

Statistical Images Representations I

(Everything is Texture!)



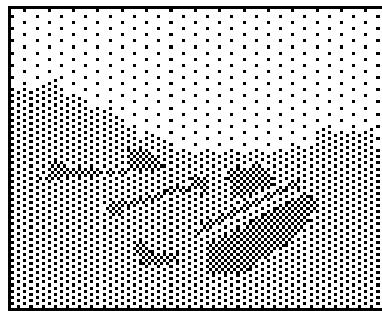
Claude
Shannon



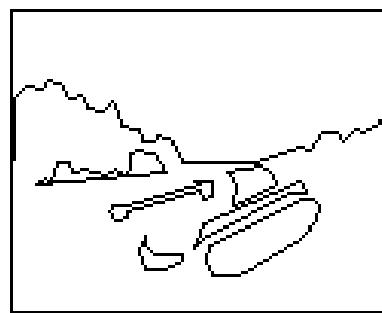
Bela
Julesz

Things used to be so easy...

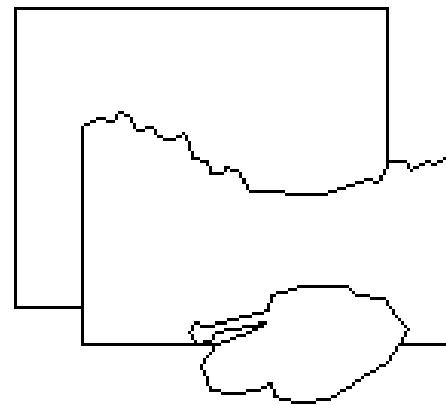
input image



edge image

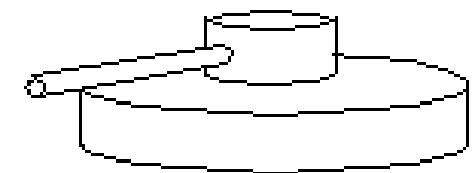


$2\frac{1}{2}$ -D sketch



(primal sketch)

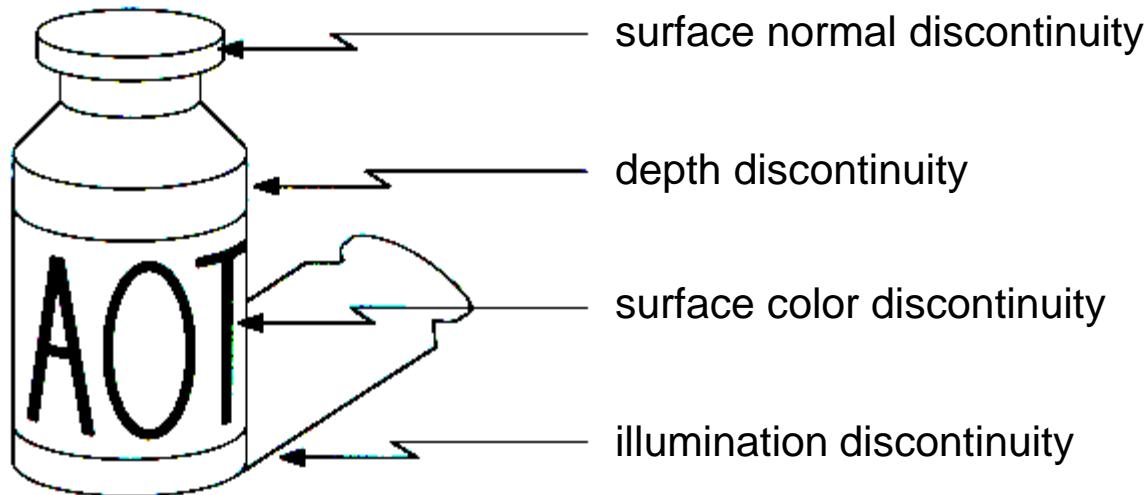
3-D model



David Marr's View of Vision (1980s)

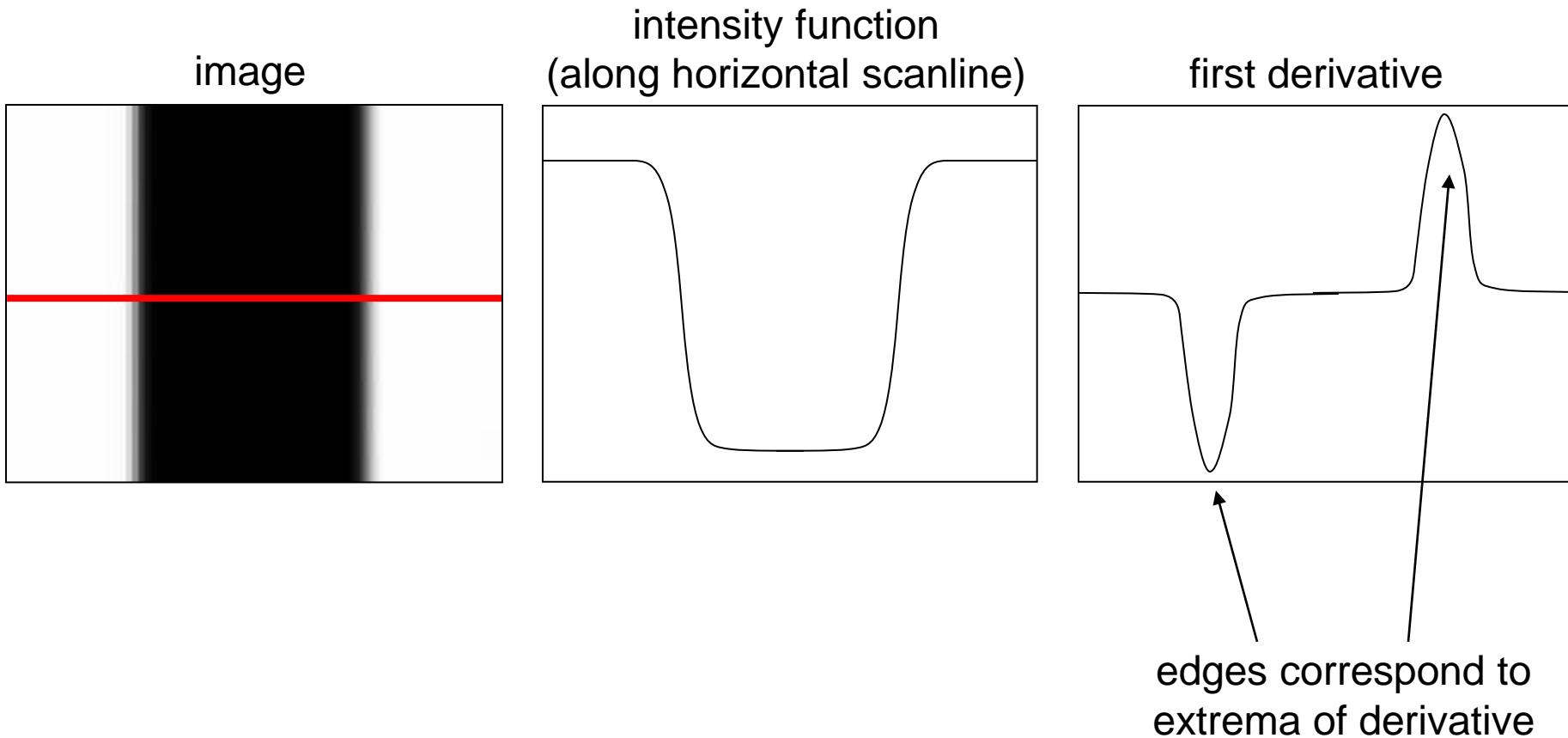
Origin of edges

Edges are caused by a variety of factors:



Characterizing edges

- An edge is a place of rapid change in the image intensity function



An edge is not a line...

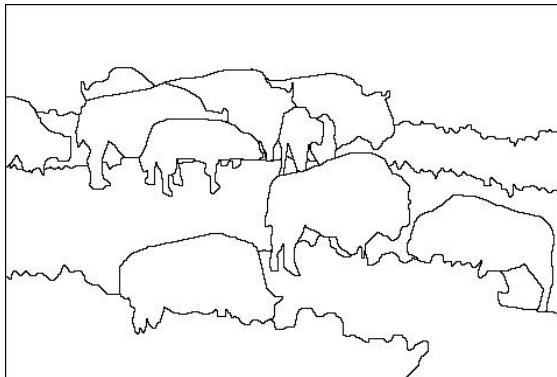


State-of-the-Art Edge Detection

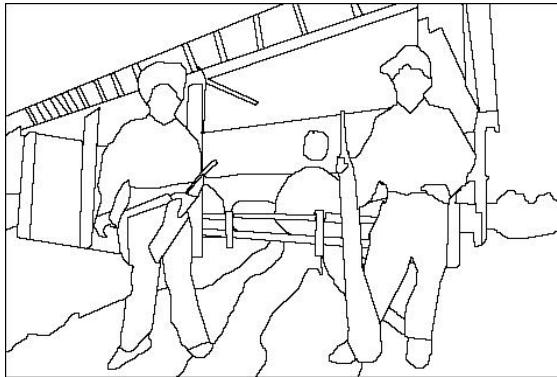
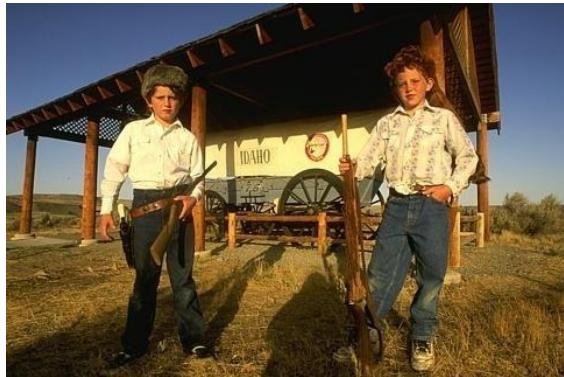
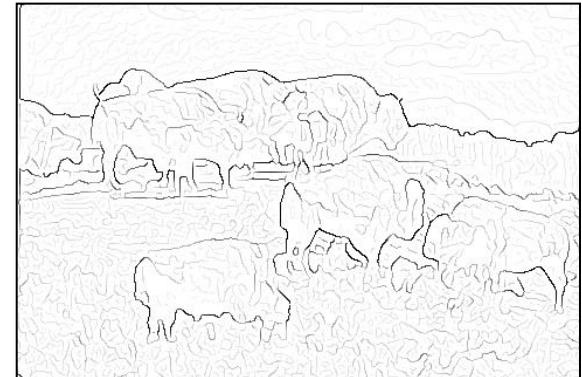
image



human segmentation



gradient magnitude



Berkeley segmentation database:

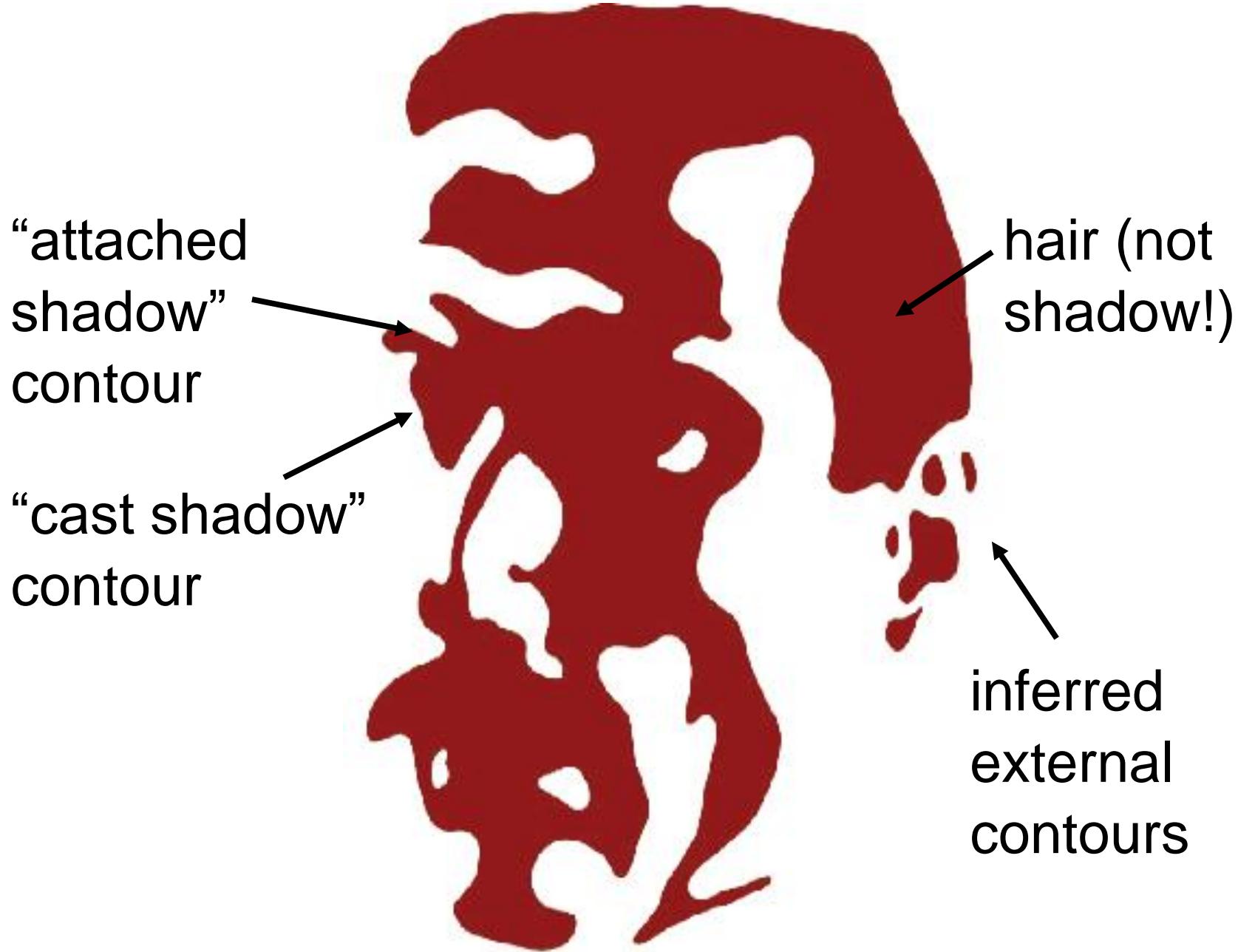
<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>

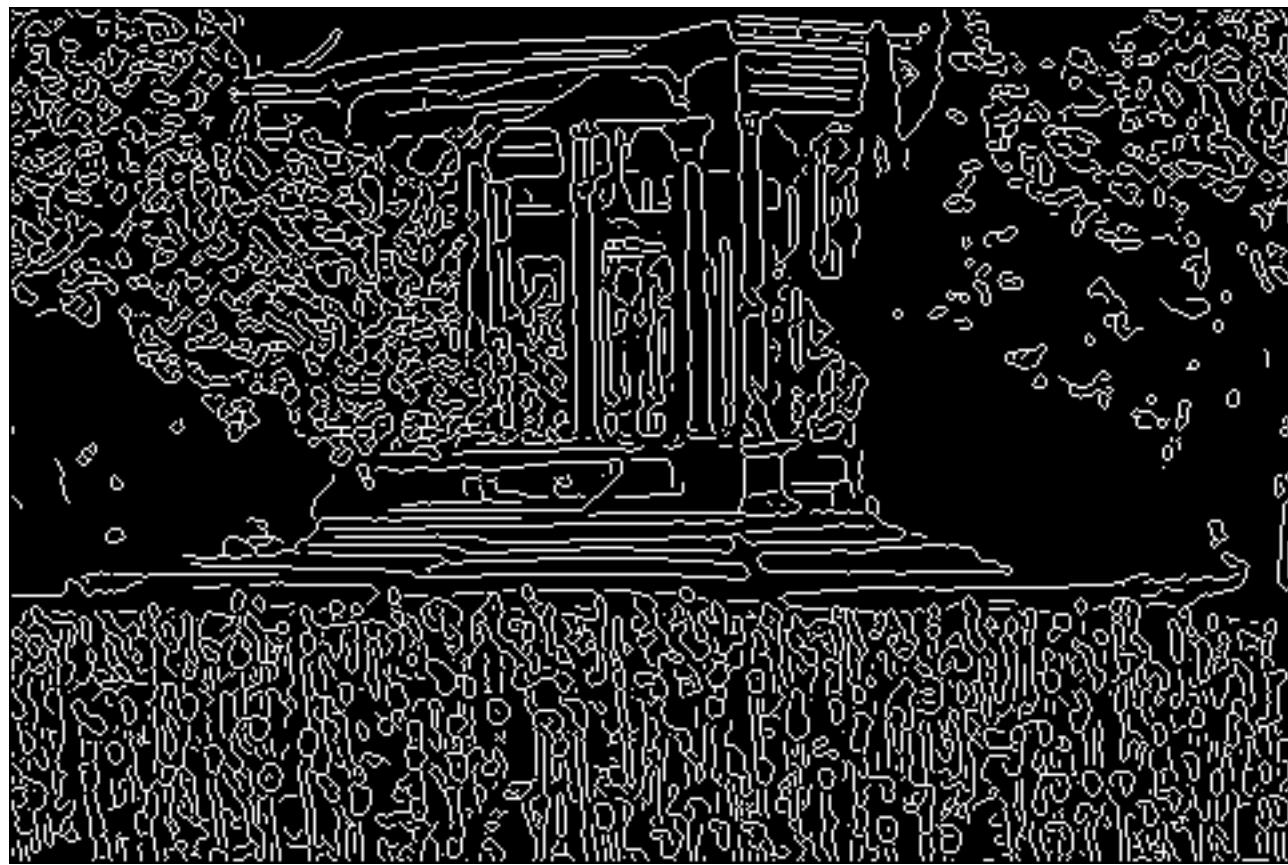


two-tone images









Need a more statistical representation



Texture as statistical image representation

- Narrowly: texture depicts spatially repeating patterns



radishes



rocks



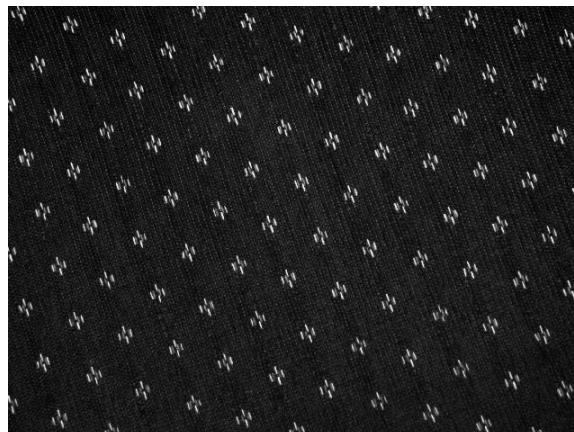
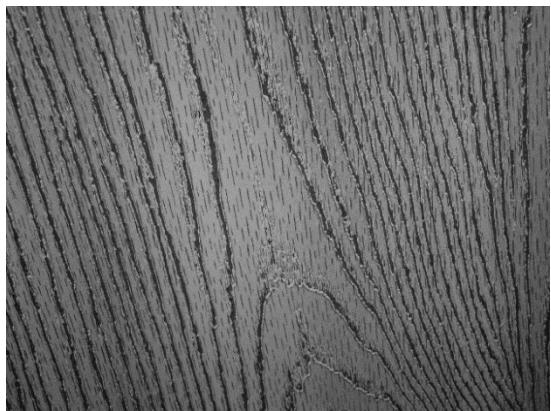
yogurt

