

计算机视觉

——一种现代方法(第二版)

Computer Vision: A Modern Approach
Second Edition

COMPUTER VISION
A MODERN APPROACH



FORSYTH | PONCE

英文版

[美] David A. Forsyth 著
Jean Ponce

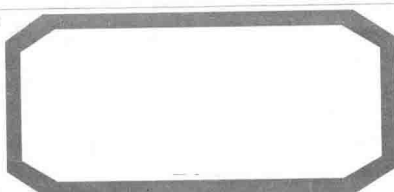


中国工信出版集团



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



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 Jean Ponce 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

计算机视觉是研究如何使人工系统从图像或多维数据中“感知”的科学。本书是计算机视觉领域的经典教材,内容涉及摄像机的几何模型、光照及阴影、颜色、线性滤波、局部图像特征、纹理、立体视觉、运动结构、聚类分割、分组与模型拟合、跟踪、配准、平滑表面及其轮廓、深度数据、图像分类、物体检测与识别、基于图像的建模与渲染、人形研究、图像搜索与检索、优化技术等。与前一版相比,本书简化了部分主题,增加了应用示例,重写了关于现代特征的内容,详述了现代图像编辑技术与物体识别技术。

本书可作为计算机几何学、计算机图形学、图像处理、模式识别、机器人学等专业高年级本科生和研究生的双语教材或参考书,也可供从事相关领域研究的工程技术人员参考阅读。

Original edition, entitled *Computer Vision: A Modern Approach*, Second Edition, 9780136085928 by David A. Forsyth and Jean Ponce, published by Pearson Education, Inc., publishing as Prentice Hall, Copyright © 2012 by 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.

China edition published by PEARSON EDUCATION ASIA LTD., and PUBLISHING HOUSE OF ELECTRONICS INDUSTRY copyright ©2017.

This edition is manufactured in the People's Republic of China, and is authorized for sale and distribution only 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:第二版=Computer Vision: A Modern Approach, Second Edition: 英文/(美)戴维·A.福赛斯(David A. Forsyth),(美)简·泊斯(Jean Ponce)著.一北京:电子工业出版社,2017.6

国外计算机科学教材系列

ISBN 978-7-121-31826-9

I. ①计… II. ①戴… ②简… 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(2017)第129527号

策划编辑:冯小贝

责任编辑:冯小贝

印刷:北京京师印务有限公司

装订:北京京师印务有限公司

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:45.75 字数:1370千字

版次:2017年6月第1版(原著第2版)

印次:2017年6月第1次印刷

定价:128.00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, 88258888。

质量投诉请发邮件至zltts@phei.com.cn, 盗版侵权举报请发邮件至dbqq@phei.com.cn。

本书咨询联系方式:fengxiaobei@phei.com.cn。

前 言

计算机视觉是一个处于知识前沿的领域。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对这一领域进行有条理的分析。

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个研究领域。因此,从我们的角度来看,视觉是指,在透彻理解照相机性能与物理成像过程的基础上(这是本书第一部分的内容),通过对每个像素值进行简单的推理(第二部分),将多幅图像中可能得到的信息综合成相互关联的整体(第三部分),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(第四部分),进而使用几何信息(第五部分)或概率统计技术(第六部分)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)。我们将在第七部分讨论其中的一些具体应用。

第二版的新内容

为增强本书的有用性,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。目前,互联网上已经有广泛公开的代码和数据。使用其他人发布的代码来构建自己的系统,或者在其他人的数据集上进行评估是非常平常的事情,至少在直观感受上是这样的。在本书的某些章节中,我们将给出一些可用的联机资源指南。互联网上的这些 URL 资源并非一直有效;我们试图提供足够的信息,以便读者可用作者名、数据集或代码的名称在互联网中搜索,并得到正确的结果。

其他改进包括:

