

## 第二章 Marr 关于视觉的计算理论

### 2.1 引言

#### 2.1.1 视觉是信息处理问题

#### 2.1.2 视觉系统的表象

### 2.2 初始简图

#### 2.2.1 未处理的初始简图

##### 2.2.1.1 亮度变化的检测与最优滤波器

##### 2.2.1.2 Marr 的亮度变化检测理论的生理实验根据

##### 2.2.1.3 不同通道信息的合并

##### 2.2.1.4 过零点的尺度理论 (scaling theorems for zero-crossing)

#### 2.2.2 完全的初始简图 (full primal sketch)

### 2.3 2.5 维简图

#### 2.3.1 图象分割

#### 2.3.2 2.5 维简图

### 2.4 三维模型

### 2.5 视觉系统的组织

### 2.6 小结

## 2.1 引言

早期的计算机视觉研究主要集中于积木世界的理解以及与此有关的低层视觉信息处理。积木世界理解就是根据线画图来理解由多面体构成的景物。这方面的研究开始于 Roberts<sup>[Rob 65]</sup>，到七十年代中期已经能处理相当复杂的线画图，并且对光照和阴影条件可以几乎不作限制<sup>[Wa 72][Huf 77]</sup>。但要从实际景物或图片获得线画图不可避免地要通过低层视觉信息处理。当时这方面的研究缺乏科学的理论指导。在很大程度上低层视觉信息处理的方法是一些技巧，而不是科学，并且已经面临严重的困境。当时学术界普遍认为低层视觉处理从本质上来看就难以得到各种有用的描述。因此需要用关于景物的先验知识和启发式知识来指导。但在实际建立由知识作自顶向下导引的视觉处理和理解程序时并未取得预想的成功，而是遇到难以克服的矛盾。总之，到七十年代中期计算机视觉研究迫切需要理论指导以推动进一步的发展。

另一方面，自 60 年代以来对人类视觉的研究也已取得了一系列重要的成果。Hubel 和 Wiesel<sup>[Hub, Wie 62, 68]</sup>关于大脑视皮层细胞感受野的论述，以及 Campell 和 Robson<sup>[Cam 68]</sup>关于视觉系统空间频率通道的论述是其中两项最重要的研究成果。如果把视觉看成是一种信息处理过程，上述结果就形成了两种不同观点：一些人认为视皮层是特征检测器，另一些人则认为视皮层是某种空间 Fourier 分析器。同时，这两种观点又都有不少无法解释的实验结果。因此，

对人类视觉的研究迫切地需要理论的指导。美国麻省理工学院（MIT）人工智能实验室的 D.Marr 教授就是在这种情况下，立足于计算机科学，总结了心理物理学、神经生理学、临床神经病学等在视觉研究上已取得成果的基础上提出了视觉计算理论。他的理论对人工智能和神经科学都产生了深远的影响。

西方的思维科学来源于逻辑和计算理论，根据这个理论，Newell 和 Simon<sup>[new 81]</sup>提出了物理符号系统的概念。他们认为无论计算机与人脑在硬件结构上如何不同，但从计算理论的层次上来说它们都是具有产生、操作和处理抽象符号能力的物理符号系统。物理符号系统对思维科学来说就象进化论对任何生物学那样是基础。在这样的计算理论看来，任何一种复杂的精神智力活动在本质上都是以离散符号处理为核心的计算过程。这种以逻辑和计算理论为核心的一系列概念加深了我们对智力本质的认识。Marr 提出的视觉计算理论与此是一脉相通的。他的计算理论使人们对视觉机理的研究提高到一个新水平。

Marr 的视觉计算理论是首次提出的阐述视觉机理的系统理论，并且对人类视觉和计算机视觉的研究都产生了深远的推动作用。但它还远未解决建立人类视觉计算理论的问题。如以后的章节中将会讨论的那样在实践中也已遇到了一些严重困难。因此这个理论还需要通过研究不断改进和完善。

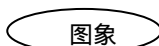
## 2.1.1 视觉是个信息处理任务

Marr 认为“看”就是人通过观察认知有什么东西和它们在什么地方。换句话说，视觉是一种处理过程，经过这个过程从图象中发现外部世界中有什么东西和它们在什么地方。因此，视觉首先是一个信息处理任务。一个信息处理任务可以，并且应该从三个不同层次来研究和理解。Marr 指出正是由于不能正确地地区分这三个层次中涉及的不同的研究问题阻碍了人工智能和心理学、神经生理学中对视觉的研究。第一个层次是信息处理的计算理论。在这个层次研究的是对什么信息进行计算和为什么要进行这些计算；第二个层次是算法。在这个层次研究的是如何进行所要求的计算，也就是要设计特定的算法；第三个层次是实现算法的机制或硬件，在这个层次上研究完成某一特定算法的计算机机构。例如，根据 Fourier 分析理论，任意连续函数可用它的 Fourier 频谱来表示，因此 Fourier 变换是属于第一层的理论。而计算 Fourier 变换的算法，如各种快速 Fourier 算法是属于第二个层次的。至于实现快速 Fourier 算法的阵列处理机就属于硬件的层次。在设计实现计算理论的算法以前必须先要有计算理论。因为只有确切地了解所进行计算的含义以后才能认真考虑设计算法。在视觉研究中有必要强调上述观点，这样有助于明确研究的目的。在人工智能领域中的研究应该主要在第一个层次，在这一层次上研究的问题应该是类似于这样的问题：“从原理上讲，如何才能根据图象中的影调信息获得物体的形状？”，“在解决视觉问题中光度学可起什么作用？”，“如果应用光度学还不足以解决问题，那么对环境的本质需做什么假设，以便用作求解时的约束条件”等。相比之下，在图象处理中经常被采用的松弛标记法所涉及的问题是根据应用环境把合理的约束直接写入到算法中去，它只不过是一种编写程序的风格而已，并不涉及视觉理论本身。因此，对松弛标记法的研究是属于信息处理的第二个层次，它不应是人工智能领域中研究的重点。

## 2.1.2 视觉系统的表象 (representation)

如前所述, Marr 认为视觉是一个信息处理任务。Marr 同时又指出视觉和处理并不完全相同。因为如果我们能够认识外部世界有什么东西在什么地方, 那么我们的大脑必须以某种方式表象 (representing) 这种信息——以尽可能丰富的色彩、形状、美感、运动和细节等多方面来表象这种信息。所以, 视觉研究不仅必须探讨我们是如何从图象中提取外部世界中对我们有用的各个方面, 而且还必须探讨内部的表象。所谓表象是指一种能把某种实体或某几类信息表达清楚的形式化系统, 以及说明该系统如何行使其职能的若干规则。使用某一种表象描述一个给定的实体所得到的结果称为该实体在这种表象下的一个描述。例如, 阿拉伯数制, 二进制数制, 它们都是表象“数”的形式化系统。阿拉伯数字表象由数集  $\{0, 1, 2, \dots, 8, 9\}$  中所抽取的一串符号组成; 构造特定整数  $n$  的描述规则是: 先把  $n$  分解为数符与位值的乘积之和, 位值用 10 的幂表示, 再按幂的大小把数符自左向右排成一行。因此叁拾柒等于  $3 \times 10^1 + 7 \times 10^0$  变成 37, 即该数在阿拉伯制下的描述。这个描述明确表达了数按 10 的整数幂所作的分解。在二进制数制中, 数叁拾柒的描述是 100101; 它明确表达的则是数按 2 的幂所作的展开, 但这样的描述不可能一蹴而就, 而要通过构造一连串的表象来实现。这些表象开始于直接从图象中得出的描述, 而每一个表象都将有助于尔后进行的处理工作, 经过一步一步的处理, 逐步恢复物体形状的越来越客观的物理性质。按照 Marr 的设想, 从图象推得形状信息的过程分成三个表象阶段: 初始简图 (primal sketch)、2.5 维简图 (2.5 dimensional sketch) 和三维模型 (3D model)。

决定图象亮度的主要因素有四个。它们是: (1) 几何关系; (2) 可见表面的反射情况; (3) 景物的照明情况; (4) 观察点的方位; 在图象中所有这些因素是混杂在一起的。某些灰度变化是由某一种因素引起的, 而另一些变化则是由另一种因素引起的, 还有一些变化则是由几种因素共同作用的结果。在视觉处理的第一阶段 (即低层视觉阶段) 所作处理的目的就是要分清楚哪一些变化是由哪一些因素引起的, 进而建立一些表象, 用这些表象把上述四个因素区分开来。大体上说, 可通过两个步骤来实现这一目的。第一步, 由图象中的变化和结构获得适当的表象。这包括检测亮度的变化, 表象并分析局部的几何结构, 以及检测光源、强光部 (highlight) 和透明度等照明效应。总之, 这是关于二维图象性质的表象。这一步得到的表象称为初始简图。第二步, 对初始简图进行一系列处理运算, 推导出一个能反映可见表面几何特征的表象。其中包括表面朝向, 观察者的距离, 以及朝向和距离的不连续性, 表面的反射情况, 以及对主要照明情况的某种粗略的描述。第二步得到的表象称为 2.5 维简图。初始简图和 2.5 维简图都是在以观察者为中心的坐标系中构成的, 正因为它们的这一结构特征, 所以被称为图。它所表达的空间关系是指在观察者视网膜上的二维关系, 不是观察者周围的外部世界相对于观察者的三维关系。第三个阶段是被观察形状的三维结构组织在以物体为中心的坐标系中表象, 以及在这种坐标系下对物体表面性质的一些描述, 这个阶段称为三维模型表象。这个视觉系统的表象层次总结成图 2.1。在第 2.2 节到第 2.4 节中我们将对这个表象系统作较详细的叙述。



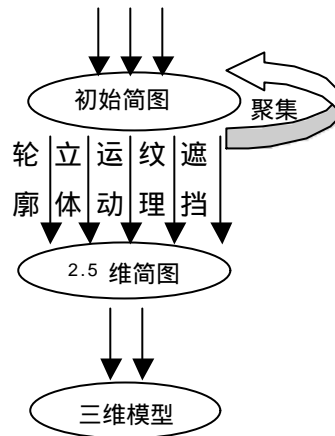


图 2.1 视觉系统的三个表象层次

## 2.2 初始简图 (primal sketch)

人类视觉系统的输入是外部世界投射在视网膜上的图象，这个图象实质上是一个二维的亮度阵列（在计算机视觉系统中，外部世界在胶片或摄像机的感光面上产生亮度阵列），而不能直接表示其它信息，因此需要进一步处理。画家作速写时用边、线、点这样的符号就可以勾画出景物的素描。这样的素描图与实际景物在视网膜上产生的图象亮度阵列是很不相同的，但人们可以毫无困难地识别它们。这说明视觉对图象所作的第一个运算是把它转换成一些原始符号构成的描述。这些描述所反映的不是亮度的绝对值大小，而是图象中的亮度变化和局部的几何特征。Marr 把这样的描述称为初始简图。初始简图是一种基元表象，它由两部分组成。第一部分，用边缘段 (edge segments)、线 (bar)、斑点 (blob) 和端点 (terminations) 这些基元构成对图象中亮度变化的描述，这被称为未处理的初始简图。第二部分利用虚拟线来完全而明确地表示几何关系，并且通过各种方式对原始的基元进行选择、聚合和概括等过程来构成更大、更为抽象的标记 (tokens)。最后得到的描述层次可以复盖某一尺度范围，这被称为完全的初始简图。图 2.2 中所示为亮度图象、未处理的初始简图和在生产完全的初始简图时所进行的聚集过程。

