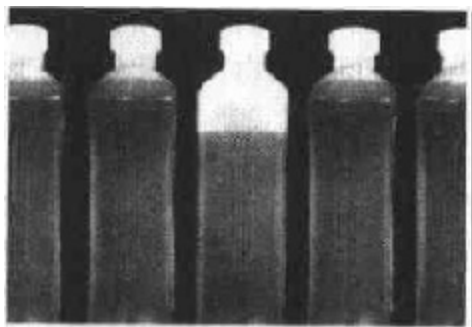


(b)“右边的两个像素”。

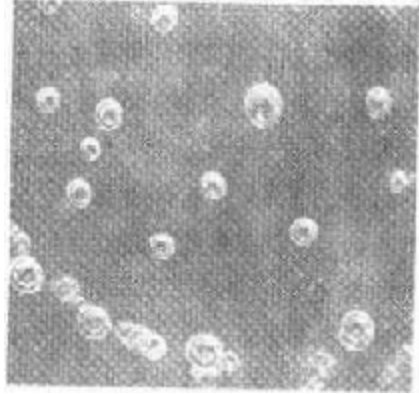
假设左上角的像素具有 0 值。

- 11.17 证明式(11.4.7)、式(11.4.8)和式(11.4.9)的正确性。
- ★11.18 在例 11.10 中提到,用两幅与最大特征值相联系的主分量图像重构 6 幅原图的近似图像是较为可靠的做法。这样做均方误差会受到什么影响?将答案表示成一个最大可能误差的百分比。
- 11.19 对一组大小为 64×64 的图像,假设式(11.4.9)给出的协方差矩阵证明是单位矩阵。则原图和仅用原特征向量的一半由式(11.4.11)重构的图像间的均方误差如何表示?
- ★11.20 如果 11.2.1 节定义的边界主轴和其本征轴相等,希望有什么条件?
- 11.21 给出一个由黑白相间的方块组成的棋盘图形的空间关系及对应的树形表达。假设左上角元素为黑色且树根对应此元素。构造的树从每一个节点辐射的分支不要多于两个。
- ★11.22 要求设计一个检测某种固态晶体圆片内部缺陷的图像处理系统。晶片使用 X 射线成像系统检测,产生分辨率为 512×512 的 8 位图像。在没有缺陷的地方图像显示为平均灰度级是 100 及方差是 400 的“均匀区”。有缺陷的地方呈现如斑点状的区域。这个区域有大约 70% 的像素偏离 50 这一灰度级或均值小于 100。如果有缺陷的区域所占面积超过 20×20 个像素则认为此晶片有缺陷。对此提出一个基于纹理分析的系统。
- 11.23 一个用瓶子装各种工业化学品的装瓶公司听说你成功解决了成像问题,便雇佣你设计一种检测瓶子未装满的方法。当瓶子在传送带上运动并通过自动装填机和封盖机时如下图所示。当液体平面低于瓶颈底部和瓶子肩部的中间点时,认为瓶子未装满。瓶子的横断面上的倾斜部分及侧面定义为瓶子的肩部。瓶子在不断移动,但公司有一个图像系统,装备了有效捕捉停止运动图像的前端闪光照明设备。所以可以得到非常接近于这里显示的样例的图像。基于以上得到的资料,提出一个检测未完全装满的瓶子的解决方案。清楚地表述所做的所有很可能对你提出的解决方案产生影响的假设。



- 11.24 听说你成功地解决了装瓶问题,你和一家公司签订合同,这家公司希望在某个过程中自动进行气泡计数以便进行质量监控。他们已经解决了成像问题并可以得到分辨率为 700×700 像素的 8 位数据图像,如下页所示。每幅图像代表 7 cm^2 的面积。公司希望用每幅图做两件事:(1)确定图像中气泡的面积占图像总面积的

比率,和(2)计算可分辨的气泡数目。基于以上提供的资料,提出一个解决方案。在你的解决方案中,务必阐述方案中可检测的最小气泡的物理尺寸。清楚地说明所做的所有很可能对你提出的解决方案产生影响的假设。



第12章 对象识别

这里通过对对象识别技术的介绍总结一下数字图像处理所涵盖的范围。正如在1.1节注意到的,我们已经定义了包括识别单个图像区域在内的数字图像处理过程所涵盖的范围。在本章中我们称单个图像区域为对象或模式。

