



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
2020-10-27

确定大程检查时间



浙江大学2020-2021学年秋冬学期校历

周 星期	学 期 月 日	暑假						秋学期										冬学期										寒假	
		八月						九月			十月				十一月			十二月				一月			二月				
		暑假						一	二	三	四	五	六	七	八	九	十	一	二	三	四	五	六	七	八	九	十	寒假	
星期一	3	10	17	24	31	7	14	21	28	5	12	19	26	2	9	16	23	30	7	14	21	28	4	11	18	25	1	8	
星期二	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17	24	1	8	15	22	29	5	12	19	26	2	9	
星期三	5	12	19	26	2	9	16	23	30	7	14	21	28	4	11	18	25	2	9	16	23	30	6	13	20	27	3	10	
星期四	6	13	20	27	3	10	17	24	1	8	15	22	29	5	12	19	26	3	10	17	24	31	7	14	21	28	4	11	
星期五	7	14	21	28	4	11	18	25	2	9	16	23	30	6	13	20	27	4	11	18	25	1	8	15	22	29	5	12	
星期六	8	15	22	29	5	12	19	26	3	10	17	24	31	7	14	21	28	5	12	19	26	2	9	16	23	30	6	13	
星期日	9	16	23	30	6	13	20	27	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17	24	31	7	14	

2020年

9月10日	本科生新生、老生报到注册
9月11日-10月10日	本科生新生始业教育
9月11日	研究生新生、老生报到注册
9月12-30日	研究生新生始业教育
9月14日	秋学期开始上课
9月21-30日	秋季研究生毕业教育及离校
10月1-8日	国庆节、中秋节放假调休（9月27日、10月10日分别与10月7日周三、10月8日周四的课对调，10月6日补10月3日的课）
10月23-25日	秋季校运动会（11月13日补10月23日周五课）
11月9-12日	分别补10月1日（周四）、10月2日（周五）、10月5日（周一）、10月6日（周二）的课
11月16-20日	秋学期考试

11月23日	冬学期开始上课
12月21-30日	冬季研究生毕业教育及离校
12月31日	浙江大学学生节（1月18日补12月31日周四课）

2021年

1月1日	元旦放假（1月19日补1月1日周五课）
1月20-28日	全校停课考试
1月29日	学生寒假开始（2月12日春节）

时段	节次	星期一	星期二	星期三	星期四	星期五	星期六	星期日	上课时间
上午	1								8:00-8:45
	2								8:50-9:35
	3								9:50-10:35
	4								10:40-11:25
	5								11:30-12:15
下午	6								13:15-14:00
	7								14:05-14:50
	8								14:55-15:40
	9								15:55-16:40
	10								16:45-17:30
晚上	11								18:30-19:15
	12								19:20-20:05
	13								20:10-20:55

 法定节假日
 周末
 寒假、暑假
 考试周

德才兼备 全面发展 求是创新 追求卓越

Why Do We Need DAM?



- Average creative person looks for a media file 83 times per week
- Fails to find it 35% of the time
- DAM reduces failure to 5%

Digital assets are not
simple bits.



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**

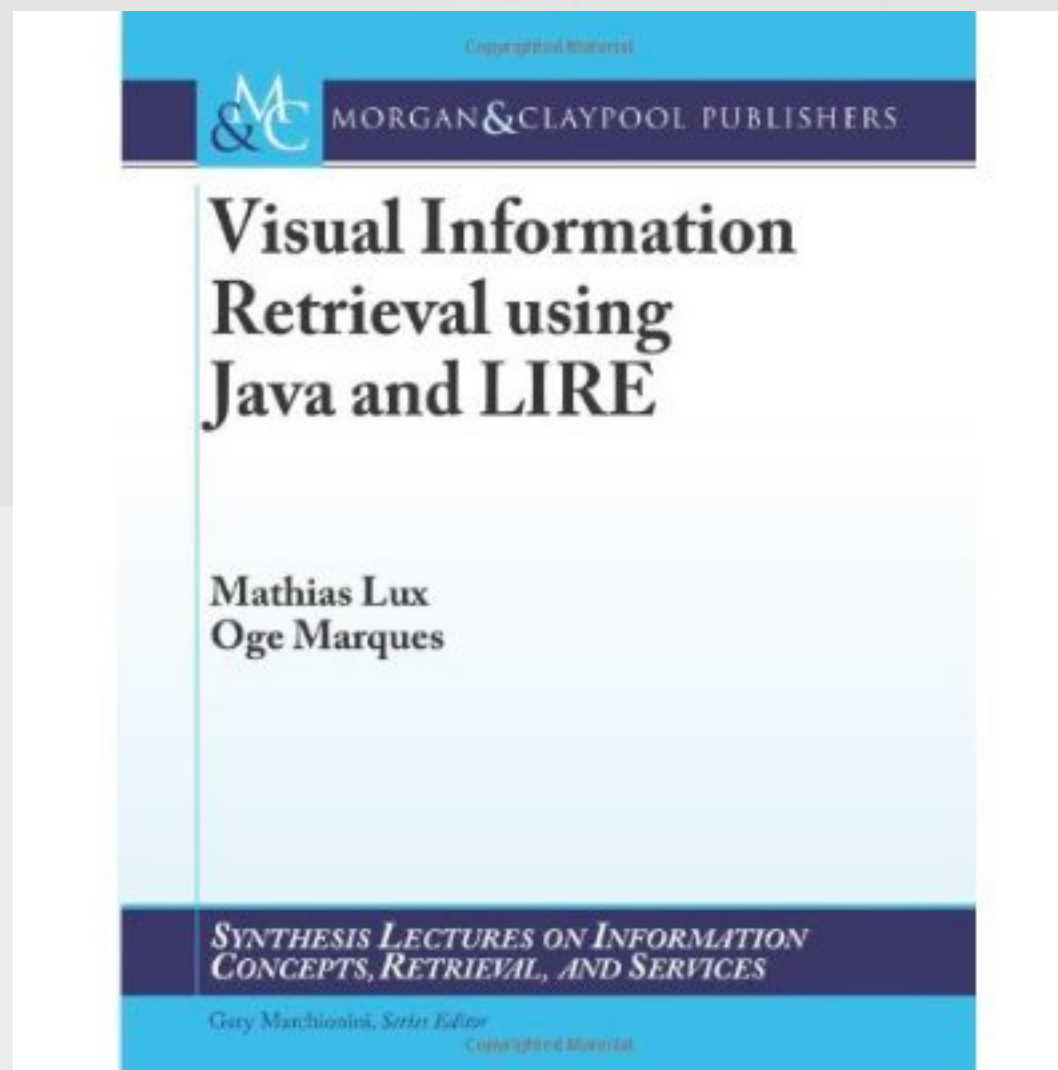
- <http://www.gnu.org/software/gift/>



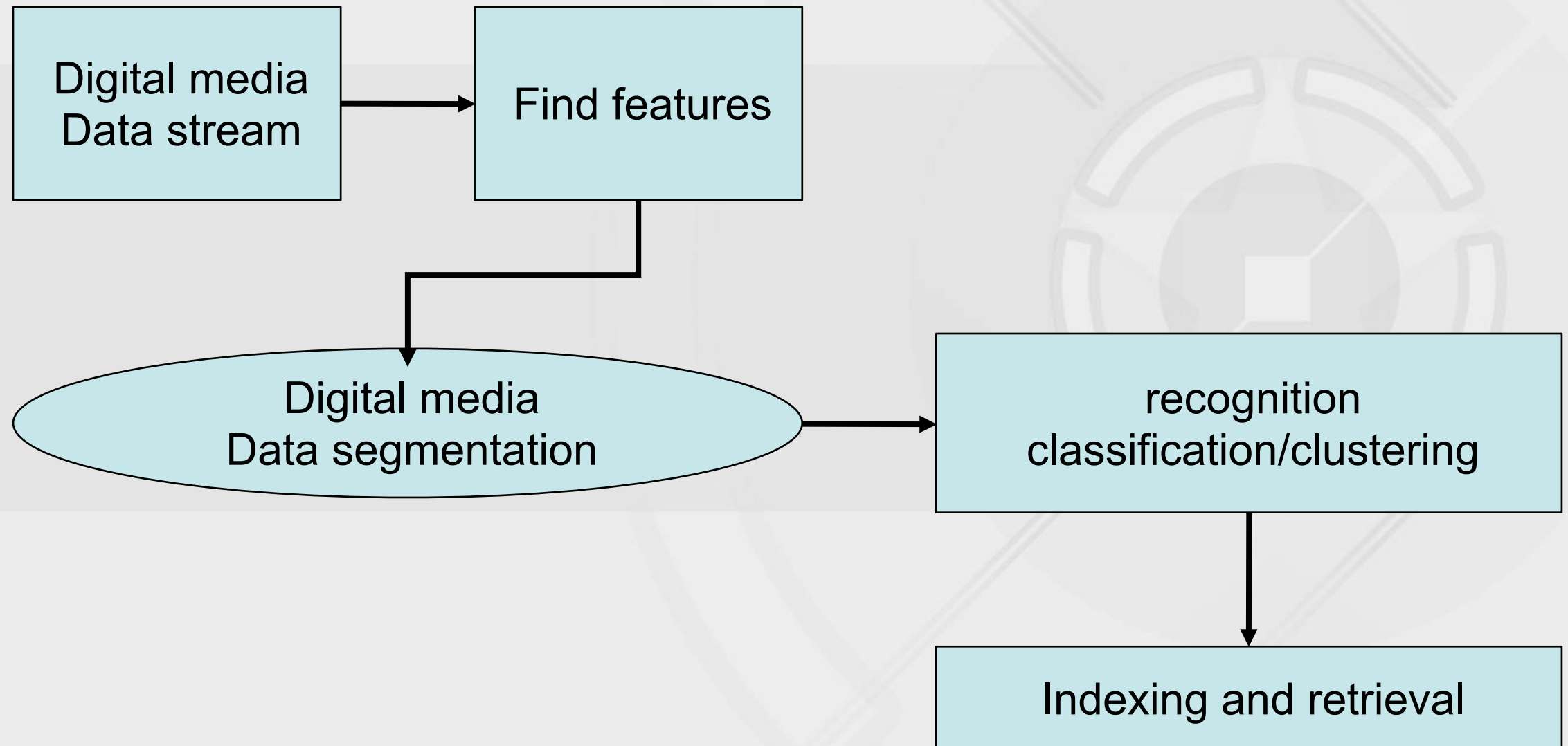
LIRE

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

Query Image



Weights: Perceptual Grouping = 0.2, Color = 0.4, Texture = 0.4, L, A, B channels.

Retrieved Images



CIRES

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

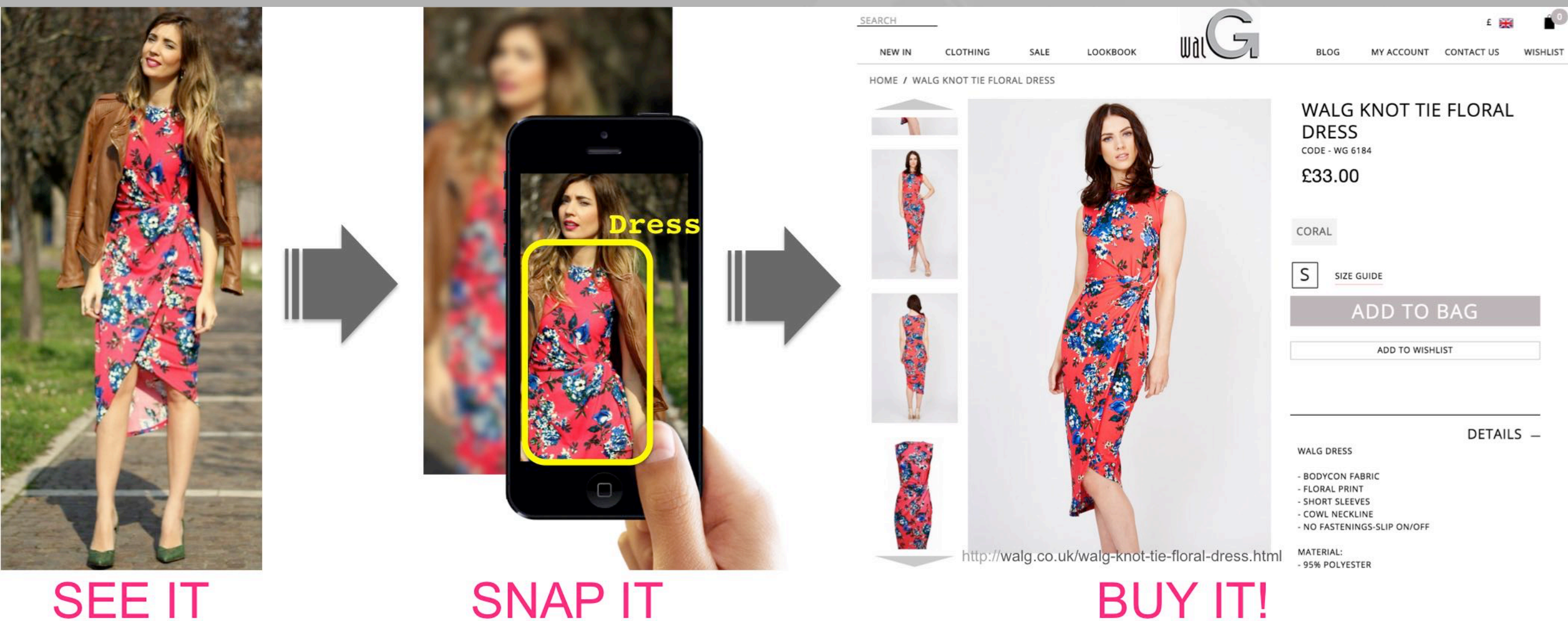
DEMO from the RGB group



iPad APP: “服饰绘”
(InSide system)



