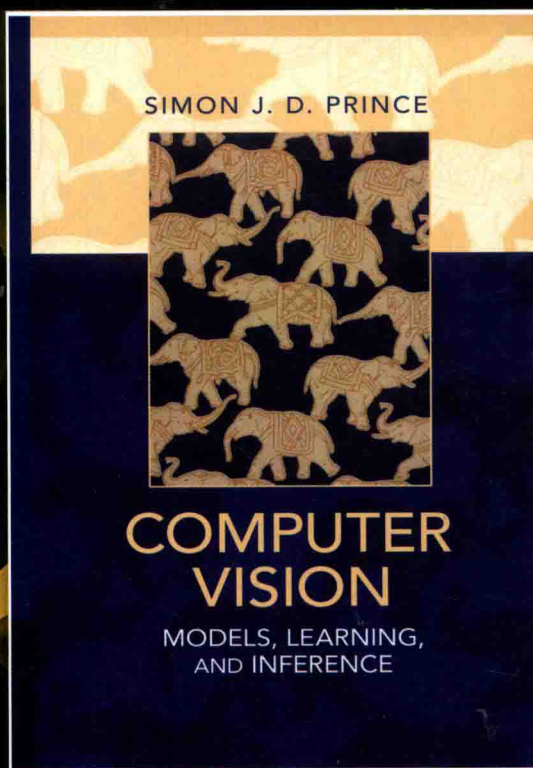


计算机视觉

模型、学习和推理

[英] 西蒙 J. D. 普林斯 (Simon J. D. Prince) 著
苗启广 刘凯 孔韦韦 许鹏飞 译

Computer Vision
Models, Learning, and Inference



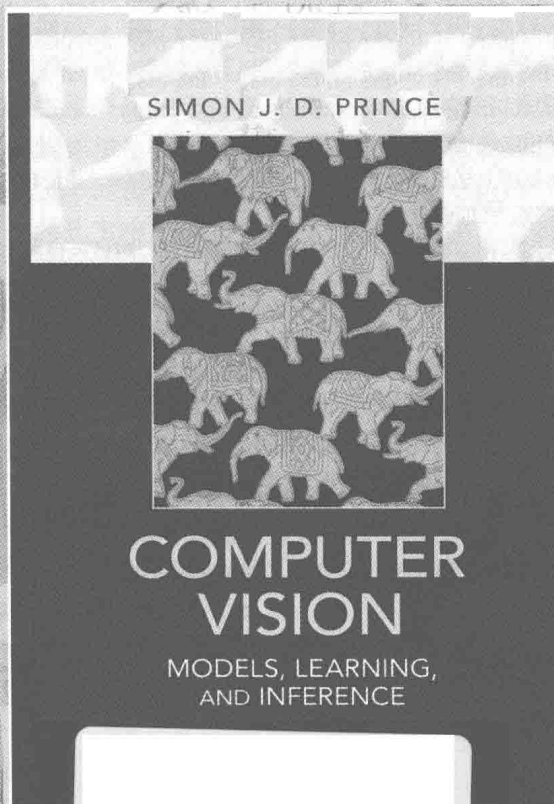
计 算 机 科 学 丛 书

计算机视觉

模型、学习和推理

[英] 西蒙 J. D. 普林斯 (Simon J. D. Prince) 著
苗启广 刘凯 孔韦韦 许鹏飞 译

Computer Vision
Models, Learning, and Inference



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

计算机视觉: 模型、学习和推理 / (英) 普林斯 (Prince, J. D.) 著; 苗启广等译. —北京: 机械工业出版社, 2015.9

(计算机科学丛书)

书名原文: Computer Vision: Models, Learning, and Inference

ISBN 978-7-111-51682-8

I. 计… II. ①普… ②苗… III. 计算机视觉—研究 IV. TP302.7

中国版本图书馆 CIP 数据核字 (2015) 第 232990 号

本书版权登记号: 图字: 01-2015-0868

This is a Chinese simplified edition of the following title published by Cambridge University Press:

Simon J. D. Prince, Computer Vision: Models, Learning, and Inference, 978-1-107-01179-3.

© Simon J. D. Prince 2012.

This Chinese simplified edition for the People's Republic of China (excluding Hong Kong, Macau and Taiwan) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press and China Machine Press in 2017.

This Chinese simplified edition is authorized for sale in the People's Republic of China (excluding Hong Kong, Macau and Taiwan) only. Unauthorized export of this simplified Chinese is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of Cambridge University Press and China Machine Press.

本书原版由剑桥大学出版社出版。

本书简体字中文版由剑桥大学出版社与机械工业出版社合作出版。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。

此版本仅限在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 销售。

本书是一本从机器学习视角讲解计算机视觉的优秀教材, 主要讲述计算机视觉中模型、学习和推理三个方面的内容, 揭示计算机视觉研究中“模型”和“算法”之间的区别, 并且对每一种新的视觉算法提出新的见解。本书图文并茂, 算法描述由浅入深, 主要包括概率、机器视觉的机器学习、局部模型、图像预处理、几何模型、视觉模型等方面的内容, 适合作为高年级本科生或研究生的计算机视觉和机器学习教材, 也可供计算机视觉方面的专业人士参考。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 谢晓芳 盛思源 缪杰

责任校对: 殷虹

印刷: 北京诚信伟业印刷有限公司

版次: 2017 年 6 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 29.25 (含 0.5 印张彩插)

书号: ISBN 978-7-111-51682-8

定价: 119.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

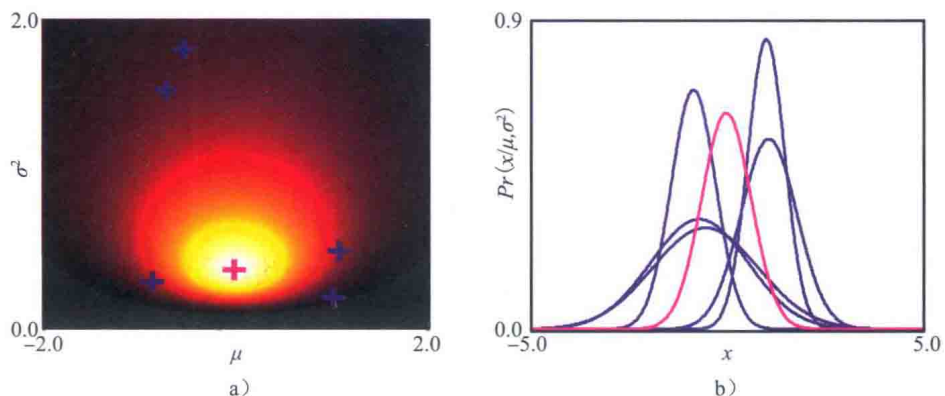


图 4-4 正态分布中参数的先验值。a) α 、 β 、 $\gamma = 1$ ， $\delta = 0$ 的正态逆伽马提供一个一元正态分布参数的宽先验分布。红色的十字表示这个先验分布的峰值。蓝色的十字是随机从分布中抽出的 5 个样本。b) 峰值和样本可以通过绘出它们所代表的正态分布来直观表示