在这一章中阐述的模式识别方法分为两个主要方面:理论判别和结构判别。第一类方法处理的是使用定量描绘子描述的各种模式。定量描绘子有:长度、面积和纹理,等等。第二类方法处理的是大部分由定性描绘子描述的各种模式。比如,11.5节讨论的关系描绘子。

识别的中心问题是从样本模式“学习”这一概念。在接下来的材料,我们将对研究理论判别和结构判决两类方法的“学习”技术进行深入讨论并给出图示加以说明。

12.1 模式和模式类

模式即描绘子的组合,例如那些在第11章中讨论过的符号。特征在有关模式识别的著作中经常被用于表示一个描绘子。模式类是一个拥有某些共同性质的模式族。模式类用 $\omega_1, \omega_2, \dots, \omega_W$ 表示,在这里 W 是模式类的数量。由机器完成的模式识别包括对不同的模式分配各自所属类的技术——这种技术是自动的并且尽可能地减少人的介入。

实践中三种常用的模式组合是矢量(用于定量描述)、串和树(用于结构描述)。模式矢量用粗体小写字母表示,比如: $\mathbf{x}, \mathbf{y}$ 和 $\mathbf{z}$,并采取下列形式:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (12.1.1)$$

这里,每个分量 x_i 代表第 i 个描绘子, n 是与模式有关的符号总数。模式矢量用列向量表示(即 $n \times 1$ 阶矩阵)。因此模式矢量可以用式(12.1.1)中所示的形式或用公式 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 的等价形式表示,这里 T 表示矩阵转置。读者可以在11.4节了解这种表示法。

模式矢量 $\mathbf{x}$ 中的元素性质取决于描述物理模式自身所采用的方法。让我们用一个简单的例子来说明并给出度量分类的历史意义。在一篇经典的论文中,费希尔(Fisher [1936])发表了后来称为判别式分析的新技术(在12.2节讨论),它通过测量花瓣的宽度和长度(图12.1)来识别三种不同类型的鸢尾属(Iris)植物的花(Iris Setosa, Iris Virginica 和 Iris Versicolor)。

