Feature Extraction 2

李东晓

lidx@zju.edu.cn

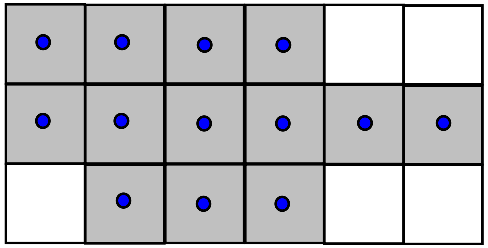
Contents

- Region Feature Descriptors
- Principle Components as Feature Descriptors
- Whole-Image Features
- Scale-Invariant Feature Transform (SIFT)

Simple Descriptors

1. 区域面积

$$A = \sum_{(x,y)\in R} 1$$



区域面积计算

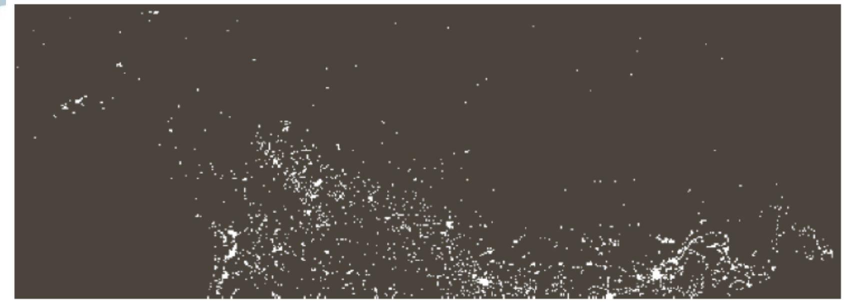
2. 区域重心(区域形心)

$$\bar{x} = \frac{1}{A}\sum_{(x,y)\in R} x$$

$$\bar{y} = \frac{1}{A}\sum_{(x,y)\in R} y$$

A是区域R内的像素点数，即面积

Infrared images of the Americas at night



Region no. (from top)	Ratio of lights per region to total lights
1	0.204
2	0.640
3	0.049
4	0.107

Earth at Night

Shape Descriptor

1. 形状因子（旋转、尺度不变）

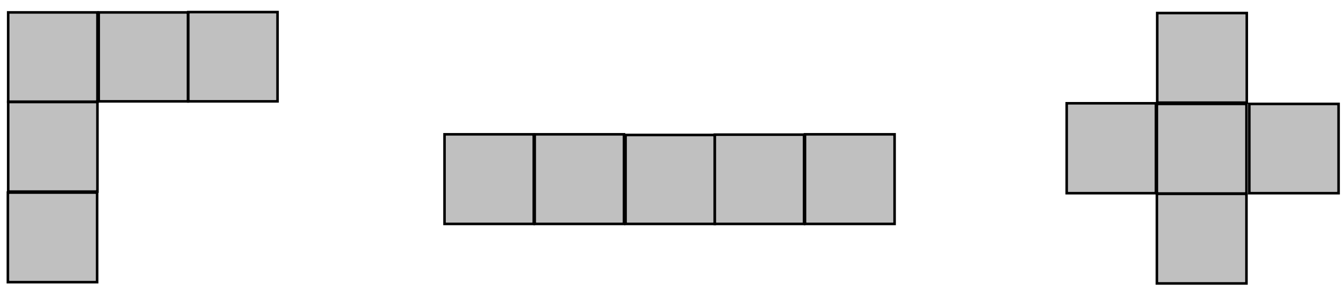
–区域的周长为P，面积为A，则形状因子（或叫圆形度）为：

$$F = \frac{P^2}{4\pi A} = \frac{(2\pi r)^2}{4\pi(\pi r^2)}$$

–如果区域为圆形，则形状因子值最小，F=1

–对数字图像，F＝1：

- 正八边形（四连通）；正菱形（八连通）



有相同形状因子的三个区（周长=12,面积=5）

2. 边界能量

– 设目标的周长为P，变量p表示边界上的一个点与起点间的弧长，即p的取值为[0,P]

– p点的曲率半径为r(p)，曲率函数为K(p) = 1/r(p)，

– 单位长度的边界平均能量定义为

$$E = \frac{1}{P}\int_0^P \left|K(p)\right|^2 dp$$

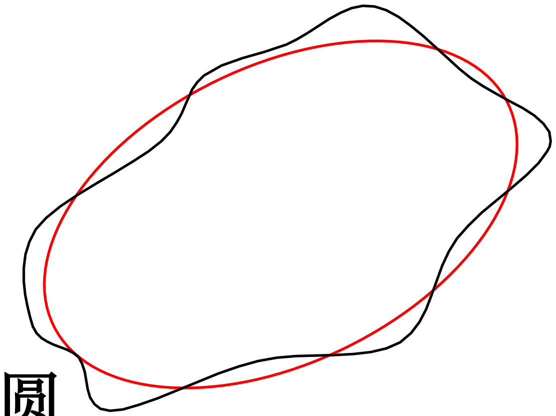
– 面积相同时，圆形区域有最小的平均边界能量

$$E_0 = \left(\frac{2\pi}{P}\right)^2 = \frac{1}{R^2}$$

R是该圆的半径

3. 拟合椭圆

- 拟合椭圆：用一个椭圆形状去拟合一组边界点
  - –它的长短轴、中心位置和主轴方向都可作为区域的特征，其中长短轴是旋转和位移不变的
  - –长短轴之比也可用于表示圆形程度



- 二次曲线的一般方程为

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

  - –如果满足 $b^2 - 4ac < 0$ ，则表示一个椭圆
  - –令**a=1**，使方程归整化，边界上的点有如下均方误差：

$$\varepsilon^2 = \sum_i \left(x_i^2 + bx_i y_i + cy_i^2 + dx_i + ey_i + f\right)^2$$

  - –令上式对参数**b, c, d, e, f**偏导数等于**0**，就可求出使均方误差最小的参数 （拟合）

4.球状性

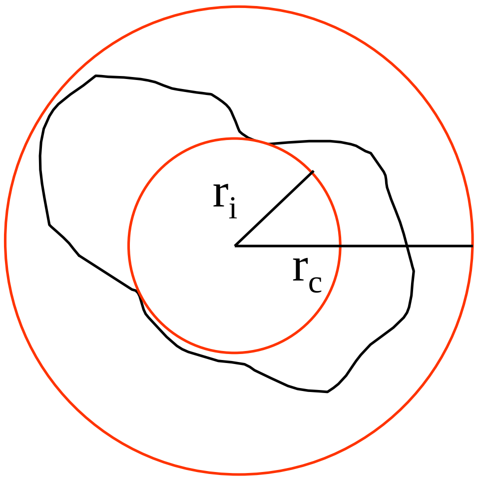
– 对于一个二维区域的边界，可找到它的一个外接圆和一个内接圆,二个圆均以区域的重心为圆心

– 设内接圆的半径为r_i，外接圆的半径为r_c，则定义球状性（sphericity）：$S = \frac{r_i}{r_c}$

– 若区域为圆形，则S=1，否则S<1

– 它在平移、旋转和尺度变化时不变

– 只要将圆换成球，可扩展到三维



5.圆形性

–先定义从区域重心到边界点的平均距离：

$$\mu_R = \frac{1}{K}\sum_{k=0}^{K-1} d_k$$

- K是边界上的像素点数，d_k是边界上点到重心的距离

–定义从区域重心到边界点的距离的方差：

$$\sigma_R = \frac{1}{K}\sum_{k=0}^{K-1}\left[d_k - \mu_R\right]^2$$

- 圆形性定义为 $C = \dfrac{\mu_R}{\sigma_R}$

在区域趋于圆时，$\sigma_R \to 0 \quad C \to \infty$

6. 矩形性

–先找到区域的面积最小外接矩形（**MER**）：

将区域的边界以**3**°左右的增量旋转**90**°，从而得到一水平放置的外接矩形**:**

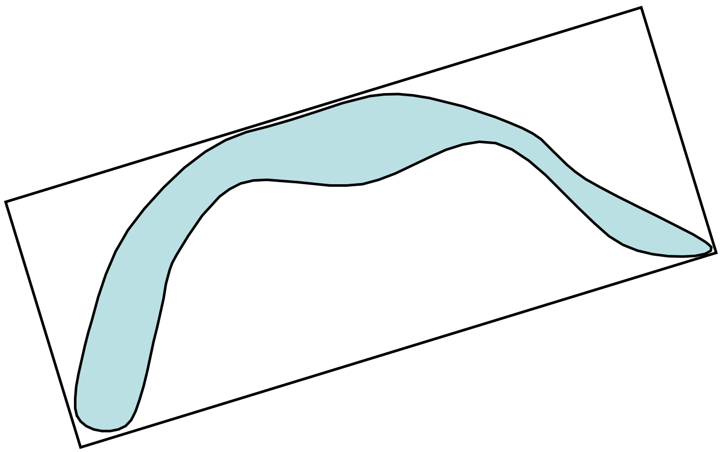
计算时只要找出**x,y**的最大和最小值就可计算外接矩形的面积

–设区域的最小外接矩形（**MER**）的面积为A_R，目标的实际面积为A_O，

•矩形拟合因子为：

$$R = A_O / A_R$$

- 矩形区域，R取得最大值1.0
- 圆形区域的R为π/4
- 对于细长并弯曲的目标，有较小的R
  - 最小外接矩形的宽长比也可作为区域形状的一个特征

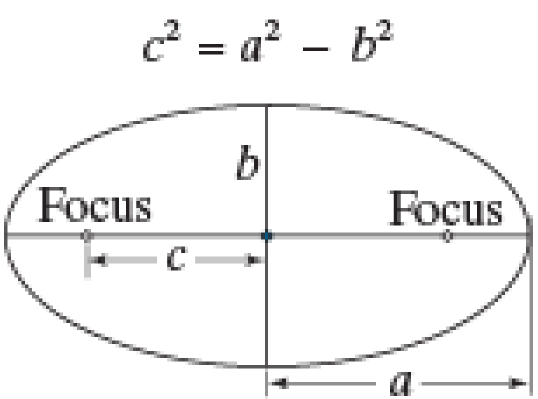


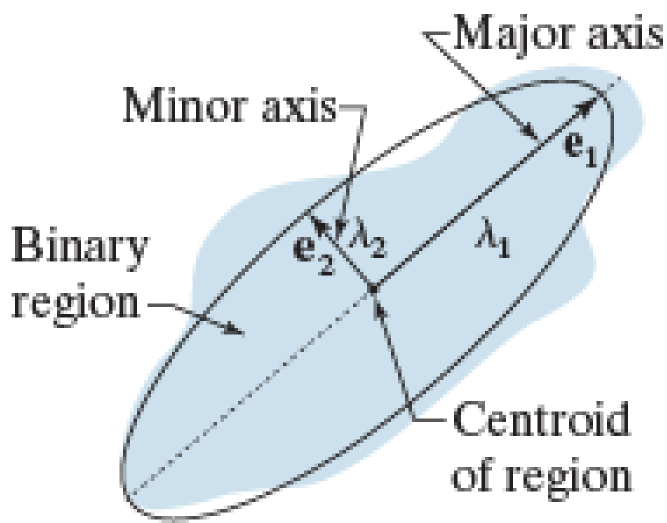
Basic Descriptors (DIP4E)

$$\text{compactness} = \frac{p^2}{A}$$

$$\text{circularity} = \frac{4\pi A}{p^2}$$

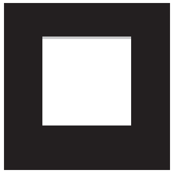
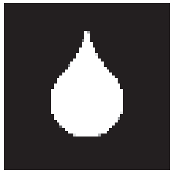
$$\text{eccentricity} = \frac{c}{a} = \frac{\sqrt{a^2 - b^2}}{a} = \sqrt{1 - (b/a)^2} \quad a \geq b$$

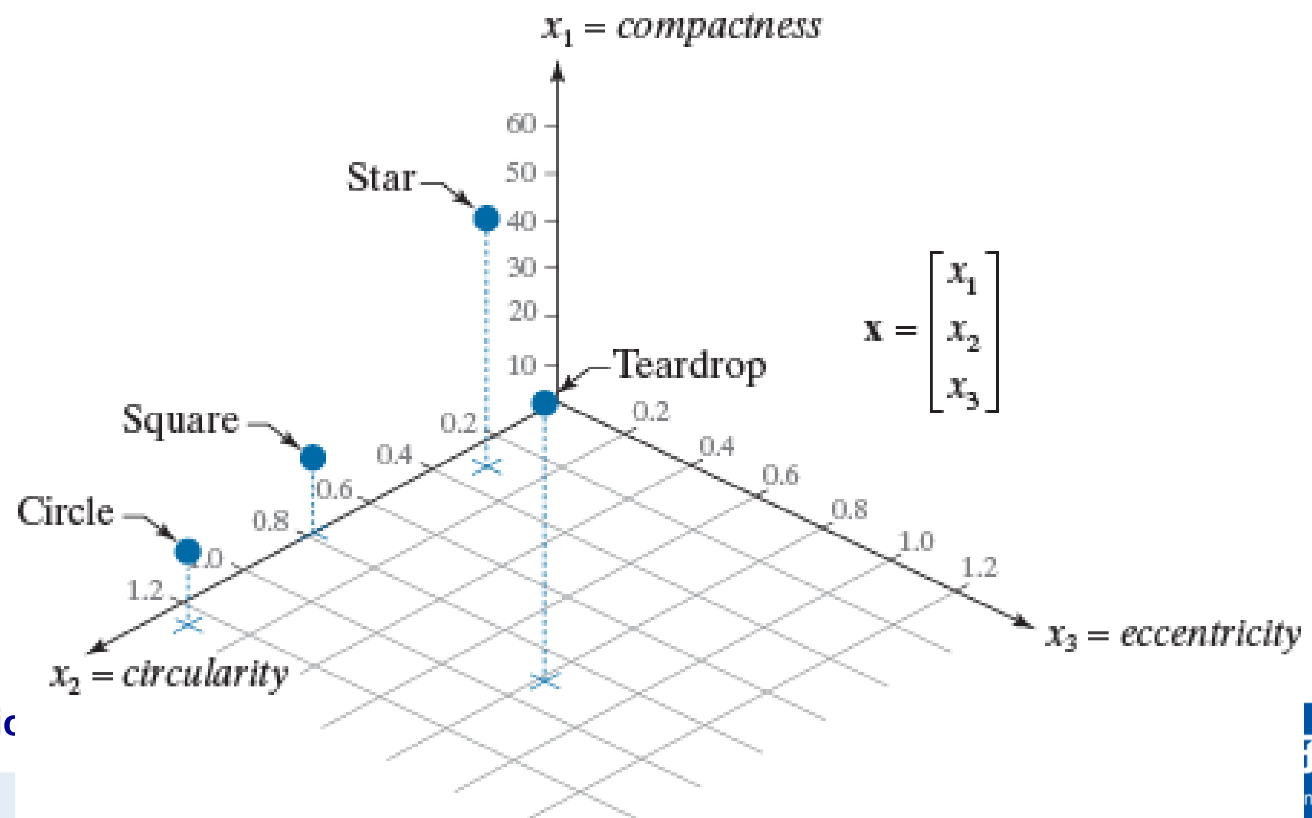




$\mathbf{e}_1$ λ_1 and $\mathbf{e}_2$ λ_2 are the eigenvectors and corresponding eigenvalues of the covariance matrix of the coordinates of the region

