浙江大学计算机学院
数字媒体与网络技术

Digital Asset Management
数字媒体资源管理

6. Introduction to Digital Media Retrieval

任课老师：张宏鑫
2015-11-03

Origin of digital media retrieval

- IR (Information Retrieval)
  - To retrieve information that users want based on some keys or hints
  - Support:
    - daily life use
    - authoring
    - thinking and designing

Main methods of digital media retrieval

- **Text-based** digital media retrieval
  - –Boolean model
  - –Clustering model
  - –Vector model
  - –Probability model
- **Content-based** digital media retrieval
  - **–Query By Examples**
- **Semantic-based** digital media retrieval

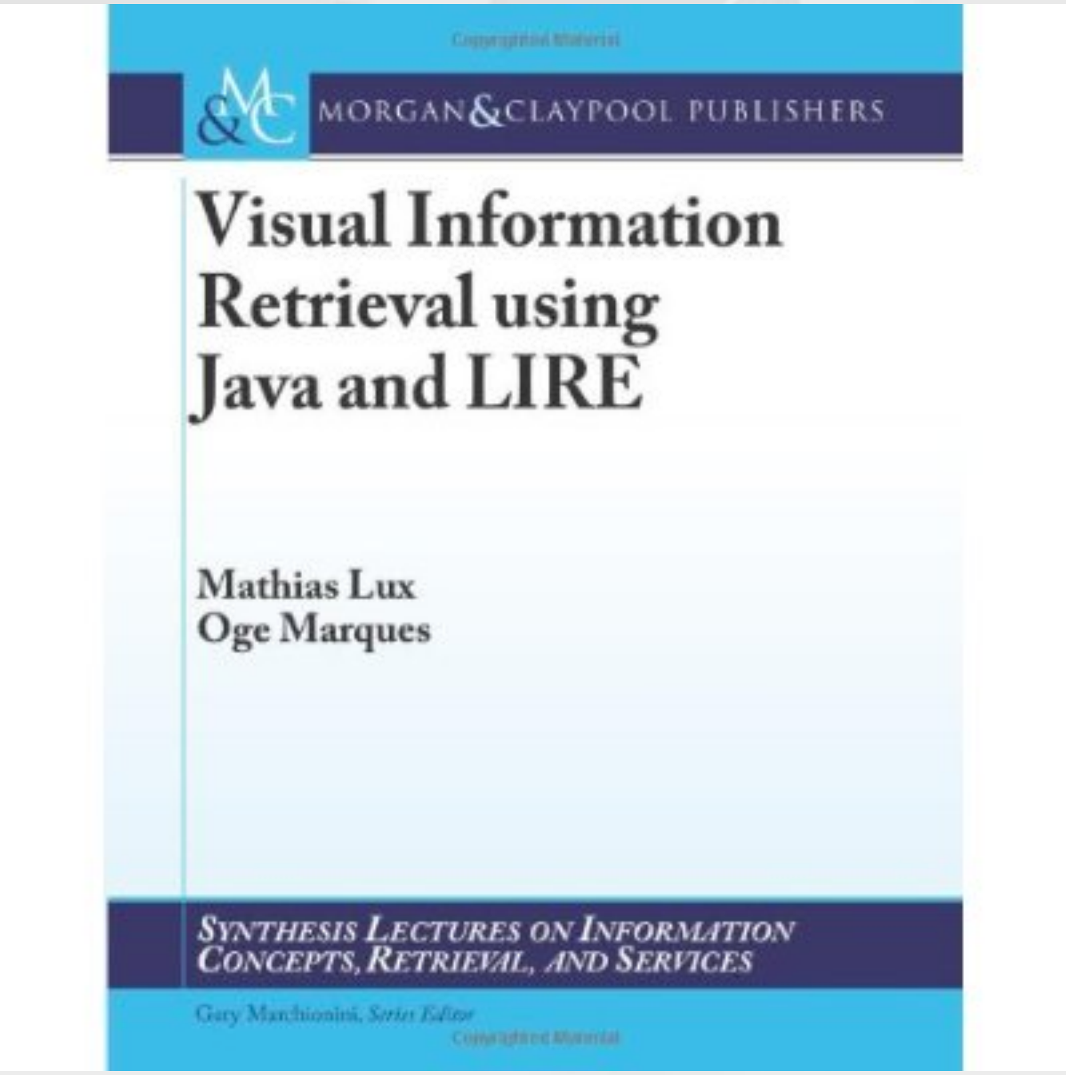
Content based digital media retrieval

- Query by example on multimedia-data

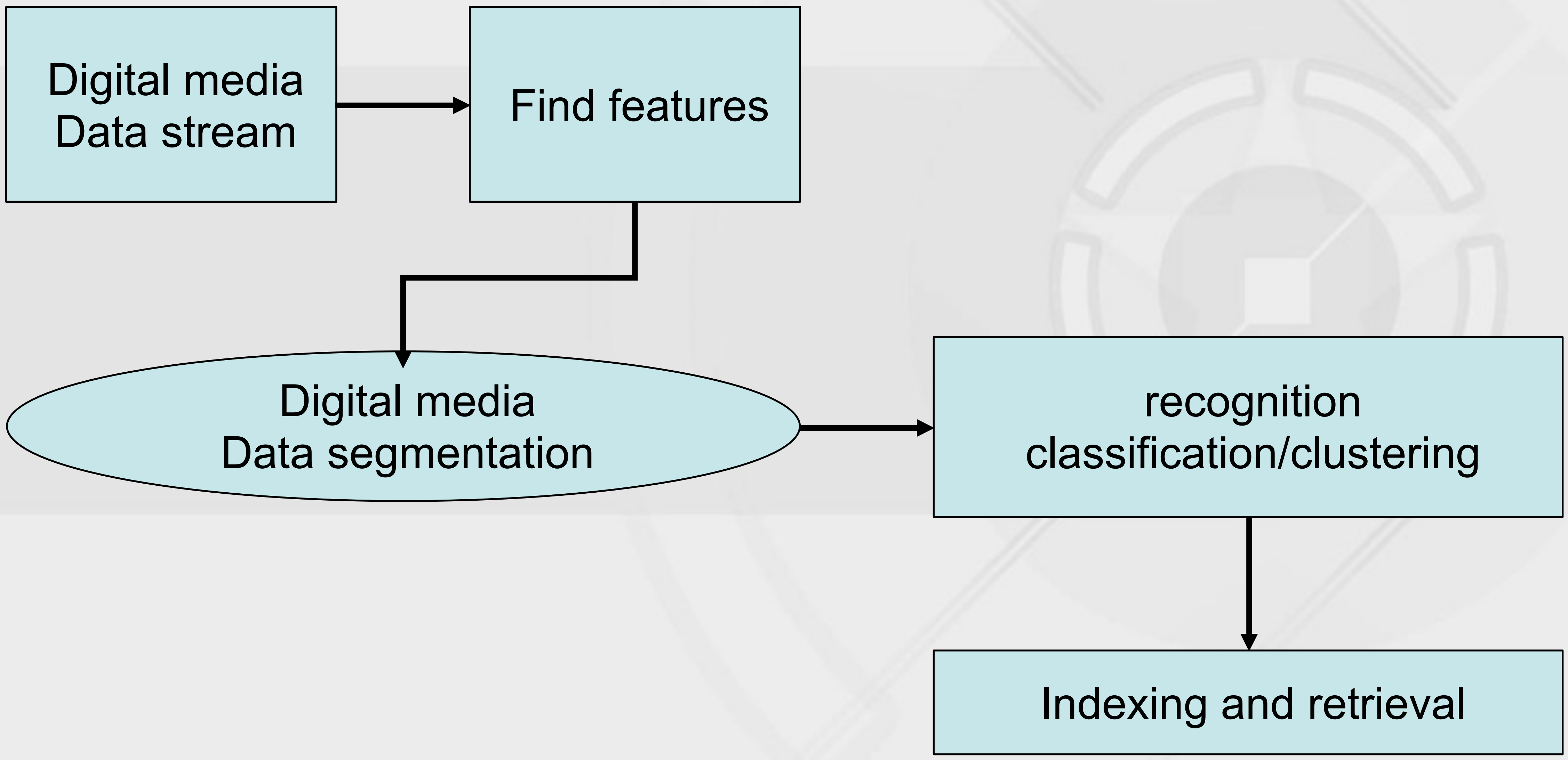
- Demo:
  - **The GNU Image-Finding Tool**
  - http://www.gnu.org/software/gift/

LIRE
Image Search Engine with Lucene

- https://code.google.com/p/lire/



The workflow of digital media analysis and retrieval



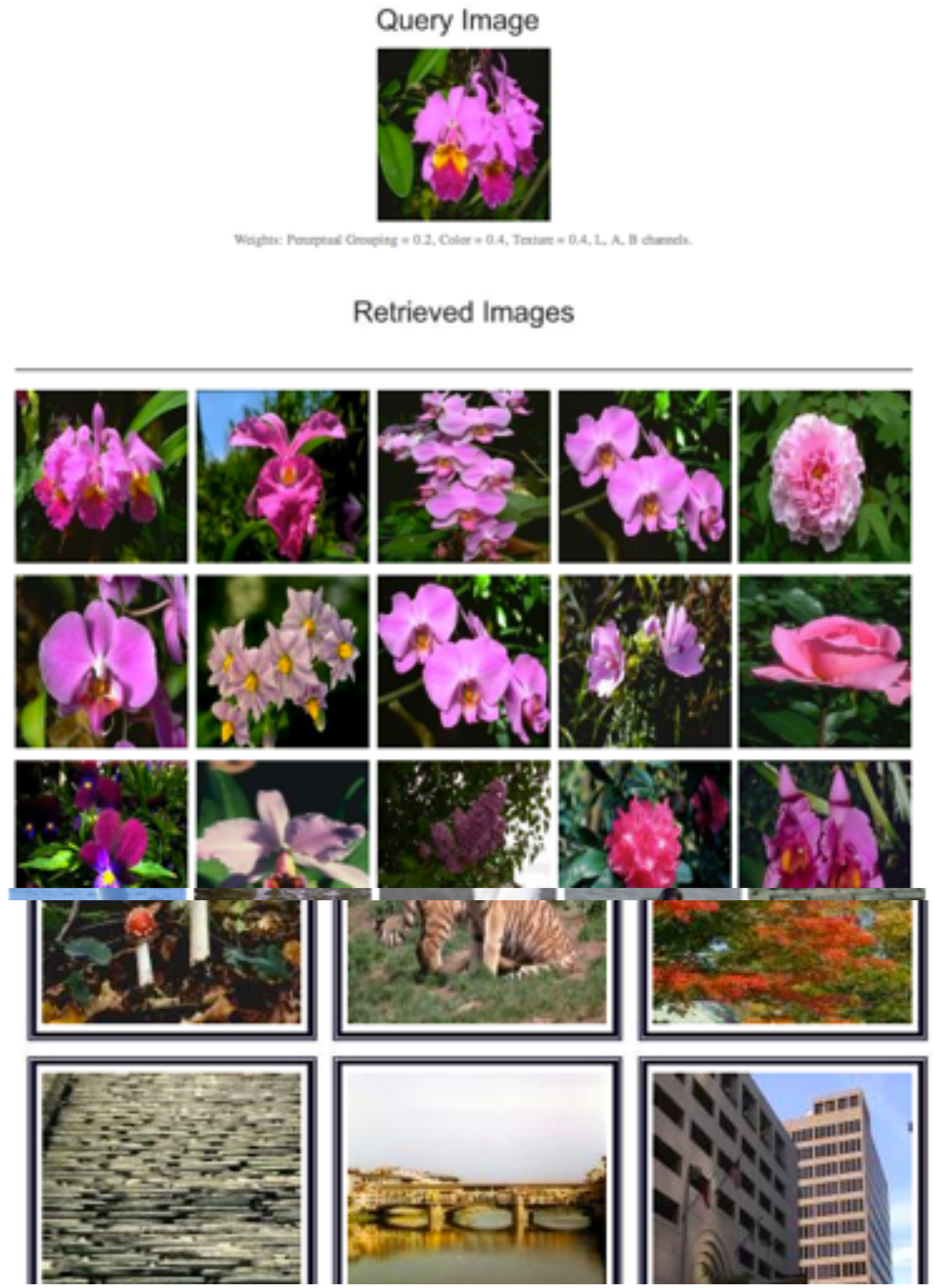
Content-based digital media retrieval

- In this lesson, we will know …
  - –Content-based image retrieval
  - –Content-based video retrieval
  - –Content-based audio retrieval
  - –Content-based graphics retrieval

