译 者 序

人们利用自己的眼、耳、鼻、嘴以及其他器官从周围的环境中获取信息，从而在纷繁的世界中生存与发展。因此，一旦产生了计算机，人们也希望计算机能够替代他们与周围的环境交换信息，这就导致了计算机视觉这门学科的产生与发展。

计算机视觉成为一门独立的学科，至少可以从美国麻省理工学院 Marr 教授这一代人所做的奠基工作开始，因此计算器视觉这门学科至少已有 20～30 年的历史。在此期间，由于计算机视觉具有的潜在应用十分广泛，所涉及的学科知识极其繁多，研究的问题又极富挑战性，因此它一直是计算机学科中的一门热门学科，并吸引了许多从事心理学、神经科学、生理学、生物物理学、数学与计算机等各种学科研究人员的关注，从而把图像处理、模式识别、人工智能、数学、认知科学、机器学习、计算机图形学等各方面的研究成果融会进来。

要从事计算机视觉的研究，就要先学好这门学科，掌握这门学科的基础理论与治学方法，掌握最必要的数学工具。因此选择一本合适的教材与参考书是十分关键的，我们翻译的这本书是近年来一本关于计算机视觉的新书，与先前的各种介绍计算机视觉原理的教科书与著作相比较，作者从他们多年从事研究的经历出发，敏锐地感到计算机视觉这门学科已经日趋成熟，已经能在实际应用中发挥极其重要的作用，他们抓住服务于应用这条线，对计算机视觉发展过程中积累的理论与成果进行了认真的挑选，从而编写了这样一本颇具特色的教科书：

- 作者力求该书既做到系统条理，又能各章相对独立，便于学生通读或选择部分内容阅读；
- 既讲清基本原理，又密切联系应用，使学生既能掌握基本原理又能与实际应用联系起来；
- 既不乏经典理论，又侧重近年的新鲜成果，使学生既了解计算机视觉的发展历史，又能把主要精力放在被实践证实为有效的近年研究的新成果上；
- 该书将必要的数学知识融入各相关章节中，具有深入浅出的效果；
- 作者为不同需求的学生设计了若干种不同的教学计划。该书的网络简版已在美国若干学校试用，反映颇好。

总之，这是近年来较成功的一本计算机视觉教材，也非常适合有兴趣的专业人才自学。

本书的主要译著者从事计算机视觉的研究与教学工作已有 20 年的历史，我们感到把这本书翻译成中文，对我国学习与研究计算机视觉的研究人员与学生会有帮助。

本书的第 1～5 章、第 12 章、第 13 章、第 19 章、第 20 章、第 22～26 章由林学訚及其学生马赓宇、杜杨洲、李旸和高华翻译，林学訚校阅；第 6～11 章、第 14～18 章、第 21 章由王宏及其学生董斌、于骞、安东、蔡文超、何英华、钱文杰、杨阳、叶明江翻译，王宏校阅。全书由林学訚校订。

在翻译过程中，我们力求忠实、准确地把握原著，同时保留原著风格。但是由于翻译时间仓促，原著涉及内容又十分广泛，难免有翻译不当乃至错误之处，殷切希望读者在学习过程中发现问题并随时与译著者联系，以便今后重印或再版时更正改进。

林学訚
王　宏
清华大学计算机系

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；在该领域经常可以看到缺乏权威性的现象；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，但是经常可以看到它们之间缺乏联系。尽管如此，我们还是力图在这本书中对这个领域给出一个相当有条理的分析。

我们认为计算机视觉，或简称为“视觉”，是一项事业，它与研究人类或动物的视觉是不同的。它借助于几何、物理和学习技术来构筑模型，从而用统计的方法来处理数据。因此从我们的角度看，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），视觉对每个像素值进行简单的推理（第二篇），将在多幅图像中可能得到的信息综合成和谐的整体（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），使用几何信息（第五篇）或概率统计技术（第六篇）来识别物体。计算机视觉具有很广泛的应用，有早期研究较多的应用（如移动机器人导航、工业检测、军事侦察），也有较新的应用（如人机交互、数字图书馆中用的图像检索、医学图像处理以及计算机图形学中合成场景的逼真绘制）。我们将在第七篇中讨论一些具体应用。

为什么要研究视觉

计算机视觉主要的诀窍在于从图像或图像序列中获取对世界的描述。毫无疑问，这是很有使用价值的。摄取图像通常不带破坏性，是安全的。它又是一件不费力的事，现在的成本也不昂贵。用户希望从图像中获取的描述对不同的应用可能相差很大。例如，一种称为从运动求取结构（Structure from motion）的技术可以从图像序列中获取所见物体的描述以及摄像机的运动规律。娱乐产业中的人们利用这种技术来构造建筑物的三维模型，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，人们一般关注运动，而将结构舍弃。这是因为，一般仅知道机器人工作的区域的某些方面，而并不知道机器人在这个区域的确切位置。可以从固定在机器人上的摄像机的运动信息中确定机器人的位置。

计算机视觉还有许多其他方面的重要应用，其中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。另一种应用是：检验人们对物体拍摄的图像，以便确定它们是否符合规定。第三种应用是卫星图像的理解。这既可用于军事目的，如编制程序来确定近来有没有与军事有关的现象在给定的地区内发生，或估计轰炸所引起的损害，也可以服务于民用的目的。今年的玉米收成会怎样，有多少雨林被保存下来？第四个应用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图书馆的文本（尽管这仍然是具有未解决难题的课题），但确实不知道对于图片或视频图书馆应该如何做。

计算机视觉正处于一个自身发展的关键时期。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；计算机和图像系统的价格已经很便宜。在不太长时间之前，得到好的数字彩色图像要花费上万美元，而现在至多几百美元就够了。在不太久之前，彩色打印机是很难看到的，如果有的话也往往在研究实验室中才能见到，而现在它们出现在许多家庭中。这就意味着搞研究工作变得容易了。这也意味着有许多人产生了一些需要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以及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以及如何运用它这一点已经得到了极大改善。我们正开始有可能解决许多人关注的问题，但是还没有一个困难的问题已得到解决，并且还有许多较容易的问题仍然没有解决(打算解决困难问题时要保持清醒的头脑)。现在正是研究这个主题的时候！

本书的内容

这本书包含的内容，我们认为从事计算机视觉研究的人们都应该知道。然而这本书是面对更广大的读者的。我们希望包括从事计算几何、计算机图形学、图像处理(总的说来与图像有关的)以及机器人学者会感到这是一本有益的参考书。我们试图做到让这本书对视觉这一课题有中等兴趣的本科高年级学生或研究生一年级的学生感到能接受。每一章覆盖这个题目不同的部分，而各章之间是相对独立的，这从表1就可瞥见。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。一般说来，我们已努力做到使每一章从容易的内容开始，把深奥难懂的内容放在最后。每一章末尾都有一个小结，包含历史性资料与相应的观点。我们努力使这本书叙述有用的概念，或者在今后或许会有用。我们把重点放在理解成像的基本几何学和物理学知识上，但是也力图把它们与实际应用联系起来。总的说来，这本书反映了几何学以及多种形式的应用统计学在近年来对计算机视觉的多方面的影响。

如果读者从头到尾阅读本书，是会有收获的。当然也会很累，这本书包含的内容在一个学期学习是太多了。当然，未来的(或现正从事)计算机视觉专业人员应该逐字阅读、做每一个练习、报告所发现的问题，以便在第二版中更正。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。我们已经努力使得该书能自成体系，这是指具有工科高年级数学理科水平的读者，对本书的内容会感到合适而不需要去参考其他课本。我们也注意到将数学降低到必不可少的程度，因为这本书是关于计算机视觉的，而不是应用数学方面的。我们还选择了将所保留的数学内容穿插在主要的章节内而不是单独的附录中的做法。

总而言之，我们已经努力减少各章之间的联系，以免只对某些专门话题感兴趣的读者在整本书中漫游。但做到各章完全自成体系是不可能的，各章之间的联系在表1中列出。

表 1　各章之间的关联程度："必读章节"一列给出的章节必须很好地理解，"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

部　分	章　号	必读章节	有助章节
第一部分	1.摄像机		
	2.摄像机的几何模型	1	
	3.摄像机的几何标定	2	
	4.幅射学——光亮度度量		
	5.光源、阴影与影调		4,1
	6.颜色		5
第二部分	7.线性滤波		
	8.边缘检测	7	
	9.纹理	7	8
第三部分	10.多视角几何学	3	
	11.立体视觉	10	
	12.从运动估计仿射模型	10	
	13.从运动估计投影模型	12	
第四部分	14.基于聚类的分割方法		9,6,5
	15.基于模型拟合的分割		14
	16.使用随机方法的分割与拟合		15,10
	17.基于线性动态模型的跟踪		
第五部分	18.基于模型的视觉	3	
	19.平滑曲面及其轮廓	2	
	20.外观图	19	
	21.距离数据		20,19,3
第六部分	22.利用分类器建立模板		9,8,7,6,5
	23.基于模板间关系的识别		9,8,7,6,5
	24.基于空间关系的几何模板	2,1	16,15,14
第七部分	25.应用:在数字化收藏库中查找		16,15,14,6
	26.应用:基于图像的绘制	1	13,12,11,6,5,3,2,1

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人感兴趣不是一件容易的事。为此我们不得不删掉一些素材，去掉一些主题等。在最后一刻，删掉了两个完整的章节：一个是对概率论和推理的介绍，另一个是用非线性动力学跟踪物体的方法的叙述。这些章出现在本书的网页 http://www.cs.berkeley.edu/~daf/book.html 上。

去掉某些主题取决于我们个人的判断，或者是由于我们已感到筋疲力尽而停止了撰写某些方面，或者是由于我们得知它们太晚了以至于无法将它们加进去，也可能由于我们不得不缩

短某些章，或许还有其他种种理由。我们有意忽略那些主要与历史有关的细节，而将历史的评论放在每章的末尾。我们不认为自己是个成功的知识考古学家，这意味着一些概念可能比我们所表达的有更深的历史意义。我们并没有回避撰写变形模板与拼图这两个具有相当重要实用意义的话题，我们将试图把它们放到第二版中去。

致谢

在筹备这本书的过程中，我们已经欠下了一大笔债。许多不知其名的评阅者已经读过这本书的几种草稿版本，并且对这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。编辑 Alan Apt 在 Jake Warde 的帮助下组织了评阅，对他们表示感谢。Integre Technical Publishing 的 Leslie Galen，Joe Albrecht 和 Dianne Parish 帮助解决了校对和说明方面的许多问题。书中使用的一些图像是从 IMSI's Master Photos Colloction 公司得到的。在筹集参考文献的过程中，我们大量使用了 Keith Price 提供的有关计算机视觉方面极好的参考文献，这可以从网上查到，地址为 http://iris.usc.edu/vision-notes/bibliogrophy/contents.html。

一些同事评阅了整本书的内容以及若干章节，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 Kobus Barnard，Margaret Fleck，David Kriegman，Jitendra Malik 和 Andrew Zisserman。我们的许多学生也在提建议、图示创意、校对评论以及其他方面做出了贡献。我们要感谢 Okan Arikan，Sébastien Blind，Martha Cepeda，Stephen Chenney，Frank Cho，Yakup Genc，John Haddon，Sergey Ioffe，Svetlana Lazebnik，Cathy Lec，Sung-il Pae，David Parks，Fred Rothganger，Attawith Sudsang 以及在加州大学伯克利分校 UIUC 上视觉课的一些学生们的贡献。幸运的是，一些大学的同事们在视觉课中用了这本书的草稿版本。使用过这本书初稿版本的学校有 Carnegie-Mellon 大学、Stanford 大学、在 Madison 的 Wisconsin 大学、在 Santa Barbare 的 California 大学以及 Southern California 大学，也可能有一些我们不知道的其他学校。我们对所有使用这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 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。这本书还受益于 Aydin Alaylioglu，Srinivas Akella，Marie Banich，Serge Belongie，Ajit M. Chaudhari，Navneet Dalal，Richard Hartley，Glen Healey，Mike Heath，Hayley Iben，Stéphanie Jonqnières，Tony Lewis，Benson Limketkai，Simon Maskell，Brian Milch，Tamara Miller，Cordelia Schmid，Brigitte 和 Gerry Serlin，Ilan Shimshoni，Eric de Sturler，Camillo J. Taylor，Jeff Thompson，Claire Vallant，Daniel S. Wilkerson，Jinghan Yu，Hao Zhang 和 Zhengyou Zhang。如果读者发现了明显的印刷排版错误，请给 DAF 发电子邮件(daf@cs.berkeley.edu)，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个发现者。

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，S. Mahamud，R. Mohr，H. Moravec，H. Murase，Y. Ohta，M. Okutami，M. Pollefeys，H. Saito，C. Schmid，S. Sullivan，C. Tomasi 以及 M. Turk，感谢他们为本书的某些插图提供了原件。

对教学大纲的建议

整本书从第一页开始讲起，可以分为两个学期讲，内容还是比较紧凑的。可以安排应用中的一章（如关于“基于图像绘制”这一章）放在第一学期，而将有关应用的另一章放在第二学期。但是估计有的系不会需要有如此详细的序列课。我们已在编排这本书时考虑到教师可以按他们的喜好选择其中的内容。表 2 到表 6 列出一些用于一学期 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。学生们可以学到这个领域的多个方面，包括在数字图书馆以及基于图像绘制等方面的应用。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，并建议在第二周或第三周读一下概率论的背景知识（网上的这本书有关于其的一章）。我们将应用章节放到书的末尾，但许多人可能会选择在第 10 周讲第 20 章，在第 6 周讲第 21 章。

表 2　对计算机科学、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一个学期计算机视觉的导论课的内容

周　次	章　号	节　号	主要内容
1	1,4	1.1,4(只限小结)	针孔摄像机，辐射学术语
2	5	5.1～5.5	局部影调模型，点、线和面光源，光度学体视
3	6	全部	颜色
4	7,8	7.1～7.5,8.1～8.3	线性滤波器，平滑抑制噪声，边缘检测
5	9	全部	纹理，用滤波器输出表示其统计量，纹理合成；由纹理推断形状
6	10,11	10.1,11	基本的多视角几何，立体视觉
7	14	全部	用聚类实现分割
8	15	15.1～15.4	拟合直线与曲线，用最大似然率进行拟合，鲁棒性
9	16	16.1,16.2	隐变量与 EM 算法
10	17	全部	用卡尔曼滤波来跟踪，数据相关
11	2,3	2.1,2.2,第 3 章全部	摄像机标定
12	18	全部	使用特征对应和摄像机标定的基于模型的视觉
13	22	全部	使用分类器的模板匹配
14	23	全部	基于关系的匹配
15	25,26	全部	在数字图书馆中检索图像，基于图像的绘制

表 3 所列教案是为计算机图形学的学生设计的。他们想知道与他们的课题有关的视觉基础知识。我们在此强调了能从图像提供信息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，并且对概率论有一定的了解。

表3　适用于计算机图形学学生用的教学大纲，他们想知道视觉与他们的课题有关的一些内容

周　次	章　号	节　号	主要内容
1	1,4	1.1,4(只限小结)	针孔摄像机，辐射学术语
2	5	5.1～5.5	局部影调模型，点、线和面光源，光度学体视
3	6.1～6.4	全部	颜色
4	7,8	7.1～7.5,8.1～8.3	线性滤波器，平滑抑制噪声，边缘检测
5	9	9.1～9.3	纹理，用滤波器输出表示其统计量，纹理合成
6	2,3	2.1,2.2,第3章全部	摄像机标定
7	10,11	10.1,11	基本的多视角几何，立体视觉
8	12	全部	由运动恢复仿射结构
9	13	全部	由运动恢复射影结构
10	26	全部	基于图像的绘制
11	15	全部	拟合，鲁棒性，RANSAC
12	16	全部	隐变量与EM算法
13	19	全部	表面与轮廓
14	21	全部	距离数据
15	17	全部	跟踪，卡尔曼滤波与数据相关

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，也可以在第2～3周时安排一个关于概率论的背景阅读。

表4　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	1,4	1.1,4(只限小结)	针孔摄像机，辐射学术语
2	5,6	5.1,5.3,5.4,5.5,6.1～6.4	局部影调模型；点、线和面光源；光度学体视；颜色-物理学，人的感知，颜色空间
3	2,3	全部	摄像机模型及其标定
4	7,9	第7章全部,9.1～9.3	线性滤波器，用滤波器输出表示纹理的统计量，纹理合成
5	10,11	全部	多视角几何，以立体视觉为例
6	12,13	全部	由运动推断仿射结构，由运动推断射影结构
7	13,26	全部	由运动推断射影结构，基于图像的绘制
8	14	全部	由聚类方法进行分割，重点在镜头边界检测与背景差分
9	15	全部	拟合直线、曲线，鲁棒性，RANSAC
10	16	全部	隐变量与EM算法
11	25	全部	在数字图书馆中检索图像

（续表）

周　次	章　号	节　号	主要内容
12	17	全部	跟踪，卡尔曼滤波与数据相关
13	18	全部	基于模型的视觉
14	22	全部	使用分类器检测模板
15	20	全部	距离数据

表5的教案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。学生需要在第2周或第3周学一些概率论的内容。

表5　对认知科学或人工智能学生的大纲，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	1,4	1,4(只限小结)	针孔摄像机，镜头，摄像机与人眼，辐射度学术语
2	5	全部	局部影调模型；点、线和面光源；光度学体视，互反射，光亮度计算
3	6	全部	颜色：物理，人的感知，空间，图像模型颜色恒常性
4	7	7.1～7.5,7.7	线性滤波器；采样；尺度
5	8	全部	边缘检测
6	9	全部	纹理；表达式，合成，由纹理推断形状
7	10.1,10.2	全部	基本的多视角几何
8	11	全部	立体视觉
9	14	全部	用聚类方法实现分割
10	15	全部	拟合直线、曲线，鲁棒性，RANSAC
11	16	全部	隐变量与EM算法
12	18	全部	基于模型的视觉
13	22	全部	用分类器检测模板
14	23	全部	基于模板间关系进行识别
15	24	全部	用空间关系建立几何模板

表6的教案是为那些对应用数学、电气工程或物理学有强烈兴趣的学生设计的。这个教案使一学期的内容很紧凑，进展很快，并且假设学生能够适应许多教学内容。我们假设学生具有宽广的背景知识，并且能够在第2周或第3周给他们指定阅读一些概率论的内容。我们在这样一个相当概括和要求甚高的教案中，安插了对数字图书馆的简短综述作为缓冲，也可以用基于图像绘制这一章或距离数据来代替。

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大纲

周　次	章　号	节　号	主要内容
1	1,4	全部	摄像机,辐射度学
2	5	全部	影调模型;点、线和面光源;光度学体视,互反射以及影调基元
3	6	全部	颜色-物理学,人的感知,空间,颜色恒常性
4	2,3	全部	摄像机参数与标定
5	7,8	全部	线性滤波器与边缘检测
6	8,9	全部	边缘检测;纹理:表达式,合成,推断形状
7	10,11	全部	多视角几何,以立体视觉为例
8	12,13	全部	由运动推断结构
9	14,15	全部	用聚类方法进行分割;拟合直线、曲线;鲁棒性;RANSAC
10	15,16	全部	拟合;隐变量与 EM 算法
11	17,25	全部	跟踪;卡尔曼滤波器,数据相关,数字图书馆中检索图像
12	18	全部	基于模型的视觉
13	19	全部	表面与它们的轮廓
14	20	全部	外观图
15	22	全部	模板匹配

编程作业和源程序

本书中给出的编程作业经常需要数值线性代数、奇异值分解以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 以及一些公共图书馆中得到,例如 LINPACK、LAPACK 和 MINPACK,它们可以从 Netlib 库中下载(http://www.netlib.org/)。我们在本书的网页 http://www.cs.berkley.edu/~daf/book.html 上也给出了一些至其他软件的链接。也可在此处找到编程作业的数据集或到数据集的链接。

目　　录

第一部分　图像生成与图像模型

第 1 章　摄像机 …… 2

1.1　针孔照相机 …… 2

1.2　带镜头的摄像机 …… 6

1.3　人的眼睛 …… 10

1.4　信号感应 …… 12

1.5　注释 …… 15

习题 …… 16

第 2 章　摄像机的几何模型 …… 17

2.1　欧几里得解析几何基础 …… 17

2.2　摄像机参数和透视投影 …… 23

2.3　仿射摄像机和仿射投影方程 …… 26

2.4　注释 …… 29

习题 …… 30

第 3 章　摄像机的几何标定 …… 32

3.1　最小二乘法的参数估计 …… 32

3.2　使用线性方法进行摄像机标定 …… 37

3.3　径向畸变 …… 40

3.4　分析摄影地形测量法 …… 42

3.5　应用:机器人定位 …… 43

3.6　注释 …… 44

习题 …… 45

第 4 章　辐射学——光亮度度量 …… 46

4.1　空间中的光 …… 46

4.2　到达表面的光 …… 50

4.3　重要的特殊情况 …… 53

4.4　注释 …… 56

习题 …… 57

第 5 章　光源、阴影与影调 …… 59

5.1　定性辐射学 …… 59

5.2　光源及其产生的效果 …… 60

5.3 局部影调模型 …… 65
5.4 应用:光度学体视 …… 68
5.5 互反射:全局影调模型 …… 74
5.6 注释 …… 79
习题 …… 81

第6章 颜色 …… 83
6.1 物理学中的颜色 …… 83
6.2 人类的颜色感知 …… 87
6.3 颜色表示 …… 90
6.4 图像颜色的一个模型 …… 97
6.5 从图像颜色中找到表面颜色 …… 103
6.6 注释 …… 109
习题 …… 111

第二部分 低层视觉:使用一幅图像

第7章 线性滤波 …… 114
7.1 线性滤波和卷积 …… 114
7.2 移不变线性系统 …… 118
7.3 空间频率和傅里叶变换 …… 123
7.4 采样和折叠失真 …… 126
7.5 滤波器与模板 …… 132
7.6 技术:归一化相关和检测模式 …… 133
7.7 技术:尺度和图像金字塔 …… 135
7.8 注释 …… 137
习题 …… 138

第8章 边缘检测 …… 140
8.1 噪声 …… 140
8.2 导数估计 …… 143
8.3 对边缘进行检测 …… 148
8.4 注释 …… 157
习题 …… 159

第9章 纹理 …… 161
9.1 纹理表示 …… 161
9.2 使用有方向性金字塔的分析(和合成) …… 167
9.3 应用:合成纹理来绘制 …… 174
9.4 由纹理得到形状 …… 176
9.5 注释 …… 179
习题 …… 180

第三部分　低层视觉:使用多幅图像

第 10 章　多视角几何学 …… 182
10.1　双视角 …… 182
10.2　三视图 …… 188
10.3　更多的视图 …… 192
10.4　注释 …… 195
习题 …… 197

第 11 章　立体视觉 …… 199
11.1　重建 …… 200
11.2　人类的立体视觉过程 …… 201
11.3　双目融合 …… 204
11.4　使用多个摄像机 …… 209
11.5　注释 …… 211
习题 …… 213

第 12 章　从运动估计仿射模型 …… 214
12.1　仿射几何基础 …… 215
12.2　仿射结构和两幅图之间的运动 …… 220
12.3　从多幅图像估计仿射结构和运动 …… 224
12.4　从仿射到欧氏图像 …… 226
12.5　仿射运动分割 …… 229
12.6　注释 …… 231
习题 …… 232

第 13 章　从运动估计投影模型 …… 234
13.1　投影几何基础 …… 234
13.2　从双目对应估计运动和投影结构 …… 243
13.3　多线性约束估计投影运动 …… 246
13.4　多幅图像恢复运动和投影结构 …… 247
13.5　从投影图像到欧氏图像 …… 250
13.6　注释 …… 252
习题 …… 252

第四部分　中 层 视 觉

第 14 章　基于聚类的分割方法 …… 256
14.1　什么是分割 …… 256
14.2　人类视觉:分类和格式塔原理 …… 258
14.3　应用:镜头的边界检测和背景差分 …… 263
14.4　基于像素点聚类的图像分割 …… 266

14.5 基于图论的聚类分割 …… 269
14.6 注释 …… 277
习题 …… 278

第 15 章 基于模型拟合的分割 …… 280
15.1 哈夫变换 …… 280
15.2 直线拟合 …… 283
15.3 拟合曲线 …… 286
15.4 作为概率问题的拟合 …… 290
15.5 鲁棒性 …… 291
15.6 举例:用 RANSAC 来拟合基础矩阵 …… 296
15.7 注释 …… 299
习题 …… 300
第 16 章 使用随机方法的分割与拟合 …… 302
16.1 丢失数据问题、拟合和分割 …… 302
16.2 EM 算法的应用 …… 306
16.3 模型选择:哪个模型拟合得最好 …… 315
16.4 注释 …… 317
习题 …… 318

第 17 章 基于线性动态模型的跟踪 …… 320
17.1 把跟踪作为一个抽象的推理问题 …… 320
17.2 线性动态模型 …… 322
17.3 卡尔曼滤波 …… 326
17.4 数据相关 …… 333
17.5 应用和例子 …… 336
17.6 注释 …… 340
习题 …… 340

第五部分 高层视觉几何方法

第 18 章 基于模型的视觉 …… 344
18.1 初始假设 …… 344
18.2 通过位姿一致性获取假设 …… 345
18.3 位姿聚类获得假设 …… 349
18.4 采用不变量获得假设 …… 351
18.5 校验 …… 357
18.6 应用:医学图像系统的对准 …… 359
18.7 曲面与对准 …… 363
18.8 注释 …… 363
习题 …… 365

第 19 章　平滑表面及其轮廓 …… 367
19.1　微分几何的基本要点 …… 368
19.2　表面轮廓几何学 …… 375
19.3　注释 …… 379
习题 …… 379

第 20 章　外观图 …… 381
20.1　视觉事件:微分几何的补充 …… 383
20.2　计算外观图 …… 391
20.3　外观图与物体定位 …… 395
20.4　注释 …… 398
习题 …… 399

第 21 章　距离数据 …… 401
21.1　主动距离传感器 …… 401
21.2　距离数据的分割 …… 402
21.3　距离图像的匹配和模型获取 …… 410
21.4　物体识别 …… 414
21.5　注释 …… 419
习题 …… 421

第六部分　高层视觉:基于概率和推理的方法

第 22 章　利用分类器建立模板 …… 424
22.1　分类器 …… 424
22.2　基于类直方图创建分类器 …… 431
22.3　特征选择 …… 434
22.4　神经元网络 …… 443
22.5　支持向量机 …… 451
22.6　注释 …… 455
习题 …… 457
22.7　附录 I:向后传播算法 …… 457
22.8　附录 II:线性不可分数据集上的支持向量机 …… 460
22.9　附录 III:非线性支持向量机 …… 461

第 23 章　基于模板间关系的识别 …… 463
23.1　通过对模板间关系投票检测物体 …… 463
23.2　利用概率模型及搜索的关系推理 …… 468
23.3　利用分类器简化搜索 …… 471
23.4　隐马尔可夫模型 …… 474
23.5　应用:基于隐马尔可夫模型的手语理解 …… 483
23.6　应用:基于隐马尔可夫模型的人体检测 …… 486

23.7 注释 …… 489

第 24 章 基于空间关系的几何模板 …… 491

24.1 物体与图像之间的简单关系 …… 491

24.2 基元、模板与几何推理 …… 498

24.3 后记:物体识别 …… 511

24.4 注释 …… 513

习题 …… 514

第七部分 应 用

第 25 章 应用:在数字化收藏库中查找 …… 518

25.1 背景知识:组织收藏的信息 …… 519

25.2 整幅图的概要表示 …… 522

25.3 图片的分部表示 …… 527

25.4 视频 …… 533

25.5 注释 …… 535

第 26 章 应用:基于图像的绘制 …… 536

26.1 从图像序列构造三维模型 …… 536

26.2 基于迁移的基于图像绘制方法 …… 543

26.3 光线场 …… 548

26.4 注释 …… 551

习题 …… 552

第一部分　图像生成与图像模型

- 第 1 章　摄像机
- 第 2 章　摄像机的几何模型
- 第 3 章　摄像机的几何标定
- 第 4 章　辐射学——光亮度度量
- 第 5 章　光源、阴影与影调
- 第 6 章　颜色

第1章 摄 像 机

成像设备有许多种类，从动物的眼睛到视频摄像机和雷达望远镜，它们可以装有镜头，也可以没有镜头。例如，16世纪发明的最早的照相机暗箱模型并没有镜头，而是使用一个针孔将光线聚焦到墙上或半透明的屏幕上，并且演示了在一个世纪前 Brunelleschi 发现的透视规律。早在1550年，针孔已被越来越复杂的镜头所代替，而现代的照相机或数字摄像机仍是一个摄像机暗箱，但它能够将照射到底板的每一个小区域的光强度记录下来(见图1.1)。

图1.1 图像成像在一个照相机的底板上，图像来自美国海军基础光学和光学仪器手册

一般的照相机的成像面是矩形的，而人类视网膜的形状接近于球的表面，全景照相机则安装了柱形的视网膜。成像传感器具有不同的特性，它们可以记录空间离散的图像(像我们眼睛中的视杆细胞和圆锥细胞、35毫米照相机的颗粒以及数字摄像机的矩形图像元素或像素)或连续的图像(例如早期类型的电视摄像管)。图像传感器在它的视网膜上每一点记录的信号可以是离散量或连续量，它可以由单个数字组成(黑白照相机)，或由若干个数字组成(如彩色照相机的红、绿、蓝成分的强度或人眼三种类型圆锥细胞的响应)，或者由许多数目的数字组成(例如超光谱传感器的响应)，或由波长的连续函数组成(光谱仪基本上属于这种情况)。探讨这些特性是本章的主题。

1.1 针孔照相机

1.1.1 透视投影

可以想像一下，将一个盒子的一侧扎一个小孔，然后将另一侧改成一块半透明板。如果在一个较暗的屋子里将这个盒子放在你面前，将针孔对准某种光源(譬如说蜡烛)，你可以在半透明板上看到颠倒的蜡烛图像(见图1.2)，这个图像是从景物投射到盒子的光线形成的。如果

假设针孔可以缩小成一个点的话(当然在物理上是不可能的),那么就只有惟一的一条光线穿过三个点:成像板的平面(或称为成像面)上的一个点、针孔以及景物中的某个点。

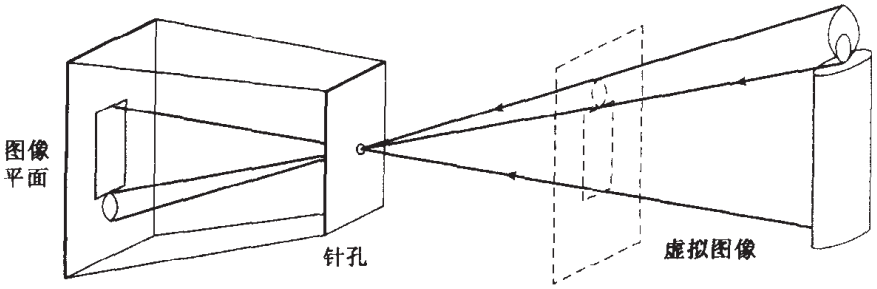


图 1.2　针孔成像模型

在现实中,针孔(不管多小)总不是无限小的,成像平面上的每个点收集的是具有一定角度的锥形光束的光线,因此严格说来理想化的、极其简单的成像几何模型是不成立的。再加上实际的照相机一般都配备有镜头,因此使得事情更加复杂。然而,15 世纪初由 Brunelleschi 首先提出的针孔透视投影模型(或称中心透视投影)在数学上是很方便的。这个模型尽管简单,但是它对成像过程的近似程度往往是可以接受的。透视投影产生的是一幅颠倒的图像,因此有时设想一个虚拟图像会方便一些,这幅图像落在一个处于针孔前面的平面上,它到针孔的距离等于实际成像面到针孔的距离(见图 1.2)。这幅虚拟图像除了图像是倒立的以外,与实际图像是完全等价的。根据所考虑的情况选择其中任一种会显得更加方便。图 1.3(a)表明了透视投影的明显效果:所观察到的物体的大小取决于它们的距离。例如,杆 B 和杆 C 的图像,B' 和 C' 具有相同的高度,但实际上杆 A 与杆 C 的尺寸只是杆 B 的一半。图 1.3(b)表现了另一个众所周知的现象:同一平面 Π 上的两条平行线的投影在成像面上将会聚到(在成像面上)一条水平线 H 上,H 这条线是穿过针孔与 Π 平行的平面与成像面相交的交线。还需指出的一点是,平面 Π 上与成像面平行的线 L 在成像面上没有图像。

这些性质很容易用纯几何方式证明,然而使用参考框架、坐标和方程式来推理也很方便(尽管并不十分优雅)。例如,将一个坐标系$(O,\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$附加到一个针孔摄像机上去,它的原点 O 与针孔重合,而向量 $\boldsymbol{i}$ 与 $\boldsymbol{j}$ 组成一个与图像平面 Π' 平行的向量平面的基,Π' 平面位于沿 $\boldsymbol{k}$ 向量正方向距离针孔 f' 处(见图 1.4)。通过针孔又垂直于 Π' 的线称为光轴,其穿过 Π' 的点 C' 称为图像中心。这个点可以作为图像平面坐标系统的原点,这在摄像机定标过程中起重要作用。

如果用 P 表示景物中坐标为(x,y,z)的一点,P' 是它的图像,坐标为(x',y',z')。因为 P' 处在图像平面中,所以有 $z'=f'$。又因为 P,O,P' 这三个点共线,则应有$\overrightarrow{OP'}=\lambda\overrightarrow{OP}$,$\lambda$为某个数,所以

$$\begin{cases} x'=\lambda x \\ y'=\lambda y \\ f'=\lambda z \end{cases} \Longleftrightarrow \lambda=\frac{x'}{x}=\frac{y'}{y}=\frac{f'}{z}$$

因此有

$$\begin{cases} x'=f'\dfrac{x}{z} \\ y'=f'\dfrac{y}{z} \end{cases} \tag{1.1}$$

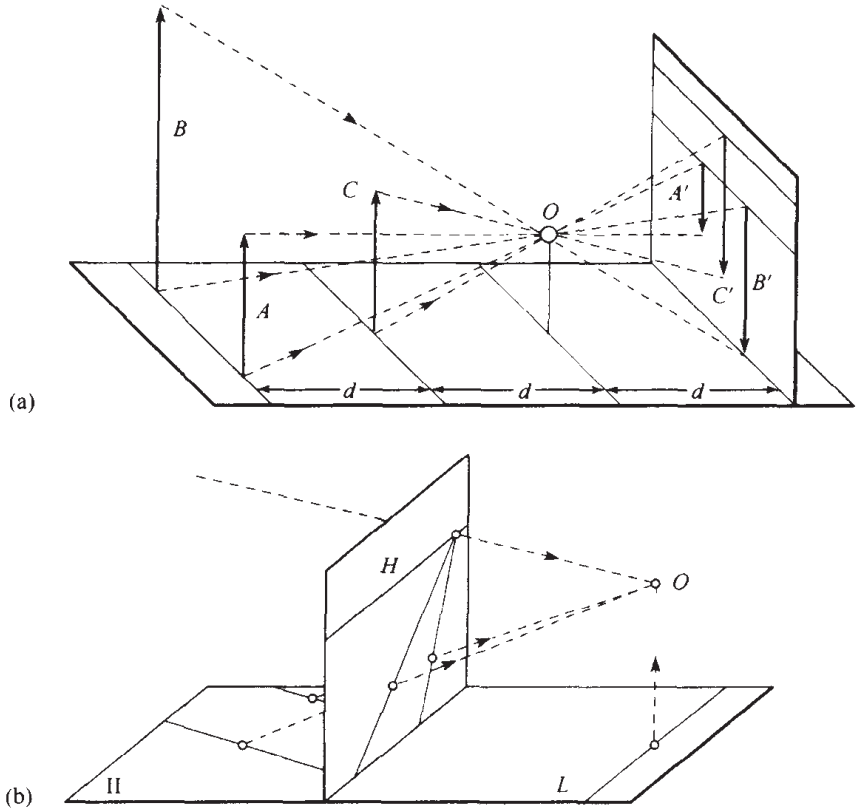


图 1.3 透视效果:(a)距离远的物体看上去比近的物体小:从针孔 O 到包含 C 的平面的距离是到包含 A 与 B 平面距离的一半;(b)平行线的图像与一条水平线相交。注意图像平面在(a)中处在针孔后面(物理视网膜),在(b)中处在它的前面(虚拟成像面)。这一章以及本书其余章节中的大部分图都采用物理成像面,但是只要合适,也采用虚拟图像平面

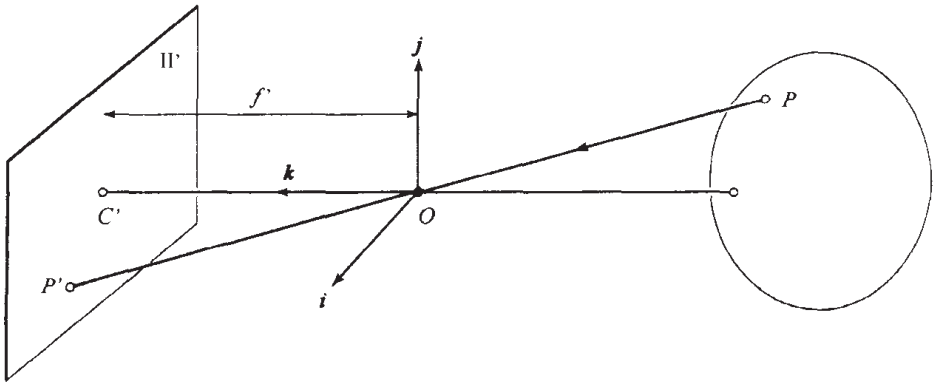


图 1.4 从点 P、它的图像点 P' 与针孔 O 三个点共线推出这一节的透视投影方程式

1.1.2 仿射投影

正如前一节中指出的,针孔透视仅仅是成像过程中几何关系的一种近似。本节讨论的是一种更加粗略的近似,称为仿射投影模型,它在某些情况下也是很有用的。我们的注意力将集中在两种特定的仿射模型——弱透视和正交投影上,第三种称为类透视模型,将在第 12 章讨

论，同时也将说明仿射投影的含义。

考虑定义在 $z = z_0$ 上的朝前平行的平面 Π_0（见图 1.5），对 Π_0 平面上的任一点 P，式(1.1)的透视投影可以改写成

$$\begin{cases} x' = -mx \\ y' = -my \end{cases} \quad 其中 \quad m = -\frac{f'}{z_0} \tag{1.2}$$

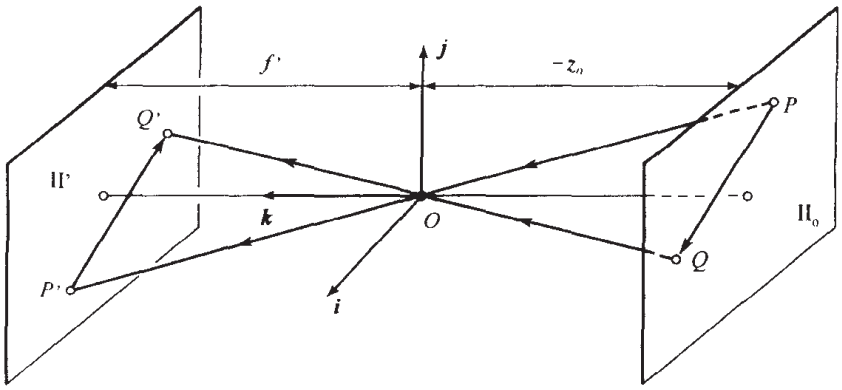


图 1.5　弱透视投影：Π_0 平面上的所有线段以同等放大率投影到图像上

由于物理上的约束，z_0 必为一个负值（该平面必然处在针孔的前面），与平面 Π_0 关联的放大率 m 是个正值，称 m 是放大率是因为以下理由：设想 Π_0 平面上的两个点 P 与 Q，以及它们的图像 P' 与 Q'（见图 1.5），显然 $\overrightarrow{PQ}$ 与 $\overrightarrow{P'Q'}$ 平行，因此有 $|\overrightarrow{P'Q'}| = m|\overrightarrow{PQ}|$，这就是前面提到的图像尺寸与物体距离之间的依赖关系。

当景物深度与它们到摄像机的平均距离相比很小时，这个放大率可以看做是一个常数，这种投影模型称为弱透视模型或按比例的正交。如果预先知道摄像机到景物的距离大体保持常数，则可以进一步将图像坐标归一化，使得 $m = -1$。这就是正交投影，定义为

$$\begin{cases} x' = x \\ y' = y \end{cases} \tag{1.3}$$

在这种情况下，所有光线与 $\boldsymbol{k}$ 轴平行，并与图像平面 Π' 正交（见图 1.6）。尽管弱透视投影在许多成像条件下是可以接受的，但假设纯正交投影通常是不现实的。

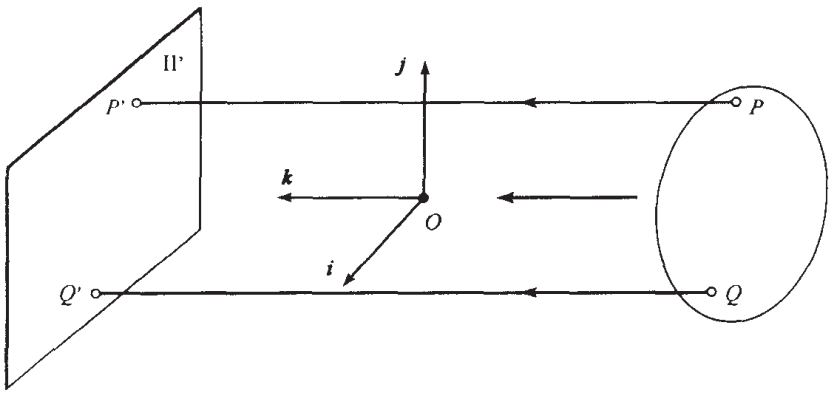


图 1.6　正交投影：与成像过程的其他模型不一样，正交投影没有包含图像特征的倒置，因此放大率定义为负值。这显得略有些不自然，但简化了投影方程

1.2 带镜头的摄像机

大部分摄像机都配备有镜头。使用镜头主要有两个理由：第一个理由是收集光线，因为在理想的针孔模型中，一条光线会投影到图像平面上的一个点。当然，实际的针孔是有尺寸的，所以图像平面中的每一个点是由一定角度范围射来的锥形光束照亮的。针孔的尺寸越大，这个锥形光束越宽，因此图像也越明亮，但是大尺寸的针孔会导致图像模糊。缩小针孔能使图像锐化，但减少了到达图像平面光的总量，并且可能产生衍射现象。使用镜头的第二个理由是能保持图像锐化聚焦，同时又可从较大面积中收集光线。

如果忽略衍射、交叉反射和其他物理光学现象，镜头的性能则服从于几何光学的定律（见图 1.7）：(1)在均匀介质中，光以直线反射（光射线）；(2)当一条光线从一个表面反射出来时，这条射线与它的反射光以及该表面的法线是共平面的，法线与这两条光线之间的夹角是对称的（译者注：原文是互补的）；(3)当光线从一种介质进入到另一种介质时，要发生折射（也就是它的方向会改变）。按照 Snell 规则，如果投射到两个透明材料界面的光线为 r_1，这两个材料的折射率分别为 n_1 与 n_2，r_2 表示折射光，那么 r_1，r_2 与该界面的法线共平面，法线与这两条光线的夹角 α_1 与 α_2 之间的关系是

$$n_1 \sin\alpha_1 = n_2 \sin\alpha_2 \tag{1.4}$$

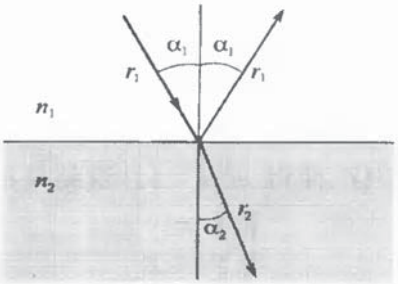


图 1.7 两均匀介质之间的界面上发生的反射与折射，折射系数分别为 n_1 与 n_2

本章仅考虑折射的效果，而忽略反射的效果。换句话说，我们只考虑透镜，而不考虑反射折射的光学系统（例如望远镜），后者可能包括反射（镜像）和折射成分。假设光线与镜头的折射表面之间的夹角较小时，跟踪光线传播路径是比较简单的。下一节将讨论这种情况。

1.2.1 近轴几何光学

在这一节考虑符合近轴（或一阶）几何光学的情况，也就是说，所有进入镜头的光线之间与镜头折射表面法线之间的夹角相对较小。此外我们还假设镜头围绕一个称为光轴的直线是旋转对称的，并且所有折射表面都是圆形的，这种对称性设置使我们能把镜头看做是在包含光轴的平面中具有圆形边界的情况来讨论投影几何。

设想有一入射光穿过光轴上的点 P_1，并在点 P 折射，圆形折射界面的半径为 R，两个透明介质的折射系数分别为 n_1 与 n_2（见图 1.8），折射的光线第二次穿过光轴时的点表示成 P_2（P_1 与 P_2 的作用完全是对称的），图形介面的中心为 C。

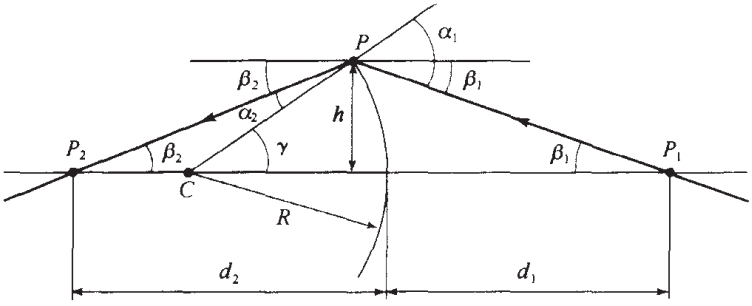


图 1.8 近轴折射:一条穿过 P_1 的光线在点 P 与一个圆形界面相交,并发生折射。折射光穿过点 P_2,界面中心在光轴上的点 C,半径为 R。假设角度 α_1,β_1,α_2 与 β_2 都很小

α_1 与 α_2 分别表示连接 C 到 P 的弦与两条射线之间的夹角,β_1(相应的 β_2)是光轴与连接 P_1(相应的 P_2)到 P 的连线之间的夹角,则光轴与 C 到 P 的连线之间的夹角,按图 1.8 所示为 $\gamma = \alpha_1 - \beta_1 = \alpha_2 + \beta_2$。如果用 h 表示点 P 到光轴的距离,R 是圆形界面的半径,假设所有角度都很小,那么按一阶近似,它们都等于它们的正弦和正切值,因此有

$$\alpha_1 = \gamma + \beta_1 \approx h\left(\frac{1}{R} + \frac{1}{d_1}\right) \quad 和 \quad \alpha_2 = \gamma - \beta_2 \approx h\left(\frac{1}{R} - \frac{1}{d_2}\right)$$

将 Snell 规则用到小角度情况,就得到近轴折射方程式:

$$n_1\alpha_1 \approx n_2\alpha_2 \Longleftrightarrow \frac{n_1}{d_1} + \frac{n_2}{d_2} = \frac{n_2 - n_1}{R} \tag{1.5}$$

需要指出的是,d_1 与 d_2 之间的关系取决于 R,n_1 与 n_2,而与 β_1 及 β_2 无关。这是近轴假设引出的主要简化。显而易见,当 d_1,d_2 和 R 的某些值或全部值变为负值时,式(1.5)仍然有效,这相当于点 P_1,P_2 或 C 换到了另一边。

当然,实际镜头至少有两个折射表面为界。光线的路径可以使用近轴折射方程迭代构造。下一节将讨论薄透镜条件下的情况。

1.2.2 薄透镜

以下考虑一个透镜具有两个半径为 R 的球形表面,折射系数为 n 的情况。假设镜头处在真空中(在空气中是一个很好的近似),折射系数为 1,是薄透镜(也就是说进入透镜的光线从右边边界折射后立即又在左边界上再次折射)。

假设一个不在光轴上的点 P 处在(负)深度 z 处,(PO)表示穿过该点与透镜中心 O 的射线(见图 1.9),正如练习中所示,它遵循 Snell 规则和式(1.5),因此(PO)射线没有折射,而其他穿过点 P 的射线都由薄透镜聚焦到沿(PO)处于 z' 深度的点 P',满足

$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f} \tag{1.6}$$

其中,$f = \dfrac{R}{2(n-1)}$ 是透镜的焦距长度。

需要指出的是,如果我们取 $z' = f'$,则表示 P 与 P' 位置之间关系的方程式与针孔透视投影条件下的情况完全相同,因为 P 与 P' 处在穿过透镜中心的射线上,但是,当图像平面处在透镜一边距点 O 的距离为 z' 处时,只有距点 O 的距离为 $-z$ 的点满足式(1.6)(薄透镜方程)并

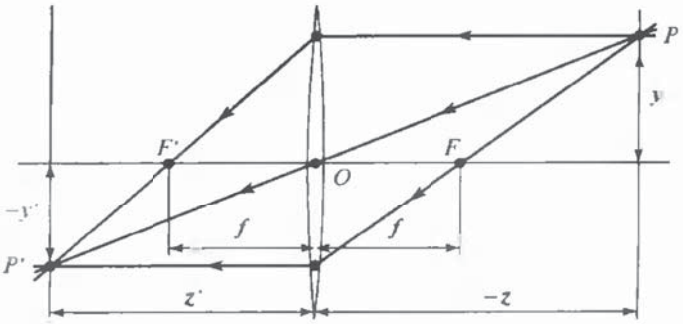


图1.9 薄透镜。穿过点 O 的光线不发生折射。平行于光轴的光线聚焦到焦点 F'

严格聚焦。如果令 $z \to -\infty$，就像星星处在 $z=-\infty$ 处，那么 f 就是透镜中心与聚焦物体所在平面之间的距离。光轴上距离透镜中心为 f 的两个点 F 和 F' 称为透镜的焦点。

实际上，在某个距离范围内的物体聚焦的程度是可以接受的(称为视野深度或聚焦深度)。正如习题中显示的那样，视野深度随着透镜 f 数值的增加而增加(也就是镜头聚焦长度与它的直径之间的比值)。摄像机的视野是指景物中实际投影到摄像机视网膜上的部分，它不仅取决于聚焦长度，还取决于视网膜的有效面积(例如照相机中能曝光的胶片的面积，或数字摄像机中 CCD 传感器的面积，见图 1.10)。

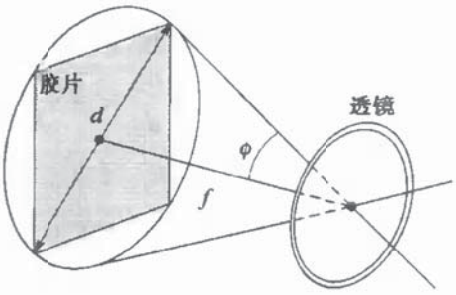


图1.10 摄像机的视野是 2ϕ，此处 $\phi \stackrel{\text{def}}{=} \arctan \frac{d}{2f}$，$d$ 是传感器的直径(胶片或 CCD 芯片)，f 是摄像机的焦距。当 f 远小于 d 时，构成广角镜头，射线方向可以偏离光轴多至45°。望远镜透镜视野很小，产生的图像接近于仿射图像

1.2.3 实际透镜

更为实际的简单光学系统是厚透镜。描述厚透镜性能的方程式很容易从近轴折射方程式导出，除了计算 z 与 z' 有一个偏移之外，它们在其他方面与针孔透视模型以及薄透镜投影方程式相同(见图 1.11)：如果 H 与 H' 表示透镜的主点，那么将点 P 与穿过点 H 且垂直于光轴的平面的距离表示成 $-z$，点 P' 与穿过点 H' 且垂直于光轴的平面的距离表示成 z' 时，式(1.6)成立。在这种情况下，惟一没发生折射的光线是沿着光轴的光线。

简单透镜会产生若干种像差。为了理解其中的原因，要记住，近轴折射公式(1.5)只是近似的，只有当光线与透镜光轴之间夹角 α 较小且 $\sin\alpha \approx \alpha$ 时它才有效。对较大的角度，利用正弦函数的三阶泰勒展开式可以改进近轴方程式：

$$\frac{n_1}{d_1}+\frac{n_2}{d_2}=\frac{n_2-n_1}{R}+h^2\left[\frac{n_1}{2d_1}\left(\frac{1}{R}+\frac{1}{d_1}\right)^2+\frac{n_2}{2d_2}\left(\frac{1}{R}-\frac{1}{d_2}\right)^2\right]$$

其中，h 与图 1.8 中相似，表示入射光线与界面交点到光轴之间的距离。需要强调的是，入射光击中界面的点距光轴越远，聚焦越接近界面。

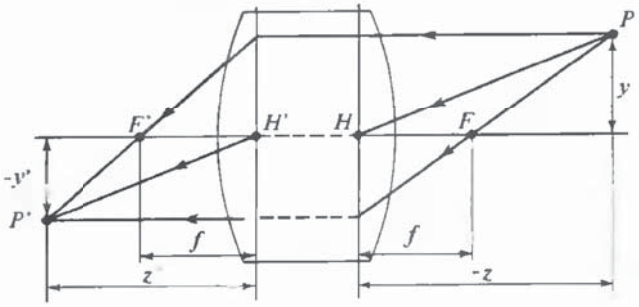


图 1.11 具有两个球形表面的简单厚透镜

对一个透镜也有相同的现象，它会引起两种类型的球形像差[见图 1.12(a)]。考虑光轴上一点 P 与它的近轴图像 P'。从点 P 发出的光线经过透镜反射后，其与光轴的交点与 P' 之间的距离称为该光线的轴向球形像差。需要指出的是，如果图像平面 Π' 矗立在 P'，则该折射光线会在与该平面相交处偏离轴线，这称为该光线的横向球形像差。这样一来，所有过点 P 的光线经透镜折射后与 Π' 平面相交的点形成了以 P' 为中心的模糊圈。当平面 Π' 沿光轴移动时，这个圈的尺寸也会改变，其中具有最小直径的圈称为最小模糊圈，一般情况下它不在 P' 处。

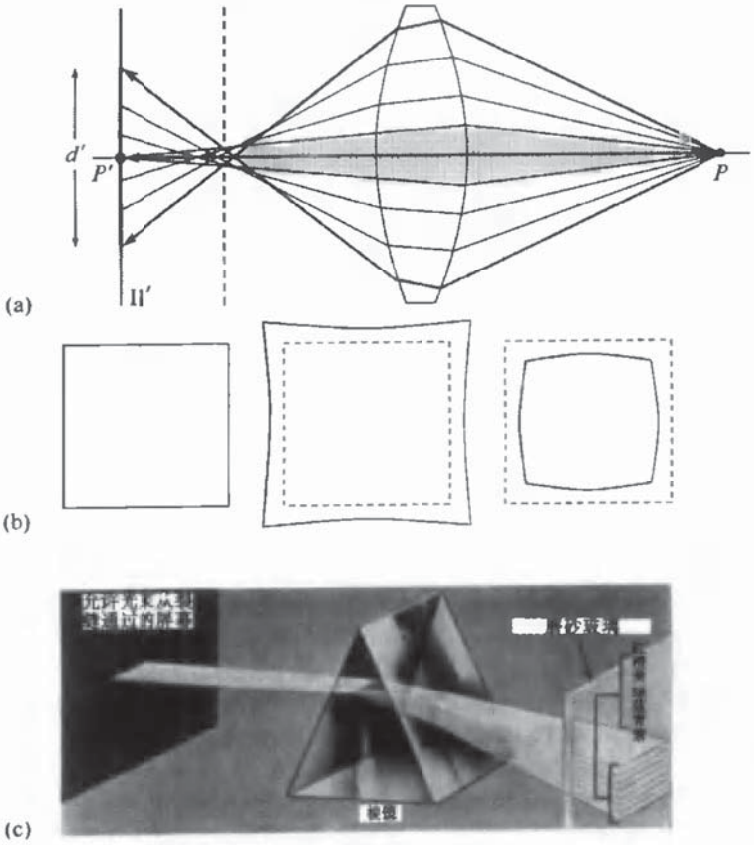


图 1.12 像差。(a)球面像差：灰色区域是近轴带，在该带中从点 P 发出的光线交到近轴图像点 P'。如果图像平面 Π' 在 P' 矗立，P' 在该平面的图像形成一个直径为 d' 的模糊圈。得到最小模糊圈的聚焦平面用虚线表示；(b)畸变：从左到右：正面平行的正方形、枕形失真与桶形失真；(c)色差：透明介质的折射系数取决于入射光的波长(或色彩)。图中一个棱镜将白色光分解成一个调色板

除了球形像差外，还有4种因一阶与三阶光学之差引起的其他类型的基本像差：慧形像差、散光、场曲率和畸变。对这些像差的定义不是本书的范围，只要知道，它们与球面像差一样会使物体每一点的图像变模糊而造成图像变差。而畸变的作用不太一样，它改变整个图像的形状[图1.12(b)]，这种效果由于透镜的不同区域的焦距略有不同。以上提到的像差是单色的(也就是说，透镜对不同波长的响应是独立的)。然而透明介质的折射系数与波长有关[图1.12(c)]，根据薄透镜方程式(1.6)，焦距也取决于波长。这导致色差现象：根据不同的波长，折射光线与光轴相交在不同的点(纵向色差)，并在同一图像平面上形成不同的模糊圈(横向色差)。

像差可以用几片简单透镜装配起来的方法来减小，这些简单透镜的形状与折射系数要选择适当，并且要用合适的光圈隔开。这种复合镜头仍然可以用厚透镜方程作为模型。对机器视觉来说，这种透镜还有一个不足：从物体点射出的光束有一部分会被各种光阑阻挡，这些光阑放在透镜内部来限制像差(见图1.13)。这种现象称为渐晕效应，其使图像周边区域的光亮度降低。渐晕效应可能对图像的自动分析程序产生影响，但对照片并不很重要，这是因为人眼对光亮度平滑的梯度很不敏感。谈到眼睛，现在是我们对这个非凡的器官观察得稍微细致一些的时候了。

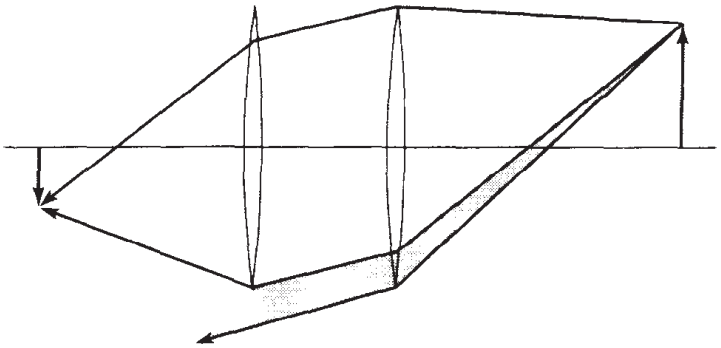


图1.13 两个透镜系统中的渐晕效应。光束的阴影部分到不了第二个透镜。透镜中的附加光圈与光阑使渐晕效应加重

1.3 人的眼睛

这里简单概括人眼的解剖学结构。这部分主要基于Wandell(1995)的著作，有兴趣的读者如想了解更多细节，请阅读一下这本优秀的书。图1.14(左)是眼球沿其垂直对称面部分的轮廓图，展示了眼睛的主要成分：虹膜和瞳孔控制投影到眼球上的光亮；角膜与水晶体透镜共同作用将光折射生成视网膜图像；最后面的是视网膜，图像在这里形成。

尽管眼睛呈球状，它的功能与一个摄像机相似，其视野覆盖宽160°、高135°的区域。与任何其他光学系统一样，它也有各种各样的几何与色彩像差。人们已提出若干个服从一阶几何光学规律的人眼模型，图1.14(右)表示了其中的一个，称为Helmoltz图解眼。其中只有三个折射表面、无限薄的角膜以及均质的透镜。在图1.14中给出的数字是眼睛聚焦于无穷远时的状况(未调节的眼睛)。这个模型只是眼睛实际光学特征的一个近似。

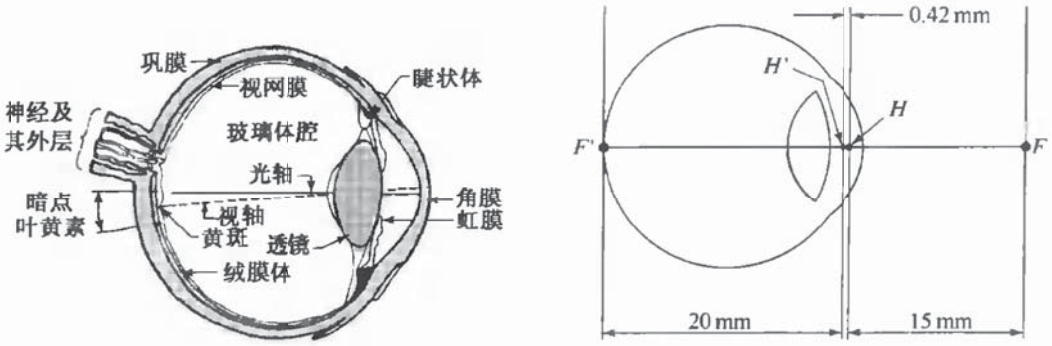


图 1.14 左图:人眼的主要组成;右图:按 Laurance 修改的 Helmoltz 图解眼(仿照 Driscoll与Vaughan,1978)。角膜极点到前主平面的距离是1.96毫米,角膜、透镜的前、后表面的半径分别为8毫米、10毫米和6毫米

下面对人眼的组成分层进行进一步的分析:角膜是透明的、高度弯曲的折射窗口;光线通过它进入眼内,随后有部分被带色的不透明的虹膜表面所阻挡。瞳孔是一个处于虹膜中心的孔,它的直径随照明改变而改变,可从 1 毫米变到 8 毫米。光线暗时它扩张,以便增加到达视网膜的能量,在正常光照条件下它收缩,以限制眼内因球面像差引起的图像模糊程度。眼睛的折射率(焦距的倒数)主要是空气与角膜界面处的折射效果,它可以通过晶体透视的变形来微调,使得眼前的物体聚焦清晰。对一个健康成人,它从 60 屈光度单位(未调节眼)变到 68 屈光度单位(1 屈光度单位 = 1 米$^{-1}$),对应焦距范围为 15 到 17 毫米。视网膜本身是一个多层薄膜,布满了两种光感受器—— 视杆细胞和圆锥细胞,它们响应为 330 ~ 730 纳米波长范围(紫到红)的光。正如在第 6 章中提到的,有三种不同类型的圆锥细胞,分别具有不同的谱敏感度,它们在感受彩色中起关键作用。一只人眼中大约有上亿个视杆细胞及 500 万个圆锥细胞,它们在视网膜上的分布是不均匀的。视网膜中心区域是黄斑,在那里圆锥细胞密集程度特别高。当眼睛注视一个物体时,它的图像清晰地聚焦在那里(见图 1.14)。圆锥细胞最集中的区域在黄斑中心视网膜凹斑,密度峰值高达 1.6×10^5/毫米^2,两个相邻的圆锥细胞中心仅相隔半分视角,但稍微偏离中心密度就急剧下落(见图 1.15)。与之相反的是,在视网膜凹中没有视杆细胞,但视杆细胞的密度在趋向于视场的边缘部分增加。视网膜上还有一个盲点,神经中枢细胞轴突从那里离开视网膜,组成视神经。

视杆细胞是非常灵敏的光感受器,它们甚至能响应单个量子。它们的数量尽管很大,但获得空间细节的能力相对较差,这是因为许多杆细胞在视网膜内会聚到相同的神经上。与之相反,圆锥细胞在较高亮度时活跃起来,视网膜凹上每个圆锥细胞的信号输出是由若干个神经编码的,因此在这个区域得到高的分辨率。一般说来,对神经响应产生影响的视网膜区域称为感受场,尽管这个词现在也经常表示神经对光实际的电响应。

当然人眼还有许多方面可以(也应该)讨论,例如,我们的两只眼睛是如何会聚与瞄准目标的,如何在立体视觉中相互合作的,等等。除此之外,视觉只是从我们头脑中的这个摄像机开始,并引导到许多令人神往(以及仍然没很好解决)的问题,破译我们脑子中各部分在人类视觉中所起的作用。本书的后续章节还将讨论人类这种努力的一些方面。

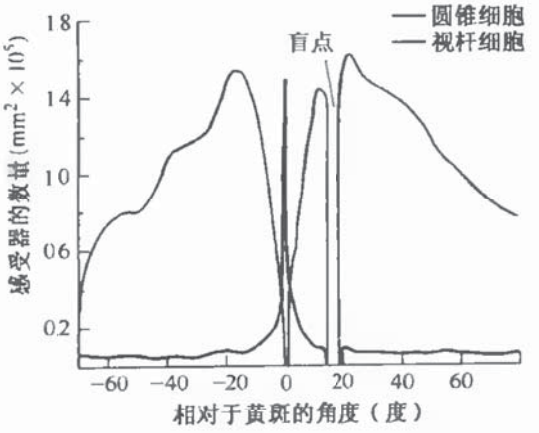


图 1.15 视杆细胞和圆锥细胞在视网膜上的分布

1.4 信号感应

现代照相机与 17 世纪的便携式照相机暗盒的区别在于记录成像于它的背板的能力。尽管某种银盐在阳光作用下会迅速变黑这个事实早在中世纪已为人们所知，但是直到 1816 年 Niepce 才获得第一张真正的照片，它将银氯化物浸泡过的纸放在照相机暗盒的成像平面上用光线曝光，然后用硝酸定影。最初的相片是负片，而 Niepce 不久就转向使用别的光敏化学材料获得正片。最早期的照片已经找不到了，而被保存下来的第一张照片是 la table servie(餐桌)，重印于图 1.16。

图 1.16 保存下来的第一张照片 la table servie，由 Nicéphore Niepce 在 1822 年得到

Niepce 发明了照相术，但 Daguerre 却是推广这项技术的人。他们俩在 1826 年结成伙伴之后，Daguerre 继续开发他自己的照相事业，他用含汞烟雾放大和显示在镀碘化银的铜底版上形成的隐含图像。当 Arago 于 1839 年在法国科学院展示 Daguerre 的工作时，银板照相法立刻取得成功，这是 Niepce 去世后三年的事。在漫长的照相术历史中，其他里程碑包括 Legray 和 Archer 在 1850 年创造的负/正湿片法，它产生很出色的负片，但要求图片及时处理；Maddox 于 1870 年发明的胶板技术，不需要即刻处理；Eastman 于 1889 年引入的照相胶片(在大部分现代应用中取代了玻璃板)；以及 Lumière 兄弟于 1895 年发明的照相机及 1908 年的彩色照相术。

由 Baird，Farnsworth 以及 Zworykin 等人于 20 世纪 20 年代发明的电视机显然是开发电子传

感器的起源。摄像管是普通类型的电视真空管，它是一个玻璃容器，一头安装有电子枪，另一头是荧光屏。荧光屏背面有一薄层光导材料，其上又覆盖一层充正电金属的透明薄膜，这个双重覆盖物构成靶面。真空管用调焦和偏转线圈缠绕着，它们的用途是用由电子枪产生的电子束反复扫描这个靶。电子束投射一层电子到靶面上以平衡其上的正电子。当荧光屏的一小块区域被光线击中，电子流产生并局部消耗了靶面的电荷。当电子束扫描到这个区域，它补充了丢失的电子，产生了一个与入射光强成正比的电流。摄像管电路随后将电流变化转换成电视信号。

1.4.1 CCD 摄像机

电荷耦合器件(CCD)摄像机于 1970 年推出，它已经在大部分现代应用中取代了摄像管摄像机，从消费者用的摄像机到适合显微镜或天文学应用的专用摄像机，CCD 传感器采用敷设在薄硅片上组成矩形网格的电荷收集晶格，来记录到达每个晶格的光能总量的某种度量(见图 1.17)。每个晶格是通过在硅片上生长一层二氧化硅，然后在二氧化硅上渗入导电门结构的方式组成的。当光子击中硅片时，电子-空穴对就产生了(光电转换)，而电子则被加载有正电压的门所形成的势能阱捕获。每个晶格将在一个固定时间间隔 T 内产生的电荷收集起来。

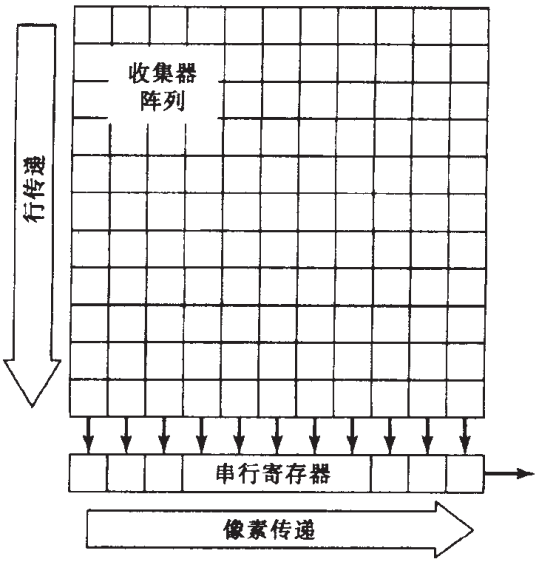


图 1.17 CCD 装置

此时，存储在每个晶格内的电荷使用电荷耦合方式往外传递，通过控制门电势的方法，每个晶体存储的电荷成组地从一个晶格传输至另一晶格，组与组之间保持一定间隔。图像从 CCD 按一次一行方式读出，每行平行地往一个串行输出寄存器传输，每次传一列的一个元素。在读出两行之间，寄存器将电荷逐个传递到一个输出放大器中，放大器随之产生一个与它所收到的电荷量成正比的信号。这个过程直到整幅图都读出才结束。对视频应用场合，图像读出过程每秒重复 30 次(电视帧率)，而在天文学等低亮度应用中，这个过程较慢，以便留出成倍的时间(秒、分甚至小时)收集电荷。要提醒的是，大部分 CCD 摄像机的数字输出在内部先转换成模拟电视信号，再送到图像帧采集卡中，才构成最后的数字图像。

日常生活使用的 CCD 摄像机实际上与黑白摄像机使用相同的芯片，所不同的是让传感器相继的行或列，分别敏感红、绿或蓝色光。经常使用的方法是使用过滤镀膜阻挡它们的补色。

也可采用其他滤波方式，如用 2×2 组成的块的拼镶嵌结构，每块用 2 个绿光、1 个红光与 1 个蓝光接受器(Bayer 格式)。单个 CCD 摄像机的分辨率自然是有限的，而较高质量的摄像机使用光束分裂器，将图像传送到三个使用不同彩色滤波器的 CCD 上。随后每一个颜色通道或分别数字化(RGB 输出)，或结合成复合彩色视觉信号(在美国用 NTSC 制，欧洲用 SECAM，日本用 PAL 制)，或组合成分量视频格式，将彩色与亮度信息分开。

1.4.2 传感器模型[①]

为简单起见，这一节我们只讨论黑白 CCD 摄像机，彩色摄像机可以以相似的方式对待，只是将每个颜色通道分开考虑，并且将相应的滤波器响应考虑进去。

CCD 矩阵第 r 行 c 列的元件上，所存储的电子数目 I 可以表示成

$$I(r,c)=T\int_{\lambda}\int_{p\in S(r,c)}E(p,\lambda)R(p)q(\lambda)dp\,d\lambda$$

其中，T 是电子采集时间，积分是对该元件对应空间域 $S(r,c)$ 与 CCD 有响应的波长范围计算。在该积分式中，E 是点 p 上单位面积与单位波长的功率(也就是辐照度，见第 4 章的正式定义)，R 是晶格的空间响应，q 是设备的量化效率(quantum efficiency，也就是每个单位的入射光能量所产生的电子数目)。一般说来，E 与 q 取决于光的波长 λ，E 和 R 与点 p 在 $S(r,c)$ 中的位置有关。

CCD 的输出放大器将每个晶格收集到的电荷转换成可度量的电压。在大多数摄像机中，这个电压又被摄像机电子线路转换成低通滤波的视频信号，其幅度与 I 成正比。该模拟图像可以再用帧采集器转换成数字图像，通过对视频信号进行空间采样，并对每个图像采样点或像素(该词从图片元素得到)上的亮度值加以量化。

有一些物理现象使上述理想摄像机模型有些变化。例如当太强的光源照射到电荷收集晶格时，就会使存在该晶格的电荷溢出到相邻的晶格中，这可以通过控制照明来避免，但是其他一些因素，诸如制作缺陷、热与量子效应以及量化引起的噪声等在成像过程中是固有的。从下面的分析中可以看到，这些因素可用简单的统计模型来描述。

量子物理学效应在每个晶格的光电转换过程中引入固有的不确定性(散粒噪声)。更精确地说，光电转换过程中所产生的电子数目可用一个服从泊松分布的随机整数变量 $N_I(r,c)$ 来描述，其均值为 $\beta(r,c)\ I(r,c)$，其中 $\beta(r,c)$ 是一个在 0 与 1 之间的数字，用来反映图像中空间响应与数量效率变化，包括缺损像素的影响。由热能引起从硅中释放的电子附加到每个集电晶格的电荷中。它们的作用称为暗电流，可以用一个随机整数变量 $N_{DC}(r,c)$ 来描述，其均值 $\mu_{DC}(r,c)$ 随着温度的增加而增加。暗电流的影响可以通过让摄像机保持冷却来控制。在 CCD 电子学中引入附加电子的概念(偏移)，它们的数目也可用泊松分布随机变量 $N_B(r,c)$ 表示，均值为 $\mu_B(r,c)$。输出放大器增加了一项读出噪声，它可用一个服从高斯分布、以均值 μ_R 与标准差 σ_R 表示的实数随机变量 R 来建模。

还有一些不确定性的其他源(例如电荷转移效率)，但经常可以忽略不计。最后还有一个由帧采集器对模拟电压进行离散化引入的几何效应(行抖动)与量化噪声。行抖动可以通过定

① 粗略地讲是指在空间上或时间上取平均值，随后的章节有更详细的讨论。

标来校正，量化噪声可以建模成均值为零的随机变量 $Q(r,c)$，它均匀分布在量化间隔为 δ 的 $[-\frac{1}{2}\delta,\frac{1}{2}\delta]$ 间隔内，方差为 $\frac{1}{12}\delta^2$。综合起来的数字信号 $D(r,c)$ 复合模型可以表示成

$$D(r,c)=\gamma(N_I(r,c)+N_{DC}(r,c)+N_B(r,c)+R(r,c))+Q(r,c)$$

其中，γ 是放大器和摄像机电路的综合放大倍数。这个模型的统计特性可借助摄像机的辐射度定标来估计，例如，暗电流可通过在暗环境（$I=0$）中取一定数量的采样图像进行估计。

1.5 注释

由 Hecht (1987)撰写的经典教科书是几何光学很好的入门书，它不仅对近轴光学有详细的讨论，同时对本章中简述的各种像差都有详细的讨论（也可参考 Driscoll 与 Vaughan 于 1978 年撰写的图书）。Horn(1986)和 Russ(1995)讨论了图像模糊。Wandell(1995)对人类视觉系统的成像过程进行了出色的叙述。Driscoll 和 Vaughan(1978)对人眼的 Helmoltz 图示模型进行了详细讨论。

CCD 设备在 Boyle 和 Smith(1970)以及 Amelio 等(1970)的著作中进行了介绍。CCD 摄像机在显微镜和天文学中的科学应用在 Aiken 等(1989)，Janesick 等(1987)，Snyder 等(1993)以及 Tyson(1990)中有详细讨论。这一章叙述的传感器统计模型是基于 Snyder 等(1993)的，而对量化噪声的讨论取自 Healey 与 Kondepudy(1994)。这两篇文章还涉及将传感器模型用在天文学中的图像复原，以及机器视觉中的摄像机辐射度学标定。

本章在对提到的一些概念的基本重要性进行讨论的基础上，所推导的主要式子都列在表 1.1 中，以供参考。

表 1.1 参考表：摄像机模型

透视投影	$\begin{cases} x'=f'\frac{x}{z} \\ y'=f'\frac{y}{z} \end{cases}$	x,y：世界坐标（$z<0$） x',y'：图像坐标 f'：针孔到视网膜的距离
弱透视投	$\begin{cases} x'=-mx \\ y'=-my \\ m=-\frac{f'}{z_0} \end{cases}$	x,y：世界坐标 x',y'：图像坐标 f'：针孔到视网膜的距离 z_0：参考点深度（<0） m：放大倍数（>0）
正交投影	$\begin{cases} x'=x \\ y'=y \end{cases}$	x,y：世界坐标 x',y'：图像坐标
Snell 规则	$n_1\sin\alpha_1=n_2\sin\alpha_2$	n_1,n_2：折射系数 α_1,α_2：射线与法线间的夹角
近轴折射	$\frac{n_1}{d_1}+\frac{n_2}{d_2}=\frac{n_2-n_1}{R}$	n_1,n_2：折射系数 d_1,d_2：点到界面的距离 R：界面半径
薄透镜方程	$\frac{1}{z'}-\frac{1}{z}=\frac{1}{f}$	z：物点深度（<0） z'：像点深度（>0） f：焦距

习题

1.1 试推导处于针孔前面 f' 处的虚拟图像的透视方程式投影。

1.2 试从几何上证明，某个平面 Π 中两条平行线的投影会聚到一条水平线 H 上，该水平线是图像平面与过针孔点平行于 Π 的平面的交线。

1.3 用透视投影式(1.1)从代数上证明与上题相同的内容。为了简单起见，可以假设该平面 Π 与图像平面平行。

1.4 试用 Snell 规则说明过薄透镜中光心的射线没有折射现象，并推导薄透镜方程。

提示：考虑一条过点 P 的射线 r_0，并分别构造透镜的右轮廓和左轮廓对 r_0 折射而得到两条射线 r_1 与 r_2。

1.5 考虑一个用薄透镜配备的摄像机，图像平面在 z' 位置，而平面上的景物点聚焦在 z 处。现假设图像平面移动至 $\hat{z}'$，证明相应的模糊圆的直径为

$$d\frac{|z'-\hat{z}'|}{z'}$$

其中，d 是透镜的直径。使用以上结果来说明视场深度(也就是使模糊圆的直径低于某个阈值 ε 的最近与最远平面之间的距离)可按下式计算

$$D=2\varepsilon fz(z+f)\frac{d}{f^2d^2-\varepsilon^2z^2}$$

并且做出结论，即对一个固定的焦距长度，视场深度随透镜直径减小而增加，f 数也因而增加。

提示：解出图像聚焦在图像平面上 $\hat{z}'$ 位置的点的深度 $\hat{z}$；要考虑 z' 比 $\hat{z}'$ 大与小两种情况。

1.6 在一个薄透镜的两个焦点分别为 F 与 F' 的条件下，用几何方法构造点 P 的图像 P'。

1.7 推出厚透镜两个球面边半径相同的厚透镜方程。

第 2 章　摄像机的几何模型

以摄像机为中心的基本透视投影原理已经在第 1 章中介绍。在建立图像坐标与世界坐标之间的关系时要用到的一些必要公式的解析表示，将在这一章中介绍。首先回忆一下欧几里得解析几何的基本公式，然后将引入一些参数（内参数和外参数）表示图像坐标系与世界坐标系之间的联系，并推出通用的透视投影公式。最后还将推出仿射模型的简洁表达，仿射模型是远距离物体针孔成像的一种近似。除此之外，还将进一步介绍第 1 章讨论过的正投影和弱透视投影。

2.1　欧几里得解析几何基础

在阅读本章前，读者应该对欧氏几何及线性代数比较熟悉。这一节将介绍一些必要的解析几何概念，如坐标系、齐次坐标、旋转矩阵等。

2.1.1　坐标系和齐次坐标

第 1 章中我们已经用过三维坐标系，现在要给出它的正式定义。坐标系要确定度量单位，这里我们使用的单位长度，可以是米或英寸。

在三维欧氏空间$\mathbb{E}^3$ 中选取一个点 O 以及三个相互垂直的单位向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$，这个四元组 $(\boldsymbol{O}, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$就定义了一个正交坐标系 $\boldsymbol{F}$。$\boldsymbol{O}$ 称为坐标系$\boldsymbol{F}$ 的原点，$\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 称为基向量。右手坐标系是指，$\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 分别对应右手的手指，拇指向上，食指向前，中指向左，如图 2.1 所示①。

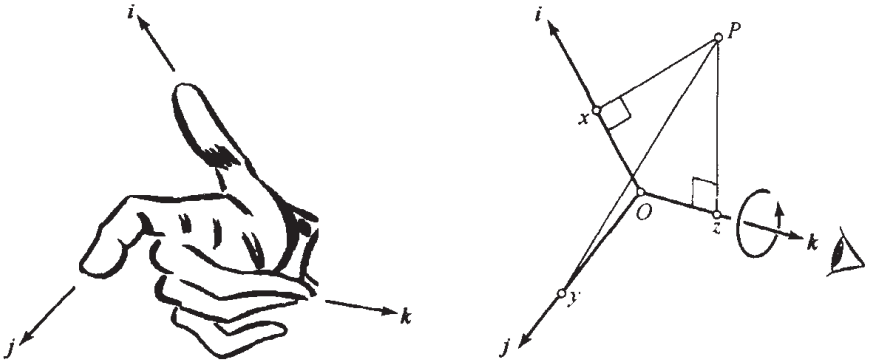


图 2.1　一个右手坐标系以及点 P 的 x, y, z 坐标

点 P 的 x, y, z 坐标定义为向量$\overrightarrow{OP}$在向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 上的正交投影（见图 2.1 右则），有

① 这是传统的右手系定义。本书的一位左撇子作者总觉得很别扭，他更喜欢如下的定义：当对着 $\boldsymbol{k}$ 轴的方向观察 $(\boldsymbol{i}, \boldsymbol{j})$平面时，$\boldsymbol{i}$ 轴相对于$\boldsymbol{j}$ 轴逆时针转了 90°（见图 2.1）。左手系相当于顺时针旋转。习惯左手或右手的读者可能觉得这种定义也很有用。

$$\begin{cases} x = \overrightarrow{OP} \cdot \boldsymbol{i} \\ y = \overrightarrow{OP} \cdot \boldsymbol{j} \\ z = \overrightarrow{OP} \cdot \boldsymbol{k} \end{cases} \Longleftrightarrow \overrightarrow{OP} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$$

列向量

$$\boldsymbol{P} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

称为坐标系 F 中点 P 的坐标向量。任意一个自由向量的坐标向量也可以按这种方法定义，即在向量 $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ 上的正交投影，显然这个向量与原点 O 的选取无关。假设有一平面 Π，Π 上任一点 A，以及一个垂直于该平面的单位向量 $\boldsymbol{n}$。Π 上的点都满足

$$\overrightarrow{AP} \cdot \boldsymbol{n} = 0$$

坐标系 F 中点 P 的坐标为 x, y, z，$\boldsymbol{n}$ 的方向为 a, b, c，上式可写成 $\overrightarrow{OP} \cdot \boldsymbol{n} - \overrightarrow{OA} \cdot \boldsymbol{n} = 0$，或

$$ax + by + cz - d = 0 \tag{2.1}$$

其中，$d \stackrel{\text{def}}{=} \overrightarrow{OA} \cdot \boldsymbol{n}$ 与 Π 上 A 点的选取无关，只是 O 到平面 Π 的距离（见图 2.2）。

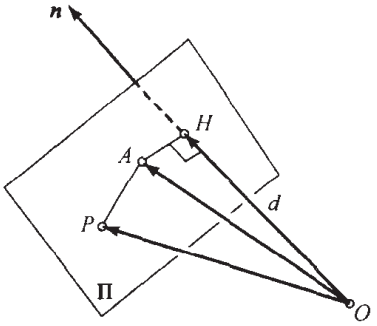


图 2.2 平面方程的几何定义。原点和平面间的距离 d 等于原点到点 H 的距离

有时，也可以用齐次坐标表示点、向量和平面。齐次坐标的定义将在第 12 章、第 13 章介绍仿射和投影几何时正式给出，目前只是对式(2.1)稍加修改，

$$(a, b, c, -d)\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = 0$$

更简洁的表示是

$$\boldsymbol{\Pi} \cdot \boldsymbol{P} = 0, \quad \text{其中,} \quad \boldsymbol{\Pi} \stackrel{\text{def}}{=} \begin{pmatrix} a \\ b \\ c \\ -d \end{pmatrix}, \quad \boldsymbol{P} \stackrel{\text{def}}{=} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2.2}$$

向量 $\boldsymbol{P}$ 为坐标系(F)中点 P 的齐次坐标，只要把原始的三维坐标添加一个 1 即可。类似地，向量 $\boldsymbol{\Pi}$ 是坐标系(F)中平面 $\boldsymbol{\Pi}$ 的齐次坐标，方程(2.2)称为该坐标系中 Π 的平面方程。记住，$\boldsymbol{\Pi}$ 定义的只是各分量的比例，若将其各个分量都放大相同的倍数，对方程的解没有影响。不论是表示点还是平面，我们都约定齐次坐标定义成比例关系。在第 13 章将详细地证明这个

结论。要从齐次坐标回到普通非齐次坐标，只需同除以第 4 项即可。

还要注意，虽然我们目前关心的都是三维欧氏几何，但以上讨论的概念可以应用到任意平面几何上：坐标系由原点和右手系正交基定义$(\boldsymbol{i},\boldsymbol{j})$；点 p 在这个坐标系下的定义为 $x=\overrightarrow{op}\cdot\boldsymbol{i}$ 和 $y=\overrightarrow{op}\cdot\boldsymbol{j}$，同样也可以定义齐次坐标；特别地，直线 δ 的方程可写成

$$ax+by-d=0 \Longleftrightarrow \boldsymbol{\delta}\cdot\boldsymbol{p}=0,\quad 其中,\quad \boldsymbol{\delta}=\begin{pmatrix}a\\b\\-d\end{pmatrix},\quad \boldsymbol{p}=\begin{pmatrix}x\\y\\1\end{pmatrix}$$

a、b、d 依次表示直线 δ 的单位法向量和 o 到直线 δ 的有向距离。

重新回到三维几何空间。齐次坐标不仅可以表示平面和直线，还能表达更复杂的形状①。例如一个以原点为圆心、半径为 R 的球面 S，其面上点的坐标要满足的充要条件是

$$x^2+y^2+z^2=R^2$$

这等价于

$$(x,y,z,1)\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&-R^2\end{pmatrix}\begin{pmatrix}x\\y\\z\\1\end{pmatrix}=0$$

二次曲面更通用的表示是

$$a_{200}x^2+a_{110}xy+a_{020}y^2+a_{011}yz+a_{002}z^2+a_{101}xz+a_{100}x+a_{010}y+a_{001}z+a_{000}=0$$

显然它等价于下面的条件

$$\boldsymbol{P}^T\mathcal{Q}\boldsymbol{P}=0,\quad 其中,\quad \mathcal{Q}=\begin{pmatrix}a_{200}&\frac{1}{2}a_{110}&\frac{1}{2}a_{101}&\frac{1}{2}a_{100}\\\frac{1}{2}a_{110}&a_{020}&\frac{1}{2}a_{011}&\frac{1}{2}a_{010}\\\frac{1}{2}a_{101}&\frac{1}{2}a_{011}&a_{002}&\frac{1}{2}a_{001}\\\frac{1}{2}a_{100}&\frac{1}{2}a_{010}&\frac{1}{2}a_{001}&a_{000}\end{pmatrix}\tag{2.3}$$

在这个方程中，$\boldsymbol{P}$ 是点 P 的齐次坐标。$\mathcal{Q}$是一个 4×4 的对称阵，它也是在比例意义下定义的。

2.1.2　坐标系变换和刚体运动

要同时考虑多个坐标系时，我们沿用 Craig(1989)的符号，坐标系 F 中点 P(向量 $\boldsymbol{v}$)的坐标向量记为${}^FP({}^F\boldsymbol{v})$，即

$${}^FP={}^F\overrightarrow{OP}=\begin{pmatrix}x\\y\\z\end{pmatrix}\Longleftrightarrow \overrightarrow{OP}=x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}$$

点、向量、矩阵的这些上标、下标最初看起来比较别扭，但在本章的后面会体现出它们的优越性。我们考虑两个坐标系的情况：$(A)=(O_A,\boldsymbol{i}_A,\boldsymbol{j}_A,\boldsymbol{k}_A)$和$(B)=(O_B,\boldsymbol{i}_B,\boldsymbol{j}_B,\boldsymbol{k}_B)$。下面将介绍如何把${}^BP$ 表示成${}^AP$ 的函数。首先考虑一种简单情况(即，$\boldsymbol{i}_A=\boldsymbol{i}_B$，$\boldsymbol{j}_A=\boldsymbol{j}_B$ 且 $\boldsymbol{k}_A=\boldsymbol{k}_B$)，两个坐标系的基向量平行 O_A 和 O_B，但是原点位置不同(见图 2.3)。

---

① 有的读者会考虑$\mathbb{E}^3$ 中的直线。一条直线可以完全定义为两个平面的交线。$\mathbb{E}^3$ 中的直线还可以用 Plücker 坐标表示成更通用的方式，请参见习题。

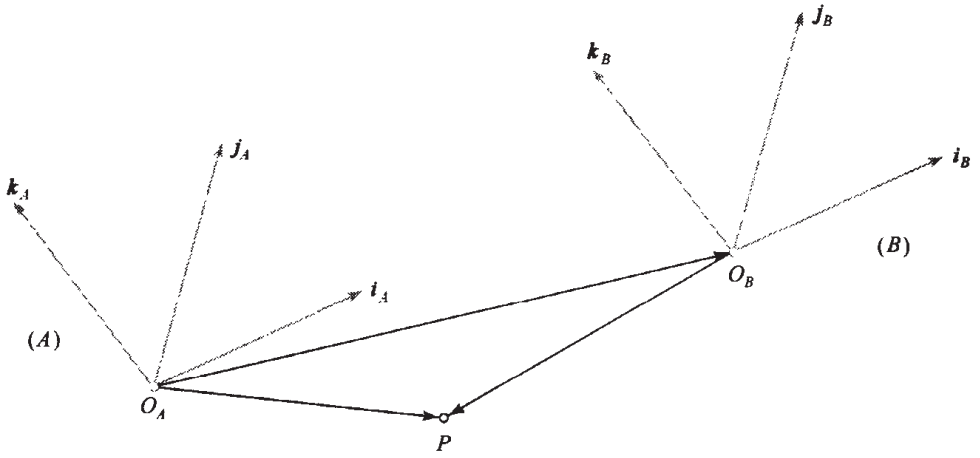


图 2.3 坐标系变换:纯平移

在这种情况下,我们称两个坐标系之间是纯平移关系,且有$\overrightarrow{O_BP} = \overrightarrow{O_BO_A} + \overrightarrow{O_AP}$,则

$$^{B}P = {}^{A}P + {}^{B}O_A$$

若两个坐标系的原点重合(即 $O_A = O_B = O$)但是基向量不同,我们称它们之间是纯旋转关系(见图 2.4)。旋转矩阵是个 3×3 的数组,定义为$^{B}_{A}\mathcal{R}$

$$^{B}_{A}\mathcal{R} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{i}_A\cdot\boldsymbol{i}_B & \boldsymbol{j}_A\cdot\boldsymbol{i}_B & \boldsymbol{k}_A\cdot\boldsymbol{i}_B \\ \boldsymbol{i}_A\cdot\boldsymbol{j}_B & \boldsymbol{j}_A\cdot\boldsymbol{j}_B & \boldsymbol{k}_A\cdot\boldsymbol{j}_B \\ \boldsymbol{i}_A\cdot\boldsymbol{k}_B & \boldsymbol{j}_A\cdot\boldsymbol{k}_B & \boldsymbol{k}_A\cdot\boldsymbol{k}_B \end{pmatrix}$$

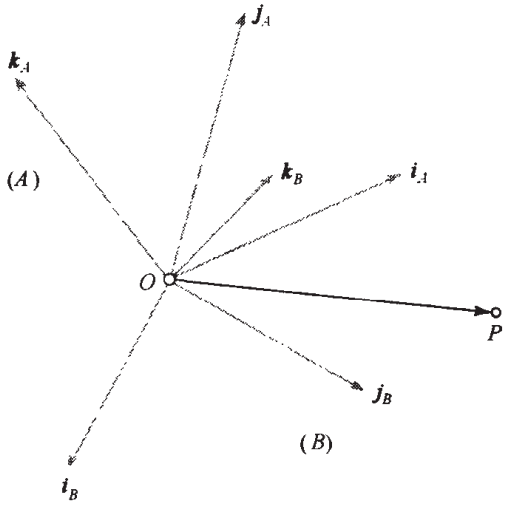


图 2.4 坐标系变换:纯旋转

注意$^{B}_{A}\mathcal{R}$的第一列是 $\boldsymbol{i}_A$ 在基($\boldsymbol{i}_B,\boldsymbol{j}_B,\boldsymbol{k}_B$)下的坐标。同样,这个矩阵的第三列是 $\boldsymbol{k}_B$ 在基($\boldsymbol{i}_A,\boldsymbol{j}_A,\boldsymbol{k}_A$)下的坐标。更简洁的表示方法是:

$$^{B}_{A}\mathcal{R} = \begin{pmatrix} {}^{B}\boldsymbol{i}_A & {}^{B}\boldsymbol{j}_A & {}^{B}\boldsymbol{k}_A \end{pmatrix} = \begin{pmatrix} {}^{A}\boldsymbol{i}_B{}^{T} \\ {}^{A}\boldsymbol{j}_B{}^{T} \\ {}^{A}\boldsymbol{k}_B{}^{T} \end{pmatrix}$$

其满足$^{A}_{B}\mathcal{R} = {}^{B}_{A}\mathcal{R}^{T}$。

前面说过,这些上标和下标最初看起来有些混乱。为了清楚起见,必须记住,在坐标变换

中，下标指示被表示的对象，而上标指示用来表示对象的坐标系。例如，${}^{A}P$ 是坐标系 A 下点 P 的表示，${}^{B}\boldsymbol{j}_A$ 是坐标系 B 下向量 $\boldsymbol{j}_A$ 的坐标，${}^{B}_{A}\mathcal{R}$ 是坐标系 A 在坐标系 B 中的旋转矩阵。

给出一个纯旋转的例子：假如 $\boldsymbol{k}_A = \boldsymbol{k}_B = \boldsymbol{k}$，用 θ 表示 $\boldsymbol{i}_A$ 和 $\boldsymbol{i}_B$ 间的夹角，当 $\boldsymbol{i}_A$ 绕 $\boldsymbol{k}$ 旋转 θ 后到达 $\boldsymbol{i}_B$（见图 2.5）。我们有

$$
{}^{B}_{A}\mathcal{R} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.4}
$$

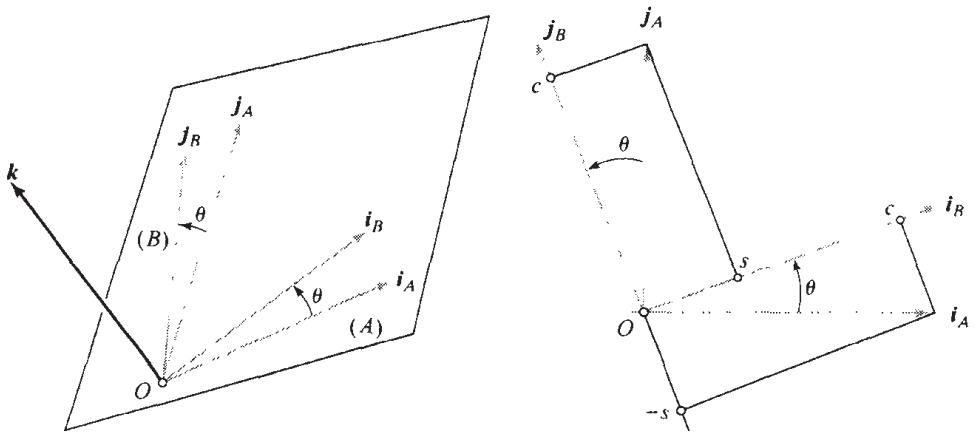


图 2.5　两个坐标系之间的夹角是 θ（相对于公共的基向量 $\boldsymbol{k}$）。如右图所示，$\boldsymbol{i}_A = c\boldsymbol{i}_B - s\boldsymbol{j}_B$，$\boldsymbol{j}_A = s\boldsymbol{i}_B + c\boldsymbol{j}_B$，其中 $c = \cos\theta$，$s = \sin\theta$

两个坐标系之间的旋转也可以通过逐步绕 $\boldsymbol{i}_A$ 或 $\boldsymbol{j}_A$ 旋转得到，这样可以写出类似的方程（见习题）。一般来说，旋转矩阵可以分解为绕 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 旋转的基本旋转矩阵的乘积。

我们来分析一下由任意旋转矩阵确定的坐标变换的特点。写出

$$
\overrightarrow{OP} = \begin{pmatrix} \boldsymbol{i}_A & \boldsymbol{j}_A & \boldsymbol{k}_A \end{pmatrix} \begin{pmatrix} {}^{A}x \\ {}^{A}y \\ {}^{A}z \end{pmatrix} = \begin{pmatrix} \boldsymbol{i}_B & \boldsymbol{j}_B & \boldsymbol{k}_B \end{pmatrix} \begin{pmatrix} {}^{B}x \\ {}^{B}y \\ {}^{B}z \end{pmatrix}
$$

由 ${}^{B}_{B}R$ 是单位阵可以知道，在坐标系 B 中满足

$$
{}^{B}P = {}^{B}_{A}\mathcal{R}\,{}^{A}P
$$

注意，右部第一项的下标和第二项的上标是相同的。这个性质对一般的坐标变换也是适用的，可以在应用中建立两个坐标系的对应点方程。

旋转矩阵有下面的特点（见习题）：(1)旋转矩阵的逆矩阵和它的转置相同；(2)旋转矩阵的模是 1；由定义可知，一个旋转矩阵的各列构成了右手正交坐标系。由于满足上述两个特点，一个旋转矩阵的各行也构成了同样的坐标系。

另外还要说明，所有旋转矩阵的集合，在矩阵乘法的作用下，构成了一个群。也就是说，满足：(a)两个旋转矩阵的乘积仍是旋转矩阵；(b)旋转矩阵的乘法满足结合率——对任意旋转矩阵 $\mathcal{R}$，$\mathcal{R}'$ 和 $\mathcal{R}''$，满足 $(\mathcal{R}\mathcal{R}')\mathcal{R}'' = \mathcal{R}(\mathcal{R}'\mathcal{R}'')$；(c)存在单位元，单位元是一个 3×3 的特殊旋转矩阵 Id，满足 $\mathcal{R}\,\mathrm{Id} = \mathrm{Id}\,\mathcal{R} = \mathcal{R}$；(d)任意旋转矩阵 $\mathcal{R}$，$\mathcal{R}^{-1} = \mathcal{R}^{T}$，存在逆元 $\mathcal{R}^{-1}$ 满足 $\mathcal{R}\mathcal{R}^{-1} = \mathcal{R}^{-1}\mathcal{R} = \mathrm{Id}$，但是这个群不是交换群（即给定两个矩阵 $\mathcal{R}$ 和 $\mathcal{R}'$，乘积 $\mathcal{R}\mathcal{R}'$ 和 $\mathcal{R}'\mathcal{R}$ 一般是不等的）。

若两个坐标系的原点和基向量都是不同的，我们称这两个坐标系之间是一般的刚体变换

(见图 2.6),且有

$$^{B}P = {}^{B}_{A}\mathcal{R}\,^{A}P + {}^{B}O_A \tag{2.5}$$

其中,${}^{B}_{A}\mathcal{R}$和${}^{B}O_A$前面已经定义过了。类似的方法也可以定义平面的齐次坐标和二次曲面的对应矩阵各自的变换关系(见习题)。

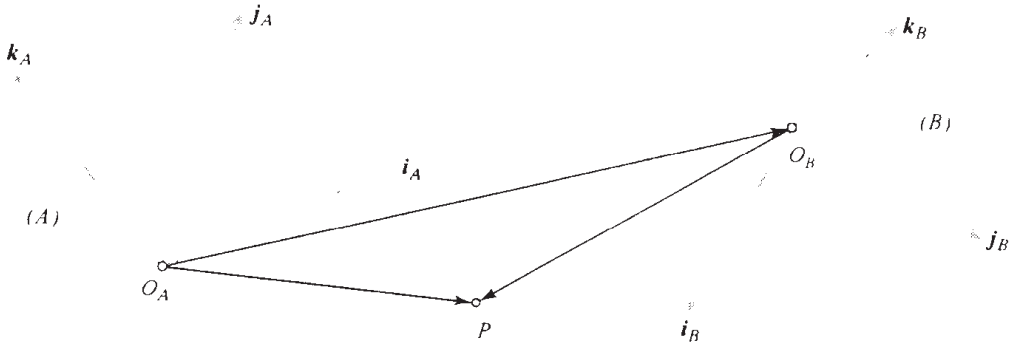


图 2.6 两个坐标系之间的变换:一般的刚体变换

在齐次坐标情况下,方程(2.5)可以写成矩阵乘积的形式:需要用到矩阵的分块乘法,即,如果

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix} \quad 且 \quad \mathcal{B} = \begin{pmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} \\ \mathcal{B}_{21} & \mathcal{B}_{22} \end{pmatrix} \tag{2.6}$$

其中,子阵$\mathcal{A}_{11}$与$\mathcal{A}_{21}$的列数和$\mathcal{B}_{11}$与$\mathcal{B}_{12}$的行数相同($\mathcal{A}_{12}$与$\mathcal{A}_{22}$的列数和$\mathcal{B}_{21}$与$\mathcal{B}_{22}$的行数相同),则有

$$\mathcal{AB} = \begin{pmatrix} \mathcal{A}_{11}\mathcal{B}_{11} + \mathcal{A}_{12}\mathcal{B}_{21} & \mathcal{A}_{11}\mathcal{B}_{12} + \mathcal{A}_{12}\mathcal{B}_{22} \\ \mathcal{A}_{21}\mathcal{B}_{11} + \mathcal{A}_{22}\mathcal{B}_{21} & \mathcal{A}_{21}\mathcal{B}_{12} + \mathcal{A}_{22}\mathcal{B}_{22} \end{pmatrix}$$

我们可以把式(2.5)写成式(2.6)的形式,

$$\begin{pmatrix} ^{B}P \\ 1 \end{pmatrix} = {}^{B}_{A}\mathcal{T}\begin{pmatrix} ^{A}P \\ 1 \end{pmatrix}, \quad 其中, \quad {}^{B}_{A}\mathcal{T} \stackrel{\text{def}}{=} \begin{pmatrix} ^{B}_{A}\mathcal{R} & ^{B}O_A \\ \mathbf{0}^T & 1 \end{pmatrix} \tag{2.7}$$

$\mathbf{0} = (0,0,0)^T$。这样,我们就可以用一个 4×4 矩阵和一个四维向量表示任意的坐标系变换。在矩阵乘法意义下,由式(2.7)定义的刚体变换集合也是一个群。

刚体变换把一个坐标系映射到另一个坐标系。对于给定的坐标系 F,也可以认为是 F 上的点到点的一个映射关系——P 映射到 P',

$$^{F}P' = \mathcal{R}\,^{F}P + \boldsymbol{t} \Longleftrightarrow \begin{pmatrix} ^{F}P' \\ 1 \end{pmatrix} = \begin{pmatrix} \mathcal{R} & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{pmatrix}\begin{pmatrix} ^{F}P \\ 1 \end{pmatrix} \tag{2.8}$$

其中,$\mathcal{R}$是一个旋转矩阵,$\boldsymbol{t}$ 是$\mathbb{R}^3$ 中的一个向量(见图 2.7)。将刚体变换集看做是在$\mathbb{E}^3$ 空间映射到自身,以及一组将刚体变换进行组合的规则,再次表明它们组成一个群。显然刚体变换是保长度和保夹角的。但是,这个刚体变换对应的 4×4 矩阵会随着坐标系 F 的不同选取而有所不同。

例如,考虑坐标系 F 绕 $\boldsymbol{k}$ 轴旋转 θ 角。在习题中会看到,这个映射可以写成

$$^{F}P' = \mathcal{R}^{F}P, \quad 其中, \quad \mathcal{R} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

如果 F'是 F 旋转以后得到的坐标系,则由方程(2.4),我们有${}^{F'}P = {}^{F'}_{F}\mathcal{R}\,^{F}P$ 和$\mathcal{R} = {}^{F'}_{F}\mathcal{R}^{-1}$。

更通用的结论是，描述两个坐标系之间变化的矩阵是把第一个坐标系映射到第二个坐标系的矩阵和它的逆。

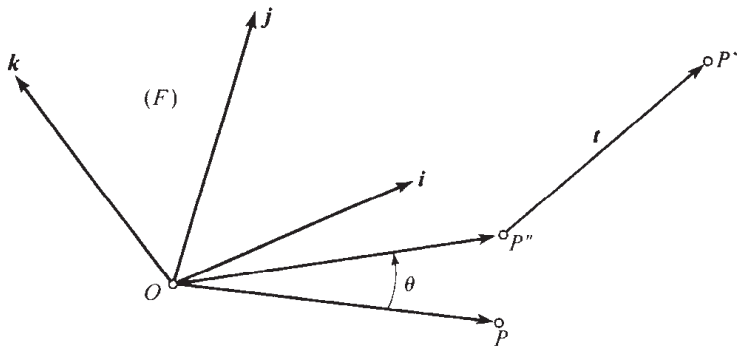


图 2.7　旋转变换$\mathcal{R}$把 P 映射到 P''，然后平移 $\boldsymbol{t}$ 把 P''映射到 P'。
在本图的例子中，$\mathcal{R}$表示绕坐标系 F 中的 $\boldsymbol{k}$ 轴转 θ 角

若$\mathcal{R}$被任意一个非奇异 3×3 矩阵$\mathcal{A}$替代，情况会怎样呢？这时方程(2.8)仍然表示点到点的映射(和坐标系间的坐标变换)，但是长度和角度不再是不变的(新的坐标系不一定能保证坐标轴长度为 1 且正交)。我们称这个 4×4 矩阵

$$\mathcal{T}=\begin{pmatrix}\mathcal{A} & \boldsymbol{t}\\ \mathbf{0}^T & 1\end{pmatrix}$$

表示一个仿射变换。若$\mathcal{T}$是非奇异的任意 4×4 矩阵，我们称其为投影变换。仿射变换和投影变换都构成群。第 12 章和第 13 章将给出更详细的说明。

2.2　摄像机参数和透视投影

在第 1 章中已经说过，对于针孔摄像机，物点 P 的世界坐标 x,y,z 与它在成像平面上的坐标 x',y' 满足透视方程(1.1)。实际上，这个方程成立是有条件的，所有距离必须在摄像机坐标系下衡量，而且图像坐标的原点要和摄像机的光心(摄像机的对称轴穿过成像平面的点)重合。在实际应用中，用一些物理参数表示摄像机坐标系与世界坐标系的关系，包括焦距、像素大小、主点的位置、摄像机的位置和方向。

本节将逐个说明这些参数。我们要区分摄像机内部参数与外部参数，前者是摄像机坐标系与前面第 1 章介绍的理想坐标系之间的关系，而后者表示摄像机在世界坐标系里的位置和方向。

对于透镜成像，只有当摄像机的焦距、物体深度和成像平面到光心的距离三者满足薄透镜方程(1.6)时，才能呈现清楚的像。本章将忽略这一点，而假设总能呈现清楚的像，而且使用方程(1.1)的时候也不考虑透镜引起的非线性畸变。这些畸变在本章中都不考虑，但是在第 3 章估计摄像机内外参数的方法(称为摄像机几何标定)时，会重新引入径向畸变。

2.2.1　内参数

我们可以为摄像机建立一个归一化图像平面，这个平面平行于摄像机的物理成像平面，且到针孔的距离为单位长度。接着在这个平面上建立一个坐标系，原点定在光轴和这个平面的

交点处，即图 2.8 中的 $\hat{C}$ 点。则透视投影方程(1.1)可以写成如下形式：

$$\begin{cases} \hat{u} = \dfrac{x}{z} \\ \hat{v} = \dfrac{y}{z} \end{cases} \Longleftrightarrow \hat{\boldsymbol{p}} = \frac{1}{z}(\mathrm{Id} \quad \boldsymbol{0})\begin{pmatrix} \boldsymbol{p} \\ 1 \end{pmatrix} \tag{2.9}$$

其中，$\hat{\boldsymbol{p}} \stackrel{\text{def}}{=} (\hat{u}, \hat{v}, 1)^T$ 是点 P 投影到这个平面上的点 $\hat{p}$ 的齐次坐标表示。

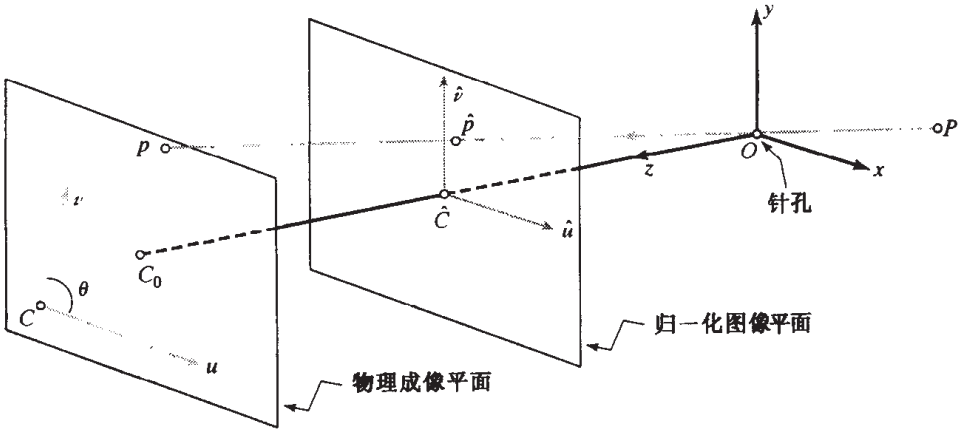


图 2.8 物理成像和归一化成像坐标系

物理成像平面的位置一般与我们定义的归一化平面位置不同(见图 2.8)：它位于距离针孔 $f \neq 1$ 的位置上①，图像坐标(u, v)一般用像素表示(而不是米等单位)。而且像素一般不是正方形，而是个长方形，所以需要用两个额外的比例因子 k 和 l，且

$$\begin{cases} u = kf\dfrac{x}{z} \\ v = lf\dfrac{y}{z} \end{cases} \tag{2.10}$$

首先说明单位长度：假设 f 是用米来表示的距离，像素的大小是$\frac{1}{k} \times \frac{1}{l}$，其中，$k$ 和 l 的单位是像素 $\times \mathrm{m}^{-1}$。参数 k、l 和 f 是相关的，如果用像素单位表示，有 $\alpha = kf$ 和 $\beta = lf$。

一般把图像的一角而不是中心定为摄像机坐标系的原点(例如在图 2.8 中，原点是左下角，则图像上一个像素的坐标就是它所在的行数和列数，有时原点也取左上角)；则 CCD 阵列的中心与光轴穿过的主点 C_0 不同。需要添加两个参数 u_0 和 v_0 来定义 C_0 在摄像机坐标系内的位置。则式(2.10)改为

$$\begin{cases} u = \alpha\dfrac{x}{z} + u_0 \\ v = \beta\dfrac{y}{z} + v_0 \end{cases} \tag{2.11}$$

最后，由于制造误差，摄像机坐标系可能会产生偏歪，即两个坐标轴不完全垂直(但和直角不会相差很大)。在这种情况下，式(2.11)改为

① 从现在起，我们都假设摄像机聚焦在无穷远处，则针孔到图像平面的距离等于焦距。

$$\begin{cases} u = \alpha\dfrac{x}{z} - \alpha\cot\theta\dfrac{y}{z} + u_0 \\ v = \dfrac{\beta}{\sin\theta}\dfrac{y}{z} + v_0 \end{cases} \tag{2.12}$$

通过式(2.9)和式(2.12),我们可以写出物理摄像机坐标系和归一化的图像坐标系之间的关系:

$$\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}, \quad 其中,\ \boldsymbol{p} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad 且 \quad \mathcal{K} \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha\cot\theta & u_0 \\ 0 & \dfrac{\beta}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.13}$$

最终得到

$$\boldsymbol{p} = \frac{1}{z}\mathcal{M}\boldsymbol{P}, \quad 其中,\ \mathcal{M} \stackrel{\text{def}}{=} (\mathcal{K} \quad \boldsymbol{0}) \tag{2.14}$$

表示 $\boldsymbol{P} = (x, y, z, 1)^T$ 在摄像机坐标系里的齐次坐标。换句话说,利用 3×4 矩阵$\mathcal{M}$,齐次坐标可以表示从四维到三维的投影变换。

像素大小和偏歪角度对每个摄像机是固定的,而且可以在制造过程中测量得到(这些信息不一定可以得到,例如对已经录制好的录像,或者不知道采集卡的采样精度)。对变焦镜头,焦距随时可能改变,如果光轴不完全垂直于图像平面的话,图像中心也可能变化。摄像机焦距的改变还会影响放大倍数,因为透镜到感光面的距离改变了。但是我们已经假设摄像机聚焦到无限远,因此在本章中将忽略这些影响。

2.2.2　外参数

摄像机坐标系 C 在世界坐标系 W 中的位置记为

$$\begin{pmatrix} {}^C P \\ 1 \end{pmatrix} = \begin{pmatrix} {}^C_W\mathcal{R} & {}^C O_W \\ \boldsymbol{0}^T & 1 \end{pmatrix} \begin{pmatrix} {}^W P \\ 1 \end{pmatrix}$$

代入式(2.14)得到

$$\boldsymbol{p} = \frac{1}{z}\mathcal{M}\boldsymbol{P}, \quad 其中,\ \mathcal{M} = \mathcal{K}(\mathcal{R} \quad \boldsymbol{t}) \tag{2.15}$$

$\mathcal{R} = {}^C_W\mathcal{R}$是旋转矩阵,$\boldsymbol{t} = {}^C O_W$ 是平移向量,$\boldsymbol{P} = ({}^W x, {}^W y, {}^W z, 1)^T$ 表示向量 P 在坐标系 W 下的齐次坐标。

这是一般形式的透视投影方程,可以利用它得到摄像机光心 O 在世界坐标系中的位置。事实上,在习题中会证明,光心的齐次坐标 $\boldsymbol{O}$ 满足$\mathcal{M}\,\boldsymbol{O} = 0$(显然,光心是惟一一个图像不惟一确定的点)。若$\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$,其中$\mathcal{A}$是非奇异 3×3 矩阵,$\boldsymbol{b}$ 是一个三维向量,则 O 的非齐次坐标为 $-\mathcal{A}^{-1}\boldsymbol{b}$。

式(2.15)中的深度 z 与$\mathcal{M}$及$\mathcal{P}$是相关的。令 $\boldsymbol{m}_1^T$, $\boldsymbol{m}_2^T$ 和 $\boldsymbol{m}_3^T$ 分别表示$\mathcal{M}$的三行,则由式(2.15)可知 $z = \boldsymbol{m}_3 \cdot \boldsymbol{P}$。有时为了方便起见,会把式(2.15)式写成

$$\begin{cases} u = \dfrac{\boldsymbol{m}_1 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}} \\ v = \dfrac{\boldsymbol{m}_2 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}} \end{cases} \tag{2.16}$$

投影矩阵可以由 5 个内参数($\alpha,\beta,u_0,v_0,\theta$)和 6 个外参数(三个表示旋转$\mathcal{R}$的角度,另三个表示平移 $\boldsymbol{t}$)显式表达,

$$\mathcal{M}=\begin{pmatrix}\alpha\boldsymbol{r}_1^T-\alpha\cot\theta\boldsymbol{r}_2^T+u_0\boldsymbol{r}_3^T & \alpha t_x-\alpha\cot\theta t_y+u_0t_z\\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T+v_0\boldsymbol{r}_3^T & \dfrac{\beta}{\sin\theta}t_y+v_0t_z\\ \boldsymbol{r}_3^T & t_z\end{pmatrix}\tag{2.17}$$

其中,$\boldsymbol{r}_1^T,\boldsymbol{r}_2^T$ 和 $\boldsymbol{r}_3^T$ 表示$\mathcal{R}$的三行,t_x,t_y 和 t_z 是向量 $\boldsymbol{t}$ 的坐标。若$\mathcal{R}$写成三个基旋转的积,则 $\boldsymbol{r}_i(i=1,2,3)$可以显式地用三个角表示。

2.2.3 透视投影矩阵的性质

本节将讨论满足何种条件的 3×4 矩阵$\mathcal{M}$可以写成式(2.17)的形式。不失一般性,可设 $\mathcal{M}=(\mathcal{A}\quad\boldsymbol{b})$,其中$\mathcal{A}$是一个 3×3 矩阵,$\boldsymbol{b}$ 是$\mathbb{R}^3$ 的一个元素,我们用 $\boldsymbol{a}_3^T$ 表示$\mathcal{A}$的第三行。显然,如果$\mathcal{M}$是式(2.17)的一个实例,则 $\boldsymbol{a}_3^T$ 必然是一个单位向量,因为它等于 $\boldsymbol{r}_3^T$,是旋转矩阵的第三行。注意,对任意$\lambda\neq0$,用$\lambda\mathcal{M}$代替$\mathcal{M}$不影响图像中的实际坐标。因此我们在本书的剩余章节中把投影矩阵看成齐次对象,即只定义到比例关系这一层。因此我们可以选择适当的比例系数使得$|\boldsymbol{a}_3|=1$。式(2.17)中的 z 只有在$\mathcal{M}$是这种规范形式时才能理解为 P 点的深度。还要注意,内参数和外参数的数目正好对应(齐次)矩阵$\mathcal{M}$的 11 个自由变量。

我们称可以写成式(2.17)形式的 3×4 矩阵为透视投影矩阵,其中的变量为摄像机的内外参数。在实际应用中,一般对内参数加一些限制,因为前面说过,其中的一部分参数是已知的。若式(2.17)中的 $\theta=\pi/2$,则我们称这个 3×4 矩阵为无偏歪透视投影矩阵。若同时满足 $\theta=\pi/2$和 $\alpha=\beta$,则称其为无偏歪单位长宽比透视投影矩阵。可以通过适当的图像坐标系变换,把已知偏歪角度和长宽比的变换转变为无偏歪单位长宽比的变换。那么,是不是任意的 3×4 矩阵都是透视投影矩阵呢?下面的定理将回答这个问题。

定理 1 设$\mathcal{M}=(\mathcal{A}\quad\boldsymbol{b})$是一个 3×4 矩阵,用 $\boldsymbol{a}_i^T(i=1,2,3)$表示由$\mathcal{A}$的最左边 3 列构成的矩阵$\mathcal{M}$的各行。

- $\mathcal{M}$是透视投影矩阵的充要条件是 $\mathrm{Det}(\mathcal{A})\neq0$。
- $\mathcal{M}$是无偏歪透视投影矩阵的充要条件是 $\mathrm{Det}(\mathcal{A})\neq0$ 且

$$(\boldsymbol{a}_1\times\boldsymbol{a}_3)\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)=0$$

- $\mathcal{M}$是无偏歪单位长宽比透视投影矩阵的充要条件是 $\mathrm{Det}(\mathcal{A})\neq0$ 且

$$\begin{cases}(\boldsymbol{a}_1\times\boldsymbol{a}_3)\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)=0\\(\boldsymbol{a}_1\times\boldsymbol{a}_3)\cdot(\boldsymbol{a}_1\times\boldsymbol{a}_3)=(\boldsymbol{a}_2\times\boldsymbol{a}_3)\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)\end{cases}$$

定理中的条件显然是必要的:由式(2.15)有$\mathcal{A}=\mathcal{K}\mathcal{R}$,则$\mathcal{A}$和$\mathcal{K}$的行列式一样且非零。而且,可以验证式(2.17)中的$\mathcal{K}\mathcal{R}$满足不同假设下的条件。Faugeras(1993)和习题中会证明该命题的充分性。

2.3 仿射摄像机和仿射投影方程

当场景的深度大小变化远小于场景距摄像机的距离时,可以用仿射投影近似成像过程。

其中包括了第 1 章介绍的正射投影和弱透视投影，以及本节将要介绍的平行投影和类透视投影。它们将在第 12 章中正式定义。

2.3.1　仿射摄像机

在正投影下，成像过程就是一个简单的垂直投影过程。若景物距离很远，且深度变化不大，正投影是透视投影的合理近似。平行投影包括了正投影，并且考虑了物体不在光轴上的情况。在这个模型中，视线是相互平行的，但是不一定和成像平面垂直。

弱透视投影和类透视投影是更一般化的模型，它们考虑了物体相对于摄像机的深度变化（见图 2.9）。O 为摄像机光心，R 为场景中的参考点。点 P 的弱透视投影分为两个步骤：先把 P 按照垂直投影投到平面 Π_r 上的点 P'，Π_r 平行于 Π' 且过点 R；然后用透视投影把 P' 映射到像点 p（图 2.9 上图）。由于 Π_r 是平行于成像平面的，因此第二步的投影过程就是坐标系的放缩变换。类透视投影同时考虑了参考点不在光轴上的影响和景物的深度变化；继续使用前面的符号，且用 Δ 表示光心 O 到参考点 R 之间的连线。先用平行投影把 P 映射到 Π_r 上的点 P'；然后再用透视投影把 P' 映射到像点 p。

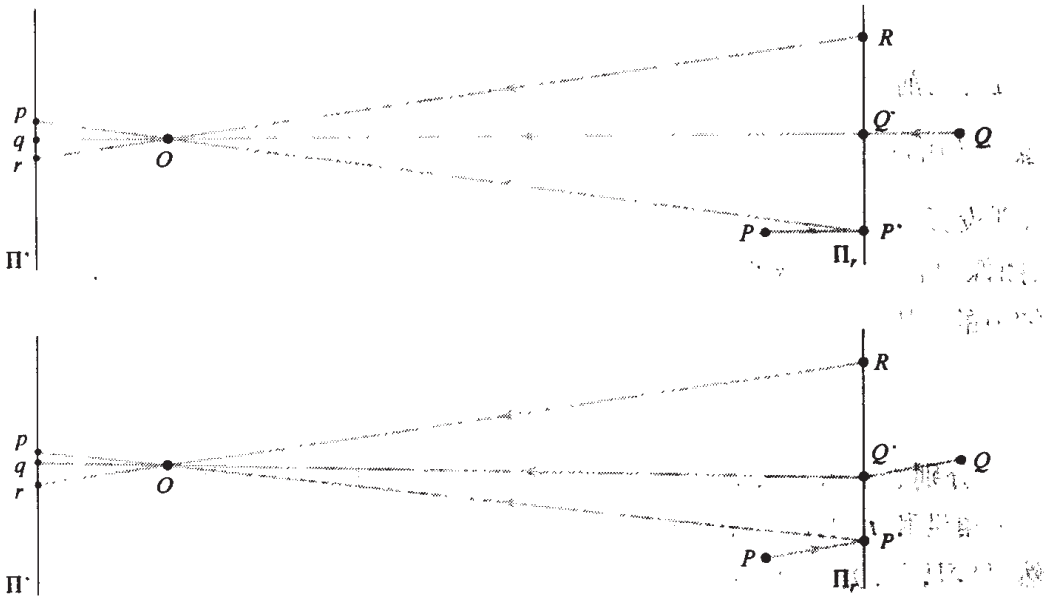


图 2.9　仿射投影模型：(上)弱透视投影；(下)类透视投影

2.3.2　仿射投影方程

下面来推导弱透视投影方程。若 z_r 代表参考点 R 的深度，可以用下面的齐次坐标表示投影过程的两个步骤 $P \rightarrow P' \rightarrow p$：

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \longrightarrow \begin{pmatrix} x \\ y \\ z_r \end{pmatrix} \longrightarrow \begin{pmatrix} \hat{u} \\ \hat{v} \\ 1 \end{pmatrix} = \begin{pmatrix} x/z_r \\ y/z_r \\ 1 \end{pmatrix}$$

或者用矩阵形式，

$$\begin{pmatrix} \hat{u} \\ \hat{v} \\ 1 \end{pmatrix} = \frac{1}{z_r} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & z_r \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

引入摄像机的标定矩阵$\mathcal{K}$以及外参数$\mathcal{R}$和 $\boldsymbol{t}$,可以得到更一般形式的投影方程:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z_r}\mathcal{K}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & z_r \end{pmatrix}\begin{pmatrix} \mathcal{R} & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{pmatrix}\begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix} \tag{2.18}$$

其中,$\boldsymbol{P}$ 表示点 P 在世界坐标系中的非齐次坐标。最后还要注意 z_r 是一个常量,可以得到

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_2 & \boldsymbol{p}_0 \\ \boldsymbol{0}^T & 1 \end{pmatrix}, \quad 其中,\ \mathcal{K}_2 \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha\cot\theta \\ 0 & \dfrac{\beta}{\sin\theta} \end{pmatrix} \quad 且 \quad \boldsymbol{p}_0 \stackrel{\text{def}}{=} \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}$$

则式(2.18)可以写成

$$\boldsymbol{p} = \mathcal{M}\begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix}, \quad 其中,\quad \mathcal{M} = (\mathcal{A} \quad \boldsymbol{b}) \tag{2.19}$$

$\boldsymbol{p} = (u, v)^T$ 是点 p 的非齐次坐标,$\mathcal{M}$是一个 2×4 投影矩阵[和通用透视投影式(2.15)比较]。在这里,2×3 矩阵$\mathcal{A}$和二维向量 $\boldsymbol{b}$ 分别定义为

$$\mathcal{A} = \frac{1}{z_r}\mathcal{K}_2\mathcal{R}_2 \quad 且 \quad \boldsymbol{b} = \frac{1}{z_r}\mathcal{K}_2\boldsymbol{t}_2 + \boldsymbol{p}_0$$

其中,$\mathcal{R}_2$ 是$\mathcal{R}$的前两行组成的 2×3 矩阵,$\boldsymbol{t}_2$ 是 $\boldsymbol{t}$ 坐标的前两维。

注意,在$\mathcal{M}$的表达式中并没有出现 $\boldsymbol{t}_z$,且 $\boldsymbol{t}_2$ 和 $\boldsymbol{p}_0$ 是关联的:若以 $\boldsymbol{t}_2 + \boldsymbol{a}$ 代替 $\boldsymbol{t}_2$,以 $\boldsymbol{p}_0 - \frac{1}{z_r}\mathcal{K}_2\boldsymbol{a}$ 代替 $\boldsymbol{p}_0$,则投影矩阵不变。这种冗余性使我们可以任意选取 $u_0 = v_0 = 0$。也就是说,在弱透视投影中,图像中心没有实际的意义。在$\mathcal{M}$中,z_r,α 和β 也是关联的,且在大部分应用中 z_r 是无法事先知道的。因此有

$$\mathcal{M} = \frac{1}{z_r}\begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}(\mathcal{R}_2 \quad \boldsymbol{t}_2) \tag{2.20}$$

其中,k 和 s 分别表示像素长宽比和偏歪。弱透视投影可由两个内参数(k 和 s)、5 个外参数($\mathcal{R}_2$ 的三个角度和 $\boldsymbol{t}_2$ 的两个坐标)和一个场景相关的结构参数 z_r 确定。

显然(见习题),类透视投影也能写成形如式(2.19)的形式:

$$\mathcal{M} = \frac{1}{z_r}\begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\left(\begin{pmatrix} 1 & 0 & -x_r/z_r \\ 0 & 1 & -y_r/z_r \end{pmatrix}\mathcal{R} \quad \boldsymbol{t}_2\right) \tag{2.21}$$

其中 x_r,y_r 和z_r 表示参考点 R 在摄像机归一化坐标系中的坐标。若 $x_r = y_r = 0$,则方程(2.21)退化为方程(2.20)的弱透视投影方程。按照方程(2.21),类透视投影矩阵包括两个内参数(k,s)、5 个外参数(定义$\mathcal{R}$的三个角度与 t_2 的两个坐标)和三个结构参数 x_r,y_r 和 z_r。实际应用中,参考点常取在图像中可见的一个点。它的三维坐标 x_r,y_r 和z_r 显然是不可知的,但是投影到图像上的二维坐标 u_r 和v_r 是可以得到的。式(2.21)变为

$$\mathcal{M} = \frac{1}{z_r}\left(\begin{pmatrix} k & s & u_0 - u_r \\ 0 & 1 & v_0 - v_r \end{pmatrix}\mathcal{R} \quad \begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\boldsymbol{t}_2\right) \tag{2.22}$$

在这个公式中,类透视投影方程由 4 个内参数(k,s,u_0,v_0)、5 个外参数($\mathcal{R}$的三个角度和 $\boldsymbol{t}_2$ 的两个坐标)和 1 个结构参数 z_r 决定。

如果弱透视投影和类透视投影中的深度值固定，z_r（如 $z_r = 1$）就可以从方程(2.20)、方程(2.21)和方程(2.22)得到正投影和平行投影的方程。若用多个正投影（平行投影）观察同一景物（也可理解为从变焦摄像头录下的图像序列），图像的大小就变成是相对值，方程(2.20)[方程(2.21)或方程(2.22)]中必须用$\mathcal{K}_2$ 代替原有的简化的标定矩阵。

2.3.3　放射投影矩阵的性质

仿射投影矩阵是一个 2×4 矩阵$\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$，其中$\mathcal{A}$是一个任意的二阶 2×3 矩阵，$\boldsymbol{b}$ 是二维实数空间的任意变量。要满足秩为 2 这个条件是因为：若秩为 1，则所有物点都会投影到图像上的一条直线上。由方程(2.20)、方程(2.21)和方程(2.22)得到的弱透视投影和类透视投影的投影矩阵$\mathcal{A}$的秩一定为 2，因为它们都可以表示成几个二阶矩阵的乘积。

弱透视投影和仿射投影都是由 8 个独立参数确定的，而类透视投影却有 10 个自由度。显然弱透视投影和类透视投影都属于仿射投影。比较参数个数使我们想到，也许可以把任意的仿射投影矩阵写成弱透视投影或类透视投影形式。但是若不增加额外限制，类透视投影不是惟一的。这些由下面的定理保证。

定理 2　仿射投影矩阵可以惟一写成（不记符号）式(2.20)的弱透视投影矩阵形式，也可写成式(2.21)或式(2.22)中定义的类透视投影矩阵，但要限制 $k = 1, s = 0$。

这个定理留给读者在习题中证明，Faugeras(2001，Propositions 4.26 和 4.27)等人已经证明了这个定理。它表明任意仿射投影矩阵可以写成弱透视投影或类透视投影，且几何性质不变。例如，第 12 章中显示，弱透视投影不改变直线的平行性，定理 2 表明任意的仿射投影都有这样的性质。在用类透视投影表示任意 2×4 矩阵时限制 $k = 1, s = 0$，并不表示类透视投影中的像素长宽比和偏歪没有作用。

2.4　注释

Craig(1989)详细介绍了坐标系表示和运动学。严密的几何摄像机模型可以在 Faugeras(1993)，Hartley 和 Zisserman(2000)以及 Faugeras 等(2001)中找到。Ohta、Maenobu 和 Sakai(1981)把类透视投影引入计算机视觉，Aloimonos(1990)研究了它的性质。Basri(1996)分析了类透视投影和仿射投影的关系。Faugeras 和 Papadopoulo(1997)推导出了用 Plücker coordinates 坐标系表示的直线的透视投影方程。下一章将用本章介绍的原理做摄像机标定（例如，从基准点的图像位置计算摄像机内外参数），这些基础也是第 10 章立体视觉和第 13 章运动分析中的关键。本章中的重要方程在表 2.1 中列出。

表 2.1　参考卡：摄像机几何模型

平面方程（齐次）	$\boldsymbol{\Pi}\cdot\boldsymbol{P} = ax + by + cz - d = 0$
二次曲面方程（齐次）	$\boldsymbol{P}^T\mathcal{Q}\boldsymbol{P} = 0,\ \mathcal{Q} = \begin{pmatrix} a_{200} & \frac{1}{2}a_{110} & \frac{1}{2}a_{101} & \frac{1}{2}a_{100} \\ \frac{1}{2}a_{110} & a_{020} & \frac{1}{2}a_{011} & \frac{1}{2}a_{010} \\ \frac{1}{2}a_{101} & \frac{1}{2}a_{011} & a_{002} & \frac{1}{2}a_{001} \\ \frac{1}{2}a_{100} & \frac{1}{2}a_{010} & \frac{1}{2}a_{001} & a_{000} \end{pmatrix}$

（续表）

旋转矩阵	${}^{B}_{A}\mathcal{R}=\begin{pmatrix} \boldsymbol{i}_A\cdot\boldsymbol{i}_B & \boldsymbol{j}_A\cdot\boldsymbol{i}_B & \boldsymbol{k}_A\cdot\boldsymbol{i}_B \\ \boldsymbol{i}_A\cdot\boldsymbol{j}_B & \boldsymbol{j}_A\cdot\boldsymbol{j}_B & \boldsymbol{k}_A\cdot\boldsymbol{j}_B \\ \boldsymbol{i}_A\cdot\boldsymbol{k}_B & \boldsymbol{j}_A\cdot\boldsymbol{k}_B & \boldsymbol{k}_A\cdot\boldsymbol{k}_B \end{pmatrix}$
坐标系变换	${}^{B}P={}^{B}_{A}\mathcal{R}\,{}^{A}P+{}^{B}O_A$
透视投影方程（齐次）	$\boldsymbol{p}=\frac{1}{z}\mathcal{M}\boldsymbol{P}$
内参数矩阵	$\mathcal{K}=\begin{pmatrix} \alpha & -\alpha\cot\theta & u_0 \\ 0 & \beta/\sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}$
透视投影矩阵	$\mathcal{M}=\mathcal{K}(\mathcal{R}\quad \boldsymbol{t})$
仿射投影方程（非齐次）	$\boldsymbol{p}=\mathcal{M}\begin{pmatrix}\boldsymbol{P}\\1\end{pmatrix}=\mathcal{A}\boldsymbol{P}+\boldsymbol{b}$
弱透视投影矩阵	$\mathcal{M}=(\mathcal{A}\quad \boldsymbol{b})=\frac{1}{z_r}\begin{pmatrix}k & s\\0 & 1\end{pmatrix}(\mathcal{R}_2\quad \boldsymbol{t}_2)$
类透视投影矩阵 I	$\mathcal{M}=\frac{1}{z_r}\begin{pmatrix}k & s\\0 & 1\end{pmatrix}\left(\begin{pmatrix}1 & 0 & -x_r/z_r\\0 & 1 & -y_r/z_r\end{pmatrix}\mathcal{R}\quad \boldsymbol{t}_2\right)$
类透视投影矩阵 II	$\mathcal{M}=\frac{1}{z_r}\left(\begin{pmatrix}k & s & u_0-u_r\\0 & 1 & v_0-v_r\end{pmatrix}\mathcal{R}\quad \begin{pmatrix}k & s\\0 & 1\end{pmatrix}\boldsymbol{t}_2\right)$

习题

2.1 B 是坐标系 A 依次绕 $\boldsymbol{i}_A$，$\boldsymbol{j}_A$ 和 $\boldsymbol{k}_A$ 转 θ 角以后的坐标系，求旋转矩阵 ${}^{A}_{B}\mathcal{R}$。

2.2 证明旋转矩阵有如下性质：(a)旋转矩阵的逆和转置相同；(b)行列式为 1。

2.3 证明刚体变换的矩阵在矩阵乘法作用下是群。

2.4 ${}^{A}\mathcal{T}$ 表示坐标系 A 中变换 $\mathcal{T}$ 的矩阵，

$$ {}^{A}\mathcal{T}=\begin{pmatrix} {}^{A}\mathcal{R} & {}^{A}\boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{pmatrix} $$

用 ${}^{A}\mathcal{T}$ 与 A 和 B 之间的刚体变换关系求 $\mathcal{T}$ 在 B 下的表示 ${}^{B}\mathcal{T}$。

2.5 坐标系 A 做刚体变换 $\mathcal{T}$ 后得到 B，证明 ${}^{B}P=\mathcal{T}^{-1}{}^{A}P$

2.6 证明，坐标系 F 绕 $\boldsymbol{k}$ 轴旋转 θ 可以表示为

$$ {}^{F}P'=\mathcal{R}\,{}^{F}P,\quad 其中,\quad \mathcal{R}=\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} $$

2.7　证明坐标系的刚体变换保持距离和角度。

2.8　证明:若摄像机坐标系有偏歪,两个图像轴的夹角不是 90 度,则方程(2.11)变为方程(2.12)。

2.9　用 $\boldsymbol{O}$ 表示光心在参考坐标系内的齐次坐标,$\mathcal{M}$表示对应的透视投影矩阵,证明$\mathcal{M}(\boldsymbol{O})=0$。

2.10　证明定理 1 的条件是必要的。

2.11　证明定理 1 的条件是充分的。这里的定理 1 和 Faugeras(1993)及 Heyden(1995)有一些区别。$\mathrm{Det}(\mathcal{A})\neq 0$ 改为了 $\boldsymbol{a}_3\neq 0$,显然,$\mathrm{Det}(\mathcal{A})\neq 0$ 推出 $\boldsymbol{a}_3\neq 0$。

2.12　${}^A\Pi$ 表示坐标系 A 上的平面 Π 齐次坐标,坐标系 B 中它的表示${}^B\Pi$ 是什么?

2.13　${}^A\mathcal{Q}$表示坐标系 A 内一个二次曲面的对称矩阵,坐标系 B 中它的表示${}^B\mathcal{Q}$是什么?

2.14　证明定理 2。

2.15　**线性 Plücker 坐标系。**$\mathbb{R}^4$ 空间上两个向量 $\boldsymbol{u},\boldsymbol{v}$ 的外积定义为

$$\boldsymbol{u}\wedge\boldsymbol{v}\stackrel{\text{def}}{=}\begin{pmatrix}u_1v_2-u_2v_1\\u_1v_3-u_3v_1\\u_1v_4-u_4v_1\\u_2v_3-u_3v_2\\u_2v_4-u_4v_2\\u_3v_4-u_4v_3\end{pmatrix}$$

在给定的坐标系中,$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 表示$\mathbb{E}^3$ 中的两个向量,则 $\boldsymbol{L}=\boldsymbol{A}\wedge\boldsymbol{B}$ 称为 A 到 B 交线的坐标。

(a) 记 $\boldsymbol{L}=(L_1,L_2,L_3,L_4,L_5,L_6)^T$,$O$ 表示坐标原点,H 是 O 在 L 上的投影位置。用$\overrightarrow{OA}$和$\overrightarrow{OB}$表示非齐次坐标。证明,$\overrightarrow{AB}=-(L_3,L_5,L_6)^T$ 和$\overrightarrow{OA}\times\overrightarrow{OB}=\overrightarrow{OH}\times\overrightarrow{AB}=(L_4,-L_2,L_1)^T$,并且直线的$\mathbb{R}^4$ 坐标满足二次约束 $L_1L_6-L_2L_5+L_3L_4=0$。

(b) 证明 A 和 B 在直线 L 上移动只改变 $\boldsymbol{L}$ 的全局比例,Plücker 坐标是齐次坐标。

(c) 证明对$\mathbb{R}^4$ 上的点 $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ 和 $\boldsymbol{t}$,有

$$(\boldsymbol{x}\wedge\boldsymbol{y})\cdot(\boldsymbol{z}\wedge\boldsymbol{t})=(\boldsymbol{x}\cdot\boldsymbol{z})(\boldsymbol{y}\cdot\boldsymbol{t})-(\boldsymbol{x}\cdot\boldsymbol{t})(\boldsymbol{y}\cdot\boldsymbol{z})$$

(d) 证明 Plücker 坐标 $\boldsymbol{L}$ 表示的直线和它在图像上的齐次坐标 $\boldsymbol{l}$ 满足如下关系

$$\rho\boldsymbol{l}=\tilde{\mathcal{M}}\boldsymbol{L},\quad 其中,\quad \tilde{\mathcal{M}}\stackrel{\text{def}}{=}\begin{pmatrix}(\boldsymbol{m}_2\wedge\boldsymbol{m}_3)^T\\(\boldsymbol{m}_3\wedge\boldsymbol{m}_1)^T\\(\boldsymbol{m}_1\wedge\boldsymbol{m}_2)^T\end{pmatrix}\tag{2.23}$$

$\boldsymbol{m}_1^T,\boldsymbol{m}_2^T$ 和 $\boldsymbol{m}_3^T$ 表示$\mathcal{M}$的各行,ρ 是比例系数。

提示:L 是过 A 和 B 的直线,A 和 B 的投影点用 a 和 b 表示,齐次坐标分别是 $\boldsymbol{a},\boldsymbol{b}$。$a,b$ 在 l 上,用 $\boldsymbol{l}$ 表示 l 的齐次坐标,则有 $\boldsymbol{l}\cdot\boldsymbol{a}=\boldsymbol{l}\cdot\boldsymbol{b}=0$。

(e) 已知 L 的 Plücker 坐标为 $\boldsymbol{L}=(L_1,L_2,L_3,L_4,L_5,L_6)^T$,点 P 的齐次坐标向量为 $\boldsymbol{P}$。证明 P 在 L 上的充要条件是

$$\mathcal{L}\boldsymbol{P}=0,\quad 其中,\quad \mathcal{L}\stackrel{\text{def}}{=}\begin{pmatrix}0&L_6&-L_5&L_4\\-L_6&0&L_3&-L_2\\L_5&-L_3&0&L_1\\-L_4&L_2&-L_1&0\end{pmatrix}$$

(f) Π 是平面 Π 的齐次坐标,证明直线 L 在平面 Π 上的充要条件是

$$\mathcal{L}^*\boldsymbol{\Pi}=0,\quad 其中,\quad \mathcal{L}^*\stackrel{\text{def}}{=}\begin{pmatrix}0&L_1&L_2&L_3\\-L_1&0&L_4&L_5\\-L_2&-L_4&0&L_6\\-L_3&-L_5&-L_6&0\end{pmatrix}$$

第 3 章　摄像机的几何标定

本章将进一步讨论摄像机的内外参数估计问题(称为几何标定)。在本章中我们假设摄像机观察到的特征(点或线)都在世界坐标系中有确定的位置(见图 3.1)。在这个假设下,摄像机标定可以看做是一个优化问题,优化的目标是使摄像机观察到的特征与理论位置(由第 2 章讨论的投影方程推导出的位置)之间的距离最小。

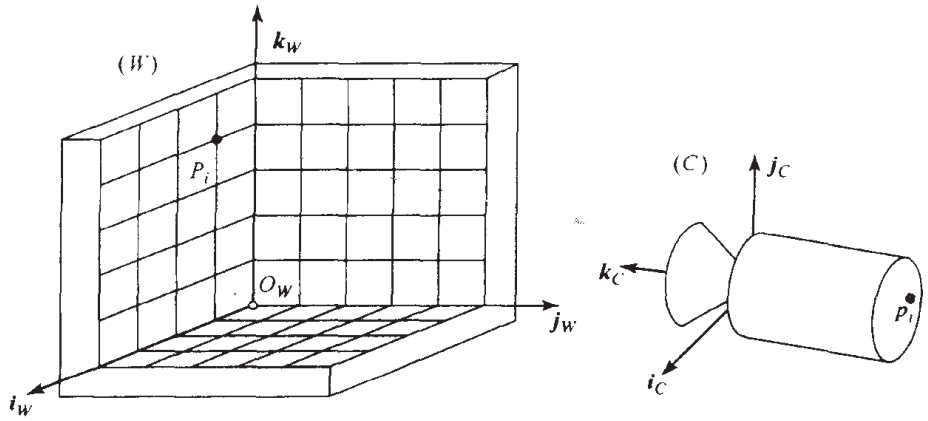


图 3.1　摄像机标定装置:这里用三个相互垂直的棋盘格平面构成标定框架。可以用其他样式的框架,也可以用包括直线在内的其他几何图形作为特征

我们将首先介绍用于优化问题的最小二乘法,然后还将介绍多种线性和非线性优化方法。摄像机标定之后,对每一个成像点都可以确定从光心射出且经过这个点的一条射线,也就可以从图像得到精确的三维测量。在本章的最后将介绍其在移动机器人定位中的应用。

3.1　最小二乘法的参数估计

前面提到,摄像机标定问题等价于使理论预测值与观察值之间的均方误差最小。这一节将介绍如何使用最小二乘法来解决这个问题,这个方法还将在本书的其他部分使用。

3.1.1　线性最小二乘法

假设有 p 个关于 q 个未知数的线性方程组:

$$\begin{cases} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1q}x_q = y_1 \\ u_{21}x_1 + u_{22}x_2 + \cdots + u_{2q}x_q = y_2 \\ \cdots \\ u_{p1}x_1 + u_{p2}x_2 + \cdots + u_{pq}x_q = y_p \end{cases} \Longleftrightarrow \mathcal{U}\boldsymbol{x} = \boldsymbol{y} \tag{3.1}$$

在这个方程中组,令

$$\mathcal{U}=\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1q} \\ u_{21} & u_{22} & \cdots & u_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ u_{p1} & u_{p2} & \cdots & u_{pq} \end{pmatrix}, \quad \mathrm{x}=\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_q \end{pmatrix}, \qquad \mathrm{y}=\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_p \end{pmatrix}$$

由线性代数理论可得到下面(一般情况下)的结论:

- 当 $p<q$ 时,方程的解集构成一个$(q-p)$维的$\mathbb{R}^q$ 上的子空间;
- 当 $p=q$ 时,有惟一解;
- 当 $p>q$ 时,方程无解。

当$\mathcal{U}$的秩(线性独立的行或列的最大数目)最大(等于 $\min(p,q)$,这就是一般情况的意思)时,以上结论成立。当秩小于 $\min(p,q)$时,解的存在性取决于 $\boldsymbol{y}$ 的取值以及它是否在$\mathcal{U}$构成的空间内(由各列张成的$\mathbb{R}^p$ 的子空间)。

正则方程和伪逆　本节的剩余部分将假设$\mathcal{U}$是过约束的,即 $p>q$,且秩为 q。在这种情况下没有标准的解,只能找到使误差最小的向量 $\boldsymbol{x}$,误差定义为

$$E \stackrel{\text{def}}{=} \sum_{i=1}^{p}(u_{i1}x_1+\cdots+u_{iq}x_q-y_i)^2=|\mathcal{U}\boldsymbol{x}-\boldsymbol{y}|^2$$

由于 E 正比于方程的均方误差,因此最小二乘法的意思就是使均方误差最小化。

我们可以把误差写成 $E=\boldsymbol{e}\cdot\boldsymbol{e}$,其中 $\boldsymbol{e}\stackrel{\text{def}}{=}\mathcal{U}\boldsymbol{x}-\boldsymbol{y}$。为了找到使 E 最小的$\boldsymbol{x}$,$\boldsymbol{x}$ 的每一个分量 $x_i(i=1,\cdots,q)$的导数必须为 0,即

$$\frac{\partial E}{\partial x_i}=2\frac{\partial \boldsymbol{e}}{\partial x_i}\cdot\boldsymbol{e}=0 \quad , \quad i=1,\cdots,q$$

若$\mathcal{U}$的各列为 $\boldsymbol{c}_j=(u_{1j},\cdots,u_{mj})^T(j=1,\cdots,q)$,则有

$$\frac{\partial \boldsymbol{e}}{\partial x_i}=\frac{\partial}{\partial x_i}\left[\begin{pmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_q \end{pmatrix}\begin{pmatrix} x_1 \\ \cdots \\ x_q \end{pmatrix}-\boldsymbol{y}\right]=\frac{\partial}{\partial x_i}(x_1\boldsymbol{c}_1+\cdots+x_q\boldsymbol{c}_q-\boldsymbol{y})=\boldsymbol{c}_i$$

特别注意,从$\partial E/\partial x_i=0$ 可以导出 $\boldsymbol{c}_i^T(\mathcal{U}\ \boldsymbol{x}-\boldsymbol{y})=0$。把这些约束从上到下排列起来就变成了下面的形式:

$$\boldsymbol{0}=\begin{pmatrix} \boldsymbol{c}_1^T \\ \cdots \\ \boldsymbol{c}_q^T \end{pmatrix}(\mathcal{U}\boldsymbol{x}-\boldsymbol{y})=\mathcal{U}^T(\mathcal{U}\boldsymbol{x}-\boldsymbol{y}) \Longleftrightarrow \mathcal{U}^T\mathcal{U}\boldsymbol{x}=\mathcal{U}^T\boldsymbol{y}$$

当$\mathcal{U}$满秩(q)时,矩阵 $\mathcal{U}^T\mathcal{U}$是可逆的,则方程 $\boldsymbol{x}=\mathcal{U}^\dagger\boldsymbol{y}$(其中$\mathcal{U}^\dagger\stackrel{\text{def}}{=}[(\mathcal{U}^T\mathcal{U})^{-1}\mathcal{U}^T]$)的解,就是所求的 $\boldsymbol{x}$,$q\times q$ 矩阵 $\mathcal{U}^\dagger$ 称为矩阵 $\mathcal{U}$的伪逆矩阵。当 $\mathcal{U}$是非奇异方阵时,$\mathcal{U}^\dagger$ 和$\mathcal{U}^{-1}$相等。在解线性最小二乘法时,不一定要计算伪逆矩阵的实际值,QR 分解或分解 SVD 方法(第 12 章后半部分介绍)都是很好的数值解法。

齐次系统和特征值问题　我们对原始问题稍加变化,同样是 p 个线性方程,q 个未知数,但是矩阵 $\boldsymbol{y}$ 为 0,

$$\begin{cases} u_{11}x_1+u_{12}x_2+\cdots+u_{1q}x_q=0 \\ u_{21}x_1+u_{22}x_2+\cdots+u_{2q}x_q=0 \\ \cdots \\ u_{p1}x_1+u_{p2}x_2+\cdots+u_{pq}x_q=0 \end{cases} \Longleftrightarrow \mathcal{U}\boldsymbol{x}=\boldsymbol{0} \tag{3.2}$$

这是关于 $\boldsymbol{x}$ 的齐次方程(若 $\boldsymbol{x}$ 为解,则对任意$\lambda \neq 0$,$\lambda \boldsymbol{x}$ 也是解)。当 $p=q$ 且矩阵$\mathcal{U}$非奇异时,方程(3.2)有惟一解 $\boldsymbol{x}=\boldsymbol{0}$。当 $p \geqslant q$ 且$\mathcal{U}$奇异(秩小于 q)时,方程才有非零解。这种情况下,如果不引入附加限制,则原始的误差最小化 $E=|\mathcal{U}\boldsymbol{x}|^2$ 没有意义,因为总能让 $\boldsymbol{x}$ 接近 $\boldsymbol{0}$ 使得误差很小。由方程的齐次性,有 $E(\lambda \boldsymbol{x})=\lambda^2 E(\boldsymbol{x})$,因此要求$|\boldsymbol{x}|^2=1$ 是一个合理的限制,这样可以避免没有意义的解,而且可以得到惟一解。

误差 E 可以写成$|\mathcal{U}\boldsymbol{x}|^2=\boldsymbol{x}^T(\mathcal{U}^T\mathcal{U})\boldsymbol{x}$,这个 $q \times q$ 的矩阵$\mathcal{U}^T\mathcal{U}$是对称半正定的(它的特征值为正数或零),且可以通过特征值分解过程对角化,特征向量为 $\boldsymbol{e}_i(i=1,\cdots,q)$,特征值为 $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_q$。可以把任意单位向量 $\boldsymbol{x}$ 用这些特征向量表示,$\boldsymbol{x}=\mu_1\boldsymbol{e}_1+\cdots+\mu_q\boldsymbol{e}_q$,其中 $\mu_1^2+\cdots+\mu_q^2=1$,且有

$$
\begin{aligned}
E(\boldsymbol{x})-E(\boldsymbol{e}_1) &= \boldsymbol{x}^T(\mathcal{U}^T\mathcal{U})\boldsymbol{x}-\boldsymbol{e}_1^T(\mathcal{U}^T\mathcal{U})\boldsymbol{e}_1=\lambda_1^2\mu_1^2+\cdots+\lambda_q^2\mu_q^2-\lambda_1^2 \\
&\geqslant \lambda_1^2(\mu_1^2+\cdots+\mu_q^2-1)=0
\end{aligned}
$$

这说明使 E 最小的$\boldsymbol{x}$ 就是$\mathcal{U}^T\mathcal{U}$的最小的特征值对应的特征向量 $\boldsymbol{e}_1$,而且误差 E 此时达到最小值λ_1^2。计算对称矩阵的特征值和特征向量有很多种方法,包括 Jacobi 变换、通过 QR 分解三角化等。SVD 分解是另一种计算特征值和特征向量的方法,在这个方法中,不需要算出矩阵$\mathcal{U}^T\mathcal{U}$。

在举例说明齐次线性最小二乘法之前,让我们考虑一下一个更普遍的优化问题:在$|\mathcal{V}\boldsymbol{x}|^2=1$ 约束下求$|\mathcal{U}\boldsymbol{x}|^2$ 的最小值,其中$\mathcal{V}$是一个 $r \times q$ 的矩阵(当$\mathcal{V}=\mathrm{Id}$ 是单位阵时,问题退化为齐次线性最小二乘法)。使得下式

$$\mathcal{U}^T\mathcal{U}\boldsymbol{x}=\lambda\mathcal{V}^T\mathcal{V}\boldsymbol{x}$$

成立的向量 $\boldsymbol{x}$ 和系数λ称为关于对称阵$\mathcal{U}^T\mathcal{U}$和$\mathcal{V}^T\mathcal{V}$的广义特征向量和对应的广义特征值。在习题中会看到,所考虑问题的解恰好是使广义特征值最小(半正定性保证这个值最小是 0)情况下的特征值和特征向量。计算关于两个对称矩阵的广义特征值和特征向量,也已经有了有效的方法。

例 3.1 对平面点集进行直线拟合

假设平面上有 n 个点$p_i(i=1,\cdots,n)$,坐标依次为(x_i,y_i)(见图 3.2)。和这些点拟合最好的直线是什么?为了回答这个问题,先要确定一个标准,来衡量对点集直线拟合的好坏,或者说,定义一种误差函数 E 来衡量线和点之间的差异。那么使 E 最小的直线就是最佳拟合。

点到直线的最小均方距离是一种合理的误差函数(见图 3.2)。在第 2 章中讲过,单位法向量为 $\boldsymbol{n}=(a,b)^T$,距原点距离 d 的直线的方程为 $ax+by=d$。很容易验证点$(x,y)^T$ 到直线的距离为$|ax+by-d|$。则我们可以用

$$E(a,b,d)=\sum_{i=1}^{n}(ax_i+by_i-d)^2$$

作为误差测量,则直线拟合问题就转变为求 a、b、d 使E 最小,且满足 $a^2+b^2=1$ 限制的问题。E 对d 求导就可以得到 $0=\partial E/\partial d=-2\sum_{i=1}^{n}(ax_i+by_i-d)$,则

$$d=a\bar{x}+b\bar{y},\quad \text{其中},\quad \bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\quad,\quad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i \tag{3.3}$$

$\bar{x}$ 和 $\bar{y}$ 就是输入点的重心。把上式推导出的 d 代入E,得到

$$E=\sum_{i=1}^{n}[a(x_i-\bar{x})+b(y_i-\bar{y})]^2=|\mathcal{U}\boldsymbol{n}|^2 \quad \text{其中，} \quad \mathcal{U}=\begin{pmatrix} x_1-\bar{x} & y_1-\bar{y} \\ \cdots & \cdots \\ x_n-\bar{x} & y_n-\bar{y} \end{pmatrix}$$

于是原来的问题就转化为求 $\boldsymbol{n}$ 使得 $|\mathcal{U}\boldsymbol{n}|^2$ 最小，且 $\boldsymbol{n}$ 满足 $|\boldsymbol{n}^2|=1$。这是前面研究过的齐次线性方程最小二乘法求解问题，它的解是 2×2 矩阵 $\mathcal{U}^T\mathcal{U}$ 的最小特征值对应的特征向量。得到 a 和 b 后，计算 d 可以按式(3.3)进行。注意 $\mathcal{U}^T\mathcal{U}$ 可以写成

$$\begin{pmatrix} \sum_{i=1}^{n}x_i^2-n\bar{x}^2 & \sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y} \\ \sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y} & \sum_{i=1}^{n}y_i^2-n\bar{y}^2 \end{pmatrix}$$

这个式子是点集 p_i 的二阶惯性矩矩阵。实际上，在本节中定义的最佳拟合直线就是基础力学中定义的最小惯性轴。

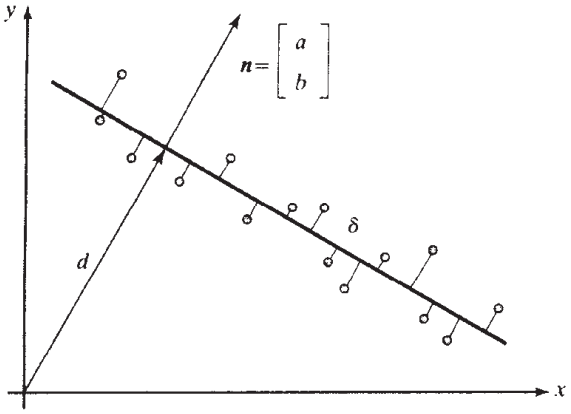


图 3.2　平面上 n 个点的最佳拟合直线，定义为到点集的均方垂直距离最小的直线 δ（在本图中，短的平行线段的均方长度就是直线 δ 到点集的距离）

3.1.2　非线性最小二乘法

考虑一般情况下的 p 个方程和 q 个未知数：

$$\begin{cases} f_1(x_1,x_2,\cdots,x_q)=0 \\ f_2(x_1,x_2,\cdots,x_q)=0 \\ \cdots \\ f_p(x_1,x_2,\cdots,x_q)=0 \end{cases} \Longleftrightarrow \boldsymbol{f}(\boldsymbol{x})=\boldsymbol{0} \tag{3.4}$$

其中，$f_i(i=1,\cdots,p)$ 是 $\mathbb{R}^p$ 到 $\mathbb{R}$ 上的可微函数。我们用缩写 $\boldsymbol{f}=(f_1,\cdots,f_p)^T$ 和 $\boldsymbol{x}=(x_1,\cdots,x_q)^T$ 表示。一般情况下，有

- 若 $p<q$，解构成一个 $\mathbb{R}^q$ 上的 $(q-p)$ 维的子集；
- 若 $p=q$，则有有限个解；
- 若 $p>q$，则无解。

这与线性情况有几处明显的不同：在欠约束情况下，解的维数仍是 $q-p$，但是不再构成一个向量空间，它的结构由 f_i 决定。在 $p=q$ 情况下，不再是惟一解，而是有若干个解。$f_i(i=1,\cdots,p)$ 究竟要满足哪些条件才能保证原来的结论成立，这已经超出了本书的范围。

问题是没有一个通用的方法可以找到方程(3.4)在 $p=q$ 时的所有解,或 $p>q$ 时误差为全局最小值的解

$$E(\boldsymbol{x}) \stackrel{\text{def}}{=} |\boldsymbol{f}(\boldsymbol{x})|^2 = \sum_{i=1}^{p} f_i^2(\boldsymbol{x})$$

因此我们要介绍的是一些把原问题简化为局部线性问题的迭代方法,来力求找到至少一个合理解。它们都建立在 $\boldsymbol{x}$ 邻域上 f_i 的一阶泰勒展开式基础上:

$$f_i(\boldsymbol{x}+\delta\boldsymbol{x}) = f_i(\boldsymbol{x}) + \delta x_1 \frac{\partial f_i}{\partial x_1}(\boldsymbol{x}) + \cdots + \delta x_q \frac{\partial f_i}{\partial x_q}(\boldsymbol{x}) + O(|\delta\boldsymbol{x}|^2) \approx f_i(\boldsymbol{x}) + \nabla f_i(\boldsymbol{x}) \cdot \delta\boldsymbol{x}$$

其中,$\nabla f_i(\boldsymbol{x}) = (\partial f_i/\partial x_1, \cdots, \partial f_i/\partial x_q)^T$ 是 f_i 在点 $\boldsymbol{x}$ 的梯度值。在忽略二阶项 $O(|\delta\boldsymbol{x}|^2)$ 的情况下,马上可以得到

$$\boldsymbol{f}(\boldsymbol{x}+\delta\boldsymbol{x}) \approx \boldsymbol{f}(\boldsymbol{x}) + \mathcal{J}_{\boldsymbol{f}}(\boldsymbol{x})\delta\boldsymbol{x} \tag{3.5}$$

其中,$\mathcal{J}_f(\boldsymbol{x})$ 是 $\boldsymbol{f}$ 的 Jacobian 矩阵,定义为一个 $p \times q$ 矩阵

$$\mathcal{J}_{\boldsymbol{f}}(\boldsymbol{x}) \stackrel{\text{def}}{=} \begin{pmatrix} \nabla f_1^T(\boldsymbol{x}) \\ \cdots \\ \nabla f_p^T(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_1}{\partial x_q}(\boldsymbol{x}) \\ \cdots & \cdots & \cdots \\ \frac{\partial f_p}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_p}{\partial x_q}(\boldsymbol{x}) \end{pmatrix}$$

牛顿法:非线性方程的方阵系统 方程(3.4)在 $p=q$ 时会产生多个解(一般情况)。虽然没有一般的方法可以对任意 $\boldsymbol{f}$ 找到所有这些解,但可以以式(3.5)为基础构造一个迭代方法找到其中一个解:已知解的当前估计值为 $\boldsymbol{x}$,我们的想法是对 $\boldsymbol{x}$ 加一个扰动 $\delta\boldsymbol{x}$,使 $\boldsymbol{f}(\boldsymbol{x}+\delta\boldsymbol{x}) \approx \boldsymbol{0}$,或者,根据式(3.5),

$$\mathcal{J}_{\boldsymbol{f}}(\boldsymbol{x})\delta\boldsymbol{x} = -\boldsymbol{f}(\boldsymbol{x})$$

当这个 Jacobian 矩阵非奇异时,只要解这个线性方程就可以找到 $\delta\boldsymbol{x}$ 的合适解,重复这个过程直到收敛。

牛顿法在接近解的地方收敛很快:它按照平方速度收敛(第 $k+1$ 步的误差与第 k 步误差的平方成正比)。当起始点离解很远时,上边介绍的牛顿法可能失败。很多方法可以提高它的鲁棒性,但是在本书中没有篇幅讨论这些方法。

牛顿法:非线性方程的过约束系统 当 p 大于 q 时,要找一个均方误差 E 的局部极小解。在这个极小值点,E 的导数为零,可以利用这个特征来使用牛顿法。令 $\boldsymbol{F}(\boldsymbol{x}) = \frac{1}{2}\nabla E(\boldsymbol{x})$,用牛顿法可以找到非线性方程组 $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}$ 的一组解。由 E 的微分可得到

$$\boldsymbol{F}(\boldsymbol{x}) = \mathcal{J}_{\boldsymbol{f}}^T(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}) \tag{3.6}$$

那么 $\boldsymbol{F}$ 的 Jacobian 矩阵就是

$$\mathcal{J}_{\boldsymbol{F}}(\boldsymbol{x}) = \mathcal{J}_{\boldsymbol{f}}^T(\boldsymbol{x})\mathcal{J}_{\boldsymbol{f}}(\boldsymbol{x}) + \sum_{i=1}^{p} f_i(\boldsymbol{x})\mathcal{H}_{f_i}(\boldsymbol{x}) \tag{3.7}$$

其中,$\mathcal{H}_{f_i}(\boldsymbol{x})$ 是 f_i 的 Hessian 矩阵,由 f_i 的二阶导数组成

$$\mathcal{H}_{f_i}(\mathrm{x}) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial^2 f_i}{\partial x_1^2}(\mathrm{x}) & \cdots & \frac{\partial^2 f_i}{\partial x_1 \partial x_q}(\mathrm{x}) \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 f_i}{\partial x_1 \partial x_q}(\mathrm{x}) & \cdots & \frac{\partial^2 f_i}{\partial x_q^2}(\mathrm{x}) \end{pmatrix}$$

在牛顿法中，$\delta \boldsymbol{x}$ 满足$\mathcal{J}_F(\boldsymbol{x})\delta_{\boldsymbol{x}} = -F(\boldsymbol{x})$。同样，从式(3.6)和式(3.7)可以推出 $\delta \boldsymbol{x}$ 满足

$$\left[\mathcal{J}_f^T(\boldsymbol{x})\mathcal{J}_f(\boldsymbol{x}) + \sum_{i=1}^{p} f_i(\boldsymbol{x})\mathcal{H}_{f_i}(\boldsymbol{x})\right]\delta \boldsymbol{x} = -\mathcal{J}_f^T(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}) \tag{3.8}$$

高斯牛顿法和 Levenberg-Marquardt 方法　牛顿法中需要计算函数 f_i 的 Hessians 矩阵，这不但困难，而且费时。下面要介绍的两种非线性优化方法不需要计算 Hessians 矩阵。先介绍高斯牛顿法：在这种方法中，还是利用 $\boldsymbol{f}$ 的一阶泰勒展开式逼近 E 的极小值。但对给定的 $\boldsymbol{x}$，要找到 $\delta \boldsymbol{x}$ 使 $E(\boldsymbol{x}+\delta \boldsymbol{x})$最小。把式(3.5)代入式(3.4)得到

$$E(\boldsymbol{x}+\delta \boldsymbol{x}) = |\boldsymbol{f}(\boldsymbol{x}+\delta \boldsymbol{x})|^2 \approx |\boldsymbol{f}(\boldsymbol{x}) + \mathcal{J}_f(\boldsymbol{x})\delta \boldsymbol{x}|^2$$

现在，问题转化为求线性最小二乘解的问题，$\delta \boldsymbol{x}$ 可以通过解线性方程$\mathcal{J}_f^{\dagger}(\boldsymbol{x})\delta \boldsymbol{x} = -\boldsymbol{f}(\boldsymbol{x})$或通过伪逆得到，

$$\mathcal{J}_f^T(\boldsymbol{x})\mathcal{J}_f(\boldsymbol{x})\delta \boldsymbol{x} = -\mathcal{J}_f^T(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}) \tag{3.9}$$

比较式(3.8)和式(3.9)，可以看出高斯牛顿法是牛顿法的一种近似，它忽略了 Hessians 矩阵$\mathcal{H}_{f_i}$部分。当 f_i 在解的附近取值(残差)很小时，这种近似是可以的，因为在式(3.8)中 Hessians 矩阵$\mathcal{H}_{f_i}$要乘以残差。在这种情况下，高斯牛顿法和牛顿法基本相同，都是(近似)平方收敛的。若残差很大，高斯牛顿法可能收敛很慢或根本不收敛。

对式(3.9)稍加修改有

$$[\mathcal{J}_f^T(\boldsymbol{x})\mathcal{J}_f(\boldsymbol{x}) + \mu \mathrm{Id}]\delta \boldsymbol{x} = -\mathcal{J}_f^T(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}) \tag{3.10}$$

其中，μ 在每步迭代中可以取不同的值，就是计算机视觉中常用的 Levenberg-Marquardt 方法。这里包含 Hessians 的项，用单位阵的倍数取代。Levenberg-Marquardt 方法和高斯牛顿法有相同的收敛速度，但是它更加鲁棒：即使在 Jacobian 矩阵$\mathcal{J}_f$ 不满秩以及伪逆不存在的情况下，也能使用。

3.2　使用线性方法进行摄像机标定

下面回到摄像机标定的问题。标定用装置在摄像机中成像，在本节中假设从图像中取了 n 个特征点 $P_i(i=1,\cdots,n)$，这些点的齐次坐标 $\boldsymbol{P}_i$ 是已知的，这可以通过自动或手工的方法得到。标定过程分为两步，(a)计算这个坐标系下摄像机的投影矩阵$\mathcal{M}$；(b)从投影矩阵估计摄像机的内外参数。在 3.2.3 节中，将介绍如何去除退化点，这些可能会导致标定的第一步失败。投影矩阵共有 11 个独立无关的变量，每个点 P_i 和它的投影点 p_i 可以产生一个线性约束关系。当 $n>11$ 时，方程一般是没有公共解的，但是用 3.1.1 节中介绍的方法可以在最小二乘的意义上找到解。

3.2.1 估计投影矩阵

假设摄像机的歪斜不为零，则按照第 2 章的定理 1，投影矩阵$\mathcal{M}$可以是非奇异的任何矩阵。由式(2.16)可得

$$\begin{cases} (\boldsymbol{m}_1 - u_i\boldsymbol{m}_3)\cdot\boldsymbol{P} = 0 \\ (\boldsymbol{m}_2 - v_i\boldsymbol{m}_3)\cdot\boldsymbol{P} = 0 \end{cases}$$

则所有 n 个点的约束可以写成一个 $2n$ 维的线性方程$\mathcal{M}$，$\mathcal{P}\boldsymbol{m} = 0$，方程参数为投影矩阵的 12 个分量。

$$\mathcal{P} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{P}_1^T & \boldsymbol{0}^T & -u_1\boldsymbol{P}_1^T \\ \boldsymbol{0}^T & \boldsymbol{P}_1^T & -v_1\boldsymbol{P}_1^T \\ \cdots & \cdots & \cdots \\ \boldsymbol{P}_n^T & \boldsymbol{0}^T & -u_n\boldsymbol{P}_n^T \\ \boldsymbol{0}^T & \boldsymbol{P}_n^T & -v_n\boldsymbol{P}_n^T \end{pmatrix} \quad , \quad \boldsymbol{m} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \boldsymbol{m}_3 \end{pmatrix} = 0$$

若 $n \geqslant 6$，则可以用齐次线性最小二乘法计算单位向量 $\boldsymbol{m}$ 的值(且因而得到矩阵$\mathcal{M}$)，得到使$|\mathcal{P}\boldsymbol{m}|^2$ 最小的解。

3.2.2 估计内外参数

投影矩阵可以用摄像机的内外参数表示[第 2 章式(2.17)]，因此在得到投影矩阵$\mathcal{M}$后，可以用下面的方法恢复这些参数：把投影矩阵写成$\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$的形式，其中 $\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \boldsymbol{a}_3^T$ 表示$\mathcal{A}$的各行。我们可以得到

$$\rho(\mathcal{A} \quad \boldsymbol{b}) = \mathcal{K}(\mathcal{R} \quad \boldsymbol{t}) \Longleftrightarrow \rho\begin{pmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \boldsymbol{a}_3^T \end{pmatrix} = \begin{pmatrix} \alpha\boldsymbol{r}_1^T - \alpha\cot\theta\boldsymbol{r}_2^T + u_0\boldsymbol{r}_3^T \\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T + v_0\boldsymbol{r}_3^T \\ \boldsymbol{r}_3^T \end{pmatrix}$$

引入未知的比例系数 ρ，使得$\mathcal{M}$的模为 1($|\mathcal{M}| = |\boldsymbol{m}| = 1$)。

若加上旋转矩阵各行都是单位长度且相互垂直的约束，又可得到

$$\begin{cases} \rho = \varepsilon/|\boldsymbol{a}_3|, \\ \boldsymbol{r}_3 = \rho\boldsymbol{a}_3, \\ u_0 = \rho^2(\boldsymbol{a}_1\cdot\boldsymbol{a}_3), \\ v_0 = \rho^2(\boldsymbol{a}_2\cdot\boldsymbol{a}_3), \end{cases} \quad \text{其中,} \quad \varepsilon = \mp 1 \tag{3.11}$$

由于 θ 取值在$\pi/2$ 的邻域之间，其正弦值总是正的，则有

$$\begin{cases} \rho^2(\boldsymbol{a}_1\times\boldsymbol{a}_3) = -\alpha\boldsymbol{r}_2 - \alpha\cot\theta\boldsymbol{r}_1 \\ \rho^2(\boldsymbol{a}_2\times\boldsymbol{a}_3) = \dfrac{\beta}{\sin\theta}\boldsymbol{r}_1 \end{cases} \quad , \quad \begin{cases} \rho^2|\boldsymbol{a}_1\times\boldsymbol{a}_3| = \dfrac{|\alpha|}{\sin\theta} \\ \rho^2|\boldsymbol{a}_2\times\boldsymbol{a}_3| = \dfrac{|\beta|}{\sin\theta} \end{cases} \tag{3.12}$$

于是

$$\begin{cases} \cos\theta = -\dfrac{(\boldsymbol{a}_1\times\boldsymbol{a}_3)\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)}{|\boldsymbol{a}_1\times\boldsymbol{a}_3||\boldsymbol{a}_2\times\boldsymbol{a}_3|} \\ \alpha = \rho^2|\boldsymbol{a}_1\times\boldsymbol{a}_3|\sin\theta \\ \beta = \rho^2|\boldsymbol{a}_2\times\boldsymbol{a}_3|\sin\theta \end{cases} \tag{3.13}$$

由于 α 和 β 的符号可以事先得到，因此一般都设为正的。

从式(3.12)的第二部分可以计算得到 $\boldsymbol{r}_1$ 和 $\boldsymbol{r}_2$，

$$\begin{cases} \boldsymbol{r}_1 = \dfrac{\rho^2 \sin\theta}{\beta}(\boldsymbol{a}_2 \times \boldsymbol{a}_3) = \dfrac{1}{|\boldsymbol{a}_2 \times \boldsymbol{a}_3|}(\boldsymbol{a}_2 \times \boldsymbol{a}_3) \\ \boldsymbol{r}_2 = \boldsymbol{r}_3 \times \boldsymbol{r}_1 \end{cases} \tag{3.14}$$

注意，根据 ε 值的不同，$\mathcal{R}$ 有两种选择。由 $\mathcal{K}\boldsymbol{t} = \rho\boldsymbol{b}$ 得到 $\boldsymbol{t} = \rho\mathcal{K}^{-1}\boldsymbol{b}$，可以确定平移参数 t。在实际应用中，t_z 的符号是事先知道的(取决于世界坐标系的原点是在摄像机前还是摄像机后)，这样就可以确定惟一的摄像机参数了。

3.2.3　特征点的退化问题

特征点 $P_i(i=1,\cdots,n)$ 的一些退化情况可能会引起标定失败。我们只考虑数据点 $\boldsymbol{p}_i(i=1,\cdots,n)$ 的取值没有误差的理想情况，来探讨矩阵 $\mathbb{R}^{12}$ 的零空间，也就是满足 $\mathcal{P}\boldsymbol{l}=\boldsymbol{0}$ 的所有 $\boldsymbol{l}$ 在 $\mathbb{R}^{12}$ 上构成的子空间。

$\boldsymbol{l}$ 是满足上述约束的向量。令 $\boldsymbol{l}$ 由下面的四元组构成，$\boldsymbol{\lambda}=(l_1,l_2,l_3,l_4)^T$，$\boldsymbol{\mu}=(l_5,l_6,l_7,l_8)^T$，$\boldsymbol{\nu}=(l_9,l_{10},l_{11},l_{12})^T$，则有

$$\boldsymbol{0} = \mathcal{P}\boldsymbol{l} = \begin{pmatrix} \boldsymbol{P}_1^T & \boldsymbol{0}^T & -u_1\boldsymbol{P}_1^T \\ \boldsymbol{0}^T & \boldsymbol{P}_1^T & -v_1\boldsymbol{P}_1^T \\ \cdots & \cdots & \cdots \\ \boldsymbol{P}_n^T & \boldsymbol{0}^T & -u_n\boldsymbol{P}_n^T \\ \boldsymbol{0}^T & \boldsymbol{P}_n^T & -v_n\boldsymbol{P}_n^T \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}_1^T\boldsymbol{\lambda} - u_1\boldsymbol{P}_1^T\boldsymbol{\nu} \\ \boldsymbol{P}_1^T\boldsymbol{\mu} - v_1\boldsymbol{P}_1^T\boldsymbol{\nu} \\ \cdots \\ \boldsymbol{P}_n^T\boldsymbol{\lambda} - u_n\boldsymbol{P}_n^T\boldsymbol{\nu} \\ \boldsymbol{P}_n^T\boldsymbol{\mu} - v_n\boldsymbol{P}_n^T\boldsymbol{\nu} \end{pmatrix} \tag{3.15}$$

由式(2.16)和上式联合求解，可得

$$\begin{cases} \boldsymbol{P}_i^T\boldsymbol{\lambda} - \dfrac{\boldsymbol{m}_1^T\boldsymbol{P}_i}{\boldsymbol{m}_3^T\boldsymbol{P}_i}\boldsymbol{P}_i^T\boldsymbol{\nu} = 0 \\ \boldsymbol{P}_i^T\boldsymbol{\mu} - \dfrac{\boldsymbol{m}_2^T\boldsymbol{P}_i}{\boldsymbol{m}_3^T\boldsymbol{P}_i}\boldsymbol{P}_i^T\boldsymbol{\nu} = 0 \end{cases}, \quad i = 1,\cdots,n$$

将分式改写并整理后，得到

$$\begin{cases} \boldsymbol{P}_i^T(\boldsymbol{\lambda}\boldsymbol{m}_3^T - \boldsymbol{m}_1\boldsymbol{\nu}^T)\boldsymbol{P}_i = 0 \\ \boldsymbol{P}_i^T(\boldsymbol{\mu}\boldsymbol{m}_3^T - \boldsymbol{m}_2\boldsymbol{\nu}^T)\boldsymbol{P}_i = 0 \end{cases}, \quad i = 1,\cdots,n \tag{3.16}$$

显然，由 $\boldsymbol{\lambda}=\boldsymbol{m}_1$，$\boldsymbol{\mu}=\boldsymbol{m}_2$，$\boldsymbol{\nu}=\boldsymbol{m}_3$ 组成的 $\boldsymbol{l}$ 是一组解。那么还有其他的解吗？

第一种情况是考虑 $P_i(i=1,\cdots,n)$ 都在一个平面 Π 上的情形，即存在一个四维向量 $\boldsymbol{\Pi}$，使得 $\boldsymbol{\Pi}\cdot\boldsymbol{P}_i=0$。显然，$(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu})$ 取 $(\boldsymbol{\Pi},\boldsymbol{0},\boldsymbol{0})$，$(\boldsymbol{0},\boldsymbol{\Pi},\boldsymbol{0})$ 或 $(\boldsymbol{0},\boldsymbol{0},\boldsymbol{\Pi})$ 或它们的线性组合都是解。换句话说，$\mathcal{P}$ 的零空间是由这些向量及 $\boldsymbol{m}$ 张成。这说明在实际中点集 P_i 不能共面。

一般来说，对给定的非零向量 $\boldsymbol{l}$，满足式(3.16)的点 $P_i(i=1,\cdots,n)$ 处在这两个方程确定的二次曲面的交线上。仔细观察式(3.16)可以发现，由 $\boldsymbol{m}_3\cdot\boldsymbol{P}=0$ 和 $\boldsymbol{\nu}\cdot\boldsymbol{P}=0$ 确定的交线同时满足这两个方程，这表明两个二次曲面的交线是一条直线和一条过原点的三次曲线。三次曲线可以由它上面的 6 个点完全决定，因此 7 个随机选取的点一般不会在该曲线上。由于原点也是该曲线上的一个点，故只要取 $n\geqslant 6$ 就可以保证 $\mathcal{P}$ 的秩为 11，且投影矩阵有惟一解。

3.3 径向畸变

前面一直假设摄像机使用的是完美的透镜。正如第 1 章中提到的，真实透镜会受多种畸变的影响。这一节专门讨论径向畸变，这是一种由于目标点偏离光轴而引起的畸变。不妨设图像中心已知 $u_0 = v_0 = 0$，则投影过程为

$$\boldsymbol{p} = \frac{1}{z}\begin{pmatrix} 1/\lambda & 0 & 0 \\ 0 & 1/\lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}\mathcal{M}\boldsymbol{P} \tag{3.17}$$

其中，$\boldsymbol{\lambda}$为图像中心与像点 p 之间平方距离 d^2 的多项式函数。在大多数应用中，用一个低阶的多项式(例如，$\boldsymbol{\lambda} = 1 + \sum_{p=1}^{q}\kappa_p d^{2p}$，其中 $q \leqslant 3$)就足够了，且畸变相关系数 $\kappa_p(p = 1, \cdots, q)$都设得很小。$d^2$ 是在归一化的图像坐标上计算的(例如，$d^2 = \hat{u}^2 + \hat{v}^2$)。将 $u_0 = 0$ 和 $v_0 = 0$ 代入式(2.13)，经过简单的代数运算后，可以把 d^2 表示成 u 和 v 的函数，

$$d^2 = \frac{u^2}{\alpha^2} + \frac{v^2}{\beta^2} + 2\frac{uv}{\alpha\beta}\cos\theta \tag{3.18}$$

式(3.18)以 u 和 v 显式表示$\boldsymbol{\lambda}$，这样，式(3.17)中的 $q + 11$ 个摄像机参数就得到一个高度非线性的约束。虽然理论上这些参数都可以通过下一节介绍的非线性最小二乘法得到，但是我们更倾向于推荐下面的两步修正方法：先去掉式(3.17)中$\boldsymbol{\lambda}$的影响，然后再用线性最小二乘法求解摄像机的 9 个参数。剩下的 $q + 2$ 参数个可以用简单的非线性方法由式(3.17)和式(3.18)式求得。

3.3.1 投影矩阵估计

径向畸变改变了图像点 p 到图像中心的距离，但是没有改变图像点相对于中心的直线方向。Tsai(1987a)首先引入了这条径向对准约束，它可以表示为

$$\lambda\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \dfrac{\boldsymbol{m}_1 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}} \\ \dfrac{\boldsymbol{m}_2 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}} \end{pmatrix} \Longrightarrow v(\boldsymbol{m}_1 \cdot \boldsymbol{P}) - u(\boldsymbol{m}_2 \cdot \boldsymbol{P}) = 0 \tag{3.19}$$

若已知 n 个基准点，就可以列出关于 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 的线性方程，

$$\mathcal{Q}\boldsymbol{n} = 0, \quad 其中，\ \mathcal{Q} \stackrel{\text{def}}{=} \begin{pmatrix} v_1\boldsymbol{P}_1^T & -u_1\boldsymbol{P}_1^T \\ \cdots & \cdots \\ v_n\boldsymbol{P}_n^T & -u_n\boldsymbol{P}_n^T \end{pmatrix}, \quad \boldsymbol{n} = \begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \end{pmatrix} \tag{3.20}$$

注意它和前面分析过的问题很相似。若 $n \geqslant 8$，则这个方程是超定的，可以用最小二乘法得到模为 1 的解。

3.3.2 估计内外参数

一旦得到 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$，就可以像前面那样定义相应的 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$，且有，

$$\rho\begin{pmatrix}\boldsymbol{a}_1^T\\ \boldsymbol{a}_2^T\end{pmatrix}=\begin{pmatrix}\alpha\boldsymbol{r}_1^T-\alpha\cot\theta\boldsymbol{r}_2^T+u_0\boldsymbol{r}_3^T\\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T+v_0\boldsymbol{r}_3^T\end{pmatrix}$$

计算 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 的模和点积，可以马上得到摄像机的像素长宽比和歪斜角，

$$\frac{\beta}{\alpha}=\frac{|\boldsymbol{a}_2|}{|\boldsymbol{a}_1|}\quad,\quad\cos\theta=-\frac{\boldsymbol{a}_1\cdot\boldsymbol{a}_2}{|\boldsymbol{a}_1||\boldsymbol{a}_2|}\tag{3.21}$$

由于 $\boldsymbol{r}_2^T$ 是旋转矩阵的第 2 行，所以一定是单位向量，则有

$$\alpha=\varepsilon\rho|\boldsymbol{a}_1|\sin\theta\quad,\quad\beta=\varepsilon\rho|\boldsymbol{a}_2|\sin\theta\tag{3.22}$$

其中，$\varepsilon=\mp1$。代数推导后可以得到下面的方程：

$$\begin{cases}\boldsymbol{r}_1=\dfrac{\varepsilon}{\sin\theta}\left(\dfrac{1}{|\boldsymbol{a}_1|}\boldsymbol{a}_1+\dfrac{\cos\theta}{|\boldsymbol{a}_2|}\boldsymbol{a}_2\right)\\ \boldsymbol{r}_2=\dfrac{\varepsilon}{|\boldsymbol{a}_2|}\boldsymbol{a}_2\end{cases}$$

把该式和 $\boldsymbol{r}_3=\boldsymbol{r}_1\times\boldsymbol{r}_2$ 联立求解，可以得到旋转矩阵$\mathcal{R}$，但有两个可能的解。两个平移参数也可以通过下式求解：

$$\begin{pmatrix}\alpha t_x-\alpha\cot\theta t_y\\ \dfrac{\beta}{\sin\theta}t_y\end{pmatrix}=\rho\begin{pmatrix}b_1\\ b_2\end{pmatrix}$$

其中，b_1 和 b_2 是 $\boldsymbol{b}$ 的前两个分量。然后就可以接着计算 t_x 和 t_y 的值，

$$\begin{cases}t_x=\dfrac{\varepsilon}{\sin\theta}\left(\dfrac{b_1}{|a_1|}+\dfrac{b_2\cos\theta}{|a_2|}\right)\\ t_y=\dfrac{\varepsilon b_2}{|a_2|}\end{cases}$$

若没有其他约束，不可能得到深度信息 t_z 和实际的摄像机放大倍数，即不能从 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 得到 ρ。要估计这几个参数，要用到原始的投影方程。可以把式(3.19)的左式改写为：

$$\begin{cases}(\boldsymbol{m}_1-\lambda u\boldsymbol{m}_3)\cdot\boldsymbol{P}=0\\ (\boldsymbol{m}_2-\lambda v\boldsymbol{m}_3)\cdot\boldsymbol{P}=0\end{cases}\tag{3.23}$$

其中，$\boldsymbol{m}_1$ 和$\boldsymbol{m}_2$ 是已知的，由方程(2.17)可得 $\boldsymbol{m}_3^T=(\boldsymbol{r}_3^T\quad t_z)$，其中$\boldsymbol{r}_3$ 也是已知的。把式(3.21)和式(3.22)得到的 α，β 和 $\cos\theta$ 代入式(3.18)中可以得到 d^2，

$$d^2=\frac{1}{\rho^2}\frac{|u\boldsymbol{a}_2-v\boldsymbol{a}_1|^2}{|\boldsymbol{a}_1\times\boldsymbol{a}_2|^2}$$

再把这个值代入式(3.23)得到一个关于 ρ，t_z 和 $\kappa_p(p=1,\cdots,q)$的非线性方程组。只要有足够的数据点，就可以用 3.1.2 节介绍的非线性最小二乘法求解这些参数。这些方法都是迭代方法，需要初始化未知变量。我们可以先假设$\lambda=1$，用线性最小二乘法来估计 ρ 和 t_z 的初值。对于畸变参数，可以设它们的初值为 0。前面分析过，若规定了 t_z 的方向，就可以求解 t_z 的不确定性。

3.3.3　退化情况

下面分析在哪些情况下不能确定惟一的 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$。任给零空间中的一个向量 $\boldsymbol{l}$，我们定

义向量$\boldsymbol{\lambda}=(l_1,l_2,l_3,l_4)^T$和$\boldsymbol{\mu}=(l_5,l_6,l_7,l_8)^T$,则有

$$\mathbf{0}=\mathcal{Q}\boldsymbol{l}=\begin{pmatrix} v_1\boldsymbol{P}_1^T & -u_1\boldsymbol{P}_1^T \\ \cdots & \cdots \\ v_n\boldsymbol{P}_n^T & -u_n\boldsymbol{P}_n^T \end{pmatrix}\begin{pmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\mu}\end{pmatrix}=\begin{pmatrix} v_1\boldsymbol{P}_1^T\boldsymbol{\lambda}-u_1\boldsymbol{P}_1^T\boldsymbol{\mu} \\ \cdots \\ v_n\boldsymbol{P}_n^T\boldsymbol{\lambda}-u_n\boldsymbol{P}_n^T\boldsymbol{\mu}\end{pmatrix}$$

把上式整理为关于u_i和v_i的形式,将分式改写并整理后,得到

$$\boldsymbol{P}_i^T(\boldsymbol{m}_2\boldsymbol{\lambda}^T-\boldsymbol{m}_1\boldsymbol{\mu}^T)\boldsymbol{P}_i=0 \quad , \quad i=1,\cdots,n \tag{3.24}$$

由$\boldsymbol{\lambda}=\boldsymbol{m}_1$和$\boldsymbol{\mu}=\boldsymbol{m}_2$得到的$\boldsymbol{l}$显然是方程的一个解(无噪声,所有图像点都是准确的)。若所有的点$P_i(i=1,\cdots,n)$都在同一平面Π上,即$\boldsymbol{\Pi}\cdot\boldsymbol{P}_i=0$(其中$\Pi$是一个四维向量),则$(\boldsymbol{\lambda},\boldsymbol{\mu})$可以取$(\boldsymbol{\Pi},\mathbf{0})$,或$(\mathbf{0},\boldsymbol{\Pi})$,或这两个向量的任意线性组合构成式(3.24)的解。零空间$\mathcal{Q}$是由$\boldsymbol{l}$这些向量展成的三维向量空间。故而处于一个平面上的参考点,不能用于标定。

普遍说来,已知$\boldsymbol{\lambda}$和$\boldsymbol{\mu}$,若点集P_i处在式(3.24)定义的二次曲面上,就是一种退化情况。注意这个面中包括下面4条直线:$\boldsymbol{\lambda}\cdot\boldsymbol{P}=\boldsymbol{\mu}\cdot\boldsymbol{P}=0$,$\boldsymbol{\lambda}\cdot\boldsymbol{P}=\boldsymbol{m}_1\cdot\boldsymbol{P}=0$,$\boldsymbol{\mu}\cdot\boldsymbol{P}=\boldsymbol{m}_2\cdot\boldsymbol{P}=0$和$\boldsymbol{m}_1\cdot\boldsymbol{P}=\boldsymbol{m}_2\cdot\boldsymbol{P}=0$。因此,它一定由两个平面或一个圆锥,或双曲面的一支,或抛物面组成。只要非退化的点足够多,就一定可以用最小二乘法得到惟一解。

3.4 分析摄影地形测量法

迄今为止介绍的所有标定方法都忽略了标定过程中的一些约束。例如,在3.2节中假定摄像机歪斜可以为任意值,而不是一个很接近0的值。本节要介绍的方法考虑了所有这些相关的约束,这个方法是从摄影地形测量法借鉴而来的。摄影地形测量法是一种从一幅或多幅图像定量恢复几何信息的方法,在测绘、军事、城市规划等很多领域都有应用。很多年来,摄影地形测量法综合运用几何、光学和数学方法来从图像恢复场景的三维信息,20世纪50年代计算机的出现使得这个问题更容易解决了。分析摄影地形测量学主要研究的是由摄像机内参数确定的内方向和外参数确定的外方向。

还是同样假定有n个参考点$P_i(i=1,\cdots,n)$,已知它们的世界坐标,优化的目标是使参考点的实际成像位置(u_i,v_i)与理论位置$(\tilde{u}_i,\tilde{v}_i)$的距离最小,估计值由投影方程得到,其中的摄像机参数向量$\boldsymbol{\xi}=(\xi_1,\cdots,\xi_q)^T(q\geqslant 11)$包含了摄像机内外参数和各种畸变的系数。最小方差可写成

$$E(\boldsymbol{\xi})=\sum_{i=1}^{n}[(\tilde{u}_i(\boldsymbol{\xi})-u_i)^2+(\tilde{v}_i(\boldsymbol{\xi})-v_i)^2]$$

其中,

$$\tilde{u}_i(\boldsymbol{\xi})\stackrel{\text{def}}{=}\frac{\boldsymbol{m}_1(\boldsymbol{\xi})\cdot\boldsymbol{P}_i}{\boldsymbol{m}_3(\boldsymbol{\xi})\cdot\boldsymbol{P}_i} \quad , \quad \tilde{v}_i(\boldsymbol{\xi})\stackrel{\text{def}}{=}\frac{\boldsymbol{m}_2(\boldsymbol{\xi})\cdot\boldsymbol{P}_i}{\boldsymbol{m}_3(\boldsymbol{\xi})\cdot\boldsymbol{P}_i}$$

与以往的情况不同,误差和各个参数$\boldsymbol{\xi}$的关系不是线性的,而是复杂的多项式与三角函数的组合关系。要优化这个误差函数,需要用到3.1.2节中介绍的非线性最小二乘法。沿用那一节使用的符号,误差函数可以写为

$$E(\boldsymbol{\xi}) = |\boldsymbol{f}(\boldsymbol{\xi})|^2 = \sum_{j=1}^{2n} f_j^2(\boldsymbol{\xi}),\ \text{其中,}\ \begin{cases} f_{2i-1}(\boldsymbol{\xi}) = \tilde{u}_i(\boldsymbol{\xi}) - u_i \\ f_{2i}(\boldsymbol{\xi}) \quad = \tilde{v}_i(\boldsymbol{\xi}) - v_i \end{cases},\quad i = 1, \cdots, n$$

3.1.2 节介绍的高斯牛顿法和 Levenberg-Marquard 方法需要用到函数 f_j 的导数，牛顿法不但需要导数，还需要 Hessian 矩阵。但是本节中介绍的方法只需要计算导数，即 Jacobian 矩阵。设 $\tilde{x}_i = \boldsymbol{m}_1 \cdot \boldsymbol{P}_i, \tilde{y}_i = \boldsymbol{m}_2 \cdot \boldsymbol{P}_i, \tilde{z}_i = \boldsymbol{m}_3 \cdot \boldsymbol{P}_i (i = 1, \cdots, n)$，有 $\tilde{u}_i = \tilde{x}_i / \tilde{z}_i$ 和 $\tilde{v}_i = \tilde{y}_i / \tilde{z}_i$。则

$$\begin{cases} \dfrac{\partial f_{2i-1}}{\partial \xi_j} = \dfrac{\partial \tilde{u}_i}{\partial \xi_j} = \dfrac{1}{\tilde{z}_i}\dfrac{\partial \tilde{x}_i}{\partial \xi_j} - \dfrac{\tilde{x}_i}{\tilde{z}_i^2}\dfrac{\partial \tilde{z}_i}{\partial \xi_j} = \dfrac{1}{\tilde{z}_i}\left(\dfrac{\partial}{\partial \xi_j}(\boldsymbol{m}_1 \cdot \boldsymbol{P}_i) - \tilde{u}_i \dfrac{\partial}{\partial \xi_j}(\boldsymbol{m}_3 \cdot \boldsymbol{P}_i)\right) \\ \dfrac{\partial f_{2i}}{\partial \xi_j} \quad = \dfrac{\partial \tilde{v}_i}{\partial \xi_j} = \dfrac{1}{\tilde{z}_i}\dfrac{\partial \tilde{y}_i}{\partial \xi_j} - \dfrac{\tilde{y}_i}{\tilde{z}_i^2}\dfrac{\partial \tilde{z}_i}{\partial \xi_j} = \dfrac{1}{\tilde{z}_i}\left(\dfrac{\partial}{\partial \xi_j}(\boldsymbol{m}_2 \cdot \boldsymbol{P}_i) - \tilde{v}_i \dfrac{\partial}{\partial \xi_j}(\boldsymbol{m}_3 \cdot \boldsymbol{P}_i)\right) \end{cases}$$

又可写为

$$\begin{pmatrix} \dfrac{\partial f_{2i-1}}{\partial \xi_j} \\ \dfrac{\partial f_{2i}}{\partial \xi_j} \end{pmatrix} = \frac{1}{\tilde{z}_i}\begin{pmatrix} \boldsymbol{P}_i^T & \boldsymbol{0}^T & -\tilde{u}_i \boldsymbol{P}_i^T \\ \boldsymbol{0}^T & \boldsymbol{P}_i^T & -\tilde{v}_i \boldsymbol{P}_i^T \end{pmatrix} \mathcal{J}\boldsymbol{m}$$

其中，$\boldsymbol{m}$ 是$\mathcal{M}$中的 12 维实数向量，$\mathcal{J}\boldsymbol{m}$ 是 $\boldsymbol{\zeta}$ 的 Jacobian 矩阵。最终得到 $\boldsymbol{f}$ 的 Jacobian 矩阵为

$$\mathcal{J}_{\boldsymbol{f}} = \begin{pmatrix} \dfrac{1}{\tilde{z}_1}\boldsymbol{P}_1^T & \boldsymbol{0}^T & -\dfrac{\tilde{u}_1}{\tilde{z}_1}\boldsymbol{P}_1^T \\ \boldsymbol{0}^T & \dfrac{1}{\tilde{z}_1}\boldsymbol{P}_1^T & -\dfrac{\tilde{v}_1}{\tilde{z}_1}\boldsymbol{P}_1^T \\ \cdots & \cdots & \cdots \\ \dfrac{1}{\tilde{z}_n}\boldsymbol{P}_n^T & \boldsymbol{0}^T & -\dfrac{\tilde{u}_n}{\tilde{z}_n}\boldsymbol{P}_n^T \\ \boldsymbol{0}^T & \dfrac{1}{\tilde{z}_n}\boldsymbol{P}_n^T & -\dfrac{\tilde{v}_n}{\tilde{z}_n}\boldsymbol{P}_n^T \end{pmatrix} \mathcal{J}\boldsymbol{m}$$

上式中的 $\tilde{u}_i, \tilde{v}_i, \tilde{z}_i$ 和 $\boldsymbol{P}_i$ 是由样本点决定的，但是$\mathcal{J}\boldsymbol{m}$ 只取决于摄像机的内外参数，这个方法需要旋转矩阵$\mathcal{R}$的显式表达。第 2 章介绍了一种这样的表达方法，习题和第 21 章中还将介绍其他的表示方法。

3.5　应用：机器人定位

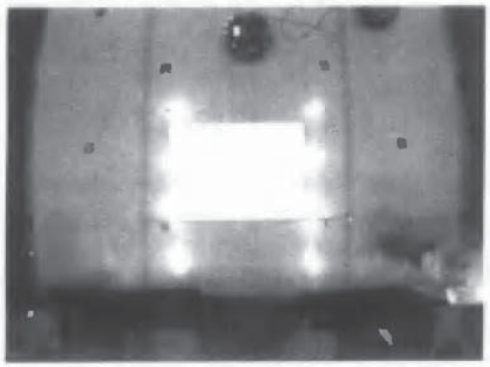
本章介绍的标定方法在很多系统中都有应用，从测量学到立体视觉，包括自动化中的物体定位。我们将简要介绍 Devy 等(1997)提出的非线性摄像机标定方法及其在自动机器人中的应用。与前面介绍的系统不同，这个方法使用了很多幅图像(20 帧)来标定位置固定的摄像机，参考点是一个平面上的网格点(见图 3.3)，其中一张图是把网格放在地面上照的，用来确定世界坐标系。在简单的手工初步粗定位后，用网格点附近的灰度参数模型，准确地找到每幅图像中的网格点，定位精确可以达到 1/10 像素。

图像的几何模型如 3.3 节所述，有三个参数表示径向畸变，没有歪斜。标定方法可以恢复一个摄像机内参数，并确定每幅图像的摄像机外参数。内参数的初始值可以从摄像机上的标识和采集卡的制造商处得到。外参数的初值可以用 Tsai 在 1987 年提出的方法改进得到。假设世界坐标系的 z 轴垂直于标定用的网格平面，然后用线性最小二乘法可以求得投影矩阵。

则式(3.20)变为

$$Q'\boldsymbol{n}' = 0, \quad 其中, \quad Q' = \begin{pmatrix} v_1x_1 & v_1y_1 & v_1 & -u_1x_1 & -u_1y_1 & -u_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ v_nx_n & v_ny_n & v_n & -u_nx_n & -u_1y_n & -u_n \end{pmatrix}$$

且 $\boldsymbol{n}' = (m_{11}, m_{12}, m_{14}, m_{21}, m_{22}, m_{24})^T$。



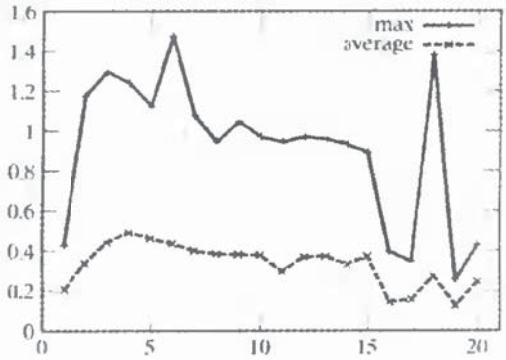


图 3.3 标定实验。左图:20 张输入图像中的一张;注意有很强的径向畸变。机器人在图像上方,有一个LED阵列;右图:20幅图像的平均和最大投影误差

注意,显式指定 $z_i = 0$ 可以避免上一章遇到的退化情况。因为内参数已知,知道 $\boldsymbol{n}'$ 后可以很容易得到外参数(见习题)。有了内参数和外参数的初始值后,就可以用非线性优化方法(这里用的是 Levenberg-Marquardt 方法)来优化观察值和预测值之间的均方误差。

实验中采用的硬件是 576×768 的摄像机和 4.5 mm 透镜,图 3.3 显示了网格点的成像位置与估计位置之间的误差。摄像机标定后,可以用来监视移动机器人在地面参考坐标系内的位置。每个机器人由不同的发光二极管标记(见图 3.3)。在定位实验中,摄像机前安装了一个过滤镜,可以过滤掉发光二极管外的所有光线。用简单的模式匹配方法就可以区分每个机器人,它们的位置和方向可以由发光二极管的位置和摄像机参数得到。摄像机距地面 4 m,定位误差平均为 2 cm,不超过 5 cm,方向误差平均为 1 度,不超过 5 度。

3.6 注释

3.2 节介绍的线性标定方法摘自 Faugeras(1993),引入径向畸变的改进方法由 Tsai(1987*a*)提出,Haralick 和 Shapiro(1992)同时提出了分析摄影地形测量法。手工相机测量是最标准的方法,其后的研究者(包括本书作者)研究的独创性理论和精确的方法都是原方法的不同描述(Thompson 等,1966,Slama 等,1980)。我们将在第 10 章介绍多幅图像的摄影地形测量法。Luenberger(1985)、Bertsekas(1995)和 Heath(2002)都介绍了通用的优化方法。Triggs 等(2000)发表了一篇关于最小二乘法在分析摄影地形测量法中应用的综述和讨论。若假设样本点符合正态分布,则最小二乘法的结果满足统计学上的最大相似度准则。我们将在第 15 章继续解释这个结论,还要重新分析平面内的直线拟合问题。

习题

3.1　证明，在 $|\mathcal{V}\boldsymbol{x}|^2=1$ 约束下，使 $|\mathcal{U}\boldsymbol{x}|^2$ 最小的向量 $\boldsymbol{x}$ 就是对称矩阵 $\mathcal{U}^T\mathcal{U}$ 和 $\mathcal{V}^T\mathcal{V}$ 上的最小广义特征值对应的特征向量。

提示：等价于下面的没有约束的最小化问题：求 $\boldsymbol{x}$ 使 $E(\boldsymbol{x})=|\mathcal{V}\boldsymbol{x}|^2/|\mathcal{V}\boldsymbol{x}|^2$ 最小。

3.2　证明在 3.1.1 节中得到的 2×2 矩阵 $\mathcal{U}^T\mathcal{U}$ 就是点集 $p_i(i=1,\cdots,n)$ 的惯性矩。

3.3　把 3.1.1 节中的直线拟合方法扩展到在 $\mathbb{E}^3$ 空间寻找最佳拟合平面。

3.4　推导下面两个式子的 Hessian 矩阵：$f_{2i-1}(\boldsymbol{\xi})=\tilde{u}_i(\boldsymbol{\xi})-u_i$，$f_{2i}(\boldsymbol{\xi})=\tilde{v}_i(\boldsymbol{\xi})-v_i(i=1,\cdots,n)$。

3.5　欧拉角是这样定义的：先绕 z 轴旋转 α，再绕 y 轴旋转 β，最后再绕 z 轴旋转 γ。

证明欧拉角可以在原坐标系中用下面的矩阵表示：

$$\begin{pmatrix} \cos\alpha\cos\beta\cos\gamma-\sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma-\sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma+\cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma+\cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{pmatrix}$$

3.6　证明 Rodrigues 公式。设 $\mathcal{R}$ 是绕 $\boldsymbol{u}$ 的旋转，转角为 θ，则

$$\mathcal{R}\boldsymbol{x}=\cos\theta\boldsymbol{x}+\sin\theta\boldsymbol{u}\times\boldsymbol{x}+(1-\cos\theta)(\boldsymbol{u}\cdot\boldsymbol{x})\boldsymbol{u}$$

提示：旋转不会改变向量在与旋转轴垂直的平面上的投影值。

3.7　利用 Rodrigues 定理证明，$\mathcal{R}$ 的坐标变换矩阵为

$$\begin{pmatrix} u^2(1-c)+c & uv(1-c)-ws & uw(1-c)+vs \\ uv(1-c)+ws & v^2(1-c)+c & vw(1-c)-us \\ uw(1-c)-vs & vw(1-c)+us & w^2(1-c)+c \end{pmatrix}$$

其中，$c=\cos\theta$ 和 $s=\sin\theta$。

3.8　若已知摄像机内参数，如何从 3.5 节中介绍的向量 $\boldsymbol{n}'$ 求摄像机外参数。

提示：旋转矩阵的各列都是单位向量。

3.9　假设用摄像机拍摄了 n 条 Plücker 坐标已知的刻线，

(a)证明若 $n\geqslant 9$，可以恢复第 2 章习题中的投影矩阵 $\tilde{\mathcal{M}}$；

(b)若 $\tilde{\mathcal{M}}$ 已知，则投影矩阵 $\mathcal{M}$ 也能用最小二乘法得到。

提示：$\mathcal{M}$ 的各列 $\boldsymbol{m}_i$ 分别表示三个平面 Π_i，而 $\tilde{\mathcal{M}}$ 的各行表示三条直线。利用线和平面间的关系可以写出 $\tilde{\boldsymbol{m}}_i$ 的约束。

编程作业

3.10　用最小二乘法在三维空间中找点集的最佳拟合平面。

3.11　在平面 $\mathbb{R}^2$ 内找点集 $(x_i,y_i)^T(i=1,\cdots,n)$ 的最小二乘拟合的圆锥曲线 $ax^2+bxy+cy^2+dx+ey+f=0$。

3.12　实现 3.2 节中介绍的线性标定算法。

3.13　实现 3.3 节中介绍的带径向畸变的标定算法。

3.14　实现 3.4 节中介绍的非线性标定算法。

第 4 章　辐射学——光亮度度量

这一章介绍一个对我们描述光的性能有用的词库。这一章中没有视觉的算法，但是其中的定义与概念对后续学习是很有用的。有些读者会感到本章的细节比自己想要的要多。为了使这些读者感到方便，表 4.1 对主要术语的简要定义给出了摘要。

4.1　空间中的光

对光进行度量是辐射学要研究的领域。为了描述能量如何从光源传输到表面块，以及能量到达表面时所发生的现象，我们需要一系列的单位。首先需要学习的是空间光的性质。

4.1.1　透视缩小效应

当光源相对于光线运行的方向倾斜时，对于面对这个光源的一块表面来说，它会显得偏小。同样对光源来说，如果被照射表面相对光的照射方向倾斜时，它也会显得偏小。这种现象称为透视缩小效应。

透视缩小效应是很重要的，因为远距离光源对表面的影响与从表面的角度观察光源所呈现的状态有关。我们可以以表面上与某一点有关的半球方向来考察这个很重要的事实（见图 4.1），因为只有这半个球的方向辐射到达该点。如果两个光源从每个入射方向到达某表面点的辐射总量完全相同，它们对该表面点的影响也就完全相同，因此从该表面观察它们是无法区分它们的。

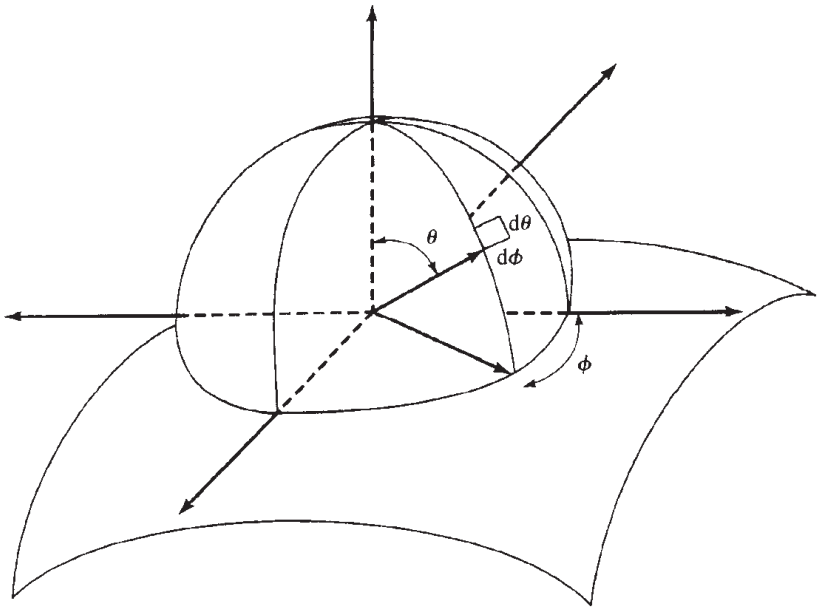


图 4.1　表面上一点沿着以它为中心的半球方向观察世界，表面的法线用来描绘半球的方向，给出 θ，ϕ 坐标系统。我们统一用该系统给出半球的角坐标。在辐射学的问题中，我们通常用从各方面来的辐射的总和来计算表面的亮度，因此尽管我们无法确定当 $\phi = 0$ 时的量，但这并不构成问题

这个结论对从光源角度看表面块，也同样适用。如果被光源照射的某块表面用另一块替代，并且这两块不同的表面对光源来说完全一样，也就是说从表面出发的每个方向以同样的性质影响其周围，那么光源也无法区分它们，正因为如此，能量以同样的传输率传给它们。

4.1.2　立体角

一个光源对输入半球所产生的效果可以用该光源对应的立体角描述。立体角的定义方法与平面上角度的定义相似。

平面上一根长度为 dl 的无穷小线段对应点 $\boldsymbol{p}$ 的角度可以通过将该线段在以点 $\boldsymbol{p}$ 为中心的单位圆上的投影来计算，投影长度就是所求的、以弧度表示的角度值(见图 4.2)。因为该线段无限短，它对应一个无限小的角度，该角度既取决于到圆心的距离，也与线的方向有关：

$$d\phi = \frac{dl\cos\theta_1}{r}$$

曲线对应的角度可以通过将其分解成无限小线段并求其总和(积分)得到。

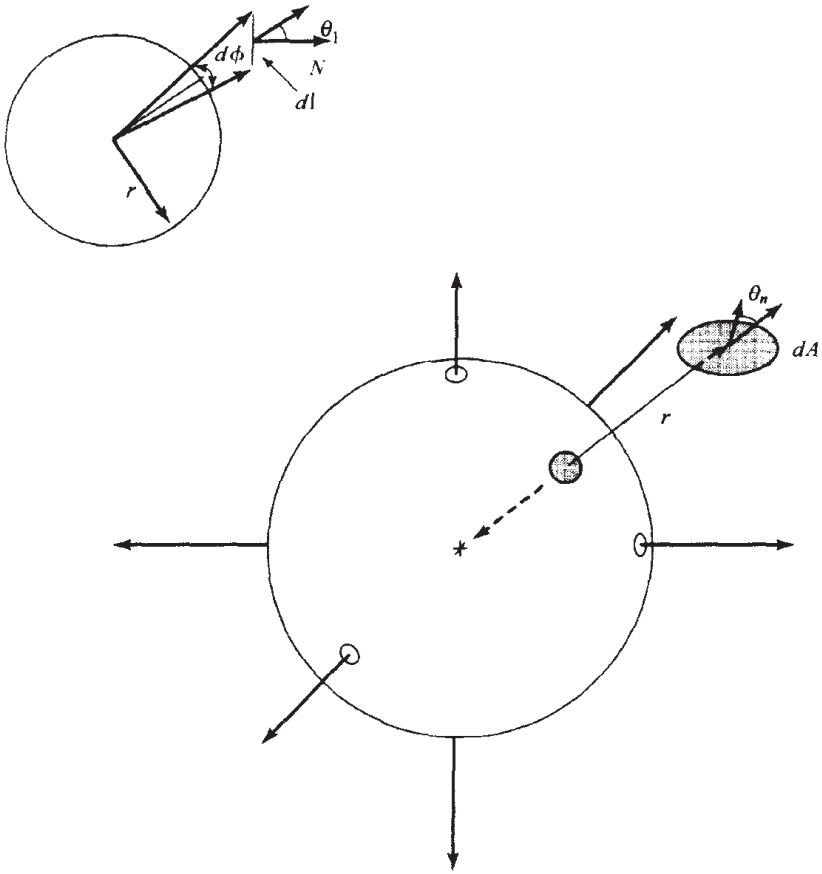


图 4.2　上图：一曲线段对应于某点的角度可通过将该曲线投影到以该点为中心的单位圆，并度量该投影的长度来计算。对一个小线段，这个角度为 $d\phi = (1/r)\,dl\cos\theta_1$；下图：说明立体角概念的球。围绕坐标轴的小圆用来帮助读者体会到该图是三维表面。一个无限小表面块投影到以某点为中心的单位球上，所得到的面积就是该表面块的立体角。此时表面块及其投影面积都很小，而立体角是 $(1/r^2)\,dA\cos\theta_n$

同样，点 $\boldsymbol{x}$ 对应于一表面块的立体角可通过将表面块在以为 $\boldsymbol{x}$ 中心的单位球上的投影计算，该投影的面积是所求的立体角，单位是球面度。通常用符号(ω)表示立体角。要指出的

是,立体角吸收了透视缩小效应的直观概念——在输入半球看来是相同的表面块对应相同的立体角。

如果表面块的面积 dA 很小(表示成无限小形式),那么它所对应的无限小立体角就很容易用表面块面积以及到它的距离计算:

$$d\omega = \frac{dA\cos\theta_n}{r^2}$$

其中,所使用的术语已在图 4.2 中给出。

立体角可以用常见的在球(如图 4.2 所示)上的角坐标方式表示。从图 4.1 以及圆弧长的表达式可得到,角 θ 与 ϕ 的无限小增量($d\theta, d\phi$)在球上切出的立体角区域为

$$d\omega = \sin\theta d\theta d\phi$$

上面两个式子都值得记住,因为它们在许多应用中是十分有用的。

4.1.3 辐射度

光在空间的分布是位置与方向的函数。例如,夜晚在一个空屋子里点亮一把狭窄光束的手电时,我们需要知道手电在哪里发光,以及它沿什么方向发光。这种照明的效果可以用一块无穷小表面接收到的功率表示,可想像成空间某点附近有一块具有某种朝向的表面。我们用**辐射度**定义:度量光的分布的合适单位是辐射度,它定义为在某点的单位立体角,垂直于传输方向的单位面积上沿传输方向传输的功率(单位时间的能量总量)

辐射度的单位是每平方米每球面度上的瓦特数($W \times m^{-2} \times sr^{-1}$)。辐射度的定义看起来可能有些奇怪,但是它与辐射学中最基本的现象是一致的:一个小面块从正面对着光源收集的能量,比该面块沿接近切线方向收集的要多,一面块从光源收集到的能量数量既取决于从面块看光源有多大,也取决于从光源看面块的面积有多大。需要记住的是,辐射度的单位中的平方米是符合透视缩小效应的(指与传输方向垂直算的面积)。

辐射度是位置与方向的函数(狭窄光束的手电筒是值得记住的好模型——人们可以移动手电筒的位置,改变光线的方向)。对给定的一个点 P 以及一个(可能是非单位的)向量 $\boldsymbol{v}$,点 P 在方向 $\boldsymbol{v}$ 的辐射度一般用 $L(P,\boldsymbol{v})$ 表示,点 P 可以处在自由空间或表面上。对后者有时用 $L(P,\theta,\phi)$ 表示点 P 在球坐标 θ 与 ϕ 方向的辐射度,该球坐标系统的轴取自表面的法线向量。

沿直线的辐射度是常数 对绝大多数重要的视觉问题,假定光线在传输过程中没有遇到介质是安全的,也就是说假设我们处在真空中。这样一来辐射度就具有一个十分理想的性质,即对空间两个点 P_1 与 P_2 来说(它们之间没有遮挡),从 P_1 离开到 P_2 方向去的辐射度,等于从 P_2 方向到达点 P_1 的辐射度。

下面的证明乍看起来似乎是不必要的,其实仔细研究一下是很值得的,因为它对许多其他计算问题十分关键。图 4.3 表示了一个表面沿某个方向发射光的情况。根据定义,如果在该面块的辐射度为 $L(P_1,\theta,\phi)$,那么从该面块在时间间隔 dt 内沿 θ,ϕ 方向的无限小立体角区域 $d\omega$ 发射的能量为

$$L(P_1,\theta,\phi)(\cos\theta_1 dA_1)(d\omega)(dt)$$

这是辐射度乘以经透视缩小效应后的面块面积,再乘以通过的立体角以及传输时间间隔得到的结果。

现在可以设想有两个面块，一个位于 P_1，面积为 dA_1，而另一个在 P_2，面积为 dA_2（见图 4.3）。为了避免与角度坐标混淆，将从 P_1 到 P_2 的方向写成 $\overrightarrow{P_1P_2}$，角度 θ_1 与 θ_2 的定义见图 4.3。

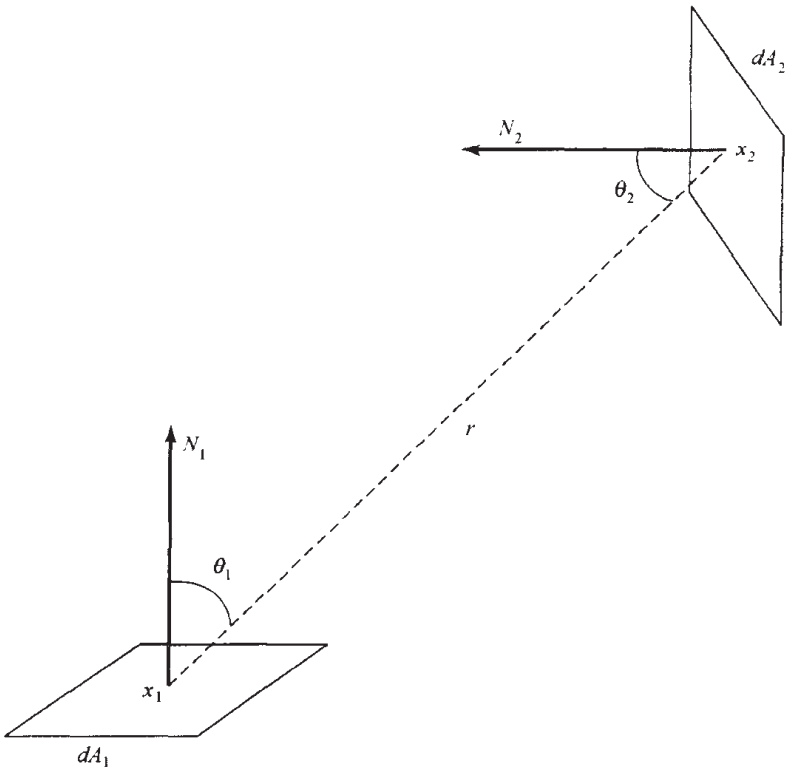


图 4.3　光的强度最好用辐射度来度量，因为在真空中或在清澈的空气中的合理距离内，沿直线路径传播的辐射度不会下降。这一点课文中已借助于从面块 dA_1 传输到面块 dA_2 的能量用能量守恒方式说明

从 P_1 离开朝 P_2 去的辐射度为 $L(P_1, \overrightarrow{P_1P_2})$，而沿 P_1 方向来到达 P_2 的辐射度是 $L(P_2, \overrightarrow{P_1P_2})$。

这意味着，在时间间隔 dt 中从 P_1 离开到达 P_2 的能量为

$$d^3E_{1\to 2} = L(P_1, \overrightarrow{P_1P_2})(\cos\theta_1 dA_1)(d\omega_{2(1)})(dt)$$

其中，$d\omega_{2(1)}$ 是面块 1 对应于面块 2 的立体角（发射到这个立体角范围内的能量到达面块 2，而其他的都消失到空间了）。符号 $d^3E_{1\to 2}$ 表示涉及三项无穷小量。由立体角表达式有

$$d\omega_{2(1)} = \frac{\cos\theta_2 dA_2}{r^2}$$

所以从 1 到达 2 的能量是

$$d^3E_{1\to 2} = L(P_1, \overrightarrow{P_1P_2})\left(\frac{\cos\theta_1\cos\theta_2}{r^2}\right)dA_2 dA_1\, dt$$

因为介质是真空的，它没有吸收能量，因此从 1 到达 2 的能量等于沿到达 2 方向离开 1 的能量。从 1 到达 2 的能量是

$$
\begin{aligned}
d^3E_{1\to 2} &= L(P_2, \overrightarrow{P_1P_2})(\cos\theta_2 dA_2)(d\omega_{1(2)})(dt) \\
&= L(P_2, \overrightarrow{P_1P_2})\left(\frac{\cos\theta_2\cos\theta_1}{r^2}\right)dA_1dA_2dt \\
&= d^3E_{1\to 2} \\
&= L(P_1, \overrightarrow{P_1P_2})\left(\frac{\cos\theta_2\cos\theta_1}{r^2}\right)dA_2dA_1dt
\end{aligned}
$$

这就意味着 $L(P_2, \overrightarrow{P_1P_2}) = L(P_1, \overrightarrow{P_1P_2})$，即沿(无遮挡的)直线传输的辐射度是常数。

4.2 到达表面的光

当光线到达一个表面时，它可能会被吸收、传输出去或散射，通常是这三种方式的组合。例如，到达皮肤的光可以在不同深度散射到细胞组织中去，或从血液中或其中的黑色素反射，也可被吸收，或者沿皮肤的油膜层切向散射，随后又在某处逸出。

这种情况会因为某些表面会吸收某种波长的光，再发射出另一波长的光而变得更复杂。这种现象称之为荧光，是很常见的：蝎子在 X 射线照射下会发出可见光荧光；人类牙齿在紫外线照射下会发出暗蓝色荧光(尼龙内衣也有荧光作用；而假牙则没有，这会带来不必要的尴尬，因而在跳舞的夜总会上不用紫外线)；洗衣业通过带荧光的洗衣粉使衣服在紫外光线中更明亮；一个表面加温到一定程度也会发出可见光。

4.2.1 简化假设

通常都设所有效果都是局部的，因此可用一个没有荧光和发射的宏观模型来解释现象，称这种模型为局部反应模型。对视觉中常见的表面类型与所要解决的问题来说，这是一个合理的模型。在这个模型中：

- 表面上离开某点的辐射度仅取决于到达该点的辐射度(尽管光线在该点会改变方向，但假定它不会从一点转移到另一点)；
- 假设离开一个表面给定波长的光是因为到达的光是这种波长；
- 假设表面不会从内部发出光以及分别对待不同的光源。

4.2.2 双向反射分布函数

我们需要描述输入照明源与反射光之间的关系，这是一个涉及入射光方向与反射光方向的函数。

辐照度 适用于表示入射功率的单位是辐照度，它的定义是表面单位面积(不计透视缩小效应)的入射功率。

按此定义，由从角度为(θ_i, ϕ_i)的微小立体角 $d\omega$ 处来的辐射度 $L_i(P, \theta_i, \phi_i)$所照射的面积为 dA 的表面受到的辐照度为

$$(1/dA)(L_i(P, \theta_i, \phi_i))(\cos\theta_i dA)d\omega = L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega$$

这表明，表面辐照度是通过将入射辐射度乘以透视缩小效率因素以及入射立体角度得到的。这个单位的主要特点是，表面上某点的入射功率可以通过从整个输入半球得到的辐照度总和

来计算，这使得它成为输入功率的适用单位。

BRDF　最通用的局部反射模型是双向反射分布函数，简写为 BRDF（bidirectional reflectance distribution function），它的定义是：输出方向的辐射度与输入方向的辐照度的比率。

因此，如果一个表面接受来自位于角度(θ_i, ϕ_i)的微小立体角 $d\omega$ 照射的辐射度为 $L_i(P, \theta_i, \phi_i)$，而它发出的辐射度为 $L_o(P, \theta_o, \phi_o)$，则它的 BRDF 是

$$\rho_{bd}(\theta_o, \phi_o, \theta_i, \phi_i) = \frac{L_o(P, \theta_o, \phi_o)}{L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega}$$

BRDF 单位的量纲是球面度的逆（sr^{-1}），它可以从零（在该方向没有反射光）变化至无穷大（输出方向的单位辐射度来自于输入方向任意小的辐射度）。BRDF 在输入与输出方向是对称的——这是著名的 Helmholtz 互易原理。

用 BRDF 计算从表面离开的辐射度　来自某一特定方向的辐照度中离开该表面的辐射度可以从 BRDF 的定义获得

$$L_o(P, \theta_o, \phi_o) = \rho_{bd}(\theta_o, \phi_o, \theta_i, \phi_i) L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega$$

从一个表面的入射辐照度（不管其入射方向）引起的输出辐射度更有意义，这可以从所有输入方向的作用求总和得到

$$L_o(P, \theta_o, \phi_o) = \int_\Omega \rho_{bd}(\theta_o, \phi_o, \theta_i, \phi_i) L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega$$

这里，Ω 指输入半球。

对 BRDF 的约束条件　BRDF 并不是一个具有 4 个变量的任意对称函数。为了弄清这一点，假设有一个面积为 dA 的表面块，受到 $L_i(P, \theta_i, \phi_i)$（$W \times m^{-2} \times sr^{-1}$）的照射，则在时间间隔 dt 中到达该表面的总能量是

$$\left(\int_{\Omega_i} L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega_i\right) dA\, dt = \left(\int_0^{2\pi}\int_0^{\frac{\pi}{2}} L_i(P, \theta_i, \phi_i)\cos\theta_i \sin\theta_i d\theta_i d\phi_i\right) dA\, dt$$

因为已经假设从表面一点离开的能量取自于到达该点的能量，并且表面内部没有产生能量，因此在该段时间内，从该表面离开的能量必须少于或等于到达的能量总量。故应有

$$\left(\int_{\Omega_i} L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega_i\right) dA\, dt \geqslant \left(\int_{\Omega_o} L_o(P, \theta_o, \phi_o)\cos\theta_o d\omega_o\right) dA\, dt$$

但是从表面离开的能量可写成

$$\left(\int_{\Omega_o}\int_{\Omega_i} \rho_{bd}(\theta_o, \phi_o, \theta_i, \phi_i) L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega_i \cos\theta_o d\omega_o\right) dA\, dt$$

这就是说对给定的 BRDF，不管 L_i 的具体值，必有

$$\int_{\Omega_i} L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega_i \geqslant \int_{\Omega_o}\int_{\Omega_i} \rho_{bd}(\theta_o, \phi_o, \theta_i, \phi_i) L_i(P, \theta_i, \phi_i)\cos\theta_i d\omega_i \cos\theta_o d\omega_o$$

因此，尽管对某些输入与输出角度，BRDF 可以较大，但对大多数来说它不能大。事实上平均值必须相当小。这个事实可以用来证明：对一个 BRDF，它的最大值是 $1/\pi$，并且与角度无关。

4.2.3　例子：薄透镜的辐射度学

为了说明迄今为止已经介绍的某些概念，我们来讨论用一个薄透镜将中心为点 P 的景物

片断δA发出的光线，集中到以P'为中心的图像块$\delta A'$上的情况（见图4.4），分析点P物体的辐射度L与点P'图像辐照度E之间的关系①。如果用$\delta\omega$表示从透镜中心O对应δA（或$\delta A'$）的立体角，则有

$$\delta\omega = \frac{\delta A'\cos\alpha}{(z'/\cos\alpha)^2} = \frac{\delta A\cos\beta}{(z/\cos\alpha)^2}，于是有\ \frac{\delta A}{\delta A'} = \frac{\cos\alpha}{\cos\beta}\left(\frac{z}{z'}\right)^2$$

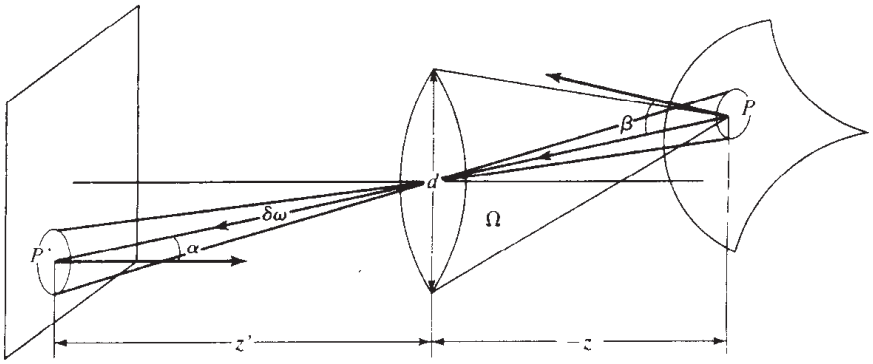


图4.4 物体辐射度与图像的辐照度

由于直径为d的透镜面积为$\frac{\pi}{4}d^2$，加上用Ω表示P对应于透镜的立体角，则有

$$\Omega = \frac{\pi}{4}\frac{d^2\cos\alpha}{(z/\cos\alpha)^2} = \frac{\pi}{4}\left(\frac{d}{z}\right)^2\cos^3\alpha$$

从δA发出并投到透镜的功率δP可以用物体辐射度L表示成

$$\delta P = L\Omega\delta A\cos\beta = \frac{\pi}{4}\left(\frac{d}{z}\right)^2 L\delta A\cos^3\alpha\cos\beta$$

功率δP被透镜会聚到图像平面的表面块$\delta A'$，这是到达该块仅有的功率。因此图像的辐照度为

$$E = \frac{\delta P}{\delta A'} = \frac{\pi}{4}\left(\frac{d}{z}\right)^2 L\frac{\delta A}{\delta A'}\cos^3\alpha\cos\beta$$

将$\delta A/\delta A'$的值代入上式，得到

$$E = \left[\frac{\pi}{4}\left(\frac{d}{z'}\right)^2\cos^4\alpha\right]L \tag{4.1}$$

以下几点使上述关系值得关注：首先，它表明图像辐照度与物体的辐射度成正比，换句话说，我们度量的（E）与我们关注的（L）成正比。其次照射度与透镜面积成正比，与透镜中心到图像平面的距离成反比。$a = d/f$是相对孔径，是早先定义的数的倒数。式（4.1）表明当透镜聚焦到无限远（此时$z' = f$）时，E与a^2成正比。最后辐射度与$\cos^4\alpha$成正比，并且它随光线偏

① 此处为了简洁，省略了光通量与辐照度的参数，请记住，L表示点P沿P'方向的辐射度。

离光轴下降。在 α 值较小时，人们以及图像分析算法几乎注意不到这种现象，因为前者对平滑亮度梯度十分不敏感，而后者通常对渐晕效应更加敏感，而这种现象在大多数成像过程中与 $\cos^4\alpha$ 的下降相比占主导地位。

4.3 重要的特殊情况

辐射度是一个相当精细的量，因为它与角度有关。这种一般性经常是很主要的。例如手电筒这个例子中所描述的光在空间中的分布。作为另一个例子，可以考虑用手电筒光照射一个光盘的情况，可以看到它表面反射的光的强度和色彩，是与观察表面的视角以及照射角度密切相关的。这个照射光盘的例子是值得试一试的，因为它展示了反射表面会有怎样奇特的性能；它也表明我们对那些没有这种性质的表面是如此地习惯。对许多表面来说，反射光对角度是不敏感甚至是无关的，棉布就是一个例子，因此需要与角度无关的单位系统。

4.3.1 光通量

如果离开表面的辐射度与输出角无关，那么使用一个明显与方向有关的单位是没有意义的。一个合适的单位是光通量(radiosity)，其定义为：表面上一点单位面积发出的总能量。

光通量通常写成 $B(P)$，单位是瓦特数每平方米($W \times m^{-2}$)。为了获取表面某点的光通量，可以将该点在整个输出半球发出的辐射度求和。因此，如果 P 是表面上的一个点，它发射的辐射度为 $L(P,\theta,\phi)$，则在该点的光通量为

$$B(P) = \int_{\Omega} L(P,\theta,\phi)\cos\theta d\omega$$

其中，Ω 是输出半球，$\cos\theta$ 项反映因透视缩小效应对面积的影响，$d\omega$ 可像以前一样用与 θ,ϕ 有关的项等表示。

有恒定辐射度表面的光通量　对辐射度与角度无关的一个表面，其光通量与辐射度的关系是要记住的一个结果。在这种情况下，$L_o(x,\theta_o,\phi_o) = L_o(P)$，那么光通量可以通过将所有输出方向离开该表面的辐射度求和得到

$$\begin{aligned} B(P) &= \int_{\Omega} L_o(P)\cos\theta d\omega \\ &= L_o(P)\int_0^{\frac{\pi}{2}}\int_0^{2\pi}\cos\theta\sin\theta d\phi d\theta \\ &= \pi L_o(P) \end{aligned}$$

4.3.2 方向性半球反射

双向反射分布函数 BRDF 也是一个精细的量，对它的度量一般也是困难的、昂贵的并且重复性不好。这是由于表面尘土与老化过程对 BRDF 度量会产生明显的影响。例如，接触一个表面会将油脂传给它，这通常会形成一个小的隆起，产生类似透镜的作用，以至于使表面的定向性能有显著改变。

对许多表面来说，输出的光往往与输出角度无关。在这种情况下，对这种表面反射特性的

一种自然的度量是方向性半球反射率。通常写成 ρ_{dh}，它定义为：从某给定方向入射的辐照度中被表面所有方向反射出的部分。

表面方向性半球反射率可通过表面所有方向射出的辐射度之和除以从照明方向上输入的辐照度得到：

$$\begin{aligned}\rho_{dh}(\theta_i,\phi_i) &= \frac{\int_\Omega L_o(P,\theta_o,\phi_o)\cos\theta_o d\omega_o}{L_i(P,\theta_i,\phi_i)\cos\theta_i d\omega_i} \\ &= \int_\Omega \left\{\frac{L_o(P,\theta_o,\phi_o)\cos\theta_o}{L_i(P,\theta_i,\phi_i)\cos\theta_i d\omega_i}\right\} d\omega_o \\ &= \int_\Omega \rho_{bd}(\theta_o,\phi_o,\theta_i,\phi_i)\cos\theta_o d\omega_o\end{aligned}$$

这种性质是无量纲的，它的值在 0 到 1 之间。

对任何表面都可计算方向性半球反射率。对某些表面来说，它随入射方向剧烈变化。带有细微的对称三角形纹路的表面就是一个很好的例子，它的一面是黑的，而另一面是白的。如果这些纹路足够细，则从宏观描述来看表面是平坦的。当指向白的一面时，它的方向性半球反射率是大的，但若指向黑的一面，则是小的。

4.3.3 朗伯表面和漫反射系数

对某些表面来说，方向性半球反射率与照明方向无关，这种表面的例子包括棉布、多种地毯、无光纸和无光涂料等。这种表面的正式定义是它的 BRDF 与输出方向无关（基于互逆定理，也与输入方向无关），这意味着表面输出的辐射度与角度无关。这种表面称为理想漫散表面或朗伯表面（以 Johan Lambert 命名，它是第一个提出这个概念的）。

显然，使用光通量作为单位来描述朗伯表面发出的能量是很自然的。对朗伯表面来说，方向性半球反射率与方向无关。在这种情况下，经常称方向性半球反射率为漫射反射率或漫射系数，写成 ρ_d。对一个 BRDF 为 $\rho_{bd}(\theta_o,\phi_o,\theta_i,\phi_i)=\rho$ 的朗伯表面，可得到

$$\begin{aligned}\rho_d &= \int_\Omega \rho_{bd}(\theta_o,\phi_o,\theta_i,\phi_i)\cos\theta_o d\omega_o \\ &= \int_\Omega \rho\cos\theta_o d\omega_o \\ &= \rho\int_0^{\frac{\pi}{2}}\int_0^{2\pi}\cos\theta_o\sin\theta_o d\theta_o d\phi_o \\ &= \pi\rho\end{aligned}$$

而更常用的形式是

$$\rho_{brdf} = \frac{\rho_d}{\pi}$$

这个式子是很有用的，值得记住。

由于我们对明亮程度的感觉粗略地对应于辐射度的度量值，因而朗伯表面从任何方向看起来都具有同样亮度，不论照明是何种方向。这可作为测试表面是否近似于朗伯表面的粗略方法。

4.3.4　镜面反射表面

第二种重要的表面类型是玻璃或像镜子一样的表面，通常称为镜面反射表面（源自拉丁语 speculum，镜子）。理想的镜面反射器的性能像一面理想的镜子，某特定方向的入射光只能向一个镜面反射方向反射，从与入射光方向相反的方向射出。通常一部分入射的辐射被吸收了，对一个理想镜面反射表面来说，任何方向吸收入射光的比例都是相同的，而未被吸收部分沿镜面反射方向射出。理想的镜面反射表面的 BRDF 具有奇特的形式（练习），因为某一方向的入射光只能从一个方向射出。

镜面反射带　能够近似为理想镜面反射器的表面是很少的。看一个平坦表面能否近似为理想镜面反射器，可用它能否确实起到镜子的作用来检验。在过去，要制造出好的镜子是相当困难的，一般通过用抛光金属来造镜子。除非金属被高度抛光并很好地维护，否则某个方向射入的入射光通常会沿着反射方向周围的一小束方向反射出去，这会导致典型的模糊效果。平坦的馅饼金属锅底就是一个很好的例子。如果锅底比较新，你可以看到在表面有你自己变形的脸，但当镜子用就很勉强，而磨损后的锅底只能反射部分扭曲了的模糊图像。

有较大的镜面反射瓣，就意味着反射图像变形更加严重，反射的光线也不很明亮（因为入射光强被分散到一组反射方向上），往往只能看到相对较亮的物体，如光源等的镜面反射。因此在亮光涂料或塑料表面，人们看到的是沿光源镜面反射方向的明亮光团，经常称它为光斑（specularity），而几乎没有其他镜面反射效应。通常不一定要为光瓣的形状建模。如果要对光瓣建模，常用的模型是 Phong 模型。它假定镜面反射的是点光源（见图 4.5）。在这个模型中，从镜面反射表面反射出的光强正比于 $\cos^n(\delta\theta) = \cos^n(\theta_o - \theta_s)$，其中 θ_o 是射出角度，θ_s 是镜面反射方向，而 θ_n 是一个参数。大的 n 值对应狭窄光瓣与锐利的小光斑；而小的值导致宽广的光瓣及边界略显模糊的大光斑。

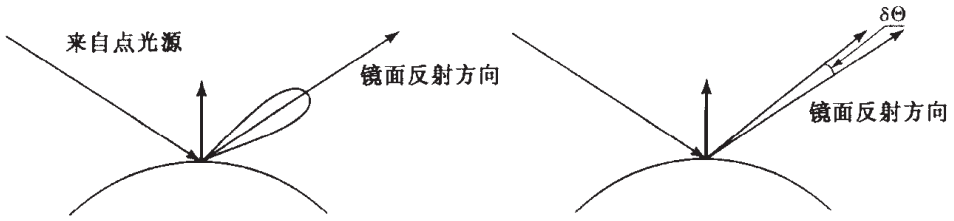


图 4.5　镜面反射表面通常围绕镜面反射方向小邻域的方向反射光，反射光强度如左图所示，与方向有关，Phong模型可用来描述光瓣形状与偏离镜面反射方向角之间的关系

4.3.5　朗伯 + 镜面模型

理想的漫反射或完美的镜面反射表面是相对较少的。许多表面的双向反射分布函数可以用朗伯成分与具有狭窄光瓣的镜面反射成分的结合体来近似表示。通常镜面反射成分用一个镜面反射率来加权。由于我们并不打算仔细审视这一镜面反射光斑，而对光瓣的形状未加关注。在这种情况下，某给定方向的表面的辐射度（此时它与方向有关）一般近似为

$$L(P,\theta_o,\phi_o) = \rho_d(P)\int_\Omega L(P,\theta_i,\phi_i)\cos\theta_i d\omega + \rho_s(P)L(P,\theta_s,\phi_s)\cos^n(\theta_s-\theta_o)$$

其中，θ_s，ϕ_s 描述镜面反射方向，ρ_s 是镜面反射率。我们很快会注意到，一般情况下不大理会镜面反射项的精确幅值。

使用这种模型隐含着将“太窄”的镜面反射光瓣排除在外，因为多数算法会遭遇到来自光源偶发的小的狭窄镜面反射光斑。具有太窄镜面小瓣（镜子）的表面会产生光斑过多的细节，同样也不考虑“太宽”的光瓣，因为此时很难辨认出镜面反射光斑。

表 4.1 参考卡：辐射学术语

话 题	要 点
透视缩小效应	沿擦地角度看的大面积与正面看的小面积相当，这是十分重要的，因为如果两个不同的接收器从光源来看是一样的话，它们将接收到相同的辐射，两个不同的光源从接收器角度看是相同的话，它们对接收器的效果是相同的
辐射度（Radiance）	**定义**：单位时间内沿某个方向，在与其垂直的单位面积、单位立体角上传输到某点的能量；**单位**：$W \times m^{-2} \times sr^{-1}$；**用途**：表达在自由空间传输的光，或表达从表面反射出来与方向密切相关的光
辐照度（Irradiance）	**定义**：单位表面积上的入射总功率；**单位**：$W \times m^{-2}$。**用途**：表示到达一个表面的光强
光通量（Radiosity）	**定义**：从表面某点射出的单位表面积的总功率；**单位**：$W \times m^{-2}$；**用途**：表达从漫反射表面反射的光强度
双向反射分布函数（BRDF）	**定义**：沿反射方向反射光强与入射光之比值；**单位**：sr^{-1}；**用途**：表示与方向密切相关反射的一般表面的反射特性
方向性半球反射	**定义**：从某给定方向上入射光被表面向所有方向反射的部分；**单位**：无量纲；**用途**：表示反射方向不重要的表面的反射特性
反射率	**定义**：漫反射表面的方向性半球反射率；**单位**：无量纲；**用途**：描述漫反射表面属性
漫反射表面； 朗伯表面	**定义**：双向反射分布函数是常数的表面；**例子**：棉布、涂料与纸张、表观亮度与观察角度无关的表面
镜面反射表面	**定义**：性能像镜子的表面；**例子**：镜子、抛光金属
光斑	**定义**：由双向反射分布函数中的镜面反射成分导致的表面小亮点。**例子**：各种塑料表面，抛光金属表面，某些涂料以及发光的衣料；**重要特征**：当你的头摆动时，光斑在表面也改变位置

4.4 注释

我们强烈推荐 François Sillion 所著的一本很杰出的书（Sillion 1994），它清晰地分析了辐射度计算。有许多更为详细的出版物可供参考（Nayar，Ikeuchi 与 Kanade，1991*b*）。我们对反射问题的讨论是很肤浅的，镜面反射加漫反射的模型源于 Cook，Torrance 与 Sparrow（Torrance 与 Sparrow，1967，Cook 与 Torrance，1987）。在计算机视觉与计算机图形学界对这个模型有不少修改版本。反射模型可以通过将表面粗糙度的统计描述与电磁条件结合起来（如 Beckmann 与 Spizzichino，1987）或采用散射模型（如 Torrance 与 Sparrow 的工作 1967，和 Cook 与 Torrance，1987 的工作）而获得。

在没有讨论的光学效应目录上，处于第一位的是偏离镜面的反射，紧随其后的是镜面反向散射。偏离镜面的反射通常发生在擦亮表面，这种表面上相当部分的面积与宏观表面法线之间有明显的角度，这就导致了由这些区域产生的第二个镜面光瓣。如果要求从镜面反射精确

推断表面形状，这种效果会给从镜面反射推断表面形状的算法带来麻烦。镜面反向散射发生在表面将光线沿光源方向反射，一般来说，产生这种现象的原因与偏离镜面的反射相似。这种效果同样会给利用镜面反射推断表面形状的算法带来麻烦。将这些特性考虑进去的某些反射模型在 Tagare 与 de Figueiredo(1991)的书中说明。

一般人都认为粗糙表面是朗伯表面。这种信念具有明显的主观臆想成分，因为粗糙表面经常出现局部阴影效应，它使得反射出的光强与照明角度密切相关。例如，用接近切向角度的光照射一堵抹灰墙，会在墙上显示出清晰的光源形状以及黑暗区域，这是因为表面的微小面块或者正对光源，或者被阴影覆盖。如果沿法线方向照射同一堵墙，这些图像一般是看不到的。在较细尺度上类似效应的平均效果使粗糙表面与朗伯表面有明显差异(详细情况可参考 Koenderink，van Doorn，Dana 和 Nayar，1999，Nayar 和 Oren，1993，1995，Oren 与 Nayar，1995 以及 Wolff，Nayar 和 Oren，1998)。

另一个不支持简单宏观表面模型的例子是开满鲜花的田野。一个在远处的观察者可将该田野抽象成“表面”；然而这种抽象会使这种表面具有很奇特的性质。如果人们沿法线方向观察这个田野，看到的主要是花卉，但从切线方向看过去会发现有杆与花，这意味着色彩会随观察方向剧烈改变(这种效果是 Leung 与 Malik 于 1997 年发现的)。

习题

4.1　半球的球面角度是多少？

4.2　已经证明在非吸收介质中，沿直线传播的辐射度不会减小，这使得它成为一个有用的单位。如果采用每平方米的透视变小面积上的功率(这是辐照度)，这个单位会随直线上的距离改变而改变。这种差异有怎样的意义？

4.3　**一个吸收型介质**：假设这个世界充满着一种迷向性吸收性介质。一种既好又简单的介质模型可以通过考虑沿一条直线传输的辐射度获得。如果沿线点 x 的辐射度是 N，在 $x+dx$ 处为 $N-(\alpha dx)N$。

(a) 写出在这种介质中从一表面块传播到另一表面块的辐射度的表达式；

(b) 定性地描述一下充满这种介质的房间内的光的分布，α 分别取小的与大的正值；该房间是一立方体，光源是位于天花板中心的单个小块光源。要记住，当 α 是大的正值时，实际上没有光到达房间的墙壁。

4.4　对针孔直径为 d 的针孔照相机，推导出景物辐射度与图像辐照度之间的关系。

4.5　对针孔直径为 d 的球形照相机，写出景物辐射度与图像辐照度之间的关系。

4.6　用 CD 的底面作为例子来辨认普通表面既不是朗伯表面也不是纯镜面反射表面。有许多生物学例子的颜色是蓝的。举出两个不同的理由，说明一有机物具有非朗伯表面是有好处的。

4.7　证明一个理想漫反射表面的方向性半球反射是一个常数，证明如果一个表面的方向性半球反射是常数，则它是理想漫反射的。

4.8　证明理想镜面反射表面的双向反射分布函数是

$$\rho_{bd}(\theta_o,\phi_o,\theta_i,\phi_i)=\rho_s(\theta_i)\{2\delta(\sin^2\theta_o-\sin^2\theta_i)\}\{\delta(\phi_o-\phi\pi)\}$$

其中，$\rho_s(\theta_i)$是离开表面的部分辐射。

4.9 为什么光斑要比漫反射明亮?

4.10 一个表面具有固定的双向反射分布函数,该常数可能的最大值是什么?假设已知该表面吸收入射光的20%(其余的反射出来),那么此时双向反射分布函数的值是什么?

