

清华大学电子与信息技术系列教材



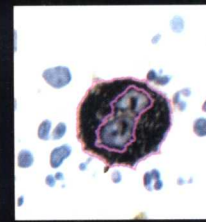
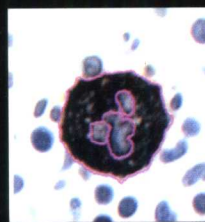
图象工程 下册

图象理解 与计算机视觉

章毓晋 编著

清华大学出版社

<http://www.tup.tsinghua.edu.cn>



清华大学电子与信息技术系列教材

图象工程（下册）——

图象理解与计算机视觉

章毓晋 编著

清华大学出版社

(京)新登字158号

内 容 提 要

本书主要介绍图象理解和计算机视觉的基本概念、研究内容和实用技术,以及近年来国际上有关的最新研究成果和应用实例。

本书包括五部分。第一部分(第1, 2章)概述了一些有关图象和视觉的基础知识;第二部分(第3, 4章)是对一些图象采集、处理和分析技术的推广;第三部分(第5, 6章)讨论有关立体匹配和场景恢复的技术;第四部分(第7, 8, 9章)介绍有关对场景进行解释所涉及的知识表达、匹配理解等内容;第五部分(附录A和附录B)介绍了一些相关技术,包括统计和结构模式识别、人工神经网络、数学形态学等,以及有关的参考文献目录。书中还提供了大量例题与习题。

本书可作为信息和信号处理、通信与电子系统、模式识别、机器人视觉、生物医学工程等学科专业课教材,也可供上述学科及信息工程、电子工程、计算机科学与技术、机器人自动化、遥感和军事侦察等领域的科技工作者和高等院校的师生参考。

图书在版编目(CIP)数据

图象工程 下册, 图象理解与计算机视觉/章毓晋编著. —北京: 清华大学出版社, 2000
清华大学电子与信息技术系列教材

ISBN 7-302-03980-1

I.图… II.章… III.① 图象处理-计算机应用-高等学校-教材 ② 计算机视觉-高等学校-教材
IV.TP391.41

中国版本图书馆CIP数据核字(2000)第 35833 号

出版者: 清华大学出版社(北京清华大学学研大厦, 邮编 100084)

[http:// www. tup. tsinghua. edu. cn](http://www.tup.tsinghua.edu.cn)

印刷者: 世界知识印刷厂

发行者: 新华书店总店北京发行所

开 本: 787×1092 1/16 印张: 17.75 字数: 425 千字

版 次: 2000 年 8 月第 1 版 2000 年 10 月第 2 次印刷

书 号: ISBN 7-302-03980-1/TN·112

印 数: 4001~9000

定 价: 20.00 元

前 言

图象是用各种观测系统以不同形式和手段观测客观世界而获得的，可以直接或间接作用于人眼并进而产生视知觉的实体。科学研究和统计表明，人类从外界获得的信息约有75%是从图象中获得的。随着计算机的发展，数字图象技术近年来得到极大的重视和长足的进展，出现了许多有关的新理论、新方法、新算法、新手段和新设备，并已在科学研究、工业生产、医疗卫生、教育、娱乐、管理和通信等方面得到了广泛的应用，对推动社会发展，改善人们生活水平都起到了重要的作用。

为了对各种图象技术进行综合研究、集成应用，我们有必要建立一个整体框架——图象工程。图象工程是一门系统地研究各种图象理论、技术和应用的新的交叉学科。它的研究方法数学、物理学、生理学、心理学、电子学、计算机科学等学科相互借鉴；它的研究范围与模式识别、计算机视觉、计算机图形学等专业互相交叉；它的研究进展与人工智能、神经网络、遗传算法、模糊逻辑等理论和技术密切相关；它的发展应用与生物学、遥感、通信、文档处理等许多领域紧密结合。

图象工程是一门综合学科。考虑到其内容多、跨度大、覆盖面广以及教学的特点，本书分为相对独立又互相联系的上下两册：《图象工程（上册）—— 图象处理和分析》和《图象工程（下册）—— 图象理解与计算机视觉》。

本书在编写的方针上，力求具有理论性、实用性、系统性、实时性。考虑到图象技术的飞速发展，本书在内容上，既认真选取了有代表性的经典内容，又注意选取了20世纪90年代国际上的最新研究成果。本书既参考了许多有关文献，也结合了作者的一些研究工作和成果以及教案。本书主要介绍图象工程的基本内容，但又有一定的深度和广度，希望读者能通过本书的学习，独立地和全面地了解该领域的基本理论、技术、应用和发展。本书在内容叙述上，力求理论概念严谨，论证简明扼要。考虑到图象技术的应用特点，书中列举了大量实例使原理和概念具体化，引导读者把这些概念和原理应用于实际。

考虑到专业课教学的特点，本书从方便读者学习的角度出发做了两项工作：

第一，在主要内容的编排上充分注意了理论的系统性和内容的结构化。全书将不同的图象技术分支均纳入图象工程大框架进行介绍，而且正文都统一按照技术分章节，互相关系清晰明确。除每册书将第1章作为该书总纲外，之后每章（包括附录）又将其第1节作为该章总纲，并以结构图表示所属内容及分类。希望这些能使读者在学习时既有一个总体概念，又可时刻把握当前学习内容在整个结构中的位置。

第二，增加了两方面的内容。其一是为方便读者进一步查阅（尤其是联网查阅）英文原文资料增加了英文目录，这也相当于提供了英文关键词；其二是本书不仅列举了书中直接引用的具体文献，还增加了一些重要专著、期刊和国际会议的名称。这不仅对阅读本书有益，为进一步深入学习提供了线索，而且对将来从事科研工作也会有帮助。

