

21世纪高等学校计算机规划教材

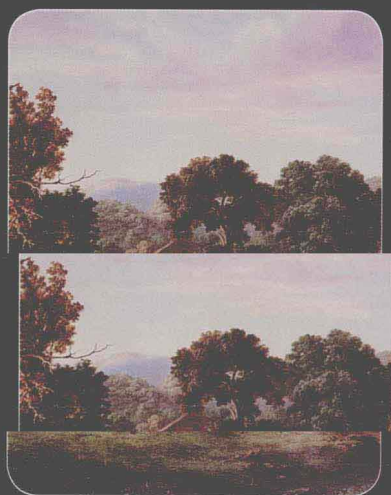
21st Century University Planned Textbooks of Computer Science

# 计算机 视觉教程

A Course of Computer Vision

章毓晋 编著

- 选材全面、原理清晰、方法详尽、注重实用
- 突出重点、解释直观、定义精炼、示例丰富
- 各章内容平衡、方便课堂教学、提供术语索引



 人民邮电出版社  
POSTS & TELECOM PRESS

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

# 计算机 视觉教程

A Course of Computer Vision

章毓晋 编著



名家系列

人民邮电出版社

北京

## 图书在版编目(CIP)数据

计算机视觉教程 / 章毓晋编著. -- 北京 : 人民邮电出版社, 2011. 3  
21世纪高等学校计算机规划教材  
ISBN 978-7-115-24290-7

I. ①计… II. ①章… III. ①计算机视觉—高等学校—教材 IV. ①TP302.7

中国版本图书馆CIP数据核字(2010)第242071号

## 内 容 提 要

本书系统地介绍计算机视觉的一些基本原理、典型方法和实用技术,内容包括视觉和视知觉、图像采集、图像预处理、基元检测、目标分割、目标表达和描述、形状特性分析、立体视觉、三维景物恢复、运动分析、景物识别、广义匹配、场景解释。读者可从中了解计算机视觉的基本原理和典型技术,并能据此解决计算机视觉应用中的一些具体问题。本书提供了许多讲解例题,每章均有要点和小结、参考文献介绍以及练习题(为部分练习题提供了解答)。

本书可作为信息科学、计算机科学、计算机应用、信号与信息处理、通信与信息系统、电子与通信工程、模式识别与智能系统等学科大学本科或研究生的专业基础课教材,也可作为远程教育或继续教育中计算机应用、电子技术等专业的研究生课程教材,还可供涉及计算机视觉技术应用行业(如工业自动化、人机交互、办公自动化、视觉导航和机器人、安全监控、生物医学、遥感测绘、智能交通和军事公安等)的科技工作者自学或科研参考。

21世纪高等学校计算机规划教材

### 计算机视觉教程

- 
- ◆ 编 著 章毓晋  
责任编辑 武恩玉
  - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号  
邮编 100061 电子函件 315@ptpress.com.cn  
网址 <http://www.ptpress.com.cn>  
三河市海波印务有限公司印刷
  - ◆ 开本: 787×1092 1/16  
印张: 18.5 2011年3月第1版  
字数: 487千字 2011年3月河北第1次印刷

ISBN 978-7-115-24290-7

定价: 36.00元

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154

广告经营许可证: 京崇工商广字第0021号

# 前 言

本书是一部介绍计算机视觉的基本原理、典型方法和实用技术的专门教材，希望为普通高等工科院校的计算机及相近专业开设第一门计算机视觉课程服务。

本书选材比较精练，但基本覆盖了计算机视觉的主要内容，同时也兼顾了不同专业背景的学习者和自学读者学习的需要。书中尽量少涉及信号处理的基础知识，也不通过从 1-D 信号向 2-D 推广来介绍内容。因为计算机视觉技术现已是一门比较实用的技术，所以本书比较注重实用性。本书没有过多地强调理论性，尽量减少了公式推导。本书提供了较多的例题，通过直观的解释来帮助读者理解抽象的概念。

本书编写时，从读者学习的角度出发，对概念尽可能先给出比较精练的定义再进行具体介绍，对所涉及的技术和方法除给出原理外，还提供了比较详细的描述。本书中的各部分相对独立，对每个概念或方法尽量一次讲解清楚，基本不需参引书中的其他部分。

本书从方便课堂教学的角度出发，设计了内容比较均衡，并且篇幅基本一样的 14 章（每章均分 4 节讲解内容），可以安排每章一次课，用于一个学期的教学。对专业基础较好或较高年级的学生，可考虑每章用 3 个学时，对其他一些相近专业或较低年级的学生，可考虑每章用 4 个学时。本书各章后的练习题也是等数量的，同样考虑了复习的均衡。

本书每章后都配有总结和复习，一方面总结了该章各节的要点，帮助复习；另一方面，有针对性地介绍一些相关的参考文献，帮助学有余力的学生进一步深入探讨。本书还在后面给出了术语索引（文中标为黑体）。其中对每个术语给出了对应的英文，既方便读者对本书进行查阅，也方便读者联网搜索相关资料。

本书从结构上看，共有 14 章正文，以及部分练习题答案、参考文献和索引。在这 17 个一级标题下共有 70 个二级标题（节），再之下还有 112 个三级标题（小节）。全书共有文字（包括图片、绘图、表格、公式等）约 45 万字，有编了号的图 261 个、表格 30 个、公式 540 个。为便于教学和理解，本书共给出各类例题 75 个，思考题和练习题 168 个，对其中的 54 个练习题提供了参考解答（它们有些补充了正文内容，有些给出了更多的示例）。另外最后列出了所介绍的 200 多篇参考文献的目录和用于索引的 600 多个术语。

