180206081104P2009H

图像处理

第七讲:图像的矩阵表示；图像的空间与统计描述

内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

什么是图像？(IV)

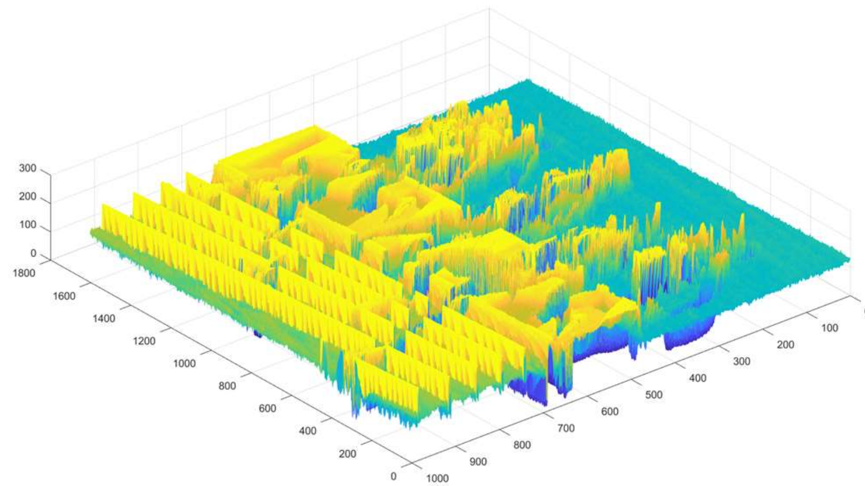
- 一幅二维图像是一个像素阵列。

- 一幅二维图像是一个矩阵。

  → 数学工具: 线性代数, 神经网络

- 一幅二维图像是一个曲面 (经平滑处理)。

  → 数学工具: 微积分，微分几何…



矩阵的特征值分解

$$A \in R^{n \times n}$$

$$\boxed{A = PDP^{-1}}$$

$$Ap_i = \lambda_i p_i$$
$$(A - \lambda_i I)p_i = 0$$

$$P = [p_1, p_2, \dots p_n]$$

$$A[p_1, p_2, \dots p_n] = [p_1, p_2, \dots p_n]\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

奇异值分解 (I)

$$A \in R^{n\times m}, \quad r = rank(A) \le \min\{n, m\}$$

$$\boxed{A = U\Sigma V^T}$$

$$U = [u_1, \dots, u_n] \in R^{n\times n}, \qquad V = [v_1, \dots, v_m] \in R^{m\times m}$$

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \in R^{n\times m}, \Sigma_r = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_r \end{bmatrix}, \qquad \sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$$

https://www.youtube.com/watch?v=Nx0lRBaXoz4
Lecture on SVD by Prof. Gilbert Strang
http://open.163.com/special/opencourse/daishu.html

奇异值分解的基本特性 (I)

$$UU^T = I_{n\times n}$$
$$V^TV = I_{m\times m}$$

$\{u_i\}_{i=1,\ldots,n} \rightarrow$ left singular vectors
$\{v_i\}_{i=1,\ldots,m} \rightarrow$ right singular vectors
$\{\sigma_i\}_{i=1,\ldots,r} \rightarrow$ singular values of A

$\{u_i\}_{i=1,\ldots,n} \rightarrow$ eigenvectors of AA^T
$\{v_i\}_{i=1,\ldots,m} \rightarrow$ eigenvectors of A^TA
$\{\sigma_i\}_{i=1,\ldots,r} \rightarrow$ square roots of non$-$zero eigenvalues of AA^T or A^TA

奇异值分解的基本特性 (II)

$$A = U\Sigma V^T$$

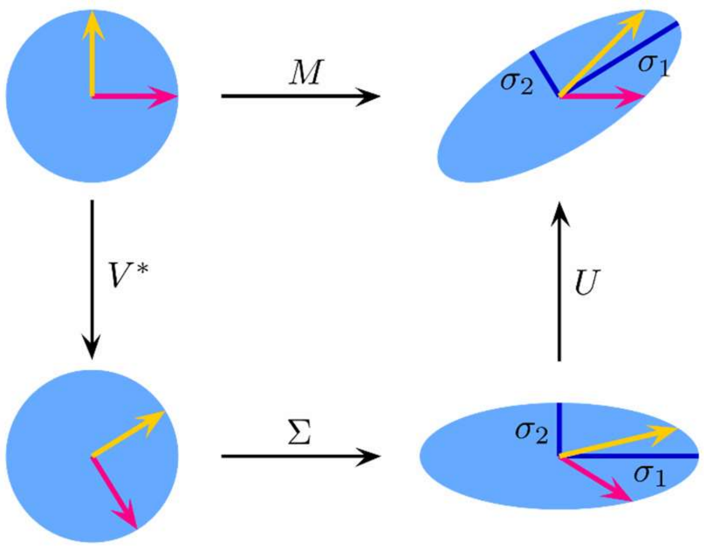
$$AV = U\Sigma$$

$$A[v_1, \dots, v_m] = [u_1, \dots, u_n]\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}$$

$$Av_i = \sigma_i u_i \;\; for\; i = 1, \dots, m$$

$$\boxed{u_i = \frac{1}{\sigma_i} A v_i \;\; \text{for } i = 1, \dots, r}$$

奇异值分解的直观解释



$$M = U \cdot \Sigma \cdot V^*$$

https://en.wikipedia.org/wiki/Singular_value_decomposition

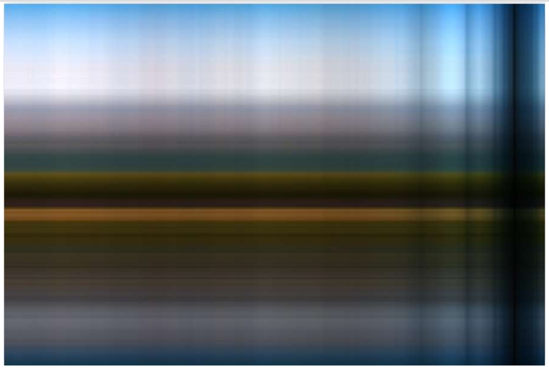
图像的SVD逼近 (I)

$$
\begin{aligned}
A &= U\Sigma V \\
&= [u_1, \dots, u_r]\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}\begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix} \\
&= u_1\sigma_1 v_1^T + \cdots + u_r\sigma_r v_r^T \\
&= \sum_{i=1}^{r} \sigma_i u_i v_i^T \\
&= \sum_{i=1}^{r} A_i
\end{aligned}
$$

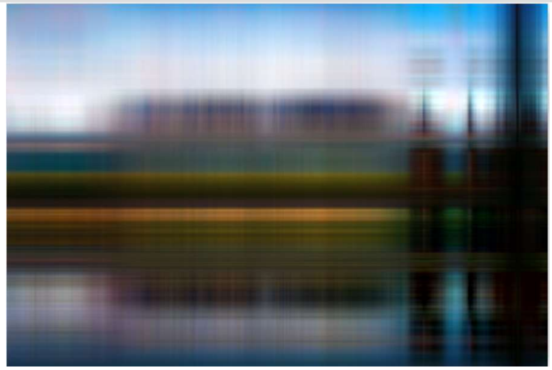
矩阵A表示为秩为1的矩阵的加权和

图像的SVD逼近 (II)

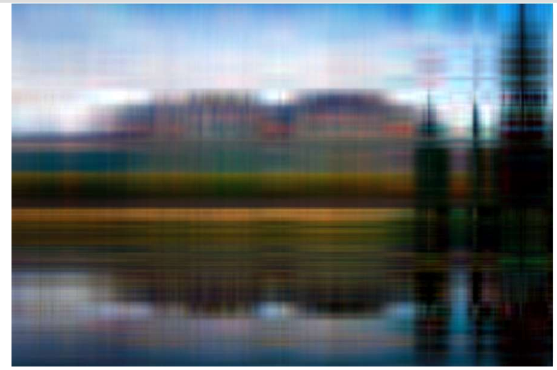
N=1



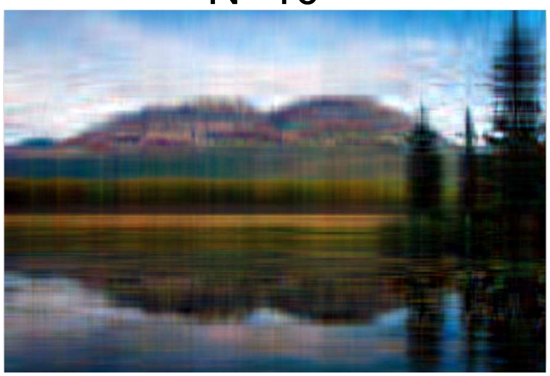
N=2



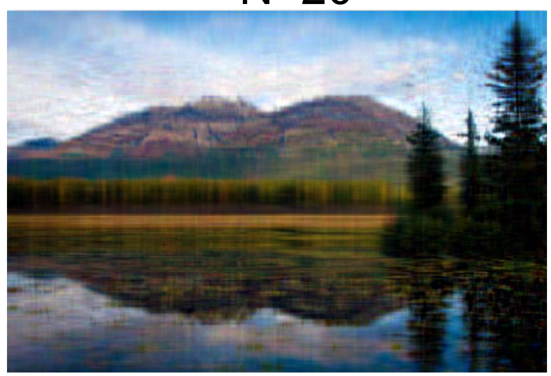
N=3



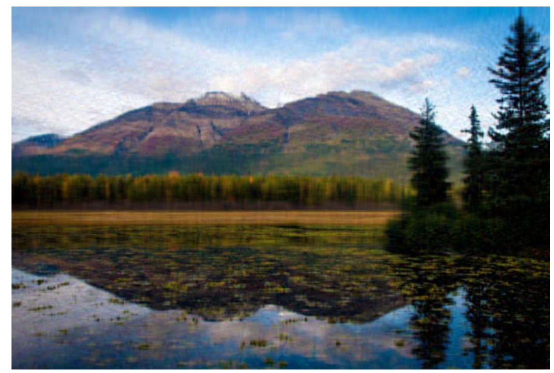
N=10



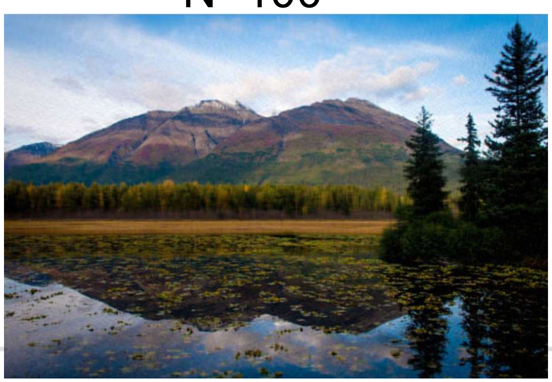
N=20



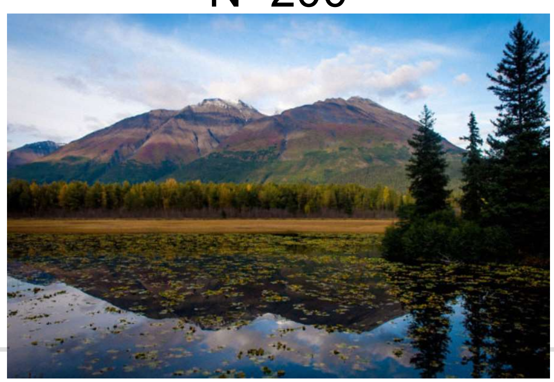
N=50



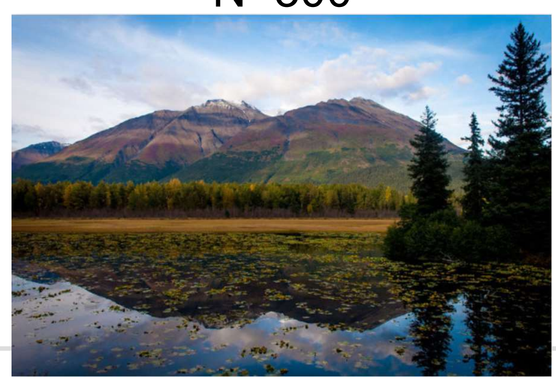
N=100



N=200



N=300



http://timbaumann.info/svd-image-compression-demo/

矢量的各种范数

$$V \in R^{n\times 1}$$

$$V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$\|A\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{1/p}$$

$$\|A\|_1 = \sum_{i=1}^{n} |v_i|$$

$$\|A\|_2 = \left(\sum_{i=1}^{n} |v_i|\right)^{1/2}$$

$$\|A\|_\infty = \max_i |v_i|$$

矩阵的各种范数

$A \in R^{n\times m}$

$$\|A\|_1 = \max_{1\le j\le m} \sum_{i=1}^{n} |a_{ij}| \ \rightarrow \text{maximum absolute column sum of } A$$

$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{m} |a_{ij}| \ \rightarrow \text{maximum absolute row sum of } A$$

$$\|A\|_2 = \sigma(A) \quad \text{maximum singular value of }_{\max}$$

$$\|A\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2 \right)^{1/2} \ \rightarrow \text{Frobenius norm } \ A$$

