

计算机视觉——算法与应用

Computer Vision: Algorithms and Applications

(美) Richard Szeliski 著
艾海舟 兴军亮 等译



计 算 机 视 觉

——算法与应用

(美) Richard Szeliski 著
艾海舟 兴军亮 等译

清华大学出版社
北 京

内 容 简 介

《计算机视觉——算法与应用》探索了用于分析和解释图像的各种常用技术,描述了具有一定挑战性的视觉应用方面的成功实例,兼顾专业的医学成像和图像编辑与交织之类有趣的大众应用,以便学生能够将其应用于自己的照片和视频,从中获得成就感和乐趣。本书从科学的角度介绍基本的视觉问题,将成像过程的物理模型公式化,然后在此基础上生成对场景的逼真描述。作者还运用统计模型来分析和运用严格的工程方法来解决这些问题。

本书作为本科生和研究生“计算机视觉”课程的理想教材,适合计算机和电子工程专业学生使用,重点介绍现实中之有效的基本技术,通过大量应用和练习来鼓励学生大胆创新。此外,本书的精心设计和编排,使其可以作为计算机视觉领域中一本独特的基础技术参考和最新研究成果文献。

Translation from the English language edition:

Computer Vision: Algorithms and Applications, 1st Edition by Richard Szeliski

Copyright © Richard Szeliski 2010

Springer is a part of Springer Science + Business Media

All Rights Reserved.

本书中文简体字翻译版由德国施普林格公司授权清华大学出版社在中华人民共和国境内(不包括中国香港、澳门特别行政区和中国台湾地区)独家出版发行。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。

北京市版权局著作权合同登记号 图字:01-2011-0579

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

计算机视觉——算法与应用/(美)塞利斯基(Szeliski,R.)著;艾海舟,兴军亮等译.

—北京:清华大学出版社,2012.1

书名原文:Computer Vision: Algorithms and Applications

ISBN 978-7-302-26915-1

I. ①计… II. ①塞… ②艾… ③兴… III. ①计算机视觉—应用 IV. ①TP302.7

中国版本图书馆 CIP 数据核字(2011)第 187910 号

责任编辑:文开琪

封面设计:杨玉兰

版式设计:北京东方人华科技有限公司

责任校对:周剑云

责任印制:王秀菊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

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

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:清华大学印刷厂

装 订 者:三河市金元印装有限公司

经 销:全国新华书店

开 本:185×260 **印 张:**42.5 **插 页:**16 **字 数:**1020 千字

版 次:2012 年 1 月第 1 版 **印 次:**2012 年 1 月第 1 次印刷

印 数:1~4000

定 价:99.00 元

译者的话

在近两年“计算机视觉”课程的教学过程中，我向学生推荐这本当时尚未正式出版的教材(网上有不断更新的电子版草稿)作为参考书，我觉得这是一本难得的好教材。

为了帮助学生扫除阅读英文版时可能碰到的障碍，我承担了本书的翻译工作。有了中文版，广大读者——尤其是无法流畅阅读英文版的读者——学习起来无疑更轻松，从而使本书能够充分发挥作用。

在本书的翻译过程中，对于英文版中明显存在的少数纰漏(主要是排印或疏漏)，我们逐一进行了订正。这些错误一般都很明显，因此译文中我们没有专门声明。有英文阅读能力的读者，在阅读本书的过程中，不妨参照英文版，这样做不仅可以加深对相关专业术语的理解，还能通过这一实践方式提升专业英语阅读能力。

参与本书翻译的人员如下(按照工作量的大小排序)：

艾海舟(序、目录、第1章、第2章、第15章和术语)

兴军亮(第4章和第11章)

段根全(第6章和第14章，14.2节除外)

陈先捷(第10章)

曹翀(第3章的3.5~3.9节)

苏延超(第12章，14.2节)

忻海(第8章)

杜宇宁(第3章的3.1~3.4节)

王楠(第5章)

卜鹏洋(附录A~C)

张晨光(第9章)

刘力为(第7章)

刘之方(第13章)

全部译稿的最后审定由艾海舟负责。由于译者水平有限，书中难免存在纰漏，欢迎广大读者批评指正。在阅读过程中，如果发现问题，请发送电子邮件告知，以便今后重印时加以订正。

艾海舟

清华大学计算机系

电子邮件: ahz@mail.tsinghua.edu.cn

主页网址: <http://media.cs.tsinghua.edu.cn/~ahz>

序

本书萌芽于 2001 年，当时，华盛顿大学的 Steve Seitz 邀我和他一起讲一门课，课程名称是“面向计算机图形学的计算机视觉”。那个时候，计算机图形学领域正在越来越多地使用计算机视觉技术，用它来创建基于图像的真实物体的模型，用于产生视觉效果，用于通过计算摄影学技术来合并真实影像。我们决定聚焦于计算机视觉在若干有趣问题中的应用，例如使用个人照片的图像拼接和基于照片的 3D 建模等，这一想法引起了学生们的共鸣。