(Application) Research



<http://www.tamaraberg.com/street2shop/>



Examples in TAOBAO

中国大陆

亲, 请登录

免费注册

手机逛淘宝

0-0 天猫双11主会场

我的淘宝

购物车0

收藏夹

商品分类

卖家中心

联系客服

网站导航

全球狂欢节 2017 主会场

手机会场 爆款直降千元

电视家影 60寸电视3699

个护家清 1元秒百元券

医药健康 买即赠更划算

住宅家具 买家具送家具

主会场

会场导航

大促优惠

我的嘉年华

宝贝 天猫 店铺

Q 小钢珠糖果

搜索

新款连衣裙 四件套 潮流T恤 时尚女鞋 短裤 半身裙 男士外套 墙纸 行车记录仪 新款男鞋 耳机 更多

主题市场

天猫 聚划算 天猫超市 淘抢购 电器城 司法拍卖 中国质造 兴农扶贫 飞猪旅行 智能生活 苏宁易购

距双11开幕剩 05 天

女装 / 男装 / 内衣

鞋靴 / 箱包 / 配件

童装玩具 / 孕产 / 用品

家电 / 数码 / 手机

美妆 / 洗护 / 保健品

珠宝 / 眼镜 / 手表

运动 / 户外 / 乐器

游戏 / 动漫 / 影视

美食 / 生鲜 / 零食

鲜花 / 宠物 / 农资

房产 / 装修 / 建材


家具 / 家饰 / 家纺

汽车 / 二手车 / 用品

办公 / DIY / 五金电子

百货 / 餐厨 / 家庭保健

如何来找英国女王同款?



来源: https://www.baidu.com/home/news/data/newspage?nid=15305344363915675011&n_type=0&p_from=1

会员俱乐部

注册 开店

息举报专区

论坛 安全 公益

云带来无限发展潜力

回首8年双11引发回忆杀

车险 游戏

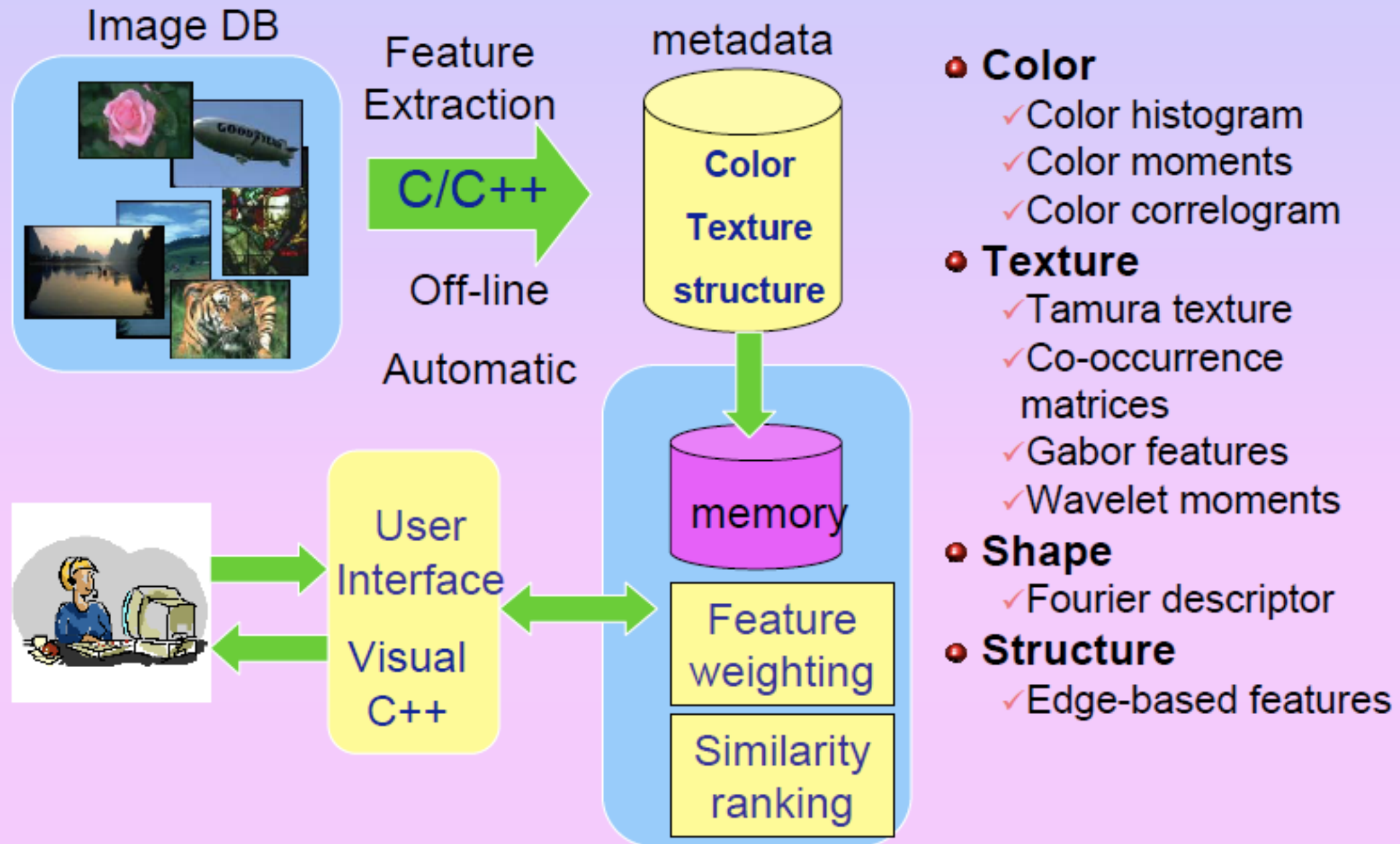
酒店 理财

购物车

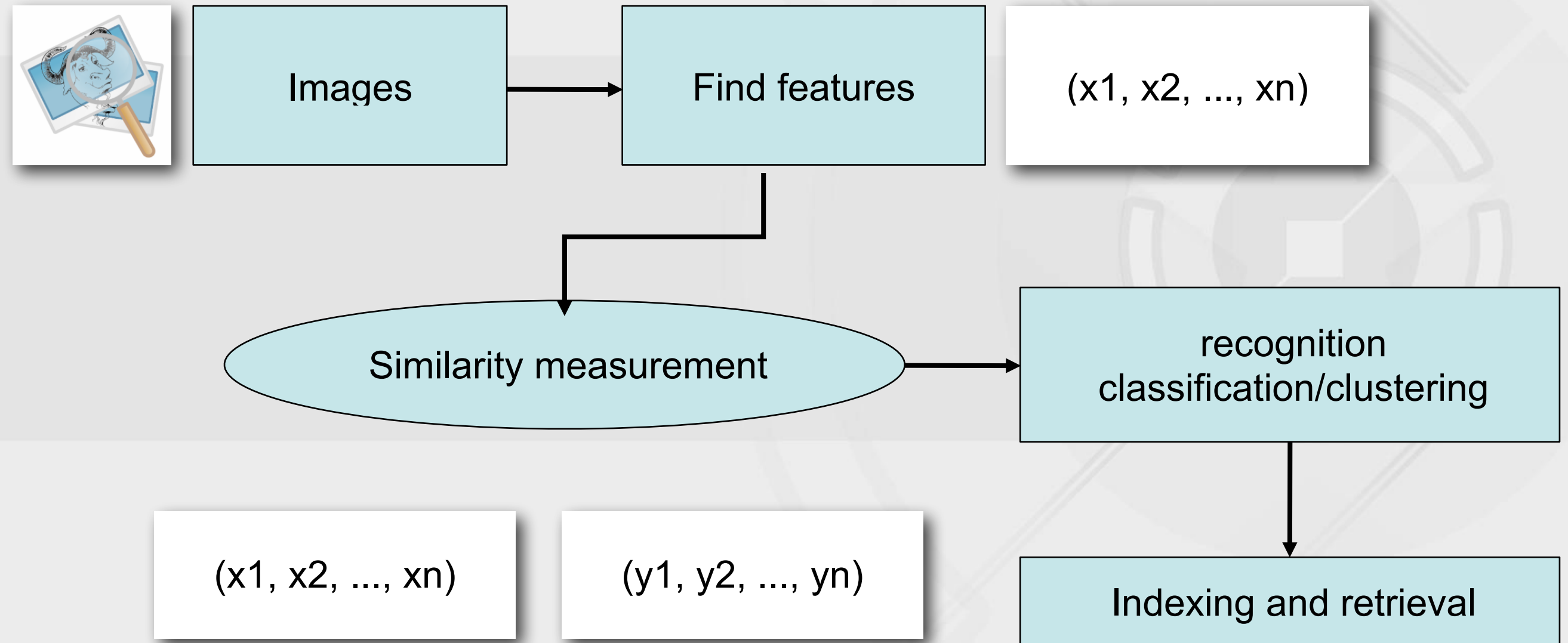
购物车

Multimedia Information Retrieval

• Content-based Image Retrieval



Workflow of CBIR



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

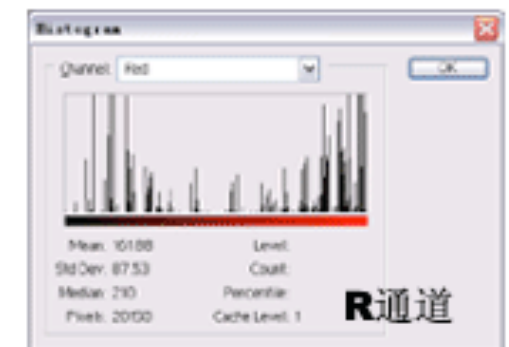
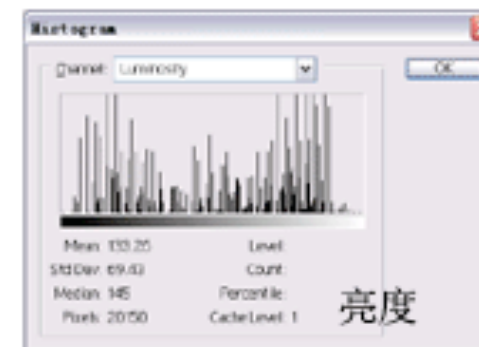