本册书为下册，主要介绍了图象理解与计算机视觉的一些基本概念、基本原理、典型方法、实用技术以及国际上有关研究的若干新成果。读者从中可了解和掌握常用的图象理解与计算机视觉技术手段，并能据此解决实际图象应用中的具体问题。与上册配合，能使读者建立一个比较完整的图象理论体系。

本册书主要包括五部分。第一部分（包含第 1, 2 章）是图象和视觉基础，论述了图象技术整体概况和分类、图象工程高层技术的特点以及有关视感觉和视知觉模型等；第二部分（包含第 3, 4 章）介绍了 3-D 世界立体成象和一些扩展图象分析技术；第三部分（包含第 5, 6 章）介绍了场景重建的基本原理和技术，包括双目和多目立体匹配，借助阴影、纹理、运动等线索的 3-D 目标形状恢复；第四部分（包含第 7, 8, 9 章）介绍了场景理解的基本内容，包括知识和表达、匹配和理解、以及视觉信息系统等；第五部分（包含附录 A 和附录 B）介绍相关技术和补充材料。

本册书共有 9 章正文和 2 个附录，在 11 个一级标题下有 59 个二级标题（节），再之下还有 103 个三级标题（小节），从内容上大体覆盖了图象理解和计算机视觉的各个重要分支和一些相关的基本技术。全册书共有文字（也包括图象、绘图、表格、公式等）约 43 万字。本册书共有编了号的图 227 个、表格 28 个、公式 562 个。为便于教学，本册书共给出各类例题 100 个。为便于检查教学效果，各章后均有练习题，本册书共有练习题 142 个。另外，书末列出了提供参考和直接引用的 100 多篇文献的目录。

本册书的编写始于 1993 年，1994 年开始在清华大学电子工程系作为教材为研究生新开课程使用。开始篇幅比较小，笔者每年参考当年有关的新进展并结合教学实践逐步扩充和增加了许多内容，如此六年才完成此书。值得指出的是，书中汇集了几年来不少听课学生的贡献，许多例题和练习题是在历届学生作业和课堂讨论的基础上提炼出来的，一些图片还直接由学生帮助制作，在选材上也从学生的反馈中受到许多启发。另外，本册书还采用了贾波、刘忠伟和姚玉荣等学生在学位论文工作中收集和实验得到的图片。

本册书初稿曾经北方交通大学阮秋琦教授和清华大学林行刚教授审阅，作者愿借此机会表示衷心的感谢。

全书在编写过程中得到王仁康编辑认真的审阅和精心的修改，这既使全书增色不少，也使作者受益匪浅，作者自行排版也得到了她的热心支持。

最后要指出，作者妻子何芸的直接帮助和牺牲精神，家人在各方面的理解和支持，是作者最终能把本书写到这一页的保证！

章毓晋

2000 年春节于清华园

通信：北京清华大学电子工程系，100084
电话：(010) 62781430
传真：(010) 62770317
电子邮件：zhangyj@ee.tsinghua.edu.cn
个人主页：www.ee.tsinghua.edu.cn/staff/zhangyujin
研究室网：image.ee.tsinghua.edu.cn

目 录

第 1 章 绪论	1	第 3 章 立体成象	25
1.1 图象工程概述	1	3.1 概述和分类	25
1.2 图象理解与计算机视觉	3	3.2 图象采集装置	26
1.2.1 图象理解	3	3.2.1 采集装置的性能指标	26
1.2.2 计算机视觉	4	3.2.2 灰度图采集和 CCD 摄象器件	26
1.2.3 图象理解与计算机视觉 的关系	4	3.2.3 深度图采集和结构光 测距装置	27
1.2.4 和图象理解与计算机视觉 相关的其它学科	5	3.3 成象变换和摄象机模型	29
1.2.5 图象理解与计算机视觉 的应用领域	6	3.3.1 世界坐标与摄象机坐标 重合时的摄象机模型	30
1.3 本书的安排和概要	6	3.3.2 世界坐标与摄象机坐标 分开时的摄象机模型	31
1.3.1 整体安排	6	3.3.3 通用摄象机模型	34
1.3.2 各章概要	7	3.4 摄象机校准	37
练习题	9	3.5 立体成象方式	39
第 2 章 视感觉和视知觉	10	3.5.1 立体成象方式分类	39
2.1 概述和分类	10	3.5.2 单目成象和畸变	40
2.2 视觉过程与特性	11	3.5.3 双目成象和视差	42
2.2.1 视觉过程	11	3.5.4 结构光成象和成象高度	44
2.2.2 视觉的时间特性	11	练习题	46
2.2.3 视觉的空间特性	12	第 4 章 扩展分析技术	48
2.3 形状知觉	13	4.1 概述和分类	48
2.3.1 轮廓	13	4.2 广义哈夫变换	49
2.3.2 图形和背景	15	4.2.1 基本哈夫变换原理	49
2.3.3 几何图形视错觉	18	4.2.2 广义哈夫变换原理	50
2.4 空间知觉	20	4.2.3 完整广义哈夫变换	52
2.4.1 非视觉性深度线索	20	4.3 亚像素边缘检测	54
2.4.2 双眼深度线索	21	4.3.1 基于矩保持的亚像素 边缘检测	54
2.4.3 单眼深度线索	22		
练习题	24		

