

第三章 基于推理的视觉理解

3.1 基于推理的视觉理解概述

3.1.1 逆向和不适应问题

3.1.2 感知组织现象

3.1.3 视觉识别和推理

3.2 感知组织的基本原理

3.2.1 根据图象关系的显要性进行聚类

3.2.1.1 图象关系非偶然性产生的概率

3.2.1.2 限制计算的复杂性

3.2.2 通过求能量极小进行聚类

3.2.2.1 计算策略

3.2.2.2 感知聚类中的表象

3.2.2.3 聚类能量的计算

3.2.3 根据图象特征推论三维空间结构

3.3 景物结构的模型

3.3.1 部件模型和自然形状表示法

3.3.2 部件识别理论 (Recognition-By-Component, RBC)

3.1 基于推理的视觉理解

在研究基于推理的视觉理解以前先要对视觉信息处理和理解中的一些主要特点和特性有基本的了解。其中最主要的是要认识到视觉问题从本质上说是不适应的 (ill-posed)，如果没有附加信息就不能找到解答。这在很大程度上是由于许多视觉任务从本质上来说是逆向的问题。例如，图象是三维景物的二维投影，如果已知景物的三维模型，在一定的几何条件下通过投影 (正向过程) 产生的图象是唯一的。但是同一幅图象可以是无穷多种三维物体的投影的结果。因此根据图象不可能唯一地确定它是什么景物投影的结果。例如，图 3.1 中所示的 M 形图象，它可以是由图 3.1(a) 中的空间某 M 形物体投影的结果，但它同样也可以是图 3.1(b) 中所示的三根在空间互不相交的一些空间曲线投影的结果。

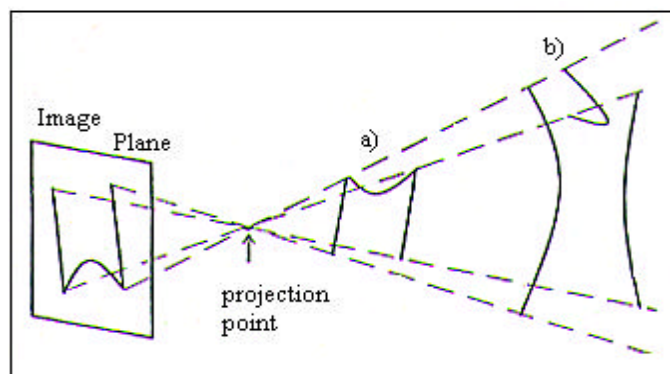


图 3.1 不同的形体产生相同的图象

但是当人们看到(a)中所示的图象时，并不感到它是多义的，这说明人类视觉在理解这

样的图象时要么是利用了附加的高层信息（例如，知道图中是哪一类物体的知识），要么是利用了某些可以去除多义性解释的通用约束。心理物理学研究的结果表明人类同时使用了这两种策略，但令人惊奇的是在消除多义性方面，高层知识提供的信息要比低层的通用约束提供的少。这说明人类视觉在理解图象时必定利用了某些通用的约束。当然这些通用约束中包括景物和物体成象的基本物理性能，但还不止这些。其中很可能还包括所谓的感知组织（perceptual organization）。如果不对这些约束有透彻的了解，要建立通用的视觉系统是困难的。以下对逆向和不适应问题，以及感知组织作简要介绍。

3.1.1 逆向和不适应问题

通常逆向问题涉及在只知道输出和所进行处理的本质的条件下，重构由于某种操作所畸变的数据。例如，已知 $x+y=17$ ；求解满足上述公式的 x 和 y ，就是一个逆向问题。这时所进行的处理是已知的（加法），操作输出是已知的（17），但其输入 x 和 y 是未知的。许多低层计算机视觉问题是逆向的，因为这些问题要求根据从三维到二维变换后的输出（例如灰度图象），重构三维形状、边缘位置或运动等。

1. 适定性定义

1923 年 Hadamard 定义当一个数学问题满足下述条件时是适定的：

(1) 存在一个解；(2) 这个解是唯一的；(3) 解连续地取决于输入数据，即输入数据的很小变化对输出也只引起一个小的变化。

虽然许多逆向问题是不适定的，如上述简单例子所示。这是因为在这些问题中存在许多可能的解，因此不满足条件(2)。此外，实际中只能得到观察数据，而不是实际数据，所以不能保证满足(3)，因此需要附加的信息来使这问题成为适定的。附加信息可以是定量的，这就得到近似解（quasisolution）。另一种方法是使问题正则化，这时需要应用诸如解的特征这样的定性信息。

2. 正则理论

正则理论提供一个解决不适应问题的框架。这涉及通过扩展问题的定义来限制解的空间，具体方法是增加体现解所希望特性的约束。这些附加的约束被称为稳定函数（stabilizing function），设： A 是已知线性算子， y 是已知数据， g 是要恢复的未知数据。逆向问题定义为恢复 g ，使得：

$$y = Ag$$

通常， y 是由测量得到，但伴随着噪声。所以，只知道它的近似值 y^* ，因此，所能得到的是由下式求得近似解：

$$y^* = Ag$$

这个近似解可通过寻找使得下式为极小的 g 来求得：

$$\|Ag - y^*\|^2$$

其中 $\| \cdot \|$ 是合适的模。设： g^* 是体现解的所希望特性的通解。因此，所希望的解 g 应使下式为极小：

$$P(g) = \|g - g^*\|^2$$

$P(g)$ 被称为稳定函数（stabilizing function）。

设： $P(g)$ 至少是半单调的（quasimonotone），把上述两条件综合起来，求解就可表示为寻找使下式为极小的 g

$$\|Ag - y^*\| + \lambda_p [P(g)]$$

其中 λ_p 是控制上述两项相对重要性的正则化参数。这里第一项描述所求解经变换后的符合程

度，第二项说明所求解与所希望的解的特性的接近程度。通过应用稳定函数，在某种程度的数据范围内，可保证解的唯一性和强制所得结果连续地取于所观察的数据。这样就使问题成为适定的并在现在这种极小化问题形式下可求解。

3.1.2 感知组织

感知组织是指人类视觉系统所具有的在不知道图象内容的先验知识条件下，从图象获得相对的聚类和结构的能力。例如，人能从随机分布的图象元素的背景中迅速地检测出对称性、共线性、平行性、连通性和重复纹理等特性。对感知组织研究的全盛时期是在 1920~30 年代，这个时期在感知研究中占主导的是 Gestalt 理论。这个理论主要的研究内容就是感知组织现象。Gestalt 学家的基本原则是整体要大于部分之和。例如，纸上画的两个点具有每个点单独时所没有的方向性。所以方向性被认为是形成特性（emergent property）。因此，Gestalt 研究如何把简单的敏感输入组织或聚类成为复合的稳定感知。这样的感知表现出形式和结构。Gestalt 这个词本身的含意就是整体（whole）和结构（Configuration）。

Gestalt 心理学家对我们理解感知组织的主要贡献是进行了大量的感知组织现象的验证，并把它们进行分类（见图 3.2），Gestalt 心理学家认为下述这些因素在结构的感知中起重要作用：

- (1) 接近性（Proximity）：较为接近的元素倾向于聚集在一起；
- (2) 相似性（Similarity）：颜色、方向、或大小这样的物理属性相似的元素相聚集；
- (3) 封闭性（Closure）：曲线段在形成完整曲线时有形成封闭区域的倾向；
- (4) 连续性（Continuation）：位于同一条直线或平滑曲线上的元素相聚集；
- (5) 对称性（Symmetry）：任何横向对称于某个轴的元素相聚集；
- (6) 熟知性（familiarity）：我们经常看到它们在一起的元素相聚集。

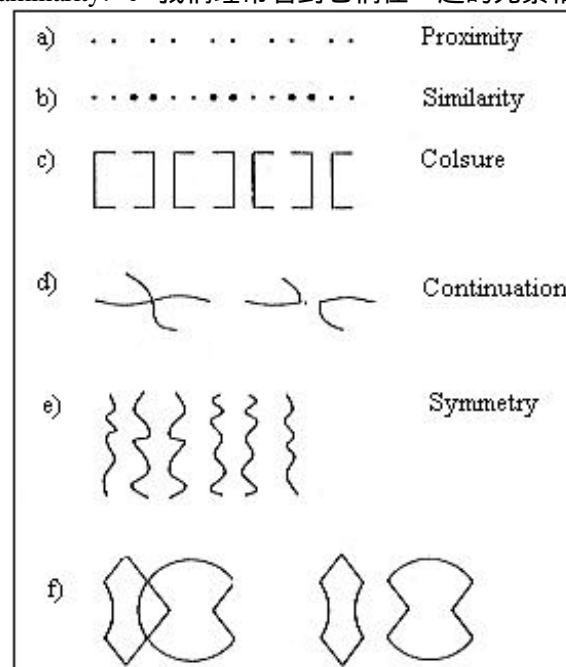


图 3.2 聚类现象的分类

此外还有如相同方向上以同一速度移动的点、同一运动结构的各部分应被聚成一类。Gestalt 心理学家研究的一种影响是引起了对感知组织基本原理的研究。初始的 Gestalt 学家对这方面的研究不很成功，他们把组织的规则归结为称为完形（pragnanz）的单个规则，Pragnanz 这个德文字的含意是“简单”或“完善”的形式。但这只是一个没有定量公式的

转圈的定义。总的来说 Gestalt 理论只是描述性的，而不是定量的。例如，根据相似性或熟知性这样很通用的术语是难以得到定量的理论的。因此当代的感知组织研究是要寻找聚类因素的定量和客观的测量。例如，Hochberg 和 Brooks(1960)提出了直线线画图中角的数量能很好地反映形象的复杂性，和所感知物体的立体性。Hochberg (1981) [Hoc 81]和 Rock(1975) 提出简单性原理，即通常被感知的是要求最少量信息来规定的组织。但不幸的是简单性本身也是不适定的，因为描述一个形象的简单性的程度完全取决于所用的描述语言。所以，Lowe 和 Binford(1982)[Low 82] 以及 Witkin 和 Tenenbaum(1983)[Wit 83] 提出了非偶然性概念 (non-accidentalness) 。即所感知的是最不可能由于偶然机会产生的形状。McCafferty(1990)[McC 90]提出了用求能量极小进行聚类的方法。他们都提出一种可通过计算来对聚类过程进行评价和衡量的方法。这两种方法将在下一节中分别介绍。此外值得注意的是某些新发现的感知组织现象，这些现象有助于加深对感知组织的理解。其中一个是由 Kanizsa (1979) [Kan 79]提出的形状完整性 (Shape Completion) 或错觉轮廓 (illusory Contour)，图 3.3 上所示为这种现象的一例子。从图上人们通常可以感觉到一个把四角分别放在四个圆上的正方形，而不是四个单独的圆。这个感觉到的正方形大部分是由在物理上并不存在，但是可感知的轮廓构成。这个现象是重要的，因为这表明感知组织可以通过并无直接物理起因的推理输入产生，与此类似的还有由 Glass(1969)和 Stevens(1978)提出的虚拟线 (Virtual lines)。

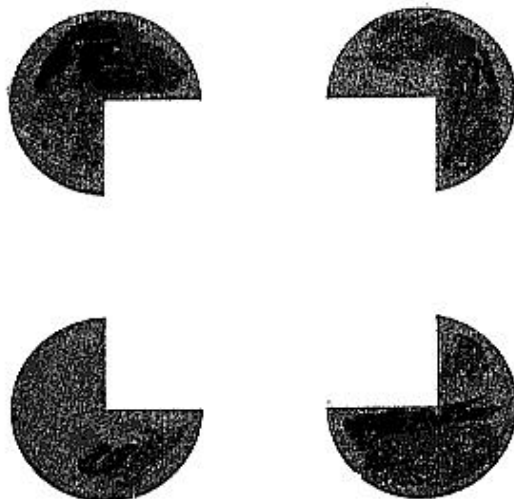


图 3.3 错觉轮廓

还有一个不清楚的问题是感知组织的各个方面与立体视觉融合或物体识别这样的其它视觉处理过程之间的关系。这个关系是就它们在视觉通路上的相对位置而言的。Marr (1982, 1976) [Mar 82]提出：基于边缘点的相似性之上的聚类和线段的连接，发生在产生完全的初始简图的过程中，这说明聚类发生在根据各种视觉线索恢复形状的处理过程以前。对这个观点 Marr 没有给出什么心理物理学或心理学的证明，他只是给出了一些例证。