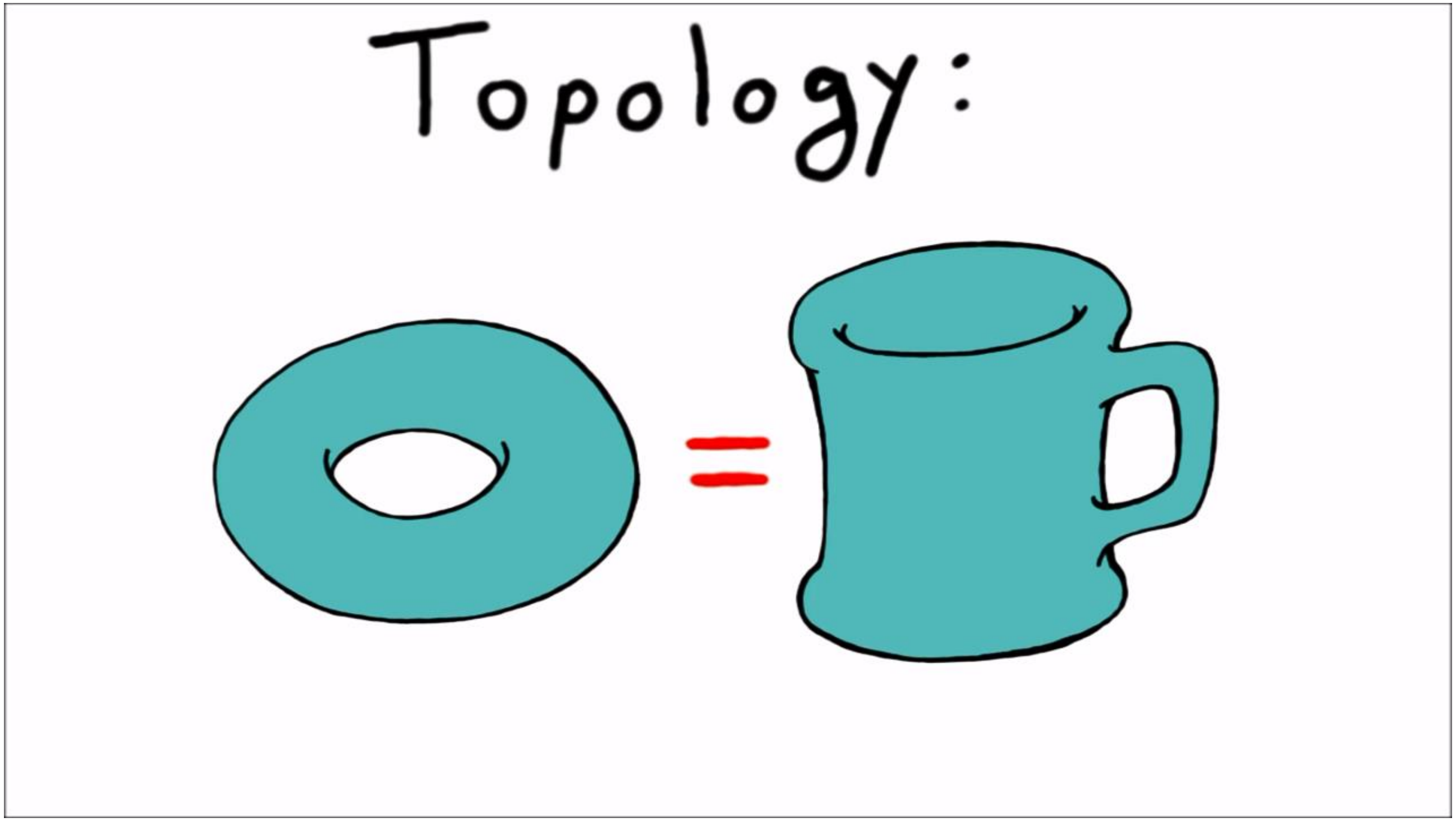
Comparison of Feature Descriptors

Descriptor				
Compactness	10.1701	42.2442	15.9836	13.2308
Circularity	1.2356	0.2975	0.7862	0.9478
Eccentricity	0.0411	0.0636	0	0.8117

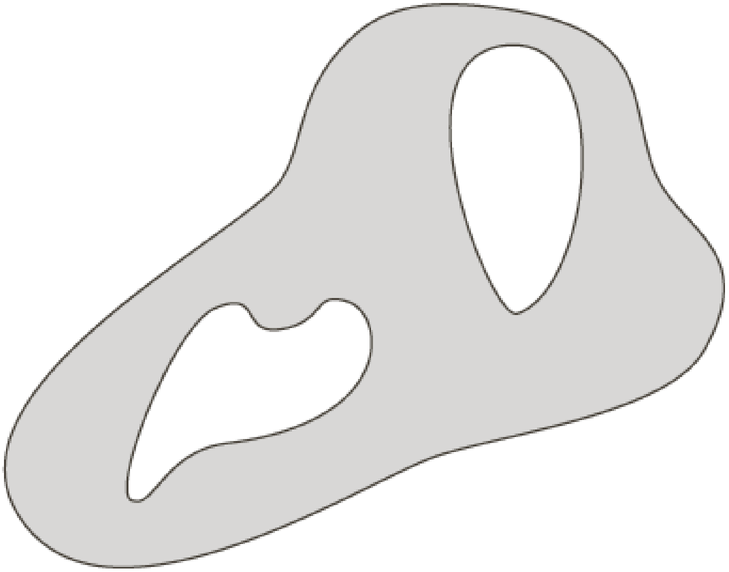


Topological Descriptors

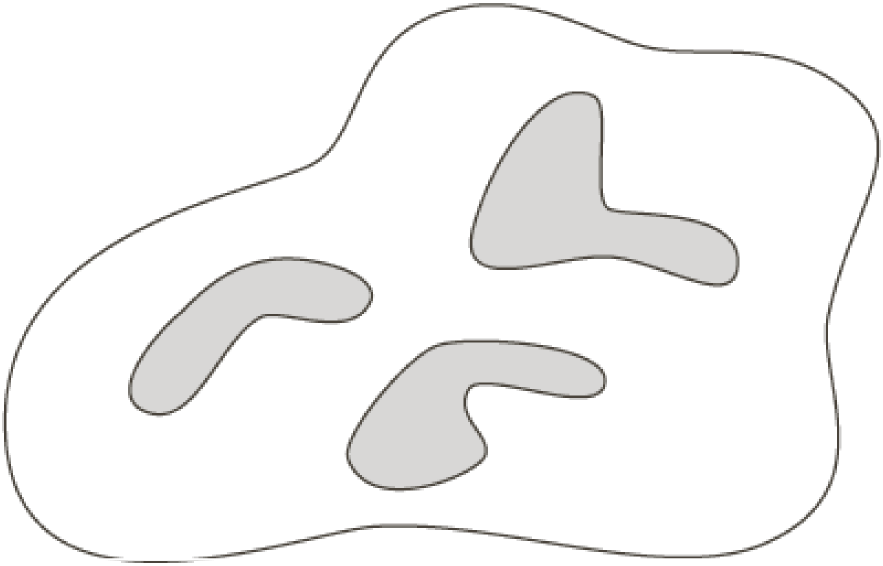
- 图形的拓扑性质
  - 是指在不发生撕裂或粘连的情况下，**不因图像橡皮膜变形（rubber-sheet distortions）而改变的性质**



- 区域的两种拓扑性质：
  - 区域的连通数C
  - 区域的孔数H

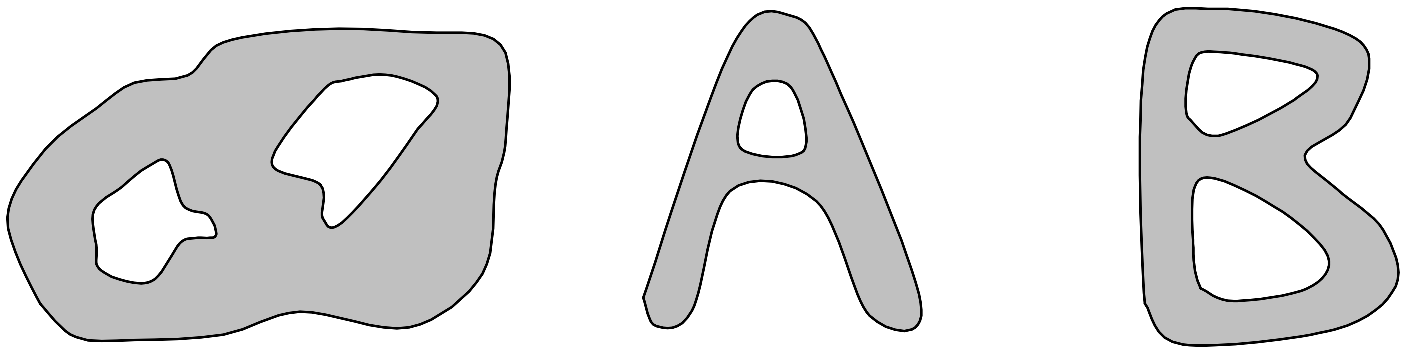


A region with two holes



A region with three connected components

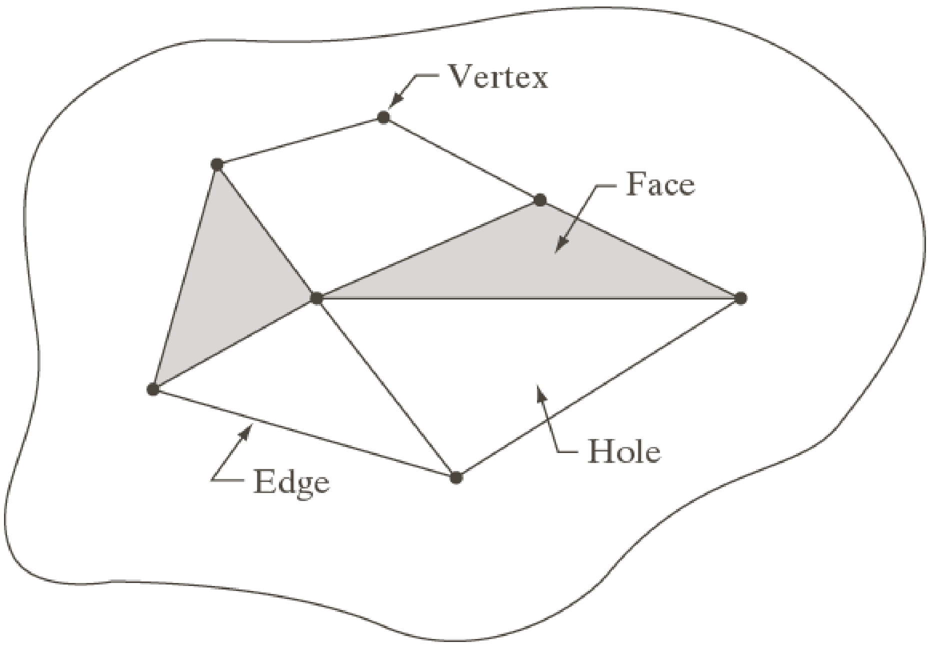
- 另一种拓扑性质：
  - 欧拉数（Euler number）：E=C-H



(a) C=1, H=2, E= -1 (b) C=1, H=1, E=0 (c) C=1, H=2, E= -1

欧拉公式

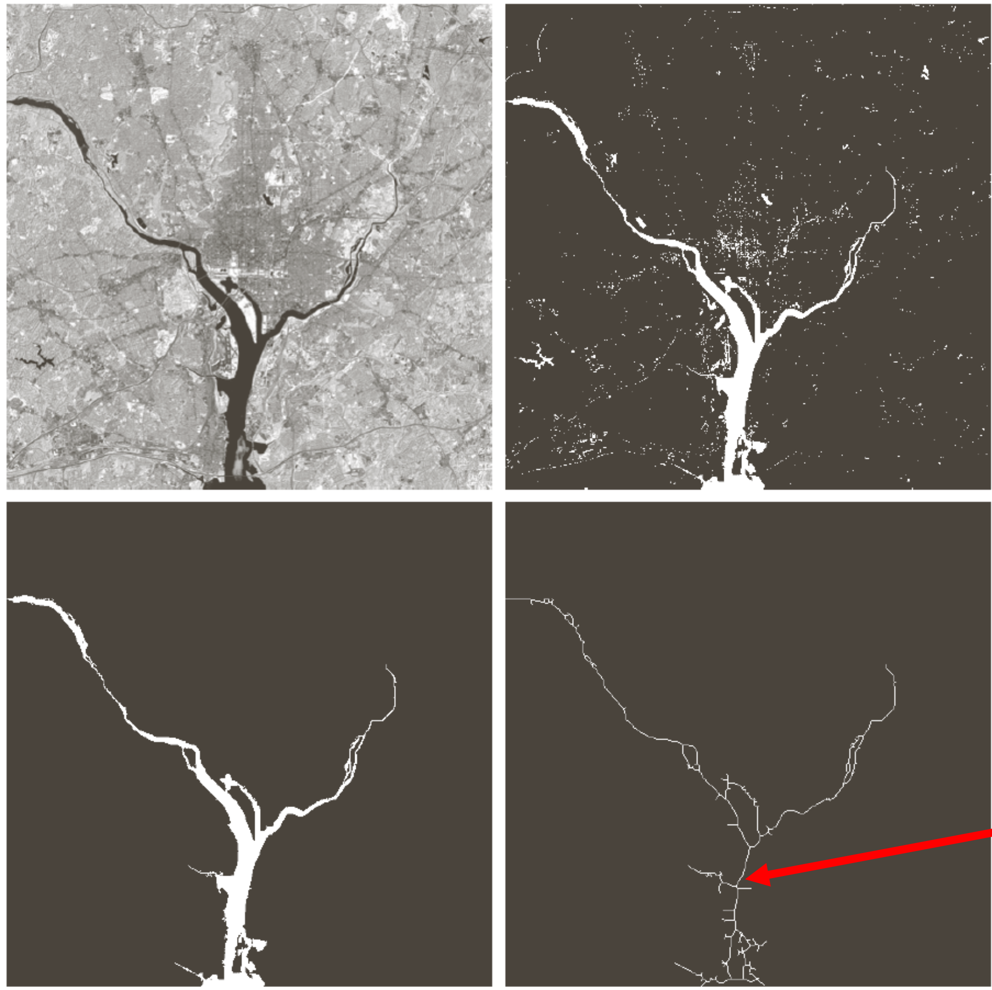
- 对于一个直线段构成的区域（称为多边形网络）
  - 若顶点数为V，边数为Q，面数为F
  - 则有欧拉定理： V-Q+F = E = C-H
  - 例：V=7，Q=11，F=2，C=1，H=3
    - 欧拉数E= -2



Use of connected components for extracting the largest features in a segmented image

Infrared image of the Washington D.C.