4.3.2 利用一阶微分期望值的亚像素边缘检测	55	第6章 三维形状信息恢复	108
4.3.3 利用切线信息的亚像素边缘检测	56	6.1 概述和分类	108
4.4 3-D边缘检测和阈值分割	59	6.2 光度立体学	109
4.4.1 3-D边缘检测	59	6.2.1 场景亮度和图象亮度	109
4.4.2 3-D阈值分割	62	6.2.2 表面反射特性和亮度	111
4.5 16-邻域	63	6.2.3 目标表面朝向	113
4.5.1 术语和定义	64	6.2.4 反射图	114
4.5.2 N_{16} 空间中的连通性和最短通路	66	6.2.5 光度立体学求解	116
4.5.3 T_i 变换	69	6.3 从运动求取结构	119
4.5.4 N_{16} 空间中的距离	70	6.3.1 光流和运动场	119
4.6 3-D表面的表达	72	6.3.2 光流约束方程	121
4.6.1 3-D轮廓线的表达	72	6.3.3 光流计算	121
4.6.2 3-D目标表面的表达	72	6.3.4 光流与表面取向	126
4.6.3 表面面元插值3-D轮廓	73	6.4 从阴影恢复形状	128
4.7 3-D实体的表达	75	6.4.1 阴影与形状	128
4.7.1 表达方案	75	6.4.2 利用单目图象求解照度方程	132
4.7.2 广义圆柱体	76	6.5 纹理与表面朝向	136
4.7.3 扩展高斯图	78	6.5.1 基本方法	136
练习题	79	6.5.2 线段纹理消失点的确定	140
第5章 立体视觉	81	练习题	141
5.1 概述和分类	81	第7章 知识和表达	143
5.2 立体视觉模块	82	7.1 概述和分类	143
5.3 双目立体匹配	84	7.2 场景知识	144
5.4 多目立体匹配	86	7.3 过程知识	145
5.4.1 多目图象	86	7.4 知识表达基础	147
5.4.2 倒距离	87	7.4.1 对知识表达的要求	147
5.5 正交多目立体匹配	90	7.4.2 知识表达类型	148
5.5.1 基本原理	90	7.4.3 图象理解系统中的知识模块	149
5.5.2 基于梯度分类的正交三目立体匹配	94	7.4.4 图象理解中的知识表达	150
5.5.3 正交多目立体匹配	97	7.5 逻辑系统	152
5.6 视差图误差检测与校正	98	7.5.1 定义	152
5.7 亚像素级视差	102	7.5.2 利用定理证明来推理	155
练习题	106	7.6 语义网络	157
		7.7 产生式(专家)系统	160
		练习题	162

第 8 章 匹配和理解	163	9.5 典型系统比较	204
8.1 概述和分类	163	9.6 讨论和展望	205
8.2 模板匹配	164	练习题	209
8.3 目标匹配	165	附录 A 相关技术	210
8.3.1 字符串匹配	165	A.1 概述和分类	210
8.3.2 特征点匹配	166	A.2 统计模式识别	211
8.3.3 形状数匹配	167	A.2.1 模式和分类	211
8.3.4 惯量等效椭圆匹配	168	A.2.2 最小距离分类器	213
8.4 动态模式匹配	169	A.2.3 最优统计分类器	215
8.5 关系匹配	171	A.3 结构模式识别	219
8.6 线图同构	175	A.3.1 字符串结构识别	219
8.7 特征内容匹配	176	A.3.2 树结构识别	222
8.7.1 一般框架和系统模块	176	A.3.3 学习和推理	224
8.7.2 颜色匹配	177	A.4 人工神经网络	227
8.7.3 纹理匹配	180	A.4.1 发展和特点	227
8.7.4 形状匹配	182	A.4.2 用于两个模式类的 感知机模型	228
8.7.5 综合特征匹配	184	A.4.3 多层前向神经网络	232
8.7.6 语义内容匹配	186	A.4.4 多层网络决策面的 复杂度	238
8.7.7 基于内容的多媒体信息检索 与国际标准 MPEG-7	187	A.5 图象代数与形态分析	241
练习题	188	A.5.1 二值数学形态学	241
第 9 章 视觉理论和信息系统	190	A.5.2 灰度数学形态学的 一些基本算法	246
9.1 概述和分类	190	A.5.3 灰度数学形态学的 一些实用算法	249
9.2 视觉计算理论	191	A.5.4 图象代数	255
9.2.1 马尔理论	191	练习题	258
9.2.2 关于马尔理论的讨论	195	附录 B 参考文献	261
9.3 视觉信息系统模型	196	B.1 主要相关书籍	261
9.3.1 系统模型结构	196	B.2 本书引用的中文文献	262
9.3.2 多层次串行结构	197	B.3 本书引用的英文文献	263
9.3.3 以知识库为中心的 辐射结构	198	英文目录	269
9.3.4 以知识库为根的树结构	199		
9.3.5 多模块交叉配合结构	200		
9.4 具体系统分析	201		
9.4.1 VISIONS 系统	201		
9.4.2 ACRONYM 系统	202		
9.4.3 KB Vision 系统	203		

第1章 绪 论

1.1 图象工程概述

图象是用各种观测系统以不同形式和手段观测客观世界而获得的, 可以直接或间接作用于人眼并进而产生视知觉的实体[章 1996a]。人的视觉系统就具有这样的观察能力。人们还制造了许多利用各种辐射对客观场景成象的系统。图象是表达视觉信息的一种物理形式。数字图象采集的最终结果是某种能量的样本阵列, 所以常用矩阵或数组来表示, 其中每个元素的坐标代表场景点的位置而元素的值代表场景点的某个特性量[Zhang 1996]。

人们可用各种技术方式和手段对图象进行加工以获得需要的信息。图象技术从广义上可看作是各种图象加工技术的总称。它包括利用计算机和其它电子设备进行和完成的一系列工作, 例如图象的采集、获取、编码、存储和传输, 图象的合成和产生, 图象的显示、绘制和输出, 图象的变换、增强、恢复(复原)和重建, 图象的分割, 特征的提取和测量, 目标的检测、表达和描述, 序列图象的校正, 图象数据库的建立、索引、查询和抽取, 图象的分类、表示和识别, 3-D 景物的重建复原, 图象模型的建立, 图象知识的利用和匹配, 图象和场景的解释和理解, 以及基于它们的推理、判断、决策和行为规划等。另外, 图象技术还可包括为完成上述功能而进行的硬件和系统设计及制作等方面的技术[章 1999a]。

