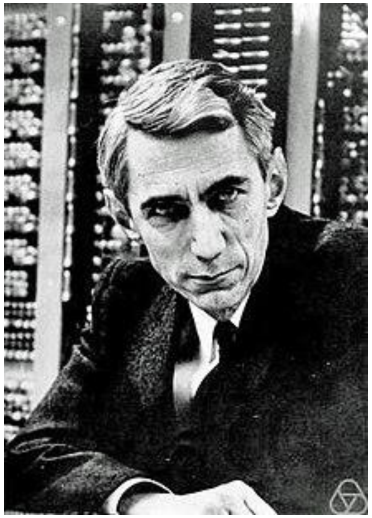
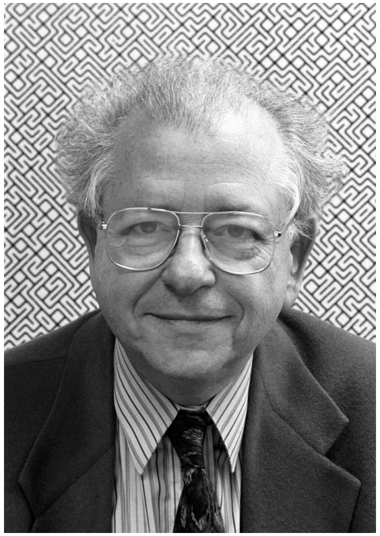
Statistical Images Representations I
(Everything is Texture!)



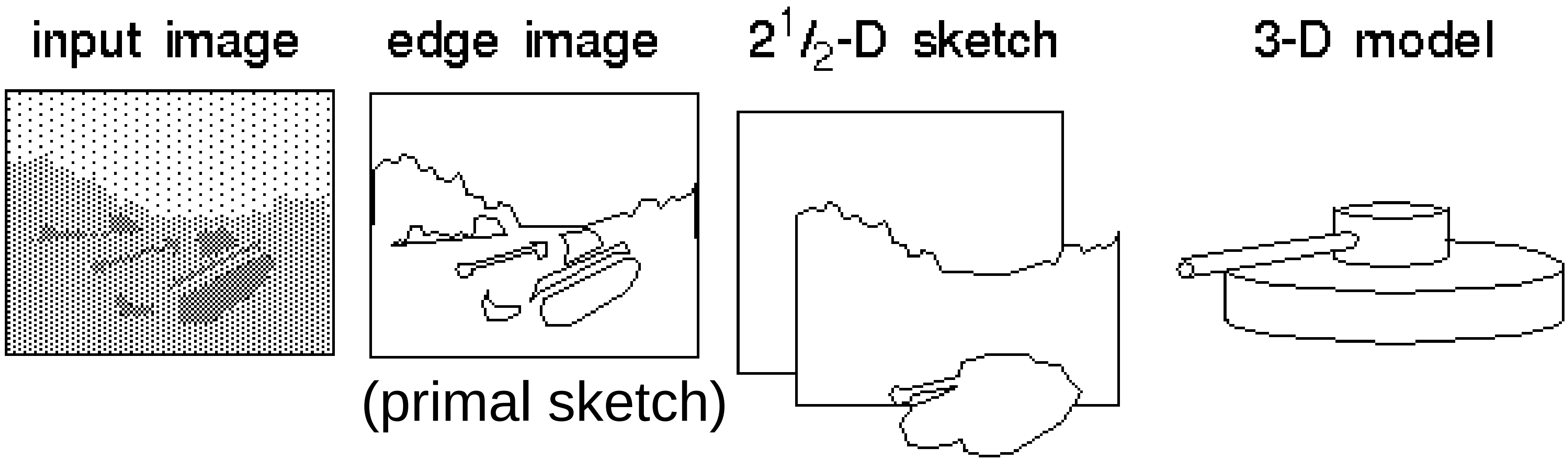
Claude Shannon



Bela Julesz

Alexei Efros, CS280, Spring 2023

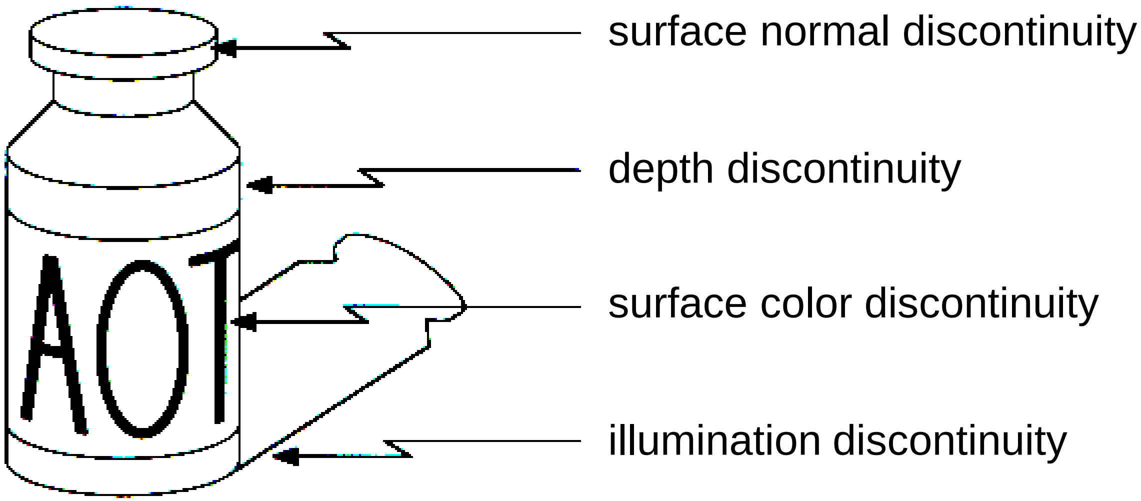
Things used to be so easy…



David Marr's View of Vision (1980s)

Origin of edges

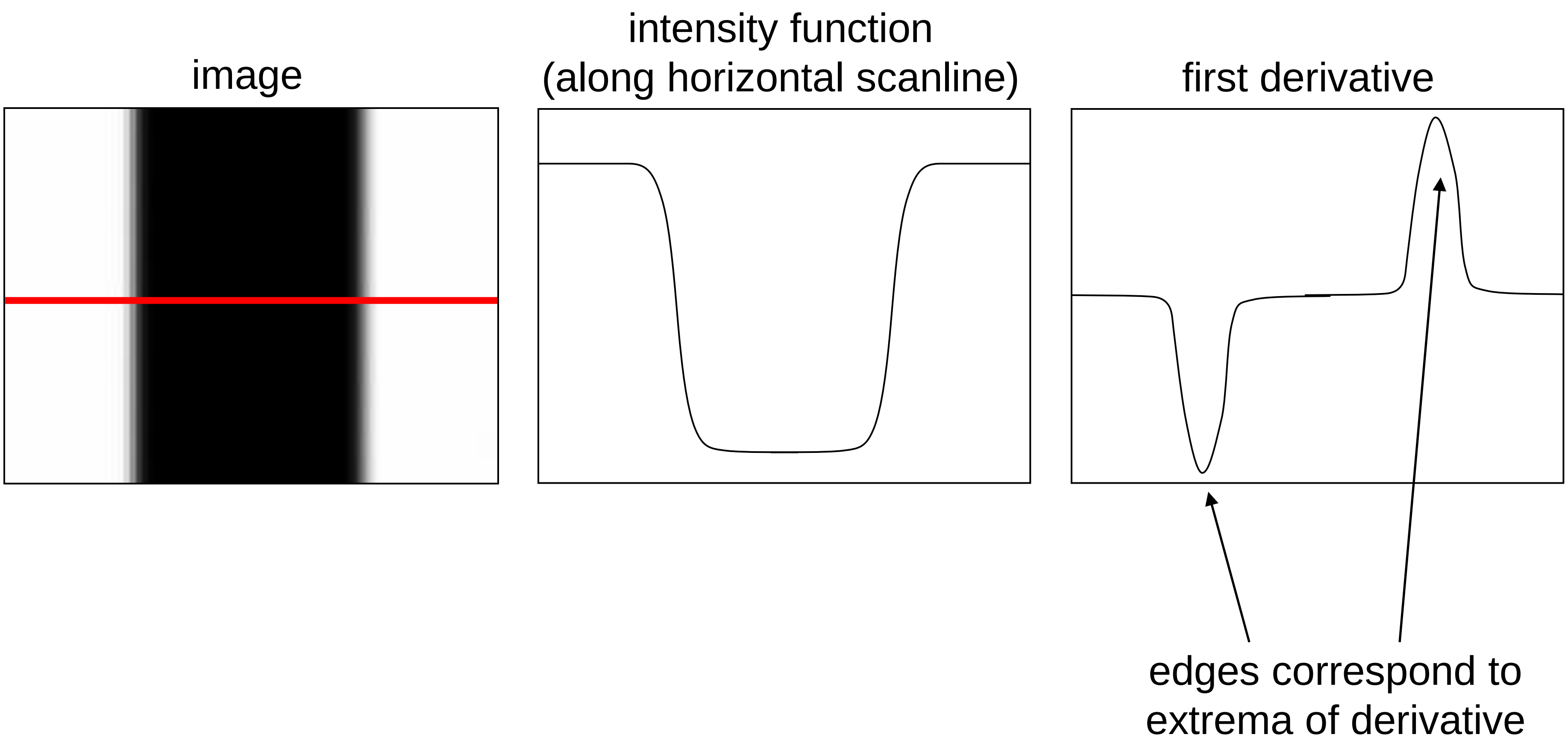
Edges are caused by a variety of factors:



Source: Steve Seitz

Characterizing edges

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



An edge is not a line...

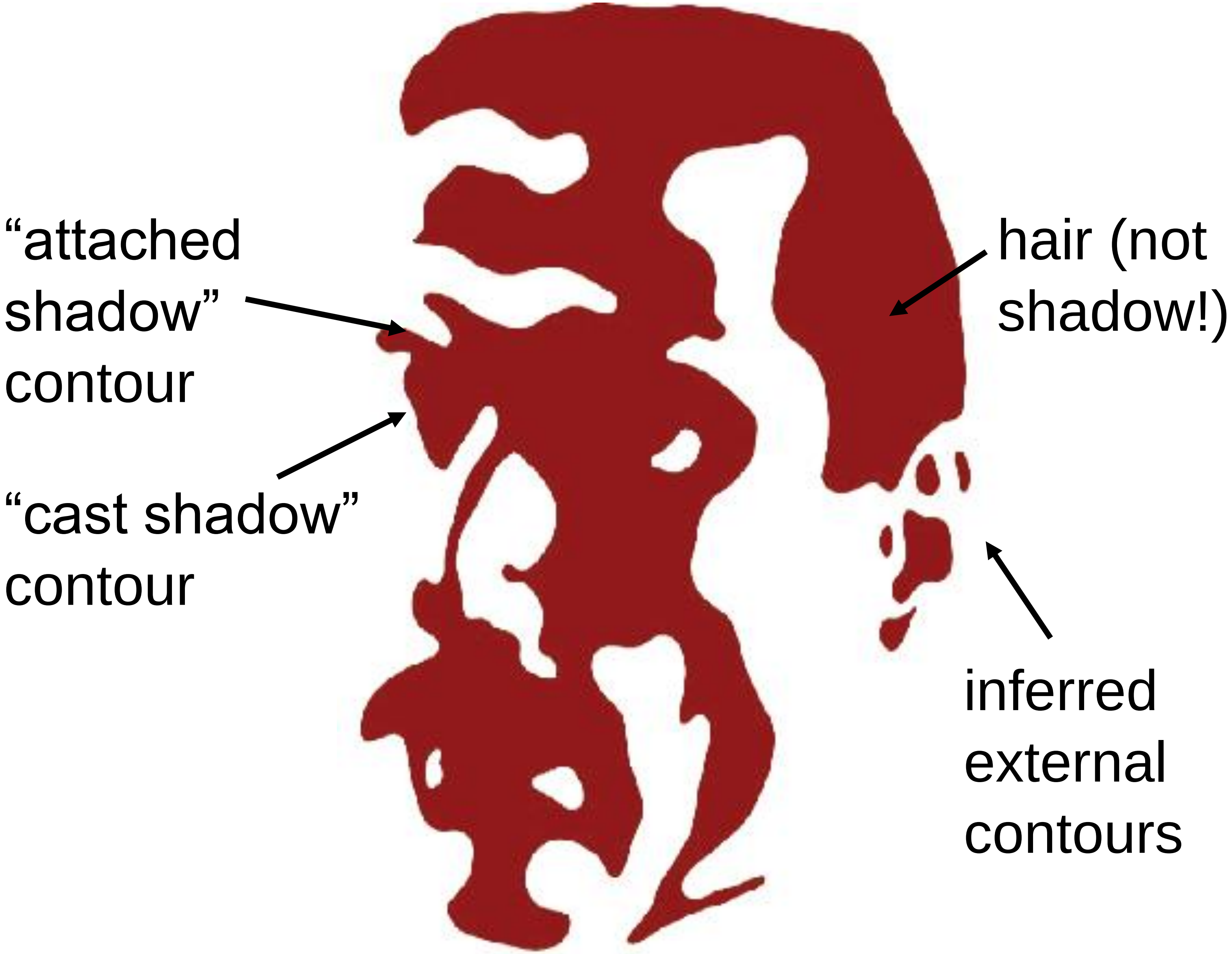
State-of-the-Art Edge Detection

image | human segmentation | gradient magnitude

Berkeley segmentation database:

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

two-tone images


"attached shadow" contour
"cast shadow" contour
hair (not shadow!)
inferred external contours

Need a more statistical representation

Texture as statistical image representation

- Narrowly: texture depicts spatially repeating patterns

radishes

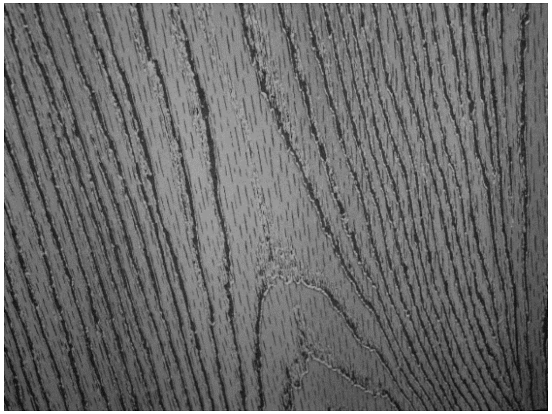
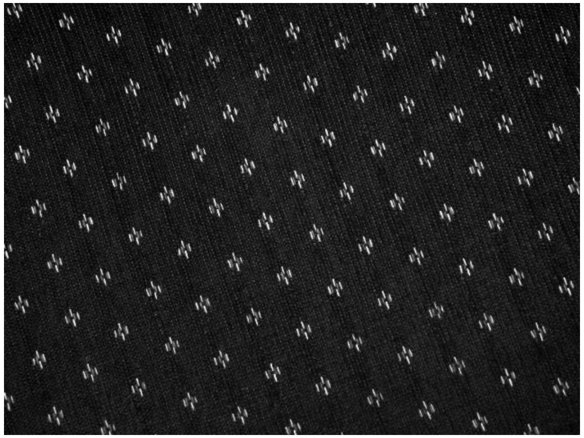
rocks

yogurt

Texture as “stuff”

