

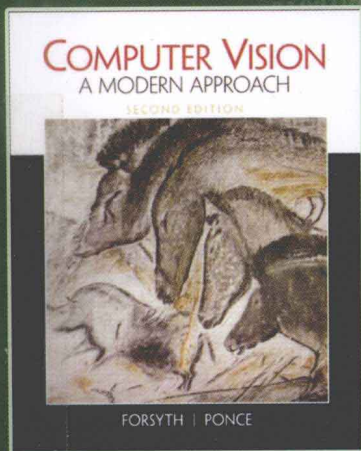
国外计算机科学教材系列

PEARSON

计算机视觉

——一种现代方法(第二版)

Computer Vision: A Modern Approach
Second Edition



英文版

[美] David A. Forsyth 著
Jean Ponce



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

国外计算机科学教材系列

计算机视觉

——一种现代方法

(第二版)(英文版)

Computer Vision: A Modern Approach
Second Edition

[美] David A. Forsyth 著
Jean Ponce

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

计算机视觉是借助于几何、物理和学习理论来建立模型,从而使用统计方法来处理数据的一种方法。本书是近年较为成功的一本计算机视觉教材,内容涉及几何摄像机模型、光照与着色、彩色、线性滤波器、局部图像特性、纹理、立体视觉、从运动求取结构、聚类分割、组合与模型拟合、跟踪、配准、平滑曲面及其轮廓、距离数据、分类、图像分类、图像目标检测、目标识别专题、基于图像的建模与渲染、图像中人的研究、图像搜索与检索、优化技术等。全书条理清楚,系统性强,且各章相对独立;此外,全书理论联系实际,并纳入了近年来该领域的最新研究成果。

本书可作为高等院校计算几何、计算机图形学、图像处理、机器人学等专业学生的教材,也可供相关的专业人士阅读。

Original edition, entitled *COMPUTER VISION: A MODERN APPROACH*, 2E, 9780136085928 by DAVID A. FORSYTH and JEAN PONCE published by Pearson Education, Inc., publishing as Prentice Hall, Copyright © 2012 Pearson Education, Inc.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

English edition published by PEARSON EDUCATION ASIA LTD., and PUBLISHING HOUSE OF ELECTRONICS INDUSTRY, Copyright © 2012.

This edition is manufactured in the People's Republic of China, and is authorized for sale in the mainland of China exclusively (except Taiwan, Hong Kong SAR and Macau SAR).

本书英文影印版专有出版权由 Pearson Education (培生教育出版集团) 授予电子工业出版社。未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

本书在中国大陆地区出版, 仅限在中国大陆发行。

本书英文影印版贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

版权贸易合同登记号 图字: 01-2012-3065

图书在版编目(CIP)数据

计算机视觉: 一种现代方法: 第2版: 英文/(美)福赛斯(Forsyth, D. A.), (美)泊斯(Ponce, J.)著.
北京: 电子工业出版社, 2012.5

(国外计算机科学教材系列)

书名原文: *Computer Vision: A Modern Approach*

ISBN 978-7-121-16830-7

I. ①计… II. ①福… ②泊… III. ①计算机视觉—高等学校—教材—英文 IV. ①TP302.7

中国版本图书馆 CIP 数据核字 (2012) 第 074460 号

策划编辑: 冯小贝

责任编辑: 冯小贝

印刷: 三河市鑫金马印装有限公司

装订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开本: 787×980 1/16 印张: 49.5 字数: 1268 千字

印次: 2012 年 5 月第 1 次印刷

定价: 95.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。服务热线: (010) 88258888。

出版说明

21世纪初的5至10年是我国国民经济和社会发展的重要时期,也是信息产业快速发展的关键时期。在我国加入WTO后的今天,培养一支适应国际化竞争的一流IT人才队伍是我国高等教育的重要任务之一。信息科学和技术方面人才的优劣与多寡,是我国面对国际竞争时成败的关键因素。

当前,正值我国高等教育特别是信息科学领域的教育调整、变革的重大时期,为使我国教育体制与国际化接轨,有条件的高等院校正在为某些信息学科和技术课程使用国外优秀教材和优秀原版教材,以使我国在计算机教学上尽快赶上国际先进水平。

电子工业出版社秉承多年来引进国外优秀图书的经验,翻译出版了“国外计算机科学教材系列”丛书,这套教材覆盖学科范围广、领域宽、层次多,既有本科专业课程教材,也有研究生课程教材,以适应不同院系、不同专业、不同层次的师生对教材的需求,广大师生可自由选择 and 自由组合使用。这些教材涉及的学科方向包括网络与通信、操作系统、计算机组织与结构、算法与数据结构、数据库与信息处理、编程语言、图形图像与多媒体、软件工程等。同时,我们也适当引进了一些优秀英文原版教材,本着翻译版本和英文原版并重的原则,对重点图书既提供英文原版又提供相应的翻译版本。

在图书选题上,我们大都选择国外著名出版公司出版的高校教材,如Pearson Education培生教育出版集团、麦格劳-希尔教育出版集团、麻省理工学院出版社、剑桥大学出版社等。撰写教材的许多作者都是蜚声世界的教授、学者,如道格拉斯·科默(Douglas E. Comer)、威廉·斯托林斯(William Stallings)、哈维·戴特尔(Harvey M. Deitel)、尤利斯·布莱克(Uyless Black)等。

