

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 (第2版)

A Course of Computer Vision (2nd Edition)

章毓晋 编著

- 选材全面、原理清晰、方法详尽、技术实用
- 重点突出、解释直观、定义精练、示例丰富
- 内容选择灵活、习题部分解答、术语提供索引



名家系列



中国工信出版集团



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 (第2版)

A Course of Computer Vision (2nd Edition)

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. — 2版. — 北京 : 人民邮电出版社, 2017.3
21世纪高等学校计算机规划教材. 名家系列
ISBN 978-7-115-44154-6

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(2016)第296815号

(双S策) 计算机视觉

内 容 提 要

本书系统地介绍了计算机视觉的基本原理、典型方法和实用技术, 内容包括绪论、图像采集、图像预处理、基元检测、目标分割、目标表达和描述、纹理分析、形状分析、立体视觉、三维景物恢复、运动分析、景物识别、广义匹配、时空行为理解、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。本书提供了许多例题, 每章均有要点小结、参考文献介绍和练习题(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)的科技工作者自学及科研参考。

◆ 编 著 章毓晋

责任编辑 曾 斌

执行编辑 刘 尉

责任印制 杨林杰

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

三河市中晟雅豪印务有限公司印刷

◆ 开本: 787×1092 1/16

印张: 22.25

2017年3月第2版

字数: 588千字

2017年3月河北第1次印刷

定价: 65.00元

读者服务热线: (010)81055256 印装质量热线: (010)81055316

反盗版热线: (010)81055315



人民邮电

前言

本书为《计算机视觉教程》的第2版，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，旨在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。

本次再版仍然保持了原版的基本特点和风格，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。本书没有过多地强调理论性，尽量减少了公式推导，对先修知识也没有特别要求。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，读者基本上不再需要去参考本书中的其他部分。本书在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。本书在书后给出了术语索引（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，也方便读者在网上搜索相关资料。

本次再版在内容上有一些增删调整，取消了原第2章（视觉和视知觉），增加了新的第7章（纹理分析）和新的第14章（时空行为理解）。另外还增加了哈里斯兴趣点算子、椭圆定位和检测、特征空间聚类等节，以及边界细化、轮廓形状矩阵、背景建模、运动目标跟踪、知识库为根的树结构等小节。

本次再版增加了许多例题，内容也比较多样，其中有些是介绍一些新的选学内容，有些是从其他的角度来解释本书中已介绍的概念和方法，也有些是把原有一些细节介绍内容改为例题。这些例题多数可根据课时安排、学生基础情况等选择使用，比较灵活。相比上一版，练习题仍保持为每章12题，其中有一些是新增的，对一些保留的也进行了一些改动，并且把提供参考答案的题目数量限制在每章2题，以供教师布置作业时有更多的选择。本次再版对参考文献进行了全面更新，大量增加了21世纪发表的新文献，大幅削减了20世纪发表的文献，还尽量对有新版的书籍进行了版本更新。

本书如果用于课堂教学，可考虑每一次课讲一章，用于约一个学期的教学。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（或将例题留到课外自选学习）。

本书从结构上看，共有15章正文，以及部分练习题解答、参考文献和索引。在这18个一级标题下共有79个二级标题（节），再之下还有127个三级标题（小节）。全书（包括图片、绘图、表格、公式等）折合文字近60万字，有编了号的图345个、表格35个、公式712个。为便于教学和理解，本书共给出各类例题145个，思考题和练习题168个，对其中的24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近 200 篇参考文献的目录和用于索引的近 700 个术语。

