

计算机视觉基本理论和方法

计算机视觉基本理论和方法

吴健康 肖锦玉 编

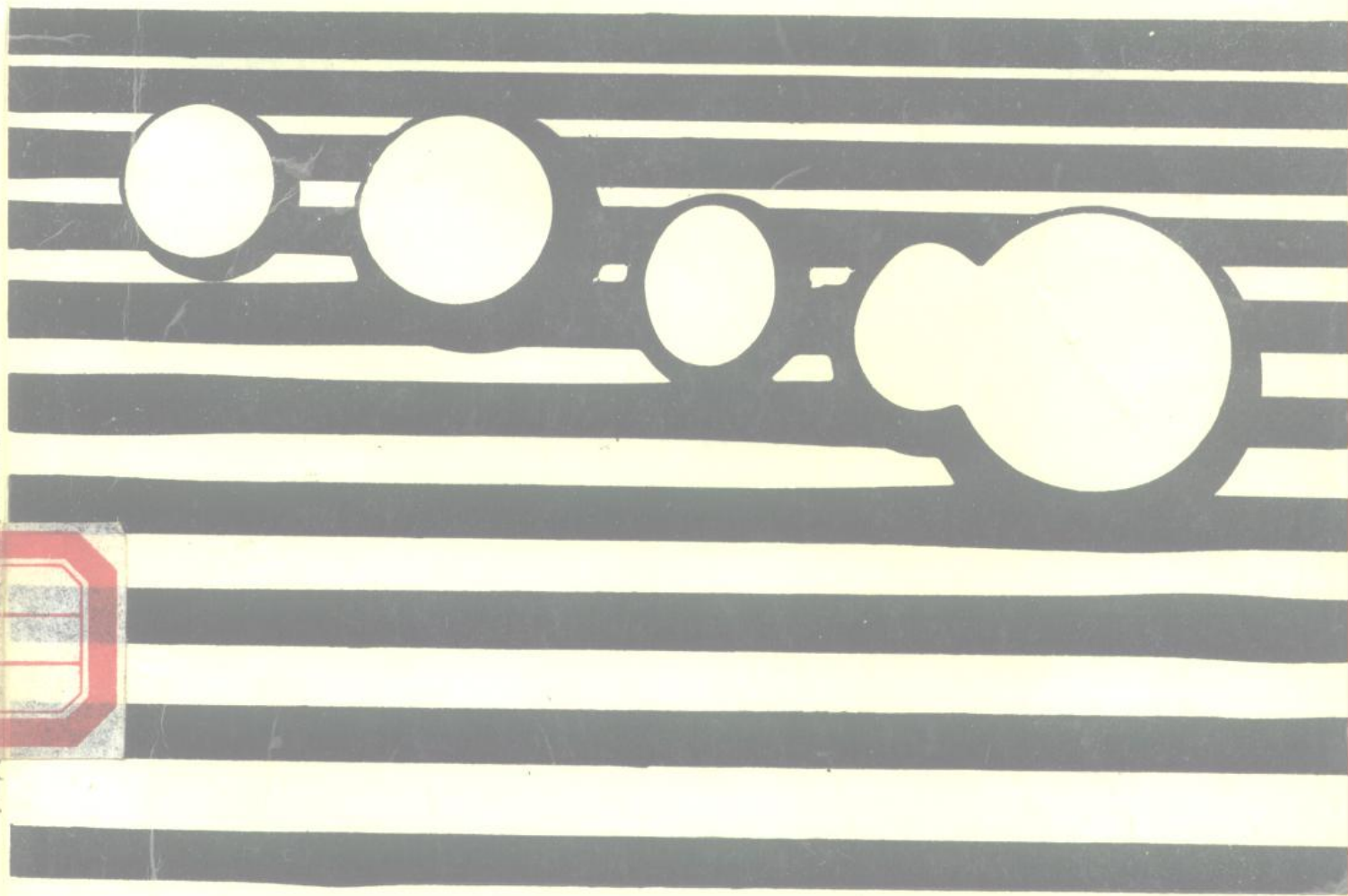
中国科学技术大学出版社

中国科

87

8

版社



计算机视觉基本理论和方法

吴健康 肖锦玉 编

中国科学技术大学出版社

1993·合肥

(皖)新登字08号

计算机视觉基本理论和方法

吴健康 肖锦玉 编

*

中国科学技术大学出版社出版

(安徽省合肥市金寨路96号, 230026)

合肥炮兵学院印刷厂印刷

安徽省新华书店发行

*

开本:787×1092/16 印张:12.75字数:310千

1993年4月第1版 1993年4月第1次印刷

印数: 1—3000册

ISBN7-312-00408-3/TP·47

定价: 9.50元

(凡购买中国科大版图书, 如有白页、缺页、倒页者, 由本社发行部负责调换)

内 容 简 介

计算机视觉是通过计算机技术对人类视觉信息处理的一种探索及模拟，当今计算机技术和模式识别的飞速发展，使计算机视觉的研究进入了一个崭新的阶段。本书论述了计算机视觉的基本原理和方法，并高度概述了近年来研究的最新成果及发展动向。全书共分七章，即：计算机视觉三十年；图像数据的统计模型；视觉中的近似理论和正则化理论；三维景物模型；利用单阶二维灰度图像进行三维物体的识别；使用网络的形状理论；计算机视觉的神经网络方法。