感谢出版社编辑的精心组织，认真审阅和细心修改。最后，作者感谢妻子何芸、女儿章荷铭以及父母家人在各方面给予的理解和支持。

章毓晋

2010 年暑假于蓝旗营

通信：北京清华大学电子工程系，100084

电话：(010)62781430

传真：(010)62770317

邮箱：zhang-yj@tsinghua.edu.cn

主页：www.ee.tsinghua.edu.cn/~zhangyujin/

# 目 录

<b>第 1 章 绪论</b> .....	1	<b>第 3 章 图像采集</b> .....	38
1.1 计算机视觉.....	1	3.1 采集模型.....	38
1.1.1 视觉.....	1	3.1.1 几何成像模型.....	38
1.1.2 计算机视觉概述.....	2	3.1.2 亮度成像模型.....	43
1.1.3 相关学科.....	3	3.2 采集装置.....	44
1.1.4 应用领域.....	4	3.2.1 采集装置及性能指标.....	44
1.2 图像基础.....	4	3.2.2 空间和幅度分辨率.....	46
1.2.1 图像.....	4	3.3 采集方式.....	47
1.2.2 图像表达和显示.....	6	3.3.1 成像方式一览.....	47
1.2.3 图像存储.....	8	3.3.2 结构光法.....	48
1.3 像素间联系.....	10	3.4 摄像机标定.....	50
1.3.1 像素邻域.....	11	3.4.1 标定程序和步骤.....	50
1.3.2 像素间距离.....	12	3.4.2 两级标定法.....	52
1.4 本书内容提要.....	14	总结和复习.....	55
1.4.1 计算机视觉系统及模块.....	15	<b>第 4 章 图像预处理</b> .....	57
1.4.2 如何学习使用本书.....	16	4.1 坐标变换.....	57
总结和复习.....	18	4.1.1 基本坐标变换.....	57
<b>第 2 章 视觉和视知觉</b> .....	20	4.1.2 几何失真校正.....	59
2.1 视觉过程和特性.....	20	4.2 灰度映射.....	61
2.1.1 视觉过程.....	20	4.2.1 灰度映射原理.....	62
2.1.2 视觉的时间特性.....	22	4.2.2 灰度映射示例.....	62
2.1.3 视觉的空间特性.....	24	4.3 直方图修正.....	64
2.2 形状知觉.....	25	4.3.1 直方图均衡化.....	64
2.2.1 形状的感知.....	25	4.3.2 直方图规范化.....	66
2.2.2 轮廓.....	26	4.4 空域滤波.....	70
2.2.3 图形和背景.....	28	4.4.1 原理和分类.....	70
2.3 空间知觉.....	30	4.4.2 线性平滑滤波.....	71
2.3.1 非视觉性深度线索.....	30	4.4.3 线性锐化滤波.....	72
2.3.2 双目深度线索.....	31	4.4.4 非线性平滑滤波.....	73
2.3.3 单目深度线索.....	33	4.4.5 非线性锐化滤波.....	74
2.4 运动知觉.....	35	总结和复习.....	74
总结和复习.....	36	<b>第 5 章 基元检测</b> .....	77
		5.1 边缘检测.....	77

5.1.1 检测原理	77	7.3 基于边界的描述	125
5.1.2 一阶导数算子	78	7.3.1 边界长度和直径	125
5.1.3 二阶导数算子	80	7.3.2 边界形状数	126
5.1.4 边界闭合	82	7.4 基于区域的描述	127
5.2 SUSAN 算子	83	7.4.1 区域面积和密度	127
5.2.1 USAN 原理	83	7.4.2 区域形状数	128
5.2.2 角点和边缘检测	84	7.4.3 不变矩	129
5.3 哈夫变换	87	7.4.4 拓扑描述符	131
5.3.1 基本哈夫变换原理	87	总结和复习	132
5.3.2 广义哈夫变换原理	88	<b>第 8 章 形状特性分析</b>	135
5.3.3 完整广义哈夫变换	90	8.1 形状紧凑性描述符	135
5.4 位置直方图技术	92	8.2 形状复杂性描述符	141
总结和复习	94	8.3 基于多边形的形状分析	143
<b>第 6 章 目标分割</b>	96	8.3.1 多边形的获取	143
6.1 轮廓搜索	96	8.3.2 多边形描述	145
6.1.1 图搜索	96	8.4 基于曲率的形状分析	146
6.1.2 动态规划	98	8.4.1 轮廓曲率	146
6.2 主动轮廓模型	99	8.4.2 曲面曲率	149
6.2.1 主动轮廓	99	总结和复习	151
6.2.2 能量函数	100	<b>第 9 章 立体视觉</b>	153
6.3 基本阈值技术	103	9.1 立体视觉模块	153
6.3.1 原理和分类	103	9.2 双目成像和视差	155
6.3.2 全局阈值的选取	104	9.2.1 双目横向模式	155
6.3.3 局部阈值的选取	106	9.2.2 双目横向会聚模式	157
6.3.4 动态阈值的选取	109	9.2.3 双目纵向模式	158
6.4 特色阈值方法	110	9.3 基于区域的立体匹配	159
6.4.1 多分辨率阈值	110	9.3.1 模板匹配	160
6.4.2 过渡区阈值	111	9.3.2 双目立体匹配	161
总结和复习	114	9.4 基于特征的立体匹配	165
<b>第 7 章 目标表达和描述</b>	116	9.4.1 点对点的方法	165
7.1 基于边界的表达	116	9.4.2 动态规划匹配	167
7.1.1 链码	116	总结和复习	168
7.1.2 边界段和凸包	118	<b>第 10 章 三维景物恢复</b>	171
7.1.3 边界标记	119	10.1 由光移恢复表面朝向	171
7.2 基于区域的表达	121	10.1.1 表面反射特性	172
7.2.1 四叉树	121	10.1.2 目标表面朝向	174
7.2.2 围绕区域	122	10.1.3 反射图	174
7.2.3 骨架	122		

