# VisuaTexture (in human and machine)

Somewhere in Cinque Terre, May 2005

CS194: Image Manipulation, Comp. Vision, and Comp. Photo

Alexei Efros, UC Berkeley, Spring 2020

# What is Texture?

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

radishes

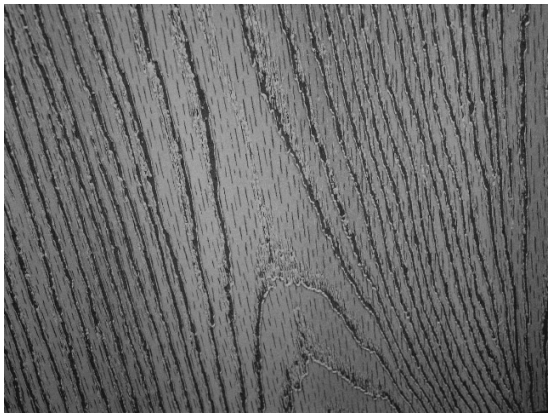
rocks

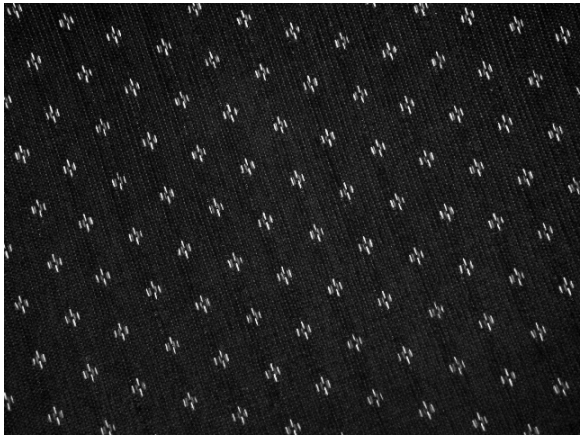
yogurt

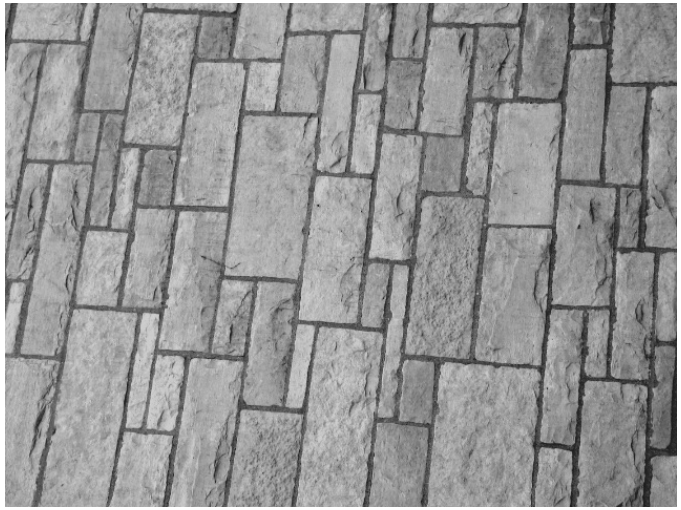
# Texture as “stuff”

Source: Forsyth

# Texture and Material







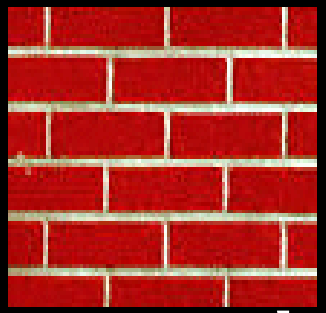
http://www-cvr.ai.uiuc.edu/ponce_grp/data/texture_database/samples/

# Texture Synthesis

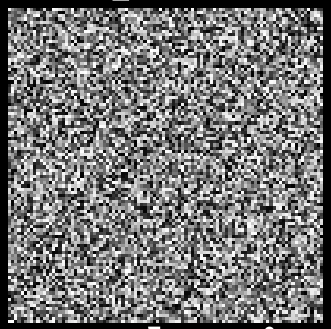
- Goal of Texture Synthesis: create new samples of a given texture
- Many applications: virtual environments, hole-filling, texturing surfaces

# The Challenge

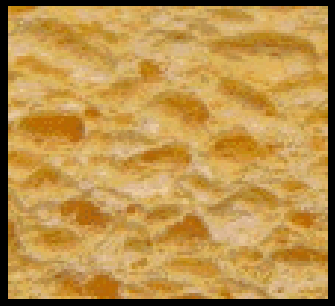
- Need to model the whole spectrum: from repeated to stochastic texture



**repeated**

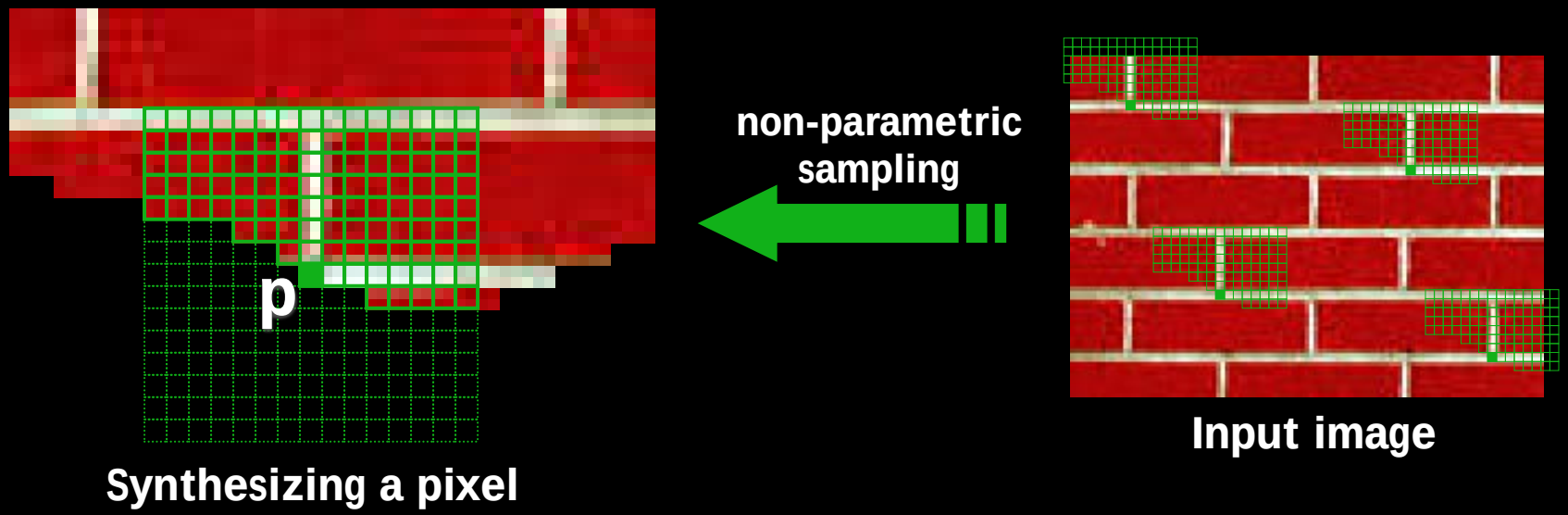


**stochastic**



**Both?**

# Efros & Leung Algorithm

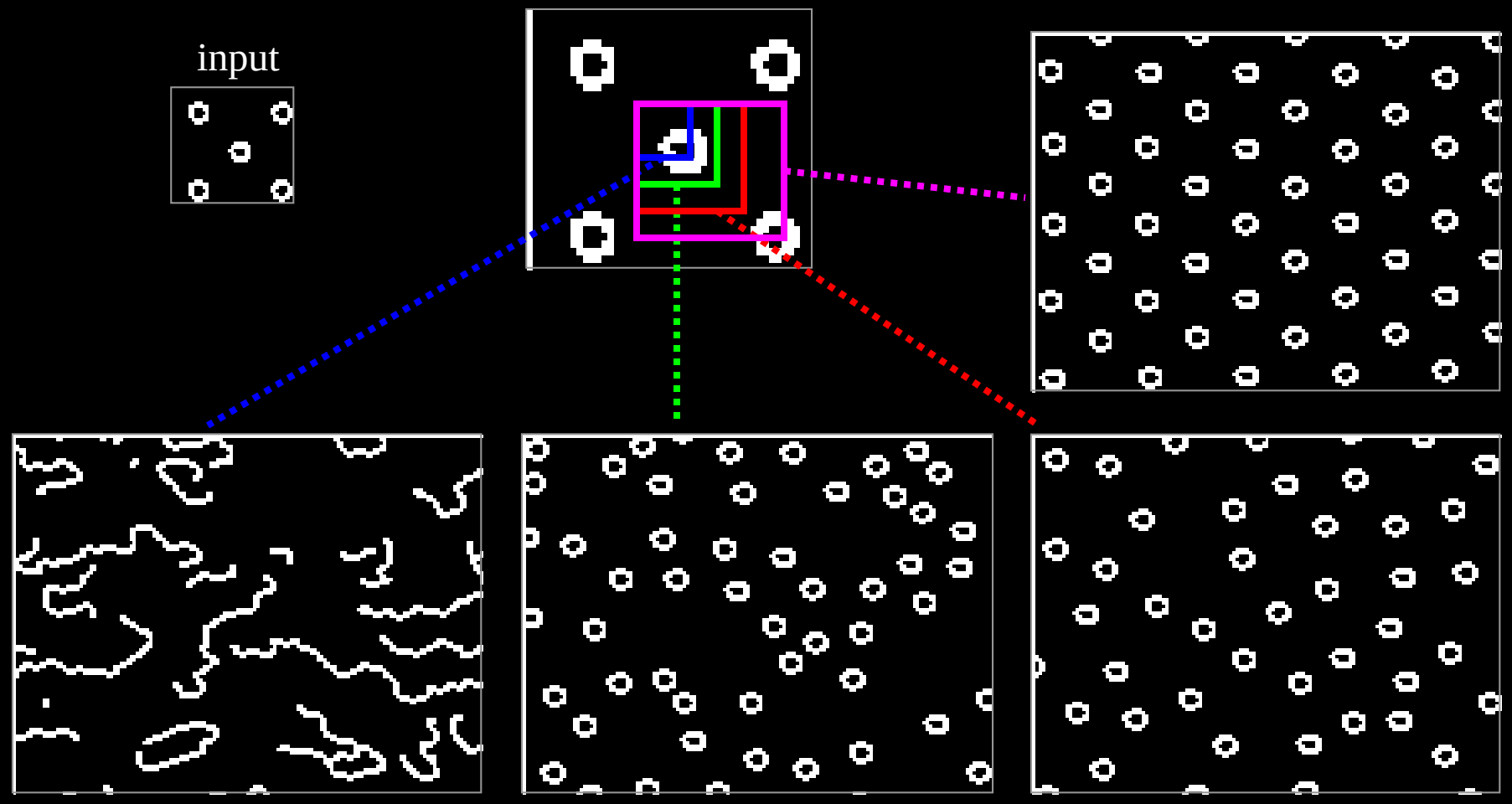



- Assuming Markov property, compute P(**p**|N(**p**))
  - Building explicit probability tables infeasible
  - Instead, we *search the input image* for all similar neighborhoods — that's our pdf for **p**
  - To sample from this pdf, just pick one match at random

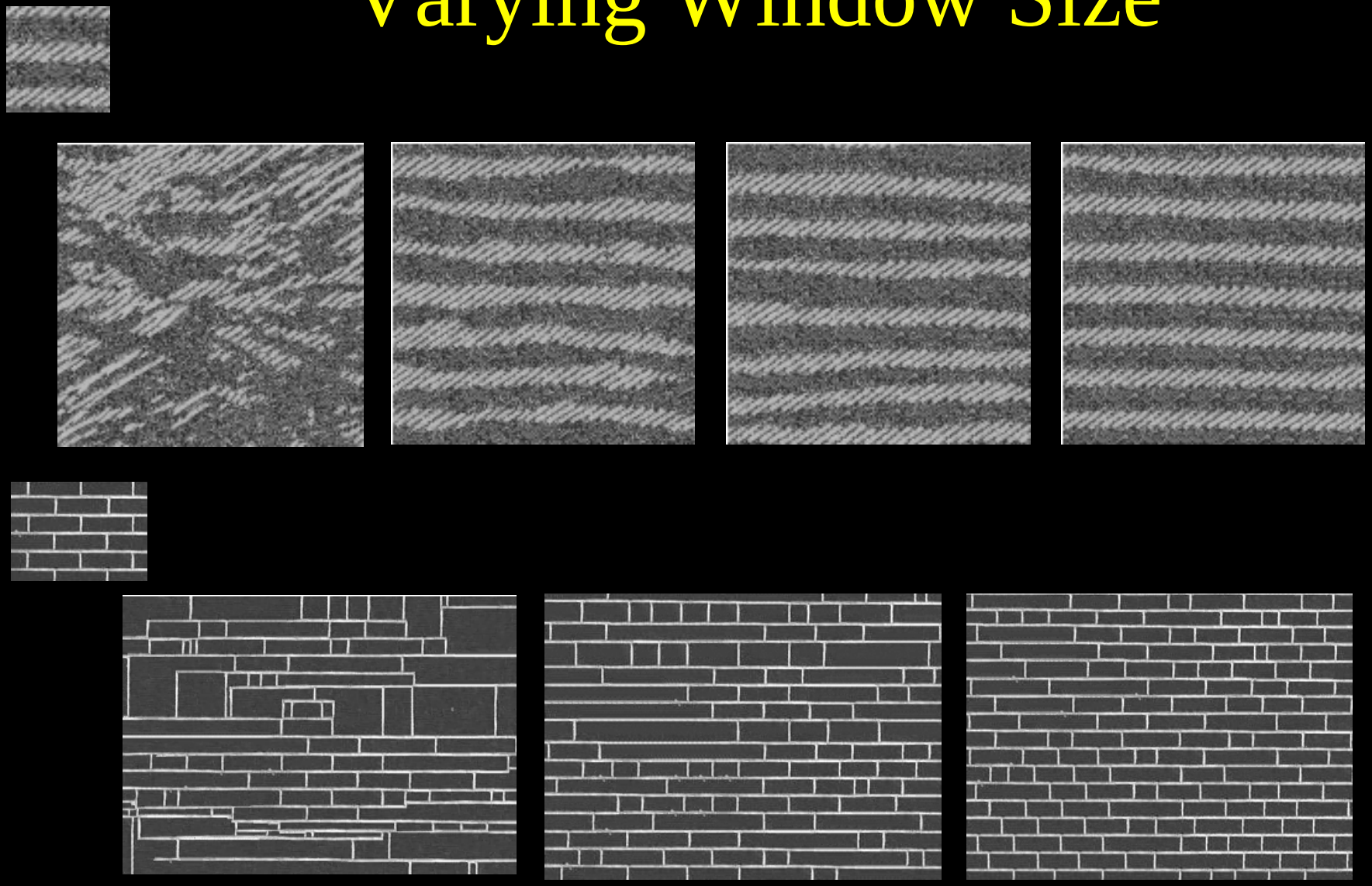
# Some Details

- Growing is in “onion skin” order
  - Within each “layer”, pixels with most neighbors are synthesized first
  - If no close match can be found, the pixel is not synthesized until the end
- Using *Gaussian-weighted* SSD is very important
  - to make sure the new pixel agrees with its closest neighbors
  - Approximates reduction to a smaller neighborhood window if data is too sparse

# Neighborhood Window



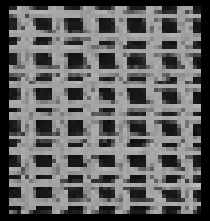
# Varying Window Size

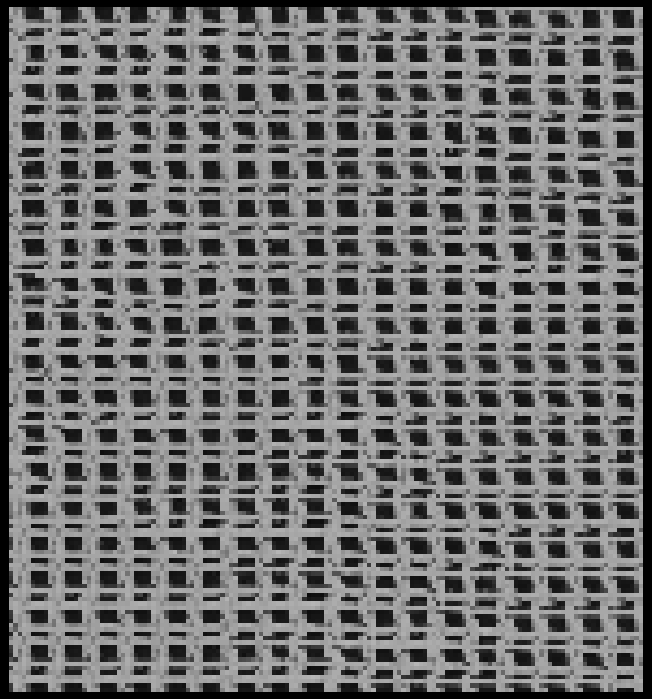


Increasing window size

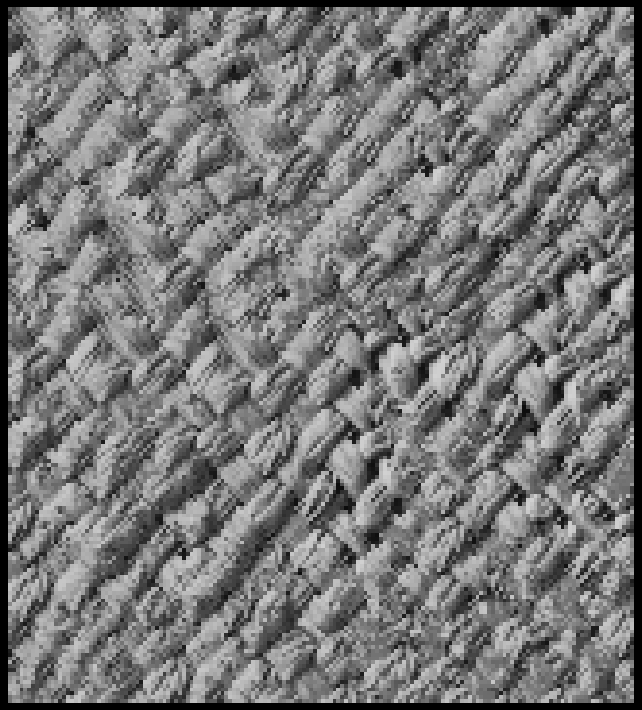
# Synthesis Results

french canvas



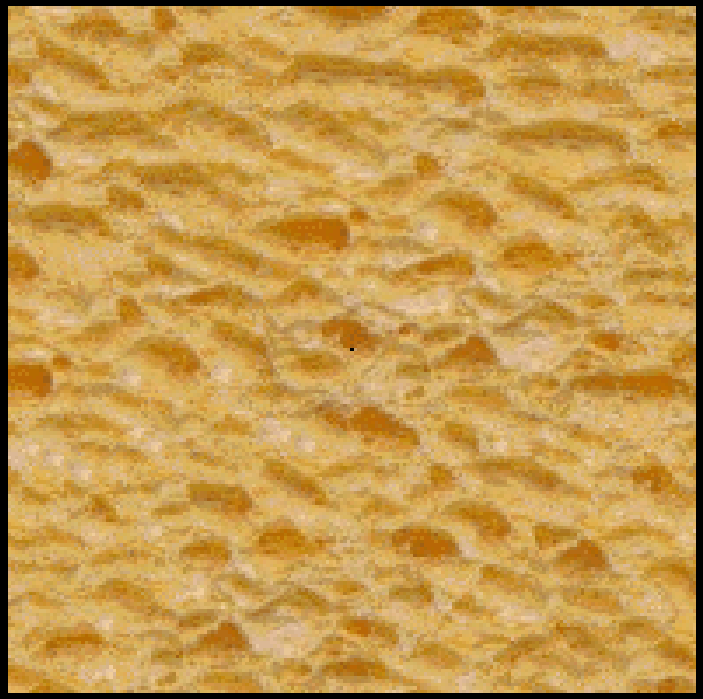


rafia weave

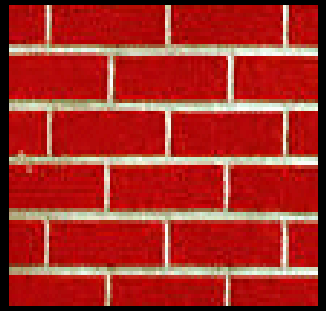


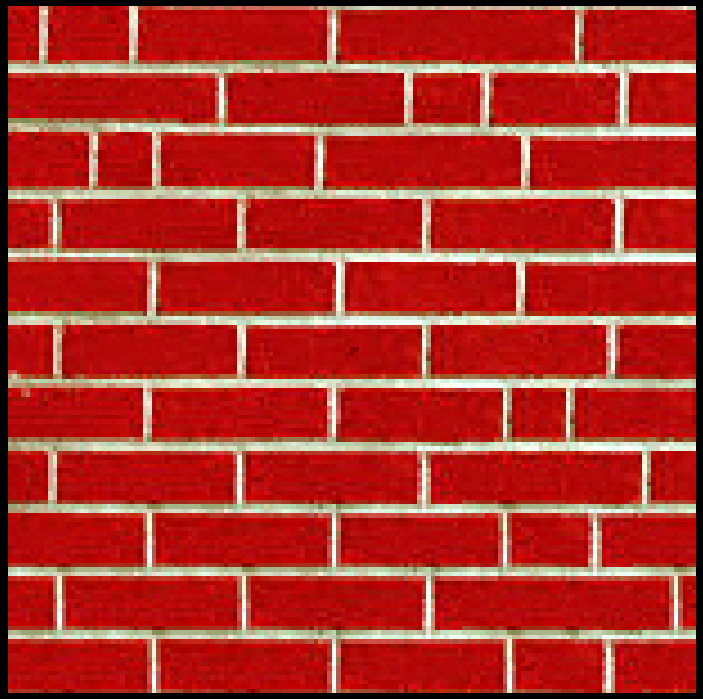
# More Results

white bread

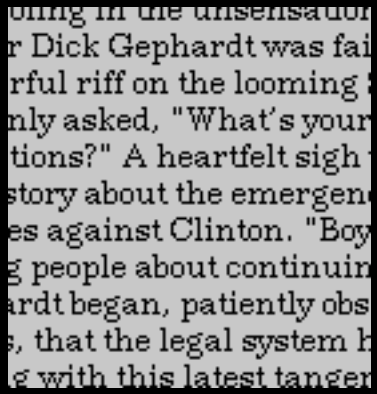


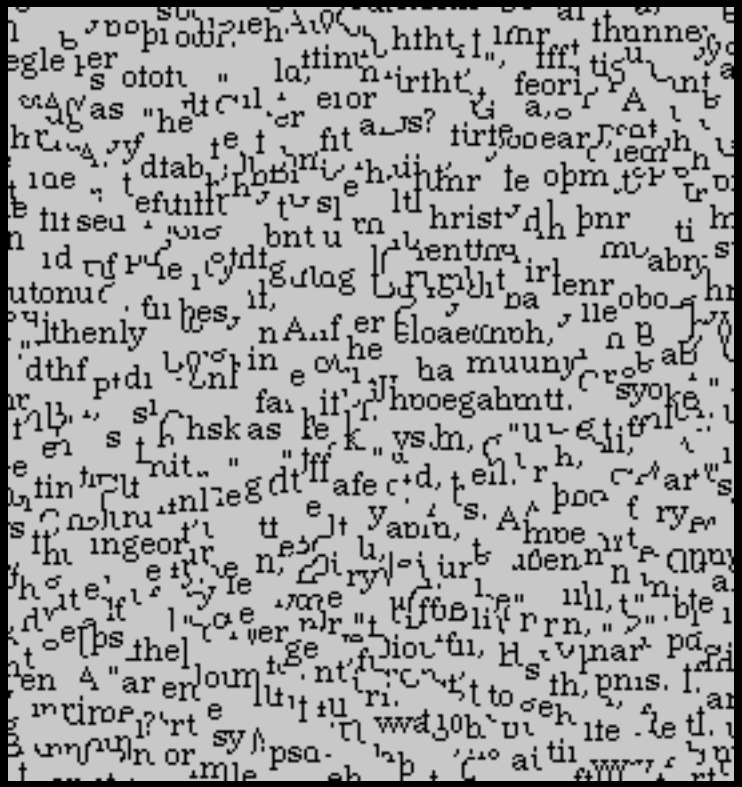
brick wall

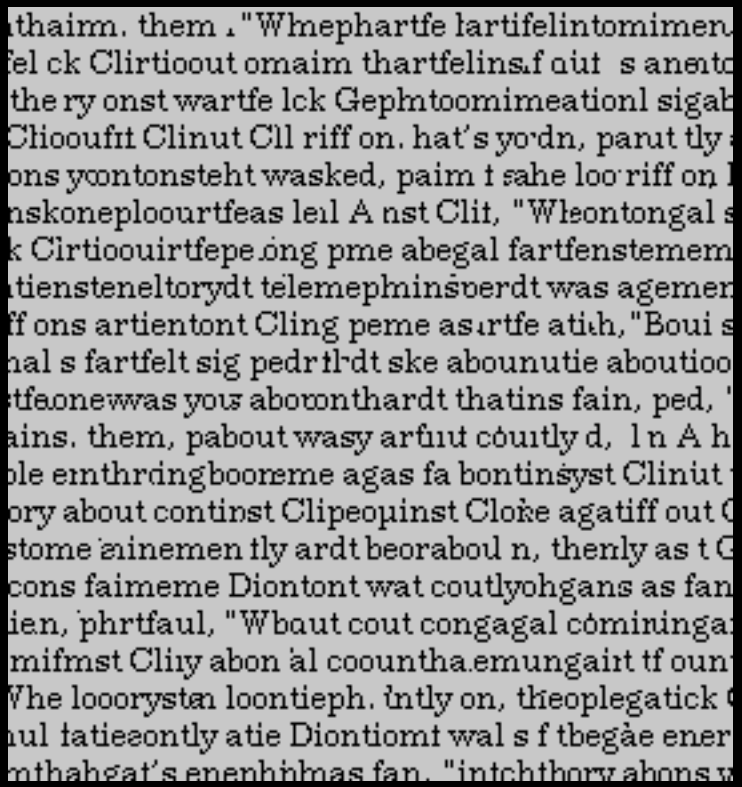




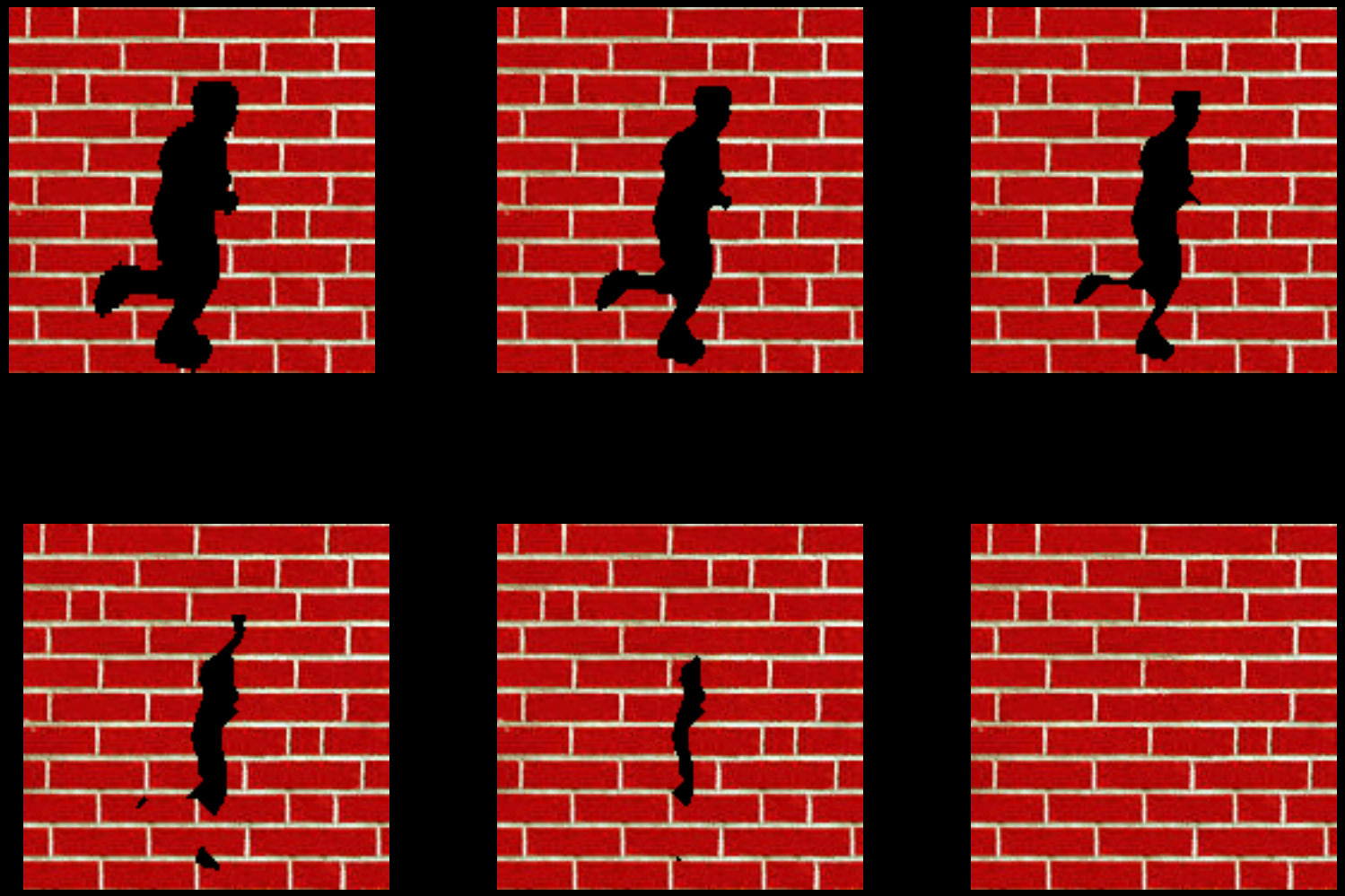
# Homage to Shannon



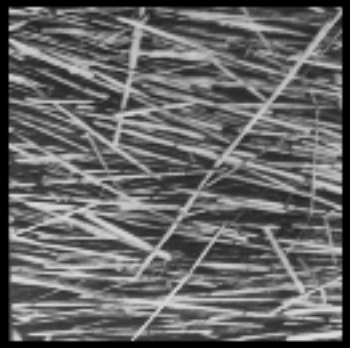



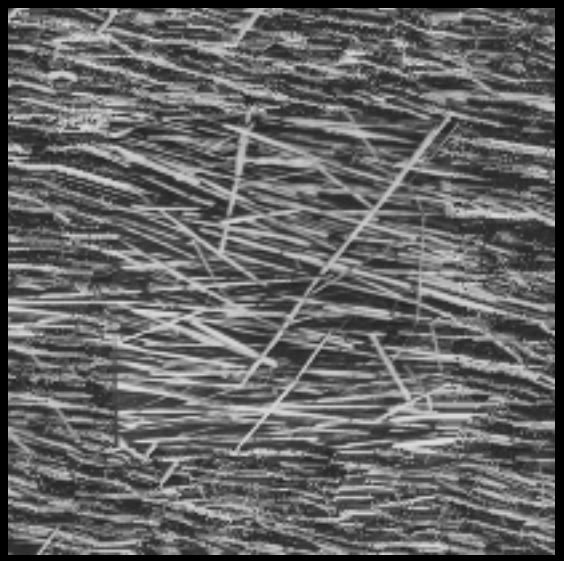


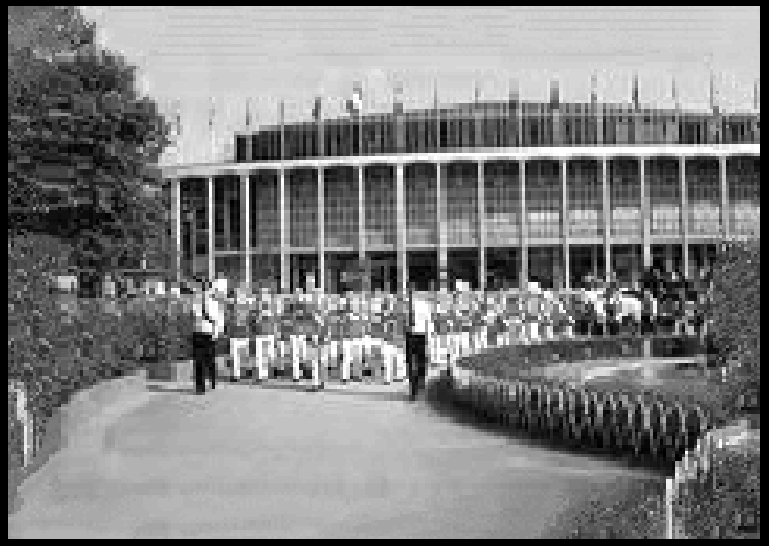
# Hole Filling



# Extrapolation



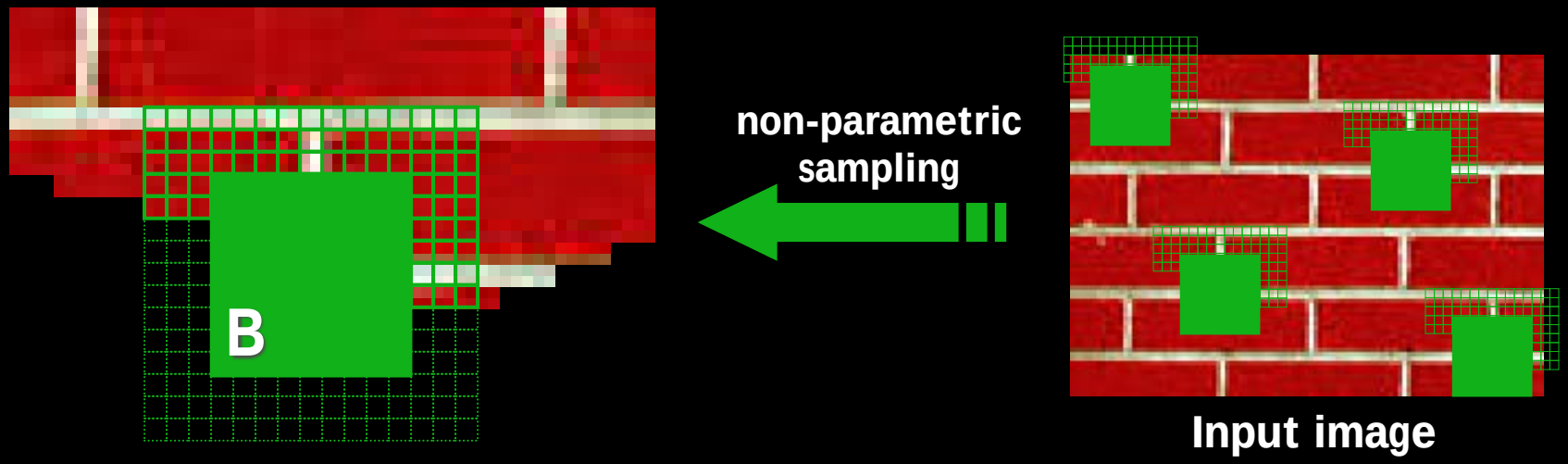




# Summary

- The Efros & Leung algorithm
  - Very simple
  - Surprisingly good results
  - Synthesis is easier than analysis!
  - …but very slow

# Image Quilting [Efros & Freeman]

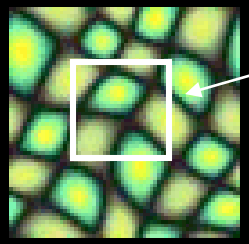



Synthesizing a block

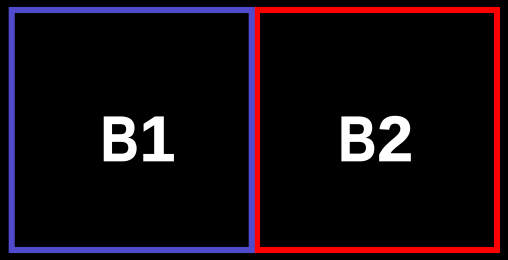
- Observation: neighbor pixels are highly correlated

**Idea: unit of synthesis = block**

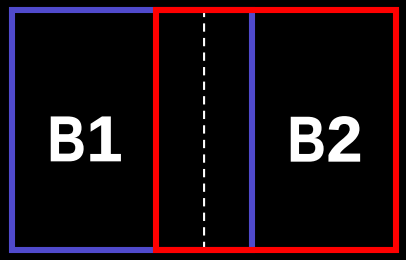
- Exactly the same but now we want P(**B**|N(**B**))
- Much faster: synthesize all pixels in a block at once
- Not the same as multi-scale!



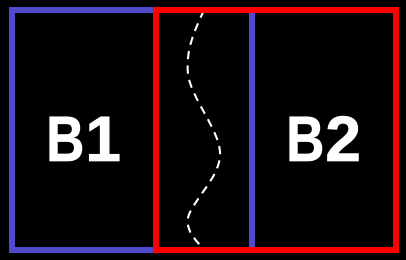

**Input texture**



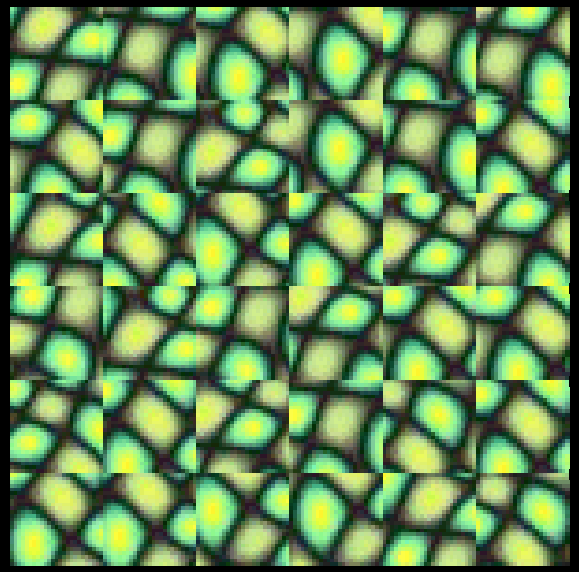

**Random placement of blocks**

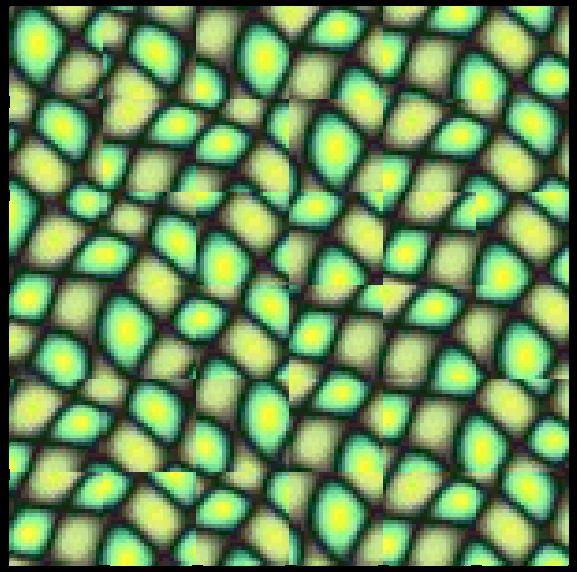


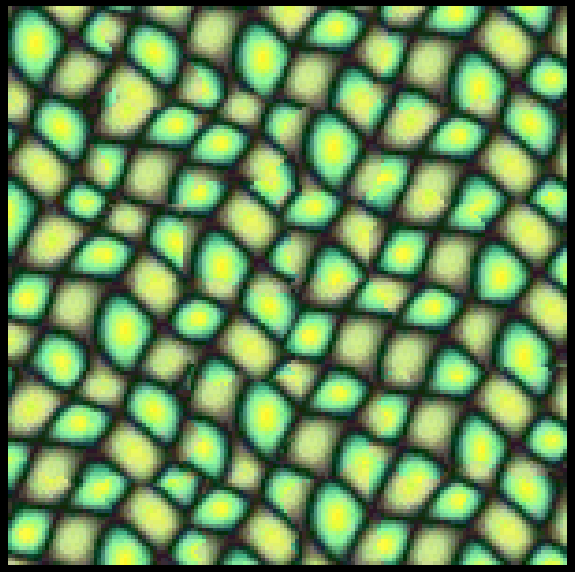

**Neighboring blocks constrained by overlap**



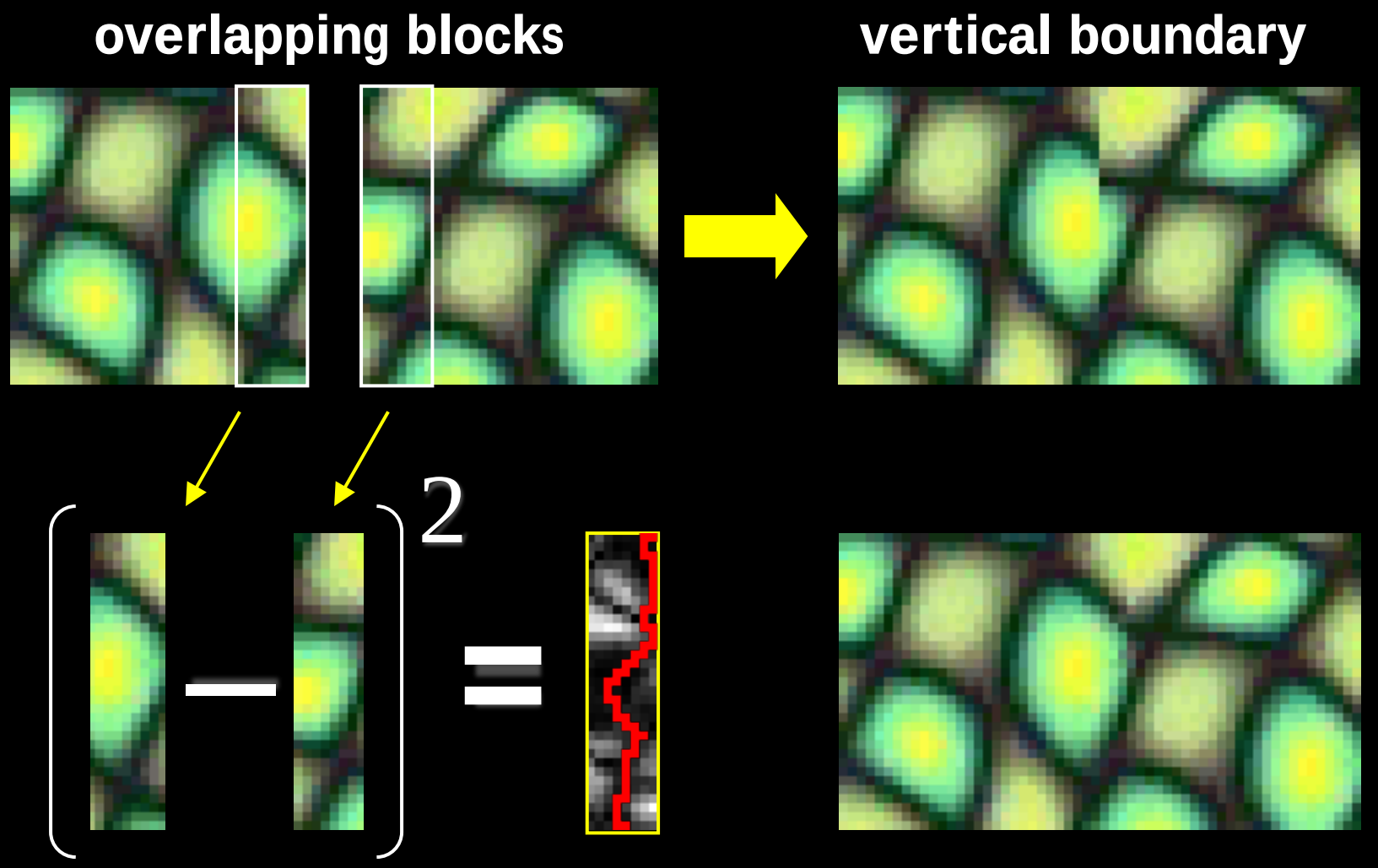

**Minimal error boundary cut**





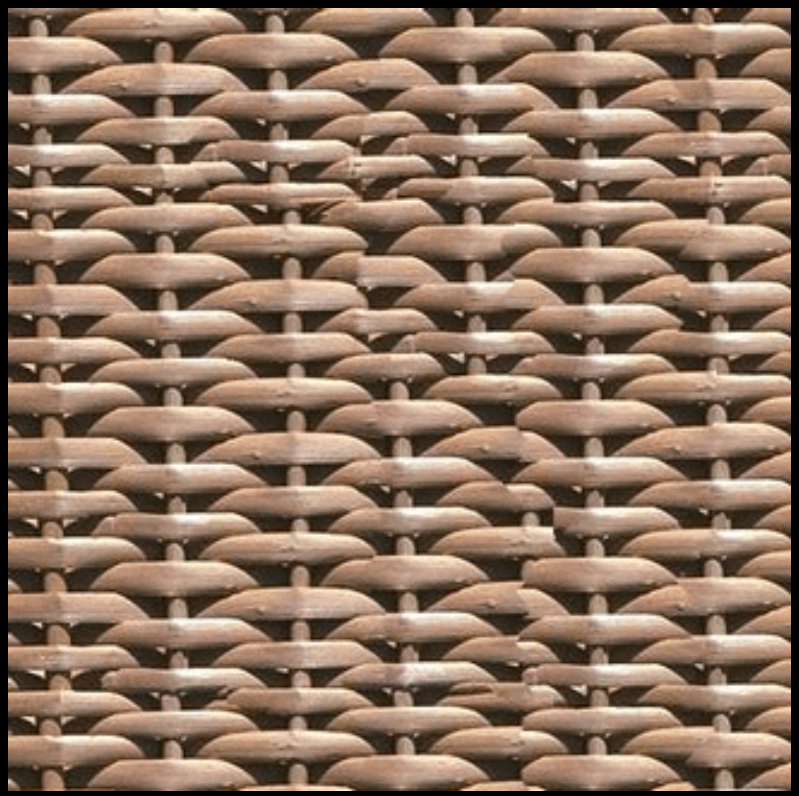


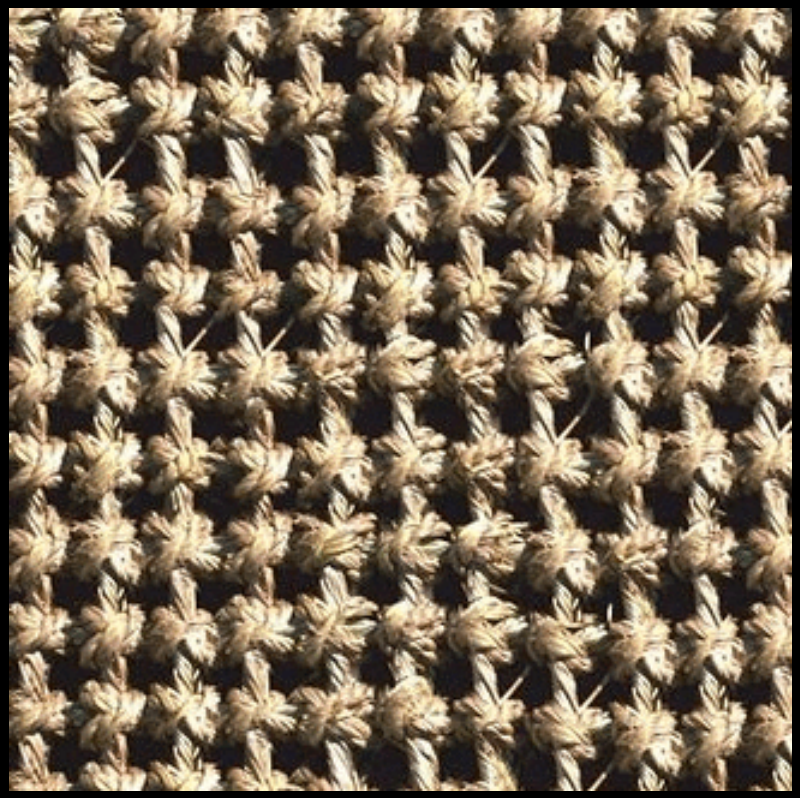
# Minimal error boundary



# Our Philosophy

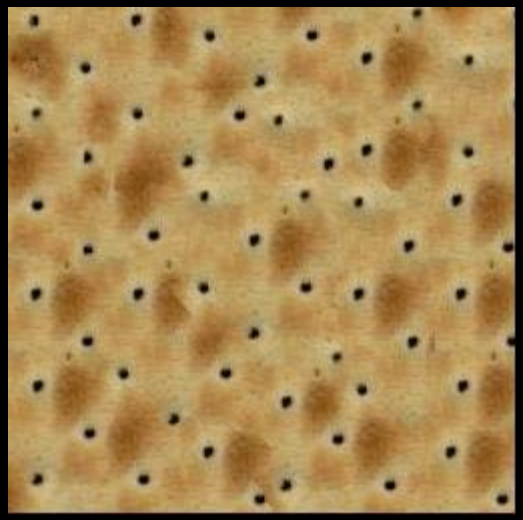
- The “Corrupt Professor’s Algorithm”:
  - Plagiarize as much of the source image as you can
  - Then try to cover up the evidence
- Rationale:
  - Texture blocks are by definition correct samples of texture so problem only connecting them together