4.11 眼睛对辐射度值产生响应。解释为什么经常把朗伯表面看做是具有亮度与观察角度无关性质的表面。

4.12 证明一个半径为ϵ的球对应于距球心 r 处的点的立体角近似为$\pi\left(\frac{\epsilon}{r}\right)^2$,设 $r>>\epsilon$。

第5章　光源、阴影与影调

物体的表面呈现或亮或暗，其影响因素有两个：它们的反射率以及接收到的光的总量。一个用来描述表面光亮度产生机理的模型通常称为影调模型。这个模型是重要的，因为使用一个适当的影调模型可以解释像素的值。如果使用了正确的影调模型，就有可能只从几幅图来重构物体与它们的反射率。此外，我们还能够解释阴影，并且解释它们令人迷惑的现象以及在室内景物几乎总能见到的现象。

5.1　定性辐射学

我们一定很想知道在各种光照条件下表面会有多“亮”，想知道“亮度”与局部表面属性、表面形状以及照明之间的关系。正如第4章提到的，透视缩小效应意味着不同光源对表面可以产生同样的效果，分析这个问题的最强有力工具是从表面的角度考虑光源是什么样子的。定性辐射学是一种利用这种技巧的一种技术，它并不很复杂，没有很难的数学知识，但是非常有效。在有些场合，这种技术对“亮度”给出定性的描述，但很可能甚至不知道这个词的意义。

回顾4.1.1节以及图4.1提到的现象，即一个表面块是通过其半球方向来看世界的。从某一方向到达该表面的辐射穿过该半球的一个点。如果两个表面块具有等价的输入半球，它们就具有相同的输入辐射量，而不用考虑外部世界的具体情况。这也意味着，具有相同输入半球的表面块之间的亮度差异是表面属性不同造成的。特别要指出的是，如果两个具有同样双向反射分布函数的表面块具有相同的输入半球，那么它们输出的辐射量也一定相同。

朗伯曾明确指出了在阴天环境下，处于无限高的墙底部的一个均匀一致平面上“亮度”的分布情况（见图5.1）。在这种情况下，平面上每一点具有同样的半球，它的半个观察球被墙挡住了，而另外半个见到了天空。天空是均匀的，平面也是均匀的，所以每个点必定有相同的“亮度”。

第二个例子更巧妙一些。在一个无限大平面上有一堵无限薄的黑墙，它在一个方向上无限长（见图5.2），我们希望知道具有等“亮度”的曲线的定性描述。不难想像图5.2上通过某点的所有线上的所有点具有相同的输入半球，所以具有相同的“亮度”。除此之外，平面上“亮度”分布还应该以墙脚线为对称轴，因此可以预期最亮的点是沿墙脚线的延伸方向，而最暗的则在墙的底部。

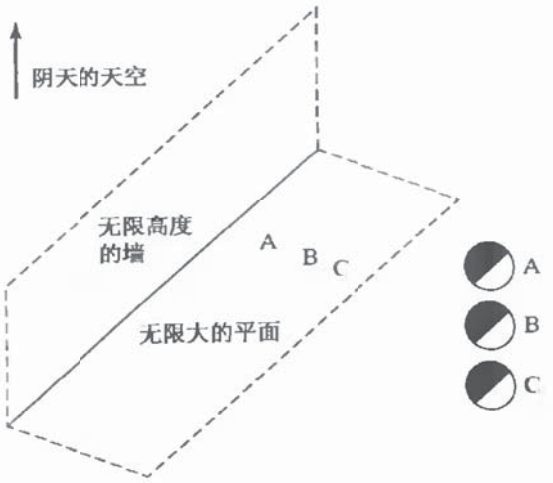


图 5.1 一个可以说明从表面块角度观察世界而获得定性辐射学的几何设置。我们想知道处在无限高度的墙底部两个不同点的亮度。在这个几何设置中，一堵无限高度的无光泽的黑墙挡住了阴天天空的视角，阴天天空是一个无穷大半径具有均匀"亮度"的半球。图的右半部展示了相应点看得见或看不见该光源的方向，这是通过将半球沿一个圆周的方向展平才得到的(或者与从上面看下来的说法等价)。因为每个点都有相同的输入半球，它们的亮度是一致的

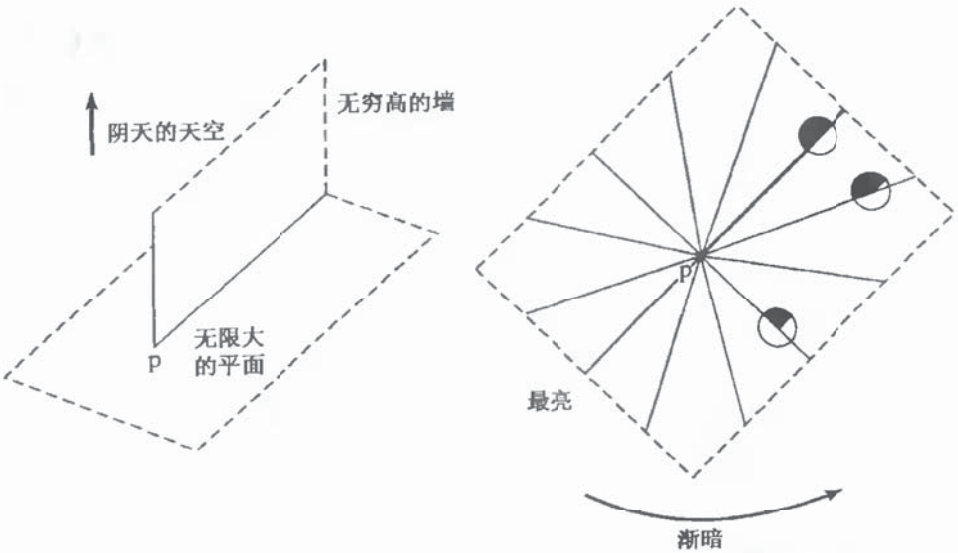


图 5.2 在一个无限的白平面(显示于左图)上有一堵无光泽的黑墙，它无限薄，一半是无限延伸的。这个几何设置也处在无穷半径，均匀"亮度"的阴天照明环境下。文中指出了如何确定平面上具有相似"亮度"的线。这些线用一个对平面俯视的方式，显示于右图，粗线表示墙。叠加在某些线上的是某些等照度线上的输入半球表示。半球沿着这些线是固定不变的(与几何因素有关)，但它们从线到线是变化的

5.2 光源及其产生的效果

5.2.1 光源的辐射学属性

我们将光源定义成将其内部产生的光发射出来的任何东西(也就是说不是反射光源)。为了描述一个光源，需要对它向每个方向射出光通量进行描述。通常由内部产生的辐射度与反

射辐射度是分别对待的。这是因为尽管一个光源可以反射光，但它反射出的光取决于环境，而内部产生的光源则独立于环境。

一般说来，要求对光源向各个方向射出的辐射度进行完整描述是很少出现的，通常将光源描述成向每个方向发出相同的辐射度（可能某些方向是零，就像聚光灯那样），此时一个合适的量是发射度，它的定义是：发光表面单位面积单位时间内所发射出的能量，能量是内部产生的。

发射度与光通量是相似的，可以按下式计算

$$E(P)=\int_{\Omega} L_e(P,\theta_o,\phi_o)\cos\theta_o\, d\omega$$

除了描述光源的发射度外，还需要描述光源的几何形状，它对光源周围的光在空间变化以及接近光源的物体产生的阴影有很重要的作用。光源一般使用十分简单的几何表示，这有两个理由：第一，许多人造光源可以用点光源、线光源或面光源有效地表示；第二，简单几何形状的光源仍然能产生令人吃惊的复杂效果。

5.2.2　点光源

经常将光源用一个非常小的球，实际上是一个点来近似，这种光源称为点光源。使用这种模型是很自然的，因为许多光源的大小与它们所处的环境相比是很小的。点光源产生的效果，可以用在一个很小的球面上每个点都发出等量的光来描述。

假设有一个表面块在距离一个半径为ϵ的球 r 处远，并且有$\epsilon \ll r$（见图 5.3）。对实际光源经常可以使用相对其半径而言球远离表面的假设。此时光源覆盖的立体角是 Ω_s，它可近似正比于

$$\frac{\epsilon^2}{r^2}$$

光源产生在这个半球上的照明分布也会成比例地变化。当球逐渐离开时，从表面块射向球体的射线趋于集中，而整个光通量只略微有变化（有一小组新的射线加入到球的边缘处，但这些射线产生的影响是较小的，因为它们来自于与球相切的方向）。在极限情况趋向于零时，就没有新的射线加入了。

光源产生的光通量可以通过将光源产生的分布乘以 $\cos\theta_i$ 并积分求得（译者注：θ_i 是光源球面上的面块射线的角度）。当ϵ趋于零时，输入半球上所照射的面块缩小，而 $\cos\theta_i$ 也接近常数。如果 ρ 是表面的反射率，则由点光源产生的光通量是

$$\rho\left(\frac{\epsilon}{r}\right)^2 E\cos\theta$$

其中，E 是有关光源对一小块发射度的积分。对一般情况不需要详细的表达式 E（若要求这种表达式，则必须求这个积分，而这正是我们故意回避的）。

近点光源　要获得近点光源的标准模型，需要用到表面单位法线向量 $N(P)$与从 P 到光源的向量 $S(P)$，它的长度为$\epsilon^2 E$ 。此时有

$$\rho_d(P)\frac{N(P)\cdot S(P)}{r(P)^2}$$

这是一个很方便的模型，因为它给出了光通量与形状（法线向量）之间的显式关系，其中 S 一般称为光源向量。通常在这个模型中忽略到光源距离的依赖关系（这些是不正确的）。

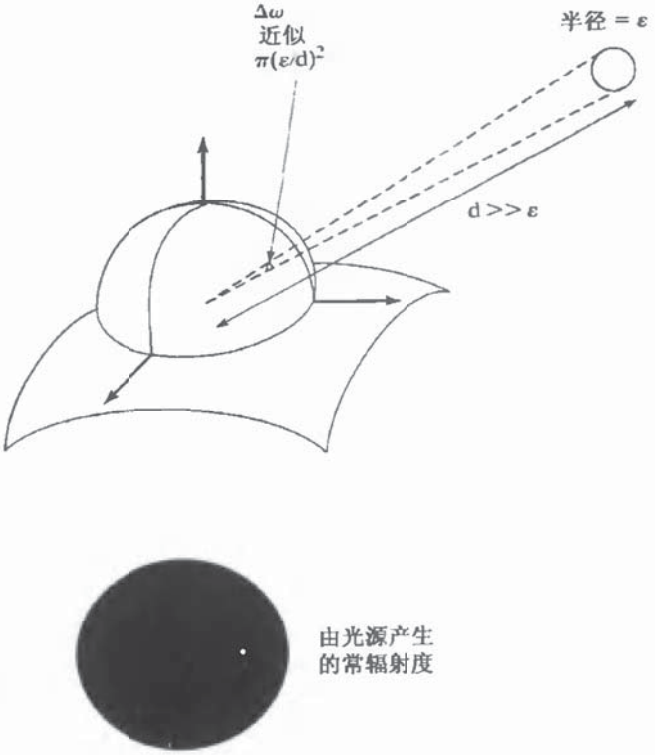


图 5.3 一个表面块注视着具有小半径、远距离的球；该球在表面输入半径上产生了一个被照亮的小块。文中通过分析远距离的球的运动，对这小块尺度造成的影响获得点光源性能的描述

无穷远处的点光源 太阳是很远的，因此 $1/r(P)^2$ 与 $\boldsymbol{S}(P)$ 实质上是个常数，这种情况的光源称为无穷远处的点光源。如果所关注的所有表面块与它们到光源的距离相比是紧靠在一起的，则 $r(P)=r_0+\Delta r(P)$，其中 $r_0 \gg \Delta r(P)$；同时 $\boldsymbol{S}(P)=\boldsymbol{S}_0+\Delta\boldsymbol{S}(P)$，其中 $|\boldsymbol{S}_0| \gg |\Delta\boldsymbol{S}(P)|$，此时有

$$\frac{\boldsymbol{N}\cdot\boldsymbol{S}(P)}{r(P)^2}=\frac{\boldsymbol{N}\cdot(\boldsymbol{S}_0+\Delta\boldsymbol{S}(P))}{(r_0+\Delta r(P))^2}\approx\frac{\boldsymbol{N}\cdot\boldsymbol{S}_0}{r_0^2}$$

由于 $\boldsymbol{S}_0$ 与 r_0 是常数，因此在表示式中将它们明显地表示出来没什么必要。这样一来，我们可以用 $\boldsymbol{S}=(1/r_0^2)\boldsymbol{S}_0$ 来取代，而由无限远的点光源带来的光通量成为

$$B(P)=\rho_d(P)(\boldsymbol{N}\cdot\boldsymbol{S})$$

符号 $\boldsymbol{S}$ 再一次称为光源向量。这个模型的典型应用是推理一个合适的光源向量，而不是从光源的发射度和几何因素计算它的数值。

选择一个点光源模型 处于无限远的点光源是一个好的模型，可以太阳作为例子，因为太阳所对应的立体角很小，并且基本是常数，不管它处在视场的哪一处（通过这个例子的测试表明所做的近似是有效的）。如果所使用的线性传感器放大倍数未知，可将光源的强度与未知的放大倍数都归结到光源向量中，这样做而不加以说明是很经常的。

从以上推导过程中也可以预期，如果物体之间的距离与到光源的距离在数量级上相当，那么无穷远点光源就不是一个好模型。在这种情况下，我们就不能再用这种近似，必须顾及光源的光通量会随光源距离增加而下降这一事实。

如果我们从不同表面块的角度看光源，问题的核心就不难理解了。对较近的表面块，光源看上去就显得大，不管它的半径有多小；这就意味着光源产生的光通量必须增加。如果光源足够远，就像太阳这个例子那样，我们就可以忽略这种效应，因为对任何可能的运动来说，它的表观尺寸并没有改变。

然而对于挂在房间中部的灯泡这种结构的光源来说，光源所对应的立体角与距离平方的倒数成正比，这就意味着有光源的光通量也会增加。这种情况下正确的光源模型是 5.2.2 节中的点光源。使用这种模型的困难在于光通量在空间中改变很快，并且与实际体验到的不一致。例如，如果点光源放置在一个立方体的中央，按照这个模型的预测，房间角上的光通量应该近似等于每个墙面中心处光通量的三分之一。但是，实际房间的角落并不会那么暗。在实际中，对较近的点光源经常用缩短距离项的方法来应付这种情况。这种做法从辐射度学来说是不正确的，但却趋向于是一个较好的模型。对这种看上去矛盾的现象，只有在讨论了影调模型后才能解释清楚。

5.2.3　线光源

线光源的几何形状呈线形，一个很好的例子是单根的日光灯管。线光源在自然景物或人工环境中并不十分自然，所以我们只简略地讨论一下。对它们感兴趣主要是作为辐射学问题中的一个例子；尤为特殊的一点是接近线光源的表面的光通量，其随着到光源距离的倒数（而不是距离的平方）变化，其中的原因比这一效果更有兴趣。我们用一个半径为ϵ的细圆柱体作为线光源的模型。我们暂且假设线光源是无穷长的，并且考虑表面正对光源的情况，如图 5.4 所示。

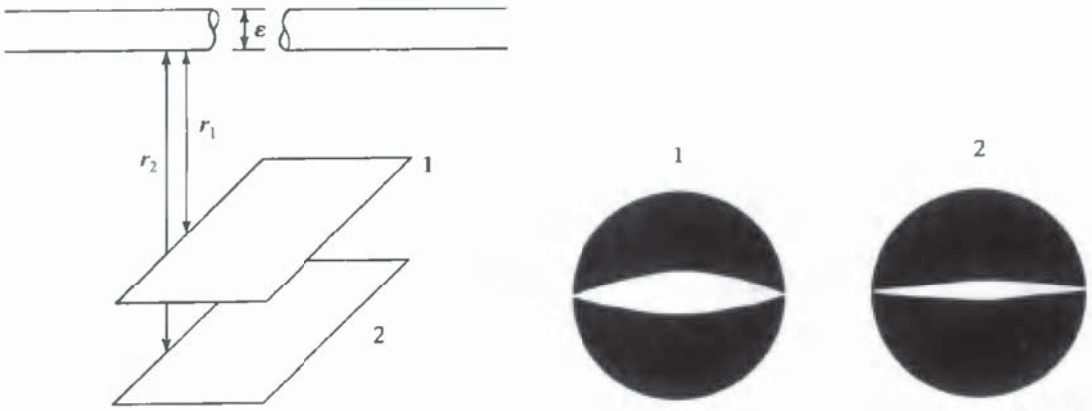


图 5.4　对较接近光源的点来说，光源产生的光通量随着距离倒数关系而下降。左图中有两块表面对着一个无穷长直径的窄圆柱体，发出恒定发射度的光源。右图是从每个表面块角度看光源的形状，画成由下往上看输入半球背面的情况。要强调的是在半球上光源的长度没有改变，但宽度随（ϵ/r）改变，这就得到了相应的结果

图 5.4 画出了从表面块 1 与表面 2 观察光源所呈现情况的草图；半球上照射区域的宽度随到光源的距离而变化，但长度不随其改变（因为光源无穷长）。因为该宽度近似于ϵ/r，因此光源照射的光通量会随着距离倒数的关系下降。显然对不是无穷长的光源来说，这种现象也适用于较近的表面块。

5.2.4 面光源

面光源是指面形发光区域。面光源的重要性体现在两个方面：首先在自然景物中它们经常出现——阴天天空就是一个很好的例子，而在人工环境中，办公室天花板上的日光灯束就是一个例子；其次研究面光源使我们能够说明各种阴影与互反应的效应。一般用表面块所发射出的辐射度与位置及方向无关的模型来描述面光源，它们可以用发射度来描述。

用一个与线光源中用过的论证相似的方法表明，对离光源不太远的点来说，由面光源产生的光通量并不随到光源的距离改变而改变。其理由是，如果面光源的面积与到光源距离相比足够大的话，不管是否靠近光源，某个输入半球被光源覆盖的面积是基本不变的。这说明了为什么在照明工程中如此广泛使用面光源的缘故——它们往往会得到相当均匀一致的照明。对我们的应用来说，需要对面光源照射的光通量有一个更精确的描述，因此需要将积分写出来。

由面光源产生的精确度光通量 假设有一个面光源，其上的点 Q 的发射度为 $E(Q)$，它照射着一个漫反射表面块。用 $\overrightarrow{QP}$ 表示从 Q 到 P 的方向，而不用角度坐标（另一些符号显示在图 5.5 上）。表面的光通量通过对所有入射方向的入射辐射度求和进行计算。这个积分可以转换成对光源所占区域的积分

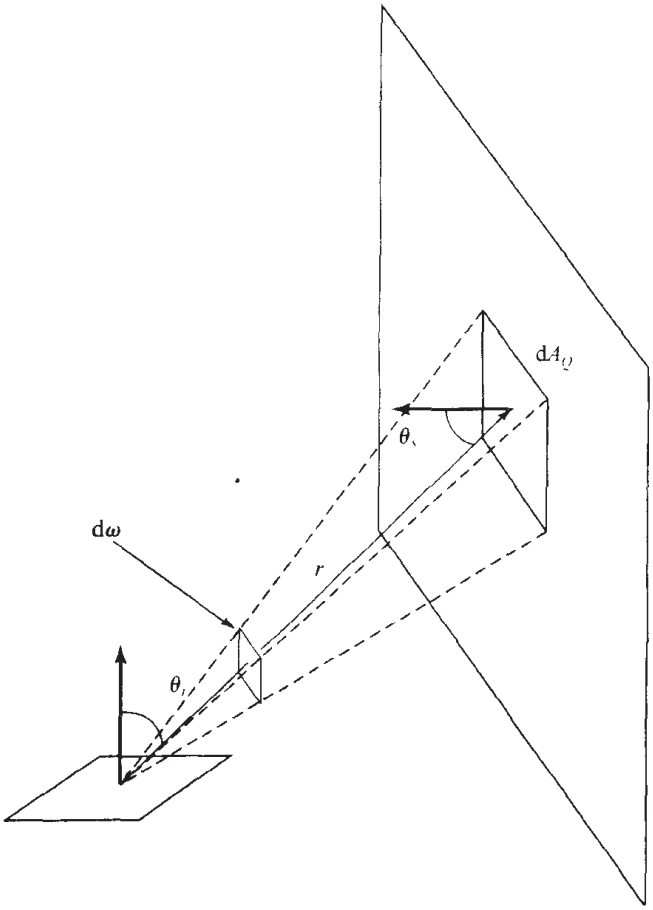


图 5.5 一个漫射光源照亮一个漫射表面，光源发射度为 $E(Q)$，我们希望计算光源照射表面块的光通量。我们通过将表面输入辐射度的积分计算转换成对光源面积的积分计算。这种转换是方便的，因为它能使我们避免对不同的表面使用不同的角度坐标系。然而，它仍然导致一个通常没有封闭形式表示的积分计算

$$
\begin{aligned}
B(P) &= \rho_d(P)\int_{\Omega} L_i(P, \overrightarrow{QP})\cos\theta_i\, d\omega \\
&= \rho_d(P)\int_{\Omega} L_e(Q, \overrightarrow{QP})\cos\theta_i\, d\omega \\
&= \rho_d(P)\int_{\Omega}\left(\frac{1}{\pi}E(Q)\right)\cos\theta_i\, d\omega \\
&= \rho_d(P)\int_{\text{Source}}\left(\frac{1}{\pi}E(Q)\right)\cos\theta_i\left(\cos\theta_s\frac{dA_Q}{r^2}\right) \\
&= \rho_d(P)\int_{\text{Source}} E(Q)\frac{\cos\theta_i\cos\theta_s}{\pi r^2}\, dA_Q
\end{aligned}
$$

这个转换之所以成立，是因为沿直线的辐射度是常数，且有 $E(Q)=(1/\pi)L_e(Q)$。这是很有用的，因为这样做我们就不需要担心角坐标系统的一致性问题。尽管我们采用这种转换，但是描述面光源效果的积分一般是很困难的，也不可能有封闭解。

5.3　局部影调模型

我们已经研究了光的物理学，这是因为我们想知道物体会有怎样的亮度及其原因，以期待从这些模型中抽取物体的信息。到目前为止，我们了解了光源照射表面块的光通量，但是这不是影调模型。辐射可以通过别的方式到达物体表面块（例如可以从别的表面块反射得到），我们需要知道要考虑哪些成分。

最方便运作的模型是局部影调模型，它只考虑将光源所发射的光通量求和计算表面块的光通量。这意味着使用这样一个假设：光并不是从表面反射到另一表面而来的，它来自于光源，到达某个表面，并进而直达摄像机。这个模型显而易见是违反物理学的，但是它容易分析。这个模型支持一系列的算法与理论（见 5.4 节）。遗憾的是，它常常产生相当不准确的推测。更糟的是几乎没有可靠的信息说明在什么情况下使用这个模型是安全的。

一个替代它的模型是考虑所有的辐射量（5.5 节），既考虑从光源来的辐射度，也考虑从辐射表面来的辐射度。这个模型在物理上是准确的，但通常难于操作。

5.3.1　点光源下的局部影调模型

若干点光源照明条件下的局部影调模型，可以通过将由光源所发射的光通量写出来而得到。于是有

$$
B(P) = \sum_{s\in\text{sources visible from } P} B_s(P)
$$

其中，$B_s(P)$是由光源 s 产生的光通量。这个表示式的意义并不大，但是值得注意的是，如果所有光源都是处于无穷远的点光源，就有

$$
B(P) = \sum_{s\in\text{sources visible from } P} \rho_d(P)\boldsymbol{N}(P)\cdot\boldsymbol{S}_s
$$

如果我们只关注所有点都能看到相同光源的区域，可以将所有的光源向量加起来得到一个具有相同效果的虚拟光源。此时形状与影调之间的关系就会相当直接——光通量是表面法

线向量中某个分量的度量。

对不是处于无限远的点光源来说，这个模型变成

$$B(P)=\sum_{s\in\text{sources visible from }P}\rho_d(P)\frac{N(P)\cdot S(P)}{r_s(P)^2}$$

其中，$r_s(P)$是从光源到点 P 的距离，这一项的出现意味着形状与影调之间的关系会更加含糊一些。

阴影的出现 在局部影调模型中，如果一个表面块没有被全部光源照到，就会出现阴影。在这种模型中，点光源产生一系列带清晰轮廓的阴影，看不到所有光源的阴影区域就显得特别暗。由单个光源形成的阴影可能是很清晰的或很暗的，这取决于光源的尺寸与其近旁表面的反射率(它们可能把光反射到阴影中并因而使边界变得模糊)。在 19 世纪，把阴影投射到纸上并画下来，曾是流行的娱乐项目，所得到的剪影在古董店偶尔还能看到。

点光源投射到平面上的阴影的几何关系与透视投影摄像机成像的几何关系很相似(见图 5.6)。如果从表面到光源的射线穿过一个物体，那么平面上的这块表面就处在阴影中。这表明有两种阴影边界。在自遮挡阴影边界处，表面开始转向背离光源的方向，从边界处表面块到光源的射线与表面是相切的。在投射阴影的边界处，从表面的角度看光源突然消失在遮挡物的后面。然而投射到曲面上的阴影会有十分复杂的几何关系的。

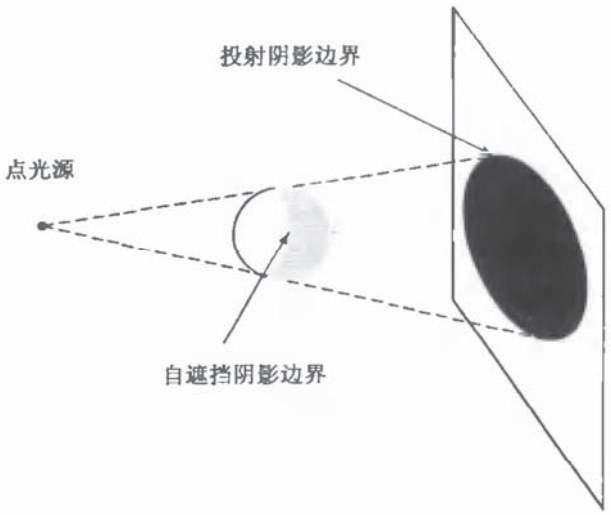


图 5.6 由点光源投影到平面的阴影是相对简单的。自遮挡阴影边界发生在表面朝向转变成背对光源的方向处，投射阴影边界发生在远处表面遮住了光源的视角的情况下

如果有多个光源，阴影就不会那么暗(除了那些看不到任何光源的点)，并且可能会有许多性质上不同的阴影区域(每个光源投影各自的阴影——有些点可能看不到多于 1 个的光源)。这种效果的一个例子发生在电视转播足球比赛的情况下。因为足球场四周均匀分布了许多明亮的照明灯，类似于远处的点光源，每个足球运动员的脚部有一组围绕着他的均匀散射的阴影。当运动员跑来跑去时，这些影子一般会变亮或变暗，这是因为其他光源和互反射产生的照明强度变亮或变暗造成的。

5.3.2 面光源及其阴影

对一组面光源来说，局部影调模型要复杂得多，因为一个表面块很可能只看到某个光源的一部分。这个模型表示成

$$B(P) = \sum_{s \in \text{all sources}} \left\{ \int_{\text{visible component of source } s} \text{Radiosity due to source} \right\}$$

$$= \sum_{s \in \text{all sources}} \int_{\text{visible component of source } s} \left\{ E(Q) \frac{\cos\theta_Q \cos\theta_s}{\pi r^2} dA_Q \right\}$$

其中使用了图 5.5 的术语；通常我们假设 E 对每个光源都是常数。

面光源不会产生有清晰边界的暗阴影，这是由于从被照表面的角度看，光源是从遮挡物体背后逐渐出现的(可想像一下月蚀出现的现象，它们是十分相似的)。通常将点分成两种：全影区——完全看不见光源的区域，半影区——能看到部分光源的区域。在室内环境中，绝大部分光源是这种或那种面光源，所以这种效果是很容易见到的。例如，靠近墙壁举起胳膊并观察它投射的影子，影子中有一个暗的核心，胳膊靠墙越近，这块面积就越大，这是全影区。核心部分被较亮的区域包围(半影区)，边界是模糊的。图 5.7 说明了其几何关系。

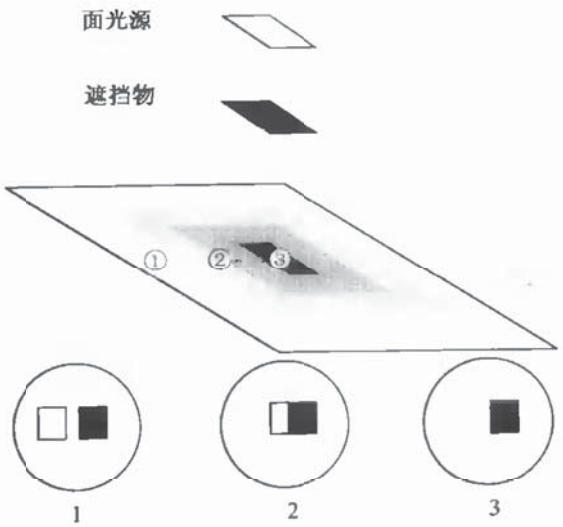


图 5.7　面光源产生复杂边界平滑过渡的阴影区，这是因为从表面块的观察角度看，光源逐渐消失在遮挡物的后面。完全看不到光源的区域称为全影区，能见到部分光源的区域称为半影区。想像自己背躺在一个表面上仰望世界是一个好的模型。在点1可以看见全部光源，点2看到一部分光源，而在点3则什么也看不到

5.3.3 环境光

局部影调模型有一个十分明显的问题：它们预测某些阴影区域看不到光源，因而是完全暗的。但是这种预测几乎在所有场合都是不正确的，因为阴影区能被来自其他漫散表面的光照到。这种效果有时是很明显的，在使用面光源且具有光亮墙面的房间里，只有将物体靠近墙壁或靠近光源时才能见到阴影。其原因是，墙上的表面块能看到房间内其他所有墙面。在一个物体贴近墙壁之前，它只遮住了每个面块可视半球中的很小一部分。

对某些环境来说，一表面块从别的表面块获得的辐照度近似于常数，并且在输入半球上近似于均匀分布。这种情况对一个球体的内部来说是肯定成立的(由于球的对称性)，而对有白色墙壁的房间的内部也是大体成立的(将立方体模型看做球体)。在这种环境中，别的表面块对每个表面块光通量的影响有时可以用增加一个反映环境光的影响的项来建模。有两种确定这一项目的策略。第一种策略是，如果每个表面块所看到环境的比例相同(例如球体的内部)，则可以对每块辐射度加上同样的常数。这一项的幅度通常是估计的。

第二种策略是，如果一些表面所看到的环境比其他表面看到的环境多些或少些(如果环境中的一些区域挡住了一个表面的视线，例如一个处于凹槽底部的表面)这也可以加以考虑。为此需要从所考虑的表面块的视角对环境建模。一种很自然的策略是将环境建模成一个大的、遥远的、有常数光通量的多边形，这个多边形的视线对某些表面是被遮挡了的(见图 5.8)。对看到环境较少的表面来说，其后果是反映环境光的影响的项要小些。这种模型通常比增加常数环境光的影响的项要准确些。遗憾的是，从这种模型中提取信息是非常困难的，难度很有可能与使用全局影调模型时相同。

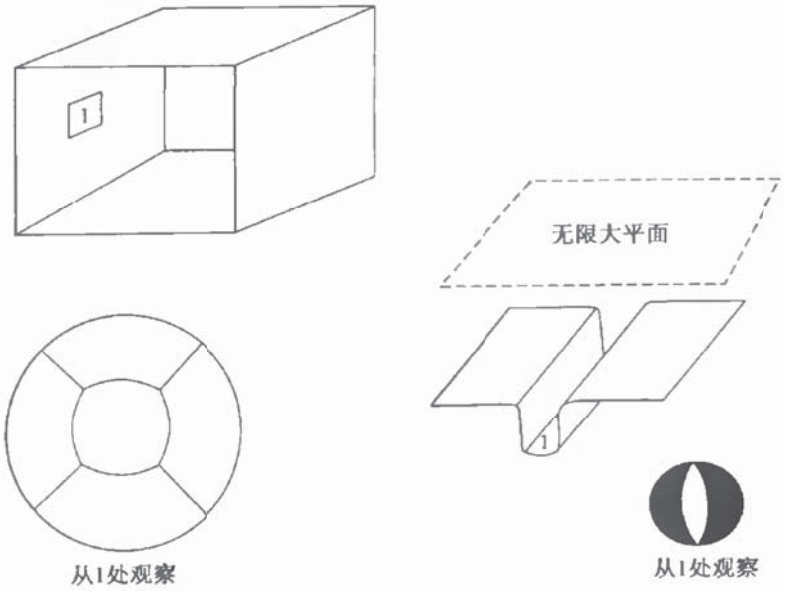


图 5.8 环境光是局部影调模型中的一个附加项，它用来对光通量进行预测时考虑从远距离反射表面的影响。对一个球体内部或一个立方体内部的环境(图的左面)来说，其中一个面块上每一点看到大致相同的事情，此时用一个常数的环境光的项是合适的。在更加复杂的情况下，某些表面块所看到的周围环境比其他面块要明显少。例如在右图凹槽底部所看到的外部环境相对少，对此可建模成具有常数发射度的多边形，其输入半球如右图下部所示

5.4 应用：光度学体视

我们用一个表面在不同光照条件所拍摄的图像序列来重构这个表面块的形状。为简单起见，我们采用正交投影摄像机，并且选择坐标系统使空间点(x,y,z)投影到图像上的点(x,y)(我们采用的方法对第 1 章中叙述的其他摄像机模型也适用)。

在这种情况下，要度量表面的形状，需要获得表面的深度。受此启发，可用$(x,y,f(x,y))$

来表示这个表面——这种表示方式称为 Monge 表面，以首次使用这种模型的法国军事工程师的名字命名(见图 5.9)。这种表示方法之所以具有吸引力，是因为给定表面点的图像坐标可以惟一确定这个点。需要提醒的是，如果要获得一个刚体的度量，需要重构不止一个面块，因为我们还需要观察物体的背面。

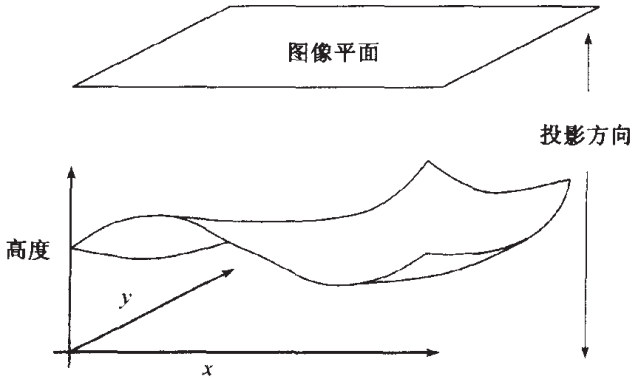


图 5.9　Monge 表面将表面块用高度的函数来表示，对于光度体视这个例子假定使用的是正交投影摄像机注视着Monge表面，它将空间点(x, y, z)映射到图像点(x, y)。这意味着表面的形状可以用图像中的位置的函数表示

光度学体视法是一种从图像恢复 Monge 表面表达式的方法。这个方法对一个表面在若干幅图像中的图像强度值进行分析推理，这些图像从同一视角观察该表面但用不同的光源照明。该方法恢复与每个像素相对应的表面上各点的高度，在计算机视觉领域中将所得到的表达式称为高度图、深度图、或致密深度图。

将摄像机与表面的位置固定好，并使照明用的点光源到表面的距离比表面块尺寸大得多。计算采用一个局部影调模型，并且假设没有环境光(后面会再次提到)，这样一来表面上一点 P 的光通量为

$$B(P) = \rho(P)\boldsymbol{N}(P) \cdot \boldsymbol{S}_1$$

其中，$\boldsymbol{N}$ 是单位表面法线，$\boldsymbol{S}_1$ 是光源向量。结合我们使用的摄像机模型，每个图像点(x, y)只有一个表面点 P，因此可用 $B(x, y)$表示 $B(P)$。我们假设摄像机对表面光通量的响应是线性的，所以在(x, y)处像素的值为

$$\begin{aligned} I(x, y) &= kB(x, y) \\ &= k\rho(x, y)\boldsymbol{N}(x, y) \cdot \boldsymbol{S}_1 \\ &= \boldsymbol{g}(x, y) \cdot \boldsymbol{V}_1 \end{aligned}$$

其中，k 是表示摄像机对输入辐射度响应的一个常数 $\boldsymbol{g}(x, y) = \rho(x, y)\boldsymbol{N}(x, y)$，以及 $\boldsymbol{V}_1 = k\boldsymbol{S}_1$。

在这些等式中，$\boldsymbol{g}(x, y)$描述表面，$\boldsymbol{V}_1$ 是有关照明与摄像机属性的一个量，图像像素的值是向量 $\boldsymbol{g}(x, y)$与向量 $\boldsymbol{V}_1$ 的点积。有了足够多的点积值，就可以重构 $\boldsymbol{g}$ 以及这个表面。

5.4.1　由多个视图获取表面法线与反射率

如果我们具有 n 个光源，每个光源对应的 $\boldsymbol{V}_i$ 已知，则可将这些 $\boldsymbol{V}_i$ 重叠到一个矩阵$\mathcal{V}$中，该矩阵是已知的

$$\mathcal{V}=\begin{pmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \\ \cdots \\ \boldsymbol{V}_n^T \end{pmatrix}$$

对每一个图像点，也可将其度量值写成一个向量形式

$$\boldsymbol{i}(x,y)=\{I_1(x,y),I_2(x,y),\cdots,I_n(x,y)\}^T$$

需要说明的是，每个图像点有一个向量，每个向量包含了该点在不同光源条件下观察到的所有图像亮度值。现在有

$$\boldsymbol{i}(x,y)=\mathcal{V}\boldsymbol{g}(x,y)$$

而 $\boldsymbol{g}$ 通过解这个线性系统获得，图像中每个点有一个线性系统。一般情况下，要求 $n>3$，这样可以得到最小二乘解，它的好处在于解的残留误差提供了对度量值的检验。

这种方法的困难在于表面的相当一部分区域可能处于这个或那个光源的阴影之下(见图 5.10)。有一种对付阴影的窍门，这种窍门是，如果的确没有环境光，那么可以从图像向量生成一个矩阵，并且用这个矩阵乘上等式的两边，则可以将处在阴影中的任何一个点的方程置成零。所生成的矩阵为

$$\mathcal{I}(x,y)=\begin{pmatrix} I_1(x,y) & \cdots & 0 & 0 \\ 0 & I_2(x,y) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & I_n(x,y) \end{pmatrix}$$

以及

$$\mathcal{I}\boldsymbol{i}=\mathcal{I}\mathcal{V}\boldsymbol{g}(x,y)$$

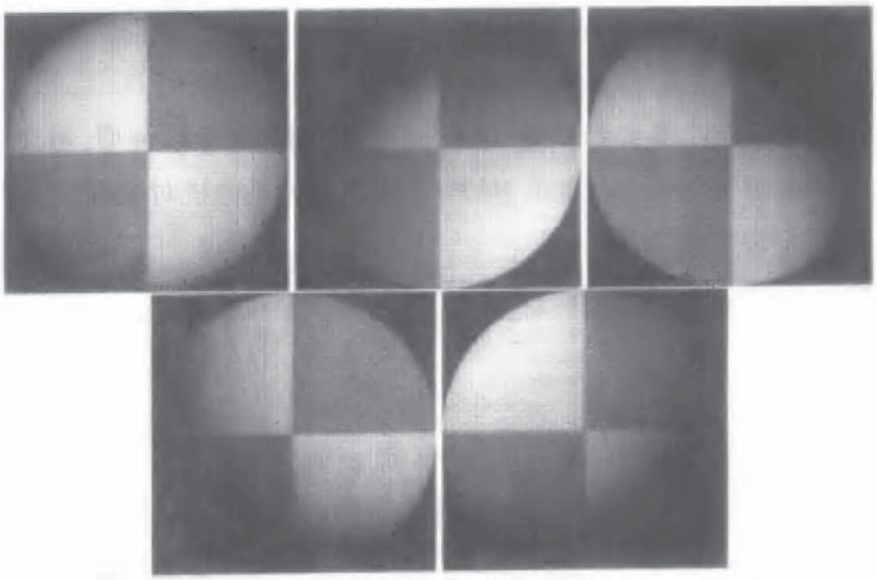


图 5.10 一个球面组成的 5 幅图像，是使用正交投影从同一视角得到的。这些图像的影调使用的是局部影调模型与远处的点光源。这是一个凸面物体，所以只有当光源方向与视角平行时，才没有阴影。不同光源条件下亮度的变化给出了有关表面形状的信息

矩阵$\mathcal{I}$能起到使阴影区域的效果置成零的作用，因为该矩阵中对应于阴影中的点的元素是零。此时图像中每个点都有一个线性系统，解这个线性系统就可恢复每个点的 $\boldsymbol{g}$ 向量。

度量反射率　因为 $\boldsymbol{N}$ 是单位法向量，因此从 $\boldsymbol{g}$ 的度量中就可以提取出反射率，这意味着 $|\boldsymbol{g}(x,y)| = \rho(x,y)$。这也给度量提供了检验的机会，因为反射率的数值范围只能从零到 1，任何像素$|\boldsymbol{g}|$大于 1 者都值得怀疑——或者是该像素无法计算或者$\mathcal{V}$是错误的。用这种方法求得的图 5.10 中图像的反射率显示在图 5.11 中

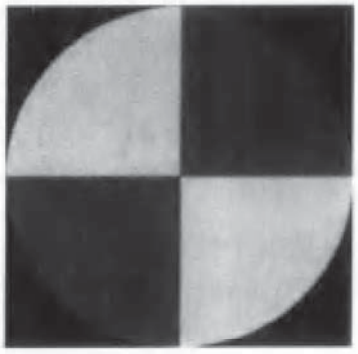


图 5.11　用一个图来表示，从图 5.10 的输入数据获得的向量场 $\boldsymbol{g}(x,y)$的幅值，它是表面的反射系数

恢复法线向量　因为法线向量是单位向量，它也可以从 $\boldsymbol{g}$ 中得到

$$\boldsymbol{N}(x,y) = \frac{1}{|\boldsymbol{g}(x,y)|}\boldsymbol{g}(x,y)$$

图 5.12 显示了从图 5.10 恢复的法线向量值。

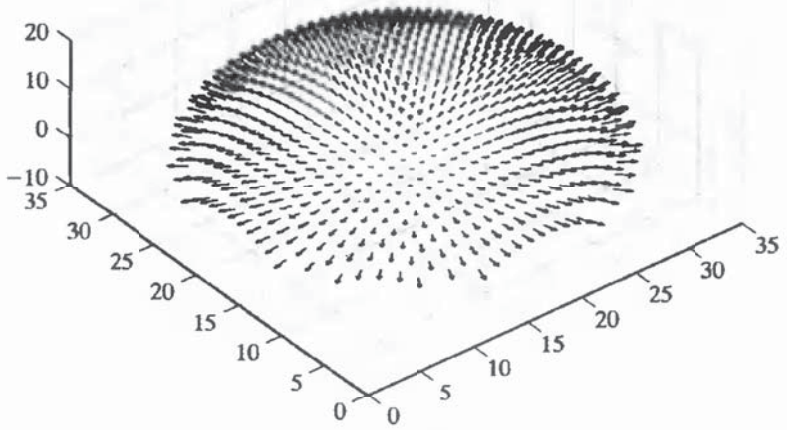


图 5.12　从图 5.10 恢复得到的法线向量场

5.4.2　由法线获取表面形状

由于表面表示成$(x,y,f(x,y))$，所以法线是(x,y)的函数

$$\boldsymbol{N}(x,y) = \frac{1}{\sqrt{1+\frac{\partial f}{\partial x}^2+\frac{\partial f}{\partial y}^2}}\left\{-\frac{\partial f}{\partial x},-\frac{\partial f}{\partial y},1\right\}^T$$

为了恢复深度图，需要从单位法线的度量值确定 $f(x,y)$。

假设在某点(x,y)的法线向量的度量值是$(a(x,y),b(x,y),c(x,y))$，那么

$$\frac{\partial f}{\partial x}=\frac{a(x,y)}{c(x,y)} \quad 和 \quad \frac{\partial f}{\partial y}=\frac{b(x,y)}{c(x,y)}$$

此外，我们又得到了对数据集的另一种检验，因为

$$\frac{\partial^2 f}{\partial x \partial y}=\frac{\partial^2 f}{\partial y \partial x}$$

所以在每个点的差值

$$\frac{\partial\left(\frac{a(x,y)}{c(x,y)}\right)}{\partial y}-\frac{\partial\left(\frac{b(x,y)}{c(x,y)}\right)}{\partial x}$$

应该很小。在原理上它应该是零，但是由于我们只能通过数值方法估计偏导数，因此差值是非零的小值是可以允许的。这种测试称为可积性测试，在视觉应用中常常简化成检验混合二阶偏导数是否相等。

算法 5.1 光度学体视

在同一视角不同照明条件下采集许多图像

从光源与摄像机信息确定矩阵$\mathcal{V}$

对反射率与法线向量(3 个分量)构造数组

p($\frac{\partial f}{\partial x}$度量值)与

q($\frac{\partial f}{\partial y}$度量值)

For 图像数组的每个点

将图像值写成一向量 $\boldsymbol{i}$

构筑对角矩阵$\mathcal{I}$

对$\mathcal{I}\mathcal{V}\boldsymbol{g}=\mathcal{I}\boldsymbol{i}$ 求解，获取该点的 $\boldsymbol{g}$

该点的反射率是$|\boldsymbol{g}|$

该点的法线向量是$\frac{\boldsymbol{g}}{|\boldsymbol{g}|}$

该点的 p 是$\frac{N_1}{N_3}$

该点的 q 是$\frac{N_2}{N_3}$

end

检验：$\left(\frac{\partial p}{\partial y}-\frac{\partial q}{\partial x}\right)^2$ 是否处处很小？

（未完待续）

（续）

```
将高度图的左上角置零
For 高度图左列的每个像素
  高度值 = 原有高度值 + 相应的 q 值
end
For 每一行
  For 每行最左元素外的所有元素
    高度值 = 原有高度值 + 相应 p 值
  end
end
```

通过积分求形状　假设偏导数通过健全测试，就可以在保留一个常深度偏差条件下重构出该表面，偏导数给出沿 x 或 y 方向走一小步时表面高度的变化，因而通过沿某种路径累计高度的变化值就可得到该表面的形状。具体说来，有

$$f(x, y) = \oint_C \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot d\boldsymbol{l} + c$$

其中，C 是从某一固定点到(x, y)的一条曲线，c 是一个积分常数，它表示在起始点处表面的(未知)高度。恢复表面与所选的曲线无关(练习)。

例如我们从(0,0)开始重构在(u, v)的表面，先从沿 $x = 0$ 到点$(0, v)$的线将 y 求和，再沿 $y = v$ 到点(u, v)将 x 求和

$$f(u, v) = \int_0^v \frac{\partial f}{\partial y}(0, y)\, dy + \int_0^u \frac{\partial f}{\partial x}(x, v)\, dx + c$$

这是算法 5.1 给出的积分路径。任何其他路径集也能奏效，但是最好的方法是使用许多不同路径计算并加以平均，目的在于削弱偏导数估计的误差。图 5.13 显示了从图 5.10 数据得到的重构结果。

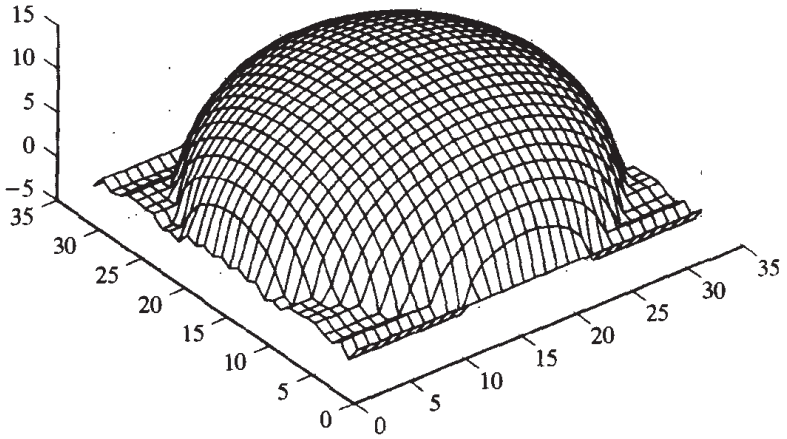


图 5.13　对图 5.12 的法线向量场，使用文中讲述的方法进行积分得到的高度场

恢复形状的另一个方法是选择一个函数 $f(x,y)$，使它的偏导数与度量的偏导数数值最相近。我们在6.5.2节的一个相似问题中再讨论这种方法。

5.5 互反射：全局影调模型

局部影调模型很可能产生误导。在现实世界中，每个表面块不仅被光源照明，也被别的表面块的反射光照明（这种现象称为互反射）。考虑了互反射效果的模型称为全局影调模型。互反射会导致各种复杂的，并且仍没有充分理解的影调效果。遗憾的是，这些效果经常发生，并且至今仍然不知怎样才能在不损害基本定性属性的条件下简化全局影调模型。

例如，图5.14展示了两间房间内部的视图。一间房间有黑色的墙，并放置有黑色物体；另一间有白的墙与白色物体。每一房间近似地用一个远处的点光源照明模型。在光源强度可适当调整的条件下，按局部影调模型的预测，这些图像的差别可以做到分辨不出来。但是实际上与白屋相比，黑屋有更暗的阴影，以及在多面体的折缝处的边缘更加鲜明。这是因为黑屋内的表面反射到别的表面的光少（它们更暗），而在白屋中别的表面是不可忽视的辐射源。在图中分别显示了摄像机对光通量的响应的一节（对漫反射表面来说与光通量成正比），它们在性质上是明显不同的。在黑屋子内表面块的光通量是常数，就像局部影调模型所预测的那样，而在白屋子中图像缓慢变化是常见的——这些发生在凹面角处，在那里物体表面把光彼此反射到对方。

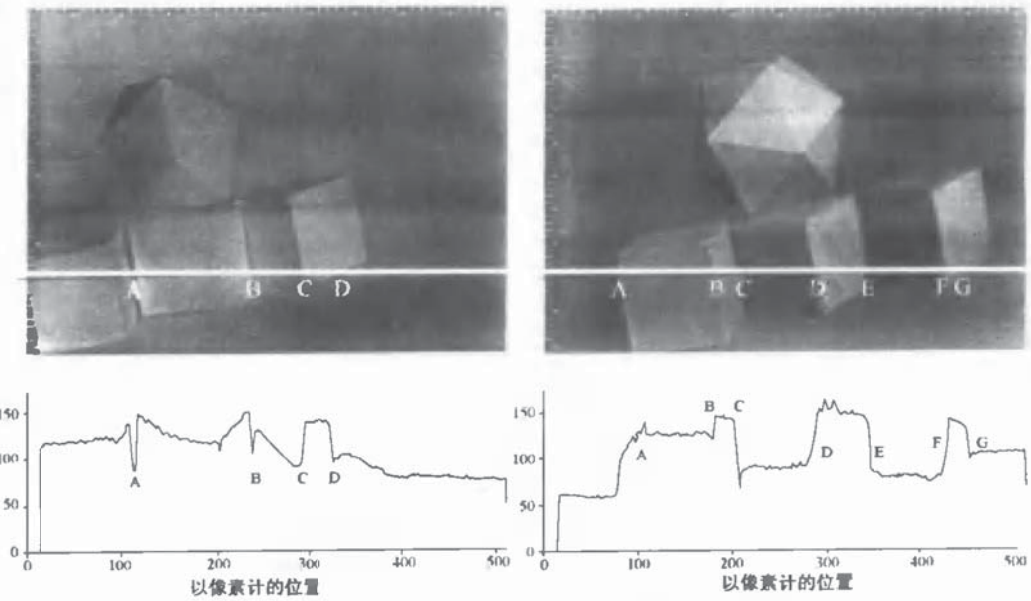


图5.14 左边一列所显示的数据是从无光黑墙的屋子取得的，在屋内有一些无光黑色的多面体物体；右图一列的数据是从一个放有白色物体的白屋内部取得的。这两幅图像在性质上是不同的，在黑屋内有更暗的阴影与鲜明的轮廓，而在白屋中的凹面角上有光的反射。该图展示了沿一条线上一段图像强度分布

这种效果解释了为什么一个用点光源照射的房间并没有像局部影调模型预测的，有鲜明的照明变化的情况（回顾 5.2.2 节）。房间的墙与地面将光线反射回来，因而使原本应该很暗的角落有变亮的趋势。

5.5.1　一个互反射模型

对一组漫反射表面块，光通量的预测机理是不难理解的：离开该表面块的总光通量是它的发射度（除了光源外它为零）加上从它反射的光通量：

$$B(P) = E(P) + B_{\text{refl}}(P)$$

从表面块的视角看，无法区分从别的表面射来的能量是该表面自身发光还是反射的光，因此可以使用面光源的表达式，获取 $B_{\text{refl}}(Q)$ 的表达式。特别地，从所关注的表面块的视角看，客观世界 R 处的一个表面块，等价于一个具有发射度 $B(R)$ 的面光源。也就是说

$$\begin{aligned} B_{\text{refl}}(P) &= \rho_d(P) \int_{\text{world}} \text{visible}(P, Q) B(Q) \frac{\cos\theta_P \cos\theta_Q}{\pi d_{PQ}^2} dA_Q \\ &= \rho_d(P) \int_{\text{world}} \text{visible}(P, Q) K(P, Q) B(Q) dA_Q \end{aligned}$$

其中的术语在图 5.15 中说明，而

$$\text{visible}(P, Q) = \begin{cases} 1, & \text{如果 } P \text{ 可以看到 } Q \\ 0, & \text{如果 } P \text{ 不可以看到 } Q \end{cases}$$

项 $\text{visible}(P, Q) K(P, Q)$ 通常称为互反射核函数，将其代入 $B_{\text{refl}}(P)$ 得到

$$B(P) = E(P) + \rho_d(P) \int_{\text{world}} \text{visible}(P, Q) K(P, Q) B(Q) dA_Q$$

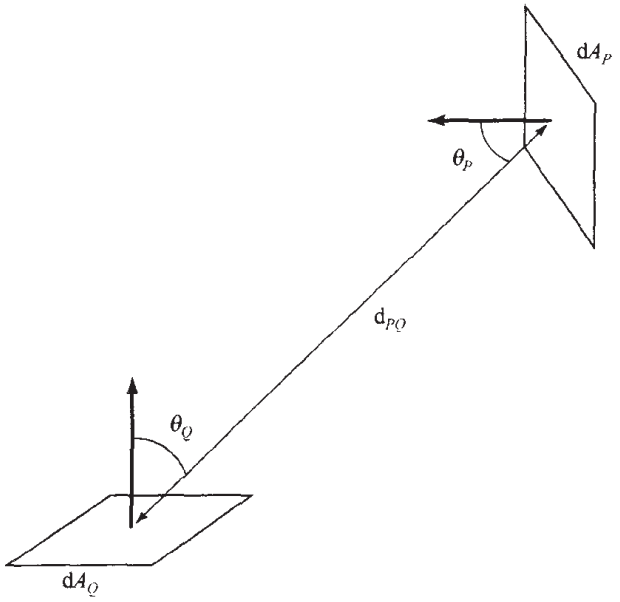


图 5.15　文中所推导的互反射核表达式中使用的术语

要强调的是问题的解在积分号内。这种形式的方程称为第二类 Fredholm 积分方程式，这种方程式是相当难对付类型方程式的一个例子，因为一般说来互反射核函数是不连续的，并且会有奇异点。这个方程式的解，可以得到相当好的描述漫反射表面外观的模型，这个话题对计

算机图形学领域的相当部分产业起了支持作用[Cohen 和 Wallace(1993)或 Sillion(1994)的工作是这个话题很好的起点]。这个模型对所观察的效果有好的预测性(见图 5.16)。

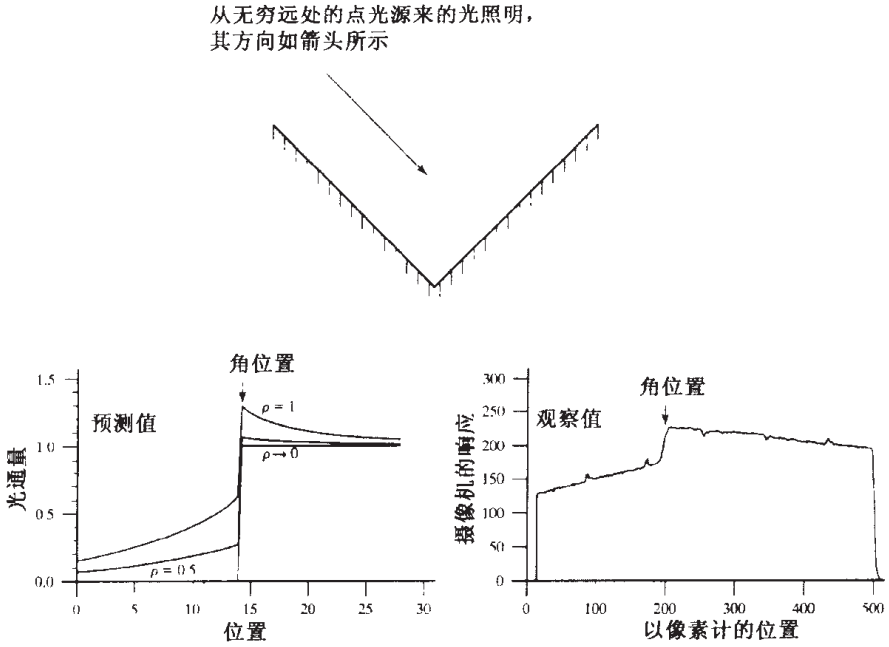


图 5.16 文中描述的模型，对相互反射能产生准确的定性预测。图的上部表示了一个由无穷远点光源照射的直角凹槽，光源方向与其中一个面平行。左下图是不同的ρ值时对这种格局进行的光通量预测。适当的比例放大了各种ρ值预测的值，以便各条曲线的走向看得更清楚。$\rho\to 0$的情况对应局部影调模型。右图表示的是对用白纸做的角的图像所观察到的图像强度，它显示出了与边有关的屋顶式的坡度，而局部影调模型预测的是一个台阶

5.5.2 光通量的求解

我们通过对一个求全局影调模型解的概要分析来说明这种方法。首先将场景划分成小的平面块，并假设每一块上的光通量近似为常数值。这种近似方法是符合逻辑的，因为对小块表面进行分析可以得到准确的表达式。在此基础上构造一个向量 $\boldsymbol{B}$，它的每个分量含每块的光通量。说得更确切些，$\boldsymbol{B}$ 的第 i 分量是第 i 块表面的光通量。

将第 i 块表面上来自第 j 块表面上的光通量的输入光通量写成

$$B_{j\to i}(P)=\rho_d(P)\int_{\text{patch } j}\text{visible}(P,Q)K(P,Q)\,dA_Q B_j$$

其中，P 是第 i 块表面的坐标，而 Q(原书上为 R)是第 j 块表面的坐标。鉴于这个表示式并不是常数，所以必须取第 i 块的平均值，得到

$$\overline{B}_{j\to i}=\frac{1}{A_i}\int_{\text{patch } i}\rho_d(P)\int_{\text{patch } j}\text{visible}(P,Q)K(P,Q)\,dA_P\,dA_Q B_j$$

其中，A_i 是第 i 块表面的面积。如果我们硬性规定每块表面的发射度也是常数，就可以得到模型

$$B_i = E_i + \sum_{\text{all } j} \overline{B}_{j \to i}$$
$$= E_i + \sum_{\text{all } j} K_{ij} B_j$$

其中，

$$K_{ij} = \frac{1}{A_i} \int_{\text{patch } i} \rho_d(P) \int_{\text{patch } j} \text{visible}(P, Q) K(P, Q) \, dA_P \, dA_Q$$

有时称矩阵中的元素为形状因子。

B_i 是由线性方程构成的系统(尽管是一个极其大的线性方程系统—— K_{ij} 可以是一个百万乘百万的矩阵)。为了高效、快速与准确地求解该系统，需要使用一些技巧。但这超出我们这本书的范围，Sillion(1994)对此做了很好的叙述，Cohen 与 Wallace(1993)的书也是如此。

5.5.3　互反射的定性效果

我们希望从光通量抽取出形状信息。使用局部模型来做这件事是相对简单的(5.4 节提供了一些细节)，但这种模型描述世界的效果不好，它对推导出的形状信息所造成影响的严重程度也几乎一无所知。使用全局影调模型提取形状信息是困难的，这有两个原因。第一，形状与光通量之间的关系是复杂的，因为它由相互反射的核函数控制。其次，几乎经常会有一些看不见的表面对视野内的物体辐射光。这些所谓“隐含的表面”的存在就意味着难以使用互反射核来考虑景物中的全部辐射情况，因为某些辐射源是看不见的，我们很可能对它们一无所知。

以上所述表明定性地了解互反射的局部效应是重要的。从这种考虑出发，我们既可以降低互反射产生的效果，但也可以充分地利用它们。尽管这个话题很大程度上还是个开放的研究话题，但是有些事情可以说一说。

平滑与区域效应　首先要指出的是，互反射自然会起到平滑的作用。如果我们打算通过彩色玻璃投射到地面上的花式来解释彩色玻璃，就会很明显地看到这种现象：地面上的花式常常是一群模糊的彩色团状物。在使用图 5.17 的粗糙的模型时，这种现象很容易看到。图中的几何关系是一个表面块面朝一个无限平面，该平面距离该表面一个单位距离远，它的光通量呈现 $\sin\omega x$ 规律。研究改变表面块到平面的距离是没有必要的，因为相互反射问题具有比例不变的解，也就是说，对一个有两个距离单位远的表面块的解，可以通过读图上 2ω 处的数就可。这个表面块面积很小，以至于它对平面光通量的影响可以忽略不计。如果这表面块相对平面的倾斜角为 σ，它所载有的光通量也接近周期性的，空间频率为 $\omega\cos\sigma$。我们称该频率分量的幅值为表面块的增益，画在图 5.17 上。这个图的重要属性是，空域高频很难跨过平面到表面块之间的间隔。这意味着具有高频与高幅度的影调效应一般不可能来自于远处的表面(除非它们超乎寻常的亮)。

由于远处表面造成空域频率项的幅值快速衰减，因此如果看到高频段有高幅度项，那么它几乎不可能是远处被动辐射源产生的效果(因为这些效果迅速淡化)。有一种区分影调的通用惯例将在 6.5.2 节中讨论。这种惯例是说：如果影调快速变化(“边”)以及动态范围相对低，则影调是反射产生的，否则就是照明所引起的。我们可以解释这个惯例：空间频率有一半基本上不受远处表面的互相照明的影响，因为增益小。这种范围内的空域频率不可能从远处被动辐射源来，除非这些辐射源有超乎寻常的高光通量。因此，这些频率范围的空间频率可以看成具有区域效应，它们只能从一定距离范围的互反射产生。

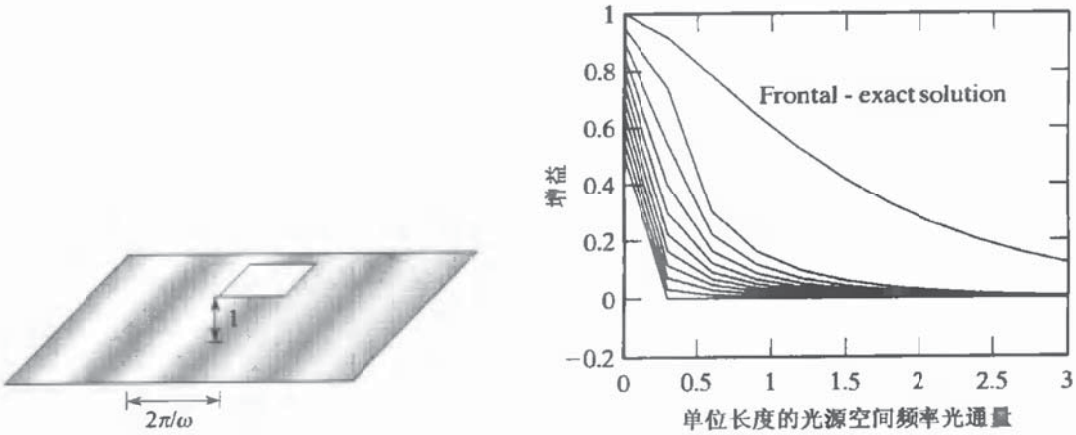


图 5.17 一小块表面注视一个平面，平面具有正弦变化幅度为一个单位的光通量。由于平面产生的效应，这块表面具有近似正弦变化的光通量。我们称该成分的幅值为表面块的增益。这个图显示了表面块倾角不同时增益的数值估计与平面上空域频率的函数关系，倾角分成10档，从0到$\pi/2$。增益下降得非常快，意味着空间高频率段的大项必定是近距离产生的效应，而不是远处放射体起的作用。这也就是为什么很难通过观察彩色玻璃窗口跟前的地面来确定彩色玻璃窗上的图案的原因

最引人注意的区域效应可能就是反光——主要在凹陷的区域出现小亮斑(见图 5.18 与图 5.19中说明)。另一个重要的现象是色彩掺和，彩色表面将光反射到别的彩色表面上。一般情况下人们并不太注意这种现象，除非特意地要观察。色彩掺和经常由画家再现出来。

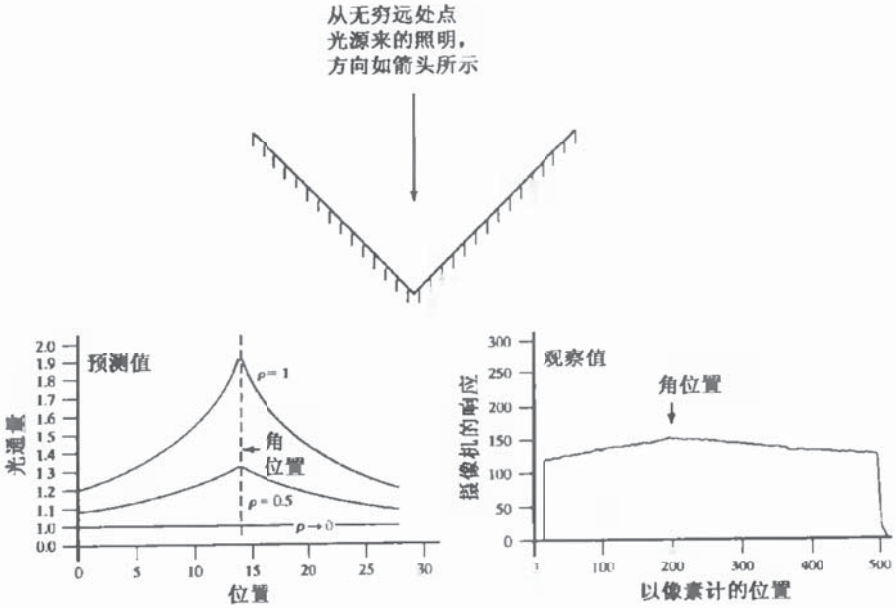


图 5.18 凹面边界上的反光现象，是相互反射现象造成的结果。图的顶部显示这里遇到的情况：一个无穷远点光源照明一个凹面直角槽，光源向量为角平分线的方向。左图表示对这种场景互反射模型的强度预测。$\rho \to 0$的情况是局部影调模型。这些曲线已对齐以便比较。随着表面反射率增加，屋顶类型的结构出现了。右图显示的是从一个实际场景中观察到的这种效应

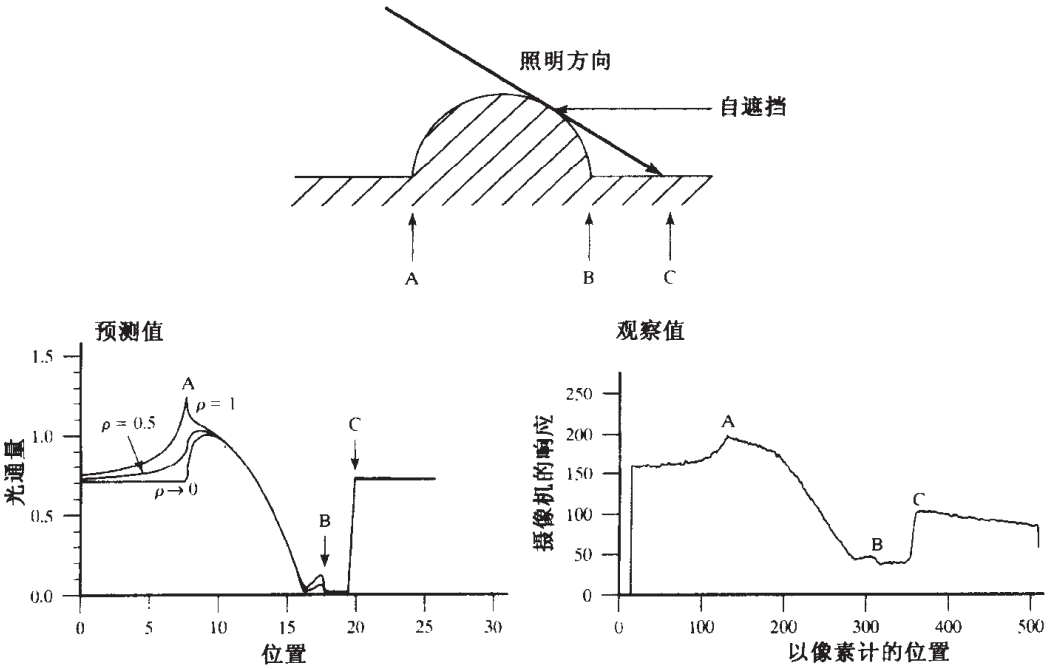


图 5.19　反光发生很普遍，它们通常是由大的反射表面在合适的视角引起的。在图上部所示的几何关系中，柱形突起的阴影区域正以相当合适的视角观察平面背景——如果背景足够大，在突起底部的表面块的半球中有接近一半看到这个平面。这意味着突起边缘以及投射阴影区域内有较大幅度的反光（而局部模型预测它是黑的）。突起的另一边也有反光，不同值的解显示在左图上（为了便于比较已经过规范化处理），右图是在一个实际场景观察到的效果

5.6　注释

在有关视觉问题的文献中，影调模型研究得很不系统，点光源近似到处滥用。使用它时要加小心，对别人的使用方法要带着疑问。我们认为我们是指出光源的物理效应与影调模型之间区别的第一人。

局部影调模型

局部影调模型最大的好处是分析方法简单。局部影调模型的主要特征是在具有常数反射率的表面，表面块的光通量只是法线向量的函数。这意味着人们可以避免对反射与光源进行分辨，取而代之的是以表面及光源属性简单编码成反射映射来表示。这种反射是一种函数，它获取法线向量的表达式，对具有该种法线向量的点返回期望的光通量值。

Horn 开创了在计算机视觉中对影调进行系统性研究，他的重要文章是使用点光源从局部影调模型恢复形状（Horn，1970，1975），一个时间更近的说明见 Horn（1990）。这种方法基本上已废而不用了（至少部分不用了，因为它们在处理由全局影调模型引起的困难方面显得无能为力），所以我们这里不再综述大量的文献。一个全面的概述可见 Horn 与 Brooks（1989）。形状与反射率是歧义的，适当改变反射率，不同形状的表面可以产生相同的图像（Belhumeur，Kriegman 与 Yuille，1999，Kriegman 与 Belhumeur，1998）。由于表面法线在局部影调模型中是个关键，因而这种模型通常在表面影调与曲率之间建立了很好的联系（Koenderink 与 van Doorn，1980）。

互反射

全局影调的效果在有关影调的文献中常常被忽略，这引起本书作者之一的强烈不满。忽略互反射的理由在于对它进行分析十分困难，尤其是使用全局影调模型的输出来推断物体属性时。如果相互反射现象的效果对模型的输出改变不多，那么忽略它们问题可能不大。遗憾的是，很少有人沿这种思路去推理，其原因在于很难证明互反射条件下所使用的方法是稳定的。对空间频率问题的讨论是遵循 Haddon 和 Forsyth(1998*a*)的思路的，它们是受 Koenderink 与 van Doorn(1983)工作而启发的。除此之外，关于相互反射影调的全面性质的知识很少，据我们所知这是一个重要的空档。另一种不同的策略是使用一个绘制模型以迭代方法重复估计形状(Nayar, Ikeuchi 与 Kanade, 1991*a*)。

Horn 是指出全局影调效果重要性的第一人(Horn, 1977)。Koenderink 和 van Doorn(1983)提出全局模型下的光通量可在通过局部模型得到的光通量的基础上，再用一个线性运算得到。人们于是研究这种运算，在某种情况下它的特征函数(经常称为几何众数)是有益的。此后 Forsyth 与 Zisserman(1989, 1990, 1991)展示了由相互反射引起的各种定性的效果。

光度学体视

光度学体视的原型来自于 Woodham，这种有用的概念有若干变化的形式(Horn, Woodham 与 Silver, 1978; Woodham, 1979, 1980, 1989, 1994)。光度学体视法有若干种发展。一个有趣的想法是用三种不同颜色的光照射表面(以及在不同的位置)以及使用彩色图像。对色彩进行适当选择，相当于得到三幅图像，因而度量过程被简化了。

一般说来，光度学体视用在照明条件很容易控制的环境中，以确保图像中没有环境光的影响。将环境光的影响加入到给定的计算格式中也是相对简单的，可以在矩阵V中附加一列 1。在这种情况下，$\boldsymbol{g}(x, y)$成为一个四维向量，其第 4 个分量是环境光项。然而这种方法并不能确保环境光的项在空间是常数，因此必须检验这一项是否是常数，如果不是则要调整这个模型。

光度学体视法只采用局部影调模型，该模型不一定限于由远处的点光源照射的朗伯表面。如果表面的光通量是表面法线满足少量约束的已知函数，那么仍可使用光度学体视。这是由于一个单幅视图像素的强度只能将法线确定成含一个参数的解，这就要求用两个视图来确定法线。已知反射率的表面由远处的点光源照射的情况是最简单的例子。

实际上如果表面的光通量是表面 k 个参数的函数，光度学体视仍然可采用。单个视角的像素值将法线定为含 $k+1$ 个参数的表示，则 $k+1$ 个视图确定了法线。要使这种方法有效，光通量需要用我们的运算能解决问题的函数表示(例如，如果表面的光通量是表面法线的常函数，这就不可能从光通量推出对法线应满足的条件)。这样就可以同时恢复形状与反射图(Garcia-Bermejo, Diaz Pernas 与 Coronado, 1996; Mukawa, 1990; Nayar, Ikeuchi 与 Kanade, 1990, 以及 Tagare 与 de Figueiredo, 1992, 1993)。

其他影调表示方法

除了从影调信号获取形状信息这种方法之外，也可以试图将它们与一组不相同的例子进行匹配，这意味着要研究一个表面能产生怎样的影调现象这个问题。能够出现的影调现象集是格外有限的(Belhumeur 与 Kriegman, 1998)。关于这种集的结构的知识很有价值，因为它使得我们知道如何对影调图像进行比较而不至于与照明变化相混淆。照明变化是在人脸检测与识别的应用

中的特殊问题(Adini,Moses 与 Ullman,1997;Phillips 与 Vardi,1996)。了解照明可能引起的变化看来是有帮助的(Georghiades,Kriegman 与 Belhumeur,1998,2000;Jacobs,Belhumeur 与 Basri,1998)。

另一种可能性是将相互反射定性分析的概念扩展而得到影调基元——一种影调模式,它是特征化的,且与形状模式稳定地相联系。例如,表面的窄槽、深孔是暗的,圆柱体的影调具有逐渐展开的特征。只有少量的基元为大家熟知,但其中一些显得挺有用的(Haddon 与 Forsyth,1998*a*,1998*b*)。

习题

5.1　如果光源是点光源,那么一个球投到平面上的阴影的形状是怎样的?

5.2　有一个方的面光源与一块方的遮挡物,它们都平行于一个平面。光源与遮挡物尺寸相同,并且它们垂直上下放置,中心对齐。

(a) 全影区的形状是怎样的?

(b) 半影区外轮廓的形状是怎样的?

5.3　有一个方的面光源与方的遮挡物,平行于一个平面。面光源的边长是遮挡物的 2 倍,它们上下垂直放置且中心对齐。

(a) 全影区的形状是怎样的?

(b) 半影区外轮廓的形状是怎样的?

5.4　有一个方的面光源与方的遮挡物,平行于一个平面,垂直上下放置且中心对齐。现光源的边长为遮挡物的一半。

(a) 全影区的形状是怎样的?

(b) 半影区外轮廓的形状是怎样的?

5.5　一小球将其阴影投射到一个大球上,请描述可能出现的阴影轮廓。

5.6　说明为什么说用阴影边界来推断形状是困难的,尤其是当阴影投到一弯曲表面上时,则更困难。

5.7　一个无穷小表面从正面沿圆形面光源的对称轴注视着该面光源,光源的发射度是常数。计算表面块从光源发射度 $E(\boldsymbol{u})$得到的光通量,写成与光源面积以及面到光源中心距离的函数。你可能要查阅积分表——如果你不用查表,值得为自己感到高兴——但是这是极少量具有封闭解的情况之一。如果你对它进行变换从而避开了余弦项查找就比较容易了。

5.8　如图 5.17 所示,一小块表面在单位距离远处注视一个无穷大平面,这个表面块足够小以至于它反射到平面上的光可以忽略。该平面的光通量为 $B(x,y)=1+\sin ax$。表面块与该平面相互平行。我们将表面块沿与平面平行的方式运动,并考虑它在各点的光通量。

(a) 如果移动这表面块,显示它的光通量随它的位置在 x 上周期性变化。

(b) 将表面块的中心固定在(0,0),为表面块该点的光通量确定一个以 a 为函数的封闭解。

5.9　如果你隔着一个大海湾在白天远望,经常会发现难以区别对面的山,而接近黄昏日落,它们则清晰可见。这种现象与空气中光的散射有关——一大片的空气实际上是一个光源。解释这一现象?我们已将空气模型成真空以断言在真空中能量沿直线传播不会丢失。

使用你的解释来估计一下在多大尺度上这个模型是可以采纳的。

5.10 请阅读下面这本书："*Colour and Light in Nature*"，由 Lynch 与 Livingstone 著，Cambridge University 出版社 1995 年出版。

编程作业

5.11 面光源可以近似为由点光源组成的网格。这种近似的弱点是半影区包含量化误差，这对我们的眼睛来说是很讨厌的。

(a) 解释这现象。

(b) 对一个方的光源与一块遮挡物投射阴影到一个无穷大平面的效果进行展示。几何关系不变时，你会发现随着点光源的数量增加，量化误差会下降。

(c) 这种近似具有很不愉快的性质，即对任何有限数量的网格，改变几何关系会产生任意大的量化误差。

5.12 由许多黑色物体与另外一些白色物体构成的场景(纸、胶水与刷子会有用)，观察互反射的效果，你能不能得到一种准则来可靠地从图像上区分它们？(如果你能，将其公布，这个问题看上去容易，其实不然。)

5.13 (这个作业要求有一些关于数值分析的知识)执行重现图 5.17 所需的数值积分。这些积分运算并不是非常简单：如果在无限平面上使用坐标，这个领域的尺度是一个麻烦事；如果将坐标转换成表面块的半球视角，在半球轮廓上的辐射度的频率会趋向无穷。估计积分最好的方法是对半球使用 Monte Carlo 方法。需要使用重要性采样，因为轮廓上的点对积分的贡献要比顶上的小一些。

5.14 一立方体内部的顶部中心有一小的方形光源，建立互反射线性方程组并求解。

5.15 实现一个光度学体视系统。

(a) 它的度量有多准确(也就是与已知的形状信息相比较有多好)？互反射有没有影响它的准确性？

(b) 度量的重复性怎样(也就是如果你很可能在不同照明条件下获得另一组图像，利用它们恢复形式，将两者进行比较结果怎样)？

(c) 将重构最小化方法与积分方法相比较，哪一个更准确、重复性更好？为什么？在实验中这些差别有没有体现出来？

(d) 改进积分方法的一种可行的方法是通过对许多不同表面块积分获得深度，然后取其平均(你需要对常数项加以小心)。这些有没有改进其准确性或可重复性？

第6章 颜　　色

颜色是一个丰富而复杂的体验，这种体验通常由对不同波长的光有不同反应的视觉系统所产生(产生的其他原因还包括眼球上的压力以及梦境)。尽管乍看起来物体的颜色对识别物体来说是一个有用的线索，但是这一点目前还难以应用。

6.1 物理学中的颜色

本节首先扩充辐射度学的词汇，使得这些词汇能够描述与波长相关的量的能量，然后描述彩色表面和彩色光源的典型属性。

6.1.1 彩色光的辐射测量：光谱量

将前面已经描述的所有物理单位通过添加短语每单位波长(per unit wavelength)，可以得到光谱的单位(spectral units)。通过这样的扩充，我们就能描述在能量、双向反射分布函数(BRDF)或漫反射系数中的差异与波长的关系了。在这里不考虑如荧光这样的交互作用，因为这种情况下的能量变化能够改变波长；因此，通过将第4章所定义的物理量增加短语“每单位波长”，就获得了大家所熟知的谱描述量(spectral quantities)。

第一个谱描述量称为谱辐射度(spectral radiance)，通常记为 $L^{\lambda}(\boldsymbol{x},\theta,\phi)$，并且将在波长范围$[\lambda,\lambda+\mathrm{d}\lambda]$所发出的辐射度记为 $L^{\lambda}(\boldsymbol{x},\theta,\phi)\mathrm{d}\lambda$。谱辐射度的单位是 $\mathrm{Wm^{-3}sr^{-1}}$——立方米是由于波长附加的因素。另外，对光源的角度分布不是很重要的那些问题，谱发射度(spectral exitance)是一个合适的属性；光谱发射度的单位是 $\mathrm{Wm^{-3}}$。

类似的，谱双向反射分布函数(spectral BRDF)是发射方向上的谱辐射度与入射方向上的谱辐照度(spectral irradiance)二者的比。由于 BRDF 是由比率所定义的，所以谱 BRDF 的单位是 $\mathrm{sr^{-1}}$。

6.1.2 颜色的来源

建立一个光源通常包含加热一个物体直到它能发光这样一个过程。首先来研究这个过程的理想化情况，然后描述日光的谱功率分布，并且讨论一些人造光源。

黑体辐射器　不反射光的物体——通常称为黑体——是最有效的照明辐射器。一个加热的黑体可发出电磁辐射。值得引起注意的是，辐射的谱功率分布只依赖于被加热体的温度，因此可通过下面的方式建立一个非常好的黑体：获取一个中空的金属，并且在金属上有一个微小的洞能看到金属的空腔——这种结构使得只有很少的一部分光能进入洞内并到达人眼。通过加热，这样的腔可测得热黑体的光谱功率分布。如果用 T 来表示体在开氏温标下的温度，h 表示普朗克常量，k 是麦克斯韦－玻尔兹曼常量，c 是光速，λ 是波长，则可得到

$$E(\lambda) \propto \frac{1}{\lambda^5} \frac{1}{(\exp(hc/k\lambda) - 1)}$$

公式表明,对黑体辐射器存在着一个光颜色的参数族——参数指的是温度,因此可以讨论光源的色温(color temperature)问题。这里的温度指的是那些看起来很相似的黑体的温度。在相对比较低的温度下,黑体是红色的,随着温度的增加,变得越来越白,中间经历橙色直到浅黄。

太阳和天空 最重要的自然光源是太阳,通常把它建模为很远的一个亮点。太阳发出的光被天空所散射。说得更具体些,从太阳离开的光,被天空所散射,到达物体表面,然后反射到摄像机或人眼中。这意味着天空是一个重要的自然光源。把天空描述为具有恒定发射度的半球所形成的光源,是一个粗略的几何模型。发射度恒定的假设是比较差的,因为天空在水平线比在最高点要亮。假设每单位体积的空气发出恒定量的光是天空最自然的模型,这意味着,由于沿着水平线的一条可见光线通过更多的天空,天空在水平线比在最高点要亮。

白天室外物体的表面被两方面的光照射,一部分直接来自于太阳——通常称为日光(daylight),另一部分是被空气散射的阳光[有时称为天空漫反射(skylight)或大气光(airlight);云或雪还可以增加其他的光线]。日光的颜色每年和每天都随着时间而变化(见图 6.1)。这种情形已被广泛研究。

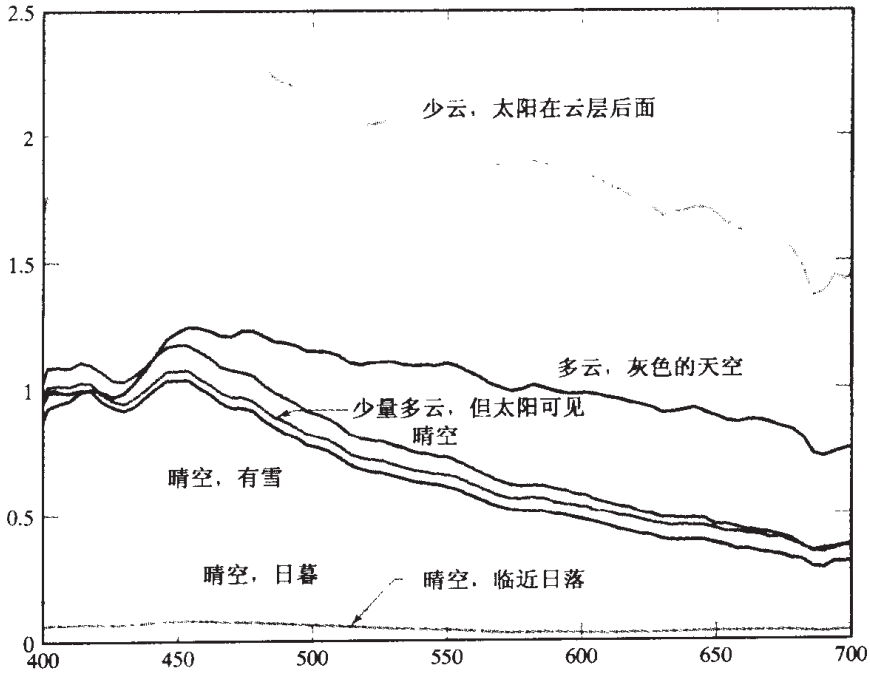


图 6.1 不同时间、不同条件下,白天所测量的日光的相对谱功率存在非常大的变化。图中显示了由 Jussi Parkkinen 和 Pertti Silfsten 提供的 7 种不同的日光测量,测量是通过日光照明钡硫酸盐样本获得的(白色表面反射率高)。数据可由下述地址获得:http://www.it.lut.fi/research/color/lutcs_database.html

对于晴空,每单位体积所散射的辐射强度与频率的 4 次方有关;因此,波长长的光比波长短的光在被散射之前能通过更长的距离(通常称为雷利散射)。由此可见,当太阳升空较高时,

到达地面的阳光中蓝色的光被散射的较多，从而使得太阳看起来更黄一些，而这些蓝光能从天空散射到眼睛中，使得天空看起来是蓝色的。对于天空的光谱辐射，存在一些标准的模型，这些模型刻画了每天不同时刻、位于不同纬度的天空。当天空中有灰尘微粒时，会产生令人吃惊的结果(粒度越大，散射效果越复杂，这通常用 Mie 散射模型进行粗略的描述，这种模型在 Lynch 和 Livingston，2001 或 Minnaert，1993 中有描述)。文献中一位作者描述了对约翰内斯堡日落的生动回忆，这种现象是由矿堆形成的空气中的尘粒产生的。

人工照明　典型的人工光源通常只有几种类型：

- 白炽灯有一个加热至高温的金属灯丝，其光谱大致满足黑体定律。在大部分实用情况下，白炽灯有一点点红色的色调，这是因为元素融化的温度限制了光源的色温。
- 荧光灯通过产生可撞击真空管中气体的高速电子而工作。这个过程反过来可释放紫外线辐射，这种辐射可引起真空管内部附着的磷荧光。通常附着层包括三或四种磷，发出荧光的波长范围很窄。虽然大部分可发荧光的真空管生成具有蓝色调的光，但类似自然光的真空管在不断增多(见图 6.2)。
- 在某些真空管中，电弧在由气体金属和惰性气体所组成的大气中撞击。从活跃状态跌至低能状态的电子可产生光。这种类型的灯的典型特点是某些波长具有很强的辐射，这些波长对应于特别的状态变迁。最常见的是钠弧灯和汞电弧灯。钠弧灯发射黄 - 橙光的效率很高，常用于高速公路的照明。汞电弧灯发射蓝 - 白光，通常用于安全照明。

图 6.2 给出了不同光源谱的样本。

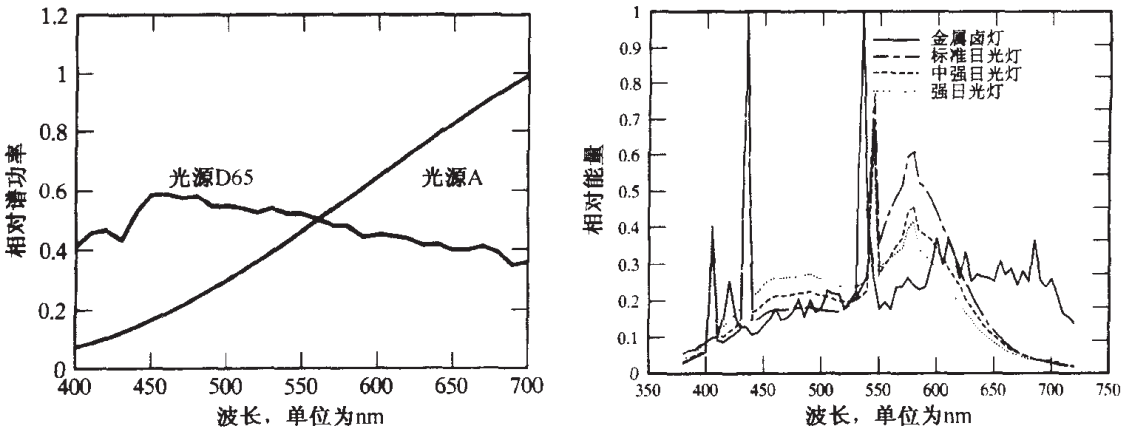


图 6.2　存在多种照明模型；左图显示的是两个标准的 CIE 模型的相对谱功率分布：光源A表示具有色温2800 K的一个100 W的钨丝灯泡；光源D65表示日光；右图给出三菱电子公司的4种不同灯的相对谱功率分布。注意图中又亮又窄的能带，这些能带是由荧光灯中的磷发射的荧光

6.1.3　表面的颜色

表面的颜色受许多因素制约，这些因素包括对不同波长光的吸收不均衡，折射、衍射以及大量散射(更详细的介绍见 Lamb 和 Bourriau，1995，Lynch 和 Livingston，2001，Minnaert，1993 或 Williamson 和 Cummins，1983 中给出的例子)。通常这些效果都包含在一个宏观的双向反射分

布函数(BRDF)模型中,是典型的朗伯加镜面反射的近似;目前称为光谱反射率(spectral reflectance,有时缩写为反射率(reflectance)或(较少用到)频谱反射系数(spectral albedo)。图 6.3 和图 6.4 给出了大量自然物体的光谱反射率。

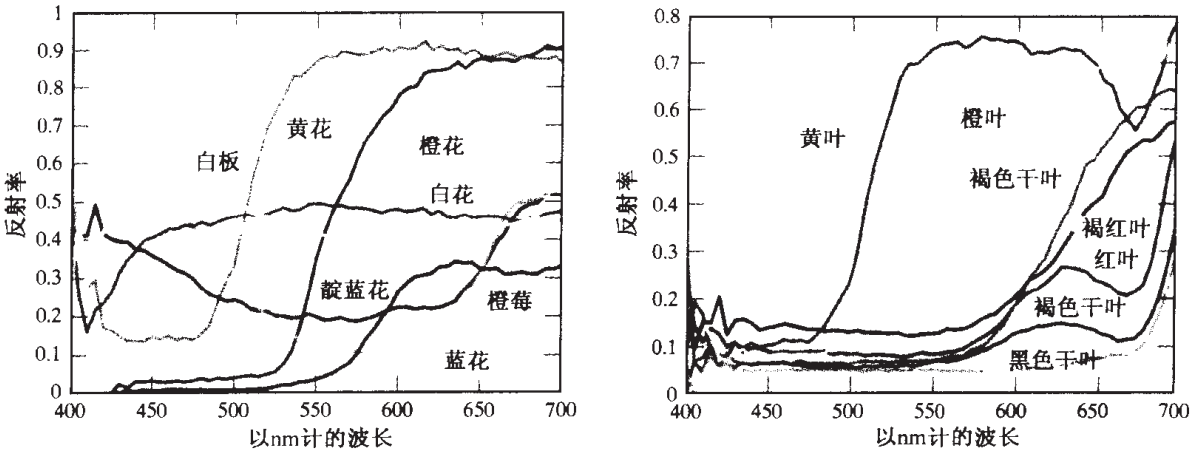


图 6.3 这幅图显示了由芬兰 Kuopio 大学物理系 Esa Koivisto 所测量的大量自然表面的光谱反射率。左图表示的是一系列不同自然表面的光谱反射率——每种都给出一个颜色名。右图表示不同颜色叶子的光谱反射率——同样每种也给出一个颜色名

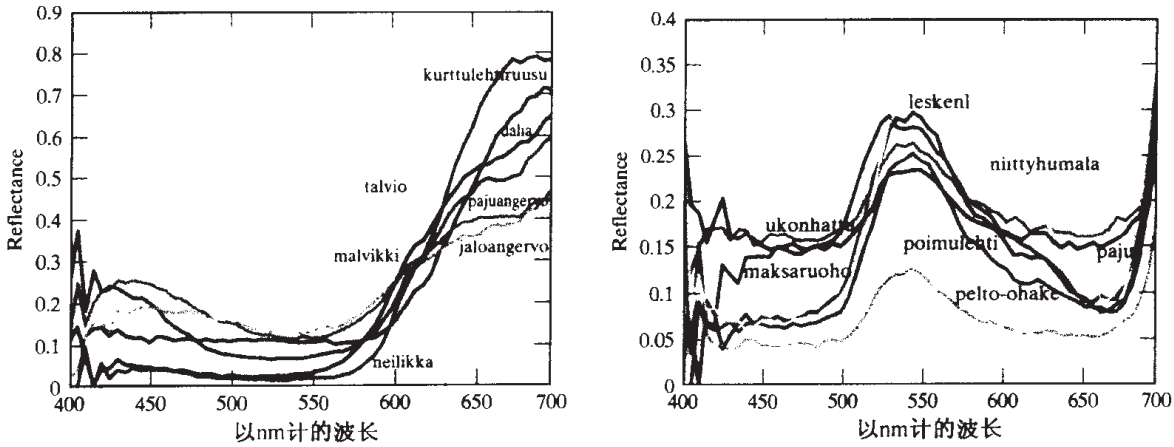


图 6.4 这幅图进一步显示了其他一些自然表面的光谱反射反照率,它们也是由芬兰 Kuopio 大学物理系 Esa Koivisto 测量的。左图表示的是一系列不同红花的光谱反射率,每种都给出一个芬兰名字。右图表示绿叶的光谱反射率,每种也都给出一个芬兰名字。由于只有较少的因素产生植物中的颜色,这些光谱反射率变化不大

进入到眼睛的光的颜色同时受到光照的谱辐射(颜色!)和表面的谱反射率(颜色!)的影响。如果使用朗伯加镜面反射模型,那么有:

$$E(\lambda) = \rho_{dh}(\lambda) S(\lambda) \times \text{geometric terms} + \text{specular terms}$$

其中，$E(\lambda)$是表面的光谱通量，$\rho_{sh}(\lambda)$是光谱反射率，$S(\lambda)$是光谱幅照度。镜面反射的颜色依赖于物体表面——需要谱镜面反射率(spectral specular albedo)这个术语。

颜色和镜面反射 通常，金属的表面有一个与波长相关的光谱分量，例如闪亮的铜钱表面具有黄色的金属光泽。然而不具有传导性的表面(绝缘体表面)的谱分量独立于波长，例如，闪亮的塑料物体表面的镜面反射是光的颜色。6.4.3 节描述了如何使用这些属性来寻找镜面反射，以及对应于金属或塑料物体的图像区域。

6.2 人类的颜色感知

为了描述颜色，首先需要知道人是如何对它们做出反应的。人类对颜色的感知是一个复杂的功能；光照、记忆、物体以及情绪均起作用。最简单的一个问题是去理解在简单的可视条件下人对哪些光谱辐射产生同样的反应(6.2.1 节)。由此产生关于颜色匹配的一个简单、线性的理论，这个理论不仅精确而且对于描述颜色来说非常有用。在6.2.2 节介绍完颜色转换之后我们粗略地给出这种机制。

6.2.1 颜色匹配

在黑色的背景下只有两种颜色可见，这就是颜色感知最简单的情形。在一个典型的实验中，观察者在一个半区看到一束彩色光——测试光(见图 6.5)，然后在另一半区调整光的混合来达到两边区域颜色的匹配。调整的方式是改变混合光中固定数目的原色(primaries)中原色的强度。以这种调整方式，为了获得相应的匹配，需要大量的光来进行实验，但许多不同的调整方式可能得到同样的匹配结果。

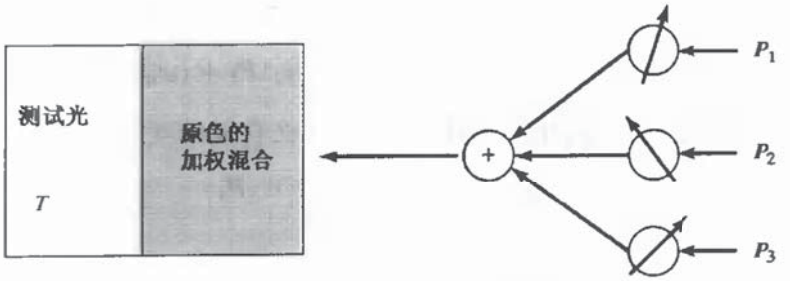


图 6.5 对人类对颜色的感知进行研究，可以要求测试者用混合颜色的方法与测试色进行匹配，这幅图画出这种实验的示意图。测试者看到一个测试光T，将混合器中三种色的量进行调配并显示在测试光旁边。要求测试者调配三种色的量，使混合色与测试色匹配，三原色的混合可以写成$w_1 P_1 + w_2 P_2 + w_3 P_3$。如果混合光与测试光匹配，则有$T = w_1 P_1 + w_2 P_2 + w_3 P_3$。一个值得注意的事实是：对多数人来说，三原色足以匹配许多颜色，而如果允许减法匹配，也就是将某些原色的量附加到测试光来实现匹配的话，则可以匹配所有颜色。有些人需要更少的原色，并且在匹配给定测试光时，大多数人选择相同的混合权重

记 T 表示测试光，等号表示匹配，非负的权重记为 w_i，原色记为 P_i。匹配可表示成如下的代数形式：

$$T = w_1 P_1 + w_2 P_2 + \cdots$$

上述公式的含义是，测试光 T 与某一特定的颜色混合（$w_1, w_2, \cdots$）相匹配。如果在匹配过程中允许减量匹配的话，那么过程可以简化：在减量匹配中，观察者可在测试光中增加某些原色的量来代替匹配。这可通过允许上面表达式的权重取负值，将减量匹配写为代数形式。

三原色 由实验可知，对于大部分的观察者来说，只需要三种原色就可以匹配一个测试光。但需要加以说明以防误解。首先，必须允许减量匹配；其次，原色之间必须是独立的，其含义是没有两种原色的混合可生成第三种原色，这种现象称为三原色原理。通常通过假设在眼中存在三种不同的颜色感受体来解释这个原理。近年来，基因研究中出现一些证据支持这种观点（Nathans, Piantanida, Eddy, Shows 和 Hogness, 1986*a*; Nathans, Thomas 和 Hogness, 1986*b*）。另外，如果给定同样的原色和测试光，大部分观测者会选择同样的原色混合方式来匹配测试光。这种现象通常被解释为三个不同类型的颜色感受体对大部分人来说是差不多的，并且从基因研究所获得的一些直接证据来看也支持这种观点。

Grassman 定律 在所描述的环境中，线性匹配可以得到相当精确的近似，满足 Grassman 定律。

首先，如果我们混合两种测试光，那么将两种测试光对应的匹配混合起来就会得到混合光的匹配。意思是，如果

$$T_a = w_{a1} P_1 + w_{a2} P_2 + w_{a3} P_3$$

且

$$T_b = w_{b1} P_1 + w_{b2} P_2 + w_{b3} P_3$$

那么

$$T_a + T_b = (w_{a1} + w_{b1}) P_1 + (w_{a2} + w_{b2}) P_2 + (w_{a3} + w_{b3}) P_3$$

第二，如果两个测试光可用同样的权重进行匹配，那么它们相互之间也可匹配——即，如果

$$T_a = w_1 P_1 + w_2 P_2 + w_3 P_3$$

且

$$T_b = w_1 P_1 + w_2 P_2 + w_3 P_3$$

那么

$$T_a = T_b$$

最后，匹配是线性的：如果

$$T_a = w_1 P_1 + w_2 P_2 + w_3 P_3$$

那么

$$kT_a = (kw_1) P_1 + (kw_2) P_2 + (kw_3) P_3$$

其中，k 是非负值。

例外情况 给定同样的测试光和原色集合，大部分人使用了相同的加权组合匹配测试光。

因此,三原色原理和 Grassman 定律与那些适用于生物系统中的定律一样,基本上是符合实际的。下面是一些例外情况:

- 由于遗传的原因具有异常颜色系统的人(可能只使用更少的原色去匹配每种情况);
- 由于神经的病变具有异常颜色系统的人(可以表现为各种效果,包括完全丧失对颜色的感知);
- 某些年长的人(由于眼中黄斑色素的增生使得对权值的选择不同于普通人);
- 非常亮的光(对于同样的光来说,色度和饱和度与较暗的光看起来不一样);
- 在非常暗的条件下(颜色转换的机制不同于在较亮条件下的情况)。

6.2.2 颜色感受体

三原色原理表明人眼中颜色的转换方式存在大量的约束。有一种假设可以令人满意的解释颜色感知现象,即假设在人眼中存在与颜色感知有关的三种不同类型的感受体。每种感受体均将入射光转变成神经信号,这些感受体的感光度可以从颜色匹配实验中得到。如果具有不同光谱的两种测试光看起来是一样的,那么它们在感受体上必定产生了同样的效果。

单度量原则 单度量原则(principle of univariance)描述的是这些感受体的活动只有一种类型(例如,它们反应很强烈或很弱,但是并不对接收到的光的波长起反应)。通过解剖光敏感细胞并且测量它们对不同波长的光的反应,或者从颜色匹配反向推理,能够通过一些实验验证这种想法。单度量是个非常好的想法,因为对于人类对彩色光的反应来说,它给我们提供了一个很好且简单的模型:如果感受器对两种光产生同样的反应的话,那么它们之间是匹配的,而不考虑它们的光谱辐射(spectral radiance)。

由于系统的匹配是线性的,感受体也必须是线性的。记 p_k 是第 k 个感受体的反应,$\sigma_k(\lambda)$ 是它的灵敏度,$E(\lambda)$是到达感受体的光,并且 Λ 是可视波长的区域。通过将到达光谱中的每种波长的反应相加,我们可获得一个感受体的全部反应:

$$p_k = \int_{\Lambda} \sigma_k(\lambda) E(\lambda) d\lambda$$

视杆细胞和视锥细胞 视网膜的解剖学研究表明两种类型的细胞对光敏感,这两种细胞由它们的形状来区分。视锥细胞(cone)的光敏感区域的形状大致为锥形,类似的,视杆细胞(rod)的形状为圆柱形。视锥细胞主要对彩色视觉起作用,并且完全占据视网膜中央凹的位置。在对光的敏感性方面,视锥细胞要弱于视杆细胞,从而在暗光下,彩色视觉很弱并且基本上难以分辨(由于视网膜中央凹不工作,不能充分感知空间的精度)。

遗传学中关于彩色视觉的研究支持下列想法:在视网膜上存在三种类型的视锥细胞,它们分别对不同波长范围的光敏感(有证据表明,对于每种类型的视锥细胞,人与人之间存在着一些微小差异)。通过将正常观测者的彩色匹配数据和缺乏一种视锥细胞观测者的彩色匹配数据二者之间进行比较,可以得到三种不同类型的感受体对不同波长的灵敏度。图 6.6 中显示了以这种方式获得的灵敏度。三种类型的视锥细胞称为 S 型视锥、M 型视锥和 L 型视锥(它们分别对短、中和长波长的光敏感)。有时候也称为蓝、绿和红视锥细胞;但是这种名称并不确切,因为对红色的感知并不是由对红色视锥细胞的刺激而引起的,其他两种也是这样。

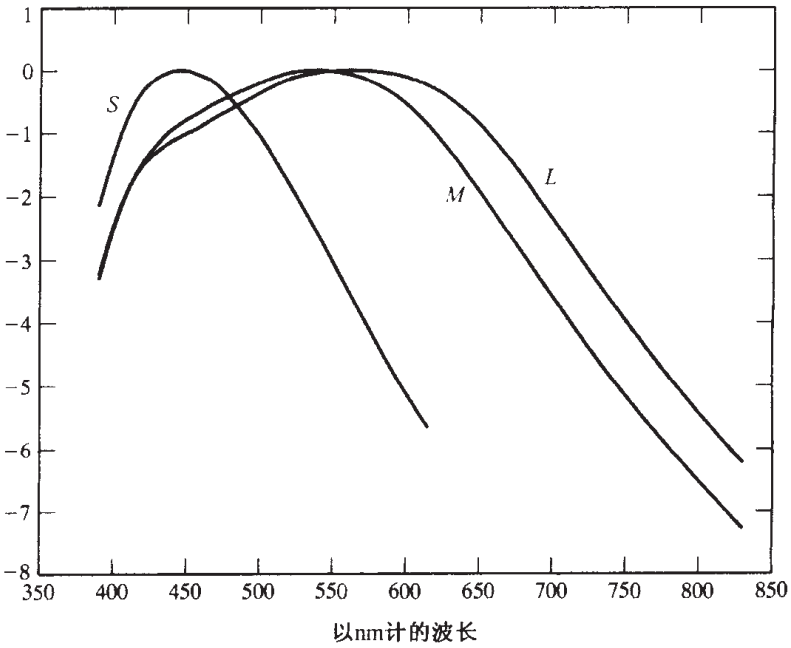


图 6.6 人眼中共有三种颜色感受器，通常叫做视锥细胞。这些感受器对所有光子感受方式相同，但接收光子的数量不同。这幅图是人眼中的三种颜色感受器对相关光谱敏感性的记录。前两种感受器（有时分别叫做红和绿视锥细胞，更恰当的名称是长波感受器和中波感受器）峰值非常接近。第三种感受器（蓝视锥细胞，或者更合适的名称，短波感受器）的峰值同前两种相差很多。可通过以下方式获得感受器对光的反应：将所有波长光谱辐射率和感受器敏感性的共同作用累加起来

6.3 颜色表示

正确描述颜色在商业应用中非常重要。许多产品同特定的颜色密切相关——举个例子，麦当劳的金色拱门，许多流行计算机的颜色和胶卷盒的颜色——制造者希望不同批次生产的产品能够具有相同的颜色，这需要一个关于颜色的标准系统。简单的命名是不够的，因为很少有人记得很多颜色名字，大部分人会把许多种颜色同一个名字联系起来。

6.3.1 线性色彩空间

有一种表示颜色的自然机制：制定一个原色的标准集，然后把所有色光都用这三个原色表示，这三个原色的权重，与人们在匹配颜色时所使用的相同。从原理上看这种用法是很方便的：为了描述一种颜色，我们做一个匹配实验，获取匹配权重。当然，如果我们用标准光照射表面的话（如果表面非常干净），这种方法同样可以扩充来表示表面的颜色。

每次要描述一个颜色时，做匹配实验是可以做到的。举个例子，这是一个油漆店要用到的技术。你拿来一片脱落的油漆，他们通过将不同的颜色油漆混合的方法，调出相同颜色的油漆。油漆店这么做是因为油漆复杂的散射效果使得对混合漆颜色进行预测非常困难。然而，格拉斯曼定律指出，有色光的混合——至少那些可以看到的——是线性的，这意味着可以获得更简单的方法。

颜色匹配函数 在线性方式混合颜色时，对某个已知谱辐射度的光源进行匹配，能够建立一个简单算法决定所确定原色的权重。光源的光谱辐射度可以看做单一波长源强度的加权混合。因为颜色匹配是线性的，与已知谱辐射度的光源匹配的组合，可以首先匹配单一波长对应的三原色值，然后将这些值加权求和得到。

如果有一套用于匹配每种波长所需的原色——一套颜色匹配函数——就能够得到任意光谱辐射率对应的参数值。这种颜色匹配函数——记为 $f_1(\lambda)$，$f_2(\lambda)$ 和 $f_3(\lambda)$——能够通过三原色 P_1，P_2，P_3 由实验得到。本质上讲，调整每种原色的权重来匹配每种波长的单元辐射率，从而为每种波长匹配单位辐射度 $U(\lambda)$ 获得一系列参数。我们把这个过程记做

$$U(\lambda) = f_1(\lambda)P_1 + f_2(\lambda)P_2 + f_3(\lambda)P_3$$

[也就是说，对每种波长 λ，$f_1(\lambda)$，$f_2(\lambda)$ 和 $f_3(\lambda)$ 为这个波长提供权重用于匹配单位辐射度]。

这种光源，记为 $S(\lambda)$，是大量单一波长光的混合，每一种光具有不同的亮度。把每种单一波长的光同原色进行匹配，再把这些匹配权重相加，得到

$$\begin{aligned} S(\lambda) &= w_1P_1 + w_2P_2 + w_3P_3 \\ &= \left\{\int_\Lambda f_1(\lambda)S(\lambda)d\lambda\right\}P_1 + \left\{\int_\Lambda f_2(\lambda)S(\lambda)d\lambda\right\}P_2 + \left\{\int_\Lambda f_3(\lambda)S(\lambda)d\lambda\right\}P_3 \end{aligned}$$

线性颜色空间的一般性问题 线性颜色命名系统能够通过指定原色(获取相应颜色匹配函数)，或者通过指定颜色匹配函数(获取相应暗示原色)实现。这种方式的不便之处在于，如果原色是真实光，则对某些波长至少有一个颜色匹配函数是负的，这并不违反自然规律；这只是意味着无论使用怎样的原色，都要求有负匹配。当然这是令人讨厌的。

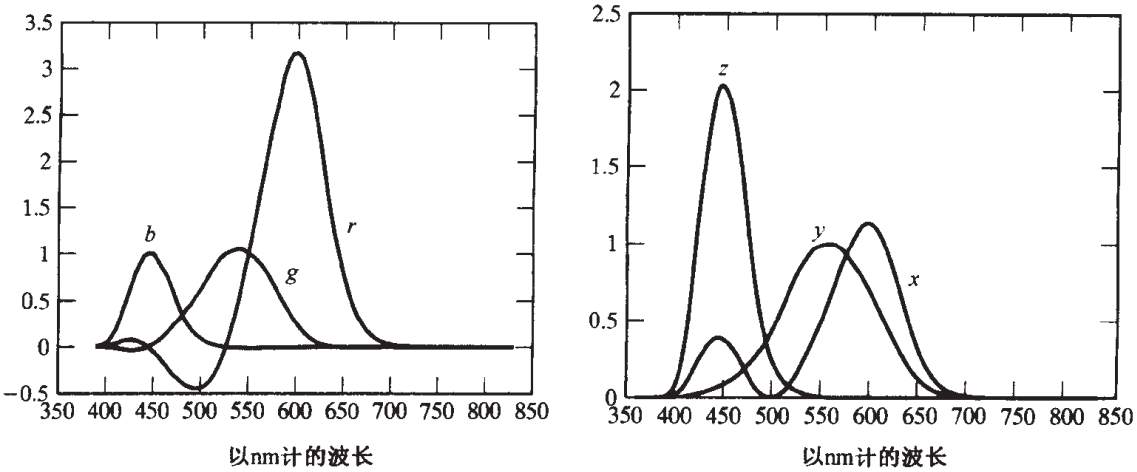


图 6.7 左图是 RGB 三原色系统的颜色匹配函数。负值意味着需要使用对这些波长上的光进行 RGB 三原色参数的负匹配。右图中，颜色匹配函数对应国际照明委员会的X，Y和Z三原色参数；匹配函数总是正的，但是这些原色并不真实

避免这个问题的一个方法是设立一套始终为正的颜色匹配函数，这样导致原色必然是虚构的，因为一些波长的光谱辐射率为负。

尽管看起来是个问题：怎么产生一种具有虚构原色的真实颜色？但实际上并没有问题。因为颜色命名系统很少这样做。通常，我们可以简单地对比权重来判断颜色是否相近，为此知道颜色匹配函数就足够了。国际照明委员会 CIE(commission international d'éclairage，一个制订

标准的机构)已对许多不同的系统实现了标准化。

国际照明委员会 XYZ 颜色空间 国际照明委员会 XYZ 颜色空间是一种非常流行的标准。每个点的颜色匹配函数都是正的,所以任何真实光的坐标总是正的。不可能获得国际照明委员会 X,Y 或 Z 三原色,因为一些波长的光谱辐射度是负的。然而,只要稳定颜色匹配函数,就能够指定一个颜色的 XYZ 坐标并且描述它。

线性颜色空间允许以许多有效的图形学方法进行构造,但在三维空间中画出比在二维空间中画困难许多,所以通常将 XYZ 空间同平面 $X+Y+Z=1$ 相交(如图 6.8 所示)并使用坐标

$$(x, y) = \left(\frac{X}{X+Y+Z}, \frac{Y}{X+Y+Z}\right)$$

画出结果图。这个空间在图 6.9 和图 6.10 画出。一些更有用的解释显示在图 6.11 中。国际照明委员会 xy 坐标在视觉和图形学教材以及一些应用中广泛使用,尽管现在许多研究者认为它已经过时了。

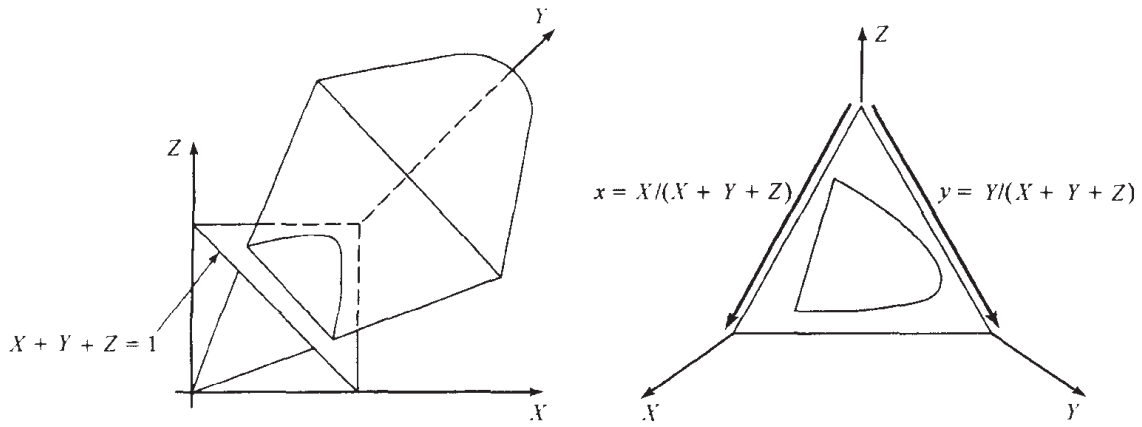


图 6.8 国际照明委员会 XYZ 坐标空间的所有可视颜色值范围,顶点在原点。通常抑制掉颜色亮度方便一些,因为对颜色感知是线性的是一个很好的近似,抑制颜色亮度用平面 $X+Y+Z=1$ 截断圆锥来实现,从而获得图6.9和图6.10所示的国际照明委员会xy空间

RGB 颜色空间 颜色空间通常根据实际需要而发明,所以彼此之间有很大不同。RGB 颜色空间是线性颜色空间,按正规使用单一波长原色(R 是 645.16 nm,G 是 526.32 nm,B 是 444.44 nm,见图 6.7)。一般将显示器上所使用的磷光体作为 RGB 的原色。一般将可得到的颜色表示成一个立方体,通常称为 RGB 立方体,边缘代表 R,G,B。立方体见图 6.12。

CMY 和 Black 手写时代人们的直觉建议,基本颜色应该是红、黄和蓝;红和绿混合得到黄。这个直觉不能用于显示器的原因同涂料有关(有负的混合)而不是因为光。涂料吸收了入射光中的其他颜色再从纸上反射出来。于是,红墨水实际上是吸收了绿色光和蓝色光的染料——入射红光穿过这染料从纸面反射回来。

这样的负映射的颜色空间会非常复杂。在最简单的情况下,混合是线性的(或者接近线性)而且使用 CMY 空间。这个空间中有三个参数:青色(蓝绿色)、洋红(略带紫色的颜色)和黄色。这些参数应该视为白光的减色参数:青色是白色减去红色,洋红是白色减去绿色,黄色是白色减去蓝色。混合色的表现可以根据 RGB 颜色空间得到。例如,洋红和青色混合得到

$$(W-R)+(W-G)=R+G+B-R-G=B$$

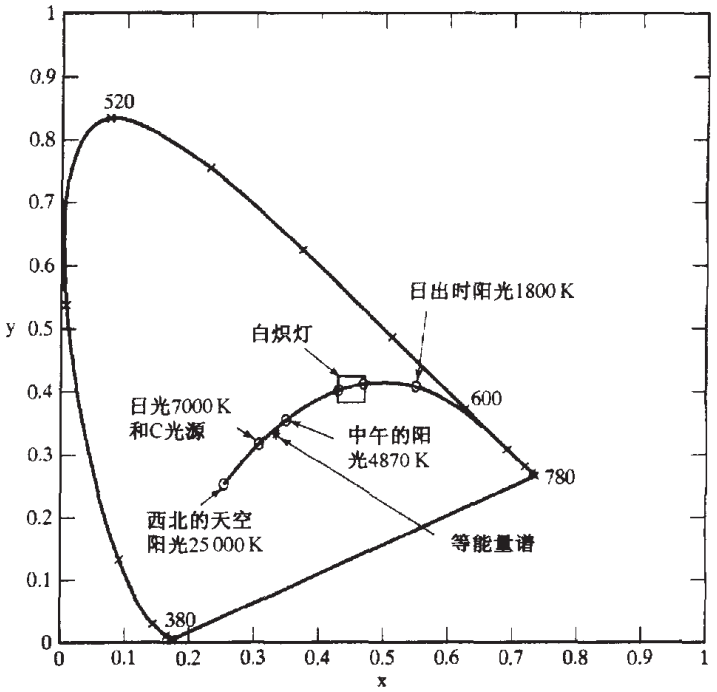


图 6.9 本图给出了国际照明委员会 1931 标准 xy 颜色空间的恒常亮度切图。这个空间有两个坐标轴。图中曲线边界通常称为谱轨迹，表示了单一波长的光所感受到的颜色。图中给出了不同温度下黑体辐射的轨迹和天空不同颜色的轨迹。图标的中心位置是一个中性点，形成该点颜色的三原色的权重相同。国际照明委员会选择这样的原色使得这种光是消色差的。一般说来，远离中性点的颜色更加饱和(深红和浅粉色的区别)，而围绕中性点转动时有色调的变化(绿和红的区别)

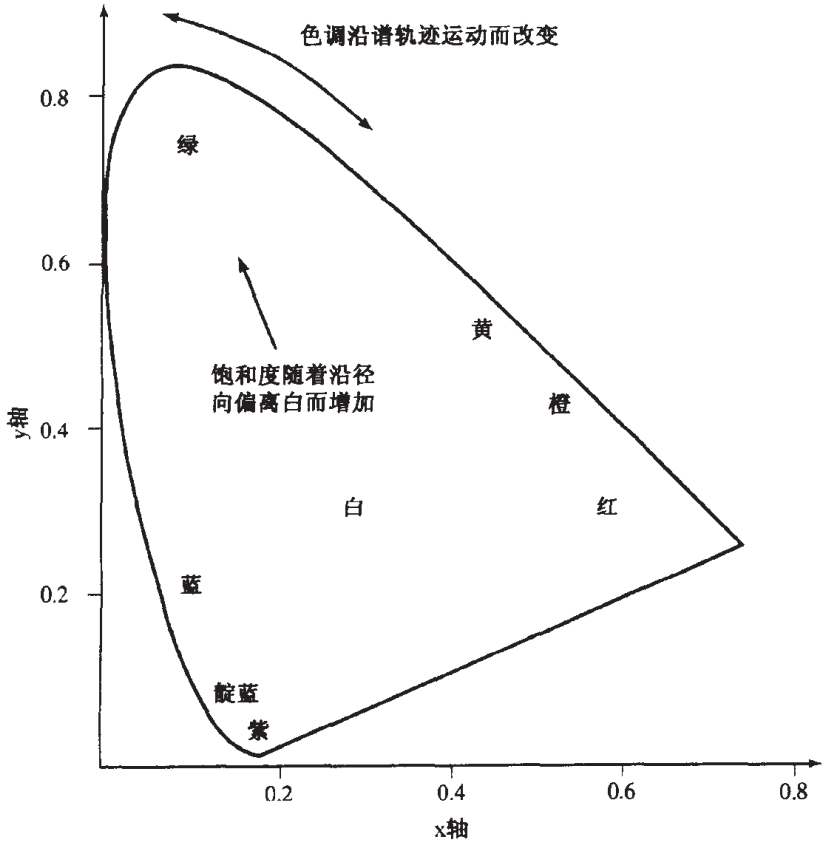


图 6.10 本图给出了国际照明委员会 1931 标准 xy 颜色空间的恒常亮度切图。颜色名称在图中标出。通常，远离中性点的颜色更加饱和(深红和浅粉色的区别)。围绕着中性点转动时有色调的变化(绿和红的区别)

这是蓝色。注意到 $W+W=W$,因为假设不管怎么涂改,墨水都不能让纸看起来更亮。实际印刷设备使用最少4种墨水(青色、洋红、黄色和黑色),因为混合墨水只能产生很淡的黑色,三种颜色的配色方案也会因为对不准使文字周围产生有色晕圈,而且彩色墨水比黑墨贵很多。要想获得很好效果的彩色印刷是很困难的:不同的墨水光谱特性不同,不同的纸张也有不同的光谱特性,墨水的混合也是非线性的。

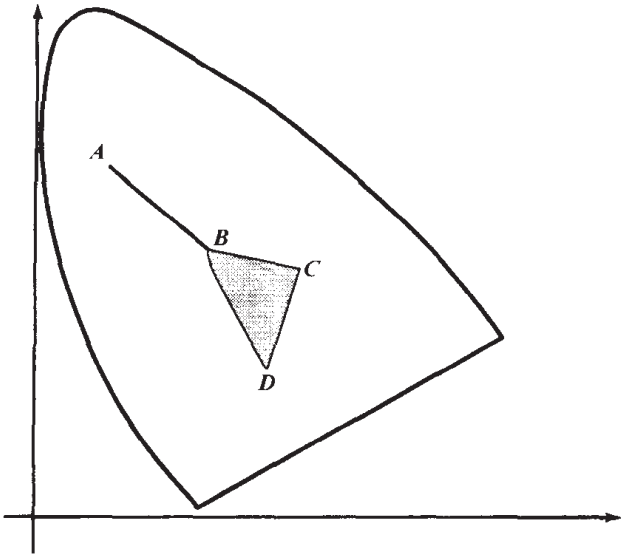


图 6.11 颜色系统的线性模型允许许多有用的配置。如果有两个光,它们的国际照明委员会坐标是 ***A*** 和 ***B***,通过 ***A*** 和 ***B*** 的非负组合获得的所有的颜色能够用 ***A*** 和 ***B*** 的连线表示,而可以由 ***B***,***C*** 和 ***D*** 混合得到的颜色在三个点构成的三角形内。在设计显示器时这一点是很重要的——每个显示器有三种磷光粉,每种磷光粉越饱和,就越能显示更丰富的图像。这也解释了为什么同样的图像在不同的显示器上看起来不完全相同。谱轨迹的曲率说明了如果没有负匹配,不存在三个真实存在的原色能够显示所有的颜色

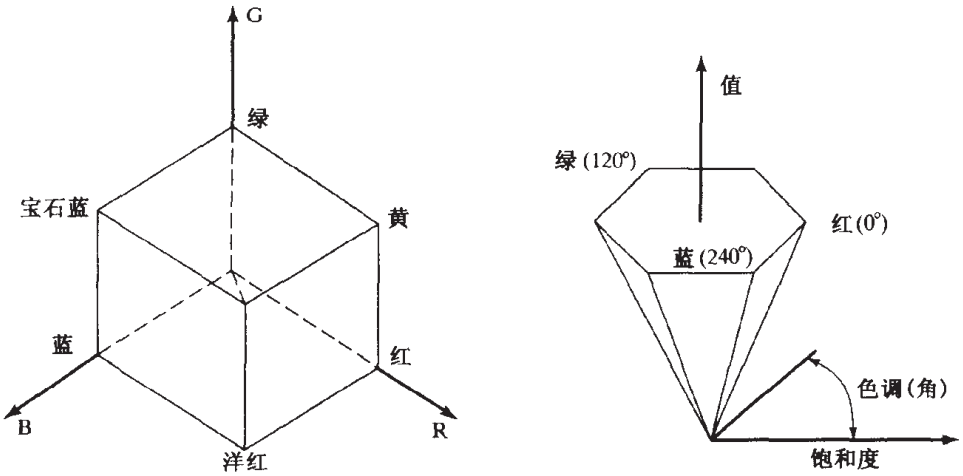


图 6.12 左图是 RGB 立方体;使用三个原色(R,G 和 B——通常由显示器的颜色定义)能够获取的所有颜色(权重值在0和1之间)的空间。沿着从原点到点(1,1,1)的中性轴观察这个立方体,可以看到图中的六边形。这个六边形将色调用角度表示(从红到绿过程中颜色性质的改变),可以在直觉观察上得到。右图中有一个圆锥,圆锥上的点到水平面的距离是颜色的亮度,圆周上的旋转角度是色调,离开中心轴的距离是颜色的饱和度

6.3.2 非线性颜色空间

线性空间的颜色坐标可能并不是必要的编码属性,虽然其在常用语言或者实际应用中是非常重要的颜色属性。有用的颜色术语包括:色调——从红过渡到绿的过程中颜色的改变,饱和度——从红过渡到粉红的过程中颜色性质的改变,亮度(有时叫做光亮度)——从黑过渡到白的过程中颜色性质的改变。例如,如果我们对检测一个颜色是否位于特定的红色区域内感兴趣,我们希望直接检测颜色的色调。

线性颜色空间的另一个困难在于,坐标不符合人类对颜色拓扑的直觉;通常直觉认为色调形成一个圈,色调从红色到橙色到黄色到绿色,再到青、蓝、紫,最后回到红。另一个想法是局部色彩关系:红色在紫色和橙色之间,橙色在红色和黄色之间,黄色在橙色和绿色之间,绿色在黄色和青色之间,青色在绿色和蓝色之间,蓝色在青色和紫色之间,紫色在蓝色和红色之间。这些局部关系中的每一个都成立,但整体来看色调就形成了一个圈。这意味着线性颜色空间的坐标无法模拟色调,因为它的坐标的最大值和最小值相差很远。

色调、饱和度和亮度 处理这个问题的一个标准方法是构造一个颜色空间来反映这些关系,这可以通过使用一个到 RGB 空间的非线性变换而得到。有很多种这样的空间,一种叫做 HSV 空间(色调、饱和度和亮度),通过沿 RGB 立方体的中心轴往下看得到。因为 RGB 是一个线性空间,所以用偏离原点的尺度代表亮度。可以把 RGB 空间平展成二维的常亮度空间,并变形为一个六边形,得到图 6.12 所示的结构。其中,色调通过沿着中心点旋转改变的角度得到,同时饱和度随着点远离中心点而改变。

在线性颜色空间之间,或者线性与非线性颜色空间之间(Fairchild,1998 是很好的参考),有许多可能的其他坐标变换。使用一套坐标代替另一套很难有明显的改进(尤其是如果坐标系统间只是一对一的转换),除非考虑到编码、比特率或者感知的一致性。

均匀颜色空间 通常无法准确地重构颜色。这意味着知道人眼对颜色差异的区分能力是重要的;通常比较微小的颜色区别的影响是有用的,而试图比较大的颜色区别通常是困难的,譬如回答"蓝色和黄色之间的区别大,还是红色和绿色的区别大?"之类的问题。

要确定略能察觉到的颜色差异,可以通过修改观察者看到的颜色,直到他们发现同原始颜色不同时为止。将这些区别绘制到颜色空间中,就形成了与原颜色无法区分的颜色区域的边界。通常将可察觉的差异表示为椭圆。在国际照明委员会 xy 空间,这些椭圆取决于它们在空间发生的不同位置,如同图 6.13 中所示的麦克亚当椭圆。

这意味着在(x,y)坐标中,用$((\Delta x)^2+(\Delta y)^2)^{(1/2)}$给出略能察觉的颜色差异的尺度只是一个很弱的指示器(如果它是一个很好的指示器,那么椭圆代表的无区别颜色区域将是一个圆)。在均匀颜色空间中,坐标空间中的距离反映了两种颜色不同的程度——在这样的颜色空间里,如果坐标空间的距离在某些阈值之下,人眼将无法区分这些颜色。

一个更均匀的颜色空间能够从国际照明委员会的 XYZ 坐标,通过使用投影变换校正椭圆得到。这就得到了一个在图 6.14 中说明的国际照明委员会 $u'v'$空间,其坐标是:

$$(u',v')=\left(\frac{4X}{X+15Y+3Z},\frac{9Y}{X+15Y+3Z}\right)$$

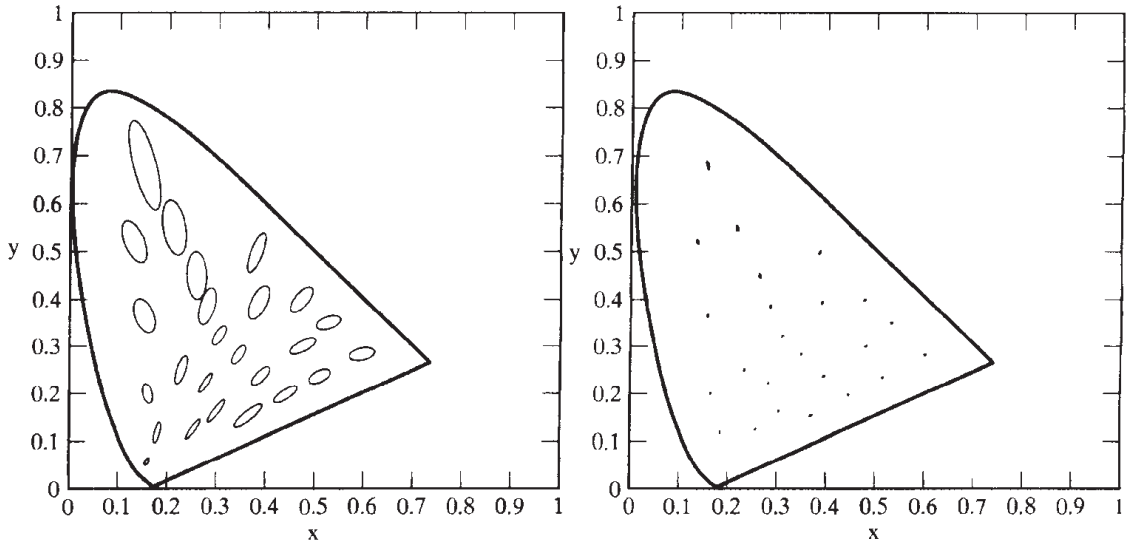


图 6.13 这幅图显示了国际照明委员会 x,y 空间中颜色匹配的变化。检测光在椭圆的中心,椭圆的尺寸代表人类能够区别的光的范围,边界表示的是能察觉到的颜色区别。为了能看清楚,左图中的椭圆幅值扩大了10倍,右图中的椭圆是原始尺寸。这些椭圆称为麦克亚当椭圆。注意到上面的椭圆比下面的要大,上升过程中椭圆也在旋转。这意味着 x,y 坐标差异的幅度很难说明颜色的差异

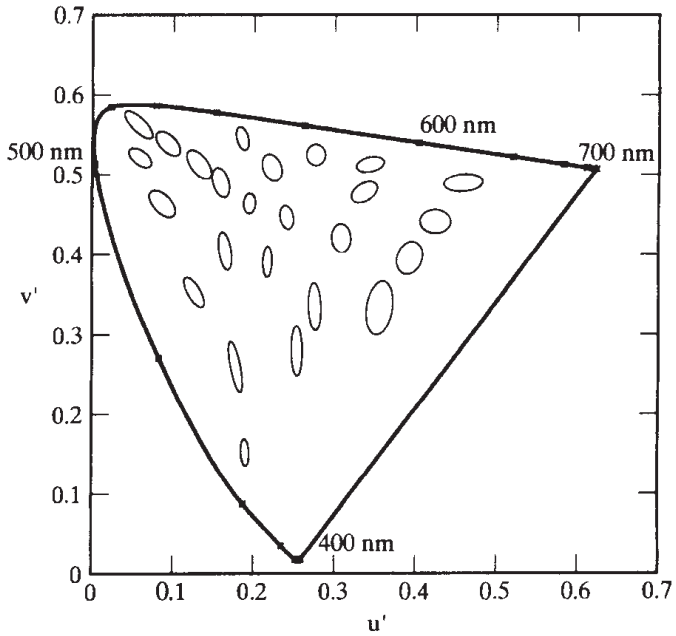


图 6.14 这幅图显示了国际照明委员会 1976 年 u′v′空间,它是通过对国际照明委员会 x,y 空间进行投影变换得到的。目的是使麦克亚当椭圆变成圆——这就形成了一个均匀颜色空间。许多非线性变换能够使这个空间更加均匀(细节参见Fairchild(1998))

通常 u′,v′空间坐标间的距离是区分两种颜色的相当好的指示器,但是这里忽略了亮度上的区别。国际照明委员会 LAB 是现在最通用的均匀颜色空间。LAB 中颜色的坐标通过对 XYZ 坐标的非线性变换得到:

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16$$

$$a^* = 500\left[\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}\right]$$

$$b^* = 200\left[\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}}\right]$$

其中，X_n，Y_n，Z_n 是基准白标的 X，Y，Z 坐标值。之所以考虑 LAB 空间是因为该空间实质上是均匀的。在一些问题中，重要的是人类观察者能够注意到的颜色区别，LAB 坐标中的区别能够给出很好的帮助。

6.3.3 空间和暂时效应

预测复杂颜色的（也就是说，比一对光更令人感兴趣的刺激）组合的感观是非常困难的。如果视觉系统在一段时间内曝光在一个特定发光体前，则将导致颜色系统产生适应性——这个过程叫做色适应。适应导致颜色结构图歪斜，对两个适应了不同发光体的观察者来说，它们可能将具有不同染色性的光谱看成同样的颜色。适应也可能由视场中的表面片造成。其他的重要机制包括同化（使得表面片的色彩在周围色彩的影响下趋向周围的色彩）和对比（使得表面片的色彩在周围色彩的影响下远离周围色彩）。这种效果与视神经以及色彩恒常性中的编码问题相关。

6.4 图像颜色的一个模型

为了解释摄像机获得的颜色值，需要理解摄像机做了什么以及希望对什么样的物理效果建模。我们的模型支持好几种非常简单但功能强大的推理算法。

6.4.1 摄像机

大部分彩色摄像机都具有一个简单的成像装置。在每一个传感器单元，三个滤波器中的每一个提供所要的感光函数（粗略地说，就是红色、绿色和蓝色）。这些滤波器按照马赛克式的镶嵌形式排列，使用的模式也是各种各样的。然后将 CCD 摄像机的输出进行处理，从而重构出完整的红绿蓝图像。当然，在该过程中某些信号的信息会丢失，丢失的信息取决于具体的摄像机和镶嵌方式。特别的，这些信息丢失并不影响观察者所感知的图像的质量，但是它们会极大地影响各种空间计算——比如，边缘检测算法定位边缘的能力（因为亮度的空间分辨率可能并非所看到的那样）。

如果 CCD 单元系统和滤波器具有一个在 CIE XYZ 颜色匹配函数的线性变换领域内的感光函数，那么就很令人满意了，该属性意味着摄像机使用了与人相同的方式进行颜色匹配。这个要求似乎非常苛刻，在实际中是难以达到，因为实验证据表明大部分摄像机并不具备这一可取的属性。

CCD 摄像机本质上是线性设备。然而，大部分摄像机都用在拍摄影片中，因而趋向于压缩输入的动态范围（范围高端的亮度差别被减少了，低端也是如此）。线性设备的输出总是看起来很粗糙（暗的部分太暗，而亮的部分太亮），因此设备制造商对输出使用了各种形式的压缩算

法。最常用的是伽马校正。这是一种最初用于对付监视器的非线性的压缩形式，特别地，如果电子枪的输入电压为 V_{in} 监视器的亮度为 V_{in}^{γ}（CRT 监视器的 γ 值一般是 2.2）。在大部分计算机显示设备中，提供给电子枪的电压是帧缓冲区中的值的线性函数。意思是，如果我们希望亮度为 I，那么帧缓冲器中的值就与 $I^{1/\gamma}$ 成正比。一般摄像机是经过伽马校正的，从而摄像机提供的值可直接放入帧缓冲区以获得相应的亮度，因此摄像机的输出不是输入的线性函数。

曾经有段时间，CCD 摄像机带有一个分离的控制电路盒，该电路盒带有一个开关，可用于关闭非线性计算；但那些快乐的日子已经过去了。现在的摄像机的输入－输出关系一般需要校正，因为可能存在各种各样的非线性计算（Barnard 和 Funt，2002；Holst，1998；或者 Vora，Farell，Tietz 和 Brainard，1997）。偶尔也会用一些冒险的手段去修正摄像机电路，这种手段怕事者一般不会用。在以后的讨论中，假设摄像机的输入输出关系已知，并且近似线性。

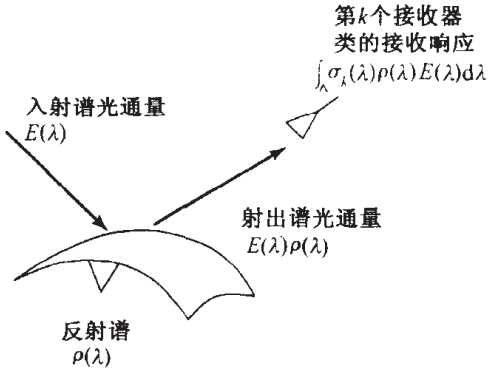


图 6.15 如果用一个光谱为 $E(\lambda)$ 的光源照射一个漫反射谱反射率 $\rho(\lambda)$ 的标准漫反射的表面，反射光的光谱是 $\rho(\lambda)E(\lambda)$（乘以与表面方向有关的某些常数，此处不予考虑）。因此，如果第 k 类的线性感光器看到了该表面片，其响应将是 $p_k = \int_\Lambda \sigma_k(\lambda)\rho(\lambda)E(\lambda)d\lambda$，其中，$\Lambda$ 是相关的波长的范围，并且 $\sigma_k(\lambda)$ 是第 k 个感光器的感光度

6.4.2 图像颜色的一个模型

到达摄像机的光的颜色由两个因素决定：第一是光所离开的表面的谱反射率，第二是落于该表面的谱辐射度。如果一个谱漫反射率为 $\rho(\lambda)$ 的标准漫反射表面被一个光谱为 $E(\lambda)$ 的光源所照射，那么反射光的光谱是 $\rho(\lambda)E(\lambda)$（乘以与表面方向有关的某些常数，此处不予考虑）。因此，如果第 k 个线性感光器看到了该曲面片，其响应将是（见图 6.15）：

$$p_k = \int_\Lambda \sigma_k(\lambda)\rho(\lambda)E(\lambda)d\lambda$$

其中，Λ 是相关的波长的范围，并且 $\sigma_k(\lambda)$ 是第 k 个感光器的感光度。

到达物体表面的光的颜色变化范围比较广泛，从室内的蓝色荧光灯到暖橙黄色的钨灯，到太阳光中的橙黄色甚至红色光，因此到达摄像机的光的颜色不能很好地表示所注视表面的颜色（图 6.16，图 6.17，图 6.18 和图 6.19）。

不考虑第 5 章所讨论的物理模型的某些细节，可将摄像机某一像素的值建模为

$$\boldsymbol{C}(\boldsymbol{x}) = g_d(\boldsymbol{x})\boldsymbol{d}(\boldsymbol{x}) + g_s(\boldsymbol{x})\boldsymbol{s}(\boldsymbol{x}) + \boldsymbol{i}(\boldsymbol{x})$$

在该模型中，

- $d(x)$是在同一光照射下一个等价的平坦正面的表面图像颜色
- $g_d(x)$是随空间而变化且对因表面方向改变而导致亮度的变化的项
- $s(x)$是来自等价的平坦正面的表面的镜面反射的图像颜色
- $g_s(x)$是随着空间而变化且表示了镜面反射能量变化的项
- $i(x)$是一个表示颜色互反射和光照空间变化等量的项

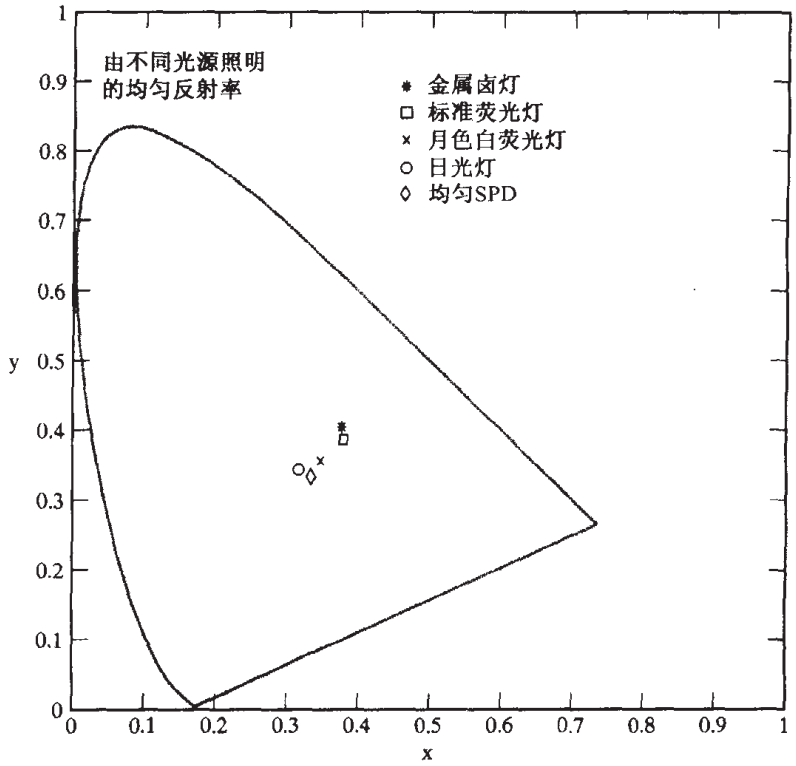


图 6.16　光源可以具有非常广泛的颜色范围。本图展示了图 6.2 中的 4 个光源的颜色，与均匀的谱功率分布的颜色的比较，本图使用CIE x，y坐标进行绘制

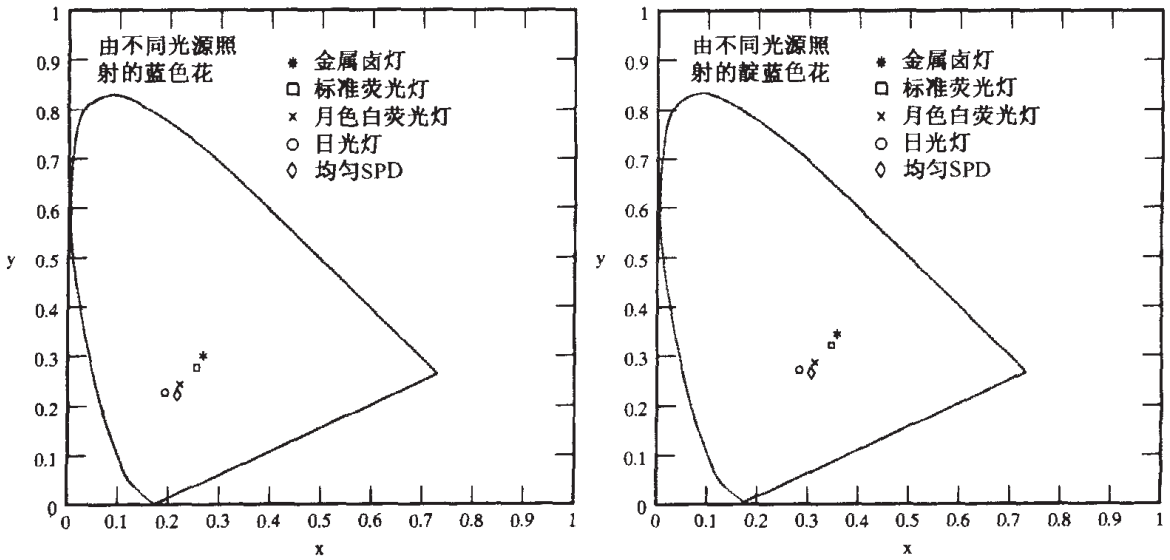


图 6.17　在不同的光照射下表面具有明显不同的颜色。这些图像显示了图 6.3 中的蓝色和靛蓝色花在图 6.2中的不同光源，以及在均匀谱功率分布下看到的颜色

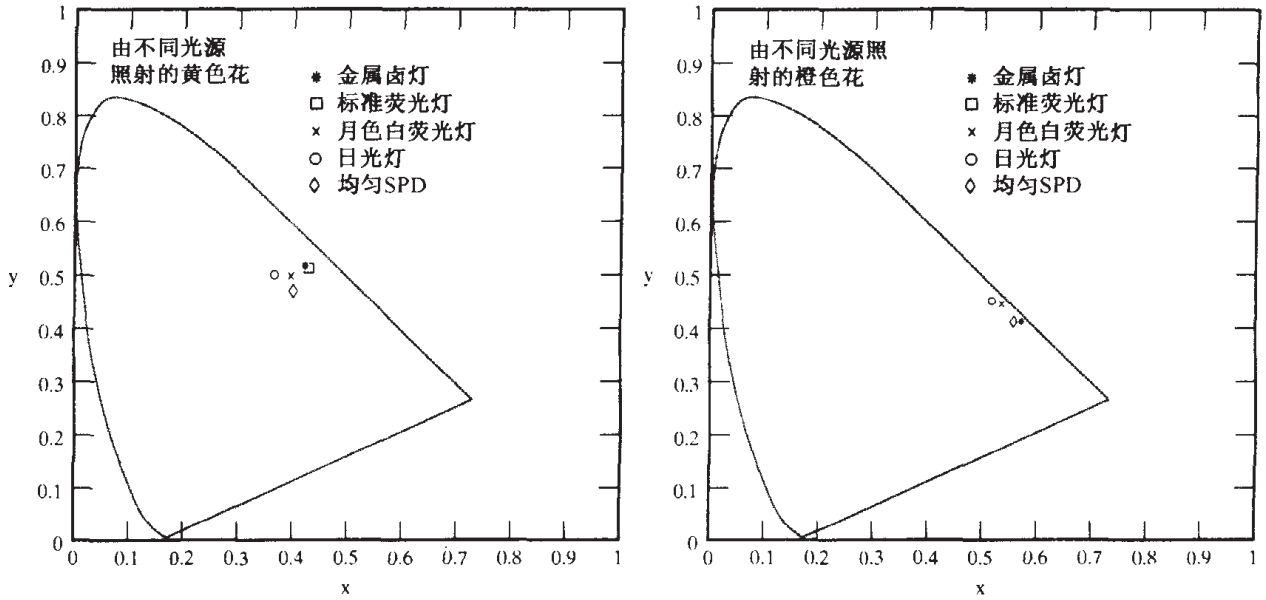


图 6.18 在不同的光照射下表面具有不同的颜色。这些图像显示了图 6.3 的黄色和橙色花在图 6.2 中的不同光源，以及在均匀谱功率分布下的颜色

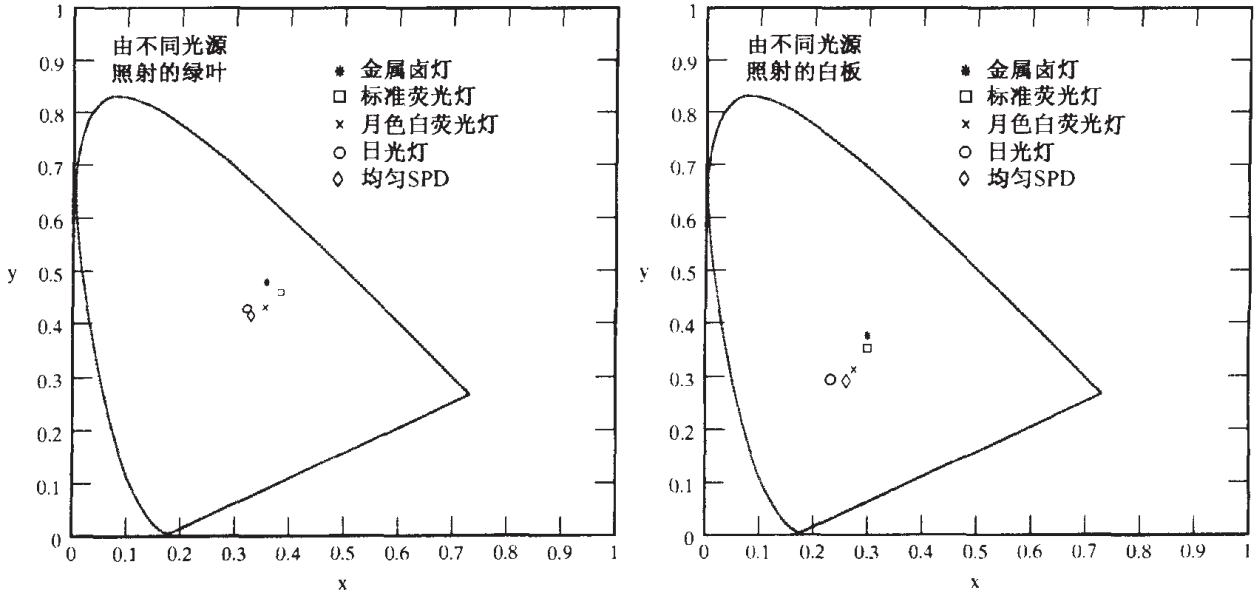


图 6.19 在不同的光照射下表面具有不同的颜色。这些图像显示了图 6.3 的白板以及图 6.4中一片绿叶在图6.2中的不同光源，以及在均匀光谱能量分布下的颜色

由于主要关心的是在局部范围内从颜色中提取信息，因此忽略项 $g_d(\boldsymbol{x})$和 $\boldsymbol{i}(\boldsymbol{x})$ 的细节结构。对于如何从 $\boldsymbol{i}(\boldsymbol{x})$ 中提取信息，目前一无所知，所有的证据都表明这种提取是很困难的。有时候该项相对于其他项是非常小的，并且在空间常常缓慢变化。我们在处理中将忽略该项，因此必须保证它很小(或者它的存在对我们的算法影响不会太大)，而镜面反射表现为小而亮的区域，可以被找到。

6.4.3 应用：找到镜面反射

镜面反射对于物体的外观具有很强的影响，它们一般表现为小的亮片，称为高光。高光对

于人类对表面特性的感知具有重要的影响;在图像上附加小的、类似于高光的区域可以使被描绘的物体看起来光滑明亮。镜面反射常常太亮,以至于摄像机饱和而无法度量其颜色。然而,由于镜面反射的外观是强约束的,并且有许多有效的方案来标示它们,因此检测的结果可以用于提供形状的线索。

实际所见的反射率的动态范围相对较小,具有很高和很低的反射率的表面是难以制造的。均匀的光照也很常见,同时,大部分摄像机在它们的操作范围内可合理地接近为线性。所有这些意味着很亮的区域不会由漫反射产生,它们要么是光源(这种或者那种形式的,可能是有光在背后的有污点的玻璃窗),要么是镜面反射。此外,镜面反射一般较小。因此,检测小的亮片是找到镜面反射的有效方法(Brelstaff 和 Blake,1988)。

在彩色图像中,绝缘体和导体物质上的镜面反射常常看起来明显不同。这与导电性有关,因为电场不能穿透导体(内部的电子只是不断移动来取消电场),从而打在金属表面的光或者被吸收或者被镜面反射出去。阴暗的金属表面看起来阴暗是因为其表面粗糙,而闪亮的金属表面具有特定色彩的闪亮片,是因为导体对不同的波长吸收不同量的能量。然而,入射在绝缘体表面的光,却可以进入表面。

许多绝缘体的表面可以建模为一个镶嵌有随机分布色素的透明矩阵,这对于塑胶和某些油漆而言是一个很好的模型。在该模型中,存在两种反射成分,分别对应于镜面反射和漫反射概念:体反射,产生于进入矩阵的光,撞击各种色素颜料,然后离开表面;表面反射,产生于表面的镜面反射。假设色素是随机分布的(很小,而且不在表面上)且矩阵是合理的,体反射成分的作用类似于漫反射,谱反射率取决于色素的成分,而表面成分与波长无关。

假设注视一个单色的绝缘体。设想互反射项可以忽略不计,则摄像机的像素亮度模型变成

$$\boldsymbol{p}(\boldsymbol{x}) = g_d(\boldsymbol{x})\boldsymbol{d} + g_s(\boldsymbol{x})\boldsymbol{s}$$

其中,$\boldsymbol{s}$ 是光源的颜色,$\boldsymbol{d}$ 是漫反射光的颜色,$g_d(\boldsymbol{x})$ 是一个与表面方向有关的几何项,而 $g_s(\boldsymbol{x})$ 项给出镜面反射的范围。如果物体是弯曲的,那么 $g_s(\boldsymbol{x})$在曲面的大部分地方都很小,仅仅在镜面反射处很大:$g_d(\boldsymbol{x})$随表面方向缓慢变化。现在把表面产生的颜色映射到接收器的响应空间,看一看其表现出的结构情况(见图 6.20)。

由于 $g_d(\boldsymbol{x})\boldsymbol{d}$ 表示的是与接收器的响应相同的向量乘以随着空间变化的常数项,所以将生成的直线延长后会通过原点。如果存在镜面反射,那么就又有一条对应于项 $g_s(\boldsymbol{x})\boldsymbol{s}$ 的直线,一般它不会通过原点(由于漫反射项)。对应的是一条直线,而非一个平面区域,这是因为 $g_s(\boldsymbol{x})$仅仅在表面法线的一个很小的范围内较大。之所以如此设想,原因在于表面是弯曲的,$g_s(\boldsymbol{x})$只对应于表面很小的一个区域,其中项 $g_d(\boldsymbol{x})$ 在该区域应该大致不变。我们期望镜面反射区域是条直线,而不是孤立的像素,因为我们期望曲面有(也许很窄)镜面反射带,也就说是镜面反射系数有一个取值范围。第二条直线可能会与有色立方体的表面相交,并且被截断。

所得到的"狗腿"(dog-leg)模式几乎立即能导出镜面反射标记算法:找到这个模式,然后找到镜面反射线。该线上的所有像素都是镜面反射像素,漫反射项和镜面反射成分可以很容易地估计出来。为了使算法更有效,我们必须确保像素集合只表示一个物体。可使用图 6.21 展示的局部图像窗口来判断。即使表面不是单色的,该方法依然成立——比如一个有图片的咖啡杯子,但是在颜色空间中寻找上述结构仍是一个困难的问题,据我们所知目前尚未解决。

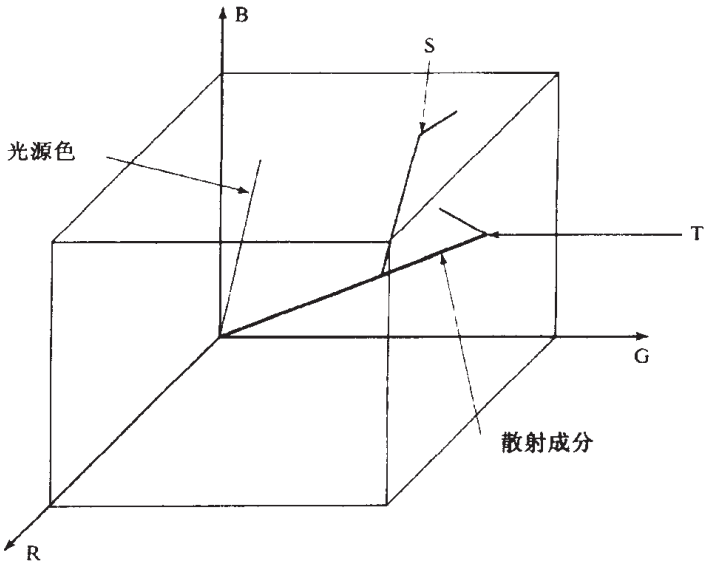


图 6.20 假设有一幅颜色均匀分布的单色表面图。我们的光反射模型会产生如图所示的谱分布范围。假定反射光包括漫反射项和镜面反射项，而且镜面反射项就是光源的颜色。曲面上的大部分点不具备较强的镜面反射项，而是相同的较亮或者较暗的漫反射颜色值。在某种点，镜面反射项很大，这就在谱分布范围中产生了“狗腿”(dog-leg)模式，因为在漫反射项中加入了光源项。如果漫反射很亮，有一两个颜色通道就会饱和(点T)；类似的，如果镜面反射很强，同样有一两个颜色通道会饱和(点S)

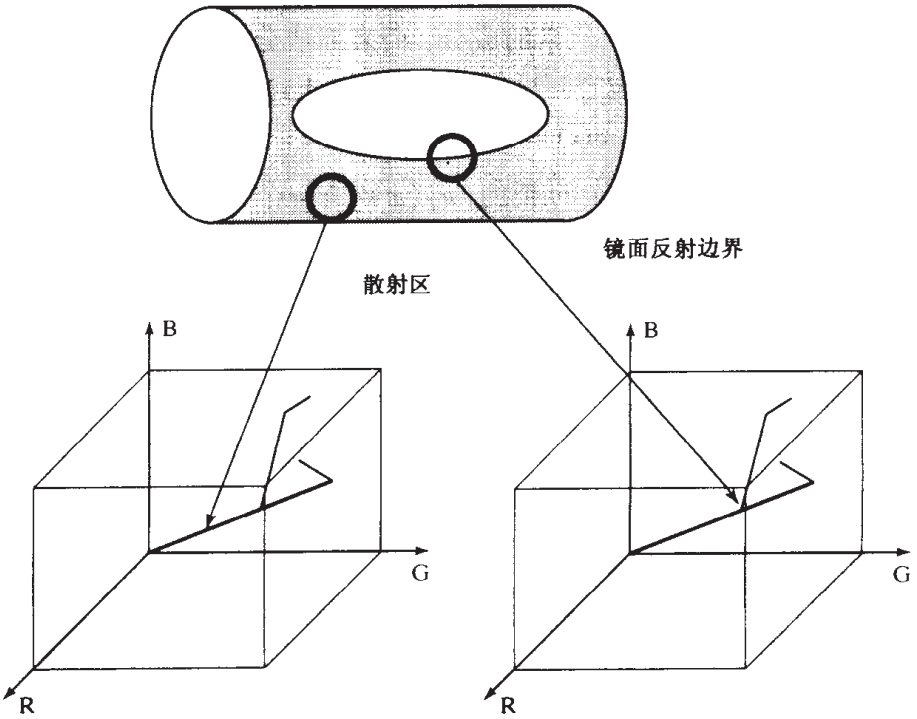


图 6.21 塑胶物体表面的镜面反射引起的线性聚类，可以通过对图像窗口内像素进行推理来获得。在一个背景为黑色、由塑胶物体组成的世界里，一个背景窗口的像素在颜色空间中大概对应于一点，因为所有的像素具有相同的颜色。面位于物体表面上的窗口在颜色空间中产生了一个类似直线聚类的点，因为它们的亮度有变化，但是颜色没有变。在镜面反射区的边界处，窗口内产生了类似于平面的聚类，因为点是两种不同颜色的加权组合(镜反射颜色和体颜色)。最后，在镜反射区域的内部，窗口可以产生类似于体的聚类，这是因为摄像机饱和，以及窗口的范围同时包含边界和饱和点。区域是类似于直线、平面、还是体，通过检查像素的协方差的特征值可以很容易地确定

6.5　从图像颜色中找到表面颜色

一个颜色恒常性算法若能获取一幅图像、排除光影响并给出所看到的表面的实际颜色，这是很有吸引力的。颜色恒常性算法是比较广泛的视觉问题中一个有趣的子问题：从不确定的图像度量中决定某些参数，必须使用一个模型来处理这些度量，还应该报告从度量中得到的不确定性表示（也许包括置信区间、后验概率的协方差或者一系列代表性的解）。

6.5.1　人类对表面颜色的感知

在人类视觉系统中，存在某些颜色恒常性算法，而人们一般不会感觉到这些。没有经验的摄影者常常感到奇怪的是，荧光灯下室内拍摄的场景照片蕴含蓝色，而室外的场景照片则蕴含暖橙色。

一般对颜色恒常性与光照恒常性加以区别。颜色恒常性通常考虑那些色彩中与亮度无关的因素，如色调和饱和度；而光照恒常性是人类用于区分表面是白色、灰色还是黑色（表面的亮度）的技能，无论照明强度（亮度）如何变化。颜色恒常性不是绝对精确的，也不是不可避免的。人类可以报告以下信息：

- 白光照射下表面的颜色（一般称为表面色）
- 到达眼睛的光的颜色，该技能使得艺术家能够在有色光照射的表面上涂上油彩
- 有时能报告落于表面上的光的颜色

所有这些都可看做颜色恒常性处理过程的副产品。

6.3 节的色度学理论可以预测在给定光能分布的独立点光源下观察者将感知到的颜色。人类的颜色恒常性算法似乎可以从复杂的场景结构中获得信息，从而告诉我们，如果点光源是较大的复杂场景的一部分，那么从色度理论中获得的预测就非常不准确。Land 和 McCann (1971)的演示（见图 6.22）给出了该种效应的一个有说服力的例子。令人奇怪的是，在复杂的场景中很难预测人类会看到哪种颜色（Fairchild，1998；Helson，1938*a*，1938*b*，1934，1940），这就是难以产生非常好的颜色重现系统的诸多困难之一。

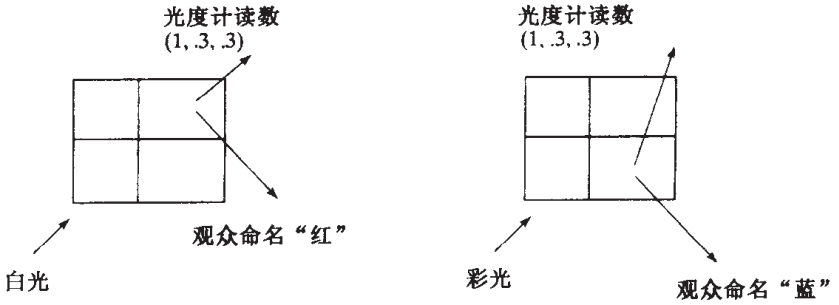


图 6.22　Land 给观众展示了由无光泽彩色矩形铺成的图形，因与著名荷兰画家蒙德里安（Mondrian）的一幅作品相像而得名，他还使用了三个幻灯片投影仪，分别投射红色、绿色和蓝色光。他使用了一个光度计来度量在三种通道下离开同一个点的光能，对应于人眼的三类感受器。他记录了度量结果，并且要求观众给纸片命名，假设结果是“红色”（左边）。Land然后调整幻灯片投影仪，使得别的某个纸片能够得到相同的光度度量，并要求观众给该纸片命名。回答给出的是纸片在白光下的颜色描述，如果纸片在白光下是蓝色，答案就是“蓝色”（右边）。在后来的实验中，Land把楔形的非彩色的过滤器放入投影仪中，使得落于纸板上的光的色彩在纸条上缓慢变化。虽然从纸条的一端到另外一端光度计度量变化得很缓慢，观众看到的纸条仍然是颜色不变的

人类在颜色恒常性上的能力令人非常难以理解。在人身上做的主要实验(Arend 和 Reeves,1986;McCann,McKee 和 Taylor,1976)并不能解释所有情况。比如,我们还不知道颜色恒常性的鲁棒性,或者高层次色调对颜色的判断有多大的作用。颜色恒常性有时候完全失败(否则就不会有电影业了),但是在什么情况下它会失败还没有完全弄清楚。对于非彩色的刺激,存在大量的关于表面光亮度感的数据。由于表面的亮度随着方向以及光照的强度而变化,因而我们可能会感觉人类的亮度恒常性是很弱的。而事实上,它在较大的光源范围内工作得非常好。

6.5.2 亮度推理

有大量的证据表明人类的亮度恒常性包括两个过程:其中一个过程对各个图像块的亮度进行比较,并使用比较的结果确定哪一个图像块更亮,哪一个更暗;第二个过程建立起这些比较可以参照的绝对标准(Gilchrist,Kossyfidis,Bonato,Agostini,Cataliotti,Li,Spehar,Annan 和 Economou,1999)。我们首先讨论亮度算法,因为它们看起来比颜色恒常性算法简单。

图像亮度的一个简单模型 落于一个像素上的辐射度取决于看到的表面的照明度、表面的双向反射分布函数(BRDF)、表面相对于光源的位置以及摄像机的响应。如果假设场景是平面且正面摆放、表面满足朗伯特性且摄像机对辐射的响应是线性的,那么情况就相当简单了。

基于上述假设获取了摄像机在点 $\boldsymbol{X}$ 响应的一个模型 C,它是照明度、反射率和摄像机增益常数的乘积:

$$C(\boldsymbol{x}) = k_c I(\boldsymbol{x})\rho(\boldsymbol{x})$$

如果我们取对数,就得到

$$\log C(\boldsymbol{x}) = \log k_c + \log I(\boldsymbol{x}) + \log \rho(\boldsymbol{x})$$

另外做如下一些假设:

- 首先,假设只有反射率在空间可以变化很快,意思是反射率的集合看起来就像把不同灰度的纸拼在了一起。该假设很容易证明:世界上反射率连续变化是较少的(最好的例子是成熟的水果),反射率的变化主要发生在一个物体遮挡另外一个物体时(因此可以预料到变化很快)。这意味着项 $\log \rho(\boldsymbol{x})$ 的空间导数项或者是零(反射率不变)或者很大(反射率发生变化)。
- 其二,照明度可以在空间发生缓慢变化。该假设比较接近现实。比如,来自点光源的照明变化相对比较缓慢,除非点光源靠得非常近,太阳是一个极好的例子。另外一个例子是,室内的光源变化非常慢,因为白色的墙壁类似于面光源。然而,该假设在阴影边界完全不成立。我们不得不把这作为一个特殊的例子,并假设或者不存在阴影边界,或者我们已经知道它们在哪里。

从模型中恢复亮度 设计使用上述模型的算法比较简单。最简单的算法,即 Land 和 McCann(1971)的 Retinex 算法,已经不再使用。一个很自然的方法是对对数变换求微分,抛弃小的梯度,并对结果求积分(Horn,1974)。丢失的仅仅是积分的常数项,因此亮度比是可以得到的,但是绝对亮度度量得不到。图 6.23 说明了在一维图像上的处理过程,其中的微分和积分计算非常简单。

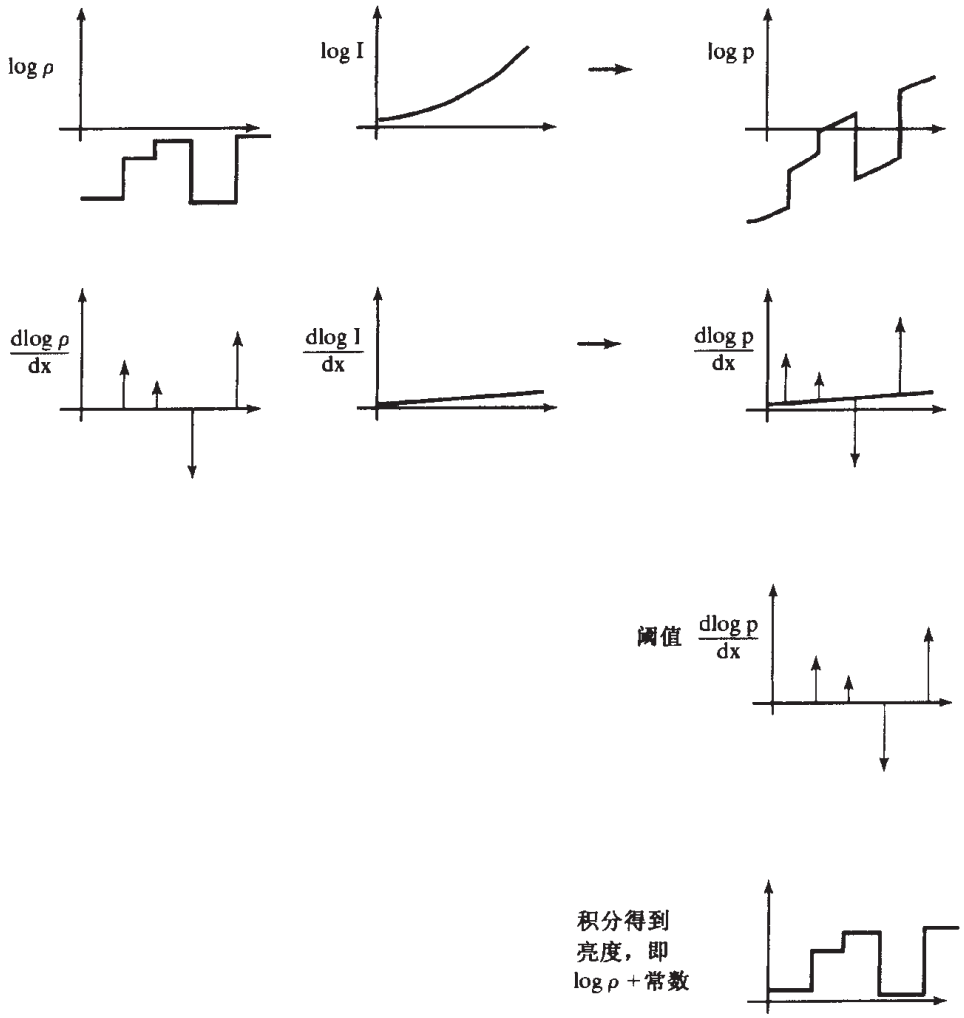


图 6.23　使用一维图像，可以很容易地说明亮度算法。在上一行，左图表示 $\log \rho(x)$，中图表示 $I(x)$，右图表示两者的和，即 $\log C$。图像亮度的对数在表面反射的变化处具有较大的导数，面在仅仅因为光照梯度引起的反射变化处梯度较小。通过对对数亮度求导，阈值处理以去除较小的导数，并积分来恢复亮度，损失的仅仅是积分常数

该方法可以扩展到二维的情况。微分和阈值计算非常简单：在每一个点计算梯度幅度的大小；如果幅度小于阈值，就把梯度向量设置为零，或者保持不变。困难在于积分这些梯度从而得到对数形式的反射率图。阈值处理后的梯度也许不再是图像的梯度，因为二阶的混合偏微分可能并不相等（可积性，与 5.4.2 节比较）。

问题可以改写为最小化问题：选择梯度最接近于阈值处理后的梯度的对数反射率图。这是一个相对简单的问题，因为计算图像的梯度是一个线性运算。阈值处理后的梯度的 x 分量放到一个向量 $\boldsymbol{p}$ 中，y 分量放到向量 $\boldsymbol{q}$ 中，把表示对数反射率的向量写成 $\boldsymbol{l}$。现在得到 x 导数过程就是线性的，因此存在某一矩阵 $\mathcal{M}_x$，满足 $\mathcal{M}_x\boldsymbol{l}$ 是 x 的导数；对于 y 的系数，对应矩阵写做 $\mathcal{M}_y$。

问题变化为寻找使下式的值为最小的向量 $\boldsymbol{l}$

$$|\mathcal{M}_x\boldsymbol{l}-\boldsymbol{p}|^2+|\mathcal{M}_y\boldsymbol{l}-\boldsymbol{q}|^2$$

这是一个二次最小化问题，通过线性化过程可以得到结果。需要用到一些小技巧，因为对 $\boldsymbol{l}$ 增

加常数向量，不会改变导数，所以该问题不存在惟一解。我们在练习中研究该最小化问题。

算法 6.1 确定图片的亮度
形成图像对数的亮度梯度
在每个像素上，如果梯度值小于某一阈值，令梯度值为零
解决本文描述的最小化问题来重新构造对数反射率
得到积分常数
把常数加入对数反射率，并取幂

积分常数必须从其他假设中得到。有两个明显的可能性：

- 假设最亮的图片是白色的；
- 假设平均亮度是常数。

我们将在练习中研究这些模型的结果。

有限维的线性模型 6.4.2 节的图像颜色模型是

$$\boldsymbol{C}(\boldsymbol{x}) = g_d(\boldsymbol{x})\boldsymbol{d}(\boldsymbol{x}) + g_s(\boldsymbol{x})\boldsymbol{s}(\boldsymbol{x}) + \boldsymbol{i}(\boldsymbol{x})$$

我们决定忽略互反射项 $\boldsymbol{i}(\boldsymbol{x})$。原则上，可以使用 6.4.3 节的方法来生成不包含镜面反射的图像，这就产生了项 $g_d(\boldsymbol{x})\boldsymbol{d}(\boldsymbol{x})$。假设 $g_d(\boldsymbol{x})$是常数，那么看到的是一个平坦的正对着我们的表面。

所得到的项 $\boldsymbol{d}(\boldsymbol{x})$，把世界建模为正面平坦的漫反射有色表面的拼接。假设有一个在整幅图像上颜色不变的光源，则该项就是光源、接收器、反射信息的融合。在现实世界中不可能完全分离它们。然而，如果给定一个分布合理的有色表面的世界和一个合理的光源，现有的算法可以从图像颜色中给出表面颜色的非常有用的估计。

项 $\boldsymbol{d}(\boldsymbol{x})$来自光源的谱辐照度，表面的谱反射率，摄像机的敏感度之间的互相作用。需要一个模型来解释这些交互作用。在 6.4.2 节中讨论过，如果具有谱漫反射率 $\rho(\lambda)$的标准漫反射表面被一个光谱为 $E(\lambda)$的光源照射，反射光的光谱就是 $\rho(\lambda)E(\lambda)$(乘以与表面方向有关的常数，已决定忽略它)。

因此，第 k 类的线性感光器收到来自该表面的光，它的响应就是：

$$p_k = \int_\Lambda \sigma_k(\lambda)\rho(\lambda)E(\lambda)d\lambda$$

其中，Λ 是相关的波长的范围，$\sigma_k(\lambda)$是第 k 类感光器的敏感度(见图 6.15)。

该响应与表面反射率和照明度都成线性关系，这就意味着对可能的表面反射和光源族使用线性模型。一个有限维线性模型把表面谱反射率和光源辐射度建模为有限数目的基函数的加权和。对于谱反射率和光源不必使用相同的基。

如果表面谱反射率的一个有限维的模型是世界的合理描述，任何表面反射率都可以写成

$$\rho(\lambda) = \sum_{j=1}^{n} r_j \phi_j(\lambda)$$

其中，$\phi_j(\lambda)$是反射率模型的基函数，而 r_j 随着表面的不同而变化。

类似的，如果光源的一个有限维的线性模型是合理的，那么任何光源都可以写成

$$E(\lambda)=\sum_{i=1}^{m} e_i \psi_i(\lambda)$$

其中，$\psi_i(\lambda)$是光源模型的基函数。

同时使用两个模型，第 k 类接收器的响应就是

$$\begin{aligned} p_k &= \int \sigma_k(\lambda)\left(\sum_{j=1}^{n} r_j \phi_j(\lambda)\right)\left(\sum_{i=1}^{m} e_i \psi_i(\lambda)\right) d\lambda \\ &= \sum_{i=1,j=1}^{m,n} e_i r_j \left(\int \sigma_k(\lambda)\phi_j(\lambda)\psi_i(\lambda)\right) d\lambda \\ &= \sum_{i=1,j=1}^{m,n} e_i r_j g_{ijk} \end{aligned}$$

其中，

$$g_{ijk}=\int \sigma_k(\lambda)\phi_j(\lambda)\psi_i(\lambda) d\lambda$$

已知，因为它们是世界模型的组成成分（它们可以从观察中得到，见练习）。

6.5.3　从有限维的线性模型中获得表面颜色

每一个索引项都可以解释为一个向量的分量，使用 $\boldsymbol{p}$ 来表示该向量，向量的第 k 个分量表示为 p_k。可以直接使用相关系数 $\boldsymbol{r}$ 的向量来表示表面颜色，或者计算出 $\boldsymbol{r}$，然后确定在白色光源下表面看起来的情况，并使用它来间接表示表面颜色。后一个表示在实际中更有用，此外，结果也更容易解释。

平均反射率规范化　假设在所有场景中，反射率的空间平均值是一个常数并已知（比如，可以假设所有的反射率的空间平均值都是全灰色）。对于反射率的有限维的基，可以把平均值写为

$$\sum_{j=1}^{n} \overline{r_j} \phi_j(\lambda)$$

既然假设平均反射率是常数，那么接收器响应的平均值一定也是一个常数（如果成像过程是线性的——见注释中的讨论），并且第 k 类接收器的响应的平均值可以写成：

$$\overline{p_k}=\sum_{i=1,j=1}^{m,n} e_i g_{ijk} \overline{r_j}$$

如果把第 k 个分量为 $\bar{p}_k$ 的向量表示为 $\bar{\boldsymbol{p}}$，并且 $\mathcal{A}$ 是一个矩阵，其中第 k,i 个元素为

$$\sum_{j=1}^{n} \bar{r}_j g_{ijk}$$

那么可以把前一个表达式重新写成：

$$\bar{\boldsymbol{p}}=\mathcal{A}\boldsymbol{e}$$

合理地选择接收器，矩阵 $\mathcal{A}$ 将会是满秩的，那么如果光源的有限维线性模型的维数与接收器的数目相同，就可以确定 $\boldsymbol{e}$，它表示了光源的参数。当然，一旦光源的参数已知，就可以报告每一个像素上的表面反射率，或者修正图像，使它看起来是在白色光下获得的图像。

算法 6.2 从已知的平均反射率获得颜色恒常性

计算图像的颜色平均值 $\bar{\boldsymbol{p}}$

从 $\bar{\boldsymbol{p}} = \mathcal{A}\boldsymbol{e}$ 中计算出 $\boldsymbol{e}$

为了获得白光下的图像 $\boldsymbol{e}^w$

对于每一个像素，从 $p_k = \sum_{i=1,j=1m,n} e_i g_{ijk} r_j$ 中计算出 $\boldsymbol{r}$

把像素值替换为 $p_k^w = \sum_{i=1,j=1m,n} e_i^w g_{ijk} r_j$

然而假设平均反射率是一个已知常数一般并不会得到满足，因为它一般不会接近实际情况。比如，如果假设平均反射率是一个取中间值的灰度值（通常的选择；可参见 Buchsbaum，1980；Gershon，Jepson 和 Tsotsos，1986），一个叶子密布的森林的林间空地的图像报告为绿色光照明下的各种灰色物体的组合。避免该情况的一个方法是对于不同类型的场景，改变平均值（Gershon，Jepson和Tsotsos，1986），但是我们如何决定使用哪一个平均值呢？另一个方法是计算出一个不是纯粹的空间平均值意义上的平均值。比如，可以对由 10 个或者更多像素表示的颜色求平均值，但是不使用每种颜色的像素的数目对它们加权。

色域（Gamut）规范化 不是任一个可能的像素值都可以通过在白光下拍摄实际表面的图像得到的。一般来说不可能获得如下这种像素值：一个通道响应很强，而另外一个通道很弱（比如，红色通道是 255，而绿色和蓝色通道是 0），这意味着整个图像的色域（所有像素值的集合）包含了光源的信息。比如，如果某个人观察到一个像素的值为(255,0,0)，那么其光源的颜色很可能就是红色。

如果一幅图像的色域中包含两个像素值，设为 $\boldsymbol{p}_1$ 和 $\boldsymbol{p}_2$，那么就必然可以在同样的光源条件下获得包含值 $t\boldsymbol{p}_1 + (1-t)\boldsymbol{p}_2$ 的图像，其中，$0 \leqslant t \leqslant 1$（因为可以在表面混合着色剂）。这意味着图像色域的凸包包含了光源信息。这些约束可以用于限制光源的颜色。

令 G 表示给定图像色域的凸包，W 为在白光下有各种不同表面的一幅图像的色域的凸包，并且 $\mathcal{M}_{\boldsymbol{e}}$ 为一个映射，它将从光源 $\boldsymbol{e}$ 下看到的一幅图像映射到在白光下看到的一幅图像。那么需要考虑的惟一光源是那些满足 $\mathcal{M}_{\boldsymbol{e}}(G) \in W$ 的光源。如果 $\mathcal{M}_{\boldsymbol{e}}$ 具有合适的结构问题就好办了；假设 $\mathcal{M}_{\boldsymbol{e}}$ 的元素是对角矩阵就是一个很自然的例子。

在有限维线性模型的情况下，$\mathcal{M}_{\boldsymbol{e}}$ 线性依赖于 $\boldsymbol{e}$，从而满足约束的光源族也是一个凸包。该族可以通过凸包集合的交来构造，其中每一个凸包都对应于一个映射关系族，它将 G 的凸包的一个顶点映射到 W 内的某个点（或者可以写下一串对 $\boldsymbol{e}$ 的线性约束）。

算法 6.3 基于色域映射的颜色恒常性算法

获得白光下大量不同颜色表面的大量图像的色域 W（这是所有像素值的凸包）

获得图像的色域 G（这是所有像素值的凸包）

获得满足 $\mathcal{M}_{\boldsymbol{e}}$ 的光源映射族 $\mathcal{M}_{\boldsymbol{e}}\, G \in W$ 的每一个元素，这表示了所有可能的光源

选择该族中的某一些元素，并用它映射图像中的任何一个像素

一旦已经形成了这个映射族，剩下的就是找到一个合适的光源，这有各种可能的策略。如果已经知道可能遇到的特定光源的似然率，那么可以选择可能性最大的。假设大部分图片包

含各种不同颜色的表面,这会导致所选择的光源使恢复的色域的范围最广。可以对光源使用其他约束(比如,所有的光源在所有波长上都应具有非负的能量),从而增加对光源集合的约束(Finlayson 和 Hordley,2000)。

6.6 注释

计算机视觉对颜色信息的使用仍然很初步,困难之一就是不知道哪些是合适的信息。John Mollon 的关于"灵长类动物的颜色系统被认为是为辨认果树而进化的"论述可以是一种解释,但是它并没有太多用处。

有很多使用颜色信息的重要资料,我们推荐 Hardin 和 Maffi(1997),Lamb 和 Bourriau(1995),Lynch 和 Livingston(2001),Minnaert(1993),Trussell,Allebach,Fairchild,Funt 和 Wong(1997),Williamson 和 Cummins(1983)。Wyszecki 和 Stiles(1982)的文章包含大量有用的信息。

三原色理论和颜色空间

直到目前为止,对为什么应用三原色理论还没给出结论性的解释,尽管通常认为是由于眼睛中存在三种不同类型的颜色感受器。Nathans 在遗传学中对感光器的研究,可以认为解释了这些假设(见 Nathans,Piantanida,Eddy,Shows 和 Hogness,1986*a*;Nathans,Thomas 和 Hogness,1986*b*),但是还远不能给出一个完备的解释,因为 Nathans 的工作也表明许多人有不止三个感光器(Mollon,1995)。

目前存在大量的颜色空间和颜色表观模型。重要的问题并非在哪个坐标系下度量颜色,而在于如何计算差异,所以对颜色度量标准的定义还有待深思。

颜色度量标准是一个古老的话题,一般用 MacAdam 椭圆来拟合尺度张量。该方法的困难在于度量张量带有很强的隐含性:能够通过积分在很大的范围内度量差异,然而很难看到大范围内比较颜色是有意义的。另外一个要关注的是,观察者从 Maxwellian 的视角观察对差异赋予的权重,与图像颜色差异的语义重要性是两个不同的问题。

镜面反射检测

这里讨论的镜面反射检测算法出自 Shafer(1985*b*),由 Klinker,Shafer 和 Kanade(1987*a*,*b*),(1990)以及 Maxwell 和 Shafer(2000)加以改进。镜面反射也可以从它们都是很亮很小的角度进行检测(Brelstaff 和 Blake,1988),或者从它们相对于背景的颜色和运动不同的角度进行检测(Lee 和 Bajcsy,1992*a*;Lee 和 Bajcsy,1992*b*)。

亮度

Land 报告了各种关于颜色的视觉实验(Land,1959*a*,1959*b*,1959*c*,1983)。最近并没有太多关于颜色恒常性算法的研究。基本思想来自 Land 和 McCann(1971)。Horn(1974)将他们的工作形式化以适应于计算机视觉的研究。Blake(1985)给出了 Horn 算法的一个变形。这就是我们描述的亮度算法。它的最初形式有些不同,称为 Retinex 算法(Land 和 McCann,1971)。Retinex 算法最初是作为颜色恒常性算法提出的,该算法非常难以分析(Brainard 和 Wandell,1986)。

亮度技术使用的并不如设想的那么广泛,尽管有证据表明它们给出了一些关于真实图像的有用信息(Brelstaff 和 Blake,1987)。仅依赖于梯度信息将光源和反射率分开是粗糙的,它忽

略了重要的信息。其中之一就是大的亮度变化一定是光源变化引起的，无论它们变化得有多快。另一个是阴影边界的颜色变化同反射率边界的颜色变化是不一样的。哪一个变化是反射率引起的，哪一个是光源引起的对于我们来说就像是一个推理问题，一个可行的推理（写出似然模型很容易，但先验概率是哪一个并不知道）。已经有些研究通过建立局部模型试图解决这个问题（Freeman，Pasztor 和 Carmichael，2000）。

颜色恒常性

光谱反射的有限维线性模型可以借助于表面物理特性来支持，因为光谱吸收线被固态效应而变粗。对表面反射的有限维线性模型的实验验证，是Cohen（1964）对一组标准参照表面（称为蒙赛尔碎片）的度量，以及Krinov（1947）对一组标准参照物的度量。Cohen（1964）对他的数据使用了主轴分解来得到一系列基函数，Maloney（1984）使用这些函数对Krinov的数据进行了函数的加权拟合，效果不错，但带有模式引起的偏差。在每一个例子中，前三个主轴解释了样本方差的高百分比（接近 99%），因而这些函数的线性组合对所有的样本函数拟合得非常好。最近，Maloney（1986）用Cohen（1964）的基函数对大数据集，包括Krinov（1947）的数据，以及更多的蒙赛尔碎片表面反射率数据，进行了拟合，并得出结论：表面反射精确模型的阶为 5 或者 6。

有限维线性模型在颜色恒常性研究中是一个非常重要的工具，从该方法中自然衍生出大量的算法。有一些算法利用了线性空间的特性（Maloney，1984；Maloney 和 Wandell，1986；Wandell，1987）。在类似塑料这样的表面，反射光的镜面成分与光源的颜色相同。如果可以从图像中的物体中确定镜面区域，光源的颜色就知道了。该思想最近非常流行——Judd（1940）在 1960 年记录了德国人在表面颜色感知中的早期工作，并称它为“更有用观点”。该观点最近也非常流行（D'Zmura和Lennie，1986；Flock，1984；Klinker，Shafer和Kanade，1987*a*，*b*；Lee，1986）。假设颜色均值是常数，是另外一种常用的方法（Buchsbaum，1980；Gershon，1987）。

色域映射方法来自Forsyth（1990），该方法已大大增强了（Barnard，2000；Finlayson和Hordley，1999，2000）。与光照变化相关的映射族的做法，$\mathcal{M}_e$ 已经得到了深入的研究。该工作最初是由Von Kries所做（他对该问题的想法与我们不同）。他假设颜色恒常性本质上是在每一个通道上独立的亮度计算的结果，这意味着可以通过独立地缩放每一个通道来校正一幅图像。该做法一般称为Von Kries法则。在我们的表示中，该法则等于假设$\mathcal{M}_e$ 为对角矩阵。Von Kries法则被证明是一个非常好的法则（Finlayson，Drew和Funt，1994*a*）。目前最好的做法是首先对通道使用线性变换，然后使用对角映射来缩放所得到了结果（Finlayson等，1994*a*，1994*b*）。

在颜色恒常性方面，很少有工作把光源的空间变化和表面颜色的解结合起来，这就是在我们的模型中忽略了一些项的原因。理想情况下，应在阴影和表面方向上进行研究。另外，在我们看起来整个问题就像一个推理问题，却很难解决。关于这个极其重要问题的主要文章有Barnard，Finlayson和Funt（1997），Funt和Drew（1988）。其中还存在很大的研究空间。

有色表面间的相互反射导致了一种称为染色的现象，其中每一个表面都反射有色光到其他表面上。这种现象在实际中非常多。人类似乎能很好地完全忽略它们，大部分人并没有意识到这种现象的存在。忽略染色可能使用了空间的信息。举出一些好的例子需要动一番脑筋。作者在南加州偶然碰到一个很好的例子，那儿的路边有许多很大的开着白色夹竹桃的树篱。白色夹竹桃的叶子是深色的，而花是白的。在明亮的阳光下，时而可以看到开着黄色夹竹桃花的树篱；乍一想会以为这是由于停泊在路上的黄色服务卡车的颜色反射黄光到白花上所致。人忽略染色的能力被深色

的叶子所破坏,这些叶子破坏了空间的模式。染色包含了表面颜色的信息,但是很难被分离出来(见 Drew 和 Funt,1990;Funt 和 Drew,1993;Funt,Drew 和 Ho,1991)。

物体识别中颜色的应用

在以后的章节中可以看到,建立利用物体的颜色来帮助识别的系统是技巧性很强的工作。颜色恒常性被认为是一种可以改进匹配的过程,因为可以利用物体的属性,而不是特定观测下的属性(Finlayson,Chatterjee和Funt,1996;Funt,Barnard和Martin,1998;Funt和Finlayson,1995)。如果我们愿意接纳进化论的一个观点,那么均匀的颜色空间是有用的,这个观点说:理解人们能觉察到的颜色差异是有价值的,因为它们与有用的度量是相适应的。

习题

6.1 准备一包彩色纸,和一个朋友一起来比较使用的颜色名称。最好准备一大包纸——可选择那些用于绘画,或者Pantone色彩系统的样品纸,一般比较便宜。尝试的名称最好是基本的颜色名称——红色,粉红色,橙色,黄色,绿色,蓝色,紫红色,棕色,白色,灰色和黑色,这些术语(以及其他一些术语)都是非常经典的,在各种语言中得到了广泛的使用(Hardin和Maffi,1997 的论文给出该论题当前思考的一个很好的总结)。在命名过程中很容易产生一些分歧,例如哪些称为蓝色,哪些称为绿色。

6.2 导出 RGB 与 CIE XYZ 之间互相转换的公式,该变换为线性变换。写出线性变换的元素表达式即可,不必查找颜色匹配函数的实际数值。

6.3 通过选择原色,为这些原色构造颜色匹配函数来得到颜色空间。证明存在一个线性变换,该变换将一个线性颜色空间中的值变换到另外一个颜色空间中;最简单的方法是按照颜色匹配函数写出变换方程。

6.4 练习 3 告诉我们,在建立线性颜色空间时,可以任意地选择原色,但是对于颜色匹配函数的选择是有约束的。为什么?这些约束又是什么?

6.5 在一种光照下具有相同颜色、而在另外一种光照下具有不同颜色的两个表面,一般称为条件同色(metamers)。最优色是一种在某些波长值为零、而在另外一些波长值为 1 的光谱反射,或者辐射。虽然最优色在实际中并不存在,但在解释各种效果时它们还是很有用的(参见 Ostwald 的文章)。

(a) 使用最优色解释条件同色现象。

(b) 给出一个特定的光谱反射率,证明它存在无穷的条件同色的光谱反射率。

(c) 使用最优色来构造这样一个例子:一些表面在一种光照(比如,红色和绿色)下具有不同颜色,而在另外一种光照下具有相同颜色。

(d) 使用最优色来构造这样一个例子:一些表面在光照改变时交换外观颜色(比如,在光照一下,表面一看起来是红色,表面二看起来是绿色;而在光源二下,表面一看起来是绿色,而表面二是红色)。

6.6 需要将打印机的色域映射到显示器的色域上,其中每个色域中都存在另外一个色域没有的颜色。假设给定一个无法重新精确生成的显示器颜色,你可以选择最接近的打印机颜色。解释一下,为什么对于重新生成图像,这是一个不好的做法?在生成商业图片中它

是否能够工作(条图、饼图以及其他类似的,包含有大块相同颜色块的图)?

6.7 体(积)色是与有色透明物质相关的一种现象——最吸引人的例子是一杯酒。色彩来自不同波长的不同吸收系数。解释:(a)为什么一小杯颜色足够深红的酒看起来是黑色的;(b)为什么一大杯浅色的红酒看起来也是黑色的。可以做实验。

6.8 (本练习需要一些数值分析的知识)在 6.5.2 节中,我们将从表面集合中恢复对数反射率的问题描述为对下式最小化

$$|\mathcal{M}_x \boldsymbol{l} - \boldsymbol{p}|^2 + |\mathcal{M}_y \boldsymbol{l} - \boldsymbol{q}|^2$$

其中,$\mathcal{M}_x$ 表示 $\boldsymbol{l}$ 相对于 x 的导数,$\mathcal{M}_y$ 表示相对于 y 的导数。

(a) 我们确信$\mathcal{M}_x$ 和$\mathcal{M}_y$ 存在。请使用前向差分的表达式(或者中分,或者其他关于导数的一些差分近似)来形成矩阵。矩阵的所有元素几乎都是零。

(b) 最小化问题可以被重新写成如下形式

$$\text{choose } \boldsymbol{l} \text{ to minimize } (\mathcal{A}\boldsymbol{l} + \boldsymbol{b})^T(\mathcal{A}\boldsymbol{l} + b)$$

确定$\mathcal{A}$和 $\boldsymbol{b}$ 的值,并写出如何解决这个一般性的问题。注意$\mathcal{A}$不是满秩的,所以不能对它求逆矩阵。

6.9 在 6.5.2 节中曾经提到产生积分常数的两个假设。

(a) 写出如何使用这些假设来恢复反射率图。

(b) 对于每一个假设,描述一种失败的情形,以及失败的本质。所举的例子必须在有许多反射率可见的情况下可行。

6.10 阅读图书"*Colour: Art and Science*",Lamb 和 Bourriau 著,剑桥大学出版社 1995 年出版。

编程作业

6.11 关于光源和表面光谱的一些参考内容可以在 Web 上找到(http://www.it.lut.fi/research/color/lutcs_database.html)。使用主分量分析,从一些光源和表面反射中拟合出一个有限维线性模型,绘制结果模型,将所绘制的结果与精确的结果进行比较。判断在哪一处造成了最明显的错误?为什么?

6.12 使用不同的纸张在喷墨打印机上打印一幅彩色图像,并比较结果。特别关注下列情况:(a)驱动程序知道打印机在哪种纸上打印,并比较颜色的变化(哪一个是无法感知的)。(b)对打印机将在哪种纸上打印给出错误的信息(比如,在普通纸上打印却告诉打印机在相片纸上打印)。你能够解释看到的变化吗?为什么照相纸有光泽?

6.13 用有限维线性模型对光源和反射物进行拟合,这种方法不太好,因为没有办法保证交互作用可以很好地表示(无法在拟合误差中解释得到)。通过不使用基函数的拟合过程可以得到 g_{ijk}。实现该过程[细节描述见 Marimont 和 Wandell(1992)],并把结果与以前练习中得到的结果进行比较。

6.14 假设反射率的空间均值是常数,建立一个颜色恒常性算法。使用有限维线性模型。可以从练习 3 中得到 g_{ijk} 值。

6.15 在表面的彩色模型中忽略互反射。做一个实验,思考一下从颜色互反射中可能得到的颜色漂移的大小(非常大)。人类很少把颜色互反射解释为表面颜色。想一想为什么会这样?参考对亮度算法的讨论。

6.16 基于 6.4.3 节所描述的线,设计一个镜面反射检测算法。

第二部分　低层视觉:使用一幅图像

- 第 7 章　线性滤波
- 第 8 章　边缘检测
- 第 9 章　纹理

第7章 线性滤波

在一张斑马和黑白斑点狗的图片上，间隔分布着几乎相同数量的黑白像素。两者之间的区别不在于逐个像素的值，而在于小组像素的特征表现。这一章将介绍一些如何获取小组像素分布特征描述的方法。

这里主要的策略是用不同的加权模式计算像素加权和，以寻找不同的图像模式。这个方法尽管非常简单，却非常有用，它能够平滑图像中的噪声、检测边缘以及其他的图像模式。

7.1 线性滤波和卷积

许多重要的效果都能够用简单的模型进行建模。构造一个和图像同样大小的新数组，用一个加权模式来计算图像中各个位置的数值的加权和，在新数组的相应位置填上该加权值之和。不同的权重模式代表不同的处理方法。计算确定区域的局部平均就是一个例子。对于输入图像 F 的每个像素(i,j)，计算该像素为中心的 $2k+1\times 2k+1$ 邻域内每个点的平均值$\mathcal{F}$，得到输出

$$\mathcal{R}_{ij}=\frac{1}{(2k+1)^2}\sum_{u=i-k}^{u=i+k}\sum_{v=j-k}^{v=j+k}\mathcal{F}_{uv}$$

这个例子的加权模式很简单(每个像素根据同样的常数加权)，但是也可以使用其他一些更有意义的权重。例如，设置中心点的权重值很大，随着点远离中心，权重值迅速减小，这种模式用于模拟散焦镜头系统的平滑效果。

无论选取何种权重，这种方法的输出是移位不变的——这意味着输出值依赖于图像相邻区域的特征，而不是相邻区域的位置；同时输出是线性的——意味着两幅图像和的输出等于两幅图像输出的和。这种方法就是线性滤波。

7.1.1 卷积

这里先介绍一些符号表示。线性滤波器使用的加权模式通常称为滤波的核，使用滤波的过程称为卷积。有一点需要注意：后面将会解释(7.2.1 节)，卷积使得过程变得不那么显而易见。具体说来，给定一个滤波核$\mathcal{H}$，图像$\mathcal{F}$的卷积结果是一个图像$\mathcal{R}$。$\mathcal{R}$的(i,j)位置的元素值表示为

$$R_{ij}=\sum_{u,v}H_{i-u,j-v}F_{u,v}$$

这个过程定义了卷积——我们称使用$\mathcal{H}$将$\mathcal{F}$卷积到域$\mathcal{R}$。仔细观察表达式，与相关计算相比，替换变量 u(或 v)的“方向”是反的。这一点很重要，如果忘记了反向，将得出错误的结果。反向的原因将在 7.2.1 节中加以解释。我们故意避免标出求和的范围。实际上我们假设求和中的 u,v 范围足够大，以保证将所有的非零元素都包括在内，而且，假设没有提及的元素的数值

都是零;这意味着可以把核表示成大量零元素中分布少量非零元素的模型。下面将使用这个惯例。

例7.1 通过平均进行平滑

一般地,图像具有相邻像素的数值相近的性质,而噪声的影响能够合理地假设为并没有改变上述性质。举个例子:图像中偶尔有一些坏像素,或者少量均值为零的随机值叠加在像素点的数值中。很自然地,可以通过将每一个点用周围点的加权平均替代的方法,以减少噪声的影响,通常称这个过程为平滑或者模糊。

用一个点为中心的某个邻域内像素值的非加权平均值替代这个点的数值,同使用元素为常数的核进行卷积相同。需要注意邻域的范围。这个过程是一个较差的模糊模型——它的输出看起来并不像一个散焦的照相机(见图7.1)。这个原因非常明显,假设有一个图像,除了中心点数值为1外其余每个点都为零。如果通过对这幅图像的每个点进行无权重平均的方法进行光滑,结果看起来就像是一个盒子,这并不是一个散焦照相机的结果。希望采用一个小的光点对图形对称的模糊区进行模糊,光点中心的亮度要比边缘处的亮,亮度逐渐减弱。正如图7.1所示,这种形式的加权模式,能够获得令人信服的散焦模型。

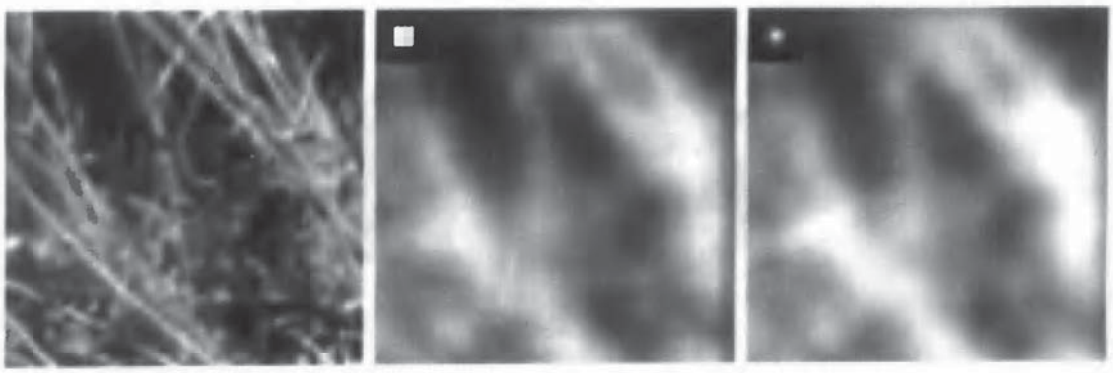


图7.1 尽管均匀局部平均看起来是一个不错的光滑模型,它产生的通常并不是散焦镜头的效果。上面的图像对比了均匀局部平均同加权平均的模糊效果。左边的图像是一个草丛。中间是均匀局部模型的模糊效果,右边是高斯加权模型的模糊效果。两种情况下光滑的程度看起来相同,但是均匀局部平均产生一些水平和垂直的细线——这种现象通常称为振铃效应。左上角表明了光滑图像使用的权重模式,它们自身呈现为一幅图像;亮点代表大数值,暗点代表小数值(这个例子中,最小的值是零)

例7.2 高斯平滑

处理这种模糊问题的较好模型是如图7.2所示的对称性高斯模型:

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)$$

σ是高斯分布的标准差(或是"sigma"),单位是像素间距,通常称为像素;常数项使得在整个平面的积分值为1,但在平滑应用中经常被忽略。这种平滑的名字来自于核的形状与概率密度分布相同方差的二维正态(或高斯)随机变量相同。

这种平滑核实现加权平均:中心点的权值要比边缘点的权值强得多。这种方法的合理性可以定性地分析:平滑抑制噪声要遵循像素值与相邻点相近的要求。对较远相邻点取较小的权重,可以确保点看起来更接近比较近的相邻点。下面是一个定性分析:

- 如果高斯分布的标准差很小——甚至小于一个像素——平滑效果将会很差，因为偏离中心的所有像素的权重都非常小。
- 如果一个大些的标准差，相邻的像素在加权平均过程中将有大一些的权重，意味着平均的结果将偏向多数相邻点的意见——这样能够得到一个像素值的较好估计，噪声随着光滑也将大大降低，但代价是图像会有些模糊。
- 最后，一个具有很大标准差的核将导致图像细节随同噪声一同消失。

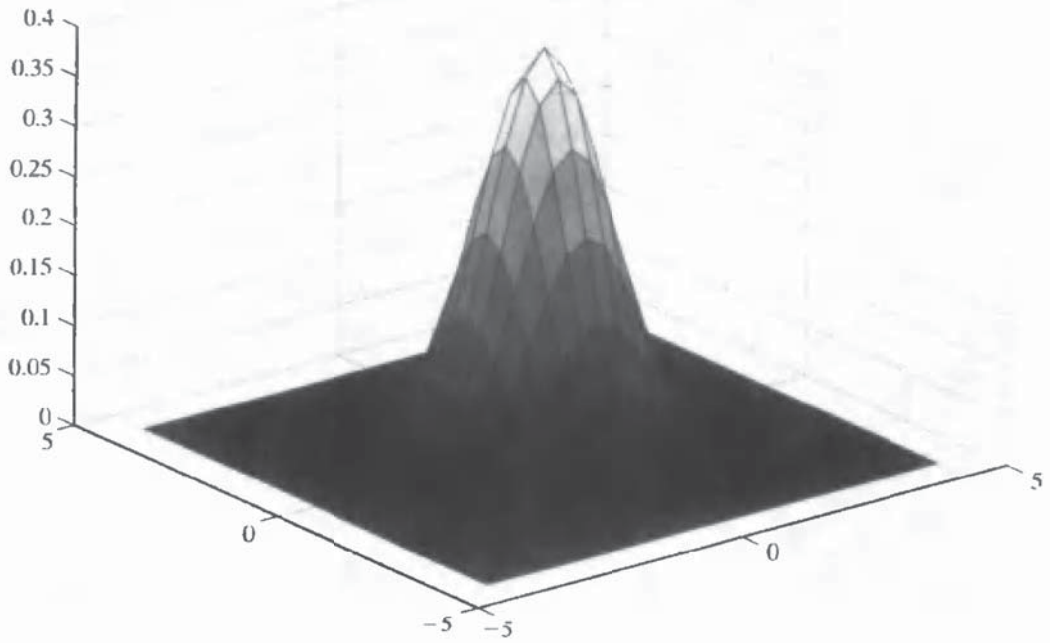


图 7.2 二维对称高斯核。图像中核的尺度按总和为 1 设置，这个放缩比例通常被忽略；核的 $\sigma = 1$。这个核的卷积形成了一个加权平均，强调位于卷积中心的点的作用，再加上边缘点的微弱贡献。注意高斯分布同关于图像模糊的点扩散函数在性质上的相似性：它是圆形对称的，中心的作用明显，边界作用逐渐衰减

图 7.3 说明了这种现象。读者可以注意到高斯平滑在抑制噪声方面是很有效的。

在应用中，一个离散的平滑核可以通过建立一个 $2k+1 \times 2k+1$ 的矩阵得到，第 (i,j) 个元素值为

$$H_{ij} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{((i-k-1)^2+(j-k-1)^2)}{2\sigma^2}\right)$$

σ 的选取值得注意，如果 σ 的值太小，矩阵中只有一个非零元素。如果 σ 太大，k 也必须大，否则将忽视周边点的贡献，如果选用合适权重，这些周边点本应对结果有所贡献。

例 7.3 导数和有限差分

图像导数可以使用另一个卷积过程进行近似。因为

$$\frac{\partial f}{\partial x} = \lim_{\epsilon \to 0}\frac{f(x+\epsilon, y) - f(x, y)}{\epsilon}$$

可以用有限差分作为一个偏导数的估计：

$$\frac{\partial h}{\partial x} \approx h_{i+1,j} - h_{i-1,j}$$

这与卷积的效果相同，卷积核为

$$\mathcal{H} = \begin{Bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{Bmatrix}$$

注意到这个核可以解释成一种模板：它对一边是正的、另一边是负的图像结构产生大的正响应，对其镜像图像产生一个大的负响应。

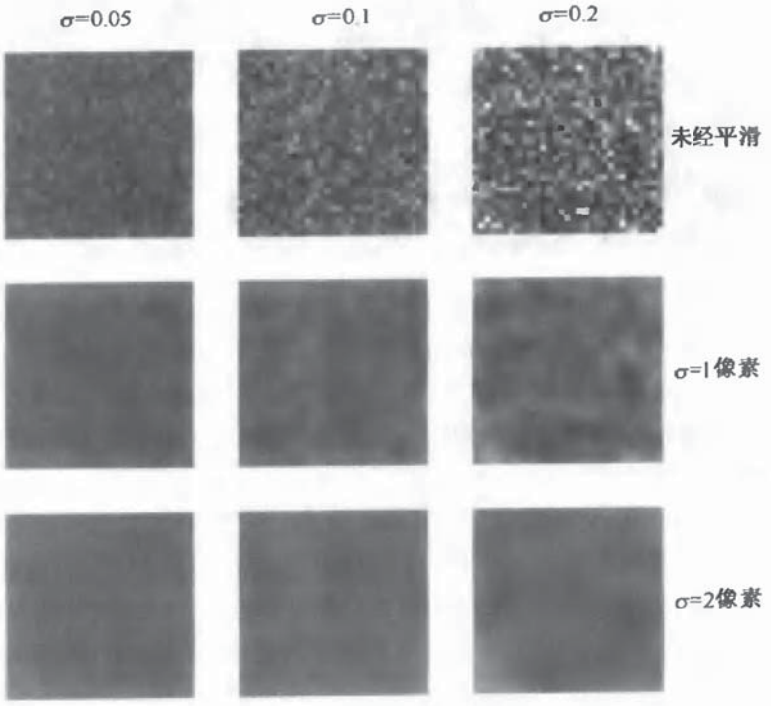


图7.3 上面一行表示一个常数值灰度图加入附加高斯噪声后的情况。在这种噪声模型中，每个像素附加一个均值为零的随机变量。像素值的范围从0到1，第1列噪声的标准差为全量程的1/20。中间一行表示使用$\sigma = 1$的高斯滤波对上面一行相应图像进行处理后的结果。注意这里符号含义的区别；这里有高斯噪声和高斯滤波，各自有相应的σ值。因为高斯滤波对高斯噪声有特别好的抑制作用，所以尽管并非始终有效，但是通过上下文可以进行区分。这是因为每个点的噪声作用是独立的，意味着它们的期望平均值将是噪声的均值。下面的一行表示使用$\sigma = 2$的高斯滤波对上面一行相应图像进行处理后的结果

正如图7.4所示，用有限差分对导数进行估计最不令人满意。这是因为有限差分对快速变化响应强烈(即，大量级的输出)，而快速变化是噪声的特征。举个例子，如果我们买一个折价的照相机，它的某些点可能会呈现或黑或白。在这些点，有限差分的输出将非常大，因为一般来说，它们同周围的点很不相同。这一切表明在有限差分前使用某些平滑方法是合适的，其细节将在8.1节和8.2节中介绍。

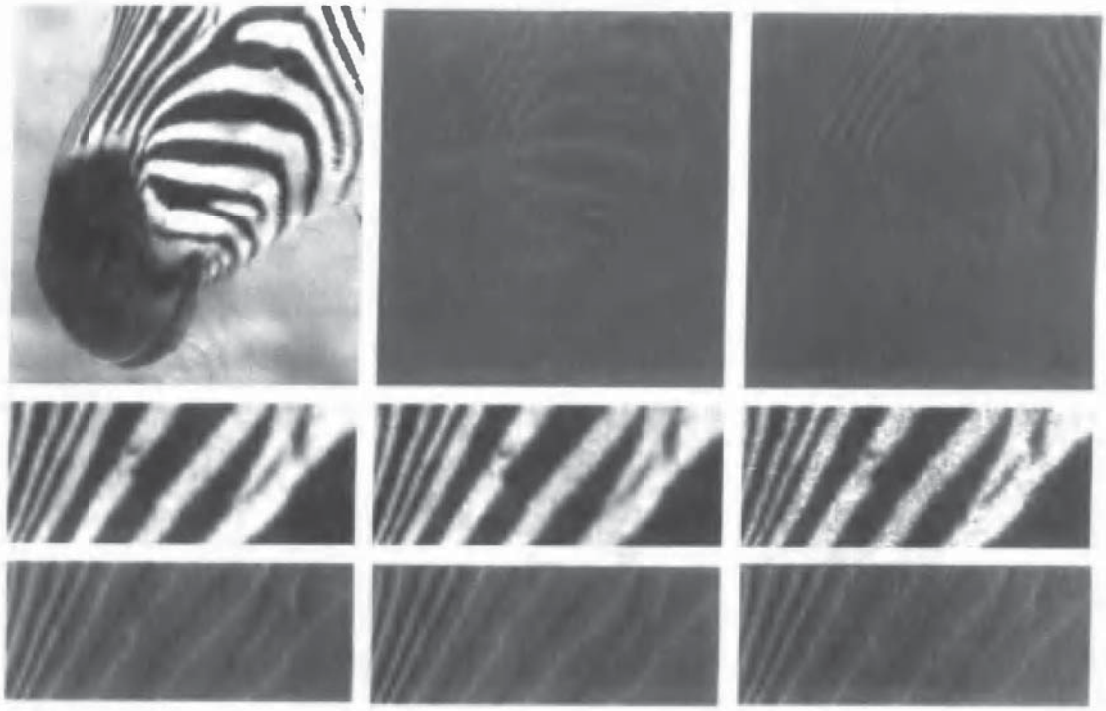


图 7.4 上面一行表示由有限差分得到的导数估计。左边的图像是一个斑马图像上的局部细节。中间的图像是y方向上的偏导数——对水平条纹反应强烈，而对竖直条纹反应微弱。左边的图像是x方向上的偏导数——对竖直条纹反应强烈，而对水平条纹反应微弱。有限差分对噪声的影响非常敏感。中间一行左边的图像是一个斑马图像上的局部细节；这一行中间的图像是原图的每个像素加上一个均值为零的随机噪声($\sigma=0.03$，像素最大值为1，最小值为0)；这一行右边的图像是原图的每个像素加上一个均值为零的随机噪声($\sigma=0.09$)。下面一行是中间一行每幅图像相应x方向上的偏导数。可以注意到，噪声对差分结果影响之大——差分图像中颗粒效应明显增加。在差分图像中，中等灰度为零值，低灰度区域为负值，高灰度区域为正值

7.2 移不变线性系统

卷积表示了一大类系统的性质。特别的，大多数图像系统可以高度近似成具有以下三个重要性质：

- **叠加性**：表示为

$$R(f+g)=R(f)+R(g)$$

  这意味着，对一组混合激励的响应等于对单个激励响应的和。

- **按比例缩放**：对零输入的响应是零。与叠加性类似，对一个按比例缩放输入的响应等于对输入响应的按比例缩放，也就是说：

$$R(kf)=kR(f)$$

  同时具有叠加性和按比例缩放性质的系统，称为线性系统。

- **位移不变性**：在位移不变系统中，对转移了位置的激励的响应等于对原激励的响应产生相同的位移。举个例子，如果投向照相机中心的光为一个小亮点，那么如果这束光投向外围，我们看到的仍然是同样的小亮点，只是转移了位置。

一个同时满足线性和移不变性的系统称为移不变线性系统，或者简称为系统。

移不变线性系统对激励的响应可以由卷积得到。我们首先在离散输入情况下证明这一点：输入为向量或者矩阵，产生离散输出。我们使用这一点，描述直线或平面上连续系统的情况，通过这些分析，能得到一些关于卷积的有用结论。

7.2.1　离散卷积

在一维情况下，移不变线性系统的输入和输出各为一个向量。这种情况最容易解决，因为需要处理的索引数较少。二维情况下，输入和输出各为一个矩阵。对于每一种情况，假设输入和输出是无限维空间，这样可以忽略一些由输入边界引起的次要因素，这一点将在 7.2.3 节中讨论。

一维离散卷积　设输入向量为 $\boldsymbol{f}$。为了方便，假设向量的维数无限，它的元素索引为整数（也就是说，有一个元素的指数为 -1 的情况）。向量的第 i 个分量是 f_i。$\boldsymbol{f}$ 可以表示成某组基的元素的加权和。一个方便的基由这样一组元素组成，每个元素中除了一个分量为 1，其他分量为 0。定义：

$$\boldsymbol{e}_0 = \cdots 0, 0, 0, 1, 0, 0, 0, \cdots$$

这个数据向量在第零个位置有一个 1，其他位置为 0。定义一种移位操作，以便把一个向量演变成该向量的移位表示。具体说来，向量 $\mathtt{Shift}(\boldsymbol{f}, i)$ 的第 j 个元素，是 f 的第 $j-i$ 个分量 $\boldsymbol{f}$。例如，移位 $\mathtt{Shift}(\boldsymbol{e}_0, 1)$ 的第一个分量为 0。现在可以写为

$$\boldsymbol{f} = \sum_i f_i \,\mathtt{Shift}(\boldsymbol{e}_0, i)$$

把系统对向量 $\boldsymbol{f}$ 的响应记做

$$R(\boldsymbol{f})$$

因为系统是移不变的，有

$$R(\mathtt{Shift}(\boldsymbol{f}, k)) = \mathtt{Shift}(R(\boldsymbol{f}), k)$$

并且，由于线性，则有

$$R(k\boldsymbol{f}) = kR(\boldsymbol{f})$$

这意味着

$$\begin{aligned} R(\boldsymbol{f}) &= R\left(\sum_i f_i \,\mathtt{Shift}(\boldsymbol{e}_0, i)\right) \\ &= \sum_i R(f_i \,\mathtt{Shift}(\boldsymbol{e}_0, i)) \\ &= \sum_i f_i R(\mathtt{Shift}(\boldsymbol{e}_0, i)) \\ &= \sum_i f_i \,\mathtt{Shift}(R(\boldsymbol{e}_0), i)) \end{aligned}$$

这又意味着，只需要知道系统对 $\boldsymbol{e}_0$ 的响应，就能够得到系统对任何数据向量的响应，通常称之为系统的脉冲响应。假设脉冲响应记为 $\boldsymbol{g}$，则有

$$R(\boldsymbol{f}) = \sum_i f_i \,\mathtt{Shift}(\boldsymbol{g}, i) = \boldsymbol{g} * \boldsymbol{f}$$

其中定义了一个操作:一维离散卷积,记为 * 。

如果考虑 $R(\boldsymbol{f})$的第 j 个元素,记做 R_i,我们有

$$R_j = \sum_i g_{j-i} f_i$$

这里沿用了(同时也解释了)7.1.1 节中使用的形式。

二维离散卷积 设矩阵$\mathcal{D}$中第(i,j)个元素记为 D_{ij}。这里对脉冲响应的适当类推是对如下激励的响应

$$\mathcal{E}_{00} = \begin{matrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{matrix}$$

如果$\mathcal{G}$是系统对这个激励的响应,与一维卷积系统一样,考虑对激励$\mathcal{F}$的响应,就是

$$R_{ij} = \sum_{u,v} G_{i-u,j-v} F_{uv}$$

记做

$$\mathcal{R} = \mathcal{G} ** \mathcal{H}$$

7.2.2 连续卷积

移不变线性系统对连续输入产生连续输出。举个例子,照相机镜头取连续一串曝光,拍出一串画面,许多镜头都是近似移不变的。对这些系统的简单研究让我们能够分析,用一个离散函数(每个像素值)近似一个连续系统(穿过图像平面上拍下的连续辐射度)产生的信息丢失。

对系统响应的描述是借助于系统对一个特殊的函数 δ 的响应实现的。我们先在一维情况下推导,以使叙述容易些。

一维卷积 通过一个离散系统的表达式,可以获得连续移不变线性系统响应的表达式。使用一个在一段时间内保持不变的离散输入,就可以得到一个连续的输入函数。缩小时间窗口,考虑趋向极限时的情况。

设系统输入是一个一维函数,返回值也是一个一维函数。我们再次把对输入 $f(x)$的响应记为 $R(f)$;当需要强调 f 是一个函数时,记为 $R(f(x))$,响应仍然是一个函数;偶尔需要强调这一点时,记做 $R(f)(u)$。表达式中线性关系记为

$$R(kf) = kR(f)$$

(k 为常数),并且通过引入一种移位操作符号表示移不变性

$$\mathtt{Shift}(f, c) = f(u - c)$$

通过移位操作符号 Shift,把移不变性表示为

$$R(\mathtt{Shift}(f, c)) = \mathtt{Shift}(R(f), c)$$

定义窗口(box)函数为:

$$box_\epsilon(x) = \begin{cases} 0 & abs(x) > \frac{\epsilon}{2} \\ 1 & abs(x) < \frac{\epsilon}{2} \end{cases}$$

$box_\epsilon(\epsilon/2)$的值对我们的目的无关紧要。输入函数为$f(x)$。构造一个均匀分布的点的网格x_i，其中$x_{i+1} - x_i = \epsilon$。我们构造一个向量$\boldsymbol{f}$，$\boldsymbol{f}$的第i个元素(记做f_i)为$f(x_i)$。这个向量能够用来表示函数。

通过$\sum_i f_i$ Shift (box_ϵ, x_i)获得一个函数f的近似表达。把这个输入到一个移不变线性系统中，响应是对窗口函数移位响应的加权和。这意味着

$$\begin{aligned} R(\sum_i f_i \,\mathtt{Shift}(box_\epsilon, x_i)) &= \sum_i R(f_i \,\mathtt{Shift}(box_\epsilon, x_i)) \\ &= \sum_i f_i R(\mathtt{Shift}(box_\epsilon, x_i)) \\ &= \sum_i f_i \,\mathtt{Shift}\left(R\left(\frac{box_\epsilon}{\epsilon}\epsilon\right), x_i\right) \\ &= \sum_i f_i \,\mathtt{Shift}\left(R\left(\frac{box_\epsilon}{\epsilon}\right), x_i\right)\epsilon \end{aligned}$$

到目前为止所做的一切都是对离散函数的推导。如果$\epsilon \to 0$，则可以获得一个近似积分。

这里介绍一个新的工具，叫做δ函数，用来表示box_ϵ/ϵ这一项。定义

$$d_\epsilon(x) = \frac{box_\epsilon(x)}{\epsilon}$$

而δ函数就是

$$\delta(x) = \lim_{\epsilon \to 0} d_\epsilon(x)$$

由于我们并不打算计算这个极限值，所以不用讨论$\delta(0)$的值。这个函数一个有趣的特性，因为，实际的移不变线性系统对δ函数的响应存在并具有紧支持(也就是说，除了在有限数量的有限长度间隔外，其他均为0值)。例如，一个非常小、非常亮的光是一个很好的二维δ函数的模型。如果使光更小更亮，同时保持整体能量为一个常数，通过散焦镜头，我们能够看到一个很小但是有限的光点。δ函数同连续情况下的$\boldsymbol{e}_0$具有相似性。

这意味着系统响应的表达式

$$\sum_i f_i \,\mathtt{Shift}\left(R\left(\frac{box_\epsilon}{\epsilon}\right), x_i\right)\epsilon$$

当ϵ趋向于0时转变为积分。可以得到

$$\begin{aligned} R(f) &= \int \left\{R(\delta)(u - x')\right\} f(x')\,dx' \\ &= \int g(u - x') f(x')\,dx' \end{aligned}$$

这里把$R(\delta)$——通常称为系统的脉冲响应——记做g，并且忽略积分的上下限。这些积分能够从$-\infty$到∞，但如果g和h有紧支持，则能够使用更紧凑的上下限。这个操作也被称为卷积，并进一步表达为

$$R(f) = (g * f)$$

卷积是对称的，意思是

$$(g * h)(x) = (h * g)(x)$$

卷积是可结合的，意思是

$$(f * (g * h)) = ((f * g) * h)$$

后面的性质意味着，可以找到一个单一的移不变线性系统，它等同于两个不同系统的组合。当讨论采样时这一点是很有用的。

二维卷积 二维的卷积推导需要更多的符号。定义窗口函数 $box_{\epsilon^2}(x, y) = box_{\epsilon}(x) box_{\epsilon}(y)$，则有

$$d_{\epsilon}(x, y) = \frac{box_{\epsilon^2}(x, y)}{\epsilon^2}$$

当ϵ趋于 0 时，δ 函数是 $d_{\epsilon}(x, y)$的极限。最终，在累加和中有更多的项。于是可以得到下述表达式

$$\begin{aligned} R(h)(x, y) &= \int\int g(x - x', y - y') h(x', y') \, dx \, dy \\ &= (g ** h)(x, y) \end{aligned}$$

这里使用两个 * 表示二维卷积。二维卷积是对称的，意思是

$$(g ** h) = (h ** g)$$

卷积是可结合的，意思是

$$((f ** g) ** h) = (f ** (g ** h))$$

二维系统脉冲响应的一个模型，可想像成照相机中看远处很小的点光源时见到的模式(覆盖一个很小的视角)。在实际镜头中会看到有点模糊的亮块，这解释了点扩展函数称呼的由来。它通常在二维系统的脉冲响应中使用。线性系统的点扩展函数通常称为它的核。

7.2.3 离散卷积的边缘效应

在实际系统中，无法获取数据的无限数组。这意味着当计算卷积时，需要考虑图像的边缘；在边缘处计算有些像素位置的卷积值时，需要虚拟并不存在的图像值。这可以采用很多策略：

- **忽略这些点**——这意味着只考虑那些只需图像位置真实存在的点计算卷积的像素。这种方法的优点是直接，但其缺点是输出比输入要小。重复的卷积可能造成图像严重收缩。
- **使用常数填充图像**——这意味着，当输出值接近图像边缘时，卷积输出对图像的依赖程度下降。这是一个卷积技巧，因为它能够保证图像不收缩，但是缺点在于会在边缘产生梯度。
- **使用其他方法填充图像**——例如，假设图像是一个双重周期函数，所以如果图像大小为 $n \times m$，那么 $m + 1$ 列——由卷积所需——将会与 $m - 1$ 列相同。这样在边缘附近能够出现较大的二阶导数值。

7.3　空间频率和傅里叶变换

以上使用的技巧，是把一个信号 $g(x,y)$ 看成许多(或者无数)小窗口函数的加权和。这个模型强调了信号是一个向量空间的一个元素——窗口函数形成卷积的基，权重是这个基的系数。需要一个新的技术来处理到目前为止没有涉及的两个相关问题：

- 一个问题是：尽管很明显，一个离散图像版本不能代表信号的全部信息，但是尚未说明失去了哪些元素；
- 另一个问题是：很明显，不能采用每隔 k 个像素取一个的简单方法压缩图像——这样会把国际象棋盘变成全白或者全黑——需要知道怎样可以安全地压缩图像。

所有这些问题同一幅图像内存的快速变化有关。例如，压缩一幅图像很可能丢失快速的变化，因为它们在样本中滑动；与此类似，快速变化时的导数非常大。

这个效果可以通过基的变化来研究。将基变为一系列正弦函数，把信号表示为无限个正弦函数的无限加权和。这意味着信号的快速变化是明显的，因为它们在新的基中对应于高频正弦项有高幅值。

7.3.1　傅里叶变换

通过傅里叶变换进行基的转化。信号 $g(x,y)$ 的傅里叶变换定义为

$$\mathcal{F}(g(x,y))(u,v)=\iint_{-\infty}^{\infty} g(x,y)e^{-i2\pi(ux+vy)}\,dx\,dy$$

并且假设存在合适的技术条件使得积分存在。g 的所有值都是有限的是积分存在的充分条件，还有大量其他可能的情况(Bracewell，1995)。这个变换的输入是一个 x,y 的复函数，返回的是 u,v 的复函数(图像是具有 0 虚部分量的复函数)。

我们暂且固定 u 和 v 的值以考虑在这个点变换的含义。指数能够改写为

$$e^{-i2\pi(ux+vy)}=\cos(2\pi(ux+vy))+i\sin(2\pi(ux+vy))$$

这些项是 x,y 平面中的正弦函数，它们的方向和频率由 u,v 确定。例如，考虑实数项，当 $ux+vy$ 是常数项时(也就是说，沿着 x,y 平面的一条直线，其斜率由 $\tan\theta=v/u$ 给出)，该实数项也是常数。这一项的梯度同 $ux+vy$ 为常数的直线正交，正弦的频率是 $\sqrt{u^2+v^2}$。这些正弦函数通常称为空间频率成分，图 7.5 显示了若干例子。

该积分应看做是点积运算。如果 u 和 v 固定，该积分的值是 x 和 y 的正弦函数与原始函数之间的点积。这种看法是十分有用的，因为点积度量出一个向量在另一个向量方向上的数量大小。

同样的方法，特定 u,v 处的变换值可以看做是对信号中特定频率和方向正弦数量进行的测量。变换将 x 和 y 的函数转换成 u 和 v 的函数，任何特定 (u,v) 处的值等于原始函数中特定正弦函数的数值。这种观点确定了傅里叶变换模型是一种基的变化。

线性性　傅里叶变换是线性的：

$$\mathcal{F}(g(x,y)+h(x,y))=\mathcal{F}(g(x,y))+\mathcal{F}(h(x,y))$$

和

$$\mathcal{F}(kg(x,y)) = k\mathcal{F}(g(x,y))$$

图 7.5 傅里叶变换基元素的实数部分表示为灰度图像。最亮的点数值为 1，最暗的点数值为 0。域的范围是 $[-1,1]\times[-1,1]$，以图像中心为原点。左图中，$(u,v)=(0,0.4)$；中间的图中，$(u,v)=(1,2)$；右图中，$(u,v)=(10,-5)$。这些是文中描述的不同频率和方向的正弦函数

傅里叶逆变换 把一个信号从它的傅里叶变换中恢复过来是很有用的。这是对基的另一种变化：

$$g(x,y) = \iint_{-\infty}^{\infty} \mathcal{F}(g(x,y))(u,v)e^{i2\pi(ux+vy)}\,du\,dv$$

傅里叶变换对 傅里叶变换在许多不同的情况下都很有用，大量的例子出现在 Bracewell 的文章中(1995)。在表 7.1 中列举了一些作为参考。表 7.1 的最后一行包含卷积定理，信号域上的卷积同傅里叶域上的乘法相同。下面(9.2.2 节)几次使用这个重要的性质。

表 7.1 一些二维函数及其傅里叶变换。这张表可以双向使用(将 u,v 和 x,y 适当的替代)，因为一个函数傅里叶变换的傅里叶变换还是这个函数。读者可能怀疑 δ 函数的无限加和结果同傅里叶变换的线性矛盾。通过仔细检查极限能够发现这两者并不矛盾(可以参见 Bracewell, 1995)。一般读者可能还注意到了 $\mathcal{F}\left(\frac{\partial f}{\partial y}\right)$ 可以通过将表的两行结合起来得到

函数	傅里叶变换
$g(x,y)$	$\iint_{-\infty}^{\infty} g(x,y)e^{-i2\pi(ux+vy)}\,dx\,dy$
$\iint_{-\infty}^{\infty} \mathcal{F}(g(x,y))(u,v)e^{i2\pi(ux+vy)}\,du\,dv$	$\mathcal{F}(g(x,y))(u,v)$
$\delta(x,y)$	1
$\frac{\partial f}{\partial x}(x,y)$	$u\mathcal{F}(f)(u,v)$
$0.5\delta(x+a,y)+0.5\delta(x-a,y)$	$\cos 2\pi au$
$e^{-\pi(x^2+y^2)}$	$e^{-\pi(u^2+v^2)}$
$box_1(x,y)$	$\frac{\sin u}{u}\frac{\sin v}{v}$

（续表）

函数	傅里叶变换
$f(ax, by)$	$\frac{\mathcal{F}(f)(u/a, v/b)}{ab}$
$\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\delta(x-i, y-j)$	$\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\delta(u-i, v-j)$
$(f**g)(x, y)$	$\mathcal{F}(f)\mathcal{F}(g)(u, v)$
$f(x-a, y-b)$	$e^{-i2\pi(au+bv)}\mathcal{F}(f)$
$f(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$	$\mathcal{F}(f)(u\cos\theta - v\sin\theta, u\sin\theta + v\cos\theta)$

相位和幅度 傅里叶变换包含一个实部和一个虚部：

$$\begin{aligned}\mathcal{F}(g(x, y))(u, v) &= \int\int_{-\infty}^{\infty} g(x, y)\cos(2\pi(ux+vy))\,dx\,dy \\ &\quad + i\int\int_{-\infty}^{\infty} g(x, y)\sin(2\pi(ux+vy))\,dx\,dy \\ &= \Re(\mathcal{F}(g)) + i * \Im(\mathcal{F}(g)) \\ &= \mathcal{F}_R(g) + i * \mathcal{F}_I(g)\end{aligned}$$

通常，在平面上画出复函数图像是很麻烦的。一种解决方法是分别画出$\mathcal{F}_R(g)$和$\mathcal{F}_I(g)$；另一种方法是考虑复函数的幅度和相位，并分别画出，因而分别称做幅度谱和相位谱。

一个函数在特定 u, v 的傅里叶变换值取决于整个函数，显然这可以从定义中得到，因为积分的域是整个函数域。这会导致一些错综复杂的性质。首先，一个函数的局部变化（例如，将一个区域的点置0）将导致傅里叶变换中每个点发生变化，这意味着傅里叶变换很难作为一个表达式用（比如，仅仅考察傅里叶变换很难说图像中具有某个特征）。其次，图像的幅度谱经常是相似的。这一点看起来是自然现象，而不是定理证明的结果。因此，图像的幅度谱提供的信息是相当有限的（图7.6就是一个例子）。

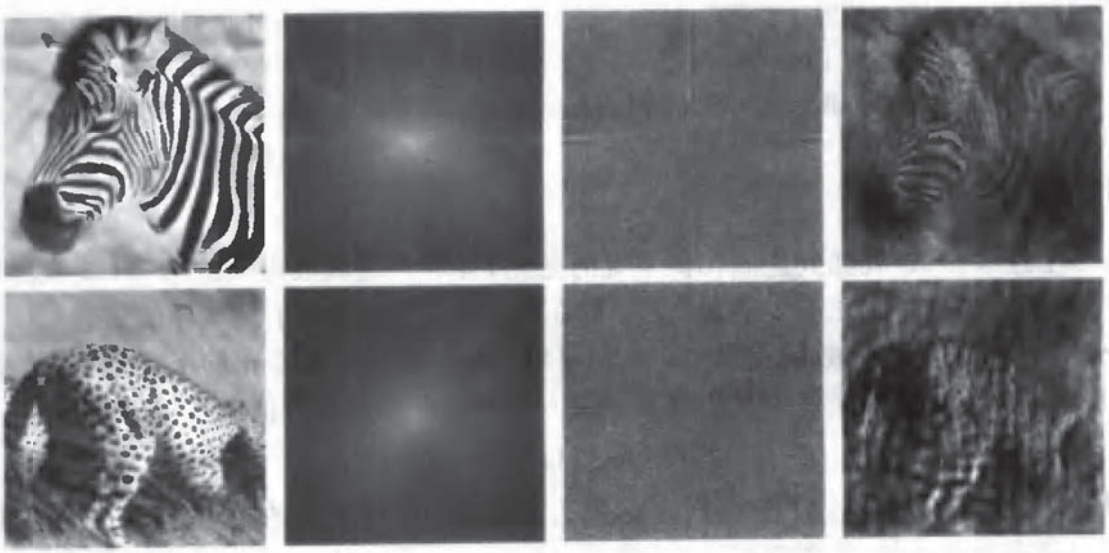


图7.6 每一行的第二张图片给出第一张图片对数表示的幅度谱；第三张图片给出相位谱：$-\pi$ 是暗的，π是亮的。最后的图片是幅度谱变换回来后的情况。尽管反变换引入一些图像噪声，但并不影响对图像的解释，同时可以看出，在理解图像方面，相位谱比幅度谱更有意义

7.4 采样和折叠失真

讨论傅里叶变换的最重要原因在于进一步认识离散和连续图像间的不同，尤其是对离散像素阵列进行运算时一些信息丢失了，但究竟丢失了什么呢？国际象棋盘是一个很好而且简单的例子，如图 7.7 所示。问题在于采样数目与函数的相对关系，在给出一个强有力的模型条件下，这个问题可以描述得相当精确。

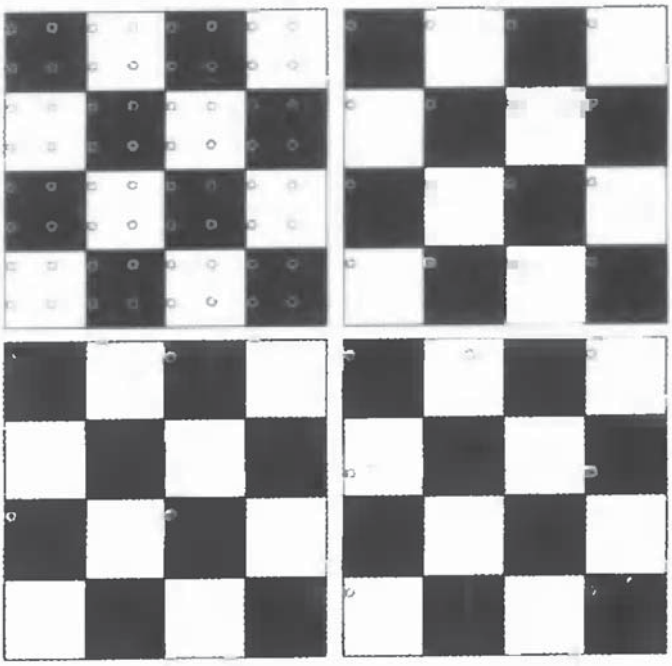


图 7.7 位于上方的两个国际象棋盘展示了一个看起来成功的采样过程(不论它是否依赖于以后将处理的一些细节)。灰色的圆圈代表采样；如果有足够的采样，采样点能够代表潜在函数的细节。下面的两种采样显然是不成功的，采样点表示的格子比实际的格子少。这说明了两个现象：首先，成功的采样有足够的采样数据；其次，不成功的采样安排导致高频信息呈现为低频信息的形式

7.4.1 采样

从一个连续函数(比如到达照相机系统的亮光)得到一组离散点上的数值(就像照片中每个像素值)称为采样。构建一个模型，可以准确地得到采样中丢失了什么样的信息。

一维采样 对一个函数的一维采样可以得到一组离散值。最重要的采样是等距的离散点采样，并假设只在整数点采样。这意味着从输入的函数返回一个数值向量：

$$\mathrm{sample}_{1D}(f(x)) = \boldsymbol{f}$$

我们对采样过程进行建模的方法是假设向量元素的值是函数 $f(x)$ 在相应采样点的函数值，同时允许向量出现负序号(见图 7.8)，这意味着 $\boldsymbol{f}$ 的第 i 个元素是 $f(x_i)$。

二维采样 二维采样同一维采样相似。尽管采样可以出现在非正规点(最好的例子是人类的视网膜)，但是在这里仍然假设采样在整数坐标上。这样就产生了一个等距的矩形网格，这对大多数摄像机都适用。采样图像就是有限尺寸的矩形数组(所有网格外的值为零)。

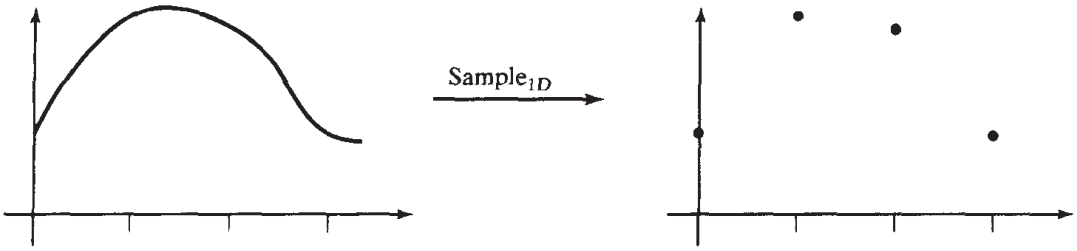


图 7.8　如上图所示，对函数的一维采样得到一个向量，每一个元素是函数在该采样点的数值。对于上述目的，自变量整数值作为采样点已经足够。允许向量是无限维，并可以有正序号和负序号

用形式化模型的术语，对一个二维函数（而不是一维函数）进行采样时，得到一个矩阵（见图 7.9）。这个矩阵中的每个方向上都允许存在负序号，并记做

$$\texttt{sample}_{2D}(F(x,y)) = \mathcal{F}$$

这里矩阵$\mathcal{F}$的第 i,j 个元素为 $F(x_i, y_j) = F(i,j)$。

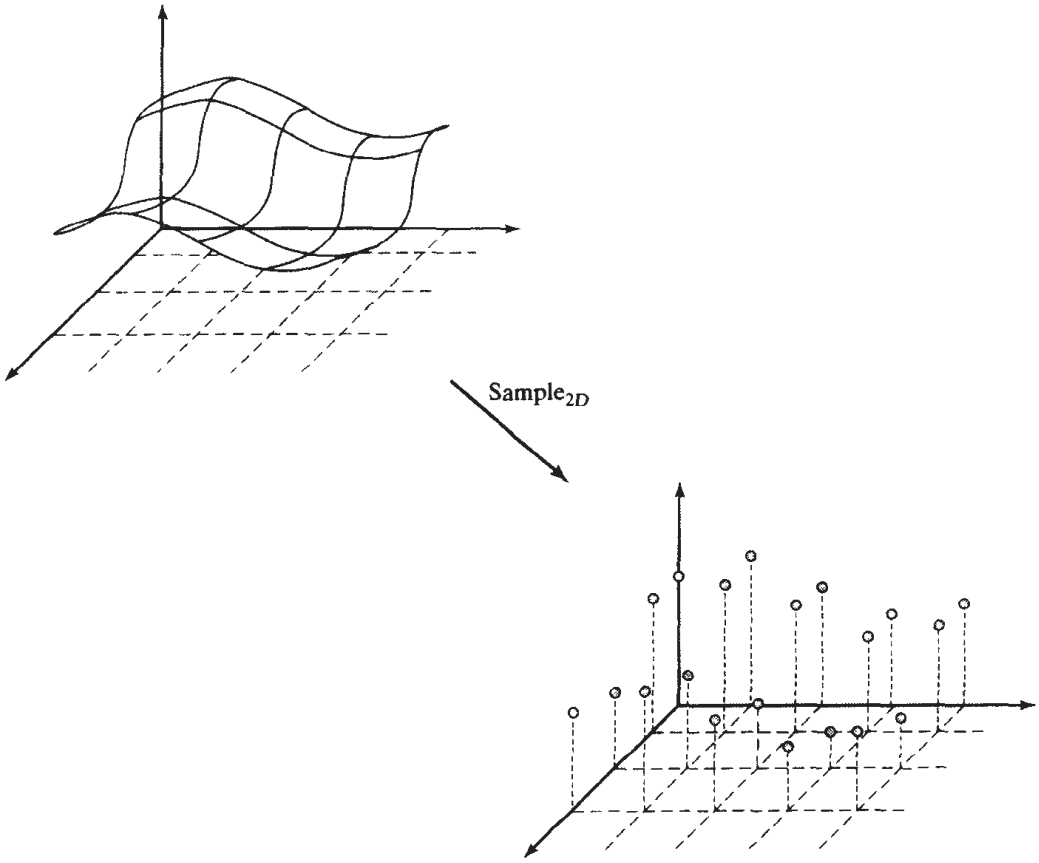


图 7.9　对函数的二维采样得到一个矩阵；矩阵仍然可以是无限维，并可以有正序号和负序号

在实际系统中，采样并不总是均匀分布的。由于电视接收机无处不在，这种情况更是经常发生出现，因为电视屏幕的可视比例是 4:3（宽度:高度）。摄像机经常通过将水平方向隔开的距离比竖直方向大（行话讲是“非方形像素”）的方式实现这种效果。

采样信号的连续模型　我们需要一个采样信号的连续模型。通常这个模型被用为估算积分——特别的，采用傅里叶变换包括对模型中复指数的积分。这个积分如何运行是清楚的——积分值应该通过把每个整数点的值加起来得到，这意味着不能建模为一个除了整数点（这

里获取信号值)外其他各点均为零的函数,因为这种函数的积分为零。

一个采样信号合适的连续模型依赖于 δ 函数的一个重要性质,

$$\begin{aligned}\int_{-\infty}^{\infty} a\delta(x)f(x)\,dx &= a\lim_{\epsilon\to 0}\int_{-\infty}^{\infty} d(x;\epsilon)f(x)\,dx \\ &= a\lim_{\epsilon\to 0}\int_{-\infty}^{\infty}\frac{bar(x;\epsilon)}{\epsilon}(f(x))\,dx \\ &= a\lim_{\epsilon\to 0}\sum_{i=-\infty}^{\infty}\frac{bar(x;\epsilon)}{\epsilon}(f(i\epsilon)bar(x-i\epsilon;\epsilon))\epsilon \\ &= af(0)\end{aligned}$$

这里使用积分的概念作为条形加和的极限。

一个采样信号合适的连续模型由在每个采样点的一个加权 δ 函数组成,权值是采样点的采样值。通过用在每个采样点的一系列 δ 函数乘以被采样信号可以获得这一模型。一维情况下,这样的函数称为梳齿形函数(因为它的图像看起来是这样的形状);二维情况下,这样的函数称为钉床函数(同样的原因)。

讨论二维情况并假设在整数点采样,可以得到

$$\begin{aligned}\mathtt{sample}_{2D}(f) &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} f(i,j)\delta(x-i,y-j) \\ &= f(x,y)\left\{\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\delta(x-i,y-j)\right\}\end{aligned}$$

该函数除了整数点外函数值为 0(因为 δ 函数除了整数点外都是 0),它的积分值也是这些整数点函数值的和。

7.4.2 折叠失真

采样中会有信息的丢失。这一节将指出,采样过慢的信号会显示错误结果;原始信号中的高频空间元素在采样信号中会表现为低频元素——这种效应称为折叠失真。

采样信号的傅里叶变换 采样信号是原始信号和钉床函数结合的产物。根据卷积定理,两个函数乘积的傅里叶变换是两个函数傅里叶变换的卷积,这意味着采样信号的傅里叶变换等于函数的傅里叶变换同另一个钉床函数傅里叶变换的卷积。

由于将一个函数同一个位移 δ 函数卷积只是位移了这个函数(见练习)。这意味着采样信号的傅里叶变换是信号的一系列傅里叶变换位移版本的和,表示为

$$\begin{aligned}\mathcal{F}(\mathtt{sample}_{2D}(f(x,y))) &= \mathcal{F}\left(f(x,y)\left\{\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\delta(x-i,y-j)\right\}\right) \\ &= \mathcal{F}(f(x,y)) ** \mathcal{F}\left(\left\{\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\delta(x-i,y-j)\right\}\right) \\ &= \sum_{i=-\infty}^{\infty} F(u-i,v-j)\end{aligned}$$

这里把函数 $f(x,y)$ 的傅里叶变换写为 $F(u,v)$。

如果这些信号傅里叶变换的位移版本并不互相交叠，则很容易从采样重构原始信号：对采样信号进行傅里叶变换，取出傅里叶变换的一个副本，再进行反变换（见图 7.10）。

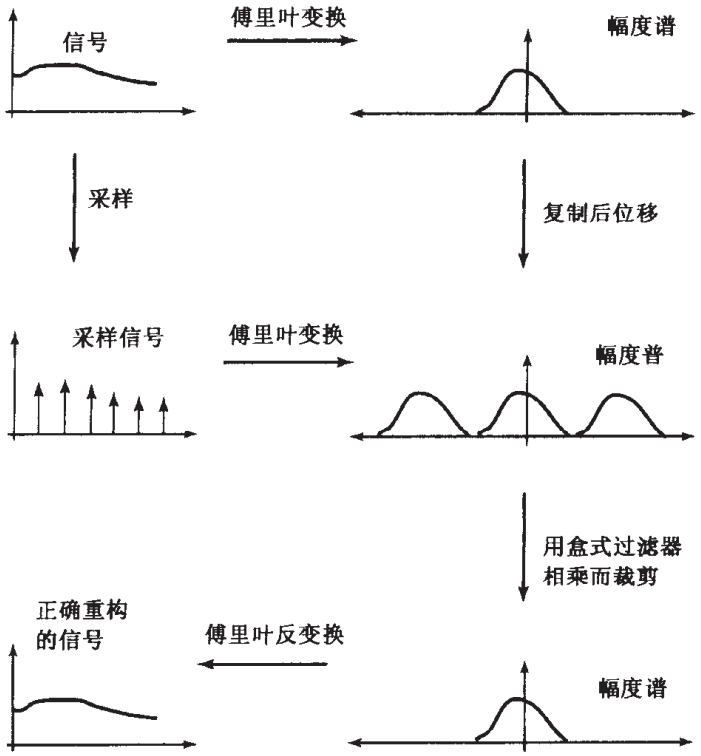


图 7.10　采样信号的傅里叶变换由原始信号的傅里叶变换根据采样频率偏移后的复制组合构成。有两种可能性：如果偏移的部分互不相交叠（这里的情况），原始信号能够根据采样信号重构（裁剪傅里叶变换中的一块，进行反变换）。如果偏移部分相交叠（见图7.11），则由于相交叠区域被叠加，无法获得单独的傅里叶变换，信号是折叠失真的

但是，如果相关区域的确相互交叠，信号就不能重建，因为无法确定相叠区域的傅里叶变换。这里不同的傅里叶变换叠加在一起，结果受到很大影响，通常称为折叠失真，高频空间呈现为低频空间（见图 7.12 和练习）。这里的讨论涉及到奈奎斯特定理——采样频率至少是信号最高频分量两倍以上时，才能够从采样信号重构原始信号。

7.4.3　平滑和重采样

奈奎斯特定理意味着，不能对一个图像取每 k 个像素的方法进行压缩（如图 7.12 所示）。相反，需要对图像进行过滤以便去除高于采样频率的空间频率。这可通过将图像的傅里叶变换与二维栅函数相乘实现，起到低通滤波器的作用，或等价的，可以把图像与 $(\sin x \sin y)/(xy)$ 这种形式的核卷积。这是复杂且昂贵的（“不可能”的一种客气说法）卷积，因为函数有无限支集。

最有趣的情况发生在需要将图像高度和宽度减半时。假设采样图像没有折叠失真（如果有折叠失真，我们将无能为力；一旦图像被采样，任何可能发生的折叠失真都将出现，如果没有一个图像模型我们所能做的也十分有限）。这意味着采样图像的傅里叶变换将是一些傅里叶变换的复制，并将中心转移到 u,v 空间的整数点。

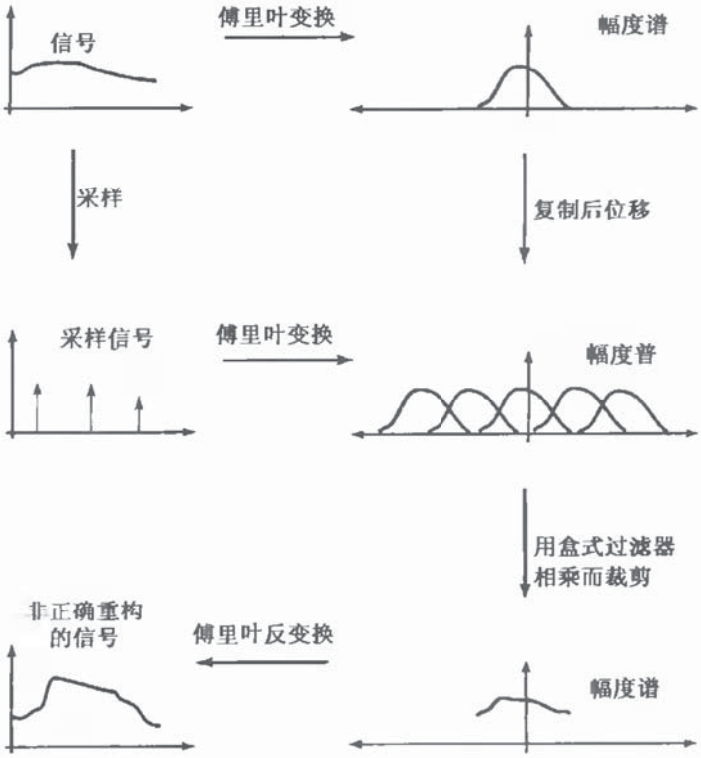


图 7.11 采样信号的傅里叶变换由原始信号的傅里叶变换根据采样频率偏移后的复制组合构成。有两种可能性：如果偏移的部分互不相交叠（见图7.10），原始信号能够根据采样信号重构（裁剪傅里叶变换中的一块，进行反变换）。如果偏移部分相交叠（这里的情况），相交叠区域被叠加，无法获得单独的傅里叶变换，信号是折叠失真的。这也说明高频信号具有向低频信号折叠失真的趋势

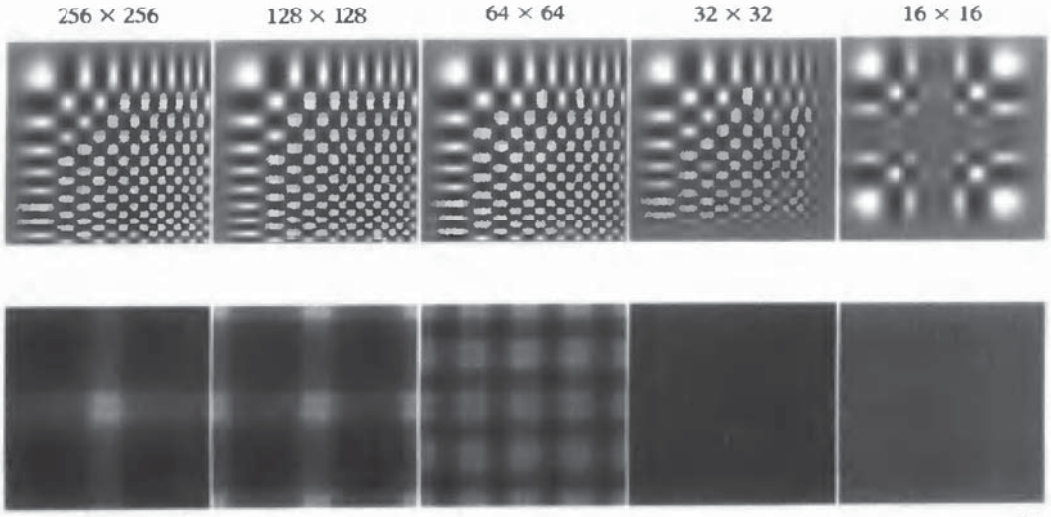


图 7.12 上面一行是网格图像与两个频率线性增长的正弦函数（一个在 x 方向，一个在 y 方向）相乘后的采样图像。这一序列中其他图像通过没有平滑的两倍重采样系数获得（也就是说，这一个是64 × 64，则下一个是128 × 128，等等。所有刻度都是相同尺寸的）。注意有明显的折叠失真；高频空间折叠失真到低频空间，最小的图像表示大图像的效果很差。下面的一行给出了每幅图傅里叶变换的对数幅度谱（取对数压缩尺度空间），中心的是常量。注意到重采样的傅里叶变换是通过原始图像比例缩小后的傅里叶变换排列获得的。原始傅里叶变换复制的相互影响意味着在某些点无法恢复；这就是折叠失真造成的

如果对信号重采样，这些复制将出现在 u,v 空间的半整点。这意味着，为了避免折叠失真，需要使用一个能够很强地减少原始傅里叶变换在区域 $|u|<1/2,|v|<1/2$ 外的内容的滤波器。当然，如果我们减少这个区域内的信号，也会损失一些信息。由于高斯函数的傅里叶变换仍然是高斯函数，并且高斯函数迅速衰减，因此，如果我们将图像和高斯函数进行卷积——或者乘上高斯函数傅里叶变换的结果，情况相同——就能够获得需要的结果。

高斯函数的选择取决于应用；如果 σ 很大，则折叠失真较小(因为范围之外的核的值非常小)，但是也会丢失信息，因为核的值在区域内不是平坦的；与此类似，如果 σ 很小，区域内的信息丢失减少，但是折叠失真将会加大。图 7.13 和图 7.14 说明了选取不同 σ 值的效果。

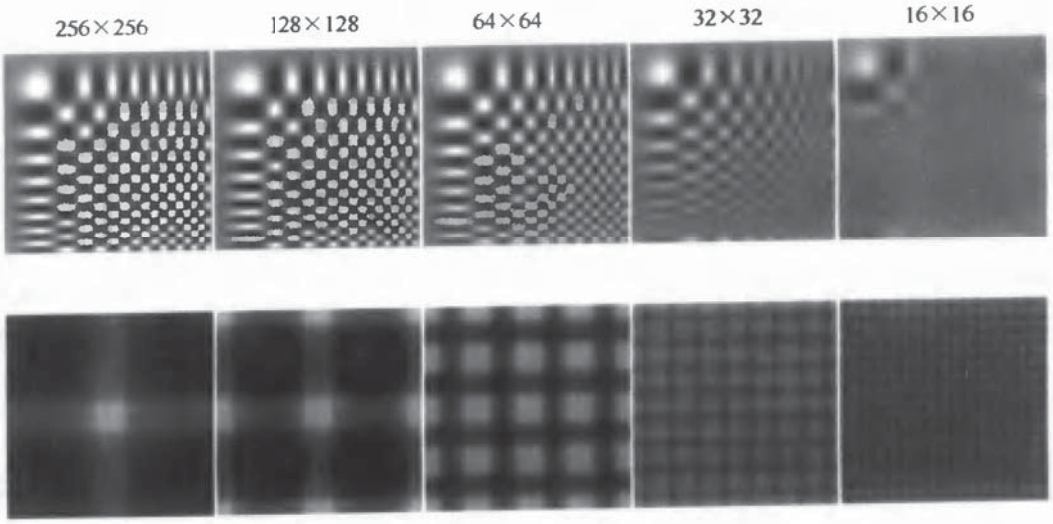


图 7.13 上面一行：图 7.12 的重采样版本，同样是两倍重采样，但是这次每幅图像重采样之前使用σ为一个像素的高斯函数平滑。这个滤波器是低通滤波抑制了高频分量，减少了折叠失真。下面一行：低通滤波的效果很容易在这些对数幅度图像中看出；低通滤波抑制了高频分量，所以分量间干扰降低，从而减少折叠失真

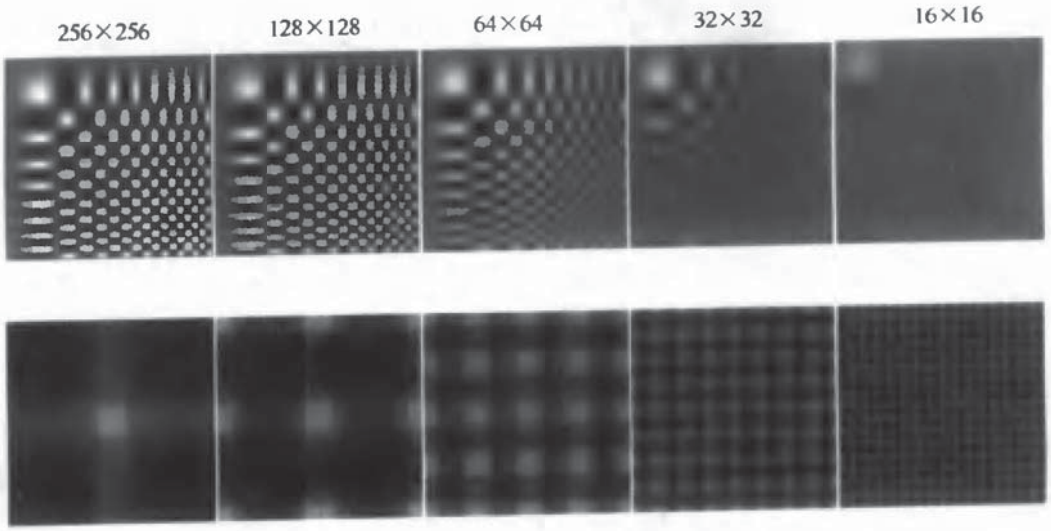


图 7.14 上面一行：图 7.12 中图像重采样版本，同样是两倍重采样，但是这次每幅图像重采样之前使用σ为两像素的高斯函数平滑。滤波器比图7.13的情况更有效地消除了高频部分。下面一行：低通滤波的效果很容易在这些对数幅度图像中看出；低通滤波抑制了高频分量，所以分量间干扰降低，从而减少折叠失真

使用一个高斯函数作为低通滤波是因为它对高频分量响应很小，低频分量响应较大。事实上，高斯函数不是一个特别好的低通滤波器。我们需要的是一个滤波器，它的响应接近于低频范围——带通；同时在高频区域响应接近于常数(零)——带阻。设计一个比高斯滤波更好的低通滤波器是可能的，设计过程包括仔细权衡波纹准则(带通和带阻的响应是否平滑)和衰减(响应衰减到零并停留在那的速度有多快)。图像重采样的基本步骤在算法 7.1 中给出。

算法 7.1 对图像的二倍重采样

对原始图像使用低通滤波器

(σ 在 1～2 个像素之间的高斯函数通常是合适的选择)

产生一个新的图像，边的维度是原始图像的一半

把新图像第 i,j 个像素的值设置为滤波后图像的第 $2i,2j$ 个像素。

7.5 滤波器与模板

滤波器为检测简单模式提供了自然直接的机制，因为滤波器对类似滤波器的模式元素有很强的响应。例如，平滑的导数滤波器在导数值很大的地方响应强烈。在这些点，滤波器的核看起来类似于要检测的特征。x 方向的导数滤波器看起来像间隔排列的竖直亮条和暗条(这些区域在 x 方向上有较大的导数值)，等等。

通常，滤波器对类似它的特征有明显响应(见图 7.15)。这是一个简单的几何结论。

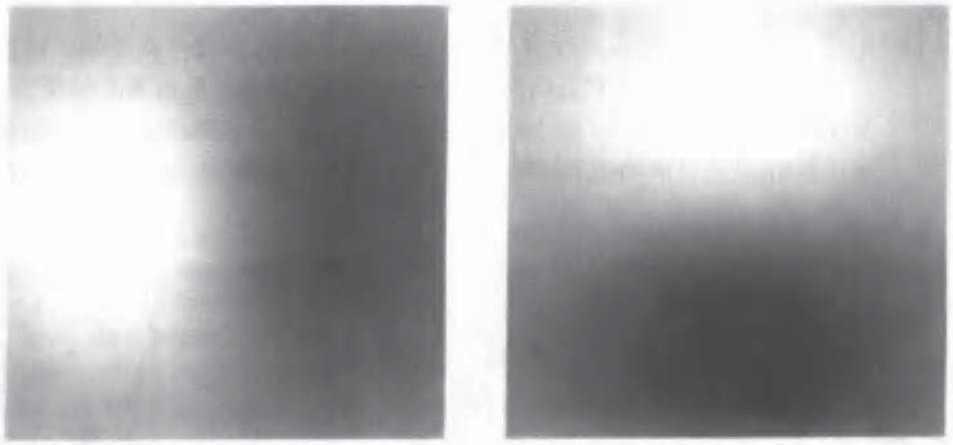


图 7.15 滤波器核与它要检测的特征看起来类似。左图中，平滑的高斯滤波器的导数检测在 x 方向的大变化(就像一个暗条旁边是一个亮条)；右图中，平滑的高斯滤波器的导数检测在 y 方向的大变化

7.5.1 卷积与点积

回顾 7.1.1 节，对一些线性滤波器的核 $\mathcal{G}$ 而言，这些滤波器对图像 $\mathcal{H}$ 的响应由下式给出：

$$R_{ij} = \sum_{u,v} G_{i-u,j-v} H_{uv}$$

当 i、j 都为 0 时，考虑滤波器在这一点的响应，则有：

$$R = \sum_{u,v} G_{-u,-v} H_{u,v}$$

滤波器响应是通过将图像元素和滤波器元素联系起来获得的，将相关元素相乘后累加。将图像扫描为一个向量，滤波器的核为另一个向量，这样相关的元素就处在向量相同位置。通过插入0元素，可以保证两个向量的维数相同。这样一来，将相关元素相乘后累加就与点积运算一样了。

这是一个非常有用的类比，因为当代表图像的向量与代表滤波器核的向量平行时，点积具有最大值。这意味着当图像模式类似于滤波器模式时，滤波器响应最大。当模式越明显时，滤波器的响应也会变大。

现在考虑在图像一些其他点上对滤波器的响应。模型中的特征没有明显改变。把图像再次扫描为一个向量，滤波器的核为另一个，这样相关元素就处在向量相同位置。同样的，应用滤波器的结果是一个点积。对这种点积有两种有用的思考方式。

7.5.2 基的改变

我们认为卷积是图像和另一个不同的向量在每一点上的点积(因为我们移动了滤波器的核，暂时搁置图像中的其他点)。这个新的向量通过重新规划老的向量，使得元素处在正确的位置，可以得到累加和。这意味着，通过将图像与滤波器进行卷积，使用图像向量空间中一组新的基表示图像——这组基通过滤波器的不同移位版本获得。原始的基元素中除了某一特定位置为1外均为0元素。新的基元素是某种模式的位移版本。

对于讨论的大部分核，这个滤波过程会损失信息——同平滑消除噪声的原因相同——而这组基的系数突显出来。当系数突显出来时，它们有效地揭示出图像的结构；这种基的转化对纹理分析是有价值的。典型的做法是选择一组包含小的、有效的特征元素的基。基的系数的较大值意味着存在一个模式元素，纹理能够通过这些模式元素之间的关系来表示，这些关系通常使用一些概率模型。

7.6 技术：归一化相关和检测模式

可以把卷积视为将滤波器与以某一点(这个点是当前关注点)为中心的一块图像进行的比较。这样，图像中对应滤波器核的邻域可以扫描为一个向量，用于同滤波器的核进行比较。点积本身很难检测到特征，因为图像区域很亮时点积值也会变得很大。通过与向量的类比，滤波器向量和图像邻域向量的夹角余弦有着特殊意义；这意味着要计算图像相关区域(位于滤波器的核范围内的图像元素)的几何平均值，并将该值除滤波器响应。

当图像区域看起来类似滤波器的核时，可以得到一个大的正值；当图像区域看起来同滤波器的核对比度相反时，得到一个小的负值。如果对比度相反的不必比较时，可使用值将被平方。这种方法叫做归一化相关，是一种简单有效的检测模式的方法。

7.6.1 通过归一化相关检测手的方法控制电视

设计一套根据人的手势来操控的某些系统是非常有趣的。例如，当你对着灯挥手时，可以使房间照明变亮，指点空调器使室内的气温改变，或者对电视屏幕上一个令人讨厌的政客做一个合适的动作就可以改变频道。在典型的实际应用中，对可获得的计算量有着严格的限制，这意味着问题的关键在于手势识别系统必须足够简单。当然，这样的系统其功能通常是非常有限的。

控制电视机 一种典型的情况是，用户界面处在某种状态——可能正在显示一个菜单，一

个事件发生了——可能遥控器给出一个指示，这个事件导致用户界面改变状态——一个新的菜单项被突显出来，整个进程继续进行。在某些状态中，一些事件驱使系统控制某种动作——频道可能改变了。所有这些意味着对于用户界面而言，状态机是一个很自然的模型。

要让视觉适合这种模型的一种方法是提供事件。由于只有较少的几种事件因而这是可行的。我们知道在某种特定状态下，系统应该关注哪种事件。因此，视觉系统只需要确定有没有事件发生，少量特定事件的哪种事件发生。构造满足这些约束条件的系统往往是能够做到的。

要设计少量的动作以模拟远程控制；需要类似按键的事件（例如，开关电视），类似点击的事件（例如，调大音量，这也可能使用按键完成）。通过这些事件，可以打开电视，并得到屏幕菜单系统导引。

检测手 Freeman，Anderson 等人（1998）研制了一套能空手遥控开电视的界面，该系统之所以是鲁棒的，是因为系统需要做的只是确定视野中是否存在一只手。而用户通过抬起手臂张开手进行操作。因为用户一般在距离摄像机较远处操作——手的尺寸是大体知道的，所以没有必要改变尺度搜索。在电视机前，需要搜索和确定手的区域是很小的。

开电视时，手必须以相当标准的姿势和方向抬起，与电视的距离也是确定的（所以我们知道它应是怎样的情形）。这意味着归一化相关值足以找到手的位置。相关图像任何一个点的归一化相关值足够高时，它对应于手。这种方法也可以用来控制音量和其他操作，比如打开电视机和关闭电视机。这样，需要一些确认手应该怎么动作的规则——向一边动表示增大音量，向另一边动表示减小音量——这一点可以通过对比前一帧同当前帧的位置得到。系统通过一个图标显示手的位置，所以使用者能够得到系统正在做什么的反馈信号（见图 7.16）。注意到这种方法的一个显著特征是自校正。应用这种方法时，可在安装电视机后，坐在电视前试着移动你的手几次，以便让系统估计手可能出现的位置和手的尺寸。

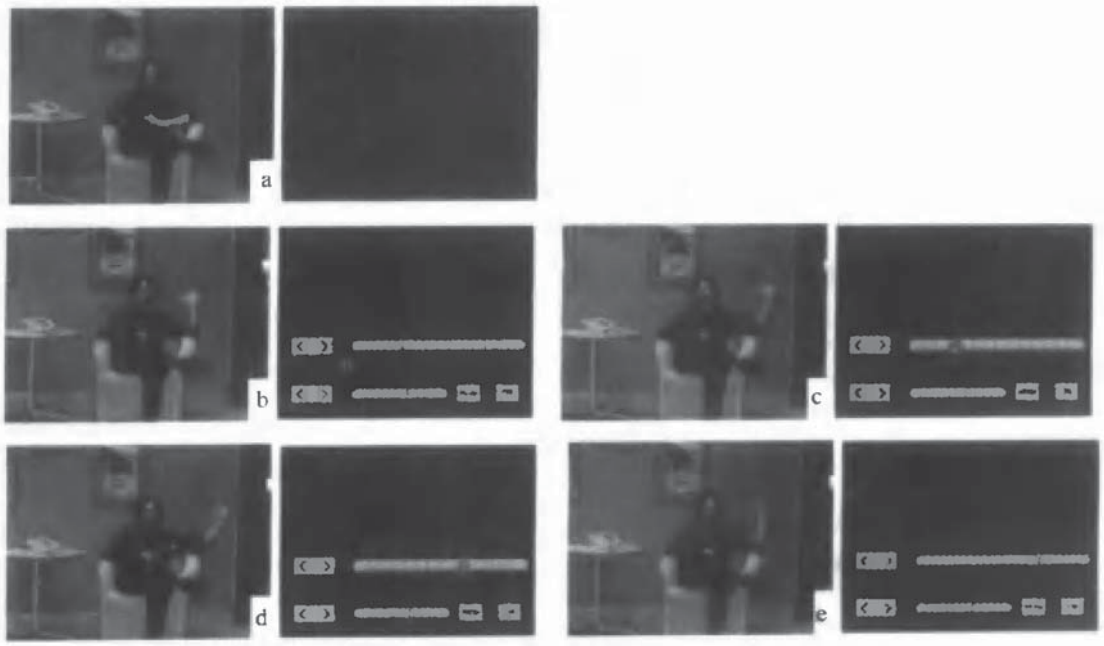


图 7.16 Freeman 等人控制电视机系统的例子。每个状态中，左边是电视机看到的使用者，右边是使用者看到的电视机的情况。(a)中，电视机没有打开，但一个线程在注视使用者。(b)中，一只张开的手使得电视机打开，出现用户界面控制面板。(c)中，面板中聚焦在使用者手的运动上。(d)中，使用者通过操作移动屏幕上的图标，切换电视频道。(e)中，手合上，关闭电视机

7.7 技术:尺度和图像金字塔

不同尺度的图像看起来非常不同。例如,图7.17中斑马的鼻子能够被描述出每根绒毛——它需要根据对几个像素的小尺度有向滤波器的响应进行编码——或者是斑马身上斑纹的形式。在斑马的例子中,我们不希望用大滤波器寻找斑纹,因为这些滤波器可能出现不真实的精度——不需要表示斑纹中的每一个绒毛,而且也不便构造,应用起来也会很慢。与使用大滤波器相比,更实用的方法是对平滑重采样后的图像使用小滤波器。

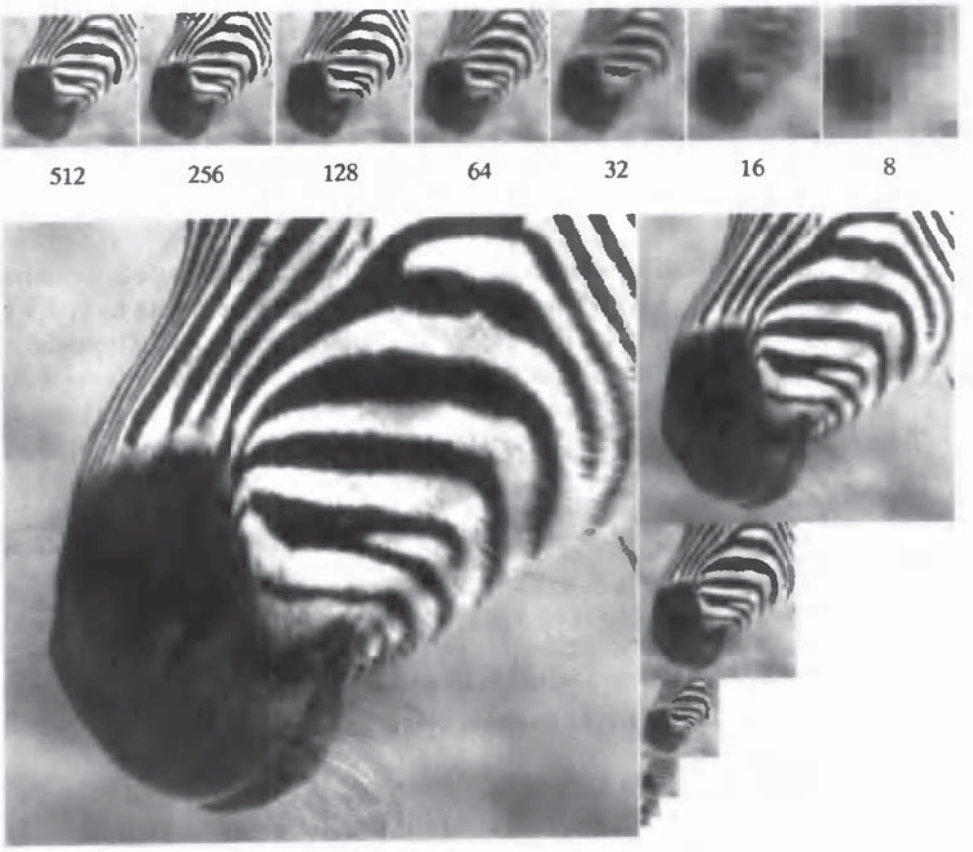


图7.17 图像的高斯金字塔从512×512到8×8。上面一行中,把每张图以同样的尺寸显示(其中的一些像素尺寸较大),图的下面一部分是每张图的真实尺寸。注意到,如果将每张图与一个固定尺寸滤波器进行卷积,将获得截然不同的现象。最细粒度的图像中一个8×8像素的块将包含一些绒毛;在粗一些的粒度中可能包含一个完整的斑纹,同时在最粗粒度中,包含动物的鼻子

7.7.1 高斯金字塔

图像金字塔是对一幅图像表示的集合,其名字来源于视觉上的相似性。典型的,金字塔的每一层是前一层宽度和高度的一半;如果新的一层构建在前一层上面,就可以形成一个金字塔。在高斯金字塔中,每一层使用一个对称的高斯核进行平滑,同时进行重采样以获得下一层(见图7.17)。如果图像尺寸是2的幂,或者是2的幂的倍数,那么构建这些金字塔通常是很方便的。最小的图像得到最好的平滑,通常将这些层称为图像的粗尺度。

算法 7.2 形成高斯金字塔
将图像设定为最细尺度层
对每一层,从最细层的上一层到最粗层
　　通过对上一个最细一层使用高斯函数进行平滑,然后重采样获得这一层
结束

借助一个小符号能够写出高斯金字塔各层的简单表达式。符号 $S^{\downarrow}$ 对图像重采样;具体说来,$S^{\downarrow}(\mathcal{I})$的第 j,k 个元素是$\mathcal{I}$的第 $2j,2k$ 个元素。金字塔 $P(\mathcal{I})$的第 n 层表示为$P(\mathcal{I})_n$。使用这个记号,便有

$$P_{\text{Gaussian}}(\mathcal{I})_{n+1} = S^{\downarrow}(G_\sigma * * P_{\text{Gaussian}}(\mathcal{I})_n)$$
$$= S^{\downarrow}G_\sigma(P_{\text{Gaussian}}(\mathcal{I})_n)$$

(其中,用 G_σ 作为对图像进行线性操作,而其将图像与高斯函数做卷积)。细节最丰富的一层是原始图像:

$$P_{\text{Gaussian}}(\mathcal{I})_1 = \mathcal{I}$$

7.7.2 多尺度表示的应用

高斯金字塔很有用,因为它使得提取同一幅图像中不同类型的结构表示变为可能。存在三种典型的应用。

多尺度搜索 许多物体能表示为小图像模式,正面脸部图像就是一个例子。典型的,低分辨率时,脸部特征具有特色鲜明的模式:眼睛形成暗圈,在一条暗条下面(眉毛),一个亮条分开(镜子对鼻子的反射),它又在一个暗条之上(嘴)。有很多种使用这些特征检测脸形的方法(见第 22 章)。这些方法都假设脸位于一个小的尺度范围内。所有的脸形可通过搜索金字塔找到。为了找到比较大的脸形,寻找粗一些的层;为了找到比较小的脸形,寻找细一些的层。这个有用的方法应用于许多不同类型的特征,这些在后面章节中可以看到。

空间搜索 另一个应用是空间搜索,其是计算机视觉中的一个常见主题。典型的,一张图片中有一个点,需要在第二张图片中找到一个点对应于这个点。这个问题出现在立体观测中——因为两张图从不同视角获得,所以点的位置发生了改变;或是发生在运动分析时——由于摄像机的运动,或者点处在运动物体上而造成图像点的运动。

在原始图像中搜索匹配是低效的,因为不得不处理大量的细节。现在一个非常流行且更好的方法是,先在高度平滑和重采样后的图像中寻找匹配,继而在图像详细版本中提高匹配的精确度。举例来说,将一个 1024×1024 的图像变为 4×4 的版本,进行匹配,然后在 8×8 版本中精炼(在一个粗糙匹配中,很容易进行精炼);继而在 16×16 版本中精炼,等等,一直到恢复到 1024×1024。它提供了一个特别有效的搜索,因为 4×4 版本中的一个像素对应于 1024×1024 版本中的 256 个像素。这个方法叫做由粗到精的匹配。

特征跟踪 平滑的粗糙层中找到的大多数特征,与较大的且高对比的图像事件相关,因为对于粗糙尺度下标出的图像特征,需要细尺度图像中很多像素对它进行确认。典型的,寻找粗糙尺度现象会错误估计特征的尺寸和位置。例如,粗糙尺度下一个像素的误差,表示了精细尺度下成倍增加的像素误差。

在精细尺度下会有许多特征，其中的一些同较小的、低对比图像事件相关。改进精细尺度下获得的一系列特征的一种方法是，跟踪特征到较粗糙尺度，接收在较粗糙尺度下能找到对应的精细尺度特征。这个方法，一般称为特征跟踪，其能够抑制纹理区域（通常称为噪声）引起的特征和真实噪声引起的特征。

7.8 注释

在对线性系统的介绍中不可能没有遗漏。但如果不能领会这一章中的一些核心概念，则很难读懂视觉中滤波器的关键内容。这里给出了一个简洁直接的总结，更多的细节可参阅 Bracewell 的论著（1995），（2000）。

真实图像系统同移不变线性系统的对比

图像系统只是近似线性的。胶片是非线性的——对一个弱激励没有响应，对一个强激励会达到饱和——但是在一定的合适范围内仍然可以建立一个线性模型。CCD 摄像机在工作范围内是线性的，但作为热噪声影响的结果，对 0 输入会给出一个很小的非 0 响应（这就是为什么天文学家需要冷却摄像机的缘故），并且对很强的激励会达到饱和。CCD 摄像机通常包含电子转换线路，使得输出效果类似胶片，因为消费者更习惯于胶片。移不变是近似的，因为镜头在图像边缘处会产生失真。一些镜头——比如鱼眼镜头——不是移不变的。

尺度

尺度空间和尺度表示方面有大量的文献，它起源于 Witkin（1983）的工作，并由 Koenderink 和 van Doorn（1986）发展。从那以后，出现了大量的相关工作（一些开始于 ter Haar Romeny，Florack，Koenderink 和 Viergever，1997；或者 Nielsen，Johansen，Olsen 和 Weickert，1999）。这里只给出了最简要的描述，因为分析需要特殊的技巧。技术的用途也是目前热烈争论的议题。

针对不同方向与特性的尺度化

尺度空间模型的一个重要困难在于对称性高斯平滑过程趋向于过分光滑了边缘。例如，天边有两棵相邻的树，在代表每棵树细节的小尺度快完成合并之前，对应于两棵树的大尺度就可能已经合并了。这意味着，在边缘点应该采用不同的平滑方法。例如，我们可以估计梯度的幅值和方向；对大梯度，使用有向平滑操作，沿垂直于梯度方向进行平滑；对小梯度，则使用对称平滑操作。这种方法通常称为保留边缘的平滑。

现在更常规的版本是 Perona 和 Malik（1990a，b）的工作，注意到尺度空间表示家族是一个扩散方程的解决方案

$$\begin{aligned}\frac{\partial \Phi}{\partial \sigma} &= \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} \\ &= \nabla^2 \Phi\end{aligned}$$

初始值为

$$\Phi(x, y, 0) = \mathcal{I}(x, y)$$

如果等式在同样的初始条件下改为如下形式

$$\begin{aligned}\frac{\partial \Phi}{\partial \sigma} &= \nabla \cdot (c(x, y, \sigma)\nabla \Phi) \\ &= c(x, y, \sigma)\nabla^2 \Phi + (\nabla c(x, y, \sigma)) \cdot (\nabla \Phi)\end{aligned}$$

那么,如果 $c(x,y,\sigma)=1$,就得到前面那个扩散方程,如果 $c(x,y,\sigma)=0$,就没有了平滑。这里,假设 c 不依赖于 σ。如果知道图中边缘的位置,就能够构造一个模板,由 $c(x,y)=1$ 与 $c(x,y)=0$ 的区域组成,$c(x,y)=0$ 是沿着边缘的值的区域,将 $c(x,y)=1$ 的区域分隔开,这样一来,只在被隔开的区域内进行平滑,而不会跨过边界。尽管我们不知道边缘在哪里——否则练习为空——我们能够从图像梯度的幅值中,对 $c(x,y)$ 做出合理的选择。如果梯度较大,c 应该小,反之亦然。有不少参考文献讨论这种方法;ter Haar Romeny(1994)的工作可以作为一个起点。

习题

7.1 证明下式给出的非加权局部平均运算

$$\mathcal{R}_{ij} = \frac{1}{(2k+1)^2} \sum_{u=i-k}^{u=i+k} \sum_{v=j-k}^{v=j+k} \mathcal{F}_{uv}$$

是一个卷积。这个卷积的核是什么?

7.2 一幅写为$\mathcal{E}_0$ 的图像除了中心为 1,其余各点都是 0,证明将这幅图像同核

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{((i-k-1)^2 + (j-k-1)^2)}{2\sigma^2}\right)$$

(一个离散高斯函数)卷积后,形成一个对称的圆形的光滑斑点。

7.3 证明:将一幅图像同离散可分离二维滤波器的核卷积,与同两个一维滤波器的核卷积等价。估计用此方法对一个 NxN 的图像和一个 $2k+1 \times 2k+1$ 的核操作节省的操作数目。

7.4 证明一个函数与 δ 函数卷积后重新得到原函数。证明将一个函数与偏移的 δ 函数卷积相当于对函数的偏移。

7.5 我们说将一幅图像同形如$(\sin x \sin y)/(xy)$的核进行卷积是不可能的,因为该种函数有无限支集。那么为什么不可以对图像进行傅里叶变换,将傅里叶变换乘上窗口函数,再进行傅里叶反变换呢?提示:考虑支集区域。

7.6 折叠失真将高频空间变成低频空间。解释为什么会发生如下现象:

(a) 老的西部牛仔电影中四轮马车在移动,轮子看起来似乎静止甚至反方向转动(也就是说马车自左向右运动,轮子看起来反时针方向转动)。

(b) 电视里带有很细暗条纹的白衬衣经常产生一个微光的颜色矩阵。

(c) 在射线跟踪图像中,斑驳的区域产生温和的影子。

编程作业

7.7 获得高斯核的一种方法是对一个常数核自身卷积多次。对比这种方法并评估一个高斯核。

(a) 为了获取一个合适的近似需要重复多少次卷积?(需要确定合适近似的概念,可能需要表示近似的质量和卷积重复次数的关系。)

(b) 这样做能获得怎样的好处?(提示:并非所有的电脑都有浮点运算器 FPU。)

7.8 写一个程序,使其产生图像的高斯金字塔。

7.9 一个采样高斯核必然会有折叠失真,因为核包含任意高频分量。假设核在一个无限网格上采样。当标准方差变小时,折叠失真的能量必然增加。表示折叠失真的能量同高斯核以像素表示的标准差之间的关系。现在假设高斯核用一个 7×7 的网格给出。如果折叠失真能量必须保持与由截断的高斯函数产生的误差幅值相同的量级,那么这个网格上能够表达的最小标准差是多少?