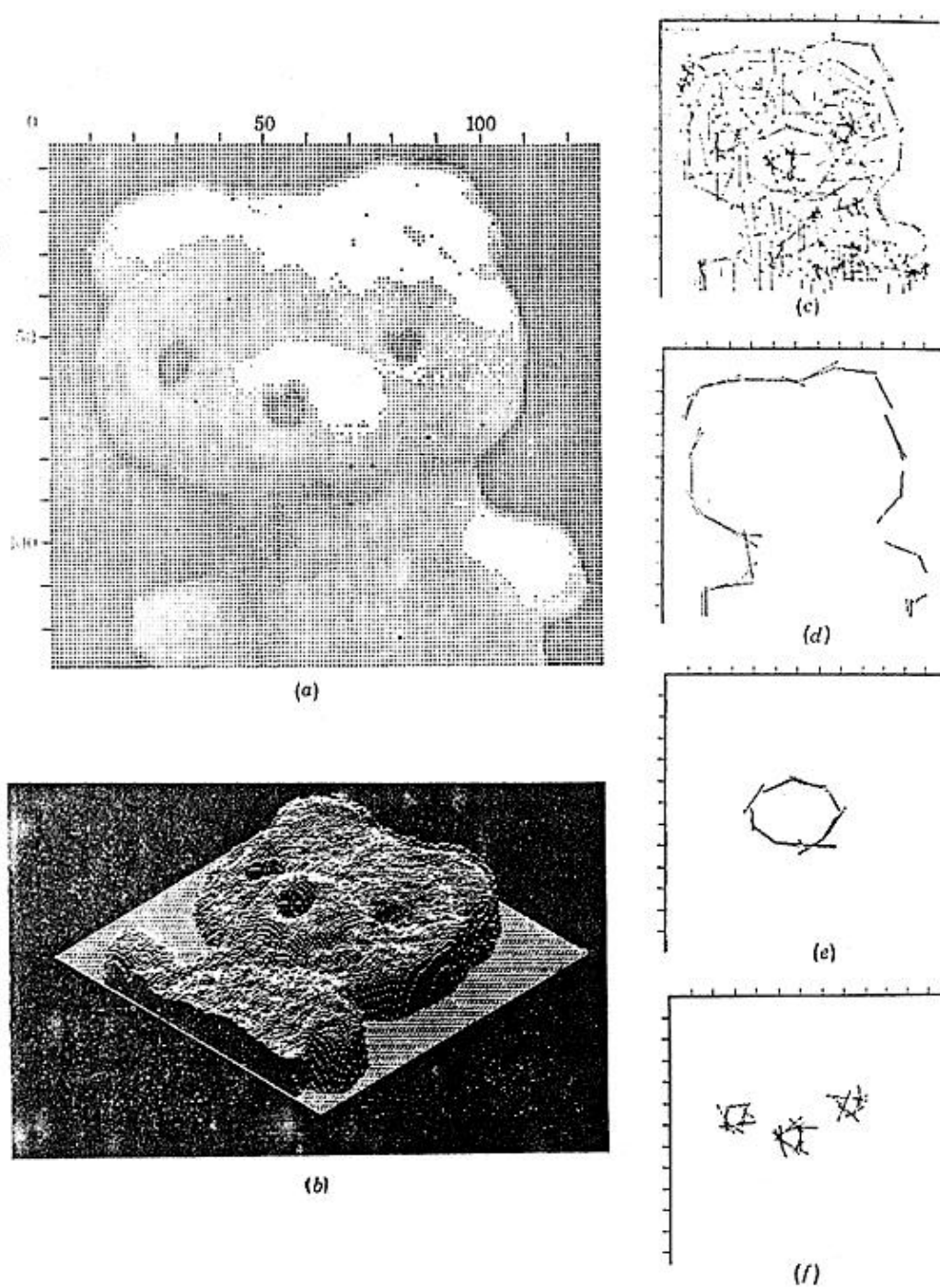


图 2.2 初始简图 (AI. App. p.212. fig 8-3)

(a)亮度图象；(b)图象的亮度；(c)未处理的初始简图的分割，这时未进行许多聚集处理；  
 (d)(e)(f)是抽取完全的初始简图过程中找到的聚集。

## 2.2.1 未处理的初始简图 (raw primal sketch)

计算未处理的初始简图涉及的问题是检测亮度的变化、确定亮度变化的空间位置，以及如何用符号来表示这些亮度变化。

## 2.2.2 亮度变化的检测与最优滤波器

### 1. 亮度变化的检测

亮度的变化可由亮度的梯度和亮度梯度的变化来描述。为简单起见，以一维的情况为例说明。如用一阶差分来近似，那么可通过把该点左、右两边的亮度值相减求得。这相当于把此点一边各点的亮度乘以+1，把此点另一边的亮度乘以-1，然后把两边的乘积相加，所得结果即为该点的亮度梯度。在数字图象处理的文献中把这样的运算称为与一阶差分空间算子做卷积。这样的运算是一种局部运算。现在的问题是仅对左、右相邻点，还是在更大的范围内求亮度变化。如图 2.3(a)所示是缓慢变化，那么为了检测亮度梯度就需要在更大的范围内使左、右两边的亮度求差，即采用较大的空间算子。与此相反，如果亮度分布是图 2.3(b)中所示的急剧变化，那么用大的空间算子就会丢失这些急剧的亮度变化，为此要采用较小的空间算子。总之，如果图象中的亮度在某个范围内变化，那么就要采用相应大小的不同的空间算子。

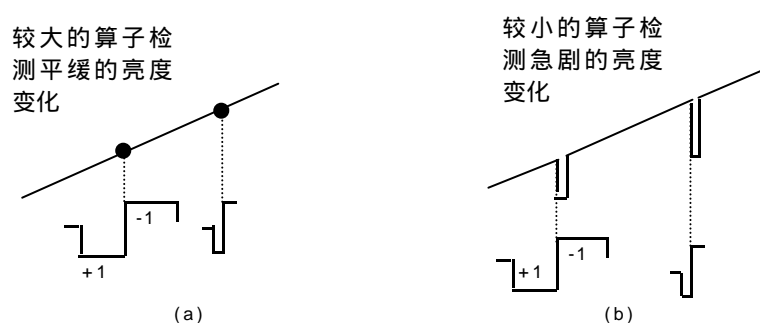
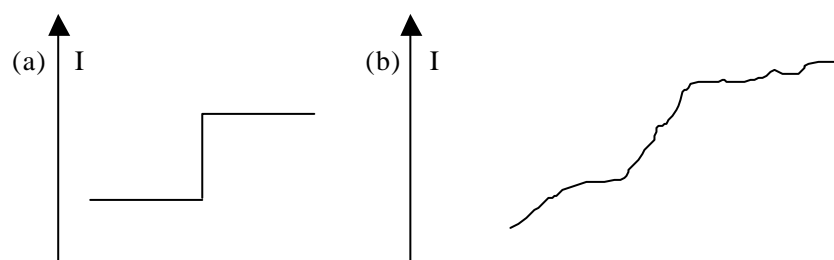


图 2.3 采用不同大小的算子检测不同尺度的亮度梯度

亮度梯度的变化也是对亮度变化的有用描述。一阶微分可检测亮度梯度，二阶微分可检测亮度梯度的变化。如果亮度分布是图 2.4(a) 中所示的台阶状，那么在亮度的一阶微分图形上出现一个峰点(b)，而在二阶微分的图形上相应地出现一个过零点(c)。通过确定亮度二阶微分上的过零点就可以确定亮度台阶的位置。



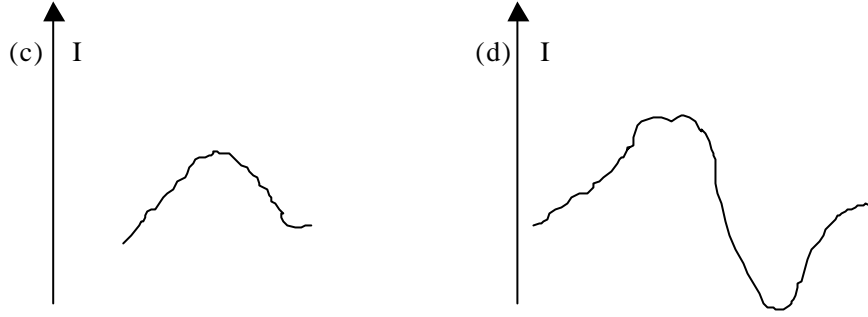


图 2.4 台阶状亮度分布(a)的一阶和二阶微分，(b)是有噪声情况下的实际边缘，(c)和(d)则是有噪声情况下的一阶和二阶微分。

在实际景物的图象中，在两个平面之间的边缘处，亮度急剧地由一个值变为另一个值，亮度的分布呈台阶状。实际上，由于不可避免地存在噪声，台阶状的变化经常被模糊成图 2.4(b)中所示的情况。为了确定这样的有噪声的亮度变化的位置需要经过以下四步处理：

第一步从图象建立平均亮度阵列。用局部的亮度平均来减少噪声的影响。 $I_i$  是  $i$  点的图象亮度， $A_i$  是  $i$  点周围的亮度的局部平均值：

$$A_i = \frac{I_{i-1} + I_i + I_{i+1}}{3}$$

第二步，从平均亮度阵列建立平均一阶差分阵列。为此，要把右邻差分  $A_{i+1} - A_i$  和左邻差分  $A_i - A_{i-1}$  相平均。这相当于把  $A_{i+1}$  和  $A_{i-1}$  相平均。 $F_i$  是平均亮度的平均一次差分：

$$F_i = \frac{(A_{i+1} - A_i) + (A_i - A_{i-1})}{2} = \frac{A_{i+1} - A_{i-1}}{2}$$

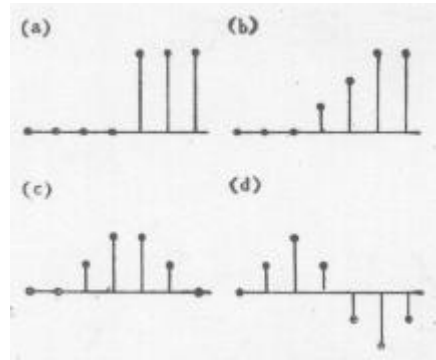


图 2.5 (a) 理想边缘的亮度分布；(b) 对(a)进行平均运算的结果；(c) 对(b)进行平均差分的结果；(d) 对(c)进行平均差分的结果。

第三步，从平均一次差分阵列建立平均二次差分。这相当于求一次差分的平均。 $S_i$  是平均亮度的平均二次差分：

$$S_i = \frac{(F_{i+1} - F_i) + (F_i - F_{i-1})}{2} = \frac{F_{i+1} - F_{i-1}}{2}$$

第四步，在所求得阵列上记下峰点、陡峭的斜率和过零点以寻找边缘。

平均过程把理想的台阶和被噪声所模糊的台阶都转换成平滑的台阶（图 2.5(b)），一次差分过程又把平滑了的台阶转换成凸缘（图 2.5(c)）。而二次差分又把凸缘变成了 S 形曲线（图 2.5(d)）。这个 S 形曲线在正的和负的尖峰之间陡峭地穿过零线。