10.1.4 光度立体学求解	176	<b>第 13 章 广义匹配</b>	226
10.2 从明暗恢复形状	177	13.1 目标匹配	227
10.2.1 明暗与形状	177	13.1.1 匹配的度量	227
10.2.2 求解亮度方程	180	13.1.2 字符串匹配	229
10.3 纹理变化与表面朝向	182	13.1.3 惯量等效椭圆匹配	231
10.3.1 三种典型变化	182	13.2 动态模式匹配	233
10.3.2 确定线段的纹理消失点	184	13.3 关系匹配	235
10.4 根据焦距确定深度	185	13.3.1 关系表达和距离	235
总结和复习	187	13.3.2 关系匹配模型	237
<b>第 11 章 运动分析</b>	189	13.4 图同构匹配	238
11.1 运动分类和表达	189	13.4.1 图论基础	238
11.2 全局运动检测	193	13.4.2 图同构和匹配	241
11.2.1 利用图像差的运动检测	193	总结和复习	242
11.2.2 基于模型的运动检测	196	<b>第 14 章 场景解释</b>	245
11.3 运动目标分割	197	14.1 线条图标记解释	245
11.4 运动光流和表面取向	198	14.2 体育比赛视频排序	248
11.4.1 光流约束方程	199	14.3 计算机视觉系统模型	252
11.4.2 光流计算	199	14.3.1 多层次串行结构	252
11.4.3 光流与表面取向	203	14.3.2 以知识库为中心的辐射结构	253
总结和复习	205	14.3.3 多模块交叉配合结构	254
<b>第 12 章 景物识别</b>	207	14.4 计算机视觉理论框架	255
12.1 统计模式分类	207	14.4.1 马尔视觉计算理论	256
12.1.1 模式分类原理	208	14.4.2 对马尔理论框架的改进	258
12.1.2 最小距离分类器	208	14.4.3 新理论框架的研究	259
12.1.3 最优统计分类器	209	总结和复习	262
12.2 感知机	213	<b>部分练习题解答</b>	263
12.3 支持向量机	216	<b>参考文献</b>	273
12.4 结构模式识别	219	<b>索引</b>	282
12.4.1 字符串结构识别	219		
12.4.2 树结构识别	222		
总结和复习	224		

# 第 1 章

## 绪论

计算机技术的发展,使计算机具有视觉功能进一步得到广泛关注和研究,理论和实践都越来越成熟。同时,与计算机视觉相关的课程和教材也在不断更新。本书是一本专门用于计算机视觉课程教学的教材,主要介绍一些基本和典型的计算机视觉技术。

计算机视觉作为一门学科,与数学、物理学、生理学、感知心理学、神经科学以及计算机科学等都有密切的联系。虽然具有相关学科的基础对学习计算机视觉很重要,很有利,但本书作为一本基础教材,也可以自成体系地进行学习。本书主要介绍计算机视觉的内容,在需要时也对相关预备知识给予了概括介绍。

本章作为本书的绪论,先对计算机视觉学科的总体内容和范围给予概括描述,然后对密切相关的基本图像概念和基础知识给予具体介绍,并详细阐述了图像中像素之间的联系,以便为后续各章的学习奠定基础。

本章各节安排如下。

1.1 节对计算机视觉给予概括介绍,包括密切联系的人类视觉、计算机视觉的研究方法和研究目标、几个主要的相关学科,以及应用领域等。

1.2 节介绍有关图像的基本概念和基础知识,除列举了多种常见的图像类型以外,还对图像的表达和显示设备,以及图像的存储设备和文件格式进行了讨论。

1.3 节详细描述了图像中像素之间的联系,包括与像素邻域相关的几个概念,像素间距离的计算,以及距离和邻域尺寸之间的关系。

1.4 节结合对计算机视觉系统框架的描述,列出了本书的主要内容、结构和安排,对各章的重要内容给予了概括介绍,并在此基础上对如何选取书中内容用于教学提出了一些建议。

## 1.1 计算机视觉

下面对计算机视觉的起源、目的、范围、内涵、特点、应用等进行概括介绍。

### 1.1.1 视觉

说到计算机视觉,需要先介绍一下人类视觉,即一般所说的视觉。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。人类视觉过程可看做是一个复杂的从感觉(感受到的是对 3-D 世界之 2-D 投影得到的图像)到知觉(由 2-D 图像认知 3-D 世界的内容和含义)的过程。



视觉是人们非常熟悉的一种功能，它不仅帮助人们获得信息，而且还帮助人们加工信息 [Thibos 1989]。视觉进一步可分为视感觉和视知觉。感觉是较低层次的，它主要接收外部刺激。知觉则处于较高层次，它要将外部刺激转化为有意义的内容。一般感觉对外部刺激是基本不加区别地完全接收，而知觉则要确定由外界刺激的哪些部分组合成所关心的“目标”。

视感觉主要是从分子的层次和观点来理解人们对光（即可见辐射）反应的基本性质（如亮度、颜色），它主要涉及物理、化学等学科。视感觉中主要研究的内容有：①光的物理特性，如光量子、光波、光谱等；②光刺激视觉感受器官的程度，如光度学、眼睛构造、视觉适应、视觉的强度和灵敏度、视觉的时空特性等；③光作用于视网膜后经视觉系统加工而产生的感觉，如明亮程度、色调等 [郝 1983]。

视知觉主要论述人们从客观世界接收到视觉刺激后如何反应以及反应所采用的方式。它研究如何通过视觉形成人们关于外在世界空间的表象，所以兼有心理因素。视知觉作为对于当前客观事物反映的一种形式，只依靠光投射到视网膜上形成的网像和人们已知的眼或神经系统的机制是难以把全部（知觉）过程解释清楚的。视知觉是在神经中枢进行的一组活动，它把视野中一些分散的刺激加以组织，构成具有一定形状的整体以认识世界。早在两千年前，亚里士多德就定义视知觉的任务是确定“什么东西在什么地方”（What is where） [Finkel 1994]。

视觉的最终目的从狭义上说是能对客观场景做出对观察者有意义的解释和描述，从广义上讲，还包括基于这些解释和描述并根据周围环境和观察者的意愿来制定出行为规划，以作用于周围的世界，这实际上也就是计算机视觉的目标。

## 1.1.2 计算机视觉概述

计算机视觉是指用计算机实现人类的视觉功能，即对客观世界中三维场景的感知、加工和理解。视觉研究的原始目的是把握和理解有关场景的图像，辨识和定位其中的目标，确定它们的结构、空间排列和分布以及目标间的相互关系等。计算机视觉的研究目标是根据感知到的图像对客观世界中实际的目标和场景做出有意义的判断 [Shapiro 2001]。