由于图象技术近年来得到极大的重视和长足的进展, 出现了许多有关的新理论、新方法、新算法、新手段、新设备。图象界一致认为亟需对它们进行综合研究和集成应用。我们以为这个工作需要在一个整体框架下进行, 这个框架就是图象工程[章 1996a]。众所周知, 工程是指将自然科学的原理应用到工业部门而形成的各学科的总称。图象工程学科则是将数学、光学等基础科学的原理, 结合计算机技术及在图象应用中积累的技术经验而发展起来的一个对整个图象领域进行研究应用的新学科。事实上, 图象技术多年来的发展和积累为图象工程学科的建立打下了坚实的基础, 而各类图象应用也对图象工程学科的建立提出了迫切的需要[章 2000e]。

图象工程的内容非常丰富, 应用也非常广泛, 根据抽象程度、研究方法、操作对象和数据量等的不同可分为三个层次(见[章 1999a]图 1.1.2): 图象处理、图象分析和图象理解。图象处理是比较低层的操作, 它主要在图象像素级上进行处理, 处理的数据量非常大。图象分析则进入了中层, 分割和特征提取把原来以像素描述的图象转变成比较简洁的非图形式的描述。图象理解主要指高层操作, 基本上是对从描述抽象出来的符号进行运算, 其处理过程和方法与人类的思维推理可以有许多类似之处。在这里, 随着抽象程度的提高, 数据量是逐渐减少的。具体说来, 原始图象数据经过一系列的处理过程, 逐步转化为更有组织和用途的信息。在这个过程中, 语义不断引入, 操作对象也逐步发生变化。另外, 高层操作对低层操作有指导作用, 能提高低层操作的效能, 完成复杂的任务。

概括地说，图象工程是既有联系又有区别的图象处理、图象分析及图象理解三者的有机结合，另外，还包括对它们的工程应用。从概念上讲，图象工程既能较好地将许多相近学科兼蓄并容，也进一步强调了图象技术的应用，所以我们选用图象工程来概括整个图象领域的研究应用。

图象工程是一门系统地研究各种图象理论、技术和应用的新的交叉学科。从它的研究方法来看，它与数学、物理学、生物学、生理学（特别是神经生理学）、心理学、电子学、计算机科学等许多学科可以相互借鉴；从它的研究范围来看，它与模式识别、计算机视觉、计算机图形学等多个专业又互相交叉。另外，图象工程的研究进展与人工智能、神经网络、遗传算法、模糊逻辑等理论和技术都有密切的联系，它的发展应用与生物医学、材料、遥感、通信、交通管理、军事侦察、文档处理和工业自动化等许多领域也是不可分割的。

图象工程是全面系统研究图象理论方法，阐述图象技术原理，推广图象技术应用及总结生产实践经验的新学科。结合本书内容，图象工程的主要构成可用图 1.1.1 的整体框架来表示。这里要用到各种图象技术以帮助人们从场景获得信息。首先要进行的就是利用各种方式从场景获得图象。接下来对图象的低层处理主要是为了改善图象的视觉效果或在保持视觉效果的基础上减少图象的数据量，处理的结果主要是为用户观看的。进一步对图象的中层分析主要是对图象里感兴趣的目标进行检测、提取和测量。分析的结果能为用户提供描述图象目标特点和性质的数据。最后对图象的高层理解是要通过对图象里各目标的性质和它们之间相互关系的研究，了解把握图象内容并解释原来的客观场景，理解的结果能为用户提供客观世界的信息，从而可以指导和规划行动。这些从低层到高层所用的图象技术都得到了包括人工智能、神经网络、遗传算法、模糊逻辑、图象代数等新理论、新工具、新技术的有力支持。为完成这些工作还要采取合适的策略进行控制。

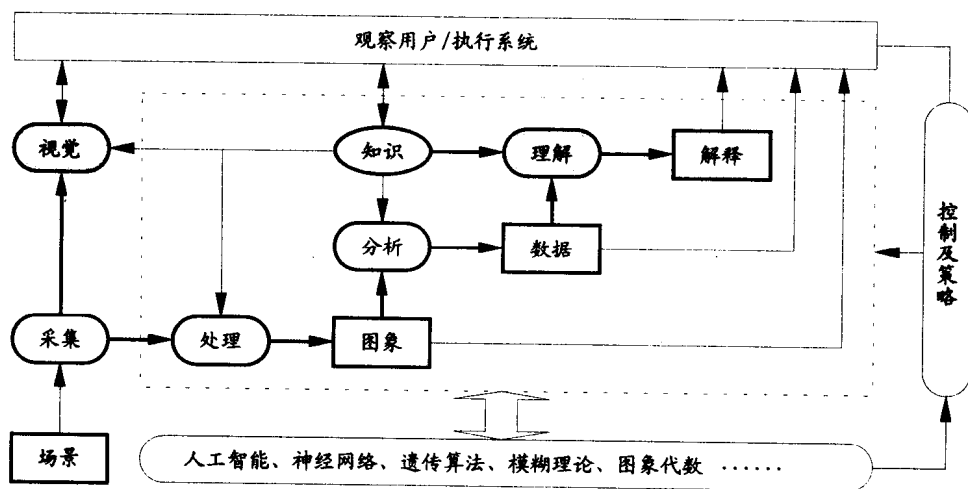


图 1.1.1 图象工程整体框架

上面所述关于低层图象处理中各重要分支的理论、技术和方法，以及中层图象分析的基本原理和技术已在上册[章 1999a]里进行了详细的介绍，而高层中有关视觉（和视知觉）、知识及应用、图象理解、对场景的解释以及为完成这些工作所采用的控制和策略将是本册