从那时起，华盛顿大学和斯坦福大学就一直使用类似的课程大纲和项目导向的课程结构来进行常规计算机视觉课程的教学(在斯坦福大学，在 2003 年这门课程由我和 David Fleet 共同讲授)。类似的课程大纲也被其他很多大学所采用，并被纳入计算摄影学相关的更专业的课程。(有关如何在课程中使用本书的建议，请参见 1.4 节的表 1.1。)

本书还反映了我在企业研究实验室(DEC 剑桥研究实验室和微软研究院)这二十年的计算机视觉研究经历。在从事研究的过程中，我主要关注在真实世界中具有实际应用的问题和在实践中行之有效的办法(算法)。因此，本书更强调在真实世界条件下有效的基本方法，而较少关注内在完美但难以实际应用的神秘的数学内容。

本书适用于计算机科学和电子工程专业高年级本科的计算机视觉课程。学生最好已经修过图像处理或计算机图形学课程，这样一来，便可以少花一些时间来学习一般性的数学背景知识，多花一些时间来学习计算机视觉技术。本书也适用于研究生的计算机视觉课程(通过专研更富有挑战性的应用和算法领域)，作为基本技术和近期研究文献的参考用书。为此，我尽量尝试引用每个子领域中最新的研究进展，即便其技术细节过于复杂而无法在本书中涉及。

在课程教学过程中，我们发现，要使学生从容应对真实图像及其带来的挑战，让他们尝试实现一些小的课程设计(通常一个建立在另一个基础之上)，是很有帮助的。随后，要求学生分成组选择各自的主题，完成最终的课程设计。(有时，这些课程设计甚至能转换为会议论文!)本书各章最后的习题包含有关小型中期课程设计题目的很多建议，也包含一些更开放的问题，这些问题的解决仍然是活跃的研究课题。只要有可能，我都会鼓励学生用他们自己的个人照片来测试他们的算法，因为这可以更好地激发他们的兴趣，往往会产生富有创造性的衍生问题，使他们更熟悉真实影像的多样性和复杂性。

在阐述和解决计算机视觉问题的过程中，我常常发现从三个高层途径获取灵感是有帮助的。

- **科学层面：**建立图像形成过程的详细模型，为了恢复感兴趣量而构建其逆过程的数学方法(必要时，做简化假设使其在数学上更容易处理)。

- **统计层面**：使用概率模型来量化产生输入图像的未知量先验似然率和噪声测量过程，然后推断所期望量的最可能的估计并分析其结果的不确定程度。使用的推断算法往往与用于逆转(科学的)图像形成过程的优化方法密切相关。
- **工程层面**：开发出易于描述和实现且已知在实践中行之有效的方法。测试这些方法，以便于了解其不足和失效模态，及其期望的计算代价(运行时的性能)。

以上这三个途径相互依存，并且贯穿本书始终。

我个人的研究和发展哲学(本书中的习题亦然)非常强调算法测试。在计算机视觉领域，提出一个算法在少数几幅图像上使某件事似乎可以做而不是把某件事做对，这太容易了。要想使算法有效，最理想的途径是使用一种“三部曲”策略。

首先，在干净的合成数据上测试算法，因为已知其精确结果。其次，在该数据上增加噪声，评测性能是怎样作为噪声水平的函数退化的。最后，在真实世界数据上测试算法，优先取自广泛输入源的数据，比如万维网上的照片。只有这样，我们才能确信该算法能够处理真实世界的复杂性，即不符合某种简化模型或假设的图像。

为了在这一过程中帮助学生，本书附带大量补充阅读材料，这些都可以在本书网站找到，网址为 <http://szeliski.org/Book>。具体资源类别(参见附录 C 的描述)如下：

- 指向万维网上可以找到的问题的常用数据集的链接；
- 指向软件库的链接，可帮助学生从基本任务入手，比如读/写图像或创建和操作图像；
- 与本书素材对应的幻灯片；
- 本书所引用的论文文献列表。

在本领域发表新论文的教师和研究人員可能对后两项资源更感兴趣，但即便是普通学生，迟早也会发现它们是很有用的。有些软件库包含广泛的计算机视觉算法的实现，能帮助你应对更难的项目(征得导师同意的情况下)。

致谢

我要感谢对本书写作有帮助的所有人，他们的研究热情、咨询和鼓励帮助我写就本书。

McGill 大学的 Steve Zucker 是第一个引导我涉足计算机视觉领域的人，他教导我们所有的学生要敢于质疑和辩论研究结果和研究方法，鼓励我攻读这个领域的研究生。

我的博士论文导师，卡内基·梅隆大学的 Takeo Kanade(金出武雄)和 Geoff Hinton，教给我良好的研究、写作和报告的基本方法。他们激发了我对视觉处理、3D 建模和统计方法的兴趣，与此同时，Larry Matthies 让我见识了卡尔曼滤波和立体匹配。