图 3.4 是一个例证，说明感知组织如何能为根据影调恢复形状的分析提供标记。图上所示的是一些分布在黑色背景上的非黑色离散点，这些非黑色点的灰度是按照一个圆柱图象的灰度分布曲线来确定的。这些非黑色点被聚类，以恢复原来的灰度分布情况，并呈现出圆柱的三维形状。这表明在进行形状分析以前单独的点必须以某种方式被聚类成单元，这是因为，正是由这样的聚类产生的区域而不是单独点本身产生三维形状。

总之，有证据表明感知组织的不同方面是在视觉通路的不同阶段上进行的。这说明感知组织是视觉通路上的许多阶段都涉及的功能，这个功能的目的是试图发现所通过数据中的结构。还有一个重要的问题是低层视觉处理输出应采取的形式。这个问题很重要，因为聚类处理是递归进行的，这点可由图 3.4 所示来证明，其中非黑色的像素被聚类成块，块又被聚类成圆周。聚类处理可以是在另一种聚类处理输出的基础上进行的事实说明，在所有的聚类过

程中应使用相同的表达。

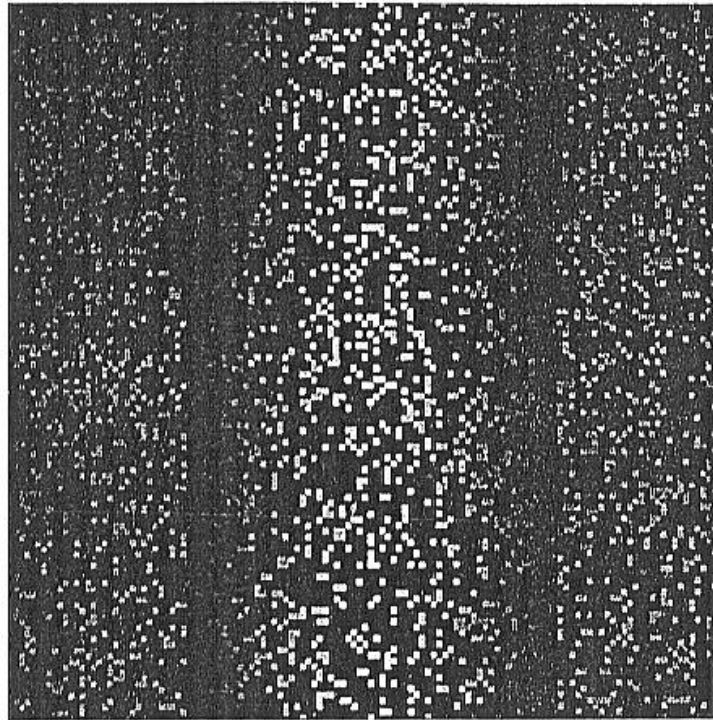


图 3.4 随机点图案，其中非黑色点的灰度是按圆柱图象的灰度分布取的。

1.3 视觉识别与推理

在计算机视觉中解决视觉任务的方法可分成两大类：基于重构（reconstruction）的方法和基于识别（recognition）的方法。基于重构的方法试图根据图象恢复和重构外部视觉环境的物理参数。例如，物体表面的深度或方向，物体的边界，以及光源的方向等，Marr 的视觉计算理论基本可以归入这一类。基于识别方法的目标是物体的识别和描述。识别是指发现图象中的成份与环境中的物体的先验表示之间的对应。所以就识别而言，并不需要重构环境完整的物理参数，在识别过程中关于环境的先验知识将起极其重要的作用。我们在日常生活中可能看到的物体和景物数量巨大、种类繁多，如果没有这些先验期望的约束作用，许多视觉问题就可能因为约束不充分而不能被解决。识别使我们能超越图象中的数据，因为我们可根据小部分预期的对应达到可靠的识别，然后应用知识来推论由视觉数据没有直接提供的景物的特性。这说明视觉信息处理中在尽可能早的阶段中应用知识的价值和必要性，也说明视觉理解可以通过推理来完成。因此基于识别的方法也可以称为基于推理的方法。

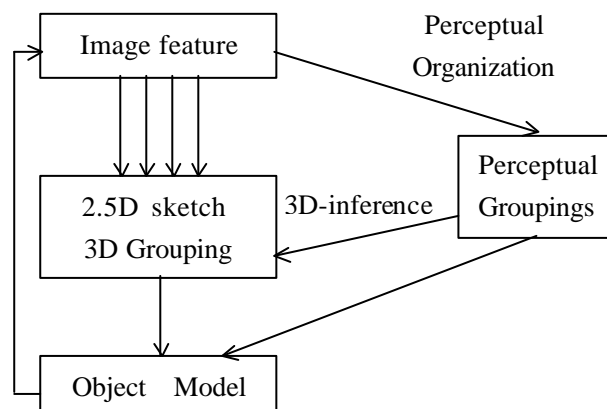


图 3.5 Lowe 的视觉识别模型

识别可通过多种预测的特性与实际检测特性之间的对应来实现，其中包括形状、颜色、纹理、连通性、上下文、运动或影调。这里要强调的是空间对应性，即图象中特征的检测位置准确地与某种物体特征在特定的投影下的位置相吻合。

Lowe^[Low 85]提出了进行视觉识别的模型，如图 3.5 所示。在此模型中除了经过深度和表面表象的通路以外，还有由所谓的感知组织原理形成的通路。感知组织可以从二维图象特征直接形成，并且可以用作基于搜索的识别过程的输入。至于图象解释正确性的验证也可直接通过校验三维知识和图象二维位置之间的一致性来完成，而不需要通过深度表象。

感知组织和推理在视觉理解中的作用可以用以下心理物理学实验来证明，在此实验图象中，为进行感知组织的必要信息被故意丢失了。如图 3.6 所示，图中是一幅自行车的线画图，但是它只完成了大约 50%，并且使可进行自底向上聚集的大多数可能都被隐去（即隐去了大多数显著的共线性、平行性、对称性和靠近的端点等）。实验证明，当被试验者对



图 3.6 当自底向上的图象特征聚类机会被排除时，线画图就难以识别

图中的内容并无先验知识时，要识别这幅图是相当困难的，在一个有 10 位试验者的小组中有 9 位在 90 秒的时间内没能识出物体，第 10 位试验者则花了 45 秒完成识别。我们可以在图中逐步地增加进行感知组织的线索，并继续进行实验，以观察这时能否使识别所需时间缩短，在图 3.7 中只对图 3.6 增加了一根线条，但它的位置具有战略的重要性，使它能与其



图 3.7 增加一条提供聚类证论的线就使识别变得容易

它的线条聚集成为共线曲线。这第二幅图所需的识别时间就大为降低。10 名试验者中有 3

名可在 5 秒以内完成识别，其余 7 名可在 60 秒以内完成识别。从可以通过控制形成感知组织来影响识别时间说明这个过程具有搜索的本质。在形成初始的感知组织聚类产生了有关线条可形成一个圆形曲线的假设以后，使自行车的识别问题迎刃而解，说明了这是一个由假设—检验组成的推理过程。

Kanizsa^[Kan 79]提出形状的感知可分成两个阶段：初始阶段，把视觉输入聚成具有空间和时间规则性的区域；第二个阶段是完整性、完全性和集成性的感知推理，这个阶段使我们能超出传感数据直接给出的信息范围，填补实现感知所缺少的信息。

3.2 感知组织的基本原理

Gestalt 心理学家证明和强调了感知组织在视觉中的重要性，他们虽然认识到了在感知组织计算中应该计算什么，但是并没有令人信服地解答为什么要进行和如何进行这样的计算的问题。也就是没有解决感知组织的基本原理的问题。对感知组织基本原理的研究要解决的问题是聚类过程的目的是什么，以及如何进行聚类过程。对此问题研究者已进行了大量的工作。下面我们介绍其中有代表性的两种观点：(1) Lowe^[Low 85]和 Witkin 等认为聚类过程的目的是发现图象元素之间的因果关系或非偶然性关系。图象之间的关系不太可能是由于偶然因素产生的程度决定了元素之间关系的显要性 (Significance)，因此感知组织可被看成是对图象特性的每一种可能的聚类赋以显要性的过程。(2) McCafferty^[McC 90]认为和大多数逆向问题一样，在其原始状态下聚类是一个不适定问题。可以通过引入稳定函数来限制解的空间，这些稳定函数表征了所期望的解的性质。感知组织的 Gestalt 规则可用来描述人类视觉系统具有的聚类特征，因此可通过引入稳定函数来实现 Gestalt 规则。按照正则理论，这时聚类就成为一个求能量极小的问题。在以下的章节中，我们将分别讨论这两种聚类方法。

3.2.1 根据图象关系的显要性进行聚类

Lowe 认为可把图象中元素之间的关系分成两类：一类是由于偶然巧合的视点或位置产生的关系；另一类是由景物中的某些有意义的 (即可预测的) 关系所产生。例如，图 3.8(a) 和 (b) 中所示的由三个点所组成的两种关系。在 (a) 中三个点组成了等间距的共线关系，当视点在相当大的范围内变化时，这样的关系都可以得以保持，因此当看到图象中的点之间具有等间距的共线关系，可以推测它们是空间等间距共线点的成象。这样的关系被认为是因果的、非偶然的。与此相反 (b) 图中三个点形成的是等边三角形的关系，如果三个点在空间形成等边三角形的关系，那么只有在某一特定视角下，它们的图象才能保持等边三角形的关系。同时对于空间中的任意三个点都存在一个特殊的视角，在该视角下这三个点在图象平面中的投影是等边三角形。因此三个点之间的等边三角形关系就不是因果关系，而是偶然关系，这时我们就不能从图象中点之间的等边三角形关系推论它们在三维空间中也保持这样的关系。因此 Lowe 和 [Wit & Ten 83] 认为：(1) 聚类过程的目的是发现图象元素之间的因果关系 (Causal relations)，或非偶然性关系 (non-accidental relations)，这些关系在以后的解释过程中原封不动地保存下来，解释过程中的许多处理只是给原始的聚类加上标志。所以，从图象恢复三维结构的主要计算工作已由聚类过程完成。例如，在上述例子中三个点之间在空间形成的等间距共线关系在聚类过程中已经被推论得知。(2) 图象中元素之间关系的显要性 (Significance) 取决于这种关系不太可能是由于偶然性因素产生的程度。例如，在图象中平行曲线关系被认为是高度显要的，这不是由于投影产生平行曲线结构的机会比不是这种结构的多，而是因为两条不平行的曲线通过投影成为平行可能性很小。(3) 图象关系的显要性也提供了进行聚类的方法。感知组织可被看成是对图象特征的每一种可能的聚类赋以显要性程

度的过程。

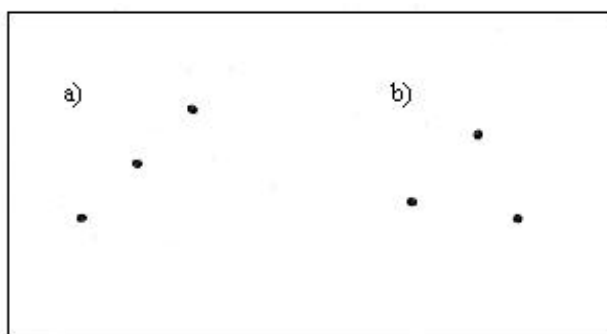


图 3.8 (a)三点成等间距关系 (b)三点成等边三角形关系

3.2.1.1 图象关系非偶然性产生的概率

Lowe 认为每种图象关系都包含了表示这种关系是非偶然性产生的统计信息，并且正是这种非偶然性的程度形成了对它赋以显要性程度的基础。由于在图象中可能存在的图象关系类型有无穷多种。（例如，对任何给定的 N ，图中所有的“直线”对之间形成相对角度为 N 度的关系。）在这些可能的关系中只有一小部分是显要和值得检测的。而确定一种关系是否值得检测的关键就是要计算这种图象关系是非偶然性产生的概率（probability of accident occurrence），以下我们来研究这个问题。

在计算图象关系的非偶然性概率时要考虑以下多种因素：

(1) 关于图象投影过程的知识告诉我们，只有某些类型的图象关系通常不是由于偶然性产生的。因此从统计上讲，只有这样的关系是可探测的；(2) 在作非偶然性统计估计时可利用对于每种关系产生概率的先验知识；(3) 在建立图象关系事件的模型时可假设图象元素的位置和方位是独立的；(4) 对任何关系来说，相似特征的背景密度决定了某个给定程度接近性的显要性；(5) 初始的关系可递归地组合成新的关系，此新关系可影响原始估计的显要性。