计算机视觉的研究方法目前主要有两种：一种是仿生学的方法，即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；另一种是工程的方法，即从分析人类视觉过程的功能着手，并不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、可行的手段来实现系统的功能。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，它们互相联系和补充。第一个研究目标是建立计算机视觉系统来完成各种视觉任务。换句话说，要使计算机能借助各种视觉传感器（如 CCD、CMOS 摄像器件等）获取场景的图像，从中感知和恢复 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，进而做出判定和决断。这里主要研究的是完成这些工作的技术机理。目前这方面的工作集中在构建各种专用的系统，完成在各种实际场合中提出的专门视觉任务，而从长远来说则要建成更为通用的系统，完成一般的视觉任务 [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机理的掌握和理解（如计算神经科学）。这里主要研究的是生物学机理。长期以来，人们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，特别是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。需要指出的是，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]。本书主要考虑第一个研究目标。

由此可见,计算机视觉要利用计算机实现人的视觉功能,同时其研究又从人类视觉那里得到了许多启发。计算机视觉方面的许多重要研究都是通过理解人类视觉系统而实现的,典型的例子如用金字塔作为一种有效的数据结构,利用局部朝向的概念,以及滤波技术来检测运动等。另外,借助对人类视觉系统巨大理解能力的研究也可帮助人们开发新的图像理解和计算机视觉算法。

### 1.1.3 相关学科

作为一门学科,计算机视觉与许多学科都有着千丝万缕的联系,特别是与一些相关和相近的学科交融交叉。下面简单介绍几个最接近的学科。

(1) 图像理解。图像理解是图像工程的高层阶段,其重点是在图像分析[章 2005b]的基础上,结合人工智能和认知理论,进一步研究图像中各目标的性质和它们之间的相互联系,并理解图像内容的含义以及解释对应的客观场景,从而指导和规划行动。图像理解与计算机科学有密切的联系(作为图像理解基础的图像处理和图像分析与计算机科学有密切的联系)。图像理解与计算机视觉有相同的目标,都是借助工程技术的手段,通过对客观场景所获得的图像来实现对场景的认识和解释。它们可以看做是专业和背景不同的人习惯使用的不同术语。

(2) 机器视觉。机器视觉/机器人视觉与计算机视觉有着千丝万缕的联系,很多情况下都作为同义词使用。具体地说,一般认为计算机视觉更侧重于对场景分析和图像解释的理论和算法,而机器视觉/机器人视觉则更关注通过视觉传感器获取环境的图像,构建具有视觉感知功能的系统以及实现检测和辨识物体的算法。

(3) 模式识别。模式是指有相似性但又不完全相同的客观事物或现象所构成的类别。模式包含的范围很广,图像就是模式的一种。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为对客观世界中的物体和过程进行分类、描述的学科。目前,对图像模式的识别主要集中在对图像中感兴趣的内容(目标)的分类、分析和描述,在此基础上还可以进一步实现计算机视觉的目标。同时,计算机视觉的研究中也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)并不能把计算机视觉全部包括进去。

(4) 人工智能。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、解决问题等的功能。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的功能的能力和技术。视觉功能是人类智能的一种体现,所以计算机视觉与人工智能密切相关。计算机视觉的研究中使用了许多人工智能技术,反过来,计算机视觉也可看做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。

(5) 计算机图形学。计算机图形学研究如何由给定的描述生成“图像”,它与计算机视觉也有密切的关系。一般人们将计算机图形学称为计算机视觉的反/逆(inverse)问题,因为视觉从2-D图像提取3-D信息,而图形学使用3-D模型来生成2-D场景图像。需要注意的是,与计算机视觉中存在许多不确定性相比,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。在许多实际应用中,人们更多关心的是图形生成的速度和精度,即在实时性和逼真度之间取得某种妥协。

除以上相近学科外,从更广泛的领域看,计算机视觉要借助工程方法解决一些生物的问题,完成生物固有的功能,所以它与生物学、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。近年计算机视觉研究者与视觉心理、生理研究者紧密结合,已获得了一系列研究成果。计算机视觉属于工程应用科学,与工业自动化、人机交互、办公自动化、视觉导航和机器人、

安全监控、生物医学、遥感测绘、智能交通和军事公安等学科也密不可分。一方面，计算机视觉的研究充分利用了这些学科的成果；另一方面，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。

### 1.1.4 应用领域

近年来计算机视觉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、办公自动化、邮政自动化、邮件分检、金相分析、无损探测、印刷板质量检验、精细印刷品缺陷检测以及在各种危险场合工作的机器人等。将视觉技术用于工业生产自动化，可以加快生产速度，保证质量的一致性，还可以避免由于人的疲劳、注意力不集中等产生的误判。

(2) 人机交互：让计算机借助人的手势动作（手语）、嘴唇动作（唇读）、躯干运动（步态）、人脸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，也可增加交互的方便性和临场感等。

(3) 安全监控：如支票、签名辨伪，人脸识别，罪犯脸形的合成、识别和查询，指纹、印章的鉴定和识别等，可有效地监测和防止许多类型的犯罪。

(4) 军事公安：如军事侦察、合成孔径雷达图像分析、战场环境/场景建模表示等。

(5) 遥感测绘：如矿藏勘探、资源探测、气象预报、自然灾害监测监控等。

(6) 视觉导航：如太空探测、航天飞行、巡航导弹制导、无人驾驶飞机飞行、自动行驶车辆的安全操纵、移动机器人、精确制导、公路交通管理以及智能交通的各个方面等，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。

