第十六章 物体识别

目前，物体识别的基本方法是建立物体模型，然后使用各种匹配算法从真实的图像中识别出与物体模型最相似的物体。物体识别的正式定义如下：给定一幅包含一个或多个物体的图像和一组对应物体模型的标记，机器应将标记正确地分配给图像中对应的区域或区域集合。物体识别和图像分割是紧密相关的，因为没有物体或物体局部识别，分割就无法进行，而没有分割，物体识别也是不可能的。

16．1 识别系统的基本组成

可以认为，物体识别系统包括四个主要模块：即模型库、特征检测器、假设生成(hypothesis formation)和假设验证(hypothesis verification)等模块。图 16.1 给出了系统不同模块之间的作用和信息流图。

模型库包含有所有的已知模型。模型库的信息取决于物体识别方法，可以是定量、定性或函数描述，也可以是精确的几何曲面信息。在大多数情况下，物体的模型是抽象的特征矢量。特征是物体的一种属性，比如，尺度、色彩和形状等，特征在描述和识别物体过程中起着十分重要的作用。

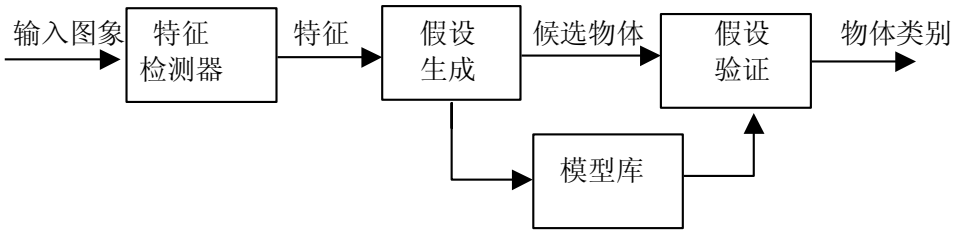


图 16.1 物体识别系统组成示意图

输入图象可以是灰度图象、彩色图象、深度图象或它们的组合。特征检测器对输入图象的特征进行检测，并对特征进行定位，这样有助于假设生成。物体特征的选取取决于待识别物体的类型和模型库数据结构。假设生成模块使用图像特征来给场景中的物体分配一个似然值，这一步可以大大减小物体识别的搜索空间。一般来说，模型库是一种索引图，它有利于从所有可能的物体集合中去除那些不可能的候选者。假设验证模块使用物体模型来验证假设，并进一步给出精确似然值。在所有证据的基础上，选用具有最大似然值的物体作为识别结果。

假设生成和假设验证在不同的识别方法中，其重要性是不一样的。一些系统仅仅使用假设生成，然后选择具有最高似然值的物体作为识别结果。模式分类方法是此种方法的一个很好的例子。另一方面，许多智能系统很少依靠假设生成，更多的工作是在验证阶段。实际上，经典模式识别方法之一的模板匹配方法就没有假设生成阶段。

为了实现上述步骤，物体识别系统必须选择合适的手段和方法。对于特定的应用，在选择合适的方法时，必须考虑许多因素和问题。在设计物体识别系统时必须考虑的问题有：

(1) 模型表示

模型表示涉及到物体具有那些重要属性或特征以及这些特征如何在模型库中表示。对

于大多数物体来说，几何特征描述是可以很有效的；但对于另外一些物体，可能需要更一般的特征或函数来表示。物体的表示应该包含所有相关信息，但没用任何冗余信息，并且将这些信息以某种方式组织起来，使得物体识别系统的不同组元能够容易访问这些信息。

(2) 特征提取

特征提取的算法有很多，根据应用对象，应选择可靠的特征检测方法和特征定位方法。许多特征可以在二维图像中计算出来，但它们与物体的三维特征有关。由于图像生成过程的特性，有些特征可以很容易地计算出来，而其它特征计算起来则非常困难。

(3) 特征模式匹配

特征模式匹配是指图像中的物体特征同模型库中的模型相匹配。在许多物体识别任务中，待识别的物体的数量较多，每一个物体拥有的特征也有许多。显然，穷举匹配方法可以解决识别问题，但识别效率太低，不是很实用。因此，在建立匹配方法时，必须考虑特征的有效性和匹配算法的高效率。

(4) 假设生成

为了有效地提高识别效率，可以根据物体特征首先建立可能的物体集合，并给每一可能的物体分配置概率值。“假设生成”过程基本上是一种启发式过程，由此可以减小搜索空间。假设生成过程特别注重使用应用域知识，将某种置信概率值分配给该应用域中的不同物体。

(5) 物体验证

如何使用物体模型，从给定图像中的可能物体集合中选择最有可能的物体？每一个可能物体的存在可以用它们的模型来证明。我们必须测试每一个可能的假设来验证一个物体的存在或忽略这个物体的存在。如果模型是几何模型，则很容易用摄像机的位置和其它场景参数来验证物体。

按照问题的复杂性，图 16.1 的一个或多个模块可能变得不重要，这主要取决于问题的复杂度。举例来说，基于模式识别的物体识别系统不使用任何基于特征的匹配或物体验证；它们直接给物体指定概率并选择具有最大概率的物体。

16．2 物体识别的复杂度

我们知道，场景图像与照明、摄像机参数、摄像机位置等因素有关，因此，要从一幅图象中识别物体，特别是从包含多个物体的图像中识别特定的物体，必须考虑这些因素。物体识别的复杂度依赖于以下几个因素：

(1)　场景的不变性

场景的复杂度取决于获取图像时的条件(照明、背景、摄像机参数和观察点)是否同模型建立条件相似。如前几章所置述，场景的条件显著地影响同一物体的图像。在不同的场景条件下，不同特征检测器的性能显著不同。因此必须考虑背景、其它物体以及照明的特性，以决定哪种特征可以得到有效而可靠地检测。

(2)　图像模型空间

在某些应用中，三维物体可以近似地认为是二维物体，此时的物体模型可以用二维特征来表示。如果模型是三维且不能忽略透视效应，那样情况就变得很复杂。在这种情况下，特征是在二维图像空间中检测的，而物体的模型可能是在三维空间中表示的。这样，同一个三

维空间特征可能在三维图像中表现为不同的特征。在动态图像分析中，由于物体运动，这种情况也会发生。

(3) 模型库中物体的数目

如果物体的数目很少，则可以直接使用顺序穷举匹配方法，无需假设生成阶段。如果物体的数目很大，则假设生成阶段是很重要的。用于物体识别的特征选择计算量也随着物体数量的增加而迅速地增加。

(4) 图像中物体的数目和遮挡问题

如果图像中只有一个物体，它可能是完全可见的。随着图像中物体的数目增加，遮挡概率也随之增加。在许多图像分析中，遮挡是一个严重的问题。遮挡导致了原先特征点的消失，新特征点的产生。因此，在假设验证阶段就应该考虑遮挡问题。一般来说，识别任务的难度随着图像中物体数目的增加而增大。图像中遮挡物体的存在也使图像分割难度增大。

根据物体识别任务所在的空间，常把物体识别分为二维识别和三维识别。

(1)二维

在许多应用中，图像是从足够远的距离上获取的，因此可以认为图像是通过正交投影生成的。如果物体总是在场景中的一个稳定位置，那么也可以认为是二维情况。在这些应用中，可以使用二维模型数据库。二维物体识别一般有两种可能的情况：

- 物体没有被遮挡 ，如遥感和许多工业应用场合。
- 物体被其它物体遮挡或者只有部分可见，如识别堆放物体问题。

(2) 三维

从不同的视角获取同一物体的图像可能是完全不同的，此时识别物体需要三维模型。在物体识别过程中，还要考虑投视投影以及获取图像的视角的影响。对于三维情况，有两种用于物体识别任务的信息：

- 灰度图像　　灰度图像没有明显包含物体表面信息，用灰度图像可以识别对应于物体三维结构的特征
- 2.5 维图像　在许多应用中，以观察者为中心的坐标系中的物体表面可以直接通过测距成像传感器获取的距离图像或通过立体灰度图像对计算出来的深度图来表示，这里的深度图和距离图像即为 2.5 维图像。物体的曲面信息可以有效地用于物体识别任务。

16．3 图像矩不变量特征表示

矩不变量特征主要是针对二维识别情况提出来的。人是很容易从图象中识别出特定的物体形状；但对于机器视觉来说却是一件难事。一方面，图象分割受到背景与物体之间的反差影响以及光源、遮挡等影响，不容易实现；另一方面，摄象机从不同的视角和距离获取的同一场景的图象是不同的，这样给形状的提取和识别带来很大困难。人们对二维形状的提取和识别已经做了大量的研究，提出了许许多多的方法。本节仅仅介绍一种被广泛使用的矩不变量特征。