- 简化了公式推导。与第一版相比,本书的数学知识更简单、更清晰。特别简化了关于照相机(第1章)、阴影处理(第2章)、由两个视图重建(第7章)及由多个视图重建(第8章)的相关公式推导。
- 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)、解释深度数据(第14章)和理解人类活动(第21章)。
- 新增了关于现代特征的内容,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物体识别。
- 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)。

- 全面介绍了现代物体识别技术。首先从分类器(第 15 章)开始讨论,然后介绍图像分类技术(第 16 章)和物体检测算法(第 17 章)。最后,第 18 章回顾了物体识别领域的最新进展。
- 最后,本书提供了相关的参考文献^①。

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。拍摄的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。针对不同的应用,用户希望从图像中获取的描述可能相差很大。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的、简单准确的、令人信服的模型。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。我们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,或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,编写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。之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,之前我们也很难看到彩色打印机,即使看到,也往往是在实验室中,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为容易,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,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。

① 相关的参考文献请登录华信教育资源网(www.hxedu.com.cn)注册下载。

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。这使得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，或者今后有用的概念。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章阅读本书，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。对于具有本科高年级数学水平的读者，我们已经努力使该书能自成体系，以便读者无须参考其他课本。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要求的最低水平，同时保留了将数学内容穿插在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在整本书中漫游。但是，做到各章完全自成体系是不可能的，表 1 给出了各章之间的联系。

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程序的链接。

本书尽力提供了相关的参考文献，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太多了。

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

部 分	章 名	必 读 章 节	有 助 章 节
第一部分	1. 摄像机的几何模型		
	2. 光照及阴影		
	3. 颜色	2	
第二部分	4. 线性滤波		
	5. 局部图像特征	4	
	6. 纹理	5, 4	2
第三部分	7. 立体视觉	1	22
	8. 从运动中恢复三维结构	1, 7	22

部 分	章 名	必 读 章 节	有 助 章 节
第四部分	9. 基于聚类的分割方法		2, 3, 4, 5, 6, 22
	10. 分组与模型拟合		9
	11. 跟踪		2, 5, 22
第五部分	12. 配准	1	14
	13. 平滑表面及其轮廓	1	
	14. 深度数据		12
	15. 用于分类的学习		22
	16. 图像分类	15, 5	
	17. 检测图像中的物体	16, 15, 5	
	18. 物体识别	17, 16, 15, 5	
第六部分	19. 基于图像的建模与渲染	1, 2, 7, 8	
	20. 对人的观察		17, 16, 15, 11, 5
	21. 图像搜索与检索		17, 16, 15, 11, 5
第七部分	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、去掉一些主题等。

去掉某些主题是因为我们的个人判断,或是因为已感到筋疲力尽而停止了撰写,或是因为得知它们已太晚而无法加入书中,或是因为不得不缩短某些章节,或是其他种种理由。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我们阐述得更深。

关于计算机视觉的相关教材如下: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. (2010)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的计算机视觉方法。Dickinson et al. (2009)收集整理了物体识别领域的最新进展。Badke (2012)介绍了计算机视觉方法在特效方面的应用。

各种会议论文集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;许多大学也提供了付费阅读的论文。

致谢

在筹备这本书的过程中,我们已经欠下了一大笔人情债。许多不知名的评阅者已经阅读过本书第一版和第二版的初稿,并对本书做出了非常巨大的贡献。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 的帮助下组织了评阅工作,在此对他们表示感谢。还要感谢 Marilyn Lloyd 帮助我们解决了各种制作问题。

一些同事评阅了整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、Kevin Karsch、Svetlana Lazebnik、Cathy Lee、Binbin Liao、Nicolas Loeff、Julien Mairal、Sung-il Pae、David Parks、Deva Ramanan、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 上视觉课程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程中使用了本书的初稿版本。使用过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、加州大学圣塔芭芭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的建设性意见表示感谢,特别要感谢 Chris Bregler、Chuck Dyer、Martial Hebert、David Kriegman、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。