Image histogram

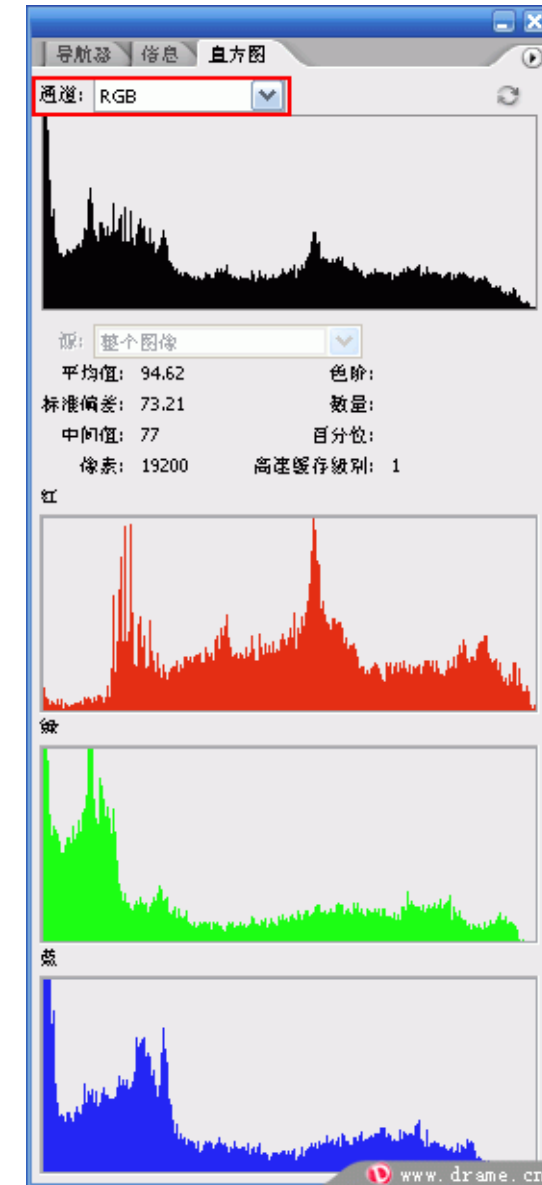
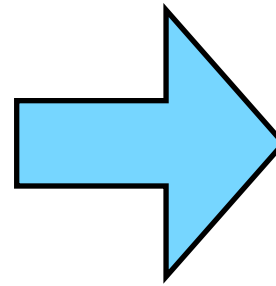
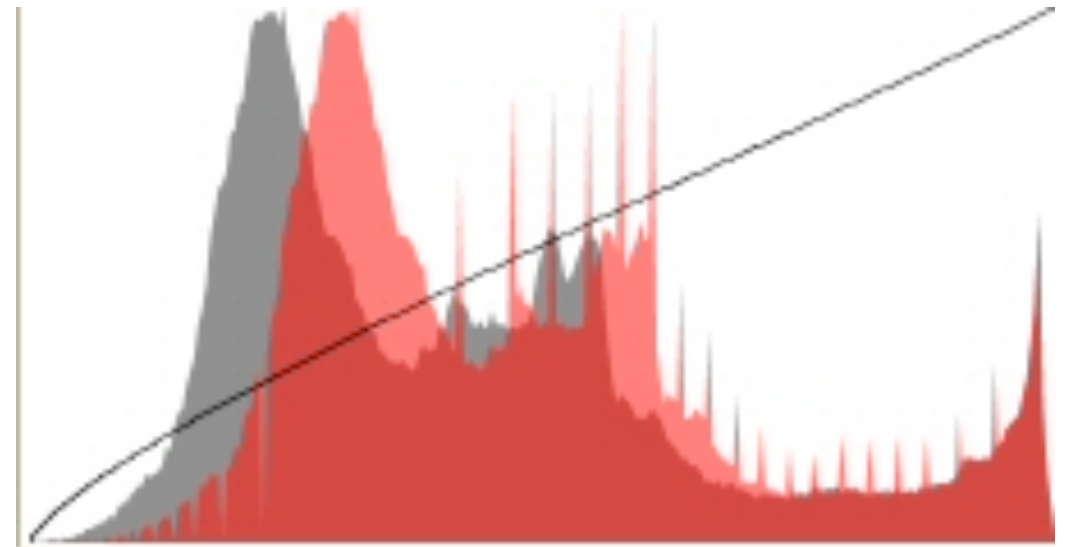
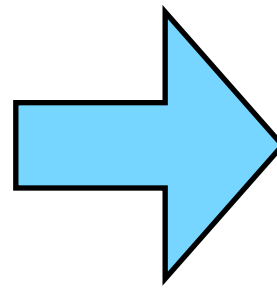


Image histogram



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

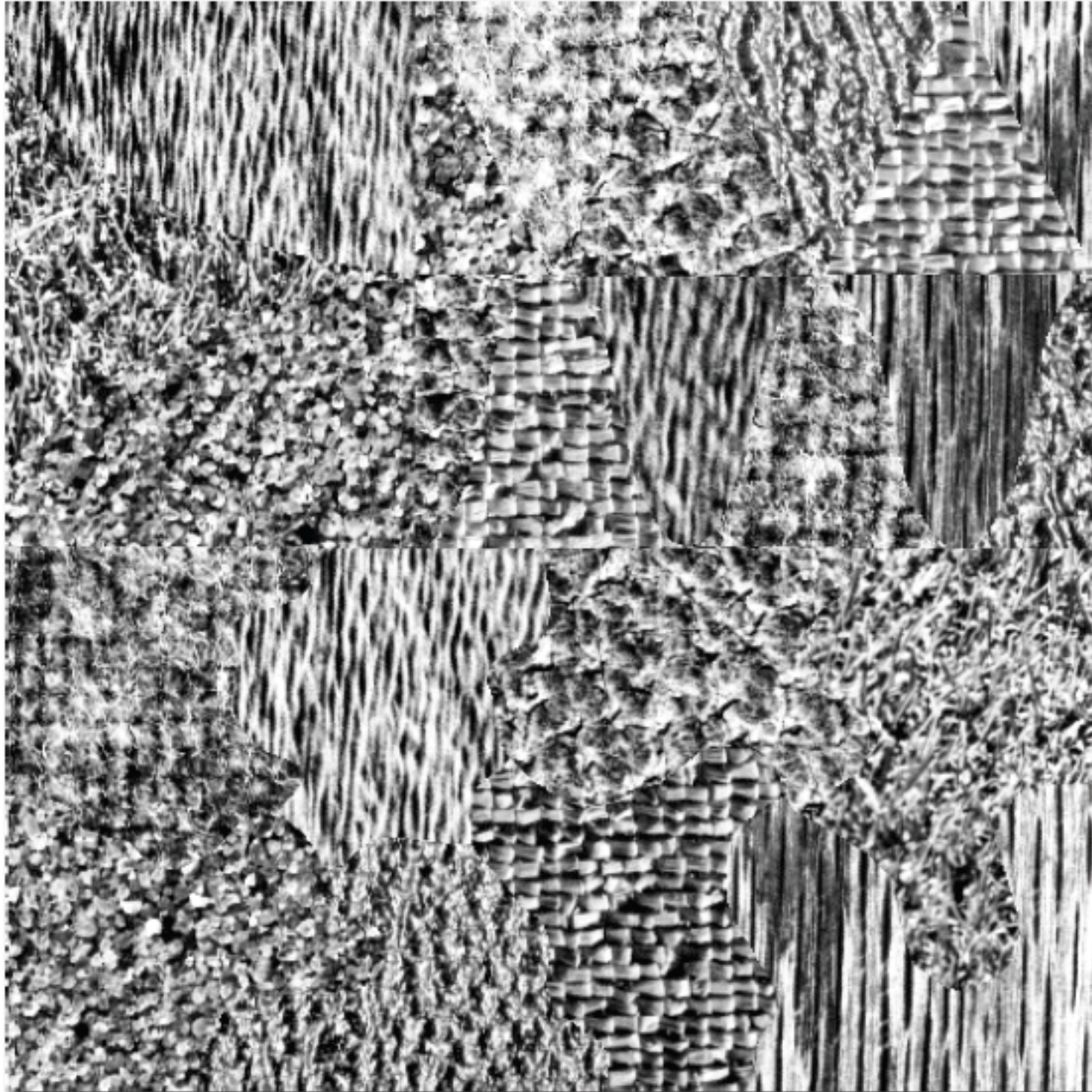


Image texture features

- 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

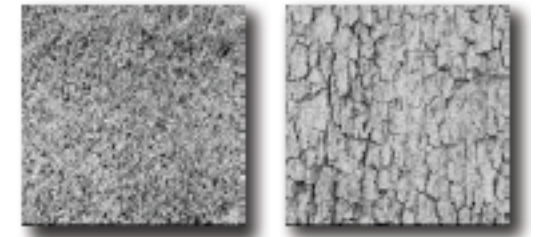


Tamura texture features

- 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

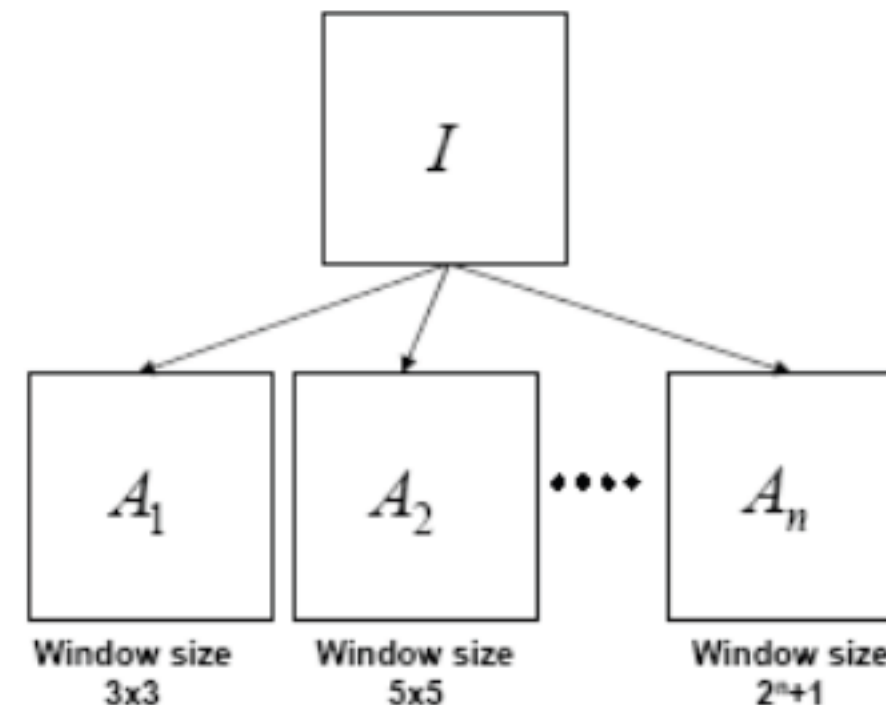
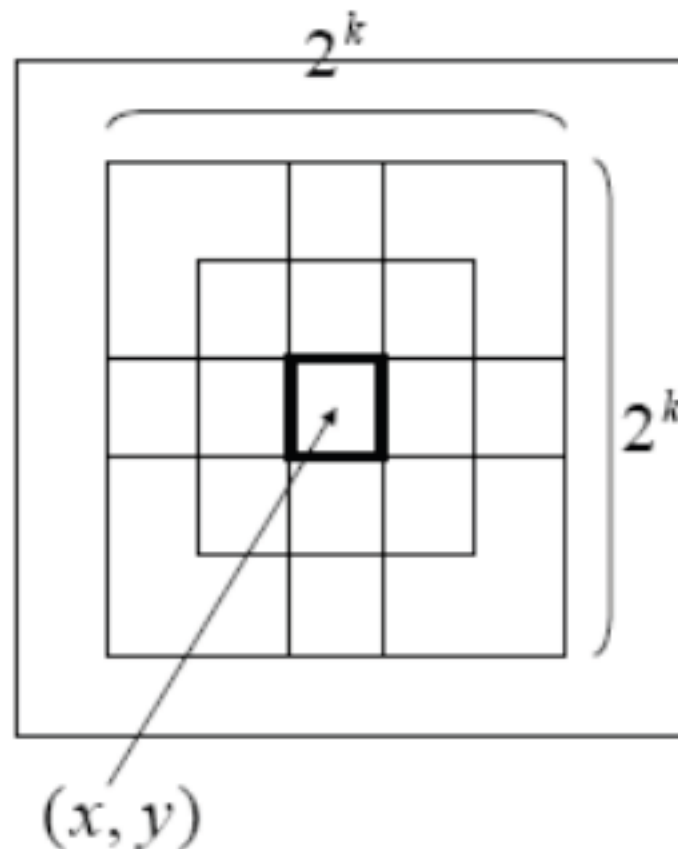


- Goal

- 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

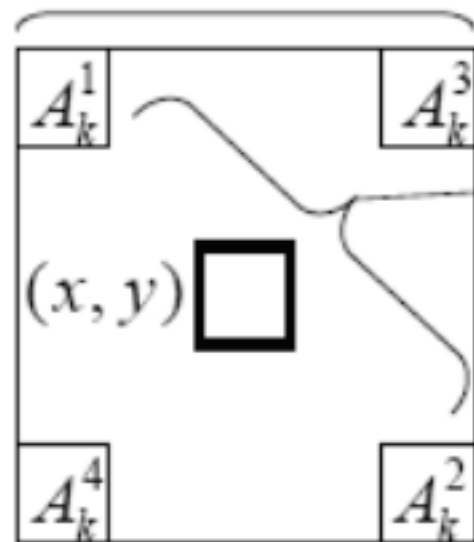
$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} I(i, j) / 2^{2k}$$



Tamura – Coarseness (cont.)

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

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



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

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

$$(x,y) \rightarrow \{E_{1,a}, E_{1,b}, E_{2,a}, E_{2,b}, \dots, E_{n,a}, E_{n,b}\}$$

Tamura – Coarseness (cont.)