在已有的术语中,每一种花都用两个度量描述,这就生成了一个二维模式矢量的形式。

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (12.1.2)$$

x_1 和 x_2 分别代表花瓣的长度和宽度。本例中的3个模式类 $\omega_1, \omega_2, \omega_3$ 分别表示 Setosa,

Virginica和 Versicolor 三种花。

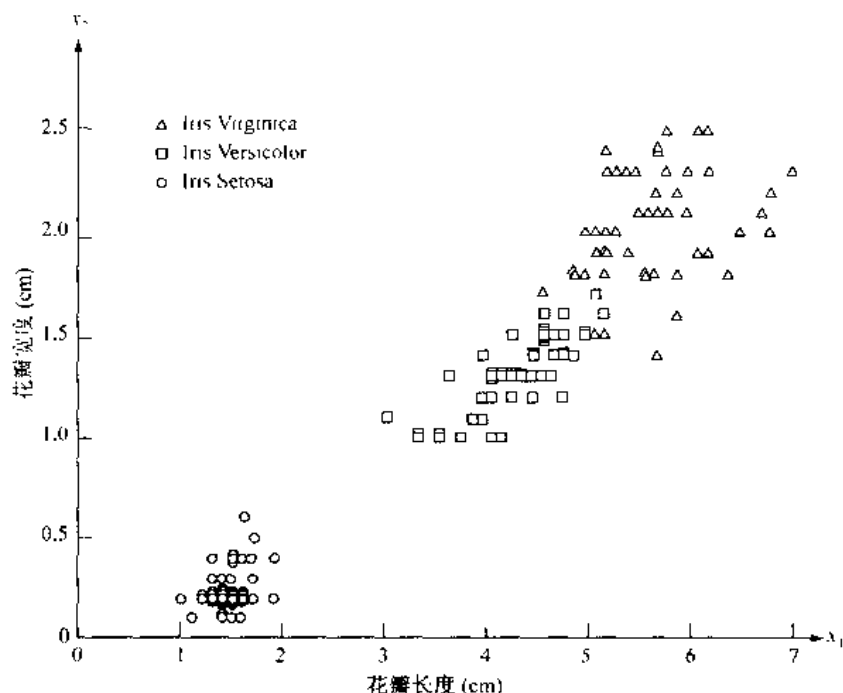


图 12.1 用两个度量描述三类鸢尾属植物的花

由于花瓣在宽度和长度上不同,描述这些花的模式矢量也会不同,这种不同不仅存在于不同种类的花瓣之间,在同一种类的花瓣间也存在。图 12.1 显示了每一类鸢尾属植物花瓣样本的长度和宽度的度量。在选定了一系列度量后(本例中选择两组数据),模式矢量的元素便成为每一个物理样本的完整描述。因此,例子中的每一朵花就由二维欧氏空间中的一个点代表。我们也注意到,本例中的花瓣长度和宽度较好地将 *Iris Setosa* 类同其他两类区分开来,但对 *Virginica* 和 *Versicolor* 类则分辨效果不佳。这种结果说明了经典特征选择方法的问题,说明类型可分离程度的高低很大程度上取决于应用的描绘子的选择。对此我们将在 12.2 节和 12.3 节中做更多的阐述。

图 12.2 显示了另一个生成模式矢量的例子。在这个例子中,关注的是不同类型的噪声波形。其中一个样本示于图 12.2(a)。如果我们选择用噪声信号代表每一个对象(见 11.1.3 节),会得到如图 12.2(b)中显示的一维信号形式。假设选择样本信号的幅度值来描述每一个信号,即以指定的间隔值 θ 对信号进行取样,表示为 $\theta_1, \theta_2, \dots, \theta_n$ 。然后通过令 $x_1 = r(\theta_1), x_2 = r(\theta_2), \dots, x_n = r(\theta_n)$ 构成模式矢量。这些矢量成为 n 维欧氏空间中的点,并且模式类被画成 n 维空间中的“云团”。

如果不直接使用信号的幅度,可以计算一个给定信号(见 11.2.4 节)的前 n 个统计矩,并使用这些描绘子作为每一个模式矢量的分量。实际上,现在很明显,可以有很多其他方法生成模式矢量。将在这一章中介绍其中的一些方法。目前,应记住的关键概念是,一个模式矢量的元素是根据什么量的描绘子进行选择,这对于基于模式矢量方法进行对象识别的最终效果有很大影响。

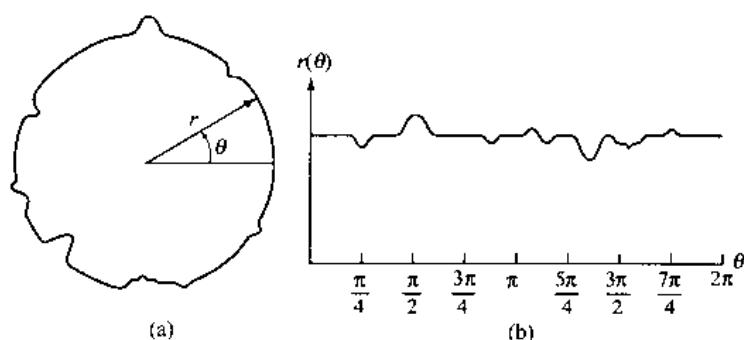
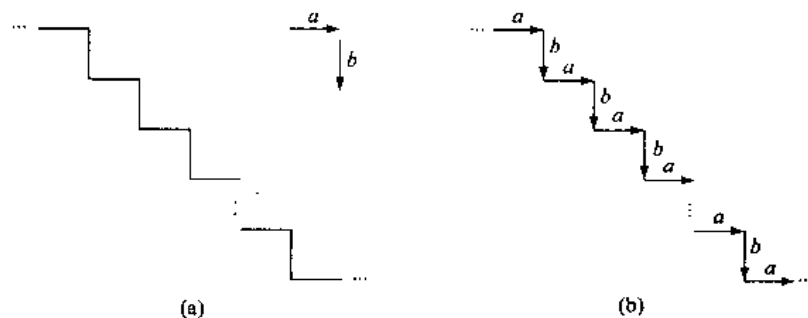


图 12.2 一个噪声对象和它对应的信号

刚才描述的这种产生模式矢量的技术可生成由定量信息描述其特性的模式类。在某些应用中,模式的特性很适于用结构关系进行描述。例如,指纹识别是基于称为细节的指纹特征的相互关系。综合指纹的相对大小和位置,这些特征是描述指纹纹路属性的主要分量,如指纹的端点、分支、合并以及不连续段。这类识别问题通常用结构性方法会得到很好的解决,因为它们的特征不仅同数量有关,而且各个特征间的空间关系也决定着它们的类别归属。此问题将在 11.5 节中介绍。在这里我们重新简略地回顾了有关模式描绘子的内容。