Thresholded



River Branch

Largest connected component

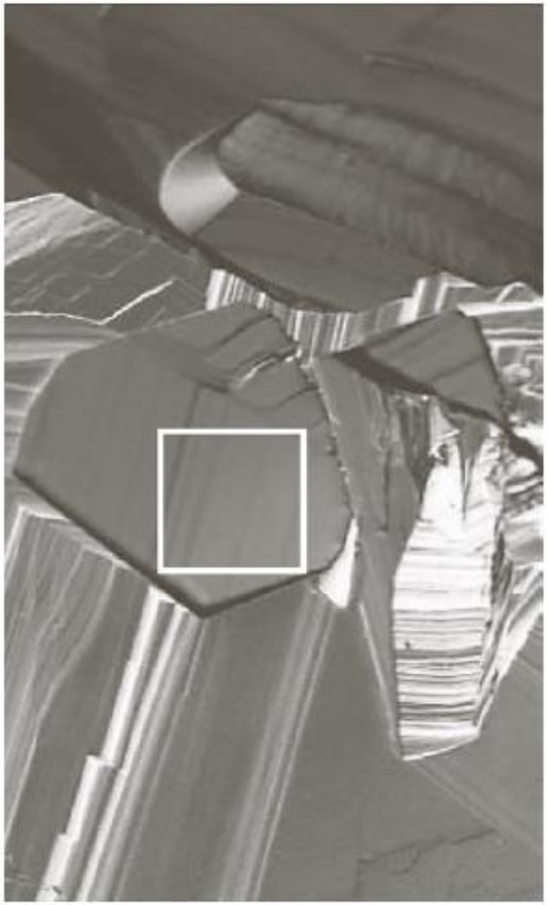
Skeleton

Texture

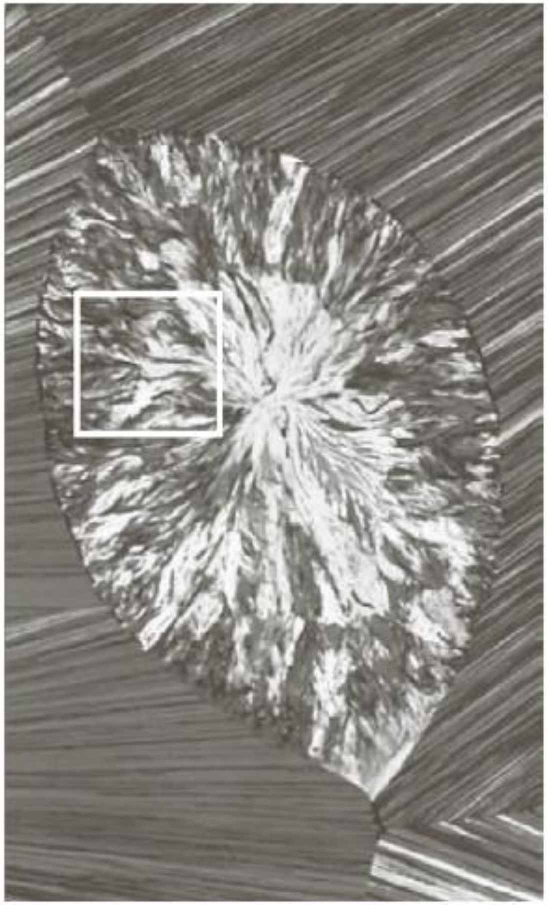
- 常用的目标纹理特征

灰度（或各彩色分量）的最大值、最小值、中值、平均值、方差及高阶矩等统计量

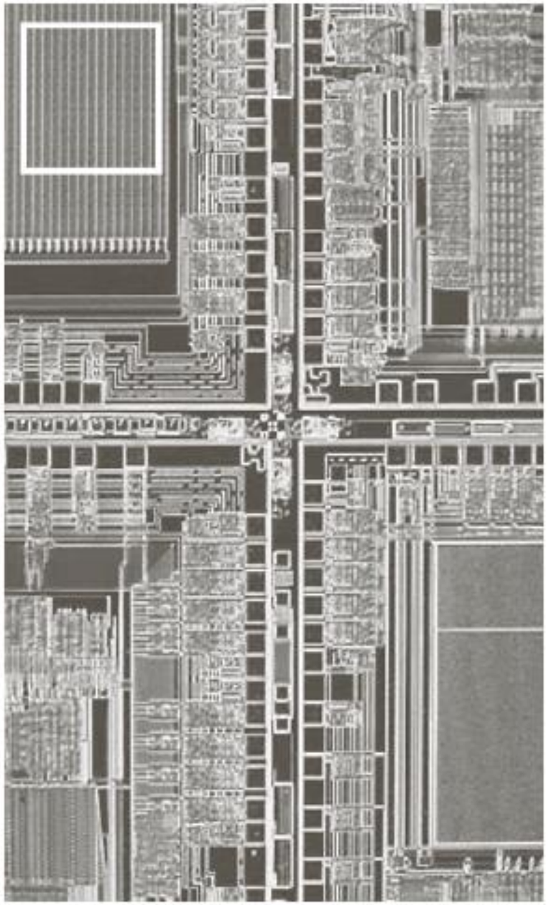
Texture Descriptors



Smooth



Coarse



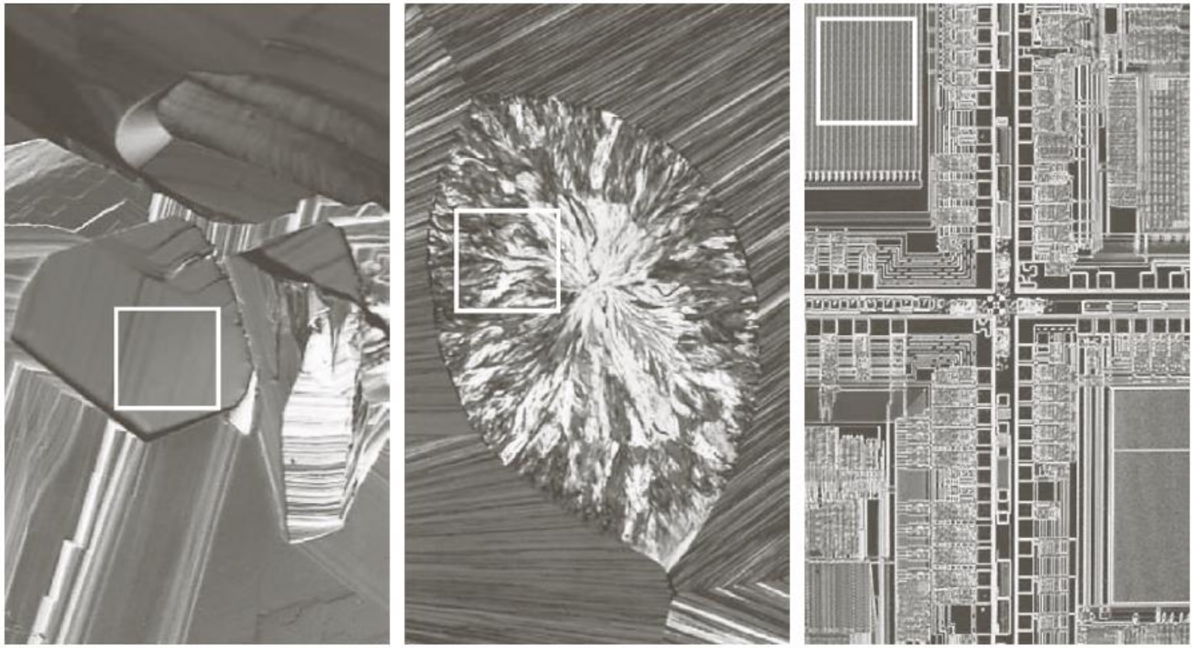
Regular

Statistical Approaches

- Statistical moments of the intensity histogram
  - Mean $$m = \sum_{i=0}^{L-1} z_i p(z_i)$$
  - *n*-th moment $$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i)$$
  - Measure of Smoothness $$R(z) = 1 - \frac{1}{1 + \sigma^2(z)}$$
  - Measure of the skewness

    asymmetry of the distribution $$\mu_3(z) = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$$
  - Measure of uniformity $$U(z) = \sum_{i=0}^{L-1} p^2(z_i)$$
  - Average entropy $$e(z) = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$$

Texture Descriptors



Smooth Coarse Regular

Texture	Mean	Standard deviation	R (normalized)	Third moment	Uniformity	Entropy
Smooth	82.64	11.79	0.002	−0.105	0.026	5.434
Coarse	143.56	74.63	0.079	−0.151	0.005	7.783
Regular	99.72	33.73	0.017	0.750	0.013	6.674

Texture Descriptors

共生矩阵（co-occurrence matrix）

- 在一幅图像中规定一种对于两个像素间的位置关系（包括方向和距离），设有这种关系的两个像素的灰度值分别为 g_i 和 g_j，
- 共生矩阵：利用这些像素对得到的二维直方图
  - 共生矩阵A的每个元素a_{ij}是具有灰度g_i和g_j的像素对在图像上出现的数目.
  - 例：g_j 是 g_i 的右下一个像素

$$
\begin{matrix}
0 & 0 & 0 & 1 & 2 \\
1 & 1 & 0 & 1 & 1 \\
2 & 2 & 1 & 0 & 0 \\
1 & 1 & 0 & 2 & 0 \\
0 & 0 & 1 & 0 & 1
\end{matrix}
$$

图像数据

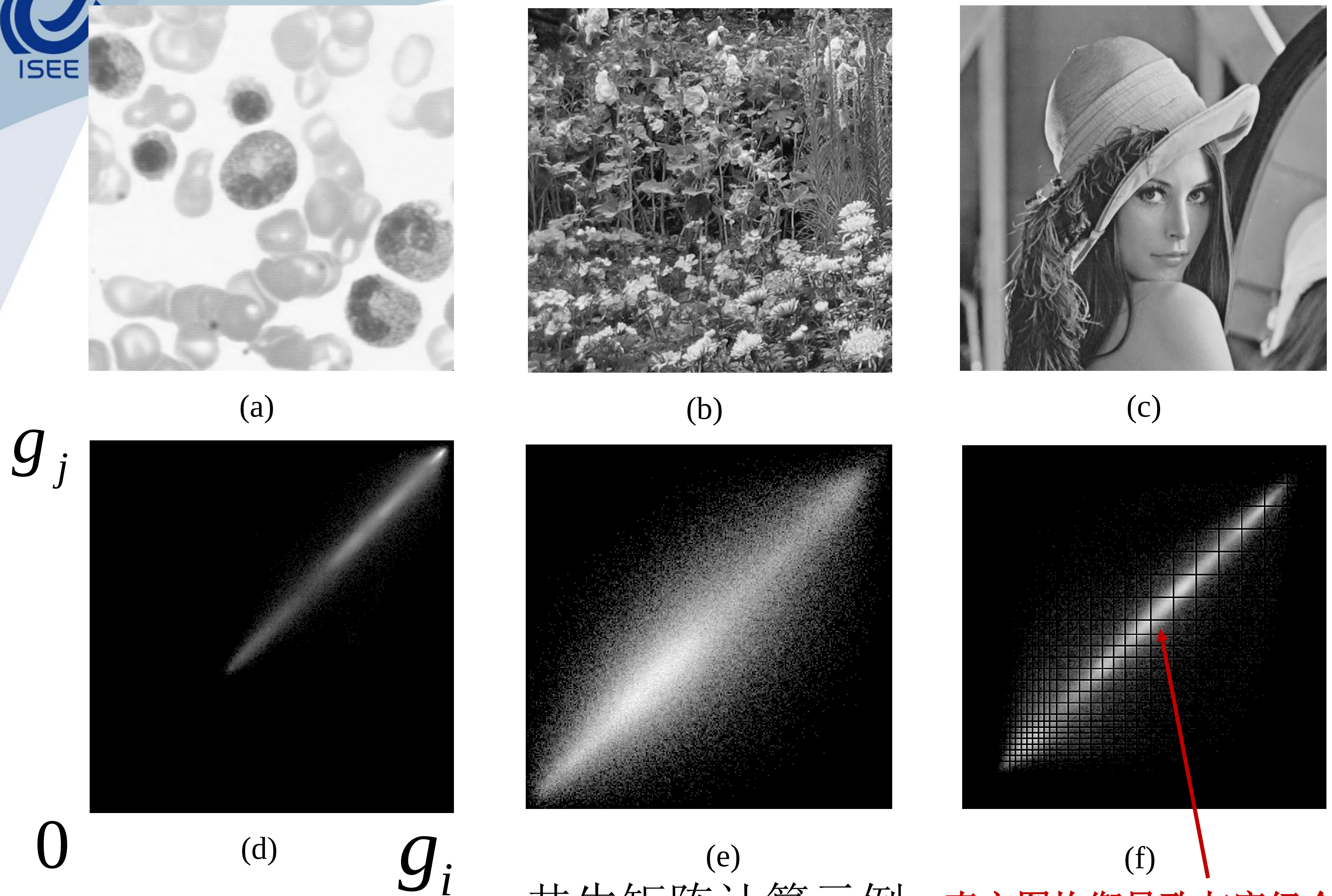
$$
A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{array}{c} \begin{matrix} g_i=0 & 1 & 2 \end{matrix} \\ \begin{bmatrix} 4 & 2 & 1 \\ 2 & 3 & 2 \\ 0 & 2 & 0 \end{bmatrix} \end{array}
$$

$\downarrow g_j = 0, 1, 2$

– 共生矩阵可定义为：

$$P(g_1, g_2) = \frac{\#\{[(x_1, y_1), (x_2, y_2)] \in S \mid f(x_1, y_1) = g_1 \ \& \ f(x_2, y_2) = g_2\}}{\# S}$$

- S为目标区域R中具有指定空间关系的像素对的集合
- # 表示数量
- 右边分子上表示具有指定关系，且灰度值分为g_i和g_j的像素偶对数
- 分母为具有指定关系的像素对的总数
- 这样得到的P是归一化值。

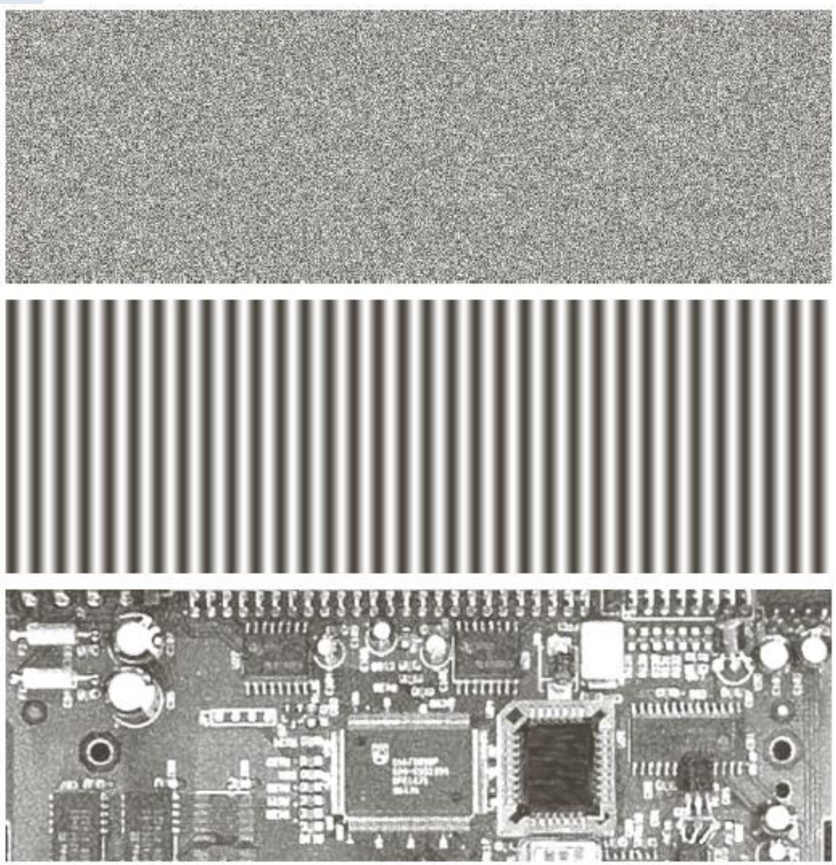


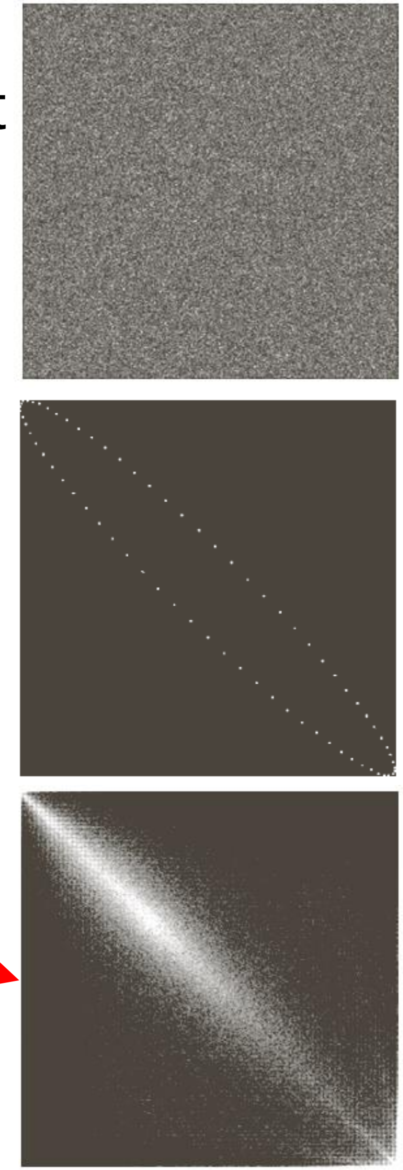
共生矩阵计算示例

直方图均衡导致灰度级合并

Co-occurrence matrix

Position operator:
one pixel immediately to the right





–利用共生矩阵定义的几个常用的纹理特征

1.最大概率

$$W_P = \max_{g_1,g_2} P(g_1, g_2)$$

–表示P的最强响应

2.一致性，或称纹理二阶矩 $W_M = \sum_{g_1}\sum_{g_2}(P(g_1, g_2))^2$

–当所有的 $P(g_1, g_2)$ 相等时，W_M有最小值，对应图像的纹理较丰富

3.熵 $W_E = -\sum_{g_1}\sum_{g_2} P(g_1, g_2)\log P(g_1, g_2)$

–当所有的 $P(g_1, g_2)$ 相等时，W_E有最大值。

4.差异度的k阶矩 $W_C = \sum_{g_1}\sum_{g_2}|g_1 - g_2|^k P(g_1, g_2)$

- 在图像平坦时，g_1和g_2相近，W_C较小

5.惯性（2阶矩） $W_I = \sum_{g_1}\sum_{g_2}(g_1 - g_2)^2 P(g_1, g_2)$

在图像平坦时，g_1和g_2相近，W_I较小

6.逆差异度的k阶矩 $W_R = \sum_{g_1}\sum_{g_2}|g_1 - g_2|^{-k} P(g_1, g_2)$

- W_R与W_C有相反的效果

7.均匀性 $W_H = \sum_{g_1}\sum_{g_2}\frac{P(g_1, g_2)}{k + |g_1 - g_2|}$

- 在图像平坦时，共生矩阵集中在g_1和g_2相近的对角线上 $k + |g_1 - g_2|$较小的概率较大，故W_H较大

TABLE 12.3
Descriptors used for characterizing co-occurrence matrices of size $K \times K$. The term p_{ij} is the ij-th term of $\mathbf{G}$ divided by the sum of the

$$m_r = \sum_{i=1}^{K} i \sum_{j=1}^{K} p_{ij}$$

$$m_c = \sum_{j=1}^{K} j \sum_{i=1}^{K} p_{ij}$$

$$\sigma_r^2 = \sum_{i=1}^{K} (i - m_r)^2 \sum_{j=1}^{K} p_{ij}$$

$$\sigma_c^2 = \sum_{j=1}^{K} (j - m_c)^2 \sum_{i=1}^{K} p_{ij}$$

Descriptor	Explanation	Formula
Maximum probability	Measures the strongest response of $\mathbf{G}$. The range of values is [0, 1].	$\max_{i,j}(p_{ij})$
Correlation	A measure of how correlated a pixel is to its neighbor over the entire image. The range of values is 1 to −1 corresponding to perfect positive and perfect negative correlations. This measure is not defined if either standard deviation is zero.	$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r \sigma_c}$ $\sigma_r \neq 0;\ \sigma_c \neq 0$
Contrast	A measure of intensity contrast between a pixel and its neighbor over the entire image. The range of values is 0 (when $\mathbf{G}$ is constant) to $(K-1)^2$.	$\sum_{i=1}^{K}\sum_{j=1}^{K}(i-j)^2 p_{ij}$
Uniformity (also called Energy)	A measure of uniformity in the range [0, 1]. Uniformity is 1 for a constant image.	$\sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij}^2$
Homogeneity	Measures the spatial closeness to the diagonal of the distribution of elements in $\mathbf{G}$. The range of values is [0, 1], with the maximum being achieved when $\mathbf{G}$ is a diagonal matrix.	$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{p_{ij}}{1+\lvert i-j\rvert}$
Entropy	Measures the randomness of the elements of $\mathbf{G}$. The entropy is 0 when all p_{ij}'s are 0, and is maximum when the p_{ij}'s are uniformly distributed. The maximum value is thus $2\log_2 K$.	$-\sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij}\log_2 p_{ij}$

TABLE 12.4
Descriptors evaluated using the co-occurrence matrices displayed as images in Fig. 12.32.