第8章 边缘检测

有很多原因使得图像中亮度的急剧变化为人们所关注。首先,物体边界通常会产生亮度急剧地变化——一个明亮的物体可能会在一个灰暗的背景之上,或者一个灰暗的物体会在一个明亮的背景之上。其次,反射率变化也通常会产生亮度的急剧变化,从而形成与众不同的图案——例如斑马的条纹和美洲豹的斑点。阴影同样也能在亮度上产生急剧的变化。最后,物体表面朝向的急剧变化通常与图像中亮度的急剧变化有关。

图像中亮度明显而急剧变化的点通常称为边缘或边缘点。我们更乐意将边缘点与物体的边缘或者其他一些有意义的变化相对应,但要精确的定义这种要加以标记的变化是很困难的——在一幅花瓣漫天飞舞的田园风景画中,哪一个区域才是一个物体的边界呢?具有代表性的是,很难从一个复杂的边界中确定出一个有明确定义的边界,如果非要这样做的话,则需要许多高层的信息。虽然如此,构建视觉系统的经验告诉我们,令人感兴趣的东西常常出现在图像的边缘处,所以知道图像中边缘的位置是很重要的。

8.1 噪声

边缘检测中一个主要的问题就是图像噪声。这是因为边缘检测器是为了响应急剧的变化而构造的;但是在图像中获得急剧变化的方法就是在像素上添加噪声(因为在每一个像素点上的噪声值通常是无关联的,或者说有可能相差很大)。如同 7.3 节指出的,噪声使得图像导数有限差分估计变得不再可用。我们用这个观察结果作为研究一般图像中噪声的推动力。

噪声这个术语通常表示无法获取信息的图像度量或者与图像中所包含的无关信息的图像度量,其余的则都是信号。认为噪声不包含信息的看法是错误的——例如,我们可以通过在一个黑暗的房间中,使用盖上镜头盖的照相机拍一张照片从而获取到该照相机的温度的估计。此外,由于在没有噪声模型的情况下,无法确定噪声的意义,所以,认为噪声是没有模型的看法是错误的。噪声是我们并不需要使用的任何信息,这就是它的全部。

8.1.1 可加性静态高斯噪声

在可加性静态高斯噪声模型中,每个像素点都相互独立地加上了一个根据同一个高斯概率分布产生的值。几乎所有这种分布的平均值都是零。标准偏差是这种模型的一个参数。这个模型用于描述照相机中的热噪声,这在图 8.1 中举例说明。

线性滤波器对可加性静态高斯噪声的响应 假设有一个核为$\mathcal{G}$的离散线性滤波器,将其用于一个由平均值为μ、标准偏差为σ的可加性静态高斯噪声所构成的噪声图像$\mathcal{N}$。滤波器在点 i,j 的响应为

$$R(\mathcal{N})_{i,j} = \sum_{u,v} G_{i-u,j-v} N_{u,v}$$

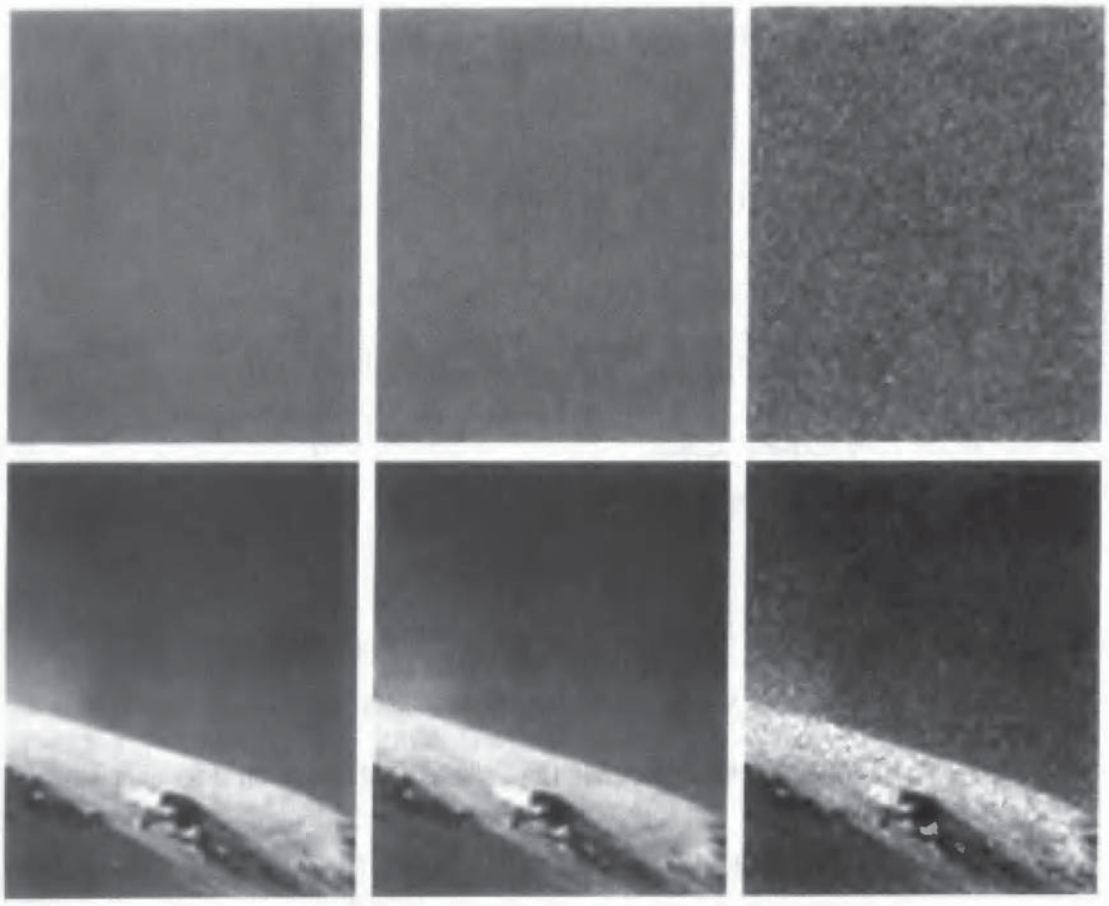


图 8.1　上面一行表示了一个施加可加性静态高斯噪声的三个结果。在图像中加入了相当于亮度值范围的一半以表现正值和负值的噪声。从左到右，噪声的标准偏差分别为完全亮度值范围的1/256，4/256和16/256。对于一个用8比特表示一个像素点的照相机，这分别相当于 0 比特、2比特、5比特的噪声。下面的一行显示将这个噪声加到同一幅图像中的结果。在每个例子中，小于零的值和大于亮度最大值的值相应地被调整为零和最大值

由于噪声是静态的，计算的数学期望值不取决于这一个点，假设 i 和 j 为 0，因而可在表达式中省略。假设核具有有限支集，所有噪声变量的某些子集对所求的数学期望值有贡献；设这些子集为 $n_{0,0},\cdots,n_{r,s}$，这个响应的期望值为

$$\begin{aligned}\mathrm{E}[R(\mathcal{N})] &= \int_{-\infty}^{\infty}\{R(\mathcal{N})\}p(N_{0,0},\cdots,N_{r,s})dN_{0,0}\cdots dN_{r,s}\\ &= \sum_{u,v} G_{-u,-v}\left\{\int_{-\infty}^{\infty} N_{u,v}p(N_{u,v})dN_{u,v}\right\}\end{aligned}$$

对公式做一些改变，并且提取出一些在求和公式的表达式中没有出现的变量。由于所有的 $N_{u,v}$ 都是独立同分布、平均值为 μ 的高斯随机变量，则有

$$\mathrm{E}[R(\mathcal{N})] = \mu\sum_{u,v} G_{i-u,j-v}$$

对噪声响应的方差很容易获得。我们要确定的是

$$\mathrm{E}[\{R(\mathcal{N})_{i,j} - \mathrm{E}[R(\mathcal{N})_{i,j}]\}^2]$$

它与下式相同

$$\int \{R(\mathcal{N})_{i,j} - \mathrm{E}[R(\mathcal{N})_{i,j}]\}^2 p(N_{0,0}, \cdots, N_{r,s}) dN_{0,0} \cdots dN_{r,s}$$

展开为

$$\int \left\{ \sum_{u,v} G_{-u,-v}(N_{u,v} - \mu) \right\}^2 p(N_{0,0}, \cdots, N_{r,s}) dN_{0,0} \cdots dN_{r,s}$$

这个表达式展开为一个两种积分和的形式。公式的项

$$\int G^2_{-u,-v}(N_{u,v} - \mu)^2 p(N_{0,0}, \cdots, N_{r,s}) dN_{0,0} \cdots dN_{r,s}$$

(对于一些 u, v)能够容易地求积,这是由于每个 $N_{u,v}$ 都是独立的;积分为 $\sigma^2 G^2_{-u,-v}$,其中 σ 是噪声的标准偏差。公式的项

$$\int G_{-u,-v} G_{-a,-b}(N_{u,v} - \mu)(N_{a,b} - \mu) p(N_{0,0}, \cdots, N_{r,s}) dN_{0,0} \cdots dN_{r,s}$$

(对于一些 u, v 和 a, b)积分结果为 0,同样是因为每个噪声项都是独立的。则有

$$\mathrm{E}\left[\{R(\mathcal{N})_{i,j} - \mathrm{E}[R(\mathcal{N})_{i,j}]\}^2\right] = \sigma^2 \sum G^2_{u,v}$$

可加性静态高斯噪声模型的难以解决的问题 照字面意义,可加性静态高斯噪声模型是图像噪声的一个简化模型。首先,该模型允许图像有任意大幅度的正的像素值(并且,更惊人的允许负值)。但如果在一些典型环境下(例如室内和白天)合适地选择标准偏差,这并不会导致太大的问题,因为在实际中这些极端的像素值几乎不会出现。在处理含噪声的图像时,这些导致问题的像素点分别被置为 0 或最大值。

其次,噪声值被认为是完全独立的,所以该模型无法获取成组的像素点具有相关影响的情况,这种情况可能是由于照相机电路的设计或是由于照相机集成电路中的过热点造成的。这个问题实际上很难解决,因为考虑到这种影响的模型在分析上很难处理。最后一点,该模型无法很好地描述那些坏死的像素(即始终表示为全黑或全白的像素点)。如果标准偏差过大并且我们限制像素值范围的话,就会出现坏死的像素,但是对于描述图像其余部分而言,标准偏差就有可能过大。可加性高斯噪声的一个重要的优点是很容易估计滤波器的响应。于是,它提供了很好的想法,可以用于确定不同的滤波器对信号响应和忽略噪声响应的有效性。

8.1.2 为什么有限差分会响应噪声

关于线性滤波器对可加性静态高斯噪声响应的讨论,提供了噪声在有限差分情况下的一些表现。假设有一幅包含平均值为 0 的静态高斯噪声的图像,并且考虑一个对导数阶数递增用来估计用的有限差分滤波器的响应的方差。使用核

$$\begin{matrix} 0 & 0 \\ 1 & -1 \\ 0 & 0 \end{matrix}$$

估计一阶导数。则二阶导数是一阶导数的一阶导数,于是核为

$$\begin{matrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{matrix}$$

稍微思考一下就可以确信，在这种情况下，一个 k 阶导数核的系数就是 Pascal 三角的第 $k+1$ 行加上间隔反复的正负号。对于每一个这样的导数滤波器，高斯噪声的响应均值是零，但是响应的方差急剧上升；对于 k 阶导数，其值就是 Pascal 三角的第 $k+1$ 行平方和与标准差的乘积。图 8.2 显示了这个结果。

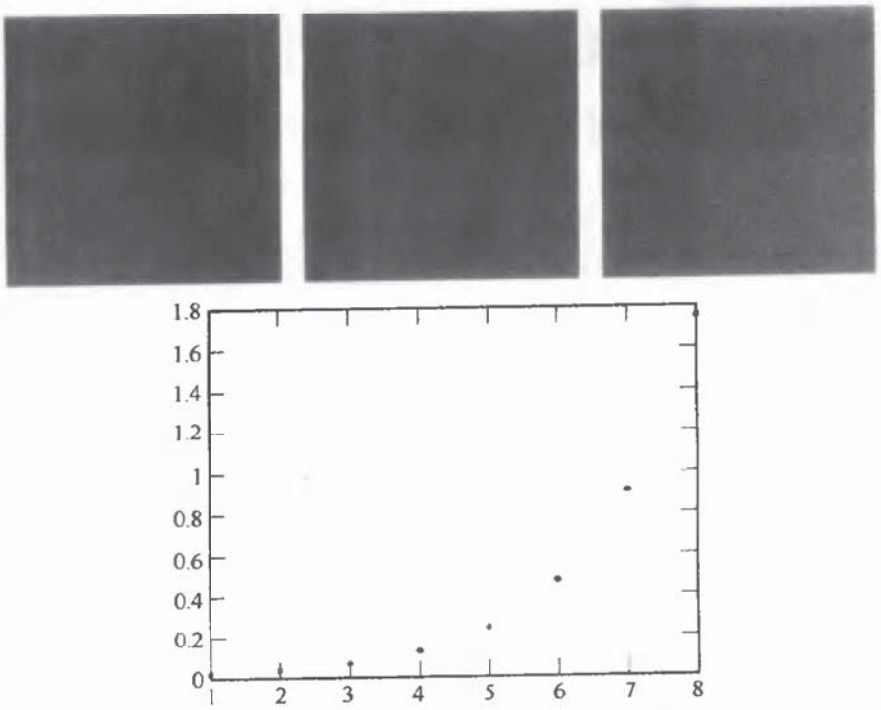


图8.2　按照 8.1.2 节的观点，有限差分能够使可加性高斯噪声充分增强。左上角的一幅图是平均值为0、标准偏差为全值范围的4/256的高斯噪声。上方居中的图显示了在 x 方向上3阶导数的有限差分估计，右上方的图则显示了 x 方向上6阶导数的有限差分估计。每一幅图中都加上了全值范围的一半以显示正值的导数和负值的导数。图像使用了同样的灰度尺度显示；在6阶导数的例子中，有一些值超过了取值范围。下方的图表显示了该噪声图像从 1 阶导数到8阶导数的标准偏差值（参数的选取基于Pascal三角）

存在另一种解释。从表 7.1 得知，对一个函数求微分相当于乘以傅里叶变换的频率变量；这表明高频分量在低频分量减弱的情况下得到了增强。直觉上这也是正确的——对一个函数求微分将会使得常数分量变为零，并且正弦函数的导数的振幅随着频率的增加而增大。此外，这个特性也是我们对导数感兴趣的原因；之所以讨论导数，正是因为快速的变化（即高频部分产生的来源）会有较大的导数。

8.2　导数估计

如同图 7.4 指出的，简单的有限差分滤波器能够对噪声产生强烈的响应，于是使用两个有限差分滤波器（每个方向一个）将是一个求梯度的不好的方法。处理这个问题的方法就是对图像平滑后求导（同样也会平滑导数）。在实际中，图像几乎总是被高斯滤波器平滑过的——事实上，有限差分操作是平滑过的，如图 8.3 所示。按照惯例，我们首先讨论这个问题，然后对于希望了解更多的读者，我们将讨论为什么平滑是有用的，以及为什么高斯滤波器是一个平滑滤波器的很好选择。

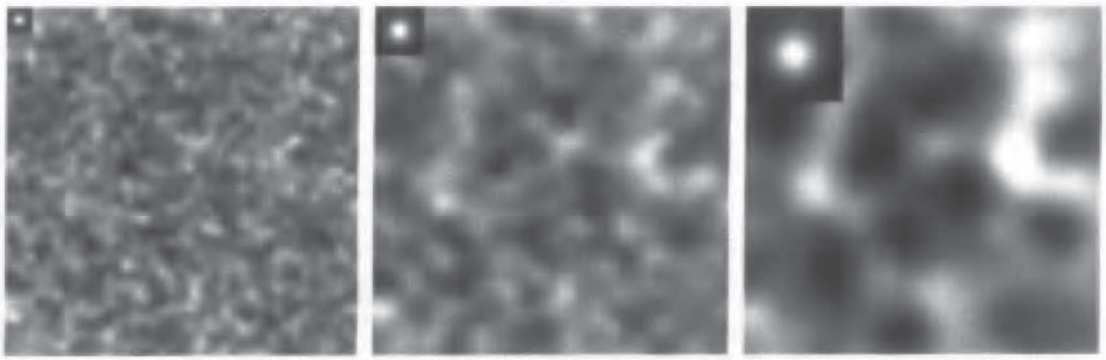


图 8.3 图中显示了对静态可加高斯噪声进行平滑，会导致信号中相邻像素值趋向相似。结果同滤波器核的尺度有关，因为滤波器的核导致相关性。图像显示，高斯滤波器的核越大，平滑效果越明显。灰色像素为零值，较深像素为负值，较浅像素为正值。图像的左上角为滤波器的核，可以看到核的空间尺度（已经定义了高斯核大小同亮度的关系，所以中心像素为白色，边缘像素为黑色）。如图像所显示的，平滑后的噪声看起来接近自然纹理

8.2.1 高斯滤波器的导数

平滑一个图像，然后求其导数相当于将其与一个平滑的核函数的导数进行卷积。只要想想连续卷积就很容易得出这个事实。

首先，微分是一个线性移不变的，这表明存在一些核——这里回避它是什么样子的问题——可以求微分。现在，给一个函数 $I(x,y)$

$$\frac{\partial I}{\partial x} = K_{(\partial/\partial x)} ** I$$

现在，希望得到一个平滑函数的导数。把平滑卷积核写为 S。由于卷积满足结合运算，则有

$$(K_{(\partial/\partial x)} ** (S ** I)) = (K_{(\partial/\partial x)} ** S) ** I = \left(\frac{\partial S}{\partial x}\right) ** I$$

这种情况在平滑函数是高斯函数时是最常见的，可以写为

$$\frac{\partial (G_\sigma ** I)}{\partial x} = \left(\frac{\partial G_\sigma}{\partial x}\right) ** I$$

于是，只需与高斯函数的导数求卷积，而不必先卷积然后求微分。平滑导致导数估计有更小的响应噪声（见图 8.4）。

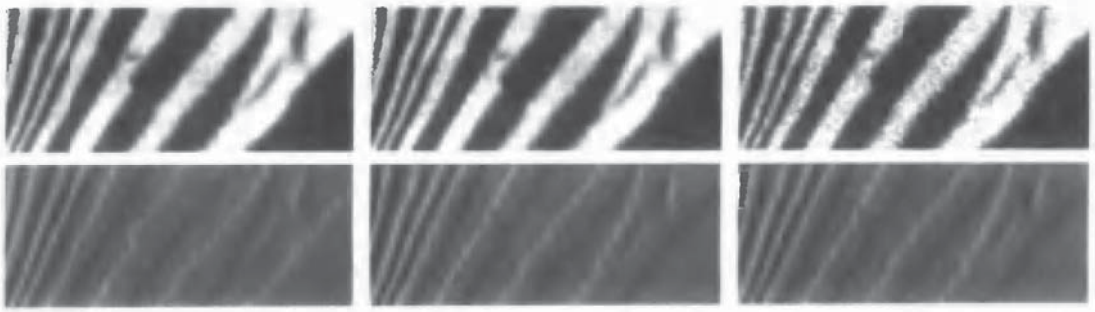


图 8.4 与有限差分滤波器相比，高斯滤波器的导数对噪声的响应更小。左上的图像是一幅斑马图片的局部细节；上方居中的是同样一幅图片掺杂了平均值为零、$\sigma = 0.03$（像素值范围从0到1）的可加性静态高斯噪声的结果。右上的是同样一幅图片掺杂了平均值为0、$\sigma = 0.09$的可加性静态高斯噪声的结果。第二行显示了每幅图片在x轴方向的偏导数，在每个例子中都用标准差对一个像素σ的高斯滤波器的导数进行估计。注意平滑是如何有助于减少噪声的影响

8.2.2 为什么平滑是有用的

一般说来，任何对我们有意义的变化都将影响一堆像素点。例如，一个物体的轮廓将会导致在图像上一长条链状的像素点有较大的导数值。对于很多噪声模型而言，一些来源于噪声的大的导数值主要是局部事件。这表明平滑一个差分过的图像更倾向于支持那些感兴趣的变化，并且抑制噪声的影响。

对为什么平滑是有用的，可有另外一种解释。假设先平滑一个噪声图像，然后进行微分。首先，噪声的方差会由于平滑核函数而减小，这是因为我们倾向于使用正的平滑核函数，且有

$$\sum_{uv} G_{uv} = 1$$

也就是说

$$\sum_{uv} G_{uv}^2 \leqslant 1$$

其次，像素有变得与相邻的像素相似的倾向——如果使用可加性静态高斯噪声来平滑，结果信号的像素值将不再是独立的。在某种意义上，这就是平滑的意义所在——平滑是根据其邻近像素值来推断一个像素值的方法。无论如何，如果像素倾向于与邻近的像素相同的话，那么其导数必然也将变小(因为导数表示的就是一个像素与其相邻像素值的差异的大小)。

另一个途径是根据空间频率来解释。可以知道，可加性静态高斯噪声在各个频率有着相同分布的能量。如果将噪声进行微分，就加重了高频。如果不试图去修正这个情况，则在梯度图中将会由于噪声的影响而产生偶然出现的过大值。使用一个高斯滤波器将抑制这些高频，就好像它对重采样所起的作用一样(见 7.4.3 节)。

平滑噪声有其实用性。如同图 8.3 指出的，平滑过的噪声倾向于同一些自然的纹理相类似，在计算机图形学应用中，平滑过的噪声作为一种纹理的来源有着广泛的应用(Ebert, Musgrave, Peachey, Worley 和 Perlin, 1998; Perlin, 1985)。

8.2.3 选择平滑滤波器

可以通过一个边缘模型和一些判定标准来选择一个对该模型具有最佳响应的方法选择平滑滤波器。但是很难将这个问题作为一个二维问题，因为二维中的边缘能够弯曲。按照惯例，将通过表示成一个一维问题来选择平滑滤波器，然后在二维情况下使用其旋转对称的版本。

一维滤波器需要通过边缘模型获取。通常使用的模型是一个带可加性静态高斯噪声的未知高度的阶跃函数，如下给出

$$edge(x) = AU(x) + n(x)$$

其中，

$$U(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}$$

[$U(0)$的值与我们的目的并不相关]，A 通常称为边缘的对比度。在一维情况下，求梯度幅值等同于求导数响应的平方。因此，通常需要找到一个导数估计滤波器而不是一个平滑滤波器(它实际上可以通过导数估计滤波器重建)。

Canny(1986)建立了通过使用连续模型来选择导数估计滤波器的准则,可以优化三个标准的组合:

- **信噪比**——滤波器应该对在 $x=0$ 处的边缘,比对噪声有更强的响应。
- **定位精度**——滤波器应该在接近 $x=0$ 处达到最大值。
- **低误报率**——在 $x=0$ 处合理的邻域内,应该只有一个最大响应。

一旦找到一个连续的滤波器,它即是离散化的。上述标准可以通过多种的方式组合而得出一些不同的滤波器。有一个值得注意的事实是,来源于大多数标准组合下的最优平滑滤波器在很大程度上像是高斯函数——直觉上这是成立的,这是因为平滑滤波器需要在靠近中间的像素上有较大的权重,而在远离的像素上有较小的权重,有点像高斯函数。在实际应用中,最优的平滑滤波器通常被高斯函数取代,而在性能上并没有什么特别重要的退化。

用于估计导数的 σ 的选择常常称为平滑的尺度。尺度对导数滤波器的响应有实质性的影响。假设在一个恒定的背景上有一道狭窄的条纹,就好像斑马的毛发。如果使用小于条纹宽度的尺度来平滑的话,滤波器将会在条纹的两边都有响应,于是能够分辨出条纹的上升沿和下降沿。如果滤波器宽度过大的话,条纹将被平滑到背景中,只产生很小的响应或完全没有响应(见图 8.5)。

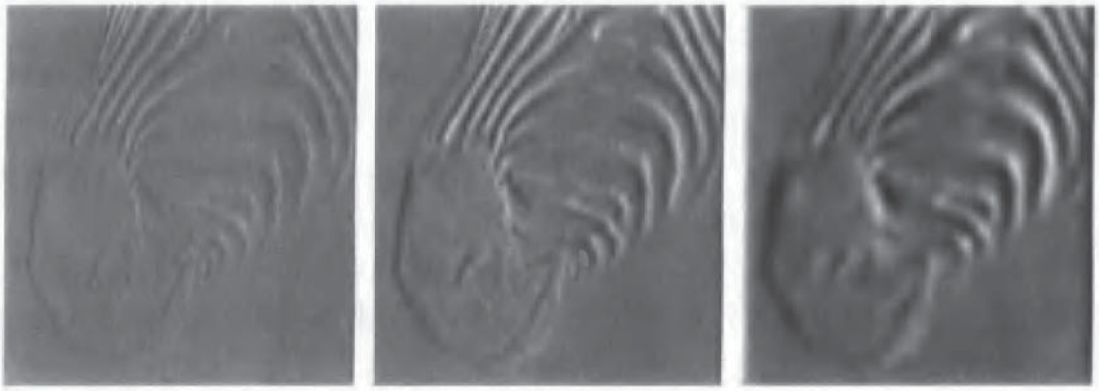


图 8.5 在高斯导数滤波器中,高斯函数的尺度(也就是 σ)对结果有显著的影响。三幅图显示了一幅斑马头的图像在x轴方向的导数,分别使用了1个像素、3个像素、7个像素的σ高斯导数滤波器(从左向右)。注意在一个较细尺度下图像是如何显示出毛发的,动物的毛发在中等尺度下未能显现出来,同时斑马口上方的细密斑纹在较粗尺度下未能显现出来

8.2.4 为什么使用高斯平滑

尽管高斯函数不是惟一可能的光滑核,但因为有一些重要的特性,所以它是实用的。首先,如果我们将两个高斯函数做卷积,得到的结果将是另一个高斯函数:

$$G_{\sigma_1} ** G_{\sigma_2} = G_{\sqrt{\sigma_1^2+\sigma_2^2}}$$

这表明,通过对一个已经平滑过的图像进行再平滑从而得到深度平滑图像是可能的。这是一个重要的特性,因为离散卷积是一个代价较高的操作(在滤波器的核很大时尤其显著),并且通常希望获得一幅图像的平滑程度不同的若干版本。

效率 考虑使用一个 σ 是一个像素的高斯核函数对一个图像做卷积。尽管高斯核在无

限域上是非零的，但由于指数的形式，在大多数区域中它是很小的。对于 σ 是一个像素的情形，在以原点为中心的 5×5 矩阵之外的点的值小于 $e^{-4}=0.018\ 4$，而在以原点为中心的 7×7 矩阵之外的点的值小于 $e^{-9}=0.000\ 123\ 4$。这表明可以忽略它们的贡献，而将离散高斯表示为一个小数组（5×5 或者 7×7，这取决于个人的爱好和分配用于表示核的位数）。

但是，如果 σ 是 10 个像素，我们则需要一个 50×50 的数组或者情形更糟。除非你认为卷积一个有着 50×50 数组的合理大小的图像没有太大问题，否则另一种选择——重复同一个小得多的核函数卷积——将会更有效率，因为我们不需要保留每一个像素的中间结果。这是因为平滑图像在一定的范围内是冗余的（大多数像素点包含它们邻接值的较显著的分量），其结果就是可以将一些像素点丢弃。于是我们有了一个十分有效的对策：平滑、间隔采样、平滑、间隔采样……其结果是一幅图像，它将会与一个高度平滑的图像拥有同样的信息，但尺寸更小、更易得到。在 7.7.1 节中，我们已经研究了这个方法的细节。

中心极限定理　图 8.6 中阐明了高斯核函数的另一个显著特性。对一个重要的函数族来说，将函数族中任意一个成员不断与自己卷积，最终将生成一个高斯函数。这意味着，如果我们选择一个不同的平滑核函数，并且重复地施加在图像之上，通过卷积的结合律性质，则最终的结果就像我们使用了高斯函数来平滑该图像一样。

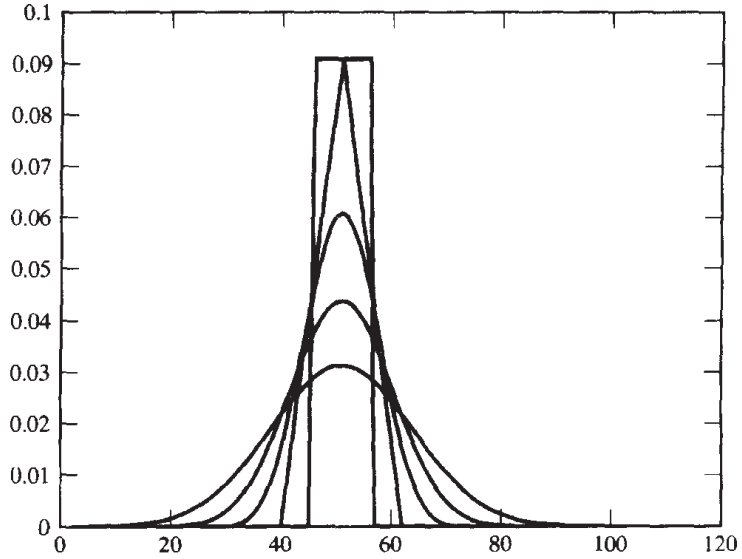


图 8.6　中心极限定理声称，如果对一个正的核函数不断卷积，最终以一个高斯函数的缩放比例为极限。图中显示了一维情况下的这个效果，图中的那个三角形是通过将一个窗函数与其自身卷积得到的，后面每一个结果都是将前一个结果与其自身卷积得到的

高斯函数是可分离的　最后一点，一个各向同性的高斯函数可以表示为

$$G_\sigma(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)$$
$$=\left(\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x^2)}{2\sigma^2}\right)\right)\times\left(\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y^2)}{2\sigma^2}\right)\right)$$

这是两个一维高斯函数的乘积。一般来说，一个分解为 $f(x,y)=g(x)h(y)$ 的函数 $f(x,y)$ 对应于一个张量积。通常将一个对应张量积的滤波器核函数称为可分离核。可分离性的确是一个有用的特性。特别地，使用一个可分离的滤波器核函数相当于同两个一维核函数卷积——

一个在 x 轴方向,另一个在 y 轴方向(见习题)。

有许多核函数是可分离的。可分离的滤波器核函数导致分解因式的离散表达。特别地,当$\mathcal{H}$是一个离散的、可分离的滤波器核函数时,有一些向量 $\boldsymbol{f}$ 和 $\boldsymbol{g}$

$$H_{ij} = f_i g_j$$

有可能使用数值线性代数的技术来确定这一特性,因为矩阵$\mathcal{H}$的秩必须是1。商业化的卷积算法包常常在将核函数应用于图像之前先测试它是否是可分离的。如果这个核函数被证明是可分离的,那么这一测试的代价将会很容易补偿。很多核函数能够被一组可分离的核函数的和近似表达。如果这些核函数的数量足够少,那么进行卷积时,这样的近似将带来使用上的节约。如果希望对一个图像使用很多不同的滤波器做卷积时,这将是一个很吸引人的策略;在这种情况下,有人试图对这些滤波器中每一个,获取用多个可分离的和函数的加权形式表示的表达式,即少量的基本元素的张量积。于是将能够使用一些基本元素对图像做卷积,从而通过不同加权和的结果得到使用不同滤波器下的图像。

间隔采样高斯函数中的折叠失真 关于折叠失真的讨论可以为我们提供一些关于可能的平滑参数的见解。我们使用的任何高斯核函数都是在一个单像素间隔网格采样的高斯函数的近似值。这表明,对于通过采样近似值重建的原始核函数,它将不会包含空间频率大于0.5像素$^{-1}$的分量。对于高斯函数这是不可能的,因为高斯函数的傅里叶变换仍是高斯函数,因此并非是有限带宽的。我们所能做的最好的事就是坚持认定折叠的信号的能量大小在某个阈值之下——这意味着在一个离散网格上的平滑滤波器的 σ 可取的最小值(如果取值小于这个最小值,则平滑滤波器将会严重的折叠,见习题)。

8.3 对边缘进行检测

边缘检测中的两个主要策略都将亮度上的快速变化作为边缘的模型。首先,我们观测到最快的变化发生在二维中二阶导数为零的地方(见8.3.1节)。尽管这个方法曾经很重要,但已不再流行了,可供选择的是明确寻找梯度幅值达到极值的像素点(见8.3.2节)。

8.3.1 使用拉普拉斯算子检测边缘

在一维情况下,一个信号的导数值为极值时,它的二阶导数为零。这表明,如果我们希望找到大的亮度变化,一个需要寻找的地方就是二阶导数为零的地方。这个方法可以扩展到二维空间中。这需要一个对应于二阶导数的量,需要旋转不变量。很容易证明拉普拉斯算子具有这个特性。在二维情况下,一个函数的拉普拉斯算子定义为

$$(\nabla^2 f)(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

在使用拉普拉斯算子之前先平滑图像是很自然的过程。注意拉普拉斯算子是一个线性操作(如果不是很确定的话,应该证明一下),表明我们可以通过使用一些核函数卷积图像来达到施加拉普拉斯算子的作用。(其中,核函数写做 K_{∇^2})。由于卷积符合结合律,于是

$$(K_{\nabla^2} ** (G_\sigma ** I)) = (K_{\nabla^2} ** G_\sigma) ** I = (\nabla^2 G_\sigma) ** I$$

与一阶导数的情况类似,这之所以很重要,是因为平滑一个图像然后施加拉普拉斯算子,等同于使用平滑核函数的拉普拉斯算子对图像做卷积。图8.7显示了所得到的核函数。

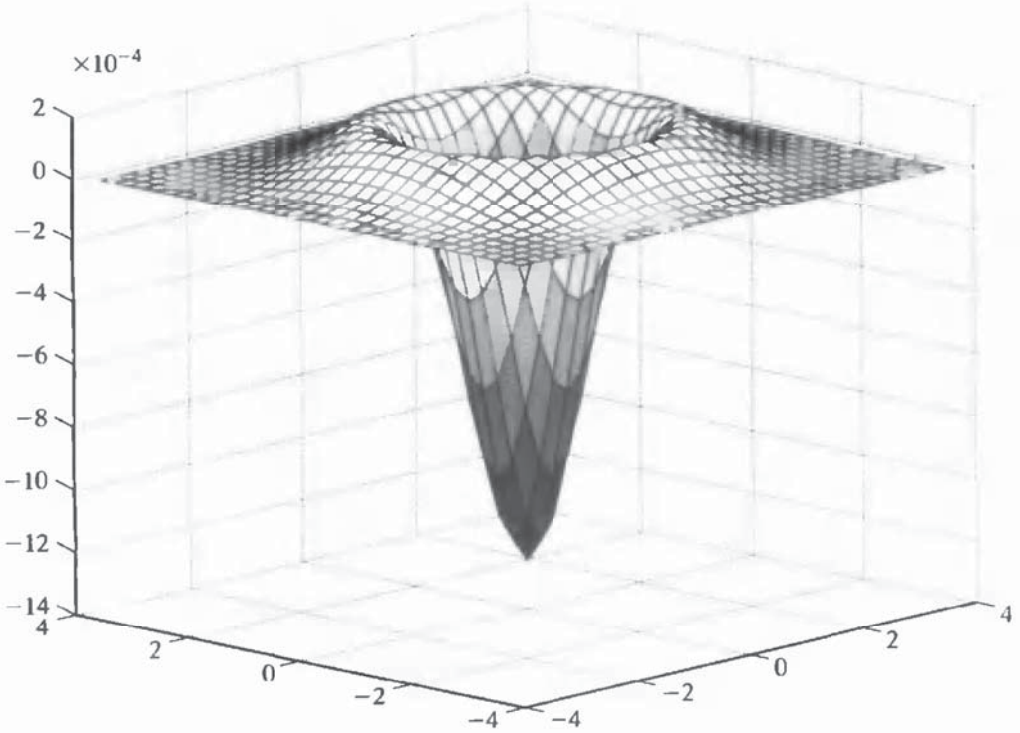


图8.7 如图显示的 σ 为一像素的高斯滤波器核函数的拉普拉斯算子，可以认为是从一个周边元素加权平均值中减去中央像素(因此与文中描述的模糊掩模相似)。通常将这个核函数替换为两个高斯函数的差——一个有较小的 σ，另一个有较大的 σ

如图8.8显示的，这导致了一个简单但在历史上很重要的边缘检测策略。我们使用某个尺度的拉普拉斯高斯算子对图像卷积，并且标记结果为零的点——过零点。应该检验这些点以确保它们的梯度值很大，该方法由 Marr 和 Hildreth 提出(1980)。

高斯滤波器的拉普拉斯算子在边缘的一侧的响应是正的，而在另一边则是负的。这表明占这个响应一定百分比的部分加回到原图上，将产生一幅边缘更加明显、细节更容易看清的图片。这个观测结果可以回溯到一个称为模糊遮掩的图像显像技术，即通过减去在该区域亮度的局部平均值，从而在明亮的区域增加细节的可视度的一个模糊的正片。这差不多相当于使用一个高斯差分滤波器过滤一幅图像，将结果乘以一个小的常数，然后将它加回到原图像上。现在，两个高斯滤波器核函数的差看起来就像是一个高斯核函数的拉普拉斯算子，并且常常使用其中一个替换另一个。这表明模糊遮掩加强了一个图像的边缘。

对于高斯边缘检测器的拉普拉斯算子已经提出了一些不同的意见。由于高斯滤波器的拉普拉斯算子不是定向的，所以它的响应由一个穿越边缘的平均和一个沿着边缘的平均组成。这表明在拐角处——边缘方向改变的地方——其表现是很差的，这将会很不准确地标记拐角处的边缘。此外，在三面点或更多面点处，正确的记录拐角的布局将会是很困难的，如图8.9所示。其次，沿着边缘的分量将会使滤波器对噪声的响应有贡献，而不一定是对边缘；这表明过零点有可能并不确切地出现在边缘上。

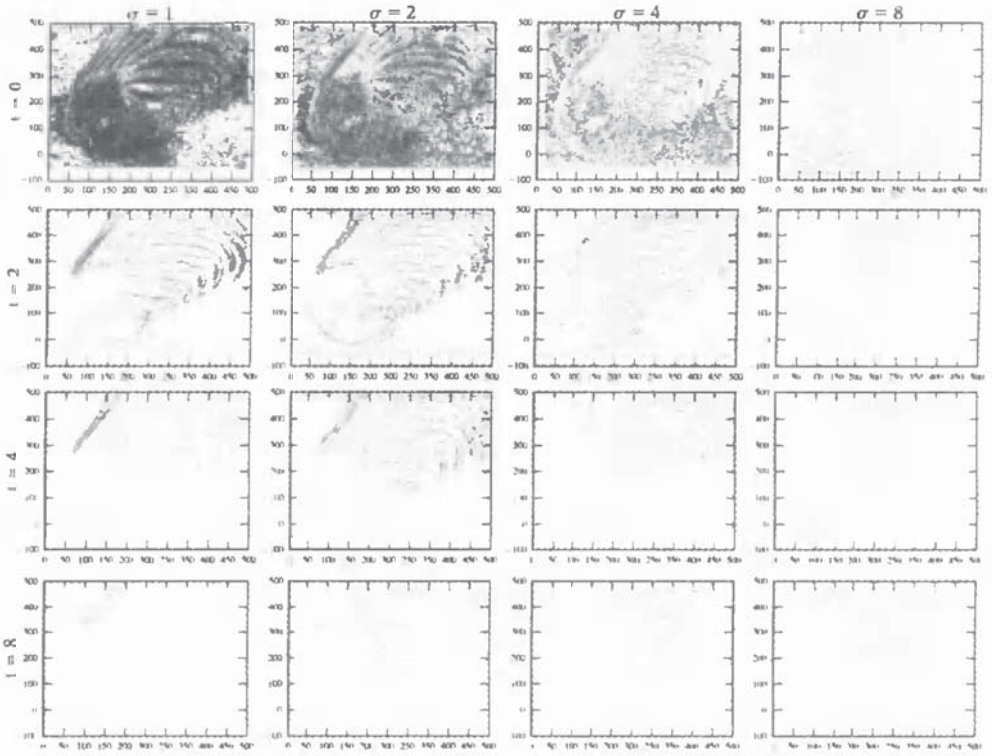


图 8.8 不同尺度下和不同梯度幅值阈值下的高斯函数的拉普拉斯算子的过零点。每一列表示一个固定的尺度，而梯度阈值t从上往下增加(每相邻两幅图之间以两倍为单位增长)。每一行表示一个固定的 t，尺度以两倍为单位从一个像素的σ增加到8个像素的σ。注意在细尺度、小阈值的情况下，边缘包含了大量有用或无用的细节信息(取决于对斑马鼻子上毛发的兴趣)。如果尺度增加的话，细节就被抑制；阈值增加的话，小区域的边缘就被丢弃。没有尺度或阈值能给出斑马头部的轮廓；所有的响应都是针对斑纹的，尽管随着尺度的增加，斑马鼻子上面细小的斑纹不再可分辨

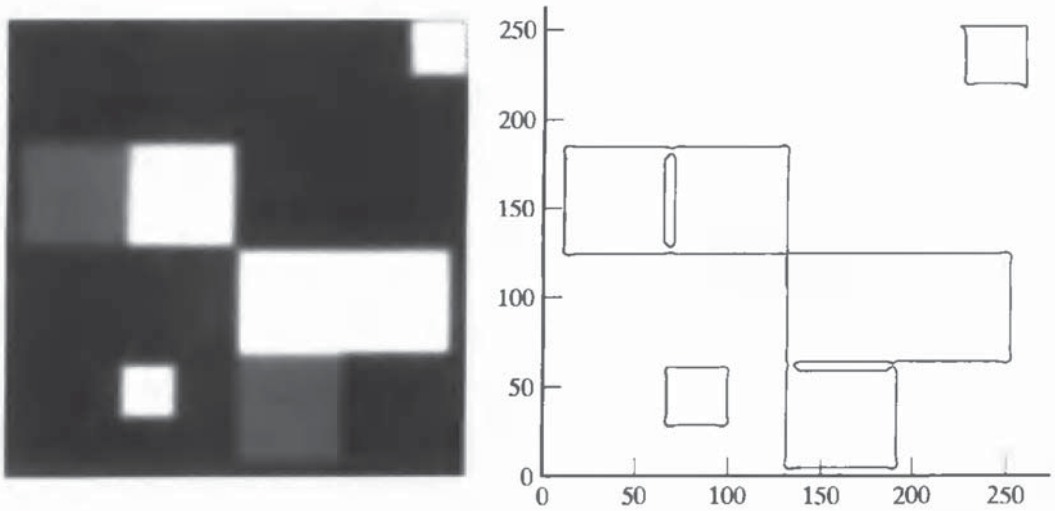


图 8.9 高斯函数的拉普拉斯算子的过零点输出在拐角处会表现得很奇特。首先，在一个直角处，过零点凸出于拐角(但是通过顶点)。这种效果并不是由于数字化或量化，并能证明在连续情况下出现。在三个或更多个边缘交界的地方，轮廓表现得很奇怪，具体的细节则取决于轮廓结构的标记算法——这个算法(即MATLAB中使用的算法)生成奇怪的环。这种现象能够通过仔细的设计轮廓标记算法来加以缓解，该算法需要结合相当细节化的顶点模型

8.3.2 基于梯度的边缘检测器

在基于梯度的边缘检测器中，我们计算梯度值的近似值——几乎总是使用高斯函数作为平滑滤波器——并且使用这个近似值来决定边缘点的位置。具有代表性的是，在图像中沿着一条踪迹的梯度值将很大(见图 8.10)。然而，由于目标的边缘是弯曲的，我们希望获取这条踪迹上最与众不同的点构成的曲线。

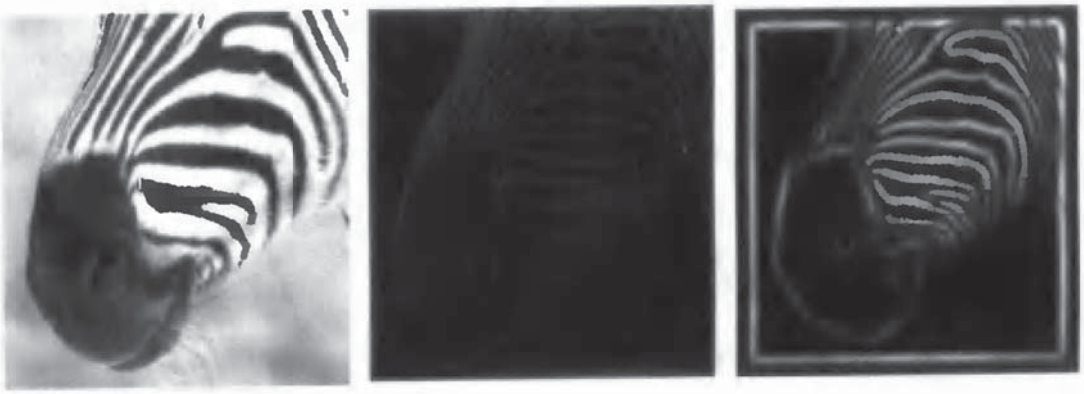


图 8.10 梯度幅值可以通过平滑一个图像并对它进行微分来近似获得，这等同于使用一个平滑滤波器核函数的导数来做卷积。平滑的程度将影响梯度幅值；在本图中，我们展示了在不同尺度下同一幅斑马图片的梯度幅值。中间的一幅使用$\sigma = 1$像素的高斯函数的导数来近似梯度值，右边的一幅则使用$\sigma = 2$像素的高斯函数的导数来近似梯度值。注意大的梯度值形成明显的踪迹

一个很自然的方法就是在寻找垂直于边缘方向上梯度值最大的点。在这种方法中，垂直于边缘的方向也已使用梯度的方向来估计(见图 8.11)。这些考虑构成了算法 8.1。大多数现在的边缘检测器遵循这个算法，但是关于具体细节的正确执行仍存在实质性的争论。

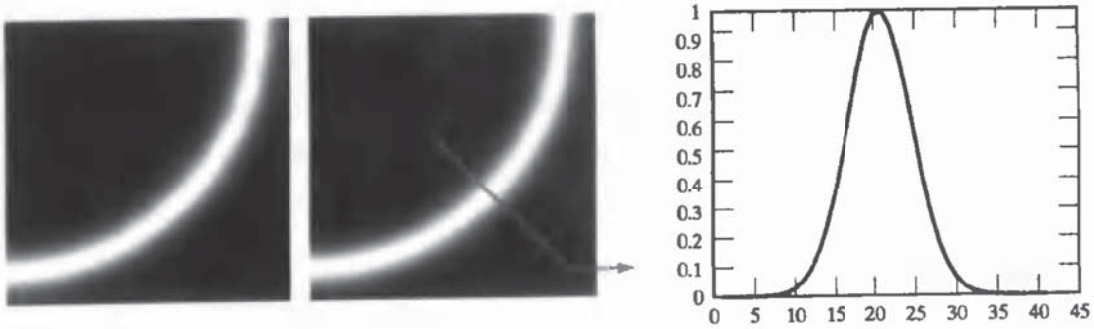


图 8.11 在图像中，大的梯度值将会形成粗的踪迹。典型做法是，我们希望将这些踪迹简化为具有代表性的边缘点组成的曲线。一个很自然的方法就是沿踪迹的垂线方向寻找最大值，我们使用梯度方向作为沿踪迹的垂线方向的估计。左边的图显示了一条有较大梯度值的踪迹；中间的图显示了一个适当的截取方向，右边的图显示了该方向的峰值

非最大值抑制 通过给出梯度幅值的估计，我们可以获取边缘点。这里我们再一次未经客观的定义，只是依靠合理的直觉来说明。那些梯度幅值能够被认为是一系列的低山丘，标记局部极值将会标记出孤立点——与小山顶相仿。一个更好的评判标准是，沿着梯度的方向即应该垂直于边缘的方向截取，标记该方向的梯度值最大的点。这将会得到一系列的沿着山脊的点，这个过程称为非最大值抑制(见图 8.12)。

算法 8.1 基于梯度的边缘检测
形成一个图像梯度的估计
从这个估计中获取梯度值
找出如下的点
　梯度值最大
　在垂直于边缘的方向上也很大
　这些点就是边缘点

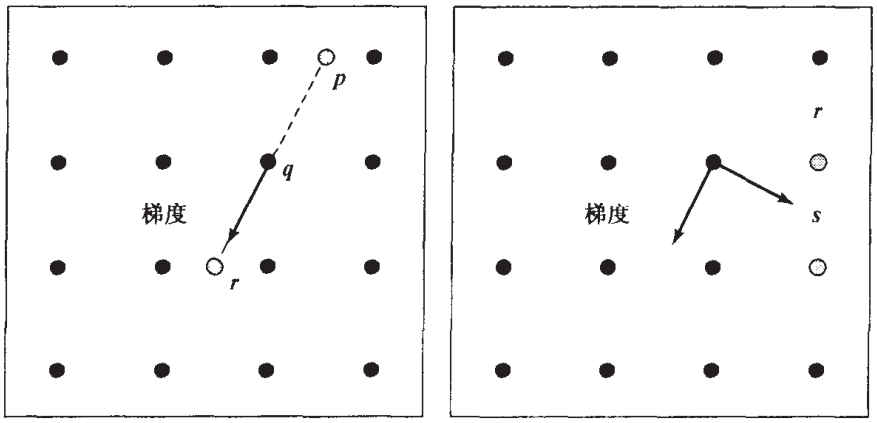


图 8.12 非最大值抑制将在梯度方向上获取梯度幅值的最大点。左边的图显示了我们是如何重建梯度值的，圆点表示像素栅格。在像素点 q 处，我们试图确定其是否为梯度值最大的地方；通过 q 的梯度方向并不通过任何附近的像素点，不论向前或向后，于是我们需要进行插值以获取在 p 和 r 处的梯度值的大小；如果 q 处的值大于其余的值，则 q 就是一个边缘点。典型的是使用线性插值来重建值的大小，在这个例子中应该分别使用 p 和 r 的左边和右边的像素值来插值计算这些点值的大小。在右边，我们描述了在给定 q 是边缘点的情况下如何寻找下一个边缘点的方法；一个适当的搜索方向是垂直于梯度的方向，于是点 s 和点 t 应该被考虑为下一个边缘点。注意，原则上，我们不需要把自己限定在图像中的像素点上，因为我们知道 s 和 t 之间的预计的位置。因此，我们又需要进行插值来获取不在栅格上的点的梯度值

边缘跟踪 典型地，我们希望边缘点存在于链状的曲线上。下面是在非最大值抑制中的一些重要的步骤：

- 决定一个给定的点是否一个边缘点
- 如果是的话，则寻找下一个边缘点

一旦这些步骤被理解的话，就很容易列举所有的边缘链。我们找到第一个边缘点，将它标记，最大限度地扩展所有通过这个点的边缘链，标记所有这些边缘链上的点，并且继续下去以寻找所有未标记的边缘点。

两个主要步骤都是简单的。暂且假设边缘将标记在像素点的位置（而不是在一些更精细的像素栅格处）。我们可以通过将任何点的值与其在梯度方向上前面或后面的一些点比较，来判断该梯度值是否是最大值（见图 8.11）。这是一个梯度距离的函数，典型的是使用下一行（或下一列）像素和前一行来判断当前像素点的梯度值是否更大（见图 8.12）。梯度方向并不总是通过下一个像素点，所以需要进行插值来决定感兴趣的点的梯度值，常用线性插值。

算法 8.2 非最大值抑制

```
while 存在较大梯度的点还没有被访问过时
    寻找一个垂直于梯度的方向上有局部最大值的点作为起始点
      消除已经访问过的点
    while 如果可能的话,通过如下的方法扩展一个通过当前点的链
      1)使用梯度垂直方向预测下一个点的集合
      2)在该梯度方向上寻找一个最大值
      3)检查在最大值处的梯度值是否足够大
      4)去除已访问的点及其相邻的点
    记录下一个点,即成为当前点
    end
end
```

如果像素点被证明是一个边缘点,那么曲线上的下一个边缘点将被预计为沿梯度垂线的方向走一小步。通常,这一步并不中止在像素上,一个很自然的对策就是查看该方向附近的像素点(见图 8.12)。这个方法将产生一个曲线的集合,并且能够通过在一个白色的背景上用相应的黑色曲线表现出来,就如同图 8.13、图 8.14 和图 8.15 中显示的一样。

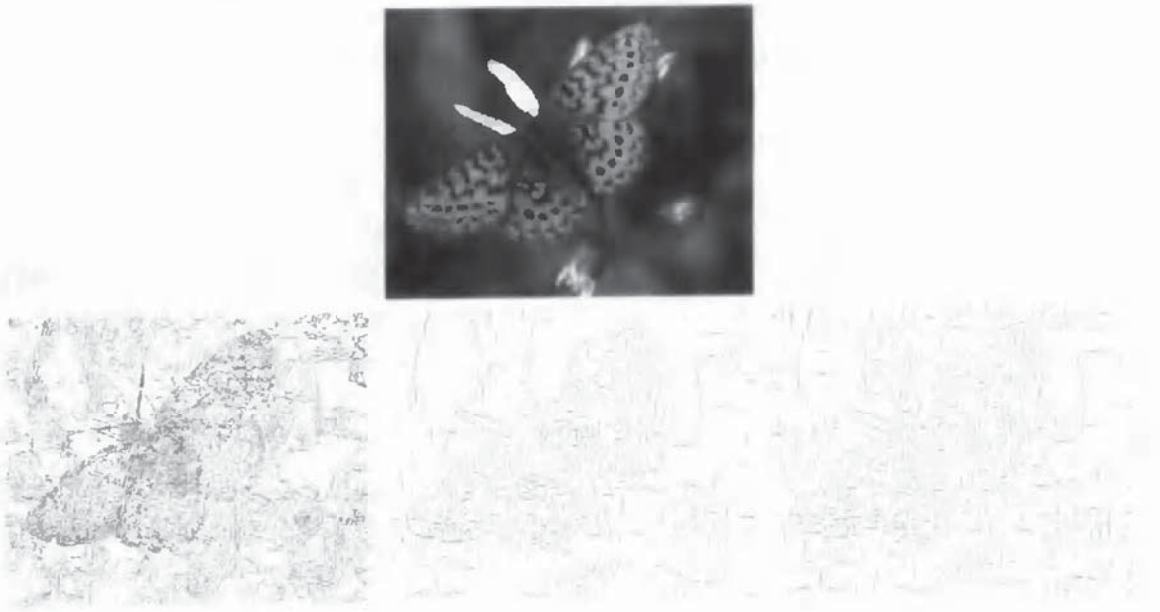


图 8.13 上图中显现的是标记在图像像素栅格上的边缘点。左边图中显示的边缘点是通过使用 σ 为1像素的高斯平滑滤波器获取的,并且通过一个较高的阈值检测这些点的梯度值来判断该点是否一个边缘点。中间图中显示的边缘点是通过使用σ为4像素的高斯平滑滤波器获取的,并且通过一个较高的阈值检测这些点的梯度值来判断该点是否一个边缘点。右边图中显示的边缘点是通过使用σ为4像素的高斯平滑滤波器获取的,并且通过一个较低的阈值检测这些点的梯度值来判断该点是否一个边缘点。在一个细微的尺度上,高对比度下微小的细节产生了出现在粗糙尺度上的边缘点。当阈值较高时,边缘点的曲线常常是断裂的,因为梯度值小于阈值;当阈值较低时,则引入了一些具有不怎么重要的边缘点

图 8.14 图示中，上面的图按照相应的像素栅格将边缘点标出。左下图的边缘点是用 σ 为一个像素的高斯平滑器对图像进行高斯平滑得到的，并且用一个很高的阈值去检测梯度值的大小以确定某一个点是否边缘点。中下图上的边缘点是用 σ 为4个像素的高斯平滑器对图像进行高斯平滑得到的，并且用一个很高的阈值去检测梯度值的大小以确定某一个点是否边缘点。右下图上的边缘点是用 σ 为4个像素的高斯平滑器对图像进行高斯平滑得到的，并且用一个很低的阈值去检测梯度值的大小以确定某一个点是否边缘点。用合适的尺度，较好的细节和高对比度可以产生的边缘点，在粗糙的尺度上会消失。当阈值取的很高时，边缘点的曲线经常遭到破坏，因为梯度值往往小于所设的阈值；而对于很低的阈值，会增加各种各样意义不确定的新边缘点

迟滞 能够成为物体边界的合理表示的曲线实在太多，这在一定程度上是由于我们仅仅标记了梯度值的最大值而没有考虑这些最大值有多大。更常见的是使用一个阈值来确保所有的最大值要大于某个下界，这将导致断裂的边缘曲线(仔细观察图 8.13 和图 8.15)。处理这种情况的常用窍门是使用迟滞；我们设置两个阈值并且在边缘链开始的时候使用较大的阈值，而在后面的时候使用较小的阈值。这个窍门常常在边缘输出中获得改进的结果(见习题)。

8.3.3 技术：方向性表示和角点

众所周知，边缘检测器在角点处会发生失败，这是由于在 x 方向和 y 方向的偏导数足以估计一个带方向的梯度的假设已不再成立。在尖锐的角点处，或者更糟糕的，在基于方向的角点处，它们偏导数的估计很糟糕，因为它们必须穿过角点。存在着各式各样特殊的角点检测器，它们寻找图像中与其相邻处梯度相差巨大的部分(见图 8.16)。更一般地，在一个图像相邻区域的梯度的统计对于该相邻区域是非常有用的。下面将不同类型的图像窗口粗略的分为 4 类：

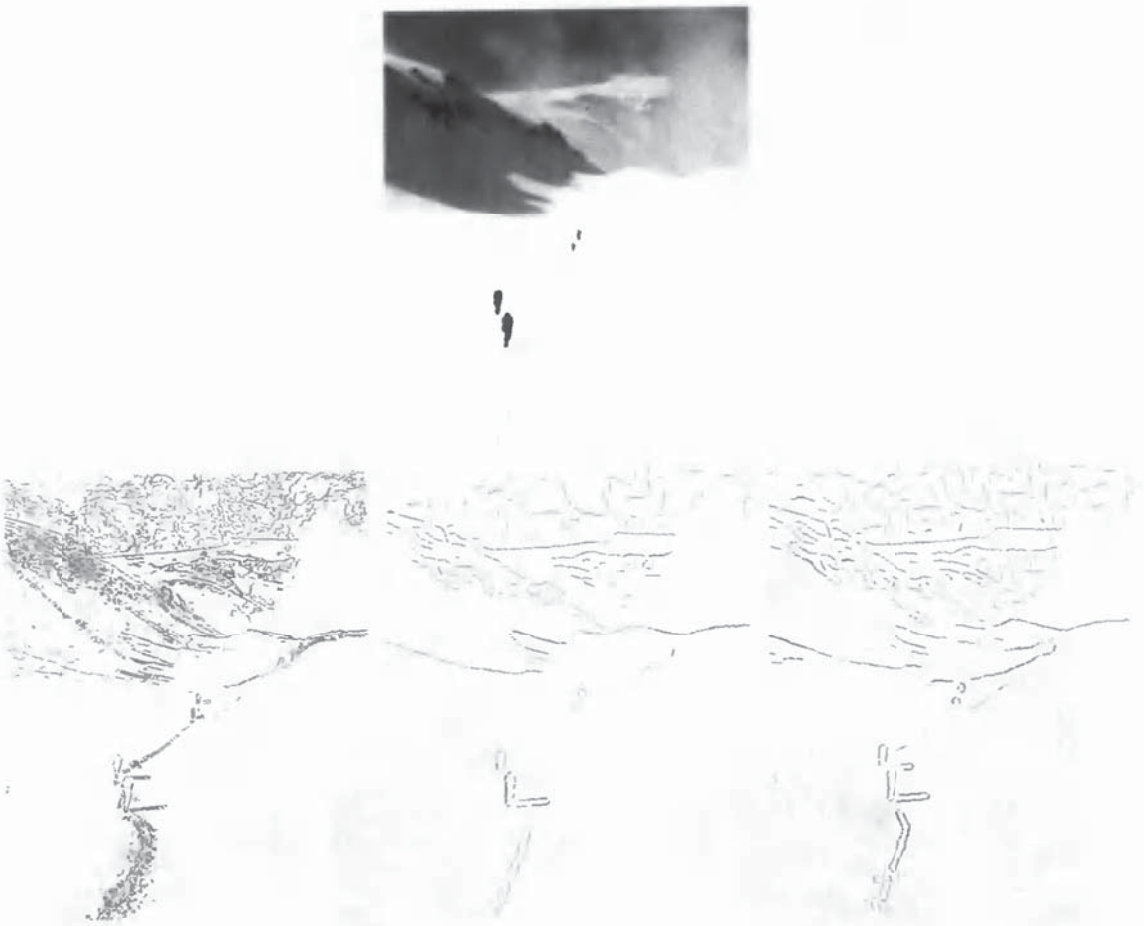


图 8.15 图示中，上面的图按照相应的像素栅格将边缘点标出。左下图的边缘点是用 σ 为 1 个像素的高斯平滑器对图像进行高斯平滑得到的，并且用一个很高的阈值去检测梯度值的大小以确定某一个点是否边缘点。中下图上的边缘点是用 σ 为4个像素的高斯平滑器对图像进行高斯平滑得到的，并且用一个很高的阈值去检测梯度值的大小以确定某一个点是否边缘点。右下图上的边缘点是用 σ 为4个像素的高斯平滑器对图像进行高斯平滑得到的，并且用一个很低的阈值去检测梯度值的大小以确定某一个点是否边缘点。用细的尺度、高对比度下的细节产生的边缘点，在粗糙的尺度上会消失。当阈值取的很高时，边缘点的曲线经常遭到破坏，因为梯度值往往小于所设的阈值；而对于很低的阈值，会增加各种各样意义不确定的新边缘点

- 静态窗口，它的灰度水平近似为常量；
- 边缘窗口，在这种窗口中，沿某单个方向，图像亮度有急剧的变化；
- 流窗口，在这种窗口中，有一些良好的平行条纹——说明是毛发还是皮肤；
- 以及二维窗口，在这种窗口中，有一些二维纹理的形态——说明是点还是拐角。

这些情形对应着图像梯度各种不同的表现。在静态窗口中，梯度向量较短；在边缘窗口中，存在着一些较长的梯度向量并且所有向量都是朝向某一个特定的方向；在流窗口中，存在着许多梯度向量朝向两个方向；在二维窗口中，梯度向量的方向摇摆不定。

这些差别可以很容易地通过窗口中不同类型的方向来刻画。特别地，矩阵

$$\mathcal{H} = \sum_{window} \left\{ (\nabla I)(\nabla I)^T \right\}$$

$$\approx \sum_{window} \begin{Bmatrix} \left(\frac{\partial G_\sigma}{\partial x} * * \mathcal{I}\right)\left(\frac{\partial G_\sigma}{\partial x} * * \mathcal{I}\right) & \left(\frac{\partial G_\sigma}{\partial x} * * \mathcal{I}\right)\left(\frac{\partial G_\sigma}{\partial y} * * \mathcal{I}\right) \\ \left(\frac{\partial G_\sigma}{\partial x} * * \mathcal{I}\right)\left(\frac{\partial G_\sigma}{\partial y} * * \mathcal{I}\right) & \left(\frac{\partial G_\sigma}{\partial y} * * \mathcal{I}\right)\left(\frac{\partial G_\sigma}{\partial y} * * \mathcal{I}\right) \end{Bmatrix}$$

给出了一个窗口的方向性的表现。在一个静态窗口中，矩阵的两个特征值都很小，因为所有的项都很小。在一个边缘窗口中，我们能够看到与边缘上的梯度相关联有一个大的特征值，另外一个特征值小，这是因为基本上没有其他方向上的梯度。在一个流窗口中，我们能够得到与边缘窗口同样的性质，并且那个大的特征值会更大，因为有许多边在起作用。最后，在一个二维窗口中，两个特征值都是大的。

图 8.16 左边的图显示的是一棵 joshua 树，右边的图表示的是用向量叠加的方法表示左边那幅图的方向，方向用小向量叠加在图中。注意到在拐角和纹理区域，方向向量摆动剧烈

这个矩阵的行为可以通过由一些小常量ϵ绘出的椭圆来理解：

$$(x, y)^T \mathcal{H}^{-1} (x, y) = \epsilon$$

这些椭圆被叠加到图像窗口中。它们的长轴和短轴都沿着$\mathcal{H}$，表示的是上面矩阵的特征向量的方向，这些椭圆沿着它们长轴和短轴方向的程度与特征值的大小相对应；也就是说，一个大圆对应于一个边缘窗口，一个很扁的椭圆指示一个边缘窗口（见图 8.17 和图 8.18）。因此，角点可以通过给定极值的大椭圆区域描点而标出。定位的准确性受到窗口大小和梯度因素的限制。提供更详细的角点模型可以得到更精确的定位（参考 Harris 和 Stephens，1988 或者 Schmid，Mohr 和 Bauckhage，2000）。

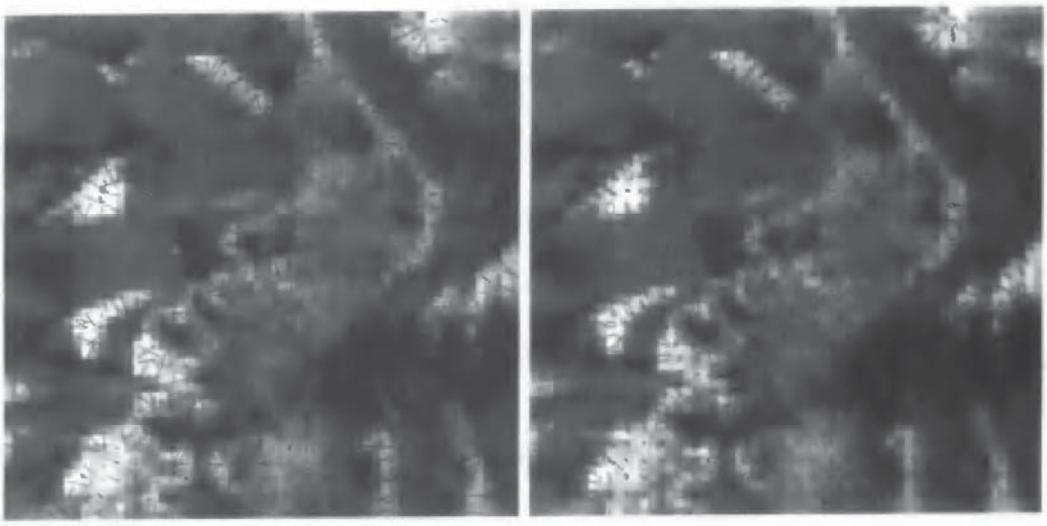


图 8.17 一个 joshua 树图细节的方向图。左边的图，方向是用向量表示并且在图中叠加，梯度值太小的方向会被去掉。右边的图显示的是用3×3的窗口表示的椭圆

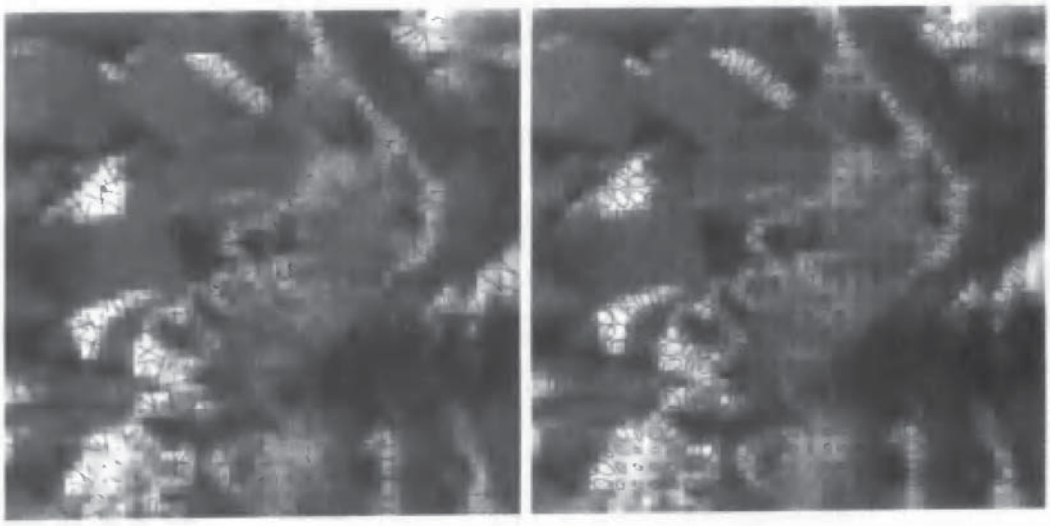


图 8.18 一个 joshua 树图细节的方向图。左边的图，方向是用向量表示的并且在图中叠加，梯度值太小的方向会被去掉。右边的图显示的是用5×5的窗口表示的椭圆

8.4 注释

有很多关于边缘检测的著作。我们所知的最早的文献是 Julez(1959)(是的，1959 年!)；而人们熟知的早期著作当始于 Davis(1975)1975 年的调查，Herskovits 和 Binford(1970 年)，还有 Horn(1971)和 Hueckel(1971)，他们建立了边缘模型并且检测这些模型。

边缘检测还是一个很有争议的题目，它的很多领域仍然是空白，我们所知的充其量只是一鳞半爪。存在许多最优边缘检测器的标准，也有更多的“最优的”边缘检测器。这个领域很关键的一篇文章的作者是 Canny(1986)，其他重要的变种归功于 Deriche(1987)和 Spacek(1986)。Faugeras 的教科书中有着主要论点的细节和附加说明。许多变种方法都归结为在测量梯度前先用类似于高斯函数的东西对图像进行平滑。

物体的边界并不像图像的值那样有很剧烈的变化。首先，物体可能很遗憾地与它们的背

景没有很强的对比度。其次,物体经常带有纹理或者形成自身边界的标记——以至于我们很难找到物体的实际边界。最后,阴影或者类似的东西可能会生成与原物体的边界毫无关系的边界。处理这些问题需要某些方法和策略。

首先,某些应用允许对光照进行控制;如果可以选择光照,选择适当可以使对比度发生巨大改变并且可以消除阴影的光源。第二,将平滑参数设置大一些和将对比度阈值设置高一些,通常可以确保纹理的边缘是平滑的而不会被标记。而这也是不确定的,因为很难选择出可靠的平滑参数值和阈值,并且将纹理纯粹认为是一种无用的东西而不是一种信息的资源是一种保守的做法。

处理边缘和物体边界之间的这种难题还有其他方法。首先,有人可能会去做更好的边缘检测器。许多重要著作都涉及这方面的工作、如定位角点拓扑等。在运用这种方法时,我们的观点倾向于这种努力的研究成果是急剧衰减的,在本章中我们仅能提供一些参考性的提示文章。读者可以从 Bergholm(1987),Deriche(1990),Elder 和 Zucker(1998),Fleck(1992*a*),Kube 和 Perona(1996),Olson(1998),Perona 和 Malik(1990*a*,*b*)或者 Torre 和 Poggio(1986)开始看起。

其次,有人对边缘检测的实用性持完全否定的态度。这种观念出自于这种观察:边缘检测的一些阶段,尤其是在非最大值抑法中,丢掉的信息是非常难以挽回的。这是因为我们做了一个很难做的决定——对阈值进行测试。随着争论的进行,取而代之的,有人希望用一种"软"方法保持这种信息。这种争论听起来很吸引人,而我们却倾向于不采用这一观点,因为当前没有实际的机制能够完整地对这种软信息的量进行操作。

最后,有人把这个看成是一种需要由整个系统体系结构综合处理的问题——这种宿命论的观点是,几乎一切的视觉过程都会见到的令人不愉快的特征,处理这个问题的正确方法是充分理解视觉信息的综合性,以至于建立一个可以容忍这种困难的视觉系统。尽管这是一件吃力不讨好的工作(确切地讲怎样建立这样的体系结构),但是我们发现,这种方法非常有吸引力并且一次又一次地讨论它。

所有的边缘检测器在角点处的性能都很差;它们之间只有细微差别。在过零点的高斯函数的拉普拉斯算子的情况下,这个问题很容易理解(Berzins,1984)。这种糟糕的情况会使讨论陷入两种危险的困境(出了什么问题,要做些什么)。角点检测器有着许多各种各样的形式并且非常复杂,这主要是因为角可以作为很好的点特征,而这对由特征对应支持的一系列算法,如立体视觉、重构以及由运动恢复等非常重要。这些已经导致对角点检测器的定位性质有相当细微的分析(例如,Schmid,Mohr 和 Bauckhage,2000)。

文献另一条线索是去确定边缘检测器做得怎么样。有人可能转而研究定位的准确性(例如,Kakarala 和 Hero,1992,Lyvers 和 Mitchell,1988)或者稳定性(例如,Cho,Meer 和 Cabrera,1997,1998);有人可能会去与人类的偏爱进行比较(例如,Bowyer,Kranenburg 和 Dougherty,1999;Dougherty 和 Bowyer,1998;Heath,Sarkar,Sanocki 和 Bowyer,1997)或者在一项特定任务背景下进行性能比较,像由运动恢复结构(例如,Shin,Goldgof 和 Bowyer,1998)或者识别(例如,Shin,Goldgof 和 Bowyer,1999)。所有的边缘检测器都面临一些共同的困难(例如,在角点,Fleck,1992*b*)。

有时我们将那些对我们的边缘检测器做出响应的边缘称为阶跃边缘,因为它们由像素值发生锐利"不连续"变化,经常用阶跃模型描述的边缘所组成。有人也对其他各种各样的边缘形式进行了研究。最经常引证的事例当数屋顶边缘,它是由一个上升段和一个下降段相遇组成的,就像有些由相互反射效果产生的反光(见图 5.16)。另外一个相互反射效果的例子是一

个阶跃和一个屋顶的合成。我们可以通过使用与上面所叙述的基本相同的台阶来得到这种现象(发现一个“最优化”的过滤器,并且对其输出做非最大值抑制)(Canny,1986;Perona 和 Malik,1990*a*,*b*)。在实际中,人们很少做这些。这看起来似乎有两个原因。第一,被采用的模型没有合适的理论(或者实践)基础。什么样的合成边缘值得去找呢?最简单的回答——那些容易推导出最优化过滤器的——是最令人满意的。第二,屋顶边缘和更多合成边缘概念的语义比阶跃边缘更加含糊不清。当发现屋顶边缘时该如何去做,我们毫无概念。

边缘不容易确定并且通常不容易被检测,但是人们可以用边缘检测器的输出结果来解决问题。屋顶边缘同样不容易确定也同样不容易被检测,我们从来没有见到用屋顶边缘检测器的输出结果解决问题。真正困难的地方似乎是没有可靠的机制提前预言什么值得去检测。我们会在后续的讨论中触及这个非常困难的问题。

习题

8.1　一幅 500×500 像素的图像$\mathcal{I}$中的每一个像素值都是一个独立的具有平均值为 0、标准偏差为 1 的标准分布的随机变量。使用前向差分估计($|I_{i+1,j} - I_{i,j}|$)估计 x 轴方向导数绝对值大于 3 的像素点的数量。

8.2　一幅 500×500 像素图像$\mathcal{I}$中的每一个像素值都是一个独立的具有平均值为 0、标准偏差为 1 的标准分布的随机变量。$\mathcal{I}$与一个 $2k+1 \times 2k+1$ 的核$\mathcal{G}$卷积。求结果中像素值的协方差。有两种方法来求解:逐个求解(例如,分别在 x 轴方向和 y 轴方向大于 $2k+1$ 的点的值明显是独立的)或一次性全部求解。不考虑边界处的像素值。

8.3　有一个输出值范围为 0 到 255 整数值的摄像机,其空间解析度是 1024×768 像素,并且每秒生成 30 帧。在某个场景中,在没有噪声的情况下,它将输出常量 128。摄像机的输出受平均值为 0、标准偏差为 1 的可加性静态高斯噪声的影响。按该模型预计,需要等多久能够见到负值出现(提示:你会发现使用对数计算答案将会很有帮助,因为直接计算 $\exp(-128^2/2)$将会得 0;窍门是使用一个大的正数和大的负数对数来约去)?

8.4　在 8.3.1 节中曾说过,相对于一维情况,二维情况下切合实际的二阶导数必须是旋转不变量。为什么?

编程作业

8.5　为什么在图像的拉普拉斯算子的过零点处检查梯度幅值是否足够大是必要的?证明对何种边缘,这个检查是重要的。

8.6　高斯函数的拉普拉斯算子看起来就像是两个不同尺度的高斯函数的差分。在不同的两种尺度值的情况下,比较这两个核函数。哪个给出了更好的近似值?使用过零法检测边缘时这个近似值的误差有多大影响?

8.7　获取一个 Canny 边缘检测器的实现(MATLAB 在图像处理工具包中也有一个实现),并且使用一些图像来显示出尺度和对比度阈值在边缘检测中的影响。是否容易找到刚好能标记出物体边缘的边缘检测器?在哪种应用中会容易找到?

8.8　在实现了迟滞的边缘检测器中,很容易使迟滞失效——即实质上把较低的阈值和较高的阈值设置成同一个值。使用这个窍门来比较有迟滞和没有迟滞的边缘检测器的结果。

有如下的问题需要考虑：

(a) 如何处理边缘检测器的输出？具有连接边缘点链有时是有益的。此时迟滞有明显的帮助吗？

(b) 噪声抑制：我们常常希望强制边缘检测器忽略一些边缘点而标记其余的。一个边缘是有用的特征是其具有高对比度(它并非是可靠的方法)。对抑制低对比度的边缘点但又不切断高对比度的边缘，使用迟滞有多大的可靠性？

第9章 纹　　理

纹理是一个非常普遍的现象，容易辨认但是很难定义。一般而言，一个效果是否被称为纹理，是由观察它的尺度决定的。一片树叶占据了图像大部分画面时只算是一个物体，但是一棵树的树冠是纹理。纹理有许多不同的来源。首先，很多小物体组成的图像被认为是纹理。例如，草、树叶、灌木丛、小卵石和头发。其次，物体表面上看起来像很多小物体组成的有规律的形状，也可以认为是纹理。例如，印度豹、美洲豹身上的斑点，老虎或斑马身上的条纹，树皮、木头和皮肤上的图案。

在纹理处理中一般有三个基本问题：

- **纹理分割**是把图片分成不同的部分，每部分内部具有相近的纹理问题。纹理分割包括表示纹理、确定分割区域的边缘的基本原理。在这一章，我们只涉及如何表示纹理(9.1 节)；第 14 章和第 16 章会讨论如何利用这些表示分割有纹理的图像。
- **纹理合成**寻找如何利用小的范例图像构造大片纹理区域的方法。我们用范例图像来建立纹理的概率模型，然后利用概率模型来获得纹理的图像。有很多建立一个概率模型的方法，9.3 节讲述了 3 个目前正在使用的较为成功的方法。
- **纹理恢复形状**包括由图像纹理恢复表面的方向和表面的形状。通过假定在同一个表面的不同点的纹理“看起来相同”来完成这一点，这意味着不同点之间纹理的差异提供了表面形状的信息。9.4 节会给出有关这方面的主要(相当技术性的)论证。

9.1 纹理表示

图像的纹理一般由一些十分规则的子元素(又称为纹理基元)以有组织的模式构成(见图 9.1 和图 9.2)。例如，图 9.1 的一个纹理中由三角形构成，类似的，该图中的另一个纹理由箭头记号组成。尝试表示纹理的一个很自然的方法就是找到纹理基元，然后描述它们放置的方式。

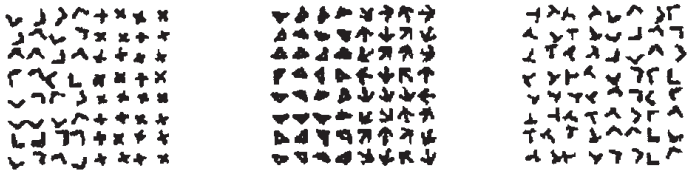


图 9.1　纹理的一些例子，用于人们在研究中显示如何容易地区分各种各样不同的纹理。注意这些纹理是由风格非常相似的子元素构成的，以某种有含义的方式重复

这种方法的困难之一就是没有已知的标准纹理基元集合，这意味着我们不清楚应该寻找什么，所以应该选取更加简单的模式元素——比如说点，条形——思考它们空间上的分布，来取代寻找箭头和三角形这个层次上的模式。这种方法的好处在于它非常容易利用对图像进行滤波找到简单的模式元素。

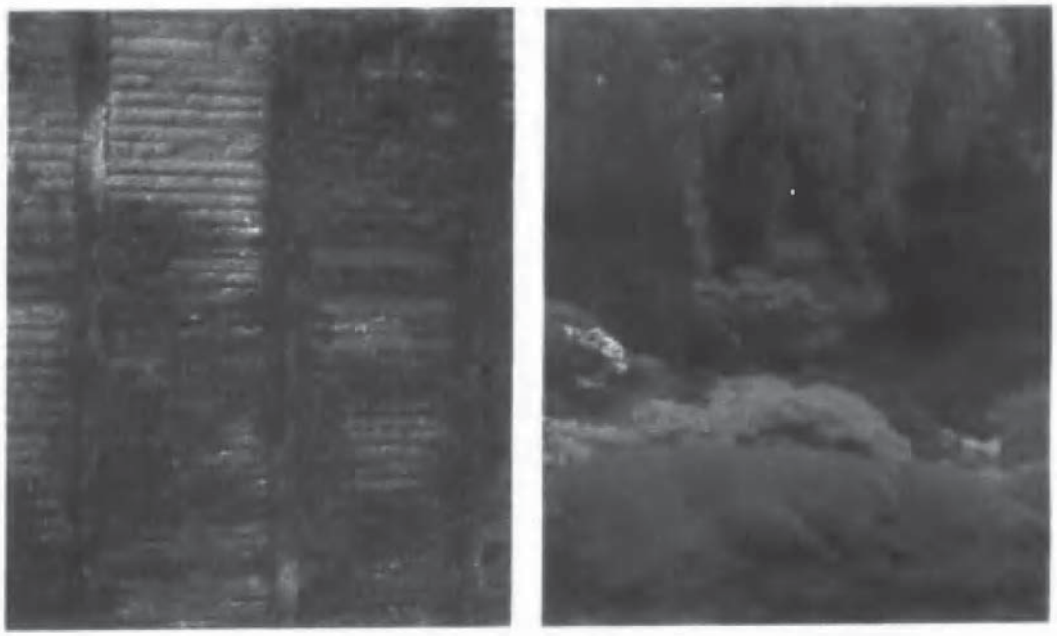


图 9.2 典型的纹理图像。像灌木丛、草丛、树叶、水这些物质，对于它们属于何种物质的理解与它们的纹理之间有非常密切的关系(左边那幅图，当你的手指在上面抹过它的表面时是什么感觉？是不是湿的？)，注意利用纹理从右边那幅图，得到多少关于植物的种类，它们的形状等信息。这些纹理同样由风格非常相似的子元素构成，按一个大致的模式排列

9.1.1 利用滤波器库提取图像结构

在 7.5 节，我们看到用一个线性滤波器卷积图像可以得到一幅图像在不同基上的表示。用一个滤波器卷积图像将该图像变换到一个新的基上的好处是，这个过程使得图像原有的结构清晰可见。这是因为当一个邻域内的图像模式和滤波器核很相近时，可以得到一个很强的响应，否则会得到一个很弱的响应。

因此可以根据一个滤波器集的响应表示图像纹理，这个不同的滤波器的集合应该包括一系列模式——通常有点和条形时——不同规模尺寸的集合(比如说，用于区分大的点和小的点、大的条形和小的条形)。卷积后图像中一点的数值表示，点的结构(spottiness)(或条形结构，等等)在特定尺度下在图像中的响应对应点上。当大量重复这种表示时，它对于揭示结构(点的结构、条形结构，等等)在一定程度上证明是很有用的。

通常，点滤波器是非常有用的，因为它们对于与其邻域不同的小区域响应强烈(例如，在边缘的任意一边或一点)。另一个吸引人之处是它可以检测不带方向的结构。另一方面，条形过滤器是有方向性的，往往对于带方向的结构响应明显。

点和条形用高斯的加权和 对于应该使用什么样的滤波器，没有标准的答案；各种各样的答案都被尝试过。类似于人的视觉皮层，一般至少使用了一个点滤波器和一个由不同方向、不同尺寸和相位的带方向的条形滤波器组成的集合。条形的相位指的是条形垂直方向的相位，类似于正弦曲线的相位(例如，当交叉点在原点过零的时候，相位就是 0 度)。

获得这些滤波器的一个方法是，在不同的尺寸比例形成不同权值的加权高斯滤波器差分，这种技术在图 9.3 的滤波器中使用，这个例子的滤波器包括：

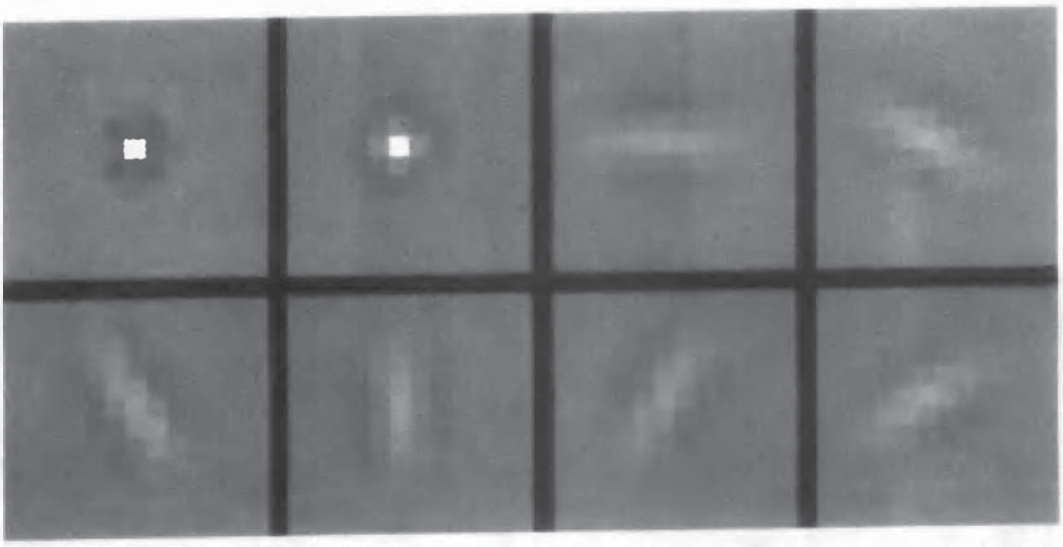


图9.3 一组8个滤波器用来将图像扩展为一系列的响应。这些滤波器在一个固定的尺寸下显示,零代表中等灰度等级,浅色一些的值代表正值,深色一些的值代表负值。它们代表两个不同的点和6个条形,这组滤波器曾被Malik和Perona(1990)使用过

- **一个点滤波器**,用三个同中心、对称的高斯滤波器加权得到。它们的权值分别是1,-2,1。相应的sigma是0.62,1,和1.6。
- **另一个点滤波器**,用两个同中心、对称的高斯滤波器加权得到。它们的权值分别是1,-1。相应的sigma是0.71和1.14。
- **一系列带方向的条形滤波器**:由三个方向性的高斯滤波器的加权和组成,彼此之间有偏移。这些条形滤波器有6种样式;每种都是一个水平的条形滤波器旋转后的样式。在水平方向的高斯滤波器权值是-1,2,-1。它们在x和y方向的sigma不同,x方向的sigma值都是2,y方向的sigma值都是1。中心点在y轴上偏移,位于(0,1),(0,0)和(0,-1)处。

应该清楚的是,如何选择滤波器的细节几乎是无关紧要的。有大量的经验建议,应该有一系列的点滤波器和不同尺寸及方向的条形滤波器——这些是集合提供的,但是并没有什么理由使人相信最优化选择的滤波器带来什么明显的好处。

图9.4和图9.5显示了一个蝴蝶的输入图像对于这个滤波器库的响应的绝对值。注意到这样一点,条形滤波器并非完全可靠的条形检测器(因为某特定方向的条形滤波器对各种尺寸与方向的条件都有响应),滤波器给出了图像数据可信的描述。一般来说,条形滤波器对于方向的条形响应强烈,对于其他模式响应很弱,而点滤波器对于孤立的一些点做出响应。用输出的绝对值给出结果,浅色的像素代表强烈的响应,图像的排列顺序与上面滤波器的排列位置相对应。

滤波器的个数以及它们的方向 我们并不知道对于有效的纹理算法用多少个滤波器是最好的。Perona(1995)列出了各种各样的系统使用的不同个数和方向的数目。滤波器数目从4到11,方向数目从2到18。各个程序的方向数目相差很大,只要至少有6个方向,似乎就没有多大影响。一般来说,点过滤器是高斯滤波器,而条形滤波器则是由带方向的高斯函数差分得到的。

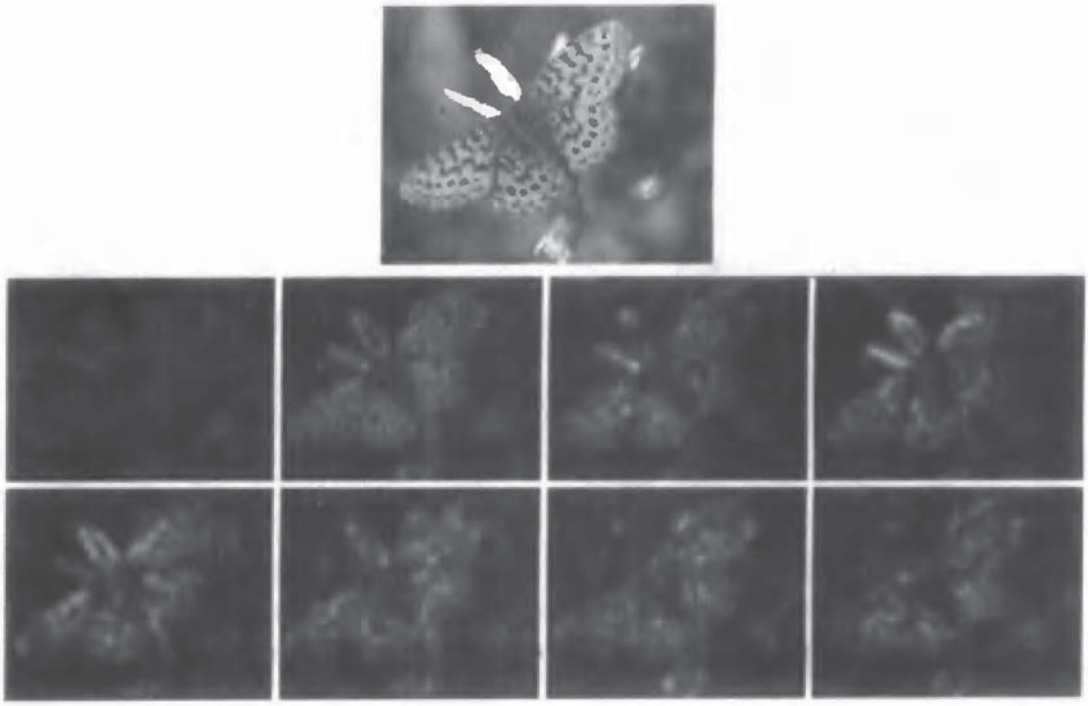


图 9.4 顶部是一个尺寸适中的蝴蝶图像，下面是使用图 9.3 的各个滤波器之后的结果

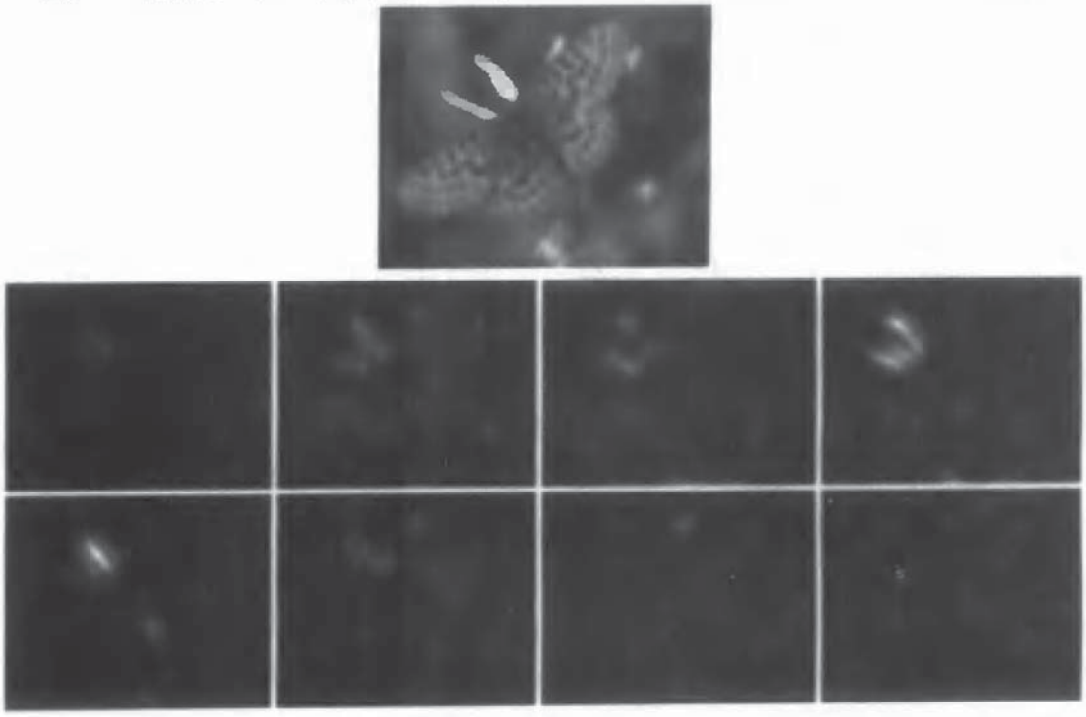


图 9.5 一个蝴蝶的输入图像和图 9.3 的滤波器的响应，该滤波器的尺度比图 9.4 的大。注意带方向的条形滤波器对于翅膀、触角和翅膀边缘上的条形响应。事实上，一个条形滤波器出现响应并不代表另一个会产生效果，但是响应的大小是对图像中条形方向的提示

同样没有任何迹象显示，利用更加复杂的一组滤波器比使用基本的点滤波器加条形滤波器组合能得到更多的好处。这里有一个矛盾：使用更多的滤波器将导致更多的细节（和冗余度大）的图像描述，但是必须将图像和所有这些滤波器卷积，这是非常耗时的。简化这个过程的一个方法是利用金字塔结构来控制要处理的冗余信息。

9.1.2 利用滤波器输出的统计表示纹理

一组滤波后的图像就其本身而言并不是纹理的表示，因为我们需要纹理元素整体分布情况的一些表示，例如，一块黄色花朵的田野中由很多黄色的点组成，还有一些垂直的绿色的条形。一个斑马是由一些黑色的条带在白色的背景上组成。这里隐含假设了一个尺度，纹理可以在这个尺度上被描述。在黄色花朵的田野上，一个小的图像窗口也许只包含一朵花；在斑马上它也许只包含一块黑色或者白色的区域。与此类似，窗口太大则除了有关系的纹理外还会包括背景。注意这里有两个尺度：滤波器的尺度和用来考虑滤波器分布的尺度。

假定用来表示纹理图像窗口的尺度已知。一个典型的描述包括这个窗口的一组滤波器输出的统计。输出通常进行平方(它的一个优点在于计算从黑色到白色的条纹和白色到黑色的条纹是一样的)。例如在图9.6中，根据水平和垂直的纹理显示了一个假定的表示。这个表示是通过对水平(垂直)的条形滤波器的输出平方后得到的，然后用一个较粗的尺度对其进行平滑。这个平滑等价于对在一些窗口中平方滤波器的输出的均值进行估计。最后，平滑后的输出被分类到某一描述纹理的类。在图9.6的示例中，纹理被分为4类中的某一类，分类的根据是它们的水平输出结果很大，或者垂直输出结果很大，或者两个输出结果都很大，或者两个输出结果都不是很大。

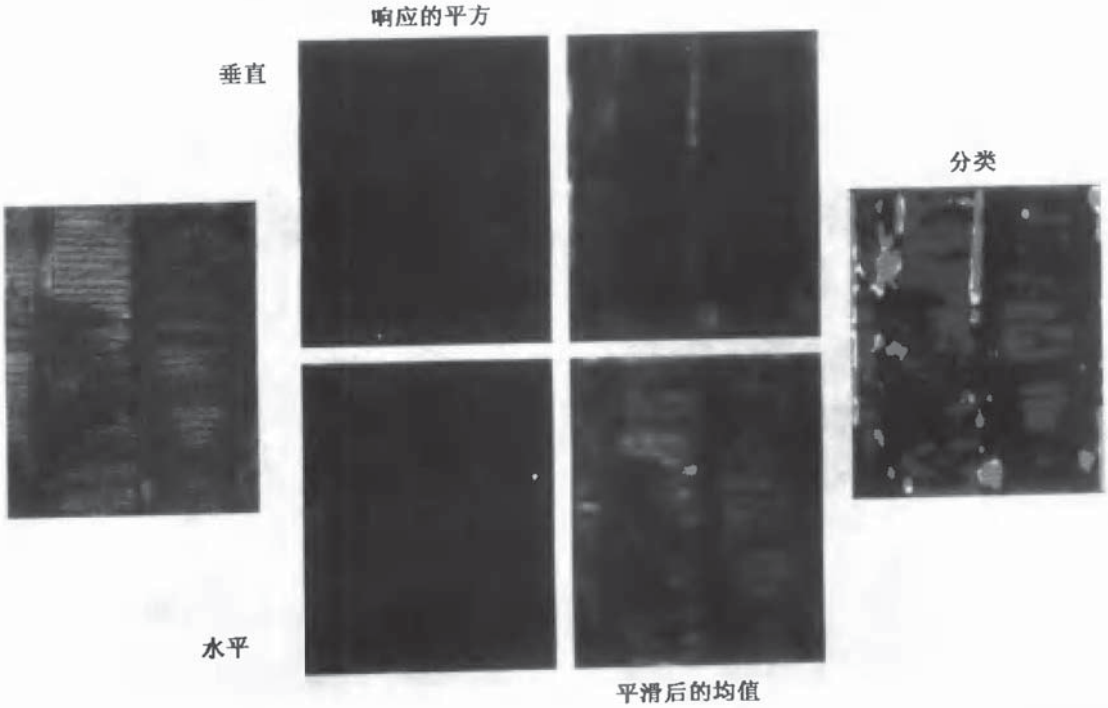


图9.6 一个根据滤波器输出结果假定的描述。我们明显地减少了滤波器的数量(有两个不同的滤波器，一个垂直，一个水平)。左边的图像是输入，注意可以将它的各部分可靠地描述为垂直的，或是水平的，或是模糊的。中间左边那一列的图像是平方后的滤波输出结果(平方是用来使从黑色到白色的转变和从白色到黑色的转变相同)。这些值用同样的线性尺寸来显示，浅一些的点代表更强的响应，中间右边的图像是平滑后的结果(可以认为是在一个小窗口的平方响应的平均值)。这种平均响应使垂直的条纹在垂直的图像中更加明显，水平的条纹在水平图像中更加明显。最后，阈值化两幅图像，然后合并它们得到右边的图像。(黑色的值既不是垂直也不是水平，深灰度的值是水平，浅灰度的值是垂直，白色的值是水平且垂直)

统计量的选择 应该收集什么样的统计量在某种程度上取决于打算描述的范围。然而，在纹理合成方面的工作说明了对选择合适模型的一些约束条件，这也正是 9.3 节在这个问题上花费很大篇幅的原因。假设现在已经决定了收集统计量的窗口尺度，一个策略就是对一组滤波器计算滤波器的输出平方的平均值(Malik 和 Perona，1989)。于是一个窗口就可用一些数字组成的一个向量来描述，这些数字中每一个都是一些滤波器在该窗口响应平方的平均值。这种方法能够将包含很多点的窗口与包含很多条纹的窗口区分开，前者对点滤波器的平均响应值高，而后者对条形滤波器的平均响应值高。这就是图 9.6 使用的方法，但是滤波器的数量要多一些。

另一个可选的方法是计算窗口上的滤波器输出的平均值和标准偏差，然后用它们作为特征向量(Ma 和 Manjunath，1996)。这种形式下的纹理描述可以用于在很多例子的基础上(见图 9.7)恢复图像窗口。这种方法是很有用的，因为在卫星图像中，一块区域是建筑物还是植物，可以很容易地由它们的纹理检测出来。因而可以说，如果能够匹配纹理，便可以找出卫星拍摄的图片上的所有区域，例如卫星图像中的植物。两个特征向量相差很远的纹理也许看起来很像，这个可以通过修改度量特征向量的差异的方法来处理。

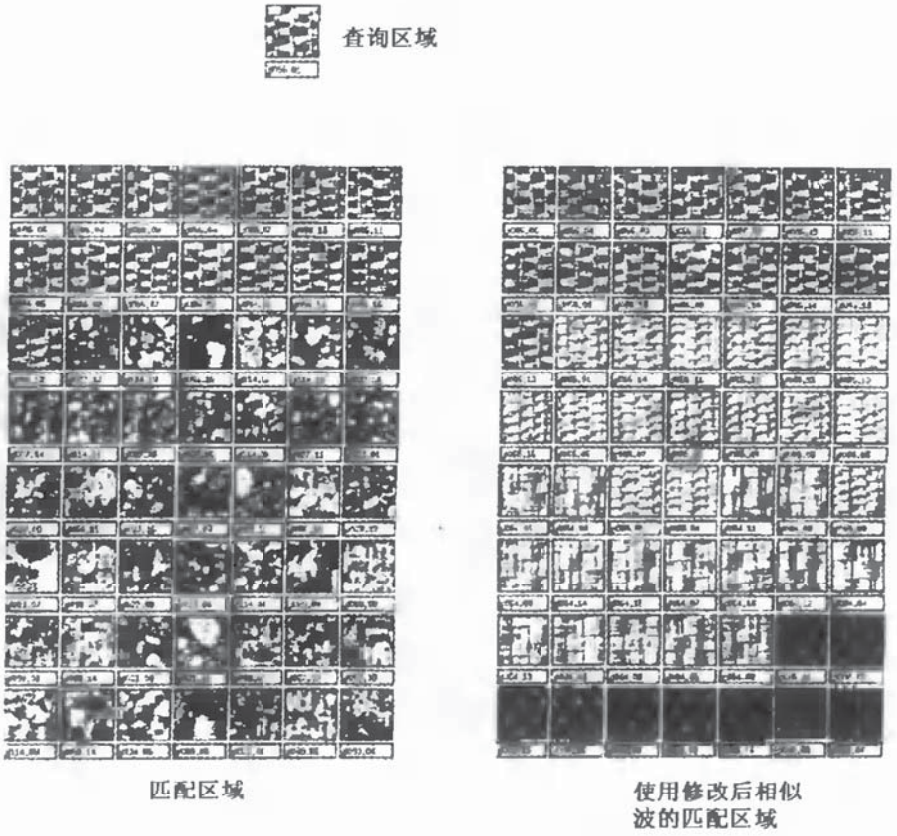


图 9.7 纹理可以用窗口上的滤波器输出结果的平均值和标准偏差描述。如果利用一些不同的滤波器，则导致一个向量描述一个窗口。很多点组成的纹理对于点滤波器会有高平均值的输出，很多条形组成的纹理对于条形滤波器也会有高平均值的输出，类似等。这表明一个图像可以利用特征向量为基础比较和其他图像之间的距离。一个单纯的欧几里得距离得到可以接受的结果(左边)，但是一个修正过的距离函数得到很好的结果(右边)

平均值、平均值加标准偏差都不是一个理想的描述,因为滤波器响应之间的关系是非常重要的。例如:想像纹理包括许多小点分布成条形——比方说,从天空上看一片卷心菜田。像这样一个纹理,小的点滤波器会强烈响应,大的条形滤波器也会强烈响应。但是响应是相关的,小的点滤波器响应强烈的地方,大的条形滤波器也会响应强烈。另一方面,如果纹理由很多大的条形和分散在背景中很多小的点组成,小的点滤波器会反应强烈,而大的条滤波器也会反应强烈,但是反应并不相关。可以尝试和记录滤波器输出的协方差——它能处理上面所述的卷心菜田的例子——但是一般会有过多的项来形成一个准确的评估。通常可以定义一些项,它们是一些在某个特定的应用中也许会有用的协方差项,然后使用它们来取代上面的方法。

尺度的选择 另一个要处理的实际问题就是选择表示纹理的尺度。通常,我们在感兴趣的点选择一个小的窗口,然后增加窗口的尺寸,直到再增加不会带来显著的变化。例如,想像在斑马的图像上选择一个像素。图像在这个像素的周围一个非常小的窗口上有一个固定的值,比方说黑色。当窗口稍微放大,就会有明显的变化,在窗口中会有一些黑色的点和一些白色的点。当窗口增大到包括一些条纹的时候,再增大窗口不会带来明显的变化,它只会变成更大的包含很多条纹的窗口(stripey window)。

一个可以用来决定何时停止增大窗口的统计量是极性。首先决定窗口的主方向——梯度的平均方向。对于每个梯度向量,首先得到梯度向量和主方向的点积,其次得到正值的点平滑后的平均值和负值的点平滑后的平均值,然后得到它们之间的差值。这个值测量在一个区域中沿主方向的梯度大小(正值点)和逆主方向的梯度大小(负值点)的相对关系。

可以对不同窗口尺寸的测量该统计量,可以在某个范围内的窗口尺寸中做这件事。先从一个合适的尺寸开始,然后考虑逐渐增大窗口直到尺寸改变的时候极性不再改变。注意这个判据有可能并不惟一。例如,想像一张分辨率很高的斑马图像,如果由一个足够小的窗口开始(可能只有一根头发粗细),这个判据会选择一个包括几根头发的窗口尺寸。如果从一个稍大些且包括很多头发的窗口开始,判据会选择一个包围着几个条纹的窗口尺寸。

9.2 使用有方向性金字塔的分析(和合成)

用一系列滤波器的输出作为统计量来表示纹理,需要把图像和很多不同尺寸的不同滤波器进行卷积。有些相当好的方法可以系统化地做到这一点,这些方法将在这一节讲述。许多读者也许不关注这个问题,那么可以直接跳到9.3节。

用一些滤波器和图像卷积的过程称为分析,用一些方向性的滤波器卷积图像,有时不很严格地描述为方向分析或者方向表示。高斯金字塔(7.7.1节)是一个用滤波器库分析图像的例子——在这个例子中,用的是平滑滤波器。高斯金字塔系统的处理尺寸是通过对图像平滑后,再间隔采样进行的。这意味着产生下一个粗糙的尺寸较为容易,因为不需要处理冗余的信息。

事实上,高斯金字塔是一个高度冗余的描述,因为每一层都是上一层低通滤波后的结果——这意味着我们必须把低频描述很多次。高斯金字塔的每一层都是更细一层外观的预测——这个预测并不精确,但它表明并不需要储存更细层的全部,只需要保持一个和预测差异的记录。这是拉普拉斯金字塔的灵感来源。

拉普拉斯金字塔可以获得冗余度较少的多尺度表示,但是它不能直接处理方向。通过在傅里叶领域采用金字塔,可以得到一个也能对方向信息编码的方法(9.2.2节)。9.2.3节描述了一个可以描述方向的方法。

9.2.1 拉普拉斯金字塔

高斯金字塔的每一个粗糙层都预测了下一层的样子，拉普拉斯金字塔利用了这一点。如果我们有一个倍频取样的操作，可以得到和下一个更细层同样尺寸的粗糙层版本，那么只需保存该预测和下一个更细层之间的差异即可。

显然，图像信息不能凭空产生，但是可以利用重复像素扩展一个粗糙的尺寸，这包括将图像从第 $n+1$ 层变为第 n 层的倍频取样操作 $S^{\uparrow}$。特别地，$S^{\uparrow}(\mathcal{I})$ 从一个图片产生一个每一维尺寸加倍的图像。输出图像在 $(2j-1,2k-1)$；$(2j,2k-1)$；$(2j-1,2k)$；$(2j,2k)$ 的 4 个元素都和 $\mathcal{I}$ 的第 (j,k) 个元素的值相同。

分析——由图像建立拉普拉斯金字塔 拉普拉斯金字塔的最粗糙的尺寸层和高斯金字塔的粗糙尺寸是一样的。拉普拉斯金字塔每一个更细尺寸的层，是高斯金字塔的层与对高斯金字塔下一层的倍频取样获得的预测之间的差异。这意味着：

$$P_{\text{Laplacian}}(\mathcal{I})_m = P_{\text{Gaussian}}(\mathcal{I})_m$$

（m 是最粗糙的层）和

$$\begin{aligned} P_{\text{Laplacian}}(\mathcal{I})_k &= P_{\text{Gaussian}}(\mathcal{I})_k - S^{\uparrow}(P_{\text{Gaussian}}(\mathcal{I})_{k+1}) \\ &= (Id - S^{\uparrow}S^{\downarrow}G_{\sigma})P_{\text{Gaussian}}(\mathcal{I})_k \end{aligned}$$

所有这些可以得到算法 9.1，虽然拉普拉斯这个名字有些误导——这里没有差分操作——但它并非不可接受，因为每一层都是高斯滤波器差分的近似。

算法 9.1 根据图像建立拉普拉斯金字塔
形成高斯金字塔
将高斯金字塔的最粗糙层设置为拉普拉斯金字塔的最粗糙层
对最粗糙层的下一层开始到最细层中的每一层，对较粗糙层用倍频采样，并从高斯金字塔这一层减去，得到拉普拉斯金字塔的这一层
end

拉普拉斯金字塔每一层可以被想像成带通滤波器的响应，这是因为所获得的图像是用一个特定分辨率的图像，减去那些可以用粗糙层的版本——对图像的低通部分响应——预测的部分。这意味着可以期待图像中一组一个特定频率的条纹，会导致在金字塔某一层响应强烈，而在其他层响应很弱（见图 9.8）。

因为金字塔不同层表示不同的空间频率，所以拉普拉斯金字塔可以用来作为一个有效的压缩机制。

合成——从拉普拉斯金字塔恢复图像 拉普拉斯金字塔有一个重要的特征，很容易由图像的拉普拉斯金字塔恢复图像。先利用拉普拉斯金字塔来恢复高斯金字塔，然后得到最精细尺寸的高斯金字塔（便得到图像）来完成这个任务。首先，拉普拉斯金字塔的最粗糙层就是高斯金字塔的最粗糙层。高斯金字塔最粗糙层的下一个层这样获得：先得到最粗糙层，对它倍频采样，然后和拉普拉斯金字塔最粗糙层的下一层相加（这样一层一层加起来）。这个过程称为合成，在算法 9.2 中描述。

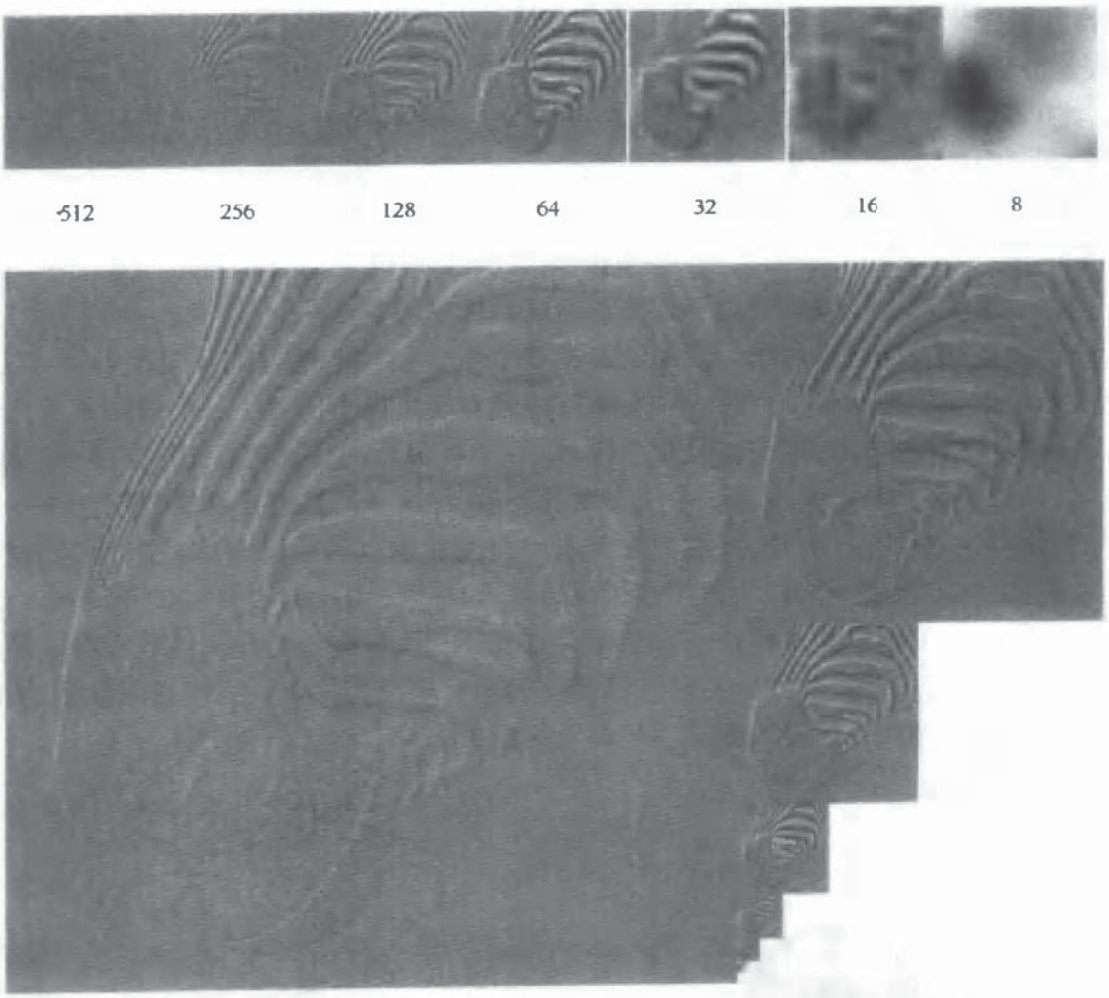


图 9.8 图像的拉普拉斯金字塔，从 512×512 到 8×8，零响应用中等灰度编码，正值颜色较浅，负值颜色较深。注意条纹对特定的尺寸给出较强的响应，因为每一层对应于一个带通滤波器的输出(大致的)

算法 9.2 合成：从拉普拉斯金字塔得到一个图像

设定现在的图像是最粗糙层

对于最粗糙层的下一层开始到最细层中的每一层，倍频取样当前图像并将当前层相加到结果中去

　设定当前图像是操作的结果

end

当前图像现在包含原始图像

9.2.2 空间频率域中的滤波器

卷积定理(在空域卷积相当于在傅里叶频域上相乘)告诉我们滤波器做了什么以及金字塔包含什么信息。下面将通过几个方面来阐述这一理论：平滑和低通滤波之间是一个自然的类推，一些带通滤波器自然的响应方向性结构，以及某种局部空间频域分析可以使用一族特定的滤波器得到。

平滑和低通滤波器 卷积定理指出，将图像和一个各向同性的、均方差为 σ 的高斯滤波器卷积，等价于将图像的傅里叶变换和一个标准差为 $1/\sigma$ 的高斯滤波器相乘。高斯滤波器衰减得非常快，特别当 σ 很大时。这说明当高频只是 $1/\sigma$ 的几倍时，结果的傅里叶变换在高频的能量极少。可以把这个认为是一个低通滤波——它在低频增益很大，在高频增益很小。这是一个很合理的认定，如果用一个均方差很小的高斯滤波器平滑，除了最高的频率之外几乎所有频率都被保留，如果利用一个标准差很大的高斯滤波器平滑，结果会很接近图像的平均值。这说明高斯金字塔本质上是图像经过一组低通滤波器后的版本。

带通滤波器和方向选择因子 带通滤波器就是对于某个范围内的频率增益很大，对于高于和低于它的频率的增益很小。有一种带通滤波器对于方向不敏感。一个很自然的例子就是用两个各向同性的高斯滤波器的差分平滑图像，一个有很大的标准差，一个有很小的标准差。在频域，这个滤波器的核看起来像一个值很大的环形(图 9.9 的左边)，这说明它选择一定范围的频率，但是对于方向性没有选择(因为在频域空间上和原点等距离的那些点，它们表示频域的基本元素，但是方向各不相同)，理想的带通滤波器在环面内是单位值，在外面是零。这样的带通滤波器会有无限的空间延伸——使它很难处理——所以差分的高斯函数看起来是一个令人满意的实际选择。当然，差分的高斯滤波器是用来获得拉普拉斯金字塔的滤波器，所以拉普拉斯金字塔是由图像经一组带通滤波器处理后的结果。

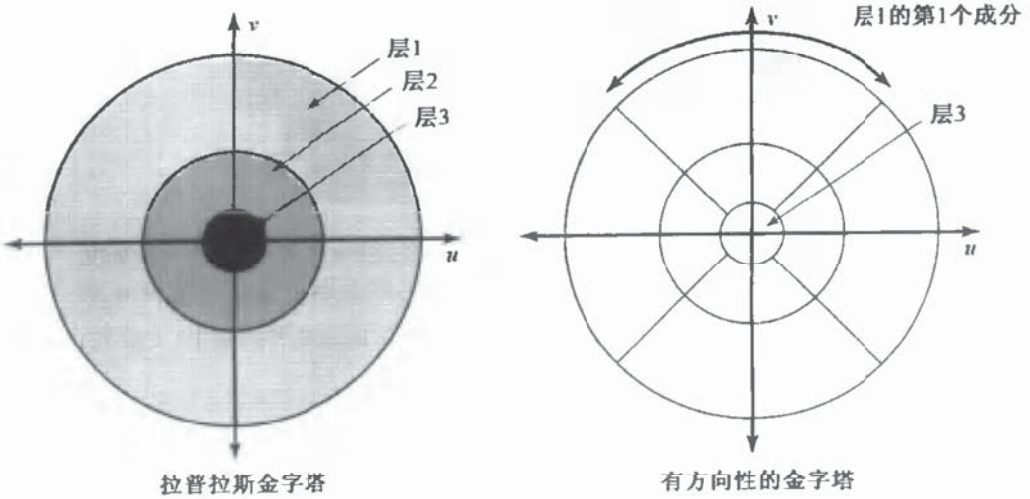


图 9.9 拉普拉斯金字塔的每一层都包含一个平滑和重抽样的图像中的元素，这个图像没有被下一个平滑层所描述。假设高斯滤波器是一个足够好的滤波器，每一层可以认为是对于图像部分在一定范围内的频率的描述，它表明每一层的傅里叶变换是傅里叶空间 (u,v) 的一个环面内的值((u,v) 向量的幅度代表频率)。这些环的和就是图像的傅里叶变换，所以每一层从图像的傅里叶变换中剪掉了一个环形。一个有方向性的金字塔将每个环形分成扇形。如果 (u,v) 空间用极坐标来描述，每一个扇形对应于一段径向值范围和一个角度值范围($\arctan(u/v)$ 给出了傅里叶基本元素的方向)

另一种类型是带通滤波器的傅里叶变换，在环形的一个扇形区域内的值很大，而在区域外的值很小(图 9.9 的右半部分)——这个滤波器是有方向选择性的，意味着它对于一些特定范围的频率和方向的信号响应强烈。

局部频域分析和 Gabor 滤波器 傅里叶变换的一个难点是傅里叶系数的确定依赖于整幅图像，傅里叶变换对于一些特定的 (u,v) 值需要用到图像的每一个像素来计算。这是一个在考虑

图像时很不方便的方法,因为失去了整个空间的信息。例如,图9.12中条纹是变得越来越粗的。如果根据局部定义的频域来考虑,那么可以用图像空间频率变化的观点来思考这个现象。在包围一个点的某个窗口中,窄的条纹看起来像是高频项,宽的条纹看起来像是低频项。

Gabor滤波器可以解决这个问题。它的核看起来像傅里叶的基乘以高斯函数,这意味着Gabor滤波器对于图像中只是在局部范围内有特定频率和方向的点响应强烈。Gabor滤波器都是成对出现的,通常被称为积分对。积分对中的一个复现某特定方向的对称性分量,另一个复现其反对称分量,对称核的数学公式是

$$G_{\text{symmetric}}(x, y) = \cos\left(k_x x + k_y y\right) \exp - \left\{\frac{x^2 + y^2}{2\sigma^2}\right\}$$

反对称核的公式是

$$G_{\text{antisymmetric}}(x, y) = \sin\left(k_0 x + k_1 y\right) \exp - \left\{\frac{x^2 + y^2}{2\sigma^2}\right\}$$

滤波器在图9.10和图9.11中说明,(k_x, k_y)给出滤波器响应强烈的空间频率,σ是滤波器的尺度。原则上,使用很多不同尺度、频率、方向的Gabor滤波器,就可以给图像一个详细的局部描述。Gabor滤波器σ为无穷时近似于傅里叶变换,这也说明了为什么有两种滤波器,而且解释了为什么可以认为Gabor滤波器给出了一个局部的空间频率分析。

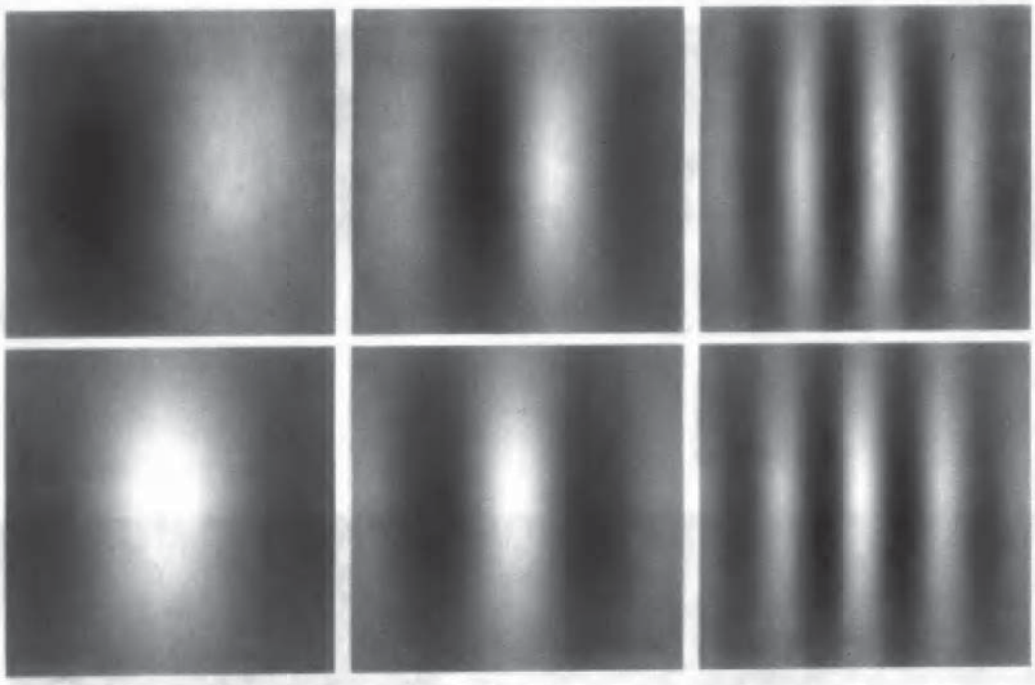


图9.10 Gabor滤波器是对称高斯函数与带方向的正弦函数相乘的产物,文中给出了核的形式。将Gabor滤波器的核当成图像显示,中等灰度的值代表零,深一点的值代表负值,浅一点的值代表正值。上面一行显示了反对称性的部分,下面一行显示了对称性的部分。对称和反对称的部分有$\pi/2$弧度的相位差,因为条纹的横截面垂线(在这个例子中是水平的)给出的正弦曲线有这个相位差。这些滤波器的尺寸是固定的,其分别展示三个不同的频率。图9.11显示了更细尺寸的Gabor滤波器

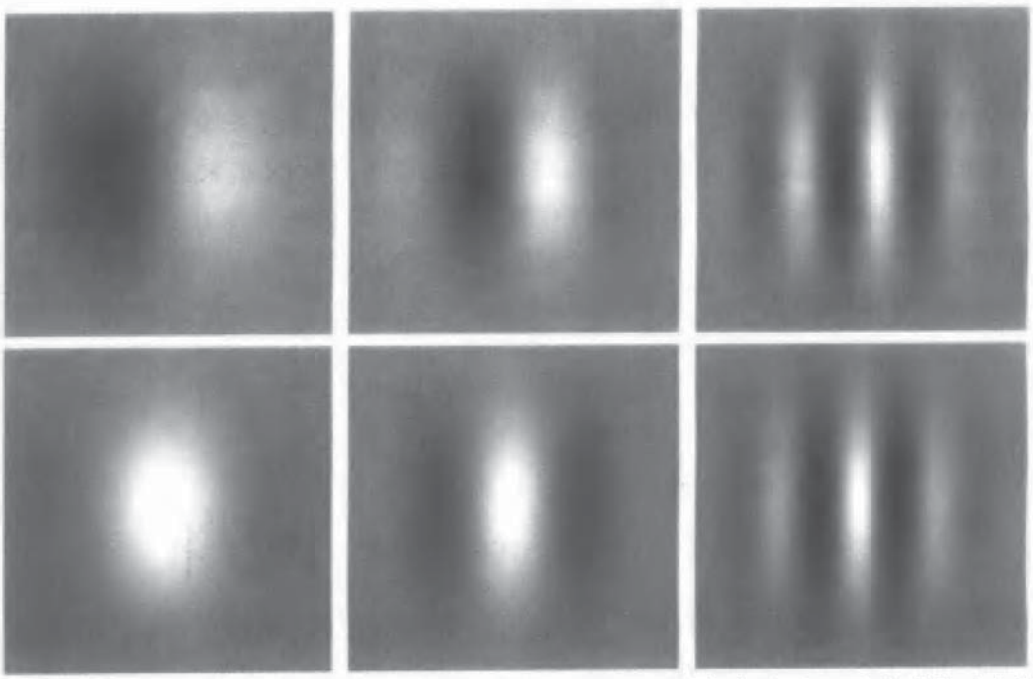


图 9.11 这个图像将 Gabor 滤波器的核当成图像显示，中灰度的值代表零，深一点的值代表负值，浅一点的值代表正值。上面一行显示了反对称性的部分，下面一行显示了对称性的部分。这些滤波器的尺寸是固定的，但分别展示三个不同的频率。图9.11显示的滤波器比图9.10的尺度更细

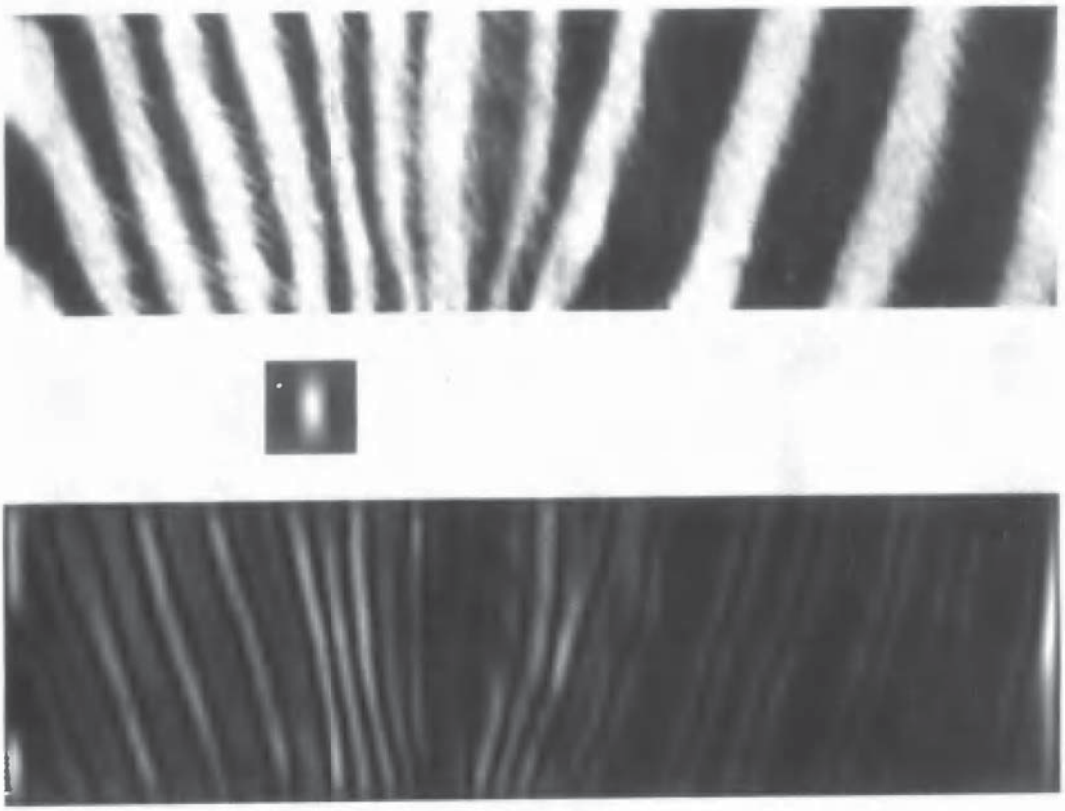


图 9.12 上面的图像显示了一张斑马图像的细节，之所以选中它是因为它有不同尺度和方向的条纹。将图像用中间的Gabor滤波器的核卷积。下面的图像显示运算的绝对值，注意当条纹的频率大致和Gabor滤波器核的高斯窗口相匹配时响应很大(即，核中的条纹具有和三条条纹大致相同的尺寸和方向)。当条纹变大或变小时响应衰减。因此，滤波器执行的是一种局部的频域分析。这个滤波器是积分对中的一个(它是对称性的部分)。反对称性的部分同样有选频性。两个响应可以认为是(复数)局部傅里叶变换的两个部分，所以可以从中提取出幅度和相位信息

9.2.3 带方向的金字塔

拉普拉斯金字塔没有包含分析图像纹理足够的信息，因为里面没有对于条纹方向的显式表示。解决这个问题的一个很自然的策略就是将每一层再分解，获得一组代表不同方向能量的各个部分，每部分可以认为是一个特定尺寸和方向的方向性滤波器的响应。结果是一个图像细节化的分析，称为带方向的金字塔(图 9.13)。

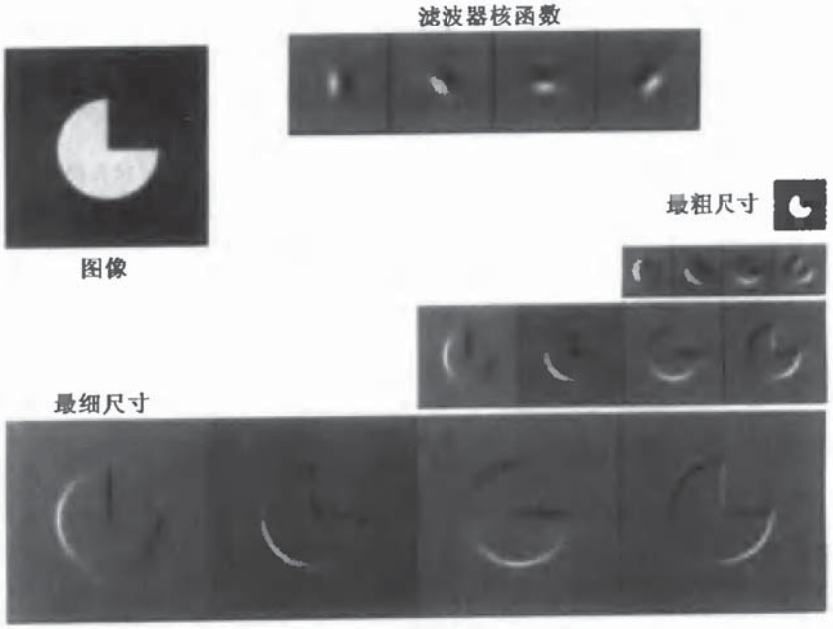


图9.13 一个带方向的金字塔，它由左上方的图像形成，每层有 4 个方向。首先把图像分为描述不同频带的子带(如同拉普拉斯金字塔)，然后应用方向性滤波器(右上图)将这些子带分解成一组不同的图像，每个表示图像中一个特定尺寸和方向的能量。注意方向性层对特定方向的边缘响应强烈，面对其他方向响应很弱。Eero Simoncelli编写的建立带方向金字塔的代码可以在http://www.cis.upenn.edu/~eero/steerpyr.html找到

深入讨论带方向的金字塔的设计超出了本书的范围。设计的第一个约束条件是滤波器应该选择一个小范围内的频率和方向，就像图 9.9 所示。设计的第二个约束条件是对滤波器的分析:应该容易合成。如果把带方向的金字塔认为是拉普拉斯金字塔的一个分解(见图 9.14)，那么合成包括重建拉普拉斯金字塔的每一层，然后由拉普拉斯金字塔合成图像。这个理想的策略必须包括一组滤波器，既有方向性响应，又容易合成。构造这样一组滤波器是有可能使得从分解的部分重建一层的过程，包括将图像再次用同样的滤波器滤波一次的过程(如图 9.15所建议的)。这些金字塔的有效实现可以在 http://www.cis.upenn.edu/~eero/steerpyr.html找到。设计过程的细节描述参见 Karasaridis 和 Simoncelli(1996)以及 Simoncelli 和 Freeman(1995)。

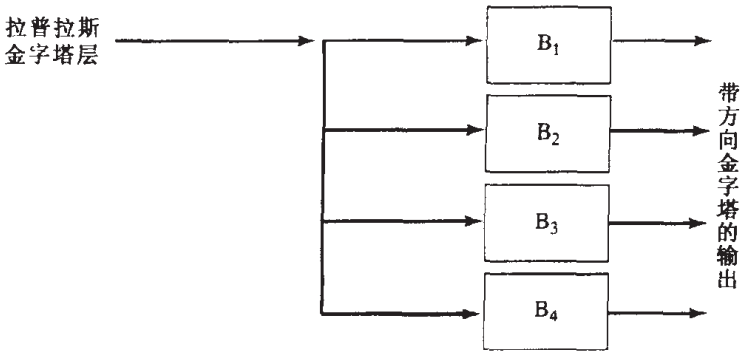


图 9.14 带方向的金字塔由拉普拉斯金字塔的每一层进行方向性滤波得到(由上面的方框示意性的画出)。拉普拉斯金字塔的每一层描述一定范围内的频域,方向性滤波器将这些范围的频域分解为一组方向因子

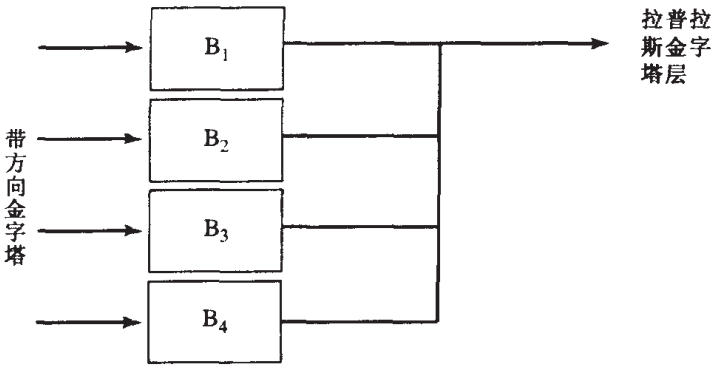


图 9.15 带方向的金字塔,可以用再滤波并相加结果的方式合成图像,正如上面示意所指出的)。这个特性可由适当的选择滤波器得到

9.3 应用:合成纹理来绘制

如果物体模型是带有纹理的,那么绘制使其看起来更加真实(这里值得思考的是为什么它应该是真实的,尽管这一点已被广泛接受)。有很多种纹理映射的技术;基本的想法就是当绘制一个物体时,用来产生影调的反射系数值,可以利用纹理图得到。对物体的表面采用某种坐标系统以便将纹理图的元素和表面上的点联系起来。选择不同的坐标系统会导致绘制效果截然不同,而且让纹理自然地贴在表面上也不容易(例如,考虑在斑马上画条纹,怎样画这些条纹才比较自然)。尽管如此,纹理映射看起来是使重现的场景看上去更真实的一种重要技巧。

纹理映射需要纹理,而映射一个很大物体的纹理需要大量的纹理图。如果希望物体接近于实际可见的景物,表面上的纹理要有高的分辨率,所以纹理图分辨率的问题变得非常明显。分片贴纹理图像可能效果很差,因为很难得到贴得很好的图像——边缘要排列整齐,而且即使这样,周期性结构的形成也令人烦恼。当然可以从很多来源中购买图像的纹理,但是理想的做法是编制一个程序,从一个很小的示例图像产生很大的纹理图像。可以得到这种方式的非常复杂的程序,它们显示了用滤波器输出描述纹理的实用价值。

9.3.1 均匀性

纹理合成的一般性策略是把纹理作为某种概率分布的样本，然后尝试由同样的分布得到其他的样本。为了使这种方法有实用性，需要通过样本纹理得到一个概率模型。第一件要做的事情就是假定纹理是均匀的，这意味着纹理的局部窗口无论是在哪里得到的，看起来都“相同”。说得更规范些，一个像素值的概率分布是由它邻域内的像素性质决定的，而不是由那个像素的位置决定的。

均匀性的假设说明我们可以根据样本区域的性质，为样本区域外的纹理建立一个模型。这个假设经常应用于允许尺寸适当变化的自然纹理中。例如，斑马背的条纹是均匀性的，但是需要记住背上的条纹是垂直的，但是腿上的是水平的。使用样本纹理来获得合成纹理的概率模型有很多方法，这里我们仅描述一种。

9.3.2 用局部模型抽样合成

正如 Efros 和 Leung(1999)指出的，样本图像可以用来作为概率模型。我们暂且假设除了一个像素外，合成图像的每一个像素已知。为了得到该像素值的概率模型，可以将这个像素的邻域和样本图像进行匹配。样本图像中的每一个进行匹配的邻域都有一个概率对应于我们感兴趣的像素，这些概率集合是一个感兴趣像素的概率分布直方图。从这个集合随机的和均匀的提取样本，可以得到和样本图像一致的值。如图 9.16 所示。

找到图像邻域的匹配 问题的本质就是从围绕感兴趣的像素取出几种形式的邻域，然后把它们和样本图像的邻域比较。领域的尺寸和形状是很重要的，因为它决定像素可以彼此直接影响的范围（见图 9.17）。Efros 使用了一个正方形的邻域，其中心是我们感兴趣的像素。

两个图像邻域的相似度可以用相应的像素的平方差之和来衡量。当邻域很相似时，这个值很小，而当它们明显不同时这个值很大(它本质上是两个向量之差的模)。当然，要合成的像素的值并未计算在平方差的和之内。

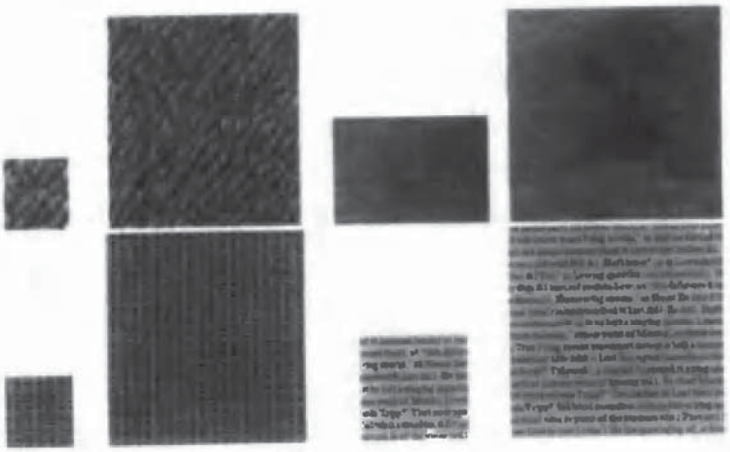


图 9.16 Efros 纹理合成算法(算法 9.3)将要合成的图像的邻域和样本图像进行比较，然后从匹配邻域后得到的概率值中随机地选择。这说明算法可以再现复杂的结构，正如这些例子所示。左边的小块是样本纹理，算法合成了右边的那块图像。注意合成的文本看起来仍然像文本，它看起来像不同长度的词如同文本一样地放置。每个词看起来像由字母组成的(虽然当人们靠近看的时候会发现这个解释不对)

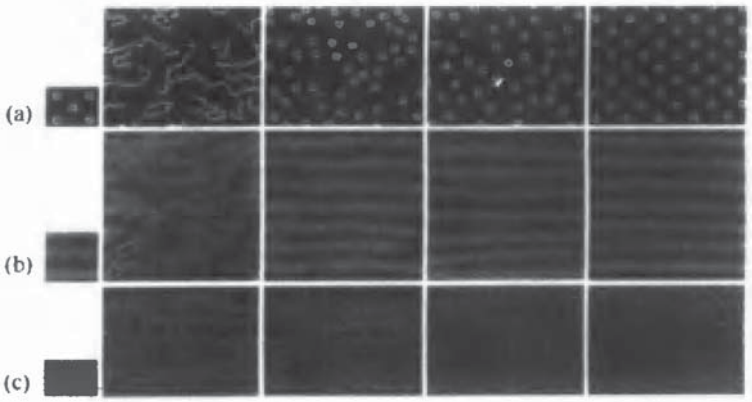


图 9.17 要匹配的图像邻域的尺寸不同使得算法 9.3 的结果明显不同。在图像中，右边的纹理利用左边的小方块合成，从左到右使用的邻域尺寸递增。如果用很小的邻域，那么算法不易再现大尺度的纹理效果。例如，对于斑点密布的纹理，如果邻域太小，就没有办法得到点的结构(所以只看到曲线单元)，算法合成由曲线片断组成的纹理。当邻域变大，算法可以得到点的结构，但并不是均匀的间距。当邻域很大时，才得到均匀间距的效果

利用邻域合成纹理 现在我们已了解如何得到一个缺失的单独像素的纹理值：均匀和随机的选择样本图像的像素值，样本图像的邻域应和我们的像素邻域匹配(也就是说，那个邻域的平方差之和小于某个阀值)。

一般，我们需要合成不止一个像素。通常，要合成的像素的邻域中一些像素的值是并不知道的——这些像素也需要合成。得到一个感兴趣的像素集合的一种方法是，在计算平方差之和时只是用那些已知的值，然后按比例判断阈值。合成过程可以从样本图像中随机选择一块像素开始，如算法 9.3 所示。

算法 9.3 非参数的纹理合成
从样本图像中随机选择一小方块像素
将这个小方块的值插入到要合成的图像中
直到要合成的图像的每一个地方都有一个值
  对于合成后的块边界上的每一个没有合成的地方
    用样本图像匹配该地方的邻域，计算匹配值时忽略没有合成的地方
    从那组匹配的邻域中响应地方的值中均匀和随机的选择一个值
  end
end

9.4 由纹理得到形状

同样一片纹理从正面与从切向角看起来很不一样，因为透视缩小效应会导致纹理元素(包括它们之间的间距)在某个方向比其他方向收缩得更严重。这使我们想到如果提供一个纹理模型，便可以从纹理中恢复出一些形状信息，人类具有这种能力(见图 9.18)。引人注目的是，

很多纹理模型提供了足够的信息来推断形状。这对于平面是非常简单的(9.4.1 节),而在曲面的情况一些细节仍不清楚,但一般性的问题是共同的。

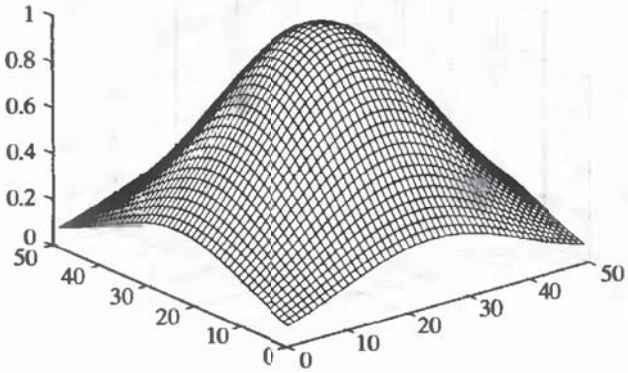


图 9.18 人们由表面纹理的外观得到表面的形状信息。左边的图像显示了这个结论,通常,除了轮廓区域,描述表面的惟一信息来源就是表面纹理的扭曲。右边灌木丛的纹理使人们产生由它们形成的圆形表面的感觉

9.4.1 由纹理得到的平面形状

如果我们知道我们在注视着一个平面,由纹理得到形状就转化为决定平面相对照相机的关系。如果我们假定一个平面的方位,则可以把图像纹理映射回平面。如果我们有一些纹理的一致性的模型,则可以通过测试反投影纹理的这个性质,得到有最佳反投影纹理效果的平面。这种一般的策略对很多纹理模型适用。我们将限制讨论在正交投影的情况。如果照相机不是正交投影的话,我们仍然可以继续讨论,不过需要多一些实际的工作和符号。我们会注意讨论其他情况。

表示一个平面 现在假设我们在正交投影的照相机注视着一个有纹理的平面,因为照相机是正交投影的,所以没有办法测量平面的深度。但是,可以考虑平面的方向。我们根据照相机坐标系统来讨论。首先需要知道有纹理的平面的法线和视觉方向之间的角度——有时称为倾斜角。然后,投影法线在照相机坐标系统的角度——有时称为俯仰角(见图 9.19)。在一张平面的图像中,有一个倾斜方向——在平行于投影法线的平面上的方向。

各向同性假设 一个各向同性纹理就是遇到该纹理因子的概率不依赖于该因子的方向。这意味着一个各向同性纹理的概率模型不需要依赖纹理平面坐标系统的方向。

如果我们假设纹理是各向同性的,则倾斜角和俯仰角都可以从图像中得到。我们可以合成一个有纹理的平面的正交投影视图,首先绕俯仰角旋转坐标视图,然后用倾斜角的余弦在某个坐标方向缩短——这个过程称为视角变换。理解这一点的最简单的方法就是假设纹理由一组在平面上散开的圆组成。在一个正交投影的视图中,这些圆会映射为椭圆,短轴会给出俯仰角,高宽比会给出倾斜角(见习题和图 9.19)。

一个各向同性的纹理的正投影视图并不是各向同性的(除非这个平面和图像平面是平行的),这是因为在倾斜角方向的收缩干扰了纹理的各向同性。指向收缩方向的因子会变得短一些。此外,因子如果有指向收缩方向的部分,这部分也会变得短些。与视角变换对应的是逆视

角变换(给出倾斜角和俯仰角,把一个图像平面变成一个物体的平面纹理)。它引出了一个检测平面方向的策略:找到一个逆视角变换,将图像纹理变成一个各向同性的纹理,然后由逆视角变换恢复倾斜角和俯仰角。

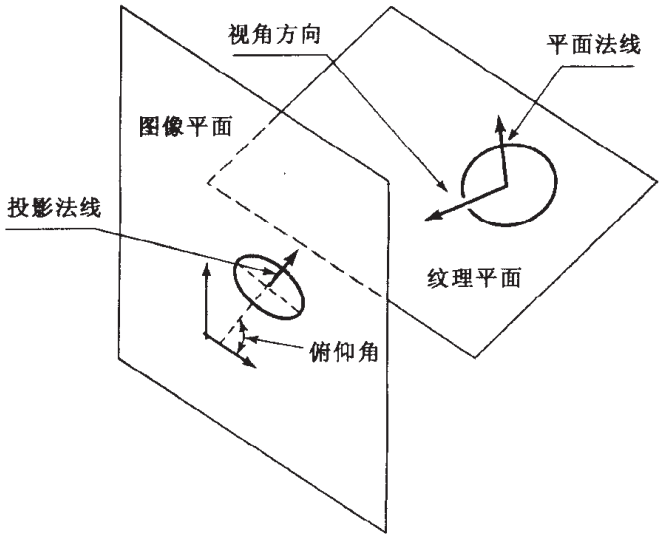


图 9.19 平面相对于照相机平面的方向可以由倾斜角给出——有纹理的平面的法线和视线方向的夹角——还有俯仰角——投影法线在照相机坐标系统的角度。该图解释了俯仰角,然后显示了一个圆投影为椭圆

有很多种方法可以找到这种视角变换。一个很自然的策略就是使用一组方向性滤波器的输出能量。这是经过平方后的响应,在图像上各处加起来。对于一个各向同性的纹理,可以认为对于任何给定的尺寸在任一方向能量输出都一样,因为该纹理遇到模式的概率不依赖于它的方向。因此,一个各向同性的度量是输出能量的标准差与方向的函数关系。我们可以把在不同尺寸上的这个度量加起来,也许用对应尺寸的能量将这个度量加权。该度量越小,纹理的各向同性越好。我们可以利用这个度量找到使图像看起来更加各向同性的逆视角变换,利用标准的优化方法。

注意这个方法可以立即扩展到透视映射、球形映射和其他视角变换。简单地,我们只需从很多变换中找到使图像纹理看起来更加各向同性的变换。其次,还要小心仔细。例如,改变一个各向同性的纹理的尺寸会导致另一个各向同性的纹理,这意味着不可能恢复出一个缩放比例的参数,尝试恢复也不是一个好想法。用各向同性的假设恢复一个平面方向的最主要问题是,世界上各向同性的纹理非常少。

均匀性假设 在一个正交投影的视图中利用纹理的均匀性(9.3.1 节的定义)假设恢复出平面的方向是不可能的,这是因为视角变换将一个均匀性的纹理变成另一个均匀性的纹理。然而,如果假设视图是透视的,这将使上述成为可能。

要理解这一点首先需注意,均匀性意味着,如果将一个很大的均匀网格加在一个平面上,那么在每个小网格中所出现的事件数目应该是(近似)相同的。例如,如果一个纹理由一个均匀性模型的很多点组成,每个正方形期望的点的数量对于每个正方形应该是一样的。然而,如果看到一个有纹理的面叠加有一个网格的图是透视视图,那么一些网格元素就会映射成大的四边形,其他就要映射成很小的四边形(除非视图是正视的)。这又意味着,映射的纹理在图像平面不可能是均匀的——因为在图像平面的某些网格的四边形中会有很多纹理事件,而其他

地方就会很少。例如一个很多点的平面,靠近平面地平线的点的投影变小了。合适的策略是选择一个变换使图像平面的纹理更加均匀,注意我们可以利用照相机平面来决定平面的方向,但是得不到它的深度,因为一个均匀性纹理的正视图也是均匀性的——所有的纹理事件均按同样尺度缩放。

9.5 注释

我们将有关纹理的文章压缩在这一章。多年来,出现了许多描述图像纹理的技术,代表性的做法是研讨模式彼此间相对关系的统计量。争论的问题是如何描述一个模式、考虑什么样的统计量。尽管认为用线性滤波器来描述模型的方法是正确的结论略显过早,这种想法现在是占优势的,主要因为它很容易用这个策略来解决问题。对纹理非常感兴趣的读者可能抱怨我们省略了关于马尔可夫随机模型。这是由于建立这种模型需要大量数学模型,以及对 MRF 还缺少满意的推理算法。感兴趣的读者可查阅 Chellappa 和 Jain(1993),Cross 和 Jain(1983),Manjunath 和 Chellappa(1991),或者 Speis 和 Healey(1996)。

另一个重要的省略就是对于用小波方法描述纹理的讨论。尽管这个方法也大体遵循本章讨论的原则——用多个滤波器的输出描述纹理——但这些滤波器后面蕴涵着一个复杂的理论。请有兴趣的读者查阅 Ma 和 Manjunath(1995),(1996)或者 Manjunath 和 Ma(1996*b*,*c*)。

滤波器、金字塔和效率

如果想用一个很大范围的滤波器(它们有很多尺寸和方向)的输出来描述纹理,则需要保证滤波时的效率。这是一个引起很多关注的话题;一般的方法就是尝试建立一个张量积基,以便很好地描述能使用的滤波器族。用一种合适的构造方法,需要把图像和少量可分离的核卷积,然后把很多滤波器的响应结果用不同的方法综合起来进行评估(因此需要基是张量积)。重要的论文包括 Freeman 和 Adelson(1991);Greenspan, Belongie, Perona, Goodman, Rakshit 和 Anderson(1994);Hel-Or 和 Teo(1996);Perona(1992),(1995);Simoncelli 和 Farid(1995);Simoncelli 和 Freeman(1995)。

纹理合成

要透彻讨论纹理合成方法会耗费大量的时间。除了 MRF 之外,这里未涉及的最重要的方法是 Zhu, Wu 和 Mumford(1998)的工作,这种方法使用复杂的熵判据,首先选择描述纹理的滤波器,然后建立纹理的概率模型。

由纹理恢复形状

只有很少的方法可以由一个表面的纹理投影来恢复一个表面模型。全局方法尝试用关于纹理元素分布的假设恢复一个完整的表面模型。合适的假设是各向同性(Witkin,1981)(这个方法的不足在于很少有各向同性的自然纹理)或者均匀性(Aloimonos,1986;Blake 和 Marinos,1990)。建立在均匀性的这些方法都假设,纹理元素是均匀的泊松点过程在平面上作用的结果;纹理元素中心的密度梯度得出平面的参数。然而,单个纹理元素的变形没有考虑。

局部方法恢复了表面上的点的一些微分几何参数(典型地,法线和曲率)。这类方法起源于 Garding(1992),已由 Malik 和 Rosenholtz(1997),Rosenholtz 和 Malik(1997)在各种各样的表面

成功地加以演示。Clerc 和 Mallat(1999)用小波的方法运作。这个方法有一个至关重要的缺点,它或者需要知道平行于被讨论点的帧场的纹理元素坐标帧,或者需要知道帧场的微小旋转(这一点见 Garding,1995,并由 Rosenholtz 和 Malik,1997 选取用来展示的纹理证实,该假设被称为纹理稳定性)。例如,如果有人想用这些方法来恢复一个撒着巧克力颗粒的油炸圈饼的曲率,这需要确保颗粒都是平行于表面的(或者颗粒间的角度是已知的)。所以,可以论证这个方法只能在很小一部分集合的有纹理的表面上有用。另一个重要的难点就是恢复的数据,这些方法估计的是局部的法线和曲率。但是曲率是法线的派生物,所以一个局部的估计也许有用,但是没有理由相信一组局部的估计会相容。这是一个可积分性的问题。表面插补的方法在很大程度上已经被计算机视觉所抛弃,因为对于那些缺乏数据区域表面的语义状态是不确定的。在由纹理得到形状的计算中,插补技术无疑是一个很重要的规则——它表示了人们这样一种设想,即表面是在很慢变化的。不完整的局部测量到的表面法线会相互约束,从而获取在某些点法线较好的全局估计。

习题

9.1 显示一个圆,它在一个正交投影视图中看起来像一个椭圆,椭圆的短轴是俯仰角方向。那么椭圆的缩放比例是多少?

9.2 我们将在已给出由均匀泊松分布过程产生的点纹理的条件下研究在一个正交投影视图中测量平面方向的方法。回想按照这种过程产生点的一个方法是:依照对点的 x 和 y 坐标进行均匀和随机的抽样来处理。假设我们处理的点都在一个单位正方形中。

(a) 证明一个点在一个特定集合的概率和该集合的面积成比例。

(b) 假设我们把区域分成不同的集合。证明每个集合中点的数目有一个多项式的概率分布。我们现在使用这些观察来恢复平面的方向。我们将图像的纹理分成不同集合的聚集。

(c) 每个集合逆向投影到纹理平面的面积,是否为平面方向的一个函数?

(d) 是否可能用这个信息决定平面的方向?可利用(c)的结果。

编程作业

9.3 **纹理合成**:实现 9.3.2 节的非参数纹理合成算法,使用你的实现来研究:

(a) 窗口尺寸对合成纹理的影响。

(b) 窗口形状对合成纹理的影响。

(c) 匹配判据对合成纹理的影响(比如说,使用平方后的加权和取代平方和)。

9.4 **纹理表示**:实现一个纹理分类器可以区分至少 6 种纹理,使用 9.1.2 节中的尺寸选择机制,然后计算滤波器输出的统计量。推荐使用至少 6 个方向性条形滤波器和一个点滤波器的输出结果的平均值和协方差。也许需要研读第 22 章的分类,使用一个简单的分类器(其中的诀窍是最近邻使用马氏距离)。

第三部分　低层视觉:使用多幅图像

- 第 10 章　多视角几何学
- 第 11 章　立体视觉
- 第 12 章　从运动估计仿射模型
- 第 13 章　从运动估计投影模型

第 10 章　多视角几何学

尽管图像中包含着丰富的信息，但是仅从一幅图像无法直接获得沿着某条投影线上的点的深度，至少需要两幅图像，才可以通过三角测量的方法得到点的深度。当然，这也是大多数动物拥有至少两只眼睛，并在寻找同伴或天敌的过程中移动头部的原因之一。基于同样的原因，需要为自动机器人装备立体视觉或运动分析系统。在建立这样一个系统之前，我们必须了解同一场景中多个视角的三维约束以及相关的摄像机配置，这就是本章的目标。特别地，我们将阐述同一场景中 2 幅、3 幅或多幅不同视角图像之间的几何和代数关系。在双目立体视觉系统中，我们知道任何一点的第一个像点必定位于它的第二个像点和两个摄像机的光学中心所形成的平面上。在摄像机的内部参数和基本矩阵已知的情况下，这种外极约束可以用代数方法描述为一个 3×3 矩阵，称为本征矩阵。同一条直线的 3 个像之间，存在另外一种约束关系——即它们与原像所形成平面的交会退化（形成一条直线——译者注）。这种几何关系可以用代数方法描述为一个 $3\times3\times3$ 三焦张量。更多的像会引入更多的约束关系，例如，同一点的 4 个像点满足某种四线性关系，其中的系数要使用四焦张量表示。明显地，不需要任何摄像机和场景的信息，就能够建立同一场景特征的多个像所满足的方程。本章将介绍一些直接根据成像数据估计它们的参数的方法。

计算机视觉作为一个科学领域，不仅仅关注多视角几何学。在第 3 章中已经提到，分析摄影地形测量法的目标就是根据多张图片准确地获得定量的几何信息。本章将通过一些例子简要讨论应用外极和三焦约束解决经典分析摄影地形测量法的转换问题（例如，通过在参考图像中某个点的位置，来推测该点在其他图像中的位置）。立体视觉和运动分析领域的更多应用，将在以后的章节进行介绍。

10.1　双视角

10.1.1　外极几何

考虑一点 P，通过光学中心点分别在 O 和 O' 的两个摄像机所成的像为 p 和 p'。这 5 个点都位于 2 条相交光线 OP 和 $O'P$（见图 10.1）所形成的外极平面上。特别地，点 p' 位于该平面与第二台摄像机的视平面 Π' 的交线 l' 上。直线 l' 是与点 p 相关联的外极线，它经过点 e'。e' 是连接两个光学中心 O 和 O' 的基线和平面 Π' 的交点。同样，点 p 位于与点 p' 关联的直线 l 上，且该直线经过基线和平面 Π 的交点 e。

点 e 和点 e' 称为两个摄像机的外极点，外极点 e' 是第二个摄像机观察到的图像中第一个摄像机的光学中心 O 的投影，反之亦然。如上文所述，如果 p 和 p' 是同一个点的不同像点，那么 p' 一定位于与 p 相关联的外极线上。这种外极线约束是立体视觉和运动分析中的基本原理。

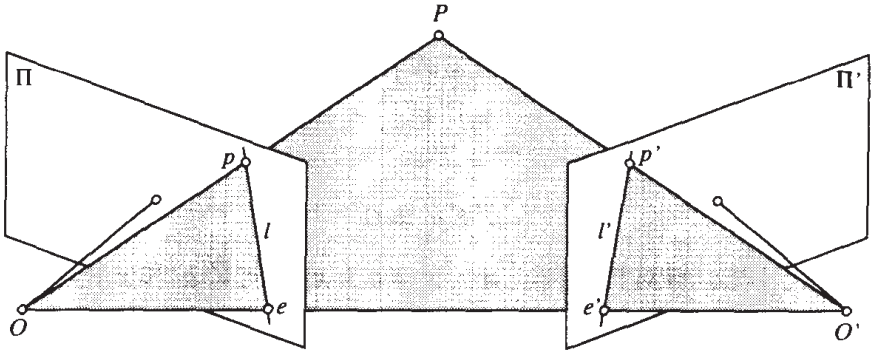


图 10.1　外极几何：点 P，两台摄像机的光学中心 O 和 O'，点 P 的两个像点 p 和 p' 都位于同一平面上。为了简化绘图，这里和本章的其他图例一样，摄像机用针孔来代表，一个虚拟成像平面置于针孔的前方。对于实际的在针孔后的物理成像平面，本章后面使用的几何与代数的论述同样成立

假定已知立体视觉中两个摄像机的内部和外部参数。如第 11 章所述，在立体视觉分析中，最困难的部分是建立两幅图像的对应关系（也就是说，确定第二幅图像中的哪些点与第一幅中的点匹配）。外极线约束在很大程度上限制了寻找这种对应关系的搜索范围：事实上，既然我们假定该设备是标定过的，那么点 p 的坐标完全决定了连接 O 和 p 的光线，并由此决定了相关联的外极平面 $OO'p$ 和外极线 l'。对匹配的搜索能够被限定在这条直线上，而不是限定在整幅图像上（见图 10.2）。在两帧运动分析中，每个摄像机的内部参数可能已标定，但是两台摄像机坐标系统间的刚性变换并不知道。在这种情况下，外极几何明显约束了移动的可能范围。下面的几节研究了这种情况下的一些变化。

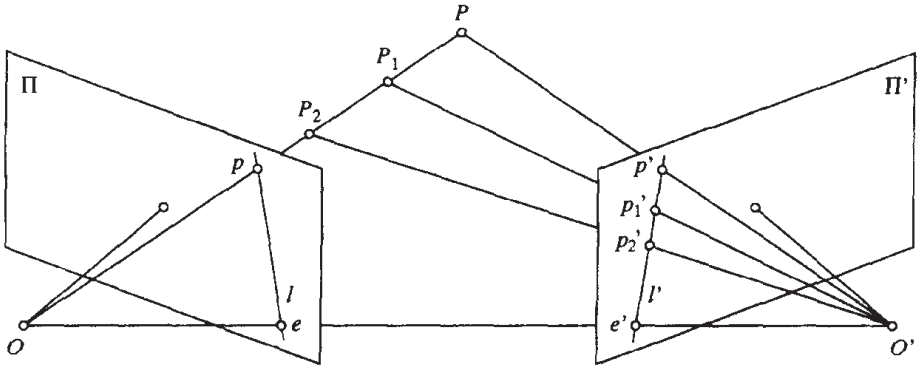


图 10.2　外极线约束：已有一个标定过的立体视觉系统，可能与点 p 相匹配的点限定于相关联的外极线 l' 上

10.1.2　标定的情况

这里我们假定每台摄像机的内部参数已知，所以 $\boldsymbol{p}=\hat{\boldsymbol{p}}$。显然，外极线约束说明了三个向量 $\overrightarrow{Op}$，$\overrightarrow{O'p'}$ 和 $\overrightarrow{OO'}$ 共面。等价地，其中一个向量在其他两个向量所在的平面上，即

$$\overrightarrow{Op}\cdot[\overrightarrow{OO'}\times\overrightarrow{O'p'}]=0$$

我们可以使用与第一台摄像机相关联的坐标系，将这个与坐标无关的方程改写为

$$\boldsymbol{p}\cdot[\boldsymbol{t}\times(\mathcal{R}\boldsymbol{p}')] \tag{10.1}$$

其中，$\boldsymbol{p}=(u,v,1)^T$ 且 $\boldsymbol{p}'=(u',v',1)^T$，表示 p 和 p' 的齐次图像坐标向量。$\boldsymbol{t}$ 是区分两个坐标

系的坐标平移向量$\overrightarrow{OO'}$。$\mathcal{R}$是旋转矩阵，即在第二个坐标系统中坐标为 $\boldsymbol{w}'$的自由向量在第一个坐标系中的坐标为$\mathcal{R}\boldsymbol{w}'$。这样，两个投影矩阵在第一台摄像机所对应的坐标系的表示为$(\mathrm{Id}\quad \mathbf{0})$和$(\mathcal{R}^T\quad -\mathcal{R}^T\boldsymbol{t})$。

方程(10.1)最后可以改写为

$$\boldsymbol{p}^T\mathcal{E}\boldsymbol{p}'=0 \tag{10.2}$$

其中，$\mathcal{E}=[\boldsymbol{t}_\times]\mathcal{R}$，$[\boldsymbol{a}_\times]$表示斜对称矩阵，$[\boldsymbol{a}_\times]\boldsymbol{x}=\boldsymbol{a}\times\boldsymbol{x}$ 是向量$\boldsymbol{a}$ 和$\boldsymbol{x}$ 的叉积。矩阵$\mathcal{E}$称为本征矩阵，Longuet-Higgins(1981)首先引入这一概念。它的9个系数只是定义到比例关系，并且可以通过旋转矩阵$\mathcal{R}$的3个自由度和决定平移向量 $\boldsymbol{t}$ 的方向的2个自由度来参数化。

$\mathcal{E}\boldsymbol{p}'$可以解释为第一幅图像中与点 p'所关联的外极线的坐标向量：事实上，直线 l 可以通过方程$au+bv+c=0$来定义，其中(u,v)表示该直线上一点的坐标。(a,b)为该直线的单位法向量，$-c$ 是原点和l 之间(带符号的)距离。另外我们也可以根据直线上一点的齐次坐标向量 $\boldsymbol{p}=(u,v,1)^T$ 和向量$\boldsymbol{l}=(a,b,c)^T$，将直线方程定义为 $\boldsymbol{l}\cdot\boldsymbol{p}=0$。在这种情况下不再需要限制条件 $a^2+b^2=1$，因为方程中 $\boldsymbol{l}$ 可以独立进行尺度变换。这样，方程(10.2)描述了点 p 位于与向量$\mathcal{E}\boldsymbol{p}'$关联的外极线上。对称地，坐标向量$\mathcal{E}^T\boldsymbol{p}$ 代表了在第二幅图像中与点 p 关联的外极线。显然，本征矩阵是奇异的，因为 $\boldsymbol{t}$ 与第一个外极坐标系中的坐标向量$\boldsymbol{e}$ 平行，所以有$\mathcal{E}^T\boldsymbol{e}=-\mathcal{R}^T[\boldsymbol{t}_\times]\boldsymbol{e}=0$。同样，很容易得到 $\boldsymbol{e}'$在$\mathcal{E}$的零空间中。事实上，正如 Huang 和 Faugeras(1989)指出的，本征矩阵是奇异的，并有两个相等的非零奇异值(见习题)。

10.1.3 微小运动

现在让我们把注意力转移到无穷小的位置变化上。我们考虑一台运动的摄像机，平移线速度为 $\boldsymbol{v}$，旋转角速度为$\boldsymbol{\omega}$。令 δt 表示像点 p 的速度或称运动场。对间隔微小时间 $\dot{\boldsymbol{p}}=(\dot{u},\dot{v},0)^T$ 的两帧，改写方程(10.2)，使用旋转的指数形式(见习题)，可以有如下关系(一阶)

$$\begin{cases}\boldsymbol{t}=\delta t\,\boldsymbol{v}\\ \mathcal{R}=\mathrm{Id}+\delta t\,[\boldsymbol{\omega}_\times]\\ \boldsymbol{p}'=\boldsymbol{p}+\delta t\,\dot{\boldsymbol{p}}\end{cases} \tag{10.3}$$

代入方程(10.2)，忽略 δt 的所有二阶和高于二阶的项，得到：

$$\boldsymbol{p}^T([\boldsymbol{v}_\times][\boldsymbol{\omega}_\times])\boldsymbol{p}-(\boldsymbol{p}\times\dot{\boldsymbol{p}})\cdot\boldsymbol{v}=0 \tag{10.4}$$

Longuet-Higgins 关系(10.2)表现了离散情况下的外极几何，而方程(10.4)只是它的瞬时形式。在只有平移的情况下，有$\boldsymbol{\omega}=0$，所以$(\boldsymbol{p}\times\dot{\boldsymbol{p}})\cdot\boldsymbol{v}=0$。就是说，三个向量 $\boldsymbol{p}=\overrightarrow{op}$，$\dot{\boldsymbol{p}}$ 和 $\boldsymbol{v}$ 一定共面。如果用 e 表示无穷小的外极点或膨胀中心(也就是说，经过光学中心并与速度向量 $\boldsymbol{v}$ 平行的直线，和成像平面的交点)，我们得到众所周知的结果，在只有平移时，运动场指向膨胀中心(见图 10.3)。

10.1.4 非标定的情况

Longuet-Higgins 关系适用于内部标定过的摄像机。当内部参数未知时(非标定的摄像机)，可以写成 $\boldsymbol{p}=\mathcal{K}\hat{\boldsymbol{p}}$ 和$\boldsymbol{p}'=\mathcal{K}'\hat{\boldsymbol{p}}'$。其中，$\mathcal{K}$和$\mathcal{K}'$是 3×3 标定矩阵，$\hat{\boldsymbol{p}}$ 和$\hat{\boldsymbol{p}}'$是规范化的像点的坐标向量。Longuet-Higgins 关系给出这些向量之间的关系：

$$\boldsymbol{p}^T\mathcal{F}\boldsymbol{p}' = 0 \tag{10.5}$$

其中，矩阵$\mathcal{F} = \mathcal{K}^{-T}\mathcal{E}\mathcal{K}'^{-1}$称为基础矩阵。通常，它不是本征矩阵，它的秩同样为 2。和前面一样，$\mathcal{F}(\mathcal{F}^T)$的与 0 特征值对应的特征向量是外极线上的点 $\boldsymbol{e}'(\boldsymbol{e})$。注意，$\mathcal{F}\boldsymbol{p}'(\mathcal{F}^T\boldsymbol{p})$代表了在第一(二)个像中与点 $\boldsymbol{p}'(\boldsymbol{p})$对应的外极线。

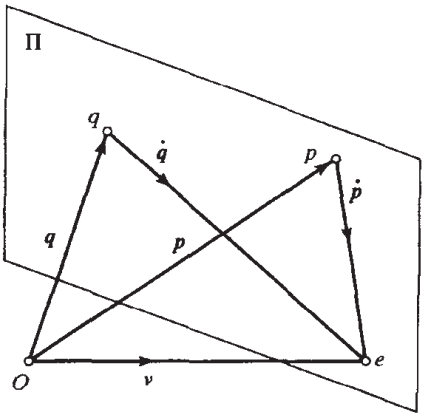


图 10.3　膨胀中心：在纯平移运动中，图像上每个点的运动场指向膨胀中心

秩为 2 这个限制意味着基础矩阵只允许有 7 个独立的参数。对基础矩阵的参数化有多种方法，最直接的方式是通过两个外极点的坐标向量 $\boldsymbol{e} = (\alpha, \beta)^T$，$\boldsymbol{e}' = (\alpha', \beta')^T$ 利用外极变换，即一组外极线映射到另一组的变换，对基础矩阵进行参数化。在第 13 章中，我们将在运动结构的讨论中考察这种变换的属性。目前我们只需知道(不需证明)，可以通过 4 个数 a、b、c、d 来决定参数(在一个尺度上)。基础矩阵可以写为

$$\mathcal{F} = \begin{pmatrix} b & a & -a\beta - b\alpha \\ -d & -c & c\beta + d\alpha \\ d\beta' - b\alpha' & c\beta' - a\alpha' & -c\beta\beta' - d\beta'\alpha + a\beta\alpha' + b\alpha\alpha' \end{pmatrix} \tag{10.6}$$

10.1.5　弱标定

如前面提到的，本征矩阵在一个尺度上可以通过 5 个独立的参数定义，因此本征矩阵（至少在理论上）可以通过式(10.2)用 5 个对应点来计算。相似地，基础矩阵通过 7 个独立的参数进行定义(式 10.6 中的参数 a, b, c, d 仅仅在一个尺度上定义)，原则上也可以通过 7 个对应点来估计。通过最少的参数估计本征矩阵和基础矩阵的方法确实存在(见本章“注释”一节)。但是它们过于复杂，不在这里进行讨论。这一节讨论更为简单的问题：对于内部参数未知的摄像机，通过两幅图中的冗余对应点集合来估计外极几何——这一过程称为弱标定。

注意到式(10.5)对于基础矩阵$\mathcal{F}$的 9 个系数是线性的：

$$(u, v, 1)\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \tag{10.7}$$

由于这个方程对于$\mathcal{F}$的系数是齐次的，我们可以设定 $F_{33} = 1$，并使用 8 个对应点 $p_i \leftrightarrow p_i'$ $(i = 1, \cdots, 8)$将方程(10.7)改写为一个 8×8 非齐次线性方程组：

$$\begin{pmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' \\ u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' \\ u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' \\ u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' \\ u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' \\ u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' \\ u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8' \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

使用这个方程组来估计基础矩阵得到的 8 点算法，最初由 Longuet-Higgins(1981)在标定摄像机的情况下提出。当相应的 8×8 矩阵为奇异时，它就失效了。像 Faugeras(1993)以及习题中所示的，这种情况只会在 8 个点和 2 个光学中心位于同一个二次曲面的情况下发生。不过这种情况发生的可能性很小，因为 9 个点就可完全确定一个二次曲面，这意味着经过任意 10 个点的二次曲面一般是不存在的(也就是说，如果相应的 8×8 矩阵是奇异的，往往改换一组对应点就可以了)。

由于以$\mathcal{F}$的系数形成的向量存在单位范数，如果存在 $n>8$ 个对应关系，可通过线性最小二乘方法，对$\mathcal{F}$的系数最小化下式：

$$\sum_{i=1}^{n}(\boldsymbol{p}_i^T\mathcal{F}\boldsymbol{p}_i')^2 \tag{10.8}$$

以此对$\mathcal{F}$进行估计。

注意到 8 点算法和利用最小二乘的方法都忽略了基础矩阵阶为 2 的特性①。为了利用这个约束，Luong 等(1993，1996)提出利用 8 点算法得到的矩阵$\mathcal{F}$作为两步估计过程的基础：首先，使用线性最小二乘方法最小化$|\mathcal{F}^T\boldsymbol{e}|^2$和$|\mathcal{F}\boldsymbol{e}'|^2$得到外极点 $\boldsymbol{e}$ 和$\boldsymbol{e}'$；然后，在式 (10.6)中代入这些点的坐标：通过外极变化的系数产生基础矩阵的一组线性参数，这样可以通过线性最小二乘方法最小化式(10.8)获得$\mathcal{F}$矩阵的估计。

采用最小二乘法的 8 点算法可以使与外极限制相关的代数距离的几何平均值最小(也就是说，所有对应点对误差函数 $e(\boldsymbol{p},\boldsymbol{p}')$的几何平均值 $=\boldsymbol{p}^T\mathcal{F}\boldsymbol{p}'$最小)。这个误差函数有其几何意义；特别地，我们有

$$e(\boldsymbol{p},\boldsymbol{p}') = \lambda \mathrm{d}(\boldsymbol{p},\mathcal{F}\boldsymbol{p}') = \lambda' \mathrm{d}(\boldsymbol{p}',\mathcal{F}^T\boldsymbol{p})$$

其中，$d(\boldsymbol{p},\boldsymbol{l})$表示点 $\boldsymbol{p}$ 和直线 $\boldsymbol{l}$(带符号)的欧氏距离，$\mathcal{F}\boldsymbol{p}$ 和$\mathcal{F}^T\boldsymbol{p}'$是与$\boldsymbol{p}$ 和$\boldsymbol{p}'$相应的外极线。比例因子 λ 和λ'是$\mathcal{F}\boldsymbol{p}'$和$\mathcal{F}^T\boldsymbol{p}$ 前两个分量组成向量的范数，他们对于观察数据点对的依赖导致了估计过程的偏差。

当然也可以去掉放缩因子，直接求得图像点与相应外极线之间的几何距离，再求出该距离的几何平均值的最小值——即，

$$\sum_{i=1}^{n}\left[\mathrm{d}^2(\boldsymbol{p}_i,\mathcal{F}\boldsymbol{p}_i') + \mathrm{d}^2(\boldsymbol{p}_i',\mathcal{F}^T\boldsymbol{p}_i)\right]$$

不管基础矩阵的参数化如何选择，它始终是一个非线性的问题。但是，我们可以使用 8 点算法的结果来初始化最小化方法。这个方法首先由 Luong 等(1993)提出，它可以得到比采用 8 点算

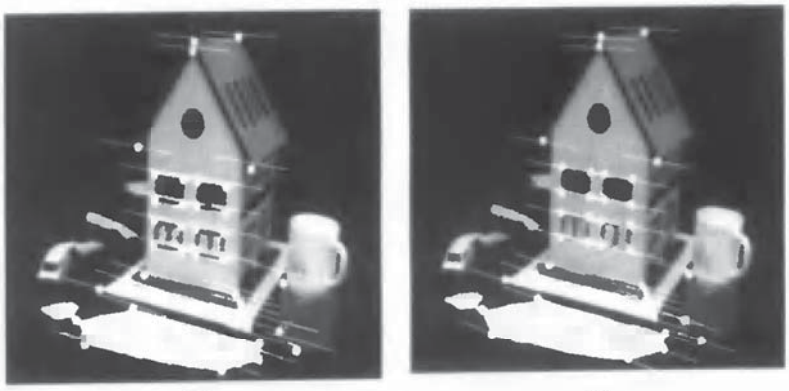
① 以前由 Longuet-Higgins 提出的算法忽视了特征矩阵秩为 2，并且有两个相等的奇异值。

法更好的结果。作为一个替代的方法，Hartley(1995)提出规范化线性 8 点算法。这个方法从下述观察出发：原有方法的不良性能通常是由于不良的数值条件$\sqrt{2}$造成的，提出通过平移和放缩使数据点集中在原点附近，并且到原点平均距离为像素。实践表明，这种规范化很好地改进了线性最小二乘估计方法的条件。具体来说，算法有 4 步：首先，通过适当的平移和放缩算子变换图像坐标：$\mathcal{T}:\boldsymbol{p}_i \rightarrow \tilde{\boldsymbol{p}}_i$ 和$\mathcal{T}':\boldsymbol{p}_i' \rightarrow \tilde{\boldsymbol{p}}_i'$。然后，使用最小二乘法计算矩阵$\tilde{\mathcal{F}}$，最小化

$$\sum_{i=1}^{n}(\tilde{\boldsymbol{p}}_i^T \tilde{\mathcal{F}} \tilde{\boldsymbol{p}}_i')^2$$

第三步，强化秩为 2 的限制；可以通过前面描述的 Luong 等的两步方法完成这件事情，但是 Hartley 使用了 Tsai 和 Huang(1984)在标定情况下提出的技术，它构造了$\tilde{\mathcal{F}}$的奇异值分解$\tilde{\mathcal{F}} = \mathcal{U}\mathcal{S}\mathcal{V}^T$。奇异值分解在第 12 章中正式定义，在这里只指出$\mathcal{S} = \mathrm{diag}(r, s, t)$是一个 3×3 的对角矩阵，且$r \geqslant s \geqslant t$，$\mathcal{U}$，$\mathcal{V}$是两个正交的 3×3 矩阵，如在第 12 章中所述，最小化$\tilde{\mathcal{F}} - \tilde{\mathcal{F}}$的 Frobenius 范数获得的秩为 2 的矩阵$\tilde{\mathcal{F}}$就是$\tilde{\mathcal{F}} = \mathcal{U}\ \mathrm{diag}(r, s, 0)\mathcal{V}^T$。算法最后一步，将$\mathcal{F} = \mathcal{T}^T \tilde{\mathcal{F}} \mathcal{T}'$作为基础矩阵的最终估计值。

图 10.4 显示了一个弱标定实验，这个实验使用了一个玩具房子的两幅图像上的 37 组对应点作为输入。数据点在图中用圆点表示，所经过的外极线用短的直线段表示。图 10.4(a)显示了使用最小二乘法的普通 8 点算法得到的输出结果，图 10.4(b)显示了使用 Hartley 变换后的该方法的输出结果。正如所期望的，第二种情况下的输出要好得多，事实上，它已经非常接近使用 Luong 等(1993，1996)提出的几何距离的标准结果。



(a)　　(b)

	最小二乘法	(Hartley，1995)	(Loung 等，1993)
平均距离	2.33像素	0.92像素	0.86像素

图 10.4　使用一个玩具房子的两幅图像上的 37 个点作为输入的弱标定实验。图中显示了外极线(a)使用最小二乘法的普通 8 点算法得到(b)使用 Hartley(1995)规范化后的该方法。注意，例如，在图(a)中靠近杯子底部的点误差要大一些。定量的比较见上面的表格，其中给出了两种方法，还包括Luong等(1993)的非线性算法中数据点之间的平均距离和相应的外极线

10.2 三视图

现在我们研究相同场景三个视图的几何约束，让我们回到 $\boldsymbol{p}=\hat{\boldsymbol{p}}$ 时的标定情形。想像一下三个透视摄像机观察同一个点 P，它的像点被记为 p_1，p_2，p_3（见图 10.5）。摄像机的镜头中心 O_1，O_2 和 O_3 确定了一个三焦平面，并且与它们的视平面相交于三条三焦直线 t_1，t_2，t_3。它们中每一条直线都经过相应的外极点（例如，和第二个摄像机对应的 t_2，通过另外两个摄像机的投影点 e_{12} 和 e_{32}）。

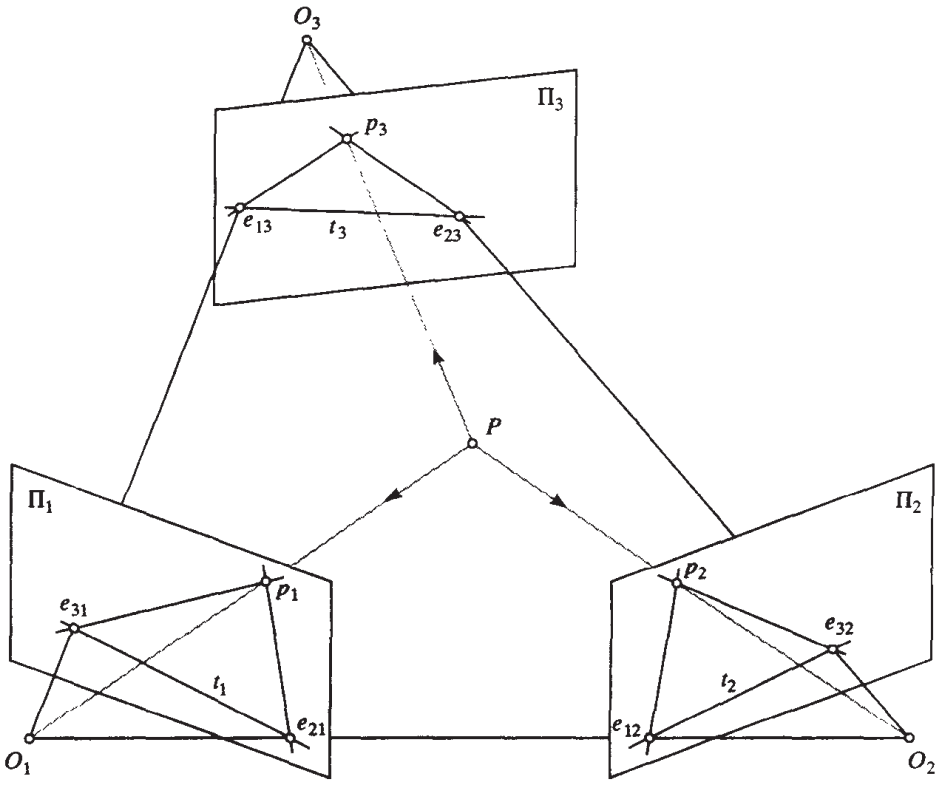


图 10.5 三目外极几何。注意，一般情况下，点 P 并不位于由点 O_1，O_2，O_3 决定的三焦平面上

每对摄像机确定一个外极约束，即，

$$\begin{cases} \boldsymbol{p}_1^T\mathcal{E}_{12}\boldsymbol{p}_2=0 \\ \boldsymbol{p}_2^T\mathcal{E}_{23}\boldsymbol{p}_3=0 \\ \boldsymbol{p}_3^T\mathcal{E}_{31}\boldsymbol{p}_1=0 \end{cases} \tag{10.9}$$

其中，$\mathcal{E}_{ij}$ 是和图像对 $i\leftrightarrow j$ 对应的本征矩阵。这三个约束并不是独立的，这是由于 $\boldsymbol{e}_{31}^T\mathcal{E}_{12}\boldsymbol{e}_{32}=\boldsymbol{e}_{12}^T\mathcal{E}_{23}\boldsymbol{e}_{13}=\boldsymbol{e}_{23}^T\mathcal{E}_{31}\boldsymbol{e}_{21}=0$（究其原因，请考虑外极点 e_{31} 和 e_{32}；它们是第三个摄像机的光学中心点 O_3 的两个像点，因此它们应该是外极点的对应点）。

方程组（10.9）中任意两个方程都是独立的。特别地，当本征矩阵已知时，就可以通过另外两个点 p_2 和 p_3 的位置预测出 p_1 的位置：确实，在方程组（10.9）中第一个和第三个约束构成了由两个线性方程组成的系统，其中 p_1 的两个坐标未知。从几何上来说，p_1 是与 p_2 和 p_3 相应的外极线的交点（见图 10.5）。这样，对于引言中提到的转换问题，三目外极几何给出了一个解决的方法。

10.2.1　三焦几何

如果用一条直线代替一个点在三幅图中成像，我们可以得到第二组约束条件：投影到像平面上直线 l 的点集构成一个平面 L，针孔（光心）和直线 l 位于该平面 L 上。我们可以用下面的方法描述这个平面：如果 $\mathcal{M}$ 表示一个 3×4 的投影矩阵，点 p 满足 $z\boldsymbol{p}=\mathcal{M}\boldsymbol{P}$，$L$ 上的一个点 P 在 l 上的投影，可以描述为下式：

$$\boldsymbol{l}^T\mathcal{M}\boldsymbol{P}=0 \tag{10.10}$$

其中，$\boldsymbol{P}=(x,y,z,1)^T$ 是 P 点的四维齐次坐标向量，$\boldsymbol{l}=(a,b,c)^T$ 是 l 的三维齐次坐标向量。显然，方程(10.10)是包含摄像机光学中心点和直线 l 的平面 L 的方程，其中，$\boldsymbol{L}=\mathcal{M}^T\boldsymbol{l}$ 是该平面的向量表示。

同一直线的两个像 l_1 和 l_2 并不限制相应的摄像机的相对位置和方向，因为相应的平面 L_1 和 L_2 总是相交的（除非它们是平行的，但是这种情况可以认为是相交于无限远；更多的讨论见第 13 章）。现在考虑同一条直线 l 的三个像 l_i，l_2，l_3，并用 L_1，L_2，L_3 表示相应的平面（见图 10.6）。

这些平面的交界处形成一条直线而不是一般情况下的一个点。从代数的观点看，这三个方程确定的系统

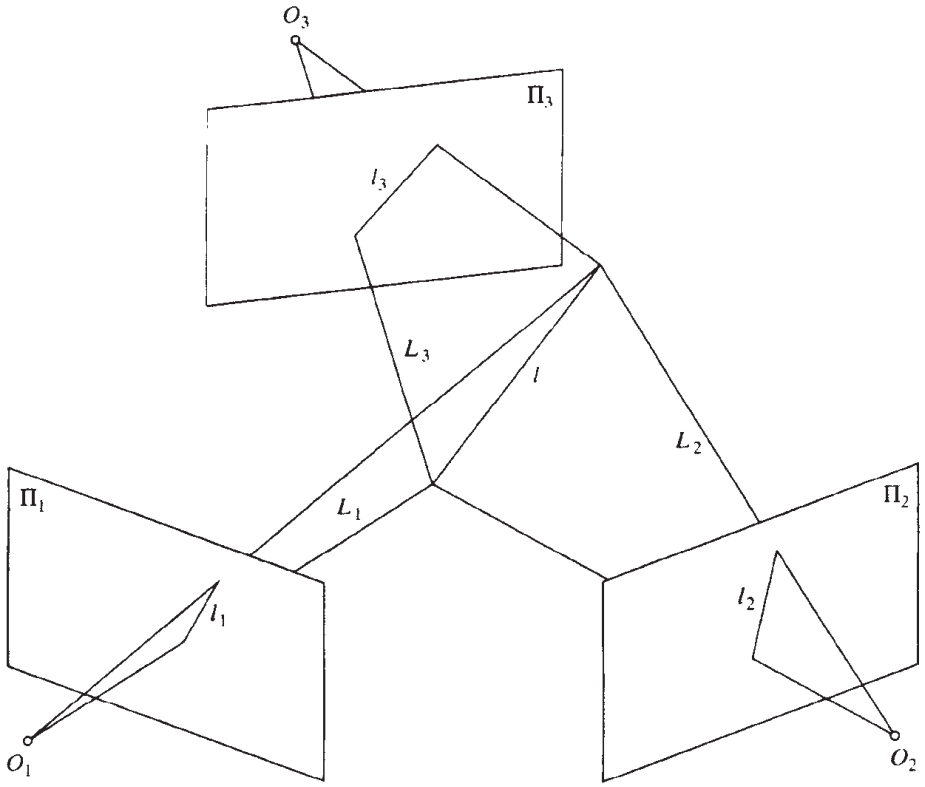


图 10.6　一条线的三个像定义它为的三个平面的（退化的）相交线

$$\begin{pmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \\ \boldsymbol{L}_3^T \end{pmatrix} \boldsymbol{P} = \boldsymbol{0}$$

其中,三个未知量 x, y, z 一定是可以约简的,或者等价地,如下 3×4 矩阵

$$\mathcal{L} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{l}_1^T \mathcal{M}_1 \\ \boldsymbol{l}_2^T \mathcal{M}_2 \\ \boldsymbol{l}_3^T \mathcal{M}_3 \end{pmatrix}$$

的秩一定是2,这也说明了它的所有 3×3 的子矩阵的行列式一定是零。明显的,这些行列式就是 $\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3$ 的坐标向量的三线性组合。如后面的内容所述,4个行列式中仅有两个是独立的。

10.2.2 标定的情形

要得到三线性约束的一个显式表示,我们取第一个摄像机的坐标系作为参考的世界坐标系,这样可以将投影矩阵记为$\mathcal{M}_1 = (\mathrm{Id} \quad \boldsymbol{0})$,$\mathcal{M}_2 = (\mathcal{R}_2 \quad \boldsymbol{t}_2)$和$\mathcal{M}_3 = (\mathcal{R}_3 \quad \boldsymbol{t}_3)$,并且改写$\mathcal{L}$为

$$\mathcal{L} = \begin{pmatrix} \boldsymbol{l}_1^T & 0 \\ \boldsymbol{l}_2^T \mathcal{R}_2 & \boldsymbol{l}_2^T \boldsymbol{t}_2 \\ \boldsymbol{l}_3^T \mathcal{R}_3 & \boldsymbol{l}_3^T \boldsymbol{t}_3 \end{pmatrix} \tag{10.11}$$

正如在习题中所示的,其中三个子矩阵的行列式可以写在一起,如下所示:

$$\boldsymbol{l}_1 \times \begin{pmatrix} \boldsymbol{l}_2^T \mathcal{G}_1^1 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^2 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^3 \boldsymbol{l}_3 \end{pmatrix} = \boldsymbol{0} \tag{10.12}$$

其中,

$$\mathcal{G}_1^i = \boldsymbol{t}_2 \boldsymbol{R}_3^{iT} - \boldsymbol{R}_2^i \boldsymbol{t}_3^T \qquad i = 1, 2, 3 \tag{10.13}$$

并且 $\boldsymbol{R}_2^i$ 和 $\boldsymbol{R}_3^i (i = 1,2,3)$表示$\mathcal{R}_2$ 和$\mathcal{R}_3$ 的列。第4个行列式等价于$|\boldsymbol{l}_1 \quad \mathcal{R}_2 \boldsymbol{l}_2 \quad \mathcal{R}_3 \boldsymbol{l}_3|$,当 L_1, L_2, L_3 的法线共面时它的值为零。相应的方程可以表示为方程(10.12)的三个行列式的线性组合(见习题)。当然,这些行列式中仅有两个是线性无关的。

三个 3×3 的矩阵$\mathcal{G}_1^i$ 定义了27个参数的 $3\times3\times3$ 的三焦张量(或者去掉比例因子是26个)。(张量是一个对应于多线性形式的多维矩阵,如二维矩阵对应于双线性形式)让第一个坐标系作为可以表示所有投影方程的坐标系,O_1 是它的原点,这样向量 $\boldsymbol{t}_2$ 和 $\boldsymbol{t}_3$ 能被转换为外极点 e_{12}和 e_{13}的齐次图像坐标。特别的,从式(10.13)可以得到,对于每一对对应的外极线 l_2 和 l_3,满足 $\boldsymbol{l}_2^T \mathcal{G}_1^i \boldsymbol{l}_3 = 0$。

方程(10.12)可以改写成如下形式

$$\boldsymbol{l}_1 \propto \begin{pmatrix} \boldsymbol{l}_2^T \mathcal{G}_1^1 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^2 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^3 \boldsymbol{l}_3 \end{pmatrix} \tag{10.14}$$

其中,$\boldsymbol{a} \propto \boldsymbol{b}$ 表示向量$\boldsymbol{a}$ 和$\boldsymbol{b}$ 只相差一个非零比例因子。这表明三焦距张量也限制了三个对应点的位置:假设 P 是 l 上的一个点。P 的第一个像位于直线 l_1 上,因此有 $\boldsymbol{p}_1^T \boldsymbol{l}_1 = 0$。特别地:

$$\boldsymbol{p}_1^T \begin{pmatrix} \boldsymbol{l}_2^T \mathcal{G}_1^1 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^2 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^3 \boldsymbol{l}_3 \end{pmatrix} = 0 \tag{10.15}$$

给定三个对应点 $p_1 \leftrightarrow p_2 \leftrightarrow p_3$（见图 10.7），把分别过点 p_2, p_3 的两条直线（如 $\boldsymbol{l}_i' = [1, 0, -u_i]^T$ 与 $\boldsymbol{l}_i' = [0, 1, -v_i]^T$，$i = 2, 3$ 代入等式(10.15)），可以得到 4 个独立的约束条件，这些约束对于 p_1, p_2 和 p_3 的坐标是三线性的。如果张量已知，就可以从 p_2，p_3 在其他图像中的位置计算 p_1 的位置。这就是转换问题的第二个解法。

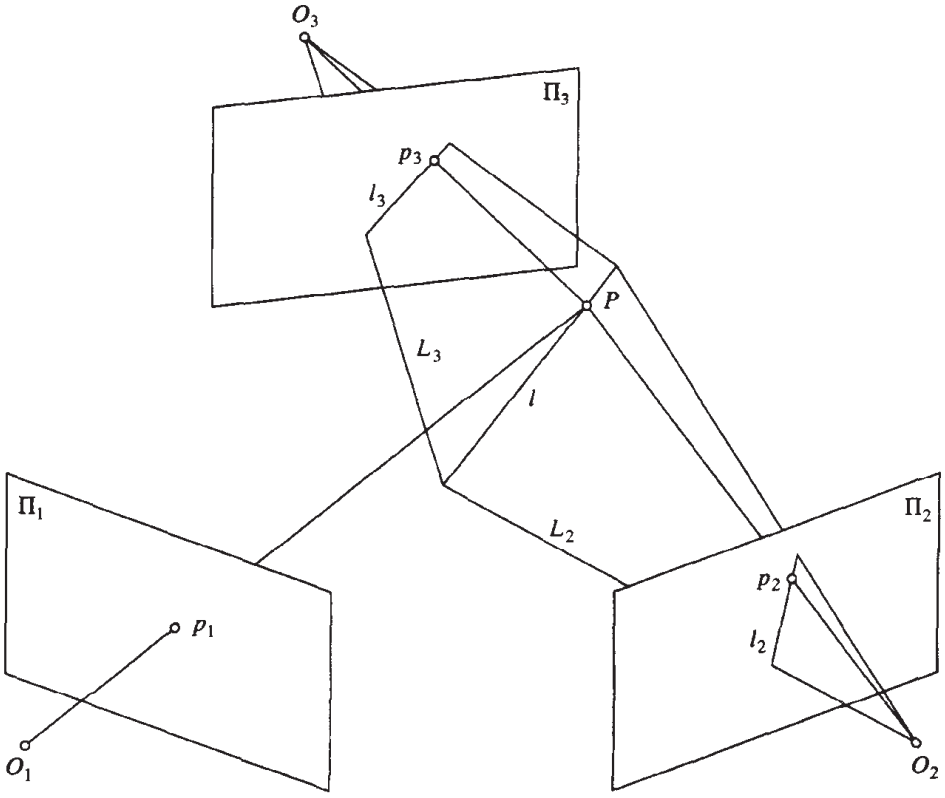


图 10.7　给出点 P 在三个图像中的对应点 p_1、p_2、p_3，以及任意两条过 p_2、p_3 的直线 l_2、l_3，则射线 O_1 和 p_1 一定与 l_2、l_3 延伸出来的平面 L_2、L_3 的交线相交

10.2.3　未标定的情况

在三个摄像机具体参数未知的情况下，我们也可以导出三线性约束。在这种情况下，由与向量 $\boldsymbol{l}$ 对应的直线为 $\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}$，以及 $\boldsymbol{l}^T\boldsymbol{p} = 0$，就可以得到 $\boldsymbol{l} = \mathcal{K}^{-T}\hat{\boldsymbol{l}}$，等价的有 $\hat{\boldsymbol{l}} = \mathcal{K}^T\boldsymbol{l}$。

特别地，当 $\boldsymbol{p}_i = \hat{\boldsymbol{p}}_i$ 且 $\boldsymbol{l}_i = \hat{\boldsymbol{l}}_i$ 时，就可以得到式(10.11)。对于一般情况我们有：

$$\mathcal{L} = \begin{pmatrix} \boldsymbol{l}_1^T \mathcal{K}_1 & 0 \\ \boldsymbol{l}_2^T \mathcal{K}_2 \mathcal{R}_2 & \boldsymbol{l}_2^T \mathcal{K}_2 \boldsymbol{t}_2 \\ \boldsymbol{l}_3^T \mathcal{K}_3 \mathcal{R}_3 & \boldsymbol{l}_3^T \mathcal{K}_3 \boldsymbol{t}_3 \end{pmatrix}$$

和

$$\mathrm{Rank}(\mathcal{L}) = 2 \Longleftrightarrow \mathrm{Rank}\left[\mathcal{L}\begin{pmatrix} \mathcal{K}_1^{-1} & 0 \\ 0 & 1 \end{pmatrix}\right] = \mathrm{Rank}\begin{pmatrix} \boldsymbol{l}_1^T & 0 \\ \boldsymbol{l}_2^T \mathcal{A}_2 & \boldsymbol{l}_2^T \boldsymbol{b}_2 \\ \boldsymbol{l}_3^T \mathcal{A}_3 & \boldsymbol{l}_3^T \boldsymbol{b}_3 \end{pmatrix} = 2$$

其中，$\mathcal{A}_i \overset{\text{def}}{=} \mathcal{K}_i \mathcal{R}_i \mathcal{K}_1^{-1}$，$\boldsymbol{b}_i \overset{\text{def}}{=} \mathcal{K}_i \boldsymbol{t}_i$，$i=2,3$。需说明的是，关于三个摄像机的投影矩阵可以表示为$\mathcal{M}_1=(\mathcal{K}_1 \quad \mathbf{0})$，$\mathcal{M}_2=(\mathcal{A}_2\mathcal{K}_1 \quad \boldsymbol{b}_2)$及$\mathcal{M}_3=(\mathcal{A}_3\mathcal{K}_1 \quad \boldsymbol{b}_3)$。特别地，$\boldsymbol{b}_2$ 和 $\boldsymbol{b}_3$ 同样可以理解为外极点 e_{12}、e_{13}的齐次图像坐标，，并且式(10.14)和式(10.15)的三线性约束依然成立。

$$\mathcal{G}_1^i = \boldsymbol{b}_2\boldsymbol{A}_3^{iT} - \boldsymbol{A}_2^i\boldsymbol{b}_3^T$$

其中，$\boldsymbol{A}_2^i$ 和 $\boldsymbol{A}_3^i(i=1,2,3)$表示$\mathcal{A}_2$ 和$\mathcal{A}_3$ 的列。前面曾经提到，对于任何一对对应的外极线 l_2 和 l_3，有 $\boldsymbol{l}_2^T\mathcal{G}_1^i\boldsymbol{l}_3=0$。

10.2.4 三焦张量的估算

现在可以通过由三幅图建立的点线对应关系来说明如何进行三焦张量的估算。定义张量的方程式是线性的，并且只依赖于对图像的测量。与弱标定情况相似，可以使用线性方法来估算张量中的这 26 个参数。每三个对应点提供 4 个独立的线性方程，每三条对应线可以提供 2 个线性约束。因此，张量参数可以由满足 $2p+l\geqslant 13$ 的 p 个点和 l 条线计算。例如，7 组点条件或者 13 组的线条件，或者 3 组点条件和 7 组线条件等，都可以计算出张量参数。与弱标定情况相似，可以通过规范化图像坐标系来提高张量估算过程的数字稳定性，在规范化后的坐标系中，数据点集中在原点周围，其平均值距原点$\sqrt{2}$个像素。

如果按照这个方法我们就忽略了一点，即 26 个三焦张量参数是不独立的。这很容易证明：本征矩阵只有 5 个独立的参数(旋转和平移参数，后者只用于同一个尺度上的定义)，基础矩阵只有 7 个。同样，定义三焦张量的参数要满足一些约束条件，包括前面提到的方程 $\boldsymbol{l}_2^T\mathcal{G}_1^i\boldsymbol{l}_3=0(i=1,2,3)$，其中任意一对匹配的外极线 l_2，l_3 都满足。很容易证明矩阵$\mathcal{G}_1^i$ 是奇异矩阵——这个性质我们将在后面的第 13 章中介绍。Faugeras 和 Mourrain(1995)发现，一个未标定的三目立体视觉设备的三焦张量参数要满足 8 个独立约束，从而将独立参数的数目减少到 18 个。Hartley(1995)提出的方法首先从线性估计三焦张量中获得外极点 e_{12}和 e_{13}(或者等价地获得式 10.13 中的 $\boldsymbol{t}_2$ 和 $\boldsymbol{t}_3$ 向量)，然后以一个线性的形式恢复满足这些约束的张量系数的集合，这个方法后验地增强了这些约束。

10.3 更多的视图

4 个视图会怎么样呢？在这一节中我们将介绍 Faugeras 和 Mourrain(1995)的方法。消去第 2 章中透视射影方程(2.16)中的分母可得

$$\begin{pmatrix} u\mathcal{M}^3-\mathcal{M}^1 \\ v\mathcal{M}^3-\mathcal{M}^2 \end{pmatrix}\boldsymbol{P}=0 \tag{10.16}$$

其中，$\mathcal{M}^1$，$\mathcal{M}^2$ 和$\mathcal{M}^3$ 表示矩阵$\mathcal{M}$中三行(请注意，这里的$\mathcal{M}^1$，$\mathcal{M}^2$，$\mathcal{M}^3$ 和原来用来表示矩阵一行的 $\boldsymbol{m}_1^T$，$\boldsymbol{m}_2^T$，$\boldsymbol{m}_3^T$ 是等价的，之所以没有用原来的方式表示矩阵的一行，主要是为了避免与后面用于表示不同矩阵不同行的符号发生混淆)。

假设有 4 个视图，它们相应的投影矩阵为$\mathcal{M}_j(j=1,2,3,4)$。由式(10.16)可得

$$\mathcal{Q}\boldsymbol{P}=0,\quad 其中,\quad \mathcal{Q}\stackrel{\text{def}}{=}\begin{pmatrix} u_1\mathcal{M}_1^3-\mathcal{M}_1^1 \\ v_1\mathcal{M}_1^3-\mathcal{M}_1^2 \\ u_2\mathcal{M}_2^3-\mathcal{M}_2^1 \\ v_2\mathcal{M}_2^3-\mathcal{M}_2^2 \\ u_3\mathcal{M}_3^3-\mathcal{M}_3^1 \\ v_3\mathcal{M}_3^3-\mathcal{M}_3^2 \\ u_4\mathcal{M}_4^3-\mathcal{M}_4^1 \\ v_4\mathcal{M}_4^3-\mathcal{M}_4^2 \end{pmatrix} \tag{10.17}$$

这个方程组有 8 个齐次方程,4 个未知数,可以得到一个非平凡解。由此可知相应的 8×4 相关矩阵$\mathcal{Q}$的秩最大为 3,也就是说,任意一个 4×4 子矩阵行列式一定为 0。从几何意义上讲,每一对方程都代表了一条过像点 p_i 的射线 R_i ($i=1,2,3,4$),而且若这些射线都交于一点 P(见图 10.8),矩阵$\mathcal{Q}$的阶必为 3。

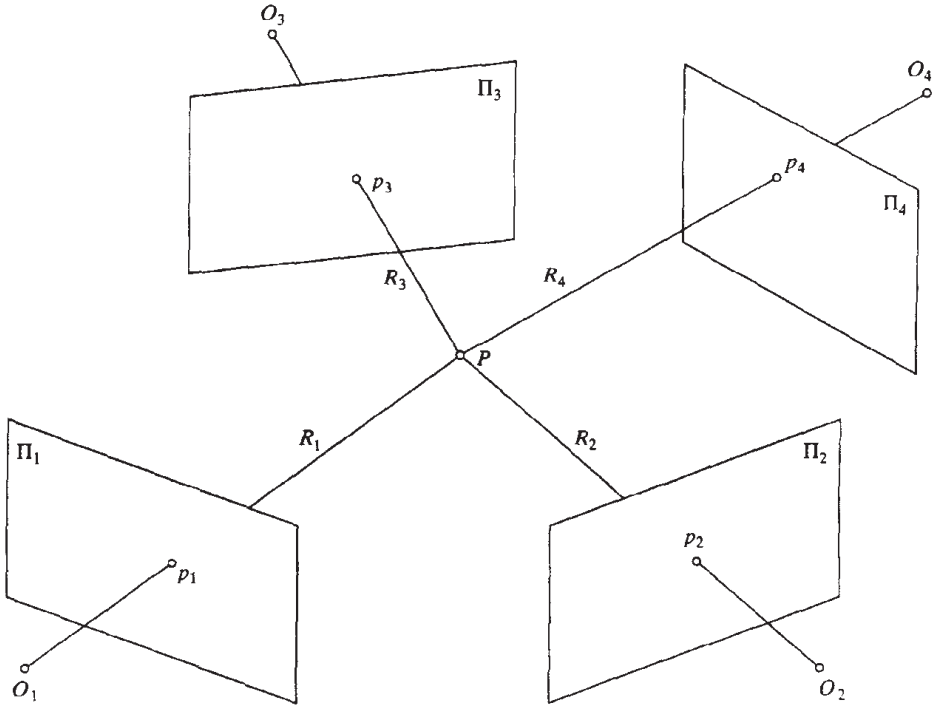


图 10.8　点 P 在 4 个视图中的像点是这 4 个像点相应射线的交点点 P 的 4 个像点P_1、P_2、P_3和P_4确定了该点是对应的4条射线R_i($i=1,2,3,4$)的交点

矩阵$\mathcal{Q}$有三类 4×4 子行列式:

1. 这类行列式中的两行取自同一射影矩阵,另外两行取自另一矩阵。这种行列式包括 6 个,例如①

$$\mathrm{Det}\begin{pmatrix} u_1\mathcal{M}_1^3-\mathcal{M}_1^1 \\ v_1\mathcal{M}_1^3-\mathcal{M}_1^2 \\ u_2\mathcal{M}_2^3-\mathcal{M}_2^1 \\ v_2\mathcal{M}_2^3-\mathcal{M}_2^2 \end{pmatrix}=0 \tag{10.18}$$

① 一般的形式可以通过用(u^1,u^2)代替(u,v)、调整字母索引、引入张量来描述。我们在这里仅用简化的形式。

这类等式对相应图像点的位置产生双线性约束。我们令$\mathcal{M}_1 = (\mathrm{Id} \quad \mathbf{0})$、$\mathcal{M}_2 = (\mathcal{R}^T \quad -\mathcal{R}^T\boldsymbol{t})$，很容易证明(见习题)，上述方程可以简化为式(10.2)的外极点约束。

2. 第二类行列式中两行取自同一射影矩阵，另外两行分别取自另外两个不同矩阵。这种方程有 48 个，例如

$$\mathrm{Det}\begin{pmatrix} u_1\mathcal{M}_1^3 - \mathcal{M}_1^1 \\ v_1\mathcal{M}_1^3 - \mathcal{M}_1^2 \\ u_2\mathcal{M}_2^3 - \mathcal{M}_2^1 \\ v_3\mathcal{M}_3^3 - \mathcal{M}_3^2 \end{pmatrix} = 0 \tag{10.19}$$

这类等式对相应像点位置产生三线性约束。如果令$\mathcal{M}_1 = (\mathrm{Id} \quad \mathbf{0})$，很容易证明(见习题)，上述方程可以转化为式(10.15)表示的三线性约束。特别地，这类行列式可以使用矩阵$\mathcal{G}_i^i$($i = 1,2,3$)表示。注意，在这里我们给出三焦约束的几何解释——与同一点的三个像点对应的射线交于空间一点。

3. 最后一类行列式中每一行都来源于不同的矩阵。这类行列式包括 16 个，例如

$$\mathrm{Det}\begin{pmatrix} v_1\mathcal{M}_1^3 - \mathcal{M}_1^2 \\ u_2\mathcal{M}_2^3 - \mathcal{M}_2^1 \\ v_3\mathcal{M}_3^3 - \mathcal{M}_3^2 \\ v_4\mathcal{M}_4^3 - \mathcal{M}_4^2 \end{pmatrix} = 0 \tag{10.20}$$

这类等式产生对相应像点 $\boldsymbol{p}_i$ = ($\boldsymbol{i} = 1,2,3,4$)位置的四线性约束。从几何上讲，矩阵$\mathcal{Q}$的每一行代表了图像上的一条直线，也可以看成是穿过相应摄像机光学中心点的某个平面。因此每个这类等式表示 4 个相关平面相交于一点(一般情况下 4 个平面是不相交的)。

现在我们来研究四线性方程。联系像点坐标考虑如式(10.20)的行列式，立刻发现四线性约束可以写成如下形式：

$$\varepsilon_{ijkl}\,\mathrm{Det}\begin{pmatrix} \mathcal{M}_1^i \\ \mathcal{M}_2^j \\ \mathcal{M}_3^k \\ \mathcal{M}_4^l \end{pmatrix} \tag{10.21}$$

其中，$\varepsilon_{ijkl} = \mp 1$，$i$，$j$，$k$，$l$ 顺序表示 1 到 3(见习题)。这些参数决定了四焦张量(Triggs，1995)。

像三焦张量一样，四焦张量也可以用点和线来解释它的几何意义。特别地，假定点 P 的 4 个像点为 p_i($i = 1,2,3,4$)，像平面上 4 个任意穿过像点的直线为 l_i($i = 1,2,3,4$)。由 4 条直线的原像组成的四个平面 L_i($i = 1,2,3,4$)一定交于 P 点，反之说明这个 4×4 的矩阵一定是 3 阶的，而且行列式为 0

$$\mathcal{L} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{l}_1^T\mathcal{M}_1 \\ \boldsymbol{l}_2^T\mathcal{M}_2 \\ \boldsymbol{l}_3^T\mathcal{M}_3 \\ \boldsymbol{l}_4^T\mathcal{M}_4 \end{pmatrix}$$

明显地，这就是对 4 条直线 l_i($i = 1,2,3,4$)参数的四线性约束。另外，因为$\mathcal{L}$的每行 $\boldsymbol{L}_i^T = \boldsymbol{l}_i^T\mathcal{M}_i$ 为相应矩阵各行的线性组合，所以利用 l_i 的坐标可以从$\mathcal{M}_i$ 推导出四焦张量的系数 $\mathrm{Det}(\mathcal{L})$，这

些系数就是式(10.21)定义的四焦张量系数。

最后,因为 $\mathrm{Det}(\mathcal{L})$在 $\boldsymbol{l}_1$ 的坐标下是线性的,因而这个行列式等于 0 也可以写为 $\boldsymbol{l}_1 \cdot \boldsymbol{q}(\boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4) = 0$,其中 $\boldsymbol{q}$ 是一个 l_i 坐标的($i = 2,3,4$)三线性函数。因为任何经过点 p_1 的直线 l_1 满足 $\boldsymbol{p}_1 \propto \boldsymbol{q}(\boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4)$,从几何意义上讲,通过 O_1, p_1 的射线一定通过 l_2, l_3, l_4 原像组成平面的交点(见图 10.9)。从代数意义上讲,给出四焦张量和任意三条穿过三个像点的直线,就可以推出第 4 个像点的位置。这就是另一种转换方法。

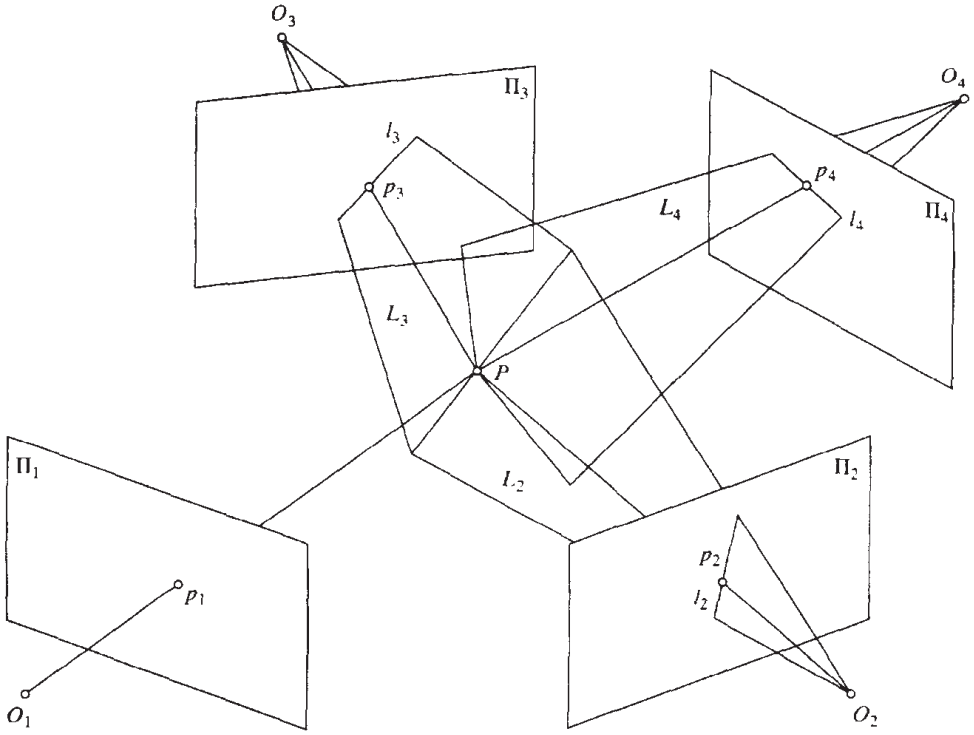


图 10.9　给定一点 P 的 4 个像点 p_1, p_2, p_3, p_4 和三条任意直线 l_2, l_3, l_4 分别通过点 p_2, p_3, p_4,通过点 O_1 和 p_1 的射线必定通过这些相交直线的像所构成的三个平面 L_2, L_3, L_4 的交点

我们并没有对矩阵 $\mathcal{M}_i$ 的形式做任何的假设,因而四焦约束可以应用于标定和未标定两种情况。四焦张量是由 81 个系数定义的(或者是在同一个尺度上的 80 个),但是可以证明这些系数要满足 51 个独立的约束,这样独立系数就只有 29 个了。还可以证明,尽管用同一点的 4 个像点可以列出 16 个相互独立的,对 80 个张量系数的约束,如式(10.20),但是在与每对点相关的 32 个方程中,存在着线性相关性。因此,用线性形式估算四焦张量必须要 6 个点。在 Hartley(1998)中可以找到完成这个任务和增强的实际四焦张量算法。最后,Faugeras和Mourrain(1995)说明了,在代数上四焦线性张量依赖于相关的特征/基础矩阵和三焦张量,所以这样并不引入新的约束。同样地,它说明了更多的视角并不再引入新的独立的限制。

10.4　注释

本征矩阵作为外极约束的代数形式,由Longuet-Higgins(1981)提出,Huang和Faugeras(1989)进一步阐明了它的特性。基础矩阵由Luong和Faugeras(1992,1996)引入。从对应点集估计基础矩阵的鲁棒方法也包括Zhang等(1995)。在第 13 章的讨论中,提出从一系列图像中恢复场景结

构和摄像机运动的问题，我们将再回到基础矩阵和外极变换的相关性质。由 10.1.3 节中推导、方程组(10.4)中定义的瞬时外极约束只适用于标定的摄像机。对于内部参数改变的摄像机的情况，请参考 Viéville 和 Faugeras(1995)。一条直线的三视角的三线性约束由 Spektakis 和 Aloimonos(1990)，Weng，Huang 和 Ahuja(1992)在内部标定的摄像机的运动分析一文中单独提出。Shashua(1995)和 Hartley(1997)扩展了未标定的情况。四焦张量由 Triggs(1995)引入，它的相关性质由 Faugeras 和 Mourrain(1995)、Faugeras 和 Papadopoulo(1997)、Hartley(1998)以及 Heyden(1998)所研究。

引言中提到的分析摄影地形测量法关心的是多幅图片定量信息的提取。在本章中，双目和三目几何约束被认为是决定一对或者三个立体摄像头内外参数(在分析摄影地形测量法中被称为内外方向参数)的条件方程之源。特别地，Longuet-Higgins 关系以一种隐形的形式表示，似乎是共面条件方程。三目约束产生了比例关系约束条件方程，并考虑了标定误差和图像测量误差(见 Thompson 等，1966，第 10 章)。在 Longuet-Higgins 关系下，通过同一个点的三个像点的射线并不能保证相交于一点(见图 10.10)。

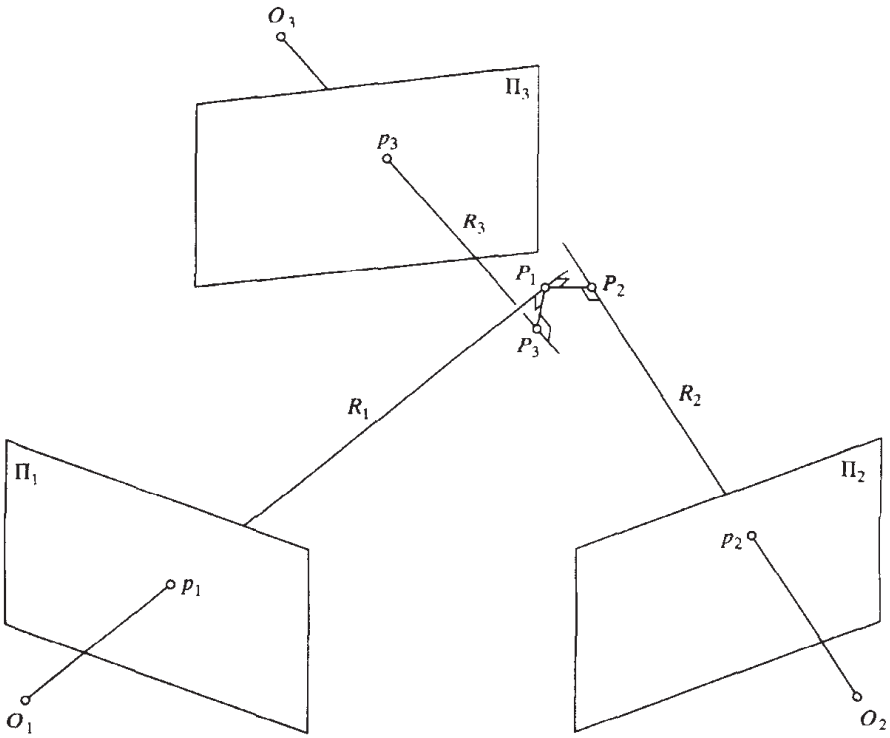


图 10.10 出现标定或者测量误差的三目约束：射线 R_1，R_2，R_3 可能不相交

设置如下：如果和像点 p_1 和 p_i 相对应的射线 R_1 和 R_i ($i=2,3$)不相交，它们之间的最小距离就是点 P_1 和 P_i 之间的距离，这样两点的连线与 R_1 和 R_i 都垂直。从代数上可以表示为：

$$\overrightarrow{O_1P_1} = z_1^i\overrightarrow{O_1p_1} = \overrightarrow{O_1O_i} + z_i\overrightarrow{O_ip_i} + \lambda_i(\overrightarrow{O_1p_1} \times \overrightarrow{O_ip_i}) \quad , \quad i = 2,3 \tag{10.22}$$

假设摄像机已经内部标定，那么与第二个、第三个摄像机对应的投影矩阵为($\mathcal{R}_2^T \ -\mathcal{R}_2^T\boldsymbol{t}_2$)和($\mathcal{R}_3^T \ -\mathcal{R}_3^T\boldsymbol{t}_3$)。方程(10.22)可以改写为第一个摄像机坐标系中的方程：

$$z_1^i\boldsymbol{p}_1 = \boldsymbol{t}_i + z_i\mathcal{R}_i\boldsymbol{p}_i + \lambda_i(\boldsymbol{p}_1 \times \mathcal{R}_i\boldsymbol{p}_i) \quad , \quad i = 2,3 \tag{10.23}$$

注意，通过包含依赖于(未知的)内部参数的条件，可以写出一个相似的方程来表示完全没有标

定的摄像机的情况。不管哪种情况，给定 $\boldsymbol{p}_1, \boldsymbol{p}_i$ 和相应摄像机的投影矩阵，方程(10.23)都能用来计算未知的 z_i、λ_i 和 z_1^i(见习题)。接下来，比例关系约束条件可写为 $z_1^2 = z_1^3$。虽然它比三焦约束还要复杂(尤其是，它不是关于点 p_1, p_2, p_3 三线性的)，但是它的条件不涉及观察点的坐标，并且它能(原则上)通过图像数据直接估计三焦几何。一个潜在的优点就是误差方程 $z_1^2 - z_1^3$ 有明显的几何意义：它是通过摄像机对 $1 \leftrightarrow 2$ 和 $1 \leftrightarrow 3$ 分别得到的点 P 的深度估计值之差。进一步研究三焦张量和比例关系约束条件以及它对三焦几何的实际应用是件很有趣的工作。

习题

10.1　证明本征矩阵的一个奇异值为 0，另外两个奇异值相等。(Huang 和 Faugeras(1989)已经表明逆命题也是成立的——即，任何 3×3 的矩阵，如果有一个奇异值为 0，另外两个奇异值相等，那么它是一个本征矩阵。)

提示：$\mathcal{E}$ 的奇异值是 $\mathcal{E}\mathcal{E}^T$ 的特征值。

10.2　旋转矩阵的指数表示。以单位向量 $\boldsymbol{a}$ 为旋转轴且旋转角度为 θ 的旋转所对应的矩阵，可以表示为 $e^{\theta[\boldsymbol{a}\times]} \stackrel{\text{def}}{=} \sum_{i=0}^{+\infty} \frac{1}{i!}(\theta[\boldsymbol{a}_\times])^i$。用这个表示来推导方程(10.3)。

10.3　方程(10.4)无穷小外极约束的推导假设观察场景是静止不动的，而摄像机是运动的。证明，当摄像机固定，场景以平移速度 $\boldsymbol{v}$ 和角速度 $\boldsymbol{\omega}$ 移动时，外极约束可以改写为 $\boldsymbol{p}^T([\boldsymbol{v}_\times][\boldsymbol{\omega}_\times])\boldsymbol{p} + (\boldsymbol{p} \times \dot{\boldsymbol{p}}) \cdot \boldsymbol{v} = 0$。注意，方程现在是在式(10.4)中出现的两项之和而不是它们的差。

提示：如果 $\mathcal{R}$ 和 $\boldsymbol{t}$ 分别表示对于移动摄像机的本征矩阵定义中的旋转矩阵和平移向量，证明对于一个静止摄像头产生的相同运动场的物体，其位移由旋转矩阵 $\mathcal{R}^T$ 和平移向量 $-\mathcal{R}^T\boldsymbol{t}$ 给出。

10.4　证明：当与 8 点算法相关的 8×8 矩阵是奇异矩阵时，8 个点和两个光学中心点位于同一个二次曲面上(Faugeras，1993)。

提示：利用这个结论：当矩阵是奇异时，它的列向量存在非平凡的线性组合等于零；同时利用该情况下第一个摄像机坐标系下的两个投影矩阵是 $(\mathrm{Id} \quad \mathbf{0})$ 和 $(\mathcal{R}^T \quad -\mathcal{R}^T\boldsymbol{t})$。

10.5　证明下面矩阵的三个 3×3 的行列式

$$\mathcal{L} = \begin{pmatrix} \boldsymbol{l}_1^T & 0 \\ \boldsymbol{l}_2^T\mathcal{R}_2 & \boldsymbol{l}_2^T\boldsymbol{t}_2 \\ \boldsymbol{l}_3^T\mathcal{R}_3 & \boldsymbol{l}_3^T\boldsymbol{t}_3 \end{pmatrix} \quad \text{可以写为} \quad \boldsymbol{l}_1 \times \begin{pmatrix} \boldsymbol{l}_2^T\mathcal{G}_1^1\boldsymbol{l}_3 \\ \boldsymbol{l}_2^T\mathcal{G}_1^2\boldsymbol{l}_3 \\ \boldsymbol{l}_2^T\mathcal{G}_1^3\boldsymbol{l}_3 \end{pmatrix} = \mathbf{0}$$

并证明第 4 个行列式是它们的线性组合。

10.6　证明当满足 $\mathcal{M}_1 = (\mathrm{Id} \quad \mathbf{0})$ 且 $\mathcal{M}_2 = (\mathcal{R}^T \quad -\mathcal{R}^T\boldsymbol{t})$ 时，方程(10.18)等价于方程(10.2)。

10.7　证明当满足 $\mathcal{M}_1 = (\mathrm{Id} \quad \mathbf{0})$ 时，方程组(10.19)等价于方程组(10.15)。

10.8　在图像坐标系下得出方程组(10.20)，并证明系数确实可以写成式(10.21)的形式。

10.9　根据 $\boldsymbol{p}_1, \boldsymbol{p}_i, \mathcal{R}_i$ 和 $\boldsymbol{t}_i$ $(i = 2,3)$，利用方程组(10.23)计算未知量 z_i, λ_i 和 z_1^i。证明 λ_i 的值直接与外极约束相关，并且表明 $z_1^2 - z_1^3$ 对于数据点的依赖度。

编程作业

10.10 使用双目点对应实现弱标定的 8 点算法。

10.11 分别在使用和不使用 Hartley 的前提条件下，实现弱标定的 8 点算法的最小二乘法算法。

10.12 实现从点对应估计三焦张量的算法。

10.13 实现从直线对应估计三焦张量的算法。

第11章 立体视觉

融合两只眼睛获得的图像并察觉它们之间的差别(或称视差)使我们可以获得明显的深度感。在这一章,我们将设计和实现能模仿人类视觉可以获得深度这一能力的算法,这部分内容称为立体视觉。可靠的立体感知算法在机器人视觉导航(见图 11.1)、地图生成、航空勘测和近距照相测量等领域都有很好的应用价值。同样,可靠的立体感知算法在用于目标识别的图像分割,以及用于计算机图形学的三维场景重建中也大有用武之地。

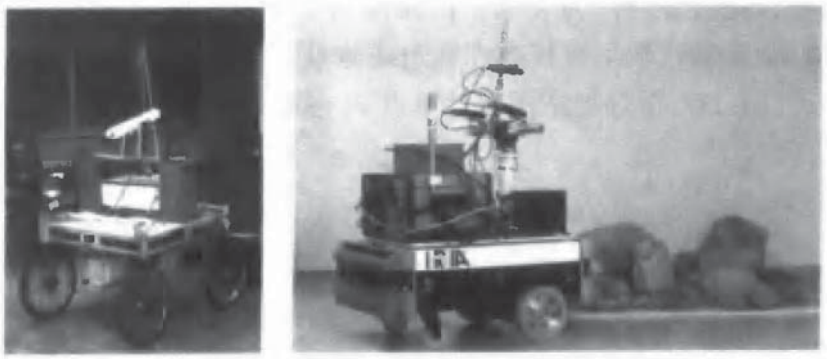


图 11.1 左图:斯坦福大学的小车展示了单个摄像机在机器人视觉导航中的应用,这个摄像机沿直线间断地运动并提供室外场景的多个快照。右图:INRIA的移动机器人使用三个摄像机对周围环境进行地图生成。就像这些例子中所示,尽管两个摄像机已经足以胜任立体融合,但是移动机器人有时会安装三个(或更多)的摄像机。本章大部分内容只涉及双目立体视觉,11.4节将讨论使用多个摄像机的算法

立体视觉包括两个过程:融合两个(或多个)摄像机观察到的特征,以及重建这些特征的三维原像。后一个过程相对简单,这是由于对应点的原像(理论上)出现在经过成像点和相应光心的射线的交点处(如图 11.2 左图所示)。因此在任意时刻,当单个图像特征被观察到时,立体视觉是简单的。然而一幅图像中往往包含几十万个像素,同时又有几万个图像特征,例如边缘等,因此必须设计一些方法来建立正确的对应以避免错误的深度测量(图 11.2 右图)。外极约束在立体匹配的过程中扮演了一个基础性的作用,因为外极约束将搜索图像中的对应关系限制在相互对应的外极线上进行。

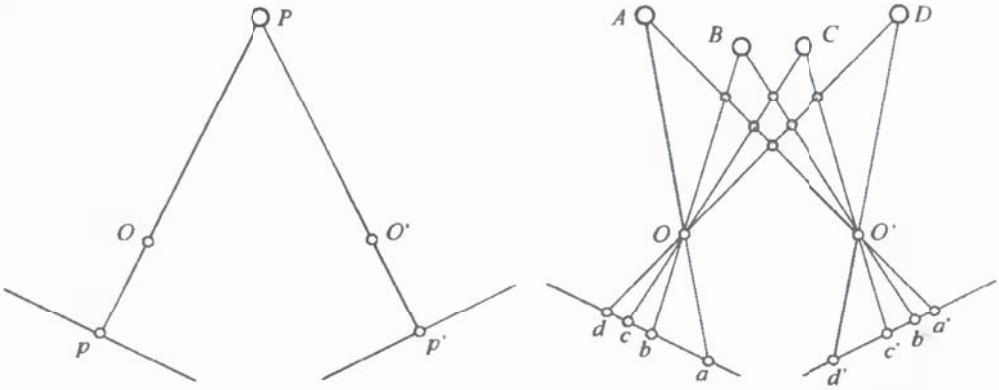


图 11.2 双目融合问题:简单的情况如左图所示,图中没有二义性,在这种情况下立体重构是一个简单的问题。更加一般的情况如右图所示,左边的像平面上任意4个点,可以和右边像平面任意 4 点对应。只有 4 个对应是正确的,其他的导致不正确的重建,在图中用灰色小圆点表示

我们假设在本章后面部分提到的摄像机都是完全标定过的，即相对于某个固定的坐标系，它们的内外参数可以准确获知(这当然也意味着相应的双目或三目摄像机的本征矩阵和三焦张量是知道的)。未标定摄像机的情况在第 12 章和第 13 章运动恢复结构的内容中介绍。

11.1 重建

已知一个标定过的摄像机和两个对应点 p 和 p'，在原理上，可以直接通过将两条射线 $R = Op$ 和 $R' = O'p'$ 相交来重建相应的场景点。然而，在实际中由于标定和特征定位的误差，射线 R 和 R' 可能永远也不会真正相交(见图 11.3)。在这种情况下，有很多合理的重建方法可以采用。例如，可以建立一条线段同时垂直 R 和 R' 并与两条射线相交：这条线段的中心 P 是最靠近两条射线的点，可以把这个点作为 p 和 p' 的原像点。读者应该注意到，在第 10 章的结尾使用了相似的构造方法，它从代数上表征存在标定或测量误差时的多视角几何。在第 10 章中导出的式(10.22)和式(10.23)，很好地适用于计算点 P 在第一个摄像机坐标系下的坐标。

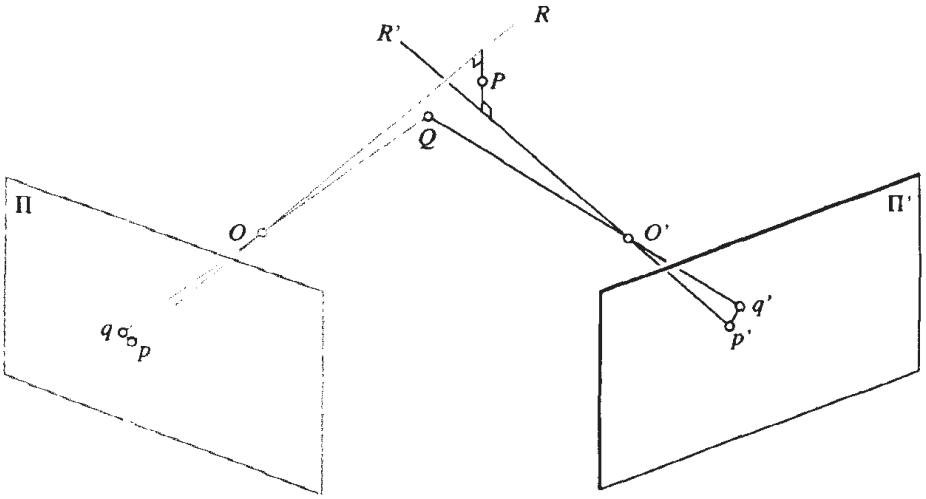


图 11.3 存在测量误差时的重建。细节内容见正文

另外，也可以使用纯代数的方法重建场景点：给定投影矩阵 $\mathcal{M}$ 和 $\mathcal{M}'$ 以及对应点 p 和 p'，可以把约束 $z\boldsymbol{p} = \mathcal{M}\boldsymbol{P}$ 和 $z'\boldsymbol{p}' = \mathcal{M}\boldsymbol{P}$ 重写成如下形式：

$$\begin{cases} \boldsymbol{p} \times \mathcal{M}\boldsymbol{P} = 0 \\ \boldsymbol{p}' \times \mathcal{M}'\boldsymbol{P} = 0 \end{cases} \Longleftrightarrow \begin{pmatrix} [\boldsymbol{p}_\times]\mathcal{M} \\ [\boldsymbol{p}'_\times]\mathcal{M}' \end{pmatrix} \boldsymbol{P} = 0$$

上式是一个过约束的方程，有 4 个关于 P 的坐标的独立线性等式。用第 3 章介绍的最小二乘法可以解这个方程。和前面方法不同的是，这个重建方法没有明显的几何解释，但是可以很容易地推广到三个或多个摄像机的情况，每增加一个摄像机只是增加两个约束。

最后，还有一种重建场景点的方法：设对应于 p 和 p' 的场景点为 Q，这个 Q 点实际的成像点是 q 和 q'，Q 点的选择要求使 $d^2(p,q) + d^2(p',q')$ 最小(见图 11.3)。与本节前面介绍的两种方法不同的是，这个方法没有重建点的解析解，必须通过第 3 章中介绍的非线性最小二乘方法来估计。前面两种方法获得的结果都可以作为这个最优化过程的初始值。这个非线性方法也适用于多幅图像的情况。

11.1.1　图像校正

当感兴趣的图像经过校正后(即用两幅等价图像代替,这两幅等价图像与平行于基线的平面共面,如图 11.4 所示,基线即连接两个光心的直线),立体视觉算法的计算量可以大大降低。校正过程可以通过将原图投影到一个新的图像平面中来实现。在适当的坐标系中,校正图像的外极线和校正图像的扫描线相同,它们都平行于基线。关于校正图像平面的选择,有两个自由度:(a) 平面和基线的距离,实际上这个自由度是无关的,这是由于调整这个自由度只改变校正图像的尺度,这个效果可以通过图像坐标轴的逆尺度变换来平衡,(b)相对于垂直基线的平面,校正平面的法向量方向;很自然的一个选择就是平行于两个像平面交线的一个平面,并且这个平面使得由于再投影过程所产生的扭曲最小。

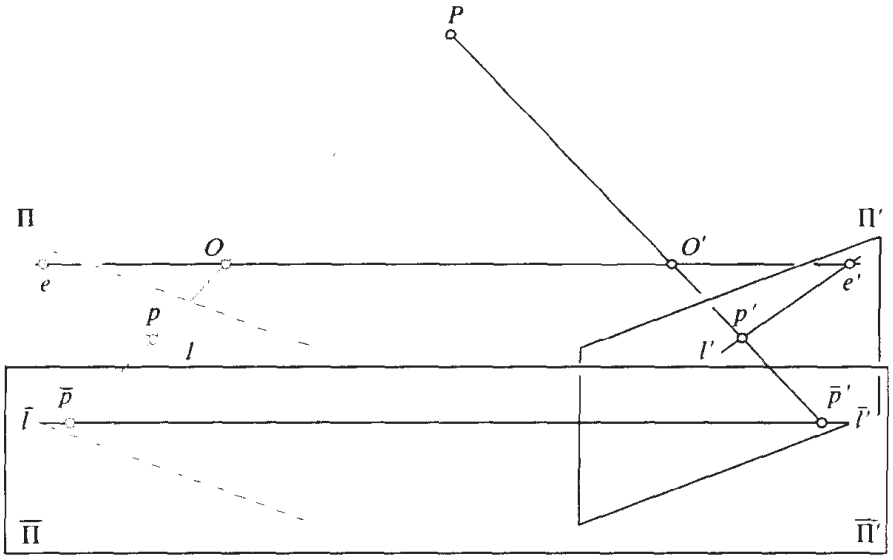


图 11.4　一个校正图像对:两个图像平面 Π 和 Π′原图,它们被再投影到同一个平行于基线的平面 $\bar{\Pi}=\bar{\Pi}'$。原图中点 p 和 p' 对应的外极线 l 和 l' 被投影到同一条扫描线 $\bar{l}=\bar{l}'$ 上,这条扫描线也平行于基线并且过再投影点 $\bar{p}$ 和 $\bar{p}'$。在现代计算机图像硬件和软件条件下,通过将输入图像看做多面网格并使用纹理映射将网格投影到平面 $\bar{\Pi}=\bar{\Pi}'$ 上,可以很容易获得地校正图像

在校正图像的情况下,前面介绍过的视差的概念有了一个准确的含义:给定左右图像中在同一扫描线上的两点 p 和 p',它们的坐标设为 (u,v) 和 (u',v),视差可以被定义为 $d=u'-u$。从现在开始假设使用的是归一化的图像坐标。如练习中所示,如果 B 代表光心间的距离,在这里也可以称为基线,那么在(归一化)左摄像机坐标系中,点 P 的深度可以表示为 $z=-B/d$。特别地,在第一个摄像机坐标系中,点 P 的坐标向量为 $\boldsymbol{P}=-(B/d)\boldsymbol{p}$, 其中,$\boldsymbol{p}=(u,v,1)^T$ 是点 p 在归一化图像坐标系中的坐标向量。对于校正图像,这里也提供了另一个重建方法。

11.2　人类的立体视觉过程

在介绍建立双目对应的算法之前,先来讨论一下人类立体视觉过程的机制。首先,应该注意到,与摄像机刚性地固定在一个立体支架上不同,人的两个眼睛可以在眼眶内转动。在每个瞬间,它们注视着空间中的一个特定点(也就是说,它们旋转使得对应图像成像在视网膜中央

凹的中心)。

图 11.5 说明了一个简化的两维的情形。如果 l 和 r 分别代表平分左右眼球的垂直平面与过同一场景点射线的(逆时针)夹角,我们定义相应的视差为 $d = r - l$(如图 11.5 所示)。这可以作为一个基本的三角学练习来证明 $d = D - F$,其仔,D 代表穿过场景点射线间的夹角,F 表示穿过注视点的射线的夹角。零视差的点位于 Vieth-Müller 圆上,这个圆经过注视点和两个眼球的前节点(即光心)。在这个圆内部的点有正的视差,在这个圆外部的点有负的视差,如图 11.5所示。所有视差为 d 的点共圆,这个圆经过两个眼球的节点,随着 d 的变化形成不同的圆。很明显,这个性质可以将注视点附近的点按照深度进行排序。然而,可以很清楚地看到,如果要获得场景点的绝对位置必须知道聚散角(vergence angle),即头部的中垂平面和两个注视射线的夹角,才能重建场景点的绝对位置。

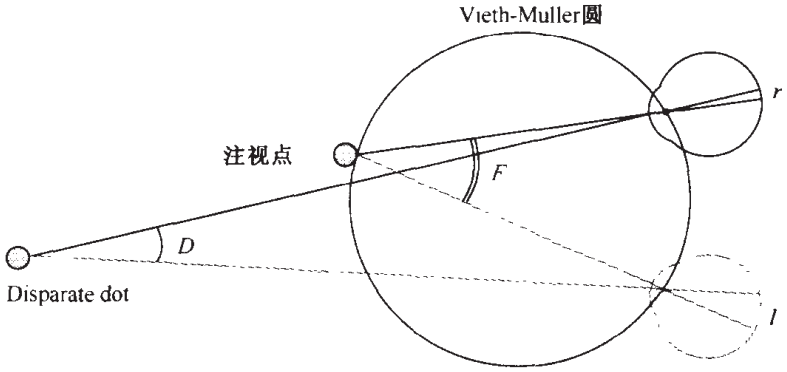


图 11.5 在这幅图中,靠近眼睛的点是注视点,它没有视差地投影到视网膜的中央。远处点的投影偏离中心位置的不同程度指示出不同的深度

三维的情况自然会复杂一些:零视差的点形成一个面,即等视面;但是一般的结论是一样的:即绝对位置的获得需要已知聚散角。在近 100 年前,Wundt 和 Helmholtz(1909)已经证明了人类的神经系统无法准确地测量这些角度。然而相对深度,也就是说在沿着视线方向上对深度的排序,可以非常准确地被我们的眼睛所判断。例如,我们通常可以判断在同一等视面附近的两个目标哪个更靠近观察者,即便两个目标的视差只有几个弧秒(立体匹配敏感度的阈值),它与人眼可以测量的最小距离(单目敏感度的阈值)相匹配。

关于建立左右眼的对应,Julesz(1960)提出了如下问题:单目过程是不是双目融合的一个基本机制[局部的亮度模式(微模式)或者点的更高一级的组织形式(宏模式)在它们被融合之前,先识别成物体]? 双目过程是不是一个基本过程呢(两幅图像在这一过程中被合并到一个单一的视场中,所有后续的处理都在这个视场中发生)? 还是单目和双目的结合呢? 一些有趣的证据暗示双目机制是基本的。这里引证 Julesz 的话:"在航空侦察中有这样一个事实,由复杂背景掩蔽的物体往往很难被单目检测到,然而在双目情况下则会被轻而易举地检测到。"为了获得更多结论性的数据并解决这个问题,Julesz 引入了一个新的方法,随机点立体图(random dot stereogram),这是一对人工合成的图像,通过随机的向白色物体上喷洒一些黑点来获得,白色物体通常用一个小方盘叠在一个大方盘上(如图 11.6 所示)。Julesz 又指出:"用单目看每一幅图像都是完全随机的。但是用双目看时,图像对明显地给人以位于环境前(或环境后)方块的形象"。结论是清楚的:人类的双目融合过程不能用直接和实际视网膜相联系的外围(译者

注:子过程,即是一个基本过程不能划分成子过程)过程来解释。相反,它涉及中枢神经系统和一个想像中的超视网膜,这个超视网膜将左右眼的激励结合成单一的整体。

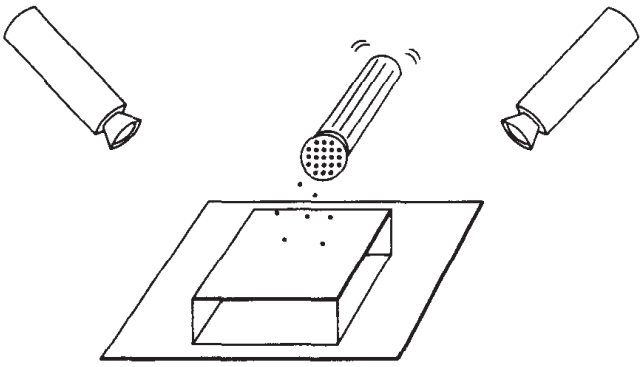


图 11.6　随机点立体图:在两个盘子上面撒(模拟的)胡椒粉

Julesz 提出的立体视觉的偶极(dipole)模型是一种按照协作方式(cooperative)工作的方法,相邻的匹配互相影响以避免二义性并促成对场景的全局分析。Marr 和 Poggio(1976)提出的方法也是按照协作方式工作的另一个方法,对于随机点立体图有很好的效果。他们的方法依靠三个约束:(a)一致性(compatibility)(黑色的点只能匹配黑色的点,或者更一般地,两个图像特征只有出自同一类物理特征才可能匹配),(b)惟一性(uniqueness)(一个黑点至多只能与另一幅图像中的一个黑点匹配),(c)连续性(continuity)(匹配点的视差在图像中绝大部分是平滑变化的)。已知在一对对应的外极线上有一些黑色点,Marr 和 Poggio 建立了一个图来体现可能的对应(如图 11.7 所示)。

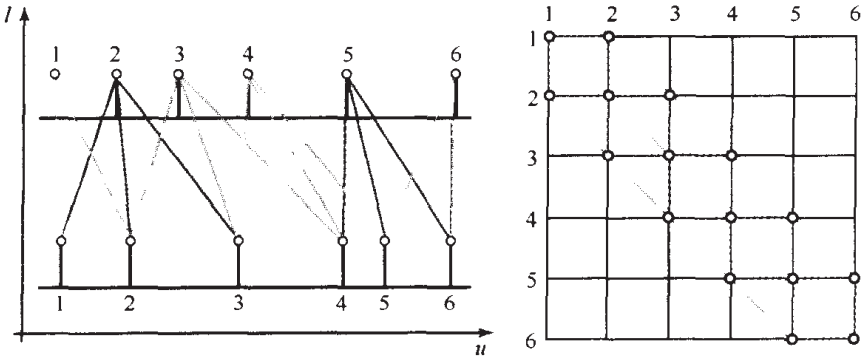


图 11.7　一个协作方式工作的立体分析方法:Marr-Poggio 算法(1976)。左图展示了两幅图像沿着同一扫描线的灰度轮廓。每个钉子状的图示代表一个黑点。连接两个黑点的线段意味着在一定的最大视差范围可能的匹配。这些匹配在右图中也有显示,右图是这些节点形成的一个图。垂直和水平弧连接的节点表示了左图或右图中同一个点。对角线弧连接了具有相似视差的节点

图中节点是在一定视差范围内的黑色点对,这反映了一致性约束;垂直和水平的弧表示受抑制的连接,这体现了惟一性约束(任何两点间的匹配不能是两个左图点的匹配——水平抑制,也不能是两个右图点的匹配——垂直抑制);对角线弧表示了受鼓励的连接,这体现了连续性约束(即优先视差相似的匹配)。

在这个方法中,对每个节点对应一个属性的测量。在一定的视差范围内,将每对可能的匹配初始为 1。匹配过程是迭代的、并行的,在每次迭代过程中,每个节点根据邻居节点值的加

权和赋值。受鼓励的连接权重赋值为 1,受抑制的连接权重赋值为 $-w$(这里 w 是一个合适的权重系数)。如果相应的加权和超过某个阈值,这个节点赋值为 1,否则就赋值为 0。这个方法对于随机点图(见图 11.8)非常可靠,但是对于自然图像效果并不理想。在下一节中,我们将回到计算机视觉的讨论并提出一些对于真实图像更有效的技术。尽管如此,原始的 Marr-Poggio 算法及其实现提供了立体视觉的早期范例,并为随机点立体图的融合提供了可能,至今这个方法还在吸引着人们的兴趣。

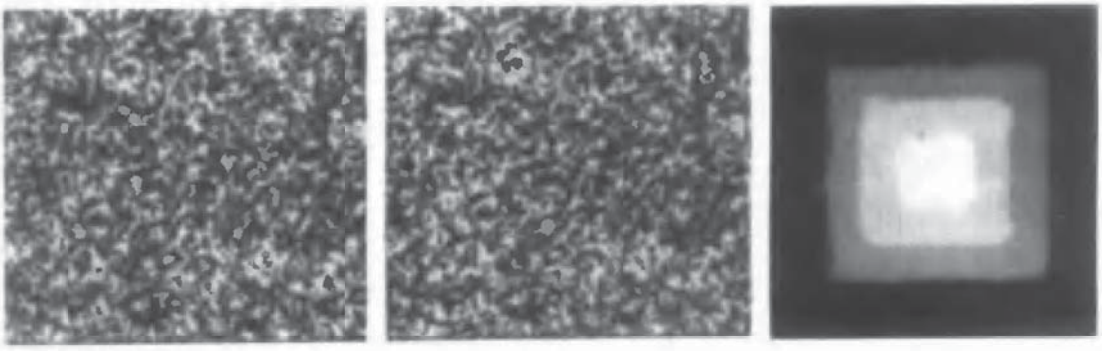


图 11.8 按照从左到右的顺序:首先是一对随机点立体图描述了 4 个不同深度的平面(就像一个婚礼蛋糕塔),然后是经过14次Marr-Poggio合作算法迭代后得到的视差图

11.3 双目融合

11.3.1 相关

通过比较可能的匹配点周围的灰度情况,相关(correlation)方法可以找到像素级的图像对应。在解决双目融合问题上,它属于本章前面提到的第一部分技术。(译者注:即融合两个或多摄像机观察到的特征。)(Kelly 等, 1977;Gennery,1980)。具体来说,让我们考虑一个校正过的图像对以及第一幅图像中一个点(u,v)。设以(u,v)为中心大小为 $p=(2m+1)\times(2n+1)$的窗口对应这样一个向量 $\boldsymbol{w}(u,v)\in\mathbb{R}^p$,该向量是通过扫描窗口每一行数值而获得(实际上扫描顺序不重要只要这个顺序是固定的)。已知在第二幅图像中存在一个匹配点$(u+d,v)$,那么可以建立第二个向量 $\boldsymbol{w}'(u+d,v)$并定义相应的归一化相关函数(normalized correlation function)如下:

$$C(d)=\frac{1}{|\boldsymbol{w}-\bar{\boldsymbol{w}}|}\frac{1}{|\boldsymbol{w}'-\bar{\boldsymbol{w}}'|}\left[(\boldsymbol{w}-\bar{\boldsymbol{w}})\cdot(\boldsymbol{w}'-\bar{\boldsymbol{w}}')\right]$$

为了简便起见,其中索引 u、v 和 d 被省略,$\bar{\boldsymbol{a}}$ 代表 $\boldsymbol{a}$ 的均值(如图 11.9 所示)。

很明显,归一化相关函数 C 的范围在 -1 和 $+1$ 之间。当两个窗口亮度值之间的关系形成仿射变换时,即 $I'=\lambda I+\mu$,其中 λ 和μ为常数且 $\lambda>0$,归一化相关函数达到最大值(见习题)。换句话说,当两个图像块相差一个偏移常量和一个比例因子时,该函数达到最大值。立体匹配可以通过在一定视差范围内寻找 C 函数(即归一化相关函数——译者注)的最大值来获得[①]。

① 在某些场合下,归一化相关函数相对于亮度函数仿射变换的不变性会给基于相关的匹配技术带来一定程度的鲁棒性,这些场合包括被观测表面不完全是朗伯表面(Lambertian surface)(译者注:常用的一种漫反射模型),或者两个摄像机有不同的增益,或者镜头具有不同的 f 值。

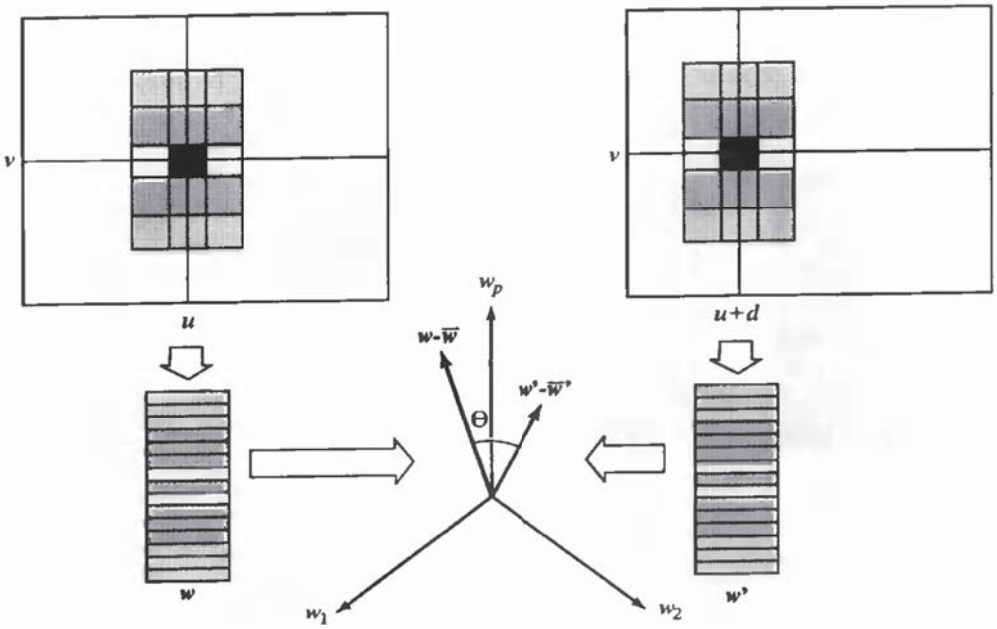


图 11.9　沿着相应的外极线两个 3×5 大小窗口的相关过程。第二个窗口位置和第一个相距 d，两个窗口对应的向量 w 和 w' 位于空间 $\mathbb{R}^{15}$。相关函数测量了向量 $w-\bar{w}$ 和向量 $w'-\bar{w}'$ 之间夹角的余弦值，这两个向量是 w 和 w' 和相应均值的差

这里让我们讨论一下基于相关的方法。首先，我们可以容易获知(见习题)，最大化归一化相关函数等价于将下式最小化：

$$\left|\frac{1}{|w-\bar{w}|}(w-\bar{w})-\frac{1}{|w'-\bar{w}'|}(w'-\bar{w}')\right|^2$$

或等价于将经过归一化的两个窗口像素值的平方差最小化。其次，尽管在一定视差范围内计算每个像素的归一化相关函数值计算量非常大，但是通过迭代技术则可以有效地实现(见习题)。最后，用于建立立体对应的基于相关的方法的最主要问题是，隐含地假设了被观测表面(局部地)平行于两个图像平面(见图 11.10)。

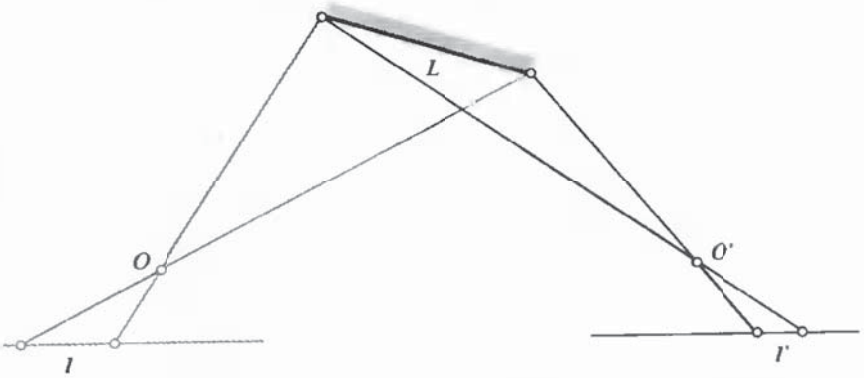


图 11.10　一个(倾斜)表面的透视缩略图由摄像机的位置决定：$l/L \neq l'/L$

这里我们推荐一个两步算法，首先用初始估计的视差扭转相关窗口以补偿两幅图像中由于透视效果造成的不均衡。图 11.11 给出了一个例子，其中对于左图中的每一个矩形，利用矩

形中心的视差及其变化率，在右图中定义了扭转窗口(Devernay 和 Faugeras，1994)。利用最优化方法找到合适的视差值及其变化率，使得左图矩形和右图窗口之间的相关函数达到最大值，右图中的值通过插值法获得。

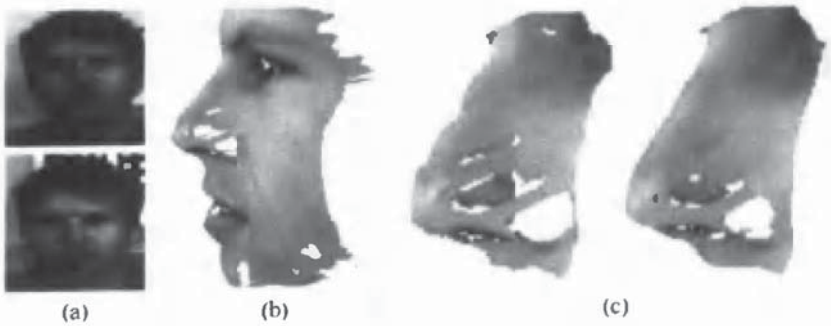


图 11.11 基于相关的立体匹配：(a) 一对立体图像对；(b) 被重建表面的纹理映射图；(c)普通相关方法(左)和经过求精的相关方法获得的鼻部重建结果的比较。显然后者的结果要好

11.3.2 多尺度的边缘匹配

倾斜的表面给基于相关的方法带来了问题，其他有关相关方法的争论可以在 Julesz(1960)和 Marr(1982)中找到。这些争论指出，应在多个尺度上寻求图像中的对应关系，在(很可能存在的)明显的图像特征如边缘上的匹配，要比在未经加工的像素灰度上的匹配更可靠。这些原则在 Marr 和 Poggio(1979)提出的算法中得到体现，如算法 11.1 所示。

算法 11.1 Marr-Poggio(1979)多尺度双目融合算法

1. 用$\nabla^2 G_\sigma$ 和(校正过的)两幅图像进行卷积，卷积核的标准方差递增，即 $\sigma_1 < \sigma_2 < \sigma_3 < \sigma_4$。
2. 沿着卷积图像的水平扫描线找 Laplacian 过零点。
3. 对每个滤波器的尺度 σ，在$[-w_\sigma, +w_\sigma]$的视差范围内，匹配梯度大小相当、方向相近的过零点，其中 $w_\sigma = 2\sqrt{2}\sigma$。
4. 在匹配点的周围，使用在较大尺度上找到的视差来平移图像，使得在更小尺度上不匹配的区域可以对应起来。

在$[-w_\sigma, w_\sigma]$视差范围内，这个算法寻找每个尺度上的匹配点，其中 $w_\sigma = 2\sqrt{2}\sigma$ 是$\nabla^2 G_\sigma$ 滤波器中心为负的部分。这是出于心理学和统计学上的考虑。特别地，假设卷积图像是高斯白噪声过程，Grimson(1981 a)证明了当匹配特征相互之间的方向在 30 度以内，在$[-w_\sigma, +w_\sigma]$视差范围内过零点错误匹配的发生概率仅为 0.2。在匹配范围内还可能存在多个匹配，一个简单的方法可以用来消除多个可能匹配[详细内容见 Grimson(1981 a)]。当然，将搜索限制在$[-w_\sigma, +w_\sigma]$的范围内，使得算法无法找到视差不在该范围内的正确的过零点匹配。由于 w_σ 正比于尺度 σ，正是在若干这样的尺度上搜索匹配，在大尺度上搜索到的视差可以控制眼球的运动(或者等价的图像偏移)，这个运动可以将有大尺度视差的过零点对转移到较小尺度上可以匹配的范围内。这个过程发生在算法 11.1 的第 4 步中，图 11.12 对这一点进行了说明。一旦

匹配找到了，相应的视差可以存放在一个缓冲区中，这个缓冲区被 Marr 和 Nishihara(1978)称为两维半草图($2\frac{1}{2}$ dimensional sketch)。Grimson(1981a)给出了这个算法的实现，并且广泛地应用于测试随机点立体图和自然图像。另一个例子在图 11.12(下)中给出。

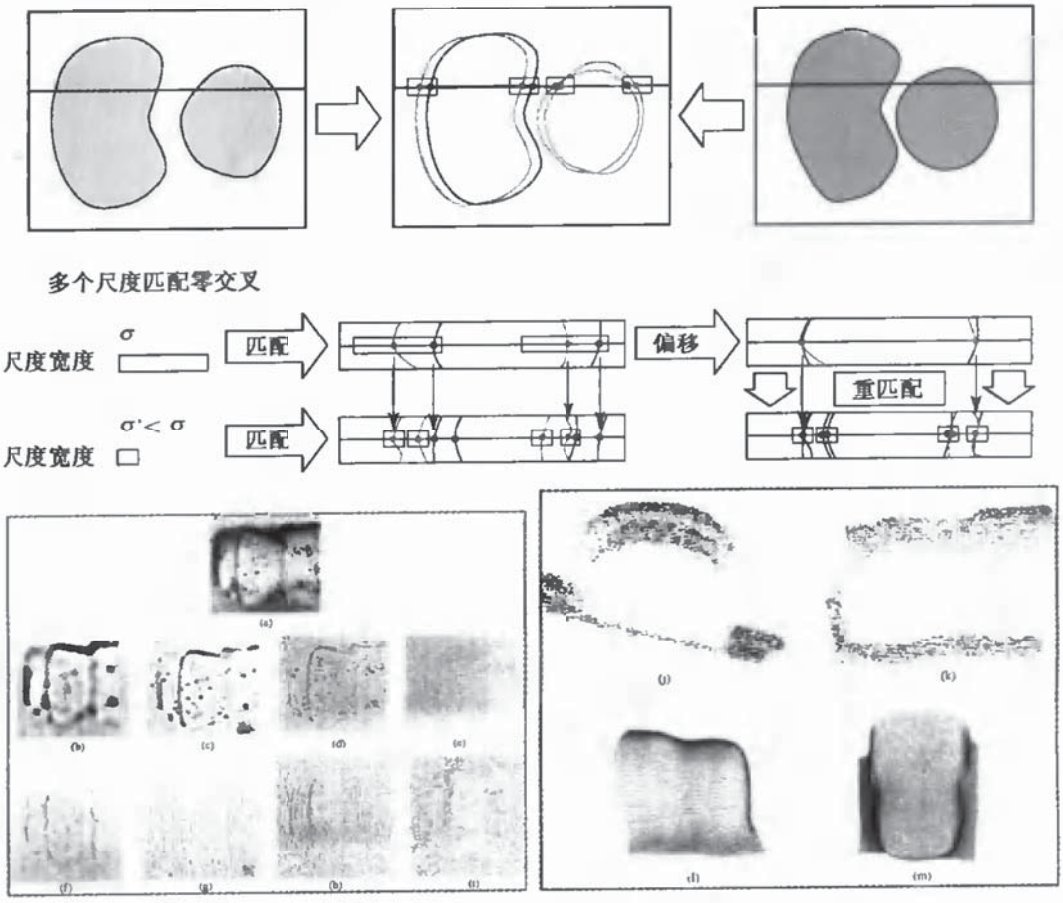


图 11.12　上：单尺度匹配；中：多尺度匹配；下：结果。下左：输入数据(包括一个输入图像，4个$\nabla^2 G_\sigma$滤波器得到的结果以及相应的过零点)；下右：由匹配过程建立的深度图的两个视图以及通过内插重建点得到的物体表面的两个视图

11.3.3 动态规划

下面我们考虑两个顺序，一个是沿着一对外极线，匹配图像特征的顺序；另一个是沿着外极平面与被观测物体表面交线，匹配表面特征的顺序。通常我们有理由认为这两个顺序是相反的(如图 11.13 左图所示)。这是 20 世纪 80 年代提出的(Baker 和 Binford，1981；Ohta 和 Kanade，1985)所谓顺序性约束(ordering constraint)。很有趣的是，在真实场景中，上述约束不一定能够满足。特别地，例如当一个小物体挡住了部分大物体(如图 11.13 右图所示)或者当涉及透明物体时，上述顺序性约束都可能不成立。但是至少在机器人视觉领域，涉及透明物体的情况很少见。

尽管有这些限制，顺序性约束还是一个合理的约束，它可以用来设计有效的基于动态规划(dynamic programming)的算法(Forney，1973；Aho 等，1974)以建立立体对应(见图 11.14)。特别地，让我们假设在对应的外极线上有一些特征点(或说边元)。这里我们的目标是匹配沿着两

个灰度轮廓分开这些特征点的间隔(如图 11.14 左图所示)。根据顺序性约束,尽管当遮挡或噪声使得对应关系丢失时,两幅图像中的特征点的间隔会退化成一个点,但是特征点的顺序必须是一样的。

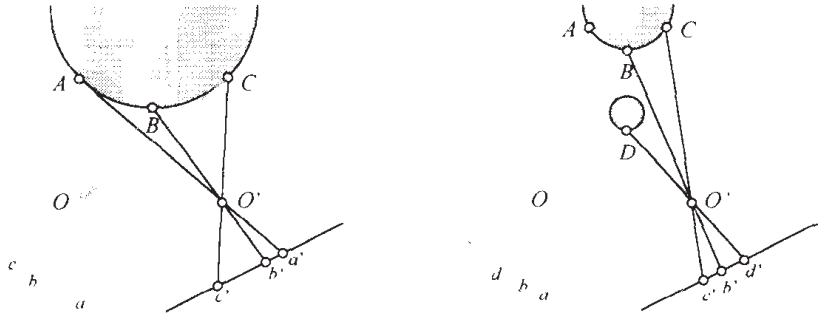


图 11.13 顺序性约束。在(通常)情况下,如左图所示,沿着两个(同一方向的)外极线上特征点的顺序是一样的。在右图所示的情况中,一个小物体位于大物体的前方。部分表面上的点在一个图像中是不可见的(例如,A在右图中是不可见的),图像点的顺序在两幅图像中是不一样的:b在左图中位于d的右边,但是b'在右图中位于d'的左边

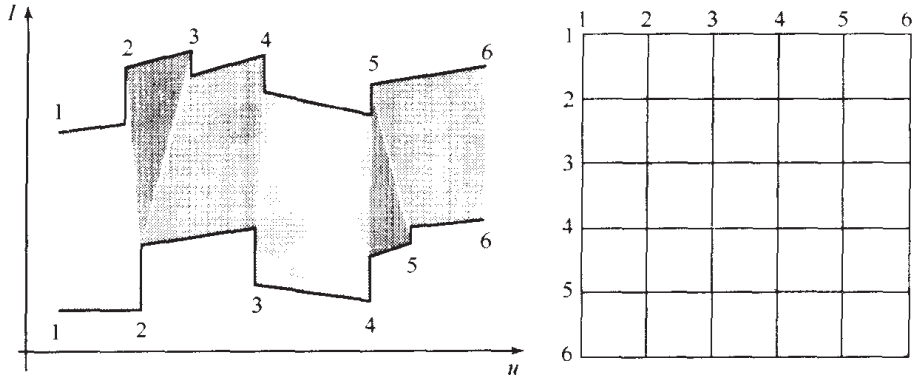


图 11.14 动态规划与立体视觉:左图显示了沿着匹配的外极线的两个灰度轮廓。多边形连接的两个轮廓意味着连续间隔上的匹配(部分匹配的间隔长度为零)。右图表示了和左图图形中相同的信息。当灰度轮廓上的间隔(i,j)和间隔(i',j')相互匹配时,用一段弧(粗线段)连接两个节点(i,i')和(j,j')

在这个背景下,我们可以重新叙述匹配问题为:在一个图上优化路径代价,这个图中的节点对应左右图像特征对;图中的弧表示在左右图像灰度轮廓间隔之间的匹配,这些间隔和对应节点特征相联系(如图 11.14 中右图所示)。一个弧的代价衡量了两个对应间隔之间的差异(例如灰度均值的平方差)。这个优化问题可以使用动态规划算法来解决,如算法 11.2 中所示。

算法 11.2 的计算复杂度为 $O(mn)$,其中 m 和 n 分别代表在左右对应扫描线上边缘点的数量①。Baker 和 Binford(1981)实现了这个方法的一个变种,他们将一个从粗到细的扫描线内的搜索过程和一个按合作方式工作的过程结合在一起,以增强扫描线间的一致性,Ohta 和 Kanade(1985)使用动态规划方法进行扫描线内和扫描线间的优化,后者的优化过程是在一个三维搜索空间中进行。图 11.15 给出了来自 Ohta 和 Kanade ,1985 的结果。

① 这个版本的算法假设所有边缘都是匹配的。考虑到噪声和边缘检测的噪声,允许匹配算法跳过有限的边缘也是合理的,但是这并不改变算法的渐近复杂度(Ohta 和 Kanade,1985)。

算法 11.2　用于在两条对应的扫描线上建立立体对应的动态规划算法，两条扫描线上分别有 m 和 n 个边缘点(为了方便起见，扫描线的端点也被包含在内)。两个辅助函数：下邻居节点函数 Inferior-Neighbors(k,l)返回节点(k,l)的邻居节点(i,j)的列表，要求 $i \leqslant k$ 且 $j \leqslant l$；弧代价函数 Arc-Cost(i,j,k,l)评价并返回匹配间隔(i,k)和(j,l)的代价。为保证算法的正确，最优代价函数 $C(1,1)$应该初始化为零。

```
    % 在所有节点(k,l)中按照升序循环。
for k = 1 to m do
  for l = 1 to n do
    %初始化最优代价函数 C(k,l)和回溯指针 B(k,l)。
    C(k,l)←+∞; B(k,l)←nil;
    %在(k,l)的所有下节点(i,j)中循环。
    for(i,j)∈Inferior-Neighbors(k,l) do
      %计算新的路径代价并在必要情况下更新回溯指针
      d←C(i,j)+Arc-Cost(i,j,k,l);
      if d<C(k,l) then C(k,l)←d; B(k,l)←(i,j) endif;
      endfor;
    endfor;
  endfor;
%通过跟踪回溯指针从(m,n)建立最优路径。
P←{(m,n)};(i,j)←(m,n);
while B(i,j)≠nil do (i,j)←B(i,j); P←{(i,j)}∪P endwhile。
```

图 11.15　五角大楼的两幅图像和通过 Ohta 和 Kanade(1985)动态规划算法得到的视差的等比例图

11.4　使用多个摄像机

11.4.1　三个摄像机

增加第三个摄像机可以消除(大部分)由双目图像点匹配造成的不确定性。本质上，第三

幅图像可以用来检查前两幅图像中假定的匹配(如图 11.16 中所示):和前两幅图像中匹配点对应的三维空间点首先被重建,然后再投影到第三幅图像。如果在第三幅图像的再投影点周围没有相容的点,那么这个匹配一定是错误的匹配。实际上,重建/再投影过程可以避免,如果我们注意到,在第 10 章中,已知三个弱标定(已经足够了)摄像机和空间某点的两个像点,总可以通过将相应的外极线取交来预测该点在第三幅图像中的位置。

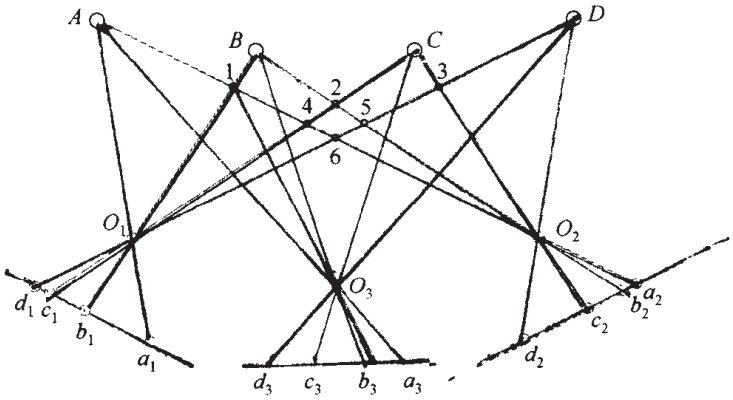


图 11.16 灰色小圆点意味着左右图像中 4 个点的不正确重建。在中间增加一个摄像头消除了匹配的不确定性:所有和真实点对应的射线都不穿过这6个点中的任何一个。换句话说,在前两幅图像中的点匹配可以通过将三维空间点再投影到第三幅图像中来检查。例如,在 b_1 和 a_2 之间的匹配很明显是错误的,因为在第三幅图像中靠近由假定重建点产生的再投影点周围没有特征点,假定重建点在图中标记为1

11.4.2 多个摄像机

在大多数三目立体算法中,首先利用两幅图像假定可能的对应,然后用第三幅图像来接受或拒绝这些对应。与此不同的是,Okutami 和 Kanade(1993)提出一个多个摄像机的算法,其中同时利用所有图像来搜索匹配。基本想法简单而精巧:假设所有图像都是被校正过的,将搜索正确的视差的操作转换为搜索正确的深度或者深度的倒数。当然,对于每个摄像头来说深度的倒数和视差成正比,然而视差由于摄像机的不同而变化,因此深度的倒数被用来作为一个通用的搜索索引。选择第一幅图像作为参考,Okutami 和 Kanade 将与所有其他摄像机相关的平方差加到一个全局评价函数 E 中(就像我们以前说明的,这当然是等价于向全局评价函数增加了与图像联系的相关函数)。

评价函数 E 是深度倒数的函数,图 11.17 画出了对于不同数量的摄像头该函数的函数值。我们应该注意到相应的图像包含了一个重复性的图案,只用两个或三个摄像机并不能产生一个单一的明确的最小值。然而,增加更多的摄像机提供了一个清楚的最小值对应于正确的匹配。

图 11.18 显示了 10 张校正过的图像序列并且根据上述算法给出了表面重建图。

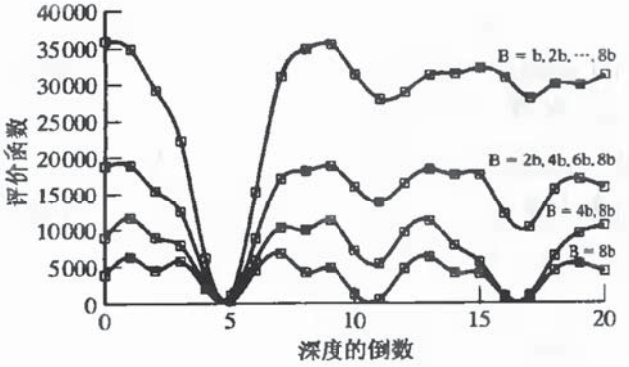


图 11.17　融合多个视图信息：在这里画出的平方差的和是深度倒数的函数，这里画了不同输入图像数量的情况。这些数据是从图11.18图像顶部的某条扫描线上获得的，它们的灰度值几乎是周期性的。这个图清楚的显示出函数的最小值随着图像数量的增加变得越来越清楚