图 12.3(a)显示了一个简单的阶梯状模式。这个模式可以从模式矢量的角度取样和表达,如同在图 12.2 中使用的方法。然而,用这种描述方法将使其基本的结构特征(即两个原始元素构成的重复出现的特征)无法表现出来。一种更有意义的描述方法是定义元素 a 和 b 并将模式定义成如图 12.3(b)中所示的字符串 $w = \cdots abababab \cdots$ 。这种特殊模式类别的结构是由要求的描述得到的,这种描述用一种首尾方式定义连接,并且只允许交换符号。这种结构构成适用于任何长度的阶梯,但不包括由原始 a 和 b 连接成的另一类字符串生成的其他类型结构。

图 12.3 (a)阶梯结构,(b)基于元素 a 和 b 的结构编码生成串描述 $\cdots ababab \cdots$

串的描述适于生成对象模式和其他实体模式,它们的结构是基于原始元素的较简单连接,通常和边界形状有关系。对许多应用来说更有效的一种方法是树形描述结构,如在 11.5 节中定义的。基本上,树形结构是一种主要的分层有序结构。例如,图 12.4 显示了一幅有稠密建筑物群的城市中心区及周围居民区的卫星照片。用 S 定义整个图像区域。图 12.5 中显示的树形表示法(从上到下)是通过应用结构关系中的“包含于”得到的。因此,树的根结点代表整幅图像。下一级的结点表示此图是由城市中心区和居民区构成的。居民区依次由住宅、高速公路和商业区构成。再下一层进一步描述了住宅和高速公路。可以一直继续这样的细分,直到到达在图像上解析不同区域的能力极限。

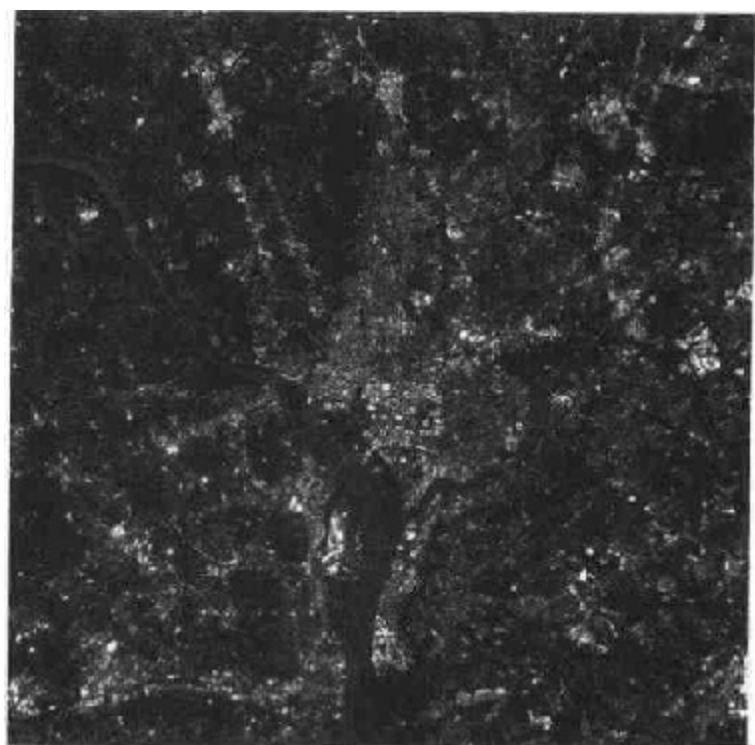


图 12.4 有大量建筑物的城市中心区(华盛顿特区)和周围的居民区的卫星图片(由 NASA 提供)