Source: Forsyth

Texture and Material

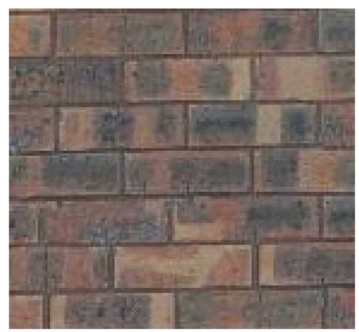



http://www-cvr.ai.uiuc.edu/ponce_grp/data/texture_database/samples/

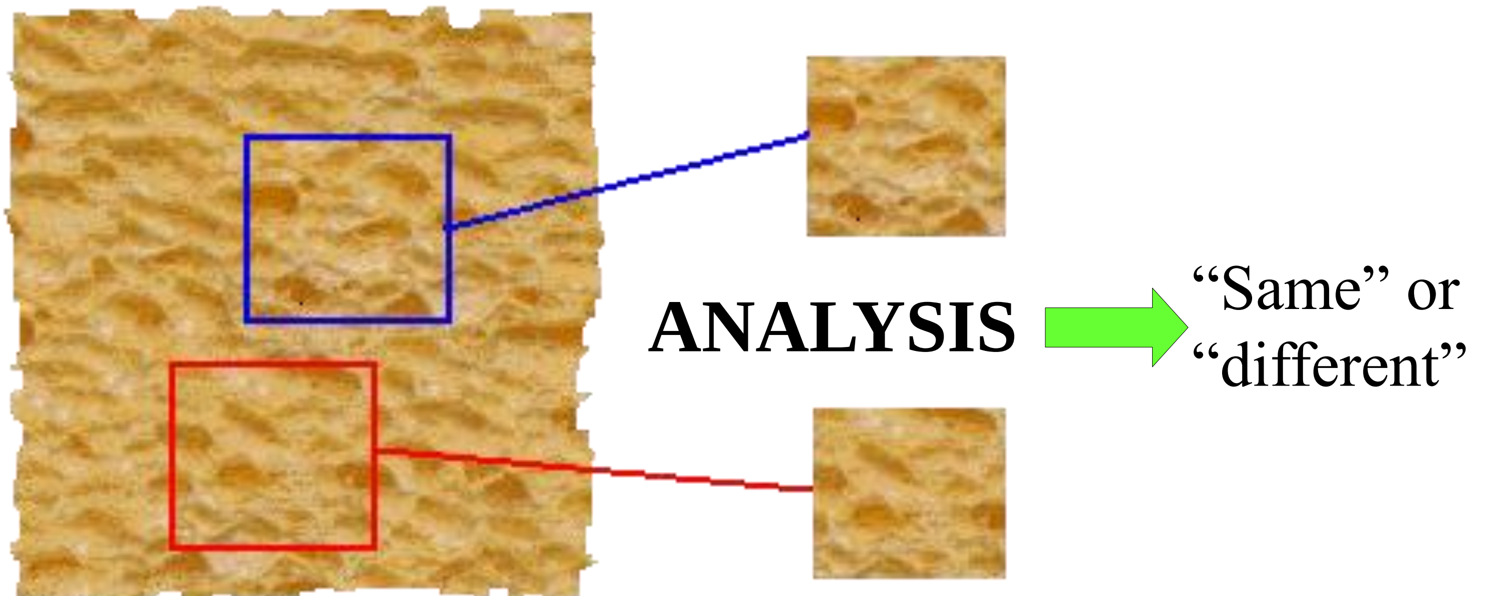
Statistical, not exact, similarity







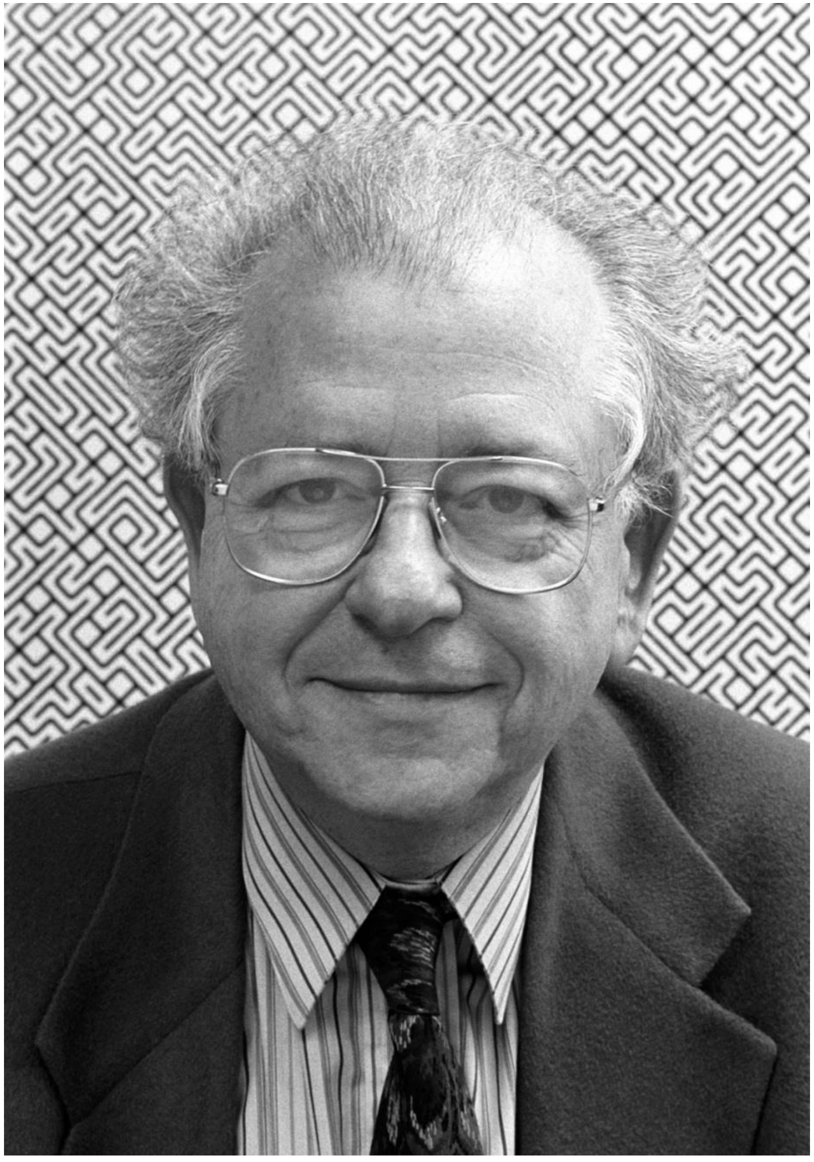
Texture Analysis



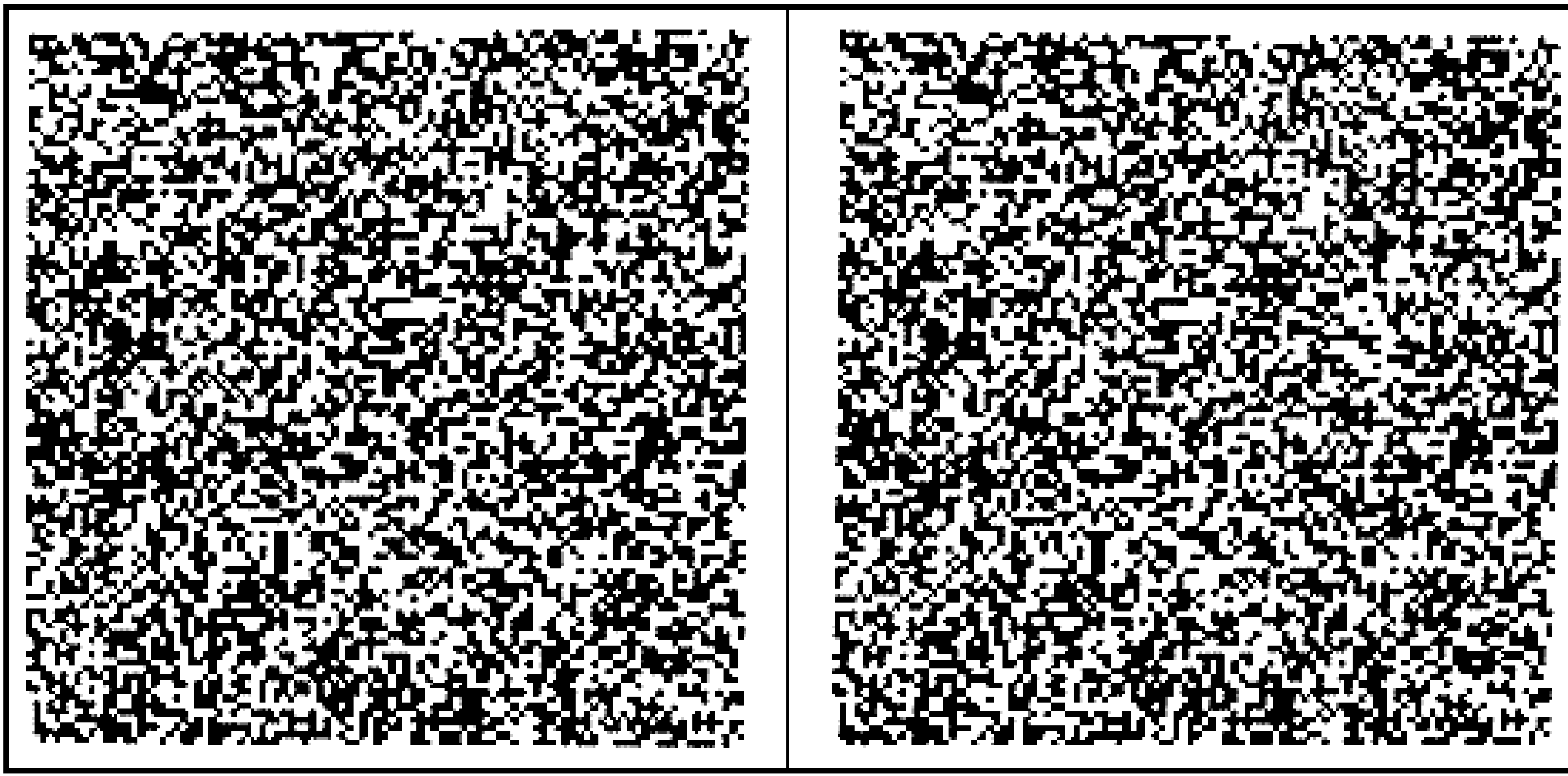
True (infinite) texture

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

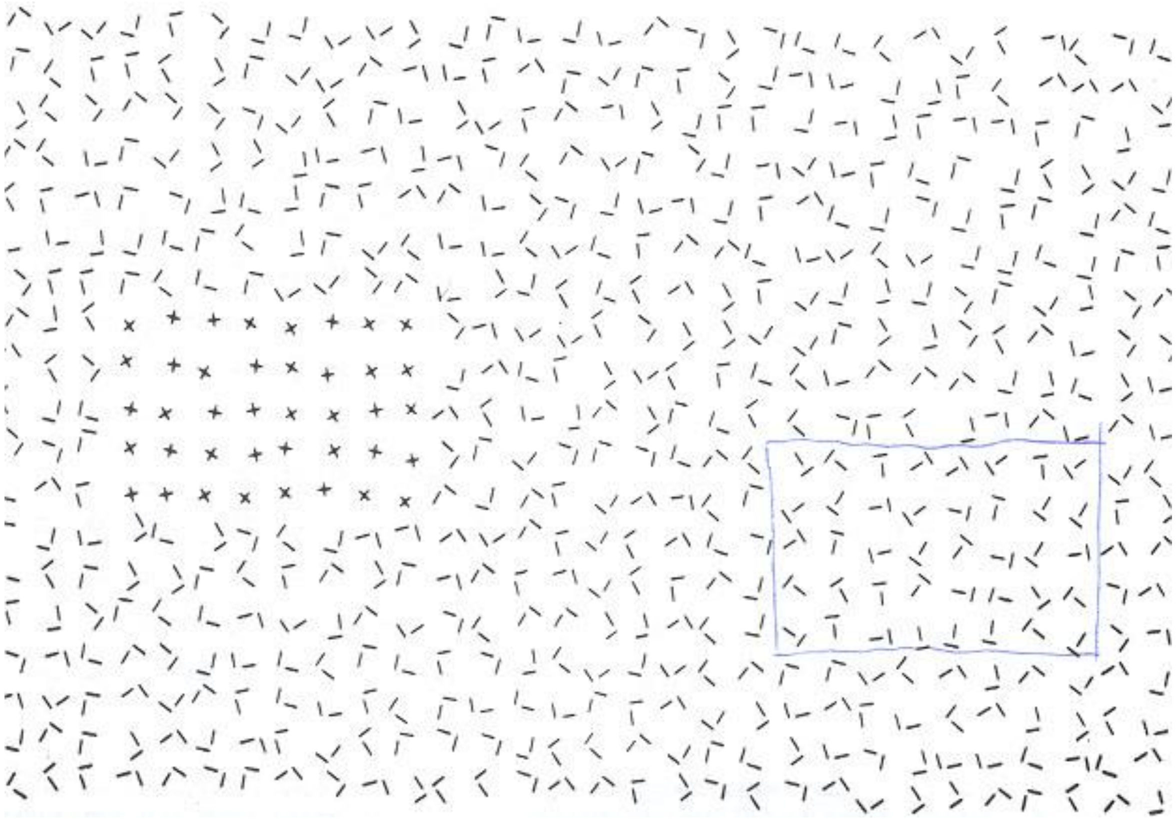
Béla Julesz, father of texture



Random Dot Stereograms

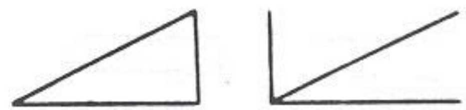


Texton Discrimination (Julesz)

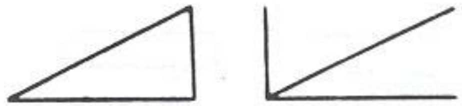


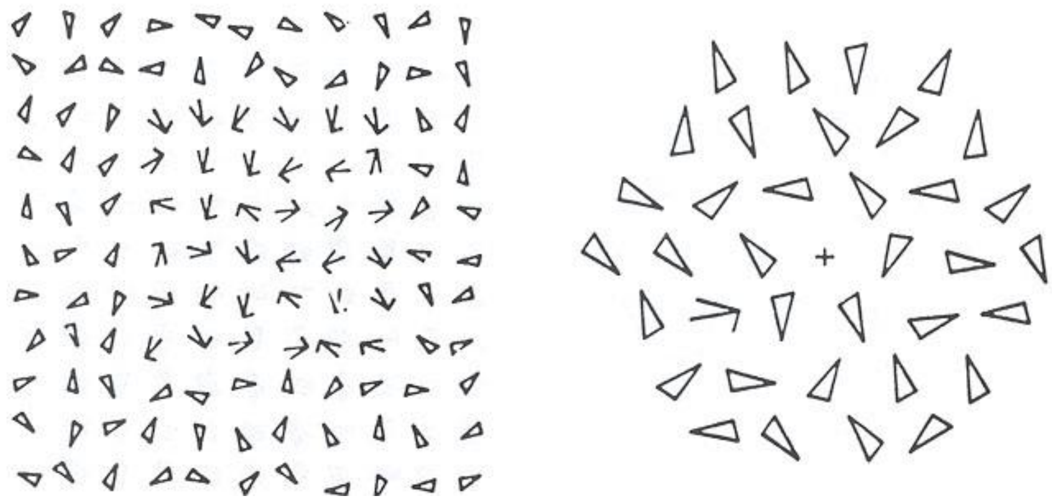
Human vision is sensitive to the difference of some types of elements and appears to be “numb” on other types of differences.

Search Experiment I



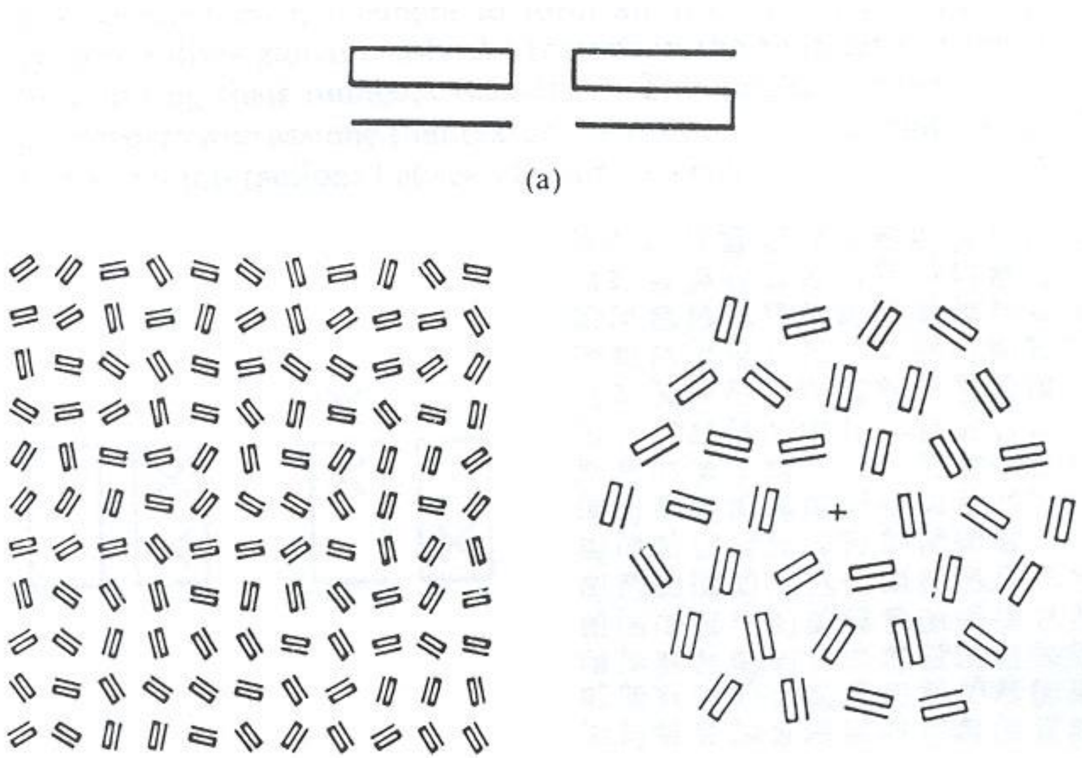
Search Experiment I





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.