Demetri Terzopoulos 是我在涉足工业界后从事第一份研究工作的导师，他教给我成功发表论文的方法。Yvan Leclerc 和 Pascal Fua，我在斯坦福研究院(SRI International)短暂停留期间的同事，在可供选择的计算机视觉研究方法方面给予我新的观点。

在 DEC 剑桥研究实验室工作的六年里，我有幸与很多同事共同工作，包括 Ingrid Carlbom, Gudrun Klinker, Keith Waters, Richard Weiss, Stephane Lavallee 和 Sing Bing Kang(江胜明)，同时也指导了最初的一大批杰出的暑期实习生，包括 David Tonnesen, Sing Bing Kang(江胜明), James Coughlan, Harry Shum(沈向洋)。正是在这里，我与 Daniel Scharstein 就此开始长期合作，他目前在 Middlebury 学院工作。

在微软研究院，我非常荣幸能和世界上最好的一些计算机视觉与计算机图形学领域的研究员一起工作，他们是：Michael Cohen, Hugues Hoppe, Stephen Gortler, Steve Shafer, Matthew Turk, Harry Shum(沈向洋), Anandan, Phil Torr, Antonio Criminisi, Georg Petschnigg, Kentaro Toyama, Ramin Zabih, Shai Avidan, Sing Bing Kang(江胜明), Matt Uyttendaele, Patrice Simard, Larry Zitnick, Richard Hartley, Simon Winder, Drew Steedly, Chris Pal, Nebojsa Jojic, Patrick Baudisch, Dani Lischinski, Matthew Brown, Simon Baker, Michael Goesele, Eric Stollnitz, David Nistér, Blaise Aguera y Arcas, Sudipta Sinha, Johannes Kopf, Neel Joshi, Krishnan Ramnath。我也非常幸运能有如此杰出的实习生，他们是 Polina Golland, Simon Baker, Mei Han(韩玫), Arno Schödl, Ron Dror, Ashley Eden, Jinxiang Chai(柴金祥), Rahul Swaminathan, Yanghai Tsing(秦漾海), Sam Hasinoff, Anat Levin, Matthew Brown, Eric Bennett, Vaibhav Vaish, Jan-Michael Frahm, James Diebel, Ce Liu(刘策), Josef Sivic, Grant Schindler, Colin Zheng, Neel Joshi, Sudipta Sinha, Zeev Farbman, Rahul Garg, Tim Cho, Yekeun Jeong, Richard Roberts, Varsha Hedau, Dilip Krishnan。

在微软工作时，我还有机会与在华盛顿大学的杰出的同事合作，我是该校的合聘教授。我要感谢 Tony DeRose 和 David Salesin，是他们最初鼓励我参与华盛顿大学正在进行的研究。我要感谢我的长期合作者 Brian Curless, Steve Seitz, Maneesh Agrawala, Sameer Agarwal, Yasu Furukawa。还要感谢我指导且和我配合很好的学生，他们是：Frederic Pighin, Yung-Yu Chuang, Doug Zongker, Colin Zheng, Aseem Agarwala, Dan Goldman, Noah Snavely, Rahul Garg, Ryan Kaminsky。正如序开始时所提到的，本书发端于 Steve Seitz 邀请我一起讲授的视觉课程，源于 Steve 的鼓励、课程笔记和编辑输入。

我还要感谢许多其他的计算机视觉研究人员，他们给了我很多关于本书的建设性建议，包括：Sing Bing Kang(江胜明)，他是我的非正式版图书编辑；Vladimir Kolmogorov，他撰写了关于 MRF 推断的线性规划方法之附录 B.5.5；Daniel Scharstein, Richard Hartley, Simon Baker, Noah Snavely, Bill Freeman, Svetlana

Lazebnik, Matthew Turk, Jitendra Malik, Alyosha Efros, Michael Black, Brian Curless, Sameer Agarwal, Li Zhang(张力), Deva Ramanan, Olga Veksler, Yuri Boykov, Carsten Rother, Phil Torr, Bill Triggs, Bruce Maxwell, Jana Kosecka, Eero Simoncelli, Aaron Hertzmann, Antonio Torralba, Tomaso Poggio, Theo Pavlidis, Baba Vemuri, Nando de Freitas, Chuck Dyer, Song Yi(宋毅), Falk Schubert, Roman Pflugfelder, Marshall Tappen, James Coughlan, Sammy Rogmans, Klaus Strobel, Shanmuganathan, Andreas Siebert, Yongjun Wu(吴勇军), Fred Pighin, Juan Cockburn, Ronald Mallet, Tim Soper, Georgios Evangelidis, Dwight Fowler, Itzik Bayaz, Daniel O'Connor, Srikrishna Bhat. Shena Deuchers 极为出色地完成了本书的排版编辑工作, 提出了很多有价值的改进建议。Springer 出版社的 Wayne Wheeler 和 Simon Rees 在本书整个出版过程中对我帮助很大。Keith Price 的 “Annotated Computer Vision Bibliography” (注解版计算机视觉参考文献) 对追溯参考文献和查找相关工作起着非常重要的作用。