我们在接下来的章节中将深入探讨在以前的段落中所讨论技术描述的对象识别方法。

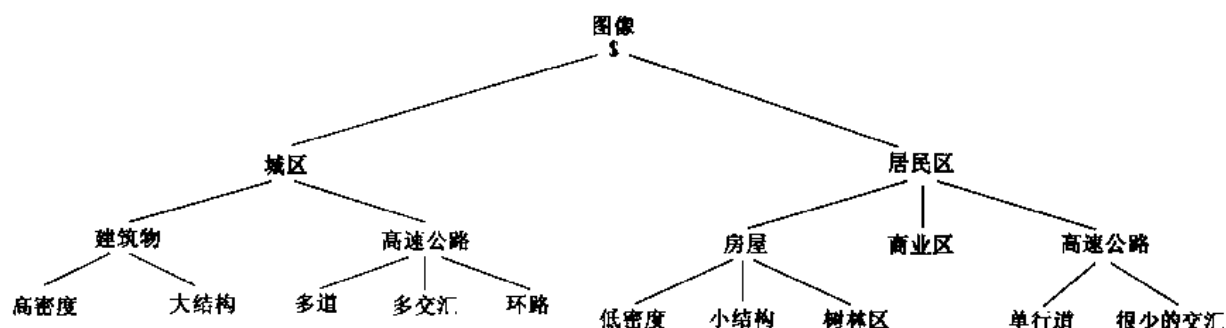


图 12.5 图 12.4 中一幅图像的树描述

12.2 基于决策理论方法的识别

使用决策理论方法的识别是基于使用决策(或判别)函数的识别。令 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 代表一个 n 维模式矢量,如在 12.1 节中讨论的。对 W 个模式类 $\omega_1, \omega_2, \dots, \omega_W$, 决策理论模式识别的基本问题是依据属性寻找 W 个判别函数 $d_1(\mathbf{x}), d_2(\mathbf{x}), \dots, d_W(\mathbf{x})$, 如果模式 $\mathbf{x}$ 属于类 ω_i , 则

$$d_i(\mathbf{x}) > d_j(\mathbf{x}) \quad j = 1, 2, \dots, W; j \neq i \quad (12.2.1)$$

换句话说,一个未知的模式 $\mathbf{x}$ 被称为属于第 i 个模式类,只有当把 $\mathbf{x}$ 代入所有判别函数后能得到 $d_i(\mathbf{x})$ 的最大值时才成立。被解决的模式和模式类之间的联系可以是任意的。

从 ω_j 中分离 ω_i 类别的决策边界由 $\mathbf{x}$ 值判断, $\mathbf{x}$ 值由 $d_i(\mathbf{x}) = d_j(\mathbf{x})$ 或等价地由下式给出:

$$d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0 \quad (12.2.2)$$

通常是通过单一函数 $d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0$ 来识别两个类之间的决策(判别)边界。因此, $d_{ij}(\mathbf{x}) > 0$ 时属于 ω_i 类的模式, 而 $d_{ij}(\mathbf{x}) < 0$ 时属于 ω_j 类的模式。本节重点探讨各种各样的寻找判别函数的方法。这些函数要满足式(12.2.1)。

12.2.1 匹配

基于匹配的识别技术通过一种原型模式矢量表示每一个类。一个未知模式被赋予一个按预先已定义的度量与其最相近的类。最简单的方法是最近距离分类器, 这种方法如其名称所暗示的一样, 要计算(在欧几里得空间中的)未知量和每一个原型矢量间的距离。选择其中的最小距离来决策。还讨论过一种基于相关的方法, 这种方法可以直接依据图像用公式表示并且相当直观。

最小距离分类器

假设我们把每一个模式类的原型定义为该类模式的平均矢量:

$$\mathbf{m}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in \omega_j} \mathbf{x}_j \quad j = 1, 2, \dots, W \quad (12.2.3)$$

这里 N_j 是类 ω_j 的模式矢量数目, 在总和式中取全部这些矢量。如以前一样, W 是模式类的数目。决定一个未知模式矢量 $\mathbf{x}$ 的类别成员的方法是, 如前边所述的那样, 将它分配给与它最接近的原型类。使用欧几里得空间距离判断紧密性简化了计算距离测度的问题:

$$D_j(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_j\| \quad j = 1, 2, \dots, W \quad (12.2.4)$$

这里 $\|\mathbf{a}\| = (\mathbf{a}^T \mathbf{a})^{1/2}$ 是欧几里得范数。 $D_i(\mathbf{x})$ 的值是最小距离时, 把 $\mathbf{x}$ 划归给类 ω_i 。也就是说, 最小距离暗示此公式表示了最好匹配。不难看出(习题 12.2)选择最小距离等同于下面的函数评估:

$$d_j(\mathbf{x}) = \mathbf{x}^T \mathbf{m}_j - \frac{1}{2} \mathbf{m}_j^T \mathbf{m}_j \quad j = 1, 2, \dots, W \quad (12.2.5)$$