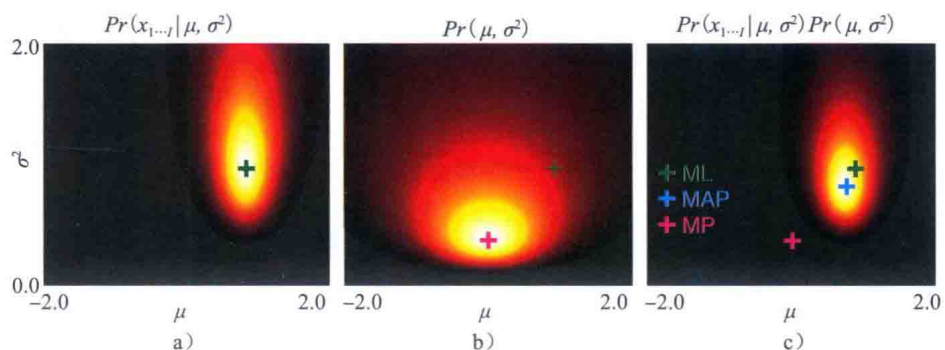


图 4-5 正态参数的最大后验推理。a) 乘以似然函数 b) 给出一个新函数的先验概率 c) 这与后验分布成正比。可以看到最大后验 (MAP) 解 (青色的十字) 在后验分布的峰值。它在最大似然 (ML) 解 (绿色的十字) 和最大前验分布 (MP, 粉红色的十字) 之间

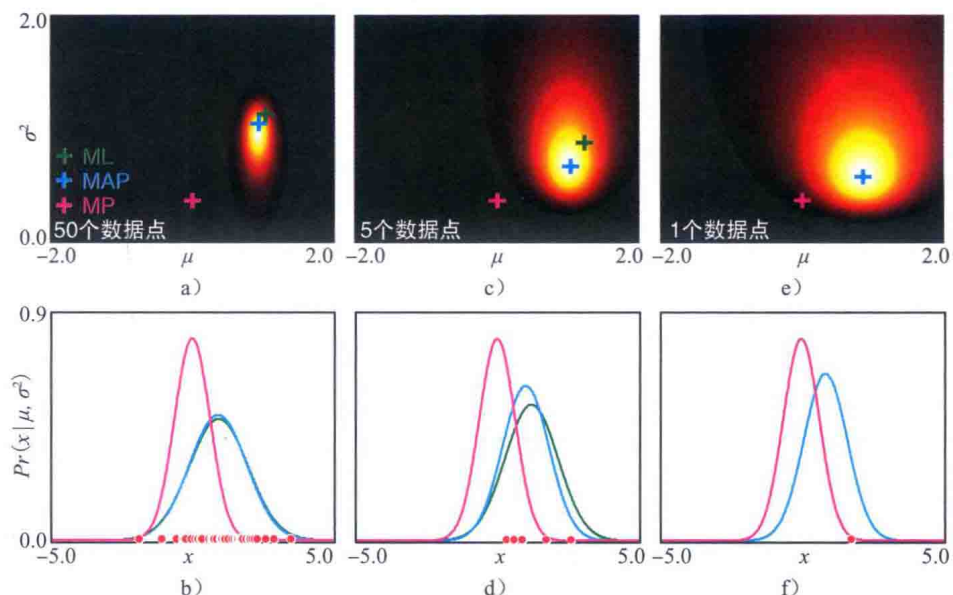


图 4-6 最大后验估计。a) 参数 μ 和 σ^2 的后验分布。MAP 解 (青色的十字) 位于 ML (绿色的十字) 和先验值顶点 (粉红色的十字) 之间。b) MAP 解、ML 解和先验值的峰值对应的正态分布。c ~ d) 由于具有很少的数据点，因此先验值对最终解有更大的影响。e ~ f) 由于只有一个数据点，因此最大似然解不能计算 (不能计算单个点的方差)。然而，MAP 估计依然可以计算

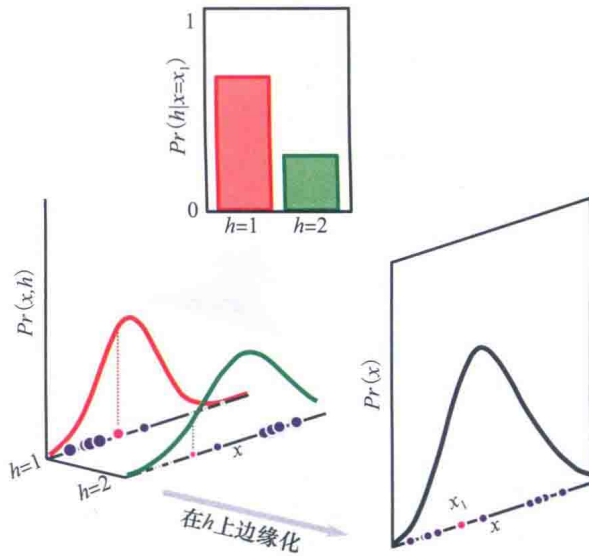


图 7-8 拟合混合高斯模型的 E 步。对于 I 个数据点 $\mathbf{x}_1, \dots, \mathbf{x}_I$ 中的每一个，计算关于隐变量 h_i 的后验分布 $\Pr(h_i | \mathbf{x}_i)$ 。 h_i 取 k 时的后验概率 $\Pr(h_{i=k} | \mathbf{x}_i)$ 可以理解为正态分布 k 对于数据点 x_i 的贡献。例如，对于数据点 x_1 (粉红色圆点)，分量 1 (红色曲线) 比分量 2 (绿色曲线) 的贡献率大于 2 倍。注意，在联合分布 (左侧) 中，投影数据点的大小表示了对应关系