Normalized Co-occurrence Matrix	Maximum Probability	Correlation	Contrast	Uniformity	Homogeneity	Entropy
G_1/n_1	0.00006	–0.0005	10838	0.00002	0.0366	15.75
G_2/n_2	0.01500	0.9650	00570	0.01230	0.0824	06.43
G_3/n_3	0.06860	0.8798	01356	0.00480	0.2048	13.58

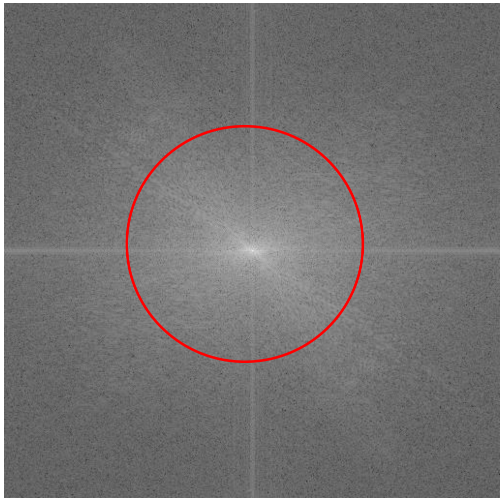
Spectral Approaches

- 图像上周期性和方向性的纹理在频谱上有相应的表现
  - 周期性纹理在频谱上有相应的能量集中区域
  - 纹理的方向在频谱上表现为能量集中区域所在的方向
- 将频谱表示成极坐标形式：**S(r,θ)**
  - 对于每个方向θ，频谱的分布可用一维函数$S_\theta(r)$表示
  - 对于每个指定r的圆环上，频谱的分布也可用一维函数$S_r(\theta)$表示
  - 在r半径上的总的能量为

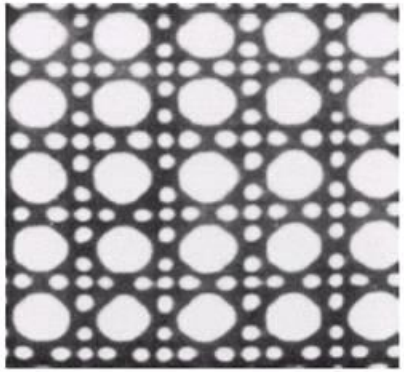
$$S(r)=\sum_{\theta=0}^{\pi} S_\theta(r)$$

  - 在θ方向的总能量为

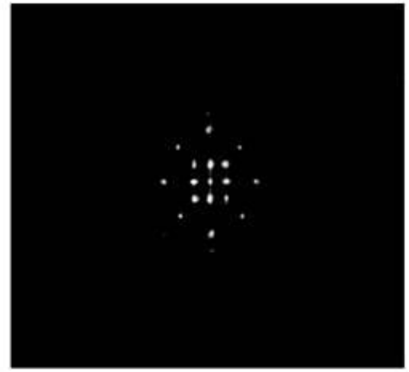
$$S(\theta)=\sum_{r=1}^{R_0} S_r(\theta)$$



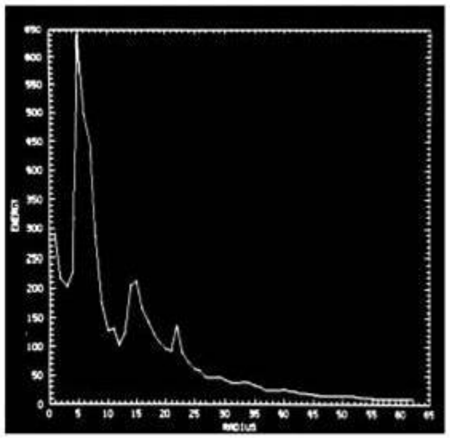
- 用S(r)和S(θ)可描述整幅图像或所选区域的纹理特性
- 也从描绘子S(r)和S(θ) 定量计算一些其它特征
  - 如最大值位置、均值、幅度的方差、均值与最大值间的距离等



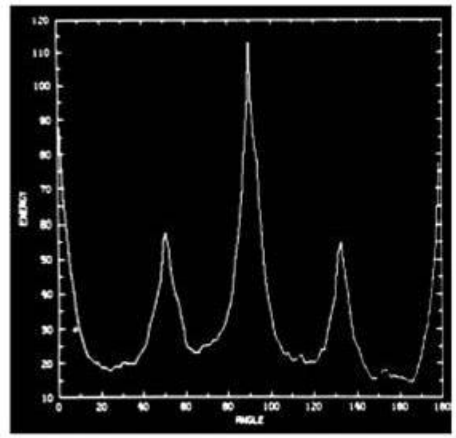
(a) 周期性纹理图



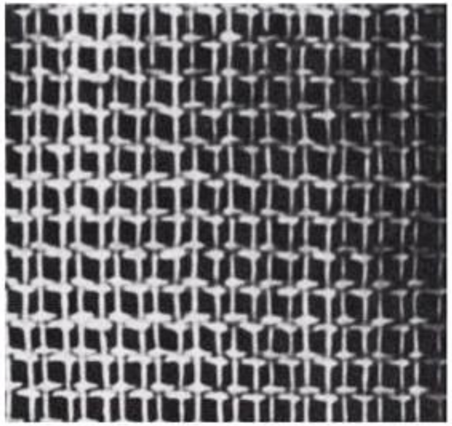
(b) 频谱图



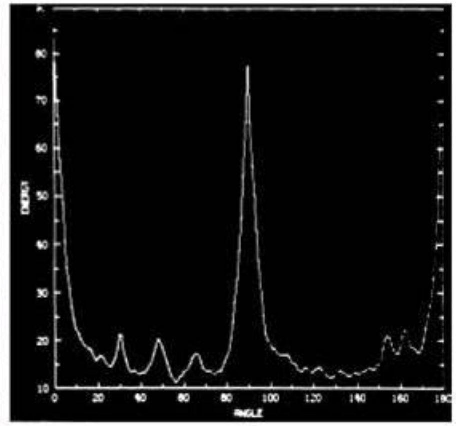
(c) S(r)



(d) S(θ)

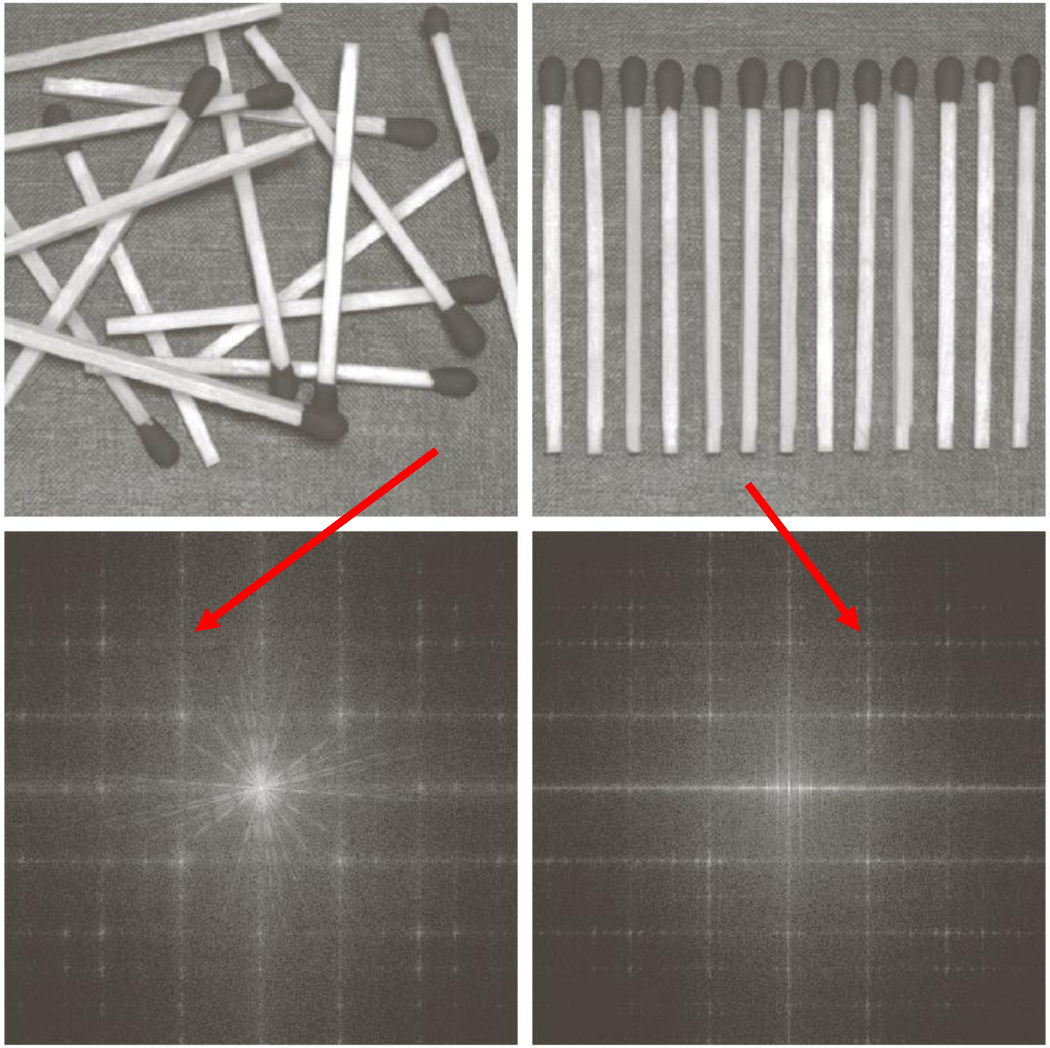


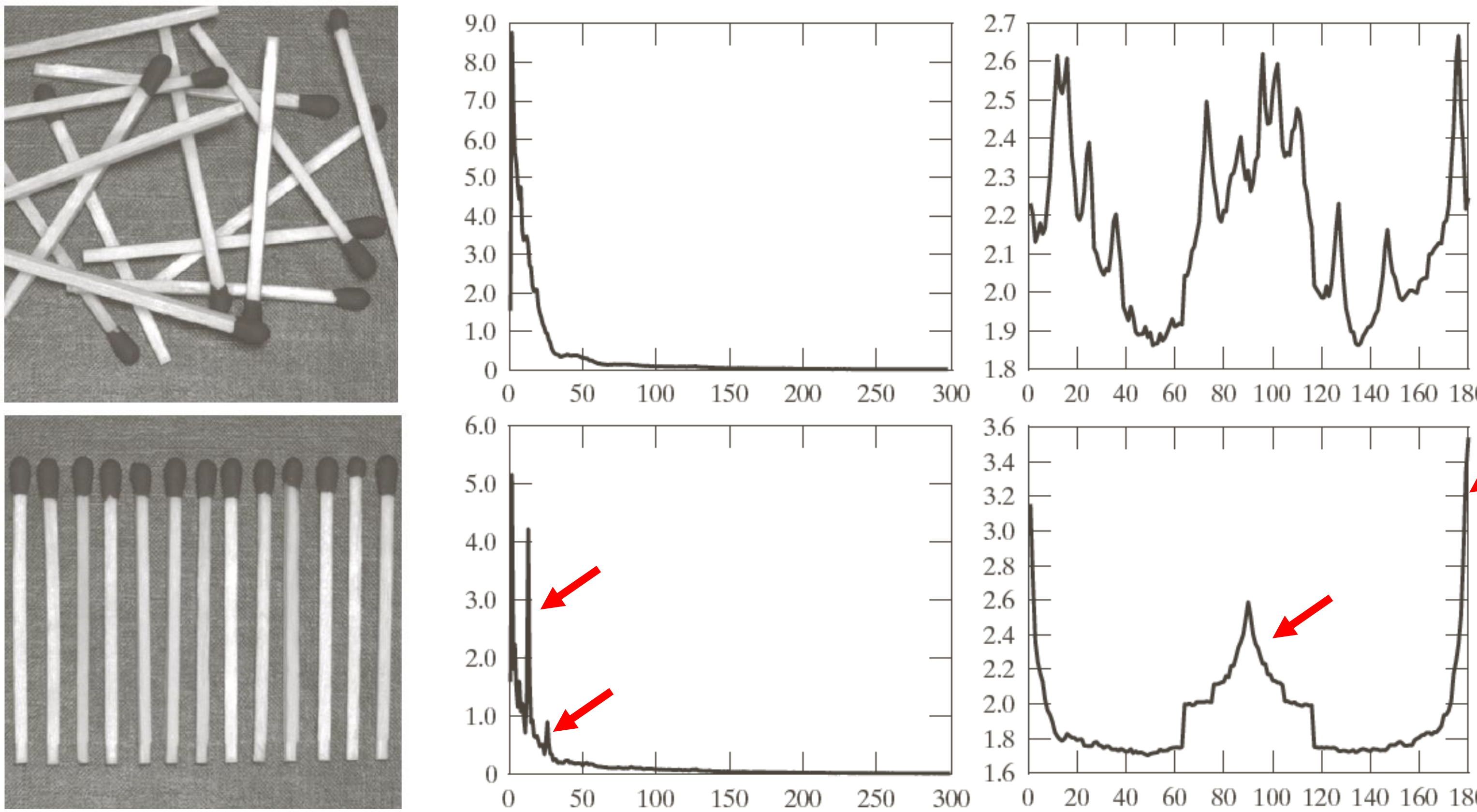
(e) 另一幅纹理图



(f) S(θ)

Fourier Spectra





9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0
0 50 100 150 200 250 300
2.7
2.6
2.5
2.4
2.3
2.2
2.1
2.0
1.9
1.8
0 20 40 60 80 100 120 140 160 180
6.0
5.0
4.0
3.0
2.0
1.0
0
0 50 100 150 200 250 300
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
0 20 40 60 80 100 120 140 160 180

$S(r)$ $S(\theta)$

Moment Invariants（不变矩）

- 一幅数字图像f(x,y)，它的p+q阶矩定义:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y)$$

–可以证明
  - m_{pq} 唯一地被f(x,y)所确定
  - 所有非负整数p和q得到的 m_{pq} 无限集也完全确定了f(x,y)

- 图像f(x,y)的p+q阶中心矩定义

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y)$$

–其中 $\bar{x} = m_{10} / m_{00}$ $\bar{y} = m_{01} / m_{00}$
是图像的重心坐标

– 前三阶中心矩可用下式计算

$$\mu_{00} = m_{00} \qquad \mu_{02} = m_{02} - \bar{y} m_{01}$$

$$\mu_{10} = 0 \qquad \mu_{30} = m_{30} - 3\bar{x} m_{20} + 2\bar{x}^2 m_{10}$$

$$\mu_{01} = 0 \qquad \mu_{03} = m_{03} - 3\bar{y} m_{02} + 2\bar{y}^2 m_{01}$$

$$\mu_{11} = m_{11} - \bar{y} m_{10} \qquad \mu_{21} = m_{21} - 2\bar{x} m_{11} - \bar{y} m_{20} + 2\bar{x}^2 m_{01}$$

$$\mu_{20} = m_{20} - \bar{x} m_{10} \qquad \mu_{12} = m_{12} - 2\bar{y} m_{11} - \bar{x} m_{02} + 2\bar{y}^2 m_{10}$$