并且在 $d_i(\mathbf{x})$ 得出最大数值时将 $\mathbf{x}$ 划归给类 ω_i 。这个公式同式(12.2.1)定义的决策函数的概念一致。

在式(12.2.2)和式(12.2.5)中, 类 ω_i 和 ω_j 之间的最小距离分类器的决策边界是:

$$\begin{aligned} d_{ij}(\mathbf{x}) &= d_i(\mathbf{x}) - d_j(\mathbf{x}) \\ &= \mathbf{x}^T (\mathbf{m}_i - \mathbf{m}_j) - \frac{1}{2} (\mathbf{m}_i - \mathbf{m}_j)^T (\mathbf{m}_i - \mathbf{m}_j) = 0 \end{aligned} \quad (12.2.6)$$

由式(12.2.6)给出的面是连接 $\mathbf{m}_i$ 和 $\mathbf{m}_j$ 线段的中垂线。 $n = 2$ 时, 中垂线是一条线; $n = 3$ 时, 是平面; $n > 3$ 时, 是一个超平面。

例 12.1 最小距离分类器的图解

图 12.6 显示了两个从图 12.1 鸢尾属植物例子中抽象出来的模式类。这两个模式类: Iris

Versicolor 和 Iris Setosa 分别由 ω_1 和 ω_2 表示, 样本平均矢量 $\mathbf{m}_1 = (4.3, 1.3)^T$ 和 $\mathbf{m}_2 = (1.5, 0.3)^T$ 。根据式(12.2.5), 决策函数是:

$$\begin{aligned} d_1(\mathbf{x}) &= \mathbf{x}^T \mathbf{m}_1 - \frac{1}{2} \mathbf{m}_1^T \mathbf{m}_1 \\ &= 4.3x_1 + 1.3x_2 - 10.1 \end{aligned}$$

和

$$\begin{aligned} d_2(\mathbf{x}) &= \mathbf{x}^T \mathbf{m}_2 - \frac{1}{2} \mathbf{m}_2^T \mathbf{m}_2 \\ &= 1.5x_1 + 0.3x_2 - 1.17 \end{aligned}$$

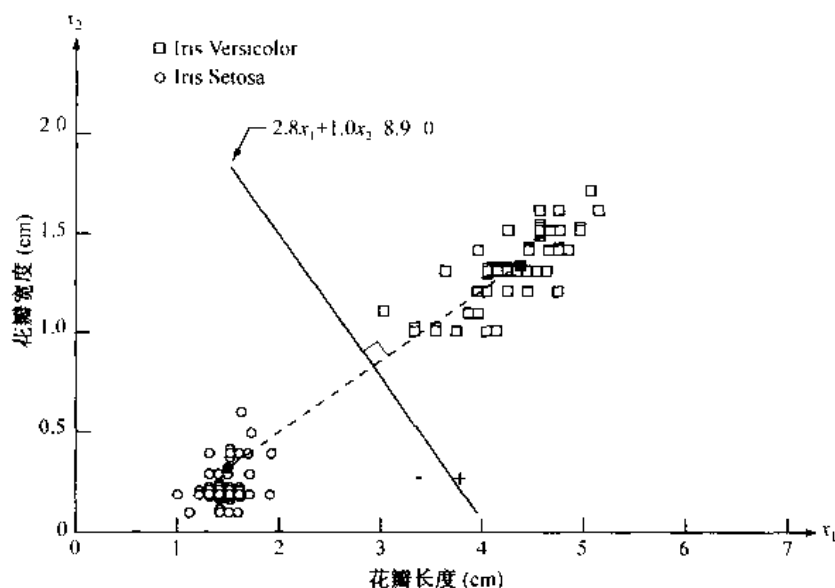


图 12.6 Iris Versicolor 和 Iris Setosa 类的最小距离分类器的决策边界。黑点和方块是平均值

由式(12.2.6)得出边界方程是:

$$\begin{aligned} d_{12}(\mathbf{x}) &= d_1(\mathbf{x}) - d_2(\mathbf{x}) \\ &= 2.8x_1 + 1.0x_2 - 8.9 = 0 \end{aligned}$$