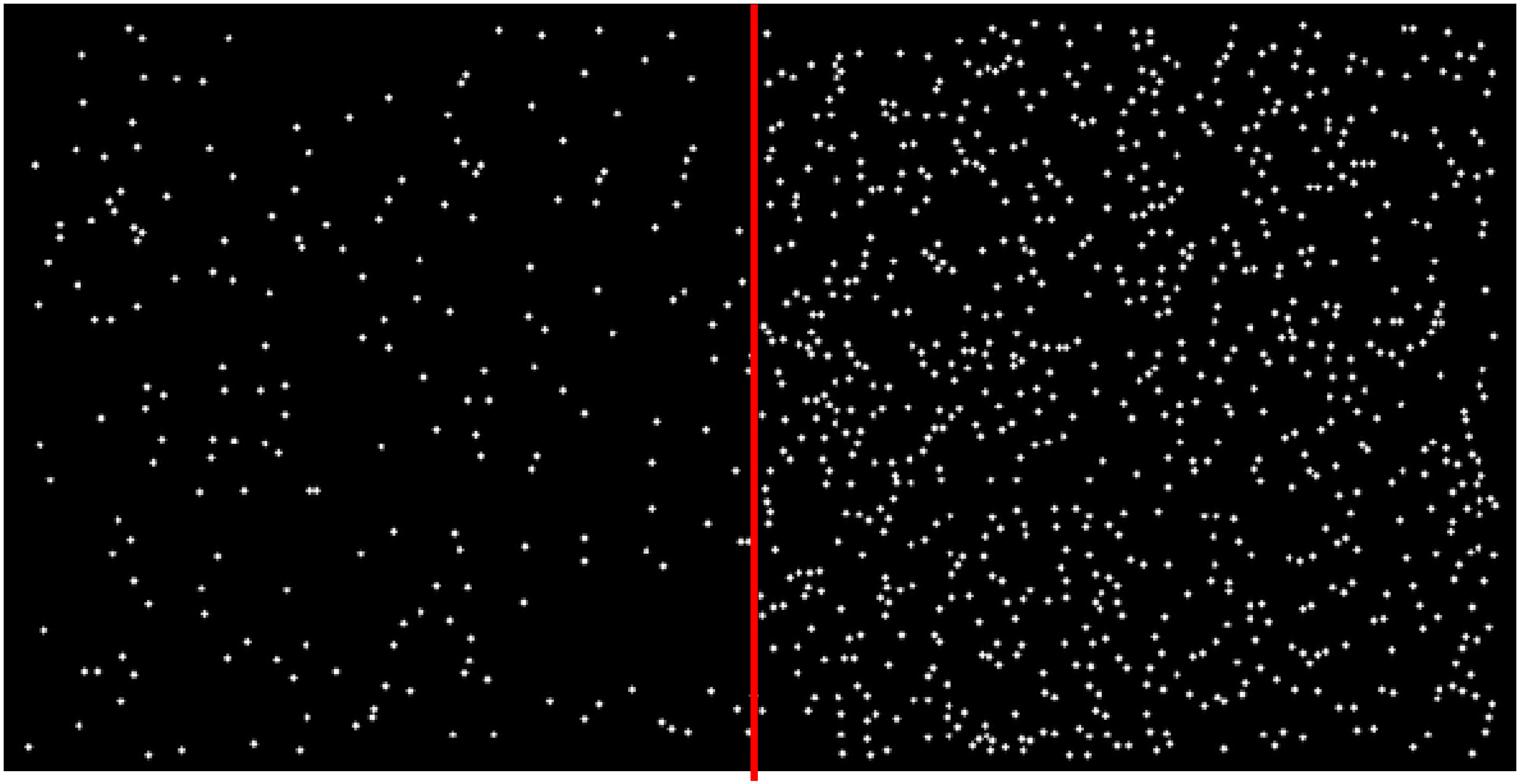
2. Attentive vision

   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.*

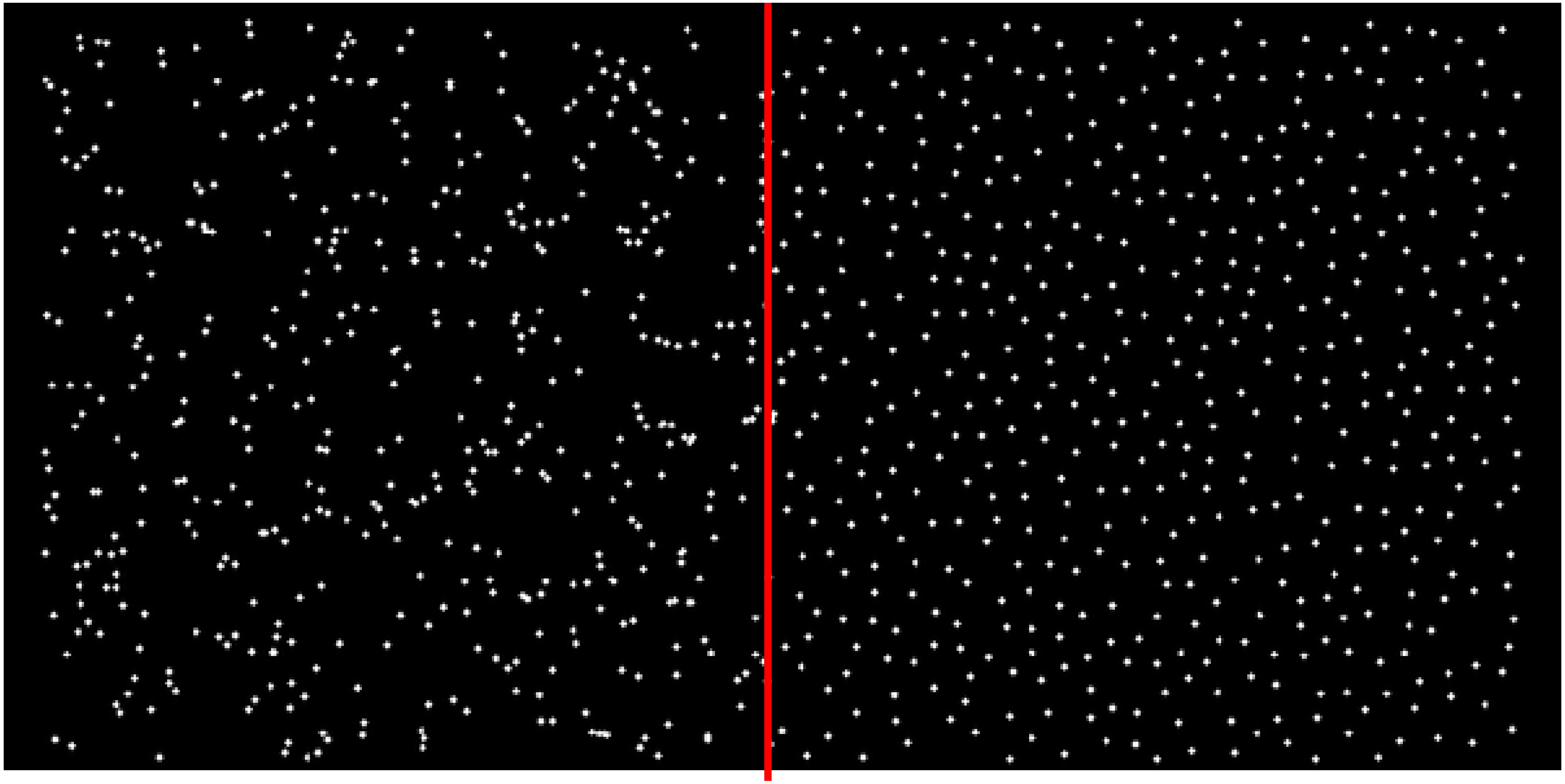
1st Order Statistics



5% white

20% white

2^{nd} Order Statistics



10% white

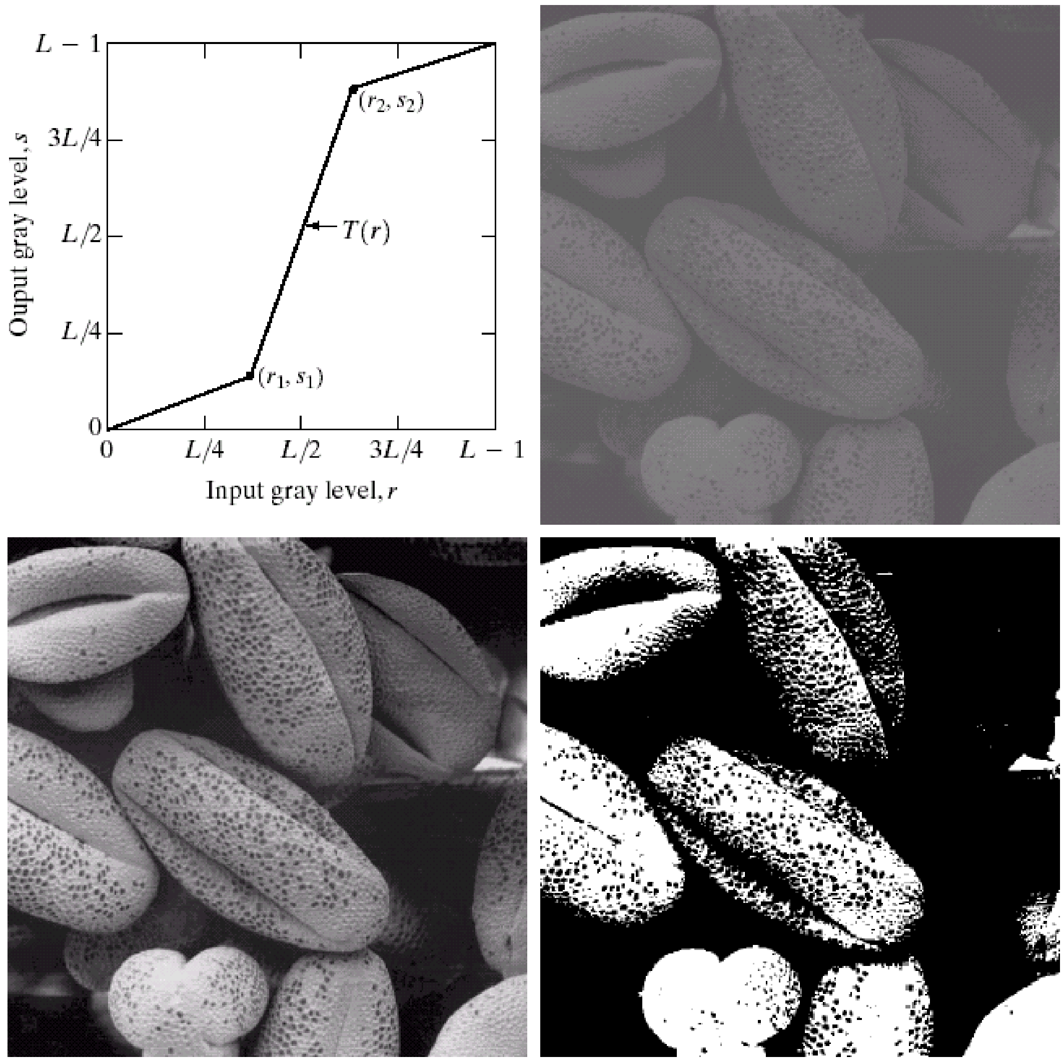
Two questions of texture modeling

- 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

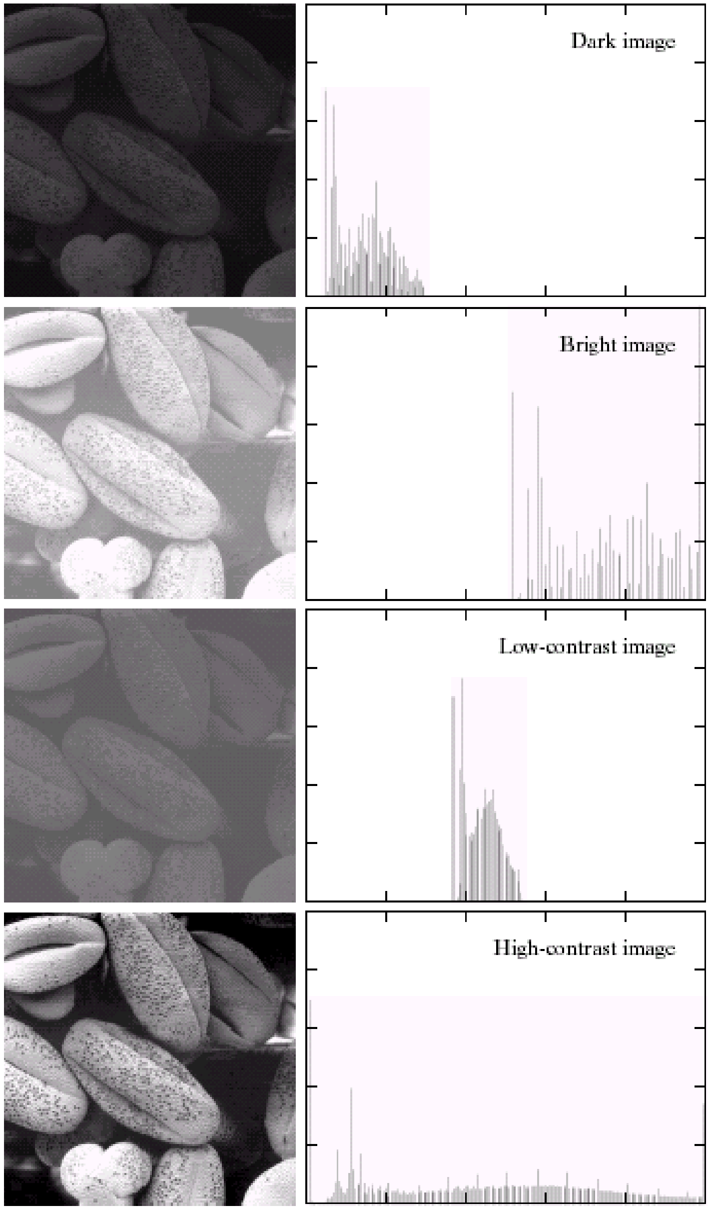
Contrast Stretching

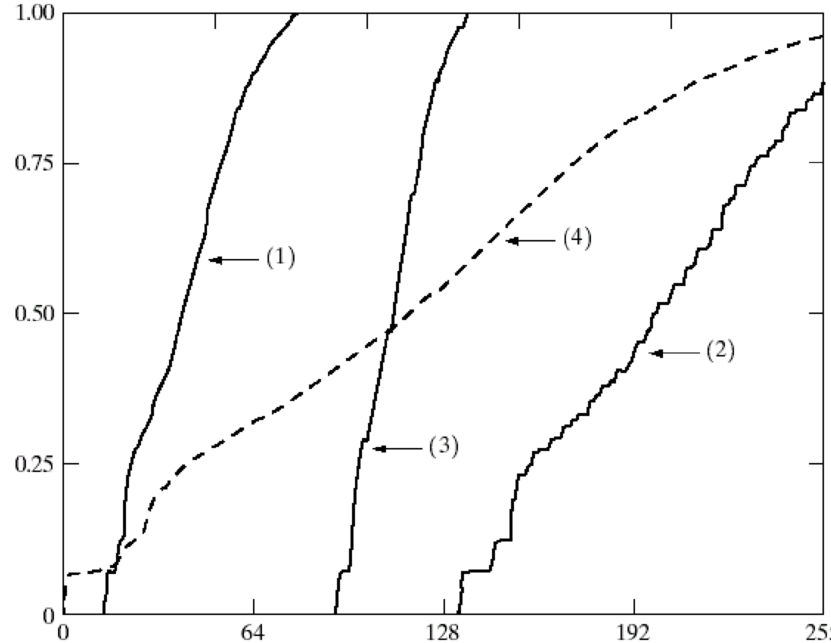


a b
c d

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





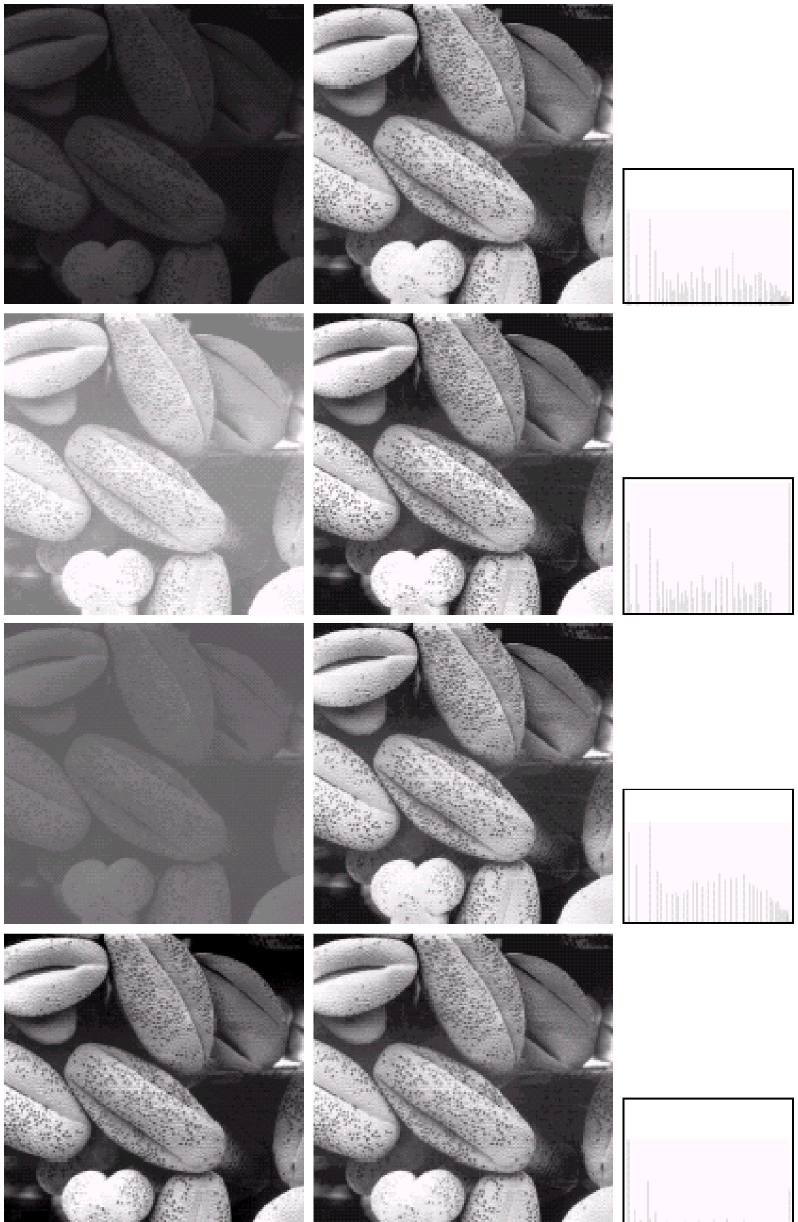
Cumulative Histograms

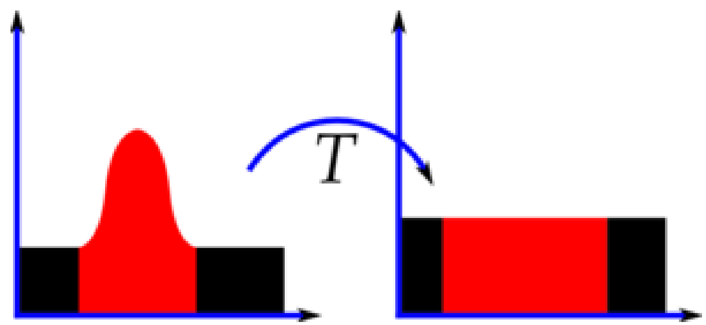
$$T(f(x, y))$$

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



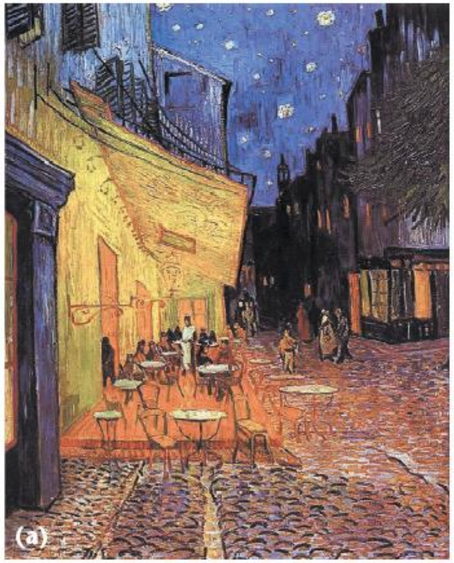


credit @ Zefram

a b c

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

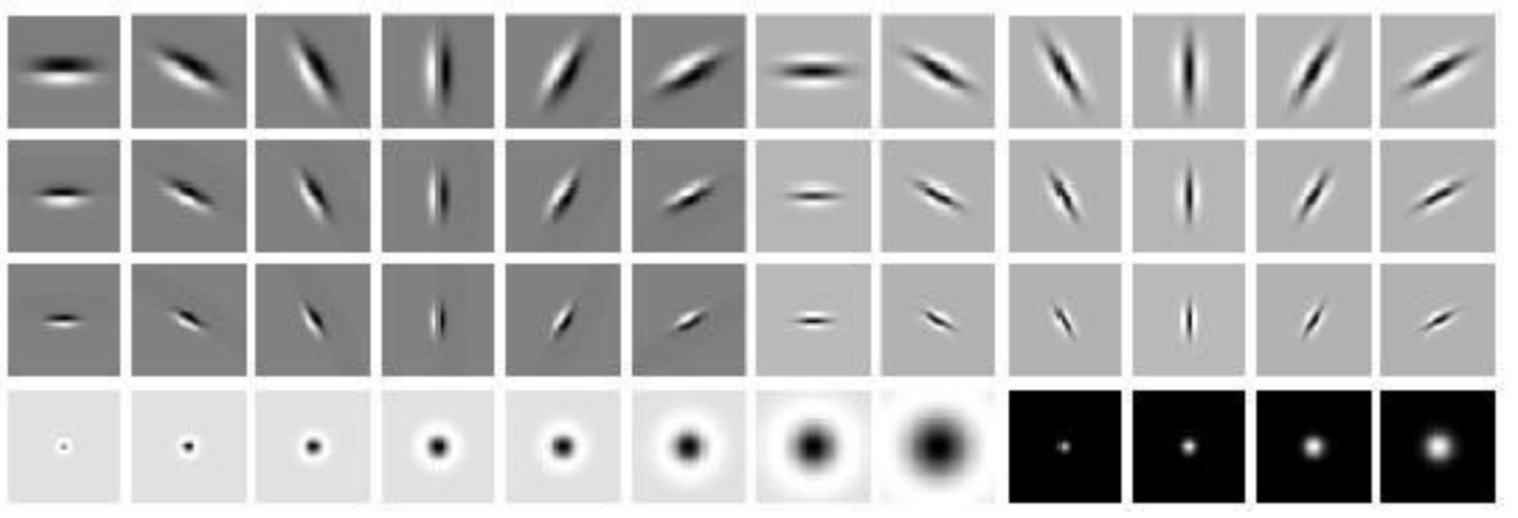
Color Transfer [Reinhard, et al, 2001]