- Step 3: select the scale with the largest variation

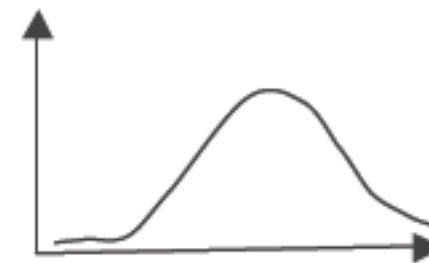
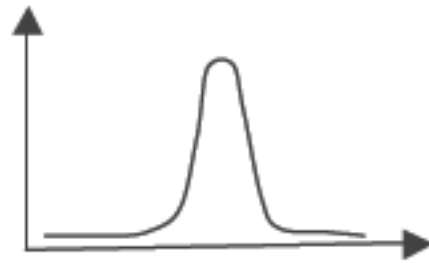
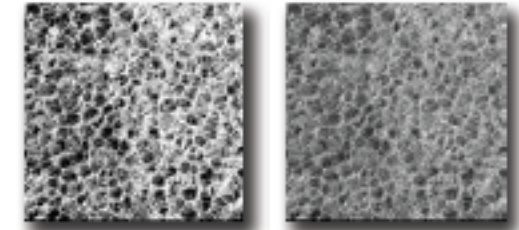
$$S_{max}(x, y) = 2^k \quad / \quad E_k = \max\{E_1, E_2, \dots, E_L\}$$

- Step 4: compute the coarseness

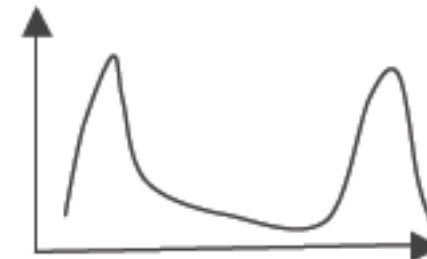
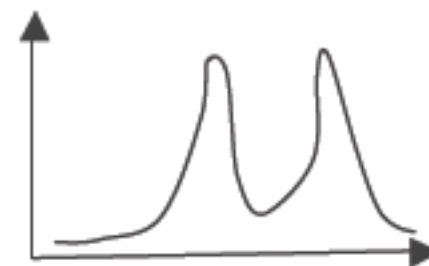
$$M_{\text{crs}} = \frac{1}{n \times m} \sum_i^n \sum_j^m S_{max}(i, j)$$

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)

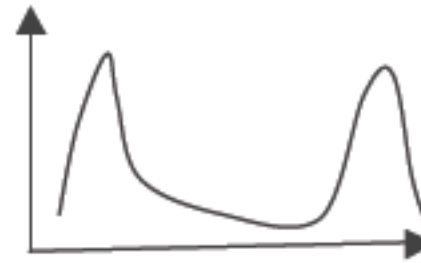
$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

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

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

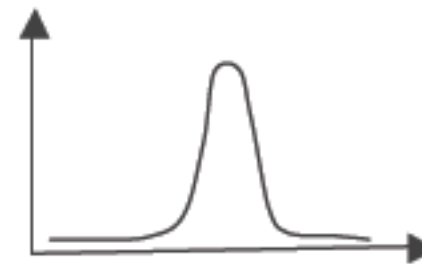
Tamura – Contrast (cont.)

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



**distribution with
two separate peaks**

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

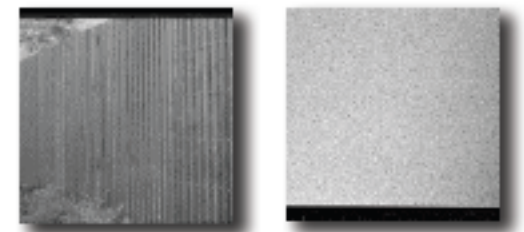


unimodal distribution

- Contrast estimate is given by:

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

Tamura – Orientation



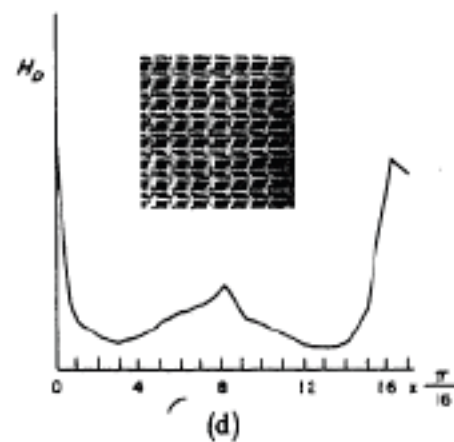
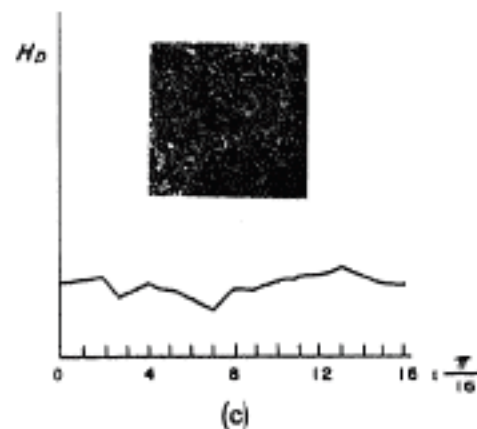
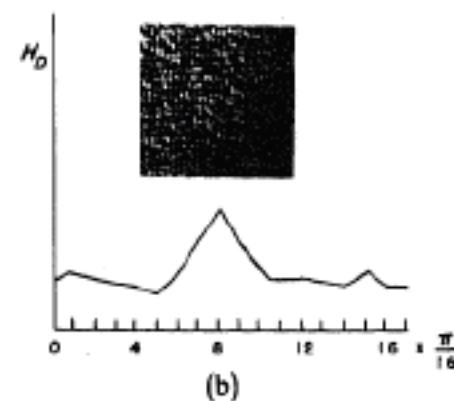
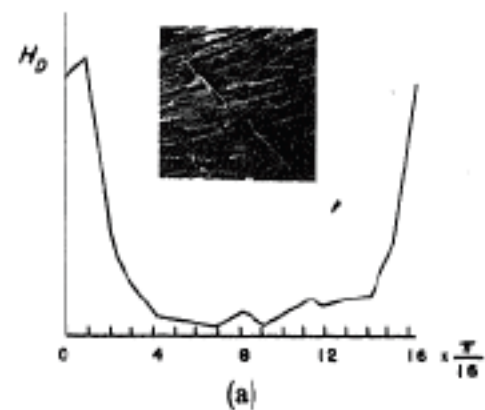
- Building the histogram of local edges at different orientations $H_D(k)$

- By deriving the edge magnitude at X and Y directions

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

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

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



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

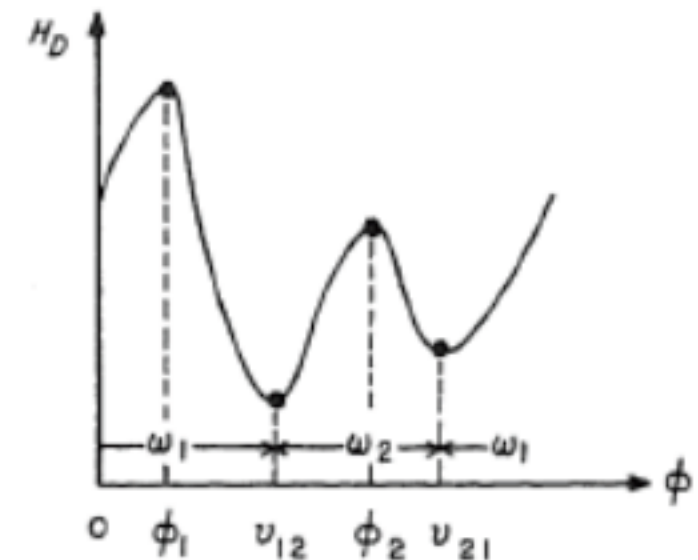
$$M_{orient} = 1 - r \cdot n_p \cdot \sum_p \sum_{\phi \in w_p} (\phi - \phi_p)^2 \cdot H_D(\phi)$$

n_p = Number of peaks

ϕ_p = Position of peak, p , in H_D

w_p = Points in peak p

r = Normalisation factor



(MR)SAR

[Mao'92]

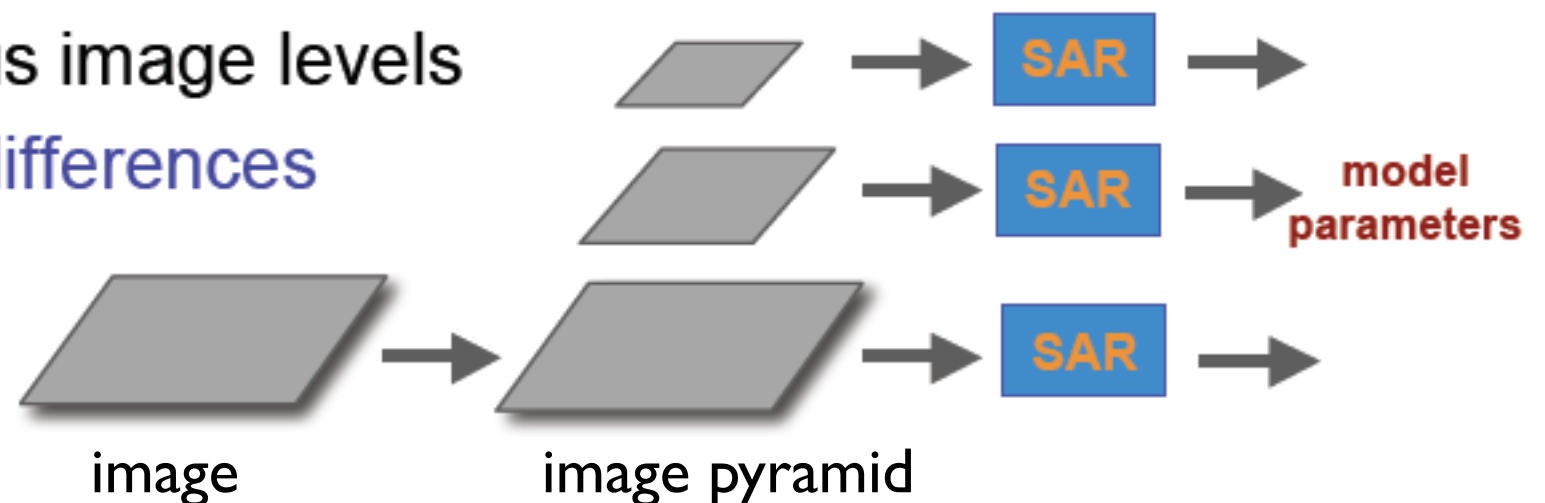
- Each pixel is a random variable whose value is estimated from its neighboring pixels + noise
 - A kind of Markov Random Field model



- **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: 0° , 45° , 90° , 135° , non direction and no edge.

- Utilizing the filters



90° edge



0° edge



45° edge



135° edge



non-directional edge

- 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

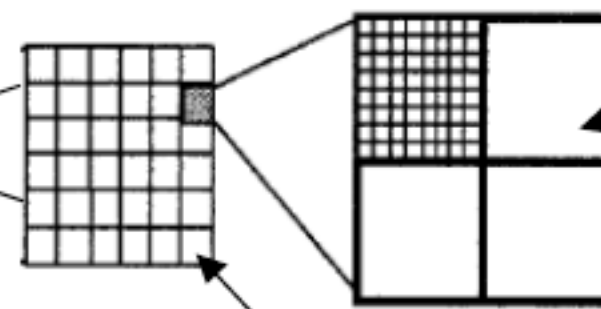
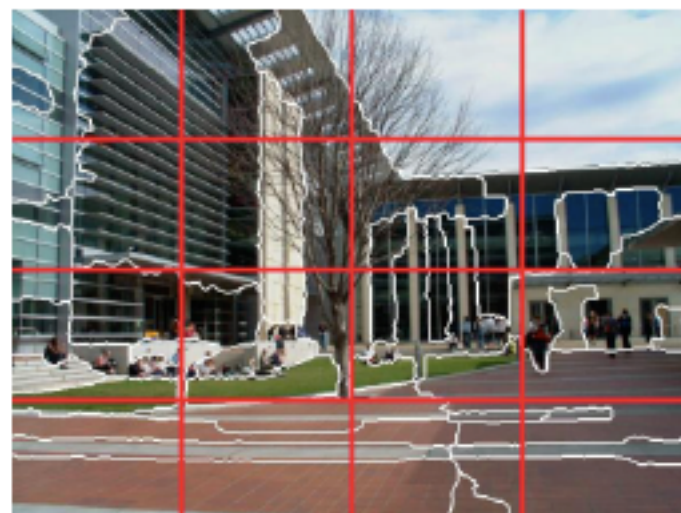