矩不变量是指物体图象经过平移、旋转以及比例变换仍然不变的矩特征量。设二维物体

的图象用 $f(x,y)$ 表示。其 $(p+q)$ 阶矩定义为：

$$m_{pq}=\sum_x\sum_y x^p y^q f(x,y) \tag{16.1}$$

$$\mu_{pq}=\sum_x\sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{16.2}$$

其中 $\bar{x}=\dfrac{m_{10}}{m_{00}}$ $\bar{y}=\dfrac{m_{01}}{m_{00}}$

零阶矩 $m_{00}=\sum_x\sum_y f(x,y)$，当 $f(x,y)$ 相当于物体密度时，则零阶矩 m_{00} 是密度的总和，即物体的质量。一阶矩 $m_{10}=\sum_x\sum_y xf(x,y)$ 和 $m_{01}=\sum_x\sum_y yf(x,y)$ 分别除以零阶矩 m_{00} 后所得的 $\bar{x}=\dfrac{m_{10}}{m_{00}}$ 和 $\bar{y}=\dfrac{m_{01}}{m_{00}}$ 是物体质量中心的坐标，或者直接表示的是区域灰度重心的坐标。

中心矩 μ_{pq} 反映区域 R 中灰度重心分布的度量。例如 μ_{20} 和 μ_{02} 分别表示 R 围绕通过灰度重心的垂直和水平轴线的惯性矩。若 $\mu_{20}>\mu_{02}$，则可能是一个水平方向拉长的物体。μ_{30} 和 μ_{03} 的幅值可以度量物体对于垂直和水平轴线的不对称性。如果是完全对称的形状，其值应为零。

$(p+q)$ 规范化中心矩记作 η_{pq}，定义为

$$\eta_{pq}=\mu_{pq}\big/\mu_{00}^r \tag{16.3}$$

其中 $$r=(p+q+2)/2$$

利用二阶和三阶规范化中心矩可以导出下面七个不变矩组：

$$\phi_1=\eta_{20}+\eta_{02} \tag{16.4}$$

$$\phi_2=(\eta_{20}-\eta_{02})^2+4\eta_{11}^2 \tag{16.5}$$

$$\phi_3=(\eta_{30}-3\eta_{12})^2+(3\eta_{21}+\eta_{03})^2 \tag{16.6}$$

$$\phi_4=(\eta_{30}+\eta_{12})^2+(\eta_{21}+\eta_{03})^2 \tag{16.7}$$

$$\begin{aligned}\phi_5=&(\eta_{30}-3\eta_{12})(\eta_{30}-\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2]\\&+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]\end{aligned} \tag{16.8}$$

$$\begin{aligned}\phi_6=&(\eta_{20}-\eta_{02})[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]\\&+4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03})\end{aligned} \tag{16.9}$$

$$\begin{aligned}\phi_7 &= (3\eta_{12}-\eta_{30})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2] \\ &+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})[3(\eta_{03}+\eta_{12})^2-(\eta_{12}+\eta_{03})^2]\end{aligned} \quad (16.10)$$

Hu M.K.在 1962 年已证明这个矩组对于平移、旋转和比例变化都是不变的。

在实际中，用上式计算形状的矩特征不变量，其数值分布范围在之间$10^0 \sim 10^{-12}$，显然，矩不变量特征值越小，对识别结果的贡献也越小。为此，可以对上述七个矩不变量进行如下修正：

$$t_1=\phi_1 \qquad t_2=\phi_2 \qquad t_3=\sqrt[5]{\phi_3^2} \qquad t_4=\sqrt[5]{\phi_4^2}$$

$$t_5=\sqrt[5]{\phi_5^2} \qquad t_6=\sqrt[5]{\phi_6^2} \qquad t_7=\sqrt[5]{\phi_7^2} \quad (16.11)$$

用上述公式得到矩特征不变量值分布范围大约在$10^0 \sim 10^{-4}$之间.

在使用矩不变量时，还要注意以下几个问题：

（1）二维矩不变量是指二维平移、旋转和比例变换下的不变量，因此，对于其它类型的变换，如仿射变换、射影变换，上述的矩不变量是不成立的，或只能作为近似的不变量。

（2）对于二值区域图像，区域与其边界是完全等价的，因此可以使用边界的数据来计算矩特征，这样可以大大提高矩特征的计算效率。

（3）矩特征是关于区域的全局特征，若物体的一部分被遮挡，则无法计算矩不变量，在这种情况下，可以使用物体区域的其它特征来完成识别任务。

16．4 三维物体模型表示

图像是场景在图象平面上的一种透视投影表示，因此在“以摄像机为中心的坐标系”，或“以观察者为中心”的坐标系中表示物体是很自然的，当然也可以在“以物体为中心”的坐标系中表示物体，或在世界坐标系中表示物体。不过，选择合适的坐标系会有利于坐标的变换、特征检测和后处理等有关算法的有效实现。

在工程研究领域，人们常常通过牺牲某一部分的代价来换取另一部分的高性能。在机器视觉领域，为了提高某一算法的有效性，通常是以增加运算量或增加计算成本(时间、存贮空间或硬件成本)为代价的。用于物体识别的表示也不例外。因此，设计者必须认真考虑系统设计问题中的参数，一般选择最好的表示。目前，人们已经开发出许多物体表示方法这些方法大致分为三大类：

- 基于表面的物体模型表示方法，如表面片、网机表示等。
- 基于体积的物体表示方法，如结构立体几何、体元或体系表示。
- 基于函数的表示方法，样条函数、超二次曲面等到。

下面讨论几种物体识别的常用表示方法。

16．4．1 多视图表示

如果要通过图像识别三维物体，则三维物体必须由若干幅图像来表示，这些图像是从空间中任意点或从特定点拍摄的。对于大多数物体来说，必须获取表示该物体各个方向的形态的大量图像才能实现有效的物体识别任务。

用图像表示物体的一种方法是朝向图(aspect graph)表示，朝向图包含了一个物体的所有稳定的视图。以及所有稳定视图之间的关系。图 16.2 给出了一个简单的物体及其朝向图，朝向图的每一个结点表示一个稳定的视图，结点连线表示从一个稳定视图到另一个稳定视图的过程。

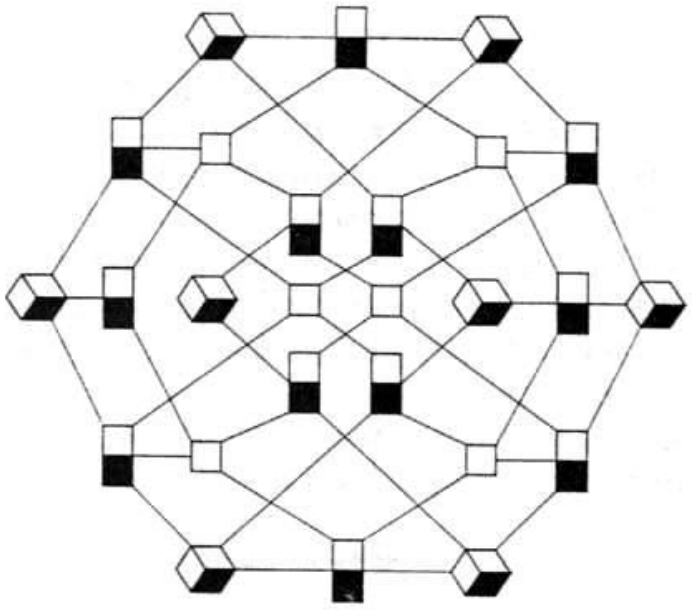


图 16.2 一个简单物体朝向图

16．4．2 结构立体几何

结构立体几何(constructive solid geometry, CGS)方法使用简单的立体基元和一组布尔运算来表示物体，立体基元包括长方体、圆锥、圆柱和球等简单的三维形状，布尔运算是指：并、交、差。CGS 表示式如下：

〈CGS 表示〉::〈立体基元〉 |

依据〈运动参数〉移动〈CGS 表示〉 |

〈CGS 表示〉〈布尔运算〉〈CGS 表示〉

利用上述运算可以将物体的结构立体几何表示用一个树结构来描述，如图 16.3 所示。叶结点表示立体基元或运动参数，其它结点表示布尔运算。通过立体基元可以构造许多“人工”物体，如图 16.4 所示。

