12 Image Pattern Classification

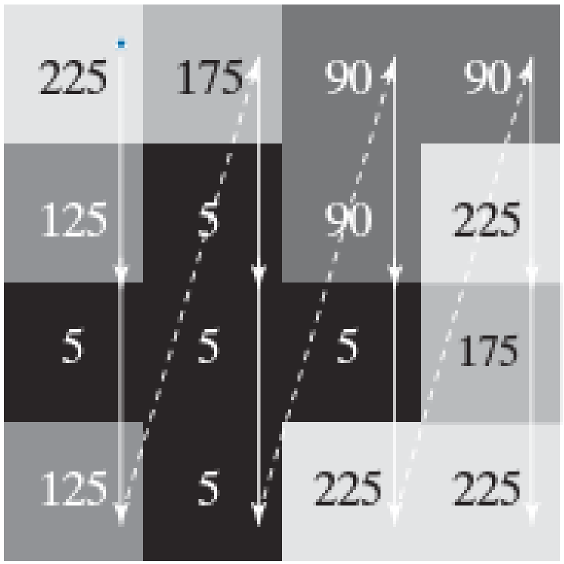
李东晓

lidx@zju.edu.cn

Contents

Patterns and Pattern Classes

- Pattern: a spatial arrangement of features
  - Quantitative Pattern
    - Vector
  - Structural Pattern
    - String
    - Tree
    - graph



$$\mathbf{x} = \begin{bmatrix} 225 \\ 125 \\ 5 \\ \vdots \\ 225 \\ 175 \\ 225 \end{bmatrix}$$

- Pattern class: a family of patterns that share some common properties

$$\omega_1, \omega_2, \ldots, \omega_W,$$

where W is the number of classes.

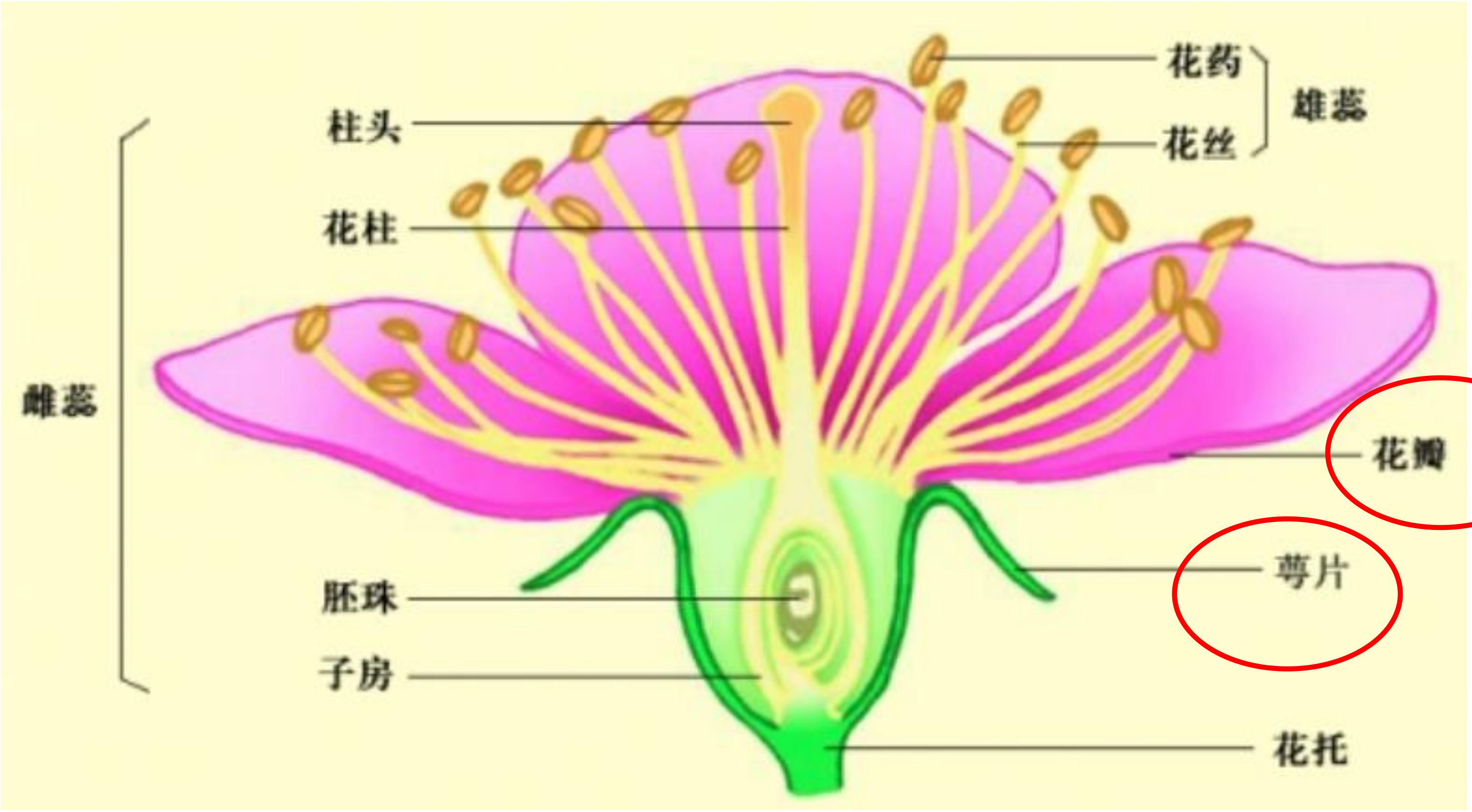
Pattern Classification = Recognition

- Patterns
  - Labeled
  - Unlabeled
- Datasets
  - Training set
  - Validation set
  - Test set
- Machine learning
  - Supervised ←→ Labeled data
  - Unsupervised ←→ Unlabeled data

Example

- Iris（鸢尾花卉）数据集，由Fisher于1936年收集整理，用于多重变量分析实验。包含150个数据，分为3类（Setosa，Versicolour，Virginica），每类50个数据，每个数据包含4个属性：花萼长度、花萼宽度、花瓣长度、花瓣宽度。
- 其它比较流行的数据集还有Adult，Wine，Car Evaluation等。

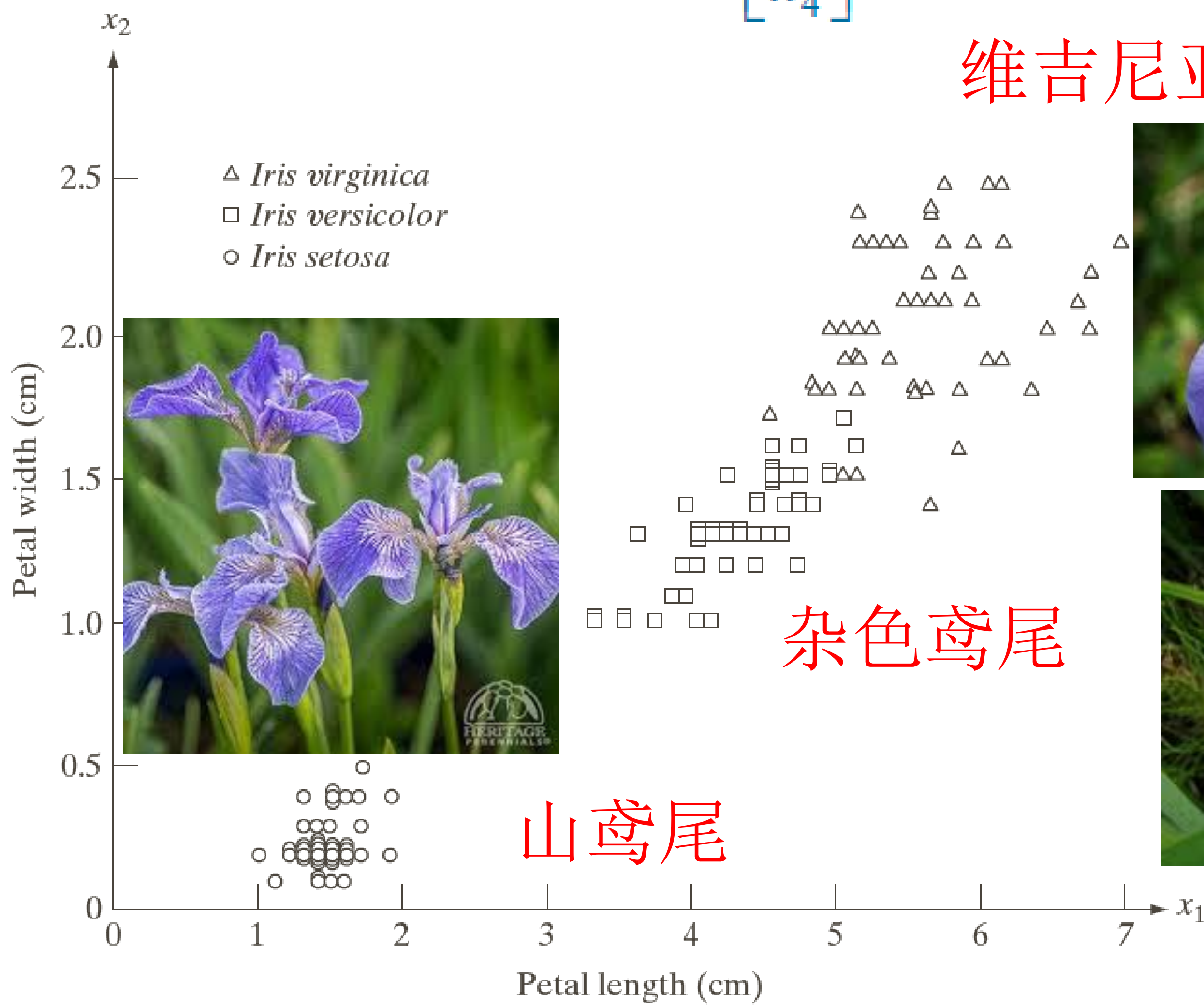
花的构造



- Pattern vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

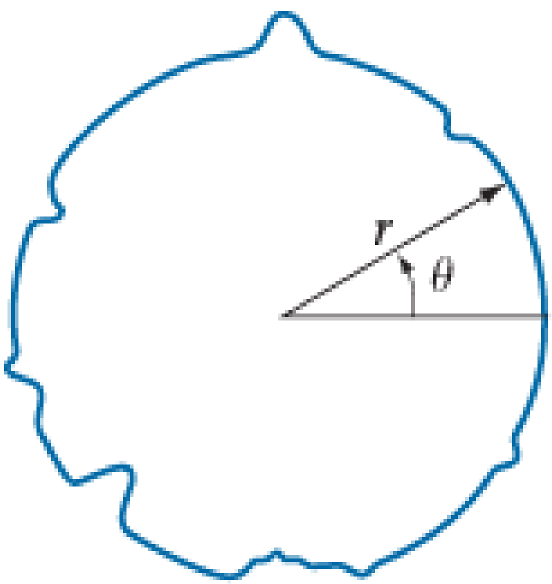
x_1 = Petal width
x_2 = Petal length
x_3 = Sepal width
x_4 = Sepal length