  - –Merging and analysis of multiple media
  - –Development and challenging

1. Content-based image retrieval

CBIR



CIRES

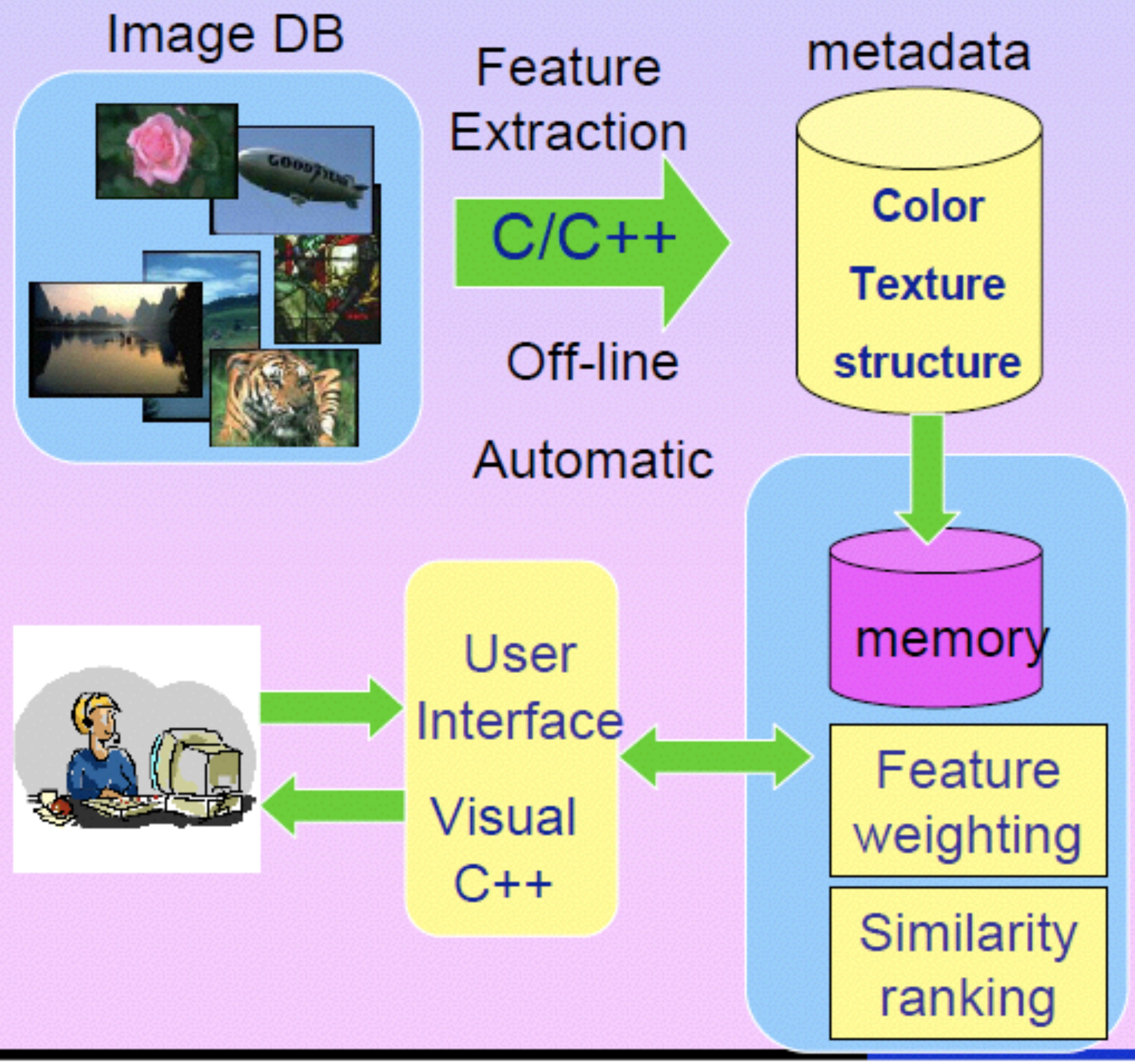
http://amazon.ece.utexas.edu/~qasim/sample_queries.htm

DEMO from the RGB group

iPad APP："服饰绘"

Multimedia Information Retrieval

- **Content-based Image Retrieval**



- **Color**
  - ✓Color histogram
  - ✓Color moments
  - ✓Color correlogram
- **Texture**
  - ✓Tamura texture
  - ✓Co-occurrence matrices
  - ✓Gabor features
  - ✓Wavelet moments
- **Shape**
  - ✓Fourier descriptor
- **Structure**
  - ✓Edge-based features

Workflow of CBIR

Images → Find features → (x1, x2, ..., xn)

Similarity measurement → recognition classification/clustering → Indexing and retrieval

(x1, x2, ..., xn)

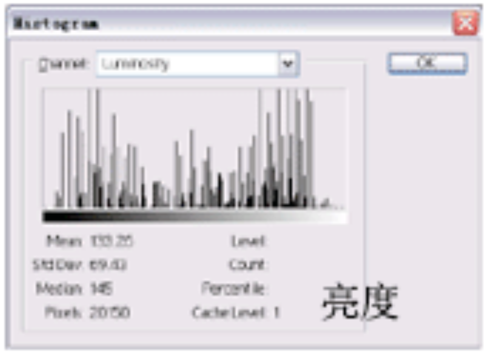
(y1, y2, ..., yn)

Features of image

- Finding out features of image is a key step of image retrieval
  - Image-based retrieval usually need to pre-construct feature database of images for retrieval
- Major image features:
  - Color features
  - Texture features
  - Shape features
  - Space relation features

Color features of image

- Color feature is a most widely used vision feature. It is mainly used to analyze color distributions in an image, including:
  - Color histogram
  - Color moments
  - Color set
  - Color clustering vectors
  - Color relation graph



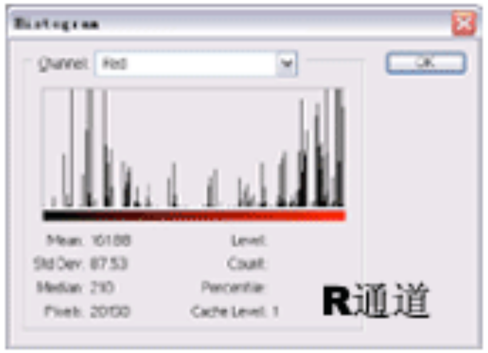
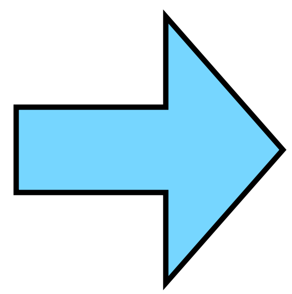


Image histogram



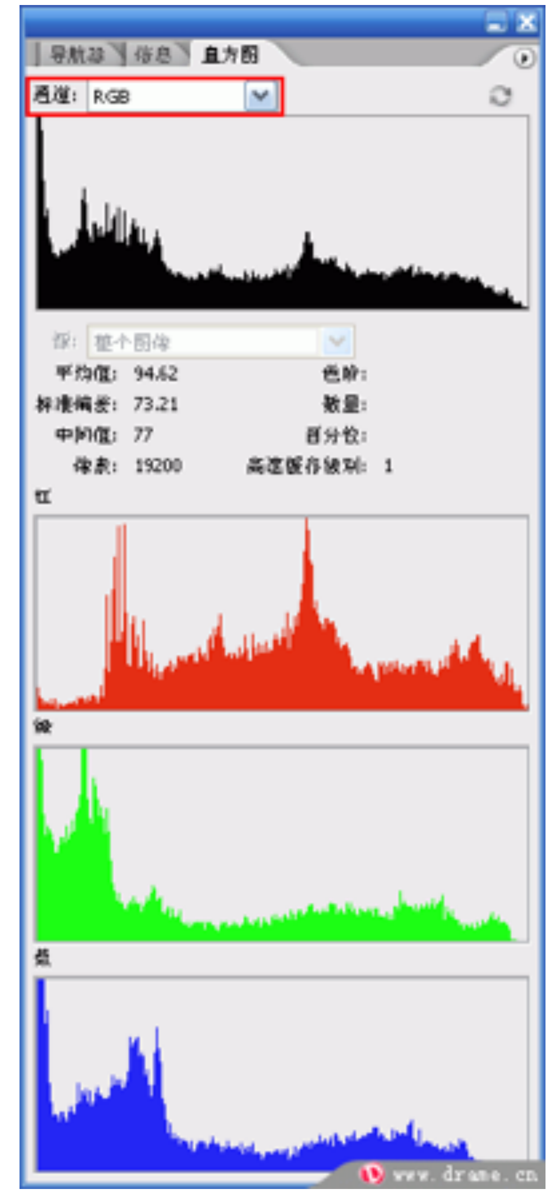
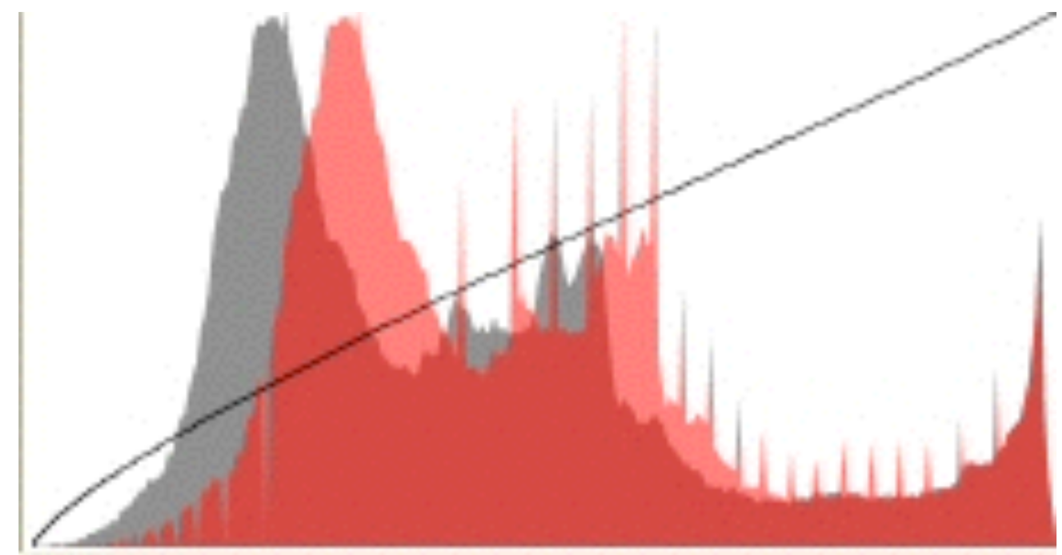


Image histogram



图像的颜色矩 (color moments)

- Color moments are global statistical features of an image, which are proposed by Stricker and Orengo.
  - First order moment (mean) $$\mu_i = \frac{1}{n}\sum_{j=1}^{n} I_{ij}$$
  - Second order moment (variance) $$\sigma_i^2 = \frac{1}{n}\sum_{j=1}^{n} (I_{ij} - \mu_i)^2$$
  - Third order moment (skewness) $$s_i^3 = \frac{1}{n}\sum_{j=1}^{n} (I_{ij} - \mu_i)^3$$
- Color moments are always applied with other image features for efficiently shrinking seeking ranges.

color moments: example

1	3	6	3	1
3	6	8	6	3
6	8	10	8	6
3	6	8	6	3
1	3	6	3	1

mean =4.72

variance =6.52

skewness =2.34

Image texture features

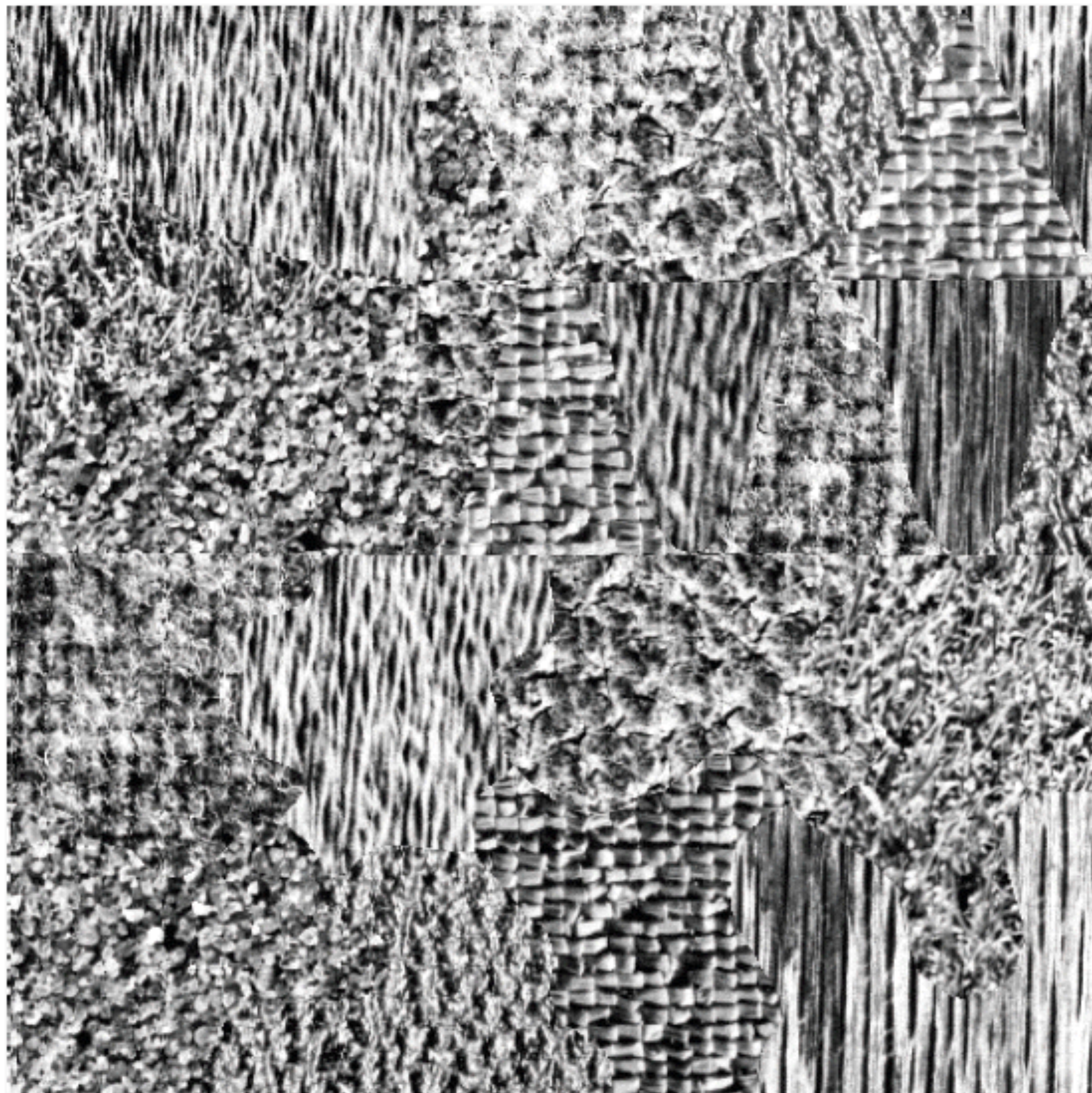


Image texture features

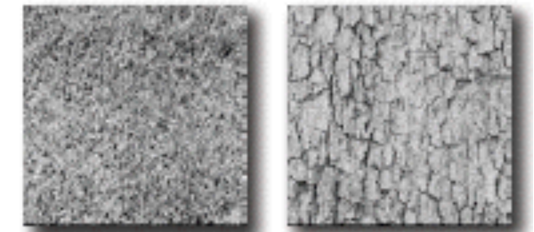
- Texture features are such vision features employed to measure homogeneous phenomenon in images. They are
  - independent to color or illuminance,
  - and are intrinsic features of object surfaces.