我期望这本书准确、信息可靠和及时, 所以如果你有任何改进本书的建议, 不妨发电子邮件告诉我。

最后, 没有家人难以置信的支持和鼓励, 本书不可能问世, 或者说不值得费时费力。谨以此书献给我的父母, Zdzisław 和 Jadwiga, 他们的爱、慷慨和成就总是激励着我; 献给我的妹妹 Basia, 因为她对我所付出的毕生的手足之情; 特别献给 Lyn, Anne 和 Stephen, 他们在所有事情(包括本书这个项目)上每天都给予我鼓励, 使所有事情都有非凡的价值。

Wenatchee 湖畔

2010 年 8 月

目 录

第 1 章 概述.....1	
1.1 什么是计算机视觉?2	
1.2 简史8	
1.3 本书概述.....16	
1.4 课程大纲样例21	
1.5 标记法说明22	
1.6 扩展阅读.....22	
第 2 章 图像形成.....25	
2.1 几何基元和变换26	
2.1.1 几何基元26	
2.1.2 2D 变换29	
2.1.3 3D 变换32	
2.1.4 3D 旋转33	
2.1.5 3D 到 2D 投影.....37	
2.1.6 镜头畸变.....46	
2.2 光度测定学的图像形成.....47	
2.2.1 照明.....48	
2.2.2 反射和阴影49	
2.2.3 光学.....54	
2.3 数字摄像机.....57	
2.3.1 采样与混叠60	
2.3.2 色彩.....63	
2.3.3 压缩.....71	
2.4 补充阅读.....72	
2.5 习题73	
第 3 章 图像处理.....77	
3.1 点算子78	
3.1.1 像素变换79	
3.1.2 彩色变换81	
3.1.3 合成与抠图81	
3.1.4 直方图均衡化83	
3.1.5 应用: 色调调整..... 86	
3.2 线性滤波 86	
3.2.1 可分离的滤波..... 89	
3.2.2 线性滤波示例..... 90	
3.2.3 带通和导向滤波器 91	
3.3 更多的邻域算子 95	
3.3.1 非线性滤波 95	
3.3.2 形态学 99	
3.3.3 距离变换 100	
3.3.4 连通量 101	
3.4 傅里叶变换 102	
3.4.1 傅里叶变换对..... 105	
3.4.2 二维傅里叶变换..... 107	
3.4.3 维纳滤波 108	
3.4.4 应用: 锐化, 模糊 和去噪 111	
3.5 金字塔与小波 111	
3.5.1 插值..... 112	
3.5.2 降采样 114	
3.5.3 多分辨率表达..... 116	
3.5.4 小波 119	
3.5.5 应用: 图像融合 123	
3.6 几何变换 125	
3.6.1 参数化变换 125	
3.6.2 基于网格的卷绕..... 131	
3.6.3 应用: 基于特征的变形 .. 133	
3.7 全局优化 133	
3.7.1 正则化 134	
3.7.2 马尔科夫随机场 138	
3.7.3 应用: 图像的恢复 147	
3.8 补充阅读..... 147	
3.9 习题..... 149	

第 4 章 特征检测与匹配157	5.4 规范图割..... 227
4.1 点和块.....159	5.5 图割和基于能量的方法..... 230
4.1.1 特征检测器.....160	5.6 补充阅读..... 234
4.1.2 特征描述子.....169	5.7 习题..... 235
4.1.3 特征匹配.....172	第 6 章 基于特征的配准 237
4.1.4 特征跟踪.....179	6.1 基于 2D 和 3D 特征的配准..... 238
4.1.5 应用: 表演驱动的动画.....181	6.1.1 使用最小二乘的 2D 配准..... 238
4.2 边缘.....182	6.1.2 应用: 全景图..... 240
4.2.1 边缘检测.....182	6.1.3 迭代算法..... 241
4.2.2 边缘连接.....187	6.1.4 鲁棒最小二乘 和 RANSAC..... 243
4.2.3 应用: 边缘编辑和增强.....189	6.1.5 3D 配准..... 245
4.3 线条.....190	6.2 姿态估计..... 246
4.3.1 逐次近似.....191	6.2.1 线性算法..... 246
4.3.2 Hough 变换.....191	6.2.2 迭代算法..... 248
4.3.3 消失点.....194	6.2.3 应用: 增强现实..... 249
4.3.4 应用: 矩形检测.....196	6.3 几何内参数标定..... 250
4.4 扩展阅读.....197	6.3.1 标定模式..... 250
4.5 习题.....198	6.3.2 消失点..... 252
第 5 章 分割205	6.3.3 应用: 单视图测量学..... 253
5.1 活动轮廓.....206	6.3.4 旋转运动..... 254
5.1.1 蛇行.....207	6.3.5 径向畸变..... 256
5.1.2 动态蛇行和 CONDENSATION.....211	6.4 补充阅读..... 257
5.1.3 剪刀.....214	6.5 习题..... 258
5.1.4 水平集.....215	第 7 章 由运动到结构 263
5.1.5 应用: 轮廓跟踪和 转描机.....217	7.1 三角测量..... 264
5.2 分裂与归并.....218	7.2 二视图由运动到结构..... 266
5.2.1 分水岭.....218	7.2.1 投影(未标定的)重建..... 270
5.2.2 区域分裂(区分式聚类).....219	7.2.2 自标定..... 271
5.2.3 区域归并(凝聚式聚类).....219	7.2.3 应用: 视图变形..... 273
5.2.4 基于图的分割.....219	7.3 因子分解..... 274
5.2.5 概率聚集.....220	7.3.1 透视与投影因子分解..... 276
5.3 均值移位和模态发现.....221	7.3.2 应用: 稀疏 3D 模型 提取..... 277
5.3.1 k-均值和高斯混合.....222	
5.3.2 均值移位.....224	