图 12.6 显示了具有这种边界的一幅图(注意轴线并未使用同一比例)。任何模式矢量代入类 ω_1 会得到 $d_{12}(\mathbf{x}) > 0$ 。相反,任何模式矢量代入类 ω_2 会得到 $d_{12}(\mathbf{x}) < 0$ 。换句话说,给出一个属于这两个类之一的未知模式,依据 $d_{12}(\mathbf{x})$ 的符号将完全可以判断模式归属的类。

实际上,就每个模式类的均值而论,当均值间的间距与每一类的分开度或随机性相比很大时,最小距离分类器的分辨效果是很好的。在 12.2.2 节,我们可以看到,当每个类的均值分布在 n 维模式空间中呈现一种球形的“超级云团”状态时,最小距离分类器会得到最佳性能(从错误分类的平均损失的最小水平上来说)。

除系统的设计者控制输入性质之外,在实际中,大的均值分离和相对小的类散布同时出现是很少见的。一个很好的例子就是设计系统来读取有固定格式特征的字体。例如,大家所熟悉的美国 Banker 协会的 E-13B 字体的字符集。如图 12.7 所示,这个特殊的字符集由 14 个字

符组成,这些字符被特意设计在 9×7 个格子中以便于读取。这些字符通常用含有精细磁性材料的墨水印刷。在读取这些字符之前,墨水经过一个磁场,它加重每个字符以便简化检测。换句话说,分割问题已经用人工突出每个字符的关键特征解决了。

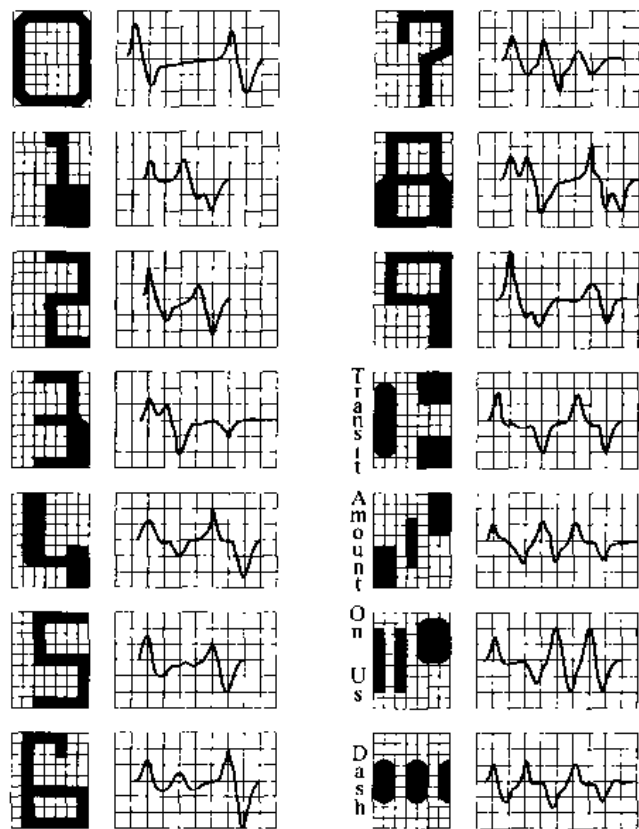


图 12.7 美国 Banker 协会的 E-13B 字体的字符集和对应波形

典型的字符扫描是在水平方向上用一個比字符长但更窄的单缝读取头进行的。当读取头扫过一个字符时,它产生一个一维的电信号,这个电信号随读取头下方的字符面积的增加或减少按比率变化。例如,考虑图 12.7 中数字 0 有关的波形。当读取头从左向右运动时,从读取头看去的字符区开始增大,从而产生一个正的变形(一个正的变化率)。当读取头开始远离 0 的左半部时,读取头下方的字符面积开始减少,产生一个负的变形。当读取头处于字符的中间区域时,字符面积近乎保持不变,产生一个 0 变形。当读取头进入字符的右半部分时,这种变形模式重复出现。字体的设计保证每个字符产生与其他字符不同的波形。同时也保证每种波形的波峰和零点大致在网格背景的垂直线上出现,这些波形同图 12.7 所显示的一样。E-13B 字体有一个特性,即仅仅在某些点采样就可以产生足够的恰当分类的特征信息。磁性墨水的使用有助于产生干净的波形从而将采样的散布减少到最小。