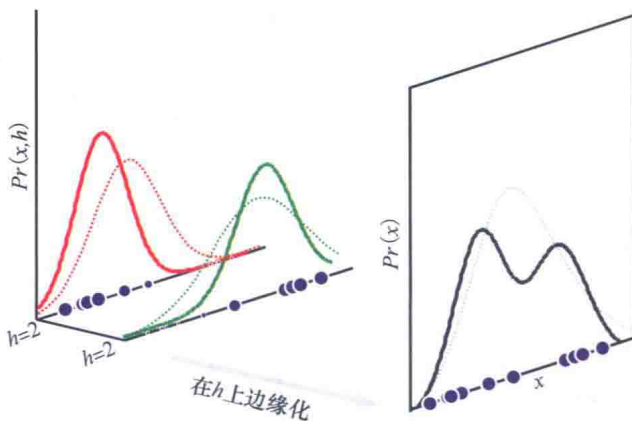


图 7-9 拟合混合高斯模型的 M 步。对于第 k 个高斯分量，更新参数 $\{\lambda_k, \mu_k, \Sigma_k\}$ 。第 i 个数据点 \mathbf{x}_i 根据 E 步中对应关系 r_{ik} (以点的大小来表示) 以协助于这些更新；与第 k 个分量关系越密切的数据点对于参数的影响越大。虚线和实线分别表示更新前后的拟合

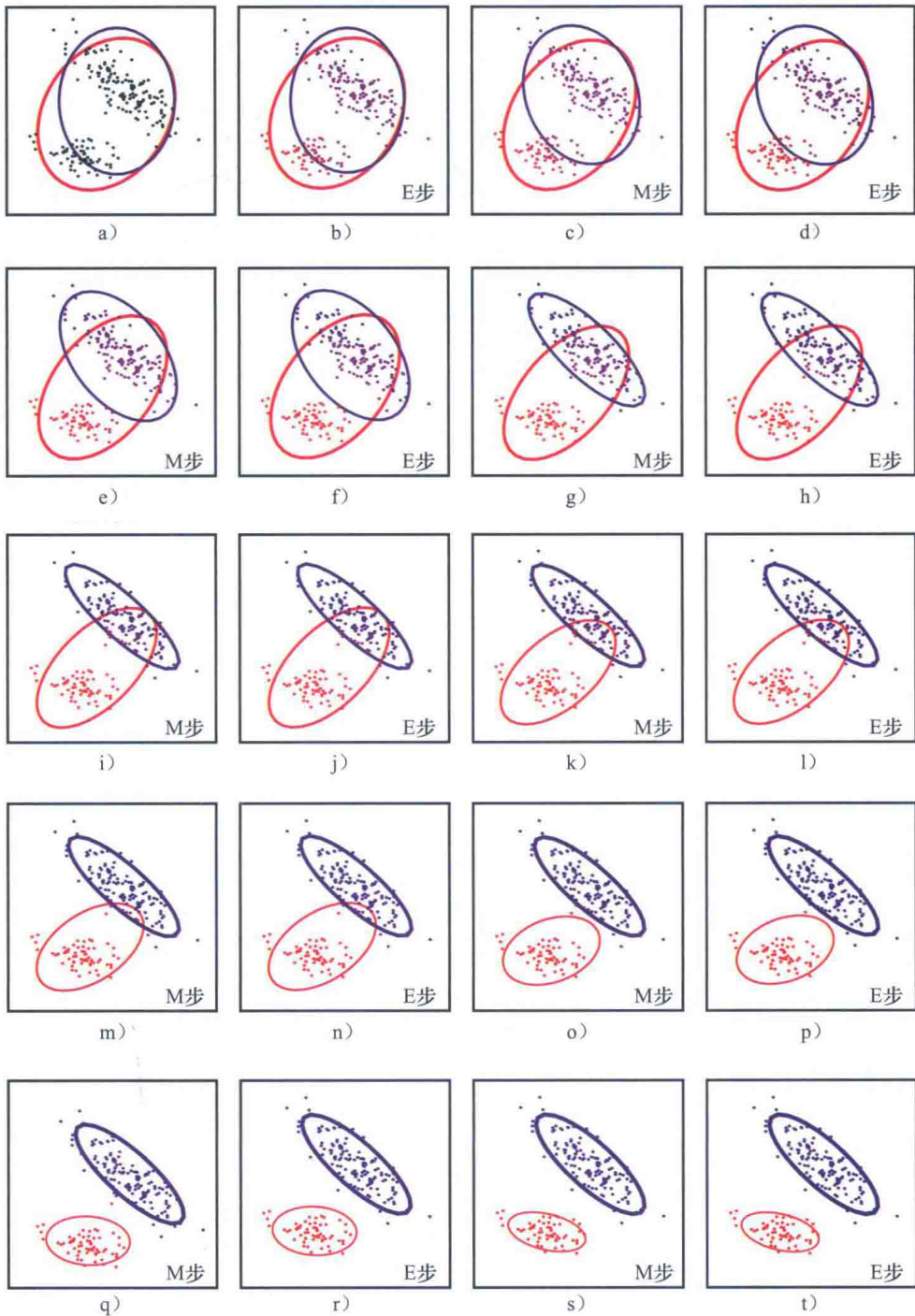


图 7-10 拟合二维数据的两个高斯混合模型。a) 原始模型。b) E 步。对于每个数据点，计算由每个高斯分布生成的后验概率（用点的颜色表示）。c) M 步。根据这些后验概率来更新每个高斯分布的均值、方差和权重。椭圆形表示两者之间的马氏距离。椭圆的权重（粗细）表示高斯权重。d ~ t) 后续的 E 步和 M 步迭代

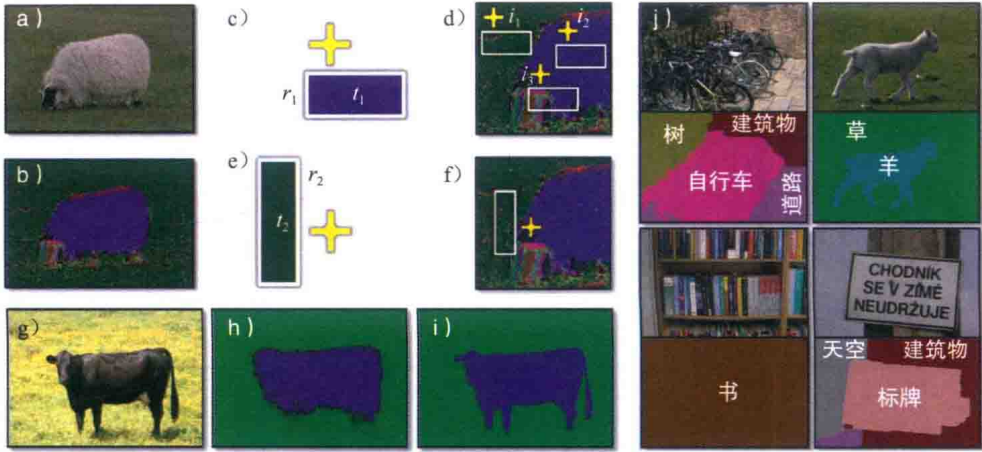


图 9-22 使用“TextonBoost”对语义图像进行标注。a) 原始图像。b) 转换为纹理基元的图像，每一个像素值都有一个离散值，表示存在该类型的纹理。c) 系统是基于弱分类器的，这种分类器统计矩形内的特定类型纹理基元数，该矩形即为当前位置的偏移（黄色十字）。d) 这不仅提供了有关对象本身（包括类似羊的纹理基元）的信息，同时也提供了附近对象（附近草的纹理基元）的相关信息。e、f) 另一个弱分类器的例子。g) 测试图像。h) 逐像素分类在物体边缘都不是很精确，因此，i) 通过一个条件随机场来改善结果。j) 结果和标准图的比较。源自 Shotton 等（2009）。© 2009 Springer