7.4 光束平差法.....278	9.1.4 缝隙消除.....333
7.4.1 挖掘稀疏性.....280	9.1.5 应用: 视频摘要和压缩...334
7.4.2 应用: 匹配运动和增强 现实.....282	9.1.6 圆柱面和球面坐标.....335
7.4.3 不确定性和二义性.....283	9.2 全局配准.....338
7.4.4 应用: 由因特网照片 重建.....284	9.2.1 光束平差法.....338
7.5 限定结构和运动.....287	9.2.2 视差消除.....341
7.5.1 基于线条的方法.....287	9.2.3 认出全景图.....343
7.5.2 基于平面的方法.....288	9.2.4 直接配准和基于特征的 配准.....345
7.6 补充阅读.....289	9.3 合成.....346
7.7 习题.....290	9.3.1 合成表面的选择.....346
第8章 稠密运动估计.....293	9.3.2 像素选择和加权 (去虚影).....348
8.1 平移配准.....294	9.3.3 应用: 照片蒙太奇.....352
8.1.1 分层运动估计.....297	9.3.4 融合.....353
8.1.2 基于傅里叶的配准.....298	9.4 补充阅读.....355
8.1.3 逐次求精.....300	9.5 习题.....356
8.2 参数化运动.....305	第10章 计算摄影学.....359
8.2.1 应用: 视频稳定化.....308	10.1 光度学标定.....361
8.2.2 学到的运动模型.....308	10.1.1 辐射度响应函数.....362
8.3 基于样条的运动.....309	10.1.2 噪声水平估计.....363
8.4 光流.....312	10.1.3 虚影.....364
8.4.1 多帧运动估计.....315	10.1.4 光学模糊(空间响应) 估计.....365
8.4.2 应用: 视频去噪.....316	10.2 高动态范围成像.....368
8.4.3 应用: 去隔行扫描.....316	10.2.1 色调映射.....374
8.5 层次运动.....317	10.2.2 应用: 闪影术.....380
8.5.1 应用: 帧插值.....319	10.3 超分辨率和模糊去除.....381
8.5.2 透明层和反射.....320	10.3.1 彩色图像去马赛克.....385
8.6 补充阅读.....321	10.3.2 应用: 彩色化.....387
8.7 习题.....322	10.4 图像抠图和合成.....388
第9章 图像拼接.....327	10.4.1 蓝屏抠图.....389
9.1 运动模型.....329	10.4.2 自然图像抠图.....391
9.1.1 平面透视运动.....329	10.4.3 基于优化的抠图.....394
9.1.2 应用: 白板和文档扫描...330	10.4.4 烟、阴影和闪抠图.....396
9.1.3 旋转全景图.....331	10.4.5 视频抠图.....397