感谢人民邮电出版社编辑的精心组稿、认真审阅和细心修改。

最后,感谢妻子何芸、女儿章荷铭等家人在各方面的理解和支持。

章毓晋

2016 年暑假于蓝旗营

通信:北京清华大学电子工程系,100084

电话:(010) 62798540

传真:(010) 62770317

邮箱:zhang-yj@tsinghua.edu.cn

主页:oa.ee.tsinghua.edu.cn/~zhangyujin/

目 录

第 1 章 绪论	1	3.1.1 基本坐标变换.....	43
1.1 计算机视觉.....	1	3.1.2 几何失真校正.....	45
1.1.1 视觉概述.....	1	3.2 灰度映射.....	47
1.1.2 计算机视觉的目标.....	2	3.2.1 灰度映射原理.....	47
1.1.3 相关学科.....	3	3.2.2 灰度映射示例.....	48
1.1.4 应用领域.....	4	3.3 直方图修正.....	49
1.2 图像基础.....	5	3.3.1 直方图均衡化.....	50
1.2.1 图像及类别.....	5	3.3.2 直方图规定化.....	52
1.2.2 图像表达和显示.....	7	3.4 空域滤波.....	55
1.2.3 图像存储.....	8	3.4.1 原理和分类.....	55
1.3 像素间联系.....	11	3.4.2 线性平滑滤波.....	57
1.3.1 像素邻域.....	11	3.4.3 线性锐化滤波.....	58
1.3.2 像素间距离.....	12	3.4.4 非线性平滑滤波.....	59
1.4 本书内容提要.....	15	3.4.5 非线性锐化滤波.....	62
1.4.1 计算机视觉系统及模块.....	15	总结和复习.....	63
1.4.2 如何学习使用本书.....	16	第 4 章 基元检测	65
总结和复习.....	19	4.1 边缘检测.....	65
第 2 章 图像采集	21	4.1.1 检测原理.....	66
2.1 采集装置.....	21	4.1.2 一阶导数算子.....	67
2.2 采集模型.....	23	4.1.3 二阶导数算子.....	70
2.2.1 几何成像模型.....	23	4.1.4 边界闭合.....	74
2.2.2 亮度成像模型.....	29	4.1.5 边界细化.....	75
2.2.3 空间和幅度分辨率.....	31	4.2 SUSAN 算子.....	76
2.3 采集方式.....	32	4.2.1 USAN 原理.....	76
2.3.1 成像方式一览.....	33	4.2.2 角点和边缘检测.....	77
2.3.2 结构光法.....	33	4.3 哈里斯兴趣点算子.....	80
2.4 摄像机标定.....	35	4.4 哈夫变换.....	82
2.4.1 标定程序和步骤.....	35	4.4.1 基本哈夫变换.....	82
2.4.2 两级标定法.....	38	4.4.2 广义哈夫变换.....	85
总结和复习.....	41	4.4.3 完整广义哈夫变换.....	87
第 3 章 图像预处理	43	4.5 椭圆定位和检测.....	88
3.1 坐标变换.....	43	4.6 位置直方图技术.....	90

总结和复习	92	6.4.2 区域形状数	128
第5章 目标分割	94	6.4.3 区域不变矩	129
5.1 轮廓搜索	94	6.4.4 拓扑描述符	131
5.1.1 图搜索	95	总结和复习	132
5.1.2 动态规划	96	第7章 纹理分析	134
5.2 主动轮廓模型	97	7.1 统计描述方法	134
5.2.1 主动轮廓	97	7.1.1 灰度共生矩阵	135
5.2.2 能量函数	98	7.1.2 基于共生矩阵的描述	136
5.3 基本阈值技术	100	7.1.3 基于能量的描述	137
5.3.1 原理和分类	100	7.2 结构描述方法	139
5.3.2 全局阈值的选取	102	7.2.1 结构描述原理	139
5.3.3 局部阈值的选取	104	7.2.2 纹理镶嵌	141
5.3.4 动态阈值的选取	106	7.2.3 局部二值模式	142
5.4 特色阈值方法	107	7.3 频谱描述方法	144
5.4.1 多分辨率阈值	107	7.3.1 傅里叶频谱描述	144
5.4.2 过渡区阈值	109	7.3.2 盖伯频谱描述	145
5.5 特征空间聚类	111	7.4 纹理图像分割	147
5.5.1 基本聚类方法	111	7.4.1 有监督纹理分割	148
5.5.2 均移确定聚类中心	112	7.4.2 无监督纹理分割	150
总结和复习	113	总结和复习	152
第6章 目标表达和描述	115	第8章 形状分析	154
6.1 基于边界的表达	115	8.1 形状紧凑性描述符	155
6.1.1 链码	116	8.2 形状复杂性描述符	160
6.1.2 边界段和凸包	117	8.3 基于多边形的形状分析	163
6.1.3 边界标记	118	8.3.1 多边形计算	163
6.2 基于区域的表达	120	8.3.2 多边形描述	164
6.2.1 四叉树	120	8.4 基于曲率的形状分析	166
6.2.2 围绕区域	121	8.4.1 轮廓曲率	166
6.2.3 骨架	122	8.4.2 曲面曲率	169
6.3 基于边界的描述	124	总结和复习	170
6.3.1 边界长度和直径	124	第9章 立体视觉	172
6.3.2 边界形状数	125	9.1 立体视觉模块	172
6.3.3 轮廓形状矩阵	126	9.2 双目成像和视差	174
6.4 基于区域的描述	127	9.2.1 双目横向模式	174
6.4.1 区域面积和密度	127	9.2.2 双目横向会聚模式	176