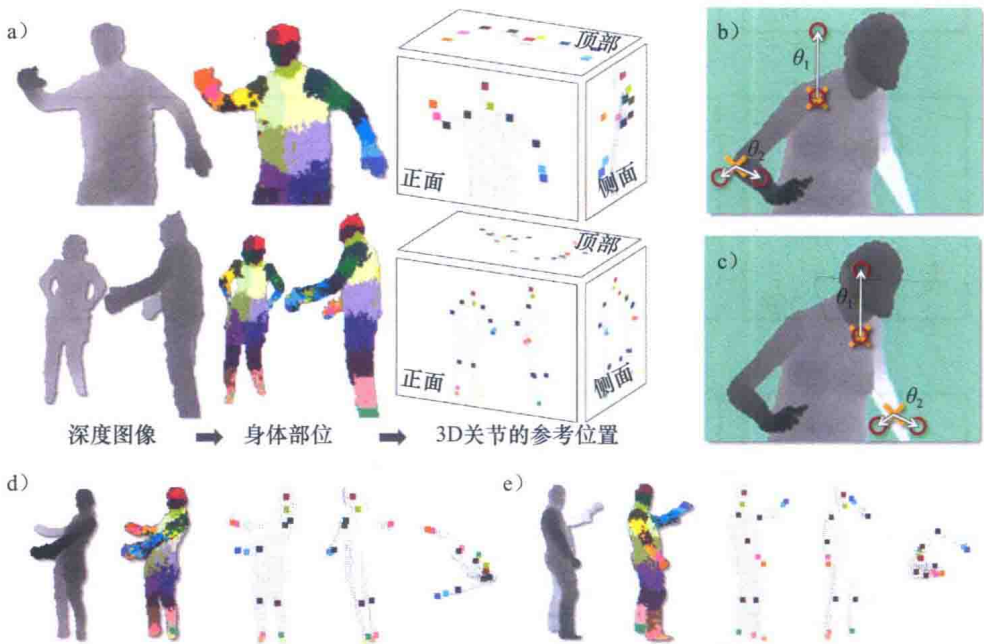


图 9-24 人体不同部位的识别。a) 该系统的目标是通过一幅深度图像 x 和分配给每一个像素的离散标签 w ，来标记可能存在的 31 个人体部位。这些深度标签可以作为 3D 关节的参考位置。b) 基于决策树的分类器，在树中的每一个点，数据都根据两点（红圈）的相对深度以及和当前像素的偏移量（黄色十字）进行划分。在这个例子中，两个情况下有很大的区别，而在 c) 中这种区别是很小的——因此这些差别提供了姿态信息，d、e) 是另外两个深度图像、标签标记和姿态预测的例子。源自 Shotton 等（2011）。© 2011 IEEE

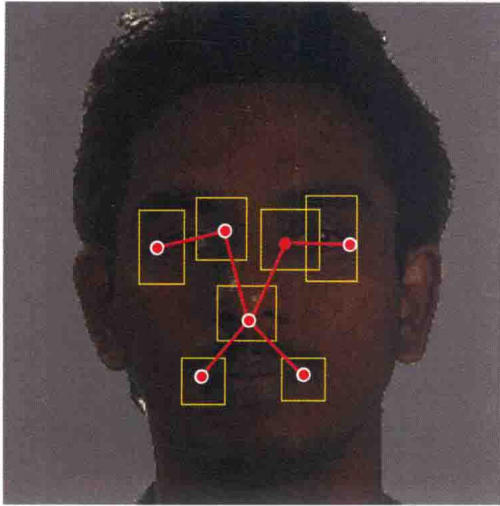


图 11-19 形象化结构。此脸部模型包括 7 个部分（红点），以树状结构（红线）连接在一起。每个部分的可能位置由方框表示。虽然每个部分可以取几个百像素，但是 MAP 位置可以利用动态规划方法通过图的树结构有效地进行推理。局部化面部特征是许多人脸识别方法的共同元素

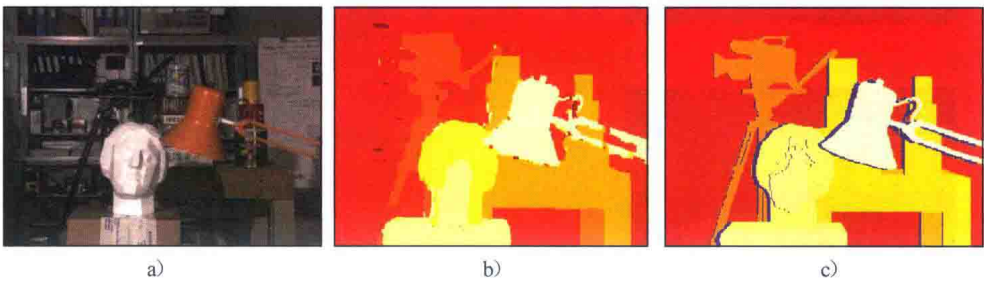


图 12-27 立体视觉。a) 原始立体对图像。b) 采用 Boykov 等（1999）方法的视差估计。c) 真实差异。蓝色像素表明在第二幅图像中被遮挡的区域，因此不具有一个有效的匹配或者视差。该算法没有对这一事实进行解释，且在那些区域中产生了噪声估计