10.5 纹理分析与合成	398	12.2.2 应用: 数字遗产	453
10.5.1 应用: 空洞填充 与修图	400	12.3 表面表达	454
10.5.2 应用: 非真实感绘制	401	12.3.1 表面插值	454
10.6 补充阅读	403	12.3.2 表面简化	455
10.7 习题	404	12.3.3 几何图像	456
第 11 章 立体视觉对应	409	12.4 基于点的表达	456
11.1 极线几何学	412	12.5 体积表达	457
11.1.1 矫正	412	12.6 基于模型的重建	459
11.1.2 平面扫描	414	12.6.1 建筑结构	459
11.2 稀疏对应	416	12.6.2 头部和人脸	461
11.3 稠密对应	418	12.6.3 应用: 脸部动画	463
11.4 局部方法	420	12.6.4 完整人体建模与跟踪	465
11.4.1 亚像素估计 与不确定性	422	12.7 恢复纹理映射与反照率	469
11.4.2 应用: 基于立体视觉的 头部跟踪	423	12.7.1 估计 BRDF	470
11.5 全局优化	424	12.7.2 应用: 3D 摄影学	471
11.5.1 动态规划	425	12.8 补充阅读	472
11.5.2 基于分割的方法	427	12.9 习题	473
11.5.3 应用: z-键控与背景 替换	428	第 13 章 基于图像的绘制	477
11.6 多视图立体视觉	429	13.1 视图插值	478
11.6.1 体积与 3D 表面重建	432	13.1.1 视图相关的纹理映射	480
11.6.2 由轮廓到形状	436	13.1.2 应用: 照片游览	481
11.7 补充阅读	438	13.2 层次深度图像	482
11.8 习题	439	13.3 光场与发光图	484
第 12 章 3D 重建	443	13.3.1 非结构化发光图	487
12.1 由 X 到形状	444	13.3.2 表面光场	488
12.1.1 由阴影到形状与光度 测量立体视觉	445	13.3.3 应用: 同心拼图	489
12.1.2 由纹理到形状	447	13.4 环境影像形板	490
12.1.3 由聚焦到形状	448	13.4.1 更高维光场	491
12.2 主动距离获取	449	13.4.2 从建模到绘制	492
12.2.1 距离数据归并	451	13.5 基于视频的绘制	493
		13.5.1 基于视频的动画	493
		13.5.2 视频纹理	494
		13.5.3 应用: 图片动画	497
		13.5.4 3D 视频	497

13.5.5 应用: 基于视频的 浏览.....	499	A.1.1 奇异值分解.....	570
13.6 补充阅读.....	501	A.1.2 特征值分解.....	571
13.7 习题.....	503	A.1.3 QR 因子分解.....	573
第 14 章 识别	507	A.1.4 乔里斯基分解.....	574
14.1 物体检测.....	509	A.2 线性最小二乘.....	575
14.1.1 人脸检测.....	509	A.3 非线性最小二乘.....	578
14.1.2 行人检测.....	515	A.4 直接稀疏矩阵方法.....	579
14.2 人脸识别.....	518	A.5 迭代方法.....	580
14.2.1 特征脸.....	518	A.5.1 共轭梯度.....	581
14.2.2 活动表观与 3D 形状 模型.....	525	A.5.2 预处理.....	582
14.2.3 应用: 个人照片收藏.....	528	A.5.3 多重网格.....	583
14.3 实例识别.....	529	附录 B 贝叶斯建模与推断	585
14.3.1 几何配准.....	530	B.1 估计理论.....	586
14.3.2 大型数据库.....	531	B.2 最大似然估计与最小二乘.....	589
14.3.3 应用: 位置识别.....	535	B.3 鲁棒统计学.....	590
14.4 类别识别.....	537	B.4 先验模型与贝叶斯推断.....	591
14.4.1 词袋.....	539	B.5 马尔科夫随机场.....	592
14.4.2 基于部件的模型.....	542	B.5.1 梯度下降与模拟退火.....	594
14.4.3 基于分割的识别.....	545	B.5.2 动态规划.....	595
14.4.4 应用: 智能照片编辑.....	548	B.5.3 置信传播.....	596
14.5 上下文与场景理解.....	550	B.5.4 图割.....	598
14.5.1 学习与大型图像收集.....	552	B.5.5 线性规划.....	601
14.5.2 应用: 图像搜索.....	554	B.6 不确定性估计(误差分析).....	602
14.6 识别数据库和测试集.....	555	附录 C 补充材料	604
14.7 补充阅读.....	559	C.1 数据集.....	605
14.8 习题.....	562	C.2 软件.....	607
第 15 章 结语	567	C.3 幻灯片与讲座.....	615
附录 A 线性代数与数值方法	569	C.4 参考文献.....	615
A.1 矩阵分解.....	570	词汇表	617

第 1 章

概 述

- 1.1 什么是计算机视觉？
- 1.2 简史
- 1.3 本书概述
- 1.4 课程大纲样例
- 1.5 标记法说明
- 1.6 扩展阅读



图 1.1 人类视觉系统可以毫不费力地解释这张照片中存在的由半透明材质和阴影形成的微妙变化，并从背景中将物体正确地分割出来

1.1 什么是计算机视觉？

作为人，我们显然很容易感知周围世界的三维结构。想一想，当我们观察身边桌上的一瓶花时，三维的感知是怎样的生动鲜明。通过光线和阴影在其表面显现出来的细微模式，可以分辨出每个花瓣的形状和其半透明性，毫不费力地从场景的背景中将每朵花分割出来(图 1.1)。观看一幅有边框的合影时，可以轻松数出照片中所有的人(并说出名字)，甚至可以从他们的面部外观猜出其情感状态。感知心理学家已经花了几十年时间试图理解视觉系统是如何工作的，尽管他们能够想出光学错觉^①来梳理其原理的某些部分(图 1.3)，但这个难题的完整解答仍然扑朔迷离(Marr 1982; Palmer 1999; Livingstone 2008)。