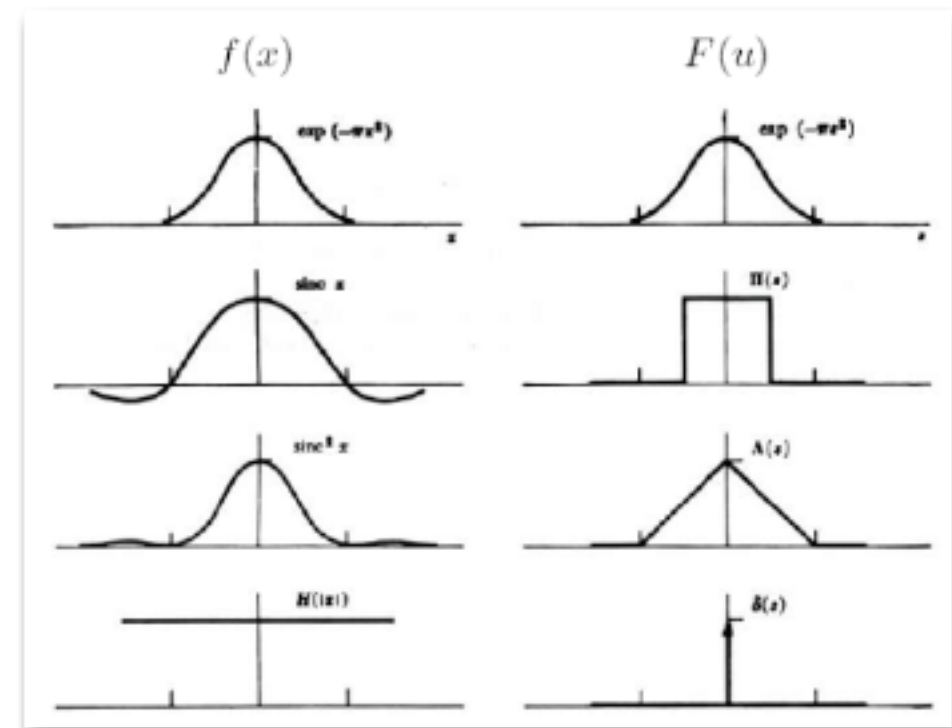
Image-block

Macro-block

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
- 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 $e^{-i2\pi(ux+vy)}$

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



Discrete Fourier Transform

- 2D DFT

$$F(k, l) = \frac{1}{N^2} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} f(a, b) e^{-i2\pi(\frac{ka}{N} + \frac{lb}{N})}$$

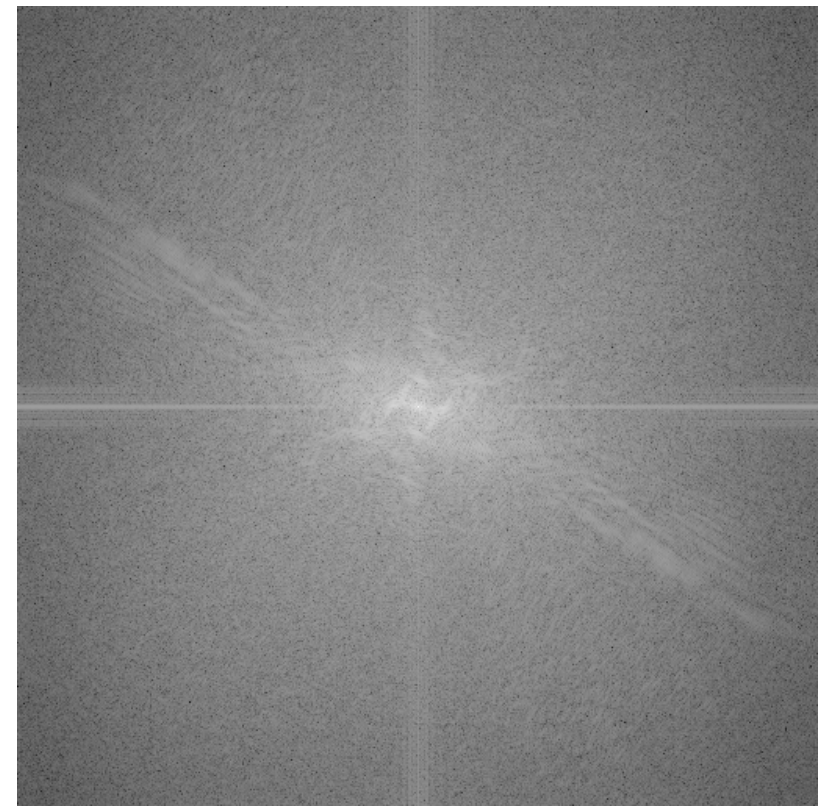
- 2D IDFT

$$f(a, b) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} F(k, l) e^{i2\pi(\frac{ka}{N} + \frac{lb}{N})}$$