– f(x,y)的归一化中心矩 $\eta_{pq}=\dfrac{\mu_{pq}}{\mu_{00}^{r}}$

其中 $r=\dfrac{p+q}{2}+1$ p+q=2,3,…

– 使二阶中心矩 μ_{11} 变得最小的旋转角θ可由下式得出

$$\tan 2\theta=\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}$$

- 若x,y轴旋转θ角得坐标轴x',y'，称为该目标的主轴
  - 如果目标在计算矩前旋转θ角，或相对于x',y'轴计算矩，则矩具有旋转不变性

- 由归一化的二阶和三阶中心矩可得到如下7个对连续**图像平移、旋转和尺度变换不变的矩**：

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

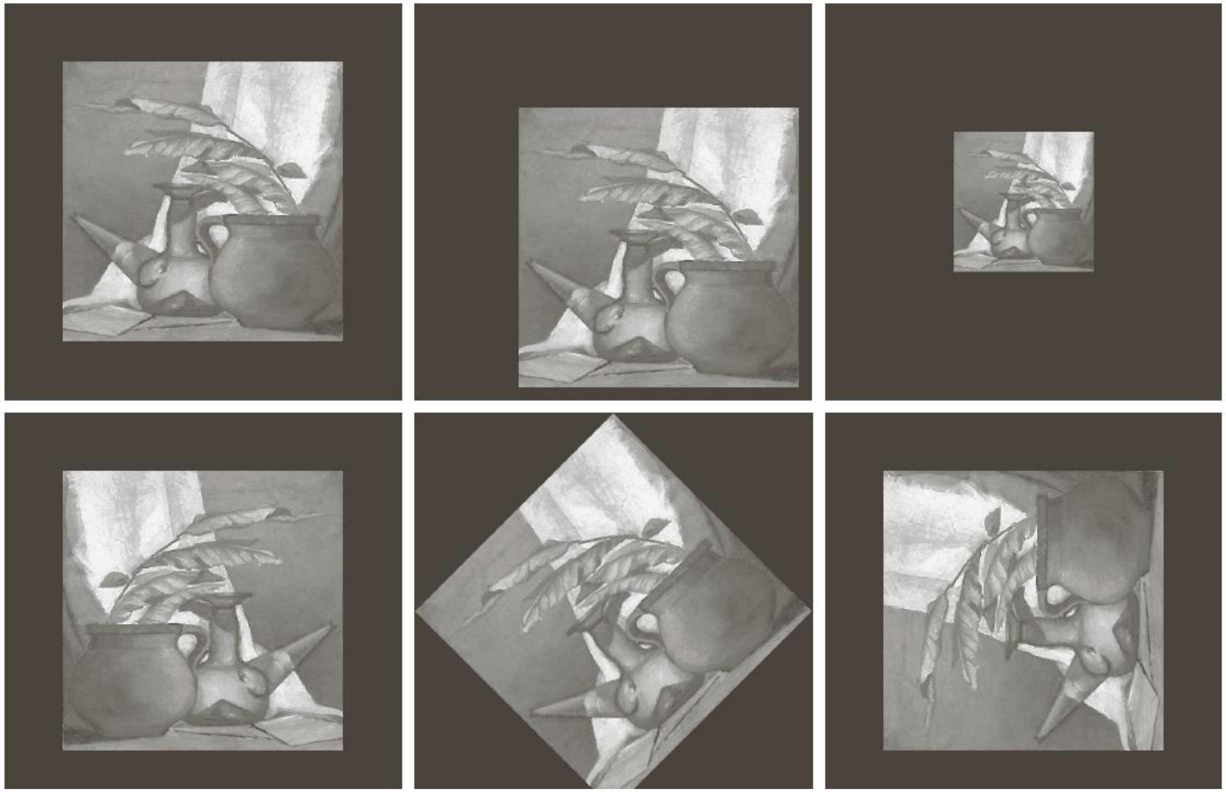
$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\begin{aligned}\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned}$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\begin{aligned}\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})\end{aligned}$$

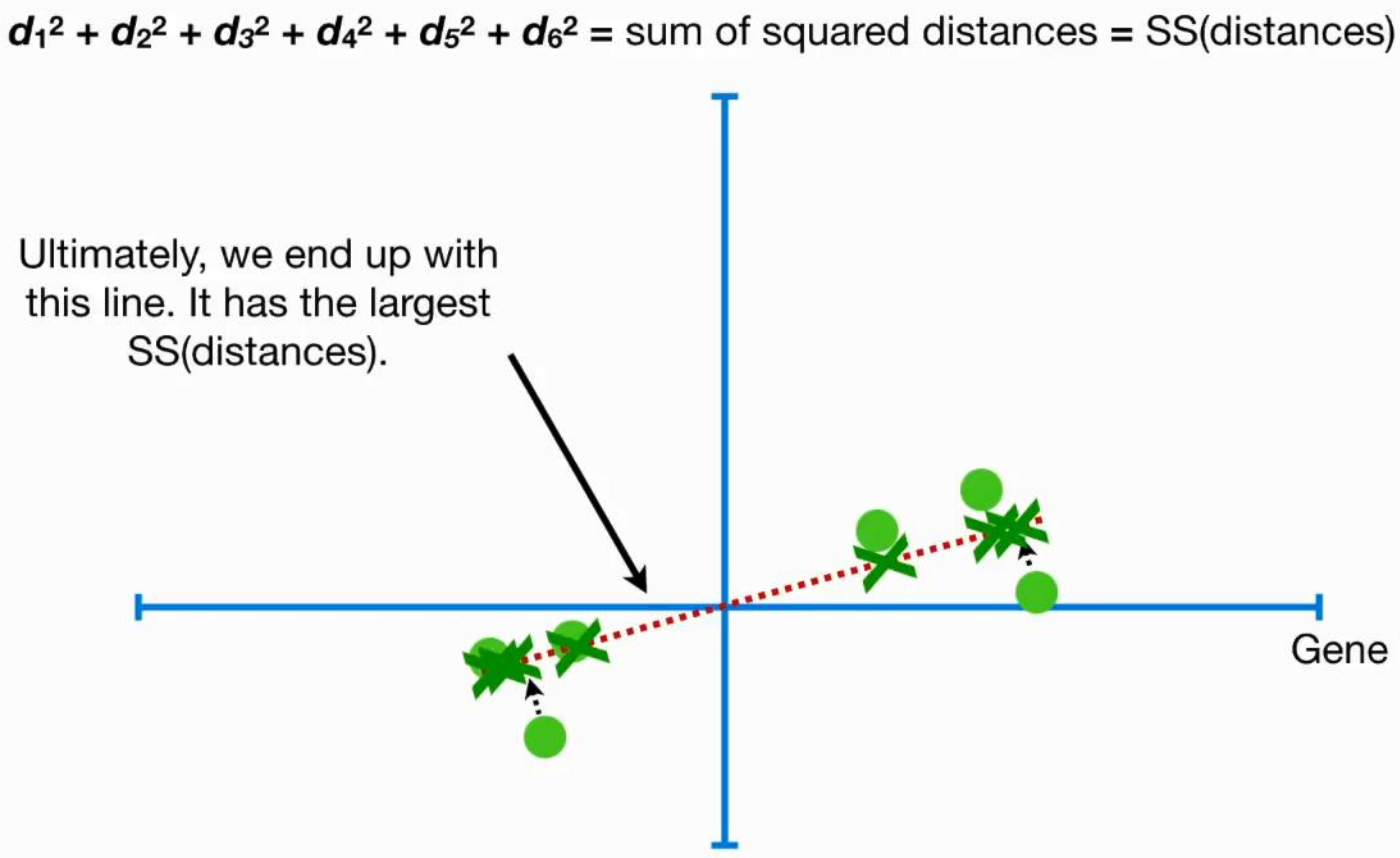
由于 $\phi_7^2 + \phi_5^2 = \phi_3\phi_4^3$ ，所以上述7个不变矩只用6个就可



Moment Invariant	Original Image	Translated	Half Size	Mirrored	Rotated 45°	Rotated 90°
ϕ_1	2.8662	2.8662	2.8664	2.8662	2.8661	2.8662
ϕ_2	7.1265	7.1265	7.1257	7.1265	7.1266	7.1265
ϕ_3	10.4109	10.4109	10.4047	10.4109	10.4115	10.4109
ϕ_4	10.3742	10.3742	10.3719	10.3742	10.3742	10.3742
ϕ_5	21.3674	21.3674	21.3924	21.3674	21.3663	21.3674
ϕ_6	13.9417	13.9417	13.9383	13.9417	13.9417	13.9417
ϕ_7	−20.7809	−20.7809	−20.7724	20.7809	−20.7813	−20.7809

可用于判断是否镜像

Principal Component Analysis (PCA)



Principal Component Analysis (PCA)

- A population of n-dimensional vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Mean vector

$$\mathbf{m}_\mathbf{x} = E\{\mathbf{x}\}$$

- Covariance matrix

$$\mathbf{C}_\mathbf{x} = E\left\{(\mathbf{x} - \mathbf{m}_\mathbf{x})(\mathbf{x} - \mathbf{m}_\mathbf{x})^T\right\}$$

PCA (cont…)

- Eigenvectors and eigenvalues

$$\mathbf{C}\mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad \lambda_j \geq \lambda_{j+1} \text{ for } j = 1, 2, \ldots, n-1$$

- $A = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_n^T \end{bmatrix}, \quad A^{-1} = A^T = [e_1^T \quad e_2^T \quad \ldots \quad e_n^T]$

- Hotelling transform

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_\mathbf{x})$$

Hotelling Transform

- Properties

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m_x})$$

$$\mathbf{m_y} = E\{\mathbf{y}\} = \mathbf{0}$$

$$\mathbf{C_y} = \mathbf{A}\mathbf{C_x}\mathbf{A}^T \quad = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

$$\mathbf{A}^{-1} = \mathbf{A}^T$$

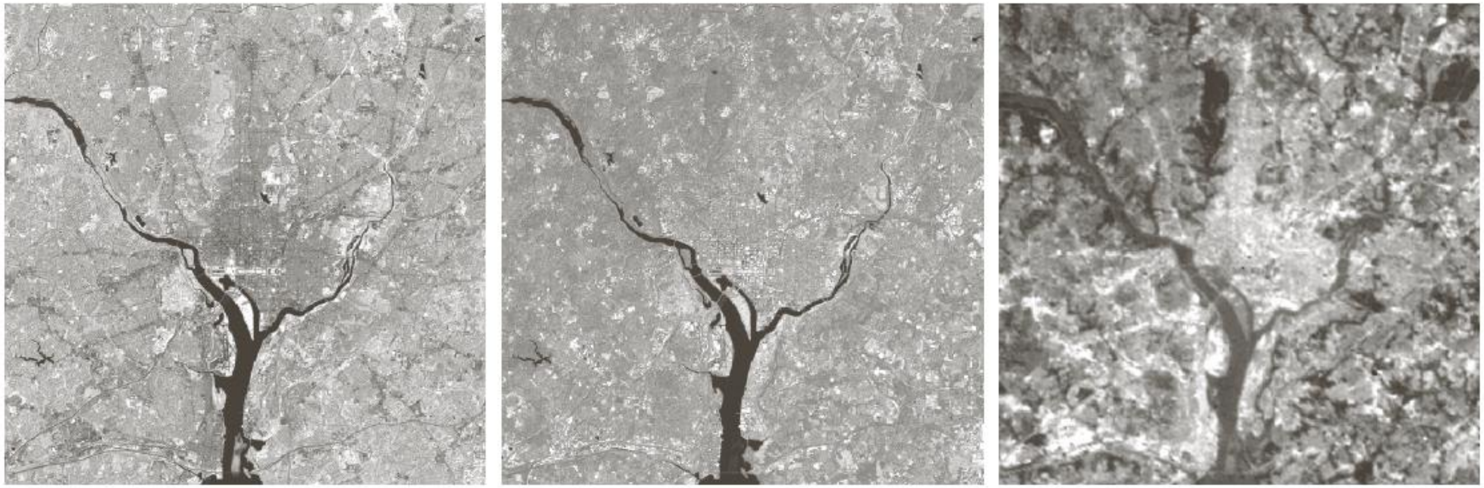
- Full reconstruction $\mathbf{x} = \mathbf{A}^T\mathbf{y} + \mathbf{m_x}$
- Partial reconstruction

$$\hat{\mathbf{x}} = \mathbf{A}_k^T\mathbf{y} + \mathbf{m_x}$$

$$e_{\mathrm{ms}} = \sum_{j=1}^{n}\lambda_j - \sum_{j=1}^{k}\lambda_j = \sum_{j=k+1}^{n}\lambda_j$$