实际上，一般曲面的物体都不能用所选的几种基元来表示，所以 CSG 在物体识别中的应用十分有限。这些表示常用于 CAD/CAM 应用中的物体表示。

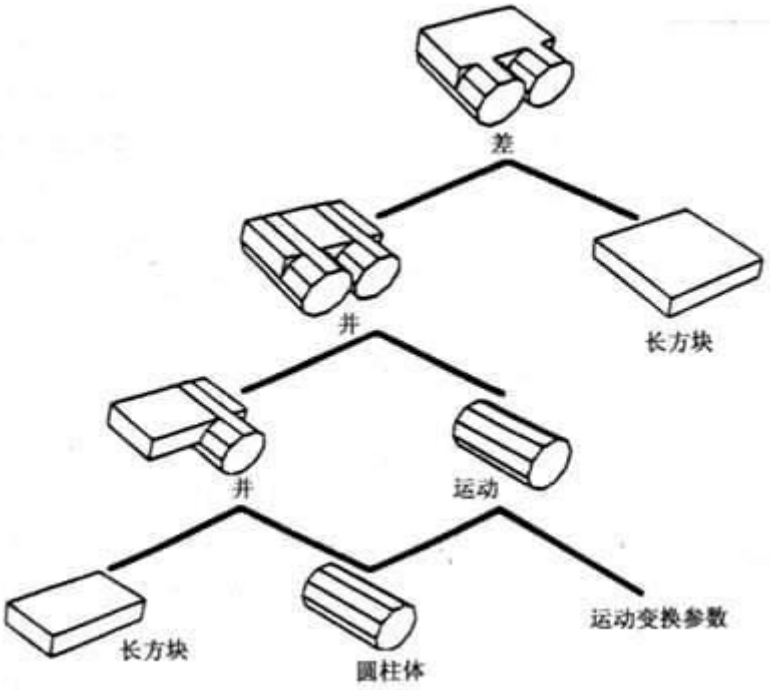


图 16.3 物体的 CSG 表示示意图

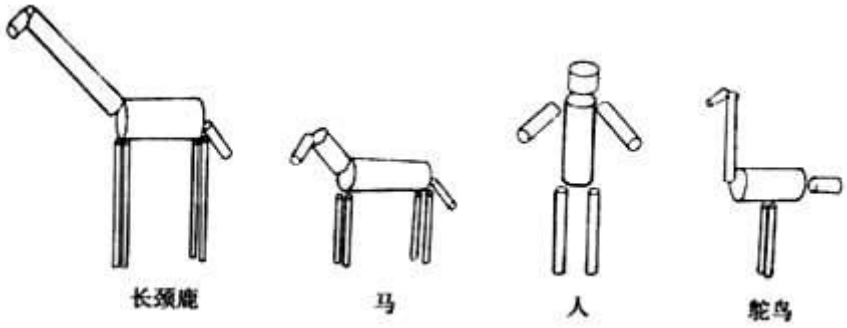


图 16.4 动物的参数化表示

16．4．3 体积表示

三维物体可以用该物体所占三维空间的非重叠子区域来表示，即空间占有量。一般非重叠子区域可以分为标准和非标准两大类。标准子区域是指正方体、长方体等基本体素（voxel），利用这种体素构造物体的方法称为体素表示。图 16.5 给出了物体的体素表示。

非标准子区域是指三维物体本身具有的特定的体积单元，利用这种体元构成的三维物体表示称之为单元分解(cell decomposition)表示方法。单元分解将三维物体分割成更小的单元，单元与单元之间不共享体积，将相邻单元之间的这种关系称为准不连接(quasi-disjoint)。唯一的运算是＂粘接＂，如图 16.6 所示。通常要求分解后的单元不包含有孔，而且能够进行简单的连接。现在的两种常用的单元分解方法是“八叉树”和“K 级树”，它们可以通过递归体积分解过程来构造。单元分解的准不连接性质和占有单元，在一些算法中是非常有用的，质量可通过计算各个简单单元的质量并求和来获得。这样可以表明立体是否连成一起，或是否有孔洞，并且能够容易地以单元分解和空间占有(Spatial occpancy)的形式表示非一致的对象（人体胸腔内部组织构造），此时在每一个单元中将保留 CT 数值，或物质的编码信息，而不是以比特表示的“实或空”的信息。

空间占有量表示方法包含了物体的详细描述，这是一种低层次的描述。这种类型的表示必须经过处理才能得到物体的特定特征，以使得假设生成过程成为可能。

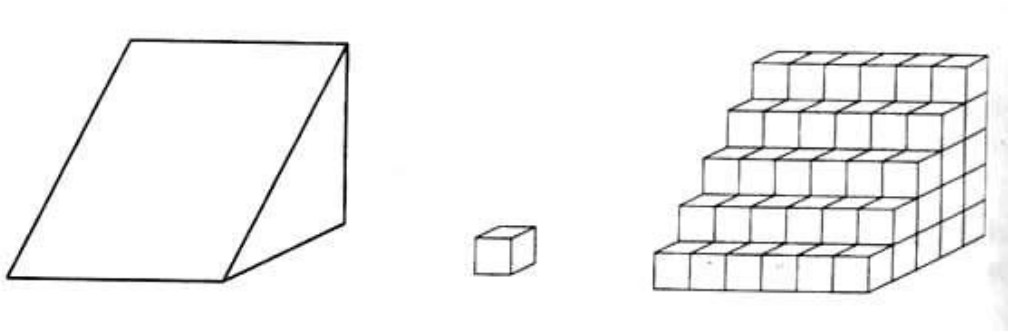


图 16.5 物体的体元表示

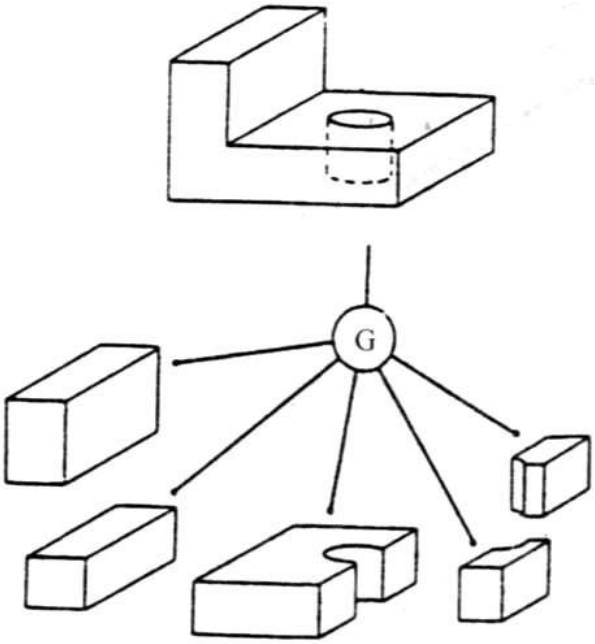


图 16.6　单元分解示意图

16．4．4　扫掠表示：广义柱面

物体形状的扫掠表示包含一条作为轴线的三维空间曲线，一个二维截面图，和定义截面如何沿空间曲线扫掠的扫掠规则，如图 16.7 所示。其中，二维截面可以沿着脊梁线光滑地变化，柱体轴是用虚线表示的，坐标轴是相对于柱体中心轴画出的，每一点处的截面垂直于柱体中心轴

对于许多工业零件或其它物体，物体的截面一般沿空间轴光滑变化，在这种情况下，这种表示方法是令人满意的。但对于任意形状的物体，光滑条件通常是不满足的，因而这种表示也是不合适的。

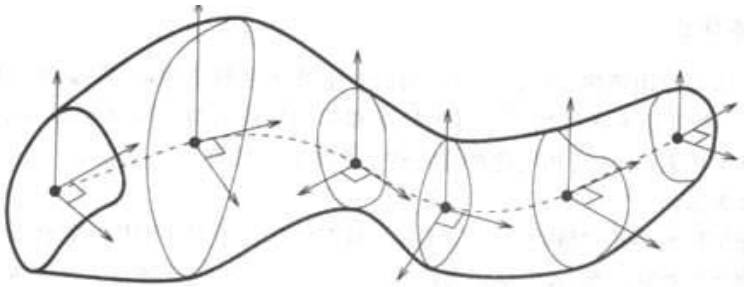


图 16.7 物体的广义柱面表示。

16．4．5 函数表示

我们在第七章和第十三章介绍了大量的函数据表示方法，比如三次曲线，双三次曲面等。这些曲线与曲面是从计算几何发展起来的，适应于曲线和曲面设计。在设计过程中，一般不需要所设计的曲面与某一已知物体形状完全一致，比如，人的手指用圆柱体近似等。但是，在物体识别领域，为了唯一地识别物体，必须使用一种准确的函数来描述一个已知物体，当然，在实现中有很大的难度，比如，常常出现相同或相近的物体形状会有完全不同的函数表示。在计算机视觉领域使用的另一类函数是广义圆柱面和超二次曲面[Pentland 1986,Bajcsy 1987]，这类函数可以用于一大类物体建模，并具有简洁性。

16．4．6 三角形网面表示

物体三维形状模型的更一般表示是多边形网面表示，其中应用最为普遍的是三角形网面表示。一个物体三维形状数据通常有两种途径得到：一种是根据实际物体的几何形状通过 CAD 方法建立，这种方法对于规则形状的物体建模十分有效，比如，机械零件、汽车、飞机等。对于形状十分复杂的物体，比如动物、天然物体，则可以利用测距成像或立体成像系统来获取，图 16.8a 就是利用激光三角测距成像得到的深度图。从物体的不同方向获取一系列深度图并链扣起来[Turk 1994]，就形成物体完整的三维形状数据，然后再用三角形网面表示出来，如图 16.8b 所示。图 16.9 是图 16.8b 网面模型的多分辨率表示[Johnson 1988]。选择适当的分辨率表示既可以保持原有物体的形状，又可以大大减少冗余数据。