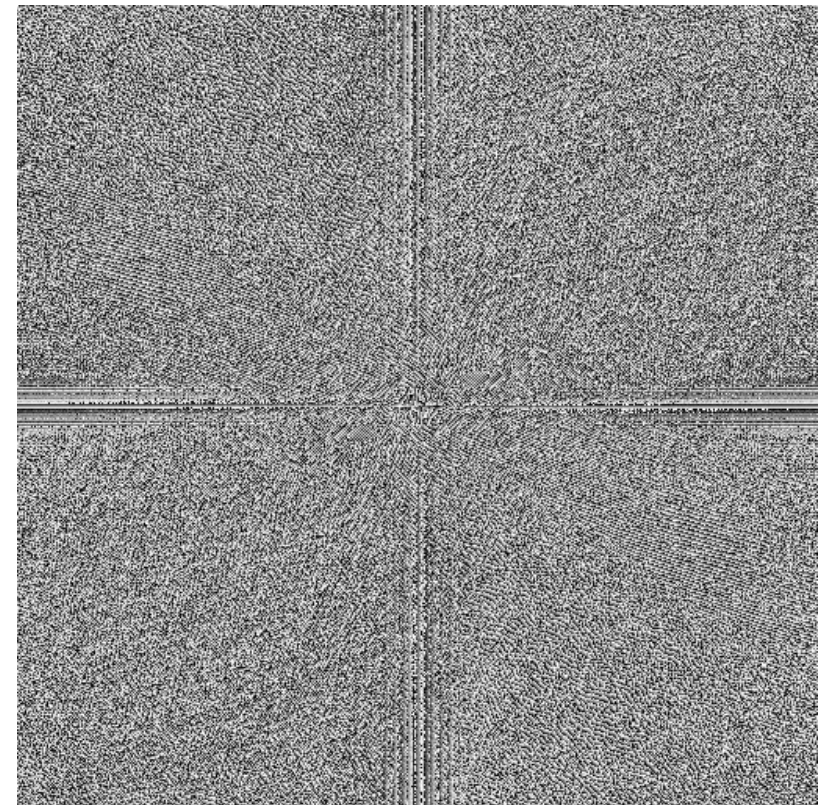


Zebra

Fourier
Transform



magnitude transform

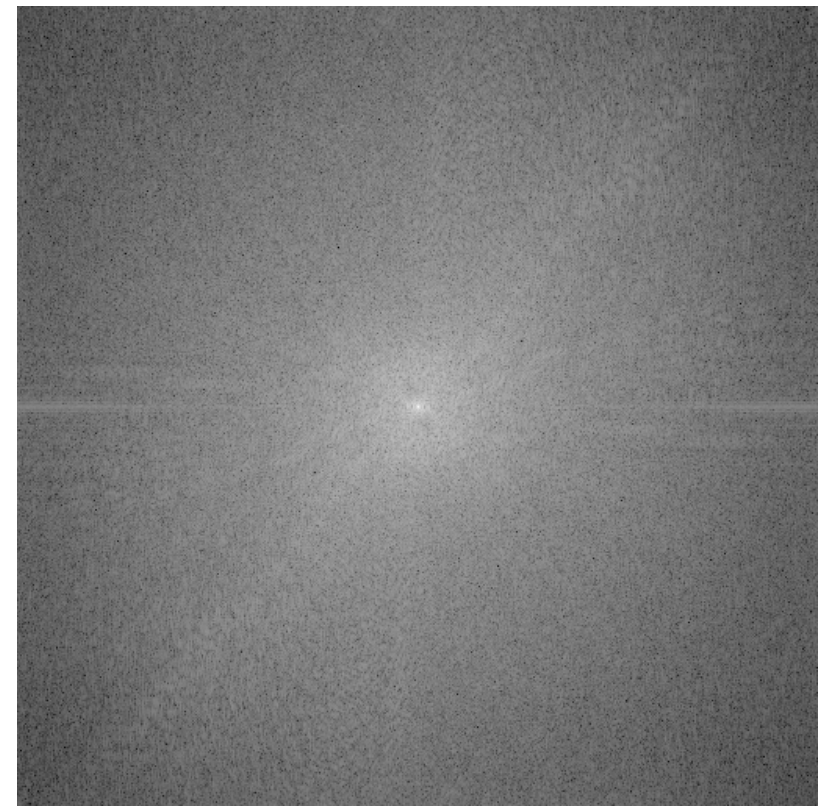


phase transform

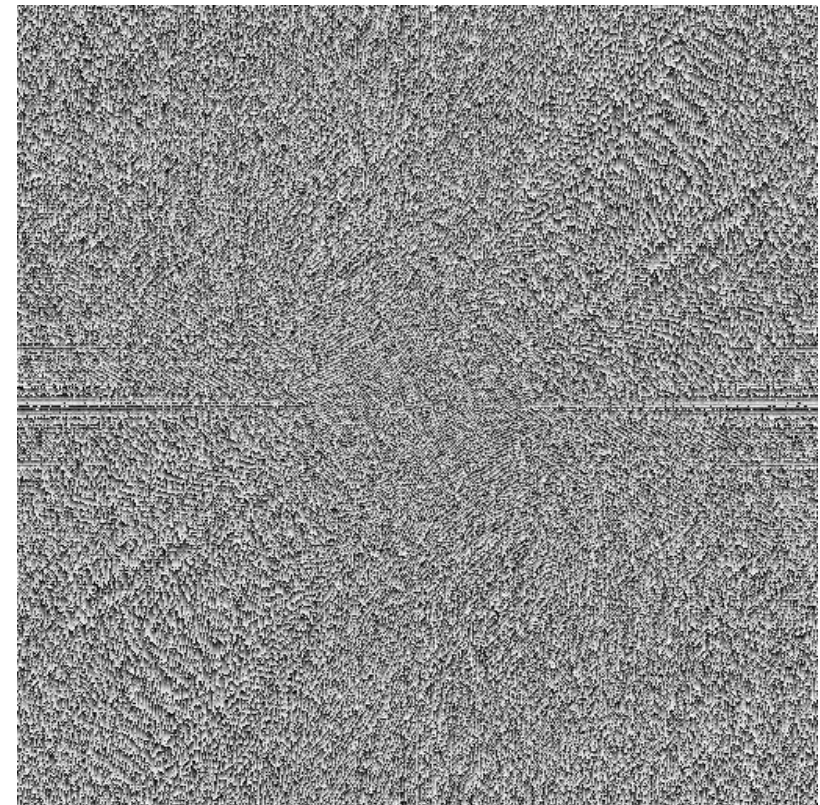


Leopard

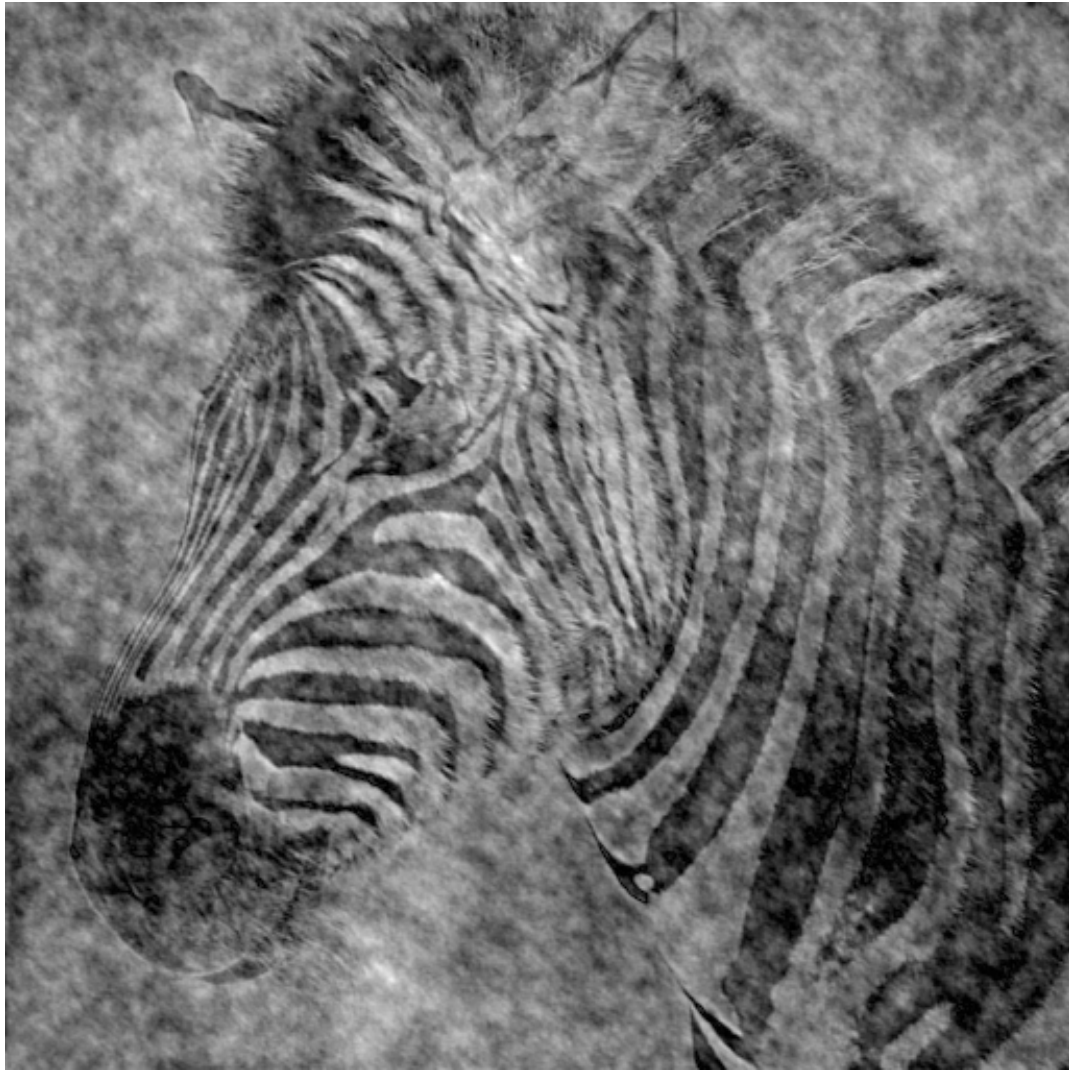
Fourier
Transform



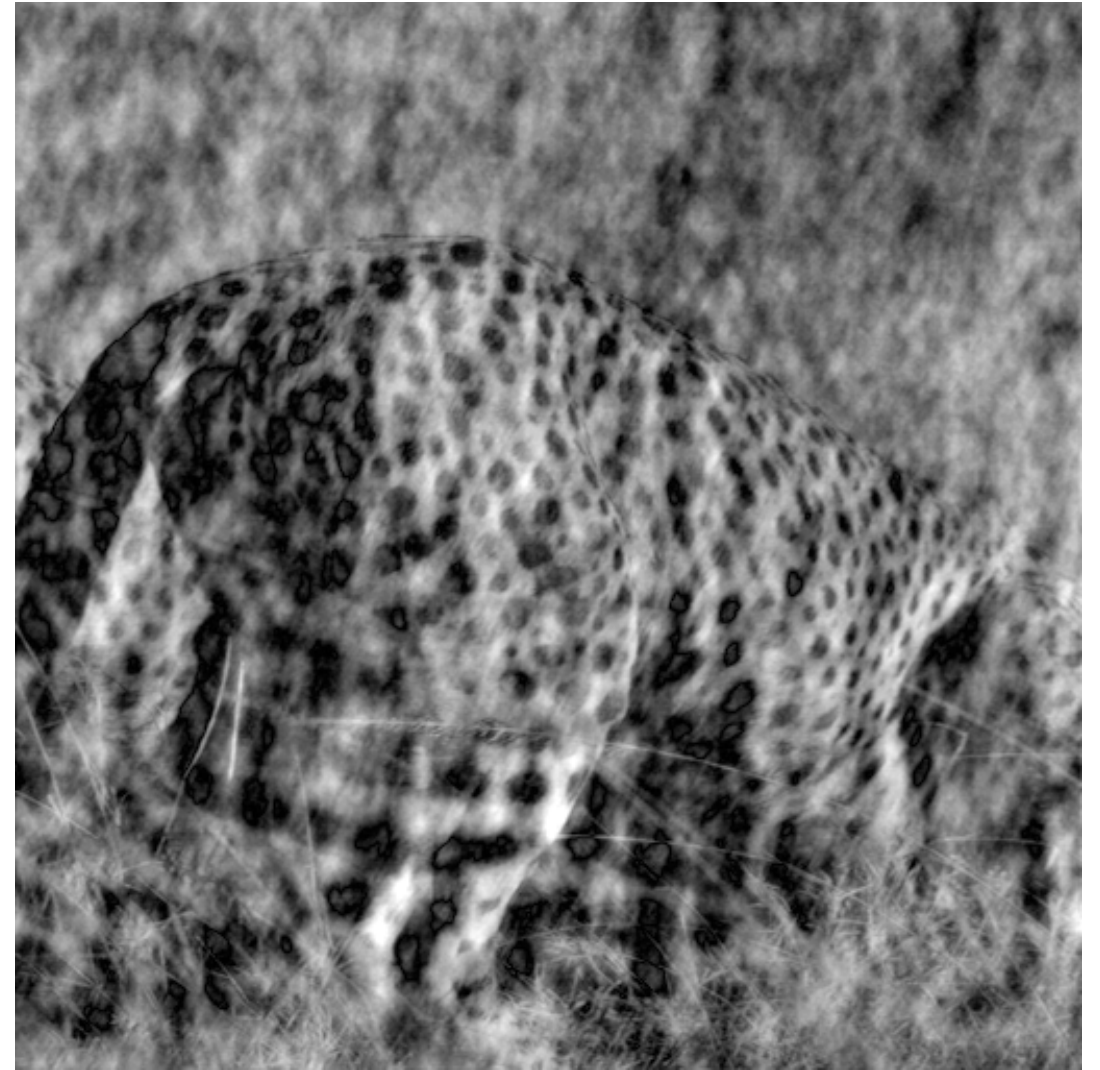
magnitude transform



phase transform

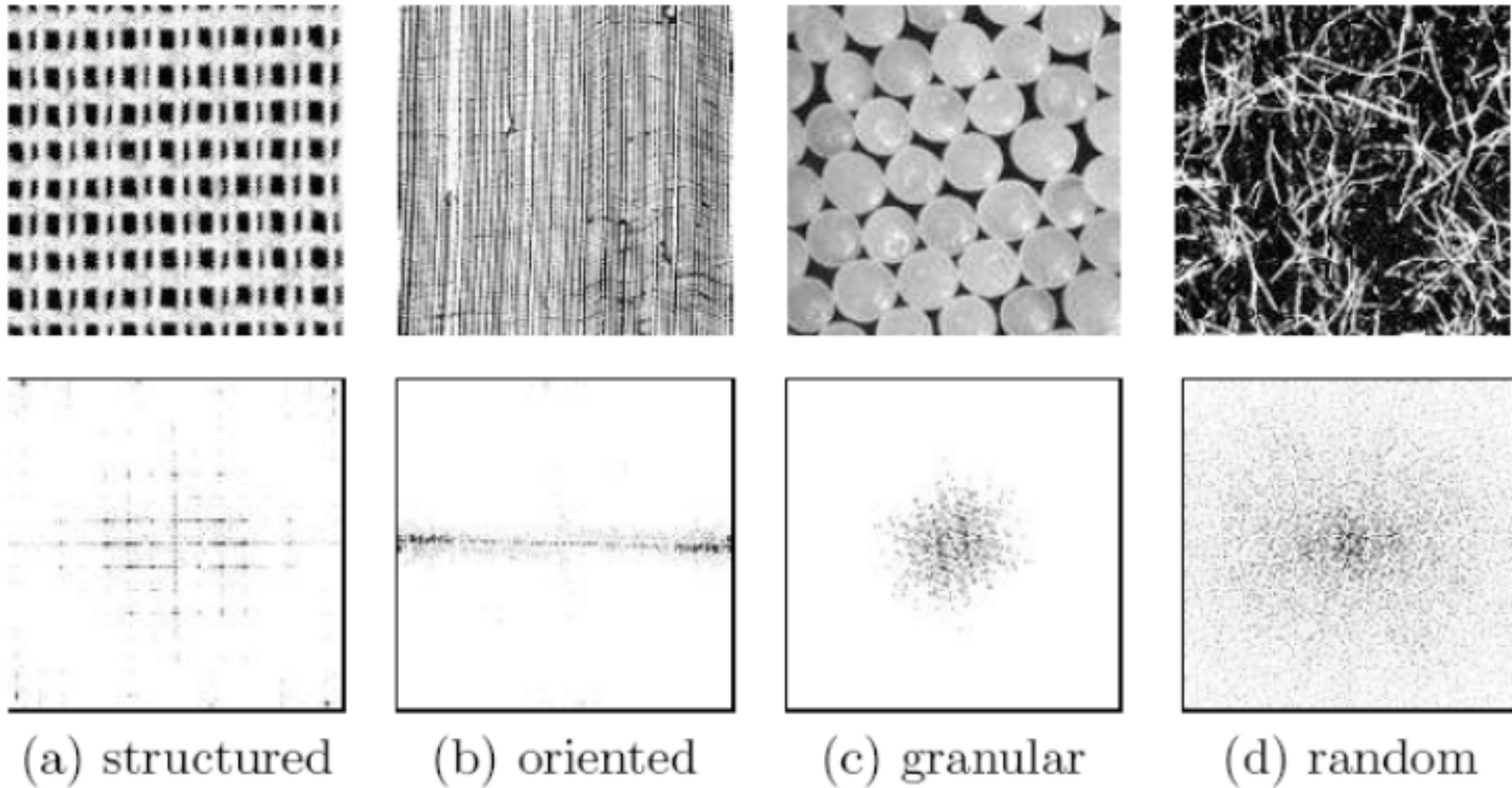


Zebra's phase
+ Leo's mag



Leo's phase
+ Zebra's mag

Natural Images and Their FT



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

Frequency Domain Features

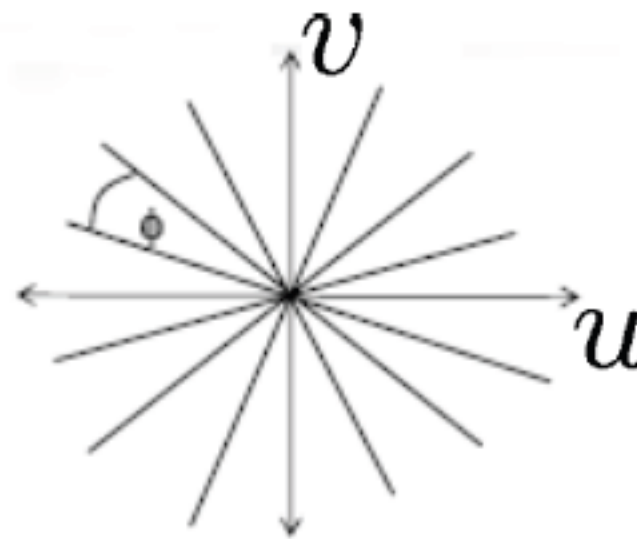
Fourier domain energy distribution

- Angular features (directionality)