矩阵的各种范数计算示例

$$A=\begin{bmatrix}-3 & 5 & 7\\ 2 & 6 & 4\\ 0 & 2 & 8\end{bmatrix}$$

$$\begin{aligned}\|A\|_1 &= \max\{|-3|+2+0;5+6+2;7+4+8\}\\ &= \max\{5,13,19\}\\ &= 19\\ \|A\|_\infty &= \max\{|-3|+5+7;2+6+4;0+2+8\}\\ &= \max\{15,12,10\}\\ &= 15\end{aligned}$$

图像的奇异值分解的应用 (I): 最优逼近

$$A \approx \sum_{i=1}^{k} A_i = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

$$\sum_{i=1}^{k} \sigma_i u_i v_i^T = \arg \min_{X \in R^{n \times m}, rank(X) \le k} \{\|A - X\|_2\}$$

$$\sum_{i=1}^{k} \sigma_i u_i v_i^T = \arg \min_{X \in R^{n \times m}, rank(X) \le k} \{\|A - X\|_F\}$$

图像的奇异值分解的应用 (II): 图像压缩和去噪声

N=1

N=29

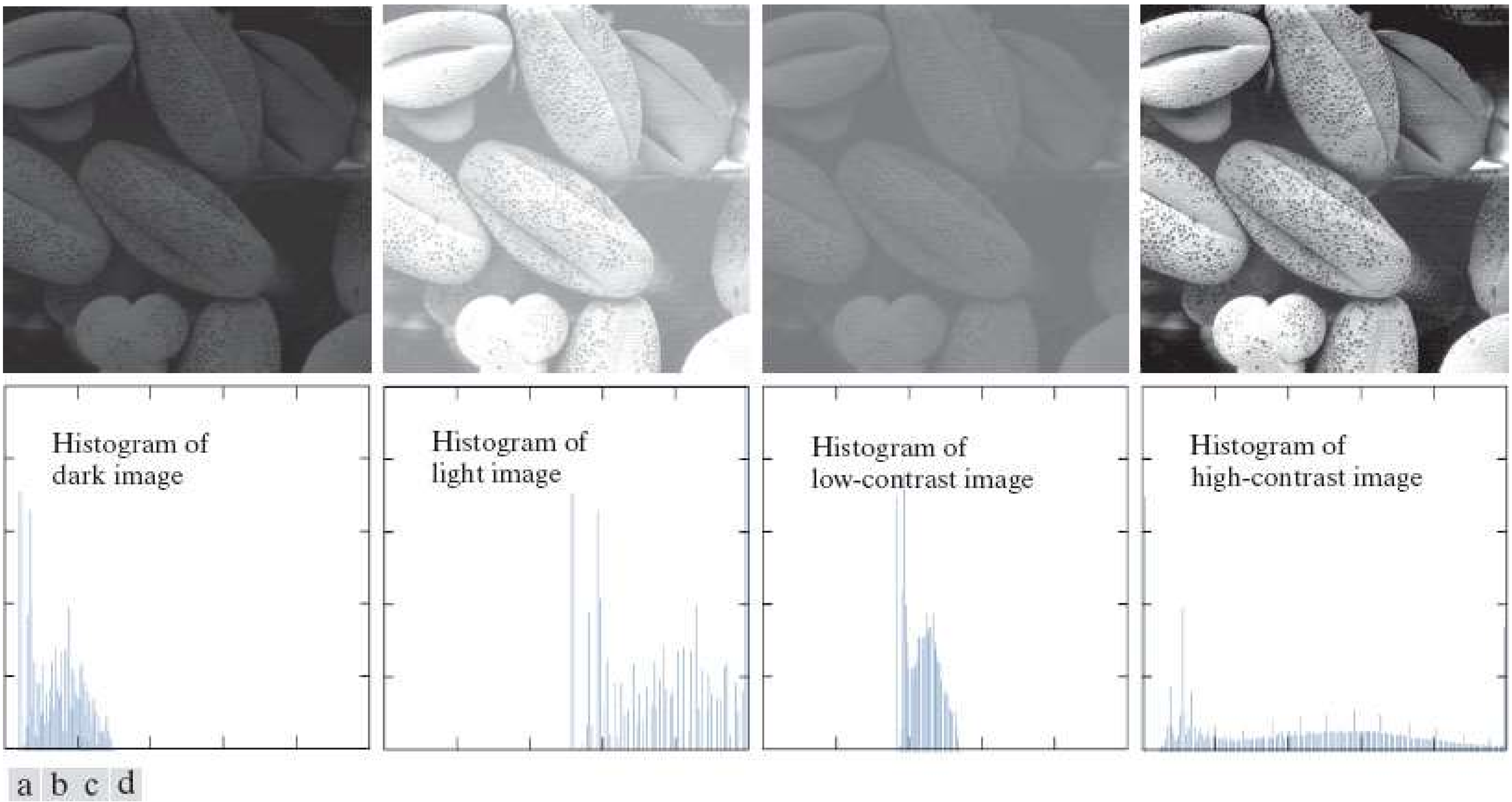
N=61

https://demonstrations.wolfram.com/ImageCompressionViaTheSingularValueDecomposition/

内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

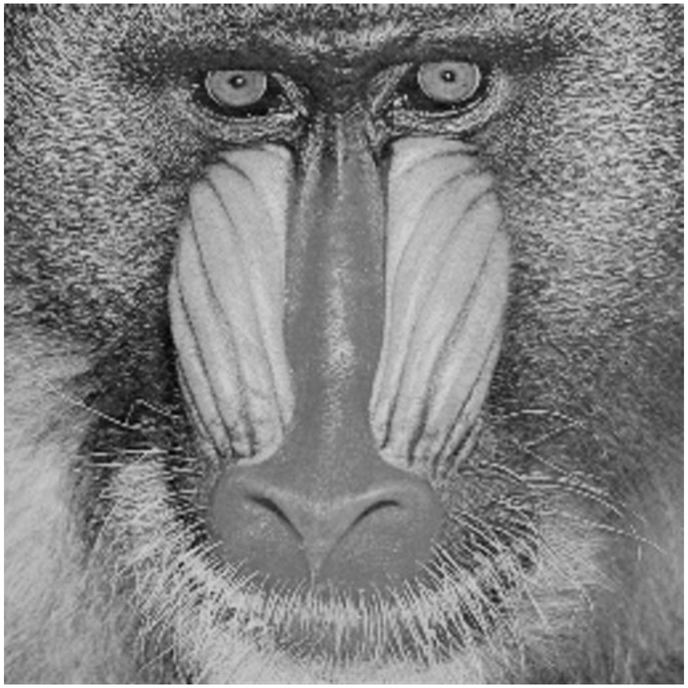
图像的灰度直方图表示


Histogram of dark image
Histogram of light image
Histogram of low-contrast image
Histogram of high-contrast image
a b c d

为什么需要对于图像进行统计描述？(I)

- 示例1：每个像素的灰度信号受到噪声的影响是一个随机变量。

- 示例2：两个或者多个像素的灰度之间的关系可以用多维随机向量的联合分布描述。



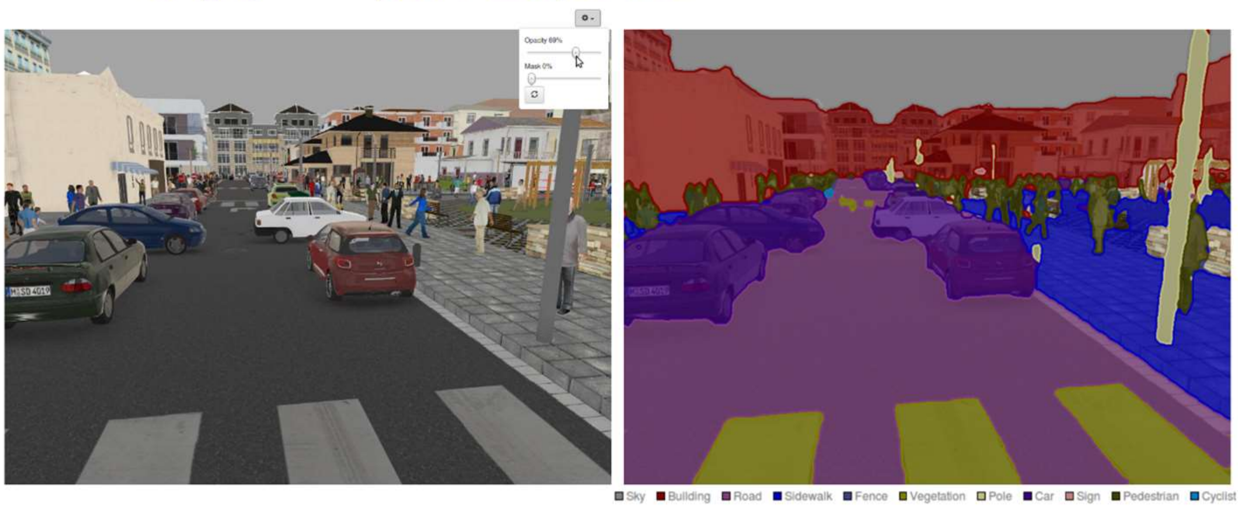
为什么需要对于图像进行统计描述？(II)

- 示例3：动态图像中单个像素的灰度信号是一个时域随机过程。
- 示例4：不同位置的像素灰度可以看成是一个一维空间域随机过程或者二维空间域随机场。

为什么需要对于图像进行统计描述？(III)

- 示例5：物体的分类与识别

- 示例6：物体的分割



图像的统计描述是图像处理与计算机视觉的核心概念与方法

对于图像进行统计描述不仅非常自然而且使得我们可以采用概率与统计方法对于图像进行处理和理解

内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

概率论复习

内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

样本空间

- 一个随机试验的所有可能的结果的集合称为该实验的样本空间。
- 例子1：抛掷一枚硬币 $S=\{H,T\}$
- 例子2：抛掷一颗骰子 $S=\{1,2,3,4,5,6\}$
- 例子3：抛掷二枚硬币

$$S=\{(H,H),(H,T),(T,H),(T,T)\}$$

概率的形式化(公理化)定义

- 对于样本空间*S*的每一个事件*E*，如果一个函数满足*P*(*E*)如果满足以下三个条件，该函数称为事件*E*的概率

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. $P\left(\overset{i=1}{\overset{\infty}{\cup}} E\right) = \sum_{i=1}^{\infty} P(E_i)$ if $E_i \cap E_j = \emptyset$ for $i \neq j$

随机变量的分布函数

- 单一随机变量的累积分布函数(cumulative distribution function, CDF)，简称分布函数的定义如下：

$$F(b) = P\{X \le b\}$$

- 两个随机变量的联合累积分布函数(joint cumulative distribution function, CDF)，简称分布函数的定义如下：

$$F(a,b) = P\{X \le a, Y \le b\}$$

随机过程的基本概念

- 一个随机过程$\{X(t), t \in T\}$是随机变量的一个集合。
- 集合*T*称为此过程的指标集 (index set)。如果*T*是可数集，该随机过程称为一个离散随机过程。如果*T*是一个实数区间，该随机过程称为一个连续随机过程。
- 随机变量$X(t)$所有可能取的值的全体称为该随机过程的状态空间。

随机过程的基本数字特征

- 均值函数 $m(t) \triangleq E\left(X(t)\right)$
- 方差函数 $D(t) \triangleq E\left\{\left(X(t) - m(t)\right)^2\right\}$
- 协方差函数 $R(s,t) \triangleq \operatorname{cov}\left(X(s), X(t)\right)$

平稳随机过程

- 宽平稳随机过程 (协方差平稳过程) wide sense stationary
  - $D(X(t))$ 存在 (二阶矩过程)
  - $E(x(t)) = m, \qquad \operatorname{cov}(X_t, X_{t+\tau}) = R(\tau)$
- 严平稳随机过程 strictly stationary

$$F\left(X_{t_1}, X_{t_2}, \dots, X_{t_n}\right) = F\left(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}\right)$$

随机过程的遍历性(各态历经性)的涵义

- 对于一个宽平稳随机过程 $\{X(t), t \in T\}$

$$\bar{X}(t) \triangleq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)dt = E\big(X(t)\big)$$

数学期望的遍历性 ergodic

$$\begin{aligned}\bar{R}(t, t+\tau) &\triangleq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \big(X(t)X(t+\tau) - \bar{X}(t)\bar{X}(t+\tau)\big)dt \\ &= R(\tau)\end{aligned}$$

