

麻省理工学院 (MIT) 机器视觉课程指定教材



Robot Vision

机器视觉

【美】伯特霍尔德·霍恩 著
Berthold Klaus Paul Horn

王亮 蒋欣兰 译

中国青年出版社

麻省理工学院 (MIT) 机器视觉课程指定教材

机器视觉

Robot Vision

【美】伯特霍尔德·霍恩 著
Berthold Klaus Paul Horn

王亮 蒋欣兰 译

中国青年出版社

(京)新登字083号

图书在版编目(CIP)数据

机器视觉/[美]霍恩(Horn, B.K.P.)著;王亮,
蒋欣兰译.—北京:中国青年出版社,2014.7

书名原文:Robot vision

ISBN 978-7-5153-2531-6

I. ①机… II. ①霍… ②王… ③蒋… 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(2014)第146784号



©1986 by The Massachusetts Institute of Technology
Simplified Chinese Edition Copyright:2014 CHINA YOUTH PRESS

北京市版权局著作权合同登记

图字:01-2014-3986

责任编辑:彭岩

*

中国青年出版社出版 发行

社址:北京东四12条21号 邮政编码:100708

网址:www.cyp.com.cn

编辑部电话:(010)57350407 门市部电话:(010)57350370

北京科信印刷有限公司印刷 新华书店经销

*

700×1000 1/16 37印张 4插页 660千字

2014年8月北京第1版 2014年8月北京第1次印刷

定价:98.00元

本书如有印装质量问题,请凭购书发票与质检部联系调换

联系电话:(010)57350337

One of the greatest pleasures in life is figuring out how to solve a problem. Machine vision presents us with many puzzles, the solution of which typically involves understanding the physics, followed by mathematical modelling, and finally algorithmic implementation.

While we have many "proofs of concept" (animals) of vision systems, there is no simple theory (yet) that tells us how to build a mechanism to make use of an image as the basis of interaction with the environment.

I hope you, the reader, will have many pleasant experiences solving machine vision problems.

A handwritten signature in black ink, appearing to read "J. P. Jones". The signature is stylized with a large, sweeping initial "J" and "P".

中文版序言

对于《*Robot Vision*》中译本的出版发行，我感到非常兴奋。因为这将使得广大的学生、教师和科研人员可以更容易地接触和了解机器视觉。自从《*Robot Vision*》首次出版以来，将近三十年过去了，机器视觉领域一直在不断地发展和变化着，但是直到现在，始终存在着一种需要，即：寻找一种关于这个领域的统一的研究方法，并且，该方法应该是能够经受住时间考验的。我相信在研究机器视觉的过程中，我们应该思索如下两个问题：1) 成像过程的基本原理是什么？2) 如何探索对成像过程“求逆”的知识和方法（具体地说，就是从一张图像、多张图像或者图像序列中恢复出关于场景或者观测者的信息）？显然，要做到这一点，我们至少需要理解：光学、运动学和坐标变换。通过物理模型，我们可以很自然地导出相应的数学公式，然后，通过数学公式，我们可以导出相应的算法。本书的主题包括：

- 建立起对成像过程的基本理解；
- 探讨对成像过程进行求逆的方法。

尽管本书包含了许多用于解决机器视觉问题的具体方法，但是，本书所关注的是一种针对不同的机器视觉问题的整体解决方法。事实上，许多成功的机器视觉技术，例如：从明暗恢复形状、光流、无源导航等，都是这种“整体解决方法”的具体应用。关注基本原理所带来的一个好处是：真正需要去理解的核心内容变少了，甚至连算法的实现细节都有可能变成一个“瞬间”就被理解或完成的工作。

最近几年，在机器视觉的应用领域取得了很多实质性的进展，这些进展部分归功于机器学习方法的使用。当然，机器学习结果的好坏依赖于：从图像中所提取出来的特征的“质量”。特征的“描述力”越强，区分效果也就越好；所提取出的特征越有“特点”，所得到的结果也就越好。从这个意义上说，本书所讲述的内容可以被看作是：用于提取“具有物理意义的特征”的方法。在写《*Robot Vision*》的序言时，我以“**机器视觉是一个年轻的、发展很快的领域**”作为开篇的第一句话。这么多年过去了，我们很高兴地看到：直到今天，这句话还像当初我写它时一样正确！