书的主要内容。事实上，高层理解还可进一步分为两个层次；在较低层是借助图象对场景本身的重建和恢复，而在较高层则是对客观景物的识别和解释。

1.2 图象理解与计算机视觉

1.2.1 图象理解

如前所述，图象工程从低到高的三个层次组成一个有机的联合体，它们在从图象获取信息方面都起着重要的作用。但目前对利用高层技术的研究和应用相对比较少[章 1997a]，[章 1998c]，[章 1999c]，[章 2000c]。

图象理解作为图象工程的高层，其重点是在图象分析的基础上，结合人工智能和认知理论，进一步研究图象中各目标的性质和它们之间的相互联系，并理解图象内容的含义以及解释原来的客观场景，从而指导和规划行动。如果说图象分析主要是以观察者为中心研究客观世界（主要研究可观察到的事物），那么图象理解在一定程度上是以客观世界为中心，借助知识、经验等来把握整个客观世界（包括没有直接观察到的事物）。

图象理解是要用计算机系统来帮助解释图象的含义，从而实现利用图象信息解释客观世界。它要确定为完成某个视觉任务需要通过图象采集从客观世界获取哪些信息，需要通过图象处理和分析从图象中提取哪些信息，以及利用哪些信息继续加工以获得需要的解释。它要研究理解能力的数学模型，并通过对数学模型的程序化，实现理解能力的计算机模拟。

限于目前计算机系统的能力和图象理解技术的水平，上述许多任务还无法完全自动实现。在多数情况下，“系统”完成较低层的工作，而人需要接着系统完成剩下的较高层的工作，如图 1.2.1 所示。

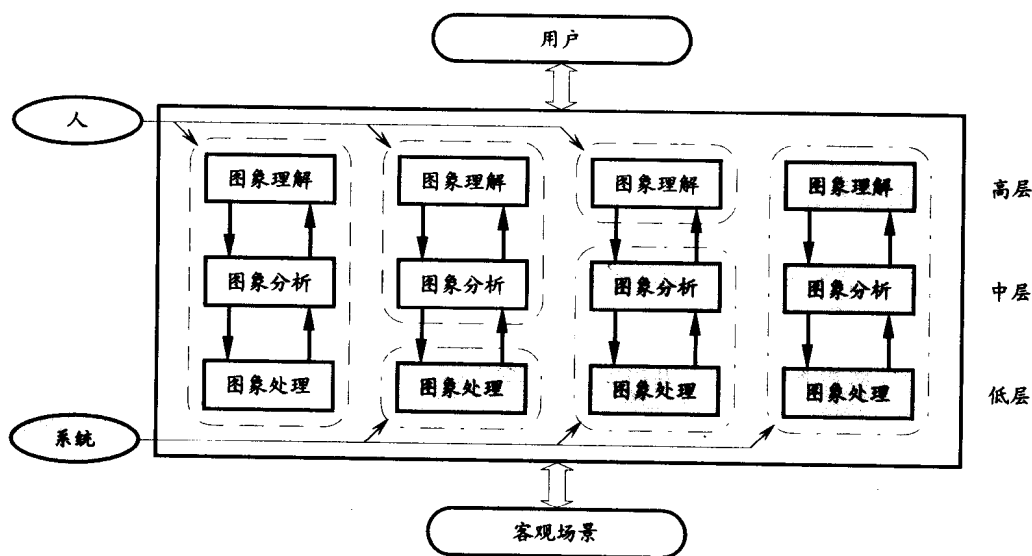


图 1.2.1 系统和用户覆盖不同的层次

图象理解是要用户通过图象认识客观世界。如果完全没有系统,人就必须执行所有各层工作。如果系统只有低层能力,那么人就需要在系统处理的基础上完成中层和高层的工作。如果系统有了低层和中层的能力,那么用户只需要在系统处理的基础上完成高层的工作就可以了。而如果系统有了从低层直到高层的能力,用户很容易就可做出决策。

随着计算机科学和技术的发展,计算机系统正在逐步实现对图象的处理、分析和理解功能,减少人的介入。目前瓶颈主要在理解层次。

1.2.2 计算机视觉

视觉是人类观察世界、认知世界的重要功能手段。人类从外界获得的信息约有 75% 来自视觉系统,这既说明视觉信息量巨大,也表明人类对视觉信息有较高的利用率。人类视觉过程可看作是一个复杂的从感觉(感受到的是对 3-D 世界之 2-D 投影得到的图象)到知觉(由 2-D 图象认知 3-D 世界内容和含义)的过程。视觉的最终目的从狭义上说是要对场景作出对观察者有意义的解释和描述,从广义上讲,还有基于这些解释和描述并根据周围环境和观察者的意愿制定出行为规划。

计算机视觉是指用计算机实现人的视觉功能——对客观世界的三维场景的感知、识别和理解。这里主要有两类方法:一类是仿生学的方法,参照人类视觉系统的结构原理,建立相应的处理模块完成类似的功能和工作;另一类是工程的方法,从分析人类视觉过程的功能着手,并不去刻意模拟人类视觉系统内部结构,而仅考虑系统的输入和输出,并采用任何现有的可行的手段实现系统功能。本书主要讨论第二种方法。