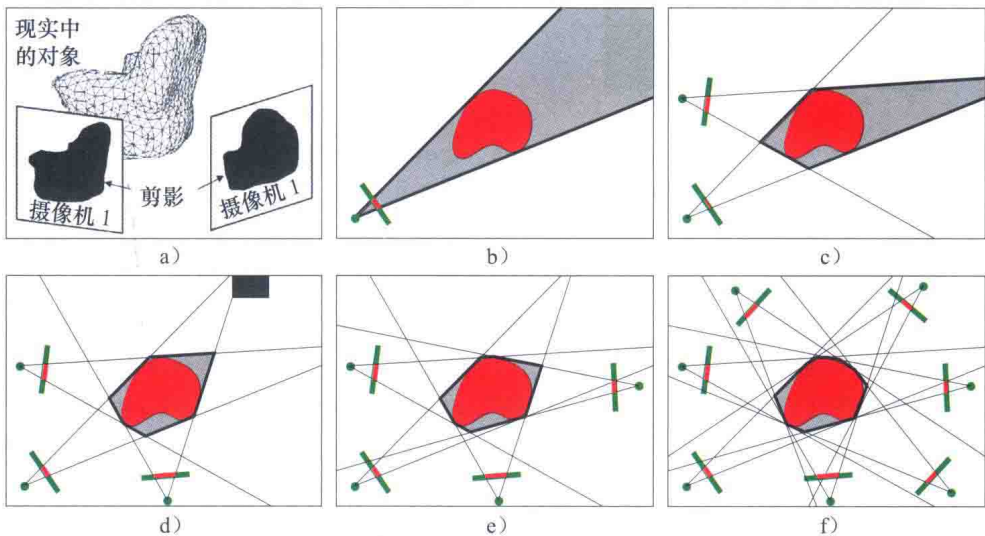


图 14-15 剪影重构。a) 目标是基于许多摄像机中的轮廓恢复出有关物体形状的信息。b) 考虑一台观察一个对象（2D 切片）的摄像机。我们知道该物体必然位于包含其轮廓的多束光线中的某一位置。c) 当我们增加第二台摄像机时，我们知道该物体依然位于包含其轮廓的多束光线的交叉区域内（灰色区域）。d ~ f) 当我们增加更多台摄像机时，就越来越接近真实的形状。但是，无论我们增加多少台摄像机，都无法捕获凹陷区域

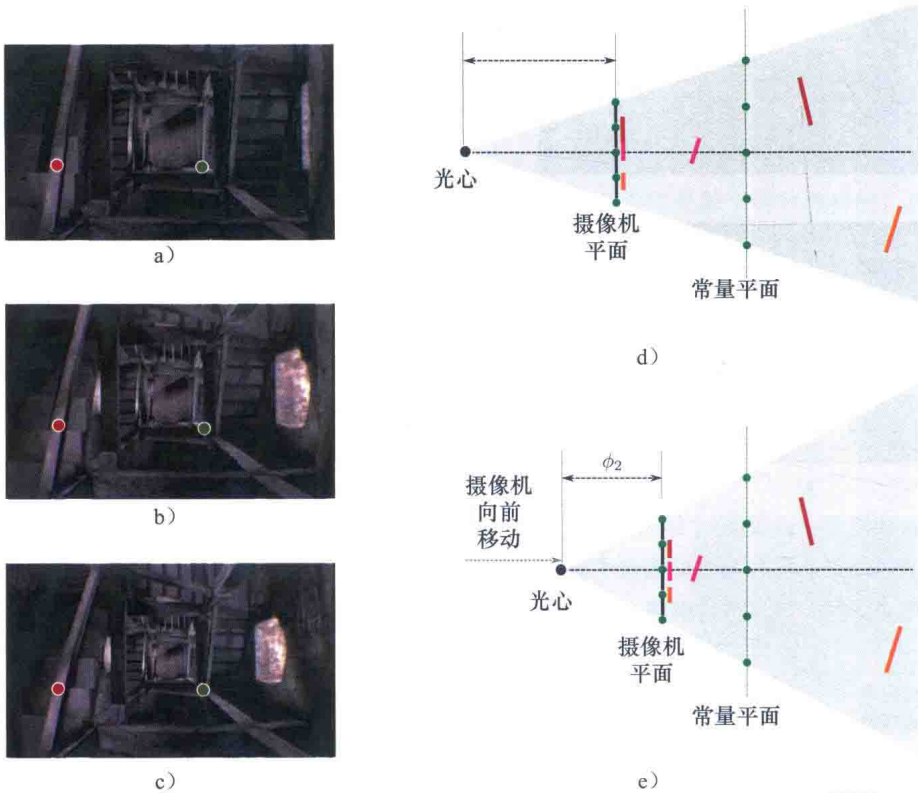


图 14-18 滑动变焦。a ~ c) 《Vertigo》中的 3 帧，其中楼梯井看起来发生了扭曲。附近的物体仍然在原位，而远处的物体系统地通过序列。为了看清这点，考虑每帧中同一个 (x, y) 位置的红色圆圈和绿色圆圈。红色圆圈仍然位于栏杆附近，而绿色圆圈在第一个图像中位于楼梯井的地板上，而在最后一个图像中却在楼梯的中间。d) 为了理解这种效应，考虑一个观察场景的摄像机，这个摄像机由一些同等深度的绿色点和一些其他的平面所构成（彩色线条）。e) 我们沿着 w 轴移动摄像机但同时改变焦距，使得绿色点在同一位置能够成像。在这些变化中，绿色点所在平面中的物体是静止的，但是场景的其他部分发生了移动，甚至可能相互遮挡

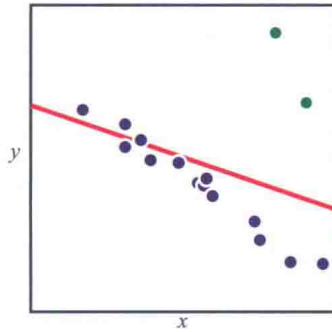


图 15-15 RANSAC 的动机。大多数的数据集（蓝色点）能够利用线性回归模型进行很好的解释，然而有两个异常点（绿色点）。但是，如果我们将线性回归模型与全部数据拟合，平均预测值（红线）将会被拖向异常点，并且无法对大多数数据进行很好地描述。RANSAC 算法通过将异常点标出并将模型拟合剩余数据的方式规避了这个问题

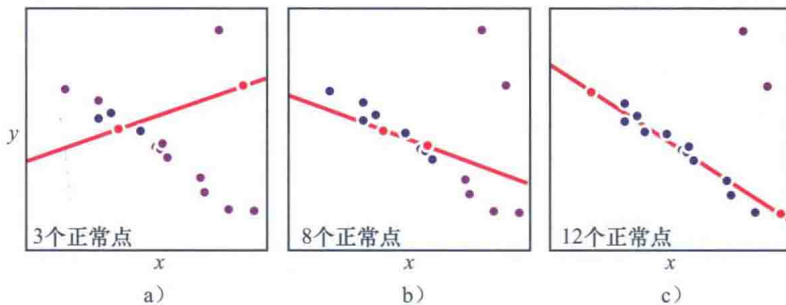


图 15-16 RANSAC 过程。a) 选择一个点的随机最小子集来拟合这条线（红色点）。我们将这条线拟合到这些点，并计数有多少其他的点与这个解（蓝色点）相符合。这些点称为正常点。这里只有 3 个正常点。b, c) 针对点的不同最小子集重复这一过程。多次迭代后，选择拥有最多正常点的拟合。我们只使用这一拟合中的正常点对线进行重拟合