Berthold K. P. Horn

2014年6月于MIT

译者序

这是一本很老的书，但是，这是一本很经典的书！这是一本出版于三十年前的书，但是，这是一本在美国麻省理工学院（MIT）的“机器视觉”课上讲了三十多年^①，并且，还将继续被使用下去的书！这是一本亚马逊上所有购书者都给予五星评价的书！这是一本被读者誉为“机器视觉的圣经”的书！这是一本引用量超过5000的书！为什么一本书会有如此强大的生命力？或许作者已经在序言中给出了答案：“相对于讨论那些针对某些特定问题的特殊方法，我更加关注于：发展出一种关于机器视觉的可靠研究方法。”无论机器视觉领域的发展更新如何迅速，一些基本的思想和方法将永远是非常重要的！我相信对于刚进入这个领域或者有志于这个领域研究的朋友，这些思想和方法应该是值得去了解、学习和吸收运用的。

作为机器视觉领域的先驱和奠基者之一，作者在该领域的许多方面都有开创性和奠基性的贡献，例如：**从明暗中恢复形状**（shape from shading）、**光流**（optical flow）、**无源导航**（passive navigation）、**扩展 Gauss 图**（Extended Gaussian Image）等，而本书也可以看作是：作者对这些原创性工作的一个总结。不同于简单的内容总结，作者用一条主线将书中所有的内容穿插在了一起，使得本书结构紧凑，条分缕析，从而成为一本优秀的教科书。这条主线也是机器视觉的核心问题：**如何从一张或多张图像中恢复出场景中所包含的信息？**沿着这条主线，作者将机器视觉的核心思想和关键技术清晰明确地讲了出来，使得本书成为一本思想性的著作。这是本书最大的特点，也是本书成为经典的原因所在！

2011年9月至2013年3月，译者在MIT进行了18个月的访问学习。在这期间，译者有幸听了Horn教授的“机器视觉”课程^②，并且，认识了Horn教授。本书以及该课程对我的冲击和震撼是难以想象的！在课堂上以及课下看这本书的每一分钟，我都有一种“相见恨晚”的感觉。这门课我学了两遍，至今仍使我深感遗憾的是：由于时间的原因，没有机会再去听第三遍！我感觉我是非常幸运的，有机会在这么短的时间内，通过本书和该课程，对该领域的很多东西有了清晰和深刻的认识。我想：对于许多想了解机器视觉以及有志于这个领域研究的朋友，这本书将会是有用的。这是我翻译此书的初衷，也是

^①本书的英文原版《Robot Vision》出版于1986年，在本书出版以前，本书的手稿作为MIT“机器视觉”课程的讲义，已经使用了5年。

^②机器视觉的课程编号为6.801，它是MIT的秋季课程。

激励我度过翻译过程中无数个漫漫长夜的真正动力所在。

我们有理由相信：既然“圣经”没有过时，这本书也同样不会！由于译者水平有限，很难将本书由“圣经”译成“圣经”，或者至少译成一本“名著”，但是，我们始终坚持：要将本书译成一本“读得懂”的书！因此，我们力求内容的清晰明确和易于理解。在翻译的过程中，时时让我感到苦恼的是：如何在内容的准确性和流畅性之间找到一种合理的平衡。因为对于学术著作的翻译，大致包括两个层面：