- Major texture features
  - Tamura texture features
  - Self-regression texture model
  - Transform based texture features
    - DWT，DFT，Garbor filter bank
  - others

Tamura texture features

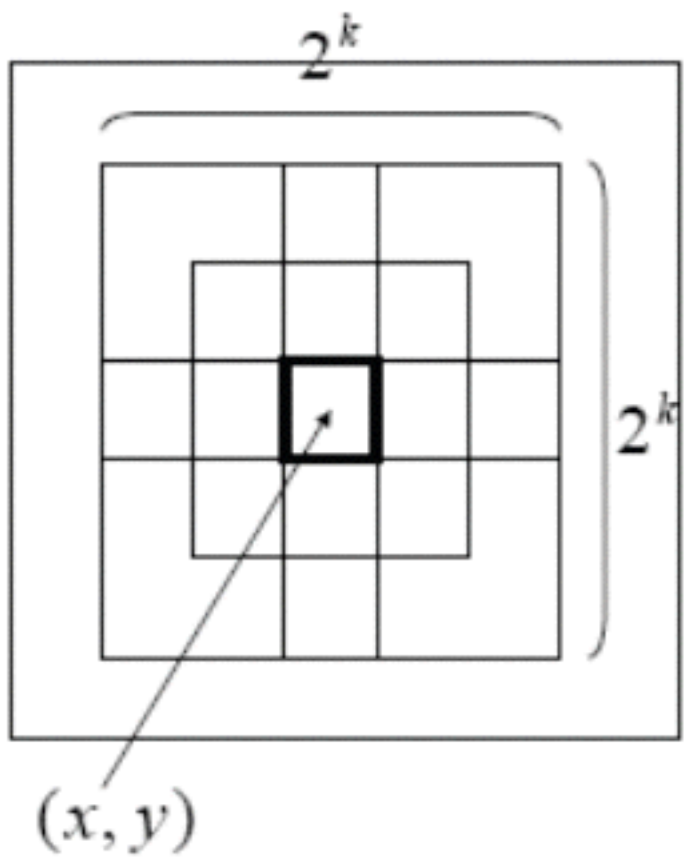
- a set of texture feature representation based on the psychology research results on human vision cognition of textures:
  - coarseness (粗糙度)
  - contrast (对比度)
  - directionality (方向度)
  - line-likeness (线相似度)
  - regularity (规整度)
  - roughness (粗略度)

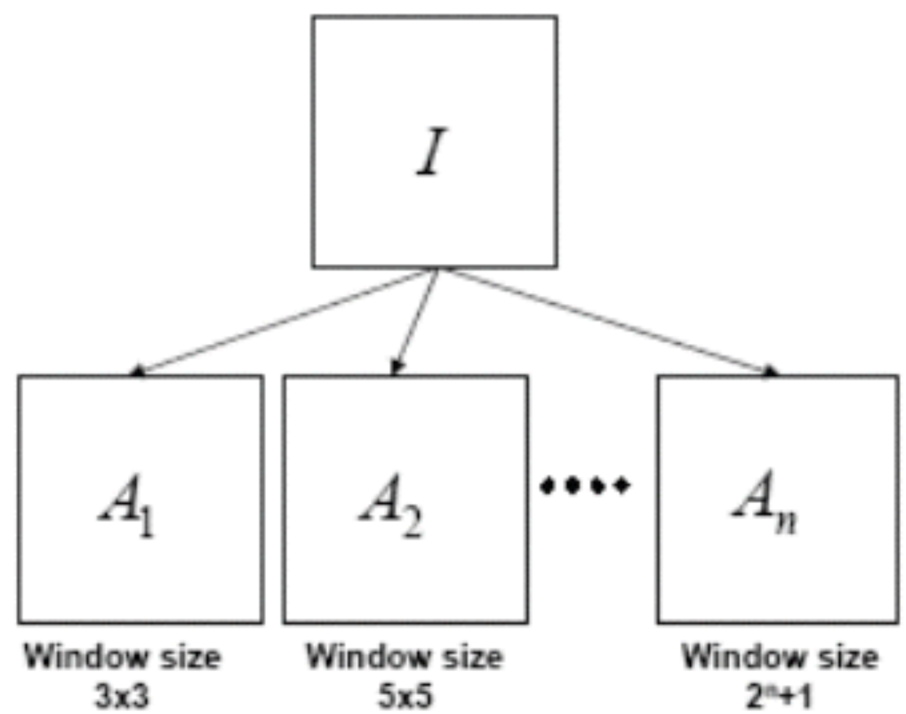
Tamura – Coarseness



- Goal
  - Pick a large size as best when coarse texture is present, or a small size when only fine texture
- Step 1: Compute averages at different scales at every points

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} I(i,j)/2^{2k}$$

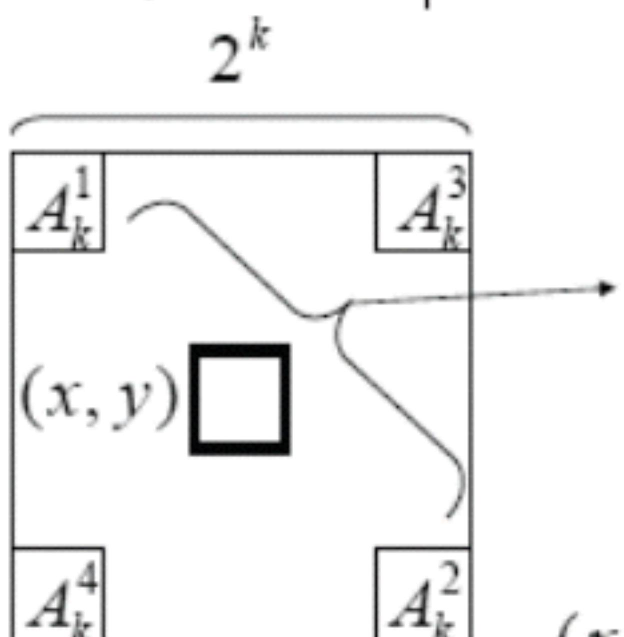




Tamura – Coarseness (cont.)

- Step 2: compute neighborhood difference at each scale on opposite sides of different directions

$$E_{k,h}(x,y) = \left| A_k(x-2^{k-1}, y) - A_k(x+2^{k-1}, y) \right|$$



$$E_{k,a}(x,y) = \left| A_k^{\ 1} - A_k^{\ 2} \right|$$

$$E_{k,b}(x,y) = \left| A_k^{\ 3} - A_k^{\ 4} \right|$$

$$(x,y) \longrightarrow \{E_{1,a}, E_{1,b}, E_{2,a}, E_{2,b}, \ldots, E_{n,a}, E_{n,b}\}$$

Tamura – Coarseness (cont.)

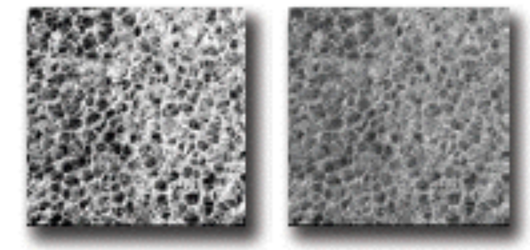
- Step 3: select the scale with the largest variation

$$S_{max}(x, y) = 2^k \quad / \quad E_k = max\{E_1, E_2, \ldots, E_L\}$$

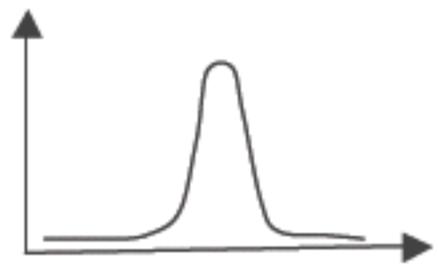
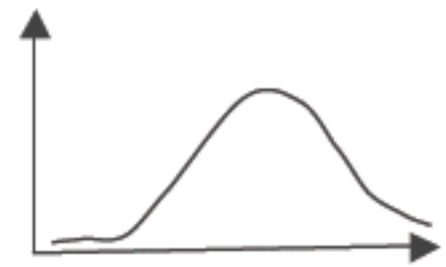
- Step 4: compute the coarseness

$$M_{\text{crs}} = \frac{1}{n \times m} \sum_{i}^{n} \sum_{j}^{m} S_{max}(i, j)$$

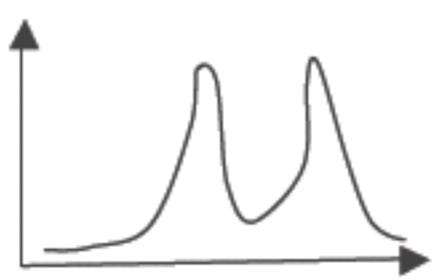
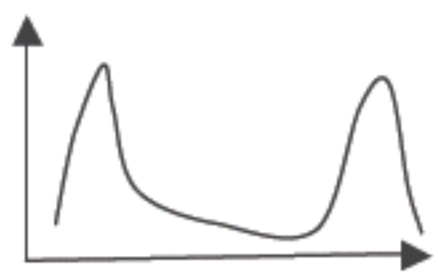
Tamura – Contrast



- Gaussian-like histogram distribution → low contrast

- Histogram polarization. Is it Gaussian? How many peaks it has? Where they are?

- Polarization can be estimated by the kurtosis (曲率度)

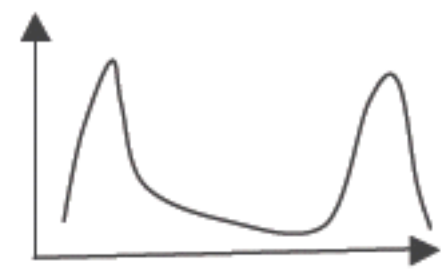
$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

$$\mu_4 = E[I^4(x, y)]$$

$$\sigma^4 = E[(I(x, y) - \mu)^4]$$

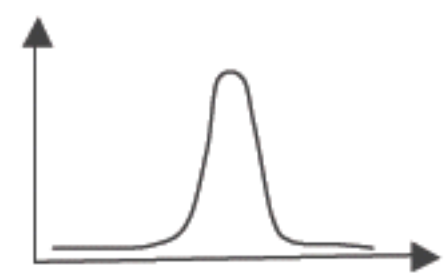
Tamura – Contrast (cont.)

$$\alpha_4 = \frac{\mu_4}{\sigma^4} \downarrow$$



distribution with two separate peaks

$$\alpha_4 = \frac{\mu_4}{\sigma^4} \uparrow$$

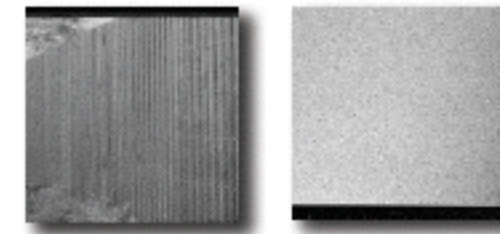


unimodal distribution

- Contrast estimate is given by:

$$M_{contrast} = \frac{\sigma}{(\alpha_4)^{\frac{1}{4}}}$$

Tamura – Orientation

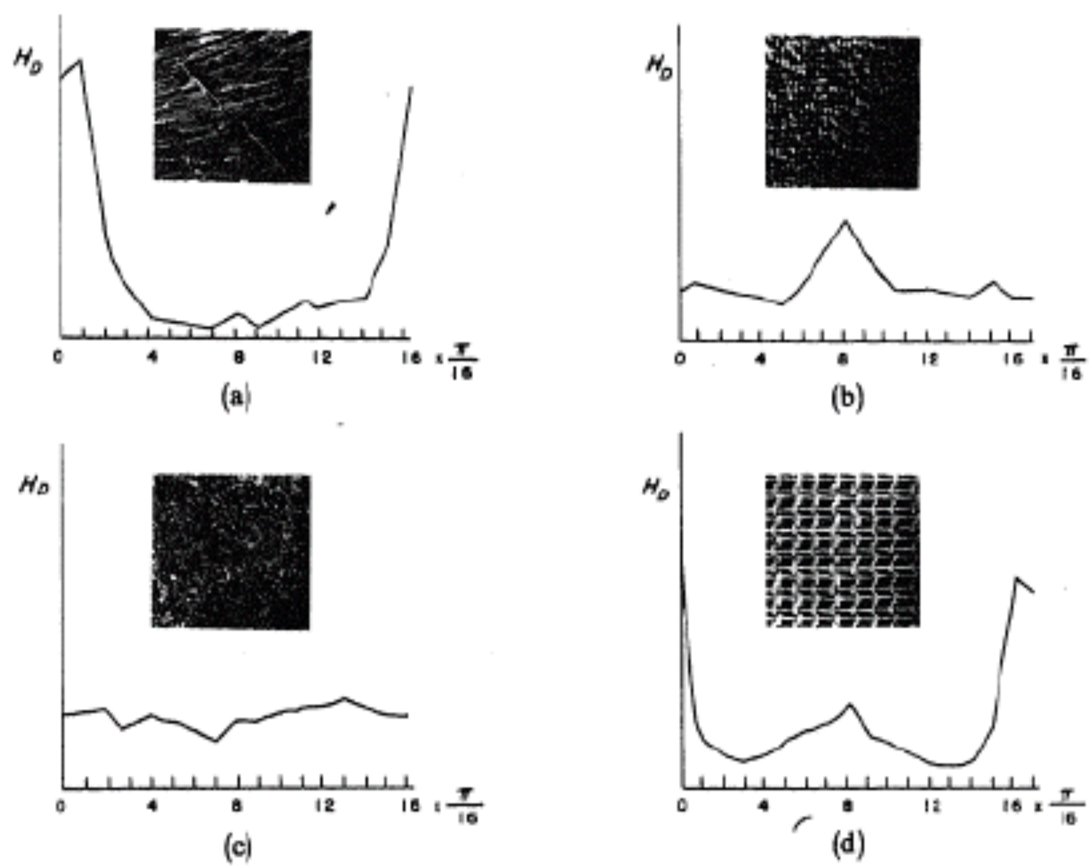


- Building the histogram of local edges at different orientations $H_D(k)$
  - By deriving the edge magnitude at X and Y directions

$$\theta = \mathrm{tg}^{-1}(\nabla_V / \nabla_H) + \frac{\pi}{2}$$

$$|\nabla G| = \left(|\nabla_V| + |\nabla_H|\right)/2$$

$$\begin{array}{cc} \nabla_V & \nabla_H \\ \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix} \end{array}$$