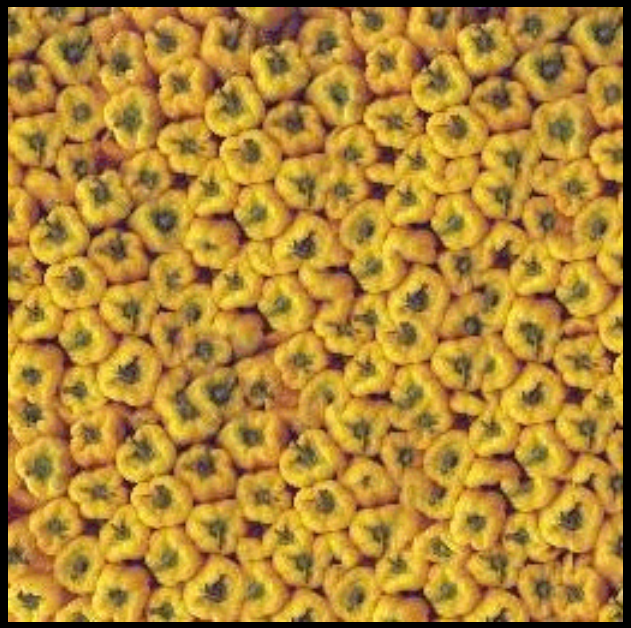

College Inn
Chicken Broth


College Inn
Chicken Broth

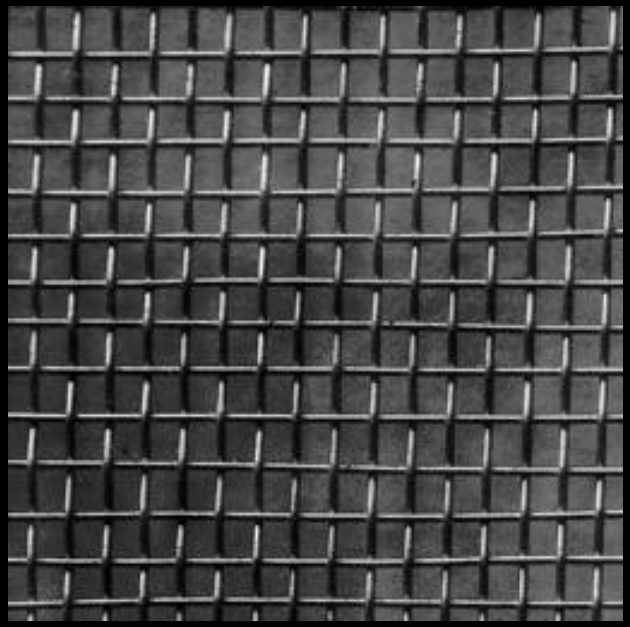





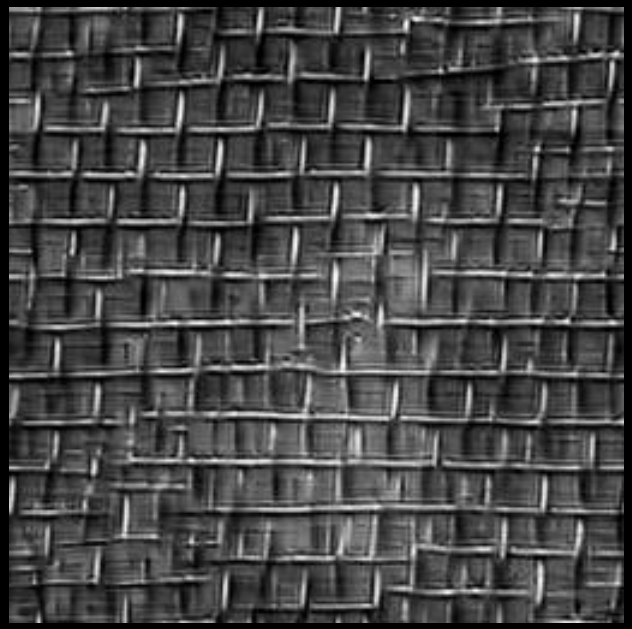




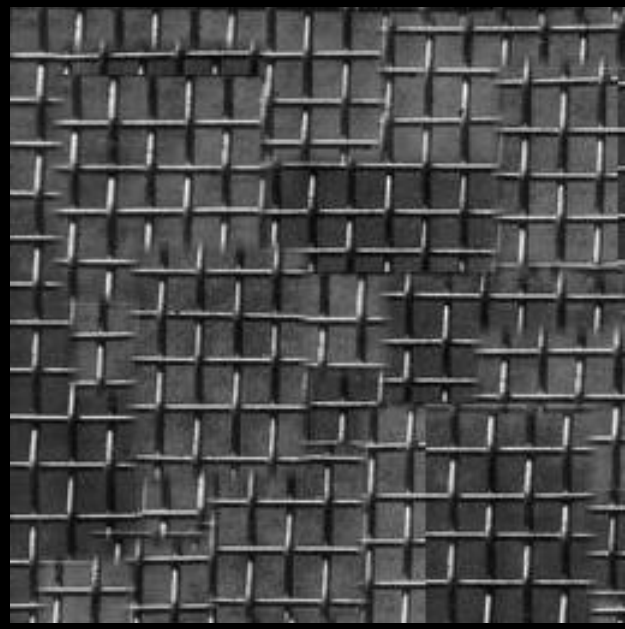
# Failures
## (Chernobyl Harvest)



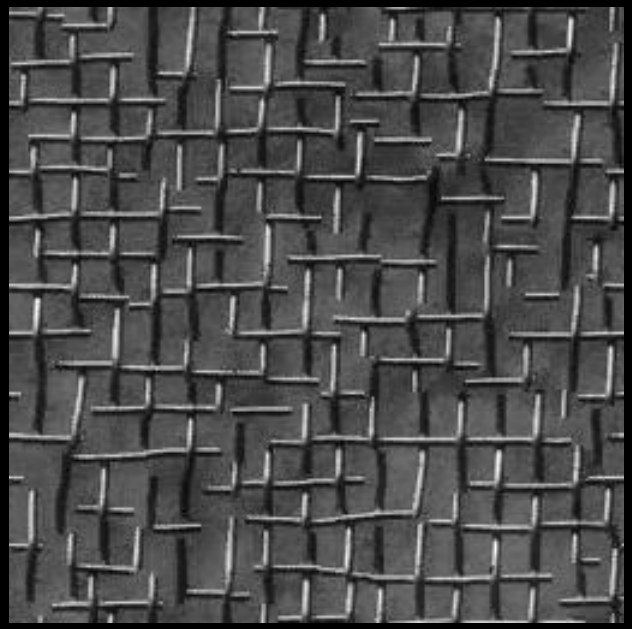
input image



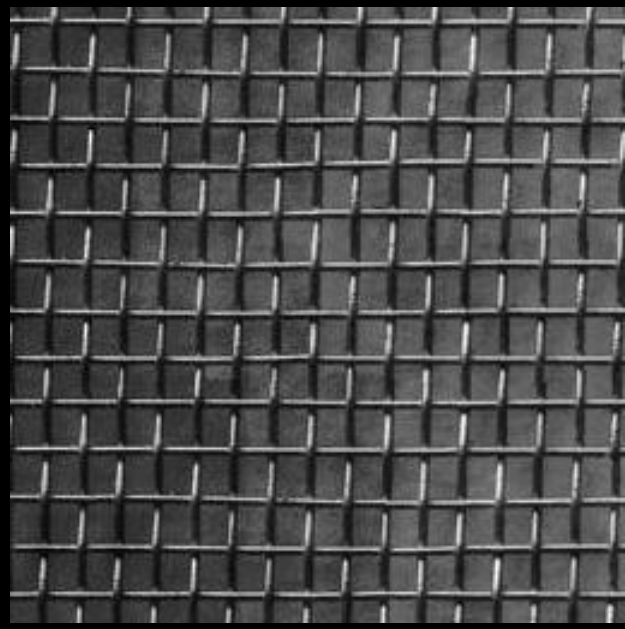
Portilla & Simoncelli



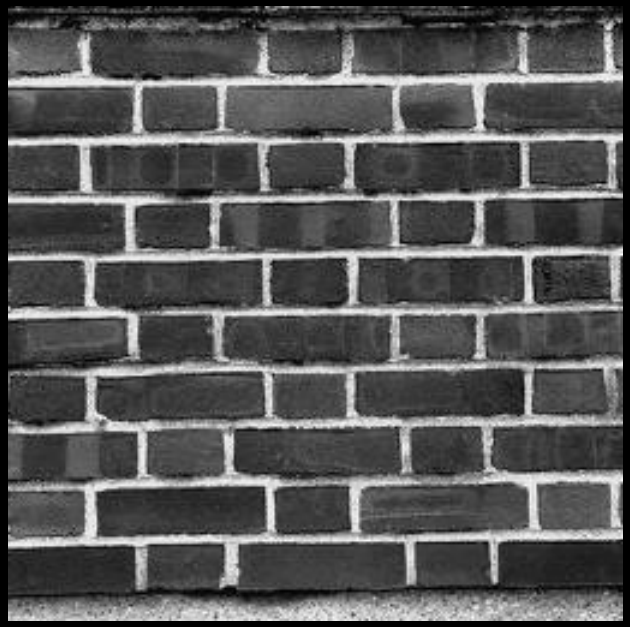
Xu, Guo & Shum



Wei & Levoy

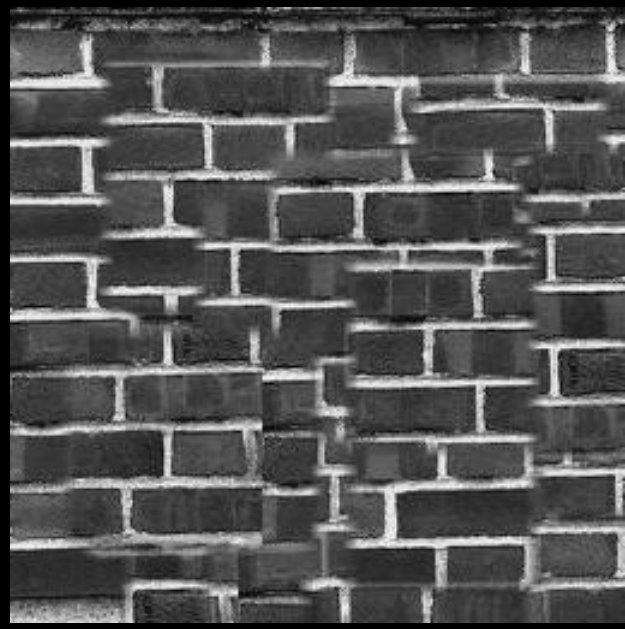


Our algorithm

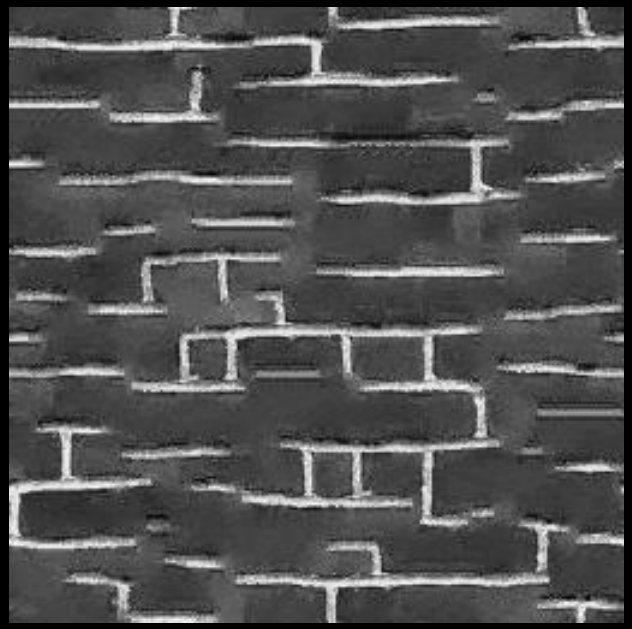


input image

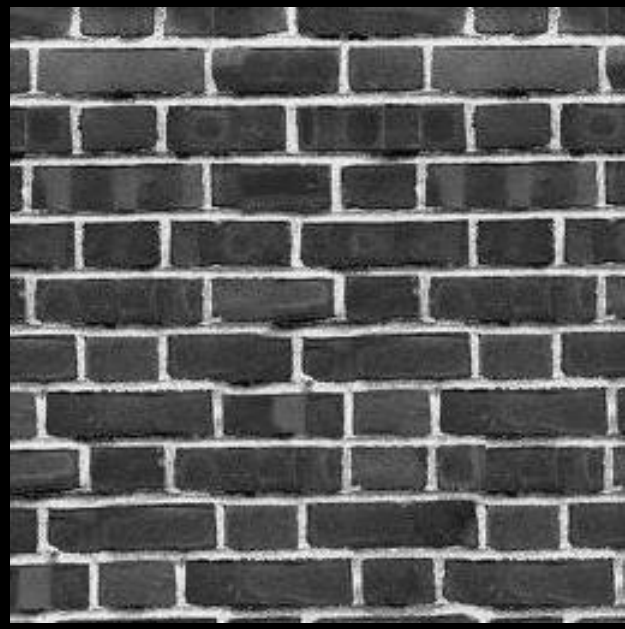
Portilla & Simoncelli



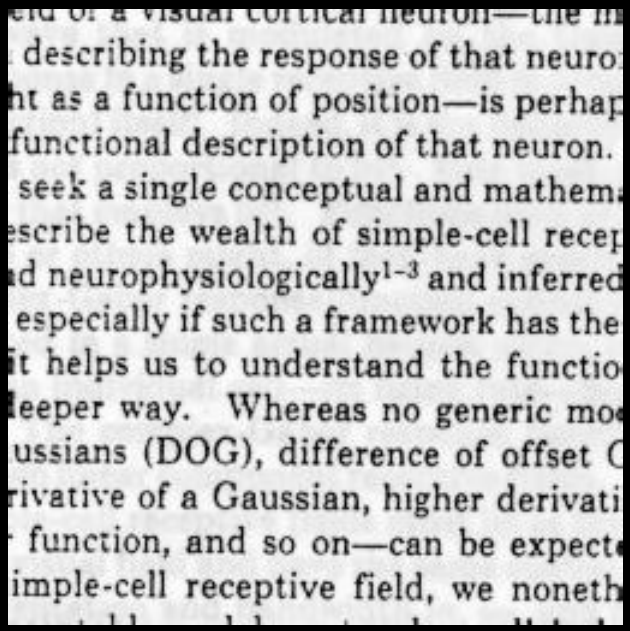
Xu, Guo & Shum



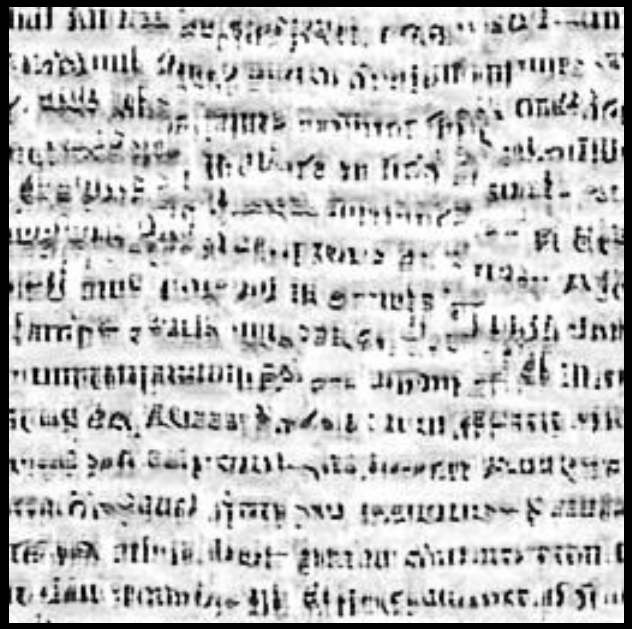
Wei & Levoy



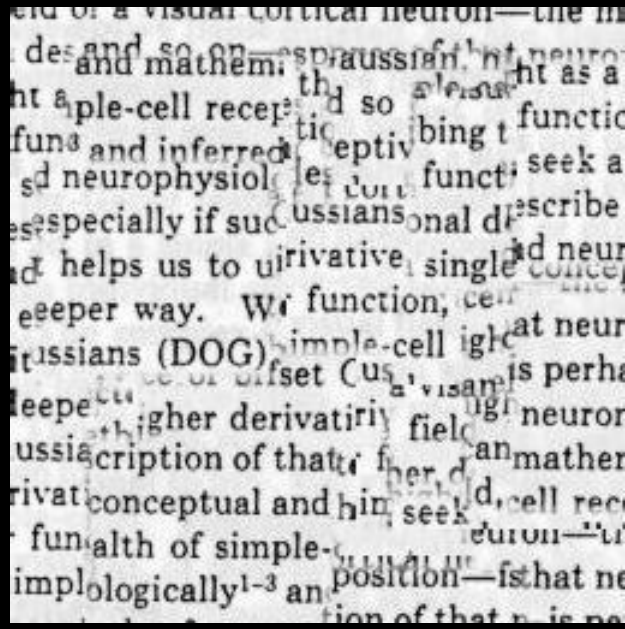
Our algorithm



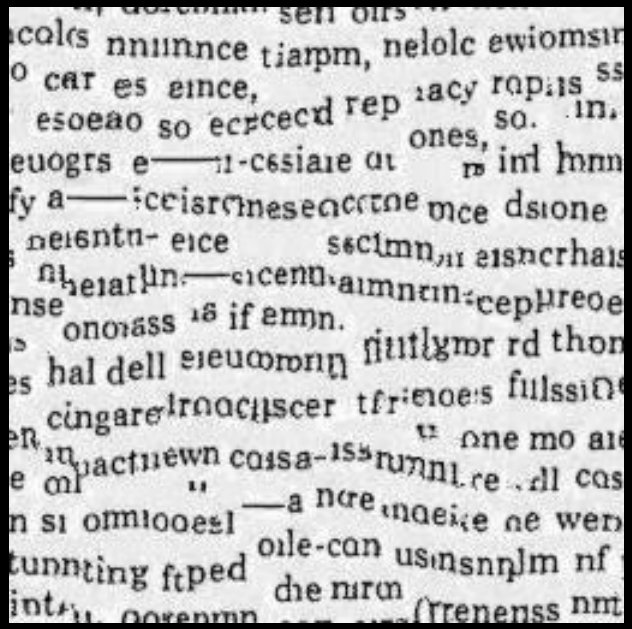

**input image**



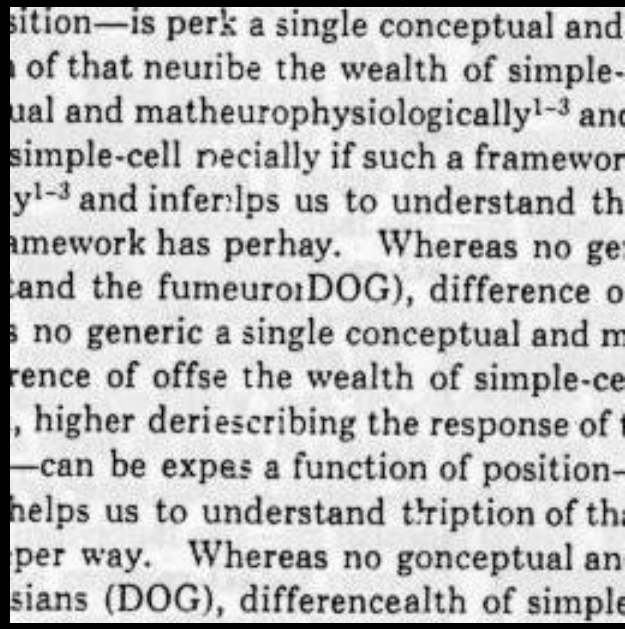
**Portilla & Simoncelli**



**Xu, Guo & Shum**



**Wei & Levoy**



**Our algorithm**

# Application: Texture Transfer

- Try to explain one object with bits and pieces of another object:

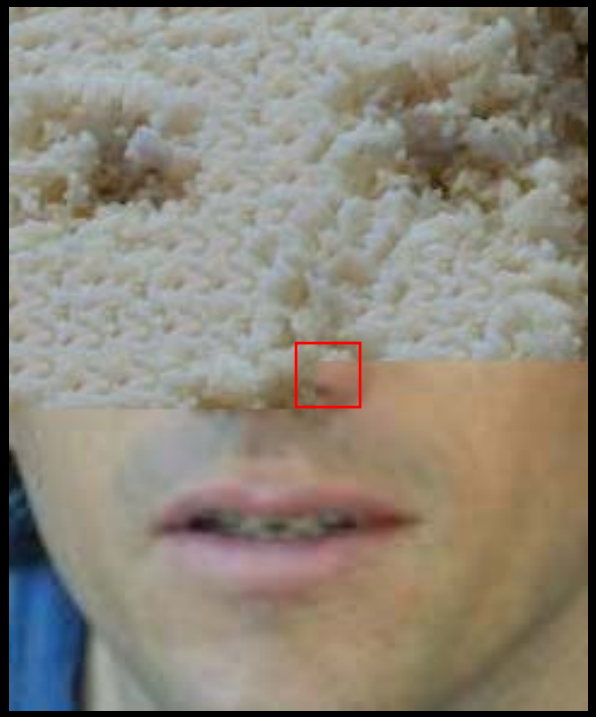
# Texture Transfer

Constraint

Texture sample

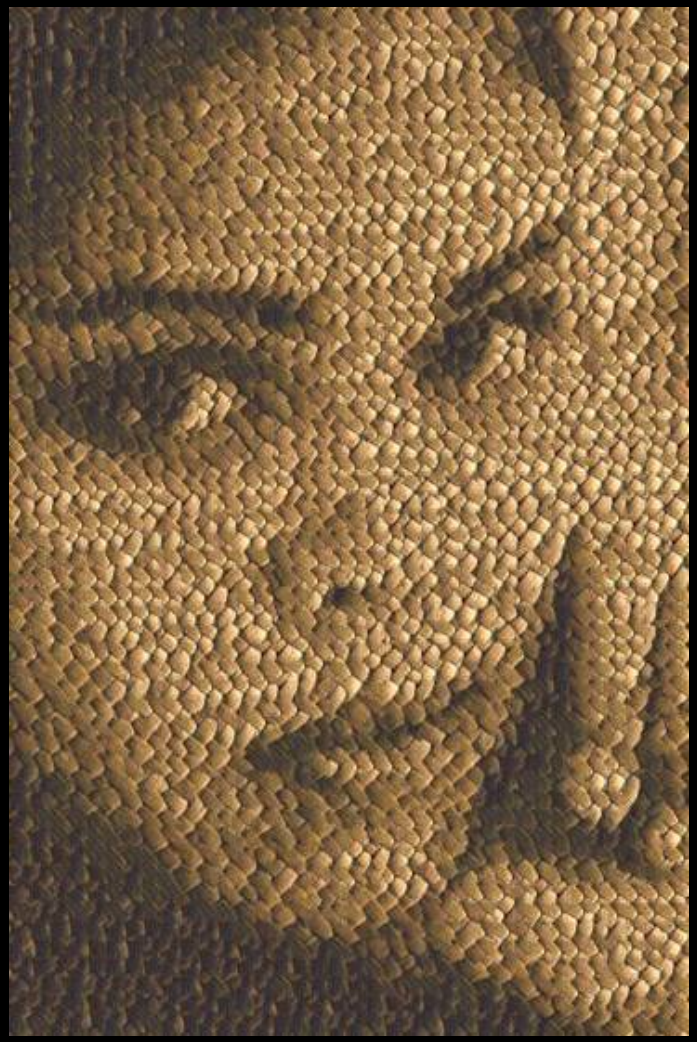
# Texture Transfer

- Take the texture from one image and “paint” it onto another object



Same as texture synthesis, except an additional constraint:

1. Consistency of texture
2. Similarity to the image being “explained”



# Image Analogies

Aaron Hertzmann[1,2]

Chuck Jacobs[2]

Nuria Oliver[2]

Brian Curless[3]

David Salesin[2,3]

[1]New York University

[2]Microsoft Research

[3]University of Washington

# Image Analogies

A

A’

B

B’

# Blur Filter

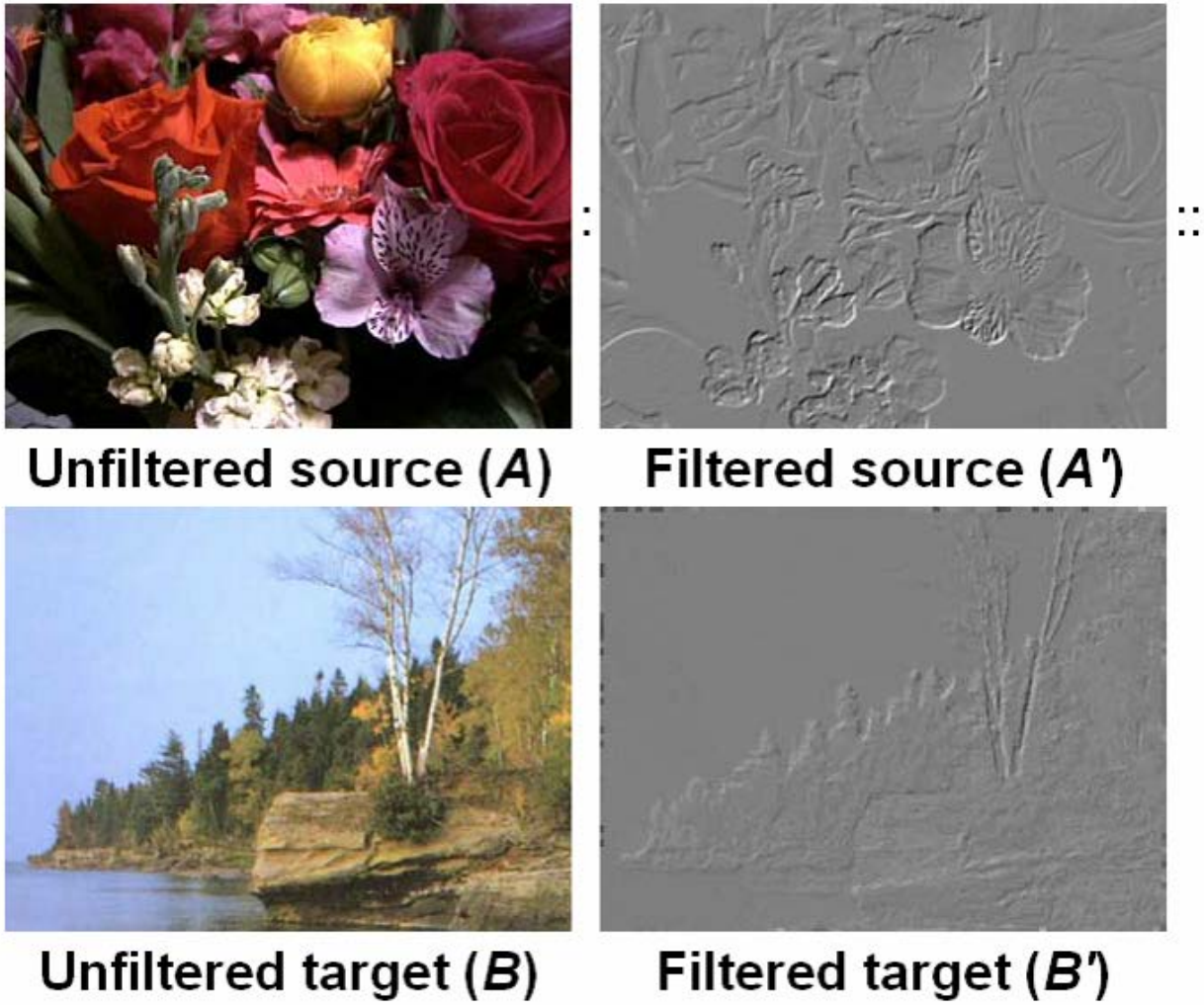
Unfiltered source ($A$)

Filtered source ($A'$)

Unfiltered target ($B$)

Filtered target ($B'$)

# Edge Filter


Unfiltered source (A)
Filtered source (A')
Unfiltered target (B)
Filtered target (B')

# Artistic Filters

A

A’

B

B’

# Image Analogies

Goal: Process an image by example

A

:

A’

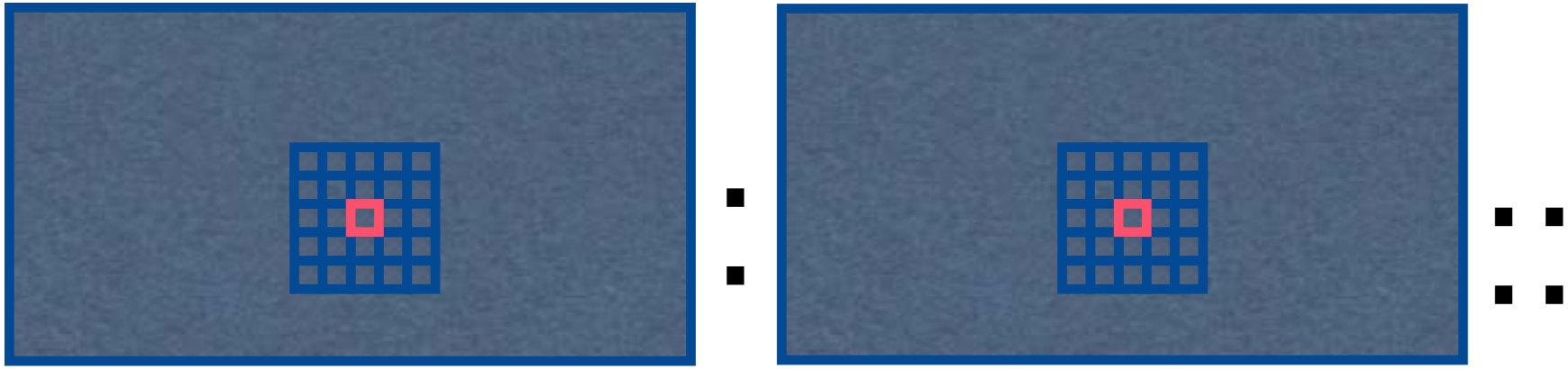
::

B

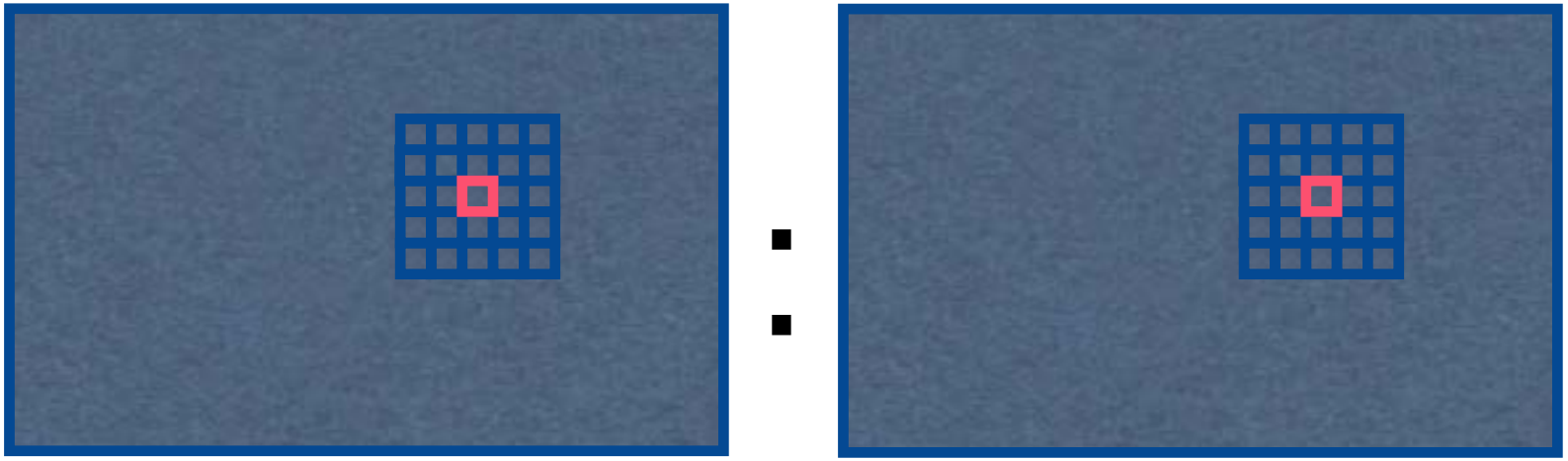
:


B’

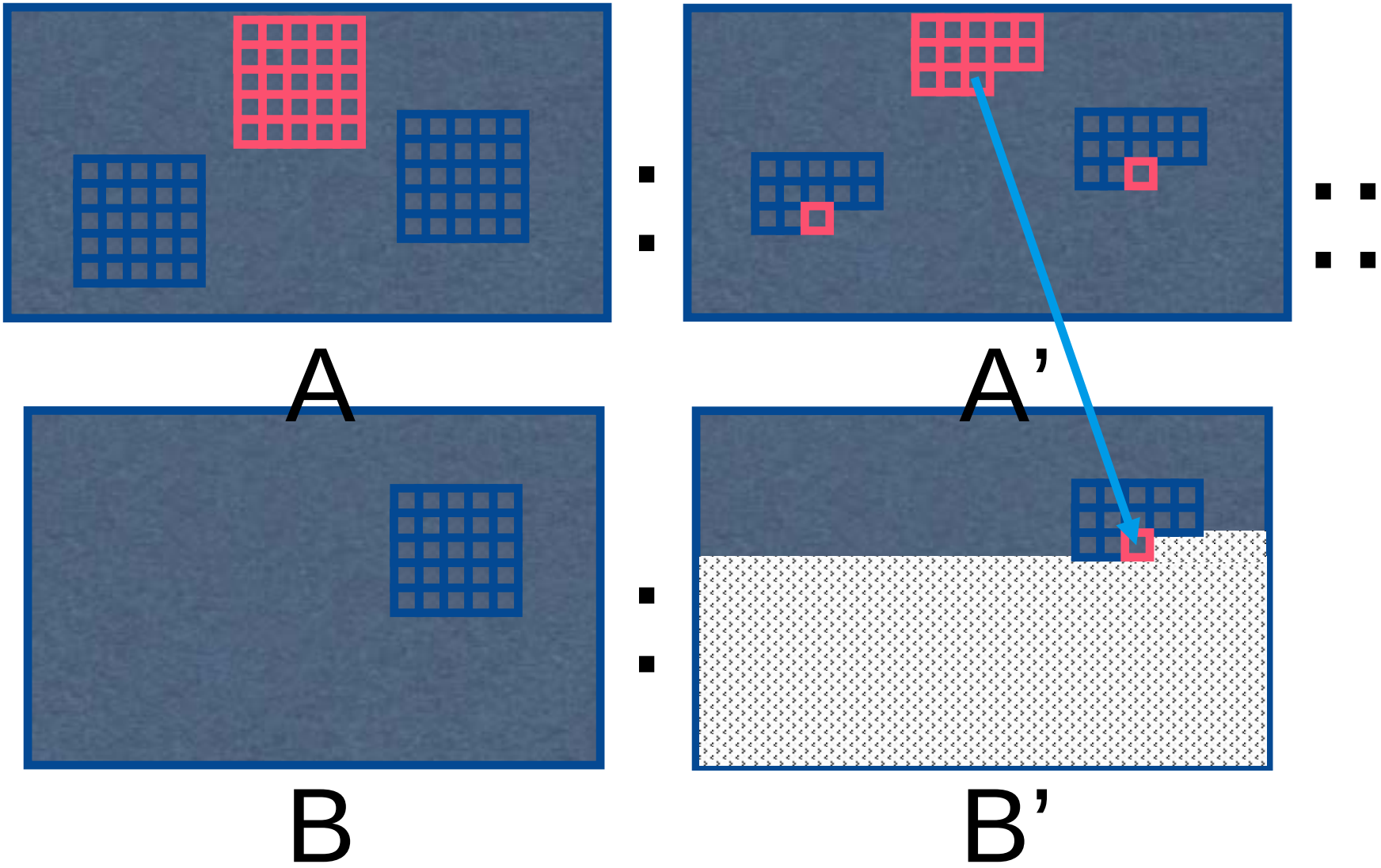
Hertzmann et al. SIGGRAPH 2001



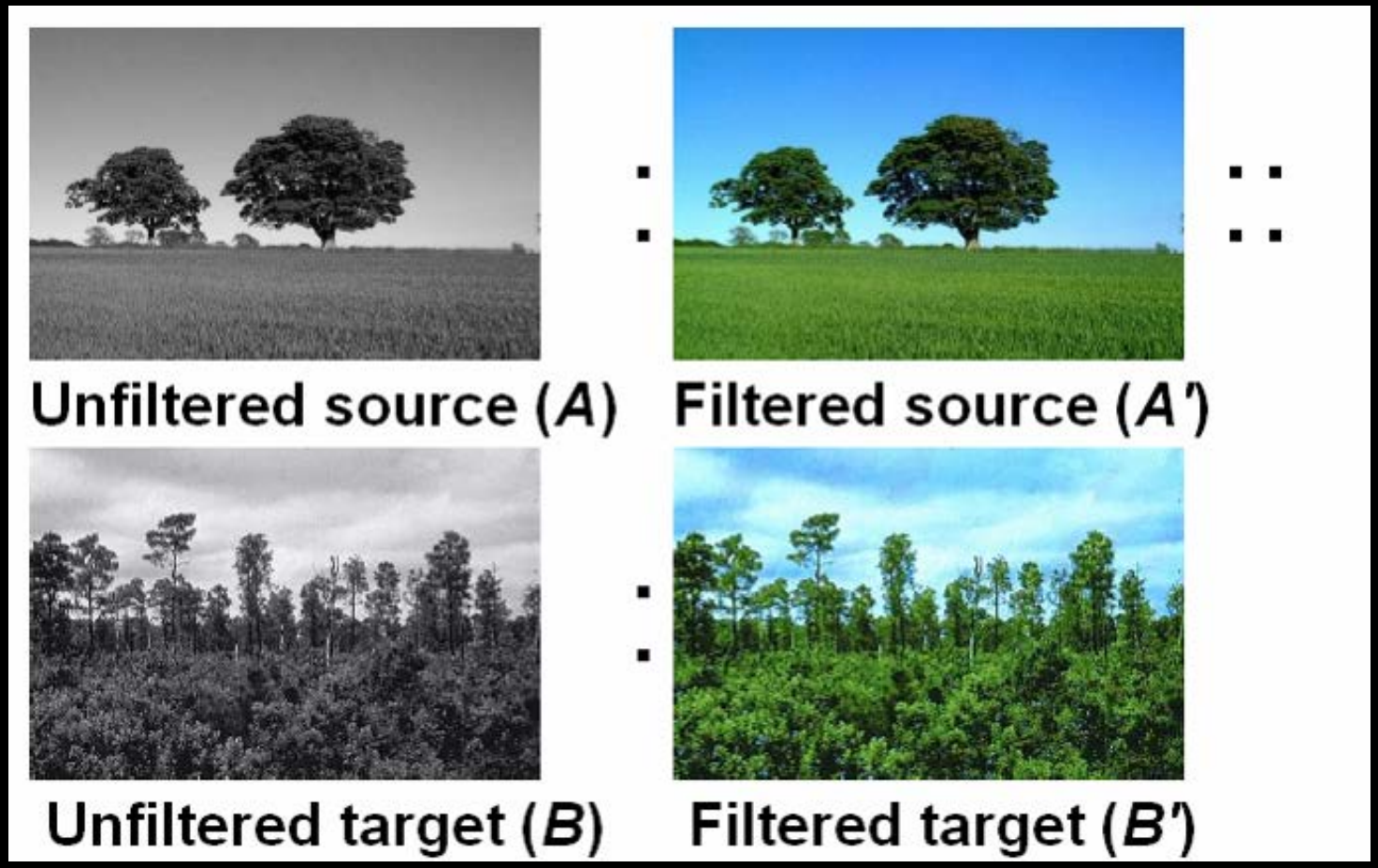
A A’



B B’


A
A’
B
B’

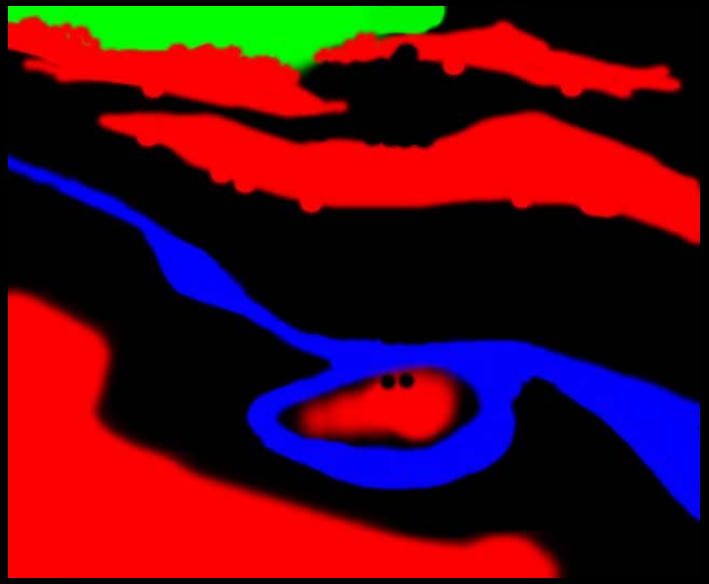
# Colorization



# Texture-by-numbers

A

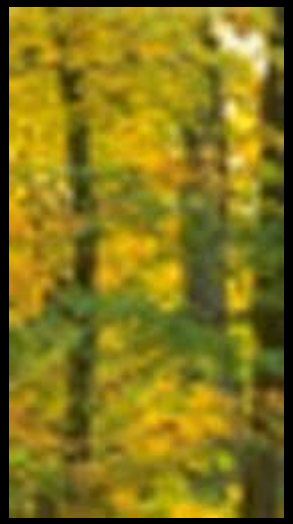
A’



B

B’