9.2.3 双目纵向模式.....	177	总结和复习.....	238
9.3 基于区域的立体匹配.....	178	第 12 章 景物识别	240
9.3.1 模板匹配.....	179	12.1 统计模式分类.....	240
9.3.2 双目立体匹配.....	180	12.1.1 模式分类原理.....	241
9.4 基于特征的立体匹配.....	186	12.1.2 最小距离分类器.....	241
9.4.1 点对点的方法.....	186	12.1.3 最优统计分类器.....	242
9.4.2 动态规划匹配.....	188	12.2 感知机.....	246
总结和复习.....	189	12.3 支持向量机.....	249
第 10 章 三维景物恢复	192	12.4 结构模式识别.....	252
10.1 由光移恢复表面朝向.....	193	12.4.1 字符串结构识别.....	252
10.1.1 表面反射特性.....	193	12.4.2 树结构识别.....	255
10.1.2 目标表面朝向.....	196	总结和复习.....	257
10.1.3 反射图.....	196	第 13 章 广义匹配	259
10.1.4 光度立体学求解.....	198	13.1 目标匹配.....	260
10.2 从影调获取形状信息.....	199	13.1.1 匹配的度量.....	260
10.2.1 影调与形状.....	199	13.1.2 字符串匹配.....	263
10.2.2 求解亮度方程.....	202	13.1.3 惯量等效椭圆匹配.....	264
10.3 纹理变化与表面朝向.....	204	13.2 动态模式匹配.....	266
10.3.1 三种典型变化.....	204	13.3 关系匹配.....	269
10.3.2 确定线段的纹理消失点.....	206	13.3.1 关系表达和距离.....	269
10.4 根据焦距确定深度.....	209	13.3.2 关系匹配模型.....	270
总结和复习.....	210	13.4 图同构匹配.....	272
第 11 章 运动分析	212	13.4.1 图论基础.....	272
11.1 运动分类和表达.....	213	13.4.2 图同构和匹配.....	275
11.2 全局运动检测.....	216	总结和复习.....	276
11.2.1 利用图像差的运动检测.....	216	第 14 章 时空行为理解	279
11.2.2 基于模型的运动检测.....	219	14.1 时空技术.....	279
11.3 运动目标检测和分割.....	221	14.2 时空兴趣点.....	281
11.3.1 背景建模.....	221	14.3 动态轨迹学习和分析.....	283
11.3.2 运动目标跟踪.....	225	14.3.1 自动场景建模.....	283
11.3.3 运动目标分割.....	229	14.3.2 路径学习.....	285
11.4 运动光流和表面取向.....	230	14.3.3 自动活动分析.....	287
11.4.1 光流约束方程.....	230	14.4 动作分类和识别.....	288
11.4.2 光流计算.....	231	14.4.1 动作分类.....	288
11.4.3 光流与表面取向.....	235	14.4.2 动作识别.....	290

14.5	活动和行为建模	293	15.3.3	知识库为根的树结构	313
14.5.1	动作建模	294	15.3.4	多模块交叉配合结构	314
14.5.2	活动建模和识别	297	15.4	计算机视觉理论框架	315
	总结和复习	301	15.4.1	马尔视觉计算理论	315
第 15 章	场景解释	303	15.4.2	对马尔理论框架的改进	318
15.1	线条图标记解释	303	15.4.3	新理论框架的研究	319
15.2	体育比赛视频排序	306		总结和复习	321
15.3	计算机视觉系统模型	311	部分练习题解答	323	
15.3.1	多层次串行结构	311	参考文献	329	
15.3.2	知识库为中心的辐射结构	312	索引	338	