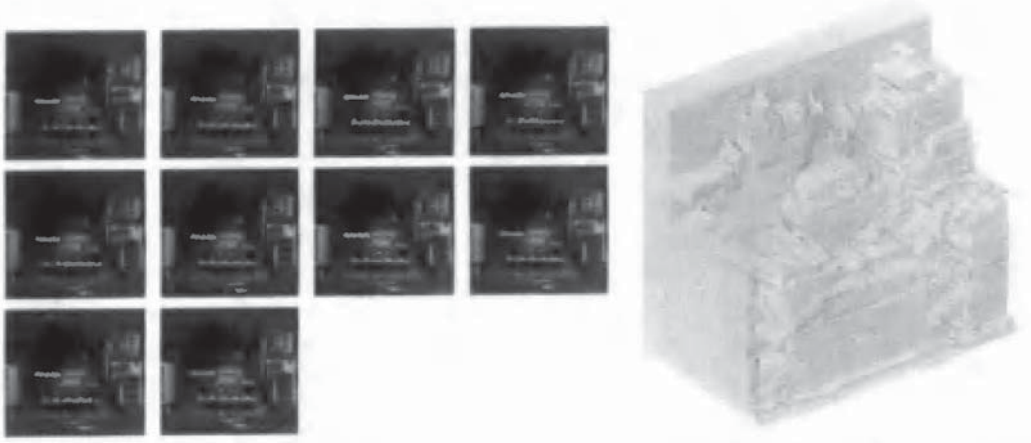


图 11.18　10 幅图像序列和相应的重建图像。靠近图像顶部的网格板造成了几乎周期性的亮度信号，从而带来立体视觉中的不确定性，如图11.17中看到的

11.5 注释

视差引发人类立体视觉的事实首先由 Wheatstone(1838)发明的立体照相机所证实。此后，Dove(1841)从电火花产生的光亮太短以致无法引起人眼聚视出发，证实了不需要眼球的移动视差足以进行立体视觉的事实(Helmholtz, 1909, 第 455 页)。人眼的立体视觉在 Helmholtz (1909)的经典书籍中被进一步讨论，对于任何对这个领域感兴趣的人，这都是一本很好的书。同样在 Julesz(1960, 1971), Frisby(1980)以及 Marr(1982)的书籍中都有对人眼立体视觉的论述。由于篇幅的限制，在本章中没有介绍的人类双目立体感知理论包括 Koenderink 和 Van Doorn (1976*a*), Pollard 等(1970), McKee 等(1990)，以及 Anderson 和 Nakayama(1994)等人的理论。

关于机器立体视觉，在 Grimson(1981*b*), Marr(1982), Horn(1986)以及 Faugeras(1993)的书籍中有很好的论述。Marr 侧重于从计算角度讲述人类立体视觉，然而 Horn 的论述强调人工立体视觉系统中的几何成像学。Grimson 和 Faugeras 侧重从几何和代数的角度讲述立体视觉。与立体匹配相关的约束在(Binford, 1984)中有所论述。在双目立体视觉中有关线段匹配的早期技术包括 Medioni 和 Nevatia(1984)以及 Ayache 和 Faugeras(1987)。三目融合的算法包括

Milenkovic和 Kanade(1985), Yachida 等(1986), Ayache 和 Lustman(1987)以及 Robert 和 Faugeras(1991)。正如在 Robert 和 Faugeras(1991)以及习题中指出的,第 10 章引入的三焦张量也可以用来预测在一幅图像中沿着某条图像曲线的切线和曲率,这需要已知在另一幅图像中相应的数值。这个事实可以用来从三幅图像中有效地匹配和重建曲线。

正如前面提到的,图像边缘通常作为建立双目对应的基础,至少一部分原因是它们可以被看做为图像处理中的物理属性,对应于如反照率、颜色或遮挡的边界。在立体匹配算法中一个点很少被考虑,这是因为,对于有光滑表面的物体,立体融合总是在沿着它们(见图 11.19)的轮廓线上失效。实际上,在这种情况下,相应的图像边缘是与视点相关的,匹配这些边缘往往造成错误的重建。正如在 Arbogast 和 Mohr(1991), Vaillant 和 Faugeras(1992), Cipolla 和 Blake(1992)以及 Boyer 和 Berger(1996)中指出的,在这种情况下,三个摄像机足以重建一个局部二阶的表面模型。

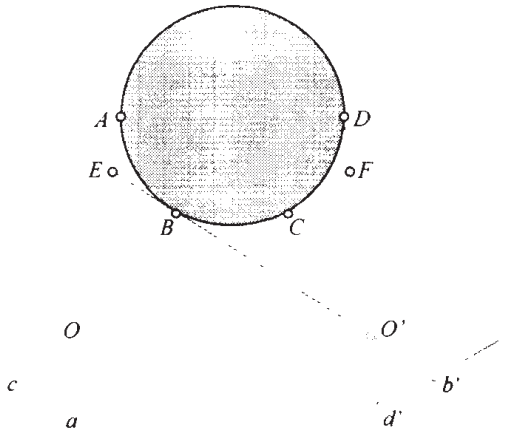


图 11.19 立体匹配在光滑物体边界处失效:对于短基线的情况,点对(c, d')和(a, b')很容易被大部分基于边缘的算法匹配到一起,在三维重建中产生了假想的点 F 和 E

目前很难说在基于特征的匹配和基于灰度级的匹配之间,哪个更胜一筹。由于靠近表面纹理前者准确,但是只产生稀疏的测量集,后者在均匀的区域下结果不好但是可以在有纹理的区域处提供稠密对应。在这种情况下,从稀疏采样点插值以获得稠密表面的问题就显得重要,尽管本章中几乎没有提及这个问题。有兴趣的读者可以参考 Grimson(1981*b*)和 Terzopoulos(1984)以获得详细的内容。

由于篇幅限制,我们没有讨论的另一个立体视觉方法,它涉及到高层次的插值过程——例如,预测/校正方法用于图形学的图像描述(Ayache 和 Faverjon, 1997),或者金字塔技术用于匹配两幅图像中的曲线、表面和体积(Lim 和 Binford, 1988)。

本章展示的所有算法(暗含地)假设了被融合的图像是非常相似的,这等价于考虑一个短基线的情况。一个用于处理长基线的有效算法可以在 Pritchett 和 Zisserman(1998)中找到。另外,基于模型的方法将在第 26 章中讨论。

最后一点,本章的讨论仅限于固定内外参数的立体摄像机。主动视觉(active vision)关心如何建立一个能够动态修改这些参数的视觉系统(例如,修改摄像机的焦距,聚散角以及在感知和机器人任务中如何利用这些能力;见 Aloimonos 等, 1987; Bajcsy, 1988; Ahuja 和 Abott 1993; Brunnström 等, 1996)。

习题

11.1　证明:对于校正过的图像对,在第一个摄像机的归一化坐标系内,P 点深度可以表示为 $z = -B/d$,其中,B 是基线,d 是视差。

11.2　利用视差的定义证明立体重建的准确度是基线和深度的函数。

11.3　给出两维条件下眼睛会聚时的重建公式。

11.4　给出一个算法,用以产生一个模糊的随机点立体图,要求这个随机点立体图可以描述两个不同的盘子浮在第三个盘子上。

11.5　证明当两个窗口的图像亮度可以用一个仿射变换 $I' = \lambda I + \mu$ 相联系时,相关函数达到最大值 1,其中,λ 和 μ 为某个常数,$\lambda > 0$。

11.6　证明对于零均值和单位 Frobenius 范数的图像,相关与平方差的和是等价的。

11.7　迭代计算相关函数

(a) 证明 $(\boldsymbol{w} - \bar{\boldsymbol{w}}) \cdot (\boldsymbol{w}' - \bar{\boldsymbol{w}}') = \boldsymbol{w} \cdot \boldsymbol{w}' - (2m+1)(2n+1)\bar{I}\,\bar{I}'$。

(b) 证明灰度的均值 $\bar{I}$ 可以迭代计算,并估计步进计算的计算量。

(c) 将先验计算推广到构建相关函数的所有元素,并估计对一对图像进行相关运算的整体计算量。

11.8　说明对于校正图像,如何使用视差函数的一阶扩展,投影右图窗口使得其对应于左图中的矩形区域。说明在这种情况下如何计算相关,使用插值来估计右图中相对于左图中心的像素值。

11.9　说明如何使用三焦张量预测沿着一条图像曲线的切线,已知在另外两幅图像中该曲线的切线测量。

编程作业

11.10 实现校正过程。

11.11 实现一个基于相关的立体分析算法。

11.12 实现一个多尺度的立体分析算法。

11.13 实现一个基于动态规划的立体分析算法。

11.14 实现一个三目立体视觉算法。

第 12 章　从运动估计仿射模型

本章还要再次分析从多幅图像估计景物三维深度的问题。在三维重构中，摄像机是标定过的，它们的内参数和外参数都已知，由外参数就可以得到它们在世界坐标系中的位置。这就大大简化了重构过程，则我们的重点放在了立体视觉系统中的双目（三目）融合问题上。本章要讨论一个更复杂的问题，摄像机的位置甚至是内参数都是事先不知道的，而且会随着时间变化。这种情况一般出现在用手持摄像机录像的情况，在摄像过程中可能重新对焦，是基于图像的绘制（image-based rendering）中的典型问题，可以获取物体形状并生成其他视角的视图（第 26 章）。相关的应用还有动态视觉系统，它的标定参数会动态变化，以及星际机器人，在起飞和降落时由于加速度的影响，它们的参数也可能产生变化。在机器人导航中，恢复摄像机位置是和估计场景形状同样重要的任务。

本章不考虑匹配问题，假设 m 幅图像中的 n 个投影点是已经匹配好的[①]。我们关注的是“从运动得到结构”的纯几何问题，从图像匹配估计匹配点的三维坐标（例如，在场景结构的坐标系内），以及相对于每个摄像机对应的投影矩阵。本章中关注的是景物深度变化相对较小的情况，可以用第 1 章和第 2 章中介绍的仿射模型近似透视投影，在下一章中再介绍完整的没有简化的运动重构问题。具体来讲，给定 n 个点 $P_j(j=1,\cdots,n)$ 在 m 个仿射摄像机上的成像点，这 mn 个点的坐标（非齐次）为 $\boldsymbol{p}_{ij}$，仿射方程（2.19）可以写为

$$\boldsymbol{p}_{ij}=\mathcal{M}_i\begin{pmatrix}\boldsymbol{P}_j\\1\end{pmatrix}=\mathcal{A}_i\boldsymbol{P}_j+\boldsymbol{b}_i\ ,\text{其中 } i=1,\cdots,m\quad;\qquad j=1,\cdots,n \tag{12.1}$$

“从运动估计仿射结构”问题可以定义为估计 m 个 2×4 矩阵 $\mathcal{M}_i=(\mathcal{A}_i\quad \boldsymbol{b}_i)$ 和从 mn 个图像坐标中估计在某个固定世界坐标系中，n 个点 $\boldsymbol{P}_j$ 的坐标 P_j。

若投影矩阵 $\mathcal{M}_i$ 可以取任意形式（如，摄像机内外参数未知，第 2 章），方程（12.1）给出了 $2mn$ 条约束，而 $\mathcal{M}_i$ 和点 $\boldsymbol{P}_j$ 的未知数个数为 $8m+3n$。当 m 和 n 足够大时，$2mn$ 大于 $8m+3n$，也就是说，有足够的参考点和足够的视角，就可以通过第 3 章介绍的最小二乘法估计景物结构和运动参数。但是要明确一点，若 $\mathcal{M}_i$ 和 $\boldsymbol{P}_j$ 是方程（12.1）的一个解，则下面的 $\mathcal{M}'_i$ 和 $\boldsymbol{P}'_j$ 同样是解，

$$\mathcal{M}'_i=\mathcal{M}_i\mathcal{Q}\quad\text{和}\quad\begin{pmatrix}\boldsymbol{P}'_j\\1\end{pmatrix}=\mathcal{Q}^{-1}\begin{pmatrix}\boldsymbol{P}_j\\1\end{pmatrix} \tag{12.2}$$

$\mathcal{Q}$ 可以是任意仿射矩阵——它可以写为（参考第 2 章和下一节）

$$\mathcal{Q}=\begin{pmatrix}\mathcal{C}&\boldsymbol{d}\\\boldsymbol{0}^T&1\end{pmatrix}\quad\text{且}\quad\mathcal{Q}^{-1}=\begin{pmatrix}\mathcal{C}^{-1}&-\mathcal{C}^{-1}\boldsymbol{d}\\\boldsymbol{0}^T&1\end{pmatrix} \tag{12.3}$$

其中，$\mathcal{C}$ 是一个非奇异的 3×3 矩阵，$\boldsymbol{d}$ 是 $\mathbb{R}^3$ 上的一个向量。换句话说，“从运动估计结构”仿射

---

① 连续运动序列以及场景多个视角图像的匹配问题在第 17 章和第 23 章介绍。

方程的解在仿射意义下是惟一的。若用 12 个参数表示仿射变换，只要 $2mn \geqslant 8m + 3n - 12$ 就可以得到有限个解。对于 $m = 2$，这说明 4 个对应点就可以（在仿射变换意义下）确定两个投影矩阵和其他点的三维坐标。这在 12.2 节将详细证明。

若已知摄像机内参数，则可以认为对应的标定矩阵是单位阵，投影矩阵的参数 $\mathcal{M}_i = (\mathcal{A}_i \quad \boldsymbol{b}_i)$ 必须满足另外的约束。例如，按照式(2.20)，（已标定的）弱透视投影摄像机对应的矩阵 $\mathcal{A}_i$ 由旋转矩阵的前两行除以对应点的深度得到。在 12.4 节将讲到，若有足够多的图像，可以用这种约束来去除仿射歧义性。仿射的“运动恢复结构”需要分成两个步骤进行：(a)第一步是用至少两幅图像建立场景的惟一的（仿射意义下）三维描述，称为仿射形状；(b)用其他的视图以及由已知的摄像机参数和确定的仿射模型约束确定场景惟一的刚性欧几里得结构。以上方法的第一步得到了解的基础部分：仿射模型是场景的完整的三维描述，在第 26 章将讲到，不需要其他的信息就可以用它合成任何角度的图像。第二步只是把这个模型表现在一个欧几里得空间里（例如，用一个仿射变换来表示场景，并把它映射到一个欧几里得空间上）。

利用三个或更多的图像后，“运动估计结构”问题是超定的，可以得到更健壮的最小二乘解。因此，本章的大部分都是介绍从多幅图像恢复场景的仿射模型，最后介绍利用物体的不同运动把数据点分割为物体的方法。

12.1　仿射几何基础

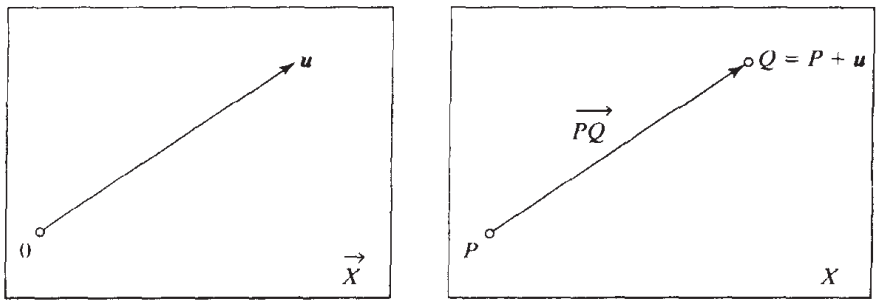
首先介绍一些仿射几何学的基础知识，其中相应的几何及代数工具可用来描述和证明仿射投影模型的一些性质。它们也是本章“运动估计结构”算法中的基本模块。

Snapper 和 Troyer(1989)指出，简单地讲，仿射几何就是把欧几里得几何学中的长度、面积和角度测量都去掉以后剩下的东西，但是平行的概念和直线上点的比例仍然保留。这里不能给出一个严格的仿射几何学的定义。我们在本章中只以非正式的定义引入仿射模型，并对以前提过的仿射模型基础进行回顾。在阅读下一节前，读者应已具有重心组合、仿射坐标系和仿射变换等相关知识。

12.1.1　仿射空间和重心组合

实仿射空间由点集 X、实向量空间 $\vec{X}$，以及从 $\vec{X}$ 到 X 的作用 ϕ 组成。向量空间 $\vec{X}$ 称为仿射空间 X 的基础。可以把作用 ϕ 看成是一个群到一个集合的映射，把群中的元素映射到集合中的一个双向关系上。在这里，ϕ 作用到 $\boldsymbol{u} \in \vec{X}$ 上，得到 $X \rightarrow X$。则对 $\vec{X}$ 上的任意 $\boldsymbol{u}$、$\boldsymbol{v}$ 和 X 上一点 P，$\phi_{u+v}(P) = \phi_u \circ \phi_v(P)$，$\phi_0(P) = P$；且对 X 上任两点 P、Q，存在惟一的向量 $\boldsymbol{u}$，使得 $\phi_u(P) = Q$。由于这个定义看起来比较抽象，所以我们来看一个具体的例子。$\mathbb{E}^3$ 是一个常见的仿射空间，其中 X 是实际的点集，$\vec{X}$ 是 X 到自身的变换集合。另一个仿射空间是 X 和 $\vec{X}$ 都取 $\mathbb{R}^n$，作用 ϕ 是 $\phi_u(P) = P + \boldsymbol{u}$，其中 P 和 $\boldsymbol{u}$ 都是 $\mathbb{R}^n$ 上的元素，“+”是普通的向量加法。

例 12.1 $\mathbb{R}^2$ 向量空间作为一个仿射平面



若选择 $X = \vec{X} = \mathbb{R}^2$，可以认为$\mathbb{R}^2$ 是仿射空间。若已知 $P = (x, y)^T$ 和 $\boldsymbol{u} = (a, b)^T$，可得 $\phi_u(P) \stackrel{\text{def}}{=} P + \boldsymbol{u} = (x + a, y + b)^T$。若已知 $P = (x, y)^T$ 和 $Q(x', y')^T$ 令 $P + \boldsymbol{u} = Q$ 成立的惟一向量 $\boldsymbol{u}$ 是 $\boldsymbol{u} = Q - P = \overrightarrow{PQ} \stackrel{\text{def}}{=} (x' - x, y' - y)^T$。

以后我们把点 $\phi_u(P)$记为 $P + \boldsymbol{u}$，$\boldsymbol{u}$ 记为$\overrightarrow{PQ}$或等价的 $Q - P$(因为 $\phi_u(P) = Q$)。可以从下面的事实证明：选择一点 O 作为 X 的原点后，我们可以通过向量 $\boldsymbol{u} = \overrightarrow{OP}$确定任意的点 P，满足 $\phi_u(O) = P$。实际上，有

$$Q = P + \overrightarrow{PQ} \Longleftrightarrow \overrightarrow{OQ} = \overrightarrow{OP} + \overrightarrow{PQ} \quad 和 \quad Q - P = \overrightarrow{PQ} \Longleftrightarrow \overrightarrow{OQ} - \overrightarrow{OP} = \overrightarrow{PQ}$$

对初学者而言，引入坐标原点可以帮助他们建立仿射模型。但是有一点需要明确，点 $P + \boldsymbol{u}$ 和向量$\overrightarrow{PQ} = Q - P$ 并不依赖于原点的选取。符号“+”和“-”也只是一种记号，并不是在向量空间这个加法群中的通常含义。

虽然可以把一个向量“加”到一个点上，或者把两个点相“减”，但是把两个点相“加”或是把一个点“乘”一个比例因子都是不可能的。另一方面，可以定义点的带约束的“线性组合”：对 $m + 1$ 个点 $A_0, A_1, \cdots, A_m$ 和权重 $\alpha_0, \alpha_1, \cdots, \alpha_m$，满足 $\alpha_0 + \alpha_1 + \cdots + \alpha_m = 1$；对应的重心为

$$\sum_{i=0}^{m} \alpha_i A_i \stackrel{\text{def}}{=} A_j + \sum_{i=0, i \neq j}^{m} \alpha_i (A_i - A_j) \tag{12.4}$$

其中，j 是 0 到 m 之间的一个整数。等式右侧是通过一个向量(向量 $A_i - A_j$ 的线性组合)和一个点相加定义了一个新的点。显然，这个定义与 j 的选取无关(习题)，这保证了点 A_i($i = 0, \cdots, m$)在定义$\sum_{i=0}^{m} \alpha_i A_i$ 中是等价的。当 $\alpha_0 + \alpha_1 + \cdots + \alpha_m = 1$ 时，也可以通过引入原点 O 和利用$\sum_{i=0}^{m} \alpha_i \overrightarrow{OA_i} = \overrightarrow{OA_j} + \sum_{i=0, i \neq j}^{m} \alpha_i (\overrightarrow{OA_i} - \overrightarrow{OA_j})$来证明这个问题。但是按照等式(12.4)定义的重心组合更好一些，因为它显然是和原点的选取无关的。

$m + 1$ 个点的重心是一种常见的重心组合，这时所有权重都是 $1/(m + 1)$。任何总和是 1 的权重集都对应一个有效的重心组合。

12.1.2 仿射子空间和仿射坐标系

确定一点 O 和$\vec{X}$ 上的向量子空间 U 后，满足以下条件的点集 $O + U \stackrel{\text{def}}{=} \{O + \boldsymbol{u}, \boldsymbol{u} \in U\}$，称为 X 上的一个仿射子空间，它的维数等于相关的向量子空间的维数。若有 U' 是 U''的子空间或 U''是 U' 的子空间，则称两个仿射子空间 $O'' + U''$和 $O' + U'$平行。一维和二维的子空间分别称为线与平面。若 $\vec{X}$ 的维数为 n，则它的维数为 $n - 1$ 的仿射子空间称为超平面。仿射空间中

的仿射线、仿射平面和仿射超平面的含义与物理的三维空间或$\mathbb{R}^n$中线、平面和超平面的含义一致。

例 12.2　两个仿射子空间的交或者为空，或者是一个仿射子空间

设有仿射空间 X 下的两个仿射子空间 $Y' = O' + U'$ 和 $Y'' = O'' + U''$，它们的交记为 Z。设 P_0 是 Z 上的某点。由定义，有 $P_0 = O' + \boldsymbol{u}'_0 = O'' + \boldsymbol{u}''_0$，其中 $\boldsymbol{u}'_0$ 属于 U'，$\boldsymbol{u}''_0$ 属于 U''。同理，对 Z 上的任一个点 P，有 $P = O' + \boldsymbol{u}' = O'' + \boldsymbol{u}''$，其中 $\boldsymbol{u}'$ 属于 U'，$\boldsymbol{u}''$ 属于 U''。则可以推出

$$P = P_0 + \boldsymbol{u}' - \boldsymbol{u}'_0 = P_0 + \boldsymbol{u}'' - \boldsymbol{u}''_0$$

这说明(a) $\boldsymbol{u}' - \boldsymbol{u}'_0 = \boldsymbol{u}'' - \boldsymbol{u}''_0$ 是 $U' \cap U''$ 上的一个元素，(b) P 是 $P_0 + U' \cap U''$ 上的一个元素。反过来，$P_0 + U' \cap U''$ 上的任意一点 P 可以通过交集 $U' \cap U''$ 上的向量 $\boldsymbol{u}$ 表示，$P = P_0 + \boldsymbol{u}$；于是有

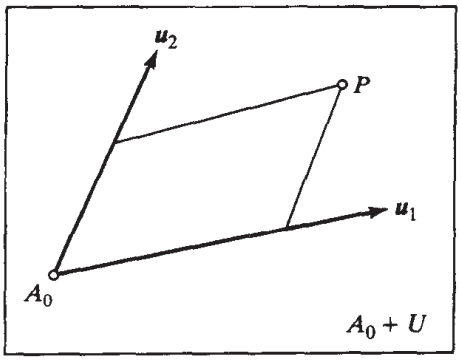
$$P = O' + (P_0 - O') + \boldsymbol{u} = O' + \boldsymbol{u}' + \boldsymbol{u} = O'' + (P_0 - O'') + \boldsymbol{u} = O'' + \boldsymbol{u}'' + \boldsymbol{u}$$

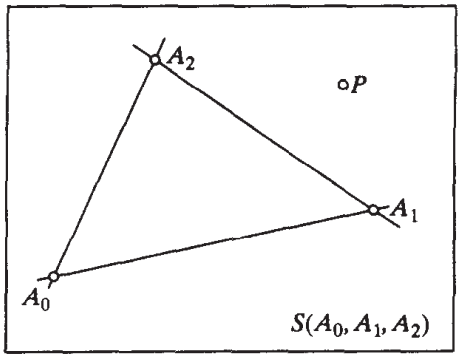
这说明 P 是 Z 中的一个元素。最后我们得到了 $Z = P_0 + U' \cap U''$。注意两个仿射空间的交可能是空的，例如两条平行线就不会相交。空间中两条错切的直线也不会相交，虽然它们并不彼此平行。

仿射空间也可以直接用点来定义：设 $S(A_0, A_1, \cdots, A_m)$ 表示 $m+1$ 个点 $A_0, A_1, \cdots, A_m$ 的所有线性组合构成的点集。可以验证 $S(A_0, A_1, \cdots, A_m)$ 实际上是一个仿射空间(见习题)，并且维数最高为 m(例如，两个点定义了一条直线，三个点[一般情况下]定义了一个平面，等等)。若 $m+1$ 个点不是处于一个维数小于等于 $m-1$ 的子空间内，则称它们独立。$m+1$ 个独立的点定义(张成)一个 m 维的子空间。

例 12.3　仿射空间的两个补充定义

设$\mathbb{R}^3$上不共线的三个点 A_0、A_1 和 A_2 构成的仿射空间。可以这样定义$\mathbb{R}^3$上的仿射平面 $\Pi = A_0 + U$：把 A_0 设为 O，向量平面 U 由两个向量 $\boldsymbol{u}_1 = \overrightarrow{A_0A_1}$ 和 $\boldsymbol{u}_2 = \overrightarrow{A_0A_2}$ 张成。





平面 Π 也可以看成$\mathbb{R}^3$上的仿射子空间 $S(A_0, A_1, A_2)$，Π 上的任意一点可以表示为 A_0，A_1，A_2 的重心组合。

仿射坐标系 $O + U$ 由 $O + U$ 上一点 A_0(称为原点)和 U 上的坐标系$(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$组成。$O + U$ 上一点 P 的仿射坐标定义为向量$\overrightarrow{A_0P}$在$(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$上的坐标。第 2 章用的(欧氏)点坐标和经典的欧氏几何都是一种仿射坐标系，理解这一点非常重要。但是用来定义那种坐标系的向量 $\boldsymbol{u}_i (i = 1, \cdots, m)$要满足额外的条件：它们必须相互垂直。一般的仿射坐标系不要求这种性质，实际上在仿射坐标系下长度和角度都是没有定义的。

例 12.4 仿射坐标变换

给定$\mathbb{E}^3$上的一个坐标系$(F)=(O,\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})$，对一个属于$\mathbb{E}^3$的点$P$，若有$\overrightarrow{OP}=x\boldsymbol{u}+y\boldsymbol{v}+z\boldsymbol{w}$，则可以用与第2章中同样的方式定义$P$的仿射坐标，${}^{F}P=(x,y,z)^T$。

对仿射空间$\mathbb{E}^3$给出两个仿射坐标系$(A)=(O_A,\boldsymbol{u}_A,\boldsymbol{v}_A,\boldsymbol{w}_A)$和$(B)=(O_B,\boldsymbol{u}_B,\boldsymbol{v}_B,\boldsymbol{w}_B)$，可以定义一个$3\times3$矩阵

$$ {}^{B}_{A}\mathcal{C}=\begin{pmatrix}{}^{B}\boldsymbol{u}_A & {}^{B}\boldsymbol{v}_A & {}^{B}\boldsymbol{w}_A\end{pmatrix} $$

其中，${}^{B}\boldsymbol{a}$表示向量$\boldsymbol{a}$在(向量)坐标系$(\boldsymbol{u}_A,\boldsymbol{v}_A,\boldsymbol{w}_A)$下的坐标向量。显然${}^{B}P={}^{B}_{A}\mathcal{C}\,{}^{A}P+{}^{B}O_A$，用齐次坐标表示，有

$$ \begin{pmatrix}{}^{B}P\\1\end{pmatrix}={}^{B}_{A}\mathcal{T}\begin{pmatrix}{}^{A}P\\1\end{pmatrix},\quad 其中\quad {}^{B}_{A}\mathcal{T}=\begin{pmatrix}{}^{B}_{A}\mathcal{C} & {}^{B}O_A\\ \mathbf{0}^T & 1\end{pmatrix} $$

上式与欧氏坐标系下的坐标系变换有明显的相似之处。但是这里的基向量并不构成正交基，因此${}^{B}_{A}\mathcal{C}$是一个普通的非奇异3×3矩阵，而不是一个旋转矩阵，${}^{B}_{A}\mathcal{T}$是一个仿射变换矩阵。

还可以有另一个定义n维坐标系X的方法。选取X上$n+1$个独立的点$A_0,A_1,\cdots,A_n$。Y上一点P的重心坐标(重心组合系数)$\alpha_i(i=0,1,\cdots,n)$是惟一的，$P=\alpha_0A_0+\alpha_1A_1+\cdots+\alpha_nA_n$。这与仿射坐标有密切的联系：在式(12.4)中令$j=0$，有

$$ P=\alpha_0A_0+\alpha_1A_1+\cdots+\alpha_nA_n=A_0+\alpha_1(A_1-A_0)+\cdots+\alpha_n(A_n-A_0) $$

上式表明，在由$A_i(i=0,1,\cdots,m)$组成的基下，点P的仿射坐标为$\alpha_1,\cdots,\alpha_m$。

当指定了仿射空间X的仿射基，X上的$m+1$个点构成p维($m\geqslant p$且$n\geqslant p$)仿射子空间的充要条件是由坐标向量$(x_{i1},\cdots,x_{in})^T(i=0,1,\cdots,m)$组成的下面这个$(n+1)\times(m+1)$矩阵

$$ \mathcal{D}=\begin{pmatrix}x_{01} & x_{11} & \cdots & x_{m1}\\ \cdots & \cdots & \cdots & \cdots\\ x_{0n} & x_{1n} & \cdots & x_{mn}\\ 1 & 1 & \cdots & 1\end{pmatrix} $$

的秩为$p+1$。实际上，若秩小于$p+1$，一定可以把上面矩阵中的任一列，表示成其他p列的重心组合，而秩大于$p+1$意味着至少$p+2$个点是独立的。

例 12.5 平面中的直线方程

假设仿射平面内三个点A_0,A_1和A_2在某个基下的坐标为$(x_0,y_0)^T$，$(x_1,y_1)^T$和$(x_2,y_2)^T$。按上一段的结论，这三个点在一个一维仿射空间(共线)上的充要条件是下面矩阵的秩为2，

$$ \mathcal{D}=\begin{pmatrix}x_0 & x_1 & x_2\\ y_0 & y_1 & y_2\\ 1 & 1 & 1\end{pmatrix} $$

或者等价于矩阵的行列式为0。注意下面的式子，

$$ \mathrm{Det}(\mathcal{D})=x_1y_2-x_2y_1+x_2y_0-x_0y_2+x_0y_1-x_1y_0=\begin{pmatrix}x_1-x_0\\ y_1-y_0\end{pmatrix}\times\begin{pmatrix}x_2-x_0\\ y_2-y_0\end{pmatrix} $$

其中，“$\times$”表示$\mathbb{R}^2$中两个向量的叉积。因此$\mathrm{Det}(\mathcal{D})=0$等价于$\overrightarrow{A_0A_1}$和$\overrightarrow{A_0A_2}$平行或三个点共线。若$A_0$和$A_1$固定，则$\mathrm{Det}(\mathcal{D})=0$可以看成是过$A_0$和$A_1$的直线方程，未知数是$A_2$的坐

标，形式为 $ax_2 + by_2 + c = 0$。这个方法可以推广到任意多个点定义的仿射空间：只要让矩阵D的相应余子式的行列式为 0 就可以得到对应的方程。

12.1.3　仿射变换和仿射投影模型

两个仿射空间 X 和 Y 之间的仿射变换是 X 到 Y 之间的双射，把 m 维的子空间映射为 m 维的子空间，把平行的子空间映射为平行子空间，并保留重心组合不变(或等价的仿射坐标不变；见图 12.1)。判断一个变换是否为仿射变换也可以用它所具有的(看起来较弱)把直线映射到直线且不改变平行线段之间带符号长度比例关系的性质决定。

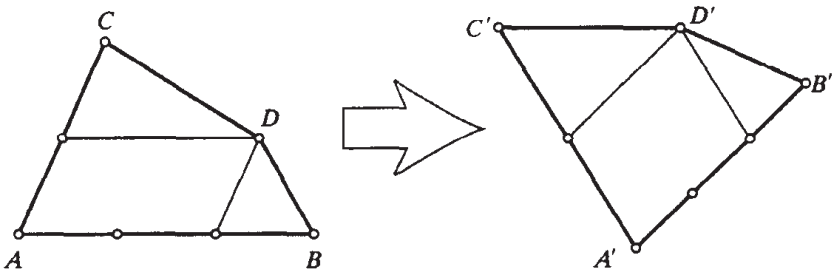


图 12.1　平面的一个仿射变换。点 A、B、C 和 D 变换到 A'、B'、C' 和 D'。D 在 ABC 组成的平面下的仿射坐标和 D' 在 $A'B'C'$ 下的坐标是一样的，都是(2/3，1/2)

两个 m 维仿射空间 X 和 Y 可以用 $m+1$ 个独立的点 $A_0, \cdots, A_m$ 和它们的像 $B_0, \cdots, B_m$ 完全决定。实际上，对于在以 A_i 为基的 X 中，仿射坐标为 $\alpha_i (i = 0, \cdots, m)$ 的任意其他点，它们在以 B_i 为基的 Y 中的像的坐标是相同的。反过来，给出 Y 中任意 m 个独立的点 $B_0, \cdots, B_m$，则存在惟一的仿射变换，将 X 中的 A_i 映射到 Y 中的 B_i。很明显，仿射变换是不保角度或距离的，图 12.1 也说明了这个结论。事实上，$\mathbb{R}^3$ 上的任意仿射变换都可以由平移、旋转、不均匀放缩和剪切这些变换组合得到。

向量和仿射空间之间的关系引发了线性与仿射变换之间的关系。具体说来，两个仿射空间 X 和 Y 之间的仿射变换$\psi: X \to Y$ 与向量空间 $\vec{X}$ 和 $\vec{Y}$ 之间的关系可以写为

$$\psi(P) = \psi(O) + \vec{\psi}(P - O)$$

其中，O 是任选的坐标原点，$\vec{\psi}: \vec{X} \to \vec{Y}$ 是 $\vec{X}$ 到 $\vec{Y}$ 上的一个线性映射，它是和点 O 的选取无关的。若 X 和 Y 的维数为 m(有限)，且以 O 为坐标原点的仿射坐标系已选定，则可以得到下面的熟悉形式

$$\psi(\boldsymbol{P}) = \boldsymbol{d} + \boldsymbol{C}\boldsymbol{P} = \boldsymbol{C}\boldsymbol{P} + \boldsymbol{d}$$

其中，$\boldsymbol{P}$ 为向量 P 在选定基下的坐标，$\boldsymbol{d}$ 表示$\psi(O)$的坐标向量，$\boldsymbol{C}$是在同一坐标系下表示$\vec{\psi}$的 $m \times m$ 矩阵。于是第 2 章定义的仿射变换和这里定义的是一致的。

平行投影的一个基本性质是它们导致了从平面到图像的一个仿射变换。首先我们看看它是不是保持共线点间的有向距离的：图 12.2(左)中三角形 OAa、OBb 和 OCc 是相似的，且对任意方向的 OC 和 Oc 都满足$\overline{AB}/\overline{BC} = \overline{ab}/\overline{bc}$。为了证明平行投影保持直线间的平行性，要用到这样的事实：一个平面与两个平行平面各自交出的交线是平行的(见习题)。现在看图 12.2 右图的情况，这里两个平行直线 Δ_1 和 Δ_2 被投影到一个平面上，由于每条直线和它们的投影定义的平面是相互平行的，因此它们与图像的交线 δ_1 和 δ_2 也是相互平行的。

从一个平面到另一个平面的弱透视投影和类透视投影也是仿射变换。由于它们是由一个平行投影和仿射变换构成的，其中仿射变换又包括与深度成反比的比例因子和摄像机内参数，故可以直接说明它们是仿射变换。第 2 章的定理 2 指出，一个仿射投影一定可以写成弱透视投影，因此，平面间的仿射投影也是仿射变换。

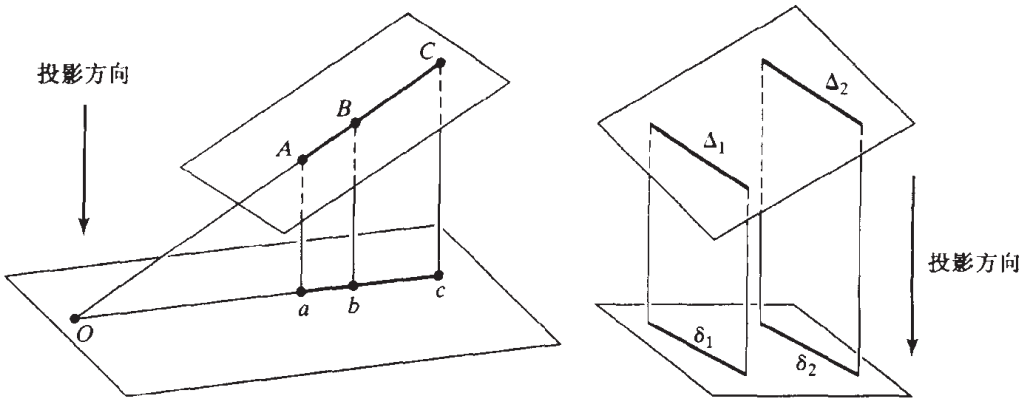


图 12.2 平行投影保持：(左)共线点之间的有向距离的比值和(右)直线间的平行性

故仿射投影保持平行和重心组合关系。尤其是它把场景中任意点集的重心映射到图像上对应点集的重心(因此可以用一个简单的方法确定类透视投影的参考点；参考第 2 章)，而且共线点之间的有向距离的比值是一个仿射投影不变量(这在物体识别中很有用；参考第 23 章)。

12.1.4 仿射形状

对仿射空间 X 上的两个点集 S 和 S'(可能是无限的)，若存在一个仿射变换 $\psi: X \rightarrow X$，使得 S' 是 S 在 ψ 作用下的像，则称两个点集是仿射等价的。可以看到仿射等价是一种等价关系，我们把 X 上 S 的仿射形状定义为所有仿射等价点集组成的等价类。从运动估计仿射结构可以看成一个从图像序列间的特征匹配恢复景物的仿射形状(或者由透视矩阵确定的等价类)的问题。现在我们已经有了解决这个问题的有效工具。

12.2 仿射结构和两幅图之间的运动

本节从已知一个场景的两幅仿射图像(以后的部分会继续分析多幅图像的情况)开始分析。本章介绍的两种从运动估计模型的方法是互补的：第一种方法用几何推导恢复场景的仿射形状(如果需要可以得到投影矩阵)，第二种方法是用代数方法估计投影矩阵(从投影矩阵可以容易地得到景物点的位置)。

12.2.1 几何场景重建

我们已经知道，已知 4 个点 A, B, C, D 的两个仿射视图后，就可以计算其他任意点在(A, B, C, D)这个基下的仿射坐标了。下面将介绍 Koenderink 和 Van Doorn(1990)提出的证明方法。一个平面到另外一个平面的仿射投影是一个仿射变换。若 P 正好在三角形 ABC 构成的平面 Π 上，在这个平面内用 ABC 作为基，得到 P 的仿射坐标，可以直接从两个图像中的任一个度量。设 E(对应的 Q)是平面 Π 与过 D 和 d'(对应 P 和 p')的直线的交点(见图 12.3)。E

和 Q 在平面 Π'' 上的投影点为 e'' 和 q''，它在基 (a'', b'', c'') 下的坐标与 d' 和 p' 在基 (a', b', c') 下的坐标是一样的。

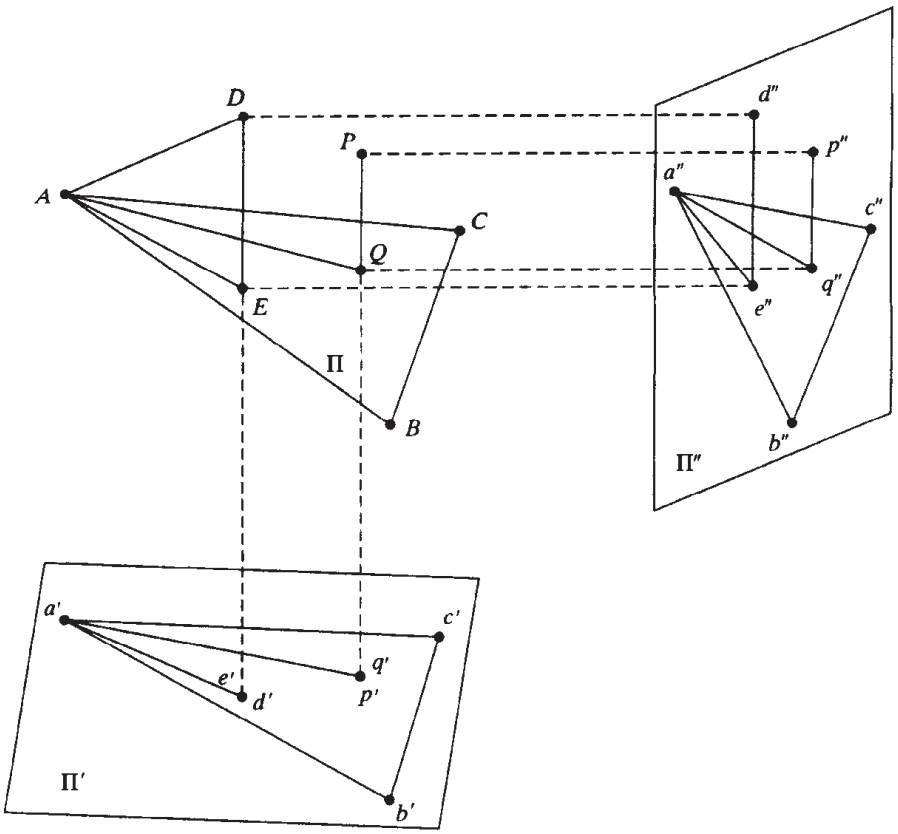


图 12.3　点 P 的仿射坐标在 A, B, C, D 四点构成的基上的几何约束。本图显示了平行投影的情况，但是推理过程适用于通用的仿射情况

另外，由于线段 ED 和 QP 平行于第一个投影方向，两条线段 $e''d''$ 和 $q''p''$ 也是相互平行的，它们之间的比例为

$$\lambda = \frac{\overline{q''p''}}{\overline{e''d''}} = \frac{\overline{QP}}{\overline{ED}}$$

其中，$\overline{AB}$ 表示在以任意朝向(但固定)连接 A 和 B 这两个点的直线上，AB 间的有向距离。

若用 $(\alpha_{d'}, \beta_{d'})$ 和 $(\alpha_{p'}, \beta_{p'})$ 分别表示 $d' = e'$ 和 $p' = q'$ 在基 (a', b', c') 下的坐标，则有

$$\begin{aligned}\overrightarrow{AP} &= \overrightarrow{AQ} + \overrightarrow{QP} = \alpha_{p'}\overrightarrow{AB} + \beta_{p'}\overrightarrow{AC} + \lambda\overrightarrow{ED} \\ &= (\alpha_{p'} - \lambda\alpha_{d'})\overrightarrow{AB} + (\beta_{p'} - \lambda\beta_{d'})\overrightarrow{AC} + \lambda\overrightarrow{AD}\end{aligned}$$

也就是说，P 在基 (A, B, C, D) 下的坐标为 $(\alpha_{p'} - \lambda\alpha_{d'}, \beta_{p'} - \lambda\beta_{d'}, \lambda)$。这就是“从运动估计放射模型”定理：给定不共面 4 点的两个仿射视图，可以惟一确定场景的仿射形状(Koenderink 和 Van Doorn，1990)。图 12.4 是在 Koenderink 和 Van Doorn 的实验中用来合成人脸的三个投影图。此外，图中还显示了从这两幅图像计算出的仿射模型。

12.2.2　代数运动估计

下面介绍另外一种完全不同的方法，它是完全基于代数运算的，而没有考虑几何上的实际

含义，这种代数方法利用了“从运动估计模型”中的歧义性简化投影矩阵的形式，从而可以推出一个非常简单的计算矩阵和对应仿射形状的方法。

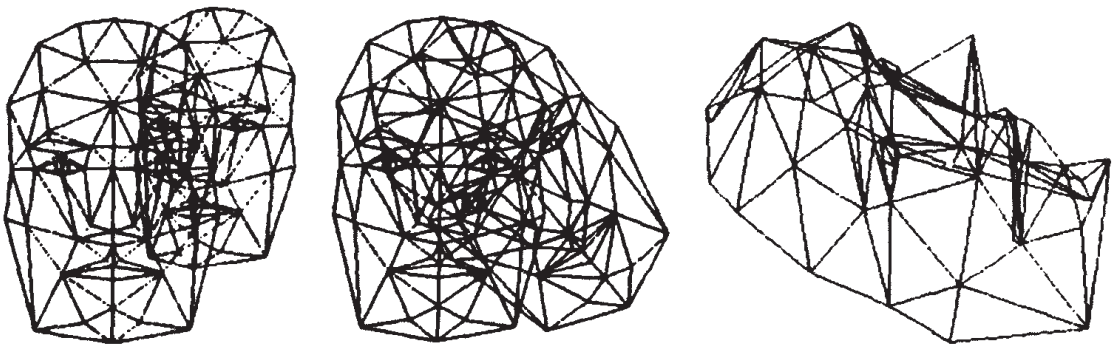


图 12.4 从两个视角仿射重构。左图和中图：人脸的三个视图；图 0 和图 1 在左图上重叠显示，图1和图2在中图上重叠显示。右图：从图像0和1重构得到的人脸仿射外形（在12.4节中的欧氏重构中会用到第三幅图像）

首先引入外极线约束的仿射等价性。考虑两个仿射图像，并把仿射方程

$$\begin{cases} \boldsymbol{p} = \mathcal{A}\boldsymbol{P} + \boldsymbol{b} \\ \boldsymbol{p}' = \mathcal{A}'\boldsymbol{P} + \boldsymbol{b}' \end{cases} \text{写成} \begin{pmatrix} \mathcal{A} & \boldsymbol{p} - \boldsymbol{b} \\ \mathcal{A}' & \boldsymbol{p}' - \boldsymbol{b}' \end{pmatrix} \begin{pmatrix} \boldsymbol{P} \\ -1 \end{pmatrix} = \boldsymbol{0}$$

的形式，这个方程有非零解的充要条件是

$$\mathrm{Det}\begin{pmatrix} \mathcal{A} & \boldsymbol{p} - \boldsymbol{b} \\ \mathcal{A}' & \boldsymbol{p}' - \boldsymbol{b}' \end{pmatrix} = 0$$

或

$$\alpha u + \beta v + \alpha' u' + \beta' v' + \delta = 0 \tag{12.5}$$

其中，$\alpha, \beta, \alpha', \beta'$ 是只与 $\mathcal{A}, \boldsymbol{b}, \mathcal{A}'$ 和 $\boldsymbol{b}'$ 有关的常量。这就是仿射外极约束。事实上，给出第一幅图像中的一点 p，它的对应点 p' 的位置是受方程 12.5 限制的，必然在直线 $\alpha' u' + \beta' v' + \gamma' = 0$ 上，其中，$\gamma' = \alpha u + \beta v + \delta$，反过来对于第二幅图像上的点也有同样的限制（见图 12.5）。

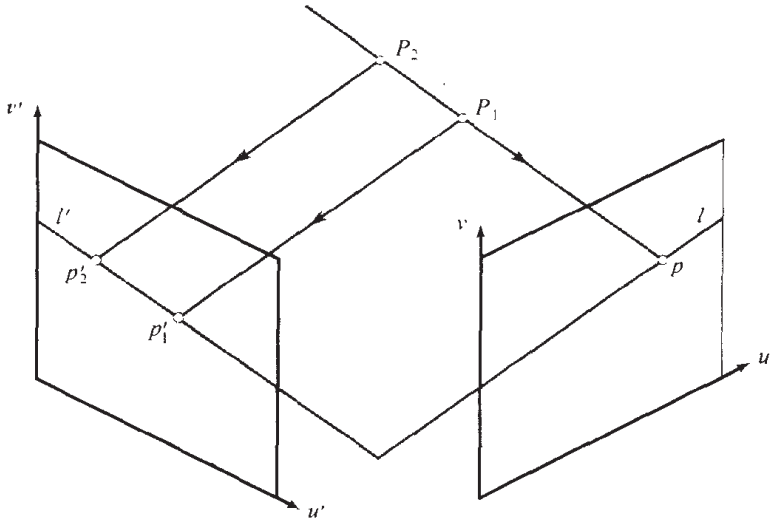


图 12.5 仿射外极几何：给定两幅平行投影图像、一幅图像上的一点 p 和两个投影方向确定了一个外极平面。外极平面和另一幅图像的交线为外极线 l'。在投影时，点 p 的对应点 p' 一定在这条直线上

注意一幅图像上的各条外极线是相互平行的：例如移动 p 会改变 γ'，也就是从原点到外极线 l' 的距离，但是不会改变 l' 的方向。

外极线约束可以写成我们很熟悉的形式

$$(u, v, 1)\mathcal{F}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0, \quad 其中 \quad \mathcal{F} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & \beta \\ \alpha' & \beta' & \delta \end{pmatrix}$$

是仿射基础矩阵。仿射外极几何可以看成透视外极几何的一种极限情况。若景物逐渐远离透视投影摄像机，拍到的图像序列确实就是仿射图像（有关细节请参考习题）。

投影矩阵可以从外极约束估计得到。“从运动估计仿射模型”中固有的仿射不确定性，能够简化计算：按照方程(12.2)和方程(12.3)，若$\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$和$\mathcal{M}' = (\mathcal{A}' \quad \boldsymbol{b}')$是问题的解，则$\tilde{\mathcal{M}} = \mathcal{M}\mathcal{Q}$和$\tilde{\mathcal{M}}' = \mathcal{M}'\mathcal{Q}$也是解，其中$\mathcal{Q} = \begin{pmatrix} \mathcal{C} & \boldsymbol{d} \\ \boldsymbol{0}^T & 1 \end{pmatrix}$是任意仿射变换。新的投影矩阵可以写成$\tilde{\mathcal{M}} = (\mathcal{A}\mathcal{C} \quad \mathcal{A}\,\boldsymbol{d} + \boldsymbol{b})$和$\tilde{\mathcal{M}}' = (\mathcal{A}'\mathcal{C} \quad \mathcal{A}'\boldsymbol{d} + \boldsymbol{b}')$。注意，按照方程(12.3)，这种变换等价于对所有景物点 P 做一个反变换，用 $\tilde{\boldsymbol{P}} = \mathcal{C}^{-1}(\boldsymbol{P} - \boldsymbol{d})$代替 $\boldsymbol{P}$。

把$\mathcal{A}(\mathcal{A}')$的两行记为 $\boldsymbol{a}_1^T$ 和 $\boldsymbol{a}_2^T$($\boldsymbol{a}'^T_1$ 和 $\boldsymbol{a}'^T_2$)，并引入向量 $\boldsymbol{b} = (b_1, b_2)^T$ 和$\boldsymbol{b}' = (b'_1, b'_2)^T$，则外极约束可以写成

$$0 = \mathrm{Det}\begin{pmatrix} \mathcal{A}\mathcal{C} & \boldsymbol{p} - \mathcal{A}\boldsymbol{d} - \boldsymbol{b} \\ \mathcal{A}'\mathcal{C} & \boldsymbol{p}' - \mathcal{A}'\boldsymbol{d} - \boldsymbol{b}' \end{pmatrix} = \mathrm{Det}\left(\begin{array}{c|c} \boldsymbol{a}_1^T\mathcal{C} & u - \boldsymbol{a}_1^T\boldsymbol{d} - b_1 \\ \boldsymbol{a}_2^T\mathcal{C} & v - \boldsymbol{a}_2^T\boldsymbol{d} - b_2 \\ \hline \boldsymbol{a}_1'^T\mathcal{C} & u' - \boldsymbol{a}_1'^T\boldsymbol{d} - b'_1 \\ \boldsymbol{a}_2'^T\mathcal{C} & v' - \boldsymbol{a}_2'^T\boldsymbol{d} - b'_2 \end{array}\right)$$

$$= \mathrm{Det}\left(\begin{array}{c|c} \boldsymbol{a}_1^T\mathcal{C} & u - \boldsymbol{a}_1^T\boldsymbol{d} - b_1 \\ \boldsymbol{a}_2^T\mathcal{C} & v - \boldsymbol{a}_2^T\boldsymbol{d} - b_2 \\ \boldsymbol{a}_1'^T\mathcal{C} & u' - \boldsymbol{a}_1'^T\boldsymbol{d} - b'_1 \\ \hline \boldsymbol{a}_2'^T\mathcal{C} & v' - \boldsymbol{a}_2'^T\boldsymbol{d} - b'_2 \end{array}\right) = \mathrm{Det}\begin{pmatrix} \mathcal{S}\mathcal{C} & \boldsymbol{q} - \mathcal{S}\boldsymbol{d} - \boldsymbol{r} \\ \boldsymbol{c}^T & v' - d \end{pmatrix}$$

其中

$$\mathcal{S} = \begin{pmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \boldsymbol{a}_1'^T \end{pmatrix}, \quad \boldsymbol{q} = \begin{pmatrix} u \\ v \\ u' \end{pmatrix}, \quad \boldsymbol{r} = \begin{pmatrix} b_1 \\ b_2 \\ b'_1 \end{pmatrix}, \quad \boldsymbol{c} = \mathcal{C}^T\boldsymbol{a}'_2 \quad , \qquad d = \boldsymbol{a}_2'^T\boldsymbol{d} + b'_2$$

若$\mathcal{S}$非奇异，可以取$\mathcal{C} = \mathcal{S}^{-1}$和 $\boldsymbol{d} = -\mathcal{S}^{-1}\boldsymbol{r}$。若 $\boldsymbol{c} = (a, b, c)^T$，可把两个投影矩阵写成规范形式

$$\tilde{\mathcal{M}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad , \quad \tilde{\mathcal{M}}' = \begin{pmatrix} 0 & 0 & 1 & 0 \\ a & b & c & d \end{pmatrix} \tag{12.6}$$

则外极约束又可以改写为

$$\mathrm{Det}\begin{pmatrix} 1 & 0 & 0 & u \\ 0 & 1 & 0 & v \\ 0 & 0 & 1 & u' \\ a & b & c & v' - d \end{pmatrix} = -au - bv - cu' + v' - d = 0$$

其中，系数 a, b, c, d 与参数$\alpha, \beta, \alpha', \beta'$满足$\delta$by$a : \alpha = b : \beta = c : \alpha' = -1 : \beta' = d : \delta$ 的关系。

已知足够的对应点之后，可以用线性最小二乘法估计系数 a,b,c,d，这与第 10 章对透视情况的分析类似。若这些参数已知，就知道了这两个仿射矩阵；那么从图像坐标估计场景的问题就是求坐标 $\tilde{\boldsymbol{P}}$ 的问题，可以通过再次运用最小二乘法解下面的方程组方程得到，

$$\begin{pmatrix} 1 & 0 & 0 & u \\ 0 & 1 & 0 & v \\ 0 & 0 & 1 & u' \\ a & b & c & v'-d \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{P}} \\ -1 \end{pmatrix} = 0 \tag{12.7}$$

注意方程组(12.7)中的前三个方程理论上足以求解 $\tilde{\boldsymbol{P}} = (u, v, u')^T$，而不需要估计系数 a,b,c,d，也不需要匹配点。这并不是非常奇怪的结论：若已知两个标定的摄像机，投影方向相互垂直，且 v 轴相互平行，则不需要欧氏重构就可以直接使用 $z = u', x = u, y = v$(可以参考图 12.5，设其为正交投影且外极线平行于 u 和 u' 轴)。在实用中，用四个方程可以得到更精确的结果。本节的方法通过仿射变换$\mathcal{Q}$把第一行$\mathcal{A}'$简化为(0,0,1)。若$\mathcal{S}$奇异(接近奇异)，可以用类似的方法对$\mathcal{A}'$的第二列进行处理。若$\mathcal{S}$和过程中构造的矩阵都奇异，说明两幅图像的成像平面是平行的，不可能恢复场景结构。

12.3 从多幅图像估计仿射结构和运动

上一节介绍的方法是从尽量少的图像恢复场景结构或/对应的投影矩阵。现在要分析从很多图像估计这些信息的方法。首先将说明任何的仿射图像几何都对应着同样的仿射结构；然后利用这条性质推导出 Tomasi 和 Kanade(1992)提出的分解方法，该方法可以很好地解决从多幅图像估计仿射结构和运动的问题。

12.3.1 仿射图像序列的仿射结构

在本节和下一节中，假设场景和 m 个仿射摄像机都是固定的，并把景物点 P 在各个成像平面上的像记为$p_1,\cdots,p_m$。把由方程(12.1)给出的 m 个对应关系写在一起，就得到

$$\boldsymbol{q} = \boldsymbol{r} + \mathcal{A}\boldsymbol{P}$$

其中，

$$\boldsymbol{q} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{p}_1 \\ \cdots \\ \boldsymbol{p}_m \end{pmatrix}, \quad \boldsymbol{r} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{b}_1 \\ \cdots \\ \boldsymbol{b}_m \end{pmatrix}, \qquad \mathcal{A} \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{A}_1 \\ \cdots \\ \mathcal{A}_m \end{pmatrix}$$

若 I 表示从 m 个摄像机拍到的像点坐标矩阵，

$$I = \{\boldsymbol{r} + \mathcal{A}\boldsymbol{P} | \boldsymbol{P} \in \mathbb{R}^3\} = \boldsymbol{r} + V_{\mathcal{A}}$$

其中，$V_{\mathcal{A}}$是 $2m \times 3$ 矩阵$\mathcal{A}$的取值空间(就是说，在$\mathbb{R}^{2m}$中由它的列向量张成的三维子空间。换句话说，I 是仿射空间$\mathbb{R}^{2m}$上的一个三维子空间)。另外，如果像以前那样描述 m 个摄像机拍到的 n 个点 $P_1,\cdots,P_n$，可以把观测数据写成$(2m+1)\times n$ 矩阵的形式：

$$\mathcal{D} = \begin{pmatrix} \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ 1 & \cdots & 1 \end{pmatrix}$$

由 12.1 节，这个矩阵的秩至多为 4。

12.3.2　用分解因子方法解“从运动估计仿射模型”问题

Tomasi 和 Kanade(1992)通过奇异值分解来估计景物的结构和相应的摄像机运动，是从仿射图像估计仿射结构的一种鲁棒分解方法(参见“插入部分”)。

技术：奇异值分解(SVD)

设$\mathcal{A}$是一个 $m \times n$ 矩阵，且满足 $m \geqslant n$，则$\mathcal{A}$总可以写成

$$\mathcal{A} = \mathcal{U}\mathcal{W}\mathcal{V}^T$$

其中，

- $\mathcal{U}$是 $m \times n$ 的列正交阵，(即，$\mathcal{U}^T\mathcal{U} = \mathrm{Id}_n$)
- $\mathcal{W}$是一个对角阵，对角元素 $w_i(i = 1, \cdots, n)$为$\mathcal{A}$的奇异值，满足 $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_n \geqslant 0$
- $\mathcal{V}$是一个 $n \times n$ 正交阵，$\mathcal{V}^T\mathcal{V} = \mathcal{V}\mathcal{V}^T = \mathrm{Id}_n$

这就是矩阵$\mathcal{A}$的奇异值分解，可以用 Wilkinson 和 Reich(1971)提出的一种分解方法来计算。下面的定理说明了矩阵的奇异值与它的平方阵的特征值与特征向量的关系。

定理 3　$\mathcal{A}$的奇异值，就是$\mathcal{A}^T\mathcal{A}$的特征值，且$\mathcal{V}$的各列就是对应的特征向量。

这个定理可以用来解$\mathcal{A}\boldsymbol{x} = \mathbf{0}$一类的超定齐次线性方程，而不需要具体计算出对应的方阵$\mathcal{A}^T\mathcal{A}$。$\mathcal{A}$进行 SVD 后，矩阵$\mathcal{V}$中与最小的奇异值对应的列向量就是方程的解。

SVD 还能用来检测矩阵是不是满秩的。设$\mathcal{A}$的秩 $p < n$，则矩阵$\mathcal{U}$、$\mathcal{W}$和$\mathcal{V}$可以展开为

$$\mathcal{U} = \begin{bmatrix} \mathcal{U}_p & \mathcal{U}_{n-p} \end{bmatrix} \quad \mathcal{W} = \begin{bmatrix} \mathcal{W}_p & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathcal{V}^T = \begin{bmatrix} \mathcal{V}_p^T \\ \mathcal{V}_{n-p}^T \end{bmatrix}$$

且

- 在由$\mathcal{A}$各列张成的空间上，$\mathcal{U}_p$ 的各列是一组基，
- 在由 $\mathcal{A}\boldsymbol{x} = 0$ 的解向量所展成的空间(解空间)上，$\mathcal{V}_{n-p}$是一组基。

$m \times p$ 和 $n \times p$ 维矩阵$\mathcal{U}_p$ 和$\mathcal{V}_p$ 都是列正交的，则必然有$\mathcal{A} = \mathcal{U}_p\,\mathcal{W}_p\,\mathcal{V}_p^T$。

下面的定理说明，SVD 还提供了一种有效的近似方法。在两种情况下，$\mathcal{U}_p$ 和$\mathcal{V}_p$ 仍旧是$\mathcal{U}$和$\mathcal{V}$最左边的 p 列，$\mathcal{W}_p$ 是由 p 个最大的特征值组成的对角阵。但是此时$\mathcal{A}$的最大秩为 n，剩下的奇异值也可以是非零的。

定理 4　若$\mathcal{A}$的秩大于 p，则$\mathcal{U}_p\,\mathcal{W}_p\,\mathcal{V}_p^T$ 是在 Frobenius 距离意义下，对$\mathcal{A}$的秩为 p 的最优估计。

这个定理是本章提出的用分解方法解“运动估计仿射模型”方法的基础。

设景物坐标系的原点为某个景物点或是所有景物点的重心，记为 P_0，则可以通过坐标变换把图像坐标系的原点移动到图像的对应位置上。变换 $\boldsymbol{p} \rightarrow \boldsymbol{p} - \boldsymbol{p}_0$ 把像点矩阵 I 的原点统一起来，则 I 就成为一个三维向量空间。就是说，对任意景物点 P 和 $i = 1, \cdots, m$，有 $\boldsymbol{p}_i = \mathcal{A}_i\boldsymbol{P}$。或等价形式 $\boldsymbol{q} = \mathcal{A}\,\boldsymbol{P}$，

$$I = \{\mathcal{A}\boldsymbol{P} | \boldsymbol{P} \in \mathbb{R}^3\} = V_{\mathcal{A}}$$

给定 n 个点 $P_1, \cdots, P_n$ 的 m 幅图像，则可以写成一个 $2m \times n$ 矩阵

$$\mathcal{D} \stackrel{\text{def}}{=} (\boldsymbol{q}_1 \quad \cdots \quad \boldsymbol{q}_n) = \mathcal{A}\mathcal{P}, \qquad \mathcal{P} \stackrel{\text{def}}{=} (\boldsymbol{P}_1 \quad \cdots \quad \boldsymbol{P}_n)$$

$\mathcal{D}$是一个 $2m \times 3$ 矩阵和一个 $3 \times n$ 矩阵的乘积，则它的秩一般为 3。若$\mathcal{U}\mathcal{W}\mathcal{V}^T$ 是它的奇异值分解，则只有三个奇异值不是零，而$\mathcal{D} = \mathcal{U}_3 \mathcal{W}_3 \mathcal{V}_3^T$，其中$\mathcal{U}_3$ 和$\mathcal{V}_3$ 分别是 $2m \times 3$ 和 $3 \times n$ 的矩阵，由$\mathcal{U}$和$\mathcal{V}$的左边三列组成，而$\mathcal{W}_3$ 是由对应的三个非零奇异值组成的 3×3 对角阵。

至此可以确定$\mathcal{A}_0 = \mathcal{U}_3$ 和$\mathcal{P}_0 = \mathcal{W}_3 \mathcal{V}_3^T$ 分别表示实际(仿射)摄像机位移和场景形状。由定义保证，$\mathcal{A}$的各列是$\mathcal{D}$张成的空间 $V_{\mathcal{A}}$ 上的一组基，而$\mathcal{A}_0$ 是这个向量空间上的另一组基。这说明存在一个 3×3 矩阵$\mathcal{Q}$使得$\mathcal{A} = \mathcal{A}_0 \mathcal{Q}$，并有$\mathcal{P} = \mathcal{Q}^{-1}\mathcal{P}_0$。则对任意的可逆 3×3 方阵$\mathcal{Q}$有 $D = (A_0 \mathcal{Q})(\mathcal{Q}^{-1}\mathcal{P}_0)$。这种线性不确定性加上世界坐标原点的选取这个自由度，进一步说明了“从运动估计结构”问题中的仿射不确定性。而 SVD 方法可以很好地估计出仿射运动和场景结构。

前面的推导只是在理想的无噪声情况下是可行的。在实际中，由于图像噪声问题，特征点的位置是有误差的，以及实际摄像机并不是仿射摄像机，方程$\mathcal{D} = \mathcal{A}\mathcal{P}$并不完全成立，矩阵$\mathcal{D}$可能是满秩的。但是在这种情况下，SVD 仍然能有效地估计仿射结构和运动：做法是调整矩阵$\mathcal{A}_i (i = 1, \cdots, m)$和 $\boldsymbol{P}_j (j = 1, \cdots, m)$，即$\mathcal{A}$和$\mathcal{P}$，使下式最小化

$$E \stackrel{\text{def}}{=} \sum_{i,j} |\boldsymbol{p}_{ij} - \mathcal{A}_i \boldsymbol{P}_j|^2 = \sum_j |\boldsymbol{q}_j - \mathcal{A}\boldsymbol{P}_j|^2 = |\mathcal{D} - \mathcal{A}\mathcal{P}|^2$$

算法 12.1 Tomasi-Kanade 的分解方法解“从运动估计结构”问题。注意在 Tomasi 和 Kanade(1992)提出的原始方法中，$\mathcal{A}_0 = \mathcal{U}_3 \sqrt{\mathcal{W}_3}$和$\mathcal{P}_0 = \sqrt{\mathcal{W}_3}\mathcal{V}_3^T$。这两种解在理论上和数值上都是等价的。

1. 计算奇异值分解$\mathcal{D} = \mathcal{U}\mathcal{W}\mathcal{V}^T$；
2. 分别取$\mathcal{U}$和$\mathcal{V}$的左边三列得到$\mathcal{U}_3, \mathcal{V}_3, \mathcal{W}_3$，对应的奇异值组成$\mathcal{W}$的 3×3 子阵$\mathcal{W}_3$。
3. 定义

$$\mathcal{A}_0 = \mathcal{U}_3 \quad , \qquad \mathcal{P}_0 = \mathcal{W}_3 \mathcal{V}_3^T$$

$2m \times 3$ 矩阵$\mathcal{A}_0$ 是摄像机位移的估计，$3 \times n$ 矩阵$\mathcal{P}_0$ 是场景结构的估计。

按照定理 4，矩阵$\mathcal{A}_0\mathcal{P}_0$ 是在 3 阶矩阵中对$\mathcal{D}$的最优估计。因为$\mathcal{A}\mathcal{P}$是 3 阶 $2m \times 3$ 矩阵$\mathcal{A}$和 3 阶 $3 \times n$ 矩阵$\mathcal{P}$的乘积，它的秩为 3，则 E 的最小只在$\mathcal{A} = \mathcal{A}_0$ 和$\mathcal{P} = \mathcal{P}_0$ 时达到，即$\mathcal{A}_0$ 和$\mathcal{P}_0$ 是摄像机位移和场景结构的最佳估计。这与“从位移估计仿射结构”问题的固有不确定性并不矛盾：所有仿射等价的解都对应同样的 E。可以按照算法 12.1 的方法用 SVD 分解来估计仿射结构$\mathcal{D}$。

12.4 从仿射到欧氏图像

若用两个正交投影摄像机观测一个固定场景，则归一化的坐标向量可以直接拿来作为像点的坐标。在这种情况下，两个摄像机之间的坐标变换就是一个欧氏变换(可以写成旋转和平移的组合)。在正交投影条件下，深度上的平移不会产生影响，在成像平面内的平移(面内平

移)只要对齐一些像点 A 就可以很容易地去除,而绕光轴的旋转也可以很容易地去掉。若过点 A 做一个正对视点的平面的平移,则此时两个视图就相差一个绕面内某个轴的旋转。Koenderink和Van Doorn(1990)指出这种旋转构成一个只有一个参数的旋转族,除了深度的比例变换和剪切还不能确定之外,可以基本确定景物形状。若加入第三个摄像机,则可以把解的个数限定到一对或两对,在正对视点的平面内具有反射对称关系的情况(见图 12.4 和图 12-6)。具体过程比较复杂,就不在这里介绍了。我们将介绍的方法是,已知摄像机的仿射投影矩阵后,如何得到场景的欧氏几何结构。

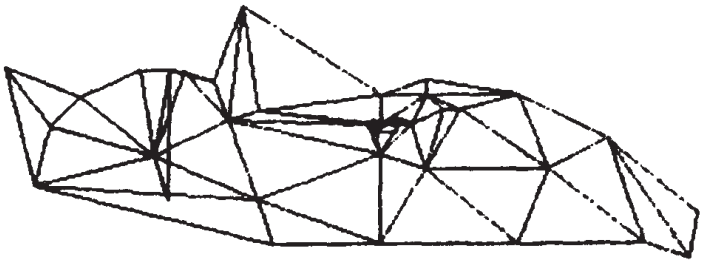


图 12.6　图 12.4 中的人脸通过三个摄像机进行欧氏重构后的结果

12.4.1　欧氏约束和标定的仿射摄像机

我们可以从另外一种角度看待图像处理中的正交投影,弱透视投影和类透视投影模型(由于实际应用中很少用到平行投影,这里就不详细介绍了),且只考虑摄像机已经标定过的情况。显然,仿射投影方程依然成立但是投影矩阵$\mathcal{M}=(\mathcal{A}\quad \boldsymbol{b})$要受一些限制。

回忆一下第 2 章的式(2.20),一个弱透视投影矩阵可以写成

$$\mathcal{M}=\frac{1}{z_r}\begin{pmatrix}k & s\\ 0 & 1\end{pmatrix}(\mathcal{R}_2\quad \boldsymbol{t}_2)$$

其中,$\mathcal{R}_2$ 是一个由旋转矩阵前两行组成的 2×3 矩阵,$\boldsymbol{t}_2$ 是$\mathbb{R}^2$ 上的一个向量。摄像机标定后,可以直接使用归一化的图像坐标。取 $k=1$ 和 $s=0$,则投影矩阵变为

$$\hat{\mathcal{M}}=(\hat{\mathcal{A}}\quad \hat{\boldsymbol{b}})=\frac{1}{z_r}(\mathcal{R}_2\quad \boldsymbol{t}_2) \tag{12.8}$$

正交投影相机是弱透视投影摄像机取 $z_r=1$ 时的特例,也满足式(12.8),$\hat{\mathcal{A}}$是旋转矩阵的一部分,两个单位行向量 $\hat{\boldsymbol{a}}_1^T$ 和 $\hat{\boldsymbol{a}}_2^T$ 是相互垂直的。换句话说,正交投影摄像机是附加两个约束的仿射摄像机,

$$\hat{\boldsymbol{a}}_1\cdot\hat{\boldsymbol{a}}_2=0\quad 和\quad |\hat{\boldsymbol{a}}_1|^2=|\hat{\boldsymbol{a}}_2|^2=1 \tag{12.9}$$

弱透视投影的情况类似,但是$\hat{\mathcal{A}}$的各行不再是单位向量。弱透视投影的限制是

$$\hat{\boldsymbol{a}}_1\cdot\hat{\boldsymbol{a}}_2=0\quad 和\quad |\hat{\boldsymbol{a}}_1|^2=|\hat{\boldsymbol{a}}_2|^2 \tag{12.10}$$

最后,由第 2 章中式(2.22)对类透视投影的参数化表示,可得类透视投影是添加下面约束的仿射摄像机

$$\hat{\boldsymbol{a}}_1\cdot\hat{\boldsymbol{a}}_2=\frac{u_r v_r}{2(1+u_r^2)}|\hat{\boldsymbol{a}}_1|^2+\frac{u_r v_r}{2(1+v_r^2)}|\hat{\boldsymbol{a}}_2|^2\quad 和\quad \frac{|\hat{\boldsymbol{a}}_1|^2}{(1+u_r^2)}=\frac{|\hat{\boldsymbol{a}}_2|^2}{(1+v_r^2)} \tag{12.11}$$

其中,(u_r,v_r)是参考点 R 在类透视投影下的投影坐标。

12.4.2 从多个视角计算欧几里得升级

这一节只考虑正交投影情况，且设场景的仿射模型和每个视图的投影矩阵都已经知道了。我们已经知道这个“运动估计模型”问题的解，在仿射意义下是一样的。若景物点在一个欧氏坐标系内的坐标为 $\hat{\boldsymbol{P}}$，且对应的投影矩阵为 $\hat{\mathcal{M}} = (\hat{\mathcal{A}} \quad \hat{\boldsymbol{b}})$，则一定存在仿射变换

$$\mathcal{Q} = \begin{pmatrix} \mathcal{C} & \boldsymbol{d} \\ \boldsymbol{0}^T & 1 \end{pmatrix}$$

使得 $\hat{\mathcal{M}} = \mathcal{M}\mathcal{Q}$ 和 $\hat{\boldsymbol{P}} = \mathcal{C}^{-1}(\boldsymbol{P} - \boldsymbol{d})$。这个变换称为欧几里得升级，因为它把一个场景的仿射形状映射成欧氏形状。

下面介绍在 $m \geqslant 3$ 的情况下如何计算这种升级。设 $\mathcal{M}_i = (\mathcal{A}_i \quad \boldsymbol{b}_i)$ 表示对应的投影矩阵，估计过程中用到了 12.3.2 节中的分解方法。若 $\hat{\mathcal{M}}_i = \mathcal{M}_i\mathcal{Q}$，则式(12.9)的正交投影约束可以写成

$$\begin{cases} \hat{\boldsymbol{a}}_{i1} \cdot \hat{\boldsymbol{a}}_{i2} = 0, \\ |\hat{\boldsymbol{a}}_{i1}|^2 = 1, \\ |\hat{\boldsymbol{a}}_{i2}|^2 = 1, \end{cases} \iff \begin{cases} \boldsymbol{a}_{i1}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i2} = 0, \\ \boldsymbol{a}_{i1}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i1} = 1, \\ \boldsymbol{a}_{i2}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i2} = 1, \end{cases} \quad \text{for} \quad i = 1, \cdots, m \tag{12.12}$$

其中，$\boldsymbol{a}_{i1}^T$ 和 $\boldsymbol{a}_{i2}^T$ 代表 $\mathcal{A}_i$ 的各列。这是一个关于 $\mathcal{C}$ 各系数的 $3m$ 个二次方程的超定系统，可以用非线性最小二乘法求解。另一个方法是把式(12.12)看做是对矩阵 $\mathcal{D} \stackrel{\text{def}}{=} \mathcal{C}\mathcal{C}^T$ 的一组线性约束。$\mathcal{D}$ 的系数可以通过线性最小二乘法得到，则用 Cholesky 分解求 $\sqrt{\mathcal{D}}$ 就能得到 $\mathcal{C}$ 了。要注意，这意味着恢复的矩阵 $\mathcal{D}$ 必须是正定的，但是这在有噪声情况下不一定成立。而且方程(12.12)的解并没有求出旋转。为了惟一确定 $\mathcal{Q}$ 同时简化计算，可以把 $\mathcal{M}_1$（也可以是 $\mathcal{M}_2$）规范化，然后再用上一节提出的方法求解。

图 12.7 是一个例子，包括从不同视角观察房子得到的 4 幅图片、一个恢复出来的场景结构以及从相近视角观察到的实际对比图片。

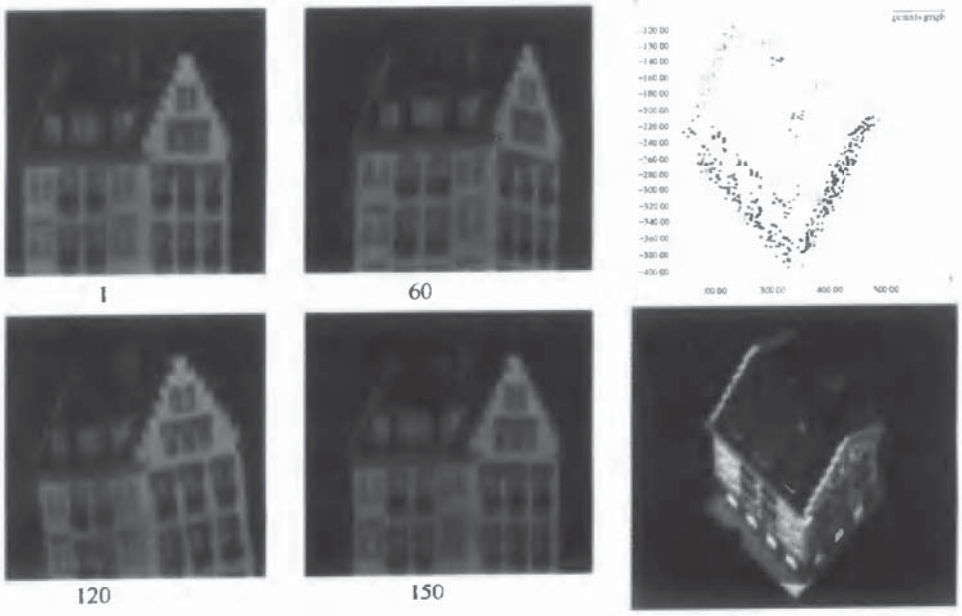


图 12.7 由运动实验得出的欧几里得结构，左：从一所房子的 150 帧序列中挑出的采样图像，右：重构结构的一个视图（上）与该房子从相似视角拍的实际图像（下）

对弱透视投影和类透视投影的欧几里得升级方法基本相同，只是新的约束式(12.10)和式(12.11)代替了原来的 $3m$ 条约束式(12.12)。注意在这种情况下不可能确定场景的实际大小，因为式(12.10)和式(12.11)是齐次坐标的约束。场景只能在相似的意义下恢复(放缩和刚体变换)。因此这里的欧氏形状是指，在相似意义下一致的点集合所构成等价类。

12.5　仿射运动分割

到目前为止都假设所有观察到的点都做一致运动。如果场景中的 k 个物体分别运动，结果会怎么样？这一节将介绍两种如何把数据点分成独立的运动物体的方法。

12.5.1　数据的梯形矩阵归约

与 12.3.1 节介绍的一样，数据矩阵定义为

$$\mathcal{D}=\begin{pmatrix} \boldsymbol{p}_{11} & \cdots & \boldsymbol{p}_{1n} \\ \cdots & \cdots & \cdots \\ \boldsymbol{p}_{m1} & \cdots & \boldsymbol{p}_{mn} \\ 1 & \cdots & 1 \end{pmatrix}$$

但是，这里的$\mathcal{D}$不再是 4 阶的，而是每个物体对应的那些列构成一个 4 阶的子阵 $D_i(i=1,\cdots,k)$，则$\mathcal{D}$的秩至多是 $4k$。Gear(1998)提出，通过计算$\mathcal{D}$规约到梯形阵(RREF)可以确定子空间 D_i 和对应的列向量，可以通过这种方法把数据点划分成若干刚体(或者，更确切地说，仿射变换下的若干物体)。

$\mathcal{U}$的 RREF 是由它的各行线性组合构成的新矩阵$\mathcal{V}$，且满足下面的性质：

1. 全是 0 的列都集中在最下面；
2. 每行的第一个非零元素都是 1，称为起始 1；
3. 下一行的起始 1 必须在上一行起始 1 的右边；且
4. 在起始 1 所在的列上，只有它自己是非零的。

基列是包含起始 1 的列。非基列上的非零元素只能在有起始 1 的行上出现，且非基列 $\boldsymbol{v}$ 都在由基列张成的子空间上，设非基列的非零元素为 $\alpha_1,\cdots,\alpha_k$，则 $\boldsymbol{v}$ 可以写成 $\alpha_1\boldsymbol{v}_{j1}+\cdots+\alpha_k\boldsymbol{v}_{jk}$。基列的个数决定矩阵的秩。

下面的例子显示了这些特性，$\mathcal{U}$是一个 7×6 矩阵，它的 RREF 是$\mathcal{V}$(为了使$\mathcal{U}$形式比较简单，我们特别选取了一个$\mathcal{V}$)：

$$\mathcal{U}=\begin{pmatrix} 1 & 0 & 1 & -5 & 2 & -9 \\ 2 & 4 & 10 & 0 & 1 & 1 \\ -1 & 1 & 1 & 3 & 0 & 1 \\ 0 & 1 & 2 & -1 & 3 & -10 \\ 3 & -2 & -1 & 0 & 1 & 3 \\ 0 & 5 & 10 & 2 & -2 & 8 \\ -2 & 3 & 4 & 1 & 0 & -3 \end{pmatrix} \longrightarrow \mathcal{V}=\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

我们用 $\boldsymbol{u}_i$ 和 $\boldsymbol{v}_i(i=1,\cdots,6)$表示$\mathcal{U}$和$\mathcal{V}$的各列。总共有 4 个基列 $\boldsymbol{v}_1$，$\boldsymbol{v}_2$，$\boldsymbol{v}_4$，$\boldsymbol{v}_5$，则$\mathcal{U}$的秩是 4。$\boldsymbol{v}_3$ 中的非零元素都在 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 中起始 1 所在的列上，$\boldsymbol{v}_3$ 在 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 张成的$\mathbb{R}^7$ 的子空间

上。$\boldsymbol{v}_3$ 的元素为 1 和 2，故 $\boldsymbol{v}_3 = \boldsymbol{v}_1 + 2\boldsymbol{v}_2$。同样地，$\boldsymbol{v}_6$ 上的非零元素在 $\boldsymbol{v}_1$，$\boldsymbol{v}_4$ 和 $\boldsymbol{v}_5$ 中起始 1 所在的列上，$\boldsymbol{v}_6$ 在 $\boldsymbol{v}_1$，$\boldsymbol{v}_4$ 和 $\boldsymbol{v}_5$ 张成的$\mathbb{R}^7$ 的子空间上。$\boldsymbol{v}_6$ 的元素为 2，1 和 -3，故 $\boldsymbol{v}_6 = 2\boldsymbol{v}_1 + \boldsymbol{v}_4 - 3\boldsymbol{v}_5$。实际上，这些等式对原矩阵$\mathcal{U}$也成立（$\boldsymbol{u}_3 = \boldsymbol{u}_1 + 2\boldsymbol{u}_2$，$\boldsymbol{u}_6 = 2\boldsymbol{u}_1 + \boldsymbol{u}_4 - 3\boldsymbol{u}_5$，计算一下就能验证）。这是因为$\mathcal{V}$的各行是$\mathcal{U}$各行的线性组合。

这些性质对任意矩阵和对应的RREF成立，这在理论上说明了数据矩阵$\mathcal{D}$的RREF可以用来做仿射运动分割。实际上，它确定了$\mathcal{D}$张成的空间上的一组基（数据矩阵的这些列对应于RREF的基列），同时也确定了由这些基的子集构成的子空间上的所有列。若这些四维子空间 D_i 之间只在原点相交（对足够大的 m 一般是对的），则可以从RREF画出对应的点之间的连通关系图。它的结点表示RREF的列，若两个结点之间存在连线，则表示至少有一行这两列的元素都不是零。

但是在实际的有噪声和数值计算误差情况下，情况变得非常复杂。一种简单的实现方法是用主元素的 Gauss-Jordan 消去法，通常得到满秩的矩阵，使得非基列都不在张成的四维空间里（见习题）。Gear（1998）给出了多种计算 RREF 的“鲁棒化”方法，包括加检验的 Gauss-Jordan 方法来去掉小的主元素，和在 QR 分解后再对三角阵$\mathcal{R}$使用 Gauss-Jordan 的方法。这些方法在合成图像和真实图像的实验中都得到了很好的分割效果。

12.5.2 形状交互矩阵

上一节介绍的方法只依赖于仿射图像的仿射结构。Costeira 和 Kanade（1998）提出了另外一种完全不同的方法，即基于数据矩阵分解的方法。这一节以两个不用运动的点集为例介绍这个方法，然后就能很容易地推广到任意多个独立运动物体。

在运动分割问题中，不可能对每个物体定义一个秩为 3 的矩阵，因为对应点集的中心是不知道的。然而，若假设数据是无噪声的，则数据矩阵$\mathcal{D}^{(i)}$（$i = 1,2$）可以写成

$$\mathcal{D}^{(i)} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{p}_{11}^{(i)} & \cdots & \boldsymbol{p}_{1n_i}^{(i)} \\ \cdots & \cdots & \cdots \\ \boldsymbol{p}_{m1}^{(i)} & \cdots & \boldsymbol{p}_{mn_i}^{(i)} \end{pmatrix}$$

其中，n_i 是物体 i 上点的个数，$n_1 + n_2 = n$。每个数据矩阵都可以写成$\mathcal{D}^{(i)} = \mathcal{M}^{(i)}\mathcal{P}^{(i)}$，故秩都是 4，其中

$$\mathcal{M}^{(i)} \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{M}_1^{(i)} \\ \cdots \\ \mathcal{M}_m^{(i)} \end{pmatrix}, \qquad \mathcal{P}^{(i)} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{P}_1^{(i)} & \cdots & \boldsymbol{P}_{n_i}^{(i)} \\ 1 & \cdots & 1 \end{pmatrix}$$

我们先定义组合起来的 $2m \times n$ 数据矩阵$\mathcal{D} \stackrel{\text{def}}{=} (\mathcal{D}^{(1)} \quad \mathcal{D}^{(2)})$和对应的 $2m \times 8$（运动）矩阵以及 $8 \times n$（结构）矩阵：

$$\mathcal{M} \stackrel{\text{def}}{=} (\mathcal{M}^{(1)} \quad \mathcal{M}^{(2)}), \qquad \mathcal{P} \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{P}^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \mathcal{P}^{(2)} \end{pmatrix}$$

有了这些定义，就得到$\mathcal{D} = \mathcal{M}\mathcal{P}$，它说明$\mathcal{D}$的秩最大为 8，则$\mathcal{P}$的各行是$\mathbb{R}^{2m}$上，由$\mathcal{D}$矩阵的各行所展成的八维子空间的一组基。Strang（1980）中提到，对于把正交投影的任意向量映射到由$\mathcal{A}$矩阵的各列张成的子空间上的运算，都可以用矩阵表示成$\mathcal{Z} \stackrel{\text{def}}{=} \mathcal{A}(\mathcal{A}^T\mathcal{A})^{-1}\mathcal{A}^T$。特别地，由于$\mathcal{P}^T$ 是一个块对角阵，则$\mathcal{D}$的各行（即$\mathcal{D}^T$ 的各列）对应的矩阵$\mathcal{Z}$也是块对角的。

显然，这里$\mathcal{P}$是未知的，但是如果其他矩阵的各行构成了$\mathcal{D}$行空间的基，它就可以当做$\mathcal{P}$。例如，若秩为 8 的矩阵$\mathcal{D}$的 SVD 分解为$\mathcal{U}_8 \mathcal{W}_8 \mathcal{V}_8^T$，则可以用$\mathcal{V}_8^T$ 的行作为基，由于$\mathcal{V}_8$ 是正交的，故有$\mathcal{Z} = \mathcal{V}_8(\mathcal{V}_8^T \mathcal{V}_8)^{-1}\mathcal{V}_8^T = \mathcal{V}_8 \mathcal{V}_8^T$。这样得到的矩阵$\mathcal{Z}$被 Costeira 和 Kanade(1998)称为形状交互矩阵，它也是块对角的。

上面的过程中认为数据点是按照所属物体的顺序排好的，一般情况则未必如此。但是可以看出矩阵$\mathcal{Z}$的值是与点的顺序无关的。改变点的顺序只是交换了$\mathcal{D}$的各列，也就交换了$\mathcal{Z}$的行和列。因此恢复正确的点的顺序(即把点分割成物体)就是通过行列变换把矩阵$\mathcal{Z}$变为块对角阵的过程。

Costeira 和 Kanade 提出了多种在噪声环境下寻找这种变换的方法，其中一种是使不在对角块上的元素的平方和最小化(参考 Costeira 和 Kanade，1998)。图 12.8 是一个实验结果，包括两个物体组成的场景的图像和对应的特征点跟踪，排序前和排序后的形状交互矩阵，以及对应的分割结果。

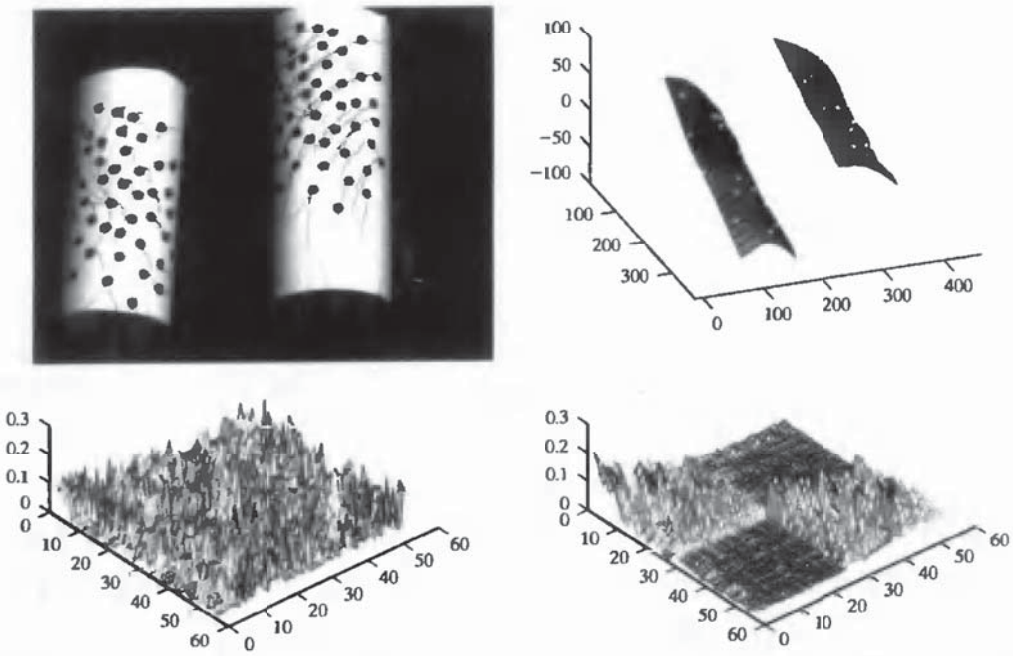


图 12.8　运动分割——实验结果。左上：由两个圆柱组成的场景的一幅图像，包括特征跟踪结果；右上：运动分割后的重构结果。左下：形状交互矩阵。右下：排序后的形状交互矩阵

12.6　注释

“运动估计结构”问题是 Ullman(1979)在标定正交投影装置时提出的。Koenderink 和 Van Doorn(1990)首先提出在仿射假定下的解。12.3.2 节介绍的分解方法是 Tomasi 和 Kanade(1992)提出的。在本章中提到，把“运动估计结构”问题分解为仿射和欧氏阶段，可以用简单且鲁棒的方法从图像序列重构场景形状。仿射阶段也很重要，因为它是 Gear(1998)和 Costeira 及 Kanade(1998)提出的基于运动分割方法的基础，见 12.5 节；Boult 和 Brown(1991)提出了另外一种方法。在第 26 章中给出了一些其他的应用，如在增强现实下的交互图像合成。各种仿射图像的仿射结构，以及等价仿射运动的 4 阶数据矩阵都说明仿射图像是三个模型图像的线性组合

(Ullman 和 Basri,1991),且景物点的投影图像是三个参考点投影的线性组合(Weinshall 和 Tomasi,1995)。Tomasi 和 Kanade(1992)提出用非线性最小二乘法解欧几里得升级矩阵$\mathcal{Q}$。对同一个问题的 Cholesky 方法是 Poelman 和 Kanade(1997)提出的;Weinshall 和 Tomasi(1995)提出了另外一种变形。这一章中提出的方法近些年又有了很多扩展,包括运动到模型的递增恢复法(Weinshall 和 Tomasi,1995;Morita 和 Kanade,1997),将仿射/欧氏分解扩展到透视/仿射/欧氏分层化(Faugeras,1995),以及对应的透视模型估计理论(Faugeras,1992;Hartley 等,1992;参考下一章),以及 Tomasi 和 Kanade(1992)提出的对透视情形(Sturm 和 Triggs,1996)的通用分解方法,和很多其他具有双线性结构的计算机视觉问题(Koenderink 和 Van Doorn,1997)。

习题

12.1 坐标系中,两个点的"加"运算和把一个点"乘"一个缩放因子必然是不独立于坐标系的,说明原因。

12.2 证明:如下定义的重心组合是与 j 的选取无关的

$$\sum_{i=0}^{m}\alpha_i A_i \stackrel{\text{def}}{=} A_j + \sum_{i=0,i\neq j}^{m}\alpha_i(A_i - A_j)$$

12.3 证明下面的等价关系

$$^{B}P = {}^{B}_{A}C\,{}^{A}P + {}^{B}O_A \quad \Longleftrightarrow \quad \begin{pmatrix} {}^{B}P \\ 1 \end{pmatrix} = \begin{pmatrix} {}^{B}_{A}C & {}^{B}O_A \\ \mathbf{0}^T & 1 \end{pmatrix}\begin{pmatrix} {}^{A}P \\ 1 \end{pmatrix}$$

12.4 说明 $m+1$ 个点 $A_0,\cdots,A_m$ 的重心组合构成了 X 的一个仿射子空间,且维数最高为 m。

12.5 推出三维空间$\mathbb{R}^3$ 中过两点的直线方程(提示:需要两个方程)。

12.6 说明一个平面和两个平行平面的交线是两条平行线。

12.7 X 和 Y 分别是向量空间$\vec{X}$ 和$\vec{Y}$ 下的两个子空间,它们之间存在仿射变换$\psi: X\rightarrow Y$。说明这个仿射变换就是$\psi(P)=\psi(O)+\vec{\psi}(P-O)$,其中 O 是任意选取的原点,$\vec{\psi}:\vec{X}\rightarrow\vec{Y}$ 是从$\vec{X}$ 到$\vec{Y}$ 且与 O 无关的线性变换。

12.8 说明仿射摄像机(对应的外极几何)可以看成是焦距离场景无限远时的透视图像。

12.9 把第 10 章介绍的多线性方法推广到仿射情况。

12.10 证明定理 3。

12.11 证明标定过的类透视投影摄像机和满足下面约束的仿射摄像机等价。

$$\hat{\boldsymbol{a}}_1\cdot\hat{\boldsymbol{a}}_2 = \frac{u_r v_r}{2(1+u_r^2)}|\hat{\boldsymbol{a}}_1|^2 + \frac{u_r v_r}{2(1+v_r^2)}|\hat{\boldsymbol{a}}_2|^2 \quad , \qquad \frac{|\hat{\boldsymbol{a}}_1|^2}{(1+u_r^2)} = \frac{|\hat{\boldsymbol{a}}_2|^2}{(1+v_r^2)}$$

其中,(u_r,v_r)表示点 R 投影后的坐标。

12.12 一个 $m\times n$ 矩阵的元素是随机生成的($m\geqslant n$),它的 RREF 是什么样子?若 $m<n$,情况又会如何?并说明原因。

编程作业

12.13 实现 Koenderink-Van Doorn 的从运动估计仿射形状方法。

12.14 实现从图像对应估计外极几何的方法，和从投影矩阵估计场景结构的方法。

12.15 实现 Tomasi-Kanade 的从运动估计仿射形状方法。

12.16 在 RREF 方法中，若对u加一个[0,0.000 1]之间的均匀分布扰动，对 RREF 结果有什么影响(用 MATLAB 中的 RREF 函数)；再用“鲁棒”版的约减方法(容许非零的 RREF)计算，并比较两个结果。

第 13 章　从运动估计投影模型

这一章还是介绍如何从 m 个图像的 n 对对应点恢复场景结构或摄像机运动，但是这里要用透视模型。给定 n 个点 $P_j(j=1,\cdots,n)$ 和它们在 m 个摄像机观察下的齐次坐标 $\boldsymbol{p}_{ij}=(u_{ij},v_{ij},1)^T$，则可以写出如下的透视投影方程

$$\begin{cases} u_{ij}=\dfrac{\boldsymbol{m}_{i1}\cdot\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\cdot\boldsymbol{P}_j} \\ v_{ij}=\dfrac{\boldsymbol{m}_{i2}\cdot\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\cdot\boldsymbol{P}_j} \end{cases} \qquad i=1,\cdots,m \quad 和 \quad j=1,\cdots,n \tag{13.1}$$

其中，$\boldsymbol{m}_{i1}^T$，$\boldsymbol{m}_{i2}^T$ 和 $\boldsymbol{m}_{i3}^T$ 表示 3×4 投影矩阵 $\mathcal{M}_i$ 的各行，$\mathcal{M}_i$ 是第 i 个摄像机在某个坐标系下的投影矩阵，$\boldsymbol{P}_j$ 是点 P_j 在同一个坐标系下的齐次坐标。“从运动估计投影结构问题”就是从 mn 对图像对应点 $\boldsymbol{p}_{ij}$ 估计 m 个矩阵 $\mathcal{M}_i$ 和 n 个向量 $\boldsymbol{P}_j$ 的问题。

若 $\mathcal{M}_i$ 和 $\boldsymbol{P}_j$ 是方程(13.1)的解，对任意非零的 λ_i 和 μ_j，$\lambda_i\mathcal{M}_i$ 和 $\mu_j\boldsymbol{P}_j$ 也是解。特别地，在第 2 章中已经说明，满足方程(13.1)的矩阵 $\mathcal{M}_i$ 只定义到比例层次，有 11 个独立的参数；向量 $\boldsymbol{P}_j$ 也类似，有 3 个独立的参数(必要时，可以化简为形如 $(x_j,y_j,z_j,1)^T$ 的规范形式，一般情况下第 4 个坐标不为零)。和仿射情况类似，“从运动估计投影结构”这个名字就道出了其中的不确定性：若摄像机未标定，由定理 1(第 2 章)，投影矩阵 $\mathcal{M}_i$ 可以是秩为 3 的任意的 3×4 矩阵。因此，若 $\mathcal{M}_i$ 和 $\boldsymbol{P}_j$ 是方程(13.1)的解，则 $\mathcal{M}_i'=\mathcal{M}_i\mathcal{Q}$ 和 $\boldsymbol{P}_j'=\mathcal{Q}^{-1}\boldsymbol{P}_j$ 也是解，其中 $\mathcal{Q}$ 是一个投影变换矩阵(任意的非奇异 4×4 矩阵)。矩阵 $\mathcal{Q}$ 是定义在比例上的，有 15 个自由参数，因为把它乘一个非零系数等价于把 $\mathcal{M}_i$ 和 $\boldsymbol{P}_j$ 除以相同的倍数。因为方程(13.1)对 $11m$ 个矩阵参数和 $3n$ 个向量 $\boldsymbol{P}_j$ 共有 $2mn$ 个约束再加上问题本身的仿射不确定性，则若是问题有有限数量的解，需要 $2mn\geqslant11m+3n-15$。对 $m=2$ 的情况，7 对对应点就足够确定(在投影意义下)两个投影矩阵和任何其他点的位置。这将在 13.2 节和 13.3 节正式证明。

在本章的后续部分，透视几何代替了仿射几何在第 12 章中的位置，它提供了类似的通用方法。(最初)先不考虑已标定摄像机的欧氏约束，因此就可以用线性方法恢复场景的投影结构和摄像机运动；然后再讨论已标定的透视摄像机的(部分或全部)几何约束，从投影重构升级为欧氏重构。

13.1　投影几何基础

在投影几何下的距离度量比仿射几何下的更基本。此处不再有仿射时的平行线间的比例，实际上，连平行的概念都没有了。仍然保留点、直线和面的概念，但是，引入了一种对共线点间距离更弱的比例测量——交比。与仿射情形类似，我们不会以定理证明的方式介绍投影几何原理，而只是停留在一种非正式的层次上。

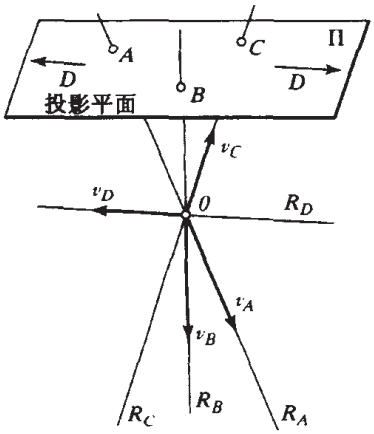
13.1.1　投影空间

考虑一个 $n+1$ 维的实向量空间 $\vec{X}$。如果 $\boldsymbol{v}$ 是 $\vec{X}$上的一个非零元素，那么集合$\mathbb{R}\boldsymbol{v}$ 是所有与 $\boldsymbol{v}$ 成比例的点组成的集合，称为射线，它可以用其中任一个非零元素代表。由 $\vec{X}$确定的 n 维实投影空间 $X=P(\vec{X})$是指 $\vec{X}$中的射线集合，或者说，在等价关系 $\boldsymbol{v}\sim\boldsymbol{v}'$下，$\vec{X}$中一个非零向量集合$(\vec{X}\setminus 0)$的商，这里的等价关系是指“$\boldsymbol{v}\sim\boldsymbol{v}'$当且仅当存在$k\in\mathbb{R}$，使得 $\boldsymbol{v}=k\boldsymbol{v}'$”。$X$ 中的元素称为点，我们称一组点是线性相关(无关)的，若代表这些射线的向量是线性相关(无关)的。映射 $\vec{X}\setminus 0\to P(\vec{X})$建立了 $\vec{X}$中任一非零向量和 X 中点$p(\boldsymbol{v})$之间的联系。

例 13.1　$P(\mathbb{R}^3)$下的模型

考虑$\mathbb{R}^3$ 下的仿射平面 Π。$\mathbb{R}^3$ 中与 Π 不平行的射线一一对应于这个平面上的点。例如，由 $\boldsymbol{v}_A,\boldsymbol{v}_B,\boldsymbol{v}_C$ 确定的射线R_A,R_B,R_C 可以映射到它们与平面 Π 的交点 A,B,C。向量 $\boldsymbol{v}_A,\boldsymbol{v}_B,\boldsymbol{v}_C$ 是线性无关的，则(由定义)A,B,C 也是线性无关的。

随着射线越来越平行于平面 Π，它与平面的交点将趋向于无穷远，事实上，可以通过在 Π 上添加一个在与无穷远处的一维点集，对应与 Π 平行的射线，这样就能构造投影平面$P(\mathbb{R}^3)$的模型了(也就是，一个二维的与$\mathbb{R}^3$ 同构的投影空间 $\hat{\Pi}$)。例如，射线 R_D 平行于 Π，则它就被映射到 $\hat{\Pi}$ 上的无穷远点 D。



由于任意仿射平面都可以通过选取仿射坐标系映射到$\mathbb{R}^2$ 上，例 13.1 说明，仿射平面，$\mathbb{E}^3$ 或其他任意仿射平面，在选取适当的点表示无穷远后，都可以嵌入到投影空间内。在 13.1.3 节将介绍这个增补过程。

13.1.2　投影子空间和投影坐标系

设 $\vec{Y}$是 $\vec{X}$上的一个$(m+1)$维向量子空间。$\vec{Y}$上的射线集合 $Y=P(\vec{Y})$称为 X 的投影子空间，且维数为 m。给出 $\vec{Y}$的一组基$(\boldsymbol{e}_0,\boldsymbol{e}_1,\cdots,\boldsymbol{e}_m)$，则可以对 Y 上的每一点 P 建立一个$\mathbb{R}^{m+1}$上带一个参数的元素族——或者说，$\boldsymbol{v}\in\vec{Y}$满足 $P=p(\boldsymbol{v})$的向量的坐标$(x_0,x_1,\cdots,x_m)^T$。这个族的元素两两之间是成比例的，且一个作为代表的元素称为点 P 的齐次投影坐标集。

齐次坐标也可以从本质上解释为 Y 上的一族点：设有 $m+1(m\leqslant n)$个线性独立的点 A_0，

$A_1,\cdots,A_m$ 和 $m+1$ 个向量 $\boldsymbol{v}_i(i=0,1,\cdots,m)$ 表示对应的射线。若另外有线性相关的一点 A^*，而 $\boldsymbol{v}^*$ 表示对应的射线，则有

$$\boldsymbol{v}^* = \mu_0\boldsymbol{v}_0 + \mu_1\boldsymbol{v}_1 + \cdots + \mu_m\boldsymbol{v}_m$$

系数 μ_i 不是惟一的，因为每个向量 $\boldsymbol{v}_i$ 只定义到了比例层次。但是，若任何的 μ_i 都不为 0（例如，$\boldsymbol{v}^*$ 不在任何 m 个向量 $\boldsymbol{v}_i$ 张成的子空间上，或等价的，对应点线性独立），我们可以惟一地定义 $m+1$ 个非零向量 $\boldsymbol{e}_i=\mu_i\boldsymbol{v}_i$，使得

$$\boldsymbol{v}^* = \boldsymbol{e}_0 + \boldsymbol{e}_1 + \cdots + \boldsymbol{e}_m$$

特别地，任何与 $\boldsymbol{v}_i$ 线性相关的向量 $\boldsymbol{v}$ 都可以惟一地写成

$$\boldsymbol{v} = x_0\boldsymbol{e}_0 + x_1\boldsymbol{e}_1 + \cdots + x_m\boldsymbol{e}_m$$

这就定义了一个 $\mathbb{R}^{m+1}$ 上射线 $\mathbb{R}(x_0,x_1,\cdots,x_m)^T$ 与 X 上子空间 S_m 之间的一一对应。S_m 实际上就是 $\vec{X}$ 上由 $\boldsymbol{v}_i$（或者等价的 $\boldsymbol{e}_i$）张成的子空间 $\vec{Y}$ 对应的投影空间。若 $P=p(\boldsymbol{v})$ 是 S_m 上与射线相关的点，则 $x_0,x_1,\cdots,x_m$ 称为 P 在由 $m+1$ 个基础点 A_i 和单位点 A^* 确定的投影坐标系下的齐次（投影）坐标。注意，由于一个射线对应的向量 $\boldsymbol{v}$ 只是定义到比例层次，所以点的齐次坐标也只到比例层次。

在这个对应的投影坐标系下，很容易验证它的基础点和单位点有非常简单的形式——如下，

$$\boldsymbol{A}_0=\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\ \boldsymbol{A}_1=\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix},\ \ldots,\ \boldsymbol{A}_m=\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix},\ \boldsymbol{A}^*=\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}$$

本节同时介绍了两种齐次坐标系的描述。它惟一的区别是在第一种方法中，坐标向量 $\boldsymbol{e}_0,\boldsymbol{e}_1,\cdots,\boldsymbol{e}_m$ 的选取已经事先给定，而第二种方法则是从一些点构造出一个投影坐标系。

例 13.2 投影坐标变换

给定三维投影空间 X 上的坐标系 $(A)=(A_0,A_1,A_2,A_3,A^*)$，我们可以定义任意点 P 的（齐次投影）坐标 ${}^AP=({}^Ax_0,{}^Ax_1,{}^Ax_2,{}^Ax_3)^T$。我们考虑另一个投影坐标系 $(B)=(B_0,B_1,B_2,B_3,B^*)$。显然（习题）对应的坐标变换可以写成

$$\rho^BP = {}^B_A\mathcal{T}\,{}^AP \tag{13.2}$$

其中，${}^B_A\mathcal{T}$ 是一个只定义到比例的 4×4 投影变换矩阵，ρ 是一个比例因子，以保证等号两边相等。现在我们来计算这个矩阵。把确定坐标系 (A) 的点代入方程(13.2)，可得

$$\rho_0{}^BA_0={}^B_A\mathcal{T}\begin{pmatrix}1\\0\\0\\0\end{pmatrix},\ \rho_1{}^BA_1={}^B_A\mathcal{T}\begin{pmatrix}0\\1\\0\\0\end{pmatrix},\ \rho_2{}^BA_2={}^B_A\mathcal{T}\begin{pmatrix}0\\0\\1\\0\end{pmatrix},\ \rho_3{}^BA_3={}^B_A\mathcal{T}\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

和

$$\rho^{*B}A^*={}^B_A\mathcal{T}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

由于矩阵${}_A^B\mathcal{T}$只定义到比例，我们可以取 $\rho^* = 1$，则有

$$ {}_A^B\mathcal{T} = \begin{pmatrix} \rho_0 {}^BA_0 & \rho_1 {}^BA_1 & \rho_2 {}^BA_2 & \rho_3 {}^BA_3 \end{pmatrix} $$

其中，因子 ρ_i 是线性方程的解：

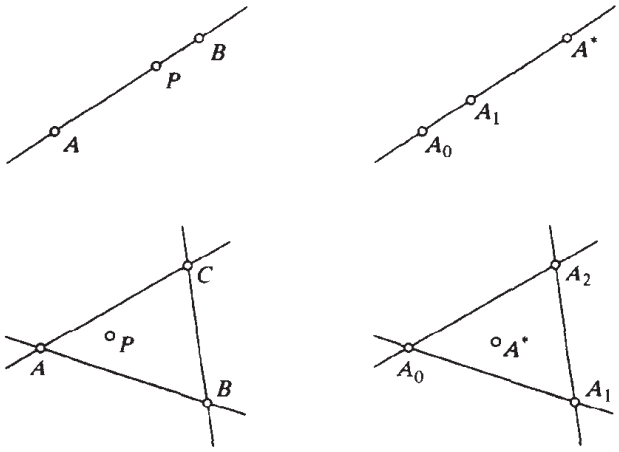
$$ \begin{pmatrix} {}^BA_0 & {}^BA_1 & {}^BA_2 & {}^BA_3 \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = {}^BA^* $$

注意这与第 2 章和第 12 章中欧氏坐标系和仿射坐标系的变换很像。任意有限维空间上的投影空间变换也可以写成类似的形式。

X 的一维投影子空间称为线。从二维到 $n-1$ 维的线性子空间分别称为平面和超平面。超平面 S_{n-1}由 n 个线性无关的点 $P_0, P_1, \cdots, P_{n-1}$的线性组合构成。

例 13.3　投影线和平面

一条投影线由它上面的两个点 A 和 B 惟一确定，但是定义一个投影坐标系需要三个不同的点 A_0, A_1 和 A^*。同样地，一个平面可以由三个点惟一确定，但是确定投影坐标系需要平面上的 4 个点：三个构成三角形的非退化的基础点 A_0, A_1, A_2 和不在三角形任意一边上的单位点 A^*。



但是，若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 表示这两个点在某投影坐标下的方程，则直线上任意一点 $\boldsymbol{P}$ 的坐标可以写成 $\boldsymbol{P} = \lambda\boldsymbol{A} + \mu\boldsymbol{B}$。由 A 和 B 代表的射线 R_A 和 R_B 是线性独立的，但是对于同一直线上的点 P，它对应的射线 R_P 处在由 R_A 和 R_B 定义的直线上。同样，若 $\boldsymbol{A}, \boldsymbol{B}$ 和 $\boldsymbol{C}$ 表示在某一坐标系下三个不共线的点，则这个平面上任一点的坐标向量 $\boldsymbol{P}$ 可以惟一写成 $\boldsymbol{P} = \lambda\boldsymbol{A} + \mu\boldsymbol{B} + \nu\boldsymbol{C}$。

13.1.3　仿射和投影空间

例 13.1(非正式地)说明了如何通过在无穷远处添加一维点集把仿射平面嵌入投影平面的方法。更通用的结论是，对 n 维仿射空间 X，可以通过在无穷远处添加与直线方向有关的点集来构造投影闭集合 $\tilde{X}$，这些点构成 $\tilde{X}$上的超平面，称为无穷远处的超平面，记为 ∞_X。

我们取 X 上的点 A，，并引入 $\tilde{X} \stackrel{\text{def}}{=} P(\vec{X} \times \mathbb{R})$，其中，$\vec{X}$是 X 的实际向量空间。我们可以通

过插入映射 $J_A: X \to \tilde{X}[J_A(P) = p(\overrightarrow{AP}, 1)]$把 X 嵌入 $\tilde{X}$中(见图 13.1)①。如前所述,在 $\tilde{X}$下,$J_A(X)$的补是超平面 $\infty_X \stackrel{\text{def}}{=} P(\vec{X} \times \{0\})$。

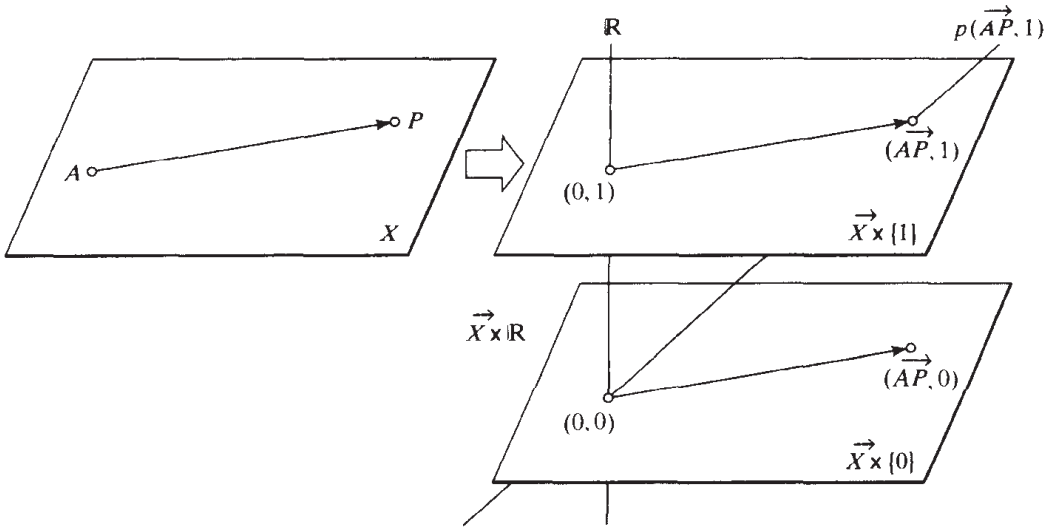


图 13.1 仿射空间的投影增补

现在考虑 X 的一个固定的仿射坐标系$(A_0, A_1, \cdots, A_n)$,并用 J_{A_0}把 X 嵌入$\tilde{X}$。向量$\overrightarrow{A_0A_i}$($i = 1, \cdots, n$)构成了 $\vec{X}$的一组基,则 $n+1$ 个向量 $\boldsymbol{e}_i \stackrel{\text{def}}{=} (\overrightarrow{A_0A_i}, 0)$($i = 1, \cdots, n$)和 $\boldsymbol{e}_{n+1} \stackrel{\text{def}}{=} (\mathbf{0}, \mathbf{1})$构成了$\vec{X} \times \mathbb{R}$上的一组基。例如,若$(x_1, \cdots, x_n)$表示点 P 在基$(A_0, A_1, \cdots, A_n)$下的仿射坐标,则有

$$
\begin{aligned}
J_{A_0}(P) &= p(\overrightarrow{A_0P}, 1) = p(x_1\overrightarrow{A_0A_1} + \cdots + x_n\overrightarrow{A_0A_n}, 1) \\
&= p(x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n + \boldsymbol{e}_{n+1})
\end{aligned}
$$

且在由向量$(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_{n+1})$构成的 $\vec{X} \times \mathbb{R}$上的基下的齐次投影坐标 $J_{A_0}(P)$为$(x_1, \cdots, x_n, 1)$。另一方面,在无穷远点中点的坐标为$(x_1, \cdots, x_n, 0)$。特别要提到的是,这个投影增补过程与第 2 章中引入的,用齐次坐标描述图像和景物的方法是一致的,并将在本书中一直沿用。

引入无穷远点的方法使投影几何学能避免在仿射模型中出现大量例外的情况。例如仿射平面 Π 下的平行直线永不相交,除非它们重合在一起。相反,投影平面下的两条不同直线一定交于一个点(这是因为对应的向量空间交于一条射线),Π 下两条平行线将交于 $\tilde{\Pi}$中的一个无穷远点,这个点由平行线的方向决定(习题)。

13.1.4 超平面和对偶

如前所述,投影平面下的两条直线只有一个交点。同样的,两个点只属于一条直线。这两

① 我们把 X 和实际的向量空间 $\vec{X}$上的一点 P 的向量记做$\overrightarrow{AP}$。这个向量化过程显然是依赖于原点 A 选取的,但是容易说明 $\tilde{X}$又是不依赖于这个选取的。一个更严格的投影的增补过程是引入与仿射空间相关的泛向量空间,这里没有篇幅详细介绍。参考 Berger(1987,第 5 章)。注意,$p(\boldsymbol{v}, \lambda)$被写为 $p(\boldsymbol{v}, \lambda)$。

个结论构成了伴生公理，可以进而推出公理体系上的投影平面。点和线是对称的，更确切地说，是对偶的。

为了引入对偶的一般定义，我们假设在 n 维仿射空间 X 上已经有了一个坐标系，以及在超平面 S_{n-1} 上的 $n+1$ 个点 $P_0, P_1, \cdots, P_n$。由于这些点是线性相关的，由它们的坐标构成的 $(n+1)\times(n+1)$ 矩阵就是奇异的。把它的行列式按最后一列展开，有

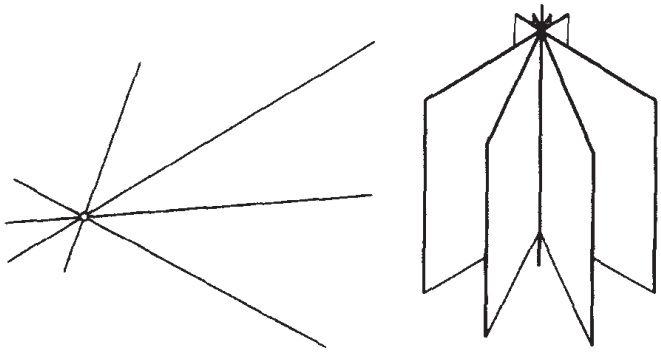
$$u_0x_0 + u_1x_1 + \cdots + u_nx_n = 0 \tag{13.3}$$

其中，$(x_0, x_1, \cdots, x_n)$ 表示 P_n 的齐次坐标，$(u_0, u_1, \cdots, u_n)$ 是 $P_0, P_1, \cdots, P_{n-1}$ 坐标的函数。注意没有设 P_n 的最后一维为 1，是为了强调系数 u_i 和 x_i 之间的对称性。

方程(13.3)对超平面 S_{n-1} 上的所有点 P_n 都成立，称为 S_{n-1} 的方程(注意与仿射情形的类似性)。反过来，任意形如式(13.3)的方程只要至少有一个系数 u_i 非零，它就是一个超平面的方程，由于式(13.3)中的 u_i 只定义到比例层次，则必定存在一个从 $\mathbb{R}^{n+1}$ 上射线到 X 上超平面间的一一映射，而且我们可以进一步用这些超平面定义第二个投影空间 $X^* = P(\vec{X}^*)$，称为 $\vec{X}$ 的对偶空间(由于 X^* 可以看成是 $\vec{X}$ 的对偶空间 $\vec{X}^*$ 的投影空间)。系数 $(u_0, u_1, \cdots, u_n)$ 定义了这个点在对应的 X^* 空间下超平面 S_{n-1} 下的齐次投影坐标，方程(13.3)也可以看成是通过点 P_n 的超平面的集合。

例 13.4　直线的对偶

在 $\mathbb{P}^2 \overset{\text{def}}{=} \tilde{\mathbb{E}}^2$ 下，点和直线是对偶的，在 $\mathbb{P}^3 \overset{\text{def}}{=} \tilde{\mathbb{E}}^3$，点和平面是对偶的，但是在 $\mathbb{P}^3$ 下，点和直线不是对偶的。那么在一般情况下，X 下直线的对偶是什么呢？直线是 X 上的一维线性子空间，它的所有元素都是直线上两点的线性组合。同样地，在 X^* 空间上直线的对偶称为一束超平面，其中所有元素都是其中两个超平面的线性组合。在平面内，直线的对偶是交在同一点的一束直线。



在三维空间下，直线的对偶是交于同一条直线的一束平面。

在本节的最后，我们不加证明地说：对 X 上点成立的任何几何定理，都会在超平面上(X^* 上的点)导出一个相对应的定理，反之亦然，这两个定理互为对偶。

13.1.5　交比和投影坐标

本节主要关心三维投影空间 $\tilde{\mathbb{E}}^3$ 的情况，一个点的非齐次投影坐标可以用交比定义。在仿射条件下，给定共线 4 个点 A, B, C, D，其中 A, B, C 不同，那么这些点的交比定义为

$$\{A, B; C, D\} \stackrel{\text{def}}{=} \frac{\overline{CA}}{\overline{CB}} \times \frac{\overline{DB}}{\overline{DA}}$$

其中，$\overline{PQ}$表示 P、Q 两点间的有向距离，方向是在连接它们的直线 Δ 上任意选取的。这个直线方向是固定的，但是可以任意选取，因为把它反相的话并不影响交比。注意，$\{A, B; C, D\}$是一个比例，只有在 $D \neq A$ 时才有意义，因为若 $D = A$ 则除数为零。我们可以扩展交比的定义到整个仿射直线，用符号 ∞ 表示非零实数被零除的结果，在整个投影直线 $\tilde{\Delta}$上 $\{A, B; C, \infty_{\Delta}\} = \overline{CA}/\overline{CB}$。反过来，给定投影直线 Δ 上三点 A，B 和 C，则存在惟一的投影变换 $h: \Delta \rightarrow \tilde{\mathbb{R}}$，把 Δ 映射到其投影增补 $\tilde{\mathbb{R}} = \mathbb{R} \cup \infty$ 上，且有 $h(A) = \infty$，$h(B) = 0$ 和 $h(C) = 1$。也可以用 $\{A, B; C, D\} \stackrel{\text{def}}{=} h(D)$定义交比。

给定投影坐标系(A_0, A_1, A^*)，直线 Δ 和线上一点 P 的齐次坐标(x_0, x_1)，可以定义非齐次坐标 $k_0 = x_0/x_1$。比例 k_0 有时称为 P 的投影系数，显然 $k_0 = \{A_0, A_1; A_2, P\}$。

前面说过，过点 O 的直线集合称为一束直线。4 条共面直线 Δ_1，Δ_2，Δ_3 和 Δ_4 的交比按如下方法定义：任取一条不过 O 点直线 Δ，与原有的 4 条直线的交点可以定义交比，可以证明这个交比与 Δ 的选择无关[图 13.2(a)]。

下面考虑同一束的 4 个平面 Π_1，Π_2，Π_3 和 Π_4，用 Δ 表示它们的公共交线，这些平面的交比定义为它们与任一个不包括 Δ 的平面 Π 的交线的交比[图 13.2(b)]。同样，这个交比也是与 Π 的选取无关的。

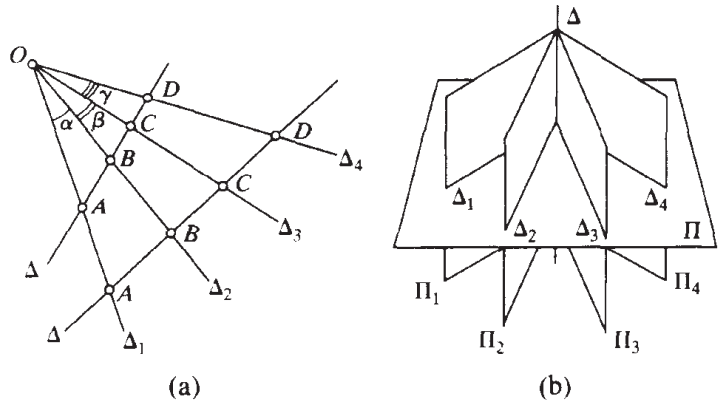


图 13.2 交比定义：(a)4 条直线，(b)4 个平面。习题中将证明，交比$\{A, B; C, D\}$只与角度 α，β和γ有关。具体说来，我们有$\{A, B; C, D\} = \{A', B'; C', D'\}$

在平面情况下，点 P 在基(A_0, A_1, A_2, A^*)下的非齐次坐标(k_0, k_1)定义为 $k_0 = x_0/x_2$，$k_1 = x_1/x_2$，且有

$$\begin{cases} k_0 = \{A_1A_0,\ A_1A_2;\ A_1A^*,\ A_1P\} \\ k_1 = \{A_0A_1,\ A_0A_2;\ A_0A^*,\ A_0P\} \end{cases}$$

其中，MN 表示过点 M 和 N 的直线，$\{\Delta_1, \Delta_2; \Delta_3, \Delta_4\}$表示一束直线 Δ_1，Δ_2，Δ_3，Δ_4 的交比。

类似的，点 P 在基$(A_0, A_1, A_2, A_3, A^*)$解的非其次坐标$(k_0, k_1, k_2)$定义为 $k_0 = x_0/x_3$，$k_1 = x_1/x_3$，$k_2 = x_2/x_3$，且有

$$\begin{cases} k_0 = \{A_1A_2A_0,\ A_1A_2A_3;\ A_1A_2A^*,\ A_1A_2P\} \\ k_1 = \{A_2A_0A_1,\ A_2A_0A_3;\ A_2A_0A^*,\ A_2A_0P\} \\ k_2 = \{A_0A_1A_2,\ A_0A_1A_3;\ A_0A_1A^*,\ A_0A_1P\} \end{cases}$$

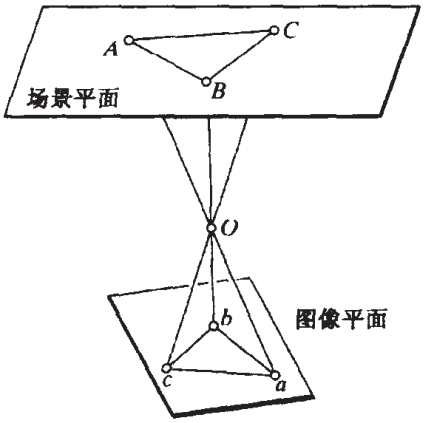
其中，LMN 是过 L、M 和 N 三点的平面，$\{\Pi_1,\Pi_2;\Pi_3,\Pi_4\}$ 表示一束平面 Π_1,Π_2,Π_3,Π_4 的交比。

13.1.6　投影变换

考虑两个向量空间 $\vec{X}$ 和 $\vec{X}'$ 之间的线性双射 $\vec{\psi}:\vec{X}\to\vec{X}'$。由于是线性映射，$\vec{\psi}$ 把 $\vec{X}$ 上的射线映射为 $\vec{X}'$ 上的射线，又由于是双射，它把非零向量映射为非零向量。若对 $\vec{X}$ 上的所有向量 $v\neq 0$ 定义 $\psi(p(\boldsymbol{v}))\overset{\text{def}}{=}p(\vec{\psi}(\boldsymbol{v}))$，则我们可以推出映射 $\psi:P(\vec{X})\to P(\vec{X}')$。映射 ψ 是一个双射，称为投影变换（或单应）。显然投影变换在映射合成的作用下组成一个群。若 $\vec{X}'=\vec{X}$，则这个群称为 $X=P(\vec{X})$ 的投影群。

例 13.5　共面点和它们的图像间的投影对应

设有 $\mathbb{E}^3$ 中的两个平面和平面外一点 O。习题中将说到，把第一个平面上（投影闭集内）的一点 A 映射到第二个平面与 AO 交点的透视投影映射是一个投影变换。



这个特性并不奇怪，因为按照例 13.1，这两个平面可以看做过 O 的射线对应的投影平面。

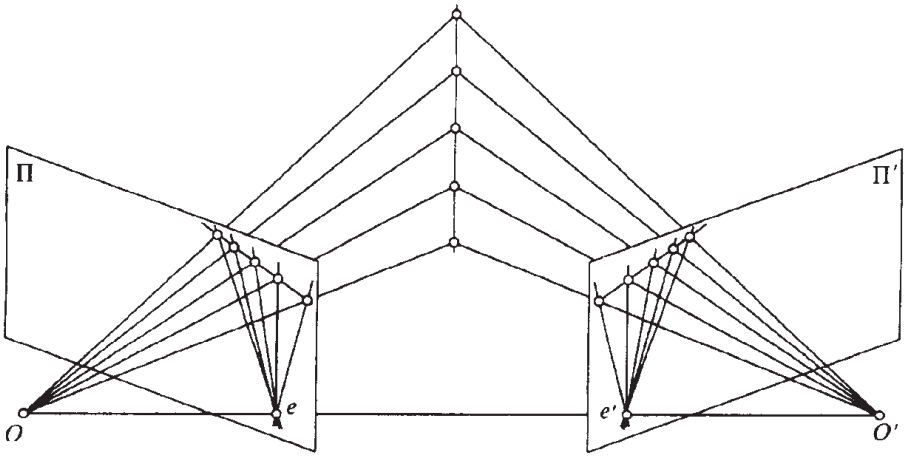
投影几何研究的是投影空间在变换中保持不变的性质。例如，其中一个不变性就是一组点的线性独立性（或相关性）。给定一个投影变换 $\psi:X\to X'$，我们考虑 $\vec{X}$ 中的 $m+1$ 个线性独立的向量 $\boldsymbol{v}_0,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m$ 以及对应的 X 中的点 $A_0,A_1,\cdots,A_m$。由于 $\vec{\psi}$ 是双射，向量 $\vec{\psi}(\boldsymbol{v}_i)$ 是线性独立的，因此点 $A'_i=\psi(A_i)$ 也是线性独立的。则有 $(A)=(A_0,A_1,\cdots,A_{n+1})$ 是 n 维投影空间 X 上的投影坐标系，$(A')=(A'_0,A'_1,\cdots,A'_{n+1})$ 也是 X' 上的投影坐标系。反过来，若已知两个 n 维投影空间 X 和 X'，它们的基分别为 $(A_0,A_1,\cdots,A_{n+1})$ 和 $(A'_0,A'_1,\cdots A'_{n+1})$，则存在惟一的变换 $\psi:X\to X'$ 使得对 $i=0,1,\cdots,n+1$，有 $\psi(A_i)=A'_i$。

投影坐标还产生了第二个不变量。实际上，由于实际映射 $\vec{\psi}$ 的线性性质，若点 P 在投影坐标系 $(A_0,A_1,\cdots,A_{n+1})$ 下的坐标为 $(x_0,x_1,\cdots,x_n)$，则点 $\psi(P)$ 在由 $A'_i=\psi(A_i)$ 构成的 X' 的坐标系下坐标相同。实际上，投影变换的一个特性就是把直线映射为直线，并保证交比不变（即投影坐标）。回到例子 13.5，若已知共面点集的像，则可以确定每个点的投影坐标系，其中坐标系是由平面上的 4 个点确定的。这在后面的章节中设计识别系统的不变量时是很有用的。

与刚体变换和仿射变换类似，若已知两个坐标系(F)和(F')后，两个 n 维投影空间之间的单应ψ可以用一个$(n+1)\times(n+1)$矩阵表示：这仍然是因为变换$\vec{\psi}$的线性性质所决定的。因此，若 $P'=\psi(P)$，则我们有${}^{F}P'=\mathcal{Q}^{F}P$，其中，$\mathcal{Q}$是一个$(n+1)\times(n+1)$非奇异矩阵，由于齐次投影坐标只定义比例关系，因此这个矩阵也只定义比例关系。

例 13.6 基础矩阵的参数化

这里重新考虑未标定摄像机确定外极几何的问题。在第 10 章中首次讨论过这个问题，那里只是不加证明地提出了基础矩阵的显式参数化描述。现在来构造这种参数化。我们把外极变换定义为从一组外极线到另一组外极线之间的映射。如下图所示，这个变换是一个单应。



实际上，过两个摄像机的外极平面分别构成了一束平面，它们的脊是连接两个光心的基线。外极平面与对应图像平面的交构成外极线，由任意 4 条外极线产生的交比与对应平面的交比是相同的。反过来，这就说明外极变换保持交比不变，因此是投影变换。

我们把外极点 e 和 e' 在对应图像坐标系内的(仿射)坐标记为$(\alpha,\beta)^T$ 和$(\alpha',\beta')^T$，且把对应外极线 l 和 l' 上的点的坐标记为$(u,v)^T$ 和$(u',v')^T$。由于外极变换这个线性映射把射线$\mathbb{R}(u-\alpha,v-\beta)^T$ 映射为射线$\mathbb{R}(u'-\alpha',v'-\beta')^T$，显然(习题)直线 l 和 l' 的斜度 τ 和 τ' 满足

$$\tau'=\frac{a\tau+b}{c\tau+d},\qquad \tau\stackrel{\text{def}}{=}\frac{v-\beta}{u-\alpha}\quad 和\quad \tau'\stackrel{\text{def}}{=}\frac{v'-\beta'}{u'-\alpha'}\tag{13.4}$$

把方程(13.4)的分母消去就得到 u、v 和 u'、v' 的双线性组合，可以写成 $\boldsymbol{p}^T\mathcal{F}\boldsymbol{p}'=0$，其中 $\mathcal{F}$ 已经在第 10 章给出，但没有证明，

$$\mathcal{F}=\begin{pmatrix} b & a & -a\beta-b\alpha \\ -d & -c & c\beta+d\alpha \\ d\beta'-b\alpha' & c\beta'-a\alpha' & -c\beta\beta'-d\beta'\alpha+a\beta\alpha'+b\alpha\alpha' \end{pmatrix}$$

13.1.7 投影形状

沿用仿射情况下的方法，若存在一个投影变换$\psi:X\to X$，使得点集 S' 是点集 S 在 ψ 下的像，我们称在投影空间 X 中两个点集 S 和 S' 是投影等价的。与仿射情形类似，显然投影等价是一个等价关系，X 中点集 S 的投影形状定义为所有投影等价的点集构成的等价类。类似地，从运动估计投影模型可以重新定义为从图像序列间的特征匹配恢复场景的投影形状(或由对

应的投影矩阵构成的等价类)问题。

13.2　从双目对应估计运动和投影结构

本章的剩余部分将关注如何从 m 幅图像上的 n 个特征点恢复场景的三维投影结构。这一节考虑两幅图像的情况,从三个或更多的视角恢复运动和结构的问题将在下两节讨论。我们假设外极点已知,在第 10 章中说过,这至少需要 7 对对应点。

13.2.1　几何场景重构

首先考虑已知外极点时用几何方法估计场景的投影模型。取合适的点做投影坐标,就可以利用投影结构固有的不确定性来简化重构问题。

假设观察到标定结构上 4 个不共面的点 A、B、C 和 D(见图 13.3)。设 O'(对应的 O'')表示第一个(第二个)摄像机的光心。对任意点 P,p'(p'')表示射线 $O'P$($O''P$)与平面 ABC 的交点。外极点为 e' 和 e'',基线交平面 ABC 于 E(显然,$E' = E'' = E$,$A' = A'' = A$)。

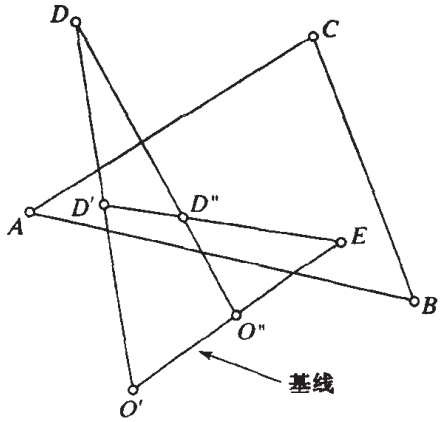


图 13.3　在 A、B、C、O'、O''构成的基下用几何重构 D 点的投影坐标

取 A、B、C、O' 和 O'' 为三维空间的基,目标是重构 D 的位置。取 a'、b'、c'、e' 为第一个图像平面的基,可以测出 d' 在这组基下的坐标,进而估计 D' 在基 A、B、C、E 下的坐标。类似的,通过第二个图像下 D'' 在基 a''、b''、c''、e'' 下的坐标 d'',可以估计点 D''。通过两条射线 $O'D'$ 和 $O''D''$ 的交,可以得到点 D。

现在用代数来描述这种几何重构。在由四面体 A,O'',O',B 和单位点 C 构成的基下计算投影坐标系和 D 点的非齐次投影坐标比较简单。坐标由下面三个交比定义

$$\begin{cases} k_0 = \{O''O'A,\ O''O'B;\ O''O'C,\ O''O'D\} \\ k_1 = \{O'AO'',\ O'AB;\ O'AC,\ O'AD\} \\ k_2 = \{AO''O',\ AO''B;\ AO''C,\ AO''D\} \end{cases}$$

通过计算这束外极平面与两个图像平面的交，可以马上从图像上的交比得到 k_0，k_1 和 k_2。

$$\begin{cases} k_0 = \{e'a', e'b'; e'c', e'd'\} = \{e''a'', e''b''; e''c'', e''d''\} \\ k_1 = \{a'e', a'b'; a'c', a'd'\} \\ k_2 = \{a''e'', a''b''; a''c'', a''d''\} \end{cases} \tag{13.4}$$

图 13.4 是使用这种方法的一个例子，两个弱标定的摄像机共有 46 对对应点。图 13.4(a)显示了输入图像和点对应，图 13.4(b)是场景的一幅投影重构结果，用来渲染场景的原始投影坐标。由于这种显示方法不是很清楚，我们在图 13.4(c)中显示的结果是在景物点上作用了一个投影变换，把三个参考点(用小圈标出)和摄像机中心映射到标定过的欧氏位置。作为比较，实际位置也一起显示出来。

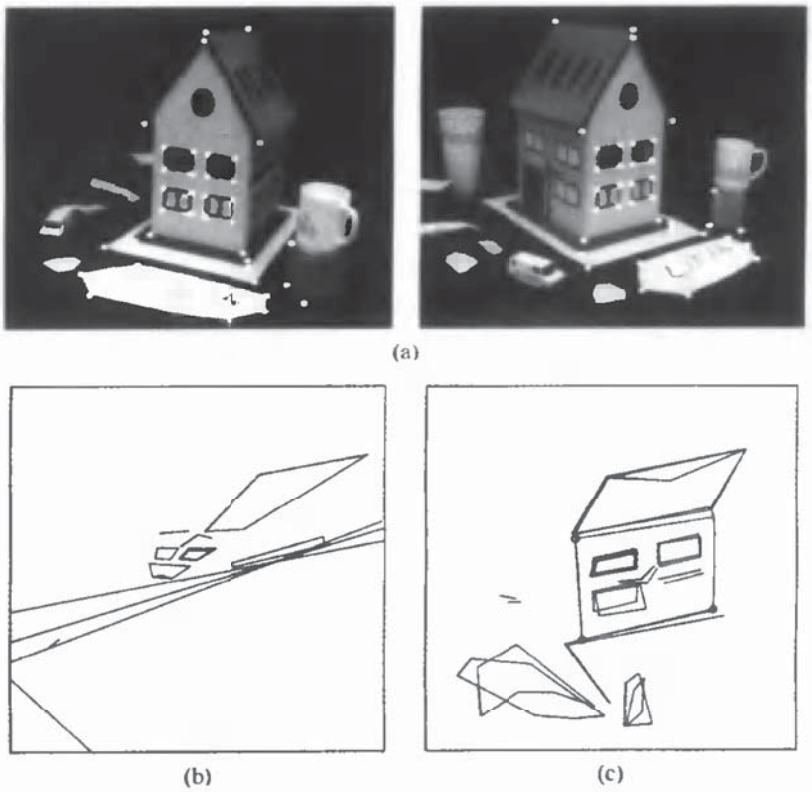


图 13.4 几何点重构：(a)输入数据，(b)原始的投影坐标，(c)修正的投影坐标

13.2.2 代数运动估计

这一节将完全从代数角度解决从双目点对应估计场景投影形状的问题，同样假设立体场景是弱标定的。则第 2 章的透视投影方程(2.15)很自然地扩展到增补的$\mathbb{E}^3$空间，且在这个空间的任意坐标系中形式都相同。实际上，若在某个欧氏坐标系(F)中方程(2.15)写成 $z\boldsymbol{p} = \mathcal{M}\boldsymbol{P}$，则在投影坐标系($F'$)下有相似的方程 $z\boldsymbol{p} = \mathcal{M}'\boldsymbol{P}'$，其中 $\boldsymbol{P}' = {}^{F'}\boldsymbol{P} = {}^{F'}_{F}\mathcal{T}\boldsymbol{P}$ 和$\mathcal{M}' = \mathcal{M}{}^{F}_{F'}\mathcal{T}^{-1}$。

具体说来，设有 5 个点 A_0, A_1, A_2, A_3, A_4，把它们当做$\mathbb{E}^3$ 的基，而 A_4 作为单位点。设一个摄像机在拍摄这些点的投影矩阵为$\mathcal{M}$，设 a_0，a_1，a_2，a_3，a_4 是这些点的像，选取 a_0 和 a_3 作为图像平面上投影的基，a_3 为单位点。又设 α, β 和 γ 为 a_4 在这个基下的坐标。

对 $i=0,1,2,3,4$，有 $z_i\boldsymbol{a}_i = \mathcal{M}\boldsymbol{A}_i$，则有

$$\mathcal{M} = \begin{pmatrix} z_0 & 0 & 0 & z_3 \\ 0 & z_1 & 0 & z_3 \\ 0 & 0 & z_2 & z_3 \end{pmatrix} \quad 和 \quad \begin{cases} z_4\alpha = z_0 + z_3 \\ z_4\beta = z_1 + z_3 \\ z_4\gamma = z_2 + z_3 \end{cases}$$

由于投影矩阵只定义比例关系，所以可以把它的系数都除以 z_3，并定义 $\lambda = z_4/z_3$，则有

$$\mathcal{M} = \begin{pmatrix} \lambda\alpha - 1 & 0 & 0 & 1 \\ 0 & \lambda\beta - 1 & 0 & 1 \\ 0 & 0 & \lambda\gamma - 1 & 1 \end{pmatrix}$$

若同样场景的第二幅图像也已知，设投影矩阵为$\mathcal{M}'$，图像点为 $a'_0, a'_1, a'_2, a'_3, a'_4$，会得到同样的结果，我们有

$$\mathcal{M}' = \begin{pmatrix} \lambda'\alpha' - 1 & 0 & 0 & 1 \\ 0 & \lambda'\beta' - 1 & 0 & 1 \\ 0 & 0 & \lambda'\gamma' - 1 & 1 \end{pmatrix}$$

这样，两个参数 λ 和λ'就完全确定了这两个摄像机的体视结构。可以利用框架的外极几何来计算这些参数。设 C 为第一个摄像机的光心，e' 是第二个摄像机成像平面上的外极点，坐标向量 $\boldsymbol{C}$ 和 $\boldsymbol{e}'$ 在对应的投影基上。我们有$\mathcal{M}\boldsymbol{C} = \boldsymbol{0}$，且有

$$\boldsymbol{C} = \left(\frac{1}{1-\lambda\alpha}, \frac{1}{1-\lambda\beta}, \frac{1}{1-\lambda\gamma}, 1\right)^T$$

把$\mathcal{M}'\boldsymbol{C} = \boldsymbol{e}'$代入，则有

$$\boldsymbol{e}' = \left(1 - \frac{\lambda'\alpha' - 1}{\lambda\alpha - 1}, 1 - \frac{\lambda'\beta' - 1}{\lambda\beta - 1}, 1 - \frac{\lambda'\gamma' - 1}{\lambda\gamma - 1}\right)^T$$

若μ'和ν'表示 e' 在由 a'_i 构成的基上的非齐次坐标，则最终得到

$$\begin{cases} \mu'(\lambda\gamma - \lambda'\gamma')(\lambda\alpha - 1) = (\lambda\alpha - \lambda'\alpha')(\lambda\gamma - 1) \\ \nu'(\lambda\gamma - \lambda'\gamma')(\lambda\beta - 1) = (\lambda\beta - \lambda'\beta')(\lambda\gamma - 1) \end{cases} \tag{13.5}$$

像式(13.5)这样的有两个未知数 λ 和λ'的二次方程组，一般会有 4 个解，可以看做平面(λ, λ')上由两个方程确定的两条圆锥曲线的 4 个交点。考察方程(13.5)可以马上得到两组解$(\lambda, \lambda') = (0,0)$和$(\lambda, \lambda') = (1/\gamma, 1/\gamma')$。很容易(有些枯燥)得到另外两组解是相同的(几何上，两条圆锥曲线在交点处是相切的)，对应的 λ 和λ'为

$$\lambda = \frac{\mathrm{Det}\begin{pmatrix} \mu' & \alpha & \alpha' \\ \nu' & \beta & \beta' \\ 1 & \gamma & \gamma' \end{pmatrix}}{\mathrm{Det}\begin{pmatrix} \mu'\alpha & \alpha & \alpha' \\ \nu'\beta & \beta & \beta' \\ \gamma & \gamma & \gamma' \end{pmatrix}}, \quad \lambda' = \frac{\mathrm{Det}\begin{pmatrix} \mu & \alpha & \alpha' \\ \nu & \beta & \beta' \\ 1 & \gamma & \gamma' \end{pmatrix}}{\mathrm{Det}\begin{pmatrix} \mu\alpha' & \alpha & \alpha' \\ \nu\beta' & \beta & \beta' \\ \gamma' & \gamma & \gamma' \end{pmatrix}}$$

这些值惟一地确定了投影矩阵$\mathcal{M}$和$\mathcal{M}'$。注意，考虑第二个外极点并不会增加独立的约束，因为外极约束就是外极线上的匹配约束。若投影矩阵已知，就只是重构场景点的问题。

13.3 多线性约束估计投影运动

上两节给出的方法是恢复场景相对于5个点的结构，那么重构的质量就和这些点的位置息息相关。与之不同，这一节介绍的方法是以一致的方式考虑了所有点，并用第10章用到的多线性约束方法以对应的投影矩阵的形式重构摄像机运动。

13.3.1 从基础矩阵估计运动

假设已经从双目对应恢复了两幅图像的基础矩阵$\mathcal{F}$，与仿射情形类似，利用从运动估计投影结构固有的不确定性，投影矩阵可以由$\mathcal{F}$的参数化表示出来。由于在投影情形下，摄像机运动和场景结构可能相差任意一个投影变换，通过一个合适的4×4矩阵$\mathcal{Q}$，可以把两个矩阵简化为规范形式$\tilde{\mathcal{M}}=\mathcal{M}\mathcal{Q}$和$\tilde{\mathcal{M}}'=\mathcal{M}'\mathcal{Q}'$。可以使$\tilde{\mathcal{M}}'$是$(\mathrm{Id}\ \mathbf{0})$的倍数，而让$\tilde{\mathcal{M}}$是任意形式。这个约减过程可以确定$\mathcal{Q}$的11个参数，并通过舍弃$\mathcal{Q}$剩下的4个自由度使$\tilde{\mathcal{M}}$简化。

下面用规范形式的$\tilde{\mathcal{M}}'$来推导基础矩阵的新形式。若$\tilde{\boldsymbol{P}}=(x,y,z,1)^T$表示点$P$的在世界坐标系中的齐次坐标向量，则可以把两个摄像机的投影方程写成$z\boldsymbol{p}=(\mathcal{A}\ \ \boldsymbol{b})\tilde{\boldsymbol{P}}$和$z'\boldsymbol{p}'=(\mathrm{Id}\ \mathbf{0})\tilde{\boldsymbol{P}}$，或等价地

$$z\boldsymbol{p}=\mathcal{A}(\mathrm{Id}\quad\mathbf{0})\tilde{\boldsymbol{P}}+\boldsymbol{b}=z'\mathcal{A}\boldsymbol{p}'+\boldsymbol{b}$$

这又可以推出$z\boldsymbol{b}\times\boldsymbol{p}=z'\boldsymbol{b}\times\mathcal{A}\boldsymbol{p}'$，再用$\boldsymbol{p}$的点积形式表示，可得

$$\boldsymbol{p}^T\mathcal{F}\boldsymbol{p}'=0\quad，其中\quad\mathcal{F}=[\boldsymbol{b}_\times]\mathcal{A}$$

注意这个表达式与第10章推出的基本矩阵的表达式很像。

具体说来，有$\mathcal{F}^T\boldsymbol{b}=0$，则(和想像的一样)$\boldsymbol{b}$是图像坐标系下的外极点的齐次坐标。矩阵$\mathcal{F}$这种新的参数表示提供了一种计算投影矩阵$\tilde{\mathcal{M}}$的简单方法。首先注意，由于$\tilde{\mathcal{M}}$整体的比例是不变的，不妨设$|\boldsymbol{b}|=1$。这样就可以用最小二乘法解$\mathcal{F}^T\boldsymbol{b}=0$得到$b$，然后，$\mathcal{A}$可以通过$\mathcal{A}_0=-[\boldsymbol{b}_\times]\mathcal{F}$得到。显然，对任意向量$\boldsymbol{a}$，$[\boldsymbol{a}_\times]^2=\boldsymbol{a}\boldsymbol{a}^T-|\boldsymbol{a}|^2\mathrm{Id}$；则有

$$[\boldsymbol{b}_\times]\mathcal{A}_0=-[\boldsymbol{b}_\times]^2\mathcal{F}=-\boldsymbol{b}\boldsymbol{b}^T\mathcal{F}+|\boldsymbol{b}|^2\mathcal{F}=\mathcal{F}$$

因为$\mathcal{F}^T\boldsymbol{b}=0$且$|\boldsymbol{b}|^2=1$，说明$\tilde{\mathcal{M}}=(\mathcal{A}_0\ \ \boldsymbol{b})$是问题的解。在练习中将看到，方程的解实际上是4个参数的族，通用形式为

$$\tilde{\mathcal{M}}=(\mathcal{A}\quad\boldsymbol{b})\quad 其中\quad\mathcal{A}=\lambda\mathcal{A}_0+(\ \mu\boldsymbol{b}\mid\nu\boldsymbol{b}\mid\tau\boldsymbol{b}\ )$$

和想像的一样，4个参数对应投影变换$\mathcal{Q}$剩下的4个自由度。一旦$\tilde{\mathcal{M}}$已知，通过用最小二乘法解方程$z\boldsymbol{p}=z'\mathcal{A}\boldsymbol{p}'+\boldsymbol{b}$中的$z$和$z'$，就可以得到任何一点$P$的坐标。

13.3.2 从三摄像机估计运动

下面写出第10章引入的三摄像机对应的三线性约束的投影模型。和上一节一样，我们可

以用合适的 4×4 矩阵对投影矩阵做后处理，使它们形如

$$\tilde{\mathcal{M}}_1=(\mathrm{Id}\quad \mathbf{0}),\quad \tilde{\mathcal{M}}_2=(\mathcal{A}_2\quad \boldsymbol{b}_2),\qquad \tilde{\mathcal{M}}_3=(\mathcal{A}_3\quad \boldsymbol{b}_3)$$

在这个变换下，$\boldsymbol{b}_2$ 和 $\boldsymbol{b}_3$ 仍然可以解释为外极点 e_{12} 和 e_{13} 的齐次图像坐标，而方程(10.14)和方程(10.15)的三线性约束仍然满足，三摄像机模型要用三个矩阵来定义①

$$\mathcal{G}_1^i=\boldsymbol{b}_2\boldsymbol{A}_3^{iT}-\boldsymbol{A}_2^i\boldsymbol{b}_3^T \tag{13.6}$$

其中，$\boldsymbol{A}_2^i$ 和 $\boldsymbol{A}_3^i(i=1,2,3)$ 表示 $\mathcal{A}_2$ 和 $\mathcal{A}_3$ 的各列。

假设已经按第 10 章的方法从点或线的对应关系估计了三焦张量，这一节的目的是恢复投影矩阵 $\tilde{\mathcal{M}}_2$ 和 $\tilde{\mathcal{M}}_3$。可以看到

$$(\boldsymbol{b}_2\times\boldsymbol{A}_2^i)^T\mathcal{G}_1^i=\left[(\boldsymbol{b}_2\times\boldsymbol{A}_2^i)^T\boldsymbol{b}_2\right]\boldsymbol{A}_3^{iT}-\left[(\boldsymbol{b}_2\times\boldsymbol{A}_2^i)^T\boldsymbol{A}_2^i\right]\boldsymbol{b}_3^T=\mathbf{0}$$

和

$$\mathcal{G}_1^i(\boldsymbol{b}_3\times\boldsymbol{A}_3^i)=\left[\boldsymbol{A}_3^{iT}(\boldsymbol{b}_3\times\boldsymbol{A}_3^i)\right]\boldsymbol{b}_2-\left[\boldsymbol{b}_3^T(\boldsymbol{b}_3\times\boldsymbol{A}_3^i)\right]\boldsymbol{A}_2^i=\mathbf{0}$$

矩阵 $\mathcal{G}_1^i$ 是奇异的(第 10 章已经提到)且向量 $\boldsymbol{b}_2\times\boldsymbol{A}_2^i$ 和 $\boldsymbol{b}_3\times\boldsymbol{A}_3^i$ 分别在它的左零空间和右零空间内。这又说明，当已知三焦张量后，我们只要计算矩阵 $\mathcal{G}_1^i(i=1,2,3)$ 的左零(右零)空间中的公共法线就可以得到 $\boldsymbol{b}_2$ 和 $\boldsymbol{b}_3$ 了。

已知外极点后，对 $i=1,2,3$ 列出方程(13.6)，则对 $\mathcal{A}_j(j=2,3)$ 的 18 个未知数可以列出 27 个方程。可以用最小二乘法在比例层次上求解。或者，可以利用对应点对和对应线对相关的三线性约束直接估计矩阵 $\mathcal{A}_j$，将三焦张量的多数写成这些矩阵的函数，就再一次导致到一个线性估计过程。

得到投影矩阵后，利用投影方程可以列出所观察的点和线的齐次坐标向量的线性约束，就可以恢复场景的投影结构了。

13.4　多幅图像恢复运动和投影结构

13.3 节利用外极约束和三焦点约束从两幅或三幅图像中，重构摄像机运动和对应的场景结构。类似地，第 10 章引入的四重焦点模型理论上也能估计每个摄像机的投影矩阵，进而估计对应的场景投影结构。但是多线性约束不能对 $m>4$ 的情况提供一种通用的方法。双目、三目、四目的运动和结构参数估计方法必须一个一个地推导。现在介绍另外一种方法，可以用非线性优化方法把所有图像综合考虑。

13.4.1　用分解因子方法解从运动估计投影模型问题

这一节，我们用一种分解因子的方法分析摄像机运动，这是第 12 章中 Tomasi-Kanade 方法在投影情况下的扩展。已知 m 幅图像和 n 个对应点，方程(13.1)可以写成

$$\mathcal{D}=\mathcal{M}\mathcal{P} \tag{13.7}$$

其中

① 严格地讲，用 $\mathcal{Q}$ 后乘这三个投影矩阵与将在第 10 章定义的(未标定的)三焦张量等式中的单位矩阵等于标定矩阵 $\mathcal{K}$，的效果是一样的，但是要注意这里并没有假设标定参数已知，而将投影坐标适当改变来简化投影矩阵的形式。

$$\mathcal{D} \stackrel{\text{def}}{=} \begin{pmatrix} z_{11}\boldsymbol{p}_{11} & z_{12}\boldsymbol{p}_{12} & \cdots & z_{1n}\boldsymbol{p}_{1n} \\ z_{21}\boldsymbol{p}_{21} & z_{22}\boldsymbol{p}_{22} & \cdots & z_{2n}\boldsymbol{p}_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ z_{m1}\boldsymbol{p}_{m1} & z_{m2}\boldsymbol{p}_{m2} & \cdots & z_{mn}\boldsymbol{p}_{mn} \end{pmatrix}, \quad \mathcal{M} \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{M}_1 \\ \mathcal{M}_2 \\ \cdots \\ \mathcal{M}_m \end{pmatrix}, \quad \mathcal{P} \stackrel{\text{def}}{=} (\boldsymbol{P}_1\ \boldsymbol{P}_2\ \cdots\ \boldsymbol{P}_n)$$

由于 $3m \times n$ 矩阵$\mathcal{D}$是 $3m \times 4$ 矩阵和 $4 \times n$ 矩阵的乘积，因此秩(最大)为 4；若投影深度 z_{ij} 已知，就能和仿射情况类似的通过$\mathcal{D}$的奇异值分解计算出$\mathcal{M}$和$\mathcal{P}$。另一方面，若$\mathcal{M}$和$\mathcal{P}$已知，则可以通过方程(13.7)读出投影深度 z_{ij}。也就是说我们可以通过一种迭代方法估计 z_{ij}，$\mathcal{M}$和$\mathcal{P}$，每一步中都假设某些量固定，来估计其他的量。

要使$\mathcal{D} - \mathcal{M}\mathcal{P}$——的 Frobenius 模的平方即

$$E \stackrel{\text{def}}{=} |\mathcal{D} - \mathcal{M}\mathcal{P}|^2 = \sum_{i,j} |z_{ij}\boldsymbol{p}_j - \mathcal{M}_i\boldsymbol{P}_j|^2$$

对未知的$\mathcal{M}_i$、$\boldsymbol{P}_j$ 和 z_{ij} 求最小。注意 E 的最小化是病态的，除非参数$\mathcal{M}_i$、$\boldsymbol{P}_j$ 和 z_{ij} 还有别的约束。前面已经说过，这些未知量之间确实不是无关的：矩阵$\mathcal{M}_i$ 和向量 $\boldsymbol{P}_j$ 只定义比例层次。若 $\mathcal{M}_i$、$\boldsymbol{P}_j$ 和 z_{ij} 是方程(13.1)的解，则对任意的比例系数 α_i 和 β_j，$\alpha_i\mathcal{M}_i$、$\beta_j\boldsymbol{P}_j$ 和 $\alpha_i\beta_j z_{ij}$ 也是解。另外，$\mathcal{M}_i = 0$、$\boldsymbol{P}_j = 0$ 和 $z_{ij} = 0$ 也总是解。实际上，方程的不合理的没有意义的解有很多，例如 $z_{ij} = 0$、$\mathcal{M}_i = \mathcal{M}_0$ 和 $\boldsymbol{P}_j = \boldsymbol{P}_0$，其中，$\mathcal{M}_0$ 是任意秩为 3 的 3×4 矩阵，$\boldsymbol{P}_0$ 是它的核中的一个单位向量。限制矩阵$\mathcal{D}$的各列 $\boldsymbol{d}_j$ 必须是单位幅值后，就可以去掉那些没有意义的解。

假设已经进行了若干步的最小化迭代，固定$\mathcal{M}$的当前值不变，对 $j = 1, \cdots, n$，计算 $\boldsymbol{z}_j \stackrel{\text{def}}{=} (z_{1j}, \cdots, z_{mj})^T$ 和 $\boldsymbol{P}_j$ 使

$$E_j \stackrel{\text{def}}{=} \sum_{i=1}^{m} |z_{ij}\boldsymbol{p}_j - \mathcal{M}_i\boldsymbol{P}_j|^2$$

最小化，这也是 E 的最小化。写出 E_j 关于向量 $\boldsymbol{P}_j$ 的导数，在最小化时导数应为 0，则有

$$0 = \frac{\partial E_j}{\partial \boldsymbol{P}_j} = 2\sum_{i=1}^{m} \mathcal{M}_i^T(z_{ij}\boldsymbol{p}_{ij} - \mathcal{M}_i\boldsymbol{P}_j)$$

或

$$\mathcal{M}^T\boldsymbol{d}_j = \mathcal{M}^T\mathcal{M}\boldsymbol{P}_j \Longleftrightarrow \boldsymbol{P}_j = \mathcal{M}^+\boldsymbol{d}_j$$

其中，$\mathcal{M}^+ \stackrel{\text{def}}{=} (\mathcal{M}^T\mathcal{M})^{-1}\mathcal{M}^T$ 是$\mathcal{M}$的伪逆。把这个值代入 E_j 的定义得到 $E_j = |(\mathrm{Id} - \mathcal{M}\mathcal{M}^+)\boldsymbol{d}_j|^2$。

现在$\mathcal{M}$是一个 $3m \times 4$ 的秩为 4 的矩阵，它的奇异值分解为$\mathcal{U}\mathcal{W}\mathcal{V}^T$，其中，$\mathcal{U}$是 $3m \times 4$ 的列正交阵，$\mathcal{W}$是 4×4 的非奇异对角阵，$\mathcal{V}^T$ 是一个正交阵。$\mathcal{M}$的伪逆是$\mathcal{M}^+ = \mathcal{V}\mathcal{W}^{-1}\mathcal{U}^T$；代入 E_j 的表达式，由 $|\boldsymbol{d}_j|^2 = 1$，则有

$$E_j = \left|[\mathrm{Id} - \mathcal{U}\mathcal{U}^T]\boldsymbol{d}_j\right|^2 = 1 - |\mathcal{U}\boldsymbol{d}_j|^2$$

这又意味着，E_j 对 $\boldsymbol{z}_j$ 和 $\boldsymbol{P}_j$ 的最小化等价于在$|\boldsymbol{d}_j|^2 = 1$ 的约束下对$|\mathcal{U}\ \boldsymbol{d}_j|^2$ 最大化。最后，由

$$\boldsymbol{d}_j = \mathcal{Q}_j\boldsymbol{z}_j, \quad \text{其中} \quad \mathcal{Q}_j \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{p}_{1j} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{p}_{2j} & \cdots & \boldsymbol{0} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{p}_{mj} \end{pmatrix}$$

说明 E_j 的最小化等价于在$|\mathcal{Q}_j\boldsymbol{z}_j|^2 = 1$ 约束下$|\mathcal{R}_j\boldsymbol{z}_j|^2$ 对 $\boldsymbol{z}_j$ 的最大化，其中$\mathcal{R}_j \stackrel{\text{def}}{=} \mathcal{U}^T\mathcal{Q}_j$。这是一个广义特征值问题，它的解是满足$\mathcal{R}_j^T\mathcal{R}_j\boldsymbol{z}_j = \lambda\mathcal{Q}_j^T\mathcal{Q}_j\boldsymbol{z}_j$ 的最大特征值λ 对应的单位向量 $\boldsymbol{z}_j$。

投影深度保持常数，而$\mathcal{M}$和 $\boldsymbol{P}$ 进行修正的最小化这一步与 Tomasi-Kanade 的从这运动计估射结构的方法是相同的。算法 13.1 是整个计算过程的概要。投影深度的初始值可设为 1，或者像前面做的一样，从外极几何的估计中得到。

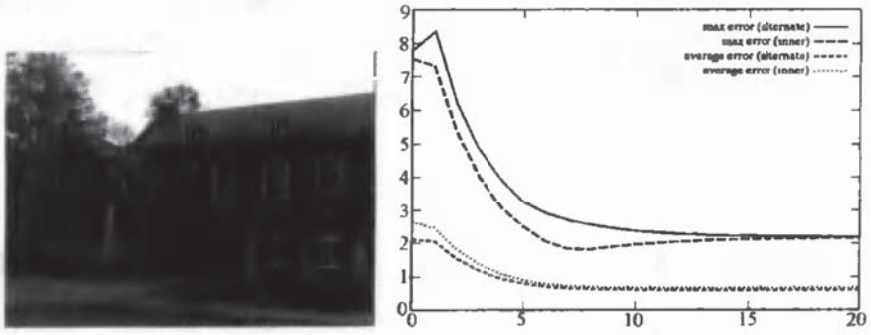
显然误差 E 是收敛到某个值 E^* 的。实际上，设 E_0 是在每次迭代开始时的误差。算法的前两步并不改变向量 z_j，而只是对 $\mathcal{M}$ 和 $\boldsymbol{P}_j$ 做 E 的最小化。若 E_2 是第二步结束时的误差，则有 $E_2 \leqslant E_0$。第三步不改变矩阵 $\mathcal{M}$，对 z_j 和 $\boldsymbol{P}_j$ 做每个误差项 E_j 的最小化。因此，这一步结束时的误差 E_3 一定小于等于 E_2。这说明误差在每一步迭代中都是单调的。由于它的下界是 0，则这个误差收敛于某个 $E^* \geqslant 0$。这种收敛性并不能保证优化算法找到的不是局部极小值。然而，基于数值分析的全局收敛定理（Luenberger，1985），Mahamud 等（2001）给出了算法 13.1 的收敛性证明。这个局部极小是不是全局极小依赖于未知数初值的选取。在 Mahamud 和 Hebert（2000）中选取的初值是 $z_{ij} = 1$，等价于从弱透视投影模型开始。

算法 13.1　分解因子法解从运动到投影形状问题

1. 计算投影深度 z_{ij} 的初始值，$i = 1, \cdots, m; j = 1, \cdots, n$
2. 对矩阵 $\mathcal{D}$ 的各列做归一化
3. 迭代
   (a) 用奇异值分解求 $2m \times 4$ 矩阵 $\mathcal{M}$ 和 $4 \times n$ 矩阵 $\mathcal{P}$，使 $|\mathcal{D} - \mathcal{M}\mathcal{P}|^2$ 最小化；
   (b) 对 j 从 1 到 n，计算矩阵 $\mathcal{R}_j$ 和 $\mathcal{Q}_j$，在 $|\mathcal{Q}_j z_j|^2 = 1$ 限制下，求 z_j 使 $|\mathcal{R}_j z_j|^2$ 最大化，用一般的特征值方法求解；
   (c) 更新 $\mathcal{D}$；

直到收敛。

图 13.5(a)是 20 幅连续图像中的第一帧。图中有手工选取的 30 个特征点，局部误差是正负一个像素。图 13.5(b)是观察图像和预测图像之间的平均误差和最大误差，其中使用了不同数目的图像序列做训练和测试。



(a)　(b)

图 13.5　用迭代法估计摄像机运动和场景结构：(a)图像序列中的一帧；(b)平均和最大重投影误差，关于迭代次数的函数图。共进行了两次实验：第一次（交错），交错取序列中的帧做训练和检测数据集；第二次（内部），前5帧和后5帧做训练集，剩下的做测试集。两种情况下，平均误差在15次迭代后都小于1个像素

13.4.2　组调整

给定矩阵 $\mathcal{M}_i (i = 1, \cdots, m)$ 和向量 $\boldsymbol{P}_j (j = 1, \cdots n)$ 的初值后，可以用非线性最小二乘法使总体误差最小

$$E=\frac{1}{mn}\sum_{i,j}\left[\left(u_{ij}-\frac{\boldsymbol{m}_{i1}\cdot\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\cdot\boldsymbol{P}_j}\right)^2+\left(v_{ij}-\frac{\boldsymbol{m}_{i2}\cdot\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\cdot\boldsymbol{P}_j}\right)^2\right]$$

这就是组调整方法，这个名字起源于摄影测量学。虽然代价可能很高，它却提供了一种方法可以利用所有数据来使物理上的误差最小——或者说，使实际图像点和用摄像机运动和场景结构估计出来的点之间的均方误差最小。

13.5 从投影图像到欧氏图像

虽然投影结构本身也很有用，但是大部分情况下我们关心的是真实场景的欧氏结构。在第 12 章已经说明，在弱透视投影或类透视图像下，即使知道摄像机的内参数，也无法恢复场景的实际比例。在透视投影下也有着同样的问题：若已知摄像机的内参数，可以取标定矩阵为单位阵，式(2.15)的投影方程在某个欧氏世界坐标系(W)中就可以写成

$$\boldsymbol{p}=\frac{1}{z}(\mathcal{R}\quad \boldsymbol{t})\begin{pmatrix}\boldsymbol{P}\\1\end{pmatrix}=\frac{1}{\lambda z}(\mathcal{R}\quad \lambda\boldsymbol{t})\begin{pmatrix}\lambda\boldsymbol{P}\\1\end{pmatrix}$$

其中，λ 是非零的比例系数。这种不确定性并不奇怪，在方程(2.15)中，$\boldsymbol{t}$ 定义为(W)的坐标原点相对于摄像机的位置：若把景物和观察的摄像机都按同样的速度远离这个点，会改变场景的深度，但不会改变得到的图像。增加新的摄像机也不会有帮助，和第 2 章定义的一样，能得到的最好结果是，在任意相似变换层次上的场景欧氏结构。

从现在开始，假设已经用 13.4 节的某个方法，从 m 幅图像的这些点得到了投影矩阵$\mathcal{M}_i$($i=1,\cdots,m$)和点的位置 $\boldsymbol{P}_j$($j=1,\cdots,n$)。我们知道任何其他的重构，特别是欧氏重构，都和投影重构差一个投影变换。换句话，若$\hat{\mathcal{M}}_i$和 $\hat{\boldsymbol{P}}_j$ 表示在欧氏坐标系下测得的形状和运动参数，则一定存在一个 4×4 矩阵$\mathcal{Q}$使得$\hat{\mathcal{M}}_i=\mathcal{M}_i\mathcal{Q}$和 $\hat{\boldsymbol{P}}_j=\mathcal{Q}^{-1}\boldsymbol{P}_j$。本节将介绍一种方法可以在已知摄像机内参数的情况下，计算欧氏升级矩阵$\mathcal{Q}$，从而从投影的形状和运动恢复欧氏形状和运动。

首先要注意，由于每个矩阵$\mathcal{M}_i$ 都是只定义比例层次，对$\hat{\mathcal{M}}_{i_i}$ 也同，则$\hat{\mathcal{M}}_{i_i}$ 可以写成(在一般情况下某些内参数未知)

$$\hat{\mathcal{M}}_i=\rho_i\mathcal{K}_i(\mathcal{R}_i\quad \boldsymbol{t}_i)$$

其中，ρ_i 表示$\mathcal{M}_i$ 未知的比例，$\mathcal{K}_i$ 是由式(2.13)定义的矩阵。特别地，若把欧氏升级矩阵写为 $\mathcal{Q}=(\mathcal{Q}_3\quad \boldsymbol{q}_4)$，其中$\mathcal{Q}_3$ 是一个 4×3 矩阵，$\boldsymbol{q}_4$ 是一个向量，则我们马上可以得到

$$\mathcal{M}_i\mathcal{Q}_3=\rho_i\mathcal{K}_i\mathcal{R}_i \tag{13.8}$$

用这个方程，若摄像机内参数已知，则矩阵$\mathcal{K}_i$ 可以变为单位阵，很简单就可以把第 12 章中使用的仿射方法改为针对投影模型的：按照方程(13.8)，3×3 矩阵$\mathcal{M}_i\mathcal{Q}_3$ 在这种情况下是放缩旋转阵。它们的各列 $\boldsymbol{m}_{ij}^T$($j=1,2,3$)相互垂直且有同样的模，则有

$$\begin{cases}\boldsymbol{m}_{i1}^T\mathcal{Q}_3\mathcal{Q}_3^T\boldsymbol{m}_{i2}=0\\ \boldsymbol{m}_{i2}^T\mathcal{Q}_3\mathcal{Q}_3^T\boldsymbol{m}_{i3}=0\\ \boldsymbol{m}_{i3}^T\mathcal{Q}_3\mathcal{Q}_3^T\boldsymbol{m}_{i1}=0\\ \boldsymbol{m}_{i1}^T\mathcal{Q}_3\mathcal{Q}_3^T\boldsymbol{m}_{i1}-\boldsymbol{m}_{i2}^T\mathcal{Q}_3\mathcal{Q}_3^T\boldsymbol{m}_{i2}=0\\ \boldsymbol{m}_{i2}^T\mathcal{Q}_3\mathcal{Q}_3^T\boldsymbol{m}_{i2}-\boldsymbol{m}_{i3}^T\mathcal{Q}_3\mathcal{Q}_3^T\boldsymbol{m}_{i3}=0\end{cases} \tag{13.9}$$

升级矩阵$\mathcal{Q}$显然只定义到任意的相似变换层次。要惟一确定的话，我们要假设世界坐标系和第一个摄像机的坐标系重合。若给出 m 幅图像，则一共有关于$\mathcal{Q}$系数的 12 个线性方程和 $5(m-1)$个二次方程。可以用非线性最小二乘法求解。

方程(13.9)是对对称矩阵$\mathcal{A} \stackrel{\text{def}}{=} \mathcal{Q}_3 \mathcal{Q}_3^T$ 系数的线性约束，则可以通过线性最小二乘法从至少两幅图像求解。注意$\mathcal{A}$的秩为 3——这是不在我们已有约束内的另外一个约束。为了恢复$\mathcal{Q}_3$ 我们还要注意，由于$\mathcal{A}$是对称的，可以通过正交基对角化为$\mathcal{A} = \mathcal{U}\mathcal{D}\mathcal{U}^T$，其中，$\mathcal{D}$是由$\mathcal{A}$的特征值构成的对角阵，$\mathcal{U}$是它的特征向量构成的正交阵。在无噪情况下，$\mathcal{A}$是半正定的，有三个正特征值和一个零特征值，$\mathcal{Q}_3$ 可以由$\mathcal{U}_3\sqrt{\mathcal{D}_3}$ 得到，其中$\mathcal{U}_3$ 是由$\mathcal{U}$中对应$\mathcal{A}$的正特征值的列构成，$\mathcal{D}_3$ 是对应的$\mathcal{D}$的子阵。但是在有噪情况下$\mathcal{A}$一般是满秩的，它的最小特征值甚至可能是负的。Ponce(2000)指出，若$\mathcal{U}_3$和$\mathcal{D}_3$取为与$\mathcal{A}$的三个最大(正)特征值对应的$\mathcal{D}$和$\mathcal{U}$的子阵，则$\mathcal{U}_3\mathcal{D}_3\mathcal{U}_3^T$ 是在 Frobenius 意义下对$\mathcal{A}$的最好的秩为 3 的半正定近似，且可以取$\mathcal{Q}_3 = \mathcal{U}_3\sqrt{\mathcal{D}_3}$。这时，$\mathcal{Q}$的最后一列 $\boldsymbol{q}_4$ 可以通过取(任取)第一个摄像机的原点作为世界坐标系的原点得到。

这个方法可以很容易地移植到只知道摄像机的部分内参数的情形：由$\mathcal{R}_i$ 是正交阵，有

$$\mathcal{M}_i \mathcal{A} \mathcal{M}_i^T = \rho_i^2 \mathcal{K}_i \mathcal{K}_i^T \tag{13.10}$$

这样，每幅图像都提供了$\mathcal{K}_i$ 和$\mathcal{A}$的约束。例如，若每个摄像机的图像中心都已知，则有 $u_0 = v_0 = 0$，矩阵$\mathcal{K}_i$ 的平方可写成

$$\mathcal{K}_i \mathcal{K}_i^T = \begin{pmatrix} \alpha_i^2 \dfrac{1}{\sin^2\theta_i} & -\alpha_i\beta_i \dfrac{\cos\theta_i}{\sin^2\theta_i} & 0 \\ -\alpha_i\beta_i \dfrac{\cos\theta_i}{\sin^2\theta_i} & \beta_i^2 \dfrac{1}{\sin^2\theta_i} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

特别地，方程(13.10)对应于$\mathcal{K}_i\ \mathcal{K}_i^T$ 中 0 值的部分是对 4×4 对称矩阵$\mathcal{A}$的 10 个系数的两个独立的线性方程。对 $m \geqslant 5$ 幅图像，可以用线性二乘法估计这些参数。若$\mathcal{A}$已知，可以像以前那样估计$\mathcal{Q}$。图 13.6 是用这种方法的变型的实验结果，得到了一个有纹理的城堡的三维模型[Pollefeys 等(1999)]。

图 13.6　合成的有纹理的城堡图像，用投影运动分析，然后再进行欧氏升级。假设主点已知

13.6 注释

本章开头部分对投影几何的简短介绍主要集中在问题的分析上面。可以参考 Todd(1946),Berger(1987)和 Samuel(1988)中对分析投影几何学的透彻介绍,Coxeter(1974)中也有很清楚的表达。从运动恢复投影形状的细节在 Hartley 和 Zisserman(2000)及 Faugeras,Luong 和 Papadopoulo(2001)中有全面的介绍。

Faugeras(1993)提到,从 7 对对应点计算外极点和外极变换的问题是 Chasles(1855)第一次提出,而 Hesse(1863)解决的。对内标定的摄像机,从 5 对对应点估计外极几何的问题是 Kruppa(1913)解决的。Faugeras 和 Maybank(1990)对 Hesse 和 Kruppa 的方法做了很全面的总结,其中引入了绝对圆锥,一种相似性不变的理想圆锥截面,由此推出的两个相切约束可以补偿缺失的对应点。这个方法只在理论上比较有意义,因为它们只依赖于很少的对应点,因此抗噪声性能比较差。第 10 章介绍的 Luong 等(1993,1996)和 Hartley(1995)提出的弱标定方法是一种可靠且准确的替代方法。

Faugeras(1992)和 Hartley 等(1992)分别独自提出了用两个未标定摄像机恢复场景投影结构的方法。其他相关重要工作还有很多,例如 Mohr 等(1992)和 Shashua(1993)。13.2.2 节介绍了 Faugeras 的原创方法,在 13.2.1 节介绍的它的几何形式是 Ponce 等(1993)提出的。在这一章介绍的双目和三目运动分析方法是 Hartley(1992,1994b,1997)和 Beardsley 等(1997)提出的。若摄像机是标定的,按照习题或 Longuet-Higgins(1981)的方法也可以从基本矩阵恢复(至多有反向的不确定性)相似型。Christy 和 Horaud(1996)提出了用迭代方法对标定的摄像机恢复投影运动和结构。用分解的方法恢复结构和运动是 Sturm 和 Triggs(1996)首次提出的。在 13.4.1 节介绍的变形方法是 Mahamud 和 Hebert(2000)提出的,Mahamud 等(2001)证明这个方法是收敛的。使用相邻两帧、三帧、四帧的方法可以在 Beardsley 等(1997)和 Pollefeys 等(1999)中找到。

在已知一些摄像机内参数的条件下计算投影重构的欧氏升级方法,很多作者都对其做了修改(例如 Heyden 和 Åström,1996;Triggs,1997;Pollefeys,1999)。在 13.5 节介绍的矩阵 $\mathcal{A} = \mathcal{Q}_3 \mathcal{Q}_3^T$ 可以在几何上解释为绝对圆锥对偶的投影描述,绝对对偶二次曲面(Triggs,1997)。和绝对圆锥类似,这个二次曲面也是相似不变的,而(对偶)$\mathcal{K}_i \mathcal{K}_i^T$ 对应的圆锥切面就是这个二次曲面到对应的图像上的投影。在不知道欧氏位置的情况下,通过对应点计算摄像机内参数的过程称为自标定。在这个领域,比较领先的是 Faugeras 和 Maybank(1992)对摄像机内参数固定时问题的研究。现在有很多可靠的自标定方法(Hartley,1994a;Fitzgibbon 和 Zisserman,1998;Pollefeys 等,1999),它们也可以用来计算投影重构的欧氏升级。Heyden 和 Åström(1998,1999),Pollefeys 等(1999)和 Ponce(2000)提出的方法可以解决投影重构的几何升级,但需要对摄像机做一些约束,如错切为零。

习题

13.1 在三摄像机情况下,要解决"从运动估计投影模型问题",至少需要多少对对应点?

13.2 说明用方程(13.2)可以表示两个投影坐标系(A)和(B)之间的变换。

13.3 说明在投影平面内的两条不同直线恰交于一点,且在仿射平面内的两条平行直线 Δ 和 Δ' 交于这个平面的投影完善上方向 $\boldsymbol{v}$ 对应的无穷远点。

提示：用 J_A 把仿射平面嵌入它的投影闭集内，把与 $J_A(\Delta)$（对应的 $J_A(\Delta')$ 对应的 $\Pi\times\mathbb{R}$ 中的向量写成向量 $(\overrightarrow{AB},1)$、$(\overrightarrow{AB}+\boldsymbol{v},1)$［对应的 $(\overrightarrow{AB'},1)$ 和 $(\overrightarrow{AB'}+\boldsymbol{v},1)$］的线性组合，其中 B 和 B' 是 Δ 和 Δ' 上任意点。

13.4　说明 $\mathbb{P}^3$ 上的两个平面间的透视投影是投影变换。

13.5　已知仿射空间 X 和仿射坐标系 $(A_0,\cdots,A_n)$，求 $\tilde{X}$ 关于向量 $e_i \overset{\text{def}}{=} (\overrightarrow{A_0A_i},0)\ (i=1,\cdots,n)$ 和 $\boldsymbol{e}_{n+1}=(\boldsymbol{0},1)$ 的投影基。点 $J_{A0}(A_i)$ 是基的一部分吗？

13.6　本题中，要证明共线 4 点 A,B,C,D 的交比为

$$\{A,B;C,D\}=\frac{\sin(\alpha+\beta)\sin(\beta+\gamma)}{\sin(\alpha+\beta+\gamma)\sin\beta}$$

其中，角 α、β 和 γ 在图 13.2 中定义。

(a) 说明三角形 PQR 的面积是

$$A(P,Q,R)=\frac{1}{2}PQ\times RH=\frac{1}{2}PQ\times PR\sin\theta$$

其中，PQ 表示点 P 和点 Q 间的距离，H 是 R 到 P,Q 连线上的投影，θ 是 PQ 和 PR 之间的夹角。

(b) 共线三点 A,B,C 的比例定义为

$$R(A,B,C)=\frac{\overline{AB}}{\overline{BC}}$$

证明

$$R(A,B,C)=A(A,B,O)/A(B,C,O)$$

其中，O 是不在直线上的一点

(c) 证明交比 $\{A,B;C,D\}$ 满足上面的公式。

13.7　证明两束外极线间的同构关系可以写成

$$\tau\to\tau'=\frac{a\tau+b}{c\tau+d}$$

其中，τ 和 τ' 是直线的斜率。

13.8　我们重新考虑三点重构的问题，已知四面体 (A,B,C,O')，如何恢复 D 在这组基下的齐次坐标和单位点 O''。注意参考点的顺序，也就是坐标的顺序，与前面用过的不同：和原来的类似，现在新的选取也是为了方便重构。

我们把点 D 的坐标（未知）记为 (x,y,z,w)，在第一个（第二个）图像平面上设置参考点 a',b',c' (a'',b'',c'') 和单位点 e' (e'')，用 (x',y',z')［(x'',y'',z'')］表示 d' (d'') 的坐标。

提示：参考图 13.3 画一个图会有帮助。

(a) 点 D' 和 D'' 以及直线 $O'D$、$O''D$ 和 $O'O''$ 穿过三角形所在平面的点 E 的齐次坐标。

(b) 求 D 的坐标关于 O' 和 D'（O'' 和 D''）和其他一些未知参数的函数。

提示：点 D,O' 和 D' 是共线的。

(c) 如何计算这些未知参数和 D 的坐标。

13.9　若 $\tilde{\mathcal{M}}=(\mathcal{A}\quad\boldsymbol{b})$ 和 $\tilde{\mathcal{M}}'=(\mathrm{Id}\quad\boldsymbol{0})$ 表示对应的基础矩阵，证明 $\mathcal{F}$ 和 $[\boldsymbol{b}_\times]\mathcal{A}$ 成比例，$\mathcal{F}^T\boldsymbol{b}=0$，且

$$\mathcal{A}=-\lambda[\boldsymbol{b}_\times]\mathcal{F}+(\ \mu\boldsymbol{b}\mid\nu\boldsymbol{b}\mid\tau\boldsymbol{b}\ )$$

13.10 本题中我们将推导基础矩阵的最小参数化方法和对应投影矩阵的计算方法。这与12.2.2节仿射情形的方法本质上是相同的。

(a) 说明两个投影矩阵$\mathcal{M}$和$\mathcal{M}'$总能通过合适的投影变换简化为归一化形式

$$\tilde{\mathcal{M}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad , \quad \tilde{\mathcal{M}}' = \begin{pmatrix} \boldsymbol{a}_1^T & b_1 \\ \boldsymbol{a}_2^T & b_2 \\ \boldsymbol{0}^T & 1 \end{pmatrix}$$

注意:为简单起见,我们假设左右矩阵都是非奇异的。

(b) 注意在投影矩阵上应用这个变换等价于在场景点 $\boldsymbol{P}$ 上应用反变换。我们用 $\tilde{\boldsymbol{P}} = (x, y, z)^T$ 表示点$\tilde{P}$在世界坐标系中的坐标,$\boldsymbol{p} = (u, v, 1)^T$ 和$\boldsymbol{p}' = (u', v', 1)^T$ 表示它在图像里的齐次坐标,证明

$$(u' - b_1)(\boldsymbol{a}_2 \cdot \boldsymbol{p}) = (v' - b_2)(\boldsymbol{a}_1 \cdot \boldsymbol{p})$$

(c) 从这个方程推出基础矩阵的8参数参数化表示,并利用$\mathcal{F}$只定义比例层次来构造最小的7参数参数化。

(d) 用这个参数化来设计方法从至少7个对应点估计$\mathcal{F}$和从场景估计投影形状。

13.11 这里我们对从基本矩阵$\mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}$恢复旋转$\mathcal{R}$和平移 $\boldsymbol{t}$ 的问题稍加修改(本题出自Andrew Zisserman)。平移很容易确定,因为 $\boldsymbol{t}$ 可以通过单位向量$\mathcal{E}^T\boldsymbol{t}$ 恢复(我们知道场景的结构只能恢复到比例层次)。

(a) 说明基础矩阵的分解可以写成

$$\mathcal{E} = \mathcal{U}\,\mathrm{diag}(1, 1, 0)\mathcal{V}^\top$$

且 $\boldsymbol{t}$ 是$\mathcal{U}$的第三列。

(b) 证明下面两个矩阵

$$\mathcal{R}_1 = \mathcal{U}\mathcal{W}\mathcal{V}^\top \quad 和 \quad \mathcal{R}_2 = \mathcal{U}\mathcal{W}^\top\mathcal{V}^\top$$

满足$\mathcal{E} = [\boldsymbol{t}]_\times\mathcal{R}$,其中

$$\mathcal{W} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

编程作业

13.12 实现13.2.1节介绍的场景投影估计的几何方法。

13.13 实现13.2.2节介绍的场景估计的代数方法。

13.14 实现13.4.1节介绍的场景估计的分解方法。

第四部分　中 层 视 觉

- 第 14 章　基于聚类的分割方法
- 第 15 章　基于模型拟合的分割
- 第 16 章　使用随机方法的分割与拟合
- 第 17 章　基于线性动态模型的跟踪

第 14 章　基于聚类的分割方法

一个引起人们广泛关注的观点认为，视觉问题是一个推理问题。所谓推理问题，也就是有一些测量结果和一个模型，我们希望知道是哪些因素导致了这些测量结果。但是仍然存在一些关键的特征使视觉问题和其他的推理问题不同：首先，前者数据量非常大，其次，不知道这些数据项中哪些部分有助于解决这个推理问题，哪些不会。例如，编写一个好的物体识别程序，其中一个较大的难点就是知道哪些像素点要识别，哪些可以忽略。如果只是简单地看像素，很难区分图 14.1 中的一个像素点是否位于表面上。这个问题可以通过采用体现图像数据重要性质的精简表示来完成。获得这样的表示的方法称为分割、分组、感知组织或者拟合。我们用分割这个术语泛指这些行为，因为尽管技术上可能不同，但是所有这些行为的动机是一样的：得到一幅图片中有用部分的一个精简的表示。要想看到一个全面的分割理论是很难的，至少有一个原因是因为哪些是感兴趣的部分取决于应用的需求。到目前为止还没有全面的分割理论，在不同的场合使用不同的术语。在本章，描述的分割方法还没有概率的表示。在下面的章节，将处理更复杂的随机概率算法。

图 14.1　如这些图像所示，视觉的一个重要部分涉及将图像信息组织成有意义的集合。人类的视觉系统做这件事情看起来非常好。上面的每一幅图像中的小圆点都被组织在一起形成纹理的表面，感觉似乎要突出纸面（你一定感觉到它们是半球状的）。这些小圆点被集合在一起，“是因为它们形成了一个表面”，根本不是一个令人满意的解释，这样就引发了计算的难题。请注意，说它们是因为形成了相同的纹理而被聚集在一起同样引发问题（我们如何知道的）。对于左图的情况，编写程序来识别单个具有一致属性的纹理将非常困难。这种形成组织的方法能被应用于很多不同的输入

14.1　什么是分割

假设要识别一幅图像中的物体。如果逐个处理每一像素，那么需要处理的像素就太多了，因此，我们希望有一些简明扼要的表示形式。至于具体表示的细节应该由具体的任务来决定，不过这些表示仍然应该有一些共同的特点。首先，由典型图片所得到的表示中的组成应该相对较少（不多于后来的算法能处理的），其次，这些组成都应该是富有启发性的。对于典型的图像，通过这些组成，明显地启示我们要寻找的物体是否出现在其中。

对于不同的数据集有不同的处理方法：一些方法适用于图像，一些方法适用于视频流，还有一些方法适用于样本——表明一幅图像中出现感兴趣模式的标识符，例如一个斑点、圆点或者边界点。事实上，视频中也一样存在样本，例如根据某个参数规则运动的点。

虽然表面上看起来这些方法很不一样，但是它们有很大的相似之处（这也是它们一起出现的理由）：每种方法都试图用某些形式的相似模型去获得数据集的精简表示（有些情况需要努力去辨认出模型）。这些共同的特征贯穿在不同的问题中。我们重温一些例子：

- **视频摘要**：用户可能希望浏览大量收藏的视频流，因此需要提供一个视频流内容的摘要。解决这件事情的一个办法就是将视频流中看起来相似的视频段划分成镜头，然后使用剪辑画面表示，每个镜头用一代表帧表示。这表明要将视频流分割成镜头。
- **检测机械零件**：假设我们希望在图像中检测一个机械零件（这种情况发生的可能比原设想的少得多）。机械零件往往包含直线（平面交界处）和圆（钻孔的地方）。这表明要分割图像成直线和圆的集合；一般来说，首先必须找到边，然后用直线和圆去匹配。
- **检测人**：假设希望在图像中检测人。这个问题至今仍有待研究，但是解决办法的一般性轮廓还是清楚的。首先应该寻找人体各部分，然后将他们组装起来。这些部分在图像中看起来像延伸的区域；如果人穿的衣服没有纹理，这样他们就是一个单色的延伸区域——我们必须寻找一个恒定色调的条形物。
- **卫星图像中检测建筑物**：建筑物绝大多数都是多面体，尤其是当它们出现在卫星图片上时。这说明可以通过背景上的一些多边形区域来表示这样的图像。一般来说，通过检测边界点组成直线，然后组成多边形。
- **图像搜索**：对于用户进行图像搜索，搜索图像必须通过一种既对用户有意义又与图片内容相关的方式来表示。因为内容一般包括出现的物体，而这些物体往往有相同的色彩和纹理，所以很自然地试图将图像分割成色彩和纹理具有一致属性的不同的区域，并且用这些区域作为图像的一个表示。

14.1.1　模型问题

分割是一个很大的话题。有各种各样的关于分割问题的例子，其中一些看起来并不是很明显的分割问题。建立自然的模型的问题就是分割问题，用多种方法解决它们将是非常有价值的，这在实际应用中经常出现。一般情况下，如果可能，我们的讨论通常涉及下面问题中的某一个：

- **形成图像的划分**：我们希望将一幅图像分解成超像素（superpixel）——就是色彩和纹理大致相同的图像区域。一般来说，这些区域的形状并不是很重要，而属性一致性却很重要。这个过程被非常广泛地研究——经常被认为是术语“分割”的惟一意思——通常认为它是识别领域中的奠基石。
- **边界点拟合直线**：如前面看到的，有很多理由说明用直线簇拟合点的集合是非常有用的。这个问题可以从非常简单（这种情况下，我们知道有多少条直线，以及每个点属于哪条直线）直到相当复杂（在大多数其他情况下）。因为将一些能拟合在一条直线上的样本组织在一起，所以这也是一个分割问题。如果将一系列的点拟合成一条直线，要非常细心，如果其中一些点并不靠近任何直线，拟合的结果将毫无意义。一般来说，需要

同时估计直线的参数和直线与点之间的对应关系。这阐明了一个重要的、非常普通的原则:不搞清对应关系就会产生噪声一样的效应。

- **特征点拟合基础矩阵:** 假设有一个特征点集的两个视图。即使得到一些提示,但是确定两个视图之间哪些点与哪个点对应仍然十分困难。其中一个重要的提示就是,如果对应正确了,将会有一个基础矩阵来将这些点连接在一起,而我们希望在不知道正确对应之前先确定这个基础矩阵。这个问题值得解决的理由有下面几条:首先,不解决这个问题,就不能通过多视角构造符合实际的形状表示。其次,这个问题的解决可以作为一个提示来判断是否一个点集正在做刚体运动。如果一个视频流中显示两个运动的物体,那么它们将有不同的基础矩阵。同样地,对应的误差在这个问题中产生噪声的效应。

用这些问题来描述各种各样的分割算法,不过你应该明白——没有一种方法能对各种模型问题都适用。

14.1.2 基于聚类的分割

分割的一个通常观点就是我们正试图知道数据集中的哪一部分应该自然地归为哪一类。这个问题就是所说的聚类,它有许多参考文献。一般情况下,可以通过两种途径来进行聚类。

但是,更多的情况下,它很复杂。一般来说,应该知道多种解决方法然后做一个明智的选择。即使这样,还要知道一个分割方法应该根据什么样的标准来决定像素点(或者样本)的分类。人类的视觉系统是一个值得借鉴的丰富资源,它需要以最一般的形式解决:

- **分级聚类:** 这里有一个较大的数据集,根据样本在数据集中的某种关系将它们划分。我们希望根据模型将它分解成小段。例如,可能
  - 将一幅图像分解成多个小区域,在各区域内色彩和纹理具有一致属性;
  - 将一幅图像分解成一些广义块,由一些色彩、纹理以及运动都具有一致属性的区域组成,看起来像肢体段;
  - 将一个视频流分解成为镜头——从相同视点表述相同内容的小段视频。
- **聚类:** 有一组不同数据组成的数据集,希望根据模型将数据项合乎情理地集合在一起。譬如,遮挡产生的效果告诉我们,同一物体在图像中是经常需要分成各部分的。聚类的例子如下:
  - 将聚在一起形成直线的样本点收集在一起;
  - 将属于同一个基础矩阵的样本点收集在一起。

当然,这里的关键问题是确定什么样的表示对于眼下的问题最适合。这个问题偶尔会很简单;这个问题,并且它对样本如何分类表现出很强并且合乎情理的倾向性。

14.2 人类视觉:分类和格式塔原理

人类视觉系统的一个主要的特征就是周围环境影响着事物的感知(例如,见图 14.2 所示)。这些观察的结果导致心理学格式塔(Gestalt)学派反对对激励反应的研究以及强调分类

作为理解视觉感知的关键。对他们而言，分类就是视觉系统将一幅图片的某些部分组合在一起并且将它们作为整体感知(这里给了上面提到的“周围环境”一个相当不精确的意思)的倾向。例如，分类导致图 14.2 所示的 Müller-Lyer 错觉——视觉系统把两个箭头各部分组成在一起，因此两条水平线彼此看起来不相同，因为它们是作为整体感知的，而不是单独作为直线感知的。此外，许多分类的效果不受主观意愿影响：例如，你不能因为希望图 14.2 中的直线看起来长度相等，而不把它们组合成箭头。

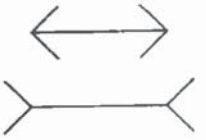


图 14.2　著名的 Müller-Lyer 错觉：两条水平线段实际上是一样长的，然而上面的图看起来短一些。显然，这样的效果是由将它们组成在一起的关系的一些特性(完形性质)引起的，而不是单个部分的特性决定的

分割的普遍经验认为，一幅图像能分解为图形——一般是，有意义的、重要的物体——和背景——图形所在的背景。然而，如图 14.3 所示，什么是图形、什么是背景可能是非常模棱两可的，也就是说这个观点还需要进一步的理论上的研究。

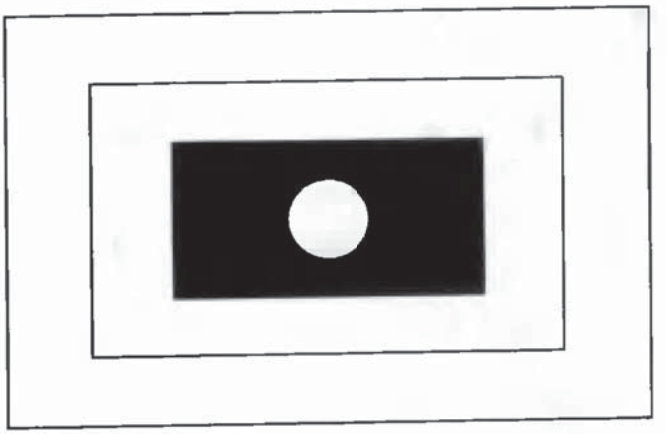


图 14.3　分割的一种观点认为分割决定图像的哪些部分是图形、哪些部分是背景。这幅图形说明这种观点会产生模棱两可的情形；白色的圆可以看做黑色矩形背景上的图形，也可以是背景的一部分，而带有一个圆孔的黑色矩形是图形的背景的一部分——这个时候白色的方框就是背景

Gestalt 学派用完形(gestalt) 这个概念(一个整体或者一组)和它的完形性质(使得各部分成为一个整体的内在关系集合)，作为他们学术思想的核心。他们工作的主要特点就是，制定一系列的规则，用来将图像元素分类和分组。他们也开发算法，不过这些算法都已经成为历史(见 Gordon，1997 年一个介绍性的说明，其中对他们工作的讨论基于更广泛的背景)。

Gestalt 心理学者确定了一系列具有规律的性质，认为这些性质事先安排好了元素集合的分组。很明显，人类视觉系统在某些方面使用了这些性质，所以它们很重要。此外，有理由相信它们确实揭示了样本什么时候归为一类的倾向，可以作为有用的中间表示。

有各种各样的规律性性质，其中一些性质是对 Gestalt 的主流的补充：

- **相似性**：相似的样本往往被归为一组。
- **相同趋向性**：具有运动一致性的样本往往被归为一组。

- **同一区域性**：位于同一封闭区域内的样本往往被归为一组。
- **平行性**：平行的曲线或样本往往被归为一组。
- **封闭性**：能形成封闭曲线的曲线段或者样本往往被归为一组。
- **对称性**：能形成对称组合的曲线段可以分成一组。
- **连续性**：能形成连续——可以形成自然的连接——曲线段可以分为一组。
- **熟悉的形状**：组合在一起能形成熟悉的形状的样本往往被归为一组。

这些性质如图 14.4，图 14.5，图 14.7 和图 14.1 所示。

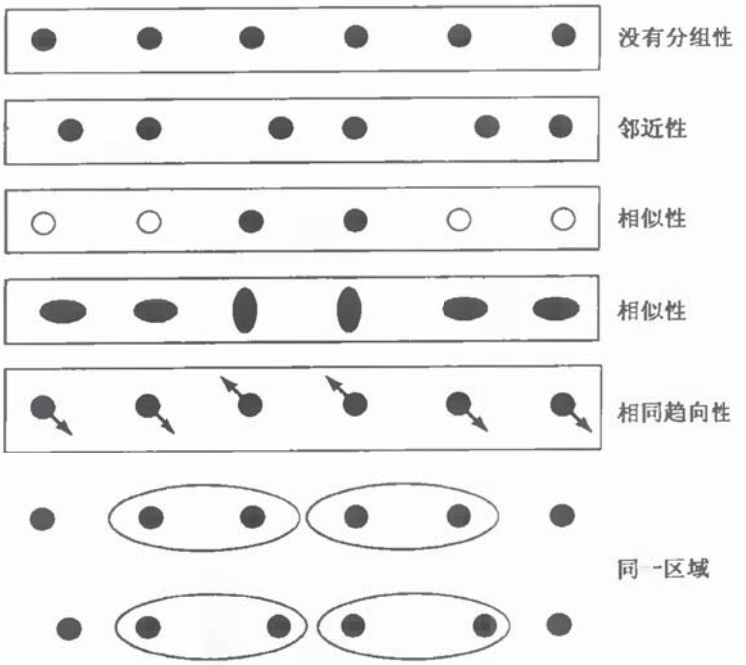


图 14.4 Gestalt 性质指导分类的例子(文中有更详细的描述)

这些规则能够很好地用来解释一些现象，但是对构造相应的算法，它们还不够充分。Gestalt心理学者对于细节问题还有很多困难，比如什么时候使用哪一条规则。为使用这些规则提供一个令人满意的算法非常困难——Gestalt 组织试图使用一个极端的原则。熟悉的形状是一个特殊的问题。关键是弄清楚在一个问题中使用什么样的形状以及怎样选择，如图 14.1。有人可能说小圆块组成一类因为它们形成了一个球。这种说法的问题在于解释这是怎样发生的——形成一个球的假设来源于哪里。通过所有物体的所有视角的搜索是一种解释，但是接下来必须弄清楚搜索怎么进行。是否要检查每一个形状的每一个球的每一个视角呢？怎样才能高效完成这件事情呢？

Gestalt 规则确实提供了一些见解，因为它们解释了在许多例子中所发生的事情。这些解释看起来很不错，因为它们说明这些规则有助于解决真实世界中普遍出现的视觉效果提出的问题——也就是，它们符合生态学现象。例如，连续性可能指出了对于遮挡问题的一种解决方法，被遮挡物体的轮廓段可以借连续性质连接起来(参见图 14.6)。

一个有公德的人将数字和按钮中间的空白填充了，如下面的图示，再也没有出现过搞混的事情了，因为有邻近性的提示，歧义性被消除了。

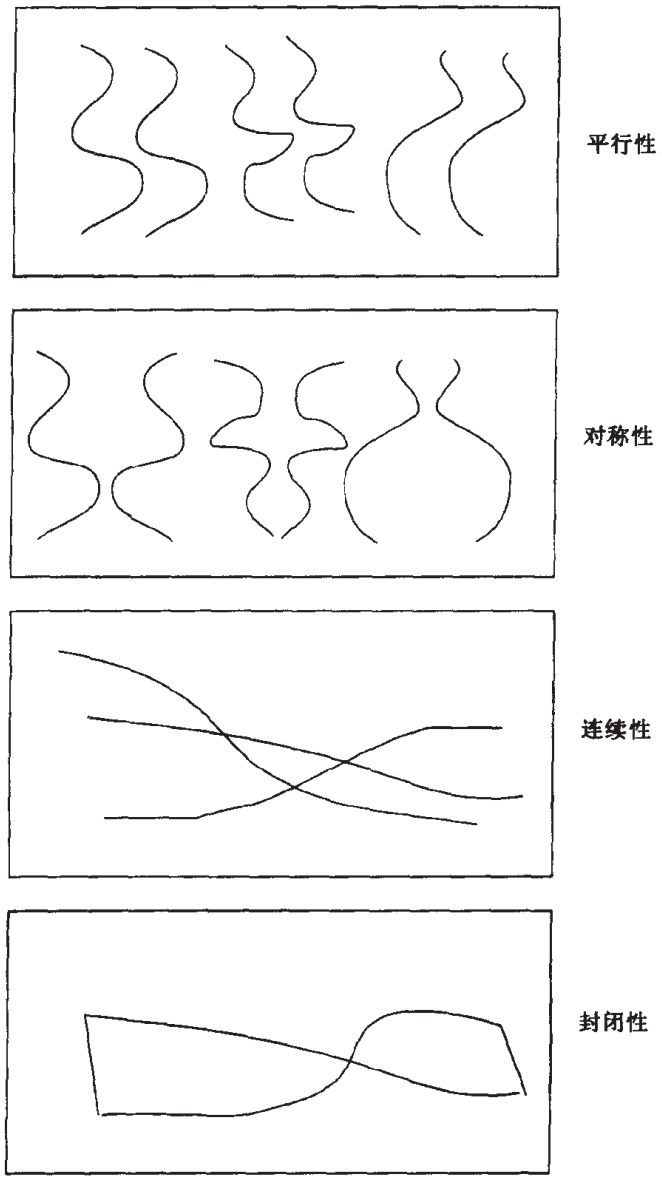


图 14.5　Gestalt 性质指导分类的例子(文中有更详细的描述)

对于偏向于用遮挡来解释一些现象的倾向有一些有趣的结果。一个就是“幻觉轮廓”(illusory contour),如图 14.8 所示。这里一些样本表明了有一个轮廓线与背景颜色没有任何反差的物体的存在。这些样本之所以被归为一类,是因为它们一起显示了一个遮挡物体的存在,这些样本如此清楚地显示了它的存在,以至于人们可以填充这个并没有明显轮廓边界的区域。

这种生态观点比较有力,因为使用它可以解释大多数归类的因素。相同趋向性可以被看做物体的各部分总是一起运动的结果。相同的,平行性是一个很有用的归类提示,因为现实中的很多物体都具有平行或者近似平行的轮廓。实质上,生态学的观点认为样本归类是因为归类的表示对于人们遇到的视觉世界有帮助。生态学的观点有一个诱人的但是模糊的统计意义。通过我们的观察,Gestalt 因素给出了让人感兴趣的提示,但是应该认为它只是一个大的归类过程的结果,而不是过程本身。

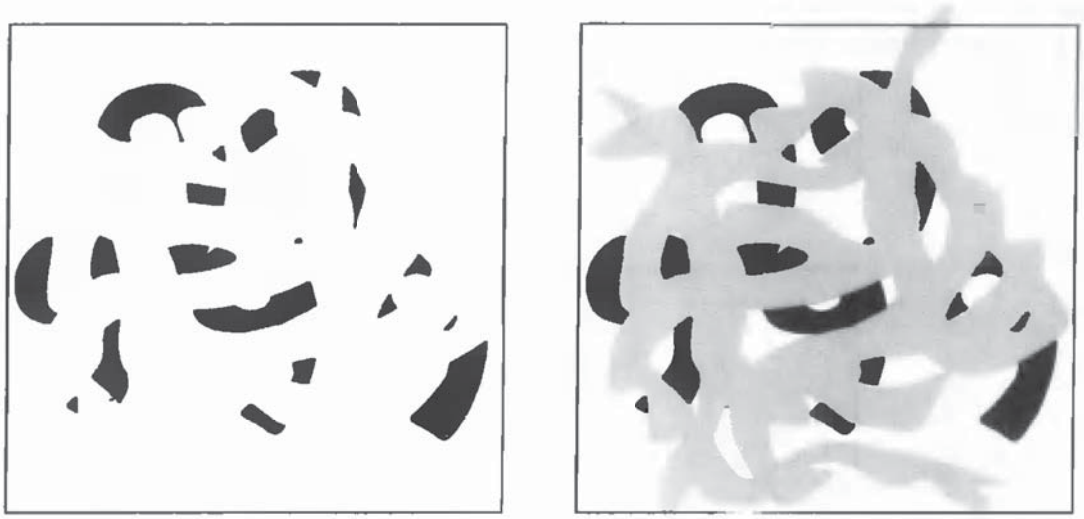


图 14.6 在归类中，遮挡似乎是一个重要的提示。有可能将左图看做几个数字；而右图可以清楚地看见一些被遮挡的数字。左图黑色的区域和右图是一样的，但是视觉系统感到有证据表明这些样本是由于某种原因分隔开，而不是零散分布的样本

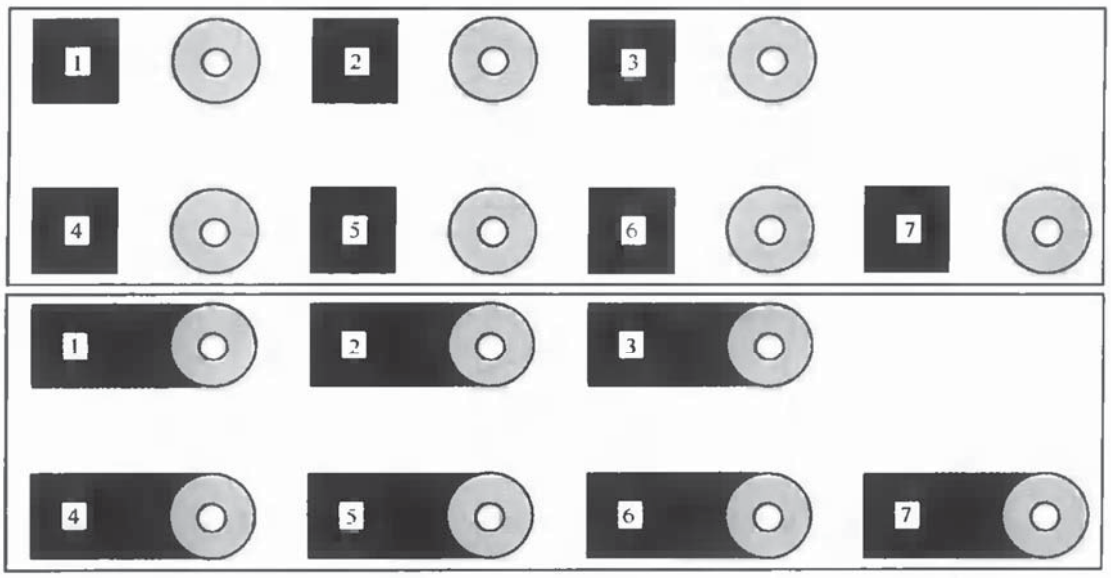


图 14.7 真实生活中归类现象的例子。加州大学伯克利分校的计算机科学大楼电梯里的按钮，过去的布局如上面的图所示。这样经常导致上错楼层，而且发现是因为按错了按钮——按钮很难准确无误地和正确的标签对应起来，短暂的一瞥很容易将它们对应错

图 14.8 这些图像中的样本表明了有边界与图像背景颜色相近的遮挡物存在。注意，其中一个对于遮挡图形的整个轮廓位置有一个清晰的印象。这些轮廓被称为“幻觉轮廓”

14.3　应用:镜头的边界检测和背景差分

简单的分割算法经常在重要的应用中很有用。一般情况下,如果很容易区分出一个有用的分解时,简单的算法能很好地发挥作用。两种重要的情形就是背景差分(任何看起来不像是一个已知背景的部分都是感兴趣的部分)和镜头的边界检测(感兴趣的是在视频中变化比较大的部分)。

14.3.1　背景差分

在很多应用中,物体总是出现在一个很稳定的背景中。标准的例子就是检测传送带上的物体。另一个例子是,在一条道路的高空视角数过往车辆——看到的道路非常稳定。另外,不是很明显的一个例子:人机交互。很常见地,一个摄像头被固定(譬如说,监视器上面)来观察一个房间。视图中看起来不像是这个房间里的物体往往是感兴趣的部分。

在这些应用中,通常可以通过从图像中减去背景图像的估计值,然后从结果中寻找绝对值比较大的部分来获得有用的分割。主要的问题在于获得一个背景图像的好的估计值。一种简单的办法就是直接取一张背景图片。这种简易的办法太不精确,因为一般情况下背景随着时间的推移是在慢慢改变的。例如,当下雨的时候,路面可能更加光亮一些,而当气候干燥的时候,可能就不是那么光亮了;人们也可能移动房间里面的书和家具,等等。

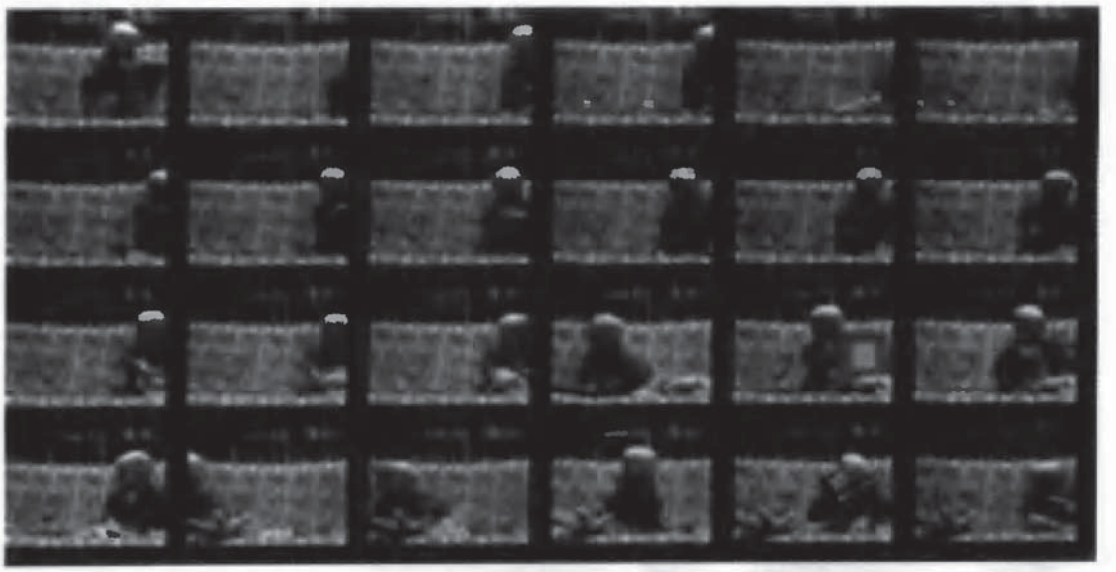


图 14.9　一个小孩在有图案的沙发上玩耍的视频,共 120 帧,上图显示的为其中每 5 帧一截取。视频帧大小为80×60,原因将在图14.11中讨论。注意在视频流中小孩从图像的一端运动到另一端

一种效果不错的办法是使用运动平均方法估计背景像素点的值。在这个方法中,计算每一个背景像素点先前值的加权平均作为它当前的估计值。一般来说,远离当前帧的像素值的权重应该是 0,越接近当前帧的权重越大。理想情况下,运动平均应该跟随背景图像变化,也就是说如果天气变化很快(或者书本被经常不确定地移动)那么只有很少帧的像素点是非零的权重,如果变化比较慢,像素点中权重非零的先前帧数目就会增加一些。这样就产生了算法 14.1。对于已经阅读过滤波器一章的读者来说,这是一个时域平滑的滤波器,我们希望抑制住

高于一般情况下背景变化频率的频率，而使位于或者低于这个频率的通过。这个方法很成功，但是需要应用于粗尺度的图像中，如图 14.10 和图 14.11 说明的。

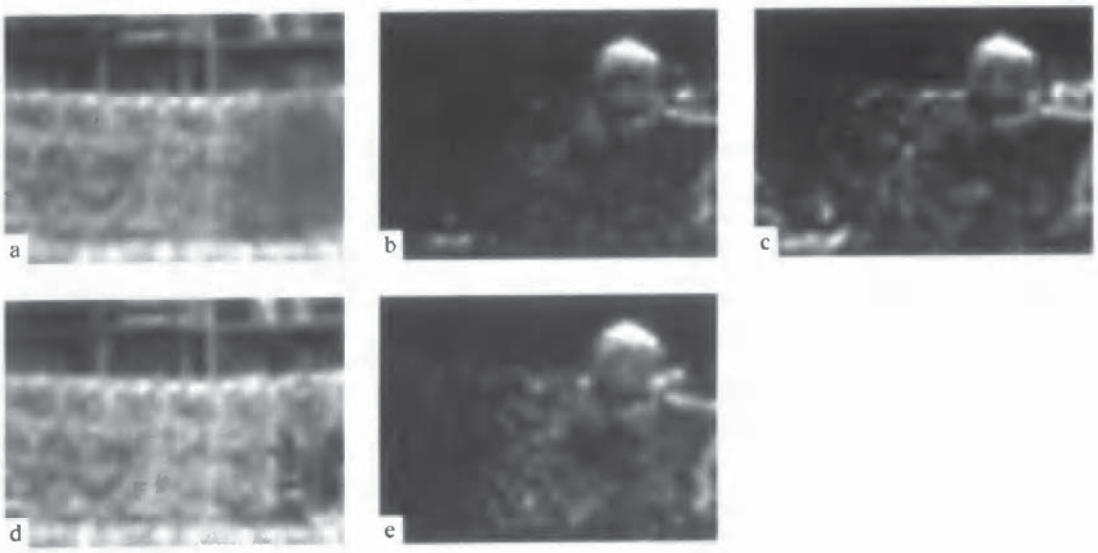


图 14.10 对图 14.9 中 80×60 大小帧的视频流采用背景差分的结果。我们比较一下两种计算背景图像的方法：(a)全部120帧的平均值——注意到小孩在沙发的一端玩的时间较在另一端玩的时间要长一些，这样导致平均帧中那个地方有点模糊不清；(b)与平均帧的差异超出阈值的像素点；(c)与平均帧的差异超出一个小一些的阈值的像素点。注意到，每种情况总有一些额外的像素点和一些丢失的像素点；(d)用更复杂一些的办法计算出来的背景图片(在16.2.5节中有简单的描述)；(e)这种方法计算出来的不同于背景的像素点，再次注意丢失的像素点

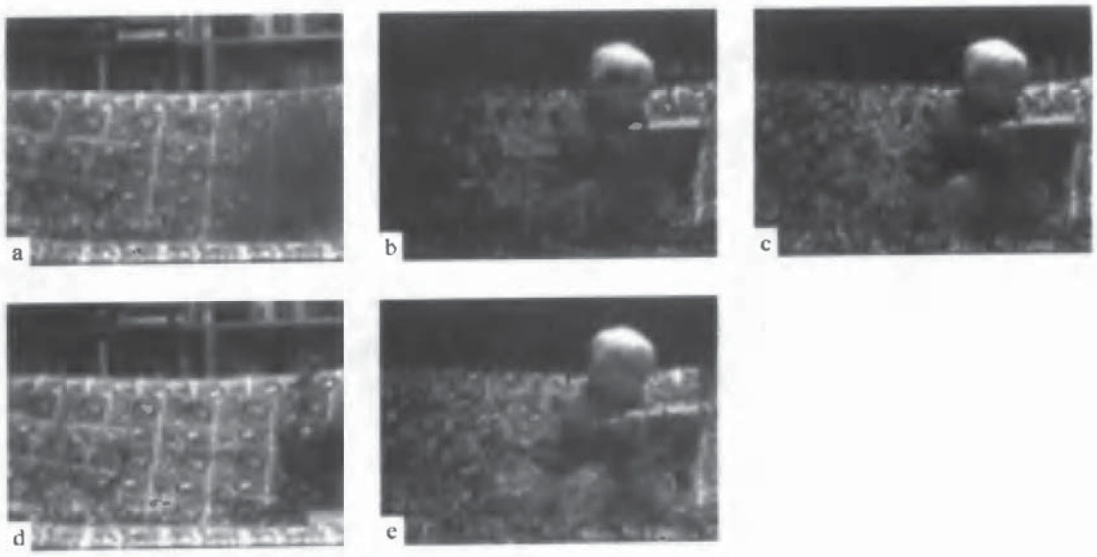


图 14.11 对准匹配在背景差分中是一件很讨厌的事情，尤其是对于纹理来说。上面这些图片是对图 14.9 用 160×120 一帧计算的结果。我们比较两种计算背景图片的方法：(a)全部120帧的平均值——注意到小孩在沙发的一端玩的时间较在另一端玩的时间要长一些，这样导致平均帧中那个地方有点模糊不清；(b)与平均帧的差异超出阈值的像素点；(c)与平均帧的差异超出一个小一些的阈值的像素点；(d)用更复杂一些的办法计算出来的背景图片(在16.2.5节中有简单的描述)；(e)这种方法计算出来不同于背景的像素点。注意错误的像素点——沙发上的花纹被误认为是小孩——的数目明显增加。这是因为轻微的运动使得沙发上空域高频的花纹配不准，导致了比较大的差距

算法 14.1　背景差分

形成一个背景估计值 $\mathcal{B}^{(0)}$。对每一帧$\mathcal{F}$

更新背景图片估计值，一般通过下面公式$\mathcal{B}^{(n+1)} = \dfrac{w_a \mathcal{F} + \sum_i w_i \mathcal{B}^{(n-i)}}{w_c}$其中，选择好权重值 w_a, w_i, w_c。

从帧中减去背景图片估计值，重新记录差值大于选定阈值的每一像素点的值。

end

14.3.2　镜头的边界检测

较长的视频流是由一系列镜头组成的——镜头指的是基本上显示的是同一物体的较短视频流。一般来说，这些镜头是编辑处理过程的产物。很少有两镜头在何处衔接的记录。用一些镜头来表示一段视频是很有用的，而每一个镜头又可以用关键帧表示。这种表示可以用于视频的检索或者概括视频内容以便用户进行浏览。

自动地寻找这些镜头的边界——镜头的边界检测——是简单分割算法的一个重要而可行的应用。镜头边界检测算法必须在视频中找出那些和上一帧相差很大的帧。检测镜头边界必须考虑到，在给定的镜头内部，物体和背景都可能在视野中移动。一般来说，这种检测采用某种形式的距离度量；如果距离大于一个给定阈值，则一个镜头边界被检测到(算法 14.2)。

算法 14.2　采用帧间差异的镜头边界检测

对于图像流中的每一帧

　计算这一帧和上一帧之间的距离

　如果距离大于某个阈值，

　　将这一帧作为一个镜头边界

end

计算距离有各种标准技巧：

- **帧差分算法**：计算视频中两帧对应点之间的差，然后求差的平方和。这些算法并不流行，因为它们太慢了——大量的差运算——并且当摄像头抖动的时候会检测到太多的镜头。
- **基于直方图算法**：计算每一帧的色彩直方图，并且计算直方图之间的差。色彩直方图之间的差是一个很好用的测量，因为它不太容易受颜色在帧空间排布的影响(例如，小的摄像头抖动将不会影响直方图)。
- **块比较算法**：将帧切分成许多小的网格，通过比较这些小的格子来比较两帧。这样避免了色彩直方图的不足，如当一个红色物体从屏幕左下角消失等价于一个红色的物体从上边界进入屏幕。一般来说，这种分块比较算法用块间距离的合成计算帧间距离——取最大值是很自然的选择，块间距离的计算采用和帧间距离一样的计算方法。
- **边缘差分算法**：计算每一帧的边缘图，然后比较这些边缘图。一般地，通过计算帧间潜在对应的边缘(附近，相似的方向，等等)的数目来比较。如果几乎没有相一致的边缘，说明这是一个镜头边界帧。可以通过转化对应的边缘的数目来计算距离。

这些方法有点就事论事，但是经常能够解决当前的问题。

14.4 基于像素点聚类的图像分割

聚类是一种方法，它将一个数据集转化成一堆聚类，聚类包含属于同一类的数据点。我们很自然地想到图像分割也是一种聚类：我们要将属于一起的像素点聚类来表示一幅图像。具体适用的标准依赖于具体的应用，归于一类的像素点可能是因为它们有相同的颜色，也可能因为它们有相同的纹理，或者它们相邻，等等。

14.4.1 基于简单聚类方法的分割

采用一种聚类方法，并由此构造出一幅图像的分割是比较容易的。本质上来说，大部分关于图像分割的作品都是由聚类的论文组成的。

简单的聚类方法 这里有两种很普通的聚类算法。在分解式聚类中，整个数据集被作为一个集合，然后集合通过递归的方法逐步分裂成适当的聚类(算法 14.4)。在凝聚式聚类中，每一个数据项都看做一个独立的类，然后这些聚类通过递归的方法逐步合并成适当的聚类(算法 14.3)。

算法 14.3 凝聚式聚类或者合并聚类
定义每个点为独立的一个类
直到聚类达到所要求的
  将类间距离最小的两类合并
end

算法 14.4 分解式聚类或者分裂聚类
定义一个包含所有点的类
直到聚类达到所要求的
  将一个类分裂成两个类，条件是所产生的两个类的类间距离最大
end

在考虑聚类的时候有两个重要的问题：

- 怎样才是一个好的类间的距离？凝聚式聚类使用类间距离来融合邻近的类，分解式聚类使用它将不够凝聚的类切分开。尽管数据点之间的一般距离是可以得到的(对于视觉问题可能不是这种情况)，但是并没有规范的类间距离。一般来说，人们选择一个对于数据集来说比较适当的类间距离计算方法。例如，可以选择两类之间最近的两个元素之间的距离作为类间距离——这样趋向于产生细长型的类(统计学家称这种方法为单连接聚类)。另一个常用的选择就是第一个聚类中的元素与第二个聚类中元素的最大距离——这样趋向于产生团状型聚类(统计学家称之为全连接聚类)。最后，可以使用聚类中元素间距离的平均值——这往往也趋向于产生“圆形的”聚类(统计学家称之为基于集团均值的聚类)。

- 应该划分多少类？如果对于产生聚类的过程没有模型，应该划分多少类是一个根本性的难点。我们已经描述过的算法生成一个聚类的层次。通常，通过树状图——一种显示类间距离的层次结构表示——给用户显示这个层次关系，用户根据这个树状图做出一个适当的聚类选择（见图 14.12 的例子）。

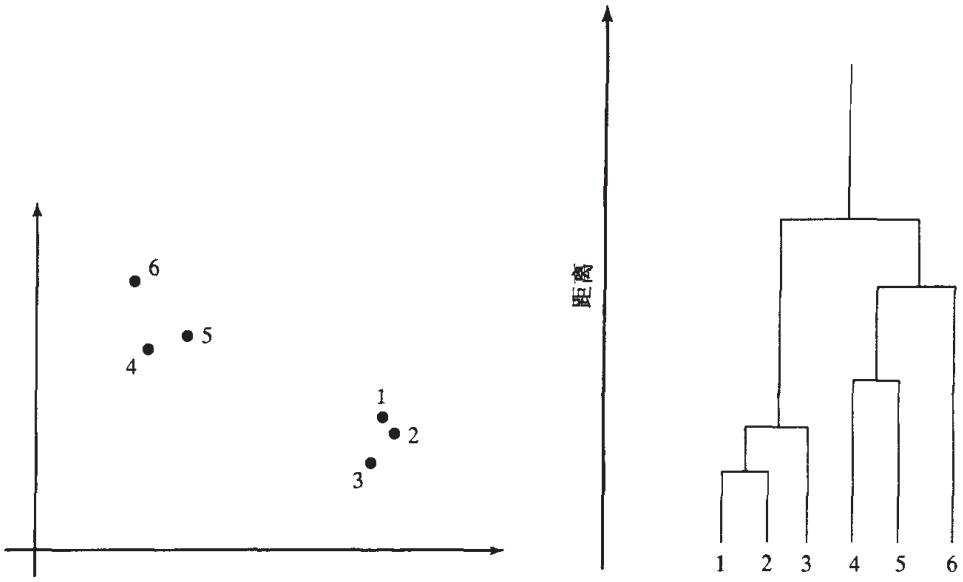


图 14.12　左边，一堆数据；右边，采用单连接凝聚式聚类得到的树状图。如果选择一个距离的特殊值，这样，对应该值的一条水平直线将树状图分割成为聚类。这种表示使我们可以看出有多少聚类并且可以看出聚类的效果如何

通过聚类方法进行分割　虽然使用的类间距离完全依赖于应用的问题，但是色彩差别和纹理差别的测量通常用做聚类距离。一般情况下，需要聚类成块状子类；这可以通过使用位置上的距离作为聚类用距离。

直接使用凝聚式聚类或者分解式聚类的主要难点在于一幅图像中含有大量的像素点。大量的数据意味着树状图将会非常庞大，这样就没有可能检查树状图了。实际上，这也意味着分割时应该通过使用某个阈值来决定什么时候停止分裂或者合并——例如，一个凝聚式聚类，如果类间距离足够的小或者聚类类数达到某个特定值，那么将停止合并。

因为像素点数很多，所以要求最好的聚类分裂（对于分解式方法）或者最好的聚类合并（对于凝聚式方法）都是不太实用的。分裂式方法经常修改成使用类的某种概要形式来确定好的分裂。一个常用的概要形式就是像素点的色彩（或灰度等级）直方图。

凝聚式方法也需要修改。首先，像素点的数目要求我们注意类间距离如何选择（重心之间的距离经常被使用）。其次，经常只合并有共同边界的类（我们可能不希望仅用三类——一类红、一类白、一类蓝来代表美国国旗）。最后，通过扫描图片直接合并距离在某个阈值以下的区域，而不是一定要去寻找最近的合并对，这样简单的区域合并非常有用。

14.4.2　使用 K-均值算法的聚类和分割

简单的聚类方法通过和已经存在的聚类进行贪心式交互达到一个好的全局性表示。例如，在凝聚式聚类中，反复地进行可能是最好的合并。然而，这些方法与要优化的目标函数之间的关系并不清楚。一个可能的办法是，先写出表示一个好的聚类的目标函数，然后构造一个

能达到这个要求的算法。

假设我们知道共有 k 类，其中 k 已知，这样就能得到一个很自然的目标函数。每一类假设有一个中心，将第 i 类的中心记做 $\boldsymbol{c}_i$。第 j 个元素被记做特征向量 $\boldsymbol{x}_j$。例如，如果聚类分散的点，那么 $\boldsymbol{x}$ 将会是点的坐标；如果分割一幅灰度图片的话，$\boldsymbol{x}$ 将会是一个像素点的灰度。

现在，假设元素都靠近它们所在类的中心，则有下面的目标函数

$$\Phi(\text{clusters, data}) = \sum_{i \in \text{clusters}} \left\{ \sum_{j \in i\text{thcluster}} (\boldsymbol{x}_j - \boldsymbol{c}_i)^T (\boldsymbol{x}_j - \boldsymbol{c}_i) \right\}$$

注意到，如果聚类中点的划分已确定，将很容易计算出每一类最合适的中心点。然而，聚类中点可能的划分种类数太大，以致不可能搜索到最小值。我们定义以下算法，它通过两步进行迭代：

- 假设聚类中心已知，并且分配每个点到最近的聚类中心。
- 假设分配已确定，选择一个新的聚类中心集。每个中心是分布在这个类中各点的平均值。

然后随机地选择聚类中心作为起始点，并轮流迭代执行这些步骤。这种方法最后收敛于目标函数的局部最小值（在每一步中，函数值要么减小，要么不变而且它是有下界的）。尽管如此，它并不一定收敛于全局最小。它也并不一定最后得到 k 类，除非修改分配方法以保证每一类都能有一些点。这种算法通常称为 k-均值算法。通过不同的 k 值使用 k-均值算法并比较其结果可以得到一个合适的聚类类数；我们将在 16.3 节中具体讨论这个问题。

应用这种方法进行图像分割时还有一个问题：分割得到的区域并不是连通的，甚至是很零散分布的（如图 14.13 和图 14.14）。可以通过将像素点的坐标作为特征来减少这种影响——一种将大的区域分割的方法（如图 14.15）。

图 14.13 左图是各种蔬菜混合的一张图片，采用 k-均值算法将它分割得到的图像如中图和右图。将每一个像素点用它所在类的均值替代；结果正如期望的是一幅自适应量化的图像。中图采用灰度信息进行分割，右图采用颜色信息进行分割。每次分割都假定为5类

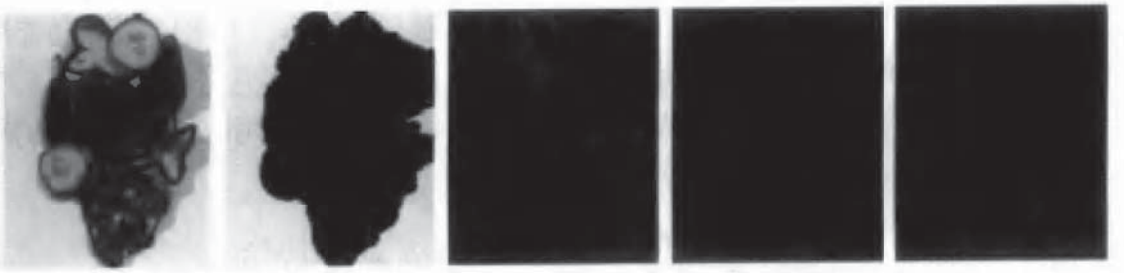


图 14.14　显示一些蔬菜图像采用 k-均值分割的结果，其中假设分类数为 11。左图显示了用均值替代所有像素点原值的所有分段。其他的图分别显示了4种划分。注意这种方法产生的分类集合不保证连通。对于这幅图像，一些划分确实非常像物体，但是一个划分也可能代表了多个物体（如辣椒）；其他的根本没有意义。因为没有考虑到纹理的问题导致将红色卷心菜的不同叶片分成了很多不同的部分

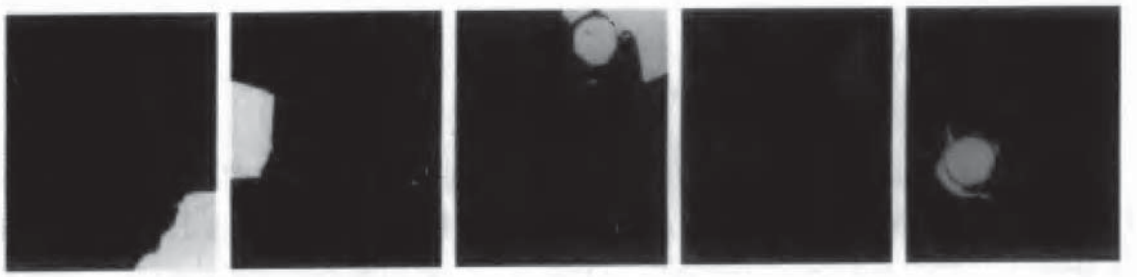


图 14.15　采用 k-均值分割蔬菜图像得到的 5 个部分，其中采用了像素的位置信息作为特征向量的一部分，并且分类数为20而不是11。注意，因为背景区域的点相距中心点太远，大片的连贯的背景区域被分割开了。每个单独的辣椒被很好地分割开来，而红色的卷心菜因为没有考虑纹理的问题仍然被分开了

算法 14.5　k-均值聚类算法

选择 k 个数据点作为类的中心点

直到类的中心不再改变退出

　将每一个数据点放到中心点离它最近的类中

　确保每一个类至少有一个数据点；解决这个问题可以从离它所在类中心较远的数据点中随机地选出一个放在空类中。

　用类中数据点的平均值替换旧的类中心点的值。

end

14.5　基于图论的聚类分割

聚类可以看做是将一个图切成合适的小片的问题。实际应用中，将每一个数据项在加权图中表示为一个顶点，如果两元素相似，那么两顶点之间的边的权值就比较大，反之亦然。然后通过去掉较小的相关权重边，将图分割成内部权重较大的各个相连的部分（每个部分和一个聚类相对应）。在这种观点下，产生了一系列不同的、非常成功的分割算法。

14.5.1　图论术语

一些术语很容易忘记，这里做一个简短的回顾：

- 图是一些顶点 V 以及连接这些顶点 E 的边的集合，被记做 $G = \{V, E\}$。每条边能用一对顶点来表示——即，$E \subset V \times V$。经常通过画一些点，以及一些连接它们的曲线来表示图。
- 有向图是边(b, a)和(a, b)不一样的图；这样的图在画的时候要使用箭头表示边的方向。
- 无向图是边(b, a)和(a, b)没有区别的图。
- 加权图是边有权值的图。
- 自环就是两端的顶点相同的一条边；自环不会出现在实际的应用中。
- 如果存在一组相连的边使得它起始于其中一点，终结于另一点，我们就说这两个顶点是相连的；如果是有向图，则要求所有箭头的方向一致。
- 连通图是一个任意两个顶点都是相连的图。
- 每个图都由一些独立的连通子图组成的——也就是说，$G = \{V_1 \cup V_2 \cdots V_n, E_1 \cup E_2 \cdots E_n\}$，其中所有的子图$\{V_i, E_i\}$都是连通的，而且不存在连接集合 V_i 和集合 V_j 中元素的边 E，其中 $i \neq j$。

14.5.2 方法概述

一个加权图能够通过一个方阵表示，理解这一点是很有用的(如图 14.16)。每一个顶点有一个行数和一个列数。矩阵的第(i, j)元素表示了连接顶点 i 到 j 的边上的权重；对于一个无向图，用一个对称矩阵表示，在对称的(i, j)和(j, i)位置的值各取权重的一半。

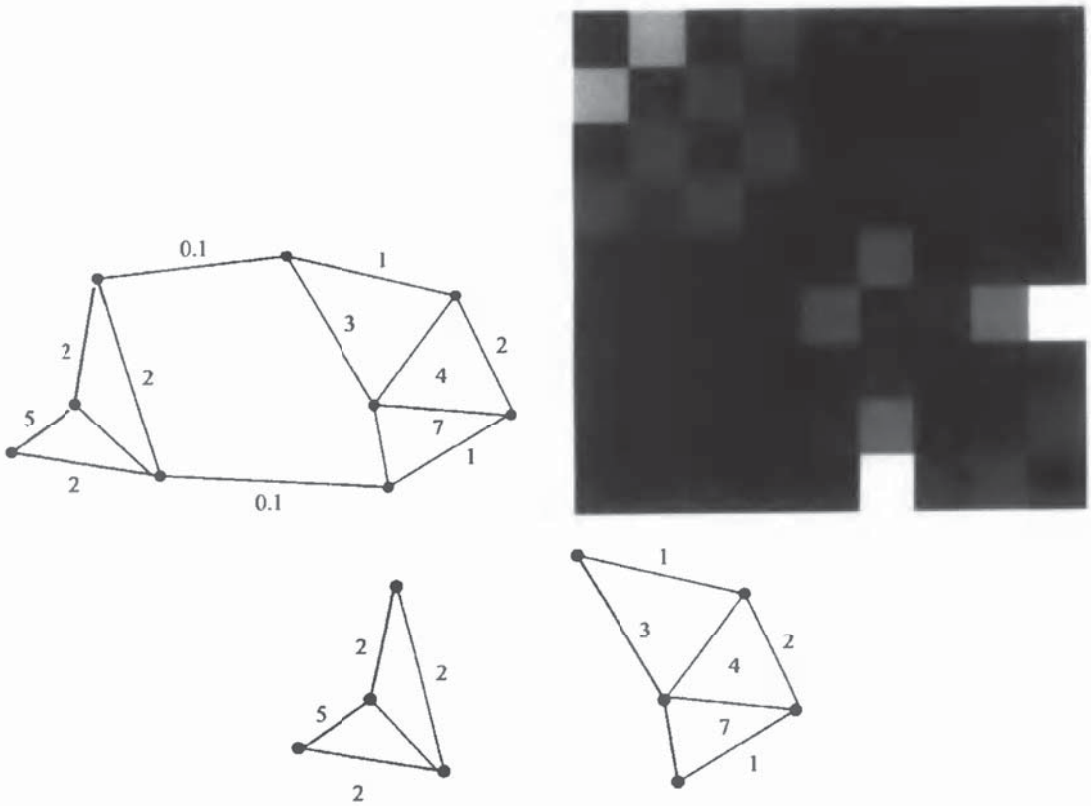


图 14.16 左上角画的是一个无向的加权图；右上角是无向图对应的加权矩阵。大的值对应亮一些的颜色。如果将点和行(和列)按不同的顺序对应，矩阵将被重新排列。我们选择了一个排列顺序，使得矩阵看起来能够突出它的块对角性。下面的图显示了将原来的图切分成为两个紧密相连的子图。这种分割将图对应的矩阵切分为对角线上的两个主块

图论在聚类中的应用如下：对每一个需要聚类的元素，将它和图上的一个顶点对应起来。构造任意两点之间的边，边上的权重代表了两个元素的相似程度。去掉一些边形成一些比较好的连通子图——理想情况下，子图内的边的权重大于子图间的连接边的权重。每个子图就是一个聚类。例如，图 14.17 显示了一些分散的点和使用某种相似性度量计算的相应的加权矩阵（也就是说，无向图只是画法上有区别）；理想的算法能使得这个矩阵看起来近似于一个块对角矩阵——因为类内相似性很强，而类间相似性较弱——再把它分解成两个矩阵，每一个是其中一个对角块。需要研究的问题就是划分子图的标准和形成连通子图的算法。

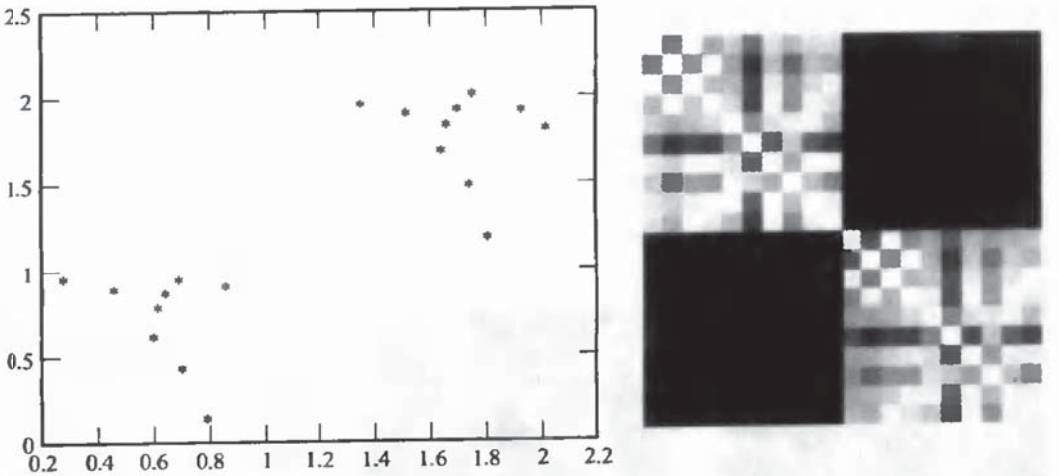


图 14.17　左图，是平面上的一些点。右图，是对于这些点使用与距离相关的衰减指数计算出来的相似性矩阵（14.5.3节），其中较大的值在颜色上浅一些，较小的值暗一些。注意这个矩阵近似于块对角结构；矩阵中有两个几乎全零的非对角块。这些对角块对应于两个很明显的类内的连接；两个非对角块对应于两个类之间的连接关系

14.5.3　相似性度量

当我们将分割看做简单的聚类时，需要提供一些测量方式来表示聚类各类之间的相似度。目前的分割模型要求给图的每一条边一个权重；这些权重在术语上通常称为相似性度量。很显然，相似性度量依赖于当前需要解决的问题。连接很相似的两个节点的弧线上的权重应该大，而连接两个不同的节点的自然要小一些。

基于距离的相似性　一旦距离超过某一个阈值，相似性应该随距离增加急剧下降。一个比较适当的表达式如下：

$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{-\left((\boldsymbol{x}-\boldsymbol{y})^t(\boldsymbol{x}-\boldsymbol{y})/2\sigma_d^2\right)\right\}$$

其中，σ_d 是一个参数，如果要将较远的点聚合起来，其值应较大，而只聚合较近的点，则其值应较小（在图 14.17 中就是使用这个表达式；注意，比例的选择对结果的影响非常大，如图 14.18 所示）。

基于亮度的相似性　当亮度值接近时相似性大，如果亮度值之差变大，则相似性减小。同样可以通过一个指数形式的表达式来表示：

$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{-\left((I(\boldsymbol{x})-I(\boldsymbol{y}))^t(I(\boldsymbol{x})-I(\boldsymbol{y}))/2\sigma_I^2\right)\right\}$$

基于颜色的相似性　需要一个均衡的颜色度量来构造有意义的颜色相似性函数。使用一

个均匀分布的颜色空间是个很好的办法，但是使用 RGB 空间并不好（原因很明显；如有问题可以重温 6.3.2 节）。一个合适的表达式如下：

$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{-\left(\mathrm{dist}(\boldsymbol{c}(\boldsymbol{x}),\boldsymbol{c}(\boldsymbol{y}))^2/2\sigma_c^2\right)\right\}$$

其中，$\boldsymbol{c}_i$ 是像素点 i 的颜色表示。

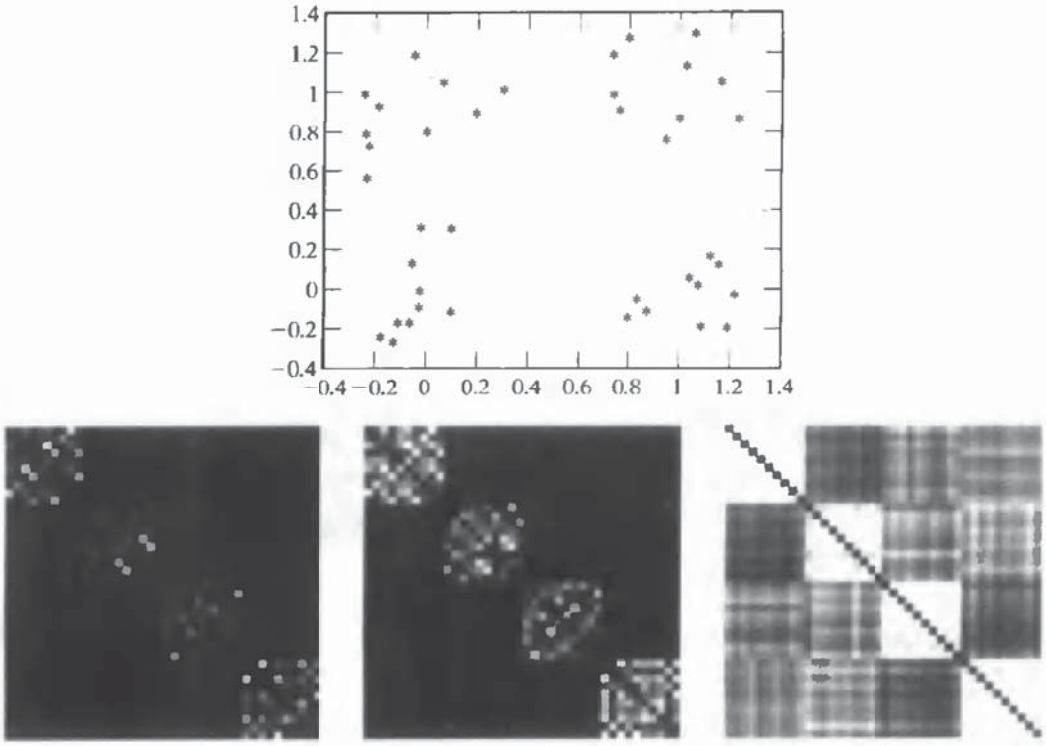


图 14.18 相似性尺度的选择对于相似矩阵的影响。上面的图给出了一些点集，共有 4 组点，每组 10 个点，它们是拥有不同的均值的旋转对称正态分布。这些点在每个方向上的标准方差为0.2。下面的图中是使用不同的σ_d值对这点计算得到的相似性矩阵。左图中，$\sigma_d = 0.1$，中间的$\sigma_d = 0.2$，右图$\sigma_d = 1$。当尺度选择最小时，所有点之间的相似性都很小；稍大一些的尺度，在相似阵中有4个很清楚的子块；对于最大的尺度，子块的数目就不很明显了

基于纹理的相似性 对于相似的纹理，相似性应该大；随着差别的增加，相似性变小。我们采用一些滤波器 $f_1,\cdots,f_n$，通过这些滤波器的输出来描述纹理，这些输出将分布于不同的尺度和方向等级上。对于大多数纹理，滤波器在每一点的输出是不同的（例如棋盘），但是根据一定邻近范围内的输出构造的直方图却很有用。这也告诉我们，首先对于每一点建立一个局部的尺度（可以通过读取粗粒度滤波器中的能量），然后计算由该尺度限定区域内滤波器输出的直方图——多半是以正被讨论的点为圆心的圆形区域。我们将直方图记为 $\boldsymbol{h}$，使用下面的指数形式：

$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{-\left((\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{f}(\boldsymbol{y}))^t(\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{f}(\boldsymbol{y}))/2\sigma_f^2\right)\right\}$$

14.5.4 特征向量和分割

在第一个例子中，假设有 k 个元素和 c 类。可以通过一个 k 元的向量表示一个类，允许各元素通过某个连续的权重与类相对应——这里对于权重一词在语义上有点模糊，但是其目的

是，如果特征向量中某个分量值比较小，则元素与这一类的相关性不是很大；反之亦然。

提取一个好的分类　一个好的分类就是类中各元素和该类有很大的相关性并且在相似性矩阵中彼此有比较大的值。将代表元素相似性的矩阵记为$\mathcal{A}$，表示元素与第 n 类相关性的权重向量记为 $\boldsymbol{w}_n$。这样我们能够构造一个目标函数

$$\boldsymbol{w}_n^T \mathcal{A} \boldsymbol{w}_n$$

这是一个各项求和的式子：

$$\{\text{元素 } i \text{ 与第 } n \text{ 类的相关性}\} \times \{\text{元素 } i \text{ 与 } j \text{ 的相似性}\} \times \{\text{元素 } j \text{ 与第 } n \text{ 类的相关性}\}$$

选择一个权重向量使得上面的目标函数最大，我们就可以得到一个分类。这个目标函数本身的值没有什么意义，因为它可以通过缩放因子 λ 乘以 $\boldsymbol{w}_n$ 使得总的结果被缩放 λ^2 倍。不过，我们可以通过 $\boldsymbol{w}_n^T \boldsymbol{w}_n = 1$ 对权重向量进行归一化。

也就是说在 $\boldsymbol{w}_n^T \boldsymbol{w}_n = 1$ 条件下，求 $\boldsymbol{w}_n^T \mathcal{A} \boldsymbol{w}_n$ 的最大值。有 Lagrangian 式：

$$\boldsymbol{w}_n^T \mathcal{A} \boldsymbol{w}_n + \lambda \left(\boldsymbol{w}_n^T \boldsymbol{w}_n - 1 \right)$$

（其中，λ 是 Lagrange 因子）。对它微分并且去掉两边的一个因子得到：

$$\mathcal{A} \boldsymbol{w}_n = \lambda \boldsymbol{w}_n$$

也就是说 $\boldsymbol{w}_n$ 是$\mathcal{A}$的特征向量。这样，可以通过寻找最大特征值对应的特征向量来得到一个分类——类的权重就是特征向量的元素值。对于聚类现象是很明显的问题，可以指望类的权重对于那些属于该类的元素取值很大，而对于其他的元素应该近乎为零（见图 14.19）。事实上，也可以通过$\mathcal{A}$的其他的特征向量得到其他类的权重。

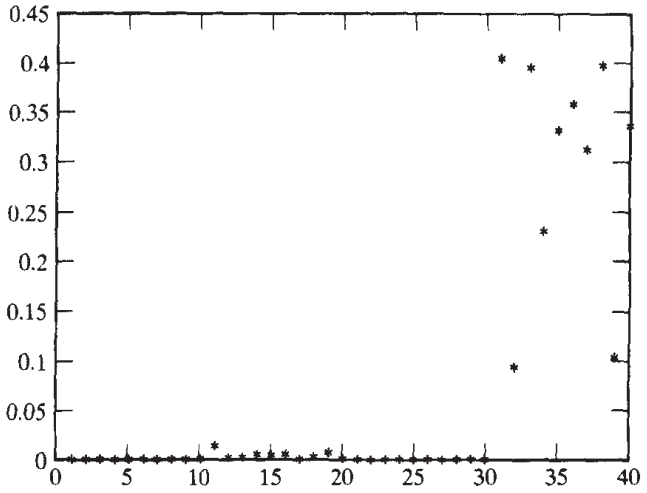


图 14.19　与图 14.18 中的数据集相对应，且 $\sigma_d = 0.2$ 的相似性矩阵最大特征值对应的特征向量。注意，大多数值都是很小的，但是有些——与该类相对应的元素——比较大。相关性幅度并不是很重要，因为特征向量放大了还是特征向量

为聚类提取权重　一般的视觉问题中，只有在较少元素之间存在较大的相关权重，我们有理由相信面临的问题是比较紧凑和明显的聚类问题。

这些属性使得相似矩阵有一个很特别的结构。特别地，如果重新标号图中结点，这样矩阵$\mathcal{A}$的行和列就会打乱。我们期望面临的问题只有少量的结点有较大权重。此外，我们还期望

这些结点形成相对连贯的、独立的一个类。也就是说，可通过重新排列矩阵 M 的行和列形成一个近似于块对角的矩阵(对角块即为一个类)。重新排列矩阵 M 只是重新排列特征向量的元素，这使我们能直接考虑变换后的矩阵 M 的特征向量(也就是，图 14.16 是一个很好的例子)。

对角块矩阵的特征向量由补零以后的对角块的特征向量组成。我们期望每一个对角块都有一个特征向量和很大的特征值对应(与聚类相对应)，而其他的特征向量都对应较小的特征值。从这一点我们知道，如果存在 c 个(其中 $c < k$)类，那么与前 c 个大的特征值对应的特征向量各代表一个类。

这也就是说，这些特征向量中的每一个就是对角块补零以后的特征向量。特别地，典型的特征向量包含一些大的分量(对应相应的对角块)和一些近乎为零的分量。我们可以期望，对每一个给定的分量，仅有其中的一个特征向量有较大的值，而其他的比较小(如图 14.20)。这样，就能将前 c 个大的特征值对应的特征向量转化为前 c 类相应的类权重。将聚类权重量化为零或 1 就得到一个离散的聚类；这就是图中所显示的。

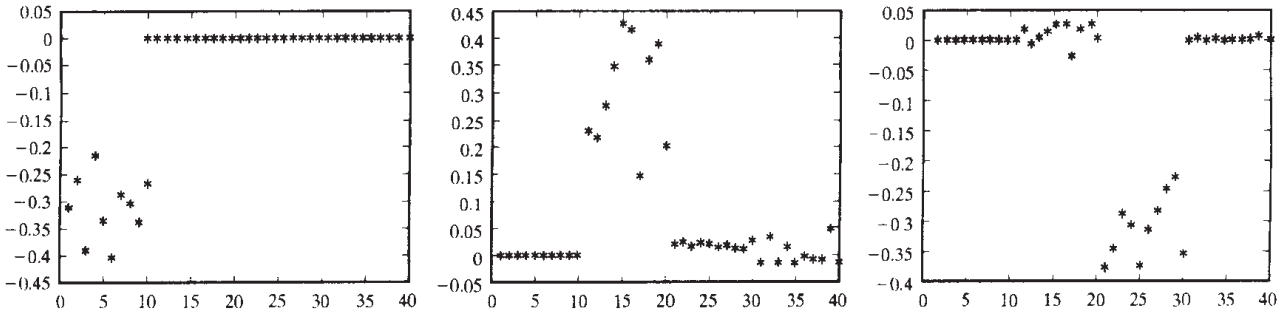


图 14.20 与图 14.18 中数据集的相似性矩阵中三个次最大特征值相对应的三个特征向量，$\sigma_d = 0.2$(与最大特征值相对应的特征向量显示于图14.19中)。注意，大部分值很小，但是对于(分开的)元素集合，对应的值很大。这也与相似性矩阵的块结构相一致。相关性幅度并不是很重要，因为特征向量放大了还是特征向量

这是一个定性的论证，有一些图片表明了这个论证的可疑之处。此外，尽管我们的讨论说明在A的频谱附近搜索是值得的——希望找到一个小的较大特征值的集合和一个大的较小特征值的集合(如图 14.21)，但是我们还不清楚如何去确定 c 的大小。

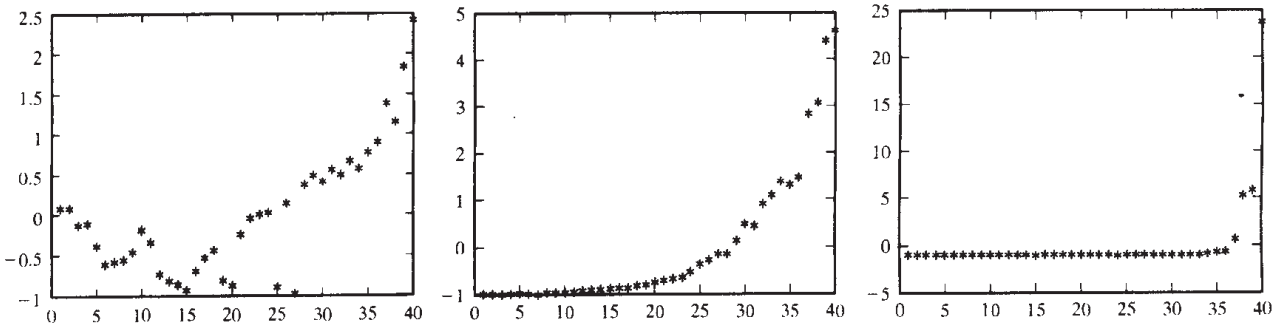


图 14.21 类的数目通过相似性矩阵的特征值反映出来。上图显示了图 14.18 中各种情况下相似性矩阵的特征值。左图对应 $\sigma_d = 0.1$；中间对应 $\sigma_d = 0.2$；右图对应 $\sigma_d = 1$。对于最细的尺度，因为所有元素之间的相似性都比较小，所以得到很多比较大的特征值；而对于下一个稍大的尺度，有 4 个比其他特征值大的特征值；对于最大的尺度，仅仅只有两个较其他特征值大的特征值

算法 14.6　基于图特征值的聚类算法
构造一个相似性矩阵
计算相似性矩阵的特征值和特征向量
直到产生足够多的分类
　取未处理过的最大特征值对应的特征向量；将所有与已经分类的元素相对应的分量清零，然后通过给剩下的部分一个阈值来决定哪些元素属于这一类，可以通过剩下的聚类分量选择一个阈值，也可以使用一个事先决定好的阈值。
　如果所有的元素都被分类，或已有足够的分类
end

14.5.5　归一化切分

前面章节的定性讨论有点不够。例如，如果对角块的特征值比较接近，就不能得到可以用来分类的特征向量，因为相同特征值对应的特征向量的任何线性组合仍然是一个特征向量(如图 14.22)。

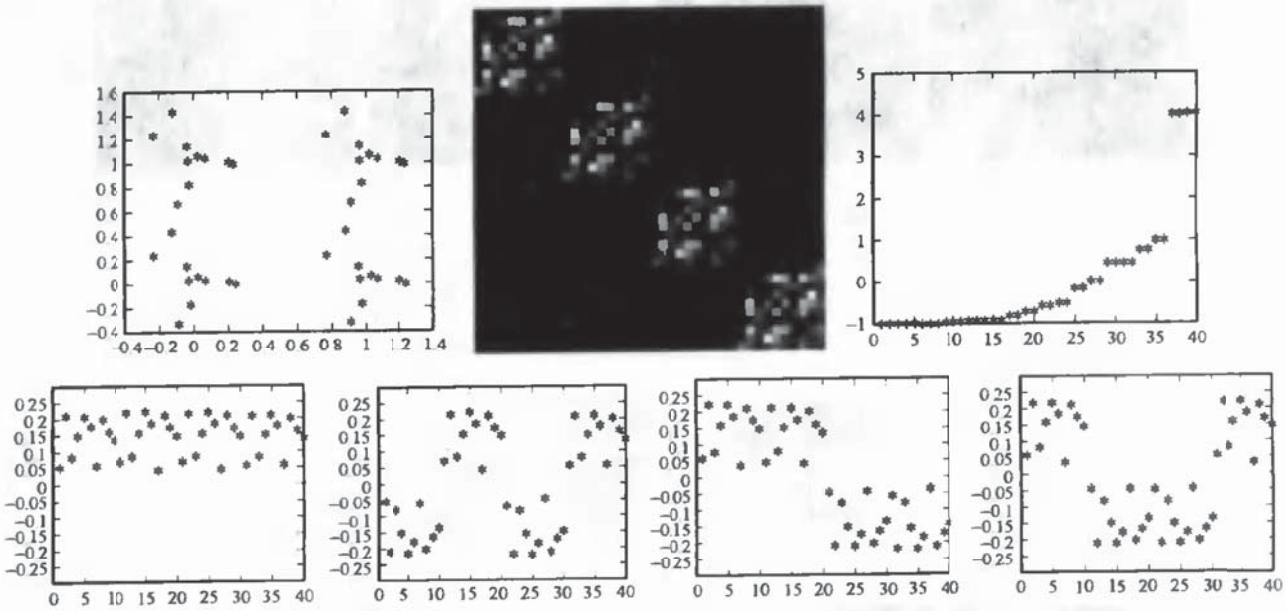


图 14.22　相似性矩阵的特征向量可能误导分类。左上角的数据集由一个点集的 4 个副本组成；这导致在相似性矩阵中重复的块结构，如上面中图所示。每一个块有相同的谱，这导致有4个(大致)相同的特征值的相似性矩阵(右上图)。下面的一行图显示了对应于4个最大特征值的特征向量；注意(a)这些值并不提示分类信息，(b)特征向量的一个线性组合可以得到很好的分类

一个可行的办法就是将图切分为两个连通的部分，使得这样切分的代价只是每个部分内总相似性的一小部分。可以将此形式化为将一个加权图切分 V 为 A、B 两个部分，并且对这种切分进行评价：

$$\frac{cut(A,B)}{assoc(A,V)}+\frac{cut(A,B)}{assoc(B,V)}$$

其中，$cut(A,B)$是图 V 中连接 A 和 B 中元素的所有边权重之和，$assoc(A,V)$是有一个顶点在 A 中的所有边的权重之和。如果切分开来的两个部分之间有很少的较小权重的连线，而两部

分内部有很多权重较大的连线，那么这个值就会很小。根据这个原则，我们希望找到使这个值最小的分割，称之为归一化切分。这个准则在实际应用中很成功（见图 14.23 和图 14.24）。

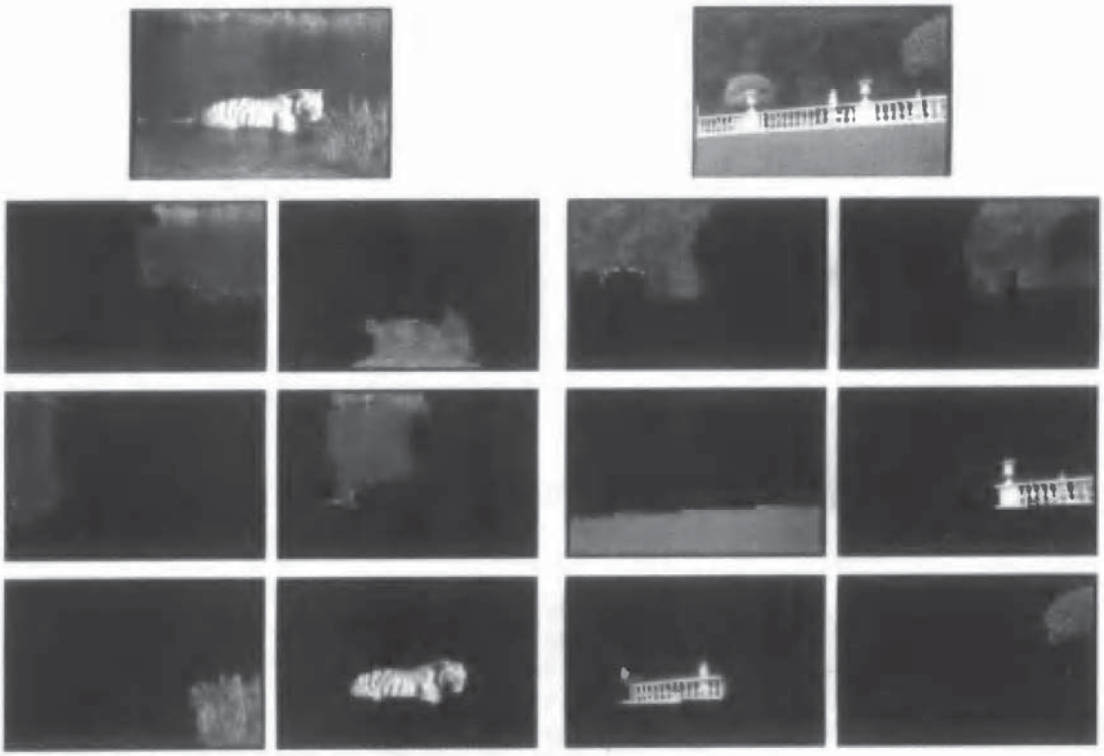


图 14.23 使用前面描述的归一化切分方法将上面的图片分割成所显示的各个部分。使用的相似性矩阵包含 14.5.3节中说的亮度和纹理。老虎游泳的图像被分割成：一个部分是老虎，一个部分是草坪，以及4个部分的湖面。同样，栏杆被分割成三个合理的连续的部分。注意，通过加入纹理可以对 k- 均值方法有所改进

图 14.24 上图：来自于一个显示一个人移动的视频流中的两帧图像。下图：通过归一化切分和时空相似函数建立的时空分割（见14.5.3节）

将元素向量记为 $\boldsymbol{y}$，每个分量对应图中一个结点，它们的值是 l 或者 $-b$。$\boldsymbol{y}$ 的值用于区分图的不同部分：如果 $\boldsymbol{y}$ 的第 i 个分量是 l，则相应的图中的结点属于同一个子图；如果是 $-b$，则结点属于另一个子图。我们将相似性矩阵记为 $\mathcal{A}$，用 $\mathcal{D}$ 表示次数矩阵（degree matrix），次数矩阵的每一个对角线元素是相应结点权重之和，即为，

$$D_{ii} = \sum_j A_{ij}$$

$\mathcal{D}$的非对角线元素为零。在这样的标记下,准则可以写做

$$\frac{\boldsymbol{y}^T(\mathcal{D}-\mathcal{A})\boldsymbol{y}}{\boldsymbol{y}^T\mathcal{D}\boldsymbol{y}}$$

我们希望找出使得上面的准则最小的 $\boldsymbol{y}$。这个问题是一个整数规划的问题。因为它完全等同于图的切分问题,所以它并没有简单多少。难点在于向量 $\boldsymbol{y}$ 的各分量是离散值——原则上,应该通过测试 $\boldsymbol{y}$ 的每一个可能值来解决这个问题,但是这涉及到一个随着像素点按指数级数增长的搜索空间,速度将会很慢(也许直到宇宙消失都还不能完成)。对于这个问题,一个常用的改进的解决办法就是计算出一个使得准则最小的实数向量 $\boldsymbol{y}$,然后通过设置一个阈值来决定每个元素应归属哪一边。这里有两个问题:首先必须得到这个实数向量;其次,必须选择一个阈值。

获得一个实数向量　实数向量很容易得到。作为一个练习,下面方程的一个解

$$(\mathcal{D}-\mathcal{A})\boldsymbol{y} = \lambda\mathcal{D}\boldsymbol{y}$$

是所求问题的一个实值解。惟一的问题就是将使用哪一个增广特征向量。显然,最小特征值肯定是零,因此,与第二小的特征值相对应的特征向量是比较合适的。得到这个特征向量的一个最简单的办法,就是变换 $\boldsymbol{z} = \mathcal{D}^{1/2}\boldsymbol{y}$,从而得到

$$\mathcal{D}^{-1/2}(\mathcal{D}-\mathcal{A})\mathcal{D}^{-1/2}\boldsymbol{z} = \lambda\boldsymbol{z}$$

于是求解 $\boldsymbol{y}$ 就简单了。同时这个问题的解也是下面方程的解

$$\mathcal{N}\boldsymbol{z} = \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}\boldsymbol{z} = \mu\boldsymbol{z}$$

其中$\mathcal{N}$有时被称为归一化仿射矩阵。

选择一个阈值　选择一个合适的阈值并不是非常困难;假设图中有 N 个结点,因此,向量 $\boldsymbol{y}$ 就有 N 个分量并且最多有 N 个不同的值。如果阈值取 v,我们将归一化切分准则的值记为 $ncut(v)$,那么最多有 $N+1$ 个 $ncut(v)$值。我们能得到所有这些值,从中选择一个阈值使得该取值最小。同时注意到这个形式实际上是一个递归的过程,结果的每一个部分也是一个图,并且这些新图同时也能再次分裂。一个更简单的标准似乎更实用一些,那就是递减地遍历特征值,使用与小的特征值对应的特征向量来获得新的聚类。

14.6　注释

分割是一个比较难的话题,存在各种各样的解决方法。过去主要的综述有 Fu 和 Mui(1981),Haralick 和 Shapiro(1985),Nevatia(1986)以及 Riseman 和 Arbib(1977)。更近一些的综述很少,其中有 Pal 和 Pal(1993)。其中一个原因就是,很难从比举例更实用的层次,评价一个分割方法。针对某具体任务的评价相对要简单一些;涉及这种类型的论文包括 Hartley,Wang,Kitchen 和 Rosenfeld(1982);Ranade 和 Prewitt(1980);Yasnoff,Mui 和 Bacus(1977)和 Zhang(1996)。

早先的聚类分割研究者是 Ohlander, Price 和 Reddy(1978)。聚类方法往往非常随意——当然,这并不是说它们没有什么用——因为确实没有太多可能的理论去预测哪些需要聚类以及怎样聚类。很显然,我们正在做的就是形成对实际应用有用的聚类,但是并没有任何有用的办

法来将这个准则规范化。在这一章中,我们试图给出概略描述而忽略其中细节,因为对于已做工作做很详细的记录是很不明智的。

各种基于图论理论的聚类算法已经应用于视觉(见 Sarkar 和 Boyer, 1998,以及 Wu 和 Leahy,1993;在 Weiss,1999 中有一个总结)。归一化的切分形式应归于 Shi 和 Malik(1997)和(2000)。一些变化涉及在运动分割中的应用(Shi 和 Malik,1998*a*)和从输出推断相似度矩阵的方法(Shi 和 Malik,1998*b*)。还有多种其他的标准(例如,Cox,Zhong 和 Rao,1996;Perona 和 Freeman,1998)。我们强调基于图论理论的聚类方法,是因为这种方法解决人们关心的任何相似性问题的能力都很有吸引力。