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

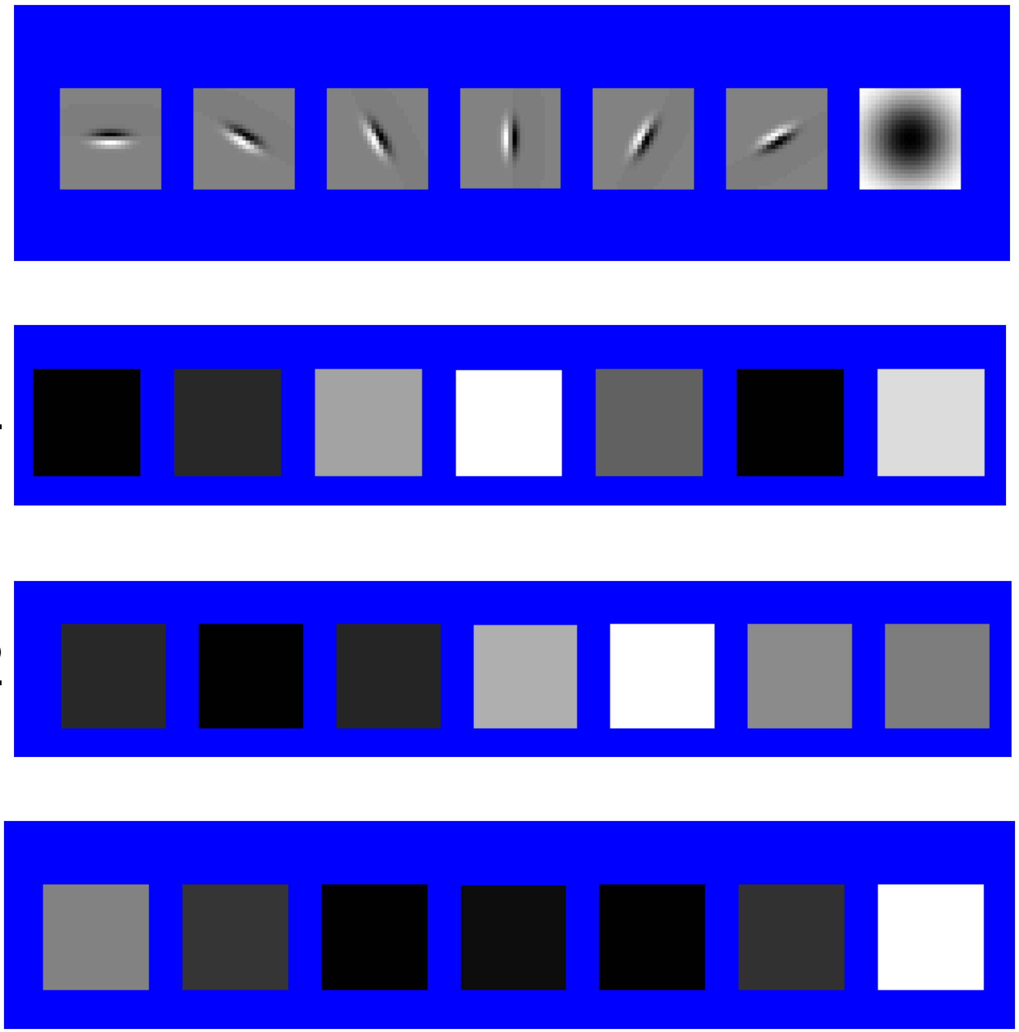
Going up from pixels: V1 filter-banks

LM Filter Bank



Can you match the texture to the response?

Filters



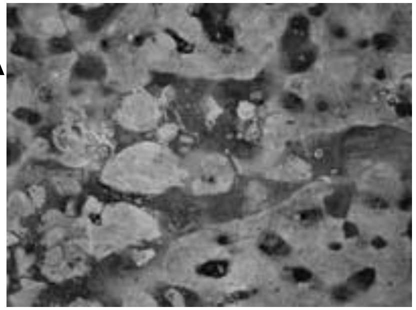
1

2

3

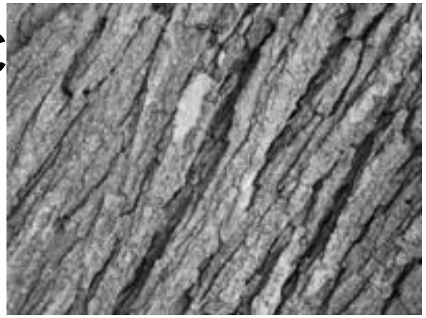
Mean abs responses

A

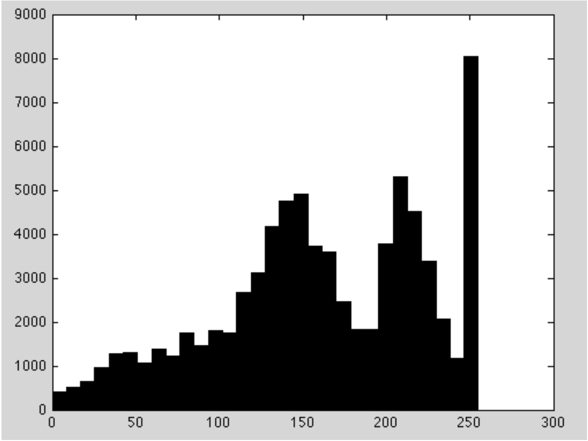


B

C

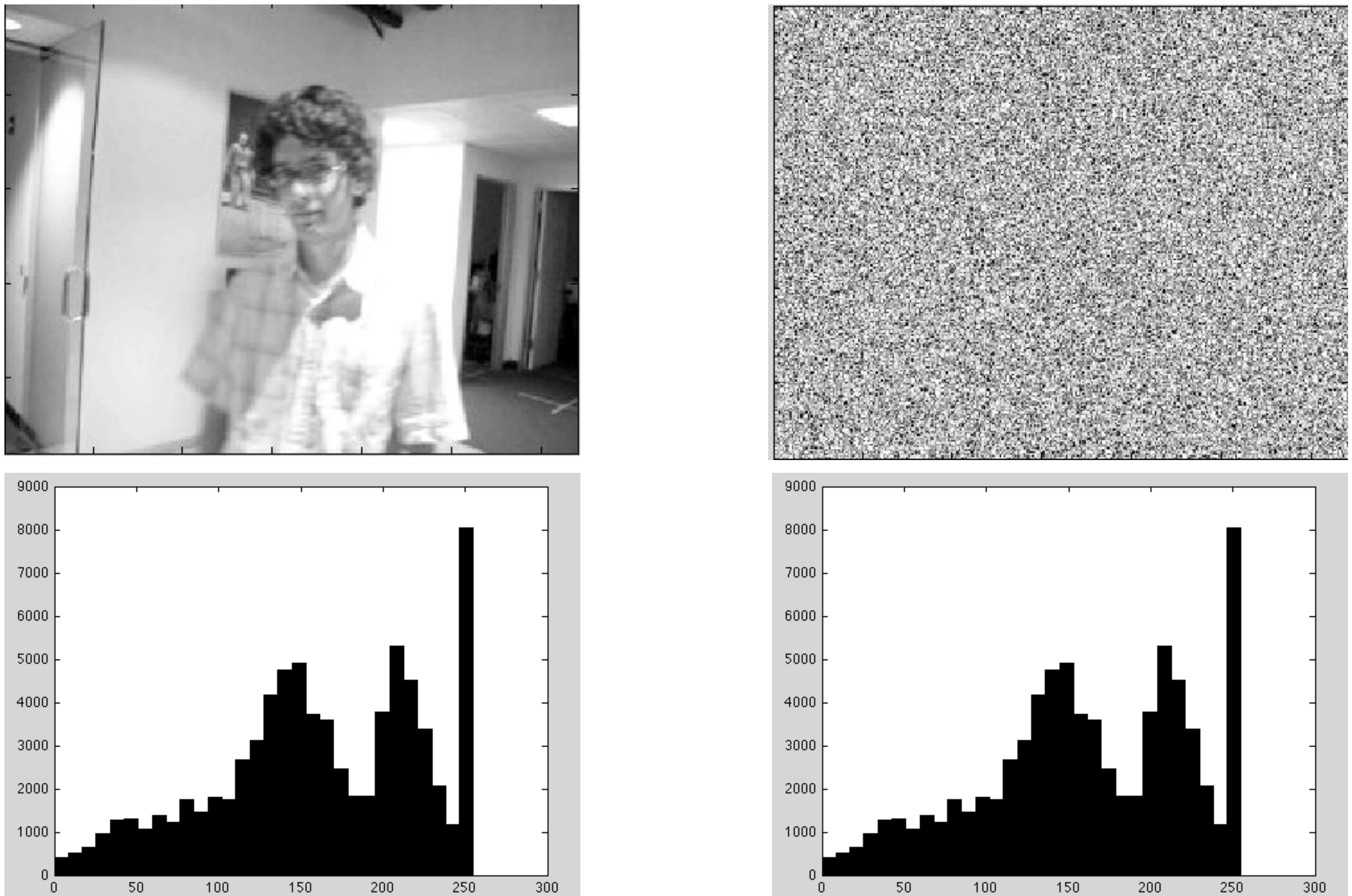


Pixel Histograms



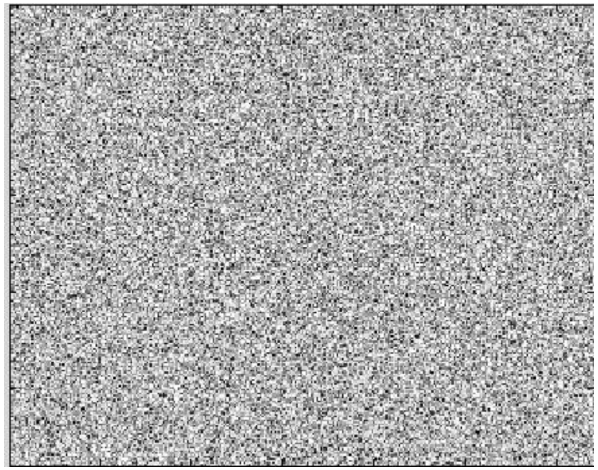
Slide by Erik Learned-Miller

Pixel Histograms



Slide by Erik Learned-Miller

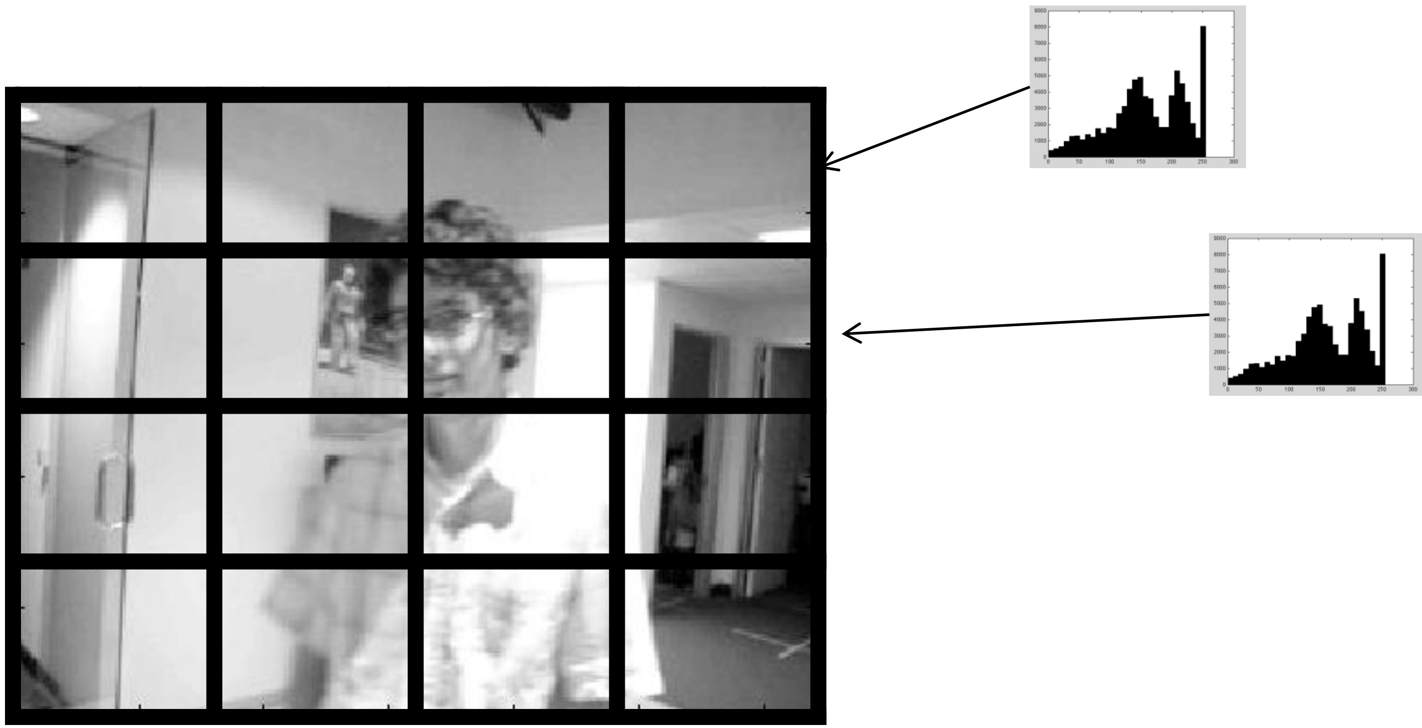
Gray value histogram comparisons



They're equal

Slide by Erik Learned-Miller

Adding spatial structure

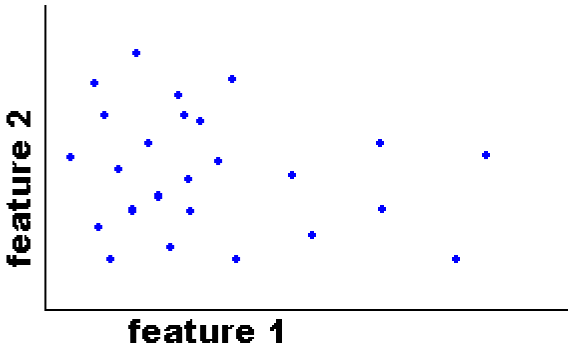


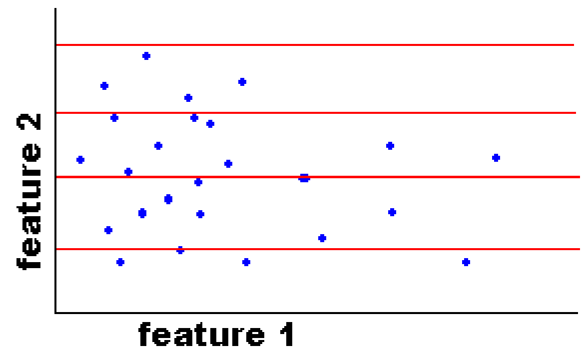
A separate histogram for each region.