进行平均以及随后差分运算的作用可以组合到一次总的平均过程中去。输入点  $I_j$  对输出点  $O_i$  的影响与两点之间相隔距离关系如下述公式所示：

$$o_i = \sum_j P_{j-i} \times I_j$$

其中，P 被称为点扩散函数，因为这个函数表示单个孤立的、亮度不为零的点如何在输出图象中扩散它的影响。当利用这个公式根据输入来确定输出时，我们说输入被点扩散函数所滤波或和点扩散函数进行卷积。例如，图 2.6(a)所示的点扩散函数相当于一维的三点平均。图 2.6(b)所示的点扩散函数相当于先平均，然后平均相邻差分。图 2.6(c)所示的点扩散函数相当于先平均，然后进行二次平均相邻差分。

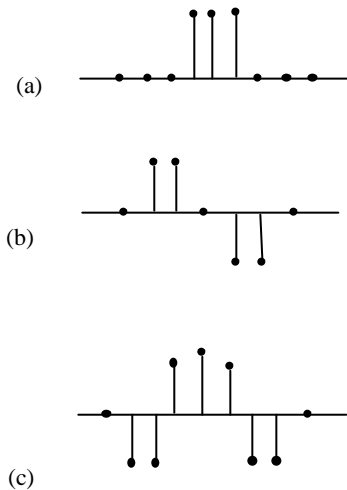


图 2.6 平均点扩散函数的一阶和二阶差分(a)一个平均函数；(b)是(a)中所示点扩散函数的平均相邻差分；(c)对平均的一阶差分进行平均的相邻差分。

以上用一维的例子说明检测图象中的亮度变化可通过对图象用某种点扩散函数（滤波器）作滤波或卷积，然后在滤波图象中寻找峰点或过零点等特征来实现。点扩散函数的扩散范围，或滤波器的大小应取决于图象中的亮度变化情况。

## 2. 最优滤波器

在自然景物的图象中亮度变化的尺度（Scale）很宽，从很缓慢的变化到很急剧的变化都可能存在。没有一个单独的滤波器可在所有的尺度下都是最优的。所以需要寻找一种分别处理在不同尺度上发生的变化的方法。根据 Campbell 和 Robson<sup>[cam 68]</sup>的研究表明人类视觉信息的处理是并行地在若干独立的空间频率通道中进行的。按这种概念首先要以各种分辨率对



图象作局部平均，然后检测发生在不同分辨率上的亮度变化。因此，在设计这种最优滤波器时要综合考虑两个实际因素。首先，由于图象滤波的目的是减小亮度变化的尺度范围，所以滤波器的频谱应该是平滑的，并在频率域中是带通的。这个条件可表示为滤波器应有小的频率方差  $\Delta w$ 。第二个因素可称为空间定位约束 (Constraint of spatial localization)。如前所述，景物中可引起图象中亮度变化的因素为：(1) 照明变化；(2) 可见表面的方向和相对观察者的距离变化；(3) 表面反射率变化。重要的是这些因素可被认为是以它们各自的尺度在空间中定位的。所以，对滤波图象中的每点的作用应该是邻近点的平滑平均的结果，而不是任何一种杂乱散布的点的平均。因此，这种滤波器在空间域中也应是平滑和准确定位的，也就是滤波器应有小的空间方差  $\Delta x$ 。

不幸的是这两个定位性要求 (一个在空间域，一个在频率域) 是相互冲突的。事实上这两者之间要满足不确定性原理 (见[Bra 65], 160-175)。不确定性原理要求  $\Delta x \Delta w \geq \frac{1}{4} p$ 。此外，只有 Gaussian 函数才能满足这种最优关系：

$$G(x) = \left[1 / s(2p)\right]^{\frac{1}{2}} \exp\left(-x^2 / 2s^2\right)$$

其中  $s$  是 Gaussian 函数的方差

在二维的情况下

$$G(r) = \frac{1}{2ps^2} \exp\left(-g^2 / 2s^2\right), g = \sqrt{x^2 + y^2}$$

Gaussian 滤波器能在这两个矛盾的要求中作最优的综合平衡。因此，计算初始简图中的第一步是使图象 (I(x, y)) 与一组不同尺度的 Gaussian 滤波器 (即不同标准方差  $s^2$  的滤波器) 作卷积。可用符号把这种计算表示为  $G * I(x, y)$ 。

在出现亮度变化的地方一阶导数就会有峰点。相应地，亮度的二阶导数上就会有过零点 (zero crossing)。因此，可通过寻找下述函数中的过零点来检测亮度的变化：

$$f(x, y) = D^2[G(r) * I(x, y)]$$

其中  $I(x, y)$  是图象， $*$  是卷积算子。根据卷积的微分规则：

$$f(x, y) = D^2 G * I(x, y)$$

算子  $D^2 G$  与图象  $I(x, y)$  卷积的结果和  $G(r)$  与  $I(x, y)$  卷积后进行二阶微分的结果完全相同。 $D^2 G$  形状看起来很像墨西哥草帽 (图 2.7)，因此有时被称为草帽形算子。

上述论述说明从原理上说，某一尺度下的亮度变化可通过把图象与  $D^2 G$  算子作卷积，并寻找方向导数，所以还要确定是在什么方向上取导数。为简化计算，如果可能的话希望使用各向同性的算子。唯一与方向无关的二阶微分算子是 Laplacian 算子  $\nabla^2$ 。Marr<sup>[Mar 80]</sup> 证明在某些很弱的假设下，使用  $\nabla^2$  算子是合适的。因此，上述二阶方向导数  $D^2(G * I)$  中的过零点可简便地在某一尺度下通过搜索  $\nabla^2 G * I(x, y)$  的过零点来求得。

$$\nabla^2 G(g) = -\frac{1}{4pw} \left[1 - \frac{g^2}{2s^2}\right] \exp\left(\frac{-g^2}{2s^2}\right)$$

其中

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}, g = \sqrt{x^2 + y^2}, w = s^4 / 4$$

这是一个旋转对称的函数，具有一个自由参数  $s$ ，它决定函数的空间尺寸大小。此函数横截面如图 2.7 所示。由图中可见，它的中央负向区域的宽度  $w_{2D}$  等于

$$w_{2D} = 2\sqrt{2}s$$

这个滤波器的大小由  $s$  的数值或等价地由  $w_{2D}$  的数值来确定。为了不使函数被过分地截短，它应在足够大的窗口内作计算，窗口宽度通常取为  $w_d \geq 3.6w_{2D}$

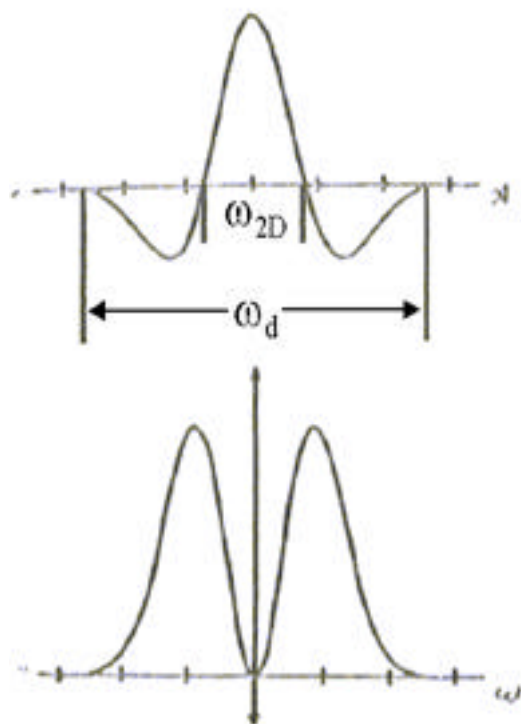


图 2.7  $\nabla^2 G(g)$  算子的横截面

图 2.8 中所示为  $\nabla^2 G$  算子与其它两种不是最优的算子的对比。因为我们是根据在频率域和空间域中同时要求定位的原则来选择 Gaussian 滤波器的，而在所示的例子中这两条要求都被严重地违背了。一个理想的一倍频程带通滤波器可满足在频率域中的定位要求，但

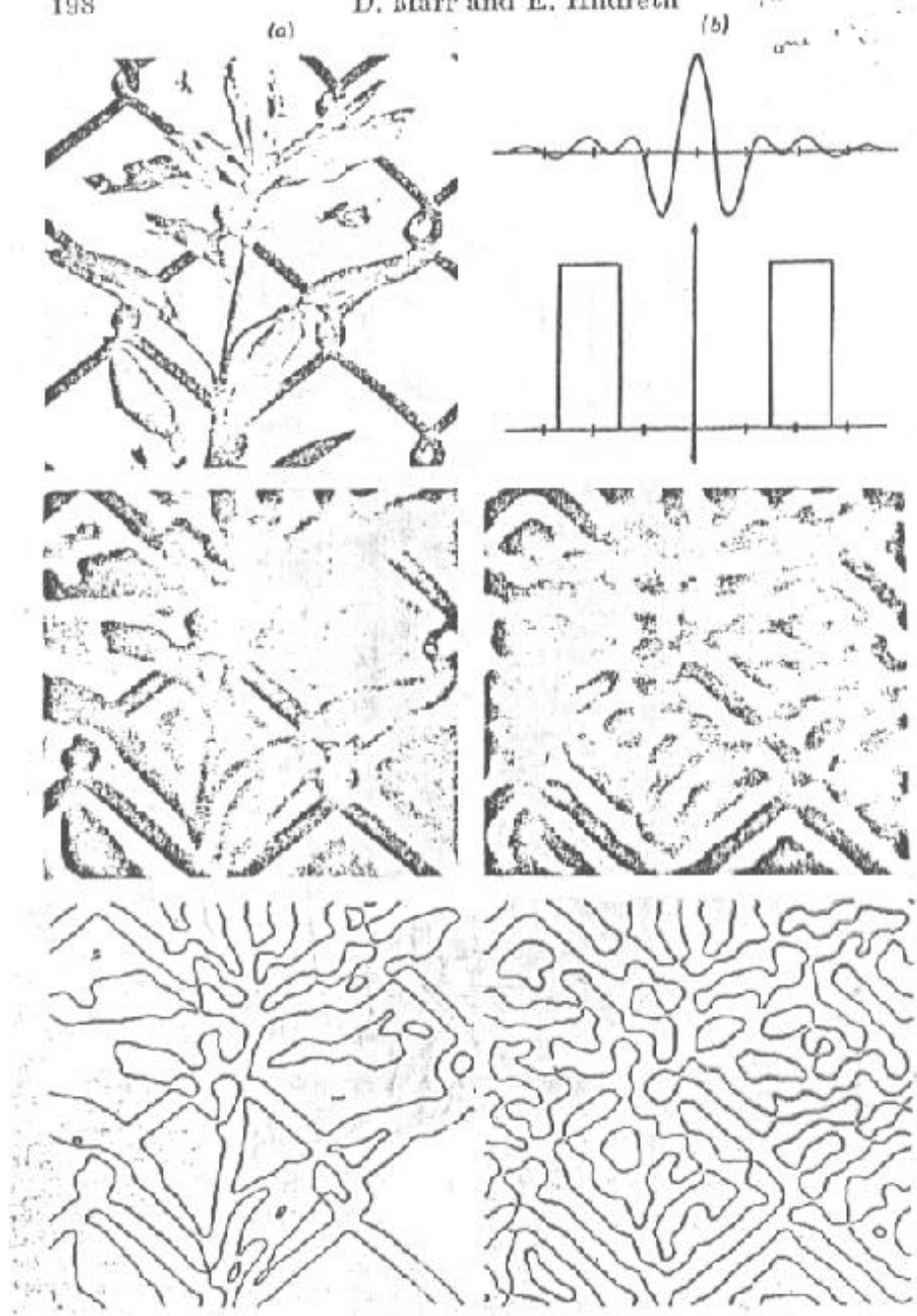


图 2.8  $\nabla^2 G$  滤波器与带通滤波器特性比较

(a) 中三幅图从上到下分别为原始图象、与  $\nabla^2 G$  作卷积所得结果、以及从卷积结果中抽取的过零点。(b) 上方第一幅图是一倍频程的带通滤波器、及其 Fourier 变换，其余两幅图的次序与(a)相同。从过零点图象中可发现，由于空间域中的旁瓣效应造成了较强的边缘反射，这样就产生了与原图无关的过零点。