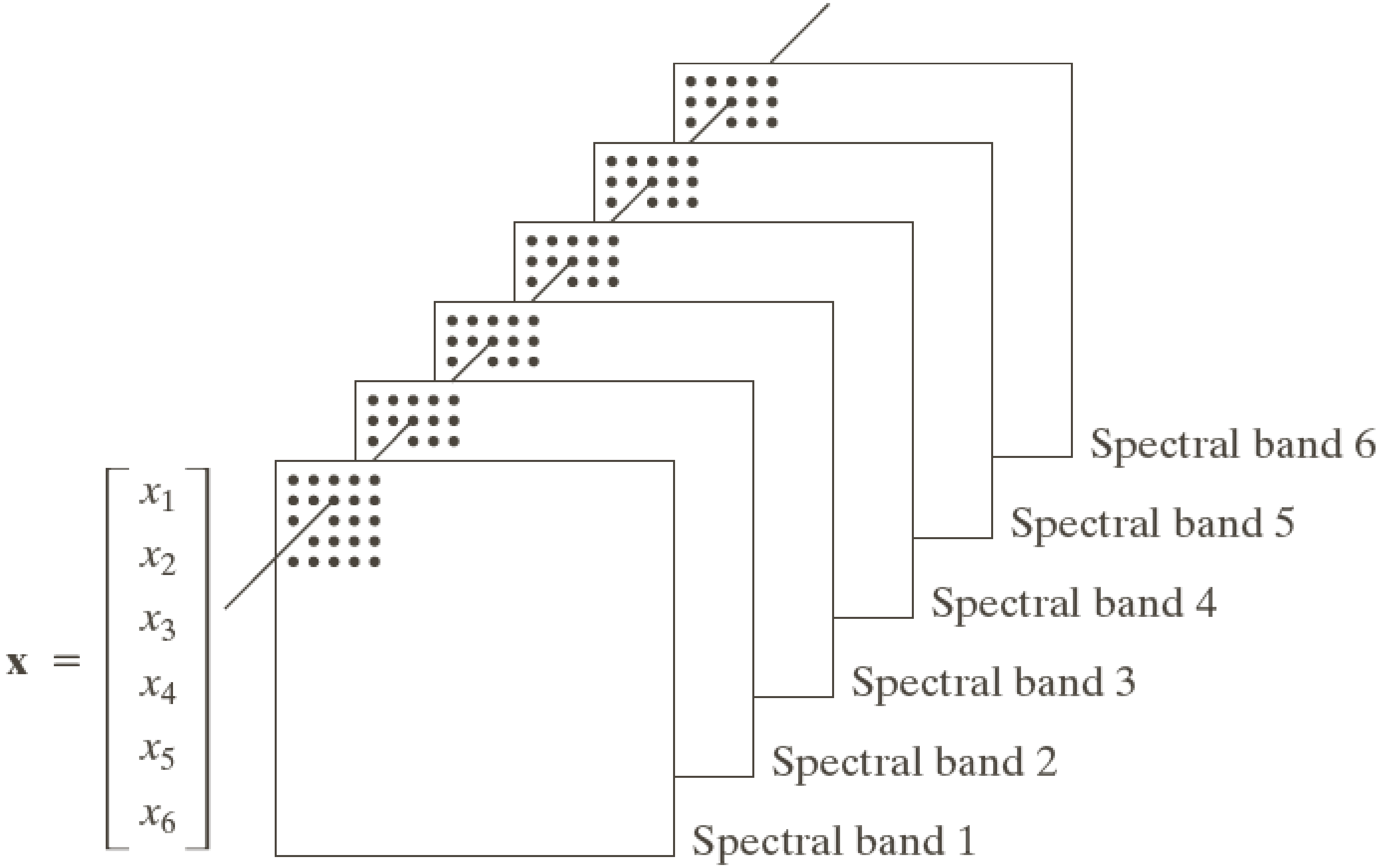
PCA Example

Visible blue Visible green Visible red



Near infrared Middle infrared Thermal infrared

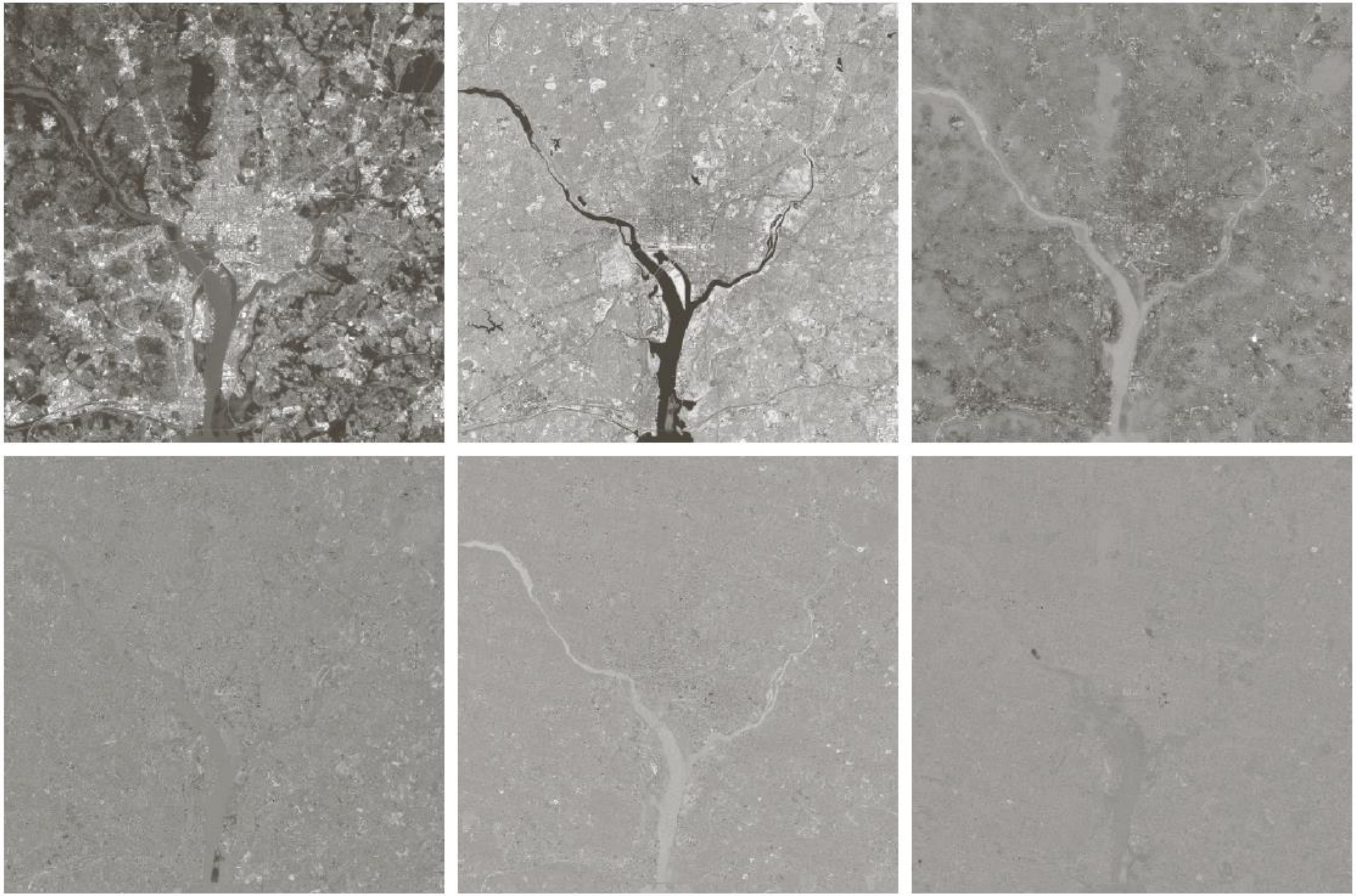
PCA Example



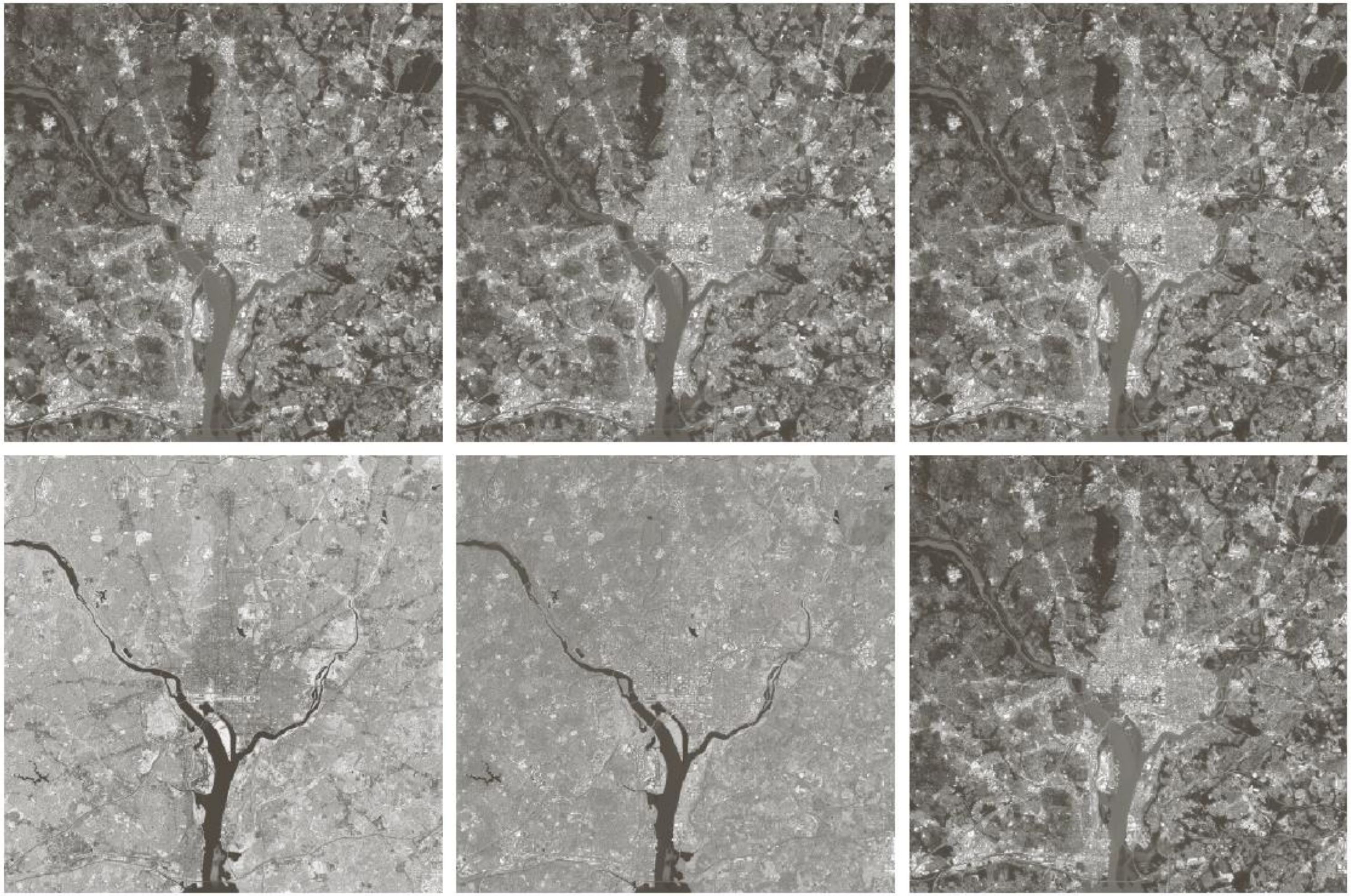
λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
10344	2966	1401	203	94	31

89% of the total variance

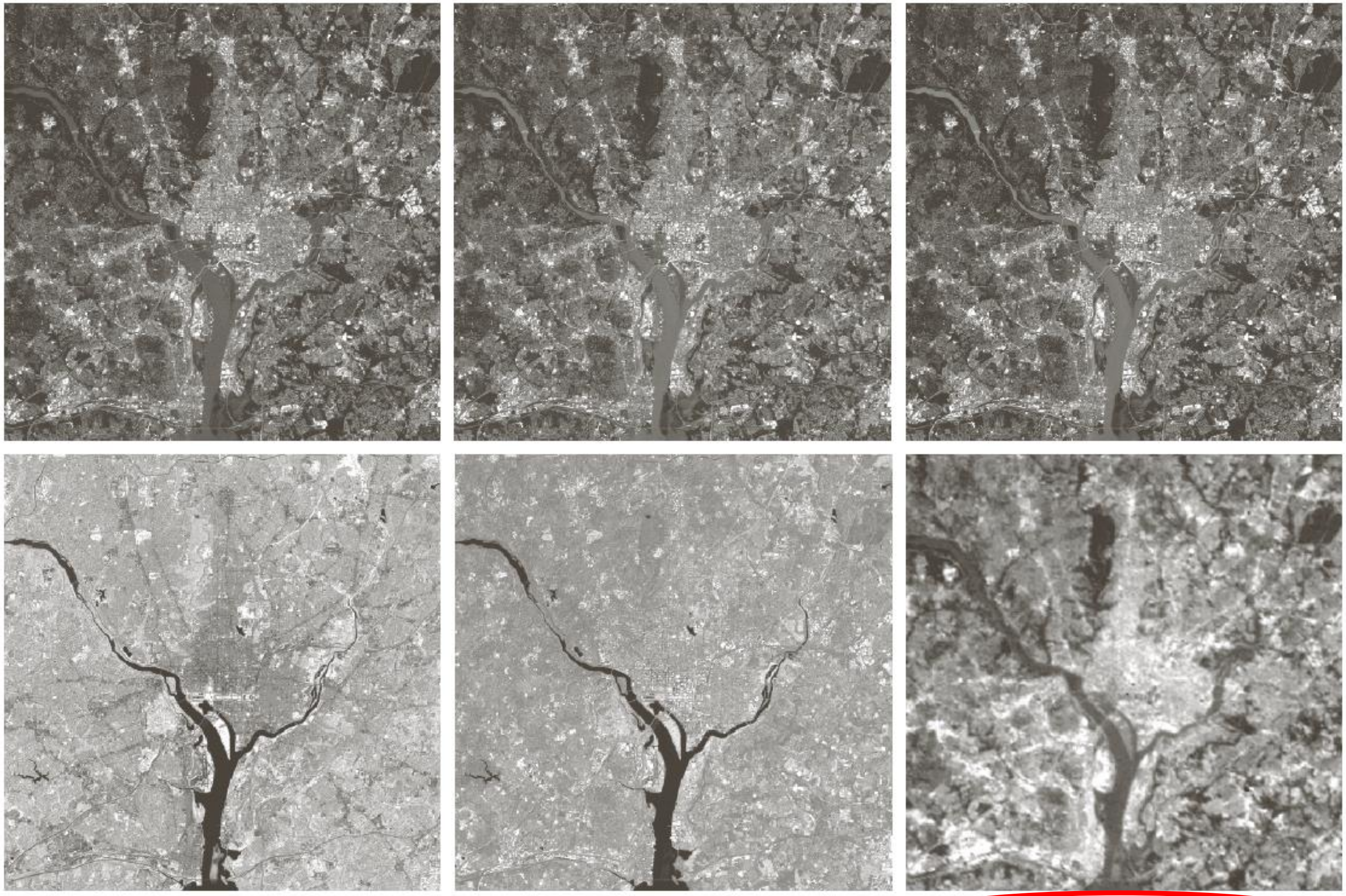
$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_x)$$



$$\hat{\mathbf{x}} = \mathbf{A}_k^T\mathbf{y} + \mathbf{m}_\mathbf{x} \qquad k = 2$$

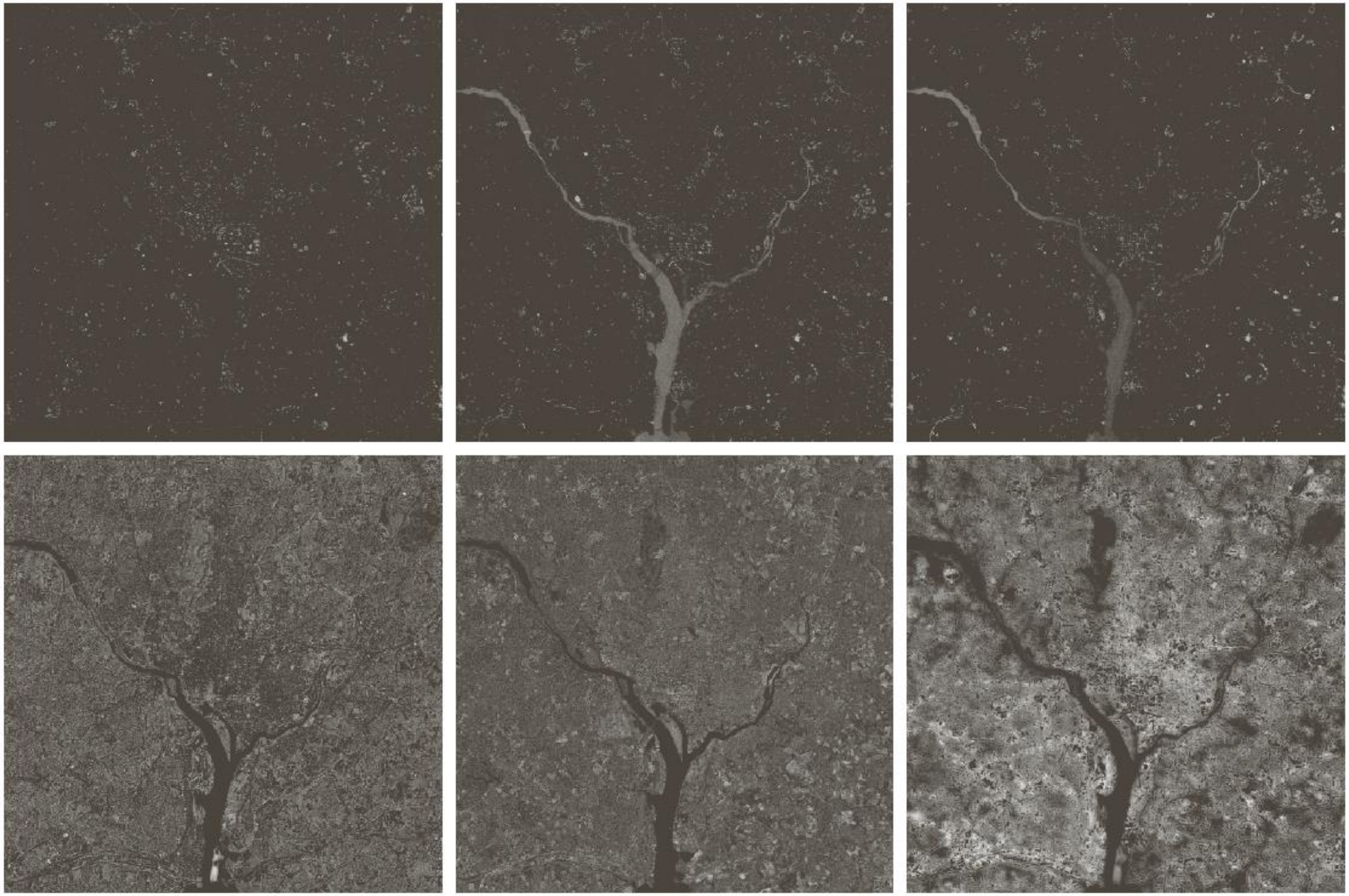


Original



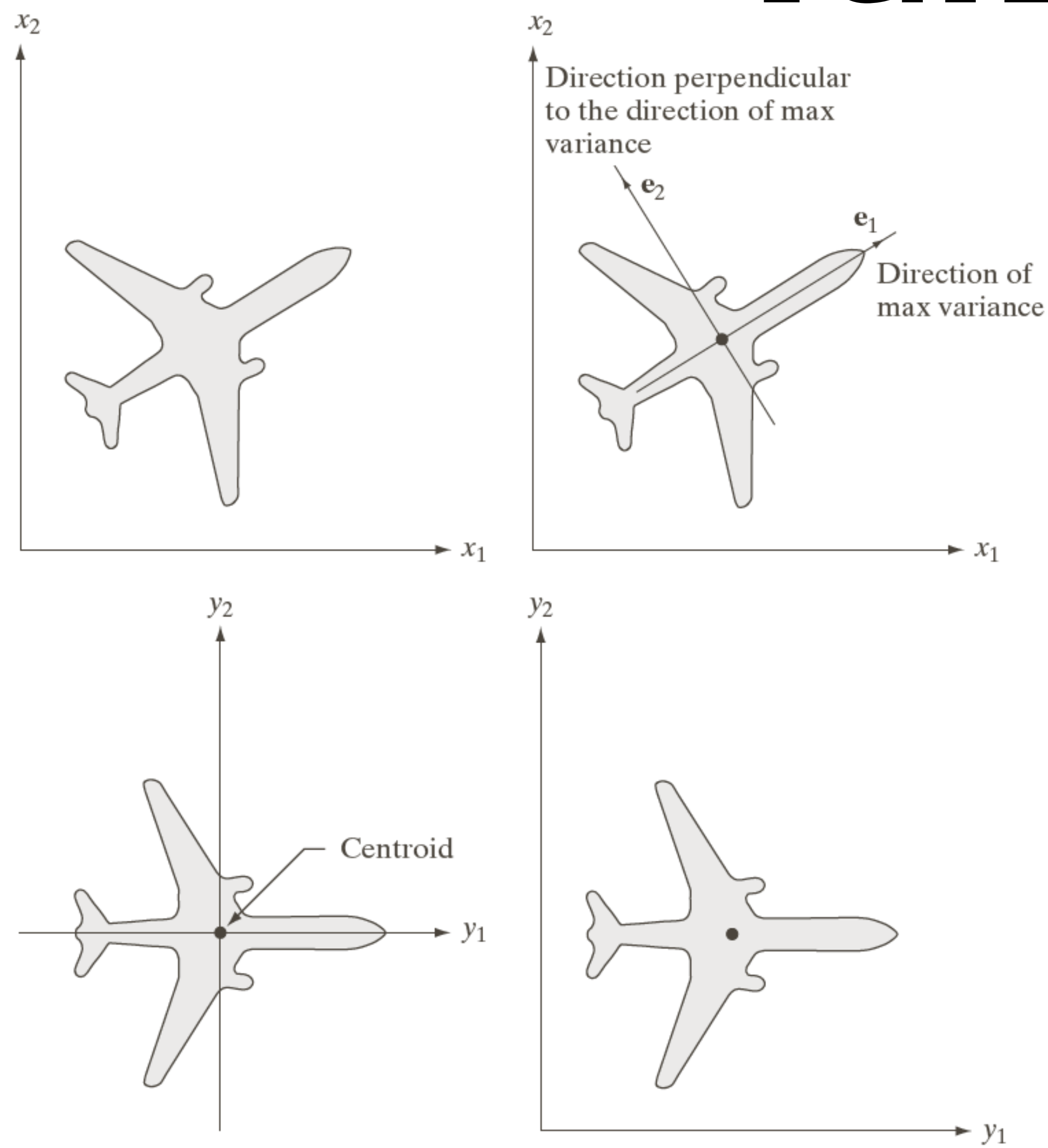
Thermal infrared

Difference Images



All images were enhanced by scaling them to the full [0,255]