Texture as “stuff”



Source: Forsyth

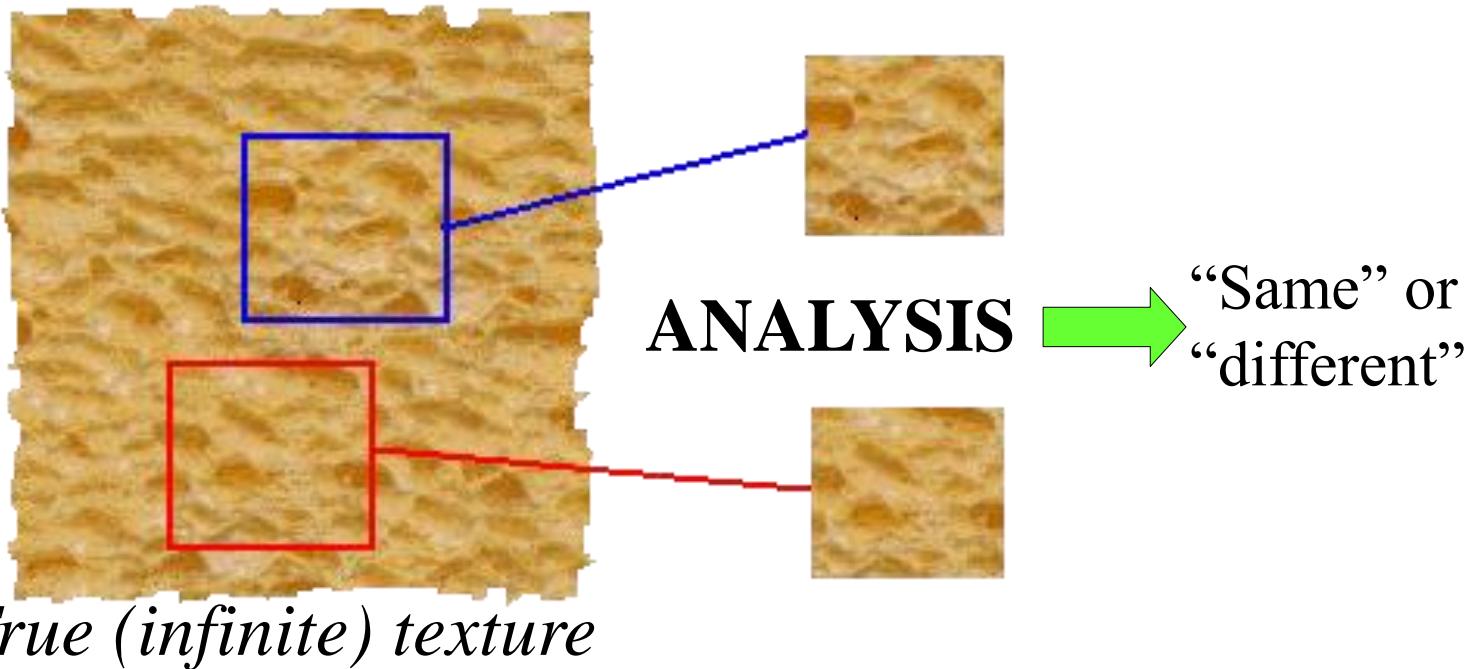
Texture and Material



Statistical, not exact, similarity

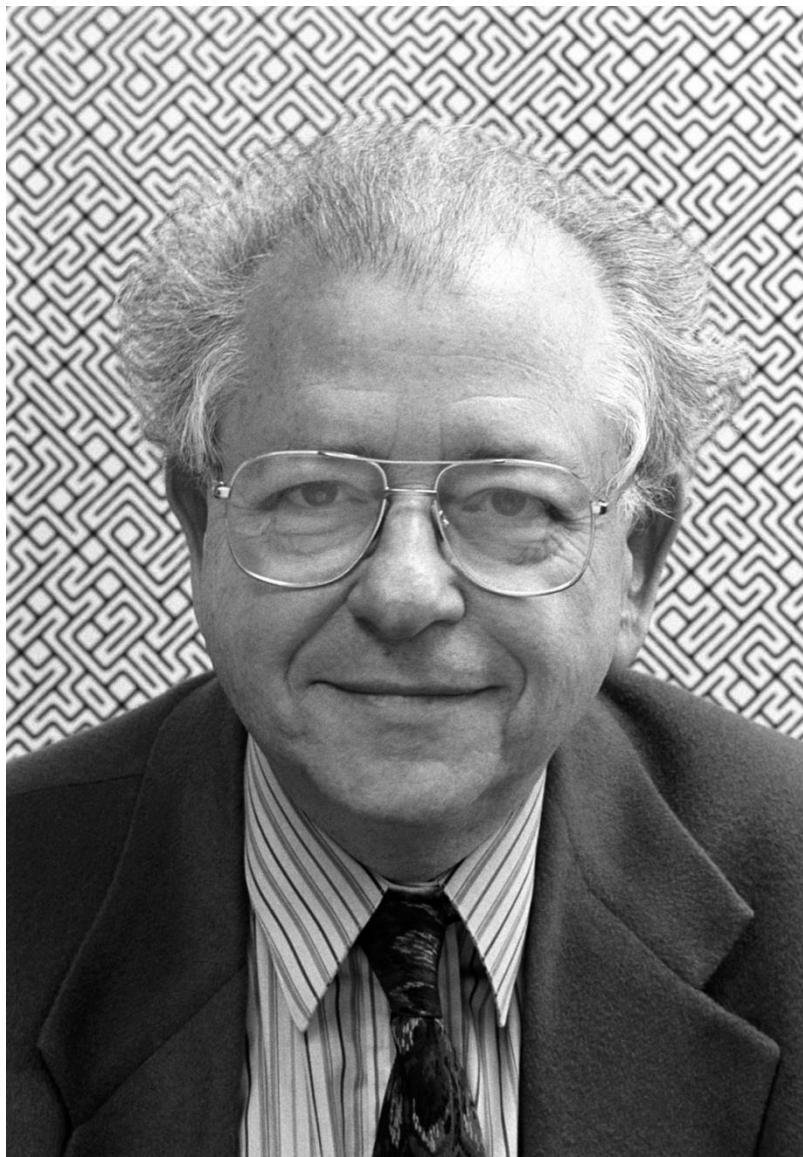


Texture Analysis

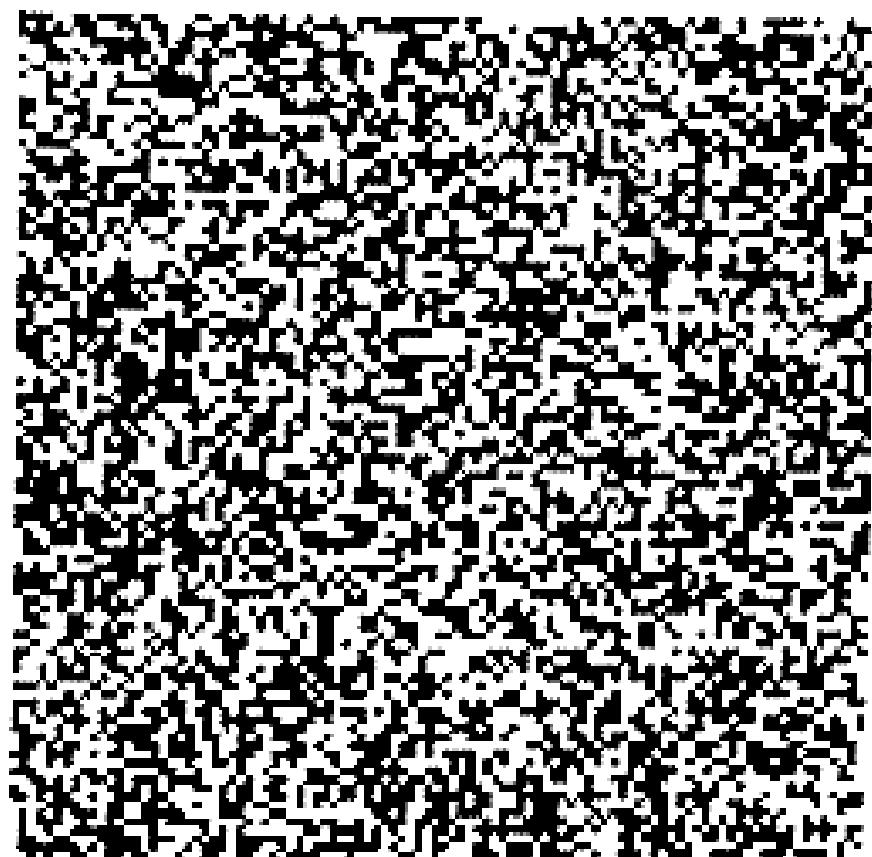
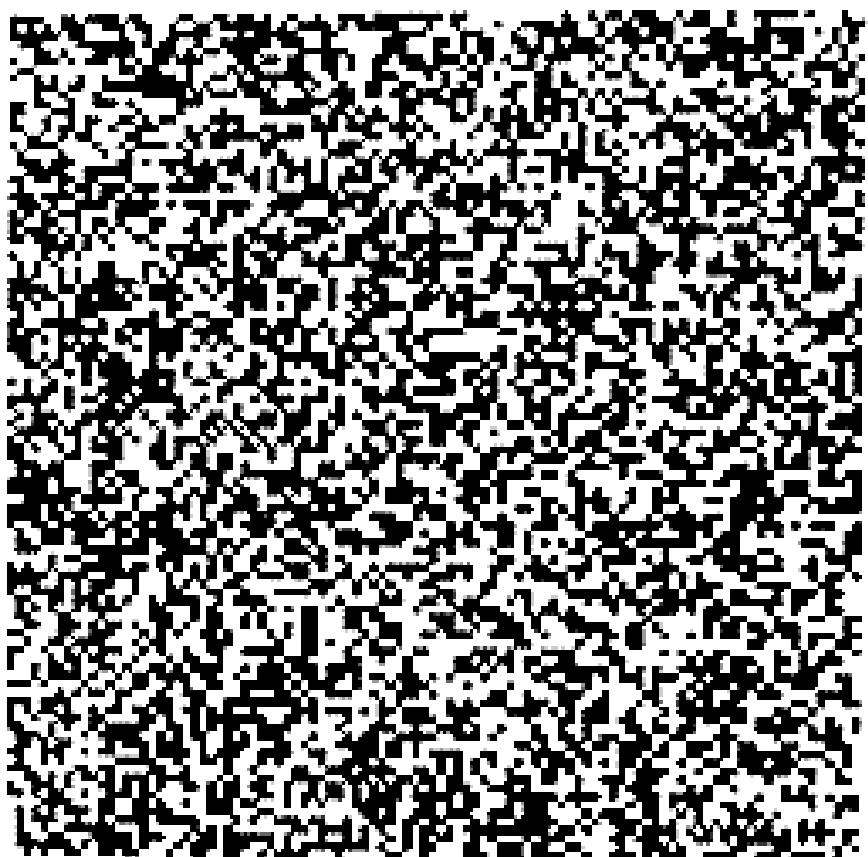


Compare textures and decide if they're made of the same “stuff”.

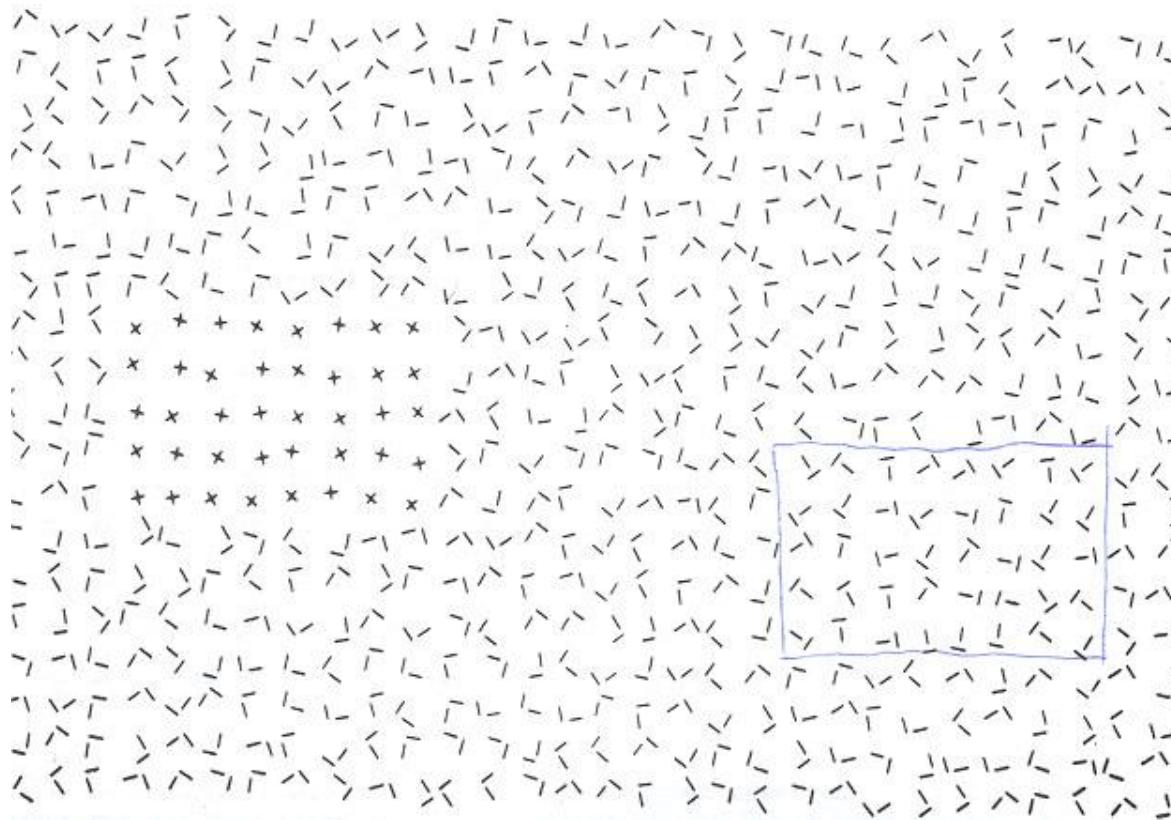
Béla Julesz, father of texture



Random Dot Stereograms

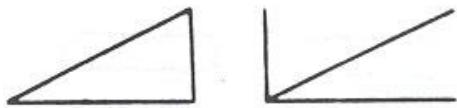


Texton Discrimination (Julesz)

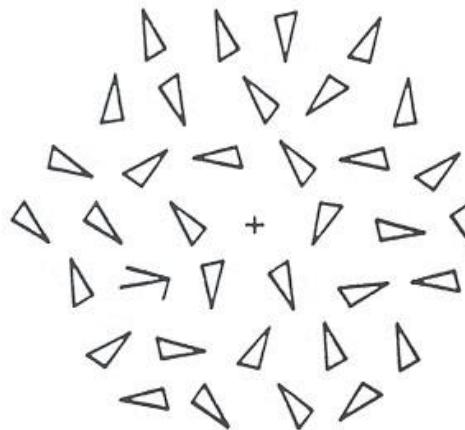
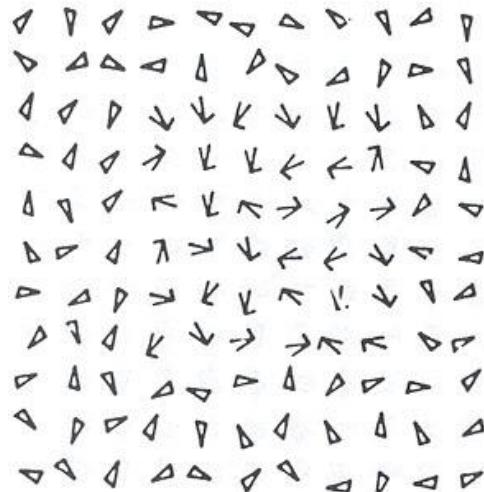
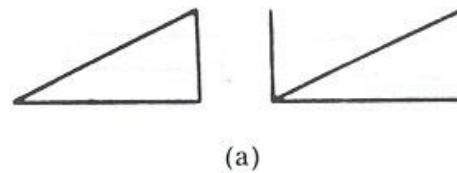


Human vision is sensitive to the difference of some types of elements and appears to be “numb” on other types of differences.