# Super-resolution



A

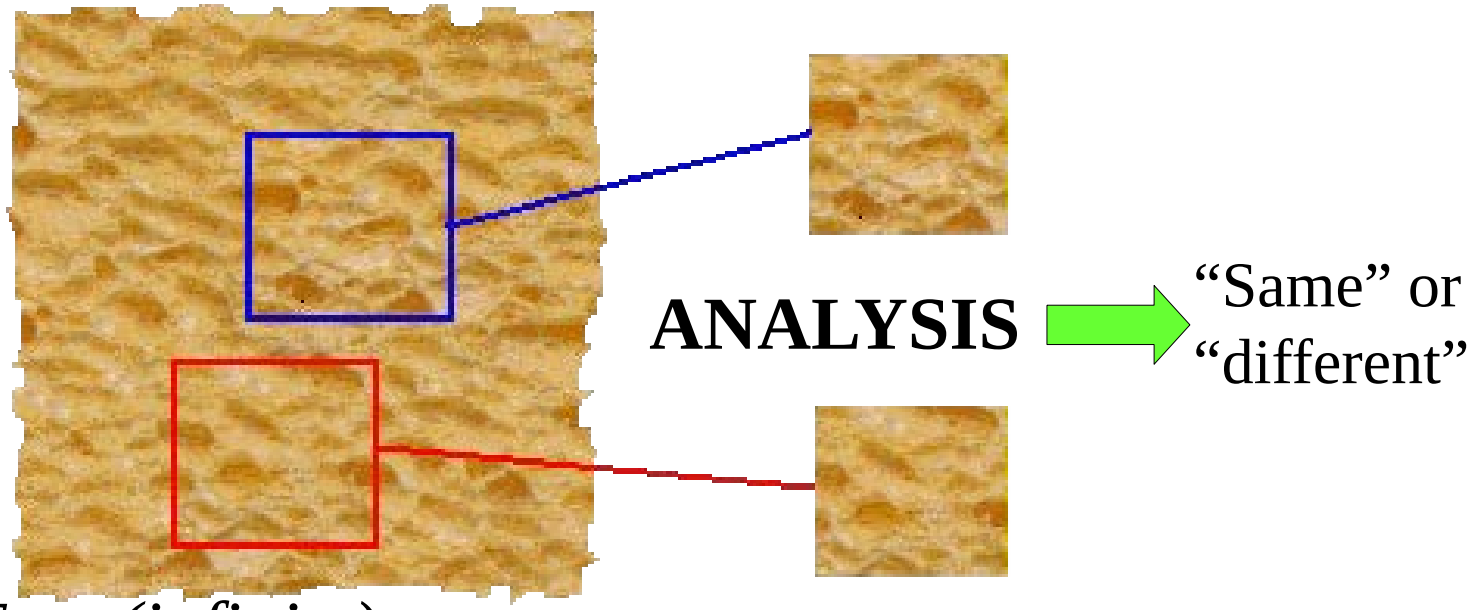
A’

# Super-resolution (result!)

B

B’

# Texture Analysis

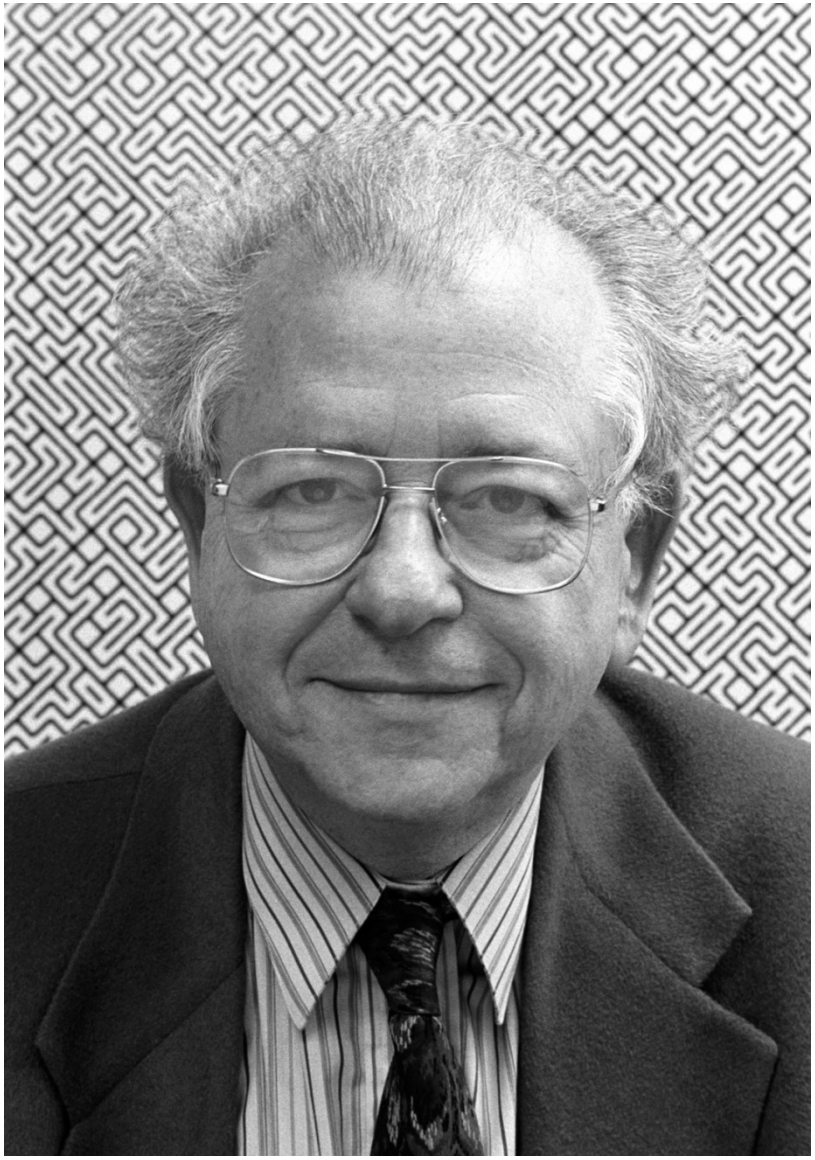



*True (infinite) texture*

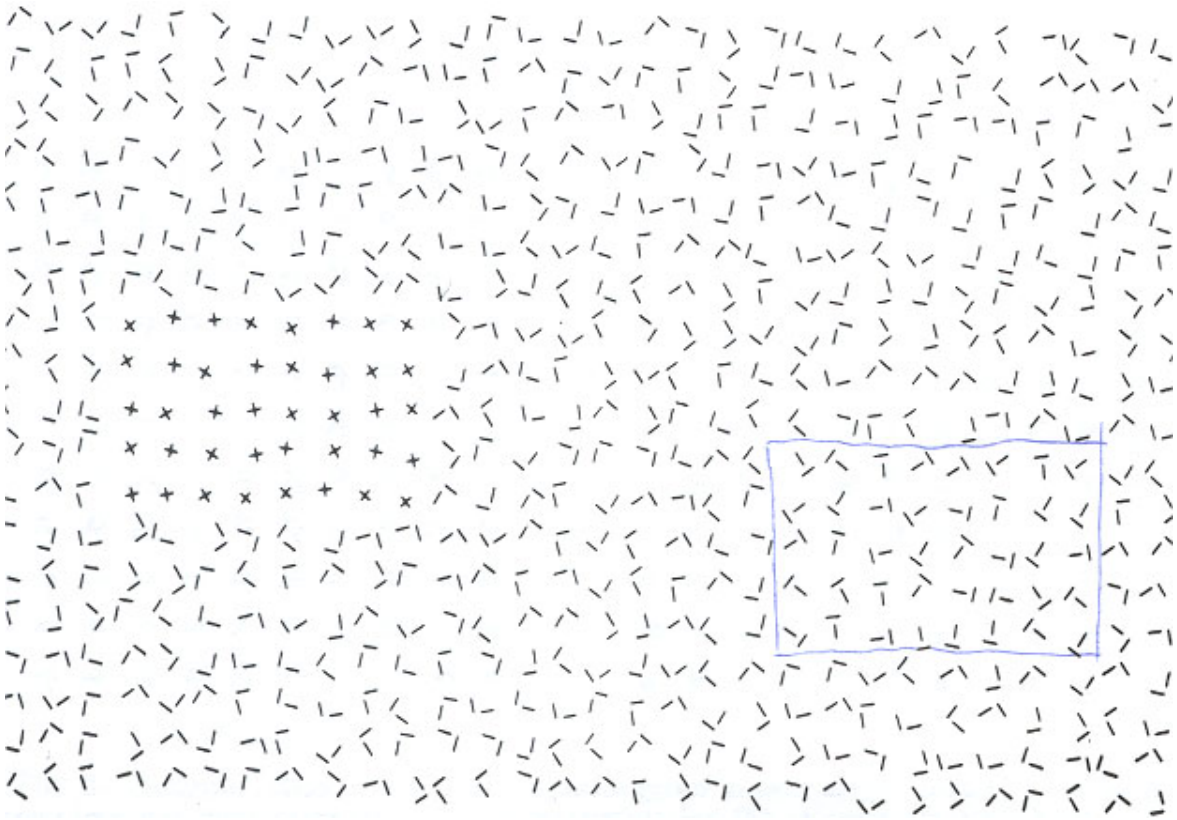
Compare textures and decide if they’re made of the same “stuff”.

# When are two textures similar?

# Béla Julesz, father of texture



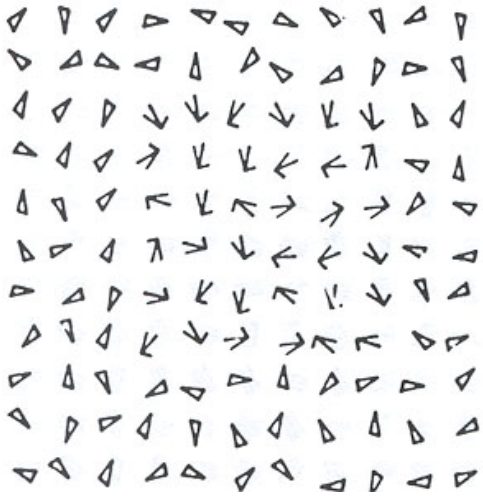
# Texton Discrimination (Julesz)

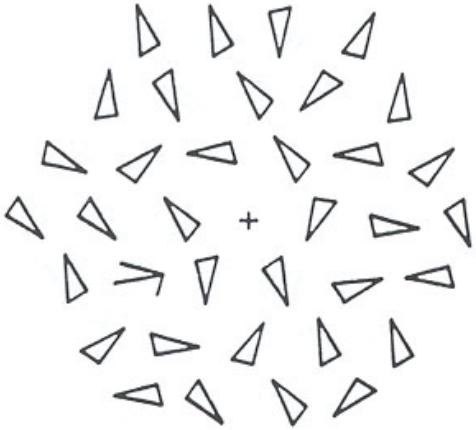


Human vision is sensitive to the difference of some types of elements and appears to be “numb” on other types of differences.

# Search Experiment I

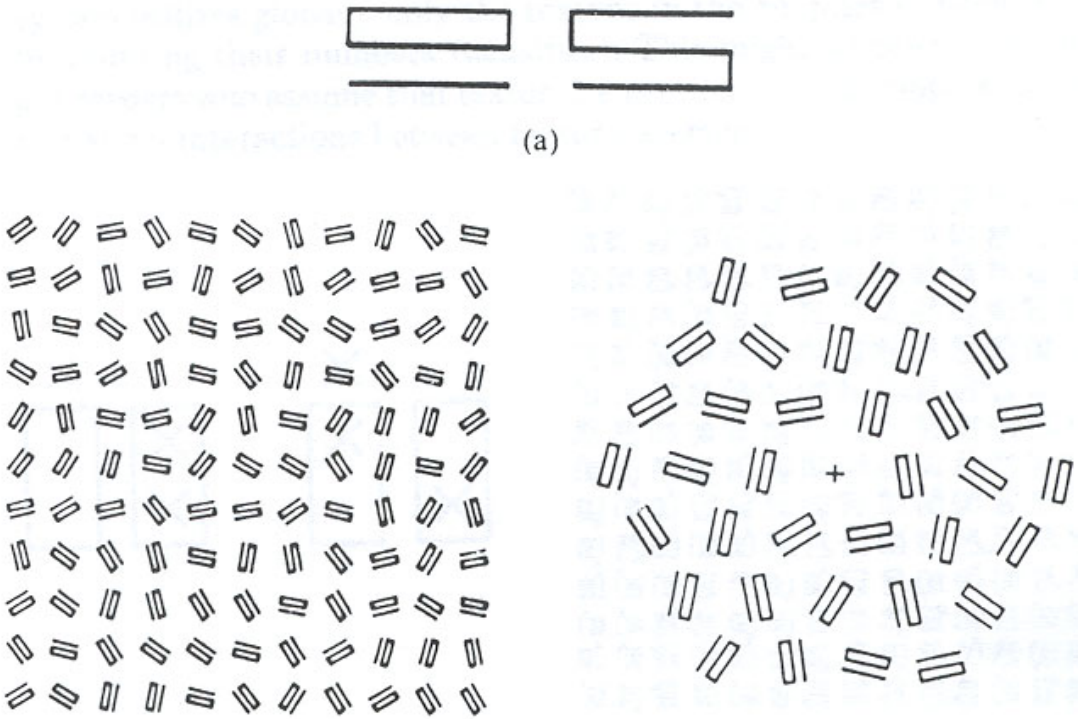
(a)





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




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.

# Heuristic (Axiom) I

Julesz then conjectured the following axiom:

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.

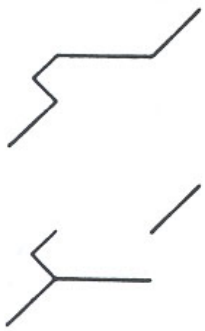
2. Attentive vision

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

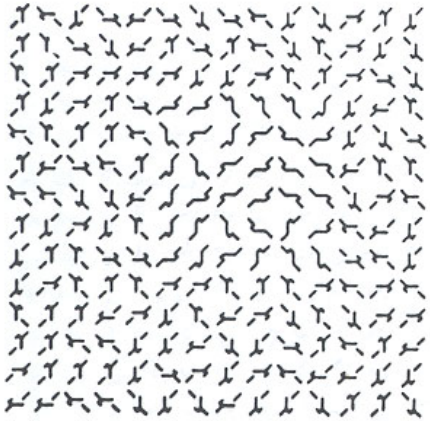
# Examples

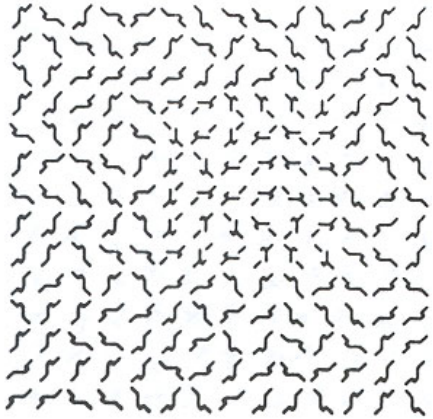
Pre-attentive vision is sensitive to size/width, orientation changes

# Examples



(a)





Sensitive to number of terminators

Left: fore-back
Right: back-fore

See previous examples
For cross and terminators

# Heuristic (Axiom) II

Textons are the fundamental elements in preattentive vision, including

1. Elongated blobs

   rectangles, ellipses, line segments with attributes
   color, orientation, width, length, flicker rate.

2. Terminators

   ends of line segments.

3. Crossings of line segments.

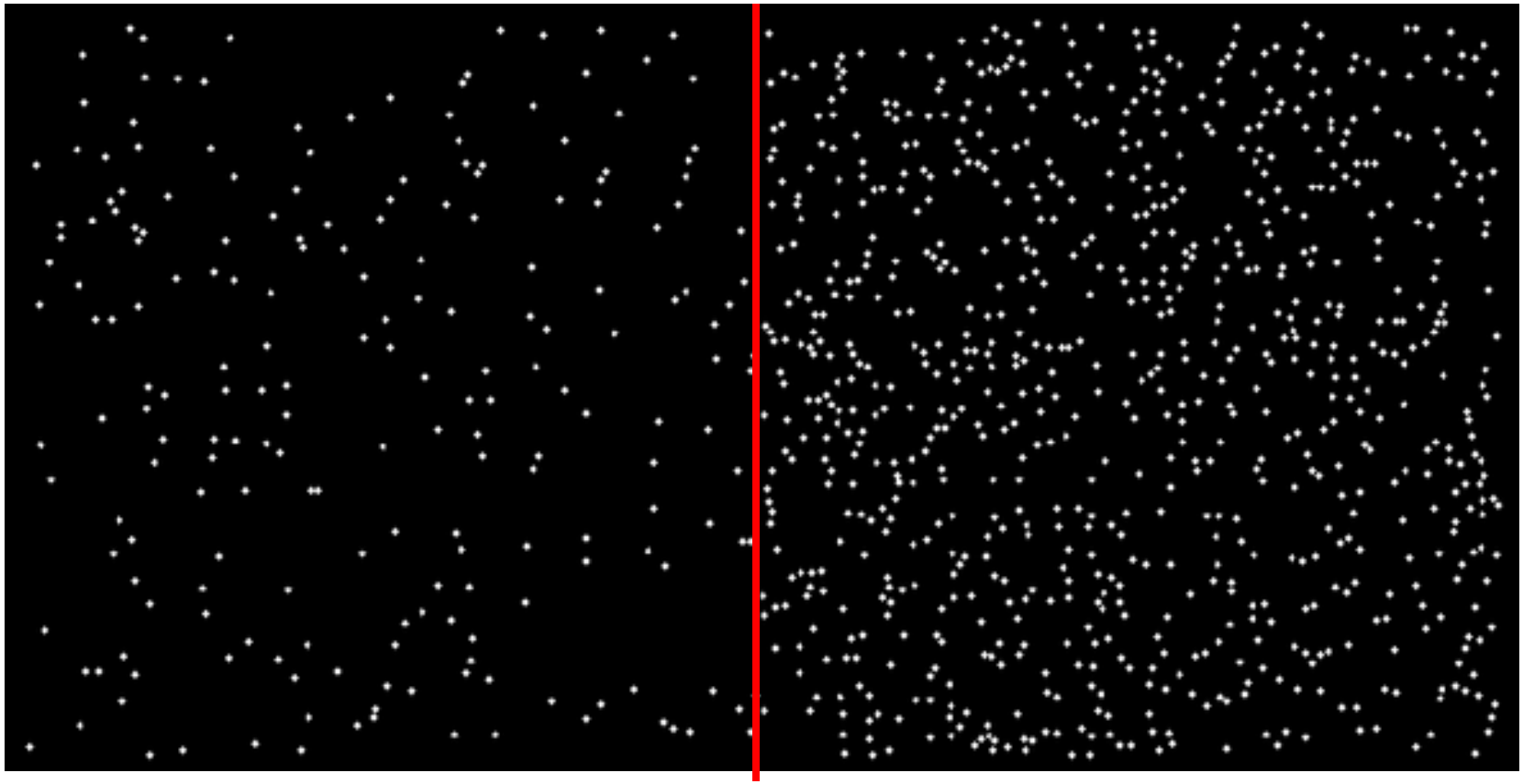
But it is worth noting that Julesz's conclusions are largely based on ensemble of artificial texture patterns. It was infeasible to synthesize natural textures for controlled experiments at that time.

# Julesz Conjecture

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

(later proved wrong)

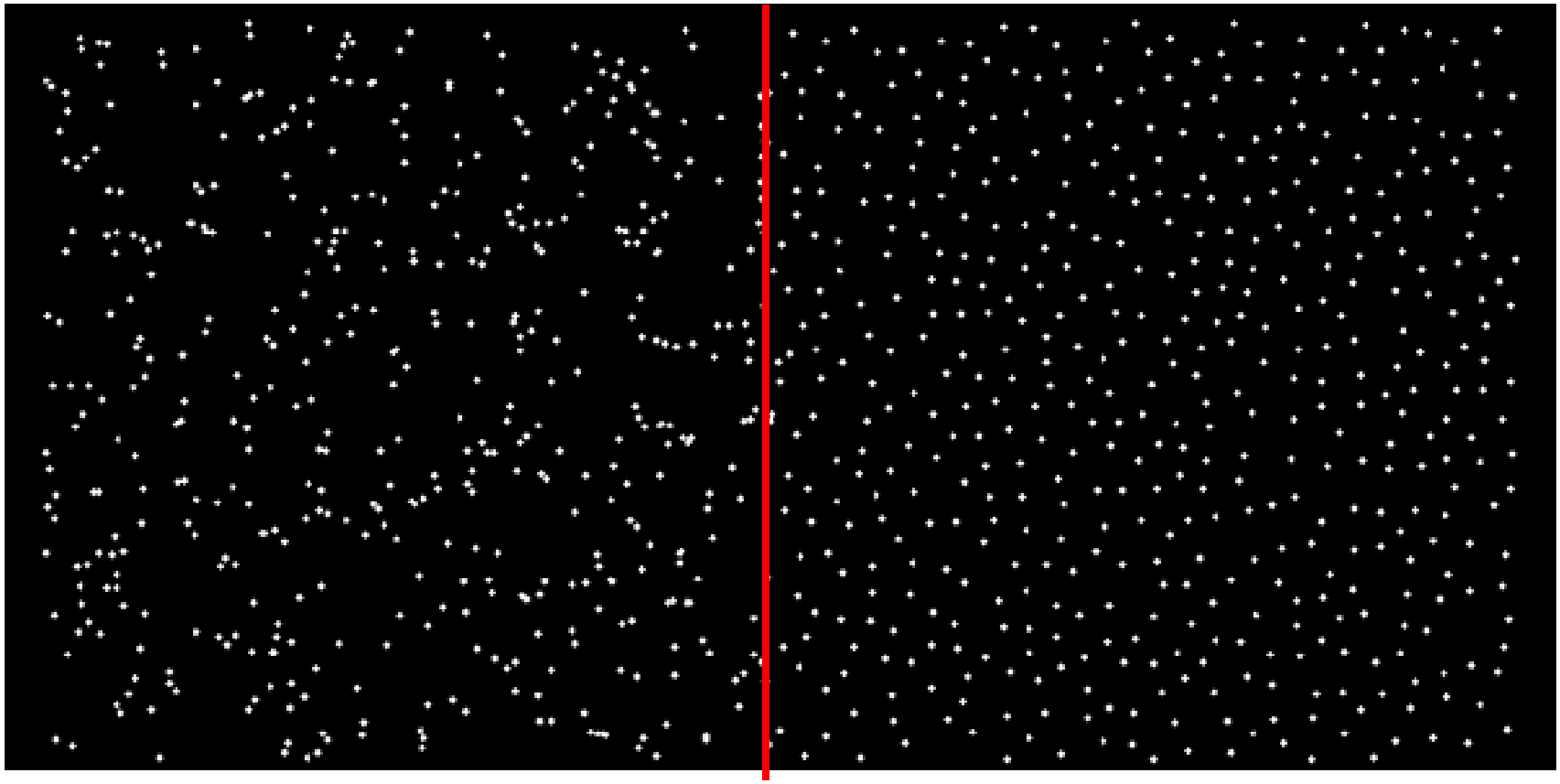
# 1st Order Statistics



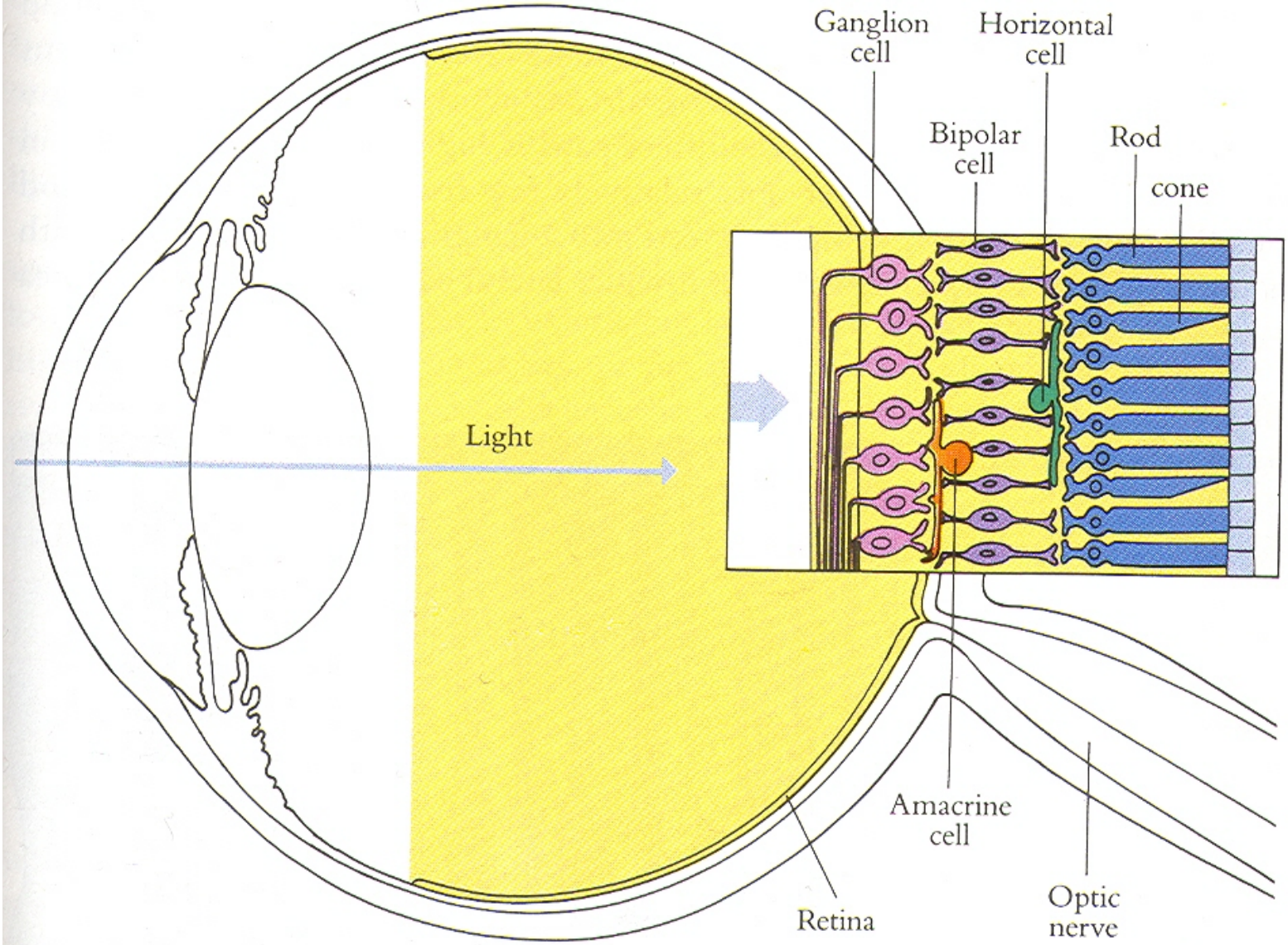
5% white

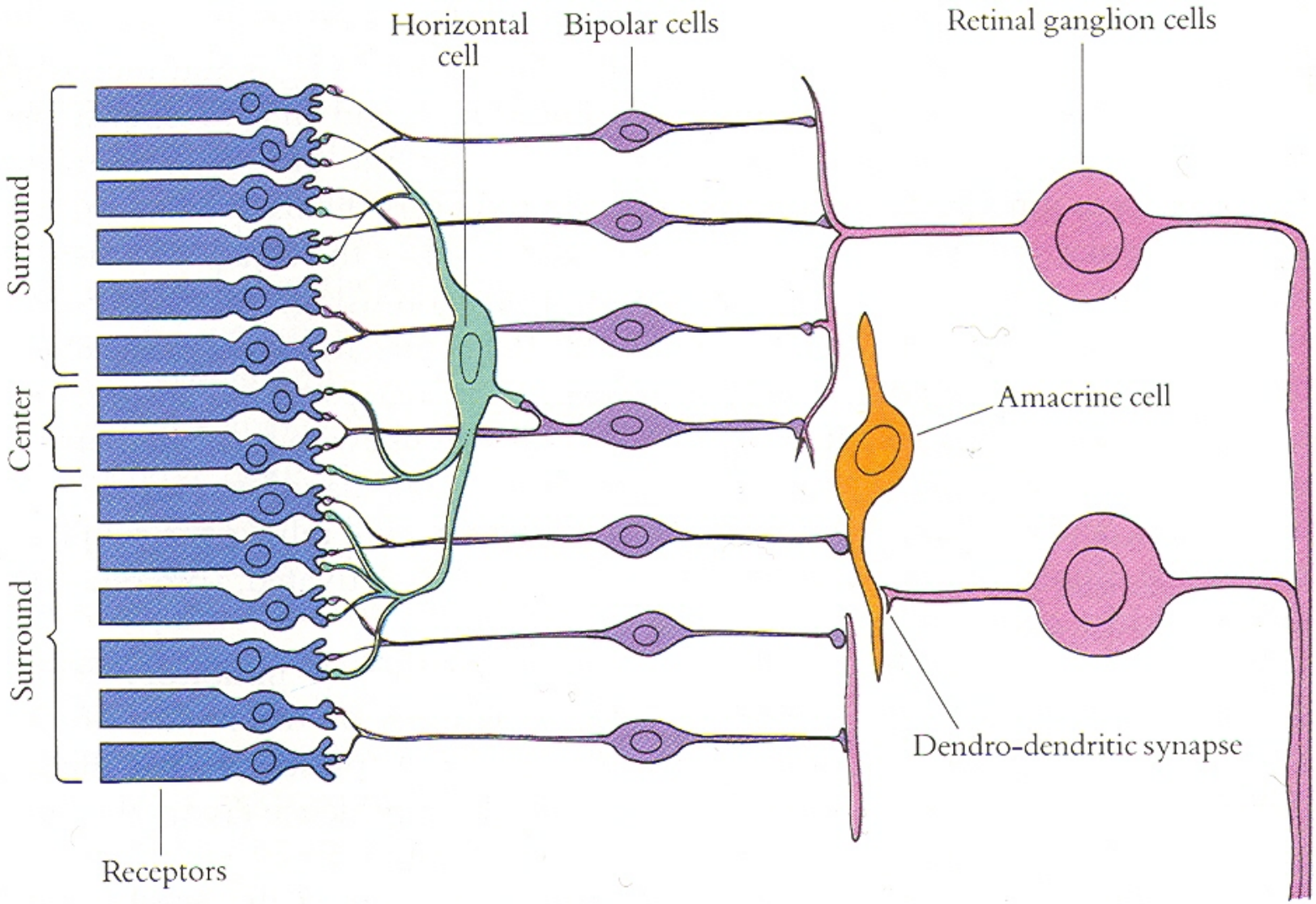
20% white

# 2nd Order Statistics

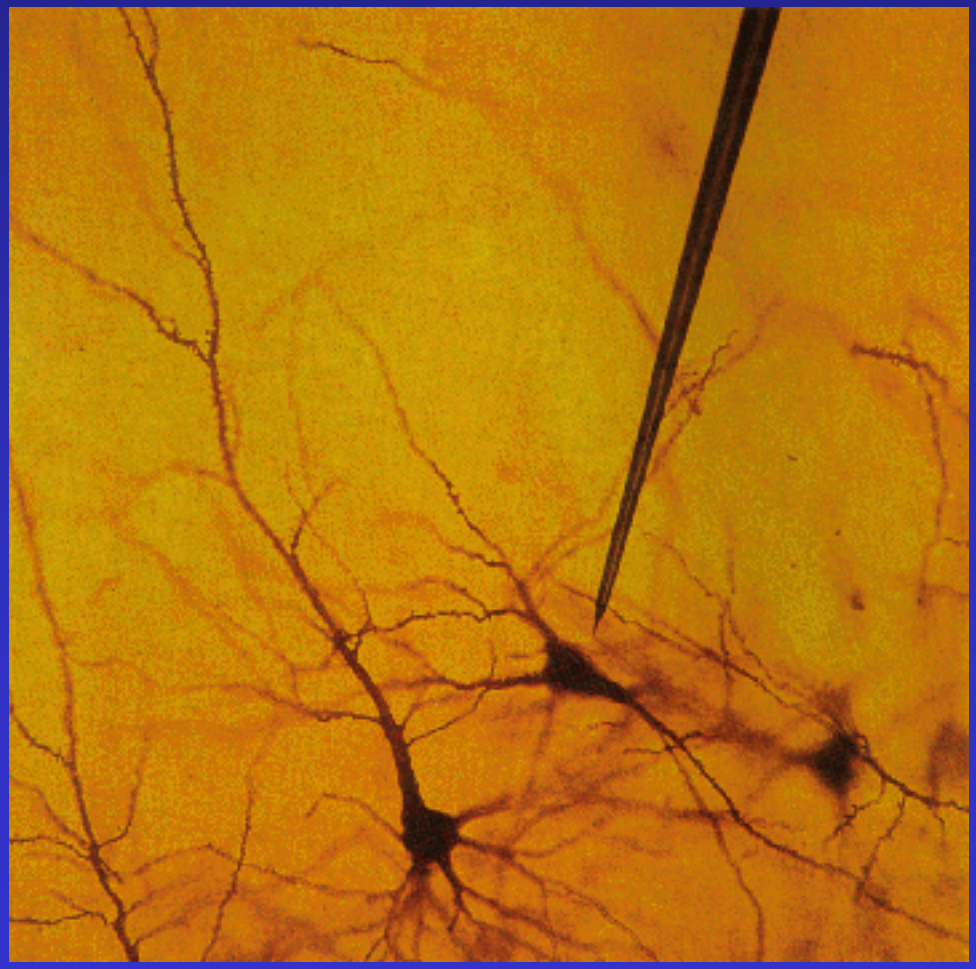


10% white


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

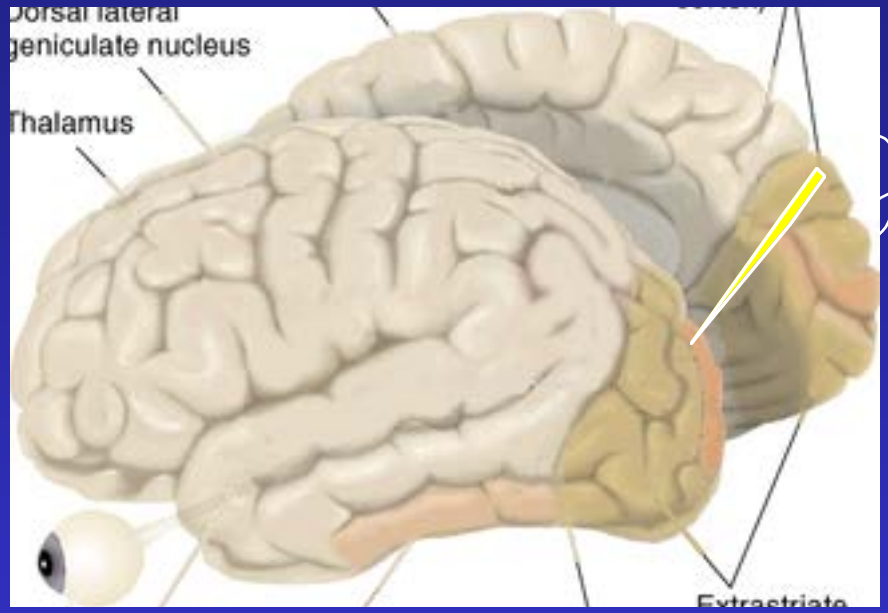

Horizontal cell
Bipolar cells
Retinal ganglion cells
Surround
Center
Surround
Amacrine cell
Dendro-dendritic synapse
Receptors

# Single Cell Recording



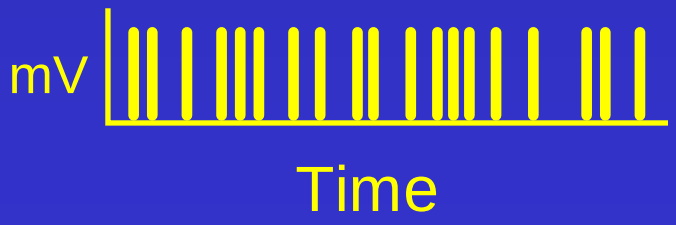
# Single Cell Recording

Microelectrode



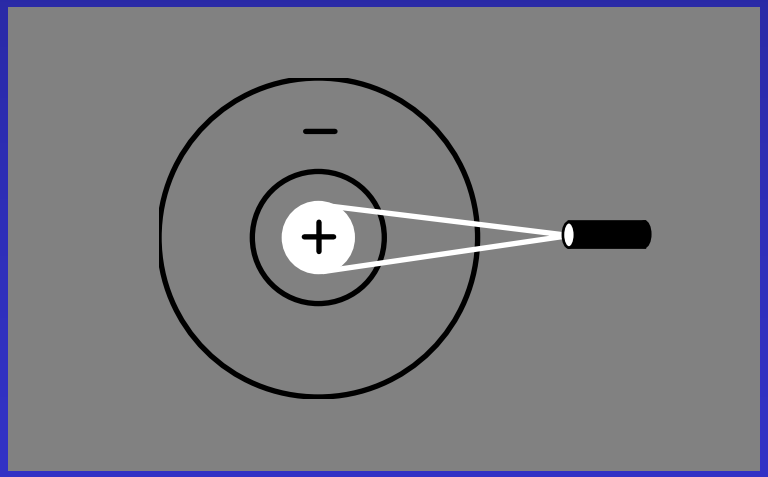

Amplifier

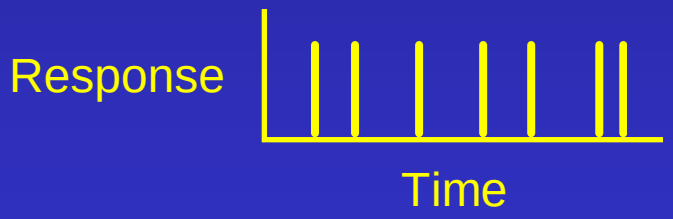
Electrical response
(action potentials)