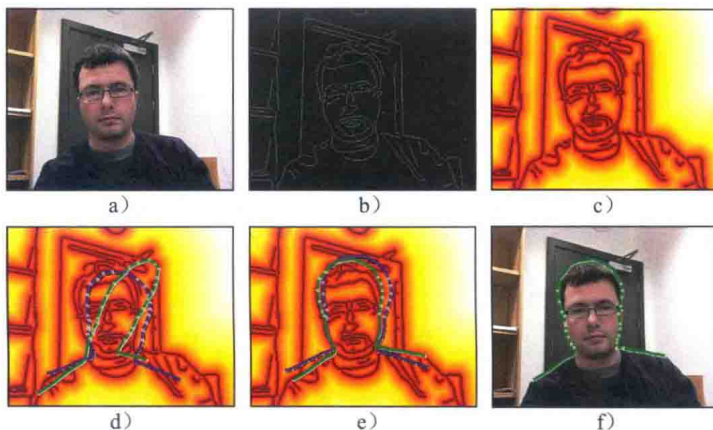


图 17-6 形状模板。已知目标的形状信息，只有图像形状的仿射变换是未知的。a) 原始图像。b) 应用 Canny 边缘检测算子得到的结果。c) 距离变换图像。图像中像素的强度代表该像素与最近边缘的距离。d) 拟合形状模板。利用随机选择的仿射变换（蓝色曲线）来对模板进行初始化。由优化后的标志点所定义的曲线（绿色曲线）已经移向距离图像中具有较低值的位置。在这种情况下，拟合过程收敛于局部最优，并且还不能确定正确的轮廓。e) 如果从接近真实的最佳值开始优化，那么它收敛于全局最大。f) 模板拟合的最终结果

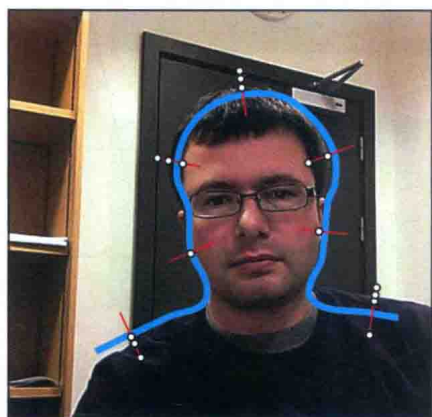


图 17-7 迭代最近点算法。将每个标志点（蓝色轮廓上红色法线处的位置）与图像中的单个边缘点相关联。在本例中，沿着法线的方向（红线）搜索轮廓。通常沿着法线方向会存在一些由边缘检测器确定的点。在每种情况下都选择最近的点——该过程称为数据关联。计算将标志点映射到最近边缘位置的变换，这就移动了边界轮廓，并且在下一次迭代中潜在地改变了最近点的位置信息

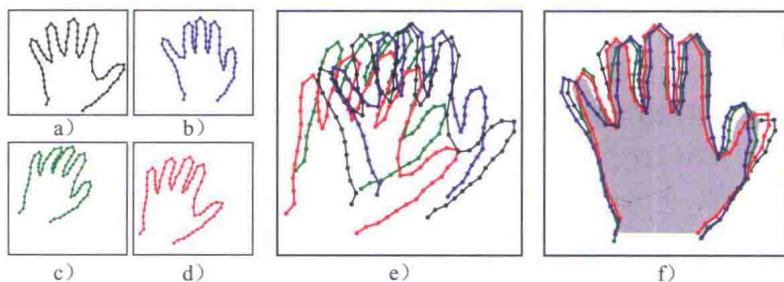


图 17-8 广义 Procrustes 分析。a ~ d) 4 个训练形状。e) 将 4 个训练形状重叠，可见这些形状没有对齐。f) 广义 Procrustes 分析的目标是使用选择的变换来同时对齐所有的训练形状。在本图中，图像通过相似的变换得以对齐（灰色区域表示平均形状）。在此过程后，残留的变化通过统计形状模型来描述

2014年上半年，我在美国宾夕法尼亚州访问时，在图书馆发现了《Computer Vision: Models, Learning, and Inference》这本书，书名使我眼前豁然一亮。书中对模型、算法、学习、推理以及基础知识等方面的介绍深入浅出、通俗易懂，深深地吸引着研究计算机视觉技术的我。作者 Simon J. D. Prince 是伦敦大学学院计算机科学系的知名教授，在生物学、计算机科学等多学科交叉领域颇有建树。一年后，机械工业出版社的策划编辑姚蕾和朱劼联系我，期望我组织翻译本书，我毫不犹豫地欣然接受。

计算机视觉和机器学习相结合，是计算机视觉领域研究的热点，也是难点。本书正是计算机视觉和机器学习相结合的典范之作，清晰明了地阐述一般计算机视觉研究者易于混淆的“模型”与“算法”的区别。本书重点关注基本的技术，强调与学习和推理有关的模型以及相应的方法，其主要特点是知识体系系统、完整，不仅包括计算机视觉的相关理论，而且包括所有的数学背景知识。更重要的是，介绍与机器学习相关的基础知识，并将计算机视觉与机器学习有机地融合起来。考虑到读者最低限度的知识储备情况，本书从概率和模式匹配的基础知识开始，逐步讲解到实际实验，读者通过完成和修改这些实验，就可以建立起可用的视觉系统。

本书包含大量插图，描述了 70 多个算法，同时详细地阐述每个算法的实现细节，细致入微地阐述计算机视觉与机器学习的方方面面，为方便读者理解现代计算机视觉提供了一个完整的概率框架。本书主要适用于计算机视觉和机器学习研究方向的高年级本科生和研究生，书中对各种方法的详细描述也能够帮助计算机视觉的初学者进行学习。

本书由西安电子科技大学的苗启广教授组织翻译，参与的译者有西安电子科技大学的刘凯教授、西安邮电大学的孔韦韦副教授和西北大学的许鹏飞老师，其中苗启广翻译了第 9、10、11 和 19 章，刘凯翻译了第 1~8 章，孔韦韦翻译了第 12~15 章和附录部分，许鹏飞翻译了第 16~18 章和第 20 章以及前言等内容，统稿和审校由苗启广和刘凯共同完成。