为确保教材的选题质量和翻译质量,我们约请了清华大学、北京大学、北京航空航天大学、复旦大学、上海交通大学、南京大学、浙江大学、哈尔滨工业大学、华中科技大学、西安交通大学、国防科学技术大学、解放军理工大学等著名高校的教授和骨干教师参与了本系列教材的选题、翻译和审校工作。他们中既有讲授同类教材的骨干教师、博士,也有积累了几十年教学经验的老教授和博士生导师。

在该系列教材的选题、翻译和编辑加工过程中,为提高教材质量,我们做了大量细致的工作,包括对所选教材进行全面论证;选择编辑时力求达到专业对口;对排版、印制质量进行严格把关。对于英文教材中出现的错误,我们通过与作者联络和网上下载勘误表等方式,逐一进行了修订。

此外,我们还将与国外著名出版公司合作,提供一些教材的教学支持资料,希望能为授课老师提供帮助。今后,我们将继续加强与各高校教师的密切联系,为广大师生引进更多的国外优秀教材和参考书,为我国计算机科学教学体系与国际教学体系的接轨做出努力。

电子工业出版社

教材出版委员会

- | | | |
|----|-----|---|
| 主任 | 杨芙清 | 北京大学教授
中国科学院院士
北京大学信息与工程学部主任
北京大学软件工程研究所所长 |
| 委员 | 王 珊 | 中国人民大学信息学院院长、教授 |
| | 胡道元 | 清华大学计算机科学与技术系教授
国际信息处理联合会通信系统中国代表 |
| | 钟玉琢 | 清华大学计算机科学与技术系教授、博士生导师
清华大学深圳研究生院信息学部主任 |
| | 谢希仁 | 中国人民解放军理工大学教授
全军网络技术研究中心主任、博士生导师 |
| | 尤晋元 | 上海交通大学计算机科学与工程系教授
上海分布计算技术中心主任 |
| | 施伯乐 | 上海国际数据库研究中心主任、复旦大学教授
中国计算机学会常务理事、上海市计算机学会理事长 |
| | 邹 鹏 | 国防科学技术大学计算机学院教授、博士生导师
教育部计算机基础课程教学指导委员会副主任委员 |
| | 张昆藏 | 青岛大学信息工程学院教授 |

前 言^①

计算机视觉是处于知识前沿的领域之一。与其他前沿领域一样，它既激动人心，又显得头绪繁多。在该领域经常出现缺乏权威性的现象，许多有用的做法并没有理论基础，而一些理论在实际应用中又毫无用处。虽然许多方面的研究已见成效，但是它们之间通常缺乏联系。尽管如此，我们还是力图在本书中较有条理地介绍这一领域。

与研究人类或动物的视觉不同，我们认为计算机视觉（或简称为“视觉”），是借助于几何、物理和学习理论来建立模型，从而使用统计方法来处理数据的一项事业。因此，从我们的角度来看，视觉是指，在透彻理解摄像机性能与物理成像过程的基础上（这是本书第一篇的内容），通过对每个像素值进行简单的推理（第二篇），将多幅图像中可能得到的信息综合成相互关联的整体（第三篇），确定像素集之间的联系以便将它们彼此分割开，或推断一些形状信息（第四篇），进而使用几何信息（第五篇）或概率统计技术（第六篇）来识别物体。计算机视觉的应用相当广泛，既有成熟的应用（如移动机器人导航、工业检测、军事侦察），也有新出现的应用（如人机交互、数字图书馆中的图像检索、医学图像分析，以及计算机图形学中合成场景的渲染）。我们将在第七篇讨论其中的一些具体应用。

第二版的新内容

为增强本书的可用性，我们对第一版做了许多改进。其中最重要的改进源自第一版出版以来计算机视觉这一学科的巨大变化，Internet 上现已广泛发布了相关的代码和数据。至少在直觉上，使用其他人发布的代码来建立系统并基于其他人的数据集来评估系统是很平常的。在本书的几章中，我们将给出一些可用的联机资源指南。Internet 上的这些 URL 并非总是一直有效；我们试图提供足够的信息，以便读者使用作者名、数据集或代码的名称在 Google 中搜索时，可以得到正确的结果。

其他改进包括：

- 简化了内容。与第一版相比，本书的数学知识更简单、更清晰。特别简化了关于摄像机（第 1 章）、着色处理（第 2 章）、由两个视图重建（第 7 章）及由多个视图重建（第 8 章）的内容。
- 给出了更宽泛的应用。具体包括基于图像的建模与渲染（第 19 章）、图像搜索（第 22 章）、建立图像马赛克（12.1 节）、医学图像配准（12.3 节）、