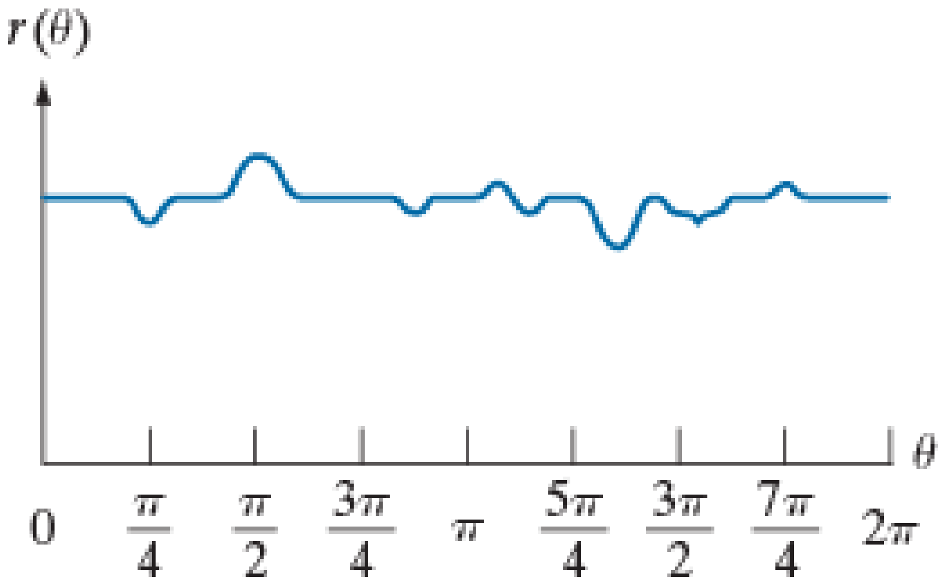


维吉尼亚鸢尾

Example of Pattern Vector

- A noisy object and its signature
  - Sampling $x_1 = r(\theta_1), x_2 = r(\theta_2), \ldots, x_n = r(\theta_n)$
  - Statistical moments





$$\mathbf{x} = \begin{bmatrix} g(r(\theta_1)) \\ g(r(\theta_2)) \\ \vdots \\ g(r(\theta_n)) \end{bmatrix}$$

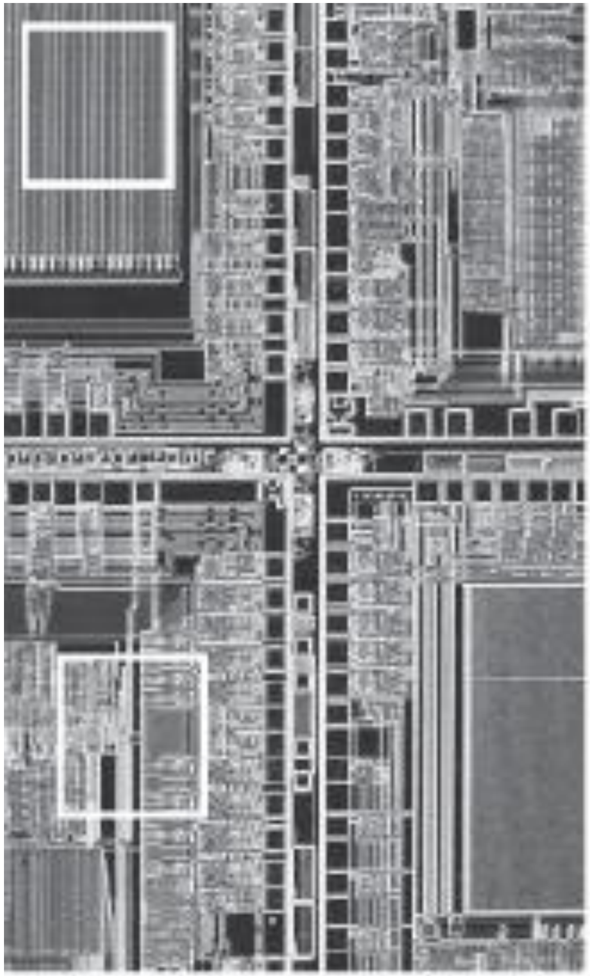
Example of Pattern Vector

Boundary and regional features

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \begin{aligned} x_1 &= \text{compactness} \\ x_2 &= \text{circularity} \\ x_3 &= \text{eccentricity} \end{aligned}$$

Example of Pattern Vector

Texture features

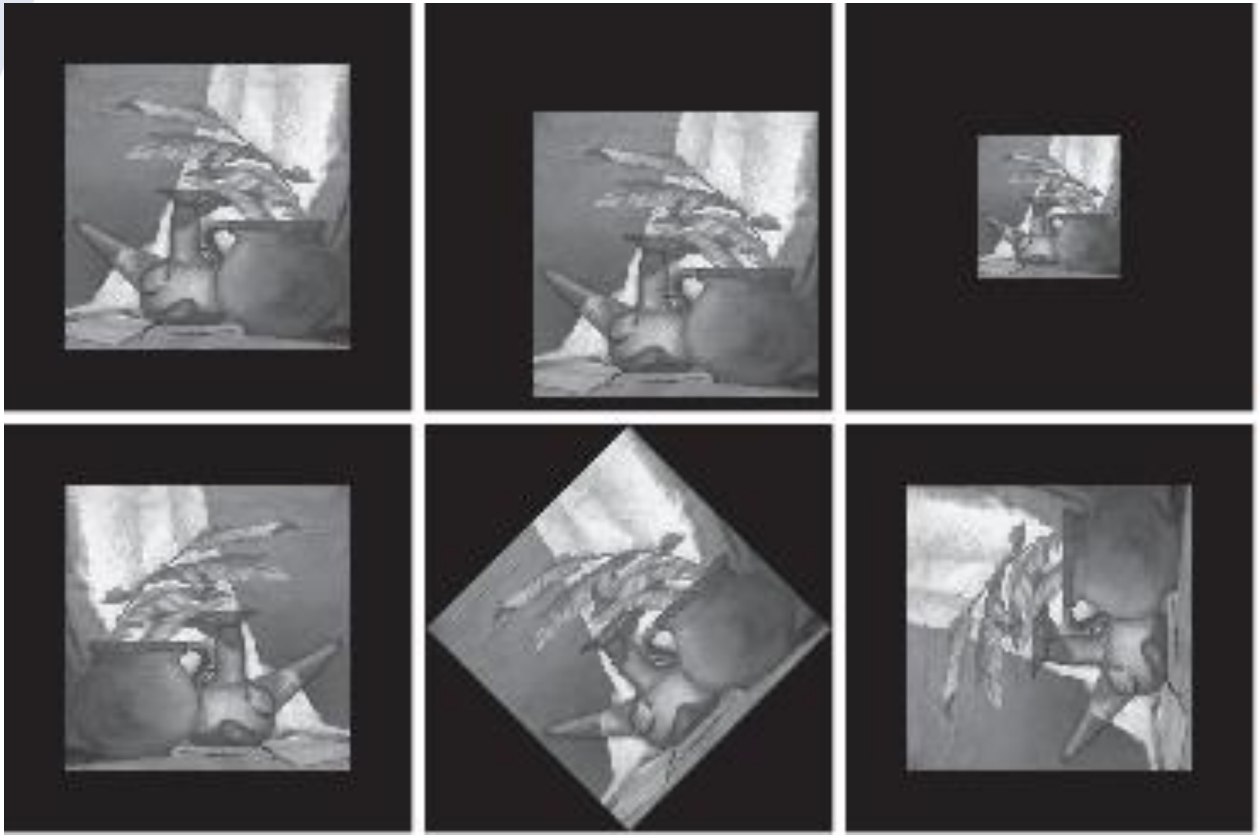


$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$$

x_1 = max probability
x_2 = correlation
x_3 = contrast
x_4 = uniformity
x_5 = homogeneity
x_6 = entropy

Example of Pattern Vector

Moment invariants



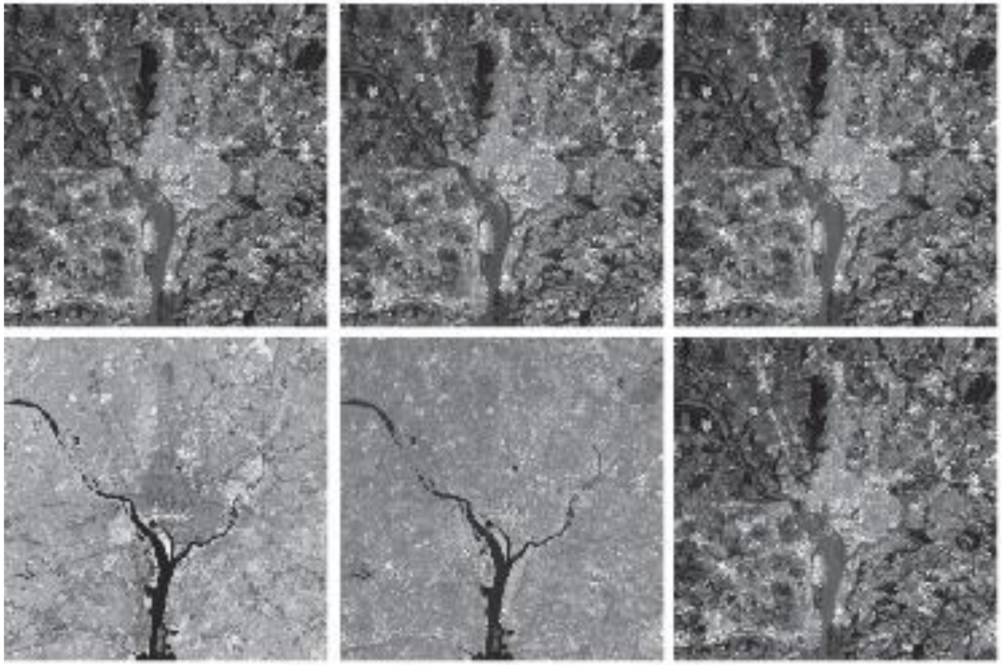
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \phi_5 \\ \phi_6 \\ \phi_7 \end{bmatrix}$$