本书的翻译工作涉及非常多计算机视觉基础知识的准确翻译，大量术语的统一，多位译者的交流沟通以及字斟句酌的研读等，既从头至尾充满着挑战，又洋溢着团结、合作的气氛，更不时分享逐步成文的喜悦。在此要感谢武警工程大学的雷阳博士，西安电子科技大学的博士研究生刘家辰、刘如意、李宇楠、程飞，硕士研究生孙尔强、范莹莹、黄玉辉、黄志新等，他们在书稿初译、译文修改、统一术语等方面做了大量的工作。感谢机械工业出版社的策划编辑姚蕾和其他编辑，他们为本书的出版付出了大量劳动，认真编辑书稿并提出了修改意见。

谨以本书献给致力于计算机视觉算法研究和开发设计的读者，翻译不足之处请批评指正！

苗启广

译者简介

Computer Vision: Models, Learning, and Inference



苗启广 西安电子科技大学计算机学院教授、博士生导师，2012年入选“教育部新世纪优秀人才支持计划”。中国计算机学会 CCF 理事，陕西省计算机学会理事，陕西省大数据与云计算产业联盟理事，CCF 计算机视觉专委会委员，CCF 人工智能与模式识别专委会委员，CCF 青年工作委员会委员，CCF YOCSEF 主席(2017~2018)，教育部工程专业认证协会计算机分委会工程专业认证专家。2005年12月获西安电子科技大学计算机应用技术博士学位，2014年在美国做高访，主要从事计算机视觉与机器学习、大数据分析以及高性能计算方面的研究。主持在研和完成国家自然科学基金、国防预先研究项目、国防 863 和武器装备基金项目 20 余项。2008/2011/2014 年分别获西安电子科技大学“十佳师德标兵”称号。近年来，在 IEEE TNNLS、TEC、TIP、TGRS、AAAI、软件学报、计算机学报等国内外重要学术期刊及国际会议上发表 SCI/EI 收录论文 70 余篇。担任 2015 年 CCF 首届中国计算机视觉大会程序委员会主席、2011 年 CCF 首届青年精英大会组委会主席。Journal of Industrial Mathematics、中国计算机学会通讯(CCCF)、物联网技术等国内外期刊编委，教育部国家科学技术奖评审专家，国防基础科研评审专家。先后获省部级奖 2 项。



刘凯 西安电子科技大学计算机学院教授，博士生导师，陕西省图像图形学会理事，中国仪器仪表学会空间仪器分会理事。主要研究领域包括图像视频压缩编码、图像识别以及视频跟踪。主持和参加了国家自然科学基金、高分辨率对地观测重大专项、探月工程以及多项横向合作项目。发表 30 多篇学术论文，获得 10 余项发明专利。



孔韦韦 博士后，硕士生导师，西安邮电大学副教授。现为 IEEE 会员，IEICE 会员，韩国 AISS 协会编委，中国计算机学会会员，陕西省计算机学会人工智能与模式识别专业委员会委员，主要研究领域为图像智能信息融合、入侵检测等。主持和参与了国家自然科学基金、国家级信息保障技术重点实验室开放基金课题，以及全国博士后特别资助项目、全国博士后基金面上项目一等资助、全军学位与研究生教育研讨会专项研究、陕西省自然科学基金项目等 10 余项课题。以第一作者在 SCI 源期刊(如中国科学(F 辑 信息科学)(英文版)、IET Image Processing、IET Signal Processing、Optical

Engineering、IET Electronics Letters、Infrared Physics & Technology 等杂志)上发表论文近 20 篇,获 2012 年度 IET 学术协会优秀学术论文成果奖,以第一申请人申请发明专利 2 项,出版专著 1 部(第一完成人)、参与编写著作 1 部(第三完成人),并担任多个 SCI 源期刊的特约审稿人。



许鹏飞 西北大学信息科学与技术学院讲师,2014 年获西安电子科技大学计算机应用技术博士学位。主要研究方向是模式识别、数字图像处理。目前已在 IEEE T Image Processing、Neurocomputing、JVCI、IET Image Processing、Integrative Zoology、Optics Communication、MTAP、电子学报、CIS 等国际与国内权威期刊和会议上发表 10 余篇学术论文。获 2014 年西安市科学技术奖三等奖。主持和参与了国家自然科学基金、国防预先研究项目、西北大学科学研究基金资助项目等。获得 10 余项国家发明专利。

我对本书从开始到后期发展的情况已经有了一个大概的了解，非常荣幸被邀请为本书写序。我是刚好参加了 BMVC 2011 之后开始写这篇序的，在该会议上，我发现已经有其他学者开始阅读本书的草稿了，并且还听到一些诸如 “What amazing figures!” “It’s so comprehensive!” 以及 “He’s so Bayesian!” 等很好的评价。

如果仅仅因为本书含有精美的插图、对每一种新的视觉算法提出新的见解或者因为它很 “贝叶斯”，那么我不建议你阅读这本书。我建议你阅读它的主要原因是：它清晰地阐明了在计算机视觉研究中最重要区别，即 “模型” 和 “算法” 之间的区别。这种区别就类似于 Marr 用三层计算理论进行阐释，而 Prince 的两层差异用概率论的语言完美清晰地阐明。

那为什么这种区别如此重要呢？让我们看看视觉领域中一个古老而又简单的问题：将一幅图像分离为 “前景” 和 “背景”。经常会听到刚刚接触视觉研究的学生像早期的视觉研究者一样通过叙述算法来处理这一问题。首先使用主成分分析 (PCA) 方法找到主要的颜色轴，然后生成一幅灰度图像，接着进行阈值化处理，最后使用形态学操作符清理孔洞。然而，当他们运用这些方法在一些测试图像上进行实验时就会发现，真正的图像是更复杂的，所以需要补充新的处理步骤：需要使用某种形式的自适应阈值处理方法，并且可以通过模糊边缘图像和求图像局部极大值来获取这些阈值。

然而，多数读者都已经知道这些方法是非常脆弱的。因为使用这些多变的 “幻数” 控制方法中相互影响的步骤，不可能获得一个能够适用于所用图像 (甚至是一个可用的子集) 的参数集合。这一问题的根源在于算法根本就没有明确其具体的目标是什么。“前景” 和 “背景” 的分离到底意味着什么？我们能够从数学模型上详细阐述这一问题吗？

当计算机视觉的研究人员开始处理这些问题时，统计语言和马尔可夫随机场能够清晰地表现出目标和算法之间的差异。我们所撰写的内容不是解决问题的具体步骤，而是问题本身，例如：求最小值的函数。本书给出了定义问题的所有概率分布的公式，并提供了依据这些分布规律来获得最终答案的相应操作。本书揭示了这种方法是怎样处理多种视觉问题的，以及如何更简单地推导出鲁棒性更强的解决方案。