计算机视觉的主要研究目标可归纳成两个,它们互相联系补充。第一个目标是建成计算机视觉系统,完成各种视觉任务。换句话说,即要使计算机能借助各种视觉传感器(如 CCD, CMOS 摄像器件等)获取场景的图象,而感知和恢复 3-D 环境中物体的几何性质、姿态结构、运动情况、相互位置等,并对客观场景进行识别、描述、解释,进而作出决断。这里主要研究的是技术机理。目前这方面的工作集中在建成各种专用系统,完成在各种实际场合提出的专门视觉任务,而从长远来说则要建成通用的系统[Jain 1997]。

第二个研究目标是把该研究作为探索人脑视觉工作机理的手段,进一步加深对人脑视觉的掌握和理解(如计算神经科学)。这里主要研究的是生物学机理。长期以来,对人脑视觉系统已从生理、心理、神经、认知等方面进行了大量的研究,但还远没有揭开视觉过程的全部奥秘,可以说对视觉机理的研究和了解还远落后于对视觉信息处理的研究和掌握。需要指出,对人脑视觉的充分理解也将促进计算机视觉的深入研究[Finkel 1994]。

1.2.3 图象理解与计算机视觉的关系

图象理解和计算机视觉密切相关。图象是表达视觉信息的一种物理形式,图象理解必须借助计算机,基于对视觉信息的处理和分析来进行。计算机视觉作为一门学科,与许多以图象作为主要研究对象的学科,特别是图象处理、图象分析、图象理解有着非常密切的联系和不同程度的交叉。计算机视觉主要强调用计算机实现人的视觉功能,这中间实际上需要用到图象工程三个层次的许多技术,虽然目前的研究内容主要与图象理解相结合。

在建立视觉信息系统,用计算机协助人类完成各种视觉任务方面,图象理解和计算机视觉都需要用到摄影几何学、概率论与随机过程、人工智能等理论。例如它们都要借助两类智能活动:① 感知,如感知场景中可见部分的距离、朝向、形状、运动速度、相互关系等;② 思维,如根据场景结构分析物体的行为,推断场景的发展变化,决定和规划主体行动等。

图象理解和计算机视觉在很多情况下内容交叉重合。事实上这两个名词也常混合使用,它们互相联系,覆盖面有所重合,在概念上或实用中并没有绝然的界限,有时可认为它们虽各有侧重但常常是互为补充的。在许多场合和情况下,将它们看作是专业和背景不同的人习惯使用的不同术语更为恰当。在本书中,我们也不去刻意区分。

例 1.2.1 从一份刊物的名称变化看图象理解与计算机视觉的联系与进展

1972 年与图象技术关系密切的一份国际期刊创刊了,它当时的名称是 CGIP (计算机图形学与图象处理)。1983 年它改名为 CVGIP (计算机视觉、图形学与图象处理)。1991 年它分成两个版本,分别称为 CVGIP-GMIP (图形模型与图象处理)和 CVGIP-IU (图象理解)。1995 年它转成两个期刊,前者命名为 GMIP (图形模型与图象处理),后者命名为 CVIU (计算机视觉与图象理解)。□

1.2.4 和图象理解与计算机视觉相关的其它学科

图象理解和计算机视觉与计算机科学有密切的联系,它是随着计算机技术的发展和深入研究而获得突飞猛进发展的。一些计算机学科,如模式识别、人工智能、计算机图形学等都对这个发展起到了并将继续起到重要的影响和作用。

模式包括的范围很广,图象就是模式的一种。识别是指从客观事实中自动建立符号描述或进行逻辑推理的数学和技术,因而人们定义模式识别是对客观世界中物体和过程进行分类、描述的学科。目前,对图象模式的识别主要集中在对图象中感兴趣的内容(目标)的分类、分析和描述。图象理解和计算机视觉中使用了很多模式识别的概念和方法,但视觉信息有其特殊性和复杂性,传统的模式识别(竞争学习模型)并不能把图象理解和计算机视觉全包括进去。

人类智能主要指人类理解世界、判断事物、学习环境、规划行为、推理思维、解决问题等的能力。人工智能则指由人类用计算机模拟、执行或再生某些与人类智能有关功能的能力和技术。视觉功能是人类智能的一种体现,所以图象理解和计算机视觉与人工智能密切相关。图象理解和计算机视觉的研究中使用了许多人工智能技术,反过来,图象理解和计算机视觉也是人工智能的一个重要应用领域,需要借助人工智能的理论研究成果和系统实现经验。

计算机图形学研究如何从给定的描述生成“图象”,与计算机视觉也有密切关系。某些图形可以认为是图象分析结果的可视化,而计算机真实感景物的生成又可以认为是图象分析的逆过程[章 1996b]。另外,图形学技术在视觉系统的人机交互和建模等过程中也起很大作用。近期两相结合的一个研究热点——基于图象的绘制(image based rendering)就是一个很好的例子。需要注意,与图象理解和计算机视觉中存在许多不确定性对比,计算

机图形学处理的多是确定性问题,是通过数学途径可以解决的问题。许多实际情况下是要在图形生成的速度和精度,即实时性和逼真度之间取得某种妥协。

图象理解和计算机视觉要用工程方法解决生物的问题,完成生物固有的功能,所以与生物学、生理学、心理学、神经学等学科也有着互相学习互为依赖的关系。图象理解和计算机视觉属于工程应用科学,与电子学、集成电路设计、通信工程等密不可分。一方面图象理解和计算机视觉的研究充分利用了这些学科的成果;另一方面图象理解和计算机视觉的应用也极大地推动了这些学科的深入研究和发 展。目前图象理解和计算机视觉研究者与视觉心理生理研究者的互相结合就是一例。

1.2.5 图象理解与计算机视觉的应用领域

图象理解与计算机视觉近年来已在许多领域得到广泛应用,下面是一些典型的例子:

(1) 工业视觉:如工业检测、工业探伤、自动生产流水线、邮政自动化、计算机辅助外科手术、显微医学操作,以及各种危险场合工作的机器人等。将图象和视觉技术用于生产自动化,可以加快生产速度,保证质量的一致性,还可以避免人的疲劳、注意力不集中等带来的误判。

(2) 人机交互:如人脸识别、智能代理等,让计算机可借助人的手势动作(手语)、嘴唇动作(唇读)、躯干运动(步态)、表情测定等了解人的愿望要求而执行指令,这既符合人类的交互习惯,也可增加交互方便性和临场感等。

(3) 视觉导航:如巡航导弹制导、无人驾驶飞机飞行、自动行驶车辆、移动机器人、精确制导等,既可避免人的参与及由此带来的危险,也可提高精度和速度。

(4) 虚拟现实:如飞机驾驶员训练、医学手术模拟、场景建模、战场环境表示等,它可帮助人们超越人的生理极限,“亲临其境”,提高工作效率。

(5) 图象自动解释:包括对放射图象、显微图象、遥感多波段图象、合成孔径雷达图象、航天航测图象等的自动判读理解。由于近年来技术的发展,图象的种类和数量飞速增长,图象的自动理解已成为解决信息膨胀问题的重要手段。

(6) 对人类视觉系统和机理、人脑心理和生理的研究等。

1.3 本书的安排和概要

1.3.1 整体安排

图象理解和计算机视觉是一门相对较新的还不太成熟的学科,由于其理论和技术上的复杂性,至今还没有建立起完整的体系,研究还在不断深入发展。从成书的角度看,这门学科目前还没有统一的框架体系,范围还在逐步扩展,所以结构次序都很难确定。此外,哪些内容是基础,哪些内容是关键,处于不断变化之中,所以选材取舍也不易把握。

本书试图介绍一些图象理解和计算机视觉的基本研究内容和实用技术,并把它们有机

地融合在通用的视觉系统框架里，一方面为在该领域深入研究打下基础，另一方面也为在实际中运用图象和视觉技术提供帮助。需要指出，尽管学科还不成熟，仍有大量的专著、期刊、会议集等已对图象理解和计算机视觉及其相关技术和应用进行了多方面的介绍（见附录 B）。

本书主要包括图象工程的最高层次——图象理解——及密切相关的计算机视觉的内容（见图 1.1.1 中视觉、知识、理解、解释、控制及策略等）。全书结构见图 1.3.1。

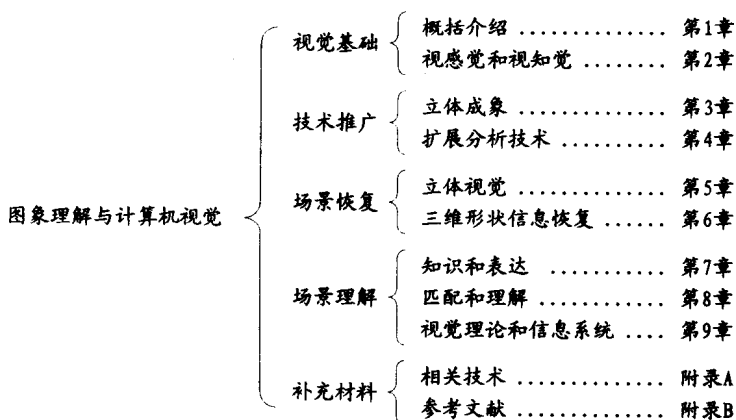


图 1.3.1 本书结构

全书共分为 9 章正文和 2 个附录。主要可分为 5 部分（见图 1.3.1）。第 1 部分概述了一些有关图象和视觉的基础知识，由前 2 章组成；第 2 部分是对一些图象采集、处理和分析技术的推广（在上册[章 1999a]的基础上），包括第 3, 4 章；第 3 部分主要讨论有关场景恢复的技术，包括第 5, 6 章；第 4 部分介绍有关对场景进行理解所涉及的内容，包括第 7, 8, 9 章。这两部分是本书的主要内容。第 5 部分包括两个附录（可看作补充材料），附录 A 介绍与图象理解和计算机视觉相关的一些技术，附录 B 给出可进一步深入学习以及本书引用的参考文献目录。

1.3.2 各章概要

第 1 章是全书总纲，试图使读者对全书，也对整个领域有概括的了解。在对图象工程作概述后，将图象理解作为其高层提出来，并讨论了计算机视觉以及图象理解与计算机视觉的关系、相关学科和应用领域，还介绍了本书的范围、整体安排和各章概要。

第 2 章介绍视感觉和视知觉的基础知识，以及视感觉和视知觉研究的一些情况。将视觉分成视感觉和视知觉，在视感觉方面主要介绍视觉过程和时间空间特性。在视知觉方面，主要介绍形状知觉和空间知觉的一些知识。

第 3 章介绍立体图象的采集。图象是计算机视觉系统的输入，也是视觉研究的出发点。对客观场景成象需要一定的图象采集设备（装置）和相应的成象方式，所以这里先介绍采集装置和摄像机模型。如果要定量地将摄像机所测量的内容和 3-D 场景中的物体联系起来，就要对摄像机进行校正标定。而为了能有效地采集和获取场景中的 3-D 信息，还需以

特定的方式采集图象，包括本征图象和非本征图象。为此，该章还介绍了摄像机校准和立体图象采集技术。

第4章介绍几种基本图象处理和分析技术的推广。图象处理和分析的自身技术在不断发展，而图象理解又建立在处理和分析基础之上，所以本章内容有承前启后的作用。所介绍的推广技术包括广义哈夫变换、亚像素级边缘检测和3-D图象分割等分割技术；还包括像素邻域的扩展（由常用的4-邻域和8-邻域扩展到16-邻域），以及目标表达的扩展（由2-D线和面推广到3-D目标表面和由2-D区域推广到3-D目标实体）。