第 1 章

绪论

视觉是人类观察世界、认知世界的重要功能手段。计算机技术的发展,使得计算机视觉——使用计算机实现视觉功能——进一步得到广泛关注和研究,理论和实践都越来越成熟。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,此次再版对原有内容根据技术发展进行了补充和更新。

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,要对计算机视觉学科的总体内容和范围给予概括描述。计算机视觉以对图像^①的各种加工为技术基础,所以本章首先对密切相关的基本图像概念和相关知识给予具体介绍,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。

根据上面所述,本章各节将如下安排。

1.1 节对计算机视觉给予概括介绍,从密切对应联系的人类视觉知识,到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各种联系,包括与像素邻域相关的几个概念、像素之间距离的计算,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整体框架的描述,列出了本书的主要内容、结构和安排,对各章的重要技术给予了概括介绍,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 视觉概述

计算机视觉源自人类视觉,即一般所说的视觉。视觉在人类对客观世界的观察和认知中起重

注:20世纪出版的中文书籍很多写为“图象”。“象”有象征、抽象、形象等含义,所以“图象”一词比“图像”一词的含义更广泛、覆盖面更宽大,不单指人像。而《现代汉语词典(第7版)》以“图像”作为标准使用。

要的作用。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。视觉进一步可分为视感觉和视知觉。这里, 感觉是较低层次的, 它主要接收外部刺激; 而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。

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视网膜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 2000 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规划, 并作用于周围的世界, 这实际上也就是计算机视觉的目标。

1.1.2 计算机视觉的目标

计算机视觉是要用计算机来实现人类的视觉功能, 即对客观世界中三维场景的感知、加工和解释。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的内部结构, 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。本书主要从工程技术的角度出发讨论第二种方法。

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出现的专门视觉任务; 而从长远来说则要建成更为通用的系统(更接近人类视觉系统), 完成一般性的视觉任务。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, 掌握和理解人脑视觉工作的机理(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。

由上可见，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功能的理解研究也可帮助人们开发新的图像理解（IU）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。下面结合图 1.1.1 简单介绍几个最接近的学科。

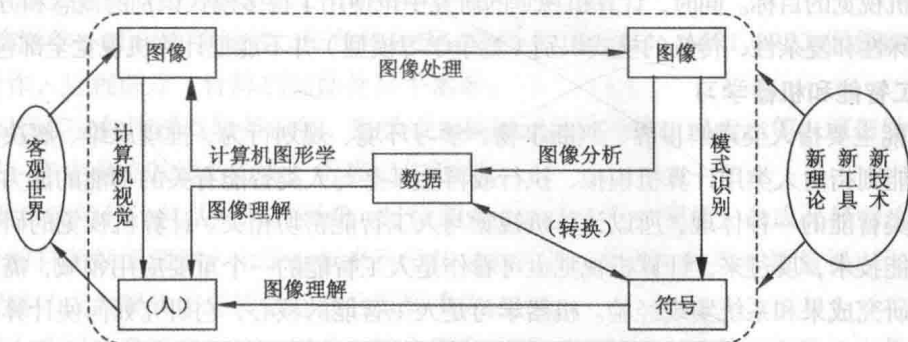


图 1.1.1 相关学科和领域的联系和区别

1. 图像工程

图像工程是一门内容非常丰富的学科，包括既有联系又有区别的三个层次：图像处理、图像分析及图像理解，另外还包括对它们的工程应用。

图像处理着重强调在图像之间进行的转换（图像入图像出）。虽然人们常用图像处理泛指各种图像技术，但比较狭义的处理主要关注的是输出图像的视觉观察效果，这包括对图像进行各种加工调整以改善图像的视觉效果并有利于后续高层加工的进行；或对图像进行压缩编码，在保证所需视觉感受的基础上减少所需存储空间或传输时间，满足给定传输通路的要求；或给图像增加一些附加信息但又不影响原始图像的外貌等。