# Retinal Receptive Fields

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



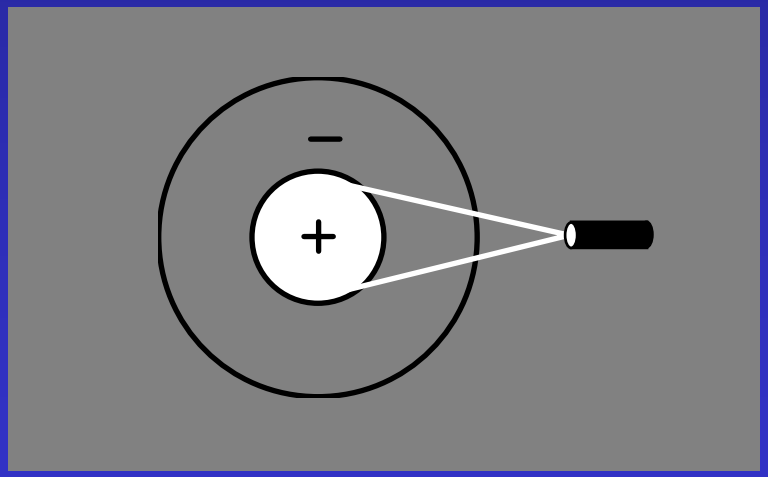



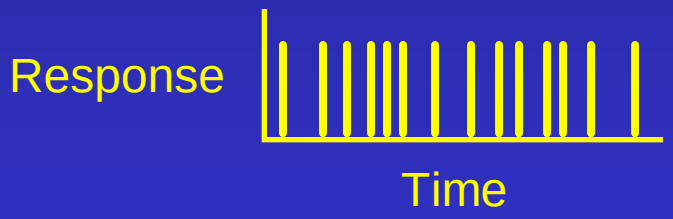

Stimulus condition

Electrical response

# Retinal Receptive Fields

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



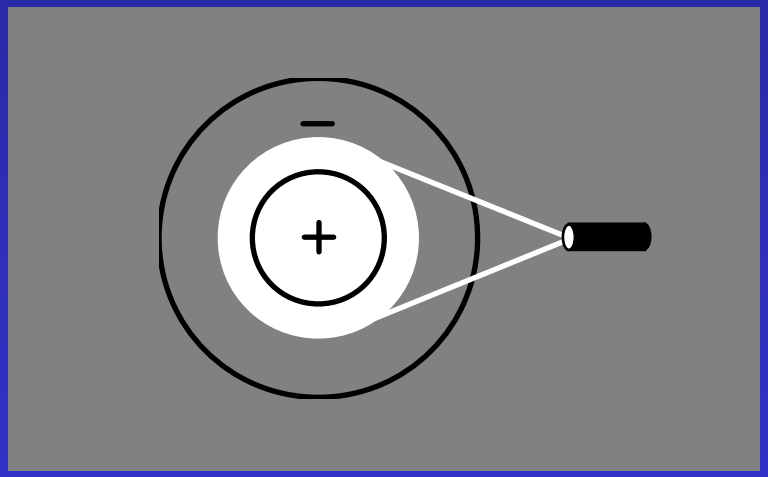



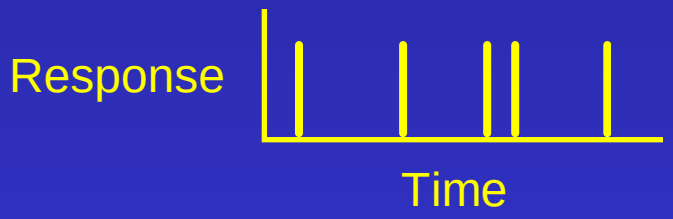

Stimulus condition

Electrical response

# Retinal Receptive Fields

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



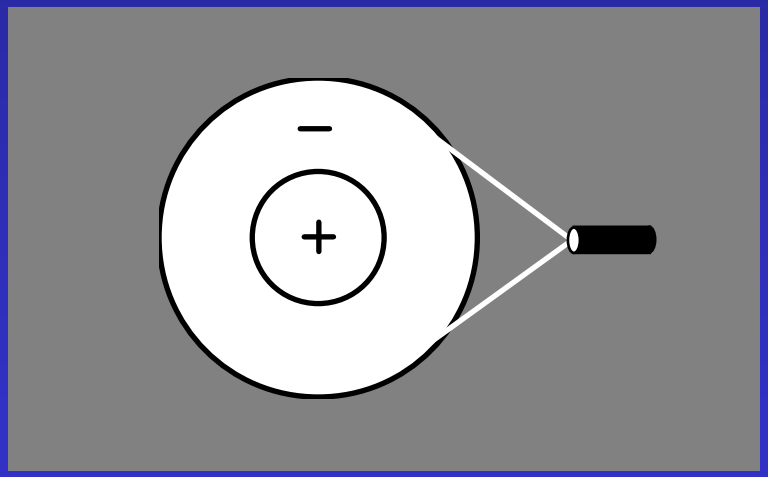



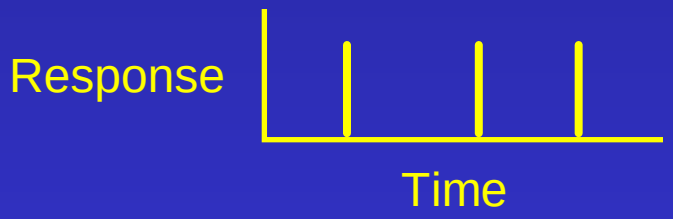

Stimulus condition

Electrical response

# Retinal Receptive Fields

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



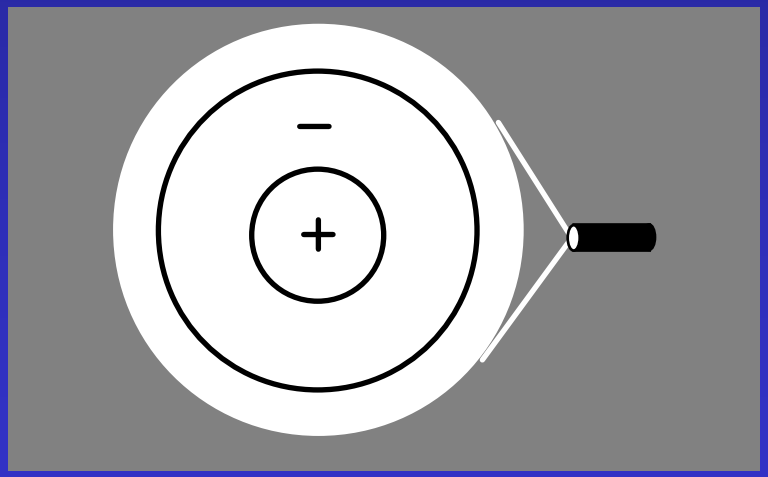



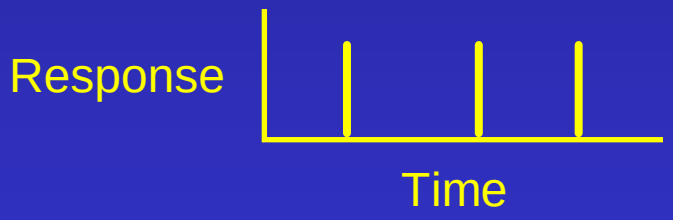

Stimulus condition

Electrical response

# Retinal Receptive Fields

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



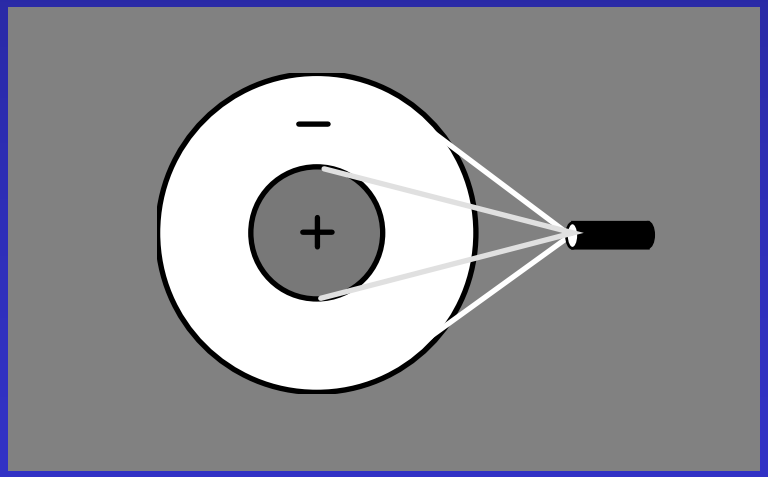



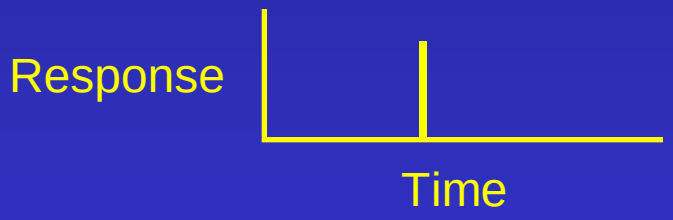

Stimulus condition

Electrical response

# Retinal Receptive Fields

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



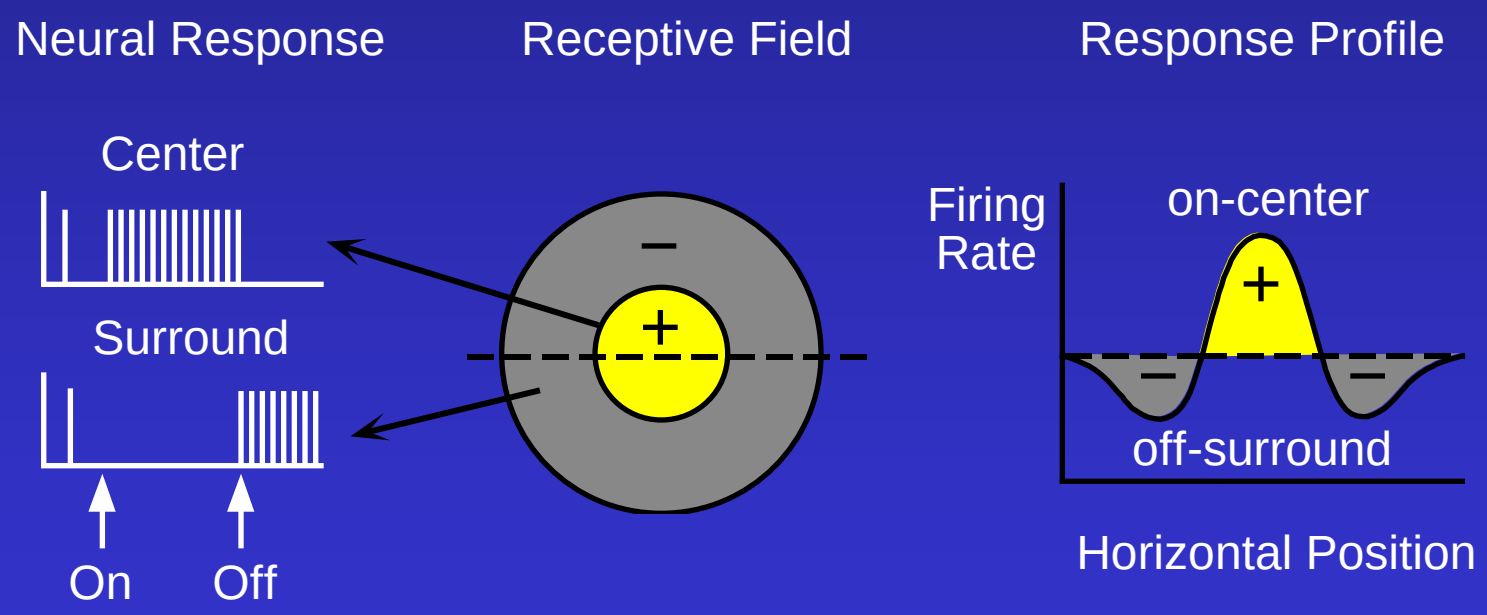




Stimulus condition

Electrical response

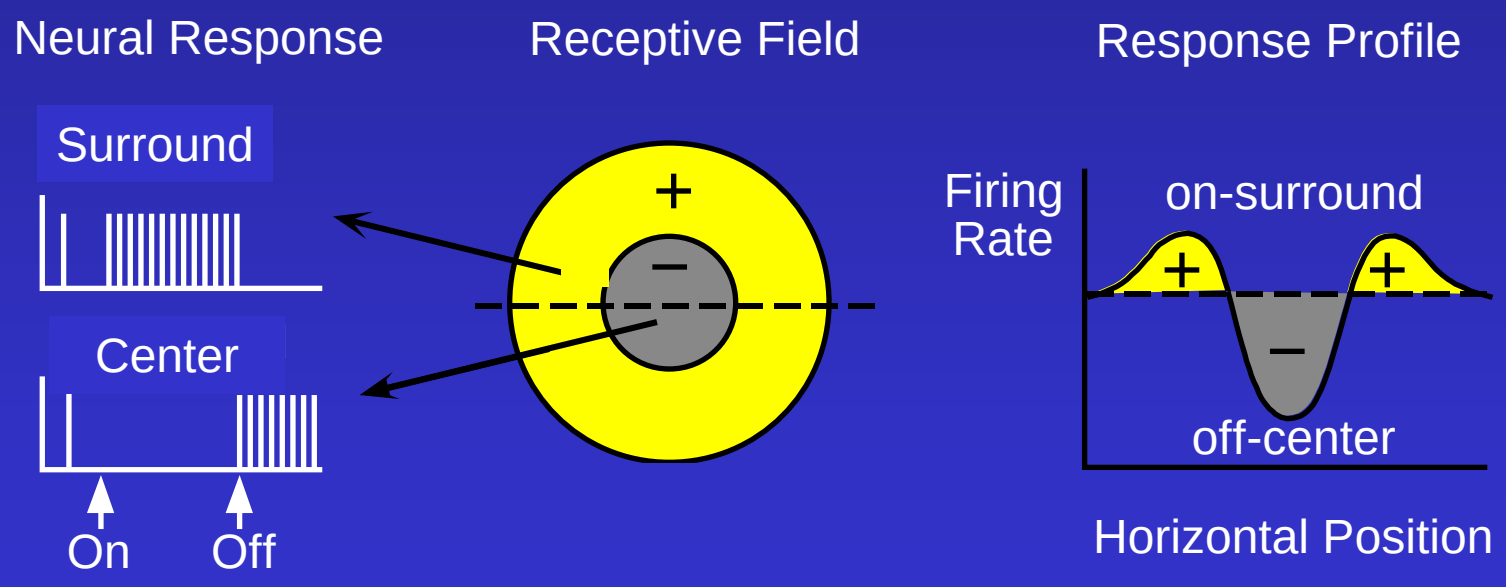
# Retinal Receptive Fields

## RF of On-center Off-surround cells

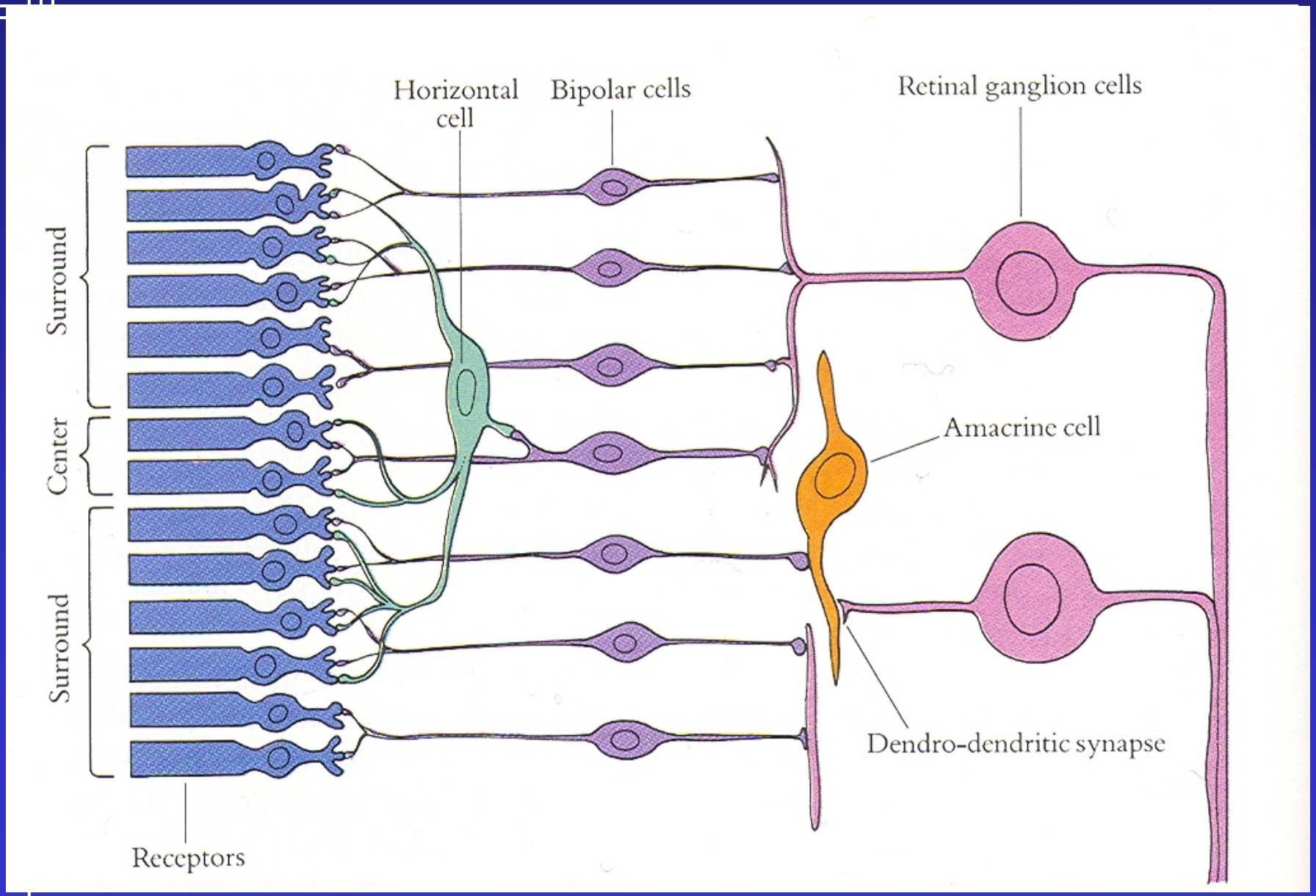


# Retinal Receptive Fields

RF of Off-center On-surround cells

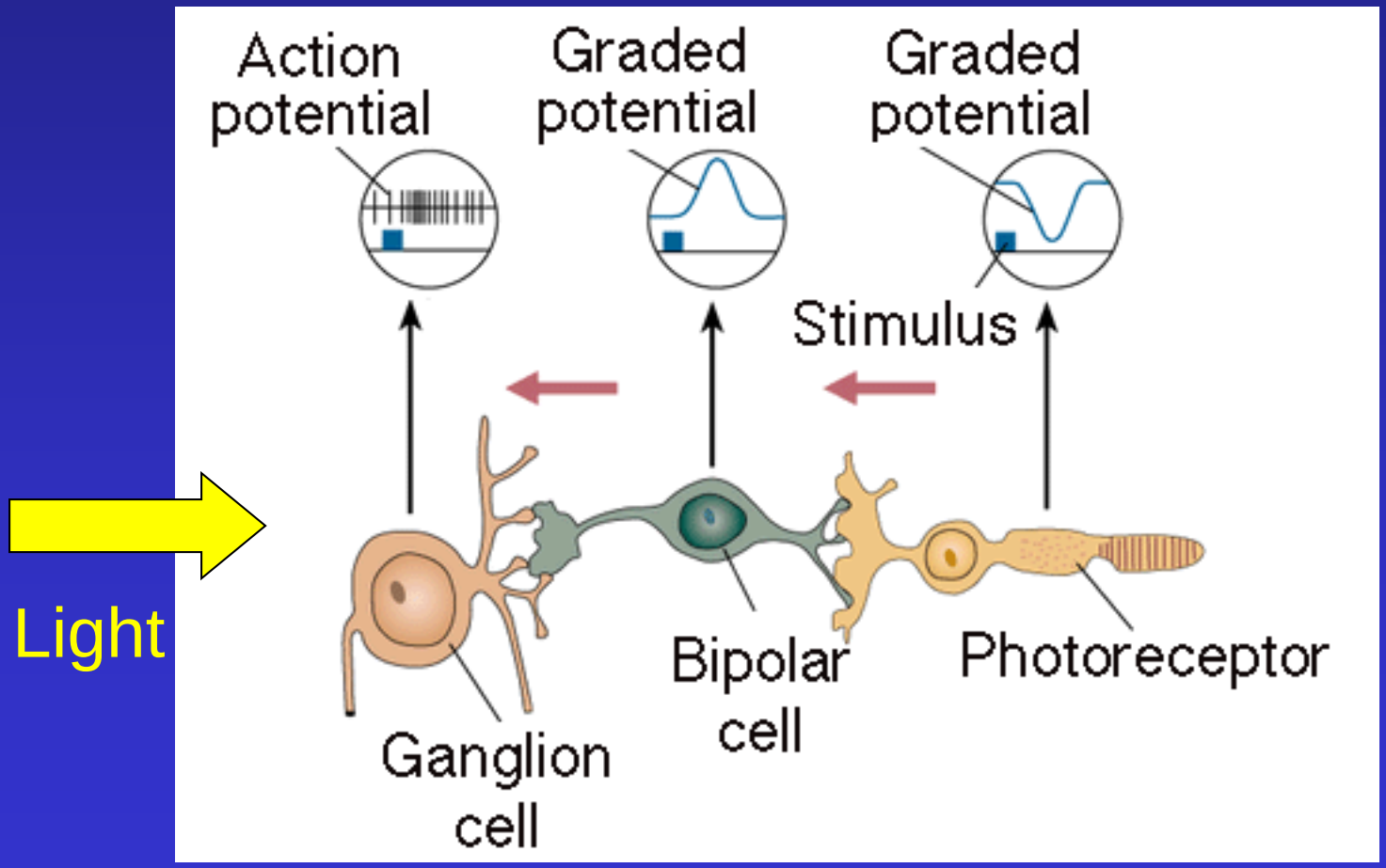


# Retinal Receptive Fields



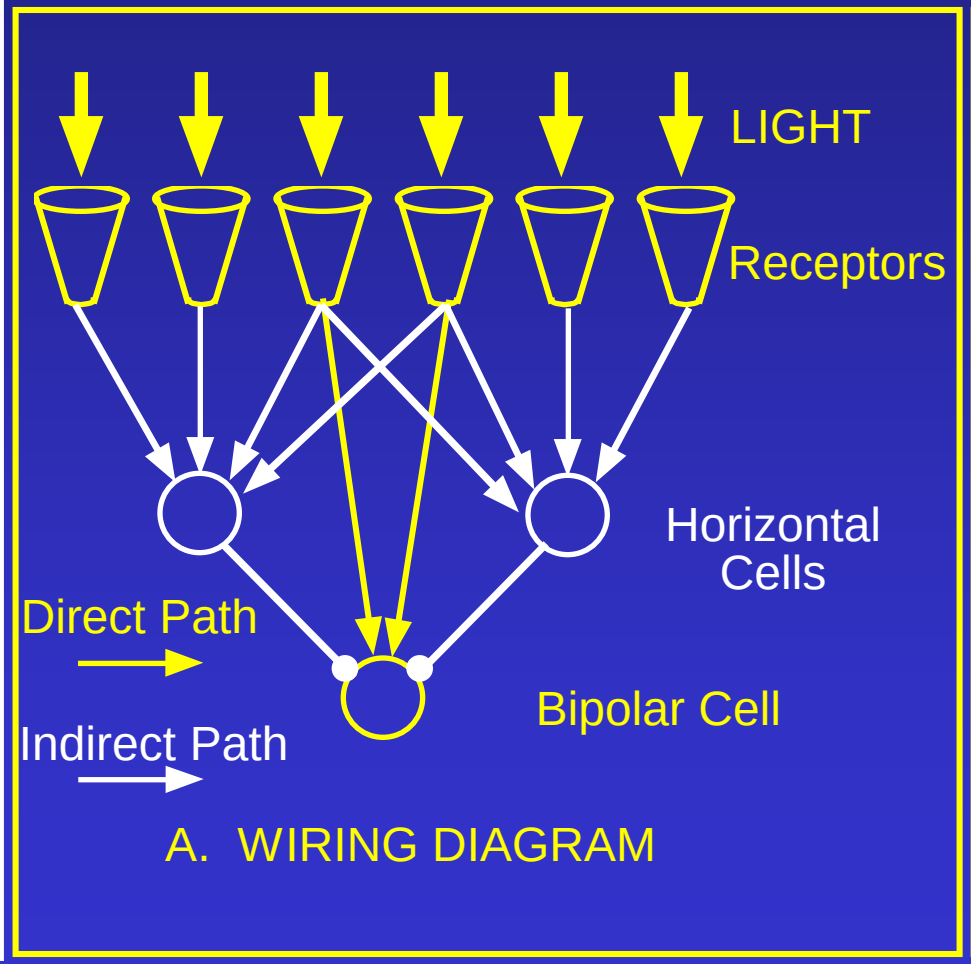
# Retinal Receptive Fields

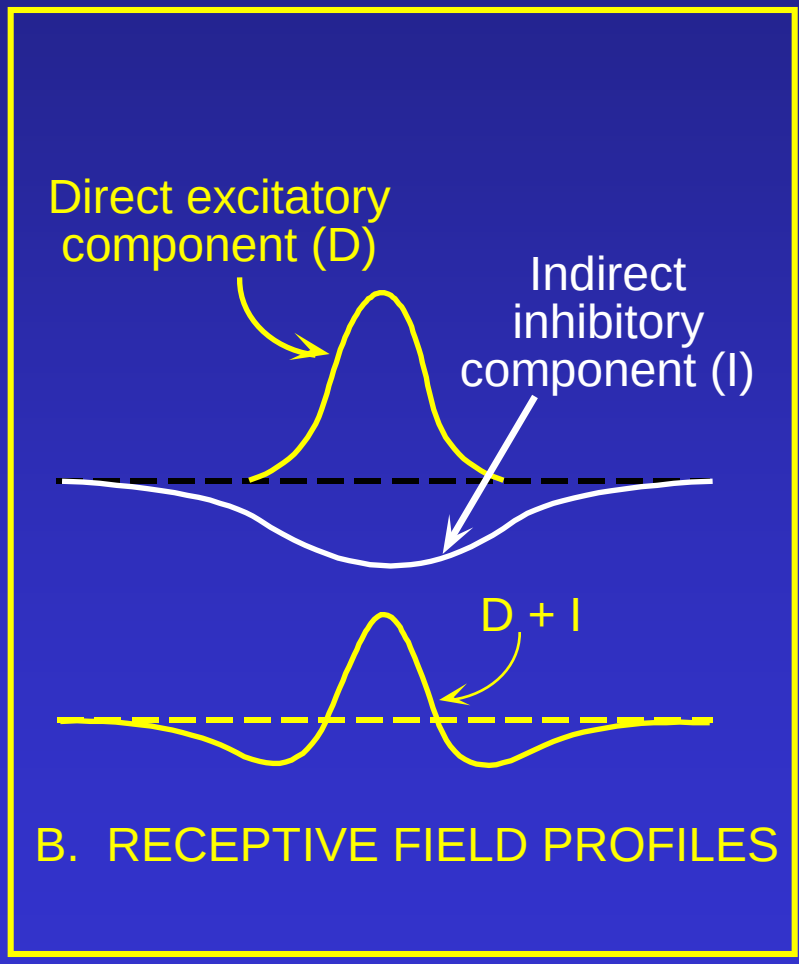
## Receptive field structure in bipolar cells



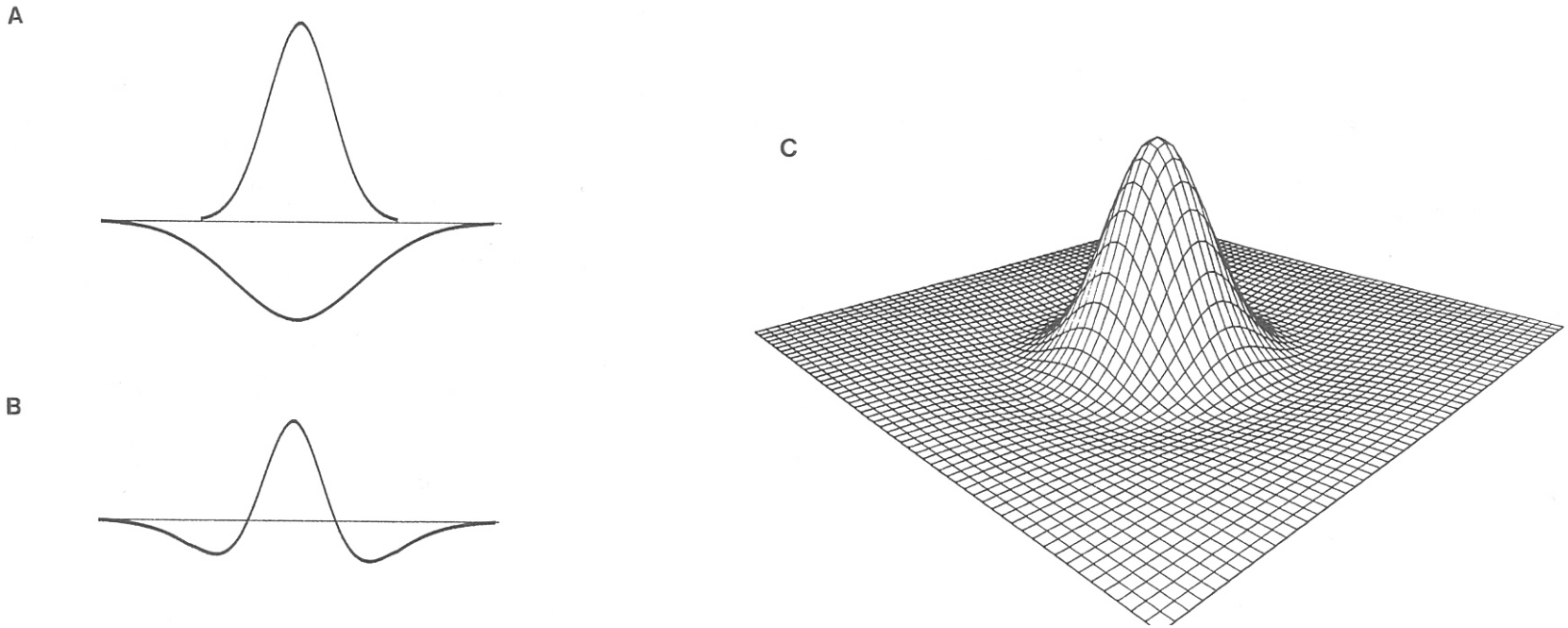
# Retinal Receptive Fields

## 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

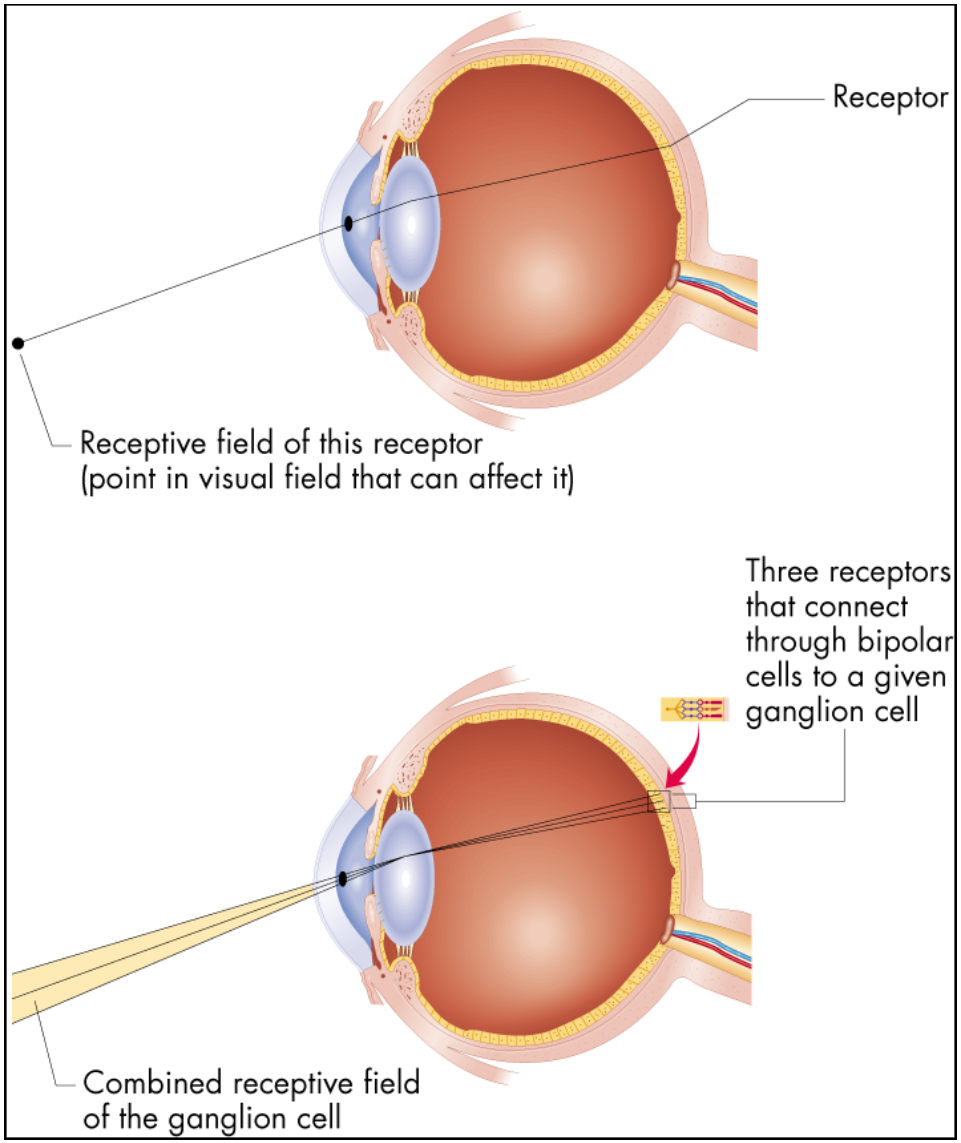
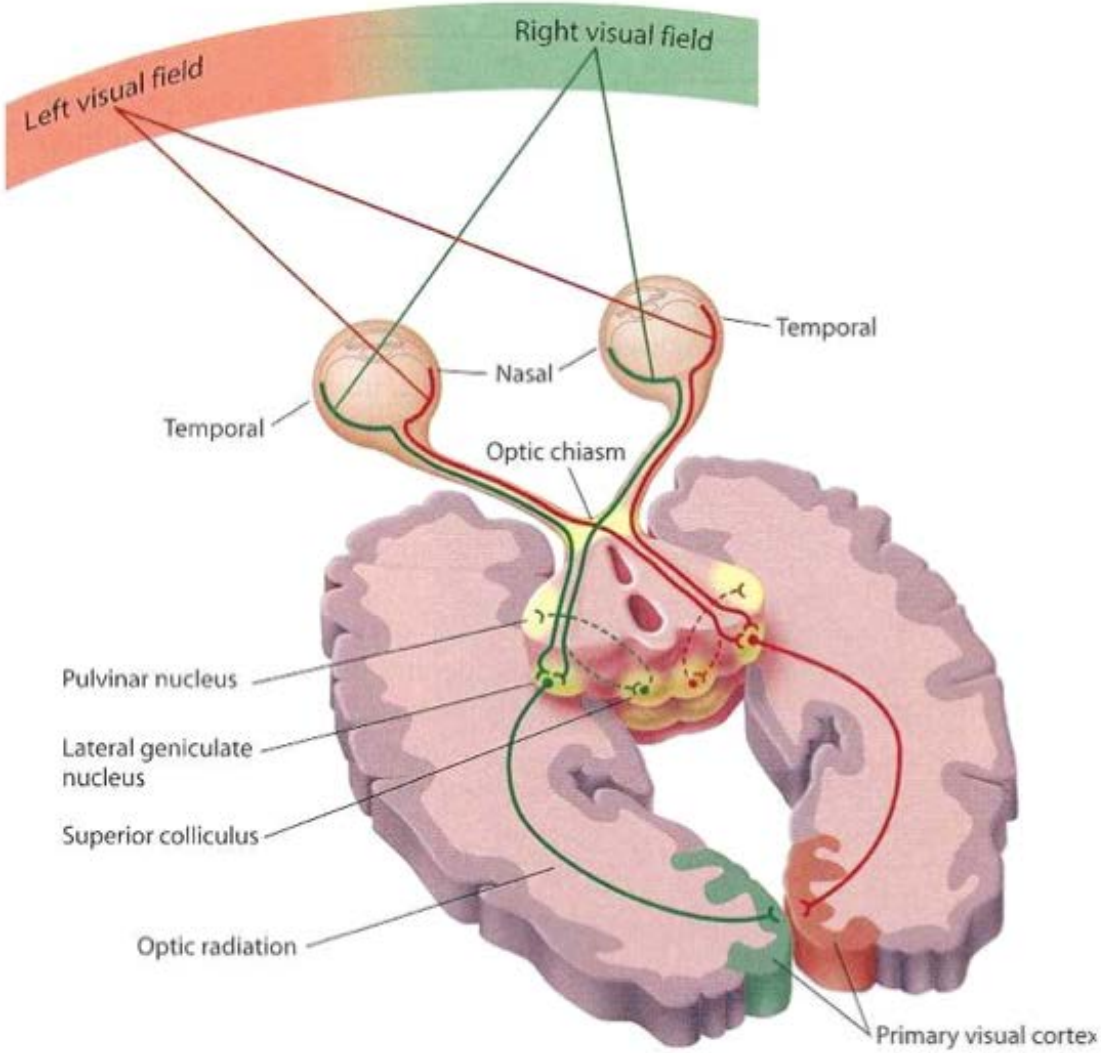



**Figure 6.16 Receptive fields**
The receptive field of a receptor is simply the area of the visual field from which light strikes that receptor. For any other cell in the visual system, the receptive field is determined by which receptors connect to the cell in question.

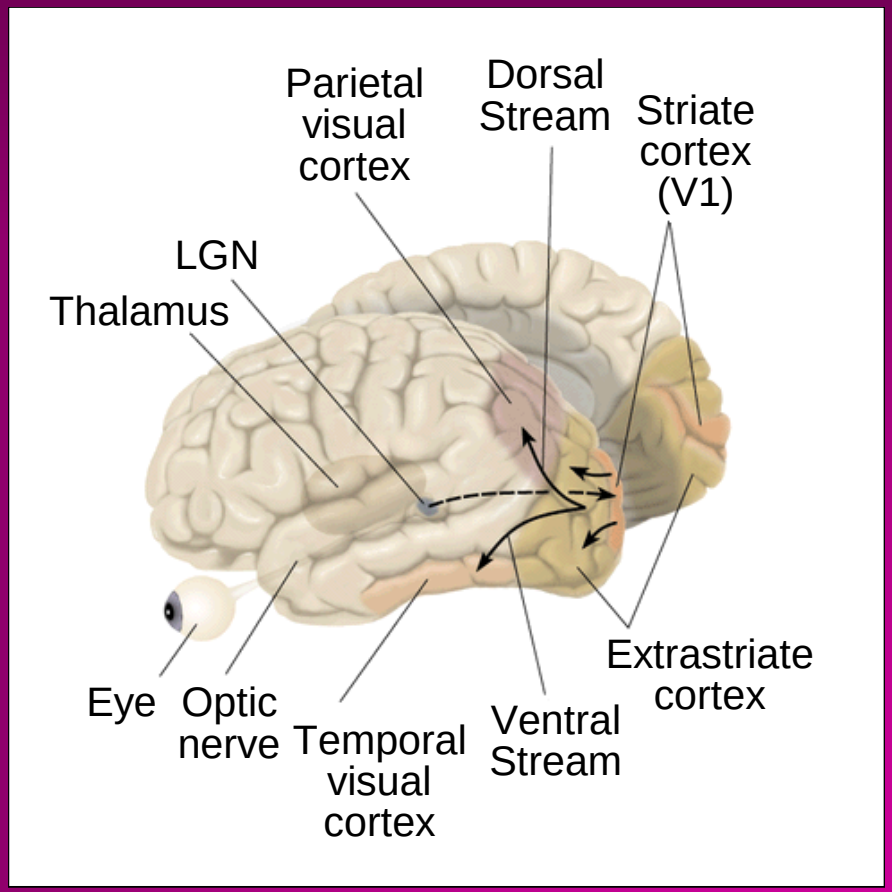
# Anatomy of Pathway to Visual Cortex



# Visual Cortex

## Cortical Area V1

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



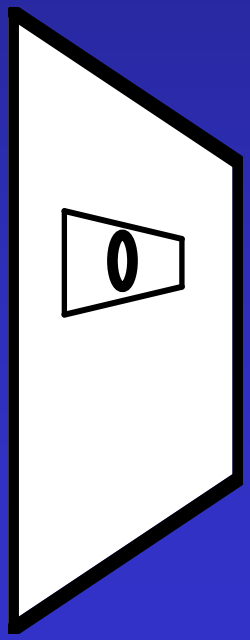
# Cortical Receptive Fields

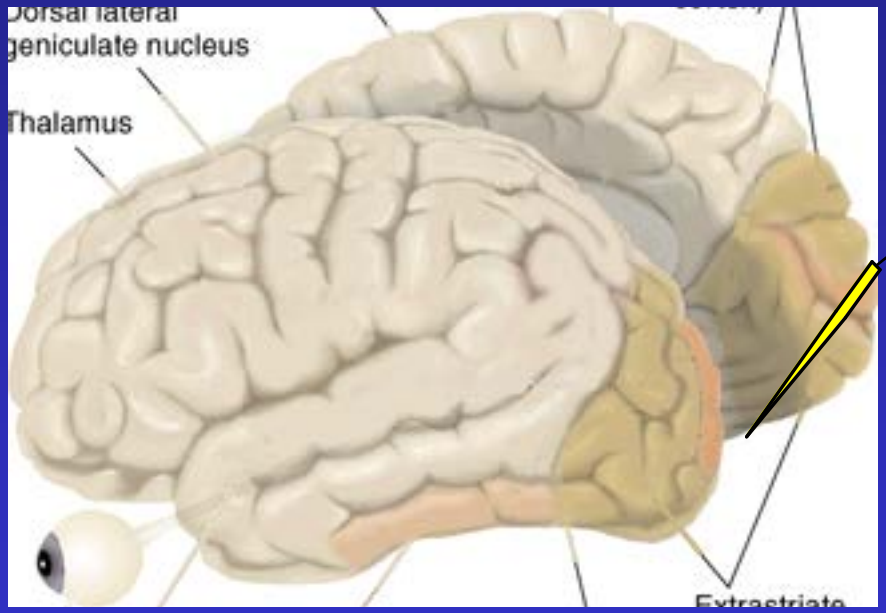
Single-cell recording from visual cortex

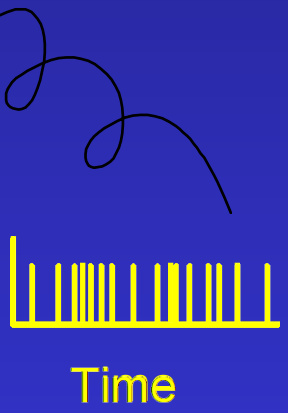
David Hubel & Thorston Wiesel

# Cortical Receptive Fields

Single-cell recording from visual cortex






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

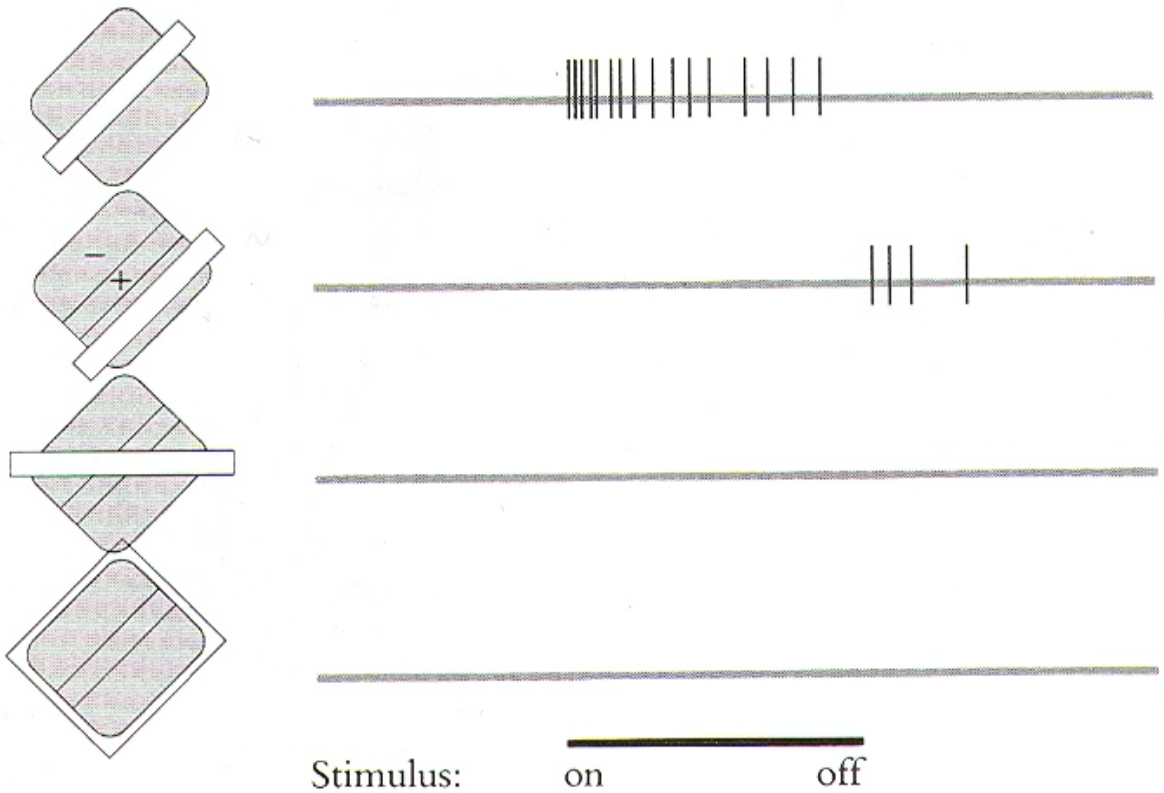
# Cortical Receptive Fields

**Three classes of cells in V1**

Simple cells

Complex cells

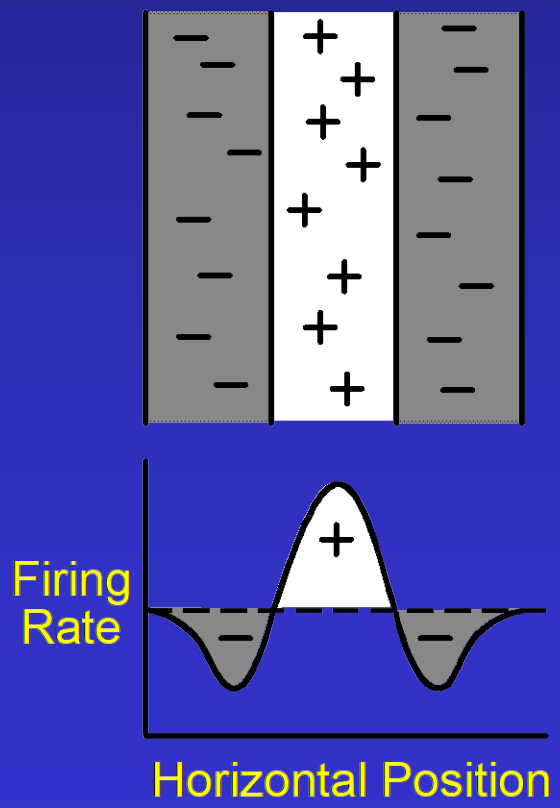
Hypercomplex cells


−
+
Stimulus:
on
off

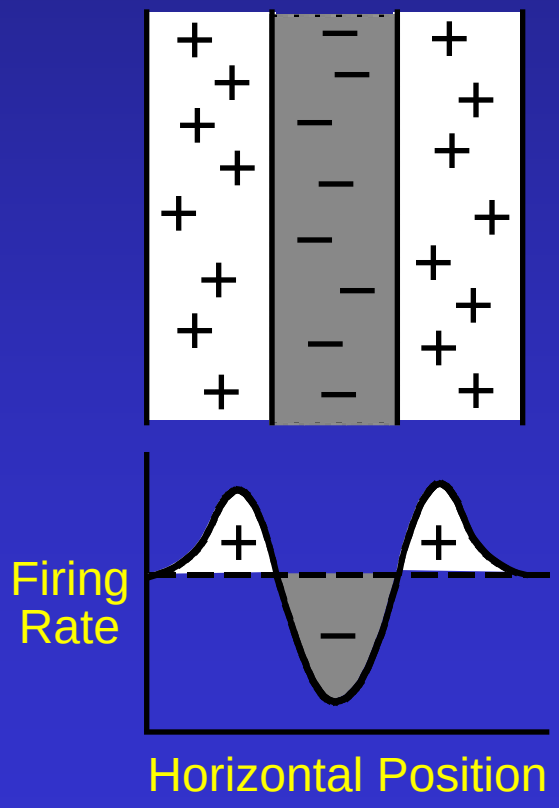
# Cortical Receptive Fields

## Simple Cells: “Line Detectors”

A. Light Line Detector



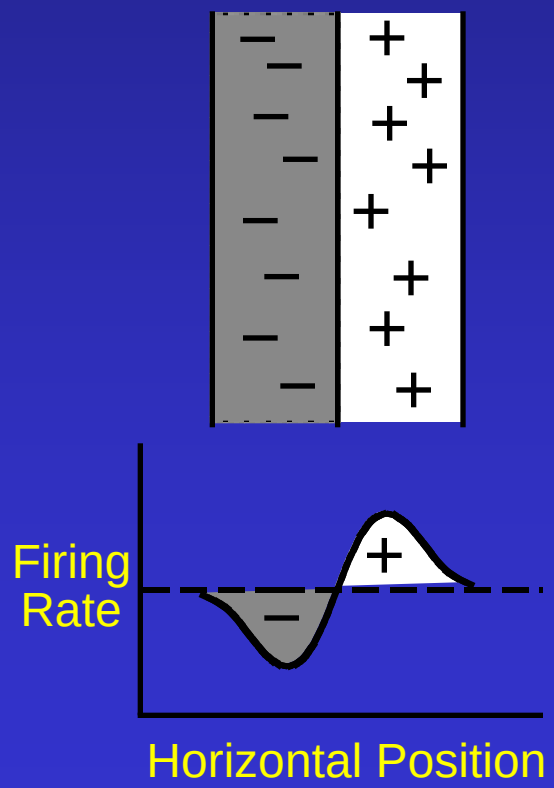

B. Dark Line Detector



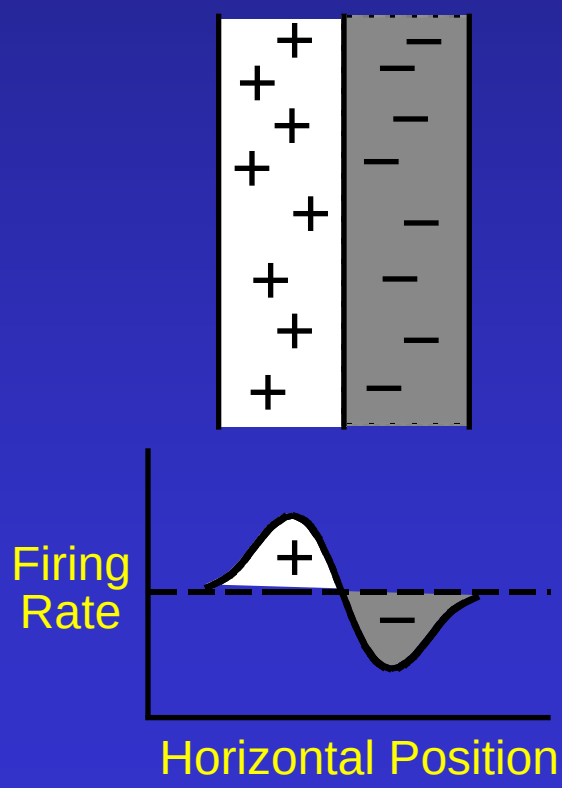
# Cortical Receptive Fields

## Simple Cells: “Edge Detectors”

C. Dark-to-light Edge Detector

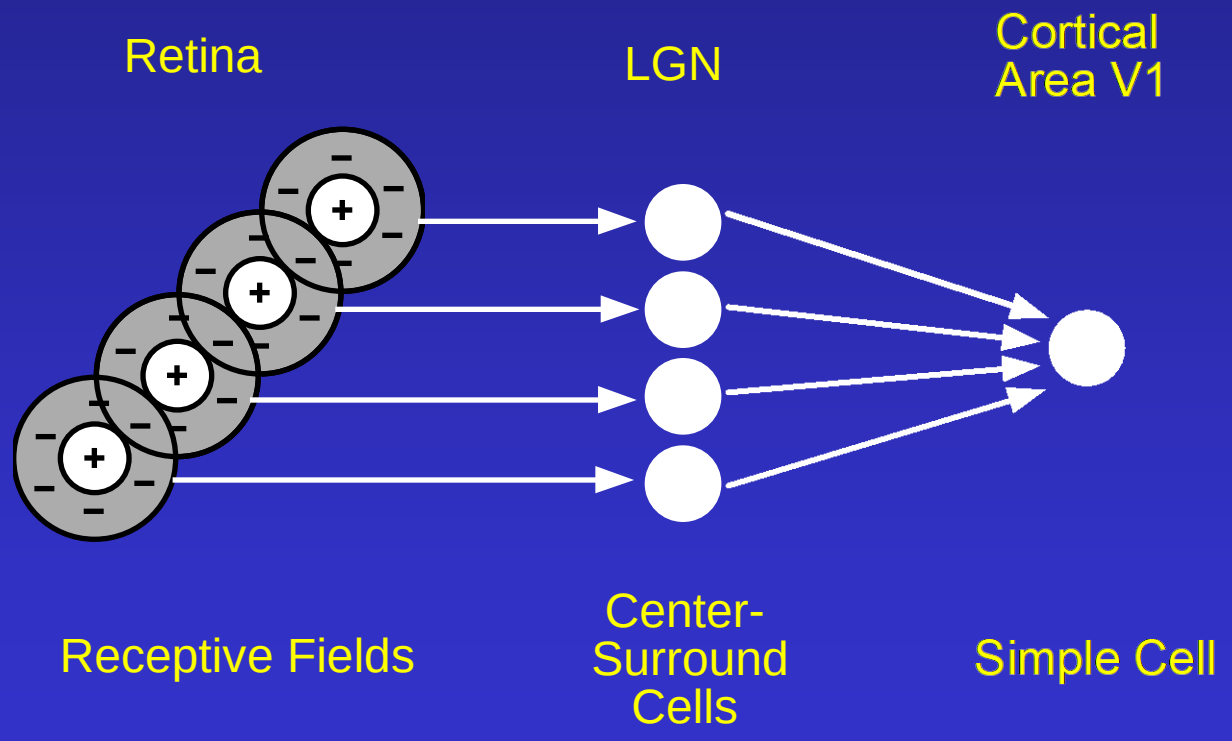



D. Light-to-dark Edge Detector

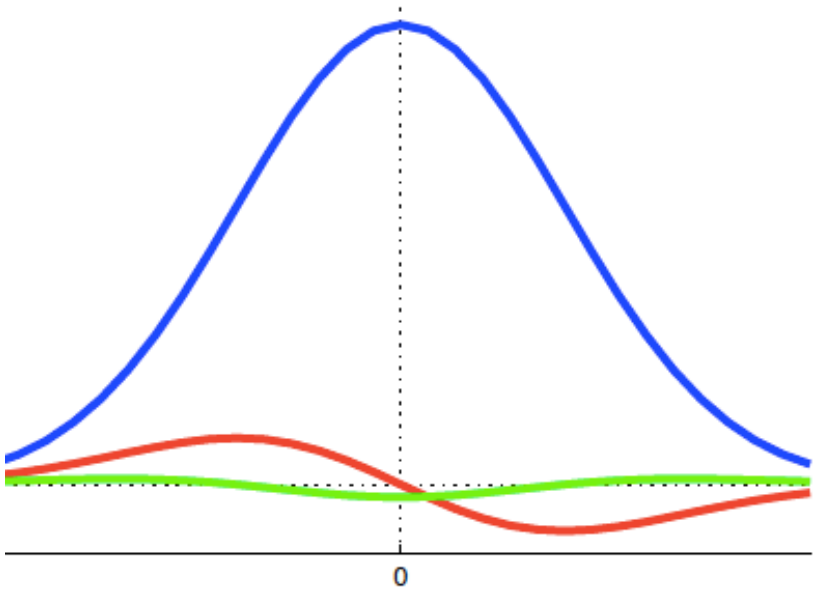


# Cortical Receptive Fields

## 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.

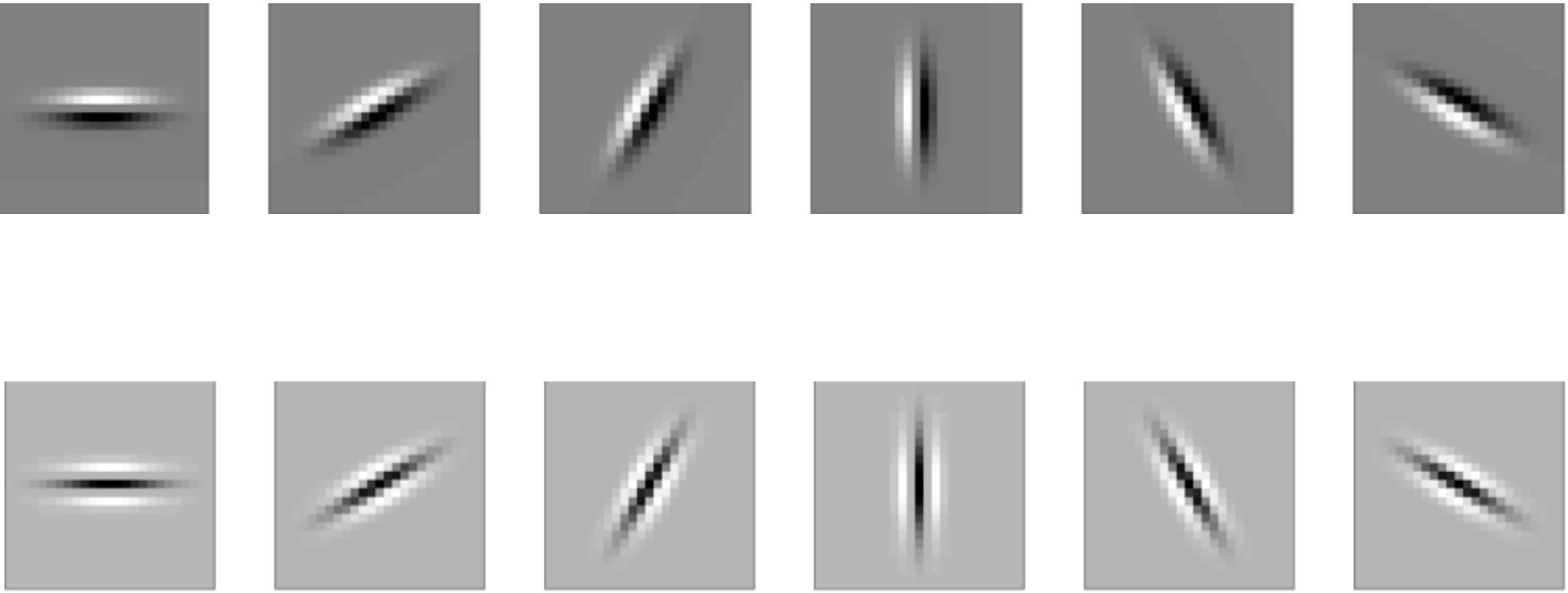
$$Rot_\theta f_1 = G'_{\sigma_1}(u)G_{\sigma_2}(v) \tag{10.6}$$
$$Rot_\theta f_2 = G''_{\sigma_1}(u)G_{\sigma_2}(v) \tag{10.7}$$