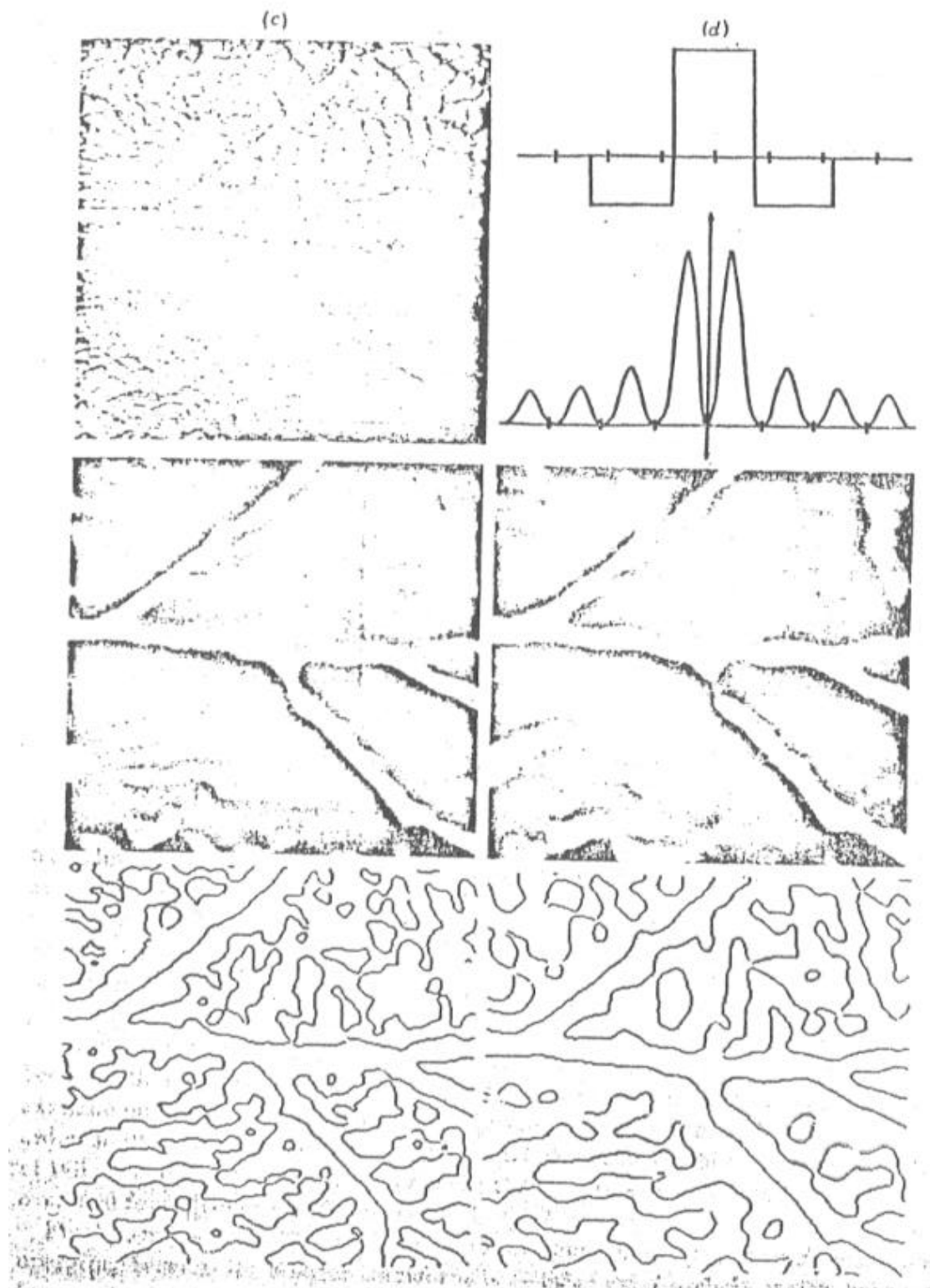


图 2.9  $\nabla^2 G$  滤波器与二阶微分（以方波近似）滤波器特性比较

(a) 中次序与图 2.8(a) 中相同；(b) 中上方所示为以方波近似的二阶微分滤波器及其 Fourier 变换。由于这种滤波器在频域中有旁瓣效应，模糊了频率的选择性，使得过零点数量减少，丢失了有用信息。

在空间域中不满足要求。因为严格的频带限制引起了空间域中的旁瓣（sideblobs）效应。因此在过零点图象中亮度的急剧变化除了产生过零点以外，还会产生反射。这些反射与图象没有直接的关系。因此在初级视觉处理中是不希望看到的。另一方面，如果滤波器在空间域中是截断的，那么在频率域中就会有旁瓣。

尽管在自然景物的图象中亮度变化的尺度很宽，但是只要用一组不同尺度（即不同方差  $\sigma$ ）的  $\nabla^2 G$  算子来检测相应尺度下的亮度变化，然后组合在一起就能处理各种情况。因此， $\nabla^2 G$  算子是一种通用的亮度变化、或边缘检测算子。

以上是在计算理论层次上对亮度变化检测方法的分析。那么如何在算法和硬件层次上实现这样的理论呢？现已发现  $\nabla^2 G$  算子可很好地用 DOG（Difference of Gaussian）函数来近似。DOG 算子是由一个宽的正向 Gaussian 函数与一个窄的负向 Gaussian 函数相加求得。这两个 Gaussian 函数的方差之比约为 1.6，如图 2.10 所示它们相加后的结果与墨西哥草帽形算子很接近。图象中的不同尺度的亮度变化需要有一组这种类型的 DOG 算子。

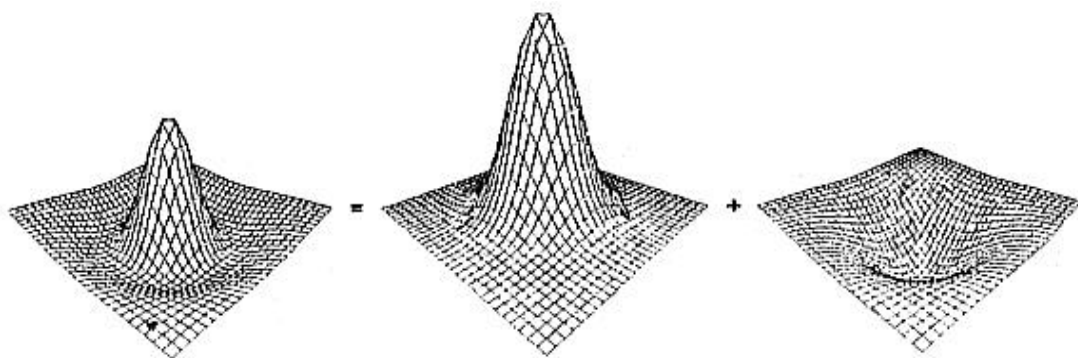


图 2.10 墨西哥草帽形算子的近似

用两个合适的 Gaussian 算子可以足够精确地近似先用 Gaussian 算子滤波，后求导所得到的算子。

#### 2.2.1.2 Marr 的亮度变化检测理论的生理实验依据

Marr 所提出的方法与在灵长目动物身上进行的神经生理学实验结果很好地符合，并且能对人的生理视觉系统的某些早已知道、但未能很好理解的特性作出了相当满意的解释。

关键的实验如图 2.11 所示。让实验动物观察各种正在移过白色背景的试验图案，试验图案中包括一条窄的黑带，一条宽的黑带和单个从白到黑的边缘。记录探针确定各种神经的反应，再把神经的反应和根据草帽形算子作出的预计相比较。图 2.12 所示为比较的结果。图 2.12(a)所示为三个从左向右移动的试验图案的轮廓。图 2.12(b)表示以合适的草帽形算子对给定的亮度分布作卷积得到的结果。图 2.12(c)中是在所谓的 X 神经节细胞上取得的实验数据。实际上有两种 X 细胞：一种 X 细胞的激发率与用草帽形算子作卷积所得图象的正向部分相似；另一种 X 细胞的激发率则与卷积所得图象的负向部分相似。为便于比较，图 2.12(c)

所示实际上是正的和负的细胞的激发率之差。图 2.12(b)和图 2.12(c)比较的结果表明二者惊人地相似。这意味着灵长目动物的视网膜确实进行了某种类似于草帽形算子卷积的处理工作。

如果对草帽形算子稍作修改，就可以改善相似性。前面已经说明草帽形算子可以用一个窄的正向 Gaussian 函数与一个宽的负向 Gaussian 函数之差来精确地近似。如果把宽的负向 Gaussian 算子的卷积结果相对于窄的正向 Gaussian 算子卷积结果稍加延迟，那么对试验图案的反应就如图 2.12(d)所示。这些新的曲线与灵长目动物的实验结果之间的相似程度大为改进。这两者的相似程度足以使我们根据作以下猜测：

- (1) 由灵长目动物的视网膜进行的部分处理在计算上相似于用草帽形算子作卷积。
- (2) 因为激发率不能是负的，所以必定存在两种视网膜细胞。一种细胞传输卷积图象的正向部分，另一种细胞传送负向部分。
- (3) 对两个细胞中的每一个来说，草帽形算子是通过激发和抑制两种运算的组合来实现卷积的。总之，激发和抑制运算的组合可实现一个卷积算子的功能，这个算子等效于取两个二维 Gaussian 算子作卷积所得图象之差。周围的抑制作用对于中央的激发作用要稍为延迟一些。

灵长目动物与人类在视觉系统方面很相似。因此，上述实验结果有力地支持了 Marr 对人类视觉系统中初级处理所提出的理论。然而，作为一种通用的边缘检测理论来说，虽然

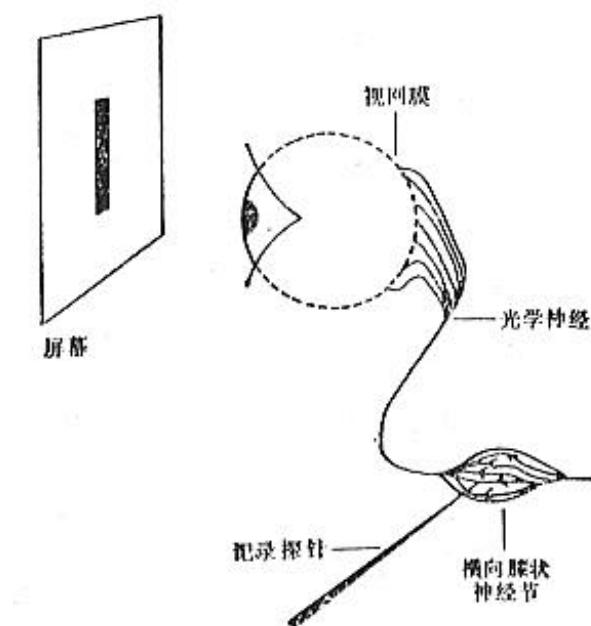


图 2.11 一个研究灵长目动物视网膜的输入输出特性的实验

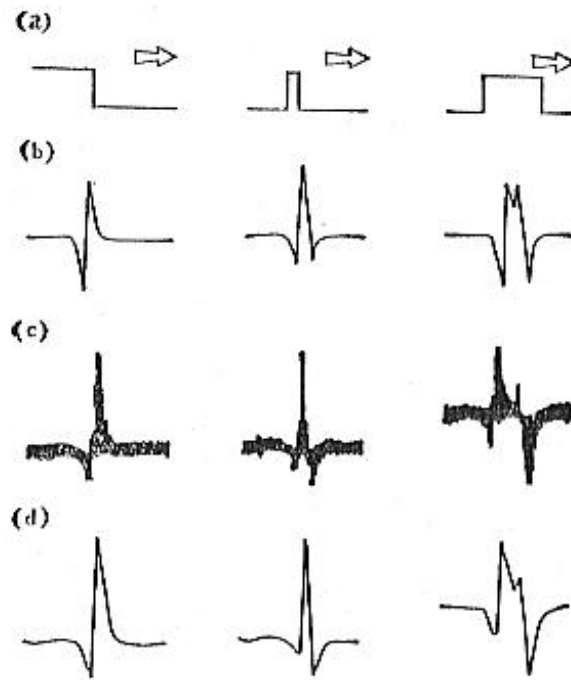


图 2.12 视网膜输入—输出特性的实验结果和草帽形算子的卷积结果的比较

(a) 试验图案的亮度分布；(b) 所处理图象中某一点预计的反应（作为时间的函数）；  
(c) 从被试验的猴子取得记录；(d) 加进延迟修正的预计反应。

很有说服力，但仍存在不少争论。Haralik<sup>[Har 80]</sup>和 Binford<sup>[Bin 81]</sup>等对他的理论提出了不少批评意见。此外，Mayhew<sup>[May 81]</sup>和 Frisby 对过零点的应用也提出了异议。他们发现人类视觉系统似乎也应用峰点，至少在照明变化和视差变化混合的情况下，在立体视觉中就是这样的。

### 2.2.1.3 不同通道信息的合并

为用符号描述图象的亮度变化需要用一组不同大小的滤波器（频率通道）来检测所有空间尺度上的亮度变化。大的滤波器检测宽的边缘，小的滤波器检测亮度急剧变化的细节。在同一频率通道内部，相邻的过零点通过聚集过程连结成小的过零点段，并被赋以方向。那么来自不同通道的信号应如何结合呢？一般来讲，如果没有附加的约束条件，这些通道的信号是互不相关的。但事实上在景物成象过程中存在着物理的约束条件。引起图象中亮度变化的物理因素主要包括：照明、可见表面的反射率、表面形状以及离观察者的距离和方位。这四个因素在图象中被编码，但它们有一个共同的性质，即它们在空间是有确定位置的。因此，亮度的变化就存在约束条件。Marr 称之为空间一致性假设（spatial coincidence assumption）：如果过零点边缘段出现在一组毗邻的独立  $\nabla^2 G$  通道中，并且这些边缘段在每个通道中的方向和位置相同，那么这样的过零点边缘段组表示的图象中的亮度变化就是由单一的物理因素引起的。

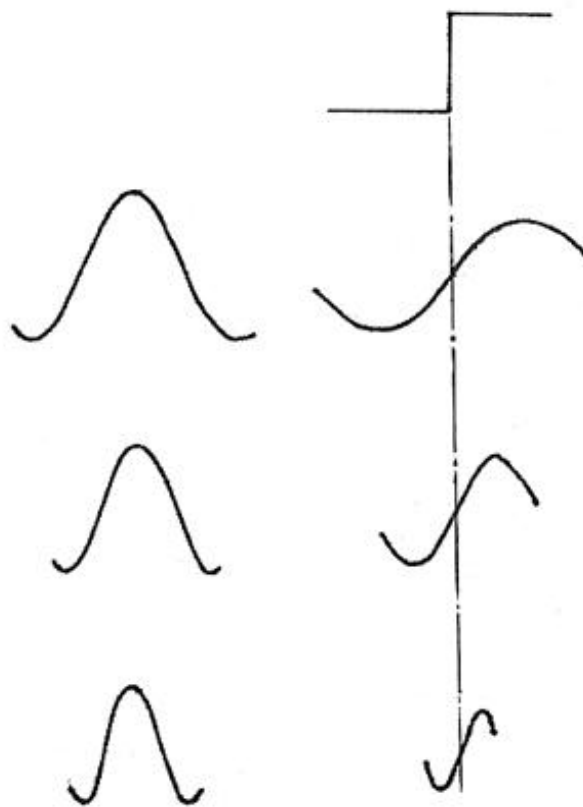


图 2.13 孤立边缘在不同的  $\nabla^2 G$  通道中产生的过零点

图 2.13 所示为孤立的边缘在不同的通道中产生的过零点。这些过零点在空间上是一致的，根据这一假设就可以把各通道的过零点合并到一起形成边、线、斑点和线段端点这样一些表示亮度变化的基元。基元的性质由它们的位置、方向、大小（长和宽），对比度来确定。未处理的初始简图（图 2.14）就是由这些基本符号构成的。



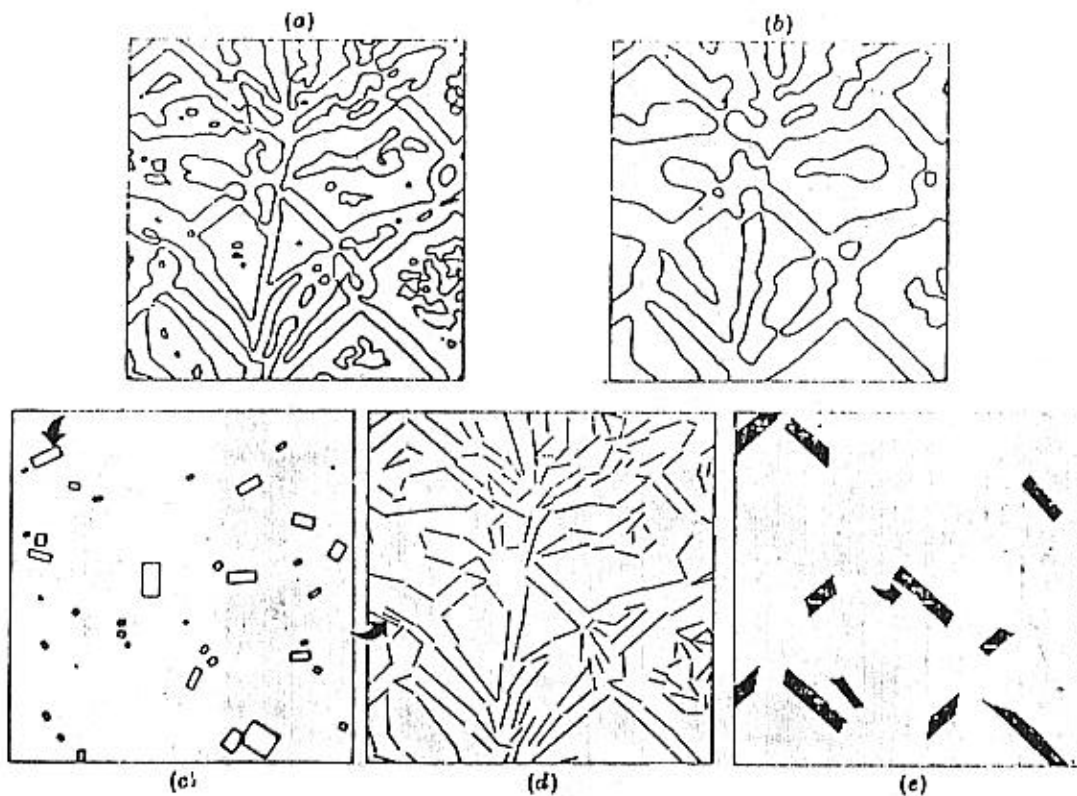


图 2.14 不同通道信息的合并 (Marr & Hildreth p.204 Fig 8)

图 2.14 所示为来自两个通道信息的合并, (a)(b)  $\omega=9$  和 18 的  $\nabla^2 G$  算子 ( $\omega=2S$ ) 处理图象所得的过零点。因为在较大的通道中不存在不与较小通道中对应的过零点。所以, 在合并的描述中边缘就与(a)中所示一样。(c)、(d)和(e)中所示为(a)中标注的位置处的符号描述: (c)表示斑点; (d)表示赋予边缘段的局部方位; (e)表示线。这些图形只表示包含在描述中的空间信息。以下所示为描述的例子。

(斑点	(位置 146 21)	(边 (位置 104 23)	(线 (位置 118 134)
(方位 105 )		(方位 120 )	(方位 120 )
(对比度 76 )		(对比度 -25 )	(对比度 -25 )
(长度 16 )		(长度 25 )	(长度 25 )
(宽度 6 ) )		(宽度 ) )	(宽度 4 )

上述描述相应于图 2.14 中箭头所指处。

#### 2.2.1.4 过零点的尺度理论 (Scaling Theorems for Zero Crossing)

大多数物理现象在空间或时间结构上都在很大的尺度范围内变化。因此, 尺度的概念对用符号描述图象或其它类型信号中的显著变化来说是至关重要的。必须在不同的尺度范围内以不同的详尽程度检测信号的变化。一般来说不同的物理过程在尺度变化时具有不同特性。在图象中, 根据不同的实际起因, 灰度的变化可发生在许多空间尺度上。通过分析信号在不

同尺度时的变化可得到关于基本物理过程本质的宝贵信息。例如，如果用不同尺度的  $\nabla^2 G$  算子对图象作卷积后检测到的过零点在空间是重合的，则说明存在一条实际的边缘，而不是表面上的标记或阴影。不仅有必要检测和描述在不同尺度下的变化，而且通过组合不同尺度下的描述可以得到许多有用信息。为检测在不同尺度下的亮度变化，Marr 提出了用不同大小的  $\nabla^2 G$  算子作卷积，并提出用某些启发式规则来合并从不同通道得到的信息。以后 Terzopoulos<sup>[Terz 82]</sup>又成功地把多级算法应用于重构视觉表面的问题，但如何把初级视觉处理中不同尺度的信息有效地组合起来这个重要的问题仍然没有解决。

最近 Witkin<sup>[Wit 83]</sup>提出一种描述在不同尺度下过零点特性的新方法。通过与各种大小的 Gaussian 滤波器作卷积来对一维信号进行平滑，然后确定其二阶导数的过零点位置，并跟踪当滤波器的大小增加（减小）时过零点位置的变化。这个过程产生了在  $x-s$  平面上由过零点构成的图形。其中  $x$  是过零点的空间位置， $s$  是 Gaussian 滤波器的方差。Witkin 通过分析过零点轮廓图形的形状对过零点作分类和标志，这样就得到了用于识别和对准所需的有效信息描述。之所以能这样做的原因主要是由于过零点轮廓的几何形状非常简单。过零点轮廓的形状在滤波器大小从小到大时不是线性的，而是碗形封闭曲线。当尺度增大时从来不会产生新的过零点。Babaud 等人<sup>[Bab 83]</sup>证明了在一维的情况下 Gaussian 滤波器是唯一的具有这种重要特性的滤波器，Yuille 和 Poggio<sup>[Yui 86]</sup>又把上述结论推广到二维。