The ϕ's are moment invariants

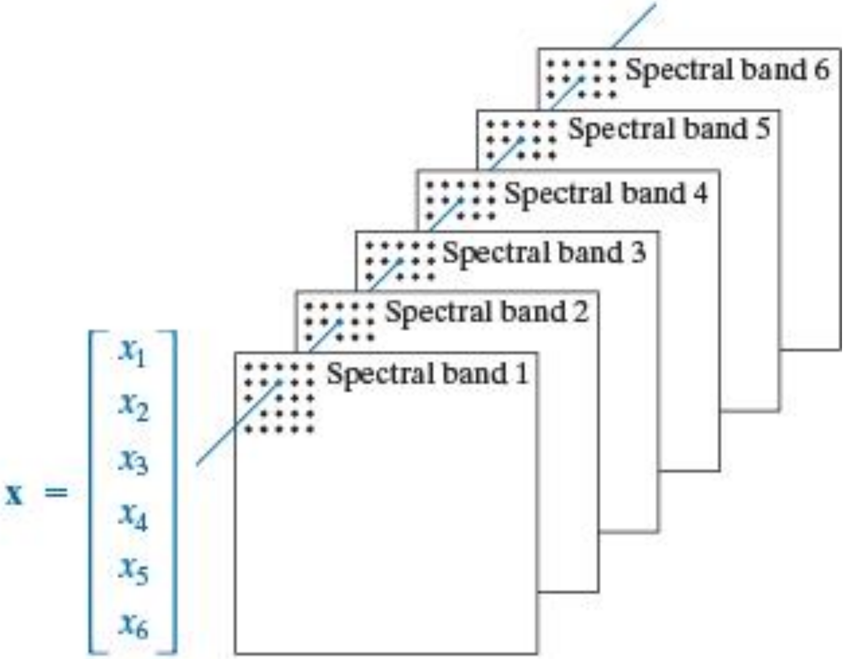
Example of Pattern Vector

A set of registered images

Images in spectral bands 1–3

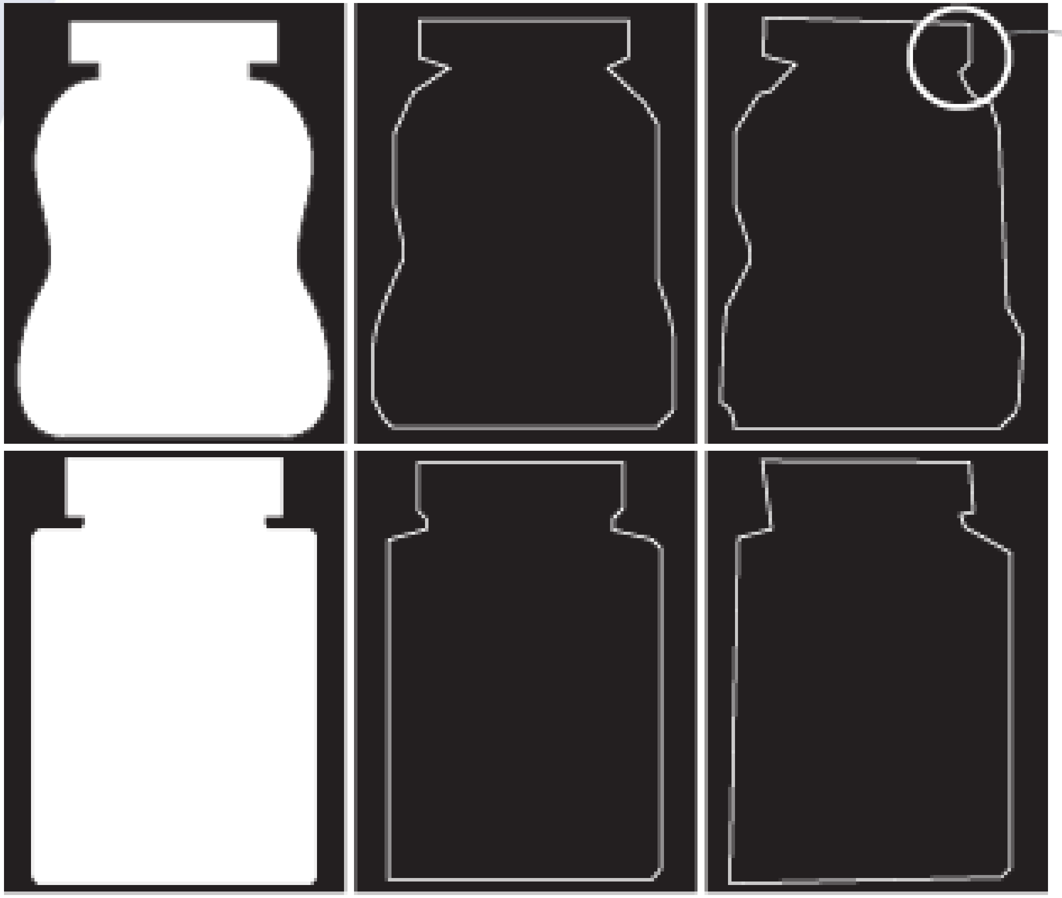


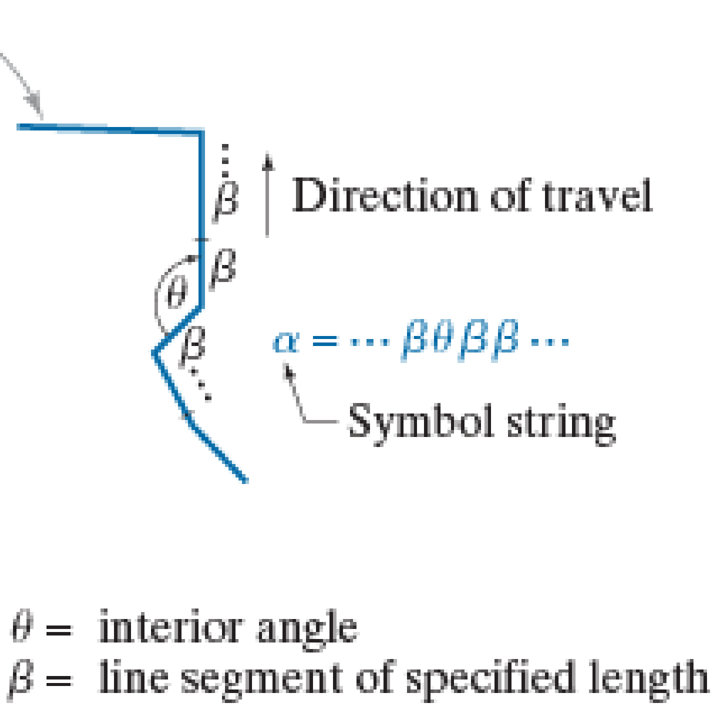
Images in spectral bands 4–6



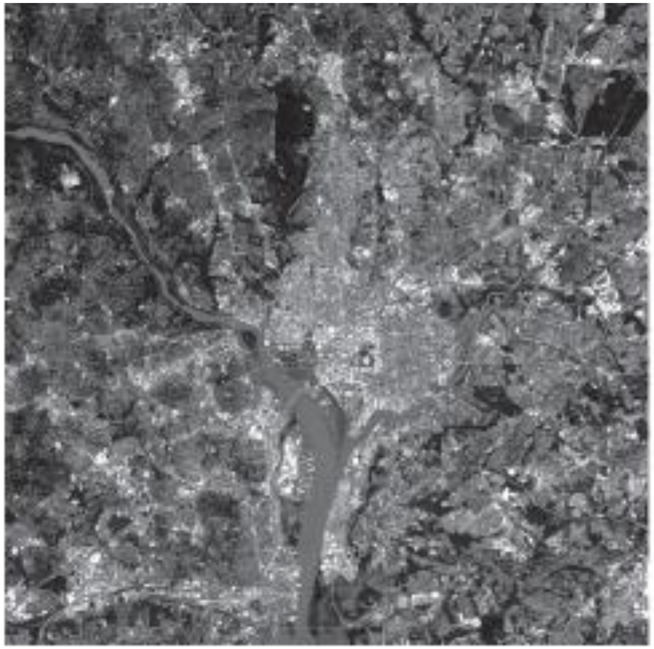
Example of Pattern String

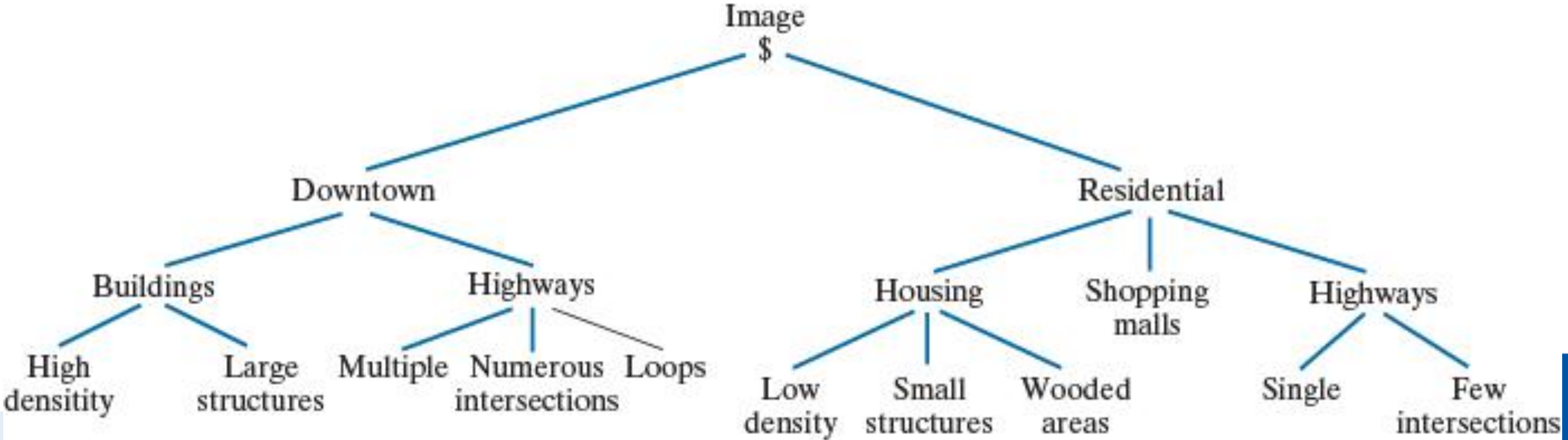
String of symbols





Example of Pattern Tree





Contents

Decision-Theoretic Pattern Recognition

- Find *W* decision (discriminant) functions

$d_1(\mathbf{x}), d_2(\mathbf{x}), \ldots, d_W(\mathbf{x})$ If $\mathbf{x} \in \omega_i$, then

$$d_i(\mathbf{x}) > d_j(\mathbf{x}) \qquad j = 1, 2, \ldots, W; j \neq i$$

- Decision boundary

$$d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0$$

- $d_{ij}(\mathbf{x}) > 0 \Rightarrow \mathbf{x} \in \omega_i$
- $d_{ij}(\mathbf{x}) < 0 \Rightarrow \mathbf{x} \in \omega_j$

Pattern Classification by Prototype Matching

- Each class ←→ a prototype pattern vector
- Pattern recognition ← closest class
  - Minimum distance classifier
  - Correlation

Minimum Distance Classifier

- Example prototype of each pattern class

$$\mathbf{m}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in \omega_j} \mathbf{x}_j \qquad j = 1, 2, \ldots, W$$

- Euclidean distance $\|\mathbf{a}\| = (\mathbf{a}^T \mathbf{a})^{1/2}$

$$D_j(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_j\| \qquad j = 1, 2, \ldots, W$$

- Decision function

$$d_j(\mathbf{x}) = \mathbf{x}^T \mathbf{m}_j - \frac{1}{2} \mathbf{m}_j^T \mathbf{m}_j \qquad j = 1, 2, \ldots, W$$

Minimum Distance Classifier

- Decision boundary

$$d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x})$$

$$= \mathbf{x}^T(\mathbf{m}_i - \mathbf{m}_j) - \frac{1}{2}(\mathbf{m}_i - \mathbf{m}_j)^T(\mathbf{m}_i + \mathbf{m}_j) = 0$$