分割在视觉问题中也是一个正在研究中的重要问题,因而对所做的工作进行比较详细的记录所涉及的工作量会很大。直到最近,还经常把识别和分割当做不同的两件事情来讨论。这种观点已经也应该过时了,因为构造某些并不实用的分类表示并没有太大意义。此外,如果我们能够确定要识别什么,那么就有可能确定分割表示应该是什么样子的。

人的分割和聚类功能

人类视觉感知中的分类问题有很多著作。标准的完形(Gestalt)手册包括 Kanizsa(1979)和 Koffka(1935)。主观的轮廓是 Kanisza 第一个描述的;Kanizsa(1976)有一个广泛的总结性讨论。Palmer(1999)的一本权威性的书给出了比这里更多的图片。Gordon(1997)还有许多关于不同视觉理论的发展和完形(Gestalt)的起源的信息。在视觉处理中某些聚类似乎很早就出现了一种称之为突显(pop out)的现象(Triesman,1982)。

感知分类

感知组织被看做是用聚类的方法将图像样本分成有用的类,分割被看做是将图像切分成不同的区域,有时候两者被认为是有区别的。我们并不接受这样的区别,将这些问题看做是相同的一件事情的好处在于,人们可以自由地将一个问题的算法转化为另一个问题所用。我们并不详细讨论感知组织问题,主要是因为我们的重点在于说明问题而不追求历史的准确性,并且这些方法也遵循统一的观点。例如,有很多关于将图像边缘点或者线连接出现可能性很大的结构的聚类文章。我们将在下面的章节中介绍其中一些方法,也将引导读者注意到 Amir 和 Lindenbaum(1996),Huttenlocher 和 Wayner(1992),Lowe(1985),Mohan 和 Nevatia(1992),Sarkar 和 Boyer(1993)以及(1994)。在构造用户接口中,(如我们前面提示的)了解什么对感知是重要的这一点非常有用(例如,Saund 和 Moran,1995)。

习题

14.1 我们希望用颜色和纹理的区别对一个像素点集进行聚类。在 14.4.2 节提到的目标函数

$$\Phi(\text{clusters, data}) = \sum_{i \in \text{clusters}} \left\{ \sum_{j \in i\text{thcluster}} (\boldsymbol{x}_j - \boldsymbol{c}_i)^T (\boldsymbol{x}_j - \boldsymbol{c}_i) \right\}$$

可能不太合适——例如,如果颜色和纹理不是按相同的尺度进行度量,那么颜色的差别将占据绝大部分的分量。

(a) 用下面形式的目标函数，扩展相应问题的 k-均值算法：

$$\Phi(\text{clusters, data}) = \sum_{i \in \text{clusters}} \left\{ \sum_{j \in i\text{'thcluster}} (\boldsymbol{x}_j - \boldsymbol{c}_i)^T \mathcal{S} (\boldsymbol{x}_j - \boldsymbol{c}_i) \right\}$$

其中，$\mathcal{S}$是一个对称的、正实数矩阵。

(b) 对于更简单一些的目标函数，必须保证每一类至少含有一个元素（否则不能计算类的中心）。对于更复杂的目标函数，一个类至少需要包含多少个元素呢？

(c) 正如在 14.4.2 节中说明的，不能保证 k-均值算法能够使得目标函数在全局范围内收敛；证明它肯定能够局部收敛。

(d) 为 k-均值聚类算法找出两个可能的局部最小值将二维的数据点集聚类。简单起见，只需要使用仅含两类的例子。不需要使用太多的数据点，只是将两种目标函数都尝试一下。

14.2 阅读 Shi 和 Malik(2000)，理解文中关于归一化切分标准是一个整数规划问题的证明。

14.3 本题旨在训练使用归一化切分得到两类以上的分类。一个办法就是为每一个独立的部分构造一个新图，然后反复递归地调用算法。你会发现这个方法和经典的分解式聚类算法非常相似。另一个办法就是寻找与最小特征值对应的特征向量。

(a) 说明为什么这两种方法不同。

(b) 假设有一个包含两个连通部分的图。请描述和最大特征值对应的特征向量。

(c) 描述与第二大特征值对应的特征向量。

(d) 在一定条件下两种方法不断地产生更多类，是否会逐渐得到相似的结果？这里一定的条件是指什么？

编程作业

14.4 编写一个使用运动平均的背景差分算法，并将它用于滤波器实验。

14.5 编写一个镜头边界检测的算法，使用任何两种所见到的方法，并且在两个视频中进行实验。

14.6 使用基于颜色和位置的 k-均值分割算法进行分割。描述分类类数选择对问题的影响以及不同的局部最小值对问题的影响。

第 15 章　基于模型拟合的分割

一种分割的观点是，如果一些像素（样本）聚在一起见证了某种模型，则把它们分在一起。这个观点与聚类很类似；主要差别在于分割中模型是已知的，而且是从更大尺度的关系看问题，而不仅仅是样本到样本的关系。例如，想像有一个程序试图把一些看上去像直线（有时并不需要精确描述）的样本划分到一个集合里，如果仅仅看两两样本之间的关系的话，是不可能做到的，相反，应该通过选择一个模型，然后确定一个拟合好坏的准则，来审视一组样本是否具有拟合该模型的属性。另一种观点是我们在聚类样本，因为它们形成了一个常见的几何形状——例如，它们都在一条直线上或都在一个圆上。不论用哪一种观点，这种行为一般叫做拟合（fitting）。

15.1　哈夫变换

直线拟合中有三个问题。第一，给定了属于某条直线的点，那么直线是什么？第二，哪些点属于哪条直线？最后，有多少条直线？哈夫变换就是一个可以解决上面三个问题的方法（尽管在实际中很难全部解决）。这是值得去理解的，因为该方法非常普遍，而且在实际中有大量的应用。

一种把在同样结构上的点聚类的方法是记录所有点能具有的所有结构，然后看看哪一个结构有最多的投票。这种（非常一般性）的技术称做哈夫变换。我们找每一个图像样本，然后确定该样本可能通过的所有结构。记录这个集合（可以把它看做投票），然后对每一个样本都这样做。我们查看投票，决定哪一个结构是真正存在的。例如，如果我们汇集所有在直线上的点，取每一个点，然后对可能穿过它的所有直线投票；对每一个点重复这一过程。那些存在的直线会突显出来，因为它们穿过很多点，所以有很多投票。

15.1.1　用哈夫变换拟合直线

哈夫变换最成功的应用是在直线检测上。我们举这个例子说明这个方法和它的缺点。一条直线很容易地参数化为点集(x, y)，使得

$$x\cos\theta + y\sin\theta + r = 0$$

现在每一个(θ, r)对代表一条惟一的直线，其中，$r \geqslant 0$ 表示该直线与原点的垂直距离，$0 \leqslant \theta < 2\pi$。我们把点对的集合$(\theta, r)$叫做直线空间，空间可以可视化为一个半无穷的柱。有一族直线通过任何一个点样本。具体说来，在直线空间中，由 $r = -x_0\cos\theta + y_0\sin\theta$ 给定的某条曲线上的直线都通过点样本(x_0, y_0)。

因为图像有已知的尺寸，因此存在某个 R 使得我们不会对 $r > R$ 的直线感兴趣，因为它们离原点太远以至于看不到它们。这意味着我们感兴趣的那些直线形成了一个平面的有界子集，用某些方便的网格（后面会讨论）将它离散化。网格的元素可以想像成放置投票的桶。桶的网格叫做累加器数组。对于每一个点样本，对该样本对应的曲线上的所有网格元素都增加一票。如果有许多点样本共线，那么可以期望会有很多票投到该直线对应的网格元素上。

15.1.2 哈夫变换在实际中的问题

遗憾的是，哈夫变换有许多严重的实际问题：

- **量化误差**：合适的网格尺寸很难选择。太粗糙的网格导致某个桶的投票值太大而无效，因为许多不同的直线对应了同一个桶。太精细的网格导致直线可能找不到，因为样本并不是准确地共线，因此所产生的投票会被记录到不同的桶里，而没有一个桶能得到大的投票（见图 15.1）。

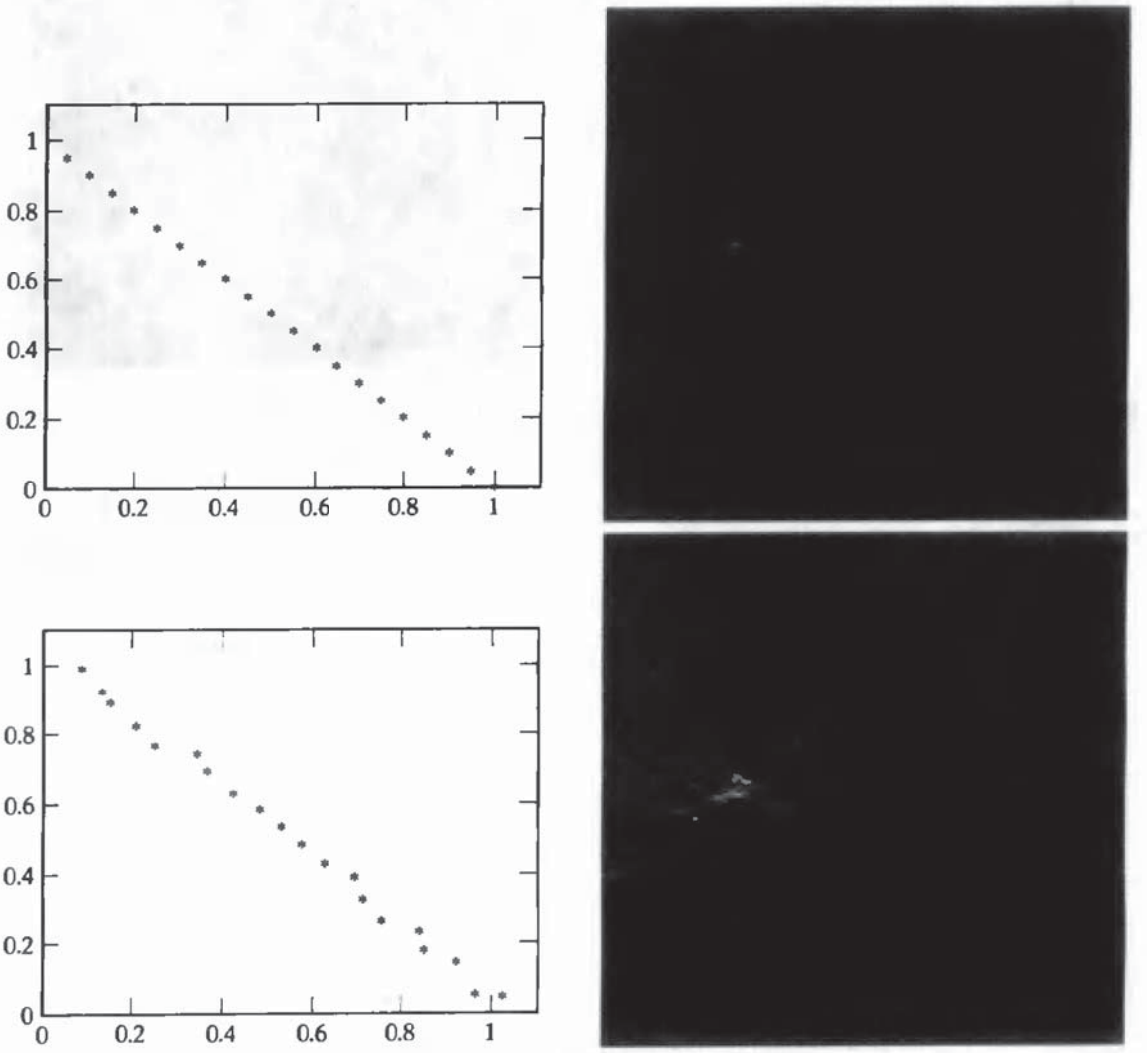


图 15.1　哈夫变换把每个点映射到通过该点的可能直线的曲线上（或者其他参数曲线）。这些图展示了检测直线用的哈夫变换。左边一列表示点，右边一列表示对应的累加阵列（投票的数目用灰度来表示，投票越多点越亮）。上面的图表示了一条直线上20个点产生的效果。右上的图是这些点的哈夫变换。在累加阵列中每个点都沿一条曲线投票；最大的投票值是20（对应最亮的那个点）。在累加阵列中的水平变量是θ，垂直变量是r；在每个方向有 200 步，r的范围是[0，1.55]。下面的图中，用一个随机的向量偏移了这些点，每个分量均匀分布在[0，0.05]之间；注意这也使得在累加阵列里的曲线偏移；最大的投票现在是6（对应图像中最亮的点，但是这个点不如上面图中的那么容易找到）

- **处理噪声的困难：**哈夫变换的好处在于它能让彼此分得很开，但靠近某个参数曲线的样本连接在一起。这同样也是一个缺点；有时候可能在一个大致均匀分布的样本中检测出一些不存在的直线（见图 15.2）。这意味着纹理区域可能在投票阵列中产生很多峰值，它的值比要检测的直线的值还大（见图 15.3 和图 15.4）。

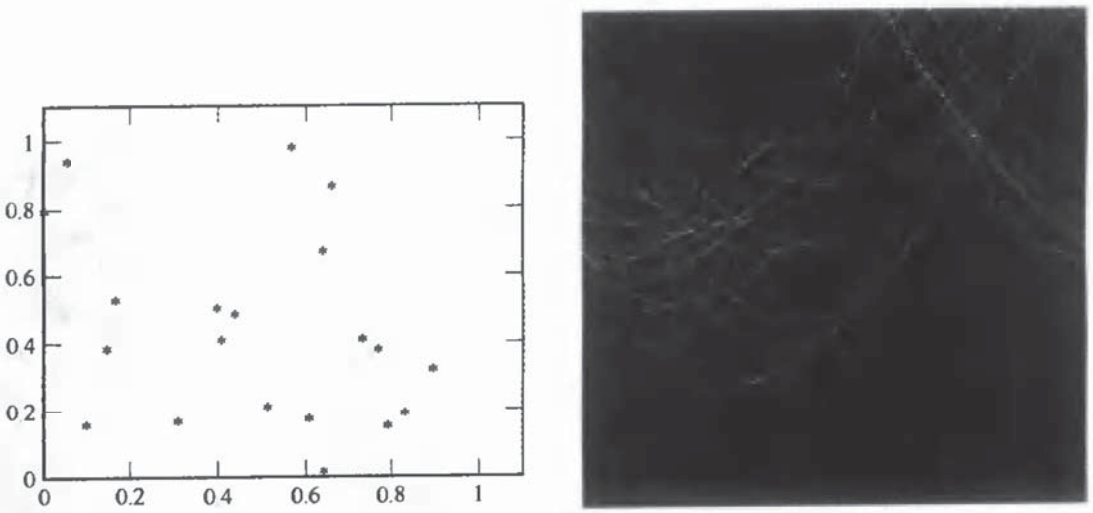


图 15.2 对于一些随机点的哈夫变换导致累加阵列里有很多投票。像图 15.1 一样，左边表示点，右边表示对应的累加阵列（投票的数目用灰度来表示，投票多的点也就越亮）。在这里，数据点是噪声（两个坐标都是[0,1]区间的均匀随机数）；累加阵列包含很多叠在一起的点，投票的最大值是4（对比图15.1里的6）。图15.3和图15.4进一步讨论噪声问题

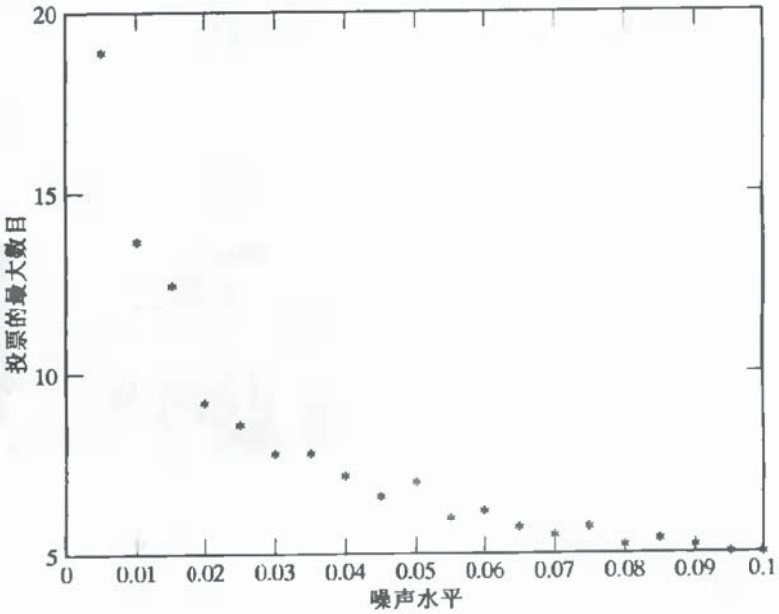


图 15.3 噪声使得哈夫变换不能鲁棒地使用。这个图显示 20 个在一条直线上的点被均匀噪声干扰后，所形成累加阵列中的最大的投票值对于噪声强度之间的关系。噪声使得曲线彼此之间发生偏移，导致投票数目的降低。这张图对10次尝试取平均。累加阵列对于每次尝试的量化程度都一样

哈夫变换尽管有这些问题，但是它仍然值得讨论，因为对于很适用的问题，它仍然可以用有价值的方法实现。实际中，它总是用来在边缘点集中检测直线。下面是一些有用的指导：

- **减少不相关样本**：可以通过调节边缘检测器来平滑纹理，设置照明产生高对比度的边缘，等等。
- **小心的选择网格**：这通常可以通过试算法获得。对某个元素投票的同时，也对所有邻近的元素投票的方法有时候会很有用。

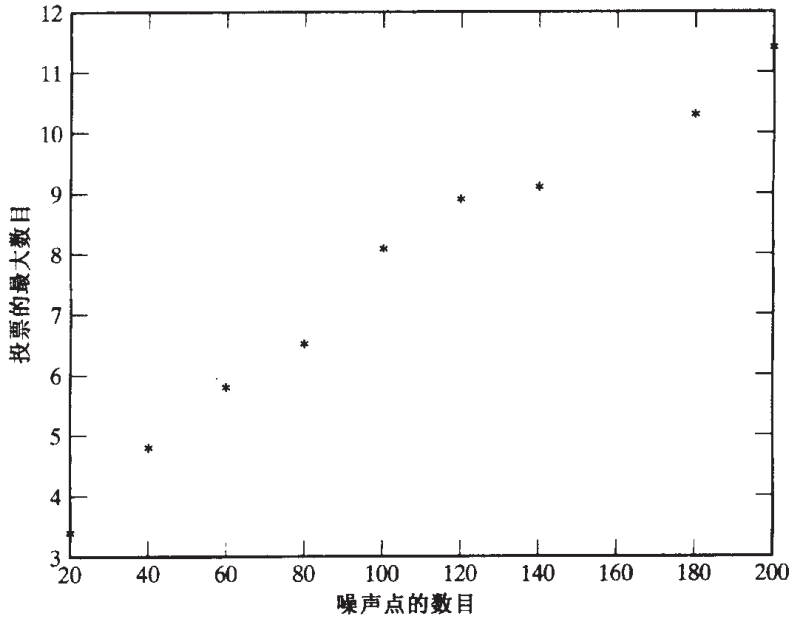


图 15.4　这个图显示哈夫变换累加阵列中最大投票数与点集点的数目之间的关系，点的坐标是 [0,1] 区间均匀分布的随机数。随着噪声的级别增长，在正确桶中的投票数下降，而在累加阵列中获得大的虚假投票值的可能性增大。这个图对10次尝试取平均。与图15.3对比，注意尺度上略有不同；比较表明，也许很难用哈夫变换从噪声中识别出直线来（因为直线产生的投票数目和噪声产生的投票数目差不多）。这些图表明在用哈夫变换之前去除噪声的重要性

15.2　直线拟合

直线拟合十分有用。在很多应用中，物体是通过直线的存在来表现的。例如，要用建筑物的图片建立建筑物的模型（像第 26 章的应用一样）。这个应用用到了建筑物的多面体模型，意味着图像中的直线非常重要。同样，许多工业零件都包含某种形式的直边；如果要识别图像中的工业零件，那么直线会很有帮助。对以上任一种情况，记录一个图像中所有直线是非常有用的分割。在这里重新回顾一下某些特殊形式的直线，以供那些未读过 3.2 节的人阅读。

15.2.1　最小二乘的直线拟合

首先假设已知所有的点属于同一条直线，并且必须找到直线的参数。这里采用符号

$$\bar{u} = \frac{\sum u_i}{k}$$

来简化表示。

最小二乘　最小二乘是一个有很长历史的拟合程序（这也是要介绍它的惟一原因）。它获

得简单的分析解，但是存在明显的偏差。对于这种方法，表示直线的形式是 $y = ax + b$。在每个数据点有(x_i, y_i)，决定选择一条直线能够通过 x 坐标的度量值最好地预测 y 坐标的度量值。

这意味着要选取直线，使得下式最小化

$$\sum_i (y_i - ax_i - b)^2$$

通过微分，直线可以是下面问题的解

$$\begin{pmatrix} \overline{y^2} \\ \overline{y} \end{pmatrix} = \begin{pmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

尽管这是一个经典问题的标准线性解，但实际上在视觉应用中没有什么用处，因为这个模型非常差。问题在于误差估计依赖于坐标系——我们把直线的纵坐标的误差作为误差，意味着接近于垂直的线会有一个很大的误差和很奇怪的拟合(见图 15.5)。事实上，这个过程太依赖于坐标系，以至于完全不能表示竖直的线。

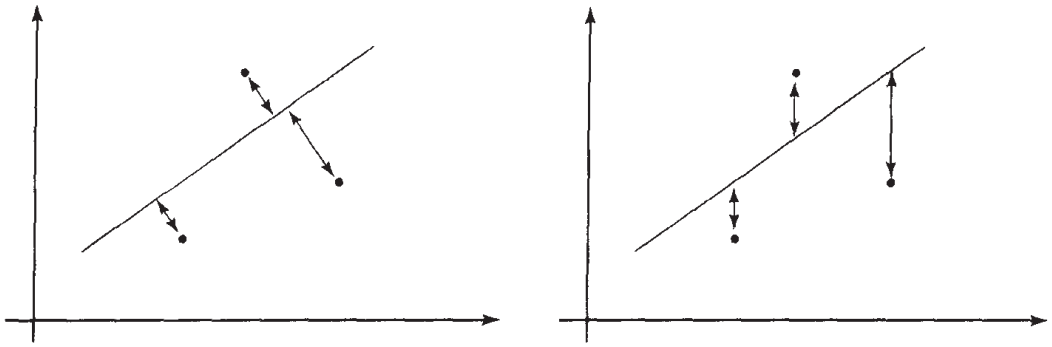


图 15.5 左图：最小二乘总误差把数据点建模为在直线上由一个抽象的点，再加上一个与直线垂直的向量所产生。希望选择一条直线可以使直线到样本的垂直距离总和最小。右图：最小二乘大体相同，但是假设误差仅出现在y坐标上。这会得到一个略微简单些的数学问题，代价是很差的拟合

最小二乘总误差 我们可以用点和线的实际距离，而不是垂直距离。这导致一个称之为最小二乘总误差(total least squares)的问题。我们可以把线作为满足 $ax + by + c = 0$ 的点集。每一条线都可以这么表示，因此可以考虑把线表示为三个数(a, b, c)。注意，用 $\lambda(a, b, c)$ $(\lambda \neq 0)$来表示的直线与(a, b, c)表示的是一样的。在练习里，要求你证明这个简单但是很有用的结论，点(u, v)到直线(a, b, c)的距离可以由下式给出：

$$\text{abs}(au + bv + c), \quad 若\ a^2 + b^2 = 1$$

根据我们的经验，这个事实非常有用，所以要牢记。

为了点与直线的垂直距离的和最小化，需要最小化

$$\sum_i (ax_i + by_i + c)^2$$

其中，$a^2 + b^2 = 1$ 并且 C 是某个无意义的标准化常数。因此，最大似然解可以通过最大化该表达式获得。用一个 Lagrange 乘子 λ，则会有一个解。如果

$$\begin{pmatrix} \overline{x^2} & \overline{xy} & \overline{x} \\ \overline{xy} & \overline{y^2} & \overline{y} \\ \overline{x} & \overline{y} & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \lambda \begin{pmatrix} 2a \\ 2b \\ 0 \end{pmatrix}$$

这意味着

$$c = -a\overline{x} - b\overline{y}$$

可以把这个替换回去，得到一个特征值问题：

$$\begin{pmatrix} \overline{x^2} - \overline{x}\,\overline{x} & \overline{xy} - \overline{x}\,\overline{y} \\ \overline{xy} - \overline{x}\,\overline{y} & \overline{y^2} - \overline{y}\,\overline{y} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \mu \begin{pmatrix} a \\ b \end{pmatrix}$$

因为这是个二维特征值问题，可以获得两个只到比例关系的封闭解（这一般只能用数值方法求解）。比例可以通过 $a^2 + b^2 = 1$ 的限制获得。这个问题的两个解是两条垂直的直线，其中一个可以使误差最小化。

15.2.2　哪些点在哪些线上

这个问题非常难，因为它涉及到搜索一个很大的组合空间。一种方法从不针对孤立点，而是从拟合边缘点出发，因此可以用边缘点的方向作为直线上下一个点的位置的暗示。如果考虑孤立点，那么可以采用 k 均值方法。

增量拟合——增量直线拟合（incremental line fitting）算法以那些边缘点形成的曲线为输入，然后用直线来拟合沿着曲线的点。边缘点所连接的曲线非常容易通过边缘检测得到，并有方向信息（见练习）。增量监测器在边缘点曲线的某一端开始，然后沿着曲线走，把拟合好的直线段截断（算法的结构在算法 15.1 中）。尽管缺少统计模型，增量直线拟合可以很好地工作，一个特点是它可以找到形成封闭曲线的直线段集合。当感兴趣的直线可以按照估计形成一个封闭曲线时（例如在某些物体识别中），这个算法便非常有用。因为它意味着算法可以很好地处理自然物体。这个算法同样会导向被遮挡的边并生成多于一条的拟合直线。这个问题可以通过后处理，寻找（大致）重合的直线对来解决，但是这个方法不怎么令人满意，因为很难给出一个准则，判断两条直线怎样才算重合。

用 *K*-均值方法将点分配给直线　假设点不包含任何它们属于哪条直线的信息（也就是说，没有颜色等信息可以利用，更严重的是那些点不是相连通的）。可以试图采用 k-均值的修改版本来确定哪些点在哪些线上。此时，采用的模型假设有 k 条线，每条线产生一些数据点子集；最佳的线与数据点的解可以通过对所有对应和线最小化下式得到：

$$\sum_{l_i \in \text{lines}} \sum_{x_j \in \text{data due to } i\text{th line}} \text{dist}(l_i, x_j)^2$$

这就有非常多的对应需要搜索。

可以非常简单地修改 k-均值法来处理这个问题。分两步：

- 将每个点分配给最近的直线
- 对于每条直线分配到的点，拟合最佳的直线

算法 15.1 增量直线拟合 通过沿着曲线走，对曲线上的点拟合直线，当残差足够大时截断曲线

```
把所有的点都放在曲线列表上，沿着曲线排序
清空直线点列表
清空直线列表
Until 曲线上点非常少
  将曲线上的最初几个点放入直线点列表
  对直线点列表拟合直线
  While 拟合的直线足够好
    把曲线上下一个点放入直线点列表然后重新拟合
  end
  把最后一个点放回曲线
  重新拟合直线
  把直线放入直线列表
end
```

这就得到了算法 15.2。可以通过查看直线的变化程度来判断收敛：标号是否已经不动（可能是最佳解），或者查看点到直线的垂直距离总和的变化。

算法 15.2 k-均值直线拟合，通过把点分配到最近的直线然后重新拟合

```
假设有 k 条直线（开始可能随机分配）
或者
假设某些点分配到某些直线，然后根据分配的点拟合出每一条直线
Until 收敛
  把每一点分配到最近的直线
  重新拟合直线
end
```

15.3 拟合曲线

原则上，拟合曲线与拟合直线类似。我们可以把点到曲线的平方距离和最小化，但是这产生了非常困难的实际问题：一般很难知道点到曲线的距离。我们或者解决这个问题，或者采用某种近似方法（一般常采用后者，因为计算较简单，而不是因为模型清晰）。我们对两种主要的曲线表示来勾画解决距离问题的一些方法。

15.3.1 隐曲线

隐曲线的坐标满足某个参数化的方程；如果方程是多项式的，那么曲线被称为代数的，这也是最普遍的情况。某些常见情况列在表 15.1 中。

表 15.1　某些视觉应用中用到的隐曲线。注意不是所有这些曲线都有点在上面。例如 $x^2+y^2+1=0$ 就没有。更高阶的曲线非常少用，因为很难得到这些曲线的稳定拟合

曲线	方程
直线	$ax+by+c=0$
圆，圆心(a,b)，半径 r	$x^2+y^2-2ax-2by+a^2+b^2-r^2=0$
椭圆(包括圆)	$ax^2+bxy+cy^2+dx+ey+f=0$ 其中，$b^2-4ac<0$
双曲线	$ax^2+bxy+cy^2+dx+ey+f=0$ 其中，$b^2-4ac>0$
抛物线	$ax^2+bxy+cy^2+dx+ey+f=0$ 其中，$b^2-4ac=0$
一般二次曲线	$ax^2+bxy+cy^2+dx+ey+f=0$

点到隐曲线的距离　现在我们要知道从一个数据点到隐曲线最近点的距离。假设曲线有 $\phi(x,y)=0$ 的形式。从隐曲线最近点到数据点的向量垂直于曲线，所以最近点可以通过找所有满足下式的(u,v)获得：

1. (u,v)是曲线上的一点，满足 $\phi(u,v)=0$；
2. $\boldsymbol{s}=(d_x,d_y)-(u,v)$正交于曲线。

给出所有 $\boldsymbol{s}$，其中最短的距离是数据点到曲线的距离。

第二个规则需要做点工作确定法线(正交)。一个隐曲线的法线是离开曲线最快的方向；在这个方向上，ϕ 必须变化非常快。这意味着(u,v)的法线是：

$$\left(\frac{\partial\phi}{\partial x},\frac{\partial\phi}{\partial y}\right)$$

如果曲线的切线是 $\boldsymbol{T}$，则有 $\boldsymbol{T}.\boldsymbol{s}=0$。因为在二维下，可以从法线确定切线，所以在点$(u,v)$上有：

$$\psi(u,v;d_x,d_y)=\frac{\partial\phi}{\partial y}(u,v)\{d_x-u\}-\frac{\partial\phi}{\partial x}(u,v)\{d_y-v\}=0$$

有两个方程和两个未知数，因此原则上可以解。然而，正如例 15.1 所示的那样，求解也并非那么简单。

例 15.1　点到二次曲线的距离

二次曲线可以由 $ax^2+bxy+cy^2+dx+ey+f=0$ 给出。给定数据点(d_x,d_y)，二次曲线上的最近点满足下面两个方程：

$$au^2+buv+cv^2+du+ev+f=0$$

和

$$2(a-c)uv-(2ad_y+e)u+(2cd_x+d)v+(ed_x-dd_y)=0$$

这两对方程最多有4个实数解(练习中,要证明这一点,给出一个获得这些解的算法,然后写出各种情况)。例如,选择椭圆 $2x^2+y^2-1=0$,产生方程

$$2u^2+v^2-1=0 \quad 和 \quad 2uv-4d_yu+2d_xv=0$$

让我们考虑一族数据点 $(d_x,d_y)=(0,\lambda)$;然后可以重新处理方程,得到:

$$2u^2+v^2-1=0 \quad 和 \quad 2uv-4\lambda u=2u(v-2\lambda)=0$$

第二个方程得到:$u=0$ 或 $v=2\lambda$。两个解是 $(0,1)$,$(0,-1)$。其他两个解可以通过解 $(2u^2+4\lambda^2-1=0)$ 得到 $-1/2\leqslant\lambda\leqslant1/2$。解的情况在图 15.6 中说明。

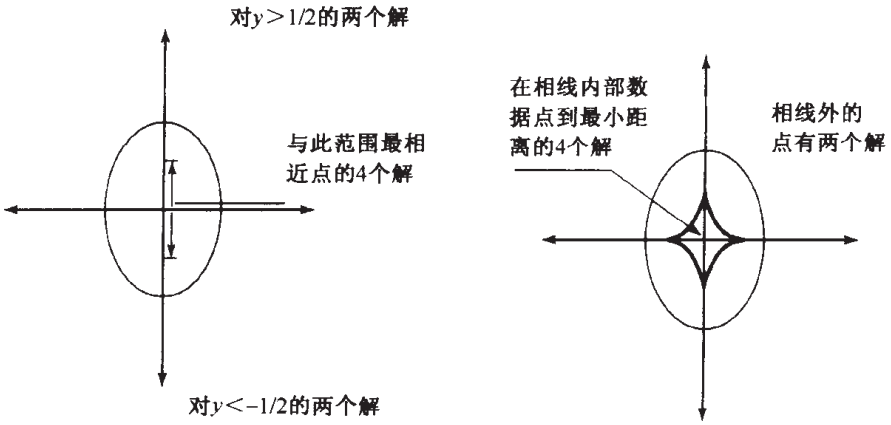


图 15.6 左图,在文中的例子,我们研究对于竖直轴上的数据点,一点与椭圆的距离的可能解的数目。右图表示对于椭圆的一般情况

近似距离 对一个相对简单的曲线,要解的问题已经比较麻烦。一个比较复杂的几何曲线[将 ϕ 选择为较高阶(d 阶)多项式而获得],将导致一个非常难解的问题。这是因为曲线上的最近点通过解两个多项式方程得到,两个都是 d 阶的。可以证明会有 d^2 个解,在实际中通常难以获得。因此实践中出现了不同的点到隐曲线距离的近似。

已知的最好近似是代数距离:此时,通过计算每一点的多项式方程获得点到曲线的距离,也就是说,我们做如下近似:

$$(d_x,d_y) \text{ 与 } \phi(x,y)=0 \text{ 的距离等于 } \phi(d_x,d_y)$$

当数据点与曲线相当接近时,这种近似可以(非常粗略的)加以证明。对于一个足够接近于曲线的点,近似到一阶时,当 (d_x,d_y) 正交于曲线移动时 $\phi(d_x,d_y)$ 增加,因为曲线的法线可以通过 ϕ 的梯度给出,而当 (d_x,d_y) 沿曲线切向移动时不增加。一个实际的困难是,代数距离定义的不好,因为许多多项式对应相同的曲线。特别地,像 $\mu\phi(x,y)=0$ 的曲线和 $\phi(x,y)=0$ 是一样的。这个问题可以通过用某种方式把多项式的系数规范化来解决。

在 15.2 节中已经看到这种方法的一个例子,拟合一条直线 $\phi(x,y)=ax+by+c=0$,使得满足约束 $a^2+b^2=1$ 的代数距离最小。在这里,代数距离和实际距离的意义一样。规范化的选择很重要。例如,如果要拟合二次曲线 $ax^2+bxy+cy^2+dx+ey+f=0$,而使用约束 $b=1$,我们就不能拟合圆。另一种近似是用

$$\frac{\phi(d_x,d_y)}{|\nabla\phi(d_x,d_y)|}$$

它的好处在于并不要求规范化常量。在直线的例子里，这个近似是精确的。注意这个近似有着和代数距离相同的属性，沿着法线方向值变大，等等。这种近似的好处是某种意义上它比代数距离更准确，因为它用法线的长度做了规范化。这表示可以(大致地)认为它给了一个距离的比例，沿着曲线到点的方向。实际中，这种近似很少用到，因为代数距离近似可以产生更简单的数值问题。

如果数据点远离曲线，两种近似都很危险，它们会表现得非常奇怪和不可理解。因此，如果数据点不是非常靠近曲线，拟合的曲线和数据点集的关系会变得有点奇怪。代数距离在实际中得到非常广泛的使用，因为它可以产生很简单的数值问题，并且用在更高维的问题上，比如估计点与隐曲面之间的距离。对于这种问题实际距离往往非常难以计算。

15.3.2　参数曲线

参数曲线的坐标通过有一个参数的参数函数给定，参数沿着曲线而变化。参数曲线有如下形式：

$$(x(t), y(t)) = (x(t;\theta), y(t;\theta)) \quad t \in [t_{\min}, t_{\max}]$$

表 15.2 给出了一些有用的参数曲线。

表 15.2　在视觉应用中经常用到一些参数曲线。普遍的做法是把一组三次曲线放在一起，对它们的系数加以限制，使其形成一个连续的微分曲线；这称为三次样条

曲线	参数形式	参数
圆心在原点的圆	$(r\sin(t), r\cos(t))$	$\theta = r$ $t \in [0, 2\pi)$
圆	$(r\sin(t) + a, r\cos(t) + b)$	$\theta = (r, a, b)$ $t \in [0, 2\pi)$
与轴对齐的椭圆	$(r_1\sin(t) + a, r_2\cos(t) + b)$	$\theta = (r_1, r_2, a, b)$ $t \in [0, 2\pi)$
椭圆	$(\cos\phi\,(r_1\sin(t) + a) - \sin\phi\,(r_2\cos(t) + b),$ $\sin\phi\,(r_1\sin(t) + a) + \cos\phi\,(r_2\cos(t) + b))$	$\theta = (r_1, r_2, a, b, \phi)$ $t \in [0, 2\pi)$
三次曲线	$(at^3 + bt^2 + ct + d, et^3 + ft^2 + gt + h)$	$\theta = (a, b, c, d, e, f, g, h)$ $t \in [0, 1]$

从点到参数曲线的距离　设有数据点(d_x, d_y)。在参数曲线上最近的点可以用参数值表示，写做 τ。这个点除了在曲线的一端或者另一端外，数据点到最近点的向量正交于曲线。这意味着 $s(\tau) = (d_x, d_y) - (x(\tau), y(\tau))$正交于切向量，使得 $s(\tau).\boldsymbol{T} = 0$。切向量是

$$\left(\frac{dx}{dt}(\tau), \frac{dy}{dt}(\tau)\right)$$

这意味着 τ 必须满足下面的方程：

$$\frac{dx}{dt}(\tau)\{d_x - x(\tau)\} + \frac{dy}{dt}(\tau)\{d_y - y(\tau)\} = 0$$

这是一个而不是两个方程，但是对参数曲线情况不会好很多。一般来说，$x(t)$和 $y(t)$都是多

项式，因为对于多项式求根要容易很多。如果 $x(t)$ 和 $y(t)$ 是多项式的比值，那么可以把方程的左端改变形式来满足这里的多项式。然而，我们同样面对可能的多根问题。

第二个困难使得拟合到参数曲线变得不普遍。不同系数的参数曲线可能代表同一条曲线，例如 $x(t), y(t), t \in [0,1]$ 和 $x(2t), y(2t), t \in [0,1/2]$ 是同一条曲线。这个情况可能很糟，与所用的参数曲线的类型有关。

15.4 作为概率问题的拟合

在本节之前，我们拟合模型的准则还比较随意。最小二乘总误差看起来像是合理的准则，但是准则(一定)必须依靠某种我们期望的误差模型上——为什么样本会不在线上呢？重新回到用直线拟合点集的问题，而该点集已知是从一直线产生的。分析结果表明最小二乘总误差(很自然的)是一个概率标准。我们从说明图像度量是如何得到的一个模型开始。

生成模型 假设我们的观测是沿着直线选择一些点产生的，然后使用高斯噪声垂直于直线做扰动。假设这个选择沿着直线的点的过程是均匀的，原则上，这不可能，因为直线是无限长，但实际中我们假设基本上是均匀的。这意味着有 k 个测量的序列 (x_i, y_i)，从以下模型获得：

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} + n \begin{pmatrix} a \\ b \end{pmatrix}$$

其中，$n \sim N(0,\sigma)$，$au + bv + c = 0$，$a^2 + b^2 = 1$。这个模型产生似然函数：

$$P(\text{测量} \mid a, b, c) = \prod_i P(x_i, y_i \mid a, b, c)$$

给定该模型，可以选择最大似然率或者最大后验直线。一般地，没有特别的原因选择某一条而放弃另一条直线，最大似然率是合适的。对数似然是：

$$-\frac{1}{2\sigma^2} \sum_i (ax_i + by_i + c)^2$$

满足 $a^2 + b^2 = 1$。似然率最大化导致点与直线的距离和最小化，就像 15.2.1 节所述的。在使用这个准则时有两个重要的现象需要处理：

- **鲁棒性**：最小二乘总误差准则对大误差赋予大权重。这可能是个严重的问题。例如，如果一个数据点离直线非常远，但是其他数据点对该直线拟合得很好(在后面会讨论一些产生这种情况的原因)，结果拟合的直线会因该数据点产生大偏差。这个现象会是一个严重的问题。例如，如果把一个数据集合拟合成基本矩阵，则需要左图和右图之间数据点的对应；如果得到一个错误的对应，则在数据集会存在潜在的大误差。我们将在后两节详细讨论这个问题。
- **数据丢失**：我们曾假设已知哪个点属于哪条直线；但是通常是不知道的。例如，有一个观测点的集合，有些来自直线而有些则是噪声。如果知道哪些点来自直线，那么确定直线就非常容易。同样，如果知道哪条直线是产生这些点的，那也会很容易确定哪些点来自直线。丢失的数据(什么是噪声而什么不是)是一个很关键的问题。大多数分割问题可以看做丢失数据问题；第 16 章的大部分内容将讨论这个问题。

15.5　鲁棒性

所有描述过的直线拟合方法都包括误差项的平方，这在实践中可能会导致很差的拟合，因为一个不适合的数据点会产生比其他数据点大得多的误差；这个误差会导致拟合过程的显著偏差(见图 15.7)。实践中很难避免这种数据点(一般叫做外点)。采集和抄录数据点中的误差是外点的一个重要来源。另一个很普遍的来源是模型的问题，可能忽视了某些罕见但却重要的效果，或者效果的程度被低估了。最后，对应中的误差也特别容易产生外点。实际的视觉问题经常包含外点。

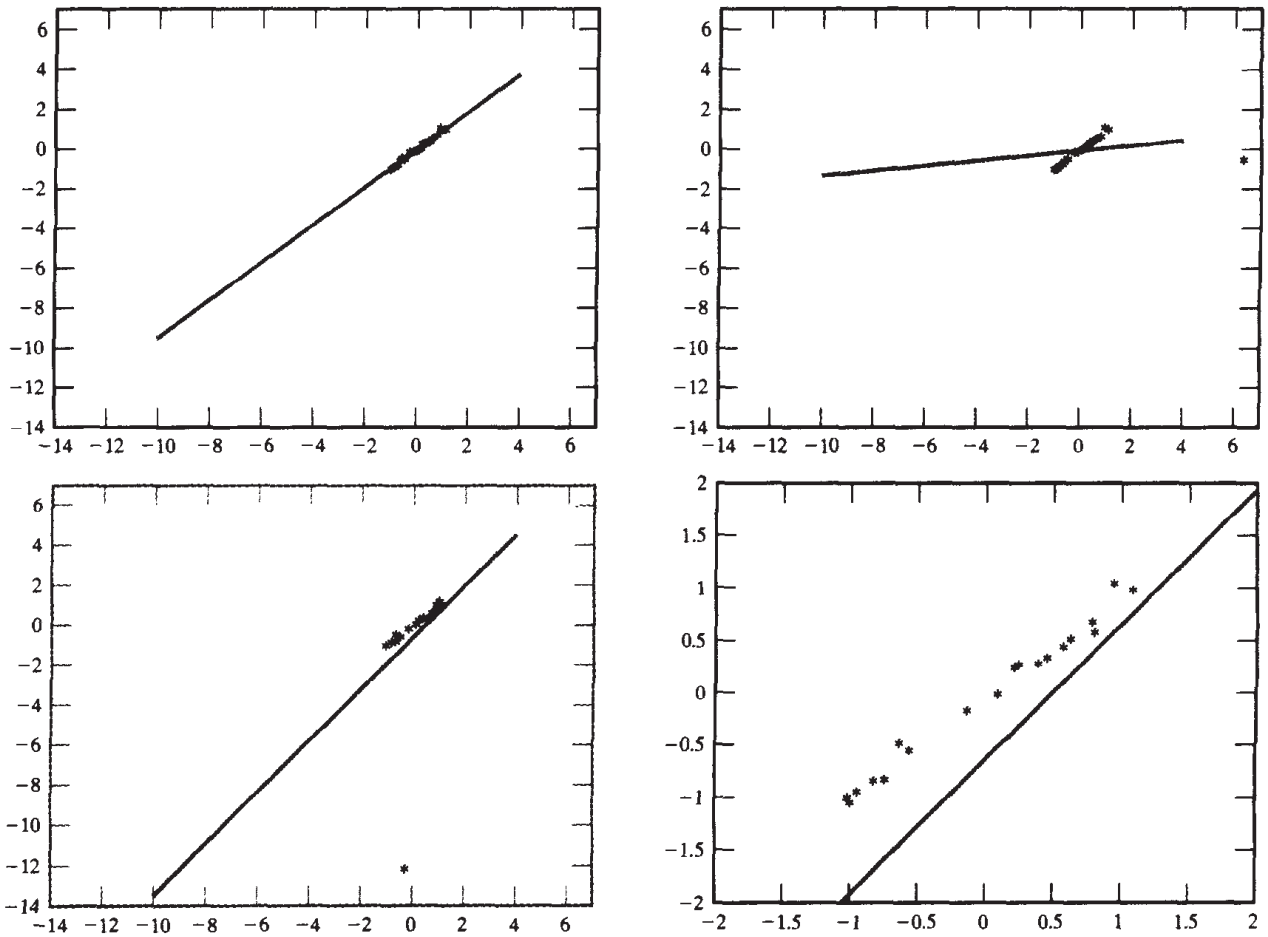


图 15.7　最小二乘直线拟合对外点非常敏感，不论是 x 坐标还是 y 坐标。左上图显示一个好的最小二乘拟合。右上图显示同样的点集，但是其中某个 x 坐标发生错误。此时，拟合直线的斜率明显偏转。左下图显示同样的点集，其中有个点的 y 坐标出现错误，造成直线的 x 截距改变了。这三幅图都是同一个数据集作为对比，但是这种坐标轴的选择没有在第三个例子中清楚地反映出拟合很差的程度。右下图显示了它的细节，可以看到拟合得相当差

处理这个问题的一个方法就是宣称这个模型是有缺陷的：模型预测外点不可能出现，而它们确实是经常出现的。很显然，增强模型的方法就是或者将噪声的影响削弱(见 15.5.1 节)，或者允许一个显式表示外点的模型。第二个策略需要研究丢失数据问题(我们不知道哪个点是外点而哪个点不是)，因此安排到 16.2.4 节讨论这个问题。另一个替代方法是搜索看上去好的点(见 15.5.2 节)。

15.5.1 M 估计

对外点的来源建模的困难在于模型可能是错误的。一般我们期待一个过程的概率模型相当接近正确的模型。假设能够保证过程模型非常接近于正确的模型，比如，密度函数之间的距离在某种意义下小于ϵ。可以用这个保证来推理关于模型参数估计过程的设计。特别地，可以选择一种估计过程，假设本质是复杂的，但从统计学上提供了信息。在这里的推理中，通过假设一个过程集合中接近于过程模型的就是真实过程，来评判一个估计的好坏，而这往往使得估计产生最坏的效果。最好的估计是在接近参数模型的最坏分布下还可以表现得非常好的估计。这是一个可以用来产生很多估计方法的准则。

一个 M 估计通过最小化下面的表达式来估计参数

$$\sum_i \rho(r_i(\boldsymbol{x}_i, \theta); \sigma)$$

其中，θ 是拟合模型的参数，$r_i(\boldsymbol{x}_i, \theta)$是模型第 i 个点的残留误差。一般地，$\rho(u;\sigma)$ 在 u 值较小时看上去像函数 u^2，当 u 增大时函数变平。一个常用的选择是：

$$\rho(u;\sigma) = \frac{u^2}{\sigma^2 + u^2}$$

σ 的参数控制函数变平的点；图 15.8 中画了一些例子。M 估计还有很多别的方法。一般地，按照它们的影响函数讨论，影响函数定义如下：

$$\frac{\partial \rho}{\partial \theta}$$

这很自然，因为我们的准则是：

$$\sum_i \rho(r_i(\boldsymbol{x}_i, \theta); \sigma) \frac{\partial \rho}{\partial \theta} = 0$$

作为所考虑的这类问题，我们希望一个好的影响函数是反对称的(略为过估计和欠估计关系不大)，并将大数值衰减掉，因为需要限制外点的影响。

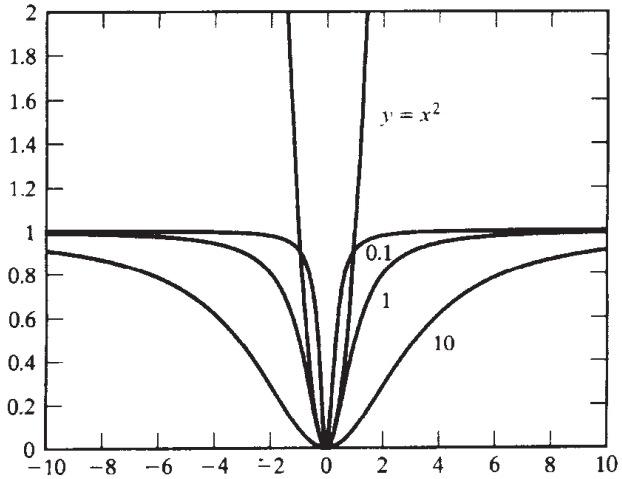


图 15.8 函数 $\rho(x;\sigma) = x^2/(\sigma^2 + x^2)$，用 $\sigma^2 = 0.1, 1, 10$ 画的，另外用 $y = x^2$ 作为对比。用 ρ 代替二次函数以降低外点对拟合的影响，一个离拟合曲线σ的数倍远的点，对拟合曲线的系数基本上不产生任何影响，因为ρ的值接近于1，且随与拟合曲线的距离变化非常缓慢

使用 M 估计有两种技巧性问题。第一，极值问题是非线性的，必须迭代求解。典型的困难是：有不止一个局部最小点，方法可能发散，方法的表现很可能对起始点敏感。对该问题的一个常用的策略是对数据集进行子采样，用最小二乘对其进行拟合，然后用这个作为拟合过程的起始点。用很多次不同的子采样，以求有很高的概率，使得子采样里至少包含一个全是好的数据点组成集合。

第二点，就像图 15.9 和图 15.10 所示，估计方法需要一个对 σ 的好的估计，其中 σ 叫做尺度(scale)。典型地，尺度估计在迭代求解过程的每一步都要使用；一个流行的尺度估计是：

$$\sigma^{(n)} = 1.4826\ \mathrm{median}_i\ |r_i^{(n)}(x_i;\theta^{(n-1)})|$$

算法 15.3　使用 M 估计来拟合一个概率模型

对于 s 从 1 到 k

　均匀随机地提取 r 个不同点组成的子集

　用最大似然(通常是最小二乘)对点集拟合获得 θ_S^0

　用 θ_s^0 来估计 σ_s^0

　直到收敛($|\theta_s^n - \theta_s^{n-1}|$非常小)

　　用最小化方法通过 $\theta_s^{n-1}, \sigma_s^{n-1}$ 得到 θ_s^n

　　计算 σ_s^n

　end

end

使用残差的中值作为准则获得这个集合的最好拟合

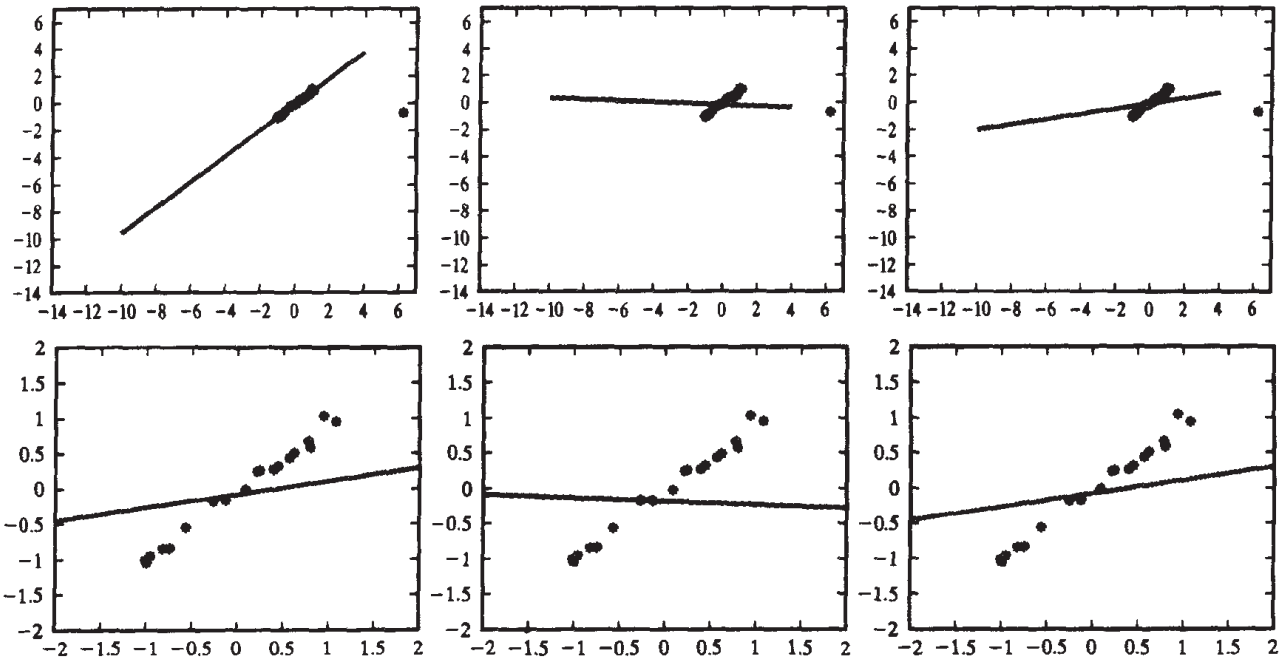


图 15.9　上面的图显示对图 15.7 第二组数据，用一种加权函数把远处的点的贡献变小(图 15.8 的 ϕ 函数)的方法拟合的直线。左图，σ 取正确值；外点的贡献被降低，拟合得很好。中图，σ 的值太小，所以拟合对所有数据点都不敏感，意味着它和数据的关系变得模糊。右图，σ 的值太大，导致外点的贡献还是很大。下面的图显示了同样例子拟合直线和非外点数据的细节

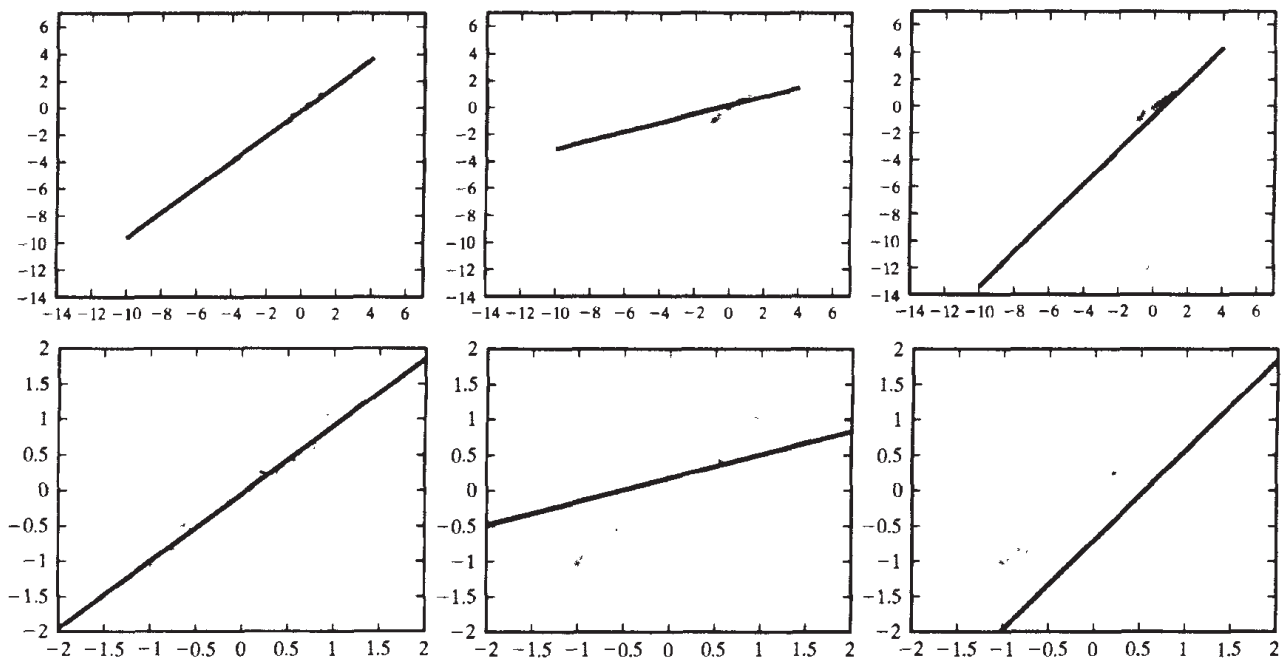


图 15.10 上面的图显示了对图 15.7 第三组数据，采用加权函数弱化远点贡献(图 15.8 的 ϕ 函数)方法的拟合情况。左图，σ 取正确值；外点的贡献被降低，拟合得很好。中图，σ 的值太小，所以拟合对所有数据点都不敏感，意味着它和数据的关系变得模糊。右图，σ 的值太大，导致外点的贡献还是很大。下面的图显示了同样例子拟合直线和非外点数据的细节

一个 M 估计可以认为是一种策略，其可以保证外点作用被衰减的概率要比它们产生平方误差的概率要高。最小化的函数对 $\boldsymbol{x}$ 取小值时看上去像距离，因此，对于合法的数据点，M 估计的表现像最大似然率估计，对 $\boldsymbol{x}$ 取大值时看上去像一个常量，意味着概率的一部分被衰减了。前一节的策略可以看做 M 估计，但是它的问题是影响函数是不连续的，因此获得最小值很麻烦。

15.5.2 RANSAC

一种修改产生模型使之有很强衰减能力的方法，是在数据点集中搜索好的点。通过迭代过程很容易做到：首先，选择一个小的数据点子集，然后对其进行拟合，查看有多少其他点可以匹配到拟合的物体上。继续这个过程直到有较大的概率找到我们想要找的结构。

例如，假设对一个包含 50% 外点的数据集拟合直线。如果把点集均匀随机地配对，那么约有四分之一的点对由好点组成。可以分辨这些好点对，因为其他很多点都会非常靠近由好的点对所拟合出来的直线。当然，可以通过对这些靠近当前直线的点进行拟合得到更好的直线估计。

这个方法导致一个算法：搜索一个随机采样，对其进行拟合，可以得到许多数据点支持。这个算法一般叫做 RANSAC，也就是随机采样一致算法(random sample consensus)，在算法 15.4 中表示。为了让该算法可行，必须选择三个参数。

算法 15.4　RANSAC:用随机采样一致拟合直线

确定:

　n——所需要的最少点数

　k——需要的迭代次数

　t——用来判断一个点是否拟合得很好的阈值

　d——判断一个点是否拟合得很好所需要的邻近点数目

直到 k 次迭代完成

　从数据中均匀随机的采样 n 个点

　对这 n 个点进行拟合

　对于在采样外的每一个点

　　用 t 比较点到直线的距离;如果距离小于 t,那么点是很靠近的

　end

　如果有 d 或更多个点靠近直线,那么是一个好的拟合。重新用这些点拟合直线

end

使用拟合误差作为准则,挑出最好的拟合

需要采样的数目　我们的采样由从数据集中均匀和随机地取出的点集组成。每个采样包含拟合所需最少数目的数据点。例如,拟合直线,抽取点对;拟合圆,抽取三个点,以此类推。假设需要抽取的最小子集有 n 个数据点,w 是这些点中好点的比例(只需要合理估计这个数)。现在获得一个好的采样所需要抽取的数目 k 的期望值如下:

$$
\begin{aligned}
E[k] &= 1P\text{一个抽取中的一个好的采样} + 2P(\text{一个抽取中的两个好的采样}) + \cdots \\
&= w^n + 2(1-w^n)w^n + 3(1-w^n)^2 w^n + \cdots \\
&= w^{-n}
\end{aligned}
$$

其中最后一步略微使用了代数学序列的处理。我们能充分肯定我们有了好的采样,所以需要抽取多于 w^{-n} 个样本;一个很自然的想法是对这个数加入一些标准偏差。k 的标准偏差可以由下式获得:

$$SD(k) = \frac{\sqrt{1-w^n}}{w^n}$$

对这个问题的另一种看法是,所有采样都是坏样本的概率为 z 时的样本数目。此时,有:

$$(1-w^n)^k = z$$

这意味着:

$$k = \frac{\log(z)}{\log(1-w^n)}$$

w 未知的情况是非常普遍的。然而,每一次拟合都包含了关于 w 的信息。特别地,如果需要 n 个数据点,可以假设成功拟合的概率是 w^n。如果观测到很长的拟合尝试序列,则可以从序列中估计 w。这意味着可以从一个比较小的 w 估计开始,产生一系列尝试性的拟合,然后改进对 w 的估计。如果拟合尝试次数已超过新的 w 估计计算的拟合尝试次数,这个过程就可以

停止。更新 w 估计的问题,变成给定一序列拟合条件下估计一个硬币投出去是朝上还是朝下的概率问题。

判断一点是否靠近拟合模型 需要判断数据点是否靠近根据采样拟合出的直线。这需要确定点与拟合出直线的距离,然后用一个阈值 d 来测试这个距离;如果距离小于阈值,则认为这个点是靠近的。一般地,确定这个参数同样也是建模过程的一部分。例如,如果用最大似然率来拟合直线,在模型中有一项 σ(在寻找最大值的处理中这一项会消失)。这一项给出了拟合模型的平均偏差的尺寸。

一般地,获得这个参数的值要相对简单。我们一般仅需要一个类似大小的估计,然后在许多不同的实验中用相同的值。决定参数可以通过尝试一些值然后查看结果决定;另一个方法是考查一些特征数据集,然后用肉眼拟合直线,最后估计平均偏差。

一致的点的数目 假设对于某个随机采样的两点数据拟合直线,需要知道直线是否是正确的。可以计算有多少点在直线的某个距离内(距离在前一章里确定)。特别地,假设知道点集里外点的概率;把这个概率写做 y。需要选择 t 个点使得 y^t 很小(比如小于 0.05)。

有两种方法可以处理。一个是根据 $y \leqslant (1-w)$,选择 t 使得 $(1-w)^t$ 很小。另一种是从某种外点模型中得到 y 的某个估计。例如,如果所有的点都在一个单位正方形中,外点均匀分布,而且距离阈值是 d,那么 $y \leqslant 2\sqrt{2}d$。

15.6 举例:用 RANSAC 来拟合基础矩阵

由三维空间的点产生两种观测,一个是左视图,另一个是右视图。把三维点的实际坐标写成 $\boldsymbol{X}_i$,在左(右)图像的坐标是 $\boldsymbol{x}_{li}(\boldsymbol{x}_{ri})$,在左(右)图像的观测坐标是 $\boldsymbol{m}_{ri}(\boldsymbol{m}_{ri})$。基础矩阵是外极线约束的一种表示。特别地,用带帽的符号表示使用齐次坐标系,对于所有的点有 $\hat{\boldsymbol{x}}_{ri}^T \mathcal{F} \hat{\boldsymbol{x}}_{li} = 0$,这里的 $\mathcal{F}$ 表示基础矩阵。

15.6.1 一种拟合误差的表示

对观测值约束一般并不严格成立,假设观测受到了有均匀旋转对称协方差的累加性高斯噪声影响。把 $\boldsymbol{m}_{li}$ 和 $\boldsymbol{m}_{ri}$ 也写成仿射(常规的,或非齐次)坐标。这意味着

$$P(\boldsymbol{m}_{li}, \boldsymbol{m}_{ri} | \boldsymbol{x}_{ri}, \boldsymbol{x}_{li}, \mathcal{F}) \propto \exp -\frac{1}{2\sigma^2} \left\{ \begin{array}{c} (\boldsymbol{x}_{ri} - \boldsymbol{m}_{ri})^T (\boldsymbol{x}_{ri} - \boldsymbol{m}_{ri}) + \\ (\boldsymbol{x}_{li} - \boldsymbol{m}_{li})^T (\boldsymbol{x}_{li} - \boldsymbol{m}_{li}) \end{array} \right\}$$

现在这是一个数据的复杂函数,有参数 $\mathcal{F}, x_{ri}, y_{ri}, x_{li}$。然而,有足够的数据点,原则上可以获得一个极值,但是目前还不能这样做,因为不知道左图像与右图像的对应。此外,我们注意到对数似然率的平方和的形式会导致鲁棒性问题。

15.6.2 对应点产生的噪声

一种处理对应的问题是假设在两个视图之间仅有小的镜头运动,因此可以假定在第二视图的特征点位置与它们相应的特征点在第一视图的位置非常接近。这是一个危险的假设;它可能导致很差的拟合,但是看上去还好。问题是一个对应误差起外点的作用。另一种策略是搜索一组与好的基础矩阵一致的对应。

如果对图像每个点附上其图像近邻的某种表示,则可以简化搜索。这意味着我们有每个

点的更多细节的描述。例如，可能对每一个点计算某一尺度范围变化的滤波器的输出集。我们可能期待一个匹配点的一个局部图像近邻，与它原点的局部图像近邻差别不会很大，因此滤波器输出集不会变化很快。这个准则将得到一组可能的对应集合，其中可能包含误差，这些误差与外点的作用相似。现在对这个假定的对应集合应用 RANSAC。

15.6.3　应用 RANSAC

下面将看到，7 点对应产生一个基础矩阵。用这个信息和一些技巧，对可能的对应应用 RANSAC 相对简单。尽管我们可能不知道对应中正确的比例，但可能通过使用拟合尝试来估计它（见前文）。确定点是否为内点的距离阈值一般是像素级（这必须依赖于摄像机的质量）。

从 7 点获得基础矩阵　设有 7 个对应假设。假设在每一点的观测值和该点的真实值一样。如果 $\boldsymbol{x}_{ri}, \boldsymbol{x}_{li}$ 已知，每个限制 $\hat{\boldsymbol{x}}_{ri}^T \mathcal{F} \hat{\boldsymbol{x}}_{li} = 0$ 产生一个基础矩阵的系数的线性方程。此外，方程 $\hat{\boldsymbol{x}}_{ri}^T \mathcal{F} \hat{\boldsymbol{x}}_{li} = 0$ 对基础矩阵的元素是齐次的。这就是说，如果 $\mathcal{U}$ 满足约束，那么 $\lambda\mathcal{U}$ 也满足。最后，回想第 10 章曾提到的，基础矩阵的秩是 2，所以 $\det(\mathcal{F}) = 0$。

这意味着仅需要 7 点就可以估计 $\mathcal{F}$。每个点产生一个 $\mathcal{F}$ 的元素的齐次方程。7 个齐次方程的解是一个两维线性空间，对于已知的 $\mathcal{F}_0$、$\mathcal{F}_1$ 和任意 λ，μ，可以写成 $\lambda\mathcal{F}_0 + \mu\mathcal{F}_1$。但是我们需要空间中一个行列式为 0 的元素，并且方程 $\det(\lambda\mathcal{F}_0 + \mu\mathcal{F}_1) = 0$ 是一个 λ，μ 的齐次三次方程。可以把两边都除以 μ，然后对 λ/μ 求解。这个方程或者有一个实根，或者有三个实根。

比较基础矩阵和其他点　现在设有一个基础矩阵的估计，需要知道是否一个特殊的假设对应与估计一致。对于这个对应，我们知道观测点 $\boldsymbol{m}_{ri}, \boldsymbol{m}_{li}$ 和某个基础矩阵的估计 $\mathcal{F}_0$。我们不知道产生观测的实际点 $\boldsymbol{x}_{ri}, \boldsymbol{x}_{li}$。然而，有：

$$P(\boldsymbol{m}_{li}, \boldsymbol{m}_{ri} | \boldsymbol{x}_{ri}, \boldsymbol{x}_{li}, \mathcal{F}) \propto \exp -\frac{1}{2\sigma^2} \left\{ \begin{array}{c} (\boldsymbol{x}_{ri} - \boldsymbol{m}_{ri})^T (\boldsymbol{x}_{ri} - \boldsymbol{m}_{ri}) + \\ (\boldsymbol{x}_{li} - \boldsymbol{m}_{li})^T (\boldsymbol{x}_{li} - \boldsymbol{m}_{li}) \end{array} \right\}$$

如果对基础矩阵的估计是好的，而且如果对应是好的，那么有一些实际点会与我们的观测很接近，并且与基础矩阵的估计一致。这意味着需要知道最近的这些点和观测的距离，如果距离很小，那么对应就是正确的。假设能够获得这个距离（后面会给出一个很好的技巧），那么通过 RANSAC 获得基础矩阵就非常简单。我们把各个部分合并到算法 15.5 中。

算法 15.5　RANSAC：采用随机采样一致算法拟合基础矩阵

确定：

　　最小的对应点对数目是 7

　　k——需要迭代的次数

　　t——用来确定点是否拟合好的阈值

　　d——确定模型拟合正确需要的附近点数目

直到 k 次迭代完成

　　从数据中均匀随机的采样 7 个点对应

（未完待续）

（续）

对这 7 个点对应拟合基础矩阵$\mathcal{F}_0$

对于在采样外的每一个点对应

用 t 检测与$\mathcal{F}_0$ 一致的最近点到观测点的距离；如果距离小于 t，那么对应是一致的

end

如果有 d 或更多的一致对应，那么是一个好的拟合。用所有这些点重新拟合基础矩阵的估计

end

使用拟合误差作为标准，挑出最好的拟合

15.6.4 获得距离

我们有一对观测点 $\boldsymbol{m}_{li}$ 和 $\boldsymbol{m}_{ri}$，希望得到这些点和与基础矩阵的估计一致的最近的点对之间的距离。这意味着最近点对必须有性质 $\boldsymbol{x}_l^T\mathcal{F}_0\boldsymbol{x}_n = 0$ 。对于这个问题我们勾画一个更好的解法，见 Hartley 和 Sturm(1997)。

注意，首先对左图和右图的图像坐标系同时进行旋转和平移不会影响距离。我们忽略最好的点对在外极点的可能性，因为这决不可能发生。我们可以把两个坐标系统都平移，使得外极点就在原点。现在右外极点有属性$\mathcal{F}\hat{\boldsymbol{e}}_r = \mathbf{0}$，左外极点也有 $\hat{\boldsymbol{e}}_l^T\mathcal{F} = \mathbf{0}^T$。这意味着，在一个左图和右图的外极点都在原点（旋转平移就可以做到，除非它们在无穷远，而这是我们所忽略的）的坐标系里：

$$\mathcal{F} = \begin{pmatrix} \mathcal{M} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

这是一个 3×3 的矩阵，因为$\mathcal{M}$是个 2×2 的块。现在如果我们对坐标系做关于原点的旋转会发生什么呢？写下 $\boldsymbol{u}_{ri} = C_{ri}\boldsymbol{x}_n$，其中

$$\mathcal{C}_{ri} = \begin{pmatrix} \mathcal{R}_{ri} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

而且，$\mathcal{R}_{ri}$是一个平面旋转（当讨论左图时，用同样的旋转，但是下标用 li），则在 $\boldsymbol{u}$（旋转过的）坐标系里，新的基础矩阵有这样的形式：

$$\mathcal{F}_0' = \mathcal{C}_{ri}^T\mathcal{F}_0\mathcal{C}_{li} = \begin{pmatrix} \mathcal{R}_{ri}^T\mathcal{M}\mathcal{R}_{li} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

这表示，通过选择合理的旋转，可以使基础矩阵对角化。

现在让我们考虑距离问题。左图里从观测点到实际点的位置和左图里从观测点到某个外极线的距离相同。这个外极线被转换成右图中的某个外极线，通过基础矩阵可以指出是哪一个。右图里从观测点到实际点的位置和右图里从观测点到这个外极线的距离相同。要做的是使得距离总和最小化。

在每个图中外极点都在原点。在左图，外极线是线族$(s, t, 0)$（表示 $sx + ty = 0$，所有穿过原点的直线，s 和 t 的改变使直线变化）。在右图里，因为我们旋转坐标系使得基础矩阵对角

化，对于某个依赖于基础矩阵的 λ，对应直线有 $(s, \lambda t, 0)$ 的形式（表示 $sx + \lambda ty = 0$）。对于 m_{xri} 的 x 坐标写出 $\boldsymbol{m}_{ri}$，现在距离是：

$$\frac{(sm_{xli} + tm_{yli})^2}{(s^2 + t^2)} + \frac{(sm_{xri} + t\lambda m_{yri})^2}{(s^2 + \lambda^2 t^2)}$$

那么我们需要把这当做 s 和 t 的函数最小化。事实上，这对于 s 和 t 是齐次的，所以需要把它考虑成一个 $u = s/t$ 的函数。我们最大化这个函数，然后考虑它作为 $v = t/s$。如果在 $t = 0$ 有最大值，最大化很简单，所以仅仅考虑 u。有一个 u 的有理函数，它的分子和分母都是 4 阶。这一定和 u 的某个有理函数相同，它的分子是三阶，分母是 4 阶，另外加一个常数（我们可以做长除法获得）。现在这个表达式的导数的分子会消失，是一个 6 阶的 u 的多项式。我们算出根，然后就可以解了。

15.6.5　对已知对应拟合基础矩阵

现在有一个已知对应的系统，想要拟合一个基础矩阵。我们希望获得对于 $\mathcal{F}$ 的最大似然解；已经有了似然率对数的表达式，但是用了未知的点集 $\boldsymbol{x}_{ri}, \boldsymbol{x}_{li}$。可以试图排除这些点（看上去很难），或者把它们解出来。

合适的策略是把它们作为三维的点解出。可以把一个摄像机固定在原点，第二个摄像机只是取基础矩阵的函数的标准形式。可以把负对数似然率作为基础矩阵和点的三维构造的函数最小化。对于后者，有一个估计作为开始点，因为知道了对应和基础矩阵。

15.7　注释

我们从很多的技术中仅仅挑出了一些重要的内容。拟合是一个经常出现的问题；几乎总能把一个问题看成拟合问题，并且经常是有帮助的。这也意味着难以提供一个非常有用的文献导读。

通常，在实际中遇到的主要困难有：(a)确定距离（实际上是非常难的）；(b)保证外点不会太多；(c)决定首先拟合什么。

很多论文都曾讨论过从一个点到曲线或曲面的距离近似：我们特别推荐 Agin(1981)；Bookstein(1979)；Cabrera 和 Meer(1996)；Porrill(1990)；Sabin(1994)；Sullivan，Sandford 和 Ponce (1994*a*，*b*)，Sullivan 和 Ponce(1998)，以及 Taubin，Cukierman，Sullivan，Ponce 和 Kriegman (1994*a*，*b*)。

鲁棒性在视觉领域的实际中所具有的影响并不像想像中应该有的那样大。较好的入门阅读材料包括 Huber(1981)；Meer，Mintz，Kim 和 Rosenfeld(1991)；Rousseeuw(1987)，以及 Stewart (1999)。这些观点除了让人很容易激动外，也确实非常有用。我们并没有深入这个题目，其主要原因在于，对这个题目的粗浅了解就已足够应付实际中的问题了。

使用RANSAC计算基础矩阵的例子，因为这是目前计算基础矩阵最好的方法。对RANSAC感兴趣的读者可以参考由运动引起的结构，特别是Hartley 和 Zisserman(2000)对算法有着很大影响的论述。我们同样推荐 Fischler 和 Bolles(1981) 以及 Torr 和 Murray(1997)。我们期望今后它可以用来形成类似 EM 的方法；后面我们会更多讨论 EM。

实际上没有强调拟合当做推理来处理。把拟合考虑成推理的好处是,人们所知道的有关拟合的任何方面都可以变成拟合的知识,越多越好。下一章将讲到这些,但是应该意识到我们在网络上关于概率章节的评论在这一章同样可以用在几乎所有方面。特别地,可以使用 MAP 推理,要做的只是对拟合误差附加一个先验项。早几年里,构造一个先验项对摆动太厉害的模型进行惩罚曾是一种流行做法。这个方法非常有用但是近几年似乎消失了。文献的例子包括 Horn 和 Schunck(1981); Bertero, Poggio 和 Torre(1988), 以及 Poggio, Torre 和 Koch(1985)。

拟合同样可以用来图像重构。把图像数据拟合成曲面(被解释为一个高度的地图),在边缘撕裂曲面。因此正确评价撕碎的代价和差的拟合的代价就很重要。这种一般性策略还可用到深度图重构、立体视觉图等。还有很多种约束可用。在Grimson(1981*b*)中,没发现特征意味着平滑(没有消息便是好消息)。最近更多的工作考虑不同形式的不连续性,包括撕裂和折缝(Blake 和 Zisserman, 1987; Mumford 和 Shah, 1985 和 1988)。

习题

15.1 证明一个简单但有用的结果:如果 $a^2+b^2=1$,则点(u,v)到一条直线(a,b,c)的垂直距离是$|(au+bv+c)|$。

15.2 从最小二乘总误差的原始模型获得特征值问题:

$$\begin{pmatrix} \overline{x^2}-\bar{x}\,\bar{x} & \overline{xy}-\bar{x}\,\bar{y} \\ \overline{xy}-\bar{x}\,\bar{y} & \overline{y^2}-\bar{y}\,\bar{y} \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\mu\begin{pmatrix} a \\ b \end{pmatrix}$$

这是个简单的练习,用最大似然率然后做一些处理就可以解,但是要保证正确并且记住,这个技术非常有用。

15.3 如何从返回方向的边缘检测器中获得边缘点的曲线?给出一个递归算法。

15.4 增量拟合的一个更稳定的变种,删除直线点列表的开始几个像素和最后几个像素,因为这些点可能来自角点。

(a) 为什么该方法可以获得改进?

(b) 怎样决定要忽略几个点?

15.5 一个二次截面由 $ax^2+bxy+cy^2+dx+ey+f=0$ 给出。

(a) 给定数据点(d_x,d_y),证明二次曲线上的最近的点(u,v)满足两个方程:

$$au^2+buv+cv^2+du+ev+f=0$$

和

$$b(u^2+v^2)+2(a-c)uv-(2a\,d_y+e)u+(2c\,d_x+d)v+(e\,d_x-d\,d_y)=0$$

(b) 有两个二次方程,将向量$(u,v,1)$写成 $\boldsymbol{u}$。证明可以把这些方程写成 $\boldsymbol{u}^T\mathcal{M}_1\boldsymbol{u}=0$ 和 $\boldsymbol{u}^T\mathcal{M}_2\boldsymbol{u}=0$,其中$\mathcal{M}_1$ 和$\mathcal{M}_2$ 是对称矩阵。

(c) 证明存在变换$\mathcal{T}$,使得$\mathcal{T}^T\mathcal{M}_1\mathcal{T}=Id$ 和$\mathcal{T}^T\mathcal{M}_2\mathcal{T}$是对角阵。

(d) 如何使用这个变换来获得方程的解集;特别地,证明可能有 4 个实数解。

(e) 证明这个方程可能有 4 个、两个或 0 个实数解。

(f) 画一个椭圆,然后指出有 4 个或两个解的点。

15.6 证明曲线

$$\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)$$

是一个圆弧(圆弧的长度依赖于参数所定义的区间)。

(a) 写出一个 t 的方程对于某个最近数据点(d_x, d_y)的最短距离。这个方程有几阶? t 可能有多少个解?

(b) 现在用 $s^3 = t$ 替换参数方程,写出对于同样数据点的最短距离点的方程。这个方程有几阶? 为什么那么高? 可以得出什么结论?

15.7 证明:对于一个圆锥体的观测圆锥体是一族平面,所有平面都穿过焦点和圆锥体的顶点。现在证明圆锥的外形由穿过顶点的直线集组成。可以用一些不需要计算的简单结论来证明。

编程作业

15.8 实现一个增量直线拟合。如果把直线点列表的头几个点和末尾几个点去掉,确定会有什么不同的结果(细致地写这个程序,因为它是实际中使用软件的一个非常有用的部分)。

15.9 实现一个哈夫直线检测。

15.10 用 HT 直线检测计算线的数目,它的效果如何?

第 16 章　使用随机方法的分割与拟合

前面各章所描述的所有分割算法本质上都涉及了局部的相似性模型。尽管某些算法尝试建立具有较好全局性的聚类，但是所用的相似性模型都是对各个独立的像素进行比较。此外，在这些算法中都没有明确地表示测量值与希望得到的值的差别的概率模型。

本章讨论用于分割的概率模型。这些方法力图使用全局模型来解释数据，并且希望能够用较少的参数去解释大量的数据。例如，用一条线去拟合一个样本集，或者尝试拟合出一对图像之间运动向量的参数集，这些参数表示的是像素如何从一幅图像运动到另一幅图像。

16.1　丢失数据问题、拟合和分割

许多重要的视觉问题都可表达为数据中丢失了有用元素的问题。例如，可以将分割问题看做确定所测到的数据分别来自于哪些数据源的问题。通常有下面一些观点：将一幅图像分割为各个区域的过程，需要确定由颜色和纹理特征所表示的像素所形成的数据源都能够生成图像中的哪些像素；将所获得的样本集分割为共线性组的过程，需要确定哪个样本在哪条线上；另外将一个运动序列分割为各个运动区域的过程，需要按照运动模型来分组运动像素。如果我们能够获取那些当前缺少的数据，那么上面所描述的这些问题将会变得比较简单（对应于上面三个问题，丢失的数据分别为：像素属于哪个区域，样本属于哪条线，像素属于哪个运动模型）。

16.1.1　丢失数据问题

丢失数据问题是一个丢失了某些数据的统计问题。在下面两种情况下丢失的数据是很重要的：第一种情况，某个数据向量中的某些项丢掉了，但这些项在其他数据向量中还存在（可能有些人在调查某些问题时通常会因为遇到这样的情况而为难）。第二种在应用中更为常见的情况是，通过使用一些值未知的变量重新去描述推理问题，使得这个推理问题更加简单。幸好，目前已有一个有效的算法可以处理丢失数据问题，该方法本质上是对丢失的数据进行预测。下面我们通过两个例子来表明这种方法和对应算法的有效性。

例 16.1　图像分割

对于图像中的每个像素，计算一个 d 维的特征向量 $\boldsymbol{x}$，$\boldsymbol{x}$ 中包括位置、颜色和纹理信息。特征向量可以包含各种不同的颜色表示，以及以某一特定像素点为中心的一系列滤波器的输出。我们将图像模型表示如下：图像中的每个像素均是由 g 个图像分割中的某一个的密度函数计算得到的。因此，为了产生一个像素，首先选择一个图像分割区域，然后通过该区域的密度函数生成所需的像素。

假设选择第 l 个分割的概率为 π_l，并且用高斯函数来建模与第 l 个区域对应的密度，模型参数为 $\theta_l=(\boldsymbol{\mu}_l,\Sigma_l)$。因此生成一个像素向量的概率可表示为：

$$p(\boldsymbol{x})=\sum_l p(\boldsymbol{x}\mid\theta_l)\pi_l$$

这种形式的模型通常称为混合模型(由于它是一个概率模型的加权和或混合;π_1 通常被称为混合权重)。在后面的讨论中,我们会经常遇到这种表达形式。

一种解释方法是把混合模型看做生成模型。基于这种观点,图像中的每个像素可通过下面的方式获取:(a)以概率 π_l 选择模型中的第 l 个分量,(b)由公式 $p(\boldsymbol{x}|\theta_l)$ 生成一个样本。可在向量空间中以密度的形式将这个模型可视化,该空间包括 g 个"块",每个块与一个图像分割相对应。我们希望能够确定以下参数:(a)每一个块的参数,(b)混合权重,(c)各个像素来源于模型中的哪个分量(从而实现对图像的分割)。

将参数统一到一个参数向量中,将混合权重记为 α_l,每个块的参数记为 $\theta_l=(\boldsymbol{\mu}_l,\Sigma_l)$,从而得到参数向量 $\Theta=(\alpha_1,\cdots,\alpha_g,\theta_1,\cdots,\theta_g)$。因此混合模型具有下面的形式:

$$p(\boldsymbol{x}\mid\Theta)=\sum_{l=1}^{g}\alpha_l p_l(\boldsymbol{x}\mid\theta_l)$$

每个分量的密度函数是通常的高斯函数:

$$p_l(\boldsymbol{x}\mid\theta_l)=\frac{1}{(2\pi)^{d/2}\det(\Sigma_l)^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right\}$$

图像的似然函数为:

$$\prod_{j\in\text{observations}}\left(\sum_{l=1}^{g}\alpha_l p_l(\boldsymbol{x}_j\mid\theta_l)\right)$$

模型中的每个分量对应于一个分割区域,并且参数向量 Θ 是未知的。

非常重要的一点是,如果已经知道了各个像素分别来源于哪个分量,那么确定 Θ 将会变得相对简单一些。首先使用最大似然估计来估计每个 θ_l,然后根据每个分量对应的区域所占图像的比例来计算 α_l。同样,如果知道了 Θ,那么对于每个像素,就能够确定最可能产生那个像素的分量,这样的结果就产生了对图像的分割。但是,问题的困难在于二者我们均不知道。

例 16.2　由点集拟合直线

在平面上有 g 条不同的直线。第 l 条直线由参数 $\boldsymbol{a}_l$ 所表示,并且以概率 π_l 生成样本。每一个样本均对应一个测量向量 $\boldsymbol{W}$,并且第 j 个测量的值用 $\boldsymbol{W}_j$ 来表示。对于第 l 条直线,可用概率密度函数来表示它产生样本的可能性,记为 $p(\boldsymbol{W}|\boldsymbol{a}_l)$。因此,能观测到某一样本的概率密度函数可用下面的公式表示:

$$p(\boldsymbol{W})=\sum_{l}\pi_l p(\boldsymbol{W}\mid\boldsymbol{a}_l)$$

这同样是一个混合模型。在这个模型下,可观测到某一样本集的概率可表示为:

$$\prod_{j\in\text{observations}}\left(\sum_{l=1}^{g}\pi_l p(\boldsymbol{W}_j\mid\boldsymbol{a}_l)\right)$$

我们希望能够推导出 $\boldsymbol{a}_l$ 和 π_l。与在分割中遇到的情况一样,如果知道哪个点是由哪条直线所产生的,问题将会变得简单。可通过直线拟合的方法来估计 $\boldsymbol{a}_l$ 并且通过计算由第 l 条直线生成的样本的个数来获得 π_l。但是困难在于只有样本的测量值,而没有表示样本和直线之间的关联的值。

丢失数据问题的形式化表述 假设有两个数据空间：完备的数据空间$\mathcal{X}$和不完备的数据空间$\mathcal{Y}$。存在从$\mathcal{X}$到$\mathcal{Y}$的一个映射f，这个映射关系使得某些数据丢失了。例如，映射关系为一个投影关系。对于图像分割的例子来说，完备的数据包含每个像素点的测量值和一个用于指示该测量值来自于混合模型中的哪个分量的变量集。所获得的不完备数据是由于丢掉了第二个表示对应关系的变量集。对于直线和样本的例子来说，完备的数据空间包括每个样本点的测量值(位置，当然，色彩和形状也可以包括进来)以及标识哪个样本来自于哪条直线的变量集，同样所获得的不完备数据是因为丢掉了第二个变量集。

模型的参数可以形成一个参数空间$\mathcal{U}$。对于图像分割的例子来说，参数空间包括混合权重以及每个混合分量的参数。对于直线和样本问题来说，参数空间包括混合权重以及每条直线的参数。我们希望在给定不完备数据的条件下能够获得这些参数的最大似然估计。如果能够获取完备的数据，就能够使用描述完备数据空间的概率密度函数，记为$p_c(\boldsymbol{x};\boldsymbol{u})$。完备数据的对数－似然函数可表示为：

$$
\begin{aligned}
L_c(\boldsymbol{x};\boldsymbol{u}) &= \log\left\{\prod_j p_c(\boldsymbol{x}_j;\boldsymbol{u})\right\} \\
&= \sum_j \log\left(p_c(\boldsymbol{x}_j;\boldsymbol{u})\right)
\end{aligned}
$$

在任何一个例子中，用对数似然函数来计算都要相对简单一些。对于图像分割的情形，问题可描述为：在给定每个像素都来源于哪个分割区域的情况下，估计每个图像分割区域的参数。在直线和样本情形中，可把问题描述为：在给定每个样本来源于哪条直线的前提下，估计混合权重和参数。

但是目前的问题是我们无法获得完备的数据。将不完备数据空间的概率密度函数表示为$p_i(\boldsymbol{y};\boldsymbol{u})$。不完备数据空间的概率密度函数，可以通过对取值为$\boldsymbol{y}$的完备数据空间中的所有值的概率密度函数进行积分来获得。即：

$$
p_i(\boldsymbol{y};\boldsymbol{u}) = \int_{\{\boldsymbol{x}|f(\boldsymbol{x})=\boldsymbol{y}\}} p_c(\boldsymbol{x};\boldsymbol{u})\,d\eta
$$

其中，η指的是满足关系$f(\boldsymbol{x})=\boldsymbol{y}$的空间$\boldsymbol{x}$中的单位变化量。不完备数据的似然函数可表示为：

$$
\prod_{j\in\text{observations}} p_i(\boldsymbol{y}_j;\boldsymbol{u})
$$

通过写出$\boldsymbol{y}$的似然函数并且最大化来估计$\boldsymbol{y}$的值，然后就可以得到对$\boldsymbol{u}$的最大似然估计。这个过程并不简单，因为无论是积分还是最大化做起来都很困难。由于积分位于对数符号内，常用的去对数的方法也不能使问题变得简单一些。有下面的计算公式：

$$
\begin{aligned}
L_i(\boldsymbol{y};\boldsymbol{u}) &= \log\left\{\prod_j p_i(\boldsymbol{y}_j;\boldsymbol{u})\right\} \\
&= \sum_j \log\left(p_i(\boldsymbol{y}_j;\boldsymbol{u})\right) \\
&= \sum_j \log\left(\int_{\{\boldsymbol{x}|f(\boldsymbol{x})=\boldsymbol{y}_j\}} p_c(\boldsymbol{x};\boldsymbol{u})\,d\eta\right)
\end{aligned}
$$

这种形式的表达是非常难于处理的。不完备数据的似然函数使得我们束手无策的原因，是我们不知道与 $\boldsymbol{ys}$ 对应的那些 $\boldsymbol{xs}$ 是否真正地与它相对应。建立不完备数据的似然函数包括了对所获得的 $\boldsymbol{xs}$ 取平均这样一个过程。

策略　对于上面所举的任意一个例子，如果已经知道了丢失的数据，那么就能够有效地估计参数。同样，如果知道了参数，就会计算出丢失的数据。因此可以使用下述迭代算法：

1. 对参数进行假设，获取丢失数据的某个估计。
2. 使用丢失数据的估计，得到自由参数的最大似然估计。

迭代这个过程直到收敛(希望能够)。对于图像分割问题来说，处理过程如下：

1. 获取对于模型分量的估计，使用对 θ_l 的估计值可以计算出每个像素来源于哪个分量；
2. 使用这个估计修改 θ_l 和混合权重。

在样本和直线的例子中，算法如下：

1. 首先给 $\boldsymbol{a}_l$ 一个猜测值，然后获取样本和直线之间的对应关系的估计。
2. 使用所估计的对应关系，修改 $\boldsymbol{a}_l$ 的估计值。

16.1.2　EM 算法

尽管我们希望给丢失数据问题设计的计算过程会收敛，但是并没有理由使得我们能够相信这些过程是收敛的。事实上，如果每一阶段均能给出合适的选择的话，它们应该会收敛。实际上，通过表明该算法是通用算法——期望最大化(expectation-maximization)算法——的特例，很容易地就能说明上述观点。

混合模型的 EM　现在假设对于丢失的变量来说，完备数据的对数似然函数是线性的。由于它和混合模型相关联，这种情况是常见的。我们给出的所有例子均具有这个性质。

在一个混合模型中，丢失的数据包括用来指示哪个数据项来源于哪个混合分量的变量(即，某样本点是来自于直线还是来自于噪声)。我们通过下面的方式来表达这一信息：对每一数据点赋予一个向量 $\boldsymbol{z}$，该向量具有 g 个元素(回忆一下，我们在前面所举的例子中每个混合模型都具有 g 个分量)。如果第 j 个数据点来源于混合模型的第 l 个分量，那么 $\boldsymbol{z}_{jl}$ 的第 l 个分量赋值为 1，否则赋值为 0。那么有 $\boldsymbol{x}_j = [\boldsymbol{y}_j, \boldsymbol{z}_j]$。现在如果我们将混合模型表示为：

$$p(\boldsymbol{y}) = \sum_l \pi_l p(\boldsymbol{y} \mid \boldsymbol{a}_l)$$

那么完备数据的对数似然函数可表示为：

$$\sum_{j \in \text{observations}} \left(\sum_{l=1}^{g} z_{lj} \log p(\boldsymbol{y}_j \mid \boldsymbol{a}_l) \right)$$

其中丢失的变量是线性的。

EM 算法的主要思想是通过用期望值来替代丢失数据，为丢失的数据获取工作变量的集合(对于 $\boldsymbol{x}$ 也是一样的)。特别是在给定值 $\boldsymbol{y}_j$ 以及参数值的情况下，在某些值上调整参数，然后计算 $\boldsymbol{z}_j$ 的期望值。接着将计算出的期望值 $\boldsymbol{z}_j$ 代入到完备数据的似然函数中，用这个函数计算相对要简单一些，然后通过最大化这个函数获得参数的值。这时，$\boldsymbol{z}_j$ 的期望值可能已经改

变了。通过交替执行期望阶段和最大化阶段，迭代直至收敛，我们给出下列算法。

描述得形式化一些，给定 $\boldsymbol{u}^s$，通过下述过程计算 $\boldsymbol{u}^{s+1}$：

1. 使用不完备的数据以及参数的当前值来计算完备数据的期望值。对于任意的 j，我们已经知道了 $\boldsymbol{y}_j$ 的期望值，现在需要计算的是 $\boldsymbol{z}_j$ 的期望值。将这个值记为 $\bar{\boldsymbol{z}}_j^{(s)}$。使用上标来表明期望依赖于参数的当前值。这个步骤称为 E 步。
2. 使用 E 步计算出的完备数据的期望值，最大化完备数据关于 $\boldsymbol{u}$ 的对数似然函数。即计算：

$$\begin{aligned}\boldsymbol{u}^{s+1} &= \arg\max_{\boldsymbol{u}} L_c(\bar{\boldsymbol{x}}^s; \boldsymbol{u}) \\ &= \arg\max_{\boldsymbol{u}} L_c([\boldsymbol{y}, \bar{\boldsymbol{z}}^s]; \boldsymbol{u})\end{aligned}$$

这一步称为 M 步。

可以证明，不完备数据的对数似然函数在每个阶段都是增长的，意思就是序列 $\boldsymbol{u}^s$ 收敛到不完备数据对数似然函数的某个（局部的）最大值［见 Dempster, Laird 和 Rubin（1977）或 McLachlan 和 Krishnan（1996）］。当然，无法保证算法可以收敛到正确的局部最大值，而且下面所给出的例子表明寻找正确的局部最大值是一件很麻烦的事。

16.1.3 通用情况下的 EM 算法

如果对于丢失的数据来说，完备数据的对数似然函数并不是线性的，那么就不能简单地替换这些变量的期望值，而必须通过在参数的当前值条件下，计算完备数据的对数似然函数的期望值来处理这些丢失的变量。假设我们知道参数 $\boldsymbol{u}^{(s)}$ 的估计值，那么在给定对参数的估计以及不完备数据的知识的条件下，就可以对完备数据的所有值求取完备数据的对数似然函数的平均值，并且对于每种情况通过所计算的概率值来分布权重。这个处理过程满足以下函数：

$$Q(\boldsymbol{u}; \boldsymbol{u}^{(s)}) = \int L_c(\boldsymbol{x}; \boldsymbol{u}) p(\boldsymbol{x} \mid \boldsymbol{u}^{(s)}, \boldsymbol{y})\, d\boldsymbol{x}$$

上式是先前估计的参数值的函数。我们关于 $\boldsymbol{u}$ 最大化这个函数，得到 $\boldsymbol{u}^{(s+1)} = \arg\max_{\boldsymbol{u}} Q(\boldsymbol{u}; \boldsymbol{u}^{(s)})$。通过简单的练习（主要为了引起注意），就可以看到这可以归于以前所描述过的线性情况的算法。

16.2 EM 算法的应用

丢失数据问题在计算机视觉中随时都可能会出现。我们收集了许多例子来说明这个常见的情况。计算通常都比较直接，一旦给出一部分的计算过程，通常会省略其他部分的计算。

16.2.1 例子：图像分割

假设存在 n 个像素，丢失的数据形成一个用 $n \times g$ 的数组表示的指示变量 $\mathcal{I}$。在每一行，除了一个元素，其他的值均为 0，这个值表示每个像素的特征向量来源于哪个块。这是上面描述的例子 16.1。

***E* 步：** 如果第 l 个像素来自第 m 个块，$\mathcal{I}$ 的 l 行 m 列元素为 1，否则为 0。意思是：

$$
\begin{aligned}
E(I_{lm}) &= 1P(\text{第 } l \text{ 个像素来自第 } m \text{ 个块}) \\
&\quad + 0.P(\text{第 } l \text{ 个像素不来自第 } m \text{ 个块}) \\
&= P(\text{第 } l \text{ 个像素来自第 } m \text{ 个块})
\end{aligned}
$$

假设第 s 层迭代的参数为 $\Theta^{(s)}$，我们有：

$$
\bar{I}_{lm} = \frac{\alpha_m^{(s)} p_m(x_l \mid \theta_l^{(s)})}{\sum_{k=1}^{K} \alpha_k^{(s)} p_k(x_l \mid \theta_l^{(s)})}
$$

记住 $\alpha_m^{(s)}$ 的意思是 α_m 在第 s 层迭代的值！

***M* 步：** 一旦得到 $\mathcal{I}$ 的期望值，剩下的就简单了。本质上是要形成 Θ^{s+1} 的最大似然估计。此外，指示变量的期望值通常并不为 0 或 1，而是取 0，1 之间的某个值。这种现象可解释为某种特定的观测通常以某种频率发生，从而在似然函数中对应于这个指示变量的项，通常以期望值的幂的量级而增加。计算加权平均和加权标准方差的表达式是熟悉的：

$$
\alpha_m^{(s+1)} = \frac{1}{r} \sum_{l=1}^{r} p\left(m \mid x_l, \Theta^{(s)}\right)
$$

$$
\mu_m^{(s+1)} = \frac{\sum_{l=1}^{r} x_l p\left(m \mid x_l, \Theta^{(s)}\right)}{\sum_{l=1}^{r} p\left(m \mid x_l, \Theta^{(s)}\right)}
$$

$$
\Sigma_m^{s+1} = \frac{\sum_{l=1}^{r} p\left(m \mid x_l, \Theta^{(s)}\right) \left\{ (x_l - \mu_m^{(s)}) (x_l - \mu_m^{(s)})^T \right\}}{\sum_{l=1}^{r} p(m \mid x_l, \Theta^{(s)})}
$$

同样记住 $\alpha_m^{(s)}$ 的意思是 α_m 在第 s 层迭代的值！

其余的问题包括指定合适的特征向量，以及讨论如何开始 EM 算法这样的问题。图 16.1 和图 16.2 所显示的结果使用了三个色彩特征——图像经过平滑后在 L * a * b * 空间的像素坐标——和三个纹理特征——这些纹理特征使用滤波器的输出来估计局部尺度，各向异性以及对比度（见图 16.1）。其他的特征可以用像素的位置，或许会更加有效。

算法 16.1　使用 EM 算法的色彩和纹理分割

选择要分割的区域的数目

创建一个支持图集，每个分割区域一个图，每个像素在每一个图中有一个元素。这些支持图将包含将像素和分割联系在一起的权重

通过下述任意一种方法初始化支持图：

　　从小块像素估计分割的参数，然后使用 E 步计算权重

　　或者

　　对支持图随机地分配任意值；

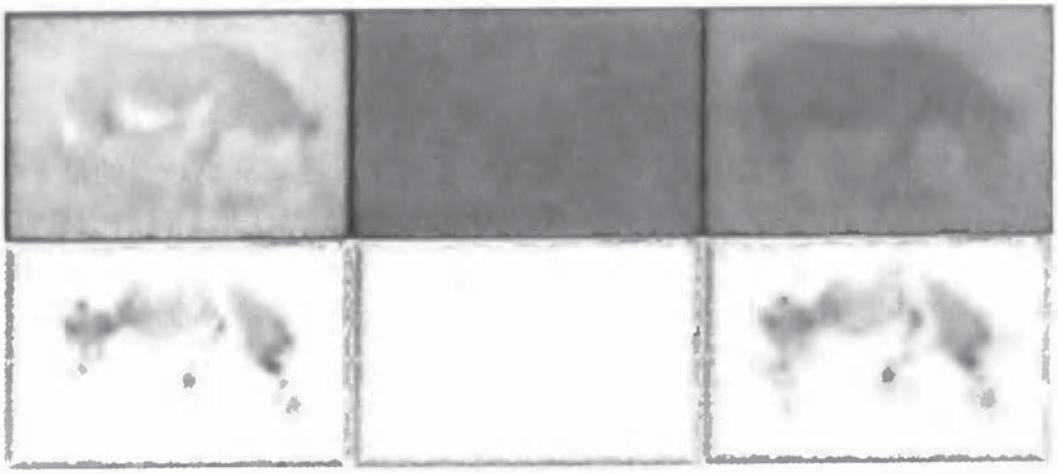
进行下列步骤直到收敛

　　使用 E 步修改支持图

　　使用 M 步修改分割的参数

end

a b



c

图 16.1 用可变的尺度对图像(a)中的斑马进行平滑从而生成图像(b)。图中的平滑使用的是对尺度的局部估计。这些尺度的度量主要是测量一个像素周围变化的尺度，在边缘的位置，尺度要窄一些，在条纹区域，尺度要宽一些。(c)图显示了所获得的特征结果；上面三幅图像显示平滑后的彩色坐标，下面三幅图像显示纹理特征

图 16.2 斑马图像(与图 16.1 所显示的图像相同)中的每个像素用 m 的值所标识，其中$p(m \mid \boldsymbol{x}_l, \Theta^s)$是产生分割的最大值。图像显示了$K=2,3,4,5$时的处理结果。对应于分割标号，每幅图像有$K$个灰度级

分割器应该提供给我们哪些信息？一种做法是给每个像素选择那些能使得 $p(m \mid \boldsymbol{x}_l, \Theta^s)$ 取得最大值的 m，另一种做法是给出它们的概率并且在它们之上建立一个推理过程。

算法 16.2　使用 EM 算法的色彩和纹理分割：E 步

对于每个像素的位置 l

　对于每个分割 m

　　在支持图中的像素的位置插入 $\alpha_m^{(s)} p_m(\boldsymbol{x}_l \mid \theta_l^{(s)})$

　end

　将支持图中的值相加得到 $\sum_{k=1}^{K} \alpha_k^{(s)} p_k(\boldsymbol{x}_l \mid \theta_l^{(s)})$，并且将每个支持图中位于位置 l 的值均除以此项

end

算法 16.3　使用 EM 算法的色彩和纹理分割：M 步

对每个分割 m

　使用以下的表达式得到分割参数的新值：

$$\alpha_m^{(s+1)} = \frac{1}{r} \sum_{l=1}^{r} p(m \mid \boldsymbol{x}_l, \Theta^{(s)})$$

$$\boldsymbol{\mu}_m^{(s+1)} = \frac{\sum_{l=1}^{r} \boldsymbol{x}_l p(m \mid \boldsymbol{x}_l, \Theta^{(s)})}{\sum_{l=1}^{r} p(m \mid \boldsymbol{x}_l, \Theta^{(s)})}$$

$$\Sigma_m^{s+1} = \frac{\sum_{l=1}^{r} p(m \mid \boldsymbol{x}_l, \Theta^{(s)}) \left\{ (\boldsymbol{x}_l - \boldsymbol{\mu}_m^{(s)})(\boldsymbol{x}_l - \boldsymbol{\mu}_m^{(s)})^T \right\}}{\sum_{l=1}^{r} p(m \mid \boldsymbol{x}_l, \Theta^{(s)})}$$

　其中，$p(m \mid \boldsymbol{x}_l, \Theta_{(s)})$ 是第 m 个支持图中位于像素位置 l 的值

end

算法 16.4　EM 直线拟合，通过使用最近的直线从而获得最高的权重的方法，给每个点设置对每条直线上的权重

选择 k 条线（可能在概率上是均等的）

或者

选择 $\overline{\mathcal{L}}$

直到收敛

　E 步：

　　用垂直距离重新计算 $\overline{\mathcal{L}}$

　M 步：

　　使用 $\overline{\mathcal{L}}$ 中的权重重新拟合直线

end

16.2.2 例子:使用 EM 进行线拟合

对前面介绍的直线拟合图像分割例子用 EM 算法;丢失的数据是指示变量$\mathcal{M}$所表示的数组,如果点 k 属于线 l,那么第 k 行,第 l 列的元素 m_{kl} 为 1,否则为 0。同前面例子的处理过程类似,通过确定 $P(m_{kl}=1|$点 k, 直线 l 的参数)来得到期望值,并且这个概率值对 l 来说正比于:

$$\exp\left(-\frac{\text{点}k\text{到直线}l\text{的距离的平方}}{2\sigma^2}\right)$$

通过下面的公式,我们能够很容易地计算出比例常数:

$$\begin{aligned}&\sum_k P(m_{kl}=1\mid \text{点}k,\ \text{直线}l\text{的参数})\\&=\sum_l P(m_{kl}=1\mid \text{点}k,\ \text{直线}l\text{的参数})\\&=1\end{aligned}$$

最大化的过程遵循的是对一个数据点集拟合出一条直线的处理形式,但是不同的是需要执行这个处理过程 g 次,并且点的坐标要通过$\overline{l_{kl}}$的值进行加权。通过观察线上的变化幅度,或者考察点距离线的垂直距离来判断算法是否收敛(执行的时候采用的是对数似然函数,见习题)。

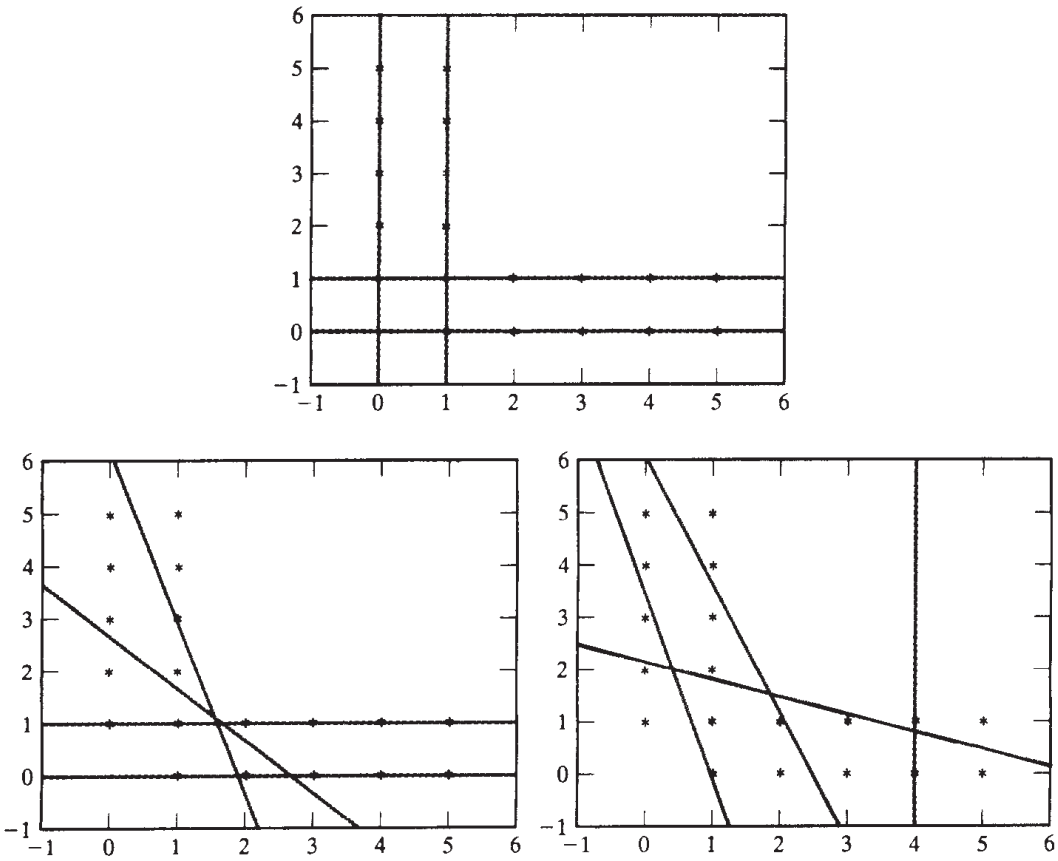


图 16.3 上面的图显示了使用 EM 线拟合算法所获得的一个很好的拟合结果。下面的图给出了两个不好的例子,在这两个例子中均用正确直线数去运行程序,但是收敛到了较差的拟合结果,这个结果却能够给数据一个较好的解释,并且的确是局部最小。实现的时候对于混合模型增加了一个分量,将数据点建模为在数据域中是随机均匀分布的;如果认为一个数据点属于这个分量的概率很大,那么该点被判为噪声点。在图16.4中我们会进一步给出一些拟合不好的例子

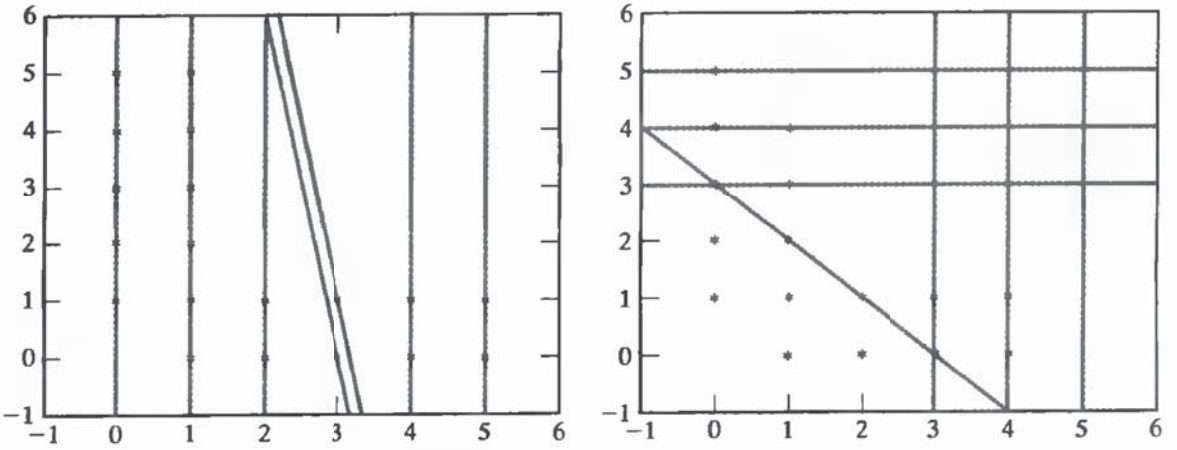


图 16.4　对图 16.3 所给出的数据，拟合得更差的例子。在这个例子中，我们尝试用 7 条直线来拟合数据集。注意所得到的拟合结果可以很好地解释数据；它们均是似然函数的局部极值。实现的时候对于混合模型增加了一个分量，将数据点建模为在数据域中是随机均匀分布的；如果认为一个数据点属于这个分量的概率很大，那么该点被判为噪声点。右图底部将某些点归于噪声，而该图对其他数据点拟合得很好

16.2.3　例子：运动分割和 EM

例如，运动序列通常包含内部具有相似运动的大区域。我们暂时假设获取了一段很短的序列——两帧——并且希望确定第一帧每点的运动场。假设运动场是用混合模型来表示的。回忆一下，通常的混合模型是密度的加权和——分量并不一定需要具有 16.1 节的高斯形式（丢失的数据、EM 算法和常见的混合模型，在视觉应用中经常一起出现）。

运动序列的生成模型通常具有下面的形式：

- 在图像中的每一点，存在一个联系该像素和下一图像中的像素的运动向量；
- 存在参数不同的运动场集合，每个具有不同的概率模型；
- 总的运动由混合模型所给出，意味着如果要确定每个像素点的图像运动，需要确定运动来源于哪个部分，然后从这个部分中得到样本。

这个模型概括了一组不同的、内部一致的运动场——运动场可能来自于，暂且说，位于不同深度的刚体的集合和一个运动的摄像机（见图 16.5）。分离的运动场通常称为层次，模型称为层次化的运动模型。

图 16.5　所采集的 MPEG 花园图像序列的第 1，15 和 30 帧，这个图像序列常常用来演示运动分割算法的有效性。图像序列是用平移运动的摄像机采集的，其中树离摄像机较近一些，房屋和地面上的花园离摄像机相对较远一些。因此，树看起来平移得相对快一些，而房屋相对慢一些。平面生成了一个仿射运动场

现在假设运动场具有参数化的形式，并且存在 g 个不同的运动场。给定一对图像，我们希望确定以下几点：(a)像素属于哪个运动场；(b)每个运动场的参数值。这些和前面两个例子看起来非常相似，因为如果知道了第一个，确定第二个就变得很容易，而如果知道了第二个，解决第一个问题也会变得简单。这还是一个数据丢失问题：丢失的数据是像素所归属的运动场，而参数是表示每个场的参数和混合权重。

假设在第一幅图像中，位于(u,v)的像素属于第 l 个运动场，该运动场的参数为 θ_l。从而在第二帧图像中这一点运动到了位置$(u,v)+\boldsymbol{m}(u,v;\theta_l)$，因此不考虑噪声的影响这两个像素点的强度应该是相等的。我们用 $I_1(u,v)$表示第一帧中位于 u,v 的像素的图像强度，其余的表示类似。丢失的数据是像素所归属的运动场。可以用指示变量 $V_{uv,l}$来表示这个含义，

$$V_{uv,l}=\left\{\begin{array}{l}1,\text{ if the } u,v\text{th pixel belongs to the } l\text{th motion field}\\ 0,\text{其他}\end{array}\right\}$$

假设在获取的强度图像中，高斯噪声具有标准方差，因此完备的数据对数似然函数为：

$$L(V,\Theta)=-\sum_{ij,l}V_{uv,l}\frac{(I_1(u,v)-I_2(u+m_1(u,v;\theta_l),v+m_2(u,v;\theta_l)))^2}{2\sigma^2}+C$$

其中，$\Theta=(\theta_l,\cdots,\theta_g)$。这里可直接使用 EM 算法。像前面一样，关键的问题是确定

$$P\left\{V_{uv,l}=1\mid I_1,I_2,\Theta\right\}$$

这些概率关系经常用支持图来表示——在图中对每个像素赋予了一个灰度级，灰度级的数值表示了该像素点最可能属于的层次(见图 16.6)。更感兴趣的问题是选择合适的参数运动模型。通常选用仿射运动模型，其中：

$$\left\{\begin{array}{c}m_1\\ m_2\end{array}\right\}(i,j;\theta_l)=\left\{\begin{array}{cc}a_{11} & a_{12}\\ a_{21} & a_{22}\end{array}\right\}\left\{\begin{array}{c}i\\ j\end{array}\right\}+\left\{\begin{array}{c}a_{13}\\ a_{23}\end{array}\right\}$$

其中，$\theta_l=(a_{11},\cdots,a_{23})$。层次化的运动表示是非常有用的，这有以下几个原因：首先，这种表示将“以同样方式”运动的那些点聚类到了一起。其次，表示能够给出运动边界。最后，新的序列能够以感兴趣的方式从层次中重建(见图 16.7)。

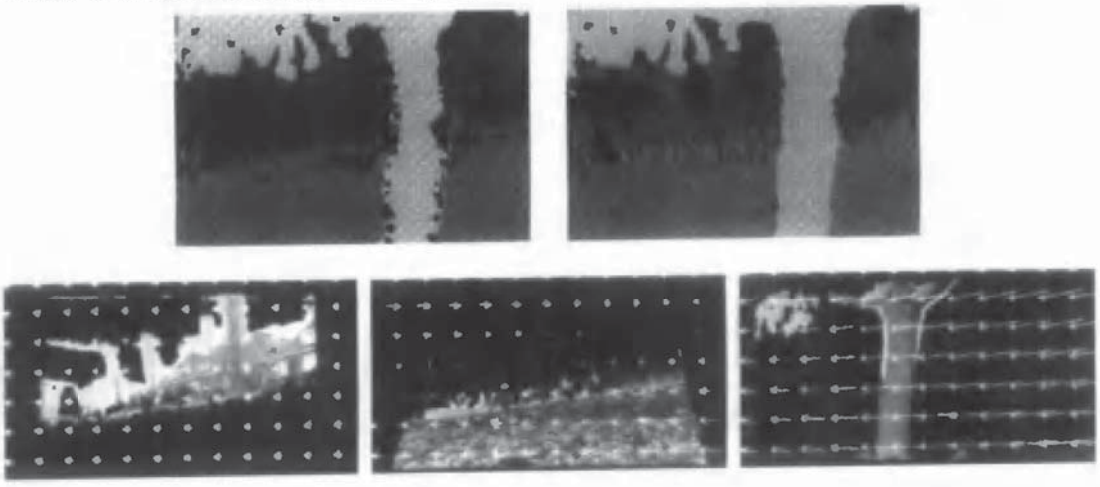


图 16.6 上层的左图表示的是花园图像序列中的像素属于哪一个层次，通过对图像运动的局部估计聚类而获得的。每个灰度级别对应一个层次，每个层次均遵循不同的仿射运动模型而运动。通过检测在时间序列上(过去的和将来的帧)周围像素运动的一致性关系可以将左图进行细化，细化的结果显示在上层的右图。下层的图显示了三个层次以及它们的运动模型

图 16.7　将运动表示为层次方法的一个特征是在去掉一些层次的情况下可以重建运动序列。在这个例子中，我们在去掉树层次的条件下，重建了MPEG花园图像序列。左图对应的是第一帧，中间的图对应的是第15帧，右图对应的是第30帧

16.2.4　例子：使用 EM 来识别出格点

前面所讨论的线拟合器在处理出格点上会遇到困难，由于它们会碰到那些模型预测不到的情况。出格点通常被认为位于概率分布的“尾部”。对于像正态分布这样的概率分布，存在大量的小概率值；这些值通常位于分布的尾部（可能由于这些值通常位于分布逐渐变小的位置）。很自然能想到的处理出格点的方法是修改模型使得分布具有明显的尾部（即，尾部的概率值要大一些）。

实现这种想法的一种方式是为出格点创建一个显式的模型，这个过程通常非常简单。我们对似然函数 P(测量|模型)和出格点项 P(出格点)求取加权和，从而得到

$$(1-\lambda)P(\text{测量}\mid\text{模型})+\lambda P(\text{出格点})$$

其中，$\lambda\in[0,1]$建模出格点出现的频率，并且 P(出格点)表示的是出格点的概率模型。如果没有更好的方法，可以在数据区域中选取均匀分布。

很容易想到的一种处理这个模型的方法是设立一个变量用来表示每个点来源于哪个部分。通过这个变量，可以用一种简单的形式建立完备数据的似然函数。当然，我们并不知道这个变量的值，这是一个丢失数据问题，我们知道如何用 EM 算法来求解这个问题（请读者在练习中给出详细的解法）。使用 EM 算法所面临的困难这里同样也会遇到（见图 16.8）。特别是很容易陷入到局部最小中去，因此需要小心对那些小概率所用的数值表示。

16.2.5　例子：使用 EM 进行背景提取

就像我们在 14.3.1 节中所看到的一样，背景估计问题是比较难的。简单地将多帧平均具有以下问题：在某一位置占据较长时间的一个物体会使得背景的平均值偏离很多。可以将这个问题看做丢失变量问题：视频中的每帧图像都是相同的（由于考虑到各种自动获取控制系统不同的调节方式，通常乘以一个常量），并且加入了噪声。将噪声建模为来源于同样的噪声源。这样会带来一些好处，属于噪声的那些像素不属于背景；通过简单地考察丢失变量在极值点上的期望值，可以获取背景估计。计算很直接，图 14.10 和图 14.11 就是以这种方式获得的。

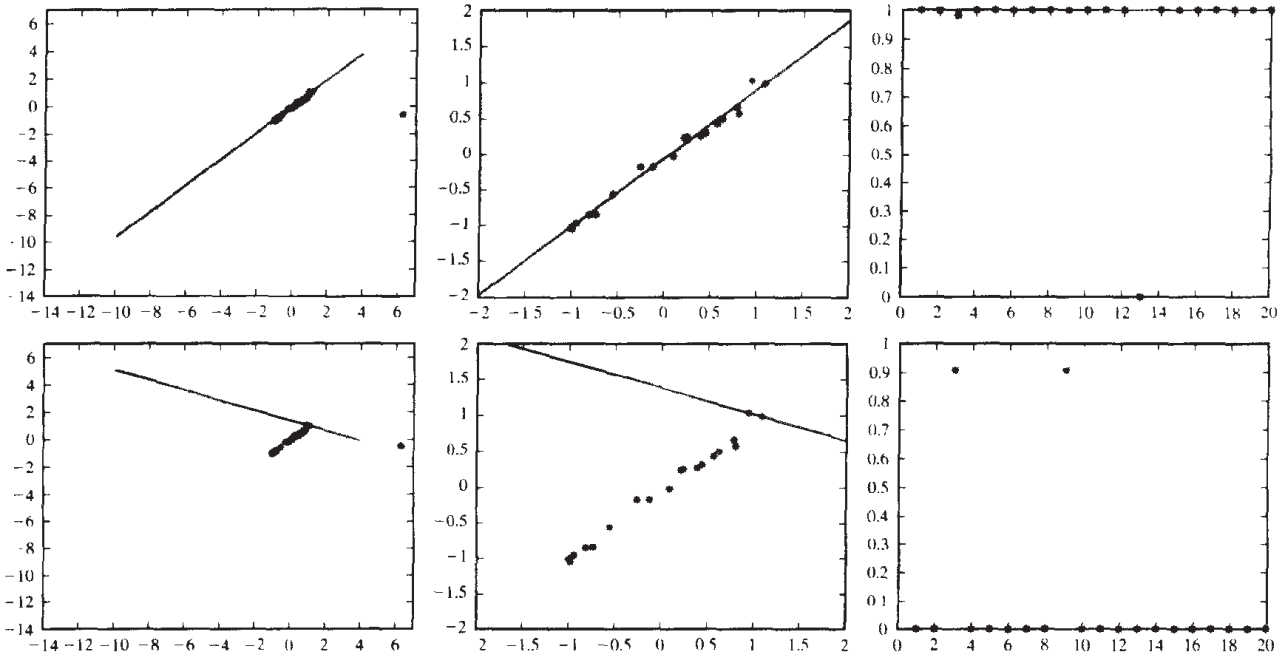


图 16.8 EM 算法也可用于排除一些出格点。现在来表明对图 15.7 的第二个数据集线拟合的有效性。上面一行显示了正确的局部最小点，下面一行显示了另外一个局部最小点。第一列显示了使用与图 15.7一样的坐标轴叠加到数据点上的线，第二列将显示限制在数据点的周围，显示了这条线的细节；第三列显示的是点来自于这些直线而不是来自于噪声模型的概率图，图是按照点的坐标绘制的。注意到在正确的局部最小的位置，除了一个点其余的点都和直线相关，但是在不正确的局部最小点，只有两个点同直线相关，其余的均被归于是噪声点

16.2.6 例子：EM 和基础矩阵

拟合基础矩阵问题同样也可以看做是丢失数据问题——现在丢失的是对应信息。假设在左、右图像上各有 n 个点，可建立一个 $n \times n$ 的矩阵C 来表示对应关系。详细解释一下，如果左图的第 i 个点与右图的第j 个点相对应，那么 c_{ij} 为 1，否则为 0。E 步和 M 步的详细形式写起来比较冗长，作为课后练习，在这里略去。矩阵C 表示丢失的数据；如果我们知道了矩阵，就能够计算完备数据的对数似然函数并使用求极值的方法来求取参数——参数为$\mathcal{A}$，对于每点来说，用 x_{ri}，y_{ri}和 x_{li}来表示。类似地，给定参数的估计值，计算C 的期望值相对直接一点（左图中的每一点可对右图中的点进行预测——这些点和它们测量之间距离的指数值可生成期望值）。另外，通常希望C 的期望值不是整数。就像前面所讨论过的一样，将权重解释为发生的频率，并且在计算似然函数的时候自乘概率到相关次幂。

除非仔细的选择初始阶段的值，这种方法一般不会给出较好的结果。遇到的问题很简单：EM 算法可能削弱丢失数据问题的组合搜索成分所带来的副作用，但是不能使得它们完全消失。由C 所表示的对应关系所占据的空间是巨大的（$n!$），这个空间相当大的部分几乎没有包含好的对应。由于在给定对应集的前提下，算法寻找可能得到的最好的基础矩阵，不好的初始化的结果可能会导致严重的后果。这是因为对应的初始集合可能是错的，可导致对基础矩阵的错误估计、得到不正确的预测，等等。一种可行的方法是使用 RANSAC 作为 EM 的起点。这种方法相比于传统 RANSAC 的好处是，通过在最后阶段使用 EM，可以较好地为相匹配的对应元素的贡献加权。

16.2.7　EM 算法的困难

EM 算法会遇到局部最小问题。这些局部最小点与所研究的问题的各个方面联系在一起,如线拟合的例子和基本矩阵的例子。这些困难来自于假设点是可以交换的(见图 16.3 和图 16.4)。注意到下列情况可以避免这个问题,因为任何一个拟合的最终结果均是初始点的确定性函数,所以应该仔细选择那些初始点。一种实现策略是从不同(自由选择)的组合出发,从结果中筛选寻找最佳的拟合。另外一个方法是使用像 Hough 变换这样的技术首先对数据进行预处理,从而猜测好的初始线拟合。但是哪种方法都不能保证能得到最好的结果。另外一种筛选的方法是注意到下列情况,很少会碰到一堆不可分辨的点,并且需要从这堆点中推导出结构。如果遇到了这种情况,通常是由于没有很好地描述所遇到的问题。如果点并不是不可区分的,并且具有某种形式的连接结构,那么选择一个较好的起始点会变得比较容易。

所遇到的另外一个困难是有些点具有非常小的期望权值。这给我们提出一个数值问题,如果将很小的权值设为 0 的话,不知道会发生什么样的情况(但这不是一个明智的做法)。相反,可以采用一些数的表示,这些表示使得可以将非常小的数相加起来,得到一个非 0 的结果。这个问题的讨论超出了本书的范围,但是不能因为在这里不详细讨论这个问题而低估了它的危害。

16.3　模型选择:哪个模型拟合得最好

在对丢失变量问题讨论的每一阶段,我们都假设混合模型的分量的数目是已知的。在实践中这个假设通常是不成立的。找到分量的数目,本质是一个模型选择问题,我们在模型集(其中不同的模型具有不同数目的分量)中寻找,从而确定哪个对数据拟合得最好。通常,负对数似然值对分量数目的指导是不好的,因为一般情况下参数较多的模型与一个数据集的拟合要比参数较少的模型好。这意味着,简单地用分量数目的函数来最小化负对数似然函数会生成太多的分量。例如,可以通过为每对数据点均设立一条通过它的直线来非常精确地拟合直线集,这样虽然可能会生成许多条直线,但是拟合误差为 0。我们在模型中添加一项来解决这个困难,该项随着分量数目的增加而增加。这种惩罚措施可以补偿由参数数目的增加所引起的负对数似然值的减少。

不存在规范的模型选择过程,理解这一点很重要。我们可以从已有的许多技术中去选择合适的技术来进行模型选择,由于每种技术使用的主要原理有所不同,所以所引起的问题也不一样(以及对标准的不同近似)。

16.3.1　基本想法

在拟合参数模型中,通常需要选择模型。模型选择问题可通过下面的步骤建立起来:有一个数据集合,是模型集合中某一参数模型的样本。我们希望确定:(a)数据集是从哪个模型中得到的,(b)模型具有什么样的参数。选择合适的参数可以用模型对将来的采样数据进行预测——测试集——即数据集(通常称为训练集)。可惜的是,这些将来的采样是不可获得的。另外使用数据集所估计出来的模型参数可能出现偏差,这是因为选择的参数只能够保证模型能够最佳地拟合训练集,而不是可能出现的数据的全部。这种现象称为选择偏差。训练集是能

够从模型所获得的数据全集的子集；只有当训练集无限大时，它才能够精确地表示模型。这就是负对数似然函数不能够指导我们选择出较好的模型的原因：拟合看起来变好了是因为它更加偏向于训练集。

所使用的正确的惩罚来自于下式表示的偏移：

2(最佳模型的对数似然函数 - 当前模型的对数似然函数)

(引自 Ripley(1996)，第 348 页)最好的模型应该是真实的模型，理想情况偏差应该为 0。上面的论述表明训练集的偏差要比测试集的偏差大。很容易想到的一个惩罚措施是，求取这个偏差在训练集和测试集上的平均的差值。这个惩罚是拟合的对数似然函数的两倍——我们不能解释两倍的原因，但是并不影响应用。我们将选择的最佳参数记为 Θ^*，对数据进行拟合的对数似然函数记为 $L(\boldsymbol{x};\Theta^*)$。

16.3.2 AIC

Akaike 提出了一种惩罚方法，常称为 AIC(为“一个信息标准，An information criterion”，而不是“Akaike 信息标准”)，目的是最小化：

$$-2L(\boldsymbol{x};\Theta^*)+2p$$

其中，p 是自由参数的数目。关于 AIC 存在大量的统计学上的争论。第一个主要的观点是缺乏一个关于数据点数目的项。这种观点是值得怀疑的，因为随着数据点的数目增加，拟合的模型和真实模型之间的偏差将变小。第二，大量的经验表明 AIC 趋向于过拟合，即为，选择一个具有很多参数模型能够较好地拟合训练集，但是这个模型在测试集上执行得并不好。

16.3.3 贝叶斯方法和 Schwartz 的 BIC

为了简单化，我们用$\mathcal{D}$表示数据，$\mathcal{M}$表示模型，θ 表示参数。应用贝叶斯规则：

$$\begin{aligned}P(\mathcal{M}\mid\mathcal{D})&=\frac{P(\mathcal{D}\mid\mathcal{M})}{P(\mathcal{M})}P(\mathcal{D})\\&=\frac{\int P(\mathcal{D}\mid\mathcal{M}_i,\theta)P(\theta)\,d\theta\,P(\mathcal{M})}{P(\mathcal{D})}\end{aligned}$$

现在选择一个后验概率比较大的模型。计算这个后验概率比较困难；然而通过一系列的近似，可得到下列标准：

$$-L(\mathcal{D};\theta^*)+\frac{p}{2}\log N$$

称为贝叶斯信息标准或 BIC。

16.3.4 描述长度

可用那些本质非统计的标准来选择模型。毕竟，我们在选择模型时可以说出为什么选择这个模型。比较自然的一个标准是选择可以清楚地编码数据集的模型。最小描述长度标准选择那些能够最有效地传输数据集的模型。为了传输数据集，需要对模型的参数编码和传输，然后在给定模型参数的条件，对数据编码和传输。如果数据与模型拟合得不好，那么由于必须对类似噪声的信号编码，后一部分的数据量会比较大。

在实际应用中对标准的改动通常超出了本书的范围。详细内容参见 Rissanen(1983)和 Rissanen(1987),或 Wallace 和 Freeman(1987);有些源于信息论的相似想法,是由 Kolmogorov 所提出的,并且在 Cover 和 Thomas (1991) 有详细阐述。令我们吃惊的是,从分析中我们可以得到 BIC,满足:

$$-L(\mathcal{D};\theta^*)+\frac{p}{2}\log N$$

16.3.5　用于估计偏差的其他方法

模型选择最主要的困难是需要使用一个无法测量的项——模型预测不在训练集中的数据的能力。给定充分大的一个训练集,可以将训练集划分为两个部分,使用其中的一个去拟合模型,另外一个去测试拟合的结果。这种方法称为交叉检验。

可以通过分化数据集,使用交叉检验去确定模型中分量的数目,用分化的一部分去拟合各种各样的模型,用另一部分去选择性能最好的模型。我们期望能用这种方法估计分量的个数,由于参数过多的模型通常会对一个数据集拟合得非常好,但与其他数据集的拟合得非常差。

只对一个数据集使用一种两划分的方法通常会引入不同形式的选择偏差,最安全的一种方式是对多次不同的两划分得到的估计值求取平均。但是当测试集很大时,这种方法并不实用,因为划分的次数非常大。最常用的方法是单留交叉检验法。在这种方法中,对每一组包含 $N-1$ 个训练集的集合拟合一个模型,计算剩下数据点的误差值,并且将这些误差值累积从而获得模型的误差。然后选择能够最小化这个估计值的模型。

据我们所知,这种方法虽然在解决其他问题时常被看做是模型选择的标准方法,但是在拟合的应用领域中还没有被使用过,该方法显然非常适于估计模型的参数。首先,假设我们对一个分量特别少以至于无法精确地描述所获得的图像的模型计算它的误差。在这种情况下,模型的误差通常很大,这是因为对很多像素来说,模型都无法很灵活地去描述那些留下来进行测试的像素。同样,如果选择一个分量较多的模型的话,对那些留下来进行测试的像素来说,模型对它们的预测也很差。

16.4　注释

显然我们认为丢失变量模型是很重要的。Dempster(1977)的统计著作中首先给出了 EM 算法的形式化描述。McLachlan 和 Krishnan(1996)中非常好地总结了这个问题,文献中描述了算法各种各样的变形。例如,没有必要去找到 $Q(\boldsymbol{u};\boldsymbol{u}^{(s)})$ 的最大值,所需要获得的是一个较好的值。又如,可使用随机积分方法去估计期望。

EM 和丢失变量模型

丢失变量模型看起来经常会发生。我们在计算机视觉中所涉及的模型均来自混合模型(所以能够得到完备数据的对数似然函数,这个函数对于丢失的变量来说是线性的),因此我们主要讨论了这种情况。很自然地会想到使用丢失变量模型去解决图像分割问题(Adelson 和 Weiss,1996;Belongie,Carson, Greenspan 和 Malik,1998; Feng 和 Perona, 1998; Vasconcelos 和 Lippman, 1997; Wells, Grimson, Kikinis 和 Jolesz, 1996)。另外还有一些研究考虑的是对多幅图像

去建立模型(例如,运动问题和立体问题)。通常的想法是将集合分解为不同的层次,每层中的元素共享同样的运动模型(Adelson 和 Weiss 1996; Dellaert, Seitz, Thorpe 和 Thrun 2000; Tao, Sawhney 和 Kumar,2000; Wang 和 Adelson,1994 和 Weiss,1997)或具有同样的深度(Baker, Szeliski 和 Anandan,1998; Brostow 和 Essa,1999; Torr, Szeliski 和 Anandan,1999*b*)或具有其他同样的属性。其他感兴趣的研究问题包括透明、镜面等产生的运动(Black 和 Anandan 1996; Darrell 和 Simoncelli 1993; Hsu, Anandan 和 Peleg 1994; Jepson 和 Black 1993; Szeliski, Avidan 和 Anandan 2000)。所得到的表示对于基于内容的绘制的研究非常有效(Shade, Gortler, Li-wei 和 Szeliski 1998)。这是一个混合模型,尽管通常不把它当做一个隐变量问题看(隐变量是像素所位于的层次,或一种等价的说法,某个层次能生成那些像素),但是很可能应该怎样看。我们期望短期内能够出现重要的研究成果。

EM 算法是非常成功的一个推理算法,但它并不是万能的。所遇到的主要困难是局部最大问题。通常如果所遇到的问题丢失的变量非常多的话,就会得到大量的局部最大点。可在正确答案的附近采用优化来解决这个问题。在应用中,在给定丢失变量数目的条件下用 EM 可以解决的许多视觉问题要比它看起来的难度要容易得多,也就意味着可以相对容易地得到解。讨论一下变量丢失问题是多么困难也是很有意义的。

模型选择

模型选择问题并未受到它应该得到的关注。在运动研究中做了许多工作,一般涉及的是应用哪个摄像机模型比较好(正交投影,透视投影等)(Kinoshita 和 Lindenbaum,2000; Maybank和 Sturm,1999 以及 Torr,1997,1999)。

同样,对于距离数据的分割方面也做了一些工作,其中的问题是哪种参数平面使得数据拟合得较好(即,有两个、三个平面等; Bubna 和 Stewart,2000)。在重建问题中,我们必须决定是否存在一个退化的摄像机运动序列(Torr, Fitzgibbon 和 Zisserman,1999*a*)。分割中标准的问题是有多少个分割(Adelson 和 Weiss,1996; Belongie 等 1998; Raja, McKenna和 Gong,1998)。如果我们使用具有预测功能的模型,那么对模型预测计算加权平均是比较好的(贝叶斯方法并不用来做模型选择)(Ripley,1996; Torr 和 Zisserman,1998)。我们只描述了那些目前可用的方法,一个未涉及的重要内容是 Kanatani 的几何信息准则(Kanatani,1998)。

习题

16.1 导出 16.1 节用于分割的表达式。一种可能的修改方式是在估计协方差矩阵的时候采用新的均值。做个实验看看在应用中这种方式有无意义。

16.2 针对用 EM 算法进行背景差分的例子,写出详细的算法步骤。如果采用较复杂的前景模型而不是均匀的自由噪声模型,会有什么帮助吗?

16.3 描述如何使用留一法的交叉验证来选择分割的数目。

编程作业

16.4 设计采用 EM 的背景差分算法。加入一个扰动项来克服 14.11 所描述的高频空间问题是否现实?

16.5 设计使用色彩和位置信息(如果能加入纹理的话会更好一些)的 EM 分割程序来分割图像;使用模型选择项来确定应该把图像分成多少个区域。局部最小的现象带来的影响有多大?

16.6 设计采用 EM 的直线拟合算法,该算法拟合固定数目的直线。考察局部最小现象的作用。一种避免陷入到局部最小的方法是分别从许多不同的点开始运行算法,然后看看哪个能得到最好的拟合结果。这种方式能很好的解决这个问题吗?对于某特定的数据集,要搜索多少个局部最小点才能够得到好的拟合结果?能够使用 Hough 变换进行改善吗?

16.7 加入一个模型选择项来扩充你所设计的直线拟合器,从而使得所设计的模型能够确定有多少条直线拟合给定的数据集。比较采用 AIC 和 BIC 方法的区别。

16.8 在你所设计的 EM 算法中加入一项噪声项,从而所设计的算法能够进行鲁棒拟合。局部最小的数目能带来什么影响?注意,如果噪声产生点的概率比较低的话,那么大部分点都是由直线所产生的点,但是拟合的结果通常比较差。如果噪声产生点的概率比较高的话,那么点是由直线所产生的当且仅当该点和直线拟合得很好。分析一下这个参数对局部最小的数目具有什么样的效果?

16.9 设计一个 RANSAC 拟合器,该拟合器对于给定的数据集,能够拟合任意数目的直线(已知)。如果扩充所设计的拟合器,使得能够选择出最佳的直线数目,还需要加入什么机制?

第 17 章　基于线性动态模型的跟踪

跟踪(tracking)是根据一组给定图像序列,对图像中物体的运动形态进行分析。好的跟踪方法已经有不少应用:

- **运动的捕捉**:如果可以精确地跟踪一个运动的人,就可以得到运动的精确记录。一旦有了这个记录,就能够使用它生成一个模拟程序,例如,我们能够控制一个卡通人物、数以千计的群众场面、甚至虚拟一个精灵。更进一步,甚至能够通过更改运动记录来获得些许不同的运动。这意味着演员可以做出他们自己不想完成的动作。
- **从运动中识别物体**:物体的运动是特征化的。我们能够根据运动特征确认物体,并能确定物体正在进行的动作。
- **监视**:知道物体在做些什么是很有用的。举个例子,机场中不同的卡车以不同的确定方式运动,如果它们不是这样运动,就可以确认出现了问题。类似地,存在一些在特定的环境中不该发生的运动特征(比如,车辆不应停在行车道内)。跟踪能够帮助计算机系统监视行为并在检测到问题时发出警报。
- **定位**:跟踪研究的一个重要部分是明确导弹的射击目标和击中目标。一般来说,本文献主要描述使用雷达或红外信号(而非视觉)来跟踪,但是基本原理是相同的——如何从已有序列判断物体的未来位置?我们应该瞄准哪里?

在典型的跟踪问题中,我们已有一个描述物体运动的模型和从一系列图像中得到的观测结果。这些观测结果可能是一些图像点的位置、一些图像区域的位置和瞬态或者其他,它们并不一定相关,一些来自感兴趣的物体而另一些可能来自其他对象甚至噪声。

跟踪问题完全可以视为推理问题。运动的物体有一些内部状态,并在每一帧图像中被观测。我们需要有效地合成观测结果来估计物体的状态。这里有两种情况。运动和观测如果都是线性的,那么跟踪的推理问题是直接的,并且有一个标准的解决方法;而非线性运动情况下,即使是运动系统中非常微小的非线性情况都会带来巨大的影响。结果,推理会变得困难,甚至在一般情况下不能实现。当状态空间维数较低时,一般能够使用有效算法进行计算。因为对非线性运动的跟踪是一个技术性问题,我们在本书的网站上为它独立设置了一章。在本章中,我们集中讨论线性动态跟踪的问题。17.1 节概述了跟踪作为一个推理问题,17.2 节讨论线性运动和卡尔曼滤波器,17.5 节设计了一些跟踪问题应用实例。跟踪人的运动是当前最热门的跟踪应用,涉及到的一些非线性运动的讨论见本书网站上的这一章中。

17.1　把跟踪作为一个抽象的推理问题

本章大部分内容讨论跟踪算法。特别地,我们认为跟踪问题是一个概率推理问题。技术上的关键困难在于,在给定观测下始终保持对物体位置的后验概率的准确表示,并且高效率。我们用一些内部状态来对物体建模;物体在第 i 帧的状态记为 $\boldsymbol{X}_i$。大写字母表示是一个随机

变量——当我们讨论这个变量的特定数值时，使用小写字母。第 i 帧图像中的观测结果是变量 $\boldsymbol{Y}_i$ 的数值；记这个观测值为 $\boldsymbol{y}_i$，有时为了强调，我们记做 $\boldsymbol{Y}_i = \boldsymbol{y}_i$。这里有三个主要问题：

- **预测**：如何根据 $\boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}$ 的观测值，获得对第 i 帧状态的预测？为了解决这个问题，需要获得 $P(\boldsymbol{X}_i \mid \boldsymbol{Y}_0 = \boldsymbol{y}_0, \cdots, \boldsymbol{Y}_{i-1} = \boldsymbol{y}_{i-1})$ 的表示。
- **数据相关**：根据第 i 帧图像中获得的某些观测信息，能够获取物体的状态。特别地，用 $P(\boldsymbol{X}_i \mid \boldsymbol{Y}_0 = \boldsymbol{y}_0, \cdots, \boldsymbol{Y}_{i-1} = \boldsymbol{y}_{i-1})$ 表示这些观测信息。
- **校正**：获得相关观测信息 $\boldsymbol{y}_i$ 后，需要计算 $P(\boldsymbol{X}_i \mid \boldsymbol{Y}_0 = \boldsymbol{y}_0, \cdots, \boldsymbol{Y}_i = \boldsymbol{y}_i)$。

17.1.1　独立假设

跟踪问题的解决离不开以下假设：

- **当前状态仅仅依赖于前一状态**：一般地，我们要求

$$P(\boldsymbol{X}_i \mid \boldsymbol{X}_1, \cdots, \boldsymbol{X}_{i-1}) = P(\boldsymbol{X}_i \mid \boldsymbol{X}_{i-1})$$

  下面我们将会看到，这个假设极大地简化了算法的设计。而且，如果能够很好地解释 $\boldsymbol{X}_i$，这个条件限制不是很严格，这一点将在下一节中说明。
- **当前观测仅仅依赖当前状态**：我们假设 $\boldsymbol{Y}_i$ 在给定 $\boldsymbol{X}_i$ 下相对于其他测量条件独立。这意味着

$$P(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \cdots \boldsymbol{Y}_k \mid \boldsymbol{X}_i) = P(\boldsymbol{Y}_i \mid \boldsymbol{X}_i) P(\boldsymbol{Y}_j, \cdots, \boldsymbol{Y}_k \mid \boldsymbol{X}_i)$$

  这也不是一个严格的或有争议的假设，但是它产生重要的简化。

这些假设意味着跟踪问题的推理结构是一个隐马尔可夫模型（状态和观测可能建立在连续域中）。可以将本章同 23.4 节中描述隐马尔可夫模型在识别中的应用进行对比。

17.1.2　用推理进行跟踪

我们用诱导的方法进行讨论。首先假设已知 $P(\boldsymbol{X}_0)$，这是在缺乏任何证据时的预测。现在校正它是很容易的：获得 $\boldsymbol{Y}_0$ 的值 $\boldsymbol{y}_0$ 后，能够得到

$$\begin{aligned} P(\boldsymbol{X}_0 \mid \boldsymbol{Y}_0 = \boldsymbol{y}_0) &= \frac{P(\boldsymbol{y}_0 \mid \boldsymbol{X}_0) P(\boldsymbol{X}_0)}{P(\boldsymbol{y}_0)} \\ &= \frac{P(\boldsymbol{y}_0 \mid \boldsymbol{X}_0) P(\boldsymbol{X}_0)}{\int P(\boldsymbol{y}_0 \mid \boldsymbol{X}_0) P(\boldsymbol{X}_0) d\boldsymbol{X}_0} \\ &\propto P(\boldsymbol{y}_0 \mid \boldsymbol{X}_0) P(\boldsymbol{X}_0) \end{aligned}$$

这一切只是贝叶斯公式，计算或是忽略比例式的常数取决于是否需要。现在假设我们得到 $P(\boldsymbol{X}_{i-1} \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1})$ 的表达式。

预测　要预测的是表达式

$$P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1})$$

由独立假设能够得到

$$
\begin{aligned}
P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) &= \int P(\boldsymbol{X}_i, \boldsymbol{X}_{i-1} \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) d\boldsymbol{X}_{i-1} \\
&= \int P(\boldsymbol{X}_i \mid \boldsymbol{X}_{i-1}, \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) P(\boldsymbol{X}_{i-1} \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) d\boldsymbol{X}_{i-1} \\
&= \int P(\boldsymbol{X}_i \mid \boldsymbol{X}_{i-1}) P(\boldsymbol{X}_{i-1} \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) d\boldsymbol{X}_{i-1}
\end{aligned}
$$

校正 校正是指获得表示

$$P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_i)$$

由独立假设能够得到

$$
\begin{aligned}
P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_i) &= \frac{P(\boldsymbol{X}_i, \boldsymbol{y}_0, \cdots, \boldsymbol{y}_i)}{P(\boldsymbol{y}_0, \cdots, \boldsymbol{y}_i)} \\
&= \frac{P(\boldsymbol{y}_i \mid \boldsymbol{X}_i, \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) P(\boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1})}{P(\boldsymbol{y}_0, \cdots, \boldsymbol{y}_i)} \\
&= P(\boldsymbol{y}_i \mid \boldsymbol{X}_i) P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) \frac{P(\boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1})}{P(\boldsymbol{y}_0, \cdots, \boldsymbol{y}_i)} \\
&= \frac{P(\boldsymbol{y}_i \mid \boldsymbol{X}_i) P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1})}{\int P(\boldsymbol{y}_i \mid \boldsymbol{X}_i) P(\boldsymbol{X}_i \mid \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}) d\boldsymbol{X}_i}
\end{aligned}
$$

17.1.3 小结

关键的算法问题是寻找相关的概率密度表示：要求我们的目的足够准确，计算简单迅速。最简单的情况是线性运动，观测模型是线性的，噪声模型是高斯模型（17.2 节）。数据相关将在 17.4 节中讨论，目前一些跟踪系统的例子将在 17.5 节中介绍。非线性引入一些不易解决的问题，我们将在网站上的教材中讨论一些现有的解决方法。

17.2 线性动态模型

线性变换和高斯概率密度之间有很好的联系。实际的效果是，如果我们关注于线性动态模型和线性观测模型，两者都有附加的高斯噪声，那么我们关注的所有概率密度都是高斯分布的，并且还可以通过使用小技巧直接确认所处理的高斯分布，避免复杂的积分运算。

在可能出现的最简单动态模型中，可以通过下面的方法确定状态：首先与已知矩阵（可能与帧有关）相乘，再加上一个均值为零、协方差已知的正态分布随机变量。类似地，观测值也可以通过首先与已知矩阵（可能与帧有关）相乘，再加上一个普通均值为 0、协方差确定的随机变量获得。我们使用符号

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma)$$

表示 $\boldsymbol{x}$ 是均值为 $\boldsymbol{\mu}$，协方差为 Σ 的随机变量的值。如果 $\boldsymbol{x}$ 是一维的，一般写做 $x \sim N(\mu, v)$，x 的标准差是 $\sqrt{v}$。则可描述动态模型如下：

$$x_i \sim N(\mathcal{D}_i \boldsymbol{x}_{i-1}; \Sigma_{d_i})$$

$$y_i \sim N(\mathcal{M}_i \boldsymbol{x}_i; \Sigma_{m_i})$$

注意到帧与帧之间，协方差和矩阵都可能有所不同。尽管这个模型看起来有局限性，但事实上是非常有用的；下面我们将介绍一些通常情况下的模型。

17.2.1　漂移点

假设 $\boldsymbol{x}$ 表示一个点的位置。如果 $\boldsymbol{D}_i = Id$，那么这个点进行随机运动——它的原有位置加上高斯噪声扰动便得到新位置。这种形态的运动好像是对静态的物体进行跟踪，所以看起来不太有用。但是某些情况下没有更好的运动模型时，这个模型得到了广泛应用，一般假设随机成分非常大并希望能逐渐摆脱它。

这个模型同时说明了观测矩阵$\mathcal{M}$的特征。要记住的最重要一点是，不需要观测每一个点在每一步时状态的每一方面。举例来说，假设三维空间的一个点：当前位置$\mathcal{M}_{3k} = (0,0,1)$，$\mathcal{M}_{3k+1} = (0,1,0)$，$\mathcal{M}_{3k+2} = (1,0,0)$每隔三帧分别对点的 z、y 或 x 坐标进行一次观测。即使在一个给定帧只观测它所在位置的一个成分，这时仍然能够期待跟踪上这个点。如果有足够的观测，我们能够重建状态——称状态是可观测的。在练习中，我们将探索这种可观测性。

17.2.2　恒速度

假设向量 $\boldsymbol{p}$ 表示点的位置，$\boldsymbol{v}$ 表示点移动的恒定速度。这样，$\boldsymbol{p}_i = \boldsymbol{p}_{i-1} + (\Delta t)\boldsymbol{v}_{i-1}$且 $\boldsymbol{v}_i = \boldsymbol{v}_{i-1}$。这表明可以把位置和速度放入一个单独的状态向量中，并且在我们的模型中使用它(见图 17.1)。特别地

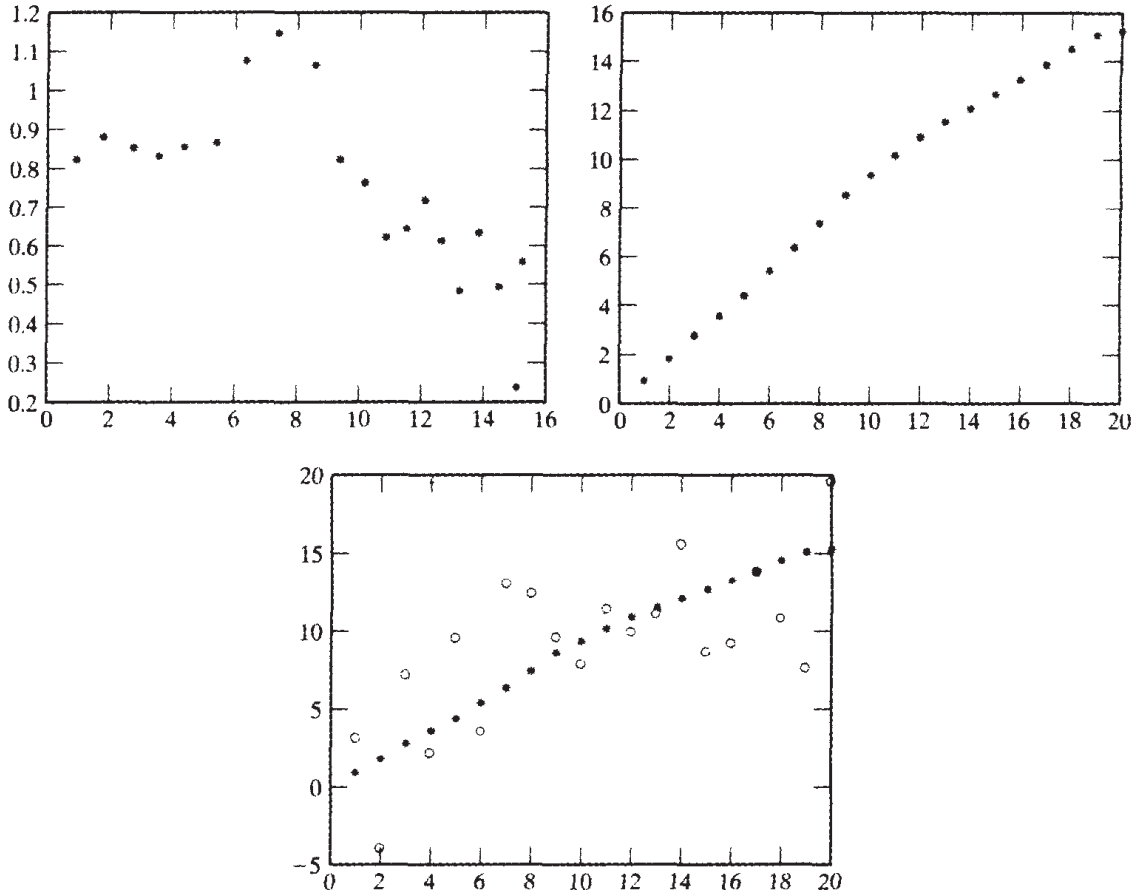


图 17.1　点的恒速直线动态模型。在这种情况，状态空间是二维的——位置坐标和速度坐标。左上角的图是一个状态图，每一个星号是一个不同状态。纵坐标表明随着横坐标速度有小的变化。这些变化只由动态模型的随机分量造成，所以速度就是一个常量加上一个随机变化。右上角的图表示状态的位置分量随时间的变化。可以看到速度可以粗略地视为常数。下图叠加了观测结果(用圆圈表示)。假定观测只对位置坐标进行。可以看出，这样并不明显地影响跟踪的效果

$$x = \begin{Bmatrix} p \\ v \end{Bmatrix}$$

和

$$\mathcal{D}_i = \begin{Bmatrix} Id & (\Delta t)Id \\ 0 & Id \end{Bmatrix}$$

再一次可以注意到，不需要观测整个状态向量来获得有用的观测。例如，在很多情况下，只要注意到在大部分情况下有

$$\mathcal{M}_i = \begin{Bmatrix} Id & 0 \end{Bmatrix}$$

也就是说，只观测点的位置。因为我们知道模型中运动的速度恒定，所以可以期待使用这些观测能够很好地估计整个状态向量。

17.2.3 恒加速度

假设向量 $\boldsymbol{p}$ 表示点的位置，向量 $\boldsymbol{v}$ 表示速度，向量 $\boldsymbol{a}$ 表示加速度恒定的点的加速度。在这种情况下，$\boldsymbol{p}_i = \boldsymbol{p}_{i-1} + (\Delta t)\boldsymbol{v}_{i-1}$，$\boldsymbol{v}_i = \boldsymbol{v}_{i-1} + (\Delta t)\boldsymbol{a}_{i-1}$ 和 $\boldsymbol{a}_i = \boldsymbol{a}_{i-1}$。同样的道理，能够把位置、速度和加速度一起放入一个单一的状态向量中，我们的模型使用该向量验证了这一点（见图 17.2）。特别地

$$x = \begin{Bmatrix} p \\ v \\ a \end{Bmatrix}$$

和

$$\mathcal{D}_i = \begin{Bmatrix} Id & (\Delta t)Id & 0 \\ 0 & Id & (\Delta t)Id \\ 0 & 0 & Id \end{Bmatrix}$$

再一次可以注意到，不需要观测整个状态向量来获得有用的观测。例如，在很多情况下，我们只要注意

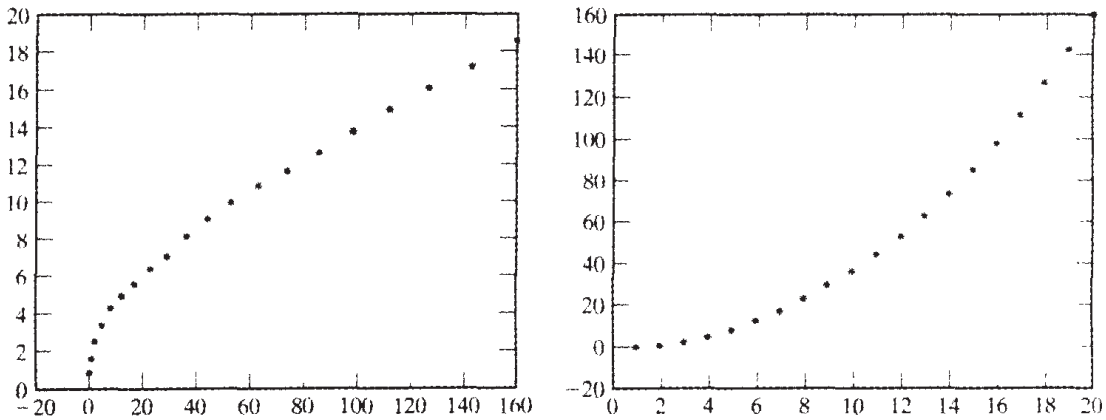


图 17.2 点的恒加速直线运动模型。左图中列出了状态的前两个分量，x 坐标代表位置，y 坐标代表速度。和预期的一样，曲线的关系是 (t^2, t)。右图是位置随时间的变化，点从初始点出发，速度迅速增加

$$\mathcal{M}_i = \{ Id \quad 0 \quad 0 \}$$

也就是说，只观测点的位置。因为我们知道模型中运动的加速度恒定，所以可以期待使用这些观测能够很好的估计整个状态矢量。

17.2.4 周期运动

假设一个点在直线上周期运动。一般地，它的位置 p 满足如下的微分方程

$$\frac{d^2p}{dt^2} = -p$$

把速度记为 v，把位置和速度合并到一个新向量 $\boldsymbol{u}$ 中，$\boldsymbol{u} = (p, v)$；这个方程就转化为一阶线性微分方程：

$$\frac{d\boldsymbol{u}}{dt} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \boldsymbol{u} = \mathcal{S}\boldsymbol{u}$$

假设使用前向欧拉方法对方程进行积分，步长设为 Δt：

$$\begin{aligned} \boldsymbol{u}_i &= \boldsymbol{u}_{i-1} + \Delta t \frac{d\boldsymbol{u}}{dt} \\ &= \boldsymbol{u}_{i-1} + \Delta t \mathcal{S}\boldsymbol{u}_{i-1} \\ &= \begin{pmatrix} 1 & \Delta t \\ -\Delta t & 1 \end{pmatrix} \boldsymbol{u}_{i-1} \end{aligned}$$

我们可以把这个式子视为状态方程，也可以使用不同的积分方法。如果使用不同的积分方法，会得到一系列 $\boldsymbol{u}_{i-1}, \cdots, \boldsymbol{u}_{i-n}$ 的表达式，需要把 $\boldsymbol{u}_{i-1}, \cdots, \boldsymbol{u}_{i-n}$ 合并到一个状态向量中，并且适当地安排矩阵(详见练习)。这种方法适合于点在平面、三维空间等上的运动(详见练习)。

17.2.5 高阶模型

考察恒速度模型的另一个方法是，扩展状态向量以规避 $P(\boldsymbol{x}_i \mid \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{i-1}) = P(\boldsymbol{x}_i \mid \boldsymbol{x}_{i-1})$ 的要求。只要我们愿意使用第 $i-1$ 个点和第 $i-2$ 个点的位置，就可以仅仅使用点所在的位置来描速恒速度模型。特别地，记坐标为 $\boldsymbol{p}$，有如下的表示：

$$P(\boldsymbol{p}_i \mid \boldsymbol{p}_1, \cdots, \boldsymbol{p}_{i-1}) = N(\boldsymbol{p}_{i-1} + (\boldsymbol{p}_{i-1} - \boldsymbol{p}_{i-2}), \Sigma_{d_i})$$

这个模型假设 $\boldsymbol{p}_i$ 和 $\boldsymbol{p}_{i-1}$ 之间的差异与 $\boldsymbol{p}_{i-1}$ 和 $\boldsymbol{p}_{i-2}$ 的差异相同，或者说，速度除了一个随机因素外是恒定的。同样的理论也可以运用于恒加速度模型，表现形式相应改为用 $\boldsymbol{p}_{i-1}$，$\boldsymbol{p}_{i-2}$ 和 $\boldsymbol{p}_{i-3}$ 表示。

我们在恒速度模型中把速度向量(用 $\boldsymbol{p}_{i-1} - \boldsymbol{p}_{i-2}$ 表示)增添到位置向量上得到状态向量。类似地，在恒加速度模型中增添速度向量、加速度向量得到状态向量，加速度向量用 $(\boldsymbol{p}_{i-1} - \boldsymbol{p}_{i-2}) - (\boldsymbol{p}_{i-2} - \boldsymbol{p}_{i-3})$ 表示。点所在的新位置，有时还取决于 $\boldsymbol{p}_{i-4}$ 点，甚至跟踪过程中更早的点。为了表示这样的运动，需要把状态向量扩展到合适的大小。注意到形象化表示模型运动特征是困难的。有两个方法可以获取 $\mathcal{D}_i$；第一个是把运动中的已知条件写下，正如前面已经做的工作；第二个需要根据数据进行学习。

17.3 卡尔曼滤波

线性动态模型的重要特征在于需要处理的所有条件概率分布都是正态分布，具体说来，$P(\boldsymbol{X}_i \mid \boldsymbol{y}_1, \cdots, \boldsymbol{y}_{i-1})$和 $P(\boldsymbol{X}_i \mid \boldsymbol{y}_1, \cdots, \boldsymbol{y}_i)$都是正态分布，这意味着它们相对较容易描述——要做的就是在预测和校正步骤中记录均值和协方差。特别地，对于预测和估计，这个模型更新均值和协方差的表达式是一个相对简单的过程。

17.3.1 一维状态向量的卡尔曼滤波

此时的动态模型为

$$x_i \sim N(d_i x_{i-1}, \sigma_{d_i}^2)$$

$$y_i \sim N(m_i x_i, \sigma_{m_i}^2)$$

我们需要维护 $P(X_i \mid y_0, \cdots, y_{i-1})$和 $P(X_i \mid y_0, \cdots, y_i)$的表示。在每种情况下，因为是正态分布，只需描述均值和标准差。

符号 把 $P(X_i \mid y_0, \cdots, y_{i-1})$的均值表示为 $\overline{X}_i^-$，把 $P(X_i \mid y_0, \cdots, y_i)$的均值表示为 $\overline{X}_i^+$——使用上标表示第 i 帧观测之前和观测之后 X_i 的置信度。类似的，我们把 $P(X_i \mid y_0, \cdots, y_{i-1})$的标准差表示为 σ_i^-，把 $P(X_i \mid y_0, \cdots, y_i)$的标准方差表示为 σ_i^+。在每种情况下，假定 $P(X_{i-1} \mid y_0, \cdots, y_{i-1})$已知，即 $\overline{X}_{i-1}^+$和 σ_{i-1}^+已知。

积分的技巧 使用正态分布的重要原因在于它们有很好的积分特性。通过改变变量获得不同参数的积分。现有的符号进行变换时可能有些困难，我们记做

$$g(x; \mu, v) = \exp\left(-\frac{(x-\mu)^2}{2v}\right)$$

在这里不再写出常数，为了方便，描述方差(记为 v)，而不是标准差。这种表达方式使一些变换写起来很简便，尤其是可以得到

$$g(x; \mu, v) = g(x-\mu; 0, v)$$

$$g(m; n, v) = g(n; m, v)$$

$$g(ax; \mu, v) = g(x; \mu/a, v/a^2)$$

我们还需要下面的式子

$$\int_{-\infty}^{\infty} g(x-u; \mu, v_a) g(u; 0, v_b)\, du \propto g(x; \mu, v_a^2 + v_b^2)$$

有几种方法证明这一结论：最简单的方法是在表中查找，复杂些的方法是直接考虑卷积，更为复杂的方法是考察两个独立随机变量之和。进一步推导得

$$g(x; a, b) g(x; c, d) = g\left(x; \frac{ad+cb}{b+d}, \frac{bd}{b+d}\right) f(a, b, c, d)$$

其中，f 的形式并不重要，只要记住它不是 x 的函数，练习中会告诉如何证明该等式。

预测　由于已经知道

$$P(X_i \mid y_0, \cdots, y_{i-1}) = \int P(X_i \mid X_{i-1}) P(X_{i-1} \mid y_0, \cdots, y_{i-1})\, dX_{i-1}$$

则有

$$\begin{aligned}
P(X_i \mid y_0, \cdots, y_{i-1}) &= \int P(X_i \mid X_{i-1}) P(X_{i-1} \mid y_0, \ldots, y_{i-1}) dX_{i-1}) \\
&\propto \int_{-\infty}^{\infty} g(X_i; d_i X_{i-1}, \sigma_{d_i}^2) g(X_{i-1}; \overline{X}_{i-1}^{+}, (\sigma_{i-1}^{+})^2) dX_{i-1} \\
&\propto \int_{-\infty}^{\infty} g((X_i - d_i X_{i-1}); 0, \sigma_{d_i}^2) g((X_{i-1} - \overline{X}_{i-1}^{+}); 0, (\sigma_{i-1}^{+})^2) dX_{i-1} \\
&\propto \int_{-\infty}^{\infty} g((X_i - d_i(u + \overline{X}_{i-1}^{+})); 0, (\sigma_{d_i})^2) g(u; 0, (\sigma_{i-1}^{+})^2) du \\
&\propto \int_{-\infty}^{\infty} g((X_i - d_i u); d_i \overline{X}_{i-1}^{+}, \sigma_{d_i}^2) g(u; 0, (\sigma_{i-1}^{+})^2) du \\
&\propto \int_{-\infty}^{\infty} g((X_i - v); d_i \overline{X}_{i-1}^{+}, \sigma_{d_i}^2) g(v; 0, (d_i \sigma_{i-1}^{+})^2) dv \\
&\propto g(X_i; d_i \overline{X}_0^{+}, \sigma_{d_i}^2 + (d_i \sigma_{i-1}^{+})^2)
\end{aligned}$$

其中我们运用了前面给出的变换，并两次改变了变量。所有这一切意味着

$$\overline{X}_i^{-} = d_i \overline{X}_{i-1}^{+}$$

$$(\sigma_i^{-})^2 = \sigma_{d_i}^2 + (d_i \sigma_{i-1}^{+})^2$$

校正　我们有

$$\begin{aligned}
P(X_i \mid y_0, \cdots, y_i) &= \frac{P(y_i \mid X_i) P(X_i \mid y_0, \cdots, y_{i-1})}{\int P(y_i \mid X_i) P(X_i \mid y_0, \cdots, y_{i-1}) dX_i} \\
&\propto P(y_i \mid X_i) P(X_i \mid y_0, \cdots, y_{i-1})
\end{aligned}$$

我们已知表示 $P(X_i \mid y_0, \cdots, y_{i-1})$ 的 $\overline{X}_i^{-}$ 和 σ_i^{-}。

使用早先给出的高斯分布的符号，可以得到

$$\begin{aligned}
P(X_i \mid y_0, \cdots, y_i) &\propto g(y_i; m_i X_i, \sigma_{m_i}^2) g(X_i; \overline{X}_i^{-}, (\sigma_i^{-})^2) \\
&= g(m_i X_i; y_i, \sigma_{m_i}^2) g(X_i; \overline{X}_i^{-}, (\sigma_i^{-})^2) \\
&= g\left(X_i; \frac{y_i}{m_i}, \frac{\sigma_{m_i}^2}{m_i^2}\right) g(X_i; \overline{X}_i^{-}, (\sigma_i^{-})^2)
\end{aligned}$$

通过与等式的模式匹配，可以得到

$$X_i^{+} = \left(\frac{\overline{X}_i^{-} \sigma_{m_i}^2 + m_i y_i (\sigma_i^{-})^2}{\sigma_{m_i}^2 + m_i^2 (\sigma_i^{-})^2}\right)$$

$$\sigma_i^{+} = \sqrt{\left(\frac{\sigma_{m_i}^2 (\sigma_i^{-})^2}{(\sigma_{m_i}^2 + m_i^2 (\sigma_i^{-})^2)}\right)}$$

算法 17.1 一维卡尔曼滤波使用给定动态模型跟踪一维状态变量时，更新所涉及的各种分布均值和协方差的估计。

动态模型：

$$x_i \sim N(d_i x_{i-1}, \sigma_{d_i})$$

$$y_i \sim N(m_i x_i, \sigma_{m_i})$$

初始假设：$\overline{x_0^-}$ 和 σ_0^- 已知

更新公式：预测

$$\overline{x_i^-} = d_i \overline{x_{i-1}^+}$$

$$\sigma_i^- = \sqrt{\sigma_{d_i}^2 + (d_i \sigma_{i-1}^+)^2}$$

更新公式：校正

$$x_i^+ = \left(\frac{\overline{x_i^-} \sigma_{m_i}^2 + m_i y_i (\sigma_i^-)^2}{\sigma_{m_i}^2 + m_i^2 (\sigma_i^-)^2} \right)$$

$$\sigma_i^+ = \sqrt{\left(\frac{\sigma_{m_i}^2 (\sigma_i^-)^2}{(\sigma_{m_i}^2 + m_i^2 (\sigma_i^-)^2)} \right)}$$

图 17.3 显示了卡尔曼滤波跟踪一个恒速度模型的结果，图 17.4 显示卡尔曼滤波跟踪一个恒加速度模型的结果。

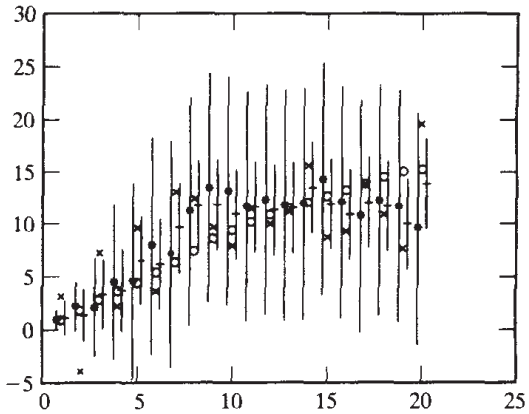


图 17.3 运动点在直线上匀速运动的卡尔曼滤波（与图 17.1 对比）。第 i 步所对应的状态用圆圈表示。* 号表示 $\overline{x_i^-}$，可以看到略微偏左，表示是观测前的估计。× 号表示观测值，+ 号表示 $\overline{x_i^+}$，略微偏右。* 号和 + 号之间的竖线是使用变量估计获得观测前和观测后的三倍标准差线，值得注意的是，当观测受到较大的噪声干扰时，竖线收缩的并不多（与图 17.4 对比）

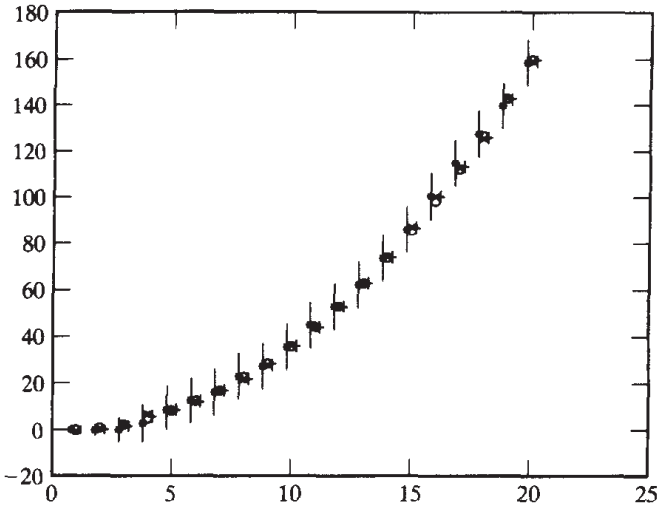


图 17.4　运动点在直线上匀加速运动的卡尔曼滤波(与图 17.2 对比)。第 i 步所对应的状态用圆圈表示。* 号表示 $\bar{x}_i^-$，可以看到略微偏左，表示是观测前的估计。× 号表示观测值，+ 号表示 $\bar{x}_i^+$，略微偏右。* 号和 + 号之间的竖线是使用变量估计获得观测前和观测后的三倍标准差线，值得注意的是，当观测没有受到太大噪声干扰时，竖线在获得观测值后快速收缩

17.3.2　一般状态向量的卡尔曼更新公式

在一维跟踪中，根据正态分布的特性，不需要进行什么积分。这一点在任意维情况下也适用。但是猜想积分的过程，比 17.3.1 节中的例子更加复杂而精细。我们忽略过于繁琐的符号推导——这对有兴趣的人来说是一项艰巨而有意思的工作(首先指出第一个点的恒等关系，其他点以次类推)——而只是简单的在算法 17.2 中给出。

算法 17.2　在使用给定动态模型跟踪某些确定维状态变量时，卡尔曼滤波更新所涉及的各种分布均值和协方差的估计。

动态模型：

$$x_i \sim N(\mathcal{D}_i x_{i-1}, \Sigma_{d_i})$$

$$y_i \sim N(\mathcal{M}_i x_i, \Sigma_{m_i})$$

初始假设：$\bar{x}_0^-$ 和 Σ_0^- 已知

更新公式：预测

$$\bar{x}_i^- = \mathcal{D}_i \bar{x}_{i-1}^+$$

$$\Sigma_i^- = \Sigma_{d_i} + \mathcal{D}_i \sigma_{i-1}^+ \mathcal{D}_i$$

更新公式：校正

$$\mathcal{K}_i = \Sigma_i^- \mathcal{M}_i^T \left[\mathcal{M}_i \Sigma_i^- \mathcal{M}_i^T + \Sigma_{m_i}\right]^{-1}$$

$$\bar{x}_i^+ = \bar{x}_i^- + \mathcal{K}_i \left[y_i - \mathcal{M}_i \bar{x}_i^-\right]$$

$$\Sigma_i^+ = [Id - \mathcal{K}_i \mathcal{M}_i] \Sigma_i^-$$

17.3.3 前向-后向平滑

$P(X_i \mid y_0, \cdots, y_i)$并不是 X_i 的最好表示方法，这是因为它无法考虑到点的未来行为。特别地，所有 y_i 之后的观测将影响我们对 X_i 的表示。这是因为这些未来的度量可能会与当前的估计相矛盾——也许未来的实际运动同位置的估计有所不同。但无论如何，$P(X_i \mid y_0, \cdots, y_i)$是在第 i 步能得到的最好的估计。

如何处理这个观测依赖于应用环境。如果应用要求迅速估计点的位置——例如观测相对驶来的一辆汽车——我们能做的事情并不多。如果离线操作——例如在法庭辩论中用，需要从录像带中得到事物最好的估计——我们能够使用所有数据点，这样就能够得到 $P(X_i \mid y_0, \cdots, y_N)$。另外一个选择是立即得到一个粗略的估计，并且能够使用后若干步的改进估计。这意味着需要表示 $P(X_i \mid y_0, \cdots, y_{i+k})$——需要等到第 $i+k$ 帧，但是它却能改善 $P(X_i \mid y_0, \cdots, y_i)$的估计。

介绍一种后向估计 现在有

$$
\begin{aligned}
P(X_i \mid y_0, \cdots, y_N) &= \frac{P(X_i, y_{i+1}, \cdots, y_N \mid y_0, \cdots, y_i) P(y_0, \cdots, y_i)}{P(y_0, \cdots, y_N)} \\
&= \frac{P(y_{i+1}, \cdots, y_N \mid X_i, y_0, \cdots, y_i) P(X_i \mid y_0, \cdots, y_i) P(y_0, \cdots, y_i)}{P(y_0, \cdots, y_N)} \\
&= \frac{P(y_{i+1}, \cdots, y_N \mid X_i) P(X_i \mid y_0, \cdots, y_i) P(y_0, \cdots, y_i)}{P(y_0, \cdots, y_N)} \\
&= P(X_i \mid y_{i+1}, \cdots, y_N) P(X_i \mid y_0, \cdots, y_i) \alpha
\end{aligned}
$$

其中，

$$
\alpha = \left(\frac{P(y_{i+1}, \cdots, y_N) P(y_0, \cdots, y_i)}{P(X_i) P(y_0, \cdots, y_N)} \right)
$$

这一项看起来会带来一个潜在的问题；实际上，可以使用一个巧妙的方法避免混乱。这种方法的关键点在于把 $P(X_i \mid y_0, \cdots, y_i)$(我们知道如何获得)和 $P(X_i \mid y_{i+1}, \cdots, y_N)$结合起来。我们实际上也知道如何获得 $P(X_i \mid y_{i+1}, \cdots, y_N)$，仅仅只要使用反向动态模型和 X_i 的预测表示，在时间上反向使用卡尔曼滤波。

合并表示 现在 X_i 有两种表示：一种通过综合到 y_i 为止的所有观测用前向滤波器获得；一种通过综合 y_i 之后的所有观测用后向滤波器获得。需要把这两种表示结合起来。我们并不直接计算 α(看起来非常复杂)，与之相反，如果把后向滤波器获得的估计看成另一种观测，就能够得到答案。具体说来，我们得到了 X_i 产生的新观测——也就是后向滤波器的结果——要同前向滤波器的估计进行结合。由卡尔曼滤波等式的特点，可以知道怎样合并由观测得到的估计。

需要引进一些符号。给前向滤波的估计结果加上 f 的上标，给后向滤波的估计结果加上 b 的上标。把 $P(X_i \mid y_0, \cdots, y_N)$的均值记为 $\overline{X}_i^*$，把 $P(X_i \mid y_0, \cdots, y_N)$的协方差记为$\Sigma_i^*$。把 $\overline{X}_i^b$ 表示作为X_i 表示的观测，均值为 $\overline{X}_i^{b,-}$ 且协方差为$\Sigma_i^{b,-}$——这里使用减号因为第 i 帧观测不能被两次使用，后向滤波预测 X_i 时使用$y_N \cdots y_{i+1}$。这个观测要与 $P(X_i \mid y_0, \cdots, y_i)$——均值为 $\overline{X}_i^{f,+}$ 且协方差为$\Sigma_i^{f,+}$——合并(代入卡尔曼滤波时，这些起观测值前表示的作用，因为

观测值现在是 $\overline{X}_i^{b,-}$)。

替换到卡尔曼等式中,可以得到

$$\mathcal{K}_i^* = \Sigma_i^{f,+}\left[\Sigma_i^{f,+} + \Sigma_i^{b,-}\right]^{-1}$$

$$\Sigma_i^* = [I - \mathcal{K}_i]\,\Sigma_i^{+,f}$$

$$\overline{X}_i^* = \overline{X}_i^{f,+} + \mathcal{K}_i^*\left[\overline{X}_i^{b,-} - \overline{X}_i^{f,+}\right]$$

在下面的算法 17.3 中,可以看到,一个小小的操作(见练习)可以产生一个更简化的形式。在图 17.5 中,可以看到前向 - 后向估计得到一个的结果明显不同。

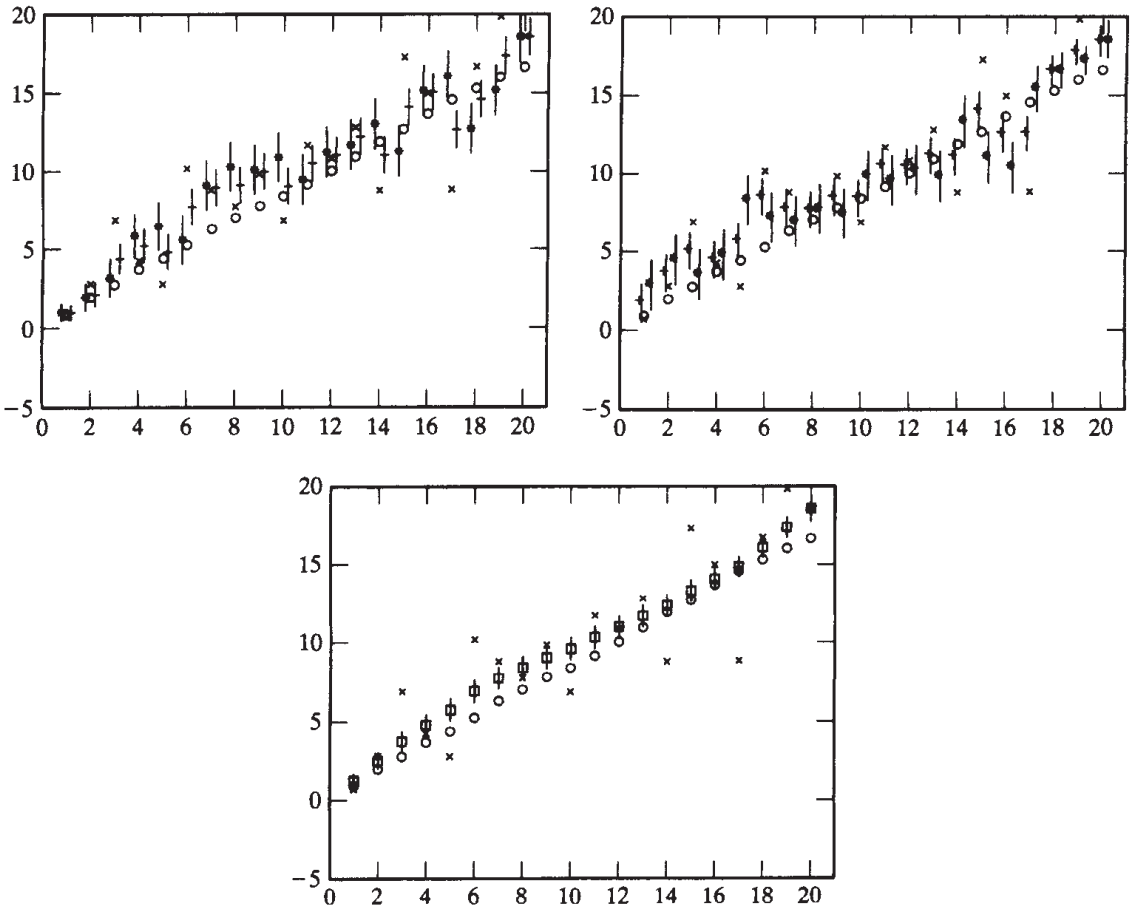


图 17.5　点在直线上匀速运动的前向 - 后向估计动态模型。这里画出了位置随时间的变化情况。左上图中是前向估计,和前面同样的约定:圆圈表示状态,× 号表示数据,* 号表示预测,+ 号表示校正,竖线表示估计的标准差。预测估计略微滞后观测,矫正估计略微超前观测。可以注意到观测是受噪声影响的。右上图中是后向估计。时间反向进行(虽然我们把曲线画在同一个轴上)所以预测估计略微超前观测,矫正估计略微滞后观测。把前向滤波的最终校正估计作为先验。同样,竖线给出每个变量的标准差。下图中是合并的前向 - 后向估计。正方形代表状态的估计。可以看到估计得到了明显改善

预测　在典型视觉应用中,进行实时前向跟踪。这导致了一种不便的非对称:我们清楚地知道物体的起点,却只能模糊地知道它的终点[也就是说,有一个很好的 $P(\boldsymbol{x}_0)$的预测,但在前向 - 后向

滤波中,很难预测 $P(\boldsymbol{x}_N)$]。一种选择是使用 $P(\boldsymbol{x}_N \mid \boldsymbol{y}_0,\cdots,\boldsymbol{y}_N)$作为预测。这样做不能令人信服,因为这个概率分布不能完全正确地反映对 $P(\boldsymbol{x}_N)$的先验置信度——我们使用了所有的观测得到它。结果只能说明这个分布低估了 $\boldsymbol{x}_N$ 的不确定性,并且引出了一个明显低估了后面状态协方差的前向-后向估计。另外一个方法是选择使用前向滤波器产生的均值,但是增大了协方差;结果是一个高估了后面状态协方差的前向-后向估计(对比图 17.5 和图 17.6)。

算法 17.3 前向-后向算法合并对状态的前向估计和后向估计得到改进的估计结果。

前向滤波:通过卡尔曼滤波得到 $P(\boldsymbol{X}_i \mid \boldsymbol{y}_0,\cdots,\boldsymbol{y}_i)$,它们是 $\overline{\boldsymbol{X}}_i^{b,+}$ 和 $\sum_i^{b,+}$。

后向滤波:通过向后使用卡尔曼滤波得到 $P(\boldsymbol{X}_i \mid \boldsymbol{y}_{i+1},\cdots,\boldsymbol{y}_N)$,它们是 $\overline{\boldsymbol{X}}_i^{b,-}$ 和 $\sum_i^{b,-}$。

合并前向估计和后向估计:把后向估计视为 $\boldsymbol{X}_i$ 的新观测值,插入卡尔曼滤波公式得到。

$$\Sigma_i^* = \left[(\Sigma_i^{f,+})^{-1} + (\Sigma_i^{b,-})^{-1}\right]^{-1}$$

$$\overline{\boldsymbol{X}}_i^* = \Sigma_i^*\left[(\Sigma_i^{f,+})^{-1}\overline{\boldsymbol{X}}_i^{f,+} + (\Sigma_i^{b,-})^{-1}\overline{\boldsymbol{X}}_i^{b,-}\right]$$

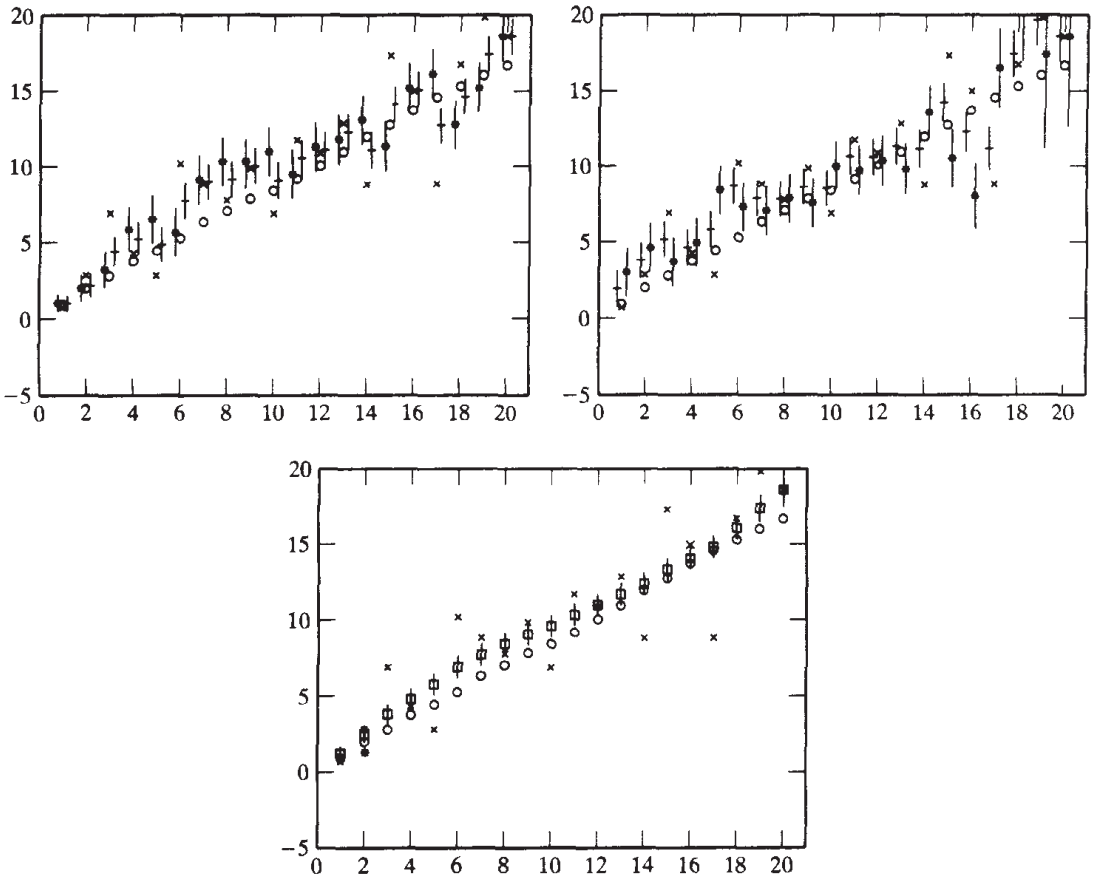


图 17.6 使用先验扩散对匀速直线运动点模型前向-后向估计最终位置的影响。这里画出了位置随时间的变化情况。左上图中是前向估计,和前面同样的约定:圆圈表示状态,× 号表示数据,* 号表示预测,+ 号表示校正,竖线表示估计的标准差。预测估计略微滞后观测,矫正估计略微超前观测。可以注意到观测是受噪声影响的。右上图中是后向估计。时间反向进行(虽然我们把曲线画在同一个轴上)所以预测估计略微超前观测,矫正估计略微滞后观测。同样,竖线给出每个变量的标准差。下图中是合并的前向-后向估计。正方形代表状态的估计。可以看到估计得到了明显改善

并不是所有的应用都是非对称的。例如，在法庭辩论中研究的录像带，我们可以手工起动进行前向和后向跟踪并且对每种情况下提供很好的预测。如果可能，能够获得更多的信息帮助保持一致性——前向跟踪结束的地方更接近于后向跟踪开始的地方。

间隔之间的平滑　虽然前向－后向平滑的公式假定了后向滤波的起点在最后一个数据点，但是在前向滤波之前若干步启动后向滤波是容易做到的。如果这样做，就能实时获得状态的估计（在观测后很快就能获得），以及在观测的若干步后得到一个改进的估计。这一点有时是非常有用的。进一步，如果能够假设比较远的未来对估计的影响远远小于比较近的未来对估计的影响，那么这是一个能从后向滤波获得显著改进的有效方法。注意，需要注意这里的后向滤波的先验估计，可以选用前向估计并且一定程度上扩大它的协方差。

17.4　数据相关

并不是观测的每个方面都能传递物体被跟踪所需的信息。事实上，我们根本没有说 y_i 究竟有什么。通常，有的观测提供有用的信息，有的观测并没有被跟踪物体有关的信息（通常称为杂物）。

决定哪种观测提供有用信息的过程通常称为数据相关。特别地，我们希望把一系列观测映射到有关目标的跟踪上，可能会忽略其他一些观测——甚至全部。这个问题的主要工作同使用雷达信号跟踪运动物体（所有敌人的飞机、导弹等）相关。特别地，在每一个给定时间里，有很多雷达信号——我们关注更新观测的物体的运动时，不需要考虑（其他）信号来自哪个物体。正如我们可以看到的，跟踪算法是复杂的，但是并不十分困难。数据相关大概是视觉应用中的最大困难，也不经常在文献中讨论。我们希望这种情况能通过实际应用而得到改善。我们的讨论局限在单一移动物体的情况下。这里的问题相当于图像中的某些像素对于物体有信息量，有些却没有——那么应该使用哪些像素进行跟踪呢？

17.4.1　选择最近——全局最近邻

在最简单的情况下，需要跟踪一个在杂乱环境下运动的单个物体。例如，在一个确定的缓慢变动的环境中跟踪一个球。我们把图像分割成区域，期望球能够产生一个区域，背景部分会随时间而改变，这应该是合理的。直觉地看，球和背景区域不容易混淆，因为我们对球的运动有很强的模型。这意味着如果一个新的背景区域既与球的区域混淆，又与运动模型混淆，运气就太差了。这就给数据相关提出了一个非常有效的策略：第 r 个区域提供观测 y_i^r，所选择的区域应该具有最优值的

$$
\begin{aligned}
P(Y_i = y_i^r \mid y_0, \cdots, y_{i-1}) &= \int P(Y_i = y_i^r \mid X_i, y_0, \cdots, y_{i-1}) P(X_i \mid y_0, \cdots, y_{i-1})\, dX_i \\
&= \int P(Y_i = y_i^r \mid X_i) P(X_i \mid y_0, \cdots, y_{i-1})\, dx_i
\end{aligned}
$$

使用卡尔曼滤波很容易确定 $P(Y_i = y_i \mid y_0, \cdots, y_{i-1})$。我们知道如何由 X_i 得到 Y_i——选取一个均值为 $\overline{X}_i^-$ 且协方差为 Σ_i^- 的正态随机变量，对它执行一个线性操作 $\mathcal{D}_i$，加上一些其他的随机变量。线性操作的输出均值为 $\mathcal{D}_i \overline{X}_i^-$ 协方差为 $\mathcal{D}_i \Sigma_i^- \mathcal{D}^T$。加上一个均值为 0 协方差为

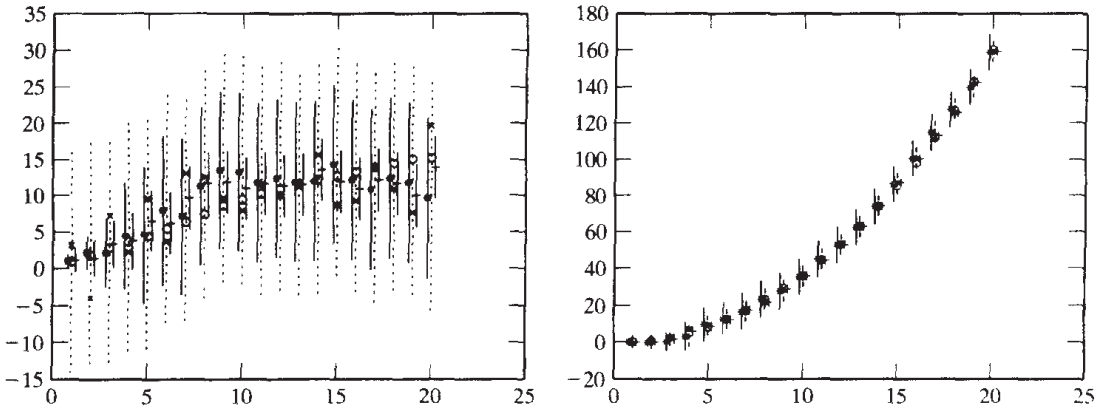
Σ_{m_i} 的随机变量，则结果的均值为

$$\mathcal{D}_i \overline{X}_i^-$$

协方差为

$$\mathcal{D}_i \Sigma_i^- \mathcal{D}^T + \Sigma_{m_i}$$

在图 17.7 中，表示了不同动态模型下卡尔曼滤波预期观测位置的变动范围。



17.7 左图用卡尔曼滤波对匀速直线运动的点求数据相关，右图是匀加速直线运动点用卡尔曼滤波求数据相关。请与图 17.1 的匀速直线运动和图 17.2 的匀加速直线运动进行比较。我们使用了图 17.3 的约定。图中叠加了观测的三标准差的直线（虚的直线经过状态）。这是根据观测前的状态估计和我们对观测过程变化的知识得到的。注意观测在这些窗口的范围内

注意这种方法的鲁棒性取决于动态模型的精确度。如果动态模型是严密的，任何容易同被跟踪物体混淆的物体的运动必须比被跟踪物体的运动更接近预测观测。这意味着一个偶然的错误辨识不会产生严重的问题，因为很难找到一个区域既同预测观测相同又能严重脱离该动态模型。在图 17.8 中，卡尔曼滤波在每一步中选择最佳观测进行跟踪；它不是总能正确地表示点的位置，但这种状态估计却总是非常好的。

注意到这里使用的只是同预测一致的观测。这可能带来危险也可能不会：这样的方法很容易跟踪一个根本不存在的物体，或者去跟踪一个根本没有任何观测的物体。如果动态模型只能提供很弱的预测（也就是说，物体并不这样运动或者 Σ_{d_i} 一直很大），我们将碰到棘手的问题，因为我们需要依赖于观测。这些问题出现的原因在于错误能够累加——很容易对一个错误的点进行长时间的跟踪，错误跟踪的时间越长，重新正确跟踪的难度越大。图 17.9 是一个在这种情况下卡尔曼滤波变得混乱的情况。

17.4.2 选通和概率数据相关

我们再一次假设跟踪一个在杂乱环境下运动的单个物体，并以在一个固定的或缓慢变动的环境中跟踪一个球为例。我们不再采用选择同预测观测最接近区域的方法，而是采取排除过于不同的区域而保留所有剩下的，并根据它们同预测的相似性对它们进行加权的方法。

第一步叫做选通（gating）。排除同预测观测过于不同的区域。“过于不同”同应用直接相关：如果排除的区域过多，可能没有区域能够保留下来。通常当观测同预期值的均值间的距离达到标准方差的一定数值的倍数时——通常是 3，将这些区域排除。如果被跟踪的物体有不

止一个动态特性，则需要更复杂的策略。例如，军事飞行器经常具有高速机动性能。这种情况下应有几重选通，选择能够通过最严格选通的所有观测。

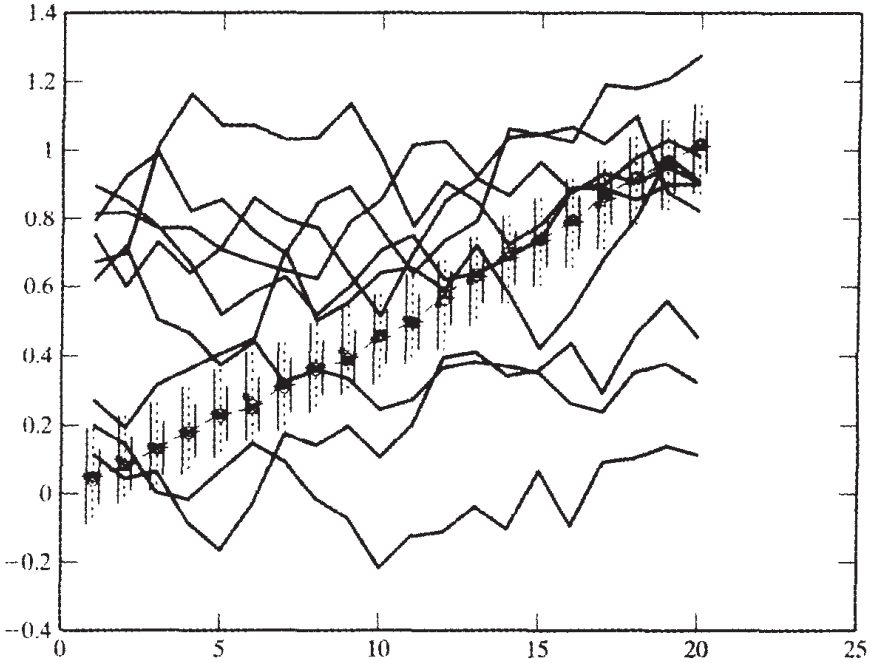


图 17.8　卡尔曼滤波中对点位置的预测能够识别"好的"观测。使用卡尔曼滤波识别一个匀速直线运动的点，每一步的 Σ_{d_i} 值很小。这里大概有10个漂移点。图中表示了漂移点（用实线表示）和被跟踪点的时间分布。被跟踪点的轨迹是一条虚线，轨迹上每个点的观测用方框表示。使用图17.3中的符号规定（也就是说，状态用圆圈表示 i。* 号表示 $\bar{x}_i^-$，它比观测略微偏左，说明估计在观测前做出。× 号表示观测值，+ 号表示 $\bar{x}_i^+$，它比观测略微偏右。* 号和 + 号之间的竖线是使用变量估计获得观测前和观测后的三倍标准差线）。滤波器选取每一步中使 $P(y_i^j \mid y_0, \cdots, y_{i-1})$ 取最大值的观测值；注意到选取的并不是每一步的实际观测值（也就是说 × 号不一定在方框内），但是它对状态保持了很好的估计（也就是说 + 号接近圆圈）

下一步被称为概率数据相关(PDA)。假设在选通中，有 N 个区域，每个区域提供一个测量向量 y_i^k（上标表示区域）。我们需要处理一系列可能的假定：或者没有区域来自物体，称之为 h_0，或者区域 k 来自物体，称之为 h_k。这时观测表示为

$$\mathrm{E}_h\left[y_i\right] = \sum_j P(h_j \mid y_0, \cdots, y_{i-1}) y_i^j$$

其中，期望值从假定空间获取（这也是为什么把它赋予 h 的下标）。没有来自物体的观测的概率取决于观测过程的细节。例如，在 Blackman 和 Popoli(1999)一书的第 4 章中，给出了一个雷达系统的例子。在其他情况下，根据训练结果，选择得到较好结果的参数值。假设已经计算出或学习到这个参数，记为 β。同样还需假设物体或者没有观测到或者只有一个观测值。则

$$\begin{aligned} P(h_j \mid y_0, \cdots, y_{i-1}) &= \int P(h_j \mid X_i) P(X_i \mid y_0, \cdots, y_{i-1}) dX_i \\ &= P(Y_i = y_i^j \mid y_0, \cdots, y_{i-1}) P(\text{被观测到的物体}) \\ &= P(Y_i = y_i^j \mid y_0, \cdots, y_{i-1})(1-\beta) \end{aligned}$$

在下面的讨论中，我们把 $P(h_j \mid \mathbf{y}_0, \cdots, \mathbf{y}_{i-1})$ 记为 p_j。实际上，这种方法经常用在卡尔曼滤波中。这样，我们把观测

$$\mathbf{y}_i' = \sum_j p_j \mathbf{y}_i^j$$

记录到卡尔曼更新公式中。注意没有得到观测的情况在 p_j 表达式中以参数 β 的形式出现，但是更新中应该选取什么样的观测的不确定性，在协方差更新中也应该出现。因此将协方差更新公式修改为

$$\begin{aligned}\Sigma_i^+ = (1-\beta)[Id - \mathcal{K}_i \mathcal{M}_i]\Sigma_i^- + \beta\Sigma_i^- \\ + \mathcal{K}_i\left[\sum_j p_j(\mathcal{H}_i\bar{\mathbf{x}}_i^- - \mathbf{y}_i^j)(\mathcal{H}_i\bar{\mathbf{x}}_i^- - \mathbf{y}_i^j)^T - \mathbf{y}_i'(\mathbf{y}_i')^T\right]\end{aligned}$$

其中，第一项是根据观测是否有效的概率对标准卡尔曼滤波的更新加权，第二项处理所有的观测都无效的情况，第三项表示了相应的不确定性。

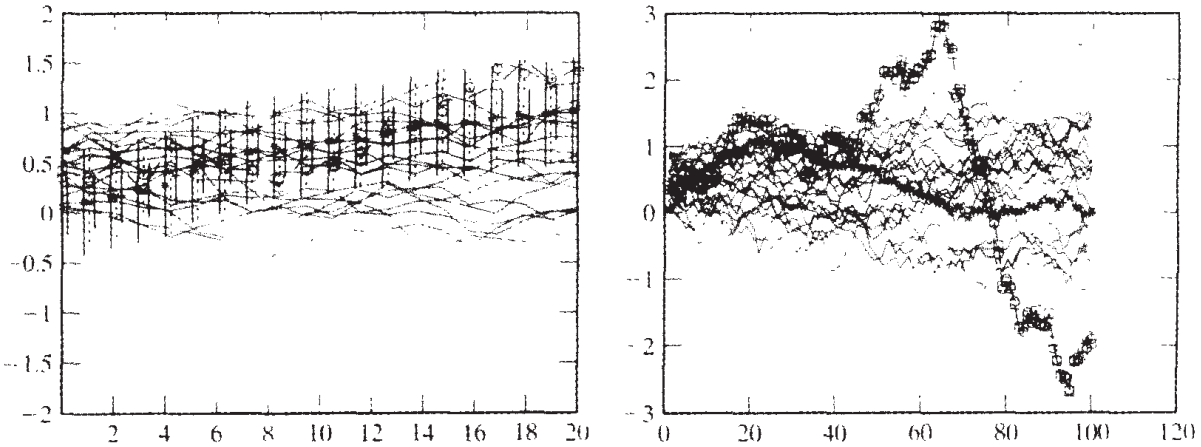


图 17.9 当动态模型约束不充分时，选择提供最好 $P(\mathbf{y}_i' \mid \mathbf{y}_0, \cdots, \mathbf{y}_{i-1})$ 的观测可能导致灾难性的后果。左图中是在有20个漂移点的背景中周期性直线运动点的卡尔曼滤波的20步。在此再次使用图17.3中的符号规定。现在 Σ_{d_i} 在每一步中都比较大，所以很容易跟踪错误的观测。看起来似乎滤波很好跟踪了状态，但实际上右图（给出100步）表明，跟踪很快就失败了

17.5 应用和例子

跟踪是一种有很多可能应用的技术。主要存在三大应用课题。

- 车辆跟踪系统能够报告交通堵塞、事故、危险或驾驶者的非法行为。交通堵塞报告对驾驶者（可以相应调整驾驶计划）和决策者（能够安排移除阻塞道路的交通线路）是非常有用的。事故报告用于紧急情况报警；如果跟踪系统能够识别车牌，能够使用危险和非法行为报告向车辆的使用者发送传票。
- 监视系统报告人们正在做的事情，通常用于针对那些做着本不该做的事情的人。例如，警察可能希望知道哪些观众向球场中投掷瓶子，或者几个银行被抢劫前是否有相同的人光顾了它们。海关希望准确地知道谁在来往于国内和国外的机场。

- 人机交互系统利用人们的动作操纵各种设备。例如，通过观察人们的行为，起居室自己决定行为——是否温柔的灯光和轻音乐是适合的。当你挥手时，电视机自己改变频道。计算机观察你在白板上写下的内容，并且在需要时将内容记录下来。

现在，最成熟的应用是车辆跟踪。这些系统在很大规模的环境中可靠地运行。这里，简单地介绍车辆跟踪系统，网站上有一章专门讨论对人的跟踪问题。

17.5.1　车辆跟踪

根据固定摄像机录像跟踪车辆的系统能够用于预测车流和交通量；目标是尽可能快地报告和预防交通问题。许多系统能够成功地跟踪车辆。最重要的问题是自动初始化跟踪。这里讨论的两种系统中，问题的解决方法是非常不同的。Sullivan，Baker，Worrall，Attwood 和 Remagnino(1997)在每一帧图像中构造感兴趣的区域的集合(ROI)。因为摄像机是固定的，可以选择覆盖每条车道的区域为相关区域(见图 17.10)；这意味着几乎所有汽车必须以确定的方向通过相关区域(如果汽车在 ROI 中选择改变道路，会有一些小问题，但是这些问题可能会忽略)。这个系统接着在 ROI 观测表示车辆存在的特征边缘信号(见图 17.10)。这些信号可能会略有混淆——典型的情况是有的车在车辆的前沿进入 ROI 时，跟踪被启动，有的是当车辆完全进入 ROI 时，跟踪被启动，第三个在车辆接近离开时被启动——因为有些车辆边缘很容易和其他的混淆。

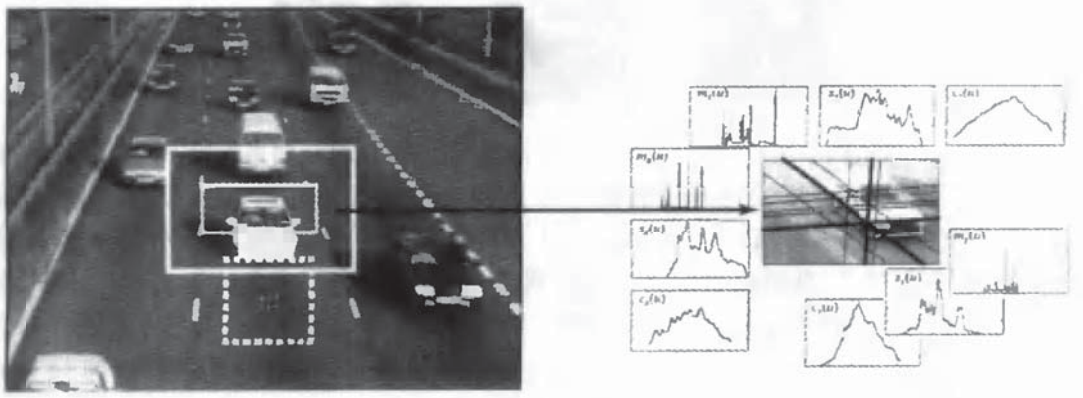


图 17.10　Sullivan 等人通过路面上的固定摄像机对车辆进行跟踪。左图中对应于路面获取一系列特殊区域。通过在一个特殊ROI中获取特征边缘信号启动跟踪；这些信号被投影到三维坐标轴——如果投影到坐标轴上的边缘同期望形式有明显相关性，跟踪就会被启动(如右图所示)

每一个被启动的跟踪要跟踪若干帧，其间要对一个定性评价进行累计——本质上是对未来位置预测的正确程度的估计。如果这个定性评价分足够高，跟踪被接受为一个假设。在每个假设中构建一个时间和空间上排他的区域，从而在这个区域中只能有一个跟踪；如果区域交迭，选择的是高质量的跟踪。要求排他的区域不会互相交迭，一般是因为两辆车不能同时占用同一个区域。一旦跟踪通过了这些测试，它在何时通过另一个 ROI 就能够预测出来。在合适时间将 ROI 同预测车辆表现的模版进行对比最终确认或是放弃跟踪。一般被启动的跟踪中只有比较少的达到这个阶段(见图 17.11)。

另外一种启动车辆跟踪的方法是跟踪车辆的各个特征，并且把这些特征组合为可能的车辆。Beymer，McLauchlan，Coifman 和 Malik(1997)成功地使用了这个策略。因为路面是平坦的，摄像机也是固定的，联系路面和摄像机的单应性能够得以确定。这个单应性能用于决定点之

间的距离，只有点之间的距离不随时间改变时，这些点能同时出现在同一个汽车上。他们的系统跟踪角点，使用二阶矩阵(8.3.3 节)进行标识，使用卡尔曼滤波进行跟踪，使用一个简单的图像提取算法对点进行组合：每个特征跟踪是一个顶点，边缘代表跟踪间的组合关系。当一个新的特征出现时——跟踪因此启动——将它同这一帧中附近的每一个特征跟踪用一个边缘连接起来。如果在某些帧，被跟踪的点之间的距离改变太大，这个边缘被丢弃。车辆离开图像的附近定义为一个离开区域。当跟踪到达离开区域时，有联系的部分被定义为车辆。这个组合算法是成功的，不仅在实验图像中(见图 17.12)而且在对长时间交通序列的参数估计也有效(见图 17.13)。

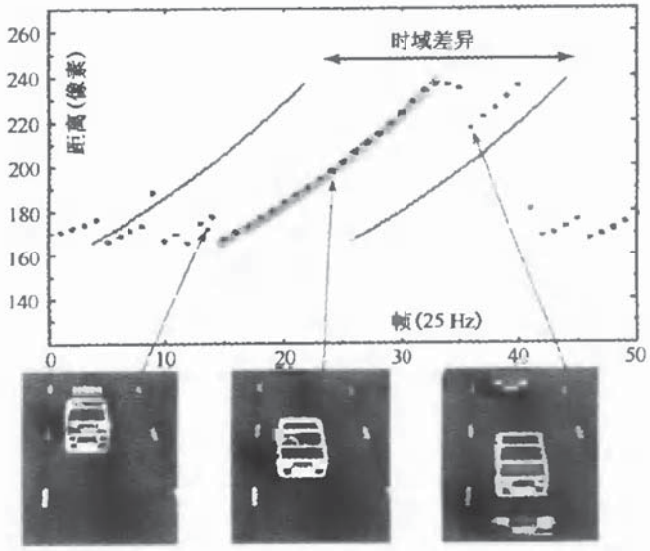


图 17.11 在 Sullivan 等人的系统中，通过将跟踪的预测同观测进行对比，如果有足够的合格观测，跟踪将继续。跟踪之间相互排斥：车辆进入区域的底部之前，系统必须决定接受哪个跟踪。这可通过对比跟踪预测和另一个 ROI 情况来完成。图像中显示了一系列的跟踪(横轴表示时间，纵轴表示位置)。注意到典型的错误跟踪(在跟踪开始时，一辆车的车头和另一辆车的车尾被连接起来当做一辆车)可能会迅速消失；有效的跟踪(和它的排他区域)会显现出来。如果两个跟踪试图互相排除，成功的将是跟踪质量最高的一方

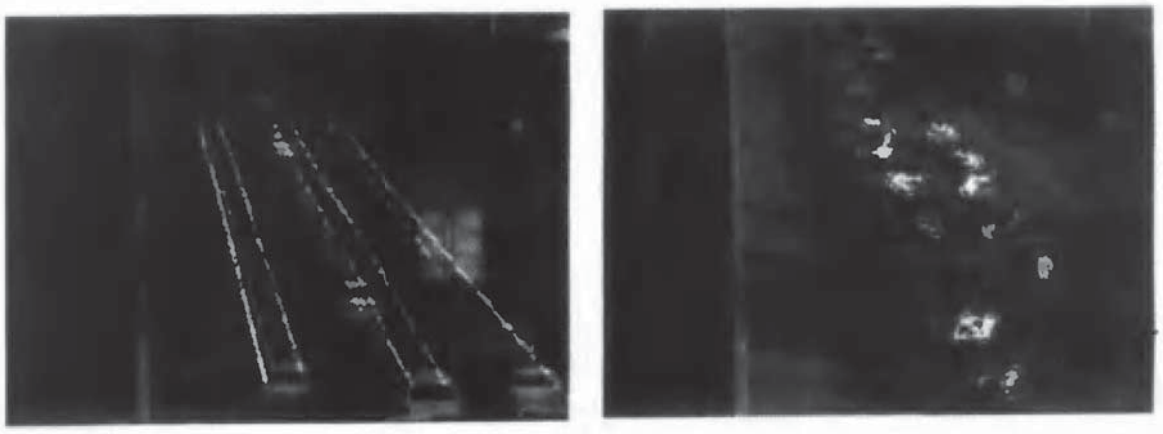


图 17.12 左图显示了 Beymer 等人系统对每个车辆的跟踪。这些跟踪由对角点使用卡尔曼滤波得到。因为摄像机相对于路面的位置已知，摄像机变换能够将平行于路面的平面上的点转换。这意味着能够确定保持常数距离的点对。右图为这些点对的组合结果(这些点对被假设为车辆)

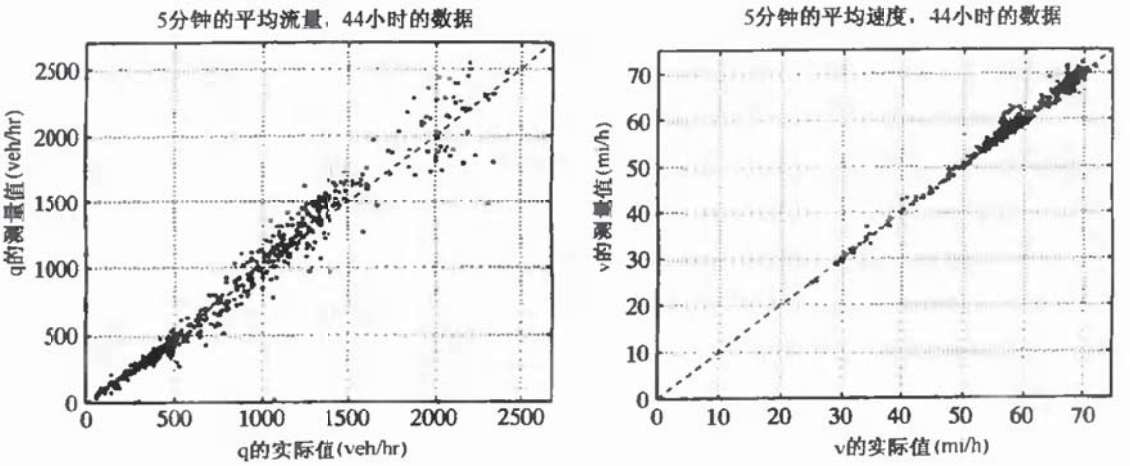


图 17.13　Beymer 等人的系统能够对交通流量和交通速度进行很精确的估计。左图中是对路面状况的估计相对于真值的散布图；右图中是对速度情况的估计相对于真值的散布图

路面到摄像机的变换能够提供许多信息。我们已经使用这个方法确定点是否在物体上(通过估算地面速度并进行速度对比)。这样可以集合物体特征。一个物体一旦被跟踪，就能够使用这种变换估算空间布局和遮挡情况。进一步讲，我们可以在运动车辆中跟踪汽车。这种情况有两个因素要考虑：首先是摄像机平台的运动(叫做自运动)；其次，其他车辆的运动。Ferryman，Maybank 和 Worrall(2000)通过匹配帧与帧之间道路情况估计自运动(见图 17.14)。使用单应和自运动的估计，我们能够把移动车辆上观测到的所有车辆情况重建到道路坐标系统，从而获得对其他运动车辆的跟踪(见图 17.14)。

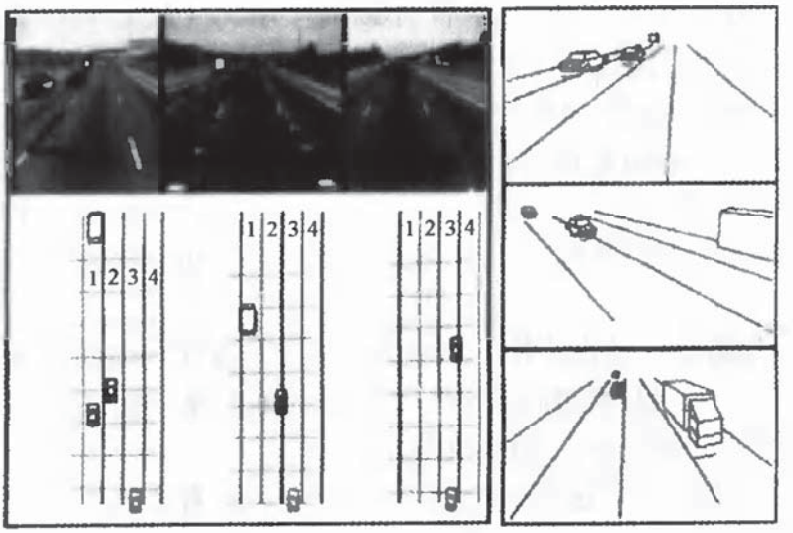


图 17.14　一旦知道了地面的单应，使用移动摄像机平台对其他车辆进行的跟踪就可以同使用本地固定摄像机平台形成坐标对应关系。左图表示了交通地图的详细重建。此外，我们能够使用地面固定物体的移动(比如白线)估计摄像机平台的运动。这一切意味着我们能够解释交通地图，举个例子：预测摄像机平台和其他车辆之间即将发生碰撞；同时能融入其他摄像机获得的交通情况(右图所示)

17.6 注释

卡尔曼滤波是一种有用的工具，它在不同领域不断以不同的形态重现。通常，非线性运动能够足够好地表示为能够运用卡尔曼滤波的线性运动。有兴趣的读者可以参看 Chui(1991)，Staff of the Analytical Sciences Corporation(1974)以及 West 和 Harrison(1997)的文章。

我们没有讨论如何获得线性动态模型。如果已知模型的顺序，使用自然状态空间，可行的观测模型，问题很容易解决；否则，情况将变得复杂——控制理论中有一个系统辨识领域讨论这个问题。推荐读者阅读 Ljung(1995)。

习题

17.1 假设有一个模型 $\boldsymbol{x}_i = \mathcal{D}_i \boldsymbol{x}_{i-1}$ 和 $y_i = \boldsymbol{M}_i^T \boldsymbol{x}_i$。对每一个 i，观测 y_i 是一个一维向量(也就是说，是一个数字)，$\boldsymbol{x}_i$ 是一个 k 维向量。如果状态能够根据任何包括 k 个观测的序列重建，我们称模型是可观测的。

(a) 证明这个要求等价于要求矩阵

$$[\boldsymbol{M}_i \mathcal{D}_i^T \boldsymbol{M}_{i+1} \mathcal{D}_i^T \mathcal{D}_{i+1}^T \boldsymbol{M}_{i+2} \cdots \mathcal{D}_i^T \cdots \mathcal{D}_{i+k-2}^T \boldsymbol{M}_{i+k-1}]$$

是满秩的。

(b) 三维空间的漂移点，$\mathcal{M}_{3k} = (0,0,1)$，$\mathcal{M}_{3k+1} = (0,1,0)$，$\mathcal{M}_{3k+2} = (1,0,0)$是可观测的。

(c) 在任何方向的匀速运动点，如果只使用观测矩阵报告点的位置，则为可观测的。

(d) 在任何方向的匀加速运动点，如果只使用观测矩阵报告点的位置，则为可观测的。

17.2 漂移动态模型中沿直线运动的点。特别地，有关系 $x_i \sim N(x_{i-1}, 1)$。起始于 $x_0 = 0$。

(a) 它的平均速度(注意，速度是有符号的)?

(b) 它的平均速率(注意，速率是有符号的)?

(c) 当它距离起始点距离超过 2 时，平均经过了多少步(也就是说，步数的期望值是多少)?

(d) 当它距离起始点距离超过 10 时，平均经过了多少步(也就是说，步数的期望值是多少)?

(e) (这个问题需要一定的思考)假设我们有两个非交叉的时间间隔，一个长度为 1，另一个长度为 2；当步数趋向无穷时比例式(时间间隔一中平均时间百分比)/(时间间隔二中平均时间百分比)的极限是多少?

(f) 或许你已经得到上一题的比例，现在进行模拟，看看多久之后这个比例看起来接近正确的答案。

17.3 我们说

$$g(x; a, b) g(x; c, d) = g\left(x; \frac{ad + cb}{b + d}, \frac{bd}{b + d}\right) f(a, b, c, d)$$

证明等式正确。最简单的方法是取对数，对分数重排列。

17.4 假设运动模型为

$$x_i \sim N(d_i x_{i-1}, \sigma_{d_i}^2)$$

$$y_i \sim N(m_i x_i, \sigma_{m_i}^2)$$

(a) $P(x_i \mid x_{i-1})$均值为 $d_i x_{i-1}$，方差为 $\sigma_{d_i}^2$ 的正态分布。那么 $P(x_{i-1} \mid x_i)$是什么？

(b) 如何使用卡尔曼滤波获得 $P(\boldsymbol{x}_i \mid \boldsymbol{y}_{i+1}, \cdots, \boldsymbol{y}_N)$的表示？

编程作业

17.5　实现一个二维卡尔曼滤波跟踪一段简单图像序列中的物体。建议使用背景差分跟踪前景点。状态空间应该包括点的位置、速度、方向——通过计算二阶矩阵获得——和它的角速度。

17.6　已有对背景的估计后，卡尔曼滤波能够通过跟踪照明变化和摄像机增益变化来改进背景差分。实现这样的卡尔曼滤波；它能够提供多大改善？注意照明变化的一个可行模型是背景同一个接近 1 的噪声项相乘——通过取对数能够将其转化为线性模型。

第五部分　高层视觉:几何方法

- 第 18 章　基于模型的视觉
- 第 19 章　平滑表面及其轮廓
- 第 20 章　外观图
- 第 21 章　距离数据

第18章　基于模型的视觉

本章将物体识别看做一个寻找对应的问题——哪一个图像特征对应哪一个物体的哪个特征？这种识别的简单想法是相当有用的，它很自然地把重点放在物体特征、图像特征以及摄像机模型的关系上。

我们讨论许多使用对应方法的不同算法。这些算法的主要依据是，物体的特征不是散乱在图像中的；如果知道了少量特征的对应关系，那么再找到更大集合的对应关系就更容易了。因为摄像机的位置一般是有规律变化的，并且只有相对很少的自由度。

理解图像特征的位置与物体特征的位置和方向之间的关系可以利用许多实际的线索。在18.6节，我们描述了一个应用，它采用了本章后面章节讲到的技术，用来将医学图像与实际病人对准，使得医生可以看到图像的哪些特征在病人身体的哪些部位。

18.1　初始假设

我们讨论的所有算法都假设，待识别的物体都有一组几何模型；这些模型一般叫做模型库。我们假设，如果关于物体的信息在算法中可用，那么它一定在模型库里。

所有在本章描述的算法都是一个类型的，通常称为假设与测试。每一个算法都要

- 假设一个图像特征集合与物体特征集合对应，然后用这个对应关系产生一个从物体坐标系到图像坐标系的投影框架的猜测。有许多不同的方法。如果知道摄像机的特征参数，那么假设就等同于对物体的位置与方向——方位进行猜想。
- 使用这个投影假设来产生一个物体的图像。这一步一般叫做反向投影(backprojection)。
- 比较图像与所产生的物体图像，如果它们两个足够的相似，那么就接受这个假设。

为了使这个方法(见图18.1)有效，必须产生相对少的假设，并且用相对快速和准确的方法来比较反向投影图与图像。这个比较的过程叫做校验(verification)，而且可能非常不可靠；在18.5节将描述校验方法。一般地，这些校验方法计算某个物体在某个方位出现的假设的分值，我们称之为校验分数(verification score)。

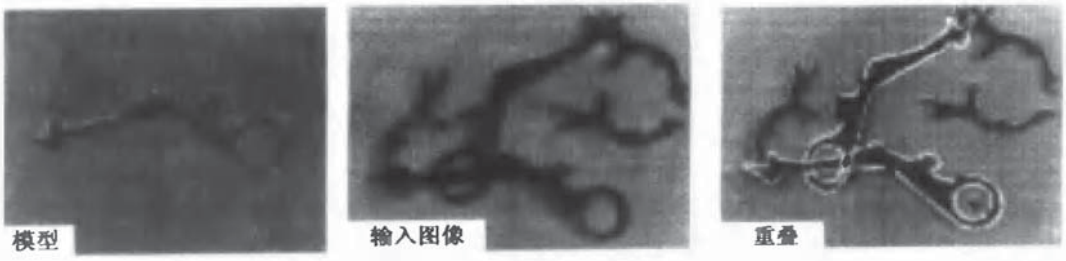


图18.1　对准方法应用在一个平面物体上。左图是一个物体的图像；中图，一个图像包含两个物体，还有一些其他物体(一个流行的术语叫杂物)。在检测出特征点后，搜索结构组之间的对应，在这里是三个点一组找对应；每个对应给出一个从模型到图像的仿射变换。在右图中，成功的对应使得许多模型边缘点与图像边缘点对准。此图中的图像来自关于本论题的最早的论文，并且受到了当时简单复制技术的影响

这个方法对点特征与曲面有效，尽管对于曲面来说细节变得非常复杂。大多数文献处理的物体模型都是由投影类似于点的几何特征组成的。这意味着物体的不同视图用相同特征集(其中某些特征可能被遮盖了)的不同视图表示，而不是不同特征集的视图。我们主要考虑这种情形(到目前为止它是实践中最有用的)。然而在 18.7 节，我们会描述一些在曲面物体图像中获取假设和校验假设的方法。

一般避免对应该匹配哪些特征的问题的细节进行讨论。我们描述的大多数算法都包括某种程度的特征搜寻——很明显，如果对特征的描述恰当，那么这种搜寻量可以减少。例如，如果特征只是简单的图像点，也许是一些边缘曲线交叉的点，那么，所有的点都一样，所以也许需要的搜索量大。如果换一种方法，点用图像的局部表示来描述(即滤波器的输出向量)，那么可能存在的对应就少很多，所以需要的搜索量也少了。

18.1.1　获取假设

算法之间的不同主要是获得假设的机制不同。最显而易见的方法是取图像中所有 M 个几何特征和 L 个物体中每个物体所有的 N 个几何特征，然后枚举物体与图像特征之间所有可能的对应关系(例如，图像特征 3 对应 7 号物体的特征 5，等等)。这是一个可怕的算法，因为枚举的总数相当惊人——$O(LM^N)$。

物体点之间的几何约束限制了空间大小。例如，如果把三维模型匹配到三维数据，我们希望模型的点对与其对应的点对距离相同。任何违反这个规则的对应关系都可以不加考虑，而不用顾及该假设的其余部门是否合理。这种推理可以认为相当于对搜索树剪枝——对一棵被强行剪枝的搜索树进行搜索的方法叫做解释树算法(interpretation tree algorithm)，源自 Grimson 和 Lozano-Perez(1984)。

当三维模型匹配到二维数据时，同样可以应用几何限制。这是因为投影模型的参数通常可以通过相对较少对点的对应确定。一旦这些参数已知，所有其他的投影特征的位置也变得已知——这是一个约束，因为这些特征的位置不再可以随意选择。因此可以利用这些约束，通过从少量对应中决定投影参数，然后用投影模型去预测其他对应(18.2 节描述这种策略，但是它的起源不清楚)。

事实上，由于下面的原因，不需要决定投影参数。一旦建立了少量物体特征与图像特征的对应(称为基集)，摄像机约束就可以用来预测其他图像特征的位置。这意味着如果基集的对应是恰当的，其他图像特征的位置相对于基集是确定的。用相对位置的合理解释，可以获得无关于投影参数的测量结果，然后用它来鉴别物体(18.4 节)。

18.2　通过位姿一致性获取假设

假设有一幅某个物体的图像所用摄像机模型已知，但是模型的参数未知(例如，在用一个已标定的透视摄像机观测物体，但是相对于物体坐标系的外参数未知)。如果假设一个足够大集合的图像特征与足够大集合的物体特征相匹配，那么可以从该假设中获取未知的参数(而且可以绘制物体的余下部分)。这种类型的方法(算法 18.1)叫做位姿一致性方法。我们将用一些例子来描述这一种算法(它的一般形式如图 18.1)。越来越多的人们称这种形式的算法为

对准(alignment);这个术语表示了物体与图像对准的思想。这个名字只是在最近版本的算法上才出现的,在文献中以不同的形式出现。只是在最近,这些不同形式算法之间的相似之处才变得明显起来。

这种类型的方法以不同的形式出现,它的细节取决于所用摄像机的模型以及物体是二维还是三维的。我们把可以用来产生摄像机假设的特征集合称为结构特征组(frame group)(物体与图像都有结构特征组)。

算法 18.1 对准:匹配物体与图像集合以推导摄像机模型

```
对于所有物体结构特征组 O
    对于所有图像结构特征组 F
        对于 F 元素与 O 元素间的所有的对应 C
          使用 F,C,O 来推导摄像机模型中的未知参数
          使用估计出的摄像机模型来绘制物体
          如果绘制与图像一致,那么物体存在
        end
    end
end
```

18.2.1 透视摄像机的位姿一致性

假设一个内参数已知的透视摄像机正在观测模型库里的一个物体。我们在基于物体的坐标系统中展开讨论,在物体的坐标系中,外参数现在只包括摄像机的位置和方向。如果一组在图像坐标系统里的图像特征和在物体坐标系统里的物体特征的对应已知,采用算法 18.1 可以确定摄像机的外参数。一旦外部参数已知,整个摄像机也就已知,那么我们就可以绘制图像的剩余部分了。

对于这个问题,有许多的结构特征组结构特征组可用。例如典型情况,好的结构特征组结构特征组包含几个不同类型的特征(为了减少要搜索的对应的数目)。流行的结构特征组结构特征组包括:

- 三点
- 三个方向(一般叫做三面顶点)和一个点(建立坐标系所必需的)
- 两面顶点(从一个点中射出的两个方向)和一个点

通常情况下,方向靠使用线段来获得,因为非常普遍的一个情况是,线段的一部分会出现,但是准确定位线段的端点很难。

内参数 在文献中,假设摄像机的内参数未知是非常普遍的。实际上问题并没有多大改变,只是需要使用更复杂的结构特征组,但是它确实提供了强制性更强的一致性推理的机会。在包括多于一个物体的图像中,可以对不同的物体使用同样的摄像机内参数。

推理的过程非常简单。首先,使用算法 18.1 识别单个物体,因此每个物体可以得到摄像机的一个相关解。对于每一对识别出的物体,比较它们摄像机解的内参数,如果它们的差别足够大,那么这两个假设不兼容。

18.2.2　仿射和投影摄像机模型

标定透视摄像机是很复杂的，因为外参数中包含旋转。一般可以使用一个允许标定简单一些的摄像机模型，但是其代价是物体识别有更多的歧义性。两个重要的简化有：

- **仿射摄像机**　把一个透视视图建模成仿射变换，然后接一个正交投影变换。
- **投影摄像机**　把一个透视视图建模成投影变换，然后接一个透视投影变换。

我们对每一种情况进行更细致的考虑。记住，惟一的现实问题是如何从一个物体的结构特征组和图像的结构特征组之间的对应中获得一个摄像机模型；其他步骤在算法 18.1 中已经提供。

仿射摄像机　可以用齐次坐标把仿射摄像机写成$\mathcal{A}$，$\Pi\mathcal{A}$是一般的仿射变换，而 Π 是正交摄像机变换。这意味着：

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

和

$$\mathcal{A} = \begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

我们用大写字母表示模型上的点，小写字母表示图像上的点，下标表示对应（因此 $\boldsymbol{p}_1 = \Pi\mathcal{A}\ \boldsymbol{P}_1$）。

一个可能的结构特征组结构特征组包含 4 个点。在这个例子里，需要从 4 个图像点与 4 个物体点对应中确定摄像机$\mathcal{A}$。现在假设已经有了 4 个图像点（$\boldsymbol{p}_i$）与 4 个物体点之间的对应（$\boldsymbol{P}_i$）。我们可以把公式 $\boldsymbol{p}_i = \Pi\mathcal{A}\boldsymbol{P}_i$ 中$\mathcal{A}$的前两行写成两个线性方程：

$$\begin{pmatrix} p_{i0} \\ p_{i1} \end{pmatrix} = \begin{pmatrix} a_{00}P_{i0} + a_{01}P_{i1} + a_{02}P_{i2} + a_{03} \\ a_{10}P_{i0} + a_{11}P_{i1} + a_{12}P_{i2} + a_{13} \end{pmatrix}$$

$\mathcal{A}$的前两行有 8 个元素；有在一般位置的 4 个点，可以解这个方程并获得对于$\mathcal{A}$前两行的惟一解。注意$\mathcal{A}$的剩余行并不参与投影，它们对于其他点的计算不必知道，因此不需要知道它的值。这意味着只需要知道$\mathcal{A}$的前两行，就可以计算反向投影了。

一些在旋转和平移下不一样的模型，在仿射摄像机下的解是含混的。假设一个模型的点集为 $\boldsymbol{P}_j$，另一个模型的点集为 $\boldsymbol{Q}_j$，且存在一个变换$\mathcal{B}$，使得对于每一个 j 都有 $\boldsymbol{P}_j = \mathcal{B}\ \boldsymbol{Q}_j$。那么这些模型在仿射摄像机下就不能区分。前一模型在仿射摄像机中的观测点集为 $\boldsymbol{p}_j = \Pi\mathcal{A}_1\boldsymbol{P}_j$，后一模型在某个不同的仿射摄像机中的观测点集为 $\boldsymbol{q}_j = \Pi\mathcal{A}_2\boldsymbol{Q}_j$。如果$\mathcal{A}_2 = \mathcal{A}_1\ \mathcal{B}$，那么

$$\boldsymbol{q}_j = \Pi\mathcal{A}_2\boldsymbol{Q}_j = \Pi\mathcal{A}_1\mathcal{B}\boldsymbol{Q}_j = \Pi\mathcal{A}_1\boldsymbol{P}_j = \boldsymbol{p}_j$$

也就是说，存在某个仿射摄像机，使得后一个模型看上去和在另一个仿射摄像机下的前一个模

型一样，所以它们不能区分。

投影摄像机 用齐次坐标可以把一个投影摄像机写成$\Pi\mathcal{A}$，$\mathcal{A}$是一般的投影变换，而Π是透视摄像机变换。则有：

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

和

$$\mathcal{A} = \begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{pmatrix}$$

同样，我们用大写字母表示模型上的点，小写字母表示图像上的点，下标表示对应（因此 $\boldsymbol{p}_1 = \Pi\mathcal{A}\ \boldsymbol{P}_1$。注意因为用齐次坐标，所以，如果 $\lambda \neq 0$，则$\mathcal{A}$与 $\lambda\ \mathcal{A}$表示同样的变换。

一个可能的结构特征组包含 5 个点。在此例中，可以由 5 个图像上的点（$\boldsymbol{p}_i$）与 5 个物体上的点（$\boldsymbol{p}_i$）的对应确定摄像机。可以把公式 $\boldsymbol{p}_i = \Pi\mathcal{A}\ \boldsymbol{P}_i$ 中$\mathcal{A}$的前两行写成两个线性方程：

$$\begin{pmatrix} p_{i0} \\ p_{i1} \end{pmatrix} = \frac{1}{a_{30}P_{i0} + a_{31}P_{i1} + a_{32}P_{i2} + a_{33}} \begin{pmatrix} a_{00}P_{i0} + a_{01}P_{i1} + a_{02}P_{i2} + a_{03} \\ a_{10}P_{i0} + a_{11}P_{i1} + a_{12}P_{i2} + a_{13} \end{pmatrix}$$

$\mathcal{A}$的前三行有 12 个元素；有 5 个在一般位置上的点，可以通过解该方程，获得对于$\mathcal{A}$前三行的惟一解。注意$\mathcal{A}$的剩余行并不参与投影，因此不需要知道它的值。这意味着只需要知道$\mathcal{A}$的前三行，就可以计算反向投影了。注意，这里所做的只是重复 18.2.2 节的工作。

一些在仿射摄像机（同样也是旋转与平移）下不一样的模型，在投影摄像机下是含混的。假设一个模型给定点集为 $\boldsymbol{P}_j$，另一个模型给定点集为 $\boldsymbol{Q}_j$，并且存在投影变换$\mathcal{B}$，使得对于每一个 j 都有 $\boldsymbol{P}_j = \mathcal{B}\ \boldsymbol{Q}_j$。这些模型在投影摄像机下就不能被区分开。前一个模型在投影摄像机下的观测是点集 $\boldsymbol{p}_j = \Pi\mathcal{A}_1\ \boldsymbol{P}_j$，后一个模型在某个不同投影摄像机下的观测是点集 $\boldsymbol{q}_j = \Pi\mathcal{A}_2\boldsymbol{Q}_j$。如果$\mathcal{A}_2 = \mathcal{A}_1\ \mathcal{B}$，那么

$$\boldsymbol{q}_j = \Pi\mathcal{A}_2\boldsymbol{Q}_j = \Pi\mathcal{A}_1\mathcal{B}\boldsymbol{Q}_j = \Pi\mathcal{A}_1\boldsymbol{P}_j = \boldsymbol{p}_j$$

也就是说，存在某个投影摄像机，使得后一个模型看上去和在另一个投影摄像机下的前一个模型一样，所以它们不能区分。

18.2.3 模型的线性组合

上面讨论的仿射摄像机用对应关系来进行显式摄像机标定。如具备一点线性代数的知识，就可以跳过摄像机标定过程。用齐次坐标可以写出一个未标定的仿射摄像机$\Pi\mathcal{A}$，其中$\mathcal{A}$是一般仿射变换，而

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

我们用大写字母表示模型上的点，用小写字母表示图像上的点，下标表示对应（因此 $\boldsymbol{p}_1 = \Pi \mathcal{A} \boldsymbol{P}_1$）。

找一个模型上的点作为原点，然后考虑从其他点到该点的偏置，从而可以忽略物体的位移。使用标记 $\boldsymbol{v}_i = \boldsymbol{p}_i - \boldsymbol{p}_0$ 和 $\boldsymbol{V}_i = \boldsymbol{P}_i - \boldsymbol{P}_0$，获取不同仿射摄像机下物体的三个视图，三个仿射变换分别为$\mathcal{A}$，$\mathcal{B}$，$\mathcal{C}$，对于物体上的第 i 个点有：

$$\boldsymbol{v}_i^A = \Pi \mathcal{A} \boldsymbol{V}_i$$
$$\boldsymbol{v}_i^B = \Pi \mathcal{B} \boldsymbol{V}_i$$
$$\boldsymbol{v}_i^C = \Pi \mathcal{C} \boldsymbol{V}_i$$

因为 Π 包含许多的 0，并且 $\boldsymbol{V}_i$ 的第 4 行是 0，可以用三个视图来大大简化计算。

把$\mathcal{A}$的第 j 行写成 $\boldsymbol{a}_j^T$，有

$$\boldsymbol{v}_i^A = (\boldsymbol{a}_0^T.\boldsymbol{V}_i, \boldsymbol{a}_1^T.\boldsymbol{V}_i, 0)^T$$
$$\boldsymbol{v}_i^B = (\boldsymbol{b}_0^T.\boldsymbol{V}_i, \boldsymbol{b}_1^T.\boldsymbol{V}_i, 0)^T$$
$$\boldsymbol{v}_i^C = (\boldsymbol{c}_0^T.\boldsymbol{V}_i, \boldsymbol{c}_1^T.\boldsymbol{V}_i, 0)^T$$

如果想要产生物体的一些任意的新视图，可以把 $\Pi \mathcal{D}$应用到点上获得，其中$\mathcal{D}$是某个新的仿射变换。为了获得视图，首先需要确定 $\boldsymbol{p}_0$ 的位置，然后我们对第 i 个点需要 $\boldsymbol{v}_i^D$。

现在 $\boldsymbol{v}_i^D = (\boldsymbol{d}_0^T.\boldsymbol{V}_i, \boldsymbol{d}_1^T.\boldsymbol{V}_i, 0)$。假设$\mathcal{A}$，$\mathcal{B}$，$\mathcal{C}$是一般性的，那么 $\boldsymbol{d}_j$ 一定是一个 $\boldsymbol{a}_j$，$\boldsymbol{b}_j$，$\boldsymbol{c}_j$ 的固定线性组合，

$$\boldsymbol{d}_j = \lambda_{(\boldsymbol{a}_j)}\boldsymbol{a}_j + \lambda_{(\boldsymbol{b}_j)}\boldsymbol{b}_j + \lambda_{(\boldsymbol{c}_j)}\boldsymbol{c}_j$$

则有：

$$\boldsymbol{v}_i^D = (\lambda_{(\boldsymbol{a}_0)}\boldsymbol{a}_0^T.\boldsymbol{V}_i + \lambda_{(\boldsymbol{b}_0)}\boldsymbol{b}_0^T.\boldsymbol{V}_i + \lambda_{(\boldsymbol{c}_0)}\boldsymbol{c}_0^T.\boldsymbol{V}_i, \lambda_{(\boldsymbol{a}_1)}\boldsymbol{a}_1^T.\boldsymbol{V}_i + \lambda_{(\boldsymbol{b}_1)}\boldsymbol{b}_1^T.\boldsymbol{V}_i + \lambda_{(\boldsymbol{c}_1)}\boldsymbol{c}_1^T.\boldsymbol{V}_i, 0)$$

这说明，给定物体三个未知的仿射视图，可以通过确定这些 λ 的值重构第 4 个仿射视图。

这个策略叫做模型的线性组合。用这种办法产生假设同样需要搜索对应关系；选择某些图像上的点 $\boldsymbol{p}_0$，$\boldsymbol{p}_1$ 等，然后解出 λ 值。一旦这些都已知，就可以绘制该物体，尽管还需要考虑隐藏线去除的问题。注意这个方法仅仅是仿射摄像机标定的另一个版本，它的优点在于物体模型用一种简单方式从物体的三个视图中构造。实际上 18.4 节中讨论的方法也能从物体的三个视图中构造物体模型。

18.3　位姿聚类获得假设

大多数物体有许多结构特征组，这意味着物体与图像结构特征组之间有许多对应关系假设能顺利地通过校验，其中每一个对应都对物体相对于摄像机（或者摄像机相对于物体）的位置和方向进行大致相同的估计。然而，来自于噪声[或者杂物（clutter），这个术语用来描述不想要的且不在模型库里的东西]的图像结构特征组中，很容易产生不相干的位姿估计。这种现象

导致在校验之前用聚类方法过滤假设的做法。

对于每个物体,建立一个表示位姿空间的累加器数组——累加器数组中的每个元素对应位姿空间中的一个桶(bucket)。对于每一个图像的结构特征组,与每一个物体的每一个结构特征组产生对应假设,并确定相应的位姿参数,然后在累加器数组表示相应位置的元素增加一票。如果在物体的累加器数组中的任何一个元素有很多票数,那么这就是该物体在这个位姿上出现的证据;这个证据可以再用校验方法验证。这个方法(在算法 18.2 里给出)与哈夫变换(15.1 节)非常类似。

算法 18.2 位姿聚类:对位姿、对应和身份投票

```
对于所有物体 O
    对于所有物体结构特征组 F(O)
        对于所有图像结构特征组 F(I)
            对于所有 F(I)的元素和 F(O)的元素之间的对应 C
                用 F(I),F(O),C 来推导物体的位姿 P(O)
                在 O 的位姿空间相应的桶[对应于 P(O)]中增加一票
            end
        end
    end
end
对于所有物体 O
    对于所有 O 位姿空间有足够投票的元素 P(O)
        使用 P(O)和摄像机模型的估计来绘制物体
        如果绘制与图像一致,那么物体存在
    end
end
```

这个方法存在两个问题(在哈夫变换里也有类似的实际问题):

1. 在包含噪声和纹理的图像中,可能产生了许多错误的结构特征组,对应物体真实位姿的累加器数组元素中的投票也许还没有错误的投票多(细节在 Grimson,Huttenlocher 的 1990*b* 文献中提到)。
2. 很难选择位姿阵列桶的尺寸;非常小的桶使得投票无法累计(因为精确计算位姿很难),太大的桶导致太多的桶有足够的累计值以至于启动校验过程。

采用对明显不可靠的投票不进行统计的方式,可以增强此方法的抗噪声能力。例如,假设物体在那个位姿上出现,而相应的特征结构组却不可见,这显然是不可靠的情况。这些改进足以产生可工作的系统(见图 18.2)。

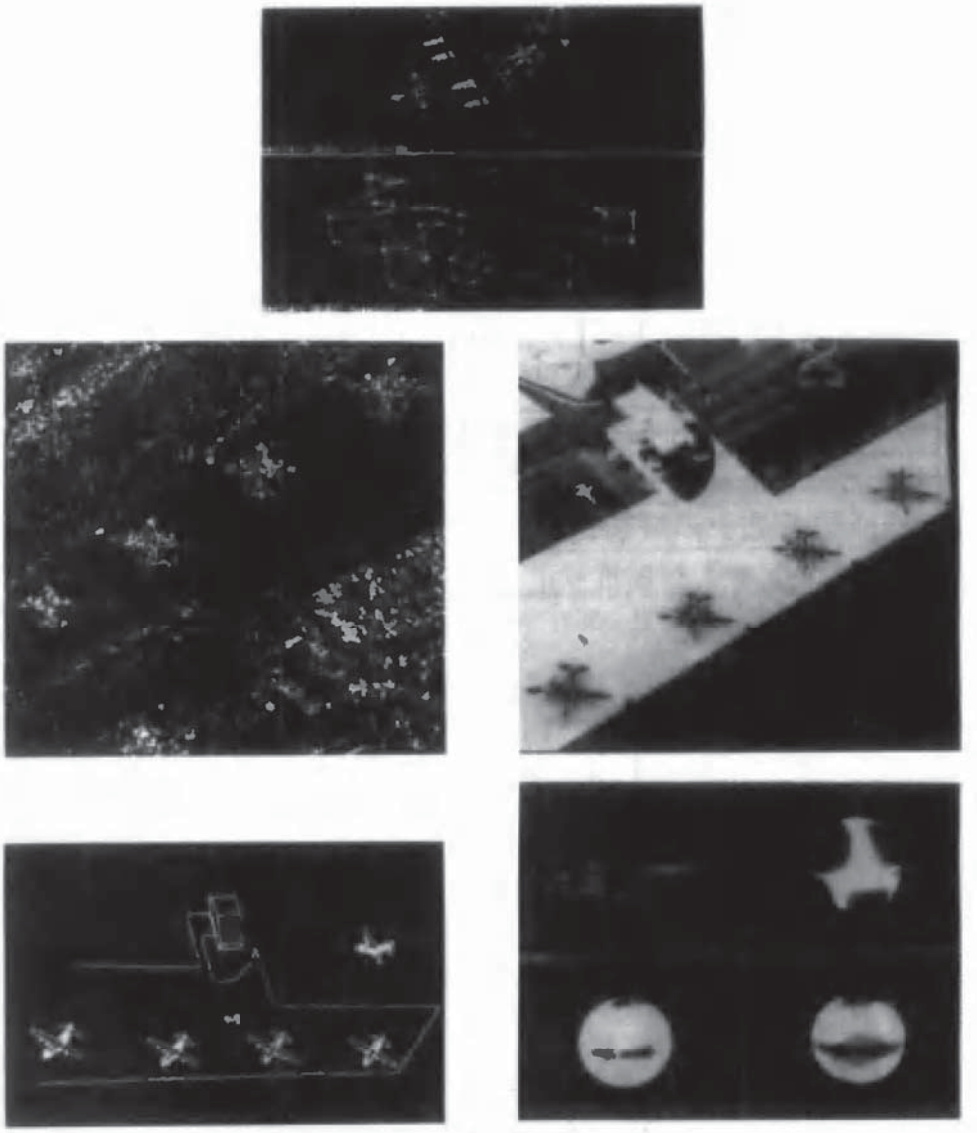


图 18.2　位姿聚类方法使用结构特征组产生位姿估计，然后对这些估计进行聚类。上图：两个在早期位姿聚类系统中的模型。中左图：用做测试的图像中标记的边缘点。中右图：找到的边缘覆盖在图上。下左：用模型在空间布局的一个新视图表示它们的位姿；注意在跑道外的那架飞机的古怪姿态。下右：对于每个结构特征组，在某些视图估计位姿更加稳定因而比其他视图好；在飞机模型旁是一个表示不同的视角的球，亮的区域表示模型上标记的特征对应错误率高

18.4　采用不变量获得假设

位姿聚类方法将导致类似摄像机位姿假设的特征对应收集在一起，另一个获得位姿假设的方法是采用与摄像机属性无关的观测。这种方法对于平面物体来说非常简单，但也可以应用于其他情形，如 Forsyth，Mundy，Zisserman 和 Rothwell(1992)；Forsyth(1996)；Huang(1981)。

18.4.1　平面图形的不变量

前面提到，一个仿射摄像机可以写成$\Pi\mathcal{A}$，其中$\mathcal{A}$是一般仿射变换，Π是正交摄像机变换。

假设有一组在同一平面上的模型点 $\boldsymbol{P}_j$；不失一般性，假设它们在 $z=0$ 平面。则有：

$$\begin{pmatrix} p_{i0} \\ p_{i1} \\ 1 \end{pmatrix} = \Pi\mathcal{A}\begin{pmatrix} P_{i0} \\ P_{i1} \\ 0 \\ 1 \end{pmatrix}$$

使用 18.2.2 节的记号，替换掉Π和$\mathcal{A}$得到：

$$\begin{pmatrix} p_{i0} \\ p_{i1} \\ 1 \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & a_{03} \\ a_{10} & a_{11} & a_{13} \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} P_{i0} \\ P_{i1} \\ 1 \end{pmatrix}$$

这一点非常重要；它表示，同一平面的点在仿射摄像机中的不同视图，可以用平面仿射变换获得。这意味着我们可以将摄像机抽象化，仅考虑这些变换在模型上的效果。

类似的结果可以应用在投影摄像机中共面点的视图上，此时的变换是平面投影变换。为了得到这个结果，我们回忆投影摄像机可以写成$\Pi\mathcal{A}$，其中$\mathcal{A}$是一般的投影变换，Π是透视摄像机变换。假设我们有一组模型点 $\boldsymbol{P}_j$，都在同一平面；不失一般性，假设它们都在 $z=0$ 平面则有：

$$\begin{pmatrix} p_{i0} \\ p_{i1} \\ 1 \end{pmatrix} = \Pi\mathcal{A}\begin{pmatrix} P_{i0} \\ P_{i1} \\ 0 \\ 1 \end{pmatrix}$$

使用 18.2.2 节的记号。替换掉Π和$\mathcal{A}$得到：

$$\begin{pmatrix} p_{i0} \\ p_{i1} \end{pmatrix} = \frac{1}{a_{20}P_{i0} + a_{21}P_{i1} + a_{23}}\begin{pmatrix} a_{00} & a_{01} & a_{03} \\ a_{10} & a_{11} & a_{13} \end{pmatrix}\begin{pmatrix} P_{i0} \\ P_{i1} \\ 1 \end{pmatrix}$$

由于采用齐次坐标，更方便的形式是：

$$\begin{pmatrix} p_{i0} \\ p_{i1} \\ p_{i2} \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & a_{03} \\ a_{10} & a_{11} & a_{13} \\ a_{20} & a_{21} & a_{23} \end{pmatrix}\begin{pmatrix} P_{i0} \\ P_{i1} \\ P_{i3} \end{pmatrix}$$

同样，这意味着共面点集在投影摄像机下的不同视图可由平面投影变换产生。这意味着我们可以将摄像机抽象化，仅考虑这些变换在模型上的效果。

共面点的仿射不变量 假设有一个共面点集组成的模型。选择其中三个点 $\boldsymbol{P}_0, \boldsymbol{P}_1, \boldsymbol{P}_2$，这可以确定一个坐标系，而模型中的任何点 $\boldsymbol{P}_i$ 可以表示为 $\boldsymbol{P}_0 + \mu_{i1}(\boldsymbol{P}_1 - \boldsymbol{P}_0) + \mu_{i2}(\boldsymbol{P}_2 - \boldsymbol{P}_0)$。仅需少量线性代数知识就可以计算与每一点相关的$\mu$值。

现在摄像机把模型点 $\boldsymbol{P}_i$ 变成图像点$\boldsymbol{p}_i$。对于观察一个共面点集的未标定仿射摄像机来说，摄像机的效果可以写成(未知的)平面仿射变换。把摄像机写成$\mathcal{C}$。有

$$\begin{aligned}
\boldsymbol{p}_i &= \mathcal{C}\boldsymbol{P}_i \\
&= \mathcal{C}(\boldsymbol{P}_0 + \mu_{i1}(\boldsymbol{P}_1 - \boldsymbol{P}_0) + \mu_{i2}(\boldsymbol{P}_2 - \boldsymbol{P}_0)) \\
&= (1 - \mu_{i1} - \mu_{i2})(\mathcal{C}\boldsymbol{P}_0) + \mu_{i1}(\mathcal{C}\boldsymbol{P}_1) + \mu_{i2}(\mathcal{C}\boldsymbol{P}_2) \\
&= (1 - \mu_{i1} - \mu_{i2})\boldsymbol{p}_0 + \mu_{i1}\boldsymbol{p}_1 + \mu_{i2}\boldsymbol{p}_2 \\
&= \boldsymbol{p}_0 + \mu_{i1}(\boldsymbol{p}_1 - \boldsymbol{p}_0) + \mu_{i2}(\boldsymbol{p}_2 - \boldsymbol{p}_1)
\end{aligned}$$

这意味着 μ_{ij} 以独立于视图的方式描述了物体的几何(即,如果在模型平面或者某些仿射视图中计算 μ_{ij},可以得到相同结果)。具有该性质的测量一般被称为仿射不变量(affine invariants)(在练习里有其他构造仿射不变量的题目)。

共面点与直线的投影不变量　在齐次坐标下,可以把图像点与(平面)模型点的关系写成 $\boldsymbol{p}_i = \mathcal{A}\,\boldsymbol{P}_i$,其中,$\mathcal{A}$ 是 3×3 的矩阵。我们有:

$$\frac{\det\left[\boldsymbol{p}_i\boldsymbol{p}_l\boldsymbol{m}_k\right]}{\det\left[\boldsymbol{p}_i\boldsymbol{p}_j\boldsymbol{p}_l\right]}\frac{\det\left[\boldsymbol{p}_i\boldsymbol{p}_l\boldsymbol{p}_m\right]}{\det\left[\boldsymbol{p}_i\boldsymbol{p}_k\boldsymbol{p}_m\right]}$$

它等于

$$\frac{\det\left[(\mathcal{A}\boldsymbol{P}_i)(\mathcal{A}\boldsymbol{P}_j)(\mathcal{A}\boldsymbol{P}_k)\right]}{\det\left[(\mathcal{A}\boldsymbol{P}_i)(\mathcal{A}\boldsymbol{P}_j)(\mathcal{A}\boldsymbol{P}_l)\right]}\frac{\det\left[(\mathcal{A}\boldsymbol{P}_i)(\mathcal{A}\boldsymbol{P}_l)(\mathcal{A}\boldsymbol{P}_m)\right]}{\det\left[(\mathcal{A}\boldsymbol{P}_i)(\mathcal{A}\boldsymbol{P}_k)(\mathcal{A}\boldsymbol{P}_m)\right]}$$

也等于

$$\frac{\det\mathcal{A}\det\left[\boldsymbol{P}_i\boldsymbol{P}_j\boldsymbol{P}_k\right]}{\det\mathcal{A}\det\left[\boldsymbol{P}_i\boldsymbol{P}_j\boldsymbol{P}_l\right]}\frac{\det\mathcal{A}\det\left[\boldsymbol{P}_i\boldsymbol{P}_l\boldsymbol{P}_m\right]}{\det\mathcal{A}\det\left[\boldsymbol{P}_i\boldsymbol{P}_k\boldsymbol{P}_m\right]}$$

它又等于

$$\frac{\det\left[\boldsymbol{P}_i\boldsymbol{P}_j\boldsymbol{P}_k\right]}{\det\left[\boldsymbol{P}_i\boldsymbol{P}_j\boldsymbol{P}_l\right]}\frac{\det\left[\boldsymbol{P}_i\boldsymbol{P}_l\boldsymbol{P}_m\right]}{\det\left[\boldsymbol{P}_i\boldsymbol{P}_k\boldsymbol{P}_m\right]}$$

只要 i,j,k,l 和 m 两两互不相等,则任意三个点不共线。也有其他行列式排列同样也是不变量(见练习)。

平面代数曲线和投影变换　代数曲线由平面上使得一个多项式为零的所有的点组成。直线就是一条代数曲线。如果在齐次坐标系下把点写为 $\boldsymbol{p}_i = [p_{i0}, p_{i1}, p_{i2}]^T$,直线是使得 $l_0p_0 + l_1p_1 + l_2p_2 = 0$ 的点 $\boldsymbol{p}$ 的轨迹。我们可以写成 $\boldsymbol{l}^T\boldsymbol{p} = 0$,其中用 $\boldsymbol{l}$ 代表直线。如果点的变换为 $\boldsymbol{p} = \mathcal{A}\,\boldsymbol{P}$,那么直线的变换就是 $\boldsymbol{l} = \mathcal{A}^{-T}\boldsymbol{L}$。从下面式子很容易看出:

$$\boldsymbol{l}^T\boldsymbol{p} = \boldsymbol{l}^T\mathcal{A}\boldsymbol{P} = \boldsymbol{L}^T\mathcal{A}^{-1}\mathcal{A}\boldsymbol{P}$$

所以如果 $\boldsymbol{p}$ 在直线 $\boldsymbol{l}$ 上,那么 $\boldsymbol{P}$ 也在直线 $\boldsymbol{L}$ 上。

因为直线变换也类似点变换,则有

$$\begin{aligned}\frac{\det\left[\boldsymbol{l}_i\boldsymbol{l}_j\boldsymbol{l}_k\right]}{\det\left[\boldsymbol{l}_i\boldsymbol{l}_j\boldsymbol{l}_l\right]}\frac{\det\left[\boldsymbol{l}_i\boldsymbol{l}_l\boldsymbol{l}_m\right]}{\det\left[\boldsymbol{l}_i\boldsymbol{l}_k\boldsymbol{l}_m\right]} &= \frac{\det\left[(\mathcal{A}^{-T}\boldsymbol{L}_i)(\mathcal{A}^{-T}\boldsymbol{L}_j)(\mathcal{A}^{-T}\boldsymbol{L}_k)\right]}{\det\left[(\mathcal{A}^{-T}\boldsymbol{L}_i)(\mathcal{A}^{-T}\boldsymbol{L}_j)(\mathcal{A}^{-T}\boldsymbol{L}_l)\right]}\frac{\det\left[(\mathcal{A}^{-T}\boldsymbol{L}_i)(\mathcal{A}^{-T}\boldsymbol{L}_l)(\mathcal{A}^{-T}\boldsymbol{L}_m)\right]}{\det\left[(\mathcal{A}^{-T}\boldsymbol{L}_i)(\mathcal{A}^{-T}\boldsymbol{L}_k)(\mathcal{A}^{-T}\boldsymbol{L}_m)\right]}\\ &= \frac{\det\mathcal{A}\det\left[\boldsymbol{L}_i\boldsymbol{L}_j\boldsymbol{L}_k\right]}{\det\mathcal{A}\det\left[\boldsymbol{L}_i\boldsymbol{L}_j\boldsymbol{L}_l\right]}\frac{\det\mathcal{A}\det\left[\boldsymbol{L}_i\boldsymbol{L}_l\boldsymbol{L}_m\right]}{\det\mathcal{A}\det\left[\boldsymbol{L}_i\boldsymbol{L}_k\boldsymbol{L}_m\right]}\\ &= \frac{\det\left[\boldsymbol{L}_i\boldsymbol{L}_j\boldsymbol{L}_k\right]}{\det\left[\boldsymbol{L}_i\boldsymbol{L}_j\boldsymbol{L}_l\right]}\frac{\det\left[\boldsymbol{L}_i\boldsymbol{L}_l\boldsymbol{L}_m\right]}{\det\left[\boldsymbol{L}_i\boldsymbol{L}_k\boldsymbol{L}_m\right]}\end{aligned}$$

只要 i,j,k,l,m 两两互不相等,任意三条线不过同一点。

事实上,代数不变量在投影中是常见的,平面二次曲线的例子特别有用。一个平面二次曲线是使得 $\boldsymbol{x}^t\,\mathcal{M}\,\boldsymbol{x} = 0$ 的点 $\boldsymbol{x}$ 的轨迹,$\boldsymbol{x}$ 是齐次坐标描述点的一个向量,矩阵 $\mathcal{M}$ 包含二次曲线

的系数。如果把坐标做某种变换(在一个摄像机里观察这些点),那么对于某些平面投影变换 $\boldsymbol{P}$,我们有 $\boldsymbol{x}' = \mathcal{P}\,\boldsymbol{x}$。在新坐标系下的二次曲线的方程可以通过下面得到:新方程必须在原来旧方程等于零的每一点上等于零(换句话说,对于每一点在旧坐标系下旧二次曲线的方程等于零)。特别地,如果逆变换,然后把结果点放入旧方程,会得到零。这个推理得到$\mathcal{M}' = \mathcal{P}^{-t}\mathcal{M}\mathcal{P}^{-1}$就是新坐标系下的二次曲线的方程。

现在假设有两个二次曲线,$\mathcal{M}$和$\mathcal{N}$。每个都是以这种方式进行变换的,通过模型来观察$\mathcal{A}_{MN} = \mathcal{M}^{-1}\mathcal{N}$,通过图像观测$\mathcal{A}'_{MN} = \mathcal{P}\mathcal{M}^{-1}\mathcal{N}\mathcal{P}^{-1}$。这意味着$\mathcal{A}'_{MN}$的特征值与$\mathcal{A}_{MN}$的一样。$\mathcal{A}_{MN}$和$\mathcal{A}'_{MN}$只是相差了一个常数因子。这表示所观测的特征值可能扩大了一个未知的因子,然而,特征值的比例是不变的。一个有用的例子是 $\mathrm{trace}(\mathcal{A}_{MN})^3/\det(\mathcal{A}_{MN})$。

对于点与直线的混合集合,构造不变量也很容易。例如,假设有一组点 $\boldsymbol{p}_i$ 和一组直线 $\boldsymbol{l}_j$。则有

$$\frac{(\boldsymbol{l}_i^T\boldsymbol{p}_k)(\boldsymbol{l}_j^T\boldsymbol{p}_l)}{(\boldsymbol{l}_i^T\boldsymbol{p}_l)(\boldsymbol{l}_j^T\boldsymbol{p}_k)} = \frac{(\boldsymbol{L}_i^T\mathcal{A}^{-1}\mathcal{A}\boldsymbol{P}_k)(\boldsymbol{L}_j^T\mathcal{A}^{-1}\mathcal{A}\boldsymbol{P}_l)}{(\boldsymbol{L}_i^T\mathcal{A}^{-1}\mathcal{A}\boldsymbol{P}_l)(\boldsymbol{L}_j^T\mathcal{A}^{-1}\mathcal{A}\boldsymbol{P}_k)}$$

$$= \frac{(\boldsymbol{L}_i^T\boldsymbol{P}_k)(\boldsymbol{L}_j^T\boldsymbol{P}_l)}{(\boldsymbol{L}_i^T\boldsymbol{P}_l)(\boldsymbol{L}_j^T\boldsymbol{P}_k)}$$

这意味着这个表达式也是不变的(只要 i,j 不等,k,l 不等)。

对于点、直线、二次曲线的混合的投影不变量也是熟知的,并且成功地运用在物体识别中(在练习中会有一些例子)。高阶平面代数曲线的投影不变量也是已知的,但是实践上没有什么重要性,因为实际中很难遇到这样的曲线,并且它们很难准确地拟合。

18.4.2 几何散列

几何散列是一种利用几何不变量来投票决定物体假设的算法,与位姿聚类方法有类似之处(要投票),但是不对位姿投票,而是对几何投票。这个思想原本用在未标定的平面模型的仿射视图,用它来解释最容易。

对于模型上给定的任意三点,可以用 18.4.1 节的技术来计算模型上其他每一点的μ_1 和μ_2 的值。可以按μ_1 和μ_2 的值建立一个索引表。对于在模型库里的每一个模型,和该模型上每一个三点集合,计算每一个其他点的相应μ值。用这些μ值作为索引,记下相应的模型名和得到这些μ值的模型上的三点集。因此,一对μ值的作用相当于关于模型名和该模型上相应三点集的假设。

在获得该表的基础上,可以用搜索对应来查找模型。在找到任意三个点后,计算所有图像中其他点的μ值。用这些μ值为索引检索出索引表的内容。如果三个点对应着物体上的三个点,相应物体和三个点的组合可以获得许多投票。可以设想噪声产生的投票是不相关的,这意味着有很多票散乱地投到不同物体上的各种三点集上,和很多票投到一个物体的三点组合上。这暗示,如果对相同物体和物体上相同三点有很多投票,那么这个物体就很可能是存在的。注意,这三点的集合可以作为一个用来校验的结构特征组。在算法 18.3 里描述了按μ值投票的方法。

算法 18.3　几何散列:对模型名和点标记投票

对于图像上所有三点组合 $T(I)$

　对于所有其他图像点 p

　　从 p 和 $T(I)$ 计算 μ 值

　　通过这些值获得表项,如果存在,对 $T(I)$ 里的三点标记上物体名和这些特殊点的名字

　　对这些标记聚类;如果有足够的标记,那么反向投影并进行校验

　　end

　end

end

这个算法可以泛化至其他几何组合而不仅是点组合(见编程作业)。如果我们有三维物体的未标定的仿射视图,那么对于每个点有三个 μ 值,每个点不能惟一确定,但是这个方法可以扩展(也见编程作业)。与位姿聚类和哈夫变换一样(实际上本方法也是哈夫变换),很难选择桶的尺寸,并且难以知道多少才是足够的,也有导致表被阻塞(clogged)的危险(Grimson 和 Huttenlocher, 1990*a*)。

18.4.3　不变量与索引

几何散列搜索对应是通过从散列表检索出可接受的标签而实现的。几何散列的主要特点是在识别时不需要搜索模型,因为散列表可以预先读入。这是非常好的特性,通常称之为索引(indexing)。在对准算法中,当不同的模型有不同类别的结构特征组时,就用到了索引。很明显,某种类型的图像结构特征组只需要对这种结构特征组适用的模型检验。

几何散列中的技巧是查看那些包含独立于物体位姿和随物体改变的信息的图像集合(μ 值),这些 μ 值可以产生物体信息。几何散列考虑了所有可能的点集,这个技巧可以扩展到所有类别的其他几何特征上。把那些包含独立于物体位姿和随物体改变的信息的特征集合叫做不变量基集(invariant bearing groups),在 18.4.1 节中可以看到一些例子。假设能得到的不同类别的不变量基集已知,对准算法可以修改成算法 18.4。

算法 18.4　采用不变量基集的不变量索引

对于不变量基集的每一个类型 T

　对于类型 T 的每一个图像集合 G

　　确定 G 的不变量的值 V

　　对于每一个类型 T 的模型特征集合 M(M 的不变量也有值 V)

　　确定使 M 变为 G 的变换

　　用该变换绘制该模型　　(未完待续)

(续)

```
            把结果与图像比较,如果相似则可接受
        end
    end
end
```

如果不变量关系集合有区别,就是说没有不变量值相同的模型特征集合,那么这个方法非常有效。不变量的值也需要准确地测量。注意必须查看与图像特征集合具有相同值的每一个模型特征集合,因为不知道哪一个组是。同样这也是一个对于对应关系的搜索,并且期望要搜索的对应会非常少。

在未标定的透视视图中用直线与二次曲线索引 在以前的算法中,可以用到很多种不变量基集。对于未知透视摄像机下这个方法对平面图形是最好的,图像变换的不变量非常多。一般来说,最有用的例子是下面三种集合的不变量:5 条直线,两个二次曲线、一个二次曲线和一条直线。

给定这些函数,一个典型的系统(在图 18.3 中说明)会有:

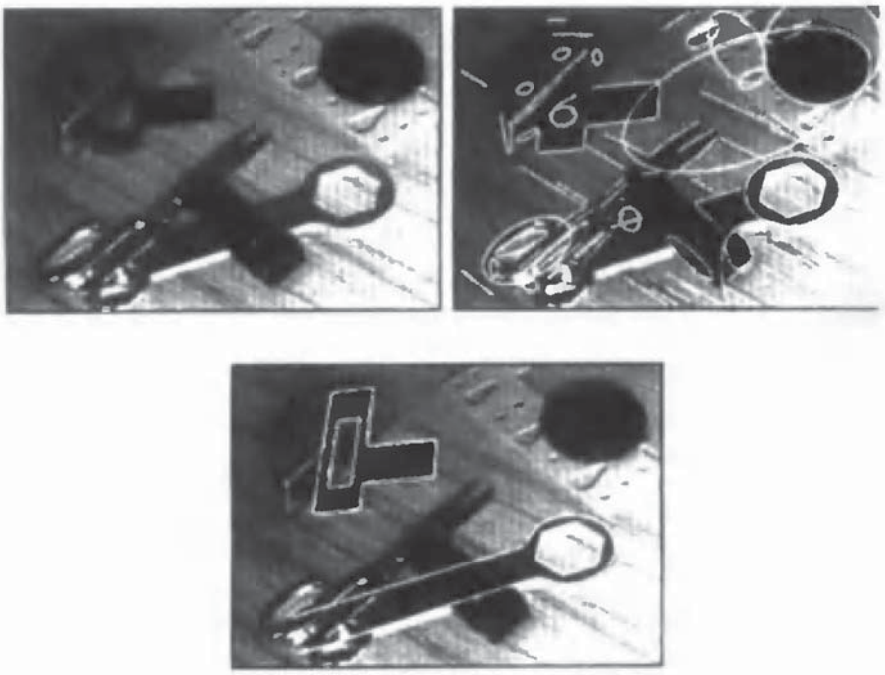


图 18.3 这些图说明了一个用不变量识别平面物体的系统的大致结构。在左上是一幅图像;右上显示了拟合成直线和二次曲线的边缘点。下图、识别出并校验过的物体轮廓结果迭加在图上

- **抽取基元集合** 边缘检测器在图像中检测边缘,然后用直线与二次曲线拟合边缘点。舍弃边缘点,剩下拟合的曲线。从 5 条直线段的所有集合来形成不变量是很繁的。这也没有必要,因为物体并不是由分散的直线组成的,因此仅仅需要线段的开曲线。它们是这样的集合:在每条线段的端点附近,一定有另一条线段的端点。
- **用不变量索引** 相关的直线与二次曲线的集合用来获得物体的假设,一般通过索引不变量的量化值的数组得到。对于每个集合类型往往只有一个或两个不变量和比较少的模型,因此使用一个阵列不是特别浪费。量化桶的尺寸一般靠尝试和修正确定,另一个

好方法是搜索所选择的桶的邻域内搜索。同样，每个集合只有一个或两个不变量也意味着这并不是非常浪费。

- **反向投影和校验**　对于每个物体假设，模型集合到图像集合的变换是确定的。用该变换对模型进行反向投影，然后进行校验。

平面曲线的不变量索引　一个几何不变量的来源是用协变构造(covariant constructions)，构造和变换是可交换的，也就是说，如果先构造，然后变换构造结果，和先变换再构造的结果是一样的。在这个方法里，构造产生一个坐标系，然后通过变换将其转换成方便的世界坐标系，再在这个坐标系[叫做规范坐标系(canonical frame)]里进行观测。因为观测是在固定坐标系里进行，所以在规范坐标系里的测量属性是不变量。应该注意到这个方法在概念上和几何散列的相似性，在几何散列方法中的μ值都是在规范坐标系里测量的。

例如，取一个曲线，然后构造一个在曲线两个不同点相切的直线(封闭不凸的平面曲线有这样的切线，但是凸封闭曲线没有，开曲线就不能保证)。现在用一个变换使得一个切点在原点而另一个切点在 x 轴上。在这个坐标系统里，你想要做的任何观测都是不变量，除非细心否则你可能不知道原点是哪个点(见图 18.4)。这种情形带来一些麻烦：假设可以做许多不同的不变量观测，那么应该做哪一个观测呢？

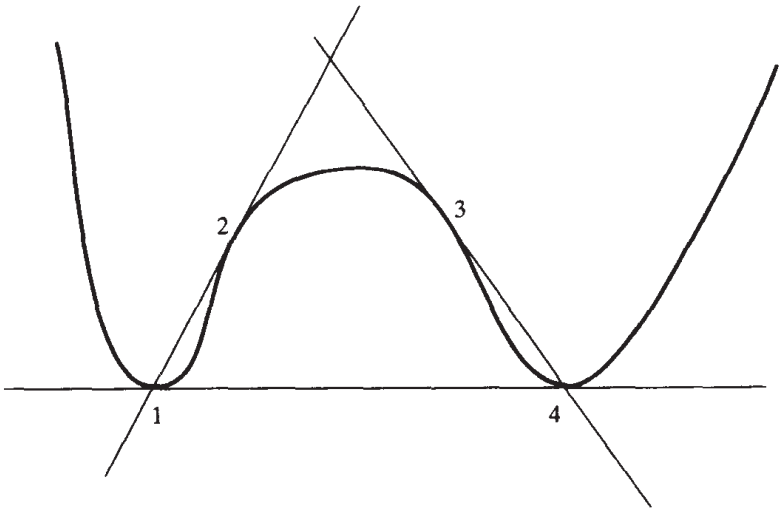


图 18.4　因为相切和连接是协变的(即，在一个坐标系里的相切，在另一个坐标系里还是相切)，在相切与连接的基础上的构造产生规范坐标系。这个图说明了叫做 M 曲线构造的一种构造。对于一个像 M 形状的曲线(倒过来的 M)，一个双切线产生两个点(1和4)，从这两点可以产生经过曲线上另两个点2和3的切线；这样产生了4个点。因为没有三点共线的4点可以映射到另外4点，我们可以把这4点作为平面上的单位方格，然后在规范坐标系下观测这个曲线。在这个系统下的任何观测都是不变量

18.5　校验

准确的校验需要好的测试方法以便判断所绘制的物体模型是否和图像相似。选择测试方法依赖于对产生绘制的环境信息掌握的程度。例如，如果环境照明和摄像机的响应都清楚地知道(对于在传送带上检测零件的系统来说，这是可能的)，期望绘制能够准确地预测图像的像

素值是合乎情理的,因此比较像素值是一个不错的测试方法。

通常,我们只知道照明的亮度能允许产生某种假设。这意味着比较对亮度变化应该是鲁棒的。惟一在实践中用到的测试是产生物体的剪影,然后用剪影边缘与图像的边缘点比较。这一节还要描述一些其他可能的测试。

18.5.1 边缘近似度

一种自然的测试方法是使用摄像机模型将物体的剪影边缘覆盖到图像上,然后把剪影边缘与实际图像边缘进行比较来给这个假设打分。通常分值用预测的剪影边缘中位于真实图像边缘附近的长度比率计算。对于摄像机系统的旋转与位移这是不变量,这是这种方法的优点,但它会随尺度变换而变化,这并不一定是坏事。通常只允许边缘点的方向与相比较的物体轮廓线的方向相近时才算分。这里的原则就是,对边缘点的描述越详细,对于它是否来自那个物体的信息就知道得越多。

在打分中包括不可见的剪影部分是不利的,所以绘制也应能去除隐藏线。使用剪影边缘是因为在剪影内部的边在不合适的光照下对比度有可能很低。这意味着它们的消失可能是光照问题,而不是物体是否真的不存在。

边缘近似度测试很可能非常不可靠,即使使用方向信息也不能真正克服这些问题。当我们把模型边缘的集合投影到图像,而在邻近处并没有相应边缘,这可以说明模型不在那里,但是边界附近的边缘的存在并不能说明物体在那里。这是因为边缘的起因有很多,不能保证在计分中用到的边缘是正确的边缘。

一个坏的位姿估计可能导致反向投影物体上的剪影边缘离实际图像边缘非常远。例如,如果一个物体(扳手)的投影图像是一个细长的区域,我们用它的一端来估计图像平面的方向,则在另一端的反向投影边缘可能离相应的图像边缘非常远(见图 18.5)。这是一个误差扩大的例子。本质上,摄像机估计对于那些接近它的特征比较有利,而越远的特征则越坏。下面有一些补救方法:

- **最大化位姿估计** 如果对于位姿估计求校验分数最大化,有时候情况会改进。但它有时候会失效:如果原来的位姿估计很差,那么物体也可能确实靠近边缘,但是是靠近错误的边缘,这可以想像在一个有纹理的背景中检验一个物体是否存在。再加上优化过程有时候很困难,特别是邻近性测试用距离的阈值表示。
- **仅对那些摄像机估计可靠的边缘计数** 据我们所知,这个方法还没在实际中试过。这个方法的优点是它可以更容易地处理像图 18.5 这样的例子;缺点是它在校验过程中没有很好地利用物体的大部分信息,意味着有可能把错误的当成真的。

把错误的边计算到得分中也是一个主要的困难。在一个有纹理的区域里,在小范围会有很多边缘点。校验的主要想法是用模型把在假设形成阶段很难连接的图像证据连在一起,所以不能从得分中把边缘点的小集合的贡献排除。这意味着在纹理很多的区域里,可能对于任何位姿的任何模型都得到高分(见图 18.5)。注意,即使把边缘方向相似信息算进来,也没有什么效果。

可以调整边缘检测器使得纹理更平滑一些,希望纹理区域可以消失。这是一个危险的方法,因为它可能改变对比度,使得物体也消失。它可以应用在许可使用的场合,并且已经被广泛应用。

图 18.5　这幅图说明了边缘方向可能在校验中引入虚假信息。在图像上标出的边缘来自一个扳手的模型，其中有52% 的边缘因与附近图像边缘点的方向匹配而通过识别与校验。遗憾的是，图像边缘来自桌上具有方向的纹理，并不是来自扳手的实例。正如文中所指出的，这个问题是可以避免的，如果把扳手的内部是非纹理的描述加进去，它就不会与桌子的方向性纹理匹配了

18.5.2　纹理、模式、亮度的相似度

如果对边缘匹配打分，那么纹理是一个麻烦事。然而，一些物体，像伪装涂料，有着截然不同的纹理而应该加以利用。可以用纹理描述子描述模型区域(就像第 9 章中区域滤波器输出的统计)，然后与图像进行比较。这种比较需要估计图像和反向投影物体区域之间差别的大小。最有希望的方法是，比较图像区域被物体区域用这种纹理覆盖的概率，和当物体不在时得到这种纹理的概率。

比较剪影边缘忽略了很多有用信息。如果物体是有图案的，意味着有大范围的着色区域，类似可乐罐上的标记，因此也可以比较反向投影的图案的边。一个更复杂的方法是用纹理描述和可能的色彩与饱和度的描述，对反向投影的图案区域和图像区域进行比较(如果不存在纹理同样可以)。

通常没有足够的光照信息去预测物体的亮度。所以在校验的实践中，亮度往往被忽略。这是错误的。亮度模式的不同通常与光源无关，例如，只有奇怪的光源，例如电影放映机，才会产生纹理亮度的图案。这意味着可以通过判断绝对亮度的差异，来自于实际中能出现的差异，还是实际中不可能出现的差异来获得校验分。

18.6　应用：医学图像系统的对准

有许多问题中，位姿估计远比识别更重要。许多识别算法都设计成，针对一些工业部件散落在箱子里或者桌子上的情况；然而产品工程师保证了它们不会混在一起，但是关注的是位姿的准确测量。医学应用也很类似，因为它通常知道所观察的是什么，关键是要准确地测量它们在哪里。

18.6.1 成像方式

有许多可用的成像技术，包括核磁共振成像(MRI)，它使用磁场来测量质子的密度，一般用来描述器官和软组织；计算机析层成像技术(CTI 或 CT)，测量所吸收的 X 光密度，一般用来测量骨骼；原子医疗成像(NMI)，用来测量各种注入的放射性原子密度，一般用于功能性的成像；超声成像(USI)，测量超声传播的速度变化，一般用来获得关于移动器官的信息(图 18.6 阐述了这些方式)。所有技术都用来获得切片数据，然后可以用做三维重建。一个标准问题是把物体分割成不同的结构。图 18.7 显示了脑、脑室和分段肿瘤的 MRI 图像。因为肿瘤相对于头骨和皮肤是固定的，数据给出了肿瘤相对于头的位置。

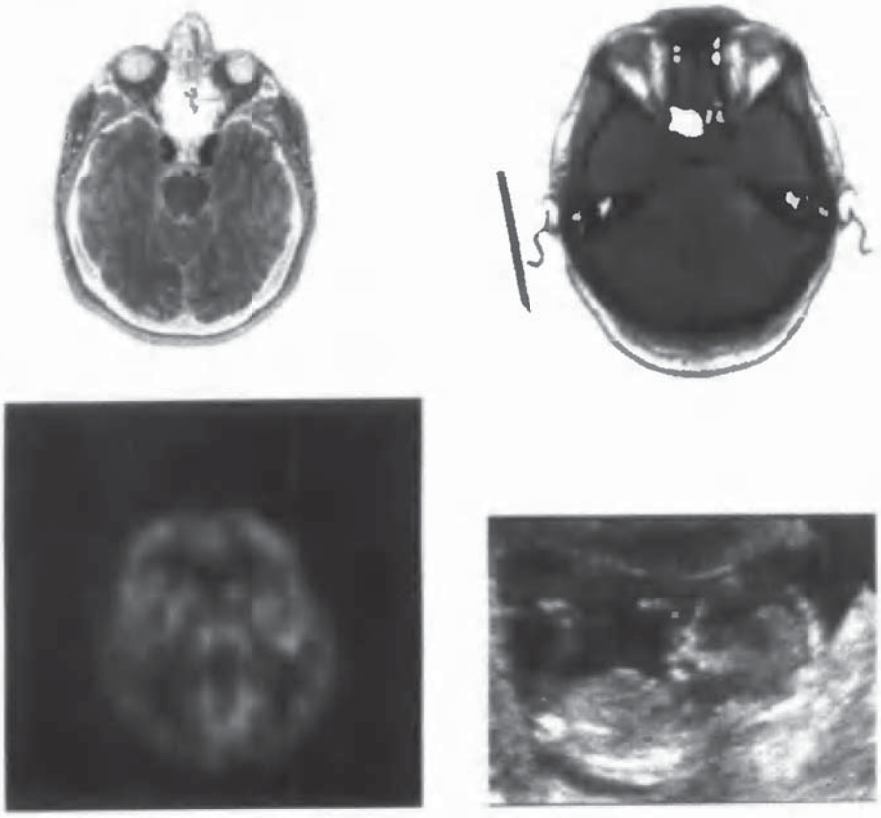


图 18.6 四种不同成像方式获得的图像。左上是一个头骨横切面的 MRI 图像；右上是头骨的 CTI 图像；左下是脑的 NMI 图像；右下是子宫里胎儿的 USI 图像。注意每一种方法如何以不同的方式表示不同的细节；MRI图像与CTI图像有较高分辨率。NMI图像是低分辨率，但是它反映了器官的功能，因为反映强烈的区域吸收了部分试剂。最后，USI图像有很多的噪声，但是显示了软组织的细节，可以从胎儿上面认出腿、身体、头和手

医学图像的对准是一个三维到三维的问题，要考虑的变换只是三维旋转和平移。几何散列是占主导地位的方法，因为它可以有效的用来搜索对应关系。这些文献主要是在用什么数据上有很大不同。

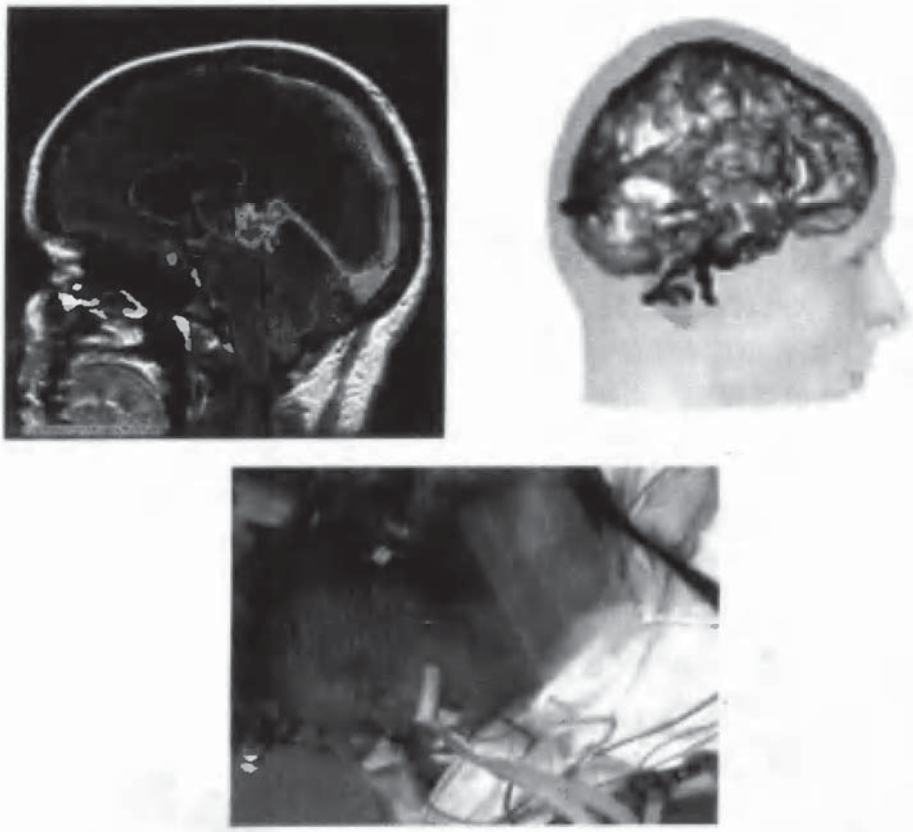


图 18.7　左上图是 MRI 数据的一个切片图附上自动获得的分割。分出了脑的轮廓，脑中的液泡和肿瘤。MRI产生了一系列的切片，从而形成体模型；右上图显示了分割好的体模型，不同颜色表示不同的区域。一旦获得数据，就可以对准到一个躺在手术台上的病人。对准使用了激光测距仪获得的深度信息；下图显示了一个摄像机看到的病人，上面叠加激光测距数据

18.6.2　对准的应用

在脑外科的应用中，医生试图在对病人伤害最少的情况下切除肿瘤。这里所举的是Grimson和他的同事的例子。一般的方法是获得病人大脑的图像，分割这些图像以便显示肿瘤，然后展示给医生看。图像覆盖在躺在手术台上的病人的图片上，而该图片是由医生视角的摄像机拍摄的，以便使医生知道肿瘤的正确位置。有很多种方法可以把反映操作的标签贴在脑图像上(一般采用刺激脑的一个区域，然后看会发生什么)，这些信息显示给医生，目的是为了将对病人的伤害降到最小。这里的问题纯粹是位姿估计；需要知道脑图像相对于在手术台上的人的位姿和脑的测量。

重构外科　提供类似的应用。例如，在面部重构手术的规划阶段，医生可以在病人头骨的一个可视化模型上进行一系列的操作，可以显示给医生看可视化的结果。当医生做手术时，也要对准到病人的视图上。

诊断应用　提供三维可视化手段以显示从不同成像方式中获得的结果。例如，医生可能有 MRI 的图像(一般有较高分辨率)和 PET(一般与功能属性相联系)。他自然希望将这些图像融合，因此同样需要对准。

18.6.3 医学成像中的几何散列技术

在应用中的各种算法之间的主要不同是几何散列中用到的测量的类别。下面分别讨论几种情况：

点对应关系 我们已经知道如何搜索点对应关系。比如说，头部 MRI 数据可以对准到手术台上的病人头部，通过用激光测距仪获得头的三维距离，然后用来和 MRI 数据的皮肤点对准。可以把 MRI 的皮肤点对按照它们分开的距离和法线夹角进行散列，然后从激光束距离进行点对查询。分开距离和角度相似的点对，会产生响应。有了一个对应点对，就可以估计位姿，然后检查位姿估计的总误差。由于 MRI 图像的皮肤点和激光测距数据都不是由孤立点组成的，它们只有表面的采样，因此并没有真实对应，尽管如此，只要采样足够密集，很有可能获得好的位姿初始假设。如果能较好地算出对准误差，仅仅算入垂直于表面的部分以避免因为小的定位误差的影响，可以通过最小化误差获得极好的位姿估计。Grimson 是这种系统的主要倡导者，其中一个系统在图 18.8 中展示。

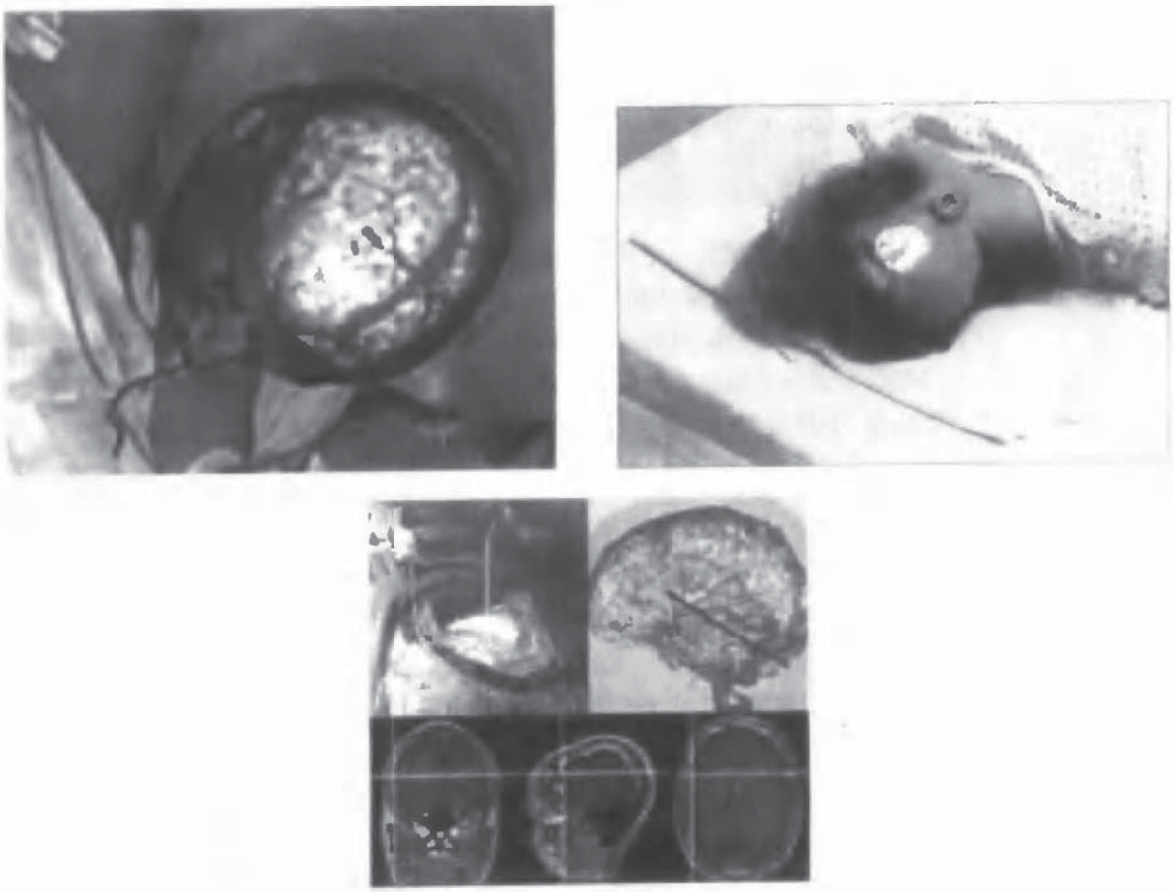


图 18.8 左上图显示了在一个病人的图上叠加的皮肤信息，以显示对准过程的成功应用。一旦数据对准到病人，就可以有很多应用。在右上图，是一个在手术台上的病人用MRI成像显示他的大脑的一部分和肿瘤信息。下图是把手术器械的位置与MRI数据对准在一起生成的图像。这意味着不管手术器械在软组织里有多深，手术器械的位置都可以显示给医生看

曲线 曲线也能用于驱动几何散列。在这种情形下，将数据拟合成曲面，然后在曲面上标记重要的曲线。使用抛物线是一种不切实际的方法，因为一些数据集会有很多平坦区域。

Ayache(1995*a*,*b*)成功地使用了最大法曲率——沿着曲面上最大法曲率的曲线取最大法曲率的极值。不管用到哪一种曲线,在曲线上的任意点,可以有一个完整的三维坐标系(Serret-Frenet 坐标系),然后可以用这个规范坐标系来获得几何散列的测量。自然选择包括曲线的曲率和挠率,以及曲线法向量与表面法向量的夹角。

坐标架构对　给定两个坐标架构,从一个坐标轴到另一个的变换对于共享的变换是不变量,因此也可以用做散列的索引。这个方法称为框架对。一个获取坐标架构对的方法是,将数据拟合成曲面,识别重要的曲线与点,然后使用局部坐标系。对于表面上一个有意义的点(脐点)可以用法线方向和两个曲率极值的方向组成一个坐标系。对于重要曲线上的任意点,可以用 Serret-Frenet 坐标系,或者表面上的坐标系(或比较这两者)。Ayache 和他的工作组着重强调了曲线和坐标架构对的使用; Ayache(1995*a*,*b*) 有更广泛的回顾。

18.7　曲面与对准

曲面同样也可以对准。产生假设的过程可能更复杂,但是绘制和校验却是直接的推广。

很自然的策略是找一些作用类似于点的结构特征组。例如,如果曲面有绘制在上面的点或者三个面不连续地交汇于一个点,那么产生假设过程就像以前描述的一样。然后一个表面的几何模型可以投影在图像上,并按照下面提到的方法来校验。

一个更复杂的方法把对准看做是最小化,而不是以前所讨论的模型的线性组合方法。在这个方法里,表面的轮廓用一个表面位姿的函数预测。可以把物体轮廓与所选择边缘点的最小距离之和作为一个目标函数,然后相对于位姿使该目标函数最小化,像图 18.9 所示。对于代数曲面的情形,预测轮廓曲线的方法可以简化;所有文献中的例子都采用代数曲面作为示例也概因于此。在 Kriegman 和 Ponce(1990*b*)中有详细描述,代数曲面的刚性使得仅用简单的轮廓就足以完全确定曲面的几何性质。感兴趣者可以参考 Forsyth(1996)。

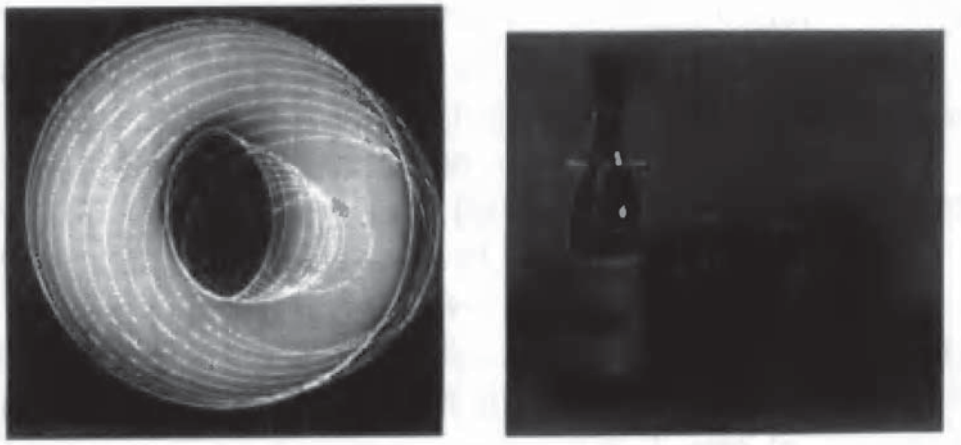


图 18.9　通过标定的透视摄像机观察一个代数形式描述的表面。产生的轮廓是一个清晰的代数曲线(因为我们可以写出该曲线所满足的一簇多项式方程),同时它的外形也是一个代数曲线。该曲线是一个表面位姿的方程。左图显示了一簇曲线,覆盖了图像中曲面的外形,通过最小化选择的边缘点与外形边缘的距离得到其位姿。曲线由最小化过程中不同的点组成。右图显示了两个不同的代数曲面成功的对准了图像的轮廓;曲面识别了两个不同的瓶子

18.8 注释

用这章描述的算法所建立的系统，一般能在杂物很多的场景下识别少量物体。这是非常重要的，因为位姿恢复和对准在应用中都非常有用，同时也因为它们的弱点指出了识别中的重要问题。

术语对准来自 Huttenlocher 和 Ullman(1990)，这是一个描述一类关于位姿一致性算法的非常好的术语。很难确定是谁先使用的这个方法，似乎是 Roberts(1965)；也可能是 Faugeras, Hebert, Pauchon 和 Ponce(1984)。同时代还有 Chin 和 Dyer(1986) 的综述文章。对准是具有相当一般性的，因为对于大多数可想像的物体的可想像的图像来说，摄像机某些形式的一致性限制是合适的，并且可以加以利用。因为只需要相对较少的图像证据就可以用来建立物体假设，对准算法抗噪声的能力也是相当强的，所以能在非常多杂物的图像中找到物体。鲁棒地测试一大堆假设比组装一个大的假设要容易些，因为待测试的假设对图像证据有更多的约束。一些对准算法的抗噪声性能已仔细研究过(Grimson, Huttenlocher 和 Alter, 1992; Grimson, Huttenlocher 和 Jacobs, 1994; Grimson, Lozano-Perez 和 Huttenlocher, 1990; Sarachik 和 Grimson, 1993)。因此对准算法被广泛应用，并且有大量变种。

然而，在模型数目增加之后，对准算法效率不高。模型数目线性增长因为模型库的组织是平面形的(flat)，没有层次关系，因此所有模型都被同等对待。此外，对于一个已经存在的模型的约束搜索可能非常有效，要证明模型不存在是很麻烦的(Grimson, 1992)。

位姿聚类出自 Thompson 和 Mundy(1987)。和哈夫变换的相似形，表明在有噪声的情况，算法的效果会差很多(Grimson 和 Huttenlocher, 1990*b*)。不同形式的几何散列见 Lamdan, Schwartz 和 Wolfson(1990); Wolfson(1990) 以及 Wolfson 和 Lamdan(1988)。使用各种不变量来索引识别假设见 Forsyth, Mundy, Zisserman, Coelho, Heller 和 Rothwell(1991)文献中描述；以及文集、书籍(Mundy 和 Zisserman, 1992; Mundy, Zisserman 和 Forsyth, 1993; Rothwell, 1995)和很多论文中(Barrett, Payton, Haag 和 Brill, 1991; Forsyth 等, 1992; Rothwell, Zisserman, Forsyth 和 Mundy, 1995; Rothwell, Zisserman, Marinos, Forsyth 和 Mundy, 1992; Zisserman, Forsyth, Mundy, Rothwell, Liu 和 Pillow, 1995*a*)。从多个、未标定的视图里检测三维物体的不变量性质，是一个本书没有描述但现在非常活跃的领域(Barrett, Brill, Haag 和 Payton, 1992)。另一个是在某些模型知识条件下计算不变量值的过程(Shashua, 1995; Weinshall, 1993)。还有一些视觉领域的工作研究获得不变量的方法(Csurka 和 Faugeras, 1998, 1999)。

运用位姿一致性可以有很多不同的形式。例如，识别假设产生摄像机内部参数的估计。这表示如果在一个图像中有多个物体，那么所有物体都必须有一致的摄像机内部参数(Forsyth, Mundy, Zisserman 和 Rothwell, 1994)。

每个算法都试图获得足够的信息来执行校验，试图尽可能减少错误的校验尝试次数。这种对图像层次(image level)校验的依赖性与模型提取的要求是不相称的；如果没有这个物种(鱼)和外形的详细信息，是不能依靠像素值或边缘来判断图片里是否包含一只鱼的。但是大多数识别应用非常需要抽象。例如，如果想在 Internet 上搜寻罗马教皇的图片，必须预先知道它的详细几何细节是不可想像的。同样的，如果想在公路上自动驾驶摩托车，就必须决定什么时候转弯以避免碰撞和什么时候必须减速。匹配的不是特征(点和直线)，而是标记特征(标记

特征可能是有毛的斑点,眼睛,或者诸如此类)和包含标记特征的模型。这里,抽取过程是在标记特征里完成的(Ettinger, 1988; Ullman, 1996)。

校验的主要角色是寻找一个假设的证据,用其他方法是收集不到的。由于收集证据的方法目前还理解得不够,目前的识别系统只有当校验工作正常时才正常工作。当有足够证据被使用且被适当计分时,校验才能很好地工作。遗憾的是,很难把这些老生常谈转化成算法。对于一个对识别系统,性能很关键的内容,校验没有被很好地研究(可以参考 Grimson 和 Huttenlocher, 1991)。基于一般证据的校验(边缘点)的困难在于,我们不知道如何统计证据。同样,如果用一些特殊证据(一个特殊伪装的图案),就很难进行抽象化。在第 22 章详细讨论的模板匹配和基于表象的视觉,可以认为是一种在校验过程中包含更多种证据的方法。

医学应用

这个题目我们无法很权威地论证。Ayache(1995*a*, *b*), Duncan 和 Ayache(2000), 以及 Gerig, Pun 和 Ratib(1994)是很有价值的综述。

三个主要的题目有:分割,用来鉴别对应于某些特定器官的图像区域;对准,用来构造不同模式图像之间和图像与病人之间的对应关系;分析,形态学分析(有多大、是否生长)和功能分析。McInerney 和 Terzopolous(1996)综述了可变形模型的应用。关于对准方法与讨论,见 Lavallee(1996) 以及 Maintz 和 Viergever(1998)。对准输出与"实际情况"的比较见 West, Fitzpatrick, Wang, Dawant, Maurer, Kessler, Maciunas, Barillot, Lemoine, Collignon, Maes, Suetens, Vandermeulen, van den Elsen, Napel, Sumanaweera, Harkness, Hemler, Hill, Hawkes, Studholme, Antoine Maintz, Viergever, Malandain, Pennec, Noz, Maguire, Pollack, Pelizzari, Robb, Hanson 和 Woods(1997)。

习题

18.1　假设用一个标定过的透视摄像机看物体,并且希望用位姿一致性算法识别。

(a) 证明三个点是一个结构特征组;

(b) 证明一条线和一个点不是结构特征组;

(c) 解释为什么让结构特征组包含不同种类的特征是一个好主意;

(d) 一个圆和不在它圆心的一个点是一个结构特征组吗?