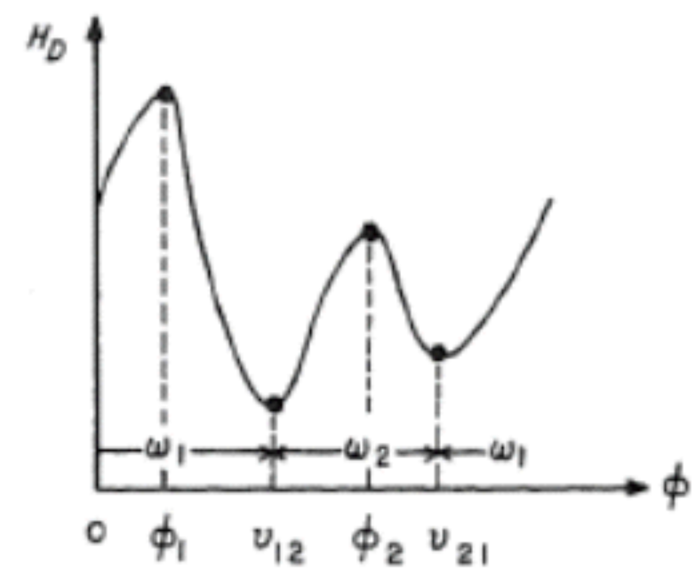


Tamura – Orientation (cont.)

- Compute the estimate from the sharpness of the peaks
  - By summing the second moments around each peak
    e.g., flat histogram
    → large 2nd moment (variance)
    → small orientation

$$M_{orient} = 1 - r \cdot n_p \cdot \sum_{p}^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 \cdot H_D(\phi)$$

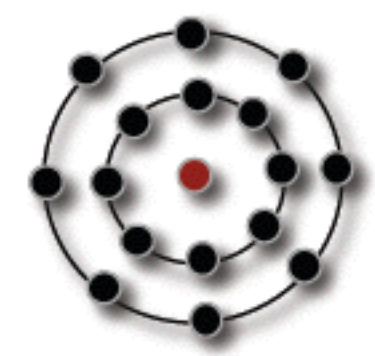
$n_p =$	*Number of peaks*
$\phi_p =$	*Position of peak, p, in* H_D
$w_p =$	*Points in peak p*
$r =$	*Normalisation factor*



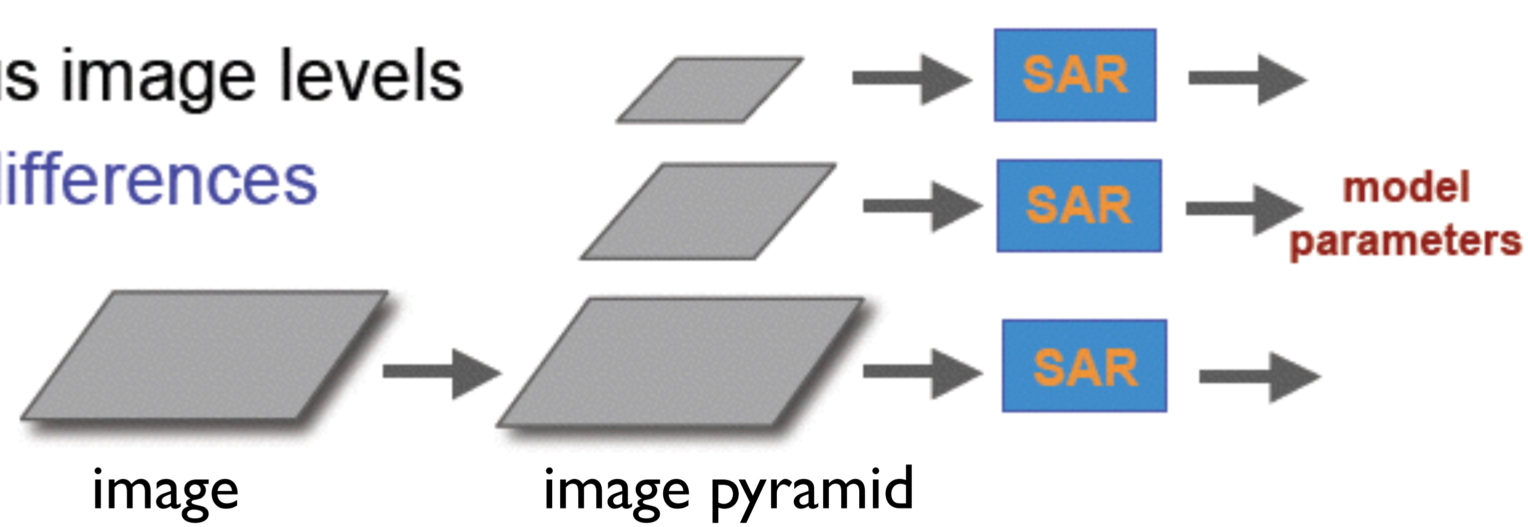
(MR)SAR

[Mao'92]

- Each pixel is a random variable whose value is estimated from its neighboring pixels + noise
  - A kid of Markov Random Field model



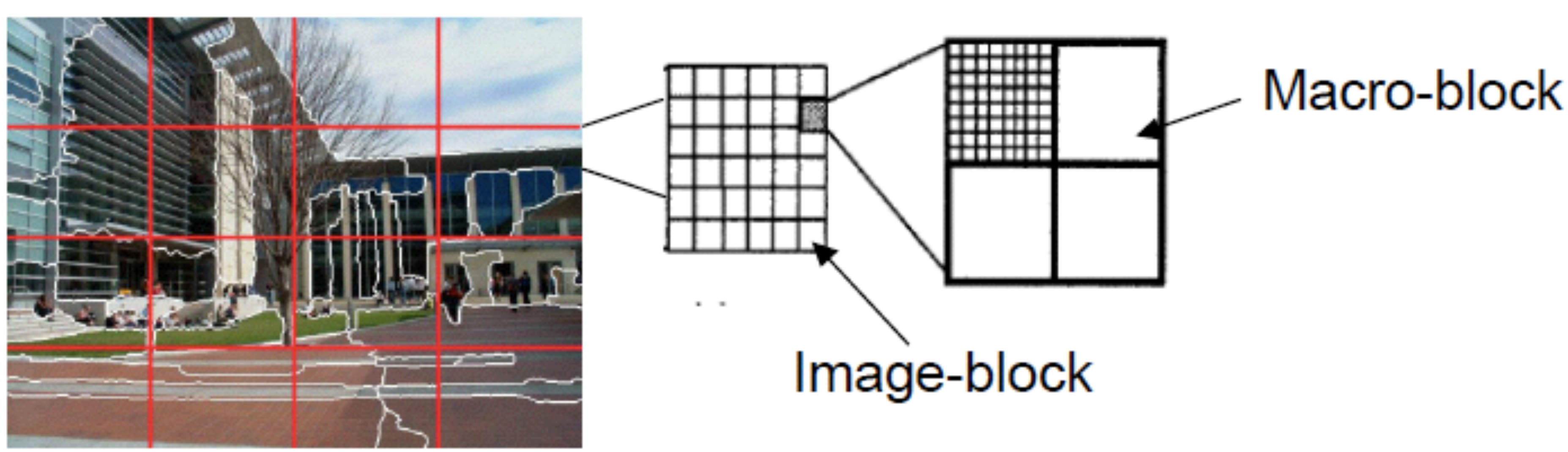
- **SAR** Model (Simultaneous Autoregressive)
  - Describes each pixel in terms of its neighboring pixels.
- **MRSAR** Model (MultiResolution SAR)
  - Describing granularities by representing textures at variety of resolutions
  - SAR applied at various image levels
  - Metric → parameter differences



Edge Histogram

- Edge histogram (EHD)
- Captures the spatial distribution of the edge in six statues: 0º, 45º, 90º, 135º, non direction and no edge.
  - Utilizing the filters

  - Global EHD of an image: Concatenating 16 sub EHDs into a 96 bins
  - Local EHD of a segment
    - Grouping the edge histogram of the image-blocks fallen into the segment

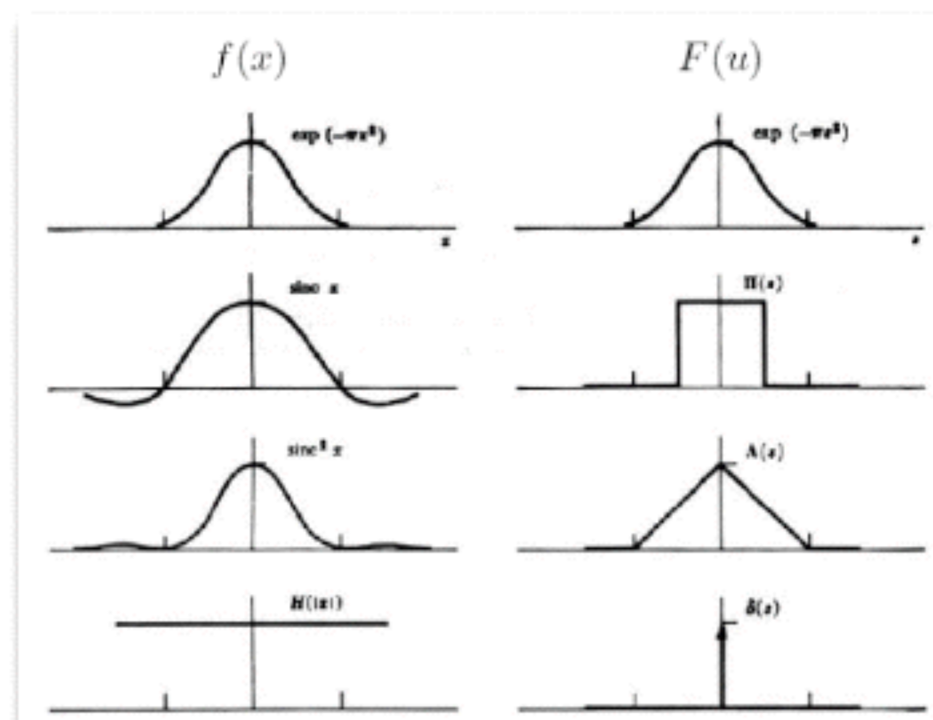


The Fourier Transform

- Represent function on a new basis
  - Think of functions as vectors, with many components
  - We now apply a linear transformation to transform the basis
    - dot product with each basis element

$$F(g(x,y))(u,v) = \iint_{\mathbb{R}^2} \underline{g(x,y)}\,\underline{e^{-i2\pi(ux+vy)}}\,dxdy$$

- In the expression, u and v select the basis element, so a function of x and y becomes a function of u and v
- basis elements have the form $e^{-i2\pi(ux+vy)}$



Discrete Fourier Transform

- 2D DFT

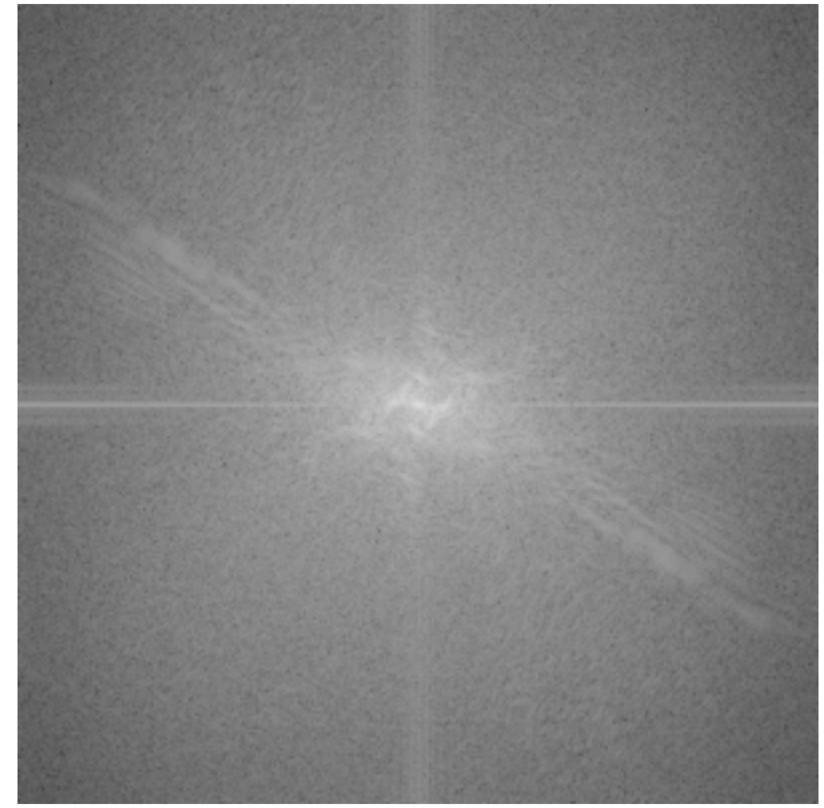
$$F(k,l) = \frac{1}{N^2} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} f(a,b)\, e^{-\iota 2\pi \left(\frac{ka}{N} + \frac{lb}{N}\right)}$$