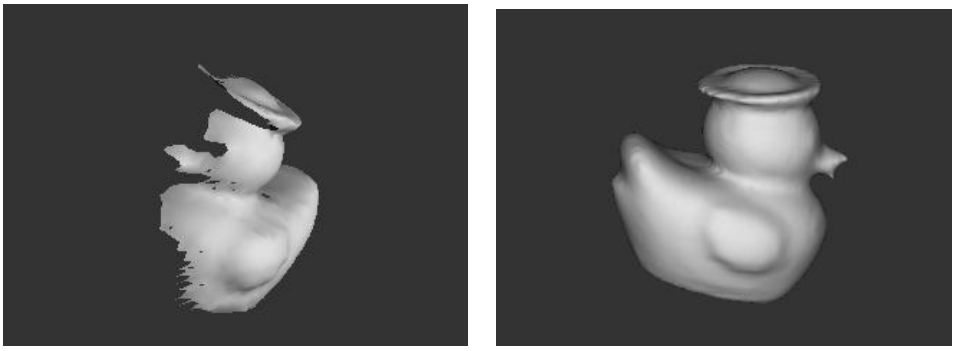
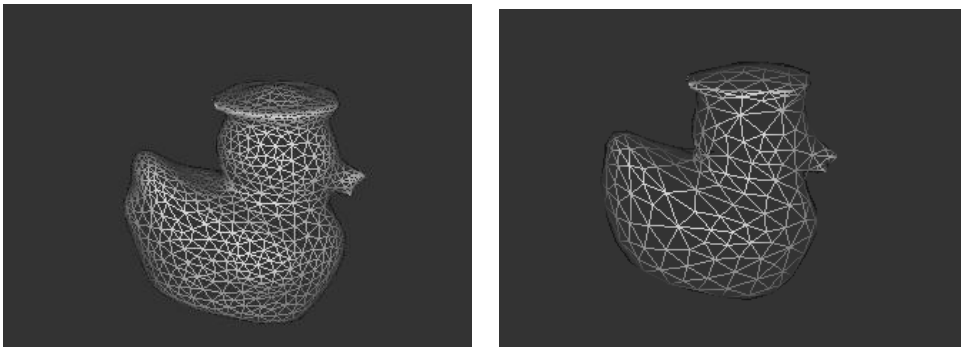


图 16.8 视觉方法建立小鸭玩具模

(a)深度图像序列中的一幅 2.5 维图像



(b)用链扣技术得到的完整三维形状模型

图 16.9 三维模型多分辨率表示

16．5 特征检测与识别策略

进行物体识别的第一步是物体特征检测，然后，基于检测出来的图像特征对图像中可能的物体建立假设公式，并使用物体模型来验证假设。并不是所有的物体识别方法都需要很强的假设公式和验证步骤。大部分识别策略已经演化，将假设和验证这两步以不同的比例组合起来。图 16.10 所示的是假设和验证的三种不同可能组合方法。即使在这些组合中，应用竞争（由本节前面讨论因素来描述）决定如何实现其中的一步或两步。下面我们将讨论几种常用的特征以及用于识别不同环境中物体的基本策略。

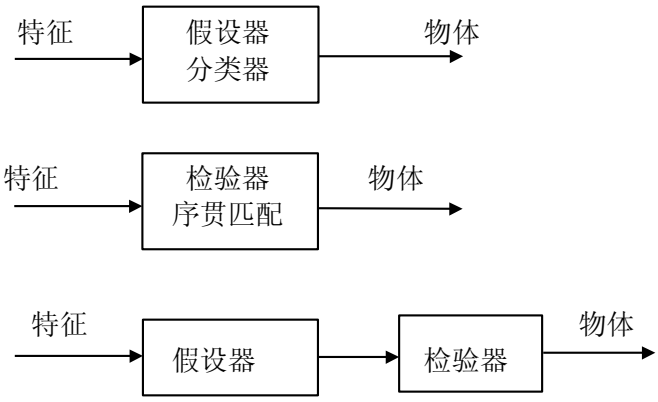


图 16.10 识别策略需要同时使用假设生成步骤和验证步骤或其中的一个步骤，取决于问题的复杂度

16．5．1 特征检测

用于物体识别的特征有许许多多，但大部分特征是基于图像中的区域或边界。假设区域或封闭的边界对应于一个实体，该实体或者是一个物体，或者是物体上的一部分。下面介绍三类常用的特征。

(1)全局特征

全局特征通常是图像区域的一些特征，如面积、周长、傅里叶描述子和矩特征等。全局特征可以通过考虑区域内的所有点来得到，或只考虑区域边界上的所有点来得到。在每一种情况下，目的都是为了找到描述子，该描述子是通过考虑所有点位置、强度特性和空间关系来得到。这些特征在本书不同的章节中都讨论过。

(2)局部特征

局部特征通常位于物体的边界上或者表示区域中可分辨的一个小曲面，比如曲率及其有关的性质就属于局部特征。曲率可能是边界曲率，也可能是从曲面上计算出来的。曲面可以是强度曲面，或是 2.5 维空间曲面。高曲率点，也叫做角点(Corner)，在物体识别中起着重要的作用。局部特征可能包含一个小边界段或是一个表面片的特定形状。一些常用的局部特征是曲率、边界段和角点。在有遮挡或图像不完整的情况下，使用物体的局部特征比用物体的全局特征更有效。图 16.11 所示的是一个物体的局部特征以及特征的图表示。

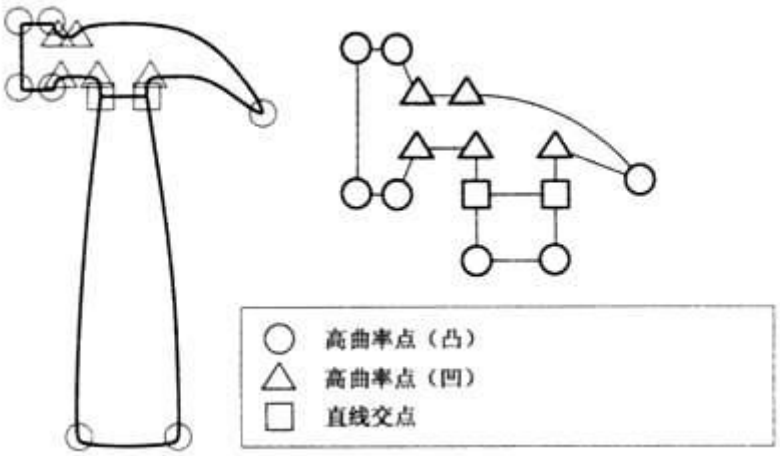


图 16.11　物体局部特征及其图表示

(3)关系特征

关系特征是基于区域、封闭轮廓或局部特征等不同实体的相对位置建立的。这些特征通常包括特征之间的距离和相对方位测量值，它们在基于使用图像区域或局部特征来识别和描述多个实体或物体时是非常有用的。在多数情况下，图像中不同实体的相对位置就完全定义了一个物体。完全相同的特征，但关系特征稍微不同，则可能表示完全不同的物体。

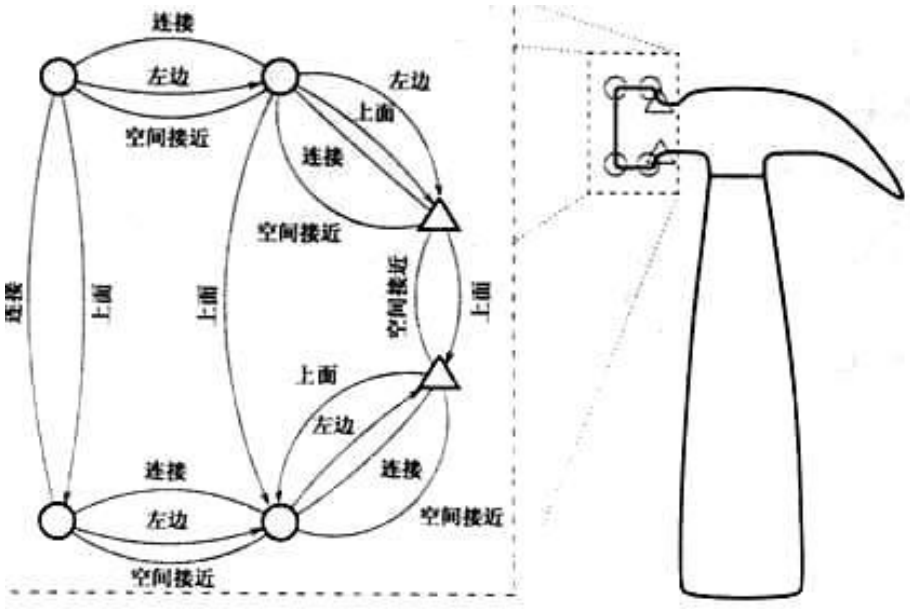


图 16.12 使用多局部和全局特征实现物体的局部表示

在图 16.12 中，我们给出物体和用特征进行物体描述的方法。局部特征和全局特征都可以用于描述一个物体。物体之间的关系可以用于生成复合特征。

16．5．2 特征分类

分类的基本思想是基于特征的匹配和识别。模式识别方法就属于此种类型，并在许多领域中得到广泛的应用。神经元网络方法也属于此种类型。这里简单地讨论一些常用的分类方法。假设 N 个特征已经从图像中检测出来，并被规范化，以便可以表示在同一度量空间。接下来假设一个物体的特征可以表示为 N 维特征空间中的一个点，其中 N 维特征空间是为特定物体识别任务而定义的。