① 采用本书作为教材的教师可获取英文原版教辅：习题解答，PPT。获取方式参见书后的教学支持说明。

解释距离数据（第 14 章）和理解人类活动（第 21 章）。

- 新增了关于现代特性的内容，特别是 HOG 和 SIFT（均在第 5 章中出现），使得应用范围从建立图像马赛克扩展到目标识别。
- 详细介绍了现代图像编辑技术，包括删除阴影（3.5 节）、填充图像中的空洞（6.3 节）、去噪（6.4 节）和交互式图像分割（9.2 节）。
- 全面介绍了现代目标识别技术。讨论先从分类器（第 15 章）开始，然后介绍图像分类技术（第 16 章）和目标检测（第 17 章）的标准方法。最后，第 18 章回顾了目标识别领域的最新进展。
- 最后，本书提供了非常详细的索引及最新的参考文献。

视觉研究的目的

计算机视觉研究的主要目的在于从图像或图像序列中提取对世界的描述。毫无疑问，这是很有使用价值的。摄取图像通常不具有破坏性，因而是安全的。同时，这也是一件不费力的事情，并且现在的成本也很低廉。用户希望从图像中获取的描述对于不同的应用可能相差很大。例如，一种称为从运动求取结构的技术，可以从图像序列中获取所见物体的描述及摄像机的运动规律。娱乐产业中，人们利用这种技术来建立建筑物的三维模型，此时人们关注结构而忽略运动信息。这些模型可以应用到实际建筑物无法使用的场合，如火灾、爆炸等场合。只要利用数量很少的一组照片就可以构造出良好、简单准确、令人信服的模式。而用这种技术来控制移动机器人时，人们一般只会关注运动而将结构舍弃。这是因为，一般只知道机器人工作区域的某些信息，而不知道机器人在这一区域的确切位置，可以由固定于机器人之上的摄像机的运动信息来确定机器人的位置。

计算机视觉还有许多其他方面的重要应用。应用之一是医学图像处理与理解，人们可以设计软件系统来增强图像，鉴别重要的现象或事件，或通过成像获得可视化信息。应用之二是，检验人们对物体拍摄的图像，以便确定它们是否符合规定。应用之三是卫星图像的理解，这既可用于军事目的（例如，编制程序来确定某一地区近来是否有与军事有关的现象发生，或估计轰炸所引起的损害），也可以服务于民用目的（例如，今年的玉米收成会怎样，有多少雨林被保存下来）。应用之四是对收集的图片加以组织与结构化。我们知道如何去搜索与浏览文本库（尽管这仍然是难以解决的课题），但确实不知道如何处理图像或视频库。

计算机视觉自身正处于发展的关键时期。从 20 世纪 60 年代起，人们就想利用计算机视觉的原理构造出有用的计算机系统，但这只是在最近才成为可能。这种繁荣的局面是受多方面的因素驱动的：计算机与成像系统的价格已经很便宜。不久前，要得到较好的数字彩色图像，需要花费上万美元，而现在至多几百美元就已足够。同样，以前很难见到彩色打印机，往往都是在研究实验室中使用，而现在它们已出现在许多家庭中。这意味着进行研究工作变得更加容易，也意味着许多人产生了一些需要使用计算机视觉方法来解决的问题。例如，人们希望将收集的图片组织起来，为他们所在的周围世界构造三维模型，并且管理与编辑收集

的视频。我们对视觉中的基本几何学和物理学的理解及如何运用它这一点已经得到了极大改善。我们开始有可能解决许多人关注的问题，但难题并未得到解决，并且许多较容易的问题也未得到解决（打算解决难题时要保持清醒的头脑）。现在正是研究这一主题的时候。

本书的内容

我们认为计算机视觉从业人员都应知道本书所包含的内容，但要强调的是，本书是面向更广泛的读者的。我们希望从事计算几何、计算机图形学、图像处理、普通成像、机器人等工作的人员会感到这是一本有益的参考书。我们试图使本书适合于对视觉这一课题感兴趣的本科高年级学生或研究生一年级的学生。每一章覆盖这一课题的不同部分，且各章之间是相对独立的，如表 1 所示。这就意味着读者不仅可以阅读整本书，也可着重于某一部分。一般来说，我们已努力做到使每一章从容易的内容开始，而把深奥难懂的内容放在最后。每一章末尾都有一个小结，包含历史性资料及相应的观点。我们努力使本书叙述有用的概念或今后有用的概念。我们把重点放在理解成像的基本几何和物理学知识上，但也力图把它们与实际应用联系起来。总之，本书反映了几何学和多种形式的应用统计学近年来对计算机视觉的多方面的影响。

阅读方式建议

虽然从头到尾阅读本书会很累，但会有许多收获，全书的内容对于一个学期的教学来说有点太多了。当然，未来（或现正）从事计算机视觉工作的专业人员应该逐字阅读本书，完成每一个练习，并报告所发现的问题以便作者在第三版中进行修订。尽管学习计算机视觉并不要求学生具有很深的数学知识，但它的确要求学生熟悉多种不同的数学概念。对于具有工科高年级数学水平的读者，我们已经努力使得该书能自成体系，以便读者无须参考其他教材。因为本书是关于计算机视觉而不是关于应用数学的，因此我们也试着将数学知识降低到最低要求的水平，同时保留了将数学内容穿插在主要章节正文中而不在附录中列出的做法。