18.2　有一组平面点 $\boldsymbol{P}_j$,它们满足平面仿射变换。证明:

$$\frac{\det\left[\boldsymbol{P}_i\boldsymbol{P}_j\boldsymbol{P}_k\right]}{\det\left[\boldsymbol{P}_i\boldsymbol{P}_j\boldsymbol{P}_l\right]}$$

是一个仿射不变量(只要 i,j,k,l 两两不同,任意三个不共线)。

18.3　用上题的结果构造一个仿射不变量

(a) 四条直线

(b) 三个共面点

(c) 一条直线和两个点

18.4　在一个斜面匹配的过程中,如果从某个或所有邻居到一条边的距离已知,那么一个点可以更新。Borgefors 算出来一个点到垂直或水平近邻的距离是 3,到对角线近邻的距离

是 4,并且保证像素值是整数。为什么这意味着$\sqrt{2}$近似为 4/3? 有没有更好的近似?

18.5 一种改进位姿估计的方法是采用一个校验分数、然后把它当做位姿的函数来最优化。我们说,如果判断一个反向投影的曲线是否接近一个边缘点的测试是距离的一个阈值,这个优化可能会面临特殊的困难。为什么这会导致一个很难的优化问题?

18.6 证明:对于未标定的仿射摄像机下的一个平面点集,摄像机的效果可以写成一个未知的平面仿射变换。如果是未标定的透视摄像机来看平面点集呢?

18.7 除了我们讨论过的散列,总结一下在医学成像中用到的对准方法。应该记住实际限制,指出你喜欢哪个方法,为什么?

18.8 总结非医学应用的对准和位姿一致性。

18.9 把一个物体表示成模型的线性组合经常称为抽象,因为可以将调整系数看做是获得不同模型的相同视图。此外,可以通过对空间加一些基本元素,得到带参数的模型。建立一个系统,匹配矩形的建筑物,其中建筑物的高、宽和长度都是未知的参数。应该扩展线性组合的思想来处理正交摄像机,包含对表示旋转的系数的限制。

第 19 章　平滑表面及其轮廓

本书的前几章已对诸如点、线、面等简单几何图形及其图像投影的参数之间的关系做了定量研究，本章我们将对三维图形进行定性研究，重点研究具有平滑表面的固体的轮廓，特别强调的是，假定这类固体的表面及其反射函数也是平滑的。忽略阴影面(假设一个点光源及摄像机处在远处场景同一点)，我们认为除了固体轮廓外其他地方也是光滑的。本章中固体轮廓也称为固体剪影(如图 19.1 所示)。

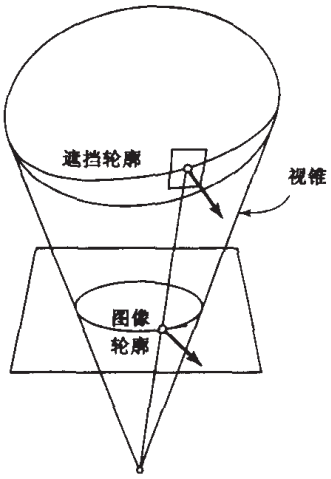


图 19.1　光滑表面的遮挡边界

这个轮廓是通过用一个视锥(或正交投影下的一个圆柱体)与视网膜(retina)的正交而形成的，视锥的顶点正好与针孔摄像机重合并且它的表面沿着一个叫遮挡轮廓(occluding contour)或边缘的曲线擦过(graze)物体。我们可以看到这个遮挡轮廓大体上也是一个平滑曲线，该曲线由视射线与固体表面相切形成的许多折返点(fold point)和一些离散的歧点(cusp point)组成，在这些歧点上射线不仅与表面相切，还与遮挡轮廓(occluding contour)相切。图像轮廓是分段平滑的，它的奇异性是一些由歧点的投影形成的歧点和由一些成对的折返点的交错重叠形成的 T 型结点(T-junction)(如图 19.2 所示)。这些术语的本意是相当明显的：折返点是指在这个点上表面因折返而离开与它相切的视线；轮廓歧点是指在这个点上轮廓的走向突然从另一路径返回，但沿着同一切线方向的点(仅适用于透明物体，不透明物体结束于歧点上，如图 19.2 所示)；同样地，两个平滑的轮廓交叉后形成的叫 T 型结点。

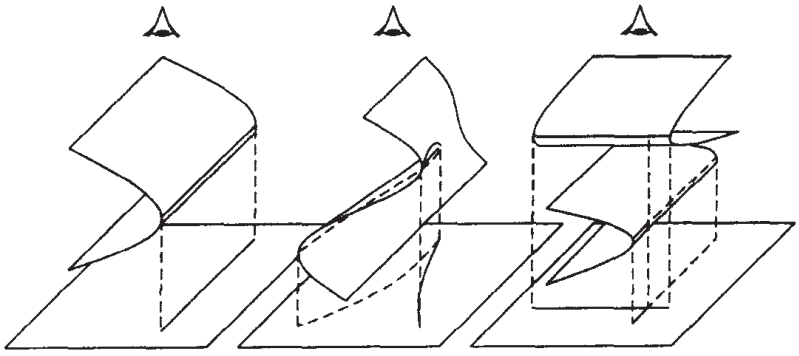


图 19.2　轮廓元素：折返点、歧点和 T 型结点

有趣的是,附属阴影是由于光源有关联的遮挡轮廓勾画出来的,而投射阴影(cast shadows)是以相应的物体轮廓为界的,这样,我们也可以知道它们的样子[提示:在阴影被投射的物体上也有弯曲的表面(如图 19.3 所示),然而,即使在这种情况下,附属阴影的边界也只是遮挡轮廓,当然,光源很少有点状的,这进一步增加了事物的复杂性]。

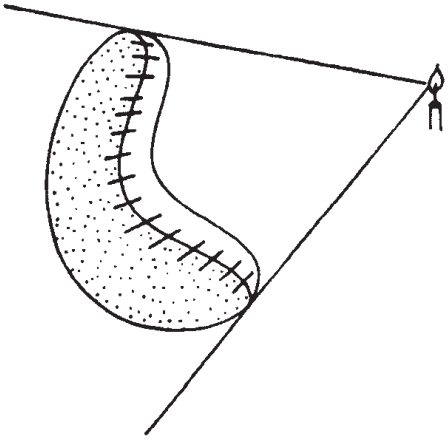


图 19.3 阴影边界和遮挡轮廓

一个固体形状的图像轮廓将它限制在相关的视锥内,但是并不能显示它的遮挡轮廓的深度,对于平滑表面的固体,它提供了额外的信息,特别要提到的是由眼睛与图像轮廓切线定义的平面与这个表面是相切的,这样,轮廓方向决定了沿着这个遮挡轮廓的表面方向。概括说来,这章主要研究曲线表面以及它的投影之间的几何关系和可以从轮廓几何推出来的有关表面几何的信息类型,比如说,轮廓曲率也显示了有关表面曲率的信息。我们从介绍有关微分几何的一些基本概念开始,这能给我们的学习提供一种自然的数学环境。在下一章研究因视点改变而导致的物体外观改变时,微分几何也是很有用的。

19.1 微分几何的基本要点

这一节将介绍有关欧氏微分几何的初步知识,这对学习光线和固体之间的局部关系是很有必要的。我们将讨论限定在三维欧氏空间$\mathbb{E}^3$的有界紧凑固体范围内,讨论的话题自然是技术性很强的,但我们力图用非正式的方式讨论问题,强调画法几何的方式,而不是以解析几何的方式。特别要强调的是,我们没有在空间$\mathbb{E}^3$选择一个全局坐标系统,尽管在某些场合在曲线或表面上一点的邻域中附加了局部坐标系统。这对定性的几何推理来说是很合适的,这也是这一章的重点。解析微分几何则留到第 21 章中对距离数据的(定量)分析中再讨论。

19.1.1 曲线

让我们先来研究在平面上的曲线。在 P 点的附近研究曲线 γ,并假设 γ 本身不相交,即使相交的话,也是在点 P 相交。如果通过 P 画一条直线 L,通常该直线会与 γ 相交于某点 Q,因而定义了一条割线(如图 19.4 所示)。当 Q 趋向于 P 时,割线 L 绕 P 旋转到达一个极限位置 T,那么称这条线为 γ 在 P 点的切线。

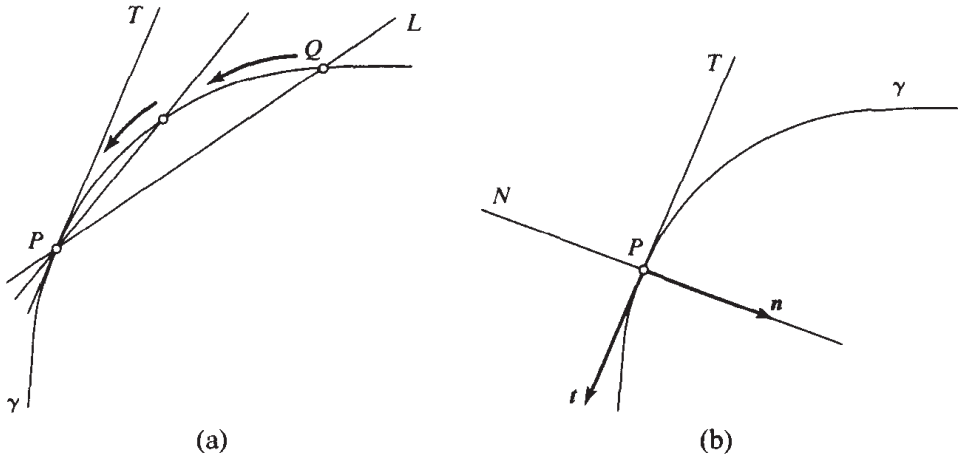


图 19.4　切线和法线:(a)切线定义为割线的极限;(b)由切线和法线定义的坐标系

通过以上构建,切线 T 与 γ 的接触比其他任何一条通过 P 的线都更密切。通过 P 画第二条线 N 并且与 T 垂直,称这条线为 γ 的法线。沿着 T 定义一单位切线向量 $\boldsymbol{t}$,这样可以建立一个右旋坐标系,原点是 P,坐标轴是 $\boldsymbol{t}$,及沿 N 的单位法线向量 $\boldsymbol{n}$,这个局部坐标系统非常适合研究 P 点附近的曲线;坐标轴将这个平面分为 4 个象限,这 4 个象限沿逆时针方向,如图 19.5所示,选择第一象限的方法是使其中某点沿这条曲线 γ 向原点靠近,那么,该点在通过 P 之后将在哪一个象限停止呢?

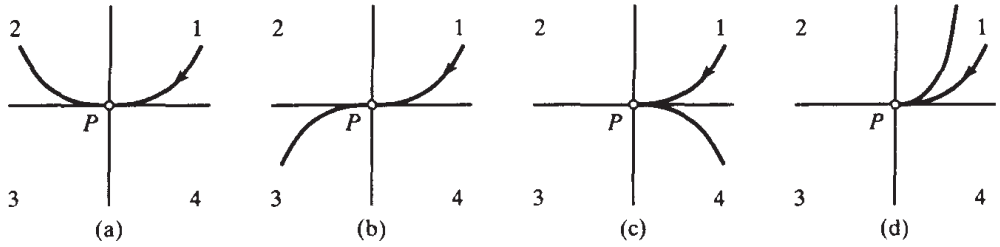


图 19.5　曲线点的分类:(a)规则点;(b)拐点;(c)第一类歧点;(d)第二类歧点,注意在规则点上曲线在切线的同侧

如图所示,这个问题会有 4 种可能的答案,它们描述了点 P 附近曲线图形的特征,我们称移动点在第二象限停止的点 P 是规则的,反之就是奇异的。当这个点通过切线后停止在第三象限,则称点 P 为曲线的拐点。在剩下的两种情况中,我们称点 P 是相应的第一类和第二类歧点,这种分类与 γ 的所选方向无关,它表明几乎所有曲线的所有点都是规则的,而奇异性只发生在一些孤立点上。

正像以前所说的,γ 在点 P 的切线是所有通过此点对 γ 最好的线性近似,而建立最接近的圆周近似,可以定义点 P 的曲率——曲线的另一种基本特性。设一点 P' 沿着曲线趋向于 P,并且让 M 表示在 P 与 P' 的法线 N 和 N' 的交点(如图 19.6 所示),当 P' 趋向 P 时,M 沿着法线 N 到达一个极限位置 C,称它为 γ 在点 P 的曲率中心。

同时,如果用 $\delta\theta$ 表示 N 和 N' 之间的角度,δs 表示连接 P 和 P' 的长度,那么当 $\delta s \to 0$,$\delta\theta/\delta s$ 趋向于一个极限 κ,κ 就称为曲线 γ 在点 P 的曲率,这表明 κ 恰好是 C 和 P 之间的距离 r 的倒数(这点类似于当角度 u 很小时,$u \approx u$),以 C 为中心,称半径为 r 的圆为点 P 的曲率圆,r 为曲率半径。

也可以看到,通过 P 及两个相邻的点 P' 和 P'' 的圆,当 P' 和 P'' 趋向 P 时,这个圆的圆心趋向于曲率的中心,这个圆确实是通过 P 与 γ 最接近的圆。在拐点曲率是 0,并且曲率圆在那儿退化成为一条直线(切线),拐点是沿曲线最平直的点。

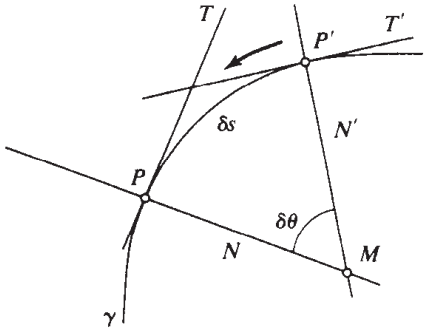


图 19.6 曲率中心定义为通过 P 邻域的法线相交点的极限

让我们来介绍一种工具——高斯图,它被证明在学习曲线和平面时是相当重要的。我们为曲线 γ 选一个方向,并且将 γ 上的每一个点 P 的单位法向量与一个单位圆上的点 Q 相联系,这是相应法向量的端点落在单位圆上的点(如图 19.7 所示),从 γ 到单位圆的映射即 γ 的高斯图[①]。

让我们再来看一看这个定义曲率的极限过程。当 P' 沿曲线趋向 P 时,P' 的高斯图像 Q' 趋向于 P 的映像 Q,N 和 N' 的夹角等于单位圆上 Q 和 Q' 的弧长,因此,当高斯图上相应弧长与曲线上相应弧长接近于零时,曲率即两个弧长的比率的极限。

高斯图也向我们提供了以前曾介绍过的曲线点分类的解释。考虑一个沿着曲线移动的微粒以及它的高斯图的移动。在规则点及拐点,γ 的遍历方向不变,但是在两种类型的歧点上改变了方向(如图 19.5 所示),另一方面,在规则点及第一类歧点,高斯图的遍历方向保持不变,但是在拐点及第二类歧点则改变方向(如图 19.7 所示),这表明了在奇异点附近单位圆是双重覆盖的,我们说高斯图在这些点上发生折叠。

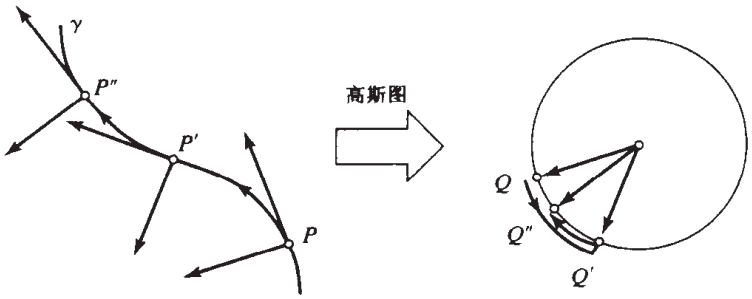


图 19.7 平面曲线的高斯图,观察高斯图的遍历方向如何在曲线拐点 P' 上改变方向,也要注意 P' 两侧的邻近点有平行切线或法线,高斯图在相应的 Q' 点折返

可以通过为曲线选择方向来为平面曲线 γ 上任何一点的曲率选择一个符号标志,例如,凸点的曲率为正:它的曲率中心与有向法向向量的顶点处在 γ 的同侧;凹点的曲率为负,它的上述两点位于 γ 的异侧,因此,曲率在拐点上改变方向,并且改变曲线的方向也会改变曲率的符号。

空间曲线比平面曲线要复杂得多,尽管切线可以像以前那样,定义为割线的极限,但是在点 P 有无穷多的直线与切线垂直,形成曲线在该点的法平面(如图 19.8 所示)。

一般说来,在一个点附近的空间曲线并不在一个平面上,但是的确存在一个惟一的平面与它最贴近,这就是密切平面,它定义为一个平面的极限情况,该平面包括 P 点切线以及一个趋向 P 的邻近点 Q。在点 P 画一个与法平面和密切平面都垂直地从切面就完成了对局部坐标系的建立过程。这个坐标系的坐标轴称为移动三面体或 Frénet 框架,分别由切线、法平面和密切平面相交形成的主法线以及由法平面和从切面相交形成的副法线组成(如图 19.8 所示)。

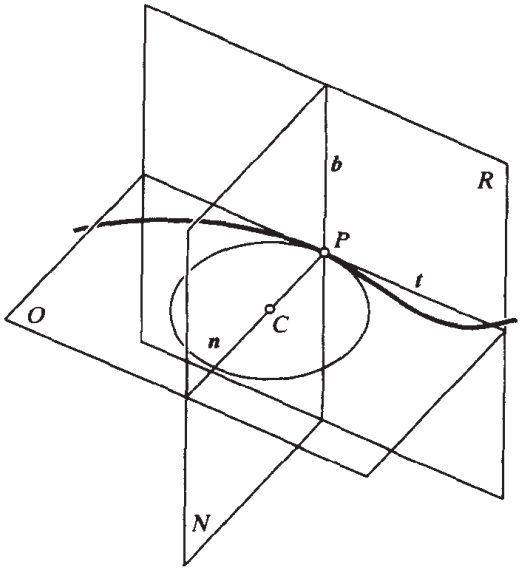


图 19.8　空间曲线的局部几何图形，N、O、R 分别是法平面、密切平面以及伸长平面，t、n、b分别是切线、主法线和副法线，C是曲率的中心

与平面情况相同，空间曲线的曲率可以用多种方式来定义，比如说，可以定义为当三个曲线点彼此接近时的极限圆半径的倒数（这个曲率圆位于密切平面），或定义为两个相邻点的距离趋向于0时，它们的切线夹角与这些点的距离的极限比率。高斯图的概念也可同样延伸到空间曲线，但这时切线、主法线和副法线的端点在一个单位球体上画曲线。应当注意的是，给空间曲线一个有意义的符号是不可能的。一般来说，这样一种曲线不会有拐点并且它的曲率在各点都是正的。

曲率可设想为沿着曲线的切线方向变化速度的量度。我们也可定义密切平面方向沿空间曲线变化的速率：设想 P 和 P' 是曲线上的两个相邻的点，我们可以测量两个相关的密切平面之间的夹角，或相关的副法线之间的夹角，并除以两个点之间的距离，当 P' 趋向 P 时，这个比率的极限称为曲线在点 P 的挠率（torsion）。不足为奇的是，它的倒数是曲线的相应弧长和副法线的球面图像长度之间比率的极限，见图 19.9。

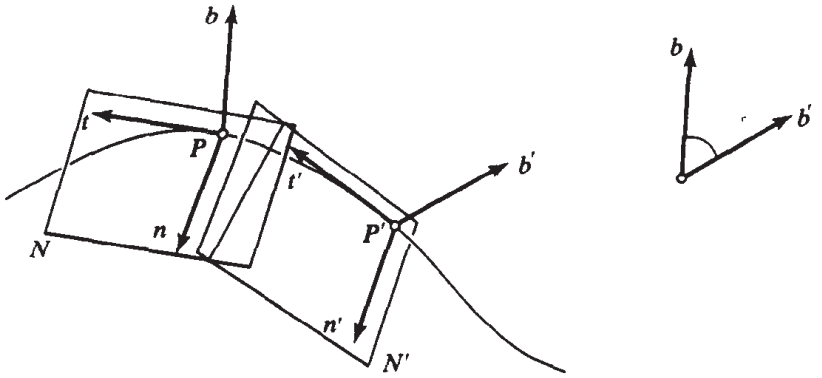


图 19.9　挠率的几何定义：当副法线之间的夹角及相应两点之间的距离都趋于 0 时，二者之比的极限

空间曲线的方向可以通过把曲线认为是移动微粒的轨迹，并为微粒选定一个方向来确定。更进一步说，可以在曲线上选择任意一个参考点 P_0，并且定义与任意点 P 相联系的弧长作为 P_0 和 P 之间的曲线弧的长度，尽管这个弧长依赖于 P_0 的选择，但是它的微分却不依赖于 P_0（因为沿着曲线移动 P_0，对弧长来说增加的是一个常量），并且可以很方便地使曲线成为弧长

的函数，而初始点 P_0 的选择可以不确定。特别是，点 P 的切线向量 $\boldsymbol{t}$ 是单位速率 $\frac{d}{ds}P$（定义为：当 P' 趋于 P 且二者之间的距离趋于 0 时，向量 $\frac{1}{\delta s}\overrightarrow{PP'}$ 的极限）。s 反向也导致 $\boldsymbol{t}$ 反向。可以看到加速度 $\frac{d^2}{ds^2}P$、曲率 κ、与主法线 $\boldsymbol{n}$ 之间有以下关系：

$$\frac{d^2}{ds^2}P = \frac{d}{ds}\boldsymbol{t} = \kappa\boldsymbol{n}$$

注意，κ 和 $\boldsymbol{n}$ 都与曲线方向无关（由曲线的反向遍历带来的负号已在微分时消除），曲率即加速率的幅度，副法线向量可定义成 $\boldsymbol{b} = \boldsymbol{t} \times \boldsymbol{n}$；与 $\boldsymbol{t}$ 类似，$\boldsymbol{b}$ 也依赖于为曲线所选择的方向。通常，有如下表示：

$$\begin{cases} \frac{d}{ds}\boldsymbol{t} = \kappa\boldsymbol{n} \\ \frac{d}{ds}\boldsymbol{n} = -\kappa\boldsymbol{t} + \tau\boldsymbol{b} \\ \frac{d}{ds}\boldsymbol{b} = -\tau\boldsymbol{n} \end{cases}$$

此处，τ 代表点 P 的挠率。与曲率不同，对一般空间曲线而言，挠率可以为正、负或 0。它的符号取决于为曲线所选择的遍历方向，并且有一个几何意义：一般来说，曲线以一个非 0 的挠率通过密切平面的每一点，并且当挠率为正时，曲线出现在密切平面的正面（副法线那面）；反之，当挠率为负时，出现在反面。当然，平面曲线的挠率处为 0。

19.1.2 表面

关于平面曲线和空间曲线的局部特征的讨论，很大部分可以推广到对表面的讨论中。设想表面 S 上一点 P 以及所有通过 P 且位于 S 上的曲线。可以证明这些曲线的切线位于同一平面 Π，P 的切平面上[如图 19.10(a)所示]。过 P 和平面 Π 与垂直的直线 N 称为在 S 上 P 点的法线，可以取单位法向量 N 的方向为表面 S 的局部方向（与曲线不同，表面在每一点上允许有惟一的法线和无限条切线），覆盖固体的表面的规范化方向可以定义为指向物体外部的法向量。

通过使表面与包括 P 点法线的平面相交，得到一个具有单参数的平面曲线族，称为法截线[如图 19.10(b)所示]。一般来说，这些曲线在点 P 是规则的或者显示出拐点。法截线的曲率也被称为相应的切线方向上的表面法曲率。通常，当法截线与指向内部的表面法线位于切平面的同一侧时，称法曲率为正，否则为负。当然，当 P 是相应法截线的拐点时，其法曲率为 0。

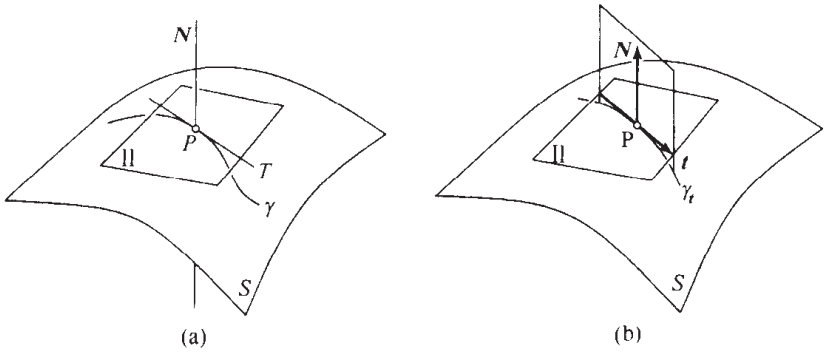


图 19.10 切面和法截线：(a)表面上点 P 的切面 Π 以及法线 N；γ 是通过 P 的表面曲线，它的切线 T 位于 Π；(b)表面 S 与带有法向量 N 和切向量 $\boldsymbol{t}$ 的平面形成的 S 的一个法截线 γ_t

有了这种规定，当截面沿着表面法线旋转时便可记录法曲率值，通常在切平面的某个确定方向呈现其最大值 κ_1，在另一个确定方向达到最小值 κ_2。这两个方向被称为在 P 点的主方向，并且可以看到它们是相互正交的，除非法曲率在所有可能的方向上具有固定的值。主曲率 κ_1 和 κ_2 及其方向为表面定义了一个最佳的局部二次曲面近似，说得更具体一些，如果我们在 P 点建立一个坐标系：沿主方向是 x 轴和 y 轴，沿外向法线建立 z 轴，那么在这种框架下，表面可用抛物线描述为：$z = -1/2(\kappa_1 x^2 + \kappa_2 y^2)$。

一个表面点的邻近区域，根据主曲率的符号可有三种不同的形状(如图 19.11 所示)，两个曲率符号相同的点 P 被认为是椭圆形的，它附近的表面是蛋形的[如图 19.11(a)所示]，它并不跨越它的切平面，看起来像鸡蛋的外壳(正曲率)或者是被打破的蛋壳的内部(负曲率)，我们说 P 在前一种情况下是凸的，在后一种情况下是凹的。当主曲率有两个相反的符号时，就有了一个双曲点，这个表面是马鞍形的，并且沿着两条曲线通过它的切平面[如图 19.11(b)所示]，相应的法截线在点 P 有个拐点，它的切线被称为表面在点 P 的渐近方向，它们被主方向分开。椭圆点和双曲形的点在表面上形成块，这些区域在一般情况下，被由抛物点形成的曲线所隔开，在这些点上主曲率之一消失。相应的主方向也是渐近方向，而表面与它的切平面相交处在那个方向上(一般)有一个歧点[如图 19.11(c)所示]。

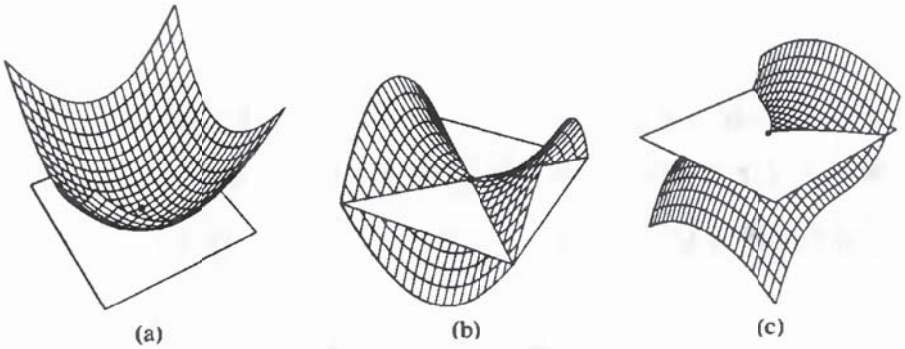


图 19.11　表面的局部形状：(a)椭圆形点；(b)双曲线点；(c)抛物点(实际上抛物点有两类，我们将在第20章来讨论)

我们自然也可以定义表面的高斯图——在相应的单位法线穿过单位球体的地方画点(后边将称其为高斯球)。对于平面曲线，高斯图在规则点的附近是一对一的，但是在一些奇异点的附近高斯图的遍历方向将改变方向。同样也可看到，在椭圆点和双曲点的区域，高斯图是一对一的，以椭圆点为中心的小封闭曲线的方向在高斯图上维持不变，但是以双曲点为中心的封闭曲线的方向是改变的(如图 19.12 所示)。

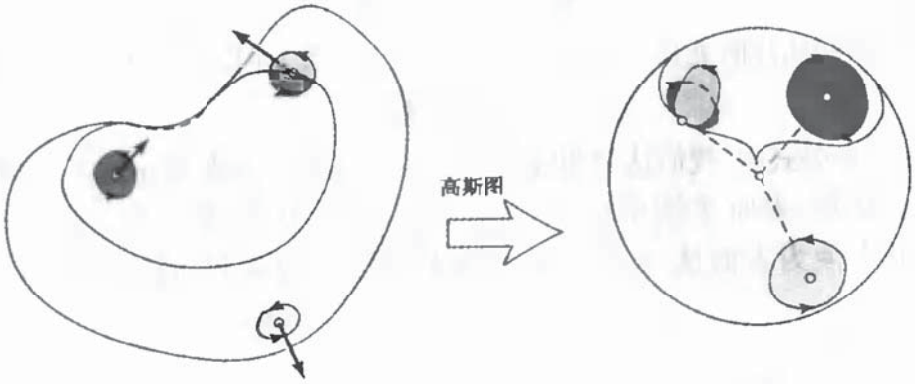


图 19.12　左边：肾形表面，它由凸域、双曲线区域组成，由抛物曲线将它们分开；右边：相应的高斯图像，重阴影部分表明是双曲线区，轻阴影部分是椭圆形区域，注意该表面形状并不是凸的，也不是凹的

在抛物点上的情形要复杂一些。在这种情况下，任何小的区域都包含了有平行法线的点，显示了在抛物点附近球体的双重覆盖(如图 19.12 所示)。我们说，沿着抛物曲线的高斯图是折叠的。请注意它与平面曲线拐点的相似性。

现在考虑一个通过点 P，且用点 P 附近弧长 s 的函数表示的表面曲线 γ。由于限制 γ 的表面法线有一个固定的长度，那么对 s 的导数位于切平面 P 内，而且这个导数的值只与 γ 的单位切线 $\boldsymbol{t}$ 有关而与 γ 本身无关。这样，我们就可以定义一个映射 $d\boldsymbol{N}$，它将在 P 的切平面内每一个单位向量 $\boldsymbol{t}$，与相应的平面法线的导数联系起来(如图 19.13 所示)。当 $\lambda \neq 1$ 时，使用规则 $d\boldsymbol{N}(\lambda \boldsymbol{t}) \stackrel{\text{def}}{=} \lambda d\boldsymbol{N}(\boldsymbol{t})$，可以将 $d\boldsymbol{N}$ 延伸到一个在整个切平面上定义的线性映射，并称其为在点 P 的高斯图微分。

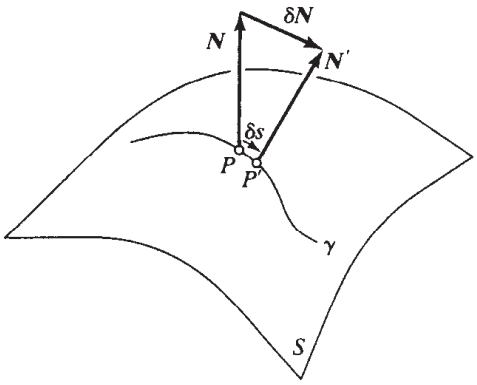


图 19.13 表面法线的方向导数：如果 P 和 P' 是曲线 γ 上的两个邻近点，$\boldsymbol{N}$ 和 $\boldsymbol{N}'$ 表示与之相关的表面法线，令 $\delta \boldsymbol{N} = \boldsymbol{N}' - \boldsymbol{N}$，则表面法线的方向导数定义为：当 P 和 P' 之间的弧长 δs 趋向于 0 时 $\frac{1}{\delta s}\delta \boldsymbol{N}$ 的极限

点 P 的第 2 种基本形式是双线性形式，该形式将切平面内任意两个向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 的关系表示为①：

$$\mathrm{II}(\boldsymbol{u}, \boldsymbol{v}) \stackrel{\text{def}}{=} \boldsymbol{u} \cdot d\boldsymbol{N}(\boldsymbol{v})$$

因为很容易证明 II 是对称的，也就是说，$\mathrm{II}(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{II}(\boldsymbol{v}, \boldsymbol{u})$，因此与任何切向量 $\boldsymbol{u}$ 有关的映射值 $\mathrm{II}(\boldsymbol{u}, \boldsymbol{u})$ 是一个二次方程式，这个二次方程式与通过点 P 的表面曲线的曲率密切相关。实际上，表面曲线上的切线 $\boldsymbol{t}$ 在每一点都与表面法线 $\boldsymbol{N}$ 相垂直，两个向量的点积对曲线弧长求微分，可得：

$$\kappa \boldsymbol{n} \cdot \boldsymbol{N} + \boldsymbol{t} \cdot d\boldsymbol{N}(\boldsymbol{t}) = 0$$

其中，$\boldsymbol{n}$ 代表曲线的主法线，κ 代表它的曲率，这还可写成

$$\mathrm{II}(\boldsymbol{t}, \boldsymbol{t}) = -\kappa \cos \phi \tag{19.1}$$

其中，ϕ 是表面和曲线法线的夹角。对法截线来说，$\boldsymbol{n} = \mp \boldsymbol{N}$，并且在 $\boldsymbol{t}$ 方向上的法曲率是

$$\kappa_t = \mathrm{II}(\boldsymbol{t}, \boldsymbol{t})$$

和以前一样，在这个公式中，我们认为曲线的主法线与表面法线点的方向相反时曲率是正的。此外，式(19.1)说明：表面法线的曲率 κ 与法曲率 κ_t 在切线 $\boldsymbol{t}$ 的方向上有如下关系：$\kappa \cos \phi = -\kappa_t$，其中，$\phi$ 为表面法线的主法线与表面法线的夹角，这就是 Meusnier 定理(如图 19.14 所示)。

① 有时第二基本制定义为 $\mathrm{II}(u, v) = -u \cdot dN(v)$。我们的定义使我们可以对凸形体的表面的指向外部的法线赋以正的法曲率。

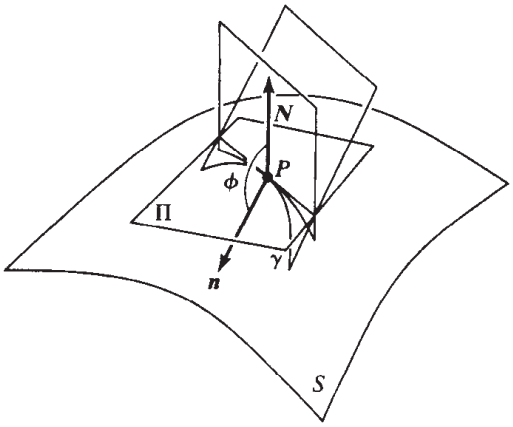


图 19.14　Meusnier 定理

可以看出，主方向是线性映射 dN 的特征向量，主曲率是相关的特征值，线性映射的行列式值 K 称为高斯曲率，它等于主曲率的乘积，因此，高斯曲率的符号决定了表面的局部形状：当 $K>0$ 时，该点是椭圆形的，当 $K<0$ 时是双曲的，当 $K=0$ 时是抛物形的，如果 δA 是表面 S 上以 P 为中心的一块小区域，$\delta A'$ 是 S 在高斯图上相应的一块区域，那么当两个区域的面积趋向于 0 时，$\delta A'/\delta A$ 的极限即高斯曲率。

19.2　表面轮廓几何学

在研究表面轮廓几何学之前，先讨论一下空间曲线 Γ 的局部形状和它在某平面 Π 上的正交投影 γ 之间的关系(如图 19.15 所示)。我们用 α 表示平面 Π 与 Γ 的切线 $\boldsymbol{t}$ 之间的夹角，用 β 表示平面 Π 与 Γ 的密切平面之间的夹角(等价于 Π 的法线与 Γ 的副法线 $\boldsymbol{b}$ 之间的夹角)，这两个角完整地定义了与图像平面相关的曲线的局部方向。

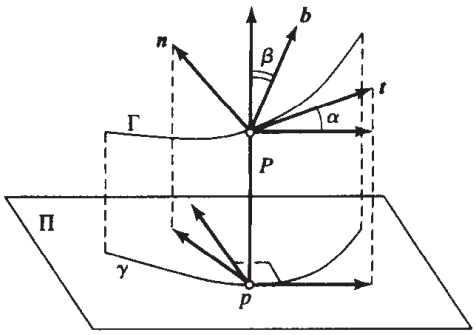


图 19.15　空间曲线及其投影。注意：γ 的切线是 Γ 的切线的投影(即切线是微粒沿着曲线移动的速率)，通常，γ 的法线 $\boldsymbol{n}$ 并不是 Γ 的法线的投影

如果 κ 表示 Γ 上的某一点的曲率，κ_a 为它的表观曲率(即：γ 在相应图像点的曲率)，那么，可以很容易地得到：

$$\kappa_a = \kappa \frac{\cos\beta}{\cos^3\alpha} \tag{19.2}$$

特别地，当视线方向在密切平面内时($\cos\beta=0$)，表观曲率 κ_a 消失，曲线图像得到一个拐

点。另外，如果视线方向与曲线相切（$\cos\alpha = \cos\beta = 0$），$\kappa_a$ 就无法定义，曲线的投影呈现为一个歧点。

空间曲线投影的这两个特性是众所周知的，在微分几何课本里都已描述，那么是否有可能用一种相似的方式把环绕固体的表面的局部形状同它的图像轮廓的形状连接起来呢？答案当然是肯定的，Koenderink（1984）在他的论文“遮挡轮廓到底告诉我们固体形状什么了？”中就给出了肯定的回答。在开始讲述并证明 Koenderink 文章中的定理之前，先来讲述一下图形轮廓的一些基本特性，最后以讨论 Koenderink 文章中定理的一些含义来结束本节。

19.2.1 遮挡轮廓和图像轮廓

如前所述，具有平滑表面的固体的图像被一条图像曲线所包围，该曲线又叫固体的轮廓、剪影或外形轮廓。这条曲线是通过用一个视锥与视网膜（retina）的相交而形成的，圆锥的顶点正好与针孔摄像机重合，并且它的表面沿着一个叫遮挡轮廓或边缘的曲线擦过物体（如图 19.1 所示）。

假定在本节讨论中使用正交投影，在这种情况下，针孔摄像机可以移动至无穷远，此时视锥变成了圆柱，圆柱的母线与视线方向平行，沿着每一条母线，表面法线是固定的，并且与图像平面平行（如图 19.16 所示），在遮挡轮廓某一点的切平面投影到图像轮廓上是一条切线，并且该图像轮廓的法线与在遮挡轮廓相应点的表面法线是相同的。

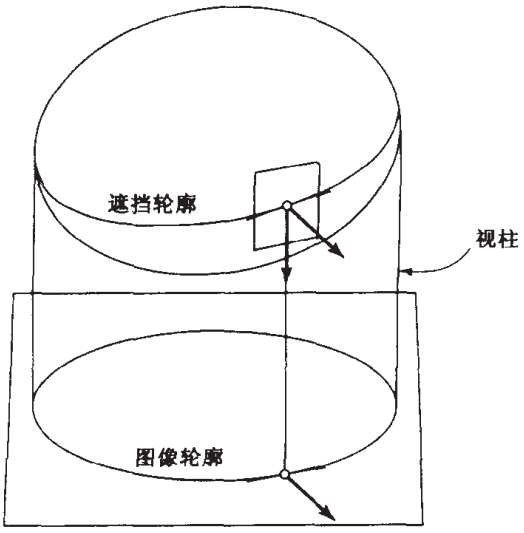


图 19.16 正交投影下的遮挡边界

必须注意的是，视线方向 $\boldsymbol{v}$ 与遮挡轮廓切线 $\boldsymbol{t}$ 通常是不垂直的（就像 Nalwa 在 1988 年曾经说过的：倾斜圆柱的遮挡轮廓与其轴平行，而并非与图像平面平行），如下一节所示，这两个方向是对偶的——这是遮挡轮廓的一个极其重要的性质。

19.2.2 图形轮廓的端点和拐点

当 $\Pi(\boldsymbol{u}, \boldsymbol{v}) = 0$ 时，切面的两个方面 $\boldsymbol{u}, \boldsymbol{v}$ 被认为是对偶的，比如，主要方向是对偶的，因为它们是 $d\boldsymbol{N}$ 的正交特征向量，并且渐近线是自对偶的。

显而易见，遮挡轮廓的切线 $\boldsymbol{t}$ 与相应的投影方向 $\boldsymbol{v}$ 总是对偶的，实际上，$\boldsymbol{v}$ 在遮挡轮廓的

每一点上与表面相切，用曲线的弧长对等式 $N\cdot v=0$ 微分，得：

$$0=\left(\frac{d}{ds}N\right)\cdot v=dN(t)\cdot v=\mathrm{II}(t,v)$$

假设 P_0 为双曲点，然后把这个平面投射到与渐近方向垂直的一个面上，因为渐近方向是自对偶的，那么在 P_0 处的遮挡轮廓必然沿这个方向，如式(19.2)所示，在这种情况下轮廓的曲率必定是无穷的，这个轮廓得到了一个一类歧点。

稍后我们将引出 Koenderink(1984)的一个定理，该定理说明了图像轮廓的曲率与表面的高斯曲率之间的数量关系，在此期间，我们将(非正式地)验证一个较弱的但很显著的结果。

定理 5：在正交投射下，轮廓的拐点是抛物点的映像(如图 19.17 所示)。

为了说明该定理为什么能成立，首先要注意的是，在正交投影下，遮挡轮廓某点的表面法线与图像轮廓相应点的法线相同。因为高斯图在抛物点折叠，那么图像轮廓的高斯图必定在这个点上改变方向。如前所示，平面曲线的高斯图在它的拐点及第二类歧点上改变方向，容易看出，后者奇异性在一般视点下是不会发生的，这就验证了结果。

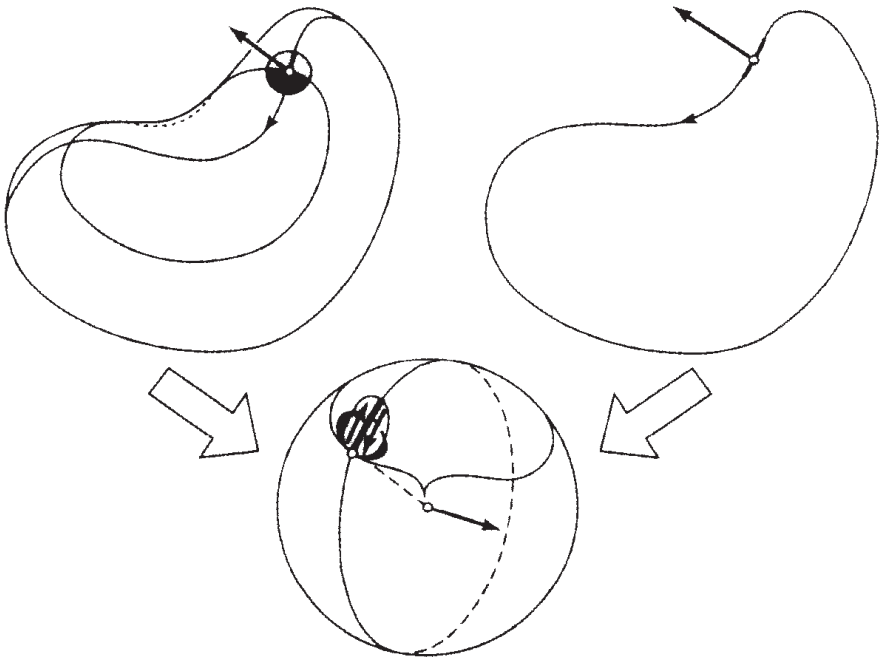


图 19.17　轮廓的拐点是抛物线点的映像：左上图为带有重叠的遮挡轮廓的肾状表面；右上图为相应的图像轮廓；下图为高斯图在抛物线点的折叠，对由遮挡轮廓和图像轮廓的映像的大圆有同样的限制

总之，遮挡轮廓是由一些点组成的，在这些点上，视线方向 v 与表面相切(在序言中已提及过的歧点)。偶尔，它在双曲歧点变为 v 的切线或穿过抛物线，这时在轮廓上会相应地出现第一类歧点或拐点。和至今提到的曲线不同的是，当遮挡轮廓的两个不同分支投射到相同的图像点时，图形轮廓也会跨越它自己，形成 T 结点(如图 19.1 所示)。在一般视点下，以下情况是惟一能发生的：例如，既没有第二类的歧点，也没有任何切线的自交。回头我们将在下一章研究一些特殊的视点以及相应轮廓的奇异性。

19.2.3 Koenderink 定理

下面讨论前面已多次提到的 Koenderink(1984)定理。像以前一样假设在正交投影下,表面 S 的遮挡轮廓上有一点 P,用 p 表示轮廓上它的图像。

定理 6: S 上点 P 的高斯曲率 K 和轮廓曲率 κ_c 有以下关系:

$$K = \kappa_c \kappa_r$$

其中,κ_r 为径向曲线的曲率,该曲线由表面 S 与某一类平面相交形成,该类平面由 S 在点 P 的法线和投影方向定义(如图 19.18 所示)。

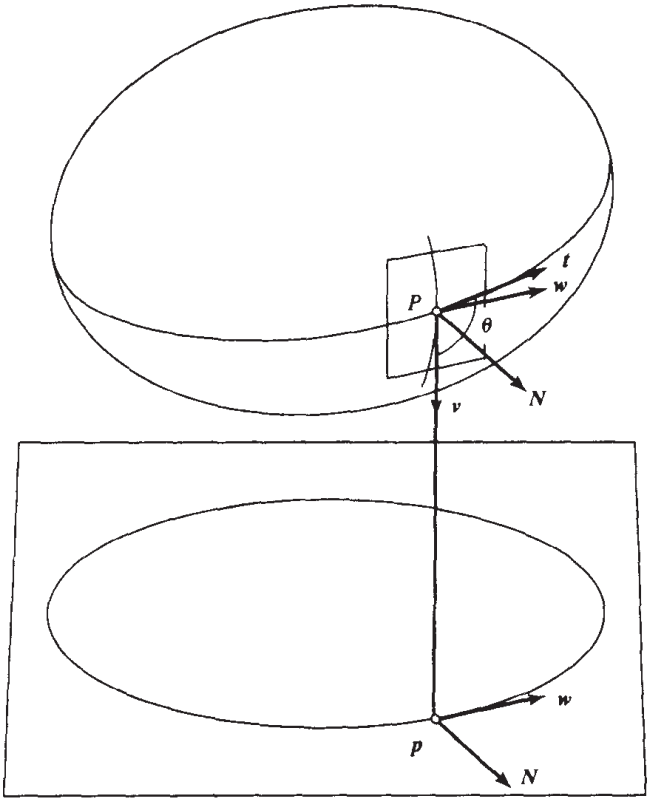


图 19.18 遮挡轮廓和图像轮廓:视线方向 $\boldsymbol{v}$ 与遮挡轮廓的切线 $\boldsymbol{t}$ 是对偶的,对于非透明固体来说,轮廓上可视点的径向曲率通常是非负的

这个看起来很简单的关系式有几个重要的结论(当然是从定理 5 开始的):因为在 $\kappa_r < 0$ 的任意点,投影光线都局部位于成像物体内部,所以沿着遮挡轮廓的径向曲率 κ_r 是正的(或 0),因而,当高斯曲率为正时,κ_c 是正的,反之,κ_c 为负。尤其是,通过这个定理可看出,轮廓的凸面对应于表面的椭圆点;而凹面对应双曲点,拐点对应抛物点。

在所有的椭圆形表面点中,因为它们的切面完全位于固体内部,所以凹点从不出现在不透明物体的遮挡轮廓上,这样,轮廓的凸性也与表面的凸性相符合。同样,当视线方向在双曲线点上是渐近方向时,轮廓上有歧点。对于不透明的物体,这就意味着轮廓的凹弧在这样的歧点结束,在这个歧点处轮廓的一个分支变成被遮挡的,由此可知 Koenderink 的定理加强并精化了图像轮廓的几何特性。

下面我们来验证一下这个定理,这与对偶方向的一般性质有关。如果 κ_u 和 κ_v 在对偶方

向 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 中代表法曲率，K 代表高斯曲率，则：

$$K\sin^2\theta = \kappa_u\kappa_v \tag{19.3}$$

其中，θ 是 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 之间的夹角，这种关系很容易通过这样一个事实得到验证，即以对偶方向形成的切平面为基础的第二基本形式矩阵是对角的（见练习题），很明显，对主要方向（$\theta=\pi/2$）和渐近线方向（$\theta=0$）来说，它是满足的。

通过 Koenderink 定理，我们可以得到

$$K\sin^2\theta = \kappa_r\kappa_t$$

其中，κ_t 代表沿遮挡轮廓方向 $\boldsymbol{t}$ 的表面的法曲率（当然，与遮挡轮廓的实际曲率有区别）。为了完成这个定理的证明，我们用表面的另一常用性质，即：具有切线 $\boldsymbol{t}$ 的任何表面曲线的表观曲率为

$$\kappa_a = \frac{\kappa_t}{\cos^2\alpha} \tag{19.4}$$

其中，像从前一样，α 代表 $\boldsymbol{t}$ 和图像平面间的夹角，像练习中那样，这个性质可以很容易从式（19.2）和Meusnier 定理得到。

换句话说，用切线和图像平面夹角余弦的平方除以相应的法曲率即可得到任意表面曲线的表观曲率。注意，κ_c 是遮挡轮廓的表观曲率，从而有：

$$\kappa_c = \frac{\kappa_t}{\sin^2\theta} \tag{19.5}$$

因为 $\alpha=\theta-\pi/2$。用式（19.5）代入式（19.3），即可完成定理证明。

19.3 注释

关于微分几何有许多优秀的教材，包括 do Carmo（1976）和 Struik（1988）的书中的一些容易理解的内容，我们的描述类似于 Hilbert 和 CohnVossen（1952）的优秀作品 Geometry and the Imagination 中的关于微分几何的引言部分。

人们并非总是认识到图像轮廓含有一些关于表面形状的重要信息（见 Marr，1977 和 Horn，1986 的与对手辩论的内容），本章证明的定理澄清了事实并首次在 Koenderink（1984）的书中出现。我们的证明过程与原始的不同，但与 Koenderink（1990）在他的 Solid Shape（与 Hilbert 和 Cohn-Vossen 的书一样，对任何真正对计算机视觉的几何特性感兴趣的人来说，这都是一本必读之书）这本书的证明相近，之所以这样选择是因为我们不愿意用任何要求特殊坐标系的公式，关于投影几何的各种不同的证明方法可在如下作者的书中找到：Brady 等（1985*a*），Arbogast 和 Mohr（1991），Cipolla 和 Blake（1992），Vaillant 和 Faugeras（1992），Boyer（1996）。

习题

19.1　通过透视照相机，得到的球体的剪影是什么形状？

19.2　通过正交投影照相机，得到的球体的剪影是什么形状？

19.3　证明：在点 P 的平面曲线的曲率 κ 是这一点的径向曲率 r 的倒数。

提示：当 u 很小时，$u \approx u$。

19.4 给定一个固定坐标系，假定用坐标向量来辨识$\mathbb{E}^3$ 点，并设参数曲线为 $x : \boldsymbol{I} \subset \mathbb{R} \rightarrow \mathbb{R}^3$，并非必须以弧长为参数，证明曲率 κ 为：

$$\kappa = \frac{|\boldsymbol{x}' \times \boldsymbol{x}''|}{|\boldsymbol{x}'|^3} \tag{19.6}$$

其中，$\boldsymbol{x}'$和$\boldsymbol{x}''$分别表示与定义它的参数 t 有关的 $\boldsymbol{x}$ 的一阶和二阶导数。

提示：用弧长对 $\boldsymbol{x}$ 重新参数化，并用微分反映参数的变化。

19.5 求证：除非法线曲率在所有可能方向上都是常量，否则主方向是互相垂直的。

19.6 求证：第二类基本形式是双线性与对称的。

19.7 用 α 表示曲线的切线 Γ 与平面 Π 之间的夹角，β 表示 Π 的副法线与 Γ 的法线之间的夹角，用 κ 表示 Γ 上某点的曲率，证明：如果 κ_a 表示的相应点图像的表观曲率，那么：

$$\kappa_a = \kappa \frac{\cos\beta}{\cos^3\alpha}$$

[注释：该结论可从 Koenderink (1990)的第 191 页找到]，

[提示：用向量 $\boldsymbol{t}$，$\boldsymbol{n}$，$\boldsymbol{b}$ 构造坐标系，其 z 轴与图像平面正交，用公式(19.6)计算 κ_a]。

19.8 用 κ_u 和 κ_v 表示点 P 的成对方向 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 的法线曲率，用 K 表示高斯曲率，证明

$$K \sin^2\theta = \kappa_u \kappa_v$$

其中，θ 表示 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 的夹角，

提示：参考根据切平面得到的第二基本形式的表达式，切平面分别由对偶方向和主方向产生。

19.9 证明：遮挡轮廓是一条与自身不相交的平滑曲线。

(提示：使用高斯图)。

19.10 证明：任何带有切线 $\boldsymbol{t}$ 的表面曲线的表观曲率为：

$$\kappa_a = \frac{\kappa_t}{\cos^2\alpha}$$

其中，α 是图像平面和 $\boldsymbol{t}$ 的夹角。

[提示：用向量 $\boldsymbol{t}$，$\boldsymbol{n}$，$\boldsymbol{b}$ 构造坐标系，其 z 轴与图像平面正交，使用式(19.2)和 Meusnier 定理]。

第20章 外 观 图

这一章我们继续研究刚体形状与它们的图像之间的定性关系，但这次把焦点集中在由一个刚体所有视图组成的集合所具有的特点上，并使用微分几何与奇异值理论中的工具将具有相同拓扑结构的图像合并成等价类。为了说明这种概念，让我们首先在扁平世界(Flatland)停留。扁平世界是这样一个世界，其中一只眼的蠓虫在一个二维的场景中漫游。让我们考虑一个由平滑封闭曲线包围的平面物体，并有一个蠓虫看着它(见图 20.1)。我们将蠓虫的眼睛用全方位的圆形摄像机模型表示(也就是说是一个针孔透视投影摄像机，具有圆形成像面以及 360°)，并假设视网膜是有排列取向的(比如顺时针排列)，因此我们能用点的顺序表示其位置。

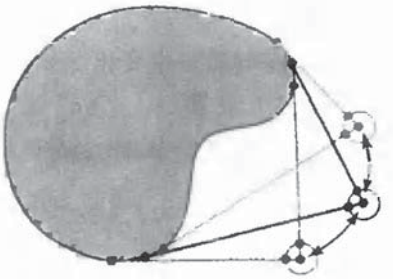


图 20.1 设想在一个扁平世界上：三个全方位圆形摄像机从相接近的视点观察一个二维物体。像这种典型的情况，轮廓点的顺序和数量(或等价地相关的投影射线)对所有三个摄像机都是相同的，这表示了从视点到图像结构的平滑映射

在这种情况下，“遮挡轮廓”是由这样一些(离散)点组成，在这些点上，来自光心发出的光线与物体的轮廓相切，而“图像轮廓”则由这些切向射线与摄像机的圆形视网膜的交点组成。扁平世界内的物体可以是透明的或不透明的；对于后者来说，贯穿部分物体边界的切射线在到达视网膜之前会被阻断。无论哪种情况，摄像机的少量运动会改变轮廓点的位置，但(一般)不会改变它们的数目或沿视网膜的顺序。

然而在某些视点，摄像机的小幅度运动会引起图像结构的剧烈变化：对透明体而言，当眼睛跨越双切线[一根线与物体边界上两个不同处相切，见图[20.2(a)]时，靠近的轮廓点合并后又以相反的顺序分开。对不透明物体来说，当眼睛穿过双切线时一个新的轮廓点出现了(或一个旧的轮廓点消失了)，因此视图在那里也发生了剧烈变化，但变化方式不同。顺序逆转被一个轮廓点的出现(或消失)所取代[见图 20.2(b)]。同样，当眼睛跨越一个透明体曲线拐点的切线时，会有一对轮廓点出现或消失[见图 20.2(c)]。而非透明体只有其中一个点可见，另一个则被物体遮挡住了[见图 20.2(d)]。

这些结构图像的改变称为视觉事件。那么是否只有双切线或拐点切线的关系才能产生视觉事件呢？两个双切线当然也会相交，一个拐点切线与双切线相交等都可能发生，导致物体外观发生更复杂的变化。三切线与拐双切线在直觉上也是可能的。然而在曲线有微小变形时三切线显然就会消失(图 20.3)，但是双切线或两根拐切线的相交却不同，它们将只是跟随变形变动而已。

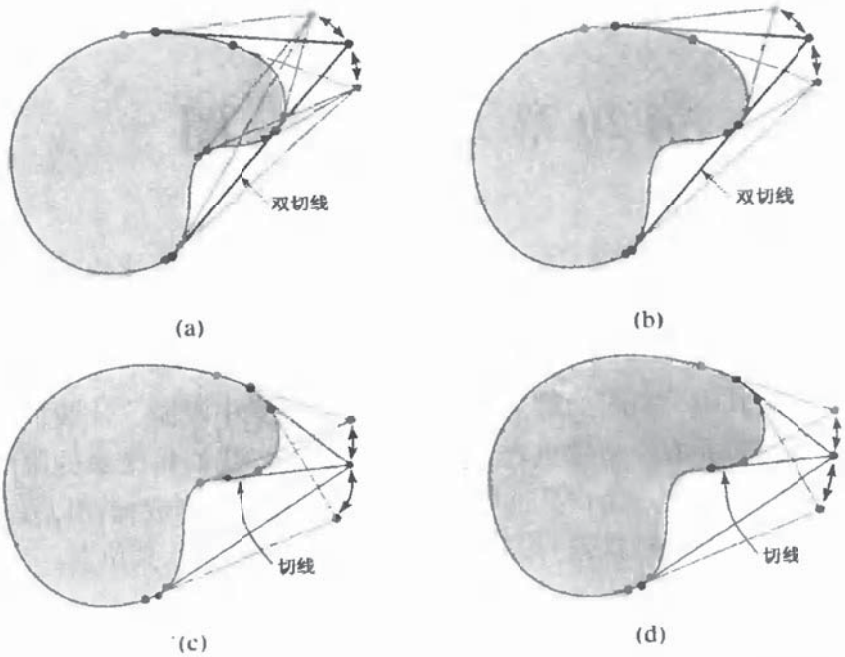


图 20.2 特殊的视角导致图像性质发生变化——例如(a)轮廓点沿视网膜的序号发生改变，或者(b)~(d)这些点的数量发生变化。为了避免杂乱将视网膜从图上略去了，详细情况见课文

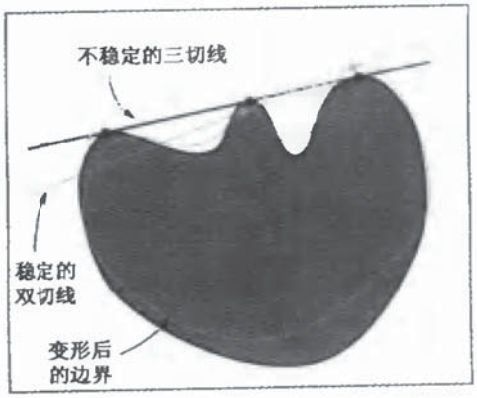


图 20.3 特殊的与一般的曲线：与双切线不一样，三切线在曲线微小变形时是不稳定的

本章我们将注意力限定在一般曲线与一般表面的范围内，它们的特征不会受小幅度变形的影响。一般性(genericity)是一种数学概念，它与一般位置这种直觉概念以及开放性、密度、贯穿性等拓扑概念有关。它的形式化定义是相当技术性的，在这里提它不合适。这里只需指出的是，尽管一般性假设把某些简单的几何图形(线，平面等)排除在外，但实际物体的边界是一般性的(现实世界中不可能画出一根真正的直线)。更值得引起注意的是，至少从本章所涉及的内容看，限定注意力到一般曲线与一般曲面，也限定了可能发生的视觉事件的数量与类型。轮廓结构的结构性变化确实可以由数学的分支——奇异性理论或突变论有规则性的预测，在扁平世界中可解释成，将一般曲线的视觉事件限制到与双切线、拐切线以及它们的相交等有关的情况。

对全方位圆形摄像机而言，物体的外观(aspect，也就是轮廓点的数量与顺序)仅取决于眼睛的位置(见图 20.2)。视觉事件关注于曲线与投影射线之间接触的何种类型，而不在乎射线

与视网膜交点的位置。于是对一个扁平世界摄像机的视点集合，可以用光心所在的平面来描述。这个平面可以划分成分格式结构，称为外观图(aspect graph)，在该图中产生视觉事件的射线以及它们之间的交将平面划分成一个个最大的分格，在每个分格内部，物体的外观没有变化(见图 20.4，左)。

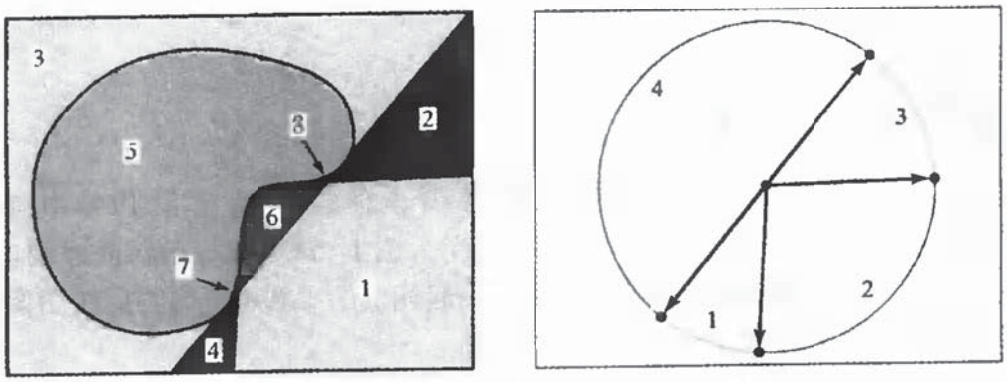


图 20.4　扁平世界中不透明物体的外观图。左图：一个由平滑封闭曲线包围的两维物体以及它的(全方位圆形)透视外观图的分格式隔间，注意用7与8标出的两个小区域。右图：相应的正交投影外观图。弧连接两个相邻的外观，但外观没有画出来。一个很容易的练习：将外观画出来。对一个透明物体该外观图会怎样变化

对远距离的观察者来说，可将目前考虑的全方位圆形摄像机用惯常用的正交摄像机代替，这是由于在这种情况下，视线方向决定了线性视网膜的方向，而图像线也总能选择到这样一种方向，使得所关注的物体完整地处在这条线的前面。在这种设置下，视点空间成为一个(常朝向的)投影向量的单位圆，而双切线与拐切线的方向将其分割成物体外观的有限集(见图 20.4，右)。

同样的原理可以用到三维的世界中：与扁平世界的情况相似，所选择的投影模型(全方位球面透视、平面透视或正交摄像机)与视点决定了物体的外观(在这种情况下，在图像轮廓的图形表达式中，用结点表示 T 形结点与歧点，弧表示结点之间的平滑轮廓片断)。所有的有效视角再一次被视觉事件边缘划分成最大分格式，在每个分格内物体的外观没有变化，这些分格用其典型外观代表并形成三维外观图的结点，而附加在视觉事件边缘上的弧将相邻的分格隔开。外观图是由 Koenderink 和 Van Doorn(1979)首先提出的，当时称为视觉势图(visual potential)。下一节将附加一些微分几何的内容，它是为理解外观图的构筑所必需的。接下来的章节讨论构筑刚体精确的与近似的外观图的算法，以及如何将后者用到仓式进料器的任务中。

20.1　视觉事件：微分几何的补充

从现在起我们假设正交投影(透视投影的情况将在 20.4 节简略地提一下)，并且用带转向的投影方向单位球为所有的视点建模。拐点、歧点与 T 结点是图像轮廓的稳定特性，它们在眼球的小幅度运动状态下一般仍能保持。作为一个例子，考虑在第 19 章展示过的轮廓拐点，它是遮挡轮廓与所关注曲面的一个抛物曲线，(一般以非零角度)相交点的投影。视点的小幅度改变使遮挡轮廓略有变形，但这两条曲线仍然在邻近的点相交，投影成一个轮廓拐点。

人们自然会问：怎样的(独特)眼球运动会使稳定的轮廓特征出现或消失？为了回答这个

问题,重新回顾一下高斯图,并引入一种渐近球面映射,这个过程说明,表面图像的边缘通过这些映射决定了轮廓上拐点、歧点的出现与消失。这为我们提供了局部视觉事件的特征(也就是说,边缘处与表面的微分几何有关的轮廓结构变化)。我们也要考虑视线与表面的多处接触。这使我们得到双切射线簇的概念,而它的边界特征使我们能理解 T 结点的生成与灭亡的原因,并引出相关的多重局部视觉事件(multilocal visual events)概念。局部与多重局部事件加在一起概括了结构轮廓变化的总体,从而决定了外观图。

20.1.1 高斯图的几何关系

高斯图为研究图像轮廓与它们的拐点提供了一个自然的机制。在第 19 章中,我们确实看到在正交投影条件下遮挡轮廓映射到单位球的一个大圆上,这个圆与抛物曲线的球面图像的相交得到了轮廓的拐点。因此,当摄像机的运动引起相应的大圆跨越抛物曲线图像时,轮廓得到(或失去)两个拐点(见图 20.5)。

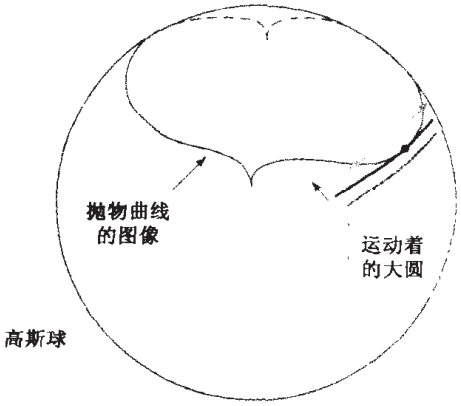


图 20.5 随着视点变化,与(正交投影)遮挡轮廓有关的高斯球大圆变成与抛物曲线的球面图像相切。此后,该圆与该曲线在两个邻近点相交,对应于两个轮廓拐点

更仔细的观察高斯图的几何关系会使我们更清楚地理解拐点对的产生过程。正如第 19 章指出的,表面在高斯球上的图像沿着它的抛物曲线图像折叠。一个在抛物曲线的一边是球的单折、而另一边是它的三重折的例子,显示在图 20.6 中。思考这种折叠产生的最简易方法是(在你脑子里)抓住一个减压气球的一部分橡皮,挤压它并将其折起来。正如该图所示,这个过程在一般情况下不仅造成球面图像的折叠,并且还含有两个歧点(在微分几何中贴切地把它的原像称为高斯歧点)。抛物曲线图像的歧点与拐点总是成对出现的(图中有两个拐点对与一个歧点对,但是也可能完全没有歧点或拐点)。该拐点将高斯图的折叠分裂成凸的与凹的两部分,它们的原像称为转接点(gutterpoints)(见图 20.6)。

当有关的大圆穿越抛物曲线的球面图像时,遮挡轮廓会出现的情况取决于穿越在哪里出现。正如图 20.6 所显示的有几种情况:当穿越沿着高斯图的凸折发生,在遮挡轮廓的球面图像上出现一个孤立点,并迅速演变成单位球上的小闭环(见图 20.6,右下部)。与此相反,如果穿越沿凹折发生,两个分开的闭环在合并之后,又以不同的连接方式分开(见图 20.6,右上部)。这种变化自然会以某种方式反映到图像轮廓上,这将在后续若干节中详细分析。

与遮挡轮廓有关的大圆也可能与抛物曲线图像在歧点处穿过。与发生在正规折叠点的交叉不同,这种交叉一般是相交的,并没有在大圆的朝向上形成相切。相交的拓扑关系并没有改变,但是在图像轮廓上会出现(或消失)两个拐点。最后一种情况是大圆在抛物曲线高斯图像

的一个拐点上穿过。在这种情况下，拓扑关系的改变就太复杂了，这里也就不提了。好在这种情况只可能在有限数量视点出现(因为一般曲面只有数量有限的转接点)。而与此相反，别的类型的折叠穿越却可能在具有一个参数的无限视点族上发生。这是因为与折叠的凸或凹部的相切穿越，可以沿着单位球延长曲线弧上任何位置发生，而与歧点的相交只发生在孤立点，但是可沿单位圆的任意朝向。我们在下一节讨论有关奇异视点族。

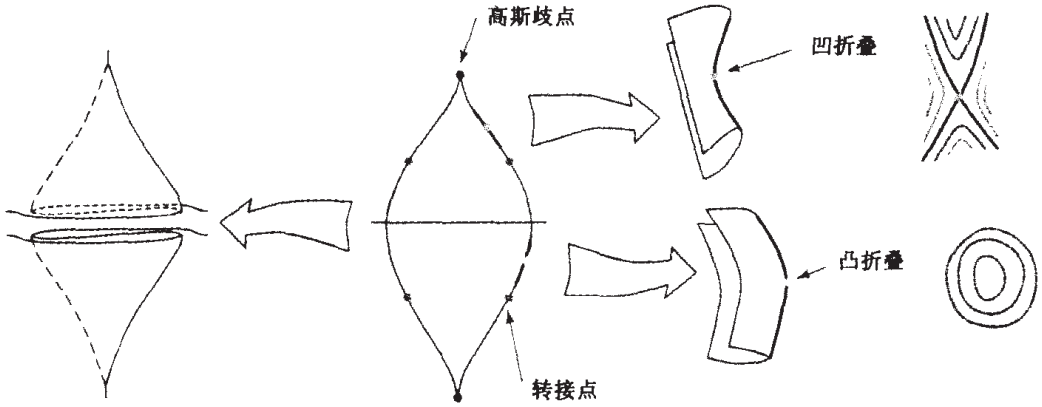


图 20.6　高斯图的折叠与歧点。转接点是抛物曲线球面图像拐点的原像。为了使折叠的结构表示更清楚，将在空间的表面折叠画在图的左部与右部。表面高斯图与大圆之间的交的拓扑关系随着圆穿越折叠时的变化显示在图的最右部

20.1.2　渐近曲线

在第 19 章已经了解到一般双曲点允许有两种不同的渐近切线存在。更为一般的情况是，双曲面片上的所有渐近切线集可以明确地划分成两族，而每一族接纳一组积分曲线的光滑场，这称为渐近曲线。沿用 Koenderink(1990)的方法给每个族一种颜色，并且讨论相关的红与蓝渐近曲线。这些曲线只涵盖表面的双曲部分，因而在它抛物边界附近必然是奇异的。一个红渐近曲线与蓝渐近曲线在一般抛物点合并时确实形成一个歧点，并且与抛物曲线以非零角度相交[见图 20.7(a)][①]。

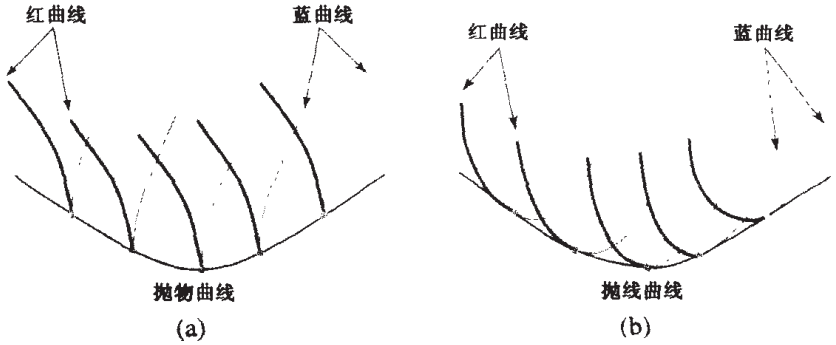


图 20.7　渐近曲线与抛物曲线的接触：(a)在表面上；(b)在高斯球上

① 这种情况在高斯歧点时是不同的，此时渐近曲线与抛物曲线相切。这种不寻常的情况也发生在非一般物体的平面抛物曲线上(例如，一个圆环侧面的顶部与底部的两个圆形抛物曲线或更一般的是旋转体的抛物线，它们与沿其轴截面高度的局部极值处)。

下面考察一下高斯图中渐近曲线的性能。回顾第 19 章曾提到的,渐近方向是自共轭的,这意味着沿渐近曲线表面法线的导数与曲线的切线是正交的,渐近曲线和它的球面图像有垂直的切线。另一方面,切平面的所有方向在一个抛物点是与渐近方向共轭的,所以过一抛物点的任何表面曲线的高斯图像与相应的渐近方向垂直。尤其要提到的是,一条抛物曲线的高斯图像是与其相交的渐近曲线图像的包络[它与这些曲线处处相切,见图 20.7(b)]。

现在我们能够叙述能引起一对交点出现(或消失)的视点的特性了。由于与遮挡轮廓相关的大圆与高斯球上抛物曲线的图像接触时变成与其相切,沿大圆法线的视线方向是沿相应的抛物曲线的渐近方向的。当大圆跨越一个高斯歧点图像时,或者等价地说当一个视线穿过该点切平面时,自然会有一对交点出现。正如早先指出的,此时图像轮廓的拓扑关系并没有变化,仅仅引起(或失去)一个波动(也就是说,在它的某个凸部有一小的凹坑,或在其某个凹部有一凸起)。下一节将讨论在其他类型的奇异点上轮廓结构是如何变化的。

20.1.3 渐近球面映射

我们知道高斯图将每个表面点与它的相应法线顶端穿过单位球的地方联系起来。我们现在来定义渐近球面映射,它将每个(双曲)点与相应的渐近方向联系起来。在进一步讨论之前先说些要点,首先对每个渐近曲线族的确有一个渐近球面图像,这两幅图像在球上可以重叠,也可不重叠。其次,椭圆点显然没有渐近球面图像,并且单位球很可能没有被双曲点图像完全覆盖。但是它也可能被充分覆盖的,或至少在局部上覆盖,它可能被同一族渐近方向的成员数次覆盖。

由于视线沿着一个渐近方向时图像轮廓会冒尖,因此当视线穿过渐近球面映射的一个折叠时,就会出现(或消失了)一个歧点对(注意这种情况与轮廓交点与高斯图折叠之间十分相似)。正如预料的,一根渐近曲线的渐近球面图像在折叠边界处是奇异的。这同样有两种可能性(图像可以以相切的方式连结边界或形成歧点),并且它们在两种类型的折叠点出现:一种与沿着抛物曲线的渐近方向有关(因为在它们椭圆边没有渐近方向),另一种与拐结点(flecnodal points)处渐近方向有关。这些点是渐近曲线投影到它们切平面的拐点[见图 20.8(a)],它们形成贯穿相应渐近曲线的曲线。与前面提到的一样,它们可以来自于两种颜色,这取决于哪种渐近线族具有拐点。渐近曲线的渐近球面图像在拐结点处出现歧点[见图 20.8(b)]。还要提到的是拐结点曲线与抛物曲线在高斯歧点处相切。

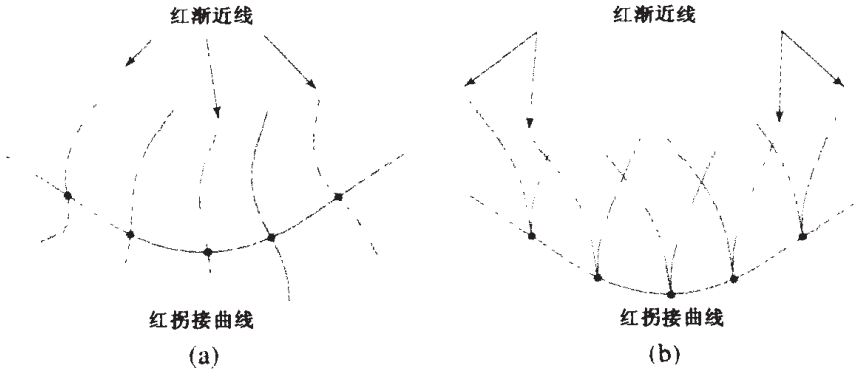


图 20.8 渐近曲线与拐结点曲线的接触:(a)在表面;(b)在渐近球面图像

显然,当视线穿过渐近球面图像的抛物线或拐接边界时轮廓的结构发生变化,这样的变化称为视觉事件,有关的边界称为视觉事件曲线。在更详细地考察各种视觉事件前,我们先提一提一

种考虑有关边界的不同但等价的方法。如果沿着抛物或拐结点表面的每一点画出奇异渐近切线,就得到由切线扫出的可展曲面(ruled surface)。当视线跨过任何一个可展曲面时,就发生视觉事件,它与球在无穷远处的交点就是视觉事件曲线,此时这个球要规范成单位球。用可展曲面的术语考虑轮廓的演变的好处是,可将视觉事件推广到透视投影(每当光心穿过它们时,视图改变了),以及有可能获得对奇异视点与表面形状之间关系清晰地可视化。

20.1.4 局部视觉事件

现在我们有条件理解在视觉事件边界上轮廓结构是如何变化的了。有三种局部视觉事件,它们是完全由局部微分表面几何的特征所决定的:唇(lip)、喙对喙(beak-to-beak)以及燕尾状(swallowtail)。它们五彩缤纷的名字是从 Thom(1972)的基本裂变分类学(catalogue of elementary catastrophes)继承来的,也与相关事件附近的轮廓形状有密切关系。

首先考察唇事件,它发生在视线穿过凸抛物点的渐近球面图像时,或等价地由有关渐近切面定义的可展曲面时(图 20.9 顶部)。前面已展示过与遮挡轮廓有关的大圆与表面的高斯图像相交的事件中会得到一个环(图 20.6 右下部),并在轮廓上创造了两个拐点与两个歧点。说得更精确些,在事件发生之前并没有图像轮廓,接着在奇异点处出现一个孤立的轮廓点,随着出现了一个封闭轮廓环,由两段轮廓在两个歧点相遇组成(图 20.9 下部),其中一支由椭圆与双曲点的投影形成,带有两个拐点,而另一支仅由双曲点的投影形成。对不透明物体,其中一支会被物体遮挡住。

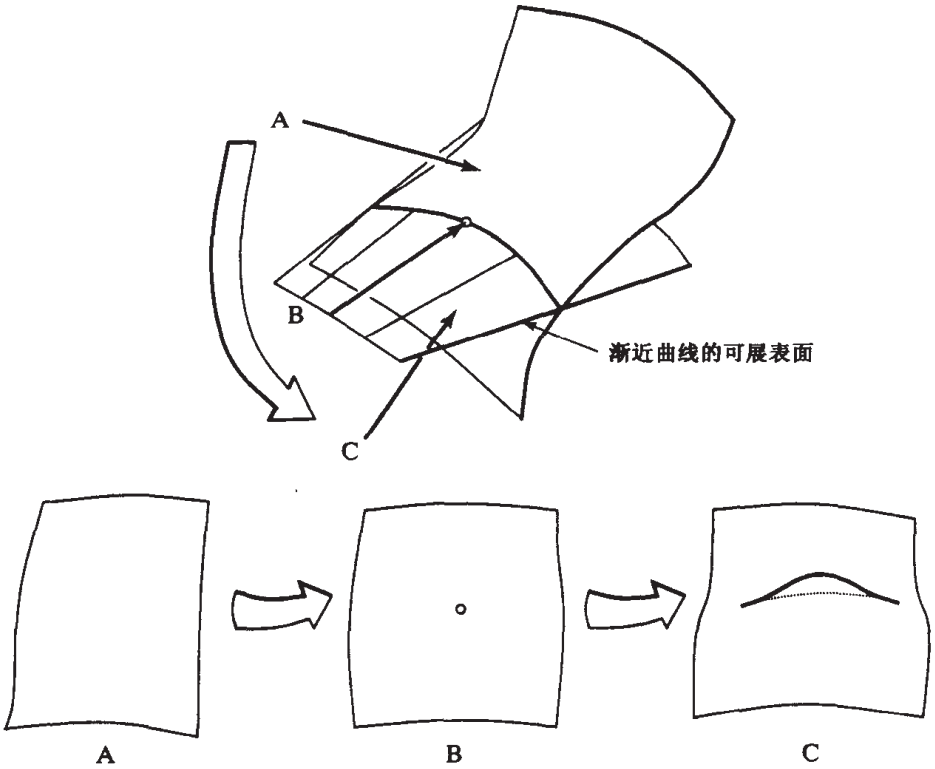


图 20.9 唇事件。这个名字与右图上的轮廓的形状有关。其中对不透明物体来说,轮廓的虚线部分由于遮挡而看不见,后面的几幅图也是如此。在这个例子中两个拐点在轮廓的可见部分,被遮挡部分全是双曲的,但是沿着视线的相反方向看情况就会反过来

喙对喙事件发生在视线穿过凹抛物点的渐近球面图像，或等价地说是相关的渐近切线定义的可展曲面时(图 20.10 上部)。正如前面所指出的，在该事件中与遮挡轮廓有关的大圆与表面的高斯图像相交的拓扑发生了变化，有两个环在合并到一起后，又发生分裂并改变了连接性(图 20.6 右上部)。在图像中轮廓的两个不同部分(每个有一个歧点与一个拐点)在图像中的一点相遇。在事件发生之前每一支都由相关的歧点划分成一个纯双曲部分与一个椭圆与双曲混合的弧，其中一个总是被遮挡着的。在事件发生之后，两个轮廓歧点与两个拐点随之消失，而轮廓分裂成两段光滑的弧并改变了连接性，其中之一是纯粹的椭圆，而另一个是纯的双曲。对不透明物体来说，其中之一总是被遮挡了的(图 20.10 下部)。当然，像所有其他的事件一样，反向的演化过程也是可能的。

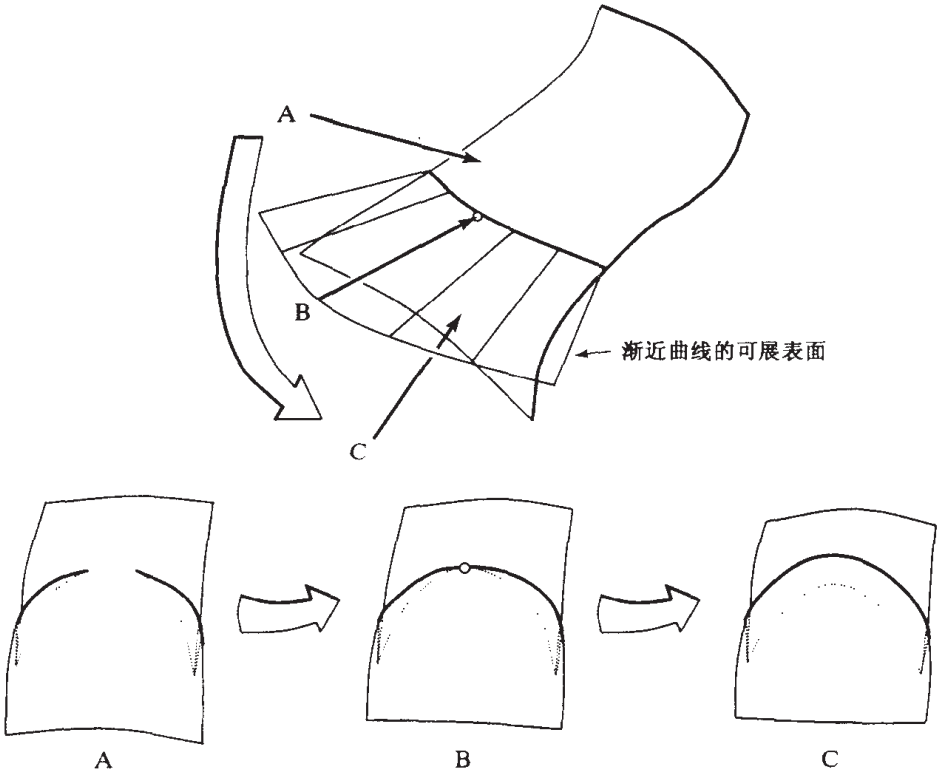


图 20.10 喙对喙事件(名字与左图轮廓的形状有关)

最后一种情况是燕尾形事件，它发生在眼睛穿过沿着相同颜色的拐结点曲线与渐近切线展出的表面时。我们知道在这个事件中两个歧点会出现(或消失)。正如图 20.11(a)~(b)所示，表面与它的切平面在一个拐结点处相交并组成两个曲线，其中之一有一个拐点。相应的渐近切线自然也是与有一拐点的渐近曲线族有关。与一般被所观察物体遮挡的渐近射线不同[见图 20.11(c)]，它是擦过物体表面的(Koenderink，1990)，这导致在奇异点处的图像轮廓产生一个锐利的 V 形。在事件变化之前轮廓是平滑的，但事件之后它得到两个歧点和一个 T 结点(图 20.11 下部)。该事件中的所有表面点是双曲的。对于不透明物体，轮廓中的一支在 T 接合点终结，而另一个在一歧点终止。

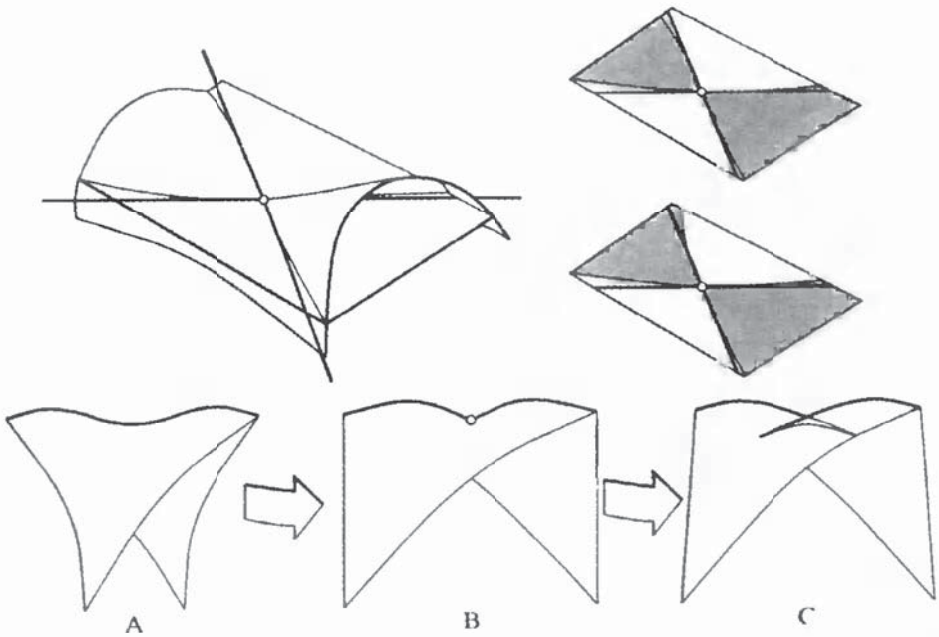


图 20.11 燕尾形事件顶部:在一拐结点附近的表面形状;(a)在这种点附近相关的物体与它的切平面相交的比较;(b)以及一个一般的双曲点;(c)底部:该事件

20.1.5 双切射线族

要记住第 19 章中提到歧点与拐点并不是惟一的稳定特征,T 结点也会在广阔的视点集中出现,它们出现在两个不同的遮挡轮廓投影到同一个图像位置时。相应的表面法线必须与连接该两点的双切视线正交,但它们之间(一般)不平行。对眼睛的微小运动来说,T 结点是稳定的,这一点在直觉上是很显然的。考虑一凸点 P 与它的切平面相交的情况(见图 20.12 左)。这个平面(一般来说)与表面会相交出一封闭(也可能是空的)曲线,并且从点 P 发出的射线会在偶数数目的点(图上的点 P' 与点 P'')与该封闭曲线相切。每个相切都会得到一个双切射线以及一个与此相关的 T 结点。眼睛的微小运动使相交的曲线有些变形,但是(一般)并不改变切点的数目。因此 T 结合点确实是稳定的。

在由所有直线所组成的四维空间中双切射线形成一个二维的双切射线簇[①]。因为在投影过程中双切线映射到 T 结点,显然这些轮廓特征在该簇的边界上会产生或消失。由于一个 T 结点在一个燕尾形特征演变过程中出现或消失,沿拐结点曲线的奇异渐开切线形成这些边界中的一个是显然的。其余的边界是由什么构成的还不清楚,这是下一节的话题。

20.1.6 多重局部视觉事件

当视线穿过双切射线簇的边界时,一对 T 结合点会出现或消失,而相应的轮廓结构变化称为多重局部视觉事件。除了在前一节提到的与穿过拐结点曲线相关的奇异性之外,在这一节将指出存在三种类型的多重局部事件——即切向交叉(tangent crossing)、歧点交叉(cusp crossing)与三重点(triple point)。

① 簇是一种拓扑概念,其将欧氏空间中定义的表面推广到更抽象的表示;此处忽略其形式化定义。双切射线簇是二维的这一点在直觉上是显然的,因为正处在观察中的这个(二维)表面的每个点所具有的双切射线数量是有限的。

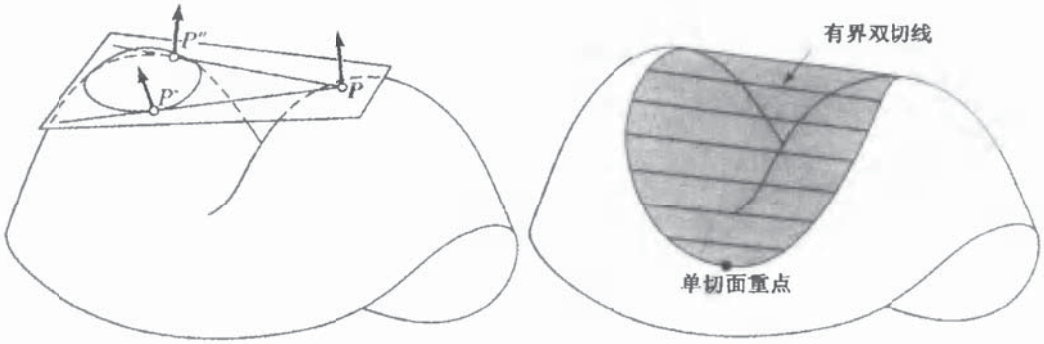


图 20.12 双切射线。左图:表面上一点 P 的切平面与表面沿一封闭曲线相交,有两条双切射线 pp' 与 pp'' 沿这曲线擦过该表面。右图:由一条直线扫出的有界双切可展表面,该线是与表面双切的一个平面两接触点的连线。在这里该可展表面与所观察表面接触处的两根曲线在一个单切面重点(unode)相切地合并在一起。这是一种类型的高斯歧点

首先观察切向交叉事件。双切射线簇一个明显的边界是由有界双切形成的(图 20.12,右),这发生在表面某点的切平面与表面其余部分相交产生的曲线缩成一个点,以及该平面与表面形成双重相切的时刻。有界双切线扫出一个可展表面,称为有界双切可展曲面。一个切向交叉出现在视线穿过这个表面(图 20.13,顶部)时,事件发生时两个原本分开的轮廓在形成两个 T 结点之前互相相切(图 20.13,底部)。对于非透明体来说,或是原先被遮挡的轮廓部分在事件演变后变成可见,或是(正如图中所示)另一个轮廓弧因遮挡而消失。

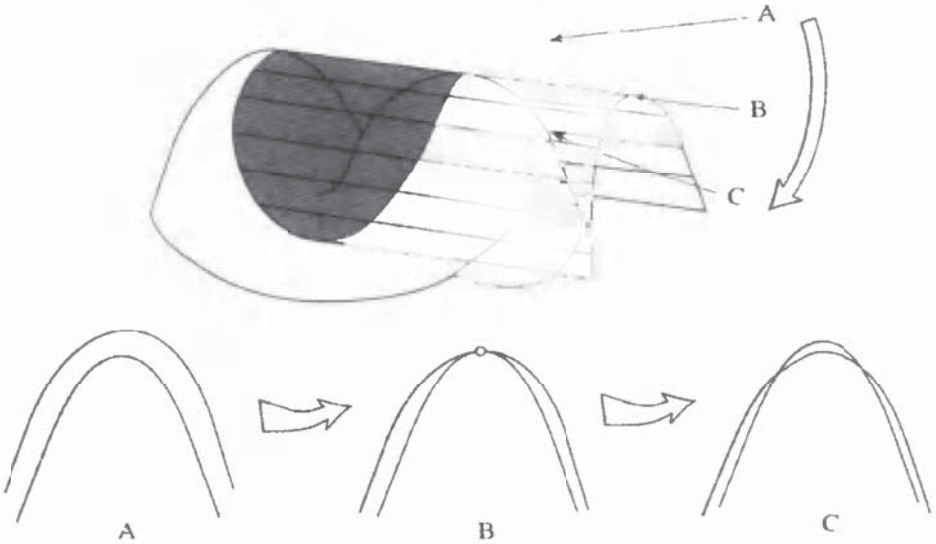


图 20.13 相切交叉事件。当视线跨越 B 中有界双切可展曲面时,遮挡轮廓在空间不同的部分之间的遮挡关系发生改变

双切射线簇还引发其他两种与双切线有关的边界,它们沿着一组曲线接触表面并扫出可展表面,这两种分别是渐近双切(asymptotic bitangents),这发生在沿着它们其中一个端点的渐近方向与表面相交[见图 20.14(a)];三重切线(tritangents)则发生在表面的三个不同点擦过[见图 20.14(b)]。当视线穿过其中一个相关的可展表面时,相应的事件就发生了,随之而来的是一对 T 结点的出现或消失:当一个光滑图像轮廓段在另一轮廓部分的歧点(或非透明体的端点)穿过时歧点交叉(cusp crossing)就发生了[见图 20.14(c)]。在此过程中两个 T 结点出现

(或消失),对非透明体来说只有其中一个能看见。当三个互相分开的轮廓段在一瞬间以非零角度聚在一起时,一个三重点事件(triple point)就形成了[见图 20.14(d)]。对透明体来说,三个 T 结点在重新分开之前叠合在奇异点。对非透明物体来说,一支轮廓与两个 T 结点消失(或出现),而另一个 T 结点出现(或消失)。

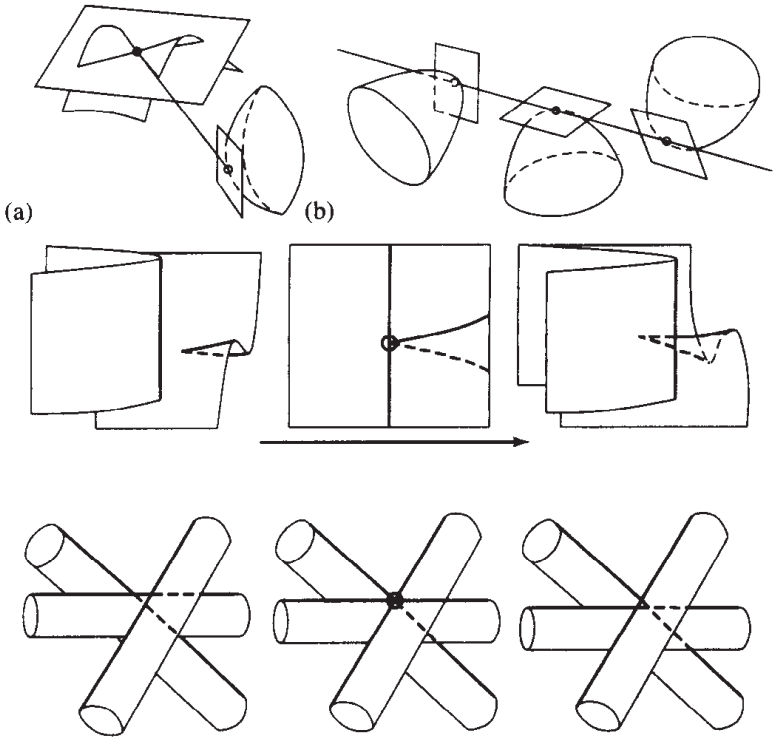


图 20.14　多重局部事件:(a)一个渐近双切射线;(b)一个三重相切射线;(c)一个歧点交叉;(d)一个三重点

20.2　计算外观图

在某种刚体边界的模型(以多面体、体密度的零集或任何你喜欢的形式)给定的条件下,人们要问的下一个问题是如何真正构筑起相应的外观图。在这一节中我们假定一个固定的欧氏坐标系统已经选定,因此$\mathbb{E}^3$中的点可以用$\mathbb{R}^3$ 中的坐标向量区分。使用给定模型给出的表面参数形式的导数(至多到三阶)项,利用 1,2 或 3 个表面点来重写抛物点、有界双切射线等的几何定义是不难的(Petitjean 等,1992)。在上述每一种情况中,早先定义过的与可展表面有关的表面曲线可以通过有 $n+1$ 个未知数的 n 个方程组在$\mathbb{R}^{n+1}$中表示:

$$\begin{cases} P_1(x_0, x_1, \cdots, x_n) = 0 \\ \cdots \\ P_n(x_0, x_1, \cdots, x_n) = 0 \end{cases} \tag{20.1}$$

其中,$1 \leqslant n \leqslant 8$ 取决于事件的类型,并与表面是否定义成参数形式或隐含式有关。例如在表面被定义为某些密度函数 $F(x, y, z) = 0$ 的零集的情况下,局部事件 $n = 2$,而对涉及三个分开的表面点的三重点 $n = 8$ (见练习)。

在给定这些方程式条件下,对一物体构筑外观图的一般方法涉及以下 4 个步骤:(a)跟踪透明体的视觉事件曲线,(b)构造由这些曲线定义的视点球的区域,(c)消除被遮挡的事件并且

合并与之关联的区域,(d)构筑相应的外观。透明体的外观图可用相同的过程来构造,但不包含(c)这一步。

这里我们对一个用代数表面(即体多项式密度的零集)包围的刚体构筑外观图的问题进行讨论:

$$S=\left\{(x,y,z)\in\mathbb{R}^3 \mid \sum_{i+j+k\le d} a_{ijk}x^i y^j z^k=0\right\}$$

代数表面的例子包含平面和二次曲面(也就是椭圆、双曲面与抛物面)以及高阶多项式密度函数的零集。最重要的是式(20.1)中定义的事件的约束条件,在这种条件下是所关注的未知数的多项式。这对执行前面描绘的一般算法的各个步骤是十分关键的,因为有关显式描述多变量多项式方程解的数值计算工具与符号计算工具都具备。

20.2.1 第一步:跟踪视觉事件

一个视觉事件与两个曲线关联:第一个称为 Γ,或取自于物体表面或(在多重局部事件情况下)取自于一个更高维的空间。第二个曲线称为 Δ,取自于相应视点集的视点球。曲线 Γ 由式(20.1)的 n 个方程式在$\mathbb{R}^{n+1}$空间定义。

这一节讨论跟踪曲线 Γ 的问题(也就是辨别它的平滑弧与奇异点,然后建立各种曲线成分的连接关系)。算法 20.1(用图 20.15 说明)给出这个问题的一个简单解。这个算法将定义 Γ 用的有 $n+1$ 个未知数的 n 个方程式为输入,输出一张图 G,它的结点是 Γ 在 x_0 方向的极值点(包括奇异点),而弧是连接它们的平滑曲线段。

算法 20.1 曲线跟踪算法

(1.1) 计算某个方向(例如 x_0)Γ 的所有极点(包括奇异点),这些点构成图 G 的结点。

(1.2) 在所有极值点计算 Γ 与 x_0 轴正交的超平面的所有交点。

(1.3) 对 x_0 轴由这些超平面定界的每一段,求 Γ 与通过该段中点的超平面之间的交点,以获得每个实际曲线段的一个采样。

(1.4) 从第(1.3)步中找到的采样点出发,用数值方法推进到第(1.2)步中找到的交点,通过泰勒展开式预测新的点并用牛顿迭代修正它们的方法。

(1.5) 将 Γ 上在非极值点的交点上相遇的平滑弧段合并,添加一个弧到 G,用来表示将一对极点(或奇异点)连接起来的曲线段。

第(1.1)步要求计算在 x_0 方向上的 Γ 的极点,这些点由含 $n+1$ 个多项式方程来描述,多项式方程组通过式(20.1)中加入了一个方程 $\mathrm{Det}(\mathcal{J})=0$ 得到,此处$\mathcal{J}$表示雅可比矩阵($\partial P_i/\partial x_j$),它们可以使用同伦延拓(homotopy continuation)找到(Morgan, 1987),它是一个计算多项式方程式方阵系统所有解的数值方法。第(1.2)与(1.3)步要求计算曲线与超平面的交。这些点也是多项式方程的解,用同伦延拓求得。对曲线的跟踪实际上在(1.4)步中进行,通过 $P_i s$ 的泰勒级数展开式使用牛顿预测/修正方法得到,其中涉及对矩阵$\mathcal{J}$求逆,这个矩阵在没有极值的间隔中确保是非奇异的。

一旦 G 构造完成,一个关于曲线 Δ 的类似的描述图 D 是容易得到的,这可通过将 Γ 的点映射到相应的渐近或双切方向上实现。

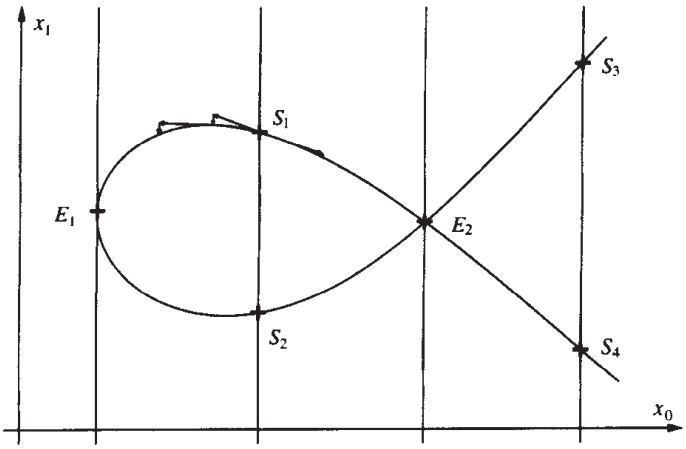


图 20.15 在二维实数域$\mathbb{R}^2$中跟踪曲线，此时超平面是直线。该曲线有两个极值点 E_1，E_2 以及 4 段一般曲线，它们的采样点分别为 S_1 到 S_4，E_2 是奇异的

20.2.2 第二步：构造区域

现在我们假定已经构造得到与所有视觉事件曲线 $\Delta_i(i=1,\cdots,p)$ 有关的图 D_i。为了构造由曲线 Δ_i 在视点球上描述的外观图区域，我们将这些曲线使用球面坐标映射到平面上，并将曲线跟踪算法改成用单元分解过程进行细化（见算法 20.2 与图 20.16），它的输出是区域的描述、它们的边界曲线以及它们的相邻关系。注意这种细化只可能对平面曲线有效。实际上，这个算法是以多边形曲线作为输入的，这通过将与图 D_i 相关的 Δ_i 曲线的离散表达式映射到平面上获得，而相应平面曲线的极值点与交点很容易从这些多边形中得到。如果需要的话，在算法中找到的视觉事件曲线与单元是很容易映射回视点球的。

算法 20.2 单元分解法。这个算法的输入是一组平面曲线，输出的是由这些曲线包围的区域以及它们的相邻关系（两个区域如共享共同的边界，也就是竖直线段或曲线段，则它们相邻）。

(2.1) 计算 x_0 方向曲线上的所有极值点。

(2.2) 计算这些曲线之间的所有交点。

(2.3) 在曲线的极值点与交点计算它与竖直线的所有交点，竖直线正交于 x_0 轴。

(2.4) 对由这些线分界的 x_0 轴的每一段做以下事情：

- (2.4.1) 求曲线与穿过该间隔段中点的竖直线的交点，以获得每个曲线段的一个采样点。
- (2.4.2) 对采样点按 x_1 值递增的序排序。
- (2.4.3) 从这些采样推进到(2.3)步找到的交点。
- (2.4.4) 在 x_0 的一个间隔的两个相连接的段以及与它们的极值连接的竖直包围一个区域。

(2.5) 对每个区域将在(2.4.1)中找到的相继采样点的中点作为区域的采样点。

(2.6) 将沿竖直线段相邻的区域合并成最大化的区域。

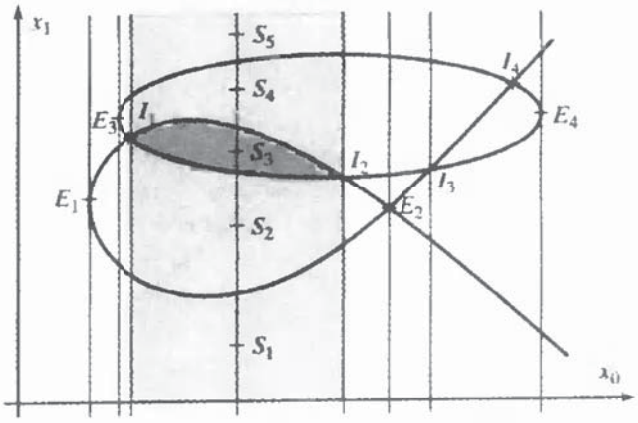


图 20.16 一个单元分解的例子，其中有两条曲线以及它们的极值点 E_i 与它们的交点 I_j。由 I_1 与 I_2 为界阴影矩形被分成5个区域，以及它们相应的采样点 S_1 到 S_5。以 S_3 代表的区域用更深的影调表示

20.2.3 算法中的其他步骤

构造外观图算法中的第(3)步与第(4)步涉及消除被遮挡的视觉事件与构筑每个区域的采样状态。这两步在概念上是简单的。要强调的是透明体的所有视觉事件是在算法的第(1)步找到的，它们的相交在第(2)步找到。对一个非透明体，某些事件是被遮挡住的，因此需要去除掉。由于一个视觉事件曲线的可见性仅仅在它的奇异点以及与其他事件相交处才改变，被遮挡的事件可以通过对在第(3)步找到的每支采样点进行射线跟踪来去除。与被遮挡事件相邻接的区域的合并也就不难了。算法的最后一步是对每个区域的一个视图确定其轮廓结构。在给定一个区域的采样视点条件下，将在 20.2.1 节叙述的曲线跟踪算法应用到图像轮廓上，以求构造一个图像结构图(image structure graph)，图的结点是轮廓的极点与奇异点(歧点和 T 结点)，而弧是连接它们的平滑段。与以前一样，射线跟踪再一次用来确定每段的采样点是否被遮挡。

20.2.4 一个例子

图 20.17(顶部)表示一个压扁形刚体的两幅素描，它的表面用多项式密度函数来定义。在图 20.17(顶部左端)上大体平行的两根曲线是这个压扁形的抛物曲线，它们将它的表面分裂成由一个鞍形区域隔开的两个凸起团状物。在图上展示的自相交曲线是拐结点曲线。图 20.17(顶部右端)展示的是与该压扁形有关的有界双切可展曲面，它的母线是压扁表面上能符合生成双切平面条件的点对的连线。抛物点曲线、拐结点曲线以及有界双切可展曲面已使用这一节讨论的曲线跟踪算法得到。在这个例子中没有渐近双切或三重切。图 20.17(底部)展示了不透明压扁体的正交投影外观图，是使用前述单元分解算法进行计算得到的。

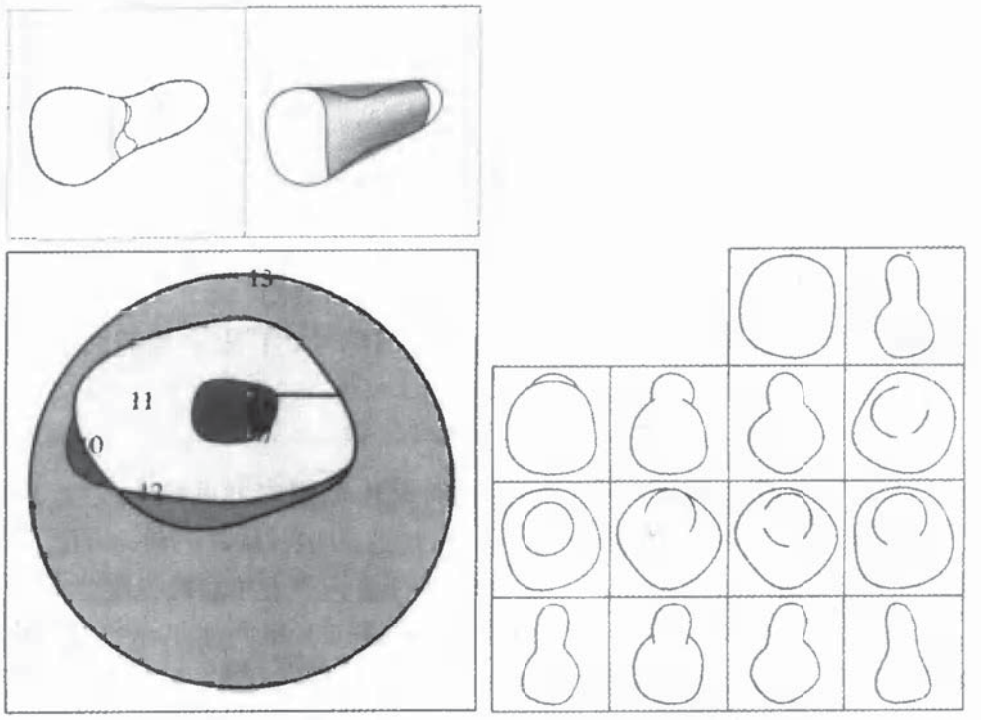


图 20.17 顶部:一个压扁形刚体以及它相应的抛物线与拐接线(左)、有界双切可展曲面(右)。底部:它的(非透明体的)正交投影外观图,(左)是视点球单元,(右)是相应的状态。注意在所示半球中实际上只有14个单元中的9个是可见的,而其中的某些(如区域7)是十分小的

20.3 外观图与物体定位

外观图在直觉上是吸引人的,在数学上也是优雅的。然而在实践中可以这样说,像前述章节所讨论的精确的外观图,仍然没有所期望的那样在物体识别等视觉任务中充分发挥作用。其中一部分原因是,诸如端点与T结点等轮廓特征要可靠地从实际图像中提取出来是非常困难的,而另一方面是因为甚至相对简单的物体都可能有非常复杂的外观图。近似的外观图却不同,它已被成功地应用到一些实际问题中,如从零件定位到物体识别。这一节我们着重考虑由 Ikeuchi 和 Kanade(1987*b*,1988)研究的仓式进料的问题。在这种场景中,一定数量的相同物体(一般是自动化工厂中的机器零件,但本节的实验中用的是塑料玩具)以随机的方位堆放在一起,等待机器人捡出。因此需要通过传感器获得的数据,确定它们的三维位置与朝向(或姿态)。我们假设这些零件的多面体模型已经具备,而传感器数据是从俯视图像使用(第5章介绍的)光度学体视技术构造的针状图。

我们把注意力集中在料箱顶部的零件上,并且假设用远处的摄像机观察(正交投影),以及该零件没有被料箱内其他物体遮住任何部分。这种假设在仓式进料器中是合适的,因为通常零件的尺寸与观察它的俯视摄像机的高度相比是相当小的。在这样的假设之下,处于顶部的零件表面上任何点的可见性可以仅由视点方向确定(这里定义成相对于附在物体上的局部坐标框架的竖直方向)。相应的外观定义为可见表面的清单,以表面面积大小按降序排列。对一个多面体零件使用简单的被遮挡表面消除技术(如 z 缓存法),来确定给定物体朝向时的外观是容易做到的。对于该例中塑料玩具这一类物体来说也没有什么特别的问题,因为它们是用

分块平滑的表面包围的，并可用多面体的网格近似。可以认为如果用来近似(曲)表面的任何平面小面块可见，那么这个(曲)表面也可见。在这种设置下物体的近似外观图可以用算法20.3方便地构造起来。

算法 20.3 一个近似外观图构造算法

1. 将观察方向的单位球进行棋盘化布局。
2. 在该棋盘化布局的每个单元的中心计算相关的外观图。
3. 将相邻的外观合并成等价类，用二进制串给出标号，“1”表示相应面可见，“零”则表示不可见。

这个算法假设在离散视点球的每个单元包含的视角范围内可见性不发生变化，因而只构造实际外观图的一个近似表示。图20.18展示了所构造的塑料玩具的外观图。此刻还要说明关于数据获取过程所受到的一些限制。尤其要提到的是一个表面能被光度学体视找到的条件是它的法线与视线之间的夹角足够小。按这个准则在图20.18所示的例子中，外观7中所有可以看到的面中没有一个能被检测出来，因而只能获得一个全零标号。尽管外观6(图20.18中标成4)中玩具的正“面”应该能检测出来，但它在实际上被对应成它表面上一个没有建模的凹陷，因此对这个外观也没有代表性视图生成。对其余的外观的每一个都计算了一个代表性视图，这是通过在相应视角范围内找到物体图形面积最大的方向，并将它的最大惯量轴与图像水平方向对齐得到的。

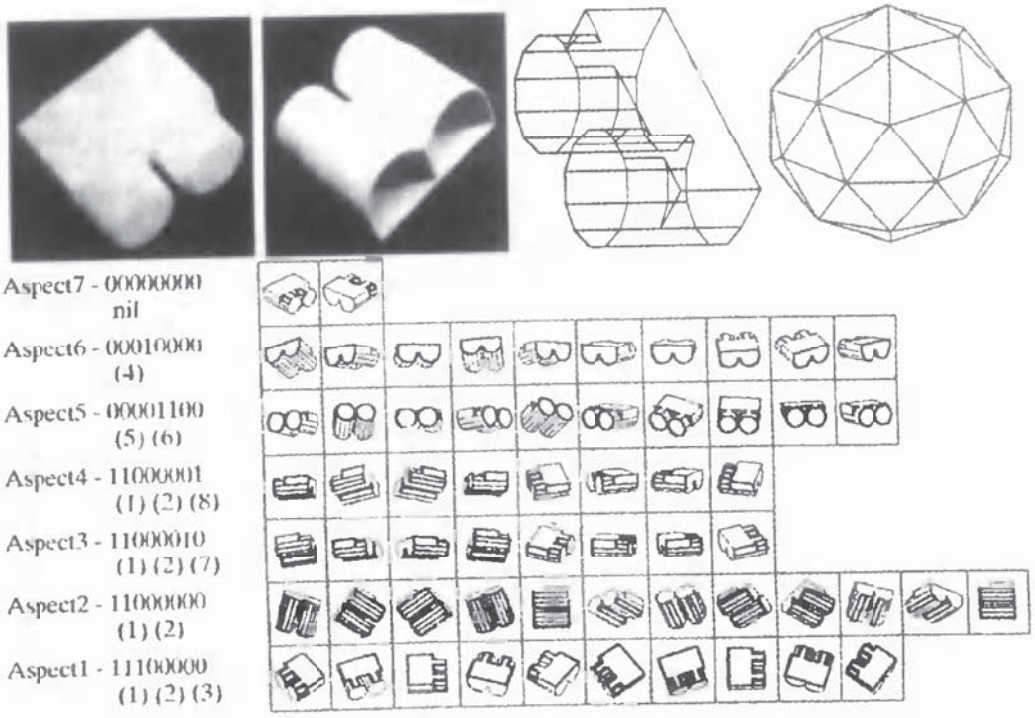


图20.18 近似的外观图。顶部，从左到右：一塑料玩具的两幅图，它的多面体模型(注意每个圆柱表面用若干平面近似，但在构造外观图时将它们考虑成单个物体面)以及一个用60个三角面拼成的半正规的单位球

为了促进定位过程，将外观组织成一个决策树是方便的，该树称为解释树（interpretation tree）。该树的每个结点包含相应视点范围、相应的外观表以及一个物体面。这棵树是以叠代方式构造的：在每个阶段，与一个结点 N 有关的表面 F 用来将相应的视点分裂成 F 能看见的一个子集 V'，与 F 看不到的子集 V''。结点 N 的左路分支赋以 V' 的视角范围，相应的外观子集以及在 V' 中某个视点可见的下一个最大面。它的右路分支以同样方式对 V'' 与相应的外观子集构造。如果在划分过程中遇到有两个面具有相同的面积，则将它们合起来使用，V' 成为它们之一可以看到的视角子集，V'' 则是它们都不可见的视角子集。这棵树从根结点开始构造，它包含物体的最大面、整个视点球以及所有可能的外观。

用以上方式构造的解释树还可以通过附加一组分类与姿态确定规则而进一步完善。这些规则是根据每块面的几何模型中几何及拓扑特征用人工方法构造的，包括它的惯性轴的方向及相应的矩、按它的总体形状分成平面、柱体、椭圆或双曲等类，它的扩展高斯图像（EGI 也就是在面内表面法线的直方图，在一个离散高斯球上计算）、从某个输入图像用边缘检测器检测的轮廓信息以及面与它的邻域之间的相邻关系。例如，一个典型的规则可以使用一个观察到的矩与模型中能得到的矩之间进行比较，以确定在解释任务中沿哪个分支继续下去。确定姿态的规则使用与已匹配的面有关的几何信息来计算物体姿态。例如，观察方向可以通过将一个面的 EGI 的质量预测中心与所观察到的对齐来计算，而物体在垂直于视线平面内的朝向可以从该面惯性轴的预测值与观察值来恢复。

图 20.19 显示了一个定位实验的结果。这个例子中有三种传感器数据源：最主要的是针状图，它是在同一视角不同光照条件下拍摄的几幅图像中提取出来的。一个粗略的深度图是通过将两个摄像机得到的针状图匹配提取出来的（双光度学体视，dual photometric stereo），而一个轮廓图是通过对其中一幅图像用边缘检测器找到的。将这三个图在同一个坐标系统中对准，而顶部区域（图 20.19 左）是从深度图确定的，并送到解释树中去。定位过程伴随在树中搜索进行、对每个结点中有关的测试用来选择正确的支路路径，并且成功地辨别所观察到的外观（图 20.19 中部）、相应的视角方位、与最终确定围绕视线应有的旋转量，从而能使预测与观察到的外观对齐（图 20.19 右）。

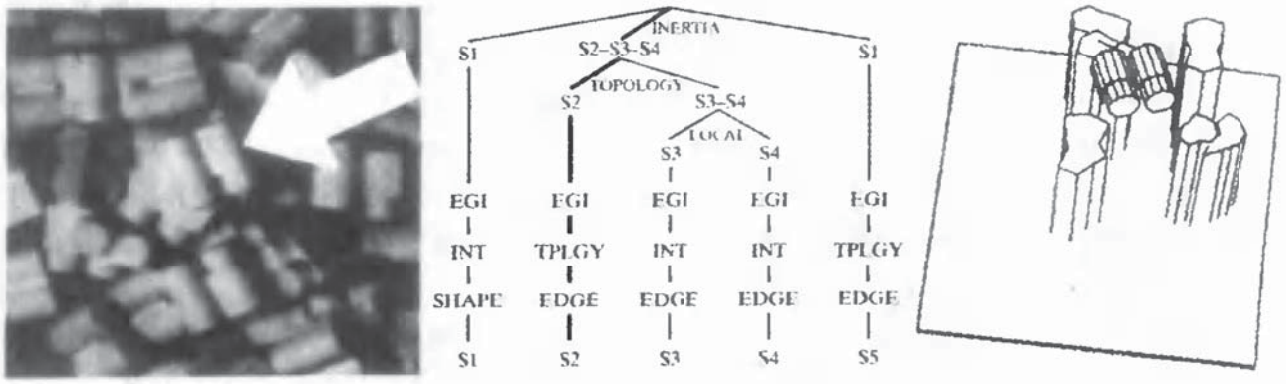


图 20.19　左：一个料箱采样图像；顶部的玩具用箭头指出。中：应用解释树到达相应图像区域所沿的路径（用粗线表示）右：用该定位算法找到的玩具，并以所估计的方位显示

20.4 注释

本章的内容主要是依据 Koenderink 与 Van Doorn 的工作，包括引入外观图概念（尽管所用名字不同）的开创性文章（Koenderink 与 Van Doorn，1976*b*，1979）。以非常容易接受的方式表述的关于形状表示方法的基础性几何知识的文章（Koenderink，1986，也可见 Platonova，1981；Kergosien，1981），以及 Koenderink（1990）著的一定程度上要求更高（但对有耐心的学生收获更多）的书。也可以阅读 Callahan 与 Weiss（1985）关于外观图所做杰出的非正式引导。一般性、奇异性与实变理论在许多书上都有讨论，其中包括 Whitney（1955），Arnol' d（1984）以及 Demazure（2000）。也可阅读 Koenderink（1990）关于为什么椅子会摆动的讨论，以及 Thom（1972）对这个问题的深入讨论。在 20.2 节中呈现的算法是出自于 Petitjean 等（1992），而这一节的题材主要取自相应的文章，它在很大程度上倚赖于 Morgan（1987）创导的数值同伦延拓方法，旨在找到多变量多项式方程方形系统的所有根（包括复数根以及在无穷远处的根）。诸如多变量消元法（Macaulay，1916；Collins，1971；Canny，1988；Manocha，1992）与柱面代数分解（Collins，1975；Arnon 等，1984）等符号运算方法也存在，Rieger（1987，1990，1992）提出的构造代数表面物体外观图的不同算法中也用了以上方法。

在本章我们集中注意力于正交投影的状态图，但是在 20.1 节中讨论的奇异切线与双切线也在透视投影中形成视觉事件边界，而相应的可展表面将三维视角空间开切成单元，它形成了透视状态图的结点。严格地讲，在透视状态图使用的成像模型仅仅对有 $360° \times 360°$ 视场的摄像，诸如在导论中讨论过的全方位球面摄像机有效①。对处在物体的凸包外的观察者，一个更方便的透视投影模型可以使用，对它视网膜的朝向要有所选择，使得物体处在针孔的前面。这一点是能做到的，因为（a）处在凸物体外的一点总可以用一平面将其与该物体分开，（b）只要物体保持在针孔前面，图像轮廓的拓扑结构就与图像平面朝向无关，这一点是很容易验证的。

关于在正交投影或透视投影条件下对多面体构造其外观图的算法已经提出很多（例如 Castore，1984；Stewman 与 Bowyer，1987，1988；Watts，1987；Gigus 与 Malik，1990；Plantinga 与 Dyer，1990；Wang 与 Freeman，1990；Gigus 等，1991），其中某些已经实现。在这种情况下只有两种视觉事件曲线——所谓的 EV 与 EEE 事件，它们发生在视线与多面体在一条边与一个顶点接触时，或在 3 个不同的边界点擦过该表面。对简单的曲面物体如由二次表面包围的刚体（Chen 与 Freeman，1991）以及回旋刚体（Eggert 与 Bowyer，1989，1991；Kriegman 与 Ponce，1990*a*）构造精确的外观图的算法也已提出并实现了。我们在本章呈现的由代数表面包围的刚体的算法以及基于柱面代数分解的方法（Rieger，1987，1990，1992）很可能是当前最一般性的，但它们无法对付由上百个双三次表面组成的 CAD 典型模型。

遗憾的是外观图的尺寸是很大的；一个具有 n 个面的多面体的正交投影外观图由 $O(n^6)$ 个单元组成（Gigus 等，1991）。对由 n 个 d 阶多项式表面组成的平滑面片表面的物体，外观图的尺寸增加到 $O(n^6 d^{12})$（Petitjean，1995）。显然对透视投影的情况这种状况会变得更糟。我们是否应该把外观图尺寸巨大的责任推到表达方式自身？还是因为作为基础的表面模型（组合的与/或 代数的）的复杂度造成（例如对相对简单自由形式表面用多面体面的数目或表面块的

① 当然，真正的全方位摄像机是罕见的，但至少由 Nayar（1997）开发的基于反折时全方位摄像机技术是可以买到的

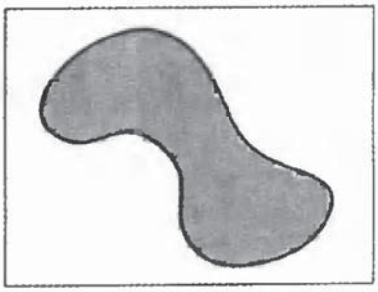
阶数)? Noble 等(1997)研究过由体密度函数的零集(例如 CT 图像中的器官的边界)定义的物体构造外观图的问题,并呈现了初步的结果。尺寸问题的另一部分因素与对实际镜头引入的光学模糊以及摄像机有限空间分辨率进行适当的建模有关。针对这个问题的初步努力可以在 Eggert 等(1993),Shimshoni 与 Ponce(1997)以及 Pae 与 Ponce(2001)中找到,其中部分是以对一般表面单参数族畸变的影响的新研究成果(Bruce 等,1996*a*,1996*b*)为基础的。

多面体的近似外观图已被成功地用到物体定位任务中。20.3 节讨论的方法是 Ikeuchi 与 Kanade(1987*b*, 1988)提出的。不同的方法还包括 Chakravarty(1982)以及 Hebert 与 Kanade(1985)。扩展高斯图是由 Horn(1984)引入的,一个双光度学体视方法在 Ikeuchi(1987*a*)中描述。

本章所叙述的用于被遮挡表面消除的射线跟踪与 z 缓冲算法可以在传统的计算机图形学课本(例如,Foley 等,1990)中找到。

习题

20.1 对下图中透明的扁平世界物体,画出其正交投影与球面透视的外观图与相应的外观:



20.2 画出该图非透明体的正交投影与球面透视外观图与相应的外观。

20.3 对一个具有单一抛物曲线的物体(如香蕉)是否可能完全没有高斯歧点? 为什么(或为什么不)?

20.4 使用方程式计算辩证法证明,对一般曲面来说线,与表面有 6 次或多于 6 次接触不会发生(提示:对定义接触的参数进行计数)。

20.5 我们看到一个渐近曲线与它的球面图像具有垂直切线。线曲率是主方向场的积分曲线。请演证这些曲线与它们的高斯图像具有平行的切线。

20.6 使用抛物曲线的高斯图像是与其相交的渐近曲线的包络这样一个事实,给出以下事实的另一个证明方法,即一对歧点在唇事件或喙对喙事件中出现(或消失)。

20.7 隐式表示曲面的唇以及喙对喙事件。可以证明(Pae 与 Ponce,2001)用某个密度函数 $F(x,y,z)=0$ 的零集隐式定义的表面的抛物曲线是由这个方程 $F^T\mathcal{A}\nabla F=0$ 描述的,其中$\nabla F=$是 F 的梯度,$\mathcal{A}$是对称矩阵

$$\mathcal{A}\stackrel{\text{def}}{=}\begin{pmatrix} F_{yy}F_{zz}-F_{yz}^2 & F_{xz}F_{yz}-F_{zz}F_{xy} & F_{xy}F_{yz}-F_{yy}F_{xz} \\ F_{xz}F_{yz}-F_{zz}F_{xy} & F_{zz}F_{xx}-F_{xz}^2 & F_{xy}F_{xz}-F_{xx}F_{yz} \\ F_{xy}F_{yz}-F_{yy}F_{xz} & F_{xy}F_{xz}-F_{xx}F_{yz} & F_{xx}F_{yy}-F_{xy}^2 \end{pmatrix}$$

还可以证明在一抛物点的渐近方向是$\mathcal{A}\nabla F$。

(a) 证明$\mathcal{A}\mathcal{H}=\mathrm{Det}(\mathcal{H})\mathrm{Id}$,其中$\mathcal{H}$表示 F 的 Hessian 矩阵。

(b) 证明高斯歧点是满足方程式 $\nabla P^T \mathcal{A} \nabla F = 0$ 的抛物点。提示:使用高斯歧点的渐近方向与抛物曲线相切,以及向量 ∇F 是由 $F=0$ 定义的表面的切平面的法线这两个事实。

20.8 隐式表面的燕尾事件。可以证明在一个双曲点的渐近方向 $\boldsymbol{a}$ 满足两个方程式 $\nabla F \cdot \boldsymbol{a} = 0$ 与 $a^T \mathcal{H} \boldsymbol{a} = 0$,其中$\mathcal{H}$表示 F 的 Hessian 矩阵。这两个方程只是表明一个表面与它的渐近切线之间的接触至少等于 3。沿拐结点的渐近切线与这表面有 4 阶接触,这可以用一个三重方程式描述,即

$$\begin{pmatrix} \boldsymbol{a}^T \mathcal{H}_x \boldsymbol{a} \\ \boldsymbol{a}^T \mathcal{H}_y \boldsymbol{a} \\ \boldsymbol{a}^T \mathcal{H}_z \boldsymbol{a} \end{pmatrix} \cdot \boldsymbol{a} = 0$$

描绘出一个对隐式表面燕尾事件跟踪的算法。

20.9 推导描述隐式曲面多重局部事件的方程式。可以使用以下事实,正如前一个练习指出的一个双曲点的渐近方向满足两个方程式 $\boldsymbol{a}^T \mathcal{H} \boldsymbol{a} = 0$ 与 $\nabla F \cdot \boldsymbol{a} = 0$。

编程作业

20.10 写一程序来揭示多重局部视觉事件:考虑两个不同半径的球并假设正交投影情况。这个程序应允许人们交互地改变视点,以及揭示与有界双切可展曲面有关的切向交叉。

20.11 为揭示歧点与它们的投影写一个相似的程序。你必须跟踪一个平面曲线。

第21章 距离数据

本章讨论距离图像(或者深度图像),这种图像存储的不是亮度和颜色信息,而是与每一个像素相关的射线与摄像机观测到的场景的第一次交点的深度信息。从一定意义上说,一幅距离图像正是立体视觉、运动或其他 shape-from-X 模块期望的输出。但是,本章集中讨论主动传感器获得的距离图像。主动传感器向场景投影某种光模式,以此来避开建立对应的困难和时间消耗问题,并构造出紧密和准确的深度图像。在对测距技术进行一个简洁的回顾后,我们将讨论图像分割、多幅图像的匹配、三维模型重构和对象识别,集中于距离数据领域中这些问题的相关方面。

21.1 主动距离传感器

基于三角测距技术的主动距离传感器可以回溯到 20 世纪 70 年代早期(比如,Agin,1972;Shirai,1972)。它们使用了与被动立体视觉系统相同的工作原理,其中一个摄像机被受控的照明光源(结构光)所替代,从而避开了第 11 章所提到的对应问题。比如,一个激光器和一对旋转镜可以用来顺序地扫描某一表面,在这种情况下,正如传统的立体视觉系统一样,激光束打在感兴趣的物体表面上的光点的位置被认为是该点与其投影点连线的投影射线与光线的交点。与传统立体视觉系统不同的是,激光点一般很容易识别,因为它比其他场景点更亮(尤其是当只能通过激光波长的滤波器放在摄像机前面时),因此避免了对应问题。类似地,激光束能够用柱状的镜头变换为光平面(见图 21.1)。这简化了距离传感器的机械设计,因为它仅仅需要一个旋转镜子。或许更重要的是,它缩短了用于获取一幅距离图像的时间,因为激光带——等价于图像的整个一列,能够在每一帧中得到。应该注意的是,这种设置不会带来匹配不确定性,因为与激光点相关的每一个图像像素相关的激光点,可以由对应投影射线与光平面的惟一交点得到。

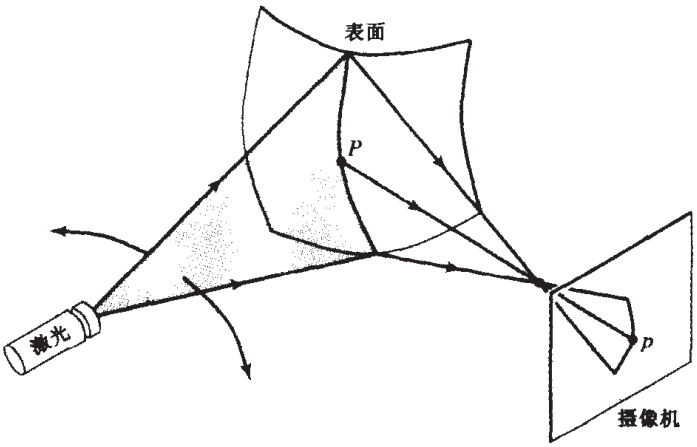


图 21.1 使用平面光扫描一个物体表面的距离传感器

这两种技术的各种变种，包括使用多个摄像机以提高测量精度和使用(也许是随时间编码的)二维模式光以提高数据获取速度。主动三角测距技术的主要缺点是数据获取速度低，在被物体遮挡的激光点处的图像数据会丢失，以及镜面反射导致的数据丢失和错误数据。后者对于所有的主动测距技术事实上是非常普遍的：纯粹的镜面反射面不会向摄像机的方向反射任何光线，除非它恰巧放在对应镜子反射的方向。最糟糕的是，反射光线可能引入二次反射，导致错误的距离度量。其他困难包括保持激光带在整个扫描过程中聚焦，以及随着距离的增加，所有三角测距技术固有的精度丢失问题(参考第 11 章的习题，本质上这是由于深度与误差成反比的事实)。现在已有市场销售的基于三角测距技术的扫描仪器(见图 21.2)。

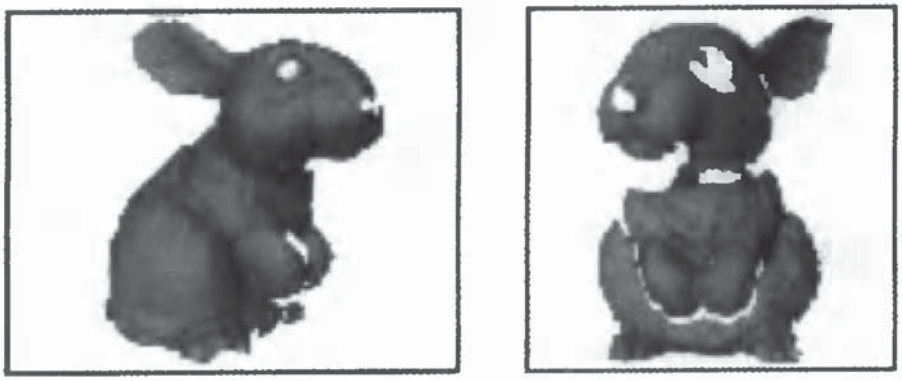


图 21.2 展示了使用美能达 VIVID 距离传感器获得的距离数据，该传感器的测量距离在 0.6 到 2.5 米之间，能够在0.6秒内获得一幅200 × 200的距离数据图像和一幅匹配好的400 × 400的彩色图像

第二种主要的主动测距技术涉及一个信号发射器、一个接收器和用于计算信号从距离传感器发射出去到与感兴趣的物体表面相交的整个过程的飞行时间。这种飞点扫描距离传感器 TOF(time-of-flight)一般装有一个扫描装置，发射器和接收器常常是共轴的，因而消除了三角测距技术中一般存在的丢失数据问题。TOF 距离传感器主要有三种：脉冲时延技术，直接度量激光脉冲的飞行时间；AM 调幅相移距离传感器，度量调频调幅激光雷达发射光束与反射光束的相位变化，该相位变化在数值上与飞行时间成正比；调频 beat 传感器，度量调频激光束及其反射的频率位移，该位移与来回旅程的飞行时间成正比。与基于三角度量的技术相比，TOF 传感器提供了更远的测量距离(可达几十米远)，适于室外机器人的导航。

新的技术仍然在不断出现，包括装有声光扫描系统的距离传感器和具有极高的图像获取速率的距离传感器，以及不使用扫描技术，使用具有大的接收器阵列的距离传感器，能够同时分析整个视场的激光脉冲的距离传感器。

21.2 距离数据的分割

本节改写了第 8 章和第 14 章介绍的一些边缘检测和分割方法，使之适用于距离图像的特定情况。正如本节将要介绍的，曲面几何的直接可用极大地简化了分割过程，因为它为曲面不连续处的定位，以及具有相似形状的相邻面片的合并提供了客观的、具有物理意义的标准。让我们从介绍偏微分几何分析学的基本概念开始，因为分析微分几何学是本节对距离图像进行边缘检测方法的基础。

21.2.1 分析微分几何学的基本元素

这里修改了第 19 章介绍的分析微分几何学的概念。假设在$\mathbb{E}^3$ 中有一个固定坐标系，并定义该空间为$\mathbb{R}^3$，定义每一个点用它的坐标向量表示。我们考虑一个参数曲面，该曲面由一个平滑的(也就是说，无限可微的)映射 $\boldsymbol{x}: U \subset \mathbb{R}^2 \to \mathbb{R}^3$ 所定义，该映射把$\mathbb{R}^2$ 的开子集 U 中任何一个二元组(u,v)与$\mathbb{R}^3$ 中的一个点 $\boldsymbol{x}(u,v)$相联系起来。为了确保切平面在曲面的任何一点存在，假定偏微分 $\boldsymbol{x}_u \overset{\text{def}}{=} \partial \boldsymbol{x}/\partial u$ 和$\boldsymbol{x}_v \overset{\text{def}}{=} \partial \boldsymbol{x}/\partial v$ 是线性无关的。事实上，令 $\alpha: I \subset \mathbb{R} \to U$ 表示一个平滑的平面曲线，其中，$\alpha(t) = (u(t), v(t))$，则 $\beta \overset{\text{def}}{=} \boldsymbol{x} \circ \boldsymbol{\alpha}$ 为曲面上的一个空间曲线。根据链式定理，在点 $\beta(t)$的曲线 β 的切向量为$u'(t)\boldsymbol{x}_u + v'(t)\boldsymbol{x}_v$，曲面在点 $\boldsymbol{x}(u,v)$的切平面平行于由向量 $\boldsymbol{x}_u$ 和$\boldsymbol{x}_v$ 张成的向量平面。因此，曲面的(单位)法向量为

$$N = \frac{1}{|\boldsymbol{x}_u \times \boldsymbol{x}_v|}(\boldsymbol{x}_u \times \boldsymbol{x}_v)$$

考虑切平面中位于点 $\boldsymbol{x}$ 的一个向量$\boldsymbol{t} = u'\boldsymbol{x}_u + v'\boldsymbol{x}_v$。容易证明第二基本范式可以由下式给出①

$$\mathrm{II}(\boldsymbol{t},\boldsymbol{t}) = \boldsymbol{t} \cdot dN(\boldsymbol{t}) = eu'^2 + 2fu'v' + gv'^2, \quad \text{其中} \quad \begin{cases} e = -N \cdot \boldsymbol{x}_{uu} \\ f = -N \cdot \boldsymbol{x}_{uv} \\ g = -N \cdot \boldsymbol{x}_{vv} \end{cases}$$

注意到向量 $\boldsymbol{t}$ 的范围一般不是单位值。我们定义第一基本形式为与切平面中两向量点积相关的双线性形式 $I(\boldsymbol{u},\boldsymbol{v}) \overset{\text{def}}{=} \boldsymbol{u} \cdot \boldsymbol{v}$。可以得到

$$\mathrm{I}(\boldsymbol{t},\boldsymbol{t}) = |\boldsymbol{t}|^2 = Eu'^2 + 2Fu'v' + Gv'^2, \quad \text{其中} \quad \begin{cases} E = \boldsymbol{x}_u \cdot \boldsymbol{x}_u \\ F = \boldsymbol{x}_u \cdot \boldsymbol{x}_v \\ G = \boldsymbol{x}_v \cdot \boldsymbol{x}_v \end{cases}$$

并且可以知道 $\boldsymbol{t}$ 方向的法曲率可由下式给出

$$\kappa_t = \frac{\mathrm{II}(\boldsymbol{t},\boldsymbol{t})}{\mathrm{I}(\boldsymbol{t},\boldsymbol{t})} = \frac{eu'^2 + 2fu'v' + gv'^2}{Eu'^2 + 2Fu'v' + Gv'^2}$$

类似地，与切平面的以$(\boldsymbol{x}_u, \boldsymbol{x}_v)$为基的高斯图的偏微分相关的矩阵可以很容易计算出来：

$$dN(\boldsymbol{t}) = \begin{pmatrix} e & f \\ f & g \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1}$$

因此，由于高斯曲率等于 dN 运算子的行列式的值，它可由下式给出：

$$K = \frac{eg - f^2}{EG - F^2}$$

渐近线的方向和主方向通过参数化同样可以很容易地得到：因为渐近线的方向满足

① 这种定义法是与 19 章中定义方向的惯例保持一致，系数 e, f, g 经常定义成相反的符号(例如 do Carmo, 1976, Struik, 1988)。

$\mathrm{II}(\boldsymbol{t},\boldsymbol{t})=0$，$u'$和$v'$对应的值就是方程$eu'^2+2fu'v'+gv'^2=0$的解。主方向满足：

$$\begin{vmatrix} v'^2 & -u'v' & u'^2 \\ E & F & G \\ e & f & g \end{vmatrix}=0 \tag{21.1}$$

例 21.1 Monge 曲面

一个很重要的参数化曲面的例子是 Monge 曲面：考虑曲面$\boldsymbol{x}(u,v)=(u,v,h(u,v))$。在这种情况下，我们有

$$\begin{cases} \boldsymbol{N}=\dfrac{1}{(1+h_u^2+h_v^2)^{1/2}}(-h_u,-h_v,1)^T \\ E=1+h_u^2,\ F=h_uh_v,\ G=1+h_v^2 \\ e=-\dfrac{h_{uu}}{(1+h_u^2+h_v^2)^{1/2}},\ f=-\dfrac{h_{uv}}{(1+h_u^2+h_v^2)^{1/2}},\ g=-\dfrac{h_{vv}}{(1+h_u^2+h_v^2)^{1/2}} \end{cases}$$

并且高斯曲率具有非常简单的形式：

$$K=\frac{h_{uu}h_{vv}-h_{uv}^2}{(1+h_u^2+h_v^2)^2}$$

例 21.2 局部曲面参数化

另一个基本例子是在由曲面主方向构成的坐标系中对该曲面的局部参数化，这是 Monge 曲面的一个特例。由于坐标系原点在切平面上，立刻可以得到$h(0,0)=h_u(0,0)=h_v(0,0)=0$。正如所期望的，原点处法线向量是$\boldsymbol{N}=(0,0,1)^T$，并且第一基本形式也是单位向量。正如在习题中所显示的，由式(21.1)可以很容易得到，参数化曲面的坐标曲线是主方向的充要条件是$f=F=0$(这意味着，可展曲面的曲率线是它的平行线和中线)。在我们讨论的情况中，已经知道$F=0$，并且这个条件退化为$h_{uv}(0,0)=0$。这种情况中的主曲率就是$\kappa_1=e/E=-h_{uu}(0,0)$和$\kappa_2=g/G=-h_{vv}(0,0)$。特别地，可以写出(0,0)邻域的泰勒展开式

$$h(u,v)=h(0,0)+(h_u,h_v)\begin{pmatrix}u\\v\end{pmatrix}+\frac{1}{2}(u,v)\begin{pmatrix}h_{uu} & h_{uv}\\ h_{uv} & h_{vv}\end{pmatrix}\begin{pmatrix}u\\v\end{pmatrix}+\varepsilon(u^2+v^2)^{3/2}$$

其中，h的导数的参数(0,0)为了简洁被省略。这表明曲面在此邻域的最佳二阶近似是由下式定义的抛物线

$$h(u,v)=-\frac{1}{2}(\kappa_1u^2+\kappa_2v^2)$$

即在第 19 章已经遇到的表达式。

21.2.2 在距离图像中寻找阶跃和顶边

本节展示了在距离图像中检测各种边缘的一个方法(Ponce 和 Brady，1987)。该技术组合了分析微分几何学和尺度空间图像分析技术，用于检测和定位距离数据中的深度和方向不连续位置。图 21.3 中是一个摩托车油瓶的距离图像，该图像用来阐明本节介绍的概念。

油瓶的曲面可以建模为基于传感器的坐标系统中的一个 Monge 曲面$z(x,y)$，它展示了两种不连续情况：阶跃，实际深度不连续的位置；顶边，深度连续但是方向急剧改变了的位置。正如下一节所示，在高斯平滑下描述阶跃和顶边的分析模型的性能是可能的，它们也相应地产生了抛物点和主曲率在对应的主方向上的极值。这是算法 21.1 多尺度边缘检测框架的基础。

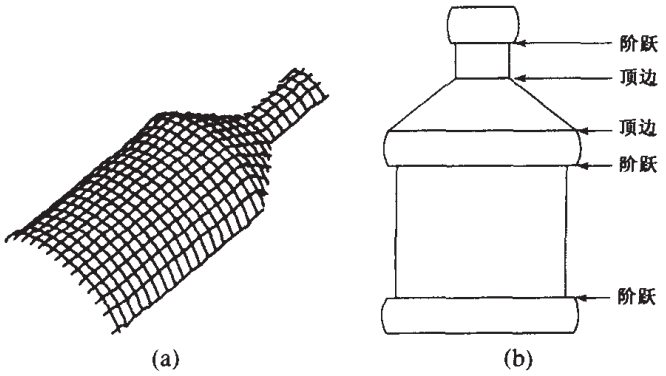


图 21.3　一个油瓶:(a) 该油瓶的一幅距离图像(背景已经被阈值消除);(b)深度和方向的不连续草图。该128 × 128图像使用INRIA距离传感器(Boissonnat和Germain,1981)获取,该传感器的深度精度约为0.5mm

算法 21.1　Ponce 和 Brady(1987)基于模型的边缘检测算法

1. 使用尺度集合 $\sigma_i(i=1,\cdots,4)$的高斯分布对距离数据平滑。计算平滑后的图像上的每一点的主方向和曲率。
2. 在每一幅平滑后的图像 $Z_{\sigma_i}(x,y)$上标出高斯曲率的过零点和支配主曲率在对应主方向的极值。
3. 使用分析的阶跃和顶边模型对各尺度上发现的特征进行匹配,并输出在曲面不连续处的点。

边缘模型　在不连续处的邻域,曲面形状在不连续的方向比在其正交的方向改变得更快。相应地,在本节后续部分我们假设不连续的方向就是主方向之一,相应的主曲率在该方向上发生急剧变化,而另外一个仍然接近于零。这使得我们可以把注意力限定于曲面不连续处的柱形模型上(比如,形式为 $z(x,y)=h(x)$的模型)。这些模型仅仅在边缘的邻域有效,其 $x-z$ 平面的方向与对应的主分量方向一致。

特别地,阶跃边缘可以用被垂直的分裂线分开的两个倾斜的半平面来建模,两个半平面的法线在 $x-z$ 平面上。该模型就是柱形模型,研究它的单变量公式(图 21.4 左)就足够了,其方程为

$$z=\begin{cases} k_1x+c & \text{当}\quad x<0 \\ k_2x+c+h & \text{当}\quad x>0 \end{cases} \tag{21.2}$$

在该表达式中,c 和 h 是常数,h 表示间隙的大小,k_1 和 k_2 表示两个半平面的斜率。引入新的常量 $k=(k_1+k_2)/2$ 和 $\delta=k_2-k_1$,很容易发现(见习题),把 z 函数与高斯函数的二阶导数进行卷积可以得到

$$z''_\sigma \stackrel{\text{def}}{=} \frac{\partial^2}{\partial\sigma^2}G_\sigma * z=\frac{1}{\sigma\sqrt{2\pi}}\left(\delta-\frac{hx}{\sigma^2}\right)\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{21.3}$$

正如第 19 章的习题所显示的,挠参数曲线的曲率是 $\kappa=|\boldsymbol{x}'\times\boldsymbol{x}''|/|\boldsymbol{x}'|^3$。在平面曲线的情况下,曲率可以赋以一个有意义的符号,并且该公式变为 $\kappa=(\boldsymbol{x}'\times\boldsymbol{x}'')/|\boldsymbol{x}'|^3$,其中的"×"这一次

表示与$\mathbb{R}^2$中的两个向量相关的坐标的行列式值操作。对应的曲率κ_σ在$x_\sigma=\sigma^2\delta/h$处消失。当$k_1=k_2$时该点只能在原点,在其他情况下它的位置是$\sigma$的二次函数。这启示我们使用主曲率之一(等价于高斯曲率)的过零点来定位阶跃边缘,其位置随着尺度的变化而改变。为了定性地刻画这些特征与σ的函数的关系,并且注意到,由于$z''_\sigma=0$在x_σ处等于零,我们有

$$\frac{\kappa''_\sigma}{\kappa'_\sigma}(x_\sigma)=\frac{z''''_\sigma}{z''_\sigma}(x_\sigma)=-2\frac{\delta}{\sigma}$$

换句话说,曲率的二阶导数与一阶导数的比与σ无关。

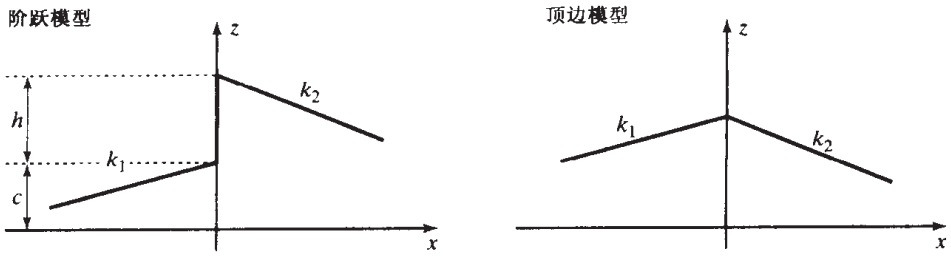


图 21.4 边缘模型:阶跃模型的两个半平面在逐点处被距离h分离,顶边模型的两个半平面在原点处连接(Ronce和Brandy,1987,图4)

顶边的一个分析模型可以在阶跃模型中(图 21.4 右)令$h=0$和$\delta\neq0$得到。在该情况下,很容易证明

$$\kappa_\sigma=\frac{1}{\sigma\sqrt{2\pi}}\frac{\delta\exp(-\frac{x^2}{2\sigma^2})}{\left[1+\left(k+\frac{\delta}{\sqrt{2\pi}}\int_0^{x/\sigma}\exp(-\frac{u^2}{2})du\right)^2\right]^{3/2}} \tag{21.4}$$

进一步可以知道,当$x_2=\lambda x_1$以及$\sigma_2=\lambda\sigma_1$时,一定有$\kappa_{\sigma_2}(x_2)=\kappa_{\sigma_1}(x_1)/\lambda$。因而,$|\kappa_\sigma|$的最大值一定是与$\sigma$成反比,它到原点的距离与$\sigma$成正比。该最大值在$\sigma$趋近于零时趋向于无穷,表明通过寻找局部曲率最大可以来定位顶边。在真实的距离数据中,应该在主方向上寻找,这与我们在曲面边缘邻域的局部形状变化的假设相吻合。

计算主曲率和主方向 根据上一节推导出来的模型,阶跃和顶边都可以通过寻找高斯曲率的过零点和在主方向上的主曲率的极值来定位。计算这些微分量要求估计距离图像中的每一点上的深度函数的一阶和二阶偏导数。由第 8 章我们知道,这可以通过把图像与高斯函数的导数进行卷积来得到。然而,距离图像与一般的图像不同:比如,在一般的图像中阶跃边缘邻域的像素值常常假设为分段不变的,该假设对于 Lambertian 物体成立,因为一阶导数意义上,曲面的形状在边缘附近是分段不变的,在此情况下,就具有分段平面的密度。相反,距离数据的分段不变(局部)模型一般是得不到满足的。类似地,在一般的图像中沿显著边缘对比度的最大值一般假定为大致相同。然而在距离数据中,有两种不同类型的阶跃边缘:将刚体彼此以及刚体及其背景之间互相分离的较大的深度不连续性,以及通常分离同一曲面各个片段的较小的间隙。本节讨论的边缘检测技术主要是针对后一种情况的。

盲目地在物体边缘使用高斯平滑将会引入急剧的形状变化，可能会破坏我们感兴趣的曲面细节(图 21.5，中间和顶部的图)。这就提示我们应该首先检测主要的深度不连续处，然后把平滑过程限制在由边界包围的曲面片中。这可以通过把距离图像与微形算子进行卷积得到(Terzopoulos，1984)，这些算子都是一些线性模版，组合在一起形成了一个 3×3 的均值掩码；比如，

$$\begin{bmatrix}1&&\\&2&\\&&1\end{bmatrix}+\begin{bmatrix}2&4&2\end{bmatrix}+\begin{bmatrix}2\\4\\2\end{bmatrix}+\begin{bmatrix}&&1\\&2&\\1&&\end{bmatrix}=\begin{bmatrix}1&2&1\\2&12&2\\1&2&1\end{bmatrix}$$

根据中心极限定理，循环使用 3×3 掩码(归一化从而保证权值和为一)与图像卷积，n 次迭代后所产生的结果，近似于对图像使用方差正比于$\sqrt{n}$的高斯进行平滑。为了避免平滑跨过不连续处，在不连续处不使用跨过不连续处的微形算子，并且对余下的再一次归一化，从而使权值的和等于 1。其效果见图 21.5(最下面一幅)。

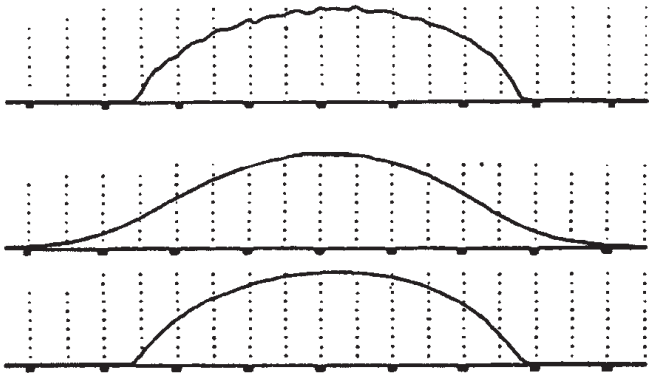


图 21.5　对距离数据图像进行平滑。顶图：图 21.3 中的距离图像的一个截面。背景已经被阈值除去。中图：高斯平滑的结果。底图：使用微形算子进行平滑

对曲面进行了平滑操作以后，就可以通过有限次差分得到高度函数的偏导数。高度函数的梯度可以通过把平滑后的图像与以下类似的掩码卷积得到：

$$\frac{\partial}{\partial x}=\frac{1}{6}\begin{bmatrix}-1&0&1\\-1&0&1\\-1&0&1\end{bmatrix}\quad 和\quad \frac{\partial}{\partial y}=\frac{1}{6}\begin{bmatrix}1&1&1\\0&0&0\\-1&-1&-1\end{bmatrix}$$

Hessian 矩阵可以通过把平滑后的图像与以下掩码卷积得到：

$$\frac{\partial^2}{\partial x^2}=\frac{1}{3}\begin{bmatrix}1&-2&1\\1&-2&1\\1&-2&1\end{bmatrix},\quad \frac{\partial^2}{\partial x\partial y}=\frac{1}{4}\begin{bmatrix}-1&0&1\\0&0&0\\1&0&-1\end{bmatrix}\quad 和\quad \frac{\partial^2}{\partial y^2}=\frac{1}{3}\begin{bmatrix}1&1&1\\-2&-2&-2\\1&1&1\end{bmatrix}$$

一旦知道了导数，主方向和主曲率可以很容易计算出来。图 21.6 显示了经过使用微形算子 20 次迭代后，检测出来的油瓶的两组主方向。正如所期望的，它们位于油瓶的子午线和旋转平行线上。

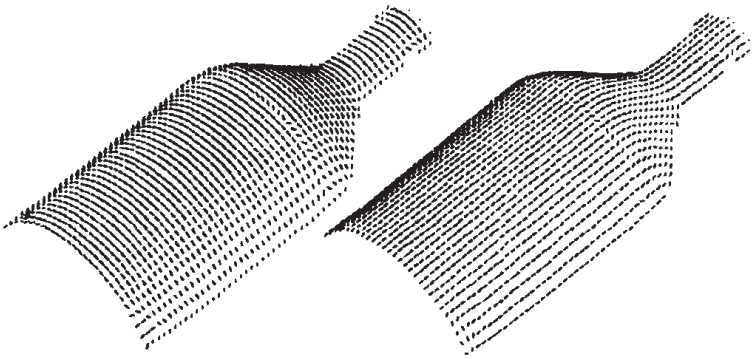
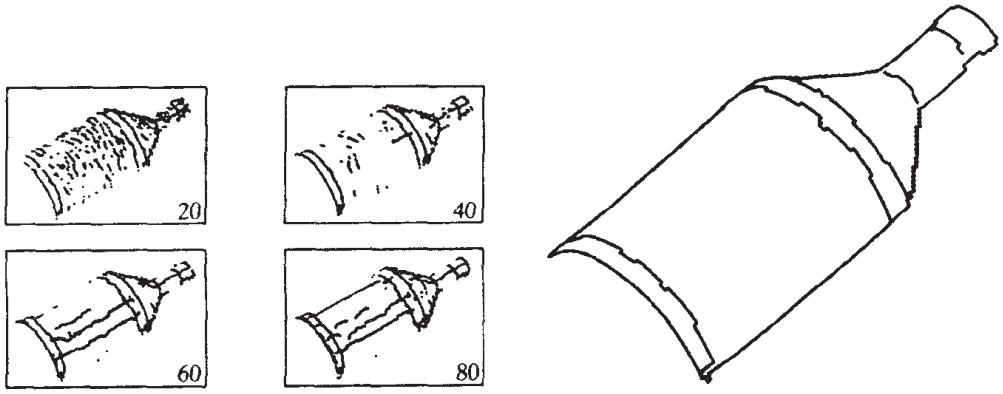


图 21.6 油瓶的两个主方向场

特征的多个尺度上的匹配 给定主曲率和主方向，抛物点可以通过高斯曲率的（无向的）过零点很容易地检测出来，主曲率在沿着对应的主方向上的局部最大值可以使用第 8 章讨论的非最大抑制技术得到。虽然在细分辨率时（比如仅仅经过一些迭代之后）存在一定量的噪声，情况会随着平滑过程而改进。由于噪声而检测到的特征也能够通过对抛物点斜率的过零点和主曲率极值的大小进行阈值操作排除，至少是部分排除。然而，实验显示平滑和阈值操作并不足以消除所有造成假象的特征。尤其是图 21.7（左）所显示的情况，平行于油瓶的曲率极值随着平滑过程越来越清晰。这是因为靠近油瓶遮挡边界的点比起靠近油瓶中心的点被微形算子平滑的程度低。

多尺度的边缘检测可以解决这个问题。由粗到精地跟踪特征，在一个给定的尺度下，所有的在较粗一级尺度中没有祖先的特征将被删除。主曲率的演变及其导数也被监视。跟踪留下来的比率 $\kappa_\sigma''/\kappa_\sigma'$ 在跨尺度时基本保持常数的抛物特征作为阶跃边缘点，而 $\sigma\kappa_\sigma$ 基本保持不变的主曲率的极值点作为阶跃边缘点输出。最后，由于这两个模型的对应过零点或者极值点与实际不连续处的距离，随着尺度增大而增大，最精细的尺度用于边缘定位。图 21.7（右）显示了使用该策略在油瓶距离图像上得到的检测结果。



21.7 在油瓶距离图像中寻找阶跃和顶边。左：分别经过 20,40,60 和 80 次平滑迭代和阈值化后检测的特征。阈值是凭经验选择的以消除大部分错误特征，同时保留了对应真实曲面不连续处的特征。但是一些假象，比如平行于油瓶轴线的曲率的极值仍然存在。右：基于模型的边缘检测的输出结果：油瓶的三个阶跃边缘和两个顶边不连续处被正确地识别出来

21.2.3　把距离图像分割为平面区域

在上一节我们看到，在照片图像和深度图像中，边缘检测使用的是完全不同的方法。在图像分割中也是类似。尤其是，有意义的分割标准在亮度邻域是难以找到的，因为像素的亮度（或者颜色）仅是形状或者反射性等物理特性的一种线索。然而，在距离领域，几何信息是直接可用的，使得我们可以使用表面点集与它们最佳拟合平面之间的距离作为有效的分割标准。本类方法的一个较好的例子是 Faugeras 和 Hebert(1986)提出的区域增长法。该算法通过运行一个图来迭代地合并平面片，图的节点是面片，弧线是连接相邻面片的公共边界。每一条弧线都有一个成本，大小等于这两个平面片上的点与它们的最佳拟合平面的平均误差。总是选择最好的弧线，并且合并对应的面片。注意，与这些面片相关联的弧线必须被删除，而连接新的面片与它们的相邻面片的弧线必须加入进来。过程参见图 21.8。

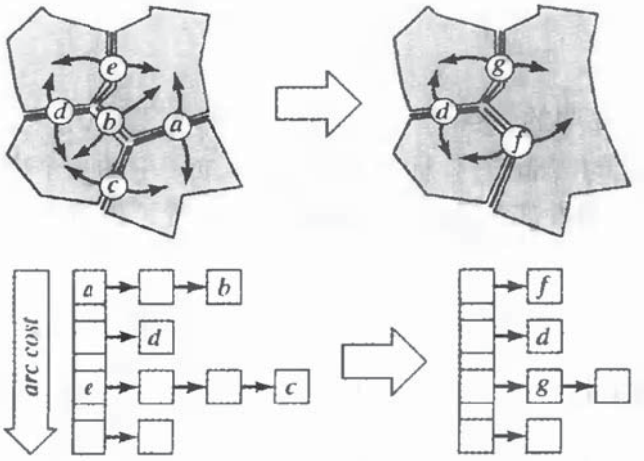
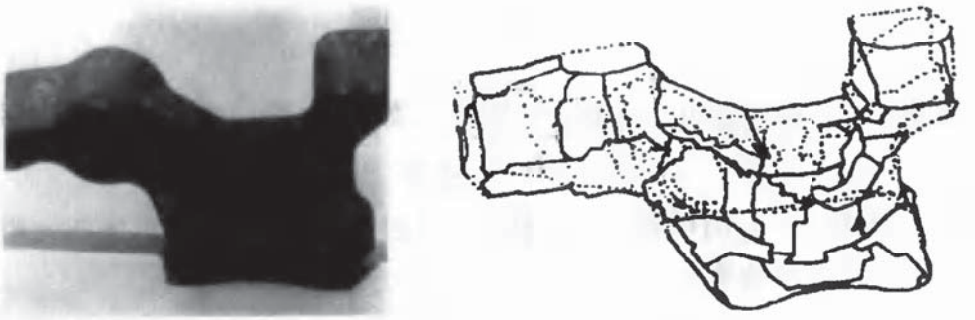


图 21.8　本图表说明了区域增长过程的一次迭代，其中以标有 a 的最小成本弧线为界的两个曲面片被合并在一起。下部的堆阵也被更新：弧线 a、b、c 和 e 被删除，两条新的弧线 f 和 g 被创建并加入图中

图的结构是使用距离数据的三角化来初始化的，并通过运行一个活动弧线的堆阵来更新。三角化或者可以直接从一幅距离图像中构造出来（通过沿着四边形的一个对角线上的像素进行分割得到），或者从由多幅图像构造出来的整体表面模型中构造出来。存储活动弧线的堆阵可以用一个桶阵列来表示，以增加的时间消耗为序，支持快速的插入和删除（图 21.8 下）。图 21.9展示了一个例子，其中的 Renault 小汽车零件用大约 60 个平面片来近似。



(a)　　(b)

图 21.9　Renault 小汽车零件　(a) 零件照片　(b) 零件模型

21.3 距离图像的匹配和模型获取

实际物体的几何模型在机械制造时非常有用(比如,用于处理和装配规划或者检验)。与本书主题更相关的是,它们同时也是许多物体识别系统的关键组成部分,并且在娱乐业中获得了越来越多的需求,比如在故事影片和视频游戏中经常出现真实物体的合成图像(第 26 章将对该问题详细讨论)。距离图像为构造物体准确的几何模型提供了非常好的数据,但是单一图片最多只能显示某一固体表面的一半,整个固体模型的构造要求集成多幅距离图像。本节讨论把多幅图像配准到同一坐标系中的问题和把这些图像提供的三维数据融合成一幅集成的曲面模型的问题。在解决这两个问题前,先介绍四元组,在本节的配准问题和下一节的识别问题中,它为从点和平面的对应中估计刚体变换提供了一个线性的方法。在本章后面我们假设在 $\mathbb{E}^3$ 中存在一个固定的坐标系,并定义该空间为$\mathbb{R}^3$,每一个点由其坐标向量定义。

21.3.1 四元组

四元组是由 Hamilton 发明的(1844)。类似于平面中的复数,它可以方便地用于表示空间中的旋转。四元组 q 由它的实部,一个标量 a,和虚部,$\mathbb{R}^3$ 中的一个向量 $\boldsymbol{\alpha}$ 定义,并且一般表示为 $\mathrm{q}=a+\boldsymbol{\alpha}$。实数可以用虚部为零的四元组表示,向量可以看做实部为零的四元组,四元组的加法定义为

$$(a+\boldsymbol{\alpha})+(b+\boldsymbol{\beta})\stackrel{\text{def}}{=}(a+b)+(\boldsymbol{\alpha}+\boldsymbol{\beta})$$

一个四元组与一个标量相乘一般表示为 $\lambda(a+\boldsymbol{\alpha})\stackrel{\text{def}}{=}\lambda a+\lambda\boldsymbol{\alpha}$,并且这两个操作给整个四元组集合一个四元向量空间的结构。也可以定义两个四元组相乘

$$(a+\boldsymbol{\alpha})(b+\boldsymbol{\beta})\stackrel{\text{def}}{=}(ab-\boldsymbol{\alpha}\cdot\boldsymbol{\beta})+(a\boldsymbol{\beta}+b\boldsymbol{\alpha}+\boldsymbol{\alpha}\times\boldsymbol{\beta})$$

四元组,加上前面定义的加和乘操作,形成了一个非交换域。其零元和单位元分别是常数 0 和 1。四元组 $\mathrm{q}=a+\boldsymbol{\alpha}$ 的共轭是$\bar{\mathrm{q}}\stackrel{\text{def}}{=}a-\boldsymbol{\alpha}$,它有一个相反的虚部。四元组的二次范式定义为

$$|\mathrm{q}|^2\stackrel{\text{def}}{=}\mathrm{q}\bar{\mathrm{q}}=\bar{\mathrm{q}}\mathrm{q}=a^2+|\boldsymbol{\alpha}|^2$$

很容易证明,对于任何四元组对 q 和 q′,有 $|\mathrm{q}\mathrm{q}'|=|\mathrm{q}||\mathrm{q}'|$。

现在可以看出四元组

$$\mathrm{q}=\cos\frac{\theta}{2}+\sin\frac{\theta}{2}\boldsymbol{u}$$

表达了在以下意义上围绕向量 $\boldsymbol{u}$ 的角度θ 的旋转$\mathcal{R}$:如果 $\boldsymbol{\alpha}$ 是$\mathbb{R}^3$ 中的某一向量,那么

$$\mathcal{R}\boldsymbol{\alpha}=\mathrm{q}\boldsymbol{\alpha}\bar{\mathrm{q}} \tag{21.5}$$

注意到$|\mathrm{q}|=1$ 并且 $-\mathrm{q}$ 同样表达了旋转$\mathcal{R}$。反过来,与某一给定单位四元组 $\mathrm{q}=a+\boldsymbol{\alpha}$,$\boldsymbol{\alpha}=(b,c,d)^T$相对应的旋转矩阵$\mathcal{R}$是

$$\mathcal{R}=\begin{pmatrix} a^2+b^2-c^2-d^2 & 2(bc-ad) & 2(bd+ac) \\ 2(bc+ad) & a^2-b^2+c^2-d^2 & 2(cd-ab) \\ 2(bd-ac) & 2(cd+ab) & a^2-b^2-c^2+d^2 \end{pmatrix} \tag{21.6}$$

该表达式可以由式(21.5)很容易地推导出来(注意:4 个参数 a,b,c,d 互相不独立,因为它们满足约束 $a^2+b^2+c^2+d^2=1$)。

最后,如果 q_1 和 q_2 是单位四元组,并且$\mathcal{R}_1$ 和$\mathcal{R}_2$ 是对应的旋转矩阵,那么四元组 q_1q_2 和 $-q_1q_2$ 都表示了旋转矩阵$\mathcal{R}_1\mathcal{R}_2$。

21.3.2 使用最近点迭代方法匹配距离图像

Besl 和 McKay(1992)提出了一个可以配准两个三维点集的算法(也就是说,计算映射第一个点集到第二个点集的刚体变换)。他们的算法简单地通过以下步骤迭代地最小化两个点集之间的平均距离:首先通过把每一个场景点与最近的模型点进行匹配来建立场景与模型特征间的对应,估计场景点到其匹配点之间的刚体变换,最后对场景应用计算出来的变换。当匹配点之间的平均距离小于某些给定阈值时迭代停止。最近点迭代(ICP)算法的伪代码如下。

容易证明,算法 21.2 在每一次迭代中都保证使得误差 E 单调减小:事实上,平均误差在匹配过程中不断减小,并且单个误差在决定最近点对的过程中也减小。算法本身不能保证收敛到全局(甚至于局部)最小值,并且需要为算法提供一个合理的刚体变换的初始猜测。为此出现了该算法的一些变形,其中包括对所有可能的变换粗略地采样和使用场景点和模型点集合的矩来估计变换。

算法 21.2 Besl 和 McKay (1992)的最近点迭代算法。辅助函数 Initialize-Registration 使用一些基于矩的全局配准方法,比如计算把场景映射到模型的刚体变换的一个初始估计。函数 Return-Closest-Pairs 返回匹配的点索引(i,j),i 和j 满足点j 是点 i 的最近点。函数 Update-Registration 从所选择的场景和模型中的点估计出刚体变换,而函数 Apply-Registration 对场景中的所有点应用刚体变换。

```
    Function ICP(Model, Scene);
  begin
  E'← +∞;
  (Rot, Trans) ← Initialize-Registration(Scene, Model);
  repeat
    E ← E';
    Registered-Scene ← Apply-Registration(Scene, Rot, Trans);
    Pairs ← Return-Closest-Pairs(Registered-Scene, Model);
    (Rot, Trans, E') ← Update-Registration(Scene, Model, Pairs, Rot, Trans);
    until |E'-E| < τ;
  return(Ret,Trans);
  end
```

寻找最近点对 在算法每一次迭代中,对给定的场景点 S 寻找模型中的最近点 M 需要(不严格的)$O(n)$时间,其中 n 是模型中的点的数目。事实上,有许多算法可以在附加预处理时间下,在 $O(\log n)$时间内在$\mathbb{R}^3$中搜索最近点,比如 k-d 树(Friedman 等 1977,log 查询时间仅仅是平均情况),或者更复杂的数据结构。Clarkson(1988)的广义随机算法需要预处理时间 $O(n^{2+\varepsilon})$,其中 ε 是任意小的正数,且查询时间是 $O(\log n)$。可以通过缓存以前的计算结果来

提高再次查询的效率。比如，Simon 等人(1994)在每一次 ICP 算法的迭代中存储每一个场景点的最近 k 个邻近点。由于刚体变换的逐步更新一般很小，很可能某一点的最近邻点在一次迭代后就存在于前一次迭代的 k 个最近邻点中。事实上，有效地确定最近邻点是否在缓存中是可能的(细节见 Simon 等，1994)。

估计刚体变换 在由旋转矩阵$\mathcal{R}$和平移向量 $\boldsymbol{t}$ 定义的刚体变换下，点 $\boldsymbol{x}$ 被映射到 $\boldsymbol{x}' = \mathcal{R}\boldsymbol{x} + \boldsymbol{t}$。因此，给定 n 对匹配点 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}'_i (i = 1, \cdots, n)$，我们寻找使得误差 E 最小的旋转矩阵$\mathcal{R}$和平移向量 $\boldsymbol{t}$

$$E = \sum_{i=1}^{n} |\boldsymbol{x}'_i - \mathcal{R}\boldsymbol{x}_i - \boldsymbol{t}|^2$$

首先注意到使 E 最小的 $\boldsymbol{t}$ 应该满足

$$0 = \frac{\partial E}{\partial \boldsymbol{t}} = -2\sum_{i=1}^{n}(\boldsymbol{x}'_i - \mathcal{R}\boldsymbol{x}_i - \boldsymbol{t})$$

或者

$$\boldsymbol{t} = \bar{\boldsymbol{x}}' - \mathcal{R}\bar{\boldsymbol{x}}, \quad 其中 \quad \bar{\boldsymbol{x}} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \quad , \quad \bar{\boldsymbol{x}}' \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}'_i \tag{21.7}$$

分别表示两个点集合 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_i'$ 的质心。

引入中心化的点 $\boldsymbol{y}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}$ 和 $\boldsymbol{y}'_i = \boldsymbol{x}'_i - \bar{\boldsymbol{x}} (i = 1, \cdots, n)$，得到

$$E = \sum_{i=1}^{n} |\boldsymbol{y}'_i - \mathcal{R}\boldsymbol{y}_i|^2$$

现在可以用四元组进行如下步骤的来最小化 E：令 q 表示与旋转矩阵$\mathcal{R}$相关的四元组。由于$|\mathrm{q}|^2 = 1$以及四元组范式的可乘属性，有

$$E = \sum_{i=1}^{n} |\boldsymbol{y}'_i - \mathrm{q}\boldsymbol{y}_i\bar{\mathrm{q}}|^2|\mathrm{q}|^2 = \sum_{i=1}^{n} |\boldsymbol{y}'_i\mathrm{q} - \mathrm{q}\boldsymbol{y}_i|^2$$

正如练习中所显示的，这允许我们改写旋转误差为 $E = \mathrm{q}^T \mathcal{B}\, \mathrm{q}$，其中$\mathcal{B} = \sum_{i=1}^{n} \mathcal{A}_i^T \mathcal{A}_i$，并且

$$\mathcal{A}_i = \begin{pmatrix} 0 & \boldsymbol{y}_i^T - \boldsymbol{y}_i'^T \\ \boldsymbol{y}'_i - \boldsymbol{y}_i & [\boldsymbol{y}_i + \boldsymbol{y}'_i]_\times \end{pmatrix}$$

注意，矩阵$\mathcal{A}_i$ 是非对称的秩为 3 的矩阵，但是矩阵$\mathcal{B}$在有噪声的情况下秩为 4。由第 3 章知道，在约束$|\mathrm{q}|^2 = 1$ 下最小化 E 是一个线性最小二乘问题，它的解就是矩阵$\mathcal{B}$的最小特征值相关对应的特征向量。一旦$\mathcal{R}$已知，$\boldsymbol{t}$ 可以由式(21.7)获得。

结果 图 21.10 展示了一个例子，其中非洲面具的两幅距离图像使用 ICP 算法配准在一起。对于该 9 cm 的物体，匹配的平均误差是 0.59 mm。

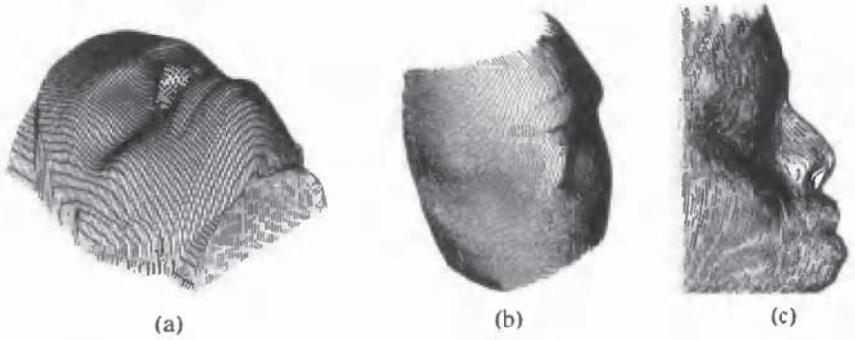


图 21.10 配准结果：(a) 一个用做模型的非洲面具的一幅距离图像；(b)模型的一个(十中取一抽取出来的)视图作为场景；(c)两个数据集配准得到的视图

21.3.3 多幅距离图像的融合

给定某一固体的配准好的距离图像的一个集合，可以构造该物体集成的曲面模型。在由 Curless 和 Levoy(1996)提出的方法中，该模型被构造为一个体积密度函数为 $D:\mathbb{R}^3\rightarrow\mathbb{R}$ 的零集合 S[也就是说，满足 $D(x,y,z)=0$ 的点(x,y,z)的集合]。类似于连续密度函数的任意水平集，S 在构造中保证了是封闭不漏水的表面，虽然它也许有几个连通件(见图 21.11)。

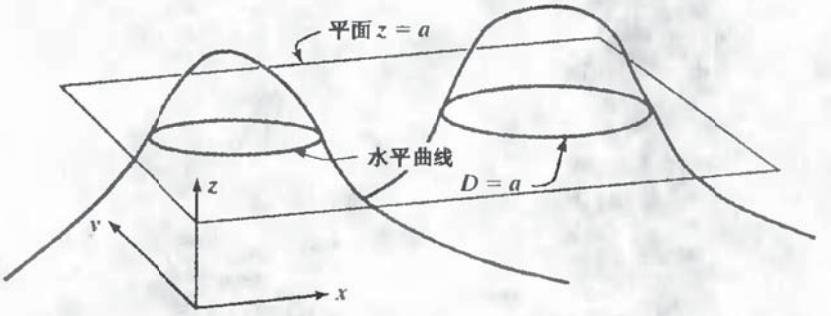


图 21.11 一个体密度函数的 2D 显示及其水平集。在本例中，“体积”显然是(x,y)平面，而“表面”是该平面中的一条曲线，例子中包含两个连通件

当然，困难主要在于从配准的距离度量中构造一个合适的密度函数。Curless 和 Levoy 把对应的表面片嵌入一个立方体网格，并对立方体网格的每一个单元，或者说体素，赋予一个中心到交于它的最近表面点的有向距离的加权和(图 21.12，左)。该平均有向距离就是期望的密度函数，其零集可以使用经典的技术获得，比如由 Lorensen 和 Cline(1987)发明的步进立方体算法，用于从体医学数据中提取同密度曲面。

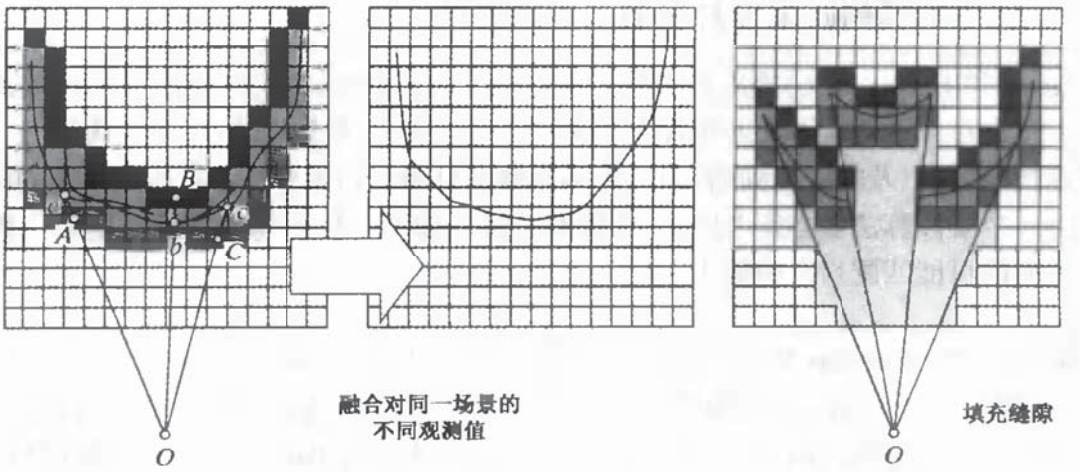


图 21.12 融合多幅距离图像的 Curless-Levoy 方法的一个 2D 例子。图的左边部分是位于点 O 的同一传感器观察到的3幅视图，通过计算体素中心(比如点A,B,C)到表面点(比如a,b,c)之间沿着视线的有向距离的加权平均的零集，被融合在一起。一般来说，使用的数据是到不同传感器的距离数据。图右边的浅灰部分是在缝隙填充过程中标示为空的体素集合

对于场景中未观察到部分的不可见曲面片，其体素初始标记为不可见，或者赋予一个等于大正数的深度值(代表无穷大)，然后像以前一样对所有接近度量表面片的体素相应地有符号距离，最后去除位于观察曲面片和传感器之间的体素(标记为空或者表示负无穷的大的深度

表示)(图 21.12,右)。

图 21.13 展示了一个例子,它是用一个 Cyberware 3030 MS 光学三角距离传感器获得的,一个佛像的多幅距离数据中建立的模型,以及通过立体成像从几何模型中构造出来的物理模型。

图 21.13 一个佛像的 3D 传真图。从左到右:佛像照片;距离图像;集成的3D模型;孔洞填充后的模型;由立体成像得到的物理模型

21.4 物体识别

现在考虑从距离图像中识别真实物体的问题。前一节介绍的配准技术在本节中介绍的两个算法中占据着关键的地位。

21.4.1 使用解释树匹配分片平面表示的表面

Faugeras 和 Hebert(1986)提出的识别算法是一个递归算法,采用刚体约束来有效地在一个解释树中搜索对应于最佳平面匹配序列的路径。基本程序参见算法 21.3 的伪代码。为了正确处理遮挡问题(以及前面提到的,距离传感器最多只能看到面对它的物体的一半的问题),在搜索的每一个阶段算法都必须考虑,一个模型平面可能不与任何场景平面匹配。一般是通过在某一平面的可能匹配列表中插入一个零标号平面。

算法 21.3 Faugeras 和 Hebert 的平面匹配算法(1986)。递归函数 Match 通过递归地访问解释树以返回最佳的平面匹配对集合。初始调用时,输入空的匹配对列表,将表示旋转和平移的参数 rot 和 trans 的值置为零。辅助函数 Potential-Matches 返回与模型平面Π相容的场景平面子集,以及映射模型平面至其场景匹配的刚体变换的当前估计(细节见正文文本)。辅助函数 Update-Registration-2 使用平面对来更新刚体变换的当前估计。

```
Function Match(model, scene, pairs, rot, trans);
begin
bestpairs ← nil; bestscore ← 0;
for Π in model do
```

(未完待续)

(续)

```
    for Π′ in Potential-Matches(scene, pairs, Π, rot, trans) do
      (rot,trans) ← Update-Registration-2(pairs, Π, Π′, rot, trans);
      (score, newpairs) ← Match(model − Π, scene − Π′, pairs + (Π,Π′), rot, trans);
      if score > bestscore then bestscore ← score; bestpairs ← newpairs endif;
      endfor;
    endfor;
  return (bestscore,bestpairs);
  end
```

选择可能的匹配 对一给定模型平面选择可能的匹配是基于各种标准的，依赖于已经建立的对应的数量，每一个新的对应提供了一个新的几何约束和更加严格的标准。在搜索开始时，我们仅仅知道面积为 A 的模型平面仅能与具有相容面积(也就是说，在范围$[\alpha A, \beta A]$内)的场景平面匹配。两个阈值的合理取值可以是 0.5 和 1.1，允许在遮挡区域之间有一些误差，以及达到 50% 的遮挡程度。

在建立起第一个对应后，估计映射模型到场景的刚体变换仍然过早，但是任何匹配平面的法线之间的角度应该大致等于第一对平面法线之间的夹角，或者说在间隔$[\theta-\varepsilon, \theta+\varepsilon]$之间。到对应平面的法线位于高斯球的一个带子内，并且可以通过对高斯球离散化，来对它们进行有效的检索，分配给每一个单元一个槽，存储法线落于其中的场景平面(见图 21.14)。

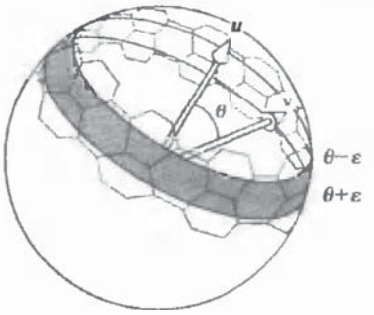


图 21.14 对于给定向量 $\boldsymbol{u}$，找到所有满足与 $\boldsymbol{u}$ 之间的夹角在$[\theta-\varepsilon, \theta+\varepsilon]$范围内的向量 $\boldsymbol{v}$。必须注意到，单位球不允许使用任意细节详细程度的规则(球体)多边形来镶嵌。图表中显示的镶嵌使用的是边长不相等的六边形(对该问题以及各种镶嵌方案的讨论见Horn，1986的第16章)

加上第二对对应足以完全确定把模型与场景中的实例分离的旋转：这在几何上是非常清楚的(在下一节中将使用分析学方法证明)，因为一对匹配向量限定旋转轴必须位于平分这些向量的平面内。两对匹配平面确定旋转轴为对应二分平面的交线。给定旋转和第三个模型平面，可以预测它与场景中的可能匹配的法线方向，它也可以使用上面提到的离散高斯球方法有效地恢复。在得到 3 对匹配后，可以估计出平移，并且使用平移估计来预测与第 4 个场景平面匹配的任何场景平面到原点的距离。这对其他更多的匹配对也是成立的。

估计刚体变换 考虑某一固定坐标系中由方程 $\boldsymbol{n}\cdot\boldsymbol{x}-d=0$ 定义的平面 Π，其中 $\boldsymbol{n}$ 表示该平面的单位法线，d 表示它到原点的有符号距离。在由旋转矩阵$\mathcal{R}$和平移向量 $\boldsymbol{t}$ 定义的刚体变换中，点 $\boldsymbol{x}$ 映射到点$\boldsymbol{x}'=\mathcal{R}\,\boldsymbol{x}+\boldsymbol{t}$，$\Pi$映射到平面$\Pi'$，$\Pi'$的方程是$\boldsymbol{n}'\cdot\boldsymbol{x}'-d'=0$，其中

$$\begin{cases} \boldsymbol{n}' = \mathcal{R}\boldsymbol{n} \\ d' = \boldsymbol{n}' \cdot \boldsymbol{t} + d \end{cases}$$

因此，估计 n 个平面Π_i 到$\Pi'_i(i=1,\cdots,n)$的映射等价于寻找旋转$\mathcal{R}$最小化误差

$$E_r = \sum_{i=1}^{n} |\boldsymbol{n}'_i - \mathcal{R}\boldsymbol{n}_i|^2$$

以及平移 $\boldsymbol{t}$ 最小化误差

$$E_t = \sum_{i=1}^{n} (d'_i - d_i - \boldsymbol{n}'_i \cdot \boldsymbol{t})^2$$

最小化 E_r 的旋转$\mathcal{R}$可以使用 21.4.1 节中的方法计算，即使用四元组表示$\mathcal{R}$，并且求解一个特征向量问题。最小化 E_t 的平移向量 $\boldsymbol{t}$ 是一个(非齐次)线性最小二乘问题的解，该解可以使用第 3 章中的技术求得。

结果 图 21.15 展示了使用图 21.9 中若干的 Renault 零件得到的检测结果。箱的距离图像使用 21.2.3 节的技术分割成平面块，匹配算法在该场景上运行了 3 次，在下一次迭代前，匹配了的面片被移除。正如图中显示的，箱中的三个零件被正确的识别，并且位姿估计过程的精确度通过根据计算出的位姿再次投影到模型的距离图像中得到验证。

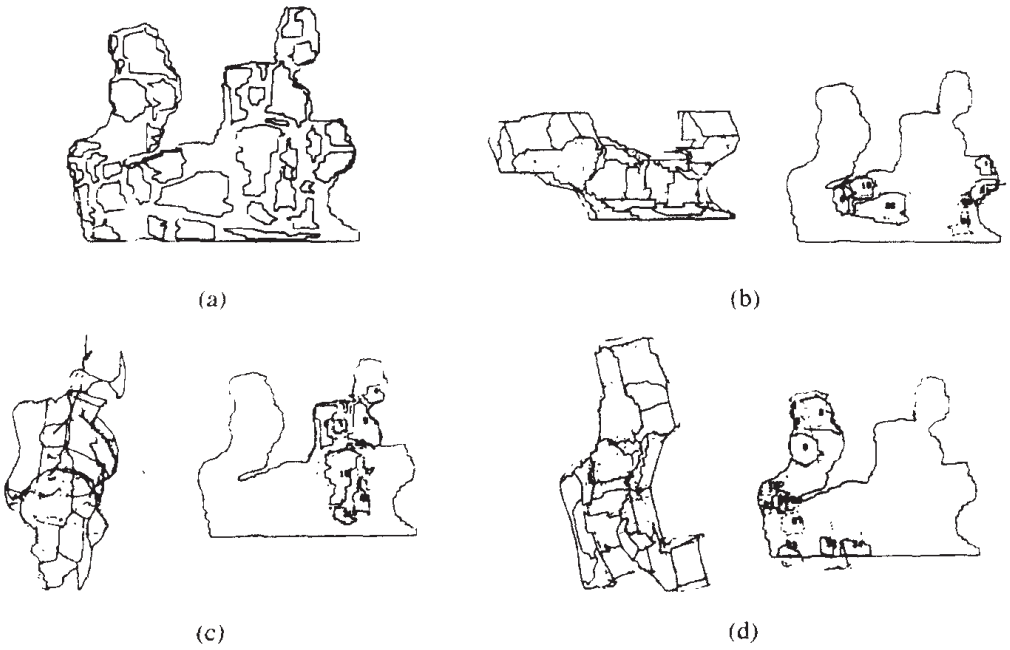


图 21.15 识别结果：(a)一个零件箱；(b)～(d)在零件箱中发现的三个零件。每一个例子中都显示出按照算法估计出的位置和方向的零件模型，以及在该位姿下(用虚线)迭加在对应平面上的距离图像

21.4.2 使用旋转图像匹配自由形态的曲面

在 21.2.2 节提到，微分几何为描述曲面的局部形状(比如，在曲面的点的每一个较小的邻域内)提供了一个有力的工具。另一方面，21.2.3 节的区域增长算法是为了基于平面片构造一个全局一致的表面描述。本节我们介绍一个半局部曲面描述——Johnson 和 Hebert(1998，1999)提出

的自旋图像，它在每一个点的较大的邻域内描述了表面的形状。在本节后面，可以证明自旋图像在刚体变换下是不变的，并且提供了一个有效的逐点表面匹配算法，从而跳过了识别过程中的分割。

自旋图像的定义　与 21.2.3 节一样，假设兴趣表面Σ以三角网格的形式给定。在每一个顶点的外曲面法线可以通过对顶点及其邻域的拟合平面来估计，并把三角网格转化为有向点的网。给定有向点 P，任一点 Q 的自旋坐标可以定义为Q 与P 处的有向法线间的(非负)距离 α，以及 Q 到切平面的(有向)距离 β(见图 21.16)。相应的，与 P 点相关联的自旋图像 $s_P:\Sigma\rightarrow\mathbb{R}^2$对于$\Sigma$上的任一点 Q 定义为

$$s_P(Q)\stackrel{\text{def}}{=}(\underbrace{|\overrightarrow{PQ}\times\boldsymbol{n}|}_{\alpha},\underbrace{\overrightarrow{PQ}\cdot\boldsymbol{n}}_{\beta})$$

从图 21.16 中可以看出，该映射不是单射。这并不奇怪，由于自旋图像仅仅提供了一个柱坐标系统的部分描述：一般记录了切平面中的某些参考向量与$\overrightarrow{PQ}$到该平面的投影之间的夹角的第三个坐标已经不存在。主方向是这样一个参考向量的必然选择，但是把焦点放在自旋坐标上避免了它们的计算——该过程因涉及二阶导数而对噪声敏感，而且可能对于平面片或者球面片存在二义性。

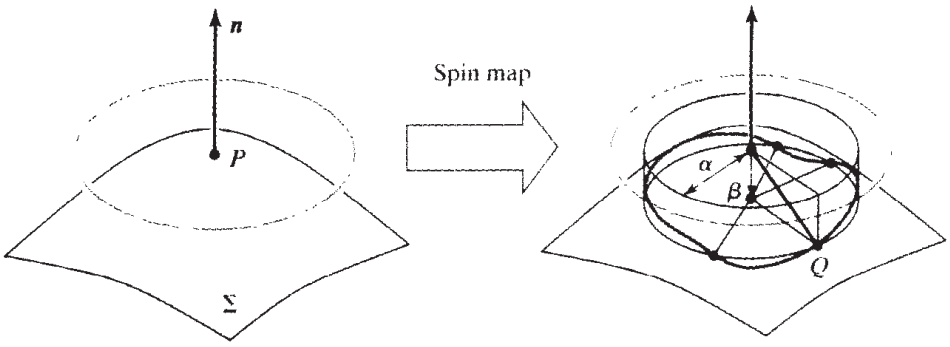


图 21.16　与表面点 P 相关的自旋图像的定义：点 Q 的自旋坐标(α,β)分别定义为$\overrightarrow{PQ}$在切平面上的投影的长度以及其在表面法线上的投影的长度。注意，在本例中，另外还有3个点与Q具有相同的(α,β)坐标

与某一个有向点相关的自旋图像是在该点邻域的 α,β 坐标的直方图。确切地说，α,β 平面被分解为 $\delta\alpha\times\delta\beta$ 个槽的矩形阵列，每个槽累计由具有 α,β 值的点在其范围内展开的表面总面积①。由 Carmichael 等(1999)以及练习可以看出，表面网格中的每一个三角形都映射到 α,β 平面的一个区域，它的边界是双曲弧线。它对自旋图像的贡献因此可以通过给该区域穿过的每一个槽赋予一个面片的面积，是与该槽相关的$\mathbb{R}^3$的环形区域相交的三角形的面片面积(见图 21.17)。这些槽可以通过扫描转换查找(Foley 等，1990)——该过程在计算机图形学中一般用于在较优的时间内，查找直边或者曲边构成的广义多边形所扫过的像素。

自旋图像由几个关键参数定义(Johnson 和 Hebert，1990)：第一个是限于以 P 为中心半径为 d 的球体范围内，用于构造该图像的支持点的支持距离 d。该球体必须足够大，以便提供较好

① 对应的点集实际上可能被分解为几个连通的部分。比如，对于足够小的值 $\delta\alpha$ 和 $\delta\beta$，在图 21.16 的例子中有 4 个连通的部分，对应于具有与 Q 相同 α,β 坐标的点为中心的小面片。

的描述能力，又必须足够小，以便在遮挡和杂乱情况下仍能识别。实践中，物体直径的十分之一可以是 d 的一个合适的选择：由此，如前所述，自旋图像实际上是在曲面的点的扩展领域内的曲面形状的半局部描述。对复杂环境的鲁棒性可以通过把支持点上的曲面法线的范围，限制在以 $\boldsymbol{n}$ 为中心半角为 θ 的锥形内。与选择支持距离相似，选择 θ 的合适的值也涉及在描述能力与对复杂环境的敏感之间的折中，由经验可知 60 度是该数值比较合适的选择。定义自旋图像的最后一个参数是槽的大小(像素单位)，或者在给定支持距离下，槽的大小(单位为米)。可以证明槽尺寸的合适大小是模型的网格顶点间的平均距离。图 21.18 展示了橡皮鸭面上的 3 个有向点相关的自旋图像。

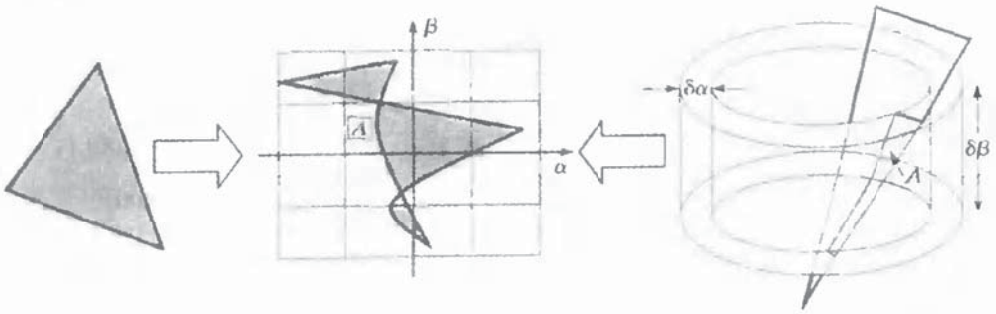


图 21.17 自旋图像的构造：图左边显示的三角形被映射到自旋图像中具有双曲边界的一个区域内。与该区域相交的每个槽的值用与该槽相关联的环相交的三角部分的面积递增

自旋图像的匹配 自旋图像的一个最重要的特点是，在刚体变换下它们是不变的。因此，原则上相关计算等图像比较技术，可以用于匹配与场景中的有向点及物体模型关联的自旋图像的匹配。然而事情并不那么简单。我们已经注意到，自旋映射不是单射的；一般来说，也不是满射，并且对于没有与物理表面点对应的 α,β 值可能出现空槽(零值像素)。遮挡也可以引起景物图像中零像素出现，而杂乱的背景也可能引入不相干的非空槽。因此，把两幅图像的比较限制在一般的非零像素是合理的。在这种情况下，Johnson 和 Hebert(1998)证明

$$S(\boldsymbol{I},\boldsymbol{J}) \stackrel{\text{def}}{=} [\text{Arctanh}(C(\boldsymbol{I},\boldsymbol{J}))]^2 - \frac{3}{N-3}$$

是对两幅重叠区域包含 N 个像素，并使用 $\mathbb{R}^N$ 中的向量 $\boldsymbol{I}$ 和 $\boldsymbol{J}$ 表示自旋图像的合适的相似性度量。在该公式中，$C(\boldsymbol{I},\boldsymbol{J})$ 表示向量 $\boldsymbol{I}$ 和 $\boldsymbol{J}$ 的规范化相关系数，并且 Arctanh 表示双曲正切函数。基于相似度量，我们可以给出一个使用自旋图像进行逐点对应的识别过程的轮廓(算法 21.4)。

该算法的各种阶段都是非常显然的。然而，需要指出的是，滤波和聚类过程依赖于把模型点相对于它本组的其他网格顶点的自旋坐标，与对应场景点相对于它本组的其他网格顶点的自旋坐标进行比较。一旦找出了具有一致性的组，模型与场景间的初始刚体变换，就可以使用 21.3.2 节中的四元组匹配技术从（有向的）点匹配中计算出来。最后，对应的相容集合可以通过下述方式进行验证，即迭代地把匹配过程扩展到邻域，同时更新将场景与模型对齐的刚体变换。

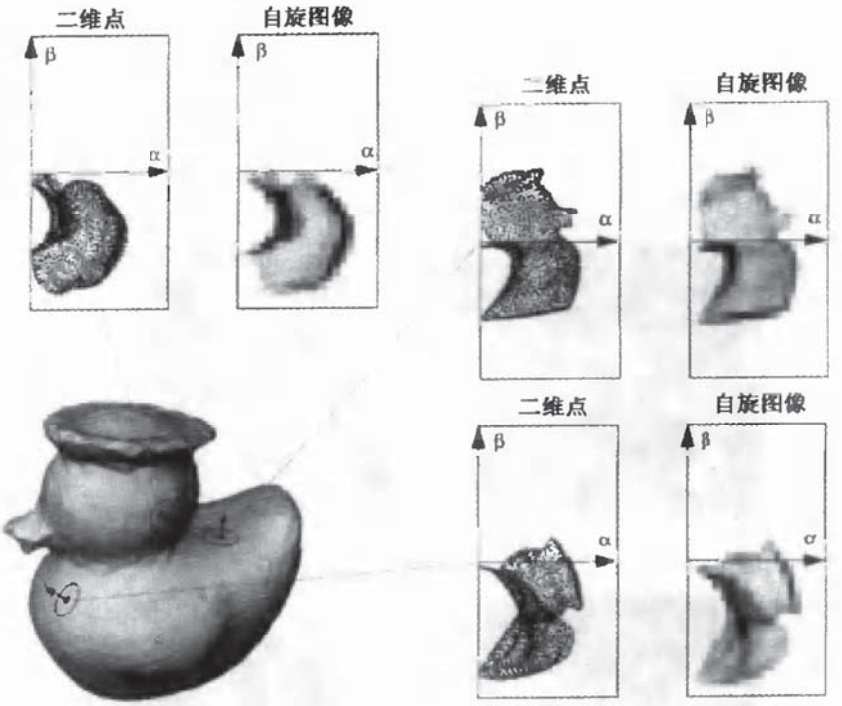


图 21.18 橡皮鸭的表面上的 3 个有向点及其对应的自旋图像。网格顶点的 a,b 坐标显示在真实的自旋图像旁

算法21.4 Johnson 和 Hebert (1998; 1999)使用自旋图像对任意曲面的逐点匹配算法。

Off-line:

Compute the spin images associated with the oriented points of a surface model and store them into a table.

On-line:

1. Form correspondences between a set of spin images randomly selected in the scene and their best matches in the model table using the similarity measure S to rank order the matches.
2. Filter and group correspondences using geometric consistency constraints, and compute the rigid transformations best aligning the matched scene and model features.
3. Verify the matches using the ICP algorithm.