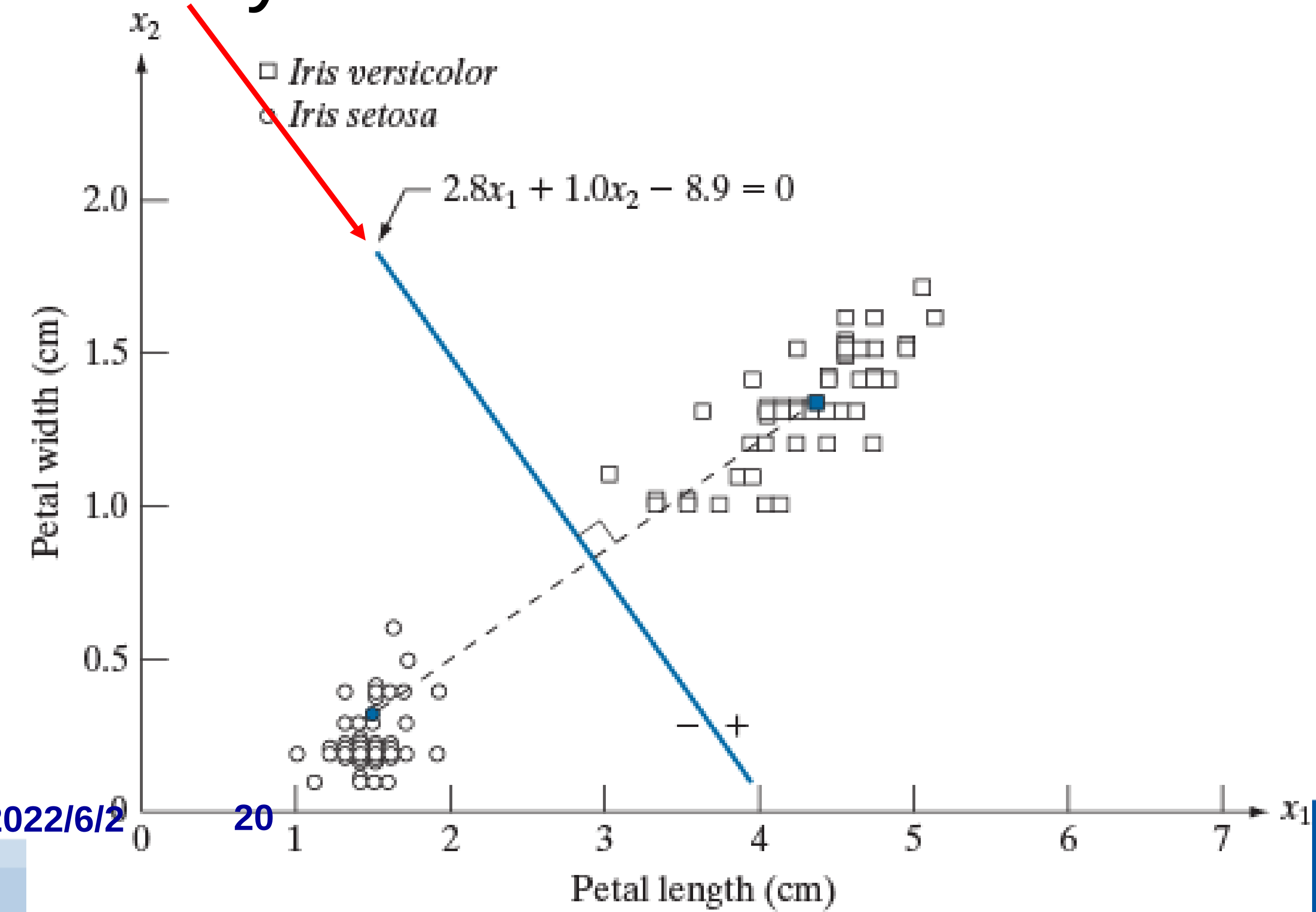
  - $n = 2$: line
  - $n = 3$: plane
  - $n > 3$: hyperplane

Example of Minimum Distance Classifier

- Mean vectors

$$\mathbf{m}_1 = (4.3, 1.3)^T \text{ and } \mathbf{m}_2 = (1.5, 0.3)^T$$

- Decision boundary



Example

- BoA Check

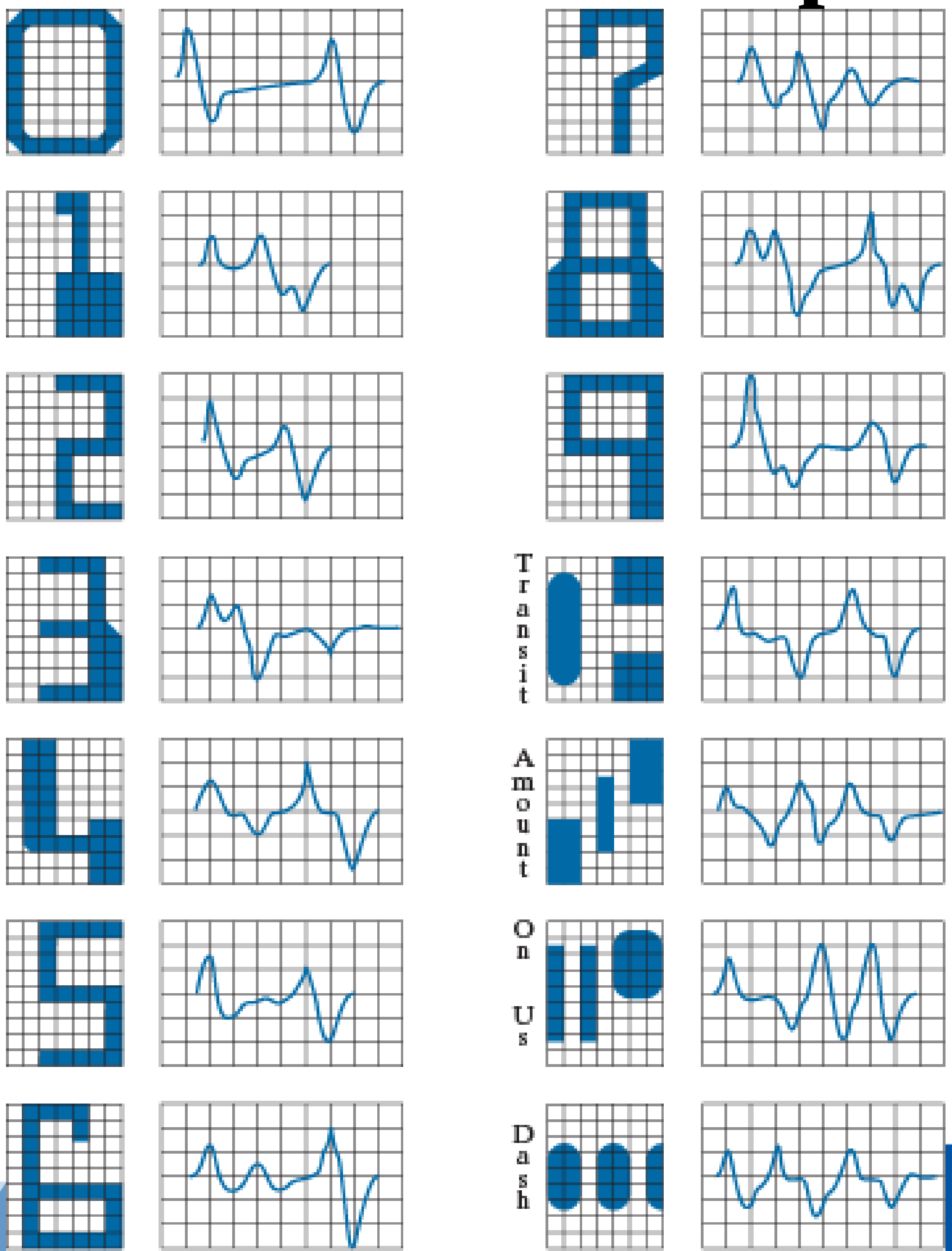
Example

- American Bankers Association E-13B font character set
- Horizontal scan
  - 9x7 grid
  - Magnetic ink
  - Single-slit reading head

→ signature waveforms

  - increase/decrease rate of character area

→ prototype pattern vector



Matching by correlation

- Correlation

$$c(x, y) = \sum_s \sum_t w(s, t) f(x + s, y + t)$$

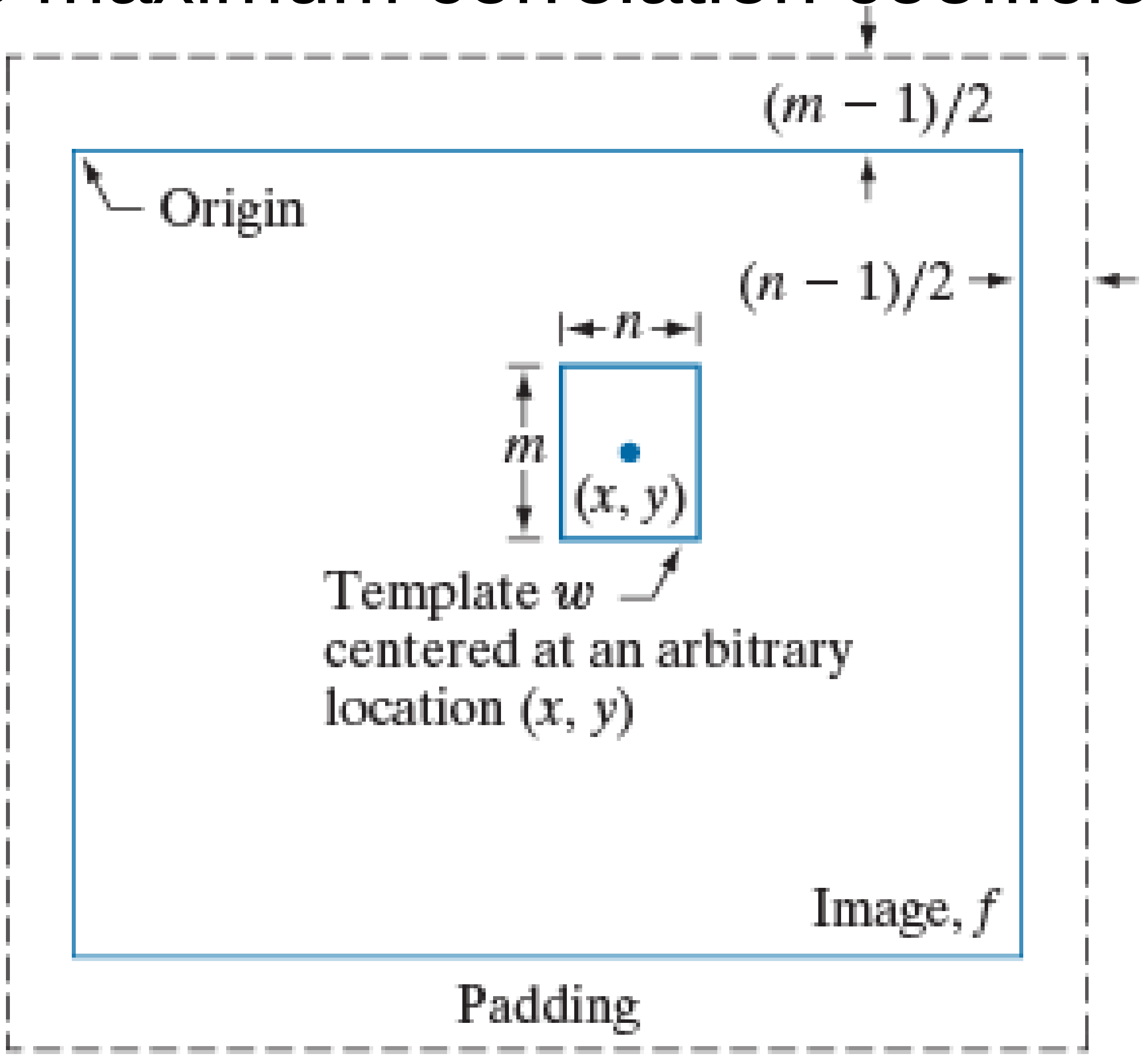
$$f(x, y) \star w(x, y) \Leftrightarrow F^*(u, v) W(u, v)$$

- Normalized correlation coefficient

$$\gamma(x, y) = \frac{\sum_s \sum_t \left[w(s, t) - \overline{w}\right] \times \left[f(x + s, y + t) - \overline{f}(x + s, y + t)\right]}{\left\{\sum_s \sum_t \left[w(s, t) - \overline{w}\right]^2 \sum_s \sum_t \left[f(x + s, y + t) - \overline{f}(x + s, y + t)\right]^2\right\}^{\frac{1}{2}}}$$