- **语言的翻译**：这句话说的是什么？
- **思想的翻译**：为什么要说这句话？

只有处理好这两者之间的关系，才能避免出现“句句翻译都正确，内容完全看不懂”的情况。这样的例子不胜枚举，例如，对于第6章6.14小节的第一句话，直接翻译应该是：“为了设计**最优滤波器**，我们必须估计出：要处理的图像的**功率谱**。”但是，“要处理的图像的**功率谱**”是可以直接计算出来的，为什么还要对其进行“估计”？事实上，这句话中的“要处理 (to be processed) 的图像”指的是：要经过处理而得到的图像，也就是说，要被复原出来的图像。再比如，对于第17章的题目：Passive Navigation & Structure from Motion，不少书籍将“Structure from Motion”直译为“从运动到结构”。这样翻译存在两个问题，首先，“到”字是译者自己加上去的，“到”字的使用极不合适。“Motion”指的是：图像亮度模式在像平面上的“运动”，具体地说，就是指运动场（或光流），而“Structure”指的是：产生运动场（或光流）的原因，包括：1) 相机和场景之间的相对运动，2) 景深。“Structure”和“Motion”之间并不是并列关系，而是因果关系，因此，我们不应该使用“到”字，而是使用“恢复”。其次，我们平时说到“运动”时所表达的意思，并不是这里的“Motion”，而是“Structure”。因此，我们将第17章的题目“简单地”译为：无源导航，以避免意思上的混淆。

此外，当语言的“优美”和思想的清晰发生矛盾时，我们毫不犹豫地选择后者！例如，在第11章中，我们探讨了“characteristic stripe”方法^①。将“characteristic stripe”译为“**特征条纹**”似乎显得很“文雅”，但是，至少对我而言，很难直观地理解：“特征条纹”这几个字究竟说的是什么。因此，我们将其译为：**特征“带**

^①给曲线上的每一个点赋予一个方向（即：一个单位向量），所有这些单位向量合在一起，就形成了空间中的一条“带子”。想象你扛着一根竹竿走路，竹竿所“扫过”的区域就是这里的stripe。

子”。尽管这种译法有些“俗气”，但是，它能够更加形象地反映出：“characteristic stripe”这个短语背后的真实想法。语言是思想的载体，而语言的使用也应该以表达思想为最终目的。因此，在翻译的过程中，我们所追求的是：形象直观、易于理解的表述方式。我们始终坚持：一定要将原著中的思想翻译出来！

当然，语言的准确和思想的清晰并不总是矛盾的，这使得我们有希望去努力让本书的中译本：

- 在语言方面 —— 让读者感到：这是一本“用中文写的书”；
- 在思想方面 —— 让读者看到：算法背后的基本思想和观念。

为此，一方面，对于书中的一些内容，我们在内容正确和忠实于作者意图的前提下，尝试着使用更加“容易被理解”的中文表述方式。例如，在第6章中，作者将：图像失焦所对应的模糊核，形象地称为“小药瓶子”。在欧美，绝大多数“药瓶子”都是圆柱体。因此，这种说法很形象贴切，但是，当我们谈到“药瓶子”时，却很难想到圆柱体，因为我们所见过的绝大多数“药瓶子”，盖子都比瓶身要小一圈。因此，我们将其意译为“小圆柱体”。此外，对于书中的内容（尤其是涉及数学分析和思想概念的内容），我们尽可能地使用短句结构。例如，对于下面的长句（即：直接翻译的结果）：“消失点是经过投影中心并且和这组平行线平行的直线与像平面的交点。”我们将其翻译为相应的短句：“消失点是：一条直线与像平面的交点；并且，这条直线必须满足如下两个条件：1) 经过投影中心，2) 和这组平行线平行。”这样做有时会失去语言上的简洁优美，但是，却更容易展现出长句中各个成分之间的逻辑关系，以便于读者理解。另一方面，对于书中的一些内容，我们加入了注释。这些注解主要针对：数学推导、专用名词以及某些结论或方法背后的思想。为了保证这些微小的调整与作者的想法和意图相符，我们一直和Horn教授进行着书信交流，对书中的内容进行了反复讨论。书中绝大多数注解的内容都是根据这些邮件的内容整理出来的。

书中每章都附有相应的参考文献。作者并不是简单地罗列出参考文献，而是通过这些参考文献介绍了该领域以及相关领域的发展过程。一个使我感到非常惊讶的事实是：尽管几十年过去了，这些参考文献中的大部分至今仍有人引用！因此，我们保留了这些参考文献的英文名称，以便于读者查找相关文章和书籍。此外，对于书中所涉及的人名，我们并没有将其翻译成中文，例如，我们将Gaussian filter、Euler equation、Fourier transform译成：Gauss滤波