协方差函数的遍历性 ergodic

各态遍历性的一个例子

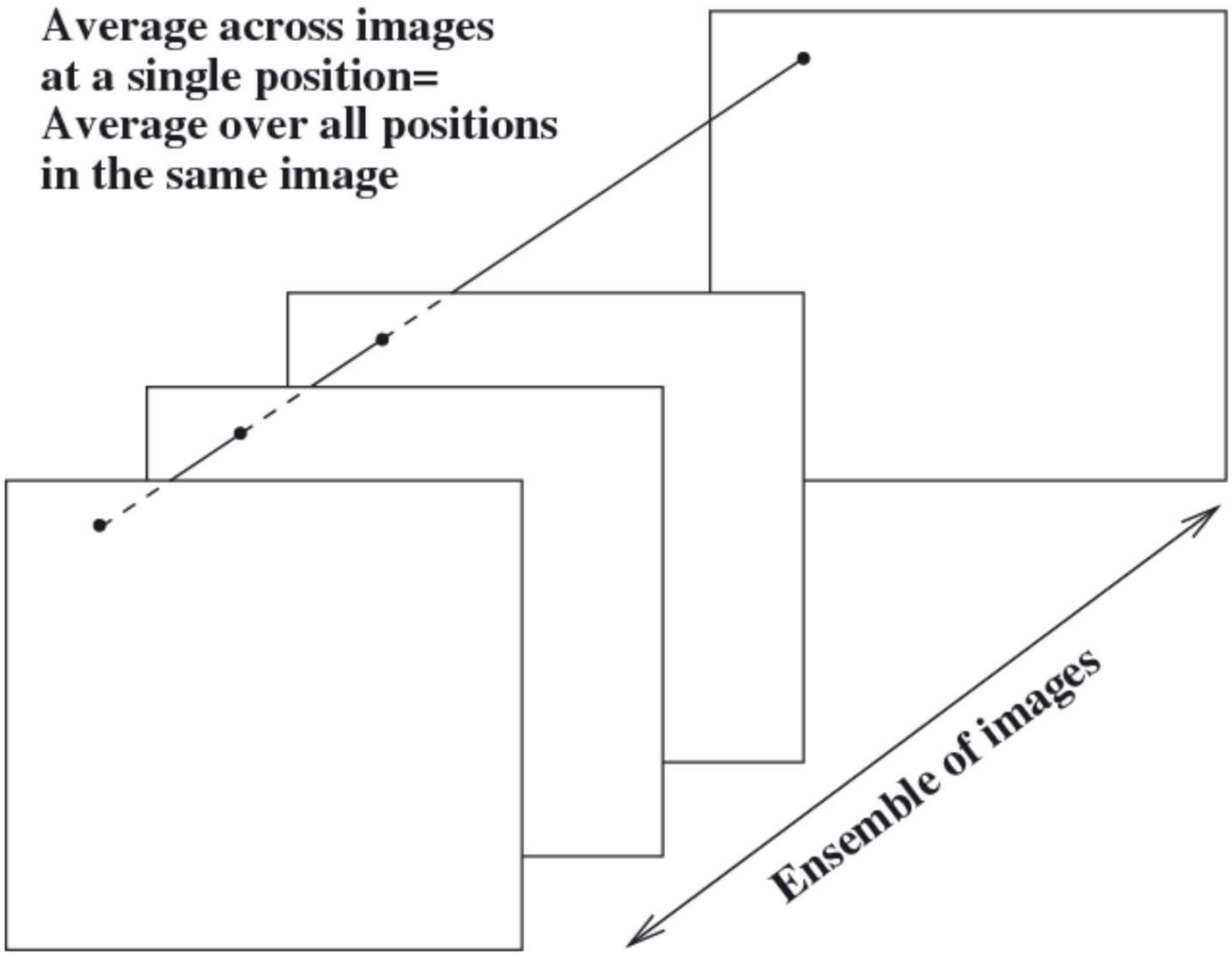


Figure 3.5: Ergodicity in a nutshell.

马尔科夫链的基本概念

- 如果随机序列 $\{X_n, n \geq 0\}$ 对于任意

$$i_0, i_1, \dots, i_n, i_{n+1} \in T, n \in S, P\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} > 0$$

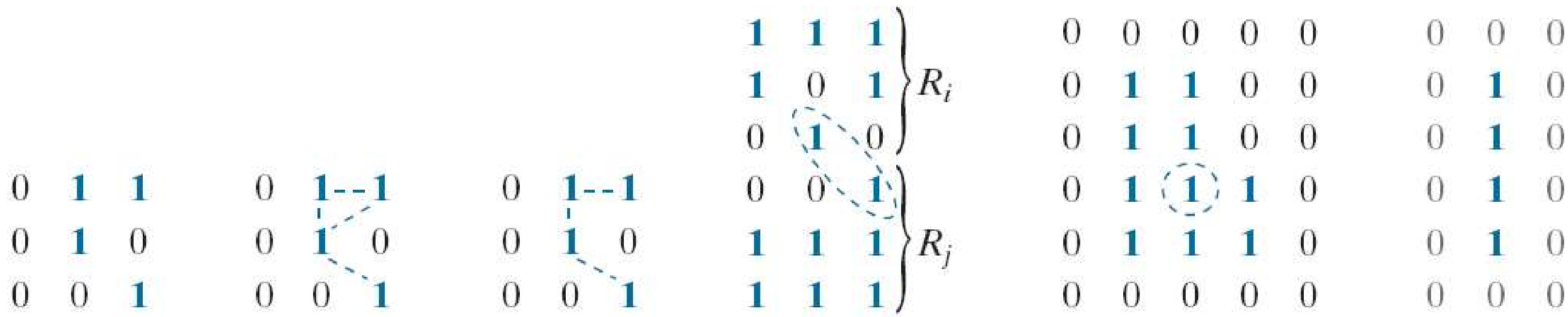
$$P\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} | X_n = i_n\}$$

$$P\{X_{n+1} = i_{n+1} | X_n = i_n\} \triangleq p_{ij}(n)$$

内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

像素间的基本空间关系：相邻像素



a b c d e f

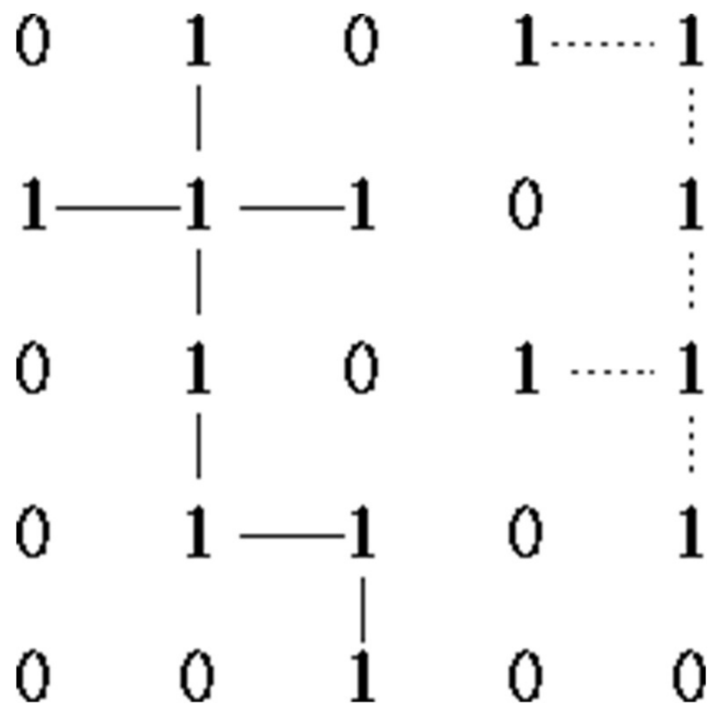
$$p(x,y) \rightarrow N_4(p) = \{(x+1,y),(x-1,y),(x,y+1),(x,y-1)\}$$

$$p(x,y) \rightarrow N_D(p) = \{(x+1,y+1),(x+1,y-1),(x-1,y+1),(x-1,y-1)\}$$

$$p(x,y) \rightarrow N_8(p) = N_4(p) \cup N_D(p)$$

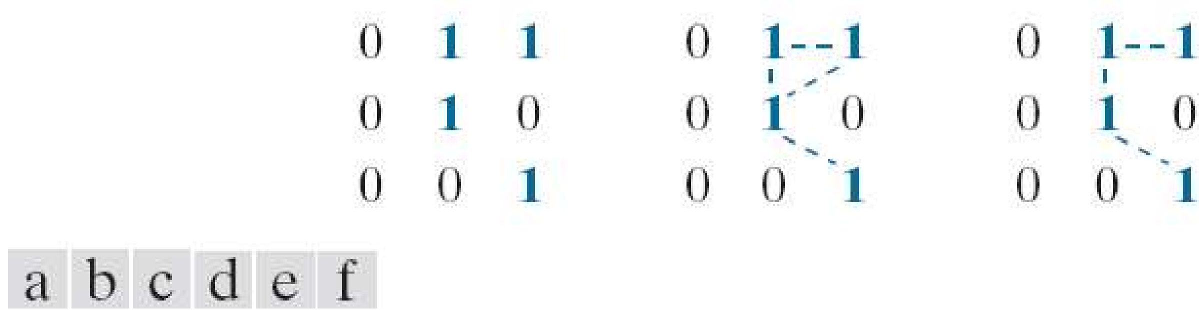
像素间的基本空间关系：邻接性，连通性 (I)

- 4邻接： $q \in N_4(p)$
- 8邻接： $q \in N_8(p)$



像素间的基本空间关系：邻接性，连通性 (II)

- 4连通：$$q \in N_4(p)$$
- 8连通：$$q \in N_8(p)$$
- m(混合)连通：



- 如果像素子集所有像素之间存在通路，该子集称为连通集。

像素间的基本空间关系：前景，背景，边界

- 如果一幅图像包含*k*个不连接的区域

$$R_k, k = 1,2, \dots, K$$

$$R_U = \bigcup_{k=1}^{K} R_k$$ 前景

$$(R_U)^C = V - R_U$$ 背景

- 区域R的边界是指其与补集相邻的点的集合。

像素间不同的距离定义

0	0	0
0	1	0
0	0	0

Image

1.41	1.0	1.41
1.0	0.0	1.0
1.41	1.0	1.41

Distance Transform

Euclidean $$D_{Eucliean} = \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2}$$

0	0	0
0	1	0
0	0	0

Image

2	1	2
1	0	1
2	1	2

Distance Transform

City block $$D_{City} = \left|x_2 - x_1\right| + \left|y_2 - y_1\right|$$

0	0	0
0	1	0
0	0	0

Image

1	1	1
1	0	1
1	1	1

Distance Transform

Chessboard $$D_{Chess} = \max\left(\left|x_2 - x_1\right|, \left|y_2 - y_1\right|\right)$$

图像物体的边界跟踪 (I)

Illustration of the first few steps in the boundary-following algorithm. The point to be processed next is labeled in bold, black; the points yet to be processed are gray; and the points found by the algorithm are shaded. Squares without labels are considered background (0) values.

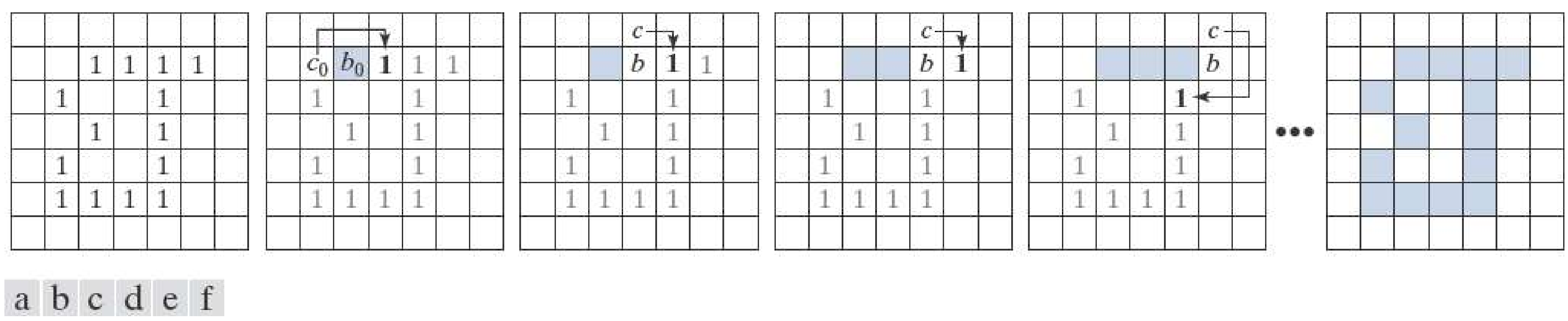
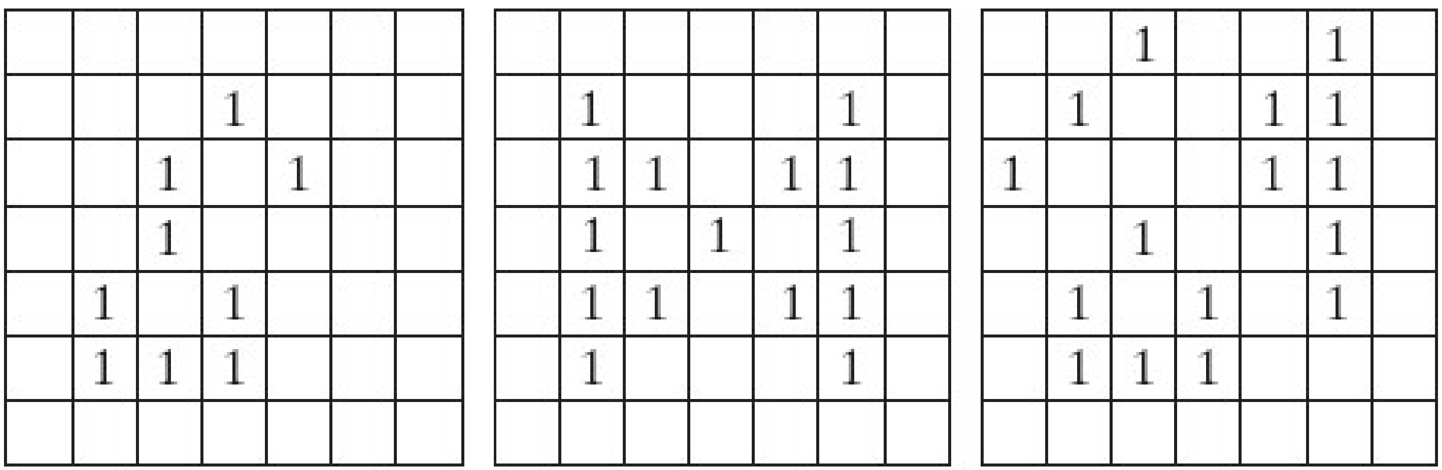