计算机视觉领域的研究人员同时也一直在研究恢复影像中物体的三维形状和外观的数学方法。目前我们已经有了可靠的方法能够从几千幅部分重叠的照片精确地计算出环境的部分 3D 模型(图 1.2a)。如果有特定物体或建筑物正面足够多的一组视图，我们就可以使用立体匹配方法创建出稠密的 3D 表面模型(图 1.2b)。我们可以在复杂的背景中跟踪运动的人(图 1.2c)。使用人脸、衣服、头发的检测和识别相结合的方法，我们甚至可以试图找到照片中所有的人并说出他们的名字，此项任务可以获得中等的成功率(图 1.2d)。尽管有这些进展，但要让计算机解释图像的能力与两岁大的孩子(例如，数出照片中所有的动物)有一样的水平，这一梦想仍然是难以实现的。视觉为什么如此困难？部分因为它是一个逆问题(inverse problem)，在信息不足的情况下，我们试图恢复一些未知量来给出完整的解答。因此，我们必须求助于基于物理的和基于概率的模型(model)来消除潜在解的歧义。然而，视觉世界的建模就其十分的复杂性远比产生语音的声道建模更困难。

在计算机视觉领域，我们使用的前向(forward)模型通常是在物理学(辐射测量学、光学和传感器设计)或计算机图形学中发展而来的。这两个领域都是在对如下过程进行建模：物体是如何运动和展示的，光线是怎样从表面反射、由空气散射、经由摄像机镜头(或人的眼睛)折射而最后投影到平的(或弯的)图像面上的。尽管计算机图形学还不完美(还没有完全由计算机生成的含有人物角色的电影可以成功地跨越这非比寻常的峡谷(uncanny valley)^②，它将真正的人与机器人和计算机动画的人相区隔)，在有限的领域，比如渲染由日常物体构成的静态场景或像恐龙那样灭绝了的动物，真实感的幻觉是很完美的。

在计算机视觉领域，我们在试图做反过程，即描述我们从一幅或多幅图像中看到的世界，比如形状、照明和色彩分布。令人惊异的是，人和动物可以毫不费力地完成，而计算机视觉算法却很容易出错。没有在该领域工作过的人常低估该问题的

^① http://www.michaelbach.de/ot/sze_muelue

^② “非比寻常的峡谷”(uncanny valley)一词是机器人学家 Masahiro Mori 创造出来用于机器人学的(Mori 1970)。它也常用于计算机制作的动画电影中，比如《最终幻想》和《极地快车》(Geller 2008)。

困难程度。(工作中同事经常向我要软件,用来找到照片中所有的人并说出他们的名字,使他们可以继续做更“有趣”的工作。)”“视觉应该是简单的”这一误解可以追溯到人工智能的早期岁月(参见 1.2 节),当时,人们最初认为智能的认知部分(逻辑证明和规划),本质上比感知部分更困难(Boden 2006)。

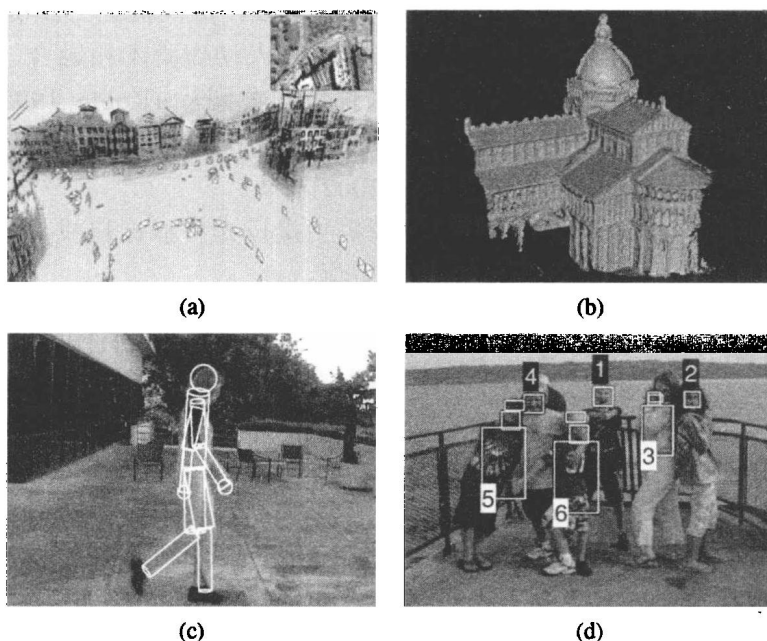


图 1.2 计算机视觉算法和应用的一些例子。(a)由运动到结构算法可以从几千幅部分重叠的照片重建起大规模复杂场景的稀疏 3D 点模型(Snavely, Seitz, and Szeliski 2006) 2006 ACM。(b)立体匹配算法可以从取自因特网的几百幅不同曝光的照片建立建筑物正面外观的详细 3D 模型(Goesele, Snavely, Curless *et al.* 2007)© 2007 IEEE。(c)人的跟踪算法可以在零乱的背景中跟踪行走的人(Sidenbladh, Black, and Fleet 2000)© 2000 Springer。(d)人脸检测算法,与基于颜色的衣服和头发检测算法相配合,可以定位并识别图像中的个人(Sivic, Zitnick, and Szeliski 2006)© 2006 Springer

