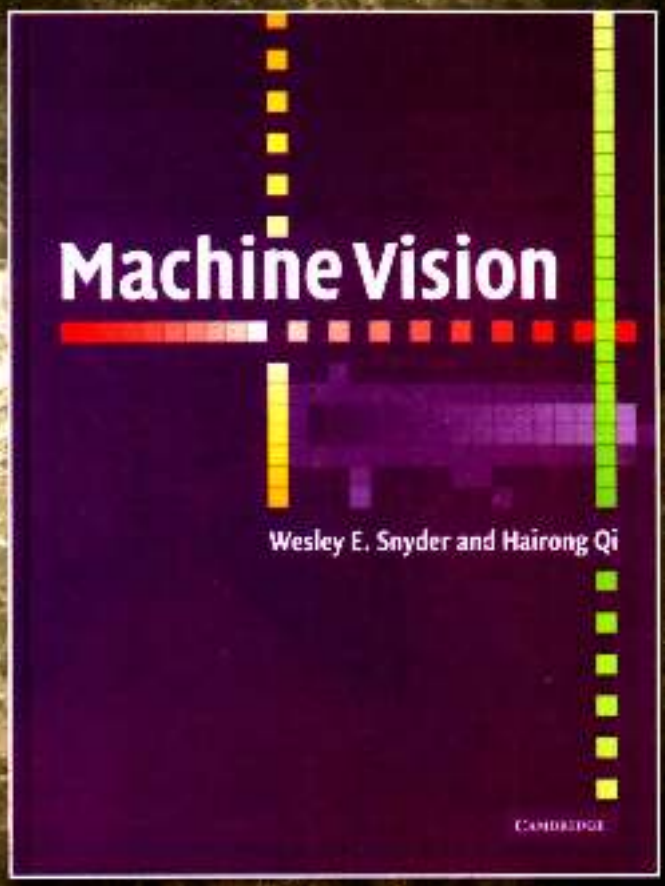
计 算 机 科 学 丛 书

机器视觉教程

(美) Wesley E. Snyder　Hairong Qi 著　林学闾　崔锦实　赵清杰　等译



Machine Vision

机械工业出版社
China Machine Press

本书对机器视觉进行了全面且通俗易懂的讲解。提供了所有必需的理论工具，并且说明如何应用到实际的图像处理与机器视觉系统中。每章包含了许多编程作业，使读者能深入了解开发实用的图像处理算法的内幕。

本书从回顾数学原理开始，然后讨论图像处理中的关键问题，还讨论了机器视觉的一些重要应用，其中包括目标的自动检测。本书强调了两个命题：一致性（解决机器视觉问题的一个主要理论架构）与优化（用来实现这些方法的数学工具）。

本书所附光盘包括书中使用的软件、数据及图像。

作者简介

Wesley E. Snyder 在美国伊利诺大学获得博士学位，现在是北卡罗来纳州立大学电机与计算机工程系教授，他曾发表过100多篇学术论文，并且是《Industrial Robots》（工业机器人）一书的作者。他曾是IEEE机器人与自动化学会以及IEEE神经网络理事会的创始人。他还担任过美国国家科学基金会、美国国家航空航天局（NASA）、Sandia实验室以及美国陆军科研办公室的顾问。

Hairong Qi 在美国北卡罗来纳州立大学获得博士学位，目前是美国田纳西大学电机与计算机工程系的助理教授。

ISBN 7-111-16790-2

9 787111 167907

HZ BOOKS

华章图书

上架指导：计算机/人工智能

华章网站 http://www.hzbook.com

网上购书：www.china-pub.com

投稿热线：(010) 88379604

购书热线：(010) 68995259, 68995264

读者信箱：hzjsj@hzbook.com

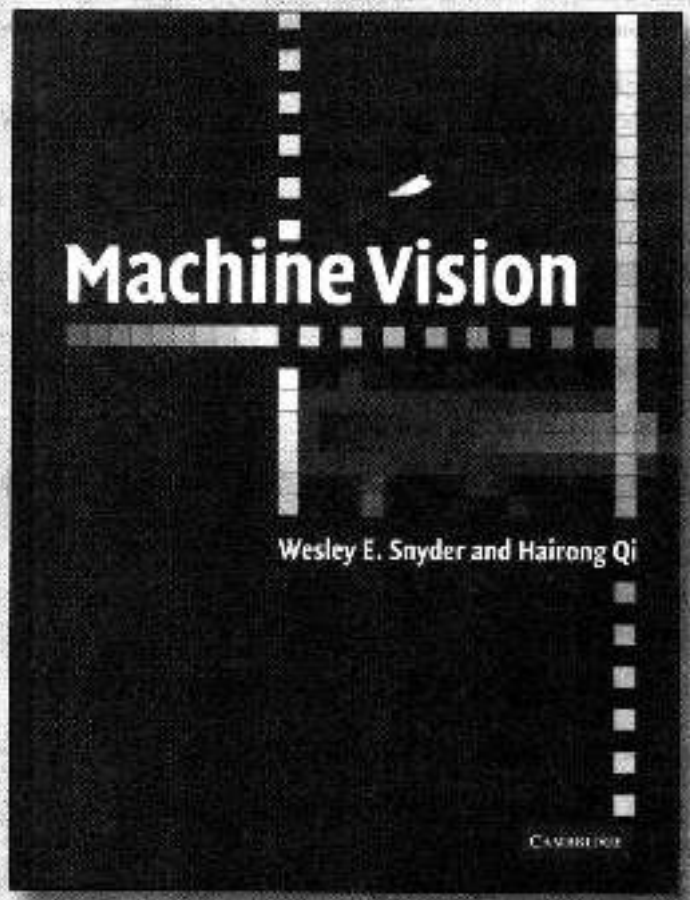
ISBN 7-111-16790-2/TP・4337

定价：49.00 元（附光盘）

计 算 机 科 学 丛 书

机器视觉教程

〔美〕Wesley E. Snyder　Hairong Qi 著　林学闾　崔锦实　赵清杰　等译



Machine Vision

机械工业出版社
China Machine Press

本书对机器视觉进行了全面且通俗易懂的讲解，是一本从实用角度出发介绍机器视觉理论与应用的教材。本书从介绍数学原理开始，然后讨论图像处理中的关键问题，如图像的描述与品质、边缘检测、特征提取、分割、纹理与形状等。讨论的内容还包括图像匹配、统计模式识别、句法模式识别、聚类、传播、自适应轮廓、参数变换以及一致性标号。本书还介绍机器视觉的一些重要应用，其中包括目标的自动检测等。

本书所附光盘包括书中使用的软件与数据。本书可作为电机工程、计算机科学与数学等专业研究生的教材，也可供技术人员参考。

Wesley E.Snyder, Hairong Qi: Machine Vision.

Originally published by Cambridge University Press in 2004.

This Chinese edition is published with the permission of the Syndicate of the Press of the University of Cambridge, Cambridge, England.

Copyright © 2004 by Cambridge University Press.

This edition is licensed for distribution and sale in the People's Republic of China only, excluding Hong Kong, Taiwan and Macao and may not be distributed and sold elsewhere.

本书原版由剑桥大学出版社出版。

本书简体字中文版由英国剑桥大学出版社授权机械工业出版社独家出版。未经出版者预先书面许可，不得以任何方式复制或抄袭本书的任何部分。

此版本仅限在中华人民共和国境内（不包括中国香港、台湾、澳门地区）销售发行，未经授权的本书出口将被视为违反版权法的行为。

版权所有，侵权必究。

本书法律顾问　北京市展达律师事务所

本书版权登记号：图字：01-2004-3285

图书在版编目（CIP）数据

机器视觉教程/（美）辛德（Snyder, W. E.）等著；林学闾等译. －北京：机械工业出版社，2005.9

（计算机科学丛书）

书名原文：Machine Vision

ISBN 7-111-16790-2

Ⅰ.机…　Ⅱ.①辛…　②林…　Ⅲ.计算机视觉－教材　Ⅳ.TP302.7

中国版本图书馆CIP数据核字（2005）第067781号

机械工业出版社（北京市西城区百万庄大街22号　邮政编码　100037）

责任编辑：毛　毅

北京瑞德印刷有限公司印刷·新华书店北京发行所发行

2005年9月第1版第1次印刷

787mm × 1092mm 1/16 · 22印张

印数：0 001-4 000 册

定价：49.00元（附光盘）

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：（010）68326294

出版者的话

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭橥了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短、从业人员较少的现状下，美国等发达国家在其计算机科学发展的几十年间积淀的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章图文信息有限公司较早意识到“出版要为教育服务”。自1998年开始，华章公司就将工作重点放在了遴选、移译国外优秀教材上。经过几年的不懈努力，我们与Prentice Hall，Addison-Wesley，McGraw-Hill，Morgan Kaufmann等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出Tanenbaum，Stroustrup，Kernighan，Jim Gray等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及庋藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力襄助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍，为进一步推广与发展打下了坚实的基础。

随着学科建设的初步完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都步入一个新的阶段。为此，华章公司将加大引进教材的力度，在“华章教育”的总规划之下出版三个系列的计算机教材：除“计算机科学丛书”之外，对影印版的教材，则单独开辟出“经典原版书库”；同时，引进全美通行的教学辅导书“Schaum's Outlines”系列组成“全美经典学习指导系列”。为了保证这三套丛书的权威性，同时也为了更好地为学校和老师们服务，华章公司聘请了中国科学院、北京大学、清华大学、国防科技大学、复旦大学、上海交通大学、南京大学、浙江大学、中国科技大学、哈尔滨工业大学、西安交通大学、中国人民大学、北京航空航天大学、北京邮电大学、中山大学、解放军理工大学、郑州大学、湖北工学院、中国国家信息安全测评认证中心等国内重点大学和科研机构在计算机的各个领域的著名学者组成“专家指导委员会”，为我们提供选题意见和出版监督。

这三套丛书是响应教育部提出的使用外版教材的号召，为国内高校的计算机及相关专业

的教学度身订造的。其中许多教材均已为M. I. T.，Stanford，U.C. Berkeley，C. M. U. 等世界名牌大学所采用。不仅涵盖了程序设计、数据结构、操作系统、计算机体系结构、数据库、编译原理、软件工程、图形学、通信与网络、离散数学等国内大学计算机专业普遍开设的核心课程，而且各具特色——有的出自语言设计者之手、有的历经三十年而不衰、有的已被全世界的几百所高校采用。在这些圆熟通博的名师大作的指引之下，读者必将在计算机科学的宫殿中由登堂而入室。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证，但我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。教材的出版只是我们的后续服务的起点。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下:

电子邮件: hzjsj@hzbook.com
联系电话: (010) 68995264
联系地址: 北京市西城区百万庄南街1号
邮政编码: 100037

专家指导委员会

（按姓氏笔画顺序）

尤晋元	王　珊	冯博琴	史忠植	史美林
石教英	吕　建	孙玉芳	吴世忠	吴时霖
张立昂	李伟琴	李师贤	李建中	杨冬青
邵维忠	陆丽娜	陆鑫达	陈向群	周伯生
周克定	周傲英	孟小峰	岳丽华	范　明
郑国梁	施伯乐	钟玉琢	唐世渭	袁崇义
高传善	梅　宏	程　旭	程时端	谢希仁
裘宗燕	戴　葵			

译 者 序

机器视觉又常称计算机视觉，这门学科的发生与发展已有几十年的历史，它是一门研究通过图像或视频数据观察周围世界的学科，主要以摄像机拍摄的图像或视频为原始数据，提取出在图像或视频中能观察到的事物。这个学科要解决的问题，与人类通过眼睛观察世界的视觉感知功能十分相似。称机器视觉是强调用机器构成的系统来处理视觉问题，而称计算机视觉则强调处理视觉的问题是一个计算问题，核心部件是计算功能强大的计算机。

随着信息时代的到来，用计算机处理各种信息的需求越来越多。多媒体信息处理技术已成为日常生活各领域的迫切需要，而人们就更希望能用计算机技术来处理视觉问题，例如利用人脸、虹膜、指纹等识别技术来处理与个人有关的一切事务。利用自动识别技术帮助盲人，利用视觉自动监控系统监视环境中发生的非常事件，如陌生人的侵入、老年人的异常行动等。再如智能交通管理系统、视频检索、用于军事目的的自动目标检测等，都需要应用机器视觉技术来解决问题。正如视觉是人类在自然环境与社会环境中生存不可缺少的最重要感知器官，机器视觉技术也是信息技术中的一门不可少的技术，因此它成为计算机学科中不可或缺的一门学科。

但是如何编写好一本机器视觉教科书却并不容易，这是由这门学科本身的特点所决定的。

几十年来，计算机视觉的研究吸引了多个学科学者与研究人员的目光，计算机视觉的研究已深入到许多不同的应用领域，例如字符识别、人脸识别、指纹识别等在基本原理与实施细节上都形成了独特的算法体系，每个领域都有数量极大的参考文献。显然要对每一个都作详细介绍，并要让读者理解，在篇幅上是不可能的，对初学者来说也是不必要的。

为了使这本书既能作为相关学科研究生与本科高年级学生学习机器视觉基础知识的教科书，并且又能作为有关专业人员自学机器视觉基本概念的参考书，作者将本书组织成基础与专题两个部分。基础部分是本书的重点，讲述每一章命题的基本概念以及基本处理方法，如图像处理、模式识别、线性代数、概率论等基础知识。学生们可以对机器视觉的基本概念及基本处理方法有一个全面的了解与掌握。本书作者还十分重视对学生通过实际动手培养基本能力，因而专门设计了一章讲解图像的一些基本操作，并要求读者能从底层的图像处理程序写起，以便真正掌握算法的实质。因此这本书是一本适合高年级本科生与研究生学习机器视觉的教科书，也是工程师与其他学科的研究人员学习机器视觉的入门参考书。

但是一旦读者在掌握了机器视觉的基础知识后，往往需要深入到机器视觉的某个应用领域中去，因此本书作者在每一章都安排了专题。在专题中讨论一些较深入的概念以及有代表性的观点，并附有丰富的参考文献目录。这样一来，读者就可以方便地深入到感兴趣的领域中去了。

此外，本书作者从始至终都强调两个命题，一个是一致性检测，另一个是求优化解。对前一个命题，作者把它看作解决机器视觉问题的基本哲理，这一点的确非常重要。当人们观察一幅数字图像时，孤立地观察每一个像素，或一个仅包含几个像素的小区域，是看不出图像中有什么东西的，而只有在审视了像素及其邻域中灰度（或颜色）值的分布特点才能发现

有价值的东西。例如，图像中有一条直线，若孤立地观察线上的每一点，并不能得出有一条直线的结论，而如果把这些点联系起来观察，就会发现有一条直线存在。这些点为什么组成一条直线呢？这是因为这些点具有一个公共的特点，它们都符合某一直线方程的条件，这就是它们之间存在的一致性。这个例子表明，仅从局部信息来看是看不出图像中有什么内容，只有从全局性角度去观察才能发现图像中究竟包含了什么。然而，图像处理的一些基本算子（如边缘算子、平滑算子等）都只能在局部进行计算，无法在大面积范围内直接进行计算是机器视觉计算的要害，而一致性检测是机器视觉中从局部计算到全局性计算的主要哲理，这点对理解机器视觉的基本处理方法是极其关键的。至于求优化解是机器视觉计算求解的基本数学工具，还拿检测直线的例子来说吧！在图像中一条直线是由若干点组成的，由于这些点是通过边缘检测算子算出来的，这些点的位置往往不十分准确，因此严格地说这些点并不构成一条理想的直线，它可以用几条不同的直线来近似。那么究竟用哪一条直线来近似呢？这就是要找一个“最优”的直线的问题。要用一句话来概括机器视觉的话，就是从带噪声的数据中求解，这种解可以是识别一个字符，检测出一张人脸，跟踪了一个物体的运动等等。由于数据往往是不完整的，是混杂在一起的（一个场景往往是多种物体的组合），带有很强的噪声，因此往往没有解析解，没有唯一解，需要从多个可能的解中找到最优解，求最优解是机器视觉中最常用的数学工具。

我们在翻译此书的过程中体会到这本书与其他一些机器视觉（或计算机视觉）的著作有明显不同，它的注意力并不放在介绍计算机视觉的最新成果上，而在于把学习机器视觉最需要、最基本的概念与处理方法有条有理地组织起来，有利于学生与初学者尽快地掌握最基本的内容。而专题部分又可以作为读者学习更深入知识的桥梁。

由于翻译时间偏短，有些部分我们理解得不一定正确，用语也不够准确，译文中难免会出现这样那样的错误与不当之处，敬请读者们原谅，并欢迎指正。

致教师

本书是讲解计算机识别景物中物体的教材，分基础内容与专题两个部分，可作为教材与教学参考书。几乎所有的章都有基础部分，可以作为教材，按教学法组织的，而专题部分包含大量有关新近发表的文章的参考文献，可以用作参考。本教材是面向电机与计算机工程系、计算机科学系或数学系的高年级本科生与研究生的。

从第4章到17章的内容包括边缘检测、形状描述、传播、自适应轮廓、参数变换、匹配与一致性标号等有关专题，还包括句法与统计模式识别，以及聚类等内容。有两个主题是贯穿所有这些章节的：一致性（解决机器视觉问题的一个主要理论架构）与优化（用来实现这些方法的数学工具）。这两个主题是无所不在的，所以在每章的结尾都要讨论它们在本章中是如何体现的。第18章介绍目标自动检测这一应用问题，结合了前面章节讨论的所有主题，展示如何解决实际问题。

本教材假设读者在线性代数与高等微积分方面具有良好的研究生程度的背景知识。成功完成本课程的学生可以承担在工业、医学、与军用机器视觉系统方面的设计任务。所附的光盘中有作者开发的软件工具以及图像数据，用来支持完成书中作业。这些软件可运行在Windows或Linux操作系统支持的PC机、在OS-X下的Macintosh计算机以及Solaris支持下的SUN计算机环境中。软件能够处理任何一种类型的图像，并能将JPEG等格式的图像转换成“标准”格式图像或进行反方向转换。

尽管本书的使用方式可以多种多样，但作者的主要设计目的是使其作为机器视觉的研究生课程教材，以及作为机器视觉的参考书。如果作为教材，希望学生能阅读本书每一章的基本部分（本书内容一个学期都学不完）。下面的表格是为研究生程度的基础课程设计的学习计划。

对学生的作业与编程必须包括一个正式的报告，由于作业中通常包含编程，完成作业的时间可以更长些，建议使用的时间列在表格第3列的括号中。本书也可用于本科高年级学生的课程，但听课的学生与讲课内容要仔细挑选。

对希望进一步深造的学生，本书的“专题”部分可以作为阅读大量文献的导引。一定要对你的学生强调这样一点（就像我们在本书中所说的），即没有任何一本教科书能提供参考文献中才能提供的细节。而对一个实际的机器视觉项目来说，则要求开发工程师深入研读出版的杂志与会议论文。正如前面所述，贯穿本书的两个主题是一致性与优化。一致性概念是贯穿整个学科的，作为解决机器视觉问题的主要理论架构。针对一个机器视觉的实际应用，工程师需要进行搜索，以便找到确定信息源之间具有一致性的方法。优化是解决机器视觉问题的主要数学工具，包括确定一致性。在每一章的末尾介绍完技术问题后，我们都要提醒学生在该章何处体现了一致性原则，以及在何处应用了哪一种优化方法。

课程安排举例

	序号内容	作业（花费周数）	阅读材料
1	引言，术语，对图像的运算，模式分类和计算机视觉，图像生成，分辨率，动态范围，像素	2.2~2.5 2.9(1)	读第2章，并确认自己具备学习该课程的基础
2	作为函数的图像，图像退化，点扩展函数，复原	3.1(1)	第1章与第3章
3	图像的特性，等亮度线，脊，连通性	3.2，4.1(2)	4.1节~4.5节
4	核算子：应用核运算估计边缘位置	4.A1， 4.A2(1)	5.1节与5.2节
5	用双二次函数拟合图像，用求向量导数的方法使函数最小化	5.1，5.2(1)	5.3节~5.4节 （跳过六边形像素）
6	图像的向量表示，图像基函数，边缘检测，高斯模糊，二阶及高阶导数	5.4，5.5(2)， 5.7，5.8， 5.9(1)	5.5节与5.6节（跳过5.7节）
7	尺度空间简介，讨论作业	5.10，5.11(1)	5.8节（跳过5.9节）
8	松弛与退火	6.1，6.3(1)	6.1节~6.3节
9	扩散	6.2(2)	6A.2节
10	MFA与扩散之间的等价性	6.7与6.8(1)	6A.4节
11	图像数学形态学	7.5~7.7(1)	7.1节
12	数学形态学（续），灰度数学形态学，距离变换	7.10(2)	7.2节，7.3节
13	填补边缘间隙，连通性	7.4(1)	7A.4节
14	用最佳阈值进行分割		8.1节，8.2节
15	连通分量标号	8.2(1)	8.3节
16	二维几何，变换	9.3(1)	9.1节，9.2节
17	二维形状特征，不变矩、傅里叶描述子，中轴	9.2，9.4， 9.10(1)	9.3节~9.7节
18	用蛇形主动轮廓与汽球主动轮廓进行分割		8.5节，8.5.1节
19	偏微分方程表达与水平集		8.5.2节
20	从X提取形状与结构光照明	9.10(1)	9A.2.2节，9A.2.3节
21	图论图像表示方法：图，区域相邻图。子图同构		第12章
22	一致性与松弛标号	10.1(1)	第10章
23	Hough变换，参数变换	11.1(2)	11.1节，11.2节，11.3.3节
24	广义Hough变换，高斯映像，电路板上检测孔的应用		11A.3节
25	图匹配，弹簧与模板，关联图	13.2与13.3(1)	13.1节~13.3节
26	统计模式识别的作用		

致谢

感谢我在北卡罗来纳州立大学的研究生们，尤其是Rajeev Ramanath，没有他的帮助，就完成不了这本书。Bilgé Karacali在校对方面给了不少帮助，并且对支持向量机这一节贡献很大。

当然要不是为了我的太太Rosalyn，所有这些都不重要了，正是她对我的鼓励才使这一切得以发生，她还不止一次地对整本书进行编辑，把原本工程化的语言转变成标准的英语。

W. E. Snyder

我首先要感谢Wesley Snyder博士邀请我作为合作者，我非常喜欢这种合作，并且从中获取了极其宝贵的经验。

在最后要交付这本书时，正值我的父母从中国来探亲。为了支持我完成最后的工作，他们一直陪伴我而放弃了在城市里观光与欢度节假日，我要深深地感谢他们。还要感谢Feiyi，给予我永远的技术支持与紧张工作后的宽慰。

Hairong Qi

目　录

译者序

致教师

第1章　引言……1

1.1　本书宗旨 ……1

1.2　读者需具备的预备知识 ……1

1.3　一些术语 ……2

1.3.1　图像处理 ……2

1.3.2　机器视觉 ……3

1.4　机器视觉系统的组成 ……3

1.5　图像的性质 ……4

1.6　图像的操作分析 ……4

参考文献 ……5

第2章　数学原理回顾……7

2.1　概率论简单回顾 ……7

2.2　线性代数简单回顾 ……8

2.2.1　线性变换……10

2.2.2　求导运算……11

2.2.3　特征值与特征向量 ……12

2.3　函数最小化简介 ……12

2.3.1　牛顿-拉弗森方法 ……14

2.3.2　局部最小与全局最小 ……15

2.3.3　模拟退火……15

2.4　马尔科夫模型 ……16

2.4.1　隐马尔科夫模型 ……17

2.4.2　维特比算法……18

2.4.3　马尔科夫输出 ……19

2.4.4　估计模型参数 ……20

2.4.5　隐马尔科夫模型的应用 ……20

参考文献 ……21

第3章　编写图像处理程序 ……23

3.1　图像文件系统软件 ……23

3.1.1　IFS头部结构 ……23

3.1.2　某些有用的IFS函数……24

3.1.3　带共性的问题 ……24

3.2　图像处理程序的基本结构……24

3.3　好的编程风格 ……25

3.4　示例程序 ……25

3.5　生成文件 ……27

第4章　图像的生成与表示方式 ……29

4.1　图像的表示方法 ……29

4.1.1　图像表示……29

4.1.2　函数表示……30

4.1.3　线性表达式……30

4.1.4　概率表示……30

4.1.5　空域频率表示 ……30

4.1.6　关系表示方法（图表示法）……32

4.2　数字图像 ……32

4.2.1　数字图像的生成 ……32

4.2.2　距离图像的生成 ……35

4.3　图像生成机制 ……37

4.4　将图像看作表面 ……38

4.4.1　等亮度线……38

4.4.2　脊 ……38

4.4.3　二值图像与中轴 ……39

4.5　邻域关系 ……39

4.6　结论 ……41

4.7　术语 ……41

专题4A　图像表示方法 ……42

4A.1　采样的变种：六角形像素 ……42

4A.2　其他形式的图像表示 ……44

参考文献 ……46

第5章　线性算子与核算子……49

5.1　什么是线性算子 ……49

5.2　核算子在数字图像中的应用 ……49

5.2.1　自变量的方向：卷积和相关 ……50

5.2.2　用核算子估计导数 ……50

5.3 通过函数拟合估计导数……51
5.4 图像的矢量表示 ……55
5.5 图像的基向量 ……56
5.6 边缘检测 ……57
5.7 用核算子表示可微函数的采样 ……59
5.8 计算卷积 ……62
5.9 尺度空间 ……64
5.9.1 四叉树 ……64
5.9.2 高斯尺度结构 ……65
5.10 量化边缘检测算子的精确性 ……67
5.11 人们的做法……67
5.12 结论 ……69
5.13 术语 ……69
专题5A 边缘检测器 ……72
5A.1 Canny边缘检测器 ……72
5A.2 改进边缘检测 ……73
5A.3 从边缘点推理线段 ……73
5A.4 空域频率表示 ……74
5A.5 术语 ……75
参考文献……77
第6章 图像松弛：复原与特征抽取 ……81
6.1 松弛……81
6.2 复原……81
6.3 最大后验方法 ……84
6.3.1 贝叶斯准则……84
6.3.2 题外话：逆问题中的问题 ……86
6.3.3 用于边缘保留型平滑的目标函数……86
6.4 均值场退火 ……87
6.4.1 选择先验项……89
6.4.2 退火：避免局部最小值 ……91
6.4.3 如何对一个含有核算子的函数求导…92
6.4.4 实际考虑：边缘保留型的平滑 ……94
6.5 结论……95
6.6 术语……95
专题6A 替代算法与等价算法……97
6A.1 GNC：一种可去除噪声的替代算法 …97
6A.2 传导率可变的扩散 ……99
6A.3 面向边缘的各向异性扩散 ……100
6A.4 对图像松弛算子的一个通用描述 …100
6A.5 与神经元网络之间的关系 ……103
6A.6 结论 ……104
参考文献 ……105
第7章 数学形态学……111
7.1 二值形态学……111
7.1.1 膨胀 ……111
7.1.2 腐蚀 ……113
7.1.3 膨胀与腐蚀的性质 ……113
7.1.4 开运算与闭运算 ……115
7.1.5 开运算与闭运算的性质 ……116
7.2 灰度形态学……117
7.3 距离变换 ……118
7.3.1 采用掩膜计算距离变换 ……118
7.3.2 Voronoi图 ……119
7.4 结论 ……119
7.5 术语 ……119
专题7A 形态学……121
7A.1 有效地计算腐蚀、膨胀 ……121
7A.2 形态学采样定理 ……124
7A.3 选择结构元素 ……126
7A.4 边缘及曲面上的缝隙弥合 ……126
7A.5 术语 ……136
参考文献 ……136
第8章 分割……139
8.1 划分图像 ……139
8.2 阈值化分割……140
8.3 连通分量分析 ……142
8.3.1 递归式区域增长算法……143
8.3.2 迭代式连通分量分析方法 ……145
8.3.3 标号图像的一种替代方法 ……149
8.4 曲线分割 ……149
8.5 主动轮廓线……150
8.5.1 能量最小化原则 ……150
8.5.2 偏微分方程方法 ……151
8.6 曲面分割 ……153
8.6.1 曲面描述 ……153
8.6.2 椭圆及椭球拟合 ……154
8.7 评估分割质量……155
8.8 结论 ……156