本书可作为理工科大学的，计算机视觉、图像处理、人工智能和计算机应用等专业的本科生和研究生的教学参考书，亦可供有关专业研究工作者参考。

D030 / 17

前 言

在过去的近三十年中，计算机视觉的研究取得了突飞猛进的发展。特别是80年代初，Marr的计算机视觉理论的提出，把计算机视觉的研究水平提高到一个新的层次。作为一个研究学科，必须有一个基本的理论框架，才能指导该学科沿着正常的方向发展。然而，计算机视觉似乎还不存在这样一个理论框架。当然，Marr的计算机视觉理论有其深刻和系统的一面，他提出了视觉信息表达的三个层次，从认知学、数学和实验方面完整地论述了原始草图的零交叉方法，论述了由体视恢复结构的方法。但是，我们不能不看到Marr理论的不足之处。首先，视觉信息是一种信息，如何从数学的方法严格地描述这种空间信息，Marr没有触及。其次，Marr较为充分地研究了早期视觉，至于视觉的高层次，即视觉知识的表达、使用、基于视觉知识的识别等等，Marr自己也承认，这一方面认知学尚不能给出任何解答，Marr自己也未论及。

试图建立计算机视觉理论框架的并非Marr一人，本书选择编写了一些有代表性的思想，并把它们分为七章。第一章是由编者之一写的综述，它涉及了计算机视觉研究的方方面面。第二章是论述图像信息的统计模型的。Belag关于方形网络上数据的统计学模型，把马尔可夫过程与Gibb分布联系起来，解决了该数学模型的实际应用方法，是迄今为止较为完善的空间数据模型。接着阐述了如何在这种数学模型指导下，使用局部图像运算，得到全局性的处理结果。这里，我们还叙述了Geman兄弟运用这一理论提出图像恢复的后验估计方法，它同时论证了模拟退火方案的收敛性，它被誉为在图像算法中运用“模拟退火”的经典论述。MIT的Poggio总结了早期视觉处理问题的共性，从数学上系统地论述了正则化理论。而且提出了相应的神经网络实现方法和学习方法，编写成第三章。第四、五、六这三章是有关三维物体识别方法的，它们分别系统地论述了三维物体模型、使用视觉组织方法直接由单一二维图像识别物体的方法、以及使用变换网络实现三维物体识别的方法。最后一章，则介绍了神经网络研究方面著名的三位学者的工作，即Grossberg提出了自己的一套计算机视觉理论，对Marr的理论提出了挑战。Fukushima和Neocognitron实现了变形手写字母的实时识别，而且具有Reinforcement等高级视觉功能。Kohonen的文章虽然不是直接论述视觉的，但他的空间自组织的概念对视觉研究很有价值。

本书前三章由吴健康编写，后四章由肖锦玉负责编写，全书最后由吴健康定稿。王立宏、徐小红、江黔夫、朱农等参加了部分工作。罗杰波、栾乐、钱传俊、余红等为本书文稿做了很多工作，作者对此表示感谢。

吴健康 肖锦玉

1991年6月24日

目 次

前言..... (i)

第一章 计算机视觉的三十年

1.1 引言	(1)
1.2 计算机视觉的基本概念	(1)
1.2.1 图像增强和恢复	(2)
1.2.2 图像数据编码和传输	(2)
1.2.3 图像分析和识别	(3)
1.2.4 图像代数.....	(12)
1.3 视觉计算理论和三维视觉.....	(16)
1.3.1 从立体视觉求深度.....	(18)
1.3.2 从运动求结构.....	(19)
1.3.3 从视觉流 (optical flow) 求结构.....	(21)
1.3.4 从阴影求形状.....	(23)
1.3.5 从纹理和表面轮廓求形状.....	(25)
1.3.6 由线条图求形状.....	(27)
1.3.7 由遮挡轮廓求形状.....	(28)
1.3.8 由空间编码求形状.....	(29)
1.3.9 结束语.....	(30)
1.4 视觉信息系统.....	(30)
1.4.1 认知学基础.....	(31)
1.4.2 基于模型的三维物体识别.....	(32)
1.4.3 空间信息的数据模式.....	(33)
1.4.4 空间信息系统的物理数据结构.....	(35)
1.4.5 使用框架系统的知识化界面.....	(35)
1.4.6 用于建立区域规划辅助决策系统的知识表达语言IDL.....	(36)
1.4.7 空间信息系统的可视查询语言.....	(36)
1.5 计算机图形学.....	(38)
1.5.1 三维物体模型.....	(38)
1.5.2 真实感景物的生成.....	(39)
1.5.3 光源模型和光照模型.....	(39)
1.5.4 光线跟踪算法产生三维真实感景物.....	(40)
1.5.5 分维几何.....	(42)
1.6 用于计算机视觉的并行体系结构.....	(44)