- 2D IDFT

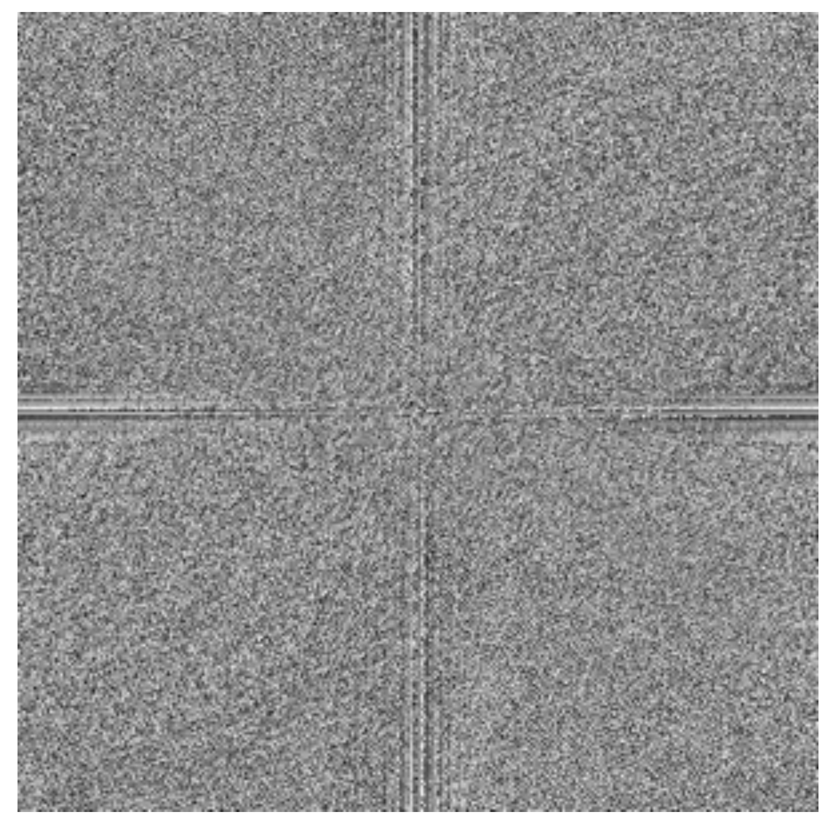
$$f(a,b) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} F(k,l)\, e^{\iota 2\pi \left(\frac{ka}{N} + \frac{lb}{N}\right)}$$