8.9 术语 ……156
专题8A 分割 ……157
8A.1 纹理分割 ……157
8A.2 使用边缘的图像分割 ……159
8A.3 运动分割 ……160
8A.4 颜色分割 ……160
8A.5 使用MAP方法的分割 ……160
8A.6 人如何完成分割 ……160
参考文献 ……160
第9章 形状 ……167
9.1 线性变换 ……167
9.2 基于协方差矩阵的变换方法 ……169
9.2.1 K-L展开的推导 ……169
9.2.2 K-L展开的性质 ……172
9.3 简单特征 ……173
9.4 矩 ……176
9.5 链码 ……177
9.6 傅里叶描述子 ……178
9.7 中轴 ……178
9.8 变形模板 ……179
9.9 二次曲面 ……180
9.10 表面的谐波函数表示 ……181
9.11 超二次曲面 ……181
9.12 广义柱体 ……183
9.13 结论 ……183
9.14 术语 ……183
专题9A 形状的描述 ……184
9A.1 求非凸区域的直径 ……184
9A.2 从图像推测三维形状 ……186
9A.3 运动分析与跟踪 ……192
9A.4 术语 ……194
参考文献 ……197
第10章 一致性标号 ……205
10.1 一致性 ……205
10.2 松弛法标号 ……208
10.2.1 利用一致性来修正标号 ……208
10.2.2 标号问题示例 ……209
10.3 结论 ……210
10.4 术语 ……211
专题10A 二维线条图的三维解释 ……211
参考文献 ……213
第11章 参数变换 ……215
11.1 Hough变换 ……215
11.1.1 垂直线带来的问题 ……216
11.1.2 如何找到交点——累加器数组 ……217
11.2 减少计算复杂度 ……218
11.3 检测圆 ……218
11.3.1 用非共线三个像素表示一个圆的定位 ……218
11.3.2 用参数变换检测圆 ……219
11.3.3 检测圆中利用梯度信息降低计算量 ……219
11.4 广义Hough变换 ……220
11.5 结论 ……220
11.6 术语 ……221
专题11A 参数变换 ……221
11A.1 检测抛物线 ……221
11A.2 检测峰值 ……223
11A.3 高斯图 ……223
11A.4 立体视觉中的参数一致性 ……223
11A.5 结论 ……224
11A.6 术语 ……224
参考文献 ……225
第12章 图和图论概念 ……227
12.1 图 ……227
12.2 图的性质 ……227
12.3 图结构的实现 ……228
12.4 区域邻接图 ……228
12.5 图匹配的应用：子图同构问题 ……230
12.6 外观图 ……230
12.7 结论 ……231
12.8 术语 ……231
参考文献 ……232
第13章 图像匹配 ……233
13.1 图像表达的匹配 ……233
13.1.1 模板匹配 ……233
13.1.2 点匹配 ……234
13.1.3 线段匹配 ……234

13.1.4 特征图像 ……235
13.2 匹配简单特征 ……237
13.3 图匹配 ……238
13.3.1 关联图 ……238
13.3.2 弹性模板 ……240
13.4 结论 ……241
13.5 术语 ……241
专题13A 匹配 ……243
13A.1 弹簧加模板回顾 ……243
13A.2 人工神经元网络用于物体识别 ……244
13A.3 图像索引 ……247
13A.4 匹配几何不变量 ……248
13A.5 结论 ……250
13A.6 术语 ……251
参考文献 ……251
第14章 统计模式识别 ……255
14.1 分类器设计 ……255
14.1.1 决策规则的选择 ……255
14.1.2 最大似然分类器 ……256
14.1.3 分类器学习算法 ……256
14.2 贝叶斯规则和最大似然分类器 ……257
14.2.1 贝叶斯规则 ……257
14.2.2 参数模式分类器 ……258
14.2.3 概率密度估计 ……259
14.2.4 均值估计 ……260
14.2.5 方差估计 ……260
14.2.6 似然比 ……261
14.3 决策区间和错误概率 ……262
14.4 条件风险 ……263
14.5 二次函数分类器 ……265
14.6 最小最大规则 ……267
14.7 最近邻方法 ……267
14.8 结论 ……268
14.9 术语 ……269
专题14A 统计模式识别 ……270
14A.1 用统计方法匹配特征向量 ……270
14A.2 支持向量机 ……272
14A.3 结论 ……275
14A.4 术语 ……276
参考文献 ……276
第15章 聚类 ……279
15.1 聚类之间的距离 ……279
15.2 聚类算法 ……281
15.2.1 合并聚类 ……281
15.2.2 k均值聚类 ……283
15.3 最优化聚类方法 ……284
15.3.1 分支定界法 ……285
15.3.2 向量量化 ……285
15.3.3 赢者全得算法 ……285
15.4 结论 ……286
15.5 术语 ……286
参考文献 ……288
第16章 句法模式识别 ……289
16.1 术语 ……289
16.2 文法类型 ……290
16.2.1 0型文法 ……290
16.2.2 1型文法 ……290
16.2.3 2型文法 ……291
16.2.4 3型文法 ……291
16.3 用文法结构进行形状识别 ……291
16.3.1 3型文法 ……291
16.3.2 2型方法 ……294
16.4 结论 ……296
16.5 术语 ……296
参考文献 ……297
第17章 应用 ……299
17.1 多谱图像分析 ……299
17.2 光学字符识别 ……299
17.3 自动诊断和辅助诊断 ……299
17.4 检测和质量控制 ……300
17.5 安全检测和入侵者识别 ……300
17.6 机器人视觉 ……300
17.6.1 机器人外科手术 ……301
17.6.2 机器人驾驶 ……301
参考文献 ……302
第18章 自动目标识别 ……307
18.1 层次体系 ……307
18.2 系统组成 ……308

18.3 算法的性能评价 ……309
18.3.1 性能表示……310
18.3.2 从训练数据生成ROC曲线 ……311
18.3.3 性能和系统评价 ……312
18.4 特有的机器视觉问题……312
18.4.1 目标信号可变性和误警率 ……312
18.4.2 跟踪 ……313
18.4.3 分割 ……314
18.4.4 特征选择……314
18.5 自动目标识别算法 ……314
18.5.1 基于模型的技术 ……316
18.5.2 基于统计的技术 ……317
18.5.3 模板匹配……317
18.6 Hough变换 ……318
18.7 形态学技术 ……318
18.8 链码 ……319
18.9 结论 ……319
参考文献 ……320
索引 ……327

第1章 引 言

The proof is straightforward, and thus omitted.

Ja-Chen Lin and Wen-Hsiang Tsai[⊖]

1.1 本书宗旨

我们把这本书写成两个层次的形式。主要层次的内容是入门引导，但“入门引导”并不意味着“容易”或“简单”，或者“不需要数学”。“入门引导”讨论的话题是在理解深入的话题之前必须要掌握的。

此外，我们还力求把这本书写成有用的参考书。例如当读者需要实现一个实际系统时，需要得到比这本书中所包含的内容更详细的资料时，我们能为他提供恰当的参考资料，使他能很顺利地研究下去。

我们写这本书的风格，力求适合学生的特点，并且具有交谈的风格。

我们还力求使这本书有很强的可读性，并令人感到愉快。某些词故意拼错，显然是为了增加幽默感，某些幽默通过夸张或使用双关语来实现，希望读者能体谅我们。

我们并不打算在这本书中包括机器视觉的所有话题。尤其是涉及字符识别与人脸识别的几乎所有文章都没有包括进来。这丝毫没有轻视这些十分重要与十分成功的应用领域的意思，而只是因为这些文章具有很强的针对性。何况我们是无法在这本书中包括每一个方面的。

有两个命题在本书中贯穿始终：**一致性**与**优化**。一致性是一种概念上的工具，体现在许多算法中，以有利于机器识别图像。其融合从局部度量中得到的信息以便获取有关图像全局性的结论。优化问题是一种数学机制，它几乎在每一章都使用，以实现这些章节要解决的问题，不管是模式分类或是图像匹配。

由于一致性与优化这两个命题是如此重要与无所不在，因此在每一章的结论中我们都要给学生指出它们在本章何处出现。所以读者有必要阅读每章的结论，也许这里包含有考试的内容呢！

1.2 读者需具备的预备知识

本书的主要读者是：电机工程、计算机工程、计算机科学、数学、统计学或物理学专业的研究生或本科高年级学生。为了学好这本书，你必须学过研究生的高等微积分，以及统计学或概率论的课程。你必须正规地上过线性代数课或掌握这方面的知识。

许多家庭作业是编程作业，要用计算机来完成。为了完成这些作业需要的硬件与软件环境需具有如下功能：

1）用C语言处理大数组（256 × 256）。

⊖ Ja-Chen Lin and Wen-Hsiang Tsai, “Feature-preserving Clustering of 2-D Data for Two-class Problems Using Analytical Formulas: An Automatic and Fast Approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5), 1994.

2）显示一幅图像。

3）打印图像。

随本书附赠的光盘中包含有你需要的所有软件，但不包括编译器、编辑器等。

我们倾向于要求读者自己编写程序，特别是从相对底层的程序开始。某些功能的程序在Matlab等软件包中已经有，尽管你通过调用这些程序可以用到一些东西，但是可以通过自己编写与调试这些程序而学到更多的知识。当然，如果某些算法的编程过于复杂，以致于妨碍了图像分析本身的学习时，可以不这样做。为此，我们为学生们提供了子程序库，使得学生能忽略数据类型、字节交换、文件存取以及与平台有关的细节，而集中精力于研究使图像分析算法能工作的逻辑问题。

尽管你需要的所有信息在这本书中都有，但是我们建议学生应该有一个导师，并且到课堂去上课。先预习一下书中课文，然后到课堂上去，之后再读课文。

请记住:

A hacker hermit named Dave

Tapped in to this course in his cave

He had to admit

He learned not a bit

But look at the money he saved

2 现在进入有关技术的话题。

1.3 一些术语

学生们有时对图像处理与机器视觉之间的区别搞不清楚，在这一节我们对一些术语下定义，以便弄清这两个学科在内容与目标上的区别。

1.3.1 图像处理

许多人认为这本书的内容是图像处理领域的一部分。然而我们最好从图像处理与机器视觉要解决的问题来区分。“图像处理”在于要使图像经处理后变得更好，图像处理系统的输出仍然是一幅图像，而“机器视觉”系统的输出却是与图像内容有关的信息。图像处理系统的功能包括增强、编码、压缩、复原与重构。

图像增强

图像增强系统所执行操作的结果是使人们觉得处理后的图像质量更好。典型的操作运算包括对比度增强（如直方图均衡化等功能）、亮度放缩、边缘锐化等。

图像编码

编码是研究图像信息的表达方式，使其更经济与有效，这包括量化方法、冗余消除。编码还可能包括研究图像信息的表达方式，使其在传输或存储图像中出现错误时仍具有鲁棒性。

图像压缩

压缩有许多技术与编码中的技术相同，但其目的在于减少存储与传输图像的比特数量。

图像复原

复原关注的是对图像中的错误进行修复。它与图像增强不同，增强只是关注使图像看上

去更好些。而要“修正”一幅图像，需要有图像退化的模型。在复原的应用中常常假设有一个确定性的模糊操作，并附有添加性的随机噪声。

图像重构

重构通常是指从若干局部图像重构成一幅完整图像的过程。例如，在计算机层析X射线摄像技术（CT）中从受试者中得到数量很大（如360幅）的X射线投影[⊖]。从这组一维信号中计算出二维图像中每一点的X射线吸收率，相同的方法还用在正电子放射层析摄像技术（PET）、核磁共振图像（MRI），以及在某些“从X恢复形状”（shape-from-X）的算法中，这些将在本书后续章节讨论。

3

1.3.2 机器视觉

机器视觉是机器（通常指数字计算机）对图像进行自动处理并报告“图像中有什么”的过程，也就是说它识别图像中的内容。图像中的内容往往是某些机器零件，而处理的目标不仅要能对机器零件定位，还要能对其进行检验。我们将在本书中详细讨论一些机器视觉的应用，如自动目标识别（ATR）和工业检验。还有各种各样的应用，在时间与书的容量上都不允许我们涉及，例如从液体流的观察中确定流的方程[1，1]。

对机器视觉，人们也经常使用术语“计算机视觉”与“图像理解”。

机器视觉由两部分组成：特征度量与基于这些特征的模式识别。

特征度量

特征度量是本书主要关注的话题。在本书中，除了第14与15章外，我们关注于对图像的元素（像素）进行处理，并从这些像素以及像素集合提取一组度量，这些度量可以表示整幅图像或其某些组成的特点。

模式分类

模式分类被定义为，对度量作出决策的过程，换句话说，即给定一个未知物体的某个度量或一组度量，根据未知物体属于哪些类的知识，从这些度量作出决策。例如，可能的物体类别是男人与女人，而某种可用来区分男女的度量是人体高度（显然，高度并不是一个区分男女的好度量，因为如果认为任何人身高超过5尺6寸，他就是男人，在许多场合会出现错误）。

4

模式识别

模式识别可以像模式分类一样定义为，对未知物体进行分类，但是一般这个术语包括产生度量值这个过程。

1.4 机器视觉系统的组成

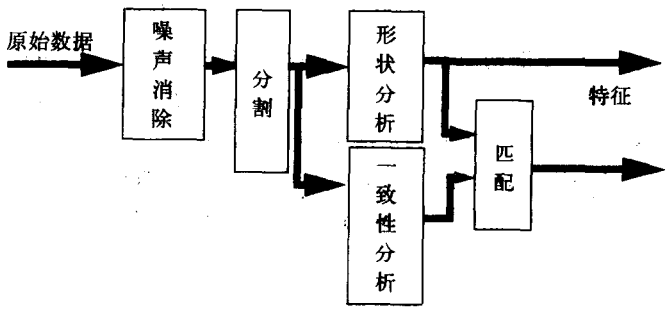
图1-1用图的方式表示了一个机器视觉系统在最基本层次上的组成。首先对未知物体进行度量，并确定一组特征的度量值。在工业应用中，这些特征包括被度量零件在图像中的长、宽和面积。一旦特征经过度量后，其数值就被送到一个实现决策规则的过程中去。这种决策规则一般用一个子程序实现。它对度量值进行计算，并根据所度量的值确定物体最可能属于的类别。

⊖ 有时CT表示“CAT扫描”，此时CAT表示“computed axial tomography”，还有一些其他类型的层析X射线摄像仪，计算轴向层析X射线摄像仪。

正如图1-1所示，一个机器视觉系统的结构相当简单。然而，每个模块的细节会相当复杂，并且对设计分类器与特征度量系统的不同选择很多。在本书中，我们会讨论分类器设计的问题，但是确定特征与对特征进行度量的过程是本书的主要话题。

“特征度量”模块可以进一步分解成较详细的操作，如图1-2所示。在这个层次上，结构流程会变得更加复杂，因为要执行的具体操作，随图像的类型以及任务要实现的目标变化而变化。并不是所有的操作在任何的应用中都要执行。

图1-1 机器视觉系统的组成框图



5

图1-2 特征度量系统的某些组成，许多机器视觉应用并不使用其中每一模块，信息也常常以别的方法流动。例如，可以直接对图像进行匹配

1.5 图像的性质

在第4章中会更加关注图像的性质，我们会看到有许多不同类型的图像以及不同的表示图像的方式。图像的类型包括称之为“图片”（picture）的二维图像。此外，还会讨论三维的图像与距离图像。我们会考虑图像不同的表示方法，如直接用图像、函数、线性以及关系表示的方式。

1.6 图像的操作分析

我们会学习许多对图像进行的不同操作。本书强调的是“图像分析”、“计算机视觉”、“机器视觉”或“图像理解”。所有这些词汇说的是同一件事情。我们感兴趣的是对图像进行度量，目的在于为机器（一般，但不限于计算机）提供识别图像中有什么东西的能力。这个过程包括几个步骤：

- 去噪声——所有图像都包含有噪声，大部分图像都是被模糊的，有些还有其他畸变。我们需要去除或降低这些畸变，以便能执行进一步的操作（运算）。本书第6章与第7章中将讨论两种去噪声的一般方法。
- 分割——需要把图像分割成有意义的区域。分割在第8章讨论。
- 特征抽取——对分割出来的区域进行度量，可以是几何度量，也可以是别的方式度量，这在第9章讨论。

6

- 一致性——从局部度量解释整幅图像的内容将在第10章与第11章讨论。
- 分类与匹配——识别物体在第12章到第16章讨论。

现在翻到下一章（你有没有注意到，这章没有作业！不要着急，在后续章节就有了。）

参考文献

[1.1] C. Shu and R. Jain, "Vector Field Analysis for Oriented Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

7

第2章　数学原理回顾

Everything, once understood,is trivial.

W. Snyder

2.1　概率论简单回顾

设想一个统计学实验：掷两个骰子。两个数加起来可从2到12，但是正如大家知道的，某些数目出现的次数要比另一些大。为了说明这一点，可以考虑得到数5的所有可能情况。

从图2-1可以看到，用两个骰子得到5可以有4种情况。每掷一个骰子的事件是**独立**的。也就是说，第二个骰子掷出一个“2”的机会（6分之1）完全与第一个骰子掷出什么无关。

事件的独立性有重要的含意，它意味着两个事件发生的联合概率等于它们单独发生的概率与条件概率的乘积：

$$Pr(a|b)P(b) = Pr(a)Pr(b) = Pr(b|a)Pr(a) = Pr(a,b) \tag{2-1}$$

式（2-1）中符号a与b代表**事件**，例如掷出一个“6”，$Pr(b)$是发生这种事件的概率，$Pr(a|b)$ 是在事件b发生条件下，发生事件a的概率。

在图2-1中，我们列出掷两个骰子可能出现的所有情况，并显示了得到从2到12的所有不同方法。可以看到有6种方法可得到数目7，由于这些事件的每一个都是等概率的（36分之1），那么，7是掷两个骰子最容易出现的情况。图2-2用图的方式表示图2-1中包含的信息。

在模式分类中，我们最感兴趣的是出现某种度量的概率。然而对一个连续性函数要画出类似于图2-2的图是有问题的。例如，“一个人身高6尺的概率是多少？”这样的问题是没有意义的，显然，对这个问题的回答是零，因为发生的概率是无穷小（“一个人身高为6.314159267尺的可能性是多少”是同样的问题）但是，从直觉上我们知道一个人身高为6尺的可能性，要比身高为10尺的可能性大。因此需要有对似然率的直觉进行量化的方法。

数字之和		得到该数字和的方法数
0		**0**
1		**0**
2	1–1	**1**
3	2–1, 1–2	**2**
4	1–3, 3–1, 2–2	**3**
5	2–3, 3–2, 4–1, 1–4	**4**
6	1–5, 5–1, 2–4, 4–2, 3–3	**5**
7	3–4, 4–3, 2–5, 5–2, 1–6, 6–1	**6**
8	2–6, 6–2, 3–5, 5–3, 4–4	**5**
9	3–6, 6–3, 4–5, 5–4	**4**
10	4–6, 6–4, 5–5	**3**
11	6–5, 5–6	**2**
12	6–6	**1**

图2-1　掷两个骰子的所有可能方法

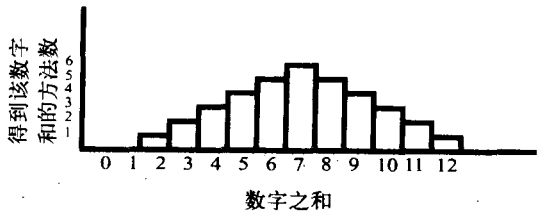


图2-2　图2-1中信息用图的方式表示

提问换成以下方式就有意义了，“一个人身高低于6尺的可能性是多少？”这种函数称为对某一度量z的概率分布函数：

$$P(x) = Pr(z < x) \tag{2-2}$$

当我们问“一个人身高低于x尺的概率是多少？”的问题时，就涉及到概率分布函数。另一个有意义的提法是“一个人身高在x与$x + \Delta x$之间的概率是多少？”对这种问题可用密度函数的术语回答。

$$Pr(x \leqslant h<x+\Delta x)=Pr\ (h<x+\Delta x) - Pr\ (h<x)=P\ (x+\Delta x) - P(x)$$

将上式除以Δx，并取极限$\Delta x \to 0$，就可以将概率密度函数定义为概率分布函数的导数：

$$p(x) = \frac{\mathrm{d}}{\mathrm{d}x} P(x) \tag{2-3}$$

9 图2-3显示掷两个骰子结果的概率分布函数。

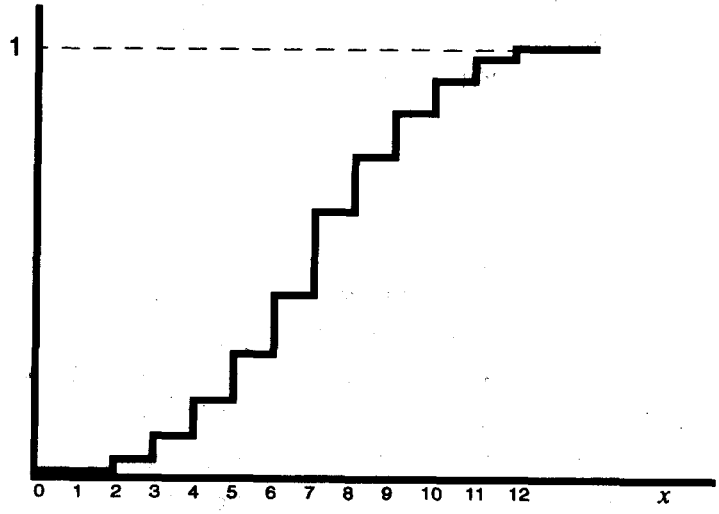


图2-3 与图2-2的概率分布，表示了掷两个骰子得到小于x数目的概率，注意这个曲线在出现可能性大的数目处较为陡峭

$p(x)$具有所期望的性质。它对连续值度量是良性定义的，并且在直觉上感觉最可能出现的度量处具有最大值。

还要注意的是：

$$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1 \tag{2-4}$$

这是必需满足的，因为某些值，不管它是哪一个事件是一定会出现的。

2.2 线性代数简单回顾

在这一节我们要简要回顾向量与矩阵的运算。一般情况下，向量用粗体表示，标量用小写罗马字表示，矩阵用大写罗马字表示。

向量通常用列向量表示。如果为了节省文件的空间，要将其写成水平形式，可采用转置

形式。例如，一个包含三个分量的向量可写成:

$$\boldsymbol{v}=[x_1 \quad x_2 \quad x_3]^{\mathrm{T}}$$

两个向量的内积是一个标量$\boldsymbol{v}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$。它的值是两个向量相应元素的乘积之和:

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=\sum_i a_i b_i$$

有时也用$<\boldsymbol{x}, \boldsymbol{y}>$表示内积，我们不大喜欢这种表示，因为它看上去与随机变量的期望值相像。有时也会看到用“点积”$\boldsymbol{x}\cdot\boldsymbol{y}$表示内积。 10

向量的幅值是$|\boldsymbol{x}|=\sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$。如果$|\boldsymbol{x}|=1$；则称其为“单位向量”，如果$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=0$，那么$\boldsymbol{x}$与$\boldsymbol{y}$“正交”。如果$\boldsymbol{x}$与$\boldsymbol{y}$是正交单位向量，称其为“正交归一”。

正交性的概念可以很容易地扩展到连续函数的情况，这只需把一个函数看作无限维的向量，列举出$f(x)$随x在a与b范围内变化的所有值即可。如果x是连续的，那么x在a与b范围内可能值的数目会有无穷多。但这并不构成问题，虽然我们不能列举它们，但我们仍然能将它们看成包含$f(x)$所有值的向量。此时，要将对有限维向量定义的求和概念换成积分，则内积可写成

$$\langle f(x), g(x)\rangle=\int_a^b f(x)g(x)\mathrm{d}x \tag{2-5}$$

正交与正交归一概念对这种内积定义仍然适用。如果这个积分为零，则该两个函数正交。所以从正交向量过渡到正交函数并不是难题。对于无限维来说，当然不能将正交可视化为“垂直”，所以要放弃把正交想像成“垂直”的这种做法，只需回顾这种定义并使用它。

设有n个向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \boldsymbol{x}_n$；如果有$\boldsymbol{v}=a_1\boldsymbol{x}_1+a_2\boldsymbol{x}_2+\cdots+a_n\boldsymbol{x}_n$则称$\boldsymbol{v}$为$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$的线性组合。

一组向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \boldsymbol{x}_n$被称为“线性独立”，条件是它们中任一个不可能写成其余向量的线性组合。