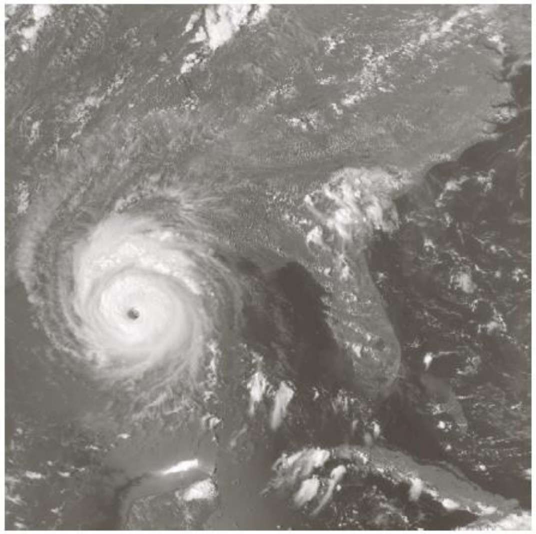
Mechanics of template matching

- Find the maximum correlation coefficient



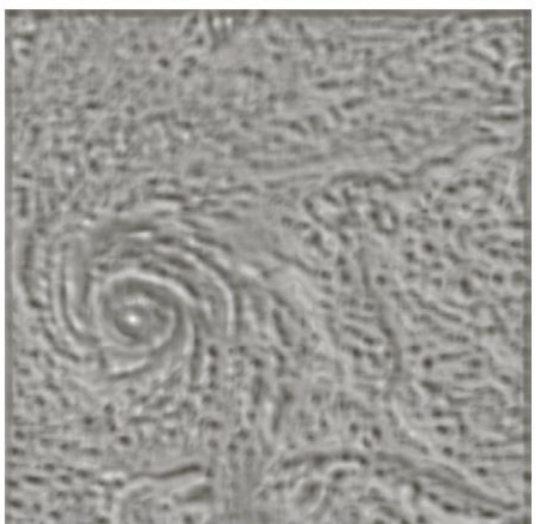
Example

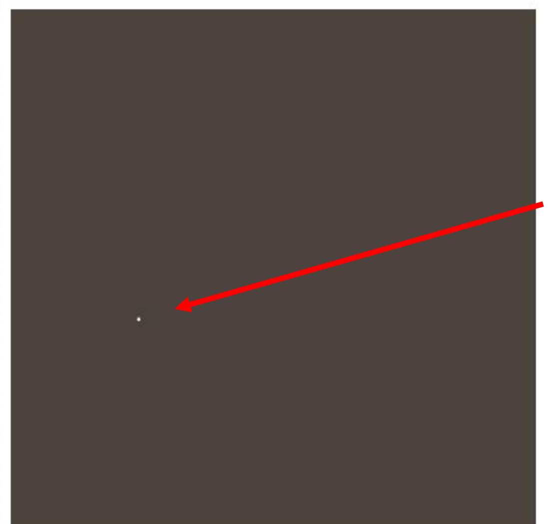
Hurricane
913x913



Template of the
eye of the storm
31x31

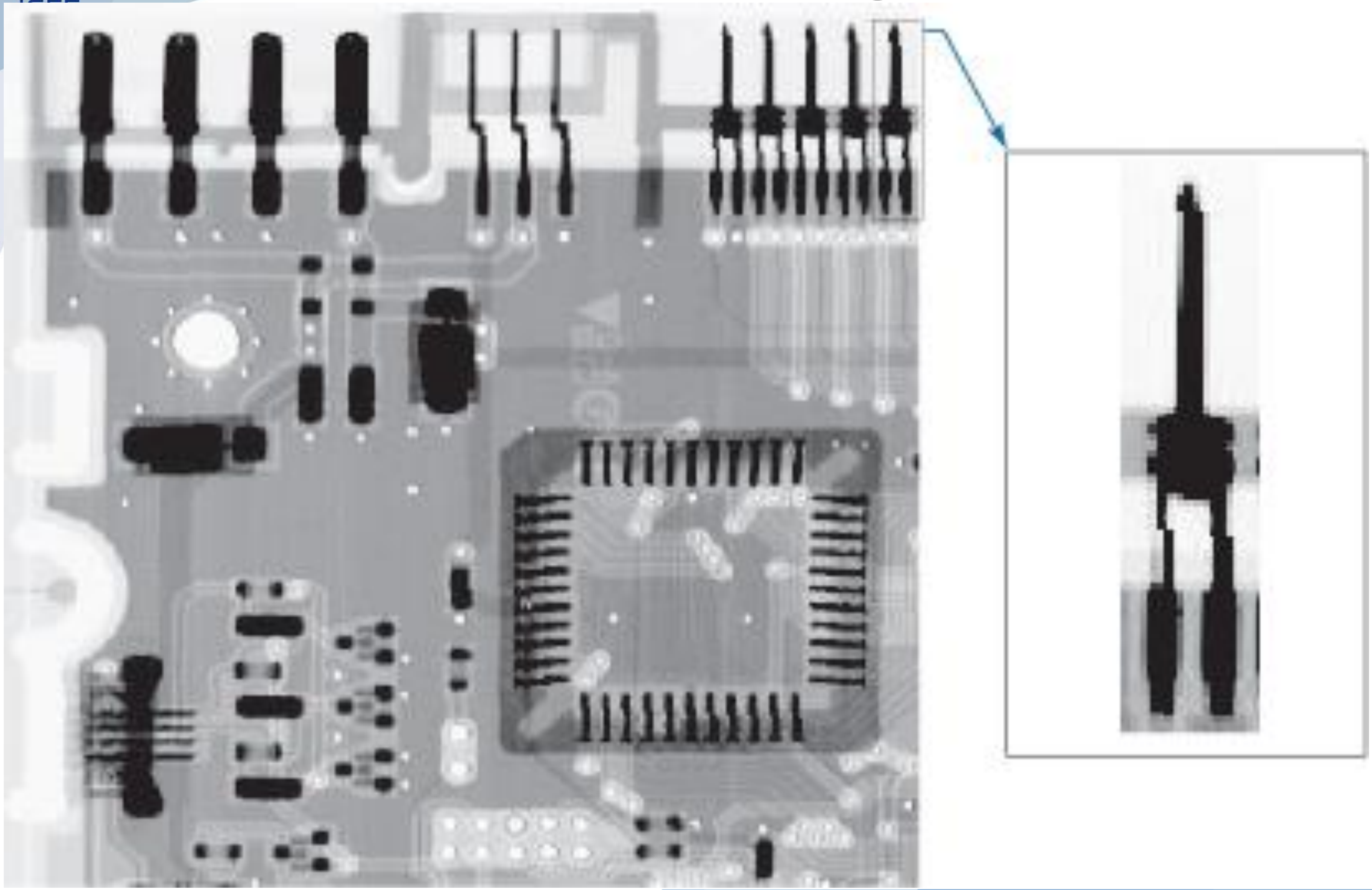
Correlation
coefficient



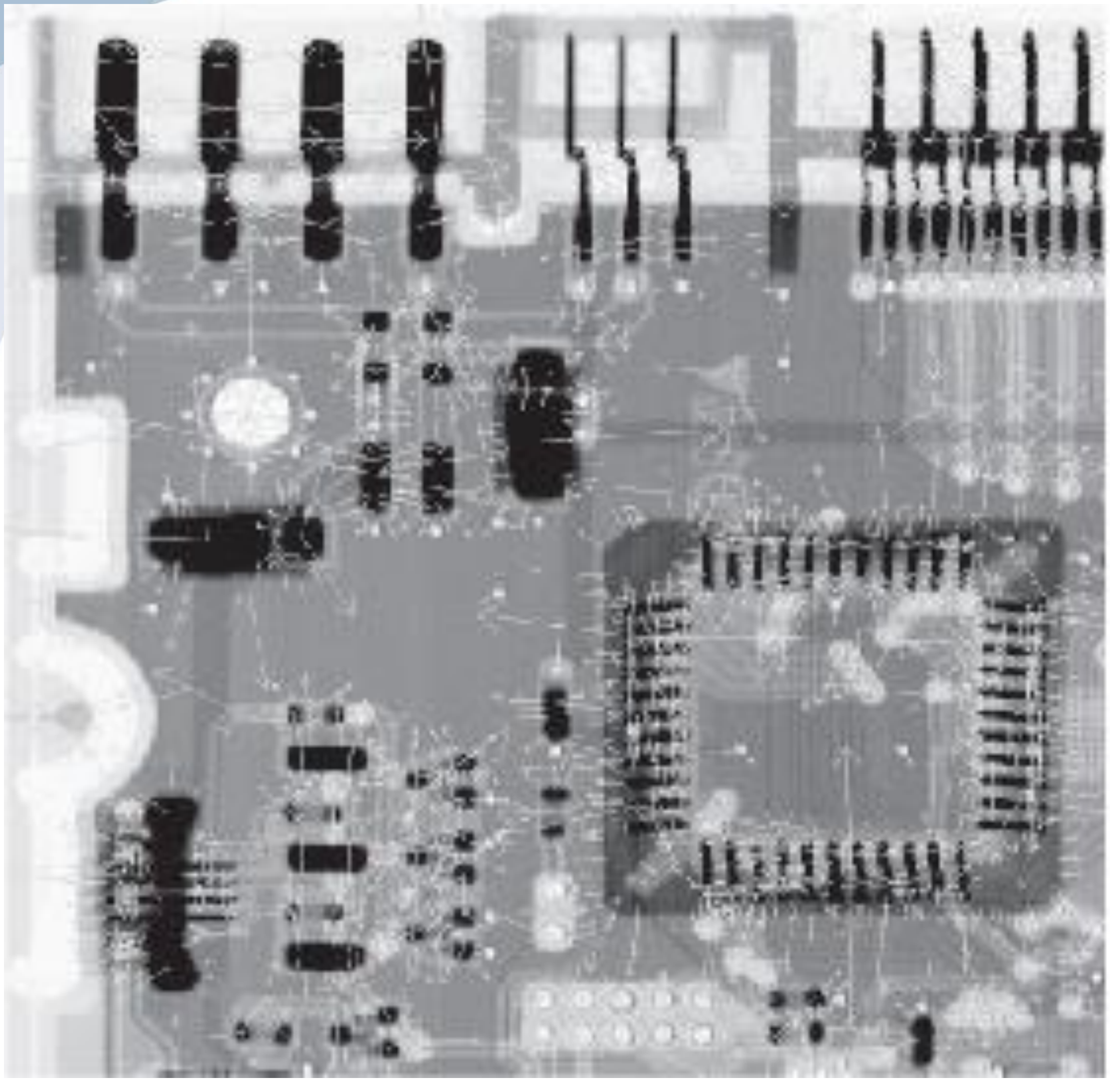


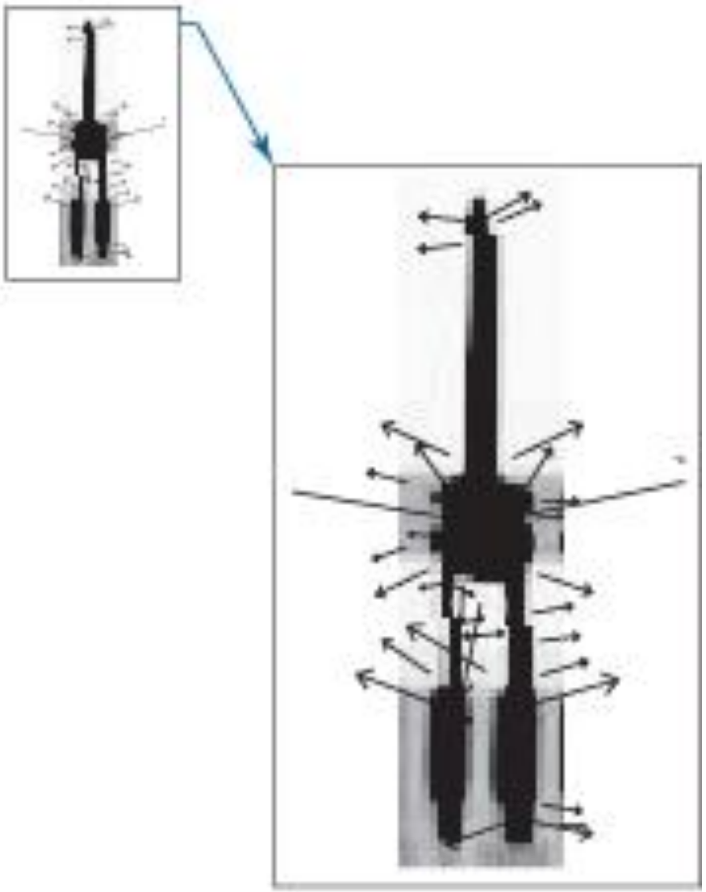
Location of
the best match

Matching SIFT features

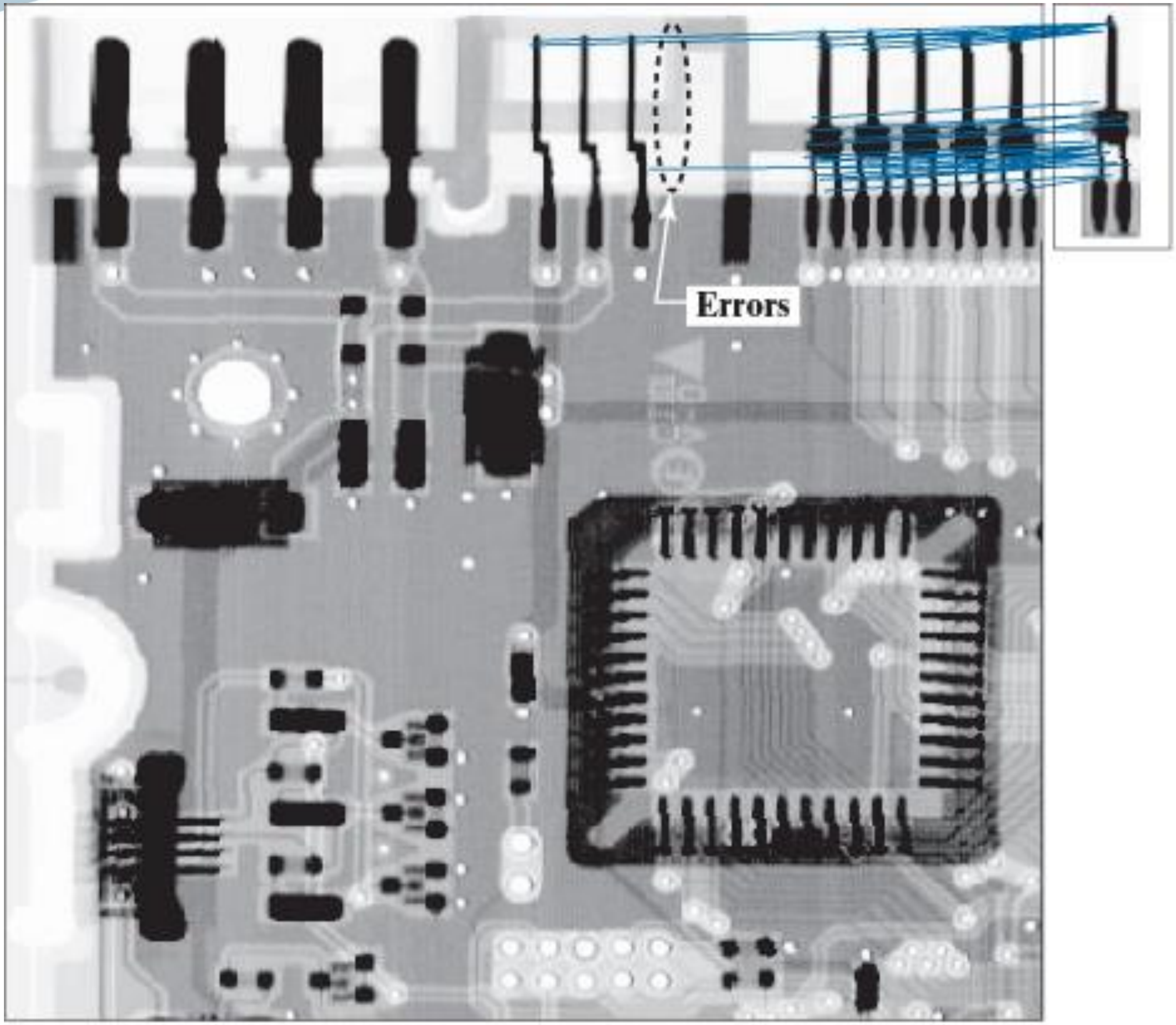


Matching SIFT features





Matching SIFT features



Matching Structural Protypes

- Shape number: 组成最小整数的差分链码
- Degree of similarity, *k*

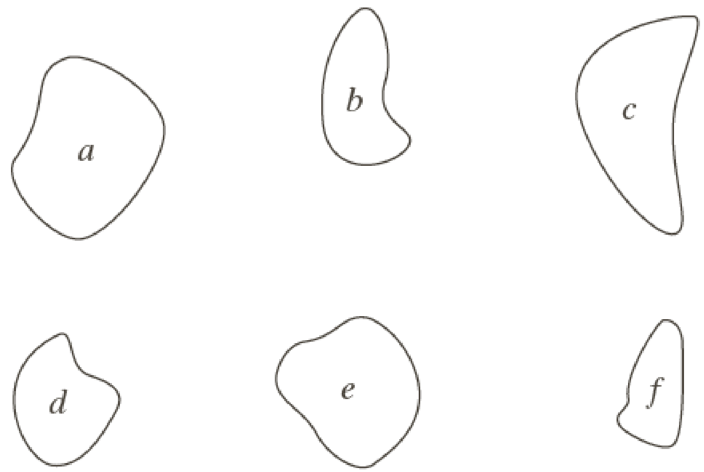
$$s_j(a) = s_j(b) \qquad \text{for } j = 4, 6, 8, \ldots, k$$

$$s_j(a) \neq s_j(b) \qquad \text{for } j = k + 2, k + 4, \ldots$$

- Distance

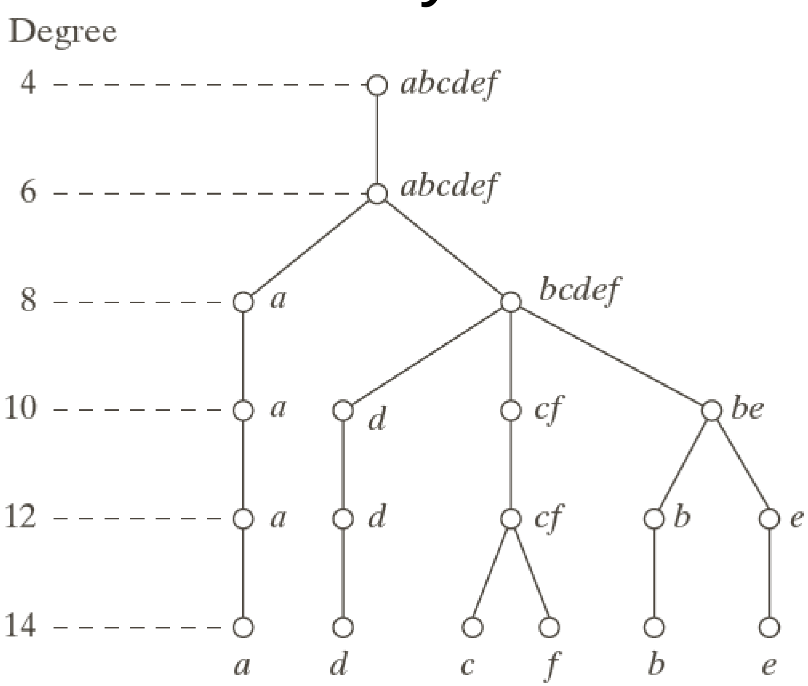
$$D(a, b) = \frac{1}{k}$$

Example



Shapes

Similarity tree



Similarity matrix

	a	b	c	d	e	f
a	∞	6	6	6	6	6
b		∞	8	8	10	8
c			∞	8	8	12
d				∞	8	8
e					∞	8
f						∞

String Matching

- α: number of matches
- β: number of mismatches

$$\beta = \max(|a|, |b|) - \alpha$$

- Similarity measure

$$R = \frac{\alpha}{\beta} = \frac{\alpha}{\max(|a|, |b|) - \alpha}$$

Sample boundaries

$$\alpha_1: 0^\circ < \theta \le 45^\circ;\ \alpha_2: 45^\circ < \theta \le 90^\circ;\ \ldots;\ \alpha_8: 315^\circ < \theta \le 360^\circ$$

Polygonal approximations