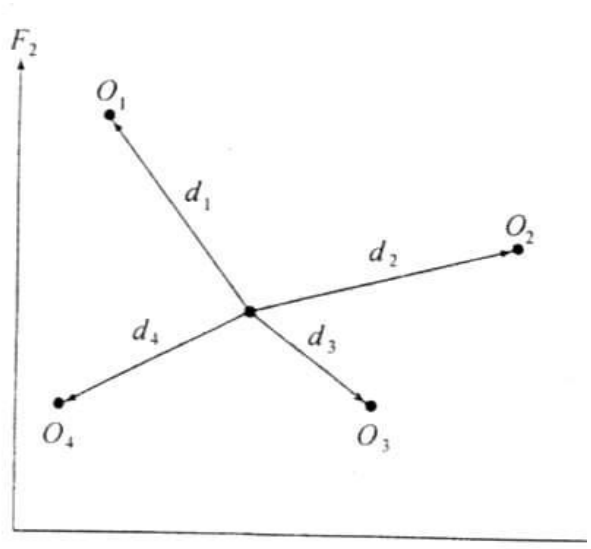


图 16.13　二维空间分类图

（1）最近邻分类器

假设有 N 类物体 $\boldsymbol{\omega}_i$，$i=1,2,...,N$。第 $\boldsymbol{i}$ 类物体模型（理想特征值）的第 j 个特征表示为 ω_{ij}，$j=1,2,...,M_i$，其中 M_i 为第 $\boldsymbol{i}$ 类物体模型的特征数。图 16.13 所示的是一个二维特征空间。为了确定一个物体所属的类别，我们可通过计算该物体特征与模型特征空间中每一类物体特征之间的距离来测量该物体与物体模型的相似性，并将该物体分配给最近的一类。此距离可能是欧几里德距离，或者是特征的任何加权组合。通常，我们计算未知物体 $\mathbf{x}$ 到 $\boldsymbol{i}$ 类物体之间的距离 d_i 如下：

$$d_i(\mathbf{x}) = \|\mathbf{x} - \boldsymbol{\omega}_i\| = \sum_{j=1}^{M_i} (x_{ij} - \omega_{ij})^2 \tag{16.12}$$

或

$$d_i(\mathbf{x}) = \|\mathbf{x} - \boldsymbol{\omega}_i\| = \sum_{j=1}^{M_i} w_{ij} (x_{ij} - \omega_{ij})^2 \tag{16.13}$$

其中，w 是一个权重系数。因为特征空间中，不同的特征对物体分类的贡献是不一样的，对于贡献大的特征，可以分配较大的权重系数，而对那些对噪声十分敏感的特征，则取较小的权重系数。上式的距离计算也可以采用其它的距离公式，如取绝对值等。

根据式(16.12)或(16.13)，物体分类决策函数为：

$$d_k(\mathbf{x}) = \min_i (d_i(\mathbf{x})) \tag{16.14}$$

则

$$\mathbf{x} \in \omega_k$$

这一决策方法称为最小近邻法。这一方法的错误分类率分析见教材[边，1988]

在实际中，找出某一特定的物体可能是很困难的，因为许多物体可能同属于一类，如图 16.14 所示，其中，特征空间中的每一簇点表示一类物体。表示物体类别簇点矩心或每一类的最近点都可认为待识别的物体类别。在这种情况下，用于分类物体的距离测度有两种：

- 将一簇点的矩心作为原型物体的特征点，计算到此点的距离。
- 计算到每一类最近点的距离。

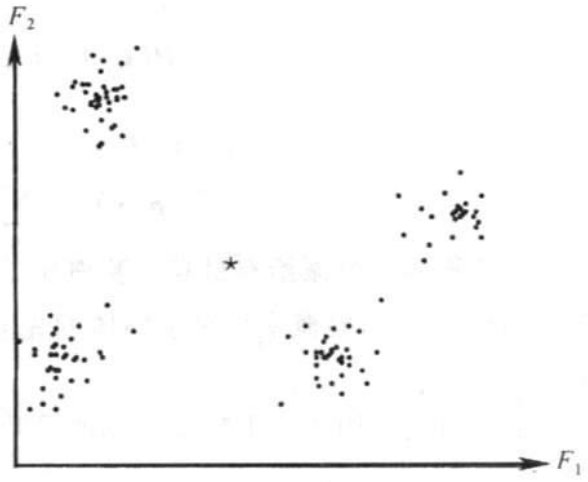


图 16.14 物体在特征空间中表示为点的示意图

（2）贝叶斯分类器

当物体在特征空间中的分布不象上面所示的那么直接时，可用贝叶斯方法来识别物体。通常情况下，不同物体的特征值有着非常严重的重叠。由图 16.15 中所示的一维特征空间可知，几个物体可能具有相同的特征值。因此，对此特征空间的一次观测可能会得到多个满足条件的候选物体类别。在此种情况下，可以用贝叶斯方法来进行决策。

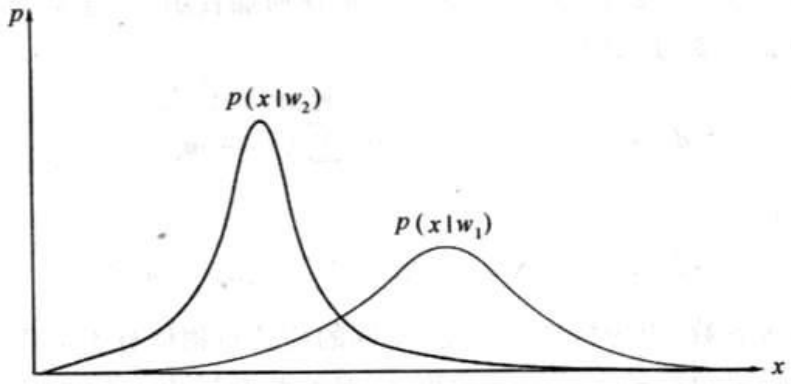


图 16.15 条件概率密度函数 $P(x|w_j)$，表示每一类物体特征值的概率

贝叶斯方法使用了有关物体特征的概率知识和物体出现的频度。假设已知 j 类物体出现的概率为 $P(w_j)$，即先验知识是 $P(w_j)$。因此，在缺乏其它知识的情况下，可以通过把未知的物体分派给 $P(w_j)$ 最大的那一类来使误差概率最小。

关于物体的类别决策通常是基于特征观测做出的。给定概率值 $P(x|w_j)$，如图 16.15 所示。条件概率 $P(x|w_j)$ 告诉我们：基于所提供的概率信息，如果观测的特征值是 x，那么此物体属于 j 类的概率为 $P(x|w_j)$。基于这种知识，我们可以计算物体的后验概率 $P(w_j|x)$。后验概率是在给定信息和观测值的情况下，未知物体属于 j 类的概率。用贝叶斯规则，此概率值为

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j)P(\omega_j)}{p(x)} \tag{16.15}$$

其中

$$p(x) = \sum_{j=1}^{N} p(x \mid \omega_j)P(\omega_j) \tag{16.16}$$

未知物体应分派给有最高后验概率 $P\left(w_j \mid x\right)$ 的那一类。从上面的公式可以看出，如图 16.16 所示，后验概率取决于物体的先验知识。如果物体的先验概率改变了，结果也会变。

上面讨论了用于一个特征识别的贝叶斯方法。这种方法很容易推广到多特征情况。

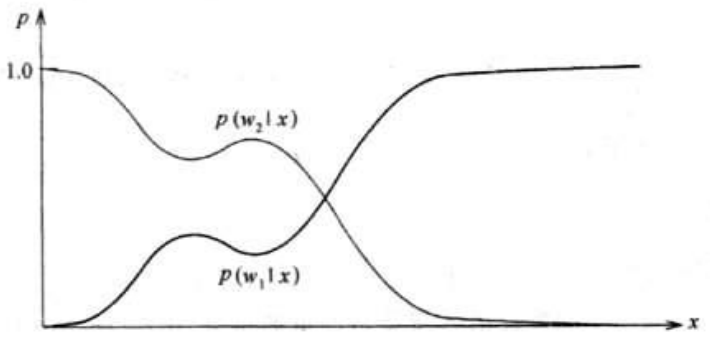


图 16.16　两种不同先验概率值对应的后验概率值示意图

16．5．3 特征匹配

分类方法使用了有效的特征和应用域知识。在许多应用中，很难得到有关特征概率和类别概率的先验知识，或得到的数据不足以设计分类器。在这种情况下，可以使用模型直接匹配未知物体，并选择最佳匹配为最终分类结果。下面讨论一些基本的匹配方法。

(1) 特征匹配

假设每一个特征类别是由它的特征来表示的。同上面一样，假设第 i 类物体的第 j 个特征值表示为 f_{ij}。对于一个未知物体，其特征表示为 u_j。该物体和第 i 类的相似性由下式给出：

$$S_i = \sum_{j=1}^{N} w_j s_j \tag{16.17}$$

其中，w_j 是第 j 个特征的权值。权值的选择是以特征的相对重要性为基础的。第 j 个特征相似值是 s_j，它可以是绝对差、规范化差或其它距离测量值。最常用的方法是用下式并考虑同特征一起使用的权值规范化。

$$s_j = \left|u_j - f_{ij}\right| \tag{16.18}$$

如果S_k是最高相似度值，则标记物体为k类。在此方法中，使用的特征可能是局部的，也可能是全局的。注意此方法没有使用特征之间的任何联系。

(2) 符号匹配

一个物体不仅可以用它的特征来表示，而且可以用特征之间的联系来表示。特征之间的关系可以是空间的，或者是其它形式的。在这样的情况下，物体可能被表示为一个图形。如图 16.11 所示，图形的每一节点都表示一个物体，弧线连结节点表示物体之间的联系。因此，物体识别问题可以认为是图形匹配问题。