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

where,

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

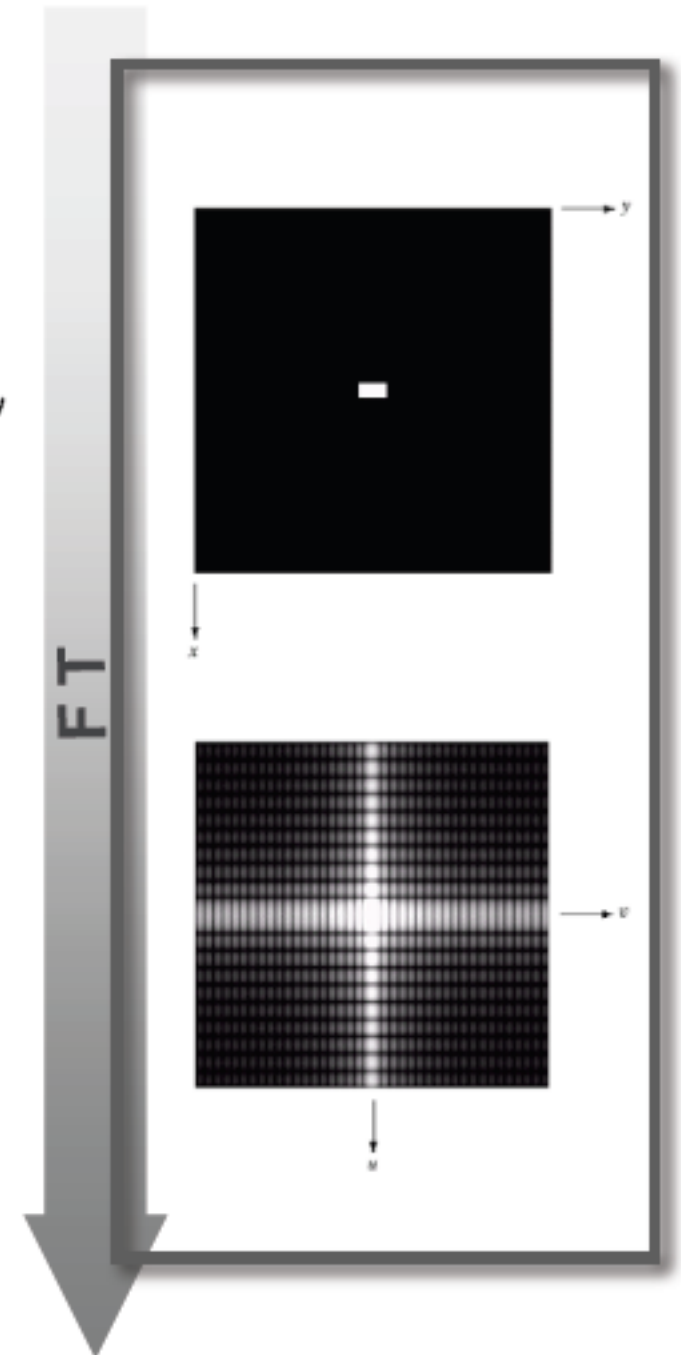
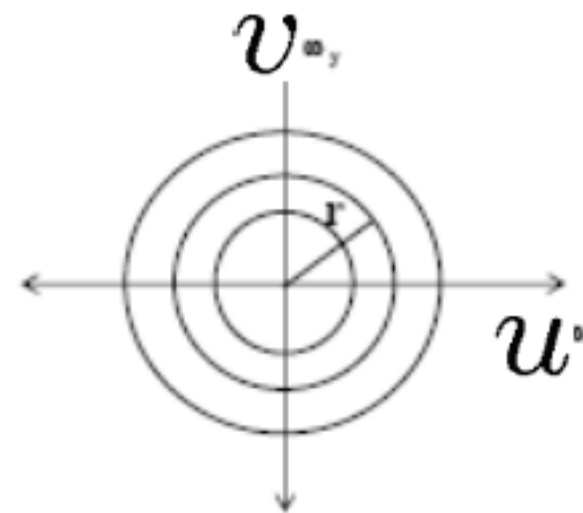


- Radial features (coarseness)

$$V_{r_1r_2}^{(r)} = \int \int |F(u, v)|^2 du dv$$

where,

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



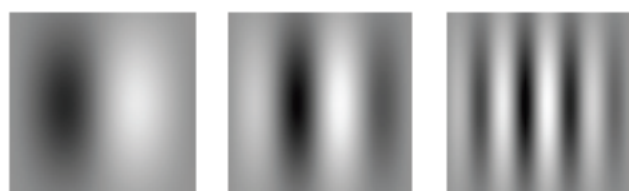
Uniform division may not be the best!!

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 $\sigma = \text{infinity}$, similar to FT
- We need to apply a number of Gabor filters are different scales, orientations, and spatial frequencies



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



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

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



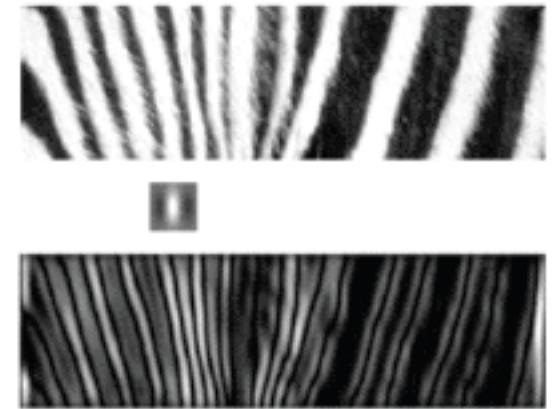
zebra image

Gabor kernel

magnitude of
the filtered image

Gabor Texture (cont.)

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

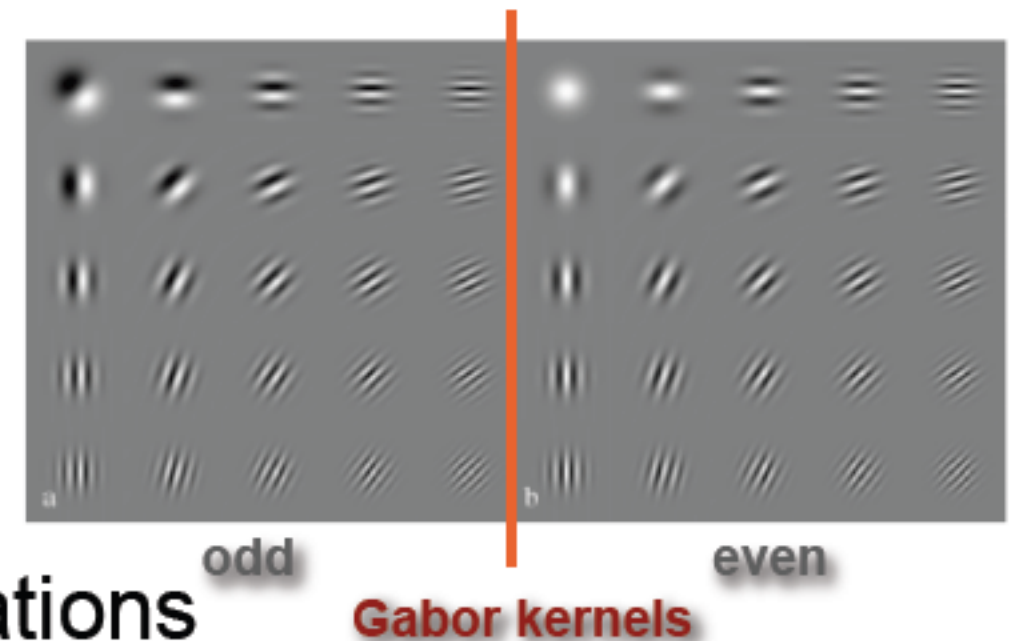


$$W_{mn}(x, y) = \int I(x_1, y_1) h_{mn}(x - x_1, y - y_1) dx_1 dy_1$$

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

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

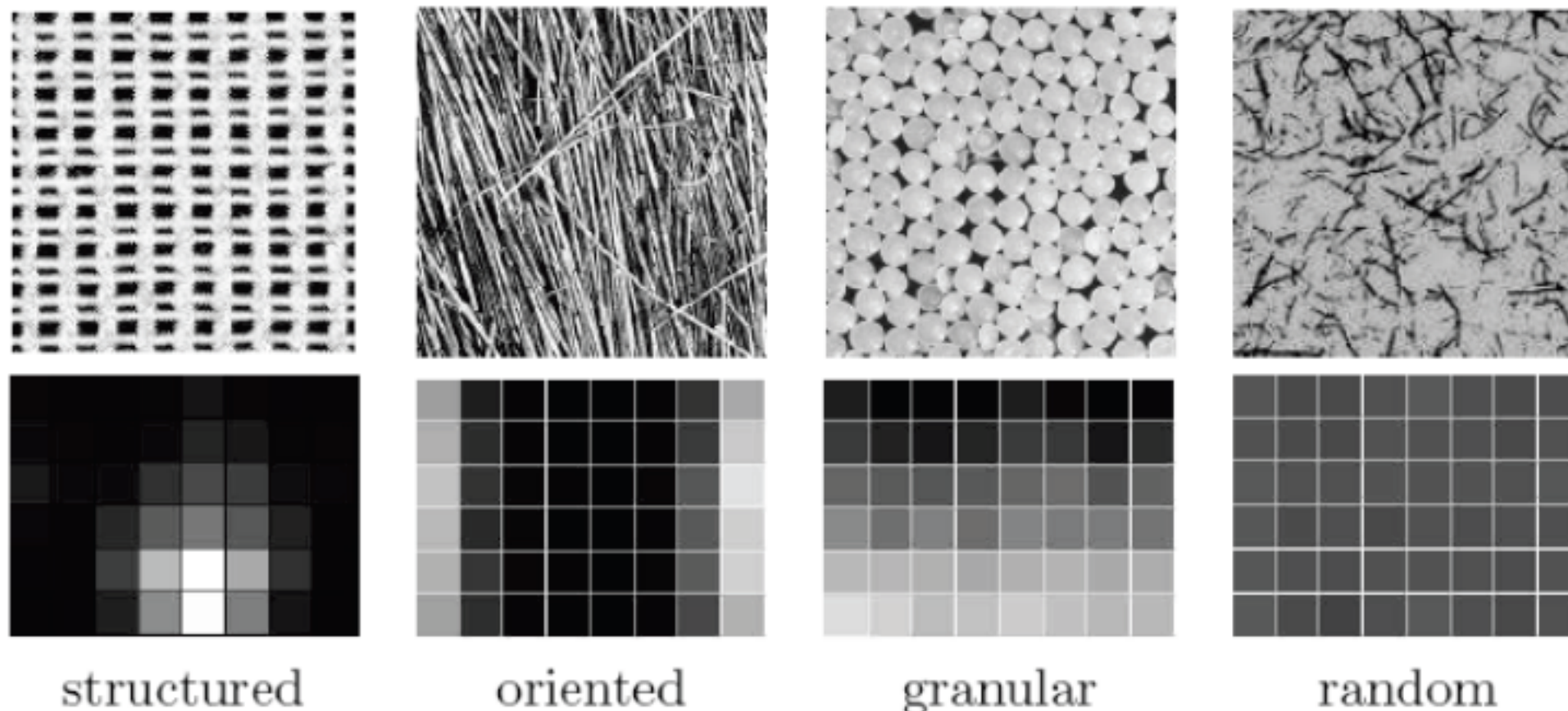
$$\sigma_{mn} = \sqrt{\int \int (|W_{mn}(x, y)| - \mu_{mn})^2 dx dy}$$



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

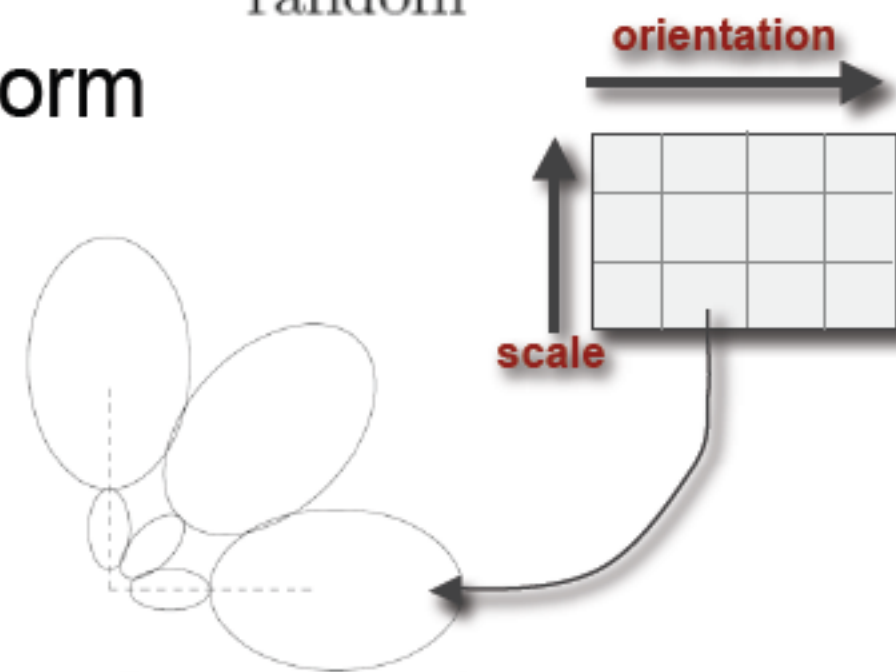
$$\bar{v} = [\mu_{00}, \sigma_{00}, \mu_{01}, \dots, \mu_{35}, \sigma_{35}]$$

Gabor Texture (cont.)