给定在$\Re^d$定义的d个d维线性独立向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_d$，那么在该空间的任何一个向量$\boldsymbol{y}$可写成$\boldsymbol{y}=a_1\boldsymbol{x}_1+a_2\boldsymbol{x}_2+\cdots+a_n\boldsymbol{x}_d$。

由于任何一个d维实数向量$\boldsymbol{y}$可以写成$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_d$的线性组合，则$\{\boldsymbol{x}_i\}$集称之为“基”集，并称这些向量“支起一个$\Re^d$空间”。任何一个线性独立的向量集都可以用作基集（符合必要充分条件）。通常正交归一的基集比较好用。

例如，下列两个向量形成一个$\Re^2$空间的基:

$$\boldsymbol{x}_1=[0\ 1]^{\mathrm{T}}\text{和}\boldsymbol{x}_2=[1\ 0]^{\mathrm{T}}$$

这是一个很熟悉的笛卡尔坐标系统。$\Re^2$的另一个基集是 11

$$\boldsymbol{x}_1=[1\ 1]^{\mathrm{T}} \quad \boldsymbol{x}_2=[-1\ 1]^{\mathrm{T}}$$

如果$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_d$覆盖$\Re^d$空间，$\boldsymbol{y}=a_1\boldsymbol{x}_1+a_2\boldsymbol{x}_2+\cdots+a_d\boldsymbol{x}_d$，那么向量$\boldsymbol{y}$的“成分”可由下式确定

$$a_i=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}_i \tag{2-6}$$

而a_i是$\boldsymbol{y}$在$\boldsymbol{x}_i$上的投影。从一个简单的笛卡尔几何角度看，式（2-6）的内积确实是一种投影，正如图2-4所示。然而只要用到式（2-6），“投影”这个词就可以用，甚至可以在更加一般的意义上用（例如，傅里叶级数的系数）。

这里所讲的向量空间中的向量是实数值的向量。

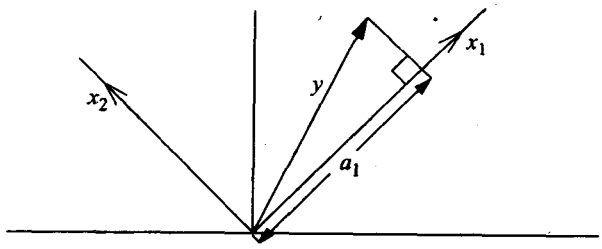


图2-4 x_1与x_2是正交归一基，y到x_1的投影长度为a_1

2.2.1 线性变换

一个“线性变换”A是一个矩阵，设A是$m \times d$的矩阵。如果将其用到一个$\Re^d$的向量$\boldsymbol{x}$，$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$，那么$\boldsymbol{y} \in \Re^m$。于是$\boldsymbol{A}$从某一个向量空间$\Re^d$取一个向量，在$\Re^m$中产生一个向量。如果向量$\boldsymbol{y}$只能由$\boldsymbol{A}$作用于$\Re^d$中某一个唯一的向量产生，则称为“一对一”的映像。假设$\Re^m$中没有一个向量不能由$\boldsymbol{A}$作用于$\Re^d$中某个向量产生，此时称$\boldsymbol{A}$为“映上”。如果$\boldsymbol{A}$是一对一映像及映上，则$\boldsymbol{A}$的逆$\boldsymbol{A}^{-1}$存在。如果两个矩阵$\boldsymbol{A}$与$\boldsymbol{B}$的乘积是符合规则的，即$\boldsymbol{C} = \boldsymbol{AB}$，则称$\boldsymbol{A}$与$\boldsymbol{B}$是“相适应”的。

有一些（经常忘掉的）重要特性：如果$\boldsymbol{A}$与$\boldsymbol{B}$相适应，那么

$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \tag{2-7}$$

如果$\boldsymbol{A}$与$\boldsymbol{B}$都可逆，则还有

12

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \tag{2-8}$$

还有一些其他的有用性质：

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{BA}) \quad 和 \quad \mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$$

条件是$\boldsymbol{A}$与$\boldsymbol{B}$都是方阵。如果矩阵$\boldsymbol{A}$满足

$$\boldsymbol{AA}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I} \tag{2-9}$$

则矩阵的转置显然就是它的逆，而$\boldsymbol{A}$也称之为“正交归一变换”(OT)，它在几何意义上相当于一个旋转。如果$\boldsymbol{A}$是一个$d \times d$正交归一矩阵，则$\boldsymbol{A}$的列是正交归一的，线性独立的，并形成覆盖$\Re^d$空间的一组基。对$\Re^3$，三个常用的OT是绕三个笛卡尔坐标轴的旋转。

$$\boldsymbol{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \quad \boldsymbol{R}_y = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \quad \boldsymbol{R}_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

如果$\boldsymbol{R}$是一个OT，且有$\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}$，则

$$|\boldsymbol{y}| = |\boldsymbol{x}| \tag{2-10}$$

矩阵$\boldsymbol{A}$是正定的，如果存在

$$\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0 \; \forall \boldsymbol{x} \in \Re^d, \quad \boldsymbol{x} \neq 0$$

$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$称之为二次型。

二次型的导数特别有用：

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^{\mathrm{T}}Ax)=(A+A^{\mathrm{T}})x$$

由于这里提到了求导数，则也可以提提另外一些有关积分的运算：

设f是x的标量函数，$x\in\mathfrak{R}^d$，那么定义

$$\frac{\mathrm{d}f}{\mathrm{d}x}=\begin{bmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_d}\end{bmatrix}^{\mathrm{T}} \tag{2-11}$$

并称之为“梯度”。只要我们谈到图像中的边缘，而$f(x)$表示空间二个方向的亮度函数，那么“梯度”会经常提到。 13

如果f是x的向量函数，那么导数是一个矩阵：

$$\frac{\mathrm{d}f^{\mathrm{T}}}{\mathrm{d}x}=\begin{bmatrix}\frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1}\\ \cdots & \cdot & \cdots & \cdots\\ \cdots & \cdot & \cdots & \cdots\\ \cdots & \cdot & \cdots & \cdots\\ \frac{\partial f_1}{\partial x_d} & \frac{\partial f_2}{\partial x_d} & \cdots & \frac{\partial f_m}{\partial x_d}\end{bmatrix} \tag{2-12}$$

该矩阵称为雅可比（Jacobian）矩阵。

此外，如果f是标量函数，$x\in\mathfrak{R}^d$，则它的二阶导数矩阵

$$\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_d}\\ \cdots & \cdots & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial^2 f}{\partial x_d\partial x_1} & \frac{\partial^2 f_2}{\partial x_d\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d^2}\end{bmatrix} \tag{2-13}$$

称为黑塞（Hessian）矩阵。

2.2.2 求导运算

在此介绍另一种表示，即一个只包含求导运算的向量：

$$\nabla=\begin{bmatrix}\frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \cdots & \frac{\partial}{\partial x_d}\end{bmatrix}^{\mathrm{T}} \tag{2-14}$$

值得注意的是，它只是一个操作符，不是向量。用它可以进行线性代数一类的事情，但它本身并没有值，甚至没有什么意义——它只在被应用时才有意义。在本书中主要涉及二维图像，使用运算符的二维形式。

$$\nabla=\begin{bmatrix}\frac{\partial}{\partial x} & \frac{\partial}{\partial y}\end{bmatrix}^{\mathrm{T}} \tag{2-15}$$

将其用到一个标量f，就可得到有含义的向量，即f的梯度：

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}^{\mathrm{T}}. \tag{2-16}$$

14 与此相似，可以使用内（点）积定义散度。（在下述定义中，只采用在式（2-16）中定义del运算符的二维形式，但相同的概念可以用到任意维数的运算中）：

$$\mathrm{div} f = \nabla f = \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} \tag{2-17}$$

同样可以将del运算符用到与矩阵的外积中：

$$\nabla \times f = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{bmatrix} [f_1 \quad f_2] = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_2}{\partial x} \\ \frac{\partial f_1}{\partial y} & \frac{\partial f_2}{\partial y} \end{bmatrix} \tag{2-18}$$

2.2.3 特征值与特征向量

如果矩阵A与向量x是相适应的，则可以写出“特征方程式”：

$$Ax = \lambda x, \ \lambda \in \Re \tag{2-19}$$

由于Ax是一个线性运算，A可以看作将x映像到它自身，只在长度上有所变化。满足式（2-19）的“特征值”λ可以不止一个。对$x \in \Re^d$，A有且只有d个特征值（它们并不一定不等）。这些值可用$\det(A - \lambda I) = 0$求解得到。（但对$d>2$,我们不推荐这种方法，而建议使用一个数值运算包来代替。）

给定一个满足式（2-19）的特征值λ，对应的x称为它对应的“特征向量”。

2.3 函数最小化简介

函数最小化在工程问题中是无所不在的：人们经常设法找到一组参数，使得这些参数的某个函数值最小。这个问题可以符号化地表示为：找到向量x，使某个函数$H(x)$达最小值：

$$\hat{H} = \min_x H(x) \tag{2-20}$$

15 其中x是某个d维参数向量，H是x的标量函数，常称为“目标函数”，通常将能使H最小化的x表示为：

$$x = \arg\min_x H(x) \tag{2-21}$$

使函数最小化的最直接的方法是将其导数设为零：

$$\nabla H(x) = 0 \tag{2-22}$$

其中∇是梯度算子——一组偏导数，式（2-22）得到一组方程式，每一个对应于x的一个成分。它们必须同时满足：

$$\frac{\partial}{\partial x_1}H(\boldsymbol{x})=0$$
$$\frac{\partial}{\partial x_2}H(\boldsymbol{x})=0$$
$$\cdots \tag{2-23}$$
$$\frac{\partial}{\partial x_d}H(\boldsymbol{x})=0$$

这种方法只有在式（2-23）有解时才实用。这只是在d=1或H至多是$\boldsymbol{x}$的二阶函数时才有用。

练习

找到使式$H = ax_1^2 + bx_1 + cx_2^2 + dx_3^2$，最小化的向量$\boldsymbol{x} = [x_1, x_2, x_3]^{\mathrm{T}}$，其中$a$、$b$、$c$与$d$是已知常数。

解

$$\frac{\partial H}{\partial x_1}=2ax_1+b$$
$$\frac{\partial H}{\partial x_2}=2cx_2$$
$$\frac{\partial H}{\partial x_3}=2dx_3$$

在$x_3 = x_2 = 0$，$x_1 = \dfrac{-b}{2a}$时最小化。

如果H是阶高于2的某个函数，或是超越函数，将导数设为零的技术就没有用（至少一般来说），则我们必须求助于数值运算。其中最常用的是梯度下降。

在一维时，梯度的用途是显而易见的，在某点$x^{(k)}$（图2-5）导数指向偏离最小值方向。也就是说在一维时，在斜坡为上坡时其符号为正。 16

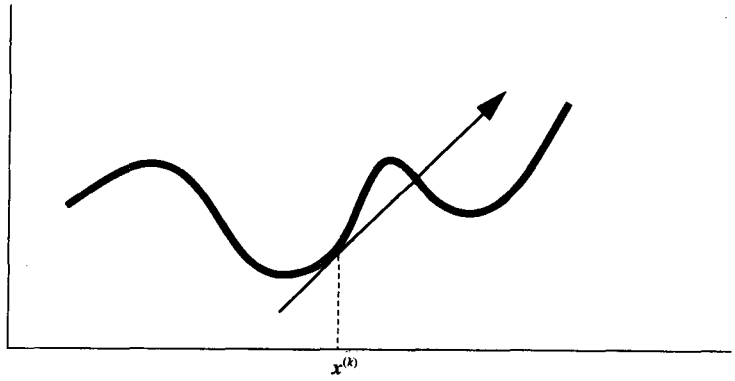


图2-5 导数的符号总是指向偏离最小值的方向

于是为了找到一个新的点x^{k+1},令

$$x^{(k+1)}=x^{(k)}-\alpha\left.\frac{\partial H}{\partial x}\right|_{x^{(k)}} \tag{2-24}$$

其中α是某个“小”常数。

对一个具有d个变量的问题，则有

$$x^{(k+1)} = x^{(k)} - \alpha \nabla H(x)|_{x^{(k)}} \tag{2-25}$$

2.3.1 牛顿-拉弗森方法

在式（2-25）中如何选α的值并不明确。如果α值太小，式（2-25）的迭代收敛过程会太长；如果α值取得太大，算法可能会不稳定，并且找不到最小值。

可以考虑用有名的牛顿-拉弗森（Newton-Raphson）求根的方法估计α值：（在一维情况下）将$H(x)$函数在$x^{(k)}$处展开成泰勒级数，并认为所有高阶项都为零而截掉：

$$H(x^{(k+1)}) = H(x^{(k)}) + (x^{(k+1)} - x^{(k)})H'(x^{(k)})$$

由于期望$x^{(k+1)}$处的H函数为零，令

$$H(x^{(k)}) + (x^{(k+1)} - x^{(k)})H'(x^{(k)}) = 0 \tag{2-26}$$

并且找到$x^{(k+1)}$来估计这个根，要利用

$$x^{(k+1)} = x^{(k)} - \frac{H(x^{(k)})}{H'(x^{(k)})} \tag{2-27}$$

然而在优化问题中，我们并不是要求根，而是要使一个函数值最小化，所以怎样才能知道求根能帮助我们呢？由于函数值最小化是它导数的根，则算法变为：

17

算法：梯度下降法

$$x^{(k+1)} = x^{(k)} - \frac{H'(x^{(k)})}{H''(x^{(k)})} \tag{2-28}$$

在高维情况下，式（2-28）变成：

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{H}^{-1}\nabla \boldsymbol{H} \tag{2-29}$$

其中H是二阶导数的黑塞矩阵，它在本章已提到过，即

$$\boldsymbol{H} = \left[\frac{\partial^2}{\partial x_i \partial x_j} H(\boldsymbol{x})\right] \tag{2-30}$$

例

给定一组x, y数据对$\{(x_i, y_i)\}$以及一个如下形式的函数

$$y = ae^{bx} \tag{2-31}$$

找到参数a与b，使下式最小化

$$H(a,b) = \sum_i (y_i - ae^{bx_i})^2 \tag{2-32}$$

解：通过观察$\ln y = \ln a + bx$，以及让$g = \ln y$与$r = \ln a$，可以使该问题求解用线性方法解决。利用以上替代式（2-32）变成：

$$H(r,b) = \sum_i (g_i - r - bx_i)^2 \tag{2-33}$$

$$\frac{\partial H}{\partial b} = 2\sum_i (g_i - r - bx_i)(-x_i) \tag{2-34}$$

$$\frac{\partial H}{\partial r}=2\sum_i(g_i-r-bx_i)(-1) \tag{2-35}$$

令式（2-34）为0，则有

$$\sum_i g_i x_i-\sum_i rx_i-\sum_i bx_i^2=0 \tag{2-36}$$

或

$$r\sum_i x_i+b\sum_i x_i^2=\sum_i g_i x_i \tag{2-37}$$

从式（2-35）又有

$$\sum_i g_i-r\sum_i 1-b\sum_i x_i=0 \tag{2-38}$$

18

或

$$Nr+b\sum_i x_i=\sum_i g_i \tag{2-39}$$

其中N是数据点的数目。式（2-37）与式（2-39）是两个联立线性方程，有两个未知数，这是很容易求解的（[2.2，2.3，2.4]中有更复杂的下降算法，如共轭梯度法）。

2.3.2 局部最小与全局最小

梯度下降法有一个严重的问题：它的解强烈依赖于初始点。如果在一个“谷”中开始，它将找到该谷的底。但是无法保证该最小值是最低的或“全局”最小值。

在继续讨论之前，先区分以下两种非线性最佳问题对我们而言是比较有益的。

- 组合最优：在这种情况中，变量具有离散值，一般为0与1，对由d个二值变量组成的$\boldsymbol{x}$，存在2^d种可能值。那么在原理上可以通过产生$\boldsymbol{x}$的每一个可能的值，以及计算相应$H(\boldsymbol{x})$来选择最小值。这种“穷尽搜索”一般是不现实的，因为可能的值会按指数激增。模拟退火是一个解决组合优化问题的好方法。
- 图像优化：图像具有一种特殊性质：每个像素只受其邻域影响（这将在后文中详细说明）。然而，像素值是连续值，而且会有成千上万个值。而均值场退火对此类问题的求解是最合适的。

2.3.3 模拟退火

在后面讨论的最小化技术中大部分都以“模拟退火”（SA）为基础（欲知更多细节，请参看Aarts和Van Laarhoven的书[2.1]）。

算法：模拟退火

（随机）选择一个$\boldsymbol{x}$的初始值，以及一个T的初始值$T>0$，当 $T>T_{\min}$，执行以下步骤

1) 产生一个$\boldsymbol{x}$的邻近点$\boldsymbol{y}$（邻近点的确切定义接下来会讨论）。

2) 如果$H(\boldsymbol{y})<H(\boldsymbol{x})$，则用$\boldsymbol{y}$代替$\boldsymbol{x}$。

19

3) 否则计算 $P_y=\exp\left(-\frac{H(\boldsymbol{y})-H(\boldsymbol{x})}{T}\right)$。如果$P_y>R$，那么用$\boldsymbol{y}$替代$\boldsymbol{x}$，其中$R$是在0到1之间均匀分布的随机数。

4) 略为降低T值，返回第1步。

模拟退火的工作机理

在组合优化中理解模拟退火是最容易的。在这种情况下，一个向量x的“邻近点”是另一个向量x_2[⊖]，而x_2与x只差一个分量（离散值）。于是如果x是二值的d维向量，可以选择一个邻近点$y = x \oplus z$，其中z是一个二值向量，其中只有一个元素是非零的，该元素是随机选择的。$\oplus$表示异或。

在算法的第2步，执行的是下降，于是我们“老在下山”。

但是第3步却提供了有时向山上爬的机制。先不考虑参数T的影响，而只管分子项，那么如果y代表一个上山移动，则接收y的概率正比于$e^{-(H(y)-H(x))}$。上坡移动可以发生，但是随着上山步幅增大，发生的可能性呈指数下降。上升移动的似然率还要受T的严重影响。若T值很大，那么$\frac{H(y)-H(x)}{T}$以及$P_y \approx 1$。于是所有的上升与下降移动都可能发生。随着T值逐渐减小，接收上升步的可能性会逐渐减小，而在T取低的值时（$T \ll (H(y) - H(x))$），这种移动就基本上不可能。

这种过程可以看成与一种物理过程相类似。在这种物理过程中，每个变量的状态（1与0）与粒子的自旋（上或下）相类似，在高的温度时，粒子随机改变状态，如果温度逐渐降低，就会达到最小能量状态。算法第4步中的参数T就与温度相似（因而也称为温度），因此这种最小化技术也称为“模拟退火”。

2.4 马尔科夫模型

尽管在机器视觉中主要关注的是发生在空间距离之间的交互关系，但马尔科夫（Markov）过程还是最容易用时间的术语来描述。其概念即某个事件发生的概率与刚发生过的某一件事情有关。在第6章主要讨论马尔科夫过程在去除噪声与分割中的应用，同时马尔科夫过程，在许多方面的问题中得到应用，包括字符识别[16.1]。

我们从介绍最简单的马尔科夫模型马尔科夫链开始，这种类型的模型适合于何时能够分辨一串事件的情况。例如，通过计算机网络收到一串符号，或一串字输入到自然语言处理器中。如果将接收到的符号用$y(t)$表示，其中t表示接收到该符号的（离散）时刻。于是$y(1)$在$y(2)$之前收到。如果让w表示符号所属的类别，$w \in \{w_1, w_2, \cdots, w_c\}$，$c$是可能的类别数。作为一个例子，将接收到的信号用0或1表示，就像通信系统中那样。

20

此时，我们关注的是$y(t)$属于某一具体类别的概率。例如第k个符号是“1”的概率是多少？由于这个概率与（符号串）历史有关，这个问题可表示为在给出$y(t)$前N个符号的条件下，$y(t)$是“1”的概率，其中用w_1表示类别1：

$$P(y(t) \in w_1 \mid y(t-1) \in w_{t-1}, y(t-2) \in w_{t-2}, \cdots, y(t-N) \in w_{t-N}) \tag{2-40}$$

一个具体一点的例子是，$y(t)$前4个符号是0110，现问$y(t)$为1（即w_1）的概率，这就可表示成：

$$P(y(k) \in w_1 \mid (y(k-1) \in w_0, y(k-2) \in w_1, y(k-3) \in w_1, y(k-4) \in w_0))$$

⊖ x的邻域集由与x的Hamming距离为1的所有向量组成。

实际中与历史有关的情况，有多种可能，有时$y(t)$是1的概率跟全部历史有关，有时则只跟其前若干个时刻收到的符号有关，一个我们特别感兴趣的情况是只跟其前一时刻的符号有关。此时，可以说对符号$y(t)$属哪一类的概率在给定全部历史的条件下，只取决于其之前接收到的最后一个信息。

$$P(y(k)\mid y(k-1)) = P(y(k)\mid (y(k-1), y(k-2), \cdots)) \tag{2-41}$$

式中已在符号表示上作了简化，将$y(k)$直接表示成它所属的类别。如果在所有历史条件下，事件$y(t)$的概率只取决于其之前收到的最后一个信息，这就称为（一阶）马尔科夫过程。由于本章只讨论一阶过程，因此以后就不再提一阶两个字[⊖]。

这个过程就隐含如下关系：

$$P(y(N)\in w_N, \cdots, y(1)\in w_1) = \left\{\prod_{t=2}^{N} P(y(t)\in w_t \mid (y(t-1)\in w_{t-1}))\right\} P(y(1)\in w_1)$$

21

如果假设只有两个类别，表示成0与1，那么只存在4种“转移概率”，表示成

$$P(y(t)=0\mid y(t-1)=0) \equiv P_{00}$$
$$P(y(t)=0\mid y(t-1)=1) \equiv P_{01}$$
$$P(y(t)=1\mid y(t-1)=0) \equiv P_{10}$$
$$P(y(t)=1\mid y(t-1)=1) \equiv P_{11}$$

一般情况下，可能有不止两个类别，所以可将转移概率表示成P_{ij}，并且用一个$c\times c$矩阵P来描述马尔科夫链，P的元素是P_{ij}。在第6章讨论马尔科夫随机场时，我们将从另一角度来看待马尔科夫过程。

作业2.1

矩阵P是对称的吗？为什么？P具有哪些有趣的性质？它的每一行（或列）加在一起会怎么样？

2.4.1 隐马尔科夫模型

隐马尔科夫模型在许多应用中都出现，例如在语音识别中就应用得很成功。在一个隐马尔科夫模型中，假设有不止一个转移矩阵，并且有一个无法（直接）度量的（隐）过程，控制转移矩阵之间的转换。这种转换过程本身就具有统计性质，并且一般也假设为一个马尔科夫过程。这种情况在图2-6中表示，其中开关的状态决定了输出$y(t)$是与马尔科夫过程1连呢，还是马尔科夫过程2？这种开关可以设想成受控于一个有限状态机（FSM），在每一时刻，它可以停留在同一种状态或发生转换，如图2-7所示。 22

我们的问题是：如果我们观察到一个符号序列

$$Y = [y(t=1), y(t=2), \cdots] = [y(1), y(2), \cdots]$$

那么我们能从中推断出什么？转移概率？状态序列？FSM的结构？控制FSM的规则？下面从估计状态序列开始。

⊖ 严格地讲，这是个一阶马尔科夫过程，但在本章我们不会涉及任何其他阶的马尔科夫过程。

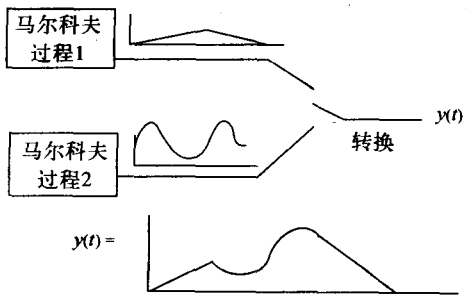


图2-6 一个隐马尔科夫模型可以看成在两个信息之间随机转换的过程

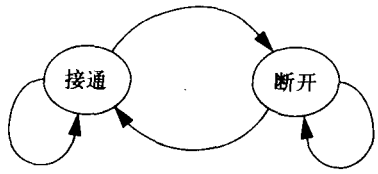


图2-7 起开关作用的有限状态机

估计状态序列

令$S(t)t = 1,\cdots, N$表示与度量$y(t)$有关的状态，并将状态序列表示成$S=[s(1), s(2), \cdots, s(N)]$，其中每个$s(t)\in\{s_1, \cdots, s_m\}$。我们要搜索出一个状态序列$S$，它在给定度量$Y$条件下成立的条件概率最大。这个概率可写成$P(S|Y)$。

使用贝叶斯规则

$$P(S|Y) = \frac{p(Y|S)P(S)}{p(Y)} \tag{2-42}$$

由于假设状态构成一个马尔科夫链，所以有

$$P(S) = \left[\prod_{t=2}^{N} P_{s(t),s(t-1)}\right] P_{s(0)} \tag{2-43}$$

我们暂且提出一个不可相信的假设，即输出的概率密度只取决于状态，用$p(y(t)|s(t))$表示这种关系。那么该序列的后验条件概率可写成:

$$p(Y|S)P(S) = \left[\prod_{t=1}^{N} p(y(t)|s(t))\right]\left[\prod_{t=2}^{N} P_{s(t),s(t-1)}\right] P_{s(0)} \tag{2-44}$$

定义$P_{s(1),\ s(0)} \equiv P_{s(0)}$，则式（2-44）可简化为

23

$$p(Y|S)P(S) = \prod_{t=1}^{N} p(y(t)|s(t))\, P_{s(t),s(t-1)} \tag{2-45}$$

回过来观察式（2-42）可以看到S的选择并不影响其分母，因此只需找到使下式最大化的序列S

$$E = \prod_{t=1}^{N} p(y(t)|s(t)) P_{s(t),s(t-1)} \tag{2-46}$$

值得高兴的是，的确有一个算法可以解这个最大化问题，这个算法是维特比（Viterbi）算法，它有许许多多的应用。它将在下一节中讨论。

2.4.2 维特比算法

这是一个用在某种特定优化问题的算法，其中每一个状态$s(t)$只有两个相邻状态$s(t+1)$与

$s(t-1)$。在这种限定条件下可以得到一个有效的算法。

首先，由于在几乎所有场合，用加法来代替乘法，总能使问题变得容易一些。所以定义一个使用对数的新目标函数：

$$L \equiv \ln E \equiv \sum_{t=1}^{N}(\Psi(t)+\Theta_{i,j}) \tag{2-47}$$

其中$\Psi(t)=\ln p(y(t)|s(t)$以及$\Theta_{i,j}=\ln p_{i,j}$。

为了方便，将一组可能的序列用如图2-8表示的图表示。某一个特定序列用图中的一条路径表示。作为一个例子，对$N=4$，$m=3$的这样一幅图来说，路径$[s_1, s_2, s_3, s_1]$表示在图2-9上。