器、Euler 方程、Fourier 变换。这样做有两个原因：1) 对于某些英文名字的翻译，现有的出版物并没有完全取得一致。例如：绝大多数的书籍将“Fourier”翻译成“傅里叶”，但是，仍然有少量的书使用“傅立叶”或“富里叶”。2) 便于读者查找相关资料。我们根据原著中的“索引表”，用“粗体”标注出了所有的重要术语，然后，在本书的最后生成了新的“索引表”。此外，我们对书中所有的公式都按章节进行了编号。这样做的一个好处是：可以将原著中大量的“上面的公式”换成“式(编号)”，以便于读者查找。只有当所指代的公式正好出现在这句话上面的时候，为了表述上的自然，我们才偶尔沿用“上面的公式”这种说法。

本书根据原著的第七次印刷版[1991年]翻译而成。在翻译的过程中，我们根据原著的勘误表，对书中的一些笔误和印刷错误进行了更正。在原著第17章中，作者使用的透视投影公式为： $x = X/Z$ 和 $y = Y/Z$ 。这个公式的基本假设是：选取相机的焦距 f 作为长度单位。但是，使用标准长度单位却更容易被理解，而且，作者也更加倾向于使用标准长度单位。因此，我们对第17章中的公式进行了大修改。透视投影公式相应地变为： $x = fX/Z$ 和 $y = fY/Z$ 。基于这个公式，我们推导出了第17章中的其他公式。虽然这样做使得有些公式看起来很“臃肿”，但是，公式中每一项的物理意义变得更加清晰了，而且，量纲也“变得”平衡了。此外，在第7章7.5节中，我们加入了：1) 循环卷积的定义，2) 循环卷积与离散 Fourier 变换之间的关系。为了便于读者的理解，我们在和 Horn 教授商讨以后，决定在第7章中加入插图 7.1；在第15章中加入插图 15.4；在第17章中加入插图 17.1 和 17.2。

在翻译本书的过程中，许多朋友给了我们很多关心和帮助，以及很多有益的建议，他们是：张明敬、陈小美、冯钊赞、龚恒风、李文钊、赵一彪、王哲、林杰、张俊三、王胜春、董克强、张宏、解宏伟、武栋、何凯、高大龙、马思源、田明杨、段世杰，Shevarl Macnamara, Arindam Biswas；北京交通大学谢桦副教授、吴发恩教授、汪成咏教授、王兵团教授和冯国臣老师；上海交通大学刘当波教授；北京联合大学李哲英教授；MIT 数学系的 Steve Jhonson 教授、David Jerrison 教授和 David Vogen 教授；MIT 计算机系的 Alan Oppenheim 教授和 Dannis Freeman 教授。我们重新绘制了书中绝大部分的插图，在这个过程中，于浩洲同学给了我们很大的帮助。在 Latex 排版过程中，吕国豪博士和吕同富教授给了我们很多帮助，在本书的最后编排过程中，彭岩编辑给了我们很多指导。王凤娇、吴洁、刘可豹、于欣妍、陈宇和王鹏飞帮助我校正了书稿中的语言文字，并且，提出了很多宝

贵的意见。本书的出版和他们的辛勤劳动是分不开的。此外，我们还要特别感谢的是：我们的恩师**罗四维**教授，他带领我们进入了这个迷人的领域；MIT 数学系的 **Gilbert Strang** 教授，他教会我如何去感受数学的美丽；MIT 计算机系的 **Berthold K. P. Horn** 教授，他给了我许多指导和帮助，不单单是知识，还有做研究的态度和方法。和他们在一起学习和工作的日子，已经成为我美好记忆中非常宝贵的一部分。最后，我要感谢我永远都报答不尽的双亲，他们的理解和支持给了我很大的勇气。

本书第 1 至 4 章由蒋欣兰翻译；其余部分由王亮翻译和整理。最后，由王亮负责全书的整理和统校。由于译者水平有限，错误和不恰当之处在所难免。如果能得到读者的帮助，使得本书不断得到改进和提高，我们将不胜感激^①。在本书的扉页中，Horn 教授亲笔题写了“致广大中国读者的寄语”，其内容为：

One of the greatest pleasures in life is figuring out how to solve a problem. Machine vision presents us with many puzzles, the solution of which typically involves understanding the physics, followed by mathematical modeling, and finally algorithmic implementation.