(7) 生物医学：红白血球计数，染色体分析，各类 X 光、CT、MRI、PET 图像的自动分析，显微医学操作，远程医疗，计算机辅助外科手术等。

(8) 虚拟现实：如飞机驾驶员训练、医学手术模拟、战场环境建模表示等，可帮助人们超越人的生理极限，产生身临其境的感觉，提高工作效率。

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

(10) 对人类视觉系统和机理，以及人脑心理和生理的研究等。这对人们理解人类视觉系统，并推动相关的发展起到了积极作用。

## 1.2 图像基础

计算机视觉是利用计算机对从客观世界采集的图像进行加工来实现视觉功能的，所以图像技术在其中起着重要的作用。本节先概括介绍一些有关图像的基本概念，然后简单介绍用计算机加工图像的系统都具有的两个模块，即存储和显示。另一个重要的模块是图像采集，将在第 3 章专门介绍。另外各种加工模块将在本书后续的章节分别进行具体介绍。

### 1.2.1 图像

图像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,进而产生视知觉的实体(参见文献[章 1996a])。人的视觉系统就是一个观测系统,通过它得到的图像就是客观事物在人心目中形成的影像。下面给出各种图像的简单定义。

### 1. 图像种类

图像可看做是对辐射(光是典型的示例)强度模式的空间分布的一种表示,是将空间辐射强度模式进行投影得到的。将3-D空间投影得到的2-D成像平面称为**图像平面**,简称**像平面**。

表示图像所反映的辐射能量在空间分布情况的函数称为**图像表达函数**。这种分布函数在广义上可以是5个变量的函数,即 $T(x, y, z, t, \lambda)$ ,其中 $x, y, z$ 是空间变量, $t$ 是时间变量, $\lambda$ 是辐射的波长(对应频谱变量)。这种情况下也称为**通用图像表达函数**。由于实际图像在时空上都是有限的,所以 $T(x, y, z, t, \lambda)$ 是一个5-D有限函数。需要注意的是这个函数中的 $T$ 可以是标量,也可以是矢量。

如果考虑上述5个变量的不同变化,图像的种类和形式是很多的。例如,考虑辐射波长的不同,有 **$\gamma$ 射线图像**(利用波长约在0.001~1 nm之间的伽马光所获得的图像)、**X射线图像**(利用波长约在1~10 nm之间的X光所获得的图像)、**紫外线图像**(利用波长约在10~380 nm之间的紫外光所获得的图像)、**可见光图像**(利用波长约在380~780 nm之间,人类视觉可感知到的光辐射作为辐射能源所产生的光所获得的图像)、**红外线图像**(利用波长约在780~1 500 nm之间的红外光所获得的图像)、**微波图像**(利用波长约在1 mm~10 m之间的微波所获得的图像)、**无线电波图像**(利用波长约在10 m~50 km之间的无线电波所获得的图像)、**交流电波图像**(利用波长约在100 km以上的交流电波所获得的图像)等。

再如考虑图像的类型,也有许多变型。相对于普通的**2-D图像**,还有**3-D图像**、**彩色图像**、**多光谱图像**(包含多个频谱区间的一组图像,每幅图像对应一个频谱区间,如多数遥感图像)、**立体图像**和**多视图图像**(包含由不同位置和朝向的多个摄像机获得的同一场景的一组图像)等;相对于**静止图像**或**单幅图像**,还有**序列图像**或**图像序列**(时间上有一定顺序和间隔,内容上相关的一组图像),**视频图像**(简称**视频**,是一种特殊的序列图像或图像序列)等;相对于常见的反映辐射量强度的**灰度图像**,还有反映场景中景物与摄像机间距离信息的(深度值)**深度图像**、反映景物表面纹理特性和纹理变化的**纹理图像**、反映景物对物质吸收值的**投影重建图像**(利用投影重建原理获得的图像)等。投影重建图像实际上是一个大类,最常见的是**计算机断层扫描图像**,其中又可分为**发射断层图像**(emission computed tomography image, CT)、**正电子发射图像**(positron emission image, PET)和**单光子发射图像**(single photon emission image, SPECT)。另外,**磁共振图像**(magnetic resonance image, MRI)和**雷达图像**(合成孔径图像就是一种常用的雷达图像)都是利用投影重建原理获得的。

### 2. 模拟图像

人从连续的客观场景直接观察到的图像是空间连续和幅度连续的**模拟图像**,也称**连续图像**。

客观世界在空间上是三维(3-D)的,但一般从客观景物得到的图像是二维(2-D)的。一幅图像可以用一个2-D数组 $f(x, y)$ 来表示,这里的 $x$ 和 $y$ 表示2-D空间 $XY$ 中一个坐标点的位置(实际图像的尺寸是有限的,所以 $x$ 和 $y$ 的取值也是有限的),而 $f$ 则代表图像在点 $(x, y)$ 的某种性质空间 $F$ 的数值(实际图像中各个位置上所具有的性质取值也是有限的,所以 $f$ 的取值也是有限的)。例如,常用的图像一般是**灰度图像**,这时 $f$ 表示**灰度值**,它常对应客观景物被观察到的亮度。图像在点 $(x, y)$ 也可有多种性质,此时可用**矢量 $f$** 来表示。例如一幅彩色图像在每一个图像点同时具有红、绿、蓝3个值,可记为 $[f_r(x, y), f_g(x, y), f_b(x, y)]$ 。需要指出的是,人们总是根据图像内不同位置的不同性质来利用图像的。

以上讨论中认为图像在空间或性质上都可以是连续的。事实上，日常所见的自然界图像有许多是连续的，即  $f, x, y$  的值可以是任意实数。

### 3. 数字图像

为了能用计算机对图像进行加工，需要把连续的模拟图像在坐标空间  $XY$  和性质空间  $F$  都离散化。这种空间和幅度都离散化了的图像是数字图像，是客观事物的可视数字化表达。数字图像可以用  $I(r, c)$  来表示，其中  $I, c, r$  的值都是整数。这里  $I$  代表离散化后的  $f$ ， $(r, c)$  代表离散化后的  $(x, y)$ ，其中  $r$  代表图像的行 (row)， $c$  代表图像的列 (column)。本书以后主要讨论计算机视觉，在不至引起混淆的情况下仍用  $f(x, y)$  代表数字图像，这里如不作特别说明， $f, x, y$  都在整数集合中取值。例如，文本图像常表示为二值图像，即此时  $f$  的取值只有两个：0 和 1。另外，计算机中的图像都是数字化了的图像，所以如不作特别说明，本书以后就用图像代表数字图像。