但是，这并不意味着我们提出模型，再让他人去求解它的各项参数，因为可能模型的空间要远多于其得到的解空间。因此，人们潜意识会记住那些已知能够求解的模型集合，并且总是为那些具有可行性解的问题寻找其相应的模型。在这一阶段，人们就会在策略上进行深入思考，例如：我能够扩展 alpha expansion 算法来求解离散的参数，然后使用高斯-牛顿法 (Gauss-Newton) 求解连续的参数，虽然这些方法可能执行效率低，但是它表明了我们努力地提出一种更高效的联合算法是值得的。这一策略是非常常见并且有用的，它能够使得人们在潜意识中一直保持创建模型的思想。

然而，即使树立本书倡导的科学态度，如今经验丰富的研究人员也陷入难以分清模型与算法之间区别的困境。他们发现自己是这样思考问题的：我将针对具体的色彩分布选择合适的高斯混合模型，然后对这些混合权值进行建模以构建 MRF 模型，并且利用图像分割的理论对模型进行更新，再返回到第一步重新执行。好的方面是，此方法能够转换为模

型。即使拟合模型的唯一已知方法是刚刚提到的方法，将它作为一个模型的学科也允许你对它进行推理，以充分利用可供选择的技术，最后进行更好的研究。阅读这本书是提高自身能力的一个可靠方式。

那么能够让我们成为更优秀的研究人员的概率语言是什么？让我们以贝叶斯定理中的工程师观点为例，我们通常听到贝叶斯和频率论之间的差异，但是我认为许多工程师都认为贝叶斯理论有很多基本问题：贝叶斯肯定在说谎。他们接近以前的平均值的估计和传感器的读数有很大的不同。例如一个能够测量身高的机器，其中的传感器有 1cm 误差的均匀分布，每当能够在 1cm 误差范围内准确预测出某人的身高就能够获得 1 英镑。根据贝叶斯定理，如果传感器读取的数据是 $200\text{cm} \pm 1\text{cm}$ ，那么应该报告 199cm；这种方式能够比猜测实际传感器的读数赚取更多的钱，因为比起那些身高 200cm 的人，身高 199cm 的人更多。因此，作为一名工程师，我认为贝叶斯理论是一种能够获得最优解的方法，并且从务实(但比我自己的要更加精细)的角度看，它在本书中是非常受欢迎的。我甚至怀疑本书是一本带有视觉样例的统计学书籍，而非一本基于概率论的视觉书籍。

如果在我写完这篇序之后还没有提及书中的插图，那将是我的错误。书中的这些插图非常好，但这并不是因为它们很漂亮(虽然它们通常很美)，而是因为这些插图甚至为那些最基本算法的工作原理和思路提供了重要的见解。第 2~4 章的插图是理解现代贝叶斯推理的基础，然而，我怀疑只有少数研究人员曾看过所有这些插图。后面的插图则能够非常清晰地表达极其复杂的思想，同时也非常清楚地表达了基础算法的实现过程，这些都真实地展示了底层模型是如何影响算法性能的。

最后，我认为这本书值得与我的同事 Richard Szeliski 最近写的一本教科书进行直接比较。那本书也是一本非常全面的计算机视觉图书，书中包含优美的插图、深刻的见解，并对大量现有的计算机视觉算法进行了有益综合。但从真正意义上来说，这两本书是分别站在教学方法的两个对立面的：Szeliski 对计算机视觉研究领域最新的技术进行了全面总结，而这本书揭示了如何在计算机视觉研究领域取得进展。在今后的几十年里，我将会一直收藏这两本书，或者一直把它们放在我的书桌上，便于自己经常翻阅。

——Andrew Fitzgibbon
Microsoft Research, Cambridge

目前，已有很多关于计算机视觉的书籍，那么还有必要再写另外一本吗？下面解释撰写本书的原因。

计算机视觉是一门工程学科，机器在现实世界中捕获的视觉信息可以激发我们的积极性。因此，我们通过使用计算机视觉解决现实问题来对我们的知识进行分类。例如，大多数视觉教科书都包含目标识别和立体视觉内容。我们的学术研讨会也是用同样的模式进行组织的。本书对这一传统方式提出了质疑：这真的是我们组织自己知识的正确方法吗？

对于目标识别问题，目前已提出多种算法解决这一问题（例如子空间模型、boosting模型、语义包模型、星座模型等）然而，这些方法没有什么共同点。任何试图全面描述知识的壮举都会转变为一个非结构化的技术列表。我们怎样让新同学把所有的技术和理论都弄懂呢？我主张使用一种不同的方式来组织知识，但首先让我告诉大家我是如何看待计算机视觉问题的。

对于一幅图像，我们不仅要观察图像中的内容，同时还需要提取其测量值。例如，我们可以直接使用RGB值，或者对图像进行滤波处理，或者执行一些更复杂的预处理。计算机视觉的目标或者需要解决的问题是使用这些测量值来推理全局状态。例如：在立体视觉中，我们尝试推断出场景的深度。在目标识别中，我们尝试推断某一特定类目标存在与否。

为了实现目标，我们建立一个模型。模型描述了测量值与全局状态之间的一系列统计关系。这一系列统计关系中的特殊成员是由一个参数集合确定的。在学习的过程中，选择这些参数，以便它们能够准确反映测量值与全局状态之间的关系。在推理的过程中，选用一组新的测量值，并利用学习后的模型来推理全局状态。学习和推理的方法包含在算法中。我认为计算机视觉应该从以下几方面来理解：目标、测量值、全局状态、模型、参数、学习和推理算法。

我们可以根据这些量选择性地组织知识，但在我看来，模型中最重要的内容是全局状态和测量值之间的统计关系。这主要有三个原因。首先，模型的类型往往超越了应用（同一个模型可用于不同的视觉任务）；其次，模型能够自然地把它自身组织成一些可分开理解的系列（例如，回归、马尔可夫随机场、相机模型）；最后，在模型层次上讨论视觉问题使得我们能够得到那些貌似不相关的算法和应用之间的关联。因此，本书的章节安排非常巧妙，每个主要的章节都讨论一系列不同的模型。

最后一点，本书中的大部分思想在第一次接触到时是难以理解的。因此，我的目标是使后续研究计算机视觉的学生更容易理解这些内容，我希望这本书能够达到这一目的，并能够激励读者深入了解计算机视觉。[⊖]

⊖ 原出版社网站和作者网站(www.computervisionmodels.com)上包含丰富的教辅和参考资源，包括教学幻灯片、插图文件、书中包含算法和模型的详细介绍、源代码、进一步阅读资源、相关课程的教学大纲以及部分习题答案。书中全部习题答案，只有使用本书作为教材的教师才可以申请，需要的教师可向剑桥大学出版社北京代表处申请，电子邮件：solutions@cambridge.org。——编辑注