While we have many working “proofs of concept” (animals) of vision systems, there is no simple theory (yet) that tells us how to build a mechanism to make use of image input as the basis of interaction with the environment.

I hope that you, the reader, will have many pleasant experiences solving machine vision problems.

希望在一开始的时候，您就从作者的手迹中“辨认”出了这几句话。最后，我们谨以 Horn 教授的这段寄语与广大读者共勉。

王亮、蒋欣兰

2014 年 3 月

^①我们的邮箱是：wangliang.bjtu@gmail.com（王亮）和 jxlzgl@sina.com（蒋欣兰）。
邮寄地址为：中国青年政治学院计算机中心 505，北京市西三环北路 25 号。邮编：100089

英文版序言

机器视觉是一个年轻的、发展很快的领域。由于新的结果不断涌现，在写这些内容时，我既感到兴奋，同时又感觉难以掌握什么时候该停下来。本书根据我在 MIT 讲授“计算机视觉”（课程编号 6.801）课程时所使用的讲义整理而成。我在 MIT 讲授“计算机视觉”课程已经 10 年了。本书的初稿被使用了 5 年^①。书中的练习题大部分来自于“计算机视觉”课程中的家庭作业以及测验。在 MIT，“计算机视觉”被定为一门“限选课”，这意味着学生可以选修这门课，也可以选修其他与人工智能相关的课程。大部分选这门课的学生都是大学新生。本书的一些章节也被用于：针对工业界以及其他学校学生的夏季课程，这种课程一般集中在一个星期左右。

10 年前，我们还有可能在一门课里同时介绍：机器人操作和机器视觉。但是，由于这两个领域的发展太快，现在，我们已经无法再将这两方面的内容合并到同一门课里（原来课程中关于机器人操作的那部分内容，已经被 Tomas Lozano-Pérez 扩充为一门新的课程“机器人操作”，课程编号 6.082）。事实上，即使是只讨论机器视觉方面的一些有趣内容，单单一个学期也似乎显得太短。

机器视觉领域的快速发展使得本书不得不精简掉那些“不是那么重要的内容”。什么是“不是那么重要的内容”？在某种程度上，这是一个因人而异的问题，而这本书也反映出我的个人倾向。本书不是一本介绍机器视觉方面所有结果的书；否则，这本书将变得非常厚，并且，难以找到一个相关的主题将这些内容“串”起来。因此，一些看似缺乏坚实理论基础的材料并没有被纳入本书。

同样的，对于一些不能被用于有效地从图像中恢复出信息的方法，也被本书所“忽略”，即使这些方法使用“先进”的数学来表现它们的合理性。本书所力求包含的内容，是那些有助于工程技术人员在“现实世界”中应用机器视觉方法的有用信息。例如，在关于二值图像处理的章节中，我们提出了如何改进许多现有的商用电子器件；在关于光度立体视觉和扩展 Gauss 图的内容中，我们所提出的方法，可能成为这个领域研究结果的商业化过程中的下一个推动力。

对于一些具体实现方法的选择和特定算法，本书并没有总是给出全部细节。依赖于所使用的计算机系统的特殊性能，所取得的好的结果，以及对于这些结果的陈述，将会偏离本书的中心。我也相信：在

^①本书出版于 1986 年，至今仍然作为 MIT“计算机视觉”课程的指定教材。作者至今一直在 MIT 讲授“计算机视觉”课程。——译者注

开始考虑实现方法之前，我们应该先解决机器视觉的基本问题！在绝大多数情况下，实现过程只是经典数值分析技巧的直接应用而已。即使这样，本书仍然包含了许多软件和硬件方面的有效实现方法的细节，例如：在关于二值图像处理的章节中，我们对具体实现方法进行了详细的讨论。