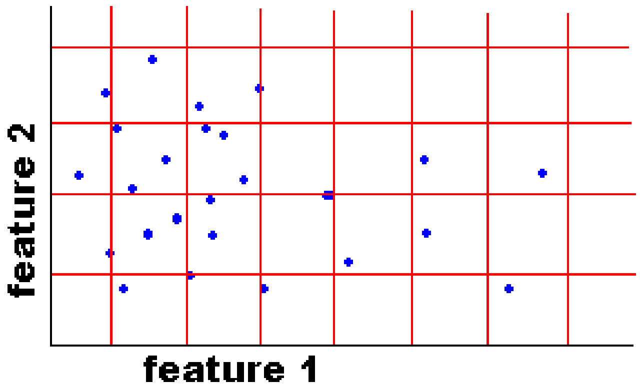
Slide by Erik Learned-Miller

Image Representations: Histograms

Images from Dave Kauchak

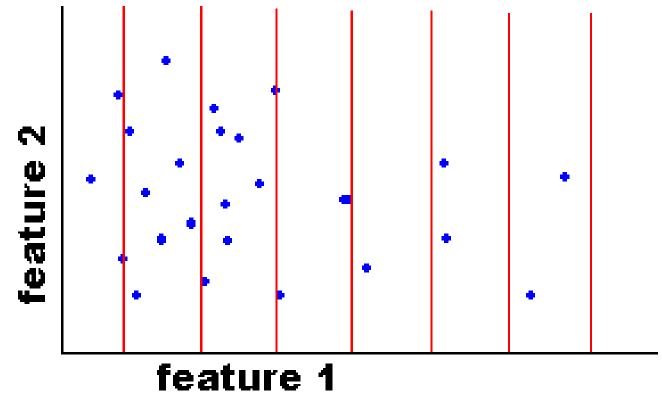






Joint histogram

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



Marginal histogram

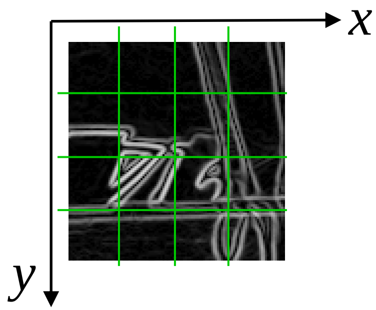
- Requires independent features
- More data/bin than joint histogram

Ex: SIFT descriptor [Lowe'99]

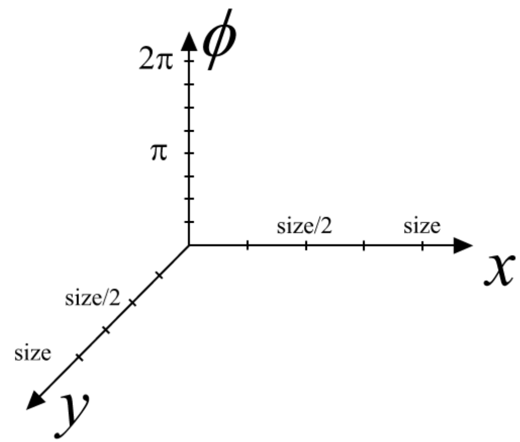
distribution of the gradient over an image patch

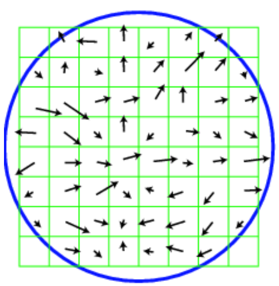
image patch

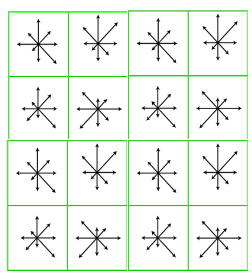
gradient



3D histogram





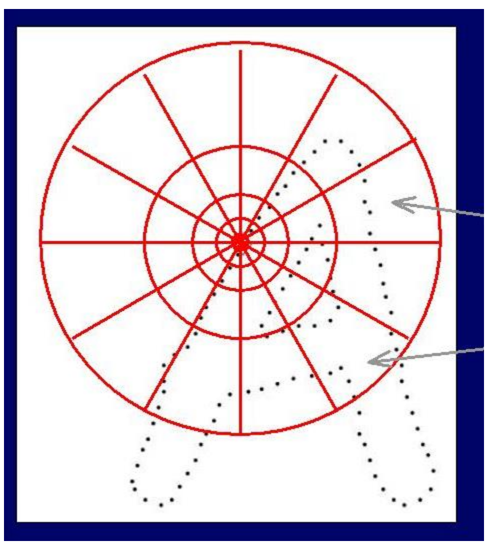


4x4 location grid and 8 orientations (128 dimensions)

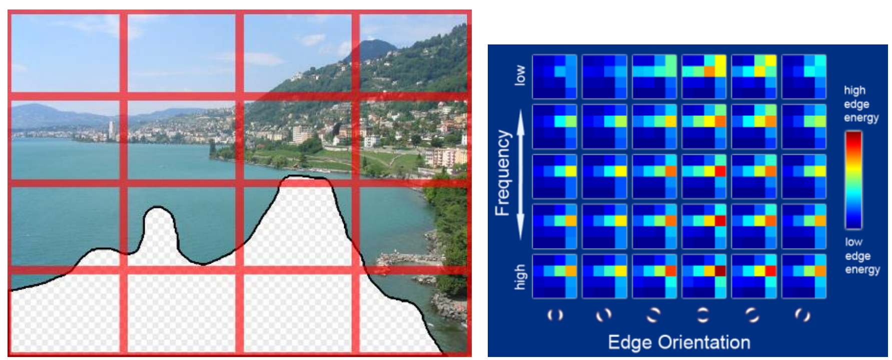
very good performance in image matching [Mikolaczyk and Schmid'03]

Gradient Histograms pop-up everywhere

HOG descriptor



Generalized Shape Context



Gist Descriptor

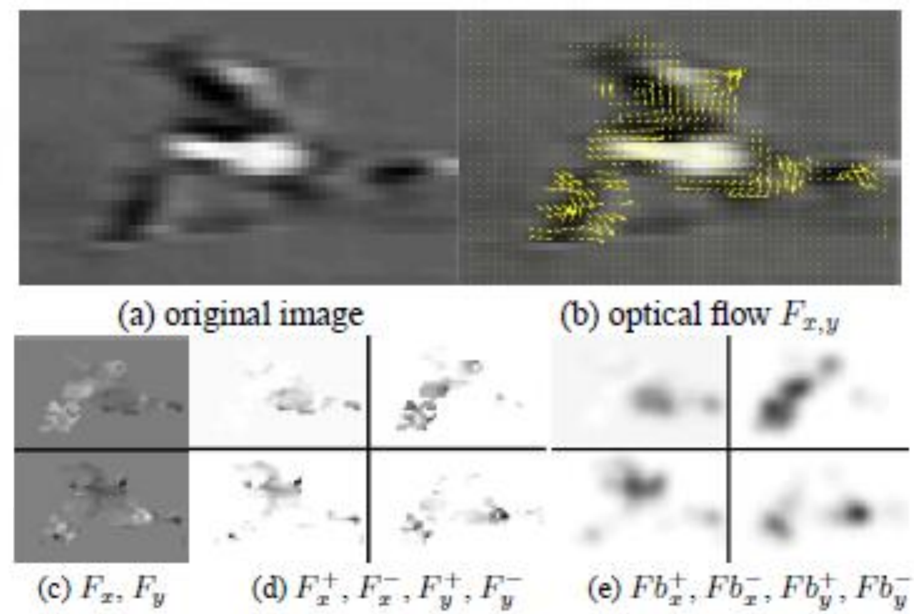
Freeman and Roth IAFGR 1995
Lowe ICCV1999
Oliva & Torralba, 2001
Belongie et al, 2001
Dalal &Triggs CVPR05

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)

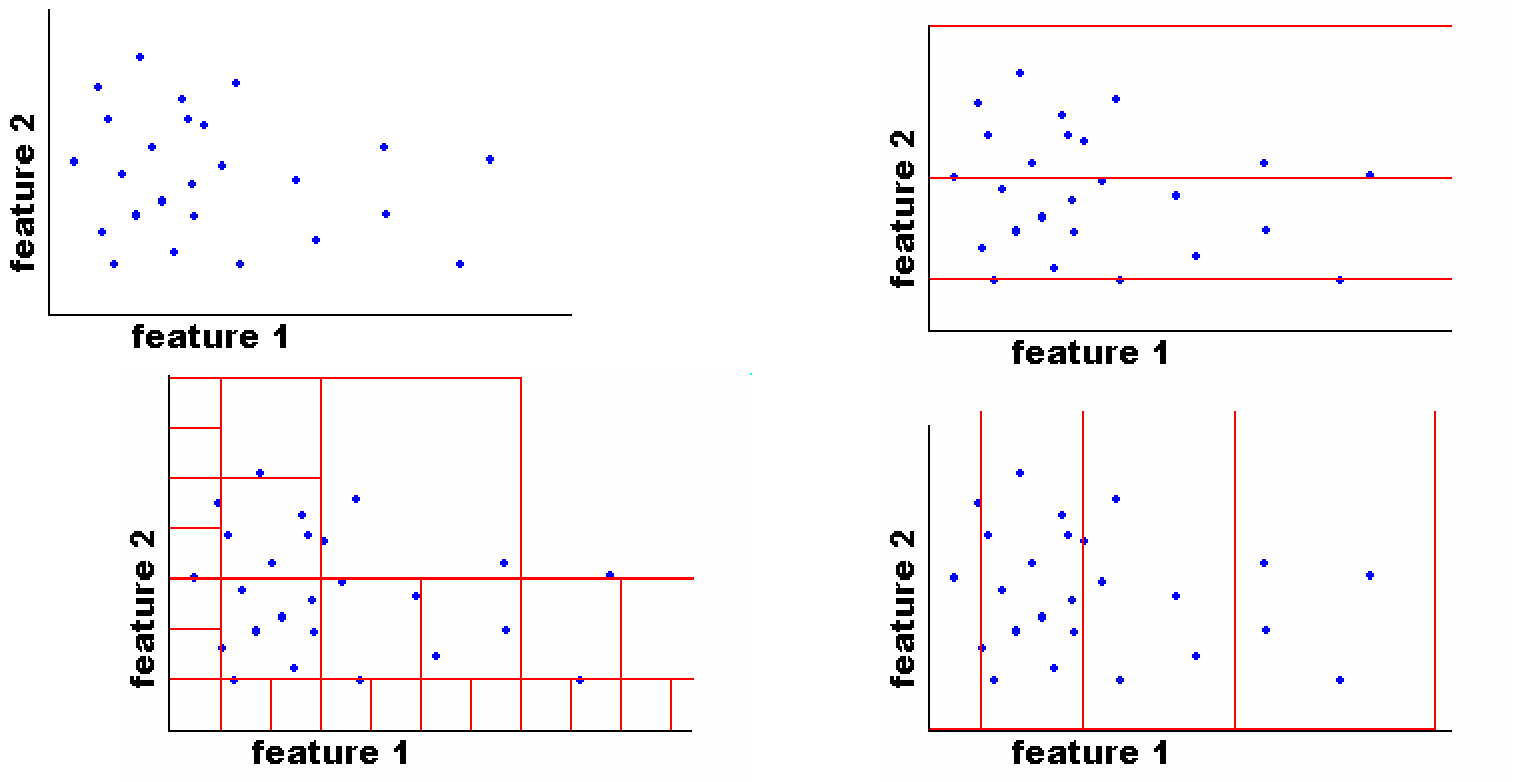
Geometric Blur (Berg et al)



Motion Descriptor (Efros e

Adaptive Representations

Images from Dave Kauchak

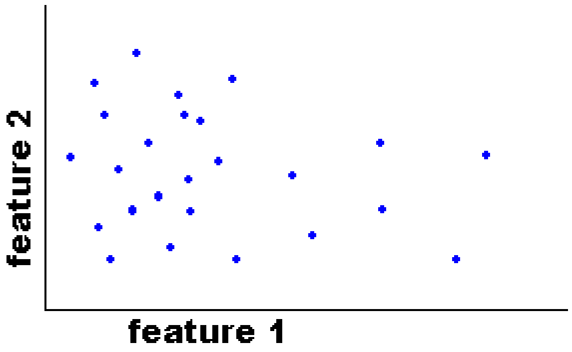


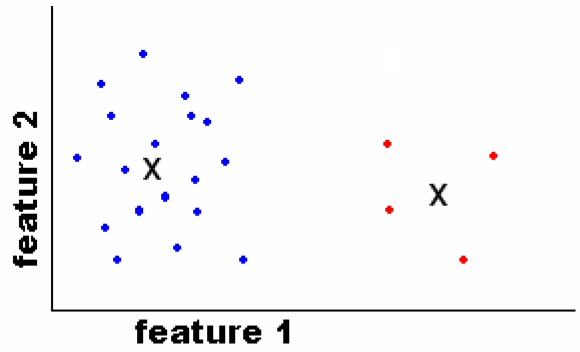
Adaptive binning

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

Clustering: very adaptive representations

Images from Dave Kauchak



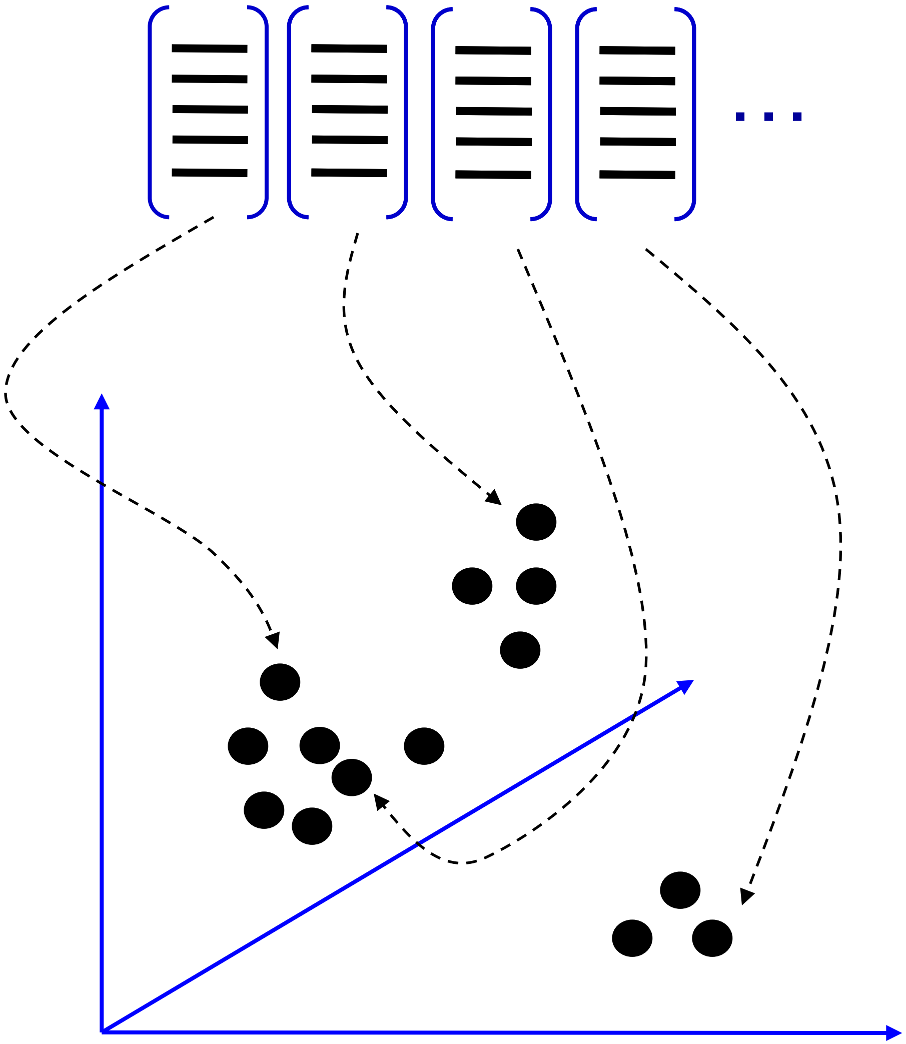


Clusters / Signatures

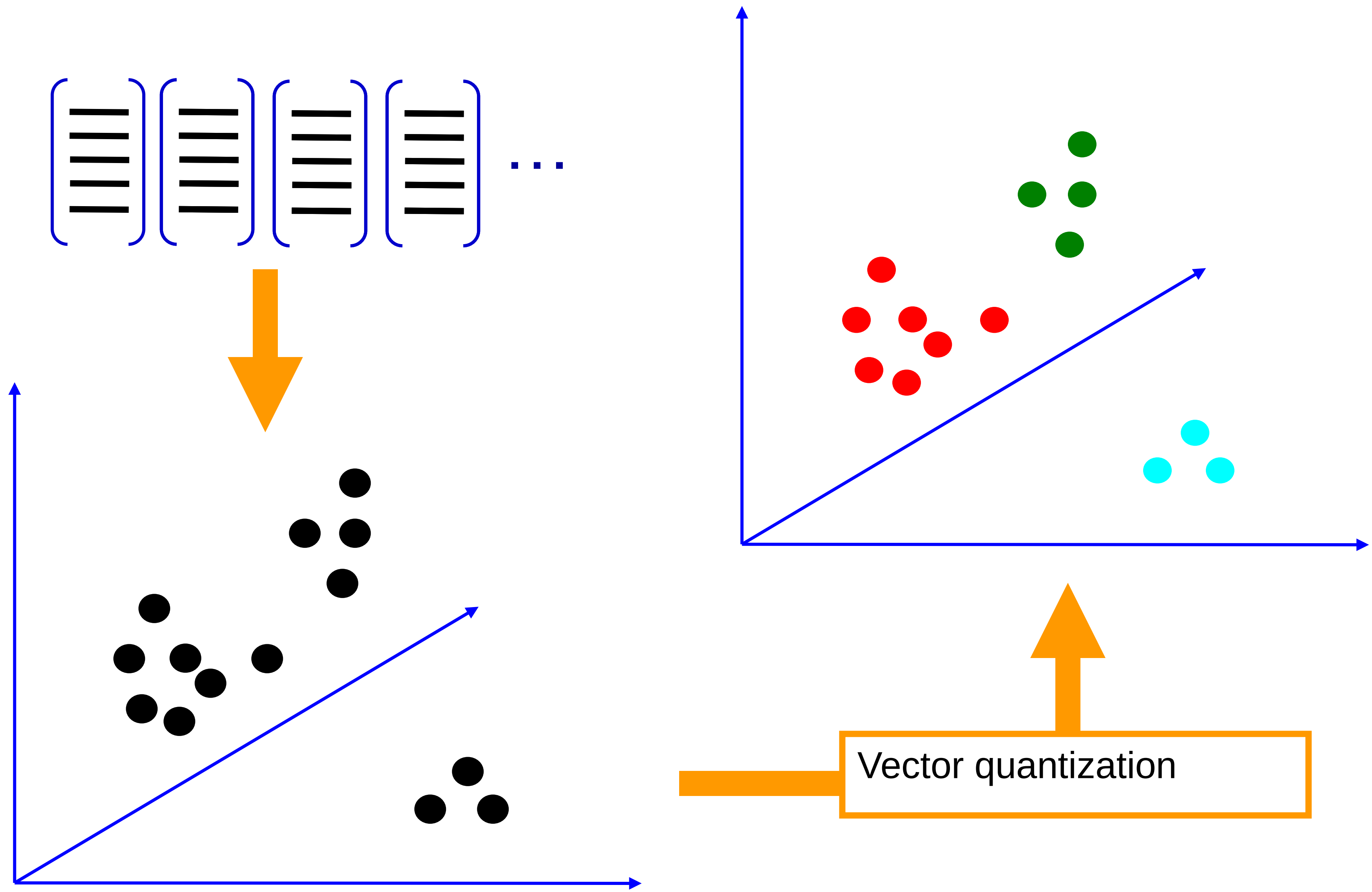
- “super-adaptive” binning
- Does not require discretization along any fixed axis