- 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)

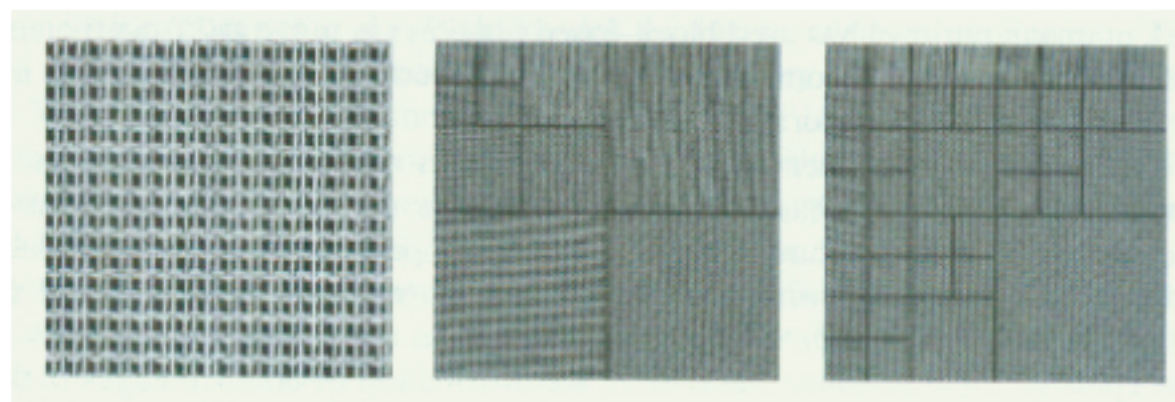
■ PWT: pyramid-structured wavelet transform

- Recursively decomposes the LL band
- Feature dimension $(3 \times 3 \times 1 + 1) \times 2 = 20$

■ TWT: pyramid-structured wavelet transform

- Some information in the middle frequency channels
- Feature dimension $40 \times 2 = 80$

LL3	HL3	HL2	HL1
LH3	HH3		
LL2		HH2	
LH1			HH1



original image

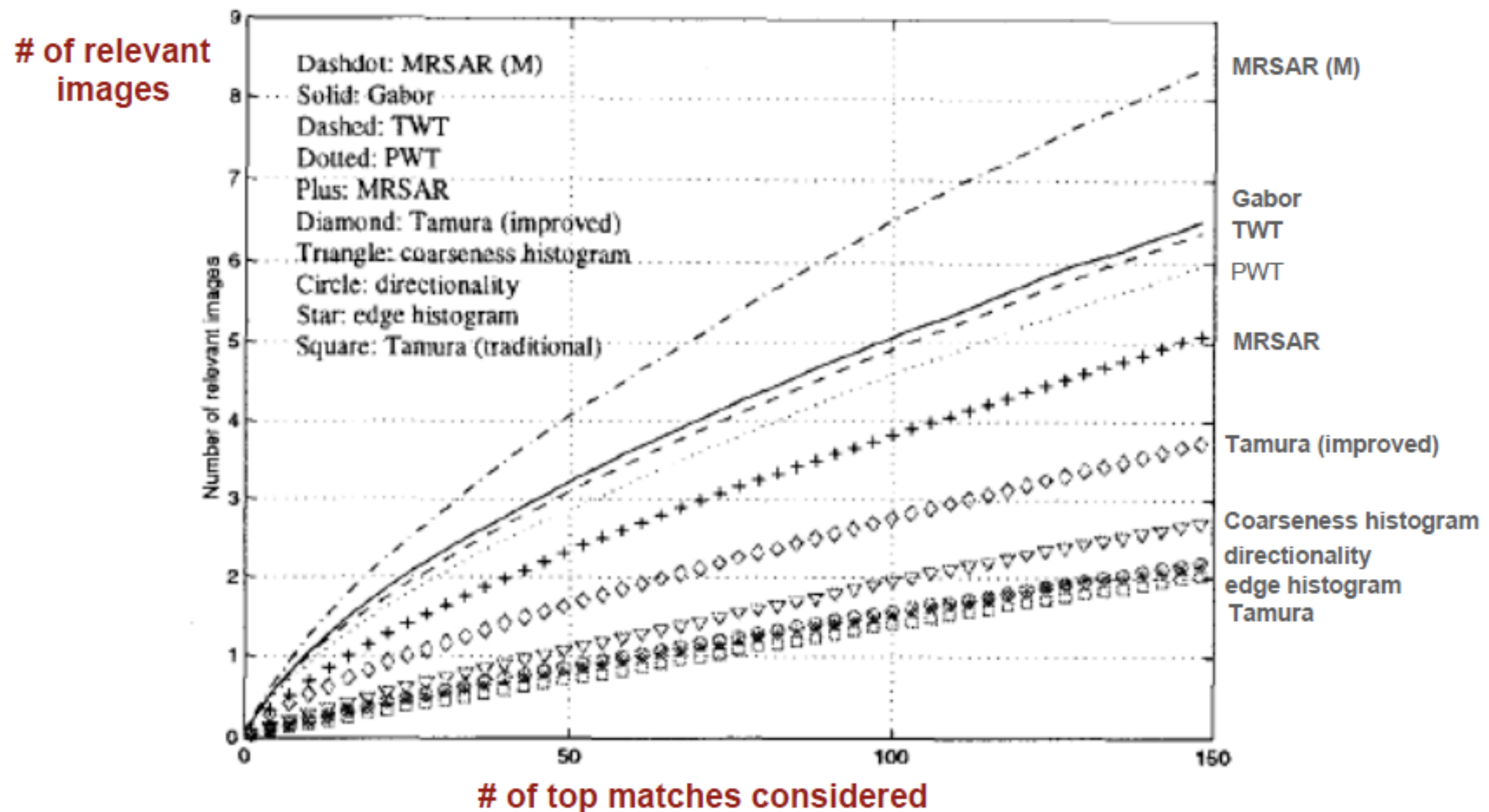
PWT

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

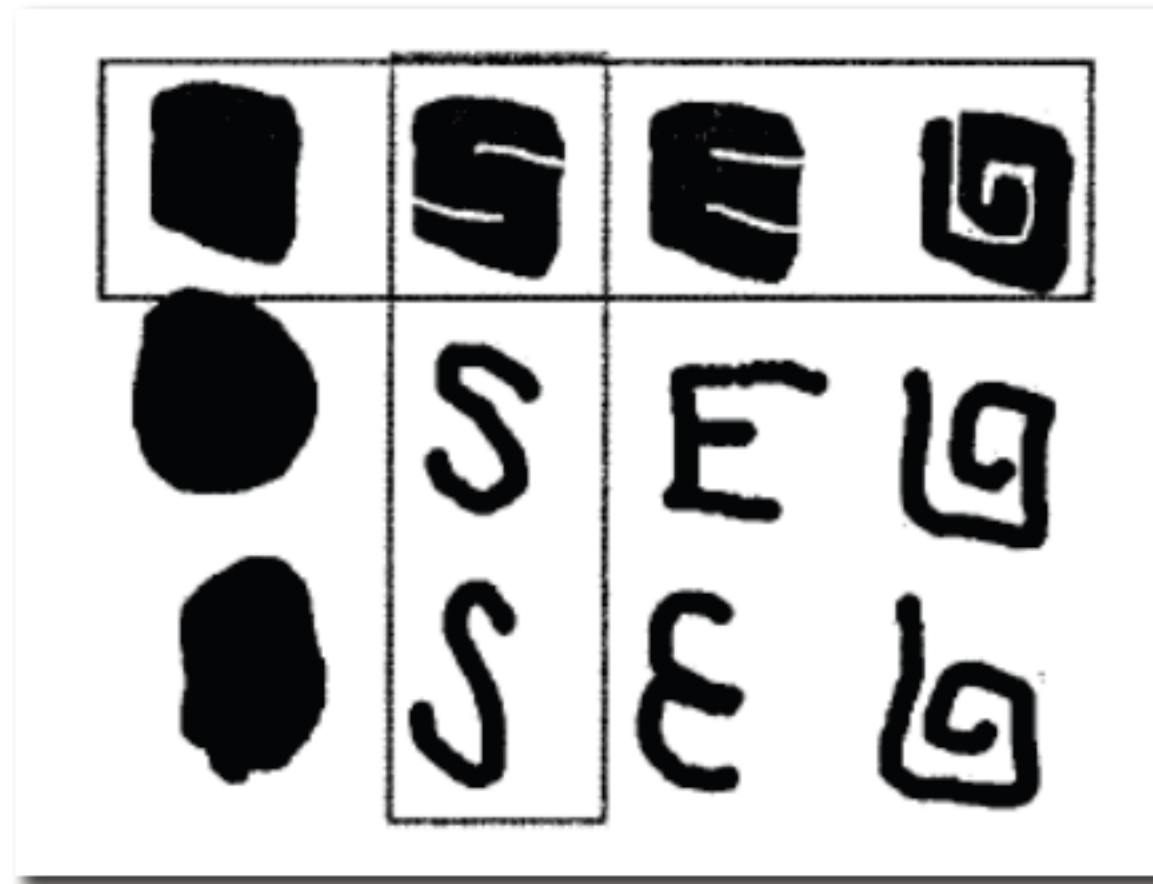


Image shape features

- 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)



(a)



(b)



(c)



(d)



(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://www.virage.com/cgi-bin/query-e>
- RetrievalWare
 - <http://vrw.excalib.com/cgi-bin/sdk/cst/cst2.bat>
- Photobook
- MARS
 - <http://jadzia.ifp.uiuc.edu:8000>

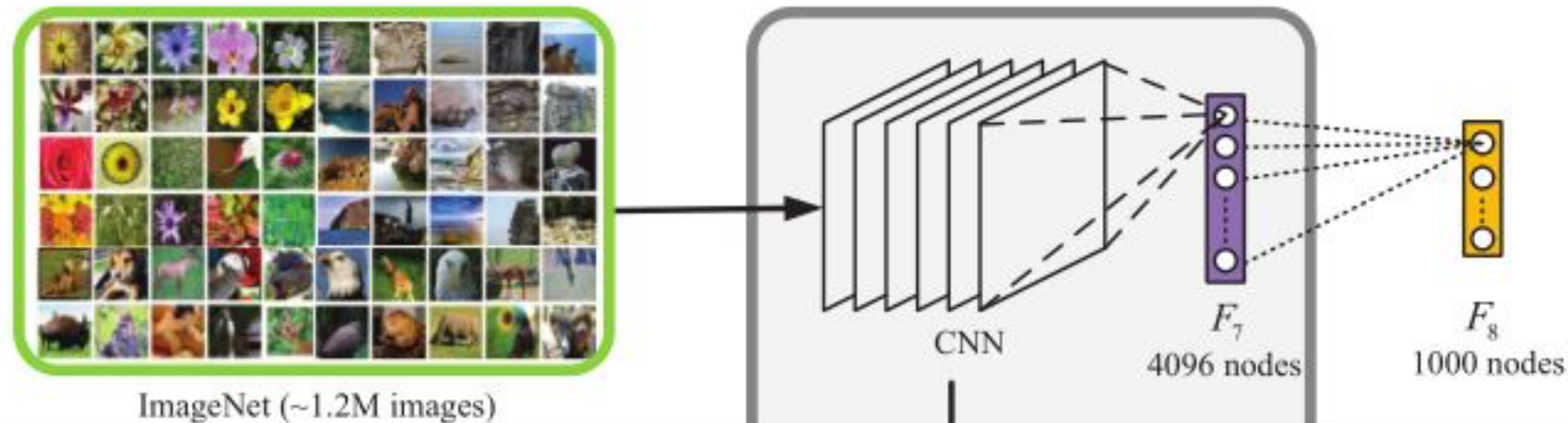
Practical image retrieval systems (cont.)

- 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

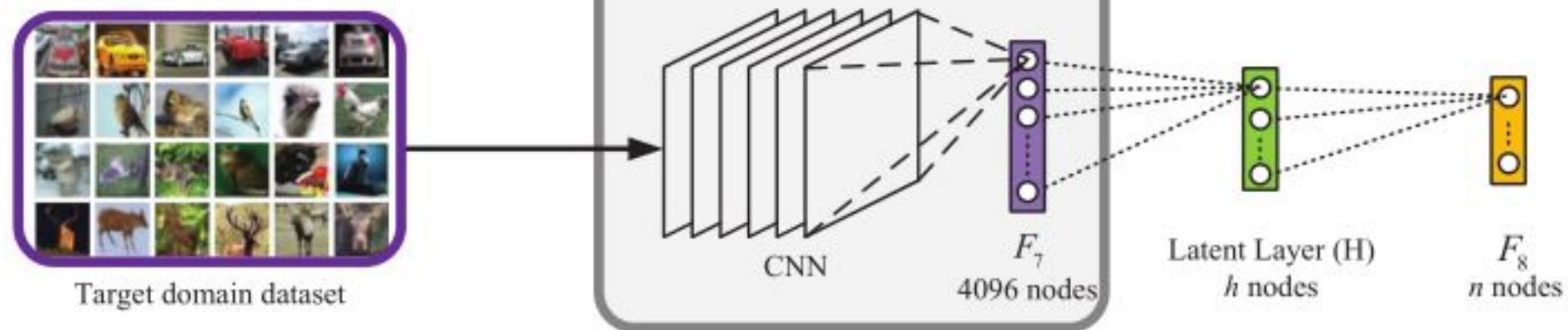


Morden Approaches

Module1: Supervised Pre-Training on ImageNet



Module2: Fine-tuning on Target Domain



<https://icodingc.github.io/xuesen/deeplearning/2017/02/21/image-retrieval.html>

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