Object 1

R	1.a	1.b	1.c	1.d	1.e	1.f
1.a	∞					
1.b	16.0	∞				
1.c	9.6	26.3	∞			
1.d	5.1	8.1	10.3	∞		
1.e	4.7	7.2	10.3	14.2	∞	
1.f	4.7	7.2	10.3	8.4	23.7	∞

Object 2

R	2.a	2.b	2.c	2.d	2.e	2.f
2.a	∞					
2.b	33.5	∞				
2.c	4.8	5.8	∞			
2.d	3.6	4.2	19.3	∞		
2.e	2.8	3.3	9.2	18.3	∞	
2.f	2.6	3.0	7.7	13.5	27.0	∞

6 samples

R	1.a	1.b	1.c	1.d	1.e	1.f
2.a	1.24	1.50	1.32	1.47	1.55	1.48
2.b	1.18	1.43	1.32	1.47	1.55	1.48
2.c	1.02	1.18	1.19	1.32	1.39	1.48
2.d	1.02	1.18	1.19	1.32	1.29	1.40
2.e	0.93	1.07	1.08	1.19	1.24	1.25
2.f	0.89	1.02	1.02	1.24	1.22	1.18

Object 1 vs Object 2

Contents

- Patterns and Pattern Classes
- Pattern Classification by Prototype Matching
- Optimum (Bayes) Statistical Classifiers
- Neural Networks and Deep Learning
- Deep Convolutional Neural Networks
- Some Additional Details of Implementation

Optimum Statistical Classifiers

- Average loss of $\mathbf{x} \leftarrow \omega_j$

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} L_{kj} p(\omega_k/\mathbf{x})$$

$$p(A/B) = [p(A)p(B/A)]/p(B)$$

$$r_j(\mathbf{x}) = \frac{1}{p(\mathbf{x})} \sum_{k=1}^{W} L_{kj} p(\mathbf{x}/\omega_k) P(\omega_k)$$

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} L_{kj} p(\mathbf{x}/\omega_k) P(\omega_k)$$

Optimum Statistical Classifiers

- Bayes classifier

$$\mathbf{x} \leftarrow \omega_i \quad \text{if } r_i(\mathbf{x}) < r_j(\mathbf{x}) \text{ for } j = 1, 2, \ldots, W; j \neq i.$$