第5章介绍立体视觉技术。立体视觉主要研究如何借助（多图象）成像技术从（多幅）图象获取场景中物体的距离（深度）信息。这里主要讨论采用双目或多目图象，通过视差计算以获得绝对深度信息的方法，其中重点在不同图象对应点匹配方面，以及正交多目的优越性。另外为了克服几何畸变和噪声干扰等影响介绍了对视差图中误差的检测和校正技术。视差的计算和深度信息的恢复是各项后续工作的基础，因此该章还介绍了可在获得一般像素级视差后进一步改善精度以达到亚像素级视差精度的方法。

第6章介绍利用单目2-D图象恢复3-D目标形状信息的技术。这与第5章的立体视觉技术相辅相成，都是为了恢复客观场景。从单目图象得到目标形状常称为“从 X 得到形状”（shape from X ），这里 X 可以代表阴影、轮廓、纹理、运动等各种线索。这里将介绍如何从图象灰度阴影和（对应物体表面）纹理的变化求取物体的三维形状，以及如何运用光度立体学方法或从物体的运动线索中得到物体三维结构的技术。

第7章介绍有关视觉知识的分类、表达和使用。知识在整个视觉过程的各个阶段都起到重要的作用，例如在分析推理和对场景作出正确的解释时需要用到各种经验知识。借助对知识的利用可以有效地提高分析理解的可靠性和效率。要有效地利用知识，就要研究有哪些知识可以利用，如何表达和管理它们。这里既介绍场景知识（有关研究对象的一些客观事实），也介绍过程知识（什么样的知识在什么样的场合需要运用、可以运用和如何运用），还对一些常用的知识表达形式，如守护程序、黑板系统、图象函数库、框架、单元、剧本形式逻辑、语义网络、产生式系统等给予了详细介绍。

第8章介绍匹配和理解技术。匹配可理解为结合不同的、已经存在的表达而建立它们的解释之间联系的技术与过程。这里研究如何把知识模型与从图象获得的描述信息匹配起来，从而达到识别目标、理解图象、认知世界的目的。因为知识可在不同的层次运用，所以匹配可在不同抽象层次上进行。本章将从图象到目标，从具体到抽象，从特征到内容渐次介绍，还结合了近年基于内容图象检索的研究成果。

第9章介绍视觉信息系统。视觉信息系统是各种图象理解和计算机视觉技术的集成和集中体现。前面各章讨论的内容主要是实现从图象来理解场景所用的一些具体技术。本章则试图将它们结合进视觉信息系统结构中，从体系上统一考虑。这里我们既介绍视觉计算理论，也介绍各种视觉系统模型，并剖析几个具体的实用系统，归纳一些典型系统的特点、优点和缺点。最后，讨论一些人们长期研究的问题、当前的研究热点，以及图象理解与计算机视觉的发展前景。

附录A介绍一些相关技术。图象理解和计算机视觉的研究离不开许多其它方面的工具和方法。这里介绍了统计模式识别和结构（句法）模式识别的基本原理，以及广泛用于

模式识别和其它许多领域的人工神经网络技术。另外，还在上册对数学形态学基本操作运算介绍的基础上，进一步讨论了图象代数和形态学分析。这些既可以作为补充材料，也可根据教学对象和专业的不同与其它章节置换。

附录 B 介绍参考文献。上册已介绍了文献查询和引用要点，所以这里仅提供一些用于进一步深入学习的重要专著以及本书引用的参考文献目录。这不仅对学习本书有益，对进一步从事科研工作也有帮助。

为了帮助读者查阅英文资料，本书后面附了英文目录。

由图 1.3.1 可见，全书内容基本按树结构安排，这种安排方式也用在其后每一章中。各章第 1 节中将介绍该章主要内容的树结构。虽然实际上各章节互相之间也是有直接联系的，但这样做尽可能保持了各章节的独立性。给出这种结构的主要目的是保证读者在学习任一章节时能知道所学内容是图象理解和计算机视觉中的哪个分支，能随时方便地确定自己“遨游”的位置。

为了帮助学习和加强理解，各章都安排了一定例题，每章章末还有集中的练习题。鉴于图象理解和计算机视觉内容的特点，有些练习题可能没有唯一的答案，有些主要是引导思维，有些可作为“项目”训练使用[章 1999b]。

练 习 题

- 1.1 为什么说图象理解处在图象工程的高层？
- 1.2 计算机视觉要达到的目的有哪些？什么是基本目的，什么是最终目的？
- 1.3 人类视觉系统与计算机视觉系统有哪些关系？
- 1.4 对图象的理解和用计算机研究实现视觉功能有哪些联系，为什么将它们结合介绍？你更喜欢用哪个名称？
- 1.5 实现图象理解有几类技术方法，各自的特点是什么？
- 1.6 图象理解与计算机视觉研究领域中有许多病态问题，列举造成病态的原因。
- 1.7 近年来国际上有几次关于图象理解与计算机视觉研究方面的大讨论，有些论文刊登在国际刊物 CVGIP 和 CVIU 上（见例 1.2.1）。请查阅有关论文，看看它们讨论的焦点是什么，对推动这方面的研究起到了哪些作用？你对此有何看法？
- 1.8 近年来仿生学方面有哪些进展对视觉研究产生了影响？
- 1.9 图象与图形既有联系又有区别，试归纳总结之。
- 1.10 查阅近年来有关图象理解与计算机视觉研究方面的文献，了解其发展趋势。
- 1.11 试举例说明一幅图象本身并不具备恢复客观场景的所有几何信息。
- 1.12 试举例说明一幅图象本身并不具有精确理解客观场景的全部信息。