Search Experiment I

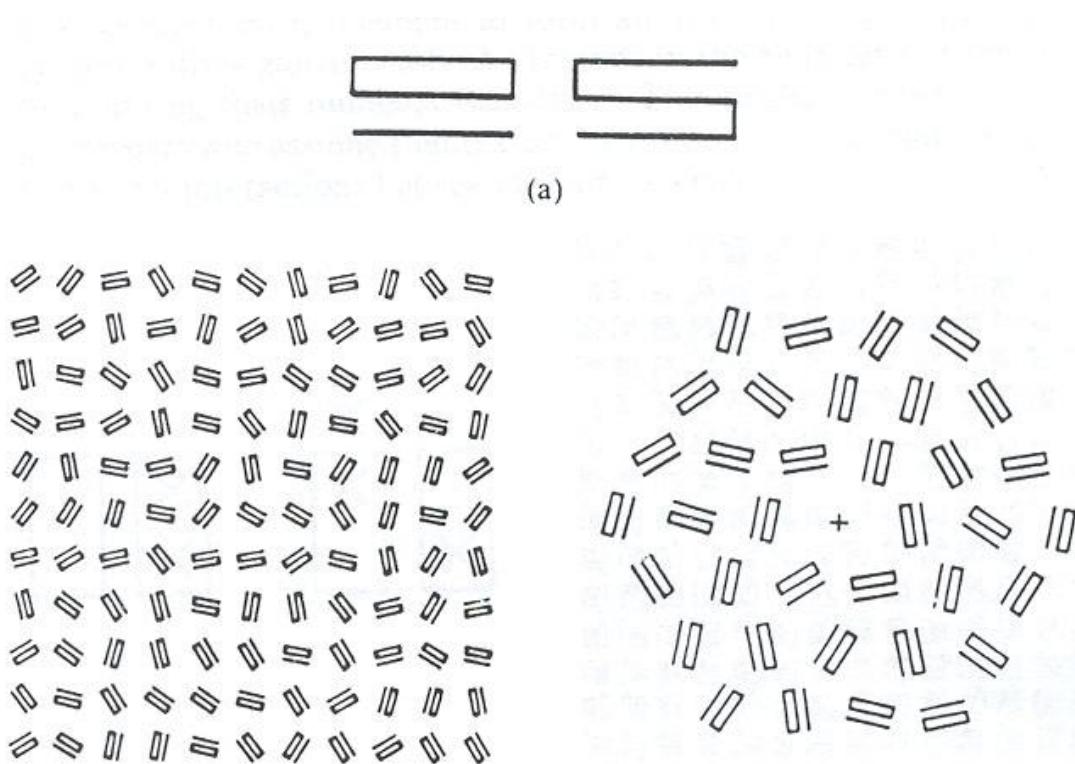


Search Experiment I



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

Search Experiment II



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

Pre-attentive vs. Attentive Vision

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.

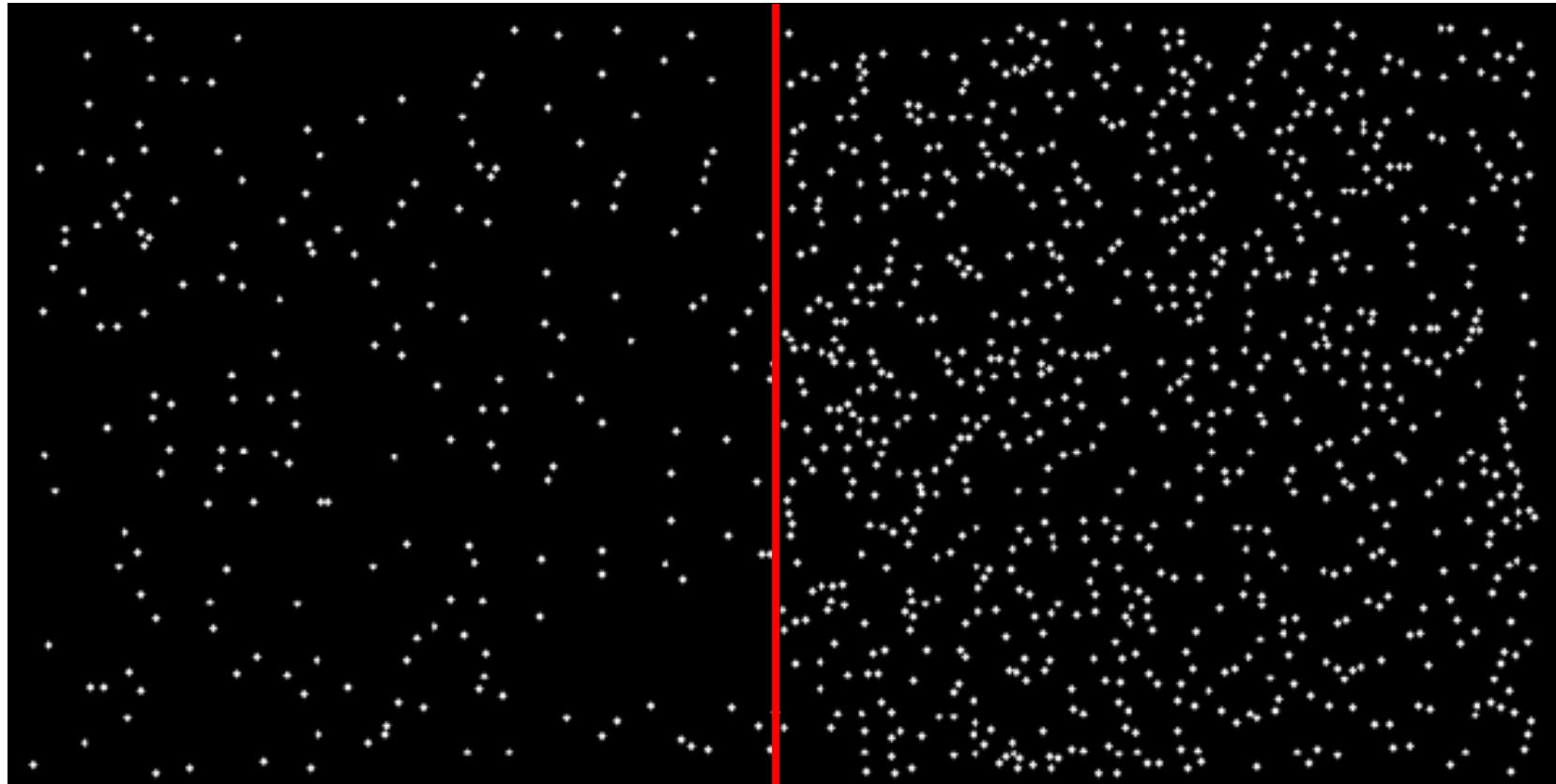
2. Attentive vision

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

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.*

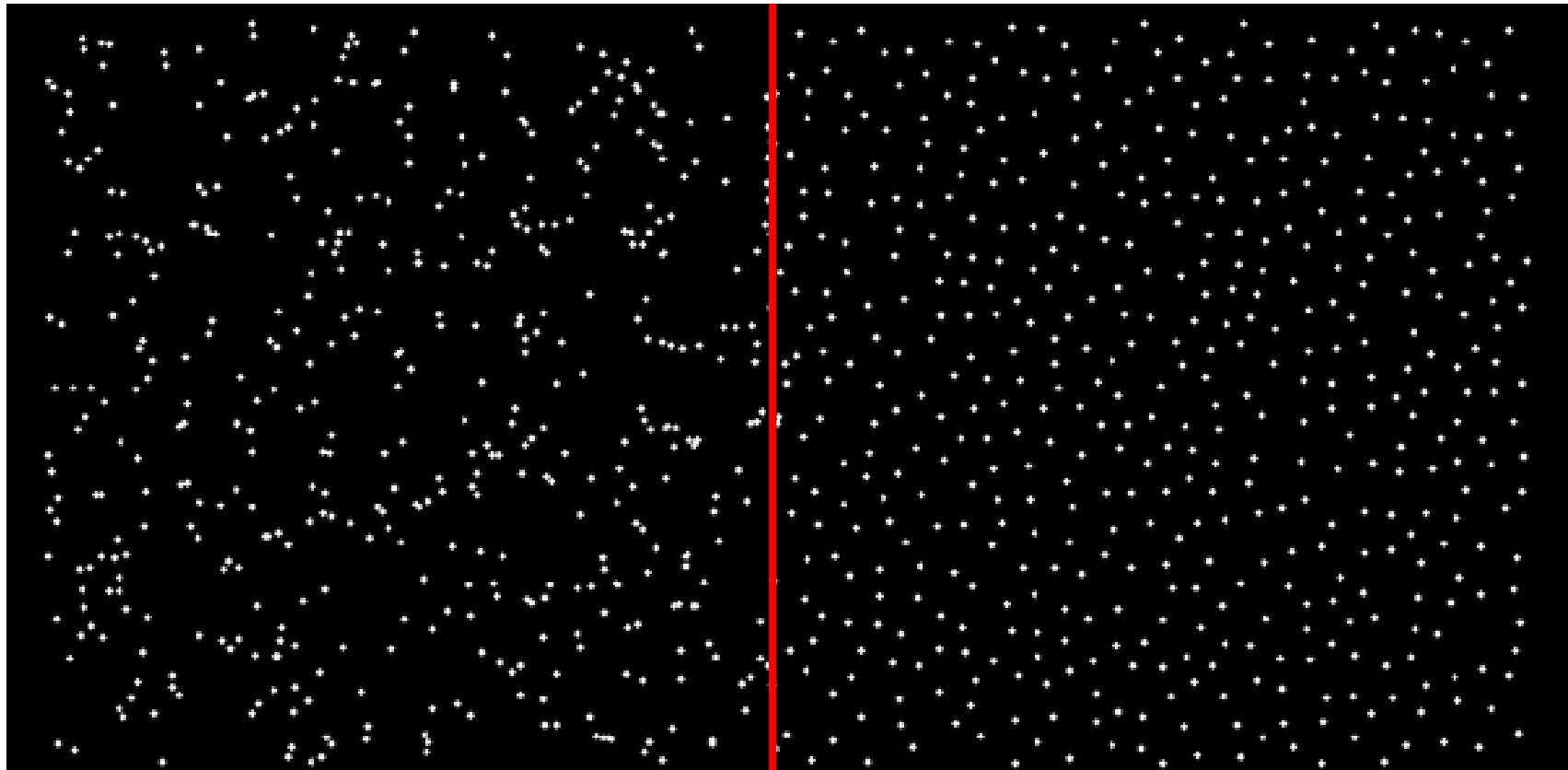
1st Order Statistics



5% white

20% white

2nd Order Statistics



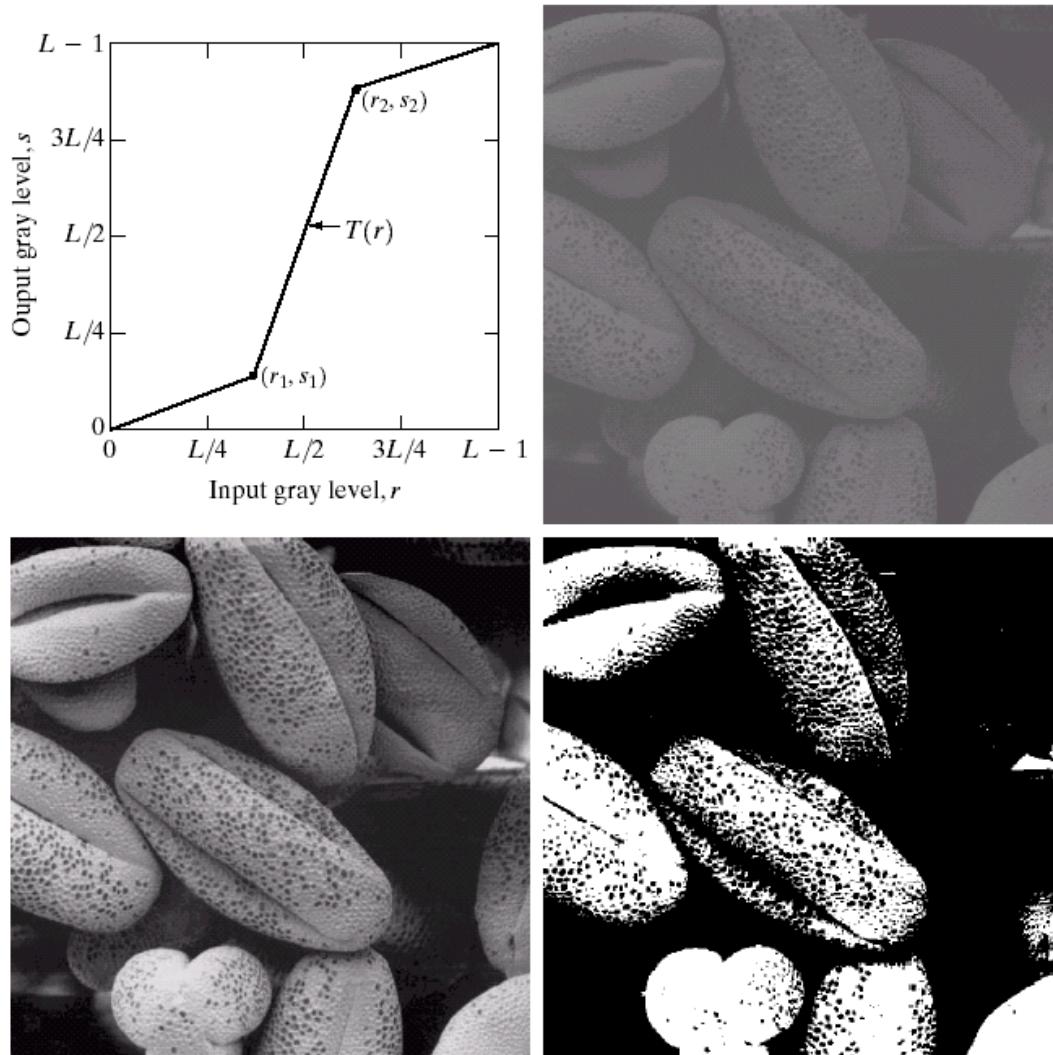
10% white

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

Histogram Review

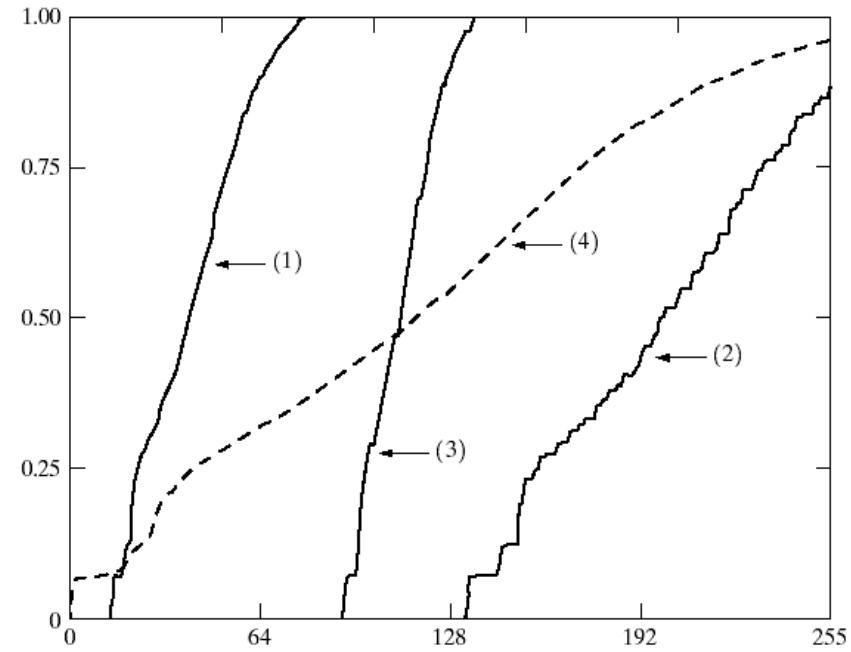
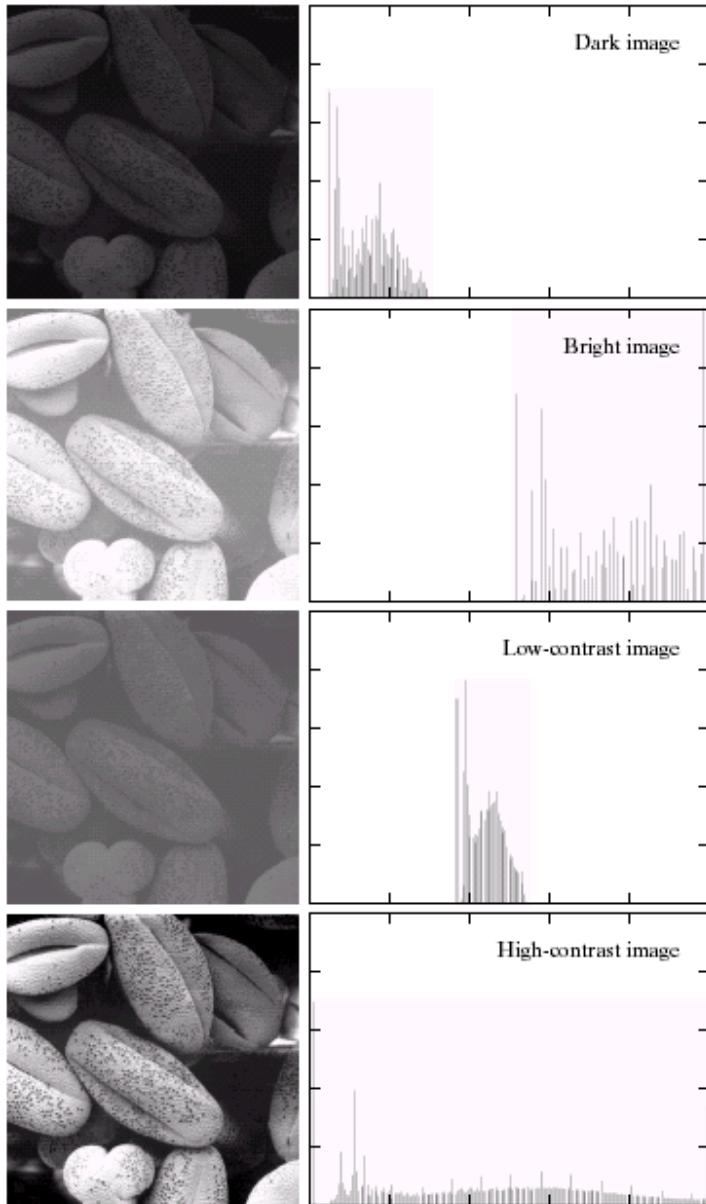
Contrast Stretching



a | b
c | d

FIGURE 3.10
Contrast stretching.
(a) Form of transformation function.
(b) A low-contrast image.
(c) Result of contrast stretching.
(d) Result of thresholding.
(Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)