为这类应用设计一个最小距离分类器很简单。只需要存储每种波形的样本值,而后令每组样本用一个原型矢量 $\mathbf{m}_j$ ($j = 1, 2, \dots, 14$) 来表示。当对一个未知字符进行分类时,按刚才所述的方法去扫描它,将波形的网格样本表示成一个矢量 $\mathbf{x}$,选择式(12.2.5)中得到最大值的原型矢量类,这个类就是矢量 $\mathbf{x}$ 所属的类。由电阻排组成的模拟电路可以实现高速分类(见习题 12.4)。

依据相关的匹配

在4.6.4节中介绍了图像相关的基本概念。把这一概念作为在一幅大小为 $M \times N$ 的图像 $f(x, y)$ 中寻找匹配的大小为 $J \times K$ 的子图 $w(x, y)$ 的基础。这里, 我们假定 $J \leq M$ 且 $K \leq N$ 。尽管相关方法可以用矢量形式表达(见习题 12.5), 但直接对一幅图像或子图形式进行操作更为直观(也更为常见)。

在其最简单的表示形式中, $f(x, y)$ 和 $w(x, y)$ 间的相关是:

$$c(x, y) = \sum_s \sum_t f(s, t) w(x + s, y + t) \quad (12.2.7)$$

$x = 0, 1, 2, \dots, M-1, y = 0, 1, 2, \dots, N-1$, 和式代表图像 w 和 f 相互重叠的地方。请注意将本式同式 (4.6.30) 相比较, 式 (4.6.30) 暗示假设函数都是实的量并且忽略常数 MN 。原因是将用一种能将常数抵消掉的归一化函数, 并且式 (12.2.7) 中给出的定义在实际使用中更为常见。并且在式 (12.2.7) 中使用符号 s 和 t 避免同 m 和 n 相混淆, 这些符号在本章中也用于其他目的。

图 12.8 用图示说明了处理步骤, 这里, 假设 f 原点在它的左上角, w 原点在它的中心。对于 $f(x, y)$ 中 x, y 的一个值, 比如说 (x_0, y_0) , 应用式 (12.2.7) 可得到量 c 的一个值。当 x 和 y 的值变化时, w 在图像区中移动, 给出函数 $c(x, y)$ 。 c 的最大值指出 w 能够最好地匹配 f 的位置。注意, 在 x 和 y 的值接近 f 的边缘处时, c 值会失去准确性。在相关中误差量与 w 的大小成比例。这是我们熟悉的边界问题, 在第 3 章已多次遇到过。

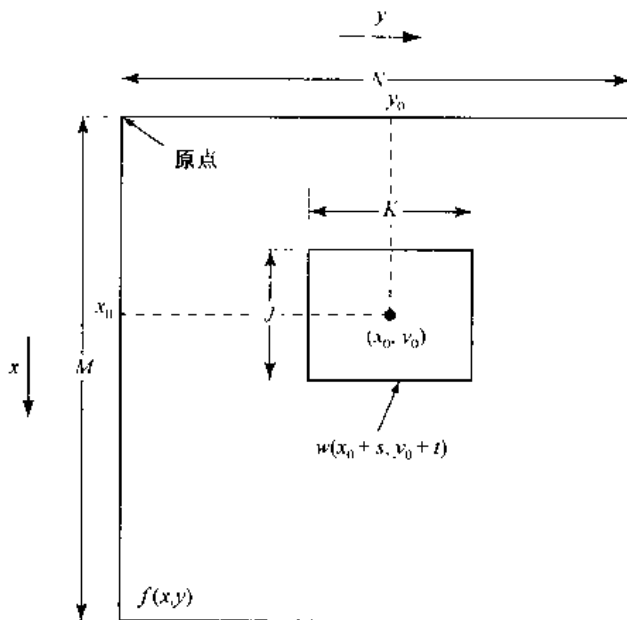


图 12.8 在点 (x_0, y_0) 得到 f 和 w 相关的排列

在式 (12.2.7) 中给出的相关函数存在着对于 f 和 w 的幅度变化过于敏感的缺点。例如, 若 f 的所有值乘 2, $c(x, y)$ 的值也加倍。为克服这个缺点而经常使用的一种方法是, 在匹配过程中使用相关系数, 其定义如下: