计算机视觉

复旦大学出版社

〔沪〕新登字 202 号

责任编辑　陆盛强
插　　图　林瑶华

计算机视觉
吴立德　著
复旦大学出版社出版
（上海国权路 579 号）
新华书店上海发行所发行　江苏省句容县排印厂印刷
开本 850×1168　1/32　印张 8　字数 230,000
1993 年 12 月第 1 版　1993 年 12 月第 1 次印刷
印数 1—3,000
ISBN 7-309-01377-8/T·110
定价：10.00 元(平装)
15.00 元(精装)

内 容 提 要

本书以 Marr 的视觉计算理论为框架，结合作者及其研究集体十余年的工作，系统地介绍了计算机视觉中的重要理论和算法。包括早期视觉中的基于滤波和马尔可夫随机场的图像分割和边缘提取方法、边缘的链码表示和分段线性表示以及边缘的树结构；中期视觉中的摄像机标定、立体视觉和运动视觉；后期视觉中的物体表示、模型库建立和三维物体的识别和定位等。

本书可供有关大学高年级学生、研究生以及研究工作者学习计算机视觉的教材或参考书。

前　言

本书是作者和他的研究生们10多年来在计算视觉与图像处理、模式识别方面所做工作的一个小结。已发表的论文见附录二，也包含了一些尚未发表的研究。

计算机视觉研究如何用计算机来部分地实现人类的视觉功能，是计算机科学和人工智能中一个十分活跃的领域。专门讨论这一领域的国际学术会议和学术期刊都在不断增长，研究人员和它的应用范围也都在不断扩大。

本书以Marr的视觉计算理论为框架，结合我们自己的工作，系统地介绍了计算机视觉中的重要理论和算法。原先这些内容都是分散地收录在许多期刊或学术会议的论文集中的，我们希望本书能对有兴趣于计算机视觉的研究人员和研究生、大学高年级同学有所裨益。

由于计算机视觉的内容十分的丰富，加上作者的学识以及本书的篇幅限制，遗漏与错误在所难免，敬请读者不吝批评指正。

我们的研究先后得到中国科学院自然科学基金、国家自然科学基金、高校博士点基金、863智能机器人主题、国家攀登计划"认知科学前沿领域中若干重大问题的研究"的资助(详见附录一)，也得到了已故清华大学常迥教授的热情鼓励和帮助，以及许多同行的支持和帮助，在此一并致以深深的谢意。

吴立德

目　录

第一章 概 论

计算机视觉是计算机科学和人工智能的一个重要分支。它的研究目的和内容有两个方面，一是如何用计算机实现部分人类视觉的功能；二是由此帮助理解人类视觉的机理。前一点侧重于功能上的相似，并不要求内部过程上也一致，有更强的应用性和工程性质；后一目的还要求其内部实现的方法和过程也一致，更带有基础性。

本章将扼要介绍计算机视觉的全貌，包括三个小节，第一节介绍人类视觉系统方面的一些发现，第二节介绍 Marr 关于视觉的计算理论的概貌，第三节介绍计算机视觉研究中发生困难的三个基本原因。

§1.1 人类视觉[1.1~1.4]

一、视觉系统

人类的视觉系统接受的是电磁波中的可见光部分，其波长大约为400纳米到760 纳米之间，不同的波长给人以不同色彩的感觉，而将所有可见光的波长按一定的比例混合起来则产生白色。

人类的视觉系统从眼球开始，入射光经过瞳孔和水晶体的调节和聚焦投射于眼底的视网膜上。视网膜中有大量的杆体细胞和锥体细胞，杆体细胞大致为锥体细胞的 18 倍，前者约为 1.2 亿个，后者约为650万个。它们是光感受器细胞，但杆体细胞只有在低照度水平下起作用，且不能感受颜色；而锥体细胞则在白天高照度水平下起作用，能感受颜色并比杆体细胞有较高的视觉敏感性。换言之，它们是相互补充的：杆体细胞在低照明水平下起作用，但不能像锥体细胞那样感受颜色和提供精细的物像；锥体细胞能辨认颜色和细节，但在低照明情况下不能起作用。

锥体细胞又可按其对不同颜色光线的感受而分成红色、绿色和蓝色三种类型，它们的组合产生各种颜色的感受。

杆体和锥体细胞在视网膜上并不是均匀分布的。视网膜的中心区域叫做视斑，视斑的中心叫做中央凹，其面积只有一平方毫米，但它是产生最清晰的视觉的地方。在这里全部都是锥体细胞，而且密度也最高。在视网膜外围，锥体细胞的密度很快地减少，而杆体细胞的数量则增加，直到视网膜的外围，杆体细胞占主导地位。

到达视网膜的光线经杆体和锥体细胞转换为神经信号，它们在传出视网膜之前，要经过视网膜中的神经节细胞的加工。

视网膜中大约有 100 万个神经节细胞，它们分为大、小两种类型，相互交错地分布着。大神经节细胞将若干光感受器细胞输出的信息相加，形成新的神经信息并送往其后的外侧膝状体，它们是"色盲"的。小神经节细胞与大神经节细胞不同，它们区别对待三种不同类型的锥体细胞，不做"加法"而是做"减法"，对颜色是敏感的。

经神经节细胞加工后的神经信号，经视交叉传到外侧膝状体，并进一步传到大脑皮层中的视觉区，如图 1.1.1 所示。

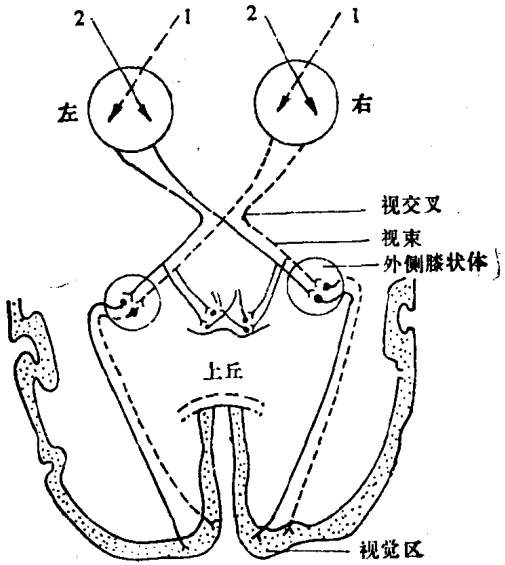

图 1.1.1 人类视觉系统

人类有左右两个外侧膝状体，位于大脑皮层下面的较深处，有花生米那么大小。每个外侧膝状体也由两种不同类型的神经元组成，它们的大小和处理的信息都不同。但跟神经节细胞不同的是，外侧膝状体中这两种不同的神经元并不是相互混杂交错的，而是在空间上分开的，形成不同的区域。外侧膝状体中小的神经元细胞（称为 parvo 细胞）部分接受来自小的神经节细胞传来的神经信号，而大的神经元细胞（称为 magno 细胞）部分则接受来自大的神经节细胞传来的神经信号。

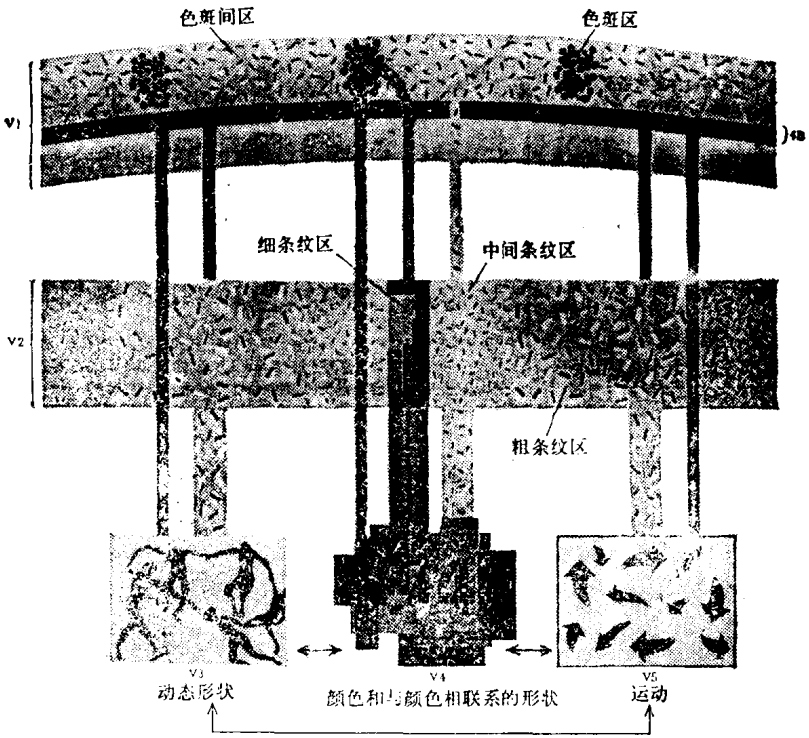

图 1.1.2 视觉区的结构特征和相互关系

上述两条通道，即 parvo 系统和 magno 系统，它们在对颜色敏感性、时间和空间分辨率方面都是有差异的。magno 系统比 parvo系统对

亮度的反差更敏感，有更快的响应时间，但对颜色不敏感，且空间分辨率较低。

外侧膝状体可分为 6 层，由底部到顶部，分别用 1,2,3,4，5，6 标出，其中第 2,3 和 5 层接受同侧眼底视网膜来的信号，而 1,4 和 6 层接受异侧眼底视网膜来的信号。

由外侧膝状体出来的信号被传入大脑皮层的视觉区。它们共由 5 个分区组成，分别称为 V1 区，V2 区，V3 区，V4 区和 V5 区，其中V5区也称为 MT 区。

这些区域的结构特征和相互关系，如图 1.1.2 所示。

总起来说，可以勾划出四个与不同视觉特征有关的并行系统；一个是针对颜色的，一个是针对运动的，两个是针对形状的。

对于运动来说，关键性的区域是 V5 区，或 MT 区，其输入信号的传递过程是从视网膜经 magno 系统到 V1 区的 4B 层，再以直接和通过 V2 区的粗条纹区两种方式传送到 V5区。

对于颜色来说，关键性的区域是 V4 区，其输入信号从视网膜经 parvo 系统到 V1 区的色斑区，然后以直接和通过 V2 区的细条纹区两种方式传送到 V4 区。

另外两个形状系统中，一个与颜色密切相关，而另一个则与颜色无关。

与颜色有关的形状系统是基于 V4 区的，其输入信号从视网膜经 parvo 系统到 V1 区的色斑区，然后经 V2 区的中条纹区两种方式传送到 V4 区。

与颜色无关的形状系统是基于 V3 区的，其输入信号从视网膜经 parvo 系统到 V1 区的 4B 层，然后以直接和通过 V2 区的粗条纹区两种方式传送到 V3 区。这一系统更多的是与动态形状，即物体运动时的形状有关。

视觉皮层内部的上述按功能的分工很自然地提出了这样一个问题：各专门化区是如何相互作用以形成统一的视觉感知的。最简单的办法也许是存在一个主区，由它对各专门化区来的信号进行合成，但解剖学已经否定了这样一个与各专门化区相连接的主区的存在。而实际

上，各专门化区都是直接或通过其它区彼此相连接的。

视觉信息的合成是在四个并行系统的各个层次上相互作用而形成的，已经发现 V3、V4、V5 不仅接受来自 V1 和 V2 的信号，而且也向 V1、V2 反馈信号，而且这种反馈的通道要比前馈的通道更为复杂和分散。例如 V5 只接受 V1 中 4B 层来的信号，而从 V5 反馈到 V1 中 4B 层的信号则很分散，其中包括对投射到 V3 去的那些细胞。正是由于这种反馈，使得不同专门化系统中的视觉信息可以得以合成而形成统一的视觉感知。可惜的是，对于这一方面，目前还知道得很少。

二、感受野

如前所述，光线到达视网膜后，由光感受器细胞(杆体细胞和锥体细胞)将它转变成神经信号，经过神经节细胞，外侧膝状体细胞等的汇集传送而到达大脑皮层的视觉区——视觉皮层。在这一过程中，一个光感受器细胞通过接受光并将它转变为输出神经信号而来影响许多神经节细胞、外侧膝状体细胞、以及视觉皮层中神经细胞。反过来，上述视觉通道上的每一个特定的神经细胞(可以是神经节细胞，外侧膝状体细胞，或视觉皮层中神经细胞)也直接或间接依赖于一批光感受器细胞(即杆体细胞和锥体细胞)。我们称直接或间接影响某一特定神经细胞的光感受器细胞全体为该特定神经元细胞的感受野。显然，在视觉通道上任一特定的神经细胞的输出仅依赖于它的感受野中的光感受器细胞所接受到的照射光，而与它的感受野之外的光感受器细胞无关。对于其感受野中的光感受器细胞而言，它们对此特定细胞的输出的影响也不一定都是相同的，有的起正的(兴奋性)作用，有的起负的(抑制性)作用，而且兴奋或抑制的大小也不尽相同。

对于视觉通道上的各种神经细胞的感受野，主要由于 D.H.Hubel 和 T.N.Wiesel 的大量工作而得到了深刻的认识。已发现，神经节细胞和侧膝状体细胞的感受野是位于视网膜不同位置处的、大小不同的同心圆型，其示意图如图1.1.3。

左边的图表示中心部分的光感受器细胞起兴奋作用，而其外围的起抑制作用，称为ON-型。右边的图正好相反，中心部分起抑制作用，称

为 OFF-型。对于具有 ON-型感受野的神经细胞而言，当在其中心部位受到光的照射，而在其外围部位不受到光的照射时，都有助于该细胞的兴奋。这类细胞实际上是对中心部分与外围的平均照明的反差敏感。

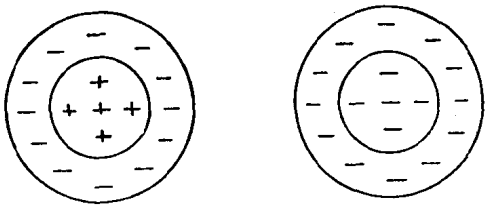

图 1.1.3 感受野

在视觉皮层中，发现有更复杂的感受野响应特性的神经元，它们对其感受野中的具有特定方向的线段和线段的运动敏感。

可以根据其感受野的响应特性对视觉皮层中的神经细胞进行分类：

一种称为简单细胞，它对其感受野中的特定位置处的具有特定方向的线段(有的是对暗背景中的明线段，有的是对明背景中的暗线段)敏感。

另一种称为复杂细胞，它也对其感受野中的具有特定方向的线段敏感，但它并不对位置敏感，并且不管线段位于其感受野中哪个位置处，只要具有特定的方向，它都敏感。因此这种复杂细胞对于保持方向的线段的移动也是敏感的。

还有超复杂细胞和极高度超复杂细胞，其中包括对感受野中出现角敏感的角检测器细胞。

所有视觉通道上的神经细胞，又可按其感受野只在一个眼的视网膜上还是同时在双眼的视网膜上，分为单眼的和双眼的。所有神经节细胞、外侧膝状体细胞和简单细胞都是单眼的，复杂细胞中约有半数为单眼，半数为双眼的。

双眼的细胞又可进一步分为右眼主导的、左眼主导的和双眼均衡的三种。

§1.2 Marr的视觉计算理论[1.5]

本世纪70年代中后期，原籍英国的D.Marr教授应邀在美国麻省理工学院的人工智能实验室创建并领导一个以博士生为主体的研究小组，从事视觉理论方面的研究，逐步形成了关于视觉的计算理论。这一理论从信息处理的角度系统地概括了心理物理学、神经生理学、临床神经病学等方面业已取得的成果，是视觉研究中迄今为止最为完善的理论，并逐步为大多数计算机视觉研究者所接受，而成为这一领域中的主导思想。它使计算机视觉的研究有了一个比较明确的体系，并大大地推动了计算机视觉研究的发展。事实上，在计算机视觉这一名词流行之前，有模式识别、图像(图片)处理、图像分析、图像理解等名词，其中图像分析、图像理解与计算机视觉的意义大致相当，正是由于Marr的计算理论促使了计算机视觉这一名词的流行。我们的工作也是在这一理论框架下进行的，将分别在后面的第二、三、四章进行介绍。

在这一理论框架下，10多年来在取得很大成功的同时，也陆续发现了一些问题。从而很自然地在计算机视觉领域中对这一领域的现状和起主导作用的Marr理论产生了不同意见，提出了许多新的理论框架，这些将在本书最后一章进行简要的介绍。

一、三个层次

这一理论把视觉过程看作一个信息处理的过程，并提出对于信息处理过程的研究应分为三个不同的层次，即计算理论的层次、表示(数据结构)与算法的层次、硬件实现的层次。这一理论还强调了当时并不受人重视的计算理论层次，并在这一层次，把视觉过程主要规定为从二维的图像信息中定量地恢复出图像所反映的场景中的三维物体的形状和空间位置。

按Marr的理论，计算理论层次要回答的问题是："计算的目的是什么？为什么这一计算是合适的？执行这一计算的策略是什么？"亦即视觉的计算理论要回答作为信息处理过程的视觉过程，它的输入是什么？

它的输出是什么？为什么由这个输入可以求得输出，或者说，输入和输出之间存在哪些内在的约束，使得我们可以由输入求得输出？等等。

而"表示(数据结构)和算法"层次要进一步回答如何实现这个计算理论？特别是输入、输出的表示(数据结构)是什么？为实现表示之间的变换应当采用什么算法？"

最后"硬件实现"层次要解决的是"在物理上如何实现这种表示和算法？"。

区分三个不同的层次，既看到它们之间的联系，又看到它们之间的区别和相对独立性，对于澄清许多问题是十分有益的。例如，人的视觉和计算机视觉在"硬件实现"的层次上是完全不同的，前者用的是由神经元组成的庞大神经网络，后者用的是由半导体器件组成的电子计算机。但如果两者在实现某一功能时，采用了相同的表示和算法，则它们在"计算理论"与"表示和算法"的层次上则可以是相同的。而如果它们实现了相同的功能，但所用的表示和算法不同，则它们仍然可以在"计算理论"的层次上相同。

由此可见，在"计算理论"和"表示与算法"的层次上，特别是在"计算理论"的层次上，人类视觉和计算机视觉是密切相关的，它们之中任一方面的进展都可以促进另一方面的进展。

二、三个阶段

人们生活于其间的世界是三维的，投射于人的视网膜上的图像则是二维的，视觉的功能则是要从感知到的二维图像中提取出有关的三维世界的信息。按 Marr 的理论，这"有关的三维世界的信息"主要地指二维图像所反映的场景中的三维物体的形状和空间位置的定量信息。

Marr 进一步将上述整个视觉过程所要完成的任务分成三个过程，如图 1.2.1 所示。

图像⟶要素图⟶2.5维图⟶三维表示

图 1.2.1 视觉过程中的三个阶段

视觉过程的第一阶段，由输入图像而获得要素图。视觉的这一阶

段也称为早期视觉。所谓要素图主要指图像中强度变化剧烈处的位置及其几何分布和组织结构，其中用到的基元包括"零交叉"、斑点、端点、边缘片断、有效线段、线段组、曲线组织、边界等。这一阶段的目的在于把原始二维图像中的重要信息更清楚地表示出来。我们将在第二章对此进行更深入的讨论。

视觉过程的第二阶段，由输入图像和要素图而获得2.5维图。视觉过程的这一阶段也称为中期视觉。所谓2.5维图指的是在以观察者为中心的坐标系中，可见表面的法向、大致的深度以及它们的不连续轮廓等，其中用到的基元包括可见表面上各点的法向、和各点离观察者的距离(深度)、深度上的不连续点、表面法向上的不连续点等等。由于2.5维图中包含了深度的信息，因而比二维要多，但还不是真正的三维表示，所以得名2.5维图。

视觉的这一阶段，按Marr的理论，是由一系列相对独立的处理模块组成的。这些处理模块包括：体现、运动、由表面明暗恢复形状、由表面轮廓线恢复形状、由表面纹理恢复形状等。我们将在第三章对此进行更深入的讨论。

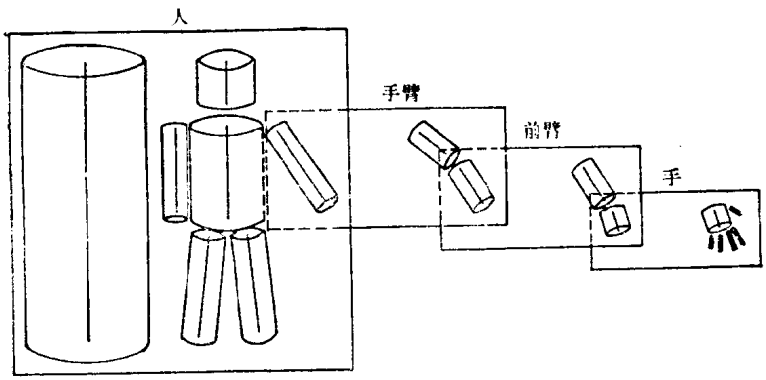

图 1.2.2　分层次的广义锥模型例子

视觉过程的第三阶段，由输入图像、要素图、2.5维图而获得物体的三维表示。视觉过程的这一阶段，也称为后期视觉。所谓物体的三维

表示指的是在以物体为中心的坐标系中，用含有体积基元和面积基元的模块化的分层次的表示，同时要给出各物体之间的空间关系的描述。Marr 建议采用的是分层次的广义锥模型。一个例子如图 1.2.2 所示。我们将在第四章对此进行深入的讨论。

§1.3 造成视觉问题困难的一些原因

如前所述，按 Marr 的理论，视觉过程可以看成是成像过程的逆过程，在成像过程中，有如下三个重要的变化。

(1) 三维的场景被投影为二维的图像，深度和不可见部分的信息被丢失了，因而也产生了同一物体在不同视角下的图像会有极大的不同，以及后面的物体被前面的物体遮挡而丢失信息等问题。

(2) 场景中的诸多因素，包括照明或光源的情况、场景中的物体的几何形状和物理性质(特别是表面的反射特性)、摄像机的特性、以及光源与物体和摄像机之间的空间关系等，都被综合成单一的图像中的像元的灰度值了。

(3) 成像过程或多或少地带入了一些畸变和噪声。

这样作为成像过程的逆过程的视觉过程，其任务是要从带畸变和噪声的、二维的、单一的灰度值中提取出尽可能不带畸变和噪声的、三维场景中的诸多因素(例如，物体的形状、物理特性以及空间位置等)的有关信息，这无疑是十分困难的。特别，当 Marr 理论还要能找到通用的、定量的精确解时，就更是如此了。

总之，由于成像过程中存在的投影、混合、畸变与噪声三个原因，使得作为成像过程的逆过程的视觉过程是不适定的和十分困难的。

参 考 文 献

[1.1] Bennett T.L., *The Sensory World: An Introduction to Sensation and Perception*, Wadsworth Publishing Company, 1978

中译本：旦明译，感觉世界：感觉和知觉导论，科学出版社，1985
[1.2] Hubel D.H., Wiesel T.N., *Brain Mechanisms of Vision*, Scientific American, 1979.9. 150-162
[1.3] Livingstone, M.S., *Art, Illusion and the Vision System*, Scientific American, Vol. 258:1(1988.1), 78-85
[1.4] Ramchandran, V., Anstis, S., *The Perception of Apparent Motion*, Scientific American, Vol. 254:6(1986.6), 80-86
[1.5] Marr, D., *Vision*, W.H. Freeman and Company, 1982
中译本：姚国正，刘磊，汪云九译，视觉计算理论，科学出版社，1988

第二章　早期视觉

如前所述，这是视觉处理过程的第一阶段，它要将输入图像中灰(强)度变化剧烈处的位置、分布和组织结构构划出来，形成要素。这一过程又可进一步分成几个子过程，一是找出图像中灰度剧烈变化的地方，称之为边素；二是将这些边素连接起来形成边缘、围线或纹理，并给出它们的适当的表示；随后根据围线，可对图像中的物体进行分割和平滑。以上的方法是基于边界的，当然也可以先对图像进行分割，形成对应于不同物体和背景的区域，而这些区域的边界即为图像中灰度变化剧烈的地方。本书主要介绍前一种方法，对后一种方法有兴趣的读者可参阅参考文献[2.1]的第4、5和9章。

早期视觉处理的好坏，直接影响以后的处理，因此是十分重要的。主要由于前面提到的"混合"和"噪声"的原因，这一处理也是十分困难，至今尚未找到十分有效的、通用的方法，也没有建立起很好的评价方法。

§2.1　边素(边过程)的检测

这里的输入是原始图像，连续时记为 $p(x,y)$，离散时记为 $p(i,j)$，$1\leqslant i\leqslant I, 1\leqslant j\leqslant J$，是一个两维数组，其中 $p(i,j)$ 表示像元(或像素)(i,j) 处的灰度值，见图2.1.1。

输出的表示有两种，其一是仍利用上述数组，当像素(i,j)处为边素时，令其为1，不是边素时，令其为0；另一种是在 Geman 兄弟[2.2]中引入的，他们称之为边过程：$h(i,j)$和$v(i,j)$，$0\leqslant i\leqslant I, 0\leqslant j\leqslant J$，其中，$h(i,j)$是像素$(i,j)$和像素$(i+1,j)$之间的水平边素，$v(i,j)$是像素$(i,j)$和$(i,j+1)$之间的垂直边素，当$h(i,j)=1$时，表示像素$(i,j)$和$(i+1,j)$