几乎从一开始，本课程就吸引了许多有志于探索这个领域的研究生。因此，本书也纳入了一些有助于拓展他们数学修养的材料。对于本科生，这些材料需要被省去，因为，在一学期的课程中，同时讲解这些数学基础以及本书中的所有材料会显得很困难。本书中的一些主题是独立于其他内容的，因此，省去书中的部分内容不会很困难。同时注意：在附录中，我们给出了关于本书所用到的绝大部分数学工具的介绍和推导。

除了一些章节是明显的两两相关以外（即：第3章和第4章、第6章和第7章、第10章和第11章、第12章和第17章、第16章和第18章），事实上，章节之间是相互独立的。对于缺乏线性系统理论知识的学生，省略掉图像处理方面的内容（第6、7、9章），会取得更好的教学效果。同样的，忽略掉关于时变图像处理的章节（第12、17章），也不会影响整本书的内容连贯性。一些章节所介绍的内容和剩下章节所介绍的相比，还没有在工业界中得到很好的应用。如果把这些章节的内容也省去的话，那么，剩下的一个由第1、2、3、4、10、11、16和18章所组成的短的基础课程，应该能够在一个学期内讲完。

这本书试图提供一个对那些我自认为理解深刻的课题的深度覆盖。这意味着：对于其他一些机器视觉方面的课题，本书并没有给出十分详尽的介绍，而另外一些我认为“过于专业”的课题也没有被纳入本书。从这方面讲，这本书可以看作是 Dana Ballard 和 Christopher Brow 所写的《*Computer Vision*》的补充。Dana Ballard 和 Christopher Brow 的《*Computer Vision*》包含了更为丰富的材料，但是内容没有本书深入。同样，在 Azriel Rosenfeld 和 Avinash Kak 的《*Digital Picture Processing*》第二版中，读者可以找到：关于本书中许多基础概念的更加详细的介绍。

本书所讨论的内容和生物视觉系统的研究有很强的联系。但我在写本书时并没有着重突出这一点，因为，这方面已经有非常杰出的著作：Davi Marr 的《*Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*》。类似地，关于边缘检测、基于特征的立体图像、以及对于时变图像某些方面的解释等

方面的内容，本书的介绍也略显不够详尽细致，因为以下书籍：Eric Grimson 的《*From Image to Surface: A Computational Study of the Human Early Visual System*》；Ellen Hildreth 的《*The Measurement of Visual Motion*》以及 Shimon Ullman 的《*The Interpretation of Visual Motion*》包含了这些方面更为详尽的细节介绍。同样的，对于模式分类，已经有非常经典的著作，例如：Richard Duda 和 Peter Hart 的《*Pattern Classification and Scene Analysis*》。但是，在整本书的写作过程中，我并没有完全坚持这种简约紧凑的写作风格，而是使用了充足的章节介绍关于图像处理的内容，尽管这些内容已经很好地包含在 William Pratt 的巨著《*Image Processing*》之中。这样做的理由是：对于理解在随后章节中所用到的处理方法，这些图像处理方面的内容实在是太重要了！

我的许多学生在初上这门课时，对以下事实感到震惊：机器视觉这个领域所需要用到的数学技巧绝非简易！在一定程度上，这是因为：他们以前看见的很多例子，都只需要使用简单的启发式的方法^①，就产生出一些表面上惊人的结果。许多这样的方法，都是在机器视觉的早期研究过程中发展起来的。它们最终导致了：对这个领域发展速度的一个虚假乐观主义的预测。后来，这些方法的巨大局限性显现了出来。显然，现在要学习机器视觉，离不开对成像理论和过程的理解。这就像学习自然语言需要一些语言学的知识一样。不久之前，这个观点还只是被少数人所接受。