1. 视点恒常性条件

景物到图象投影过程的性质向我们提供了一种对图象关系强有力和通用的约束信息。如果我们假设摄象机或眼睛的视点是与景物中的物体相独立的，那么可以证明只有某些类型的图象关系通常不是由于偶然因素产生的。这些类型的图象关系的特点是它可以在视点的一个范围内保持稳定。例如，景物中曲线的共线性可在相当宽的视点范围内投影形成图象中的共线性。任何在投影时产生的关系，如果不能在多数视点范围内保持稳定的话，就难以把这样的关系与由于偶然性产生的关系相区分。例如，景物中成直角的直线对在大多数视点下都不能投影为图象中成直角的直线对。因此，虽然我们在图象中检测到了成直角的直线对，我们也没有理由相信，这不是由于视点与某个未知空间角度相配合偶然产生的结果。

视点恒常性约束极大地限制了可作为感知组织基础的图象关系类型。只有少数类型的关系，例如，共线、连接性可以在所有的视点下保持恒常。不过，还有另外一些类型的关系，可以在相当大的范围内加以保持，因此可被认为是经常出现的。例如，平行和一系列共线特征的等间距性在透视效应的情况下，仍能在相当大的视点范围内保持。此外，由于许多物体只占据较小的视角，或者与观察者到物体的距离相比，物体本身的深度范围较小，我们还可认为其它一些关系在图象中是经常产生的。对这样的关系在使用时要小心处理。例如，在投影时曲率恒常性不能严格地保持，但对于在曲线只占较小的径向角度范围的局部区域内，基

本上可说曲率保持不变。尽管有上述这些复杂性，视点恒常性约束仍然是一种非常有用的工具，它可用于把几乎无穷的图象关系局限到少数几种候选关系，这些关系可在投影的条件下至少可以部分保持恒常性。

在视点恒常性的基础上检测图象关系的重要优点是，这样检测到的图象关系意味着它们是某个特定的空间关系投影产生的。因此就有可能根据图象关系来推论相应的空间结构。例如，如果我们已确定图象内的若干特征的共线关系不是由于偶然因素产生的，就可以推论这些特征在空间也是共线的。这个问题在下面的章节中研究。

2. 关于图象关系产生概率的先验知识

上一节讨论的视点恒常性约束是确定图象关系是否是偶然因素产生的主要因素之一，但与此同时还需考虑与图象内容有关的先验知识。当决定先验知识在判断图象关系的非偶然性时可以应用条件概率和 Bayesian（贝叶斯）推理：

设， $p(r \& a)$ 是事件 r 和 a 都是真的概率， $p(a|r)$ 是当 r 事件为真时，事件 a 的概率。因此有：

$$P(r \& a) = P(r)P(a|r) = P(a)P(r|a)$$

所以

$$P(a|r) = \frac{P(a)P(r|a)}{P(r)}$$

这就是基本的 Bayesian 公式。如果 r 是以某种精确度测得已知图象关系的事件； a 表示图象关系是偶然性产生的事件； c 表示图象关系是因果性产生的事件。那么 $P(r) = P(a) + P(c)$ （因为 a 和 c 是 r 的两种相互排斥的情况）和 $P(r|a) = P(r|c) = 1$ （因为 a 和 c 是 r 的实例），因此，根据 Bayesian 公式可得

$$P(a|r) = \frac{P(a)}{P(a) + P(c)}$$

$$P(c|r) = 1 - P(a|r) = 1 - \frac{P(a)}{P(a) + P(c)}$$

以上公式使我们能根据偶然事件和非偶然事件的先验概率来计算给定的图象关系是非偶然性的概率。对 $P(a)$ 的估计问题将在以下章节中讨论，前一节中讨论的视点恒常性条件的目标是选择 $P(c)$ 显著高的关系，但对定量估计图象关系的因果性概率来说，视点恒常性只是其中的一个因素。如何来确定 $P(c)$ 呢？一种可能的方法是通过统计的实验方法，另一种较为理论的方法是先建立视觉世界的某种通用模型，然后根据这个模型得到这种图象关系的出现概率。当然对 $P(c)$ 的估计并不需要很准确，数量级的估计就可满足应用的需要。

3. 位置独立性假设

给定以某种精度保持的图象关系，要计算这种具有一定精度的关系是偶然产生的概率，我们就必须对物体周围的分布情况作某种假设，以此为背景来判断关系的显要性。一种最通用和显然的假设是认为背景中的物体位置相互独立，由此可知在图象的背景中，物体位置也是相互独立的。这被称为位置独立性的空假设（null hypothesis）。

已知三维空间中位置和方向的独立性假设以后，就很容易计算具有给定精度的某种关系是偶然产生的概率。例如，如果两条直线平行，其平行的精度为 5° 以内，那么可算出这样的关系是由于两个独立物体偶然产生的概率是 $5/180 = 1/36$ 。

4. 背景特征密度与接近性之比

以上研究了单独给定关系的情况，当在图象中同时存在多个图象特征时，需要研究的图象关系数量就与特征数量的平方成正比。例如，已知图中有 10 条线，那么可能的线段对的数量就有 $10 \times (10-1) / 2 = 45$ 条。不难想象，可以从中发现一些相互平行的线段对。图 3.9 中的例子表示了这一点。

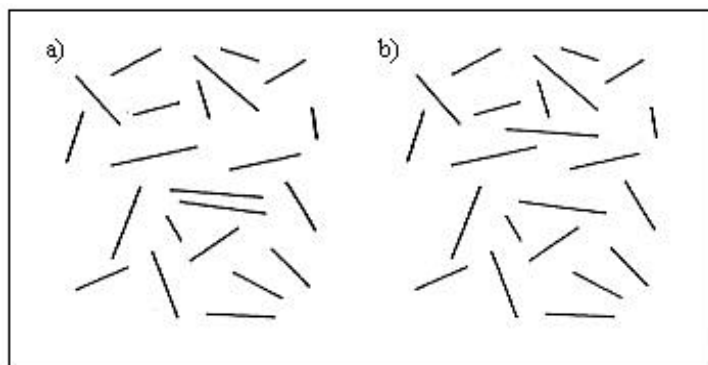


图 3.9 两条几乎平行的线段在图(a)中形成了一个显要的关系，因为这时与背景中的相似特征的密度相比，这两条线相互更为接近。但当接近性与密度之比下降时，这两条平行线段的显要性就减弱了，这说是图(b)所示的情况。

如果把背景的密度考虑进来，那么构成关系的特征之间的接近性就成为判断关系是否显要的一个主要因素，当两个特征之间的距离变得更近时，在给定的背景特征密度情况下，具有相同接近程度的其它特征的数量就急剧下降。

请注意，接近性不但是在判断其它类型图象关系显要性的一个因素，而且它本身也是一种可用于检测的非偶然性的图象关系。在空间是相互接近的特征，在各种视点下都将投影为图象中相互接近的特征，所以接近性可以通过视点恒常性试验。

5. 递归地进行构造

由于图象测量的精度有限，所以至今所讨论的简单图象关系通常不能产生很低的偶然性概率。因此难以作为识别关系的可靠证据，但是可以通过组合初步得到的关系来建立新的图象关系，这些组合图象关系的偶然性概率就会低得多。例如，我们可以把若干个共线的点聚类成线，然后又把这条线进一步组合进平行线这样较大的结构中。这些后来形成的结构对较早的聚类提供了确认。这样的过程可以一直进行下去，直到识别物体。在前面所举的识别自行车的例子中可以看到，识别了自行车就是对前面的初始聚类的强有力的确认。

3.2.1.2 限制计算的复杂性

前面已研究了若干确定某一图象关系是否是偶然性产生的因素。但是，在有些情况下人类视觉却不能检测按任何合理的统计准则来衡量都是高度显要的聚类关系。如图 3.10 中

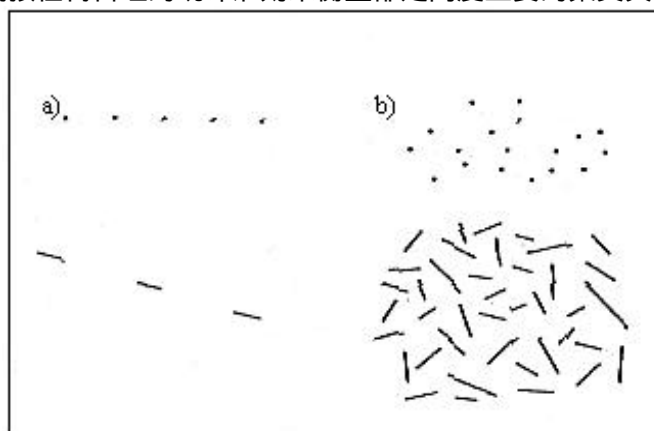


图 3.10 (a)所示为 5 个等间距共线点或 3 个等间距共线线段，如果把这些特征放在相似背景中（如图(b)所示），人类视觉就难以很快地发现这些关系

例子所示，图 3.10(a) 中严格的共线等间距的图象点或线段，不太可能是由于偶然性产生的。但人类视觉都不能在有相似特征的背景中把它们检测出来（见图 3.10(b)），不能检测这样非常显著结构的原因很可能是人类视觉的局限性，而不是一种功能特性的问题。人类视觉的这种局限性很可能是由于聚类过程中固有的计算复杂性引起的，由于要涉及检验图象中每一种可能的图象特征的子集，要在图象中发现所有可能的显著关系在计算上是行不通的。一种限制计算复杂性的方法是只检测由图象中相靠近的特征组成的关系，另一种方法是考虑在给定区域中的所有特征，并根据各种特性作直方图统计并寻找统计特性中的显著峰点，这就是大多数纹理分析方法的基础，对此这里不再详述。读者可参阅 [Low 85]。

3.2.2 通过求能量极小进行聚类

感知聚类涉及各种视觉任务，例如寻找曲线、纹理分割和完成形状等。所有这些视觉任务中的基本问题是发现数据中的组织和结构。众所周知，视觉处理的原始输入是图象数据。边缘检测的任务是发现图象中重要结构信息的位置，而物体识别则是应用景物的结构描述来推论景物的内容。感知聚类被认为是属于比边缘检测高，比物体识别低的视觉处理层次。它的任务是发现存在于现实世界景物可见事件之间的空间关系。这意味着感知聚类是一个逆向问题。因为景物中的空间关系在被投影到二维图象平面时没有直接明显地保留下来，因此，和大多数逆向问题一样，在其原始状态下聚类是一个不适定问题。如上一节所述，通过引入稳定函数可以限制解的空间，这些稳定函数表征了所期望的解。在这个问题上感知组织的 Gestalt 规则可用来描述人类视觉系统具有的聚类特征。因此可通过引入稳定函数来实现 Gestalt 规则，这样问题就变成发现使下式极小的 g

$$\begin{aligned} & \mathbf{I}_1 \|E_{prox}(g)\|^2 + \mathbf{I}_2 \|E_{sim}(g)\|^2 + \mathbf{I}_3 \|E_{clos}(g)\|^2 + \mathbf{I}_4 \|E_{cont}(g)\|^2 \\ & \mathbf{I}_5 \|E_{sym}(g)\|^2 + \mathbf{I}_6 \|E_{F.G.S}(g)\|^2 + \mathbf{I}_7 \|E_{C.F.}(g)\|^2 \end{aligned}$$

其中 $E_{prox}(g), E_{sim}(g), \dots, E_{C.F.}(g)$ 是一组稳定函数，它们分别实施由于接近、相似、封闭、连续、对称、物体一背景分离以及公共消失（Common-fate）等因素造成的聚类； $\mathbf{I}_1, \dots, \mathbf{I}_7$ 是正则参数，它们决定对每个聚类因素的加权系数。

这样聚类就成为一个求极小的问题。然而问题不在于寻找图象中单个最优解，而是要寻找聚类的最优组合，并且其中聚类的数量也不知道。

能量极小化问题在机器视觉系统中被广泛应用，它已被用于曲线检测、根据影调恢复形状、立体视觉匹配等方面。每一类极小化问题都涉及以下要点：

- (1) 解的空间：即所求解问题的解的总体集合。
- (2) 能量函数：对解空间中的每个解分配一个能量值的方法。
- (3) 最优解：在整个求解空间中能量函数为极小的解。