Pixel Histograms



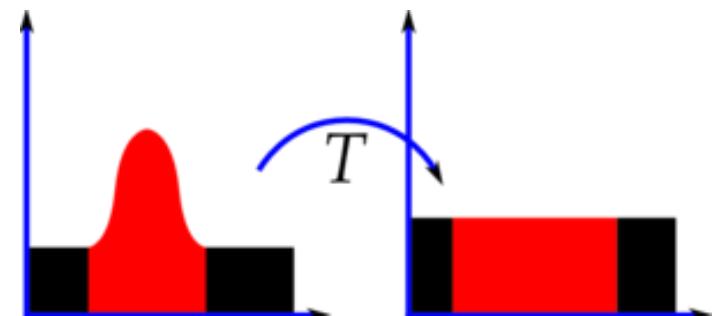
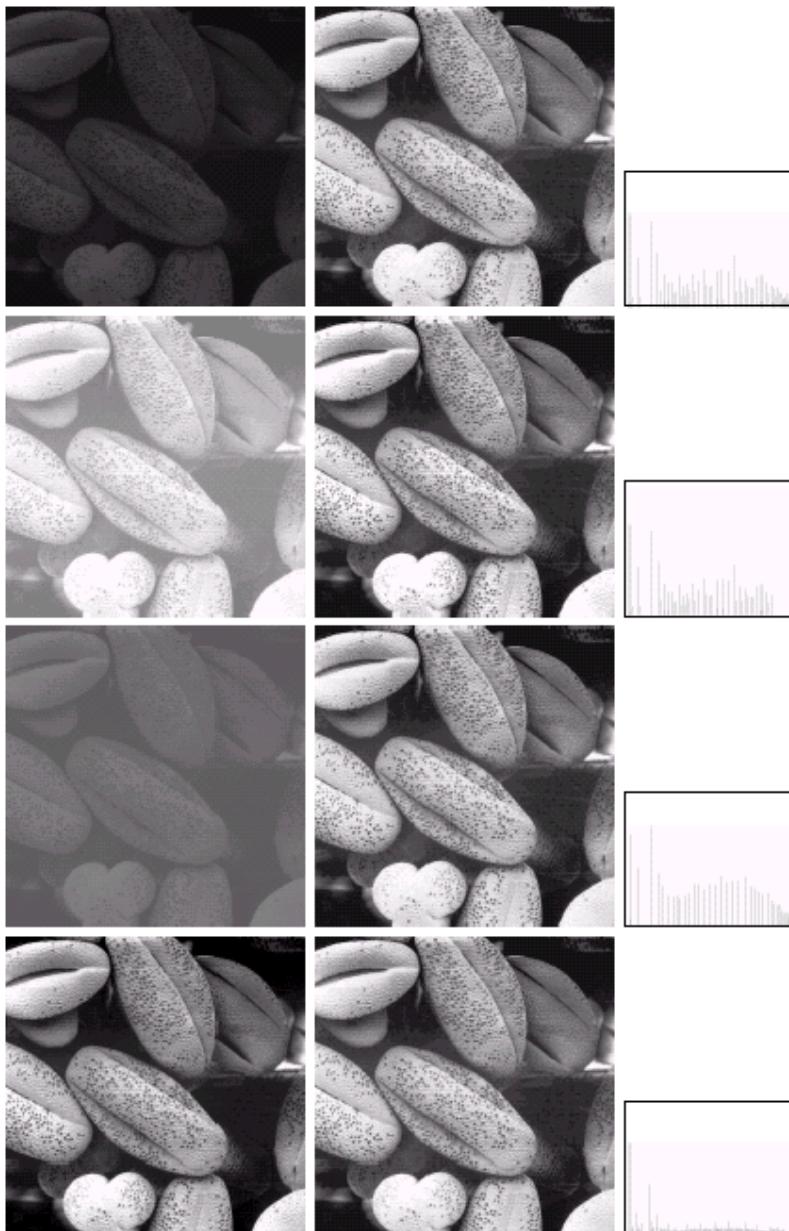
Cumulative Histograms

$$T(f(x, y))$$

a b

FIGURE 3.15 Four basic image types: dark, light, low contrast, high contrast, and their corresponding histograms. (Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)

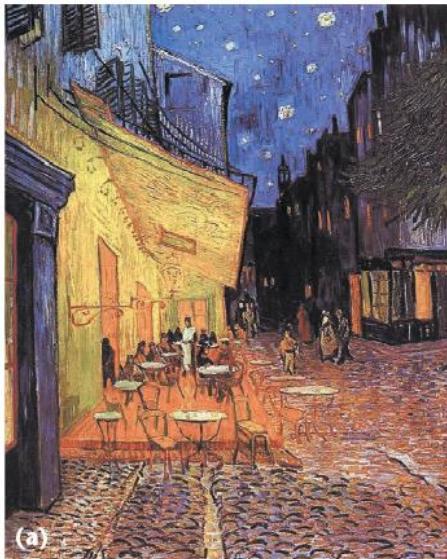
Histogram Equalization



a b c

FIGURE 3.17 (a) Images from Fig. 3.15. (b) Results of histogram equalization. (c) Corresponding histograms.

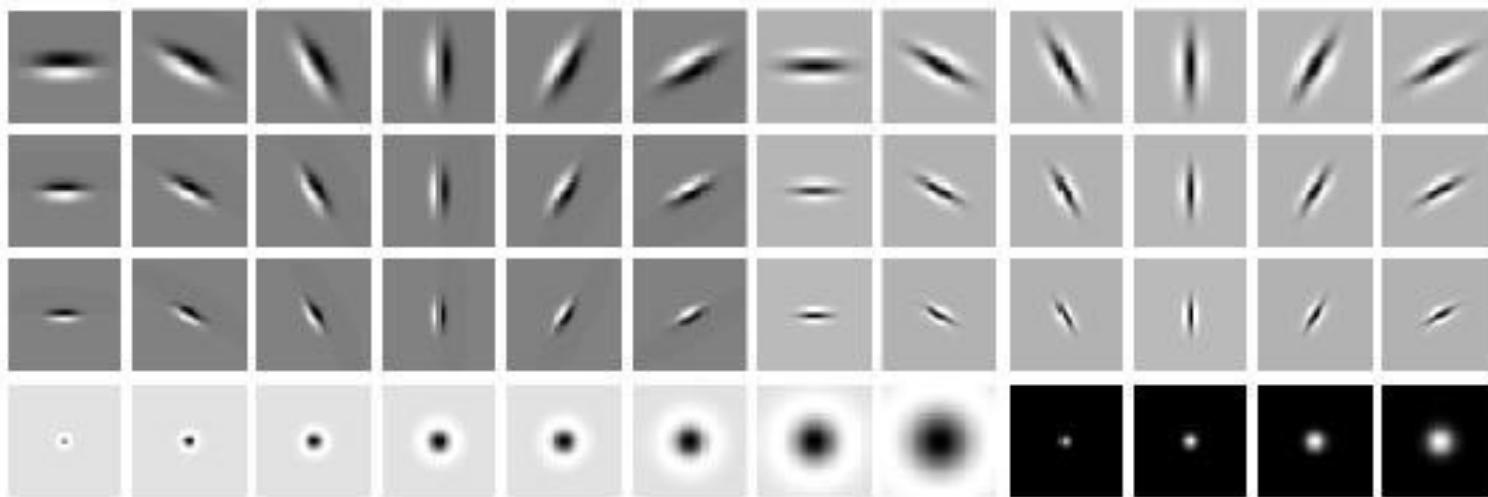
Color Transfer [Reinhard, et al, 2001]



Erik Reinhard, Michael Ashikhmin, Bruce Gooch, Peter Shirley, [Color Transfer between Images](#). *IEEE Computer Graphics and Applications*, 21(5), pp. 34–41. September 2001.

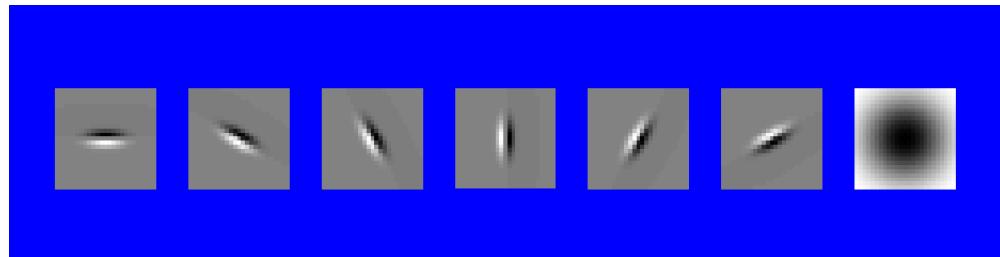
Going up from pixels: V1 filter-banks

LM Filter Bank

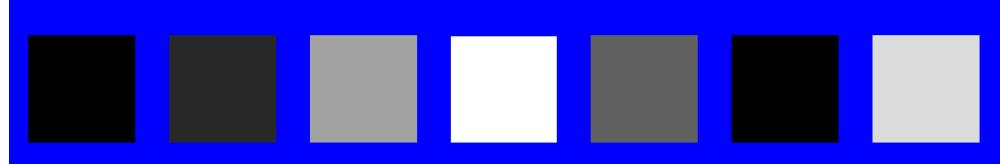


Can you match the texture to the response?

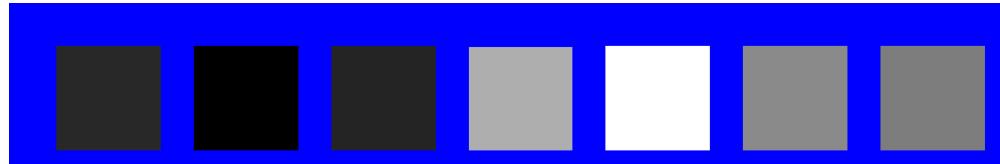
Filters



1



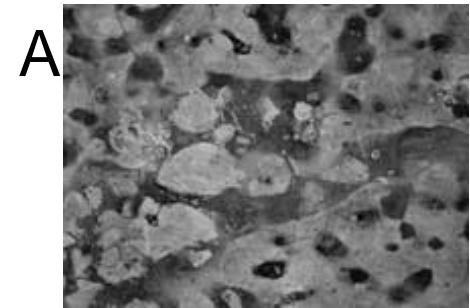
2



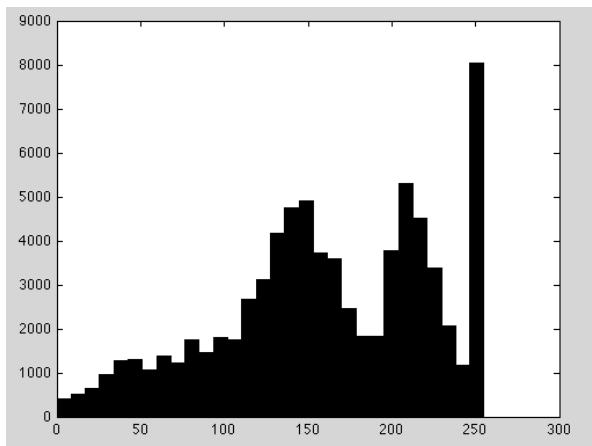
3



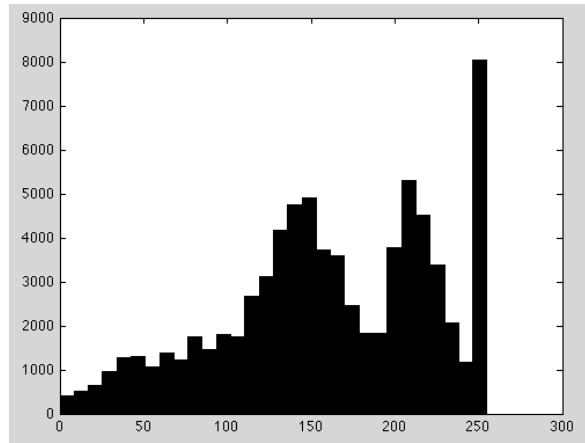
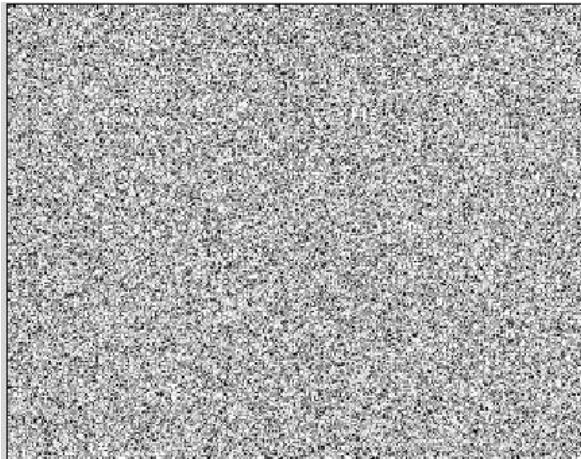
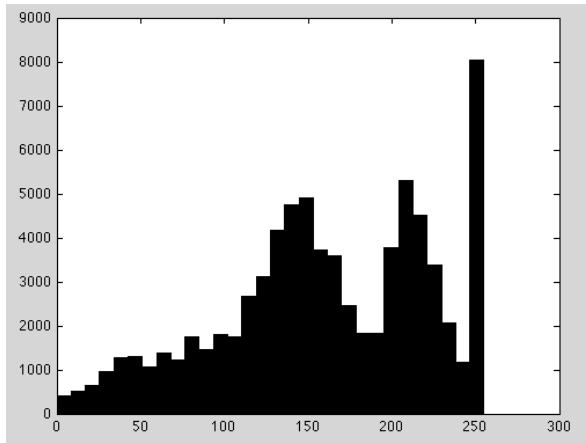
Mean abs
responses



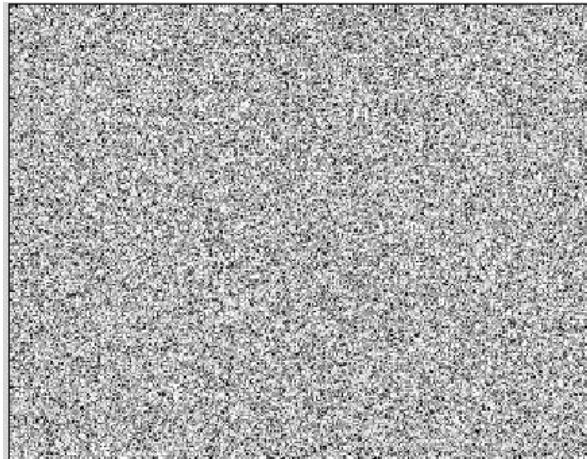
Pixel Histograms



Pixel Histograms

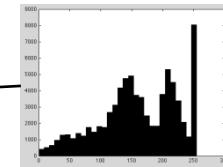
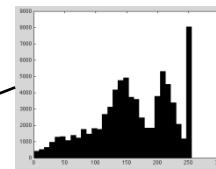
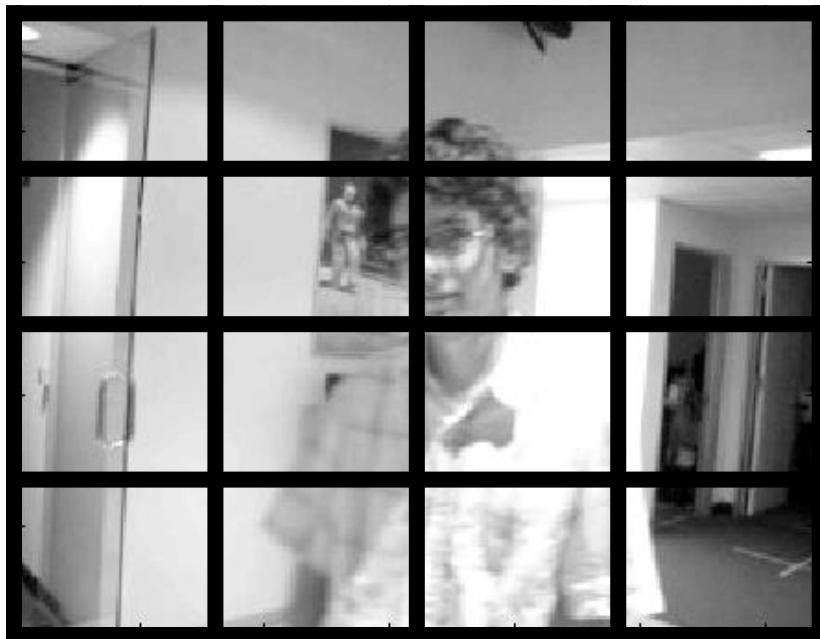


Gray value histogram comparisons



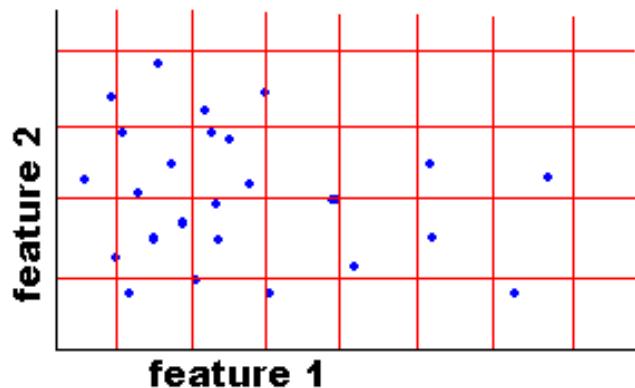
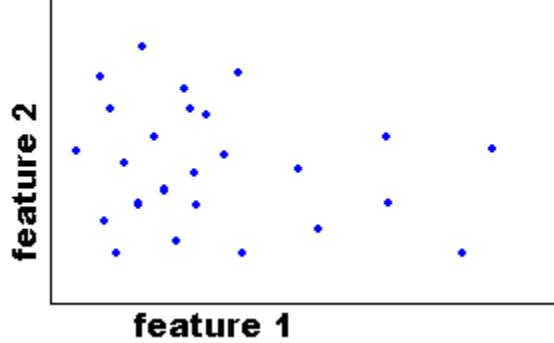
They're equal

Adding spatial structure



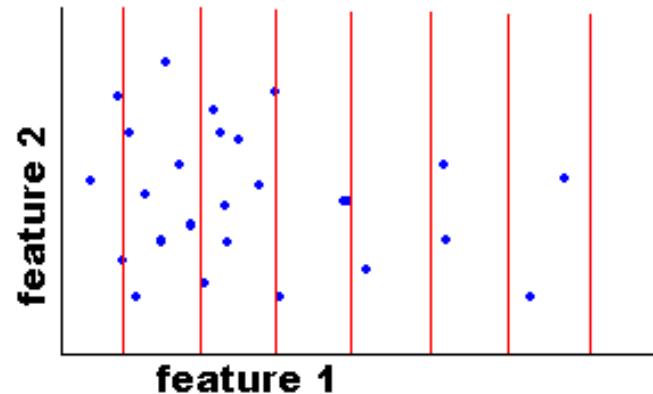
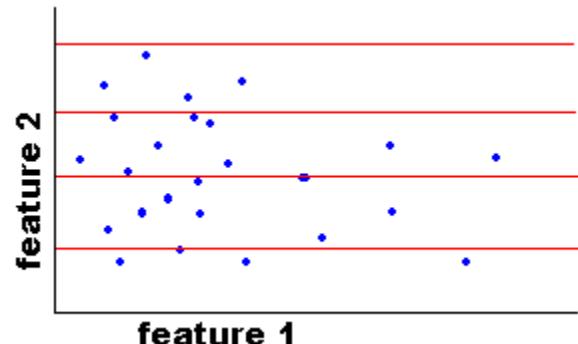
A separate histogram for each region.