Zebra

Fourier Transform



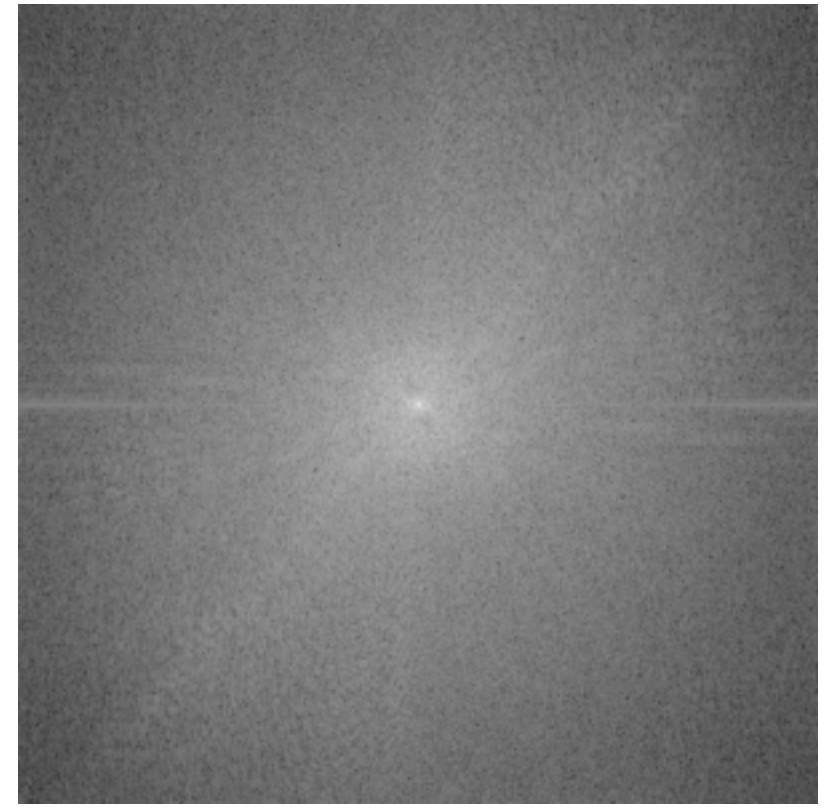
magnitude transform



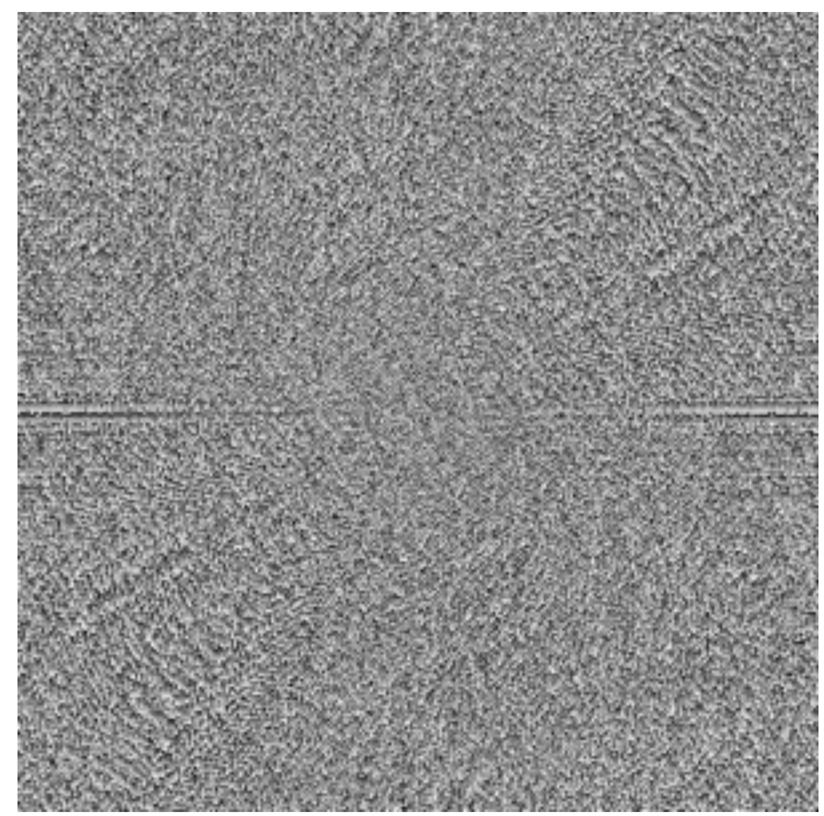
phase transform

Leopard

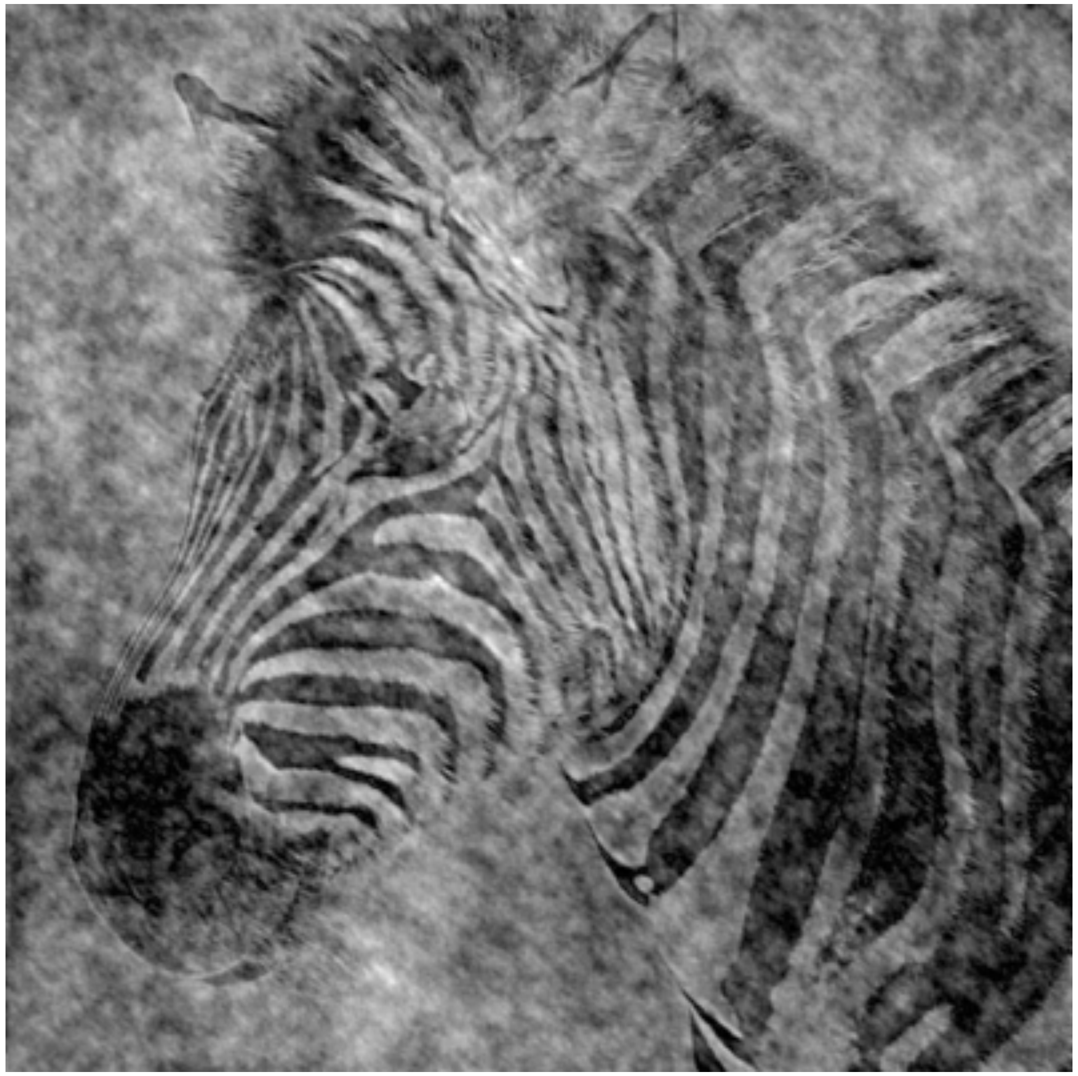
Fourier
Transform



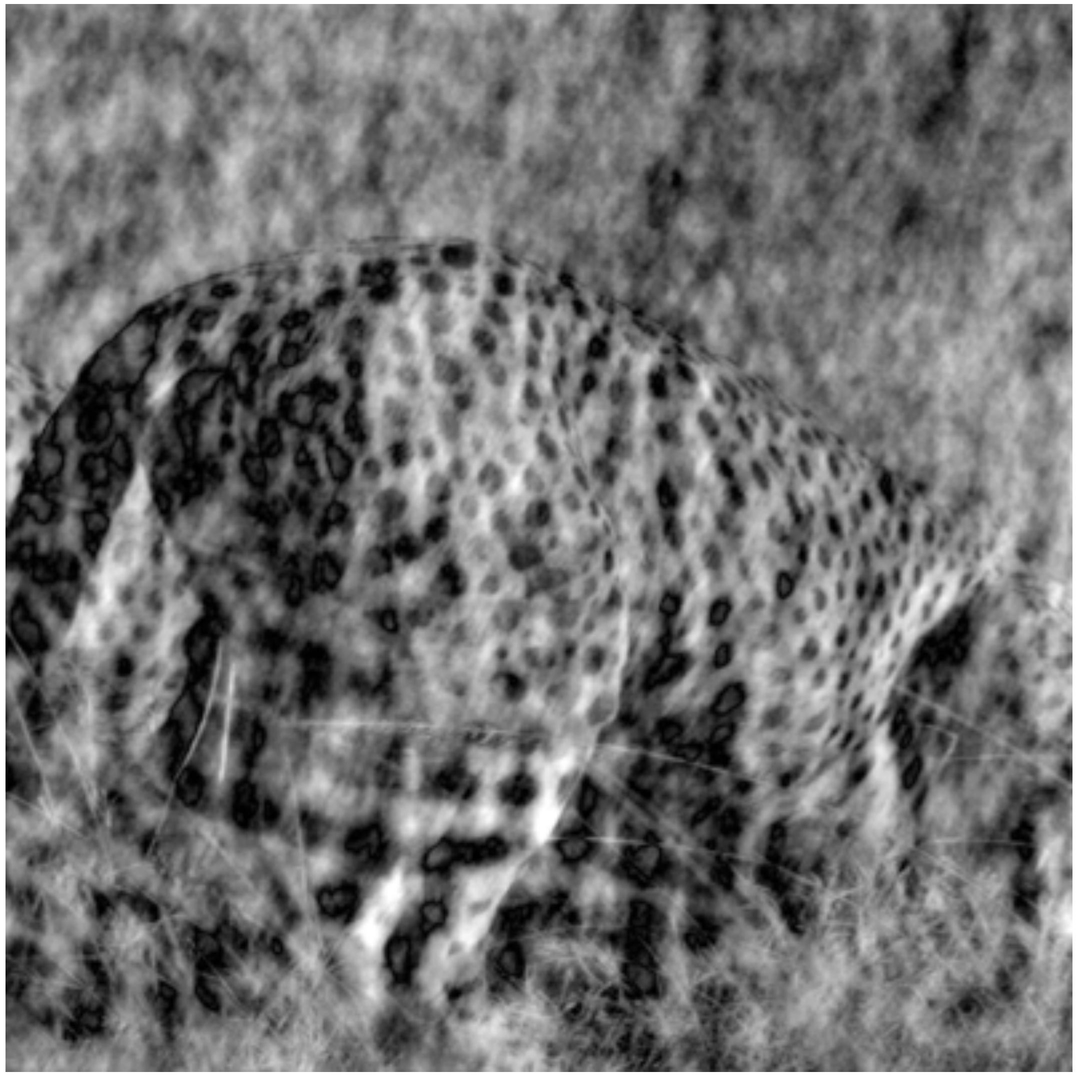
magnitude transform



phase transform

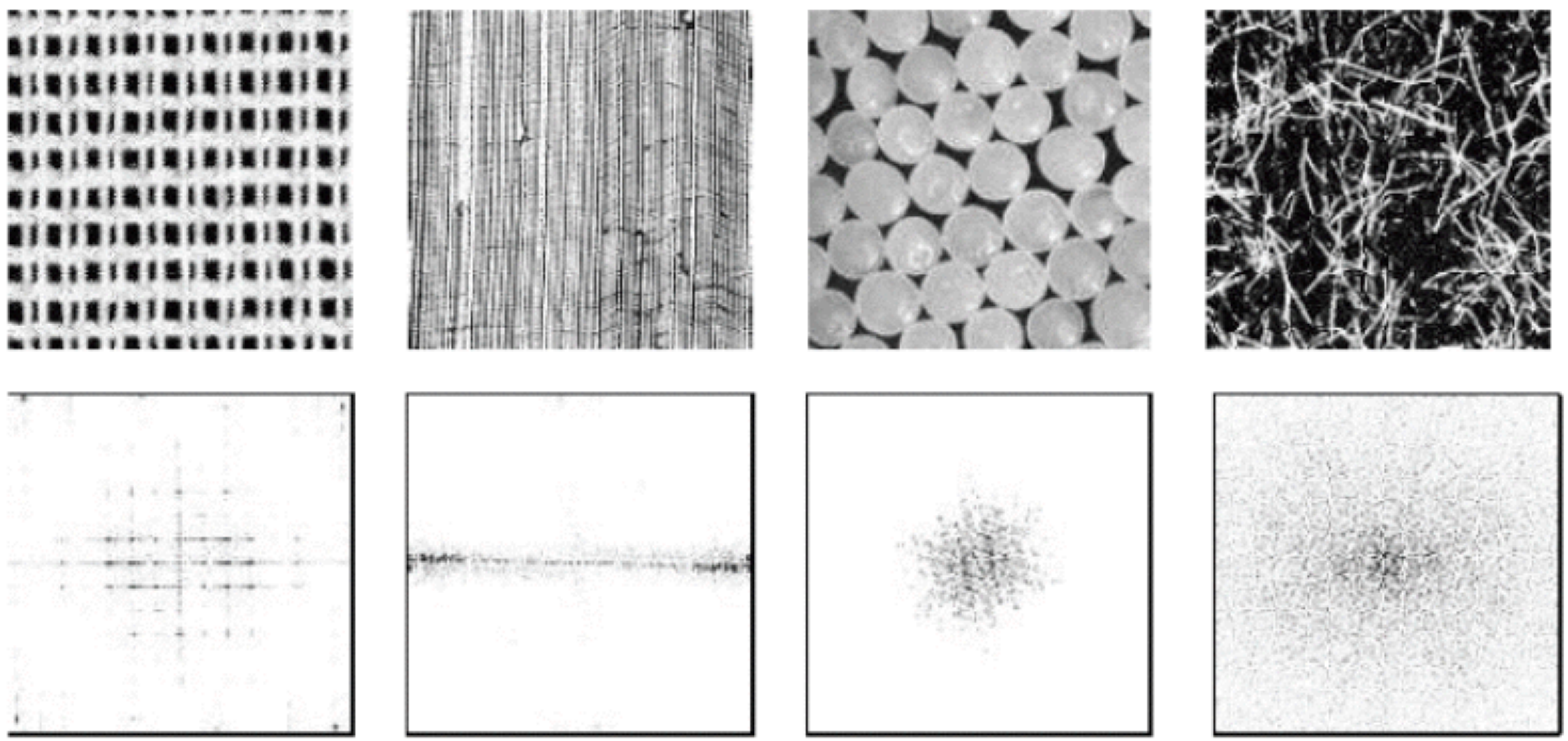


Zebra's phase
+ Leo's mag



Leo's phase
+ Zebra's mag

Natural Images and Their FT



(a) structured (b) oriented (c) granular (d) random

- What happened to the FT patterns when the texture scale and orientation are changed?

Frequency Domain Features

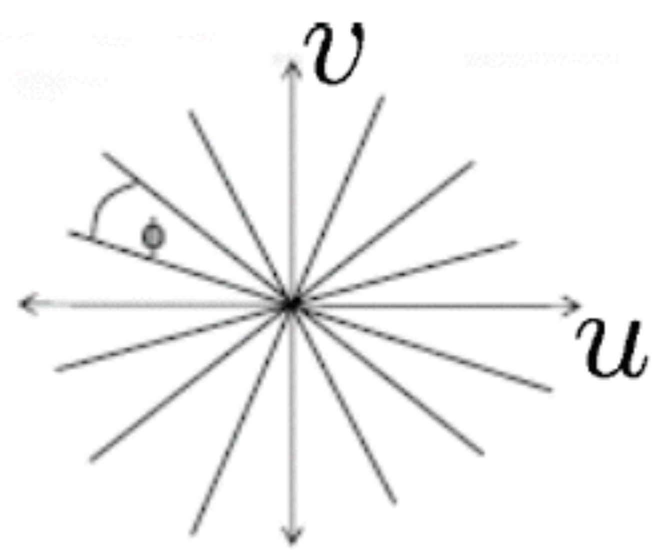
Fourier domain energy distribution

- Angular features (directionality)

$$V^{(a)}_{\theta_1\theta_2} = \int\int |F(u,v)|^2 du dv$$

where,

$$\theta_1 \leq \tan^{-1}[\frac{v}{u}] \leq \theta_2$$

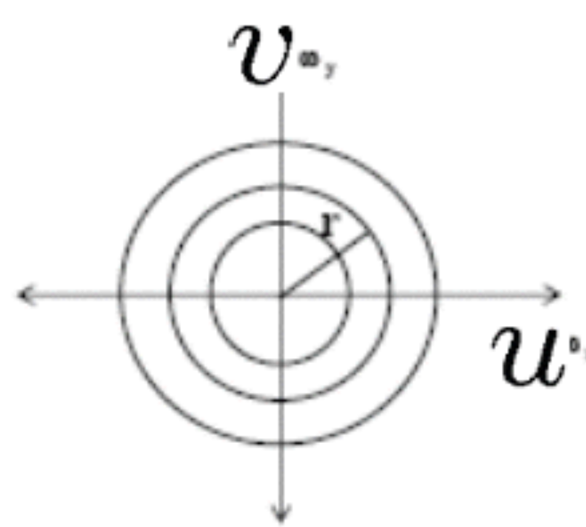


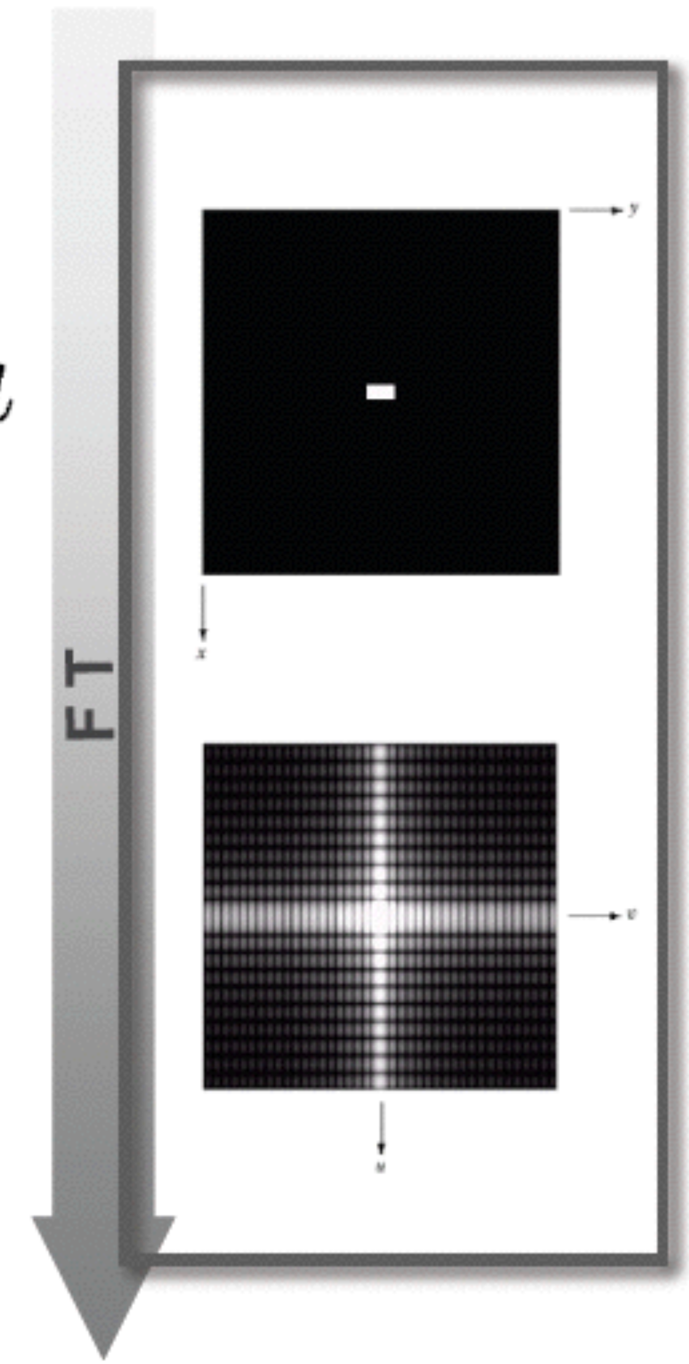
- Radial features (coarseness)

$$V^{(r)}_{r_1 r_2} = \int\int |F(u,v)|^2 du dv$$

where,

$$r_1 \leq u^2 + v^2 < r_2$$

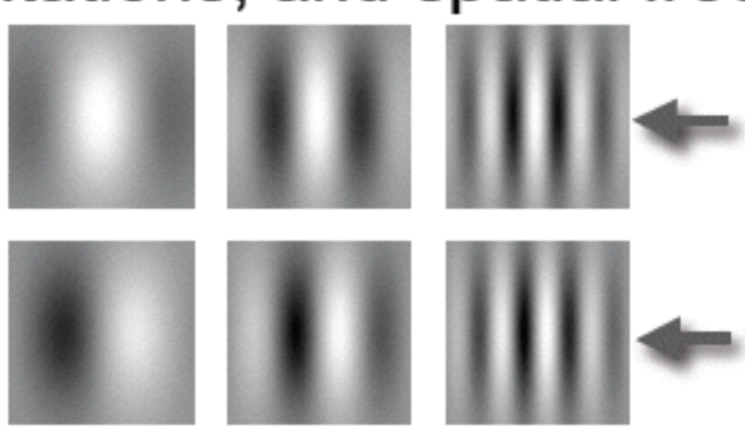




Uniform division may not be the best!!

Gabor Texture

- Fourier coefficients depend on the entire image (Global) → we lose spatial information
- Objective: local spatial frequency analysis
- Gabor kernels: looks like Fourier basis multiplied by a Gaussian
  - The product of a symmetric (even) Gaussian with an oriented sinusoid
  - Gabor filters come in pairs: symmetric and anti-symmetric (odd)
  - Each pair recover symmetric and anti-symmetric components in a particular direction
  - (k_x, k_y): the spatial frequency to which the filter responds strongly
  - σ : the scale of the filter. When σ = infinity, similar to FT
- We need to apply a number of Gabor filters are different scales, orientations, and spatial frequencies