结果 前一节的匹配算法已经在包含工业零件和各种玩具的复杂室内环境中的识别任务中广泛地测试(Johnson 和 Heert, 1998, 1999)。它还被用于室外的导航和地图生成任务中,数据集覆盖了几千平方米的区域(Carmichael 等, 1999)。图 21.19 是玩具情景简单的识别结果。

21.5 注释

主动距离感知技术非常好的综述参见 Jarvis(1983), Nitzan(1988), Besl(1989)和 Hebert (2000)。21.2.2 节基于模型的边缘检测方法,仅仅是使用微分几何学分割距离图像的众多技术之一(见 Fan 等, 1987; Besl 和 Jain, 1988)。用于距离图像平滑的微形算子计算的替换方法是

各向异性的漫射，其中每一点的平滑量依赖于梯度值(Perona 和 Malik，1990c)。21.2.3 节中把表面分割成(几乎)平面片的方法可以很容易扩展到二次曲面(见 Faugeras 和 Hebert，1986，以及练习)。扩展到高次曲面元是很有疑问的，部分是因为在这种情况下曲面拟合更加困难。有许多文献使用超二次曲面(比如，Pentland，1986；Bajcsy 和 Solina，1987；Gross 和 Boult，1988)或者代数曲面(比如，Taubin 等，1994*a*，*b*；Keren 等，1994；Sullivan 等，1994*a*，*b*)来解决后一问题。

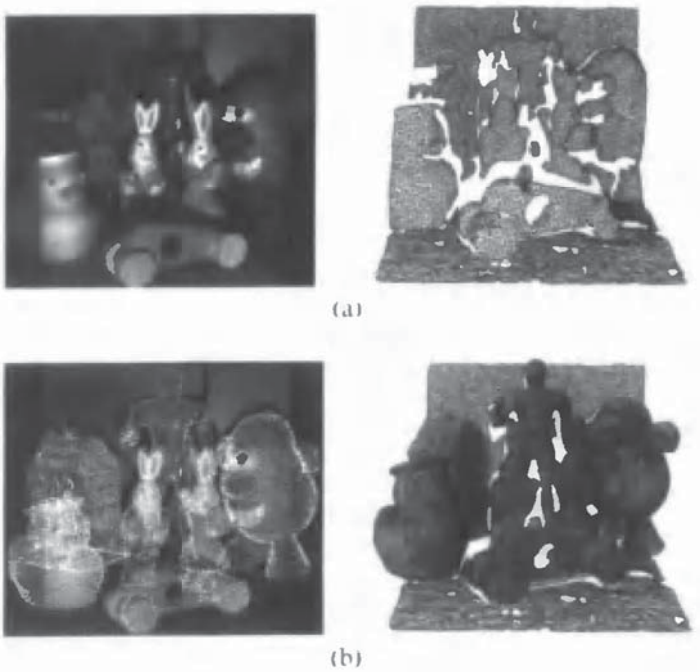


图 21.19 自旋图像识别结果：(a)一幅包含一些玩具的杂乱图像以及由相应的距离图像构造的网格图；(b) 识别出的物体迭加到原图像中

近年来，对在 21.3.2 节给出并由 Besl 和 McKay(1992)提出的 ICP 算法，出现了各种变形，包括能够处理丢失数据和/或外点的鲁棒方法(比如，Zhang，1994；Wheeler 和 Ikeuchi，1995)，这些方法已经在许多全局配准问题中得到应用(比如，Shum 等 1995；Curless 和 Levoy，1996)。

除了 Curless 和 Levoy(1996)的方法，其他多幅距离图像融合的方法还包括 Boissonnat(1984)的 Delaunay 三角法，Turk 和 Levoy(1994)的拉链多边形网格法，以及 Amenta 等人(1998)的外壳技术。本章的基于四元组的刚体变换估计方法是分别由 Faugeras 和 Hebert(1986)，以及 Horn(1987)独立开发的。21.4.1 节的识别技术，是与使用解释树在二维和三维情况下，控制特征匹配的组合耗费的其他算法密切相关的(Gaston 和 Lozano-Pérez，1984；Ayache 和 Faugeras，1986；Grimson 和 Lozano-Pérez，1987；Huttenlocher 和 Ullman，1987)。

21.4.2 节讨论的自旋图像被用于建立距离图像和曲面模型的点一级对应。与该问题相关的技术包括 Stein 和 Medioni(1992)的结构索引法，Chua 和 Jarvis(1996)提出的点签证法。在第 23 章中将从图像中识别物体的问题角度讨论该思想的一个变形(Schmid 和 Mohr，1997*a*，*b*)。总结来说，21.4.2 节描述的原算法已经扩展到各个方向：一个场景现在可以使用主分量分析(Johnson 和 Hebert，1999)，同时与多个模型匹配，另外学习技术被用于在复杂环境中删除错误匹配(Carmichael 和 Hebert，1999)。

习题

21.1　使用式(21.1)证明参数化曲面的坐标曲线是主方向的充分必要条件是 $f = F = 0$。

21.2　证明可展曲面的曲率线是它的中线和平行线。

21.3　阶跃模型:计算 $z_\sigma(x) = G_\sigma * z(x)$,其中 $z(x)$ 由式(21.2)给定。证明 z''_σ 由式(21.3)给定。证明在 $\kappa''_\sigma/\kappa'_\sigma = -2\delta/h$。得出结论:在 z''_σ 与 κ_σ 趋于零时有 x_σ 点。

21.4　顶边模型:证明 κ_σ 由式(21.4)给出。

21.5　证明四元组 $q = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}\,\boldsymbol{u}$ 表示了在式(21.6)下围绕向量 $\boldsymbol{u}$ 旋转角度为 θ 的旋转 $\mathcal{R}$。

提示:使用第 3 章练习中推导的 Rodrigues 公式

21.6　证明与给定单位四元组 $q = a + \boldsymbol{\alpha}$,其中 $\boldsymbol{\alpha} = (b, c, d)^T$ 相关的旋转矩阵 $\mathcal{R}$ 由式(21.6)给出。

21.7　证明由 21.3.2 节构造的矩阵 $\mathcal{A}_i$ 等于

$$\mathcal{A}_i = \begin{pmatrix} 0 & \boldsymbol{y}_i^T - \boldsymbol{y}_i'^T \\ \boldsymbol{y}_i' - \boldsymbol{y}_i & [\boldsymbol{y}_i + \boldsymbol{y}_i']_\times \end{pmatrix}$$

21.8　以前提到,ICP 算法可以扩展到各种类型的几何模型。在此我们考虑多面体以及分片参数面片。

a) 草拟一个方法用于计算多边形上离某点 P 最近的点 Q。

b) 草拟一个方法用于计算参数化面片 $\boldsymbol{x}: I \times J \to \mathbb{R}^3$ 上离某点 P 最近的点 Q。提示:使用牛顿迭代。

21.9　开发一个把一个点集与一个二次曲面拟合的线性最小二乘方法,二次型具有单位 Frobenius 范式的约束。

21.10　证明曲面三角在自旋图像的 α, β 空间中映射到一个具有双曲形边界的面片上。

编程作业

21.11　实现基于微形算子的平滑以及计算曲率和主方向。

21.12　实现本章描述的用于平面分割的区域增长法。

21.13　实现一个算法,从距离图像中计算出一个曲面的曲率线。提示:类似于平面分割的区域增长法,使用曲线增长算法。

21.14　实现 Besl-McKay ICP 匹配算法。

21.15　平面中的步进正方形方法:开发和实现一个算法,用于找到一个平面密度函数的零集。提示:找出曲线交于一个像素边缘的可能方法,在这些边界使用线性插值来找出零集。

21.16　实现 Faugeras-Hebert 算法的配准部分。

第六部分　高层视觉:基于概率和推理的方法

- 第 22 章　利用分类器建立模板
- 第 23 章　基于模板间关系的识别
- 第 24 章　基于空间关系的几何模板

第 22 章　利用分类器建立模板

在很多重要的物体识别问题中，需要在待识别物体上搜索一些具有简单形状和特定内容的图像窗口。什么是图像窗口呢？正面的人脸就可以看做一个椭圆形的图像窗口，在眼睛和嘴的位置上有一条深色的水平纹理，在鼻子的位置上有一条浅色的竖直纹理，而在面颊和前额的位置上则没有什么纹理。从这个角度讲，所有正面的人脸都是相似的。再比如，固定在汽车前端的照相机所看到的停车标志牌可以看做具有相同的形状和外观图像窗口。

图像窗口为我们提供了一个物体识别的方法，在识别时，可以摘要所有具有某特定形状的图像窗口，通过检验这些图像窗口，判断是不是有我们已知的某种物体存在。如果物体的大小未知，可以在不同的尺度上搜索图像窗口；如果物体的偏转方向未知，可以在不同的方向上搜索图像窗口；以此类推。这种方法通常被称做模板匹配。运用模板匹配方法，某些类物体的识别问题可以得到很好的结果，例如人脸识别和道路标志的识别。另一方面，虽然某些物体无法通过简单的模板匹配直接进行识别（例如人的识别，因为用来表征一个人的特征所需的图像窗口的数量实在是太多了），但事实证明，在这种情况下，考察不同模板间的关系仍然是一种行之有效的识别方法。在第 23 章，我们将对这一问题展开讨论。

模板匹配的关键问题是如何建立一个有效的检验图像窗口的方法，它可以告诉我们一个椭圆形的图像窗口究竟是不是人脸。通常，检验方法是从大量的样本中归纳出来的，称为分类器。分类器的输入是一个样本的特征集合，而输出是这个样本所属类别的标号。这一章中，我们将讨论一些设计分类器的方法，同时给出在视觉领域中运用这些方法的例子。我们首先介绍一些基本概念和一些常用术语（22.1 节）；接着给出两个基于直方图设计的有效分类器（22.2 节）；为了得到更复杂的分类器，需要选择一些提取样本特征的方法，所以给出两个特征提取的方法（22.3 节）；最后，介绍两种在视觉领域常用的分类方法，人工神经元网络（22.4 节）和支持向量机（22.5 节）。

22.1　分类器

所谓分类器，是根据一些类别已知样本所建立的一个规则，运用这个规则，可以赋予一个类别未知样本相应的类别标号。一般的问题可描述如下，已知一个训练样本集（$\mathbf{x}_i$，y_i）；$\boldsymbol{x}_i$ 是一个样本特征的集合，通常表示为向量的形式，向量中的每一个元素是对样本某方面特征的一个度量值，y_i 是样本所属类别的标号。另一方面，如果把一个类别的样本错误地归为其他类别会造成相对损失，这些相对损失也是已知的。我们要做的是建立一条分类规则，对任何一个类别未知样本，我们可以对它的特征向量 $\boldsymbol{x}$ 应用这条规则以判断样本所属的类别。

错分类的损失会影响分类规则的建立，在 22.1.1 节中，我们将讨论这个问题。对一般的分类器来说，判断一个样本属于不同类别的概率通常是一个关键的问题，在 22.1.2 节中，我们将讨论一些在通常情况下建立有效分类器的方法。最后，在 22.1.5 节中，我们将讨论如何评价一个分类器。

22.1.1　基于损失的决策

分类规则的选择必须依赖于错误分类所造成的损失。例如，医生诊断的过程实际上可以看做一个分类过程，他们告诉患者得了哪种病。医生可能把一个健康的人诊断为患了重病，也可能把一个重病患者诊断为没病，这两种情况都属于错误分类，显然，后者可能造成的损失要比前者严重得多。因此，对一个医生来说，当他不能确定求医者是否生病的时候，他应该更倾向于诊断求医者生了病。

不同的错误分类造成的损失也是不同的。通常，我们把错误分类表示成形如 $i \to j$ 的式子，它表示把第 i 类的样本错归为第 j 类。每个错分类都会有一定的代价，也就是所谓的损失。因此，我们可以定义损失函数 $L(i \to j)$，表示把第 i 类样本错归为第 j 类时所造成的损失。由于正确的分类不应该有任何损失，所以 $L(i \to i)$ 一定为 0，而在其他情况下，损失函数取值一定为正。

下面我们引入风险函数，对特定的分类策略来说，风险函数是应用这种分类策略所造成的损失的数学期望，而总风险值是运用某个分类器所造成的总损失的数学期望值。如果只有两个类别，运用策略 s 造成的总风险值为

$$R(s) = Pr\{1 \to 2 \mid \text{using } s\} L(1 \to 2) + Pr\{2 \to 1 \mid \text{using } s\} L(2 \to 1)$$

我们希望得到的分类策略应使得总风险值最小。

基于最小风险的两分类器　假设有一个两分类问题，损失函数已知。在这种情况下，在状态空间中存在一个分界面。这个分界面被称为决策面，状态空间中在决策面一侧的所有点可被认为是一类，而在决策面另一侧的所有点可被认为是另一类。

有一个窍门可以帮助我们确定决策面的位置。对于一个最佳分类器，位于其决策面上的点无论被归为第一类还是第二类，所造成的损失的数学期望值应该是相等的，否则，我们可以通过平移决策面得到一个更好的分类器。换句话说，决策面上的样本点可以被归为任意一类，两种分类方法造成的期望损失相同。

把状态空间中的一个点 $\boldsymbol{x}$ 归为第一类的风险是

$$\begin{aligned} P\{\text{class is } 2 \mid \boldsymbol{x}\} L(2 \to 1) + P\{\text{class is } 1 \mid \boldsymbol{x}\} L(1 \to 1) &= P\{\text{class is } 2 \mid \boldsymbol{x}\} L(2 \to 1) + 0 \\ &= p(2 \mid \boldsymbol{x}) L(2 \to 1) \end{aligned}$$

类似地，把 $\boldsymbol{x}$ 归为第二类的风险是

$$P\{\text{class is } 1 \mid \boldsymbol{x}\} L(1 \to 2) = p(1 \mid \boldsymbol{x}) L(1 \to 2)$$

在决策面上，这两个风险值相等，因此决策面由所有满足下式的 $\boldsymbol{x}$ 组成

$$p(1 \mid \boldsymbol{x}) L(1 \to 2) = p(2 \mid \boldsymbol{x}) L(2 \to 1)$$

利用贝叶斯公式，可以得到决策面方程的一个更常用的形式

$$\frac{p(\boldsymbol{x} \mid 1) p(1)}{p(\boldsymbol{x})} L(1 \to 2) = \frac{p(\boldsymbol{x} \mid 2) p(2)}{p(\boldsymbol{x})} L(2 \to 1)$$

约去等式两边的分母，可得到下式

$$p(\boldsymbol{x} \mid 1) p(1) L(1 \to 2) = p(\boldsymbol{x} \mid 2) p(2) L(2 \to 1)$$

这个等式确定了决策面上所有的点 $\boldsymbol{x}$，下面我们要做的是对不在决策面上的点进行分类。

对于不在决策面上的点，我们要使分类后的风险最小。前面提到，把决策空间中的一点 $\boldsymbol{x}$ 归为第二类的风险是

$$p(1 \mid \boldsymbol{x})L(1 \rightarrow 2)$$

归为第一类的风险类似。这意味着，如果

$$p(1 \mid \boldsymbol{x})L(1 \rightarrow 2) > p(2 \mid \boldsymbol{x})L(2 \rightarrow 1)$$

则应把 $\boldsymbol{x}$ 归为第一类，反之，如果

$$p(1 \mid \boldsymbol{x})L(1 \rightarrow 2) < p(2 \mid \boldsymbol{x})L(2 \rightarrow 1)$$

则应把 $\boldsymbol{x}$ 归为第二类。

多类别分类器 从现在开始，我们假设 $L(i \rightarrow j)$ 在 $i = j$ 的时候为 0，其他时候为 1，也就是说，各种错分类的损失相等。在一些识别问题中，我们可能选择拒识，就是不把样本归为任何一类中。这种选择同样会造成一定的损失，我们假设这样的损失为 d，且满足 $d < 1$（如果 $d > 1$，则我们永远不会选择拒识）。

对于上面定义的损失函数，最优分类器被称为贝叶斯分类器，算法 22.1 给出了贝叶斯分类算法。采用贝叶斯分类器带来的风险被称为贝叶斯风险，这是我们采用任何一种分类器所能得到的最小风险。除了在少数情况下我们可以直接给出分类规则外，一般我们很难构造一个贝叶斯分类器，因为我们无法精确地确定贝叶斯分类器中需要的几个概率。判断分类器是否有效的一个方法是考察这个分类器造成的风险随着样本数量的增加如何变化（举例来说，我们可能希望随着样本数量的增大，分类器的风险值以概率收敛于贝叶斯风险）。贝叶斯风险一般不为 0（如图 22.1 所示）。

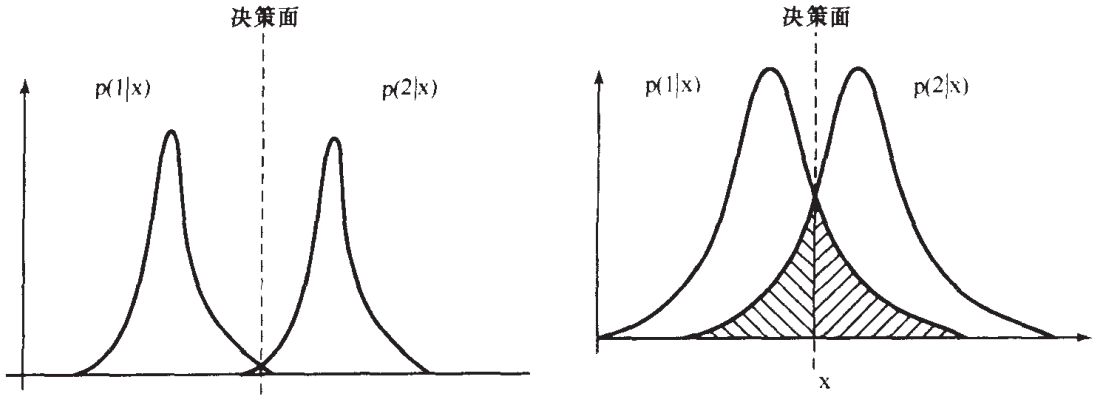


图 22.1 上图描述了一个典型的二分类问题。图上画出了两个类的后验概率 $p(\text{class} \mid x)$，在假设 $L(1 \rightarrow 2) = L(2 \rightarrow 1)$ 的条件下，我们能够得出图中虚线所示的分类器决策面。在上述条件下，贝叶斯分类器的风险是第一类的后验概率在第二类决策域的积分值与第二类后验概率在第一类决策域的积分值 的和（图中阴影部分所示）。对于左图的情况，两类的区分度比较好，因此贝叶斯分类风险也比较小；对于右图的情况，贝叶斯分类风险相对比较大[译者注：（请读者注意此图实际上是错误的，因为在两类 情况下应有 $P(1 \mid X)tP(2 \mid X) = 1$，而此图显然不满足该式，但作者要说明的意思没有错。]

算法 22.1 贝叶斯分类器利用样本属于某一类别的后验概率，以及包括拒识在内的损失函数对样本进行分类。

对下面的损失函数 （未完待续）

（续）

$$L(i \to j) = \begin{cases} 1 & i \neq j \\ 0 & i = j \\ d < 1 & \text{无决策} \end{cases}$$

最佳的分类策略是

- 如果对任意 i 不等于 k，满足 $Pr\{k|\mathbf{x}\} > Pr\{i|\mathbf{x}\}$，并且 $Pr\{k|\mathbf{x}\} > 1-d$，则判断样本属于第 k 类
- 如果存在 $k_1 \cdots k_j$ 以及对任意 i 不等于 $k_1 \cdots k_j$，满足 $Pr\{k_1|\mathbf{x}\} = Pr\{k_2|\mathbf{x}\} = \cdots = Pr\{k_j|\mathbf{x}\} > Pr\{i|\mathbf{x}\}$，则在相等概率的条件下，从 $k_1 \cdots k_j$ 中随机选择一个作为样本的类别
- 如果对于所有的 k，有 $Pr\{k|\mathbf{x}\} \leqslant 1-d$，则选择拒识。

22.1.2　分类器设计概述

通常，我们无法准确知道类条件概率密度 $Pr\{\mathbf{x}|k\}$ 和先验概率 $Pr\{k\}$，因此无法利用贝叶斯公式计算后验概率 $Pr\{\mathbf{x}|k\}$。那么如何从训练样本集设计一个分类器呢？下面我们给出两种比较普遍的方法。

- **参数化概率模型：**我们可以根据训练样本集建立一个概率模型（既可以是后验概率模型，也可以是似然比模型）。有很多方法可以完成这一工作，在后面的章节中，会介绍其中一些方法。一个最简单的情况是，假定类条件概率密度函数服从某种分布（如正态分布）。在这种情况下，需要根据训练样本集估计这种分布的参数（对正态分布来说，需要确定均值 μ 和协方差矩阵 Σ），并根据这些参数的估计值确定贝叶斯分类器。这种方法称为插件分类器（见 22.1.3 节），它涉及到一些参数化概率密度模型和参数估计的方法。需要注意的是，即便对概率模型的参数做出了一个很好的估计，可能仍然无法获得一个好的分类器，因为我们假定类条件概率密度服从的参数模型可能是不恰当的；另一方面，使用一个对训练数据来说并不准确的描述模型仍然可能得到一个好的分类器（如图 22.2 所示）。通常，我们很难得到参数较少而性能令人满意的模型。目前，一些更复杂的方法（例如将在 22.4 节详细讨论的神经元网络）提供了适应性强的概率密度模型，这类方法可以根据训练样本集自动确定一个适应于该样本集的概率密度模型。
- **直接确定决策面：**一个不准确的概率密度模型也可能构造出一个有效的分类器（如图 22.2所示），这是因为决定分类器性能好坏的因素的是决策面是否准确而不是概率密度模型是否精确（在贝叶斯分类器中，估计概率模型参数是为了确定决策面的位置）。因此，可以忽略概率模型的细节而考虑直接确定一个准确的决策面。这种方法通常是有效的，特别是当我们无法为样本特征找到一个合适的概率模型时，这种方法将显得更加重要。这类方法的一个通常的策略是假设决策面是某种类型的函数，并构造一个极值问题从这一类函数中选出最好的一个函数作为决策面的解析表达式。有时，不同类别的样本在特征空间中是线性可分的，这是一个重要的特殊情形。在这种情况下，决策面是一个超平面，将特征空间中的点带入这个超平面的方程，所有结果为正的点被归为

一类，而所有结果为负的点将被归为另一类，此时，这个超平面将是我们构造分类器的惟一参考(见 22.5 节)。

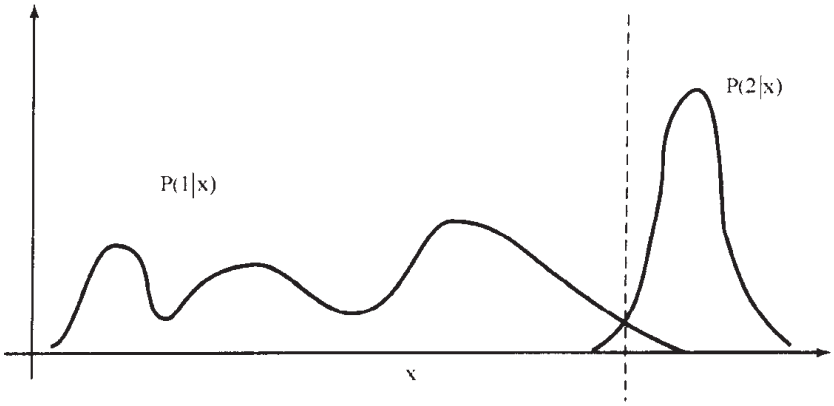


图 22.2 上图给出了两个类的后验概率，最佳决策面如图中虚线所示。尽管用正态分布去近似上述后验概率将产生较大的误差，但是分类器的性能仅由它确定的决策面的位置决定。从图上可以看出，第二类的后验概率近似为一个正态分布，而第一类根据其均值和方差用正态分布近似得到的决策面的位置与贝叶斯决策面位置很接近，因此，在上述条件下，如果我们用正态分布去近似后验概率，也能够得到一个性能较好的分类器(译者注：同图22.1此图也是错的)

22.1.3 举例：正态分布类条件概率密度条件下的插件分类器

当不同类别的样本的类条件概率密度为正态分布时，我们可以构造插件分类器。既可以根据经验确定类先验概率，也可以通过统计训练样本中不同类别的个数来估计类先验概率。接下来我们需要给出类条件概率密度的参数，这是一个参数估计的问题，需要从已知的训练样本集中估计出每个类别 k 条件概率密度的均值 $\boldsymbol{\mu}_k$ 和协方差矩阵 Σ_k。我们知道，当 $\log a > \log b$ 时，表示 $a > b$，为方便计算，我们可以取后验概率的自然对数。下面的算法 22.2 给出了分类器的设计方法。

算法 22.2 当不同类别样本的类条件概率密度为正态分布时，可以用插件分类器对样本进行分类。

假设共有 N 类样本，第 k 类的训练样本有 N_k 个，其中的第 k 类中第 i 个训练样本的特征向量记为 $\boldsymbol{x}_{k,i}$。

对每个类别 k，可以利用下面两个式子估计类条件概率密度的均值和协方差矩阵

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{i=1}^{N_k}\boldsymbol{x}_{k,i}; \qquad \Sigma_k = \frac{1}{N_k - 1}\sum_{i=1}^{N_k}(\boldsymbol{x}_{k,i} - \boldsymbol{\mu}_k)(\boldsymbol{x}_{k,i} - \boldsymbol{\mu}_k)^T$$

为了对特征向量为 $\boldsymbol{x}$ 的样本进行分类，计算 $\delta(\mathbf{x};\boldsymbol{\mu}_k,\Sigma_k)^2 - Pr\{k\} + 1/2\log|\Sigma_k|$，使上式取值最小的 k 就是该样本所属的类别。其中

$$\delta(\mathbf{x};\boldsymbol{\mu}_k,\Sigma_k) = \tfrac{1}{2}\left((\mathbf{x}-\boldsymbol{\mu}_k)^T\Sigma_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right)^{(1/2)}$$

算法中的 $\delta(\mathbf{x};\boldsymbol{\mu}_k,\Sigma_k)$ 称为 Mahalanobis 距离。这个算法在几何上可解释为：在兼顾方差

的前提下，距离待识别样本特征最近的类均值所属的类别就是该样本最可能的类别。说的具体一点，从均值出发，沿方差小的方向上的距离有较大的权值，而沿方差大的方向上的距离有较小的权值。如果假设各个类别的协方差矩阵相同(这样我们需要估计的参数就会减少)，上述算法构造的分类器可以被简化。由于 $\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}$ 对于各个类别是相等的，此时的不同类别的判别函数实际上是一个关于 $\boldsymbol{x}$ 的线性表达式(见习题)。在只有两个类别的情况下，分类的过程可被简化为判断一个关于 $\mathbf{x}$ 的线性表达式的取值是大于 0 还是小于 0。

22.1.4　举例：基于近邻法的非参数分类器

对一个类别未知的样本，可以假设其类别是在特征空间中距离这个样本最近的训练样本的类别，在大多数情况下，这个假设是合理的。近邻法正是基于这一假设来构造分类器。可以用在特征空间中距离待识别样本最近的训练样本所属的类别作为分类结果；也可以在特征空间中找出距离待识别样本最近的几个，然后用这几个训练样本的类别进行投票，以确定待识别样本最终的类别。

我们用 $A(k,l)$ 表示一个近邻分类器，它在特征空间中找到 k 个距离待识别样本最近的训练样本，并用这 k 个训练样本所属的类别进行投票，然后找到票数最高的类别，如果这个类别的票数大于 l，则把待识别样本归为这个类别，否则拒识这个样本。$A(k,0)$ 近邻分类器被称为 k 近邻分类器，而 $A(1,0)$ 近邻分类器则被称为最近邻分类器。

算法 22.3　$A(k,l)$ 近邻分类器根据特征空间中距离待识别样本最近的训练样本所属的类别对待识别样本进行分类

对一个特征向量 $\boldsymbol{x}$

1. 寻找特征空间中距离 $\boldsymbol{x}$ 最近的 k 个训练样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$；
2. 从 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$ 对应的 $y_1,\cdots,y_k$ 中找出数量最多的一个类别 c，并设这个数量为 n；
3. 如果 $n>l$，则把 $\boldsymbol{x}$ 归为类别 c，否则拒识这个样本。

近邻分类器被认为是有效的分类器，可以证明，当训练样本集的数据足够多时，近邻分类器造成的风险与贝叶斯风险的差距有一个不大的上界。随着 k 的增加，近邻分类器造成的风险与贝叶斯风险的差距以 $1/\sqrt{k}$ 的比例减小。通常，我们最多使用 3 近邻法。此外，如果贝叶斯风险值为 0，那么使用 k 近邻分类器的风险也是 0(参考文献中 Devroye, Gyorfi 和 Lugosi 1996 年的一篇论文中详细讨论了上述问题)。

近邻分类器有一些计算上的问题值得注意。首先是如何在特征空间中找到 k 个最近邻点，在高维的特征空间中，这个工作的计算量是比较大的(目前仍需要逐个计算待识别样本到每个训练样本的距离)。其实，训练样本集中的一些样本是多余的，什么样的训练样本是多余的呢？如果去掉某个训练样本，由剩下的训练样本集得到的近邻分类器对状态空间中的每个点仍然能得到相同的识别结果(换句话说决策面的位置不变)，则可以认为这个训练样本是多余的。如果去掉训练样本集中的多余样本，寻找 k 个近邻点的计算量将大大降低。但是，很难按照定义确定一个训练样本是否是多余的。(k,l) 近邻分类器的决策域是一个凸多面体，我们可以利用 Voronoi 图比较容易地确定每个决策域(确定了决策域也就确定了决策平面)，进而可以找出多余的训练样本，但是在高维的情况下，找出多余的训练样本就非常困难了。

另一个构造近邻分类器的难点是如何定义特征空间中的距离。如果特征向量中的每个特征元素描述的是同一个事物,如长度,那么使用欧式距离就可以了。但是,如果特征向量中的特征元素一个表示长度、一个表示颜色、一个表示角度该怎么办呢?此时的一个选择是利用协方差估计来计算出一个类似于 Mahalanobis 距离形式的距离。

22.1.5 分类器性能的评估与改进

显然,分类器对训练样本集是有效的,但是,这并不能说明分类器在整个特征空间上也具有好的性能。下面的例子可能有点滑稽:我们可以构造这样一个分类器,如果待识别样本的特征向量与训练集中某个样本的特征向量相同,则把待识别样本与这个训练样本归为同一类,否则,在所有的类别中随机选取一个作为待识别样本的类别。这个分类器对训练样本集能实现无错误分类,但是对其他的样本集则毫无用处。

造成这个现象的一个原因是分类器受过训练的影响,这种现象有一些不同的叫法,最常用的一个是选择偏见,实际上,它是由分类器必须对训练样本集有效这一事实决定的,而训练样本集仅仅是所有样本的一个子集。过训练描述了选择偏见现象产生的原因,分类器为了适应训练样本集,就很可能需要去适应训练样本集中某些怪癖的数据,而这些数据在其他大多数的样本集合上是很少见的,这就造成了分类器可能对该训练样本集性能很好,而对其他样本集性能很差(通常称这种现象为泛化能力差)。

通常,分类器对训练样本集表现出来的性能,会好于其对测试样本集所表现出来的性能(图 22.18 显示了一个有效分类器在训练样本集和测试样本集上的分类误差)。过训练会造成分类器对训练集的性能和其对测试集的性能之间有较大的差距,这就给我们评价一个分类器的性能带来了难题。下面介绍两种较为常用的方法:一种是隐藏一部分训练样本以测试分类器的性能(后面我们将介绍这种方法);另一种是利用理论的方法确定分类器错误率的上界(见参考文献中 Vapnik,1996 年或 1998 年的论文)。

用交叉验证方法估计总风险 如果我们把训练样本集分为两个子集,则可以用其中的一个子集构造分类器,而用另一个子集检验分类器的性能,这样就可以对分类器的期望风险做出一个直接的估计。这种方法会造成对训练样本的浪费,特别是在只有少数训练样本的时候,此外,这种方法还可能构造出一个性能较差的分类器,这是因为当训练样本很少的时候,分类器在训练集和测试集上的性能差距可能不十分明显。通常,我们会对训练样本集的任何一个可能的划分计算出分类器的风险,用所有这些风险的平均值作为对总风险值的一个估计。这个方法称为交叉验证方法,它可以帮助我们估计分类器可能表现出的性能,但是计算量比较大。

算法 22.4 交叉验证

选择训练样本集的某些子集,例如,可以使每个子集只包含一个训练样本。

对选择的每一个子集,可以根据除去这个子集的所有训练。样本构造分类器,并在这个子集上考察分类器的分类误差(或风险)。

把前面得出的所有风险值取平均,就可以估计出用整个训练集训练出的分类器的风险。

交叉验证方法的一个常用形式是每次只从训练样本集中排除一个样本,通常称为 leave-

one-out 交叉验证。分类误差的估计值通常是通过求平均值得到的，但是也可以通过其他一些更复杂的方法得到(见参考文献中 Ripley，1996 年的论文)。在这里我们并不从数学上严格地讨论交叉验证方法的优劣，但应该指出，交叉验证方法有时对训练集发生的微小变化是比较敏感的。如果一个分类器在交叉验证方法测试下表现出较好的性能，那么样本集中一些较大的子集相互之间将比较相似，这表明从样本集中提取出的概率特征也是比较准确的。

利用自举方法改善分类器性能　通常，利用更多的训练样本可以构造出更好的分类器。但是，在大的训练集上训练分类器是非常困难的，而且随着训练集的增大，分类器性能的改善程度也会逐渐减小。实际上，只有相对较少的样本决定了分类器的性能(我们将在 22.5 节详细讨论这个问题)，这些样本是处于类别边缘、相对较难分类的那部分样本，因为只有这部分样本才真正决定了决策面所处的位置。之所以需要一个大训练集来构造好的分类器，就是希望保证训练样本集中包含较多的此类样本。但是，如果在过大的训练集上进行训练就会得不偿失，因为此时的训练集中包含大量的无用样本。

这里介绍一个窍门，可以使我们避免做无用工作。可以首先在给定训练集的一个小子集上进行训练，然后对剩下的样本进行分类，如果分类错误，则把这个样本添加到训练子集中，重新训练分类器。这是因为这些错误分类的样本包含了决策面的位置信息。这个方法称为自举方法(这个名称可能容易造成混淆，因为有一种同名的随机过程，但是在这里，我们仍然使用这个名字)。

22.2　基于类直方图创建分类器

类直方图可以帮助我们建立类条件概率密度模型。如果用直方图除以总像素数，那么就可以得到一个类条件概率密度的数字表示。随着样本集的不断增大，除以总像素数后生成的直方图将逐渐收敛到类条件概率密度函数，考虑一个极端情况，如果能统计样本特征空间的所有点后生成直方图，则这个直方图实际上就是类条件概率密度函数(见参考文献 Devroye，Gyorfi 和 Lugosi，1996 年的文章以及 Vapnik，1996 年和 1998 年的论文)。如果特征空间的维数较低，这种方法将相当有效(见 22.2.1 节)。而当特征空间的维数较高时，这种方法就不是十分实用了，因为随着特征空间维数的增加，直方图格的数目将迅速增长，除非加入一些独立性假设来降低问题的复杂性(见 22.2.2 节)。

22.2.1　利用分类器识别肤色

肤色识别对于诸如基于手势的人机交互接口等系统是十分有用的。对固定的人种来说，肤色有一个相对固定的颜色范围，因此我们建立一个基于图像像素颜色的分类器来识别肤色。Jones 和 Rehg 在一个人脸图像库上统计出了一个肤色像素点 RGB 颜色值的直方图和一个非肤色像素点 RGB 颜色值的直方图(见参考文献中他们 1999 年的论文)。这两个直方图可以作为类条件概率密度的近似模型。

用 $\boldsymbol{x}$ 表示一个像素点的 RGB 颜色向量。我们可以把颜色空间划分成一些栅格，统计落入每个子区间的肤色像素的百分比作为肤色像素的直方图，这个直方图表示了肤色像素的类条件概率密度 $p(\boldsymbol{x}|\text{skin})$。类似地，可以统计落入每个子区间的非肤色像素的百分比作为非肤色像素的直方图，这个直方图表示了非肤色像素的类条件概率密度 $p(\boldsymbol{x}|\text{not skin})$。此外，还需要知道类先验概率 $p(\text{skin})$和 $p(\text{not skin})$，由于这两个类先验概率之和恒等于 1，因此只需要

知道其中的一个就可以了。知道了类先验概率，就可以构造贝叶斯分类器了(原著在这里给出了$p(\boldsymbol{x}) = p(\boldsymbol{x}|\text{skin}) + p(\boldsymbol{x}|\text{not skin})$这样一个等式，但是这个等式是不正确的，应该为$p(\boldsymbol{x}) = p(\boldsymbol{x}|\text{skin})p(\text{skin}) + p(\boldsymbol{x}|\text{not skin})p(\text{not skin})$，实际上，这里不给出这个等式也不会影响下面的内容，所以在翻译中省略了——译者注)。我们构造的贝叶斯分类器将比较下面两个值

$$\frac{p(\boldsymbol{x} \mid \text{skin})p(\text{skin})}{p(\boldsymbol{x})}L(\text{skin} \rightarrow \text{not skin})$$

$$\frac{p(\boldsymbol{x} \mid \text{not skin})p(\text{not skin})}{p(\boldsymbol{x})}L(\text{not skin} \rightarrow \text{skin})$$

由于$p(\text{skin}|\boldsymbol{x}) = 1 - p(\text{not skin}|\boldsymbol{x})$，所以可以定义如下识别规则

- 如果$p(\text{skin}|\boldsymbol{x}) > \theta$，识别为肤色；
- 如果$p(\text{skin}|\boldsymbol{x}) < \theta$，识别为非肤色；
- 如果$p(\text{skin}|\boldsymbol{x}) = \theta$，随机识别为肤色或非肤色。

其中，θ是一个与$\boldsymbol{x}$无关的表达式，其中包含了相对损失的信息。选择不同的θ值可以得到一组分类器，如果θ取合适值，分类器将是十分有效的(见图22.3)。

图22.3 上图给出了Jones和Rehg皮肤检测器的一些测试图片以及检测结果。结果图片中的黑色像素表示皮肤，白色像素表示背景。上述皮肤检测的过程是非常有效的，通过皮肤检测，可以从图像中找到人脸和人手等我们感兴趣的区域

前面构造的一组分类器中，每个分类器都有其自身的正向错误率和负向错误率，这两个错误率可以看做是θ的函数，因此可以绘制出一条错误率随θ变化的曲线，这条曲线可以描述

这组分类器的性能，称为接受操作曲线(ROC)。图示 22.4 显示了利用上述方法得到的肤色识别系统的接受操作曲线。实验证明，接受操作曲线基本上不随选择不同的类先验概率而变化，也就是说，如果选择不同的 $p(\text{skin})$，则可以通过选择其他的一些 θ 值而得到性能相当的分类器，这又给我们提供了估计类先验概率的方法。在实际应用中，通常先任意地取一些 θ 值，然后画出分类器在训练集上的损失随 θ 值变化的曲线，然后选择一个使分类损失最小的 θ 值构造最终的分类器。

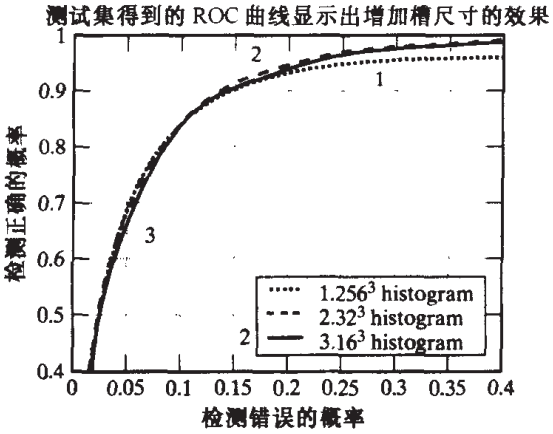


图 22.4　上图给出了 Jones 和 Rehg 皮肤检测器在不同参数 θ 下的接受操作曲线(ROC)。图像纵轴表示皮肤像素的识别率，横轴表示把非皮肤像素识别成皮肤的错误率。一个完美分类器的ROC应该是一条识别率为100% 的水平直线。通过观察不难发现，随着直方图网格数量的变化，ROC曲线的变化并不大

22.2.2　模板响应独立假设条件下的人脸检测

随着样本特征空间维数的增加，直方图模型将变得不再实用，因为直方图的格数随特征空间的维数成指数增长。但有时可以避免这种情况的发生。前面曾经提到，一些独立假设可以减少一个概率模型中需要估计的参数；同样，引入一些独立性假设可以帮助我们有效地降低使用直方图时所需的维数。尽管有人对这种方法不以为然，认为它把问题过度简化了，称它为幼稚的贝叶斯方法，但是这个方法在某些时候的确是行之有效的。Schneiderman 和 Kanade 曾经利用这种方法实现了一个人脸检测系统(见参考文献 Schneiderman 和 Kanade 1998 年的论文)。假设图像中的人脸大小大致在一个固定的范围内，并且形状规则，对正面的人脸，可能是椭圆形或近似的长方形，对侧面的人脸，可能是一些比较复杂的多边形。要做的第一件事情是给人脸部分的图像模式建立一个概率模型，我们希望估计出类概率密度 $P(\text{image pattern} \mid \text{face})$。通常采用产生式模型(产生式模型是分类学中的一个概念，这里指对图像模式进行分类)来估计这个概率密度。因为人脸部分的图像模式多种多样，无法逐个处理，但是把图像模式划分成子集，把每个子集看做一个类别，这样类别数量就大大地减少了，从而避免了繁琐的处理过程。

划分子集比较常用的方法是在一个大的图像模式的集合上进行动态聚类。例如，可以采用 k 均值算法进行动态聚类。聚类后，用每个子集的中心代表这个子集典型形式。我们知道，由于图像存在着噪声，因此相似的图像模式之间可能会存在着一些差异，而上述做法的好处在于它可以抑制这种差异。

有很多模型可以被应用于这种方法，下面介绍一个最简单的实用模型。可以假设：在确定

出现人脸的条件下，不同模式类别是完全独立的，也就是说，需要的类概率密度可以表示为如下的形式

$$
\begin{aligned}
P(\text{image}\mid\text{face}) &= P(\text{label 1 at }(x_1,y_1),\cdots,\text{label }k\text{ at }(x_k,y_k)\mid\text{face})\\
&= P(\text{label 1 at }(x_1,y_1)\mid\text{face})\cdots P(\text{label }k\text{ at }(x_k,y_k)\mid\text{face})
\end{aligned}
$$

其中，每一个 $P(\text{label } k \text{ at } (x_k, y_k) \mid \text{face})$ 都可以通过对大量样本图像标号并构造直方图的方法学习得出，由于这时的直方图是二维的，因此统计直方图的方法是可行的。类似地，可以估计另一个类概率密度 $P(\text{image} \mid \text{no face})$。直到了这两个类概率密度，就可以构造分类器了。事实证明，上述方法是有效的，Schneiderman 和 Kanade 正是利用这个方法实现了人脸检测系统和汽车检测系统(如图 22.5 所示)。

图 22.5 上图给出了 22.2.2 节介绍的人脸检测方法的一些检测结果。不同尺度的图像窗口根据从训练样本中学习出的人脸似然率模型被划分成为正面人脸、侧面人脸及非人脸。图像窗口中的一些子区域被划分成为一些类别，这些类别是从训练样本中学习得到的。人脸似然率模型认为，一个图像窗口中处于不同位置的子区域属于哪个类别是相互独立的，模型最终给出了图像窗口对各个类别的后验概率，我们可以根据这个后验概率判断图像窗口属于哪个类别

22.3　特征选择

假设需要对一个图像块进行分类，应该如何提取这个图像块的特征呢？一个简单的办法是提取图像块上所有像素的颜色值作为特征，这样可以把尽可能多的图像信息告诉分类器，但是这样做也会带来一些问题。

首先，在高维特征空间中，需要大量的样本才能比较准确地估计出类概率密度的模型。例如，一张低分辨率的人脸图像的结构是非常简单的，可以粗略地认为是在一个光滑的背景上有几个深色的条（如眉毛和眼睛）和几个浅色的条（如鼻子和前额）。但是，如果处理的是高分辨率的人脸图像，上述简单的结构将不再适用，因为需要考虑不同图像肤色纹理间的差别，这就需要提供大量的样本数据。因此，有时会采用一些方法（如人为降低图像分辨率）使样本结构变得简单，这样构造的样本特征空间将更加实用。

其次，有时可能事先掌握了关于样本模式的某些知识，如，光照对物体图像的影响是已知的。如果要求分类器使用样本对我们已经掌握的知识再次建模，则会造成对样本的浪费。因此我们希望所使用的特征与已知的知识是一致的，这需要对样本进行一些预处理（例如，除去样本图像由于光照的不同而产生的差异），或者选择一些不随某些变换而改变的特征（例如统一样本图像的大小）。

特征选择和模型选择（见 16.3 节）之间有相似之处。模型选择中，我们希望建立一个模型尽可能好地反映数据集的特征；而在特征选择中，希望找到一个描述数据集样本的特征以帮助实现对数据集的分类。二者在形式上差别不大（可以把特征集合看做是模型，而把分类器看做是模型检验器），目的也是相同的。下面讨论两种常用的提取线性特征的方法，所谓线性特征，就是作用于原始特征的一个线性函数。

22.3.1　主分量分析

特征选择的一个主要目的是找出能准确反映原始特征集的最小特征集，这个最小特征集是由具体的应用决定的。其中，一种常用的特征选择方式是使得特征集尽可能多地反映出原始特征集的变化，因为如果特征集中的某个特征值可能根据其他的特征值准确地推导出来，那么这个特征值就是多余的。因此，如果希望从特征集中去掉一个特征，一个最好的选择就是去掉那个最能根据其他特征准确推导出来的一个。我们不仅可以从原始特征集中去掉特征，还可以用原始特征的函数构造新特征。

在主分量分析方法中，新特征是原始特征的线性函数。需要从一个数据点集中建立一个低维的特征子空间，这个子空间能最好地反映出这个数据点集相对于它们的平均值的差异。这个方法（有时也称为 K-L 变换）是统计模式识别中一个经典的特征提取方法（见参考文献中 Duda 和 Hart，1973 年论文，Oja，1983 年论文，Fukunaga，1990 年论文）。

假设有一个建立在 d 维欧式空间中$\mathbb{R}^d$ 的特征点集，其中有 n 个特征向量 $\boldsymbol{x}_i$（$i=1,\cdots,n$）。它们的均值向量为 $\boldsymbol{\mu}$（这个均值向量就是 n 个特征向量的重心），协方差矩阵为 Σ。我们以均值向量 $\boldsymbol{\mu}$ 为新的原点，研究特征点相对于均值的偏差（$\boldsymbol{x}_i-\boldsymbol{\mu}$）。

我们提取的新特征是原始特征值的线性组合，实际上它是上述特征点相对于均值的偏差

在不同方向上投影的结果。设单位向量 $\boldsymbol{v}$ 表示原特征空间的一个方向，如果把这个方向看做一个新特征，那么原第 i 个特征向量 $\boldsymbol{x}_i$ 在这个新特征上的值为 $v(\boldsymbol{x}_i)=\boldsymbol{v}^T(\boldsymbol{x}_i-\boldsymbol{\mu})$。一个好的特征选择应尽可能多地反映原始特征的差异，由于 v 的均值为 0，因此 v 的方差应为

$$
\begin{aligned}
\mathrm{var}(v) &= \frac{1}{n-1}\sum_{i=1}^{n} v(\boldsymbol{x}_i)v(\boldsymbol{x}_i)^T \\
&= \frac{1}{n}\sum_{i=1}^{n-1} \boldsymbol{v}^T(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{v}^T(\boldsymbol{x}_i-\boldsymbol{\mu}))^T \\
&= \boldsymbol{v}^T\left\{\sum_{i=1}^{n-1}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\right\}\boldsymbol{v} \\
&= \boldsymbol{v}^T\Sigma\boldsymbol{v}
\end{aligned}
$$

现在我们希望在满足 $\boldsymbol{v}^T\boldsymbol{v}=1$ 的条件下，使 $\boldsymbol{v}^T\Sigma\boldsymbol{v}$ 取值最大。这是一个特征值问题，矩阵 Σ 最大的特征值所对应的特征向量就是我们要求的 $\boldsymbol{v}$。如果把原特征向量点 $\boldsymbol{x}_i$ 投影到垂直于 $\boldsymbol{v}$ 的空间上，我们可以得到 $d-1$ 维的向量。此时，方差最大的方向应该是矩阵 Σ 第二大特征值所对应的特征向量，以此类推。

矩阵Σ有 d 个特征值，按从大到小的顺序排序，排序后的特征值对应的特征向量为 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_d$，其中 $\boldsymbol{v}_1$ 是最大特征值对应的特征向量，这些特征向量构成了一个新的特征集，这个特征集有如下特征：

- 每个特征是独立的(因为 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_d$ 都是正交的)
- 由前 k 个特征构成的特征集$\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k\}$是保留原始特征集最大差异的 k 维特征集

需要注意的是：主分量既可能较好地反映原始特征，也可能相反，这与源数据有关(见图 22.6、图 22.7 和图 22.9)。

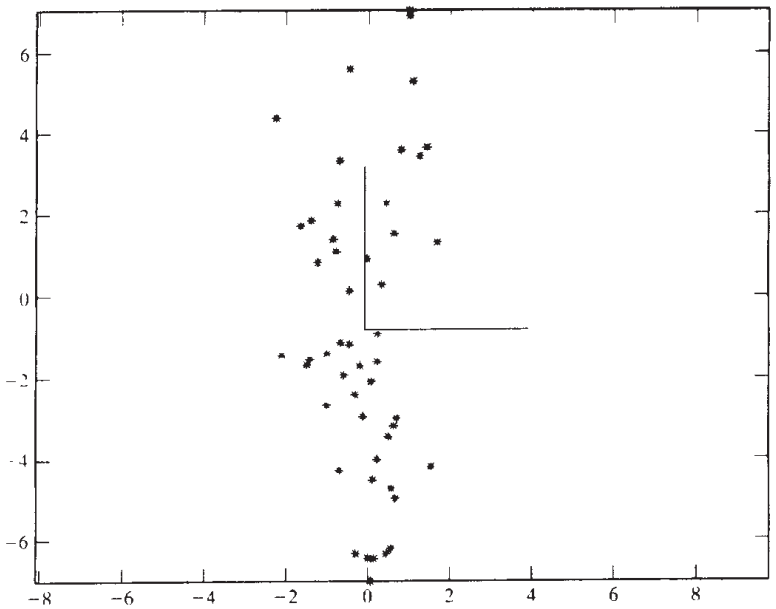


图 22.6 数据集在经过 PCA 分解后通常能够很好地被重构。上图中的两个轴表示数据集经 PCA 分解后得到的两个主分量，其中竖直方向的轴是第一维主分量，在这个方向上数据分布的方差比较大

22.3.2　基于主分量分析的身份验证方法

人类能记住并辨别很多人的相貌，如果能在计算机上实现模拟人类这一功能的自动系统，将在人机交互和安全系统等诸多领域有广泛的应用前景。Kanade 于 1973 年开发了第一套全自动的人脸识别系统，目前，人们提出了很多新的方法以解决这个领域的一些特定的问题（如固定头的转向、固定表情，等等，见参考文献中 Chellappa，Wilson 和 Sirohey，1995 年的一篇综述文章）。人脸检测方法基本上可以分为两类。一类是基于特征的匹配：通过提取和检验图像中人脸特征（如眼睛、鼻子、嘴唇）的几何参数（如长和宽）和相互关系（如相对位置）以检测人脸。另一类是基于模板的匹配：通过直接比较图像和不同标准人脸的亮度关系以检测人脸（参考文献中 Brunelli 和 Poggio，1993 年的论文讨论并比较了上述两种方法）。假设我们已经检测到了一个人脸，现在要确定这个人是谁。如果有一个可用的空间坐标系统，那么基于 PCA 的方法可以简单而有效地解决这一问题。

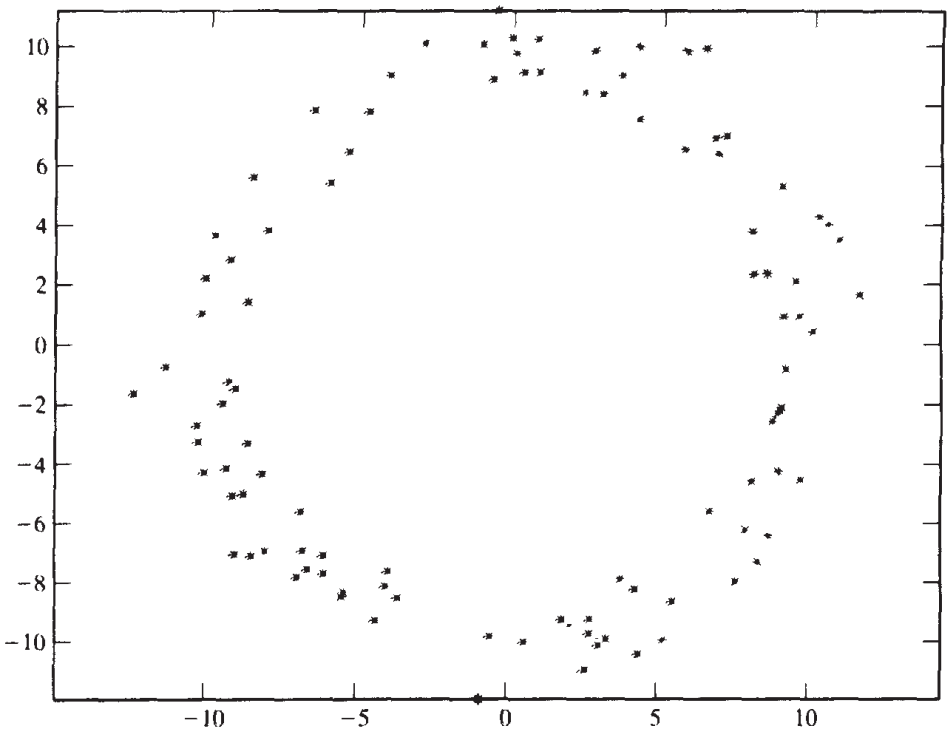


图 22.7　并非所有的数据集都能够很好地被 PCA 重构。上图所示数据集的主分量就不是很稳定，因为数据源在各个方向上的方差大致相同。从这样一个数据源采集的不同组数据集的主分量可能有较大的差异。这体现了PCA分析的一个难点——如何把圆结构的数据集投影到它的主要特征方向上

特征图像　人类可以迅速地辨认出很多人的面貌，Sirovitch 和 Kirby（见参考文献中他们 1987 年的论文）认为人类的视觉系统仅用了很少的特征去存储和检索人的相貌。因此，他们采用 PCA 方法对人脸图像进行压缩。实际上，PCA 方法帮助我们用 p 维（$p << d$）向量来表示 d 维的样本特征 $\boldsymbol{s}_i \in \mathbb{R}^d (i = 1, \cdots, n)$，而 p 维向量是原样本特征在 p 个单位向量上 $\boldsymbol{u}_j (j = 1, \cdots, p)$ 上的投影。采用原始样本特征时，需要 nd 个存储单位，而采用 PCA 后需要 $np + pd = (n + d)p$ 个存储单位（其中，np 表示 n 个新样本特征的存储空间，p_d 表示 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p$ 的存储空间）。由于向

量 $\boldsymbol{u}_j$ 与原样本特征有着同样的维数，Sirovitch 和 Kirby 称其为特征图像。实验表明，40 张特征图像能够重构出包含 115 张 128×128 的人脸图像库，误差率为 3%。如果用上面的实验结果计算存储空间（$p = 40, n = 115, d = 128 \times 128$），则采用新的样本特征后存储空间比原来减少了多一半。Sirovich 和 Kirby 还给出，对于不在前面图像库中且光照条件不利的人脸，PCA 方法重构的错误率也仅有 8%，这显示了 PCA 方法优越的可扩展性能。

算法 22.5 主分量分析建立了一个独立（译者注：应是"线性无关"）的线性特征集，并且反映原始数据集尽可能多的差异。假设我们有一个建立在 d 维欧式空间中 $\mathbb{R}^d$ 的数据点集，其中有 n 个特征向量 $\boldsymbol{x}_i (i = 1, \cdots, n)$。令：

$$\boldsymbol{\mu} = \frac{1}{n} \sum_i \boldsymbol{x}_i$$

$$\Sigma = \frac{1}{n-1} \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$$

矩阵 Σ 有 d 个特征值，按从大到小的顺序排序，排序后的特征值对应的单位特征向量为 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_d$，其中 $\boldsymbol{v}_1$ 是最大特征值对应的单位特征向量，这些特征向量构成了一个新的特征集，这个特征集有如下特征：

- 每个特征是独立（译者注：应为线性无关）的
- 由前 k 个特征构成的特征集 $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ 是保留原始数据集最大差异的 k 维特征集

特征脸 虽然 Sirovitch 和 Kirby 在论文中提到了利用特征图像进行人脸识别的观点，但是他们并没有给出一个具体的识别算法。这一工作是由 Turk 和 Pentland（见参考文献中他们 1991*a* 年的论文）完成的，他们利用这一章前面提到的近邻识别算法实现了一个完整的人脸识别系统，在他们的工作中，把"特征图像"改称为特征脸。

Turk 和 Pentland 的识别算法分为以下几个步骤：

离线阶段：

1. 建立一个包括 m 个人的人脸图像库，每个人有多张图像，包括不同表情、不同位置和不同光照条件；
2. 计算特征脸 $\boldsymbol{u}_i (i = 1, \cdots, p)$；
3. 设由 p 张特征脸 $\boldsymbol{u}_i (i = 1, \cdots, p)$ 张成的 p 维空间为 V_p，对于图像库中的每个人，提取它在 V_p 上的新特征 $\boldsymbol{w}_j (j = 1, \cdots, m)$；

在线阶段：

4. 对于系统输入的一张新图像 $\boldsymbol{t}$，计算它在 V_p 上的投影 $\boldsymbol{w}$；
5. 如果 $d = |\boldsymbol{t} - \boldsymbol{w}|$ 大于某个预先设定的阈值 ε_1，则识别图像不是人脸；
6. 否则，计算 $\boldsymbol{w}$ 与不同 $\boldsymbol{w}_j$ 之间的最近距离为 $d_k = |\boldsymbol{w} - \boldsymbol{w}_k|$，如果 d_k 小于某个预先设定的阈值 ε_2，则识别输入图像为图像库中第 k 个人；
7. 如果 $d < \varepsilon_1$ 且 $d_k \geq \varepsilon_2$，则识别输入图像为不明身份的人，然后，可以选择把输入图像加入人

脸库中，并重新计算特征脸

现在，我们来说明上述算法中第 3 步中图像库中每个人的新样本特征 w_j 是如何计算的。Turk 和 Pentland（1991*a*）的做法是取平均值，也就是说把第 j 个人的多张人脸图像都投影到 V_p 上，然后计算平均值作为 w_j（w_j 反映了第 j 个人在新特征空间中的一个平均位置）。我们也可以不计算平均值而对图像库中所有图像的新特征直接采用近邻算法。

Turk 和 Pentland 在实验中采用的是一个由 16 个人组成的包含 2500 张 128 × 128 图像的图像库，包括三种人脸偏转、三个人脸大小和三种光照条件的各种组合（见图 22.8）。

图 22.8　上图给出了 Turk 和 Pentland（1991*b*）的实验中的一部分图像数据。实际的图像集中包含了光照、朝向和大小等方面的差异，但是在上图并没有体现出来

表 22.1 给出了定量的识别结果，实验中，训练集由图像库中选取，并保证每个人的样本都出现在训练集中。训练后，对所有的图像库中的图像进行分类。实验结果统计了训练集上的标准偏差，其中，照明条件、人脸大小和人脸的偏转方向三个条件是相互独立的。

表 22.1　识别结果，所改变的实验条件是 ε_1 的值（例如 $\varepsilon_1 = \infty$ 是强迫给出分类结果），ε_1 值较低（或高）得到更准确（或更不准确）的识别结果，但未知识别结果率更高（或更低）

实　验	正确/未知识别的百分比		
实验条件	光照条件	人脸朝向	人脸尺寸变化
强制分类	96/0	85/0	64/0
强制 100% 准确	100/19	100/39	100/60
强制 20% 拒识率	100/2	94/20	74/20

22.3.3　典型变量

主分量分析建立了在特定维数下最能表达原高维数据变化的新的特征集，但是它并不能保证这个特征集能帮助我们实现有效分类。例如，如图 22.9 所示的数据集，用第一个主分量得不到一个好的分类效果，而用第二个主分量虽然不能很好地反映原始数据的差异，但能实现

一个不错的分类器。

能够明显地反映出不同类别样本间差异的线性特征称为典型变量(canonical variates)。如果有一个样本数据集 $\boldsymbol{x}_i, i \in \{1, \cdots, n\}$,如何找出典型变量。设原始样本特征是 p 维的,其中包括 g 个不同的类别,且第 j 个类别的样本特征均值为 $\boldsymbol{\mu}_j$,令 $\bar{\boldsymbol{\mu}}$ 为类平均值的平均值,即:

$$\bar{\boldsymbol{\mu}} = \frac{1}{g}\sum_{j=1}^{g}\boldsymbol{\mu}_j$$

再令:

$$\mathcal{B} = \frac{1}{g-1}\sum_{j=1}^{g}(\boldsymbol{\mu}_j - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_j - \bar{\boldsymbol{\mu}})^T$$

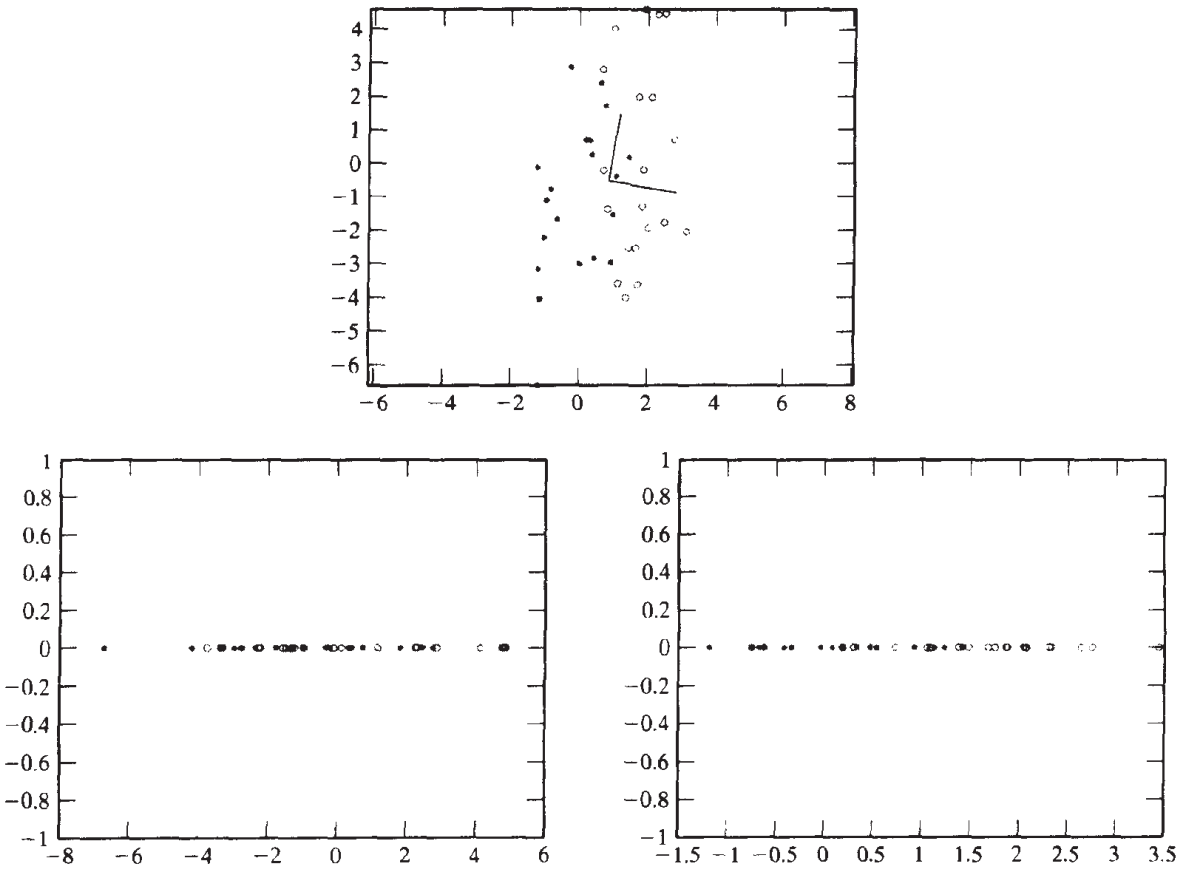


图 22.9 主分量分析并不考虑数据集中可能包含多类样本,这可能给我们构造分类器带来很大的麻烦。对于分类器来说,我们希望提取出的特征集特征个数较少,同时不同类别样本在这个特征集上的差异比较明显。上图包含了两类样本,一类用圆圈表示,另一类用星号表示。PCA建议的投影方向是图中竖直轴的方向,因为数据集在这个方向上投影的方差最大。但是利用主分量方向并不能很好地区分两个类别,因为数据集在主分量方向上的投影出现了不同程度的交叠。左下图显示了样本在第一主分量方向上的投影,样本在这个方向上的分布具有较大的方差,但是这个方向对两类样本的分类能力较差。右下图显示了样本在第二主分量方向上的投影,虽然样本在这个方向上分布的方差较小,但是这个方向对两类样本的分类能力却相对好一些

其中,$\mathcal{B}$给出了类平均值之间的方差。以下考虑最简单的情况,假设每个类别的协方差矩阵是

相同的，都为 Σ，且是满秩。我们希望找到这样的一些坐标轴，相同类别的特征点在这些坐标轴上的投影都会聚在一起，而不同类别的特征点在这些的坐标轴上的投影则是分离的。这意味着需要寻找这样一些特征，使得在新特征下，类平均值间的方差与每一类样本间的方差的比值尽可能大。类平均值间的方差通常称为类间离散度，而类样本间的方差通常称为类内离散度。

我们要找的是关于原始特征的一个线性函数，考察下面的函数

$$v(\boldsymbol{x}) = \boldsymbol{v}^T \boldsymbol{x}$$

我们希望找到一个 $\boldsymbol{v}_1$ 使得类间离散度与类内离散度的比值最大。通过与主分量分析中类似的推导，可以通过最大化下式实现上述目的：

$$\frac{\boldsymbol{v}_1^T \mathcal{B} \boldsymbol{v}_1}{\boldsymbol{v}_1^T \Sigma \boldsymbol{v}_1}$$

这个问题等同于在满足 $\boldsymbol{v}_1^T \Sigma \boldsymbol{v}_1 = 1$ 的条件下，求 $\boldsymbol{v}_1^T \mathcal{B} \boldsymbol{v}_1$ 的最大值，进一步可以推导出，$\boldsymbol{v}_1$ 应满足下式：

$$\mathcal{B}\boldsymbol{v}_1 + \lambda \Sigma \boldsymbol{v}_1 = 0$$

其中，λ 是常数。在 Σ 是满秩的条件下，这被称为广义的特征值问题，此时的 $\boldsymbol{v}_1$ 是矩阵 $\Sigma^{-1}\mathcal{B}$ 最大特征值所对应的单位特征向量（如果Σ不是满秩，我们可以求助于一些常用的数值计算软件）。

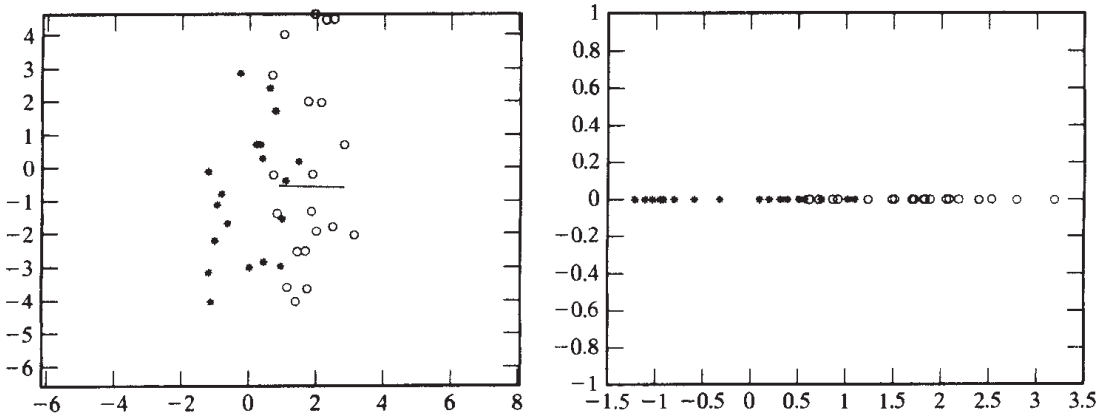


图 22.10　典型变量利用数据集中的所有样本去估计一个好的线性特征集，数据集在特征集每个特征方向上的投影都能尽可能地把不同类别的样本分开。左图显示了图22.9中的数据集以及最优典型变量的方向。右图显示了数据集在这个方向上的投影，可以看出，投影后不同类别的样本能够被较好地分离

对于其他的特征 $\boldsymbol{v}_l$，$2 \leqslant l \leqslant p$，希望能够满足同样的规则，并且保证特征 $\boldsymbol{v}_l$ 之间是独立的，这些特征都是矩阵 $\Sigma^{-1}\mathcal{B}$的单位特征向量，对应的特征值则给出了原始样本特征集在相应方向上的方差（不同的方差也是独立的）。我们可以选择矩阵 $\Sigma^{-1}\mathcal{B}$前 m 个最大的特征值对应的特征向量作为新的特征集，这样的特征集在指定的维数下最好地保持了类间分离的特性。虽然不能保证用这样的特征构造的分类器能够实现错误率最小，但它的确在保留类间结构的前提下为降低特征维数提供了可能（如图 22.11 所示）。有关这部分更详细的讨论和具体例子，请见参考文献中 McLachlan 和 Krishnan (1996)的论文以及 Ripley (1996)的论文。

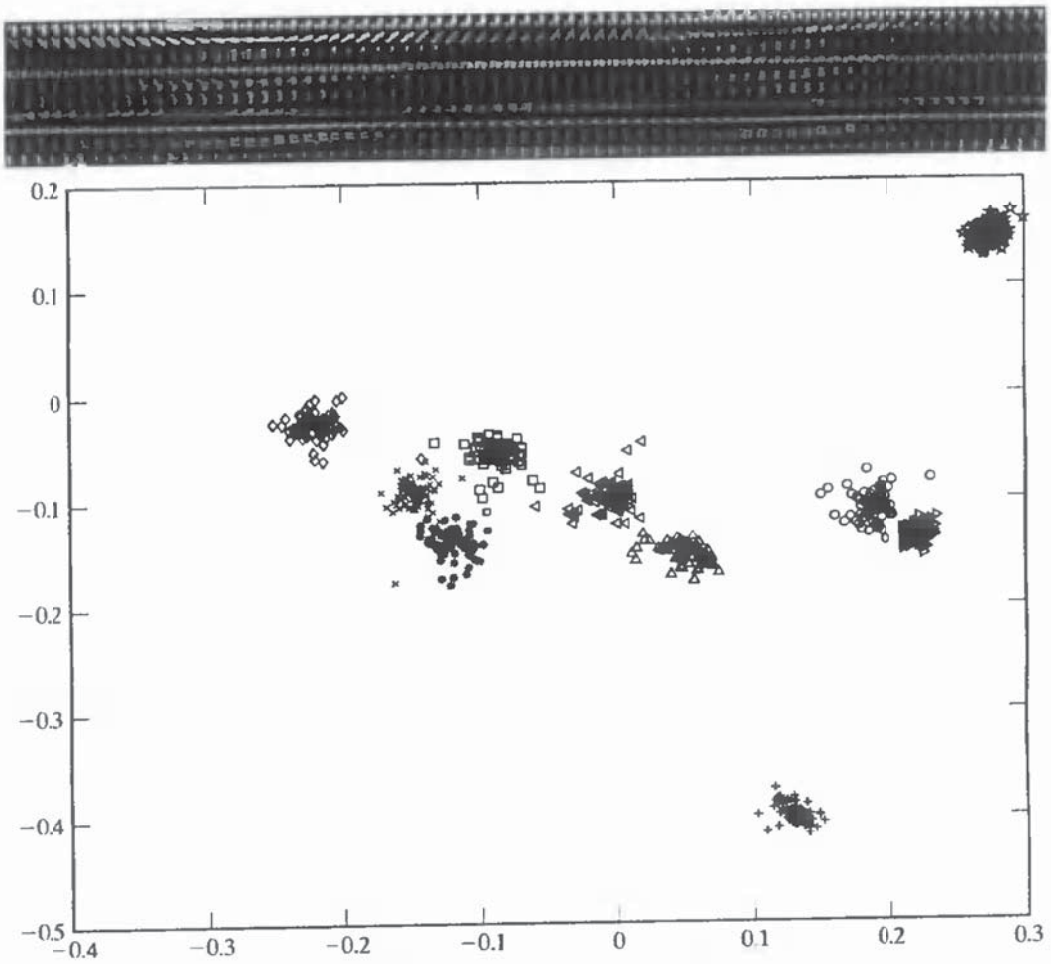


图 22.11 典型变量分析对于很多简单的模板匹配问题是十分有效的。上图显示了黑色背景下不同姿态的10种物体，这些图像是Nene和Nayar著名的COIL数据库中一部分图像平滑并重采样之后的结果。COIL数据库可从以下网址获得http://www.cs.columbia.edu/CAVE/research/softlib/coil-20.html，这是一个包含20个物体的数据库，或http://www.cs.columbia.edu/CAVE/research /softlib/coil-100.html，这是一个包含100个物体的数据库）。从上面的图像中找出一个物体是一个相对简单的问题，因为图像背景是固定的，因此不需要进行图像分割。我们用每个类别60幅图像训练典型变量。下图显示了每个类别71幅图像在前两维典型变量空间上的投影，其中每个类别包括60幅训练图像和11幅测试图像。不难发现，不同类别物体在这个空间上被很好地分开了，因此在这个二维典型变量空间上就可以构造一个很好的分类器

算法 22.6 典型变量确定了一组能够使类间尽可能分开的特征。

假设我们有 g 个类别，第 k 个类别有 n_k 个样本，分别是 $\boldsymbol{x}_{k,i}, i \in \{1, \cdots, n_k\}$，其均值为 $\boldsymbol{\mu}_j$。不妨设样本特征是 p 维的。

令 $\overline{\boldsymbol{\mu}}$ 为类平均值的平均值，即：

$$\overline{\boldsymbol{\mu}} = \frac{1}{g}\sum_{j=1}^{g}\boldsymbol{\mu}_j$$

（未完待续）

（续）

再令：

$$\mathcal{B}=\frac{1}{g-1}\sum_{j=1}^{g}(\boldsymbol{\mu}_j-\overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_j-\overline{\boldsymbol{\mu}})^T$$

假设不同类别有相同的协方差矩阵 Σ，这个 Σ 可能是已知的，也可以通过下式估计出来

$$\Sigma=\frac{1}{N-1}\sum_{c=1}^{g}\left\{\sum_{i=1}^{n_c}(\boldsymbol{x}_{c,i}-\boldsymbol{\mu}_c)(\boldsymbol{x}_{c,i}-\boldsymbol{\mu}_c)^T\right\}$$

矩阵 $\Sigma^{-1}\mathcal{B}$ 有 d 个特征值，按从大到小的顺序排列，排列后特征值所对应的单位特征向量为 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_d$，其中 $\boldsymbol{v}_1$ 是最大的特征值对应的单位特征向量。则由 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_d$ 构成的特征集具有以下特征：

- 由 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_d$ 中的前 k 个构成的特征集 $\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k\}$ 是所有 k 维特征集中保持原数据样本集类间可分性最好的一个。

如果不同类别的类协方差矩阵是不相同的，我们同样可以构造典型变量。此时，需要估计出所有样本特征相对于它本类别样本均值偏差（我们在前面提到过这个概念）的一个总体的协方差矩阵 Σ。同样，我们无法保证这样做得到的结果是最优的，但是实验证明，这样方法是行之有效的。

22.4　神经元网络

在多数情况下，简单的参数概率密度和直方图模型都不适用，此时，我们必须采用复杂的概率密度模型（见本节）或在特征空间中直接寻找决策面（见 22.5 节）。

22.4.1　基本思想

神经元网络是一种非常实用的建立概率密度模型的参数化近似技术。神经元网络的作用是逼近出输入向量 $\boldsymbol{x}$ 的一个向量函数 $\boldsymbol{f}$。通常，神经元网络是分层结构的，每层的输出是一个向量，其中的每个元素是输入向量先被仿射函数作用再被非线性函数作用的结果，仿射函数对于不同的结点是不同的，而非线性函数对于所有结点都是相同的，记做 ϕ。可以将输入向量增加一项，并把这一项的值固定为 1，这样做可以把对原输入向量的仿射变换变为对增广后输入向量的线性变换，设每一层的增广输入向量为 $\boldsymbol{u}$，则该层的输出 $\boldsymbol{v}$ 可表示为：

$$\boldsymbol{v}=[\phi(\boldsymbol{w}_1\cdot\boldsymbol{u}),\phi(\boldsymbol{w}_2\cdot\boldsymbol{u}),\cdots\phi(\boldsymbol{w}_n\cdot\boldsymbol{u})]$$

其中，$\boldsymbol{w}_i$ 是可变参数向量，通过调节 $\boldsymbol{w}_i$ 可以改变神经元网络的输出。

前面提到，神经元网络一般分为多层，每层的输入都可以看做是一个增广向量。如果希望通过一个两层网络逼近输入向量 $\boldsymbol{x}$ 的向量函数 $\boldsymbol{g}$，输出可用下式表示

$$\boldsymbol{g}(\boldsymbol{x})\approx\boldsymbol{f}(\boldsymbol{x})=\left[\phi(\boldsymbol{w}_{21}\cdot\boldsymbol{y}),\phi(\boldsymbol{w}_{22}\cdot\boldsymbol{y}),\cdots\phi(\boldsymbol{w}_{2n}\cdot\boldsymbol{y})\right]$$

其中

$$y(z) = [\phi(w_{11} \cdot z), \phi(w_{12} \cdot z), \cdots \phi(w_{1m} \cdot z), 1]$$

$$z(\boldsymbol{x}) = [x_1, x_2, \cdots, x_p, 1]$$

如果某些元素 $\boldsymbol{w}_{1k}$或$\boldsymbol{w}_{2k}$为零,则意味着 $\boldsymbol{y}$ 中的某些元素不会影响到$\boldsymbol{f}(\boldsymbol{x})$,此时第二层称为部分连接层,否则称为全连接层。当然,第一层也可能是松散耦合层或紧密耦合层。参数 n 由输出向量$\boldsymbol{f}$的维数决定,参数 p 则由输入向量$\boldsymbol{x}$ 的维数决定,而参数 m 可以自由选取,并不一定要保证 m 与n 或p 中的一个相等,通常我们选择 m 大于其中的一个。实际应用中,一般使用的是三层网络(更多层的就不太常见了)。神经元网络结构可以用图示方法表示,通常,用圆圈表示变量,用箭头表示可能不为 0 的连接,如图 22.12 所示。

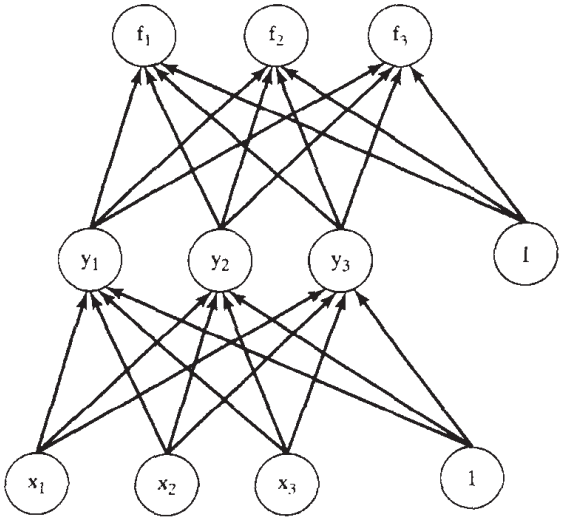


图 22.12 神经元网络通常可以用上述形式的图来描述。图中的每个圆圈表示一个神经元变量,通常圆圈中会有一个标号。网络的层次结构可以从图中表现出来,如上图所示的是一个两层神经元网络,不同层用的变量符号不同,箭头表示在仿射变换中连接两个神经元之间的非零系数。由于上图中任意两个结点都有箭头连接,因此上图所示的网络是全连接的。请注意,图中可能会出现跨层的箭头

选择非线性函数 非线性函数ϕ可以有多种选择。例如,可以选择阈值函数,当输入为正数时,输出为 1;当输入为负数时,输出为 0。阈值函数的响应在一个超平面上从 0 跳变到 1,它反映了输入向量在特征平面中相对于某个超平面的位置。由于阈值函数是不可微的,因此采用阈值函数的神经元网络训练起来是比较困难的。

通常,我们选择输出由 0 到 1 平滑变化的函数作为ϕ,一般选择 S 型函数或挤压函数。其中,S 型函数(译者注:原文中这里为 logistic 函数,但应为 S 型函数)是比较常用的一种,形式如下:

$$\phi(x; \nu) = \frac{e^{x/\nu}}{1 + e^{x/\nu}}$$

其中,ν控制了函数在 $x = 0$ 处变化的快慢。实际上,并不一定要求ϕ的输出在 0 和 1 之间,例如,另一种常用的选择是挤压函数,形式为:

$$\phi(x; \nu, A) = A \tanh(\nu x)$$

它的输出范围是 $-A$ 到 A。图 22.13 描述了上述非线性函数。

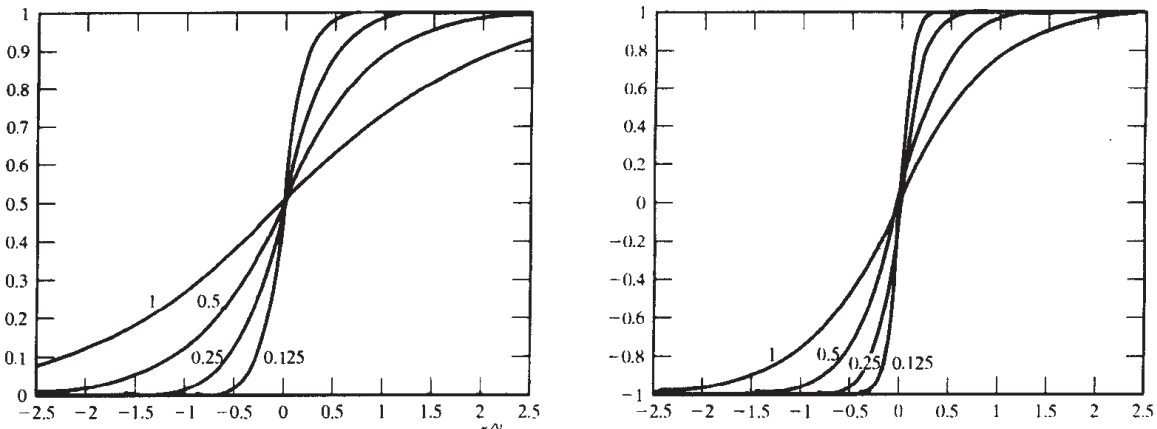


图 22.13　左图显示了型如 $\phi(x;\nu)=\dfrac{e^{x/\nu}}{1+e^{x/\nu}}$，的核函数图像，根据ν值的不同，函数图像也不同。右图显示了型如 $\phi(x;\nu,A)=A\tanh(x/\nu)$ 的核函数，同样，根据ν值的不同，函数图像也不同。一般说来，函数在其中间的部分表现出线性，而在其两边的部分表现出较强的非线性

利用神经元网络构造分类器　为了构造表示 $\boldsymbol{g}(\boldsymbol{x})$ 的神经元网络，需要搜集一些输入样本 $\boldsymbol{x}^e$，然后建立神经元网络，该网络对应 $\boldsymbol{g}$ 的每一维都有一个输出，记为 $\boldsymbol{n}(\boldsymbol{x};\boldsymbol{p})$，其中，$\boldsymbol{p}$ 表示包括所有 $\boldsymbol{w}_{ij}$ 的参数向量。然后，构造期望的输出向量 $\boldsymbol{o}^e$，一般 $\boldsymbol{o}^e=\boldsymbol{g}(\boldsymbol{x}^e)$。我们要做的是找出某个参数向量 $\hat{\boldsymbol{p}}$，使得下式的取值最小

$$Error(\boldsymbol{p})=\left(\frac{1}{2}\right)\sum_e|\boldsymbol{n}(\boldsymbol{x}^e;\boldsymbol{p})-\boldsymbol{o}^e|^2$$

适当的最优化计算软件可以帮助完成这一工作。

神经元网络一个最重要的应用是在给定训练样本集的前提下，用其输出 $\boldsymbol{g}(\boldsymbol{x})$ 去近似不同类别后验概率分布。有些时候并不知道后验概率的形式，此时可以构造一个神经元网络，对应每一个样本类别有一个输出。对于一个给定的训练样本 $\boldsymbol{x}^e$，可以这样构造期望的输出向量 $\boldsymbol{o}^e$，令 $\boldsymbol{o}^e$ 中对应训练样本所属类别的元素为 1，其他元素都为 0。这样，就可以按照前面的方法训练神经元网络，并把神经元网络的输出看做类后验概率的模型。对于任意一个类别未知的样本特征 $\boldsymbol{x}$，可以计算神经元网络 $\boldsymbol{n}(\boldsymbol{x};\hat{\boldsymbol{p}})$ 的输出，然后选择输出向量中最大的元素对应的类别作为样本的类别。

22.4.2　误差最小化

前面提到，训练神经元网络的目的是使所有训练样本的实际输出与期望输出的总差距最小，也就是使下式的取值最小

$$Error(\boldsymbol{p})=\left(\frac{1}{2}\right)\sum_e|\boldsymbol{n}(\boldsymbol{x}^e;\boldsymbol{p})-\boldsymbol{o}^e|^2$$

上式是参数向量 $\boldsymbol{p}$ 的函数。有很多方法可以帮助求出使得上式取值最小的参数向量 $\hat{\boldsymbol{p}}$，这里介绍梯度下降算法。梯度下降算法是一种迭代算法，对某个已经求出的迭代值 $\boldsymbol{p}_i$，下一个迭代值 $\boldsymbol{p}_{i+1}$ 可以通过下式求得：

$$\boldsymbol{p}_{i+1}=\boldsymbol{p}_i-\epsilon(\nabla\,Error)$$

其中，ϵ是一个小的常数。

随机梯度下降法 对于某个样本 e，令其误差为 $Error(\boldsymbol{p};\boldsymbol{x}^e)$，则总误差为 $Error(\boldsymbol{p}) = \sum_e Error(\boldsymbol{p};\boldsymbol{x}^e)$。如果采用梯度下降算法，迭代公式为：

$$\boldsymbol{p}_{i+1} = \boldsymbol{p}_i - \epsilon \nabla Error$$

（其中，总误差函数对 $\boldsymbol{p}$ 求导后在$\boldsymbol{p} = \boldsymbol{p}_i$ 处的导数值）如果ϵ足够小，则有

$$\begin{aligned} Error(\boldsymbol{p}_{i+1}) &= Error(\boldsymbol{p}_i - \epsilon \nabla Error) \\ &\approx Error(\boldsymbol{p}_i) - \epsilon(\nabla Error \cdot \nabla Error) \\ &\leqslant Error(\boldsymbol{p}_i) \end{aligned}$$

上式的等号仅在取到极小值的时候成立。这里面存在一个问题：计算误差和误差梯度的时候需要在所有训练样本上求和，实际应用中，训练样本的数量可能是很大的。我们希望尽量避免这种求和运算，因此可以采用随机梯度下降算法。在该算法中，每次迭代时，只从训练样本集中随机选取一个样本，单独计算这个样本误差的梯度，然后采用下面的迭代公式

$$\boldsymbol{P}_{i+1} = \boldsymbol{P}_i - \epsilon \nabla Error(\boldsymbol{p};\boldsymbol{x}^e)$$

（样本 $\boldsymbol{p}$ 是在等概率 $\boldsymbol{p}_i$ 的条件下随机选取的，并保证每次选取的样本不同）。采用这种算法，总误差不一定随每次迭代而下降，但如果选择合适的ϵ，可以保证总误差的期望值是下降的，如下式

$$\begin{aligned} \mathrm{E}(Error(\boldsymbol{p}_{i+1})) &= \mathrm{E}(Error(\boldsymbol{p}_i - \epsilon \nabla Error(\boldsymbol{p};\boldsymbol{x}^e)) \\ &\approx \mathrm{E}(Error(\boldsymbol{p}_i) - \epsilon(\nabla Error \cdot \nabla Error(\boldsymbol{p};\boldsymbol{x}^e)) \\ &= Error(\boldsymbol{p}_i) - \epsilon \frac{1}{n} \sum_e (\nabla Error \cdot \nabla Error(\boldsymbol{p};\boldsymbol{x}^e)) \\ &= Error(\boldsymbol{p}_i) - \epsilon \left(\nabla Error \cdot \left(\frac{1}{n} \sum_e \nabla Error(\boldsymbol{p};\boldsymbol{x}^e) \right) \right) \\ &= Error(\boldsymbol{p}_i) - \frac{\epsilon}{n}(\nabla Error \cdot \nabla Error) \\ &< Error(\boldsymbol{p}_i) \text{ if } |\nabla E| > 0 \end{aligned}$$

这样，每步选择一个样本计算梯度（再次强调，样本是在等概率的条件下随机选取的），由于总误差期望值是下降的，因此经过有限步计算后，可以达到总误差函数的极小值。梯度可以通过不同的方法计算出来，其中一个有效的方法是向后传播算法，将在 22.7 节中讨论这一算法。

算法 22.7 随机梯度下降算法（采用向后传播法计算梯度）使总误差函数最小化

随机选择一个参数向量初值 $\boldsymbol{p}_o$；

采用向后传播法计算该样本误差梯度

$\nabla Error(\boldsymbol{x}^e;\boldsymbol{p}_o)$

$\boldsymbol{p}_n = \boldsymbol{p}_o - \epsilon \nabla Error(\boldsymbol{x}^e;\boldsymbol{p}_o)$

直到$|Error(\boldsymbol{p}_n) - Error(\boldsymbol{p}_o)|$很小或$|\boldsymbol{p}_o - \boldsymbol{p}_n|$很小；

（未完待续）

（续）

$p_o = p_n$

在等概率条件下随机选择一个训练样本(x^e, o^e)

采用向后传播法计算该样本误差梯度

$\nabla Error(x^e; P_o)$

$p_n = p_o - \epsilon \nabla Error(x^e; p_o)$

end

22.4.3　何时停止训练

一般说来，梯度下降算法并不一定要持续到找到精确的最小值，有趣的是，这样做还可以增加算法的鲁棒性。可以通过考察总误差函数在最小值点附近的变化情况来解释这个问题。如果误差函数在最小值点附近变化剧烈，那么神经元网络的输出就会对参数的选取很敏感。这样的神经元网络的泛化能力是很差的，假设有一个很大的训练样本集，用其中一半的样本做训练可以得到一组参数，用另一半训练可以得到另一组参数，这两组参数间肯定会存在着微小的差别，如果误差函数在最小值点附近变化剧烈，这意味着，用一半样本训练出的神经元网络对另一半样本的分类效果将非常差，这显然不是我们所希望的。

因此神经元网络的误差函数在最小值点附近的变化应该是平缓的，换句话说，很难找到一个精确的最小值点，因为一旦接近了最小值点，训练集上的总误差将基本上不再变化。在实际应用中，一旦同时满足下列两个条件：(a)训练集上每个样本至少平均被选取了一次；(b)总误差函数的下降值低于某个预先设定的阈值，随机梯度下降算法就将停止。

另一个比较困难的问题是选择几层的神经元网络，每层多少个结点。这是一个模型选择的问题，通常需要通过实验来选择出合适的值，如果读者对这部分内容感兴趣，请看参考文献中 Ripley (1996)的论文和 Haykin (1999)的论文。

22.4.4　利用神经元网络进行人脸检测

人脸检测是分类器的一种，此处用来说明分类器的有效性。前面提到过，在大小相当的前提下，所有正面的人脸在某种意义下都是相同的——前额、面颊和鼻子等部分是浅色的区域，而眼睛、眉毛、鼻子底部和嘴等部分是深色的区域。可以在固定大小的所有图像窗口中搜索所有类似人脸的图像窗口来检测人脸。尺寸大一些或小一些的人脸可以在粗一些或细一些尺度的图像中搜索。

左侧光照射的人脸或右侧光照射的人脸图像是不一样的，因此，在检测图像窗口时，需要对图像窗口的光照效果进行校正。通常，光照效果是一个线性的斜坡(一边亮，另一边暗，两者之间是平滑的过度)，因此可以根据图像窗口的灰度值拟合出一个光照平面，然后用图像窗口的灰度值减去光照平面的值以去掉光照效果。另一个方法是取图像灰度值的对数值并将对应于线性坡度的对数值减去，其好处是利用了光照平面在对数形式中是求和的形式(这是一种粗略的模型)。但实际上，使用对数在实际应用中的结果并不明显。另一个去除光照效果的方法是直方图均衡化，这样做可以保证所有图像窗口的直方图是大致相同的(图 22.14 描述了直方

图均衡化的方法)。

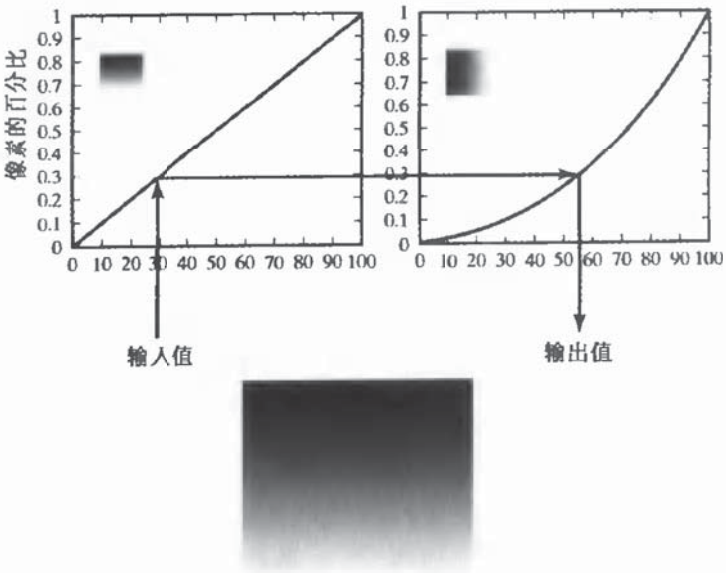


图 22.14 直方图均衡化利用累积直方图对待均衡图像进行灰度映射,目的是使待均衡图像的直方图与另一幅图像大致相同。上图显示了两幅图像的累积直方图,为了把左图的直方图变换为右图的直方图,从左图选一个灰度值,在累积直方图上找到它对应的百分比,然后把右图累积直方图上相应百分比处的灰度值作为变换后的灰度值。实际上,左图是一幅线性灰度平面(由于亮度和照明度之间的非线性关系,可能使得左图看上去是非线性的),右图是一幅立方根灰度平面,左图变换后的结果显示在下图,此时它原来的线性灰度平面已经转化为与右图类似的立方根灰度平面

对一个校正了光照效果的图像窗口,需要判断它是不是人脸。由于人脸的朝向是未知的,因此需要或者判断人脸的方向,或者构造一个对方向不敏感的分类器。Rowley, Baluja 和 Kanade(见参考文献中他们 1998*b* 年的论文)利用神经元网络建立了一个非常有效的人脸检测器,它首先用一个神经元网络判断图像窗口的方向,然后校正图像窗口使其变成正面的图像,最后把校正后的图像窗口送给另一个神经元网络以判断究竟是不是人脸(如图 22.15 所示,详细内容见参考文献中 Rowley, Baluja 和 Kanade 1996 年和 1998*a* 年的论文)。用于判断方向的神经元网络包括 36 个输出结点,分别表示 0 度、10、……、350 度。图示 22.16 显示了这个人脸检测系统的一些检测结果。

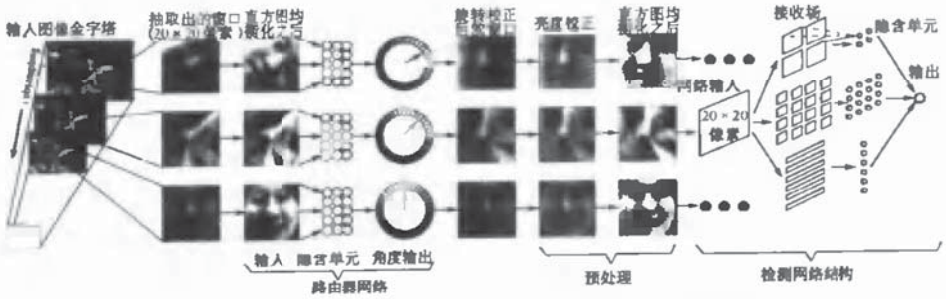


图 22.15 上图显示了 Rowley, Baluja 和 Kanade 人脸检测系统的结构。固定大小的图像窗口首先经直方图均衡化去除光照差别;然后这些图像窗口被送到一个神经元网络去估计窗口的方向。接着图像窗口被旋转相应的角度并被送到第二个神经元网络以判断当前的图像窗口是否是一个人脸

图 22.16　上图显示了 Rowley, Baluja 和 Kanade 人脸检测系统的一些检测结果。图中方框标识的区域是被认为包含人脸的区域，人脸的方向则由方框中标识眼睛的两个圆圈之间的位置关系决定

22.4.5　卷积神经元网络

神经元网络并非只有前面介绍的那一种结构，在实际应用中还会用到很多其他形式的神经元网络(参考文献中 Bishop, 1995 年的论文和 Haykin, 1999 年的论文，对各种形式的神经元网络有一个较为全面的介绍)，其中有一种形式的神经元网络对解决视觉应用中的问题非常有效，我们称它为卷积神经元网络。这种神经元网络的一个优势在于它可以用一组滤波输出描述一个图像区域，此外，还可以对一次滤波输出得到的图像描述再次滤波，得到一个复合图像描述。例如，如果希望检测手写字体，则需要一些滤波器检测图像中不同方向长条，所有检测出的长条可以构成一幅新的图像，然后需要对新的图像再次滤波找出这些长条之间的位置

关系。

这意味着有些时候我们需要建立一个滤波器系统来找出某些基元之间的某些关系，然后再利用典型的神经元网络根据前面滤波得到的描述对待识别样本进行分类。实际上，无法事先确定使用什么样的滤波器，因此，滤波器的形式也需要从训练样本集中学习。

Lecun，Bottou，Bengio 和 Haffner 于 1998 年利用卷积神经元网络实现了一些手写数字分类器（见参考文献中他们 1998 年的论文），图示 22.17 给出了这些分类器的基本结构。分类器的输入是一个 32×32 的图像窗口。第一步（见图示 22.17 的 C1）生成 6 幅特征图像，每幅特征图像是由输入图像先用一个 5×5 的模板滤波，然后加上一个常量，再作用一个 S 型函数后得到的结果，其中生成 6 幅特征图像所用的滤波模板和常量各不相同，这些参数都是通过学习得到的。

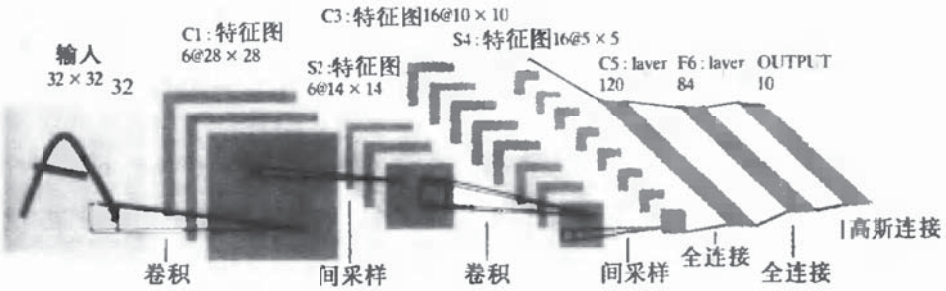


图 22.17 上图显示了 LeNet 5（一个用于检测手写字符的卷积神经元网络）的结构，其中标记为C的层是卷积层，标记为S的层是下采样层。随着神经元网络向前传递，特征数量增加同时图像窗口分辨率降低。最终图像窗口被送到一个全连接的神经元网络，它将根据待识别样本特征到标准样本特征之间的距离给出最终的分类结果

由于我们不关心图像中特征的精确位置，因此在第二步中（见图 22.17 的 S2），6 幅特征图像的分辨率被降低了，它们首先被分为若干个 2×2 的小块，然后取每个小块的平均值，乘以一个参数，再加上一个参数，然后再作用一个 S 型函数，这样就生成了 6 幅新的特征图像（新特征图像的长和宽分别是原特征图像的一半）。实际上，新特征图像相当于是原特征图像的重采样结果，其中重采样中乘以的数和加上的数也是通过学习得到的。上述先滤波再重采样的过程需要重复几次，在这个过程中，特征图像的数量增加、分辨率降低。最后我们得到一个由 84 个神经元组成的输出层，每个神经元以前面层的所有输出作为输入。

上述神经元网络可以用来识别手写字体，输出的结果是一幅 7×12 的图像，在输出图像中手写字体已经被校正成为近似的规范字体，此时可以把这个输出图像与其他规范字体进行比较从而判定是哪个字。利用这个网络可以有效地校正不规范的手写字体。对一个输入的手写字体，可以对其校正后的字体利用近邻法判断输入的是哪个字，采用这种测试方法，上述网络的识别错误率为 0.95%（如图 22.18 所示）。

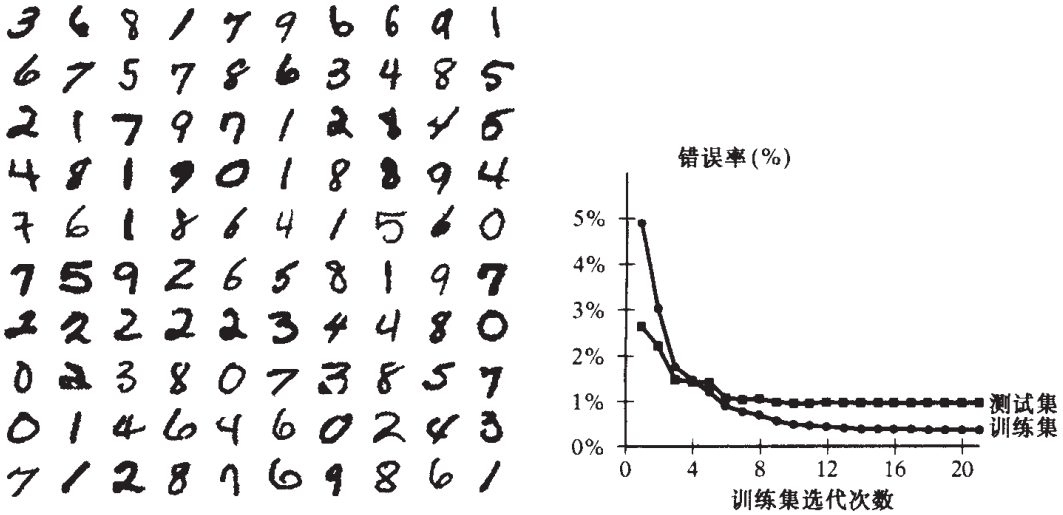


图 22.18　左图显示了用于训练和测试 LeNet 5 系统的 MNIST 手写字符数据库中的部分样本。不难发现，同类字符的不同样本之间存在着较大差异。右图显示了 LeNet 5 系统在训练样本集和测试样本集上的错误率，这个错误率被视为训练过程中对全部 60 000 个样本进行梯度下降迭代次数的函数(也就是说，横轴上的 6 表示总共进行了 360 000 次梯度下降迭代)。从右图中不难发现，在某些点处，训练集错误率降低而测试集的错误率并没有降低，这是因为训练过程只保证在训练集上达到最优，这说明神经元网络可能会出现过训练的问题

22.5　支持向量机

在视觉应用中，分类器仅仅是一种方法，并非结果，一旦出现了某种更简单、更可靠、更有效的分类技术，这种技术将立刻得到广泛的应用，支持向量机就是这样一种技术。目前，在需要应用分类器的场合，多数人会首先尝试支持向量机(除非样本分布已知，但这种情况几乎不会出现)。在这一节中，我们将介绍支持向量机的基本概念，并给出一些例子来证明这种技术是有效的。

假设有属于两个类别的 N 个样本，样本的特征用 $\boldsymbol{x}_i$ 表示，类别的编号用 y_i 表示，对其中一类样本 y_i 取 1，而对另一类样本 y_i 取 -1。于是，可以得到下面的样本集：

$$\{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_N, y_N)\}$$

需要建立这样一个规则，对任意一个输入特征 $\boldsymbol{x}$ 预测其输出 y 的符号，这个规则就是分类器。

在这一节中，我们只讨论两分类问题，这两类样本可以是线性可分的，也可以是线性不可分的，其中线性可分的情况比较简单，我们先讨论这种情况。

22.5.1　线性可分样本集上的支持向量机

在线性可分的样本集上，可以找到某个合适的 $\boldsymbol{w}$ 和 b(表示一个超平面)使得对训练集上的任意一个样本都满足：

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) > 0$$

对训练集上的每一个样本，都有上述这样一个不等式，所有的不等式构成了对 $\boldsymbol{w}$ 和 b 的约束，这个约束表明，所有 y_i 取正值的样本在某个超平面一侧，而所有 y_i 取负值的样本则在某个超平面另一侧。

由于训练集中的样本是有限的，因此可以找到满足约束条件的一组超平面，每个超平面都能把两类训练样本的凸包分开。其中一个较为保守的选择是找到一个离两类样本的凸包都较远的超平面。如何找到这样的超平面呢？我们可以连接两个类别中距离最近的点，过这条线段的中点做一个垂直于这条线段的超平面。这个超平面使得样本点到该超平面的最小距离最大化，从这个意义上讲，它满足距离两类样本点的距离都尽可能大(如图 22.19 所示)。

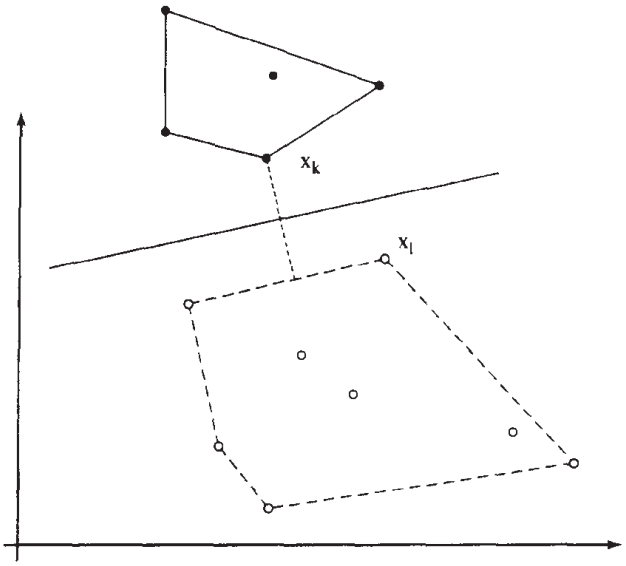


图 22.19 上图显示了支持向量机对于线性可分样本集构造的决策超平面，其中实心圆圈表示一类样本，空心圆圈表示另一类样本。我们画出了每个样本集的外包络凸多边形，对决策面的最好选择是使两类样本的外包络凸多边形到决策面的最小距离都尽可能大。可以用如下办法构造决策面，首先找到连接两个凸多边形的最短线段，然后过这条线段的中点做一个垂直于线段的超平面作为决策面。实际上，训练集中只有一部分样本决定了决策面的位置，这些样本是两类中到达决策面距离最短的样本，这些样本是我们关心的样本，可以利用它们确定决策面的位置

由于把 $\boldsymbol{w}$ 和 b 同时乘以或除以一个正数并不会影响约束条件 $y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)>0$，因此可以选择 $\boldsymbol{w}$ 和 b 使得对于训练集中的任意一个样本都满足：

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)\geqslant 1$$

对于距离超平面最近的样本点，上式等号成立。设 $\boldsymbol{x}_k$ 使等号成立，且 $y_k=1$，$\boldsymbol{x}_l$ 也使等号成立，且 $y_l=-1$。这意味着，$\boldsymbol{x}_k$ 在超平面的一侧，而 $\boldsymbol{x}_l$ 在超平面的另一侧，此外，$\boldsymbol{x}_k$ 和 $\boldsymbol{x}_l$ 是距离超平面最近点，这两个点到超平面的距离是一样的。需要注意的是，距离超平面距离最近的点可能有很多个。

由上面的假设可得：$\boldsymbol{w}\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)=2$，于是有：

$$dist(\boldsymbol{x}_k, \text{hyperplane}) + dist(\boldsymbol{x}_l, \text{hyperplane}) = \left(\frac{\boldsymbol{w}}{|\boldsymbol{w}|} \cdot \boldsymbol{x}_k + \frac{b}{|\boldsymbol{w}|}\right) - \left(\frac{\boldsymbol{w}}{|\boldsymbol{w}|} \cdot \boldsymbol{x}_1 + \frac{b}{|\boldsymbol{w}|}\right)$$
$$= \frac{\boldsymbol{w}}{|\boldsymbol{w}|} \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_2) = \frac{2}{|\boldsymbol{w}|}$$

这表明选择超平面距两类样本距离的最大值等价于取$(1/2)\boldsymbol{w} \cdot \boldsymbol{w}$的最小值，于是可以得到一个带约束条件的最小值问题

$$\text{minimize}(1/2)\boldsymbol{w} \cdot \boldsymbol{w}$$
$$\text{受限于 } to\ y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geqslant 1$$

其中，每个训练样本都有一个约束条件。

训练支持向量机　为了解决上述带约束条件的最小值问题，我们采用拉格朗日算法，引入拉格朗日系数α_i，得到如下拉格朗日函数：

$$(1/2)\boldsymbol{w} \cdot \boldsymbol{w} - \sum_1^N \alpha_i (y_i (\boldsymbol{w} \cdot \boldsymbol{x}_1 + b) - 1)$$

该拉格朗日函数要满足对$\boldsymbol{w}$和$\boldsymbol{b}$的导数取最小值，而对α_i的导数取最大值——这就是Karush-Kuhn-Tucker条件（见参考文献中 Gill，Murray 和 Wright 1981 年的最优化问题课本），经过简单的推导，可以得到下面两个式子：

$$\sum_1^N \alpha_i y_i = 0$$

$$\boldsymbol{w} = \sum_1^N \alpha_i y_i \boldsymbol{x}_i$$

上面第二个等式解释了为什么称这种技术为支持向量机。通常，作为两类分界面的超平面只由训练样本集中相对一小部分样本决定，而与其余部分样本的位置是不相关的（如图 22.19 所示）。也就是说，大部分训练样本对应的拉格朗日系数α_i的取值为 0，而那些对应α_i取值不为 0 的训练样本才真正决定了分界超平面的位置，这部分样本就称为支持向量。

将上面两个等式代入原最小值问题中，经过简单推导，可以得到如下的对偶问题：

$$\text{使} \quad \sum_i^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i (y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j) \alpha_j \text{最大化}$$

$$\text{受限于} \alpha_i \geqslant 0$$

$$\text{和} \quad \sum_{i=1}^N \alpha_i y_i = 0$$

注意上述问题的约束条件是二次的，称为二次规划问题，是一个典型的最优化问题。一些标准的软件包能帮助解决这一问题。但是这里解决的问题有其特性，虽然问题中包括大量的未知变量，但是当目标函数取到最小值时，大部分变量的取值为 0，可以利用这一点简化计算过程（见参考文献中 Smola 等 2000 年的论文）。

算法 22.8 在线性可分的样本集上建立支持向量机

符号说明:已知一个 N 样本的训练集 $\{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_N, y_N)\}$,其中 y_i 的取值为 1 或 -1,分别表示两个类别

建立 SVM:建立并解决如下的对偶优化问题:

$$\text{使} \sum_{i}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i (y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j) \alpha_j \text{ 最大化}$$

$$\text{受限于} \quad \alpha_i \geqslant 0$$

$$\text{和} \sum_{i=1}^{N} \alpha_i y_i = 0$$

然后,对所有不为 0 的 α_i,计算 $\boldsymbol{w} = \sum_1^N \alpha_i y_i \boldsymbol{x}_i$,再根据 $y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) = 1$,可以计算出 b。

分类:对任意一个新的样本特征 x,可以通过计算下面的符号对其进行分类

$$\begin{aligned} f(\boldsymbol{x}) &= \operatorname{sign}(\boldsymbol{w} \cdot \boldsymbol{x} + b) \\ &= \operatorname{sign}\left(\left(\sum_{1}^{N} \alpha_i y_i \boldsymbol{x} \cdot \boldsymbol{x}_i\right) + b\right) \\ &= \operatorname{sign}\left(\sum_{1}^{N} (\alpha_i y_i \boldsymbol{x} \cdot \boldsymbol{x}_i + b)\right) \end{aligned}$$

22.5.2 基于支持向量机的行人检测

我们可以粗略地认为,视觉应用中的行人具有大致相同的类似于棒棒糖形状的外观——上面是较宽的身体,下面是较细的腿。这样的行人可以通过支持向量机从图像中检测出来。基本的思想与 22.4.4 节中提到的人脸检测的想法类似,把固定大小的图像窗口传给分类器,分类器则告诉我们这个图像窗口中是否包含行人。图像窗口中的像素个数可能很大,其中很多像素都是不相关的。在人脸检测中,可以选择椭圆形的图像窗口,但是人的轮廓是很不规范的,因此只能选择矩形的图像窗口。

下面需要选择矩形窗口的某种特征来帮助判断这个矩形窗口中是否包括行人。从一组例子中获取一组特征是很自然的。有多种特征选择算法可供选择,Oren, Papageorgiou, Sinha, Osuna 和 Poggio(见参考文献中他们 1997 年的论文)选择了小波系数作为图像窗口的局部特征,这里的小波系数是图像窗口对某种特定滤波器的响应。他们采取了一种平均值策略。他们认为,在包含行人的图像窗口中,背景部分可以视为噪声,而不包含行人的图像窗口也可以视为噪声,此外,对于特定的滤波器,噪声响应的平均值也是知道的。对于我们选择的一个特征,可以取所有包含行人的图像窗口的特征平均值,如果这个平均值与噪声响应的平均值之间的差距很大,那么这个特征就是一个好的特征,否则这个特征就没有把握住图像窗口中行人的特点,就是一个不好的特征(如图 22.20 所示)。

特征选定之后,可以按照前面提到的方法训练支持向量机,实验证明,这个方法是有效的(如图 22.21 和图 22.22 所示)。采用 22.4 节介绍的自举方法将进一步改善检测系统的性能。

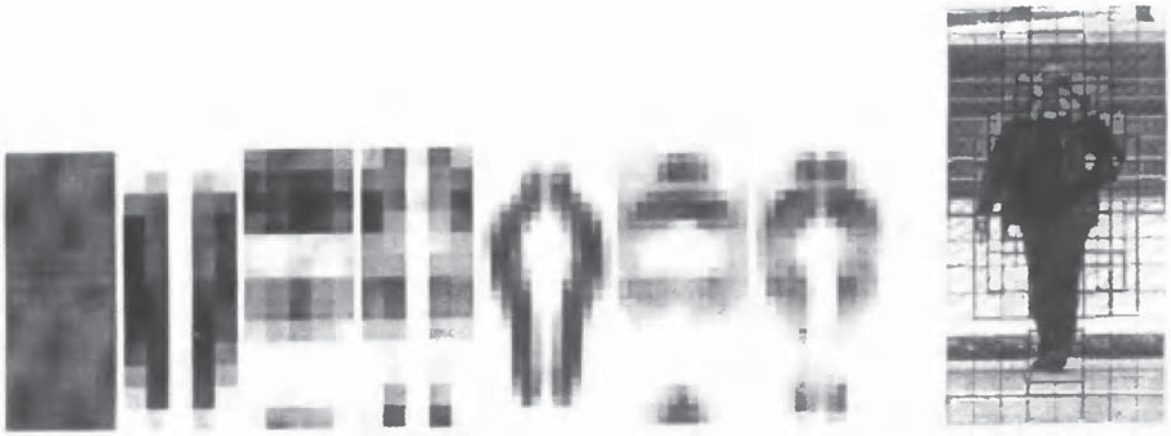


图 22.20　左图显示一些训练样本在不同图像位置上小波系数的平均值。大于平均值的系数用深色表示，小于平均值的系数用浅色表示。我们认为噪声的小波系数近似为平均值，这样较深和较浅的颜色区域就包含了大量信息，可以利用这些信息找到行人。右图显示了对训练样本的一个方格划分，对每个方格计算其特征，值得注意的是，在出现边缘的位置方格密度较大，而其他位置方格密度则较小

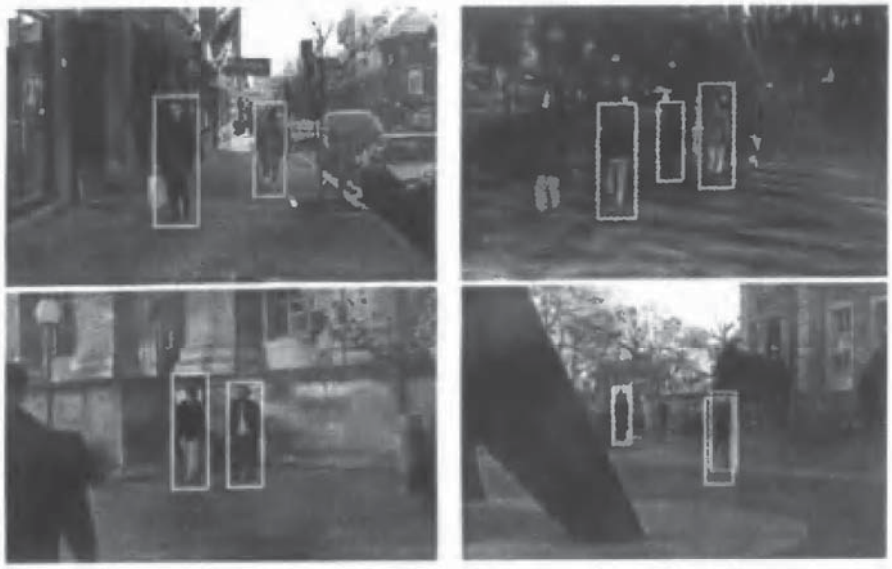


图 22.21　上图给出了利用 Papageorgiou，Oren 和 Poggio 的方法进行行人检测的一些结果，虽然不是所有的引人都检测出来，但可以看出检测率是比较高的，其接受操作曲线如图22.22所示

22.6　注释

在什么样的场合下运用哪种分类器目前还没有定论，可以根据应用的需要采用最合适的分类器。为了简单起见，在这里只介绍了有限的几种分类器，还有很多非常有用的技术并没有提到，详细内容请参阅参考文献中的下列书目——Bishop，1995 年；Hastie，Tibshirani 和 Friedman，2001；Haykin，1999；McLachlan 和 Krishnan，1996；Ripley，1996；Vapnik，1996 和 1998。

在构造分类器的技术中，直接确定决策面的方法比估计类后验概率的方法简单。但是，采用估计类后验概率的方法可以明确地给出样本属于各个类别的可能性，而直接确定决策面的方法则无法做到这一点。此外，在直接确定决策面的方法中要求每个训练样本的类别都是已

知的。对于一个实际应用中的识别问题,很难建立一个两全其美的分类器。无论采用哪种方法,为了构造有效的分类器,可能都会需要大量的训练样本,通常,如果采用的模型越强(所谓模型越强,是指模型中需要估计的参数越少),则所需要的训练样本数量就越少。

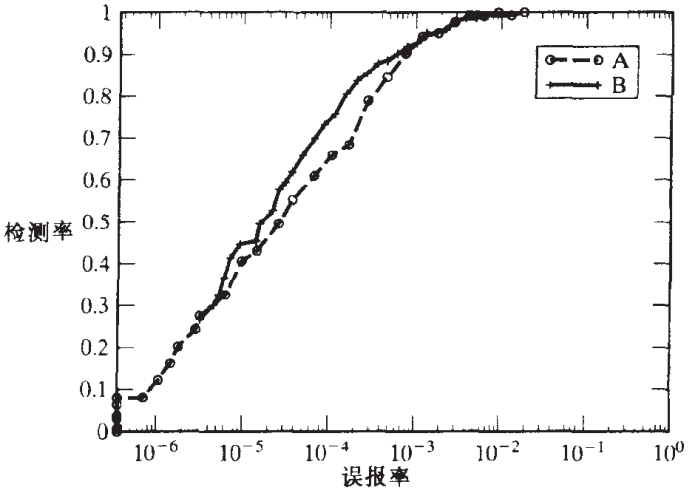


图 22.22 上述给出了 Papageorgiou,Oren 和 Poggio 行人检测系统的接受操作曲线

当样本的自由度很多,而我们又不清楚这些自由度之间关系的时候,很难建立一个有效的分类器。如果不同样本特征个数不一样,也很难利用分类器对样本进行分类。在上述两种情况中,可能需要应用结构模型对样本进行分类。在可以应用分类器的场合中,除非引入很强的假设条件外,一般很难对现在视觉应用中的高维特征样本建立概率密度模型,那些我们熟悉的简单概率模型在实际中几乎是不会出现的。此外,我们也不知道如何建立一个可以自动适应问题规模的概率模型,而这一点在视觉应用中将是非常有用的。

如何建立分类器在视觉领域和学习领域都处于研究的关键阶段。在面对一个具体问题的时候,我们无从知道哪种方法是最佳的。在这一章和下一章中,我们给出了在实际应用中曾经采用过的一些方法,其中的一些被实验证明是非常有效的,但是,到目前为止,在如何选择分类器这一问题上还没有形成一套完整的理论。

有时候可以训练多个分类器,把它们的输出结合起来以对物体进行分类,这样可以避免训练一个庞大的分类器,这种方法称为自举(boosting)方法。自举(boosting)方法中采用的分类器具有简单的决策面,容易进行训练,但分类效果一般比较差。通常,先在训练集上训练一个分类器,找出这个分类器在训练集上的错分类样本。然后,通过增加这些样本的错分类权重或在训练集中复制这些样本增加对这些样本的关注程度,之后再训练一个分类器,并找出训练集上新分类器的错分类样本。重复上面的步骤以建立多个分类器。在分类时,把所有分类器的输出结果进行加权组合,作为最终的输出结果。

肤色检测

有许多关于肤色检测的论文,上文中选择了与我们讲述的内容比较接近的一篇。肤色检测的用途包括检测人体的裸露部分(见参考文献中 Fleck,Forsyth 和 Bregler,1996 年的论文以及 Forsyth和 Fleck,1999 年的论文)以及检测和跟踪人脸和手(见参考文献中 Park,Seo,An 和 Chung,2000 年的论文以及 Yoo 和 Oh,1999 年的论文)等。肤色检测有多种方法,参考文献中

Brand 和 Mason，2000 年的论文比较了一些方法。实验证明，加入肤色检测环节的人脸检测系统其性能将得到提高，因此，在研究人脸检测问题时，我们自然而然地会去研究肤色检测的问题。

人脸检测

解决人脸检测问题需要用到模板匹配技术，通常的做法是，将一个图像窗口（一般是除去光照效果的图像窗口）传给某种形式的分类器，由分类器告诉我们这个图像窗口是不是人脸。我们已经提到了人脸检测领域的某些成果，如 Rowley，Baluja 和 Kanade 和工作以及后来的 Schneiderman 和 Kanade 的工作，为了简洁起见，我们没有提到 Sung 和 Poggio 的工作（见参考文献中他们 1998 年的论文），但他们的工作也同样重要。此外，Osuna，Freund 和 Girosi 采用支持向量机建立了一个有效的人脸识别系统（见参考文献中他们 1997 年的论文），这类方法目前已经被广泛地应用于人脸检测问题中。

人脸识别（告诉我们检测出的人脸是谁）的问题同样重要，通常情况下，人脸检测和人脸识别是连续进行的两个步骤。讨论这一问题比较重要的论文包括参考文献中 Brunelli 和 Poggio，1992 年和 1993 年的论文。此外，人脸识别问题涉及到很多技术，例如，如何根据小数量的样本进行识别（见参考文献中 Beymer 和 Poggio，1995 年的论文）；如何根据不同的视图进行识别（见参考文献中 Beymer，1994 年的论文）以及如何对不同光照条件下的人脸进行识别（见参考文献中 Adini，Moses 和 Ullman，1994 年和 1997 年的论文；Georghiades，Belhumeur 和 Kriegman，2000 年的论文；Jacobs，Belhumeur 和 Basri，1998 年的论文以及 Georghiades，Kriegman 和 Belhumeur，1998 年的论文）。

行人检测

模板匹配的方法在行人检测问题中也是适用的，用于匹配的模板的形状类似于棒棒糖。如果行人举起手他可能就无法被检测出来（因为这时他看起来就不像棒棒糖了），但是在多数情况下，行人会把手放在身体的两侧，这意味着在多数情况下，行人是有可能被检测出来的。此外，行人的运动也具有特征，同样可以被检测出来（见参考文献中 Papageorgiou，Oren 和 Poggio，1998 年的论文以及 Papageorgiou 和 Poggio，1999 年和 2000 年的论文）。

习题

22.1　在某个特征空间 S 中的度量为 $\boldsymbol{x}$。有一个开放集 D，其中任何一个元素的类别为第 1 类，而在 $S-D$ 的内部的任何元素为第 2 类。

a. 证：

$$
\begin{aligned}
R(s) &= Pr\{1 \to 2 \mid \text{using } s\}\, L(1 \to 2) + Pr\{2 \to 1 \mid \text{using } s\}\, L(2 \to 1) \\
&= \int_{S-D} p(1 \mid \boldsymbol{x})\, d\boldsymbol{x} L(1 \to 2) + \int_{D} p(2 \mid \boldsymbol{x})\, d\boldsymbol{x} L(2 \to 1)
\end{aligned}
$$

b. 在计算总风险时为什么能够忽略 D 的边界（它也是 $S-D$ 的边界）？

22.2　在 22.1 节曾提到如果每个类条件密度具有相同的协方差，算法 22.2 分类器就宽度为比较 2 个 $\boldsymbol{x}$ 的线性表达式

a. 证明以上结论

b. 如果只有两类，试证只需测试 $\boldsymbol{x}$ 的一线性表达式的正负号即可。

22.3 在 22.3.1 节中曾设置一个特征 u，第 i 个数据点的 u 的值由 $u_i = \boldsymbol{v}\cdot(\boldsymbol{x}_i - \boldsymbol{\mu})$ 给出，证明 u 具有零均值。

22.4 在 22.3.1 节中曾设置一例特征 u，第 i 个数据点的 u 的值为 $u_i = -\boldsymbol{v}\cdot(\boldsymbol{x}_i - \boldsymbol{\mu})$，则 $\boldsymbol{v}$ 是数据项协方差文即在 Σ 的特征向量。请试用对称矩阵的特征的量是正交的返一点证明不同的特征是独立的。

22.5 在 22.2.1 节中曾说接收器操作曲线不随先验选择的变化而改变，试证之。

编程作业

22.6 编制一个程序能对图像中肤色像素进行标注，要求至少比较两种不同的分类器。

22.7 选一个课文中讲过的人脸检测方法，编程并加以实现。

22.7 附录 I：回传算法

采用随机梯度下降法训练神经元网络的难点在于求误差函数的梯度 $\nabla Error$，有一个有效的方法称为回传算法。这种算法把神经元网络的层次结构看做复合函数，并计算导数值。回顾前面提到的两层网络，可表示为：

$$\boldsymbol{f}(\boldsymbol{x}) = \left[\phi(\boldsymbol{w}_{21}\cdot\boldsymbol{y}), \phi(\boldsymbol{w}_{22}\cdot\boldsymbol{y}), \cdots, \phi(\boldsymbol{w}_{2n}\cdot\boldsymbol{y})\right]$$

其中

$$\boldsymbol{y}(\boldsymbol{z}) = [\phi(\boldsymbol{w}_{11}\cdot\boldsymbol{z}), \phi(\boldsymbol{w}_{12}\cdot\boldsymbol{z}), \cdots, \phi(\boldsymbol{w}_{1m}\cdot\boldsymbol{z}), 1]$$

和

$$\boldsymbol{z}(\boldsymbol{x}) = \left[x_1, x_2, \cdots, x_p, 1\right]$$

要计算的是

$$\frac{\partial\, Error}{\partial w_{kl,m}}$$

其中，$w_{kl,m}$ 是 $\boldsymbol{w}_{kl}$ 的第 m 个分量。首先考虑输出层，此时关心的是 $w_{2l,m}$，有

$$\begin{aligned}
\frac{\partial\, Error}{\partial w_{2l,m}} &= \sum_k \frac{\partial\, Error}{\partial f_k}\frac{\partial f_k}{\partial w_{2l,m}} \\
&= \frac{\partial\, Error}{\partial f_l}\frac{\partial f_l}{\partial w_{2l,m}} \\
&= \sum_e \left\{(f_l(\boldsymbol{x}^e) - o_l^e)\frac{\partial f_l}{\partial w_{2l,m}}\right\} \\
&= \sum_e \left\{(f_l(\boldsymbol{x}^e) - o_l^e)\phi'_{2l}(y_m(\boldsymbol{x}^e))\right\} \\
&= \sum_e \left\{\delta_{2l}^e(y_m(\boldsymbol{x}^e))\right\}
\end{aligned}$$

其中

$$\phi'_{2l} = \frac{\partial\phi}{\partial u}$$

在 $u = \boldsymbol{w}_{2l} \cdot \boldsymbol{y}$ 处计算该导数，写做

$$\delta_{2l}^{e} = (f_l(\boldsymbol{x}^e) - o_l^e)\phi_{2l}'$$

这里面 $y_m(\boldsymbol{x}^e)$ 表示这一层的输入，而 δ_{2l}^{e} 表示这一层的输出。

下面再考虑隐含层，这时我们关心的是 $w_{1l,m}$，有

$$\begin{aligned}
\frac{\partial\, Error}{\partial w_{1l,m}} &= \sum_k \left\{ \frac{\partial\, Error}{\partial f_k} \frac{\partial f_k}{\partial w_{1l,m}} \right\} = \sum_{i,j} \left\{ \frac{\partial\, Error}{\partial f_i} \frac{\partial f_i}{\partial y_j} \frac{\partial y_j}{\partial w_{1l,m}} \right\} \\
&= \left\{ \sum_k \frac{\partial\, Error}{\partial f_k} \frac{\partial f_k}{\partial y_l} \right\} \frac{\partial y_l}{\partial w_{1l,m}} \\
&= \sum_e \left\{ \sum_k \left\{ (f_k(\boldsymbol{x}^e) - o_k^e) \frac{\partial f_k}{\partial y_l} \right\} \frac{\partial y_l}{\partial w_{1l,m}} \right\} \\
&= \sum_e \left\{ \sum_k \left\{ (f_k(\boldsymbol{x}^e) - o_k^e) \phi_{2k}' w_{2k,l} \right\} \frac{\partial y_l}{\partial w_{1l,m}} \right\} \\
&= \sum_e \left\{ \sum_k \left\{ (f_k(\boldsymbol{x}^e) - o_k^e) \phi_{2k}' w_{2k,l} \right\} \phi_{1l}' z_m \right\} \\
&= \sum_e \left\{ \sum_k \left\{ \delta_{2k}^{e} w_{2k,l} \right\} \phi_{1l}' z_m \right\}
\end{aligned}$$

其中

$$\phi_{2k}' = \frac{\partial \phi}{\partial u}$$

在 $u = \boldsymbol{w}_{2k} \cdot \boldsymbol{y}$ 处计算

$$\phi_{1l}' = \frac{\partial \phi}{\partial u}$$

在 $u = \boldsymbol{w}_{1l} \cdot \boldsymbol{z}$ 处计算，令：

$$\delta_{1l}^{e} = \sum_k \left\{ \delta_{2k}^{e} w_{2k,l} \right\} \phi_{1l}'$$

则：

$$\frac{\partial E}{\partial w_{1l,m}} = \sum_e \delta_{1l}^{e} z_m(\boldsymbol{x}^e)$$

不难发现，上式的形式与在输出层求出的形式是类似的——结果是求和的形式，求和的每个分量是上一层的偏导数、这一层的偏导数和这一层的输入的乘积。实际上，如果还有第三层，那么也能把误差函数关于这一层参数的偏导数表示成类似的形式——结果也是求和的形式，每个分量是第二层的偏导数、第三层偏导数和第三层输入的乘积。这意味着可以采用下面的两步算法：

1. 计算神经元网络每个结点对每个样本的输出，称为前馈步骤；
2. 利用上面的结果计算误差函数的梯度值，称为回传步骤。

上述方法求出总误差函数对于神经元网络中的每个参数的偏导数。前面提到过，采用随机梯度下降算法避免了对所有样本求和，由于梯度计算满足线性性，每步只需要计算一个样本误差

的梯度，具体的算法见算法 22.9。

算法 22.9 基于两层神经元网络的回传算法——用于计算单个样本误差函数关于神经元网络参数的偏导数

符号说明：

将一个两层神经元网络记为：

$$f(\boldsymbol{x};\boldsymbol{p}) = \left[\phi(\boldsymbol{w}_{21}\cdot\boldsymbol{y}), \phi(\boldsymbol{w}_{22}\cdot\boldsymbol{y}), \cdots, \phi(\boldsymbol{w}_{2n}\cdot\boldsymbol{y})\right]$$

$$\boldsymbol{y}(\boldsymbol{z}) = [\phi(\boldsymbol{w}_{11}\cdot\boldsymbol{z}), \phi(\boldsymbol{w}_{12}\cdot\boldsymbol{z}), \cdots, \phi(\boldsymbol{w}_{1m}\cdot\boldsymbol{z}), 1]$$

$$\boldsymbol{z}(\boldsymbol{x}) = \left[x_1, x_2, \cdots, x_p, 1\right]$$

（$\boldsymbol{p}$ 是包含所有参数的参数向量），单个样本的误差函数记为：

$$\begin{aligned} Error^e &= Error(\boldsymbol{p};\boldsymbol{x}^e) \\ &= \left(\tfrac{1}{2}\right)|\boldsymbol{f}(\boldsymbol{x}^e;\boldsymbol{p}) - \boldsymbol{o}^e|^2 \end{aligned}$$

要计算的单个样本误差函数关于所有参数分量的偏导数：

$$\frac{\partial\, Error^e}{\partial w_{kl,m}}$$

其中 $w_{kl,m}$ 是 $\boldsymbol{w}_{kl}$ 的第 m 个分量。

前馈步骤：计算 $\boldsymbol{f}(\boldsymbol{x}^e;\boldsymbol{p})$，并保留所有的中间结果。

回传步骤：首先计算

$$\delta_{2l}^e = (f_l(\boldsymbol{x}^e) - o_l^e)\phi_{2l}'$$

$$\text{在 } u = \boldsymbol{w}_{21}\cdot\boldsymbol{y} \text{ 处计算 } \phi_{2l}' = \frac{\partial\phi}{\partial u}$$

$$\frac{\partial\, Error^e}{\partial w_{2l,m}} = \sum_e \left\{\delta_{2l}^e(y_m(\boldsymbol{x}^e))\right\}$$

然后计算：

$$\delta_{1l}^e = \sum_k \left\{\delta_{2k}^e w_{2k,l}\right\}\phi_{1l}'$$

$$\text{在 } u = \boldsymbol{w}_{11}\cdot\boldsymbol{z} \text{ 处计算 } \phi_{1l}' = \frac{\partial\phi}{\partial u}$$

$$\frac{\partial E^e}{\partial w_{1l,m}} = \delta_{1l}^e z_m(\boldsymbol{x}^e)$$

22.8 附录 II：线性不可分数据集上的支持向量机

在多数情况下，决策面不是一个超平面，这称为线性不可分情况。在线性不可分的情况下，某些样本不满足前面提出的约束条件，因此需要引入一些松弛变量 $\xi_i \geqslant 0$ 来修正约束条件，修正后的约束条件可表示为：

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_1+b)\geqslant 1-\xi_i$$

此时，训练支持向量机的问题变为：

$$\text{使}\frac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w}+C\sum_{i=1}^{N}\xi_i\text{最小化}$$

$$\text{受限于 } y_i(\boldsymbol{w}\cdot\boldsymbol{x}_1+b)\geqslant 1-\xi_i$$

$$\text{和 } \xi_i\geqslant 0$$

其中，C 是某个指定的常数，它实际上起控制对错分类样本惩罚程度的作用，实现在错分类样本比例与算法复杂度之间的折中。经过简单推导，可以得到如下对偶优化问题：

$$\text{使}\sum_{i}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i(y_iy_j\boldsymbol{x}_i\cdot\boldsymbol{x}_j)\alpha_j\text{ 最大化}$$

$$\text{受限于}C\geqslant\alpha_i\geqslant 0$$

$$\text{和}\sum_{i=1}^{N}\alpha_iy_i=0$$

对于 w，仍有

$$\boldsymbol{w}=\sum_{1}^{N}\alpha_iy_i\boldsymbol{x}_i$$

但是 b 的求解方法略有不同。对于满足 $C>\alpha_i>0$ 训练样本，其松弛变量ξ_i 为 0，因此下式成立：

$$\sum_{j=1}^{N}y_j\alpha_j\boldsymbol{x}_i\cdot\boldsymbol{x}_j+b=y_i$$

对于满足上述条件的样本，可以根据上式求出 b。同样，在线性不可分情况下，训练支持向量机的关键部分还是解决一个对偶优化问题，但此时可能有大部分拉格朗日系数 α_i 不为 0。

22.9　附录 III：非线性支持向量机

对于多数样本集，采用线性决策面可能无法得到一个好的分类器，在这种情况下，需要一个几何上更复杂的决策面。解决这个问题的一个方法是把原始特征映射到新的特征空间上，使得在新特征空间上的决策面是一个超平面。例如，假设有一个两类平面点集，这两类点可以被平面上的二次曲线分开，那么可以对这个平面点集应用下面的映射

$$(x,y)\rightarrow(x^2,xy,y^2,x,y)$$

在新的特征空间中，这两类特征点的分界面就变成了一个超平面。但这种形式的映射一般不会用到，因为映射后新的特征空间的维数比原特征空间的维数高。

我们把映射记做 $\boldsymbol{x}'=\phi(\boldsymbol{x})$。对于映射后的特征点$\boldsymbol{x}'_i$，最优化问题中出现的是它们的内积运算，如下式：

$$\boldsymbol{x}'_i\cdot\boldsymbol{x}'_j$$

也可以把上式写做$\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j)$。把上式代入训练支持向量机的最优化问题，问题的解法并不会发生变化，解决最优化问题后，分类器可由下式给出：

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\sum_{1}^{N}(\alpha_i y_i \boldsymbol{x}'\cdot\boldsymbol{x}_i' + b)\right) = \operatorname{sign}\left(\sum_{1}^{N}(\alpha_i y_i \phi(\boldsymbol{x})\cdot\phi(\boldsymbol{x}_i) + b)\right)$$

假设有一个函数 $k(\boldsymbol{x},\boldsymbol{y})$，对任意的一对向量 $\boldsymbol{x},\boldsymbol{y}$，函数的取值均为正数。可以证明：一定存在某个$\phi$，使得 $k(\boldsymbol{x},\boldsymbol{y})=\phi(\boldsymbol{x})\cdot\phi(\boldsymbol{y})$，这意味着，当我们考虑映射时，不用关心$\phi$的形式，只考虑找到一个合适的函数 $k(\boldsymbol{x},\boldsymbol{y})$就可以了，用 $k(\boldsymbol{x},\boldsymbol{y})$代替$\phi$，最优化问题的形式变为：

$$\text{使}\sum_{i}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i(y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j))\alpha_j \text{ 最大化}$$

$$\text{受限于 } \alpha_i \geqslant 0$$

$$\text{和 } \sum_{i=1}^{N}\alpha_i y_i = 0$$

而分类器的形式变为：

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\sum_{1}^{N}(\alpha_i y_i k(\boldsymbol{x},\boldsymbol{x}_i) + b)\right)$$

上述公式均假设样本在由函数 k 决定的新特征空间中是线性可分的。如果是线性不可分的情况，则仍然可以引入松弛变量，这时问题的描述如下：

$$\text{使 } \sum_{i}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i(y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j))\alpha_j \text{ 最大化}$$

$$\text{受限于 } C \geqslant \alpha_i \geqslant 0$$

$$\text{和 } \sum_{i=1}^{N}\alpha_i y_i = 0$$

分类器的形式为：

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\sum_{1}^{N}(\alpha_i y_i k(\boldsymbol{x},\boldsymbol{x}_i) + b)\right)$$

有很多合适的函数可以作为 $k(x,y)$，最重要的一点是 k 必须满足：对所有的 $\boldsymbol{x},\boldsymbol{y}$，其取值必须为正。表 22.2 种给出了一些 k 的常用选择，目前我们还不知道在某种特定的情况下，如何选择 k 是最佳的，我们可以尝试不同的选择，并利用交叉验证的方法找出错误率最小的选择作为最佳选择。

表 22.2 一些支持向量机的核

核函数形式	由该核函数表示的 ϕ 的定性属性
$(\boldsymbol{x}\cdot\boldsymbol{y})^d$	ϕ 是 d 阶的所有单项式
$(\boldsymbol{x}\cdot\boldsymbol{y}+c)^d$	ϕ 是 d 阶或低于 d 阶的所有单项式
$\tanh(a\boldsymbol{x}\cdot\boldsymbol{y}+b)$	
$\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y})}{2\sigma^2}\right)$	

第 23 章　基于模板间关系的识别

某些物体可能存在着内在的自由度，这意味着该物体在不同时刻表现出来的特征可能不尽相同（例如，一个人可以活动胳膊和腿，可以像鱼一样游泳，像蛇一样扭动，等等）。在这种情况下，采用模板匹配的识别方法是很困难的，因为需要建立决策面相当复杂的分类器或建立多个不同的模板。

上述这类物体一般在某些小的部分上表现出稳定性，因此可以作为模板实现匹配。对于某个待识别物体，可以通过考察这些模板之间的某些具有启发性的关系，来确定物体所属的类别。例如，在进行人脸检测的时候，我们不是把待识别图像与一个完整的人脸模板进行比较，而是在待识别图像中寻找眼睛、鼻子和嘴，然后考察它们之间的位置关系以确定待识别图像是不是人脸。

上述方法具有一些可能的优势。首先，模板的结构简单，便于学习，不难想像，让分类器学习一个眼睛模板要比让它学习一个人脸模板容易得多。其次，可以归纳和利用简单的概率模型，因为样本特征中包含很多独立性。再次，可以用相对少的模板识别大量的样本，动物脸的识别就是一个很好的例子——几乎所有动物的脸上都有眼睛、鼻子和嘴，而且不同动物的这些器官的空间布局之间差异很小。最后，利用简单的局部模板可以刻画复杂的物体，例如，人可以活动胳膊和腿，因此，很难对人建立一个单一完整的模板，但是，可以建立一些局部模板，再建立一个概率模型表示这些局部的自由度，用这种方法刻画人会相对简单得多。

从建立一种标准的方法看，我们对这个话题还了解不够，但是到目前为止，要解决的关键问题已经非常清楚了，那就是以什么样的形式组织模板间的关系最便于我们解决问题？在这一章中，我们将介绍一些常用的方法。首先，介绍对不同关系投票的方法，通过统计票数最多的关系确定物体所属的类别（23.1 节）。此外，还可以建立概率模型，以便对某些空间关系增加权值。实际上，需要找到一个概率分布函数，当各个模板的分布情况符合要检索物体的模式时，这个概率函数取大值，否则取小值。此时，物体检测的问题就转化为搜索一组可以使上述概率密度函数取最大值的模板（23.2 节）。有时希望简化搜索的过程，我们也将讨论简化的方法（23.3 节）。但是，即使简化了搜索过程，搜索的代价可能仍然是很大的，我们将讨论一类有助于进行有效搜索的概率模型（23.4 节）。最后将给出两个例子（23.5 节和 23.6 节）。

23.1　通过对模板间关系投票检测物体

采用简单的物体模型也可以达到很好的识别效果。最简单的模型是把表现物体的图像视为由若干种不同类型的具有某种特征的图像块的集合，从而构成了相应的图像模式。为了确定当前的图像属于哪种图像模式，我们根据每个图像块的特征对所有模式进行投票，一个图像块投一票，票数最多的模式就是图像的模式，也就是物体的类别。虽然这个方法看起来很简单，但它是非常有效的。在这一节中，将首先讨论选择图像块的方法，然后再由易到难地介绍几种常用的投票方法。

23.1.1 选择图像块

通常,如果一个小图像块中很多像素处的灰度导数不为0(例如图像中表示拐角的像素,后称角点),那么这个图像块将表现出一些特有的属性。Schmid 和 Mohr(1997*a*,*b*)正是基于这一点建立了一个系统,他们选择图像中的角点(也称为兴趣点,见参考文献中 Schmid,Mohr 和 Bauckhage 三人 2000 年的论文),然后估计图像的灰度级在这些点处的导数值,并求出一组图像灰度级导数的函数,要求这些函数不随图像旋转、平移、缩放以及光照条件变化而变化,这些特征称为不变局部片断,我们这里不讨论如何详细描述这些特征。采用这种方法的价值在于:由于这种组合是不变的,因而从不同视角观察同一物体会呈现出相同的特征。

假设图像块也能分为不同的类别,从一类样本的一组图像中,可以归纳出这种类别样本典型的图像块,绝大多数情况下,相似的图像块属于相同的类别,但由于图像噪声的存在,这些图像块的不变局部片断可能存在某些程度的不同。可以把图像块人为地分成多个类别,也可以通过聚类的方法把它们分成多个类别(后者是更好的方法)。需要确定在什么时候,两组不变局部片断表示同一个类别的图像块。Schmid 和 Mohr 采用考察待识别图像块和范例图像块特征向量之间的 Mahalanobis 距离的办法,如果被考察的距离小于某个阈值,则判定待识别图像块与范例图像块属于同一个类别。

不难发现,上述方法实际上实现了一个分类器,用来判定待识别图像块所属的类别,或判定待识别图像块不属于已知类别,这些类别用范例表示,可以采用第 22 章介绍的任何一种方法实现模板匹配。当然,不要求选择的图像块特征一定旋转不变性,这样一来就要求训练集中包括了不同角度、不同位置、不同大小和不同光照的样本,而分类器就可以学习旋转、平移、缩放和光照对图像块造成的差异,使得这些差异不会影响分类结果。而采用不变局部片断的优势正在于避免了分类器学习大量样本。

23.1.2 投票和一种简单的产生式模型

按照前面的方法,对于某幅给定的图像,我们首先找到兴趣点,并对兴趣点邻域的图像块进行分类。之后如何确定图像中有什么样的模式呢?可以通过建立图像块所属类别与图像模式之间的关系回答这个问题。假设图像中有 N_i 个图像块,我们还假定图像模式只能是已知的模式集合中的一种,或者不属于已知的模式集合,同时,图像块的类别可以属于任何一种图像模式,也可以不属于所有已知的图像模式(这时,称图像块类别来自于噪声),但是,在通常情况下,一个模式不会包含所有的图像块类别,这意味着断言某种模式是否存在等价于断言某些图像块是否来自于噪声(因为只有一个模式能够出现,而这些图像块所属的类别又不在该模式中)。

现在,我们有一个简单的图像产生式模型,对于一个特定模式的图像,它的图像块只能属于某些特定的类别,而不可能属于其他类别。通过精心制作这个模型,可以得到一系列识别图像模式的算法。

上述产生式模型的一个最简单的情形是,假设一个图像模式产生了它所能产生的所有类图像块,并且要求尽可能少的图像块是由噪声产生的。基于这种假设,可以用投票的方式来确定待识别图像的模式。对待识别图像中的每个图像块,根据这个图像块的类别给每种包含这种类别图像块的模式投一票,得票最多的模式就被判定为图像所属的模式。这种方法简单而有效,但也存在着一些问题(如图 23.1 所示)。

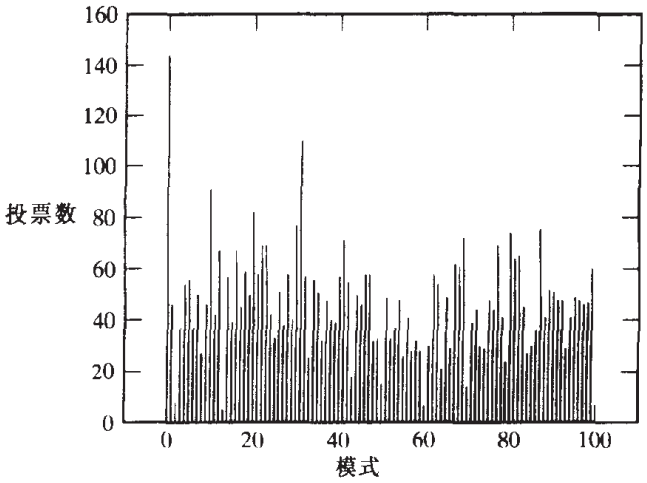


图 23.1　上图显示了某幅特定图像在简单投票策略下对不同模式的投票结果。从图示上可以看出，0号模式获得了最多票数，因此它将被判定为匹配的最终结果，同时还有另外三个模式所得的票数超过了总票数的半数

23.1.3　投票的概率模型

可以用一个概率模型解释简单投票的方法，它可以帮助理解投票方法的优点和不足。前面提到的产生式模型可以转变成概率模型，这种模型假设在物体存在的条件下，图像块的产生是独立且随机的，记做

$$P\{\text{patch of type } i \text{ appears in image} \mid j\text{th pattern is present}\} = p_{ij}$$

$$P\{\text{patch of type } i \mid \text{no pattern is present}\} = p_{i\alpha}$$

下面考虑最简单的情况，对于第 j 种图像模式，如果这种图像模式包含了第 i 种图像块类别，令 $p_{ij} = \mu$，否则令 $p_{ij} = 0$，此外，对所有 i，令 $p_{i\alpha} = \lambda < \mu$，最后，再假设观察到的每一个图像块可能产生于某个模式或噪声。设图像包括 n_i 个图像块，在上面的假设条件下，对于一种确定的图像模式，我们只需要知道有哪个图像块来自这种模式，哪个图像块出自于噪声，就可以计算出这种图像模式的似然率。假设有 n_p 个图像块属于某一具体模式，剩下的 $n_i - n_p$ 个图像块属于噪声，则该模式的似然值为：

$$P(\text{interpretation} \mid \text{pattern}) = \mu^{n_p}\lambda^{(n_i - n_p)}$$

显然，在假设 $\lambda < \mu$ 的条件下，上述似然值随 n_p 的增大而增大。但是，由于没有一种模式可以包含所有的图像块类别，因此 n_p 可能达到的最大值与具体的图像模式有关。在这种简单的产生式模型下，投票的过程实际上就是选择一种图像模式，使得上式的取值达到最大的过程。

但简单诱导模型存在着一些不足：首先，各个图像模式出现的概率可能不同，此时必须考虑图像模式的先验概率；此外，某个图像块类别来自噪声的概率可能比其来自图像模式的概率大，简单产生式模型没有考虑这种情况，这可能影响到投票结果的科学性；最后，某些图像块出现在某种物体图像中的可能性，比出现在其他物体图像中的可能性大，例如，表现图像角点的图像块会经常出现在国际象棋棋盘图像中，而在类似斑马纹理的图像中出现的可能性就比较小。

改进的产生式模型　现在讨论的模型假设：对给定的一种图像模式，各种图像块类别的出

现是独立的，这是一个相对简单的模型。若一共有 N 种图像块类别，现在假设一种图像块类别属于各种图像模式以及属于噪声的概率是不同的。假设属于第 l 类的图像块有 n_{il} 个，在属于第 l 类的 n_{il} 个图像块中，有 n_k 个来自某种图像模式，剩下的来自噪声。

此时，某种图像模式的似然率为：

$$P\left(\begin{array}{c|c} n_1 \text{ of type 1 from pattern,} & \\ \cdots, & \\ n_N \text{ patches of type } N \text{ from pattern} & j\text{th pattern} \\ \text{and } n_{i1}-n_1 \text{ of type 1 from noise,} & \\ \cdots, & \\ n_{iN}-n_N \text{ from noise} & \end{array}\right)$$

上式中来自于图像模式的项可表示为

$$P(\text{patches from pattern} \mid j\text{th pattern})P(\text{patches from noise})$$

其值为

$$P(\text{type1} \mid j\text{th pattern})^{n_1}\,\mathrm{P}(\text{type2} \mid j\text{th pattern})^{n_2}\ldots P(\text{type } N \mid j\text{th pattern})^{n_N}P(\text{noise})\text{其被评估为}$$

$$p_{1j}^{n_1}p_{2j}^{n_2}\cdots p_{Nj}^{n_N}$$

如果我们假设各个图像块类别来自于噪声的图像块个数是独立的，那么来自于噪声的项可表示为

$$P(\text{type 1} \mid \text{noise})^{(n_{i1}-n_1)}\cdots P(\text{type } N \mid \text{noise})^{(n_{iN}-n_N)}$$

其值为

$$p_{1x}^{(n_{i1}-n_1)}\cdots p_{Nx}^{(n_{iN}-n_N)}$$

因此，某个图像模式的总似然率值可表示为

$$p_{1j}^{n_1}p_{2j}^{n_2}\cdots p_{Nj}^{n_N}p_{1x}^{(n_{i1}-n_1)}\cdots p_{Nx}^{(n_{iN}-n_N)}$$

我们希望上式取到最大值。对于第 k 个图像块类别，如果 $p_{kj} > p_{kx}$，则应有 $n_k = n_{ik}$；否则，应有 $n_k = 0$，只有在满足上面的条件下，上式的似然性才可能取到最大值。我们用 π_j 表示图像属于第 j 种模式的先验概率，用 π_0 表示图像不属于任何一种模式的先验概率，此时，对第 j 种图像模式，它后验概率的最大值的形式如下

$$\left(\prod_m p_{mj}^{n_{im}}\prod_l p_{lx}^{n_{il}}\right)\pi_j$$

其中，m 表示所有满足 $p_{mj} > p_{mx}$ 的图像块类别，而 l 表示所有满足 $p_{lj} < p_{lx}$ 的图像块类别，可以对所有图像模式计算上面后验概率的结果，然后选择后验概率最大的一个作为判定的图像模式。需要注意的是，尽管表面上并没有考虑不同模板之间的几何关系，但上述模型仍然是一个关系模型。实际上，对于一个给定的图像模式，各个模板之间的联系已经隐含在条件概率中了，如果把改进的诱导模型与 22.2.2 节解决人脸检测问题所采用的模型相比较，可以发现这两个概率模型在本质上是一致的。

23.1.4　对关系投票

如果考虑模板间的几何关系，可以很容易地改进投票的方法。当我们考虑匹配的时候，不仅仅要求单一的图像块匹配，同时要求与它相邻的图像块之间也匹配，此外还要求这些匹配的图像块以“恰当的方式”排列，但是什么是“恰当的方式”可能较难判断。我们暂且在这里讨论考虑平面旋转、平移和缩放的物体匹配问题（正面人脸的检测与识别就是这一类问题）。

假设找到一个图像块与某种模式的代表物体匹配，接下来考察与这个图像块相邻的 p 个图像块，是否也与同一个代表物体匹配，如果匹配率大于 50%，再继续考察所有匹配的图像块中任意三个所构成的三角的角度是否与代表物体中相应的角度相等（这个条件称为半局部约束条件，如图 23.2 和图 23.3 所示）。如果上面两条都满足，则给这个模式投一票。可以把这种方法与几何散列法（见 18.4.2 节）相比较（请参阅参考文献中 Schmid 和 Mohr 两人1997（a，b）年的两篇论文）。很难为这种方法给出一个概率上的解释，因为此时图像块的出现概率与具体的图像模式以及图像块在图像模式中的位置都有关系。

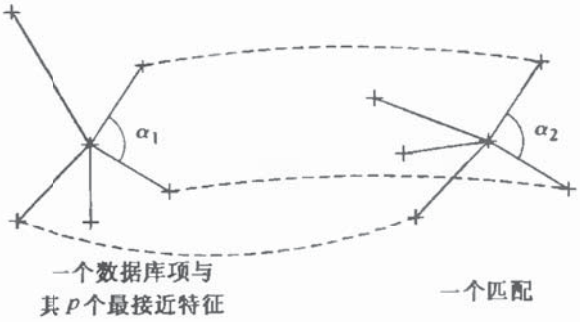


图 23.2　除了对图像中每个图像块所匹配的模式进行投票，还可以对一个图像块集合进行投票。此时，仅在某个图像块匹配上某个模式，同时，该图像块超过特定比例的相邻图像块也与相同的图像模式匹配时，才给相应的图像模式投一票。如图所示，三个匹配图像块之间的夹角应该在某个特定的范围内，这将大大减少随机匹配的数量

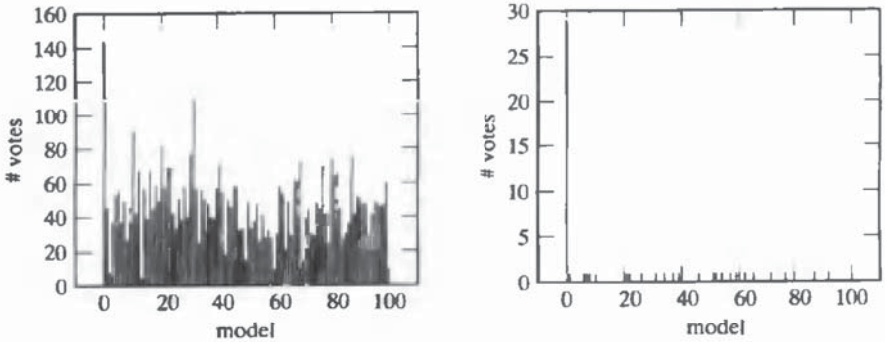


图 23.3　实际上，与原始投票策略相比，采用半局部约束将大大加强对投票结果的鉴别能力。左图显示了在原始投票策略下某幅图像每个模式的投票结果。尽管最终结果0号模式获得了最多的票数，仍有另外三个模式所得的票数超过了总票数的半数。右图显示了采用半局部约束后的投票结果，此时0号模式作为最终结果的论据更为明显

23.1.5　投票和三维物体

上一节介绍了 Schmid 和 Mohr 的方法在二维物体上的应用，其实这种方法也可以简单地

推广到三维物体上。我们可以把物体一系列视图中的每一个看做二维视图，如果二维视图足够多的话，这种方法是可行的，因为视角的微小差异造成的二维视图的差异是在匹配误差允许范围之内的。

这种把三维物体识别转化为二维物体识别的思想是很直观的，但是也存在着一些问题。最主要的一个问题是，这样做会产生大量的模式，使得投票变得非常复杂。目前，我们还无法确定要识别一个三维物体最少需要多少张二维视图。图 25.9(在第 25 章)就是采用这种方法得出的。

23.2 利用概率模型及搜索的关系推理

在前面一节中，我们假设在给定图像模式的前提下，各个模板(即图像块)所属的类别是条件独立的。尽管基于这个假设的方法在实际应用中效果不错，但这个假设对大多数物体是不成立的，因为通常在不同模板间存在着不可忽略的相互关系。例如，人脸上只有两只眼睛、一个鼻子和一张嘴，两只眼睛之间的距离和鼻子到嘴的距离大致相当，两只眼睛的连线与鼻梁和嘴中心的连线基本上垂直，等等。在 23.1.4 节中，我们曾利用类似的约束条件控制投票，但是并没有给出一个能够使用上述约束条件的基本框架。

但给出一个基本框架是比较困难的，我们需要建立一个模型，表示出各个约束条件对于推理过程的重要性。到目前为止，在这个问题上还没有一套完整的理论，因此，这里只介绍涉及的主要问题和常用的方法。

23.2.1 对应与搜索

在这一节中，我们将讨论利用概率模型进行匹配的核心问题。利用概率模型进行匹配的方法通常是获取对图像的一种采样，计算给定图像这种采样的后验概率值，然后接受后验概率最大的采样。

上面的描述忽略了很多实际问题。最基本的一个问题是，对于一幅表示特定物体的图像，我们不知道哪些图像信息来自于物体本身，哪些来自于图像噪声。通常，这是一个比较困难的问题，为了避开这个难点，可以把问题转化为对应问题，即先将图像与模板对应，然后对模板间的关系进行推理。例如，为了检测人脸，需要建立多个检测器，分别检测眼睛、鼻子和嘴，然后再确定这些元素(眼睛、鼻子和嘴)的布局。

对应 如何对应是另外一个非常重要的问题。如果找不到图像模式中包含的各个模板，就无从计算图像模式的后验概率，也就是说，首先要做的一件事情就是从待识别图像中找出基本模板(如眼睛、鼻子和嘴)，然后才能计算后验概率。在解决对应问题时，一个简单的想法是搜索图像中所有可能的对应，但可以改进搜索算法简化搜索过程。

将第 18 章的技术移植过来是相对容易的，需要利用概率论进行包装。对于识别刚性物体的问题来说，只需找到少数的对应就可以推导出很多的对应，用概率论的语言解释，对于刚性物体，对应是不独立的。

通常，概率模型估计一组变量的联合概率密度，我们称这些变量一个可能的取值为一个组合(也可以称为一个组或一种假设)。一个组合由多个模板检测器的输出组成[例如，人脸检测中的一个组合包括两个眼睛检测器的输出(左眼和右眼)，一个嘴检测器的输出和一个鼻检测

器的输出],这些输出可能与位置和方向有关,此外输出的结果还包括一个标号。这些标号表示了一个对应,这里标号是非常重要的,例如,对眼睛检测器来说,如果它不区分左眼和右眼,就必须给检测器的相应标出是左眼和右眼,但是对于嘴检测器来说,就不存在这个问题。

所有可能的对应集合的数量是很多的,不可能逐个计算它们的概率。例如,如果有一个可以检测双眼眼睛检测器、一个嘴检测器和一个鼻检测器,眼睛检测器有 N_e 个输出结果,嘴检测器有 N_m 个输出结果,鼻检测器有 N_n 个输出结果,则所有可能的取值集合的个数为 $O(N_e^2 N_n N_m)$。如果已经仔细阅读了第 18 章,会知道这是一个极端的过估计。贯通本章的论点是用数量很少的对应来预测其他对应。

递增对应组合和搜索　重新考虑对应的过程,我们要做的是构造一个对应组合,并给组合中的每一个元素确定一个标号,构造对应组合的过程实际上是一个递增的过程——通过扩充小的对应组合形成大的对应组合。每一步,我们判断当前的对应组合是否完备,如果完备(表示已经找到了一个物体),则接受这个对应集合,否则,则考虑继续扩充当前的对应集合。为了表述方便,在这里定义"对"的概念,一个"对"由一个检测器的输出以及相应的标号构成(对于人脸检测来说,标号包括"左眼"、"右眼"、"嘴巴"和"鼻子",当然,有可能存在着一些编号没有输出值与之对应),而需要构造的对应集合正是由若干个这样的"对"构成。通常,有一个对应假设集合,这个集合可能包含大量的元素,每个元素是某一个检测器输出结果,表示一个可能的对应。在构造对应集合的时候,从对应假设集合中选择一个对应,并选择一个标号与之形成一个"对",然后或者接受这个"对",把它加入到当前的对应集合中,并把这个对应从对应假设集合中去掉;或者拒绝这个"对",并把这个对应放回到对应假设集合中。直到接受当前的对应集合(找到物体)或判定不可能接受对应集合(判定当前图像不含指定物体)为止。

构造对应集合需要思考以下几个重要问题:

- **什么样的对应集合是完备的**　当我们已经对应到了某个图像模式包括的所有模板(对人脸检测来说,这意味着已经对应到了双眼、嘴和鼻子),并且已经对应到的结果满足分类器规则(表示当前的对应结果的确表示一个希望找到的物体),则此时的对应集合就是完备的。
- **什么样的对应集合可以被接受**　当对应集合已经完备的时候,可以停止搜索对应的过程,并接受当前的对应集合;此外,即使我们还没有搜索到所有需要的对应,如果当前的对应集合已经足够帮助我们判定找到了我们希望找到的物体(这样情况将在后面讨论),此时也可以接受当前的对应集合。
- **如何选择下一个对应**　在对应搜索中,需要确定下一步从对应假设集合中选择哪一个对应。通常,选择最有可能表示我们希望找到的物体的一个对应。

建立搜索问题　仍然使用人脸检测的例子来讨论这个问题。假设有左眼检测器在 $\boldsymbol{x}_1$ 位置有一个输出,右眼检测器在 $\boldsymbol{x}_2$ 位置有一个输出,嘴检测器在 $\boldsymbol{x}_3$ 位置有一个输出,鼻检测器在 $\boldsymbol{x}_4$ 位置有一个输出,其他的检测器输出都由噪声造成,被检测的图像在 $\boldsymbol{F}$ 位置有一个人脸,如果只检测图像中有没有单个人脸,这时,只需要比较下面两个后验概率:

$$P(\text{one face at } \boldsymbol{F} \mid \boldsymbol{X}_{\text{le}} = \boldsymbol{x}_1, \boldsymbol{X}_{\text{re}} = \boldsymbol{x}_2, \boldsymbol{X}_{\text{m}} = \boldsymbol{x}_3, \boldsymbol{X}_{\text{n}} = \boldsymbol{x}_4, \text{all other responses})$$

$$P(\text{no face} \mid \boldsymbol{X}_{\text{le}} = \boldsymbol{x}_1, \boldsymbol{X}_{\text{re}} = \boldsymbol{x}_2, \boldsymbol{X}_{\text{m}} = \boldsymbol{x}_3, \boldsymbol{X}_{\text{n}} = \boldsymbol{x}_4, \text{all other responses})$$

停止搜索:检测 假设在出现人脸的情况下,由噪声造成的输出与由人脸造成的输出是条件独立的,则有

$$P(\text{one face at } \boldsymbol{F} \mid \boldsymbol{X}_{\text{le}} = \boldsymbol{x}_1, \boldsymbol{X}_{\text{re}} = \boldsymbol{x}_2, \boldsymbol{X}_{\text{m}} = \boldsymbol{x}_3, \boldsymbol{X}_{\text{n}} = \boldsymbol{x}_4, \text{ all other responses})$$

其值等于

$$P(\text{one face at } \boldsymbol{F} \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) P(\text{all other responses})$$

上式的值正比于

$$P(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \mid \text{one face at } \boldsymbol{F}) P(\text{all other responses}) P(\text{one face at } \boldsymbol{F})$$

经过简单的推导,比较前面两个后验概率可以转化为比较下面两个式子

$$P(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \mid \text{one face at } \boldsymbol{F})$$

或

$$(P(\text{noise responses}) P(\text{no face}) / P(\text{one face at } \boldsymbol{F})) \text{ (term in relative loss)}$$

由于只需要判定给定的图像中有没有人脸,因此只关心这两类情况的后验概率。实际上,在一个完备的输出结果集上,可以比较这两个后验概率,同时在一个不完备的输出结果集上,也可以比较这两个概率。

一旦搜索到的对应结果满足了分类准则(图像包括人脸的后验概率大于图像不包括人脸的后验概率),就可以停止搜索了,也就是说,并不一定需要搜索到所有的特征才能判定图像中包括人脸,一旦已经对应到的结果足以判定图像中包括人脸,就可以判定图像中包括人脸,并停止搜索。

举一个例子。假设在图像中对应到了一只右眼、一张嘴和一个鼻子,希望知道这三个对应结果是否表示一张脸。这时需要计算以下联合概率密度:

$$P(\boldsymbol{X}_{\text{le}} = \text{missing}, \boldsymbol{X}_{\text{re}} = \boldsymbol{x}_2, \boldsymbol{X}_{\text{m}} = \boldsymbol{x}_3, \boldsymbol{X}_{\text{n}} = \boldsymbol{x}_4 \mid \text{one face at } \boldsymbol{F})$$

计算上式最简单的办法是认为左眼检测器没有输出,而没有输出这个事件与其他三个检测器的输出是独立的,于是

$$P(\text{missing}, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \mid \text{face})$$

的值等于

$$\int P(\text{le does not respond} \mid \boldsymbol{X}_1) P(\boldsymbol{X}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \mid \text{face}) \, d\boldsymbol{X}_1$$

如果再假设没有输出这个事件与位置也是独立的,则有

$$P(\text{missing}, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \mid \text{face})$$

的值等于

$$P(\text{le does not respond}) \int P(\boldsymbol{X}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \mid \text{face}) d\boldsymbol{X}_1$$

只有在能够确定图像中包括人脸的时候,才能够在缺少某些对应结果的情况下停止对应搜索,否则不能停止搜索,在 23.3 节中,将讨论在一定限度内简化搜索。如果在缺少某些对应结果的情况下确定了图像模式,只可能是下列两种情况之一:(a)某种对应没有被找到,但是已知的对应结果和它们的布局已经足够帮助确定图像模式;(b)某种对应没有被找到,其他的对应也没有全部找到,但其他检测器的布局特点可以确定图像模式。在(b)情况下,可能需要找到一定数量的其他对应。

23.2.2　举例:人脸检测

Perona 和他的同事采用不同的概率模型建立了一系列的人脸检测器(见参考文献中 Burl 和 Perona,1996 年的论文,Leung,Burl 和 Perona,1995 年的论文以及 Weber,Einhaeuser,Welling 和 Perona,2000 年的论文)。每个检测器都采用了递增的对应搜索方法,并采用了不同滤波器作用于图像的输出作为图像的特征,图示 23.4 描述了 Leung,Burl 和 Perona(1995)的系统。

图 23.4　Perona 等人的人脸检测器寻找人脸的局部特征,并把这些局部特征的表现根据滤波器响应进行分类,人脸的不同模式(左眼、右眼、鼻子和嘴等)的适当组合将构成一张人脸。人脸特征点之间的关系可以用特征点之间的距离以及不同模式的类条件概率密度共同表示,因为这些距离是服从高斯分布的。这意味着,一旦某个人脸模式的位置被确定以后,就可以预测出其他人脸模式的位置。左图显示了对某些人脸模式位置的预测结果,这些特征是环绕眼睛、鼻子和嘴的特征点,这些特征点之间距离的差异,来自于不同人脸结构之间的差异。右图显示了上述人脸检测器总的性能,其中左面一列显示了最佳的检测结果,而右面一列显示了次好的检测结果,通常,次好结果的后验概率比最佳结果的后验概率小一个数量级

其中,模板间的关系用模板间的距离构成的向量来表示,在图像包含人脸的前提下,前面向量服从高维的高斯分布,如果其中某个距离未知,仍然可以采用边缘化的方法求出后验概率。这种方法的一个优势在于:如果已经找到了足够多的对应,其他对应的位置可以通过下面两步估计出来:(a)找到新对应与一些已知对应的距离,能够使距离向量的联合条件概率取最大值;(b)根据这些距离估计出新的对应可能出现的位置(如图 23.4 所示)。

23.3　利用分类器简化搜索

在人脸检测问题中,假如已经检测到了由右眼、嘴和鼻子组成的组合,如果能够从概率模型中判断出此时已经无法再检测到一个合适的左眼使得所有的对应结果构成一张脸,那么在这个时候就可以停止继续扩大对应集合。前面曾经提到,如果对应到了一个左眼,会比较上述几个对应结果的联合后验概率与相对损失的大小,通常,要求联合后验概率的值应大于某个固定的值,如果确信无法找到一个位置合适的左眼对应结果,使得上述联合后验概率大于那个固定的阈值,则可以停止当前的搜索过程。

另外，如果已经对应到了一只左眼和一只右眼，但是经过推导发现无法对应到位置合适的鼻子和嘴，则也可以停止当前的搜索过程。使用这种方法可以在一定范围内简化我们搜索对应的过程，确定合适的阈值是相当技巧性的，通常可以使用一个分类器来实现这种简化。

可以认为分类器表示了一种后验概率模型，当后验概率的值低于某个阈值的时候，分类器输出 0，否则输出一个非 0 值。使用这种方法，可以很方便地判断当前的对应集合是否能够在保证联合后验概率大于指定阈值的条件下继续被扩充，如果不能，应该中止当前的搜索过程。

23.3.1 利用投影分类器判定可接受的对应集合

只有在当前对应集合尚未完备且有迹象表明它有可能被接受的条件下，才会继续扩充当前的对应集合。假设有一个可以判定一个最终的对应集合时候能够被接受的分类器，可以利用它预测一个小的对应集合是否有可能被最终接受。这个分类器可以利用一个大样本集上的训练得出。

实际上，对那些加入任何新的对应后都不可能被接受的不完整的对应集合，可以直接放弃。这就要求分类器能把决策面投影到不同的对应集合上。例如，现在对应到了一只左眼和一只右眼，是否应该继续扩充当前的对应集合呢？此时只能得到两个眼睛模板的特征，因为在当前的对应集合中还没有其他的模板。基本的思想是，是否存在其他的模板（如鼻子和嘴），使得把这些模板加入当前的对应集合后，对应集合能够通过分类器的测试。为了回答这个问题，需要把分类器在人脸 4 个模板特征（左眼、右眼、嘴、鼻子）的空间中的决策面，投影到仅仅由左右眼两个模板特征张成的空间中，如图 23.5 所示。

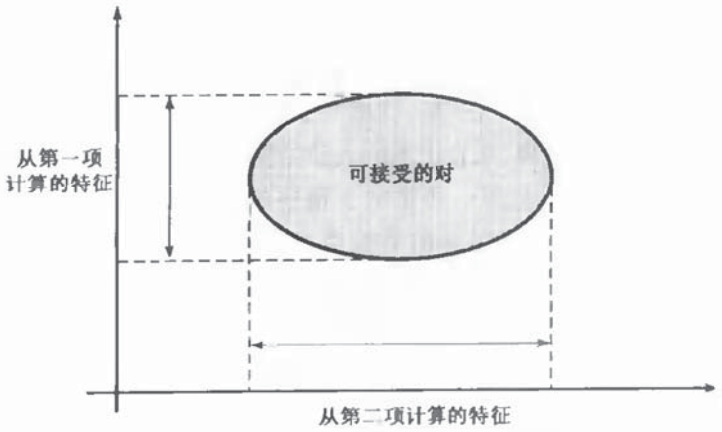


图 23.5 一个接受极大对应集合的分类器可以用于进行预测检验以判定一个小对应集合的可扩展潜力。假设希望检测包含两个元素的组，当两个元素的特征落在上图的阴影区域内部时，认为这个组是可接受的。如果第一个元素的特征不属于阴影区域在第一个特征轴上的投影区间，则可以断定包含这个特征的组肯定是不可接受的，因为不可能存在第二个元素特征，使它们的特征组合落在可接受阴影区域内。上图可以带给我们一个启示，如果把一个主分类器映射到其各个特征轴上，可以得到判断每个特征是否可接受的分类器

通过上述把分类器投影到小对应集合上的方法，可以判定哪些小对应集合在加入新的对应后有可能通过最终的分类器测试。在实际应用中，需要保证分类的投影必须准确——即原空间中的决策面，必须投影到新空间的决策面上。采用这种方法可以构造出有效与简便的分类器用来检测人和马匹（如图 23.6 和图 23.7 所示，见参考文献中 Forsyth 和 Fleck 1997 年的论文以及 Ioffe 和 Forsyth 1998 年的论文）。

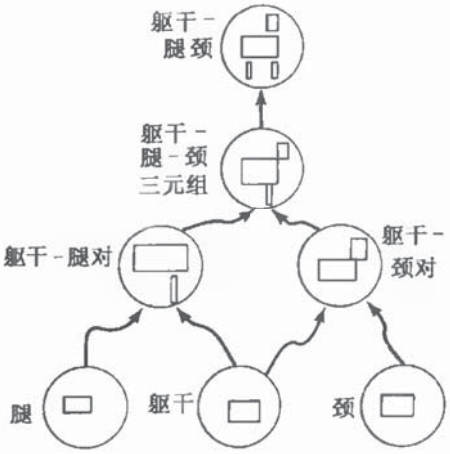


图 23.6　采用这个模型检测的马匹由两条腿、躯干和颈几部分组成。首先，需要检测图像块是否属于上述类别中的一个。接着检测躯干和腿的组合以及躯干和颈的组合，然后检测包含同一个躯干的躯干－腿组合以及躯干－颈组合是否能够组合成一个合理的躯干－腿－颈三元组。最后，检测包含同一个躯干和颈的两个三元组是否能够组成一个合理的四元组

图 23.7　可以用上面介绍的方法构造一个马匹分类器。上述图片显示了这个方法的可行性。系统首先检测出具有兽皮颜色和纹理并类似于圆柱体的图像区域（见24.1节）。一匹马由4个类似于圆柱体的部分组成——躯干、颈和双腿。而用于检测上述4部分适当组合的分类器是从样本中学习出来的。如上文所述，分类器可以被投影成为不同的分类器，每一个投影后分类器可以被用来以任意指定的顺序建立对应的组合。这个分类器的检测结果虽然不是十分精确，但是仍然很实用。上述图片是被分类器判定为包含马匹的图片，测试图片包括1086幅不包含马匹的图片和100幅包含马匹的图片，其中包含了各种各样的马匹以及很类似于马匹的其他动物。该图与图25.10是一致的（为了方便重复显示在这里）

23.3.2 举例:利用空间关系进行人体检测和马匹检测

如果希望从图片中找到人,一个自然的想法是先在图片中找到人体的各个组成部分,然后考察他们的位置关系。这里可能会用到一些特殊的方法。例如,如果要找的人体大部分是裸露的,那么可以通过皮肤检测的方法找到部分人体的组成部分(见 22.2.1 节),然后可以把已经找到的图像区域扩充为矩形(见 24.1 节)作为一个完整的组成部分。

下面就可以通过搜索对应集合的方法检测人体。例如,假设图像中的人体都是正面的,在这个假设下需要搜索 9 个对应,包括左上臂、左下臂以及躯干等。在实际应用中,全部找到 9 个对应是比较困难的,因此采用前面讨论的简化方法,它可以帮助构建一个比较可靠的人体搜索系统(见参考文献中 Ioffe 和 Forsyth 1998 年的论文)。

这个方法也可以推广到解决马匹检测的问题。首先找到一些颜色相对比较亮并且纹理较少的区域,然后把这些区域扩充为形状相对规范的区域,并认为这些区域可能是马躯体的组成部分。这里采用一个比较粗糙的模型,认为马躯体由 4 个部分组成(躯干、颈部、前肢和后肢),4 个部分的位置关系如图 23.6 所示。这里采用粗糙模型主要基于以下两点考虑:(a) 这个模型比较简单,而且便于改进;(b) 这个模型的各个部分不需要精确定位。

目前,这种方法已经被用于多种物体检测的实际应用中,如图 23.6 和图 23.7 所示,但是这里仍然存在着一些重要的问题。首先,以什么顺序搜索对应集合和测试是最好的(例如,是先对应嘴还是先对应鼻子)。可以换个方式考虑这个问题,如何投影决策面才能使投影之后的允许集体积最小(只有这样,才可能尽可能多地淘汰不可能的对应集合)。这是一个比较困难的问题,因为它涉及到不同模板特征的组合。其次,当物体类别(即图像模式)很多时如何处理。如果某个模板只在一个模式中出现,那么一旦对应到了这个模板,就可以猜测图像是那个特定的模式。但是如果模板可以属于多个模式的话,问题就会复杂很多。例如,如果希望同时检测人体和马匹,找到的图像块可能是人体的组成部分也可能是马的组成部分,那么以什么样的顺序寻找对应才能使工作量最小?最后,如果利用分类器,分类误差也会带来一些问题。

23.4 隐马尔可夫模型

到目前为止,我们引入了一些概率模型,但对识别的讨论带来的变化并不大。刚刚介绍了利用联合概率密度进行校验的方法,但仍需要同时进行对应搜索,这与第 18 章的讨论类似,而对对应搜索的简化方法也与我们在第 18 章讨论的解释树类似,然而,概率模型可以给我们更多的帮助,例如,某些概率模型可以帮助我们非常有效地找到一组最优对应。

如果希望实现一个从一段视频序列中理解美国手语的系统,这个系统必须能够推断手语者每个手势所处的内部状态。该程序根据手的位置度量来判断状态,这很可能是不很准确的,但希望能与状态有很强的依赖关系。手语的状态转移是随机的(但比较有规律),有些状态序列几乎不会出现(例如表示字母序列“wkwk”的手语状态序列就几乎不会出现)。这意味着可以既用度量又用不同手语序列出现的相对概率来推断当前手势所处的状态。

上述类型问题的要素是:

- 有一个随机变量序列(在我们的例子里是手势的状态),在给定前一个状态的情况下,下一个状态的出现是条件独立的。

- 每个时刻的随机变量有一个度量值(在我们的例子里是手的位置),而这个度量值的分布与该时刻的状态有关。

类似的问题还包括解释舞蹈运动员或武术运动员的动作。上面给出的模型对解决这一类问题是非常有用的,通常称为隐马尔可夫模型。

上面提到的随机变量序列不一定以时间为顺序,也可以按照空间关系来排列随机变量。考虑胳膊的运动,可以粗略地认为下臂的运动只与上臂的运动有关而与身体其他部位的运动无关,而上臂的运动只与躯干的运动有关而与其他部位的运动无关,等等。于是,就可以找到一组随机变量满足前面提到的条件独立条件。当然,不知道下臂的运动情况,但可以获得大量的图像,这些图像中包含了下臂运动的概率信息。

23.4.1　形式化表示

一组随机变量序列 X_n 如果满足下面条件,则称为马尔可夫链

$$P(X_n = \boldsymbol{a} \mid X_{n-1} = \boldsymbol{b}, X_{n-2} = \boldsymbol{c}, \cdots, X_0 = \boldsymbol{x}) = P(X_n = \boldsymbol{a} \mid X_{n-1} = \boldsymbol{b})$$

如果上述概率值不随 n 的变化而变化,则称其为时同态马尔可夫链。马尔可夫链几乎没有记忆性,也就是说当前的状态只依赖于前一个状态,而与再前面的状态历史就无关了。马尔可夫性可以帮助我们进行建模,因为很多实际问题中的变量都具有马尔可夫性,此外,对具有马尔可夫性的变量,可以应用很多简单的算法。离散的马尔可夫链和连续的马尔可夫链只在表示上有一些差别,这里只讨论离散马尔可夫链。

假设有一个离散的状态空间,尽管有限维的状态空间更便于想像,但状态空间的维数也可以无限。记状态空间中的元素为 s_i,并假设总共有 k 个元素。现在假设在这个状态空间上建立了一个时不变马尔可夫链,我们记:

$$P(X_n = s_j \mid X_{n-1} = s_i) = p_{ij}$$

由于上面的概率与 n 无关,因此只记为 p_{ij},可以用一个矩阵$\mathcal{P}$表示上面所有的概率,其中第 i 行、第 j 列的元素为 P_{ij},这个矩阵称为状态转移矩阵,它刻画了马尔可夫链的运动规律。假设马尔可夫链的初始状态分布为 $P(X_0 = s_i) = \pi_i$,用一个向量 $\boldsymbol{\pi}$ 表示初始分布,其中第 i 个元素为 π_i,于是有

$$\begin{aligned} P(X_1 = s_j) &= \sum_{i=1}^{k} P(X_1 = s_j \mid X_0 = s_i) P(X_0 = s_i) \\ &= \sum_{i=1}^{k} P(X_1 = s_j \mid X_0 = s_i) \pi_i \\ &= \sum_{i=1}^{k} p_{ij} \pi_i \end{aligned}$$

即状态 X_1 的分布可表示为$\mathcal{P}^T\pi$,类似地,状态 X_n 的分布可表示为$(\mathcal{P}^T)^n\boldsymbol{\pi}$。对所有的马尔可夫序列,至少存在一个分布 $\boldsymbol{\pi}^s$,使得 $\boldsymbol{\pi}^s = \mathcal{P}^T\boldsymbol{\pi}^s$,这个分布称为马尔可夫链的平稳分布。我们可以用状态图来表示一个马尔可夫序列,状态图中的边具有方向和权值,状态图中的每个结点表示一个状态,有向边上的权值则表示了转移概率(如图 23.8 所示)。

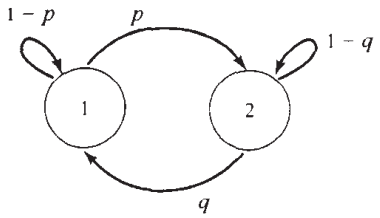


图 23.8 一个简单的、两状态马尔可夫链。其中，从状态 1 到状态 2 的转移概率为 p，状态 1 到状态1的概率为$1-p$，以此类推。我们可以用一个转移矩阵来描述这个马尔可夫链，其平稳分布是$(q/(p+q), p/(p+q))$。这意味着：如果p非常小而q近似于1的话，该马尔可夫链大部分时间会停留在状态1上；如果p和q都非常小的话，该马尔可夫链会在一个状态上停留较长一段时间然后转移到另一个状态，再在另一个状态上停留较长一段时间

如果能直接观察到 n 时刻的状态 X_n，则推断整个链是容易的，但这不是一个实用性好的模型。在实际应用中，往往无法直接得到每个时刻的状态值，于是我们得到了下面的模型。对于一个马尔可夫序列的各个时刻，只能观测到另一个随机变量 Y_n，这个 Y_n 的分布与马尔可夫序列第 n 个状态有关，有 $P(Y_n \mid X_n = s_i) = q_i(Y_n)$，可以用一个矩阵$\mathcal{Q}$表示所有的 $q_i(Y_n)$。于是，为了确定一个隐马尔可夫模型，需要给出转移概率矩阵$\mathcal{P}$，状态值和观测值之间的概率关系矩阵$\mathcal{Q}$和初始分布 π，一个隐马尔可夫模型可表示为$(\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})$。我们仍然假设状态空间中有 k 个元素。

23.4.2 隐马尔可夫模型的计算

假设在一个离散的状态空间上讨论马尔可夫链，需要解决以下两个重要问题：

- **推理**：需要判定当前的一组观测值代表了什么样的状态组合，例如，需要判定当前舞蹈者的动作含义和手语者要表达的意思。
- **拟合**：需要根据已有的观测值拟合出一个合适的隐马尔可夫模型。

上面的每个问题都有标准的有效解决办法。

网格模型 假设有 N 个来自于一个隐马尔可夫模型的观测值 $\boldsymbol{Y}_i$，可以用一个网格模型来组织这些观测值。网格模型是一个边具有方向和权值的状态图，状态图包括 N 列，每一列有 k 个结点(表示状态空间中所有 k 个状态)，而每一列又对应一个观测值。给每个结点标记一个权值，对应观测值 $\boldsymbol{Y}_j$ 的第 j 列中对应状态 $\boldsymbol{X}_i$ 的第 i 个结点的权值为 $\log q_i(\boldsymbol{Y}_j)$。

用下面的方法将状态图中的列与列连接起来。如果状态空间中状态 $\boldsymbol{X}_k$ 到 $\boldsymbol{X}_l$ 的转移概率 p_{kl} 不为 0，则把对应观测值 $\boldsymbol{Y}_j$ 的第 j 列中表示状态 $\boldsymbol{X}_k$ 的结点和对应观测值 $\boldsymbol{Y}_{j+1}$ 的第 $j+1$ 列中表示状态 $\boldsymbol{X}_l$ 的结点连接起来，这表示这两个状态之间存在着状态转移的可能，这条边的权值为 $\log p_{kl}$。图示 23.9 显示了如何用网格图表示一个隐马尔可夫模型(HMM)。

网格模型具有一个有趣的性质，每条有向路径表示了一个可能的状态序列，由于每个结点的权值为这个结点对应状态的概率的对数，而每条有向边的权值为这个状态转移概率的对数，因此，如果把某条路径上结点的权值和有向边的权值相加，就可以得到某个状态序列出现的条件概率(在观测值确定的条件下)的对数。这个算法可以方便地帮助我们找出当观测值确定时，出现的条件概率最大的状态序列，它被称为动态规划或 Viterbi 算法。

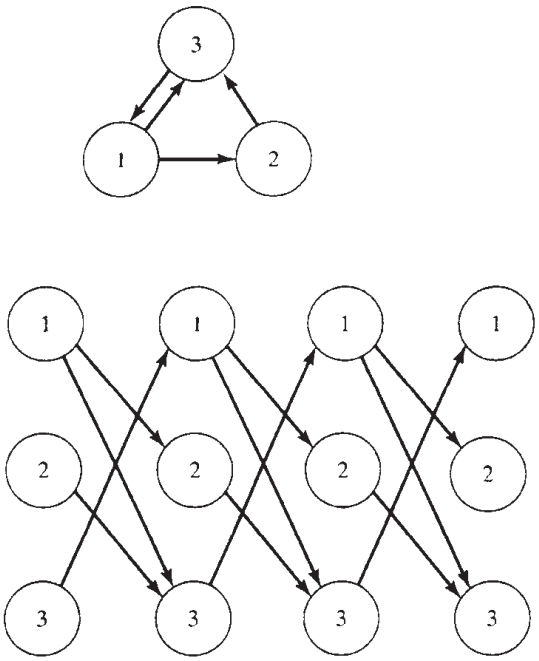


图 23.9　图的上部显示了一个简单的状态转移模型，下部则显示了这个状态转移模型相应的网络模型。实际上，网络模型上的每一条路径表示一个可能的4步状态转移序列。可以用相应状态转移概率的对数值作为每条弧的权值，用相应状态吸收概率的对数作为每个结点的权值。为了简单起见，上图中并没有标出权值

我们从网格图的最后一列开始计算，首先计算出只包含一个结点的路径的概率对数，也就是最后一列结点的权值。接下来再考虑包含两个状态的路径，这些路径从网格图的倒数第二列开始。对于这一列的所有结点，我们知道离开这个结点的每条边的权值，也知道每条边另一端结点的权值，因此可以找到一条路径使得上述权值的和最大，这条路径是离开这个结点后可能性最大的路径，把离开某个结点可能性最大的路径的权值(包括边的权值和另一端结点的权值)加到这个结点的权值上，作为这个结点的新权值，这个值表示从这个结点出发后面的路径能达到的最大值，称为结点值。

一旦计算出倒数第二列所有结点的结点值，就可以继续计算倒数第三列的结点值。对这一列的每个结点，我们知道离开它的边的权值，也知道边到达结点的结点值，可以找出上述权值和(边权值加上到达结点的结点值)最大的路径，把这个和加到倒数第三列的出发结点上作为这个结点的结点值。可以重复上面的步骤，直到计算出第一列结点的结点值。第一列结点中的最大结点值就是在确定观测值条件下状态序列的最大条件似然值。

同时可以得到取得最大似然值的路径，计算一个结点的结点值时，把最优路径之外的所有边除去，这样，当计算到第一列结点时，留下的路径就是满足似然值最大的路径。图示 23.10 展示了这个简单而有效的算法。

下面，我们将更加正式地讨论动态规划问题，可以通过考察未来或考察过去两种方法找到最优路径。在建立网格模型的时候，采用了考察未来的方法，而在下面正式的讨论中，将采用考察过去的方法。

推理与动态规划　假设有 $n+1$ 个时刻的观测值 $\{Y_0, Y_1, \cdots, Y_n\}$，要做的是找到 $n+1$ 个时刻的观测序列 $\boldsymbol{S} = \{S_0, S_1, \cdots, S_n\}$，使得下面的条件概率最大

$$P(S \mid \{Y_0, Y_1, \cdots, Y_n\}, (\mathcal{P}, \mathcal{Q}, \pi))$$

它等价于使下面的联合条件概率取值最大

$$P(S, \{Y_0, Y_1, \cdots, Y_n\} \mid (\mathcal{P}, \mathcal{Q}, \pi))$$

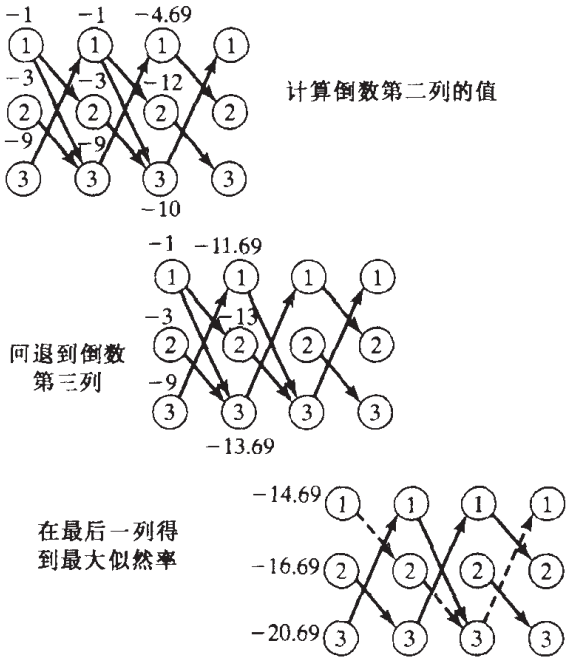


图 23.10 求出图 23.9 中所示网络模型(或其他网络模型)的最佳路径是相对简单的。假设所有1号结点的对数概率值为 - 1,所有2号结点的对数概率值为 - 3,所有3号结点的对数概率值为 - 9,同时假设离开每个结点的概率是一样的。计算过程是这样的,倒数第二列的每个结点的权值是该结点的权值加上离开这个结点的最佳路径的权值,这个计算是非常简单的。接下来,用类似的方法继续计算倒数第三列结点的权值,以此类推,直到计算到网络模型中的第一列结点。这列结点的权值表示了由该结点出发的最佳路径的权值的对数。由于已经找到了每条路径所经过的结点,因此也就找到了全网络模型的最佳路径(如上图虚线所示)

有一个标准的算法专门用来解决这类问题,称为 Viterbi 算法。我们在 k 个状态中搜索一条包含 $n+1$ 个元素的路径,一共有 k^{n+1} 种可能的选择(这里假设任意两个状态间的转移概率不为0,更规范的说法是,共有 $O(k^{n+1})$ 种可能的选择)。我们无法考察每一条路径,实际上也不需要这么做。采用如下的方法:如果对于状态集合中的任意一个状态 s_l,已经找到以这个状态 s_l 结束的 n 步状态序列 p_l(p_l 是一个 n 步状态序列,$l=1,\cdots,k$)满足使目标联合条件概率最大,则满足使目标联合条件概率最大的 $n+1$ 步状态序列必定由 p_l 中的一个加上另外的一步组成,要做的只是找出这另外的一步。

可以把上述问题理解为一个归纳问题。对第 $t-1$ 个状态 S_{t-1} 的任意一个可能的值 j,我们知道以 $S_{t-1}=j$ 结束的最优序列,记为:

$$\delta_{t-1}(j) = \max_{S_0, S_1, \cdots, S_{n-2}} P(\{S_0, S_1, \cdots, S_{t-1}=j\}, \{Y_0, Y_1, \cdots, Y_{t-1}\} \mid (\mathcal{P}, \mathcal{Q}, \pi))$$

于是有

$$\delta_t(j) = \left(\max_i \delta_{t-1}(i) P_{ij}\right) q_j(Y_t)$$

不仅需要得到最大概率值，还需要找出取去最大概率值的路径，为此定义另外一个变量

$$\psi_t(j) = \arg\max\left(\delta_{t-1}(i)P_{ij}\right)$$

$\psi_t(j)$记录了 $s_t = j$ 最佳路径在 $t-1$ 次度量时是哪一个状态。因此从$\psi_t(j)$我们知道，t 时刻$s_t = j$ 最佳路径是从 $t-1$ 时刻哪个状态过来的，而对 $t-1$ 次度量又可继续回溯，例如 $\psi_t(j) = i$，则回溯$\psi_{t-1}(i)$。

算法 23.1　Viterbi 算法找出隐马尔可夫模型中联合条件概率的最大值以及取最大值的状态路径。这里 δ 和ψ是为了方便起见引入的两个标记变量，p^* 是联合条件概率的最大值，q_t^* 是 t 步最优路径

1. 初始化：

$$\delta_0(j) = \pi_j b_j(Y_0) \quad 1 \leqslant j \leqslant N$$

$$\psi_0(j) = 0$$

2. 递归：

$$\delta_t(j) = \left(\max_i \delta_{t-1}(i)P_{ij}\right) q_j(Y_t)$$

$$\psi_t(j) = \arg\max\left(\delta_{t-1}(i)P_{ij}\right)$$

3. 结束：

$$p^* = \max_i (\delta_n(i))$$

$$q_n^* = \arg\max_i (\delta_n(i))$$

4. 路径回溯：

$$q_t^* = \psi_{t+1}(q_{t+1}^*)$$

利用 EM 算法拟合隐马尔可夫模型　假设有一个数据集 $\boldsymbol{Y}$ 被认为来自于一个隐马尔可夫模型，那么如何找到一个合适的隐马尔可夫模型呢？当然希望能找到一个尽量符合数据集特性的模型，这里采用第 16 章介绍过的一种最大期望算法。假设已经有了一个隐马尔可夫模型$(\mathcal{P},\mathcal{Q},\boldsymbol{\pi})$，我们希望用这个模型和数据集估计出一个新参数的模型$(\bar{\mathcal{P}},\bar{\mathcal{Q}},\bar{\boldsymbol{\pi}})$，这里采用如下方法。我们不加证明地给出如下一个事实：在估计过程中只可能出现下列两种情况之一，$\mathcal{P}(\boldsymbol{Y}|(\bar{\mathcal{P}},\bar{\mathcal{Q}},\bar{\boldsymbol{\pi}})) > \mathcal{P}(\boldsymbol{Y}|(\mathcal{P},\mathcal{Q},\boldsymbol{\pi})$或$(\bar{\mathcal{P}},\bar{\mathcal{Q}},\bar{\boldsymbol{\pi}}) = (\mathcal{P},\mathcal{Q},\boldsymbol{\pi})$。

根据下面的式子对模型参数进行更新

$$\overline{\pi_i} = \text{expected frequency of being in state } s_i \text{ at time 1}$$

$$\overline{P_{ij}} = \frac{\text{expected number of transitions from } s_i \text{ to } s_j}{\text{expected number of transitions from state } s_i}$$

$$\overline{q_i(k)} = \frac{\text{expected number of times in } s_j \text{ and observing } Y = y_k}{\text{expected number of times in state } s_j}$$

需要计算上面的表达式，首先，需要计算下式的值

$$P(X_t = s_i, X_{t+1} = s_j \mid \boldsymbol{Y}, (\mathcal{P}, \mathcal{Q}, \pi))$$

把上式的结果记做$\xi_t(i,j)$，一旦计算出$\xi_t(i,j)$，则有

$$\begin{aligned}
\text{从 } s_i \text{ 到 } s_j \text{ 转移数目的期望值} &= \sum_{t=0}^{n} \xi_t(i,j); \\
\text{在 } s_i \text{ 次数的期望值} &= \text{从 } s_i \text{ 转移的期望值} \\
&= \sum_{t=0}^{n} \sum_{j=1}^{k} \xi_t(i,j); \\
\text{在 0 时刻状态为 } s_i \text{ 的期望频率} &= \sum_{j=1}^{k} \xi_0(i,j); \\
\text{在 } s_i \text{ 次数并观察到}(Y = y_k)\text{ 的期望值} &= \sum_{t=0}^{n} \sum_{j=1}^{k} \xi_t(i,j)\delta(Y_t, y_k);
\end{aligned}$$

其中，$\delta(u,v)$当 u、v 相等时取 1，否则取 0。

为了计算$\xi_t(i,j)$，需要两个中间变量，分别称为前序变量和后序变量。前序变量是 $\alpha_t(j) = P(Y_0, Y_1, \cdots, Y_t, X_t = s_j \mid (\mathcal{P}, \mathcal{Q}, \pi))$，后序变量是 $\beta_t(j) = P(\{Y_{t+1}, Y_{t+2}, \cdots, Y_n\} \mid X_t = s_j, (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi}))$。

如果已经计算出这些变量的值，则有

$$\begin{aligned}
\xi_t(i,j) &= P(X_t = s_i, X_{t+1} = s_j \mid \boldsymbol{Y}, (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \\
&= \frac{P(\boldsymbol{Y}, X_t = s_i, X_{t+1} = s_j \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi}))}{P(\boldsymbol{Y} \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi}))} \\
&= \frac{\left\{\begin{array}{c} P(Y_0, Y_1, \cdots, Y_t, X_t = s_i \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \\ \times P(Y_{t+1} \mid X_{t+1} = s_j, (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \\ \times P(X_{t+1} = s_j \mid X_t = s_i, (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \\ \times P(Y_{t+2}, \cdots, Y_N \mid X_{t+1} = s_j, (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \end{array}\right\}}{P(\boldsymbol{Y} \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi}))} \\
&= \frac{\alpha_t(i) p_{ij} q_j(Y_{t+1}) \beta_{t+1}(j)}{P(\boldsymbol{Y} \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi}))} \\
&= \frac{\alpha_t(i) p_{ij} q_j(Y_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) p_{ij} q_j(Y_{t+1}) \beta_{t+1}(j)}
\end{aligned}$$

前序变量和后序变量都可以通过归纳的办法计算出来，可以按照如下的方法计算前序变量 $\alpha_t(j)$：

$$\begin{aligned}
\alpha_0(j) &= P(Y_0, X_0 = s_j \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \\
&= \pi_j q_j(Y_0); \\
\alpha_{t+1}(j) &= P(Y_0, Y_1, \cdots, Y_{t+1}, X_{t+1} = s_j \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \\
&= P(Y_0, Y_1, \cdots, Y_t, X_{t+1} = s_j \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) P(Y_{t+1} \mid X_{t+1} = s_j) \\
&= \sum_{l=1}^{k} \big[P(Y_0, Y_1, \cdots, Y_t, X_t = s_l, X_{t+1} = s_j \mid (\mathcal{P}, \mathcal{Q}, \boldsymbol{\pi})) \\
&\quad \times P(Y_{t+1} \mid X_{t+1} = s_j) \big]
\end{aligned}$$

$$= \left[\sum_{l=1}^{k} P(Y_0, Y_1, \cdots, Y_t, X_t = s_l \mid (\mathcal{P}, \mathcal{Q}, \pi)) P(X_{t+1} = s_j \mid X_t = s_l)\right]$$
$$\times P(Y_{t+1} \mid X_{t+1} = s_j)$$
$$= \left[\sum_{l=1}^{k} \alpha_t(l) p_{lj}\right] q_j(Y_{t+1}) \qquad 1 \leqslant t \leqslant n-1$$

后序变量的归纳计算方法如下：

$$\beta_n(j) = P(\text{no further states} \mid X_n = s_j, (\mathcal{P}, \mathcal{Q}, \pi))$$
$$= 1;$$
$$\beta_t(j) = P(\{Y_{t+1}, Y_{t+2}, \cdots, Y_n\} \mid X_t = s_j, (\mathcal{P}, \mathcal{Q}, \pi))$$
$$= \sum_{l=1}^{k} \left[P(\{Y_{t+1}, Y_{t+2}, \cdots, Y_n\}, X_t = s_l \mid X_{t+1} = s_j, (\mathcal{P}, \mathcal{Q}, \pi))\right]$$
$$= \left[\sum_{l=1}^{k} P(X_t = s_l, Y_{t+1} \mid X_{t+1} = s_j)\right] P(\{Y_{t+2}, \cdots, Y_n\} \mid X_{t+1} = s_j, (\mathcal{P}, \mathcal{Q}, \pi))$$
$$= \left[\sum_{l=1}^{k} p_{jl} q_l(Y_{t+1})\right] \beta_{t+1}(j) \qquad 1 \leqslant t \leqslant k-1$$

算法 23.2 描述了前面介绍的简单拟合算法。

算法 23.2　根据数据集 $\boldsymbol{Y}$ 拟合隐马尔可夫模型的 EM 算法。首先假设一个模型 $(\mathcal{P},\mathcal{Q},\boldsymbol{\pi})_i$，然后利用递归算法逐步计算实际模型的参数。这种递归算法保证了概率 $P(\boldsymbol{Y} \mid (\mathcal{P},\mathcal{Q},\boldsymbol{\pi}))$ 收敛到它的最大值。

如果 $(\mathcal{P},\mathcal{Q},\boldsymbol{\pi})_{i+1}$ 与 $(\mathcal{P},\mathcal{Q},\boldsymbol{\pi})_i$ 相等，则结束，否则执行下面的步骤

采用算法 23.4 和 23.5 计算前序变量 α 和后序变量 β

计算 $\xi_t(i,j) = \dfrac{\alpha_t(i) p_{ij} q_j(Y_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) p_{ij} q_j(Y_{t+1}) \beta_{t+1}(j)}$

利用算法 23.3 更新模型参数，记新参数为 $(\mathcal{P},\mathcal{Q},\boldsymbol{\pi})_{i+1}$

end

算法 23.3　在拟合隐马尔可夫模型时按照下面的方法更新模型参数

$$\overline{\pi_i} = \text{expected frequency of being in state } s_i \text{ at time } 0$$
$$= \sum_{j=1}^{k} \xi_0(i,j)$$
$$\overline{p_{ij}} = \frac{\text{expected number of transitions from } s_i \text{ to } s_j}{\text{expected number of transitions from state } s_i}$$
$$= \frac{\sum_{t=0}^{n} \xi_t(i,j)}{\sum_{t=0}^{n} \sum_{j=1}^{k} \xi_t(i,j)}$$

（续）

$$\overline{q_i(k)} = \frac{\text{expected number of times in } s_i \text{ and observing } Y = y_k}{\text{expected number of times in state } s_i}$$

$$= \frac{\sum_{t=0}^{n}\sum_{j=1}^{k}\xi_t(i,j)\delta(Y_t,y_k)}{\sum_{t=0}^{n}\sum_{j=1}^{k}\xi_t(i,j)}$$

其中，$\delta(u,v)$当 u、v 相等时取 1，否则取 0。

算法 23.4 在拟合隐马尔可夫模型时按照下面的方法计算前序变量

$$\alpha_0(j) = \pi_j q_j(Y_0)$$

$$\alpha_{t+1}(j) = \left[\sum_{l=1}^{k}\alpha_t(l)p_{lj}\right]q_j(Y_{t+1}) \qquad 0 \leqslant t \leqslant n-1$$

算法 23.5 在拟合隐马尔可夫模型时按照下面的方法计算后序变量

$$\beta_n(j) = 1$$

$$\beta_t(j) = \left[\sum_{l=1}^{k}p_{jl}q_l(Y_{t+1})\right]\beta_{t+1}(j) \qquad 0 \leqslant t \leqslant n-1$$

23.4.3 隐马尔可夫模型的变型

我们还没有讨论隐马尔可夫模型的拓扑图。这可以是一张全向图（任意两个结点间都有两条不同方向的边），但是并不一定需要这样一张全向图，因为如果采用全向图，将有很多的参数需要估计。在实际应用中，会采用其他一些有效的拓扑图（如图 23.11 所示）。

这里介绍左右模型，也称为 Bakis 模型，它具有下面的特性。从状态 i 出发，只能到达比 i 大的状态 j，换句话说，对任意的 $j < i$，$p_{ij} = 0$。此外，$\pi_1 = 1$，并且对任意 $i \neq 1$，有 $p_i = 0$。这意味着状态转移矩阵$\mathcal{P}$是一个上三角矩阵，此外，N 状态的左右模型一旦到达第 N 个状态，它将永远停留在第 N 个状态。实际上，这个模型给事件发生的顺序做了一些规定，这些规定对于用隐马尔可夫模型去模拟各种运动信息是非常有帮助的。

然而，采用普通的左右模型意味着大量的事件可能被错过，这可能并不一定符合一个运动模型的特点，因为可以错过一两个观测值或状态值，但是不能错过大量的状态。因此，经常会对普通的左右模型做一些改进，加入如下的约束条件：对任意 $j > i + \delta$，$p_{ij} = 0$，这里 δ 是一个小正整数（通常取 2）。

令人欣慰的是，采用改进拓扑结构后的隐马尔可夫模型并不影响前面介绍算法的实现。需要保证参数估计算法的 0 不变性，也就是说，如果从一个一般的模型开始，这个模型的转移矩阵中的某些位置为 0，那么计算出一组新的模型参数后，应该保证新的状态转移矩阵在同样的位置也为 0。

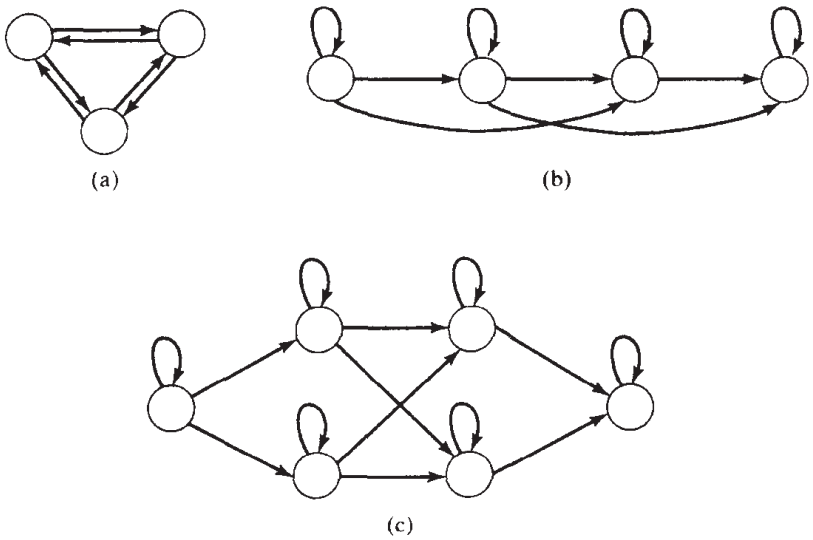


图 23.11　很多其他拓扑结构的隐马尔可夫模型已经成功地应用于实际中。最上方显示的是各态同性的模型，也称为全连接模型。中间显示的是一个4状态的左右模型。这个模型有如下一个约束条件：对任意 $j > i + 2$，有 $p_{ij} = 0$。最下面显示的是一个6状态并行路径的左右模型，这个模型实际上包括两个左右模型，中间有一个可选择的开关

23.5　应用：基于隐马尔可夫模型的手语理解

手语是用手势表示意思的语言，对不熟悉手语的普通人来说，理解手语是非常困难的。如果有一套可以翻译手语的系统，则将大大方便聋哑人与普通人之间的交流。理解手语问题与理解语音语言问题是非常类似的，而人们在理解语音语言问题的研究上已经取得了相当大的进展。

隐马尔可夫模型被成功地应用于理解语音语言，其中模型的状态值描述了语音语言系统，而观测值则是对不同语音的度量值。其实，每一个词都是一个小的隐马尔可夫模型，而这些小的模型又通过一个大的语言模型联系起来，这个语言模型实际上是一个大的隐马尔可夫模型，它给出了在某些词出现的情况下，另外一些词出现的概率。首先得到一些对语音的度量值，然后应用一些对语言模型的推理算法就可以得出一个句子。这个模型应用在理解语音语言的问题上是非常成功的，因此自然而然想到把这个模型推广到理解手语的问题上。

人类的手势与语音在某种程度上是十分相似的，它们都可以认为是事件的序列，可以通过某种方法度量这些事件，但无法确定这些事件。虽然无法证明人类的手势符合隐马尔可夫模型中条件独立的假设，但是同样没有证明人类的语音符合那个假设，而隐马尔可夫模型在理解语音语言时是十分有效的，因此在理解手语的问题上，隐马尔可夫模型也值得一试。

我们将介绍基于隐马尔可夫模型的手语理解系统，实际上这个系统也可以应用于理解一般的手势。例如，如果希望识别开窗和关窗两个动作，并认为手向远离身体的方向运动是开窗，向靠近身体的方向移动是关窗，则也可以把这两个固定的动作视为手语中的两种语义。

有许多基于隐马尔可夫模型的手语理解系统，通常，单词模型是一个增加约束条件的左右模型，使它不能跳过过多的状态（对任意 $j > i + \Delta$，$|\Delta|$ 较小，$p_{ij} = 0$）。每个单词模型只有少数几个状态，在状态图上，每个状态有一个自环，表示每个状态上可以停留任意个周期。此外，单

词模型中还包括跳过若干个状态的状态转移，表示单词模型可以很快地跨越若干状态。目前，还有两个问题需要解决，一是需要确定几种形式的语言模型把单词模型连接起来，二是需要明确度量什么。

23.5.1 语言模型：由单词组成的句子

要理解手语，就不仅仅需要分离手语表达的每一个单词，更需要识别它要表达的意思，因此需要确定单词排列的方式。语言模型是一个单词模型的排列规范，它给出了一个句子的隐马尔可夫模型。一个最简单的语言模型认为当前出现的单词与前一个出现的单词独立。语言模型可以用图示的方法表示出来，如图 23.12 所示。在这个模型中，由一个起始状态出发，根据不同的单词模型匹配一个单词，然后或者结束或者返回起始状态去匹配另外一个单词。需要从这个状态图中找到一个极值路径，这里仍然可以采用 Viterbi 算法。

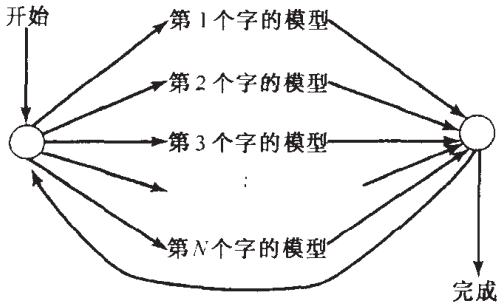


图 23.12 如果给词汇表中的每一个单词建立一个隐马尔可夫模型，然后把这些模型用独立的出现概率组合起来，就得到一个语言模型。这是一个最简单的模型，模型中单词是独立的，而且句子的长度没有限制。虽然模型中没有考虑语法，但是它仍然可以帮助我们进行简单的推理

在实际应用中，需要更复杂的语言模型，如果对英语采用上面的简单语言模型，那么可能出现一个句子中连续出现多个“and”或多个“the”的情况，但是复杂的语言模型会带来很多计算上的问题。首先要确定一个二重关联语言模型，在这个语言模型中，一个单词出现的概率与前一个出现单词有关。图示 23.13 描述了这样一个模型。更复杂的语言模型可能包含三重关联（此时一个单词出现的概率与前面两个出现的单词有关）或更复杂的图（此时一个单词出现的概率将与更前面出现的单词有关）。采用这种方法的难点在于，对于有相当词汇量的语言来说，模型状态的数量可能十分庞大，因此 Viterbi 算法搜索可能需要相当长的时间。采用某些技术可以简化搜索过程，但这已经超出了本书的讨论范围，有兴趣的读者可以参阅参考文献中 Jelinek（1999）的一本书（这本书的第 5 章给出了一个简化搜索的方法）或 Manning 和 Schütze 两人 1999 年的论文。

特征与性能水平 一些人已经编写了手语理解的程序，其中 Starner 曾经利用不同的特征提取方法编写了几个程序（见参考文献中 Starner，Weaver 和 Pentland，1998 年的论文）。最简单的方法要求手语者右手戴一只黄手套，左手戴一只红手套，这样做是为了区分手和身体的其他部分；另一种方法是根据皮肤的颜色来找到手，这要求手的附近不再有其他与皮肤相近的颜色。从图像中找出指定颜色的像素（红色和黄色的像素或皮肤颜色的像素），然后检测这个像素周围 8 个像素的颜色是否也是指定颜色，重复这个步骤直至找到我们期望的表示左右手的两个像素块。

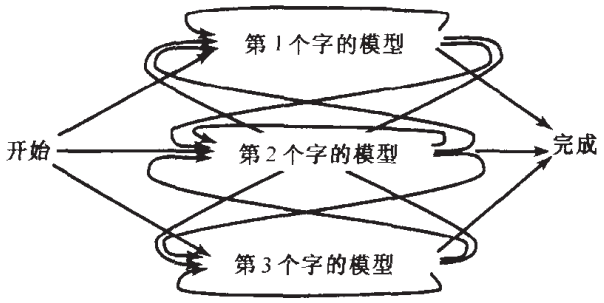


图 23.13　在二重关联的语言模型中，一个单词出现的概率与它前面出现的单词有关系。这个语言模型仍然可以通过组合单词的隐马尔可夫模型得到。此时语言模型的拓扑结构相对复杂，为了简单起见，这里只显示了三个单词

这两个像素块可以提供不同的特征。两个像素块的重心可以作为两个特征，两个像素块的重心相对于前一帧图像的偏移又可以作为两个特征，像素块的方向和面积可以通过下面的二阶矩阵度量，并作为一个特征

$$\begin{pmatrix} \int x^2\,dx\,dy & \frac{1}{2}\int xy\,dx\,dy \\ \frac{1}{2}\int xy\,dx\,dy & \int y^2\,dx\,dy \end{pmatrix}$$

其中，上述矩阵两个特征值的比表示了像素块的离心率，大的特征值表示了像素块的主方向长度，而大的特征值所对应的特征向量的方向则表示了像素块的方向。

Starner 的系统工作在一个包括 40 个词汇的手语集上，隐马尔可夫模型的拓扑图已经给出，而模型参数的估计采用了第 16 章介绍的 EM 算法。系统对分离单词的识别和由 5 个单词组成句子（句子的结构限定为代词 + 动词 + 名词 + 形容词 + 代词）的识别分别进行了测试，识别率大约在 90% 左右，与特征提取的方式和识别内容有关。见图 23.14 所示。

Vogler 和 Metaxas（见参考文献中他们 1998 年和 1999 年的论文）开发了另一套手语理解系统，这个系统通过固定在人身上的物理传感器或一个能够对手臂进行精确定位的三视角摄像机系统获取手臂位置信息。系统采用独立的语言模型对包括 53 个单词的手语集进行了测试，识别率也达到了 90%。

展望　我们介绍的系统采用的都是比较简单的语言模型，测试语言集包含的词汇量也比较少。虽然如此，隐马尔可夫模型仍然被认为是解决手语理解问题的首选模型。我们还无法确定采用简单的特征能否识别复杂的手语，一个可能的特征提取改进方向是在特征中加入对手指运动的估计。

好的语言模型和合适的推理算法是现代语音语言识别系统成功的关键。获取这样的语言模型往往需要统计每个单词在特定的语境中出现的频率（例如三重关联频率）。为了精确估计包括某些不常见语境中单词出现的相对频率，需要数量庞大的语言数据。例如，Jelinek（见参考文献中他 1999 年的论文）指出，如果要估计 15 000 个单词出现的频率，至少需要 640 000 个单词的语言数据。这意味着某些单词是十分常用的，而在一个小的语言数据集上进行这种估计通常是十分不准确的。语音语言的识别问题和自然语言的识别问题在进行这些统计的时候都遇到过很多难题，有时不得不采用一些复杂的方法解决这些难题（见参考文献中Jelinek 1999 年的论文以及Manning和Schütze 1999 年的论文）。在将来手语理解问题的研究中，我们可能需要借鉴这些方法并把它们移植到视觉领域。

图 23.14 Starner 和他的同事应用隐马尔可夫单词模型建立了一个手语识别系统。上图显示了一个手语者在检测该系统，该图是由房顶上照相机拍摄的

23.6 应用：基于隐马尔可夫模型的人体检测

在解决手语识别问题时，采用隐马尔可夫模型是比较直观的，因为手势是按照一定的约束条件随机出现的，但是在人体识别问题中采用隐马尔可夫模型的原因就不那么直观了。实际上，隐马尔可夫模型的本质并不是一个按时间顺序的序列，而是它的条件独立特性，即在给定状态 X_i 的条件下，状态 X_{i+1} 与 X_i 之前的状态是条件独立的。

这类条件独立特性可能出现在很多场合，人体检测问题就是其中之一。假设出现在图像中的人体类似于一个洋娃娃，可粗略地认为由 9 个部分组成（分别是左上、左下、右上、右下臂和腿以及躯干），每个部分可以表示为一个矩形，称为一个组成块。通常，可以认为左下臂的位置只与左上臂有关而与其他部位无关，而左上臂的位置只与躯干有关而与其他部分无关，可以简单地把上述假设推广到右臂和左右腿，这就出现了一个隐马尔可夫模型。为了表示这个模型，我们记左上臂地位置为 X_{lua}，并以此类推，于是有

$$
\begin{aligned}
&P(X_t, X_{lua}, X_{lla}, X_{rua}, X_{rla}, X_{lul}, X_{lll}, X_{rul}, X_{rll}) \\
&\quad = P(X_t)P(X_{lua} \mid X_t)P(X_{lla} \mid X_{lua})P(X_{rua} \mid X_t)P(X_{rla} \mid X_{rua})P(X_{lul} \mid X_t) \\
&\qquad \times P(X_{lll} \mid X_{lul})P(X_{rul} \mid X_t)P(X_{rll} \mid X_{rul})
\end{aligned}
$$

可以用一个树形图表示上面的依赖关系（如图 23.15 所示）。

现在有一些包含人体的图像，假设人体的每个部位都可能处于一个数量有限的状态中的一个，譬如把左上臂处于某个特定状态这一事件记为 $X_{lua} = \boldsymbol{x}_{lua}$，在这个事件发生的条件下，所获得的相对度量的条件概率为 $P(M = \boldsymbol{m} \mid X_{lua} = \boldsymbol{x}_{lua})$，以此类推。通常，在图像中能匹配到多个人体的组成块，其中有些的确来自于人体，有些则由图像噪声造成。假设在图像中总共匹配

到了 N_s 个组成块，再假设来自于噪声的组成块与来自于人体的组成块是独立的，然后就可以对所有来自于人体的组成块计算似然函数。记 $\{i_1, \cdots, i_9\}$ 表示第 i_1 个组成块为人体躯干，第 i_2 个组成块为左上臂……第 i_9 个组成块为右小腿，其他的组成块都来自于噪声，再记 $\boldsymbol{m}_{i_k}$ 表示对第 i_k 个组成块的一个度量值。

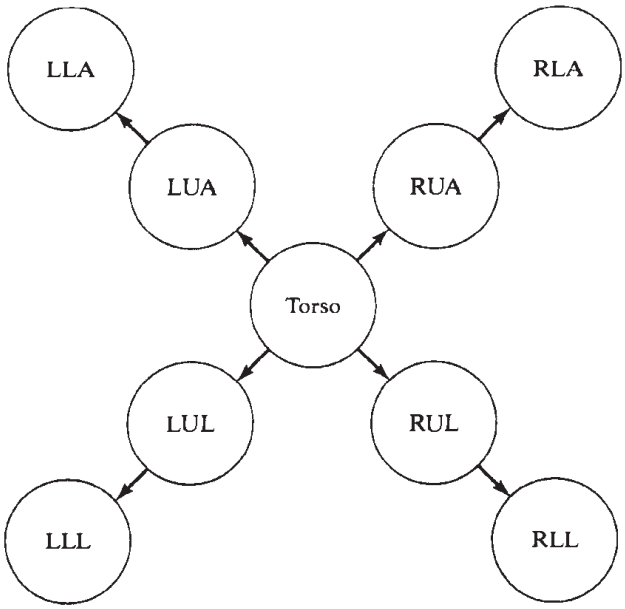


图 23.15　可以把人体想像成为由隐马尔可夫模型组成，上图所示的树状结构是人体模型一个可能的结构：其中躯干是树的根结点，由躯干产生了不同的分支，包括手臂和腿。所有上臂和大腿位置的概率在给定躯干位置的前提下是条件独立的。而在给定大腿位置的前提下，小腿位置的概率与身体其他部分的位置是条件独立的，以此类推。可以把这些条件独立的性质表示在一幅有向图上，图的结点表示随机变量，有向边指向的结点的状态直接依赖于有向边出发的结点。如果按照上述规则构造的有向图是一棵树，就得到了一个隐马尔可夫模型。需要注意这幅图的含义与图23.8有些不同，图23.8表示了所有可能的状态转移以及出现转移的概率，而这幅图表示了状态之间的依赖性

暂且假设在每个阶段所有人体组成块都在图像中出现，我们希望确定哪一种匹配选择表示一个人的肢体为以及该人的姿态，考察下面对数形式的条件概率：

$$\begin{aligned}&\log P(\{i_1, \cdots, i_9\} \mid X_t = \boldsymbol{x}_t, \cdots, X_{rll} = \boldsymbol{x}_{rll})\\&\quad = \log P(\boldsymbol{m}_{i_1} \mid \boldsymbol{x}_t) + \log P(\boldsymbol{m}_{i_2} \mid \boldsymbol{x}_{lua}) \cdots + \log P(\boldsymbol{m}_{i_9} \mid \boldsymbol{x}_{rll})\\&\qquad + (N_s - 9) P(\text{image segment from noise})\end{aligned}$$

如果把 $P(X_{lua} = \boldsymbol{x}_{lua} \mid X_t = \boldsymbol{x}_t)$ 记为 $P(\boldsymbol{x}_{lua} \mid \boldsymbol{x}_t)$，则对数形式的联合概率为

$$\begin{aligned}&\log P(\{i_1, \cdots, i_9\}, X_t = \boldsymbol{x}_t, \cdots, X_{rll} = \boldsymbol{x}_{rll})\\&\quad = \log P(\{i_1, \cdots, i_9\} \mid X_t = \boldsymbol{x}_t, \cdots, X_{rll} = \boldsymbol{x}_{rll})\\&\qquad + \log P(\boldsymbol{x}_t, \boldsymbol{x}_{lua}, \boldsymbol{x}_{lla}, \boldsymbol{x}_{rua}, \boldsymbol{x}_{rla}, \boldsymbol{x}_{lul}, \boldsymbol{x}_{lll}, \boldsymbol{x}_{rul}, \boldsymbol{x}_{rll})\end{aligned}$$

$$
\begin{aligned}
&= \log P(\{i_1, \cdots, i_9\} \mid X_t = \boldsymbol{x}_t, \cdots, X_{rll} = \boldsymbol{x}_{rll}) + \log P(\boldsymbol{x}_t) P(\boldsymbol{x}_{lua} \mid \boldsymbol{x}_t) \\
&\quad + \log P(\boldsymbol{x}_{lla} \mid \boldsymbol{x}_{lua}) + \log P(\boldsymbol{x}_{rua} \mid \boldsymbol{x}_t) \\
&\quad + \log P(\boldsymbol{x}_{rla} \mid \boldsymbol{x}_{rua}) + \log P(\boldsymbol{x}_{lul} \mid \boldsymbol{x}_t) \\
&\quad + \log P(\boldsymbol{x}_{lll} \mid \boldsymbol{x}_{lul}) + \log P(\boldsymbol{x}_{rul} \mid \boldsymbol{x}_t) \\
&\quad + \log P(\boldsymbol{x}_{rll} \mid \boldsymbol{x}_{rul}) \\
&= \log P(\boldsymbol{m}_{i_1} \mid \boldsymbol{x}_t) + \log P(\boldsymbol{m}_{i_2} \mid \boldsymbol{x}_{lua}) \cdots \\
&\quad + \log P(\boldsymbol{m}_{i_9} \mid \boldsymbol{x}_{rll}) + (N_s - 9) P(\text{image segment from noise}) \\
&\quad + \log P(\boldsymbol{x}_t) + \log P(\boldsymbol{x}_{lua} \mid \boldsymbol{x}_t) + \log P(\boldsymbol{x}_{lla} \mid \boldsymbol{x}_{lua}) \\
&\quad + \log P(\boldsymbol{x}_{rua} \mid \boldsymbol{x}_t) + \log P(\boldsymbol{x}_{rla} \mid \boldsymbol{x}_{rua}) \\
&\quad + \log P(\boldsymbol{x}_{lul} \mid \boldsymbol{x}_t) + \log P(\boldsymbol{x}_{lll} \mid \boldsymbol{x}_{lul}) \\
&\quad + \log P(\boldsymbol{x}_{rul} \mid \boldsymbol{x}_t) + \log P(\boldsymbol{x}_{rll} \mid \boldsymbol{x}_{rul})
\end{aligned}
$$

很容易建立动态规划算法的网格模型，对任意一个组成块，建立一组结点，每一个结点表示一个组成块的一个状态，每个结点有一个类似于 $\log P(\boldsymbol{m}_{i_1} \mid \boldsymbol{x}_t)$ 形式的权值，由于所有度量值和所有状态，因此所有结点的权值都可以计算出来。前面已经给出了人体组成块的树形结构图，对于树形结构图中的每一个结点，网格图中都有一列结点与之相对应，网格图中与树形图中父结点对应的列称为父列，与子结点对应的列称为子列。父列中的每一个结点与其子列中的每一个结点有一条有向边相连，这条边的权值类似于 $P(\boldsymbol{x}_{lua} \mid \boldsymbol{x}_t)$ 的形式（如图 23.16 所示）。我们希望找到权值和最大的一条路径。

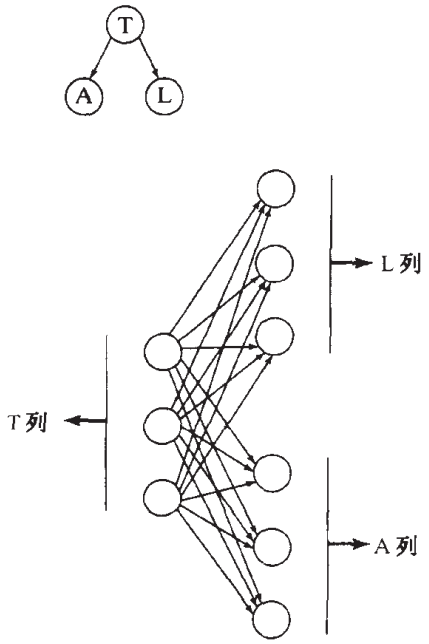


图 23.16 上图显示了一个基于树结构简单人体模型的网格模型，其中 L 表示腿，A 表示胳膊，T 表示躯干。每列的一个元素表示图像块、人体模型块和人体模型配置变量之间一个可能的对应。例如，一个结点可能表示在某个位置的第2号图像块是人的躯干。不难发现，“T”组中的每个结点包括两个子结点，这与前面例子中一个结点仅有一个子结点不同，但是这并不影响我们后面的分析过程。对“T”组的每个结点，可以确定其出现的概率，并从中选择一个出现概率最大的结点。这需要首先计算“A”组中出现概率最大的结点和“L”组中出现概率最大的结点，进而可以计算“T”组每个结点的出现概率。这提示可以对任意一个数结构模型应用动态规划算法

采用这种结构的网格模型意味着有些结点可能存在不止一个子结点，但这并不影响我们的动态规划算法。动态规划算法只要求每个结点知道从它开始向后的最优选择，使用这个模型同样可以得到这些信息。像前面介绍过的一样，从叶结点向根结点计算，当一个结点有多个子结点时，从所有分支中找出一个最大值，图示 23.16 描述了一个简单网格图的动态规划过程。

采用上述形式的模型可以很直接地从图像中检测人体。需要一个人体组成块的布局模型，例如前面给出的似然函数模型。Felzenszwalb 和 Huttenlocher(2000)假设人体组成块的颜色已知，通常是皮肤颜色或是蓝色。在这样的假设条件下，他们实现了一个令人满意的人体检测器(检测结果如图 23.17 所示)。

图 23.17　上面的左图显示了一个基于树结构的人体模型，模型中的每个部分用该部分颜色的期望值表示。这个模型试图在图像中找到模型中的11个部分在图像中的一种配置关系(其中肢体9个部分加上脸和头发共11个部分)，它满足：(a)颜色匹配；(b)相对位置关系类似人体。搜索过程可以利用上文中提到的动态规划实现。其他的三幅图片显示了利用上述方法得到的一些匹配结果

23.7　注释

隐马尔可夫模型是一个非常重要的模型，在写这本书的时候，有很多关于隐马尔可夫模型的研究工作正在进行着，当你阅读这本书的时候，可能有一些研究的前沿成果无法包含在内。我们认为，用搜索方法实现推理是其中的智慧结晶，并代表了当前隐马尔可夫模型方面的一些研究成果。隐马尔可夫模型最吸引人的地方在于利用它可以使我们的推导过程变得简便(换句话说，可以使匹配搜索过程变得简单)。

隐马尔可夫模型

可以利用不同方式的动态规划进行人体检测。Ioffe和Forsyth(见参考文献中他们 1999 年的论文)采用隐马尔可夫模型根据一个组成块是否是共享的，来帮助一个模块接受或拒绝组成块，其中条件独立假设使得模型不一定要求各个组成块是不重叠的。Song, Goncalves, di Bernardo和Perona(见参考文献中他们 1999 年的论文和 2000(a, b)年的两篇论文)把利用运动信息进行人体检测的问题描述成用概率模型实现匹配搜索问题，并采用动态规划算法解决

搜索的问题。

把隐马尔可夫模型应用于视觉问题有一些难点。首先,采用隐马尔可夫模型不一定能够获得好的结果。其次,如果希望推导过程变得简便,则需要对模型增加一些约束条件,但是条件独立假设有时限制给模型增加一些很自然的约束条件。例如,在进行人体检测时,就不能假设不同位置的图像块表示不同的组成块(因为这就要求在如图 23.17 所示的树上增加一条边,这可能导致其结果不在是一棵树)。此外,如果模型的包含复杂条件独立关系,则相应的推导过程也会变得复杂。目前,物体识别领域一个重要的研究课题就是如果找到一些满足以下条件的模型:(a) 符合客观实际,能够帮助实现有效的识别;(b) 允许采用简便的推导算法;(c) 能够通过简单组合这些模型构造一个新的模型。

上面的第三点要求是很重要的,但是目前还没有得到足够的重视。实际上,如果希望对多类物体进行识别,采用顺序识别的方法通常是比较困难的,但是,如果每一类物体提取的特征集合不相同的话,就只能采用顺序识别的方法(先看物体是不是属于第一个类别,再看是不是属于第二个类别,以此类推)。一个更理想的模型能够使得特征具有统一的形式,并能应用到不同的物体上去,而这些类型之间的区分来自于特征之间的关联以及特征的细节度量。但是这样的模型具体如何工作目前还没有成熟的研究结果。

第 24 章　基于空间关系的几何模板

在前面的章节中，模板主要是由具有一些特征外观的小块像素组成，而更加复杂的模板是非常有用的。例如，球体在实际中使用的透视投影摄像机中具有大致圆形的轮廓。这意味着如果在检测一个球体，那么将形成一个圆的边缘点收集到一起是一件很有用的事情。这是一种通过观察图像中各个成分之间的关系而得到模板的一种方法。这种模板提供了匹配的下列思路：

- 确定对所关注物体外观起约束作用的关系
- 在图像中检测满足这些关系的结构
- 基于满足条件的关系建立匹配关系

由于多数物体具有复杂的几何结构，把物体从整体角度来考虑的方法是效率不高的（因为这些关系会十分复杂）。因此不得不用物体是由结构简单的各个部件组成的观点，来取代用整个物体实现匹配的想法，并通过检测部件，将它们连接起来，形成物体的步骤实现模板匹配。

这种思考问题的方法式是很有吸引力的，因为它将物体的表达方式与分割联系起来，并且考虑了如何识别许多物体的问题，即假定物体都是由部件组成，但以不同方式进行装配，因此检测出部件，再推理它们的结构。这种思路吸引人的另一个理由是用一系列推理来实现识别。

24.1　物体与图像之间的简单关系

制约物体外轮廓的最简单情况是多面体，它的图像轮廓由一组线段链接组成。在大多数应用中，我们关注的多面体的表面块，具有较大的面积，因此图像中的线段由许多像素组成。这样一来在图像中不通过检测线段来寻找多面体是不可能的。换句话说，如果要检测出多面体，先形成线段链是好的想法。如果多面体的内轮廓能够利用的话，对多面体的轮廓可以有更进一步的约束，内部轮廓是指在图像中都看得见的表面块之间的边界（Sugihara，1986；Huffman，1977；Clowes，1971；Rothwell，Forsyth，Zisserman 与 Mundy，1993）。一般情况下，内部轮廓往往不能可靠检测出来，因而影响了这些约束的效率。

24.1.1　弯曲表面中的关系

回顾一下，表面在图像中的轮廓是由视锥与图像平面相交截面的轮廓。这种射线锥是由穿过透视摄像机焦点并与表面相切的射线组成的，称这种锥为视锥。对仿射摄像机来说这些射线是平行的。如果仿射摄像机又是正交投影的，那么视锥与图像平面的截面垂直于射线，这是最普通的情况。一般说来用视锥分析问题比用轮廓要容易些。

锥体与柱体　锥体表面是过一个点的射线沿一个平面曲线扫射形成的，这个点就是锥体的顶点，该平面曲线称为母线。值得注意的是，上述定义要比正圆锥更具一般性，而人们经常将正圆锥不恰当地简称“锥”；正圆锥具有旋转对称性（见图 24.1）。一个锥体是由它的母线经

比例放缩以及移位复制组成。如果选择一个坐标系框架使得母线能表示成$(x(t), y(t), 1)$，那么锥体就可表示成

$$(x(t)s, y(t)s, s)$$

而顶点表示成(0,0,0)。柱体是锥体的一种特殊情况，它的顶点处于无穷远处，这就意味着在一种合适的坐标系中柱体可写成

$$(x(t), y(t), s)$$

正圆柱的母线是一个园。

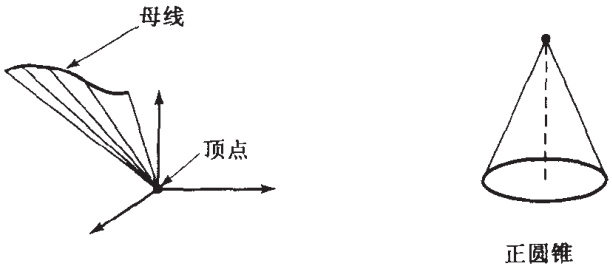


图 24.1 锥体表面是穿过其顶点沿母线扫射而成的，而正圆锥是一个特殊的锥体，它的母线是一个圆，而连接其顶点以及圆心的线与圆所在平面正交

对一个锥体的视锥是一组平面族，这些平面既穿过焦点，又穿过锥体的顶点(见图24.2)。因此对它的组合约束(grouping constraint)是：锥体的轮廓是由一组穿过其顶点的直线组成的(图 24.2)。

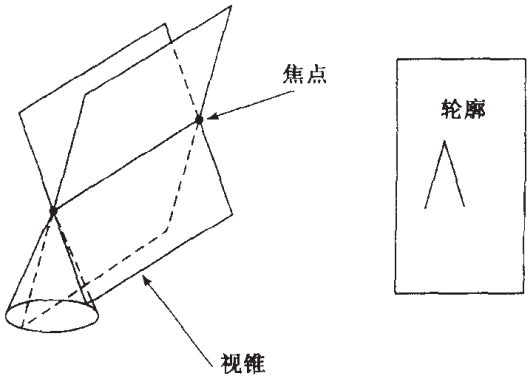


图 24.2 锥体的视锥由一组同时穿过锥体顶点与摄像机焦点，并与锥体相切的平面组成。将该视锥与一个平面相交可以得到锥体的轮廓，它是穿过一个点的一组直线

上面讲述的内容是很显然的(练习中要求一个证明)，但是却十分有用。因为一个柱体与一个锥体惟一的不同点是，柱体的顶点在无穷远处，因此该组合约束也适用于柱体。要提醒的是，柱体的图像顶点并不一定在无穷远处，除非图像是正交投影得到的。

管状表面 如果将一个固定半径的球体的球心沿一条曲线扫射，球体形成的包络是一种表面，称为管状表面，而该曲线一般称为生成曲线。管状表面一般看起来像弯折的管子(见图 24.3)，但也有呈现出奇异性的情况。用一个固定半径的球的球心沿一个圆扫过就是一个很容易说明的例子。如果球体的半径小于该圆的半径，就可得到一个圆环，但是如果球体半径

大于圆的半径，就会得到沿两个圆自相交的表面。

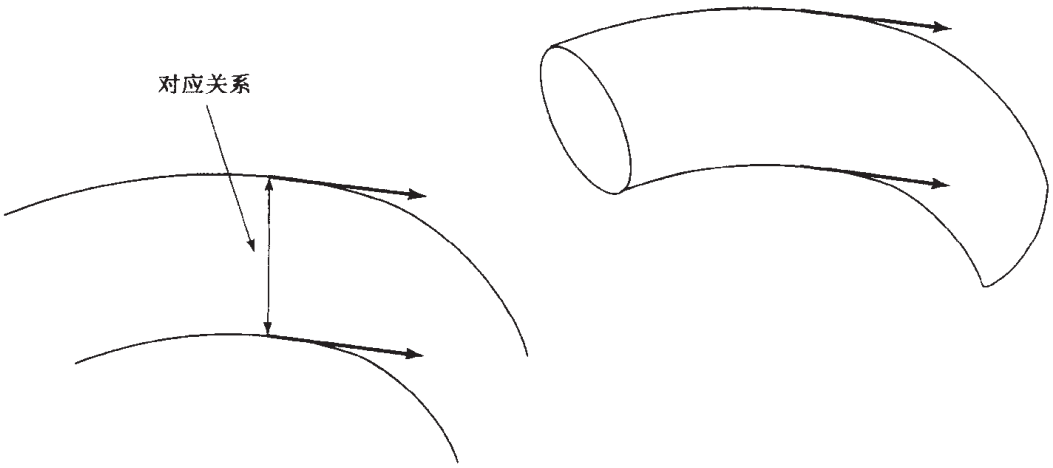


图 24.3　如果一对曲线之间存在平滑的对应关系，每一对对应点的切线相平行（左图），则这对曲线是平行的。在正交投影视图中管状曲线的轮廓由一组平行对称曲线组成（右图）

一个管状表面的局部是正圆柱（再次强调一下也可称为局部是一正圆锥，但其顶点在无穷远处）。假设使用正交投影，则柱体的轮廓是两条平行的线，这意味着从局部看管状表面的轮廓也是两根平行的线，所谓“局部”是指管状表面的轮廓的切线在对应点平行。这并不一定意味着曲线是可相互平移的类型（见图 24.4），却给我们提供了下述组合的线索：管状表面的轮廓很可能由两条曲线组成，其中一个与另一个平行（见图 24.3）。

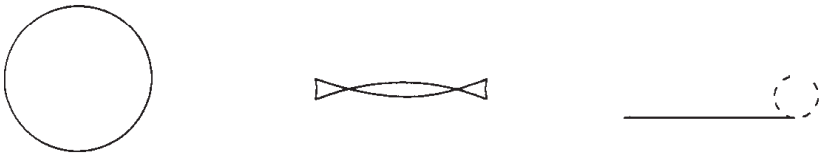


图 24.4　从不同视角看一个圆环的正交投影轮廓。需要指出的是，尽管在轮廓的两边建立起1对1的对应关系，使得一边的每个切线与另一边对应点的切线相平行，但是这并不意味着其中之一能平移至另一边（图中的情况）。在图中对应的一边用灰线表示，而另一边用黑线表示。在右图上对应的曲线自然地形成两种。在图上用虚线表示。在图上没有将不可见的轮廓段去掉，而在实际上这些会使构造对应关系的过程复杂

组合的过程可如下方式进行：先检测彼此平行的轮廓线段，然后试图将这些线段组合成轮廓。第一步，找到一个边缘点以及与其方向相同的所有其他点；然后沿着第一个边缘点的方向前进，继续按保持平行的要求搜索可能的匹配，直到保持平行的点都搜索完为止。注意要保持两边前进速度的同步。第二步，找端点相邻近并且大体指向同一方向的线段，并将它们联接在一起。这是考虑到这些线段之所以分成段，可能是由于边缘检测器的信号丢失，或由于可见性发生改变造成。

如果图像是透视投影，则可假设物体的深度范围与摄像机到物体的距离相比很小，因而透视投影可以近似成比例正交投影。由于以上做法与比例无关，因此上述做法不需要增加什么

步骤。然而对轮廓的拐点在透视投影下确实有一些附加的约束关系,但是对具有平面曲线的表面来说,它们仅能提供一些较弱的组合线索(Zisserman, Mundy, Forsyth, Liu, Pillow, Rothwell 与 Utcke, 1995*b*)。

旋转表面 旋转表面(Surface of Revolution, SOR)是通过一个圆扫过一条直的轴线得到的表面,该轴垂直于圆所在的平面,并穿过圆心,而圆在扫描过程中可以放大或缩小。因此在某种坐标系中旋转表面可写成

$$(f(s)\cos t,\ f(s)\sin t,\ g(s))$$

(这种形式允许表面可以往回扫,因而可以成为奇异的)。

旋转表面是圆形对称的:当表面沿它的轴旋转时,表面上的每个点仍处在该表面上,这表明旋转表面的视锥具有对称性。可以想像有一个穿过旋转表面轴以及摄像机焦点的平面,视锥对该平面具有翻转对称性。这并不意味着旋转表面的轮廓具有镜像对称性,因为旋转表面的轮廓是视锥与图像平面相交得到的。如果图像平面与对称平面并不成 90 度,那么旋转表面的轮廓没有镜像对称性(见图 24.5)。但是在图像中确有某种形式的对称性,这有时称为共轭对称性(conjugate symmetry)。对实际的摄像机图像平面与对称平面的夹角是接近 90 度,否则该物体会处在视场之外。摄像机的透镜系统也经常调整到这样一种状况,使得透视效应在远端产生的畸变减小。所有这些意味着对实际摄像机来说,透视投影带来的畸变量是小的,因此旋转表面的轮廓可以看成具有镜像对称性。

上述分析表明旋转表面轮廓的两边的对应点可以通过以下组合约束来鉴别:旋转表面的轮廓具有近似的镜像对称性。

它又可以表述成另一种非常有用的局部形式:图像曲线上两点的切线与该两点连线之间夹角相同,则它们可能处于对称性的两边(见图 24.6),也因此很可能处在一个旋转表面上。

我们称这种图像结构为局部对称性,而连接这两点的线为对称线。因此检测旋转表面的轮廓可以通过搜寻局部对称性来实现,这种对称性体现在,它的中点处在与它们的对称线几乎垂直的直线上。这种策略的主要困难在于许多图像包含很多很多的对称性,因而可能有许多组对称性满足这种约束条件。

直齐次广义柱 直齐次广义柱(straight homogeneous generalised cylinder, SHGC)是旋转表面的一种扩展,它是一个平面生成曲线沿一条与该平面垂直的直线上扫射得到的,该平面曲线在扫的过程中还可膨胀与收缩(见图 24.7)。在选择一个合适的坐标系中,母曲线可写成 $(x(t), y(t), 0)$,而该表面则写成

$$(x(t)f(s),\ y(t)f(s),\ g(s))$$

这种形式也意味着表面可以朝相反方向回扫。SHGC 并没有旋转表面所有的对称性,但是具有旋转表面的其他属性。旋转表面从局部看是一个正圆形的锥体,SHGC 在局部看是一个锥体。要说明这一点,可以设 s 的值为 s_0,并考虑从 s_0 到 $s_0+\epsilon$ 之间的表面带。如果这条带足够窄,那么可以近似为 $us+v$,而 $g(s)$ 可以近似为 $cs+d$, u, v, c, d 都是一些常数(这也就是导数的概念)。d 可以通过移动而排除,而 c 则通过重新参数化消除,这就是说,在 $s=s_0$ 处的所有切线向量组成一个锥体。它的顶点在我们的坐标系统中很容易确定,当 $us+v=0$,那么不管 t 是何值,所有的切线向量都通过 $(0,0,g(s))$。这表明每个切线锥的顶点在该坐标系统的 z 轴上。

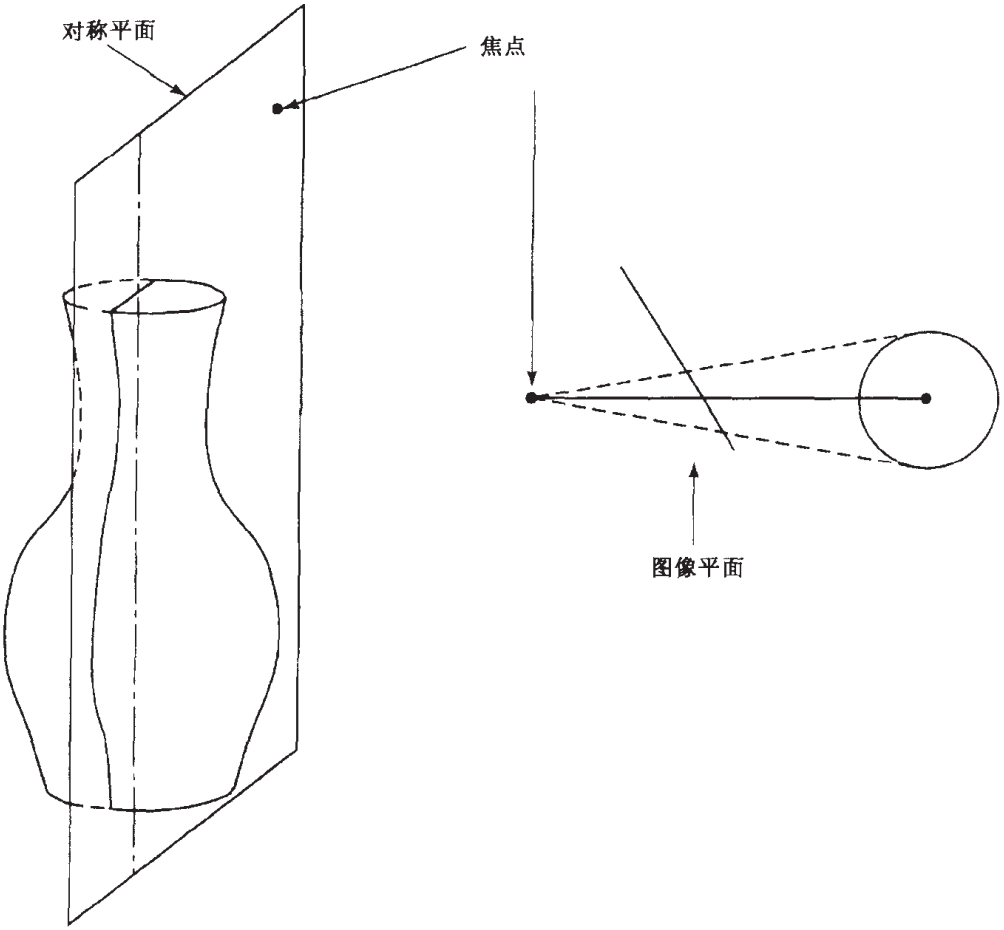


图 24.5　一个旋转表面加上一个焦点得到一个对称平面，该平面通过焦点与旋转表面的轴（左图）。在这个平面中轮廓母线具有镜像对称性。但这并不意味着图像的轮廓具有严格的镜像对称性，因为图像轮廓是视锥与一个与对称平面非直角相交的平面相交产生的。视锥在图中没有画出以免引起混淆。在右图上表示的是从上往下看的视图，在图中视锥用虚线表示，视锥被图像平面截成了旋转表面的轮廓。在该图中图像平面与对称平面的夹角略偏离90度，也就是说，在图像平面上物体有点偏离摄像机中心，因而旋转表面的轮廓并不具备严格的镜像对称，而具有共轭对称。一般情况下，摄像机具有相对小的视场，意味着图上所示的极端情况在实际情况中很少出现。对绝大多数实际摄像机来说，共轭对称与镜像对称是很难区分的

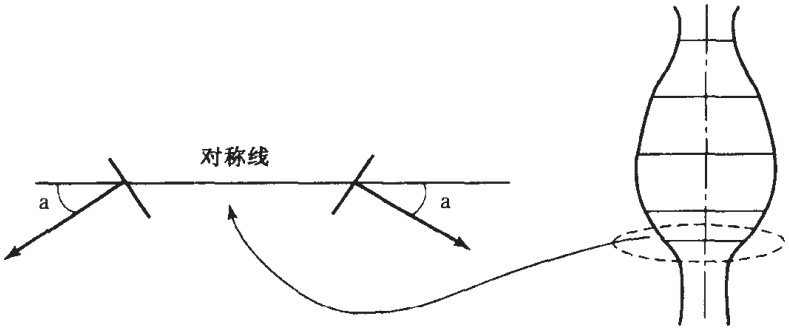


图 24.6　局部对称性是指一对轮廓点处的切线与连接该两点的连线（对称线）之间的夹角大致相同。这种对称性经常在旋转表面的轮廓上看到

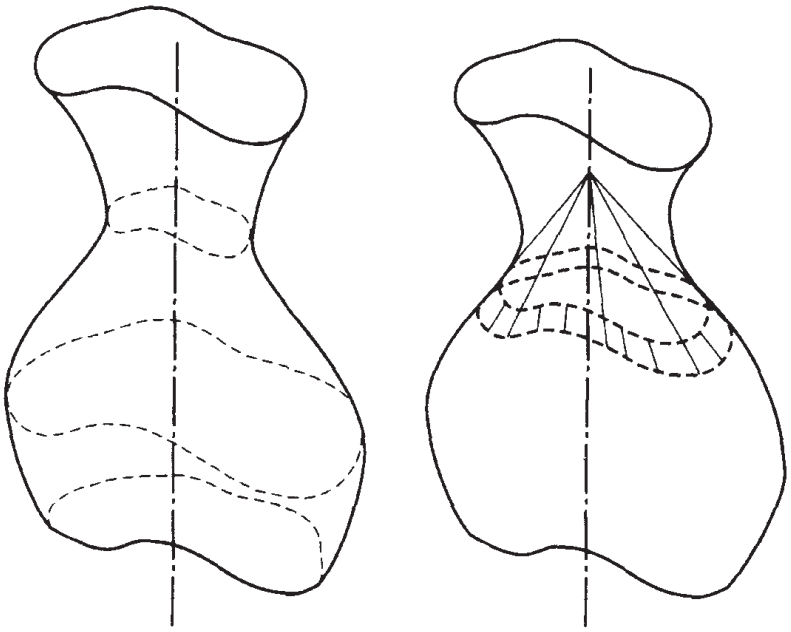


图 24.7 一个直齐次广义柱(SHGC)是通过将一平面母线,沿与平面垂直的一条线扫射得到的,在扫射过程中母线膨胀与收缩(左图),所得到的表面从局部看是锥体,一个沿着任何母线与表面相切的锥体。在右图上,表面上两条母线之间切出一条窄带并画出它们的切线。这些切线交于轴上的一个顶点,这表示SHGC上每一个这样的带子是锥体的一段(也就是说从局部看它是锥体)

尽管我们用坐标系来得到上述结论,但这只是为了便于说明。因为这个结果是切线的固有属性(incidence properties),它不受坐标改变的影响,它对任何 SHGC 在任何坐标系中适用。所以对任何一个 SHGC,在任何一个坐标系内,都存在一个明确定义的轴,所有切线锥体顶点都落在该轴线上。

由于表面在局部上是一个锥体,它的轮廓也是锥体的(局部)轮廓。因此,如果我们在对应于某个 s 值的轮廓上的点获取其切线,它们会相交于一个点上。随着 s 值的变化,切线锥的投影顶点沿一条直线运动,因此得到了以下的组合约束条件:在轮廓的分为之间存在某种对应,它体现在所产生的相应边的切线会交汇到同一条直线上。

这是一条分割用的准则,因为并不是所有的曲线都能满足上述条件的。但是使用这种形式需要某种技巧。

平面直常截面广义柱 另一种有清晰的组合线索的旋转表面是平面直常截面广义柱(planar right constant cross-section generalized cylinder,缩写成 PRCGC 或称管子)。这种表面是通过将一固定的平面曲线沿垂直于某个轴线方向扫射产生。如果这个平面曲线是一个圆,那么该表面就是管状表面,但平面直常截面广义柱可以推广到形状更加复杂的截面曲线。

与管状表面一样,PRCGC 的局部是一柱体,但并不一定是正圆柱体。这就是说从局部看视锥是一个柱体的视锥,因此在正交投影条件下视锥是一组互相平行,且又与轴平行的薄片。鉴于以上分析,在正交投影条件下,可得到以下组合约束:PRCGC 的轮廓是由一组平行曲线组成(也就是说在曲线上的点之间存在平滑的对应,以至在对应点的切线是相互平行的,见图 24.8)。

具有这种性质的一组曲线称为平行对称曲线，我们已经讨论过对一组平行对称曲线组合的问题。

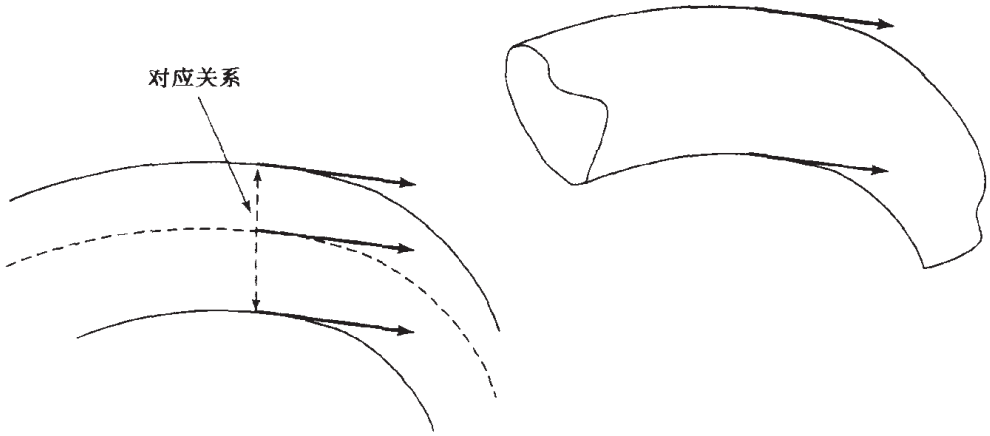


图 24.8　一对曲线称为是平行对称的，其条件是它们之间有平滑的对应，使得相对应的点的切线是平行的（左图）。在正交投影条件下PRCGC的轮廓是一组平行对称曲线（右图）。图中画出了表面的一端，以突出该图上画的是轮廓

24.1.2　基于类型的组合

以上所述的约束条件是对各种物体类型适用的，这就意味着在进行特征组合时，并不需要知道要组合的是什么物体。因为不同类型的物体生成不尽相同的图像约束，必须使用不同的策略将来自于不同类型物体的轮廓可组合起来，因此也称这种组合为基于类别的组合。其实这并不是一个问题，事实上这是一个很有吸引力的方法的例子。因为在组合过程的早期，从图像中得到的信息能使我们及时地把注意力集中到某些类型的图像曲线与物体类别上。例如，如果图像中根本没有直线，也就没有必要去组合多面体了；如果镜像对称不存在，那么也没有必要去检测与组合旋转表面。同样组合管状表面的过程仅需寻找具有平行切线的边缘点对，因而可以忽略相当多的其他边缘信息；当将点对扩展到曲线段时，还可以删掉许多虚假的信息。

于是搜索过程分成两个阶段进行。先搜索一小组物体类型以确定组合策略，每种类型的物体产生不同的轮廓点集，然后对这些点集采用相应组合策略来鉴别图像中出现的物体。通过两阶段搜索的方法来减少搜索的总量，例如图像中没有直线段，就不再需要搜索多面体模型。尽管这里讨论的例子规模不大，但所使用的原理是很有效的——在识别过程的初期检测少量信息，从而大大减少了需要检测的模型的数量以及对全幅图像的处理量。图 24.9 与图 24.10表明了这种类型组合器的性能，它是由 Zisserman，Mundy，Forsyth，Liu，Pillow，Rothwell 与 Utcke（1995*b*）建造的。

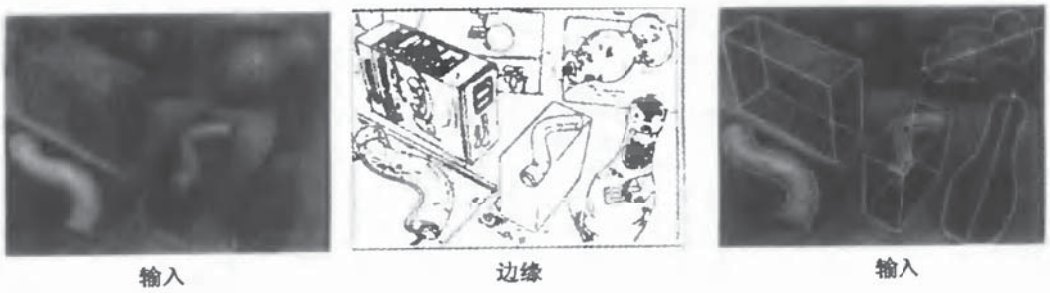


图 24.9 一个基于类别组合系统的例子：左图是一幅输入图像，它的边缘图表示在中图上。在右图上是表示不同类型的物体轮廓的曲线组。从图像中检测出的边缘往往有断点、遮挡点等。但它们可在对具体类型物体的处理过程中加以修补（例如，多面体组合器修复线段上的问题，旋转表面组合器使用镜像对称来修补镜像对称曲线对中丢失的信息等）。因为每类使用略为不同的组合器，轮廓的分类在过程的结局中确定，但是分类只取决于物体的类型，而不是具体的哪一个物体，更加详细的例子显示在图24.10

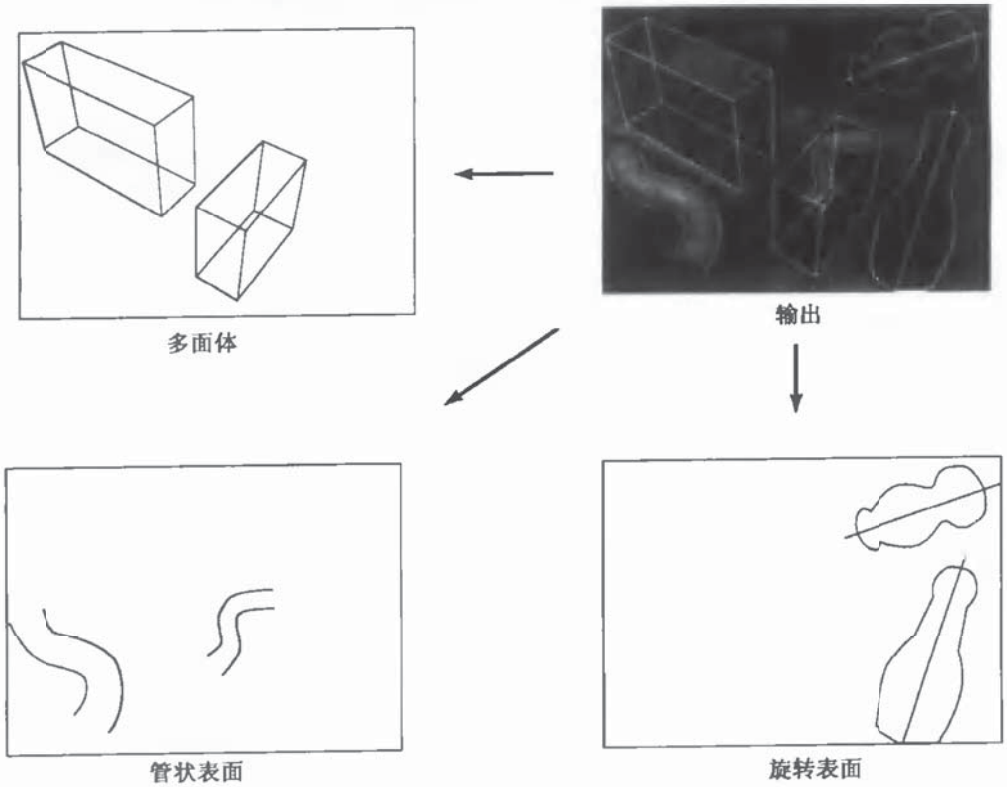


图 24.10 图 24.9 中图像的边缘图有断点、遮挡点等，这些边缘送到不同的组合器。多面体组合器通过检测直线，并将其组成链的方式构造组合以及修复线段；而管状表面组合器，通过检测局部平行的边缘点，并将其组合成平行曲线的方式进行组合与修复边缘。旋转表面组合器，通过检测具有镜像对称性的成对曲线来构造组合以及修复边缘。修复边缘之所以 可能是因为已找到的边缘点可以根据约束在组合假设的帮助下起修复断点的职责

24.2 基元、模板与几何推理

上一节的例子表明了推理路线的吸引力，但限制在相当简单形状的物体上。实际中有许

多更有兴趣的形状的例子，但表面与其图像之间的关系没有像前面所述那样紧密。对这些情况的推理引出了一些很知名的物体识别系统。

24.2.1　用广义柱体作为体基元

体基元因提供了一种几何模板而令人感兴趣：一些图像成分由于体现了符合某种约束条件的形式被收集在一起，作为一种证据表示某种基元(primitive)的存在，随后从这些证据中寻找它们之间的关系。如果基元能够以忽略不重要的细节的方式表示是比较理想的[例如我们手指尖上的由脉络与筋腱产生的拐点和脊(ridge)，并不影响将手指表示成柱形结合体]，但是我们早先讨论的基元还是相当有限的。

Binford 在 1971 年一篇著名文章中提出的广义柱(generalized cylinder，GC)是当今最普及形式的体基元。最初用的术语为“广义移动不变性”。概而言之，这种术语是指 GC 是一种刚体，它由一个截面以一维参数的方式扫掠而成，截面在扫掠过程中平滑地改变自己的形状。这种定义概括了这样一种思路，即许多人们感兴趣的物体在某种适当的细节层次上都可用扫描体描述，管状物、柱体等是很容易想到的例子，就连汽车也可想像成用一个矩形沿其驱动轮系方向扫射得到，矩形的尺寸可变大或变小(对轮子需要另做安排)。为了使广义柱的定义更容易被接受，一种方法是将管状表面的定义扩展成允许扫射球体的半径，在扫描过程中增加或减小(见图 24.11 左)。当然并不是所有表面都可表示成广义管状表面的。对广义管状表面，如果我们构造它最大内接球的球心轨迹，可以期望得到一条曲线，但是对多数表面来说，这种变换得到的都是一个表面(例如图 24.11)。

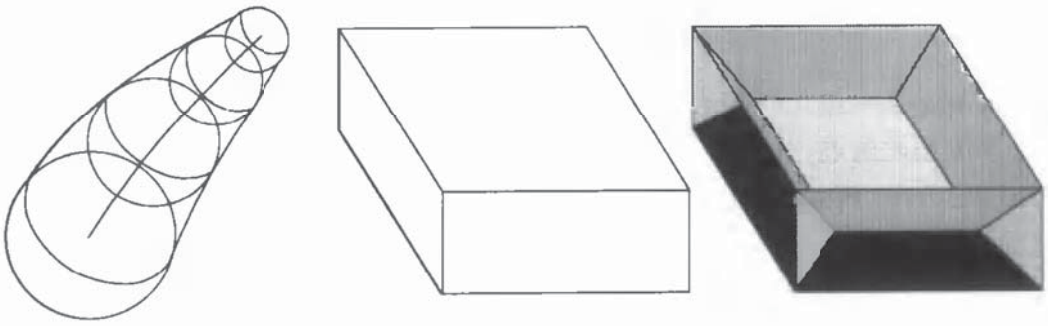


图 24.11　广义管状表面实现的可能性。左图是一个广义管状表面，它是通过沿一个空间曲线扫描一个球，并在过程中放大或缩小球的半径得到的。母球沿圆与表面接触，但并不是任何表面都可表示成广义管状表面的。这可通过构造表面最大内接球的球心轨迹来理解这一点。表面 的最大内接圆是指至少有两点与表面相切的圆。一般称这种球心的轨迹为骨架(skeleton)。对广义管状表面，它的骨架包含一条曲线，沿着该曲线扫掠一个球可以得到整个表面。显然并不是所有表面都具备这样一种性质的。在右图显示的是一个平行六面体以及它的骨架，六面体是由多边形表面组成的。略微思考一下就会明白从这些表面是得不到任何曲线具备扫出管状表面的性质，这也就是说平行六面体并不是广义管状表面

使上述定义显得更加牢靠的另一种方法是用一个平面曲线沿一个空间曲线扫描，并允许在扫描过程中改变平面曲线的尺寸以及它的倾侧角。这就是通常所说的广义柱的定义(图 24.12，左)。这显然是一大类表面，因为可以自由地选择截面比例放缩的规则、母曲线、倾侧角度等。根据这种定义，广义柱几乎可以适用于任何表面，并且对截面还可有多种选择，至

少对某些广义柱来说是这样的，一个球就是一个很好的例子。依据所选择的母线形状以及所允许的变形类型，可以得到许多不同的派生定义。如使用盘状或任意的平面区域作母线，变形可以是简单的二维放缩或其他的平滑过渡的变形。但是不论是哪一种情况，怎样从表面计算截面、放缩规则及母曲线并不很容易，从图像数据进行计算则更难了。

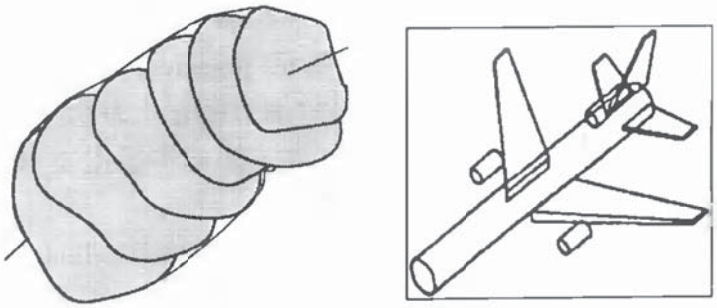


图 24.12 广义柱：左图：广义柱的初始定义：平面截面不断地发生变形；右图：一个飞机(L-1011)的ACRONYM模型。这是一个一般的宽体飞机模型的例子，它由柱体与锥体基元装配而成，这些基元的截面是圆形的与多边形的

24.2.2 带状基元

广义柱与它的图像之间的关系并不那么严密(但这并没有影响，因为人们认为广义柱定义得很精确并没有意义)，因此我们不需要在上一节中用到的那种精确的几何推理。然而如果接受把广义柱看成由变形截面扫出的表面这种观点，我们就可设想广义柱的图像是由一变形线段沿一曲线扫出的区域，这意味着要找一种能使用扫射方式得到的区域。这种区域一般称为带状基元(ribbon)。

一个带状基元是一组几何图形沿某个轨迹扫过的包络，这些几何图形是带状基元的母线，该轨迹线是带状基元的筋，几何图形在扫掠的过程中可缩小或膨胀(见图 24.13)。有两种情况一般是加以区分的，Brooks 带状基元使用的母线是线段，而 Blum 带状基元使用的母线图形是一个圆(见图 24.14)。

如果说这种扫射形式与图像中表面的扫射形式相符的话，那么可以考虑以下三个问题：

- **解释**：给定一个平面区域，能不能用带状基元给出该平面区域一个有用的表达式？
- **分割**：能否从图像中检测出带状基元，这样做是否对检测相应物体有帮助？
- **匹配**：能否用基于带状基元的表达式检测出物体？

上述三个问题是互相关联的，如果任何一组图像特征都可得到一个带状基元的话，那么带状基元对分割出有关物体起的作用就不大。

解释 大多研究工作集中在解释问题上，即对一给定平面区域用带状基元的术语确定一种有用的表达式。集中在解释问题上可能是认为它比较容易解决，并且能为物体表达式提供更进一步的了解。对 Blum 带状基元可以得到相当简单的结果。特别有意义的是，对 Blum 带状基元可以定义 Blum 变换，俗称草原烈火(grassfire)变换。在一个二维区域内，构筑一组内接圆，这组圆的每一个必须至少有两个点与该区域的边界相切。这些圆的圆心轨迹就是该区域形状的 Blum 变换，如果我们把角点看做有多重切线的点，Blum 变换也可扩展到角点。

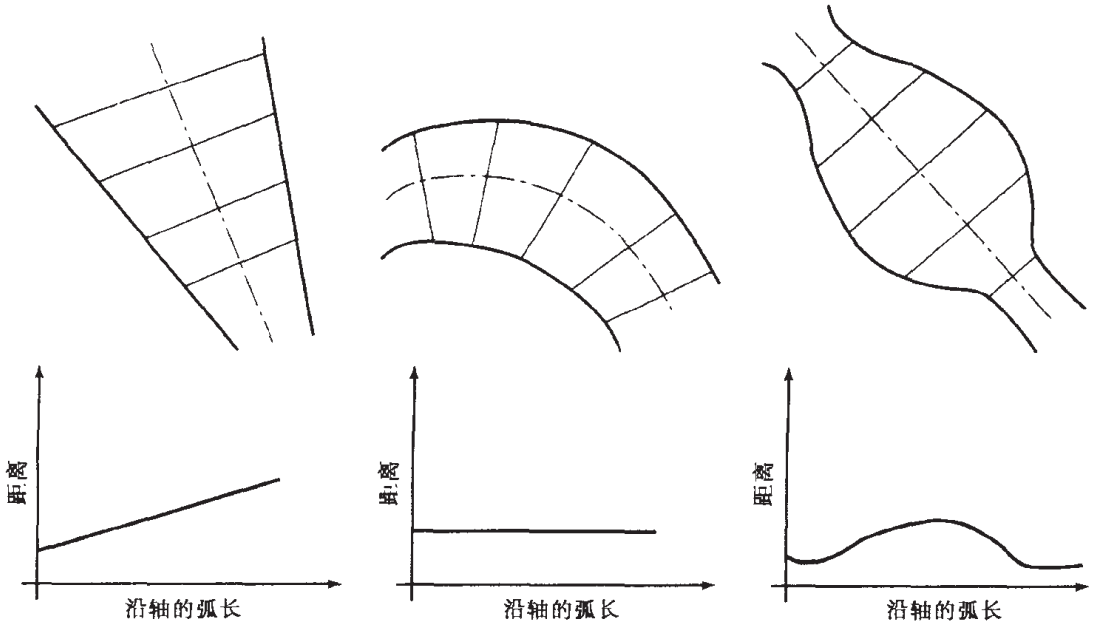


图 24.13　我们感兴趣的物体在图像中看上去经常像带状区域；对带状基元的一个合理模型可以用一根曲线为轴，沿曲线上每点的法线向两侧延伸等长距离的方式得到。每点延伸的距离可以沿着曲线变化。对图中的每个例子，用曲线与若干截线段来表示其结构。在图中每个例子下面，画出截线长度随曲线变化的曲线

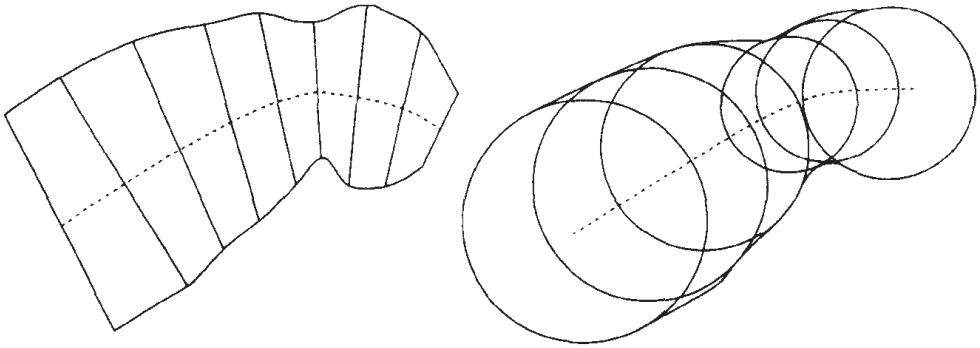


图 24.14　Brooks 带状基元(左)与 Blum 带状基元(右)。对每个基元来说，其中心线是带状基元的脊，用它的母线的中心轨迹定义。对Brooks带状基元，脊通常是母线段中点的轨迹

因此对一个平面曲线 C，可用 Blum 变换的术语写成

$$C \in \text{Blum Transform}(\text{Blum Ribbon}(C, \text{shrinking rule}))$$

换句话说，如果让一个以任何一种放缩规则放缩的圆沿 C 移动，就会得到一个区域，它的 Blum 变换包含 C。如果把某个放缩规则写成 R，则就可得到

$$\text{Blum Ribbon}(\text{Blum Transform}(\text{Blum Ribbon}(C, \text{shrinking rule})), R)$$

与下式是相同的

$$\text{Blum Ribbon}(C, \text{shrinking rule})$$

所有这些是指 Blum 变换与一个放缩规则加在一起给出了一个区域形状的描述，可以用一种简单的算法实现 Blum 变换。这种算法用叠代方式腐蚀一个数字图像区域，并最终获得该区域的

骨架(或称中轴)曲线,它是由区域的双切圆的中心组成的。这种性质使我们又从另一角度来说明 Blum 变换:如果把这个区域看做是一块草地,并在其边缘点上火,那么 Blum 变换就是由火焰向前推进时其前端相遇的点组成。这就是"草原烈火变换"这个名词的由来。骨架的分叉(也就是几支光滑分支相遇的多重点)定义了形状分解成各个部分,这种分解在直觉上可能感到很恰当,但也可能认为不合适(见图 24.15 与图 24.20)。Blum 带状基元在物体识别中显现的优点不幸地被它对边界噪声的敏感性而抵消了。例如一个矩形边上的任意小凹坑会引起骨架结构的明显改变(图 24.15,右)同样,骨架的分叉在边界的扰动下也相当不稳定。那么相似的 Brooks 变换是否能够定义得更加牢靠些?尽管在练习中描述了一种设想,但这个答案并不清楚。在这一节的余下部分我们集中讨论 Blum 变换,Brooks 带状基元与它在三维空间中与广义柱的对应关系,将在下一节继续讨论。

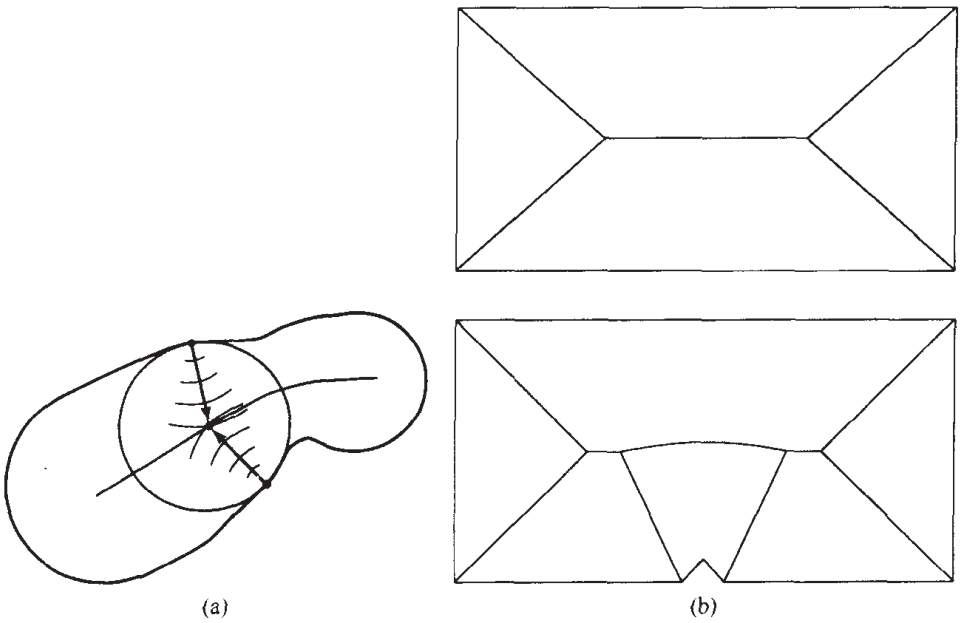


图 24.15 Blum 变换:(a) 一个平面区域的骨架是内嵌在该区域内的双切圆的圆心轨迹;(b) 一个矩形以及它带噪声形式的骨架。对平面区域内的多数点,它们到区域边界的最小距离是边界上一个点决定的,但是在骨架上的点却不同,该最小距离是由两个点同时决定的。于是骨架上的点可以想像成这样一些点的轨迹,在这些点上从区域轮廓上掀起的波浪前沿相遇在一起。在离散的图像中用这种概念得到一个求骨架的算法,该算法通过一层一层地去掉边界点的叠代方法腐蚀该区域,直到只有骨架点被保留了下来。在练习中有相关细节

分割 分割是一个有趣但比较困难的问题。当带状区域的形状有严格的限制时,就像上面的例子所示,则在图像中就可能得到较少的带状基元,也因此比较容易检测到。当对形状的约束较少时,就需要用一些启发式手段使得注意力放在有关的带状基元上。下面叙述一个由 Mohan 与 Nevatia(1989 与 1992)建造的系统。

第一步是检测边界段之间的结合关系;以便将被边缘检测遗漏的点,平滑地连接起来,或根据边界段端点相近的原则,将两条曲线形成的角找到。接着检测曲线之间的部分段是否近似平行,就像前一节组合管状表面的方法一样。这样一来就可得到一堆假设的母曲线,一般说来,这种所假设的母曲线数量太大,以致对下一步分析帮助不大(见图 24.16)。

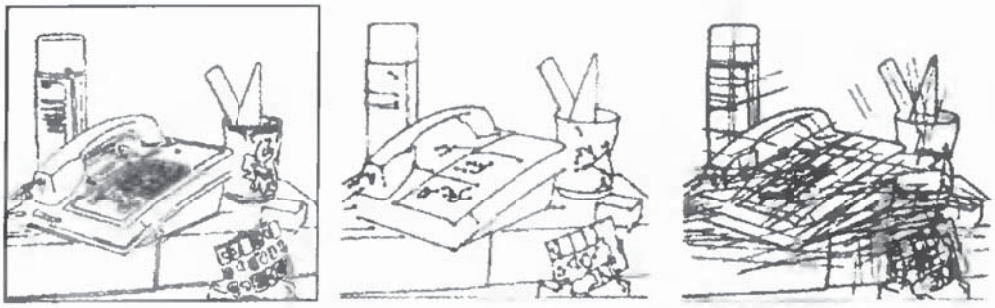


图 24.16　左图：对一些常用物体图像获得的边缘；这些边缘常有断点，物体的轮廓也没有形成相连接的边缘曲线，况且还有一些边缘并不是物体的轮廓。中图：边缘曲线已用两个准则连结起来。一个准则是间隙必须很小，并且连接后的边缘必须平滑（曲线组合），另一准则是间隙必须小（毗邻连接）。右图：在所有平行曲线对中检测出中点并构成轴线。这样的轴线数量会很大，其中大部分是没用的，用一致性准则可以剔除相当多的轴线

有一些解释彼此是不一致的。例如图 24.17 中所示两根产生曲线不能共享同一根轮廓线，否则会隐含一种非一般性的视角，一种巧合的视角。理想的情况是每个带状基元有一个管端盖帽，因此当出现矛盾时这种约束条件能帮助确定哪一个产生曲线能保留下来。此外，由于一般希望忽略物体上的一些标记，因此希望删去完全包含在别的带状物内部的带状基元（但是这里含有遮挡推理的效果）。这样一来就可以确定一个求优化解的问题：按照带状基元中连接的平滑性等评分，并按违反一致性的程度打折扣。依据这种评分，找到使评分值达到最大的一组带状基元（该方法的作者使用松弛法技术，实际上还可用其他一些技术）。正如图 24.17 所示，这种方法能够获得成功。看来三维物体与二维图像之间的关系用不怎么严格的方式表示对解决问题并无大碍。这种求最优解的框架的好处在于从原理上看，能够抑制导致获得不合适的物体的假设。当然，这会使求优化解的问题变得很困难。

图 24.17　左图：一个一致性准则的说明；两根轴如图中 A 与 B 不能共享曲线中标为 ab的同一片断（否则的话这个视点会是一种巧合）。必须从众多轴线中鉴别出最好的一组轴线，利用连续性等指标进行评价并考虑违反一致性准则所起的负面作用。这是一个求最佳的问题，可以用各种求优化解算法（Mohan 与 Nevatia 使用松弛法）。中图：展示了一组获得极值的轴线。可以将它们与相应的边界联系起来，得到一组表面假设，经过与遮挡关系推理有关的进一步处理后得到的一组假设表面展示在右图上

匹配　对称性看来确实能得到一些实用的表达式，例如在分辨率有限的条件下用基于对称性的表达式确定图像中有没有未被遮挡的人。这种系统的细节在 23.3.2 节中讨论过，在这

里摘要说明这种描述方法是为了运用基元的观点加以观察。

这个推理过程可以表示为:人体可粗略地认为是由柱状片断组成的,这些柱体在图像中形成长的直带状基元。带状基元的某些结构预示了有人出现在景物中,而有些结构则不是。所以要检测到人首先要检测带状区域,然后检测出符合人体的一些带状基元组合。相同的策略也可用到检测马的场合,如图 24.18 所示。

这种推理线路是很吸引人的:它首先检查相对少的图像组成,以判断它们是否是带状基元的一部分;然后将符合条件的组合成带状基元组,并在一个更大的范围观察是否像一个人体。这就是说,使用一些物体的部分模型将所需的证据组合到一起,以判断是否有此类物体存在。说得更确切些,在每一步只用少量的模型数据帮助我们筛选掉相当多的可能性。要强调的是,尽管这里没有用具体的概率模型,但看来的确进行了类似推理的工作。

24.2.3 带状基元能描述什么

假设有了物体的剪影(这样就躲过了分割问题),那么能用带状基元做什么呢? 由 Zhu 与 Yuille(1996)开发的 FORMS 2D 识别系统使用它们来表示人体与动物。物体通过将它们的剪影分解成带状基元来表达。困难在于如何避免歧义性的分解;Zhu 与 Yuille 通过检测与剪影多处相切的圆来避免这种问题。他们要求这种圆有尽可能大的半径,但又几乎没覆盖任何背景。通过求极值过程找到的这些圆作为进一步增长出表达式的种子点。这个算法使用递归算法,先从种子点开始跟踪骨架的关联分支,然后在骨架到边界的距离函数中寻找骨架的分叉。这些分叉表示骨架开始分裂,就像两个手指在与手掌连接处开始分裂;接着对每一个分量又开始寻找骨架。一旦找到了骨架,每个分支成为物体的一个部件,而与其末端关联的圆包含在部件的描述中(见图 24.19)。

图 24.18 一个能从接近平行边的直线段,得到某种表达式的例子。左图:场景中有两匹马的彩色图像。中左图:具有毛皮颜色与纹理属性的像素保留下来,其余的都屏蔽掉。中右图:这些像素的边缘;右图:使用上述机制得到具有接近平行边的所有直线段。这些平行线段叠加在获得这些线段的边上。要注意的是,这些成分并不精确地对应于驱体的片断,看上去与期望的结果有距离。尽管这种方式表达图像信息相当不充分,但是这些表达式能够用来找到马匹(25.3.3节)

正如图 24.19 所示,这种方法对脊椎动物,如哺乳动物与鱼等的侧视图给出了直观上相当满意的结果。一旦部件找到后,他们就转向寻找两种类型的基元:杆状与圆形扇形。杆状基元表示两个分叉点之间的脊骨段或物体具有细长形状的末端,圆形扇区则表示其余的部件(见图 24.19 上鱼的分解)。每个基元由一个 5 维向量表示,这 5 维是杆状基元的长度(或圆形扇区的角度),加上 4 个变形参数分别表示带状基元的宽度或半径函数,这些参数是对模型数据库用主分量分析的方法进行计算得到的。

匹配是利用图匹配方式进行的,使用度量部件之间的相似度的散列表与投票方式有效地检索出可能性最大的模型部件与物体种类。一个模型部件 m 与一个观察到的基元 o 之间相

似度定义为 $s(m,o)=\exp(-|\boldsymbol{m}-\boldsymbol{o}|^2)$，其中 $\boldsymbol{m}$ 与 $\boldsymbol{o}$ 是与两个基元有关的 5 维特征向量。

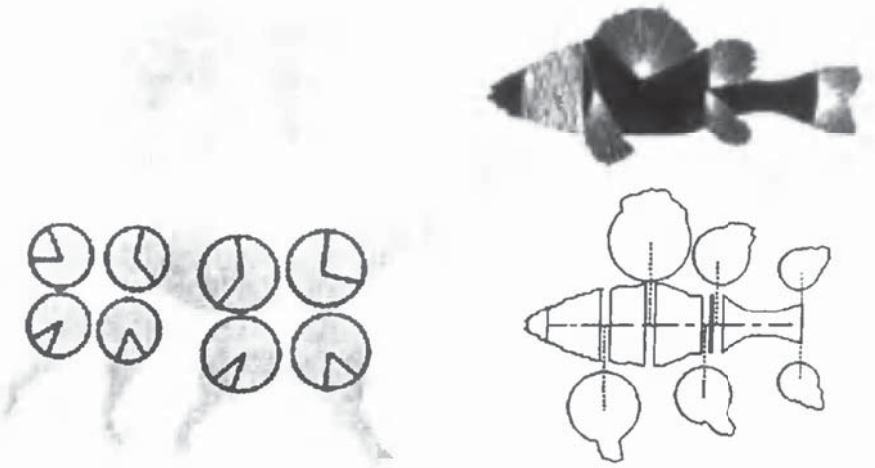


图 24.19　FORMS 工作例子：左上图上狗的剪影的分解图显示在左下图。分解过程是通过找，与轮廓多重相切的最大圆得到的。这些圆的圆心被认为是骨架分叉之处，因而可以用来将剪影划分成部件。在右上图中的鱼的剪影用同样的方法分解成部件，并展示在右下图。要注意的是一个部件是可以放在另一部件之上的。说得更具体一些，鱼身部件的顶部边界（看不见）是从看得见的凸起部分使用对称性假设推断的

每个物体的图像用一个图表示，其结点是骨架的结合点极端点，而图的边是它们之间的基元部件。每个物体种类用若干个骨架图描述，它们对应于所能观察到的图像。利用基于相似度的散列与投票选择出与可能性最大的模型相关的骨架图，然后使用一个图匹配过程将候选模型与所观察的数据进行匹配，将匹配代价超过某个阈值的匹配候选删除掉。通过自适应匹配使得匹配过程鲁棒性强，自适应匹配允许模型的骨架图在匹配过程中能在骨架运算子的作用下发生改变，以便针对自底向上进行骨架抽取的步骤中可能发生的误差做相应调整（例如由于遮挡而丢失某些分支；由于背景混杂而产生额外的分支；由于边界噪声引起结合点的分裂，或将两个结合点合并成一个）。这些运算子的每次应用要计算代价，而我们寻找代价最小且低于某阈值的匹配。

图 24.21 表示了一个识别的例子，在 FORMS 的实验中使用了一个含 17 种，35 个例子的数据库。表明这种方法取得不错的效果，这表明如果能得到物体的预先分割好的视图，我们也许能够用这类带状基元表达相当数量的物体。

24.2.4　利用柱体长度已知的条件从二维图像获取三维物体

如果有一幅用任意摄像机照的人体图像，能否从中得知该人在三维景物中的状况？答案是肯定的，但带一些定性性质。这是因为人体的先验结构对三维结构与其二维图像之间的关系有很强的约束关系，Talor(2000)指出了这一点。

可以将人体用一组柱体来建模，并且可以假设这些柱体的长度是已知的，至少它们的相对关系已知。这是这种方法的一个关键。可以从各种分类学出版物获取这些柱体的相关长度，例如从分类学研究项目(1978)中获取数据。

假定我们使用的摄像机模型是比例正交投影，而比例因子 s 已知。假设对一个已知长度为 l 的柱体的投影我们能够检索到柱体的两端，因而可以将端点的影响去掉。在这些近似条

件下所投影的柱体的长度是

$$sl\cos\theta$$

图 24.20 一旦 FORMS 确定了骨架的分叉点，就像图 24.19 所示，就在相应分支补充骨架。图上显示了一些剪影，所确定的骨架看来挺合理，因为其反映了我们的直觉，动物有头、躯体、腿与尾巴，手有手指等

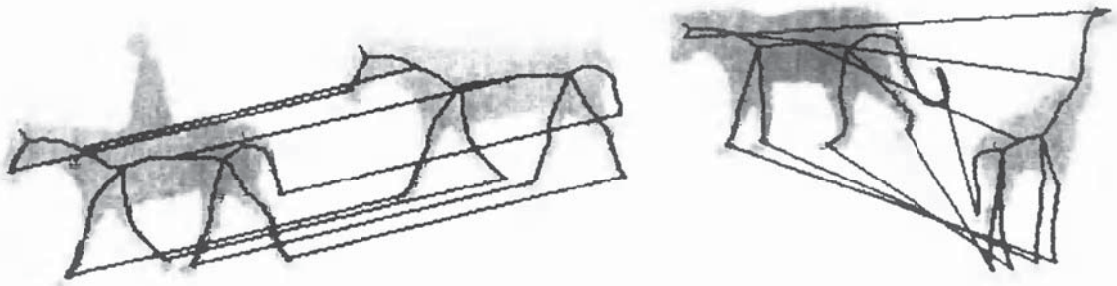


图 24.21 FORMS 识别例子

此处 θ 是柱体轴线与图像平面之间的夹角。这样一来就可以推算柱体在三维空间中的姿态，但有两种可能性(是朝向摄像机还是反方向)。因此，如果有几个这样的柱体组成的链，由于柱体是连接在一起的，而每一个有两种歧义性，那么可以得到 2^n 种不同的情况。由于人体的结构一般用 9 个柱体与一个球体来建模(脖子与头也可以合成一个柱体，但一般不这么做)，这种歧义性当中有许多是不能成立的，因为有些状态在运动学上是不可能的，因而可以排除掉。

实际情况下我们往往不知道 s 的值，因此不能直接使用上式，但是却可以用该方程式的派生形式。具体说来，假设已知两个柱体长度分别为与 l_1 与 l_2，那么它们的长度之比为 l_1/l_2，而它们图像长度之比为

$$r_{12}=\left(\frac{l_1}{l_2}\right)\left(\frac{\cos\theta_1}{\cos\theta_2}\right)$$

其中，与 θ_1 与 θ_2 是这些段与图像平面之间的夹角。如果假定有一个已知每段长度的链，而角度之比可以从图像中读出，这意味着如果确定了其中一个角度，例如 θ_1，就可以知道所有其他

的值。况且对 θ_1 也是有限制的，因为可以度量长度比 r_{ij}，那么就可有一组很有用的约束条件

$$\cos\theta_1 = r_{1k}\cos\theta_k\left(\frac{l_k}{l_1}\right)$$
$$\leqslant r_{1k}\left(\frac{l_k}{l_1}\right)$$

因此对 $\cos\theta_1$ 的任一值，只要它满足这些约束条件，就可以重构出一组已知长度比柱体的三维状态，但会有若干歧义性。

显然这些重构取决于 θ_1，但这些选择不仅受上述约束条件限制，它还要受人体结构的运动学限制。Taylor 研究的结论是选最大的值为假设是一种好的选择(见图 24.22 与图 24.23)。

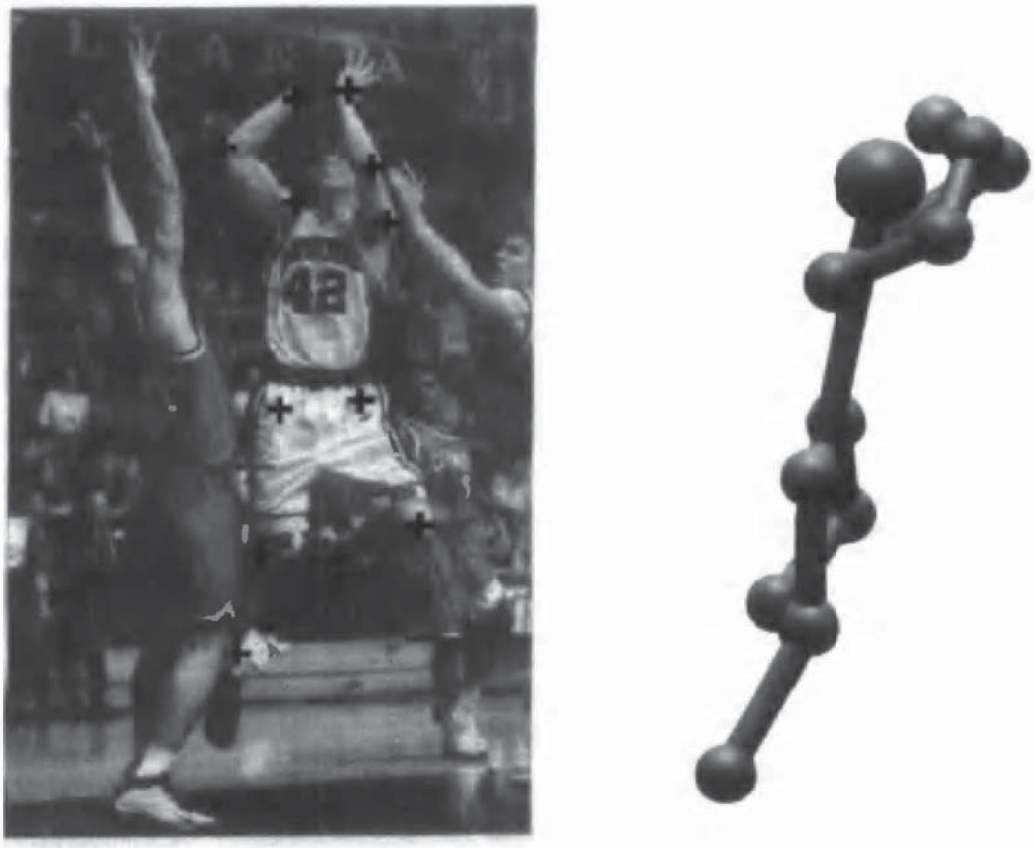


图 24.22　左图：一个人体图像，其中叉号表示由人工标出的身体各部分的末端。由于这些分段的相对长度是已知的，使用比例正交投影的假设可得到人体状态的三维重构，但这含有一些歧义性。右图：一个重构的侧视图，其中某些歧义性显示在图24.23

24.2.5　使用显式几何推理从图像数据推理三维物体

广义柱的定义是有点模糊的，广义柱与相应图像信息之间的几何关系也不很严密。但实际上还是有可能利用这些关系来建立系统，用它从图像数据推理由简单柱体组合的情况。

我们用柱体或锥体的层次结构为物体建模，柱体或锥体的截面可以是圆形的或多边形。每类物体用独特的组合结构表示(也就是用不同数量的不同类型的柱体)，而每个具体的物体则用不同的参数值表示(例如轴的长度、截面半径、一个柱体相对于另一个的位置等)。而类型

结构可以用这些参数的一组不等式表示。

图 24.23 在顶部是一个篮球运动员屈膝的图像，身体各段的末端是手工标注的。重构的歧义性是由一系列离散的选择以及一个连续参数组成，这个参数表示一段身体与图像平面之间的夹角。在底部显示由参数的不同选择获得的不同的重构。大体上说选择一个较大的角度值得到重构的躯干较长，膝也屈得更厉害一些

我们已学习过单个物体零部件所产生的轮廓的一些属性。一个物体类型的组合结构中只有某些组成部分会在图像中看到，因为有零部件会在投影中丢掉，而得不到。如果有摄像机模型，那么我们还能够利用定义物体类型的不等式进行推理。

由 Brooks，Greiner 与 Binford（1979 与 1981a）开发的 ACRONYM 系统用的就是这种方法。ACRONYM 通过一系列的预测、描述与解释步骤来识别物体。这些步骤是由一个几何推理系统驱动的，该系统能推断在三维（物体间）与二维（图像成分间）之间存在的空间关系。

例子：机身与机翼之间的关系 作为一个例子，考虑一个简化的飞机模型，它的机身是一个高度为 H，直径为 D 的中心柱体表示，机翼由两个高度为 H'，直径为 D' 与机身夹角为 θ 的对称柱体表示。用弱透视投影摄像机对飞机摄像，观察方向 $\boldsymbol{v}$ 用一个球坐标（α,β）表示，该坐标系统的 x 轴与机身的脊对齐，而 y 轴处在包含机身的轴与两机翼的平面内（图 24.24，左）。

如图 24.24 右所示，一个高度为 H，直径为 D 的柱体与视角方向 $\boldsymbol{v}$ 呈角的方式正交投影成一个 Brooks 带状基元，它的高度为 $h = H\sin\phi$ 直径为 $d = D$（忽略带状基元的椭圆端点）。在带放大系数的弱透视投影中，这些值变成 $h = \mu H\sin\phi$ 与 $d = \mu D$。于是表示机身与两个机翼的三个柱体所投影的带状基元的脊长，为 h, h'_l 和 h'_r，直径为 d, d'_l 和 d'_r，

$$\begin{cases} h = \mu H\sqrt{1-\sin^2\beta\cos^2\alpha}, \\ h'_l = \mu H'\sqrt{1-\sin^2\beta\cos^2(\theta+\alpha)}, \\ h'_r = \mu H'\sqrt{1-\sin^2\beta\cos^2(\theta-\alpha)}, \end{cases} \qquad \begin{cases} d = \mu D \\ d'_l = \mu D' \\ d'_r = \mu D' \end{cases} \tag{24.1}$$

因而，立即能得到下列与视点无关的图像约束：

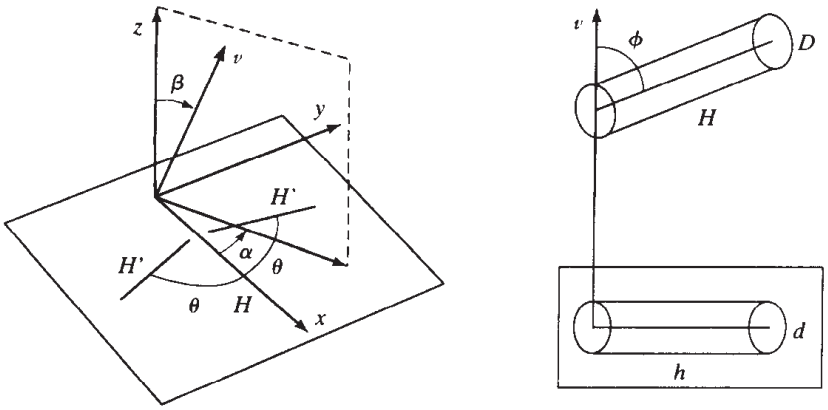


图 24.24　定义一个简单飞机模型投影的参数:(左)模型,机身与机翼只标出它们的脊,以免图像显得零乱;(右)柱体的正交投影

$$d_l' = d_r' \quad 和 \quad \frac{d_l'}{d} = \frac{d_r'}{d} = \frac{D'}{D} \tag{24.2}$$

将表示机身的带状基元的轴与左、右翼脊线之间的角度定义为 θ_l 和 θ_r。则很容易得到(见练习)

$$\begin{cases} \tan\theta_l = \dfrac{\cos\beta\sin\theta}{\cos\theta - \sin^2\beta\cos(\theta+\alpha)\cos\alpha} \\ \tan\theta_r = \dfrac{\cos\beta\sin\theta}{\cos\theta - \sin^2\beta\cos(\theta-\alpha)\cos\alpha} \end{cases} \tag{24.3}$$

这些式子表示一个平面双向对称的结构并没有投影成另一个双向对称的结构(也就是说在一般情况下,$\tan\theta_l \neq \tan\theta_r$)。

如果进一步假设是 β 小角度,忽略与有关的二阶及高阶项,就可得到:

$$\theta_l = \theta_r = \theta \tag{24.4}$$

同样精确到 2 阶有

$$h_l' = h_r' \quad 和 \quad \frac{h_l'}{h} = \frac{h_r'}{h} = \frac{H'}{H} \tag{24.5}$$

在 ACRONYM 系统中利用了各种不变量、准不变量以及对尺寸与角度的界限来进行预测。这些量是从物体与摄像机模型中推导出来的,推导过程中使用了一组启发式规则与一个功能很强的几何推理系统,该系统能够对复杂的三角与代数表达式进行运算与简化,并且能获取这些表达式的上界与下界。

构筑几何演绎　使用式(24.1)到式(24.3)有两种方法。第一种方法:给定一些对摄像机的约束条件,可以使所关注问题有关的带状基元的搜索范围变窄。第二种方法:给定所检测的带状基元的参数以及对应假设,可以得到对摄像机的约束。显然需要考虑图像描述过程中的误差。ACRONYM 使用了现在已过时的方法来传播界限,因此没有必要进一步讨论这些细节。我们更感兴趣的是它使用的处理过程,并可以与第 18 章相应的推理进行比较。假设一组带状基元与可能对应的物体已经确定。因此可以进一步得到摄像机的约束,并因而使对其他带状

基元的解释产生假设。这就意味着一小组对应可以派生出数量较大的识别假设。这样一来我们面对的是一致性标号问题:希望找到与几何约束一致的最大假设集。

ACRONYM 用扩展对应假设的方法来处理这个问题。它构筑一个表示可能的物体状态的预测图以及表示图像数据的观察图,然后寻找它们之间的匹配关系(算法 24.1 给出细节)。首先假设每一个带状基元的对应,然后使用相关的约束进行剪枝。然后用预测图中的弧上所要求的一致性约束条件来寻找带状基元对之间的匹配。然后一致性检验又进入在三个基元之间进行,以此类推。在每一步过程中通过约束传播维护全局的一致性。如此进行下去,直到处理到终点,由符合一致性连接的成分组成的解释图就形成了,它们对应于候选的物体模型(例如机场上的飞机)。最后进行一次全局性的一致性检测,来搜索符合全局约束条件的最大连接集(例如在同一机场上的所有飞机应该获得协调一致的视角参数)。图 24.25 说明这种方式使用的策略。

图 24.25 用 ACRONYM 对机场的俯视图进行解释。输入到系统图像(左上图)中有典型的宽体客机与L-1011型飞机(图24.12)以及标定过的航空摄像机模型,摄像高度在1 000 ~ 12 000米之间。在右上图是输入图像的边缘图。左下图是ACRONYM用它的预测模块找到的带状基元。这些带状基元的尺寸符合拍摄高度下相应机身、机翼、机尾的尺寸的限制。所识别的飞机在右下图显示。显然有几个飞机没识别出来。主要的原因是:(a)在分割时有些带状基元没有检测到;(b)由几何推理系统传播的约束界值超过相当限度,以致无法确定它们是否是合适的假设

算法 24.1　ACRONYM 识别系统

1. **预测**:构筑一幅预测图,它的结点是预测的图像带状基元,并有相应的参数范围,它的弧表示相邻带状基元之间的关系。
2. **描述**:构筑一个相似的观察图,它的结点是图像中的带状基元以及相应参数范围,以检测与预测图是否兼容。
3. **解释**:
   3.1 构筑一幅解释图,它的结点是预测与观察带状基元可能的匹配,使用预测图中有关的约束以确保参数的兼容性。
   3.2 使用约束传播方法来检测解释树的连通成分,要符合相关约束条件的一致性要求。
   3.3 使用组合搜索来检测一致连通成分的最大集。

24.3　后记:物体识别

计算机视觉从 1960 年开始已走过相当长的路,它的某些进步来自于计算机快速计算与图像系统价格的急剧下降,而更主要的是来自于对其各个问题加深了理解。因此,现在已形成许多实际问题都可以用计算机视觉研发的技术解决的局面。这的确是成功的,但是有许多核心问题还没有解决,以及很难用卓有成效的方法去思考。这些核心问题涉及物体的表达方式与识别。

怎样才能把识别表示成最一般方式呢?答案很可能是一层一层地进行模板匹配。但是模板又是什么、关系又如何表达、物体的层次表达又是怎样的,等等,仍然不清楚。按这种观点,基本的识别过程需要从图像信息开始,把在景物中不太可能出现的物体假设删除掉;能在场景中存在的物体种类就进入组合程序;这样可以搜集到更多的信息,又可在更深入层次上进行剪枝。这个过程一直持续到检测出物体。这完全是一种假设构造法,尽管这种思想很吸引人,但是人们还无法指出已存在一种能工作的程序,并且会取得成功。这是因为实现这种程序还充满着困难:例如,什么是组合线索?物体表达式如何组织?如何将新的物体附加到已有的表达式中?如何保证剪枝过程是在恰当的起点开始的,并且不会无功而返?