1.6.1 视觉计算机的结构要求	(45)
1.6.2 并行结构分类	(47)

第二章 图像数据的统计模型

2.1 晶格系统的空间作用与统计分析	(52)
2.2 用条件概率方法表达空间过程	(53)
2.3 马尔科夫场和Hammersley-clifford定理	(54)
2.4 与指数类型相关的一些空间组织	(56)
2.4.1 自模型	(56)
2.4.2 一些特殊的自模型	(57)
2.5 一些二维的空间组织及它们的应用	(59)
2.5.1 有限晶格组织	(59)
2.5.2 特殊的有限晶格组织	(60)
2.5.3 非晶格系统	(61)
2.5.4 联立自回归组织	(61)
2.5.5 在无限晶格上的自正态过程	(62)
2.6 晶格系统的统计分析	(63)
2.6.1 矩形晶格上的编码方法	(63)
2.6.2 矩形晶格上的单向近似	(64)
2.6.3 自正态系统的最大似然估计	(64)
2.7 随机松弛、Gibbs分布与贝叶斯最大后验图像恢复	(66)
2.7.1 退化图像模型	(67)
2.7.2 图和邻域	(68)
2.7.3 马尔可夫随机场和Gibbs 分布	(68)
2.7.4 GIBBS 抽样	(72)
2.7.5 图像恢复的例子	(73)
2.8 结束语	(74)

第三章 视觉中的近似理论和正则化理论

3.1 引 言	(75)
3.2 学习和近似网络	(76)
3.2.1 最佳近似特性	(77)
3.2.2 网络与近似方法	(78)
3.3 超表面重建学习	(79)
3.3.1 近似、正则化和归一化样条	(79)
3.3.2 用于学习的正则化技术	(79)
3.3.3 学习、贝叶斯理论和最小长度原理	(80)
3.3.4 由超表面重建到网络	(80)

3.4 正则化理论与正则化网络	(80)
3.4.1 正则化理论	(81)
3.4.2 正则化网络	(84)
3.4.3 网络与最佳近似	(85)
3.5 正则化方法的推广	(87)
3.5.1 移动中心: 对正则化解的近似	(87)
3.5.2 不同类型的根基函数与多尺度	(88)
3.5.3 加权范数与正则化	(89)
3.5.4 用不可靠样本学习	(89)
3.5.5 怎样学习中心的位置和范数权重	(90)
3.5.6 一个实际算法	(92)
3.6 高斯基函数与科学—假说神经生物	(92)
3.6.1 可分解根基函数	(92)
3.6.2 计算网络和生理学预言	(93)
3.7 一些应用	(94)
3.7.1 由其透视图识别三维物体	(94)
3.7.2 学习动力学系统	(94)
3.7.3 理解和发动机任务学习	(94)
3.8 结 论	(94)
3.8.1 正则化网络是怎样真正工作的	(94)
3.8.2 与其它方法的联系	(95)
3.8.3 网络与学习: 维数问题	(96)
3.8.4 总 结	(96)
3.9 计算视觉与正则化理论	(96)
附录A 定义与基本原理	(101)
附录B 高斯网络与Stone定理	(101)
附录C 正则化网络可任意地近似平滑函数	(103)
附录D 根基函数: 综 述	(104)

第四章 三维景物模型

4.1 景物结构模型	(107)
4.1.1 图像构成模型	(107)
4.1.2 特殊模型	(108)
4.1.3 部分和过程模型	(109)
4.2 自然形态的表达	(110)
4.2.1 生物形态	(112)
4.2.2 复杂的非生物形态	(112)
4.2.3 小 结	(114)

4.3 初级感知: 识别模型	(115)
4.3.1 识别超二次曲面模型基元	(116)
4.3.2 把视觉看作优化问题	(116)
4.3.3 学习物体模型	(119)
4.3.4 识别分形表面	(119)
4.3.5 分形尺寸的度量	(122)
4.3.6 超定束测试该模型的可用性	(123)
4.4 小结	(123)

第五章 利用单个二维灰度图像进行三维物体的识别

5.1 引言	(125)
5.2 深度重建在人类视觉中的地位	(126)
5.3 空间对应的解决	(127)
5.3.1 牛顿方法的运用	(128)
5.3.2 利用线—线对应关系	(131)
5.3.3 在匹配中使用参数确定	(132)
5.3.4 扩展匹配的初始集	(133)
5.4 感知组织	(133)
5.4.1 分组操作的导出	(134)
5.4.2 基于邻近的分组	(135)
5.4.3 基于平行的分组	(136)
5.4.4 基于共线的分组	(136)
5.4.5 分组操作的实现	(137)
5.4.6 连结点到直线段的分割	(137)
5.5 SCERP 视觉系统	(138)
5.5.1 模型匹配	(140)
5.6 进一步的研究方向	(141)
5.7 结论	(143)

第六章 使用变换网络的形状理论

6.1 引言	(144)
6.1.1 物体为中心的坐标系	(145)
6.1.2 从视觉输入计算以物体为中心的坐标系	(146)
6.1.3 用任务坐标系计算目标变换	(147)
6.1.4 理论基础和概述	(149)
6.2 背景知识参数网络分层体系	(149)
6.2.1 值单元原理	(149)
6.2.2 用值单元计算	(150)