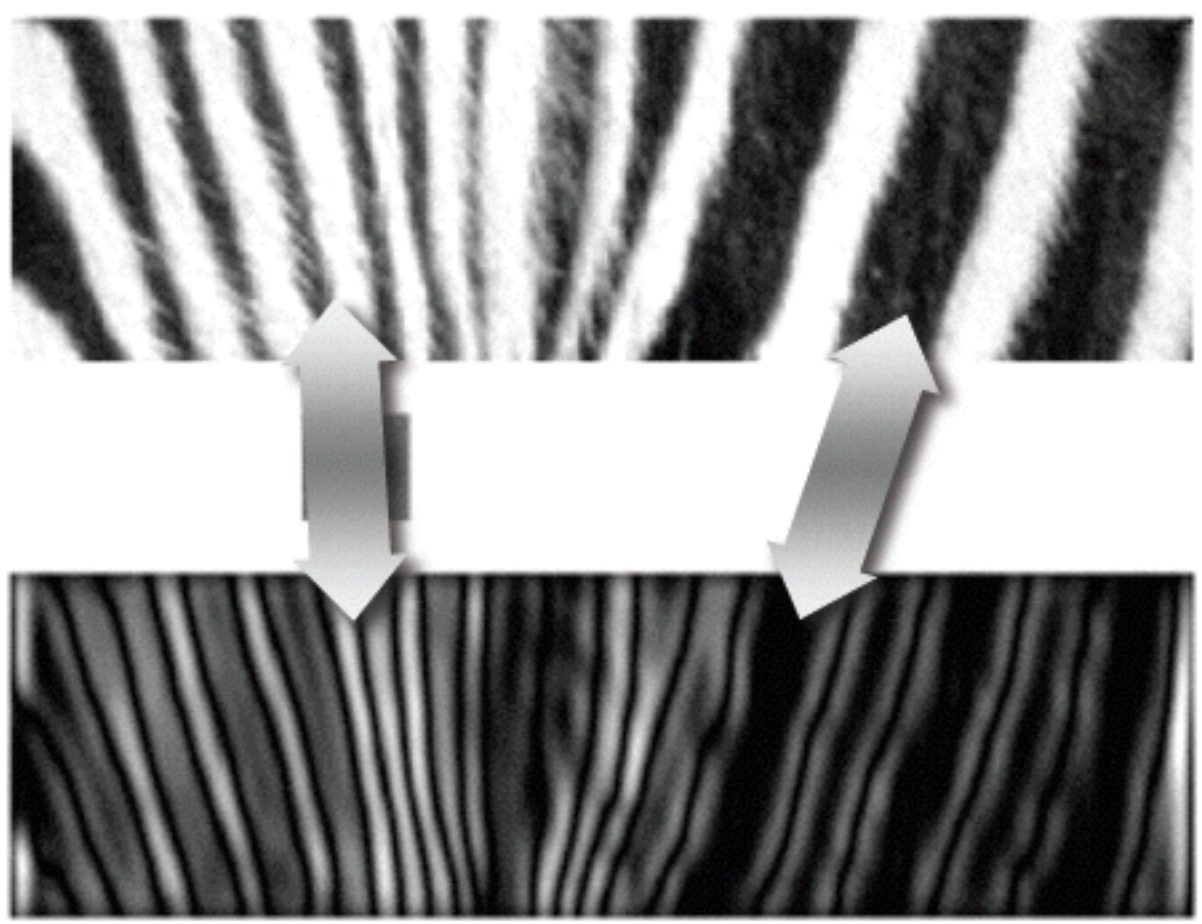


$$G_{symmetric}(x,y) = \cos(k_x x + k_y y) \exp -\frac{x^2+y^2}{2\sigma^2}$$

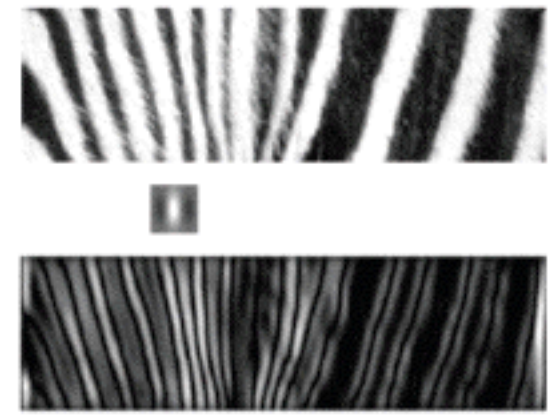
$$G_{anti-symmetric}(x,y) = \sin(k_x x + k_y y) \exp -\frac{x^2+y^2}{2\sigma^2}$$

Example – Gabor Kernel

- Zebra stripes at different scales and orientations and convolved with the Gabor kernel
- The response falls off when the stripes are larger or smaller
- The response is large when the spatial frequency of the bars roughly matches the windowed by the Gaussian in the Gabor kernel
- Local spatial frequency analysis



Gabor Texture (cont.)



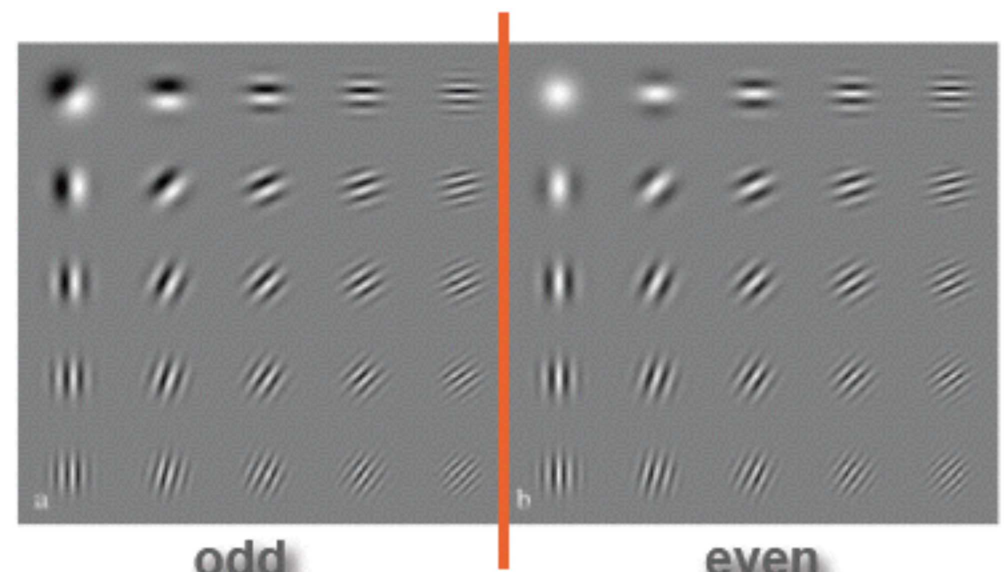
- Image $I(x,y)$ convoluted with Gabor filters h_{mn} (totally M x N)

$$W_{mn}(x,y) = \int I(x_1,y_1)h_{mn}(x-x_1,y-y_1)dx_1dy_1$$

- Using first and 2nd moments for each scale and orientations

$$\mu_{mn} = \int\int |W_{mn}(x,y)|dxdy$$

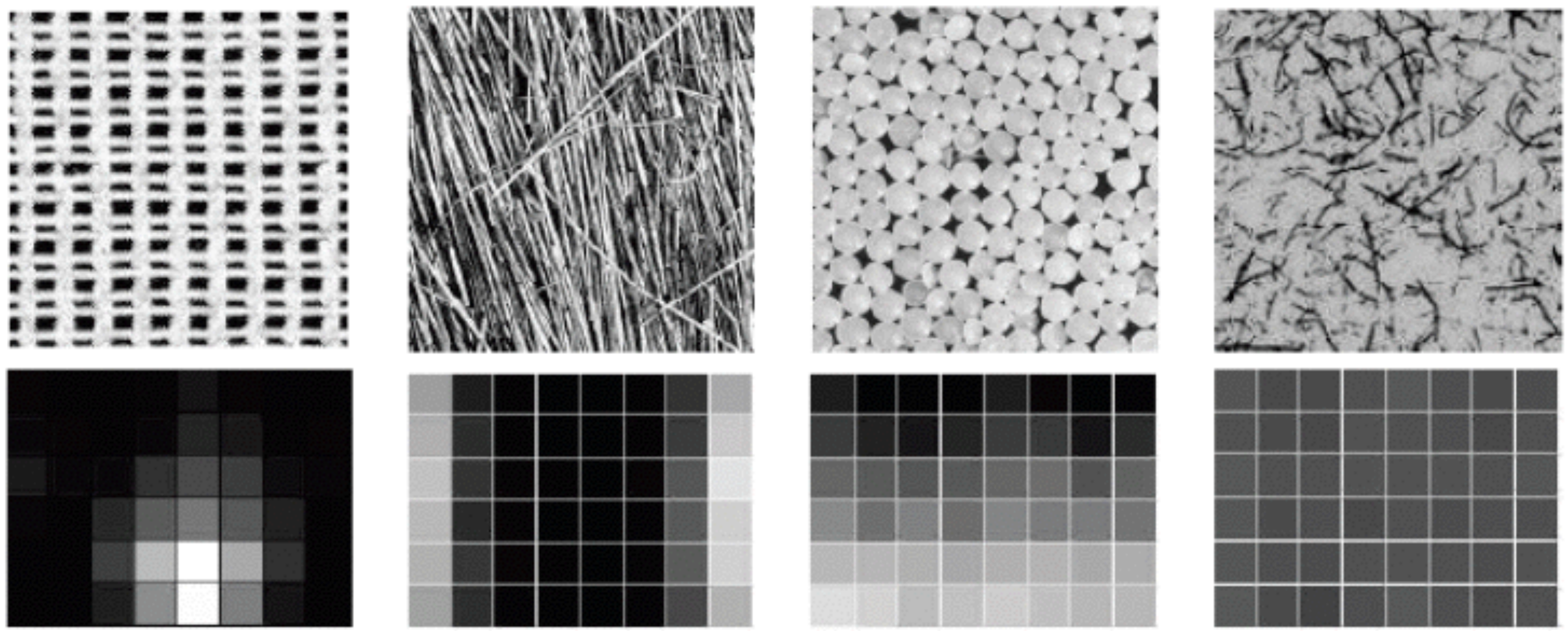
$$\sigma_{mn} = \sqrt{\int\int (|W_{mn}(x,y)| - \mu_{mn})^2 dxdy}$$



- Features: e.g., 4 scales, 6 orientations → 48 dimensions

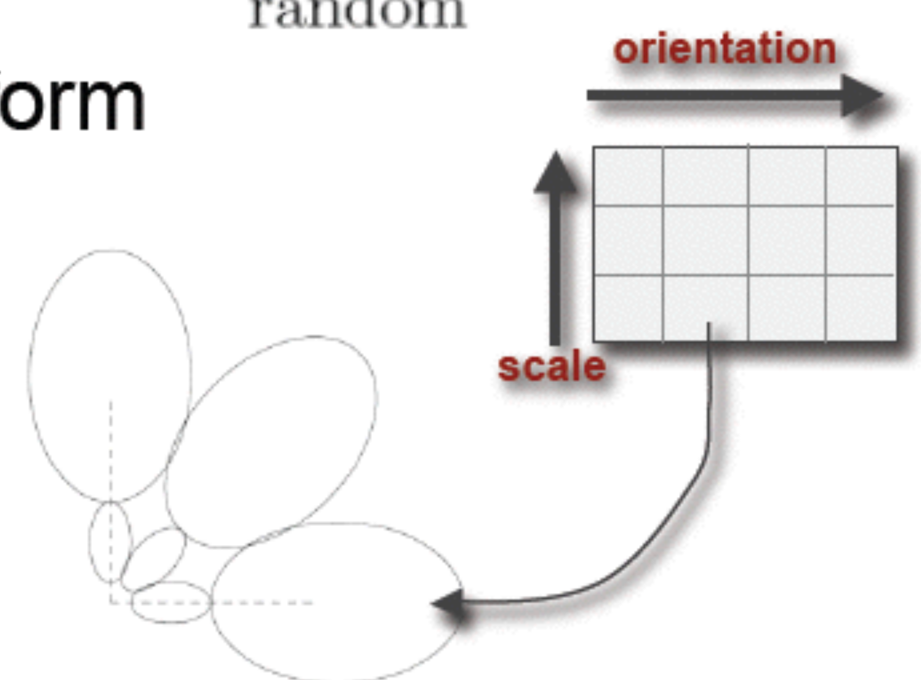
$$\bar{v} = [\mu_{00}, \sigma_{00}, \mu_{01}, \ldots, \mu_{35}, \sigma_{35}]$$

Gabor Texture (cont.)



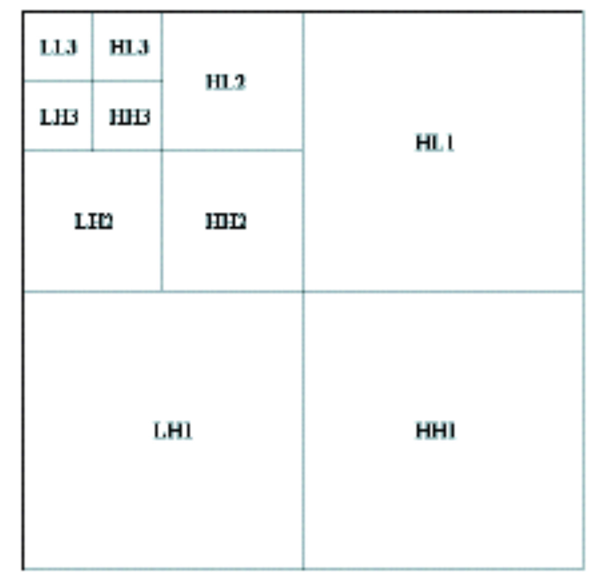
structured oriented granular random

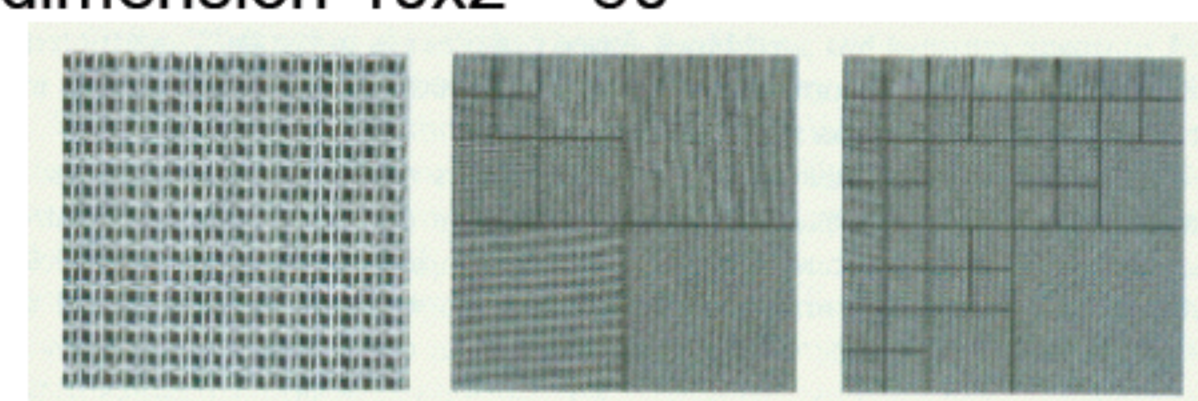
- Arranging the mean energy in a 2D form
  - structured: localized pattern
  - oriented (or directional): column pattern
  - granular: row pattern
  - random: random pattern



Wavelet Features (PWT, TWT)

- Wavelet
  - Decomposition of signal with a family of basis functions with recursive filtering and sub-sampling
  - Each level, decomposes 2D signal into 4 subbands, LL, LH, HL, HH (L=low, H=high)
- PWT: pyramid-structured wavelet transform
  - Recursively decomposes the LL band
  - Feature dimension (3x3x1+1)x2 = 20
- TWT: pyramid-structured wavelet transform
  - Some information in the middle frequency channels
  - Feature dimension 40x2 = 80