Figure 12.1

图像物体的边界跟踪 (II)

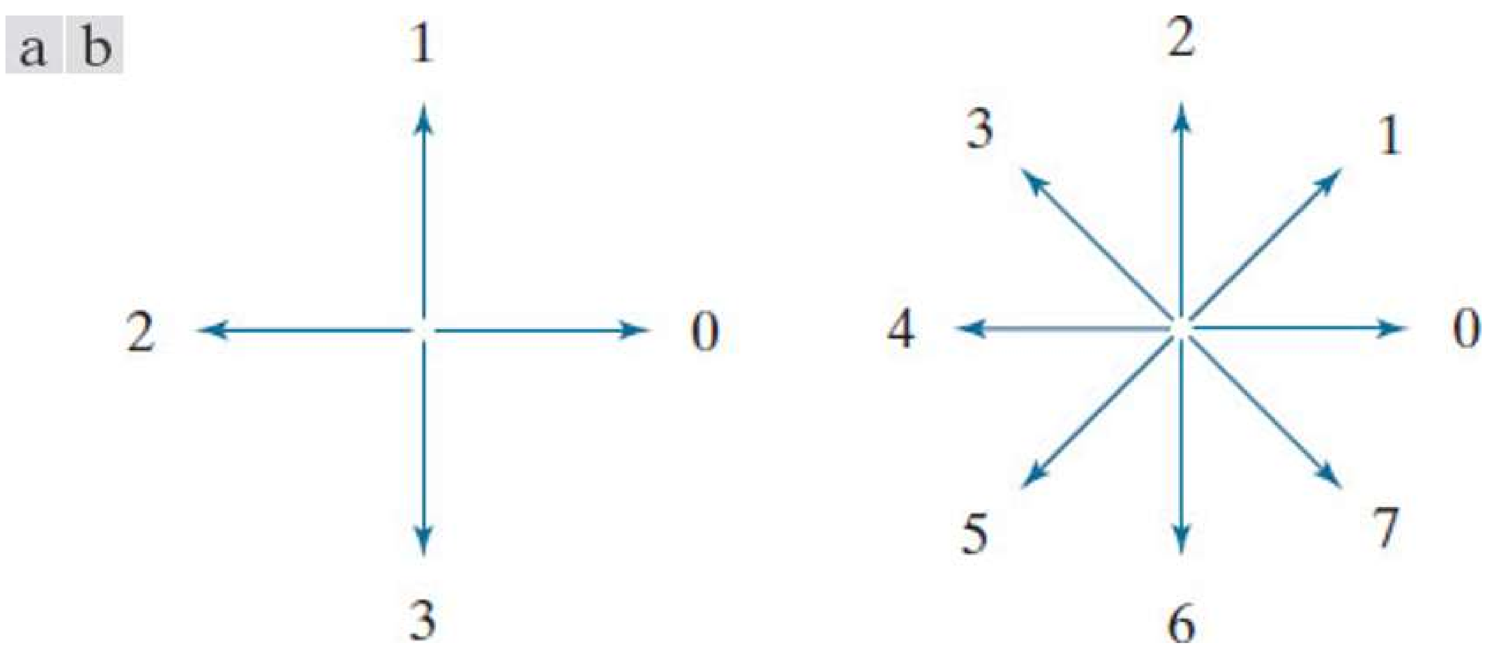
Figure 12.2 Examples of boundaries that can be processed by the boundary-following algorithm. (a) Closed boundary with a branch. (b) Self-intersecting boundary. (c) Multiple boundaries (processed one at a time).



a b c

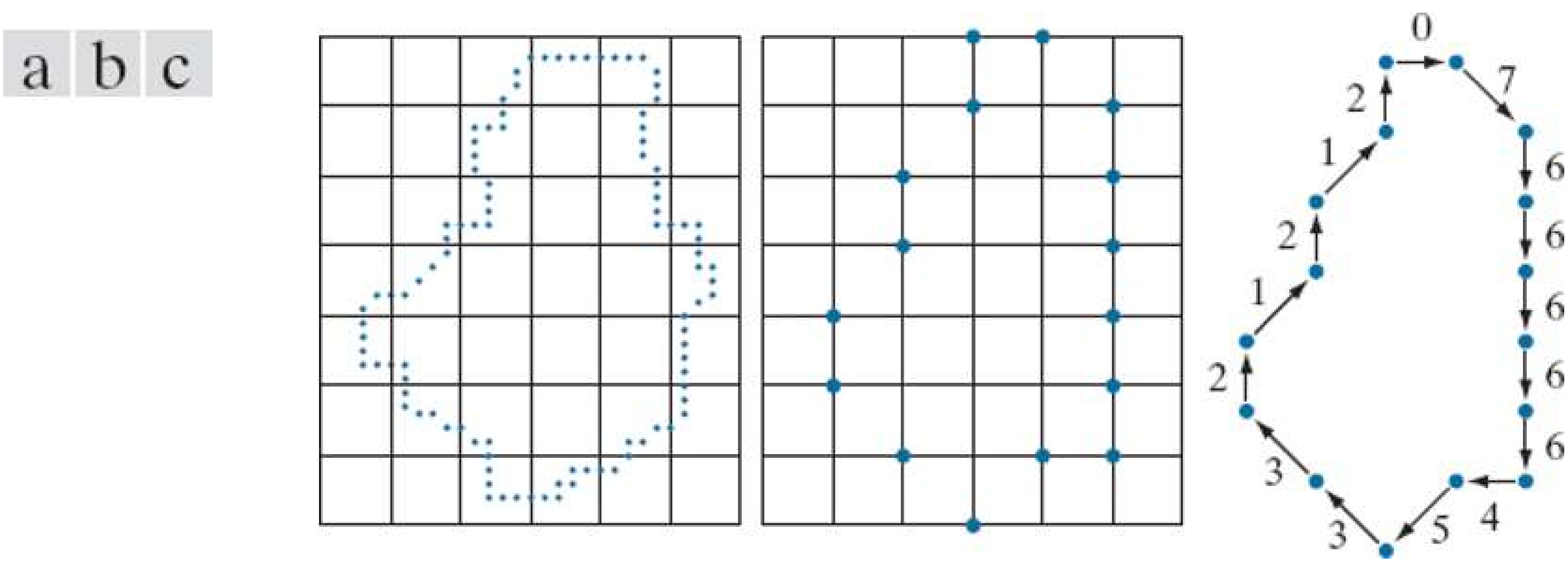
图像物体的边界跟踪 (III)

Direction numbers for (a) 4-directional chain code, and (b) 8-directional chain code.



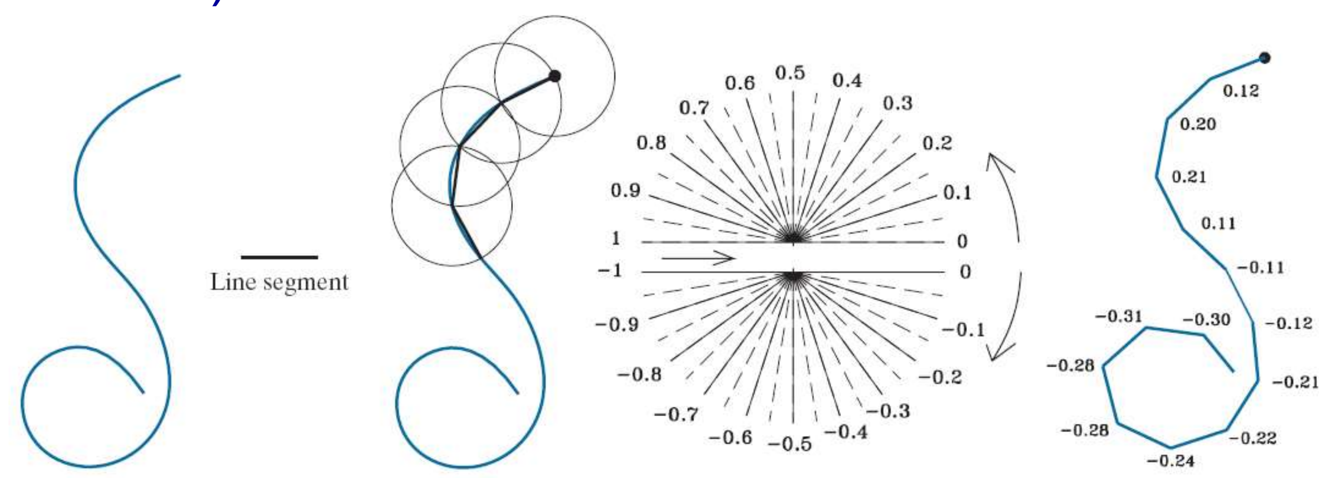
图像物体的边界跟踪 (IV)

(a) Digital boundary with resampling grid superimposed. (b) Result of resampling. (c) 8-directional chain-coded boundary.



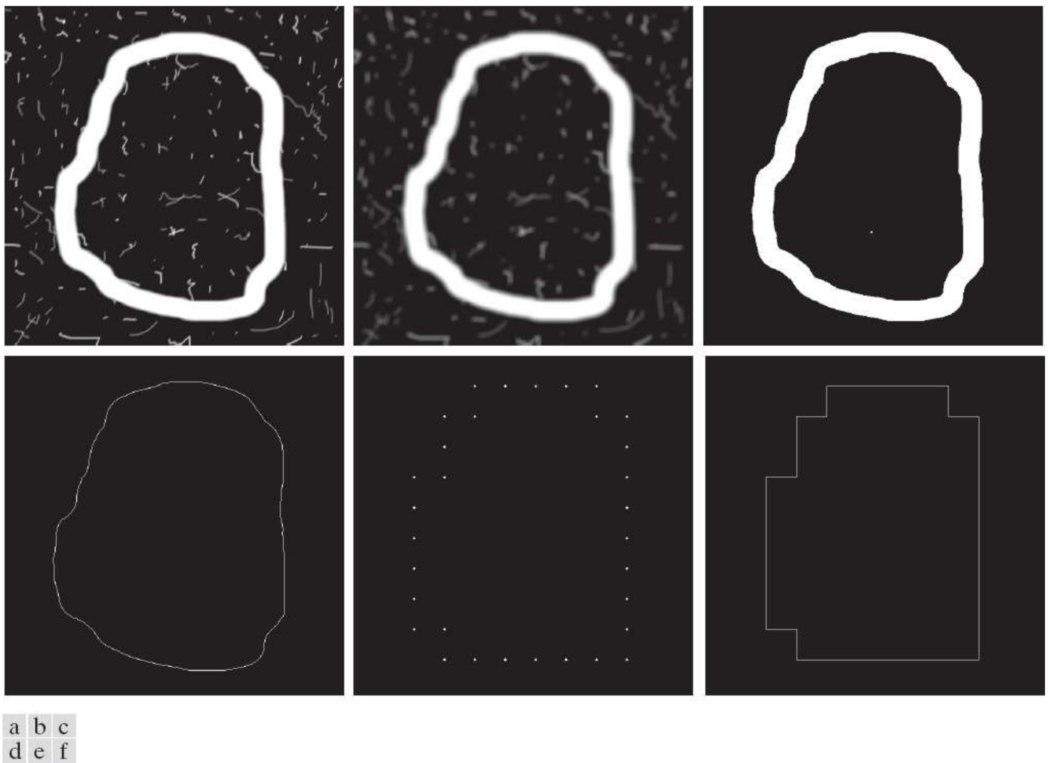
图像物体的边界跟踪 (V)

(a) An open curve. (b) A straight-line segment. (c) Traversing the curve using circumferences to determine slope changes; the dot is the origin (starting point). (d) Range of slope changes in the open interval (−1,1) (the arrow in the center of the chart indicates direction of travel). There can be ten subintervals between the slope numbers shown.(e) Resulting coded curve showing its corresponding numerical sequence of slope changes. (Courtesy of Professor Ernesto Bribiesca, IIMAS-UNAM, Mexico.)