6.2.3 连续值的离散过程	(152)
6.2.4 内插和最小化	(153)
6.2.5 内插值和Hough变换	(153)
6.2.6 实际设计	(154)
6.2.7 多面体限制	(155)
6.2.8 建立二维线段	(156)
6.2.9 建立三维线框模型	(157)
6.2.10 三维视网膜数据与原型的匹配	(159)
6.3 网络和高层控制	(159)
6.4 一个初步的试验	(160)
6.5 结 论	(161)

第七章 计算机视觉的神经网络方法

7.1 表面知觉的模型问题	(164)
7.1.1 表面理解理论解释的必要条件	(164)
7.1.2 定向反差检测器	(166)
7.1.3 OC 滤波器和短范围竞争阶段	(166)
7.1.4 大范围合作：边界完成	(167)
7.1.5 利用合作竞争反馈网络的边界选择：CC 环	(168)
7.1.6 从模拟反差到数字结构	(170)
7.1.7 利用可视反差的识别	(171)
7.2 神经识别机 (Neocognitron)：一个能进行视觉模式识别的分层神经网络	(172)
7.2.1 网络结构	(173)
7.2.2 网络的自组织	(175)
7.2.3 手写体数字的识别	(178)
7.2.4 讨 论	(184)
7.3 神经网络的自组织图	(185)
7.4 总 结	(190)

第一章 计算机视觉的三十年

1.1 引言

作为计算机科学的一个重要学科分支，计算机视觉在最近三十年有了突飞猛进的进展。其应用领域遍及工业、农业、科学研究、军事等，对这些领域的自动化水平和机动性的提高发挥了重要作用。可以预言，这种发展势头还将保持一个相当长的时期。原因很清楚，“百闻不如一见”，视觉信息在人类活动所涉及的各种信息中所占比重最大，由于它的独特的空间特性和结构特性，也不能为其它任何信息所代替。

国际上有时使用计算机视觉和图形、图像处理这一名称。通常，计算机视觉的研究范围较宽，包含了图像处理的绝大部分内容，但是一般不包括图像编码传输以及图像综合。图形学是专门研究图形的生成，而当今的图形学更多的涉及了真实景物的生成，这应该是图像综合的内容。虽然说图形是图像的一种，从概念上说图像综合包含了图形学的内容，然而图形学已经相当壮大，到了可以自立门户的地步了。人们也常使用“图像分析”这一名称，这是介于图像处理与计算机视觉之间的一个名词。另外，模式识别与机器智能、机器视觉等名词也屡见不鲜。坚持模式识别的学者争辩说，视觉信息是模式的一种，当然模式识别就可以包括计算机视觉了。事实上，在计算机视觉中确实使用了很多模式识别的概念和方法，而视觉信息有其特殊性，传统模式识别包含不了计算机视觉，视觉是人类智能的一种，计算机视觉当然可以说是一种机器智能。在美国 MIT 人工智能实验室主任 Winston 所著的“人工智能”一书，以及另一本著名的“人工智能手册”中，计算机视觉是其中重要组成部分。我们说，计算机视觉中使用了很多人工智能技术，由于计算机的重要性与日俱增，使得很多相近学科大有不包括计算机视觉就有不完整之感。

计算机视觉是一个边缘学科，这一研究领域集中了来自认知科学、通信、计算机科学以及各应用领域的科学家和工程技术人员，他们从各自不同的角度、以各自不同的背景、进行了卓有成效的研究工作。因此认识这一研究领域、接受现有研究成果，需要有比较宽的知识面。另一方面，视觉信息是有形的，在很多情况下，必须亲自在计算机上操作一遍，才能对所论方法有真正的认识。

这一章旨在将计算机视觉和图形、图像处理的历史和现状呈现给读者的同时，把主要篇幅放在叙述其中重要的研究方面以及进展和难点上，在下面的叙述中，为了方便起见，我们将计算机视觉和图形、图像处理简称为计算机视觉。

1.2 计算机视觉的基本概念

图像处理研究可以追溯到40年代的美英之间越过大西洋的图像传输。然而真正的计算机图像处理应从1964年美国喷气发动机实验室(JPL)用计算机处理了宇宙飞船拍摄的月球照片。用计算机增强的月球照片表面构造清晰可见，引起了很大的轰动。此外，美国西部的JPL W. Pratt领导的南加州大学图像处理研究所、美国东部的A. Rosenfeld领导的马里兰

大学的实验室相继开展了图像的增强和恢复、图像编码以及图像的边缘提取和分割等一系列初级图像处理技术的研究。这一系列研究工作在70年代已经取得重要进展，一些重要论著相继出现，如Rosenfeld和Pratt分别出版了数字图像处理专著，英文名为“Digital Picture Processing”和“Digital Image Processing”。