- Assume $\boxed{L_{ij} = 1 - \delta_{ij}}$

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} (1 - \delta_{kj}) p(\mathbf{x}/\omega_k) P(\omega_k)$$

$$= p(\mathbf{x}) - p(\mathbf{x}/\omega_j) P(\omega_j)$$

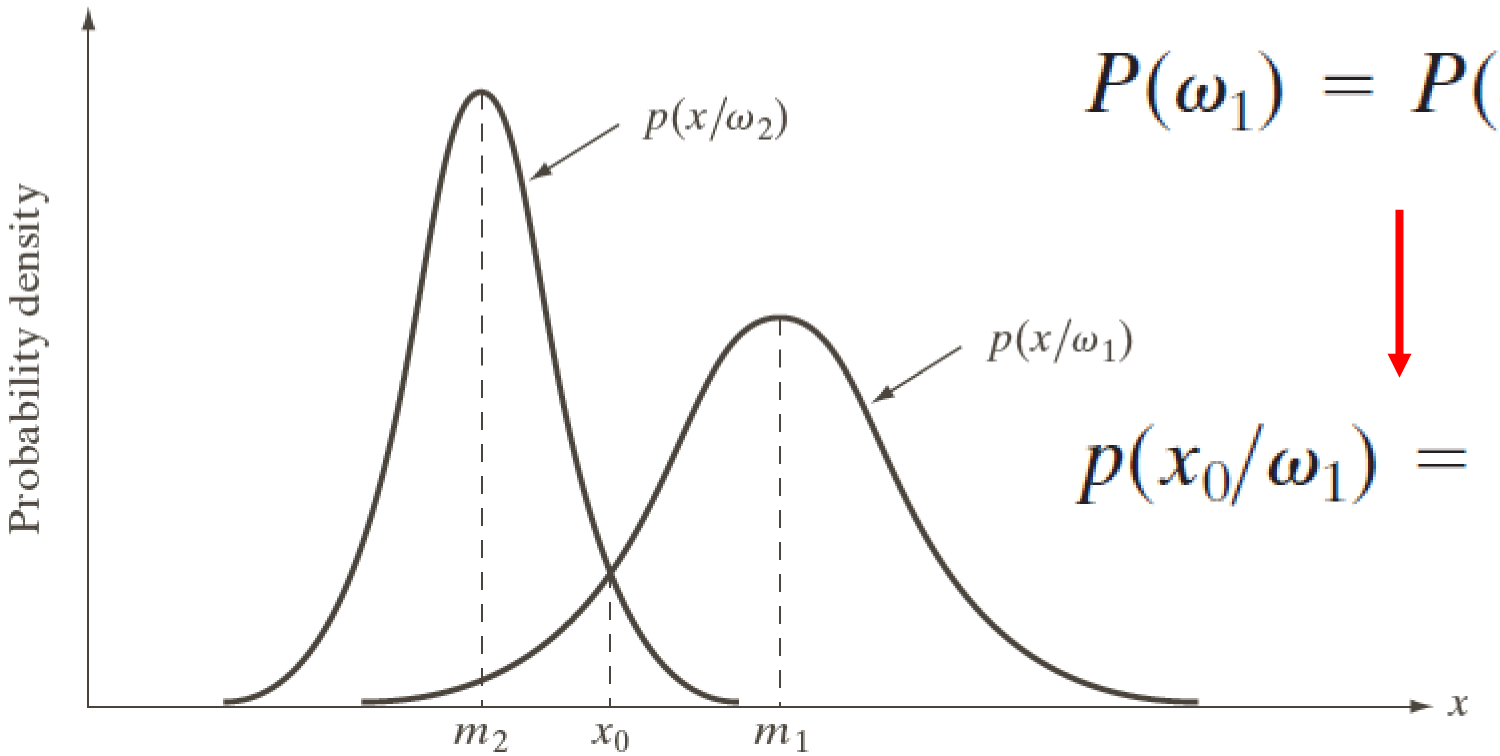
$$p(\mathbf{x}/\omega_i) P(\omega_i) > p(\mathbf{x}/\omega_j) P(\omega_j)$$

$$d_j(\mathbf{x}) = p(\mathbf{x}/\omega_j) P(\omega_j)$$

Bayes Classifier for Gaussian Pattern Classes

- Bayes decision function

$$d_j(x) = \boxed{p(x/\omega_j)P(\omega_j)} = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-m_j)^2}{2\sigma_j^2}} P(\omega_j) \qquad j = 1, 2$$



$$P(\omega_1) = P(\omega_2) = 1/2$$

$$p(x_0/\omega_1) = p(x_0/\omega_2)$$

- Extension to the *n*-dimensional case

$$p(\mathbf{x}/\omega_j) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}_j|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_j)^T\mathbf{C}_j^{-1}(\mathbf{x}-\mathbf{m}_j)}$$

$$\mathbf{m}_j = E_j\{\mathbf{x}\} = \frac{1}{N_j}\sum_{\mathbf{x}\in\omega_j}\mathbf{x}$$

$$\mathbf{C}_j = E_j\{(\mathbf{x}-\mathbf{m}_j)(\mathbf{x}-\mathbf{m}_j)^T\} = \frac{1}{N_j}\sum_{\mathbf{x}\in\omega_j}\mathbf{x}\mathbf{x}^T - \mathbf{m}_j\mathbf{m}_j^T$$

$$\begin{aligned} d_j(\mathbf{x}) &= \ln\left[p(\mathbf{x}/\omega_j)P(\omega_j)\right] \\ &= \ln p(\mathbf{x}/\omega_j) + \ln P(\omega_j) \end{aligned}$$

Bayes decision functions for Gaussian pattern classes with 0-1 loss function

$$p(\mathbf{x}/\omega_j) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}_j|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_j)^T \mathbf{C}_j^{-1}(\mathbf{x}-\mathbf{m}_j)}$$

$$d_j(\mathbf{x}) = \ln P(\omega_j) \boxed{- \frac{n}{2}\ln 2\pi} - \frac{1}{2}\ln|\mathbf{C}_j| - \frac{1}{2}\left[(\mathbf{x} - \mathbf{m}_j)^T \mathbf{C}_j^{-1}(\mathbf{x} - \mathbf{m}_j)\right]$$

$$d_j(\mathbf{x}) = \ln P(\omega_j) - \frac{1}{2}\ln|\mathbf{C}_j| - \frac{1}{2}\left[(\mathbf{x} - \mathbf{m}_j)^T \mathbf{C}_j^{-1}(\mathbf{x} - \mathbf{m}_j)\right]$$

$$\mathbf{C}_j = \mathbf{C}, \text{ for } j = 1, 2, \ldots, W$$

$$d_j(\mathbf{x}) = \ln P(\omega_j) + \mathbf{x}^T\mathbf{C}^{-1}\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{C}^{-1}\mathbf{m}_j$$

Bayes decision functions for Gaussian pattern classes with 0-1 loss function

$$\mathbf{C} = \mathbf{I} \qquad P(\omega_j) = 1/W$$

$$d_j(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{m}_j \qquad j = 1, 2, \ldots, W$$

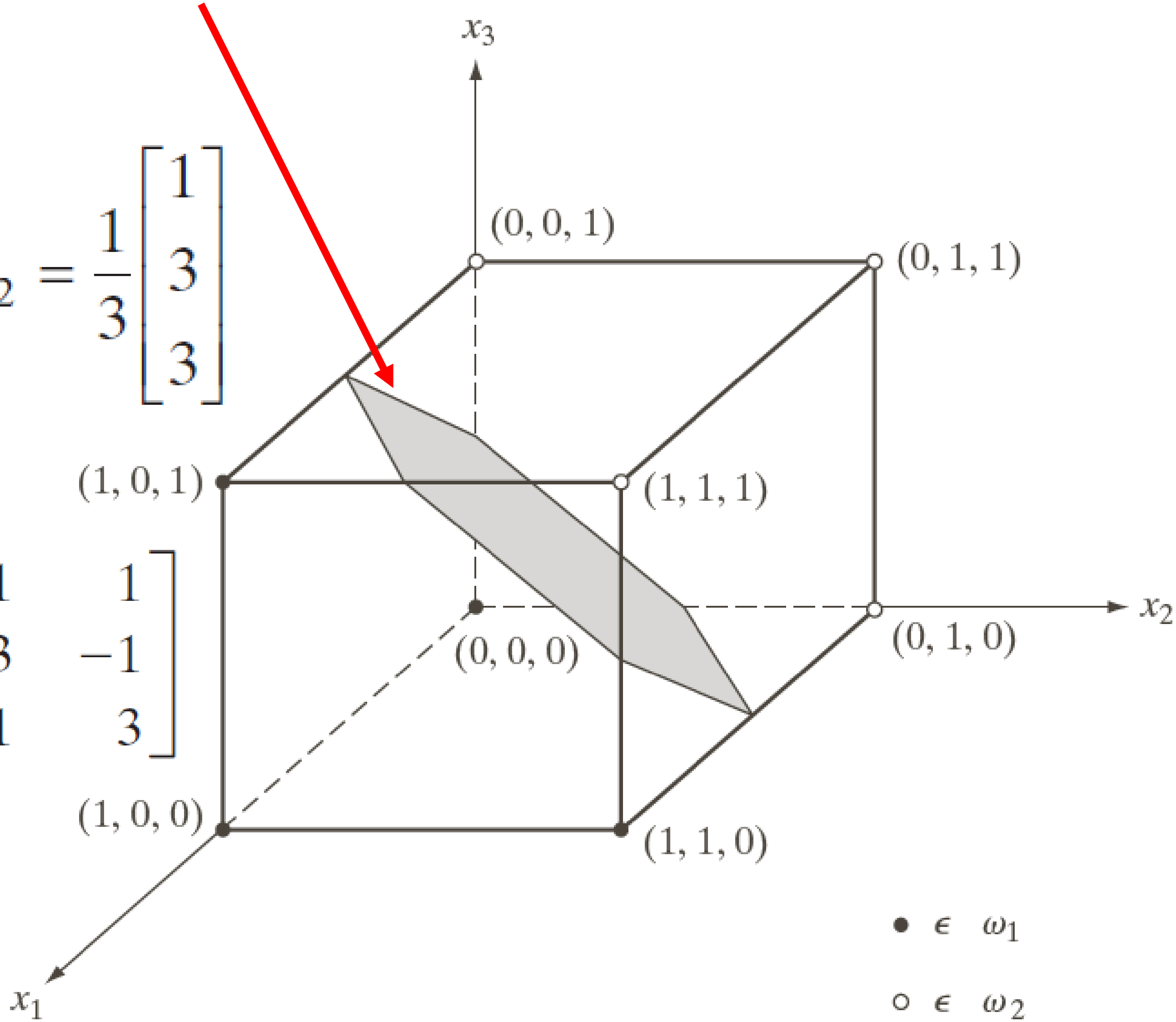
- Bayes → Minimum distance classifier
  - Pattern classes are Gaussian
  - All covariance matrices are Identity matrices
  - All classes are equally likely to occur