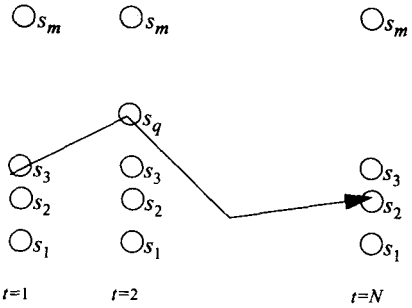


图2-8 每种可能的序列可以设想成穿过这样一幅图的一个路径

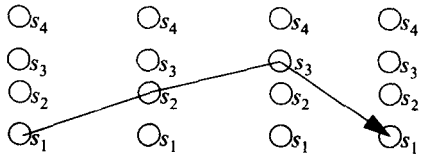


图2-9 在一个有4种状态与4个时刻的值的问题中的一条路径

24

像这样的一条路径，意味着函数有一组值。对图中的每一个结点，都附有Ψ的一个值。假设度量值$y(1)=2$，$y(2)=1$，$y(3)=2.2$，$y(4)=1$，则函数Ψ的值

$$\Psi=\ln p(y(1)=2\,|\,s(1)=s_1)+\ln p(y(2)=1\,|\,s(2)=s_2)$$
$$+\ln p(y(3)=2.2\,|\,s(3)=s_3)+\ln p(y(4)=1\,|\,s(4)=s_1)$$

在图中每条边也都有一个值，函数Θ由相应的转移概率决定。所以贯穿该图的每一个可能的路径都有一个相应的目标函数值L。我们采用归纳法来叙述该算法搜索最佳路径的过程：假设在时刻t我们已经找到每个结点的最佳路径，其成本为$LB_i(t)$，$i=1$，…，m。那么，可通过下式计算出从t时刻每个结点到$t+1$时刻每个结点的成本（m^2次计算）：

$$L_{ij}^{t}(t+1)=LB_i(t)+\Psi(y(t+1)\,|\,s_j(t+1))+\Theta_{i,j} \tag{2-48}$$

而在$t+1$时刻到达j结点的最佳路径是对应于其中的最大值。当最终到达时刻N时，终止最佳路径的结点就是最终的结点。

该算法的计算复杂度是Nm^2，它远比m^N要小，而m^N是穷尽搜索全部可能路径算法的复杂度。

2.4.3 马尔科夫输出

在上述讨论中，曾假定某个输出的概率只取决于其状态，其实根本不需要这种限制。输出本身可以是一个马尔科夫过程。如果状态发生改变，第1个输出仅取决于该状态，但随后如果该状态保持不变，输出则可服从马尔科夫链。这种情况下可将问题表示成与以上相同的方式，并用维特比算法来求解。

2.4.4 估计模型参数

剩下的最后一个问题是：在给定观察序列Y条件下，如何估计条件输出概率$p(y(k)|s(k))$，以及转移概率p_{ij}？

为此我们首先需要将$y(t)$取连续值的情况转换成离散值的情况，使得我们可以使用概率而不是概率密度。假设$y(t)\in\{y_1, y_2, \ldots, y_r\}$，那么就可定义一个输出概率矩阵$\prod = [\pi_{k,l}]$，$k = 1, \cdots, m$；$l = 1, \cdots, r$，它表示如果状态为$s_k$，观察输出为$y_l$的概率。

25

定义

$$P_{i,j|Y}(t) = Pr((s(t-1) = i, s(t) = j)|Y) \tag{2-49}$$

也就是说要计算在给定观察序列条件下，即在时刻t时从状态i转移到状态j的概率。如果知道转移概率P_{ij}以及输出概率$\pi_{k,l}$，就可以使用2.4.2节的方法计算该值。假设这些的确已知，那么可以通过对所有输入得到的概率的均值来估计转移概率。

$$P_{i,j} = \frac{\sum_{t=2}^{N} P_{i,j|Y}(t)}{\sum_{t=2}^{N} P_{j|Y}(t)} \tag{2-50}$$

其中，由于到达j结点可来自其他所有结点，则有：

$$P_{j|Y}(t) = \sum_{i=1}^{N} P_{ij|Y}(t) \tag{2-51}$$

于是通过再次平均所有观察概率，来重新估计观察概率[⊖]：

$$\pi_{k,l} = \frac{\sum_{t=1, y(t)=g_l}^{N} P_{k|Y}(t)}{\sum_{t=1}^{N} P_{k|Y}(t)} \tag{2-52}$$

这是一个迭代算法，在每次迭代中使用（2-50）与（2-52）式来修改参数，然后用（2-49）与（2-51）式来修改条件概率。该过程重复下去直至收敛。

2.4.5 隐马尔科夫模型的应用

26

隐马尔科夫模型在语音识别与文件内容识别[17.29]中找到了许多应用。

作业2.2

在哪本小说中曾出现一个名叫Markov Chaney的人物？

作业2.3

确定一个绕z轴旋转30°的正交变换，并证明得到的矩阵构成$\Re^3$的基。

作业2.4

试证：式（2-10）。

⊖ 疑原书式（2-52）笔误，故纠正。——译者注

提示：使用式（2-7）；式（2-9）也可能有用。

作业2.5

试证：一个正定矩阵具有正的特征值。

作业2.6

是否有一个唯一的值x，能使函数$y=xe^{-x}$达到最小？如果是的话，你能否通过求导并设导数为0找到x值？假设这个问题需要用梯度下降法，写一个相应算法。

作业2.7

我们需要使用梯度下降法解一个最小化问题。要最小化的函数是$\sin x + \ln y$，下面式子中哪一个是在梯度下降法中需要用的梯度表达式？

(a) $\cos x + \dfrac{1}{y}$ (b) $y = -\dfrac{1}{\cos x}$ (c) $-\infty$

(d) $\begin{bmatrix} \cos x \\ 1/y \end{bmatrix}$ (e) $\dfrac{\partial}{\partial y}\sin x + \dfrac{\partial}{\partial x}\ln y$

作业2.8

（a）写一个使用梯度下降的算法，该算法能找到向量$[x, y]^{\mathrm{T}}$，使函数$z = x\exp(-(x^2 + y))$达到最小。 27

（b）为此算法编写计算机程序。

作业2.9

确定函数$\sin x$与$\sin 2x$是否是正交归一或正交函数。

参考文献

[2.1] E.H.L. Aarts and P.J.M. van Laarhoven. *Simulated Annealing: Theory and Applications*, Dordrecht, Holland, Reidel, 1987.

[2.2] R.L. Burden, J.D. Faires, and A.C. Reynolds. *Numerical Analysis*, Boston, MA, Prindle, Weber and Schmidt, 1981.

[2.3] G. Dahlquist and A. Bjorck. *Numerical Methods*, Englewood Cliffs, NJ, Prentice-Hall, 1974.

[2.4] B. Gottfried and J. Weisman. *Introduction to Optimization Theory*, Englewood Cliffs, NJ, Prentice-Hall, 1973.

28

第3章 编写图像处理程序

Computer Science is not about computers any more than astronomy is about telescopes.

E. W. Dijkstra

人们可以用两种方法编写图像处理软件，这取决于人们要求在哪一方面实现最优。可以考虑以优化/最小化编程者时间的方式，或考虑使计算机时间最小的方式。在本课中不考虑计算机时间的优化，而认为编程人员的时间是宝贵的。因此我们遵循这样一个编程哲学，即在编程时间最短条件下编出正确可操作的程序。

本书作业规定使用C或C++语言，而不是MATLAB或JAVA。这是作者经过深思熟虑后做出的决定。对用户来说MATLAB隐藏了许多有关数据结构以及数据管理的细节。在教授这门课的过程中，此课程的变化已有多年，但作者发现这些方面的许多细节正是学生需要掌握的，它们能使得学生真正理解图像处理（尤其在像素这一级）的内涵。

3.1 图像文件系统软件

要达到能很快地写出好的软件这个目标，可以通过使用图像文件系统（IFS）中的图像访问子程序来实现。IFS是一组子程序，以及基于这些子程序的应用的集合，它支持开发图像处理软件。

IFS的优点包括以下几点:

- IFS支持任何数据类型，包括字符（char）、无符号字符（unsigned char）、短整数、无符号短整数、无符号整数、浮点、双精度、复数浮点、复数双精度、复数短、以及结构。
- IFS支持任何图像尺寸，以及任何维数。可以把信号看作一维图像而进行信号处理。
- IFS在大多数现有计算机系统都有，包括PC机上的Windows，PC机上的Linux、SUN的Unix以及Macintosh上的OS-X[⊖]。在一个平台上写的文件，可以在任何一个别的平台读。将格式转换成适应别的平台的格式是由读程序完成的，不用用户干预。
- 有大量可用程序，包括两维傅里叶变换、滤波器、分割运算等。

3.1.1 IFS头部结构

所有的IFS图像都包括一个头部，其中包括各种有关图像的信息，诸如图像中点的数目、图像的维数、数据格式、每一维的单位与扫描方向等等。与图像有关的是图像的实际数据，图像头部包括到图像数据的指针。

用户操作一幅图像是通过调用IFS库中的某些函数；这些函数的一个变量是头部地址，根据头部提供的信息，IFS库的函数自动确定数据在何处，以及如何存取数据。除了存取图像数据外，IFS程序自动地管理在存储器中的空间分配来储存数据与头部。任何事情都是动态进行操作的；并没有固定维数的数组。这就使得用户可以摆脱在数组中存取数据的困难，在数组

⊖ 抱歉，IFS不支持OS-X之前的Macintosh操作系统。

尺寸不固定的情况下，用户自己做这种事是挺麻烦的。

图像的头部结构在文件<ifs.h>中定义，通过IFSHDR获取。为了操作一幅图像，用户仅仅需要发布一个指针到图像的头部结构（如IFSHDR*your_image; 或IFSIMG your_image; ），然后，用户只需调用某些IFS函数构造一幅新的图像，并且将指针设置到由这些函数返回的值中去。

3.1.2 某些有用的IFS函数

可以用ifspin将图从盘中读出。这个子程序会计算图像的尺寸，以及需要多大的存储器，并确定图像是用怎样的计算机写的。它将执行所有必要的数据转换（字节交换，浮点格式转换等），以及将其以适合你正使用的计算机的格式读进来。你用不着知道这些数据转换操作是如何进行的。与此相似，可以用ifspot将图像写到盘上。这些子程序会把IFS头部以及相关的图像数据从内存写到磁盘中去。

可以用ifsigp或ifsfgp以及ifsipp或 ifsfpp存取图像。这些子程序的名字表示IFS获取整数像素数据，IFS获取浮点数像素数据，IFS放置整数像素数据，以及IFS放置浮点像素数据。字Integer 与Floating指的是返回或写入的数据类型。例如

30

v=ifsfgp(img, x, y)

那么返回的是浮点数，而与图像数据类型无关。也就是说，子程序为你进行数据转换。同样ifsigp会返回一个整数，而不管内部数据是什么类型。当然这也会给你带来麻烦。设想一下内部数据类型是浮点数，而一幅图像由小于1的数据组成。那么从浮点转换成整数的过程会将图像数据截成零。

某些编程作业可能要用三维数据。这意味着用户必须用一组别的子程序存取图像：ifsigp3d、ifsfgp3d、ifsipp3d以及 ifsfpp3d。例如

y=ifsigp3d(img, frame, row, col)

3.1.3 带共性的问题

当学生初次使用IFS软件时，通常会有两个共同的问题：

1) **ifsipp(img, x, y, exp(－t*t))**会引起麻烦，因为ifsipp期待第4个变量是整数，而exp将返回一个双精度数。此时应该用ifsfpp。

2) **ifsigp(img, x, y, z)**是一个不恰当的格式。ifsigp期待3个变量，并且不通过检查输入图像的维数来确定变量数目。为了存取一个三维图像，可以使用指针或用ifsigp3d(img,x,y,z)，其中第2个变量是帧号。

3.2 图像处理程序的基本结构

图像可以看作一个二维数组。它们一般以光栅扫描的方式一个像素接一个地进行处理。为了操作一幅图像，两个嵌套的for循环是最常用的程序结构，如图3-1所示。

在这个例子中，用两个整数（row与col）作为图像中行与列的索引。row与col每增加1，就执行了图像像素的扫描，从左到右，从上到下。

如果图像有两个以上维数（例如多谱图像），则有第3个整数用来作为维数的索引。相应的也要使用三个嵌套for循环，如图3-2所示。

```
.....
int row, col;
.....
for ( row = 0; row < 128; row++)
{
    for( col = 0; col < 128; col++)
    {
        /* pixel processing */
        .....
    }
    .....
}
```

图3-1 基本程序结构：两个嵌套的for循环

```
.....
int row, col, frame;
.....
for (frame = 0; frame < 224; frame++)
{
    for ( row = 0; row < 128; row++)
    {
        for( col = 0; col < 128; col++)
        {
            /* pixel processing */
            .....
        }
        .....
    }
}
```

图3-2 基本程序结构：三个嵌套的for循环

31

3.3 好的编程风格

编写程序时要遵循一个重要的编程原则：所有编写的程序要能适应任何尺寸的图像。这里并不要求所写的程序能适应多种维数的图像（适应不同维数的程序也是可能的），或任何数据类型，只要求图像尺寸。这意味着对所编程序有一个要求，即不能发布一个静态数组，并将所有数据拷入数组内，相反必须使用图像存取子程序。

另一个重要的指导原则是，除了在极个别的场合外，不要使用全局变量。全局变量是在子程序外定义的一个变量，使用这种全局变量是一种不良的编程技巧，会比所有其他编程技巧引起更多的缺陷。好的编程实践要求子程序需要知道的任何一件事情都包括在它的参数表内。

遵循这些简单的编程指导原则，会使你易于高效率地写出通用性强的程序。随着编程技巧的增长，如果需要，可以考虑采用指针管理功能，以增加程序的运行速度。 32

除了以上提到编程指导原则外，对学生来说遵循缩排、留空格、和注释规则，会使得编码的“可读性”好。

常用的缩排格式有4种，即K&R格式（或核格式）、Allman格式（或BSD格式）、Whitesmiths格式以及GNU格式，如图3-3所示。本书采用Allman格式。在不同段落之间加些空行也能改善可读性。要强调注释的重要性。但是不要加太多的注释，以免打断编码的流畅性。一般说来，要在每个函数的顶部加一段注释，包括这个函数做什么用的描述，谁写的这个函数，如何调用该函数，以及该函数返回的是什么等内容。还应该增加对每个变量发布的描述。

```
for (<cond>) {
    <body>
}
```

a) K&R格式

```
for (<cond>)
{
    <body>
}
```

b) Allman格式

```
for (<cond>)
    {
        <body>
    }
```

c) Whitesmiths格式

```
for (<cond>)
    {
        <body>
    }
```

d) GNU格式

图3-3 4种常用的缩排格式

3.4 示例程序

再看一看IFS手册，它会帮助你理解这些示例程序。同时，还要注意我们使用的编程风格。示例程序中的注释可能偏详细了一些，但这只是为了教学用的。

图3-4是一个典型的程序，它也许是可以写成的最简单例子了。图3-5列出另一个例子，它实现与图3-4程序相同的功能，但是写的方式更加灵活，也因此能适用于不同尺寸的图像。

这两个示例程序都使用子程序调用ifsigp、ifsipp、ifsfgp与ifsfpp来存取图像，并用整数值或浮点数据。这些子程序的优点就是方便：不管图像以什么样的数据类型存储，ifsigp返回整数，而ifsfgp返回浮点数。在内部，这些子程序确定数据存在哪里、存取这些数据并进行相应转换。当然所有这些操作都消耗计算机时间，作为教学作业，我们强烈推荐使用这些子程序。然而用作产品，IFS支持使用指针对数据直接存取的方法，以增加编程人员时间来缩短运行时间。

33

```
/* Example1.c
    This program thresholds an image. It uses a fixed image size.
    Written by Harry Putter, October, 2006
*/
#include <stdio.h>
#include <ifs.h>
main( )
{
    IFSIMG img1, img2;          /* Declare pointers to headers */
    int len[3];                 /* len is an array of dimensions, used by ifscreate */
    int threshold;              /* threshold is an int here */
    int row,col;                /* counters */
    int v;

    /* read in image */
    img1 = ifspin("infile.ifs");    /* read in file by this name */

    /* create a new image to save the result */
    len[0] = 2;                 /* image to be created is two dimensional */
    len[1] = 128;               /* image has 128 columns */
    len[2] = 128;               /* image has 128 rows */
    img2 = ifscreate("u8bit",len,IFS_CR_ALL,0);        /* image is unsigned 8 bit */
    threshold = 55;             /* set some value to threshold */

    /* image processing part - thresholding */
    for (row = 0; row < 128; row++)
        for (col = 0; col < 128; col++)
        {
            v = ifsigp(img1,row,col);                  /* read a pixel as an int */
            if (v > threshold)
                ifsipp(img2,row,col,255);
            else
                ifsipp(img2,row,col,0);
        }

    /* write the processed image to a file */
    ifspot(img2, "img2.ifs");       /* write image 2 to disk */
}
```

图3-4 一个IFS程序的例子，用规定好的维数值以及预定数据类型来对图像进行阈值化

```
/* Example2.c
    Thresholds an image using information about its data type and the dimensionality.
    Written by Sherlock Holmes, May 16, 1885
*/
#include <stdio.h>
#include <ifs.h>
main( )
{
    IFSIMG img1, img2;          /* Declare pointers to headers */
    int *len;                   /* len is an array of dimensions, used by ifscreate */
    int frame, row, col;        /* counters */
    float threshold, v;         /* threshold is a float here */
```

图3-5 一个IFS程序的例子，使用输入图像确定的维数、每一维的尺寸以及数据类型来对图像进行阈值化

```
    img1 = ifspin("infile.ifs");        /*read in file by this name*/
    len = ifssiz(img1);                 /* ifssiz returns a pointer to an array of image dimensions*/
    img2 = ifscreate(img1->ifsdt,len,IFS_CR_ALL,0);
                                        /* output image is to be the same type as the input */
    threshold = 55;                     /* set some value to threshold */

    /* check for one, two or three dimensions */
    switch (len[0]) {
    case 1:                                         /* 1d signal */
        for (col = 0; col < len[1]; col++)
        {
            v = ifsfgp(img1,0,col);                 /* read a pixel as a float */
            if (v > threshold)
                ifsfpp(img2,0,col,255.0);           /* write a float */
            else                                    /* if img2 not float, will be converted*/
                ifsfpp(img2,0,col,0.0);
        }
        break;
    case 2:                                         /* 2d picture */
        for (row = 0; row < len[2]; row++)
            for (col = 0; col < len[1]; col++)
            {
                v = ifsfgp(img1,row,col);           /* read a pixel as a float */
                if (v > threshold)
                    ifsfpp(img2,row,col,255.0);     /* store a float */
                else
                    ifsfpp(img2,row,col,0.0);
            }
        break;
    case 3:                                         /* 3d volume */
        for (frame = 0; frame < len[3];frame++)
            for (row = 0; row < len[2]; row++)
                for (col = 0; col < len[1]; col++)
                {
                    v = ifsfgp3d(img1,frame,row,col);       /* read a pixel as a float */
                    if (v > threshold)
                        ifsfpp3d(img2,frame,row,col,255.0);
                    else
                        ifsfpp3d(img2,frame,row,col,0.0);
                }
        break;
    default:
        printf("Sorry I cannot do 4 or more dimensions\n");
    } /* end of switch */

    ifspot(img2, "img2.ifs"); /* write image 2 to disk */
}
```

图3-5 （续）

3.5 生成文件

编程应该使用makefiles（生成文件），这比只用输入命令的方式要好得多。如果使用Microsoft C++，Lcc或其他某种编译器来开发软件，那么makefiles是隐藏的，但知道它们如何操作仍是有益的。makefiles的主要功能是规定如何构造用户的程序，我们用图3-6的例子说明这一点。

图3-6的例子只是一个再简单不过的例子。它说明名字为myprogram的可执行文件只取决于一件事情——目标模块myprogram.o。然后它显示怎样由myprogram.o与IFS库生成myprogram。

```
myprogram: myprogram.o
    cc -o myprogram myprogram.o /CDROM/Solaris/ifslib/libifs.a

myprogram.o: myprogram.c
    cc -c myprogram.c -I/CDROM/Solaris/hdr
```

图3-6 一个生成文件（makefile）的例子，它编译一个程序并将其与IFS库链接

同样，myprogram.o是对源文件myprogram.c进行编译（没有连接）生成的，它使用了一个名字为hdr的头文件，这是在CDROM的“include”目录中找到的。注意：为了规定一个库，就像在link（链接）步骤中一样，必须规定库的名字（例如libifs.a），但是为了规定一个include文件（例如ifs.h），只要规定该文件所在的目录，因为文件名在#include程序指令中给出。

WIN32中的makefile如图3-7例子所示。该程序演示了生成程序的许多符号定义能力，也明确规定了编译器的位置。

由IFS生成的程序（ifsview除外）是要有平台支持的，也就是说，运行他山之石要在PC的MSDOS内部、在Linux，Solaris下的终端内部、或在Mac中使用OS-X操作系统。

```
CFLAGS=-Ic:\lcc\include -g2 -ansic
CC=c:\lcc\bin\lcc.exe
LINKER=c:\lcc\bin\lcclnk.exe
DIST=c:\ece763\myprog\lcc\
OBJS=c;\ece763\myprog\objs\
LIBS= ifs.lib -lm
# Build myprog.c
myprog:
    $(CC) -c $(CFLAGS) c:\ece763\myprog\mysubroutine1.c
    $(CC) -c $(CFLAGS) c:\ece763\myprog\mysubroutine2.c
    $(CC) -c $(CFLAGS) c:\ece763\myprog\myprog.c
    $(LINKER) -subsystem console -o myprog.exe myprog.obj mysubroutine1.obj
mysubroutine2.obj $(LIBS)
```

图3-7 一个WIN32makefile的例子

作业3.1 学会如何使用工具

Windows用户：

如果你使用的是WIN32计算机，为建立所有IFS文件的目录，打开MSDOS以及cd。你用什么样的程序在一幅图像中添加高斯噪声？（提示：查看INDEX）。把该程序应用到angio128.ifs的图像中。用ifsview显示结果。思考如何打印结果，并递交问题的答案与图像。

作业3.2 学会如何使用IFS

这个作业的目的是使你学会用计算机，并开始写程序。

使用ifsview来观察下列图像：

images / echo1

images / echo2

images / echo3

使用any2any将上述每个转换成TIFF（键入“any2any -h”）。

例如可以用：

any2any echo1.ifs echo1.tif tiff

重要的按语：echo1是三维的，TIFF是二维图像格式，当你照上面所示方法使用any2any时，你得到的只是第0帧！

你会看到三个echo图像是相似的，完成作业，并用一段文字描述图像看上去是怎样的。你不知道这些结构是可以理解的，所以你只需描述你所见到的，包括一幅图像（想想如何打印一个文件，我们希望你已经会做这些）。如果你恰巧知道这些结构，你的导师会对你有好印象的。

34 ~ 37

第4章 图像的生成与表示方式

Computers are useless. They can only give us answers.