1.2.1 图像增强和恢复

图像的增强和恢复属于改善图像视觉质量，突出图像中重要细节的处理，它用于调整图像的对比度、去除噪声、变模糊为清晰。目前，图像增强处理在很多图像处理设备上已有专门硬件，可以实时地完成。

图像的直方图是一个重要的图像数据统计特性，它与对比度紧密相连，是图像灰度分布概率密度函数。大家知道，所谓数字图像，不同于通常我们所说的照片，是这些照片在计算机中的表示形式。一般说来，一幅二维的数字图像表示为一个矩阵，矩阵的列序号对应于图像的水平坐标，行序号对应于垂直坐标。显然，这两个坐标是离散化了的正整数。矩阵中的元素是位于相应坐标位置的图像灰度值，它们也是离散化了的整数，通常取值 $0, 1, 2, \dots, 255$ 。这一方面由于计算机中一个字节的数值表示范围的原因，另一方面，人眼也只能分辨32个左右的灰度级，在常规情况下，用一个字节表示灰度已经足够。

这样一个图像的直方图就是这幅图像中各灰度级的象素数。如果一幅图像的灰度分布集中在为数不多的几个灰度级上，则表明该图像对比度太低。这时可以用拉伸图像灰度的方法去校正它。

滤波是图像增强和恢复中的重要方法。例如图像增强中用非线性中值滤波器去除孤立噪声点，用高通滤波器突出图像细节。图像恢复在于去除成像过程中因成像设备和环境所造成的失真，如散焦失真、运动模糊、大气扰动所造成的卫星图像和天文照片失真等。图像恢复则是找出成像系统的传递函数，或称点展函数，使用逆滤波、卡尔曼滤波等方法恢复出原始图像来。实际上，由于成像系统的传递函数总不可能精确确定，恢复结果也往往因此而不理想。知道信号和噪声的统计特征也有助于图像恢复，甚至有人利用信号和噪声的统计特性，使用迭代方法比较理想地将图像从噪声中恢复出来。

1.2.2 图像数据编码和传输

数字图像的数据量是十分可观的，若将一帧电视图像数字化为一幅 $512 \times 512 \times 8$ 比特的图像，其数据量为256KByte。若以每秒25帧计，则传输数字图像的信道的速率应为每秒52.4288MBit。高信道速率意味着高投资，也意味着难以普及。图像编码的重要任务在于图像数据的压缩。图像数据的可压缩性在于图像数据存在空间相关性和序列（时间序列、谱序列）相关性。一幅人头图像，其背景的灰度值是同一的，或者相邻象素间的灰度值变化是有规则的、平稳的，这就是空间相关性。如果电视镜头很长时间一直对着一个场景，各帧电视图像之间，除了少数运动的人物之外，其它内容没有变化或变化较小，这就是时间序列相关性。相关性是与冗余度紧紧相关的。数据越相关，其冗余度也就越大，数据的压缩余地也就越大。

图像数据压缩的方法一般可分为三大类，一类是预测编码。它将图像数据的空间变化和序列变化规律用一个预测公式表示，当知道了某一象素的前面各相邻象素之后，该象素值可以用公式预测出来。由于预测公式对于收发双方都是已知的。图像数据的传输只需传输起始

值和预测误差。误差的动态范围比起图像数据的动态范围来要小得多，往往只要用1—2比特表示即可，因此，预测编码方法通常可以将每像素8比特压缩到每像素2比特左右。

变换是图像数据解相关的一种重要手段。人称最优变换的Kohonen-Loeve变换可以将一幅图像的“能量”压缩到为数不多的几十个分量上去。这样，只需要传输这为数不多的分量，在接收端再变换回去就可以达到数据压缩的目的。只不过在实际应用中，人们常用余弦变换代替K-L变换，变换也不对整个图像进行。是将整幅图像分成一个个小的（常取 16×16 ）数据块，然后将这些数据块分类、变换、量化，从而构成一个自适应变换压缩系统，图1-1所示是作者使用神经网络方法进行数据压缩的结果，其数据率仅用每像素0.167比特，压缩比达48:1。



(a) 原图像 $512 \times 512 \times 8\text{bit}$

(b) 压缩后的图像 0.167 bit/pixel

图 1-1 神经网络图像自适应编码

1.2.3 图像分析和识别

前面所述图像的增强、恢复和图像编码与传输，是属于从图像到图像的处理，其目的是为了改善图像质量或是变换图像表示形式和地点。它只能代表图像信息的某一方面应用。随着图像处理研究工作的深入，自70年代起，图像的分析 and 识别越来越多地受到重视。一幅图像中总是有一定主题的，分析和识别图像中的主题，进而指导下一步行动，这是人类活动中的主流。图像识别由简到繁可以分成三类。第一类问题最为简单，图像中的像素就表示某一物体的某种特性，例如，遥感图像中的某一像素就表示对于地面某一位置的地物的某一光谱波段的反射特性。通过其反射特性（即该图像的灰度）判别出该地物的种类就是遥感图像解译的目标。美国著名的普渡大学是这一研究领域的国际带头单位。第二类识别问题相对复杂。其中待识别物体是有形的整体，表现在二维图像中的除了物体的形状外尚有表面纹理等等。说它不是最复杂，是因为对于这类问题，二维图像信息已经足够识别物体。这类例子有研究很早的文字识别、染色体配对、矽肺病分类、某些工件识别等具有稳定视表面的三维物体识别。说这类识别问题比较复杂，因为它不容易表示成特征矢量，从而像遥感图像识别那