Image Representations: Histograms



Joint histogram

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



Marginal histogram

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

Images from Dave Kauchak

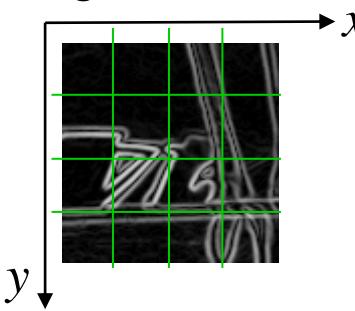
Ex: SIFT descriptor [Lowe'99]

distribution of the gradient over an image patch

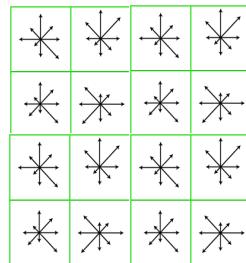
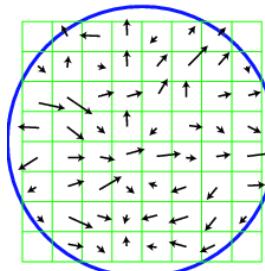
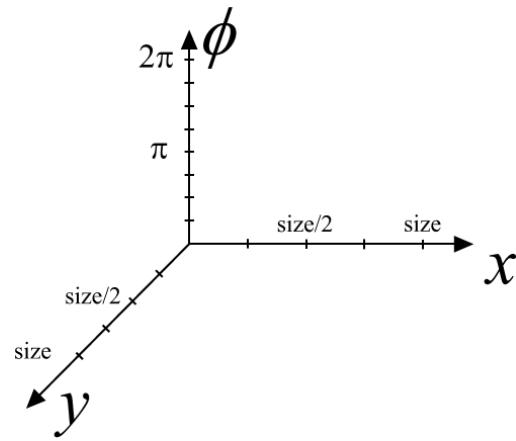
image patch



gradient



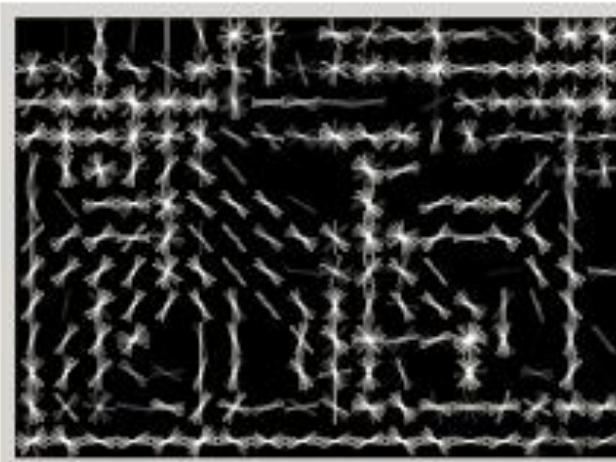
3D histogram



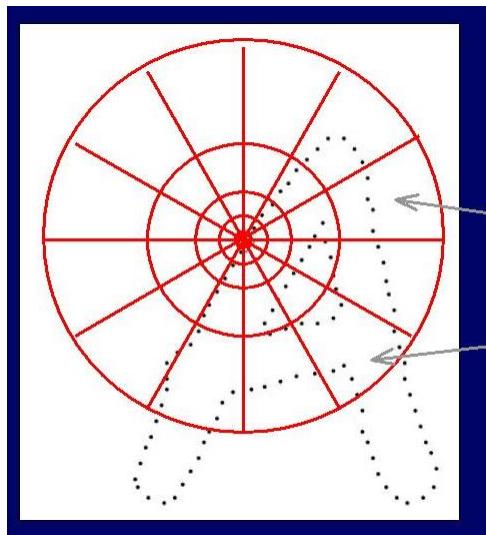
4x4 location grid and 8 orientations (128 dimensions)

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

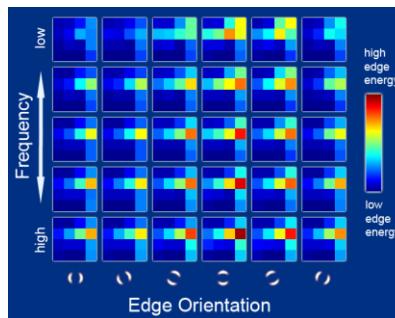
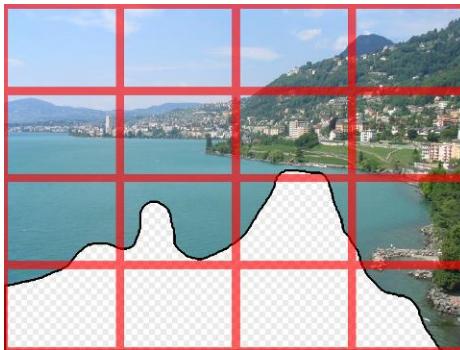
Gradient Histograms pop-up everywhere



HOG descriptor



Generalized Shape Context



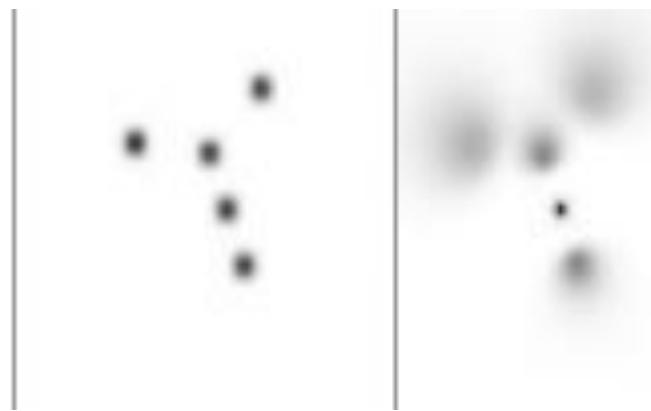
Gist Descriptor

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

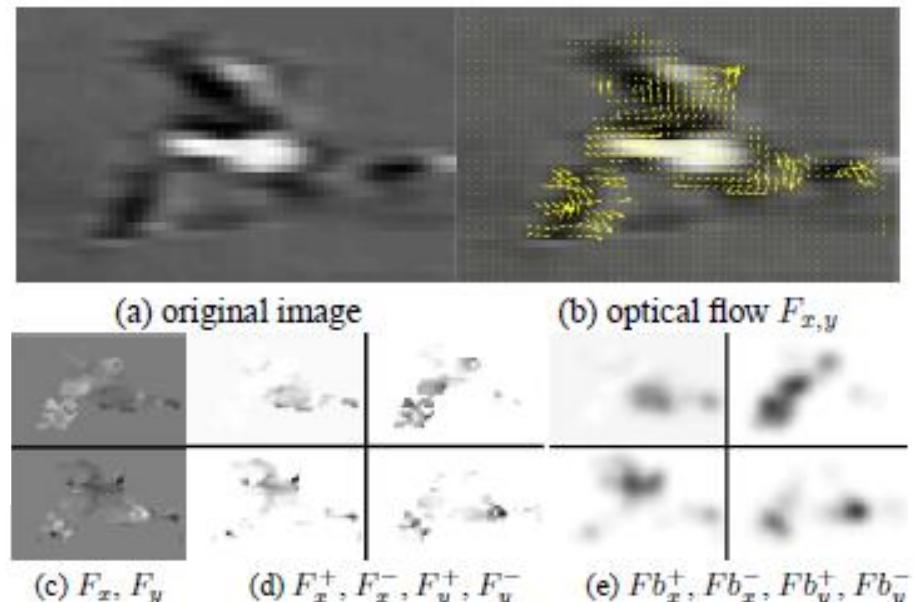
Binning achieves invariance to small patch offsets

How to add (a bit of) spatial invariance?

1. Define a spatial neighborhood
2. Compute statistics over the neighborhood by
 - Binning (spatial histograms)
 - Averaging (blurring)
 - Subsampling (e.g. average-pool, max-pool)

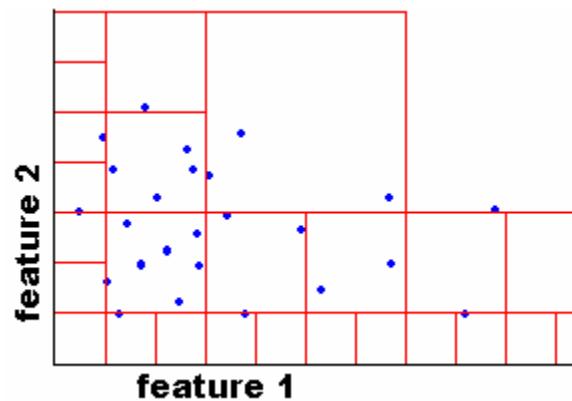
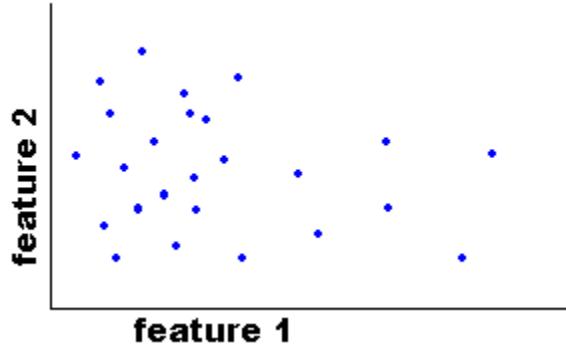


Geometric Blur (Berg et al)

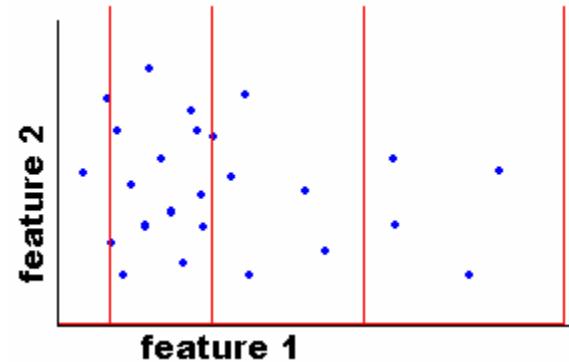
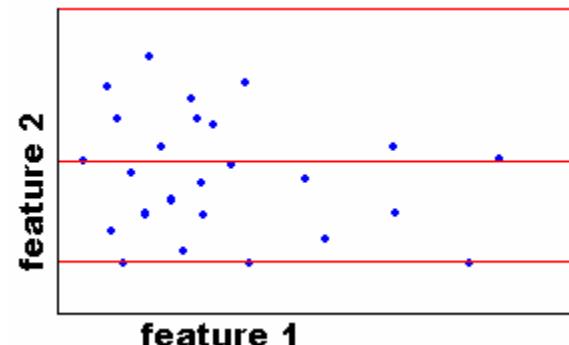


Motion Descriptor (Efros et al.)

Adaptive Representations



Images from Dave Kauchak

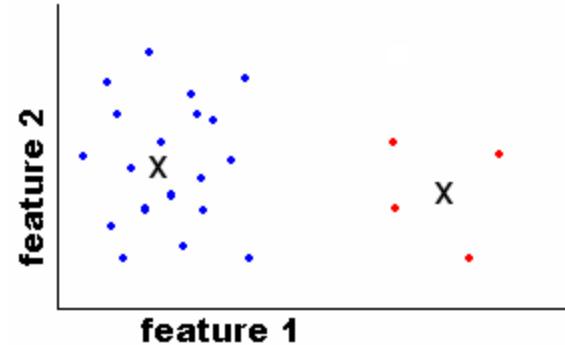
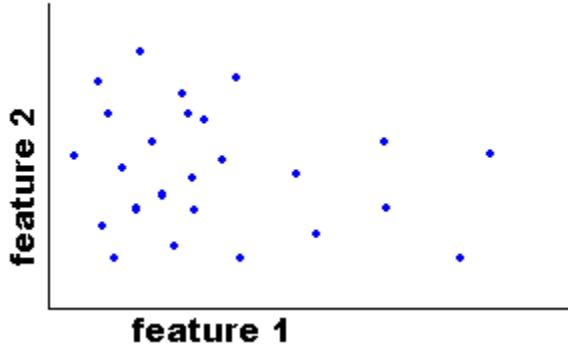