Pablo Picasso

这一章我们描述图像是如何生成的，以及它们是以什么方式表示的。表示方式包括两部分，一部分是包含在图像内的信息的数学表示方法，另一部分是图像在数字计算机中如何存储与操作的方法。在本章中还介绍一种思考图像的方法，即将图像看作高度有变化的一个表面。这种表示方法对描述图像的性质与对图像的操作都是一个强有力的工具。

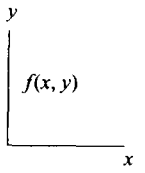
4.1 图像的表示方法

这一节讨论表达图像中信息的几种方法。这些表示方法有：图像形式、函数型式、线性、概率方法、空域频率以及关系表示。

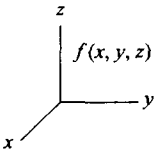
4.1.1 图像表示

一幅图像中信息的图像表示(iconic representation)是一幅图像。“噢，是的，一朵玫瑰就是一朵玫瑰。”（译者注：作者的意思是指似乎什么也没有说）。当人们了解了函数型式、线性与关系表示的含义时，就会体会到要用一个词表示一种表示方法，该种表示自身就是一幅图像。图像形式表示方法的例子有以下几种。

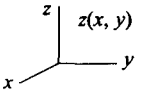
- 2维灰度图像，也称亮度图像。这种表示一般就称为“图像”，可能是彩色的或灰度级的。（要当心使用“黑白（图像）”，因为它们可能会理解成“二值的”。）一般用$f(x, y)$表示图像中一点$<x, y>$的亮度。注意x与y可以是整数（此时，指的是采样后的图像中离散的点；这些点称为“像素”（pixel）是“图像元素”的缩写形式），或实数（此时，把图像看成一个函数）。



- 3维（密度）图像。这种数据结构通常出现在医学图像中，例如CT（计算机层析图像），MRI（核磁共振图像）、超声图像等。在典型的三维图像中，每个像素表示一点的密度（从技术上讲，无法只在一点度量密度。而只能在该点邻域范围内度量密度，但其含义是清楚的）。通常将在某点$<x, y, z>$的密度用$f(x, y, z)$表示。



- $2\frac{1}{2}$维（距离）图像。在距离图像中，（物体）表面每一点的值表示一种距离，通常是指到摄像机的距离，或包含摄像机的一个平面的法向距离。于是在距离图像中表面点的位置用$z(x, y)$表示。例设传感器是激光飞点扫描器，那么图像的“亮度”实际上正比于一个激光脉冲到达表面，再反弹回到光源处一检测器所需的时间。实际上这种传感器测量的距离是两个偏转角的函数$r(\theta, \phi)$，从而生成一幅距离图像。然而从这些坐标转换成$z(x, y)$是很简单的。一般得到的数据（至少在本书中）都是经过校正生成的高度图（altitude image）。



4.1.2 函数表示

可以对任何一组数据点拟合一个函数，当我们对一幅图像进行度量时，它是由一个离散有限度量集$f(i, j)$组成，其中i与j是离散的像素坐标（整数），分别沿x与y方向。通过某种运算可以找到一个连续函数，例如，最小二乘运算能对这组数据进行最佳拟合。于是就可以用一个方程式，双元二次函数（biquadratic），来表示一幅图像（至少图像中的一个小区域）：

$$z = ax^2 + by^2 + cxy + dx + ey + f \tag{4-1}$$

或二次方程式（quardic）：

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0 \tag{4-2}$$

在式（4-1）中，一个变量是用其他变量定义的，故通常称为显式表达式，而式（4-2）是一个隐式表达[4.23]，它可以等价地用零集这个术语（zero set）$\{(x, y, z): f(x, y, z) = 0\}$表示。隐多项式有某些方便的性质。例如设想不在$f(x, y)$的零集内的一点（$x_0, y_0$），$f(x, y)$的零集的定义是：

$$f(x, y) = x^2 + y^2 - R^2 = 0 \tag{4-3}$$

那么如果将x_0与y_0代入$f(x, y)$，就得到一个非零的结果（因为该点不在其零集内）。如果这个值是负的，则知道（x_0, y_0）这个点是在曲线内部，否则在其外部[4.3]。这种内部/外部性质对所有可用多项式表达的封闭曲线（以及曲面）都成立。

39

4.1.3 线性表达式

将一幅用矩阵表示的图像展开成向量形式表示。例如所示的一幅2 × 2图像可以写成向量形式$f=[5\ 10\ 6\ 4]^T$。

5 10

6 4

4.1.4 概率表示

在第6章中，将把一幅图像表示成一个随机过程的输出。这样做的好处是可以使用强有力的数学工具，对一组带噪声的图像进行度量，给出其最佳版本的图像。

4.1.5 空域频率表示

设想用眼睛从左到右扫描图4-1中的两幅图像，并确定它们中哪一幅亮度变化更快一些。显然右边一幅图像的亮度变化更快一些。可以把亮度变化的概念，量化为每一寸纸的周期数表示。一个周期是指亮度从最大变成最小值又返回到最大值的过程。当然，如果你用一种固定的速率扫描图像，在你眼睛中看到的亮度变化用每秒周期数（赫兹）表示，是与所扫视图像内容的空域频率成正比的。

40

一幅图像的空域频率内容可以用带通滤波器来修改。例如图4-2显示了一张原图像与该图像经过低通滤波器过滤过的图像表示。该滤波器允许输入中的低频通过并输出，而高频

被压缩、可以看到频率响应是一种表征锐利度的方法。有许多高频成分的图像看起来会比较锐利。

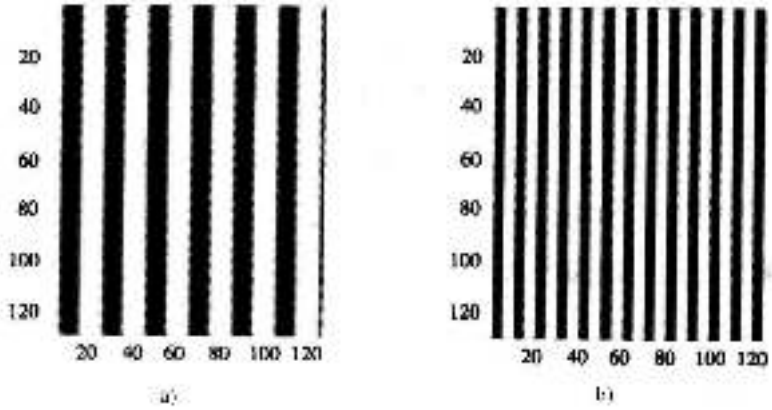


图4-1 a) 一幅图像的内容在水平方向有较低的频率 b) 一幅图像的内容在水平方向有较高的频率

图4-2 （左）一幅图像，（右）这幅图像的一幅低频图像表示。右图只在水平方向模糊，这种模糊是拍摄图像时摄像机移动造成的。注意水平边缘仍然是锐利的

尽管在本书内不大使用图像的频域表示，但读者需要频域的一些概念。

首先，正如你已经观察到的，空域频率是随方向而变化的。图4-1在水平方向比在垂直方向有更快的变化，更高的频率。并且一幅图像一般包含许多不同的空域频率，可以用二维傅里叶变换来提取图像的空域频率成分：

$$F(u,v)=\frac{1}{K}\sum_{x}\sum_{y}f(x,y)\exp[-i2\pi(ux+vy)] \tag{4-4}$$

其中K是一个适当的规范常数。傅里叶变换是用积分方式定义的，而式（4-4）的形式只定义在二维离散网格上（如，一维采样图像）。值得注意的是，傅里叶变换将x与y的函数转换成另一种函数，它的变量是u与v，u与v分别对应x与y方向的空域频率成分。如果对某一组值u_0与v_0的$F(u_0,v_0)$是一个大数，那么在该特定空域频率有大的能量。可以把术语“能量”理解为“所涉及的像素数目”或“该频率成分是否明显”。

第二个结论是，空域频率在图像中的分布是变化的，也就是说如果在图像中选一些子图像，就会发现这些子图像的傅里叶变换有明显的变化。

第三，考虑一下式（4-4）中的计算复杂度，一幅图像的傅里叶变换是一个空域频率u与v的函数，因而也可以看作是一幅图像（它的值是复数，但这不是什么问题）。如果一幅$N \times N$的图像，则某个空域频率u、v对的函数值，要通过对所有x与y对的函数值计算求和得到，其复杂度为N^2。如果频域采样数也是$N \times N$，则计算傅里叶变换的总复杂度是N^4。好在有一个称为快速傅里叶变换的算法，它计算每对u、v的函数值的复杂度为$N \log_2 N$，而不是N^2，因而计算量明显下降，因此有时某些计算在频域中要快一些。

最后还要提一下（空域中的）卷积运算与频域中的乘积运算等价，卷积将在下一章中讨论，而这种等价关系将在5.8节讨论。

4.1.6 关系表示方法（图表示法）

图是描述图像中内容的一个常用方法，这将在第12章进一步讨论。

4.2 数字图像

假设图像经过采样，也就是说x与y只取离散的整数值，同时假设f是量化过的（f只取一定范围内的整数值）。这种图像可以存储在计算机的存储器中，称为“数字图像”。

4.2.1 数字图像的生成

在讨论成像过程的文献中，提到许多种成像设备，包括析象管（dissector）、飞点扫描器、光导摄像管（vidicons）、正摄像管（orthicon）、氧化铅摄像管（plumbicon）、电荷耦合器件（CCD）、以及其他装置[4.6，4.8]。现在CCD在市场上占主导地位。成像设备在它们的成像方法以及所生成图像的性质这两方面都不相同。但是，所有这些装置都是将光能以相似的方式转换成电压。由于本章的目的是介绍图像分析的基本概念，因此我们只选择一种装置，即CCD，讨论它获取数字图像的方法。参考文献中有更精确的细节，相关例子可参考[4, 19]。

42

用硅装置成像

在CCD的表面用一组镜头来成像。当一个具有适当波长的光量子撞击装置的特定材料时，一个带电量子就产生了（形成电子-空穴对）。由于该材料的传导率很低，这些电荷趋向于停留在它们产生的区域。因此在CCD的局部区域内的电荷q可以用下式很好地近似：

$$q = \int_0^{t_f} i dt$$

其中i是入射光的强度，它用每秒光量子的数目度量。如果入射光在积分时间中是常数，则有$q = it_f$，其中t_f称为帧时间。

在摄像管类的装置中，累加的（正）电荷被扫描电子束中和，这种中和过程产生了电流，经放大成为视频信号。在CCD装置中电荷随着数字钟同步地从一个单元移至下一个。这种读出电荷的机制，不管是电子束还是电荷耦合，都是经过精心设计的，从而能使尽可能多的电荷被置0。累加电荷从0开始积分。电荷以一个正比于局部光强的速率累积，然后被读出。于是在一点度量到的信号既与该点的光强成正比，又与读操作之间的时间间隔成正比。

由于我们感兴趣的只是光强度，而不是积分时间，因此可以将积分时间的影响消除掉，

做法是使它在图像的每一处都相同。这种过程称为扫描，要求对设备上的每一点周期性地重复访问，将累积的电荷清零。一个可能是最直接，并且最常用的方法是一个从上到下，从左到右的扫描过程，称为光栅扫描（raster scanning），见图4-3。

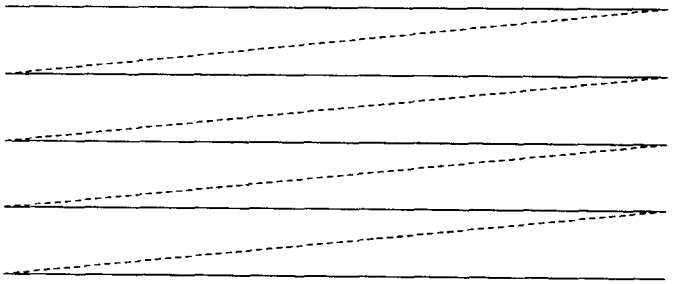


图4-3 光栅扫描：有效视频用直线表示，回扫用虚线表示

43

如图4-3，在电子束装置中电子束在回扫时是关闭的，回扫在CCD中没有物理的意义，而只是为了兼容性。图上所示是表示没有隔行扫描的最简单情况。

为了与电子束装置时代制定的扫描标准保持一致性，在电视信号的每个扫描行的末尾有一个脉冲称为消隐脉冲，这个脉冲在当时是为关闭电子束需要的时间设置的。电荷在检测器的底部被移出去，在顶部电荷又开始积累。因为电荷在所有时间内在检测器的整个表面持续积累，就要求读/传递电荷的过程必须立即回到检测器的顶部并又开始传递过程。扫描过程在一秒内重复多次。在美国制式中扫描一帧要33.3毫秒（在欧洲，每帧40毫秒）。

可用下式准确计算电子束移动的速度：

$$\frac{1\text{秒}}{30\text{帧}} \div \frac{525\text{线}}{\text{帧}} = 63.5\text{微秒/线} \tag{4-5}$$

欧洲标准是625线，每秒25帧，其速度为每行64微秒。63.5微秒中不仅包括图形信号，还包括消隐周期，大约占18%的行时间。将这部分时间减去，图形信息的时间是每行52微秒。

图4-4表示的是电视摄像机连续扫描三行的输出。显然光栅扫描过程实质上是将图像从一个二维信号转换成一维信号的过程，信号电压是时间的函数。图4-4显示的是复合视频与非复合视频信号，两者的差别取决于是否包含同步与消隐时间脉冲。

对电视信号的运作来说，同步信号是必不可少的，但对我们理解数字图像处理来说，关系并不大。而消隐信号是光栅扫描系统中最重要的定时信号。消隐是指没有视频信号的时间，有两个不同的消隐事件：水平消隐（发生在每行的末尾）以及垂直消隐（发生在图像的底部）。在数字系统中两个消隐事件分别用不同数字电平脉冲表示。通过将特定时刻的负脉冲加到视频信号中去的方法合成复合视频信号。

44

由于水平消隐表示了新一行视频信号的开端，则我们的注意力可以放在计算机是如何从这一行的电压信号编码上获取亮度信息的。

采样过程

在检测器上的电荷通过一个电阻转换成电压，并随之放大。该信号转换成数字表示是用模拟到数字的转换器实现的。模拟到数字转换器同时执行两个功能——采样和量化。

采样过程在规定的时刻度量视频信号。在每个离散时刻对视频信号进行度量，并在下一

个采样时刻到来之前保持该数值。图4-5表示了模拟电压信号与时间的函数关系，图4-6则显示了该电压经过采样之后的情况。

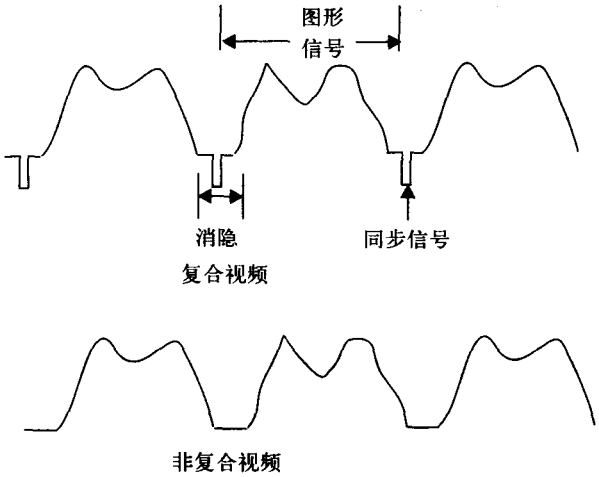


图4-4 电视摄像机复合与非复合视频输出，电压是时间的函数

图4-5 视频信号，并显示了采样时刻

图4-6 采样后的视频信号

分辨率

每一视频行上采样的数目定义为视频系统的*水平分辨率*。与其相似，一幅图像的行数定义为*垂直分辨率*。欧洲视频每幅图有625根线，垂直分辨率比美国制式要高。所以用户觉得欧洲电视的图像比美国的好。

*分辨率*这个术语也可以用来说明图形系统能清楚成像的最小单元的物理尺寸。例如乳腺X射线胶片的分辨率约为50微米，这意味着胶片上如此小的点都可以被发现。

计算机监视器有许多种分辨率标准，但计算时钟率的方法是相同的。

动态范围

采样的模拟信号用一个*量化过程*转换成数字信号，该过程显示在图4-7上。一个信号的数字电平值只有有限种可能，它用输出字的比特数来定义。视频信号经常量化成8比特的精度，因而一个信号总共只不过有256种不同的值。

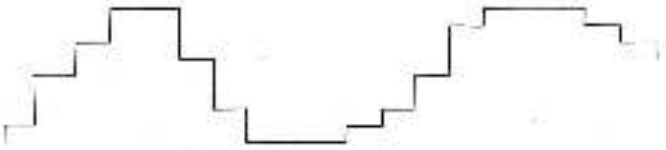


图4-7 采样及量化后的视频信号

图像系统的动态范围的一种定义是表示数字的比特数。动态范围的另一种定义是针对输入信号的，在该范围内摄像机能正常操作。两种定义都被人们经常采用，但它们的含义要根据其使用的环境确定。

由于数字图像是用光栅扫描及采样获取的，因此在时间与空间上是一一对应的。也就是说，如果我们说采样时刻，则必须将其表示成相对于图像信号顶部垂直消隐时刻的方式，这种时间关系与屏幕上的某个特定位置一一对应。

采样定义

如果对一个模拟信号进行采样与存储，那么能不能从其采样信号中精确重构出原信号呢？答案是，精确重构要求采样率至少要比信号中的最高频率成分高两倍。 46

在机器视觉中，通常并不十分关心图像中最细微细节的精确重构，而只是希望从中提取出我们所需的信息来实现手头的任务。

当一个连续值模拟信号被划分成离散值时，术语量化误差用来描述信息丢失。量化误差经常可以从图像中虚假轮廓中看出来，图4-8显示了这种情况。

图4-8 一幅人脸图像，在左图用16级灰度表示，而在右图用8级灰度(3比特)表示

4.2.2 距离图像的生成

生成距离图像有两种主要的方式——使用立体视觉与结构光。

立体视觉

许多动物有两只眼睛，并且从人们的经验也知道用两个视图，可以提取出三维信息。在几何上解释这种现象并不困难（见图4-9）。

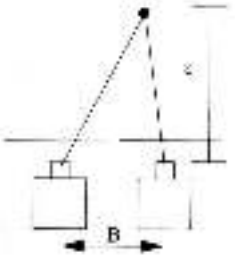


图4-9 从两个摄像机采集的图像中可以计算任何一点在三维空间的位置，条件是对应问题能够解决

如果知道摄像机之间的距离，每个摄像机的观察角度（在大多数立体视觉系统中，摄像机的中线是平行设

置的)，并且能够度量到景物中特定点在两幅图像中的位置，那么就能计算到景物中特定点的距离，称它为“距离”(range)。如果在一般性的情况下能做到这一点，也就是说，如果总能确定左图中的任何点与右图像中哪一点对应，那么对应问题就解决了。

47

立体视觉中经常使用的一个简化假设是，两个摄像机是精确平行的，并且认为如果一个点在左图上的某一行，那么它也出现在右图上相同的“外极线”(epipolar line)上。换句话说，一条外极线将一个点与它的对应点联系在一起。这种假设可以使对应问题计算的复杂度明显降低。

参考文献中有许多关于对应问题的文章，大部分集中在点匹配上，也就是说，在第二幅图的外极线上找到一点，与第一幅图像中的一点相似，相似度以某种方式度量。例如Bokil 与Khotanzad[4.5]，扩展Marr与 Poggio [4.27]的工作，也采用外极线假设。他们通过建立灰度相容矩阵（GLCM）的方法完成点匹配。左图与右图的像素值标在矩阵的左边与底部[⊖]。列在第i行的左图像素值与列在第j列的右图像素之间差异的绝对值写在矩阵的第i，j元素内。

分层匹配方法可以使对应问题容易些[4.26]（两个低层特征，例如外极边，要形成对应关系，只有当它们所属的区域是对应的才行)。

有一些方法在三维空间找曲线，这些方法并不显式地要求先解决对应问题。例如Cohen与Wang[4.10，4.11]求最佳匹配曲线，而不是逐个点。

摄像机标定[4.28，4.37]对立体视觉[4.31]是重要的，而立体视觉技术又能用来进行摄像机标定，因为它在二维图像和三维世界之间建立起联系。人们在确定标定摄像机所需的对应最小集[4.1，4.23]或其他关系[4.33，4 .34]方面做了许多工作。

从一组（来自同一物体的）对应特征就可计算一个变换，从而确定景物中一个物体的位姿（pose)，这是指物体在三维空间中的位置与朝向。在同一场景中可以有不止一组对应特征[4.20，4.38]。

立体视觉有许多不同种类的应用[4.13]，包括如何从全景摄像机中获取体视信息。

第11章在学习了参数变换的概念如何有助于解决对应问题之后，会进一步讨论立体视觉技术。

结构光照明

48

将立体视觉中的一个摄像机用一个光源（例如用激光穿过柱面透镜）代替，可以避免对应问题。这样一来这种方法就不再是立体视觉了，而一般称为*结构光照明*（structured illumination)。要理解它是如何工作的，可以回顾一下图4-9，并把其中一台摄像机用一个投影器取代。该投影器向景物中投射一条很窄但很亮的光，如图4-10所示。其中从投影器投射的角度θ已知，而另一角度ϕ是在图像中检测到亮点之后，根据其像素位置，利用像素位置与角度之间关系计算出来的。再加上投影器到摄像机之间的距离d这个已知条件，就可解出图4-10中的三角关系，而景物中亮点到摄像机的距离也就可以算出来。

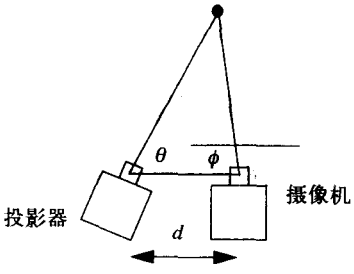


图4-10 结构光照明

图像的描述问题可能会对这种方法有影响。当用结构光去照射有镜面反射现象的场景时

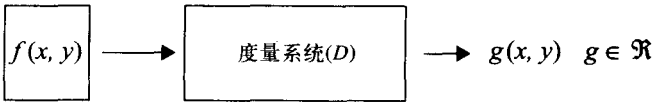
⊖ 原书是底部与左边。——译者注

会有些问题，比如照射金属表面就会出现这种现象。此时可能会没有足够强的反射出现或太多强反射出现，偏振光滤波器可能会有帮助[4.29]。

在本书讨论有关从X获取形状信息的章节（9A.2节）中会进一步讨论使用结构光照明的技术。

4.3 图像生成机制

首先将亮度定成两个空间变量的函数，$f(x,y)$是点(x,y)处的亮度；$x,y\in\Re$，f为实数$f\in\Re$。



用一个度量系统度量图像会使输入图像质量失真（退化），得到一个度量后的图像

$$g(x,y)=D(f(x,y)) \tag{4-6}$$

49

其中D是某种退化函数，一般会包含某些随机噪声过程，在一种我们比较感兴趣的退化现象中噪声是添加性的，并且独立于信号，此时g可以写成：

$$g(x,y)=\iint_{\infty}^{\infty} f(\alpha,\beta)h(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta+n(x,y) \tag{4-7}$$

这是一个卷积积分。如果任何一个退化操作D是线性以及移不变的（在空间任何地方操作都一样），那么可以证明它就能以式（4-7）这样的卷积进行计算。以下面一个一维的情况为例（只是为了方便）来证明这一点，推广到二维情况是很自然的：首先观察在点x计算的函数f，可以写成：

$$f(x)=\int_{-\infty}^{\infty} f(x')\delta(x-x')\mathrm{d}x' \tag{4-8}$$

其中$\delta(r)$表示delta（增量函数）函数，它的定义是：当其变量r为非零时，该函数值为零；而当变量r为零时，其值为无穷大，并且对该函数在$-\infty$到$+\infty$积分，其积分值为1。式（4-8）看起来似乎并没有什么奥妙之处，它只是定义了一个函数的delta函数采样[⊖]。进一步假设函数f被某种操作引起退化，从而改变了每一点x的函数值。则有

$$D(f(x))=D\left(\int_{-\infty}^{\infty} f(x')\delta(x-x')\mathrm{d}x'\right) \tag{4-9}$$

如果D是线性运算，那么可以变换式（4-9）中操作符D与积分符号的次序，得到：

$$D(f(x))=\int_{-\infty}^{\infty} D(f(x')\delta(x-x'))\mathrm{d}x' \tag{4-10}$$

D是作用在x的函数上的，而不是x'的函数。对D而言任何与x'有关的都是常数，所以可以将$f(x')$提到该操作运算之外，得到：

⊖ 但其作用会逐渐显现出来。——译者注

$$D(f(x)) = \int_{-\infty}^{\infty} f(x')D(\delta(x-x'))\mathrm{d}x' \tag{4-11}$$

此式表明D可能取决于x，或取决于x与x'的差值，总而言之它是只对脉冲函数应用的退化运算。因此f的任何线性退化可以写成f与一个函数乘积的积分，该函数是对delta函数的退化。由于delta函数的幅度为无穷大、宽度为零，在一维时称其为脉冲，而$D(\delta(x-x'))$称为脉冲响应（impulse response），而二维delta函数是一个亮点，所以称应用它带来的退化为点扩散函数（point spread function）。脉冲响应与点扩散函数实际上是一回事，只是用处不同。

50

由于脉冲响应可能取决于x与x'两者，为了方便引入一个新的表示：

$$h(x,x') = D(\delta(x-x')) \tag{4-12}$$

如果再加上另一种假设，则可得到一个更简单的表示式：设D与x无关，而只取决于x与x'的差，在此种情况可以写成$h(x,x') = h(x-x')$，而式（4-11）可简化成：

$$g(x) = D(f(x)) = \int_{-\infty}^{\infty} f(x')h(x-x')\mathrm{d}x' \tag{4-13}$$

此式中引入了g，它是系统的输出。这就是卷积积分。有许多理由说明这个积分是很重要的，其中包括它可以用快速傅里叶变换（FFT）进行计算。更重要的一点是，图像的任何退化（只要它是线性与移不变的），都可用这一类的积分计算。

4.4 将图像看作表面

在这一节讨论将图像看作三维空间的一个表面的问题。用这种角度看待图像可以把图像的属性看成（表面点的）高度。

4.4.1 等亮度线

将$f(x, y)$的值看作空间的一个表面，表示成$z= f(x, y)$，那么可用有序的三元素$[x, y, f(x, y)]^T$描述表面。对每一点(x, y)在第3维有一对应值。很重要一点在于对任何(x, y)点只有一个z值（也就是$f(x, y)$是一个函数）。因此z是一个表面。

现在考虑满足$f(x, y)=c$（c为某常数）的所有点组成的点集。如果f表示亮度，那么这个点集中的点具有相同的亮度，因此称其为“等亮度线”（isophote）。

定理

51

在图像的任何点(x, y)，通过该点的等亮度线与该点的梯度相垂直。[⊖]

4.4.2 脊

下面将空间一表面$z(x, y)$看成一座山（见图4-11）。如果画出这座山的地质轮廓图，图上的线是等海拔高度线。但是如果把亮度看作海拔高度，那么这些轮廓线就是等亮度线。站在山上的某点，沿其梯度方向观望，那么你所看到的方向就是上升最陡峭的方向。

⊖ 梯度向量在式2.11中定义，并将在式5.22中详细描述。

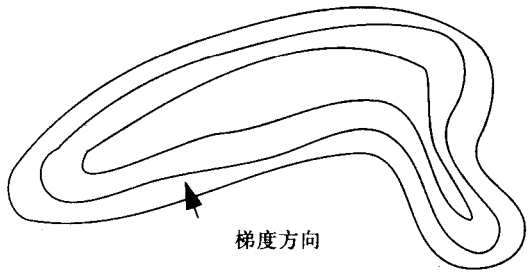


图4-11 海拔图上的轮廓与等亮度线等价，该点的梯度向量与该点的等亮度线垂直

此时若你朝你左边或右边看，这就是沿等亮度线方向看。要提醒的是梯度方向是该点最陡峭的方向，但不一定指向顶峰。

假设我们在爬山时，每一小步都是沿梯度方向走的，那么在脊线上会发生什么？怎样知道你处于脊线？能否用数学描述这个过程？

假设我们沿梯度方向迈步，在到达脊线之前每一步有大致相同的方向，然而到达脊线之后，方向会发生显著变化。所以对脊线的一个有用的定义是：脊线是由这样一组点组成，在这些点梯度方向的变化率是局部最大值。也就是说，需要找$\partial\theta/\partial v$ 取最大值的点，其中v表示梯度方向，而它是沿梯度方向的求导。在笛卡尔坐标中，

$$\frac{\partial\theta}{\partial v}=\frac{2f_xf_yf_{xy}-f_y^2f_{xx}-f_x^2f_{yy}}{(f_x^2+f_y^2)^{3/2}} \tag{4-14}$$

Maintz等人[4.24]曾指出，该式基本上与一个略为简单的式子等价，该式子只是基于亮度沿v方向的二阶导数，这就导出要将下式最大化

$$\frac{f_y^2f_{xx}-2f_xf_yf_{xy}+f_x^2f_{yy}}{(f_x^2+f_y^2)} \tag{4-15}$$

52

式中下标表示对该变量的二阶偏导数。在三维数据中，脊的概念是相同的，只是难以可视化。此时梯度是一个三维向量，指向增加密度的方向。等亮度点集就不再是曲线，而是曲面。Maintz等人在[4.24]中同样讨论了三维数据中脊的概念。需要的读者可参考该文章。