总之，我们已经努力降低了各章之间的联系，以方便只对某些专题感兴趣的读者阅读本书。但是，做到各章完全自成体系是不可能的，表 1 中给出了各章之间的联系。

我们尽力制作了丰富的索引，以便读者遇到新术语时，可以按索引在本书中找到它们。计算机视觉现在有着丰富的知识资源。人们广泛共享了数据集和软件，因此本书的相关章节中提供了许多有用数据集和软件的链接；读者也可查阅“software”和“datasets”词条下或通用主题下的索引。

我们尽力提供了最新的参考文献，但不能提供任何主题的完整参考文献，因为提供的文献数量已经足够。

表 1 各章之间的关联性：“必读章节”一列给出的章节必须很好地理解，而“有助章节”一列给出的章节对学习该章是有帮助的

篇	章 节	必 读 章 节	有 助 章 节
第一篇	1. 几何摄像机模型		
	2. 光照与渲染		
	3. 彩色	2	
第二篇	4. 线性滤波器		
	5. 局部图像特性	4	
	6. 纹理	5, 4	2
第三篇	7. 立体视觉	1	22
	8. 从运动求取结构	1, 7	22
第四篇	9. 聚类分割		2, 3, 4, 5, 6, 22
	10. 组合与模型拟合		9
	11. 追踪		2, 5, 22
第五篇	12. 配准	1	14
	13. 平滑曲面及其轮廓	1	
	14. 距离数据		12
	15. 学习分类		22
	16. 对图像进行分类	15, 5	
	17. 检测图像中的目标	16, 15, 5	
	18. 目标识别的主题	17, 16, 15, 5	
第六篇	19. 基于图像的建模与渲染	1, 2, 7, 8	
	20. 研究图像中的人		17, 16, 15, 11, 5
	21. 图像搜索与检索		17, 16, 15, 11, 5
第七篇	22. 优化技术		

本书未包含的内容

计算机视觉的参考文献数量是十分巨大的，因此要写出一本书能让普通读者感兴趣并不是一件容易的事情。为此，我们不得不删掉一些素材、去掉一些主题。

去掉某些主题主要依据个人判断，或是因为我们不得不缩短某些章节，还有其他种种理由。我们有意忽略了那些主要与历史有关的细节，而将历史评论放在每章的末尾。

在讲解概念时，我们力图做到最大的包容性与细致性，但由于计算机视觉本身是一个非常庞大的主题，因此这并不意味着我们具有渊博的知识，而意味着一些概念的历史意义可能比本书阐述的更深。

关于计算机视觉的最新教材如下：Szeliski (2010) 讲解整个视觉；Parker (2010) 主要关注算法；Davies (2005) 和 Steger *et al.* (2008) 致力于特殊的应用，特别是配准；Bradski and Kaehler (2008) 注重于对计算机视觉程序开源包 OpenCV 的介绍。

还有许多更为专业的参考文献。Hartley and Zisserman (2000a) 详细介绍了多

视图几何及多视图参数的估计。Ma *et al.* (2003b) 介绍了三维重建方法。Cyganek and Siebert (2009) 介绍了三维重建与匹配。Paragios *et al.* (2011) 介绍了计算机视觉中的数学模型。Blake *et al.* (2011) 总结了计算机视觉领域的最新马尔可夫随机场模型。Li and Jain (2005) 详细介绍了面部识别。Moeslund *et al.* (2011) 在本书写作时还未出版, 该书将详细介绍用于观察人类的计算机视觉方法。Dickinson *et al.* (2009) 收集整理了目标识别领域的最新进展文献。Badke (2012) 这本即将出版的著作将介绍计算机视觉方法在特效方面的应用。

各种会议论文集中也有许多关于计算机视觉的文献。主要的三个会议如下: IEEE 计算机视觉与模式识别会议 (CVPR), IEEE 计算机视觉国际会议 (ICCV), 欧洲计算机视觉会议。地区性会议上也出现了大量文献, 特别是亚洲计算机视觉会议 (ACCV) 和英国机器视觉会议 (BMVC)。网络上也出现了大量发表的论文, 使用搜索引擎可以搜索到它们。许多大学提供了付费论文, 还有很多论文也可以免费获取。

致谢

许多不知名的评阅者已经阅过本书第一版和第二版的几种草稿版本, 并对本书做出了非常有用的贡献。感谢他们所花费的时间与精力。

第一版的编辑 Alan Apt 在 Jake Warde 的帮助下, 组织了本书的评阅工作, 在此对他们表示感谢。Integre Technical Publishing 公司的 Leslie Galen、Joe Albrecht 和 Dianne Parish 帮助解决了第一版的校对和书中插图的许多问题。

第二版的编辑 Tracy Dunkelberger 在 Carole Snyder 的帮助下组织了本书的评阅工作, 在此对他们表示感谢。还要感谢 Marilyn Floyd 帮助我们解决了各种制作问题。

一些同事评阅了本书的全部内容或若干章节, 他们对这些章节的修订提出了宝贵且详细的建议。我们要感谢 Narendra Ahuja、Francis Bach、Kobus Barnard、Margaret Fleck、Martial Hebert、Julia Hockenmaier、Derek Hoiem、David Kriegman、Jitendra Malik 和 Andrew Zisserman。