这本书还受益于 Karteek Alahari、Aydin Alaylioglu、Srinivas Akella、Francis Bach、Marie Banich、Serge Belongie、Tamara Berg、Ajit M. Chaudhari、Navneet Dalal、Jennifer Evans、Yasutaka Furukawa、Richard Hartley、Glenn Healey、Mike Heath、Martial Hebert、Janne Heikkilä、Hayley Iben、Stéphanie Jonquière、Ivan Laptev、Christine Laubenberg、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,并使用短语“book typo”告诉我们细节,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处理方式。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。

在此还要感谢 P. Besl、B. Boufama、J. Costeira、P. Debevec、O. Fangeras、Y. Genc、M. Herbert、D. Huber、K. Ikeuchi、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有 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,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、结论或建议只是作者个人提供的,与 NSF 或 ONR 无关。

DAF 要感谢更大范围的贡献者。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 虽然不能记起小时候的事情,但他要对 Olivier Faugeras、Mike Brady 和 Tom Binford 表示感谢。他还要感谢 Sharon Collins 的帮助,没有她,本书不可能完成。两位作者还要感谢 Jan Koenderink 对本书写作的指导。

插图:本书中所用的一些插图源自 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插图进行了扩充使用,图题中对此进行了说明。感谢那些让我们使用这些插图的版权所有者的。

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(如第19章“基于图像的建模与渲染”)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。有的系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~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、电气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。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 对于计算机科学、电子工程或其他工科或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的导论课程的内容

周次	章号	节号	主要内容
1	1, 2	1.1, 2.1, 2.2. x	针孔照相机, 像素阴影模型, 阴影示例的一个推论
2	3	3.1 ~3.5	人类色彩感知, 颜色物理学, 颜色空间, 图像颜色模型
3	4	全部	线性滤波器
4	5	全部	构建局部特征
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中恢复结构
8	9	9.1 ~9.3	分割概念、应用, 采用聚类方法对像素进行分割
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, 学习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泛的背景知识, 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	针孔照相机, 像素阴影模型, 光度立体
2	3	3.1 ~3.5	人类色彩感知, 颜色物理学, 颜色空间, 图像颜色模型
3	4	全部	线性滤波器
4	5	全部	构建局部特征
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中恢复结构
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, 采用聚类方法对像素进行分割
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	深度数据
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设定的。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,也可以安排背景知识阅读。

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	针孔照相机, 像素阴影模型, 光度立体
2	3	3.1~3.5	人类色彩感知, 颜色物理学, 颜色空间, 图像颜色模型
3	4	全部	线性滤波器
4	5	全部	构建局部特征
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	来自立体摄像机运动的结构, 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	深度数据
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	对人的观察
15	21	全部	图像搜索与检索

表5中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。这个教案显得不那么步步紧逼,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	针孔照相机, 像素阴影模型, 阴影示例的一个推论
2	3	3.1~3.5	人类色彩感知, 颜色物理学, 颜色空间, 图像颜色模型
3	4	全部	线性滤波器
4	5	全部	构建局部特征
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, 采用聚类方法对像素进行分割
9	11	11.1, 11.2	简单的跟踪策略, 使用匹配进行跟踪, 流跟踪
10	15	全部	分类
11	16	全部	图像分类
12	20	全部	对人的观察
13	21	全部	图像搜索与检索
14	17	全部	检测
15	18	全部	物体识别

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使人畏惧。表6给出的教学大纲是为那些对应用数学、电气工程或物理学有浓厚兴趣的学生设计的。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、电气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; 2.1~2.4	照相机, 阴影
2	3	全部	颜色
3	4	全部	线性滤波器
4	5	全部	构建局部特征
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中恢复结构
8	9	全部	采用聚类方法对像素进行分割
9	10	全部	拟合模型
10	11	11.1~11.3	简单的跟踪策略, 使用匹配进行跟踪, 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	图像分类
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