一个图形匹配问题可以定义如下：有两个图形G_1和G_2，包含N_{ij}个节点，其中i表示图形数，j表示节点数，节点j和节点k之间的联系表示为R_{jk}。在图形上定义一个相似性测量值，该测量值包含了所有节点和函数的相似性。

在机器视觉的多数应用中，待识别的物体可能是部分可见的。因此，一个识别系统必须能从物体的部分视图来识别它们。那些使用全局特征和要求所有特征都存在的识别方法在这些应用中是行不通的。从某种意义上，部分视图识别问题和图形学中研究的图形嵌入问题是类似的。但当我们开始考虑节点相似性和节点之间关系时，物体识别中的问题与图形学问题就不同了。

我们将在 16.6 节中，详细讨论这种匹配。

16．5．4 特征标记

如果物体的数量很大，并且无法使用特征空间划分来求解，那么索引方法就变得很有吸引力了。上面讨论的符号匹配方法是一种序贯方法，需要未知物体和所有物体进行比较。显然，这种方法无法用于含有大量物体的情况。对于含有大量物体的情况，应该使用假设生成方法来减小搜索空间。然后在减小后的特征空间中，每一个物体模型与图像进行比较来实现识别物体。

特征索引方法使用了物体的特征值来构造模型数据库。当在一幅图像中检测到索引集中的一个或多个特征时，则可以用此特征来减小搜索空间，从而减小用于物体识别的总的时间。

索引集中的特征必须用模型库知识来确定。如果这样的知识无法得到，就应该分析每一个来自特征集中特征出现的频率，并在特征频率的基础上，生成用于构造数据库的索引集。

在索引数据库中，除了物体的名字和它们的模型外，有关物体在特征空间中出现的位置和方向信息都应该保存，因为这种信息在验证阶段很有帮助。

一旦生成候选物体集，就应该进入验证阶段，以选择最佳候选物体。

16．6 验证

给定一幅物体的图像，在图像中找出某类物体出现的数量及出现的位置，这是基本的验证问题，而不是物体识别问题。显然，可以用验证算法来一个一个地穷举来验证每一个模型在模型库中的存在。但是，这样的穷举方法在模型库较大时不是有效的方法。实际上用于验证的方法有许多，这里只讨论一些常用的方法。

16．6．1 模板匹配

假定有一个模板 $g(i,j)$，我们希望检测图像 $f(i,j)$ 中的模板情况。显而易见，把模板放置在图像中的某一位置，通过比较模板中的强度值和图像中对应值，可以检测模板在哪一位置的存在。因为强度值很少能够很好地匹配，我们需要测量模板强度值同对应图像值之间的不相似度。下面定义几种测量手段：

$$\max_{[i,j]\in R} | f-g | \tag{16.19}$$

$$\sum_{[i,j]\in R} | f-g | \tag{16.20}$$

$$\sum_{[i,j]\in R} (f-g)^2 \tag{16.21}$$

其中 R 是模板区域。

误差平方和方法是最流行的测量方法。在模板匹配的情况下，这种方法可以间接计算，计算成本也可以大幅度降低。几种测量定义如下：

$$\sum_{[i,j]\in R} (f-g)^2 = \sum_{[i,j]\in R} f^2 + \sum_{[i,j]\in R} g^2 - 2\sum_{[i,j]\in R} fg \tag{16.22}$$

既然假设 f 和 g 是定值，那么 $\sum fg$ 就是一种误匹配测量方法。获取模板所有的位置和情况的合理策略是移动模板，并在图像中的每一点使用匹配测量方法。这样，对于 $m\times n$ 的模板，我们计算：

$$M[i,j] = \sum_{k=1}^{m}\sum_{l=1}^{n} g[k,l]f[i+k, j+l] \tag{16.23}$$

其中 k,l 是对应于模板在图像中的位移。这种算子称为 f 和 g 之间的互相关。

我们的目的是找到是局部最大并且超过某一阈值的位置。然而，当假设 f 和 g 是常量时，上述计算将会引入一个小问题。将这一计算作用于图像上时，模板 g 是常数，但 f 会变化。由于 M 值取决于 f，因此它无法在不同位置上指示出正确的匹配。这一问题可以通过归一化互相关方法来求解。匹配测量值 M 可以使用下式计算：

$$C_{fg}[i,j] = \sum_{k=1}^{m}\sum_{l=1}^{n} g[k,l]f[i+k, j+l] \tag{16.24}$$

$$M[i,j] = \frac{C_{fg}[i,j]}{\{\sum_{k=1}^{m}\sum_{l=1}^{n} f^2[i+k, j+l]\}^2} \tag{16.25}$$

由上式可见，在 $g=cf$ 时，M 在 $[i,j]$ 处取最大值。在图 16.17 中，我们给出了一幅图像，一个模板，及使用上式计算的结果。应该指出，在模板的位置上，我们得到的是局部最大值。

在二进制图像中，上面的计算可用大大地简化。在光学计算中，模板匹配方法是一种非

常流行的方法：用卷积的频域特性来简化算式。

模板匹配的主要局限是模板只能进行平行移动。在旋转或大小变化的情况下，它是无效的。在物体只有部分是可视图的情况下，它也无法工作。

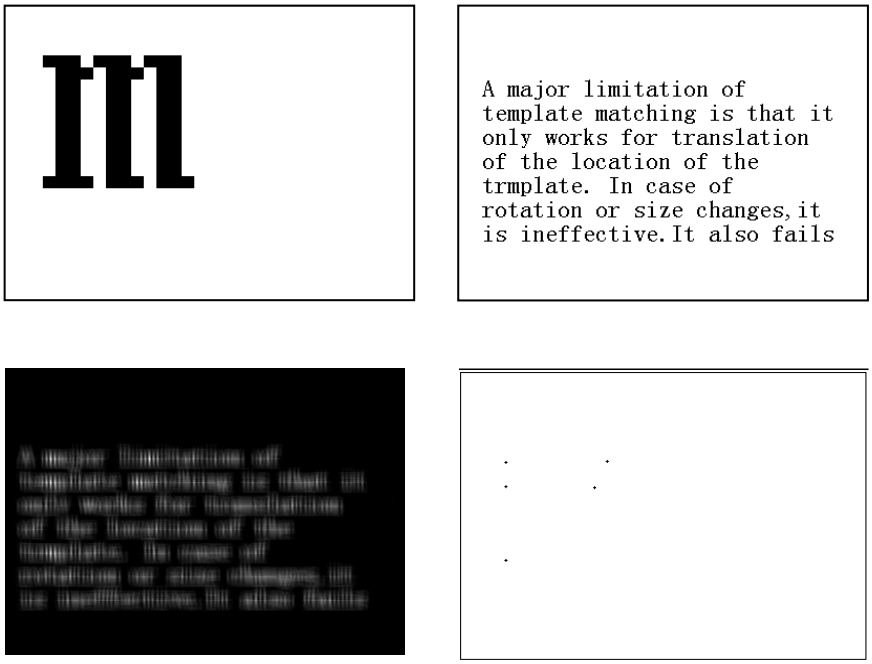


图 16.17　模板匹配实验结果

16．6．2 形态方法

形态方法也可以用来检测模板的存在及其位置。对于二进制图像，使用结构元素作为模板并打开图像，将产生与模板匹配的所有位置。对于灰度图像，可以使用灰度图像形态学。这些结果见图 16.18，(a) 结构元素，(b)一幅图像，(c) 同构开放。

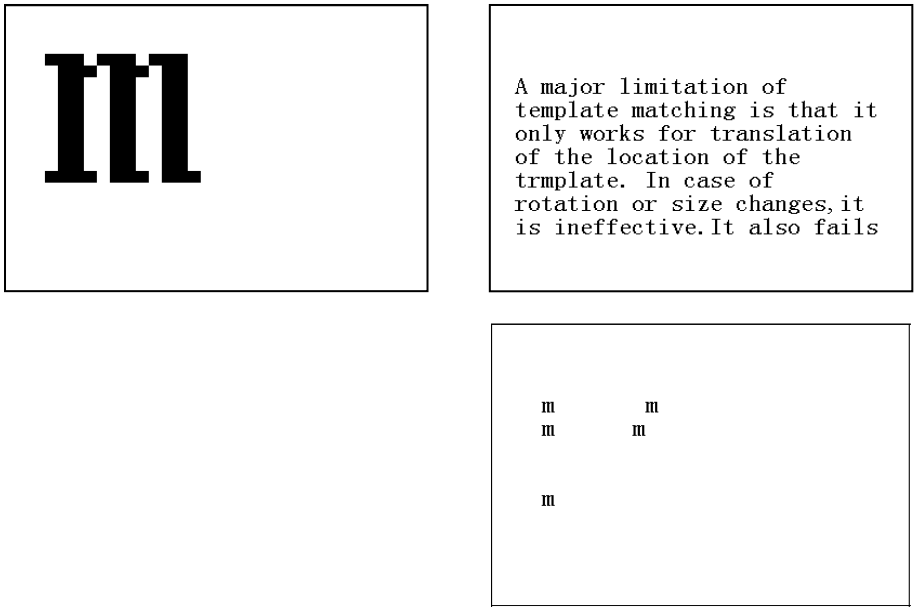


图 16.18 形态方法匹配示意图

16．6．3 符号

如上面所讨论的，如果物体模型和未知物体表示为图形，那么就必须使用一些方法来匹配这种图形表示。在此，我们将定义这些方法所基于的基本概念。

(1) 图形同构性

给定两个图形 (V_1,E_1) 和 (V_2,E_2)，在 V_1 和 V_2 之间找到一个 $1:1$ 的映射（同构） f，那么对于 $\theta_1,\theta_2 \in V_1,V_2$， $f(\theta_1)=\theta_2$，以及对于连结任意一对节点 θ_1 和 $\theta_1' \in V_1$ 的 E_1 中的每一个边缘，有一条连结 $f(\theta_1)$ 和 $f(\theta_2)$ 的 E_2 的边缘。