早期英文书籍里一般用 picture 来指图像，随着数字技术的发展，现在都用 image 代表离散化了的数字图像，因为“计算机存储人像或场景的数字图像 (computers store numerical images of a picture or scene)” (参见文献[Zhang 1996])。这样看来，应该使用“数字图像”而不是“数字图像”。事实上，“图像”一词比“图像”一词的含义更广，覆盖面更宽。

## 1.2.2 图像表达和显示

根据应用领域的不同，可以有多种不同的方法来表达和表示图像，或将图像以一定的形式显示出来。图像表达是图像显示的基础，而图像显示是计算机视觉系统的重要模块之一。

要对图像进行表达和显示，需要对图像的各个单元进行表达和显示。图像中的每个基本单元叫作图像元素，在早期用 picture 表示图像时称为像素 (picture element)。对 2-D 图像，英文里常用 pixel 代表像素。如果采集一系列的 2-D 图像或利用一些特殊设备还可得到 3-D 图像。对 3-D 图像，英文里常用 voxel 代表其基本单元，简称体素 (volume element)。近年来由于都用 image 代表图像，所以也有人建议用 imel 统一代表像素和体素。

### 1. 图像表达

前面提到，一幅 2-D 图像可以用一个 2-D 数组  $f(x, y)$  来表示。实际中还常将一幅 2-D 图像写成一个 2-D 的  $M \times N$  矩阵 (其中  $M$  和  $N$  分别为图像的总行数和总列数)：

$$F = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & f_{2N} \\ \vdots & \vdots & & \vdots \\ f_{M1} & f_{M2} & \cdots & f_{MN} \end{bmatrix} \quad (1.2.1)$$

式 (1.2.1) 就是图像的矩阵表达形式，矩阵中的每个元素对应一个像素。

一幅 2-D 图像也可以用矢量来表示，一般写成：

$$F = [f_1 \quad f_2 \quad \cdots \quad f_N] \quad (1.2.2)$$

其中

$$f_i = [f_{1i} \quad f_{2i} \quad \cdots \quad f_{Mi}]^T \quad (i=1, 2, \cdots, N) \quad (1.2.3)$$

式 (1.2.3) 就是图像的矢量表达形式，各个矢量中的每个元素也对应一个像素。

图像的矢量表达和矩阵表达是图像的两种常用表示形式，它们之间可以方便地互相转换。

### 2. 图像显示设备

计算机视觉系统中使用计算机对图像进行加工，加工后的结果通常仍是图像，且主要用于显

示给人看，所以图像显示对计算机视觉来说是非常重要的。将图像显示出来是计算机视觉系统与用户交流中的重要步骤。

常用的计算机视觉系统的显示设备主要包括可以随机存取的阴极射线管（cathode ray tube, CRT），电视显示器（TV monitor）和液晶显示器（liquid crystal display, LCD）。在 CRT 中，电子枪束的水平、垂直位置可由计算机控制。在每个偏转位置，电子枪束的强度是用电压来调制的。每个点的电压都与该点所对应的灰度值成正比，这样灰度图就转化为光亮度变化的模式，这个模式被记录在阴极射线管的屏幕上。输入显示器的图像也可以通过硬拷贝转换到幻灯片、照片或透明胶片上。

除了显示器，各种打印设备如各种打印机（printer）也可看做是图像显示设备。打印设备一般用于输出较低分辨率的图像。早年在纸上打印灰度图像的一种简便方法是利用标准行打印机的重复打印能力。输出图像上任一点的灰度值可由该点打印的字符数量和密度来控制。近年来使用的各种热敏、热升华、喷墨和激光打印机等具有更高的能力，可打印具有较高分辨率的图像。

### 3. 表达和显示方式

图像的表达和显示是密切相关的，图像显示是对图像的可视表达方式。对 2-D 图像的显示可以采取多种形式，其基本思路是将 2-D 图像看做在 2-D 空间中的一种幅度分布。根据图像的不同，采取的显示方式也不同，如对二值图像，在每个空间位置的取值只有两个，可用黑白来区别，也可用 0 和 1 来区别。

图 1.2.1 所示为对同一幅 2-D 的二值图像的 3 种不同的显示方式。在图像表达的数学模型中，一个像素区域常用其中心来表示，基于这些中心的表达形式就是将图像显示成平面上的离散点集，对应于图 1.2.1 (a)。如果将像素区域用其所覆盖的区域来表示，就得到图 1.2.1 (b) 所示的图形。把幅度值标在图像中相应的位置，就得到图 1.2.1 (c) 所示的类似矩阵表达的结果。用图 1.2.1 (b) 所示的形式也可表示有多个灰度的图像，此时需要用不同深浅的色调表示不同的灰度。用图 1.2.1 (c) 所示的形式也可表示有多个灰度的图像，此时将不同灰度用不同的数值表示。

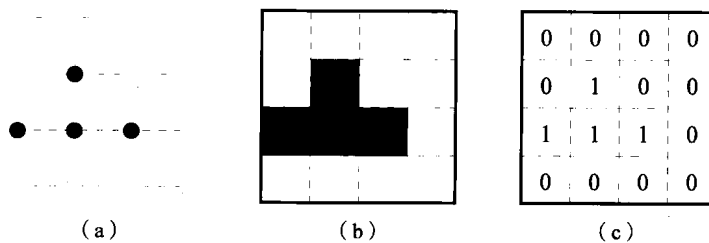


图 1.2.1 3 种表达同一幅  $4 \times 4$  的二值图像的方式

#### 例 1.2.1 灰度图像显示示例。

图 1.2.2 所示为两幅典型的灰度图像，它们都是常用的公开图像，或称标准图像。图 1.2.2 (a) 所用的坐标系常在屏幕显示中采用（屏幕扫描是从左向右，从上向下进行的），它的原点（origin） $O$  在图像的左上角，纵轴标记图像的行，横轴标记图像的列。 $f(x, y)$  既可代表这幅图像，也可表示在  $(x, y)$  行列交点处的图像值。图 1.2.2 (b) 所用的坐标系常在图像计算中采用，它的原点在图像的左下角，横轴为  $X$  轴，纵轴为  $Y$  轴（与常用的笛卡儿坐标系相同）。同样， $f(x, y)$  既可代表这幅图像，也可表示在  $(x, y)$  坐标处像素的值。