能量极小是发现最优解（或尽可能接近最优解）的过程。在感知聚类中，图象中所有可能聚类的集合是解空间，可以定义一个能量函数来评价聚类的各个方面。为了公式的完整性，最优解被定义为使图象中的能量的总和为极小的聚类集合（Collection）。

设： i —数据点的数量； j —图象中的集合数量； D —图象中数据点的集合。

$$D = \{d_1, \dots, d_i\}$$

设 g —是聚类数据点的非空集合； G —图象中所有聚类的集合

$$G = \{g_1, \dots, g_j\}$$

因此所有的点都在聚类之中

$$g_1 \vee g_2 \vee \dots \vee g_j = D$$

根据前式对每个聚类都定义一个能量函数

$$E_{Total}(g) = \mathbf{I}_1 \|E_{prox}(g)\|^2 + \mathbf{I}_2 \|E_{sim}(g)\|^2 + \mathbf{I}_3 \|E_{clos}(g)\|^2 + \mathbf{I}_4 \|E_{cont}(g)\|^2 \\ \mathbf{I}_5 \|E_{sym}(g)\|^2 + \mathbf{I}_6 \|E_{F.G.S}(g)\|^2 + \mathbf{I}_7 \|E_{C.F.}(g)\|^2$$

这样聚类的问题就变为寻找满足

$$\min \sum_{k=1}^j E_{Total}(g_k)$$

这个公式的主要优点是它把聚类系统的目标和如何达到目标的过程明确地相分离。聚类的目标由能量函数定义，这个函数表征所期望的感知聚类的特征。这些函数可被分别确定，如果需要的话可被改变。实现目标是搜索策略的问题。这个策略试图使聚类总体能量极小。也就是说试图寻找具有所希望特征的聚类，能量极小方法的另一个优点是它与任何表示方法相独立。

在通过求能量极小进行聚类的方法中还有两个重要的问题，即计算策略和表象的问题。这将在以下的章节中介绍，在此基础上再进一步讨论这种方法实际的应用情况。

3.2.2.1 计算策略

试研究图 3.12 所示的情况，显然在图中可发现多个组织层次，在聚类的最低层是黑色像素聚类成线段，与此同时白色像素聚类成背景，根据上一节讨论的原理，这意味着对这组数据来讲，在这个聚类状态下，聚类的能量要比任何其它可能的聚类状态要小。这节要讨论为给定数据寻找极小能量聚类状态时的计算策略问题，几乎每种计算极小策略都包括以下步骤：

- (1) 起始状态。这可以是预先选择，总要用到的状态或是这些数据每次都要终结的状态。
- (2) 选择下一个状态的方法。这可以是选择在状态空间中的相邻状态。或者，虽然通常是按与每种状态相关的能量来选择，但也可可是纯粹的随机选择。
- (3) 终结条件。可以是在所有可能的状态都被考虑的状况下结束或在其它状况下，例如，当当前状态的能量小于预定的阈值时结束。

以下再对上述步骤作稍详细的说明。预选的起始状态可以是以下几种：(1) 所有的数据点都被聚类成一个类；(2) 没有点被聚类在一起，也即类数等于点数；(3) 可先对数据进行简单的分割，这样就可以得到一些分类的结果，并以此作为起始状态。选择下一个状态的问题涉及寻找另一个比当前状态能量值低的状态的问题。这可以通过使类合并、分裂，和把数据点从一个类转移到另一类等操作来实现。

当假设聚类是一个寻找全局能量极小状态的问题时，一个简单的对能量取阈值的方法是不适合于作为终结条件的。因为在状态空间中能量的分布很可能是多模的，所以寻找局部的极小也不适合作为终结条件，由于状态空间可能是很巨大的，所以要评价每一种可能的聚类状态是不太现实的。因此与许多其它的优化问题相似，很难知道已找到的能量最低的状态是否是全局的极小，最多能做到的是继续搜索更低的能量状态，与此同时保持至今已发现的最优状态。因为发现全局极小的概率与已评价的不同状态数量直接相关。

这并不是说问题是不可解决的或者对问题的描述方法不对，有时人类视觉也不能发现最优的全局聚类状态。例如，对图 3.11 所示图象，很难感觉到一个稳定的最佳聚类。

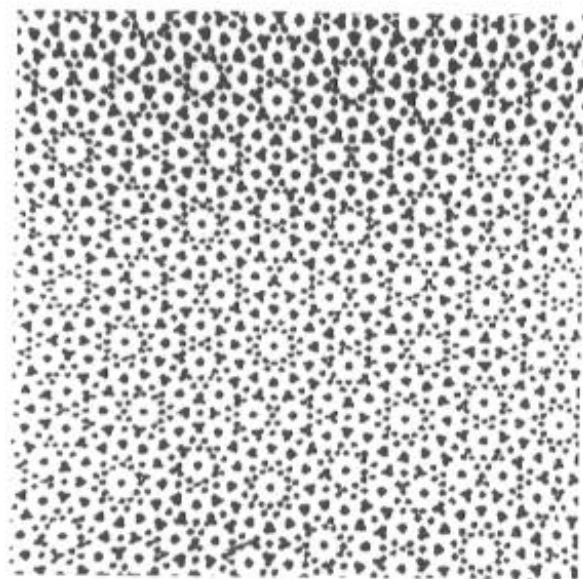


图 3.11 人类视觉的主动本质的证明（取自 Marroquin 1976）
圆圈和矩形从点的图案中不断出现又消失

3.2.2.2 感知聚类中的表象（representation）

对于在视觉信息中发现结构来说，表象很重要，特别是感知聚类涉及把图符（icon）类型的数据转换成符号类型的数据。由于感知聚类具有递归的本质，因此必须仔细地考虑浮现特性的表象问题。因为它既是聚类处理的输入也可以是输出，一个明显的问题是数据应该是以图符还是以符号形式存储。McCafferty 提出了称为“参数空间”（Parameter Space）的新方法。它可表示浮现特性，因为它兼有图符和符号表示的优点。试考虑图 3.12 所示的图象，对此图你所作的感知聚类的第一阶段是把黑色的像素聚成黑色的线划。方向是聚类

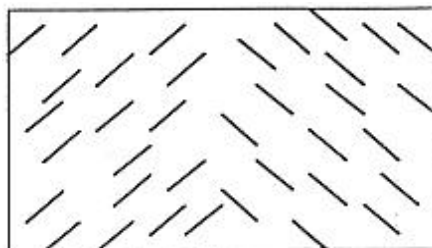


图 3.12 两个层次的聚类，首先黑色的点聚成线段，接着聚成两个区产生的浮现特性。为以图符方式表示这些新信息，在每个象素点上必须要有一个有效的值来表示浮现特性，但对白色象素点地方来说就不是这样。因为在这些地方并无“浮现特性”。在这些地方需要用符号标记，表示“这些点并无数据”。这说明表象不能是纯粹图符类型的。此外图符类型还有一个困难是，有时浮现特性会过密，这时难以用二维的图来表示。例如，在两条线的交点上同一点有两个方向值的数据，这在纯图符的方法中就难以表示。另一方面纯符号类型的数据又不能保持点之间的拓扑特性。而特征空间则可把这两种方法的优点结合起来。它是一个由二维 (x, y) 地图和附加的参数轴 z 组成的三维空间（对多元的聚类参数，如运动等则可以用多参数轴，这样就产生一个多维空间），参数轴可以是灰度、方向、深度等。空间中的 (X_0, Y_0, Z_0) 单元中的一个“位置标记”（place token）表示在二维地图的 (X_0, Y_0) 位置上有一个其值为 Z_0 的数据。因此如果在同一位置上有一个以上的数据，可简单地通过三维空间的任何 (X_0, Y_0) 栏中有一个以上的标记来表示。在某一特定栏中没有标记说明在这个位置没有定义参数值。应用参数空间作为聚类过程的输入和输出表象时数据

流就如图 3.13 所示。图中表示原始的图符数据先被转换到参数空间，然后对这些表象进行聚类处理。当可以得到浮现特性时，就以特征空间来表示这些浮现特性，以便进一步的聚类。这样的聚类不断的进行以得到更高层次的符号描述。

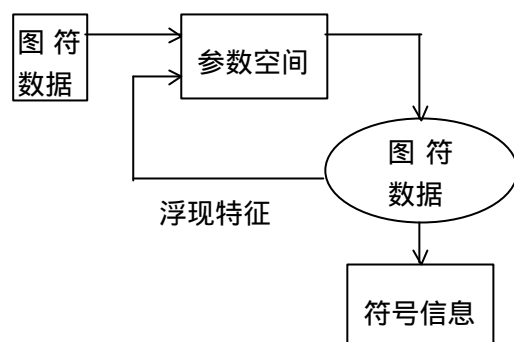


图 3.13 数据流程图，其中矩形框表示数据，椭圆表示处理

3.2.2.3 聚类能量的测量

如前所述对聚类的能量作测量是通过能量极小化进行聚类的基础，本节中将讨论根据 Gestalt 的感知组织规则对各项聚类的能量单独进行测量的方法。

1. 连续性 (Continuity)

聚类中的连续性因素是指人类视觉系统倾向于空间的连续性，而不是突然的改变。与此相关的物理现象是，现实世界中的表面边界和曲线通常是缓慢变化的。要确定聚类的连续性引起了以下若干问题。第一个问题是究竟对什么进行测量。聚类的连续性可以是指局部的方向，也可以指全局的方向。例如，对图 3.14(a)中的曲线来说，连续性测量可以是基于图 3.14(b)所示的局部曲率或是图 3.14(c)所示的全局曲率。

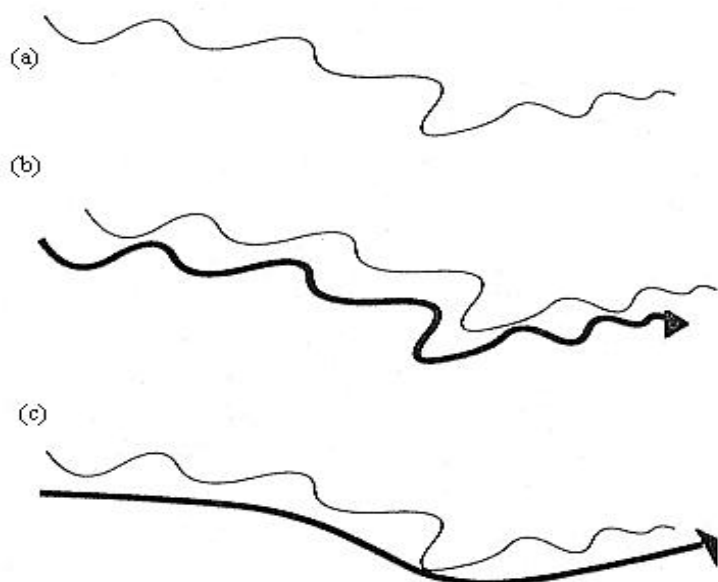


图 3.14 (a)原始曲线 (b)粗线是局部连续性测量 (c)粗线是全局总的连续性测量

另一个重要的问题在什么对象上作连续性测量。简单曲线连续性测量是相对简单的概念，而对区域或粗的曲线来说，连续性就不那么简单了，区域和曲线有一点重要的不同，这就是曲线上的点有一个自然的排列次序。除此以外，在人类视觉系统中似乎可以认为区域和

曲线具有相同的性质，而在机器视觉系统中有时对区域和曲线的概念也难以区分。例如，在用一阶微分算子检测边缘时，由于模糊或数字化的影响，边缘检测经常得到宽度超过一个像素的曲线。所以 McCafferty 提出把区域和曲线的连续性用相同的方法来处理，也即都是在聚类的定界周边（bounding envelope）上作连续性测量。这是因为在定界周边上也可以有自然的排列次序，可用于连续性测量。在曲线的情况下定界周边与曲线本身非常接近。如图 3.15 所示。而对区域状稀疏分布的数据点来说，也可以用下述算法得到定界周边并测量连续性。