where we set

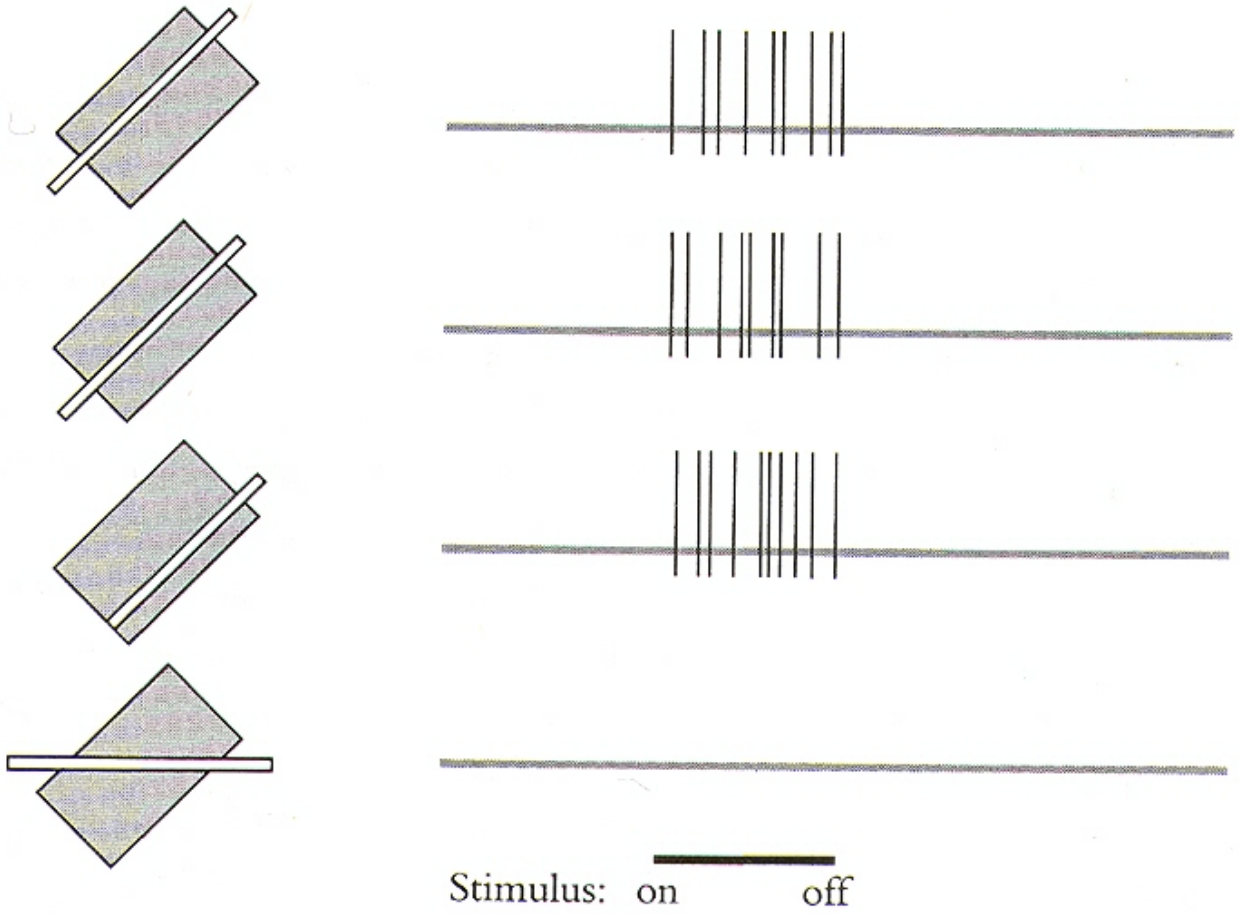
$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

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

# Oriented Gaussian First and Second Derivatives

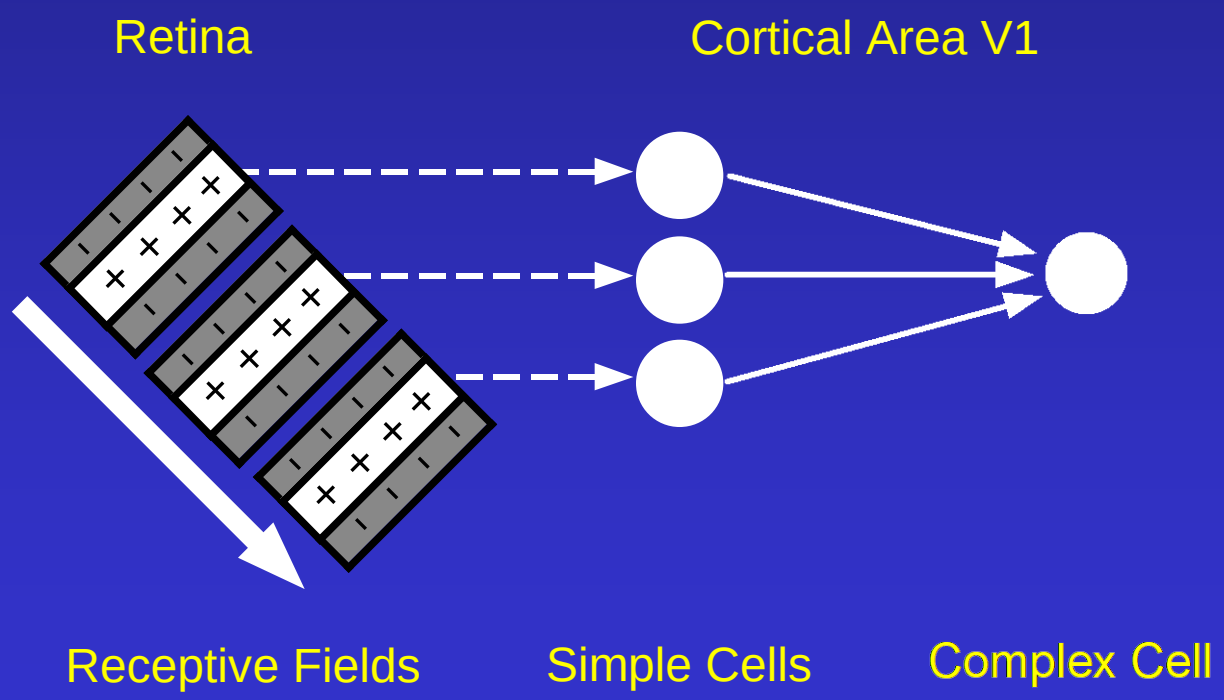


# Receptive fields of complex cells

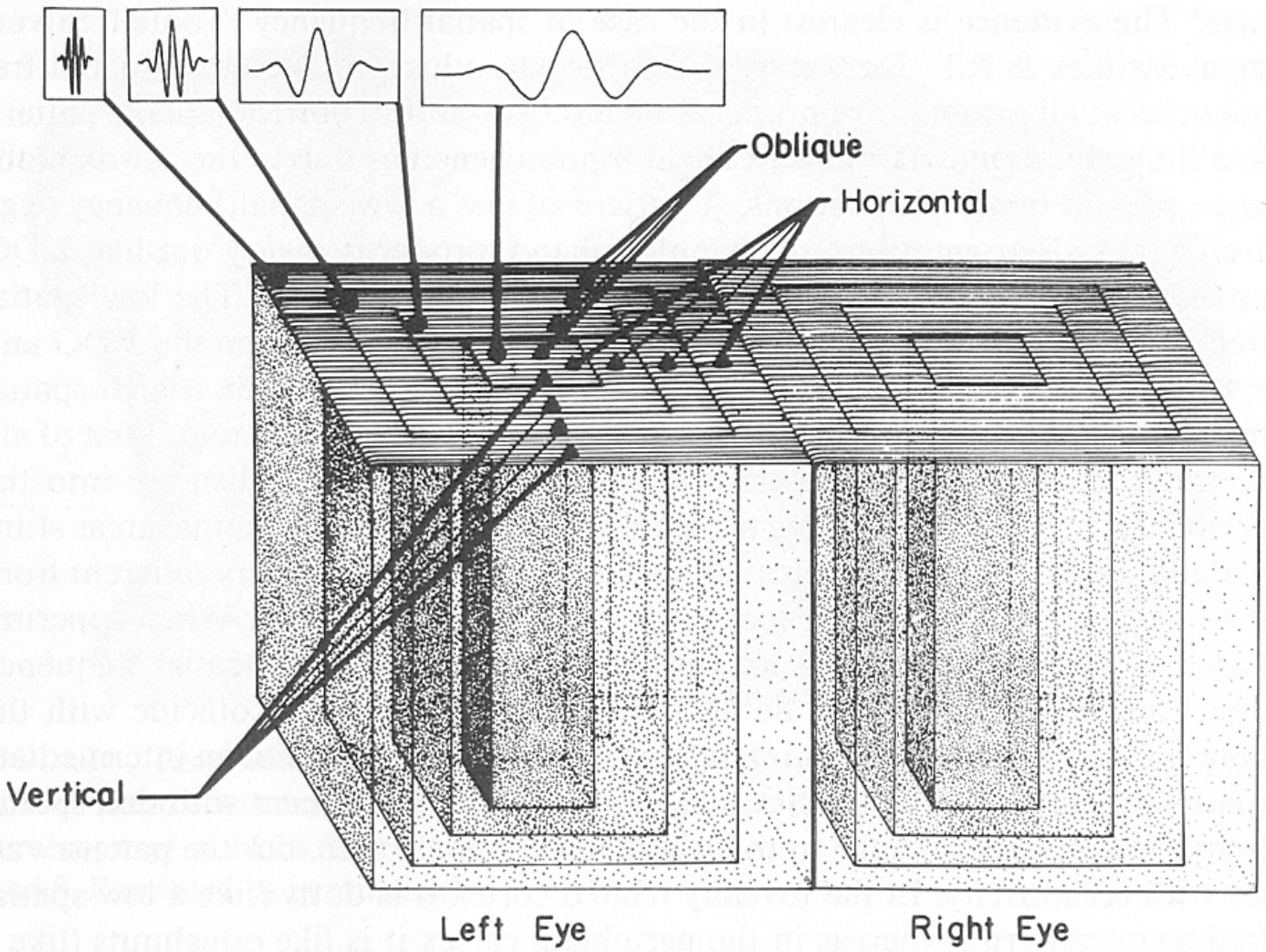


# Cortical Receptive Fields

## Constructing a Complex Cell



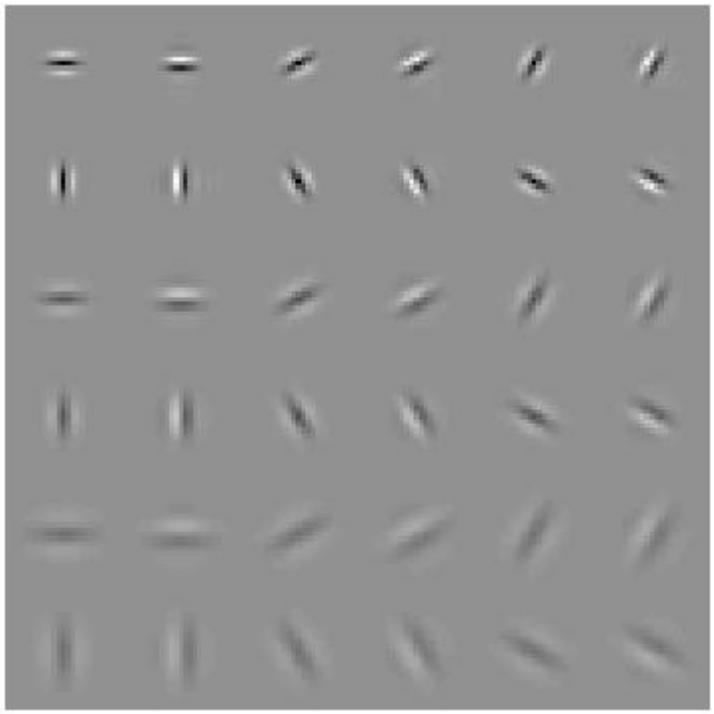
# Hypercolumns in visual cortex



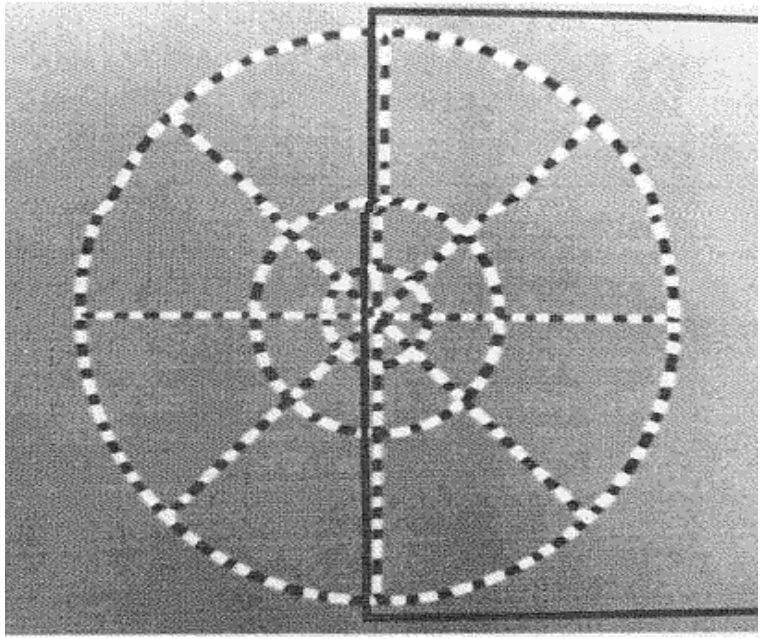

Model of Striate Module in Monkeys

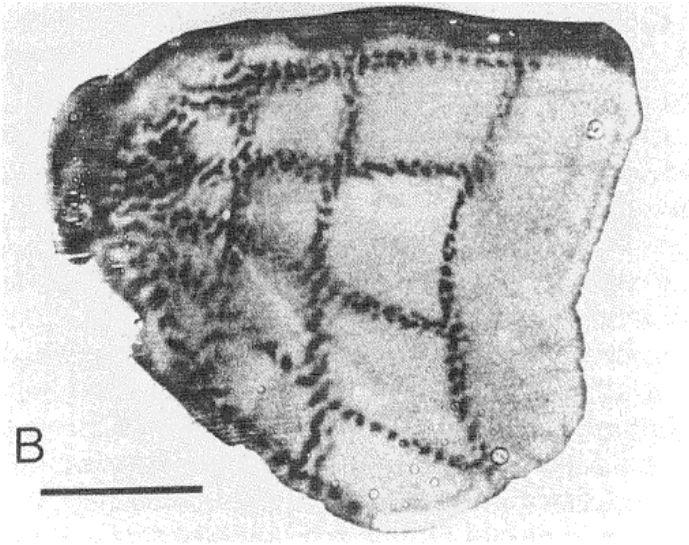
# Modeling hypercolumns

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



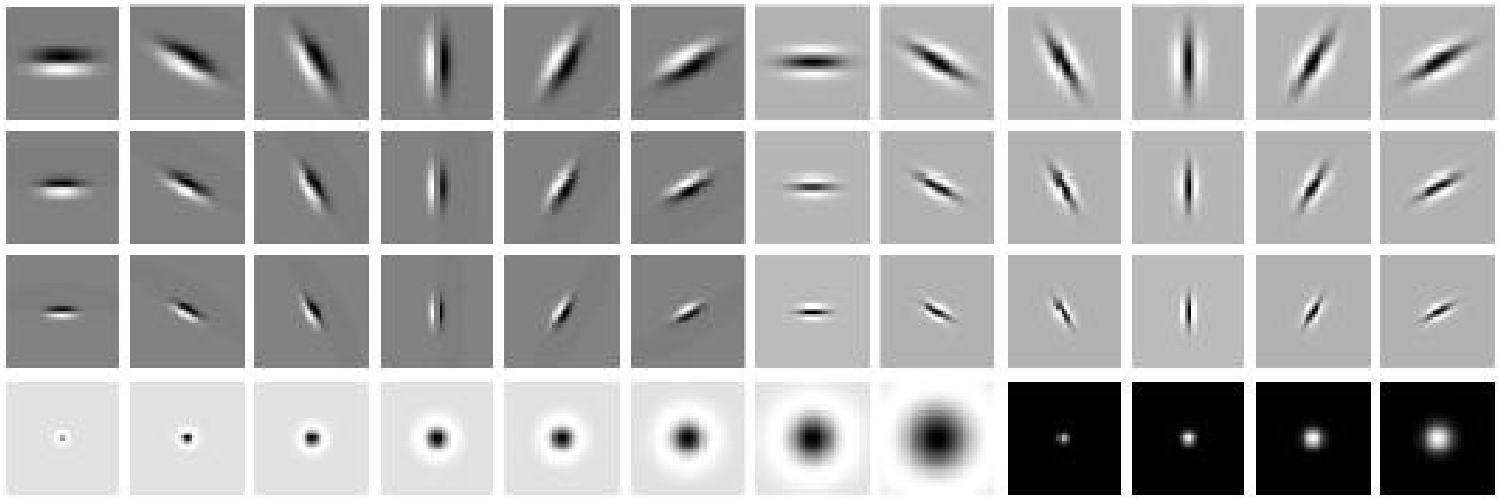
# Mapping from Retina to V1





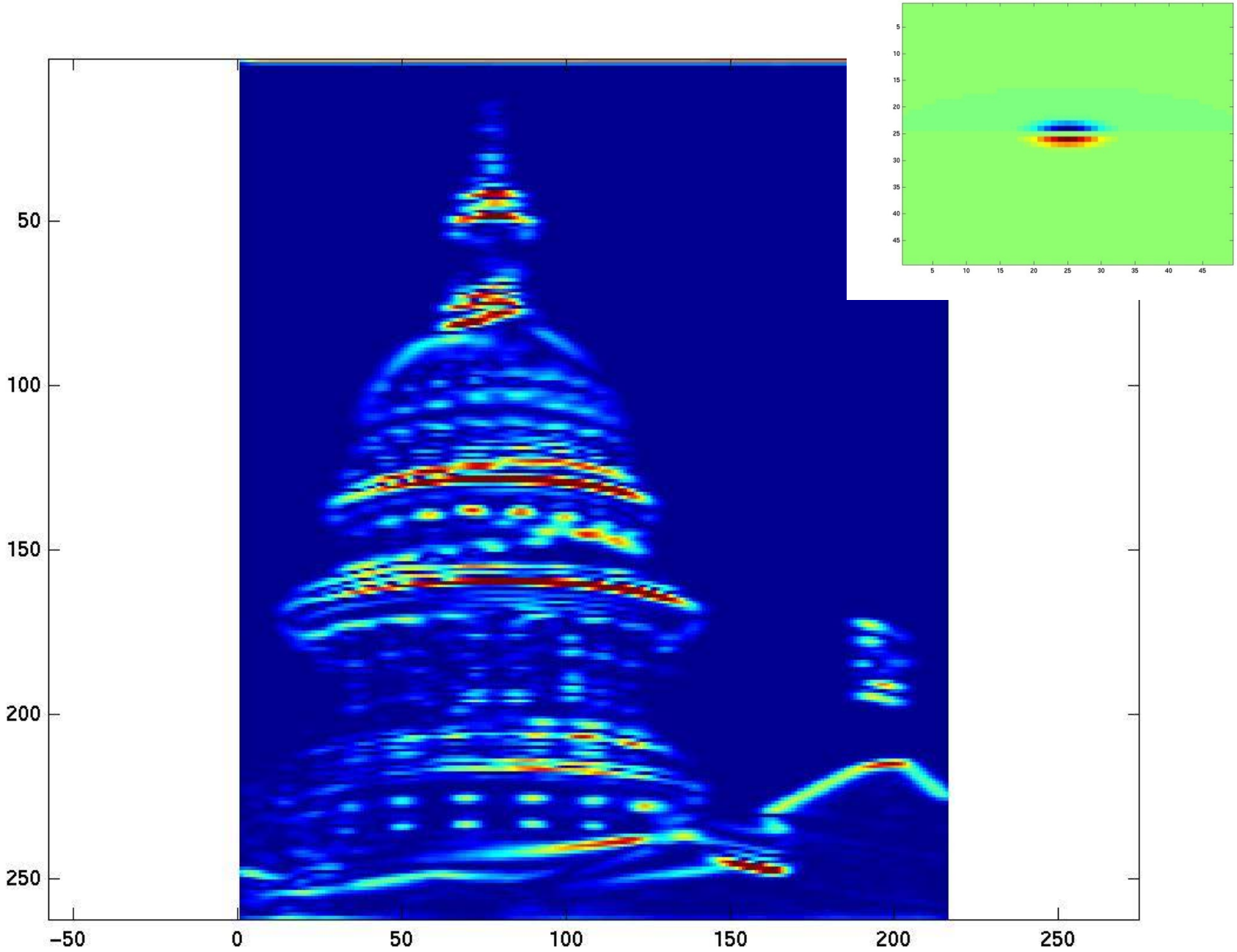
# Overcomplete representation: filter banks

LM Filter Bank

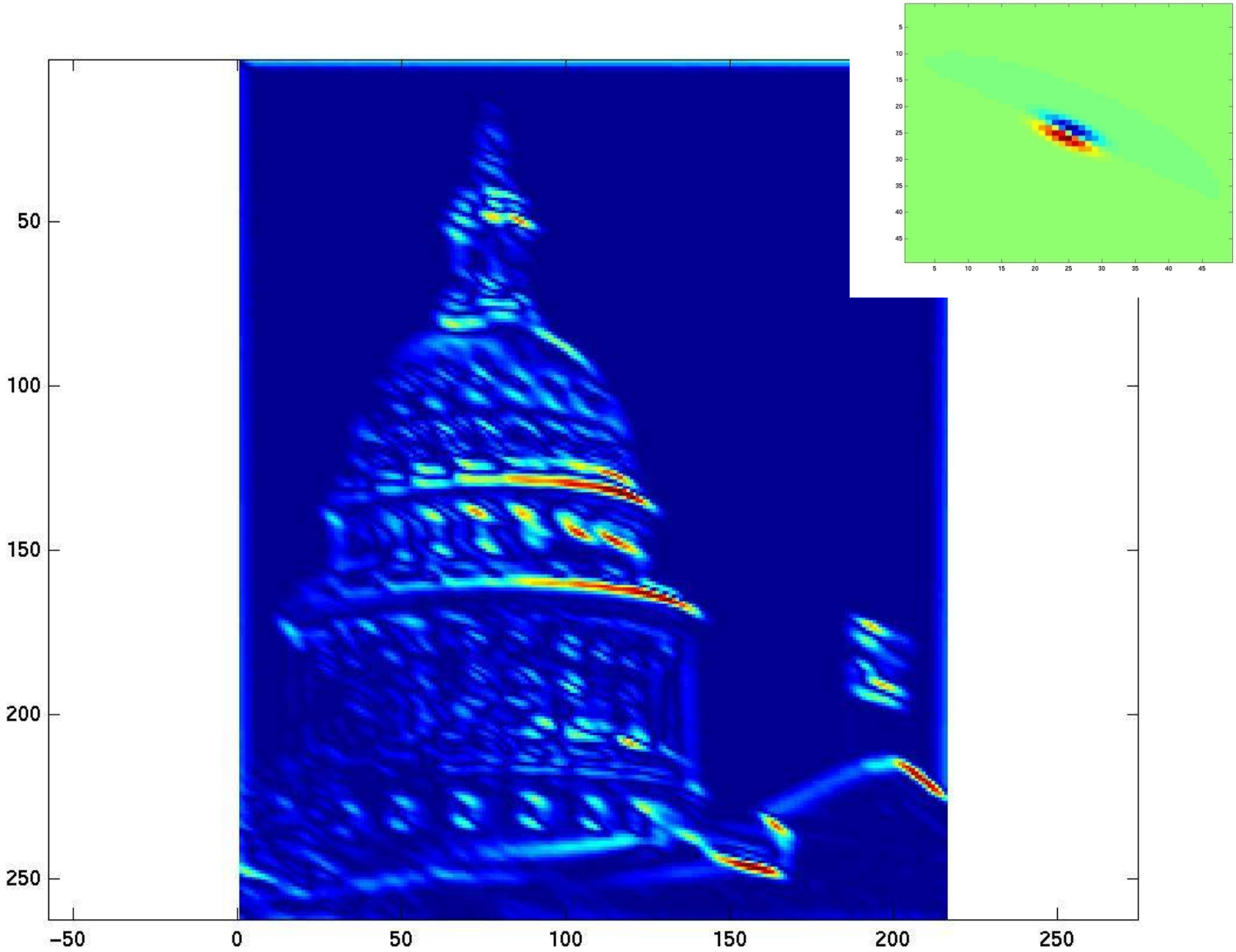


Code for filter banks: *www.robots.ox.ac.uk/~vgg/research/texclass/filters.html*

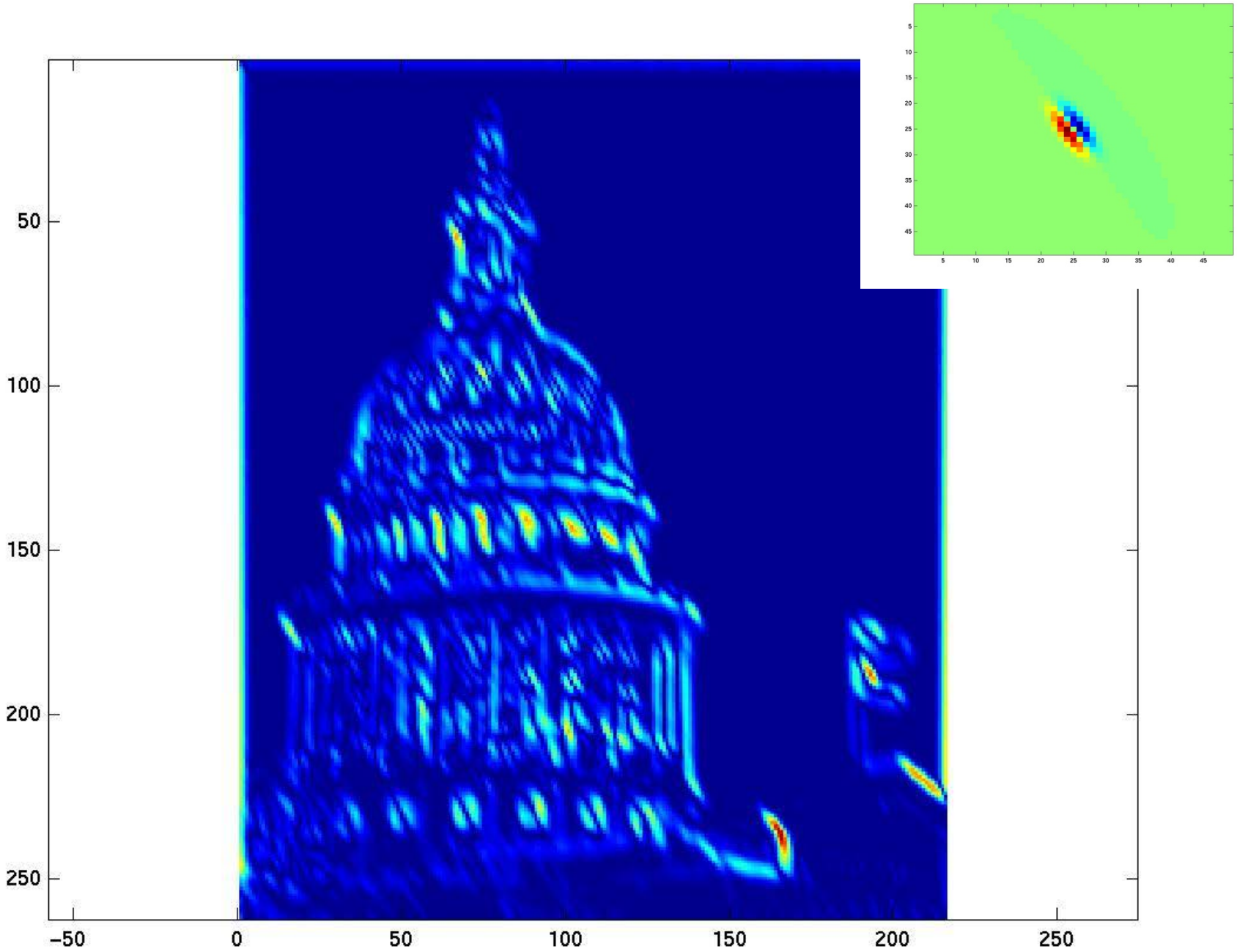
Image from http://www.texasexplorer.com/austincap2.jpg