PCA Example 2



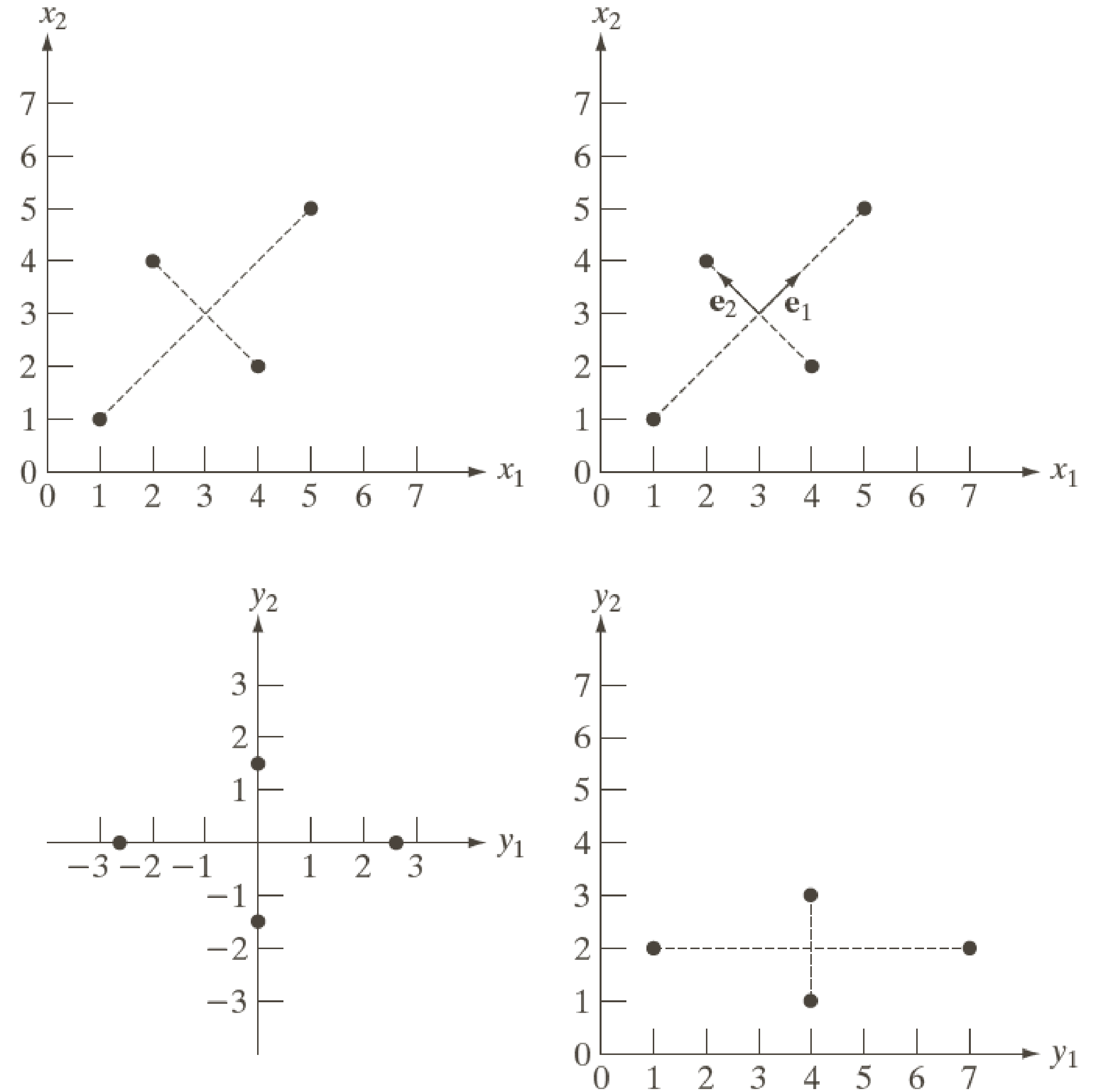
PCA Example 3

$$\mathbf{m_x} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\mathbf{C_x} = \begin{bmatrix} 3.333 & 2.00 \\ 2.00 & 3.333 \end{bmatrix}$$

$$\mathbf{e}_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$$

$$\mathbf{e}_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$$



Scale-Invariant Feature Transform (SIFT)

Scale space

$$L(x,y,\sigma) = G(x,y,\sigma) \star f(x,y)$$

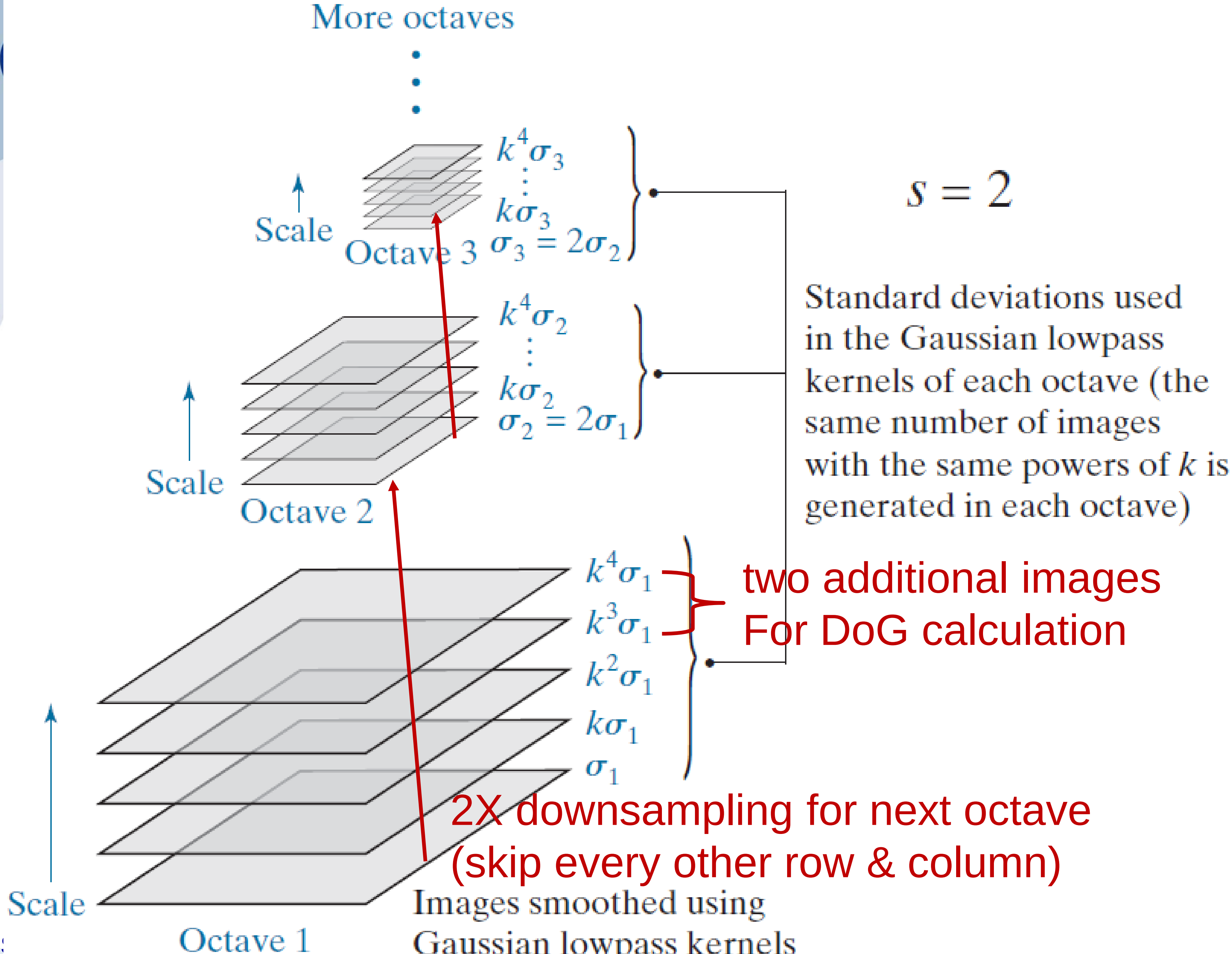
Generate a stack of smoothed images

Gaussian kernel

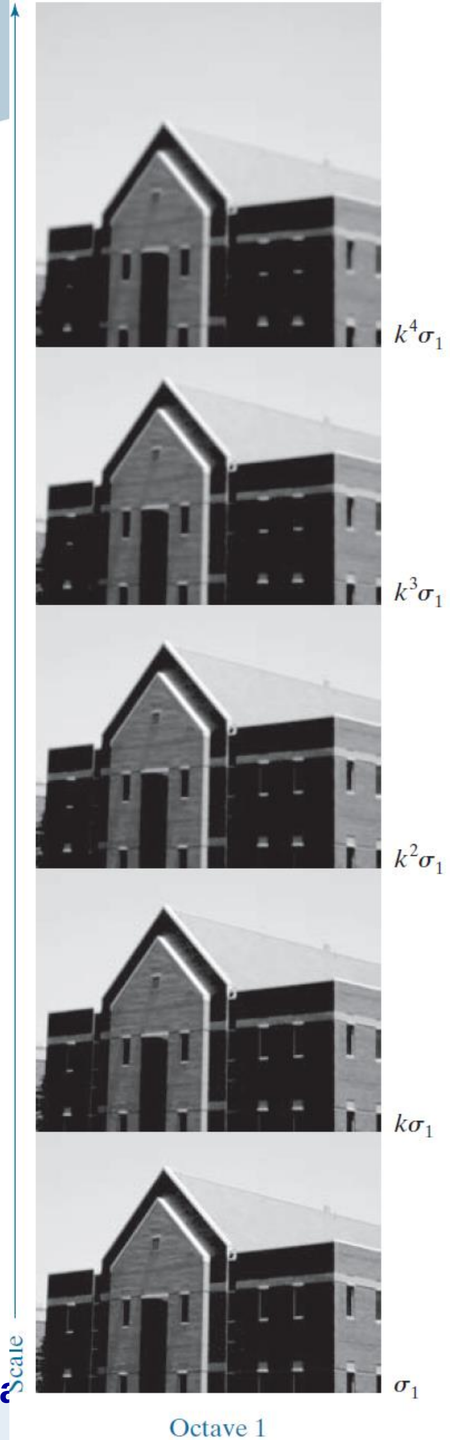
$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

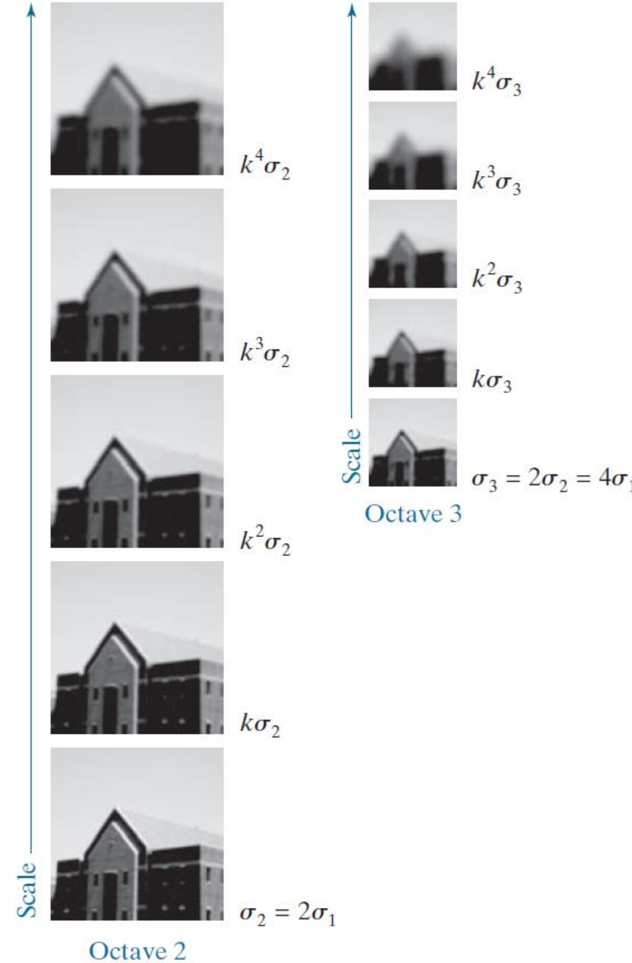
$$\sigma, k\sigma, k^2\sigma, k^3\sigma, \ldots$$

Octaves: $k^s\sigma = 2\sigma \Longrightarrow s = 2,\ k = \sqrt{2}$



More octaves
$k^4\sigma_3$
$k\sigma_3$
$\sigma_3 = 2\sigma_2$
Scale
Octave 3
$s = 2$
$k^4\sigma_2$
$k\sigma_2$
$\sigma_2 = 2\sigma_1$
Scale
Octave 2
Standard deviations used in the Gaussian lowpass kernels of each octave (the same number of images with the same powers of k is generated in each octave)
$k^4\sigma_1$
$k^3\sigma_1$
two additional images
For DoG calculation
$k^2\sigma_1$
$k\sigma_1$
σ_1
2X downsampling for next octave
(skip every other row & column)
Scale
Octave 1
Images smoothed using Gaussian lowpass kernels





$\sigma_1 = \sqrt{2}/2 = 0.707 \quad k = \sqrt{2} = 1.414$

Octave	Scale 1	2	3	4	5
1	0.707	1.000	1.414	2.000	2.828
2	1.414	2.000	2.828	4.000	5.657
3	2.828	4.000	5.657	8.000	11.314

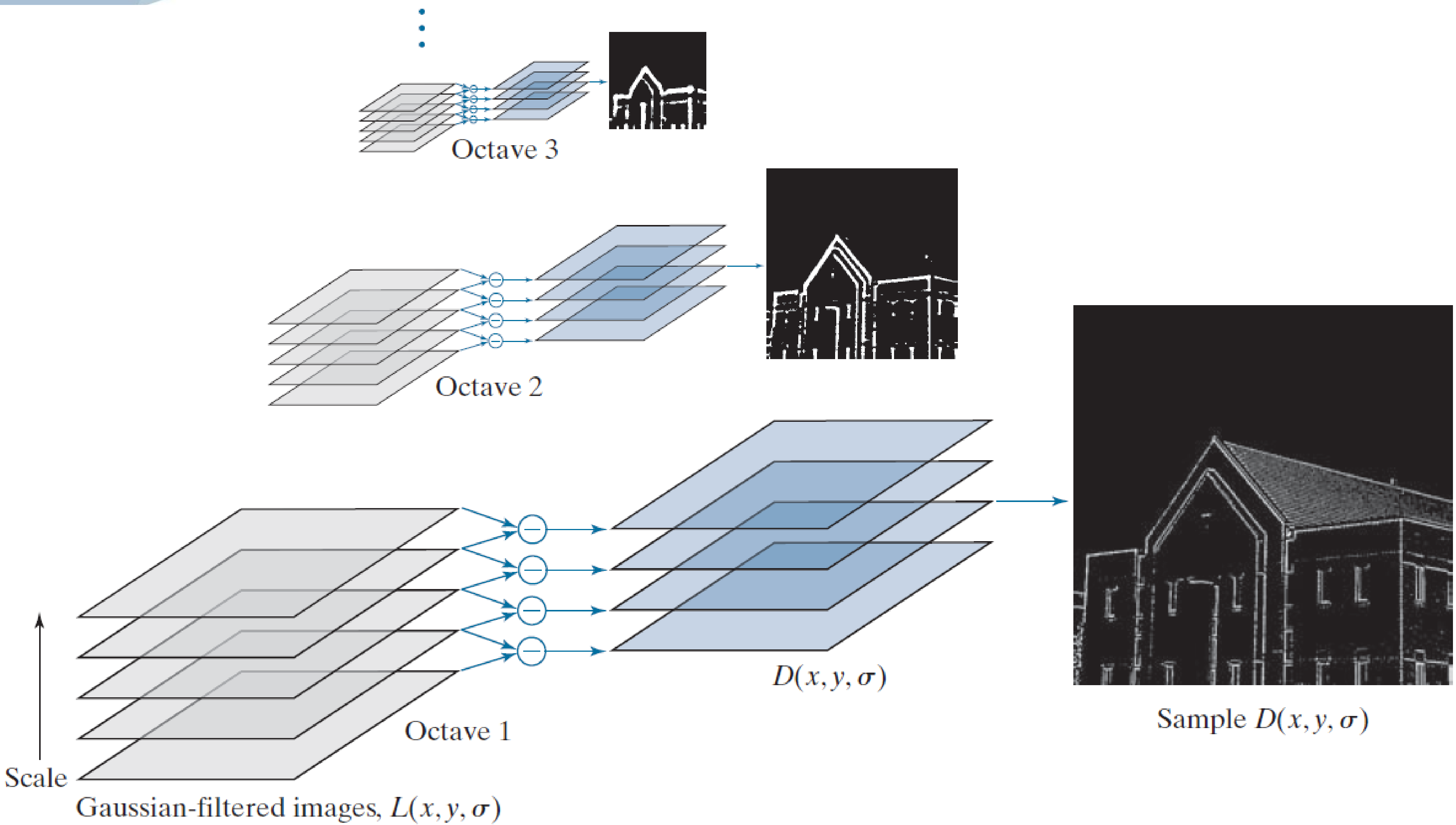
Find the Initial Keypoints

- Detect extrema in the difference of Gaussians of two adjacent scale-space images in an octave

$$
\begin{aligned}
D(x,y,\sigma) &= \left[G(x,y,k\sigma) - G(x,y,\sigma)\right] \star f(x,y) \\
&= L(x,y,k\sigma) - L(x,y,\sigma) \\
&\approx (k-1)\sigma^2\nabla^2 G
\end{aligned}
$$