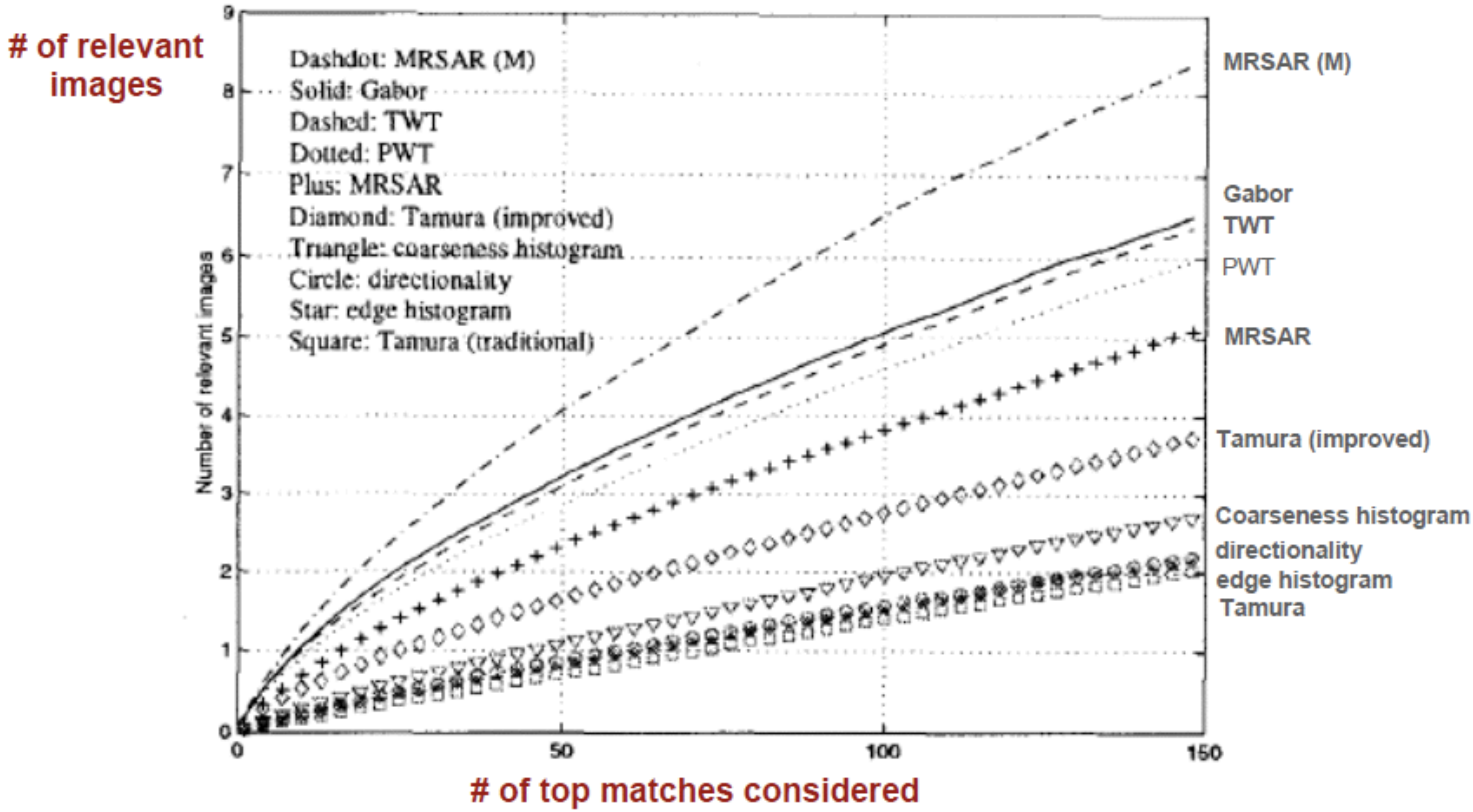


original image PWT TWT

Texture Comparisons

[Ma'98]

- Retrieval performance of different texture features according to the number of relevant images retrieved at various scopes using Corel Photo galleries



Texture directionality

- Gradient:

-1	0	1
-1	0	1
-1	0	1

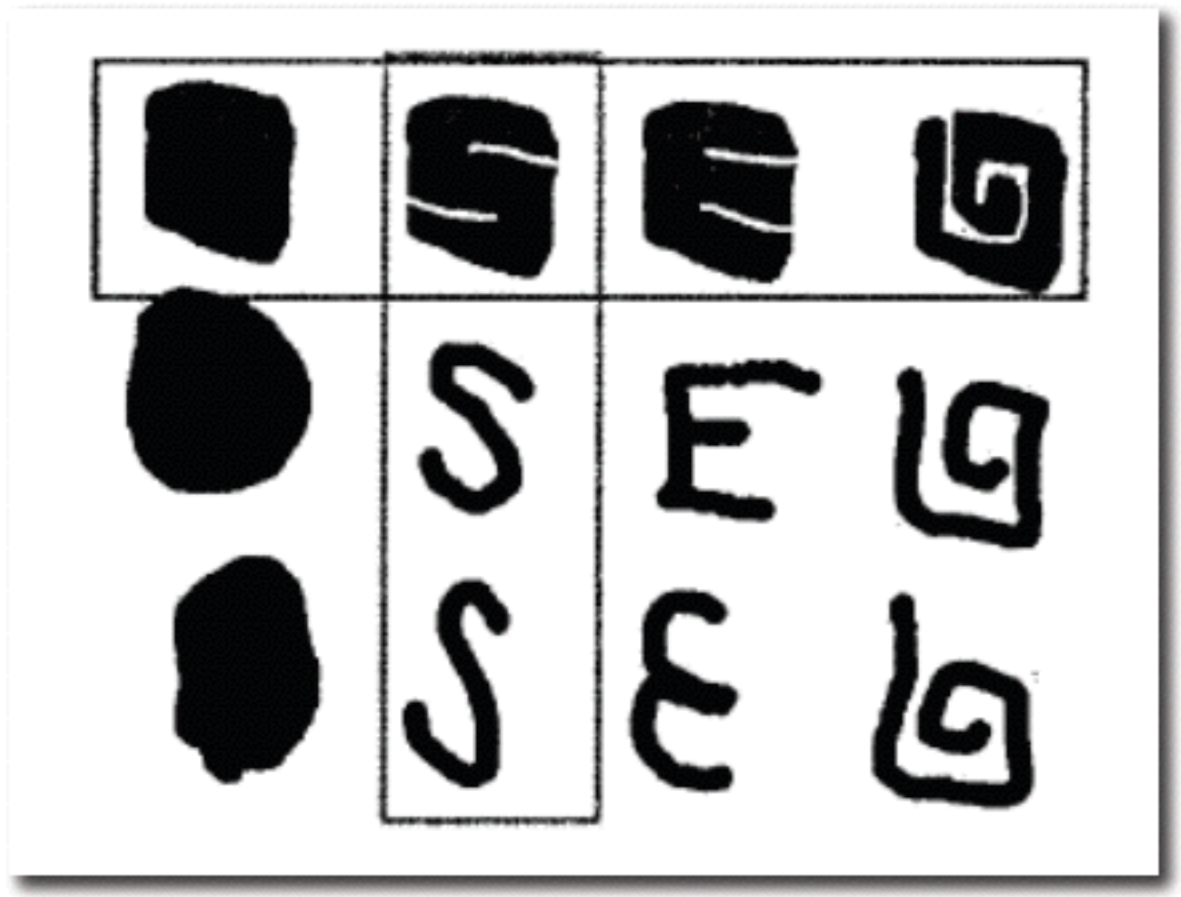
1	1	1
0	0	0
-1	-1	-1

Image shape features

- Shape features are computed out based on object segments or regions, mainly including
  - contour features
  - and regions features.

- Typical approaches include
  - Fourier shape description
  - Moment invariants

Region-based vs. Contour-based Descriptor



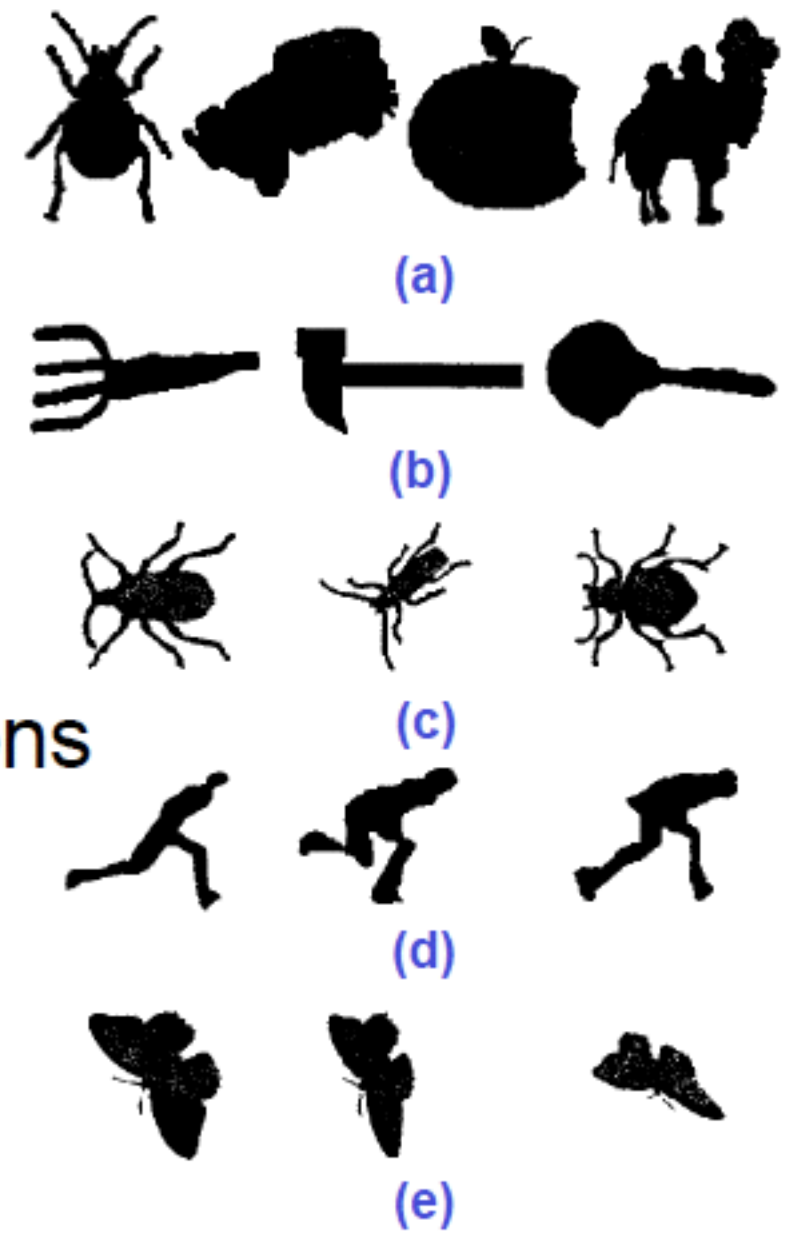
- Columns indicate contour similarity
  - Outline of contours
- Rows indicate region similarity
  - Distribution of pixels

Region-based Descriptor

- Express pixel distribution within a 2D object region
- Employs a complex 2D Angular Radial Transformation (ART)
  - 35 fields each of 4 bits
- Rotational and scale invariance
- Robust to some non-rigid transformation
- L_1 metric on transformed coefficients
- Advantages
  - Describing complex shapes with disconnected regions
  - Robust to segmentation noise
  - Small size
  - Fast extraction and matching

Contour-based Descriptor

- It's based on Curvature (曲率) Scale-Space (CSS) representation
- Found to be superior to
  - Zernike moments
  - ART
  - Fourier-based
  - Turning angles
  - Wavelets
- Rotational and scale invariance
- Robust to some non-rigid transformations
- For example
  - Applicable to (a)
  - Discriminating differences in (b)
  - Finding similarities in (c)-(e)



Problems in Shape-based Indexing

Many existing approaches assume

- Segmentation is given
- Human operator circle object of interest
- Lack of clutter and shadows
- Objects are rigid
- Planar (2-D) shape models
- Models are known in advance

Dimensional reduction for image features

In image retrieval system, increasing feature dimension can enhance precision of retrieval greatly. However, high feature dimension will lead to high computation cost. Hence it is important to reduce the redundant in feature data.

- Image feature space reduction
  - Linear dimensional reduction techniques: PCA …
  - Nonlinear dimensional reduction techniques: Isomap, LLE …
  - Clustering based feature reduction methods
- High-dimensional feature indexing
  - Database oriented high-dimensional data indexing
    - Bucketing grouping searching techniques，K-d tree, R tree …
  - Clustering methods
  - SOM

Image similarities

- How to measure similarity of different images base on features?
  - Image features always form into a fixed-length feature vector.
  - The similarity therefore can be measure by
    - Euclidian distance
    - Histogram intersection
    - Quadratic distance
    - Mahalanobis distance (马氏距离)
    - Non-geometrical similarity

Similarity and distance

- Similarity: ↑
- distance: ↓

Practical image retrieval systems

- QBIC (Query By Image Content)
  - http://www.qbic.almaden.ibm.com/
- Virage
  - http://wwwvirage.com/cgi-bin/query-e
- RetrievalWare
  - http://vrw.excalib.com/cgi-bin/sdk/cst/cst2.bat
- Photobook
- MARS
  - http://jadzia.ifp.uiuc.edu:8000

Practical image retrieval systems (cont.)

- Most existing image retrieval systems have one or more of following functions features:
  - –Random browsing
  - –Classified browsing
  - –Example based retrieval
  - –Sketch based retrieval
  - –Texture based retrieval

Future of image retrieval

- Human-computer interaction
- Semantic speech
- Web-oriented
- High dimensional data
- Perspective
- Multiple media channels
- Image feature mapping
- Standards of performance measurements
- Construction of test sets