在全书中采用如下的符号表示:点、线和面用斜体罗马字母或希腊字母表示(如 P 、 Δ 或 Π)。向量通常用粗斜体罗马字母或希腊字母表示(如 $\boldsymbol{v}$ 、 $\boldsymbol{P}$ 或 $\boldsymbol{\xi}$),但连接两个点 P 和 Q 的向量通常用 $\overrightarrow{PQ}$ 表示。小写字母通常用于表示图像平面中的几何图形(如 p 、 $\boldsymbol{p}$ 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用 $\mathbb{R}^n$ 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用 $\mathbb{R}^{m \times n}$ 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,用带有系数 u_{ji} 的 $n \times m$ 矩阵 U^T 表示。 $\mathbb{R}^n$ 的元素通常用列向量或 $n \times 1$ 矩阵来标识,例如, $\boldsymbol{a} = (a_1, a_2, a_3)^T$ 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 $\boldsymbol{a} = (a_1, \cdots, a_n)^T$ 和 $\boldsymbol{b} = (b_1, \cdots, b_n)^T$ 的点积(或内积)定义为 $\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + \cdots + a_n b_n$,它也可写为矩阵乘积的形式,即 $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T \boldsymbol{b} = \boldsymbol{b}^T \boldsymbol{a}$ 。用 $\|\boldsymbol{a}\|^2 = \boldsymbol{a} \cdot \boldsymbol{a}$ 来表示向量 $\boldsymbol{a}$ 的欧几里得范数的平方,用 d 表示由 $\mathbb{E}^n$ 中欧几里得范数引出的距离函数,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 $\boldsymbol{a}$ 的范数为 1 时,点积 $\boldsymbol{a} \cdot \boldsymbol{b}$ 等于 $\boldsymbol{b}$ 在 $\boldsymbol{a}$ 上的投影长度。更一般地,有

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

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 $\mathbf{b} = (b_1, b_2, b_3)^T$ 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,

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

$\mathbb{R}^3$ 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 及一些公共库中得到, 如 LINPACK、LAPACK 和 MINPACK, 它们可以从 Netlib 库(<http://www.netlib.org/>)中下载。在本书的正文中, 通常会给出网上所发布的源程序和数据集的链接。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了电气工程学士学位, 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, 在加州大学伯克利分校任教 10 年, 再后在伊利诺伊大学任教。2000 年和 2001 年任 IEEE 计算机视觉与模式识别会议(CVPR)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: 于 1988 年在巴黎奥赛大学获得计算机科学博士学位。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《计算机视觉与图像理解》、《计算机图形学与视觉发展及趋势》、IEEE《机器人和自动化学报》、计算机视觉国际会议(2003 年至 2008 年为首席编辑)、SIAM《成像学报》的编委会成员。1997 年, 任 IEEE 计算机视觉与模式识别会议执行主席, 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

2.5	Notes	61
3	Color	68
3.1	Human Color Perception	68
3.1.1	Color Matching	68
3.1.2	Color Receptors	71
3.2	The Physics of Color	73
3.2.1	The Color of Light Sources	73
3.2.2	The Color of Surfaces	76
3.3	Representing Color	77
3.3.1	Linear Color Spaces	77
3.3.2	Non-linear Color Spaces	83
3.4	A Model of Image Color	86
3.4.1	The Diffuse Term	88
3.4.2	The Specular Term	90
3.5	Inference from Color	90
3.5.1	Finding Specularities Using Color	90
3.5.2	Shadow Removal Using Color	92
3.5.3	Color Constancy: Surface Color from Image Color	95
3.6	Notes	99
II	EARLY VISION: JUST ONE IMAGE	105
4	Linear Filters	107
4.1	Linear Filters and Convolution	107
4.1.1	Convolution	107
4.2	Shift Invariant Linear Systems	112
4.2.1	Discrete Convolution	113
4.2.2	Continuous Convolution	115
4.2.3	Edge Effects in Discrete Convolutions	118
4.3	Spatial Frequency and Fourier Transforms	118
4.3.1	Fourier Transforms	119
4.4	Sampling and Aliasing	121
4.4.1	Sampling	122
4.4.2	Aliasing	125
4.4.3	Smoothing and Resampling	126
4.5	Filters as Templates	131
4.5.1	Convolution as a Dot Product	131
4.5.2	Changing Basis	132
4.6	Technique: Normalized Correlation and Finding Patterns	132