4.4.3 二值图像与中轴

中轴概念用来描述图像内的一个区域，在二值图像中梯度向量已经没用了，因为它几乎到处为零，因此用中轴是很有效的。在区域中有这样性质的一些点，在这些点上能够画出一个与区域的边界至少有两个切点的圆，所有这些圆的中心组成的点集就是中轴。

中轴很容易定义，但很难计算。这将在第9章引出数学形态学运算与距离变换后再进行讨论。

通过使用图像的尺度空间表示可以将脊与中轴联系起来，这将在9.7.1节讨论。

4.5 邻域关系

邻域的定义有多种方式，但最常用与最直观的方法是4邻域或8邻域，4邻域是指共享一条边的两个点[⊖]有相邻关系，8邻域则是指共享一条边或一个顶点的两个点有相邻关系。一个像

⊖ 将每个点看作一个正方形。——译者注

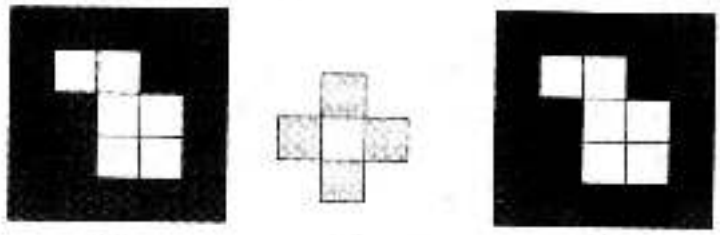
素的邻域是指与其有相邻关系的所有点组成的集合。一个中心像素的4邻域如图4-12所示。一个点v的邻域用N_v表示。下面将讨论对点集、点的邻域以及邻域集的运算。例如令A与B是图像中的点集，而v是A中一个点，则可以定义A集与B集通过邻域关系N_v进行某种运算，得到一个新的点集，例如

$$O_B(A)=\bigcup_{v\in A}(N_v\cap B)$$

图4-12 中心像素的4邻域用阴影标出

文献[4.15]把上式计算得到的结果称为A集相对B集的流域（aura），它实质上是A集中部分点的集合，这些点与B中的点有相邻关系，至于“相邻”的概念由所关注的问题定义。图4-13a显示了一幅图像包含一个集A（阴影像素的集合）、一个B集（空白像素的集合）；图4-13b表示相邻关系的定义；图4-13c则显示了B集在A集的流域[4.15]。当讨论到数学形态学运算时会涉及更多类似的关系运算。

53

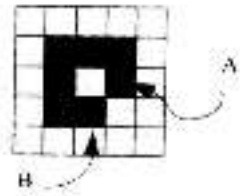


a) b) c)

图4-13 a) A集（阴影像素）与B集（空白像素）。b) 邻域关系，阴影像素是中心像素的邻域。c) a)图中空白像素集在阴影像素集中的流域，用黑色标注。在该例中用4邻域作为邻域关系的定义，但是并不一定要求相邻者必须在空间毗邻

题外话：连通性难题

在左图中前景像素是黑色的，背景像素是白色的。图像中的前景是一个环，多半是一个垫圈的低分辨率图像。这个环是封闭的吗？也就是说能否从一点开始，用已学过的邻域定义，绕着环走，不走回头步，从一个像素过渡到一个连通的邻域呢？（在这里用路径来定义连通性。）如果能够绕这个区域走下去，就说该区域是封闭的。我们已学过4邻域与8邻域，若使用4邻域关系，这个区域是封闭的吗？如果你说不，就对了！在4邻域系统中，无法从像素A过渡到像素B。如果它是不封闭的它就应该是开放的，对不对？如果它是开放的，那么其内部与外部应该是连通的，对不对？但是使用4邻域关系的定义，其内部与外部却是分开的区域。



看来我们只好放弃4邻域定义了，因为它导致一个互相矛盾的结论。那么试试8邻域关系，此时前景是封闭的对吗？（对！）但是内部与外部区域却也是连通的！这怎么可能呢？如果该区域是封闭的，它的内部与外部（从逻辑上讲）必定是分开的，所以8邻域也行不通。

对这个特定的难题，有一个解决办法，至少是部分的解决办法。这就是用8邻域或4邻域中的一个作为前景连通性的定义，而另一个给背景用。

这种修补方法对二值图像是适用的，但是对有不止两个亮度级的图像来说，这种不一致性又会出现。举这个例子的目的，是要说明在数字图像中，直觉并不总是对的。例如有许多相似问题的其他例子，是与区域周长的度量方法有关的。要警惕会有离奇的事发生，并记住

54

直觉并不总是正确的。

在4A.1节中讨论的六角形表示解决了连通性难题，它可能还能解决其他问题。

作业4.1 处理与观察距离图像

这个习题的目的是，使你习惯于使用计算机以及这门课需要的软件包。

1) 打开CDROM 的“ Image”目录（Unix用户打开系统盘）在“leadhole”子目录你会看到一些图像，名字为“site...”用适当的工具来观看这些图像，并将其中你喜欢的拷贝到你的目录中，这些图像都不大，不会占太多空间的。

2) 对所拷贝的每一幅图像，运行ifs viewpoint（为了得到这变量，键入“ifs viewpoint -h”）。这个程序接受距离图像为输入，并允许用户能从直对的视角以外的视角来观察这些图像。当从别的角度看时，你可看到距离图像是三维的。为了选择视点的运动，对“enter transform”做一个类似于下列命令行的回答:

movex – 10 movey 10 roll 30 pitch 0 yaw 5 movex 3 movez 10 end

3) 如果你愿意，你可以输入这些到相继行上。它在得到“end”之前会持续地读运动命令。roll（侧滚）、pitch（倾斜）和yaw（侧转）是旋转的命令，它们的变量单位是度，继续玩到你能回答以下问题时为止。

a) movex是否移动视点或图像，在什么方向（行或列）？

b) roll 是绕z轴转，pitch是绕什么轴旋转？

4) 取一幅图像，产生6幅新图像，分别对应不同的roll角（0、30、60、90、120与150度）。有没有将其中某个转出图像外？使用一个变换来弥补这个问题（提示: 可以考虑用一移动命令把原图移至图像的中心再旋转它）。使用程序ifs stack把这六幅图像转换成一个三维图像。如果你使用Unix则用imp来观察图像，并且演示如何用imp像演电影一样来旋转物体。（提示: 使用“volume”按钮）。如果你使用PC，你可以使用ifs2avi。AVI图像可以用许多种PC程序观看。对 .avi图像的图符连续快速按两次就可实现。 55

5) 下面学习如何使用程序 ifs spin。[⊖]演示你所生成的复杂的电影。注意: ifs spin实际上是运行viewpoint，Unix版会产生相当多的暂存文件，完成后它会自动清除，只是需要有暂存的盘空间。

写下你的结果，并向你的导师演示你的结果。

提示: 在你的CDROM的“leadhole”目录中你会找到一幅名叫 spinout.avi的图像，你所生成的图像应该大体上与它的输出相像。

4.6 结论

在本章，讨论了图像以及图像中信息的各种表示法，以便在后续章节我们会使用这些表示方法编制算法来提取信息，并对信息进行分类。

4.7 术语

对应问题　　correspondence problem

⊖ 要学会如何使用IFS程序，可以输入program_name-h或阅读手册。

曲率	curvature
动态范围	dynamic range
函数表式	functional representation
图	graph
图像表示	iconic representation
等亮度线	isophote
线性系统	linear system
中轴	medial axis
概率表示	probabilistic representation
量化	quantization
距离图像	range image
光栅扫描	raster scan
脊	ridge
分辨率	resolution
采样	sampling
空域频率	spatial frequency
立体视觉	stereo
结构光照明	structured illumination

专题4A 图像表示方法

4A.1 采样的变种：六角形像素

在一些文章中[4.36]，提到使用六边形结构的像素数组的成像传感器（见图4-14），当一堆互相相切的圆在互相挤压条件下，发生变形时六边形是其最小能量解。蜂窝是自然界这类结构中最著名的，也有一些别的例子，包括人视网膜中的视锥细胞。

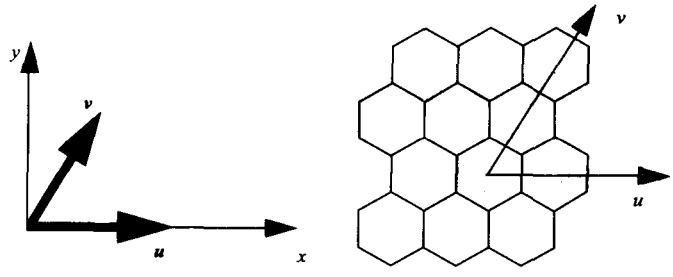


图4-14 一种适用于平面上六面体镶嵌的坐标系统。其中u与v的方向是不正交的，u与v的单位向量用于描述这个坐标系统

56 ~ 57

传统的电子成像传感器被安排成矩形数组状，这主要是因为电子束需要以光栅扫描方式扫描，而后出现的电荷耦合器件安排成矩形结构也较方便。然而矩形数组在定义邻域关系时有歧义性。而在六边形连通分析中却没有连通疑题：每个像素正好有六个近邻，无论前景，背景或别的彩色。

符号表示

在R^2中的一点用$\boldsymbol{p}=u\boldsymbol{u}+v\boldsymbol{v}$表示，其中粗体表示单位向量，非粗体的字母表示相应幅值，在讨论两个或多个点的情况时，对不同的向量使用下标，其相应分量也用相同下标，例如：$\boldsymbol{p}_i=u_i\boldsymbol{u}+v_i\boldsymbol{v}$。此外，点还可用列向量形式表示：

$$P_i=[u_i,v_i]^{\mathrm{T}} \tag{4-16}$$

在某种情况下，我们关心点在笛卡尔表示中的位置，即$[x, y]^{\mathrm{T}}$，在此种情况下也会采用下标，例如$\boldsymbol{P}_i=[u_i, v_i]^{\mathrm{T}}=[x_i, y_i]^{\mathrm{T}}$，表示对应的单位向量$u, v, x, y$的值。

引理1

任何一个有序对$[u, v]$只对应一对$[x, y]$。

证明：

使用简单的三角关系，并且注意cos 60°的值是1/2，则有：

$$x=u+\frac{v}{2} \quad 和\ y=\frac{\sqrt{3}v}{2} \tag{4-17}$$

引理2

任何笛卡尔坐标有序对$[x, y]$与一对$[u, v]$唯一对应。

证明：

解式（4-17）得

$$u=x-\frac{y}{\sqrt{3}} \quad 和 \quad v=2\frac{y}{\sqrt{3}} \tag{4-18}$$

58

一组向量$\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_d$称为$\Re^d$向量空间的基，条件是$\Re^d$中的任何向量可写成$\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_d$的线性组合，如果$\boldsymbol{b}_i$是正交归一的（$\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{b}_j=0$当$i\neq j$，以及$\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{b}_j=1$）则它们构成一组基的条件是充分的。但是正交归一并不是必要条件，$\boldsymbol{u}$与$\boldsymbol{v}$组成一组基，它们是归一的，但不正交。这种非正交性在式（4-19）中可以看出来，因为$\boldsymbol{u}$，$\boldsymbol{v}$在笛卡尔坐标中的内积并不为零。

$$\boldsymbol{u}=\boldsymbol{x},\boldsymbol{v}=\frac{1}{2}\boldsymbol{x}+\frac{\sqrt{3}}{2}\boldsymbol{y},\ 所以\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}=[1,0]\begin{bmatrix}\frac{1}{2}\\ \frac{\sqrt{3}}{2}\end{bmatrix}=\frac{1}{2} \tag{4-19}$$

定理

向量$\boldsymbol{u}$与$\boldsymbol{v}$形成$\Re^d$中的基

证明：

由于$\boldsymbol{x}$与$\boldsymbol{y}$显然是$\Re^d$的一组基，则可以将$\Re^d$的任一点$\boldsymbol{p}$写成

$p=[x, y]^{\mathrm{T}}=x\boldsymbol{x}+y\boldsymbol{y}$，然而从式（4-19）有

$$\boldsymbol{p}=x\boldsymbol{u}+y\frac{(2\boldsymbol{v}-\boldsymbol{u})}{\sqrt{3}}=\left(x-\frac{y}{\sqrt{3}}\right)\boldsymbol{u}+\frac{2y}{\sqrt{3}}\boldsymbol{v}$$

于是R^2中任何一点可以写成$\boldsymbol{u}$与$\boldsymbol{v}$的加权和。证毕。

鉴别一个像素的邻域

给定一个像素的$\boldsymbol{u}$、$\boldsymbol{v}$坐标（假设为整数）其近邻的坐标在图4-15中表示。可以使用下列循环来有效地存取像素（u, v）的所有6个近邻。

注意：无论中心像素出现在奇数或偶数行中，循环中并不要求使用“if”语句。

```
ou={-1,-1,0,1,1,0}
ov={0,1,1,0,-1,-1}
for i=1 to 6
  nu=u+ou[i]
  nv=v+ov[i]
  value=image[nu][nv]
```

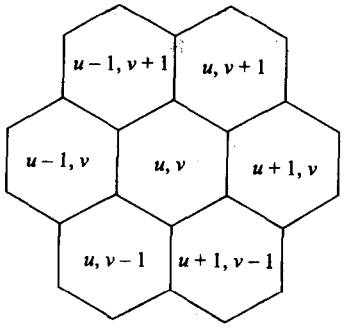


图4-15 一个像素的邻域是绝对对称的，没有4邻域或8邻域的歧义性

我们发现这种访问一个像素的近邻的方法也适用于矩形网格的8邻域定义，而且比双索引循环的方法更有效。

```
for i=-1 to 1
for j=-1 to 1
if((i!=0)or(j!=0))
value=image[u+i][v+j]
```

注意：ou与ov是相互成循环的。

4A.2 其他形式的图像表示

到目前为止,我们只是用亮度来表示图像的属性。当然我们曾提到 $2\frac{1}{2}$ 维图像，它表示距离是x与y的函数。然而，还可以计算一些其他量，作为图像的属性。曲率就是其中之一。

4A.2.1 曲率

在图像中的每一点可以计算局部曲率，对 $2\frac{1}{2}$ 维图像（表面）曲率无法用单个标量确切描述，而以采用矩阵形式为好。（想要了解这方面细节的读者可以参考doCarmo的书[4.12]或其他微分几何的书籍。）

$$K=\begin{bmatrix}E & F\\ F & G\end{bmatrix}^{-1}\begin{bmatrix}e & f\\ f & g\end{bmatrix} \tag{4-20}$$

其中

$$E=1+\left(\frac{\partial z}{\partial x}\right)^2, F=\frac{\partial z\partial z}{\partial x\partial y}, G=1+\left(\frac{\partial z}{\partial y}\right)^2$$

$$e=\left(\frac{\partial^2 z}{\partial x^2}\right)^2\Big/H, f=\left(\frac{\partial^2 z}{\partial x\partial y}\right)^2\Big/H, g=\left(\frac{\partial^2 z}{\partial y^2}\right)^2\Big/H$$

而最后得到:

$$H=\sqrt{\left(\frac{\partial z}{\partial x}\right)^2+\left(\frac{\partial z}{\partial y}\right)^2+1}$$

主曲率K_1与K_2定义为矩阵K的两个特征值，而相应的特征向量确定曲率的方向。

在许多视觉应用中，需要对曲率有一个标量描述，并且与视点无关，如此定义的标量表示有两种。平均曲率

59 ~ 60

$$K_M = \frac{1}{2}(K_1 + K_2) = \frac{1}{2}\mathrm{Tr}(K) \tag{4.21}$$

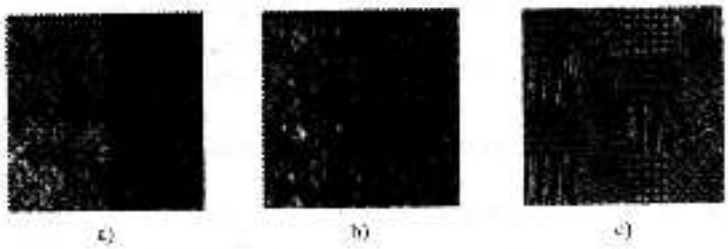
以及高斯曲率

$$K_G = K_1 K_2 = \det(K) \tag{4.22}$$

由于高斯曲率是乘积，因此只要两个主曲率中任一个为零，它的值就为零。由于这种情况在工业零件中经常出现，因此我们很少使用高斯曲率。

4A.2.2 纹理

纹理这个词似乎每个人都知道，但是它却没有确切的定义。对纹理至少有两种定义："自然"纹理，它用随机过程描述最确当；"规则"纹理，它用频域表达式最合适（见图4-16）。



a) b) c)

图4-16 a）羊毛纹理的例子[4.2] b）树皮纹理的例子[4.2] c）自然的与规则纹理的比较[4.4]

Haralick和 Shapiro[4.18]对纹理的描述是"其包含一种或多种以下属性：细度、粗糙、平滑度、颗粒状、随机性、线状条纹或有斑纹、不规则或丘状起伏"。由于像素的组合是不同的，所以我们可以采用很多特征，其中包括功率谱的矩[4.4, 4.14]、分形维[8.12]、复倒谱[4.35]。所谓的纹理分析，通常先用包含空域以及空域频域的信息来描述图像，然而用这些信息检测具有相同特征的区域[4.7, 4.14, 4.16]。

一些文章提到纹理可以用自相似（分形）过程有效表示，其中最初的文章出现在Mandelbrot 与Van Ness[4.25]的经典著作。而Kaplan与 Kuo[4.22]则指出实际纹理并不一定具备在不同尺度下保持相同纹理的性质，因而自相似的概念应作一些修改。

作业4.A1

假设图像$f(x, y)$用$f(x, y) = x^2/4 - x^2 + y^2$描述，对点$x=1$，$y=2$来说，以下式子中哪一个是穿过该点指向沿等亮度线的方向的单位向量？

(a) $\begin{bmatrix} \frac{2}{(\sqrt{5})} & \frac{1}{\sqrt{5}} \end{bmatrix}^T$ (c) $\begin{bmatrix} \frac{-1}{(\sqrt{5})} & \frac{2}{\sqrt{5}} \end{bmatrix}^T$ (e) $[2\ \ 1]^T$

(b) $\begin{bmatrix} \frac{1}{(\sqrt{5})} & \frac{2}{\sqrt{5}} \end{bmatrix}^T$ (d) $[-2\ \ 4]^T$ (f) $\begin{bmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}^T$

61

作业4.A2

想像你站在一个表面上，你无法看到整个表面，但能看到它的大部分。如果你度量所有你能见到的点的曲率，你发现主曲率之一为零，而另一个主曲率沿一个方向单调变化，

你无法精确地度量，但你推测沿这个方向的曲率变化是线性的，试问你站在什么类型的表面上？

参考文献

[4.1] T. Alter, "3-D Pose from 3 Points Using Weak-perspective," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), 1994.

[4.2] D. Badler, J. JáJá, and R. Chellappa, "Scalable Data Parallel Algorithms for Texture Synthesis and Compression using Gibbs Random Fields," *IEEE Transactions on Image Processing*, **4**(10), 1995.

[4.3] R. Bajcsy and F. Solina, "Three Dimensional Object Representation Revisited," *International Conference on Computer Vision*, London, May, 1987.

[4.4] J. Bigün and J. du Buf, "N-folded Symmetries by Complex Moments in Gabor Space and Their Application to Unsupervised Texture Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[4.5] A. Bokil and A. Khotanzad, "A Constraint Learning Feedback Dynamic Model for Stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[4.6] K. Castleman, *Digital Image Processing*, Englewood Cliffs, NJ, Prentice-Hall, 1996.

[4.7] J. Chen and A. Kundu, "Rotation and Gray Scale Transformation Invariant Texture Identification using Wavelet Decomposition and Hidden Markov Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(2), 1994.

[4.8] R. Chien and W. Snyder, "Hardware for Visual Image Processing," *IEEE Transactions on Circuits and Systems*, **22**(6), 1975.

[4.9] D. Clausi, "Texture Segmentation Example," Web publication, http://www.eng.uwaterloo.ca/~dclausi/texture.html, Spring 2001.

[4.10] F. Cohen and J. Wang, "Part I: Modeling Image Curves Using Invariant 3-D Object Curve Models – A Path to 3-D Reconstruction and Shape Estimation from Image Contours Using B-Splines, Shape Invariant Matching and Neural Network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[4.11] F. Cohen and J. Wang, "Part II: 3-D Object Recognition and Shape Estimation from Image Contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[4.12] M. doCarmo, *Differential Geometry of Curves and Surfaces*, Englewood Cliffs, NJ, Prentice-Hall, 1976.

[4.13] U. Dhond, and J. Aggarwal, "Stereo Matching in the Presence of Narrow Occluding Objects using Dynamic Disparity Search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(7), 1995.

[4.14] D. Dunn, W. Higgins, and J. Wakeley, "Texture Segmentation using 2-D Gabor Elementary Functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(2), 1994.

[4.15] I. Elfadel and R. Picard, "Gibbs Random Fields, Co-occurrences, and Texture Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[4.16] H. Greenspan, R. Goodman, R. Chellappa, and C. Anderson, "Learning Texture Discrimination Rules in a Multiresolution System," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

[4.17] M. Gürelli and L. Onural, "On a Parameter Estimation Method for Gibbs–Markov

Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[4.18] R. Haralick and L. Shapiro, *Computer and Robot Vision*, Volume I, Reading, MA, Addison-Wesley, 1992.

[4.19] G. Healey and R. Kondepudy, "Radiometric CCD Camera Calibration and Noise Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), 1994.

[4.20] Y. Hel-Or and M. Werman, "Pose Estimation by Fusing Noisy Data of Different Dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 1995.

[4.21] A. Jain and K. Karu, "Learning Texture Discrimination Masks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(2), 1996.

[4.22] L. Kaplan and C. Kuo, "Texture Roughness Analysis and Synthesis via Extended Self-similar (ESS) Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[4.23] D. Keren, D. Cooper, and J. Subrahmonia, "Describing Complicated Objects by Implicit Polynomials," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[4.24] J. Maintz, P. van den Elsen, and M. Viergever, "Evaluation of Ridge Seeking Operations for Multimodality Medical Image Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[4.25] B. Mandelbrot and J. Van Ness, "Fractional Brownian Motions, Fractional Noises, and Applications," *SIAM Review*, **10**, October, 1968.

[4.26] S. Marapan and M. Trivedi, "Multi-primitive Hierarchical (MPH) Stereo Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), 1994.

[4.27] D. Marr and T. Poggio, "Cooperative Computation of Stereo Disparity," *Science*, **194**, pp. 283–287, October, 1976.

[4.28] P. McLauchlan and D. Murray, "Active Camera Calibration for a Head-eye Platform Using Variable State-dimension Filter," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(1), 1996.

[4.29] N. Page, W. Snyder, and S. Rajala, "Turbine Blade Image Processing System," In *Advanced Software for Robotics*, ed. A Danthine, Amsterdam, North-Holland, 1984.

[4.30] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic Stereo Imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3), 2001.

[4.31] L. Quan, "Invariants of Six Points and Projective Reconstruction from Three Uncalibrated Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995

[4.32] A. Shashua, "Projective Structure from Uncalibrated Images: Structure from Motion and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), 1994.

[4.33] A. Shashua, "Algebraic Functions for Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[4.34] A. Shashua and N. Navab, "Relative Affine Structure: Canonical Model for 3D From 2D Geometry and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(9), 1996.

[4.35] P. Smith and N. Nandhakumar, "An Improved Power Cepstrum Based Stereo Corre-

spondence Method for Textured Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), 1996.

[4.36] W. Snyder, H. Qi, and W. Sander, "A Hexagonal Coordinate System," *SPIE Medical Imaging: Image Processing*, Pt. 1–2, pp. 716–727, February, 1999.

[4.37] G. Wei and S. Ma, "Implicit and Explicit Camera Calibration: Theory and Experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

62 ~ 64

[4.38] X. Zhuang and Y. Huang, "Robust 3-D – 3-D Pose Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), 1994.

第5章　线性算子与核算子

Now I see through a glass darkly, but then, face to face.