Patch Features

dictionary formation

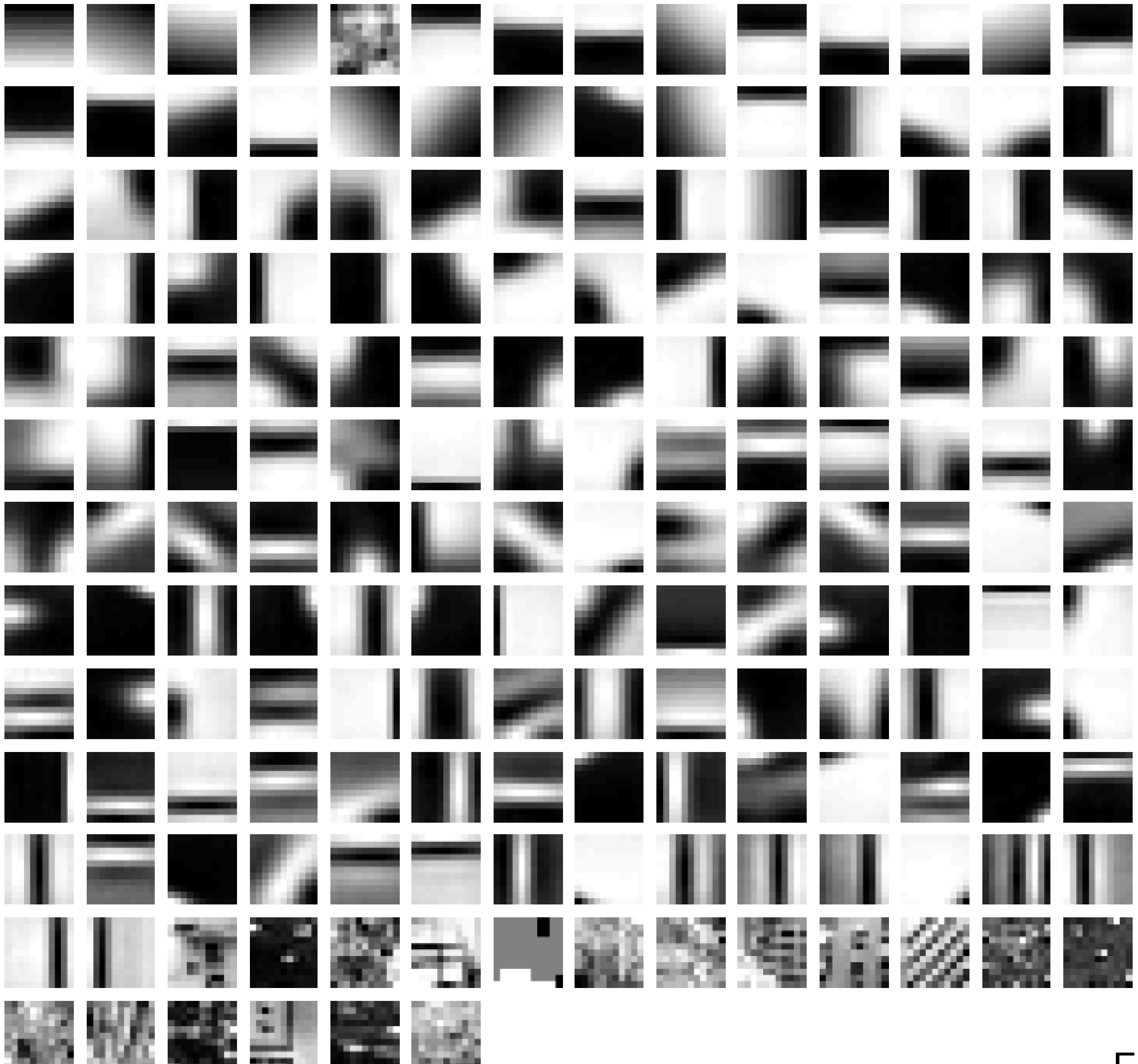


Clustering (usually k-means)



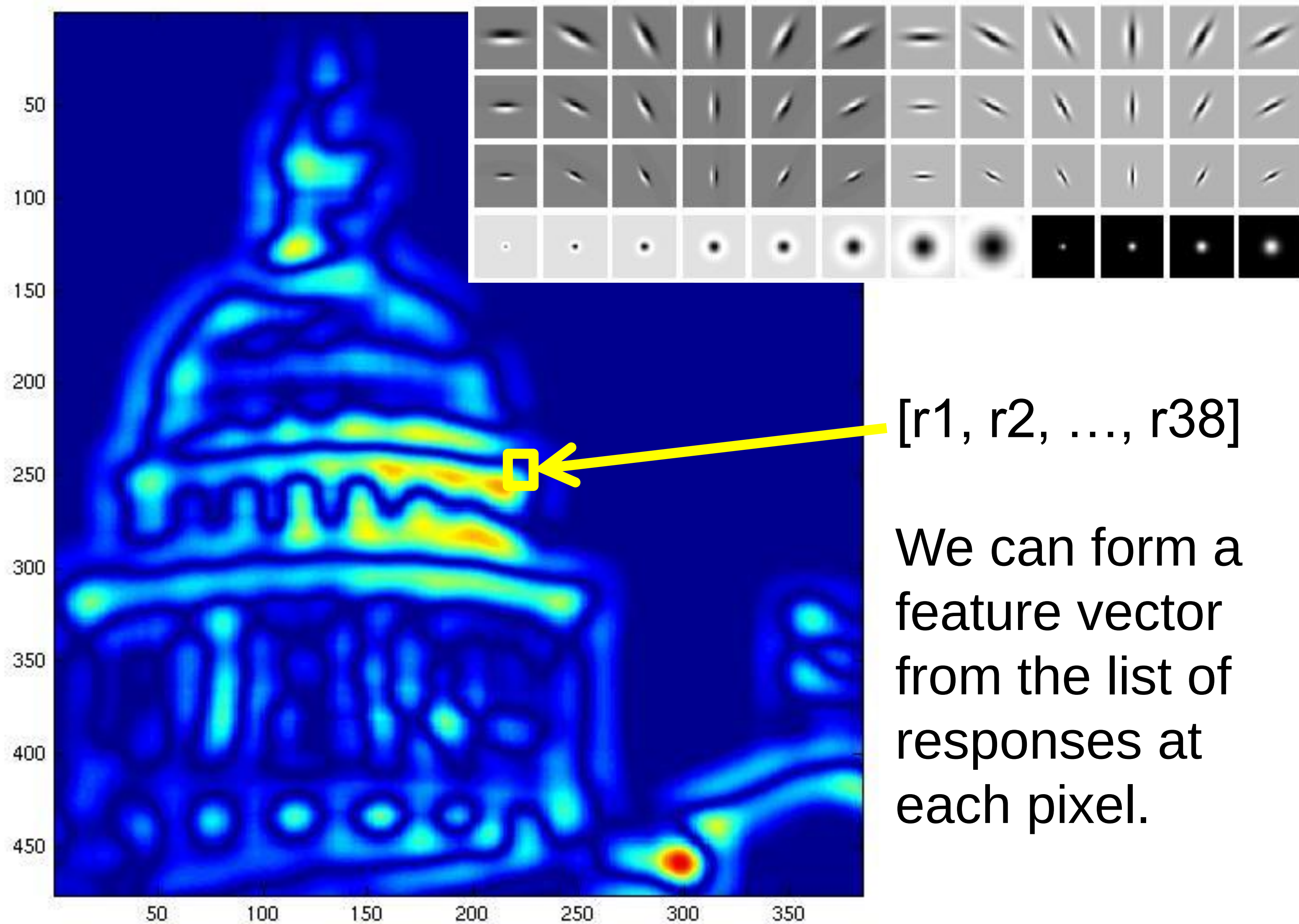
Slide credit: Josef Sivic

Clustered Image Patches



Fei-Fei et al. 2005

Filter Response Vectors



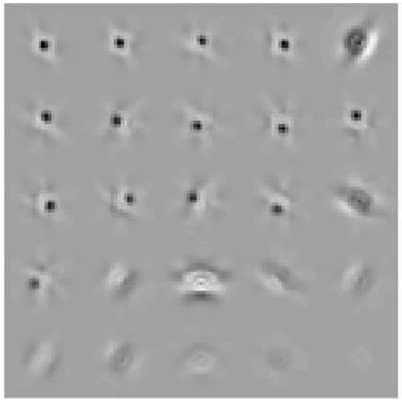
[r1, r2, …, r38]

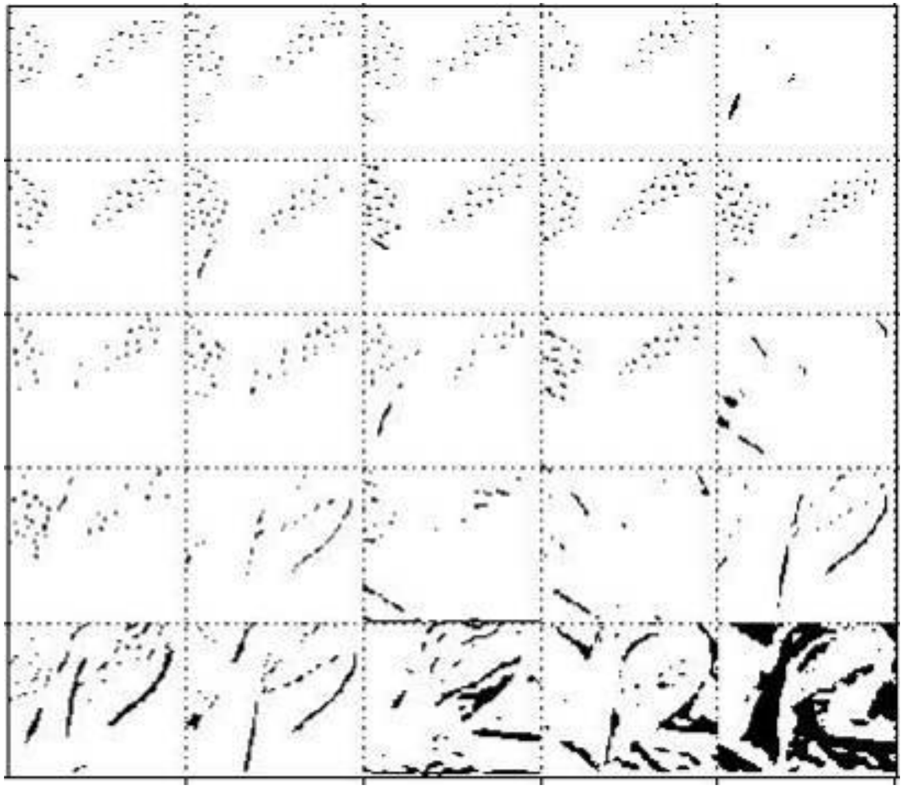
We can form a feature vector from the list of responses at each pixel.

Kristen Grauman

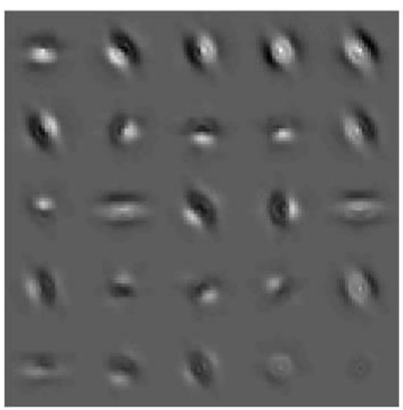
Textons (Malik et al, IJCV 2001)

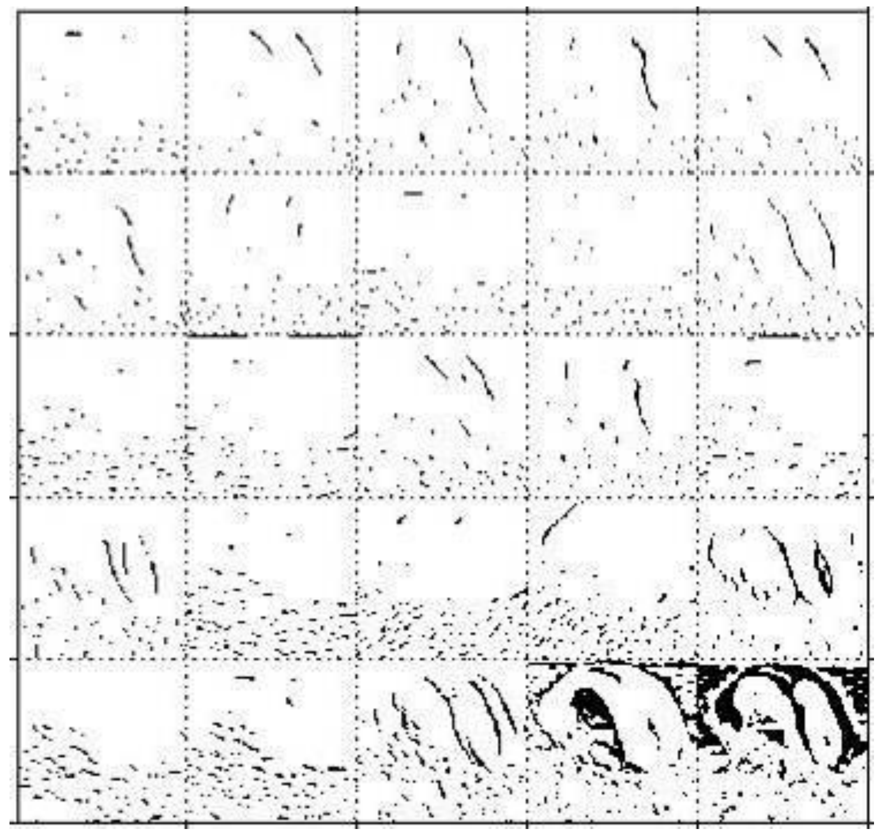
Cluster vectors of filter responses





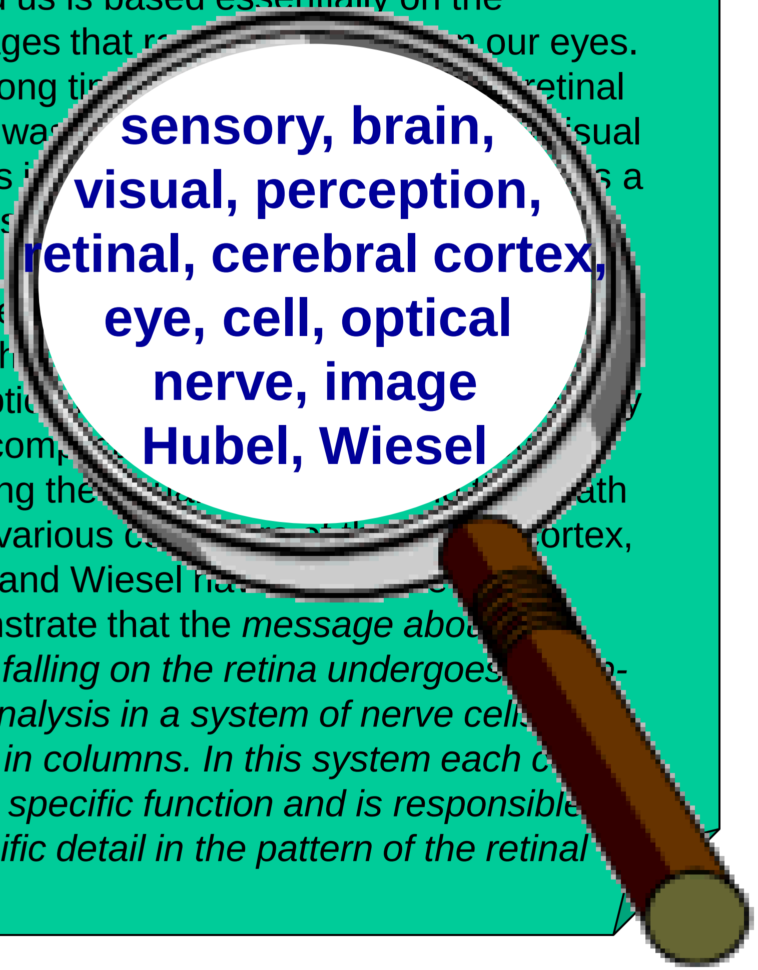
Textons (cont.)





“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 r... n our eyes. For a long ti... retinal image wa... isual centers i... s a movie s... image ... discove... know th... perceptio... y more comp... following the ... ath to the various c... ortex, Hubel and Wiesel ha... demonstrate that the *message abou... image falling on the retina undergoes ... -wise analysis in a system of nerve cells stored in columns. In this system each c... has its specific function and is responsible ... a specific detail in the pattern of the retinal image.*