好消息是计算机视觉如今正广泛应用于各种各样的实际应用之中,如下所示。

- **光学字符识别(OCR):** 阅读信上的手写邮政编码(图 1.4a) 和自动号码牌识别(ANPR)。
- **机器检验:** 为保证质量而快速检验部件,使用立体视觉在专门的光照下测量飞机机翼或汽车车身配件的容差(图 1.4b),或使用 X 光视觉检查钢铸件的缺欠。
- **零售:** 针对自动结账通道的物体识别(图 1.4c)。
- **3D 模型建立(摄影测量学):** 在诸如 Bing Maps 等系统中使用的从航空照片全自动地构建 3D 模型。
- **医学成像:** 注册手术前和手术中的成像(图 1.4d),或进行关于人老化过程中大脑形态的长期研究。

- **汽车安全**：在诸如雷达或激光雷达等主动视觉技术不好用时检测意外的障碍物，比如街道上的行人(图 1.4e，完全自动驾驶的例子请参见 Miller, Campbell, Huttenlocher *et al.* (2008); Montemerlo, Becker, Bhat *et al.* (2008); Urmson, Anhalt, Bagnell *et al.* (2008))。
- **匹配运动**：通过跟踪源视频中的特征点来估计摄像机的 3D 运动 and 环境的形状，将计算机生成的影像(CGI)与实景真人动作脚本相融合。这样的技术在好莱坞得到广泛使用(例如电影《侏罗纪公园》)(Roble 1999; Roble and Zafar 2009)；他们还需要使用精确的抠图来在前景和背景之间插入新的元素(Chuang, Agarwala, Curless *et al.* 2002)。
- **运动捕捉(mocap)**：使用从多台摄像机拍摄反光材料标记或其他视觉方法来捕捉演员的动作，以便用于计算机动画。
- **监视**：监控入侵者，分析高速公路的交通状况(图 1.4f)，监控游泳池以防溺水事件的发生。
- **指纹识别和生物测定学**：用于自动准入身份验证以及司法应用。

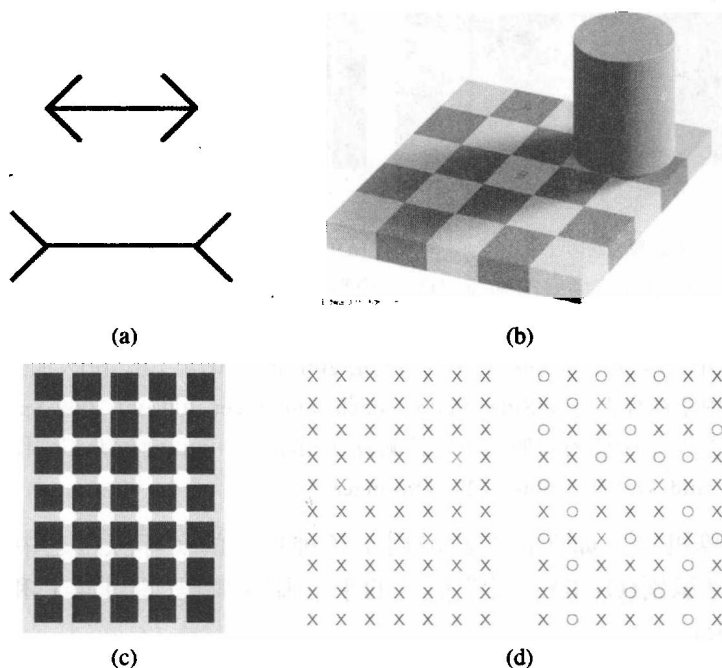


图 1.3 一些常见的光学错觉和其可以给我们关于视觉系统的启示。(a)经典的 Muller-Lyer 错觉，两条水平线条的长度显得不同，可能是由于想象的透视效果。(b)阴影中的“白”块 B 和亮处的“黑”块 A 实际上具有相同的绝对亮度值。该感觉是由于亮度恒常性，即视觉系统当解释颜色时试图减少亮度的特性。图像承蒙 Ted Adelson 的允许，http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html。(c)Hermann 网格错觉的一个变型，承蒙 Hany Farid 许可使用，<http://www.cs.dartmouth.edu/~farid/illusions/hermann.html>。当你移动眼睛扫视该图时，灰色的斑点会出现在交叉处。(d)在图的左半部分数一数红色的 X。现在，在右半部分数。是不是明显困难了？其解释与弹出(pop-out)效应有关 (Treisman 1985)，它告诉我们大脑中有关并行感知和集成通道的操作