我们的许多学生也在提出建议、图示创意、校对评论及其他方面做出了贡献。我们要感谢 Okan Arikan、Louise Benoit、Tamara Berg、Sébastien Blind、Y-Lan Boureau、Liang-Liang Cao、Martha Cepeda、Stephen Chenney、Frank Cho、Florent Couzinie-Devy、Olivier Duchenne、Pinar Duygulu、Ian Endres、Ali Farhadi、Yasutaka Furukawa、Yakup Genc、John Haddon、Varsha Hedau、Nazli Ikizler-Cinbis、Leslie Ikemoto、Sergey Ioffe、Armand Joulin、Keivin Karsch、Svetlana Lazebnik、Cathy Lec、Binbin Liao、Nicolas Loeff、Julien Mairal、Sung-il Pae、David Parks、Fred Rothganger、Amin Sadeghi、Alex Socokin、Attawith Sudsang、Du Tran、Duan Tran、Gang Wang、Yang Wang 和 Ryan White, 以及在加州大学伯克利分校、UIUC 和 ENS 上视觉课的一些学生们的贡献。

所幸的是, 许多大学的同事们在视觉课中使用了本书的草稿版本。使用过这本书初稿版本的学校有卡内基·梅隆大学、斯坦福大学、威斯康星大学、加州大学圣

巴巴拉分校及南加州大学，也可能有一些我们不知道的其他学校。我们对所有使用本书的读者所提出的建设性意见表示感谢，特别要感谢 Chris Bregler、Chuck Dyer、Martial Hebert、David Kriegnum、B. S. Manjunath 和 Ram Nevatia，他们提供了许多详尽的、非常有帮助的评论与改正意见。

以下各位也为本书提供了宝贵意见：Kartek Alahari、Aydin Alayliouglu、Srinivas Akella、Francis Bach、Marie Banich、Serge Belongie、Tamara Berg、Ajit M. Chaudhari、Naveneet Dalal、Jennifer Evans、Yasutaka Furukawa、Richard Hartley、Glenn Healey、Mike Heath、Martial Hebert、Janne Heikkila、Hayley Iben、Stéphanie Jonquières、Ivan Latev、Christine Laubenberer、Svetlana Lazebnik、Yann LeCun、Tony Lewis、Benson Limketkai、Julien Mairal、Simon Maskell、Brian Milch、Roger Mohr、Deva Ramanan、Guillermo Sapiro、Cordelia Schmid、Brigitte Serlin、Gerry Serlin、Ilan Shimshoni、Jamie Shotton、Josef Sivic、Eric de Sturler、Camillo J. Taylor、Jeff Thompson、Claire Vallat、Daniel S. Wilkerson、Jinghan Yu、Hao Zhang、Zhengyou Zhang 和 Andrew Zisserman。

在第一版中，我们曾提及：

如果读者发现了明显的印刷排版错误，请发电子邮件至 daf@cs.berkeley.edu，使用标题“book typo”并告知我们细节，我们将在第二版中感谢每个错误的第一位发现者。

业已证明，这是无效的。DAF 并不具有管理与保护电子邮件日志的能力。我们要感谢发现错误的所有人员；我们已力图修正这些错误并对所有帮助过我们的人员表示谢意。

还要感谢 P. Besl、B. Boufama、J. Costeira、P. Debevec、O. Fangeras、Y. Genc、M. Hebert、D. Huber、K. Ikenchi、A. E. Johnson、T. Kanade、K. Kutulakos、M. Levoy、Y. LeCun、S. Mahamud、R. Mohr、H. Moravec、H. Murase、Y. Ohta、M. Okutami、M. Pollefeys、H. Saito、C. Schmid、J. Shotton、S. Sullivan、C. Tomasi 和 M. Turk，感谢他们为本书的某些插图提供了原件。

DAF 要感谢美国国家科学基金 (NSF) 的支持。对本书写作有直接贡献的基金项目包括 IIS-0803603、IIS-1029035 和 IIS-0916014；其他项目在此处略去。DAF 感谢来自美国海军研究所 (ONR) 的研发支持，正在研究的项目有 N00014-01-1-0890 和 N00014-10-1-0934，它们是 MURI 规划的一部分。这些材料中的任何意见、结论或建议只代表作者本人，与 NSF 或 ONR 无关。

DAF 要感谢其他领域的贡献者。这些贡献者包括 Gerald Alanthwaite、Mike Brady、Tom Fair、Margaret Fleck、Jitendra Malik、Joe Mundy、Mike Rodd、Charlie Rothwell 和 Andrew Zisserman。JP 要对 Olivier Faugeras、Mike Brady 和 Tom Binford 表示感谢。他还要感谢 Sharon Collins 的帮助。没有她，本书不可能顺利完成。两位作者还要感谢 Jan Koenderink 对本书写作的指导。

图形：本书中所用的一些图形源自 IMSI 的主相片集 (1895 Francisco Blvd. East, San Rafael, CA 94901-5506)。我们对来自自己出版文献的图形进行了扩充使用；图题中对此进行了说明。感谢那些让我们使用这些图形的版权所有者的。