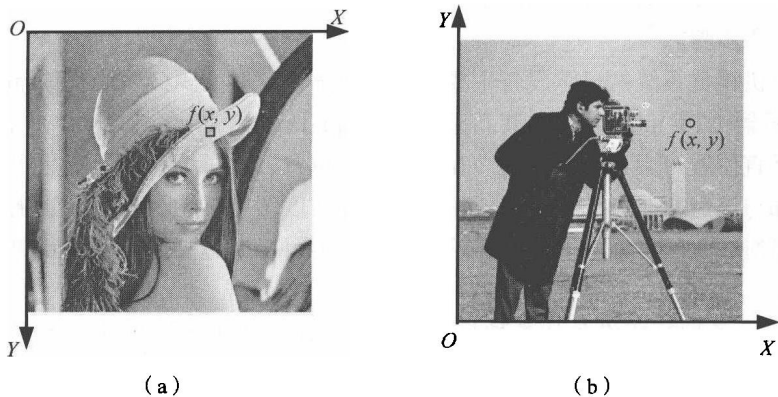


图 1.2.2 标准图像显示示例

### 1.2.3 图像存储

要将采集的图像输入计算机系统，需要有相应的图像存储器。在对图像的加工中，也常需要存储其中间结果和最终结果（很多为图像形式）。所以图像存储也是计算机视觉系统的重要模块之一。

#### 1. 图像存储器

图像包含有大量的信息，因而存储图像也需要大量的空间。在图像处理和分析系统中，大容量和快速的图像存储器是必不可少的。在计算机中，图像数据最小的量度单位是比特（bit）。存储器的存储量常用字节（1 byte = 8 bit）、千字节（k byte）、兆（ $10^6$ ）字节（M byte）、吉（ $10^9$ ）字节（G byte）、太（ $10^{12}$ ）字节（T byte）等表示。例如，存储一幅 1 024 像素×1 024 像素的 8 比特图像需要 1 兆字节的存储器。

图像存储器包括磁带（magnetic tape）、磁盘（magnetic disk）、闪存存储器（flash memory）、光盘（optical disk）、磁光盘（magneto-optical disk）等。用于图像处理的存储器可分为如下 3 类。

- （1）处理过程中使用的快速存储器。
- （2）可以比较快地重新调用的在线或联机存储器。
- （3）不经常使用的数据库（档案库）存储器。

计算机内存就是一种提供快速存储功能的存储器。目前一般微型计算机的内存常为几百到几千兆字节。另一种提供快速存储功能的存储器是特制的硬件卡，也叫帧缓存。它可存储多幅图像，并可以视频速度（每秒 25 或 30 幅图像）读取。它也可以允许对图像进行放大和缩小，以及垂直翻转和水平翻转。目前常用的帧缓存容量可达几吉字节。

磁盘是比较通用的在线存储器，常用的 Winchester 磁盘可存储上百吉字节的数据。另外还有磁光（MO）存储器，它可在 5/4 英寸的光片上存储上吉字节的数据。在线存储器的一个特点是需要经常读取数据，所以一般不采用磁带一类的顺序介质。对于更大的存储要求，还可以使用光盘塔和光盘阵列，一个光盘塔可放几十张到几百张光盘，利用机械装置插入或从光盘驱动器中抽取光盘。

数据库存储器的特点是要求非常大的容量，但对数据的读取不太频繁。这里常用磁带和光盘。长 13 英尺的磁带可存储达到吉字节的数据，但磁带的储藏寿命较短，在控制很好的环境中也只有 7 年。一般常用的一次写多次读（write-once-read-many, WORM）光盘可在 12 英寸的光盘上存储

6 吉字节的数据, 在 14 英寸的光盘上存储 10 吉字节的数据。另外, WORM 光盘在一般环境下可储藏 30 年以上。在主要是读取的应用中, 也可将 WORM 光盘放在光盘塔中。一个存储量达到太字节级的 WORM 光盘塔可存储上百万幅百万像素的灰度和彩色图像。

## 2. 图像文件格式

在计算机系统中, 图像常以文件形式存储。图像文件指包含图像数据的文件, 文件内除图像数据本身以外, 一般还有对图像的描述信息等, 以方便读取、显示图像。

表示图像常使用两种不同的方式, 一种是矢量表示形式, 另一种是光栅 (也称位图或像素图) 表示形式。图像文件可以采用任一种形式, 也可以结合使用两种形式。

在**矢量表示形式**中, 图像是用一系列线段或线段的组合体来表示的, 线段的灰度 (色度) 可以是均匀的或变化的, 在线段的组合体中各部分也可用不同灰度填充。矢量文件就像程序文件, 里面有一系列命令和数据, 执行这些命令就可根据数据画出图案。矢量文件主要用于图形数据文件。

图像数据文件主要使用**光栅表示形式**, 该形式与人对图像的理解一致 (一幅图像是许多图像点的集合), 比较适合色彩、阴影或形状变化复杂的图像。它的主要缺点是缺少对像素间相互关系的直接表示, 且限定了图像的空间分辨率。后者带来两个问题, 一个是将图像放大到一定程度就会出现方块效应, 另一个是如果将图像缩小再恢复到原尺寸, 则图像会变得比较模糊。

图像数据文件的格式有很多种, 不同的系统平台和软件常使用不同的**图像文件格式**。例如, Macintosh 机上普遍使用 MacPaint 格式 (固定大小, 宽为 576 像素, 高为 720 像素), PC Paintbrush 支持 PCX 格式 (包括单色、16 色、256 色), Digital Research (现 Novell) 支持 GEM IMG 格式, Sun Microsystems 支持 Sun 光栅格式等 (参见文献[凯 1994])。