样借助于传统而有效的统计模式识别方法。首先必须将待识别物体正确地 从图像中的背景和其它物体中区别开来，然后设法将其形状特征和纹理特征转换成能在模式识别方法中表示的形式。也就是说，这类物体的识别必须分 以下几步完成：1) 借助于边缘提取或分割方法将待识别物体从图像中分离出来；2) 建立物体的形状和纹理的描述；3) 使用相应的识别或匹配方法完成识别或解释工作。

值得提出的是，著名的美籍华人，普渡大学教授傅京荪先生是这一方面世界公认的第一号领袖人物。他不但在传统模式识别理论方面造诣很深，而且是句法模式识别方法及其在图像识别上应用的创始人和倡导者。他所领导的实验室还在图像信息系统、图像处理机等多方面处于领导地位。

遥感图像分类 目前所指的遥感图像包括三种：航空摄影图像、气象卫星图像、资源卫星图像。航空图像处理 and 识别与普通图像一样，这里就不单独叙述。气象卫星图像也无特殊之处。另外还有海洋卫星，它装有合成孔径雷达，由于它工作于微波波段，且成像机理特殊，目前有很多用这种图像分析海浪、浅滩、重构三维表面的研究工作。美国资源卫星早期是多光谱 (MSS) 图像，现在则是精度更高 (25米×25米) 光谱波段更多的TM数据。这类图像数据被广泛应用于找矿、森林、农作物调查、估产、自然灾害测报、资源和生态环境监测等许多方面。日本东京大学高木干雄教授就领导着一个由几个大学和研究所，包括生物、地理、海洋等各学科教授组成的庞大研究队伍，从事利用卫星遥感进行生态环境监测和改造的研究项目。所有这些应用归结为使用多光谱图像分类方法，正确地判定所对应地面的地物。找矿则是在图像上找出有可能孕育矿藏的线性构造。所有这些应用的第一步都应该是将图像配准到地形图上去，以保证图像内容与实际地物对应关系的正确性。

多光谱资源卫星图像，包括MSS，TM和法国的spot卫星图像，是用多个光谱传感器同步地摄取地球表面的图像序列。当你拿到某一地区的TM数据的时候，这一组图像中的相同位置的象素值就表示地面相应点在七个光谱波段 (如红外、可见光、紫外等) 的反射特性。不同的地物，其各光谱波段的反射特性是不一样的。例如，绿色植物在绿色可见光波段有较强的反射，由于水表面温度受太阳照射会升高，因此在红外波段图像灰度值较高，等等。因此若对这一组七个光谱图像的每一点，将其相应七个波段图像的灰度值组成一个七维矢量，使用统计模式识别方法就可以对图像进行逐个像素分类，从而确定其所对应地物。进而可以统计某类地物的面积，估算作物产量等。

分类方法有无监督分类和有监督分类两种。所谓监督即用户首先向分类器提供训练样本，告诉分类器，什么样的地物具有什么样的光谱特征，以训练分类器正确的分类能力，训练完成之后，该分类器就可以用于分类了。现以常用的贝叶斯分类器为例，说明分类步骤。贝叶斯分类器的定义式为：

$$D_i(\mathbf{X}) = \ln(P(\omega_i)) - \frac{1}{2} \ln|C_i| - \frac{1}{2} \{(\mathbf{X} - \mathbf{m}_i) C_i^{-1} (\mathbf{X} - \mathbf{m}_i)\} \quad (1-1)$$

其中 C_i ， \mathbf{m}_i ， $P(\omega_i)$ 分别是第*i*类的协方差矩阵，均值矢量和该类出现的概率。 $D_i(\mathbf{X})$ 是具有特征矢量 \mathbf{X} 的点 X 属于第*i*类的判决函数。算出所有各类的判决函数之后，将 \mathbf{X} 分给判决函数最大的一类。分类步骤为：1) 将待研究区域的卫星图像与地形图配准；2) 根据地面

实况调查,选出各地物的典型地块,并把它标到数字地形图上,这样,由地形图与卫星图像的配准关系,可以得到各地物类型的光谱特征集合;3)由各地物类型光谱特征集合统计出均值矢量和协方差矩阵,并估计出在待研究地区各地物面积的比例,由此定义出各类出现的概率。这样分类器的各项参数都已确定;4)使用定义好的分类器,对卫星图像作逐像素分类,得分类图。与地形图迭合成图后,就可以得到最后可以直接使用的土地利用图。

无监督分类不需要训练数据,它使用分类程序根据光谱特征矢量的分布特性,并根据某种判据把像素分成一定数目的类,然后再将分类结果与地面实况核对,以确定所分类与实际地物的对应关系。