图形的同构性只用于物体完全可见的情况下。如果一个物体是部分可见，或一个 2.5 维描述与一个三维描述进行匹配，则使用图形嵌入方法或子图同构性方法。

（2）子图同构性

在一个图形 (V_1,E_1) 和另一个图形 (V_2,E_2) 的子图之间找出同构性。

这些方法在用于匹配时存在的问题是图形同构性问题。对于任何合理的物体描述，匹配所需的时间大得不能接受。幸运的是，我们可以使用比图形同构算法所使用的更多的信息。根据节点的性质，这一信息是可以得到的。目前，人们提出了许多启发式方法来求解图形匹配问题。这些匹配方法考虑了如下问题：

- 性能和关系的变化
- 性能和关系的缺乏

- 模型是一类物体的抽象表示
- 情况可能包含额外信息

16．6．4 类比法

两条曲线之间相似性测量可以在同一个参考系坐标下通过比较二者,如图 16.19 所示.并直接计算每一点处二者的差值来实现[Jain 1995]。注意，在图 16.19 中，差值是沿 x 轴的每一点测量的。将总是沿某一轴进行测量的。总的差值是绝对误差值的总和或者是误差平方和。如果没有给定准确的配准，那就必须使用一些基于相关方法的变异公式。

为了使用三维模型识别物体，你可以使用计算机图形学的渲染方法（rendering)来在图像中找出物体的外观，然后同原始图像进行比较，以验证物体的存在。由于用于渲染物体的参数通常是未知的,因此常常考虑三维模型上的一些显著的特征,在图像中检测这些特征,并进行匹配，以验证模型在图像中是否存在。这也导致了研究物体三维表面特性及三维物体投影的理论发展，以确定用于物体识别的不变性。不变性通常是图像中的特征和特性，它们常常对物体的方位和场景照明非常敏感。这些特征在从它们的二维投影中检测三维物体是非常有用的。

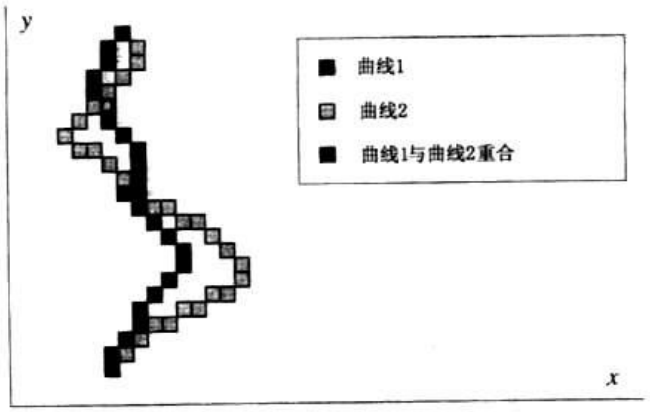


图 16.19 通过直接测量两个实体的误差来实现两个实体的匹配示意图

16．7 物体定位

物体识别通常是指从一幅图像中确定某一已知物体是否存在以及该物体在图像中的位置和方向。人们通常将物体在图像中的位置和方向估计称为物体定位估计(pose estimation)。确定物体在图像中的位置具有重要的实用价值，比如，实现与场景交互作用，分析场景几何关系，描述场景，推理场景等。目前，物体定位算法已经用于改进物体识别算法[Grimson 1991]，比如，通过验证策略(testing strategy)精确验证物体识别假设；也可以用于有效地识别和跟踪时变图像序列中的物体[wheeler 1996,贾 1996]；还可以用于检测和推理有关遮挡问题。

一般视觉定位系统的输入是距离图像(3D)序列和高度图像(ID)序列，而物体的模型一般是三维模型(3D-model)，这样就出现了两种最富挑战性的定位问题：3D 模型在 3D 图像中的定位问题，简称 3D-3D 定位估计；3D 模型在 2D 图像中的定位问题，简称 3D-2D 定位估计。下面介绍这两种定位估计方法。

16．7．1 三维－三维物体定位

3D-3D 定位算法的基本思想是在两组给定的 3D 点集中，寻找对应关系，一组是 3D 模型上的点集，另一组是 3D 图像中的点集。3D-3D 定位问题可以分为两个阶段，第一个阶段是粗定位(rough pose estimation)，第二个阶段是精细定位(pose refinement)。由于这两个阶段定位物体的条件和目标不全一样，因此，所创立的算法也不一样。粗定位是指在深度图像中确定物体出现区域和大致的方向，而精细定位是指给定一幅 3D 图像和一个物体的粗略位置，然后建立 3D 模型与 3D 图像之间的匹配目标函数，最佳匹配对应的 3D 模型定位参数就是物体在图像中的位置和方向。显然，粗定位需要更多的应用领域知识和启发式搜索等算法。目前的部分 3D-3D 物体定位基本上都假设物体的粗略位置是已知的，实际上该位置是由人来确定的。

3D-3D 精细定位可以简单地表述如为：给定模型上的一点和模型的当前位置，在 3D 图像中找出对应点。求对应点的最直接方法是在三维直角坐标空间中求最近距离的点。数学上，3D 模型中的一点 **x** 与 3D 图像点 **y** 的最近距离定义为

$$\mathbf{y} = \arg\min_{\mathbf{y}\in\mathbf{D}} \|\mathbf{x}-\mathbf{y}\|$$

其中，$\mathbf{D}$ 是 3D 图像点集，上式中搜索最近点的理论复杂度为 O(|$\mathbf{D}$|)。

如果图像上一组点 $\mathbf{y}_i$ 与模型上一组点 $\mathbf{x}_i\,(i=1,2,.....,n)$ 的距离都达到最近，则图像与模型对正(alignment)。从模型的初始位置一直到对正位置，实际上是一个刚体变换。刚体变换仍然用一个矢量对 $\langle \mathbf{R},\mathbf{t}\rangle$ 表示，$\mathbf{R}$ 是一个 3×3 旋转矩阵，$\mathbf{t}$ 是一个 3D 平移矩阵矢量。每一个对应相对于定位参数提供 3 个线性约束

$$\mathbf{y_i} = \mathbf{R}\mathbf{x_i} + \mathbf{t} \tag{16.27}$$

通常，3D 数据点补噪声污染

$$\mathbf{y}_{\mathrm{i}} = \mathbf{y}_{\mathrm{i}}^{\mathrm{a}} + \boldsymbol{\beta}$$

其中 $\boldsymbol{\beta}$ 是一个随机 3D 变量，假定 β 服从均值为 0 的正态分布 $(0,\sigma^2)$，则对于 n 个对应点，

求解定位参数变为对最小二乘方误差求极小化

$$e(\mathbf{R},\mathbf{t}) = \sum_i \|\mathbf{R}x_i + \mathbf{t} - \mathbf{y}_i\|^2 \tag{16.28}$$

上式看起来比较容易求解。显然旋转矩阵只有 3 个自由度，因此必须满足

$$\mathbf{R}\mathbf{R}^T = \mathbf{I}$$

$$|\mathbf{R}| = 1$$

其中第一个约束表示 $\mathbf{R}$ 的各列是正交的，第二个约束条件保证旋转变换是刚体变换。在实际中，要考虑这些约束又要使用线性求解的方法有相当的困难，因此，通常使用四元数矢量 $\mathbf{q}$ 来表示旋转变换(见第十二章)，$\mathrm{q}=(\mathrm{q}_{0,}\mathrm{q}_1,\mathrm{q}_2,q_3)$，其中 q_3 是一个标量，这样，刚体变换可

用 7 个矢量 p 来表示

$$\mathbf{p}=\left[\mathbf{q}^T,\mathbf{t}^T\right] \tag{16.29}$$

[Sanso 1973]首先将四元数表示用于摄影测量领域 3D-3D 定位问题，然后由[Faugeras 1986] 引入计算机视觉领域求解物体定位问题。使用四元数表示求解式(16.28)可以得到解析解，研究这一工作的还有[Horn 1987]、[Haralick 1989]和[Arun 1987]。