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

Object → **Bag of ‘words’**

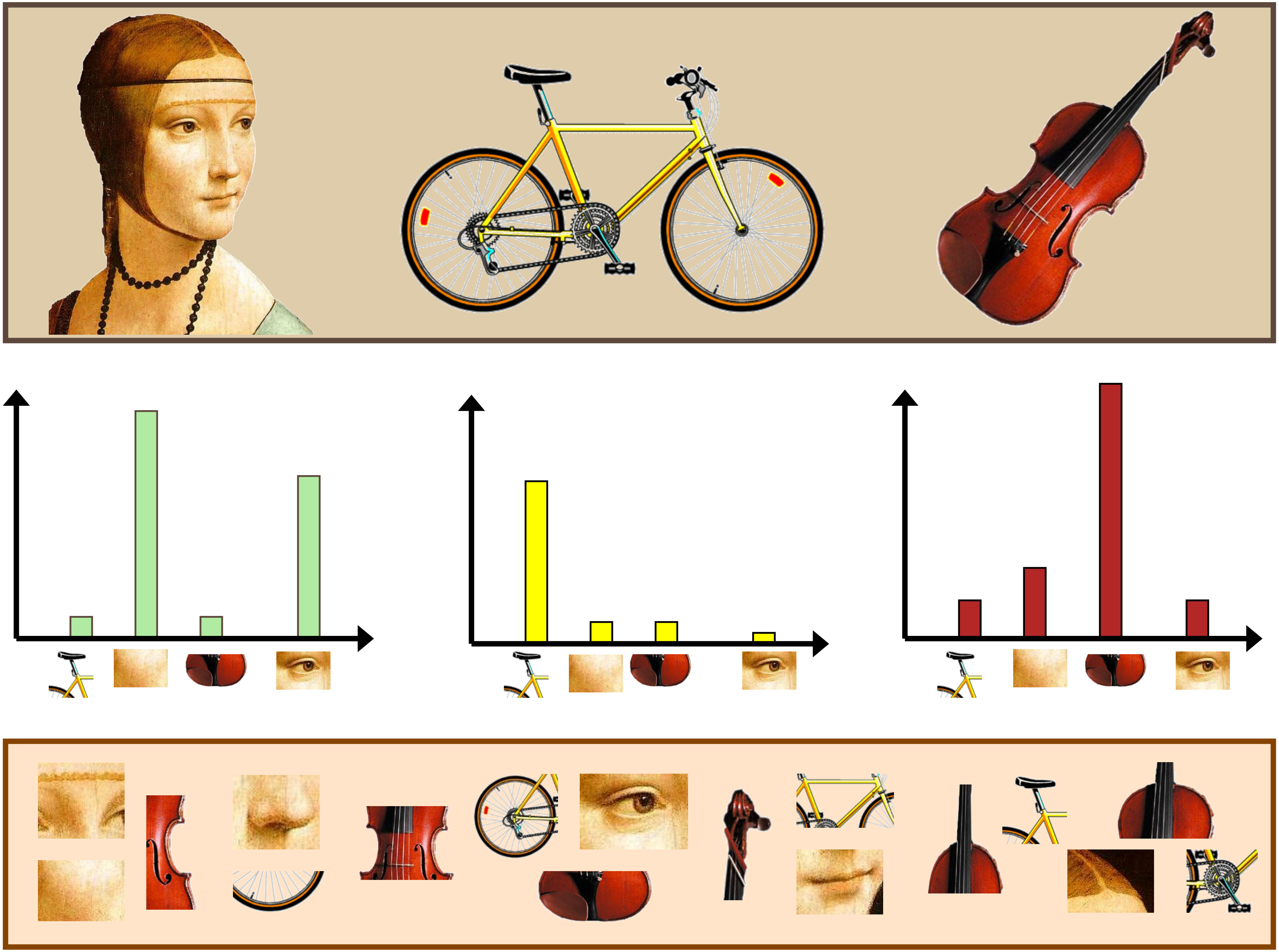
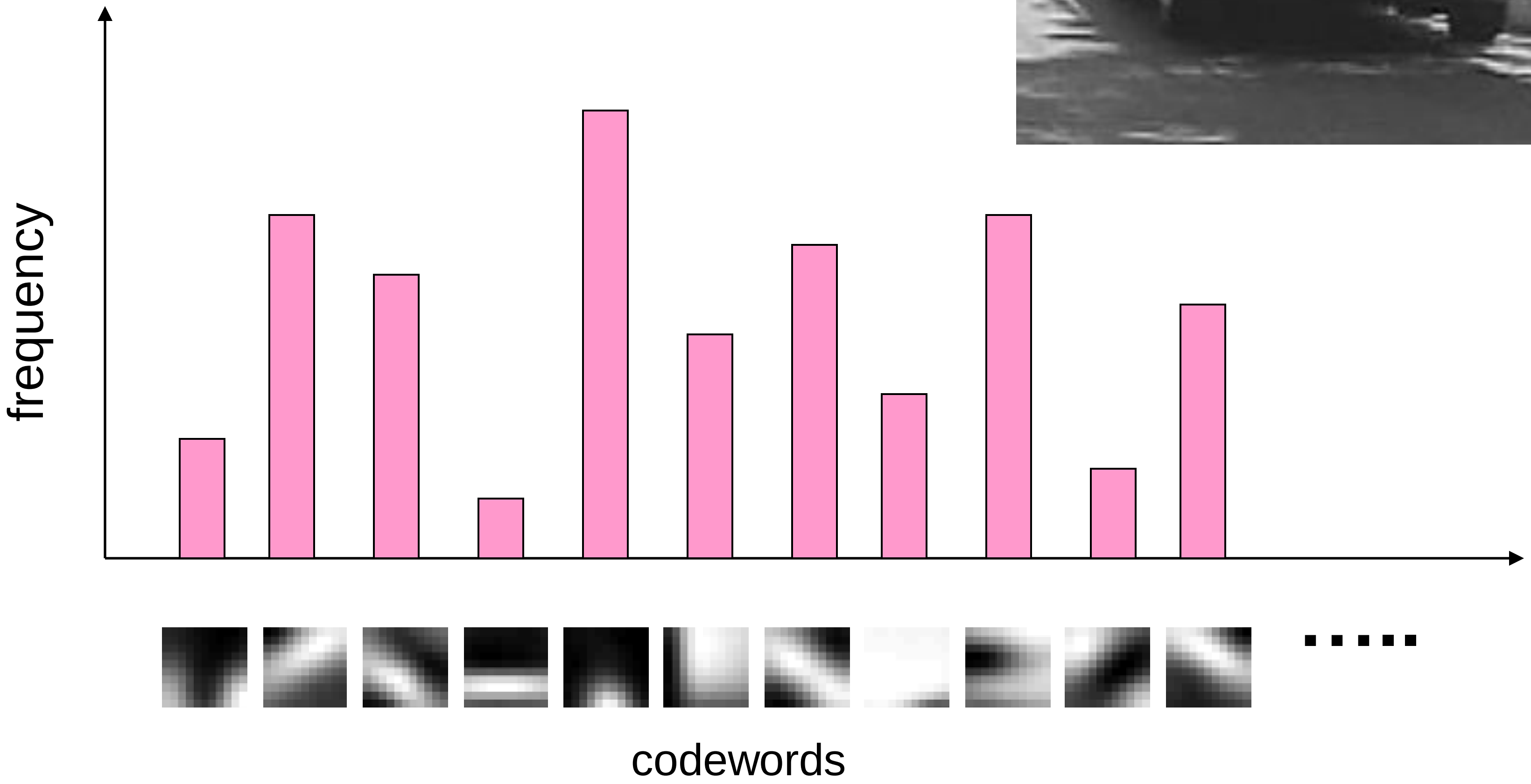


Image representation



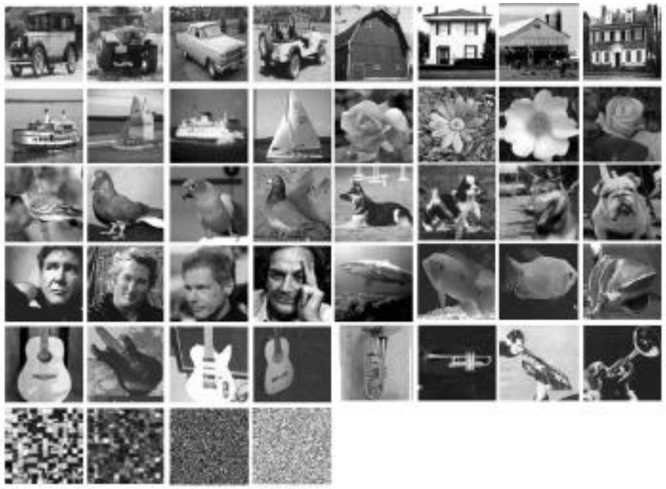
Scene Classification (Renninger & Malik)

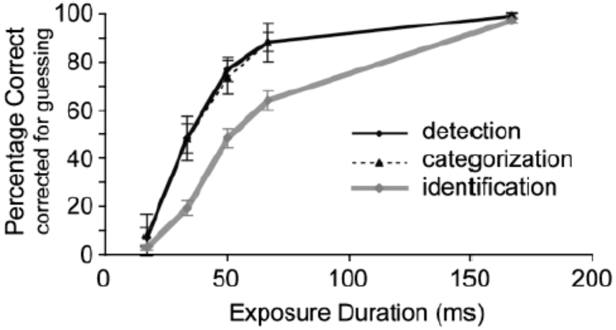
Image classification can be pre-attentive!

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

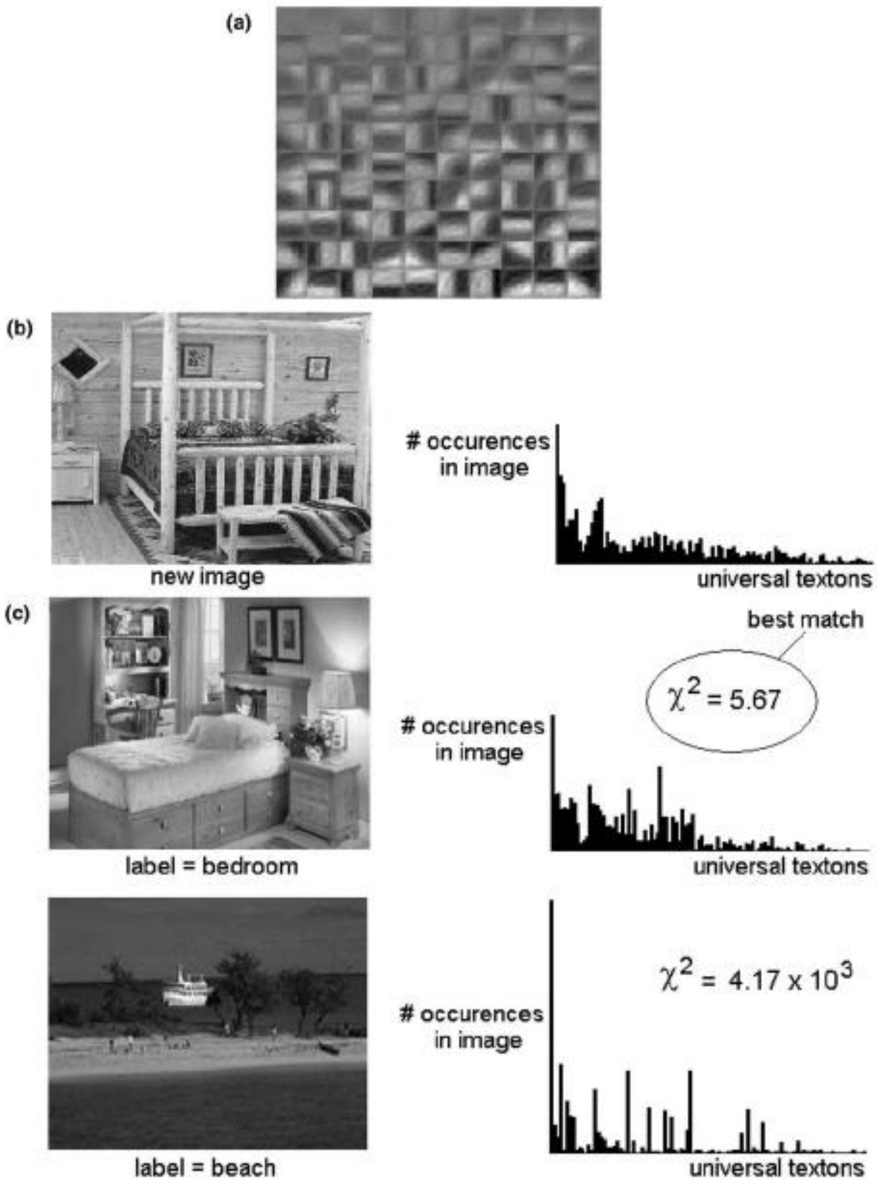
- Comparable to synaptic delay in the retina, LGN, V1, V2, V4, IT pathway.
- Doesn't rule out feed back but shows feed forward only is very powerful