a b c d e

图像物体的边界跟踪 (VI)



内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

图像的统计特征 (I)

$$i \rightarrow p(i)$$

$$Mean: \quad \sum_{i=0}^{L-1} i \cdot p(i)$$

$$Variance: \quad \sum_{i=0}^{L-1} (i-\mu)^2 \cdot p(i)$$

$$Skewness: \quad \sum_{i=0}^{L-1} (i-\mu)^3 \cdot p(i)$$

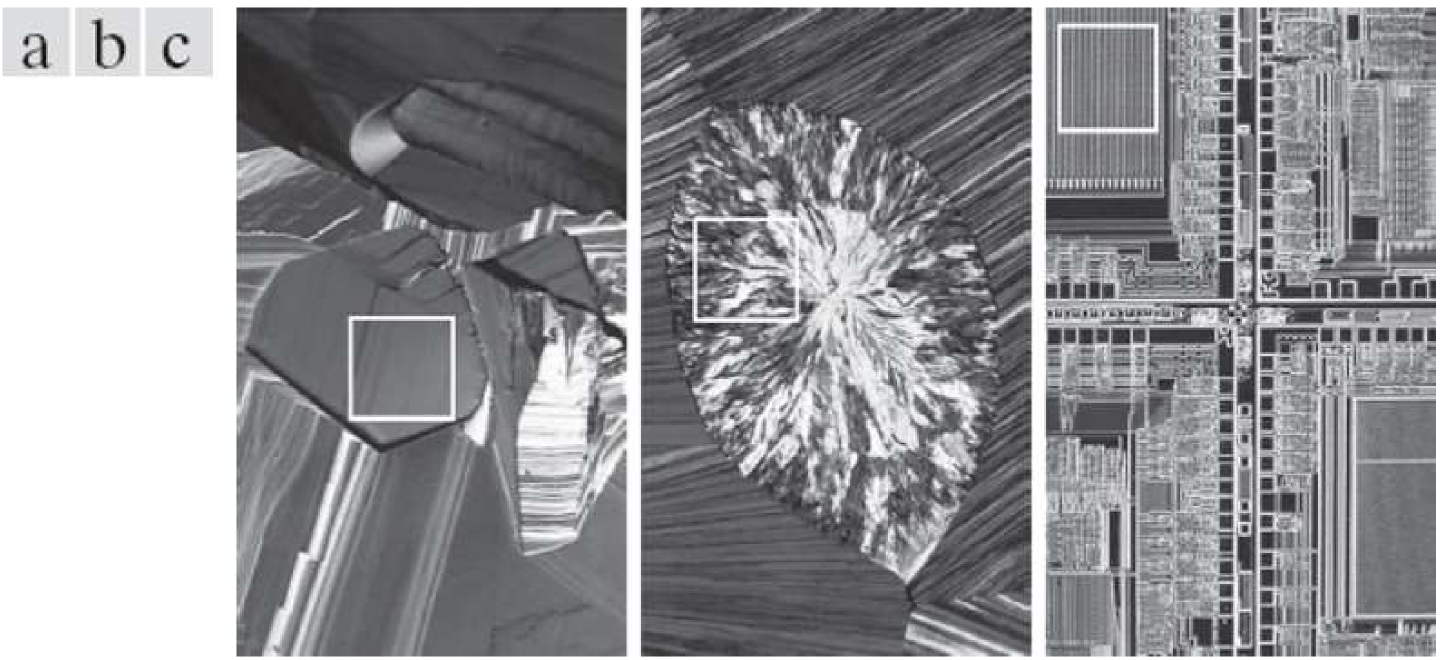
$$Kurtosis: \quad \sum_{i=0}^{L-1} (i-\mu)^4 \cdot p(i)$$

$$Uniformity: \quad \sum_{i=0}^{L-1} p^2(i)$$

$$\text{Entropy:} \quad \sum_{i=0}^{L-1} p(i) \log_2 p\,(i)$$

图像的统计特征 (II)

The white squares mark, from left to right, smooth, coarse, and regular textures. These are optical microscope images of a superconductor, human cholesterol, and a microprocessor. (Courtesy of Dr. Michael W. Davidson, Florida State University.)



图像的统计特征 (III)

Statistical texture measures for the subimages in Fig. 12.29.

Texture	Mean	Standard deviation	*R* (normalized)	3rd 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

图像的统计特征 (IV)

1	1	7	5	3	2
5	1	6	1	2	5
8	8	6	8	1	2
4	3	4	5	5	1
8	7	8	7	6	2
7	8	6	2	6	2

Image *f*

	1	2	3	4	5	6	7	8
1	1	2	0	0	0	1	1	0
2	0	0	0	0	1	1	0	0
3	0	1	0	1	0	0	0	0
4	0	0	1	0	1	0	0	0
5	2	0	1	0	1	0	0	0
6	1	3	0	0	0	0	0	1
7	0	0	0	0	1	1	0	2
8	1	0	0	0	0	2	2	1

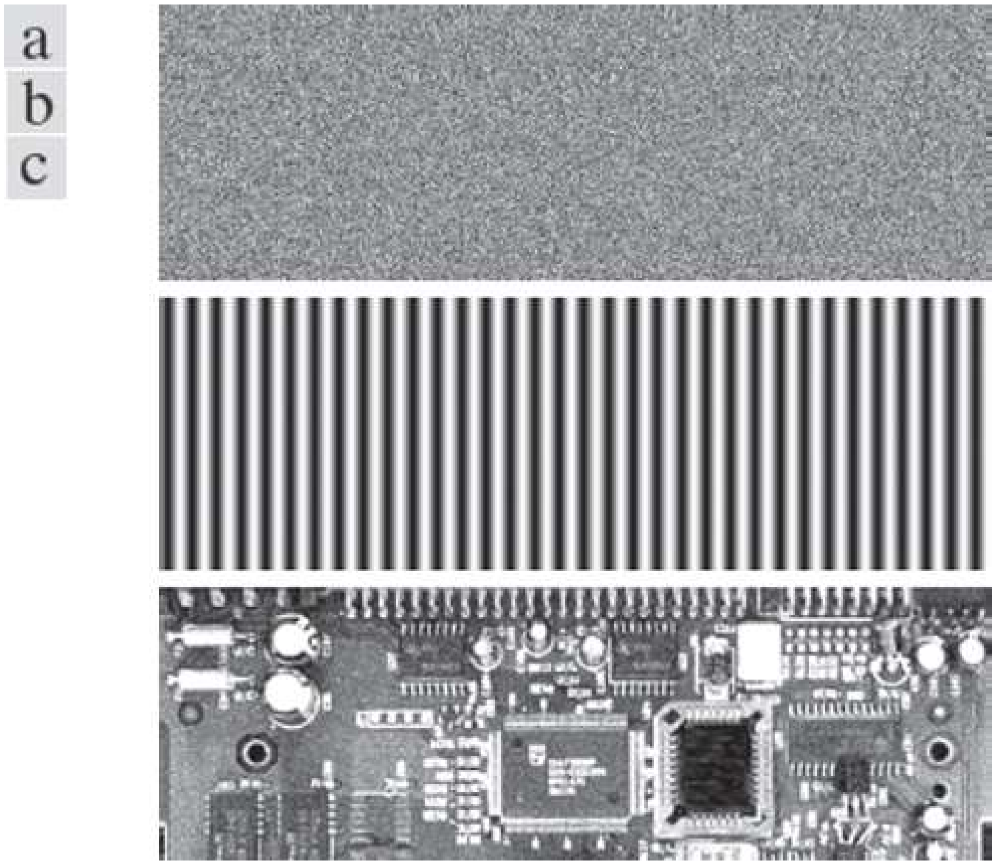
Co-occurrence matrix **G**

图像的统计特征 (V)

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

Descriptor	Explanation	Formula
Maximum probability	Measures the strongest response of **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 **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 **G**. The range of values is [0, 1], with the maximum being achieved when **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 **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 $2\log_2 K$.	$-\sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij}\log_2 p_{ij}$

图像的统计特征 (VI)



图像的统计特征 (VII)

a b c

图像的统计特征 (VIII)

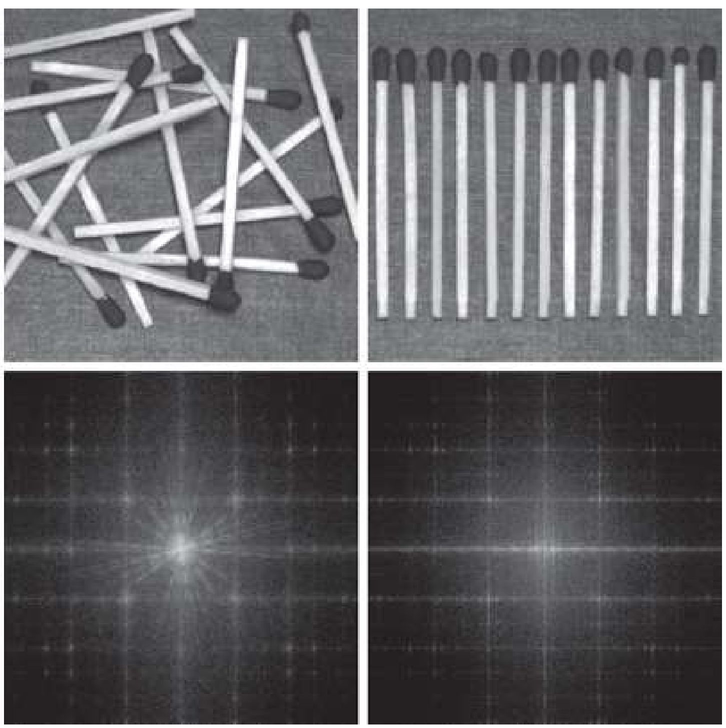
Descriptors evaluated using the co-occurrence matrices displayed as images in Fig. 12.32.

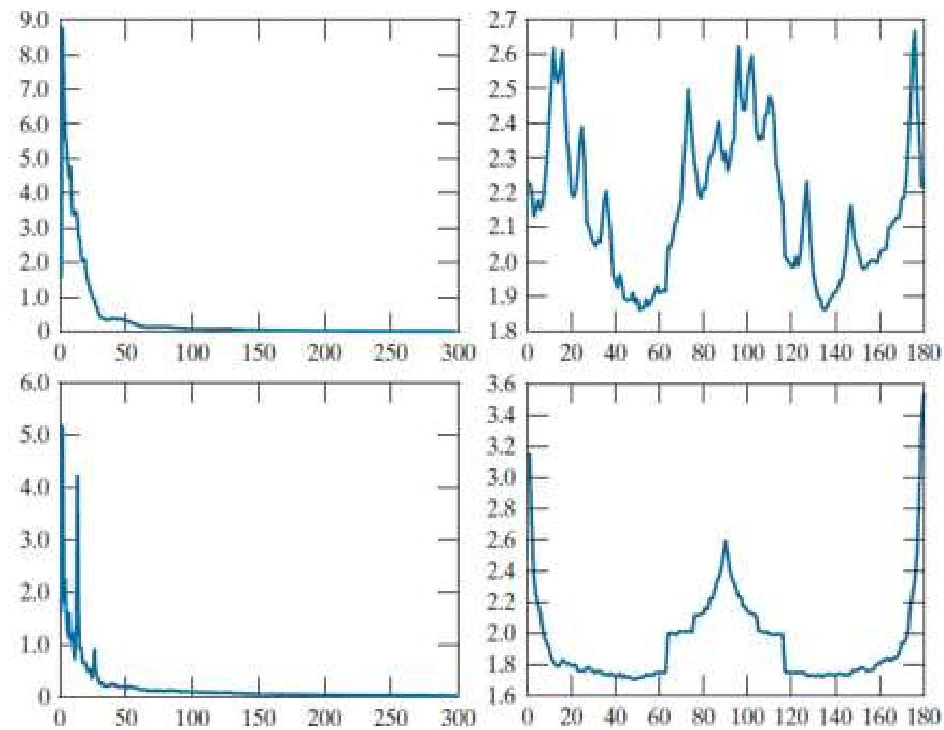
Normalized Co-occurrence Matrix	Maximum Probability	Correlation	Contrast	Uniformity	Homogeneity	Entropy
$\mathbf{G}_1/n_1$	0.00006	−0.0005	10838	0.00002	0.0366	15.75
$\mathbf{G}_2/n_2$	0.01500	0.9650	00570	0.01230	0.0824	06.43
$\mathbf{G}_3/n_3$	0.06860	0.8798	01356	0.00480	0.2048	13.58

图像的统计特征 (IX)

$$S(r)=\sum_{\theta=0}^{\pi}S_{\theta}(r) \qquad S(\theta)=\sum_{r=1}^{R}S_{r}(\theta)$$

a b
c d





内容提要

- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- **随机场的基本概念**
- 图像的图表示

基本概念 (I)

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. PAMI-6, NO. 6, NOVEMBER 1984 721

Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images

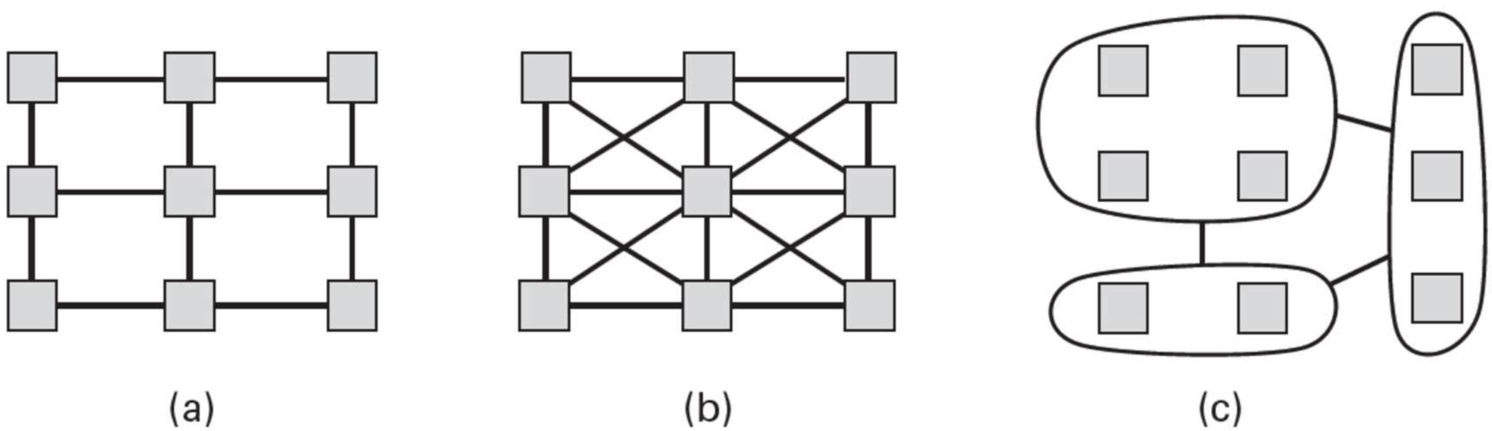
STUART GEMAN AND DONALD GEMAN

- 将每个像素的灰度值视为网格式物理系统中原子/分子的状态。
- 采用统计物理的概念、方法与模型对于图像进行建模。
- 采用贝叶斯概率模型。

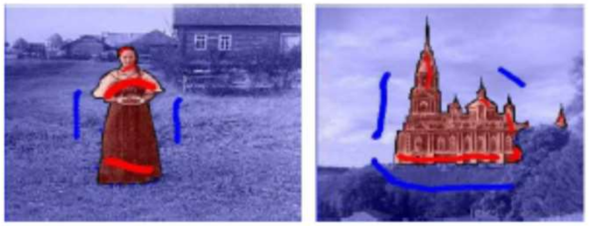
随机场的基本概念

- 随机场是随机过程的一个广义化扩展。
- 给定一个概率空间(Ω, F, P)，如下的随机变量集合称为一个随机场，其中 *T*是一个广义的拓扑空间。

$$\{F(t), t \in T\}$$



图像随机场模型的一些代表性应用 (I)



Interactive figure-ground segmentation (Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006)

Surface context (Hoiem et al., 2005)

Semantic labeling (He et al., 2004; Shotton et al., 2006; Gould et al., 2009)

Stereo matching (Scharstein and Szeliski, 2002)

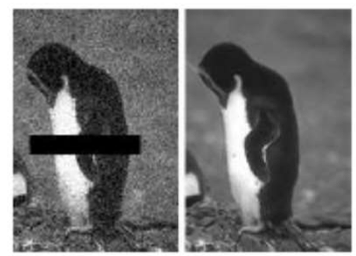


Image denoising (Felzenszwalb and Huttenlocher, 2004; Szeliski et al., 2008)

http://users.cecs.anu.edu.au/~sgould/

图像随机场模型的一些代表性应用 (II)

(Agarwala et al., 2004)

http://users.cecs.anu.edu.au/~sgould/

Presented at SSIP 2011, Szeged, Hungary

Markov Random Fields in Image Segmentation

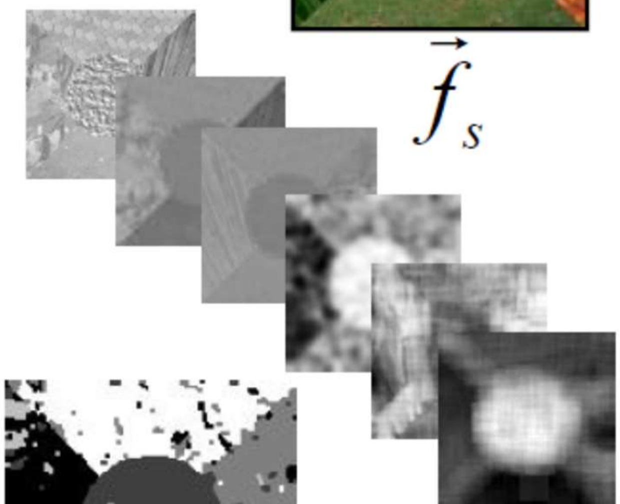
Zoltan Kato

Image Processing & Computer Graphics Dept.
University of Szeged
Hungary


sky
bldg
foreground
grass

Segmentation as a Pixel Labelling Task

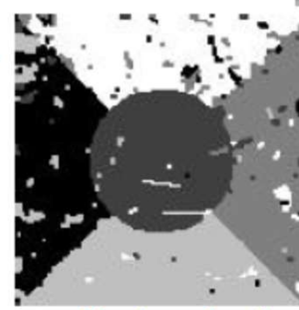
$\vec{f}_s$



1. Extract features from the input image
   - Each pixel *s* in the image has a feature vector
   - For the whole image, we have

$$f = \{\vec{f}_s : s \in S\}$$

2. Define the set of labels Λ
   - Each pixel *s* is assigned a label $\omega_s \in \Lambda$
   - For the whole image, we have



$$\omega = \{\omega_s , s \in S\}$$

- For an $N \times M$ image, there are $|\Lambda|^{NM}$ possible labelings.
   - **Which one is the right segmentation?**

Probabilistic Approach, MAP

- Define a *probability measure* on the set of all possible labelings and select the most likely one.
- $P(\omega \mid f)$ measures the probability of a labelling, given the observed feature f
- Our goal is to find an optimal labeling $\hat{\omega}$ which *maximizes* $P(\omega \mid f)$
- This is called the *Maximum a Posteriori* (MAP) estimate:

$$\hat{\omega}^{MAP} = \arg\max_{\omega \in \Omega} P(\omega \mid f)$$

Bayesian Framework

- By Bayes Theorem, we have

$$P(\omega \mid f) = \frac{P(f \mid \omega)P(\omega)}{P(f)} \propto \underbrace{P(f \mid \omega)}_{\text{likelihood}}\underbrace{P(\omega)}_{\text{prior}}$$

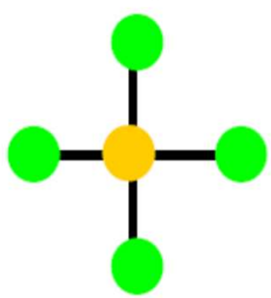
- $P(f)$ is constant
- We need to define $P(\omega)$ and $P(f \mid \omega)$ in our model
  - We will use Markov Random Fields

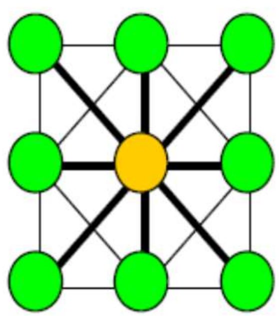
Why MRF Modelization?

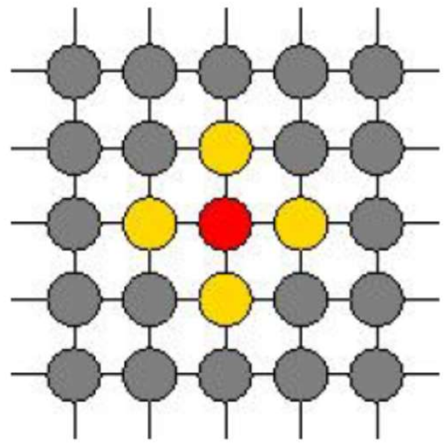
- In real images, regions are often homogenous; neighboring pixels usually have similar properties (intensity, color, texture, …)
- Markov Random Field (MRF) is a probabilistic model which captures such contextual constraints
- Well studied, strong theoretical background
- Allows MCMC sampling of the (hidden) underlying structure ➔ Simulated Annealing
- Fast and exact solution for certain type of models ➔ Graph cut [Kolmogorov]

Definition – Neighbors

- For each pixel, we can define some surrounding pixels as its neighbors.
- Example : 1st order neighbors and 2nd order neighbors





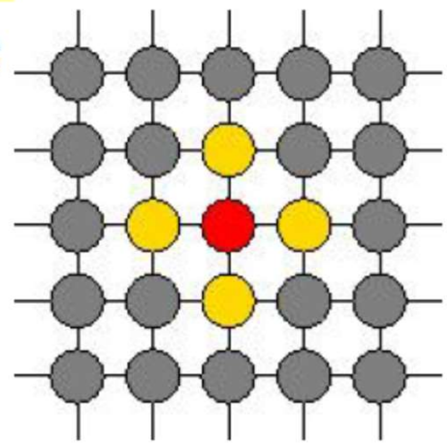


Definition – MRF

- The labeling field *X* can be modeled as a Markov Random Field (MRF) if
  1. For all $\omega \in \Omega : P(X = \omega) > 0$
  2. For every $s \in S$ and $\omega \in \Omega$:

     $$P(\omega_s \mid \omega_r, r \neq s) = P(\omega_s \mid \omega_r, r \in N_s)$$

  N_s denotes the neighbors of pixel *s*

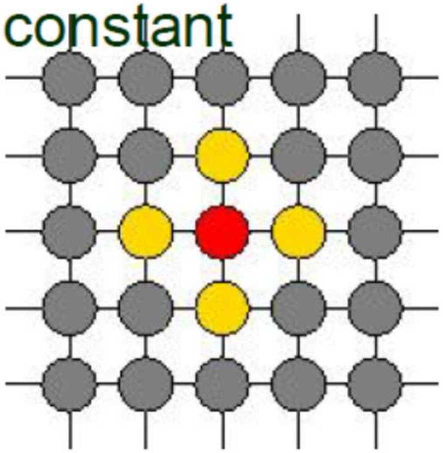


Hammersley-Clifford Theorem

- The Hammersley-Clifford Theorem states that a random field is a MRF if and only if $P(\omega)$ follows a Gibbs distribution.

$$P(\omega) = \frac{1}{Z}\exp(-U(\omega)) = \frac{1}{Z}\exp(-\sum_{c\in C} V_c(\omega))$$

- where $Z = \sum_{\omega\in\Omega} \exp(-U(\omega))$ is a normalization constant
- This theorem provides us an easy way of defining MRF models via *clique potentials*.



MRF-Gibbs 等价性关系

- Gibbs分布

$$P(X=x)=\frac{1}{Z}\exp(-E(x))$$

$$Z=\sum_x \exp(-E(x))$$

- Hammersley-Clifford定理

任何条件随机分布存在一个Gibbs分布如果满足如下三个条件

Positivity: $P(X=x)>0$

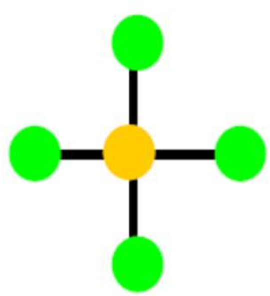
Locality: $P(X_s=x_s \mid X_t=x_t, t\neq s, t\in S)=P(X_s=x_s \mid X_t=x_t, t\in G_s)$

Homogeneity: $P(X_s=x_s \mid X_t=x_t, t\in G_s)$ is the same for all sites s

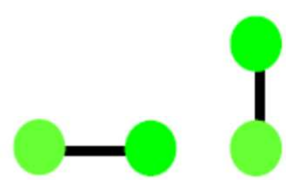
Definition – Clique

- A subset $C \subseteq S$ is called a *clique* if every pair of pixels in this subset are neighbors.
- A clique containing *n* pixels is called *n^{th} order clique*, *denoted by* C_n .
- The set of cliques in an image is denoted by