参考文献: 在准备参考文献的过程中, 我们扩充了 Keith Price 关于计算机视觉的优秀文献, 读者可在网址 <http://iris.usc.edu/Vision-Notes/bibliography/contents.html> 找到它们。

教学建议

本书可从第一页开始讲起, 对于两个学期的教学, 内容还是比较紧凑的。可以将应用中的一章(如“基于图像的渲染”的相关章节)放在第一学期讲授, 而将有关应用的另一章放在第二学期讲授。有的专业可能不需要如此详细的课程。我们在编排本书时, 考虑到了教师可以按自己的偏好来选择讲授其中的内容。表 2 到表 6 列出了一些用于一学期 15 周课的教学大纲案例, 它们是根据我们的设想来安排的。我们鼓励(并希望)教师们按自己的兴趣来重新安排。

表 2 列出的教学大纲, 是为计算机科学、电子工程或其他工程与自然科学学科的本科高年级学生或一年级研究生设计的计算机视觉导论课, 历时一个学期。学生们可以学到该领域的多个方面的知识, 包括数字图书馆和基于图像的渲染等方面的应用。尽管最难的理论部分被略去了, 但是成像的基本几何和物理学知识是较深入的。我们假设学生具有较广的背景知识, 并建议具备概率论的背景知识。我们将应用章节放到本书的末尾, 但许多人可能会选择提前讲授这些内容。

表 2 对计算机科学、电子工程或其他工科和理科的一年级研究生或本科高年级学生开设一个学期计算机视觉导论课的内容

周次	章号	节号	主要内容
1	1, 2	1.1, 2.1, 2.2.x	针孔摄像机, 像素着色模型, 着色示例的一个推论
2	3	3.1~3.5	人类彩色感知, 彩色物理学, 彩色空间, 图像彩色模型
3	4	全部	线性滤波器
4	5	全部	建立局部特性
5	6	6.1~6.2	来自滤波器和向量量化的纹理表示
6	7	7.1~7.2	双目几何, 立体视觉
7	8	8.1	使用立体摄像机从运动求取结构
8	9	9.1~9.3	分割概念、应用, 通过聚类像素进行分割
9	10	10.1~10.4	霍夫变换, 拟合线, 鲁棒性, RANSAC
10	11	11.1~11.3	简单的跟踪策略, 通过匹配跟踪, 卡尔曼滤波器, 数据相关
11	12	全部	配准
12	15	全部	分类
13	16	全部	对图像进行分类
14	17	全部	检测
15	选修	全部	第 14 章、第 19 章、第 20 章、第 21 章(应用主题)之一

表 3 所列的教学大纲是为计算机图形学专业的学生设计的。他们想知道与自己的课题有关的视觉基础知识。我们在此强调了由图像信息不定期恢复物体模型的方法, 了解这些内容需要了解摄像机和滤波器的运作机理: 跟踪在图形学领域

变得很有用，其对运动分析十分重要。我们认为学生已具有很广的背景知识，并对概率论有一定的了解。

表3 适用于计算机图形学专业学生的教学大纲，他们想知道视觉与自己的课题有关的一些内容

周次	章号	节号	主要内容
1	1, 2	1.1, 2.1, 2.2.4	针孔摄像机，像素着色模型，光度学体视
2	3	3.1~3.5	人类彩色感知，彩色物理学，彩色空间，图像彩色模型
3	4	全部	线性滤波器
4	5	全部	建立局部特性
5	6	6.3, 6.4	纹理合成，图像去噪
6	7	7.1, 7.2	双目几何，立体视觉
7	7	7.4, 7.5	高级立体方法
8	8	8.1	使用立体摄像机从运动求取结构
9	10	10.1~10.4	霍夫变换，拟合线，鲁棒性，RANSAC
10	9	9.1~9.3	分割概念、应用，通过聚类像素进行分割
11	11	11.1~11.3	简单的跟踪策略，通过匹配跟踪，卡尔曼滤波器，数据相关
12	12	全部	配准
13	14	全部	距离数据
14	19	全部	基于图像的建模与渲染
15	13	全部	表面和轮廓

表4中的教学大纲主要是为对计算机视觉应用感兴趣的学生制定的。该教学大纲覆盖了与应用直接有关的内容，我们假定这些学生已具有相当广泛的背景知识，也可以安排背景阅读。

表4 适用于对计算机应用感兴趣的学生的教学大纲

周次	章号	节号	主要内容
1	1, 2	1.1, 2.1, 2.2.4	针孔摄像机，像素着色模型，光度学体视
2	3	3.1~3.5	人类彩色感知，彩色物理学，彩色空间，图像彩色模型
3	4	全部	线性滤波器
4	5	全部	建立局部特性
5	6	6.3, 6.4	纹理合成，图像去噪
6	7	7.1, 7.2	双目几何，立体视觉
7	7	7.4, 7.5	高级立体方法
8	8, 9	8.1, 9.1~9.2	来自立体相机运动的结构，分割思想与应用
9	10	10.1~10.4	霍夫变换，拟合线，鲁棒性，RANSAC
10	12	全部	配准
11	14	全部	距离数据
12	16	全部	对图像进行分类
13	19	全部	基于图像的建模与渲染
14	20	全部	研究图像中的
15	21	全部	图像搜索与检索