Adaptive binning

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

Clustering: very adaptive representations

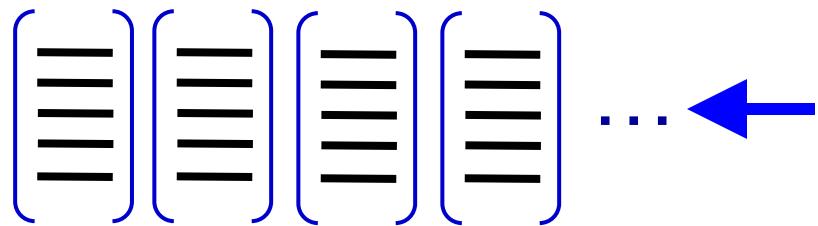
Images from Dave Kauchak



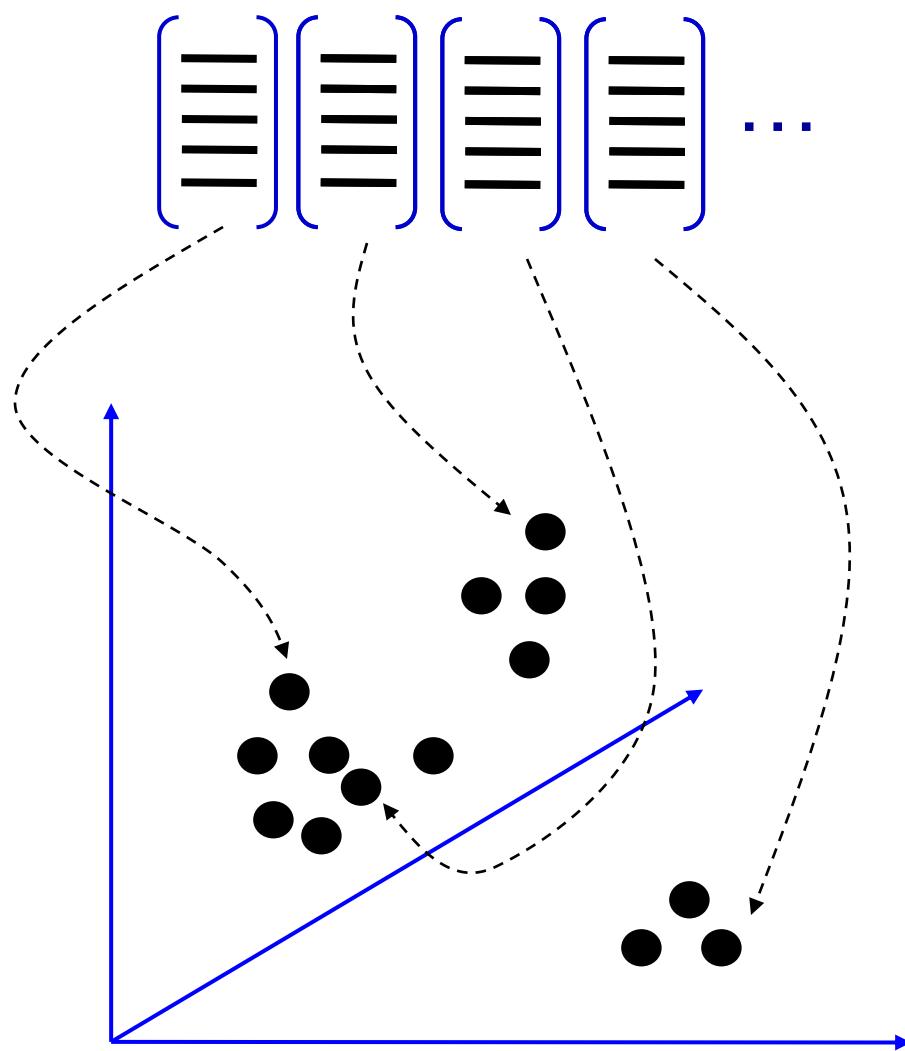
Clusters / Signatures

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

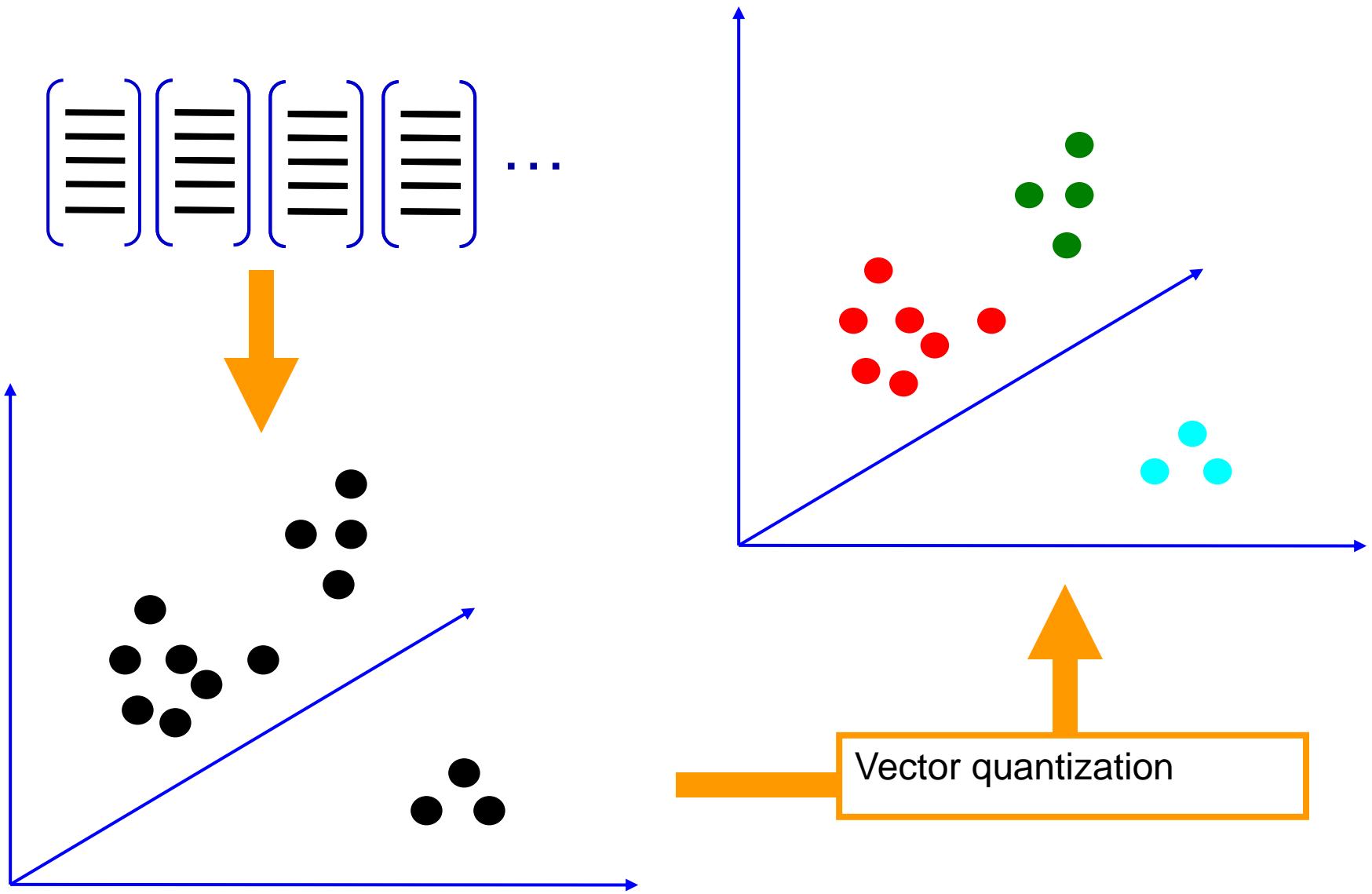
Patch Features



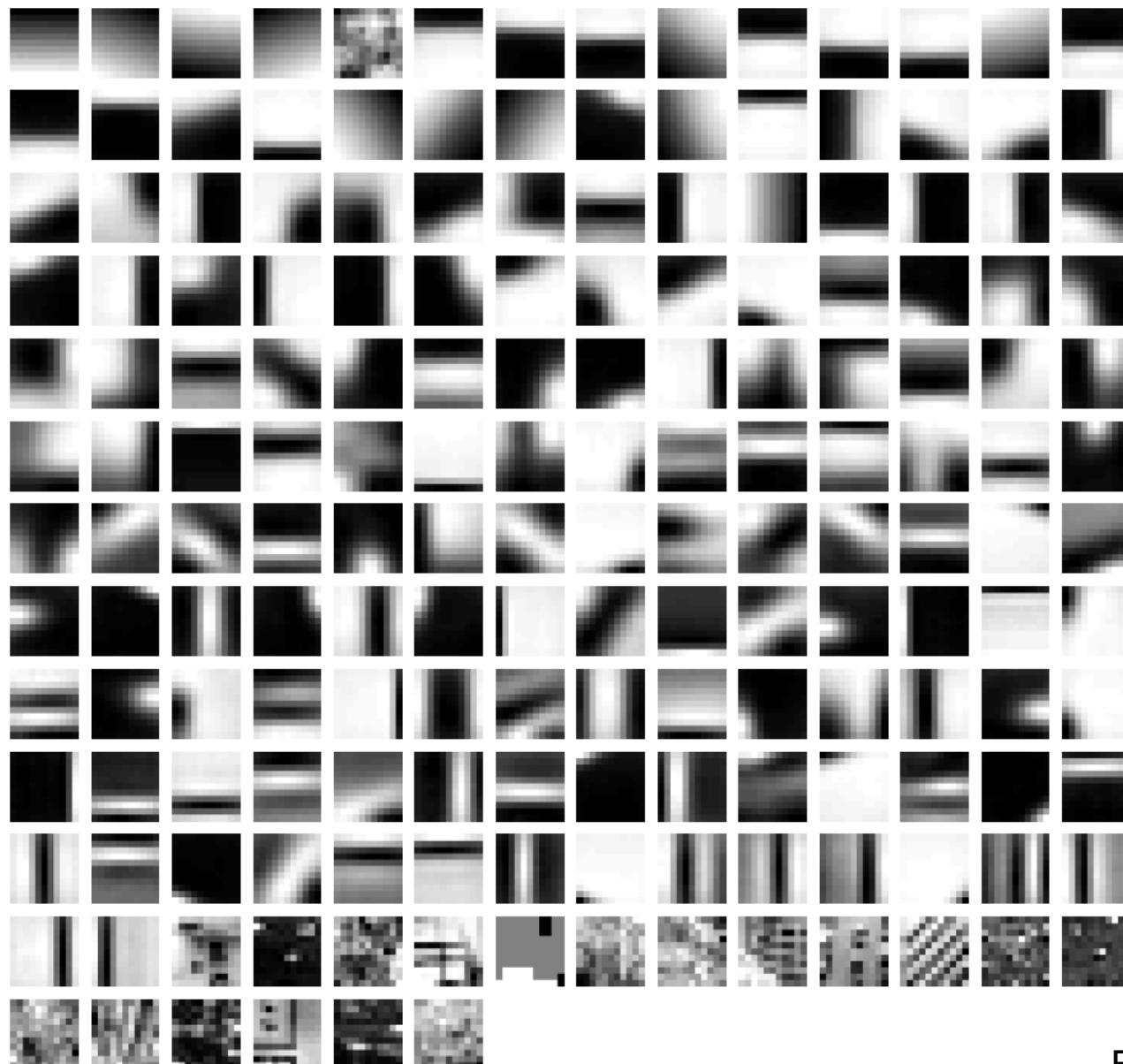
dictionary formation



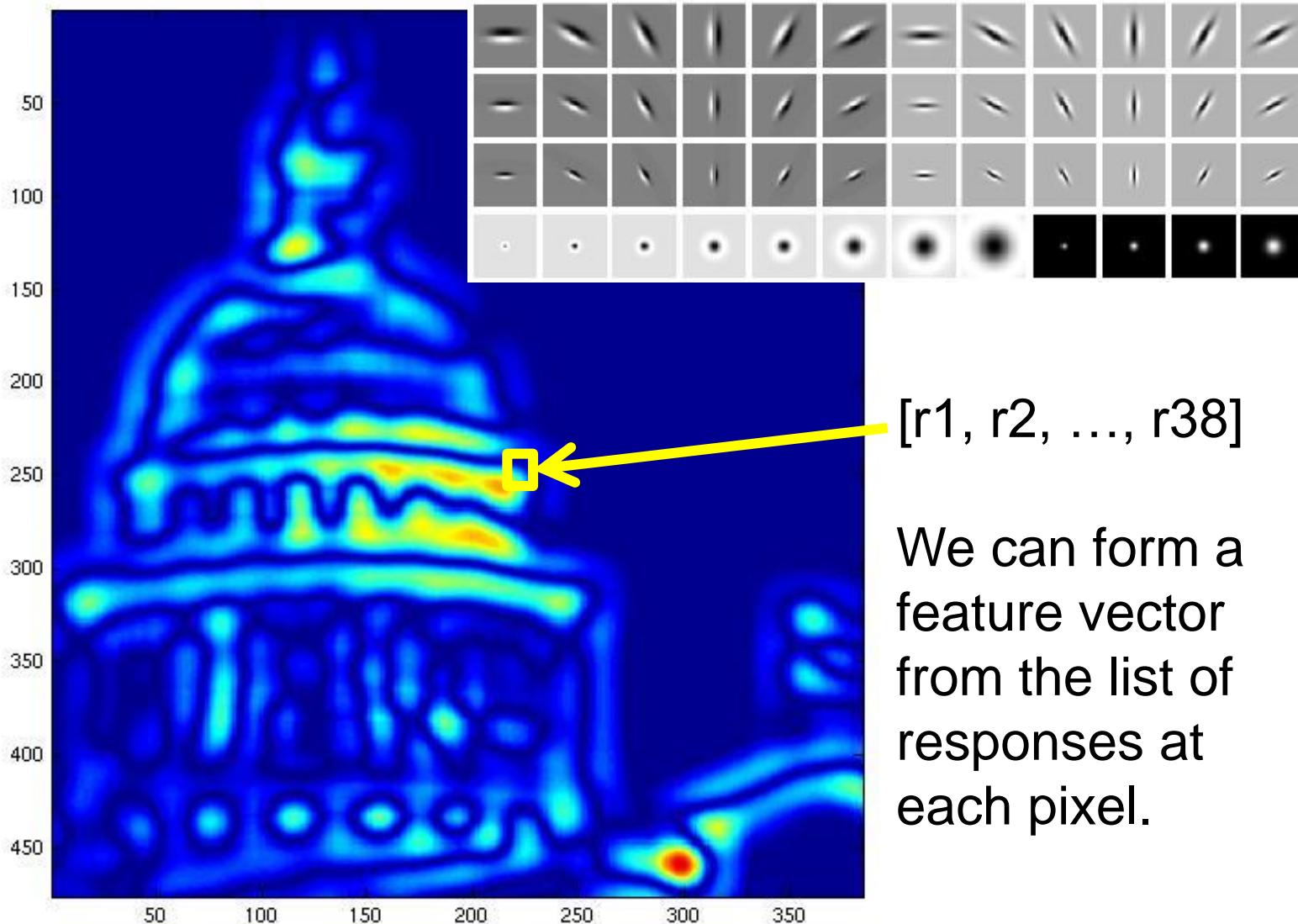
Clustering (usually k-means)



Clustered Image Patches



Filter Response Vectors

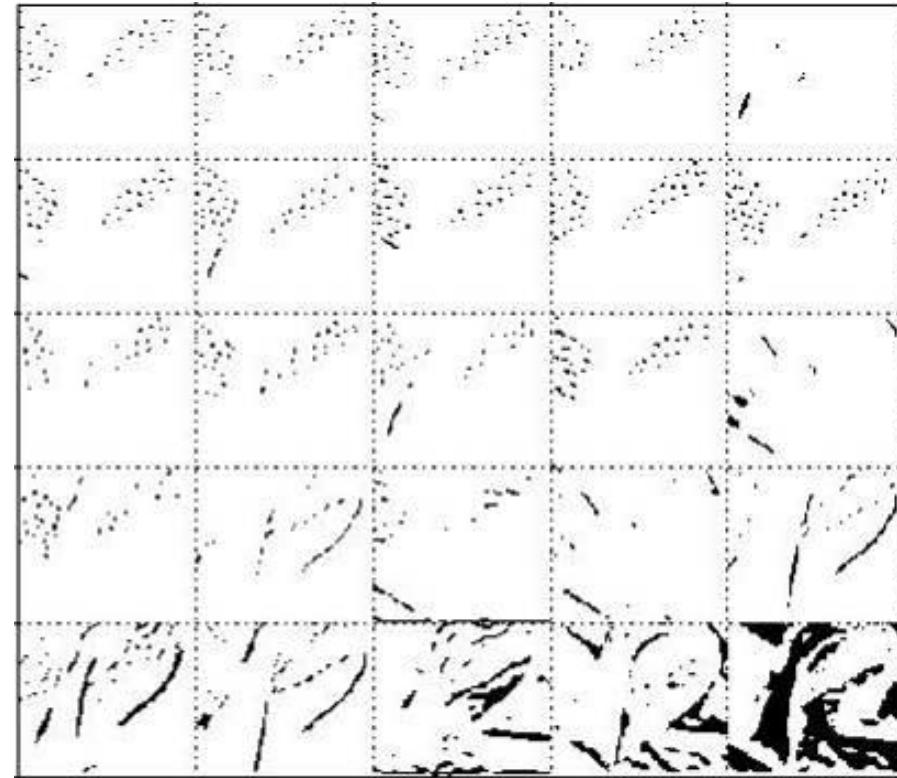
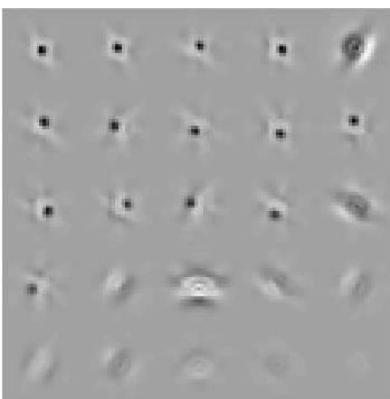


$$[r_1, r_2, \dots, r_{38}]$$

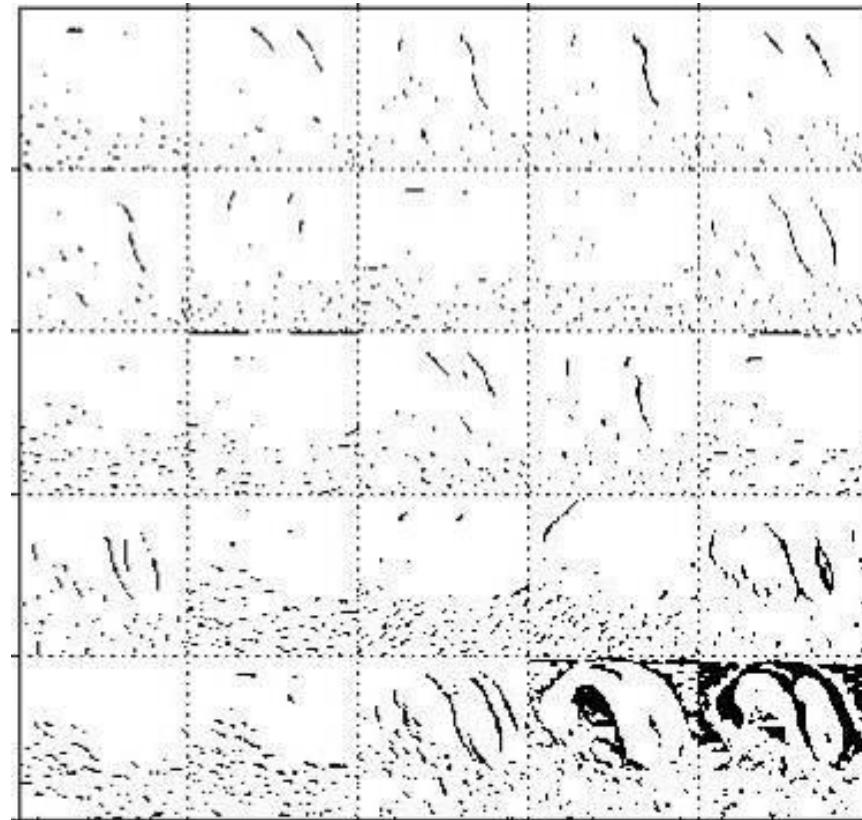
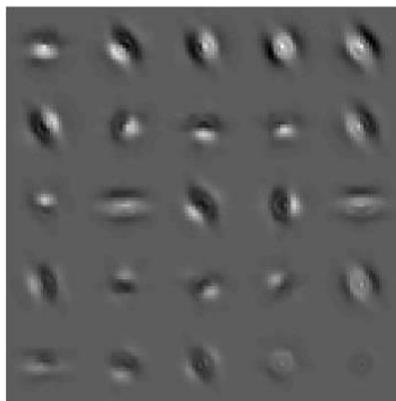
We can form a feature vector from the list of responses at each pixel.

~~Textons~~ (Malik et al, IJCV 2001)

Cluster vectors of filter responses



Textons (cont.)



“bag of words”: cluster histograms

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach our brain from our eyes. For a long time it was believed that the retinal image was processed directly in the visual centers in the cerebral cortex. It was a movie screen on which the image was projected. In 1960, however, two American scientists, David Hubel and Torsten Wiesel, discovered that the visual system is more complex than that. Following the flow of the optic nerve through the brain to the various cortical areas of the cerebral cortex, Hubel and Wiesel have shown that the message about the image falling on the retina undergoes a top-down analysis. The image is broken down into its component parts and analyzed in a top-down, column-wise analysis in a system of nerve cells. The results of this analysis are stored in columns. In this system each column has its specific function and is responsible for a specific detail in the pattern of the retinal image.