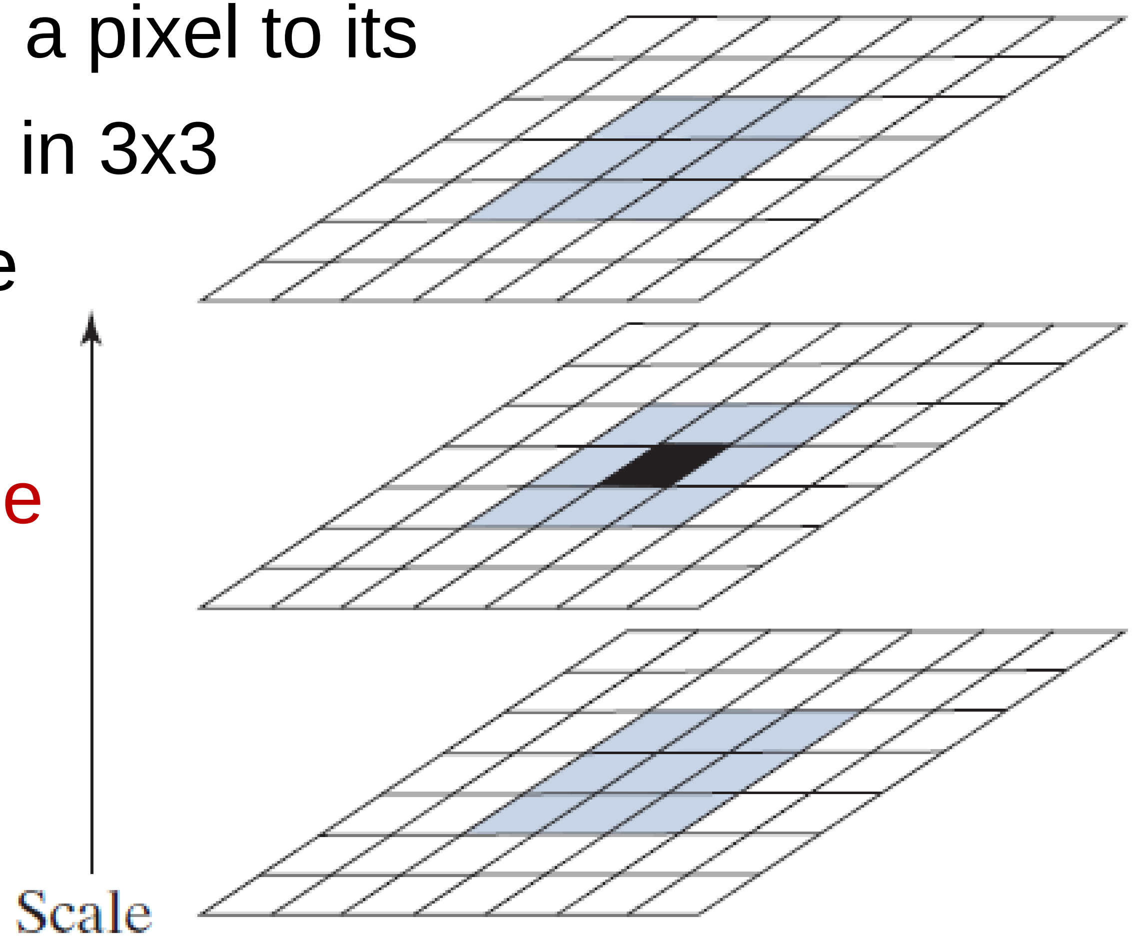
DoG: approximation to LoG

$s + 2$ difference functions



Detect local extrema (maxima or minima)

- Comparing a pixel to its 26 neighbors in 3x3 regions at the current and adjacent scale images



Corresponding sections of three contiguous $D(x, y, \sigma)$ images

Achieve subpixel accuracy

- Taylor series expansion of $D(x, y, \sigma)$

$$D(\mathbf{x}) = D + \left(\frac{\partial D}{\partial \mathbf{x}}\right)^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial}{\partial \mathbf{x}}\left(\frac{\partial D}{\partial \mathbf{x}}\right)\mathbf{x}$$

$$= D + (\nabla D)^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \mathbf{H}\,\mathbf{x}$$

Gradient operator

offset

$$\nabla D = \frac{\partial D}{\partial \mathbf{x}} = \begin{bmatrix} \partial D/\partial x \\ \partial D/\partial y \\ \partial D/\partial \sigma \end{bmatrix}$$

$$\mathbf{x} = (x, y, \sigma)^T$$

Hessian matrix

$$\mathbf{H} = \begin{bmatrix} \partial^2 D/\partial x^2 & \partial^2 D/\partial x \partial y & \partial^2 D/\partial x \partial \sigma \\ \partial^2 D/\partial y \partial x & \partial^2 D/\partial y^2 & \partial^2 D/\partial y \partial \sigma \\ \partial^2 D/\partial \sigma \partial x & \partial^2 D/\partial \sigma \partial y & \partial^2 D/\partial \sigma^2 \end{bmatrix}$$

Achieve subpixel accuracy

- Location of the extremum

$$\hat{\mathbf{x}} = -\mathbf{H}^{-1}\left(\nabla D\right)$$

If the offset is greater than 0.5 in any of its three dimensions, move to the closer integer point and redo interpolation

- Extremum

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2}\left(\nabla D\right)^{T}\hat{\mathbf{x}}$$

Eliminating Edge Response

- Quantify difference between edges and corners

Eigenvalues of Hessian matrix are proportional to the local curvature of *D*

r < Threshold

$$\mathbf{H} = \begin{bmatrix} \partial^2 D/\partial x^2 & \partial^2 D/\partial x \partial y \\ \partial^2 D/\partial y \partial x & \partial^2 D/\partial y^2 \end{bmatrix} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$$

Largest Smallest Eigenvalue

$$\mathrm{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta$$

$$\alpha = r\beta$$

$$\mathrm{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta$$

$$\frac{[\mathrm{Tr}(\mathbf{H})]^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r}$$

Eliminating Edge Response

$$\frac{[\mathrm{Tr}(\mathbf{H})]^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r}$$

Increases with $r \geq 1$, with minimum at $r = 1$

$r < 0$? Discard this point

- Keep "corner-like" point if

$$\frac{[\mathrm{Tr}(\mathbf{H})]^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r + 1)^2}{r}$$

e.g. $r = 10$

Example of SIFT keypoints

Keypoint orientation

- Gradient magnitude

$$M(x,y)=\left[\left(L(x+1,y)-L(x-1,y)\right)^2+\left(L(x,y+1)-L(x,y-1)\right)^2\right]^{\frac{1}{2}}$$

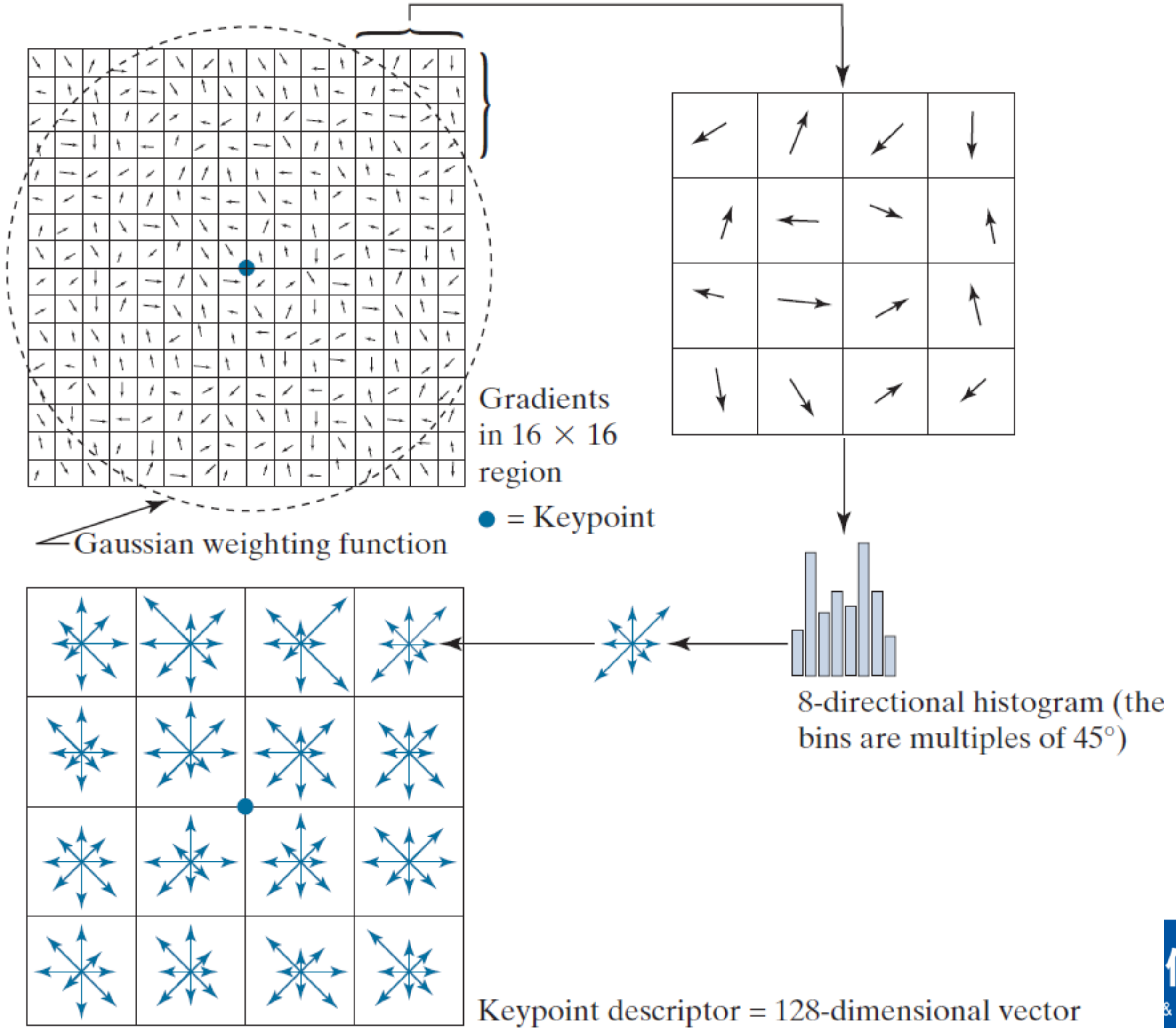
- Orientation angle

$$\theta(x,y)=\tan^{-1}\left[\left(L(x,y+1)-L(x,y-1)\right)/\left(L(x+1,y)-L(x-1,y)\right)\right]$$

- Histogram of orientations
  - Neighborhood of each keypoint
  - Weighted by its gradient magnitude
  - By a circular Gaussian function with 1.5 σ
  - 360 degrees → 36 bins
- Highest peak and ≥ 80% in the histogram
- Parabola fit to interpolate the peak position

Keypoint descriptors

- Keypoint: location, scale, orientation
- Descriptor: local region around each keypoint



Summary of SIFT algorithm

$\sigma = 1.6,\ s = 2,$
three octaves

1. Construct the scale space
2. Obtain the initial keypoints
3. Improve the location of keypoints
4. Delete unsuitable keypoints
   – Low value of *D*
   – Edge
5. Compute keypoint orientations
6. Compute keypoint descriptors

   128-dimensional feature vector

Image matching using SIFT

54 keypoints

643 keypoints

Matched: 36 keypoints
3 are incorrect

Image matching using SIFT

49 keypoints

Rotated by 5 degrees
547 keypoints

Matched: 26 keypoints
2 are incorrect

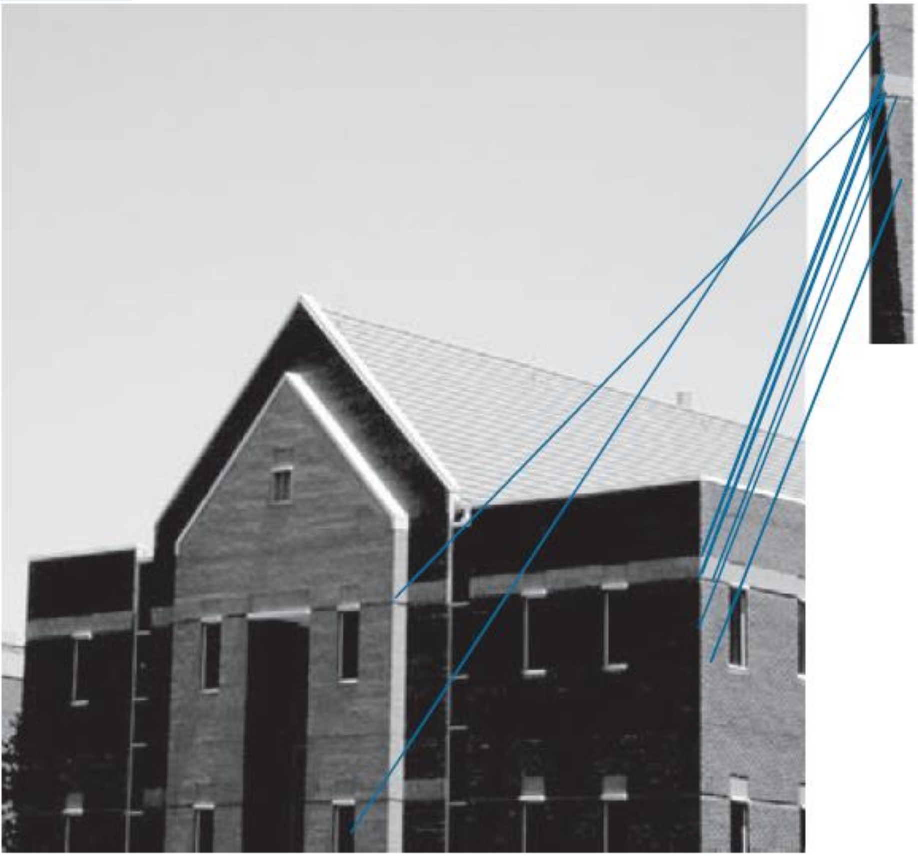
Image matching using SIFT

24 keypoints

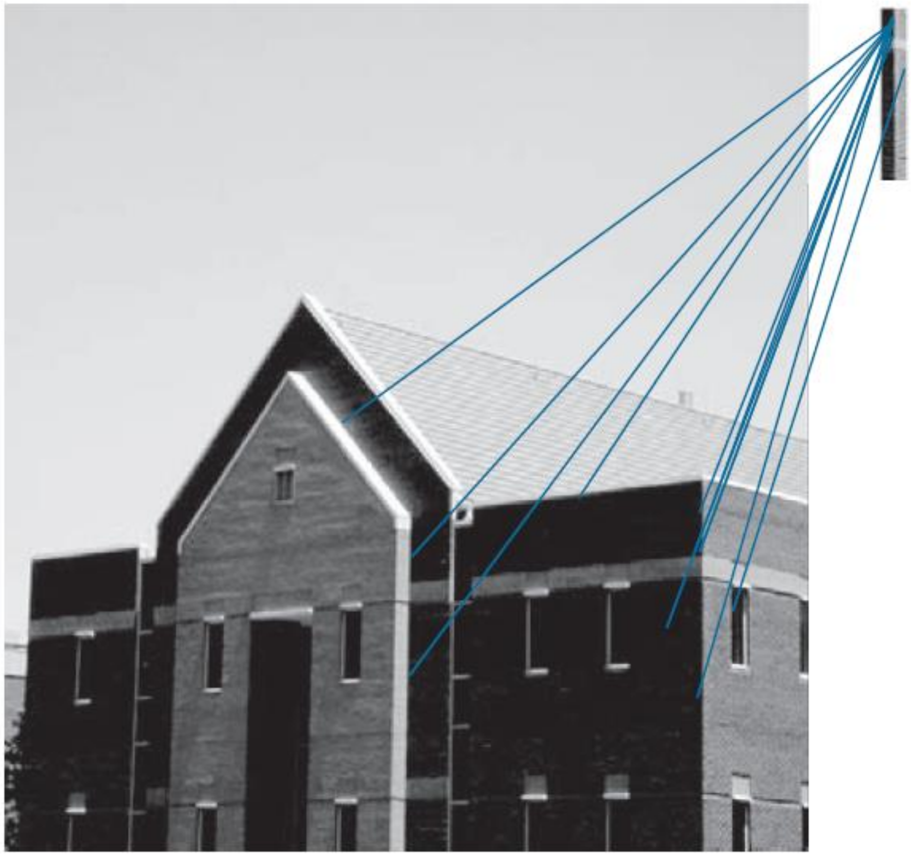
Half size in both directions
195 keypoints

Matched: 7 keypoints
1 is incorrect

Image matching using SIFT



Matched: 10 keypoints
2 are incorrect



Matched: 11 keypoints
4 are incorrect

Assignments

- 11.21, 11.24, 11.27, 11.32, 11.35

课后作业题目请对照参考第4版英文原版