Example of Bayes Classifier

$$d_1(\mathbf{x}) - d_2(\mathbf{x}) = 8x_1 - 8x_2 - 8x_3 + 4 = 0$$

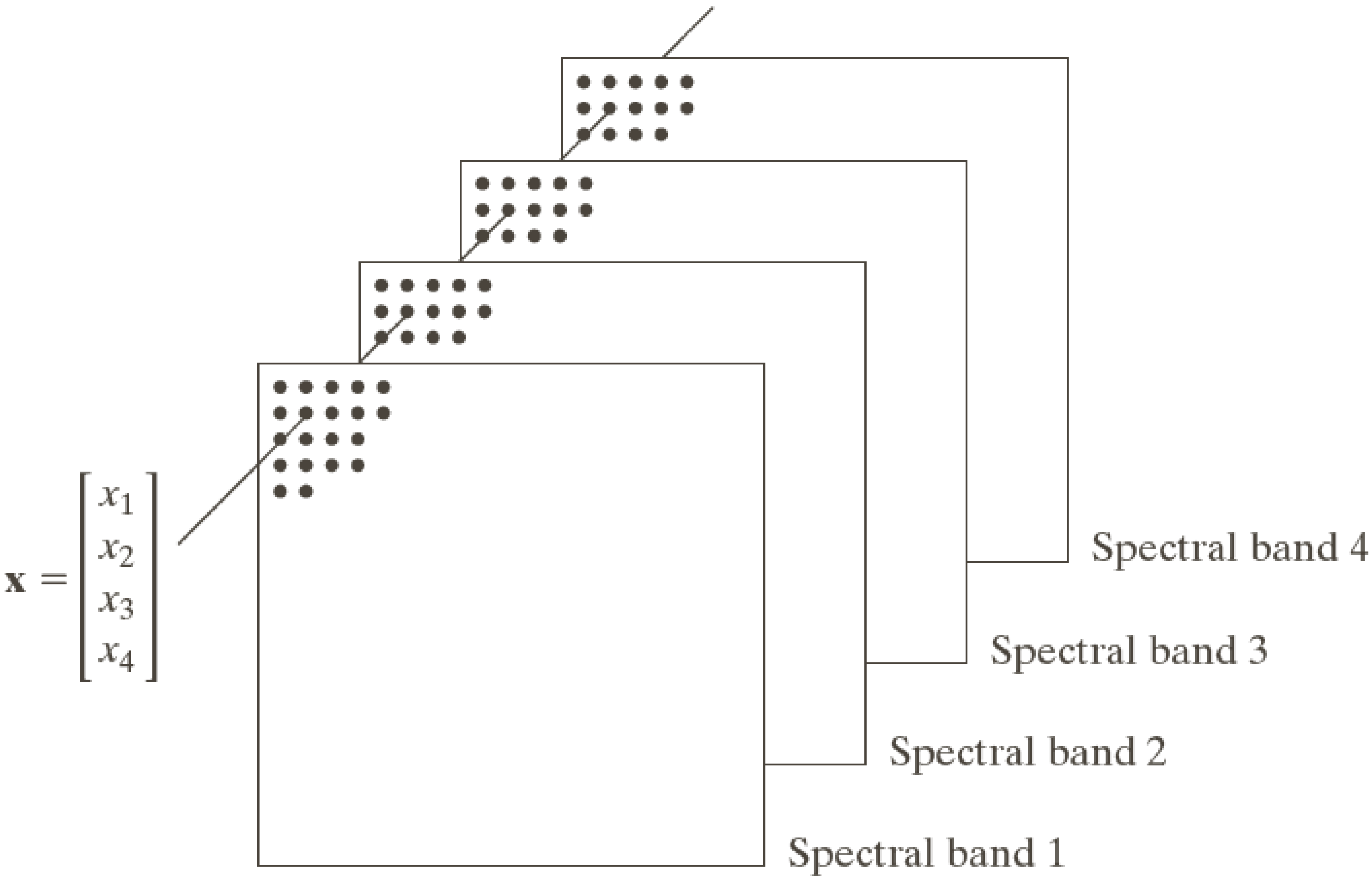
$$\mathbf{m}_1 = \frac{1}{3}\begin{bmatrix}3\\1\\1\end{bmatrix} \quad \text{and} \quad \mathbf{m}_2 = \frac{1}{3}\begin{bmatrix}1\\3\\3\end{bmatrix}$$

$$\mathbf{C}_1 = \mathbf{C}_2 = \frac{1}{16}\begin{bmatrix}3 & 1 & 1\\1 & 3 & -1\\1 & -1 & 3\end{bmatrix}$$

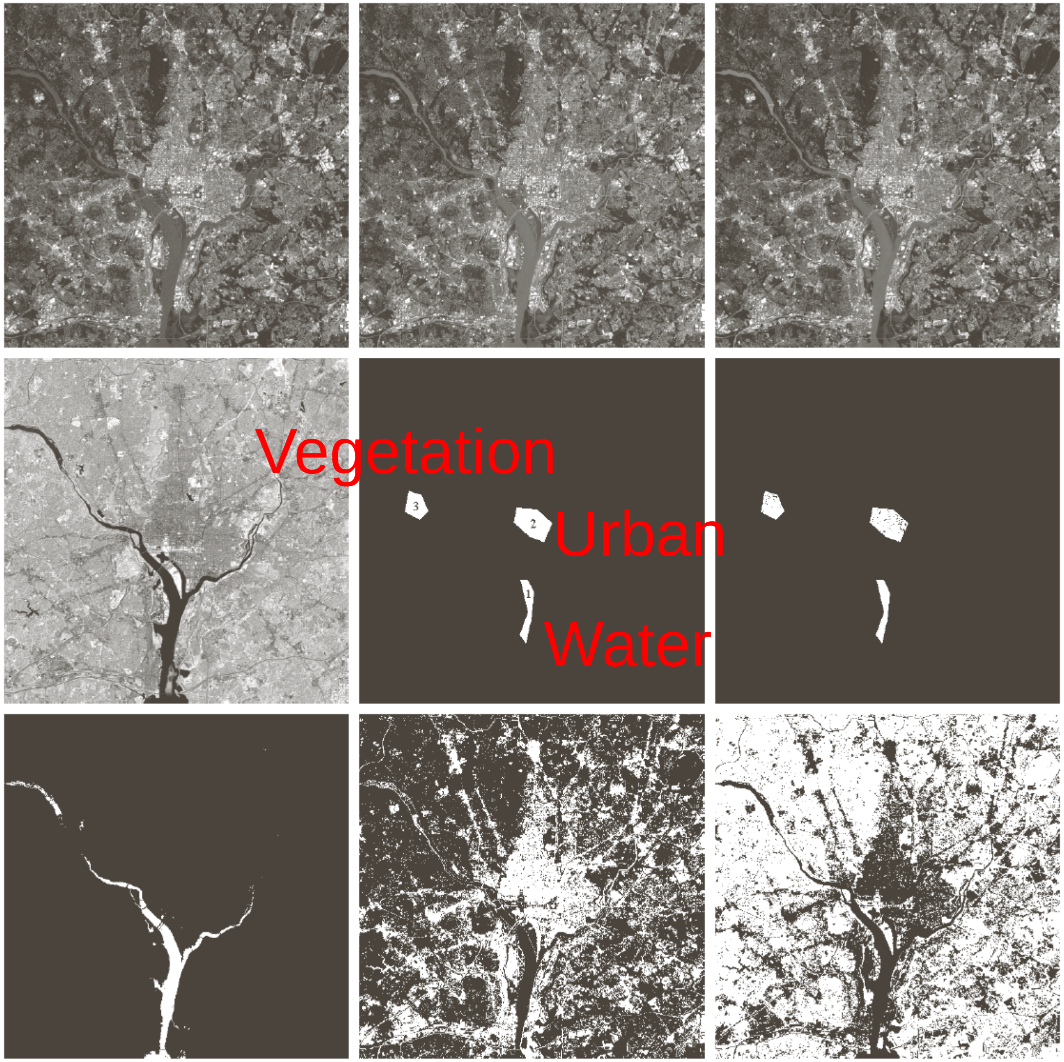


Example: Bayes Classification of Multispectral Data

- Pattern vector formation



Example: Bayes Classification of Multispectral Data



- Sample regions:

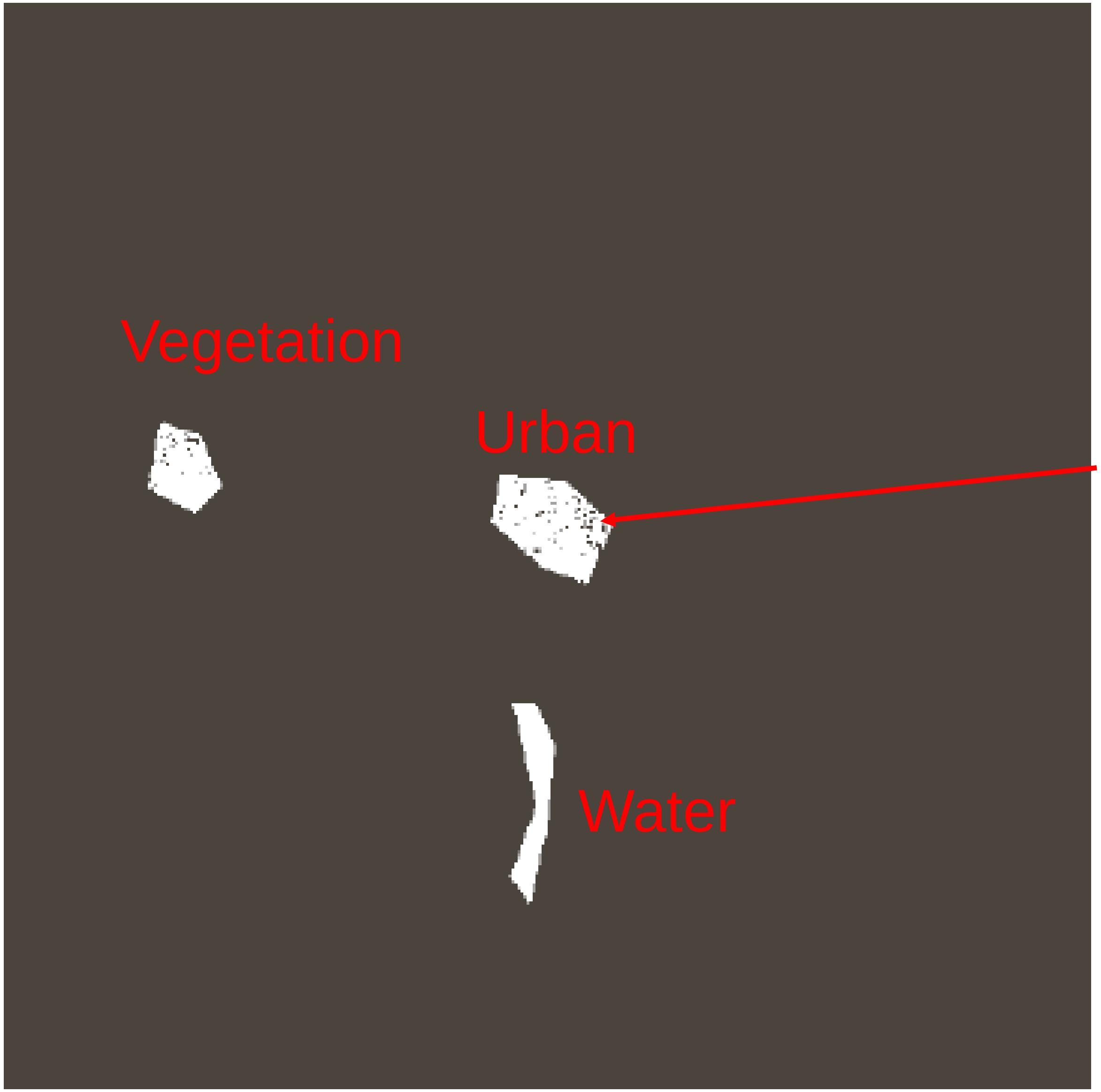
half for training, and half for testing

TABLE 12.1

Bayes classification of multispectral image data. Classes 1, 2, and 3 are water, urban, and vegetation, respectively.

Training Patterns						Test Patterns					
	No. of	Classified into Class			%		No. of	Classified into Class			%
Class	Samples	1	2	3	Correct	Class	Samples	1	2	3	Correct
1	484	482	2	0	99.6	1	483	478	3	2	98.9
2	933	0	885	48	94.9	2	932	0	880	52	94.4
3	483	0	19	464	96.1	3	482	0	16	466	96.7

- Thresholding in segmentation may be viewed as a Bayes classification problem



Black dots denote incorrect classification

Assignments

- 12.2, 12.9, 12.16, 12.30

课后作业题目请对照参考第4版英文原版

- 第7次编程作业

在华为昇腾社区

https://www.hiascend.com/edu/experiment，

自选一个感兴趣的在线实验，独立完成.

Assignments

每个编程作业要求递交1份实验报告，命名“学号姓名_prjX.pdf”，内容提纲包括：

- 实验任务： 描述本次实验的任务。
- 算法设计：理论上描述所设计的算法。
- 代码实现：描述编程环境，给出自己编写的核心代码
- 实验结果：描述具体的实验过程，给出每个小实验的输入数据、算法参数和实验结果，并对结果做简要的讨论。
- 总结：简要总结本次实验的技术内容，以及心得体会