## 2.2.2 完全的初始简图（full primal sketch）

未处理的初始简图能明确表达的只是局部信息。因此还需要根据这些局部信息构成更大尺度的轮廓和区域。视觉系统要实现这一点远不只是依次把点连接起来这么简单。首先，过零点段的数量非常多；其次，这些过零点段不只是一个简单的数字，而是一组描述。例如，聚集运算确定以下两组描述是否应该合并：

（边（对比度 4）	边（对比度 3）
（方位 15）	（方位 16）
（模糊度 5）	（模糊度 6））

因此，这里需要的聚集运算要以格式塔（Gestalt）聚集原理中的连续性、相似性等为基础。这样的聚集运算有两个目的：第一是构造记录了表面反射函数的较大尺度结构的位置标记（place tokens）。其次是检测与这些标记有关的参数的各种变化，以便于检测可见表面的方位以及离观察者距离的变化。一般来讲，完全初始简图是根据未处理的初始简图用递归方式计算求得的。

在未处理初始简图中的聚集过程大致是这样进行的：首先在图中选择大致相似的基元，把它们聚集起来形成线、曲线、较大的斑点或小区域，以表示在某一尺度上的空间结构。如图 2.15 所示，按照接近性、相似性原理先把点聚集成团。然后，根据连续性原理团又形成新的更大尺度上的标记一线。这些线又进一步形成箭头，用于形状分类。通过这种不断聚集的过程形成不同尺度上的标记或基元（primitive）。重复应用聚集过程的结果是产生更为抽象的描述。当然在图象中并不存在这样的虚拟线。但利用这样的抽象描述可作进一步的处理，

赋以特性和被聚集，或作为以后处理的输入。

未处理的初始简图也可通过纹理区别过程来聚集。这时所用的基本概念是对构成相邻区域的元素类型和密度作对比，当发现有差别时就认为存在不同的区域。

聚集运算遵循最小允诺原则（least commitment）：因为出错要付出很大代价，所以不要做任何没有把握的事情。因此聚集运算不是一次完成，而是经过多次迭代。通过邻域之间相互作用的迭代来达到对图象全局的协调的解释。这样的过程称为松弛过程（relaxation process）<sup>[Dav 81]</sup>

Marr 把聚集运算和纹理区别过程的最终结果称为完全的初始简图。这种表象完全和明确地表示区域的边界、物体的轮廓、以及基本的形状或结构。

Marr 理论的一个重要特点是完全初始简图的计算不需要知道某种物体是否出现在图象中。Marr 在这个阶段所用的方法是力图描述图象的属性，而不需要对景物中有什么物体作出假设。从这个意义上说是不需要知识的。图 2.15 中所示为完全初始简图的实例。图中的轮廓描述是在不知道图象内容的情况下求得的。

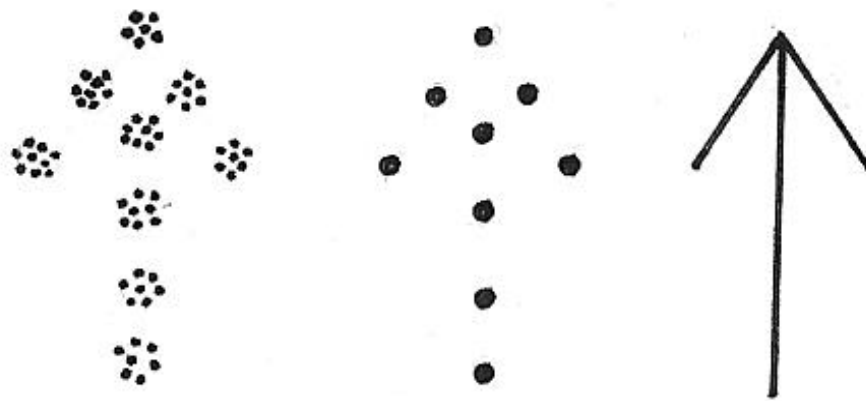


图 2.15 聚集过程

## 2.3 2.5 维简图

如何从包含在完全的初始简图 2.14 中的二维数据获得外部世界在以物体为中心的坐标系中的三维形状表象是接下来要研究的问题。研究表明如果试图从二维的图象数据直接得到三维表象，在对环境不作任何限制的通用条件将会遇到各种严重困难。以下通过分析对图象数据在作分割时遇到的矛盾来说明这个问题。

### 2.3.1 图象分割

神经生理学家和心理学家认为把图形和背景分开是视觉中的基本问题之一。反映在计算机视觉研究中这就是实现图象分割的过程。图象分割就是把图象或是分成对所研究的目的有

明确含义的区域，或是分成相应于实际物体或部分物体的区域。

尽管长期以来人们为研究这个问题作出了很大努力，但由于以下两个原因使分割的理论和实践都还是很原始的。首先，若要按图象甚至按实际世界来准确定义分割的目标简直是不可能的。例如，物体是什么？使它作为图象的一个区域与众不同的特征是什么？鼻子是物体吗？头是物体吗？如果头与物体连在一起是否还认为是一个物体？那么在马背上的人又怎么样呢？其次，图象的结构非常复杂，通常很难根据图象的亮度或某些类似于初始简图这样的数据按照相似性或其他纯视觉的线索用聚集的方法获得所希望的区域。因为，对物体识别来说重要的区域不一定具有任何可区分的视觉特征。物体的边界可能由于阴影，或其它物体的遮挡变得难以分清。这些因素使图象分割得到的区域不一定对应于景物中的物体。

那么分割这个概念有什么毛病呢？最明显的谬误似乎是：“物体”和“所希望的区域”几乎根本不是视觉的基本结构。因此，如果没有专门的附加知识无法根据初始简图或类似的早期表象重构它们。而这里研究的是通用的视觉系统，无法应用与专门领域有关的知识。所以，Marr 认为有必要从根本上避开这个传统的图象分割问题。在把景物分解成物体以前，先建立景物中可见表面的充分的描述。这就是 2.5 维简图的概念。现在的问题是如何建立景物可见表面的表象。

### 2.3.2 2.5 维简图

在人类的视觉系统中存在体视（双目立体视觉）、运动、影调、纹理等相互独立的低层视觉处理功能模块。这些模块能提供关于表面形状的多方面信息。其中最直接的是立体视觉和运动视觉。用图 2.16 中所示的随机点“立体图对”清楚地说明了体视信息处理的特点。图 2.16 中的右图与左图相同，除了中间有一小块正方形区域向左移动了一小段距离，人为地制造出一个视差。用单眼看时这两幅图没有差别，但在立体镜下左右图融合后，就能清晰地看到中间的正方形凸出图面。因此，体视能产生表面的深度信息。除了体视和运动以外，单幅图中的表面轮廓也能有效地提供形状信息。此外，还有一些次要的信息来源。表 2.1 所示为低层视觉过程中能从图象中抽取的各种有关几何结构变化的信息类型。

表 2.1

信息源	信息类型
立体视觉	视差，因而可得到 $d_g, \Delta_g$ 和 $s$
方向选择性	$\Delta g$
从运动恢复结构	$g, dg, \Delta g$ 和 $s$
光源	$g$ 和 $s$
遮挡轮廓	$\Delta g$
其它遮挡线索	$\Delta g$
表面方向轮廓	$\Delta s$
表面纹理	可能有

表面轮廓	$\Delta g$ 和 $s$
影 调	$d_s$ 和 $\Delta s$

在表中:  $g$  — 相对深度 (按垂直投影), 就是观察者到表面点的距离。

$dg$  —  $g$  的连续或小的变化

$\Delta g$  —  $g$  的不连续点

$s$  — 局部表面方向

$d_s$  —  $s$  的连续或小的变化

$\Delta s$  —  $s$  的不连续点。

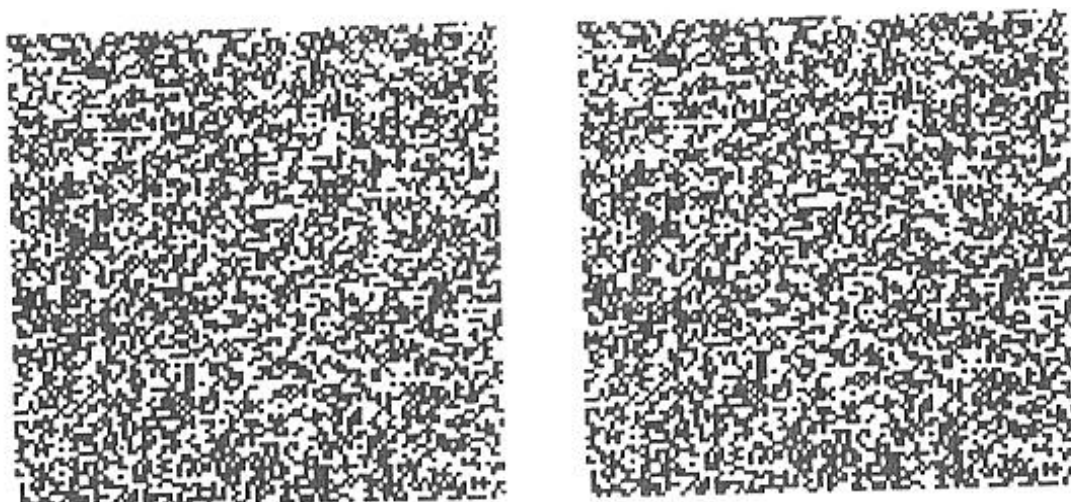


图 2.16 随机立体图对

为解决如何用合适的表象来表达与表面有关的信息, Marr 作了以下观察和分析。当我们观察一个复杂的表面, 例如一张被揉皱了的报纸时, 我们能轻易地说出表面上任何部分的方向和这一小块是否比其邻域更靠近观察者。尽管我们对表面方向的感知是很敏锐的, 但对表面形状的记忆却是较差的。此外, 如果表面中有的部分接近与视线平行, 那么用单眼观察这些部分时感到的表面方向不同于用双眼观察得到的结果。例如, 如果你通过一个很细的管子观察有花纹的天花板时, 会感到所看到的表面是与你的视线相垂直。尽管你的常识告诉你这是不对的。根据上述些观察, Marr 对表面表象的性质作出以下推测:

1. 景物中表面上的各点至少有一个深度的或表面方向的中间表象, 或同时有这两者的表象。

2. 因为对不熟悉的形状也能感知它的表面方向。所以, 得到表面表象是在把景物分解为物体以前。

3. 因为所检测到的表面单元的方向与是单眼或双眼观察有关, 所以表面表象可能完全由感知过程所支配, 而受关于表面实际是什么的专门知识的影响很小。

4. 此外，很可能不同的信息源可影响同一个表面表象。

总的来看，为了最有效地利用这些不同的，并经常是相互补充的信息源，需要把它们组合到能完全而明确地表示这些信息的表面表象中去。这个表象应不仅能完全而明确地、定量地表示深度、局部表面方向、以及它们的不连续点的信息，而且应建立和保持一个全局的深度描述。这个全局的深度描述与由上述信息所提供的局部描述相协调。Marr 把这样的表象称为 2.5 维简图。