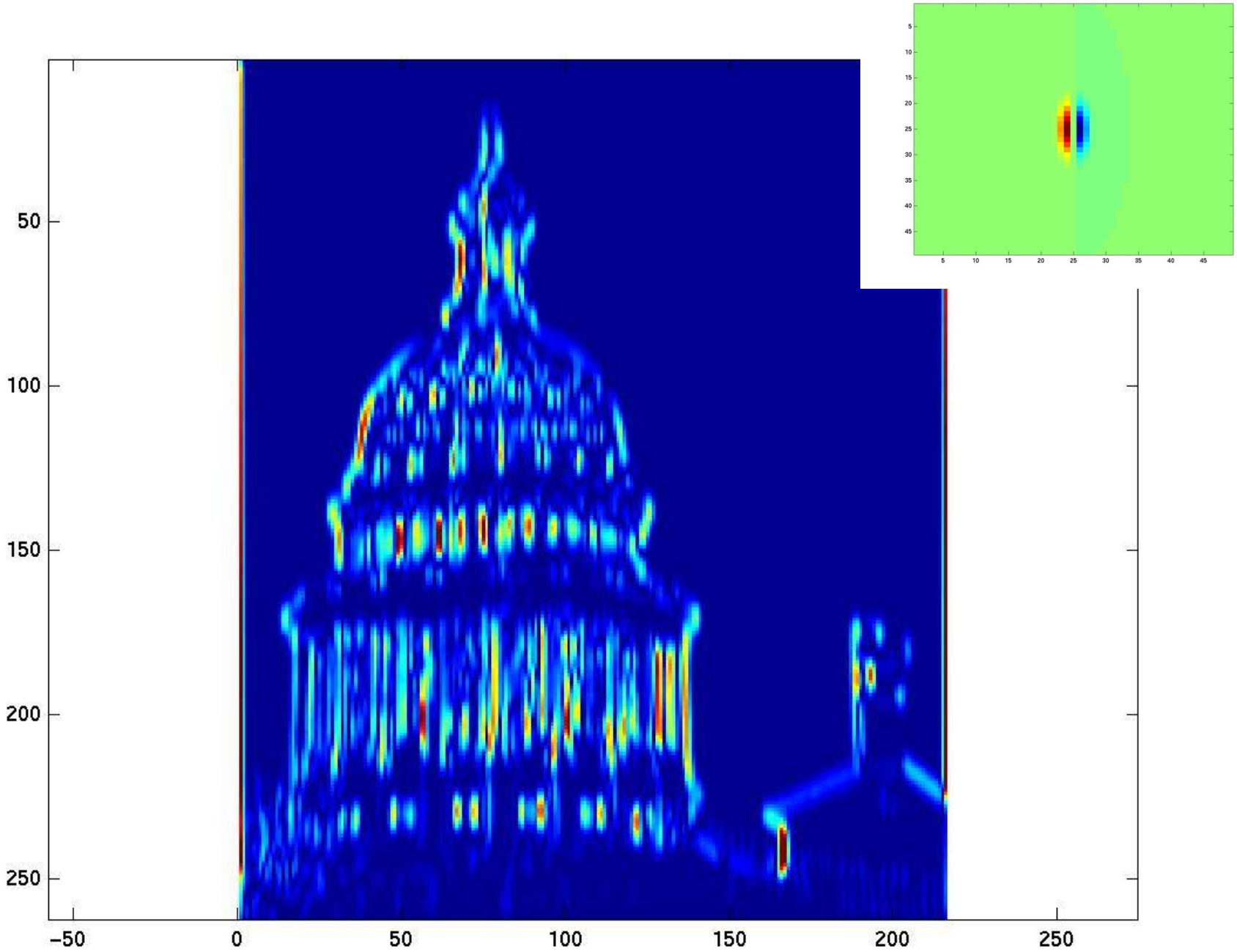

Kristen Grauman



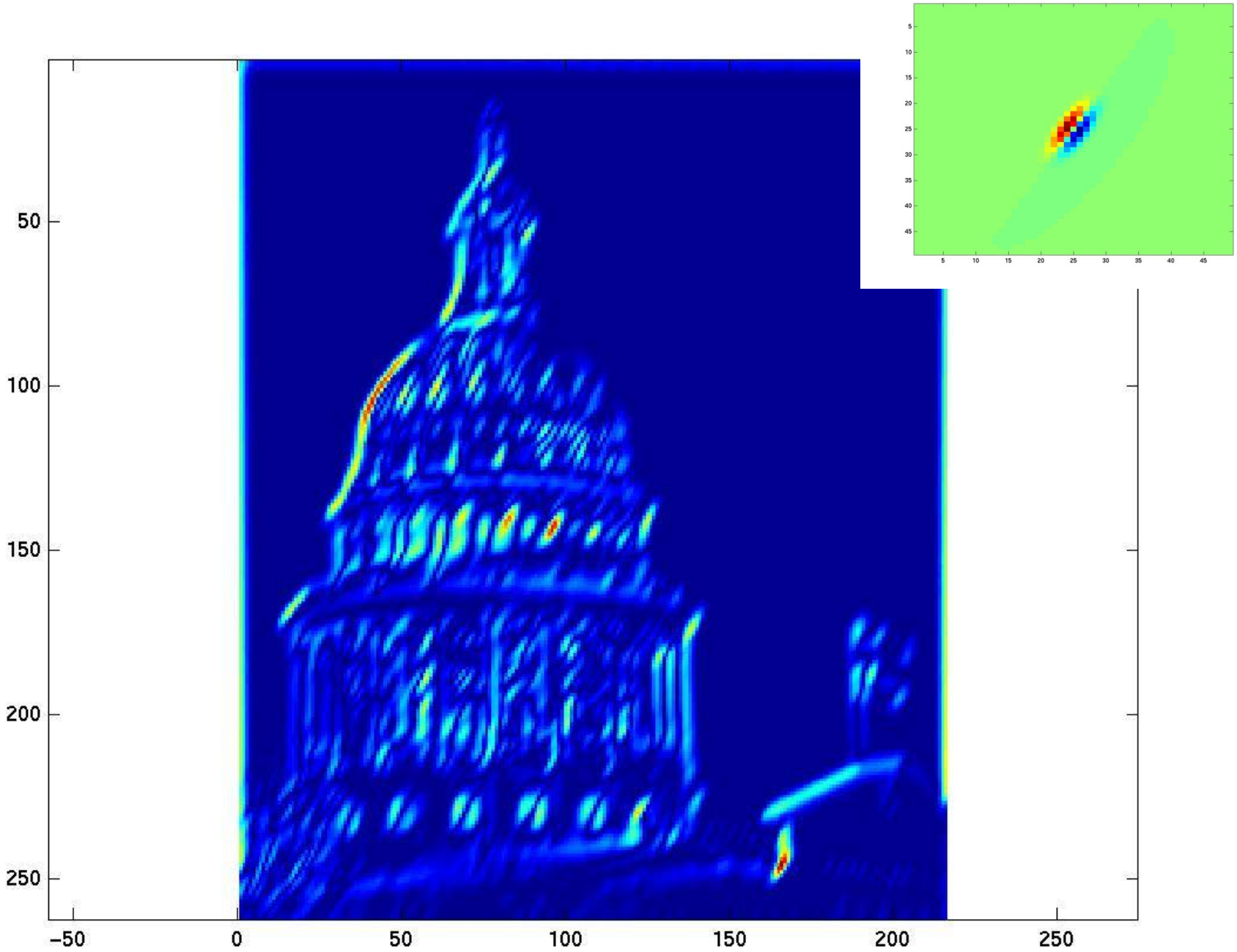

Kristen Grauman



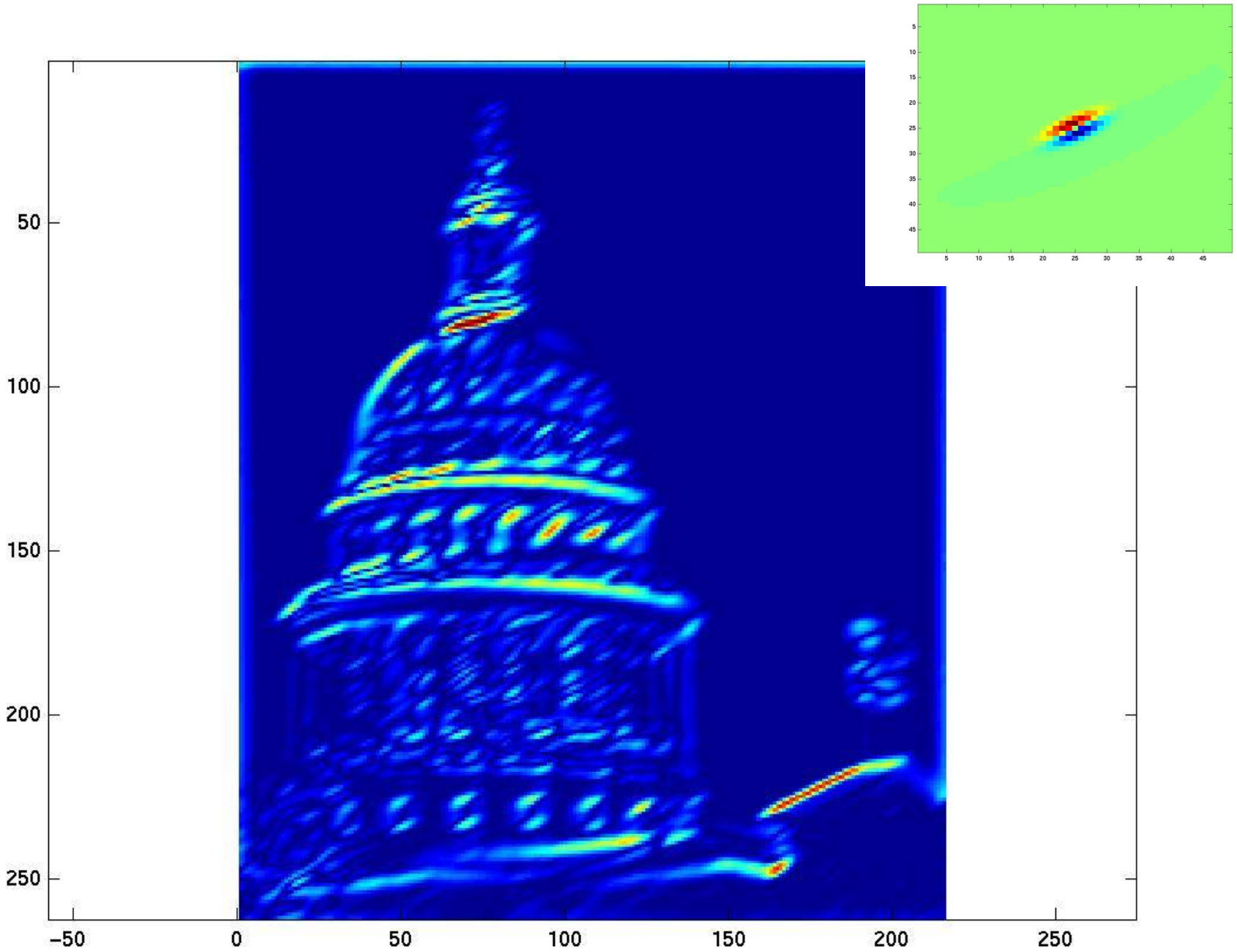

Kristen Grauman



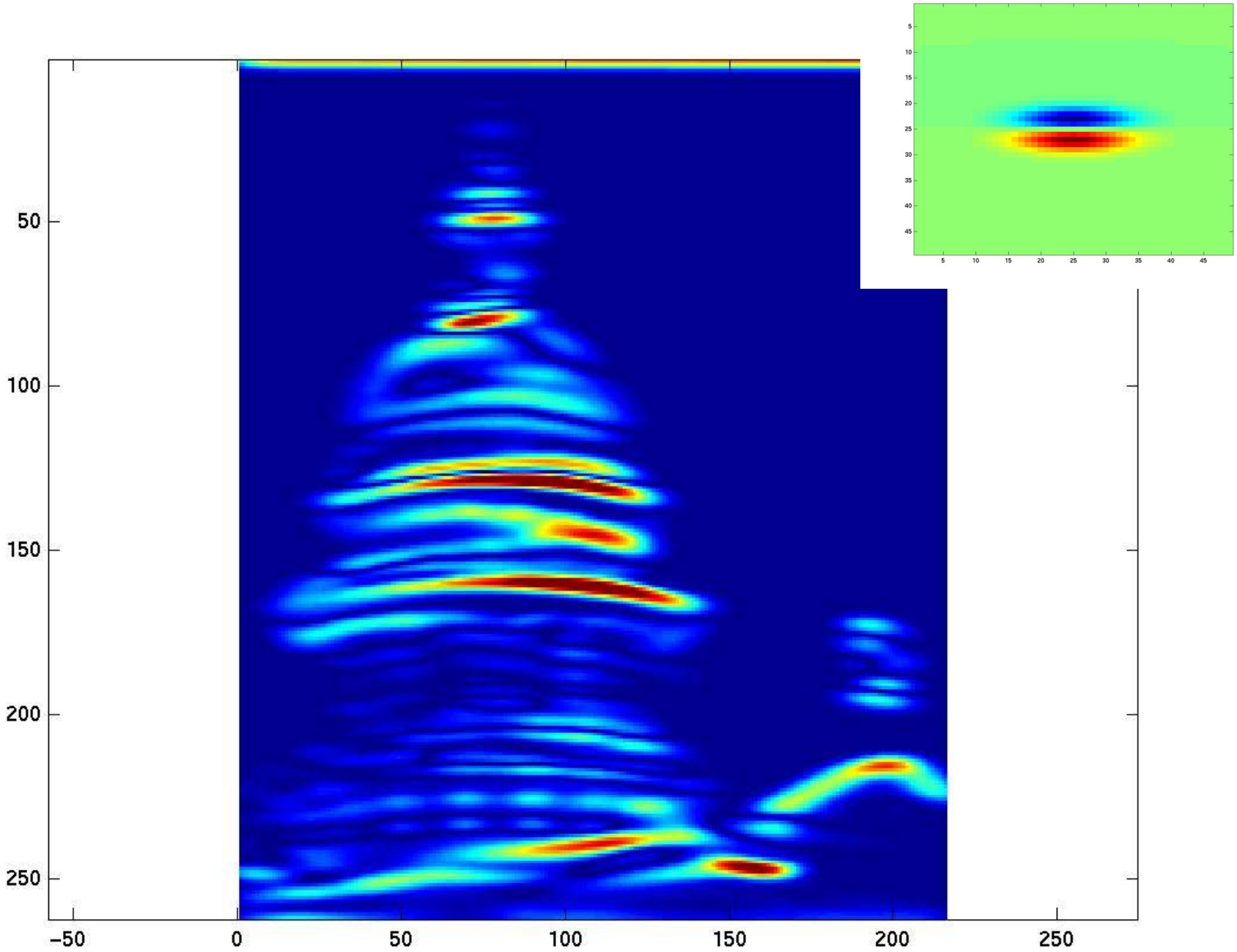

Kristen Grauman



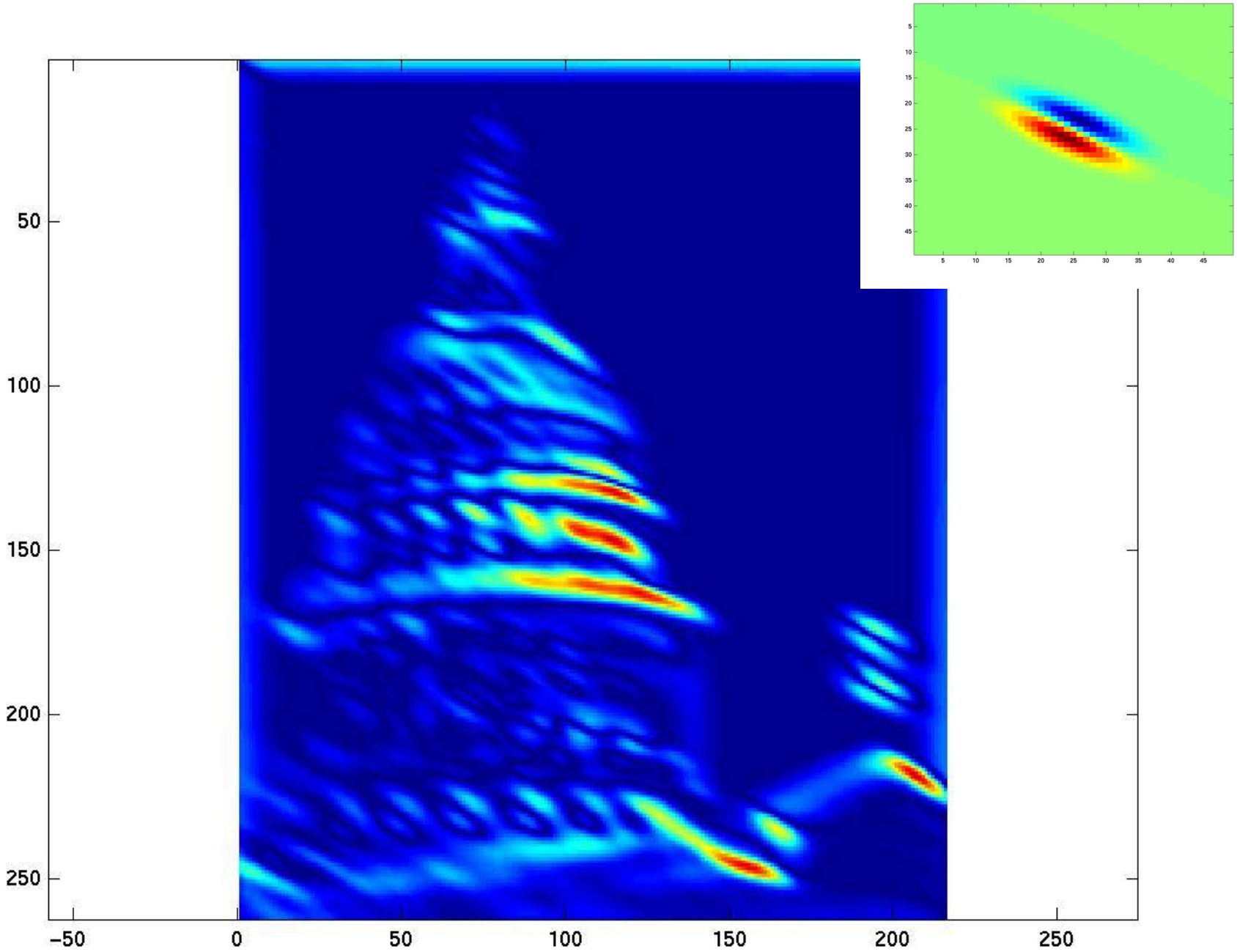

Kristen Grauman



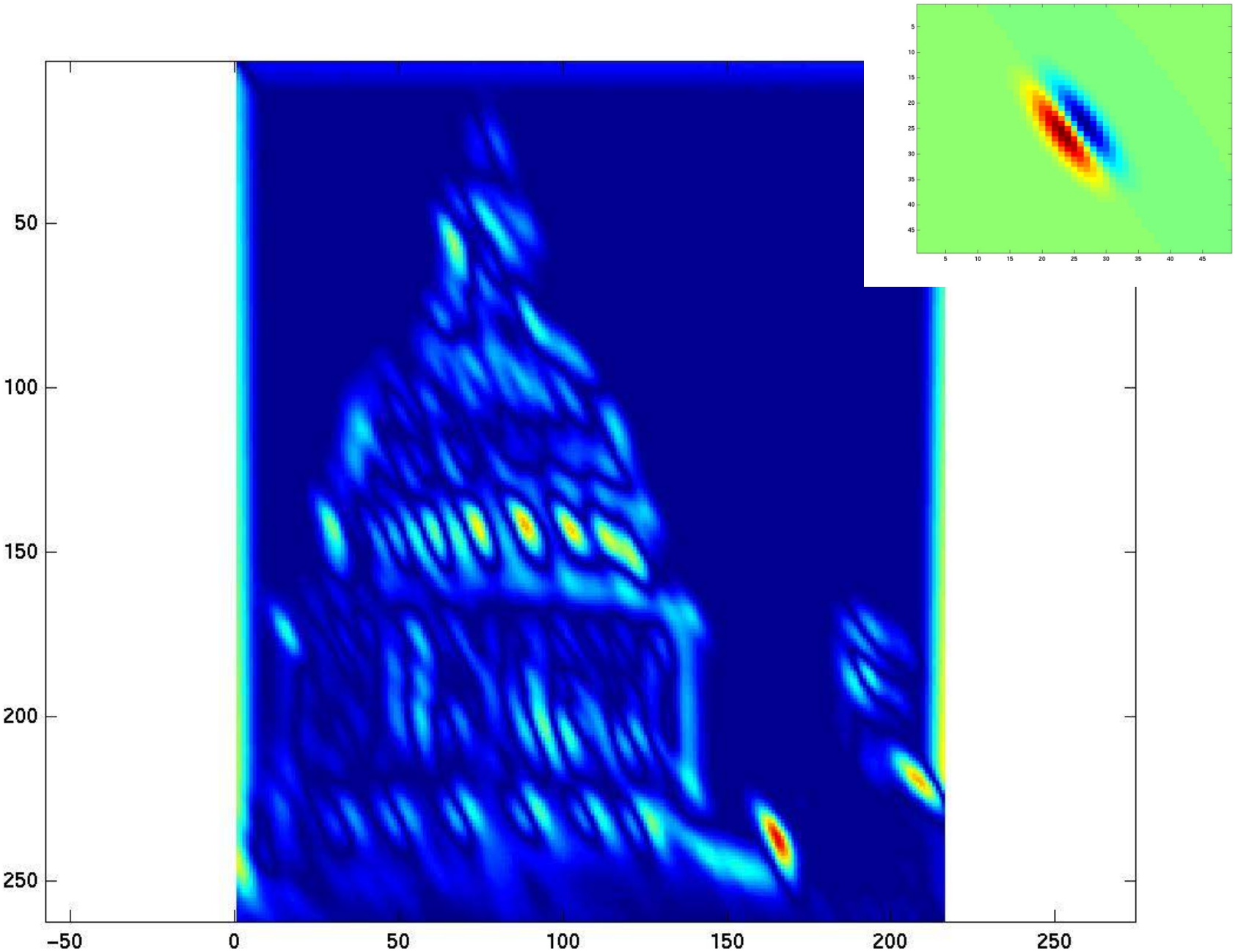

Kristen Grauman



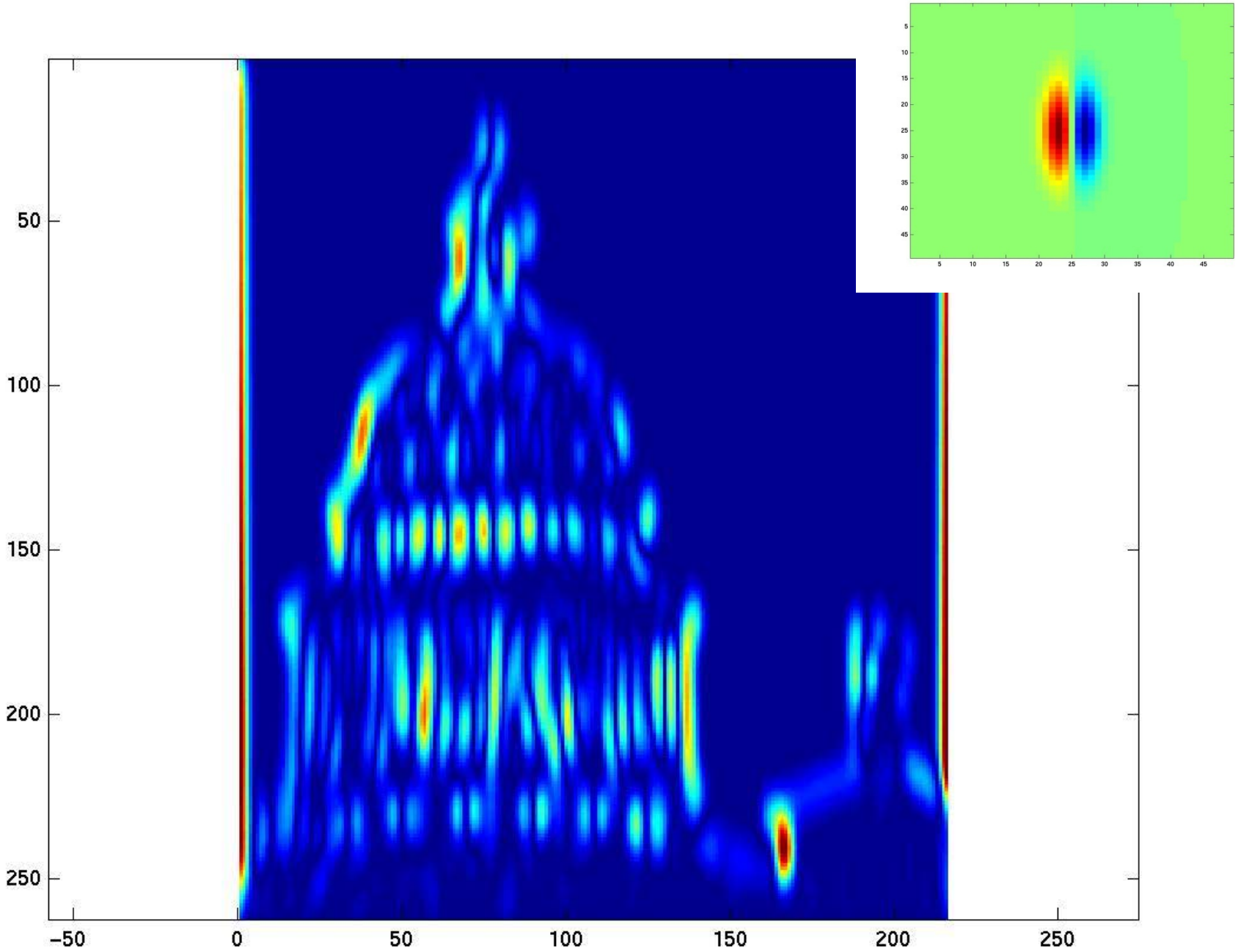

Kristen Grauman



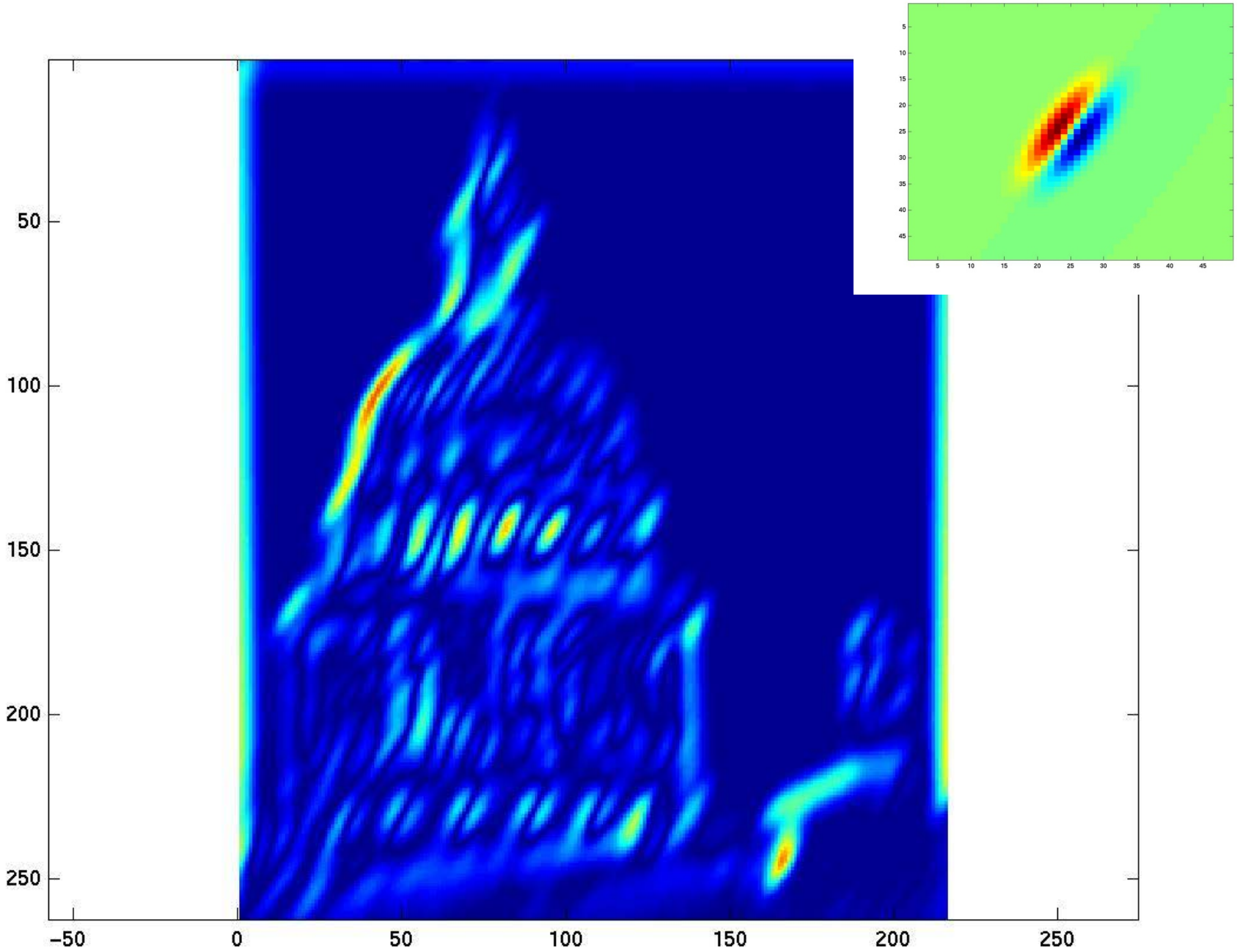

Kristen Grauman



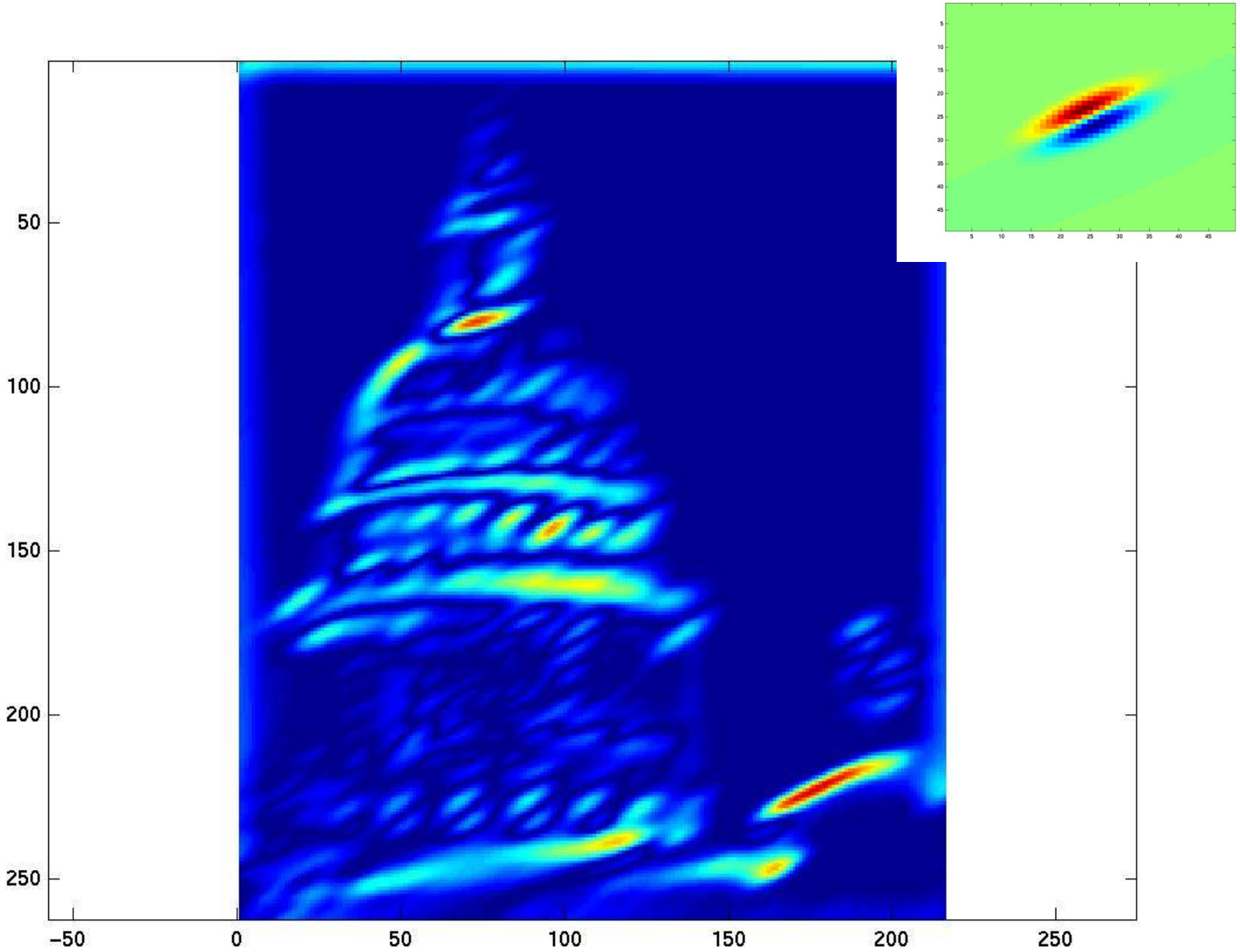

Kristen Grauman



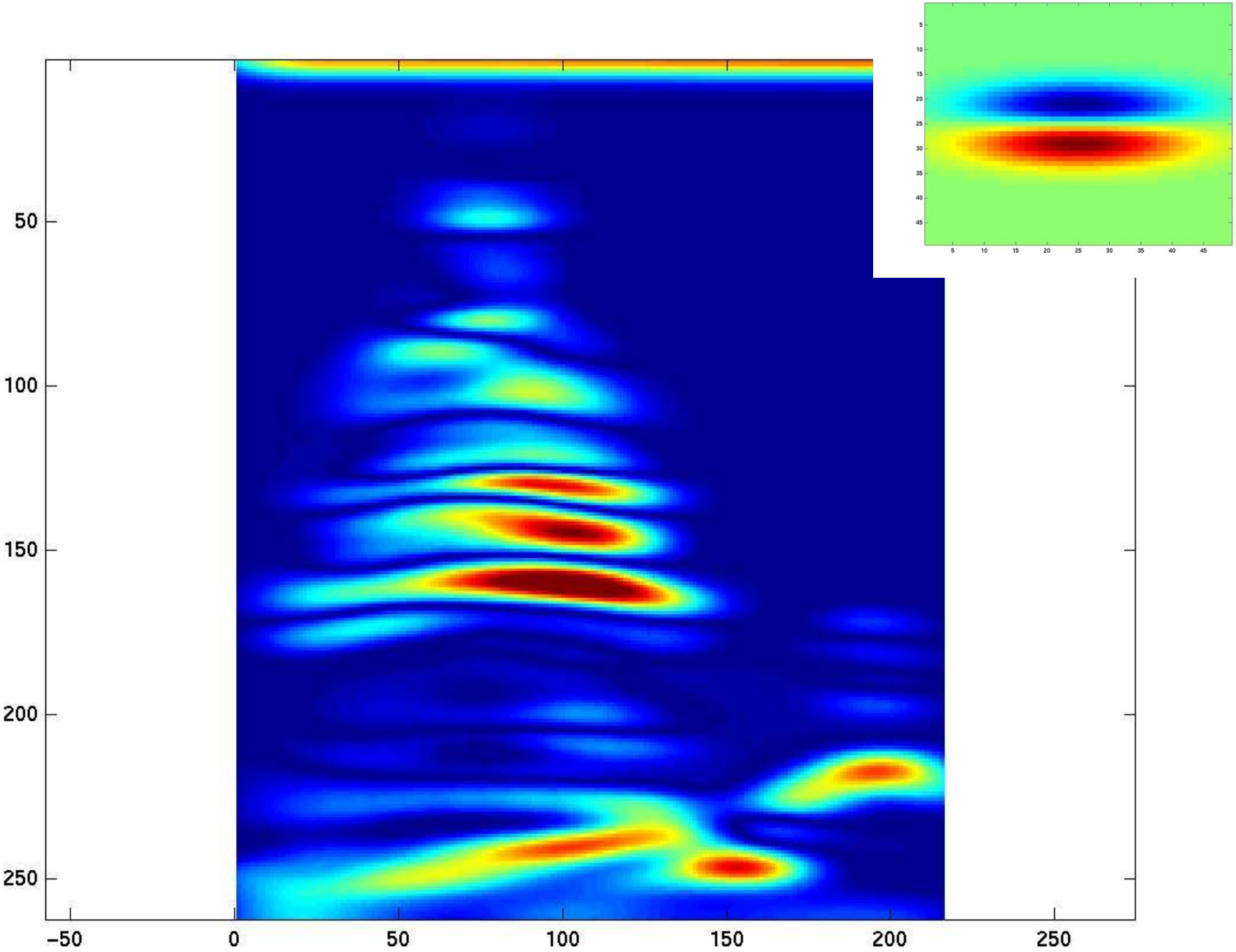

Kristen Grauman



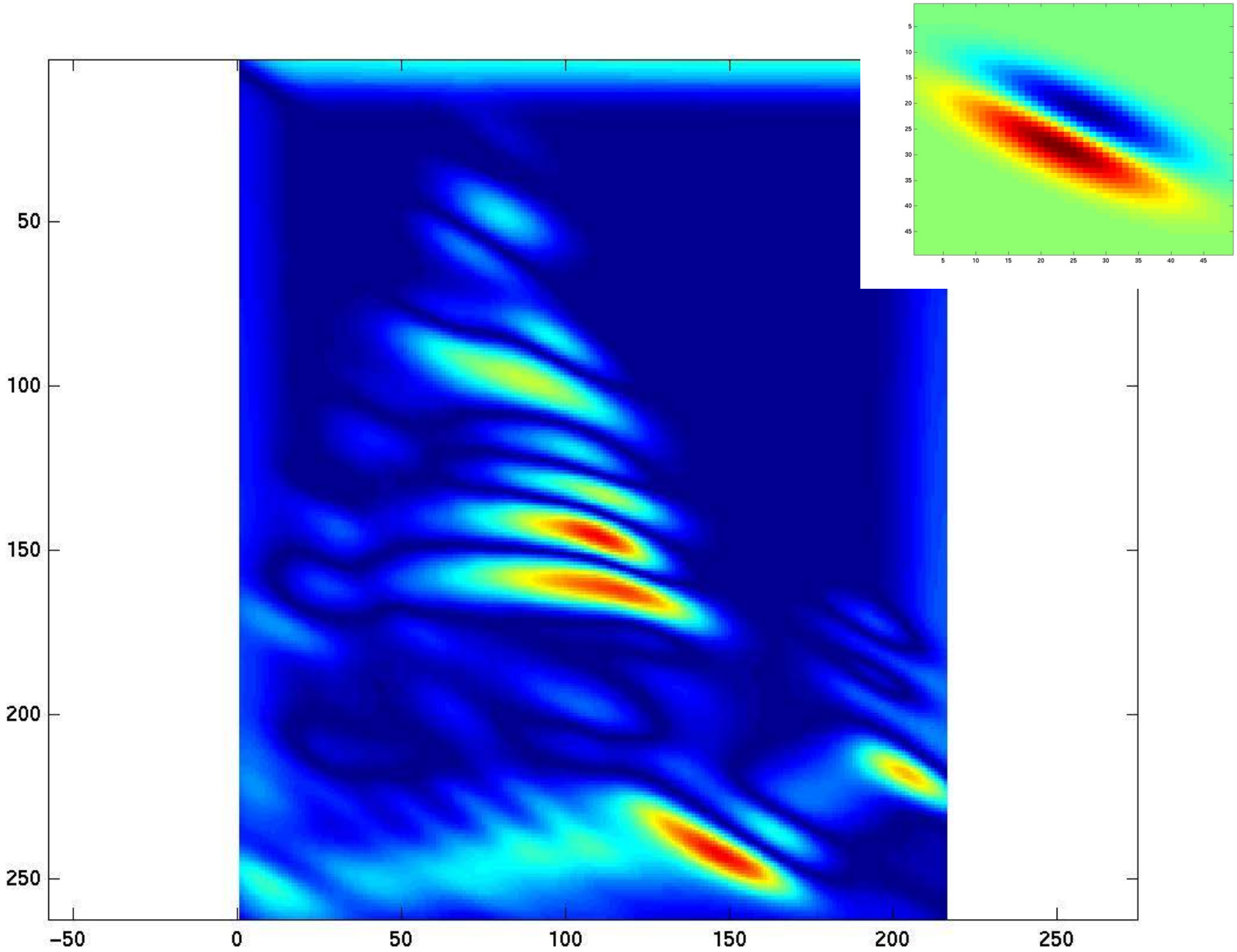

Kristen Grauman



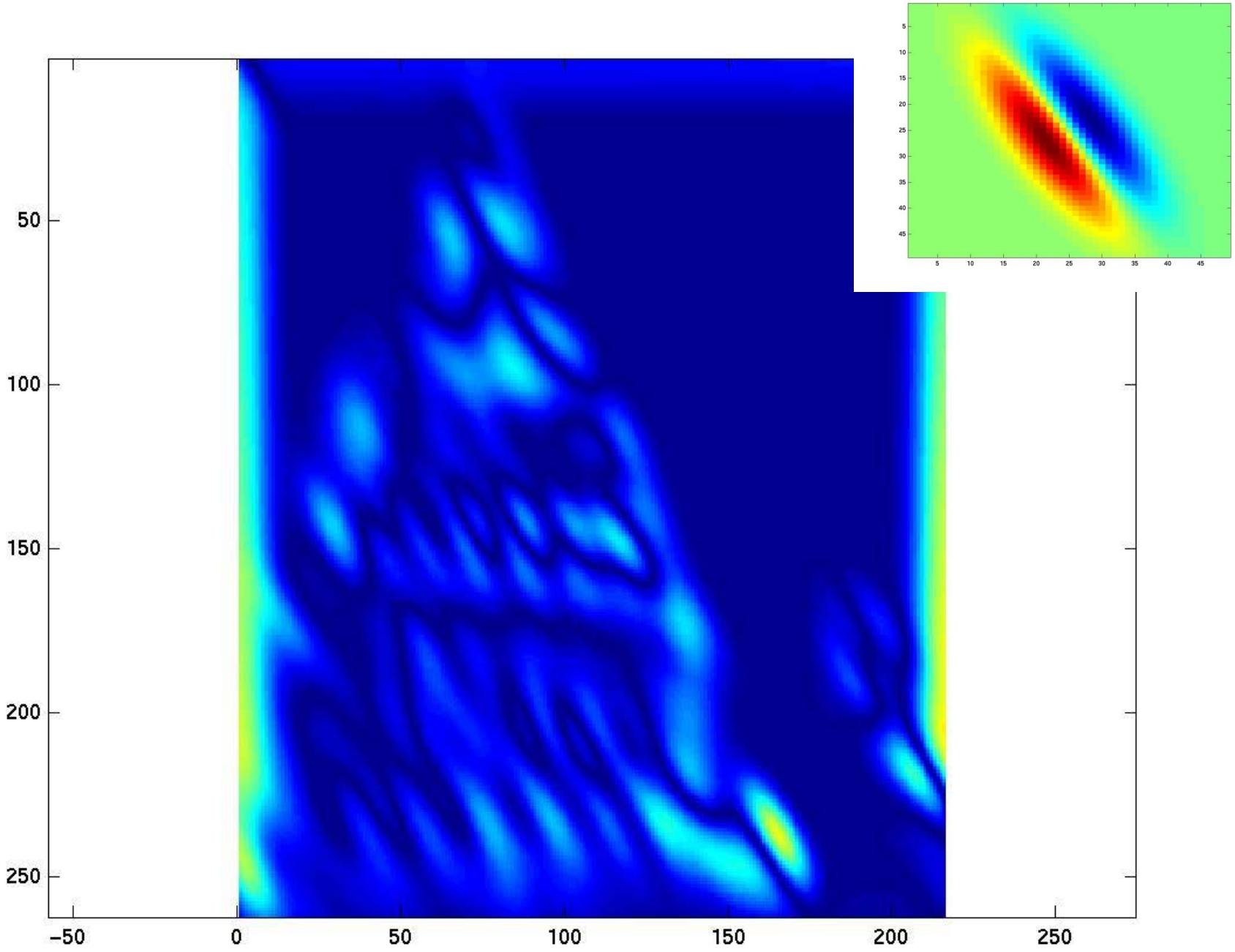

Kristen Grauman



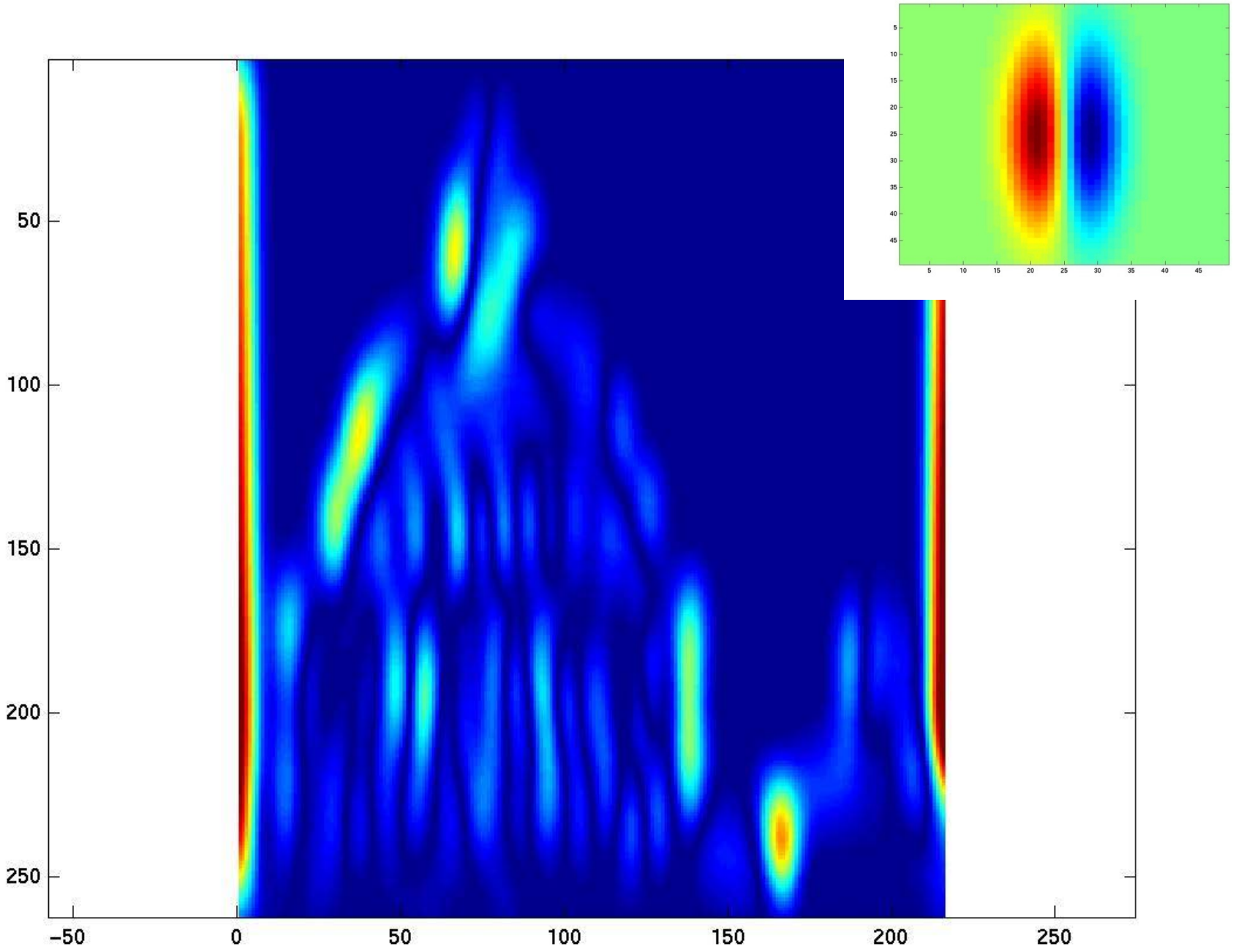

Kristen Grauman



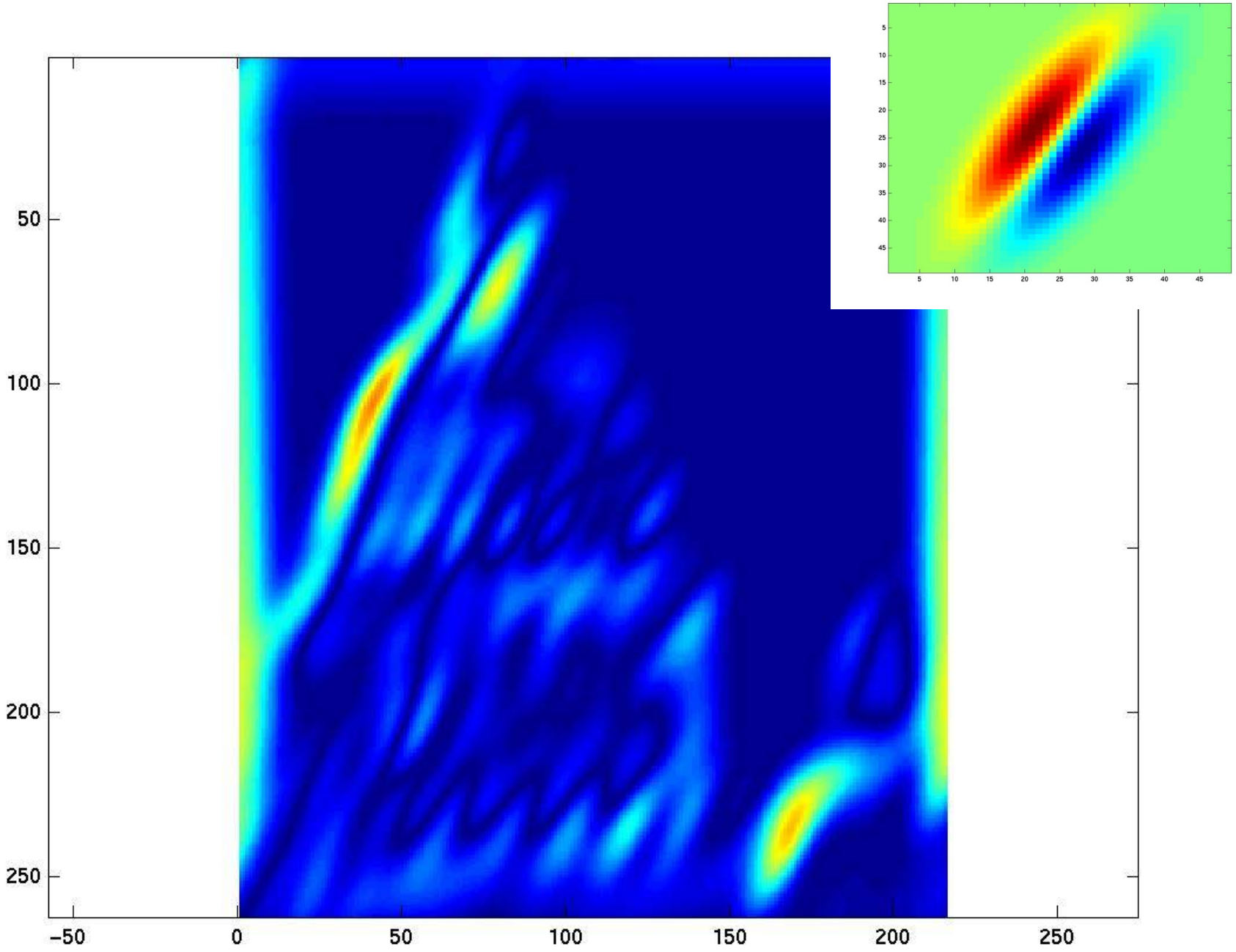

Kristen Grauman



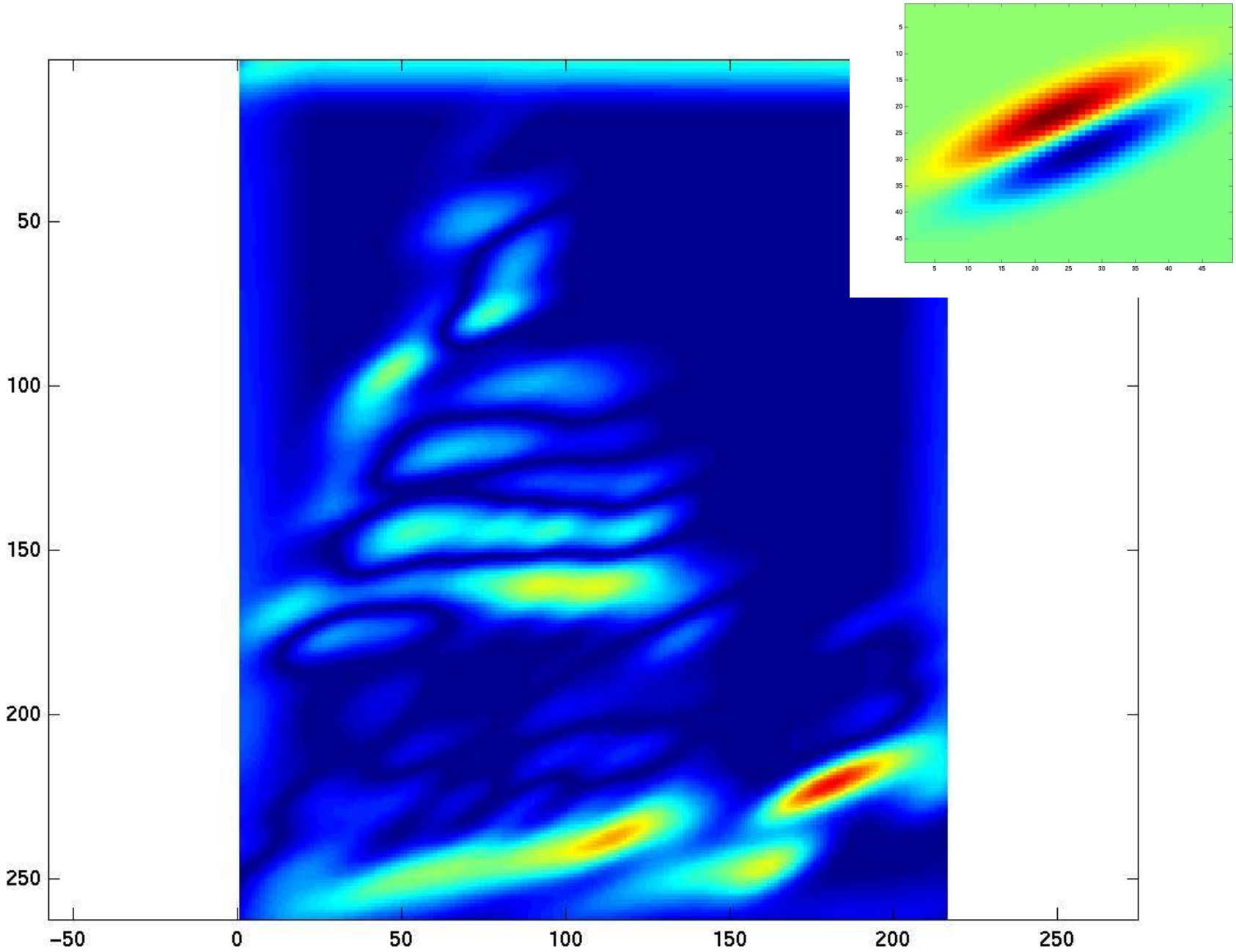

Kristen Grauman



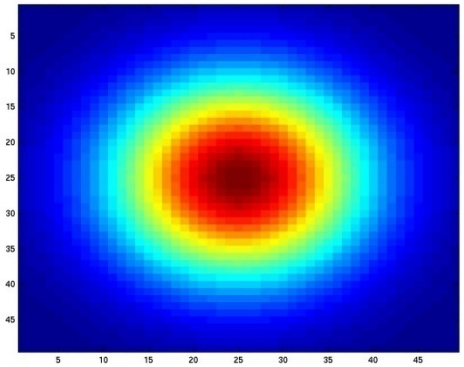

Kristen Grauman




Kristen Grauman




Kristen Grauman


Kristen Grauman

# What does it capture?

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

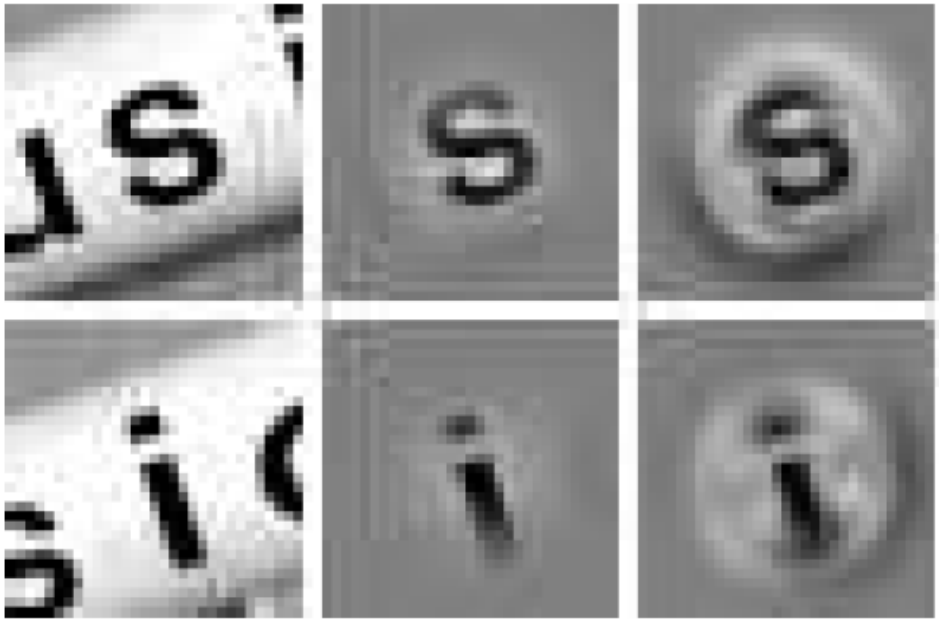


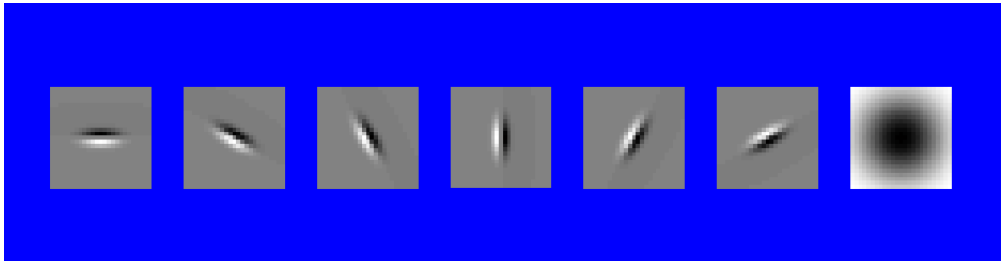
**Fig. 3.** Image reconstruction. Two example image patches (*left*), were reconstructed (*right*) from spatial filter responses at their center. Original image patches masked by a Gaussian (*middle*) are shown for comparison.