表5的教学大纲是为认知科学或人工智能学科的学生设计的,他们需要对计算机视觉重要概念的基本梗概有所了解。这种学模式比较宽松,对学生在数学方面的要求也较少。

表5 适用于认知科学或人工智能学科的学生的教学大纲,他们希望对计算机视觉的重要概念有一个基本的了解

周次	章号	节号	主要内容
1	1,2	1.1, 2.1, 2.2.x	针孔摄像机, 像素着色模型, 着色示例的一个推论
2	3	3.1~3.5	人类彩色感知, 彩色物理学, 彩色空间, 图像彩色模型
3	4	全部	线性滤波器
4	5	全部	建立局部特性
5	6	6.1, 6.2	来自滤波器和向量量化的纹理表示
6	7	7.1, 7.2	双目几何, 立体视觉
8	9	9.1~9.3	分割思想、应用, 通过聚类像素进行分割
9	11	11.1, 11.2	简单的跟踪策略, 使用匹配光学流程进行跟踪
10	15	全部	分类
11	16	全部	对图像进行分类
12	20	全部	研究图像中的人
13	21	全部	图像搜索与检索
14	17	全部	检测
15	18	全部	目标识别中的主题

对于计算机视觉的教学,我们的经验是,学习单独的概念不会出现任何困难,尽管其中的有些概念学习起来要难一些。难点在于这门学科有着太多的新概念。每个子问题看起来都要求学生进行思考,并需要使用新工具来处理它们。这就使得学习这门课程相当令人畏惧。表6给出的教学大纲是为那些对应用数学、电子工程或物理学有强烈兴趣的学生设计的。该教案使得一学期的内容很紧凑,进展很快,并且假设学生能够适应大量的教学内容。

表6 适用于对应用数学、电子工程或物理学有浓厚兴趣的学生的教学大纲

周次	章号	节号	主要内容
1	1,2	全部	摄像机, 着色
2	3	全部	彩色
3	4	全部	线性滤波器
4	5	全部	建立局部特性
5	6	全部	纹理
6	7	全部	立体视觉
7	8	全部	使用立体摄像机从运动求取结构
8	9	全部	通过聚类像素进行分割
9	10	全部	拟合模型
10	11	11.1~11.3	简单的跟踪策略, 使用匹配光学流程进行跟踪, 卡尔曼滤波器, 数据相关

(续表)

周次	章号	节号	主要内容
11	12	全部	配准
12	15	全部	分类
13	16	全部	对图像进行分类
14	17	全部	检测
15	选修	全部	第14章、第19章、第20章、第21章之一

符号表示

我们在全书中使用如下符号表示：点、线和面由斜体罗马或希腊字母表示（如 P 、 Δ 或 Π ）。向量通常由粗斜体罗马字母或希腊字母表示（如 \mathbf{v} 、 \mathbf{P} 或 $\boldsymbol{\xi}$ ），但连接两个点 P 和 Q 的向量通常由 \overrightarrow{PQ} 表示。小写字母通常用于表示图像平面中的几何图形（如 p 、 \mathbf{p} 和 δ ），而大写字母用于表示场景对象（如 P 和 Π ）。矩阵由书写体的罗马字母表示（如 U ）。

我们熟悉的三维欧几里得空间由 \mathbb{E}^3 表示，由 n 个实元形成的向量空间由 \mathbb{R}^n 表示，该空间满足加法与标量相乘定律，其中 $\mathbf{0}$ 用于表示零向量。同样，由 $m \times n$ 矩阵形成的具有实数项的向量空间由 $\mathbb{R}^{m \times n}$ 表示。当 $m = n$ 时， Id 用于表示单位矩阵，即对角项为 1 而非对角项为 0 的矩阵。带有系数 u_{ij} 的 $m \times n$ 矩阵 U 的转置矩阵，由带有系数 u_{ji} 的 $n \times m$ 矩阵 U^T 表示。 \mathbb{R}^n 的元素通常由列向量或 $n \times 1$ 矩阵来标识，例如， $\mathbf{a} = (a_1, a_2, a_3)^T$ 是 1×3 矩阵（或行向量）的转置矩阵，即一个 3×1 矩阵（或列向量），或 \mathbb{R}^3 中的一个等效元素。

\mathbb{R}^n 中两个向量 $\mathbf{a} = (a_1, \dots, a_n)^T$ 和 $\mathbf{b} = (b_1, \dots, b_n)^T$ 的点积（或内积）定义为

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + \dots + a_n b_n$$

它也可写为矩阵乘积的形式，即 $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a}$ 。我们用 $|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a}$ 来表示向量 \mathbf{a} 的欧几里得范数的平方，用 d 表示由 \mathbb{E}^n 中欧几里得范数引起的距离函数，即 $d(P, Q) = |\overrightarrow{PQ}|$ 。给定 $\mathbb{R}^{m \times n}$ 中的一个矩阵 U ，则通常使用 $|U|$ 表示其 Frobenius 范数，即其各项的平方和的平方根。