4.6.1	Controlling the Television by Finding Hands by Normalized Correlation	133
4.7	Technique: Scale and Image Pyramids	134
4.7.1	The Gaussian Pyramid	135
4.7.2	Applications of Scaled Representations	136
4.8	Notes	137
5	Local Image Features	141
5.1	Computing the Image Gradient	141
5.1.1	Derivative of Gaussian Filters	142
5.2	Representing the Image Gradient	144
5.2.1	Gradient-Based Edge Detectors	145
5.2.2	Orientations	147
5.3	Finding Corners and Building Neighborhoods	148
5.3.1	Finding Corners	149
5.3.2	Using Scale and Orientation to Build a Neighborhood	151
5.4	Describing Neighborhoods with SIFT and HOG Features	155
5.4.1	SIFT Features	157
5.4.2	HOG Features	159
5.5	Computing Local Features in Practice	160
5.6	Notes	160
6	Texture	164
6.1	Local Texture Representations Using Filters	166
6.1.1	Spots and Bars	167
6.1.2	From Filter Outputs to Texture Representation	168
6.1.3	Local Texture Representations in Practice	170
6.2	Pooled Texture Representations by Discovering Textons	171
6.2.1	Vector Quantization and Textons	172
6.2.2	K-means Clustering for Vector Quantization	172
6.3	Synthesizing Textures and Filling Holes in Images	176
6.3.1	Synthesis by Sampling Local Models	176
6.3.2	Filling in Holes in Images	179
6.4	Image Denoising	182
6.4.1	Non-local Means	183
6.4.2	Block Matching 3D (BM3D)	183
6.4.3	Learned Sparse Coding	184
6.4.4	Results	186
6.5	Shape from Texture	187
6.5.1	Shape from Texture for Planes	187
6.5.2	Shape from Texture for Curved Surfaces	190

6.6	Notes	191
III EARLY VISION: MULTIPLE IMAGES		195
7	Stereopsis	197
7.1	Binocular Camera Geometry and the Epipolar Constraint	198
7.1.1	Epipolar Geometry	198
7.1.2	The Essential Matrix	200
7.1.3	The Fundamental Matrix	201
7.2	Binocular Reconstruction	201
7.2.1	Image Rectification	202
7.3	Human Stereopsis	203
7.4	Local Methods for Binocular Fusion	205
7.4.1	Correlation	205
7.4.2	Multi-Scale Edge Matching	207
7.5	Global Methods for Binocular Fusion	210
7.5.1	Ordering Constraints and Dynamic Programming	210
7.5.2	Smoothness and Graphs	211
7.6	Using More Cameras	214
7.7	Application: Robot Navigation	215
7.8	Notes	216
8	Structure from Motion	221
8.1	Internally Calibrated Perspective Cameras	221
8.1.1	Natural Ambiguity of the Problem	223
8.1.2	Euclidean Structure and Motion from Two Images	224
8.1.3	Euclidean Structure and Motion from Multiple Images	228
8.2	Uncalibrated Weak-Perspective Cameras	230
8.2.1	Natural Ambiguity of the Problem	231
8.2.2	Affine Structure and Motion from Two Images	233
8.2.3	Affine Structure and Motion from Multiple Images	237
8.2.4	From Affine to Euclidean Shape	238
8.3	Uncalibrated Perspective Cameras	240
8.3.1	Natural Ambiguity of the Problem	241
8.3.2	Projective Structure and Motion from Two Images	242
8.3.3	Projective Structure and Motion from Multiple Images	244
8.3.4	From Projective to Euclidean Shape	246
8.4	Notes	248