Paul of Tarsus

在这一章，我们将研究图像中的线性算子，首先是微分算子，这可能是线性算子中最常见的一类算子。然后将讨论扩大到边缘检测，同时考虑各种实现边缘检测的方法。

5.1　什么是线性算子

假定D是一个作用在图像f上的算子，并生成图像g。如果D满足

$$D(\alpha f_1 + \beta f_2) = \alpha D(f_1) + \beta D(f_2) \tag{5-1}$$

这里f_1、f_2表示图像，乘数α、β是标量，那么我们说D是一个“线性算子”。

思考实验

考虑图像算子D

$$g = D(f) = af + b \quad a,b \in \Re$$

那么D是一个线性算子么？

我们建议你在看解答之前先自己计算一下。乘以一个常数之后又加了一个常数，看起来似乎应该是线性的。如果变量f是一个标量，那么D就表示一个直线方程，它自然就应该是线性的（不是吗？）！好的，让我们来证明一下。根据等式（5-1），我们可得：

$$D(\alpha f_1 + \beta f_2) = a(\alpha f_1 + \beta f_2) + b = a\alpha f_1 + a\beta f_2 + b$$

检查一下是否与下面等式一样：

$$\begin{aligned}\alpha D(f_1) + \beta D(f_2) &= \alpha(af_1 + b) + \beta(af_2 + b)\\ &= a\alpha f_1 + \alpha b + a\beta f_2 + \beta b\\ &= a\alpha f_1 + a\beta f_2 + b(\alpha + \beta)\end{aligned}$$

所以除非有$\alpha+\beta=1$，D并不是一个线性算子！这看起来似乎与直觉是矛盾的。对于这个问题我们在以后还会再探讨，并且看看能否弄明白其原因。在本章后面部分，我们将讨论图像中的线性算子。

5.2　核算子在数字图像中的应用

图像f是数字的，所以许多作者习惯将f写成矩阵f_{ij}的形式，而不是像$f(x, y)$这样的函数表达式，但我们则更喜欢采用x、y的形式，原因在以后自然会清楚。后面，我们将发现采用简单的下标f_i来表示则更为方便，但现在仍采用$f(x, y)$的形式。这里，x和y仅是取值较小的整数，例如：$0<x<511$。

考虑一维图像f和h，f有5个像素，而h有3个像素，在这儿我们称h为核算子，如图5-1所示。

将核算子的中心h_0移到图像f的一个像素f_2上，就可以得到核算子与图像对应元素的乘积之和，即：$g_2=f_1h_{-1}+f_2h_0+f_3h_1$。了解了这些后，考虑式（5-2）这样一个最一般的例子，这是3×3核算子的应用实例：

$$g(x,y)=\sum_{\alpha}\sum_{\beta}f(x+\alpha,y+\beta)h(\alpha,\beta) \tag{5-2}$$

这里α和β仅取－1、0和1，那么式（5-2）可展开为：

$$\begin{aligned}g(x,y)=&f(x-1,y-1)h(-1,-1)+f(x,y-1)h(0,-1)\\&+f(x+1,y-1)h(1,-1)+f(x-1,y)h(-1,0)\\&+f(x,y)h(0,0)+f(x+1,y)h(1,0)+f(x-1,y+1)h(-1,1)\\&+f(x,y+1)h(0,1)+f(x+1,y+1)h(1,1)\end{aligned} \tag{5-3}$$

66

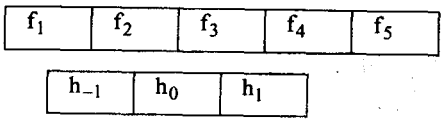


图5-1 5个像素大小的一维图像和3个像素大小的一维核算子。下标为像素的x坐标

表5-1 核算子元素的值

$h(-1,-1)$	$h(0,-1)$	$h(1,-1)$
$h(-1,0)$	$h(0,0)$	$h(1,0)$
$h(-1,1)$	$h(0,1)$	$h(1,1)$

为了更好地理解式（5-3）的本质，我们将h写成3×3的数字网格（grid）（是的，我们故意用了“网格”这个词，而不是“矩阵”），如表5-1所示。假设将这个网格放到图像上面使其中心正对像素点$f(x,y)$，然后网格中每个h值将与图像中相应像素点相乘。今后我们将h值的网格称为“核算子”。

5.2.1 自变量的方向：卷积和相关

让我们重新讨论两个重要的方程。第一个是核算子，也就是式（5-2），另一个则是二维离散卷积：

$$g(x,y)=\sum_{\alpha}\sum_{\beta}f(x+\alpha,y+\beta)h(\alpha,\beta) \tag{5-4}$$

$$g(x,y)=\sum_{\alpha}\sum_{\beta}f(x-\alpha,y-\beta)h(\alpha,\beta) \tag{5-5}$$

细心的人会注意到式（5-4）与式（5-5）中乘积的顺序是不同的。在卷积的正式定义中，如式（5-5）所示，自变量是反相的，即核算子最右边的元素h_1与相应图像最左边的像素f_2相乘，而在式（5-4）中，则是将核算子移到图像上使相应元素相乘。如果相应元素直接相乘，即左边与左边相乘，右边与右边相乘，就可以得到相关运算。不幸的是，许多文献却用词不当，误将两者都称为“卷积”，这里提醒大家注意这一点。在许多著作中，作者所采用的术语“卷积”其实是指“乘积和”。在本书中，为了避免混淆，除非指真的像式（5-5）的例子，我们尽量避免使用“卷积”这个词，而对于式（5-4），我们称之为“核算子”。

5.2.2 用核算子估计导数

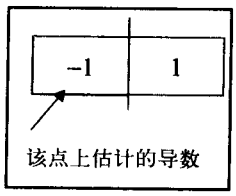
67

让我们通过一个例子，求图像空间导数的近似值$\partial f/\partial x$和$\partial f/\partial y$来说明这个概念。

我们可以从微积分课依稀回忆起下面这个公式:

$$\frac{\partial f}{\partial x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

从这个公式可以得到下面这个核算子（$\Delta x=1$）:



但是这个核算子看起来令人感到不舒服，x处的估计值为什么仅仅依赖于x和$x+1$处的值，却与$x-1$没有关系？我们采用另一个对称的定义，如下:

$$\left.\frac{\partial f}{\partial x}\right|_{x_a}=\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0-\Delta x)}{2\Delta x}$$

因为Δx不能比1小，因此我们将这个核算子定为:

-1/2	0	1/2

为了表示简单，我们写作:

1/2	-1	0	1

对噪声的敏感性是导数的一个主要问题，我们通过求水平差分后，做一次垂直平均来进行补偿，这产生了下面的核算子:

$$\left.\frac{\partial f}{\partial x}\right|_{x_0}=\frac{1}{6}\left(\begin{bmatrix}-1 & 0 & 1\\ -1 & 0 & 1\\ -1 & 0 & 1\end{bmatrix}\otimes f\right) \tag{5-6}$$

这儿引入了一个新的符号⊗，在以上的表达式中表示乘积和。文献中经常看到类似的核算子，它们都采用差分进行求导，并以某种方式将结果取平均，来补偿噪声的影响。这类核算子中最著名的可能是苏贝尔（Sobel）算子:

$$\left.\frac{\partial f}{\partial x}\right|_{x_0}=\frac{1}{8}\left(\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}\otimes f\right) \tag{5-7}$$

68

苏贝尔算子具有中间权重较大的优点。

5.3 通过函数拟合估计导数

这是一种利用图像$f(x,y)$连续性的方法，它将图像亮度看作两个空间坐标的函数，假设一个平面与这个亮度曲面相切于一个点，如图5-2所示。

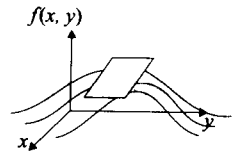


图5-2 将图像亮度看作一个双变量函数的曲面，则它的切面斜率就是两个空间偏导数

这里，我们用一个平面方程表示连续的图像：

$$f(x,y)=ax+by+c \tag{5-8}$$

然后用两个数$\partial f/\partial x=a$和$\partial f/\partial y=b$，来表示边缘强度，则点（x，y）处亮度的变化率可表示成梯度向量的形式：

$$\nabla f=\begin{bmatrix}\dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y}\end{bmatrix}^{\mathrm{T}}=[a \quad b]^{\mathrm{T}} \tag{5-9}$$

给出一幅带有噪声且模糊了的图像f，以及式（5-8）中所用的假设，下面的方法将求出a、b、c的值。

为求得这些参数，首先，注意到式（5-8）可写作$f(x,y)=A^{\mathrm{T}}X$，其中$A^{\mathrm{T}}=[a\ b\ c]$，$X^{\mathrm{T}}=[x\ y\ 1]$。假定图像中所有点集$\aleph\subset Z\times Z$亮度的观察值$g(x,y)$已经得到，而我们希望根据这些点的值拟合出一个平面，使得与这些数据拟合得最好。为了达到这个目的，根据观察数据和函数$f(x,y)$（目前还是未知的），可以得到一个误差函数：

$$E=\sum_{\aleph}(f(x,y)-g(x,y))^2=\sum_{\aleph}(A^{\mathrm{T}}X-g(x,y))^2$$

为了简化，将平方展开且合并同类项，可得：

$$E=\sum_{\aleph}(A^{\mathrm{T}}X)(A^{\mathrm{T}}X)-2A^{\mathrm{T}}Xg+g^2$$

注意：对于矢量A和X，存在$A^{\mathrm{T}}X=X^{\mathrm{T}}A$。通过求和，有：

$$\begin{aligned}E&=\sum_{\aleph}A^{\mathrm{T}}XX^{\mathrm{T}}A-2\sum_{\aleph}A^{\mathrm{T}}Xg+\sum_{\aleph}g^2\\&=A^{\mathrm{T}}\left(\sum_{\aleph}XX^{\mathrm{T}}\right)A-2A^{\mathrm{T}}\sum_{\aleph}Xg+\sum_{\aleph}g^2\end{aligned}$$

69

现在，我们希望能找到A（平面参数）使得能量函数E最小化，于是需要对E求导数，并使导数为0：

$$\frac{\mathrm{d}E}{\mathrm{d}A}=2\left(\sum_{\aleph}XX^{\mathrm{T}}\right)A-2\sum_{\aleph}Xg=0 \tag{5-10}$$

让我们来看看式（5-10）的含义。令$\sum_{\aleph}XX^{\mathrm{T}}=S$（散布矩阵），考虑关于原点对称的邻域$\aleph$，假设在这个邻域中$x$和$y$仅取值$-1$、0、1，那么

$$S=\sum_{\aleph}XX^{\mathrm{T}}=\sum_{\aleph}\begin{bmatrix}x\\y\\1\end{bmatrix}\begin{bmatrix}x & y & 1\end{bmatrix}=\begin{bmatrix}\sum x^2 & \sum xy & \sum x\\\sum xy & \sum y^2 & \sum y\\\sum x & \sum y & \sum 1\end{bmatrix}$$

对于这里所描述的邻域，有

$$\begin{bmatrix}6 & 0 & 0\\0 & 6 & 0\\0 & 0 & 9\end{bmatrix}$$

你不知道这些值是从哪儿来的吧?

好，以下就是求出这些值的方法。对于左上角的点$x=-1$, $y=-1$，有$x^2=(-1)^2=1$；对于上排中间那个点$x=0$, $y=-1$，则有$x^2=0$。这样，通过对邻域中的9个点进行计算，就可以得到$\sum x^2=6$。明白了吗?

有用的发现：如果令邻域关于原点对称，那么散布矩阵中包含x或y奇次方的项为零。

注意：一个常见的错误是把矩阵的右下角当成是1，而不是9！要仔细!

现在，我们可以得到矩阵形式的方程:

$$2\begin{bmatrix}6&0&0\\0&6&0\\0&0&9\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix}=2\begin{bmatrix}\sum g(x,y)x\\\sum g(x,y)y\\\sum g(x,y)\end{bmatrix}$$

70

这就很容易解得a:

$$a=\frac{1}{6}\sum g(x,y)\,x\approx\frac{\partial f}{\partial x}$$

因此，通过在一个邻域范围内进行拟合可以求得这个导数，对于9邻域中的每个点，将其测量值乘以x坐标，并将所有的积加起来。将所有的x坐标写成表格的形式:

−1	0	1
−1	0	1
−1	0	1

这正好是我们直观上求出的式（5-6）的核算子，现在我们正式推导出来了。

我们通过一种优化算法来求得这个结果，在这个例子中，通过最小化平方误差，来求得方程$y=f(x)$中函数$f(x)$的系数，这里f是一个多项式。回顾4.1.2节，这种形式是一种显式的函数表达。

一个术语问题：在以后的讨论中，我们将使用“核半径”这个术语。从中心到最近的那条边之间的像素个数就是半径。比如，一个3×3核的半径为1，而一个5×5核的半径就为2，依次类推。当然也可以将核算子设计成圆形的，但大多数时候，我们采用方形的核算子。

在六角形排列的像素中求图像梯度

在这一节中，我们要通过同一种方法来求得图像的梯度，但这次是在成六角形排列的像素中。参考4A.1节关于坐标系统的讨论，这是同一内容的不同表示。如果仔细地读了这两种表示方法之后，你就会对这个概念有更清晰的理解。

为了找到图像中亮度的梯度，就需要根据一个小邻域的数据来拟合出一个平面，这个平面可以表示成式（5-8）的形式。然后，在相应的方向分别对u和v求偏导数，就可以求得亮度的梯度值。对于一个中心点，取其周围6个点的邻域，并对其进行平面拟合。令该数据点集为z_i, $(i=1,\cdots,6)$，则下面的表达式表示对这6个点的平面拟合的误差，其中平面的参数为a、b、c:

$$E=\sum_{i=1}^{6}(z_i-(au_i+bv_i+c))^2 \qquad (5\text{-}11)$$

71

为了把E表示成易求导的形式，我们将其中的自变量表示成矩阵形式。定义向量$A=[a\ b\ c]^T$，$Z=[u\ v\ 1]^T$，则E可以写成:

$$E=\sum_{i=1}^{6}(z_i-A^TZ_i)^2 \tag{5-12}$$

$$=\sum_{i=1}^{6}(z_i^2-2z_iA^TZ_i+A^TZ_iA^TZ_i) \tag{5-13}$$

首先，由于$A^TZ=Z^TA$，因此可以重新改写式（5-13），为了书写方便，暂时去掉了求和的上下限:

$$=\sum z_i^2-2A^T\sum z_i'Z_i+A^T\left(\sum Z_iZ_i^T\right)A \tag{5-14}$$

式（5-14）括号中的项是个散布矩阵，即数据所在点位置的集合，并将它定义为符号S。为了找到能将E^2最小化的向量A的值，对它作关于A的偏导数:

$$\frac{\partial E}{\partial A}=-2\sum z_iZ_i+2SA \tag{5-15}$$

求S，可得:

$$S=\begin{bmatrix}\sum u_i^2 & \sum u_iv_i & \sum u_i\\ \sum u_iv_i & \sum v_i^2 & \sum v_i\\ \sum u_i & \sum v_i & \sum 1\end{bmatrix}=\begin{bmatrix}4 & -2 & 0\\ -2 & 4 & 0\\ 0 & 0 & 6\end{bmatrix}$$

（如图4-15所示，假设中心点在位置(0,0)，则对邻域中的每个点的u、v坐标进行累加，就可以得到这些数据值。）

定义

$$\sum z_iu_i=Y_u$$

$$\sum z_iv_i=Y_v$$

令式（5-15）的偏导数为0，可以得到一对联立方程:

$$\begin{aligned}4a-2b&=Y_u\\-4a+8b&=2Y_v\end{aligned} \tag{5-16}$$

其解为:

$$b=\frac{1}{6}(2Y_v+Y_u) \tag{5-17}$$

与此类似:

72

$$a=\frac{1}{6}(Y_v+2Y_u) \tag{5-18}$$

对邻域中6个像素的每个像素，将u和v的值代入，因为在u方向的梯度值为a而v方向的梯度值是b，故可以得到梯度向量。

将Y的定义式代入，等式a可以整理为:

$$a=\frac{1}{6}\left(\sum z_iv_i+2\sum z_iu_i\right)=\frac{1}{6}\sum z_i(v_i+2u_i)$$

对于与中心点紧邻的右边的像素，它的u、v坐标为(1, 0)，因此在该点处的图像亮度Z_i就应该乘以(v_i+2u_i)，即$(0+2\times 1)$为2，这就是应该代入到核算子中的值。在用这个方法的时候，应该注意的是这里所得的结果并不正确——这里所得的结果是最优估计的6倍，如果准确的结果比较重要，而不是需要一个与估计值成正比的结果，那么应该再除以6。对于每个点重复这个过程，就可以得到如图5-3所示的核算子。

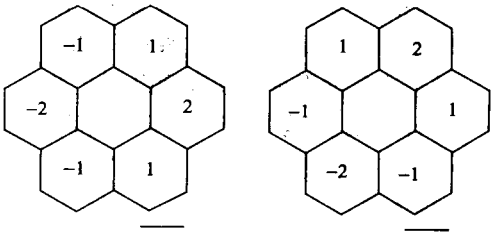


图5-3 在u和v方向用来估计亮度梯度的核算子

本节所描述的拟合的概念是很常见的，它们在整个机器视觉学科中经常出现，在拟合灰度值、表面、直线（在第9章中会提到）、曲线等时会经常需要用这类方法。甚至还有一些方法可以对已知统计参数的数据进行拟合[5.7,5.40]。

5.4 图像的矢量表示

假设将图像中的每个像素点以光栅扫描的顺序进行排列，得到一个长向量，例如，对于4×4的图像：

$$f(x,y)=\begin{array}{|c|c|c|c|}\hline 1 & 2 & 4 & 1 \\ \hline 7 & 3 & 2 & 8 \\ \hline 9 & 2 & 1 & 4 \\ \hline 4 & 1 & 2 & 3 \\ \hline \end{array}$$

$$F=[1\ 2\ 4\ 1\ 7\ 3\ 2\ 8\ 9\ 2\ 1\ 4\ 4\ 1\ 2\ 3]^{\mathrm{T}}$$

73

这称为“字典式”的表示。如果将图像写成这种形式，每个像素就可以以一种简单的索引来确定，例如，$F_0=1, F_4=7, F_{15}=3$，这里索引是从0开始的。

现在，假设要将下面的核算子作用到图像的$x=1$，$y=1$点处：

$$h=\begin{bmatrix}-1 & 0 & 2\\ -2 & 0 & 4\\ 3 & 9 & 1\end{bmatrix}$$

这里的标号也是从0开始的。

点(1,1)对应于图像中的像素F_5。这个核算子的应用可以通过将矢量F与如下的矢量做点积来实现：

$$H_5=[-1\ \ 0\ \ 2\ \ 0\ \ -2\ \ 0\ \ 4\ \ 0\ \ 3\ \ 9\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0]^{\mathrm{T}}$$

现在可以试一下，如果将核算子作用于（2, 2），该采用什么向量？你知道吗？

$$H_{10}=[0\ \ 0\ \ 0\ \ 0\ \ 0\ \ -1\ \ 0\ \ 2\ \ 0\ \ -2\ \ 0\ \ 4\ \ 0\ \ 3\ \ 9\ \ 1]^{\mathrm{T}}$$

试试对于点（2, 1）（$x=2$，$y=1$）：

$$H_6=[0\ \ -1\ \ 0\ \ 2\ \ 0\ \ -2\ \ 0\ \ 4\ \ 0\ \ 3\ \ 9\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0]^{\mathrm{T}}$$

对比一下在点（1, 1）的H_5和点（2, 1）的H_6，它们除了旋转之外是完全一样的。通过构造一个矩阵，使其每一列为这类H中的一个向量，可以对整个图像进行卷积。这样就可以产生如图所示的一个矩阵。通过乘积$G=H^TF$，G（向量形式）就是图像F卷积核H的结果。

对于这个过程可以观察到以下几点：

- 所得的矩阵是一个“循环”矩阵，每一列只是相邻列的一个旋转。
- 以这种方式应用核算子是“核算子是线性算子”这一事实的另一种体现。
- 这种形式给出了一种处理边界条件难题的方法（你考虑到这些了吗？）特别是在第一行时，这些数据是怎么乘上去的？一种办法是将图像旋转并将最下面的数据往上移。
- 矩阵H非常大，如果f是个典型256 × 256像素的图像，那么H就有(256 × 256) × (256 × 256)那么大，这是一个很大的数目。不过不要担心H有那么大，没有（几乎没有）人会去计算H并以这种方式来用。这种形式对于考虑图像并对那些关于图像算子的定理进行证明是非常有用的——这是一种概念，而不是一种计算工具。

$$
\begin{array}{c}
\begin{bmatrix}
\cdots & -1 & 0 & \cdots & 0 & \cdots \\
\cdots & 0 & -1 & \cdots & 0 & \cdots \\
\cdots & 2 & 0 & \cdots & 0 & \cdots \\
\cdots & 0 & 2 & \cdots & 0 & \cdots \\
\cdots & -2 & 0 & \cdots & 0 & \cdots \\
\cdots & 0 & -2 & \cdots & -1 & \cdots \\
\cdots & 4 & 0 & \cdots & 0 & \cdots \\
\cdots & 0 & 4 & \cdots & 2 & \cdots \\
\cdots & 3 & 0 & \cdots & 0 & \cdots \\
\cdots & 9 & 3 & \cdots & -2 & \cdots \\
\cdots & 1 & 9 & \cdots & 0 & \cdots \\
\cdots & 0 & 1 & \cdots & 4 & \cdots \\
\cdots & 0 & 0 & \cdots & 0 & \cdots \\
\cdots & 0 & 0 & \cdots & 3 & \cdots \\
\cdots & 0 & 0 & \cdots & 9 & \cdots \\
\cdots & 0 & 0 & \cdots & 1 & \cdots
\end{bmatrix} \\
\begin{array}{cc} \uparrow & \nwarrow \\ H_5 & H_6 \end{array}
\end{array}
$$

最后，通过使用傅里叶变换可以使与循环矩阵相乘变得相当快，更多的内容将会在后面介绍。

74

5.5 图像的基向量

在前一节，图像被视为一个向量。如果对一幅图像可以这么做，那么当然也可以将小的子图像视为一个向量。考虑一个点的9邻域，如果将该邻域以字典方式表示，就可以很容易构造出一个9维的向量。

从第2章可以知道，任何一个向量都可以表示成一组基向量的加权和，这里将采用这种概念。将式（2-6）改写成如下形式：

$$V=\sum_{i=1}^{9} a_i u_i$$

这里V是一个9维的向量，代表了9个像素组成的邻域，而a_i为标量加权系数，u_i是一组正交基向量的集合。

但是应该采用什么样的基向量呢？另外，什么样的基向量集才是有用的？我们经常使用的集合是笛卡尔基，即：

$$
\begin{aligned}
u_1 &= [1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0]^T \\
u_2 &= [0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0]^T \\
&\vdots \\
u_9 &= [0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1]^T
\end{aligned}
$$

这些基虽然方便简单，但是在这里对我们一点用都没有。那么还存在其他更有用的基吗？（答案是肯定的。）在给出答案之前，你记得在9维的实数空间可能有多少组基向量吗？答案是“亿万”种[⊖]。既然有这么多种选择，就应该选一些好的基向量。为了这个目的，回

⊖ 事实上，正确的答案应该是“无数”种，“亿万”种只是一种工程学上的近似。

顾一下系数a_i的作用，如果一些a_i的值比起其他的a要大的多，这意味着V与$\boldsymbol{u}_i$“非常相似”。通过计算这些a，可以得到一种方法，找到一组与图像最相似的模板邻域。

图5-4描述了一组由Frei和Chen[5.12]提出来的模板邻域。注意，邻域$\boldsymbol{u}_1$在中心水平线下面是负的，而在上面则是正的。因此，这表示存在一条水平的边缘，也就是存在一个$\partial f/\partial y$值很大的点。

$$\boldsymbol{u}_1=\begin{bmatrix}1 & \sqrt{2} & 1\\ 0 & 0 & 0\\ -1 & -\sqrt{2} & -1\end{bmatrix}\quad \boldsymbol{u}_2=\begin{bmatrix}1 & 0 & -1\\ \sqrt{2} & 0 & -\sqrt{2}\\ 1 & 0 & -1\end{bmatrix}\quad \boldsymbol{u}_3=\begin{bmatrix}0 & -1 & \sqrt{2}\\ 1 & 0 & -1\\ -\sqrt{2} & 1 & 0\end{bmatrix}$$

$$\boldsymbol{u}_4=\begin{bmatrix}\sqrt{2} & -1 & 0\\ -1 & 0 & 1\\ 0 & 1 & -\sqrt{2}\end{bmatrix}\quad \boldsymbol{u}_5=\begin{bmatrix}0 & 1 & 0\\ -1 & 0 & -1\\ 0 & 1 & 0\end{bmatrix}\quad \boldsymbol{u}_6=\begin{bmatrix}-1 & 0 & 1\\ 0 & 0 & 0\\ 1 & 0 & -1\end{bmatrix}$$

$$\boldsymbol{u}_7=\begin{bmatrix}1 & -2 & 1\\ -2 & 4 & -2\\ 1 & -2 & 1\end{bmatrix}\quad \boldsymbol{u}_8=\begin{bmatrix}-2 & 1 & -2\\ 1 & 4 & 1\\ -2 & 1 & -2\end{bmatrix}\quad \boldsymbol{u}_9=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$$

图5-4 Frei-Chen基向量

回顾一下如何求投影a_i，求一个向量V投影到一个基向量$\boldsymbol{u}_i$的标量值的方法是作内积$a_i=V^{\mathrm{T}}\boldsymbol{u}_i$。

一种求某个点的邻域与一条竖直边的相似程度的方法，是通过计算邻域向量与竖直边的基向量作内积来实现的。最后一个问题是：计算这种投影与在一个点处用核算子对图像卷积来估计$\partial f/\partial x$有什么不同呢？答案留给读者作为练习。

那么到现在，你已经知道关于线性算子和核算子的（基本上）所有知识了，下面让我们继续讨论先前所提到的核算子的应用——寻找边缘。

5.6 边缘检测

边缘是图像中亮度突然变化的区域，在那里导数（更准确地说，某一个导数）值很大。如图5-5所示，我们将边缘分成阶跃型边缘、屋脊状边缘和斜坡型边缘[5.20]。

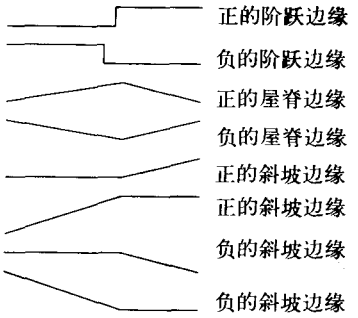


图5-5 经常出现的几类边。注意这里的术语正和负通常指的是一阶导数第一项的符号

75 ~ 76

我们已经（两次）看到过如下所示核算子是如何对关于x的偏导数进行近似的：

$$h_x = \begin{array}{|c|c|c|}\hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline \end{array} \tag{5-19}$$

类似的，

$$h_y = \begin{array}{|c|c|c|}\hline -1 & -1 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \text{或者} \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -1 & -1 \\ \hline \end{array} \tag{5-20}$$

用它来估计$\partial f/\partial y$。

在文献中还出现过其他形式的算子，我们应该知道这是因为历史原因造成的。

$$h_y = \begin{array}{|c|c|c|}\hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array} \text{或者} \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -2 & -1 \\ \hline \end{array} \tag{5-21}$$

重要提示

在软件实现中，y的正方向是向下的。从而扫描的方式是从上到下，从左到右，因此像素点（0，0）在图像的左上角，而且数字是从0开始的，而不是1。我们发现能避免这种混淆最好的方法是在程序中不用“x”和“y”，而用“行”和“列”来表示，而0行在第一行。

然而我们在文章中采用了传统的笛卡尔坐标以保证数学上的正确性，而这会使学生感到更困惑（毕竟这就是那些教授需要存在的原因，对吧？）。

消除这些困惑之后，继续我们的讨论。给定梯度向量：

$$\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix}^{\mathrm{T}} = [G_x \quad G_y]^{\mathrm{T}} \tag{5-22}$$