图像分析主要是对图像中感兴趣的目标进行检测和测量，以获得它们的客观信息从而建立对图像中目标的描述（图像入数据出）。如果说图像处理是一个从图像到图像的过程，则图像分析是一个从图像到数据的过程。这里数据可以是对目标特征测量的结果，或是基于测量的符号表示。它们描述了图像中目标的特点和性质。

图像理解的重点是在图像分析的基础上进一步研究图像中各目标的性质和它们之间的相互联系，并得出对整幅图像内容含义的理解以及对原来成像客观场景的解释，从而可以让人们做出判断，并指导和规划行动。如果说图像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），那么图像理解在一定程度上则是以客观世界为中心，并借助知识、经验等来把握和解释整个客观世界（包括没有直接观察到的事物）。（基于图像处理和分析的）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。它们可以看作是专业和背景不同的人习惯使用的不同术语。

2. 机器视觉或机器人视觉

机器视觉或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场景分析和图像解释的理论和方法，而机器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。另一方面，机器人视觉更强调机器人的机器视觉，要让机器人具有视觉感知功能。

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。（图像）模式识别与图像分析则比较相似，它们有相同的输入，而不同的输出结果可以比较方便地进行转换。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内容（目标）进行分类、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. 人工智能和机器学习

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，计算机视觉也可看作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。机器学习是人工智能的核心，它研究如何使计算机模拟或实现人类的学习行为，从而获取新的知识或技能，这是计算机视觉完成复杂视觉任务的基础。

5. 计算机图形学

图形学原本指用图形、图表、绘图等形式表达数据信息的科学，而计算机图形学研究的就是如何利用计算机技术来产生这些形式，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，而图形学里使用 3-D 模型来生成 2-D 场景图像（更一般的是从非图像形式的数据描述来生成逼真的图像）。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，计算机视觉要借助各种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、生理研究者紧密合作，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结合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

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：如人脸识别，罪犯脸型的合成、识别和查询，指纹、印章的鉴定和识别，支票、签名辨伪等，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，图像的种类和数量得以飞速增长，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 2 章专门介绍。其他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 图像及类别

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[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

图像可看作是对辐射（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， λ 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 个变量的不同取值范围和变化情况,图像的种类和形式是很多的。例如,考虑辐射波长的不同,有 γ 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(利用波长在 1~10 nm 之间的 X 光所获得的图像)、紫外线图像(利用波长在 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(利用波长在 780~1500 nm 之间的红外光所获得的图像)、微波图像(利用波长在 1 mm~10 m 之间的微波所获得的图像)、无线电波图像(利用波长在 10 m~50 km 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(分布在一个平面上图像),还有 3-D 图像(一系列 2-D 图像的集合或利用一些特殊设备得到的 3 个变量的图像)、彩色图像(给人以彩色感受的矢量图像)、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、反映景物物质对辐射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(CT 或 XCT)、正电子发射图像(PET)和单光子发射图像(SPECT)。另外,磁共振图像(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(它们的组合给出彩色的感觉),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分别在其坐标空间 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,这里如不做特别说明, f, x, y 都在整

数集中取值。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,所以如不做特别说明,本书以后就用图像代表数字图像。

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称为**图像元素**。

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):

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

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,一般写成如下的形式:

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

其中

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

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,且主要用于输出显示给人看,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(CRT)**、**电视显示器**和**液晶显示器(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也可看作是图像显示设备。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实现。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图像看作在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。

例 1.2.1 二值图像显示方式

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。

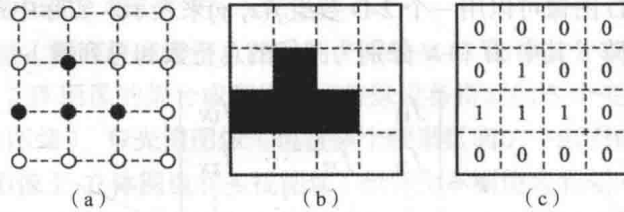


图 1.2.1 3 种表达同一幅 4×4 的二值图像的方式

例 1.2.2 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。 $f(x, y)$ 既可以代表这幅图像, 也可以表示在 (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