图像中物体的检出(边缘提取与图像分割)如前所述,为了分析和识别图像中的主体,有必要首先将图像中的主体从背景和其它物体中分离出来。一般说来,这有两种方法:边缘提取和图像分割。仔细观察各种图像之后,可以发现这样一个普遍存在的事实:物体与背景或一物体与另一物体之间的边界处往往存在图像灰度的变化,分析和利用这一灰度变化,检测出边缘来,形成完整的物体边界,达到将物体从图像中分离出来的目的。另一方面,图像中某一物体表面的灰度值具有某种同一性,例如灰度值基本均匀,纹理大体一样。根据同一性,可以将表示同一物体表面的区域检测出来。

图像分割和边缘检测的研究工作开始得很早,在这二十年左右的时间内,它一直受到重视,得到众多学者的青睐。然而,由于世间场景千姿百态、成像设备和条件各不一样、噪声的存在、照明的不均匀等等,很难找到一种普遍有效的方法,将图像中的物体检测出来。

(1) 边缘增强算子 边缘是图像中一个物体区域向另一个物体区域过渡的过渡带,常表现为灰度的阶跃和斜变,在存在噪声或图像失真的情况下,这种灰度变化趋向于缓慢和起伏。检测阶跃和斜变的一种显而易见的方法是计算灰度的梯度,找出梯度最大值及其方向,这就产生了离散情况下的拉普拉斯算子:

$$\begin{array}{ccc} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{array}$$

运算时只需将这一 3×3 窗口算子放在图像上自左向右自上向下逐像素作扫描形式的移动,在每一位置,将窗口中系数与所对应像素相乘后求和,求得中心位置像素点的梯度值。由此还派生出不少非线性窗口算子,如Robert算子、Prewitt算子、Sobel算子等。其中四个方向的Sobel算子如下:

$$\begin{array}{cccccccccc} 1 & 0 & -1 & 2 & 1 & 0 & 1 & 2 & 1 & 0 & 1 & 2 \\ 2 & 0 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 & -1 & -2 & -1 & -2 & -1 & -2 & -1 & 0 \end{array}$$

另外四个方向与此类推,算出八个方向后,取其中的最大值和方向,作为灰度的梯度和方向,图1-2(b)是一图像经过Sobel算子边缘增强后的图像。

(2) 较高级的边缘检测方法。随着图像分析应用的推广,要求边缘检测方法不仅能够噪声和模糊等情况下具有良好的性能,而且要求其检测出的边缘位置具有一定的精度。这些年来,各种新方法不断出现,它们对相应一类的图像显示了明显的效果,总的说来,这些

方法可分成以下几类：1) 滤波方法，典型的有：Marr的零交叉边缘检测方法，用形如墨西哥草帽的圆对称滤波算子 $\nabla^2 G$ 进行图像滤波，然后检测零交叉作为边缘。2) 建立边缘的灰度变化模型，然后使用最优拟合的方法拟合实际图像，从而求得其灰度梯度、方向和可信度，这类典型方法有Haralick的斜面模型(facet model)。3) 利用边缘必须连续、边缘灰度突变等先验知识，使用迭代方法检测边缘。边缘增强和检测得出的往往不是物体的完整边界，若原图像边缘灰度过滤太慢，则会出现边界过厚的现象，这时应确定最佳边界线，去除其它多余边界点。如果原图噪声太大，也会出现边界中断现象，这时应使用联接方法补齐这些边界点。我们提出了一种能一次抽取完整的、厚度为一像素的边界提取的神经网络方法，下面就简要地叙述该方法的思路和结果。

所有图像处理方法是以前述的图像数据的数学模型为基础的。我们认为边界有它们自己的特性，应该为设计更好的边缘检测和边界抽取算法提供了具有普遍适用意义的独立的模型，但是，据作者所知，除了Cooper将边界模型化为 k 阶Markov过程之外，在边界的模型方面并未进行多少工作。在Cooper的模型中，Markov过程的状态对应着 k 个边界元素的序列。这里我们提出一个相容-相斥(CC)边界模型：

二维随机场 设 $S=(S_1, S_2, \dots, S_N)$ 为一点集，将边界数据模型化为随机场 $X=(X_s, s \in S, X_s \in \Lambda)$ 。状态空间 Λ 既可取二值 $\Lambda=(0, 1)$ ，用以表征边界点的有无；也可连续取值 $\Lambda=[0, 1]$ ，用以表征边界的强度或确实为边界的概率，真实的边界数据被看作该随机场 $(X_{s_1}, X_{s_2}, \dots, X_{s_N})$ 的实现，这里 X_{s_i} 是点 S_i 的一种状态。

Markov随机场 假设 $NB=(NB_s, s \in S)$ 为 s 的一邻域， NB_s 为 s 的邻点集合，边界数据的局部特征是通过将 X 看作一Markov随机场来定义的，从而有

$$P(X_s = x_s | X_t = x_t, \text{ for all } t \in S \text{ and } t \neq s) = p(X_s | X_t, \text{ for all } t \in NB_s)$$

