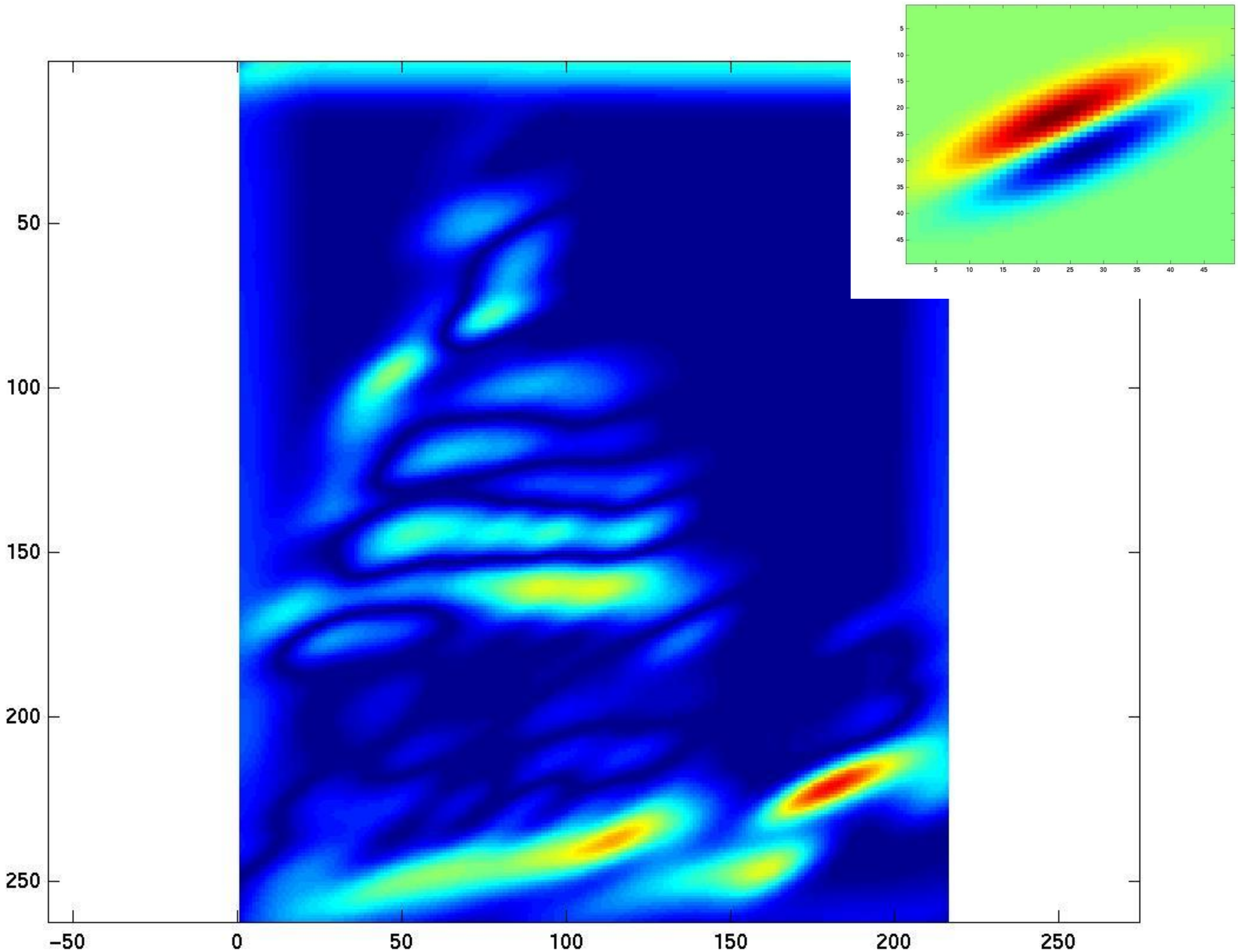
Kristen Grauman



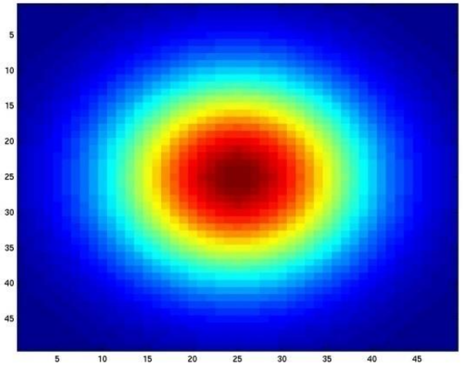
Kristen Grauman



Kristen Grauman



Kristen Grauman

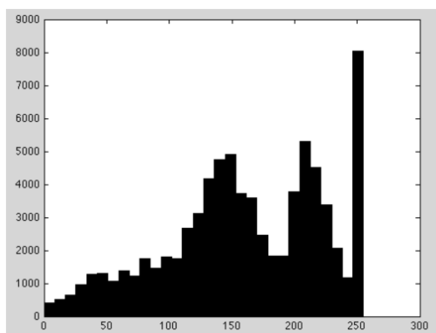

50
100
150
200
250
300
350
400
450
50
100
150
200
250
300
350


Two questions of texture modeling

- What are the texture features (textons)?
  - Pixels
  - Pixel patches
  - Outputs of V1-like filters
  - Clusters of patches / filter outputs
  - CNN features
  - Etc.

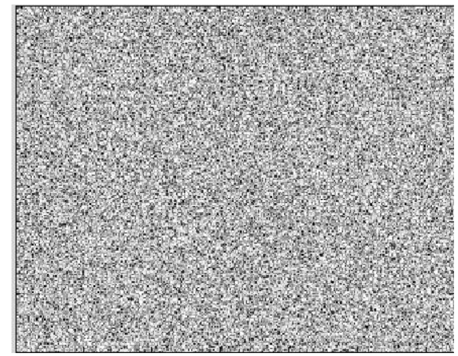
- How do we aggregate statistics
  - Various types of histograms
  - Implicit or explicit

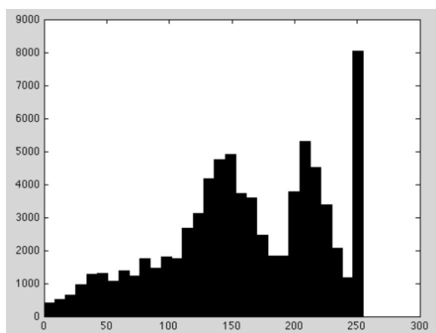
Pixel Histograms

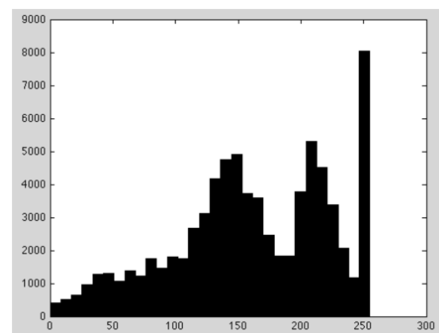


Slide by Erik Learned-Miller

Pixel Histograms

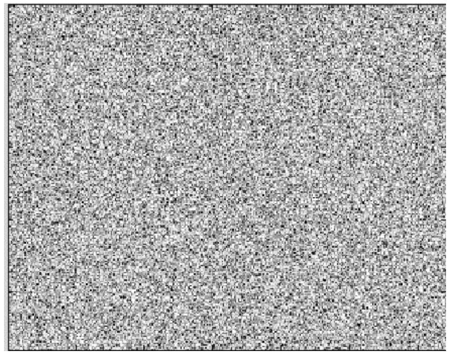






Slide by Erik Learned-Miller

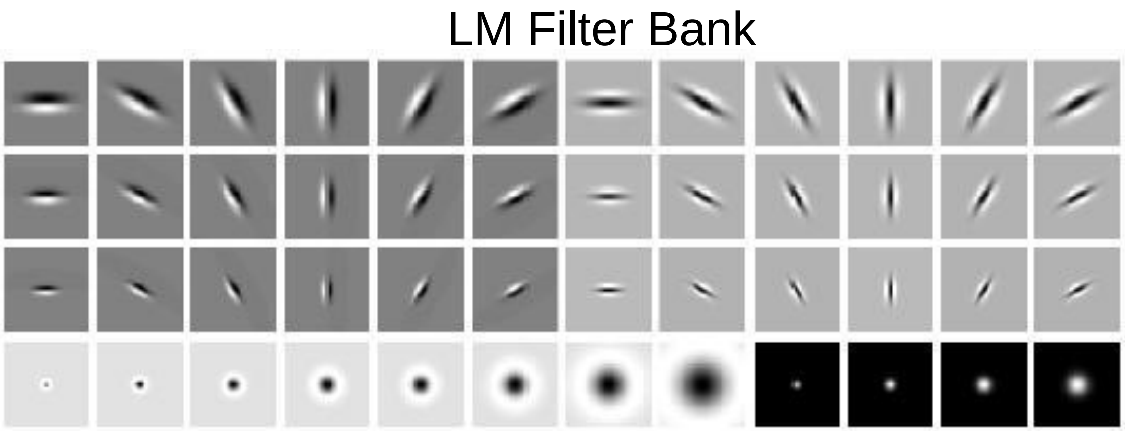
Gray value histogram comparisons



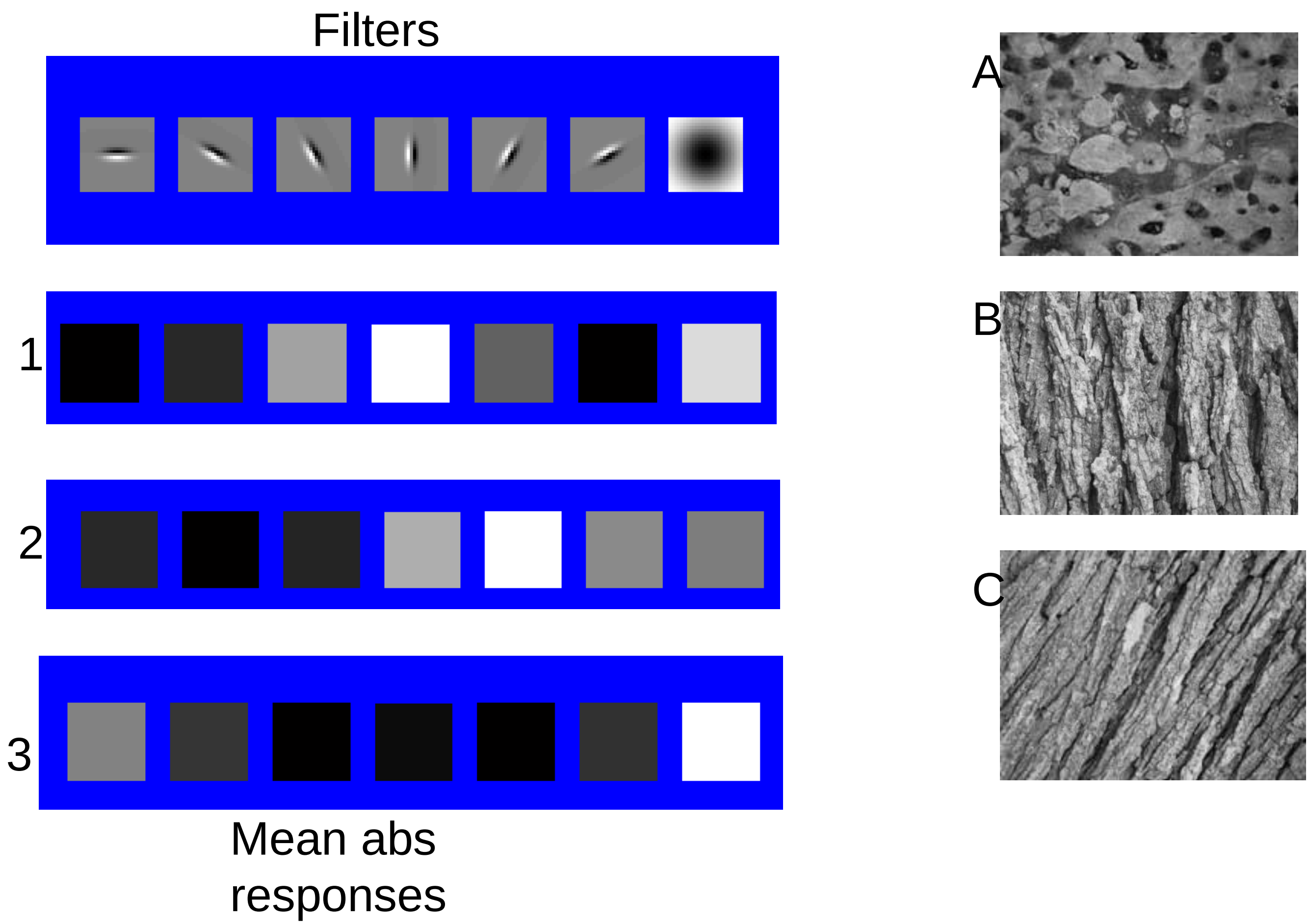
They’re equal

Slide by Erik Learned-Miller

Going up from pixels: V1 filter-banks

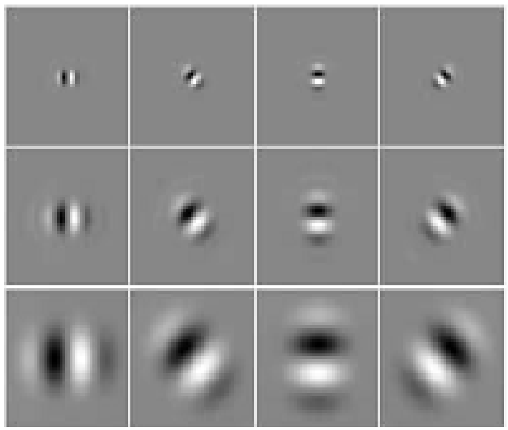


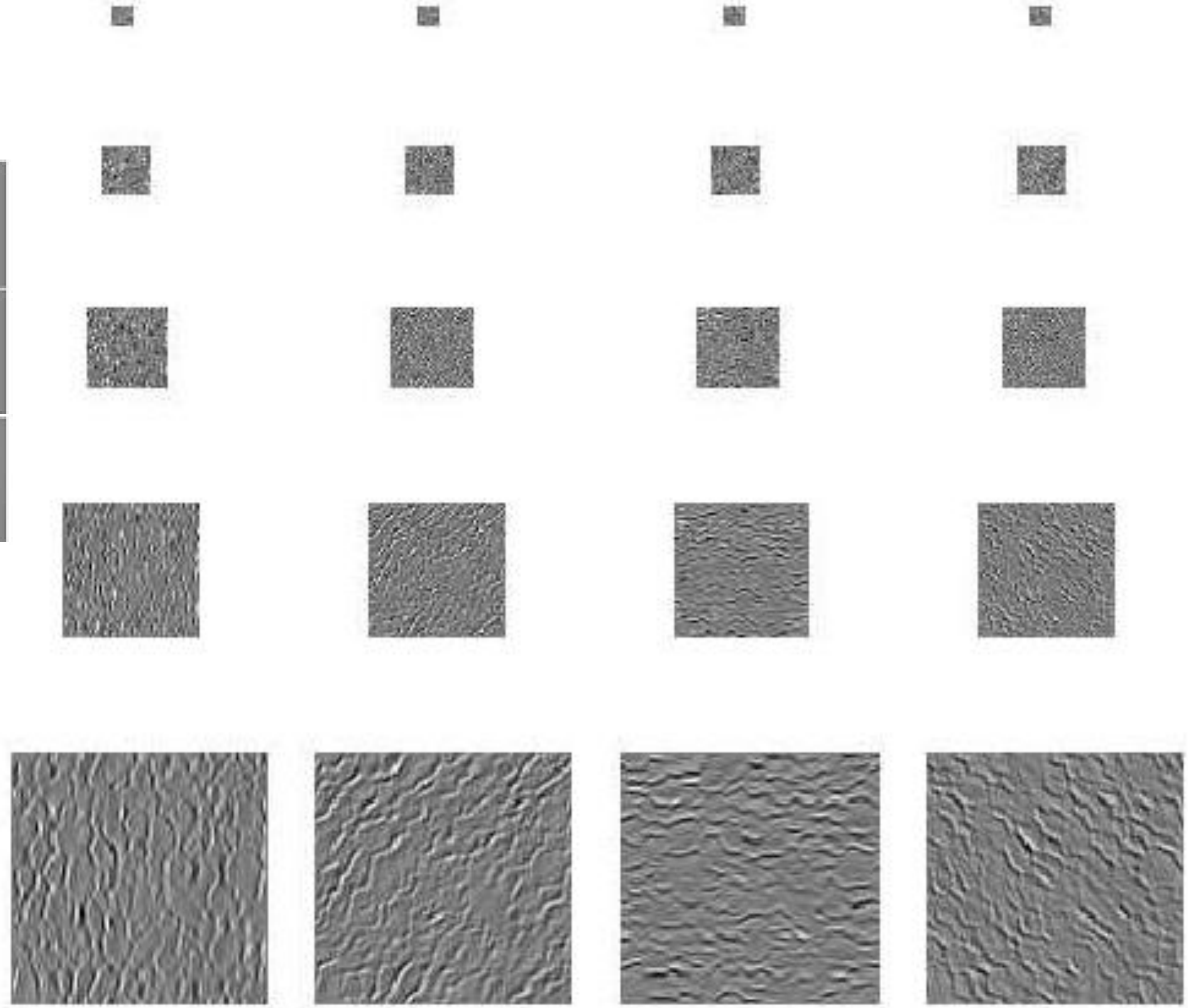
Can you match the texture to its histogram?



Slightly fancier: histogram for each filter

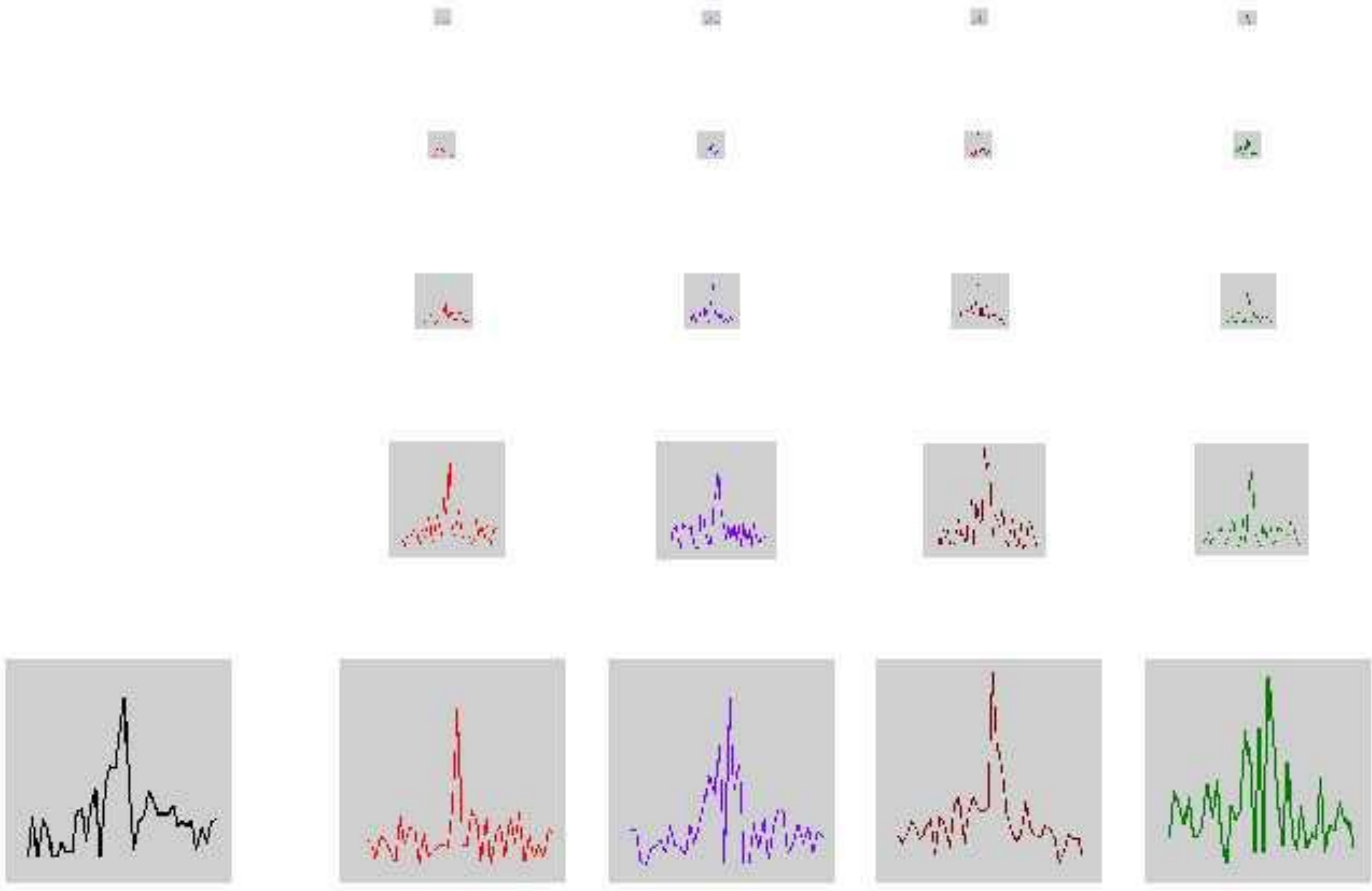
Filter bank



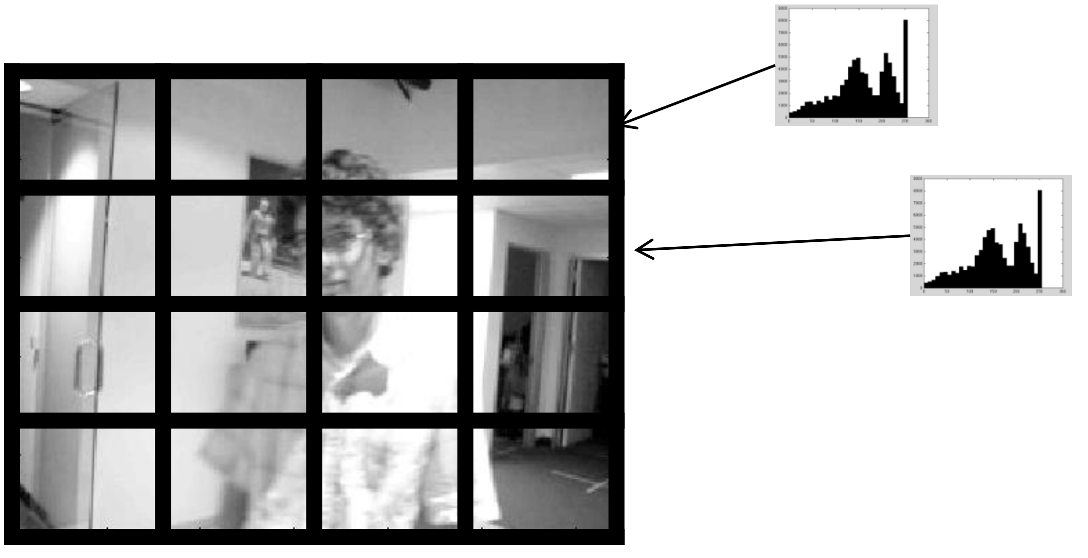


Input image

Filter response histograms



Adding spatial structure

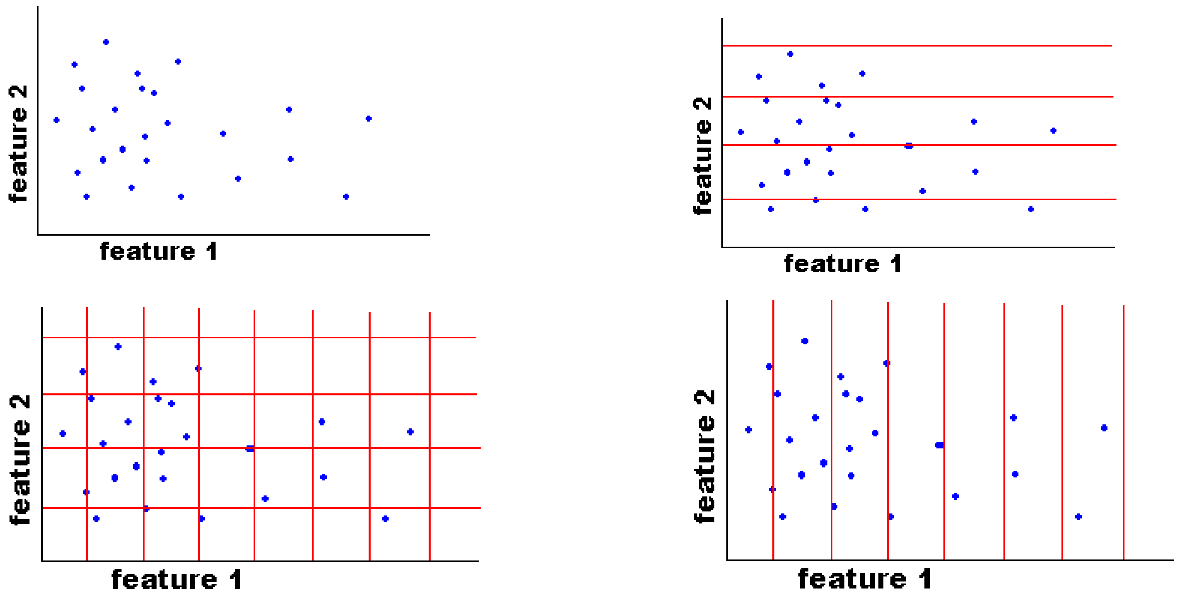


A separate histogram for each region.

Slide by Erik Learned-Miller

Image Representations: Histograms

Images from Dave



Joint histogram

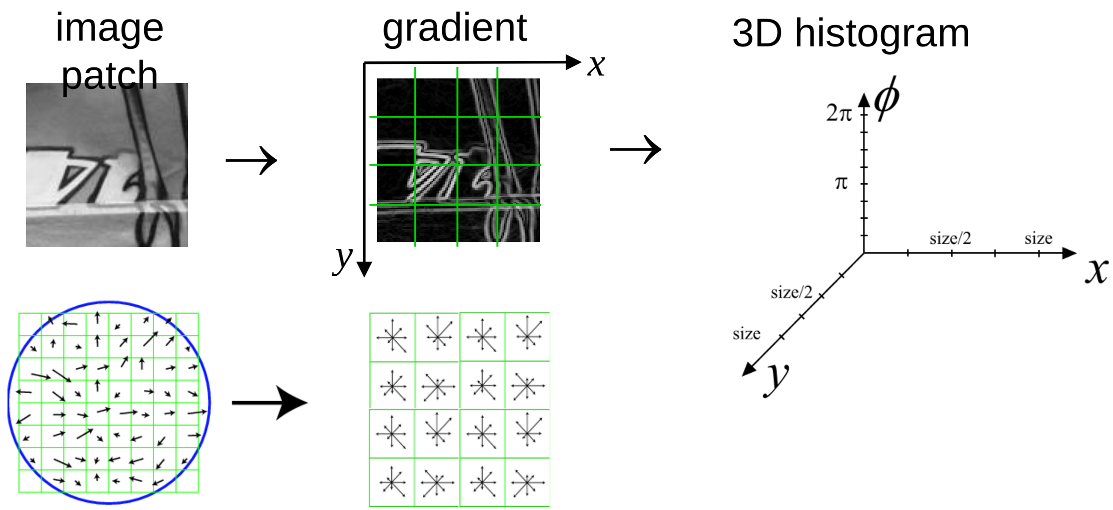
- Requires lots of data
- Loss of resolution to avoid empty bins

Marginal histogram

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

Ex: SIFT descriptor [Lowe'99]

distribution of the gradient over an image patch

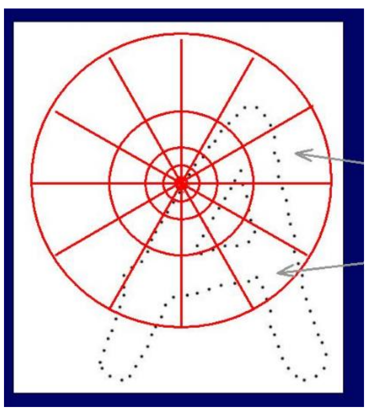


4x4 location grid and 8 orientations (128 dimensions)

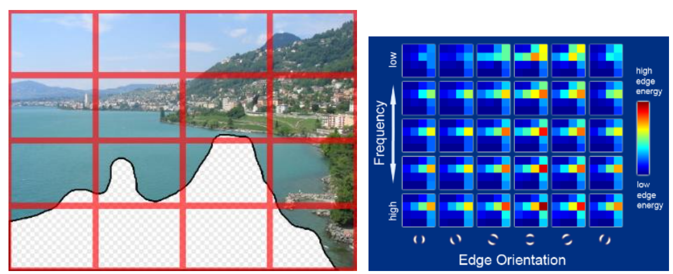
very good performance in image matching [Mikolaczyk and Schmid'03]

Gradient Histograms pop-up everywhere

HOG descriptor



Generalized Shape Context



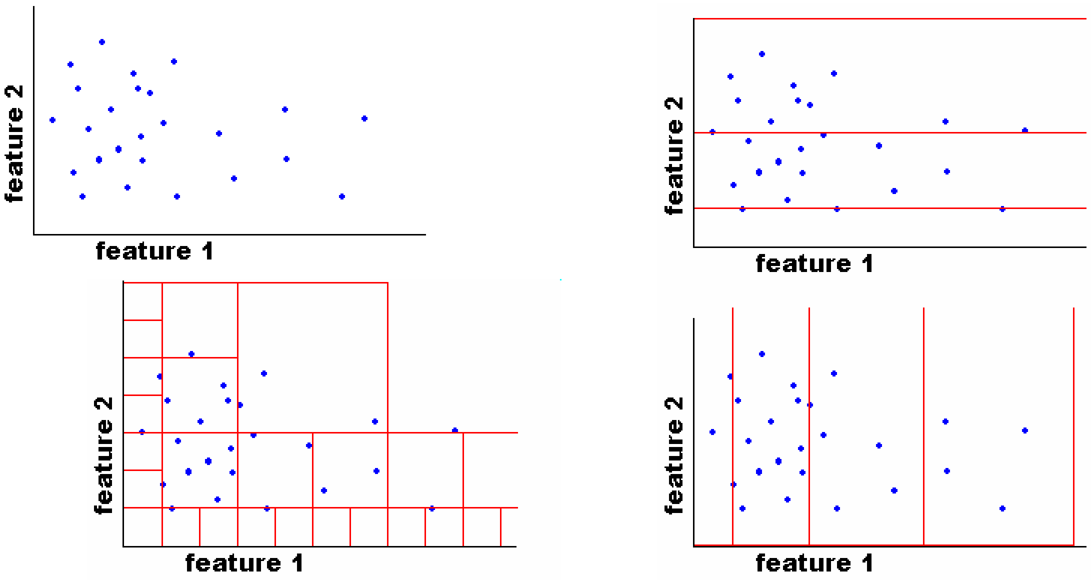
Gist Descriptor

Freeman and Roth IAFGR 1995
Lowe ICCV1999
Oliva & Torralba, 2001
Belongie et al, 2001
Dalal &Triggs CVPR05

Binning achieves invariance to small patch offsets

Adaptive Representations

Images from Dave

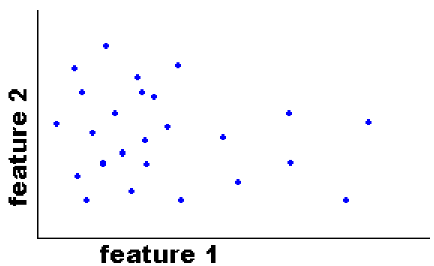


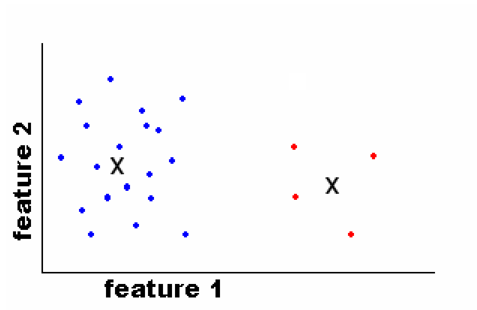
Adaptive binning

- Better data/bin distribution, fewer empty bins
- Can adapt available resolution to relative feature importance

Clustering: very adaptive representations

Images from Dave Kauchak



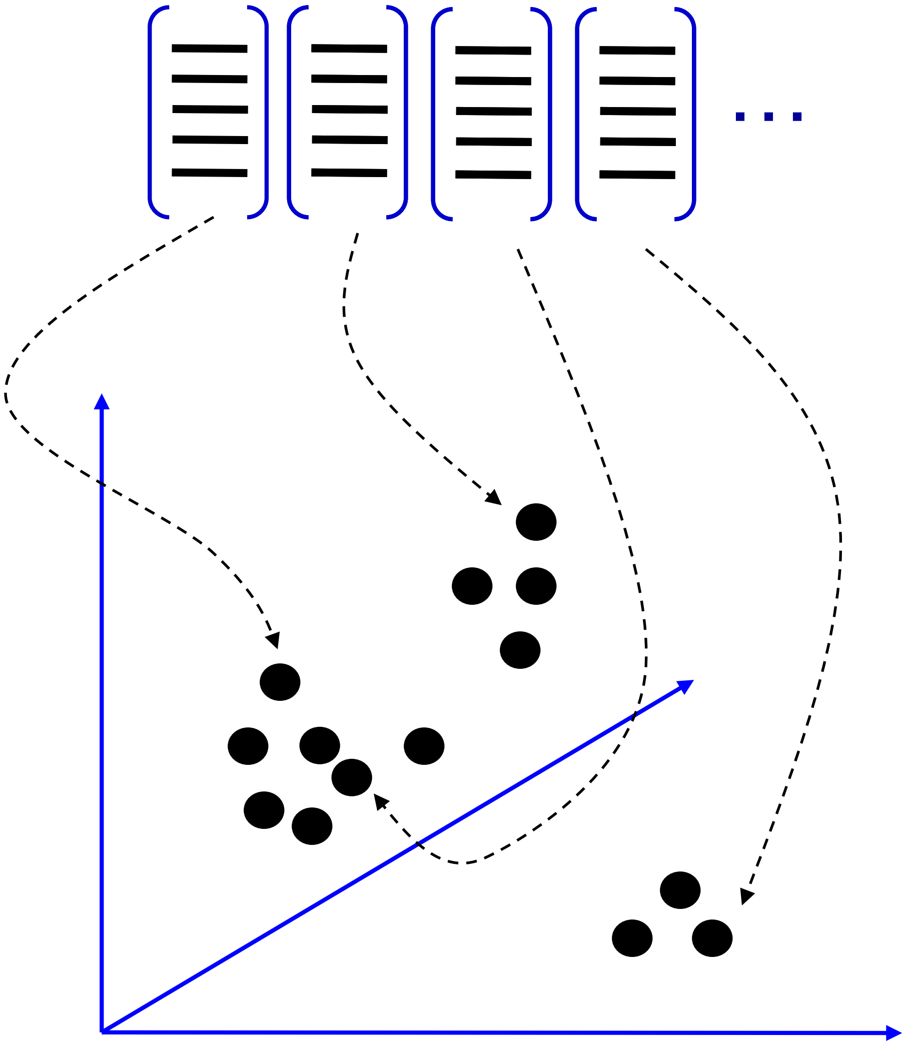


Clusters / Signatures

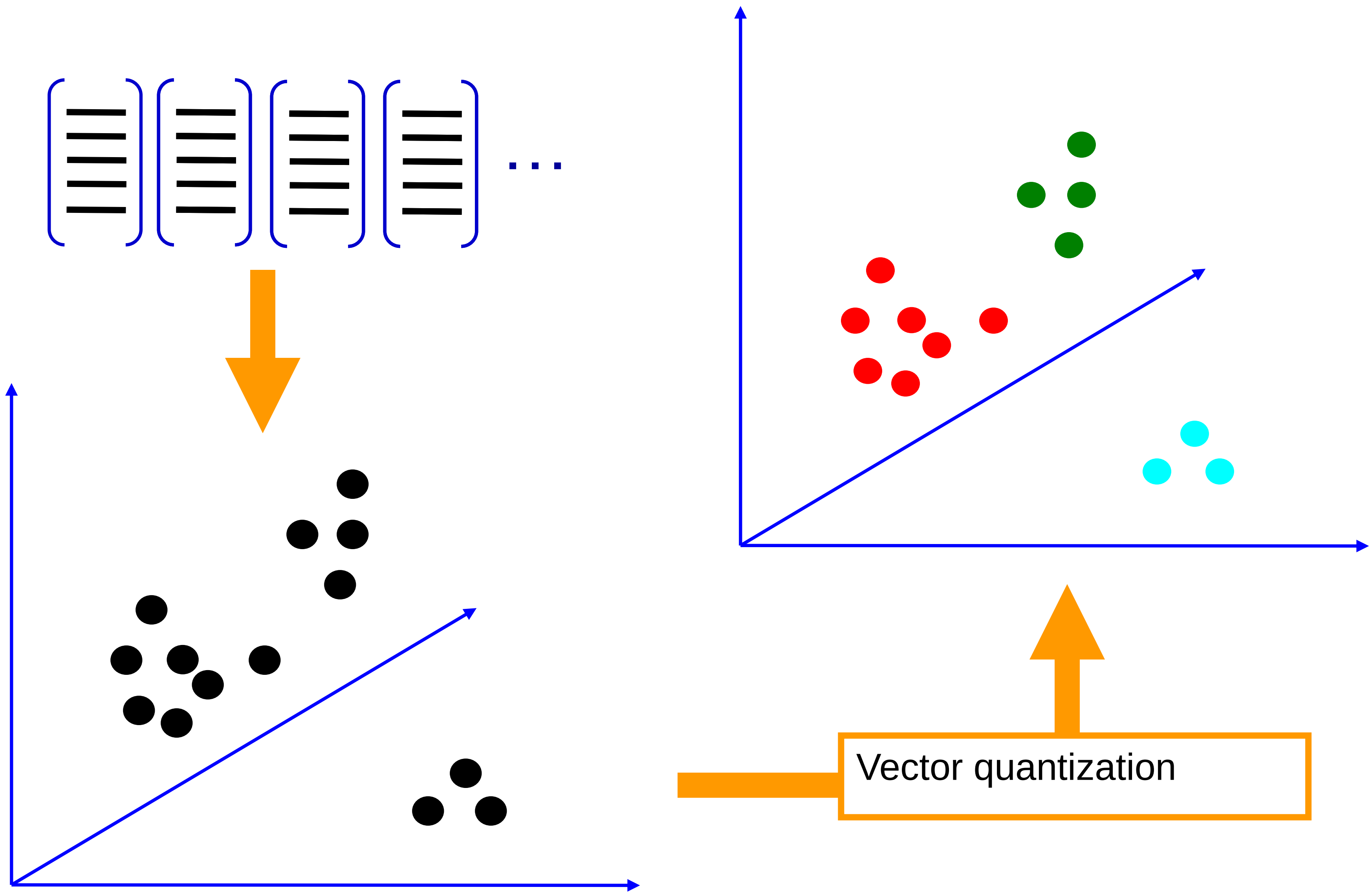
- “super-adaptive” binning
- Does not require discretization along any fixed axis

Patch Features

dictionary formation

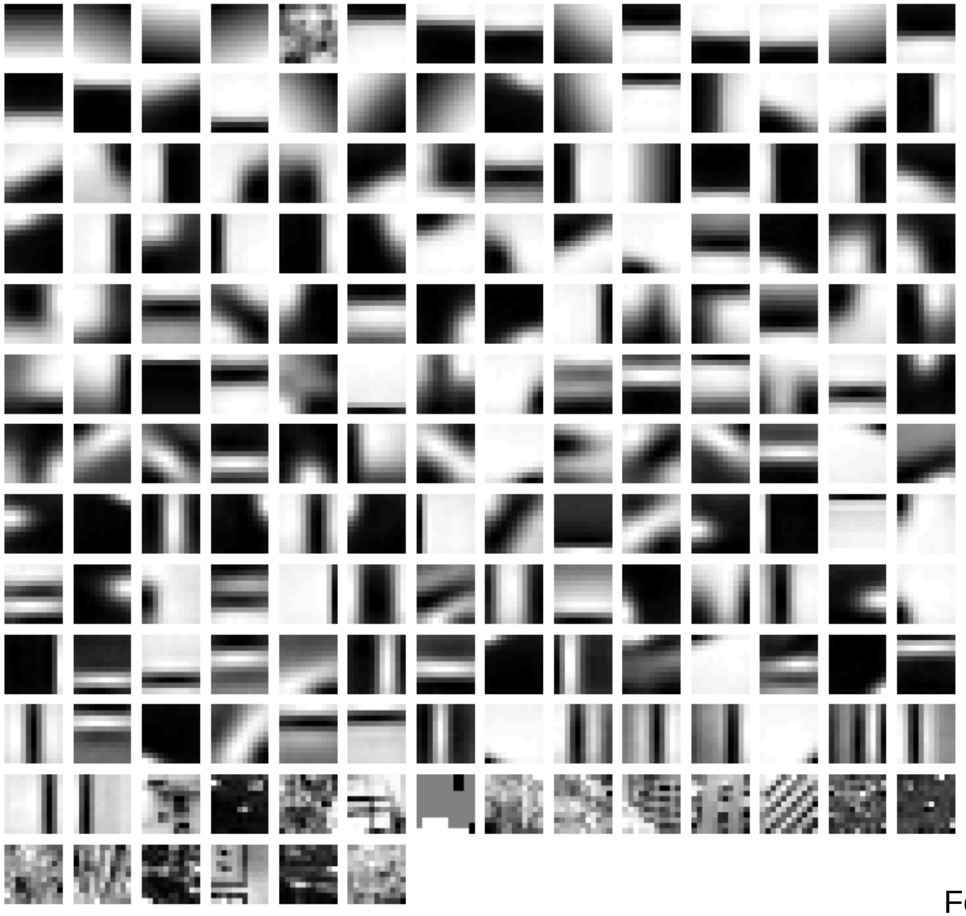


Clustering (usually k-means)



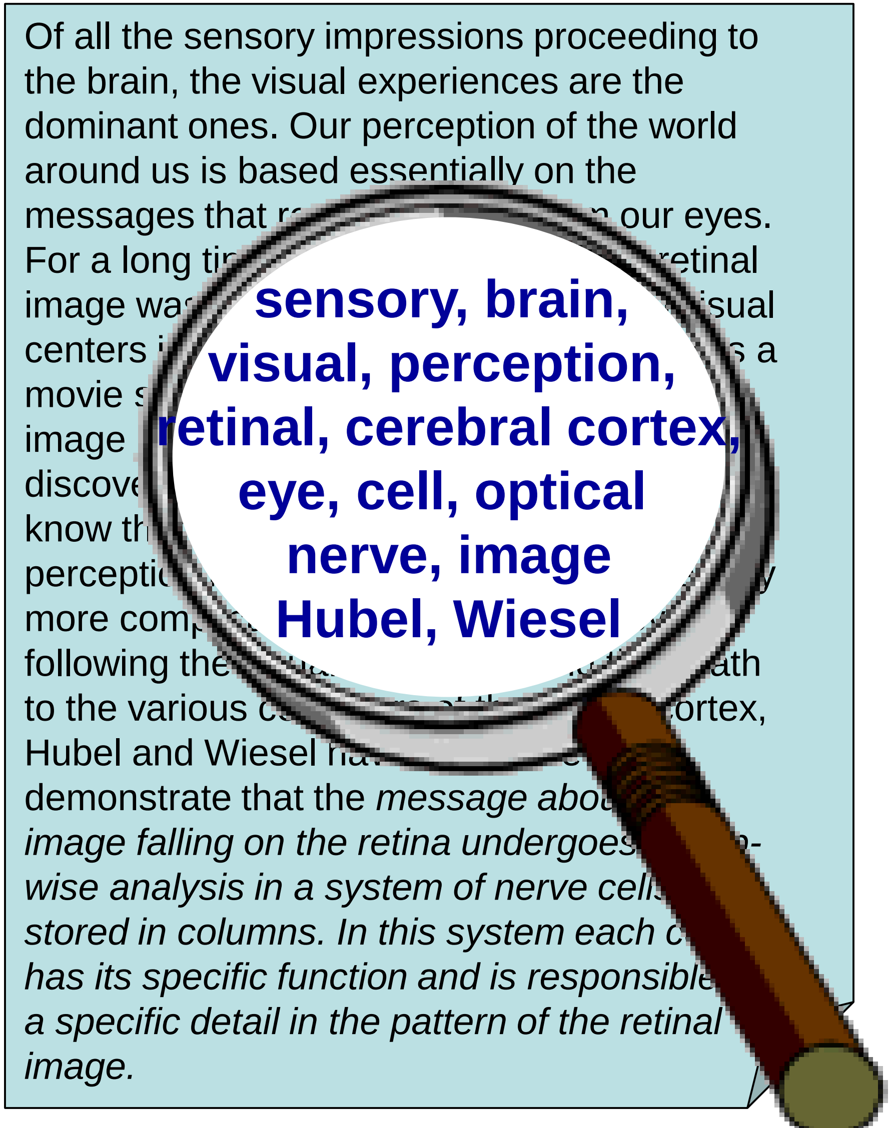
Slide credit: Josef Sivic

Clustered Image Patches ("Bag of Visual Words")



Fei-Fei et al. 2005

Analogy to documents



Object → Bag of 'words'

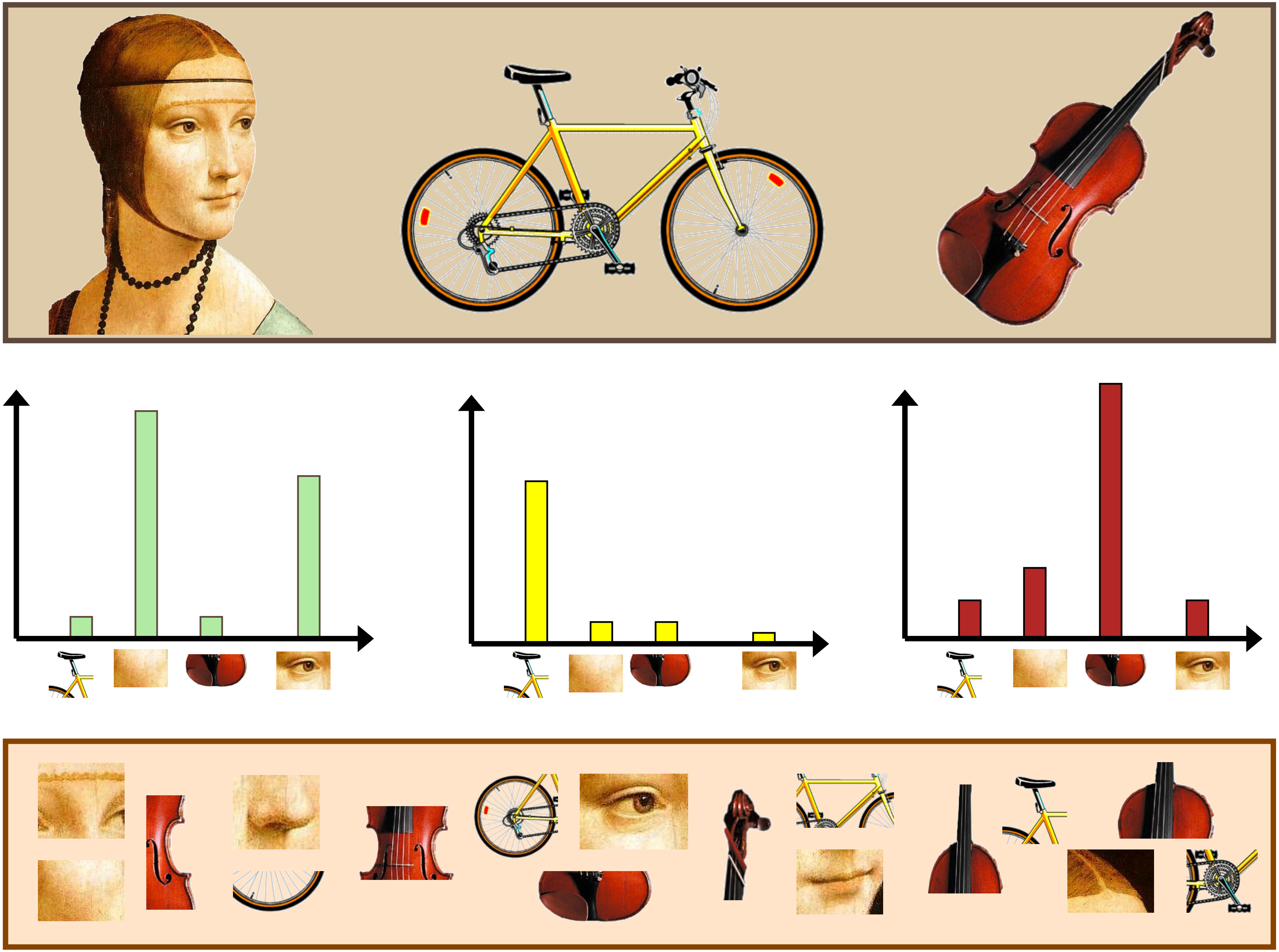
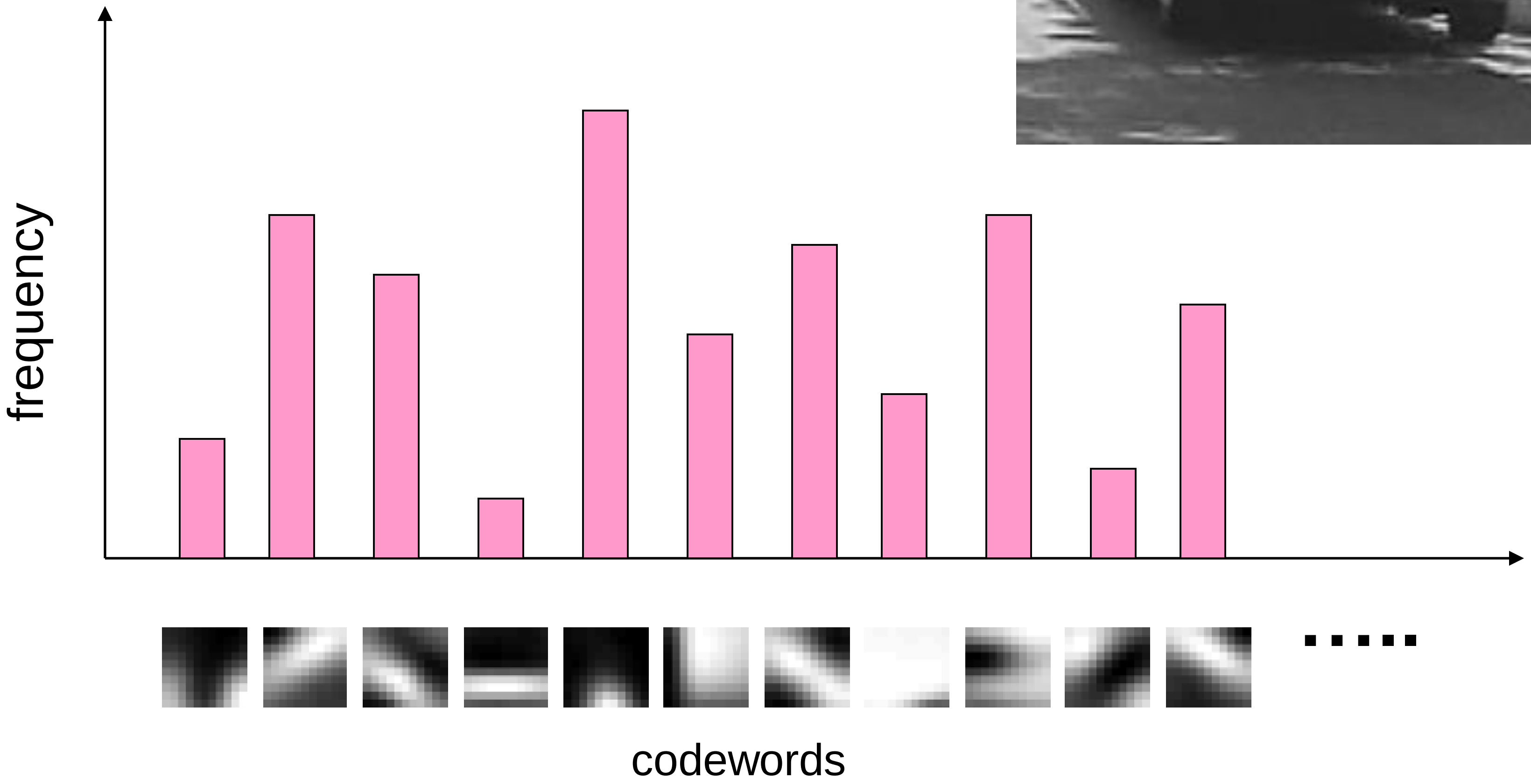


Image representation



Scene Classification (Renninger & Malik)

beach

mountain

forest

city

street

farm

kitchen

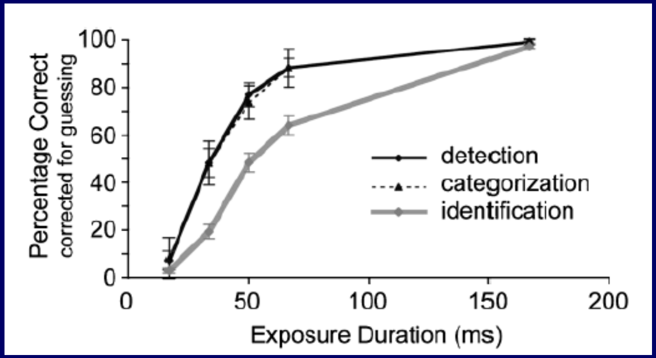
livingroom

bedroom

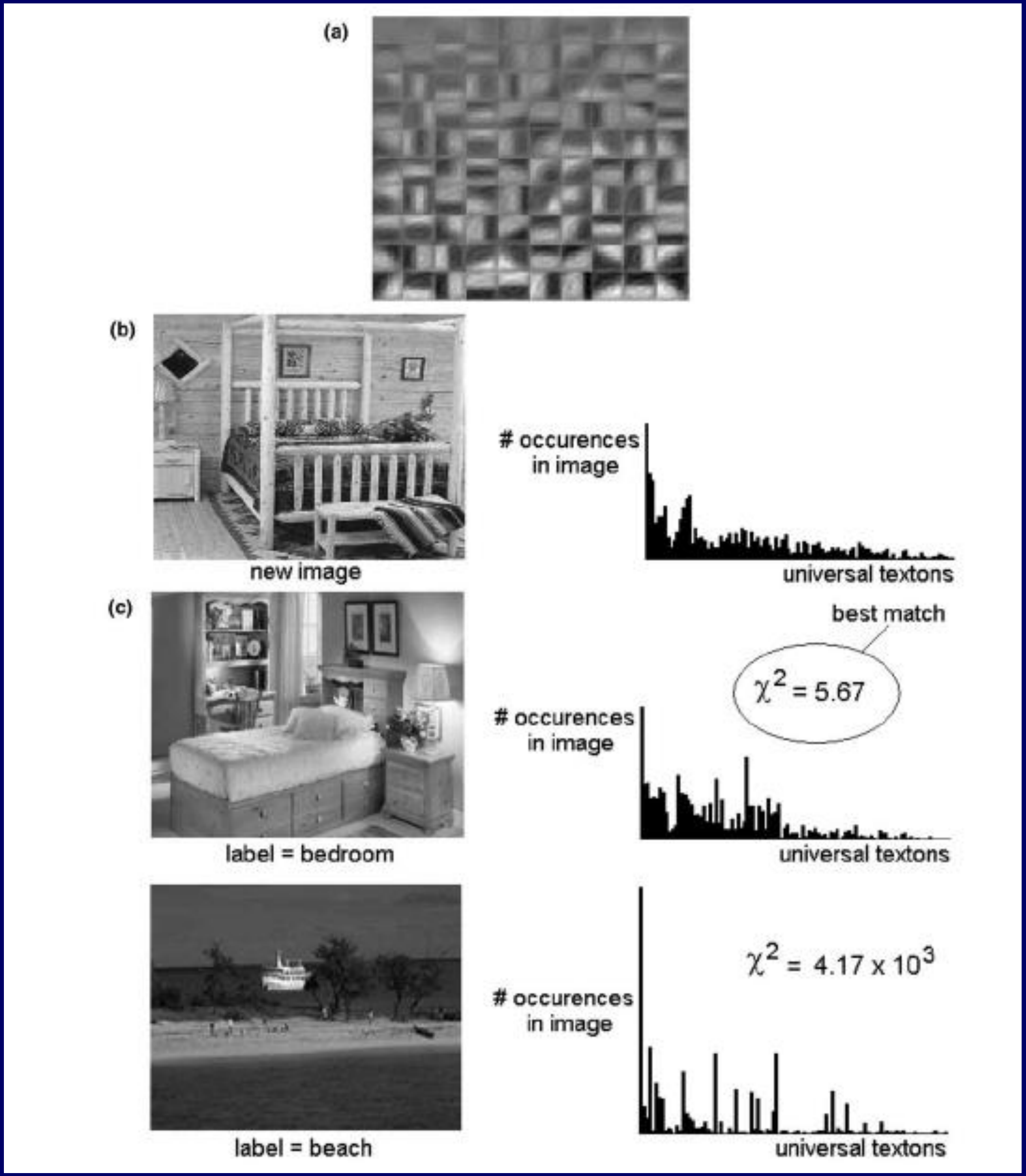
bathroom

Image classification can be pre-attentive!

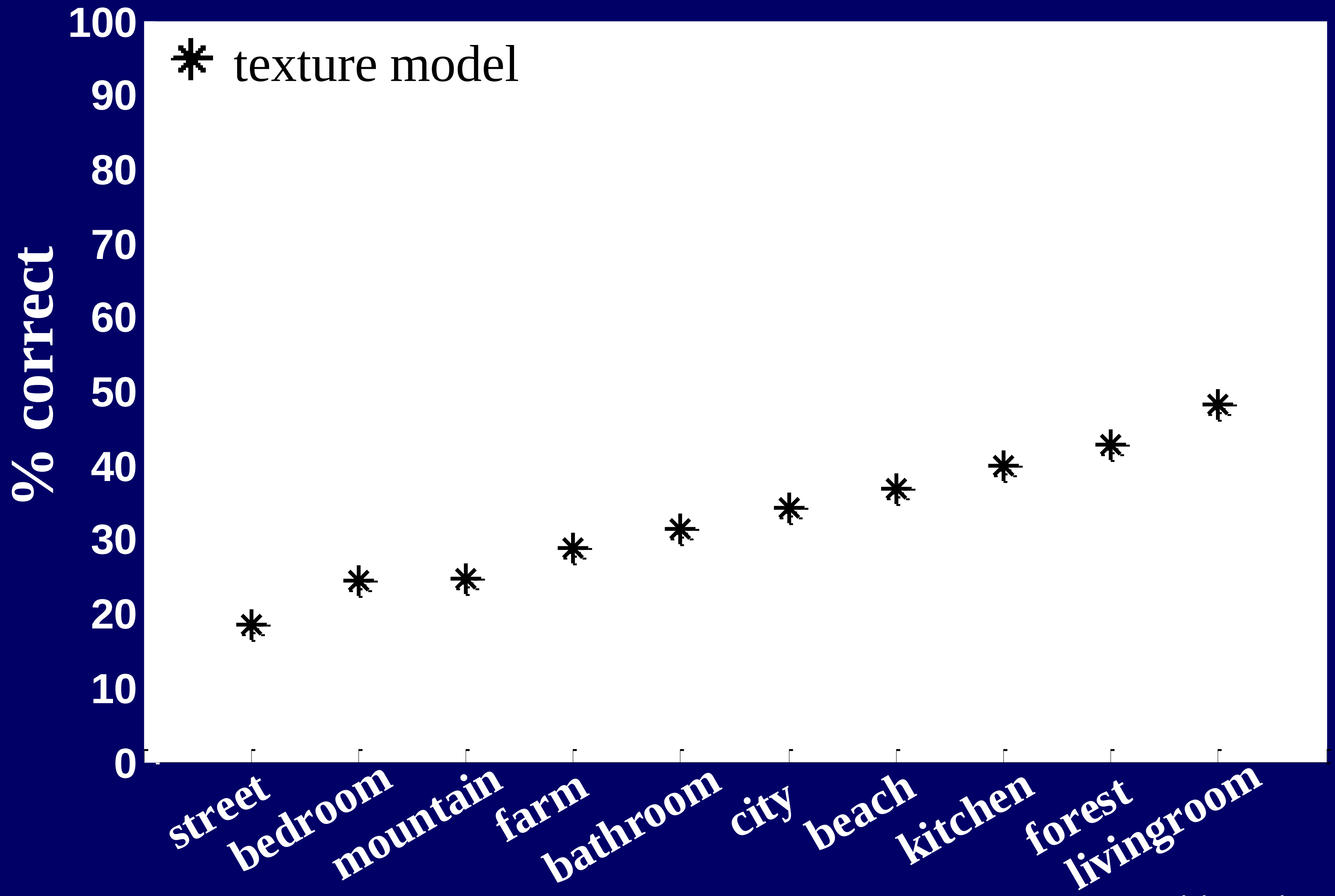
- On a task of judging animal vs no animal, humans can make mostly correct saccades in 150 ms (Kirchner & Thorpe, 2006)
  - Comparable to synaptic delay in the retina, LGN, V1, V2, V4, IT pathway.
  - Doesn't rule out feed back but shows feed forward only is very powerful
- Detection and categorization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



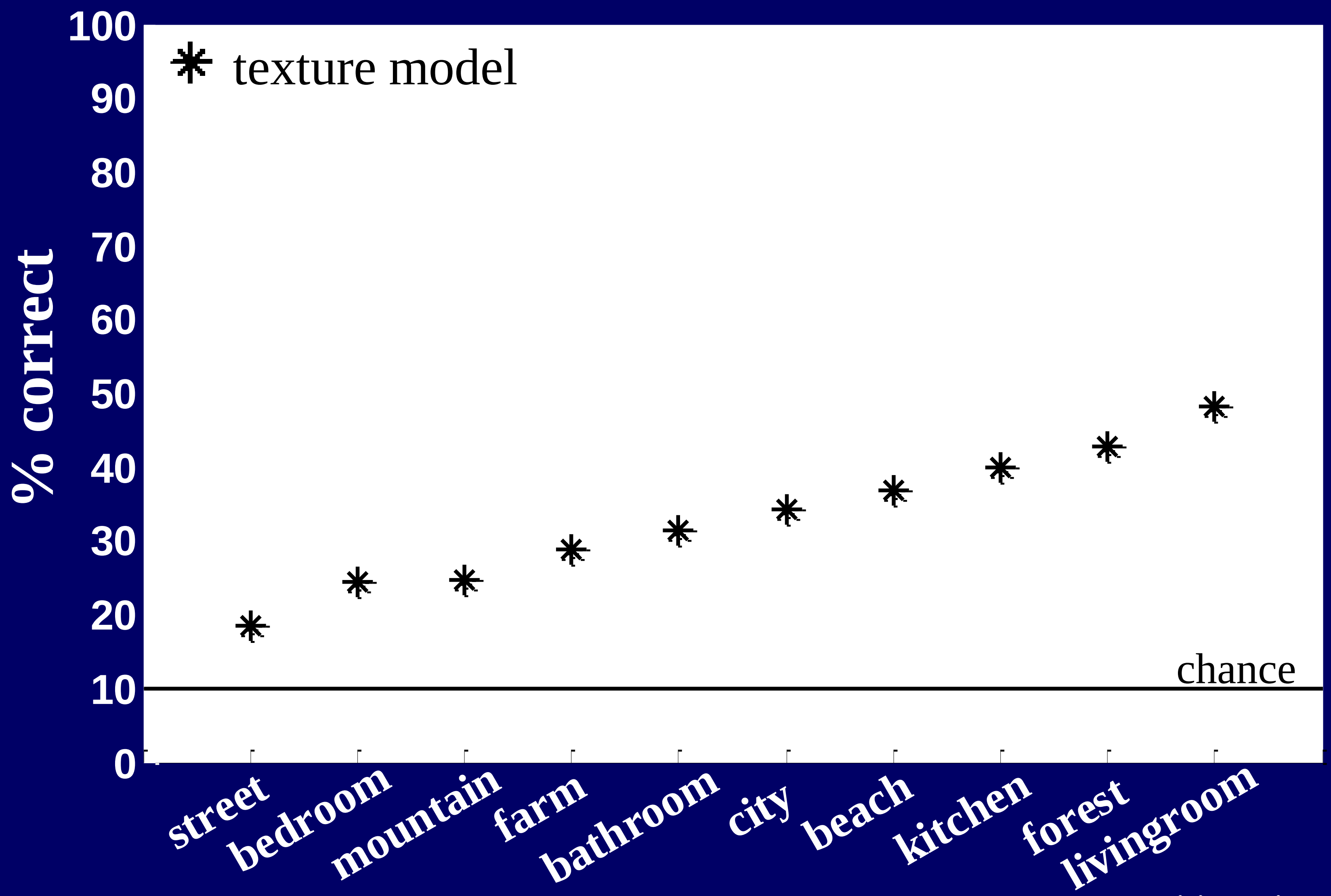
Texton Histogram Matching



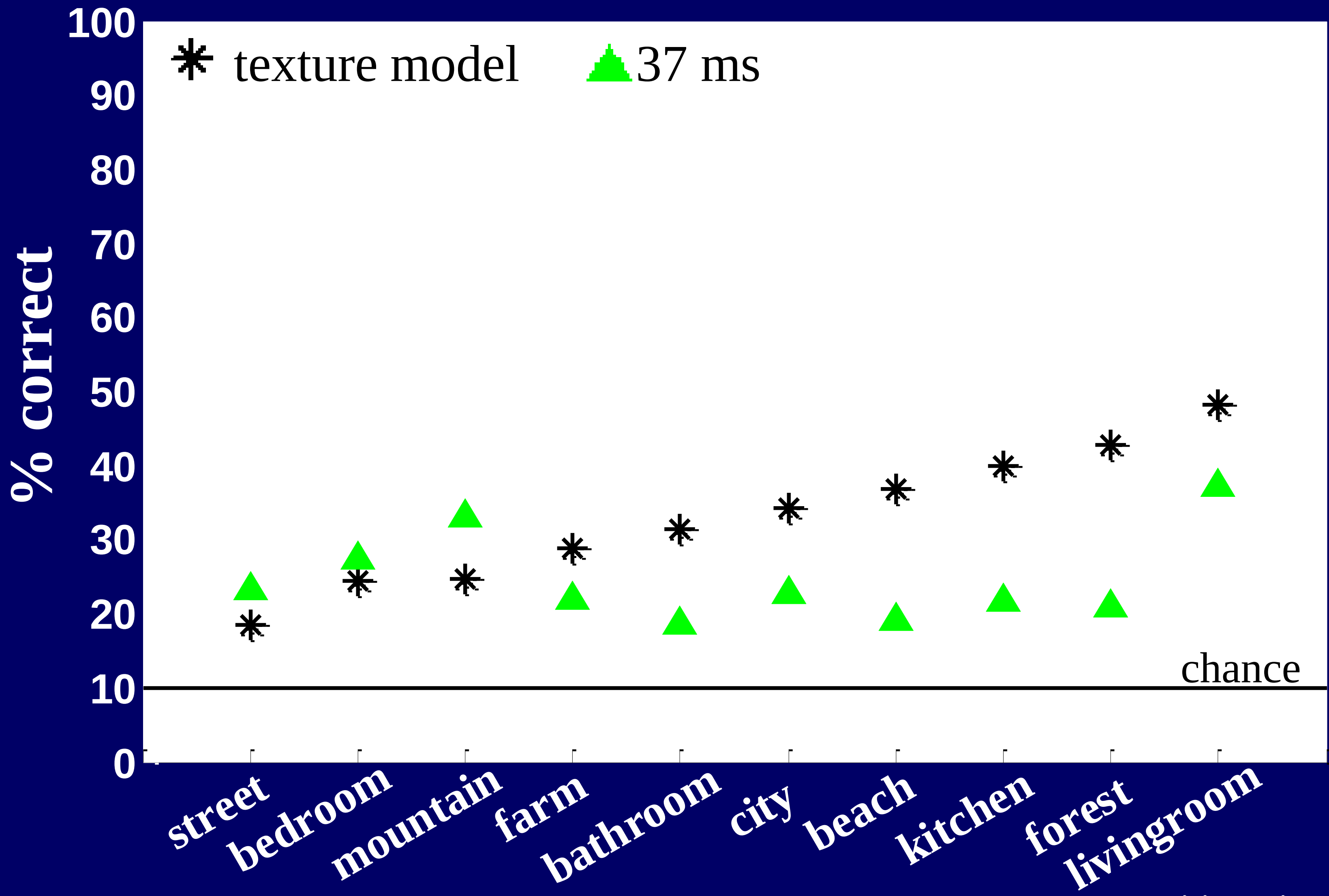
Discrimination of Basic Categories



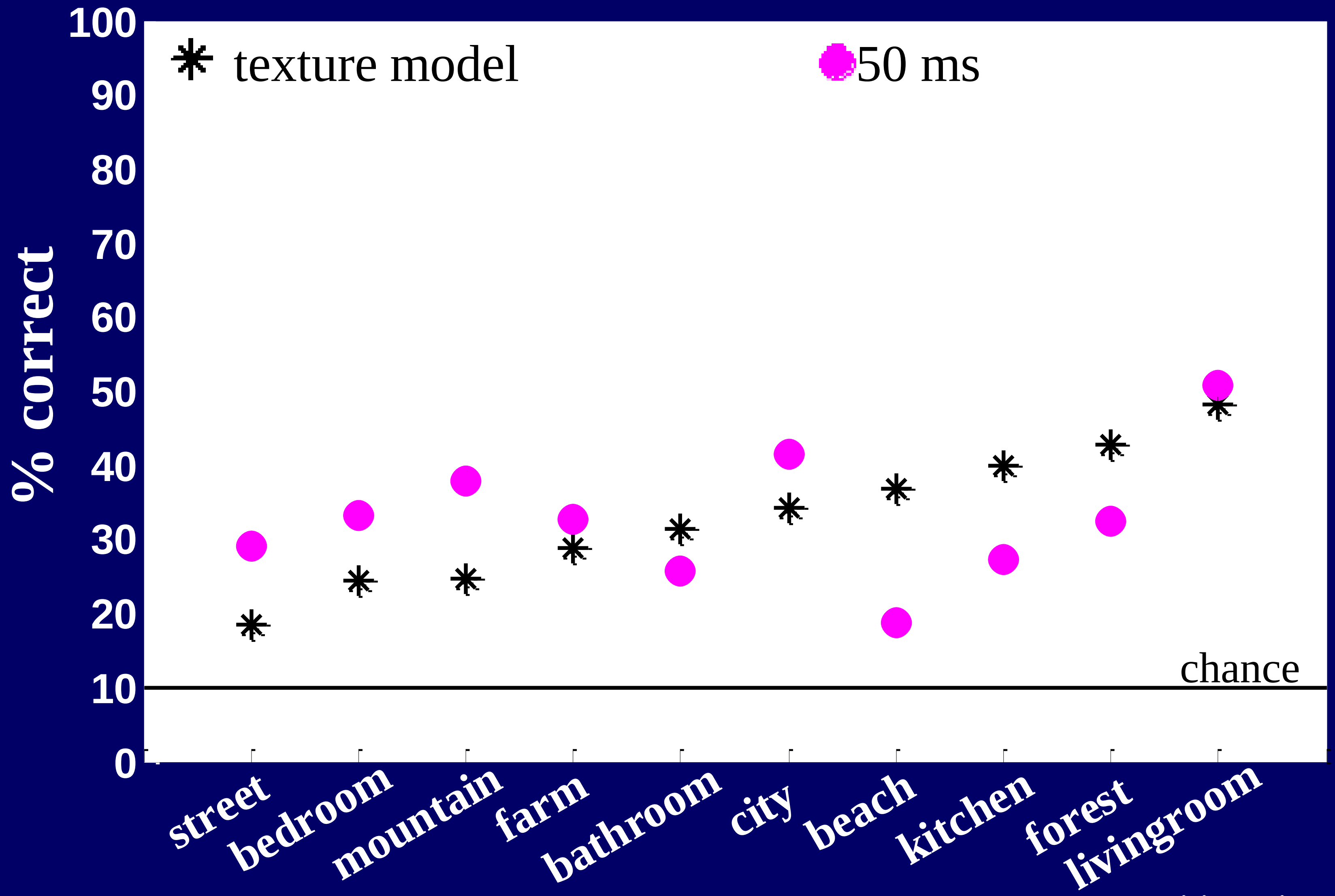
Discrimination of Basic Categories



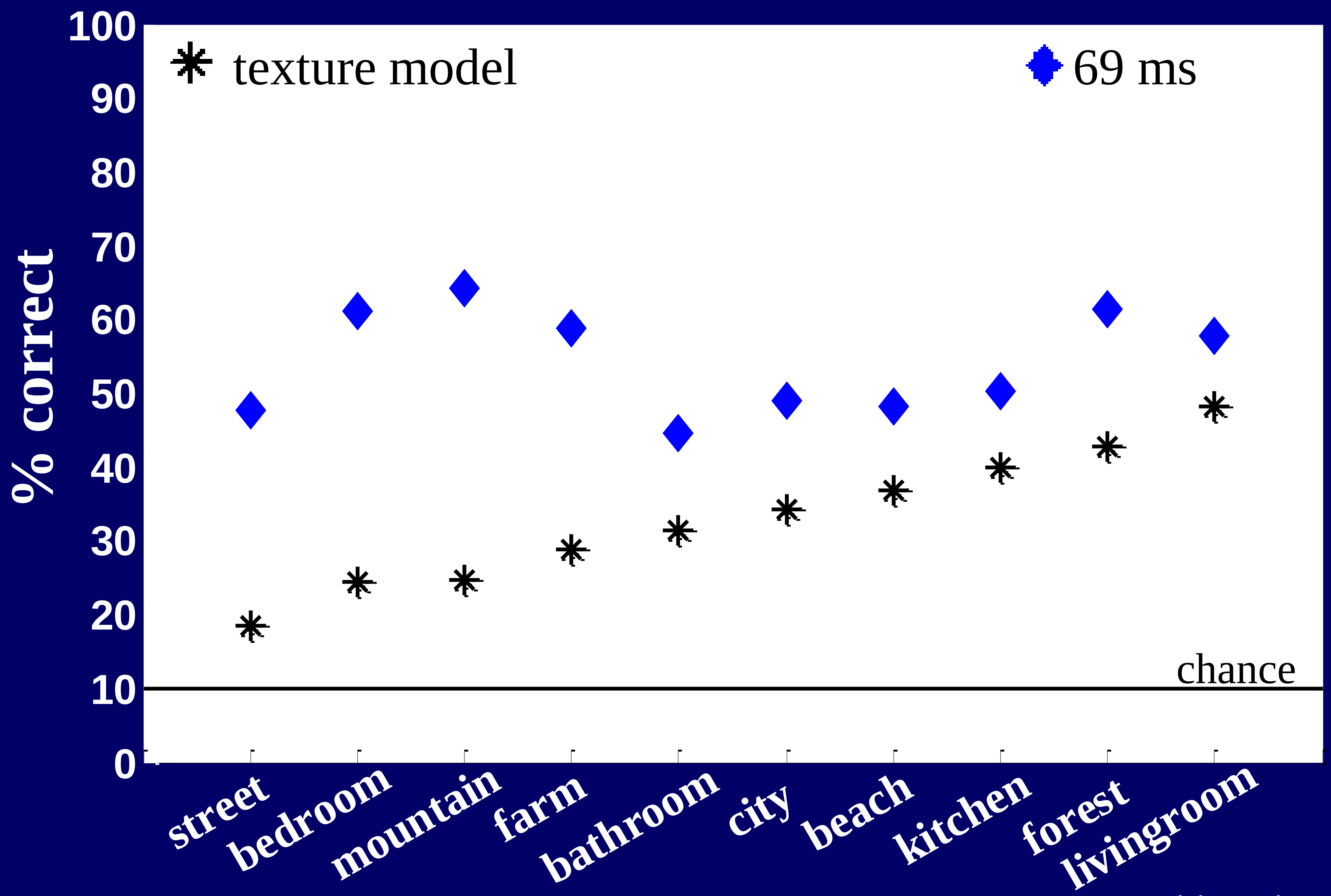
Discrimination of Basic Categories



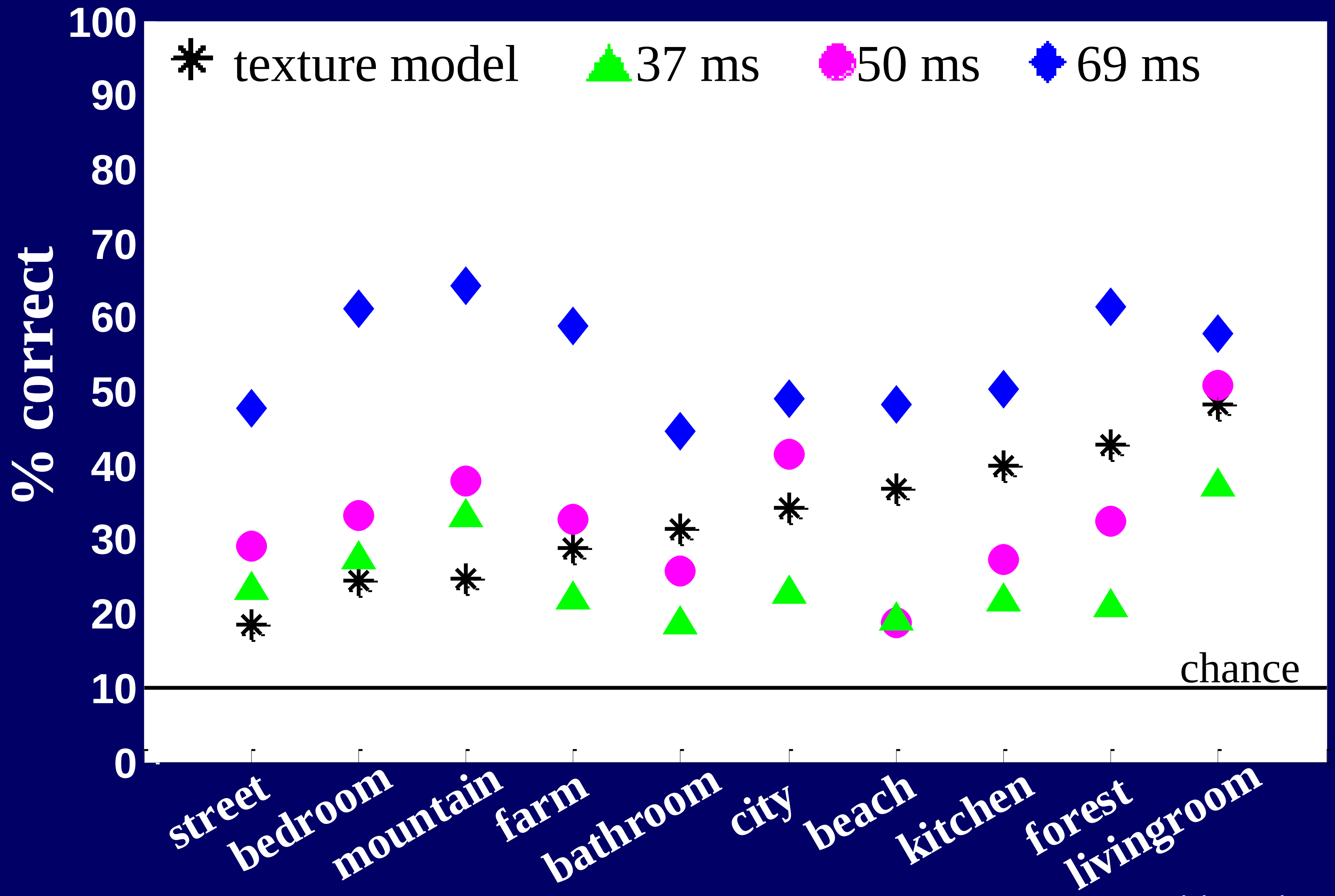
Discrimination of Basic Categories



Discrimination of Basic Categories



Discrimination of Basic Categories



Scene Recognition using Texture

Convolutional Neural Networks