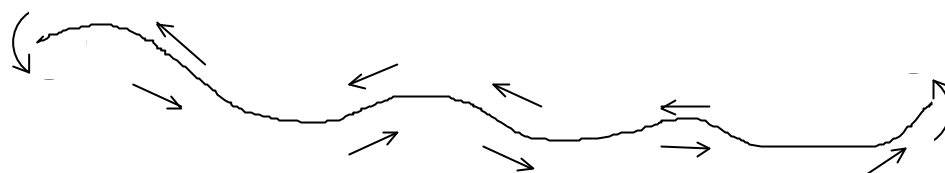


图 3.15 曲线的定界周边就在曲线的两边

定界周边检测算法：

把每个聚类看作是一张以标记（token）作为结点的图（graph）。

- (1) 发现最南面的结点，这是周边上的第一个标记。
- (2) 从水平逆时针扫描，直到弧与相邻的标记相遇。这是周边中的下一个标记。移到这个新的结点。
- (3) 从先前的方向逆时针扫描直到弧遇到新的标记，这是周边上的新标记，移到新结点。
- (4) 重复(3)，直到再次发现第一个结点，这就如图 3.16 所示。

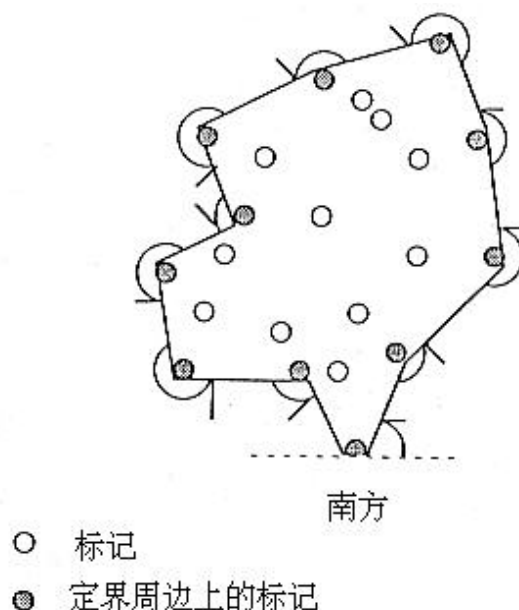


图 3.16 通过径向扫描邻域寻找定界周边

这个算法能很好地应用于曲线和区域的聚类。这里还剩下一个问题 就是对“相邻”的定义。McCarfferty 提出可用以下两种方法来定义：

- (1) 把相邻点定义为 N 最近邻点。 N 可由用户选定。
- (2) 定义邻域的大小，所有在此范围内的标记被定义为相邻。

有关的详细叙述可参见（McC90）中的第五章。

得到了定界周边就可以测量连续性，对任何曲线 $V(s) = [x(s), y(s)]$ ，其中 s 从 0（曲线的起点），到 1（曲线的终点），它的曲率是

$$E_{cont} = \int_0^1 |V_s(s)|^2 ds = \int_0^1 [|X_s(s)|^2 + |Y_s(s)|^2] ds$$

其中 $X_s(s)$ 和 $Y_s(s)$ 分别是 X 和 Y 相对于 s 的导数。这被称为膜能量，因为它描述了伸展的膜表面的能量。

在连续性的基础上进行聚类时隐含着两项能量，不仅有在使曲率极小的基础上进行的聚类，还有在使连续性变化极小的基础上进行的聚类。例如，图 3.17 上所示的曲线通常被认为是由两个不同的部分组成。这是依据每个部分不同的连续性：每个部分内的连续性变化比整个的小。

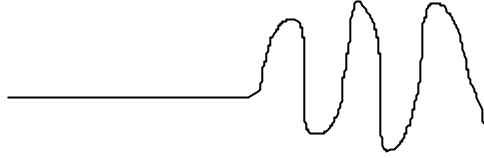


图 3.17 单报曲线被认为是由两部分组成，因此连续性变化很重要

因为连续性测量是沿定界周边进行的，所以连续性变化是用连续性的一阶差分来测量的，即 $E_{Cont\ var}$ 等于

$$E_{Cont\ var} = \frac{1}{K} \sum_{p=1}^K |Cont_{p+1} - Cont_p|$$

其中 K =定界周边上的标记数， $Cont_p$ 是周边上标记 P 的连续性测量， $Cont_{k+1} = Cont_1$ ，这样在每个标记上还要增加一项连续性变化能量。因此，在曲线发生极值的地方聚类的能量就大。

2. 接近性 (Proximity)

接近性是聚类中的位置标记在二维空间上分离程度的度量。这里的位置标记可以是数据点或是前一级聚类的输出。在接近性测量中必须考虑两个不同的方面：接近性自身和接近性差异。例如，在图 3.18 所示的图象中，人类视觉将把像素按竖的列而不是横的行聚类。这是根据接近性（聚类元素之间的最小距离）聚类的结果。

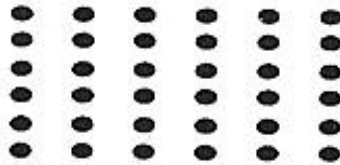


图 3.18 按接近性聚类成列

然而对图 3.19 所示的情况就要聚类成两个区域，而不是一个，其中每个区域都有不同的、但在该区域内保持不变的平均接近性值。因此，这是按每个类内部按接近性差异极小的原则来聚类成两类。所以在以接近性进行完全的聚类时要同时考虑接近性和接近性差异，在聚类能量计算时也要同时考虑这两项，接近性可计算为：

$$\frac{1}{i} \sum_{p=1}^i \frac{1}{n_i} \sum_{q=1}^{n_i} d_{pq}$$

其中 i =标记数， n_i =标记 i 的邻域数， d_{pq} =从标记 p 到标记 q 的距离。在连续性量测中已对邻域作了定义，这里可以沿用。

接近性差异可以根据接近性数据来计算，这里采用一阶差分的方法。即把接近性差异定义为：

$$\frac{1}{i} \sum_{p=1}^i \frac{1}{n_p} \sum_{q=1}^{n_p} |pr_p - pr_q|$$

其中 pr_p, pr_q = 标记 p 和 q 的接近性值, i = 标记数, n_p = 标记 p 的邻域中的点数; q = 标记 p 的邻域中的一个标记。

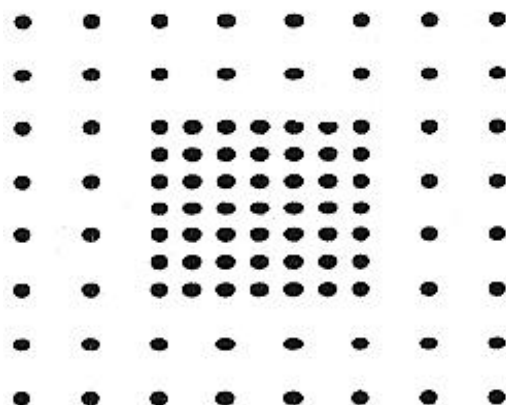


图 3.19 按接近性差异聚类成两个区域

3. 相似性 (Similarity)

人类视觉系统可在参数均匀性的基础上在许多不同的景物表象中形成聚类。这些表象包括灰度、深度、运动和浮现特征等。这样做的根据是在相同表面上的标记很可能共享某种特征。已提出了许多测量标记参数相似性的方法。一种很容易想到的方法是测量标记参数的统计方差。但是与所有基于数量的方法一样，这样就丢失了聚类中的拓扑信息。例如，图 3.20 中所示的两个图象区域看起来是相当不同的。各灰度的象素数相同，所有灰度级之间的统计方差也相同，这意味着，如果用统计方差作为相似性量测，这两个图象的相似性能量是相同，这说明统计方差不适合于作为相似性测量。所以 McCarrfferty 提出用相邻标记之间的一阶差分之和，即：

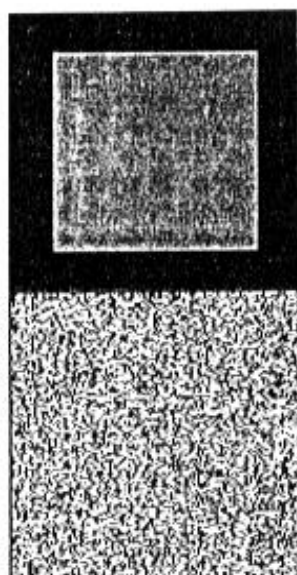


图 3.20 上、下框图象的各灰度级象素相同，而只有空间分布不同，因此，它们的系统方差相同

$$\frac{1}{i} \sum_{p=1}^i \frac{1}{n_p} \sum_{q=1}^{n_p} |param_p - param_q|$$

其中 $param_p$ = 标记 p 的参数值，在计算中也与连续性和相似性测量时相同，采用了当地的邻域的定义。

4. 封闭性 (Closure)

Gestalt 聚类因素中的封闭性是指人类视觉系统倾向于封闭的，而不是开放的形状，但封闭和开放的定义不是精确的。例如，图 3.21(a)通常被认为是封闭的。图 3.21(b)被认为是开放的，然而图 3.21(c)就既不是封闭的，也不是开放的，它从面积来看是封闭，但又看起

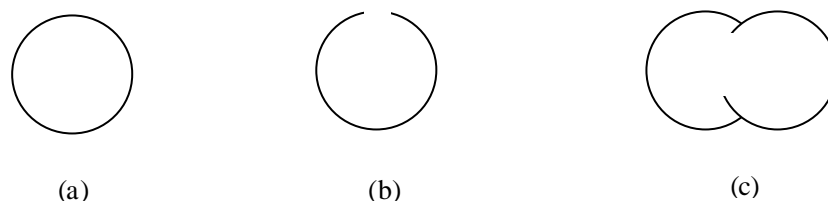


图 3.21 (a) 是封闭的 (b) 是开放的 (c) 是既不封闭、也不开放的形状来不太完全。通常封闭性是涉及把丢失的信息填补上以便使聚类更为明确 (stronger)，在许多情况下噪声是引起信息丢失的原因，此外还有遮挡的情况，这两种情况下都可以利用高层知识来填补丢失的信息。

聚类的连接性程度与主观上感觉的封闭性能够很好地对应。例如，图 3.22 中的标记 (P, Q) 只有很小的视网膜距离 (retinal distance)，但是按图形路径相距很远。封闭性比例定义为：

$$\frac{|PQ|}{\min_arc_distance(P, Q)}$$

当这个比例趋向于 1 时，(P, Q) 这对标记就相联结，形状就封闭。当这个比例趋向于 0 时，这对标记就表示聚类中封闭性不好的情况。

封闭性比例可在每个单独的标记与聚类中所有其它的标记之间计算求得。这样对每个标记都可赋予一个取决于比例的封闭性能量。对每个标记选择这些比例中的平均值或最小值以减小结果中的信息内容。选择最小的比例来标识聚类中最令人感兴趣的标记。因此，对每个标记 P 寻找能满足 $\min \frac{|PQ|}{\min_arc_distance(P, Q)}$ 的标记 Q。

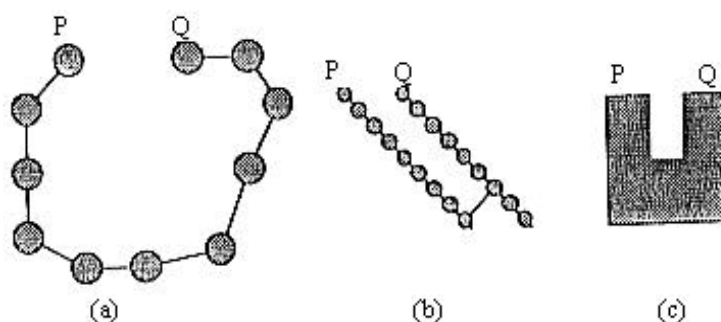


图 3.22 标记 P 和 Q 使封闭性比例最小

5. 对称性 (Symmetry)

对称性存在于许多方面，因此检测和测量对称性是复杂的问题。可能最明显的对称类型是全局对称 (global Symmetry)。全局对称通常涉及一根单独的直线轴或旋转中心。图 2.23 所示是一个全局对称的例子。在检测这样的对称性时必须对聚类中的所有标记找到一组单独的轴线。