更严重的是，很多人认为机器视觉只是“人工智能”系统中的一种分析其传感器所感知到的信息的工具。到目前为止，只有相对较少的复杂的数学处理方法被用于人工智能领域。将机器视觉和机器操作控制看作只是人工智能系统的“输入/输出”（“I/O”）的观点是错误的。机器在“观看”、“操作”以及“行动”过程中所遇到的问题，本身就是非常有趣的！这些问题非常难，因此，解决这些问题所用到的工具绝非简易！

对于一个可以成功地和周围环境进行交互的系统，我们可以通过分析：系统和环境之间进行交互的物理过程，来理解（或部分理解）这个系统。对于视觉，这意味着：如果我们想从图像中恢复出关于“世界”或场景的信息，那么，我们就应该去理解成像过程。描述物理上的相互作用，自然会得出相应的数学方程。这些方程会启发我们，从而最终得出：从图像中恢复出关于“现实世界”（或场景）的信息的算法。这就是我的主题！令人感到惊讶的是：不少学生在把从书本中学到的抽象的数学方法，应用于实际问题的过程中，找到了乐

^①所谓启发式的方法（heuristic method）是指一些小技巧和小窍门。—译者注

趣。本书的材料为他们提供了一种动机，促使他们去实践这些方法，甚至去学习一些（他们以前不会费心去学的）新概念。

本书第一部分的重点是**早期视觉**，即：如何从图像中得到简单的符号描述。我们如何将这些描述用于空间推理以及行动规划？这方面的技术还没有得到很好的发展，并且，现有的这些技术所依赖的方法，并不适用于早期视觉。本书的最后五章将探索：如何使用这些简单的符号描述来生成一个复杂的结构化描述。在最后的章节中，我们介绍了：将视觉系统集成到一个完整的机器人系统中去的细节；在这一章里，我们建立了一个用于从容器中抓取零件的系统。

在写这本书的过程中，我花了大量的气力来设计习题。这些习题有着各种不同的目的。有的是帮助读者练习该章节中所涉及的方法；有的是用于启发出更深刻的思想（这可能会用到更复杂的数学工具）；有的是为了介绍新的研究课题。各章节中的习题都是根据这个顺序进行编排的。对于一些特别难的问题，我们给出了提示。现在所“流行”的一种趋势是：使用更加紧凑的符号表示形式，但是，本书选择使用一种更加连贯的形式，即：使用一致的符号将书中的内容组织在一起。对于向量（例如：曲面的法向量或光流的速度），以前的著作（在一些情况下）会将向量的元素具体写出来，但是，现在趋向于直接使用向量形式，例如：在讲述曲面的朝向时，现在使用 Gauss 球来取代梯度空间。在本书的正文中，我一直将向量的元素具体写出来，（因为这种表述方法在刚开始的时候更容易被学生理解）。然而，在一些课后习题中，我试着展示：如何通过使用紧凑的向量形式来更加有效地解决问题。

本书中的大多数材料都在其他地方出现过。但是，通过使用一致的表示形式，本书将它们以一种更加连贯的方式组织在了一起。在一些章节的例子中，本书也提供了一些以前没有发表过的新方法。这个领域的发展实在太快了，因此，几年之后，本书中的一些内容或方法可能会被淘汰，或者至少不那么有趣。相反的，本书中没有包含的一些东西，却可能最终成为一些让人兴奋的新结果的基础。但是，这并不是一个严重的缺陷，因为，相对于讨论那些针对某些特定问题的特殊方法，我更加关注于：发展出一种关于机器视觉的可靠研究方法。

什么时候会出现一种“通用的”视觉系统？我的回答是：至少在可预知的未来不会有。但是，这并不意味着：机器视觉只是一种没有实际应用的智力活动。相反的，在以下两个方面，机器视觉已经取得了巨大的进展：

- 视觉的某些特殊方面，例如：对于**立体像对**^①的解释。
- 视觉的某些具体应用，例如：自动装配过程中的校准。

一个真正意义上的“通用”视觉系统，必须能处理视觉中的所有方面，并且，它能够被用来处理所有只需要视觉信息就能够解决的问题；此外，它还将具有探索物质世界的能力。

B. K. P. H.

^①大脑根据进入左右眼的稍具差异的两幅图像产生景深效果，从而形成立体感，这两幅图像被称为**立体像对**(stereo pairs)。详见本书第13章的内容。—译者注