这些是一些非常困难的问题,因此计算机视觉进展缓慢是自然的。值得提醒的是,物体是什么还不十分清楚——人脸是不是物体,还是眼睛、鼻子与嘴的组合是物体?泥浆是不是物体?罪恶是否物体?什么需要识别并不清楚。我们的确想识别奔跑,但很可能并不认为它是一个物体;识别罪恶是很困难的事;不管人脸是否是物体,识别人脸是个好主意。我们并不知道什么样的区别要首先考虑,什么需放在次要的位置。例如,尽管老虎与豹在某些语义层次上很有相似性,但它们看起来很不相同。与此相似,一个小的海豚与企鹅在动物学上是没有相似之处的,但它们看起来却很相像。目前并没有一个令人满意的框架来讨论这些话题。

24.3.1　考虑识别问题的依据

人们可以给几千种不同物体命名,这种功能并不受单个物体外表的变化的影响(例如改变

椅子的装饰或式样)。况且人们只需看到一个新物体的少量例子就得到它的概念,并且日后见到该物体的其他样本时将它识别出来。

如果让计算机程序具备这种技能,哪怕是部分技能都是十分有用的。人们具有这些功能,因为这是实际的需要(知道吃什么,谁为你提供食品,什么时候打架,什么时候逃跑,什么东西要吃掉你等)。对大量物体需要识别来说,构造一个性能良好的物体表达式看来是一个关键问题。

层次表示是计算机科学家处理尺度关系的习惯方式,并且人们普遍相信表达与识别大量物体的关键,是将物体以某种层次结构组织起来。一般这种层次结构往往采用先考虑高层次区别,再考虑细节描述的表示方式。典型的情况是对物体看上去像但实际上很不相同的物体,外观的区别往往放在识别过程的后部(例如小海豚与大企鹅),而物体看上去很不同,但其实质上是很相像的(如鳗鱼与鱼)物体可能在识别过程中较早时就被分开。

一个理想的物体识别系统应做到:

- 识别许多不同的物体;
  [听起来不难,实际上这是非常困难:要识别大量物体,需要知道如何将它们组织到一个数据结构中去,这种结构在给定图像数据时要容易搜索。尤其是需要知道用什么度量来区分物体,而不是区分物体的个别范例(一只猫可能是有条纹的,而另一只可能是灰的,它们都是猫)。]
- 在各种不同背景下识别出物体;
  [这也是很困难的,理想情况是一个合适的物体表达式有助于将图像组织成片断,而这些片断来自于物体的种类(而不是针对具体的样本)及与物体无关的。]
- 在一个适当的抽象层次上识别物体;
  [人们并不需要在看到一个具体的椅子之前知道它是一个椅子。理想的情况是程序能够把豹与猎豹当做有斑点的猫,然后再对它们加以区分。在什么程度上进行抽象是一种合适的层次结构,这一点仍是一个秘密。至少这种话题的一部分是与识别许多不同的物体联系在一起的。]

我们认为与上述要求相比,目前的识别策略的性能是很差的。这并不是它们很差,而是因为问题很困难。

24.3.2 当前物体识别的方法

姿态一致性方法 对图像与模型的特征之间用几何方法获取足够数量的匹配,包括在第18章中讨论的对准技术与仿射及投影不变量。在前一种方法中,匹配过程是以树搜索方式进行的,而为了使搜索成本能够加以控制,往往利用少量匹配能完全确定物体姿态这一点,并预测任何进一步的对应特征的图像位置。后一种方法用一小群特征点,直接计算与视点无关的特征向量,因而能用来搜索存储所有模型的散列表。这种方法的好处是可用低于线性的时间来搜索。

模板匹配 将物体的所有图像的描述都记录下来。正像在第22章所讨论的,它们在诸如人脸检测与三维物体识别等任务上取得成功。与纯几何方法相比,它们的主要优点在于利用了图像的亮度/彩色信息中的巨大分辨能力。但是一般要求有一个单独的分割步骤,将感兴趣的物体与图像背景区分开,还对照明等条件比较敏感。

关系匹配器　这种方法通过描述模板之间的关系来描述物体。典型做法是先检测一些规范化的图像块，然后推理它们之间的关系（像第 23 章中讨论的）。这里有两种困难：其一，正如第 23 章中指出的，某些关系模型容易匹配，而某些则困难；其二，现行方法处理局部片断（如眼睛）与简单物体（像人脸）还行。但仍然看不清楚，为利用图像块之间关系识别动物之类的问题，如何建立匹配器。

外观图　外观图将因视点不同引起物体外观的定性变化明确地记录下来（在第 20 章讨论）的基于外观图的识别技术处在基于外观与基于结构的方法之间，因为它们描述物体的外观是通过将其变化描述成视点的函数。由于相同的物体具有相似的外观，它们可以具有相似的外观图，并且了解图像的结构可以作为图像分割的指导。但是在实际问题中精确的外观图并没有达到设想的效果，这部分是因为从实际图像中可靠地抽取端点、T 结点等轮廓特征是非常困难的，而另一部分是因为即使相对简单的物体可能会有非常复杂的外观图。

24.3.3　局限性

我们讨论过的方法是相当局限性的。主要的问题是尺度问题——涉及许多不同的物体以及许多不同的背景，局限性主要可分成三类：

分割　关系匹配器很容易被大量的候选所压倒。同样地，如果在待匹配的区域内包含其他无关的信息，模板匹配的性能也会下降，待识别的基本素材容易分割出来这一点可能是重要的，这样可以在不知道正在分析的事物是什么的条件下将有关联的图像片划分在一起。

种类抽象　现在还不清楚如何在合适的抽象层次上检测物体。尽管并不期望在哺乳动物这一层次识别动物，但是在我们担心是否面对豹或猎豹之前，应该能够检测带斑点的哺乳动物，几乎不知道怎样地抽取层次是合适的，理解这一点是很重要的：如何用视觉方式描绘一个农场？至今使用的描述方法远远做不到按类别来描述。需要指出的是，尽管划分类别是一个深人的认知论的秘密，但它又是具有实用根源的现象，要使识别大量物体容易些，则要先从粗略方面的区别进行划分，然后在注意精细方面的区别。但是如何划分种类仍然是个谜。

一般性　成功的识别策略应该能用到许多类型或物体种类而不依赖精细地调整参数。理想的情况是：因为合适的方法能够按照物体结构的基本规则组织，因而就能从少量的样本学习到新物体的满意模型。已描述的方法通常能加以推广，但推广的方法并不令人特别满意。

24.4　注释

考虑用表面成型几何模板的类别已渐渐不很流行。然而在这里进行讨论是因为整个思路很可能是对的。当然不能认为这本书中的答案就是广义柱，可以认为要“把注意力放在三维结构与二维图像结构之间的关系，特别要注意任何一种能把二维图像成分组合起来的线索”。视觉的许多问题已从概率推理的概念中受到启发，但是这个部分却没有。其中一个困难在于，这个问题中的许多概念都是以相对不很精确的方式描述的，需要有些抓主要特征忽略枝节的办法。另一个困难在于不清楚向哪个方向努力，“对表面结构的约束产生对图像的约束”这种想法作为关键呢？或者“如果不能发现分割物体的约束，那么也就不能识别它”这个问题要着重解决呢？

第一个能够执行三维物体识别任务的计算机程序要回溯至 Roberts(1965)。而对在人们与计算机中的识别过程建模问题,在 Biederman(1987);Bülthoff,Tarr,Blanz 与 Zabinski(1995);Marr(1982);Palmer(1999);Rosch(1988);Tarr,Hayward,Gauthier 与 Williams(1995)以及 Ullman(1996)中讨论过,这只是一些例子。但是在人脑中用来执行识别到任务的物体模型仍然不清楚,基于视图理论的支持者(例如 Bülthoff 等,1995;Tarr 等,1995)与基于基元的表达式的支持者(例如 Marr,1982 或 Biederman,1987)之间的争论仍在继续。

骨架(或中轴)是 Blum 于 1967 年提出的,在数字图像中研究骨架的是数学形态学方法(Serra,1982)。对各种类型带状基元的比较,包括 Blum 与 Brooks 带状基元还有平滑局部对称性(Brady 与 Asada,1984),斜对称(Kanade,1981)等可在 Rosenfeld(1986)中,或 Ponce(1990)中找到。FORMS 系统是 Zhu 与 Yuille(1996)开发的。还可阅读 Siddiqi,Shokoufandeh,Dickinson 与 Zucker(1999*b*)了解相关的工作。

广义柱是由 Tom Binford(1971)提出的,它又称为广义锥(Marr 与 Nishihara,1978;Brooks,1981*a*),早期关于从图像中抽取广义柱描述的大部分工作集中在距离数据上(例如 Agin,1972)。其中 Nevatia 与 Binford(1977)的工作是尤其重要的,因为他们确实实现了一种版本的广义位移不变量:他们的算法对诸如玩偶、马与蛇等物体的所有可能的截面方向进行试探,然后选择可能的候选截面,以及相应的平滑变化的参数。ACRONYM 系统是 Brooks 与 Binford(1979,1981*a*,1981*b*)开发的。SHGC 是由 Shafer 与 Kanade(1983,1985*a*)作为广义柱分类学的一部分引进的。正如前面所述,限定所关注的 GC 的类型使得预测这些表面投影与视点无关的性质成为可能。例如,Nalwa(1987)证明在正交投影下观察的旋转物体的剪影是双向对称的,而 Ponce 等(1989)指出,在正交投影与透视投影条件下 SHGC 上对应于同一个截面的点的剪影切线交于 SHGC 轴的图像上同一点。这种经过分析得到的预测为检测图像中每一个 GC 或基于投影不变量识别 GC 提供了坚实的基础。给人留下深刻印象的结果已见报道(例如可见 Zerroug 与 Medioni,1995)。Ponce,Cepeda,Pae 与 Sullivan(1999)对三维广义柱定义一个类似于平面上用的 Blum 变换,变换做了初步尝试。中轴变换(Blum 变换,草原火灾变换)曾因为它对噪声引起的不稳定性而招致反对。但它于近来又以两种形式得到了继承。一种形式是研究震动图(Kimia,Tannenbaum 与 Zucker,1990,1995;Giblin 与 Kimia 1999;Siddiqi,Kimia,Tannenbaum 与 Zucker,1999*a*,*b*),另一种则对轮廓加以一系列小的变动,然后观察变换的变化,并以其"平均"表示变换的结果(Zhu,1999)。

讨论影调基元的文章出现在 Haddon 与 Forsyth(1998*a*),其他一些在识别任务中使用并有某种成效的基元有超二次型(Superquadrics)(例如,Pentland,1986),在物体识别中开发利用函数的作用的初始尝试见 Stark 与 Bowyer(1996)。

习题

24.1 定义 Brooks 变换:考虑一个二维形状,其边界由曲线 Γ 包围,定义为 $\boldsymbol{x}: I \rightarrow \mathbb{R}^2$ 并且用弧长作为参数。对 Γ 上任何两个点 $\boldsymbol{x}_1 \stackrel{\text{def}}{=} \boldsymbol{x}(s_1)$ 与 $\boldsymbol{x}_2 \stackrel{\text{def}}{=} \boldsymbol{x}(s_2)$ 的连线定义了该形状的一个截线,长度为 $l(s_1, s_2) = |\boldsymbol{x}_1 - \boldsymbol{x}_2|$。于是我们可以将研究该形状的截线集的问题演变成研究一个表面 S 的地貌的问题,该表面与一个高度函数有关 $h: I^2 \rightarrow \mathbb{R}^+$ 高度函数定

义为 $h(s_1, s_2) = \frac{1}{2} l(s_1, s_2)^2$。根据以上定义，由 Γ 确定的带状基元可以定义为(Ponce 等，1999)一组截线，它的末端对应于 S 的谷(根据 Haralick 1983 或 Haralick、Watson 与 Laffey，1983)，符合以下条件的点对(s_1, s_2)。条件是 h 的梯度 ∇h 是 Hessian 矩阵($\mathcal{H}$)的特征向量，而 Hessian 的另一个特征向量的特征值为正。

令 u 表示单位向量，使得 $\boldsymbol{x}_1 - \boldsymbol{x}_2 = l\boldsymbol{u}$，$\boldsymbol{t}_i$ 分别表示在$\boldsymbol{x}_i$($i = 1,2$)处的单位切线，而 θ_i 与 κ_i 则分别表示 u 与 $\boldsymbol{t}_i$ 之间的夹角以及在 $\boldsymbol{x}_i$ 处的曲率，证明与有关的带状基元是这形状的截线集，它的端点满足

$$(\cos^2\theta_1 - \cos^2\theta_2)\cos(\theta_1 - \theta_2) + l\cos\theta_1\cos\theta_2(\kappa_1\sin\theta_1 + \kappa_2\sin\theta_2) = 0$$

24.2　广义柱：在前一题中，对谷点的定义对在 n 维域定义的高度表面是有效的，并且谷点在任何维形成曲线。简要解释如何将上练习中给出的带状基元的定义扩展到广义柱的新定义。会不会有在高于二维情况才会遇到的困难？

24.3　斜对称：具有一直轴的母线与轴有一个固定角度 θ 的 Brooks 带状基元是斜对称的。斜对称在绘图分析中起重要作用，因为可以证明在正交投影下双向对称的平面图形投影成斜对称(Kanade，1981)。证明形成斜对称的两个轮廓点 P_1 和 P_2 满足下式

$$\frac{\kappa_2}{\kappa_1} = -\left[\frac{\sin\alpha_2}{\sin\alpha_1}\right]^3$$

其中，κ_i 是在 P_i($i = 1,2$)处斜对称边界的曲率，α_i 表示连接此两点的连线与此轮廓的法线之间的角度。

提示：构筑斜对称的参数表达式。

编程作业

24.4　写出一个基于腐蚀的抽骨架程序。该程序迭代地处理一个二值图直到该图形不再发生变化。每次迭代分成 8 步，第一步：对输入图像中的任何像素，如果它的邻域的图形与下面左模式匹配(其中“*”是指该位置的像素值无关紧要)则在一张辅助图的相应位置上赋值为“0”，而所有其他像素则保留在相应输入图像中的原值。

0	0	0
*	1	*
1	1	1

0	0	*
0	1	1
*	1	1

将辅助图像复制到输入图像上，用右模式重复该过程。每次迭代的下续步骤是相似的，但使用的 6 种模式是通过对原模式施行相继的 90°旋转得到的。这个程序的输出是原区域的 4 连通骨架(Serra，1982)。

24.5　实现用于骨架检测的 FORMS 方法。

24.6　实现 Brooks 变换。

24.7　写一个用于检测斜对称的程序。可以选择使用：(a)比较所有轮廓点对的幼稚方法，复杂度为 $O(n^2)$，或(b)由 Nevatia 与 Binford(1977)提出的 $O(kn)$投影算法。后一种方法可概括为：将带状基元的局部轴可能的方向离散化，对总共 k 个方向的每一个将所有点投影到容器中，在每个容器中的点检验斜对称条件，最后将得到的带状基元边对组成带状基元。

第七部分　应　　用

- 第 25 章　应用:在数字化收藏库中查找
- 第 26 章　应用:基于图像的绘制

第 25 章　应用:在数字化收藏库中查找

大量数字化图片库迅速地涌现出来。对一些图片库进行数字化是为了更好地保存、更容易发布和更好地存取。另一些本来就是数字化的,如家庭照片的个人收藏(它可以是巨大的和数字化的)、网页(它是巨大而无组织的集合)还有家庭视频(一些集合也都很大,而且现在很多都是数字化的)。

现在用来与文档或数据收藏库进行交互的工具十分复杂。一些典型的做法是,人们可以用各种文本匹配方法来搜索一个集合,可以对一个文本集合进行聚类,也可以用数据挖掘技术。数据挖掘涉及到用复杂训练过程来查找以前未知的趋势(这个有用的娱乐被称为"探索数据挖掘",一个不太令人兴奋的名字,和有时也被那些不赞成的人称为"数据挖泥")。一般地说,一个收藏库价值的重要组成部分来自于有没有这样的工具出现。要弄清其原因,可以想像一下去一个大的旧书店,里面的书按封面的尘土颜色分类,尽管这个库很大,也不能想像你会光顾这个书店,除非你无其他选择。

现在还很难以令人满意的方式组织和搜索图像库,也就是说它们有点像组织得很差的书店。困难在于建立合适的图像信息表示。用手标注每一张图是没用的,因为用一段好的文字描述一张图像是困难的。一些库非常巨大(上千万张图片;Enser, 1995)。手工索引一个大图片库要大量的工作,而且,还会出现重新索引这个库的某部分的可能性,例如,一个新闻事件会使一个以前无名的人著名起来,那就最好要知道这个库是否包含了这个人的图片。最后一点是,一般都很难知道一张图片的"含意"是什么。

尽管有这些困难,但任何有助于管理图片库的技术都有巨大的实际应用范围。一个重要的工具是检索——找到符合标准的图片——但是这绝不是我们惟一需要的。一种可能是希望以一种支持浏览的方式组织图片,使得内容相似的图片彼此邻近;另一种可能是希望发现某种趋势或具有某种工具来识别重要的变化。

典型应用如下:

- **规划和管理:**有大量的人造卫星拍摄的地球图片,可以用于为重要的政治辩论提供依据。例如,城市区域扩展了多远?雨林还剩多少?等等(例如,Smith, 1996)。
- **军事侦察:**卫星图像可以包含重要的军事信息。典型查询包括查找军事上感兴趣的变化,例如,这里有军队聚集吗?上次炸弹袭击造成了多大的破坏?今天发生了什么?等等。总之,是地球上特定地点发生的事件(例如,Mundy 和 Vrobel,1994)。
- **库存照片和库存数据:**有些商业库经常有数量巨大和品种繁多的收藏品,靠出售使用某些图像权利来赚钱(Enser,1993,1995)。
- **访问博物馆:**博物馆正在建立分辨率有限的收藏品网上视图,以此来诱惑观众亲自来博物馆观看(Holt 和 Hartwick, 1994*a*; Holt 和 Hartwick, 1994*b*; 和 Psarrou, Konstantinou, Morse 和 O'Reilly, 1997)。理想的情况应该允许观众浏览收藏品,获取博物馆里有些什么的感性知识。
- **商标和版权的维护:**随电子商务的增长,对侵害商标和版权进行自动搜索的机会也会增

加。(Eakins, Boardman 和 Graham, 1998; Jain 和 Vailaya, 1998; Kato, Shimogaki, Mizutori 和 Fujimura, 1988; Kato 和 Fujimura, 1990)。例如,在写这本书的时候就出现了这样一个组织,叫做 BayTSP,图片主人可以向它注册,让它可以在网上搜索被盗版的图片。

- **索引网页**:索引网页是个有利可图的行动。用户还希望有工具能够阻止攻击性质的图片或广告。有几个现成的工具已经能够支持在网上搜索图片,有些技术在后面会讲到(Cascia, Sethi 和 Sclaroff, 1998; Chang, Smith, Beigi 和 Benitez, 1997*b*; 或 Smith 和 Chang, 1997)。
- **医学信息系统**:发现与示例查询类似的医学图像,可以给出更多的信息,这些信息可以作为诊断或进行流行病学研究的依据(Kofakis 和 Orphanoudakis, 1991;或 Wong, 1998)。而且,人们可以按某种方式将医学图像聚类,以便为专家们提供有趣的或新颖的假设。
- **图像数据挖掘**:数据挖掘的吸引人之处在于人们可以用大量数据库探索。有时一个数据挖掘方法会提出一个真正有用或新颖的假设,并交付给相应领域的专家去验证。许多图像库可能会支持某种相似的现象。例如,有大量佛教艺术品的数字化图像集,连同地质数据(这个物体是在那里找到的吗?)以及各种专家的意见。例如,如果能够从图像中找到相应的表示,则可以跨越空间与时间的角度分析人形变化的趋势。

所有这些应用的核心问题是如何表示图像,或说图像内含表示的本质。一旦确定了一种表示,进行搜索(基于类似于某种表示来查找图像)、组织(将这种相似表示的图像组织在一起)或发现其中体现的趋势(查找这种表示下的各组成部分间的关系)就是相对容易的。理想表示最好是图像中所有物体的完全描述,但是在可预知的将来,用计算机程序产生这样描述的希望很小。然而,即使很粗糙的表示方法看来在当前也还是有用的。在这章里,我们首先回顾信息检索方面的基本知识,再讲述当前用计算机视觉工具组织和检索图像库的一系列方法。

25.1　背景知识:组织收藏的信息

信息检索是研究用各种信息从集合中获取信息条目的系统。我们对这个课题感兴趣,是因为信息检索研究者们已经成为性能分析的擅长者,而这通常是很难的。

25.1.1　信息检索系统运行状况如何评价

典型地,信息检索系统的性能用"检索率(recall)"和"准确率(precision)"来描述。前者是说实际找到的相关条目的百分比,后者是说检索到的条目里相关条目的百分比。在这里对"相关"这个词下定义是困难的,如果要对检索率和准确率给出度量的话,需要知道什么样的条目是和查询相关的。这是个有经验的人才能区分的问题。

一个系统好的标准取决于应用　当人们改变系统配置使得"检索率"上升时,典型的情况是"准确率"就会下降。一般似乎认为好的系统总是有高的"检索率"和高的"准确率",但是这不是实际情况。相反,对于一个系统好的标准取决于应用。

专利搜索:如果专利的内容与已经存在的技术和资料的概念相似,专利申请就会失败。这就是说要能找到这样的资料,能检验出所要申请专利的内容尚不存在(这就可以申请专利),或者已经存在(也就推翻了一个无效的专利申请)。申请一个专利可能是昂贵的,而专利被推翻对专利拥有者来说是悲惨的(对挑战者是有利的)。这也就是说,宁可给查询者提供较多不相关的材料,

也不要漏过相关的材料。这也就是说高的“检索率”是必要的,即使以低的“准确率”为代价。

网页和电子邮件过滤:对于网页和电子邮件过滤服务来说,有一个广泛的需求。例如,在美国,公司经常担忧大量包含色情图片的内部电子邮件会引起法律责任。一种方法是用程序搜索有嫌疑图片的电子邮件。经理希望如果程序认为图片有问题,系统就能报警并显示有问题的图片。在写作本书的时候,已经有几个这类程序的提供商,但还不清楚这是不是一个能赚钱的应用,所以当你读到这里的时候,他们可能已经不做这个生意了。在类似于这样的程序里,高“检索率”是不重要的,当然“检索率”高不会有什么坏处。如果这个程序只能有 10% 的“检索率”,它仍将面对少量图片通过这个程序的困难。高的“精度”是重要的,是因为“狼来了”的效果。人们也不喜欢产生大量错误报警的系统,不大会用那样的系统,更重要的是人们更不想去升级和维护一个低“准确率”的系统。

寻找新闻条目:现在有各种各样的服务,它们向新闻组织提供图片和视频存储。服务商总想拥有大量著名图片,可以想像一个好的照片存储服务可能会有上千张曼德拉或戴安娜王妃的图片。这种高“检索率”搜索的效果是很恼人的,没有什么图片编辑器真的想去处理成千上万张图片。典型地,图片存储组织使用他们的专门技术以及与客户交流,以便只给用户提供相关少量图片。

评估系统 合理地评估一个系统一般是很难的。通常要画出在不同“检索率”下的“准确率”曲线(靠改变匹配阈值),然后依典型的查询将这些曲线取平均。人们可以加权“检索率”和“准确率”来反映它们在不同应用中的重要程度,计算平均的效用分数。好的实验很难做,而更糟的是,坏的实验却很好做。这是因为一般很难说出什么是符合与查询相关的条件(也就是说根据一个查询应该返回什么——用“女王”这个词查询应该返回怎样的图片?),要说出在一个大收藏集里应该有多少相关的条目就更难了(想像一下在有着上千万图片的集合中计算所有与“女王”相关的图片)。用被测试系统来确定哪些图片是相关的显然是一个坏的主意(可能会导致 100% “检索率”的不准确声明),这意味着必须找到在收藏集中统计相关条目的其他方法。如果我们小心地进行计算,尤其是估计得保守一些,那么对系统“检索率”的估计就会高一些。

25.1.2 用户想要什么

对图像集的用户行为进行过最全面研究的是 Enser 在 Hulton-Deutsch 集合上所做的工作(Armitage 和 Enser, 1997; Enser, 1993; Enser, 1995)(这些论文发表后,这个集合已经被新主人获取,现在被称做 Hulton-Getty 集)。这个库是一个集照片、负片、幻灯片的集合,似乎是一个主要用于媒体专业人员使用的集。Enser 研究了客户的请求表单,他将请求分成 4 个语意类别,其依据取决于所要求的是物体类的一个实例,还是经过提炼的实例。重点包含:所使用的特定检索语言仅给出“指向 Hulton 集合的粗略的指针”(Enser, 1993, 第 35 页)和摘要语意去描述图像。例如,用户需要遗物、物理学者、腌鱼的烟的图像。所有这些概念都超越了当前图像分析技术所能达到的。基本上没有一个工具能直接陈述需要。对于可以预见的将来,设计图像检索工具的主要限制是我们对视觉十分有限的理解。

然而,对视觉十分有限的理解也可以构造有用的工具(这个非常重要的观点常被忽视),但很难去评价怎样算是成功。Enser 说 Hulton-Getty 索引系统成功的标志是,其经营组织是盈利的。在实际中进行这样的测试有些困难,但是已经有几个可利用的产品了。IBM 生产了一个图像搜索的产品——QBIC(根据图像内容查询),它已经出现在大量的市场广告中,并且看上去是成功的。相似地,Virage——一个主要产品是图像搜索引擎的公司看上去也显得很生意兴隆。

25.1.3 搜索图片

表示图像的方式粗略地分有三种:在像素级,人们对具体的像素值感兴趣;在组合级,人们关心图像的整体外观;或是在对象语意级,人们关心图像所描述的事物。

像素匹配　在像素匹配中,我们寻找尽量接近示例的那些图像,示例可以是绘制的或提供的。理想的匹配可能是在每个位置都有相同的像素值。这种形式最知名的系统当属 Jacobs, Finkelstein 和 Salesin(1995)的系统,它应用一系列不同尺度的图像滤波器,对示例图像和其他图像应用滤波器得到的响应进行比较。可用不同尺度安排这个比较的顺序,如果在粗尺度上滤波器的响应都不匹配,就不用去检察精细尺度的响应了。

像素匹配器仅在使用者知道被搜索图片是什么样子时才有用。然而,对没有知识的使用者(也就是那些不了解图片集的用户)来说,情况也不总是这样的,在一些重要的应用中像素匹配是很有用的。一种重要的应用情况是在版权保护中应用。窃取数字图像是很容易发生的,他们使用图像而不给版权所有者付费。使用合适的图像匹配器,可以很容易找到窃取者。

找到窃取者的工作过程大体是这样的:版权所有者向一个专门查找窃取者的组织注册,并付费。这个组织使用一种蜘蛛(spider)程序来搜索网页,并下载图片,然后用像素匹配器将下载图片与注册图像进行比较。一旦发现侵权图片,律师以及罚单和版权费的委托人就会发出强制信函。尽管需要一些额外操作来确保裁剪和旋转过的图像之间进行匹配,以及匹配过程的效率,基于滤波器响应的匹配对这个应用就足够了。

对像素匹配来说,还有一些其他可能的应用。例如,当前流通的大量儿童色情文学已经有些年头了,这可以追溯到 1970 年以前。调查儿童色情文学的代理处需要确定其内容是否涉及新的材料,在这种情况下,这里有比起诉持有或发行更多的事要做,因为材料文档可能正在被滥用。用已知材料的参考库匹配可以满足这个需求。而且,匹配过程可以用来将在不同诉讼案中涉及不同被告,但同一内容的原告联系在一起进行联合调查。写作本书时,法律强制代理处显然没有以这种方式使用参考集合,但是这种用处是一个广泛议论的话题。我们在这一章里将不继续讨论像素匹配,因为主要关注点是组织图像库的问题。

使用全图匹配　在一些应用中,重要的是整幅图的结构。在这些应用中,我们把图像看做是彩色像素点的排列,而不是一个物体的图片。这种抽象一般叫做“外观(appearance)”。外观和物体的语义之间的区别不容易说清,因为如果不靠外观,我们怎么能知道一张图片里面有什么?这个表示方法在实际中是重要的,因为图像本身自动地表示外观,而不需要分割图像。

当图像的组成很重要时,外观就特别有用。例如,人们可以用外观提供的线索和关键字的组合来搜索存储的图像,并要求用户能够排除图像的组成正确,而语义不正确的图像。用外观来表示图像的核心技术问题是定义一个有效的图像相似度概念。25.2 节介绍了各种不同的策略。

对象级语义　Enser 的研究表明,使用照片存储集合的人按照语义搜索图像(如熏黑的腌鱼)。基于外观的方法很难实现这种查询,也几乎不可能构造出处理语义查询的物体识别程序。然而,我们可以建立对象级语义表示。典型的做法是先分割图像,然后建立围绕分割的表

示。目前还不知道怎样建立能够处理高层语义查询的搜索工具，也不知道怎样建立通用搜索工具的用户界面。虽然如此，当前技术能够产生对各种特殊情况十分有用的工具(25.3 节)。

25.1.4 组织和浏览

图像搜索会引出一系列难题。例如，假定有一个完美的物体识别系统，怎么来描述想要的图片呢？其实，搜索通常不像看上去的那么重要。除非图片集合里有某种类型的模型存在，否则是很难制定一个有用的搜索的。譬如，想一想你到一个新的商店里会怎样做：你先试图确定商店里货物的类别，然后再找需要的东西。你是不会向一位书店售货员询问有关汽车的情况的。这说明浏览过程是重要的交互方式，当然只有在图像库经过适当地组织后它才显得有用。理想的浏览工具应该做到：

- 显示相似的图像——这是说它们看上去相似、有相似的外观、在集合中相互靠近，或包含相似的内容等——用一种方式使它们的相似性显现出来；
- 显示图像聚类的表示，这使得用户很容易知道这个库里大概有什么(相似的聚类可能彼此邻近，大的聚类可能很大，等等)。
- 提供一些交互的形式，使得人们能够在不同细节层次上观察这个库(人们可能仅仅想看到某个特定的聚类元素或看到一个聚类集附近所有图像的概要)，并且能够从不同的方向观察这个集合。

浏览工具和搜索工具本质上是互补的。用户可能先浏览这个集合而后构思一种搜索策略。用户可能在搜索完毕后，再次选择对与搜索工具返回相近的条目进行观察。

建立有用的浏览工具也需要一个有效的图像相似性概念。为这类系统构造好的用户界面是困难的(正如下面的例子所示)；理想的浏览工具应具有的特征包括：清晰简单的查询说明过程和由程序使用的清晰的内部表示，这样即使浏览失败，也就不致使人过于迷惑不解了。典型地，用户希望用提供示例图像或填写表单的方式来搜索第一幅图，然后靠单击浏览工具所提供的样本，在集合里进行浏览。

25.2 整幅图的概要表示

图像通常是高度风格化的，特别是艺术家要突出表达一个特定的物体或心境时。这说明图像的总体布局可能是其描述内容的向导，那么可以这样构造有效的查询机制：查找与示例相似的图像、查找与梗概示例相似的图像，或与外观的文本描述相似的图像。这些方法有效与否，在于从什么含义上观察图像的相似性。向用户传递图像看上去相似的含义是重要的，因为任何一点恼人的错误都可以导致令人困惑不解。一个好的相似度概念对于浏览的效率也很重要，因为能表明图像差异程度的用户界面，可以展示出一个图像布局，来表明图像集合有关部分的整体结构。

25.2.1 直方图和相关图

一个流行的相似性度量是比较图像中特定颜色类别像素的数目。例如，日落场景和田园

场景在这个度量下是不同的,因为日落场景中包含许多红色、橙色和黄色像素,田园场景里占优势的是草绿色、天蓝色,可能还有白云的颜色(如图 25.1 所示)。而且,日落场景之间都是相似的,多数都有红色、橙色和黄色像素,其他的颜色就很少。

图 25.1　加州大学伯克利分校的 Calphotos 集合中搜索田园场景时得到的查询结果。这个查询是由搜索包含许多绿色和浅蓝色像素的图像得到。如结果所示,颜色直方图是十分奏效的

颜色直方图是一幅图像或一个区域里像素点落入某个颜色空间(RGB 是通用的,其原因我们不能解释)特定量化间隔的数目的记录。如果一个物体的图像的颜色直方图与另一幅图像的直方图匹配,如果光照变化不大,那么该物体就很可能出现在那幅图像中。这种测试可能对视角和尺度很敏感,因为特定颜色像素的相对数量会剧烈改变。尽管如此,它的优点是快速和容易,并且可以应用到类似衣服这样色彩鲜明但形状不好识别的物体上。

颜色直方图匹配非常流行,它至少能追溯到 Swain 和 Ballard(1991)的工作,并且已经被用在几个实际使用的系统中(Flickner, Sawhney, Niblack 和 Ashley, 1995; Holt 和 Hartwick, 1994*b*; Ogle 和 Stonebraker, 1995)。尽管颜色直方图这种表示方法丢失了许多图像信息,但它的有效性的确有点令人吃惊。例如,在 ATT, Chappelle 等人展示了仅利用颜色直方图就能对 Corel 集合里的图像按类别分类。Corel 集是个有 6 万张图像的集合,在视觉研究中经常使用,以前可以从 Corel 公司的三个系列之中得到。

颜色直方图不记录颜色像素的空间分布。因此法国和英国国旗在颜色直方图的度量下就十分相似。每个都有相同数目的红色、蓝色和白色像素,而它们的空间分布却不一样。使用颜色直方图可能引起的一个问题是,从稍微不同角度获取的图像在颜色直方图度量下可能差异很大。考虑某种颜色的像素到另一种颜色像素的某种距离的概率可以减轻这个影响。这种概

率可以用在各种距离下像素计数的方法计算。对于摄像机的微小变动，这些概率基本上不变，因此通过这些颜色相关图的相似性得到了图像相似性的度量。要求颜色相关图相似为图像相似性提供了的另一个度量。这里的计算细节由 Zabih 和他的同事们提出（Huang，Kumar，Mitra，Zhu 和 Zabih，1997；Huang 和 Zabih，1998）。

25.2.2 纹理和纹理的纹理

颜色直方图不包含有关颜色像素空间布局的信息。显式记录位置是要做的下一步。例如，雪山的图像会在顶部有较蓝的区域，中间有较白的区域，底下也有较蓝的区域（山脚下有湖）。而瀑布的图像在左边和右边会有较暗的区域，中间有较亮的竖直带状区域。这些位置模板是由 Lipson，Grimson 和 Sinha(1997)引入的。这些位置模板可以从一组图像中学到，这种方法相对于颜色直方图来说是很大的进步。

下一步很自然地要考虑图像纹理，因为可以用纹理来区别事物。例如，可以用纹理来区别一片花（很多小的橙色块）和一朵花（一个大的橙色块），或区别 dalmation 和斑马。尽管纹理这个概念很难定义或不可能定义，但是多数人提到纹理时知道它是什么。典型的说法认为纹理是小块图案的空间排列（如，格子呢是小块方块和线的一种排列，而草场的纹理是细条的一种排列）。

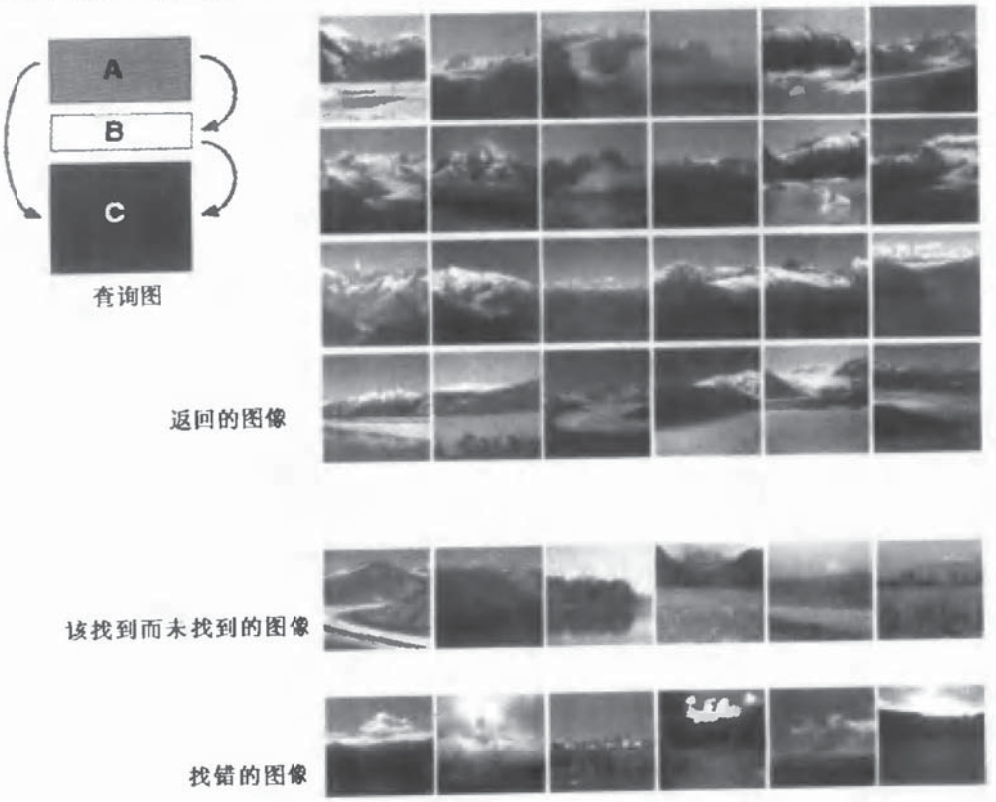


图 25.2 彩色区域的空间布局是多种类型图像内容的自然向导。左上角的图说明了雪山场景的颜色区域分局；右上图，用这个方法找到的结果里确实有雪山图片；中图：集合里存在但未被找出的雪山图片；底图：用这个方法找到的但不是雪山的图片

寻找这些子图案的通常策略是对图像应用线性滤波器（参见第 7 章或 7.6 节），这种滤波器的核看上去很像图案的基本元素。从滤波器理论看，滤波器的强烈响应说明这里有特定的图案呈现。可以应用不同种类的滤波器，根据在不同位置的响应统计，把图像分解成斑点区域和条纹

区,等等(Ma 和 Manjunath,1997*b*; Malik 和 Perona, 1990)。

首先可能想到的纹理描述是滤波器响应的直方图。例如,人们可以用几个小的黄色块查询图像。这种机制十分成功地应用于伯克利 Calphotos 集合中(http://elib.cs.berkeley.edu/photos; 其中有几千张加利福尼亚自然资源、鲜花和野生动物的图片)。

对于摄像机运动,纹理直方图存在一些问题:当摄像机接近场景时,图像中的细节变大。减小这种影响的方法是在图像上定义一组合适的变换(如,在一定尺度范围内放缩图像)。逐个应用这些变换,并将两幅图的相似性度量定义为不同变换下两幅图的最小差异。例如,我们可以用合理的因子放缩一幅图并寻找颜色和纹理直方图的最小差别。这种由 Rubner, Tomasi 和 Guibas (1998)提出的大型推土机距离(earth-mover's distance),允许各种变换。而且,它已经结合到一个组织图像的过程中,使得所显示图像之间的距离反映了集合中图像之间的差异性。这个方法允许快速和直观地浏览(见图 25.3)。

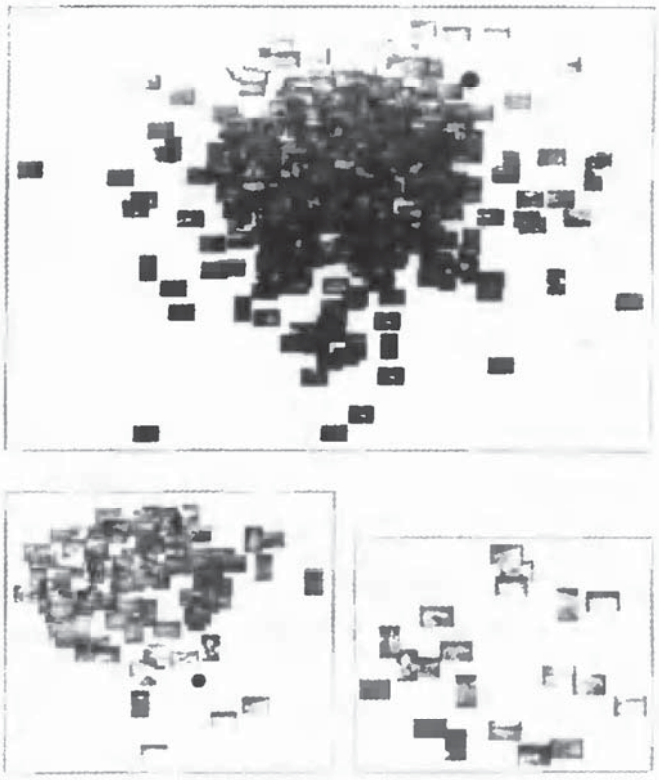


图 25.3　大型推土机距离(EarchMover's Distance,EMD)相似性度量下的图像分布。EMD 能够快速计算,于是类似这样的显示可以在线建立,而所显示的图像间距离尽可能忠实地反映出它们之间的EMD。上面的图显示了查询返回的大量图片。这种显示表示出集合的总体情况,用户只要用鼠标点击与要查找的图像相似的邻域,检索系统就知道了下一步往哪搜索(上面那幅图像中的黑色圆圈指示出用户点击的位置;这产生了右下的显示,然后同样的操作导出左下的显示)。用户用这个技术在图像数据库中浏览和导航,就像在百货公司里浏览一样。由于显示出了大量图像,以及图像间直观的距离,用户就能迅速在头脑里形成有关数据库内容的模型,并很快知道他们要找的图像在哪里

纹理的空间布局是一个强有力的线索。例如,在航空图像中,房产开发区有很特别的纹理,这个纹理的布局提供了搜索的线索。Ma 和 Manjunath 在加州大学圣·芭芭拉构造的 Netra 系统中,将纹理分成不同的风格簇(生成一个纹理辞典),用于分割大的航空图像;这个方法利用了这

样一个事实:尽管这里可能有很多的纹理簇,但只有某些纹理差别是显著的。用户可以用示例区域来查询集合中的相似图像;例如,在不同时间或日期获取特定区域的航空图像来跟踪一些事件,如发展进程,交通状况或植物生长等(见图 25.4; Ma 和 Manjunath, 1997*b* 和 1998; Manjunath 和 Ma, 1996*a* 和 1996*b*,*c*)。

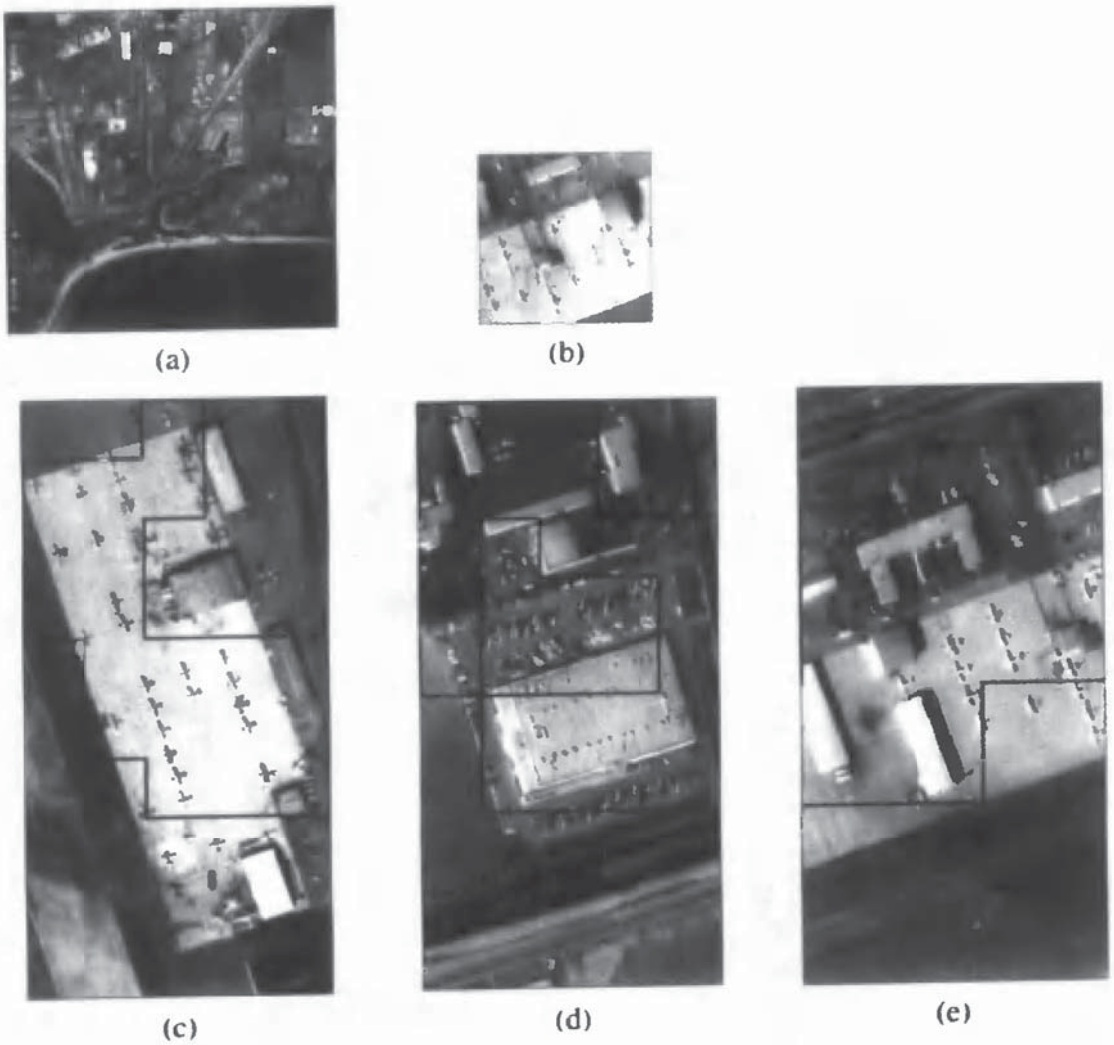


图 25.4 航空图片基于纹理的搜索。(a) 一航空图片的下采样,查询从该图开始。(b) 在查询中使用的一个区域全分辨率细节图,该区域有飞机、汽车与建筑物。(c)~(e)显示查询三个最好的结果。该结果来自三张不同的航空图片,第二与第三个结果是与查询图同一年(1972)拍摄的,第一幅是1966年的。更多细节参见(Ma, Manjunasha, 1997a)

纹理响应的区域也形成图案。例如,如果一幅图像显示了穿着带斑点衬衫的步行者,那么斑点检测滤波器将有许多强响应;强响应区粗看上去像一个大条。一组穿斑点衬衫的人将看上去像一个条的簇,它本身就是一个纹理。这些观察导致将纹理寻找滤波器应用到"纹理寻找滤波器"的输出上(可能会重复几次),并使用它们响应的相似性度量作为图像的相似性度量。这个由 de Bonet 和 Viola (1997)和 Tieu 和 Viola(2000)提出的方法涉及了大量的特征,因此要求用户填表单是不切实际的。一种切实的方法是让用户提供一些能说明所要寻找图像类型的示例图像。然后我们从这个集合里选择小的随机子集作为反例图像(这样做是可行的,因为随机选择的图像基本上不是用户要找的),然后利用正例和反例集合来建立分类器,将图像分成相关的和不相关的

两类(其中的细节有些偏离我们的介绍;第 22 章给出有关分类器的更多信息)。图 25.5 给出了使用这种方法的例子。

图 25.5　查询使用纹理的纹理方法。用户选了三张汽车照片为正例,它们对大的水平带滤波器响应比其他滤波器强烈。然后该系统随机地选取一些图像(用来表示无关的图像)这些数据用来建立一个分类器来锚列可能有关的其他图片。该查询检得到的图像中包含了若干汽车图像

25.3　图片的分部表示

这一节讲述的工具试图或多或少地直接估计对象级语义。这样的系统的典型做法是,首先要分割图像并着眼于分割出来的某些图像区域。

集合的结构对于寻找语义很有帮助,因为它可以用于指导特定搜索机制的选择。Pentland, Picard 和 Sclaroff(1996)开发的 Photobook 系统可以提供三个主要搜索类别:Shape Photobook 用轮廓的弹性变形度量形状的方法,来搜索孤立对象(如工具或鱼等);appearance Photobook 用少量主分量来查找人脸;texture Photobook 用纹理表示来查找有纹理的材质样本。

25.3.1　分割

人们将图像分解成与我们感兴趣物体相对应的区域,分类是达到这种分割的一个办法。分割是个关键的思想,因为它意味着在比较图像时可以把不相关的信息丢掉。例如,如果搜索老虎的图像,那么背景是雪还是草是无所谓的;只有老虎是要找的目标。然而,如果用整幅图像去度

量相似性，一只在草地上的老虎与一只在雪地上的老虎就不相同。这些观察说明应该先将图像分割成属于适当场景的像素区域，然后再让用户搜索特定区域的属性。最自然的分割方式应该是同一个物体的像素属于同一区域。目前，满足这个标准几乎是不可能的，因为我们不知道怎么做才能达到这个目的。然而，物体通常形成一致的颜色和纹理图像区域，使得属于同一区域的像素有希望属于相同物体。

Smith 和 Chang(1996)开发的 Visual SEEK 系统，将图像自动分割成色彩一致的区域，允许用户基于空间布局和色彩区域范围查询。对日落图像的查询可以用指定橙色背景和在橙色景上的一片黄色区域为模型。

Blobworld 是 Belongie, Carson, Greenspan 和 Malik(1998)建立的系统，它用具有一致的色彩和纹理的区域集合来表示图像(Belongie 等，1998)。这种表示呈现给用户的是，将区域颜色和纹理显示在椭圆块里的方式，来表达图像区域的形状。这些区域形状的表达很粗糙，是因为考虑到区域边界细节并不重要。使用者可以指定示例图像中哪个块是重要的和应当保持什么样的空间关系来查询系统(见图 25.6)。这些查询也可以结合文字信息。正如图 25.7 和图 25.8 所示，图像和文字是互补的。25.3.4 节将讲述一些利用这种互补特性的应用。

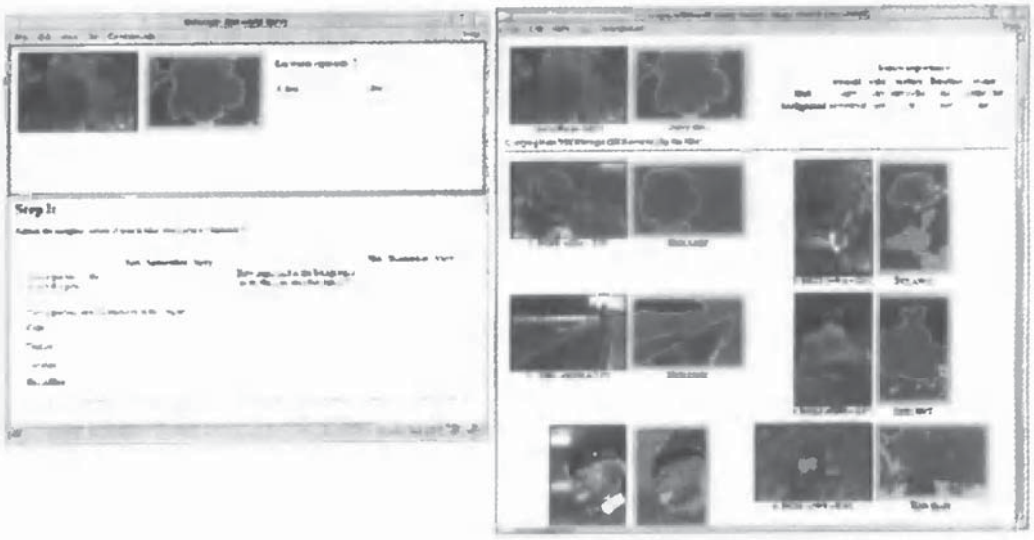


图 25.6 用 Blobworld 系统查询玫瑰花图像。图像数据库的用户一般想要找包含特定对象的图像，而不是带有某种统计数据的图像。Blobworld 系统用物体或物体的部分相对应的区域集合来表示每张图像，以便于用户查询。图像自动分割成区域，对每个区域的颜色纹理和形状进行编码。用户选择感兴趣的区域来构造一个查询(如左图)。Blobworld 系统查找图像并根据相似度评分，产生如右图所示的结果。在所示检索到的每一幅图像中，匹配区域用高亮度示出；可以看出，系统这种显示内部表示的方法使得查询结果更易理解，并有利于用户建立和完善查询。实验表明，对特殊对象的查询，如老虎和印度豹等，用 Blobworld 系统的精度，比其他基于整体颜色和纹理表示的系统要高得多

25.3.2 模板匹配

某些物体在广泛视角范围与拍摄条件下有富有特点的外观，因此可以用模板匹配的方式实现检索。模板匹配是一种物体识别策略，通过示例模板匹配图像来查找物体。我们曾在第 22 章讨论过这些细节，并着重于人脸的查找(为方便起见，我们将给一个概要)。一个模板匹配的自然应用方式是构造与特定语义类别相对应的所有图像模板(图 25.13，以及 Chang, Chen 和 Sundaram,

1998a)。这些模板可以离线构造,并允许用户使用已有的模板来简化查询,而不用去构造一个查询。

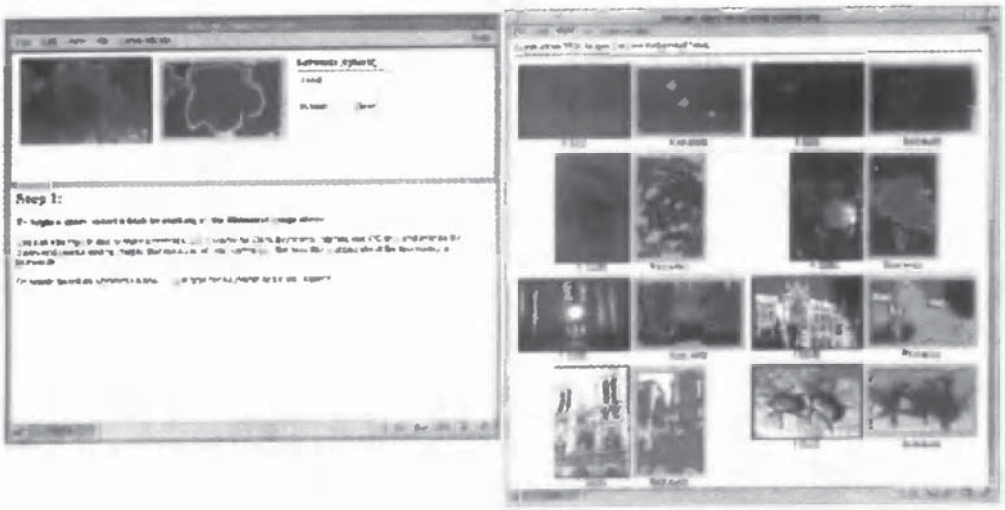


图 25.7　Blobworld 系统也允许简单文本查询。左图中,用"玫瑰"这个词来查询图像。右图所示的是靠前的几个检索结果图像。注意玫瑰色的篦鹭和玫瑰金龟子也出现在结果图像集合里,词语表达会产生歧义

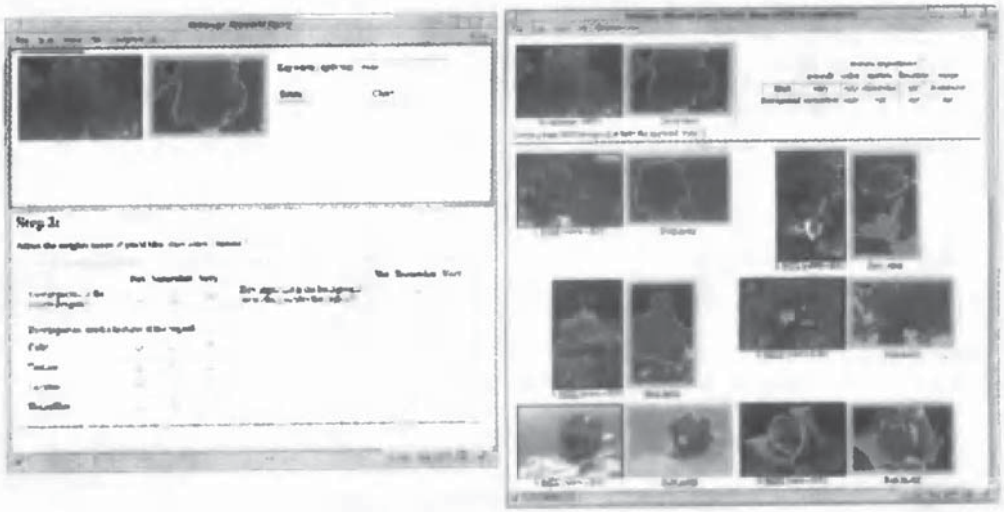


图 25.8　图像和文字互补产生出很好的结果。在图 25.6 中,有些包含红色块状的图像并不是玫瑰;相似地,在图25.7中,"玫瑰"这个单词也是不确切的。在本图左图中,查询要求包含红色块状图像,并附有"玫瑰"这个词。右图显示了查询结果的前一部分,其中大多数是花的图片

查找人脸是模板匹配一个极好的例子。人脸的正视图是十分相似的,在低分辨率尤其是如此。其主要特征是:嘴部呈现暗条,眼部呈现暗块,前额、鼻子和嘴呈现亮块。这表明,可以利用这个特征模式查找人脸图像,而不必分辨人的身份。典型的人脸查询系统取一块尺寸固定的窗口图像,修剪这些窗口使其呈椭圆形,更正光照,再用训练好的分类器识别窗口里是不是人脸(Rowley, Baluja 和 Kanade, 1998a; Rowley 等, 1998b; Sung 和 Poggio, 1998)。由于是在不同分辨率下取窗口图像,这种处理可以找到大的和小的人脸(低分辨率图下,窗口找到是大尺寸人脸;高分辨率图下找到的是小尺寸人脸)。当人脸向一边倾斜会引起人脸特征模式改变,则必须估计并校正这个倾斜;这可用机器学习机制完成(Rowley, Baluja 和 Kanade, 1998c)。获知人脸在哪里是很有用的,因为许多查询涉及图像或视频中出现的人。

25.3.3 形状和对应

如果物体外观变化，模板匹配就变得困难，因为这需要用许多模板。有一个成功的查找步行者的匹配系统，它之所以能运行，是因为在低分辨率下观察步行者，他们的手臂一般在身体两侧(Oren 等，1997)。然而，构造一个查找人的模板匹配系统是很困难的，因为衣服和人的姿态变化太大。对付这个困难的通用策略是查找小一些的模板——可能对应于整体的某部分——然后查找它们之间合理的配置。

这个技术的一种形式采用检测感兴趣点的策略，就是检测灰度值或其导数具有特殊值的那些点，如角点等。正如 Schmid 和 Mohr(1997*a*，*b*)所指出的，这些点的空间排列在许多情况下是十分不同的。例如(如图 25.9 所示)，感兴趣点之间的关系可产生十分有效的匹配，甚至对三维物体也是这样。可以将这个匹配过程扩展到生成图像到图像的变形，用它来对准图像。对准的图像可以产生进一步支持匹配的证据，并用于在特定点比较两幅图像。例如，将这个方法用于航空图像，用它来比较一天内不同时间的交通流量。

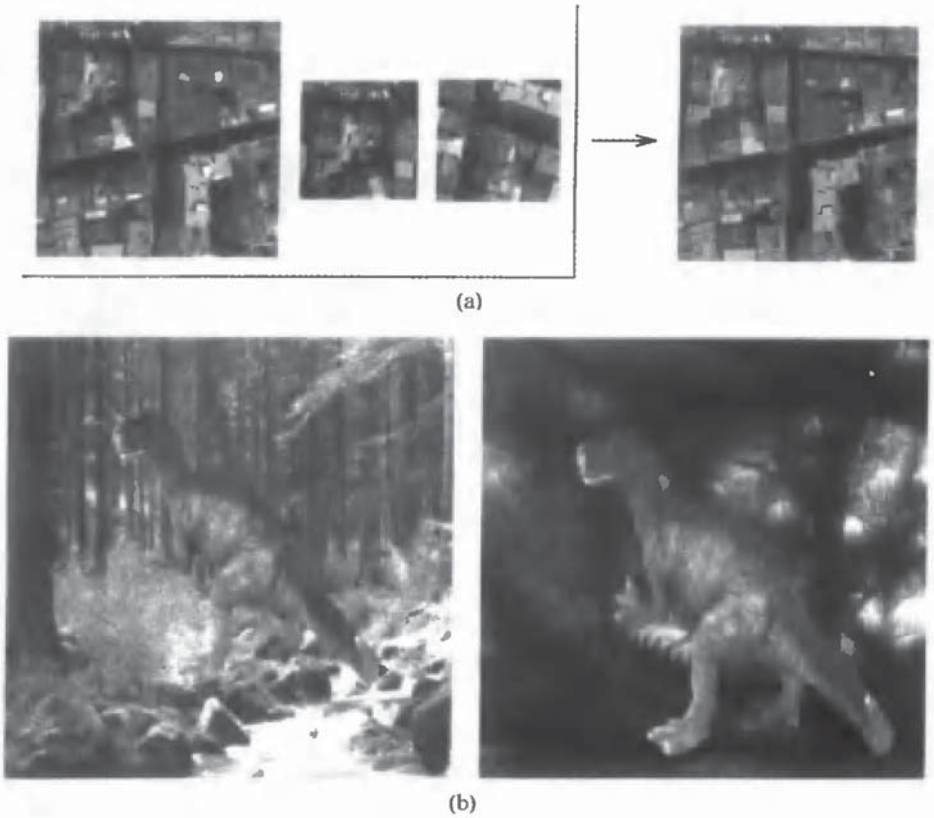


图 25.9 识别结果：(a)图像匹配解释航空图像：用左图的任何一幅图像都可以正确检索到右边的图像；(b)三维物体识别：尽管背景十分复杂，左右两图中的玩具恐龙都能被正确识别

这种形式的对应推理延伸到在更加抽象的层次上将图像部件与物体部件相匹配。例如，人和动物可以看做是对应于肢体分段的圆柱体组装体。一个自然的查找表示使用阶段组装图像部件，使之对应适当的肢体分段或其他部件。

Forsyth 和 Fleck(1999)用这种表示来识别穿衣服很少或不穿衣服人的图片。这是一个有趣的例子：第一，这比找穿衣服的人容易，因为皮肤在图像中很少有颜色和纹理变化，而衣服的

样式就会有很多变化;第二,很多人感兴趣的是,根据图像是否包含没穿衣服的人,来排除或查找图像。这个程序已经在很多不同的图像集上测试过;在一个包含 565 张穿很少衣服人的图像和 4289 张内容广泛的其他图像的集合里,这个程序标出 241 张测试图像和 182 张其他图像。第二个例子所使用的表示中的组合结构,即测试的顺序是手工建立的,但测试的内容是从数据中学习得到。这个程序识别马的图片,其细节可在 Forsyth 和 Fleck 的文章中找到(1997)。测试用了 100 张包含马的图像和 1086 张各种内容的其他图像;对于一个典型的配置,这个程序标出了 11 张马的图像和 4 张其他图像(见图 25.10)。

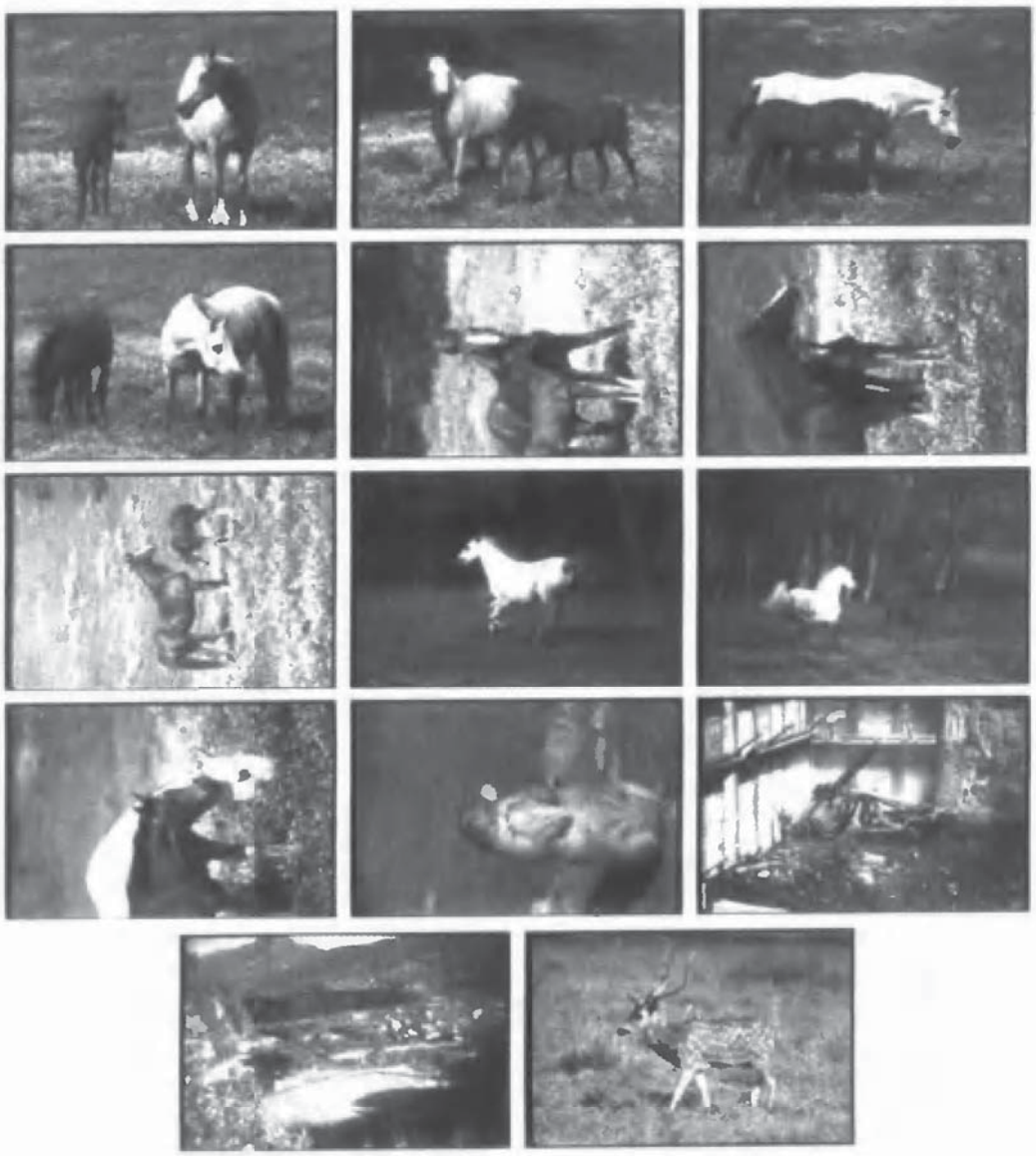


图 25.10　从 100 张包含马的图像和 1086 张不同内容其他图像的集合中,用肢体草图表示找到的马的图像。注意这个方法对外表相对不敏感,但是它容易被褐色、马状区域欺骗

25.3.4　聚类和组织集合

搜索特定片段(或具有启示关系的片段组)的另一个策略是聚类图像。典型地,我们希望生成由相似图像组成的聚类;它不仅应该包含视觉外观上的相似性,也应该包含语义上的相似性。

一个值得注意的事实是，虽然分别用文字和图像的描述会有歧义性，但它们组合起来就不是了。作家用文字描述图像时，倾向于避免提及视觉上明显的那些事物（如花的颜色等），而去描述那些使用计算机视觉很难推测的属性（如花的种类等）。这说明应该尝试同时使用图像信息和与相关文字一起来生成图像聚类。使用文字会涉及一些不属于我们职责范围的问题。第一，单词是含糊的（如"bank"一词是指涉及钱的"银行"，还是指野生百里香生长的"浅滩"）。第二，单词通常来源于句子，或甚至来源于段落，需要确定忽略哪些词，使用哪些词。第三，需要知道怎样在模型中处理单词和图片元素的联合。

Barnard 和同事们对这个有趣的题目进行了挖掘（Barnard 和 Forsyth，2001；Barnard，Duygulu 和 Forsyth，2001）。他们展示了在两个集合上做的工作：一个是 Corel 公司发行的一个图像集合，每一张图附带一个关键字集合，另一个是旧金山精美艺术博物馆里的艺术图片集，里面每一张图带有一段由志愿者写的文字注解（当写注解时，他们并不知道计算机的处理过程）。这段自由的文字被精简成名词、动词、形容词和副词的集合（Brill，1992）。用投票策略为单词选出一个词义，并基于邻近单词有相似词义的假设，将每个单词的可能词义与其邻近单词的可能词义相比较。最终，基于图像和单词聚类训练一个产生式模型。

这个产生式模型，在本质上是个混合的模型。这个混合模型的每一个组件表示出现某种带有单词语和用块状物表示的图像区域的概率，这个概率在给定组件下条件独立，并随不同部件变化，而图像区域的特征综合表示了颜色、纹理和粗略的形状。这个模型可以用 EM 算法拟合图像数据；一旦它拟合了图像数据，一幅图像就可以属于最可能产生它的那个混合物的部件的聚类。没有理由认为这是最好的进行方式，但它确实产生了相当好的聚类。图 25.11 显示了一些聚类，它们是从 corel 数据集获取的，其中比较了基于图像数据的聚类，基于文字数据的聚类以及二者结合的聚类。

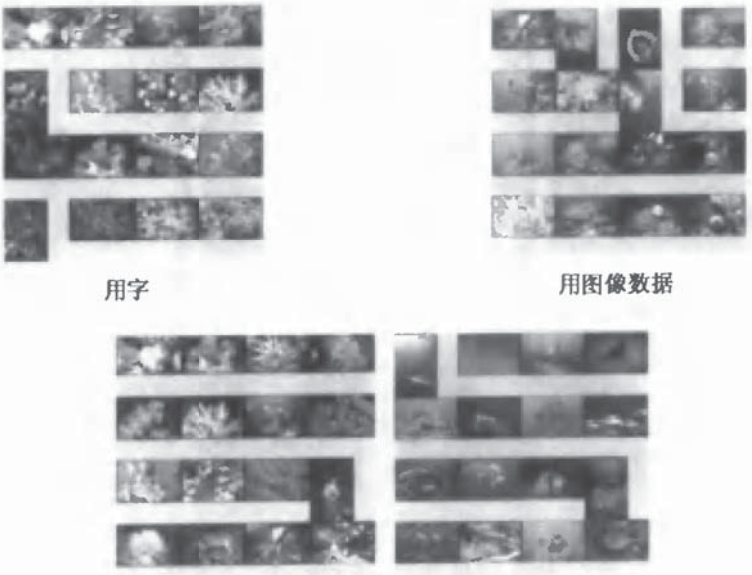


图 25.11 右上图显示了一个聚类中的一些图像集，其中的图像是仅用词语进行聚类的。应该注意到总的主题是有关海洋的；但是图像看上去十分不同；有一些潜水者，他们在蓝色海洋、珊瑚等背景上。左上图，是另一个聚类的图像集，这里的图像是仅用图像分割特征进行聚类得到的。这些图像看上去很相似，但是语义上却不一致；一些是珊瑚的图片，另一些是花的图片。左图：是同时用文字和图像分割特征得到的两个聚类。一般说来，这些图片共有一个主题，并且看上去也相似，同时具有两个所希望的属性

好的聚类可以有若干种使用方式。Barnard 等人的演示还是初步的，显然他们应该生成一种浏览机制。在这方面我们做了进一步的工作。通过训练产生式模型，我们构造了图像特征和词语的联合概率分布。这是说可以用单词搜索图像（Barnard 等人把这叫做“自动说明”）以及用图像搜索词语（“自动注释”，但你要注意这很像物体识别）。用合适的聚类，二者都能获得惊人的成功。图 25.12 显示了“自动说明”的结果。

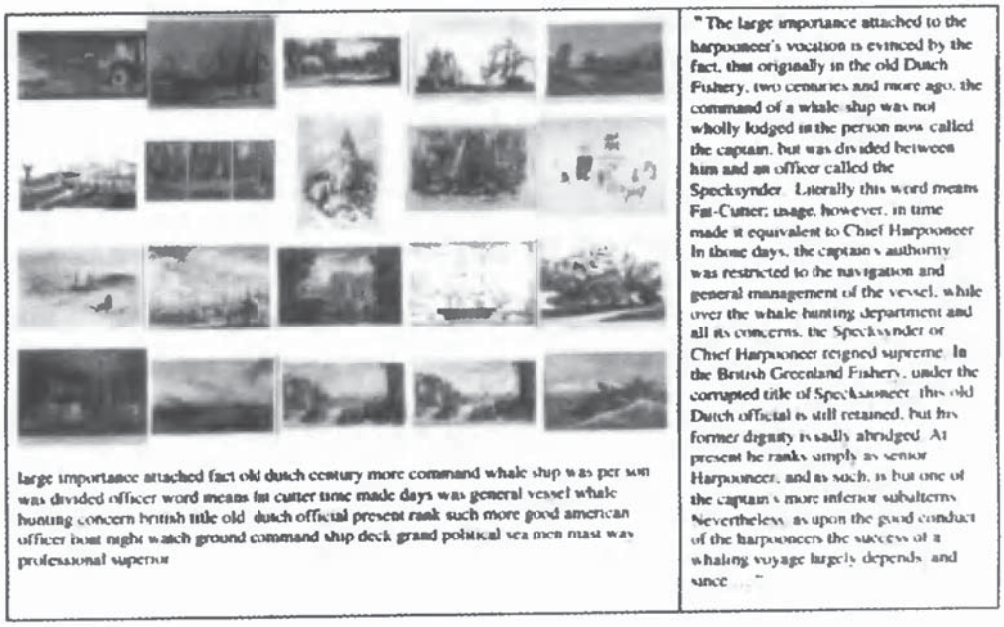


图 25.12　右图，来源于 Moby Dick 的一段文字。从这段文字中提取出名词、动词、形容词和副词，并且用投票过程为这些单词条目消除歧义。所得到的文字作为一个查询送入 Barnard 等人的联合概率模型中去，检索返回与这组词语有高度联合概率的图像。左图，是这个检索返回的图像。这个查询看来是成功的（除了有一些别的东西，这里有一张水手在划艇里捕鲸的图片）

25.4　视频

尽管视频与静态图像相比是更丰富的信息源，但是面临的问题在很大程度上相同。典型的情况是，将视频划分成镜头，即包含相似内容的简短的图像序列，然后将已经介绍的技术应用于镜头。我们曾在 14.3.2 节简短地叙述过镜头边界检测的方法。

在一段视频中，单个像素的运动经常被称为“光流”，测量方式是在下一帧里找与这一帧里相对应的像素（对应性用颜色、灰度和纹理的相似性度量）。原理上讲，每个像素都有个光流向量形成的运动场。实际上，很难在缺少特征的像素点上测量光流，因为它们可能和邻近任何像素很好地匹配。例如，考虑一个鸡蛋绕它轴心旋转时产生的光流，就很难找到鸡蛋边界内部的像素运动的信息，因为每个像素看上去都相似。

运动场可以是很复杂的；但是如果场景内没有运动物体，就能分出与摄像机拍摄相对应的运动场。例如，横扫的镜头会导致很强的横向运动，伸缩镜头会导致发射状的运动场。通常，

将所测量的运动场与参数模型族进行比较可以实现这种分类(例如,见 Sawhney 和 Ayer, 1996; Smith 和 Kanade, 1997)。

不做分割很难查询复杂的运动序列,因为许多运动都可能和这个查询不相关。例如,在足球比赛中,许多运动员的运动是不重要的。在 Chang 的系统 VideoQ 中,运动序列被分割成移动块然后用特定块的运动和颜色来查询(见图 25.13, Chang, Chen, Meng, Sundaram 和 Zhong, 1997*a*; Chang, Chen, Meng, Sundaram 和 Zhong, 1998*b*)。

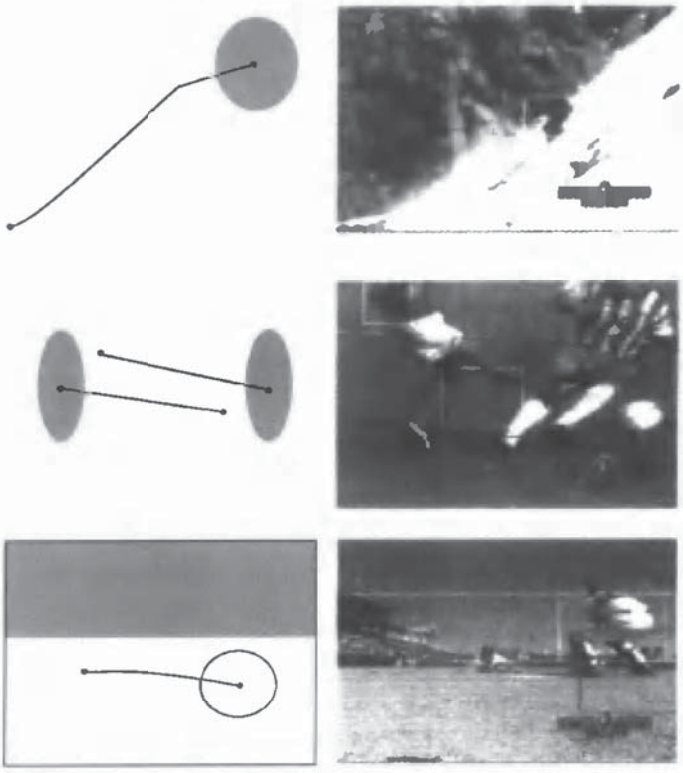


图 25.13 可以将视频表示成运动的图像块;图像序列也就可以用指定的图像块属性和期望的运动属性来查询。左边一栏是在Chang的VideoQ系统的用户界面中,用图像块各种类型的运动进行查询。右边一栏是从两个序列里返回的一些帧

Informedia 项目曾研究了编制视频序列带细节的快读描述。在这个情况下,将一段视频分割成若干镜头,并用摄像机的运动、人脸的呈现、文字呈现、配音关键词和音频等级来为镜头做注解(见图 25.14)。这些信息产生一个紧凑的表示——所谓的"快读"——它给出了视频序列中主要内容(Smith 和 Kanade, 1997; Wactlar, Kanade, Smith 和 Stevens, 1996; Smith 和 Christel, 1995; Smith 和 Hauptmann, 1995)。

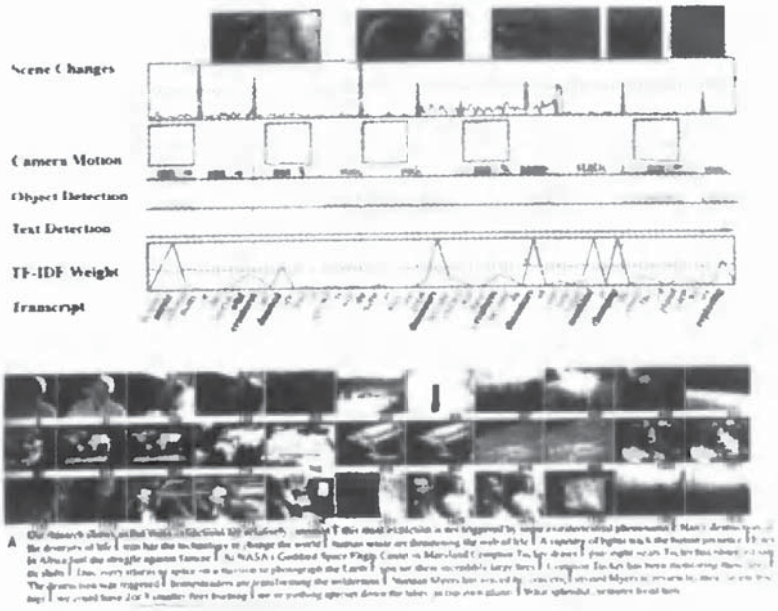


图 25.14　图的上边,描绘一段要建立一个快读的视频序列。将这个视频分割成场景,并检测出摄像机运动连同重要的物体(人脸和文字)。条形图指示出正确的结果。将条目频度与反向文档频度相比较,得到词语相关度的评价(本质上,在一个文档中喜爱的条目是常见的,但并不是在所有文档中都普遍)。所有这些数据用来提取图像帧和单词精简集,如下边的图所示。它表达了原有的视频序列,但要简短多了。使用文字、运动、人脸和镜头信息,就可能比用图像帧子采样方法获取的快读要强得多,因为很长而里面信息量又少的序列能被更好地压缩

25.5　注释

这个领域是计算机视觉中最有趣的应用,它是一个相当旧的领域(重要的早期论文有 Chang 和 Yang, 1983; Kato, Shimogaki, Mizutori 和 Fujimura, 1988; Kofakis, Orphanoudakis, 1991),但近一时期它很流行,可能是因为现在可以解决 30 年前处理不了的问题(对于这个问题,如果有更大容量的磁盘、更快的计算机等,就会有很大不同)。我们试图在叙述时给出一些参考文献,但我们知道,人们对这个领域的巨大兴趣意味着那些文献会很快过时。

颜色、纹理和图像布局与检索内容有很强相关时,对基于内容检索图像有用的工具是很多的。因为颜色、纹理和布局至多是图像内容的粗略向导,这就很容易产生令人迷惑的检索结果。目前还没有一个使这种影响最小化的检索界面设计理论。最广泛采用的策略是允许移动式浏览。

要在要求严格的应用里获得成功需要能从图像中获取概念的技术。特别地,要建立一个好用的工具,要涉及物体识别中深入和很难理解的问题。正如我们看到的,物体识别需要将图像分割成连续的片断,并对这些片断(如在人脸查找中)和片断间关系进行推理。这种有些含糊的识别观点可以用来产生分解的图像表示,实现不受背景影响的物体查找。况且,一些特殊情况的物体识别可以被明确地处理。现在还不知道怎么建立一个能搜索各种类别物体的系统,也不知道怎么为这样的系统建立能够表示丰富问题的用户界面。

第 26 章　应用:基于图像的绘制

计算机游戏、体育节目播送、电视广告和故事影片等娱乐产业每天与上亿人接触,而合成场景图像与实拍镜头相混合,在计算机游戏等是很常见的。建立这些图像就是所谓“基于图像的绘制”——在这里定义为从预先录制下的画面中合成场景的新视图——它需要从图像中恢复定量的形状信息(尽管并不必是三维的)。这一章介绍一些有代表性的基于图像的绘制方法,它们可以划分成这样几类:(a)先从图像序列中恢复三维场景,再用经典的计算机图形学工具绘制(这些方法经常与立体视觉、运动分析相关是很自然的);(b)不去恢复摄像机或场景参数,而是已经在第 10 章提到过的分析摄影地形测量法的做法,构造所观察场景所有可能图像集的一个显式表示,再用少量结点(tie points)的图像位置来指定场景的新视图,并将所有其他点迁移到新图像中;(c)用二维光线集(或更准确地说,沿这些光线的辐射亮度值)为图像建模,并用一个四维光线集,也就是光线场,为一个场景的所有图像建模(见图 26.1)。

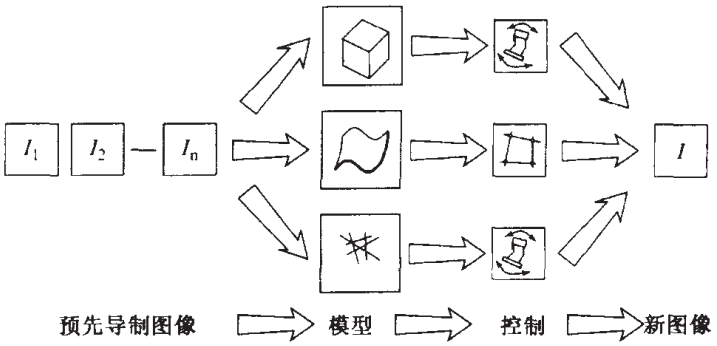


图 26.1　基于图像绘制的方法。从上到下:从图像序列中建立三维模型,基于迁移的图像合成,光线场。从左到右,是基于图像绘制的处理过程:从样本图像中建立场景模型(可以不是三维的),再用于场景的新视图绘制。绘制引擎可以受操纵杆控制(或等效地指定摄像机参数),或在基于迁移的技术中,用设置少量结点的图像位置来控制

26.1　从图像序列构造三维模型

这一节讲述从图像序列建立并绘制三维物体模型的问题。当然,用第 21 章介绍的距离扫描仪获取深度图来建立这样一个模型是可能的,但是在这里我们关心这样的情况:输入图像是一个刚性或运动场景的数字化照片或电影片断。

26.1.1　从对齐图像建立场景模型

体重构　假设一个物体已经被一组照片详细描绘,并在同一个全局坐标系中对齐。由图像边界惟一地恢复物体形状是不可能的,因为如第 19 章所述,物体表面的凹陷部分永远不会显示在图像边界上。尽管如此,我们能够从足够大的图像集中建立这个表面的一个合理估计。在这里,图像边界对刚体形状产生两个全局约束:(a)每张图像有对应视锥(viewing cone),刚体位于所有视锥的交集内;(b)这些锥体与刚体表面相切[这里有另外的局部限制;如第 19 章所

示，边界的凸起(凹陷)部分是表面凸起(凹陷)的投影]。Baumgart 在他 1974 年的博士论文里首次用到这些约束的第一种：刚体外形轮廓的近似多边形有对应的多面锥，用所有多面锥的交集可以构造各种物体的多面体模型。他的思想对许多从外形轮廓建立物体模型的方法有启发，包括这一节余下部分所介绍的技术(Sullivan 和 Ponce，1998)，这个技术也合并了与视锥相联系的相切限制。在 Baumgart 的系统里，首先由与几幅照片相联系的视锥交叉所构造观测物体的多面体近似(见图 26.2)。这个多面体的顶点用做光滑样条曲面的控制点，这个曲面经受变形直到它与视锥相切为止。这里我们关心这个曲面的构造与变形问题。

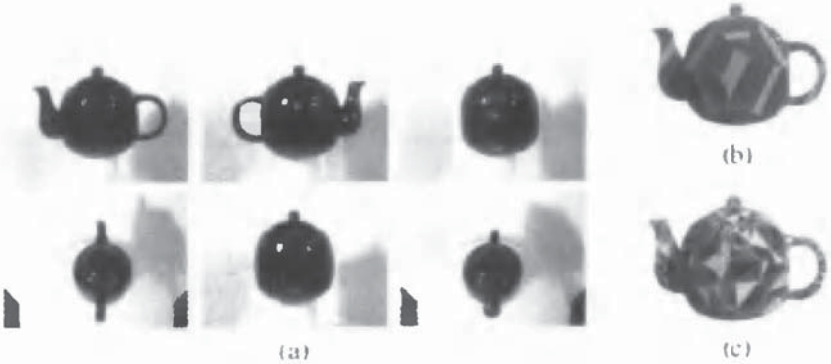


图 26.2　用(多面体)视锥相交构造物体模型：(a)一个茶壶的 6 张照片；(b)相应视锥的交；(c)三角化：将每个表面分解成三角形并简化所得到的网格

样条构造　一个样条曲线是满足某些光滑性条件的分段多项式参数曲线。例如，它可以是 C^k 的(也就是直到 k 阶连续导数可微)，k 通常取 1、2 或 G^k(也就是不需要处处可微，但要在 G^1 情况下有连续切线，在 G^2 情况下有连续曲率)。样条曲线通常用连接在一起的贝塞尔(Bézier)曲线段构造。一条 n 次贝塞尔曲线是多项式参数曲线 $P:[0,1]\to\mathbb{E}^3$，它定义为 $n+1$ 个控制点 $P_0,\cdots,P_n$ 的重心组合，

$$P(t)=\sum_{i=0}^{n} b_i^{(n)}(t)P_i$$

其中，权重 $b_i^{(n)}(t)\stackrel{\text{def}}{=}\binom{n}{i}t^i(1-t)^{n-i}$ 称为 n 次伯恩斯坦(Bernstein)多项式①。贝塞尔曲线内插它的第一个和最后一个控制点，但不对其他的控制点内插[见图 26.3(a)]。如习题中所示，端点的切线方向，是沿控制点构成控制多边形的首末线段方向。

贝塞尔曲线和样条曲线的定义可以自然地扩展到曲面：一个 n 次贝塞尔面片是参数曲面 $P:[0,1]\times[0,1]\to\mathbb{E}^3$ 它定义为三元控制点 P_{ijk} 阵列的重心组合

$$P(u,v)=\sum_{i+j+k=n} b_{ijk}^{(n)}(u,v,1-u-v)P_{ijk}$$

其中，齐次多项式 $b_{ijk}^{(n)}(u,v,w)\stackrel{\text{def}}{=}\dfrac{n!}{i!\ j!\ k!}u^iv^jw^k$ 是 n 次三变量伯恩斯坦多项式。这节的余下部分里，使用 4 次贝塞尔面片($n=4$)，每个由 15 个控制点定义[图 26.3(b)]。它们的边界是

① 这的确是一个重心组合(如第 12 章所定义)，因为伯恩斯坦(Bernstein)多项式相加总等于 1。特别地，贝塞尔曲线是个仿射构造，这是一个理想的属性，因为这些曲线仅由它们的控制点定义，独立于任何外部坐标系的选择。

4 次贝塞尔曲线 $P(u,0)$、$P(0,v)$和 $P(u,1-u)$。按定义，一个 G^1 的三角化样条是一些贝塞尔面片组成的网络，这些面片有着沿它们共同边界方向的公共切平面。对于 G^1 连续性，一个必要(但不充分)的条件是围绕共同顶点的控制点要共面。我们先构造这些控制点，然后再放置其他控制点来确保所生成的样条确实是 G^1 连续的。正如 Loop(1994)讨论的，一个共面点集 $Q_1,\cdots,Q_p$ 可以用另外 p 个一般位置的点 $C_1,\cdots,C_p$ 的重心组合来构造(在我们的情况下，它们是与输入三角化面片的一个顶点 V 邻接的 p 个三角形 T_j 的重心，图 26.4，左上)，

$$Q_i=\sum_{j=1}^{p}\frac{1}{p}\left\{1+\cos\frac{\pi}{p}\cos\left([2(j-i)-1]\frac{\pi}{p}\right)\right\}C_j$$

这个构造将点 Q_i 置于通过点集 C_i 质心 O 的平面上。平移这个平面使 O 与 V 重合产生一个新的点集 A_i，这个点集位于通过 V 的平面上(图 26.4，上边中间)。

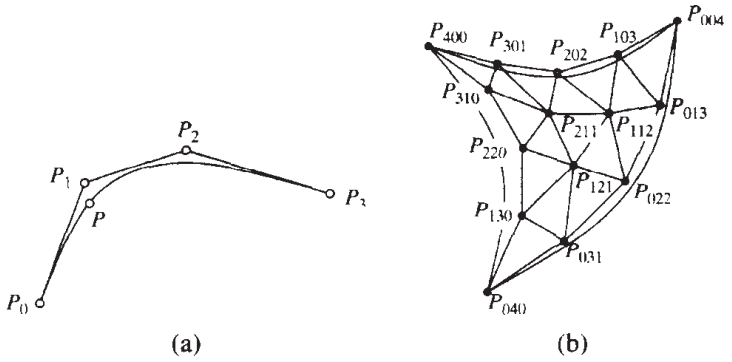


图 26.3 贝塞尔曲线和曲面：(a)三次贝塞尔曲线和它的控制多边形；(b)4 次贝塞尔三角面片和它的控制网格。张量积贝塞尔面片可以用四元组控制点来定义(Farin，1993)。三角面片更适合自由形态的封闭表面建模

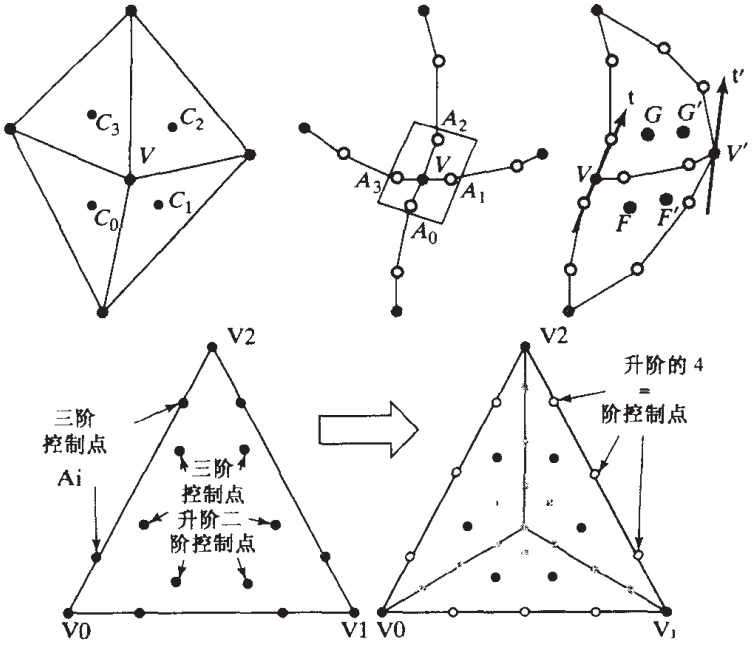


图 26.4 在三角化的多面体网格上构造三角样条。上半部分，从左到右：三次边界控制点，围绕网格顶点的边界曲线，和从相切约束构造的内部控制点。下半部分：将面片分割成三块，满足 G^1 连续：白点是由控制曲线升阶获得的控制点，灰点是其他控制点，由确保 G^1 连续性计算得到

因为三次贝塞尔曲线用 4 个点定义，我们认为两个邻近的顶点 V、V' 和与相应边相联系的点 A_i 和 A'_i，构成一条三次曲线的控制顶点。这产生了一个三次曲线段集合，它们对控制网格顶点进行插值，并形成三角面片的边界。一旦这些曲线被构造出来，就可以选择边界两侧的控制点，使之满足面片内 G^1 连续的约束。在这个构造中，交叉边界切线场在边界曲线两个端点上线性地插值切线。在端点 V 上，与包含点 A_i 的曲线相交叉的切线 $\boldsymbol{t}$，被取为与线段 A_{i-1}，A_{i+1} 平行。切线 $\boldsymbol{t}'$ 由相似的构造方式获得。内部控制点 F、F'、G 和 G'（图 26.4，右上）由与这个几何条件相联系的线性方程组来构造（Chiyokura, 1983）。然而，在一个 4 次面片上没有足够的自由度来允许同时对所有三个边界设定内部点。这样，每个面片就要分割成三块，并可使用 Shirman 和 Sequin（1987）的方法来确保新面片间的连续性：用同样形状的 4 次贝塞尔曲线来替换原来的边界，以提高边界曲线的阶次（见习题）。然后，三个 4 次三角面片可以从边界来构造，如图 26.4 下半部分所示。其结果是，在网格每个面片上获得的三个 4 次面片在穿越所有边界时实现 G^1 连续。

样条变形　我们给出一种从三角化网格构造 G^1 连续的三角样条表面近似的方法，如图 26.2(b)所示。现在介绍如何变形这个样条来确保它与输入照片的视锥相切。样条表面 S 的形状由它的控制顶点 $V_1,\cdots,V_p$ 的位置确定。我们用 $V_{jk}(k=1,2,3)$ 来代表点 $V_j(j=1,\cdots,p)$ 的欧氏坐标，并使用这 $3p$ 个系数做形状参数。在给定光线集 $R_1,\cdots,R_q$ 条件下，对 S 的参数 V_{jk} 求下式能量函数的最小化

$$\frac{1}{q}\sum_{i=1}^{q} d^2(R_i,S)+\lambda\sum_{i=1}^{r}\iint\left[|P_{uu}|^2+2|P_{uv}|^2+|P_{vv}|^2\right]du\,dv$$

其中，$d(R,S)$ 代表光线 R 到表面 S 的距离，积分是个薄壳（thin-plate）样条能量项，用于强制稀疏数据区的光滑性，λ 是用于平衡距离项和平滑项的权重常数。这个积分中的变量 u 和 v 是面片参数，求和是在构成样条曲面的 r 个面片上进行的。光线与面片的距离可以用牛顿（Newton）方法计算。对于那些与曲面不相交的光线，我们定义 $d(R,S)=\min\{|\overrightarrow{QP}|,Q\in R,P\in S\}$，并最小化 $|\overrightarrow{QP}|^2$ 来计算距离。对于那些与曲面相交的光线，我们用 Brunie, Lavallée 和 Szeliski（1992）的方法，沿相应遮挡边界点上的表面法线方向，测量到曲面上最远点的距离。在这两种情况下，用曲面 S 上的取样来初始化牛顿迭代。在曲面拟合时，用简单的梯度下降法来使样条变形，实现光线与表面间的均方距离最小化。虽然每个距离都要用数值方法计算，对表面参数 V_{jk} 的导数很容易用达到该距离的表面和光线点满足的约束进行微分来计算。

图 26.5 所示三个物体模型是用这一节介绍的方法构造的。这个技术不需要建立任何输入图片之间的对应，但它目前限于静态场景。相反，下面要介绍的方法是基于多摄像机立体观测，因而需要对应，但它不仅可以处理静态场景，也能处理动态场景。

虚拟现实　Kanade 和他的同事（1997）提出了“虚拟现实”的概念，它是一个新的视觉媒介，可以用来处理和绘制在可控环境中捕捉到的真实场景的录制图像和合成图像。在卡内基梅隆大学，这个概念的第一个物理实现由装备 10 个同步视频摄像机的测地拱顶构成。在写本书时，一个最新的实现是一个“三维房间”，里面有 $20\times20\times9$ 立方英尺的空间，用 49 个彩色摄像机拍摄，摄像机连接到一组计算机上，并在同一个世界坐标系内对齐，所有摄像机实现同步视频流实时数字化。由多摄像机立体视觉融合得到的稠密深度图可以生成三维场景模型（见

Okutami 和 Kanade, 1993,第 11 章)。这样的深度图可由一个摄像机和少量邻近摄像机(3 到 6 个)获得。每个深度图像再转换成表面网格,这个网格可以用诸如纹理映射的经典计算机图形学技术绘制。如图 26.6 所示,从单一深度图构造的场景图像可能有明显的缝隙。这些缝隙可以通过将对应相邻摄像机的网格绘制到同一图像来填补。

图 26.5 茶壶、怪兽和恐龙的形状和纹理映射模型。茶壶用 6 张对齐的照片构造;怪兽和恐龙模型分别用9幅图像构造

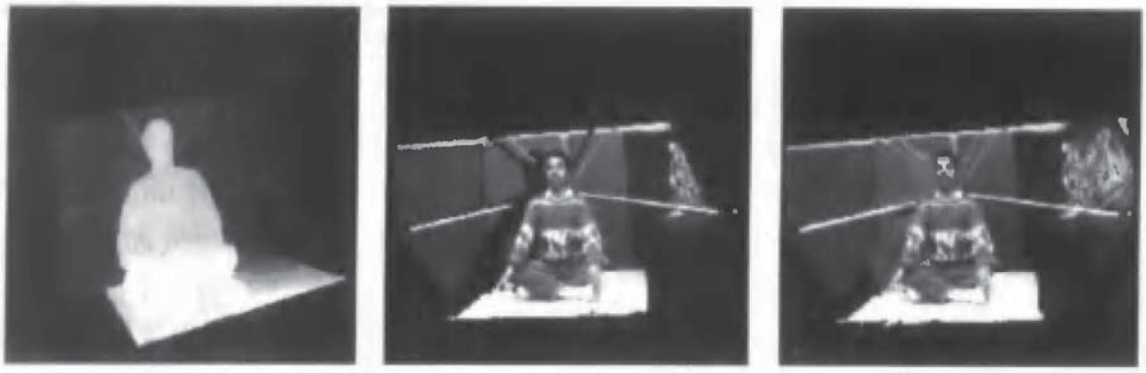


图 26.6 多摄像机立体视觉。从左到右:与一组摄像机关联的深度图;从不同视点观察到的相应网格的纹理映射图(注意深度图里暗区域表示深度不连续);从两个邻近摄像机构造的纹理映射图(注意缝隙已经被填补了)

也可以用将不同摄像机的表面网格直接合并成单一的表面模型的方法。这个任务是挑战性的,因为:(a)在几个摄像机都能观察到的区域,对同一个表面面片可能存在多个相互冲突的测量;(b)某个场景面片可能任何一个摄像机都观察不到。这两个问题可以用深度图像融合的体测量技术解决,它是由 Curless 和 Levoy(1996)提出的,在第 21 章介绍过。一旦总体表面模型被构造出来,它就可以像前面一样进行纹理映射。合成动画也可通过对输入序列中任何两幅视图插值获得。首先,表面模型用于建立两个视图的对应,第一张视图里通过任何一点

的光线与这个网格相交的交点重投影到第二张视图上，产生了所需的匹配[①]。一旦这些对应已知，新的视图可由线性插值匹配点的位置和颜色来构造。如 Saito 等人(1999)所述，这个简单算法仅提供真正透视图像的一个近似，还要添加更多逻辑来处理第一张图中可见，但第二张图中不可见的点。尽管如此，它可以用于生成遮挡图形不断变化的动态场景的真实感动画，如图 26.7 所示。

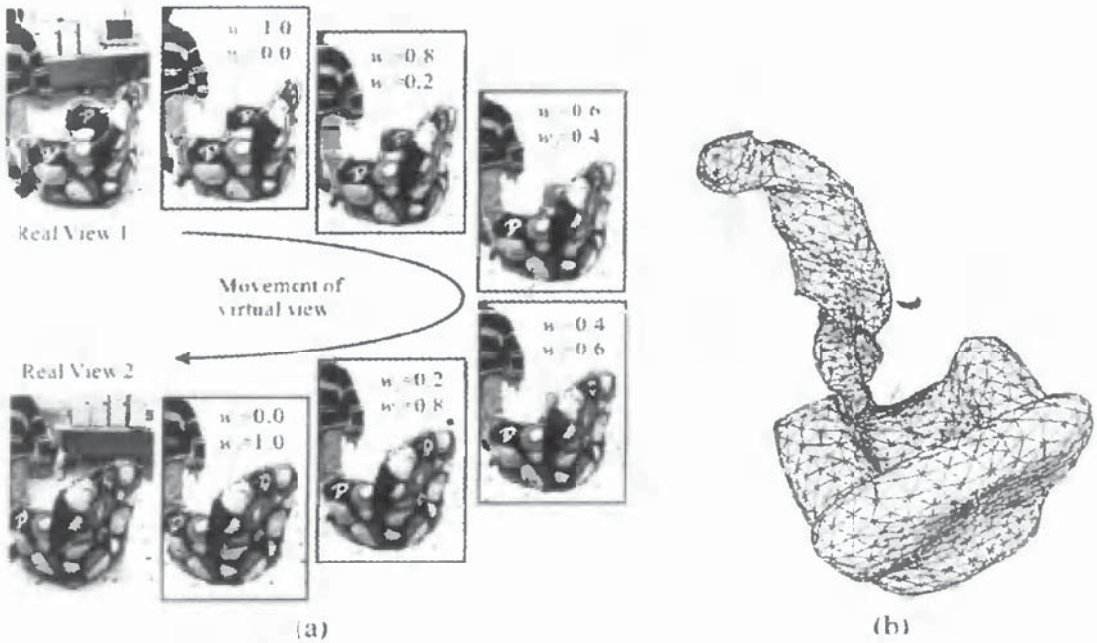


图 26.7　虚拟现实：(a)一个合成图像序列；注意第一个视图里两个椭圆区域里的遮挡已被正确处理了；(b)相应的网格模型

26.1.2　从没有对齐图像建立场景模型

这一节再次介绍从图像集合获取和绘制三维物体模型的问题，但是这次观察场景摄像机的位置事先不知道，需要用第 12 章和第 13 章介绍的方法从图像信息中恢复。显然，这一节里介绍的技术是针对计算机图形学的应用开发的。

Façade 系统　从数字化照片里建模与绘制建筑物场景的 Façade 系统是由 Debevec，Taylor 和 Malik(1996)在加州大学伯克利分校开发的。这个系统利用许多建筑物具有相对简单的整体几何学，从而简化了场景结构和摄像机运动估计，并且它使用了简单而强大的基于模型立体视觉思想，给粗略的建筑物外形增加细节。图 26.8 给出一个例子。

Faade 模型限于长方体、棱柱和旋转体这些参数化基元构造的层次结构，这些基元用少量系数(如，长方体的长、宽、高)定义，相互之间用刚体变换联系起来。定义模型的任何参数可以是常量，也可以是变量，并可在各种未知量间指定限制条件(如，两个块可以被限制为有同样的高度)。模型层次由图形用户界面交互式定义，而 Façade 系统的主要计算任务在于用图像信

① 经典的窄基线方法(如相关)在这种情况下是无效的，因为两个视图可能相距很远。一个相似的方法被用于这一章稍后介绍的 Facade 系统中，当观察表面粗略形状已知时，它建立了相距较远图像之间的对应。

息将定值分配给未知模型参数。整个系统划分成三个主要组成部分:第一部分是"分析摄影地形测量法模块",将结构和运动估计转化为一个非线性优化问题,它使照片上手工选取线段与参数模型相应部分投影的差异最小化(细节见习题)。如 Debevec 等人(1996)所示,这个过程涉及相对较少的变量,仅有拍摄建筑物的摄像机位置和方向、建筑模型的参数,并且,当某些模型边缘的方向相对于世界坐标系固定时,这些参数的初始估计可以用最小二乘法很容易地找到。

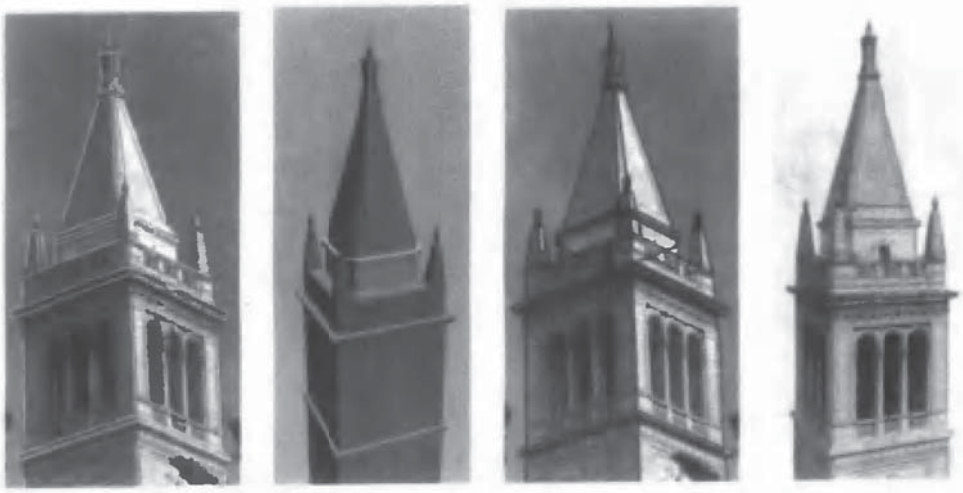


图 26.8 伯克利钟楼的 Façade模型。从左到右:带有所选取边缘的钟楼照片;从分析摄影地形测量法恢复出的三维模型;这个模型到照片上的重投影;这个模型的纹理映射视图

Façade 的第二个主要组成部分是"与视角有关的纹理映射模块",它根据用户视点,将不同照片映射到它的几何模型上,从而绘制出建筑物场景。从概念上看,是用幻灯机替换摄像机,将原始图像投影到这个模型上。当然,每个摄像机仅能看见建筑物的一个部分,而绘制完整的模型要用几张照片。通常,建筑的各个部分能够被几个不同的摄像机观察到,所以绘制器不仅要能挑选与虚拟视图合成有关的图片,而且要能对它们进行适当地合并。Façade 系统里采用的解决方法是:对新图像中每个像素用有关的输入照片预测值的加权平均赋值,其中权重与相应光线在输入视图和虚拟视图中的夹角成反比。

Façade 系统的最后一个主要组成部分是"基于模型的立体视觉模块",它在分析摄影地形测量法模块构造相对粗略的场景描述上,用立体视觉技术增加精细的几何细节。这个设置中使用立体视觉的主要困难是摄像机相距较远,这样会直接影响基于相关的匹配技术的使用。Façade 系统里采用的解决方法是利用先验的形状信息将立体图像映射到相同的参考帧中(图 26.9,上图)。具体地,给定主(key)和侧视(offset)图像,在从主摄像机视点绘制侧视图像之前,可以将侧视图像投影到场景模型上,产生与主图像相似的变过形的侧视图像(图 26.9,下图)。这样就可以使用相关来建立这两幅图像的对应,从而建立主图像与侧视的对应。一旦两幅图像的匹配建立好了,立体重构就简化为通常的三角化过程了。

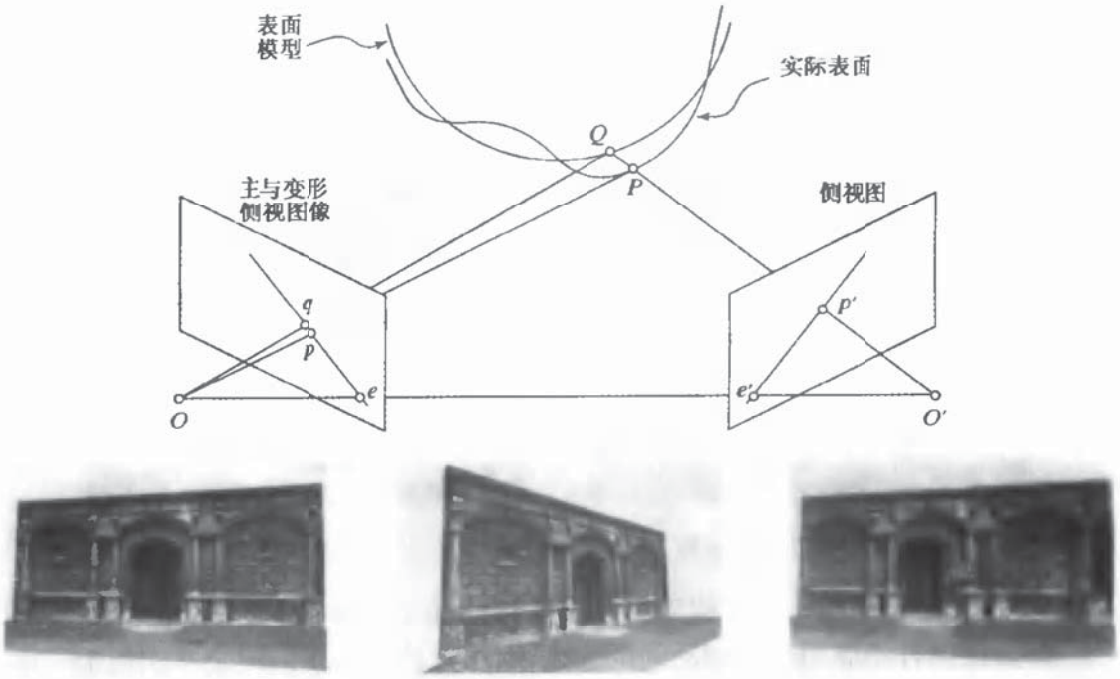


图 26.9　基于模型的立体视觉。上图:变形的侧视图像的合成。侧视图像中点 p' 被映射到表面模型的 Q 点,然后重投影到变形侧视图像的 q 点。双摄像机观察到的实际表面点 P,投影到主图像的 p 点。注意点 q 应该在外极线 ep 上,这便于常规立体视觉情况下的匹配搜索。同时也要注意,沿外极线方向 p 和 q 之间的差异度量了模型和实际表面的差异。来源于Debevec等人(1996,图15)。下图从左到右:一个主图像,一个侧视图像和相应的变形的侧视图像

26.2　基于迁移的基于图像绘制方法

这一节研究一个完全不同的基于图像绘制方法。在这个框架下,不去做显式的三维场景重构,而是直接从一个(可能很小的)视图集合建立新的图像,视图集合里的点对应已经用特征跟踪或传统立体视觉匹配方法建立好了。这个方法与第 10 章提过的分析摄影地形测量法的经典迁移问题相关。给定一些结点在参考图像集中和在新图像中的位置,并给定参考图像上结点的图像位置,预测新图像中那个点的位置。

基于迁移的基于图像绘制方法由 Laveau 和 Faugeras(1994)在投影空间中引入,他们建议先估计参考视图间的成对外极线几何,再将这些场景点重投影到一幅虚拟图像上,场景点由新光心在两幅参考图片上的投影(也就是外极点)和新视图里 4 个结点的位置指定。按定义,外极几何限制参考图像里点的可能的重投影。在新视图里,场景点的投影在与该点和两个参考图片相联系的外极线的交点上。一旦特征点被重投影了,用光线跟踪和纹理映射就合成了真实感的图像。然而,像 Laveau 和 Faugeras 指出的那样,因为缺乏与标定摄像机相关联的欧几里得约束,用任意平行投影变换绘制的图像会偏离了正确的图像,除非考虑附加的场景限制。这一节余下部分将研究能避免这个困难的基于迁移方法的两个不同仿射技术。这两个技术构造了刚体场景所有图像集的参数化表示:在第一种情况下(26.2.1 节),仿射图像空间的仿射结构用于绘制增强现实系统(an augmented reality system)里的合成物体。因为这种情况下的结点总是

几何上合理的图像特征(如,标定多边形的角点;见图 26.10),合成的图像自动地生成欧几里得图像。在第二种情况里(26.2.2 节),与标定摄像机关联的度量约束在图像空间参数化时是明确加以考虑的,再次保证正确的欧几里得图像的合成。

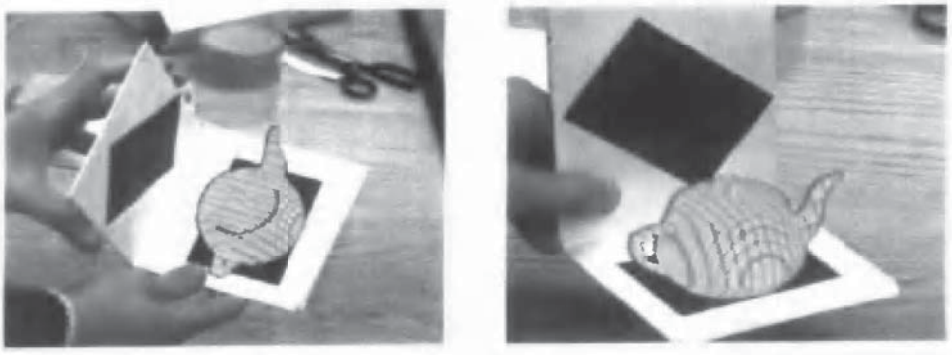


图 26.10 增强现实实验。(仿射)世界坐标系由黑色多边形角点定义

让我们再次注意已经在引言里提到的基于迁移的基于图像绘制方法的特性:因为并没有构造三维模型,操纵杆不能用于控制动画的合成,而结点位置要由用户交互式指定。这在增强现实情景中不是一个问题,但对于虚拟实现应用来说,这是否是一个可行的用户界面还有待于观察。

26.2.1 仿射视图合成

这里讲述在不显式设置三维欧氏坐标系的情况下,从场景的旧图像合成新图像的问题。回忆在第 12 章里,如果将某个世界坐标系中场景点 P 的坐标向量用 $\boldsymbol{P}=(x,y,z)^T$ 表示,并用 $\boldsymbol{p}=(u,v)^T$ 表示 P 在图像平面内投影 p 的坐标向量,方程(2.19)的仿射摄像机模型可以写成

$$\boldsymbol{p}=\mathcal{A}\boldsymbol{P}+\boldsymbol{b},\quad 其中,\quad \mathcal{A}=\begin{pmatrix}\boldsymbol{a}_1^T\\ \boldsymbol{a}_2^T\end{pmatrix} \tag{26.1}$$

$\boldsymbol{b}$ 是物体坐标系原点的图像内投影位置,$\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 是$\mathbb{R}^3$ 中的向量。

让我们考虑 4 个(不共面的)场景点,P_0,P_1,P_2 和 P_3。不失一般性地,可以选择这些点作为一个仿射参考帧,使它们的坐标向量是

$$\boldsymbol{P}_0=\begin{pmatrix}0\\0\\0\end{pmatrix},\quad \boldsymbol{P}_1=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad \boldsymbol{P}_2=\begin{pmatrix}0\\1\\0\end{pmatrix},\quad \boldsymbol{P}_3=\begin{pmatrix}0\\0\\1\end{pmatrix}$$

通常,点 $P_i(i=1,2,3)$到 P_0 的距离并不是一个单位长度,并且向量$\overrightarrow{P_0P_i}$和$\overrightarrow{P_0P_j}$也互相不正交$(i\neq j)$。这是没有关系的,因为我们操作在一个仿射环境下。由于 $\boldsymbol{P}_1$,$\boldsymbol{P}_2$ 和 $\boldsymbol{P}_3$ 构成的 3×3 的矩阵是单位阵,方程(26.1)可以改写成

$$\boldsymbol{p}=\mathcal{A}\boldsymbol{P}+\boldsymbol{b}=\begin{pmatrix}\boldsymbol{a}_1^T\\ \boldsymbol{a}_2^T\end{pmatrix}[\boldsymbol{P}_1|\boldsymbol{P}_2|\boldsymbol{P}_3]\begin{pmatrix}x\\y\\z\end{pmatrix}+\boldsymbol{b}$$

最后,因为我们将点 P_0 取为世界坐标系的原点,有 $\boldsymbol{b}=\boldsymbol{p}_0$,并且可以得到

$$\boldsymbol{p}=(1-x-y-z)\boldsymbol{p}_0+x\boldsymbol{p}_1+y\boldsymbol{p}_2+z\boldsymbol{p}_3 \tag{26.2}$$

这个结果与第 12 章讨论过的仿射图像的仿射结构有关。在基于图像绘制的情景下,由式(26.2),x,y,z 可以用最小二乘法从点 P_0,P_1,P_2,P_3 和 P 的 $m\geqslant 2$ 个图像计算出来。一旦这些值已知,指定点 p_0,p_1,p_2,p_3 并用式(26.2)来计算所有其他的点,就可以合成出新图像(Kutulakos 和 Vallino, 1998)。另外,因为场景的仿射表示的确是三维的,场景的相对深度可以计算出来,并可用于消除图形学流水线(pipeline)z 缓存部分的被遮挡表面。应该注意到,为点 p_0,p_1,p_2,p_3 指定任意位置一般会产生仿射变形图像。在增强现实应用中这不是一个问题,在那里绘制的和真实的物体共存于图像内。在这个情况下,锚点 p_0,p_1,p_2,p_3 可以在真实图像点里选择,以确保正确的欧里几得位置。图 26.10 显示了合成物体覆盖在真实图像上的例子。

当较长的图像序列可用时,这种方法的一个变通是以统一方式考虑所有场景点,它可由如下所述做法得到。假设观察一个固定点集 $P_0,\cdots,P_{n-1}$,其坐标向量是 $\boldsymbol{P}_i(i=0,\cdots,n-1)$,并让 $\boldsymbol{p}_i$ 代表相应图像点的坐标向量,对所有场景点应用式(26.1),并将它们写在一起

$$\begin{pmatrix}\boldsymbol{p}_0\\ \cdots\\ \boldsymbol{p}_{n-1}\end{pmatrix}=\begin{pmatrix}\boldsymbol{P}_0^T & \boldsymbol{0}^T & 1 & 0\\ \boldsymbol{0}^T & \boldsymbol{P}_0^T & 0 & 1\\ \cdots & \cdots & \cdots & \cdots\\ \boldsymbol{P}_{n-1}^T & \boldsymbol{0}^T & 1 & 0\\ \boldsymbol{0}^T & \boldsymbol{P}_{n-1}^T & 0 & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \boldsymbol{b}\end{pmatrix}$$

换句话说,n 个固定点的所有仿射图像集是一个嵌在$\mathbb{R}^{2n}$里的八维向量空间 V,并由向量 $\boldsymbol{a}_1,\boldsymbol{a}_2$ 和 $\boldsymbol{b}$ 参数化①。给定 $n\geqslant 4$ 个点的 $m\geqslant 8$ 个视图,这个向量空间的基可通过对下面的 $2n\times m$ 矩阵进行奇异值分解确定

$$\begin{pmatrix}\boldsymbol{p}_0^{(1)} & \cdots & \boldsymbol{p}_0^{(m)}\\ \cdots & \cdots & \cdots\\ \boldsymbol{p}_{n-1}^{(1)} & \cdots & \boldsymbol{p}_{n-1}^{(m)}\end{pmatrix}$$

这里 $\boldsymbol{p}_i^{(j)}$ 表示第 j 帧里第 i 个图像点的位置②。一旦 V 的基构造出来,给 $\boldsymbol{a}_1,\boldsymbol{a}_2$ 和 $\boldsymbol{b}$ 赋任意值即可构造出新的图像。为了便于交互式图像合成,一个更加直观的成像几何控制方法可以这样获得:和前面一样指定 4 个像点的位置,解出 $\boldsymbol{a}_1,\boldsymbol{a}_2$,和 $\boldsymbol{b}$ 的相应值,计算其余点的图像位置。

26.2.2　欧里几得视图的合成

如前所述,上节所介绍方法的一个缺点是为点 p_0,p_1,p_2,p_3 指定任意位置,这通常会产生仿射变形图像。一开始就考虑与标定摄像机相关联的欧里几得约束可以避免这个问题。我们在第 12 章里看到,一个弱透视投影摄像机是一个满足下面两个二次方程约束的仿射摄像机

$$\boldsymbol{a}_1\cdot\boldsymbol{a}_2=0\quad 和\quad |\boldsymbol{a}_1|^2=|\boldsymbol{a}_2|^2$$

前一节说过,一个固定场景的仿射图像是一个八维向量空间 V。现在如果我们限定注意力到弱透视摄像机,图像集合就成了这两个多项式约束所定义的六维子空间。相似的约束应用于类透视和

① 这与第 12 章得到的结论不矛盾,那个结论是说一个任意点集的 m 个固定视图集是$\mathbb{R}^{2m}$的三维仿射子空间

② 至少需要 8 张图像可能看上去是不必要的,因为一个场景的仿射结构可以从两幅图恢复,如第 12 章所述。确实,如习题所示,V 的一个基可由至少 4 个点的两幅图像构造

透视投影,它们也在每种情况下定义了一个六维空间(也就是,用多项式方程定义的子空间)。

假设我们观察图像上不共线的三个点 P_0, P_1, P_2。可以选择(不失一般性地)一个欧里几得坐标系,使得这三个点在这个坐标系中的坐标是

$$\boldsymbol{P}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{P}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{P}_2 = \begin{pmatrix} p \\ q \\ 0 \end{pmatrix}$$

这里,p 和 q 是非零的,但(事先)是未知数。让我们像前面一样用 p_i 表示点 P_i($i=0,1,2$)的投影。因为 P_0 是世界坐标系的原点,有 $\boldsymbol{b}=\boldsymbol{p}_0$。我们也选 p_0 作为图像坐标系的原点(这个等于将所有图像点做了一个已知的偏移),这样式(26.1)简化成

$$\boldsymbol{p} = \mathcal{A}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{a}_1^T \boldsymbol{P} \\ \boldsymbol{a}_2^T \boldsymbol{P} \end{pmatrix} \tag{26.3}$$

现在将式(26.3)应用于 P_1, P_2 和 P,得到

$$\boldsymbol{u} \stackrel{\text{def}}{=} \begin{pmatrix} u_1 \\ u_2 \\ u \end{pmatrix} = \mathcal{P}\boldsymbol{a}_1 \quad 和 \quad \boldsymbol{v} \stackrel{\text{def}}{=} \begin{pmatrix} v_1 \\ v_2 \\ v \end{pmatrix} = \mathcal{P}\boldsymbol{a}_2 \tag{26.4}$$

其中,

$$\mathcal{P} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{P}_1^T \\ \boldsymbol{P}_2^T \\ \boldsymbol{P}^T \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ p & q & 0 \\ x & y & z \end{pmatrix}$$

这意味着

$$\boldsymbol{a}_1 = \mathcal{Q}\boldsymbol{u} \quad , \quad \boldsymbol{a}_2 = \mathcal{Q}\boldsymbol{v} \tag{26.5}$$

其中

$$\mathcal{Q} \stackrel{\text{def}}{=} \mathcal{P}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & \mu & 0 \\ \alpha/z & \beta/z & 1/z \end{pmatrix} \quad , \quad \begin{cases} \lambda = -p/q \\ \mu = 1/q \\ \alpha = -(x+\lambda y) \\ \beta = -\mu y \end{cases}$$

用式(26.5)并设$\mathcal{R} \stackrel{\text{def}}{=} z^2 \mathcal{Q}^T \mathcal{Q}$,式(12.10)的弱透视投影约束可以改写成

$$\begin{cases} \boldsymbol{u}^T \mathcal{R} \boldsymbol{u} - \boldsymbol{v}^T \mathcal{R} \boldsymbol{v} = 0 \\ \boldsymbol{u}^T \mathcal{R} \boldsymbol{v} = 0 \end{cases} \tag{26.6}$$

其中,

$$\mathcal{R} = \begin{pmatrix} \xi_1 & \xi_2 & \alpha \\ \xi_2 & \xi_3 & \beta \\ \alpha & \beta & 1 \end{pmatrix} \quad 和 \quad \begin{cases} \xi_1 = (1+\lambda^2)z^2 + \alpha^2 \\ \xi_2 = \lambda\mu z^2 + \alpha\beta \\ \xi_3 = \mu^2 z^2 + \beta^2 \end{cases}$$

式(26.6)定义了系数ξ_i($i=1,2,3$)、α 和β 上的一对线性约束。这些可以改写为

$$\begin{pmatrix} \boldsymbol{d}_1^T \\ \boldsymbol{d}_2^T \end{pmatrix} \boldsymbol{\xi} = 0 \tag{26.7}$$

其中,

$$\boldsymbol{d}_1 \stackrel{\text{def}}{=} \begin{pmatrix} u_1^2 - v_1^2 \\ 2(u_1u_2 - v_1v_2) \\ u_2^2 - v_2^2 \\ 2(u_1u - v_1v) \\ 2(u_2u - v_2v) \\ u^2 - v^2 \end{pmatrix}, \quad \boldsymbol{d}_2 \stackrel{\text{def}}{=} \begin{pmatrix} u_1v_1 \\ u_1v_2 + u_2v_1 \\ u_2v_2 \\ u_1v + uv_1 \\ u_2v + uv_2 \\ uv \end{pmatrix}, \quad \boldsymbol{\xi} \stackrel{\text{def}}{=} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \alpha \\ \beta \\ 1 \end{pmatrix}$$

当 4 个点 P_0, P_1, P_2 和 P 刚性地连接在一起时,这 5 个结构系数 $\xi_1, \xi_2, \xi_3, \alpha$ 和 β 就固定了。对于 n 个点形成的刚体场景,选择其中三个点作为参考三角形并将式(26.7)应用于其余的点,产生 $2n$ 个未知数的 $2n-6$ 个二次方程组,它定义了这个场景里所有弱透视投影图像的一个参数化表示。这就是 Genc 和 Ponce(2001)的参数化图像簇(PIV)。

需要再次指出,式(26.7)的弱透视约束是 5 个结构系数的线性方程。因此给定一个图像集以及点对应,这些系数可用最小二乘法估计出来。一旦向量 $\boldsymbol{\xi}$ 被估计出来,任意像点位置就可以被指派到三个参考点。对于每个特征点,式(26.7)对未知数 u 和 v 产生两个二次方程约束。虽然由先验知识可知这个方程组有 4 个解,但只有两个解是合理解,如习题所示。实际上,给定 n 个点对应和 3 个结点的图像位置,其余 $n-3$ 个点的图像就有解析解,保留二歧义性。一旦所有特征点的位置被确定下来,就可以用三角化这些点和纹理映射这些三角形来绘制场景。有趣的是,尽管没有做显式的三维重构,消除被遮挡表面同样可以用传统的 z 缓存技术完成,做法是将相对深度值赋给三角化网格的顶点,这与第 12 章里提到的从运动中提取仿射结构定理所使用的方法紧密相关的。用 Π 表示一张输入图像的图像平面,用 Π′ 表示合成图像的图像平面。为了正确绘制合成图像,应该比较它们的深度(见图 26.11)。

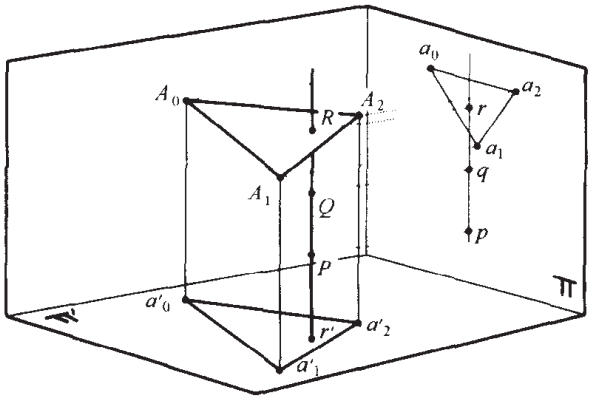


图 26.11　z 缓存技术

在图 26.11 上用 R 表示从 P 到 Q 的视线与参考点 A_0, A_1, A_2 所在平面的交点,用 p, q, r 表示 P, Q, R 在参考图像上的投影。假设当 P 和 Q 是输入图像中被跟踪的两点时,那么 p 和 q 的位置是知道的。r 的位置也容易计算,因为,Π 内参考点投影 a_0, a_1, a_2 形成仿射基下的 r 坐标,与点 A_0, A_1, A_2 形成仿射基下的 R 坐标是相同的,也因此与 Π′ 内参考点投影 a'_0, a'_1, a'_2 形成仿射基下的 r' 坐标相同。P 和 Q 相对于平面 Π 的深度比就是 $\overline{pr}/\overline{qr}$。注意确定哪个点是真正可见的,需要为 p, q, r 所在直线定向,这是与点 r' 相关联的外极线。应该为所有的外极线选择一致的朝向(这是容易的,因为它们互相平行)。注意这不需要显式计算外极线几何:给定第一个点 p',可以定出直线 pr 的方向,再对所有其他点对应使用相同的方向。所选的方向对后续帧也要一致,但这不成问题,因为外极线方向从一帧到另一帧是渐变的,因此可以简单地

选取新的方向,使得它与前一帧成锐角即可。用这个方法合成图片的例子如图 26.12 所示。

图 26.12 人脸的两幅图像,用参数化图像合成

26.3 光线场

这一节讨论基于图像绘制的另一个不同方法,它与前一节所讨论技术的相似之处仅仅在于,它不需要构造场景任何隐式或显式的三维模型。例如,用一个全景摄像机,用光学方法记录下过一点覆盖整个半球的所有光线的亮度(见 Peri 和 Nayar, 1997;图 26.13,左)。可以使用一个针孔位于该点的虚拟摄像机,将原图图像光线映射到虚拟图像上,从而可以构造出所观察到的任何图像。用户可以随意摇动(pan)和倾斜这个虚拟摄像机,并交互式地探索他的虚拟环境。相似的效果也可以这样获得:将便携式摄像机拍到的邻近图像粘合成一个拼接图,或将一个摄像机绕它光心摇动(也可能倾斜)拍到的图片组合成一个圆柱拼接图(见 Chen, 1995; 图 26.13,右)。

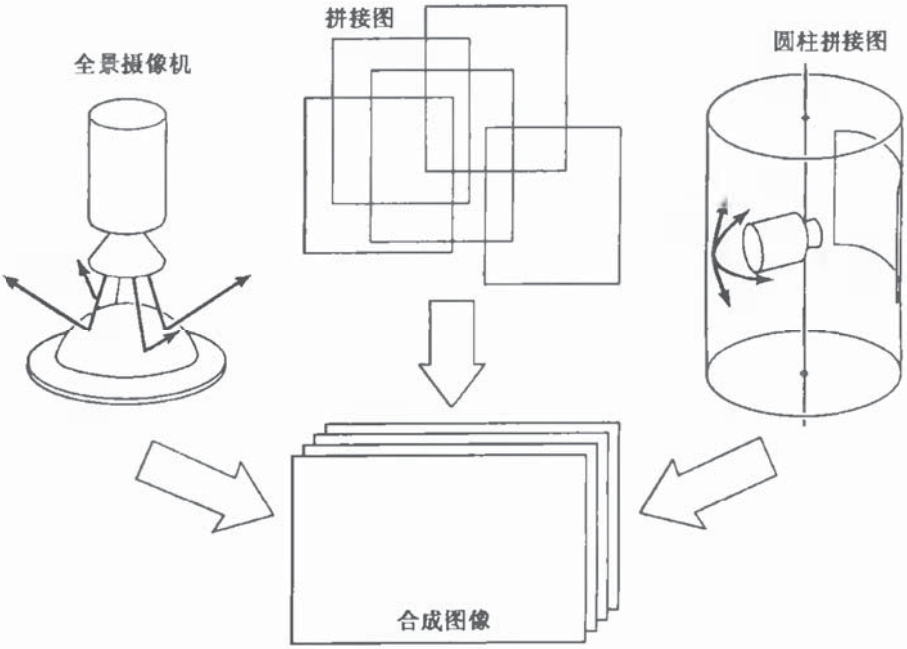


图 26.13 构造一个从固定视点观察场景的合成视图

这些技术有个弱点,它将观察者运动限制于绕摄像机光心的纯旋转。参考全光场函数(Adelson 和 Bergen, 1991)可以设计出一个更强的方法,这个函数将空间中每一点与给定时刻沿过这点光线的辐射能量相联系(图 26.14,左)。光线场(Levoy 和 Hanrahan, 1996)是在没有障碍的真空中光传播的全光场函数的快照。这使得辐射对时间以及与感兴趣点沿相应光线上的

位置的依赖关系不加考虑(假设在真空中光传播中沿直线方向辐射是不变的),并获得用沿四维参数表示的光线集的辐射表示的全光场函数。这些光线可以用多种不同方式来参数化(如,用第 3 章介绍的 Plücker 坐标),但在基于图像绘制情景下一个方便的参数化表示是光线极板(light slab):每条光线用它和两个任意平面的交点指定(图 26.14,右)。

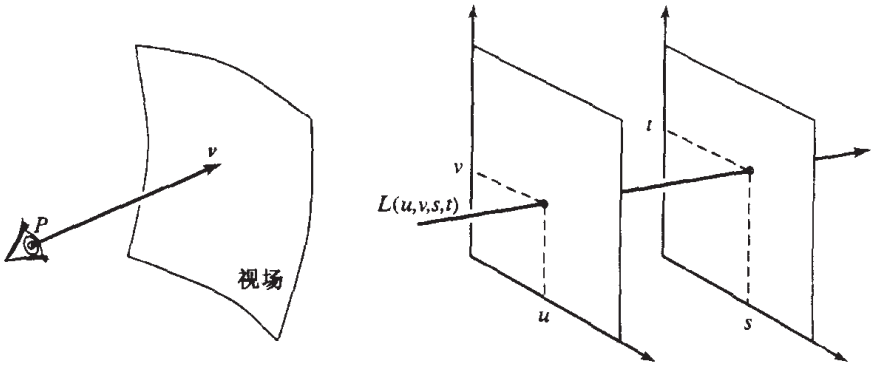


图 26.14　全光场(plenoptic)函数和光线场。左:全光场函数可以用观察者位置 P 和视线方向 v 参数表示。右:光线场可以用4个参数 u, v, s, t 定义的光线极板(light slab)参数化表示。实际上,需要用若干个光线极板为整个物体建模并得到整个球面覆盖

光线极板是一个两阶段基于图像绘制方法的基础。在学习阶段,一个场景的许多视图用于建立光线极板的离散形式,可以把它想像为一个四维查找表。在合成阶段,定义一个虚拟摄像机,从查找表里插值得到相应的视图。合成图像的质量依赖于参考图像的数量。虚拟视图离参考图像越相近,合成图像的质量就越好。注意构造光线场的光线极板模型不需要建立图像间的对应。还需指出的是,光线场技术与大多数依赖于纹理映射并假设观察表面是朗伯(Lambertian)表面的基于图像绘制方法不同,它可以用于绘制(在固定光照下)具有任意双向反射率分布函数的物体的图像。

实际上,使用大量图像并将像素坐标映射到光线极板坐标上,就能得到光线场的一个样本。图 26.15 说明了一般情况:(x, y)图像平面内任意像素与定义光线极板的(u, v)、(s, t)平面相应区域之间的映射,是一个平行透视变换。因而可以将基于硬件或软件的纹理映射方法用于将光线场传播到四维矩形网格中。在 Levoy 和 Hanrahan(1996)所述的实验里,光线场用简单的设备获取:将一个摄像机安装在平面支架上,支架可摇动和倾斜使摄像机能绕它的光心旋转,并总是指向感兴趣物体的中心。在这种情况下,将(u, v)平面取为摄像机光心被限制于其内的那种平面,就可以简化所有的计算。

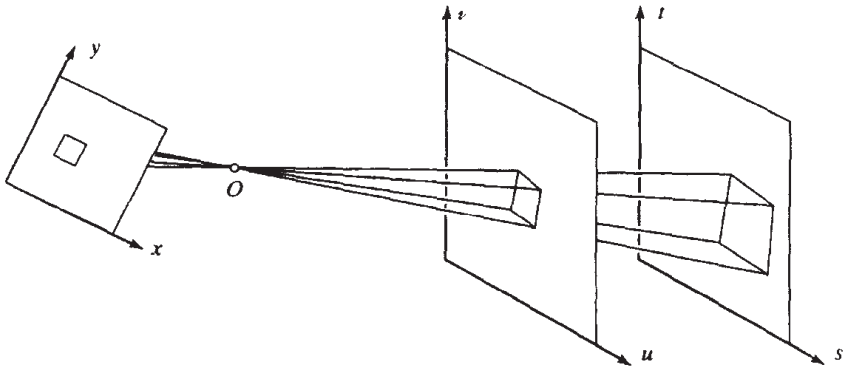


图 26.15　从图像获取一个光线极板,从光线极板合成新图像,可用(x, y)图像平面和定义光线极板的(u, v)和(s, t)平面间的透视变换建模

在绘制期间,(虚拟)图像平面与定义光线极板两个平面之间的投影映射可以再次用于有效地合成新图像。图 26.16 展示了用光线场方法生成的样例图像。上面三对图像是将各种物体的合成图片传播(populate)光线场生成的。最后一对图像是这样生成:用前面提到的平面支架来获取 2048 幅 256×256 大小的铁狮图像,聚集成 4 个光线极板,其中每个光线板是由 32×16 幅图像组成。

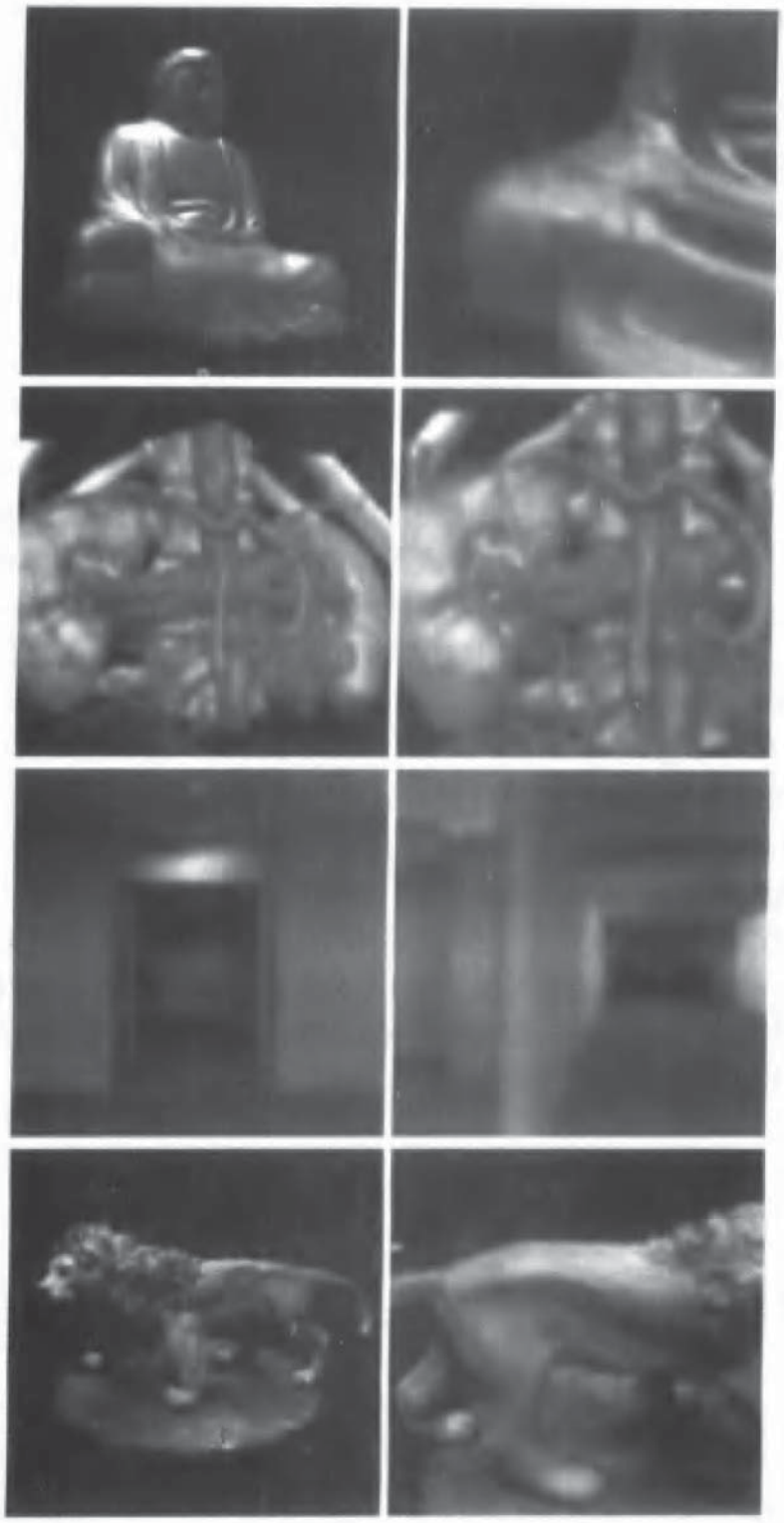


图 26.16 用光线场方法合成的图像

一个重要问题是光线极板表示的大小:铁狮的原始输入图像占用 402 MB 的磁盘空间。当然,这些图像里有许多冗余,与一个运动序列的连续帧情况相同。Levoy 和 Hanrahan(1996)提出一个简单而有效的用于图像压缩/解压缩的两级方法:光线极板先被分解成颜色值的四维超立方体(tile)。将这些超立方体用向量量化(Gersho 和 Gray, 1992)进行编码,这是一个有损压缩技术:表示原来超立方体 16 个角 RGB 值的 48 维向量被替换成相对小的再现向量集,被称为"码字",它在平方误差准则下最佳地近似了输入向量。光线极板因而被表示成为所有码字形成的码本的索引集合。在狮子这个例子中,码本相对地很小(0.8 MB),索引集大小是 16.8 MB。第二阶段压缩是将熵编码(Ziv 和 Lempel, 1977)的 gzip 实现应用于码本和索引。这个表示的最终大小仅为 3.4 MB,相应的压缩率是 118:1。在绘制期间,执行熵解码是在这个文件载入主存时进行。向量量化还原是在需要显示时执行,刷新率可以交互的确定。

26.4 注释

基于图像绘制是一个迅速发展的领域。在结束这一章之前,除前面已经讲过的方法之外,让我们提一些其他方法。Laurentini, 1995 提出用擦过刚体表面的所有视锥的交,形成刚体的视觉外壳(visual hull)的方法。一个刚体总是包含在它的视觉外壳之中,而视觉外壳又包含在这个刚体的凸包里。26.1 节介绍的从外形轮廓建物体模型的体表示法,是用有限照片集构造近似视觉外壳的方法。另有其他一些方法使用多面体或八叉树(Martin 和 Aggarwal, 1983; Connolly 和 Stenstrom, 1989; Srivastava 和 Ahuja, 1990)来表示视锥和视锥的交,其中包括一个商业化系统,Sphinx3D(Niem 和 Buschmann, 1994),它从图像自动构造物体模型。也可参考 Kutulakos 和 Seitz(1999),他们提出了一个相似的方法,叫空间雕刻,它利用亮度或色彩恒常性将空的体素迭代地删除掉。在摄像机运动已知或未知情况下,相切约束已经被用于从轮廓连续序列来构造表面的各种方法中(Arbogast 和 Mohr, 1991; Cipolla 和 Blake, 1992; Vaillant 和 Faugeras, 1992; Cipolla 等, 1995; Boyer 和 Berger, 1996; Cross 等, 1999; Joshi 等, 1999)。其他不同于 26.1 节所述的视图插值技术的方法,包括 Williams 和 Chen(1993)、Seitz 和 Dyer(1995, 1996)。除了 26.2 节所讨论的基于迁移的基于图像绘制方法之外,其他方法还包括,Havaldar 等(1996)、Avidan 和 Shashua(1997)。正如 26.3 节简短提到的,从固定视点交互式探索用户视觉环境的许多技术已经被开发出来,其中包括一个商业化的系统,由 Chen(1995)在苹果公司开发的 QuickTime VR,以及用于从特殊摄像机获取的全景图中重构针孔透视图像的一些算法(见 Peri 和 Nayar, 1997)。相似的效果也可以用控制较少的设备得到:将便携式摄像机拍得的邻近图像拼合在一起(见 Irani 等, 1996; Shum 和 Szeliski, 1998)。对于遥感地形(distant terrains)图像或绕光心旋转的摄像机所拍图像,可用平面齐次坐标对齐连续帧来构造拼接图像(mosaic)。在这个情况下,估计光流(也就是,在每个图像点的表观图像速度的向量场,这本书里大部分忽略了这个概念),对于精细图像配准和消除图像重影(deghosting)(Shum 和 Szeliski, 1998)也是很重要的。与 26.3 节所讨论的不同的光线场方法包括 McMillan 和 Bishop(1995)和 Gortler 等(1996)。一个极好的有关贝塞尔曲线曲面和样条曲线曲面的介绍可以在 Farin(1993)中找到。

习题

26.1 给定 $n+1$ 个点 $P_0,\cdots,P_n$,递归地定义参数曲线 $P_i^k(t)$:$P_i^0(t)=P_i$ 和

$$P_i^k(t)=(1-t)P_i^{k-1}(t)+tP_{i+1}^{k-1}(t) \qquad k=1,\cdots,n \quad , \quad i=0,\cdots,n-k$$

我们在这个习题中证明 $P_0^n(t)$是与 $n+1$ 个点 $P_0,\cdots,P_n$ 有关的 n 次贝塞尔曲线。贝塞尔曲线的这种构造方法叫做 de Casteljeau 算法。

(**a**) 证明伯恩斯坦多项式满足下面的递归

$$b_i^{(n)}(t)=(1-t)b_i^{(n-1)}(t)+tb_{i-1}^{(n-1)}(t)$$

这里 $b_0^{(0)}(t)=1$,并且按惯例当 $j<0$ 或 $j>n$ 时 $b_j^{(n)}(t)=0$。

(**b**) 使用归纳来证明

$$P_i^k(t)=\sum_{j=0}^{k}b_j^{(k)}(t)P_{i+j} \qquad k=0,\cdots,n \quad , \quad i=0,\cdots,n-k$$

26.2 考虑 $n+1$ 个控制点 $P_0,\cdots,P_n$ 定义的 n 次贝塞尔曲线。在这里我们提出一个问题:如何构造相同形状的 $n+1$ 次贝塞尔曲线的 $n+2$ 个控制点($Q_0,\cdots,Q_{n+1}$)。这个处理称为升阶(升高次数)。证明 $Q_0=P_0$ 和

$$Q_j=\frac{j}{n+1}P_{j-1}+\left(1-\frac{j}{n+1}\right)P_j \quad . \quad j=1,\cdots,n+1$$

提示:写出原曲线的重心点和升阶后曲线的重心点,它们相等所以可以写成等式,比较等式两边多项式系数即可。

26.3 证明 $n+1$ 个控制点 $P_0,\cdots,P_n$ 定义的贝塞尔曲线 $P(t)$的切线是

$$P'(t)=n\sum_{j=0}^{n-1}b_j^{(n-1)}(t)(P_{j+1}-P_j)$$

推论:贝塞尔曲线端点上的切线,应该沿它控制多边形的第一条线段和最后一条线段方向。

26.4 证明 26.1.1 节 Q_i 点的构造方法,其实是将这些点置于过点集 C_i 质心 O 的一个平面内。

26.5 Façade 系统的摄影测量模块。我们见到在第 3 章的习题里,用 Plücker 坐标向量 $\mathbf{\Delta}$ 表示的直线 δ 和它用齐次坐标 $\mathbf{\Delta}$ 表示的像 δ 之间的映射,可以写成 $\rho\boldsymbol{\delta}=\tilde{\mathcal{M}}\mathbf{\Delta}$。这里,$\mathbf{\Delta}$ 是模型参数的一个函数,$\tilde{\mathcal{M}}$依赖于相应的摄像机位置和方向。

(**a**) 假设直线 δ 已经与长度为 l 的图像边缘 e 匹配,将边缘 e 上点与直线 δ 间的均方距离乘以 l,可以得到预测数据与观察数据间差异的一个方便度量。将 $d(t)$定义为边缘点 $p=(1-t)p_0+tp_1$ 与直线 δ 间的带符号的距离,证明

$$E=\int_0^1 d^2(t)\,dt=\frac{1}{3}\left(d(0)^2+d(0)d(1)+d(1)^2\right)$$

这里 d_0 和 d_1 表示边缘 e 的两个端点与直线 δ 间的带符号距离。

(**b**) 如果 $\boldsymbol{p}_0$ 和 $\boldsymbol{p}_1$ 表示这些点的齐次坐标向量,证明

$$d_0 = \frac{1}{|[\tilde{\mathcal{M}}\boldsymbol{\Delta}]_2|}\boldsymbol{p}_0^T\tilde{\mathcal{M}}\boldsymbol{\Delta} \quad , \quad d_1 = \frac{1}{|[\tilde{\mathcal{M}}\boldsymbol{\Delta}]_2|}\boldsymbol{p}_1^T\tilde{\mathcal{M}}\boldsymbol{\Delta}$$

这里$[\boldsymbol{a}]_2$是个二维向量,由$\mathbb{R}^3$中向量$\boldsymbol{a}$的前两维坐标组成。

(**c**) 将摄像机和模型参数恢复,表示成非线性最小二乘问题。

26.6　当$n \geqslant 4$时,固定点集$P_0, \cdots, P_{n-1}$的所有仿射像构成了八维向量空间V。证明空间V中的一个基,当$n \geqslant 4$时,至少要由两个这些点的像构造出来。

提示:使用矩阵

$$\begin{pmatrix} u_0^{(1)} & v_0^{(1)} & \cdots & u_0^{(m)} & v_0^{(m)} \\ \cdots & \cdots & \cdots & & \\ u_{n-1}^{(1)} & v_{n-1}^{(1)} & \cdots & u_{n-1}^{(m)} & v_{n-1}^{(m)} \end{pmatrix}$$

这里$(u_i^{(j)}, v_i^{(j)})$是点P_i到图像编号j的投影坐标。

26.7　证明一个固定场景所有投影图像的集合是维度 11 的流形。

26.8　证明一个固定场景所有透视图像的集合(摄像机的内参数不变)6 维度的流形。

26.9　在这个习题中,我们证明式(26.7)只允许两个解。

(**a**) 证明式(26.6)可以改写为

$$\begin{cases} X^2 - Y^2 + e_1 - e_2 = 0 \\ 2XY + e = 0 \end{cases} \tag{26.8}$$

这里

$$\begin{cases} X = u + \alpha u_1 + \beta u_2 \\ Y = v + \alpha v_1 + \beta v_2 \end{cases}$$

e, e_1和e_2是系数,它们依赖于u_1, v_1, u_2, v_2和结构参数。

(**b**) 证明方程(26.8)的解是

$$\begin{cases} X' = \sqrt[4]{(e_1 - e_2)^2 + e^2}\cos\left(\frac{1}{2}\arctan(e, e_1 - e_2)\right) \\ Y' = \sqrt[4]{(e_1 - e_2)^2 + e^2}\sin\left(\frac{1}{2}\arctan(e, e_1 - e_2)\right) \end{cases}$$

并且$(X'', Y'') = (-X', -Y')$。

提示:替换变量将式(26.8)改写成三角函数表示的方程组。

参考文献

Adelson, E. and Bergen, J. (1991), The plenoptic function and the elements of early vision, *in* M. Landy and J. Movshon, eds, 'Computational Models of Visual Processing', MIT Press.

Adelson, E. and Weiss, Y. (1996), A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1996, pp. 321–326.

Adini, Y., Moses, Y. and Ullman, S. (1994), Face recognition: The problem of compensating for changes in illumination direction, *in* 'European Conference on Computer Vision', pp. A:286–296.

Adini, Y., Moses, Y. and Ullman, S. (1997), 'Face recognition: The problem of compensating for changes in illumination direction', *IEEE Trans. Pattern Analysis and Machine Intelligence* **19**(7), 721–732.

Agin, G. (1972), Representation and description of curved objects, PhD thesis, Stanford University.

Agin, G. (1981), Fitting ellipses and general second-order curves, Technical report, CMU Robotics Institute.

Aho, A., Hopcroft, J. and Ullman, J. (1974), *The Design and Analysis of Computer Algorithms*, Addison-Wesley.

Ahuja, N. and Abbott, A. (1993), 'Active stereo: Integrating disparity, vergence, focus, aperture, and calibration for surface estimation', *IEEE Trans. Patt. Anal. Mach. Intell.* **15**(10), 1007–1029.

Aikens, R., Agard, D. and Sedat, J. (1989), 'Solid-state imagers for microscopy', *Methods Cell Biol.* **29**, 291–313.

Aloimonos, Y. (1986), Detection of surface orientation from texture. I. the case of planes., *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 584–593.

Aloimonos, Y. (1990), 'Perspective approximations', *Image and Vision Computing* **8**(3), 177–192.

Aloimonos, Y., Weiss, I. and Bandyopadhyay, A. (1987), 'Active vision', *Int. J. of Comp. Vision* **1**(4), 333–356.

Amelio, G., Tompsett, M. and Smith, G. (1970), 'Experimental verification of the charge coupled device concept', *Bell Syst. Tech. J.* **49**, 593–600.

Amenta, N., Bern, M. and Kamvysselis, M. (1998), A new Voronoi-based surface reconstruction algorithm, *in* 'SIGGRAPH-98', pp. 415–421.

Amir, A. and Lindenbaum, M. (1996), Quantitative analysis of grouping processes, *in* 'European Conference on Computer Vision', 1996, pp. I:371–384.

Anderson, B. and Nayakama, K. (1994), 'Toward a general theory of stereopsis—binocular matching, occluding contours, and fusion', *Psych. Review* **101**(3), 414–445.

Anthropology Research Project, (1978), *Anthropometric Source Book*, Webb Associates. NASA reference publication 1024, 3 Vols.

Arbogast, E. and Mohr, R. (1991), '3D structure inference from image sequences', *Journal of Pattern Recognition and Artificial Intelligence*.

Arend, L. and Reeves, A. (1986), 'Simultaneous colour constancy', *J. Opt. Soc. America - A* **3**, 1743–1751.

Armitage, L. and Enser, P. (1997), 'Analysis of user need in image archives', *Journal of Information Science* **23**(4), 287–299.

Arnol'd, V. (1984), *Catastrophe Theory*, Springer-Verlag, Heidelberg.

Arnon, D., Collins, G. and McCallum, S. (1984), 'Cylindrical algebraic decomposition I and II', *SIAM J. Comput.* **13**(4), 865–889.

Avidan, S. and Shashua, A. (1997), Novel view synthesis in tensor space, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 1034–1040.

Ayache, N. (1995*a,b*), 'Medical computer vision, virtual-reality and robotics', *Image and Vision Computing*

13(4), 295–313.

Ayache, N. and Faugeras, O. (1986), 'Hyper: a new approach for the recognition and positioning of two-dimensional objects', *IEEE Trans. Patt. Anal. Mach. Intell.* **8**(1), 44–54.

Ayache, N. and Faugeras, O. (1987), Building, registrating, and fusing noisy visual maps, *in* 'Proc. Int. Conf. Comp. Vision', pp. 73–82.

Ayache, N. and Faverjon, B. (1997), 'Efficient registration of stereo images by matching graph descriptions of edge segments', *Int. J. of Comp. Vision* pp. 101–137.

Ayache, N. and Lustman, F. (1987), Fast and reliable passive trinocular stereovision, *in* 'Proc. Int. Conf. Comp. Vision', pp. 422–427.

Bajcsy, R. (1988), 'Active perception', *Proceedings of the IEEE* **76**(8), 996–1005.

Bajcsy, R. and Solina, F. (1987), Three-dimensional object representation revisited, *in* 'Proc. Int. Conf. Comp. Vision', pp. 231–240.

Baker, H. and Binford, T. (1981), Depth from edge- and intensity-based stereo, *in* 'Proc. International Joint Conference on Artificial Intelligence', pp. 631–636.

Baker, S., Szeliski, R. and Anandan, P. (1998), A layered approach to stereo reconstruction, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 434–441.

Barnard, K. (2000), Improvements to gamut mapping colour constancy algorithms, *in* 'European Conference on Computer Vision', 2000, pp. 390–402.

Barnard, K. and Forsyth, D. (2001), Learning the semantics of words and pictures, *in* 'Int. Conf. on Computer Vision', pp. 408–15.

Barnard, K. and Funt, B. (2002), 'Camera characterization for color research', *Color research and application*. Accepted for publication.

Barnard, K., Duygulu, P. and Forsyth, D. (2001), Clustering art, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition'. To appear.

Barnard, K., Finlayson, G. and Funt, B. (1997), 'Color constancy for scenes with varying illumination', *Computer Vision and Image Understanding* **65**(2), 311–321.

Barrett, E., Brill, M., Haag, N. and Payton, P. (1992), Some invariant linear methods in photogrammetry and model-matching, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1992, pp. 122–128.

Barrett, E., Payton, P., Haag, N. and Brill, M. (1991), 'General methods for determining projective invariants in imagery', *CVGIP: Image Understanding* **53**(1), 46–65.

Basri, R. (1996), 'Parraperspective = affine', *Int. J. of Comp. Vision* **19**(2), 169–179.

Beardsley, P., Zisserman, A. and Murray, D. (1997), 'Sequential updating of projective and affine structure from motion', *Int. J. of Comp. Vision* **23**(3), 235–259.

Beckmann, P. and Spizzichino, A. (1987), *Scattering of Electromagnetic Waves from Rough Surfaces*, Artech House.

Belhumeur, P. and Kriegman, D. (1998), 'What is the set of images of an object under all possible illumination conditions', *International Journal of Computer Vision* **28**(3), 245–260.

Belhumeur, P., Kriegman, D. and Yuille, A. (1999), 'The bas-relief ambiguity', *International Journal of Computer Vision* **35**(1), 33–44.

Belongie, S., Carson, C., Greenspan, H. and Malik, J. (1998), Color- and texture-based image segmentation using the expectation-maximization algorithm and its application to content-based image retrieval, *in* 'Proceedings, Sixth International Conference on Computer Vision', 1998, pp. 675–682.

Berger, M. (1987), *Geometry*, Springer-Verlag.

Bergholm, F. (1987), 'Edge focusing', *IEEE Trans. Pattern Analysis and Machine Intelligence* **9**(6), 726–741.

Bertero, M., Poggio, T. and Torre, V. (1988), 'Ill-posed problems in early vision', *Proceedings of IEEE* **76**(8), 869–889.

Bertsekas, D. (1995), *Nonlinear Programming*, Athena Scientific.

Berzins, V. (1984), 'Accuracy of laplacian edge detectors', *CVGIP: Image Understanding* **27**(2), 195–210.

Besl, P. (1989), 'Active optical range imaging sensors', *Machine vision and applications* **1**, 127–152.

Besl, P. and Jain, R. (1988), 'Segmentation through variable-order surface fitting', *IEEE Trans. Patt. Anal. Mach. Intell.* **10**(2), 167–192.

Besl, P. and McKay, N. (1992), 'A method for registration of 3D shapes', *IEEE Trans. Patt. Anal. Mach. Intell.* **14**(2), 239–256.

Beymer, D. (1994), Face recognition under varying pose, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1994, pp. 756–761.

Beymer, D. and Poggio, T. (1995), Face recognition from one example view, *in* 'Proceedings, Fifth International Conference on Computer Vision', 1995, pp. 500–507.

Beymer, D., McLauchlan, P., Coifman, B. and Malik, J. (1997), A real time computer vision system for measuring traffic parameters, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 495–501.

Biederman, I. (1987), 'Recognition-by-components: A theory of human image understanding', *Psych. Review* **94**(2), 115–147.

Binford, T. (1971), Visual perception by computer, *in* 'Proc. IEEE Conference on Systems and Control'.

Binford, T. (1984), Stereo vision: complexity and constraints, *in* 'Int. Symp. on Robotics Research', MIT Press, pp. 475–487.

Bishop, C. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press.

Black, M. and Anandan, P. (1996), 'The robust estimation of multiple motions: Parametric and piecewise-smooth flow-fields', *Computer Vision and Image Understanding* **63**(1), 75–104.

Blackman, S. and Popoli, R. (1999), *Design and Analysis of Modern Tracking Systems*, Artech House.

Blake, A. (1985), 'Boundary conditions for lightness computation in mondrian world', *CVGIP: Image Understanding* **32**(3), 314–327.

Blake, A. and Marinos, C. (1990), 'Shape from texture: estimation, isotropy and moments', *Artificial Intelligence* **45**(3), 323–380.

Blake, A. and Zisserman, A. (1987), *Visual Reconstruction*, MIT Press.

Blum, H. (1967), A transformation for extracting new descriptors of shape, *in* W. Wathen-Dunn, ed., 'Models for perception of speech and visual form', MIT Press.

Boissonnat, J.-D. (1984), 'Geometric structures for three-dimensional shape representation', *ACM Transaction on Computer Graphics* **3**(4), 266–286.

Boissonnat, J.-D. and Germain, F. (1981), A new approach to the problem of acquiring randomly-oriented workpieces from a bin, *in* 'Proc. International Joint Conference on Artificial Intelligence'.

Bookstein, F. (1979), 'Fitting conic sections to scattered data', *Computer Graphics Image Processing* **9**(1), 56–71.

Boult, T. and Brown, L. (1991), Factorization-based segmentation of motions, *in* 'IEEE Workshop on Visual Motion', pp. 179–186.

Bowyer, K., Kranenburg, C. and Dougherty, S. (1999), Edge detector evaluation using empirical roc curves, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1999, pp. I:354–359.

Boyer, E. (1996), Object Models from Contour Sequences, *in* 'Proceedings of Fourth European Conference on Computer Vision, Cambridge, (England)', pp. 109–118. Lecture Notes in Computer Science, volume 1065.

Boyer, E. and Berger, M.-O. (1996), '3D surface reconstruction using occluding contours', *Int. J. of Comp. Vision*. To appear.

Boyle, W. and Smith, G. (1970), 'Charge coupled semiconductor devices', *Bell Syst. Tech. J.* **49**, 587–593.

Bracewell, R. (1995), *Two-Dimensional Imaging*, Prentice Hall.

Bracewell, R. (2000), *The Fourier Transform and Its Applications, 3ed*, McGraw-Hill.

Brady, J. and Asada, H. (1984), 'Smoothed local symmetries and their implementation', *International Journal of Robotics Research*.

Brady, J., Ponce, J., Yuille, A. and Asada, H. (1985*a*), 'Describing surfaces', *Computer Vision, Graphics and Image Processing* **32**(1), 1–28.

Brady, J., Ponce, J., Yuille, A. and Asada, H. (1985*b*), Describing surfaces, *in* H. Hanafusa and H. Inoue,

eds, 'Proceedings of the 2nd International Symposium on Robotics Research', MIT Press, pp. 5–16.

Brainard, D. and Wandell, B. (1986), 'Analysis of the retinex theory of color vision', *Journal of the Optical Society of America* **3**, 1651–1661.

Brand, J. and Mason, J. (2000), A comparative assessment of three approaches to pixel-level human skin-detection, *in* 'Proceedings, International Conference on Pattern Recognition', pp. Vol I: 1056–1059.

Brelstaff, G. and Blake, A. (1987), 'Computing lightness', *Pattern Recognition Letters* **5**, 129–138.

Brelstaff, G. and Blake, A. (1988), Detecting specular reflection using lambertian constraints, *in* 'Proceedings, Second International Conference on Computer Vision', 1988, pp. 297–302.

Brill, E. (1992), A simple rule-based part of speech tagger, *in* 'Proc. Third Conference on Applied Natural Language Processing'.

Brooks, R. (1981*a*), 'Symbolic reasoning among 3-D models and 2-D images', *Artificial Intelligence Journal* **17**(1-3), 285–348.

Brooks, R. (1981*b*), Symbolic Reasoning among 3-D Models and 2-D Images, PhD thesis, Stanford University Computer Science Dept.

Brooks, R., Greiner, R. and Binford, T. (1979), Model-based three-dimensional interpretation of two-dimensional images, *in* 'Proc. International Joint Conference on Artificial Intelligence', Tokyo, Japan, pp. 105–113.

Brostow, G. and Essa, I. (1999), Motion based decompositing of video, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 8–13.

Bruce, J., Giblin, P. and Tari, F. (1996*a*), 'Parabolic curves of evolving surfaces', *Int. J. of Comp. Vision* **17**(3), 291–306.

Bruce, J., Giblin, P. and Tari, F. (1996*b*), 'Ridges, crests and sub-parabolic lines of evolving surfaces', *Int. J. of Comp. Vision* **18**(3), 195–210.

Brunelli, R. and Poggio, T. (1992), Face recognition through geometrical features, *in* Sandini (1992), pp. 792–800.

Brunelli, R. and Poggio, T. (1993), 'Face recognition: Features versus templates', *IEEE Trans. Pattern Analysis and Machine Intelligence* **15**(10), 1042–1052.

Brunie, L., Lavallée, S. and Szeliski, R. (1992), Using force fields derived from 3D distance maps for inferring the attitude of a 3D rigid object, *in* G. Sandini, ed., 'Proc. European Conf. Comp. Vision', Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 670–675.

Brunnström, K., Ekhlund, J.-O. and Uhlin, T. (1996), 'Active fixation for scene exploration', *Int. J. of Comp. Vision* **17**(2), 137–162.

Bubna, K. and Stewart, C. (2000), 'Model selection techniques and merging rules for range data segmentation algorithms', *Computer Vision and Image Understanding* **80**(2), 215–245.

Buchsbaum, G. (1980), 'A spatial processor model for object colour perception', *J. Franklin Inst.* **310**, 1–26.

Bülthoff, H., Tarr, M., Blanz, V. and Zabinski, M. (1995), 'To what extent do unique parts influence recognition across changes in viewpoint?', *Perception* **24**, 3.

Burl, M. and Perona, P. (1996), Recognition of planar object classes, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1996, pp. 223–230.

Cabrera, J. and Meer, P. (1996), 'Unbiased estimation of ellipses by bootstrapping', *IEEE Trans. Pattern Analysis and Machine Intelligence* **18**(7), 752–756.

Callahan, J. and Weiss, R. (1985), A model for describing surface shape, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 240–245.

Canny, J. (1986), 'A computational approach to edge detection', *IEEE Trans. Pattern Analysis and Machine Intelligence* **8**(6), 679–698.

Canny, J. (1988), *The Complexity of Robot Motion Planning*, MIT Press.

Carmichael, O., Huber, D. and Hebert, M. (1999), Large data sets and confusing scenes in 3-D surface matching and recognition, *in* 'Second International Conference on 3-D Digital Imaging and Modeling (3DIM'99)', pp. 358–367.

Carson, C., Thomas, M., Belongie, S., Hellerstein, J. and Malik, J. (1999), Blobworld: A system for region-

based image indexing and retrieval, *in* 'Third Int. Conf. on Visual Information Systems', Lecture Notes in Computer Science 1614, Springer-Verlag, pp. 509–516.

Cascia, M. L., Sethi, S. and Sclaroff, S. (1998), Combining textual and visual cues for content based image retrieval on the web, *in* 'IEEE Workshop on Content Based Access of Image and Video Libraries', pp. 24–28.

Castore, G. (1984), Solid Modeling, Aspect Graphs, and Robot Vision, *in* Pickett and Boyse, eds, 'Solid modeling by computer', Plenum Press, pp. 277–292.

Chakravarty, I. (1982), The use of characteristic views as a basis for recognition of three-dimensional objects, Image Processing Laboratory IPL-TR-034, Rensselaer Polytechnic Institute.

Chang, S. and Yang, C. (1983), 'Picture information measures for similarity retrieval', *CVGIP: Image Understanding* **23**, 366–375.

Chang, S.-F., Chen, W. and Sundaram, H. (1998*a*), Semantic visual templates—linking visual features to semantics, *in* 'IEEE Int. Conf. Image Processing', pp. 531–535.

Chang, S.-F., Chen, W., Meng, H., Sundaram, H. and Zhong, D. (1997*a*), Videoq—an automatic content-based video search system using visual cues, *in* 'ACM Multimedia Conference'.

Chang, S.-F., Chen, W., Meng, H., Sundaram, H. and Zhong, D. (1998*b*), 'A fully automated content based video search engine supporting spatiotemporal queries', *IEEE Trans. Circuits and Systems for Video Technology* **8**(8), 602–615.

Chang, S.-F., Smith, J., Beigi, M. and Benitez, A. (1997*b*), 'Visual information retrieval from large distributed online repositories', *Comm. ACM* **40**(12), 63–71.

Chapelle, O., Haffner, P. and Vapnik, V. (1999), 'Support vector machines for histogram-based image classification', *IEEE Neural Networks* **10**(5), 1055–1064.

Chasles, M. (1855), 'Question no. 296', *Nouv. Ann. Math.*

Chellappa, R. and Jain, A. (1993), *Markov Random Fields: Theory and Applications*, Academic Press.

Chellappa, R., Wilson, C. and Sirohey, S. (1995), 'Human and machine recognition of faces: a survey', *Proceedings IEEE* **83**(5), 705–740.

Chen, S. (1995), Quicktime VR: An image-based approach to virtual environment navigation, *in* 'SIGGRAPH', pp. 29–38.

Chen, S. and Freeman, H. (1991), On the characteristic views of quadric-surfaced solids, *in* 'IEEE Workshop on Directions in Automated CAD-Based Vision', pp. 34–43.

Chin, R. and Dyer, C. (1986), 'Model-based recognition in robot vision', *ACM Computing Surveys* **18**(1), 67–108.

Chiyokura, B. and Kimura, F. (1983), 'Design of solids with free-form surfaces', *Computer Graphics* **17**(3), 289–298.

Cho, K., Meer, P. and Cabrera, J. (1997), 'Performance assessment through bootstrap', *IEEE Trans. Pattern Analysis and Machine Intelligence* **19**(11), 1185–1198.

Cho, K., Meer, P. and Cabrera, J. (1998), 'Performance assessment through bootstrap', *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**(1), 94.

Christy, S. and Horaud, R. (1996), 'Euclidean shape and motion from multiple perspective views by affine iterations', *IEEE Trans. Patt. Anal. Mach. Intell.* **18**(11), 1098–1104.

Chua, C. and Jarvis, R. (1996), 'Point signatures: a new representation for 3D object recognition', *Int. J. of Comp. Vision* **25**(1), 63–85.

Chui, C. (1991), *Kalman Filtering: With Real-Time Applications*, Springer-Verlag.

Cipolla, R. and Blake, A. (1992), 'Surface shape from the deformation of the apparent contour', *Int. J. of Comp. Vision* **9**(2), 83–112.

Cipolla, R., Astrom, K. and Giblin, P. (1995), Motion from the frontier of curved surfaces, *in* 'Proc. Int. Conf. Comp. Vision', pp. 269–275.

Clarkson, K. (1988), 'A randomized algorithm for closest-point queries', *SIAM J. Computing* **17**, 830–847.

Clerc, M. and Mallat, S. (1999), Shape from texture through deformations, *in* 'Int. Conf. on Computer Vision', pp. 405–410.

Clowes, M. (1971), 'On seeing things', *Artificial Intelligence* **2**(1), 79–116.

Cohen, J. (1964), 'Dependency of the spectral reflectance curves of the munsell color chips', *Psychon. Sci.* **1**, 369–370.

Cohen, M. and Wallace, J. (1993), *Radiosity and realistic image synthesis*, Academic Press.

Collins, G. (1971), 'The calculation of multivariate polynomial resultants', *Journal of the ACM* **18**(4), 515–522.

Collins, G. (1975), *Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition*, Vol. 33 of *Lecture Notes in Computer Science*, Springer-Verlag, New York.

Connolly, C. and Stenstrom, J. (1989), 3D scene reconstruction from multiple intensity images, *in* 'Proc. IEEE Workshop on Interpretation of 3D Scenes', pp. 124–130.

Cook, R. and Torrance, K. (1987), A reflectance model for computer graphics, *in* 'ARPA Image Understanding Workshop', pp. 1–19.

Costeira, J. and Kanade, T. (1998), 'A multi-body factorization method for motion analysis', *Int. J. of Comp. Vision* **29**(3), 159–180.

Cover, T. and Thomas, J. (1991), *Elements of Information Theory*, Wiley-Interscience.

Cox, I., Zhong, Y. and Rao, S. (1996), Ratio regions: A technique for image segmentation, *in* 'Proceedings, International Conference on Pattern Recognition', pp. 557–564.

Coxeter, H. (1974), *Projective Geometry*, Springer-Verlag. Second edition.

Craig, J. (1989), *Introduction to Robotics: Mechanics and Control*, Addison-Wesley. Second edition.

Cross, G. and Jain, A. (1983), 'Markov random field texture models', *IEEE Trans. Pattern Analysis and Machine Intelligence* **5**(1), 25–39.

Cross, G., Fitzgibbon, A. and Zisserman, A. (1999), Parallax geometry of smooth surfaces in multiple views, *in* 'Proc. Int. Conf. Comp. Vision', pp. 323–329.

Csurka, G. and Faugeras, O. (1998), 'Computing 3-Dimensional project invariants from a pair of images using the grassmann-cayley algebra', *Image and Vision Computing* **16**(1), 3–12.

Csurka, G. and Faugeras, O. (1999), 'Algebraic and geometric tools to compute projective and permutation invariants', *IEEE Trans. Pattern Analysis and Machine Intelligence* **21**(1), 58–65.

Curless, B. and Levoy, M. (1996), A volumetric method for building complex models from range images, *in* 'SIGGRAPH'.

Darrell, T. and Simoncelli, E. (1993), 'Nulling' filters and the separation of transparent motions, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1993, pp. 738–739.

Davis, L. (1975), 'A survey of edge detection techniques', *Computer Graphics Image Processing* **4**(3), 248–270.

de Bonet, J. and Viola, P. (1997), Rosetta: An image database retrieval system, *in* 'Proc. DARPA IU Workshop', pp. 655–660.

Debevec, P., Taylor, C. and Malik, J. (1996), Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach, *in* 'SIGGRAPH', pp. 11–20.

Dellaert, F., Seitz, S., Thorpe, C. and Thrun, S. (2000), Structure from motion without correspondence, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 2000, pp. II:557–564.

Demazure, M. (2000), *Bifurcations and Catastrophes*, Springer-Verlag. Translated by D. Chillingworth.

Dempster, A., Laird, N. and Rubin, D. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society* **39** **(Series B)**, 1–38.

Deriche, R. (1987), 'Using Canny's criteria to derive a recursively implemented optimal edge detector', *International Journal of Computer Vision* **1**(2), 167–187.

Deriche, R. (1990), 'Fast algorithms for low-level vision', *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**(1), 78–87.

Devernay, F. and Faugeras, O. (1994), Computing differential properties of 3D shapes from stereopsis without 3D models, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 208–213.

Devroye, L., Gyorfi, L. and Lugosi, G. (1996), *A Probabilistic Theory of Pattern Recognition*, Springer Verlag.

Devy, M., Garric, V. and Orteu, J. (1997), Camera calibration from multiple views of a 2D object using a lgobal non-linear minimization method, *in* 'IEEE/RSJ International Conference on Intelligent Robots and Systems', pp. 1583–1589.

do Carmo, M. (1976), *Differential Geometry of Curves and Surfaces*, Prentice-Hall.

Dougherty, S. and Bowyer, K. (1998), Objective evaluation of edge detectors using a formally defined

framework, *in* 'Workshop on Empirical Evaluation Methods in Computer Vision', p. xx.

Dove, H. (1841), 'Über stereoskopie', *Annals Phys. Series 2* **110**, 494–498.

Drew, M. and Funt, B. (1990), Calculating surface reflectance using a single-bounce model of mutual reflection, *in* 'Proceedings, Third International Conference on Computer Vision', 1990, pp. 393–399.

Driscoll, W. and Vaughan, W., eds (1978), *Handbook of Optics*, McGraw-Hill.

Duda, R. and Hart, P. (1973), *Pattern Classification and Scene Analysis*, Wiley.

Duncan, J. and Ayache, N. (2000), 'Medical image analysis: Progress over two decades and the challenges ahead', *IEEE Trans. Pattern Analysis and Machine Intelligence* **22**(1), 85–106.

D'Zmura, M. and Lennie, P. (1986), 'Mechanisms of colour constancy', *J. Opt. Soc. America - A* **3**, 1662–1672.

Eakins, J., Boardman, J. and Graham, M. (1998), 'Similarity retrieval of trademark images', *IEEE Multimedia* **5**(2), 53–63.

Ebert, D. S., Musgrave, F. K., Peachey, D., Worley, S. and Perlin, K., eds (1998), *Texturing and Modeling*, Morgan Kaufmann.

Efros, A. and Leung, T. (1999), Texture synthesis by non-parametric sampling, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 1033–1038.

Eggert, D. and Bowyer, K. (1989), Computing the orthographic projection aspect graph of solids of revolution, *in* 'Proc. IEEE Workshop on Interpretation of 3D Scenes', pp. 102–108.

Eggert, D. and Bowyer, K. (1991), Perspective projection aspect graphs of solids of revolution: An implementation, *in* 'IEEE Workshop on Directions in Automated CAD-Based Vision', pp. 44–53.

Eggert, D., Bowyer, K., Dyer, C., Christensen, H. and Goldgof, D. (1993), 'The scale space aspect graph', *IEEE Trans. Patt. Anal. Mach. Intell.* **15**(11), 1114–1130.

Elder, J. and Zucker, S. (1998), 'Local scale control for edge detection and blur estimation', *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**(7), 699–716.

Enser, P. (1993), 'Query analysis in a visual information retrieval context', *J. Document and Text Management* **1**(1), 25–52.

Enser, P. (1995), 'Pictorial information retrieval', *Journal of Documentation* **51**(2), 126–170.

Ettinger, G. (1988), Large hierarchical object recognition using libraries of parameterized model sub-parts, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1988, pp. 32–41.

Fairchild, M. (1998), *Color Appearance Models*, Addison-Wesley.

Fan, T., Médioni, G. and Nevatia, R. (1987), 'Segmented descriptions of 3D surfaces', *IEEE Transactions on Robotics and Automation* **3**(6), 527–538.

Farin, G. (1993), *Curves and Surfaces for Computer Aided Geometric Design*, Academic Press, San Diego, CA.

Faugeras, O. (1992), What can be seen in three dimensions with an uncalibrated stereo rig?, *in* G. Sandini, ed., 'Proc. European Conf. Comp. Vision', Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 563–578.

Faugeras, O. (1993), *Three-Dimensional Computer Vision*, MIT Press.

Faugeras, O. (1995), 'Stratification of 3D vision: projective, affine and metric representations', *J. Opt. Soc. Am. A* **12**(3), 465–484.

Faugeras, O. and Hebert, M. (1986), 'The representation, recognition, and locating of 3-D objects', *International Journal of Robotics Research* **5**(3), 27–52.

Faugeras, O. and Maybank, S. (1990), 'Motion from point matches: multiplicity of solutions', *Int. J. of Comp. Vision* **4**(3), 225–246.

Faugeras, O. and Mourrain, B. (1995), On the geometry and algebra of the point and line correspondences between *n* images, Technical Report 2665, INRIA Sophia-Antipolis.

Faugeras, O. and Papadopoulo, T. (1997), Gaussman-Caylay algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors, Technical Report 3225, INRIA Sophia-Antipolis.

Faugeras, O., Hebert, M., Pauchon, E. and Ponce, J. (1984), Object representation, identification, and po-

sitioning from range data, *in* 'Robotics Research: The First International Symposium', MIT Press, pp. 425–446.

Faugeras, O., Luong, Q.-T. and Papadopoulo, T. (2001), *The Geometry of Multiple Images*, MIT Press.

Felzenszwalb, P. and Huttenlocher, D. (2000), Efficient matching of pictorial structures, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 2000, pp. II:66–73.

Feng, X. and Perona, P. (1998), Scene segmentation from 3D motion, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 225–231.

Ferryman, J., Maybank, S. and Worrall, A. (2000), 'Visual surveillance for moving vehicles', *International Journal of Computer Vision* **37**(2), 187–197.

Finlayson, G. and Hordley, S. (1999), 'Selection for gamut mapping colour constancy', *Image and Vision Computing* **17**(8), 597–604.

Finlayson, G. and Hordley, S. (2000), 'Improving gamut mapping color constancy', *IEEE Trans. Image Processing* **9**(10), 1774–1783.

Finlayson, G., Chatterjee, S. and Funt, B. (1996), Color angular indexing, *in* 'European Conference on Computer Vision', 1996, pp. II:16–27.

Finlayson, G., Drew, M. and Funt, B. (1994*a*), 'Color constancy: Generalized diagonal transforms suffice', *Journal of the Optical Society of America* **11**(11), 3011–3019.

Finlayson, G., Drew, M. and Funt, B. (1994*b*), 'Spectral sharpening: Sensor transformations for improved color constancy', *Journal of the Optical Society of America* **11**(5), 1553–1563.

Fischler, M. and Bolles, R. (1981), 'Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography', *Communications of the ACM* **24**(6), 381–395.

Fitzgibbon, A. and Zisserman, A. (1998), Automatic 3D model acquisition and generation of new images from video sequences, *in* 'European Signal Processing Conference', pp. 311–326.

Fleck, M. (1992*a*), 'Multiple widths yield reliable finite differences', *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**(4), 412–429.

Fleck, M. (1992*b*), 'Some defects in finite-difference edge finders', *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**(3), 337–345.

Fleck, M. (1995), Cs-tr 95-01: Perspective projection: the wrong imaging model, Technical report, University of Iowa.

Fleck, M., Forsyth, D. and Bregler, C. (1996), Finding naked people, *in* 'European Conference on Computer Vision', 1996, pp. II:593–602.

Flickner, M., Sawhney, H., Niblack, W. and Ashley, J. (1995), 'Query by image and video content: the QBIC system', *Computer* **28**(9), 23–32.

Flock, H. (1984), 'Illumination: inferred or observed?', *Perception and Psychophysics*.

Foley, J., van Dam, A., Feiner, S. and Hughes, J. (1990), *Computer Graphics: Principle and Practice*, Addison-Wesley. Second edition.

Forney, G. (1973), 'The Viterbi algorithm', *Proceedings of the IEEE*.

Forsyth, D. (1990), 'A novel approach to color constancy', *International Journal of Computer Vision* **5**(1), 5–36.

Forsyth, D. (1996), 'Recognizing algebraic surfaces from their outlines', *International Journal of Computer Vision* **18**(1), 21–40.

Forsyth, D. (1999), Sampling, resampling and colour constancy, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1999, pp. I:300–305.

Forsyth, D. and Fleck, M. (1997), Body plans, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 678–683.

Forsyth, D. and Fleck, M. (1999), 'Automatic detection of human nudes', *International Journal of Computer Vision* **32**(1), 63–77.

Forsyth, D. and Zisserman, A. (1989), Mutual illumination, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1989, pp. 466–473.

Forsyth, D. and Zisserman, A. (1990), 'Shape from shading in the light of mutual illumination', *Image and*

Vision Computing **8**, 42–29.

Forsyth, D. and Zisserman, A. (1991), 'Reflections on shading', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(7), 671–679.

Forsyth, D., Mundy, J., Zisserman, A. and Rothwell, C. (1992), Recognizing rotationally symmetric surfaces from their outlines, *in* Sandini (1992), pp. 639–647.

Forsyth, D., Mundy, J., Zisserman, A. and Rothwell, C. (1994), Using global consistency to recognise euclidean objects with an uncalibrated camera, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1994, pp. 502–507.

Forsyth, D., Mundy, J., Zisserman, A., Coelho, C., Heller, A. and Rothwell, C. (1991), 'Invariant descriptors for 3-D object recognition and pose', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(10), 971–991.

Freeman, W. and Adelson, E. (1991), 'The design and use of steerable filters', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(9), 891–906.

Freeman, W. and Brainard, D. (1997), 'Bayesian color constancy', *Journal of the Optical Society of America* **14**(7), 1393–1411.

Freeman, W., Anderson, D. and et al., P. B. (1998), 'Computer vision for interactive computer graphics', *Computer Graphics and Applications* pp. 42–53.

Freeman, W., Pasztor, E. and Carmichael, O. (2000), 'Learning low-level vision', *International Journal of Computer Vision* **40**(1), 25–47.

Friedman, J., Bentley, J. and Finkel, R. (1977), 'An algorithm for finding best matches in logarithmic expected time', *ACM Trans. on Math. Software*.

Frisby, J. (1980), *Seeing: Illusion, Brain and Mind*, Oxford University Press.

Fu, K. and Mui, J. (1981), 'A survey of image segmentation', *Pattern Recognition* **13**(1), 3–16.

Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition*, Academic Press. 2nd edition.

Funt, B. and Drew, M. (1988), Color constancy computation in near-mondrian scenes using a finite dimensional linear model, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1988, pp. 544–549.

Funt, B. and Drew, M. (1993), 'Color space analysis of mutual illumination', *IEEE Trans. Pattern Analysis and Machine Intelligence* **15**(12), 1319–1326.

Funt, B. and Finlayson, G. (1995), 'Color constant color indexing', *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**(5), 522–529.

Funt, B., Barnard, K. and Martin, L. (1998), Is machine colour constancy good enough?, *in* 'European Conference on Computer Vision', 1998, pp. 445–459.

Funt, B., Drew, M. and Ho, J. (1991), 'Color constancy from mutual reflection', *International Journal of Computer Vision* **6**(1), 5–24.

Garcia-Bermejo, J., Diaz Pernas, F. and Coronado, J. (1996), An approach for determining bidirectional reflectance parameters from range and brightness data, *in* 'Proceedings, International Conference on Image Processing', p. 16A2.

Garding, J. (1992), Shape from texture for smooth curved surfaces, *in* 'European Conference on Computer Vision', pp. 630–638.

Garding, J. (1995), Surface orientation and curvature from differential texture distortion, *in* 'Int. Conf. on Computer Vision', pp. 733–739.

Gaston, P. and Lozano-Pérez, T. (1984), 'Tactile recognition and localization using object models: The case of polyhedra in the plane', *IEEE Trans. Patt. Anal. Mach. Intell.*

Gear, C. (1998), 'Multibody grouping in moving objects', *Int. J. of Comp. Vision* **29**(2), 133–150.

Genc, Y. and Ponce, J. (1998), Parameterized image varieties: A novel approach to the analysis and synthesis of image sequences, *in* 'Proc. Int. Conf. Comp. Vision', pp. 11–16.

Genc, Y. and Ponce, J. (2001), 'Image-based rendering using parameterized image varieties', *Int. J. of Comp. Vision* **41**(3), 143–170.

Gennery, D. (1980), Modelling the environment of an exploring vehicle by means of stereo vision, PhD

thesis, Stanford University.

Georghiades, A., Belhumeur, P. and Kriegman, D. (2000), From few to many: Generative models for recognition under variable pose and illumination, *in* 'International Conference on Automatic Face and Gesture Recognition', 1900, pp. 277–284.

Georghiades, A., Kriegman, D. and Belhumeur, P. (1998), Illumination cones for recognition under variable lighting: Faces, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 52–59.

Gerig, G., Pun, T. and Ratib, O. (1994), 'Image analysis and computer vision in medicine', *Computerized Medical Imaging and Graphics* **18**(2), 85–96.

Gersho, A. and Gray, R. (1992), *Vector quantization and signal compression*, Kluwer Academic Publishers.

Gershon, R. (1987), The Use of Color in Computational Vision, PhD thesis, University of Rochester.

Gershon, R., Jepson, A. and Tsotsos, J. (1986), 'Ambient illumination and the determination of material changes', *J. Opt. Soc. America* **A-3**(10), 1700–1707.

Giblin, P. and Kimia, B. (1999), On the local form and transitions of symmetry sets, medial axes, and shocks, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 385–391.

Gigus, Z. and Malik, J. (1990), 'Computing the aspect graph for line drawings of polyhedral objects', *IEEE Trans. Patt. Anal. Mach. Intell.* **12**(2), 113–122.

Gigus, Z., Canny, J. and Seidel, R. (1991), 'Efficiently computing and representing aspect graphs of polyhedral objects', *IEEE Trans. Patt. Anal. Mach. Intell.*

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V. and Economou, E. (1999), 'An anchoring theory of lightness perception', *Psychological Review* **106**(4), 795–834.

Gill, P., Murray, W. and Wright, M. (1981), *Practical Optimization*, Academic Press.

Gordon, I. (1997), *Theories of Visual Perception*, John Wiley and Son.

Gortler, S., Grzeszczuk, R., Szeliski, R. and Cohen, M. (1996), The lumigraph, *in* 'SIGGRAPH', pp. 43–54.

Greenspan, H., Belongie, S., Perona, P., Goodman, R., Rakshit, S. and Anderson, C. (1994), Overcomplete steerable pyramid filters and rotation invariance, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1994, pp. 222–228.

Grimson, W. (1981*a*), 'A computer implementation of a theory of human stereo vision', *Philosophical Transactions of the Royal Society of London* pp. 217–253.

Grimson, W. (1981*b*), *From images to surfaces:A Computational Study of the Human Early Visual System*, MIT Press.

Grimson, W. (1992), 'The cost of choosing the wrong model in object recognition by constrained search', *International Journal of Computer Vision* **7**(3), 195–210.

Grimson, W. and Huttenlocher, D. (1990*a*), On the sensitivity of geometric hashing, *in* 'Proceedings, Third International Conference on Computer Vision', 1990, pp. 334–338.

Grimson, W. and Huttenlocher, D. (1990*b*), 'On the sensitivity of the Hough transform for object recognition', *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**(3), 255–274.

Grimson, W. and Huttenlocher, D. (1991), 'On the verification of hypothesized matches in model-based recognition', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(12), 1201–1213.

Grimson, W. and Lozano-Perez, T. (1984), 'Model-based recognition and localization from sparse range or tactile data', *International Journal of Robotics Research* **3**(3), 3–35.

Grimson, W. and Lozano-Pérez, T. (1987), 'Localizing overlapping parts by searching the interpretation tree', *IEEE Trans. Patt. Anal. Mach. Intell.* **9**(4), 469–482.

Grimson, W., Huttenlocher, D. and Alter, T. (1992), Recognizing 3D objects from 2D images: An error analysis, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1992, pp. 316–321.

Grimson, W., Huttenlocher, D. and Jacobs, D. (1994), 'A study of affine matching with bounded sensor error', *International Journal of Computer Vision* **13**(1), 7–32.

Grimson, W., Lozano-Perez, T. and Huttenlocher, D. (1990), *Object Recognition by Computer: The Role of Geometric Constraints*, MIT Press.

Gross, A. and Boult, T. (1988), Error of fit measures for recovering parametric solids, *in* 'Proc. Int. Conf.

Comp. Vision', pp. 690–694.

Haddon, J. and Forsyth, D. (1998*a*), Shading primitives: Finding folds and shallow grooves, *in* 'Proceedings, Sixth International Conference on Computer Vision', 1998, pp. 236–241.

Haddon, J. and Forsyth, D. (1998*b*), Shape representations from shading primitives, *in* 'European Conference on Computer Vision', 1998, pp. 415–431.

Hamilton, W. (1844), 'On a new species of imaginary quantities connected with a theory of quaternions', *Transactions of the Royal Irish Academy* **2**, 424–434.

Hampapur, A., Gupta, A., Horowitz, B. and Shu, C.-F. (1997), Virage video engine, *in* 'Storage and Retrieval for Image and Video Databases V—Proceedings of the SPIE', Vol. 3022, pp. 188–198.

Haralick, R. (1983), 'Ridges and valleys in digital images', *Computer Vision, Graphics and Image Processing* **22**, 28–38.

Haralick, R. and Shapiro, L. (1985), 'Image segmentation techniques', *CVGIP: Image Understanding* **29**(1), 100–132.

Haralick, R. and Shapiro, L. (1992), *Computer and robot vision*, Addison Wesley.

Haralick, R., Watson, L. and Laffey, T. (1983), 'The topographic primal sketch', *International Journal of Robotics Research* **2**, 50–72.

Hardin, C. and Maffi, L. (1997), *Color Categories in thought and language*, Cambridge University Press.

Harris, C. and Stephens, M. (1988), A combined corner and edge detector, *in* 'Alvey Conference', pp. 147–152.

Hartley, R. (1994*a*), An algorithm for self calibration from several views, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 908–912.

Hartley, R. (1994*b*), 'Projective reconstruction and invariants from multiple images', *IEEE Trans. Patt. Anal. Mach. Intell.* **16**(10), 1036–1041.

Hartley, R. (1995), In defence of the 8-point algorithm, *in* 'Proc. Int. Conf. Comp. Vision', pp. 1064–1070.

Hartley, R. (1997), 'Lines and points in three views and the trifocal tensor', *Int. J. of Comp. Vision* **22**(2), 125–140.

Hartley, R. (1998), Computation of the quadrifocal tensor, *in* 'Proc. European Conf. Comp. Vision', pp. 20–35.

Hartley, R. and Sturm, P. (1997), 'Triangulation', *Computer Vision and Image Understanding* **68**(2), 146–157.

Hartley, R. and Zisserman, A. (2000), *Multiple view geometry in computer vision*, Cambridge University Press.

Hartley, R., Gupta, R. and Chang, T. (1992), Stereo from uncalibrated cameras, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 761–764.

Hartley, R., Wang, C., Kitchen, L. and Rosenfeld, A. (1982), 'Segmentation of FLIR images: A comparative study', *IEEE Trans. Systems, Man and Cybernetics* **12**(4), 553–566.

Hastie, T., Tibshirani, R. and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Verlag.

Havaldar, P., Lee, M. and Medioni, G. (1996), View synthesis from unregistered 2D images, *in* 'Graphics Interface'96', pp. 61–69.

Haykin, S. (1999), *Neural Networks: A Comprehensive Introduction*, Prentice-Hall.

Healey, G. and Kondepudy, R. (1994), 'Radiometric CCD camera calibration and noise estimation', *IEEE Trans. Patt. Anal. Mach. Intell.* **16**(3), 267–276.

Heath, M. (2002), *Scientific Computing: An Introductory Survey*, McGraw-Hill. Second edition.

Heath, M., Sarkar, S., Sanocki, T. and Bowyer, K. (1997), 'A robust visual method for assessing the relative performance of edge detection algorithms', *IEEE Trans. Pattern Analysis and Machine Intelligence* **19**(12), 1338–1359.

Hebert, M. (2000), Active and passive range sensing for robotics, *in* 'IEEE Int. Conf. on Robotics and Automation'.

Hebert, M. and Kanade, T. (1985), The 3D profile method for object recognition, *in* 'Proc. IEEE Conf.

Comp. Vision Patt. Recog.', pp. 458–463.

Hecht, E. (1987), *Optics*, Addison-Wesley.

Hel-Or, Y. and Teo, P. (1996), Canonical decomposition of steerable functions, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1996, pp. 809–816.

Helmholtz, H. (1909), *Physiological Optics*, Dover. 1962 edition of the English translation of the 1909 German original, first published by the Optical Society of America in 1924.

Helson, H. (1934), 'Some factors and implications of colour constancy', *J. Opt. Soc. America* **48**, 555–567.

Helson, H. (1938*a*), Fundamental problems in color vision, i, *in* 'Journal of Experimental Psychology', Vol. 23.

Helson, H. (1938*b*), Fundamental problems in color vision, ii, *in* 'Journal of Experimental Psychology', Vol. 26.

Herskovits, A. and Binford, T. (1970), On boundary detection, Technical report, MIT AI Lab.

Hesse, O. (1863), 'Die cubische Gleichung, von welcher die Lösung des Problems der Homographie von M. Chasles abhängt', *J. Reine Angew. Math.* **62**, 188–192.

Heyden, A. (1995), Geometry and algebra of multiple projective transformations, PhD thesis, Lund University, Sweden.

Heyden, A. (1998), A common framework for multiple view tensors, *in* 'Proc. European Conf. Comp. Vision', pp. 3–19.

Heyden, A. and Åström, K. (1996), Euclidean reconstruction from constant intrinsic parameters, *in* 'International Conference on Pattern Recognition', pp. 339–343.

Heyden, A. and Åström, K. (1998), Minimal conditions on intrinsic parameters for Euclidean reconstruction, *in* 'Asian Conference on Computer Vision'.

Heyden, A. and Åström, K. (1999), Flexible calibration: minimal cases for auto-calibration, *in* 'Proc. Int. Conf. Comp. Vision', pp. 350–355.

Hilbert, D. and Cohn-Vossen, S. (1952), *Geometry and the Imagination*, Chelsea.

Holst, G. (1998), *CCD Arrays, Cameras and Displays*, SPIE Press.

Holt, B. and Hartwick, L. (1994*a*), 'Quick, who painted fish?': searching a picture database with the QBIC project at UC Davis', *Information Services and Use* **14**(2), 79–90.

Holt, B. and Hartwick, L. (1994*b*), Retrieving art images by image content: the UC Davis QBIC project, *in* 'ASLIB Proceedings', Vol. 46, pp. 243–8.

Horn, B. (1970), Shape from shading: A method for obtaining the shape of a smooth opaque object from one view, Technical report, MIT AI Lab.

Horn, B. (1971), The Binford-Horn line finder, Technical report, MIT AI Lab.

Horn, B. (1974), 'Determining lightness from an image', *Computer Graphics Image Processing* **3**(1), 277–299.

Horn, B. (1975), Obtaining shape from shading information, *in* 'The Psychology of Computer Vision', McGraw-Hill, pp. 115–155.

Horn, B. (1977), 'Understanding image intensities', *Artificial Intelligence* **8**(2), 201–231.

Horn, B. (1984), 'Extended Gaussian images', *Proceedings of the IEEE* **72**(12), 1671–1686.

Horn, B. (1986), *Robot Vision*, MIT Press, Cambridge, MA.

Horn, B. (1987), 'Closed-form solution of absolute orientation using unit quaternions', *J. Opt. Soc. Am. A* **4**(4), 629–642.

Horn, B. (1990), 'Height and gradient from shading', *International Journal of Computer Vision* **5**(1), 37–76.

Horn, B. and Brooks, M. (1989), *Shape from Shading*, MIT Press.

Horn, B. and Schunck, B. (1981), 'Determining optical flow', *Artificial Intelligence* **17**, 185–203.

Horn, B., Woodham, R. and Silver, W. (1978), Determining shape and reflectance using multiple images, Technical report, MIT AI Lab.

Hsu, S., Anandan, P. and Peleg, S. (1994), Accurate computation of optical flow by using layered motion representations, *in* 'Proceedings, International Conference on Pattern Recognition', pp. A:743–746.

Huang, J. and Zabih, R. (1998), Combining color and spatial information for content-based image retrieval, *in* 'European Conference on Digital Libraries'. Web version at http://www.cs.cornell.edu/html/rdz/Papers/ECDL2/spatial.htm.

Huang, J., Kumar, S., Mitra, M., Zhu, W.-J. and Zabih, R. (1997), Image indexing using color correlograms, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 762–768.

Huang, T. (1981), *Image Sequence Analysis*, Springer-Verlag.

Huang, T. and Faugeras, O. (1989), 'Some properties of the E-matrix in two-view motion estimation', *IEEE Trans. Patt. Anal. Mach. Intell.* **11**(12), 1310–1312.

Huber, P. (1981), *Robust Statistics*, Wiley.

Hueckel, M. (1971), 'An operator which locates edges in digitized pictures', *Journal of the ACM* **18**(1), 113–125.

Huffman, D. (1977), 'Realizable configurations of lines in pictures of polyhedra', *Machine Intelligence* **8**, 493–509.

Huttenlocher, D. and Ullman, S. (1987), Object recognition using alignment, *in* 'Proc. Int. Conf. Comp. Vision', pp. 102–111.

Huttenlocher, D. and Ullman, S. (1990), 'Recognizing solid objects by alignment with an image', *International Journal of Computer Vision* **5**(2), 195–212.

Huttenlocher, D. and Wayner, P. (1992), 'Finding convex edge groupings in an image', *International Journal of Computer Vision* **8**(1), 7–27.

Ikeuchi, K. (1987*a*), 'Determining a depth map using a dual photometric stereo', *International Journal of Robotics Research*.

Ikeuchi, K. (1987*b*), Precompiling a geometrical model into an interpretation tree for object recognition in bin-picking tasks, *in* 'Proc. DARPA Image Understanding Workshop', pp. 321–339.

Ikeuchi, K. and Kanade, T. (1988), 'Automatic generation of object recognition programs', *Proceedings of the IEEE* **76**(8), 1016–35.

Ioffe, S. and Forsyth, D. (1998), Learning to find pictures of people, *in* 'NIPS'.

Ioffe, S. and Forsyth, D. (1999), Finding people by sampling, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 1092–1097.

Irani, M., Anandan, P., Bergen, J., Kumar, R. and Hsu, S. (1996), 'Mosaic representations of video sequences and their applications', *Signal Processing: Image Communication*.

Jacobs, C., Finkelstein, A. and Salesin, D. (1995), Fast multiresolution image querying, *in* 'Proc SIGGRAPH-95', pp. 277–285.

Jacobs, D., Belhumeur, P. and Basri, R. (1998), Comparing images under variable illumination, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 610–617.

Jacobsen, A. and Gilchrist, A. (1988), 'The ratio principle holds over a million-to-one range of illumination', *Perception and Psychophysics* **43**, 1–6.

Jain, A. and Vailaya, A. (1998), 'Shape-based retrieval: a case study with trademark image databases', *Pattern Recognition* **31**(9), 1369–1390.

Janesick, J., Elliott, T., Collins, S., Blouke, M. and Freeman, J. (1987), 'Scientific charge-coupled devices', *Optical Engineering* **26**, 692–714.

Jarvis, R. (1983), 'A perspective on range finding techniques in computer vision', *IEEE Trans. Patt. Anal. Mach. Intell.* **5**(2), 122–139.

Jelinek, F. (1999), *Statistical Methods for Speech Recognition (Language, Speech and Communication)*, MIT Press.

Jepson, A. and Black, M. (1993), Mixture models for optical flow computation, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1993, pp. 760–761.

Johnson, A. and Hebert, M. (1998), 'Surface matching for object recognition in complex three-dimensional scenes', *Image and Vision Computing* **16**, 635–651.

Johnson, A. and Hebert, M. (1999), 'Using spin images for efficient object recognition in cluttered 3D scenes', *IEEE Trans. Patt. Anal. Mach. Intell.* **21**(5), 433–449.

Jones, M. and Rehg, J. (1999), Statistical color models with application to skin detection, *in* 'IEEE Confer-

ence on Computer Vision and Pattern Recognition', 1999, pp. I:274–280.

Joshi, T., Ahuja, N. and Ponce, J. (1999), 'Structure and motion estimation from dynamic silhouettes under perspective projection', *Int. J. of Comp. Vision* **31**(1), 31–50.

Judd, D. (1940), 'Hue, saturation and lightness of surface colors with chromatic illumination', *Journal of the Optical Society of America* **30**(1), 2–32.

Julesz, B. (1960), 'Binocular depth perception of computer-generated patterns', *The BellSystem Technical Journal* **39**(5), 1125–1162.

Julesz, B. (1971), *Foundations of Cyclopean Perception*, The University of Chicago Press.

Julez, B. (1959), 'A method of coding tv signals based on edge detection', *Bell System Tech. J.* **38**(4), 1001–1020.

Kakarala, R. and Hero, A. (1992), 'On achievable accuracy in edge localization', *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**(7), 777–781.

Kanade, T. (1973), Picture processing by computer complex and recognition of human faces, Technical report, Kyoto University, Dept. of Information Science.

Kanade, T. (1981), 'Recovery of the three-dimensional shape of an object from a single view', *Artificial Intelligence Journal* **17**, 409–460.

Kanade, T., Rander, P. and Narayanan, J. (1997), 'Virtualized reality: Constructing virtual worlds from real scenes', *IEEE Multimedia* **4**(1), 34–47.

Kanatani, K. (1998), 'Geometric information criterion for model selection', *International Journal of Computer Vision* **26**(3), 171–189.

Kanizsa, G. (1976), 'Subjective contours', *Scientific American*.

Kanizsa, G. (1979), *Organization in Vision: Essays on Gestalt Perception*, Praeger.

Karasaridis, A. and Simoncelli, E. (1996), A filter design technique for steerable pyramid image transforms, *in* 'International Conference on Acoustics, Speech and Signal Processing', pp. 2387–90.

Kato, T. and Fujimura, K. (1990), 'Trademark: Multimedia image database system with intelligent human interface', *Systems and Computers in Japan* **21**(11), 33–45.

Kato, T., Shimogaki, H., Mizutori, T. and Fujimura, K. (1988), Trademark: Multimedia database with abstracted representation on knowledge base, *in* 'Proc. Second Int Symp on Interoperable Information Systems', pp. 245–252.

Kelly, R., McConnell, P. and Mildenberger, S. (1977), 'The Gestalt photomapping system', *Photogrammetric Engineering and Remote Sensing* **43**(11), 1407–1417.

Keren, D., Cooper, D. and Subrahmonia, J. (1994), 'Describing complicated objects by implicit polynomials', *IEEE Trans. Patt. Anal. Mach. Intell.* **16**(1), 38–53.

Kergosien, Y. (1981), 'La famille des projections orthogonales d'une surface et ses singularités', *C.R. Acad. Sc. Paris* **292**, 929–932.

Kimia, B., Tannenbaum, A. and Zucker, S. (1990), Toward a computational theory of shape: An overview, *in* 'European Conference on Computer Vision', 1990, pp. 402–407.

Kimia, B., Tannenbaum, A. and Zucker, S. (1995), 'Shapes, shocks, and deformations i: The components of 2-dimensional shape and the reaction-diffusion space', *International Journal of Computer Vision* **15**(3), 189–224.

Kinoshita, K. and Lindenbaum, M. (2000), Camera model selection based on geometric AIC, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 2000, pp. II:514–519.

Klinker, G., Shafer, S. and Kanade, T. (1987*a,b*), Using a color reflection model to separate highlights from object color, *in* 'Proceedings, First International Conference on Computer Vision', pp. 145–150.

Klinker, G., Shafer, S. and Kanade, T. (1990), 'A physical approach to color image understanding', *International Journal of Computer Vision* **4**(1. January 1990), 7–38.

Koenderink, J. (1984), 'What does the occluding contour tell us about solid shape?', *Perception* **13**, 321–330.

Koenderink, J. (1986), An internal representation for solid shape based on the topological properties of the apparent contour, *in* W. Richards and S. Ullman, eds, 'Image Understanding: 1985-86', Ablex

Publishing Corp., chapter 9, pp. 257–285.

Koenderink, J. (1990), *Solid Shape*, MIT Press.

Koenderink, J. and Van Doorn, A. (1976*a*), 'Geometry of binocular vision and a model for stereopsis', *Biological Cybernetics* **21**, 29–35.

Koenderink, J. and Van Doorn, A. (1976*b*), 'The singularities of the visual mapping', *Biological Cybernetics* **24**, 51–59.

Koenderink, J. and Van Doorn, A. (1979), 'The internal representation of solid shape with respect to vision', *Biological Cybernetics* **32**, 211–216.

Koenderink, J. and van Doorn, A. (1980), 'Photometric invariants related to solid shape', *Optica Acta* **27**(7), 981–996.

Koenderink, J. and van Doorn, A. (1983), 'Geometrical modes as a general method to treat diffuse interreflections in radiometry', *Journal of the Optical Society of America* **73**(6), 843–850.

Koenderink, J. and van Doorn, A. (1986), 'Dynamic shape', *Biological Cybernetics* **53**, 383–396.

Koenderink, J. and Van Doorn, A. (1990), 'Affine structure from motion', *J. Opt. Soc. Am. A* **8**, 377–385.

Koenderink, J. and Van Doorn, A. (1997), 'The generic bilinear calibration-estimation problem', *Int. J. of Comp. Vision* **23**(3), 217–234.

Koenderink, J., van Doorn, A., Dana, K. and Nayar, S. (1999), 'Bidirectional reflection distribution function of thoroughly pitted surfaces', *International Journal of Computer Vision* **31**(2/3), 129–144.

Kofakis, P. and Orphanoudakis, S. (1991), Graphical tools and retrieval strategies for medical image databases, *in* 'Proceedings of the International Symposium on Computer Assisted Radiology', Springer-Verlag, pp. 519–524.

Koffka, K. (1935), *Principles of Gestalt Psychology*, Harcourt Brace.

Kriegman, D. and Belhumeur, P. (1998), What shadows reveal about object structure, *in* 'European Conference on Computer Vision', 1998, pp. 399–414.

Kriegman, D. and Ponce, J. (1990*a*), 'Computing exact aspect graphs of curved objects: solids of revolution', *Int. J. of Comp. Vision* **5**(2), 119–135.

Kriegman, D. and Ponce, J. (1990*b*), 'On recognizing and positioning curved 3-D objects from image contours', *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**(12), 1127–1137.

Krinov, E. (1947), Spectral reflectance properties of natural formations, Technical report, National Research Council of Canada, Technical Translation: TT-439.

Kruppa, E. (1913), 'Zur ermittung eines objektes aus zwei perspektiven mit innerer orientierung', *Sitz.-Ber. Akad. Wiss., Wien, Math. Naturw. Kl., Abt. Ila.* **122**, 1939–1948.

Kube, P. and Perona, P. (1996), 'Scale-space properties of quadratic feature-detectors', *IEEE Trans. Pattern Analysis and Machine Intelligence* **18**(10), 987–999.

Kutulakos, K. and Seitz, S. (1999), A theory of shape by space carving, *in* 'Proc. Int. Conf. Comp. Vision', pp. 307–314.

Kutulakos, K. and Vallino, J. (1998), 'Calibration-free augmented reality', *IEEE Transactions on Visualization and Computer Graphics* **4**(1), 1–20.

Lamb, T. and Bourriau, J., eds (1995), *Colour Art and Science*, Cambridge University Press.

Lamdan, Y., Schwartz, J. and Wolfson, H. (1990), 'Affine invariant model-based object recognition', *IEEE Trans. Robotics and Automation* **6**, 578–589.

Land, E. (1959*a*), 'Color vision and the natural image: Part i', *Proceedings National Academy Science USA* **45**(1), 115–129.

Land, E. (1959*b*), 'Color vision and the natural image: Part ii', *Proceedings National Academy Science USA* **45**(4), 636–644.

Land, E. (1959*c*), 'Experiments in color vision', *Scientific American* **200**, 84–89.

Land, E. (1983), 'Color vision and the natural image', *Proceedings National Academy Science USA* **80**, 5163–5169.

Land, E. and McCann, J. (1971), 'Lightness and retinex theory', *Journal of the Optical Society of America* **61**(1), 1–11.

Laurentini, A. (1995), 'How far 3D shapes can be understood from 2D silhouettes', *IEEE Trans. Patt. Anal. Mach. Intell.* **17**(2), 188–194.

Lavallée, S. (1996), Registration for computer integrated surgery: Methodology and state of the art, *in* R. Taylor, S. Lavallée, G. Burdea and R. Mosges, eds, 'Computer Integrated Surgery', MIT Press.

Laveau, S. and Faugeras, O. (1994), 3D scene representation as a collection of images and fundamental matrices, Technical Report 2205, INRIA Sophia-Antipolis.

Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998), 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE* **86**(11), 2278–2324.

Lee, H. (1986), 'Method for computing the scene-illuminant chromaticity from specular highlights', *J. Opt. Soc. Am.-A* **3**, 1694–1699.

Lee, S. and Bajcsy, R. (1992*a*), Detection of specularity using colour and multiple views, *in* Sandini (1992), pp. 99–114.

Lee, S. and Bajcsy, R. (1992*b*), 'Detection of specularity using colour and multiple views', *Image and Vision Computing* **10**, 643–653.

Leung, T. and Malik, J. (1997), On perpendicular texture or: Why do we see more flowers in the distance?, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 807–813.

Leung, T., Burl, M. and Perona, P. (1995), Finding faces in cluttered scenes using labelled random graph matching, *in* 'Proceedings, Fifth International Conference on Computer Vision', 1995, pp. 637–644.

Levoy, M. and Hanrahan, P. (1996), Light field rendering, *in* 'SIGGRAPH', pp. 31–42.

Lim, H. and Binford, T. (1988), Curved surface reconstruction using stereo correspondence, *in* 'Proc. DARPA Image Understanding Workshop', pp. 809–819.

Lipson, P., Grimson, W. L. and Sinha, P. (1997), Configuration based scene classification and image indexing, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 1007–1013.

Ljung, L. (1995), System identification, *in* W. S. Levine, ed., 'The Control Handbook', CRC Press, in cooperation with IEEE Press.

Longuet-Higgins, H. (1981), 'A computer algorithm for reconstructing a scene from two projections', *Nature* **293**, 133–135.

Loop, C. (1994), 'Smooth spline surfaces over irregular meshes', *Computer Graphics* pp. 303–310.

Lorensen, W. and Cline, H. (1987), 'Marching cubes: a high resolution 3D surface construction algorithm', *Computer Graphics* **21**, 163–169.

Lowe, D. (1985), *Perceptual Organization and Visual Recognition*, Kluwer.

Luenberger, D. (1984), *Linear and Nonlinear Programming*, Addison-Wesley. Second edition.

Luong, Q.-T. (1992), Matrice fondamentale et calibration visuelle sur l'environnement: vers une plus grande autonomie des systèmes robotiques, PhD thesis, University of Paris XI, Orsay, France.

Luong, Q.-T. and Faugeras, O. (1996), 'The fundamental matrix: theory, algorithms, and stability analysis', *Int. J. of Comp. Vision* **17**(1), 43–76.

Luong, Q.-T., Deriche, R., Faugeras, O. and Papadopoulo, T. (1993), On determining the fundamental matrix: analysis of different methods and experimental results, Technical Report 1894, INRIA Sophia-Antipolis.

Lynch, D. and Livingston, W. (2001), *Color and Light in Nature*, Cambridge University Press.

Lyvers, E. and Mitchell, O. (1988), 'Precision edge contrast and orientation estimation', *IEEE Trans. Pattern Analysis and Machine Intelligence* **10**(6), 927–937.

Ma, W. and Manjunath, B. (1995), A comparison of wavelet features for texture annotation, *in* 'Proceedings, International Conference on Image Processing', 1995, pp. II: 256–259.

Ma, W. and Manjunath, B. (1996), Texture features and learning similarity, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1996, pp. 425–430.

Ma, W. and Manjunath, B. (1997*a*), Netra: a toolbox for navigating large image databases, *in* 'IEEE Int. Conf. Image Processing', pp. 568–571.

Ma, W. and Manjunath, B. (1998), 'A texture thesaurus for browsing large aerial photographs', *Journal of the American Society for Information Science (special issue on AI Techniques for Emerging Information Systems Applications)*49(7)633–648.

Ma, W. Y. and Manjunath, B. S. (1997*b*), Edgeflow: a framework for boundary detection and image segmentation, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 744–749.

MacAdam, D. (1942), 'Visual sensitivities to small color differences in daylight', *Journal of the Optical Society of America* **32**, 247.

Macaulay, F. (1916), *The Algebraic Theory of Modular Systems*, Cambridge University Press.

Mahamud, S. and Hebert, M. (2000), Iterative projective reconstruction from multiple views, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. II–430–437.

Mahamud, S., Hebert, M., Omori, Y. and Ponce, J. (2001), Provably-convergent iterative methods for projective structure from motion, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 1018–1025.

Maintz, J. and Viergever, M. (1998), 'A survey of medical image registration', *Medical Image Analysis* **2**(1), 1–16.

Malik, J. and Perona, P. (1989), A computational model of texture segmentation, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1989, pp. 326–332.

Malik, J. and Perona, P. (1990), 'Preattentive texture discrimination with early visual mechanisms', *J. Opt. Soc. America* **7A**(5), 923–932.

Malik, J. and Rosenholtz, R. (1997), 'Computing local surface orientation and shape from texture for curved surfaces', *Int. J. Computer Vision* pp. 149–168.

Maloney, L. (1984), Computational Approaches to Color Vision, PhD thesis, Stanford University.

Maloney, L. (1986), 'Evaluation of linear models of surface spectral reflectance with small numbers of parameters', *Journal of the Optical Society of America* **3**(10), 1673–1683.

Maloney, L. and Wandell, B. (1986), 'Color constancy: A method for recovering surface spectral reflectance', *Journal of the Optical Society of America* **3**, 29–33.

Manjunath, B. and Chellappa, R. (1991), 'Unsupervised texture segmentation using markov random field models', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(5), 478–482.

Manjunath, B. and Ma, W. (1996*a*), Browsing large satellite and aerial photographs, *in* 'IEEE Int. Conf. Image Processing'.

Manjunath, B. and Ma, W. (1996*b,c*), 'Texture features for browsing and retrieval of image data', *IEEE Trans. Pattern Analysis and Machine Intelligence* **18**(8), 837–842.

Manning, C. and Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*, MIT Press.

Manocha, D. (1992), Algebraic and Numeric Techniques for Modeling and Robotics, PhD thesis, Computer Science Division, Univ. of California, Berkeley.

Marimont, D. and Wandell, B. (1992), 'Linear models of surface and illuminant spectra', *J. Opt. Soc. Am.-A* **9**, 1905–1913.

Marr, D. (1977), 'Analysis of occluding contour', *Proc. Royal Society, London* **B-197**, 441–475.

Marr, D. (1982), *Vision*, Freeman.

Marr, D. and Hildreth, E. (1980), 'Theory of edge detection', *Proceedings of Royal Society of London* **B-207**, 187–217.

Marr, D. and Nishihara, K. (1978), 'Representation and recognition of the spatial organization of three-dimensional shapes', *Proc. Royal Society, London* **B-200**, 269–294.

Marr, D. and Poggio, T. (1976), 'Cooperative computation of stereo disparity', *Science* **194**, 283–287.

Marr, D. and Poggio, T. (1979), 'A computational theory of human stereo vision', *Proceedings of the Royal Society of London* **B 204**, 301–328.

Martin, W. and Aggarwal, J. (1983), 'Volumetric description of objects from multiple views', *IEEE Trans. Patt. Anal. Mach. Intell.* **5**(2), 150–158.

Maxwell, B. and Shafer, S. (2000), 'Segmentation and interpretation of multicolored objects with highlights', *Computer Vision and Image Understanding* **77**(1), 1–24.

Maybank, S. and Faugeras, O. (1992), 'A theory of self-calibration of a moving camera', *Int. J. of Comp. Vision* **8**(2), 123–151.

Maybank, S. and Sturm, P. (1999), MDL, collineations and the fundamental matrix, *in* 'British Machine

Vision Conference'.

McCann, J., McKee, S. and Taylor (1976), 'Quantitative studies in retinex theory', *Vision Research* **16**, 445–458.

McInerney, T. and Terzopolous, D. (1996), 'Deformable models in medical image analysis: a survey', *Medical Image Analysis* **1**(2), 91–108.

McKee, S., Levi, D. and Brown, S. (1990), 'The imprecision of stereopsis', *Vision Research* **30**(11), 1763–1779.

McLachlan, G. and Krishnan, T. (1996), *The EM Algorithm and Extensions*, John Wiley & Sons.

McMillan, L. and Bishop, G. (1995), Plenoptic modeling: an image-based rendering approach, *in* 'SIGGRAPH', pp. 39–46.

Medioni, G. and Nevatia, R. (1984), 'Matching images using linear features', *IEEE Trans. Patt. Anal. Mach. Intell.* **6**(6), 675–685.

Meer, P., Mintz, D., Kim, D. and Rosenfeld, A. (1991), 'Robust regression methods for computer vision: A review', *International Journal of Computer Vision* **6**(1), 59–70.

Milenkovic, V. and Kanade, T. (1985), Trinocular vision using photometric and edge orientation constraints, *in* 'Proc. DARPA Image Understanding Workshop', pp. 163–175.

Minnaert, M. (1993), *Light and Color in the Outdoors*, Springer Verlag. Translator: L. Seymour.

Mohan, R. and Nevatia, R. (1989), Segmentation and description based on perceptual organization, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1989, pp. 333–341.

Mohan, R. and Nevatia, R. (1992), 'Perceptual organization for scene segmentation and description', *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**(6), 616–635.

Mohr, R., Morin, L. and Grosso, E. (1992), Relative positioning with uncalibrated cameras, *in* J. Mundy and A. Zisserman, eds, 'Geometric Invariance in Computer Vision', MIT Press, pp. 440–460.

Mollon, J. (1995), Seeing colour, *in* T. Lamb and J. Bourriau, eds, 'Colour Art and Science', Cambridge University Press.

Montel, P., ed. (1972), *Toute la photographie*, Librairie Larousse and Publications Montel.

Morgan, A. (1987), *Solving Polynomial Systems using Continuation for Engineering and Scientific Problems*, Prentice Hall.

Morita, T. and Kanade, T. (1997), 'A sequential factorization method for recovering shape and motion from image sequences', *IEEE Trans. Patt. Anal. Mach. Intell.*

Mukawa, N. (1990), Estimation of shape, reflection coefficients and illuminant direction from image sequences, *in* 'Proceedings, Third International Conference on Computer Vision', 1990, pp. 507–512.

Mumford, D. and Shah, J. (1985), Boundary detection by minimizing functionals, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', IEEE Press, pp. 22–26.

Mumford, D. and Shah, J. (1988), 'Optimal approximations by piecewise smooth functions and variational problems', *Communications on Pure and Applied Mathematics* **XLII**(5), 577–685.

Mundy, J. and Vrobel, P. (1994), The role of IU technology in RADIUS phase ii, *in* 'Proc. Image Understanding Workshop', pp. 251–264.

Mundy, J. and Zisserman, A. (1992), *Geometric Invariance in Computer Vision*, MIT Press.

Mundy, J., Zisserman, A. and Forsyth, D. (1993), *Applications of Invariance in Computer Vision*, Springer-Verlag.

Nalwa, V. (1987), Line-drawing interpretation: bilateral symmetry, *in* 'Proc. DARPA Image Understanding Workshop', pp. 956–967.

Nalwa, V. (1988), 'Line-drawing interpretation: A mathematical framework', *Int. J. of Comp. Vision* **2**, 103–124.

Nathans, J., Piantanida, T., Eddy, R., Shows, T. and Hogness, D. (1986*a*), 'Molecular genetics of inherited variation in human color vision', *Science* **232**, 203–210.

Nathans, J., Thomas, D. and Hogness, D. (1986*b*), 'Molecular genetics of human color vision: The genes encoding blue, green, and red pigments', *Science* **232**, 193–203.

Navy, U. (1969), *Basic Optics and Optical Instruments*, Dover. Prepared by the Bureau of Naval Personnel.

Nayar, S. (1997), Catadioptric omnidirectional camera, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 482–488.

Nayar, S. and Oren, M. (1993), Diffuse reflectance from rough surfaces, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1993, pp. 763–764.

Nayar, S. and Oren, M. (1995), 'Visual appearance of matte surfaces', *Science* **267**(5201), 1153–1156.

Nayar, S., Ikeuchi, K. and Kanade, T. (1990), 'Determining shape and reflectance of hybrid surfaces by photometric sampling', *IEEE Trans. Robotics and Automation* **6**(4), 418–431.

Nayar, S., Ikeuchi, K. and Kanade, T. (1991*a*), 'Shape from interreflections', *International Journal of Computer Vision* **6**(3), 173–195.

Nayar, S., Ikeuchi, K. and Kanade, T. (1991*b*), 'Surface reflection: Physical and geometrical perspectives', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(7), 611–634.

Nevatia, R. (1986), Image segmentation, *in* K. Fu and T. Young, eds, 'Handbook of Pattern Recognition and Image Processing', Academic Press, pp. 215–231.

Nevatia, R. and Binford, T. (1977), 'Description and recognition of complex curved objects', *Artificial Intelligence Journal* **8**, 77–98.

Nielsen, M., Johansen, P., Olsen, O. F. and Weickert, J., eds (1999), *Scale-Space Theory in Computer Vision*, Vol. 1682, Springer Verlag LNCS.

Niem, W. and Buschmann, R. (1994), Automatic modelling of 3D natural objects from multiple views, *in* 'European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production'.

Nitzan, D. (1988), 'Three-dimensional vision structure for robot applications', *IEEE Trans. Patt. Anal. Mach. Intell.* **10**(3), 291–309.

Noble, A., Wilson, D. and Ponce, J. (1997), 'On Computing Aspect Graphs of Smooth Shapes from Volumetric Data', *Computer Vision and Image Understanding: special issue on Mathematical Methods in Biomedical Image Analysis* **66**(2), 179–192.

Ogle, V. and Stonebraker, M. (1995), 'Chabot: retrieval from a relational database of images', *Computer* **28**, 40–48.

Ohlander, R., Price, K. and Reddy, R. (1978), 'Picture segmentation by a recursive region splitting method', *Computer Graphics Image Processing* **8**, 313–333.

Ohta, Y. and Kanade, T. (1985), 'Stereo by intra- and inter-scanline search', *IEEE Trans. Patt. Anal. Mach. Intell.* **7**(2), 139–154.

Ohta, Y., Maenobu, K. and Sakai, T. (1981), Obtaining surface orientation from texels under perspective projection, *in* 'Proc. International Joint Conference on Artificial Intelligence', pp. 746–751.

Oja, E. (1983), *Subspace methods of pattern recognition*, Research Study Press.

Okutami, M. and Kanade, T. (1993), 'A multiple-baseline stereo system', *IEEE Trans. Patt. Anal. Mach. Intell.* **15**(4), 353–363.

Olson, C. (1998), Variable-scale smoothing and edge detection guided by stereoscopy, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 80–85.

Oren, M. and Nayar, S. (1995), 'Generalization of the lambertian model and implications for machine vision', *International Journal of Computer Vision* **14**(3), 227–251.

Oren, M., Papageorgiou, C., Sinha, P., Osuna, E. and Poggio, T. (1997), Pedestrian detection using wavelet templates, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 193–199.

Osuna, E., Freund, R. and Girosi, F. (1997), Training support vector machines: An application to face detection, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 130–136.

Pae, S. and Ponce, J. (1999), Toward a scale-space aspect graph: Solids of revolution, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', Vol. II, pp. 196–201.

Pae, S. and Ponce, J. (2001), 'On computing structural changes in evolving surfaces and their appearance', *Int. J. of Comp. Vision* **43**(2), 113–131.

Pal, N. and Pal, S. (1993), 'A review on image segmentation techniques', *Pattern Recognition* **26**(9), 1277–1294.

Palmer, S. (1999), *Vision Science : Photons to Phenomenology*, MIT Press.

Papageorgiou, C. and Poggio, T. (1999), A pattern classification approach to dynamical object detection, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 1223–1228.

Papageorgiou, C. and Poggio, T. (2000), 'A trainable system for object detection', *International Journal of Computer Vision* **38**(1), 15–33.

Papageorgiou, C., Oren, M. and Poggio, T. (1998), A general framework for object detection, *in* 'Proceedings, Sixth International Conference on Computer Vision', 1998, pp. 555–562.

Park, J., Seo, J., An, D. and Chung, S. (2000), Detection of human faces using skin color and eyes, *in* 'Multimedia and Exposition', pp. 133–136.

Pentland, A. (1986), 'Perceptual organization and the representation of natural form', *Artificial Intelligence Journal* **28**, 293–331.

Pentland, A., Picard, R. and Sclaroff, S. (1996), 'Photobook: content-based manipulation of image databases', *Int. J. Computer Vision* **18**(3), 233–254.

Peri, V. and Nayar, S. (1997), Generation of perspective and panoramic video from omnidirectional video, *in* 'Proc. DARPA Image Understanding Workshop'.

Perlin, K. (1985), An image synthesizer, *in* 'SIGGRAPH-85', pp. 287–296.

Perona, P. (1992), Steerable-scalable kernels for edge detection and junction analysis, *in* Sandini (1992), pp. 3–18.

Perona, P. (1995), 'Deformable kernels for early vision', *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**(5), 488–499.

Perona, P. and Freeman, W. (1998), A factorization approach to grouping, *in* 'European Conference on Computer Vision', 1998, pp. 655–670.

Perona, P. and Malik, J. (1990*a,b*), Detecting and localizing edges composed of steps, peaks and roofs, *in* 'Proceedings, Third International Conference on Computer Vision', 1990, pp. 52–57.

Perona, P. and Malik, J. (1990*c*), 'Scale-space and edge detection using anisotropic diffusion', *IEEE Trans. Patt. Anal. Mach. Intell.* **12**(7), 629–639.

Petitjean, S. (1995), Géométrie énumérative et contacts de variétés linéaires: application aux graphes d'aspects d'objets courbes, PhD thesis, Institut National Polytechnique de Lorraine.

Petitjean, S., Ponce, J. and Kriegman, D. (1992), 'Computing exact aspect graphs of curved objects: Algebraic surfaces', *Int. J. of Comp. Vision* **9**(3), 231–255.

Phillips, P. and Vardi, Y. (1996), 'Efficient illumination normalization of facial images', *Pattern Recognition Letters* **17**(8), 921–927.

Plantinga, H. and Dyer, C. (1990), 'Visibility, occlusion, and the aspect graph', *Int. J. of Comp. Vision* **5**(2), 137–160.

Platonova, O. (1981), 'Singularities of the mutual disposition of a surface and a line', *Russian Mathematical Surveys* **36**(1), 248–249.

Poelman, C. and Kanade, T. (1997), 'A paraperspective factorization method for shape and motion recovery', *IEEE Trans. Patt. Anal. Mach. Intell.* **19**(3), 206–218.

Poggio, T., Torre, V. and Koch, C. (1985), 'Computational vision and regularization theory', *Nature* **317**, 314–319.

Pollard, S., Mayhew, J. and Frisby, J. (1970), 'A stereo correspondence algorithm using a disparity gradient limit', *Perception* **14**, 449–470.

Pollefeys, M. (1999), Self-calibration and metric 3D reconstruction from uncalibrated image sequences, PhD thesis, Katholieke Universiteit Leuven.

Pollefeys, M., Koch, R. and Van Gool, L. (1999), 'Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters', *Int. J. of Comp. Vision* **32**(1), 7–26.

Ponce, J. (1990), 'On characterizing ribbons and finding skewed symmetries', *Computer Vision, Graphics and Image Processing* **52**, 328–340.

Ponce, J. (2000), Metric upgrade of a projective reconstruction under the rectangular pixel assumption, *in* 'Second Workshop on Structure from Multiple Images of Large Scale Environments', pp. 18–27. Preprints.

Ponce, J. and Brady, J. (1987), Toward a surface primal sketch, *in* T. Kanade, ed., 'Three-dimensional

machine vision', Kluwer Publishers, pp. 195–240.

Ponce, J., Cass, T. and Marimont, D. (1993), Relative stereo and motion reconstruction, Technical Report UIUC-BI-AI-RCV-93-07, Beckman Institute, University of Illinois.

Ponce, J., Cepeda, M., Pae, S. and Sullivan, S. (1999), Shape models and object recognition, *in* D. Forsyth, J. Mundy, V. Gesu and R. Cipolla, eds, 'Shape, Contour and Grouping in Computer Vision', Vol. 1681 of *Lecture Notes in Computer Science*, Springer-Verlag.

Ponce, J., Chelberg, D. and Mann, W. (1989), 'Invariant properties of straight homogeneous generalized cylinders and their contours', *IEEE Trans. Patt. Anal. Mach. Intell.* **11**(9), 951–966.

Porrill, J. (1990), 'Fitting ellipses and predicting confidence envelopes using a bias corrected kalman filter', *Image and Vision Computing* **8**(1), 37–41.

Pritchett, P. and Zisserman, A. (1998), Wide baseline stereo matching, *in* 'Proc. Int. Conf. Comp. Vision', pp. 754–760.

Psarrou, A., Konstantinou, V., Morse, P. and O'Reilly, P. (1997), Content based search in mediaeval manuscripts, *in* 'TENCON-97 - Proc. IEEE Region 10 Conf. Speech and Image Technologies for Computing and Telecommunications', pp. 187–190.

Raja, Y., McKenna, S. and Gong, S. (1998), Colour model selection and adaptation in dynamic scenes, *in* 'European Conference on Computer Vision', 1998, pp. 460–474.

Ranade, S. and Prewitt, J. (1980), A comparison of some segmentation algorithms for cytology, *in* 'Proceedings, International Conference on Pattern Recognition', pp. 561–564.

Rieger, J. (1987), 'On the classification of views of piecewise-smooth objects', *Image and Vision Computing* **5**, 91–97.

Rieger, J. (1990), 'The geometry of view space of opaque objects bounded by smooth surfaces', *Artificial Intelligence Journal* **44**(1-2), 1–40.

Rieger, J. (1992), 'Global bifurcations sets and stable projections of non-singular algebraic surfaces', *Int. J. of Comp. Vision* **7**(3), 171–194.

Ripley, B. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press.

Riseman, E. and Arbib, M. (1977), 'Computational techniques in the visual segmentation of static scenes', *Computer Graphics Image Processing* **6**(3), 221–276.

Rissanen, J. (1983), 'A universal prior for integers and estimation by minimum description length', *Annals of Statistics* **11**, 416–431.

Rissanen, J. (1987), 'Stochastic complexity (with discussion)', *J. Roy. Stat. Soc. Series B* **49**, 223–239.

Robert, L. and Faugeras, O. (1991), Curve-based stereo: figural continuity and curvature, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 57–62.

Roberts, L. (1965), Machine perception of 3-D solids, *in* J. T. *et al.*, ed., 'Optical and Electro-Optical Information Processing', MIT Press, pp. 159–197.

Rosch, E. (1988), Principles of categorisation, *in* A. Collins and E. Smith, eds, 'Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence', Morgan Kauffman, pp. 312–322.

Rosenfeld, A. (1986), 'Axial representations of shape', *Computer Vision, Graphics and Image Processing* **33**, 156–173.

Rosenholtz, R. and Malik, J. (1997), 'Surface orientation from texture: isotropy or homogeneity (or both)?', *Vision Research* **37**(16), 2283–2293.

Rothwell, C. (1995), *Object Recognition through Invariant Indexing*, Oxford University Press.

Rothwell, C., Forsyth, D., Zisserman, A. and Mundy, J. (1993), Extracting projective structure from single perspective views of 3d point sets, *in* 'Proceedings, Fourth International Conference on Computer Vision', pp. 573–582.

Rothwell, C., Zisserman, A., Forsyth, D. and Mundy, J. (1995), 'Planar object recognition using projective shape representation', *International Journal of Computer Vision* **16**(1), 57–99.

Rothwell, C., Zisserman, A., Marinos, C., Forsyth, D. and Mundy, J. (1992), 'Relative motion and pose from arbitrary plane curves', *Image and Vision Computing* **10**, 250–262.

Rousseeuw, P. (1987), *Robust Regression and Outlier Detection*, Wiley.

Rowley, H., Baluja, S. and Kanade, T. (1996), Neural network-based face detection, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1996, pp. 203–208.

Rowley, H., Baluja, S. and Kanade, T. (1998*a*), 'Neural network-based face detection', *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**(1), 23–38.

Rowley, H., Baluja, S. and Kanade, T. (1998*b*), Rotation invariant neural network-based face detection, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 38–44.

Rowley, H., Baluja, S. and Kanade, T. (1998*c*), Rotation invariant neural network-based face detection, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 38–44.

Rubner, Y., Tomasi, C. and Guibas, L. J. (1998), A metric for distributions with applications to image databases, *in* 'Int. Conf. on Computer Vision', pp. 59–66.

Russ, J. (1995), *The Image Processing Handbook*, CRC Press. Second edition.

Sabin, M. (1994), *Acta Numerica 1994*, Cambridge University Press, chapter Numerical Geometry of Surfaces.

Saito, H., Baba, S., Kimura, M., Vedula, S. and Kanade, T. (1999), Appearance-based virtual view generation of temporally-varying events from multi-camera images in the 3D room, Technical Report CMU-CS-99-127, Carnegie-Mellon University.

Samuel, P. (1988), *Projective Geometry*, Springer-Verlag. English translation of "Géométrie Projective", Presses Universitaires de France, 1986.

Sandini, G., ed. (1992), *European Conference on Computer Vision*, Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag.

Sarachik, K. and Grimson, W. (1993), Gaussian error models for object recognition, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1993, pp. 400–406.

Sarkar, S. and Boyer, K. (1993), 'Integration, inference, and management of spatial information using Bayesian networks: Perceptual organization', *IEEE Trans. Pattern Analysis and Machine Intelligence* **15**(3), 256–274.

Sarkar, S. and Boyer, K. (1994), *Computing Perceptual Organization in Computer Vision*, World Scientific.

Sarkar, S. and Boyer, K. (1998), 'Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors', *Computer Vision and Image Understanding* **71**(1), 110–136.

Saund, E. and Moran, T. (1995), Perceptual organization in an interactive sketch editing application, *in* 'Proceedings, Fifth International Conference on Computer Vision', 1995, pp. 597–604.

Sawhney, H. and Ayer, S. (1996), 'Compact representations of videos through dominant and multiple motion estimation', *IEEE T. Pattern Analysis and Machine Intelligence* **18**(8), 814–830.

Schmid, C. and Mohr, R. (1997*a,b*), 'Local grayvalue invariants for image retrieval', *IEEE Trans. Patt. Anal. Mach. Intell.* **19**(5), 530–535.

Schmid, C., Mohr, R. and Bauckhage, C. (2000), 'Evaluation of interest point detectors', *International Journal of Computer Vision* **37**(2), 151–172.

Schneiderman, H. and Kanade, T. (1998), Probabilistic formulation for object recognition, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 45–51.

Seitz, S. and Dyer, C. (1995), Physically-valid view synthesis by image interpolation, *in* 'Workshop on Representations of Visual Scenes'.

Seitz, S. and Dyer, C. (1996), Toward image-based scene representation using view morphing, Technical Report 1298, University of Wisconsin.

Serra, J. (1982), *Image Analysis and Mathematical Morphology*, Academic Press.

Shade, J., Gortler, S., Li-wei, H. and Szeliski, R. (1998), Layered depth images, *in* 'SIGGRAPH 98', pp. 231–242.

Shafer, S. (1985*a*), *Shadows and Silhouettes in Computer Vision*, Kluwer Academic Publishers.

Shafer, S. (1985*b*), 'Using color to separate reflection components', *Color Res. App.* **10**(4), 210–218.

Shafer, S. and Kanade, T. (1983), The theory of straight homogeneous generalized cylinders and a taxonomy of generalized cylinders, Technical Report CMU-CS-83-105, Carnegie-Mellon University.

Shashua, A. (1993), Projective depth: a geometric invariant for 3D reconstruction from two perspec-

tive/orthographic views and for visual recognition, *in* 'Proc. Int. Conf. Comp. Vision', pp. 583–590.

Shashua, A. (1995), 'Algebraic functions for recognition', *IEEE Trans. Patt. Anal. Mach. Intell.* **17**(8), 779–789.

Shi, J. and Malik, J. (1997), Normalized cuts and image segmentation, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 731–737.

Shi, J. and Malik, J. (1998*a*), Motion segmentation and tracking using normalized cuts, *in* 'Proceedings, Sixth International Conference on Computer Vision', 1998, pp. 1154–1160.

Shi, J. and Malik, J. (1998*b*), Self-inducing relational distance and its application to image segmentation, *in* 'European Conference on Computer Vision', 1998, pp. 528–543.

Shi, J. and Malik, J. (2000), 'Normalized cuts and image segmentation', *IEEE Trans. Pattern Analysis and Machine Intelligence* **22**(8), 888–905.

Shimshoni, I. and Ponce, J. (1997), 'Finite-resolution aspect graphs of polyhedral objects', *IEEE Trans. Patt. Anal. Mach. Intell.* **19**(4), 315–327.

Shin, M., Goldgof, D. and Bowyer, K. (1998), An objective comparison methodology of edge detection algorithms using a structure from motion task, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1998, pp. 190–195.

Shin, M., Goldgof, D. and Bowyer, K. (1999), Comparison of edge detectors using an object recognition task, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1999, pp. I:360–365.

Shirai, Y. (1972), 'Recognition of polyhedrons with a range finder', *Pattern Recognition* **4**, 243–250.

Shirman, L. and Sequin, C. (1987), 'Local surface interpolation with Bezier patches', *CAGD* **4**, 279–295.

Shum, H. and Szeliski, R. (1998), Construction and refinement of panoramic mosaics with global and local alignment, *in* 'Proc. Int. Conf. Comp. Vision', pp. 953–958.

Shum, H., Ikeuchi, K. and Reddy, R. (1995), 'Principal component analysis with missing data and its application to polyhedral object modeling', *IEEE Trans. Patt. Anal. Mach. Intell.* **17**(9), 854–867.

Siddiqi, K., Kimia, B., Tannenbaum, A. and Zucker, S. (1999*a*), 'Shapes, shocks and wiggles', *Image and Vision Computing* **17**(5/6), 365–373.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J. and Zucker, S. W. (1999*b*), 'Shock graphs and shape matching', *Int. J. of Comp. Vision* **35**(1), 13–32.

Sillion, F. (1994), *Radiosity and Global Illumination*, Morgan-Kauffman.

Simon, D., Hebert, M. and Kanade, T. (1994), Real-time 3D pose estimation using a high-speed range sensor, *in* 'IEEE Int. Conf. on Robotics and Automation', pp. 2235–2241.

Simoncelli, E. and Farid, H. (1995), Steerable wedge filters, *in* 'Proceedings, Fifth International Conference on Computer Vision', 1995, pp. 189–194.

Simoncelli, E. and Freeman, W. (1995), The steerable pyramid: A flexible architecture for multi-scale derivative computation, *in* 'Proceedings, International Conference on Image Processing', 1995, pp. 444–447.

Sirovitch, L. and Kirby, M. (1987), 'Low-dimensional procedure for the characterization of human faces', *J. Opt. Soc. Am. A* **2**, 586–591.

Slama, C., Theurer, C. and Henriksen, S., eds (1980), *Manual of photogrammetry*, American Society of Photogrammetry. Fourth edition.

Smith, J. and Chang, S.-F. (1996), Visualseek: A fully automated content-based image query system, *in* 'ACM Multimedia Conference'.

Smith, J. and Chang, S.-F. (1997), 'Visually searching the web for content', *IEEE Multimedia* **4**(3), 12–20.

Smith, M. and Christel, M. (1995), Automating the creation of a digital video library, *in* 'ACM Multimedia'.

Smith, M. and Hauptmann, A. (1995), Text, speech and vision for video segmentation: The informedia project, *in* 'AAAI Fall 1995 Symposium on Computational Models for Integrating Language and Vision'.

Smith, M. and Kanade, T. (1997), Video skimming for quick browsing based on audio and image characterization, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition'.

Smith, T. (1996), 'A digital library for geographically referenced materials', *Computer* **29**(5), 54–60.

et al., A. S., ed. (2000), *Advances in Large Margin Classifiers*, MIT Press.

Snapper, E. and Troyer, R. (1989), *Metric Affine Geometry*, Dover Publications Inc. Reprinted from Academic Press, 1971.

Snyder, D., Hammoud, A. and White, R. (1993), 'Image recovery from data acquired with a charge-coupled-device camera', *J. Opt. Soc. Am. A* **10**(5), 1014–1023.

Song, Y., Feng, X. and Perona, P. (2000*a*), Towards detection of human motion, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 2000, pp. I:810–817.

Song, Y., Goncalves, L. and Perona, P. (2000*b*), Monocular percepton of biological motion: Clutter and partial occlusion, *in* 'European Conference on Computer Vision', 2000, pp. 719–733.

Song, Y., Goncalves, L., di Bernardo, E. and Perona, P. (1999), Monocular perception of biological motion: Detection and labeling, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 805–813.

Spacek, L. (1986), 'Edge detection and motion detection', *Image and Vision Computing* **4**(1), 43–56.

Speis, A. and Healey, G. (1996), 'Feature-extraction for texture-discrimination via random-field models with random spatial interaction', *IEEE Trans. Image Processing* **5**(4), 635–645.

Spetsakis, M. and Aloimonos, Y. (1990), 'Structure from motion using line correspondences', *Int. J. of Comp. Vision* **4**(3), 171–183.

Srivastava, S. and Ahuja, N. (1990), 'Octree generation from object silhouettes in perspective views', *Computer Vision, Graphics and Image Processing* **49**(1), 68–84.

Stark, L. and Bowyer, K. (1996), *Generic Object Recognition Using Form and Function*, World Scientific Publishing.

Starner, T., Weaver, J. and Pentland, A. (1998), 'Real-time american sign language recognition using desk and wearable computer based video.', *IEEE T. Pattern Analysis and Machine Intelligence* **20**(12), 1371–1375.

Stein, F. and Medioni, G. (1992), 'Structural indexing: efficient 3D object recognition', *IEEE Trans. Patt. Anal. Mach. Intell.*

Stewart, C. (1999), 'Robust parameter estimation in computer vision', *SIAM-Review* **41**(3), 513–537.

Stewman, J. and Bowyer, K. (1987), Aspect graphs for planar-face convex objects, *in* 'Proc. IEEE Workshop on Computer Vision', pp. 123–130.

Stewman, J. and Bowyer, K. (1988), Creating the perspective projection aspect graph of polyhedral objects, *in* 'Proc. Int. Conf. Comp. Vision', pp. 495–500.

Strang, G. (1980), *Linear Algebra and its Applications*, Academic Press, Inc. Second edition.

Struik, D. (1988), *Lectures on Classical Differential Geometry*, Dover. Reprint of the second edition (1961) of the work first published by Addison-Wesley in 1950.

Sturm, P. and Triggs, B. (1996), A factorization-based algorithm for multi-image projective structure and motion, *in* 'Proc. European Conf. Comp. Vision', pp. 709–720.

Sugihara, K. (1986), *Machine Interpretation of Line Drawings*, MIT Press.

Sullivan, G., Baker, K., Worrall, A., Attwood, C. and Remagnino, P. (1997), 'Model-based vehicle detection and classification using orthographic approximations', *Image and Vision Computing* **15**(8), 649–654.

Sullivan, S. and Ponce, J. (1998), 'Automatic model construction, pose estimation, and object recognition from photographs using triangular splines', *IEEE Trans. Patt. Anal. Mach. Intell.* **20**(10), 1091–1096.

Sullivan, S., Sandford, L. and Ponce, J. (1994*a,b*), 'Using geometric distance fits for 3D object modelling and recognition', *IEEE Trans. Patt. Anal. Mach. Intell.* **16**(12), 1183–1196.

Sung, K.-K. and Poggio, T. (1998), 'Example-based learning for view-based human face detection', *IEEE T. Pattern Analysis and Machine Intelligence* **20**, 39–51.

Swain, M. and Ballard, D. (1991), 'Color indexing', *Int. J. Computer Vision* **7**(1), 11–32.

Szeliski, R., Avidan, S. and Anandan, P. (2000), Layer extraction from multiple images containing reflections and transparency, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 2000, pp. I:246–253.

Tagare, H. and de Figueiredo, R. (1991), 'A theory of photometric stereo for a class of diffuse non-

Lambertian surfaces', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(2), 133–152.

Tagare, H. and de Figueiredo, R. (1992), 'Simultaneous estimation of shape and reflectance map from photometric stereo', *CVGIP: Image Understanding* **55**(3), 275–286.

Tagare, H. and de Figueiredo, R. (1993), 'A framework for the construction of reflectance maps for machine vision', *CVGIP: Image Understanding* **57**(3), 265–282.

Tao, H., Sawhney, H. and Kumar, R. (2000), Dynamic layer representation with applications to tracking, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 2000, pp. II:134–141.

Tarr, M., Hayward, W., Gauthier, I. and Williams, P. (1995), 'Is object recognition mediated by viewpoint invariant parts or viewpoint dependent features', *Perception* **24**, 4.

Taubin, G., Cukierman, F., Sullivan, S., Ponce, J. and Kriegman, D. (1994*a,b*), 'Parameterized families of polynomials for bounded algebraic and surface curve fitting', *IEEE Trans. Patt. Anal. Mach. Intell.* **16**(3), 287–303.

Taylor, C. (2000), Reconstruction of articulated objects from point correspondences in a single uncalibrated image, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 677–684.

ter Haar Romeny, B. (1994), Geometry-driven diffusion in computer vision, *in* 'Geometry Driven Diffusion in Computer Vision', Kluwer Academic Press.

ter Haar Romeny, B., Florack, L. M., Koenderink, J. J. and Viergever, M. A., eds (1997), *Scale-Space Theory in Computer Vision*, Vol. 1252, Springer Verlag LNCS.

Terzopoulos, D. (1984), Multiresolution Computation of Visible-Surface Representations, PhD thesis, Massachusetts Institute of Technology.

Thom, R. (1972), *Structural Stability and Morphogenesis*, Benjamin.

Thompson, D. and Mundy, J. (1987), Three dimensional model matching from an unconstrained viewpoint, *in* 'International Conference on Robotics and Automation', pp. 208–220.

Thompson, M., Eller, R., Radlinski, W. and Speert, J., eds (1966), *Manual of Photogrammetry*, American Society of Photogrammetry. Third edition.

Tieu, K. and Viola, P. (2000), Boosting image retrieval, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 2000, pp. I:228–235.

Todd, J. (1946), *Projective and Analytical Geometry*, Pitman Publishing Corporation.

Tomasi, C. and Kanade, T. (1991), Factoring image sequences into shape and motion, *in* 'IEEE Workshop on Visual Motion', pp. 21–28.

Tomasi, C. and Kanade, T. (1992), 'Shape and motion from image streams under orthography: a factorization method', *Int. J. of Comp. Vision* **9**(2), 137–154.

Torr, P. (1997), An assessment of information criteria for motion model selection, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 47–52.

Torr, P. (1999), Model selection for two view geometry: a review, *in* D. Forsyth, J. Mundy, V. diGesu and R. Cipolla, eds, 'Shape, Contour and Grouping in Computer Vision', Springer-Verlag, pp. 277–301.

Torr, P. and Murray, D. (1997), 'The development and comparison of robust methods for estimating the fundamental matrix', *International Journal of Computer Vision* **24**(3), 271–300.

Torr, P. and Zisserman, A. (1998), Concerning Bayesian motion segmentation, model averaging, matching and the trifocal tensor, *in* 'European Conference on Computer Vision', 1998, pp. 511–527.

Torr, P., Fitzgibbon, A. and Zisserman, A. (1999*a*), 'The problem of degeneracy in structure and motion recovery from uncalibrated image sequences', *International Journal of Computer Vision* **32**(1), 27–44.

Torr, P., Szeliski, R. and Anandan, P. (1999*b*), An integrated Bayesian approach to layer extraction from image sequences, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 983–990.

Torrance, K. and Sparrow, E. (1967), 'Theory for off-specular reflection from roughened surfaces', *Journal of the Optical Society of America* **57**, 1105–1114.

Torre, V. and Poggio, T. (1986), 'On edge detection', *IEEE Trans. Pattern Analysis and Machine Intelligence* **8**(2), 147–163.

Triesman, A. (1982), 'Perceptual grouping and attention in visual search for features and objects', *Journal of Experimental Psychology: Human Perception and Performance* **8**(2), 194–214.

Triggs, B. (1995), Matching constraints and the joint image, *in* 'Proc. Int. Conf. Comp. Vision', pp. 338–343.

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A. (2000), Bundle adjustment—a modern synthesis, *in* B. Triggs, A. Zisserman and R. Szeliski, eds, 'Vision Algorithms: Theory and Practice', Springer-Verlag, pp. 298–372. Lecture Notes in Computer Science 1883.

Triggs, W. (1997), Auto-calibration and the absolute quadric, *in* 'Proc. IEEE Conf. Comp. Vision Patt. Recog.', pp. 609–614.

Trussell, H., Allebach, J., Fairchild, M., Funt, B. and Wong, P. (1997), 'Special issue: Digital color imaging', *IEEE Trans. Image Processing* **6**(7), 897–900.

Tsai, R. (1987*a*), 'A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras', *IEEE Journal of Robotics and Automation* **RA-3**(4), 323–344.

Tsai, R. (1987*b*), 'A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras', *Journal of Robotics and Automation* **RA-3**(4), 323–344.

Tsai, R. and Huang, T. (1984), 'Uniqueness and estimation of 3D motion parameters of rigid bodies with curved surfaces', *IEEE Trans. Patt. Anal. Mach. Intell.* **6**, 13–27.

Turk, G. and Levoy, M. (1994), Zippered polygon meshes from range images, *in* 'SIGGRAPH', pp. 311–318.

Turk, M. and Pentland, A. (1991*a*), 'Face recognition using eigenfaces', *J. of Cognitive Neuroscience*.

Turk, M. and Pentland, A. (1991*b*), Face recognition using eigenfaces, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 586–591.

Tyson, J. (1990), 'Progress in low-light-level charge-coupled device imaging in astronomy', *J. Opt. Soc. Am. A* **7**, 1231–1236.

Ullman, S. (1979), *The Interpretation of Visual Motion*, MIT Press.

Ullman, S. (1996), *High-Level Vision: Object Recognition and Visual Cognition*, MIT Press.

Ullman, S. and Basri, R. (1991), 'Recognition by linear combination of models', *IEEE Trans. Patt. Anal. Mach. Intell.* **13**(10), 992–1006.

Vaillant, R. and Faugeras, O. (1992), 'Using extremal boundaries for 3D object modeling', *IEEE Trans. Patt. Anal. Mach. Intell.* **14**(2), 157–173.

Vapnik, V. (1996), *The Nature of Statistical Learning Theory*, Springer Verlag.

Vapnik, V. N. (1998), *Statistical Learning Theory*, John Wiley & Sons.

Vasconcelos, N. and Lippman, A. (1997), Empirical Bayesian EM based motion segmentation, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 527–532.

Viéville, T. and Faugeras, O. (1995), Motion analysis with a camera with unknown, and possibly varying intrinsic parameters, *in* 'Proc. Int. Conf. Comp. Vision', pp. 750–756.

Virage (n.d.), 'Virage home page at `http://www.virage.com/`'.

Vogler, C. and Metaxas, D. (1998), ASL recognition based on a coupling between HMMs and 3D motion analysis, *in* 'Proceedings, Sixth International Conference on Computer Vision', 1998, pp. 363–369.

Vogler, C. and Metaxas, D. (1999), Parallel hidden markov models for American sign language recognition, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 116–122.

Vora, P., Farell, J., Tietz, J. and Brainard, D. (1997), Digital color cameras 1: Response models, Technical Report HPL-97-53, Hewlett-Packard Laboratory.

Wactlar, H., Kanade, T., Smith, M. and Stevens, S. (1996), 'Intelligent access to digital video: The informedia project', *IEEE Computer*.

Wallace, C. and Freeman, P. (1987), 'Estimation and inference by compact encoding (with discussion)', *J. Roy. Stat. Soc. Series B* **49**, 240–265.

Wandell, B. (1987), 'The synthesis and analysis of color images', *IEEE Trans. Pattern Analysis and Machine Intelligence* **9**(1), 2–13.

Wandell, B. (1995), *Foundations of Vision*, Sinauer Associates, Inc.

Wang, J. and Adelson, E. (1994), 'Representing moving images with layers', *IEEE Trans. Image Processing* **3**(5), 625–638.

Wang, R. and Freeman, H. (1990), Object recognition based on characteristic views, *in* 'International Conference on Pattern Recognition', pp. 8–12.

Watts, N. (1987), Calculating the principal views of a polyhedron, CS Tech. Report 234, Rochester University.

Weber, M., Einhaeuser, W., Welling, M. and Perona, P. (2000), Viewpoint-invariant learning and detection of human heads, *in* 'International Conference on Automatic Face and Gesture Recognition', 1900, pp. 20–27.

Weinshall, D. (1993), Model-based invariants for 3-D vision, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1993, pp. 695–696.

Weinshall, D. and Tomasi, C. (1995), 'Linear and incremental acquisition of invariant shape models from image sequences', *IEEE Trans. Patt. Anal. Mach. Intell.*

Weiss, Y. (1997), Smoothness in layers: Motion segmentation using nonparametric mixture estimation, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', 1997, pp. 520–526.

Weiss, Y. (1999), Segmentation using eigenvectors: A unifying view, *in* 'Proceedings, Seventh International Conference on Computer Vision', 1999, pp. 975–982.

Wells, W., Grimson, W., Kikinis, R. and Jolesz, F. (1996), 'Adaptive segmentation of MRI data', *IEEE Transactions on Medical Imaging* **15**(4), 429–442.

Weng, J., Huang, T. and Ahuja, N. (1992), 'Motion and structure from line correspondences: closed-form solution, uniqueness, and optimization', *IEEE Trans. Patt. Anal. Mach. Intell.* **14**(3), 318–336.

West, J., Fitzpatrick, M., Wang, M., Dawant, B., Maurer, C.R., J., Kessler, R., Maciunas, R., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., van den Elsen, P., Napel, S., Sumanaweera, T., Harkness, B., Hemler, P., Hill, D., Hawkes, D., Studholme, C., Antoine Maintz, J., Viergever, M., Malandain, G., Pennec, X., Noz, M., Maguire, G.Q., J., Pollack, M., Pelizzari, C., Robb, R., Hanson, D. and Woods, R. (1997), 'Comparison and evaluation of retrospective intermodality registration techniques', *J. Computer Assisted Tomography* **21**(4), 554–566.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, Springer Verlag.

Wheatstone, C. (1838), 'On some remarkable, and hitherto unobserved, phenomena of binocular vision', *Philosophical Transactions of the Royal Society (London)* **128**, 371–394.

Wheeler, M. and Ikeuchi, K. (1995), 'Probabilistic hypothesis generation and robust localization for object recognition', *IEEE Trans. Patt. Anal. Mach. Intell.* **17**(3), 252–265.

Whitney, H. (1955), 'On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane', *Annals of Mathematics* **62**(3), 374–410.

Wilkinson, J. and Reinsch, C. (1971), *Linear Algebra - Vol. II of Handbook for Automatic Computation*, Springer-Verlag. Chapter I.10 by G.H. Golub and C. Reinsch.

Williams, L. and Chen, E. (1993), 'View interpolation for image synthesis', *SIGGRAPH.*

Williamson, S. and Cummins, H. (1983), *Light and Color in Nature and Art*, John Wiley & Sons.

Witkin, A. (1981), 'Recovering surface shape and orientation from texture', *Artificial Intelligence* **17**, 17–45.

Witkin, A. (1983), Scale-space filtering, *in* 'International Joint Conference on Artificial Intelligence', pp. 1019–1022.

Wolff, L., Nayar, S. and Oren, M. (1998), 'Improved diffuse reflection models for computer vision', *International Journal of Computer Vision* **30**(1), 55–71.

Wolfson, H. (1990), Model-based object recognition by geometric hashing, *in* 'European Conference on Computer Vision', 1990, pp. 526–536.

Wolfson, H. and Lamdan, Y. (1988), Geometric hashing: A general and efficient model-based recognition scheme, *in* 'Proceedings, Second International Conference on Computer Vision', 1988, pp. 238–249.

Wong, S. (1998), CBIR in medicine: still a long way to go, *in* 'IEEE Workshop on Content Based Access of Image and Video Libraries', p. 114.

Woodham, R. (1979), Analyzing curved surfaces using reflectance map techniques, *in* 'Artificial Intelligence: An MIT Perspective', MIT Press, pp. 161–182.

Woodham, R. (1980), 'Photometric method for determining surface orientation from multiple images', *Optical Engineering* **19**(1), 139–144.

Woodham, R. (1989), Determining surface curvature with photometric stereo, *in* 'International Conference on Robotics and Automation', pp. 36–42.

Woodham, R. (1994), 'Gradient and curvature from the photometric-stereo method, including local confidence estimation', *Journal of the Optical Society of America* **11**(11), 3050–3068.

Wu, Z. and Leahy, R. (1993), 'An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation', *IEEE Trans. Pattern Analysis and Machine Intelligence* **15**(11), 1101–1113.

Wyszecki, G. and Stiles, W. (1982), *Color Science: Concepts and Methods, Quantitative Data and Formulas*, Wiley.

Yachida, M., Kitamura, Y. and Kimachi, M. (1986), Trinocular vision: new approach for correspondence problem, *in* 'Proceedings IAPR International Conference on Pattern Recognition', pp. 1041–1044.

Yasnoff, W., Mui, W. and Bacus, J. (1977), 'Error measures in scene segmentation', *Pattern Recognition* **9**(4), 217–231.

Yoo, T. and Oh, I. (1999), 'A fast algorithm for tracking human faces based on chromatic histograms', *Pattern Recognition Letters* **20**(10), 967–978.

Zerroug, M. and Medioni, G. (1995), The challenge of generic object recongnition, *in* M. Hebert, J. Ponce, T. Boult and A. Gross, eds, 'Object Representation for Computer Vision', number 994 *in* 'Lecture Notes in Computer Science', Springer-Verlag, pp. 217–232.

Zhang, Y. (1996), 'A survey on evaluation methods for image segmentation', *Pattern Recognition* **29**(8), 1335–1346.

Zhang, Z. (1994), 'Iterative point matching for registration of free-form curves and surfaces', *Int. J. of Comp. Vision* **13**(2), pages 119–152.

Zhang, Z., Deriche, R., Faugeras, O. and Luong, Q.-T. (1995), 'A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry', *Artificial Intelligence Journal* **78**, 87–119.

Zhu, S. (1999), 'Stochastic jump-diffusion process for computing medial axes in markov random fields', *IEEE Trans. Pattern Analysis and Machine Intelligence* **21**(11), 1158–1169.

Zhu, S. and Yuille, A. (1996), 'FORMS: a flexible object recognition and modeling system', *Int. J. of Comp. Vision* **20**(3), 187–212.

Zhu, S., Wu, Y. and Mumford, D. (1998), 'Filters, random-fields and maximum-entropy (frame): Towards a unified theory for texture modeling', *International Journal of Computer Vision* **27**(2), 107–126.

Zisserman, A., Forsyth, D., Mundy, J., Rothwell, C., Liu, J. and Pillow, N. (1995*a*), '3d object recognition using invariance', *Artificial Intelligence* **78**(1-2), 239–288.

Zisserman, A., Mundy, J., Forsyth, D., Liu, J., Pillow, N., Rothwell, C. and Utcke, S. (1995*b*), Class-based grouping in perspective images, *in* 'Proceedings, Fifth International Conference on Computer Vision', 1995, pp. 183–188.

Ziv, J. and Lempel, A. (1977), 'A universal algorithm for sequential data compression', *IEEE Transactions on Information Theory* **IT-23**, 337–343.

看看论坛