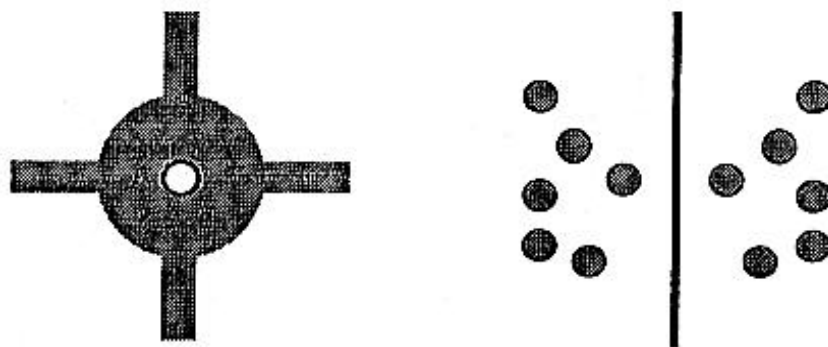


图 3.23 全局旋转对称和全局镜象对称的例子

局部对称是对称性中的另一个方面，它是指存在于形状边界内的不同部分之间局部范围的对称。图 3.24 所示为局部对称的例子，计算局部对称时只用到了局部的边界曲率信息，并对每个标记要定义一个局部的邻域。



图 3.24 局部对称的例子

对称性的另一个方面是可用于推论三维形状的斜对称性 (Skewed Symmetry)，斜对称是实际的对称性在投影到图象平面时产生的。图 3.25 所示为斜对称性的例子。对平面的表面来说，可用检测到的斜对称轴来估计平面在三维空间中的方位。

要全面地研究对称性是非常复杂的问题，这里仅限于把对称性作为感知聚类中的一种因素来研究计算方法：如果某一聚类表现高度的对称性，那么它的能量就减小。每个标记对全局对称性能量的贡献可通过测量把这个标记投影到轴的另一边时与相应标记的靠近程度来估算。

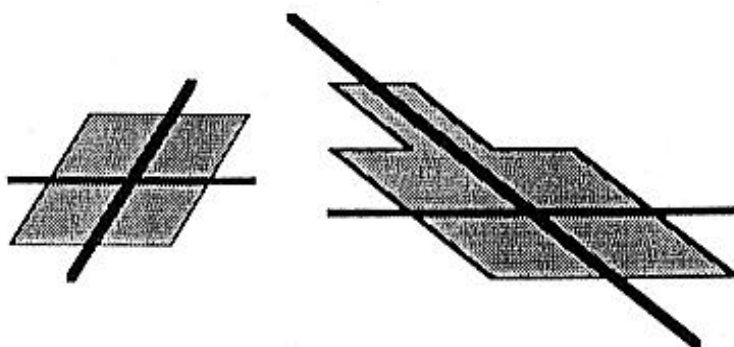


图 3.25 斜对称被认为是三维空间中的实际对称性在投影到图象平面时产生的

对称性的计算可以与前面研究的其它聚类策略相结合。例如，应用定界周边的概念可用于检测轴的两边的标记之间不存在一对一对应的对称，这就是图 3.26 所示的情况。再例如，按照递归聚类的策略，类间对称性可以成为类内对称性。

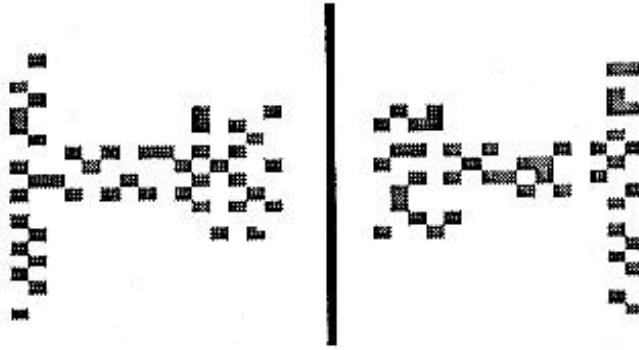


图 3.26 在稀疏数据点和没有细节对称时的对称性

6. 小结

根据上述各项聚类因素的能量测量可以得到以下的聚类能量方程

$$E_{total} = I_1 E_{prox} + I_2 E_{prox var} + I_3 E_{sin} + I_4 E_{Clos} + I_5 E_{Cont} + I_6 E_{Contrar}$$

对此公式有几点要加以说明：

- (1) 为了使聚类中同一类内的接近性和连续性的变化最小，在接近性和连续性的能量计算中增加了两个附加项 $E_{prox var}$ 和 $E_{cont var}$ 。
- (2) 为了使聚类的封闭性能量极大（而不是极小）， I_4 取负值，而其它加权系数都是正的。
- (3) 由于对称性非常复杂，所以在此公式中没有包括对称性。

在确定各种 Gestalt 聚类能量的测量方法以后，还要确定各项加权系数，来规定每种聚类因素的相对重要性，并决定总的聚类能量。对此这里不再细述，可参阅[McC90]。

3.2.3 根据图象特征推论三维空间结构

在 3.2.1 节中我们讨论了视点恒常性约束。根据视点恒常性约束不但使我们能确定哪些图象关系是可以稳定可靠检测的，并且使我们能根据二维的图象聚类推论相应的三维空间结构，在这一节中我们将详细研究各项可能的推论。在研究这些推论时，先要满足以下两个条件：(1) 假设当视点在三维空间的相当宽的范围变化时此图象关系保持恒常。(2) 观察者或摄影机的视点和光源方位与景物中的物体相独立。

1. 共直线性 (Collinearity)


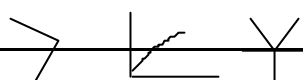
当在图象中任何三个或更多的可区分的点集合是共线时，我们可以推论，这些点在三维空间中也是共线的。这个推论可以推广到直线的情况，即图象中的直线在三维空间中也是直线。当曲线是共面的，并且碰巧摄影机位于包含曲线的平面内时，此曲线也可能投影为直线，但这只是偶然发生的情况。此外在图象中是共线的两直线，在空间也是共线的。应用这个推论就可以把由于遮挡造成的间隔连接上。

2. 共曲线性 (Curvilinear)

把共线性推论从直线推广到曲率为常数的弧线，特别是当两条曲线或 4 个甚至更多的点位于一段圆弧上时，我们可以认为它们空间也是位于同一条弧线上。但是，因为在空间是曲率为常数的弧线不一定投影为图象中一定范围内曲率为常数的弧线（例如，圆周投影为椭圆），这个推论通常只适用于有限的局部范围。

表 3.1 根据图象特征推论三维空间关系

序号	二维关系	三维关系	例子
----	------	------	----

1.	点或线段共直线	在三维空间共线	
2.	点或弧共曲线	在三维空间共曲线	
3.	两个或更多的端点在一个公共点	三维空间中多条曲线共一个端点	
4.	终结在连续曲线上	终结曲线不会比连续曲线更靠近摄象机	
5.	连续曲线交叉	两条曲线不可能同时都是遮挡边界	
6.	平行曲线	曲线在三维空间平行	
7.	三条或更多的线条会聚到一个公共点	线是平行的，或会聚到三维空间中的一个公共点	
8.	等间距的共线点或平行线	在三维空间中等间距和平行线是共面的	
9.	斜对称	在三维空间中是对称的	

3. 共端点

当图象中的两条或更多的曲线共同终结在一个端点时，可以推论这些曲线在空间也是共端点的。

4. 端点在连续曲线上

当图象中的曲线终结在一条连续的曲线上时（形成一个 T 形连接），这条终结的曲线不可能比连续的曲线距离摄象机更近一些。不然的话，这是由于摄象机的偶然位置，使得曲线的终结正好出现在另一条曲线上。如果我们具有关于这两条曲线中任何一条的类型的知识，就能通过 T 型连接推论另一条的类型。T 型连接是以下三种情况产生的：任何类型几何边界曲线的遮挡；表面上的标志，阴影或线条终结在几何边界上；或表面标志的组合。因此，如果我们已知终结的曲线是一条几何边界，那么可以推论出连续曲线也是几何边界。如果我们已知连续曲线是在终结曲线这面上的几何边界，那么可以推论终结曲线必定是表面上的标志或阴影。如果已知终结曲线是阴影，就可推论连续曲线是几何边界。

5. 连续曲线相交

当两条连续曲线相交叉（形成 X 型连接）时，这两条曲线不可能同时都是几何边界。如果已知其中一条曲线是遮挡几何边界，那么另一条曲线必然是更靠近观察者，并且或者是线或者是部分透明物体的边缘。如果已知一条曲线是阴影，那么另一条曲线就一定是同一表面上的表面标志，另一条阴影（由不同光源照投的）或是比阴影更靠近观察者的线。

6. 平行线

在图象中平行的曲线可被认为在空间中也是平行的，如果两条非平行的曲线，为要在图象中平行，那么摄像机就要被严格地放在某个特定的方位。并且当曲线越复杂时，要做到这点越困难。当然，由于透视投影产生的畸变，三维空间中的平行线在图象中的不一定是平行的。但是在许多情况下两条平行线只相距较小的视角，因此透视效应不明显。

7. 会聚在同一点线条

当三条或更多的线条会聚在一个共点时，可以推论这些线条或是在三维空间中会聚到公共点或是在三维空间中的平行线（由于透视效应，平行线会聚到消失点）。在景物中有相当多的平行线（例如在与重力一致的方向上），这就为进行这种推论提供有力的基础，作为一条推理，可以认为一旦确定了一个消失点，就可推出图象中任何指向此消失点的线在空间中有一个特定的方位。

8. 等间距

如果不考虑透视效应，那么在三维空间中一系列的等间距的共线点或线段在图象上投影产生的也是等间距的共线点或线段。

9. 斜对称性

所谓斜对称性是指沿与对称轴成某一固定角度的对称横轴寻找对称。如果在图象中发现符合对称性的形状，例如平行四边形，那么可以推论这是由在三维空间中对称的形状，例如矩形，投影产生的。并且如果斜对称形状是在同一平面上，还可以根据对称轴和对轴横轴的方向，对此平面的法线在空间的方向作出约束。

3.3 景物结构的模型

在前面的章节中我们讨论了感知组织在视觉信息处理和理解中的重要性，以及感知组织现象和基本原理。可以看到感知组织所涉及的主要是视觉信息的低层处理。在此基础上，现在有必要对视觉信息处理和理解的总框架作进一步的研究，因此，在这一节中我们将介绍 Pentland^[Pen 86]的部件模型和自然形状表示法，以及 Biederman^[Bie 85]的部件识别理论。

3.3.1 部件模型和自然形状表示法

如前所述视觉理解中的主要问题是单靠传感器数据本身不能完全地确定景物的结构。为了能根据所看到的图象对景物进行判断，我们需要有关于图象形成和景物是如何构造的知识。与此同时，有许多证据也说明周围世界的构成是有规律的：在生物界生物进化的特点是不断地重复^[Tho 42]；在非生物界受物理定律的约束，物体的形状局限于有限的几种基本模式^[Ste 74]；周围世界外表上的复杂性实际上是通过对这些基本的物体形式作各种不同的组合而产生的。

因此视觉的根本特征是识别传感数据中的规则性。所谓规则性就是景物中合理和可靠地相互联系的相对结构，也就是要依据景物的模型。模型在这里的作用是说明景物是如何构造的和景物的结构如何体现在图象的规则性之中。视觉理论是以模型为基础的，视觉对模型的需要是不能绕过的。因为正是模型把理论上的表达和计算与实际的景物状态相联系，并且解释理论的含义。一个没有模型的视觉功能理论是没有意义的。

在早期视觉中需要应用关于景物的模型来解释传感数据的观点已被许多人接受，目前所用的大多数模型可以归纳为高层的专门模型和关于图象形成的低层模型。

(1) 图象形成的模型

在计算机视觉领域已对图象形成的点状模型进行了广泛的研究。这个模型主要是从光学、材料科学和物理学中借用过来的，对它的研究是在 Marr^[Marr 82]提出的关于视觉的计算

理论的框架下进行的。按 Marr 的视觉理论，视觉信息的处理要经过一系列不同层次的视觉表象。最初的表象是直接从局部的图象特征计算得到的，而高层表象是由对小区域内的下层表象计算得出的。处理过程主要是数据驱动的。

Marr 的视觉理论虽然已获得显著的成就，但也遇到了严重困难。其中主要的问题是，为了从对图象的局部分析中得到有用的信息，要对环境的结构作诸如各向同性和平滑性这样的严格假设。而在实际景物中这样的假设经常是错误的。因为在自然景物中，成象参数从一点到另一点是以相当任意的方式改变。因此，这样的方法很难适用于自然景物的环境。

此外，这个理论更根本的困难是，即使可以得到深度图和其它表面本征特性的图象，并且这种图象形式的表象可以用于解决一定的任务，但还必须经过分割、解释等关键处理才能被理解，因此并没有能解决图象理解中的困难问题。