当向量 \mathbf{a} 的范数为 1 时，点积 $\mathbf{a} \cdot \mathbf{b}$ 等于 \mathbf{b} 在 \mathbf{a} 上的投影长度。更一般地，

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta$$

当 θ 是两个向量的夹角时，则表明两个向量正交的充分必要条件是点积为零。

\mathbb{R}^3 中两个向量 $\mathbf{a} = (a_1, a_2, a_3)^T$ 和 $\mathbf{b} = (b_1, \dots, b_3)^T$ 的叉积（或外积）是向量

$$\mathbf{a} \times \mathbf{b} \stackrel{\text{def}}{=} \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

注意 $\mathbf{a} \times \mathbf{b} = [\mathbf{a} \times] \mathbf{b}$ ，其中，

$$[a \times] \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$$

\mathbb{R}^3 中两个向量 \mathbf{a} 和 \mathbf{b} 的叉积与这两个向量正交，且 \mathbf{a} 和 \mathbf{b} 有相同方向的充分必要条件是 $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ 。如果 θ 像之前一样表示向量 \mathbf{a} 和 \mathbf{b} 间的夹角，则可证明

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}| |\sin \theta|$$

编程练习和源程序

本书中给出的编程练习有时需要数值线性代数、奇异值分解及线性与非线性最小二乘的程序。这些程序的较完整集合可以在 MATLAB 及 LINPACK、LAPACK 和 MINPACK 中得到，它们可以从 Netlib 库中下载 (<http://www.netlib.org/>)。在本书的正文中，我们通常会给出网上所发布软件和数据集的链接。OpenCV 是一个计算机视觉程序的重要开源包 [见 Bradski and Kaehler (2008)]。

关于作者

David A. Forsyth: 1984 年于威特沃特斯兰德大学取得电气工程学士学位，1986 年取得电气工程硕士学位，1989 年于牛津贝列尔学院取得博士学位。之后在艾奥瓦大学任教 3 年，并在加州大学伯克利分校任教 10 年，之后在伊利诺伊大学任教。2000 年和 2001 年任 IEEE 计算机视觉和模式识别会议执行副主席，2006 年任 CVPR 常任副主席，2008 年任欧洲计算机视觉会议执行副主席，是所有关于计算机视觉主要国际会议的常任执委会成员。他为 SIGGRAPH 执委会工作了 5 期。2006 年获 IEEE 技术成就奖，2009 年成为 IEEE 会士。

Jean Ponce: 分别于 1983 年和 1988 年在巴黎奥赛大学获得 Troisieme Cycle 和 Doctorat d'État 计算机科学学士学位。1990 年至 2005 年，作为研究科学家分别供职于法国国家信息研究所、麻省理工学院人工智能实验室和斯坦福大学机器人实验室；1990 年至 2005 年，供职于伊利诺伊大学计算机科学系。2005 年开始，成为法国巴黎高等师范学校教授。Ponce 博士还是 Computer Vision and Image Understanding、Foundations and Trends in Computer Graphics and Vision、IEEE Transactions on Robotics and Automation、International Journal of Computer Vision (2003 年至 2008 年为首席编辑)、SIAM Journal on Imaging Sciences 的编委会成员。1997 年，任计算机视觉与模式识别 IEEE 会议执行主席，2000 年任会议的大会主席。2008 年，任欧洲计算机视觉会议大会主席。2003 年，因其对计算机视觉的突出贡献，成为 IEEE 会士，并因机器人零件供给的研发工作获得美国专利。

Contents

I	IMAGE FORMATION	1
1	Geometric Camera Models	3
1.1	Image Formation	4
1.1.1	Pinhole Perspective	4
1.1.2	Weak Perspective	6
1.1.3	Cameras with Lenses	8
1.1.4	The Human Eye	12
1.2	Intrinsic and Extrinsic Parameters	14
1.2.1	Rigid Transformations and Homogeneous Coordinates	14
1.2.2	Intrinsic Parameters	16
1.2.3	Extrinsic Parameters	18
1.2.4	Perspective Projection Matrices	19
1.2.5	Weak-Perspective Projection Matrices	20
1.3	Geometric Camera Calibration	22
1.3.1	A Linear Approach to Camera Calibration	23
1.3.2	A Nonlinear Approach to Camera Calibration	27
1.4	Notes	29
2	Light and Shading	32
2.1	Modelling Pixel Brightness	32
2.1.1	Reflection at Surfaces	33
2.1.2	Sources and Their Effects	34
2.1.3	The Lambertian+Specular Model	36
2.1.4	Area Sources	36
2.2	Inference from Shading	37
2.2.1	Radiometric Calibration and High Dynamic Range Images	38
2.2.2	The Shape of Specularities	40
2.2.3	Inferring Lightness and Illumination	43
2.2.4	Photometric Stereo: Shape from Multiple Shaded Images	46
2.3	Modelling Interreflection	52
2.3.1	The Illumination at a Patch Due to an Area Source	52
2.3.2	Radiosity and Exitance	54
2.3.3	An Interreflection Model	55
2.3.4	Qualitative Properties of Interreflections	56
2.4	Shape from One Shaded Image	59