(1)M-估计

如果观察数据的误差不服从正态分布，则最小二乘法误差估计方法就不适用上述定位参数的求解，此时，可以使用 M-估计算法[Haralick 1989]， M 是指最大似然估计(Maximun likelihood estimation)。M-估计算法是一种鲁估计算法，其最一般形式为

$$E(z)=\sum_i \rho(z_i) \tag{16.30}$$

其中 $\rho(z)$ 是关于误差 的任意函数， $E(z)$ 的等价概率分布函数是

$$p(z)=e^{-E(z)} \tag{16.31}$$

这样，M-估计是 $p(z)$ 的最大似然估计。

如前所述，最小二乘估计对局外点十分敏感。最小二乘估计对应于 $\rho(z)=z^2$ 的 M-估计是

$$p(z)=e^{-\sum_i z_i^2} \tag{16.32}$$

相对于 p 求 E 的偏导数并置偏导数等于 0：

$$\frac{\partial E}{\partial p}=\sum_i \frac{\partial \rho}{\partial z_i}\frac{\partial z_i}{\partial p}=0 \tag{16.33}$$

令 $\omega(z)=\frac{1}{z}\frac{\partial \rho}{\partial z}$ 则有

$$\frac{\partial E}{\partial p}=\sum_i \omega(z_i)z_i\frac{\partial z_i}{\partial p} \tag{16.34}$$

$\omega(z_i)$ 是一个权重系数，当使用纯最小二乘方估计时， $\omega(z)=1$ ，即每一个误差值具有相等的置信度，而与误差值大小无关。为了避免局外点对估计的影响，可以使用如下阈值化条件

$$\omega(z)=\begin{cases}1 & |z|\le\theta\\ 0 & |z|>\theta\end{cases} \tag{16.35}$$

即当某点测量误差大于阈值 θ 时，就忽略该点。关于 $e(z)$ ，还有其它几种函数可供选择，

比如 Lorentz's 函数[Press 1991]等

$$\omega(z)=\frac{1}{1+\frac{1}{2}(\frac{z}{\sigma})^2} \tag{16.36}$$

(2)精确定位鲁棒法

(16.30)式可以重新写为

$$E(p)=\frac{1}{|v(p)|}\sum_{i\in v(p)}\rho(z_i(p)) \tag{16.37}$$

$v(p)$是一组模型点(相对于观察者方向是可见的)，$z_i(p)$是第 i 个对应点对之间的 3D 距离，定义为

$$Z_i(p)=\left\|\mathrm{R}(q)x_i+\mathrm{t}-\mathrm{y}_i\right\|^2 \tag{16.38}$$

$$\frac{\partial z}{\partial p}=\begin{bmatrix} z(x_i+\mathrm{t}-\mathrm{y}_i \\ -4x_i(\mathrm{t}-\mathrm{y}_i) \end{bmatrix} \tag{16.39}$$

上式建立了表示旋转和平移矢量与误差梯度之间的关系。这样，首先在初始位置上计算误差函数 E 的梯度方向$-\nabla E(p)=dp$，然后在梯度方向求目标函数极小值对应的位置$p+\lambda dp$，再求新位置的误差函数值，这样一直迭代下去。直到前后相邻两个位置对应的误差函值小于某一个预定值为止，图 16.20 是使用上述算法的实验结果[wheeler 1996]

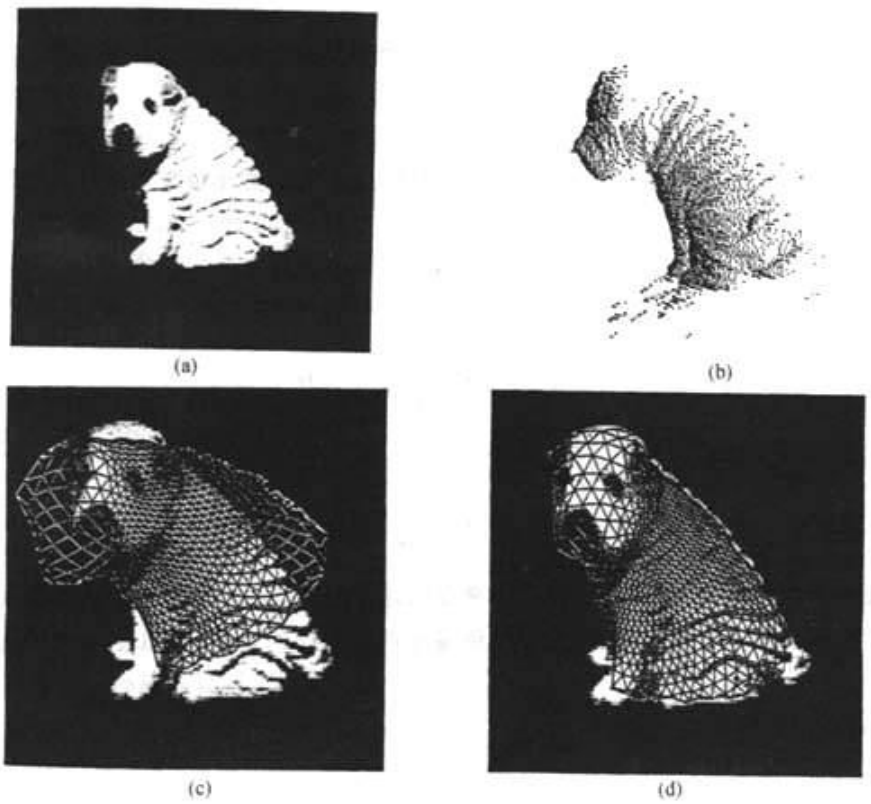


图 16.20　玩具狗 3D 定位实验结果

(a)原始高度函数　　(b)原始距离图象

(c)模型初始位置图(20mm 平移，30 度旋转) (d)最后位置估计结果

16．7．2 二维－二维物体定位

上一节讨论的 3D-3D 物体定位是在一幅距离图像中，用 3D 模型表面点去匹配图像 3D 点。本节讨论的 3D-2D 物体定位是在一幅亮度图像中，用三维物体模型与二维图像特征点进行匹配，显然 3D-2D 物体定位是一个不适定问题，其求解方法与 3D-3D 完全不同。在 3D-3D 定位中,物体 3D 模型是由表面三角片组成,匹配中图像 3D 点与 3D 模型点的最近距离实际上是图像 3D 点与 3D 模型三角片之间的最近距离。在 3D-2D 匹配中，由于输入的是一个亮度图像，只包含了物体某一个朝向(aspect)的亮度分布，而我们可用的匹配特征则是物体在此朝向时表面处的边缘特征。因此，用于 3D-2D 定位的物体模型必须包含物体边缘特征，这样，3D 模型匹配与亮度图像的匹配就变为 3D 模型的边缘与亮度图像中的边缘之间的匹配。

由上述的讨论可知，3D-3D 定位只涉及到物体固有的几何特征——3D 几何形状，并且匹配空间和数据空间都是 3D 空间，而 3D-2D 定位不仅涉及到受许多其它因素影响的亮度函数，而且匹配空间、数据空间维数不相同，因此，3D-2D 定位要比 3D-3D 难得多。

16．8 神经元网络

神经元网络方法已经用于物体识别任务。神经元网络可以实现物体的分类方法。其吸引力就在于使用类别的非线性边界来划分类别特征空间的能力。这些非线性边界可以通过网络的训练来得到。在训练阶段，需要示意许多待识别物体的许多情况。如果训练集在识别阶段得到仔细选择，以便将以后碰到的所有的情况都表示出来，然后，网络在特征空间对分类边

界进行学习。在识别阶段，网络同其它分类器就完全一样了。

神经网络最有吸引力的特点是使用非线性边界的能力和学习的能力。最大的局限是无法引入关于应用领域的已知事实以及调试操作时的困难。

思考题

16.1 列出物体识别系统的主要组成模块，并讨论它们在识别任务中的作用。

16.2 什么是朝向图？请阐述使用朝向图识别物体识别的过程。

16.3 什么是特征空间？怎样使用特征空间识别物体？

16.4 神经网络最吸引人的特点之一是它们的学习能力。它们学习的能力在物体识别中是如何使用的？哪种模型可用神经网络？你如何介绍你关于神经网络中的物体知识？

16.5 讨论模板匹配。在哪种类型的应用中你可用用模板匹配？模板匹配的主要局限是什么？

16.6 用三角面画一个 4 面体的面图。

16.7 模板 g 和图像 f，如下图所示，用归一化的相关方法匹配，求：

(1) 相关数 C_{fg}

(2) $\Sigma\Sigma f^2$

(3) 归一化相关数 M[i,j]

计算机练习题

16.1 利用一个物体识别系统从其部分视图中识别物体。图像中的物体是来自于一个大约 10 个物体的组，其中物体常可以在办公室场景中找到。只选择差不多是二维的物体(硬币、钥匙、垫子、商业卡片等)。考虑把摄像机放在桌子上 8 英尺高的地方。用多个随意的图像，其中这些物体以不同方式出现，来测试你的系统。

16.2 继续上面的例子，如今考虑三维的物体(如鼠标，订书机等)重新设计和重新使用原型物体识别系统。本系统应能从其部分视图中识别三维物体。

16.3 假设在你的模型库中有大量的物体。重新设计你的系统以有效地完成大量物体的识别任务。