# How can we represent texture?

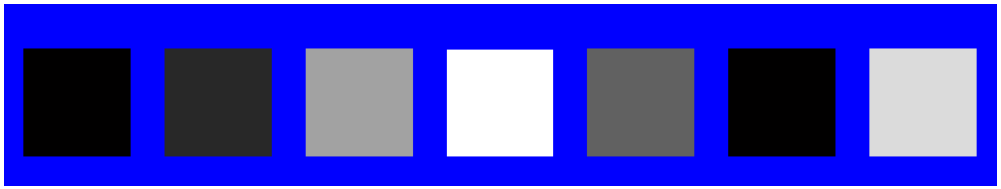
- Idea 1: Record simple statistics (e.g., mean, std.) of absolute filter responses

# Can you match the texture to the response?

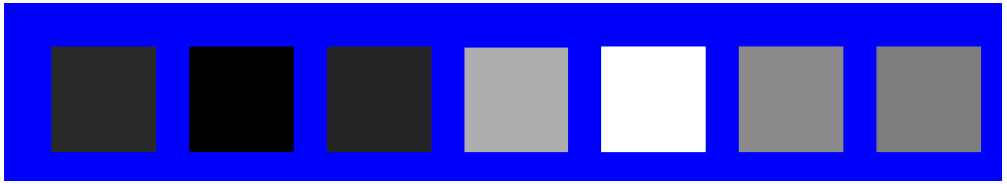
Filters



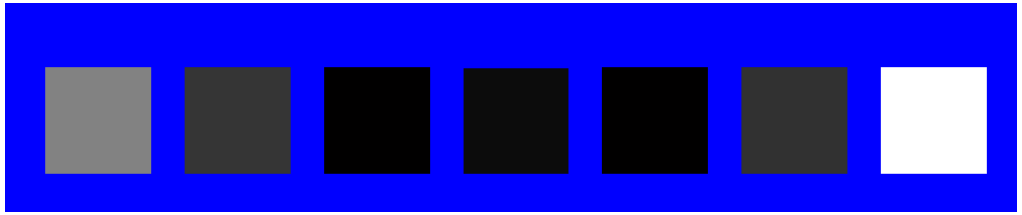
1



2

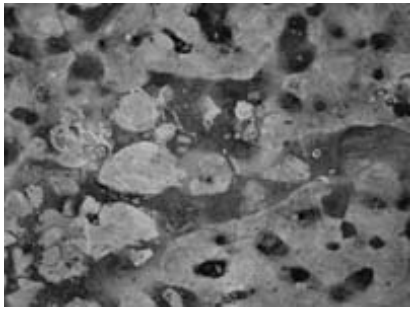


3



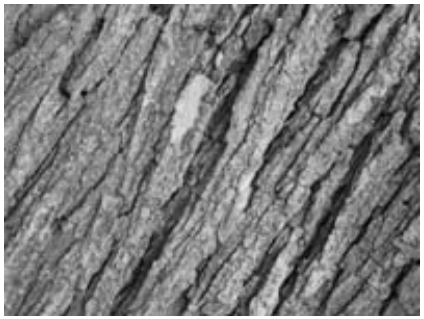
Mean abs responses

A



B

C



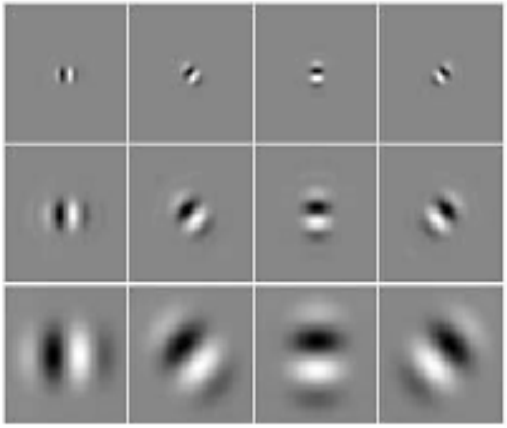
# How can we represent texture?

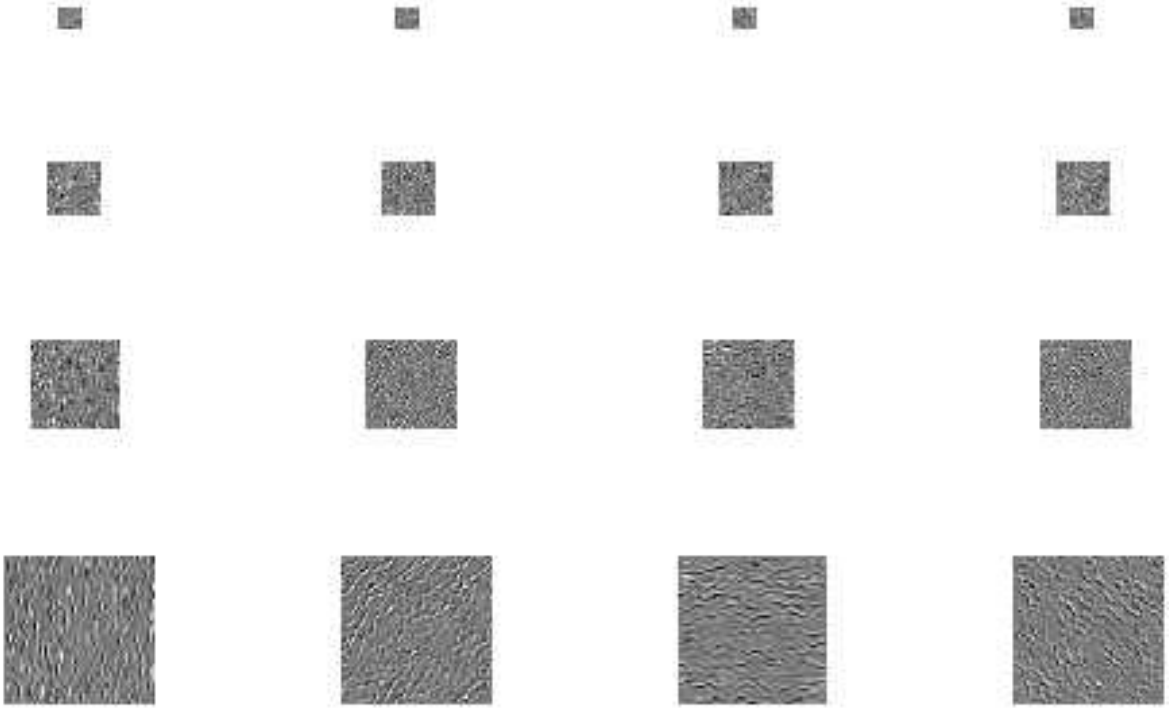
- Can be thought of as an single "orientation histogram"

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

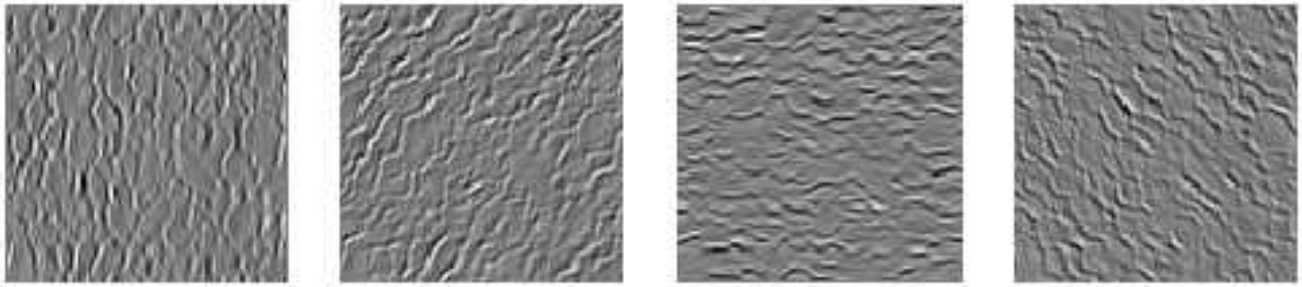
# Multi-scale filter decomposition

Filter bank

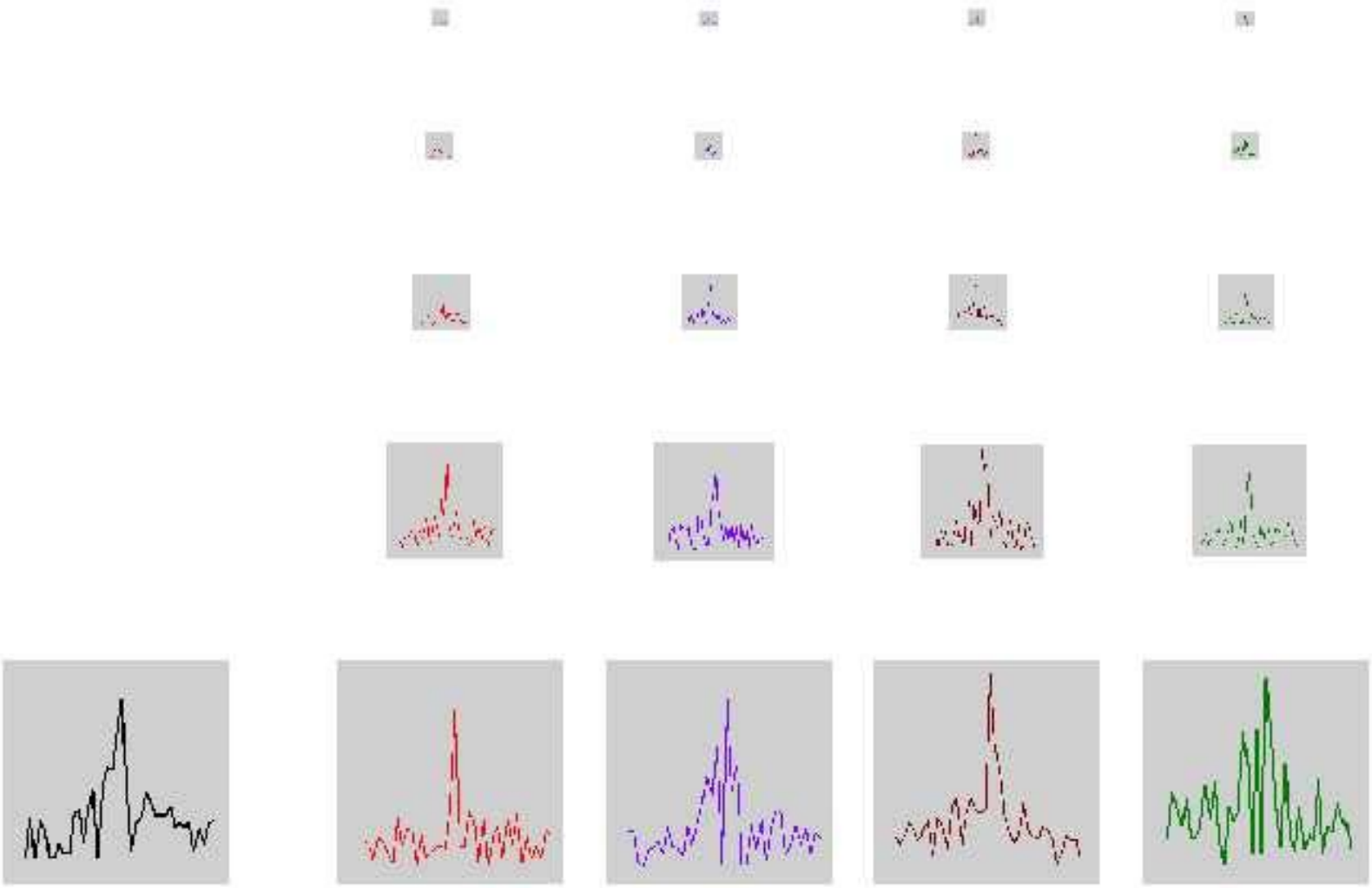




Input image



# Filter response histograms

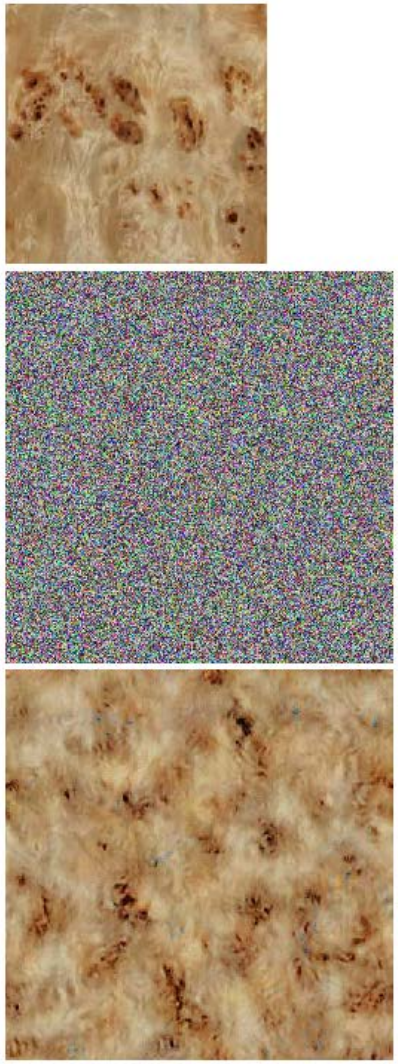


# Heeger & Bergen, SIGGRAPH‘95

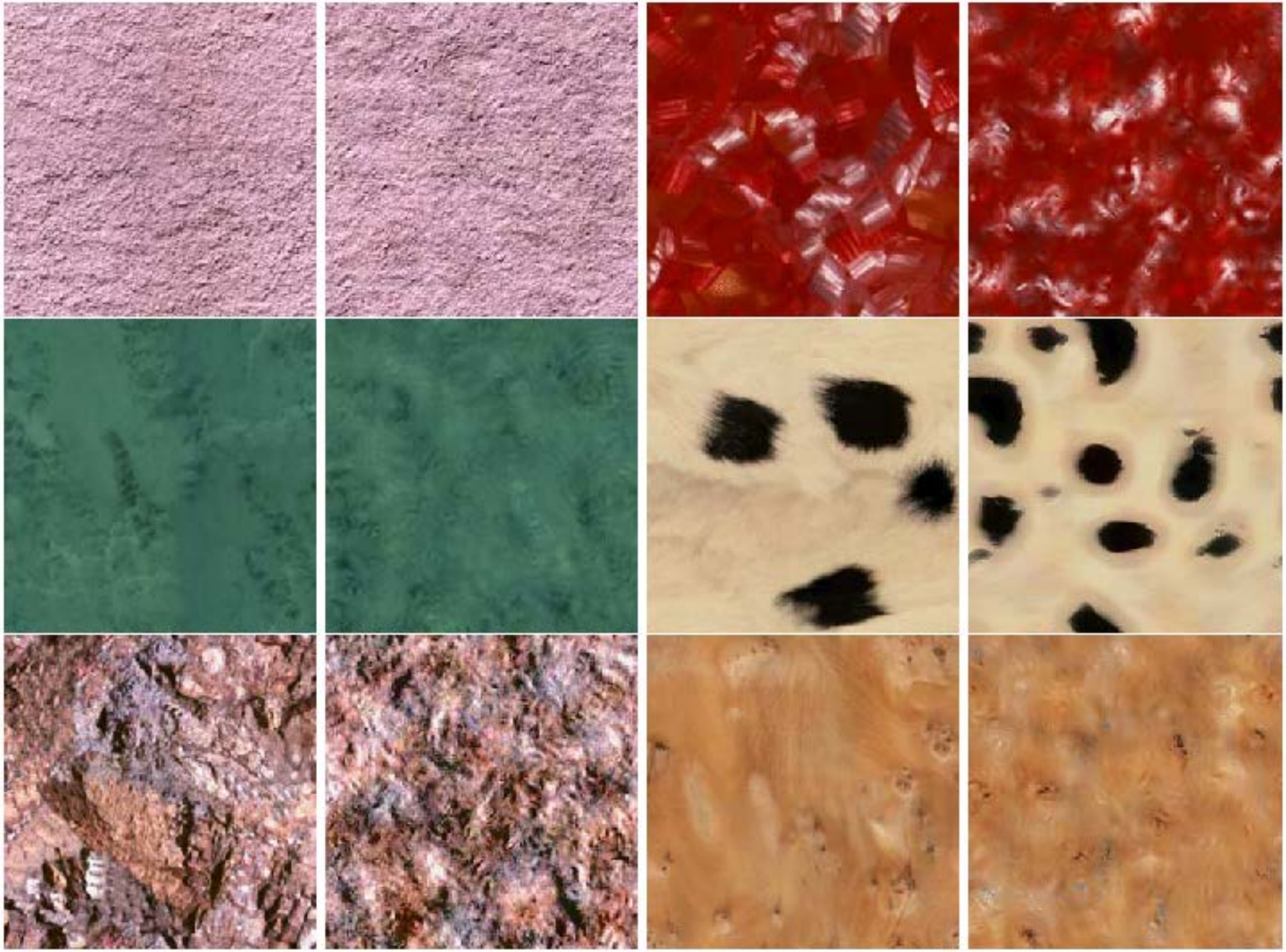
Start with a noise image as output

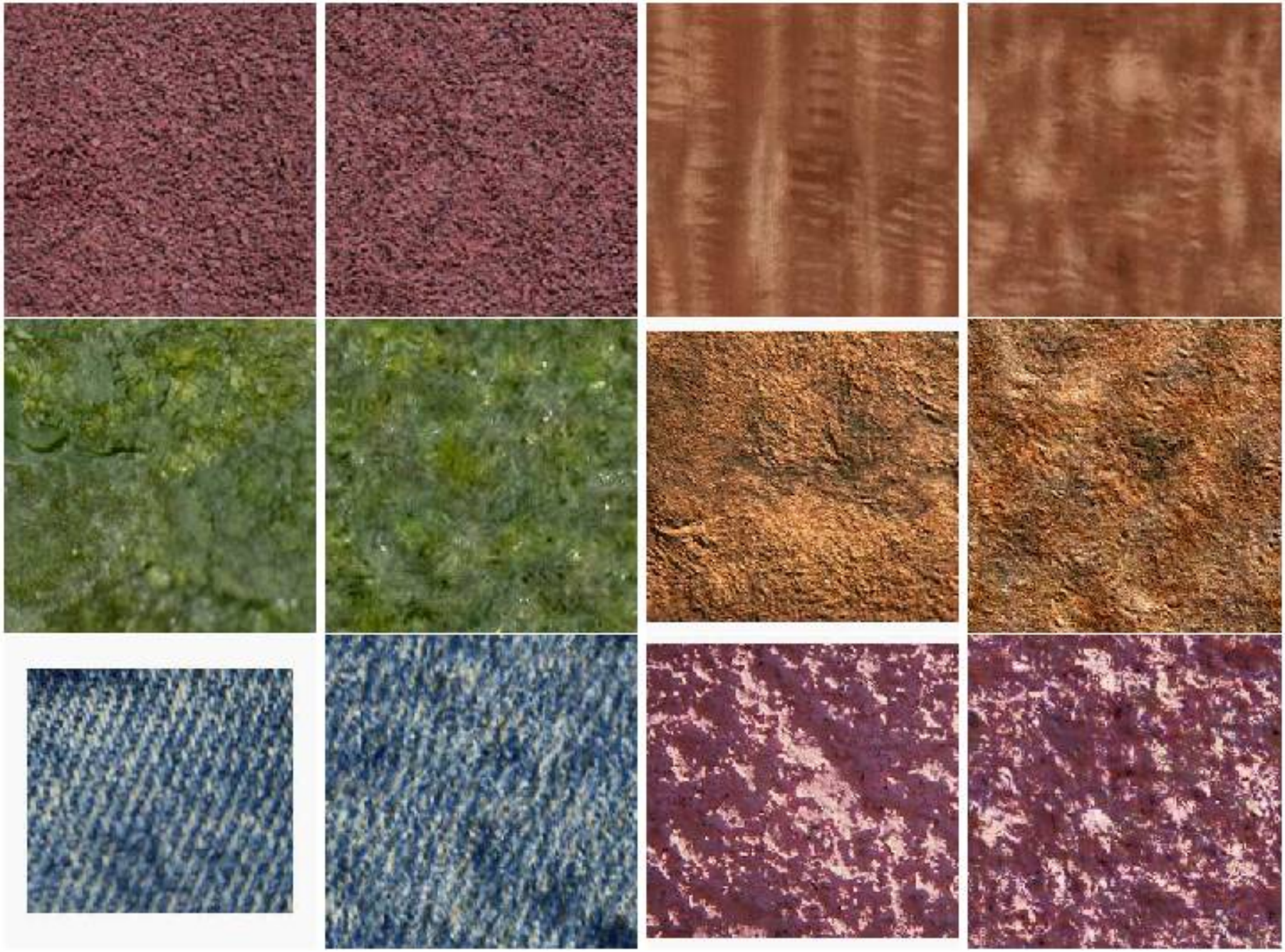
Main loop:

- 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





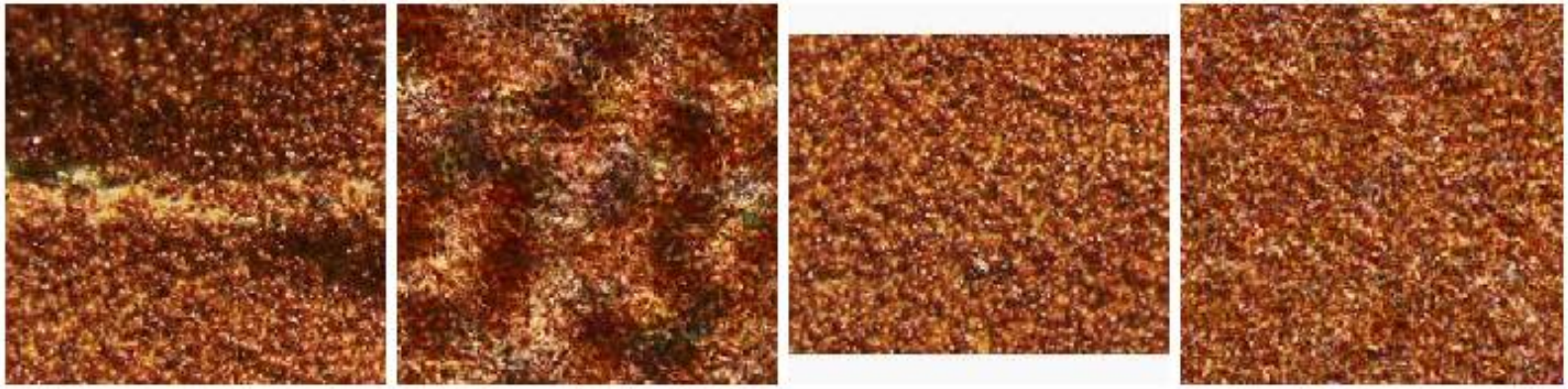


Figure 7: (Left pair) Inhomogeneous input texture produces blotchy synthetic texture. (Right pair) Homogenous input.

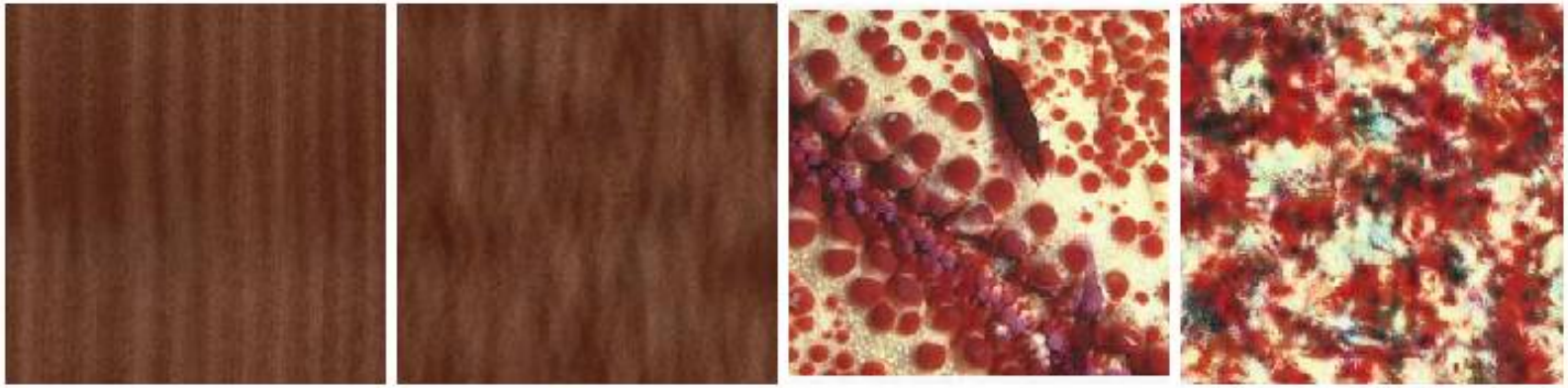


Figure 8: Examples of failures: wood grain and red coral.

Figure 9: More failures: hay and marble.

# Simoncelli & Portilla ’98+

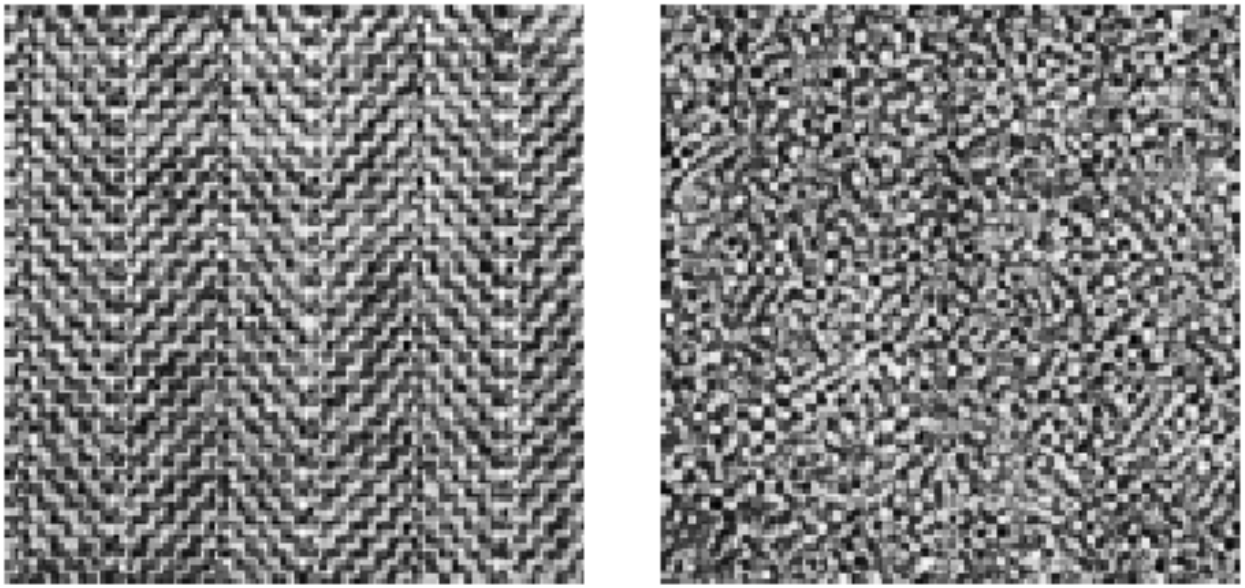


**Figure 1.** Textures with matching marginal statistics.

Marginal statistics are not enough

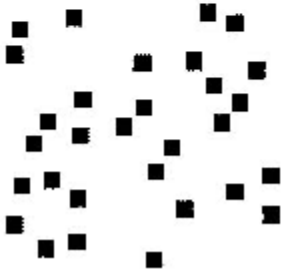
Neighboring filter responses are highly correlated

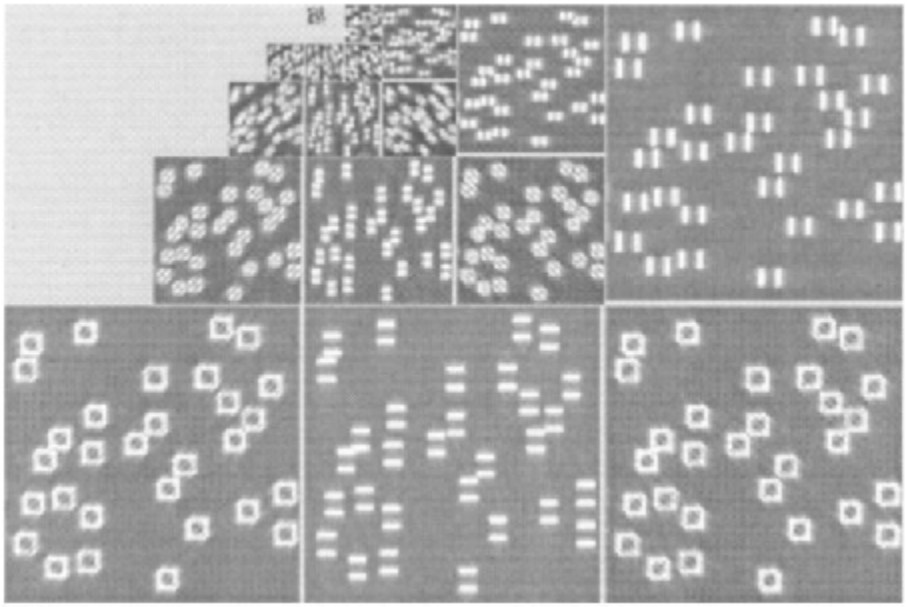
- an edge at low-res will cause an edge at high-res

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.

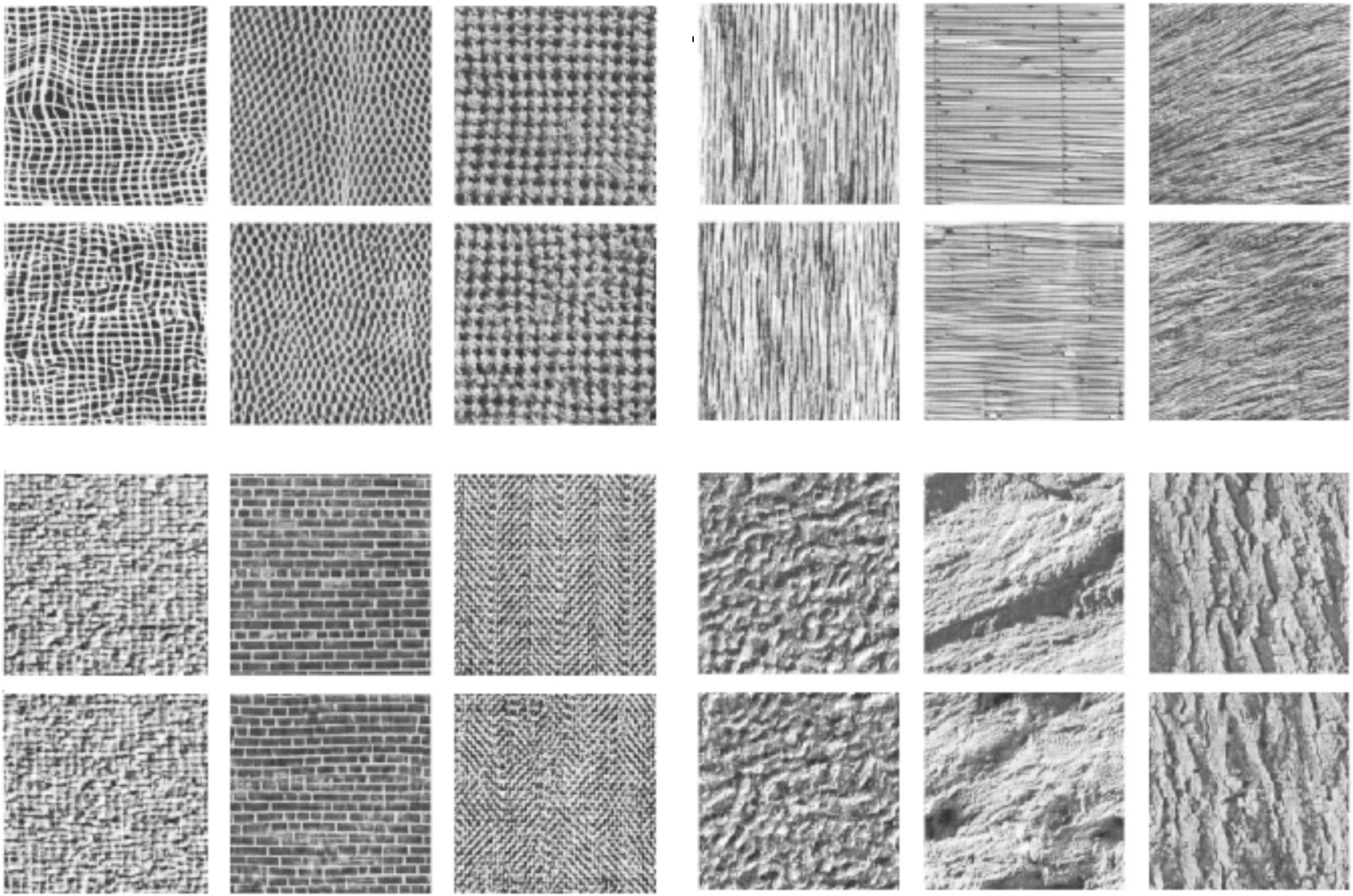
# Simoncelli & Portilla ’98+





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

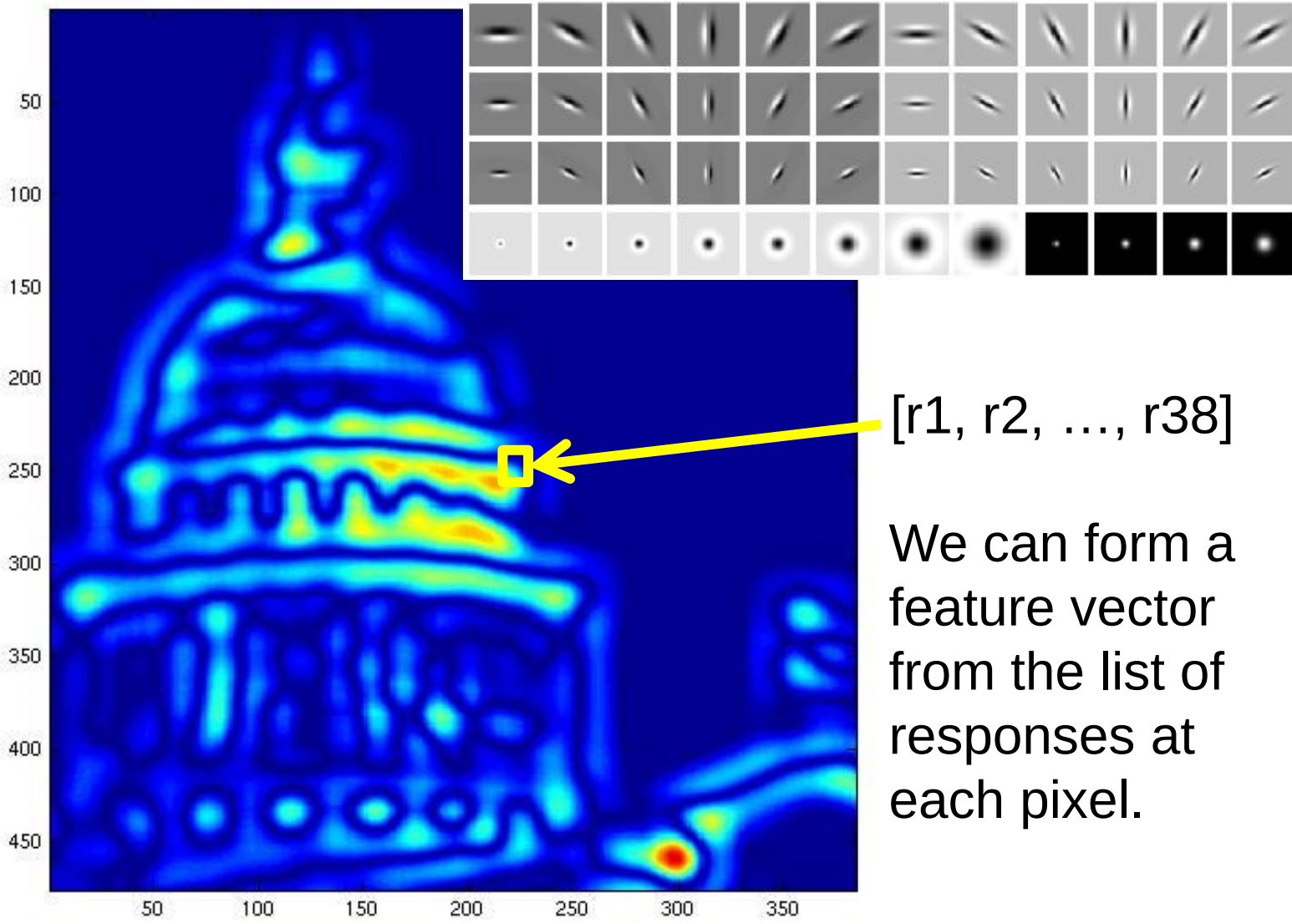
Optimize using repeated projections onto statistical constraint sufraces



# How can we represent texture?

- 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

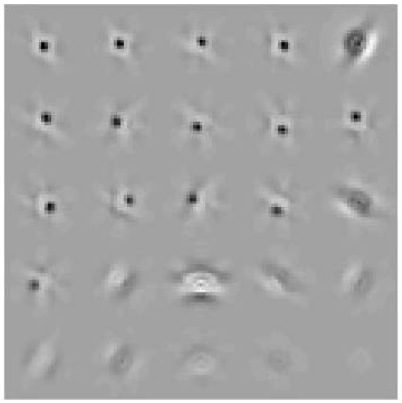


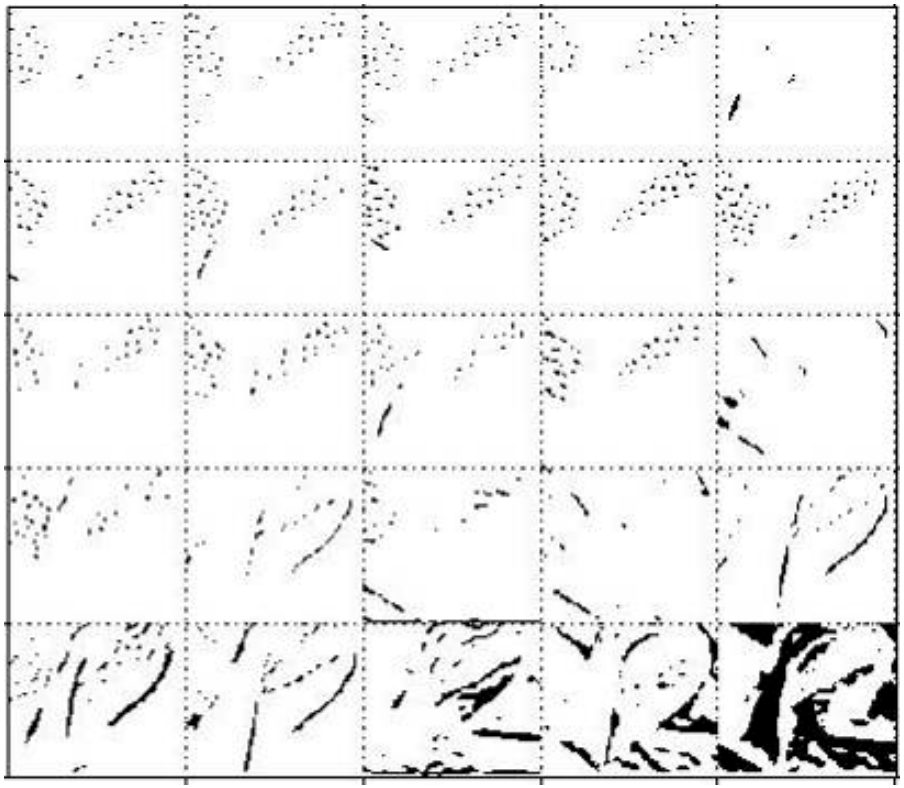

[r1, r2, …, r38]

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

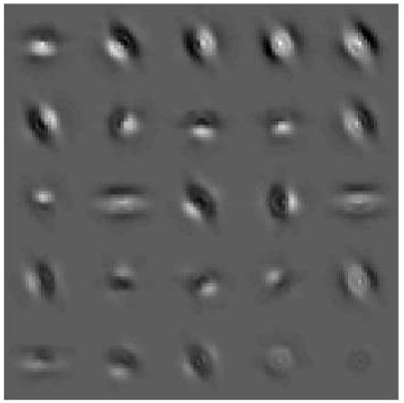
# Textons (Malik et al, IJCV 2001)

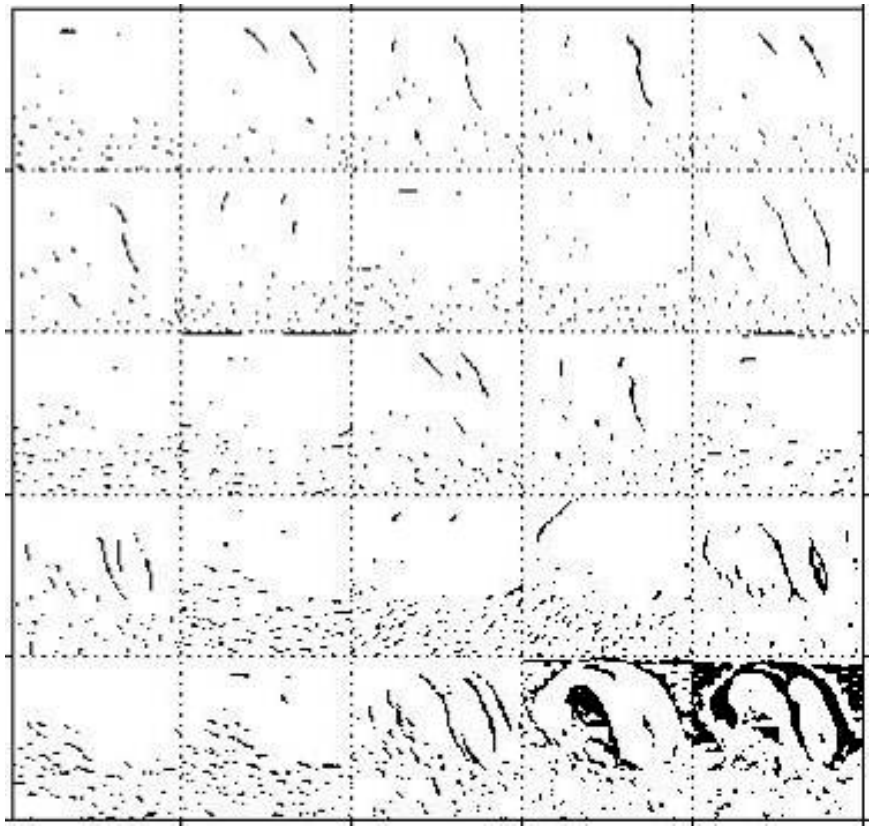
- Cluster vectors of filter responses





# Textons (cont.)





**Object** → **Bag of ‘words’**

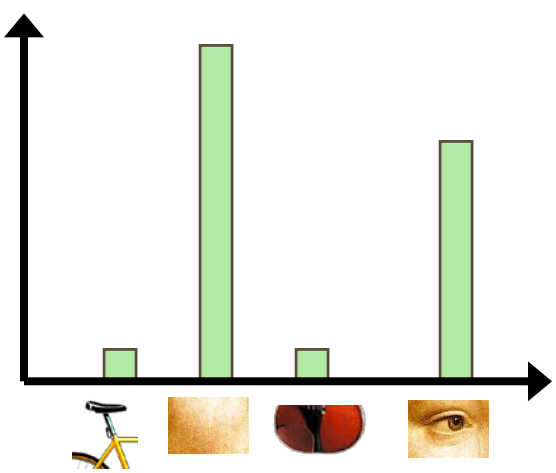
# Analogy to documents

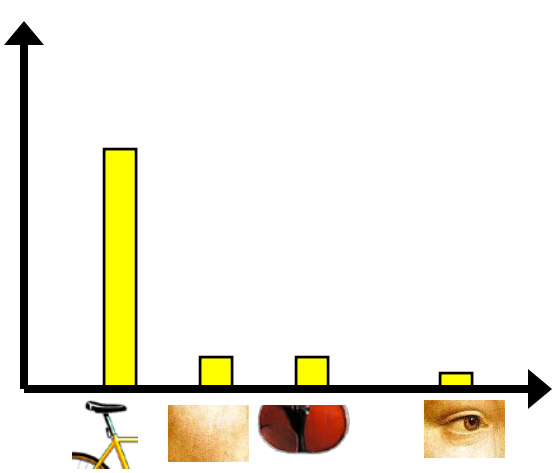
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that r[illegible] our eyes. For a long ti[illegible] retinal image wa[illegible] isual centers i[illegible] s a movie s[illegible] 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.*

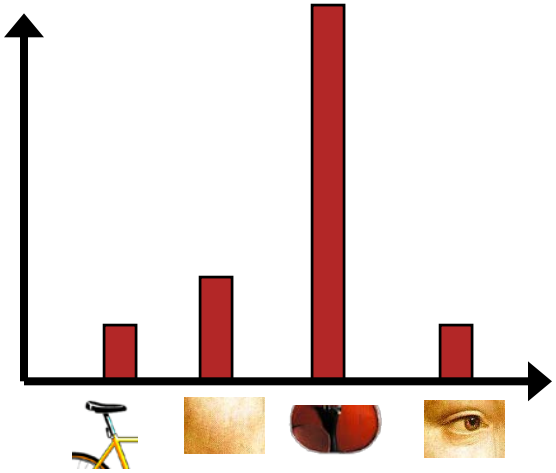
**sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel**

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.