(2) 专用的模型

另一类模型是工程风格的表示方法，例如具体物体的 CAD-CAM（计算机辅助设计和制造）模型。这样的模型本身是很详细和复杂的，所以在用图象形式表示时也是详细和复杂的，更重要的是模型中表示的是物体表面的形状，而不是物体的外表。因此当物体的方位改变时，就要根据模型来生成各个视图，但这时并未把光照和成象的条件考虑进来。把这样的模型用于图象理解是很不方便的。所以，在实际使用中经常采用它的简化方式，即线框模型，它只利用物体边缘的信息。在环境比较简单的情况下，例如，工业应用中，只用物体的边缘信息已能解决问题，所以这种模型取得了一定的成功。尽管如此，由于这种建模方法缺乏灵活性，当要识别的物体种类较多时就遇到困难。只使用这种高层专门模型的视觉系统还有一个更为本质的缺陷是缺乏学习新的物体类型的能力。

1. 部件模型

由于上述两种模型都不能满足视觉理解的要求，所以某些研究者^[Pen 86]开始寻找第三种模型。这种模型粒度介于点状的成象模型与高度的专门模型之间。他们认为有许多理由可以相信能够用这样的中间粒度的模型来准确地描述客观景物。景物可以描述为是由相对较小的一组通用过程一再发生而生成的。这是因为客观景物外表上的复杂性是通过把有限数量的基本形式以不同的方式组合起来而造成的。

人们很早就知道每当可能时，进化总是重复它的答案。这就使得在所有生物中都存在巨大的规则性：只有少数几种四肢、少数几种皮肤、少数几种叶子和少数分枝模式。例如，多得令人吃惊数量的树的模型是由伴有生成树皮和树叶的三维纹理的简单分枝过程组成的^[Smi 84]。相同的分枝模型还可用于河流、静脉和珊瑚。相似地，近来又发现受物理定律的约束，非生物形式也仅限于有限数量的基本模式^[Man 82]。Mandebrot 已证明象云彩、山坡、海岸线这些看起来复杂的形式可通过简单模式的不同尺度递归地重复来描述。Stevens 提出了有力的证据说明自然的纹理只以少数几种基本形式出现。

正是周围环境的这种内部结构使我们能得到合理的关系，也正是这种内部结构使物体的特性能聚集成类，并使我们能在常识推理中利用简化的分类描述。由此可看到也许有可能用部件（parts）：一种宏观模型，来精确地建立世界的模型。可通过相对简单的组合把部件用于建立周围物体以及它们行为的粗略但能用的模型。如果接受这种观点，那么视觉研究的中心问题就是寻找一组普通可用的部件模型，发现与单独部件合理地相关的图象规则性，然后应用这些规则性来把图象的内容识别为这些通用基元的组合。由于这样的模型比专门模型简单，所以我们可以比较容易说明它在图象中的特征；另一方面，与点状的成象模型相比，部件模型是描述宏观的结构，所以可避免由于约束不充分而被迫采用平滑性或各向同性这样强烈假设的问题。部件模型除了在复杂性和可靠性之间提供一种平衡以外，它之所以引起人们的广泛兴趣是因为描述了整个物体之间的定量关系，而不是描述局部的表面块或专门的物体，因此有可能提供一种以我们经常直接可用的粒度描述世界的“词汇”。

在形成这种部件模型时遇到的问题是：部件必须足够地复杂，以便能被可靠地识别，但同时又必须足够地简单，以便能方便地用作构成专门物体模型的构件。当前的三维视觉系统，通常用矩形体或圆柱体作为特定形状 of 模型，但要用这样的基元来自动地构成任意形状的新物体是很困难的，为了支持真正的通用视觉系统，需要开发能适用于描述任意形状的新的建模基元。为此 Pentland 提出了基于超二次型（Superquadrics）的部件模型。

2. 自然形状表示

Pentland 提出的表示系统不但可以准确描述人造的形状，而且可以描述各种自然的形状，例如，人、山、云和树等（见图 3.27）。在这个表示方法后面的想法是要提供一种模型和操作的“词汇”，使我们能把周围世界建模为由相当简单的成份——部件组成，并且这些部件是可以从图象中可靠地识别的。

这种表示方法最原始的概念是可把它想象成“一块粘土”，一种建模用的基元，它可变形和成形，但它应能大致与我们的“部件”概念相应。为此 Pentland 采用被称为超二次型的（Superquadrics）参数化的形状作为这种基本的建模基元，超二次型可表示为：

$$X(\mathbf{h}, \mathbf{w}) = \begin{pmatrix} C_h^{e_1} & C_w^{e_2} \\ C_h^{e_1} & C_w^{e_2} \\ S_w^{e_1} & \end{pmatrix}$$

其中 $\cos \mathbf{h} = C_h$; $\sin \mathbf{w} = S_w$ 。X(η, ω) 是一个扫描出表面的三维矢量，表面的参数是纬度角 \mathbf{h} 和经度角 ω ，表面的形状是由参数 e_1 和 e_2 来控制。这一族函数中包括立方体、圆柱、球面、钻石形和金字塔形以及介于这些标准形状之间的弧形边缘的形状。图 3.27 所示为这些形状的一些例子。

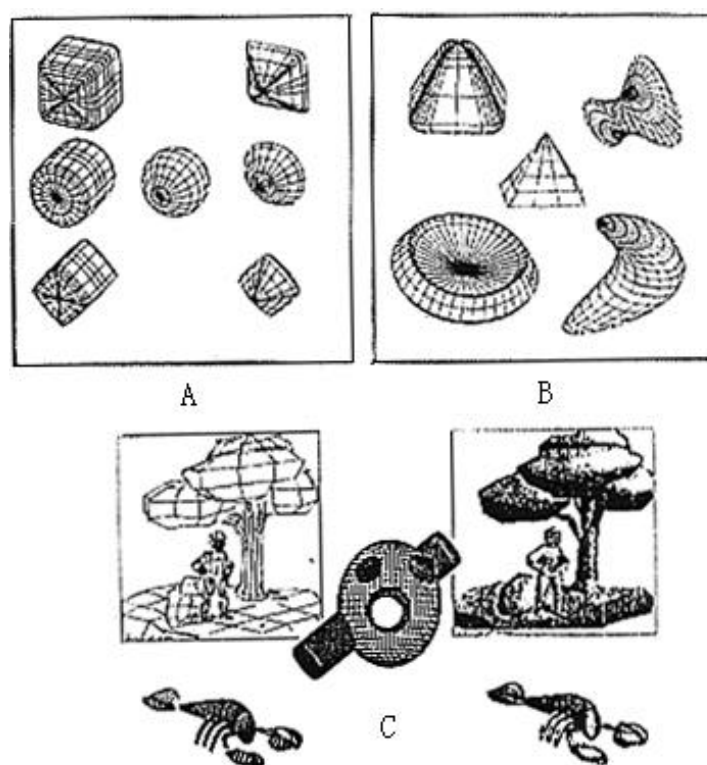


图 3.27 (A) 允许的基本形状的样本 (B) (A) 中形状的变形 (C) 基本形状的代数组合

这些基本的“粘土块”可作为原型，这些原型可通过拉伸、弯曲、扭转或变尖来变形，然后通过布尔代数运算来组合形成新的更复杂的原型，并可以递归的形式再次进行变形、作

布尔代数运算和组合。例如，椅子的靠背是一个弧形边缘的立方体，它被沿某一轴压扁，然后稍加弯曲以适合人的背部形状，椅子的座的形状与此相似，但旋转了 90°，把这两部分拼在一起，再加上长矩形的四条腿，就得到椅子的完整描述（见图 3.28）。

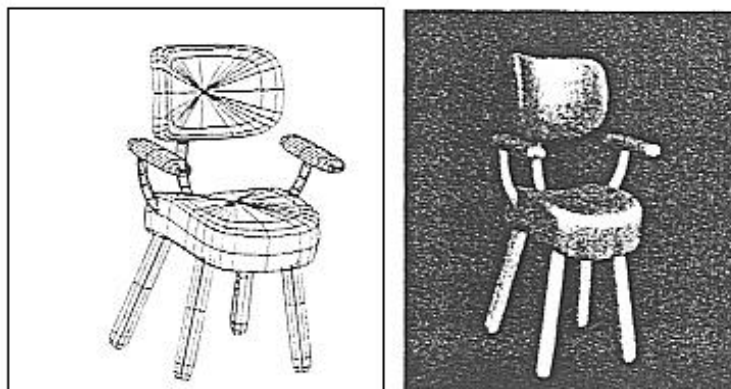


图 3.28 把合适的变形超二次形用布尔代数组组合在一起

3.3.2 部件识别理论（Recognition-By-Component, RBC）

1. RBC 理论出现的背景

部件识别理论是由 Biedeman^[Bie 85]在对人类视觉和听觉的识别能力进行分析研究的基础上提出来的。他认为人类的视觉识别能力有以下几个特点：

(1) 识别迅速；(2) 能从未经过的观察方向识别；(3) 能承受相当大的视觉噪声；(4) 在有遮挡时，仍能可靠地识别；(5) 物体是某类物体的新样本时仍能识别。

根据上述特点可以推论一个正确的物体识别理论应有以下特点：

(1) 不应根据对物体作定量测量的绝对判断，因为这样的判断不但缓慢而且容易出错。例如，当需要根据物体的曲率或长度来区分物体时，通常要比识别物体本身花更多的时间，因此在识别物体时定量的测量不应是控制因素。

(2) 应依据不受视点位置和噪声影响的信息。

(3) 应能进行部分的匹配计算。这样才能在物体被部分遮挡、缺损，或是某种类别中的新样本时仍能被可靠识别。

此外，以下两个实例也说明了部件在识别中的作用。

(1) 考察一下人们在识别一个不熟悉物体时的情形。例如图 3.29 中所示的物体。虽然人

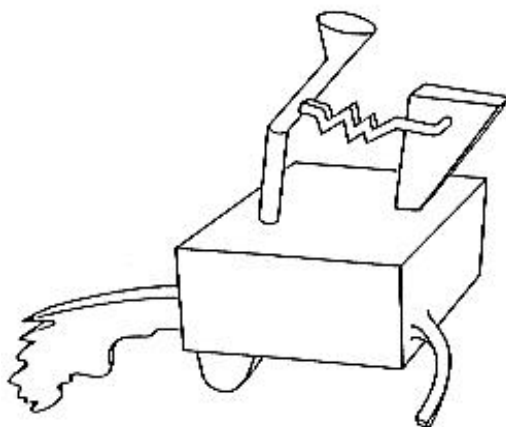


图 3.29 人们在把图中的物体分割成部件并在描述这些部件时意见相当一致
们对该图所示的物体不太熟悉，但在描述这个物体时可能意见相当一致。我们通常是在形状及凹凸变化剧烈的地方，把物体分割成部件，然后以熟悉的术语（例如，一个方块，一个圆

柱或截短的圆锥等)描述这些部件。以上虽然是对不熟悉物体情况的分析,但是对物体作分割,分析它的部件的方式并不随人们对给定的待识别物体的熟悉程度而变化。

(2) 语音中的音素数量有限,但通过不同的搭配和组合就可以产生语言中所有的词。例如,在英语中只有 38 个音素就能产生所有的词,在夏威夷语言中只需要 15 个音素。事实上只要 55 个音素就可以产生地球上所有的讲话语音。数量有限的音素的表达能力来自于它们可以相对自由地进行组合。与此类比,在视觉领域中基元不是音素,而是适当数量的简单体元。例如,圆柱、方块、三角块、圆锥等。在语音中音素之间的关系只有从左到右的顺序关系,而在视觉领域中基元之间可能的关系就要丰富得多,这样就有更为强大的表示物体形状的能力。

2. RBC 理论

RBC 理论认为在视网膜上接收到物体的图象以后,图象在凹凸变化剧烈的区域,也就是在曲率不连续的拐弯点被分割成区域。这样的分割与人类视觉的直观感觉很好地符合。然后分割得到的部件可用简单的体积部件来近似。这些体积部件可以用广义锥来建模。如前所述广义锥是通过横截面沿轴线扫描而产生,通常假设横截面与轴相垂直。第二个分割准则(确定部件的扫描轴的准则)用于提供体积部件的描述,它要求使部件的对称性和长度极大,以及使部件的横截面大小和曲率不变。