下面简单介绍 4 种应用比较广泛的图像文件格式 (参见[董 1994]和[凯 1994])。

(1) **BMP 格式**。**BMP 格式**是 Windows 环境中的一种标准 (但很多 Macintosh 应用程序不支持它), 它的全称是 Microsoft 设备独立位图 (device independent bitmap, DIB)。BMP 图像文件也称位图文件, 包括 3 部分: ①位图文件头 (也称表头); ②位图信息 (常称调色板); ③位图阵列 (即图像数据)。一个位图文件只能存放一幅图像。

位图文件头长度固定为 54 个字节, 它给出图像文件的类型、大小和位图阵列的起始位置等信息。位图信息给出图像的长和宽、每个像素的位数 (可以是 1、4、8 和 24, 分别对应单色、16 色、256 色和真彩色的情况)、压缩方法、目标设备的水平和垂直分辨率等信息。位图阵列给出原始图像里每个像素的值 (每 3 个字节表示一个像素, 分别是蓝、绿、红的值), 它的存储格式有压缩 (仅用于 16 色和 256 色图像) 和非压缩两种。

(2) **GIF 格式**。**GIF 格式**是一种公用的图像文件格式标准, 它是 8 位文件格式 (一个像素一个字节), 所以最多只能存储 256 色图像。GIF 文件中的图像数据均为压缩过的。

GIF 文件结构较复杂, 一般包括 7 个数据单元: 文件头, 通用调色板, 图像数据区, 以及 4 个补充区。其中, 表头和图像数据区是不可缺少的单元。

一个 GIF 文件中可以存放多幅图像 (这个特点对实现网页上的动画很有利), 所以文件头中包含适用于所有图像的全局数据和仅属于其后那幅图像的局部数据 (参见文献[戴 2002])。当文件中只有一幅图像时, 全局数据和局部数据一致。存放多幅图像时, 每幅图像集中成一个图像数据块, 每块的第一个字节是标识符, 指示数据块的类型 (可以是图像块、扩展块或文件结束符)。

(3) **TIFF 格式**。**TIFF 格式**是一种独立于操作系统和文件系统的格式 (在 Windows 环境和 Macintosh 机上都可使用), 便于在软件之间进行图像数据交换。TIFF 图像文件包括文件头 (表头)、



文件目录（标识信息区）和文件目录项（图像数据区）。文件头只有一个，且在文件前端。它给出数据存放顺序、文件目录的字节偏移信息。文件目录给出文件目录项的个数信息，并有一组标识信息，给出图像数据区的地址。文件目录项是存放信息的基本单位，也称为域。从类别上讲，域主要包括基本域、信息描述域、传真域、文献存储和检索域 5 类。

TIFF 格式的描述能力很强，可制定私人用的标识信息。TIFF 格式支持任意大小的图像，文件可分为 4 类：二值图像、灰度图像、调色板彩色图像和全彩色图像。一个 TIFF 文件中可以存放多幅图像，也可存放多份调色板数据。

(4) JPEG 格式。JPEG 格式源自对静止灰度或彩色图像的一种压缩标准 JPEG（参见文献 [章 2006b]），在使用有损压缩方式时，可节省的空间是相当大的，目前数码相机中均使用这种格式。

JPEG 标准只是定义了一个规范的编码数据流，并没有规定图像数据文件的格式。Cube Microsystems 公司定义了一种 JPEG 文件交换格式（JPEG file interchange format, JFIF）。JFIF 图像是一种使用灰度表示或使用  $Y, C_b, C_r$  分量彩色表示的 JPEG 图像。它包含一个与 JPEG 兼容的文件头。一个 JFIF 文件通常包含单个图像，图像可以是灰度的（其中的数据为单个分量），也可以是彩色的（其中的数据是  $Y, C_b, C_r$  分量）。 $Y, C_b, C_r$  分量与常见的  $R, G, B$  三原色的关系如下。

$$Y = 0.299R + 0.587G + 0.114B \quad (1.2.4)$$

$$C_b = 0.1687R - 0.3313G + 0.5B \quad (1.2.5)$$

$$C_r = 0.5R - 0.4187G - 0.0813B \quad (1.2.6)$$

TIFF 6.0 也支持用 JPEG 压缩的图像，TIFF 文件可以包含直接 DCT 的图像，也可以包含无损 JPEG 图像，还可以包含用 JPEG 编码的条或块的系列（这样允许只恢复图像的局部而不用读取全部内容）。

## 1.3 像素间联系

图像是一个整体概念，它本身可以分解为更小的单元。如果把一幅图像看做一个集合，那么其中的每个子集就可看做是一个子图像。子图像可大可小，一般包含一组像素，它们常组合成空间上接近的团点/团块。像素常被看做图像中的基本单元，但在有些应用中还需要考虑子像素。

一幅图像包含大量的像素，这些像素在空间是按某种规律排列的，相互之间有一定的关系。事实上，一个具有  $1000 \times 1000$  单元的图像传感器矩阵只有  $10^{-3}$  的相对分辨率。这个数值与其他测量如长度、电压、频率等相比都比较小，那些测量的相对分辨率可以远高于  $10^{-6}$ 。但是，这些测量如长度、电压、频率等技术所提供的仅是对一个点的测量，而一幅  $1000$  像素  $\times$   $1000$  像素的图像包含一百万个单元，即它是对一百万个点的测量结果。因此，图像不仅给出了空间点的信息，同时还给出了空间中变化的信息。如果采集图像序列，那么时间变化（即需研究目标的动态信息）也可以获得。另外，当空间变化是 3-D 时，那么与此对应的图像已经是 3-D 的，如再加上时间变化就成为 4-D 的了。由此可见，图像的灰度（或其他属性值）同时表达了许多时空位置及其分布的信息，而其他物理量仅反映了某一个时空位置的信息。图像的这种能力与像素间联系密切相关，所以要对图像进行有效的加工，必须考虑像素之间的联系。