**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% increase in exports to \$750bn, compared with \$660bn. The ministry also said it would annoy the US by continuing to let the Chinese yuan's value rise. China's central bank, the People's Bank, deliberately agreed to let the yuan rise against the dollar. The Chinese government also needs to encourage domestic demand so that the country can buy more from the US. China has been allowed to let the yuan rise against the dollar, and permitted it to trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

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

Object

Bag of ‘words’



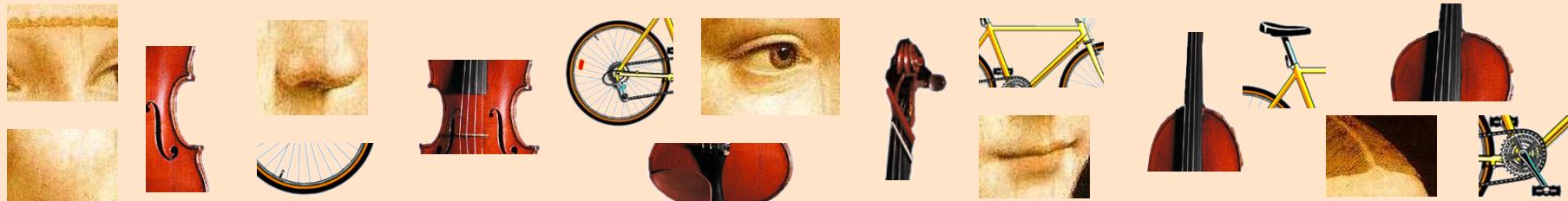
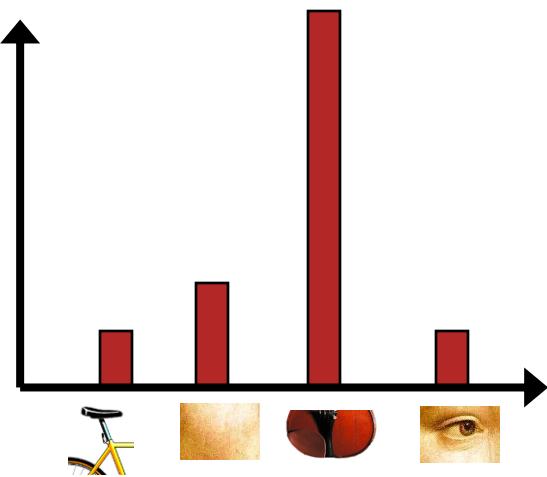
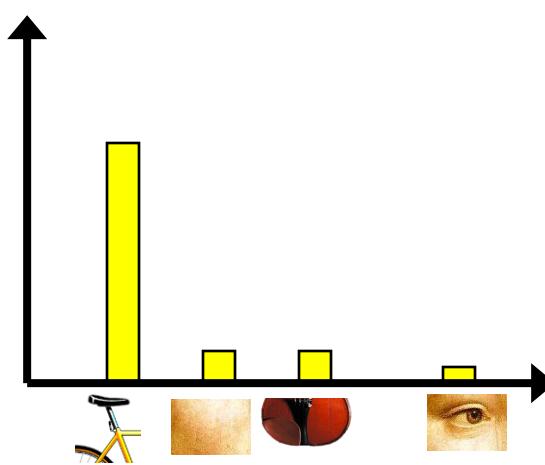
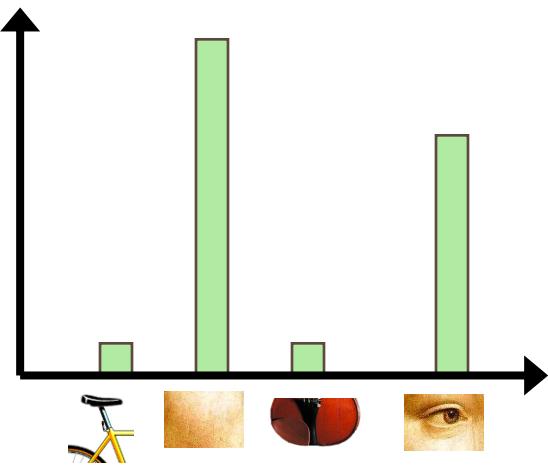
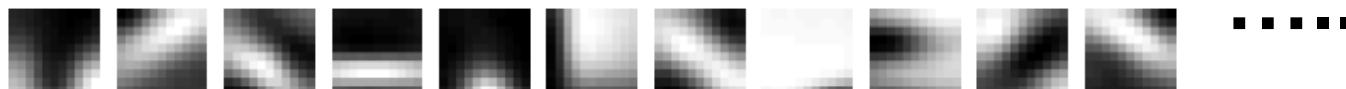
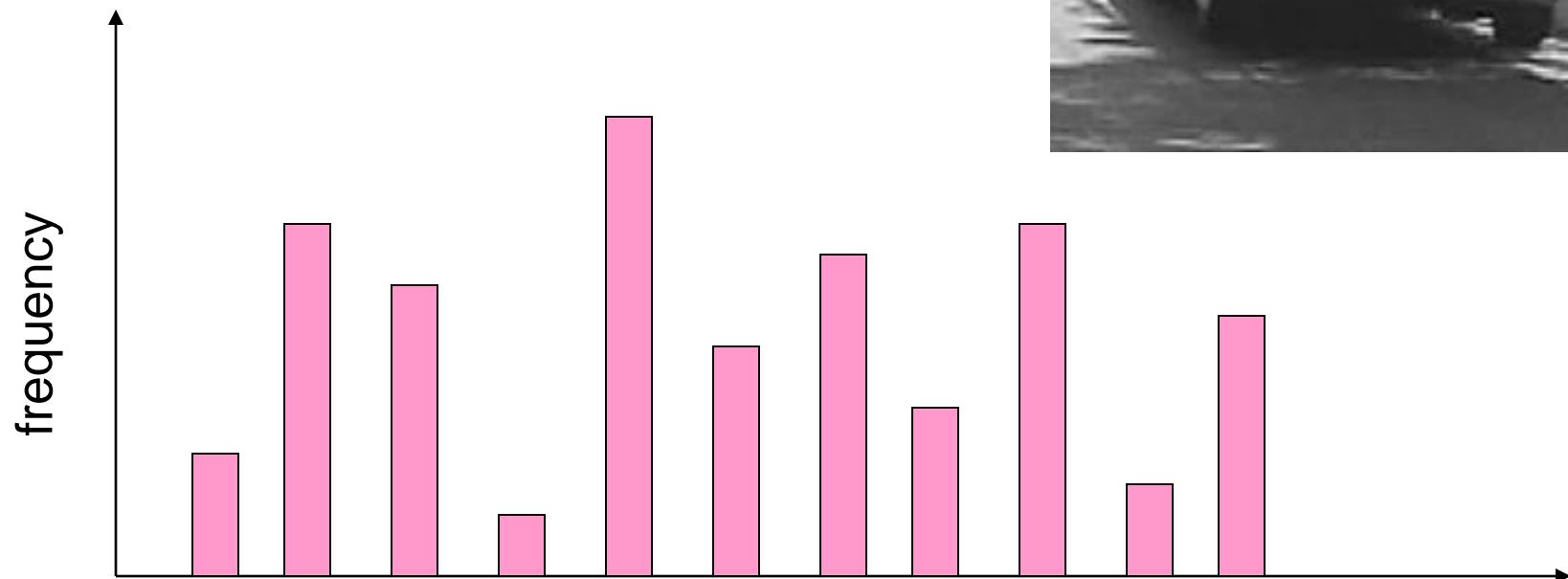


Image representation



codewords

Scene Classification (Renninger & Malik)

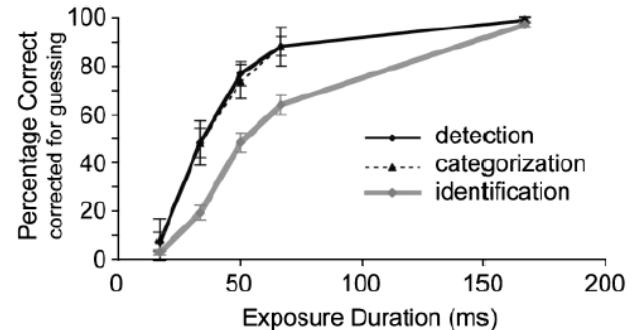
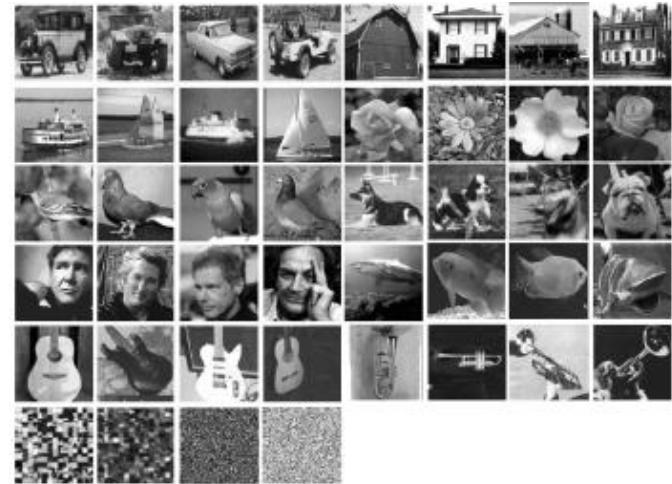


Image classification can be pre-attentive!

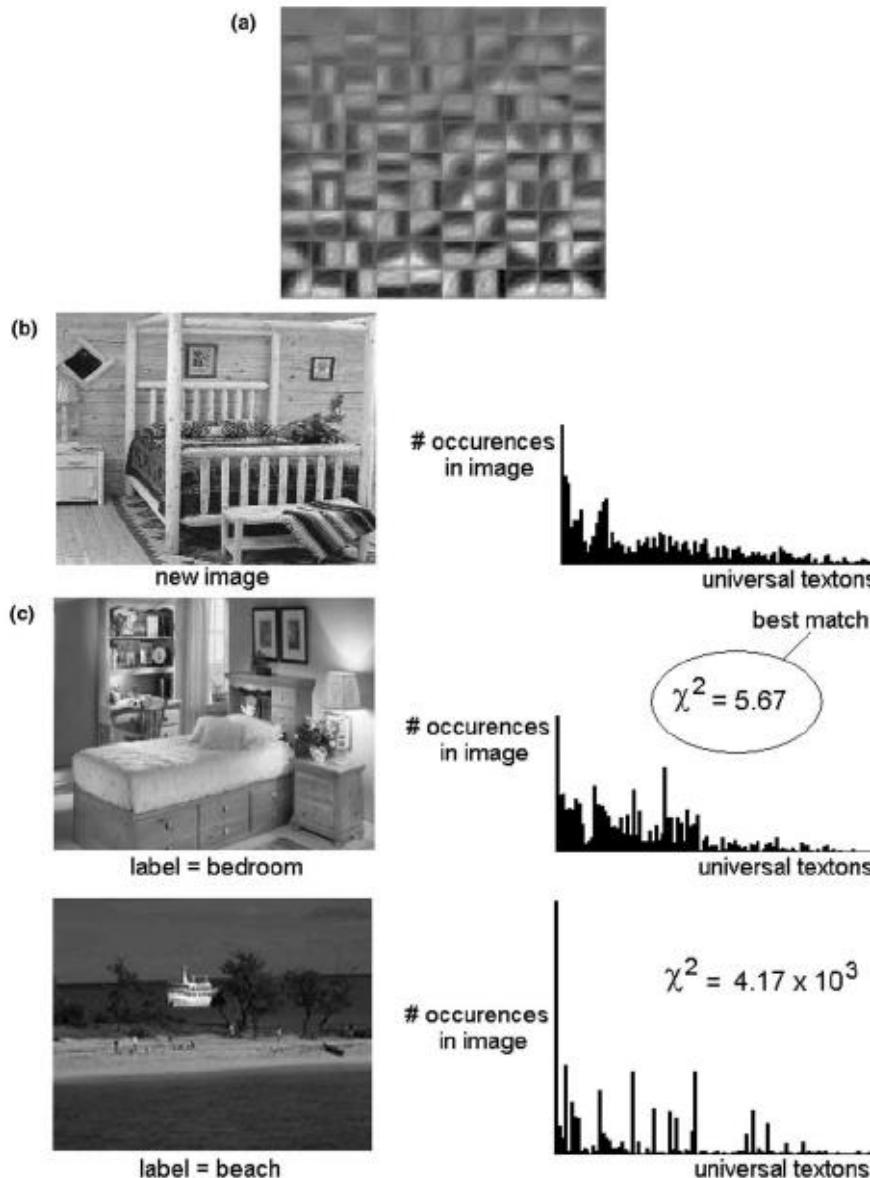
On a task of judging animal vs no animal, humans can make mostly correct saccades in 150 ms
(Kirchner & Thorpe, 2006)

- 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**

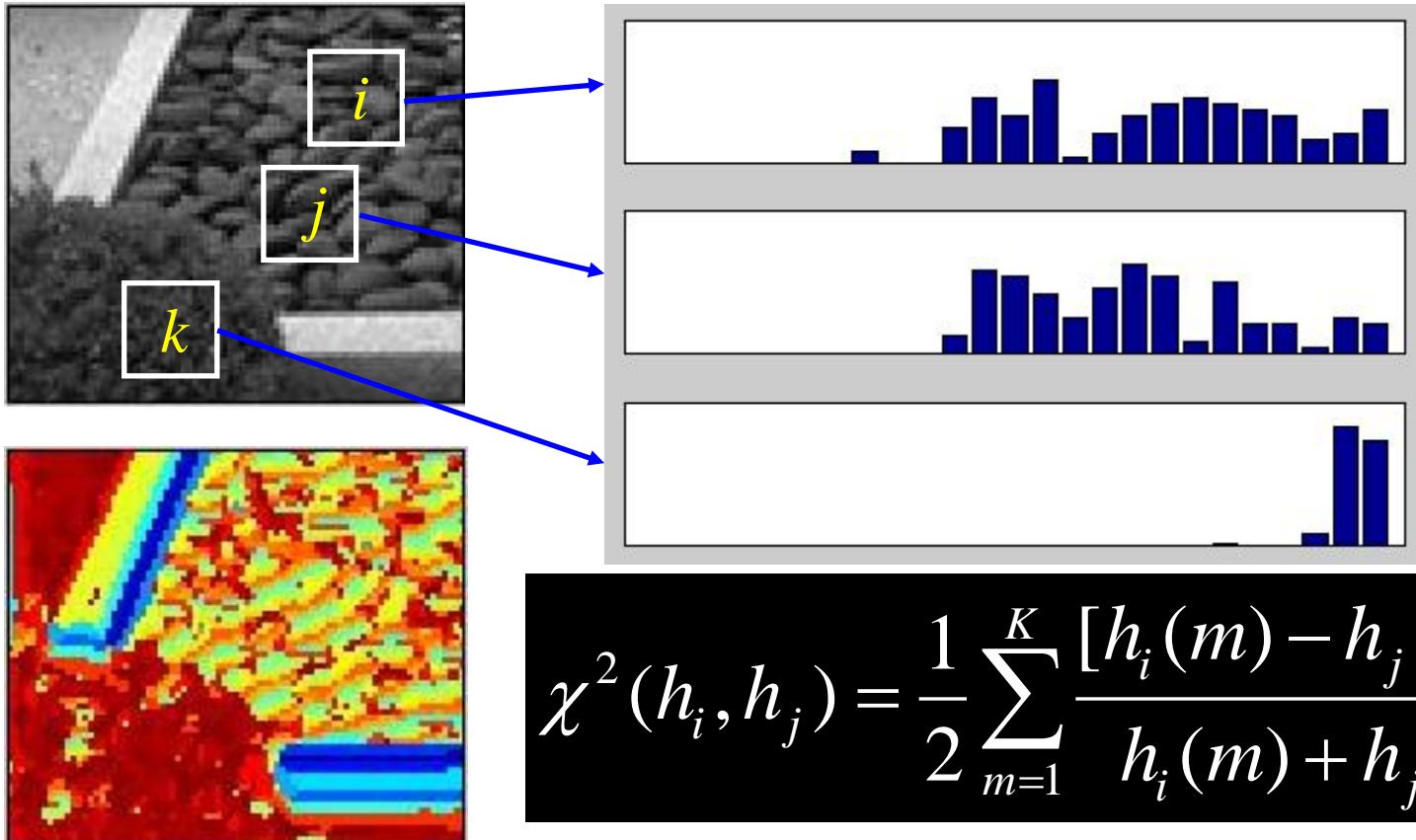
Detection and categorization are practically simultaneous (Grill-Spector & Kanwisher, 2005)