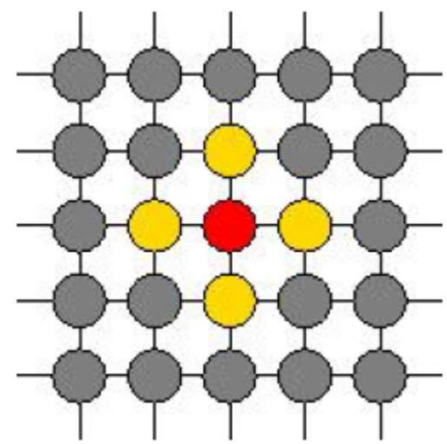
$$C = C_1 \bigcup C_2 \bigcup \ldots \bigcup C_k$$



singleton



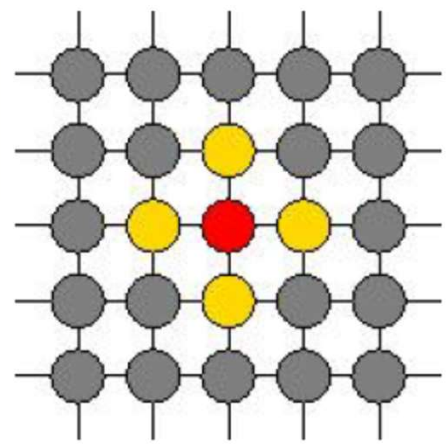
doubleton



Definition – Clique Potential

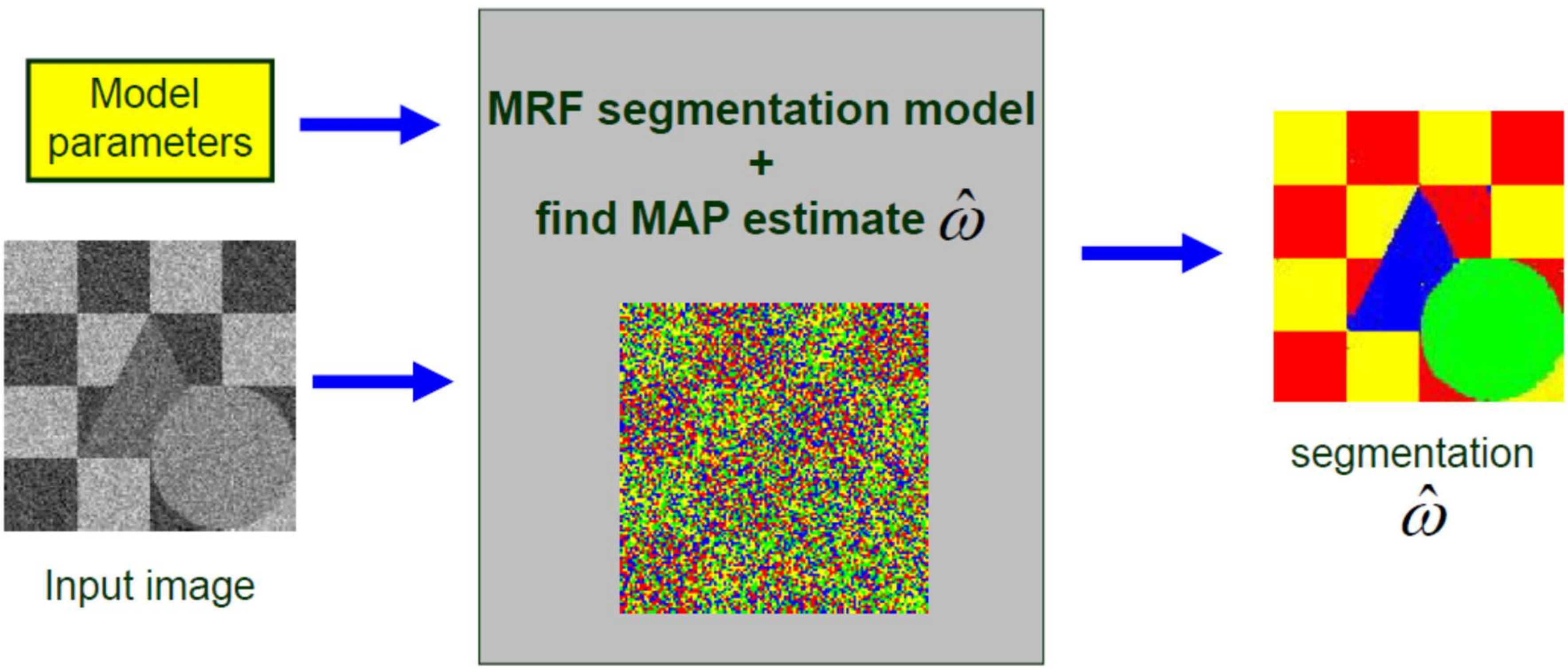
- For each clique *c* in the image, we can assign a value $V_c(\omega)$ which is called *clique potential* of *c*, where ω is the configuration of the labeling field
- The sum of potentials of all cliques gives us the energy $U(\omega)$ of the configuration ω

$$U(\omega) = \sum_{c \in C} V_c(\omega) = \sum_{i \in C_1} V_{C_1}(\omega_i) + \sum_{(i,j) \in C_2} V_{C_2}(\omega_i, \omega_j) + \ldots$$



Segmentation of grayscale images: A simple MRF model

- Construct a segmentation model where regions are formed by spatial clusters of pixels with similar intensity:



Energy function

- Now we can define the energy function of our MRF model:

$$U(\omega) = \sum_s \left(\log(\sqrt{2\pi}\sigma_{\omega_s}) + \frac{(f_s - \mu_{\omega_s})^2}{2\sigma_{\omega_s}^2} \right) + \sum_{s,r} \beta\delta(\omega_s, \omega_r)$$

- Recall:

$$P(\omega \mid f) = \frac{1}{Z}\exp(-U(\omega)) = \frac{1}{Z}\exp(-\sum_{c \in C} V_c(\omega))$$

- Hence

$$\hat{\omega}^{MAP} = \arg\max_{\omega \in \Omega} P(\omega \mid f) = \arg\min_{\omega \in \Omega} U(\omega)$$

马尔可夫随机场的基本概念 (V)

Compactness of Representation

Consider a 1 mega-pixel image, e.g., 1000×1000 pixels. We want to annotate each pixel with a label from $\mathcal{L}$. Let $L = |\mathcal{L}|$.

- There are L^{10^6} possible ways to label such an image.
- A naive encoding—i.e., one big table—would require $L^{10^6} - 1$ parameters.
- A pairwise MRF over $\mathcal{N}_4$ requires $10^6 L$ parameters for the unary terms and $2 \times 1000 \times (1000 - 1) L^2$ parameters for the pairwise terms, i.e., $O(10^6 L^2)$. Even less are required if we share parameters.

马尔可夫随机场的基本概念 (VI)

Inference and Energy Minimization

We are usually interested in finding the most probable labeling,

$$\mathbf{y}^\star = \operatorname*{argmax}_{\mathbf{y}} \mathrm{P}\left(\mathbf{y} \mid \mathbf{x}\right) = \operatorname*{argmin}_{\mathbf{y}} E\left(\mathbf{y}; \mathbf{x}\right).$$

This is known as *maximum a posteriori* (MAP) inference or *energy minimization*.

A number of techniques can be used to find $\mathbf{y}^\star$, including:

- message-passing (dynamic programming)
- integer programming (part 3)
- graph-cuts (part 2)

However, in general, inference is NP-hard.

参考文献

- 林元烈，应用随机过程，清华大学出版社，2002
- Sheldon M. Ross, 龚光鲁译，应用随机过程：概率模型导论，第11版，人民邮电出版社，2016
- Andrew Blake等著，谢昭等译，Markov随机场在视觉和图像处理中的应用，科学出版社，2014
- Stephen Gould, Tutorial titled “Markov Random Fields for Computer Vision” given at the Machine Learning Summer School (MLSS 2011), 13-17 June 2011, Singapore.
  [slides (part 1) | slides (part 2) | slides (part 3)]

内容提要

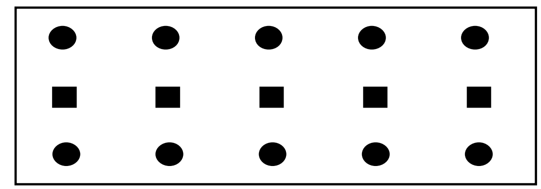
- 图像的矩阵表示与奇异值分解
- 图像的统计描述的基本概念
- 概率论基础 (复习)
- 随机变量和随机过程 (复习)
- 图像的空间描述
- 图像的统计描述
- 随机场的基本概念
- 图像的图表示

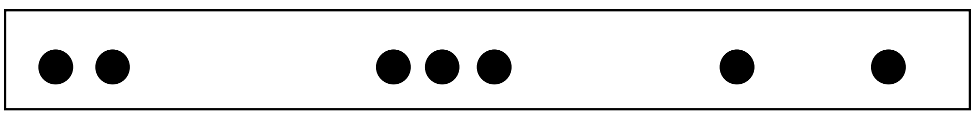
Perceptual Organization in Human Vision (I)

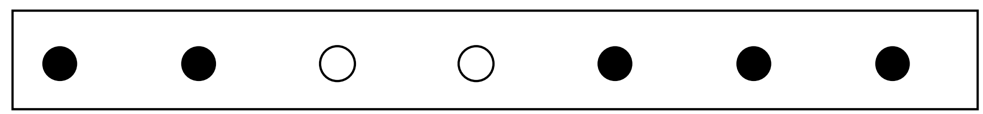
- The Gestalt school of psychology originated in 1920s-1930s in recognition of the role of perceptual organization in human vision.

- Gestalt: a German word meaning "form" or "whole"

- Some basic principles of perceptual organization

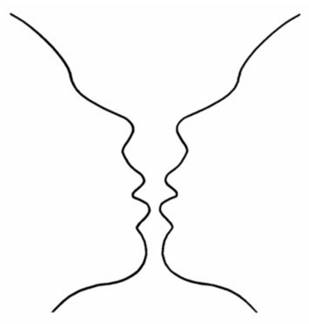


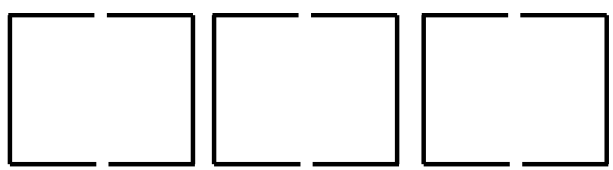


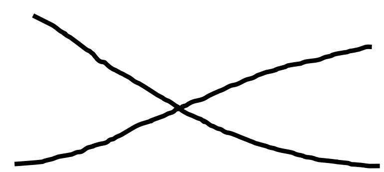


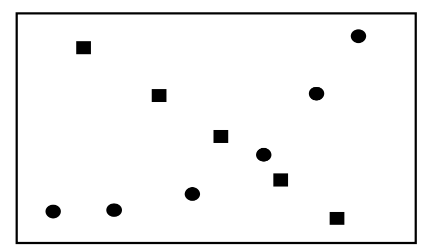
proximity

similarity









symmetry

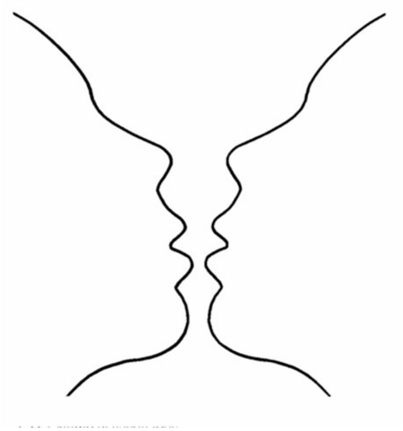
closure

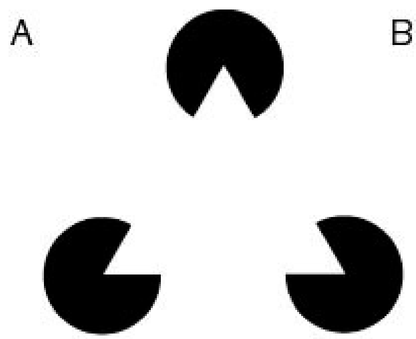
continuity

Perceptual Organization in Human Vision (II)

- Some basic questions

  - Example: The law of closure





计算的内容是什么?
如何计算?

Basic Concepts of Graphs

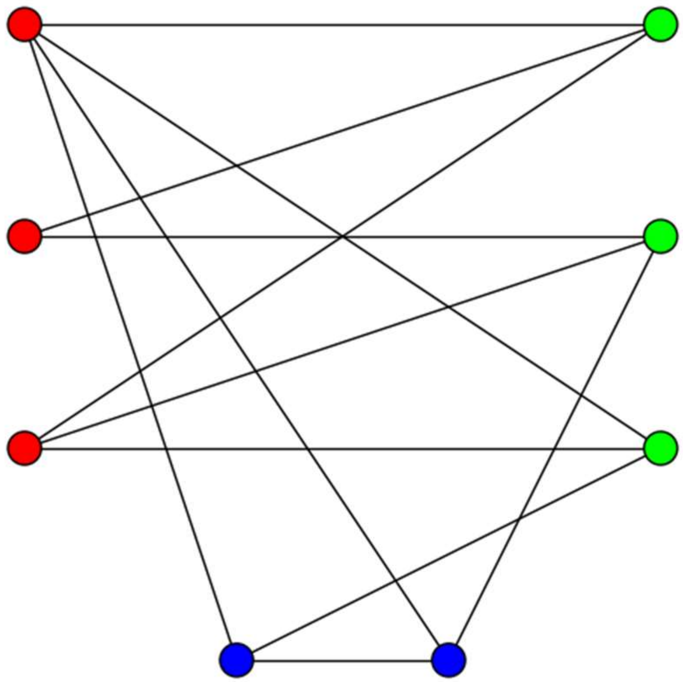
- A graph G = (V, E) is an ordered pair that consists of a set of vertices and a set of edges that connect pairs of distinct vertices.