致 谢

我要感谢选我的“机器视觉”课程的学生，他们帮助我制定和修正了这部手稿。我的助教对本书的很多内容都有贡献。Robert Sjoberg对书中的几个课题做了认真整理，遗憾的是，由于时间紧迫，我没有参与他的整理工作。Robert Sjoberg、Andy Moulton、Eric Bier、Michael Gennert 和 Jazek Myczkowski 对本书的手稿提出了很多有用的意见。

书中的一些章节是基于我和其他人一起写的文章而整理出来的。我要感谢 Michael Brooks：他对于从明暗中恢复光照（第 11 章）内容的讨论做了很多贡献；Brian Schunck：他在开发光流分析方法（第 12 章）中给了我很多帮助；Anna Bruss：她在对无源导航问题（第 17 章）的分析中做了很大贡献；Katsushi Ikeuchi：他在零件抓取系统（第 18 章）的实现过程中做了很大贡献。

对于本书的早期版本，Christopher Brown、Herbert Freeman、Eric Grimson、Ramesh Jain、Alan Mackworth 和 Lothar Rossol 提出了很多宝贵的意见。Michael Brady、Michael Brooks、Michael Gennert 和 Ellen Hildreth 浏览了本书的新近版本，并给出了许多有用的建议。对于模式识别部分（第 14 章）的内容，Michael Gennert 做了很多贡献。Boris Katz 和 Larry Cohen 仔细地阅读了本书，并且，找出了书中的许多语言问题。不幸的是，在这个过程中，我“忍不住”不断地对书中内容进行修改和增删，因此，毫无疑问，在这个过程中，难免会出现笔误。

麻省理工学院（MIT）计算机与电子工程系给了我六个月的假期，用来完成这本书的初稿。Carol Robert 对本书的书稿进行了电脑输入。Blythe Heepe 绘制了本书中大部分的插图。Phyllis Roger 帮助我整理了参考文献。Michael Gennert 帮助我完成了本书最终印刷版的准备工作。本书封面上的图由艺术家 Hajime Sorayama 友情提供^①。

Marvin Minsky 建议将我的博士论文题目定为：从图像的亮度梯度中恢复物体的形状。从此，我开始研究机器视觉。从一开始，Patrick Winston 就对我关于机器视觉的研究非常支持，当时，机器视觉还不是一个非常“热门”的领域。Marvin 是 MIT 人工智能实验室的创立者，而 Patrick 是 MIT 人工智能实验室的继任者和开拓者。在 MIT 人工智能实验室，对机器视觉的研究已经活跃了近 20 年。

在研究本书中的许多方法的过程中，我得到了美国国防部前沿研究项目中心（DARPA）和海军研究办公室（ONR）的资助。

^①英文原著采用 Hajime Sorayama（空山基）先生的画作“弹钢琴的机器人”作为封面。我们选用了新的图画来作为中译本的封面。——译者注

目 录

| | |
|------------------------|----|
| 第1章 简介 | 1 |
| 1.1 机器视觉 | 1 |
| 1.2 机器视觉的任务 | 3 |
| 1.3 机器视觉和其他领域的关系 | 5 |
| 1.4 后续章节的概要 | 6 |
| 1.5 本章参考文献 | 14 |
| 1.6 习题 | 18 |
| 第2章 成像与图像检测 | 19 |
| 2.1 成像的两个方面 | 20 |
| 2.2 亮度 | 23 |
| 2.3 透镜 | 25 |
| 2.4 我们的视觉世界 | 29 |
| 2.5 图像检测 | 31 |
| 2.6 本章参考文献 | 41 |
| 2.7 习题 | 42 |
| 第3章 二值图：几何性质 | 49 |
| 3.1 二值图 | 50 |
| 3.2 简单几何性质 | 51 |
| 3.3 投影 | 57 |
| 3.4 离散二值图 | 61 |
| 3.5 行程编码 | 62 |
| 3.6 本章参考文献 | 65 |
| 3.7 习题 | 65 |