如图 2.17 所示，在 2.5 维简图中表面方向是用二维梯度空间中的向量  $(p, q)$  来表示的，这相当于用覆盖在图象上的细针来表示。针的长度说明该点表面的倾斜程度，箭头则指向倾斜的方向。遮挡边界用线表示。表面方向的不连续边界用点线表示。深度可用标量表示，但在图 2.17 中未标出。这样的表象与观察者的视角有关。因此，它是在以观察者为中心的坐标系中关于景物三维特性的表象。

总之，2.5 维简图完全而明确地表达了初级视觉处理从图象中获得的各種信息。因此，可以认为初级视觉处理的主要目的就是构造这种表象。2.5 维简图只涉及图象中表面的各种特性，这就从根本上避免了定义“物体”、“背景”这些图象分割方法中固有的困难。

关于 2.5 维简图还有不少问题未被很好理解和解决（详见<sup>[Marr 82]</sup>）。此外，对于这个介于初始简图和三维模型之间的中间表象，其他学者也提出一些类似的设想。例如，Barrow 和 Temenbaum<sup>[Bar 78, 81]</sup>提出了本征图象（intrinsic image）的概念用以表示表面的照明、反射率、方向和距离等局部特性，他们建议的交互作用的本征图象集与 Marr 的 2.5 维简图在某些方面很相似。

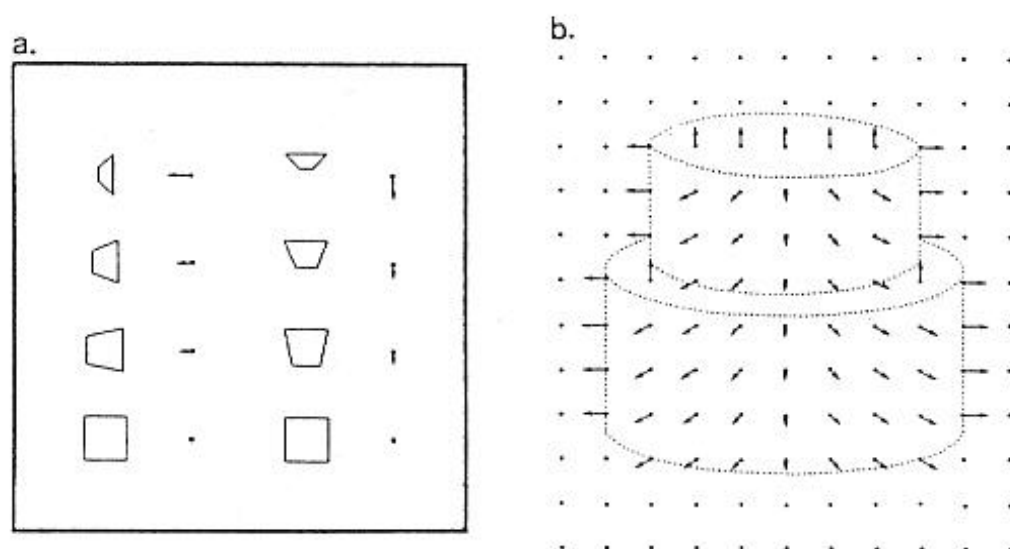


图 2.17 2.5 维简图（visual information, D.Marr P53. Fig. 12）

(a) 针的方向由表面法线在图象平面中的投影决定，长度由平面的倾斜程度决定；(b) 一幅典型的 2.5 维简图。

## 2.4 三维模型

Marr 视觉计算理论的最后部分是建立适用于视觉识别的三维形状表象。识别是指把以前看到过的物体辨认出来。识别形状的过程包括以下两部分：一是由图象建立三维表象；二是把它与人脑中储存的物体表象相匹配。匹配成功则完成了识别。2.5 维简图采用以观察者为中心的坐标系。因此，随观察者方位的不同而变化，并且没有表示形状的信息。因此，它不便于物体的识别。适合于识别的形状表象应该建立在以物体为中心的坐标系中，这样就不随观察的方位而改变，并且它还应该包含各种大小的形状基元，以便表示从小到大的各种物体的形状特性。

有两种以物体为中心的坐标系。一种是用单一坐标系来确定物体上所有点的位置。第二种是分布的坐标系。对物体上的每一个部分或单独的形状特性都有一个局部的坐标系。Marr 认为第二种更为合适。图 2.18 所示为由三维模型分级组合构成的人体三维形状表象。图中每一个方框相应于一个三维模型。框中左边是模型的轴，它可确定物体的位置、大小和方向。右边画的是模型各组成部分的轴的空间排列情况。例如，左侧第一个框中的左边所画的是人体的轴线，右边所画是躯干、头部、上肢、下肢、这些次级轴线的空间排列情况，其中每一个又指定了更次一级的坐标系。例如，上肢又进一步可分解为上臂、下臂、手掌、手指等。

接下来要解决的问题是如何确定任意形状物体的坐标系。这是一个困难的问题。很自然我们应该利用物体固有的突出的几何性质。例如，如果物体具有根据长度或对轴性可分辨出的自然的轴线，那么就应该利用这些轴线作为坐标系。对箱子形状的物体，则应该利用它的主轴，对人脸这样对称的形状，则应利用它的对称轴。

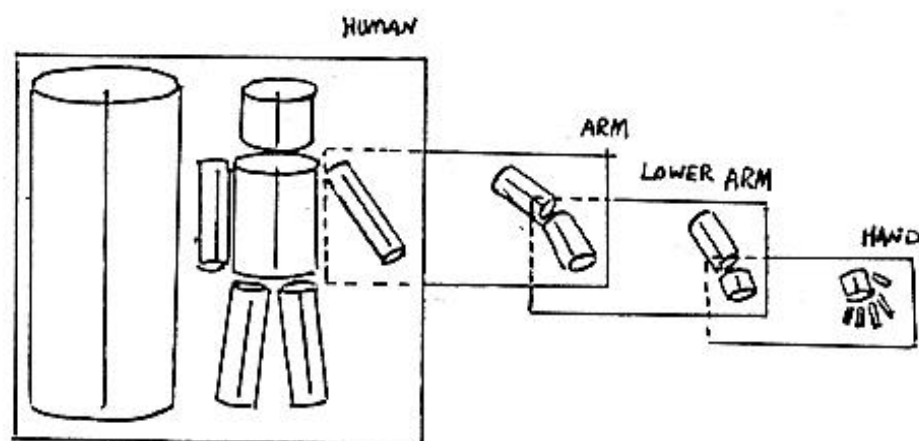


图 2.18 由三维模型分级组合构成的人体表象系统

一般来说，如果从物体的形状中能确定出轴线来，那么可以把它作为局部坐标系的基础。因此，解决确定以物体为中心的坐标系问题的一种方法是找出在其内部结构中包含轴线的那类形状。这类形状中的一种是广义锥（generalized cones）。广义锥是指某截面沿一给定的平

滑轴线移动而扫出的表面，在移动过程中截面大小可以平滑地变化，但其形状、以及与轴线之间的夹角保持不变。轴线不一定是直线（图 2.19）。许多物体，特别是那些自然生长的物体可以很方便地用一个或几个广义锥来描述。如图 2.20 中所示的动物图形，表示动物体每根棒条都是广义锥，它们分别代表动物的头、躯干、四肢、尾。许多人造的物体也可以用上述方法表示。例如，汽车是一个较大的盒子上叠加一个小盒子，建筑物是一个具有垂直轴线的大箱子等等。但是应该指出，有些表面，例如一团揉皱的报纸，是无法用广义锥来逼近的。这时就需要别的方法。然而这并不影响我们采用广义锥来作有益的理论探讨。

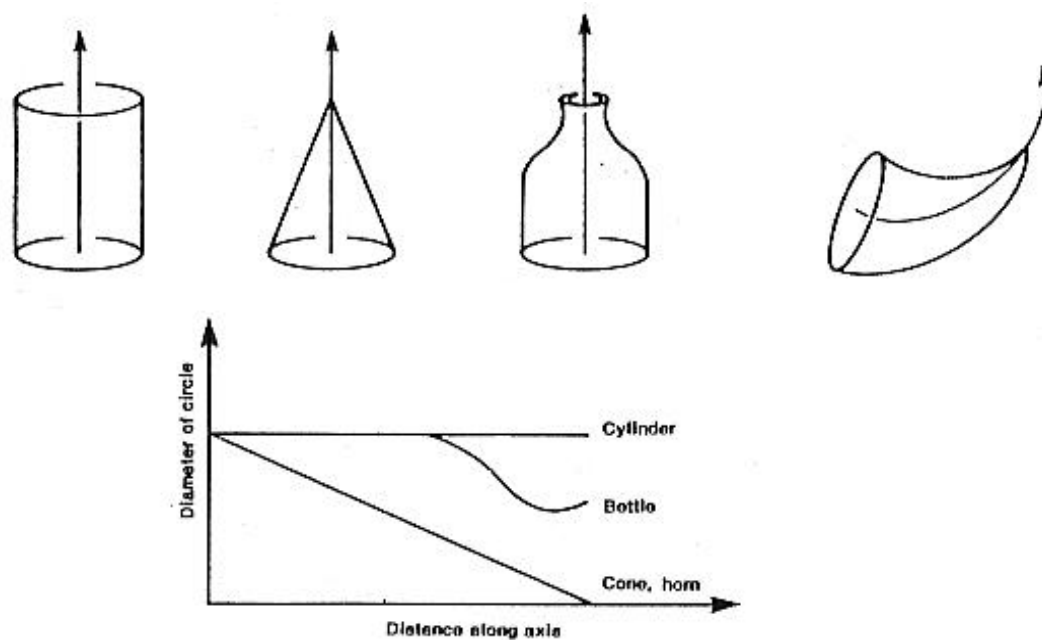


图 2.19 截面大小不固定及轴线不是直线时的广义锥



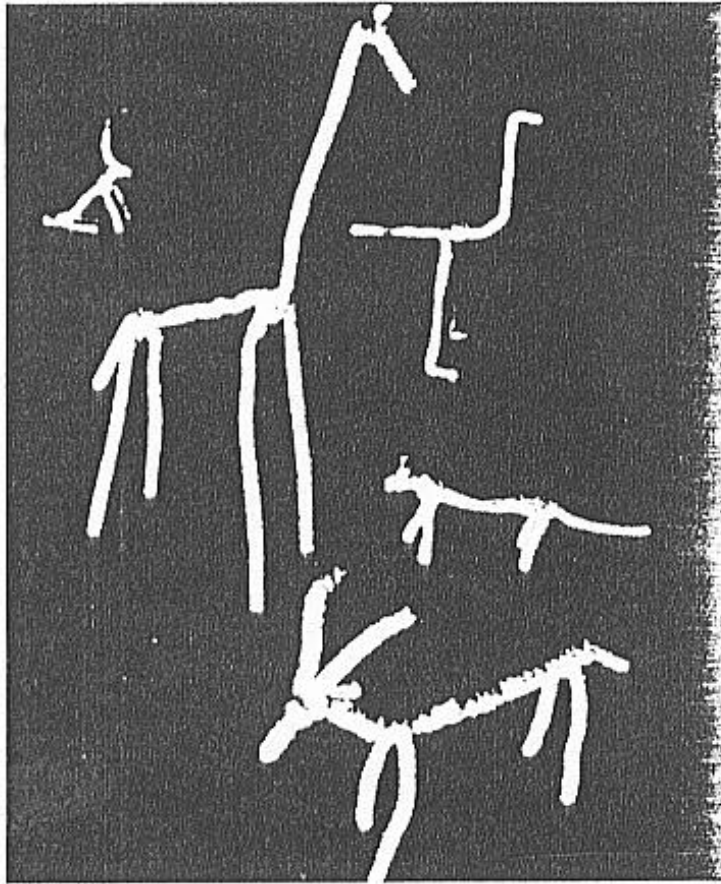


图 2.20 棒条构成的动物图形

## 2.5 视觉系统的组织

以上分别介绍了视觉系统中的各层表象。但目前还不清楚一种表象中的信息如何影响其它表象中信息的获取。一种可能是每一层的计算只取决于前面的相邻层中产生的描述。例如，按照这种自底向上的组织，2.5 维简图的计算只需要初始简图中的信息，而不需要直接从图象来的信息，也不利用任何关于从图象中可以看到什么东西的线索。在计算中所完成的每个过程都是独立或近乎独立地进行的。