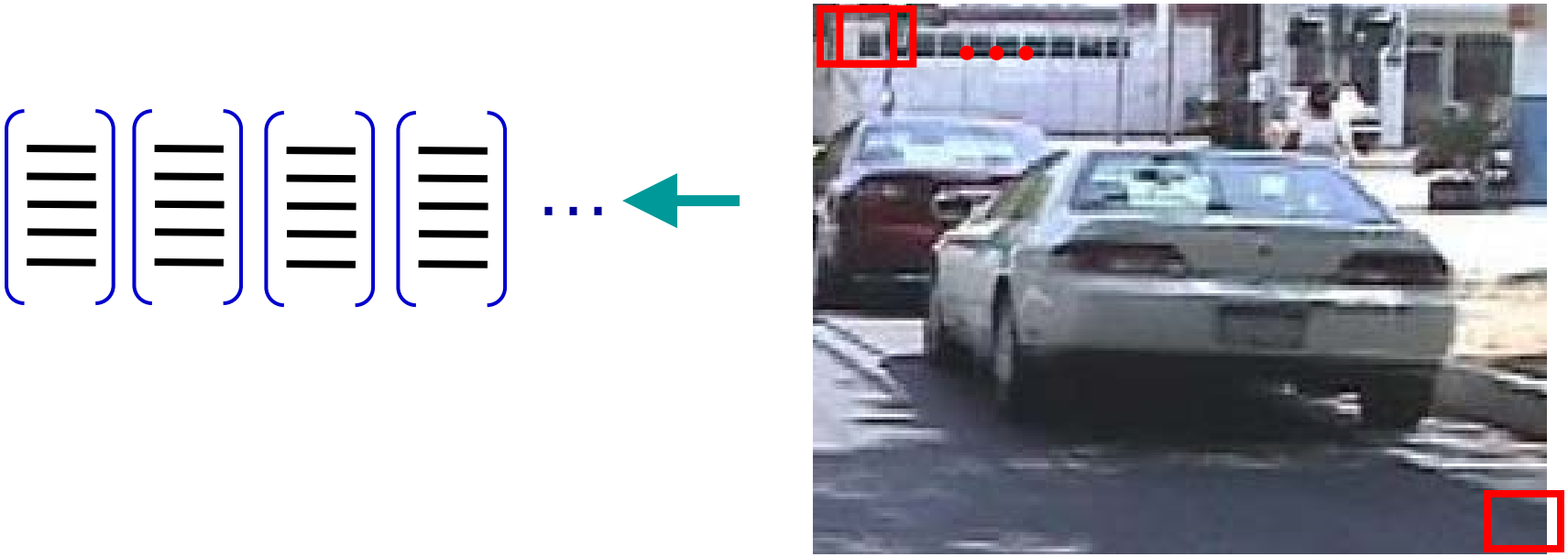
**China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value**



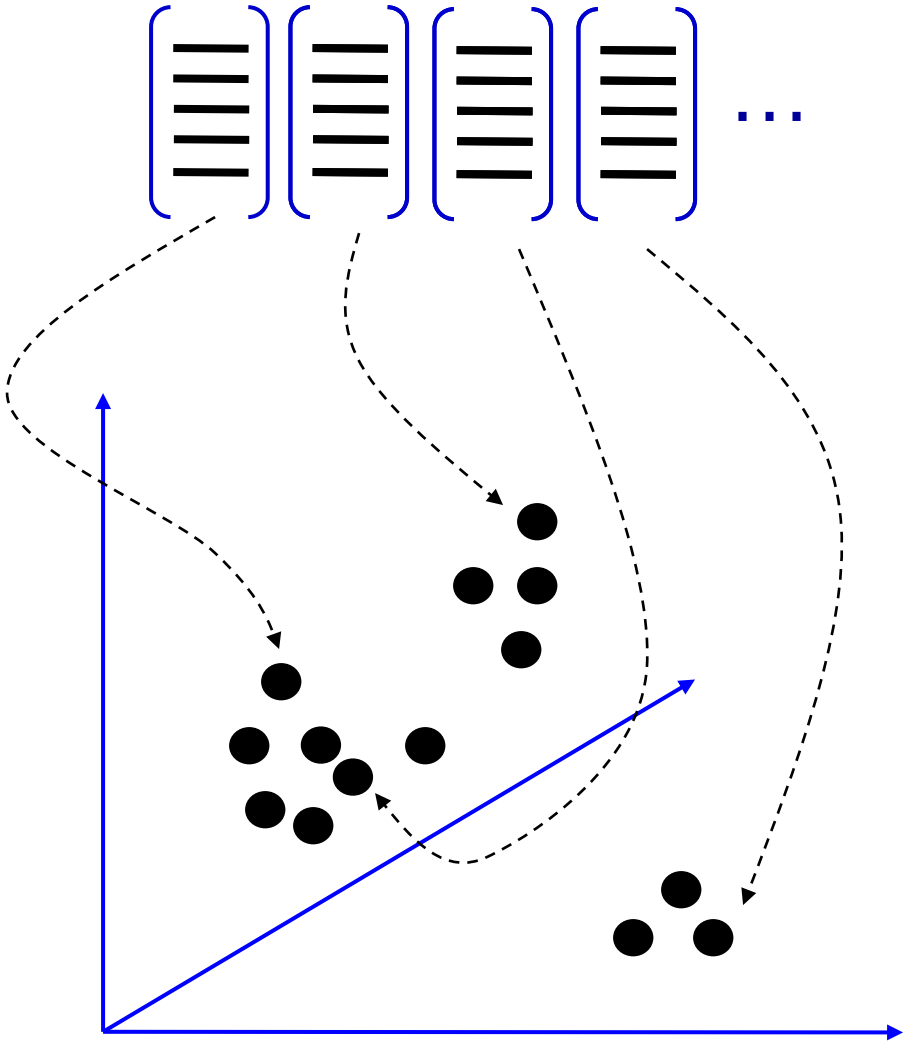




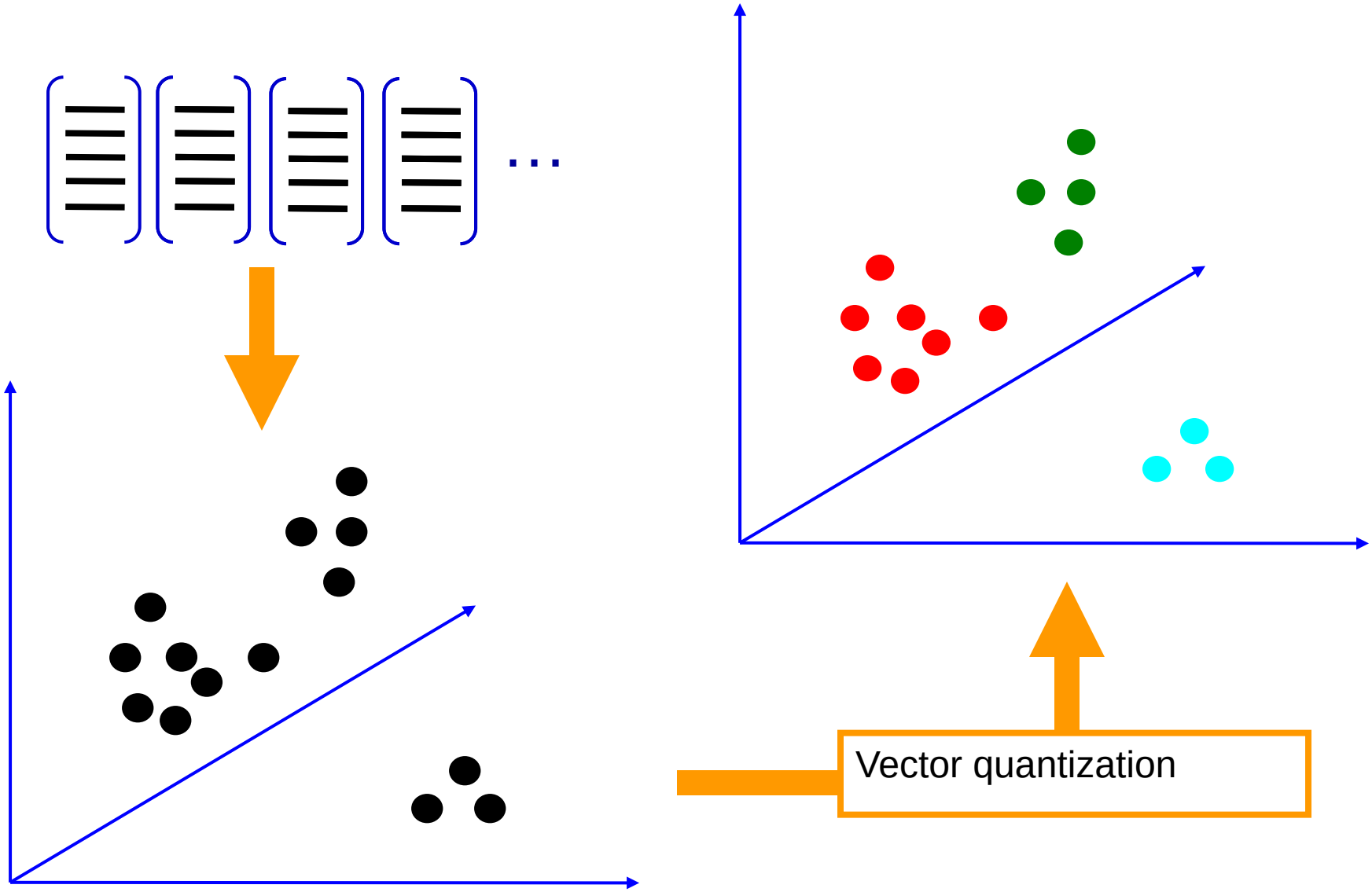
# Patch Features



# dictionary formation

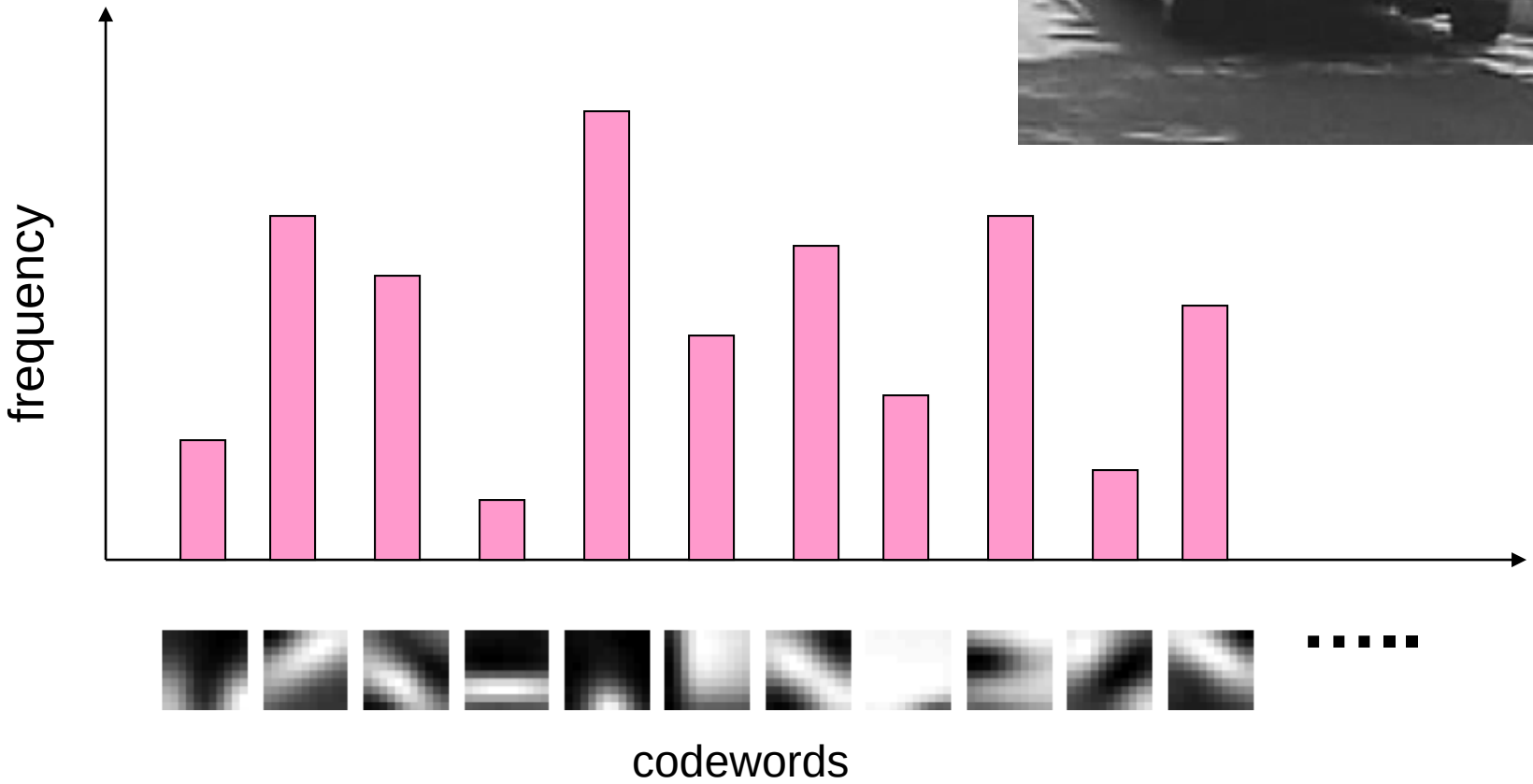


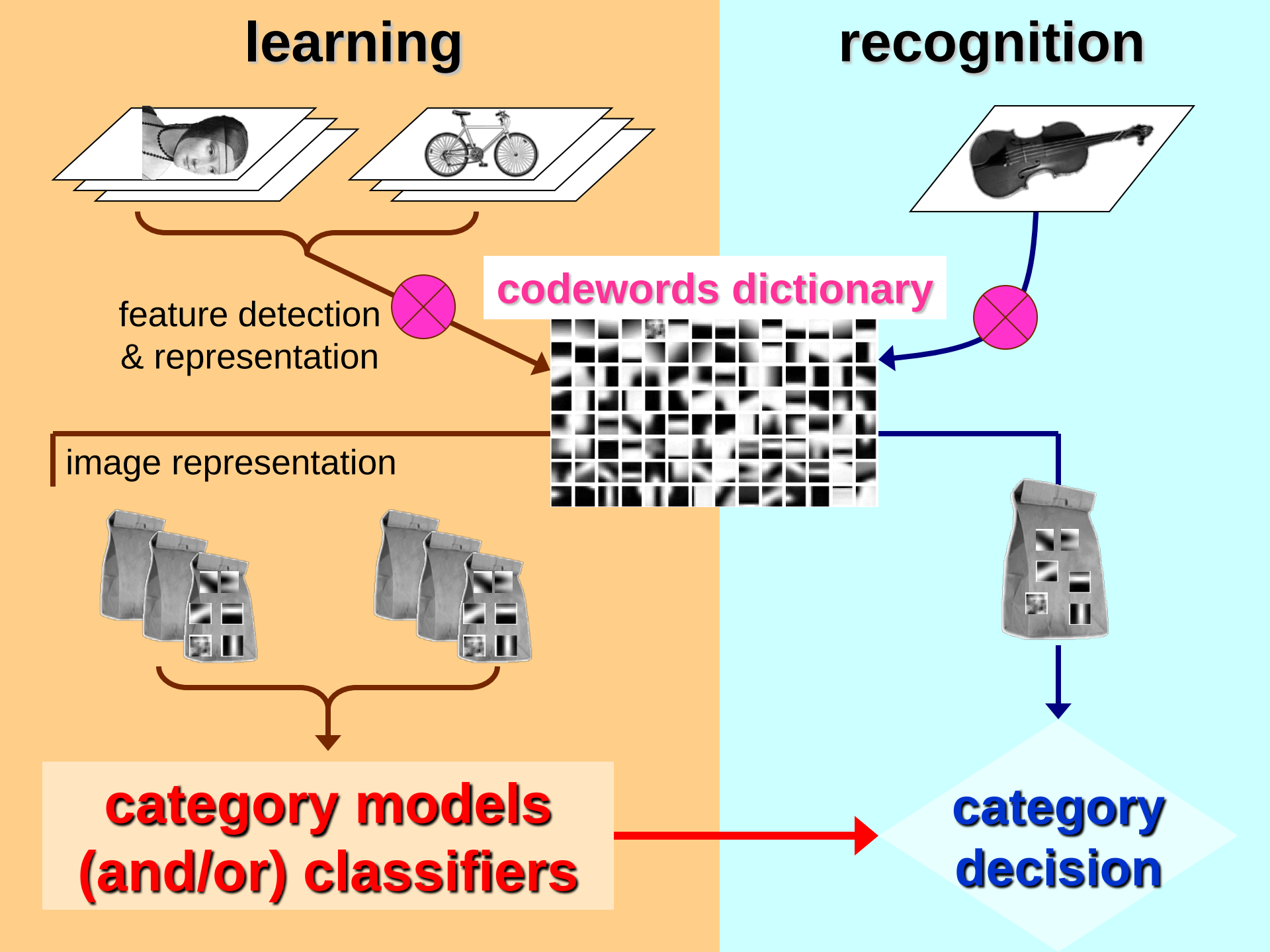
# Clustering (usually k-means)




Slide credit: Josef Sivic

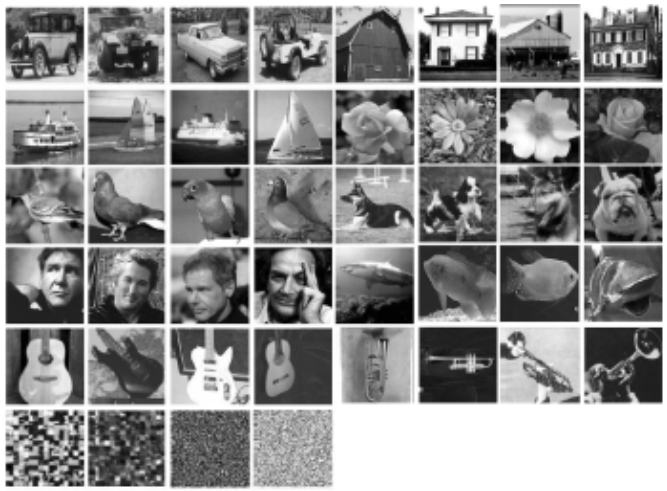
# Image representation

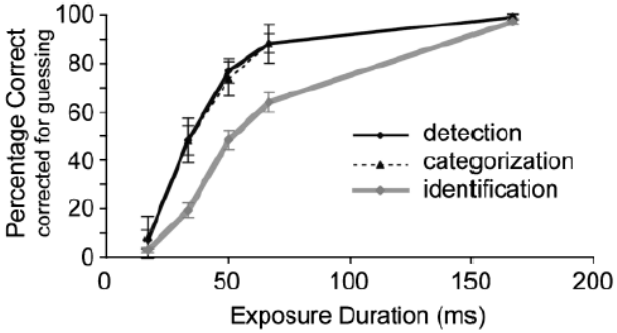



learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision

# 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)





# Scene Classification (Renninger & Malik)

beach

mountain

forest

city

street

farm

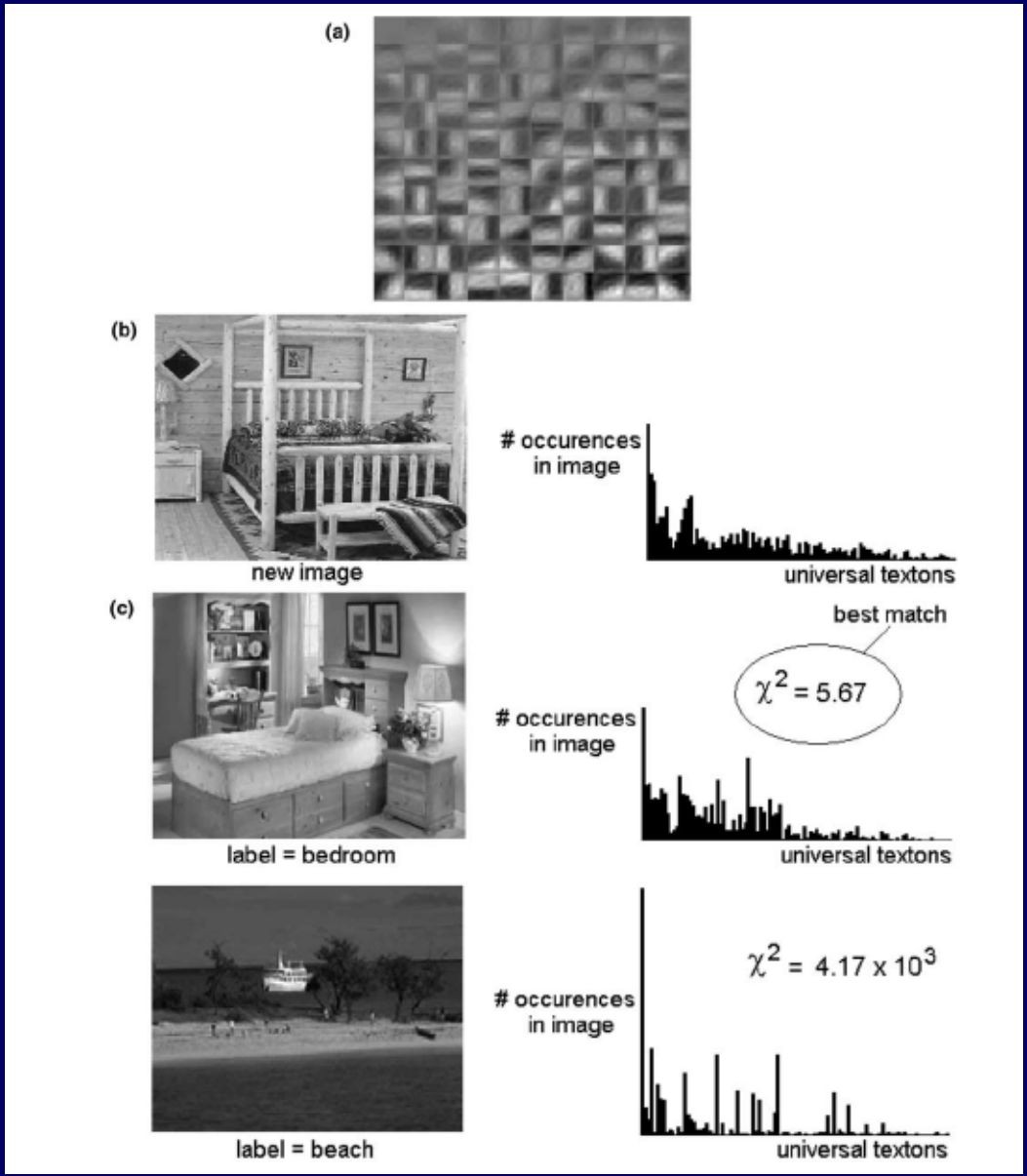
kitchen

livingroom

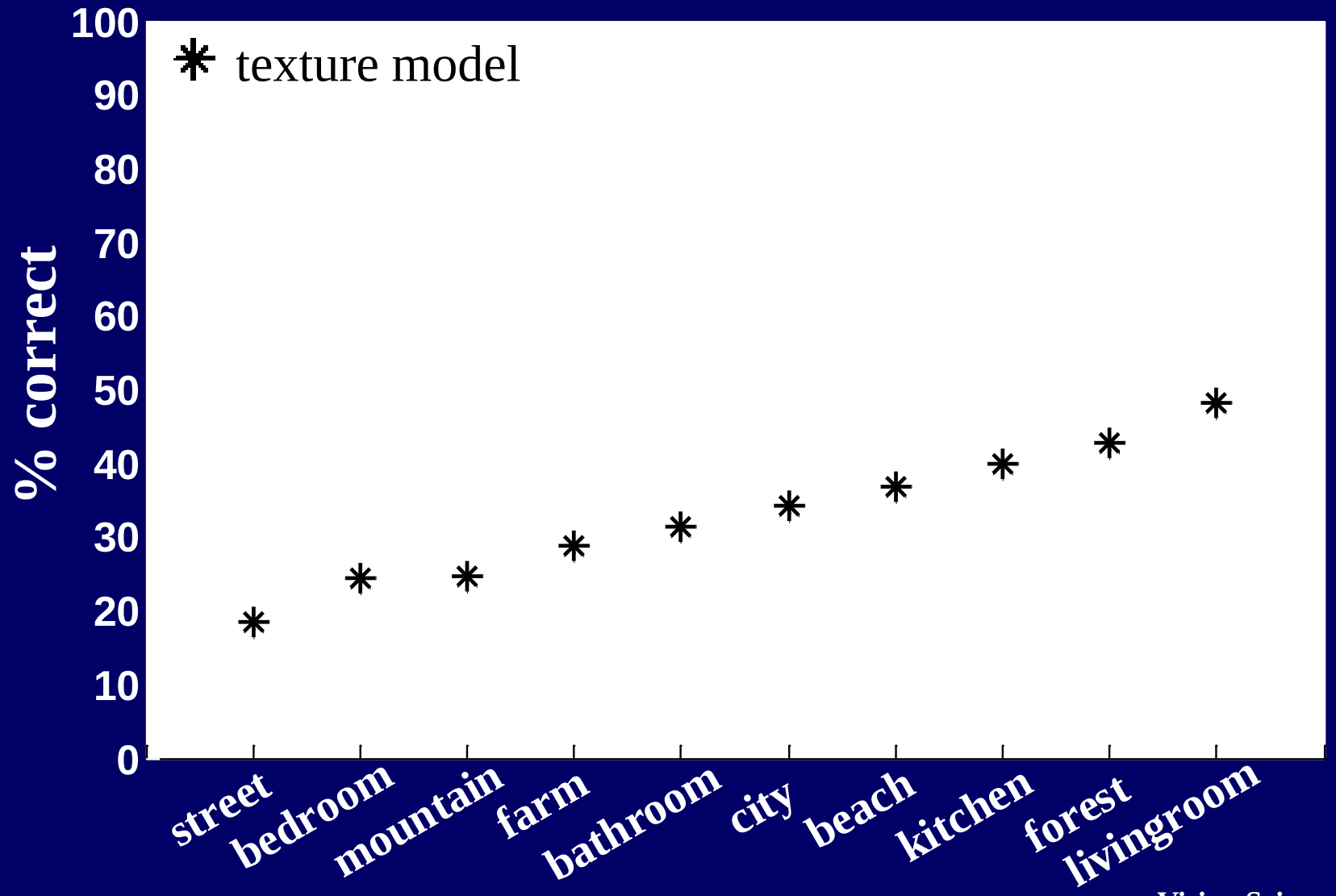
bedroom

bathroom

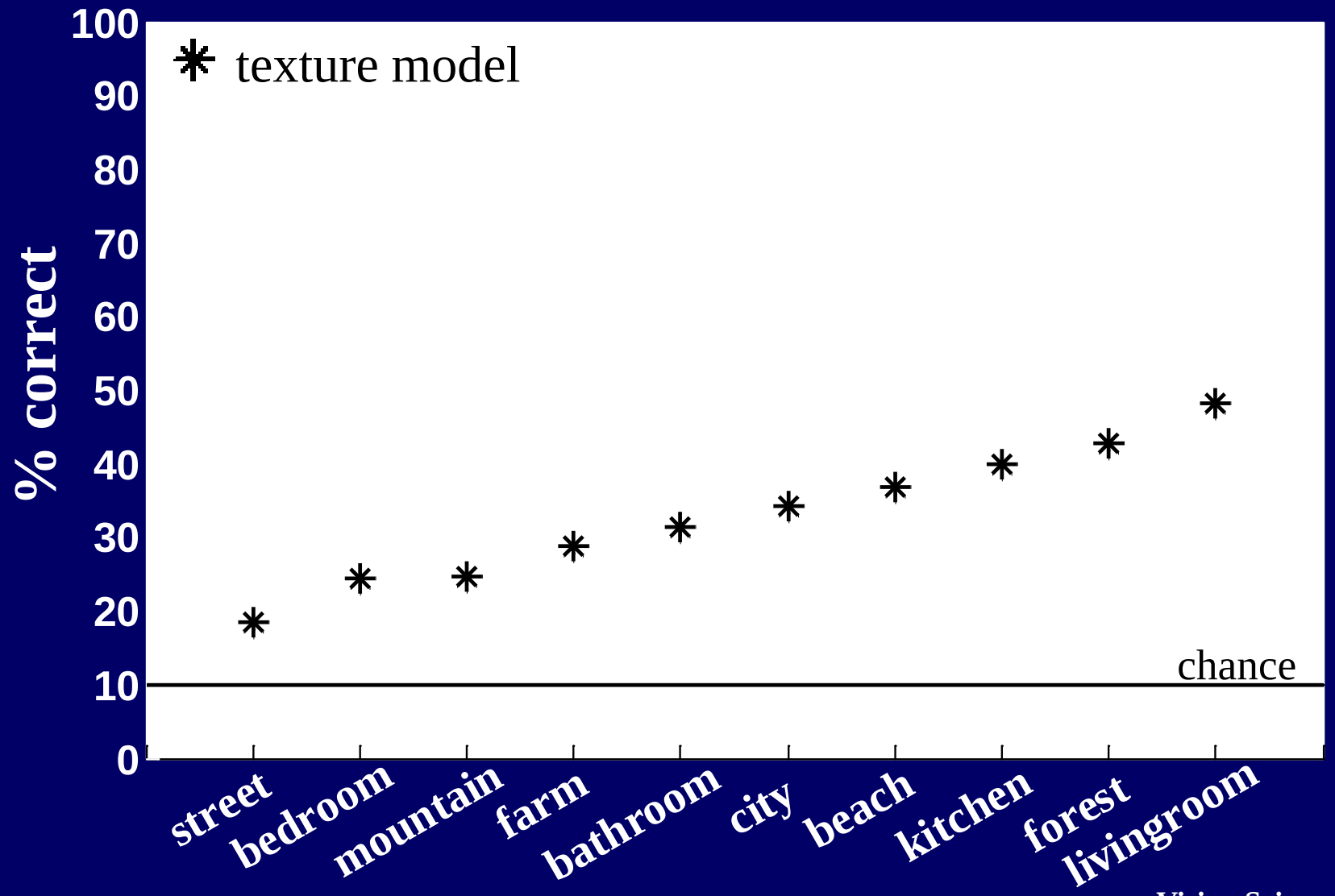
# Texton Histogram Matching



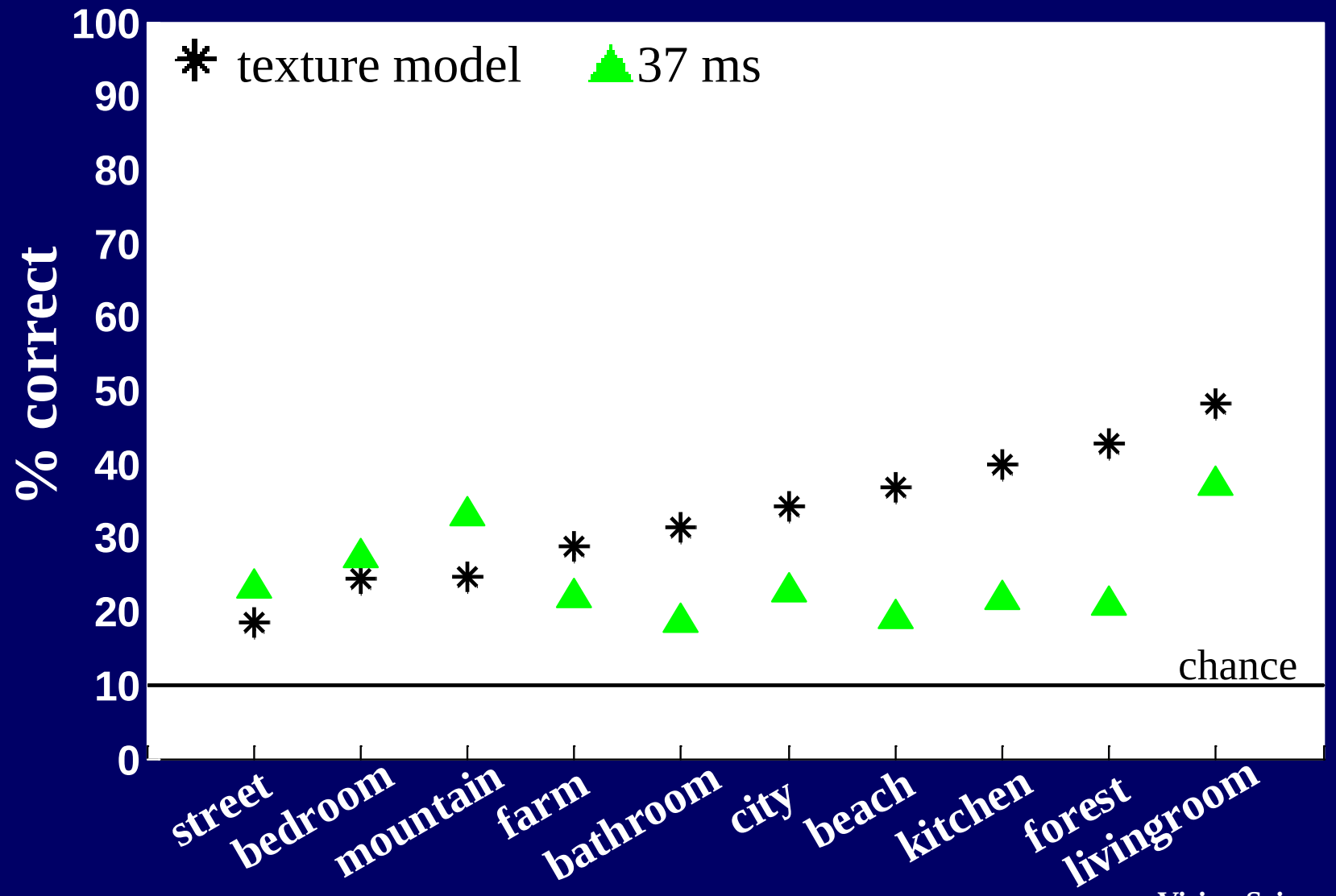
# Discrimination of Basic Categories



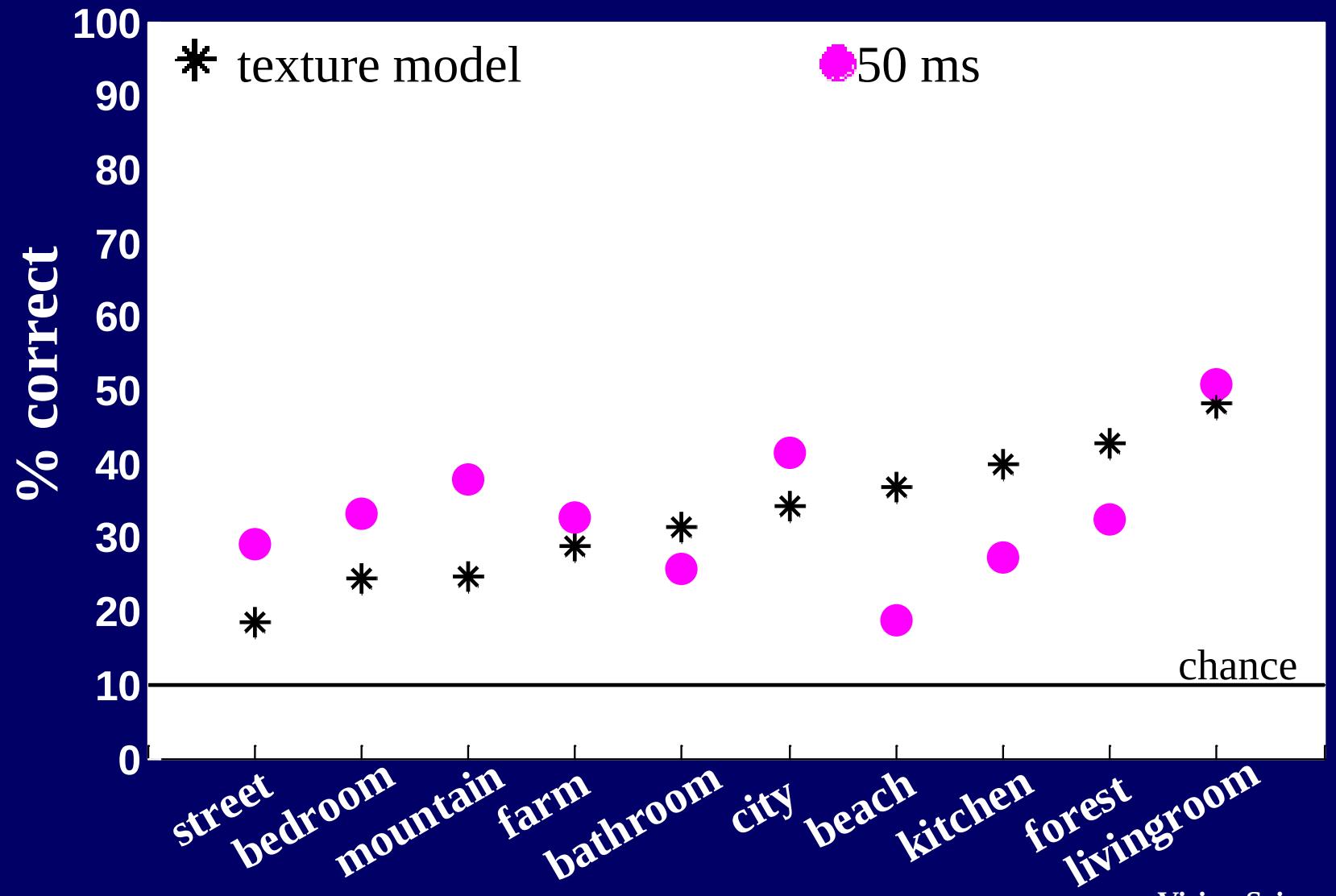
# Discrimination of Basic Categories



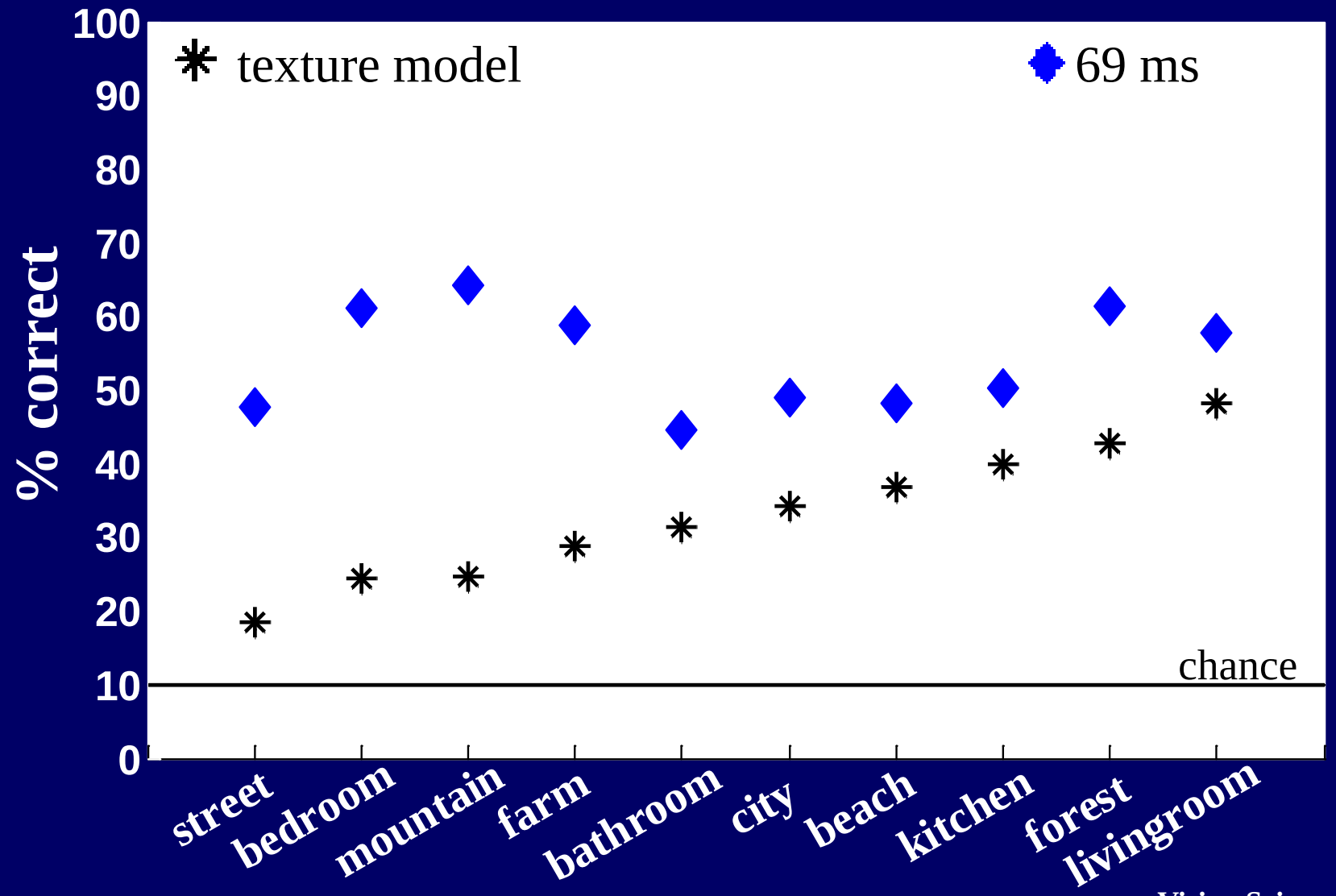
# Discrimination of Basic Categories



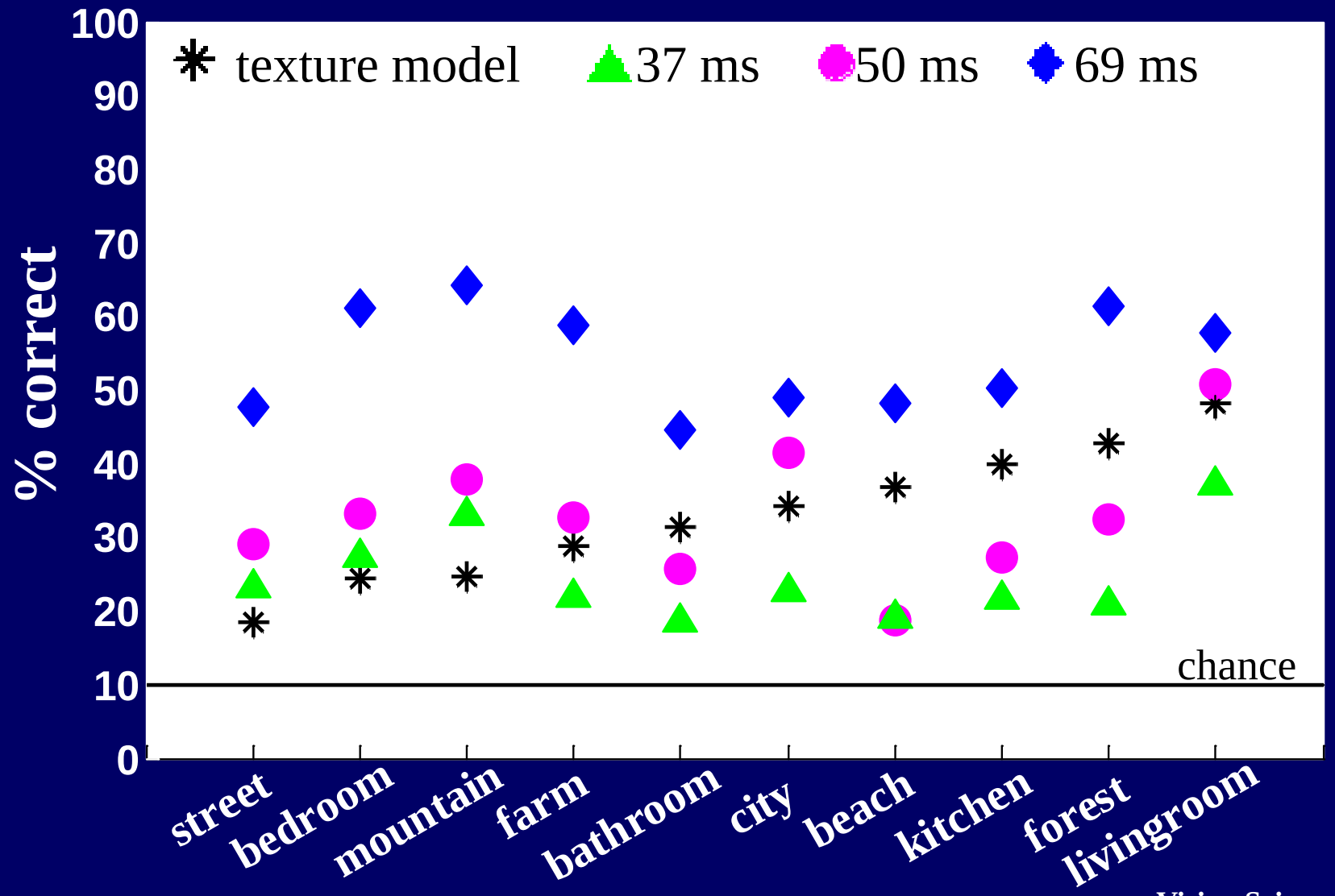
# Discrimination of Basic Categories



# 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. Olshausen[1] and David J. Field

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

where $\lambda$ 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.*