Texton Histogram Matching



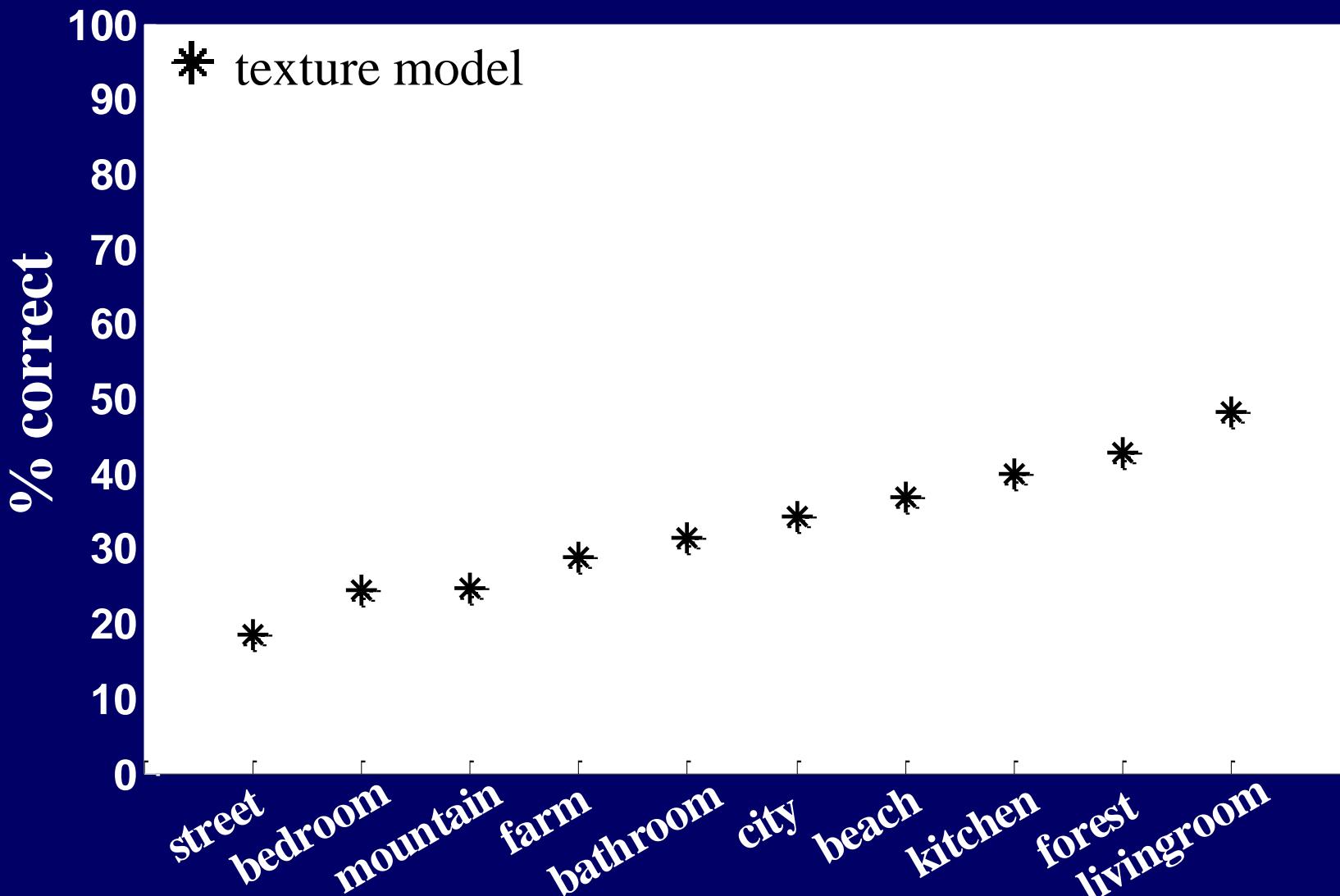
Texton histograms work on patches too



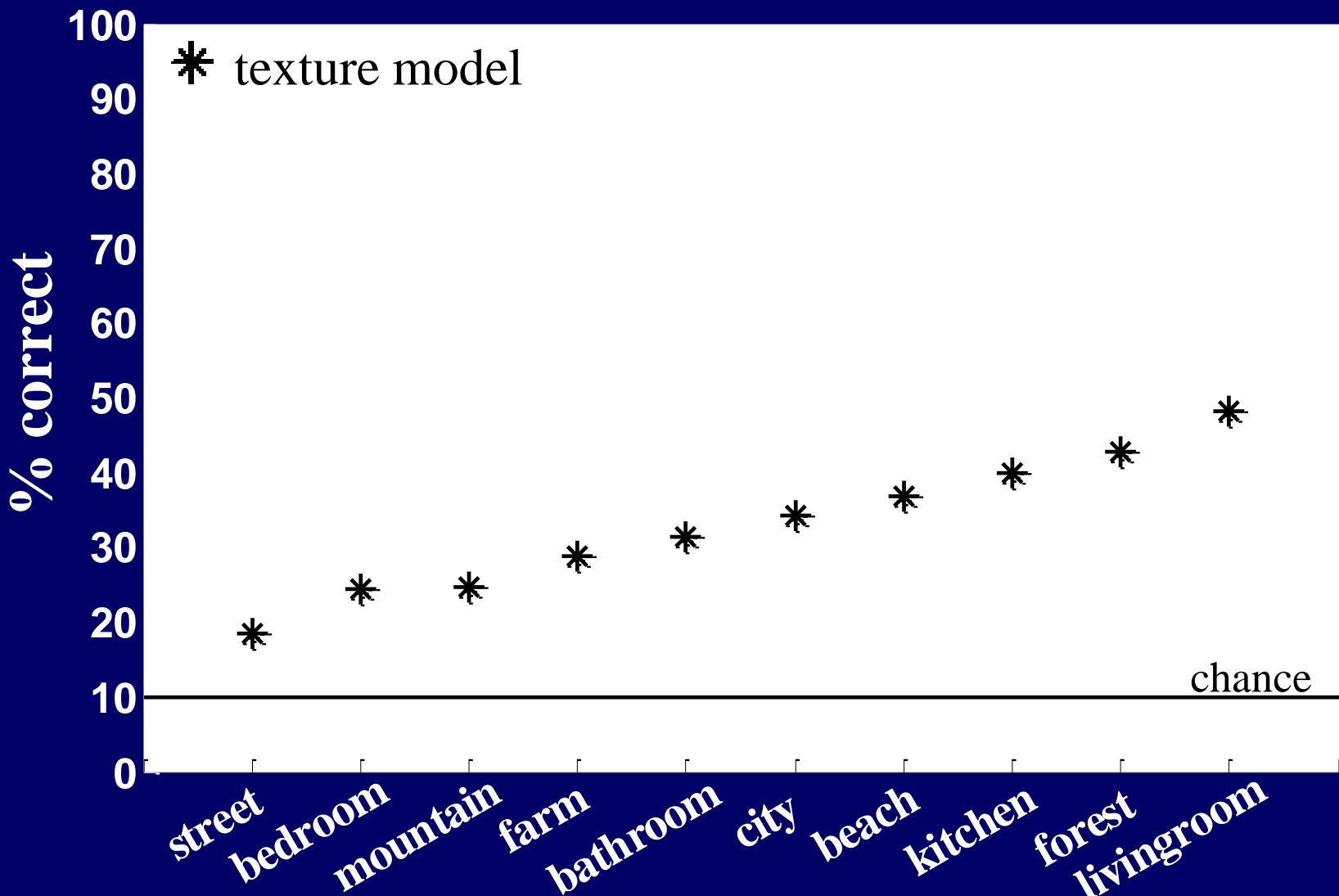
$$\chi^2(h_i, h_j) = \frac{1}{2} \sum_{m=1}^K \frac{[h_i(m) - h_j(m)]^2}{h_i(m) + h_j(m)}$$

(Malik et al, 99)

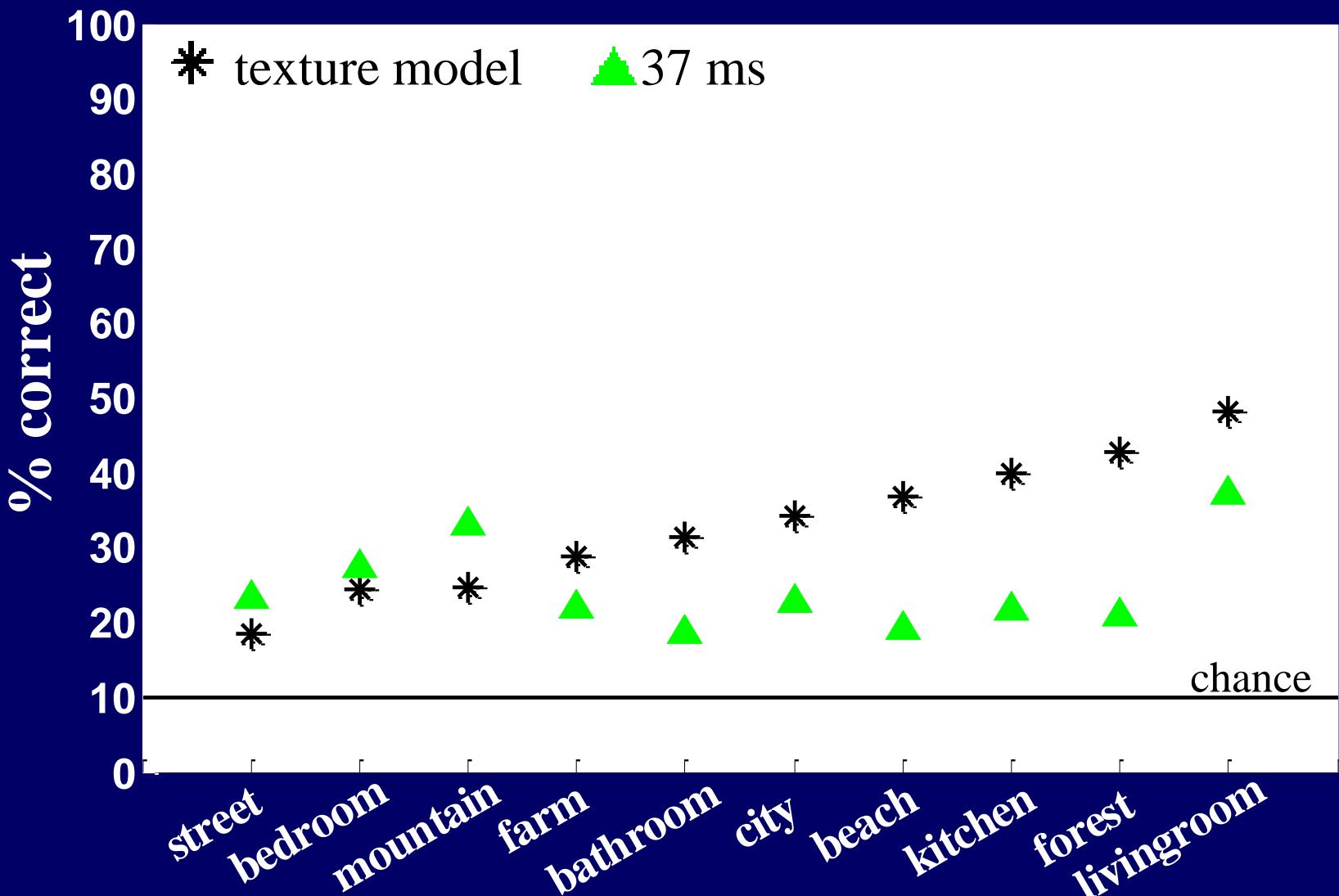
Discrimination of Basic Categories



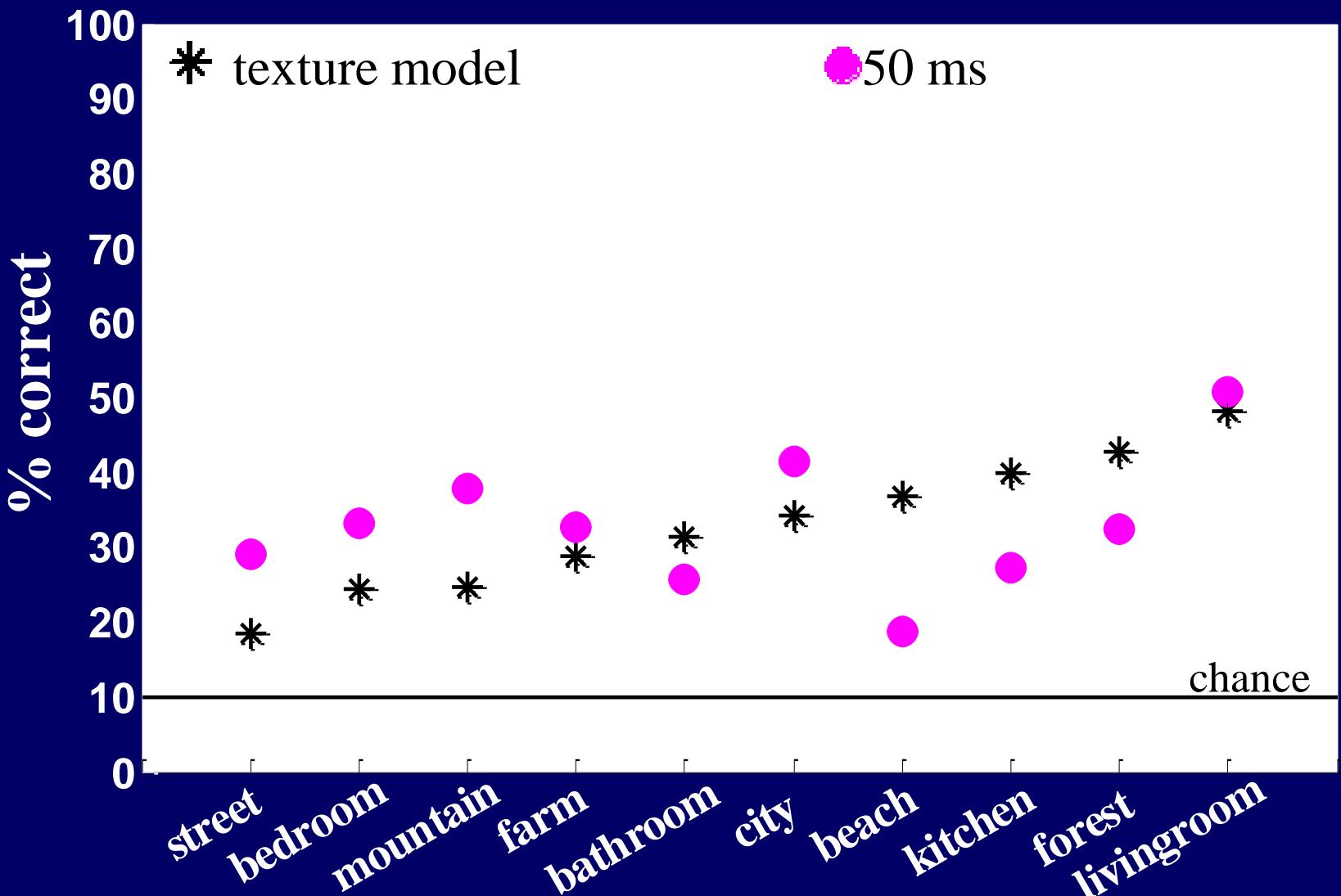
Discrimination of Basic Categories



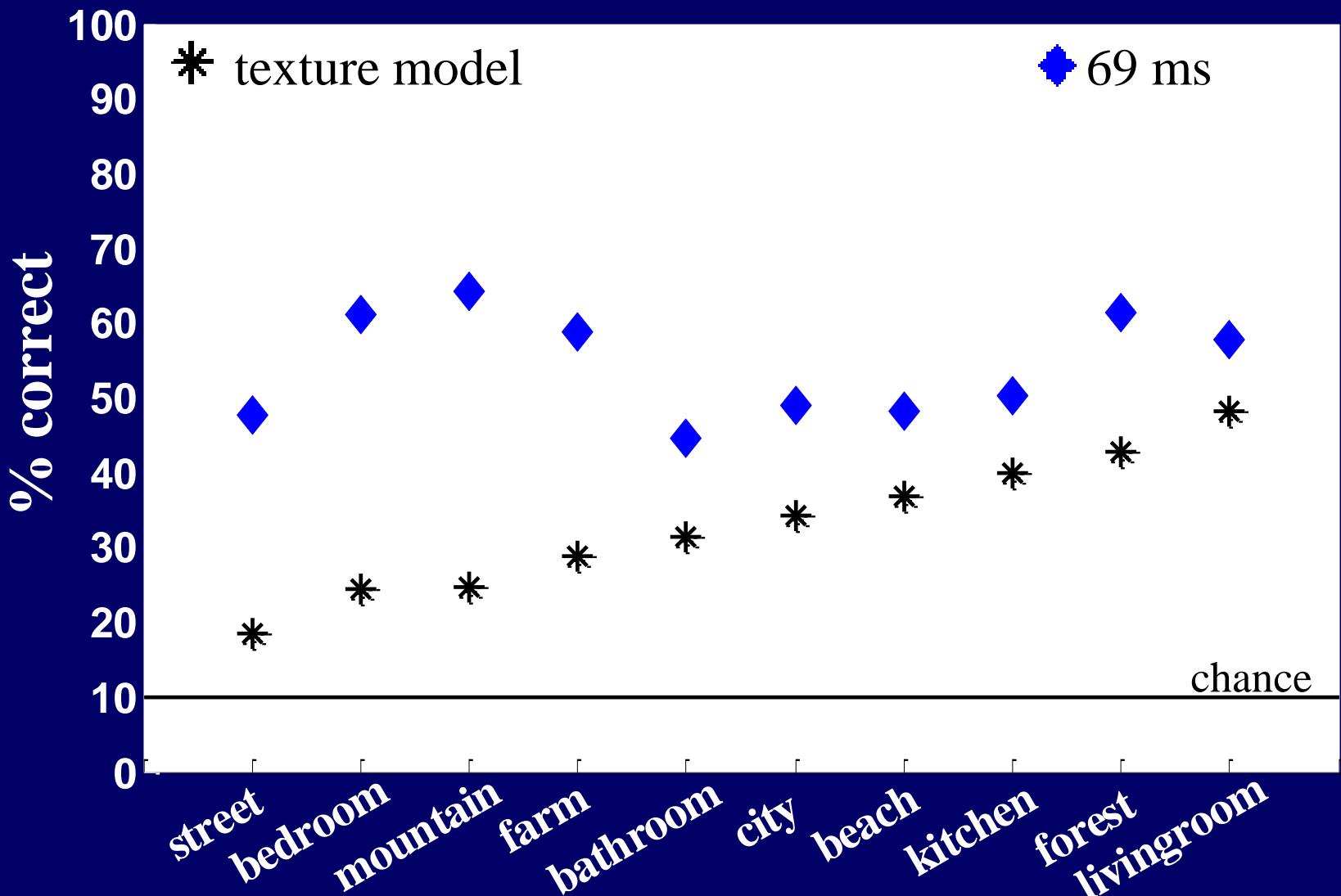
Discrimination of Basic Categories



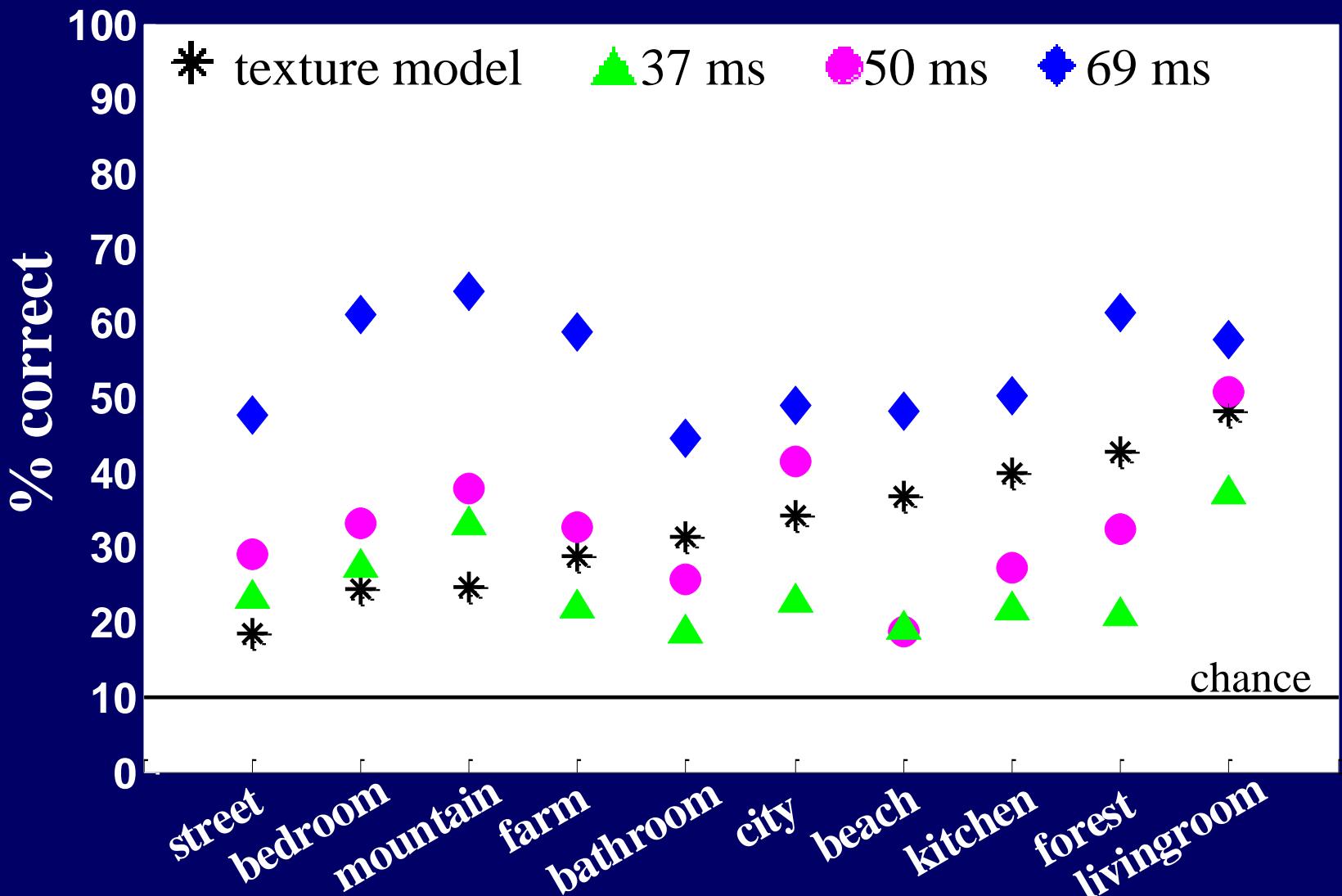
Discrimination of Basic Categories



Discrimination of Basic Categories



Discrimination of Basic Categories



Object Recognition using Texture