- A graph is weighted if a number is assigned to each edge.

- Directed graph versus undirected graph.

- A bipartite graph is a graph whose vertices can be divided into two sets such that all edges connect a vertex in one set with a vertex in the other set.

Example of a Graph that is not Bipartite

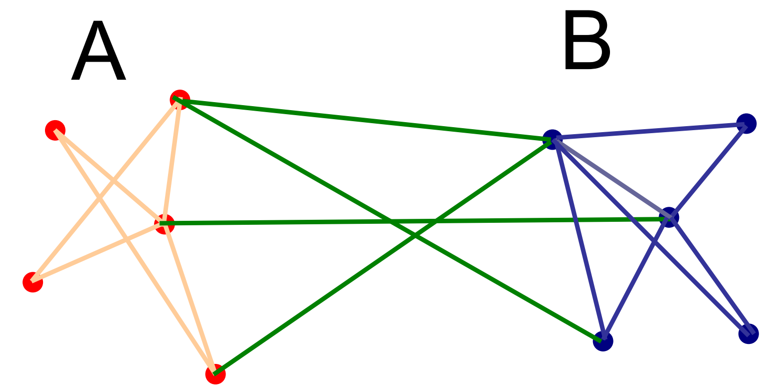


http://en.wikipedia.org/wiki/Bipartite_graph

Basic Concept of Graph Cuts

- A graph G = (V, E) can be partitioned into two disjoint sets A, B

$$A \cup B = V \qquad A \cap B = \varnothing$$



Shi & Malik, PAMI, 22:888-905, 2000

- Each vertex represents a pixel within the image.

- The weight of the edge connecting two vertices represents their similarity.

Questions?

ControlNet

Adding Conditional Control to Text-to-Image Diffusion Models

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala
Stanford University
{lvmin, anyirao, maneesh}@cs.stanford.edu

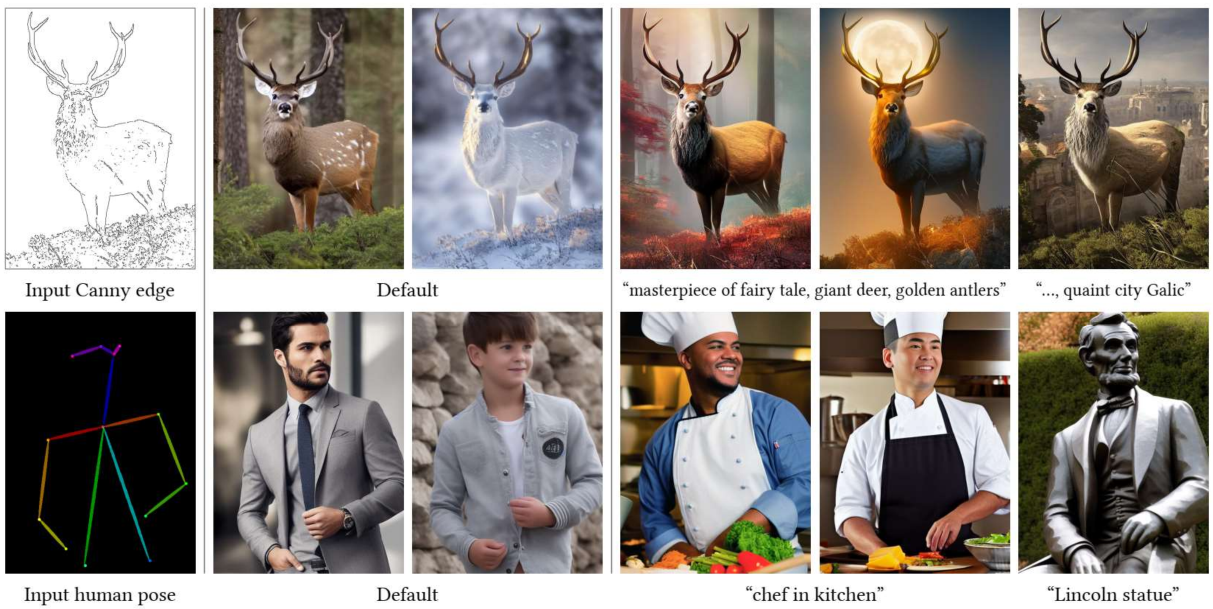


Figure 1: Controlling Stable Diffusion with learned conditions. ControlNet allows users to add conditions like Canny edges (top), human pose (bottom), *etc.*, to control the image generation of large pretrained diffusion models. The default results use the prompt "a high-quality, detailed, and professional image". Users can optionally give prompts like the "chef in kitchen".

Stable Diffusion

Stable Diffusion AI Image Generator Features FAQ WebUI Prompt Generator ControlNet Mobile App Prompts

Stable Diffusion Online

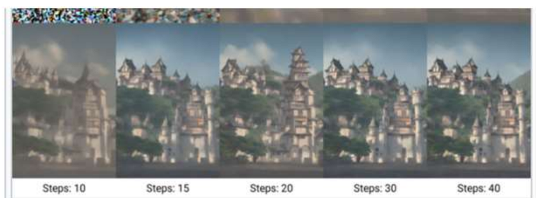
Stable Diffusion XL is a latent text-to-image diffusion model capable of generating photo-realistic images given any text input, cultivates autonomous freedom to produce incredible imagery, empowers billions of people to create stunning art within seconds.

Create beautiful images with our AI Image Generator (Text to Image) for free.

Let Your Creativity Flow. Create Unique AI Generated Images in 1-Click.

Get Started for Free ► Search Prompt

Training data [edit]



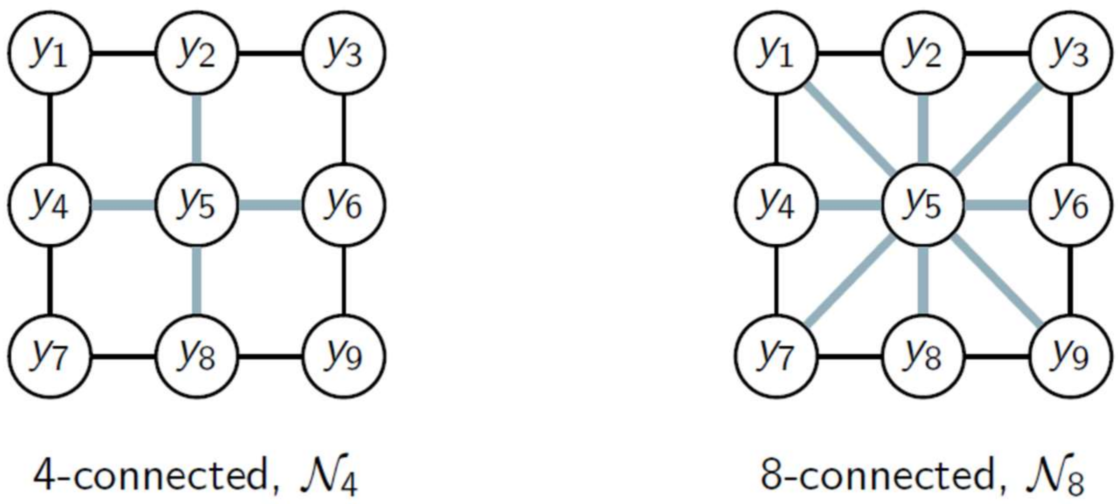
The denoising process used by Stable Diffusion. The model generates images by iteratively denoising random noise until a configured number of steps have been reached, guided by the CLIP text encoder pretrained on concepts along with the attention mechanism, resulting in the desired image depicting a representation of the trained concept.

Stable Diffusion was trained on pairs of images and captions taken from LAION-5B, a publicly available dataset derived from Common Crawl data scraped from the web, where 5 billion image-text pairs were classified based on language and filtered into separate datasets by resolution, a predicted likelihood of containing a watermark, and predicted "aesthetic" score (e.g. subjective visual quality).[18] The dataset was created by LAION, a German non-profit which receives funding from Stability AI.[18][19] The Stable Diffusion model was trained on three subsets of LAION-5B: laion2B-en, laion-high-resolution, and laion-aesthetics v2 5+.[18] A third-party analysis of the model's training data identified that out of a smaller subset of 12 million images taken from the original wider dataset used, approximately 47% of the sample size of images came from 100 different domains, with Pinterest taking up 8.5% of the subset, followed by websites such as WordPress, Blogspot, Flickr, DeviantArt and Wikimedia Commons.[citation needed] An investigation by Bayerischer Rundfunk showed that LAION's datasets, hosted on Hugging Face, contain large amounts of private and sensitive data.[20]

https://stablediffusionweb.com/

马尔可夫随机场的基本概念 (III)

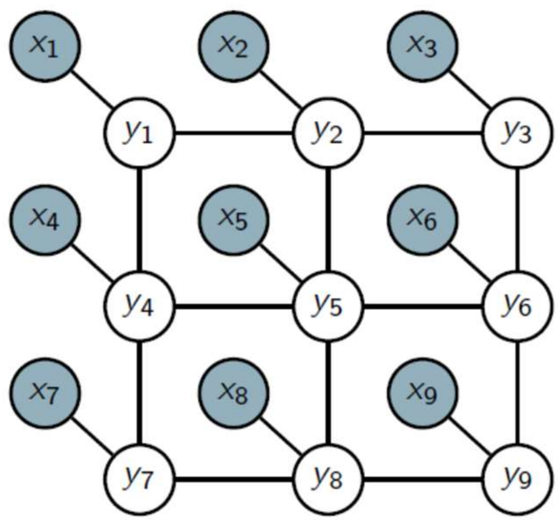
Pixel Neighbourhoods



马尔可夫随机场的基本概念 (II)

Conditional Markov Random Fields

$$
\begin{aligned}
E(\mathbf{y};\mathbf{x}) &= \sum_{c} \psi_c(\mathbf{y}_c;\mathbf{x}) \\
&= \underbrace{\sum_{i\in\mathcal{V}} \psi_i^U(y_i;\mathbf{x})}_{\text{unary}} + \underbrace{\sum_{ij\in\mathcal{E}} \psi_{ij}^P(y_i,y_j;\mathbf{x})}_{\text{pairwise}} + \underbrace{\sum_{c\in\mathcal{C}} \psi_c^H(\mathbf{y}_c;\mathbf{x})}_{\text{higher-order}}.
\end{aligned}
$$



马尔可夫随机场的基本概念 (I)

Energy Functions

Let $\mathbf{x}$ be some observations (i.e., features from the image) and let $\mathbf{y} = (y_1, \ldots, y_n)$ be a vector of random variables. Then we can write the conditional probability of $\mathbf{y}$ given $\mathbf{x}$ as

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\{-E(\mathbf{y}; \mathbf{x})\}$$

where $Z(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{L}^n} \exp\{-E(\mathbf{y}; \mathbf{x})\}$ is called the *partition function*.

The *energy function* $E(\mathbf{y}; \mathbf{x})$ usually has some structured form:

$$E(\mathbf{y}; \mathbf{x}) = \sum_c \psi_c(\mathbf{y}_c; \mathbf{x})$$

where $\psi_c(\mathbf{y}_c; \mathbf{x})$ are *clique potentials* defined over a subset of random variables $\mathbf{y}_c \subseteq \mathbf{y}$.

示例：高斯-马尔可夫模型

$$s=(i,j);\quad G(s)=\{(i,j+1),(i,j-1),(i+1,j),(i-1,j)\}$$

$$p\left(X_s=x_s \mid X_t, t\in G_s\right)=\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(x_{ij}-\frac{1}{4}\left[x_{i,j+1}+x_{i,j-1}+x_{i+1,j}+x_{i-1,j}\right]\right)^2\right]$$

马尔可夫随机场的基本概念 （IV）

Binary MRF Example

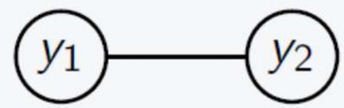
Consider the following energy function for two binary random variables, y_1 and y_2.

$$
\begin{array}{c|c|}
0 & 5 \\
1 & 2
\end{array}
\qquad
\begin{array}{c|c|}
0 & 1 \\
1 & 3
\end{array}
\qquad
\begin{array}{c|c|c|}
 & 0 & 1 \\
0 & 0 & 3 \\
1 & 4 & 0
\end{array}
$$

$$
\begin{aligned}
E(y_1, y_2) &= \psi_1(y_1) + \psi_2(y_2) + \psi_{12}(y_1, y_2) \\
&= \underbrace{5\bar{y}_1 + 2y_1}_{\psi_1} \\
&\quad + \underbrace{\bar{y}_2 + 3y_2}_{\psi_2} \\
&\quad + \underbrace{3\bar{y}_1 y_2 + 4y_1\bar{y}_2}_{\psi_{12}}
\end{aligned}
$$

where $\bar{y}_1 = 1 - y_1$ and $\bar{y}_2 = 1 - y_2$.

Graphical Model



Probability Table

y_1	y_2	E	P
0	0	6	0.244
0	1	11	0.002
1	0	7	0.090
1	1	5	0.664