另一种可能是，在初始简图和 2.5 维简图中的各种计算是靠混杂的约束传播来进行的。例如，从三维处理得到的信息可以用于改进根据灰度影调恢复表面方向的计算，反之亦然。总之，信息可以在所有方向上流动。

还有一种可能是信息自顶向下流动，对图象的理解在很大程度上取决于有控制的想象。这样的话，低层视觉就由关于应该看到什么的固定的预测所导引。目前还无法证明上述几种视觉系统的组织方式中哪一种是最适当的。

## 2.6 讨论

人和生物视觉的最显著的特点之一是选择性，也就是说观察者的注意力总是有目的地指向最感兴趣的东西。一般生物最注意环境中时常变化的东西而漠视固定不变的东西。这样就可以迅速认出什么是对自己有益的，什么是对自己有害的，从而作出攫取和躲避反应。此外，就象由格式塔心理学家发现的那样，人具有对图象数据进行组织归纳的能力，可以在多个层次上发现图象数据的规则性（regularity）、一致性（coherence）、连续性（Continuity）等整体特性。还有研究证明人类视觉具有在低层处理中抽取图形的拓扑特性的能力<sup>[陈 86]</sup>。Marr 的视觉计算理论完全不考虑视觉中的选择性和整体性（holistic），把低层视觉研究的目标确定为按照各种物理模型和附加约束条件，根据图象中各点的亮度或其它测量结果恢复景物中表面的各种有关特性，如表面方向、距离、反射率等。这些特性集中表示在 2.5 维简图中。Marr 强调在低层视觉信息处理中不需要关于图象中包含内容的先验知识，因此求解上述各项特性的方法完全是自底向上的，由数据驱动的方法。这时首先要应用成象光度学和几何学的定量模型来表示图象中某一点的测量值（亮度、颜色）与相应表面点性质之间的关系。但由于图象中每点的亮度受光照、表面材料的反射特性、景物中该点的表面方向等因素的综合影响，并且在成象过程中失去了各点的距离信息。所以，图象中某点的测量值与相应表面点性质之间的关系，从本质上来说，是不确定和多义性的。若要根据图象数据确定表面点的特性值必须增加附加的约束条件。例如，把物体限于刚体的范围，假设表面是连续的、各向同性的，或对表面性质作更为严格的约束，如表面是由平面构成的，点光源照明，表面材料的反射率是常数等。即使具备了这些条件，目前所采用的求解方法大多是类似于在求解经典边值问题（例如，确定在某一区域中的温度或电位分布问题）时所用的数值松弛方法。总的来说，这些方法都不太成熟。其中有些方法只要求作比较弱的，比较一般的假设，例如，假设各向同性，但它们的性能比较脆弱，容易出错；有些方法需要对应用场合作更严格的限制，例如，要求是光滑的理想漫反射表面，各点的反射率相同、点光源照明等，也就是适用于人造环境之中，而这样的条件在自然界，或在实际情况通常是不满足的。因此，不少学者<sup>[Kan 87]</sup>对 Marr 主张的这种不考虑视觉中的选择性和整体性，完全是自底向上，数据驱动以逐点计算的方式来处理视觉信息的可行性提出了异议。认为 Marr 视觉理论中求取本征图象或 2.5 维简图的步骤可能不是视觉信息处理中的踏脚石，而是一个走不通的陷井。

此外，Marr 的视觉信息处理方法遇到的另一个严重困难就是以逐点计算方式进行处理带来的难以承受的巨大计算量。相比之下，若按感知组织（perceptual organization）方法[Hoc 81]以整体特征作描述，则所需的计算要简捷得多。例如，在描述一个由空间点组成的圆周时只需要少数几个参数，如圆心位置、直径，以及圆周平面的方向就够了。如果不考虑这些点的整体特征，从逐点恢复各点坐标着手，计算量就会增加许多。而且计算完成以后也还不能马上显示出这些点在空间分布的整体情况。

为了克服上述困难，学者们对 Marr 的视觉计算理论进行了修正，主要的趋势有以下两种：一种趋势是着眼于利用和依靠格式塔视觉理论中的选择性和整体性，重新采用被 Marr 竭力反对的、面向分割的方法，提出了基于模型的视觉理论。这种理论认为视觉信息处理需

要辨别环境中重要的规律性 (regularities)，并把这些规律性作为感知过程中的踏脚石。因此，核心问题是找到一组通用的组件<sup>[Pen 86]</sup>。在构成这样的组件模型时要解决的关键问题是这些组件必须足够地复杂，以便被可靠地识别；同时又必须足够的简单，以便可被用作构成特定的物体模型时的构件。例如，Pentland 提出了应用不同参数的超二次锥面 (Superquadrics) 作为建模的基本元素。他希望通过把图象变换成这些宏结构组件的组合或模式，以此来简化描述和下一步的计算。从而提供一条通向图象理解的道路。由于利用了模型，所以在视觉理解中并不需要完全重构景物的三维信息，而是通过识别和推理来实现视觉理解。这就是下一章要介绍的内容。

第二种趋势是提出了视觉的连接主义模型 (Connectionist Models of Vision)<sup>[Fel 83, 85][Bal 86]</sup>。这种方法着眼于克服 Marr 理论中采用点状表象带来的复杂性，提出了具有高度并行性的神经网络。把以点阵形式表示的图象转换成整体特性的过程需要进行大量的计算。在连接主义模型中采用圆周、直线，以及其它几何形状作为模板来完成这样的转换。转换以后的整体特性用不变量 (常数，即模板) 和参数 (变量) 来表示。Hough 变换在视觉的连接主义模型中起关键作用，它提供了一种有效的途径把图象转换成一种基于某种内部模型 (模板) 的描述。并且 Hough 变换从本质上说就是可并行实现的。

此外，为了能在复杂环境下实现可靠、快速的视觉理解，吸取人类视觉中选择性的优点，不少研究者正积极从事主动视觉 (active vision) 的研究。主动视觉要使视觉系统与它支持的操作相结合。按主动视觉的方法，视觉系统可根据任务和外部输入的要求主动地控制摄像机的各种参数，如位置、方向、焦距、缩放、光圈等，以实现对视场中心的注意点、观察位置、分辨率、时间的选择性，以使得不需要对景物获取复杂的描述就可以完成所支持的任务。主动视觉中的研究方向可参阅[Swa 91]。

#### 参考文献:

- [Bab 83] Babaud, J., Wintkin, A. and Duda, R., Uniqueness of the Gaussian kernel for scale-space filtering, Fairchild Tech. Rep. 645, Flair 22, 1983.
- [Bar 78] Barrow, H. G. And Tenenbaum, J. M., Recovering scene characteristics from images, In A. R. Hanson and E. M. Riseman (Eds), Computer Vision Systems, New York, Academic Press, 3-26, 1978.
- [Bar 81] Barrow, H. G. and Tenenbaum, J. M., Computational Vision, Proceedings of IEEE, Vol. 69, no. 5, 1981.
- [Bin 81] Binford, T. O. Inferring surfaces from images, Special Issue on Vision, Artificial Intelligence 17, 205-244, 1981.
- [Bra 65] Bracewell, R., The Fourier transform and its applications, New York, MacGraw-Hill, 1965.
- [Cam 68] Campbell, F. W. C. and Robson, J., Application of Fourier analysis to the visibility of gratings, J. Physiol. (Lond) 197, 551-566, 1968.
- [Fel 83] Feldman, J. A. And Ballard, D. H., Computing with connections, In J. Beck, B. Hoope and A. Rosenfeld (Eds.), Human and Machine Vision, Academic Press, 107, 1983.

- [Fel 85] Feldman, J. A., Four frames suffice: A Provisional model of vision and space, *Behavioral and Brain Sciences* 8,1985,265-284.
- [Har84] Haraick, M., Digital edges from zero crossing of second directional derivative, *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-6, 1984, 58-68 .
- [Hoc 81] Hochberg, J., Levels of perceptual organization, In M. Kubovy and J. R. Pomerantz (Eds.), *Perceptual Organization*, 255,1981.
- [Hub,Wie 62] Hubel, D.H. and Wiesel, T. N., Receptive fields, Binocular interaction and functional architecture, In *Cat's visual cortex*, *J. Physiol. (Land)*, 166,106-154,1962.
- [Hub 68] Hubel, O. H. And Wiesel, T. N., Receptive fields and functional architecture of monkey striate cortex, *J. Physiol. (Land)* 195,215-243,1968.
- [Huf 77] Huffman, D. A., Relizable configurations of lines in itures of polyhedre, In E. W. Elcock and D. Mikhail (Eds), *Machne Intelligence* 8, Ediburgh University Press, 493-509,1977.
- [Kan 87] Kanal, L. And Tsao, T., Artificial Intelligence and natural perception, In *Intelligence Autonomous Systems*, 60-70, 1987.
- [Mar 80] Marr, D. And Hildreth, E., Theory of edge detection, *Proc. R. Soc. Land B* 207,187-217,1980.
- [Mar 82] Marr, D., *Vision*, W. H. Freeman and Company, 1982.
- [May 81] Mayhew, J.E.W. and Frisby, J. P., Psychophysical and computational studies towards a theory of human stereopsis, *Artificial Intelligence* 17,347-386,1981.
- [New 72] Newell,A. And Simon, H. A., *Human problem solving*, Englewood, Cliffs, N. J., Prentice-Hall,1972.
- [Pen 86] Pentlad, A. P., Part models, In *Proc. of Int. Conf. Pattern Recognition and Computer Vision*, Miami Beach, Florida, June 22-26,1986,242-249.
- [Rob 85] Roberts, L., Machine perception of three-dimensional solids, In J. Tippettl Eds. *Optical and Electro-optical information processing*, Cambridge, Mass, MIT Press, 159-197,1965.
- [Swa 91] Swain, M. J. and Stricker, M., Eds. *Promising directions in active vision*, The NSF Active Vision Workshop, University of Chicago, Aug. 5-7,1991.
- [Wal 72] Waltz, D., Understanding line drawings of scene with shadows, In *The Psychology of Computer Vision*, ed. By P. H. Winston, McGraw-hill Book Compny, N. Y. 1975.
- [Tez 82] Tezopoulos, D., Multilevel reconstruction of visual surfaces, *MIT Artif. Intell. Lab., A. I. Memo.* 671,1982.
- [Wit 83] Witkin, A. P., Scale-space filtering, *Proc. of The 8 th Conf. On AI* ,1019-1022,1983.
- [Yul 86] Yuile, A.L., and Poggio, T. A., Scaling theorems for zero crossings, *IEEE Trans. On Pattern Anal. Mach. Intel.* PAMI-8, no. 1,15-25,1986.