基本的部件被假设为简单的,通常是对称的和没有急剧的凹凸变化的体元,例如,方块、圆柱、球面和三角块等。RBC 的基本感知假设是部件可根据二维图象中的感知特性来区分,这些感知特性是容易检测和可相对独立于观察位置和可抗噪声和畸变的。这些感知特性中包括若干通常认为是感知组织的原理。例如好的连续性,对称性和完形(Pragnanz),RBC 在感知组织和模式识别之间建立了联系:虽然物体可能是高度复杂和不规则的,但用以识别物体的基元可以是简单和规则的,对规则性的约束不是用来表征整个物体的,而是物体的部件。

关于 RBC 理论还有以下要点:

(1) 部件之间的关系

为了表示一个物体,除了要确定部件类型以外,还要确定部件之间的相对位置。相同的部件通过不同的布置可产生不同的物体。如图 3.30 所示,把一个弧形在边上与圆柱相连就能产生一个杯子的形状,而把弧形布置在圆柱的顶部就可产生桶的形状;把一个部件布置在方块的长表面或短表面可影响到产生一个小型公文包还是一个保险箱。因此,物体的表示就是部件之间的结构描述。

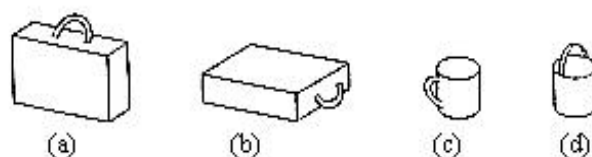


图 3.30 相同的部件经过不同的排列可产生不同的物体

(2) 处理的步骤

图 3.31 所示为物体识别的过程,通过边缘检测得到物体的线划图描述,从线划图可以检测图象的非偶然性特征。与此同时,通过检测非偶然性特征在曲率不连续区域进行物体分割,在分割区域中检测出的非偶然性特征为部件的身份提供了关键的约束。利用这些约束条件可以确定部件的种类,接着可把部件的布置与所存储的物体表示相匹配。可认为匹配可以并行地进行,并可作部分的匹配,匹配的程度取决于部件与图象和物体表示之间的相似性。上节对感知组织的研究为 RBC 中的特征检测和部件识别建立了必要的基础,其主要原理是根据图象中的非偶然性图象关系可以推理相应的空间结构特性(见 3.2.3 节),例如根据图

象中的直线可以推论三维空间中相应的物体边缘也是直线。类似的还有对称性、平行性、共端点等。

(3) 部件的定义

任何基元要作为物体识别的基础都应该是可以迅速识别和具有对视点 and 噪声的恒常性。因此根据感知组织原理检测的非偶然性特征是区分不同部件的基础。在 RBC 理论中部件的构成方法是在广义圆柱的基础上，通过四个属性在二到三层非偶然性关系上的变型来产生 36 种部件（见图 3.33），所产生的部份部件以及它们的组织如图 3.32 所示。四个属性中的 3 个描述截面的特征，即形状、对称性以及当沿轴扫描时截面的不变性。第四个则描述轴的形状。即：

① 模截面

a. 边缘：S 直线，C 曲线

b. 对称性：

- ++ 对称的，在旋转和反射时不变
- + 对称的，在反射时不变
- 不对称

c. 当沿轴扫描时横截面大小不变性

- + 保持不变
- 扩展
- 扩展和收缩

② 轴

d. 曲率

- + 直线
- 曲线

这四个属性的不同值表示非偶然性特征的不同：直线对曲线，对称对不对称，平行对不平行。横截面边缘和轴的曲率是按共线性，或曲线性来区分。横截面大小是保持不变或扩展是通过平行性来检测的：横截面不变将产生一个具有平行边的广义锥；而横截面扩展将产生不平行的边，而一个先扩展后收缩的横截面将产生椭球状的不平行边，并有一个正曲率的极大值。总之，通过这些属性的变化可描述多种多样的复杂形状，这里不再详细介绍，可参阅 [Bie 85]。

(4) 有限数量的部件是否足够

Biederman 认为为建立物体的模型只需要有限数量的部件就足够。他认为对此论断的支持来自两方面。一方面是实验性的。如前所述人类视觉所具有的迅速识别能力，说明在识别时不太可能采用基于绝对测量的定量判断。同时人们还发现人类记忆不规则形状的能力很差。容易出错，这可能是记忆时是把不规则形状规整为最相近的形状产生的。这说明识别物体主要基于定性的特征。区分这些特征只要求二到三层视点恒常差异性。

对有限数量的部件（实际上是 36 种）已经足够的论点还可以通过估算来支持。为支持这个论点需要作两方面的估算：(a) 人们能感觉到的不同类物体的数量，(b) 由 36 种部件所能表示的物体类数量。如果要证明 36 部件是足够的，那么 (b) 的数量要大于 (a) 的数量。Biederman 从两方面来对 (a) 的数量作估算。一方面是从词典查出不同的基本物体种类（大约为 1000 种）

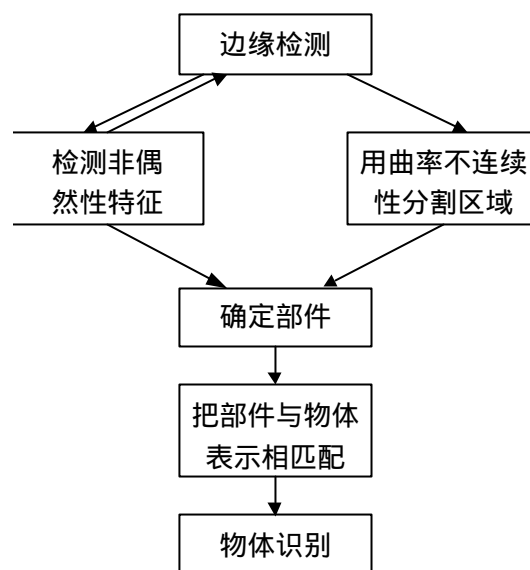


图 3.31 建议的物体识别过程

再加上一定的放大（放大 3 倍），再加上不同的子类（10 个子类），这样就有 30,000 种不同类的物体；另一方面是从人类学习新物体的速度来估计。如果每天学习 4—5 个物体，3 万种物体将要求学 18 年。人的年龄在 2—6 岁时学习的速度最快，可平均达到 4—5 个物体，而对其他年龄，特别是成年人达不到这个速度。所以 3 万种物体是一个相当留有余地的估算。下一个问题是 36 种部件可以表示多少种物体。这个问题的计算可用表 3.2 来表示。

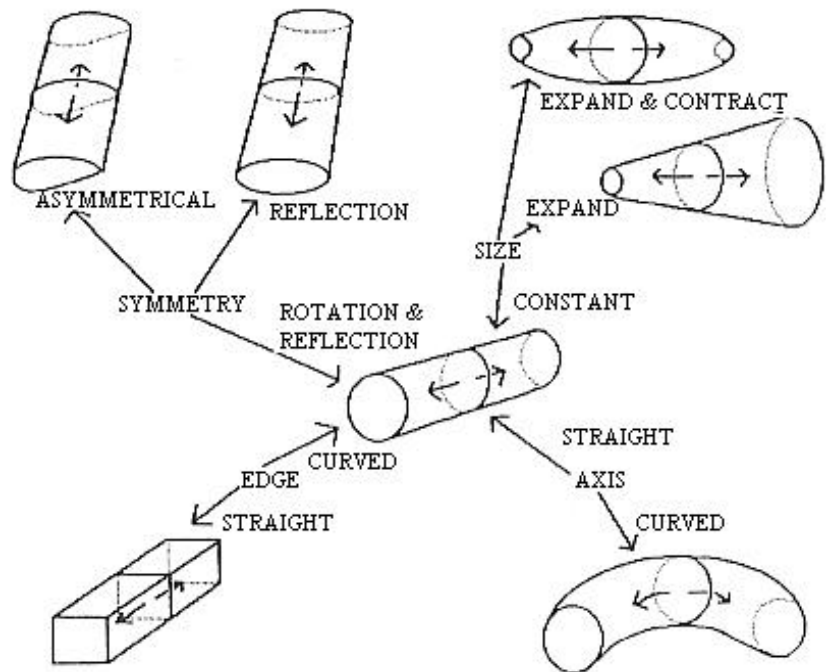


图 3.32 可通过非偶然性特征检测的广义锥变形

HUMAN IMAGE UNDERSTANDING
CROSS SECTION

Geon	Edge	Symmetry	Size	Axis
	Straight S Curved C	Rot & Ref ++ Ref + Asymm -	Constant ++ Expanded - Exp & Cont -	Straight + Curved -
	S	++	++	+
	C	++	++	+
	S	+	-	+
	S	++	+	-
	C	++	-	+
	S	+	++	+

图 3.33 根据非偶然性特征不同得到的部分部件样本

表 3.2 两种部件的表示能力

36	第一个部件 C_1
X	
36	第二个部件 C_2
X	
3	大小 [$C_1 >> C_2$, $C_2 >> C_1$, $C_1 = C_2$]
X	
1.8	C_1 在顶部或底部 (代表 80% 的物体)
X	
2	连接的本质 [端对端, 端对边]
X	
2	连接在 C_1 的长的或短的表面
=	55,987, 这是两个部件可能产生的物体数。如果有 3 个部件, 不考虑关系可有 $55987 \times 36 = 2$ 百万可能的物体。

如果我们只考虑有两种部件构成的物体数, 据保守的估算约为 55,000 种。估算时考虑了 5 种关系: (a) 部件 A 是在部件 B 的上面或下面, 据估计这可以概括 80% 的情况。(b) 任何相连的部件之间是端对端 (在连接处, 两个部件的面积相同), 就象手臂与人的身体的中接那样或端对边, 这就产生一个或两个曲率不连续点。(c) 部件 A 的大小是比部件 B 大得多、小得多, 或大致相等。(d) 每个部件是连接在它的较长边或较短边。

如果再增加第 3 个部件, 即使在不考虑第 3 个部件与其它两个部件之间的关系, 就可以表示 2 百万种物体。如果花 18 年来学这样多种物体, 那么每天要学 304 种。由于表示能力是部件或关系的乘积的函数, 所以, 部件或关系数的轻微增加就会使表示能力急剧提高。总之, 36 种部件所能产生的物体数将大大超过人们能感觉到的不同类物体数。所以 36 种部件是足够的。

参考文献

- [Bie 85] Biederman, I., Human Image Understanding: Recent Research and a Theory, Computer Vision, Graphics and Image Processing 32. (1985) 29-73.
- [Hoc 81] Hochberg, J., Levels of Perceptual organization, In M. Kubovy and J. R. Pomerantz, (eds) Perceptual Organization, 1981, 255.
- [Kan 79] Kanizsa, G., Organization in Vision, Praeger, New York, 1979.
- [Low 82] Lowe, D.G. & Benford, T.O., Segmentation and aggregation: An approach to figure-ground phenomena, proc. ARPA Image Understanding Workshop, Stanford, Calif. Sept. 1982, P46.
- [Low 85] Lowe, D.G., Perceptual Organization and Visual Recognition, Kluwer Academic Publishers, 1985.
- [Mar 82] Marr, D., Vision, W. H. Freeman and Company, 1982.
- [McC 90] McCafferty, J.O., Human and Machine Vision, Computing Perceptual Organization, Ellis Horwood, 1990.
- [Pen 86] Pentland, A.P.ed., From Pixels to Predicates: Recent Advances in Computational and Robotic vision, Ablex Publishing Corporation, Norwood, new York, 1986.
- [Pen 86] Pentland, A.P., Perceptual Organization and the Representation of Neural Form,

Artificial Intelligence, 28(1986)293-331.

[Pen 86] Pentland, A.P., Part Models, In Proceedings of Int. Conf. Pattern Recognition and Computer Vision, MiamiBeach, Florida, June 22-26, 1986, 242-249.

[Pen 88] Pentland, A.P. The Parts of Perception, in Advance in Computer Vision Vol. II, eds by Brown, C., Lawrence Erlbaum Associates, 1988.

[Sar 93] Sarkar, S. & Boyer, K.L., Perceptual Organization in Computer Vision: A Review and a Proposal for a Classificatory Structure, IEEE Tran. On System, Man and Cybernetics, Vol., 23, No. 2 March/April, 1993, 382-399.

[Wit 83] Wttin, A.P., &Tenenbacem, J.M., On the Role of Structure in Vision, in Human and Machine Vision, Beck, Hope & Rosenfield eds. Academic Press. Inc, 1983, 481-543.