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

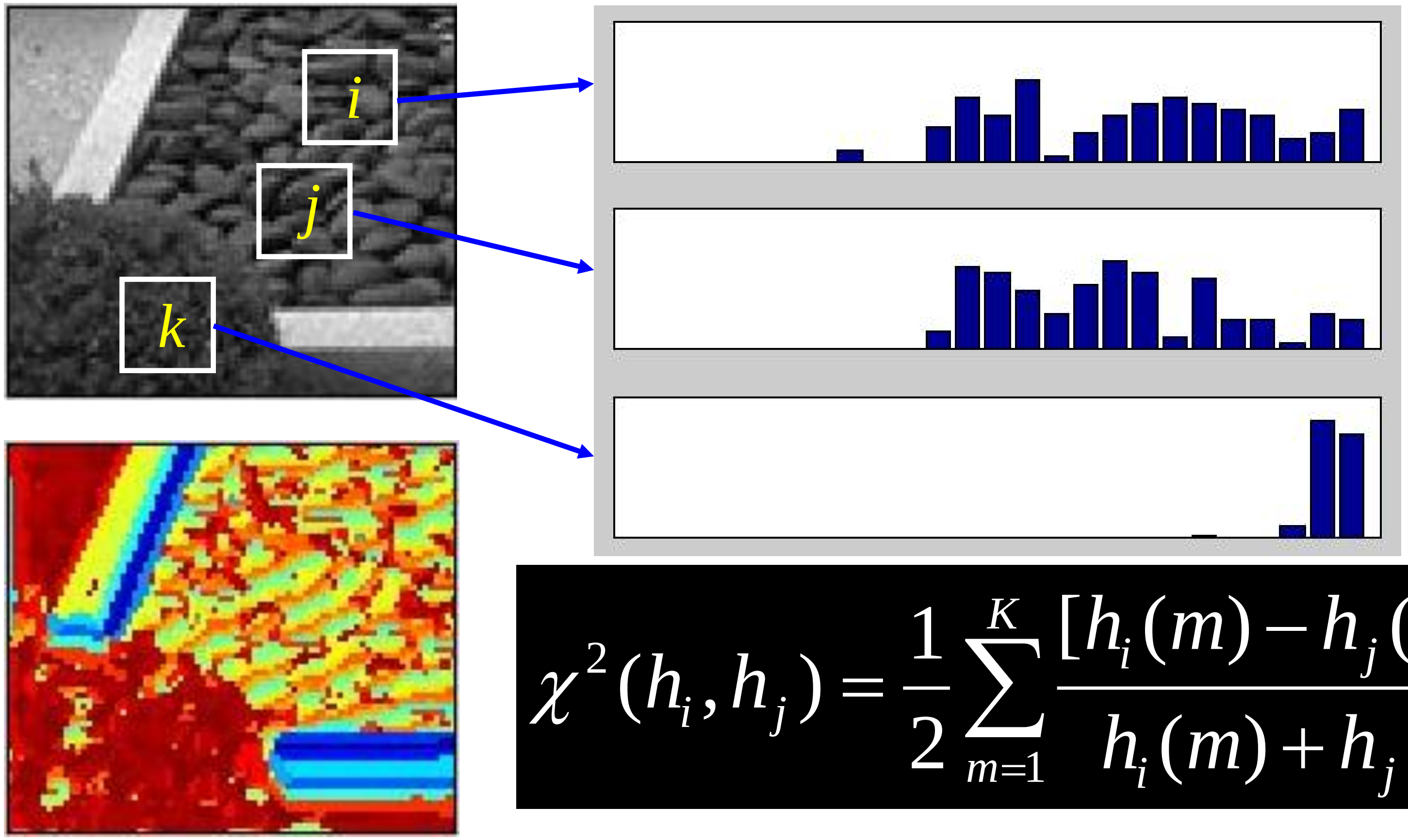




Texton Histogram Matching



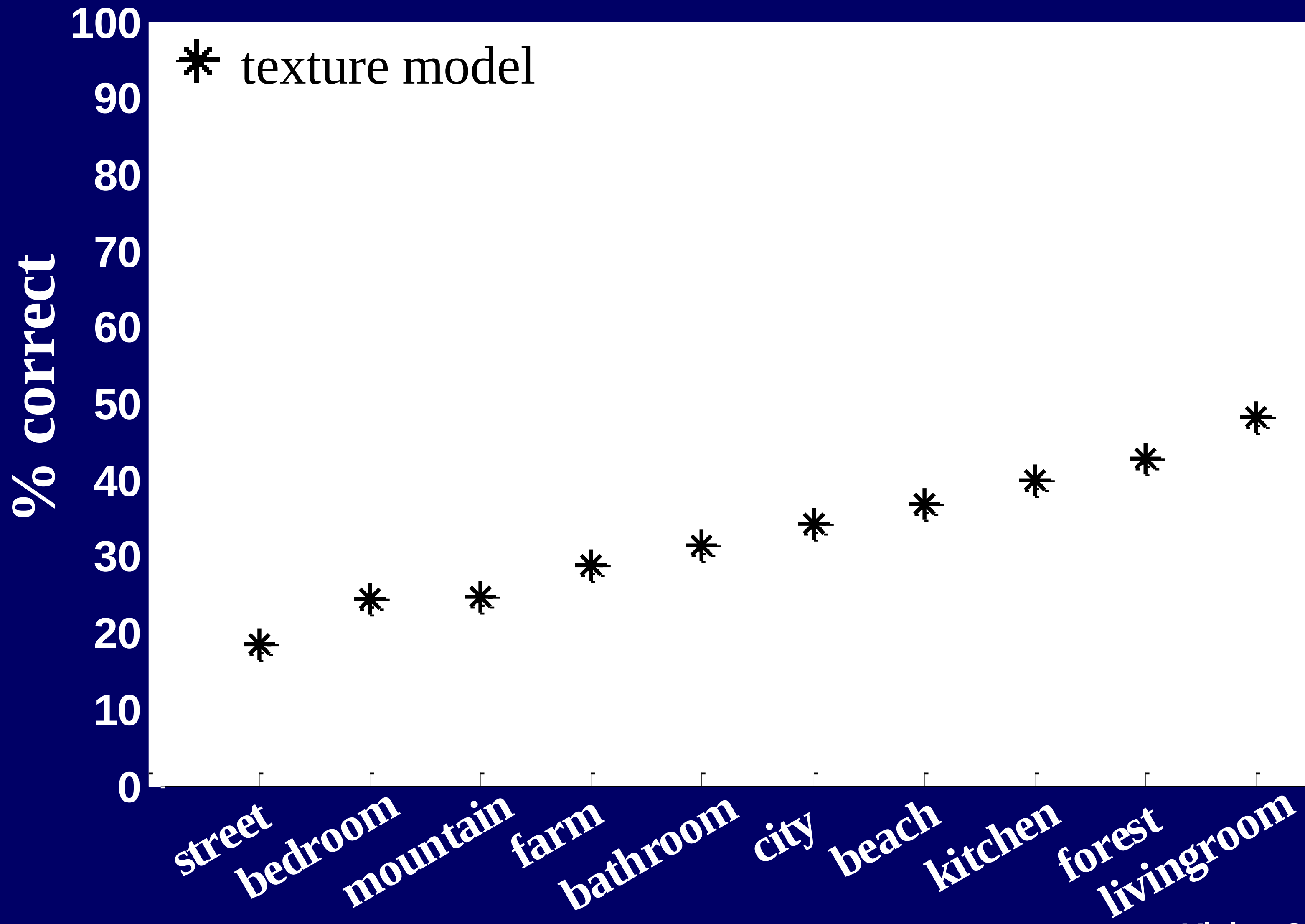
Texton histograms work on patches too



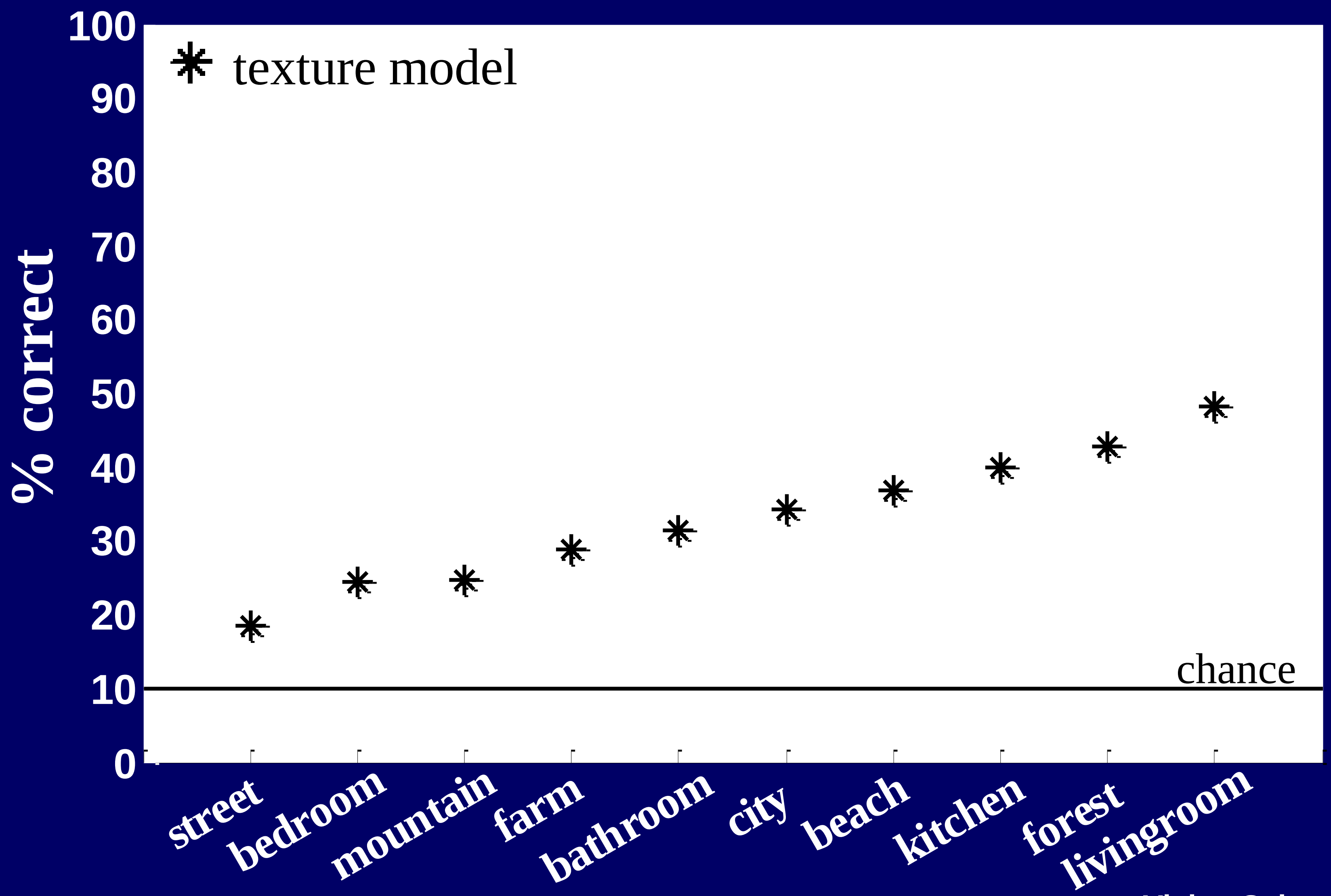
$$\chi^2(h_i, h_j) = \frac{1}{2}\sum_{m=1}^{K}\frac{[h_i(m) - h_j(m)]^2}{h_i(m) + h_j(m)}$$

(Malik et al, 99)

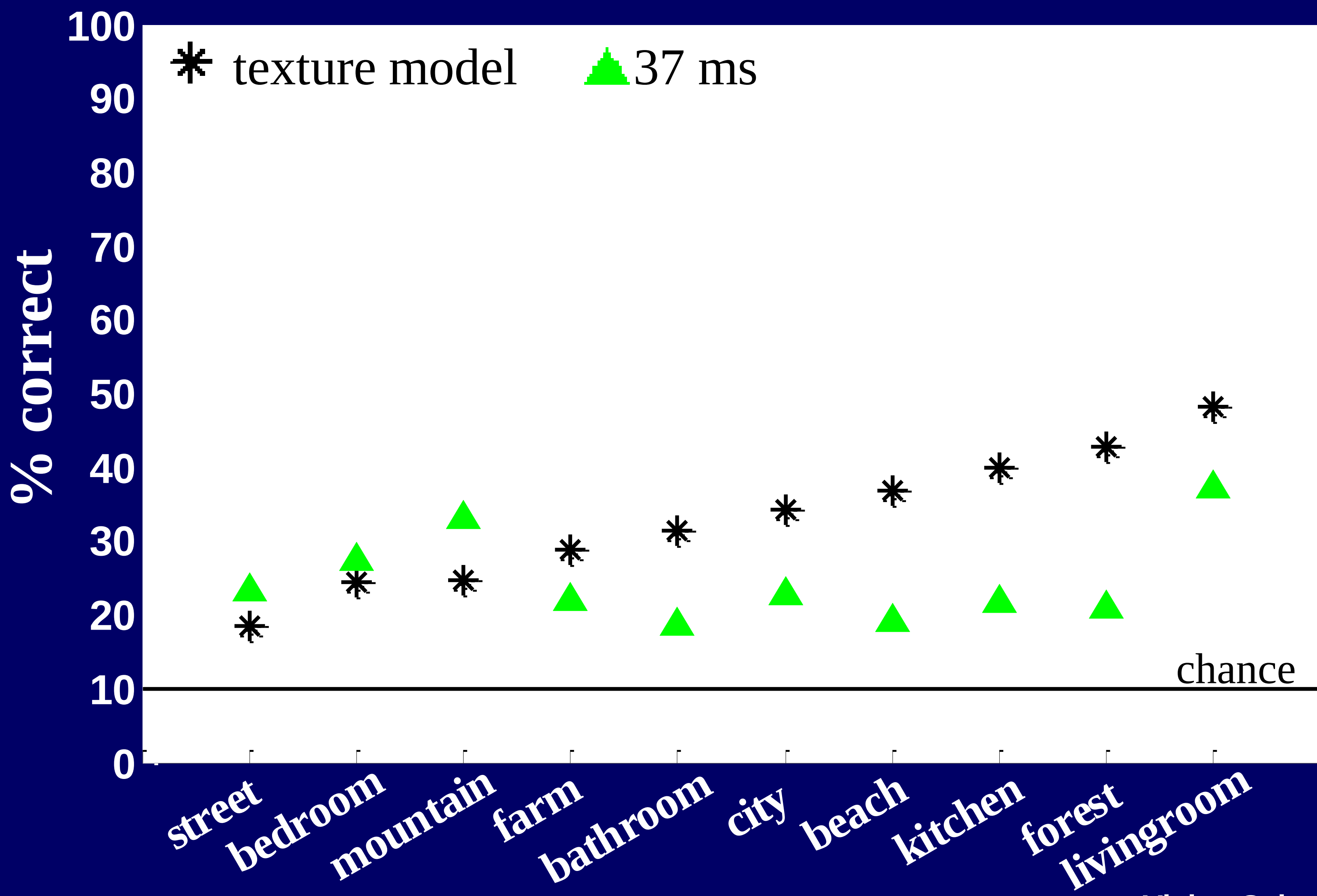
Discrimination of Basic Categories



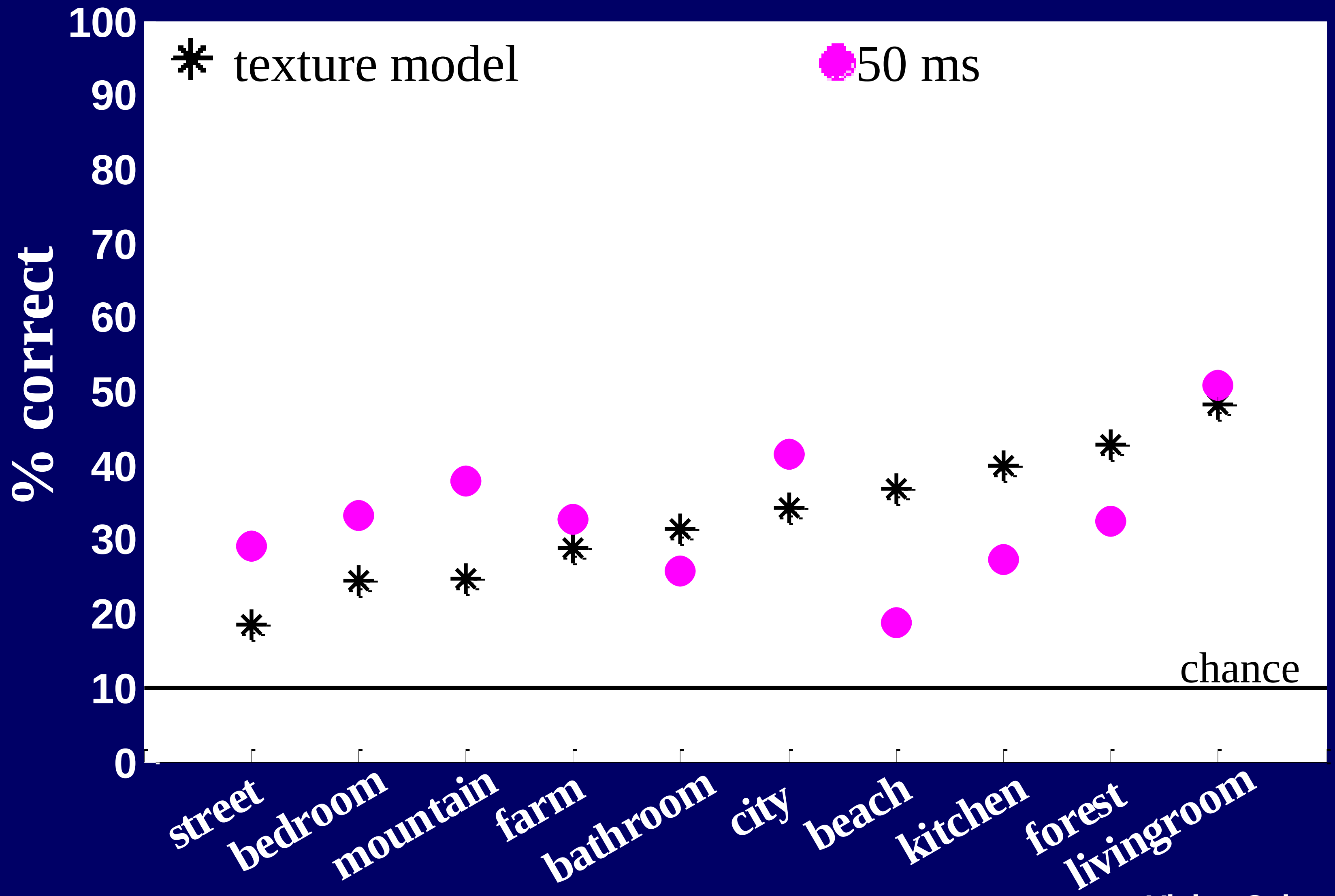
Discrimination of Basic Categories



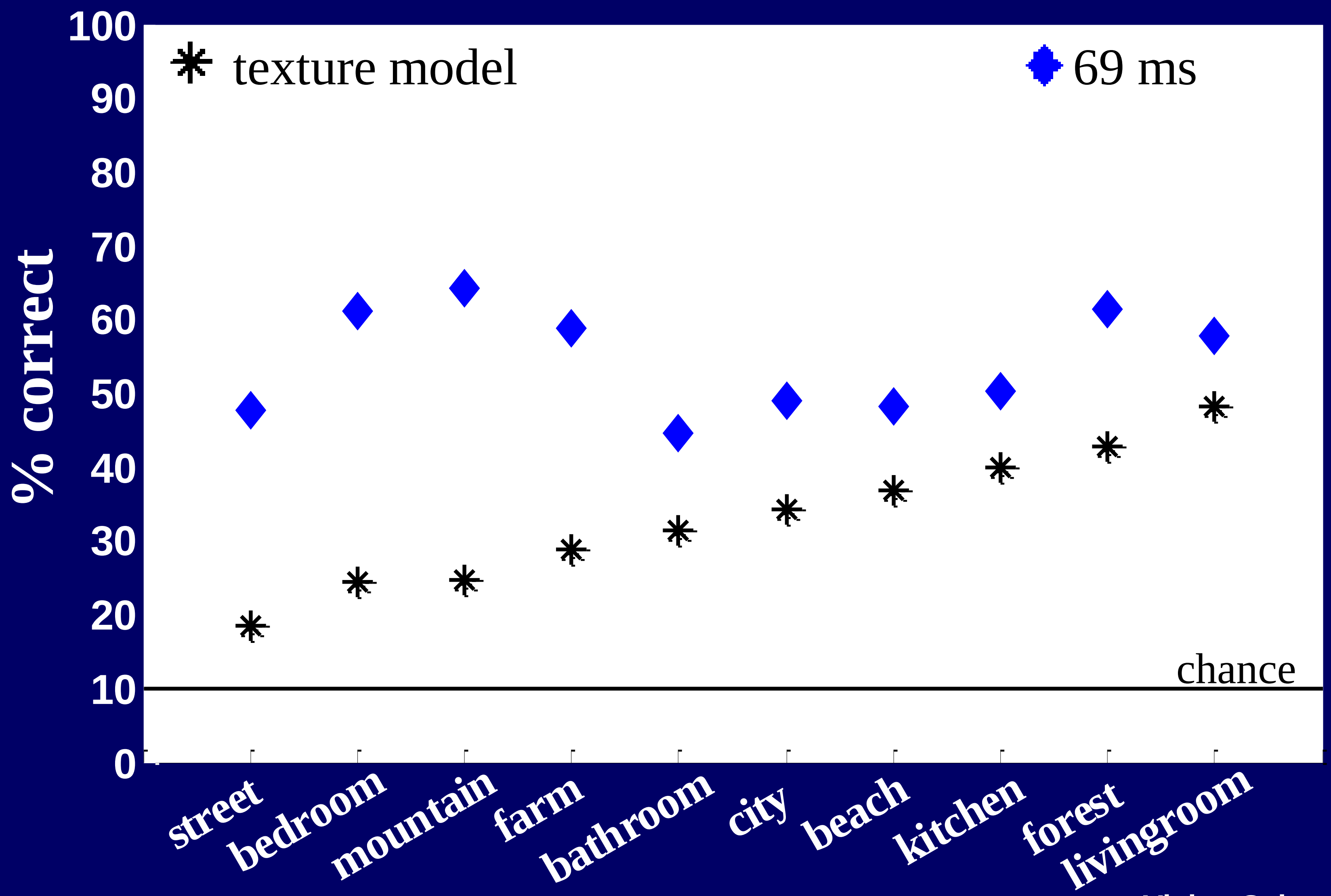
Discrimination of Basic Categories



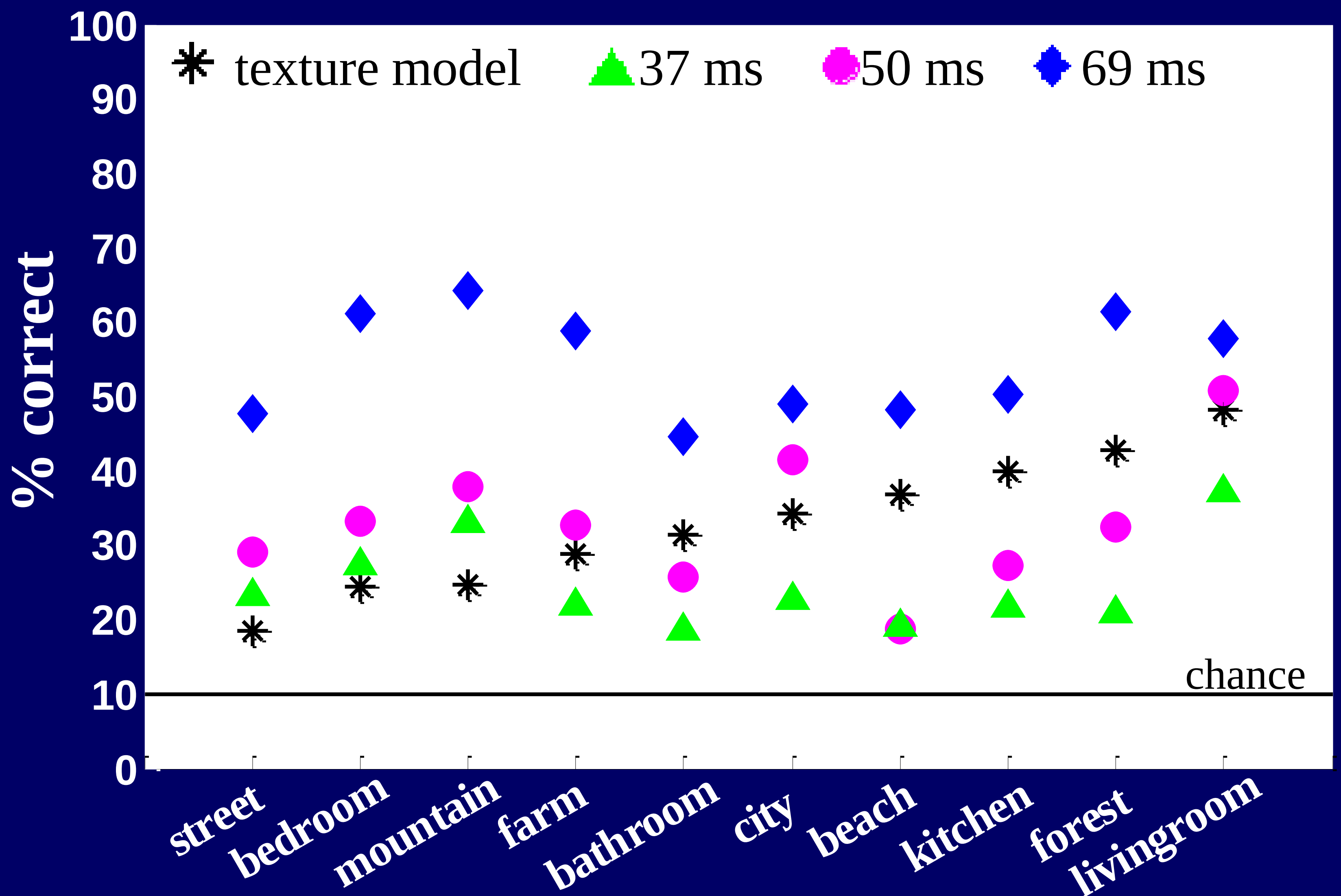
Discrimination of Basic Categories



Discrimination of Basic Categories



Discrimination of Basic Categories



Object Recognition using Texture