相容-相斥邻域 一边界元素的邻域 NB_s 可以进一步划分为两类子区域：相容区域 CS_s 和相斥区域 CD_s 。相容区域支持该边界元素的存在，而相斥区域反对它的存在，支持与反对之总和依赖于这些区域的总体结构。假定这两类区域彼此线性无关，于是有如下的Markov场和Gibbs分布形式：

$$P(X_s = x_s | X_t = x_t, \text{ for all } t \in NB_s) = \exp [f_s(X_s, X_{CS_s}, X_{CD_s})] \quad (1-2)$$

$$f_s(\cdot) = \alpha(X_s) + \Phi(X_s, CS_s) + \Psi(X_s, CD_s) \quad (1-3)$$

这里 X_s 表示点 s 的可能状况， $\alpha(\cdot)$ ， $\Phi(\cdot)$ ， $\Psi(\cdot)$ 为势能函数。由于边界具有很强的结构性，在使用Gibbs分布时，我们倾向于不使用邻域上的基团的概念。现在来看一个4-邻域的简单情况，在验证当前像素点是否存在一个水平边时，如果只考虑该像素点右边的一个点，就无法确定是存在一水平边。但是，如果该点左边和右边的像素点均是边界元素，那么该水平边就很可能存在。如果只有一边的像素点是边元素，就无法确定该点是否是边；如果两边都不是边元素，那么该边元素就可能不存在或是噪声点。另一方面，该像素点之上和之下的像素点属于相斥区域，我们假定边界只有一个像素点宽。一个水平边元素上下两个相邻位置都为水平边的事实将支持该水平边元素的存在。但是，若只有其中一个位置存在水平边元素，另一个为非边元素，这一事实将反对该水平边元素的存在，以保持一像素的边界厚度。

我们关于边缘检测和边界抽取的神经元模型与我们的数学模型相一致，在该模型中，神

神经元按照边缘网格系统排列，列中简单视皮层神经元和行中简单视皮层神经元分别具有垂直方向和水平方向的优先，相邻神经元的感受野沿轴线方向被分为平行的两个区域，分别对正和负刺激作出反应，每个受激区域内刺激的总和构成了细胞的刺激输入。

为形成结构合理的、有意义的边界，相邻神经元之间存在局部的横向相互联系。在哺乳动物大脑内，这种联系可以随着动物的成长或从环境中的学习而建立起来。神经元之间存在两种横向联系：支持当前神经元的兴奋状态的来自相容邻域的激励联系，以及来自相斥邻域的抑制联系，激励与抑制的总和是通过相容区域和相斥区域的构造性模板匹配结果来确定的，这些构造性模板反映了边界的判别标准，它们是由人或动物从经验中学习得来的。例如，邻近的平行边被抑制，连续线被激励，允许线的交叉存在。

利用竞争横向相互作用进行边界抽取 边缘网格系统中的单元具有固定的方向，或为垂直，或为水平。对于一水平边单元，左边的三个相邻单元和右边的三个相邻单元组成其相容邻域，而该单元上下的两个相邻单元属于相斥邻域，链是建立在一个单元和其邻域之间的，这样该单元就可以从相容邻域和相斥邻域接受激励信号与抑制信号。抑制信号的计算为

$$S_{i \text{ inh}} = \max_j (S_j^{\text{inh}}) \quad (1-4)$$

这里只从两个抑制相邻单元中选择一个最大的作为抑制信号。让我们考虑一水平边界轮廓上的三个水平单元，其中中间的一个单元具有最大边缘强度。若我们将两个抑制信号相加，中间边单元的抑制信号就可能最大，这当然是我们所不希望的。若只选择其中最大的一个抑制信号，就可能使中间单元在竞争中优先。

相容邻域由两个子域组成：对于水平边是左邻域和右邻域，对于垂直边是上邻域和下邻域。若我们将这两个子域内单元的输出标记为 O_j^{exc1} 和 O_j^{exc2} 则激励信号的计算式：

$$S_{\text{exc}} = \min \{ \max_j (O_j^{\text{exc1}}), \max_j (O_j^{\text{exc2}}) \} \quad (1-5)$$

对每个子域，先从子域中选择一个最大的输出，然后再从两个子域的最大输出中选择一个最小的输出作为激励信号。该公式强调边界的连续，如果当前边单元的两侧都有边元素，则当前边界单元会得到强激励。如果只有一侧边元素，则它很可能是一断点，这样我们就太不清楚当前单元是否是边界元素。

对于两侧子域都有边元素的情况，我们主要讨论其中一个子域内只有一边元素的情况，这对应于连续线或交叉线。下面的公式考虑了这种情况及连接边界缝隙的作用。

$$t_1 = \max_j (O_j^{\text{exc1}}) / \sum_j (O_j^{\text{exc1}}) \quad (1-6)$$

$$t_2 = \max_j (O_j^{\text{exc2}}) / \sum_j (O_j^{\text{exc2}}) \quad (1-7)$$

其中 C 是大于1的常数。

将一边界单元的激励信号和抑制信号分别标记为 S_{exc} 和 S_{inh} 。对于单元 i ，我们将相容区域的两个最大输出加起来作为激励信号，而从相斥区域选择一最大输出作为抑制信号。