我们对它的幅值感兴趣（这里称之为“边缘强度”）

77

$$|\nabla f| = \sqrt{G_x^2 + G_y^2} \tag{5-23}$$

以及它的方向

$$\angle \nabla f = \mathrm{a}\tan\left(\frac{G_y}{G_x}\right) \tag{5-24}$$

一种求图像边缘的方法是先计算出一幅“梯度幅值”的图像，然后对它阈值化。试一试作业5.5。

从这些实验可以了解到什么？显然，在用简单的核算子求边缘时出现了一些困难。在后面的大部分课程里面，我们将解决这个问题。首先，让我们改进这种基于核算子的边缘检测。

5.7 用核算子表示可微函数的采样

现在，我们希望你能意识到，迄今为止，你所用的所有边缘检测算子都能同时做两件事情：平滑（也叫“低通滤波”、“去噪”、“平均”或“模糊”）和微分（也叫“高通滤波”或“锐化”）。式（5-6）中的核算子实际上就是对三个所估计的导数求竖直方向的平均值，但因为它对中间像素和其上方和下方的像素采用了一样的权重，故实际上与直观感觉是相反的。

考虑作业5.6的结果，如果答案正确，那么离中心越远，核算子的值也就会越大。那就更不对了。为什么那些离求导那个点越远处的数据点反而在估计中起作用更大？这与我们给出的关于用所有的像素来拟合一个平面的假设是矛盾的，假设显然不是这样的。

因此有一种更好的方法——对中心点取更大的权重，你已经看到过这种情况了——苏贝尔算子，如式（5-7）那样。但是现在，让我们更严格一点。采用一个中间值比较大的核算子对图像进行模糊处理，然后再求微分。有很多种这样的核算子，例如：三角波形的或高斯的，但透彻的研究[5.28]已经证明采用高斯核算子进行模糊处理效果更好。这个过程可以写成：

$$d=\frac{\partial}{\partial x}(g\otimes h)$$

这里g是一幅标准的图像，h是高斯核算子，d是求得的导数图像。现在给出线性系统的一条重要定理：

对于线性算子D和$\otimes$，有：

$$D(g\otimes h)=D(h)\otimes g \tag{5-25}$$

78

式（5-25）表示并不需要首先做模糊处理然后再求微分，而只要预先计算好模糊核算子的微分，然后简单地将所得的核算子作用于图像就行了。

让我们看一下是否还记得如何对一个二维的高斯函数求导（你是否忘记二维函数了？）。

一个d维多变量的高斯函数的一般形式为：

$$\frac{1}{(2\pi)^{d/2}|K|^{1/2}}\exp\left(-\frac{[x-\mu]^{\mathrm{T}}K^{-1}[x-\mu]}{2}\right) \tag{5-26}$$

这里K是协方差矩阵，而μ是一个均值向量。因为我们需要一个中心在原点（中心像素的位置）的高斯函数使得$\mu=0$，而且没有理由说一个方向比其他的方向更好，因此选K为对角阵（各向同性）：

$$K=\begin{bmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{bmatrix}=\sigma^2 I \tag{5-27}$$

对二维的情况，式（5-26）可以简化为：

$$h(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(\frac{-[x\ \ y]^{\mathrm{T}}[x\ \ y]}{2\sigma^2}\right)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right) \tag{5-28}$$

且：

$$\frac{\partial}{\partial x}h(x,y)=\frac{-x}{2\pi\sigma^4}\exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right) \tag{5-29}$$

如果我们的目标是边缘检测，这样就可以了。但是如果需要精确的求导，特别是高阶导

数，如果采用高斯核算子，由于它会模糊图像，因此显然会引入误差，而这只能进行部分的补偿[5.39]。尽管如此，采用有效的求导核算子仍然是最简单的方法之一。

这里有些一维高斯的导数，将来可以作为参考。即使许多时候并不需要归一化因子 $\sqrt{2\pi}$（它保证高斯积分为1），这里还是将这个系数加进去了，这样就与许多文献中的表达式一致了。这里下标则用来表示导数，即：

$$G_{xx}(\sigma,x)=\frac{\partial^2}{\partial x^2}G(\sigma,x)$$

这里$G(\sigma,x)$是x的高斯函数，均值为0，标准差为σ：

$$\begin{aligned}
G(\sigma,x)&=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right)\\
G_x(\sigma,x)&=\frac{-x}{\sqrt{2\pi}\sigma^3}\exp\left(-\frac{x^2}{2\sigma^2}\right)\\
G_{xx}(\sigma,x)&=\left(\frac{x^2}{\sqrt{2\pi}\sigma^5}-\frac{1}{\sqrt{2\pi}\sigma^3}\right)\exp\left(-\frac{x^2}{2\sigma^2}\right)\\
G_{xxx}(\sigma,x)&=\frac{x}{\sqrt{2\pi}\sigma^5}\left(3-\frac{x^2}{\sigma^2}\right)\exp\left(-\frac{x^2}{2\sigma^2}\right)
\end{aligned}\tag{5-30}$$

79

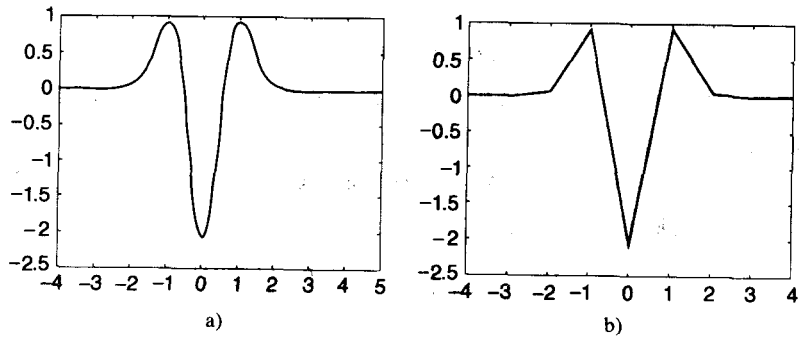


图5-6 a) 中心在0处的一维高斯的二阶导数 b) 3点近似

让我们来看一下关于如何使用这些公式，以及与它们二维形式等价的导数核算子的更多的细节。

得到这些高斯导数核算子值的最简单的方法就是简单地沿着它们的负方向将x=0, 1, 2等代入，从而产生核算子中的数值。随之出现的第一个问题是“σ应该是什么？”为了解决这个问题，我们通过对一维高斯计算二阶导数求得核算子的元素，而其他的导数则可以用同样的方法求得。看到图5-6或许会问“是否存在这么一个σ，使得二阶导数在x=－1和x=1处存在最大值”。显然是存在的，这个值即 $\sigma=1/(\sqrt{3})$。给定σ的这个值，可以计算出高斯的二阶导数在整数点x={－1, 0, 1}的值。在x=0，有 $G_{xx}(1/\sqrt{3},0)=-2.07$；在$x$=1处，有 $G_{xx}(1/\sqrt{3},1)=0.9251$。就这么完了？一点都不难吧？不幸的是，这并没有完成。非常重要的一点是，核算子的所有元素加起来的和应该是0，如果不是的话，在多次迭代之后，（如第6章中的迭代算法）就不能保持正确的亮度值。另外，核算子也应该是对称的。这在本质上定义了高斯的二阶导数。与这些条件接近而又最合理的值是{1, －2, 1}，既满足对称性，同时累加的和也为0。

然而这些并不能教给我们什么，让我们来看5 × 1的核算子，从中我们可以学到更多的东西。这里需要满足以下条件:

- 核算子的元素应尽可能接近高斯导数的值。
- 核元素累加的和为0。
- 除非需要特殊处理，核算子应该以中心对称。

可以求得5个元素的一维高斯的元素，对于x={ − 2, − 1, 0, 1, 2}，如果假定 $\sigma = 1/\sqrt{3}$，就可以得到[0.0565, 0.9251, − 2.0730, 0.9251, 0.0565]，但不幸的是，这些值加起来并不等于0。保证核算子元素累加为0这点非常重要，而对于实际的值是否精确则并不显得那么重要。因此在这种情况下应该怎么做呢？采用约束优化。一种方法是通过求得高斯的二阶导数，并在积分为0的条件下，使值与之尽可能接近。对于更复杂的情况，作者采用Interopt[5.3]的方法来解这个数值优化问题，但是你也可不用数值方法解这个问题。实现的方法如下：首先，理解这个问题（像上面提到的5个点的情况）：我们希望找到5个数，使之尽可能接近[0.0565, 0.9251, − 2.0730, 0.9251, 0.0565]，同时满足5个数相加为0的约束条件。利用对称性，实际上只有3个数，定义为[a, b, c]。为了表示方便，引入3个常量α=0.0565, β=0.9251, γ= − 2.073。为了找到与这3个数相像的a、b和c，可以写出均方误差（MSE）的形式: 80

$$H_0(a,b,c) = 2(a-\alpha)^2 + 2(b-\beta)^2 + (c-\gamma)^2 \tag{5-31}$$

利用拉格朗日乘子的概念，通过对另一个目标函数最小化可以找到最佳的a、b和c:

$$H(a,b,c) = 2(a-\alpha)^2 + 2(b-\beta)^2 + (c-\gamma)^2 + \lambda(2a+2b+c) \tag{5-32}$$

那些对于采用拉格朗日乘子进行带约束条件优化不熟悉的同学，这里还需要再解释一下。前面的λ（λ是拉格朗日乘子）是约束项，如果找到正确的a、b和c，它就应该等于0。通过最小化H，可以找到使H_0 最小化的参数，并同时满足约束条件。

为了最小化H，求导并令导数为0:

$$\begin{aligned}\frac{\partial H}{\partial a} &= 4a - 4\alpha + 2\lambda \\ \frac{\partial H}{\partial b} &= 4b - 4\beta + 2\lambda \\ \frac{\partial H}{\partial c} &= 2c - 2\gamma + \lambda\end{aligned} \tag{5-33}$$

令偏导数为0，化简并加入约束后，得到如下的线性方程组:

$$\begin{aligned}a &= \alpha - \frac{\lambda}{2} \\ b &= \beta - \frac{\lambda}{2} \\ c &= \gamma - \frac{\lambda}{2} \\ 2a + 2b + c &= 0\end{aligned} \tag{5-34}$$

解之，得如表5-2所示的解集。 81

用同样的方法可以计算出估计二维高斯偏导数的核算子。

图5-7表示一个各向同性的高斯关于x的一阶导数。你可以利用这个机会推导一些其他的核算子，作为练习。

表5-2 一维高斯的导数

一阶导数，3×1	[0.2420, 0.0, −0.2420]或[1, 0, −1]
一阶导数，5×1	[0.1080, 0.2420, 0, −0.2420, −0.1080]
二阶导数，3×1	[1, −2, 1]
二阶导数，5×1	[0.07846, 0.94706, −2.05104, 0.94706, 0.07846]

0.0261	0	−0.0261
0.1080	0	−0.1080
0.0261	0	−0.0261

图5-7 3×3的一阶导数核算子

在本章，我们探讨了基于核算子的边缘检测算子的思想。我们发现无论怎样，带噪声的图像会导致边缘:

- 在某些地方太粗。
- 在某些地方缺失。
- 在某些地方引入额外的边缘。

那就是现实——采用简单的核算子不可能得到更好的结果。在第6章，我们将探讨解决这些问题的方法。

你也肯定猜到了，正如我们希望的那样，在图像中检测边缘，除了简单地阈值化导数之外，还有其他的方法。在后面几节，我们将提到一些这一类的方法。

高阶导数

我们刚才已经看到利用高斯导数是如何求得二阶和三阶导数的。既然问题已经提出来了，而且在以后也需要知道这个术语，因此，这儿定义了两个基于二阶导数的标量算子：拉普拉斯算子和二次变分。

82 在点x，y处亮度的拉普拉斯算子为:

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

而亮度的二次变分为:

$$\left(\frac{\partial^2 f}{\partial x^2}\right)^2+\left(\frac{\partial^2 f}{\partial y^2}\right)^2+2\left(\frac{\partial^2 f}{\partial x\partial y}\right)^2$$

拉普拉斯算子可以用一些核算子进行近似，包括如下的核算子:

−1	2	−1
2	−4	2
−1	2	−1

5.8 计算卷积

回顾5.2.1节，对于任何核算子而言，它与卷积之间唯一的不同点在于x和y的方向。因此计算卷积的有效方法也就是应用核算子的有效方法。卷积运算可以按以上所说的方法直接计算出来，它是在每个像素邻域中所作的一个简单的乘积之和，然而它也可以通过傅里叶变换

求得。卷积的傅里叶变换等于两个变量各自傅里叶变换的乘积。也就是说（定义卷积为⊗），牵涉到的是:

$$g(x,y)=f(x,y)\otimes h(x,y)$$

定义两图像的傅里叶变换和卷积核算子为:

$$G(\omega_x,\omega_y)=F(g(x,y))$$
$$F(\omega_x,\omega_y)=F(f(x,y))$$
$$H(\omega_x,\omega_y)=F(h(x,y))$$

这里的符号F指做傅里叶变换。回顾4.1.5节，图像（双变量的函数）的傅里叶变换本身也是一个双变量的函数。变量ω_x和ω_y分别指x、y方向的空域频率。那么G就是F和H的乘积:

$$G(\omega_x,\omega_y)=F(\omega_x,\omega_y)\cdot H(\omega_x,\omega_y) \tag{5-35}$$

对每个空域频率值（ω_x和ω_y），两个变换的“乘积”意味着两函数值的乘积。（一般来讲，这些值都是复数。）

83

出于计算的复杂性，逐点相乘的计算量是很可观的。考虑用$L\times L$的核算子卷积$N\times N$大小图像的复杂度，在空域做卷积，对于每个像素点要做$L\times L$次乘法，共N^2L^2次。采用傅里叶变换（假定你采用快速傅里叶变换）以如下方式进行。（具体细节超出了本节范围，但我们假定一个$N\times N$图像的傅里叶变换本身也是一个相同大小的二维矩阵。）

- f的变换: $N^2\log N$。
- h的变换: $L^2\log L$。
- 进行适当的操作，比如填充，使H和F大小相同。
- 用F乘以H: N^2。
- 将结果反变换: $N^2\log N$。

尽管通过变换域并不方便，但是如果这四项的和小于N^2L^2，计算起来还是很有效的。但对于某个具体尺寸大小的图像和核算子，应该怎么做呢？幸运的是，在不同的尺寸大小下，有人已经分析了采用傅里叶变换和空域方法计算卷积的相对效率，结果如图5-8所示。从图可知，如果核算子大于15 × 15，就应该采用傅里叶方法，如果核算子小于7 × 7，则应该采用空域的方法。因为FFT要求图像的大小为2的乘方，所以区域边界需要进行特殊的变动，这样一来对于其他尺寸的图像而言就会增加计算复杂度。

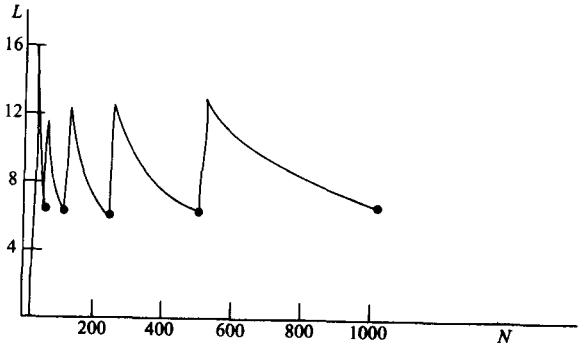


图5-8 用$L\times L$的核算子卷积$N\times N$的图像的计算效率。曲线上方的组合表示采用傅里叶方法更有效，而曲线下方是采用空域的方法更有效（取自Pratt[5.33]）

84

5.9 尺度空间

"尺度空间"（scale space）是著名的图像金字塔概念的新补充，最先在Kelly[5.19]的图像处理中被采用，后来又以各种方式进行扩展（见[5.5、5.8、5.30、5.32]等）。在金字塔方式中，产生了同一图像的一个表示序列，邻近的高一级图像通过下一级图像以2：1的下采样（或求平均）方式产生（见图5-9）。

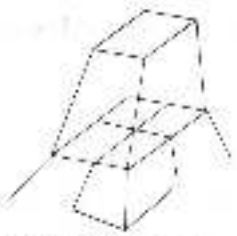


图5-9 金字塔是图像序列的一种数据结构，当前层图像的每一个像素是下一层图像4个像素的平均值

图5-10表示一个高斯金字塔，它通过在2:1的下采样之后在每一层上用高斯核制得到。当看到这幅图像之后，读者可能会提出一个很有趣的问题：从金字塔中所有的数据里，是否能重构原来的图像？答案是"不能"，因为在每一层，已经丢弃了高频信息。

虽然单独的高斯金字塔并没有包含足够的信息来构造原始图像，我们还是可以构造出一个包含足够信息的金字塔。为了达到这个目的，我们采用一个"拉普拉斯算子"金字塔，通过计算高图像的一个相似表达来进行构造，这就保留了高频信息（见图5-11）。通过组合这两组金字塔图像可以重构出原始的图像。

在现代的尺度空间的表示中，我们保留了这个概念：每一层是前一层的模糊结果，但不是子采样，每一层与前一层一样大小，但是比之更模糊。通常，每一层通过用一个方差为σ^2的高斯卷积原图像而产生，而每一层的σ是不同的，这个方差就成为了"尺度参数"。显然，在尺度的高层中（σ大），仅那些大特征是可见的。当在这章中讨论小波时，我们还将看到关于尺度空间的更多细节。

图5-10 一种高斯金字塔，通过在每一层上用高斯模糊后运行2:1的子采样进行构造而得

图5-11 这个拉普拉斯金字塔实际上是通过一个高斯差分运算得到的

85

5.9.1 四叉树

四叉树（quad tree）[5.21]是一种数据结构，在这种结构下，图像以递归的方式被分成4块，并对应于树结构的节点。这四个部分被标记成西北（NW）、东北（NE）、西南（SW）和东南（SE）。树节点和图像的对应关系可以通过一个图例很好地表示出来（见图5-12）。

在对二值图像编码的过程中，可以很直接得到产生图像四叉树的方法：如果四叉树中一幅图是均匀的（纯黑或纯白），则令它为一个叶节点，否则将它分成四份并在四叉树中再加入一层。递归重复上述过程直到图像块成为像素大小或成为均匀区域。

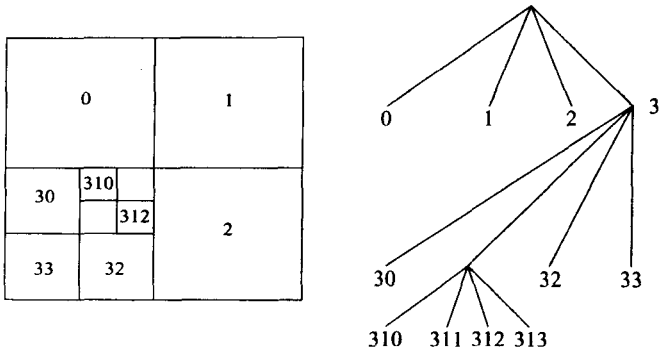


图5-12 一幅图像被分成4块。每一个不均匀的图像块将被继续分割。这种分割可以用树结构来表示 86

将四叉树集成到金字塔结构中是一件很容易的事情，对于每个节点，仅需要保存它所有子节点的均值，这样金字塔中所有的信息就存入四叉树里了。

如果图像有很大一部分区域是均匀的，那么采用四叉树似乎是一种很有效的存储和传输方式。然而通过对各种图像进行实验，即使是视频中两帧间的差分图像，这种看法也不一定正确。既然差分图像仅在物体运动的部分不是零，而大部分的区域都是零，那么似乎显然可以得出四叉树是一种很有效的图像存储方式的结论。其实并非如此，即使在那种情况下，管理树结构的开销也超过了存储带来的收益。所以，令人惊讶的是，四叉树也不是一种有效的图像压缩技术。但当其作为表示金字塔的一种手段时，在表示尺度空间时仍有优势。

采用四叉树的另外一个缺点是，物体的微小移动将导致树结构的巨大变化，也就是说，树结构并不具有旋转不变性和平移不变性。事实上，它甚至是不鲁棒的，这里“鲁棒”是指物体的小位移只在表示上相应地产生微小的变化。采用四叉树表示每个物体的子图像，而不是整个图像，可以在某种程度上避开这种缺点。

四叉树拓展到三维就叫做“八叉树”，同样的原理在那里也是适用的。

5.9.2 高斯尺度结构

我们已经知道如何模糊一幅图像。想像一个实验：对于一幅图像，采用一个标准差为1的高斯核算子对它进行模糊处理，可以得到一幅新图像，称其为图像1；采用一个标准差为2的高斯核算子对原图进行模糊，称所得图像为图像2；继续这个操作，可以得到一组图片。你可以将它们看作是一个堆栈，越上面的图像模糊得越厉害。我们称顶层图像为原图像的“大尺度”表示。显然，并不一定需要采用整数值的标准偏差，因此可以以任何需要的分辨率来建立尺度空间。尺度空间表示的基本前提是可以在不同尺度上跟踪某些图像特征，而这些特征在不同的尺度上的变化能够体现出图像的某些性质。在[5.25，5.26]中较规范地定义了尺度空间，它具有如下属性：

- 所有信号应该在同一个定义域中被定义（没有金字塔）。
- 随着尺度参数的增长，输出（图像）的表示应该越来越粗。
- 一个信号在较粗的层次中包含的结构，要比较细的层中少。如果将局部极值看作平滑度的一种度量，那么随着尺度变粗，极值是非增的。这种属性称为“尺度空间的因果关系”。
- 所有的表示都应该通过在原图上应用一个卷积核算子来产生。

最后一条性质当然是有争议的。因为卷积在形式上需要一个线性、空间不变的算子。实现尺度空间的一种有趣方法是，采用灰度值（gray scale）数学形态学方法，使用逐渐变大的 87

结构元素进行平滑，这可以避免卷积所需的条件。

尺度空间的概念除了可以表示亮度之外，还可以用来表示纹理[4.16]，甚至也可以表示概率密度函数（在这种情况下，尺度空间的表示方法成为了一种聚类算法[5.24]）。在课程中，我们还将看到一些采用尺度表示的例子。

尺度空间表示最有趣的方面之一，是我们的老朋友高斯的性能问题。当被用作核算子的时候，高斯的二阶导数（二维情况下，为高斯的拉普拉斯算子:LOG）被证明具有很好的性质。尤其值得注意的是LOG的过零点可以很好地指出边缘位置。有人很容易会问："高斯是不是一种最好的平滑算子，用它生成尺度空间需要的核算子？"另一种说法是：我们需要这么一种核算子，在更大的尺度下，它的两阶导数不会产生出新的过零点。事实上，在以后更一般的形式下，我们可以表述这个需求。

如果在一个点，某算子具有极值，最大值或最小值，则定义该点为"特征点"。尺度空间因果关系的概念表明，随着尺度的增大，图像将越来越模糊，将不会产生新的特征。高斯则是具有这种属性[5.1，5.2]的唯一的核算子（线性算子）。人们还研究了非线性算子在什么样的条件下，具有尺度空间因果关系[5.22]。

下面的例子将阐述尺度空间因果关系的思想。图5-13是在图像中沿一条直线方向的亮度

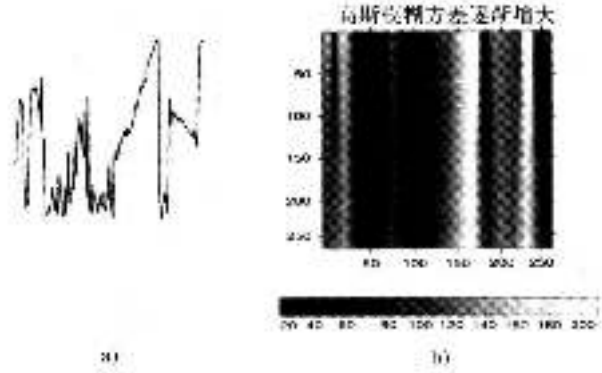


图5-13 a) 图像中沿扫描线方向的亮度曲线。b) 扫描线的尺度空间表示，向下为尺度增加的方向，可见从上到下，没有增加新的特征

曲线，而尺度空间是通过在直线方向上用一个一维且方差逐渐增大的高斯进行模糊得到的。在图5-14中，可以看到高斯的拉普拉斯算子，以及拉普拉斯算子改变正负号的点。该例子中

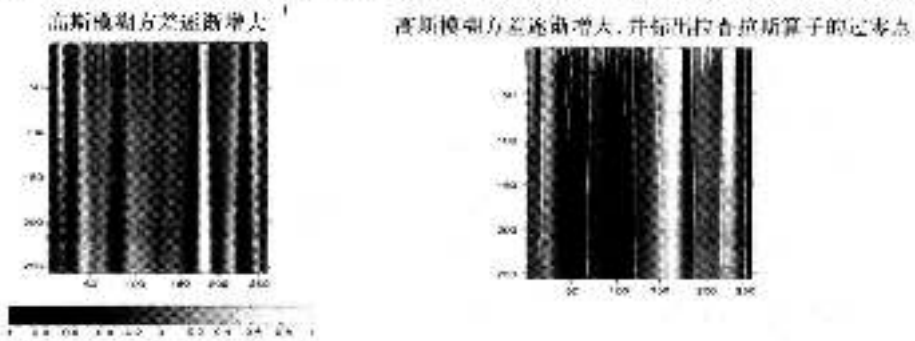


图5-14 尺度空间表示的拉普拉斯算子以及拉普拉斯算子的过零点，因为这是一维数据，因此拉普拉斯算子和二阶导数没有区别

的特征点，也就是过零点（是很好的候选边缘）显示在右图。随着尺度增大可以看到，特征点（这种情况下为过零点）并不随着尺度增大而增多。从上（低尺度）到下（高尺度），一些特征消失了，但并没有新的特征产生。

这个思想的一种很显然的应用是在图像中用它来识别重要的边缘，为了达到这个目的，我们可以沿尺度进行上溯，找到那些边，然后跟踪它们到低尺度上去。

5.10 量化边缘检测算子的精确性

在设计边缘检测算法的时候有很多种选择，因此，需要一些客观的方法来评价一个边缘检测算子是否比别的算子性能更好。Pratt[5.33]提出了一个简单的公式来解决这个问题，这个公式是在边缘点上求和:

$$R=\frac{1}{I_N}\sum_{i=1}^{I_a}\frac{1}{1+\alpha d^2} \tag{5-36}$$

这里$I_N=\max(I_I, I_A)$，I_I 定义为检测到边缘点的数目，而I_A 为图像中实际的边缘点个数。常数α是个尺度因子，d是从检测到的边缘点到最近的实际边缘的距离度量。当然，这个公式需要知道边缘点真实位置的信息。因此，它的一种重要用途是应用于合成数据，因为仅在这种数据中可以很明确地知道边缘点的实际位置（见[5.4]）。

88 ~ 89

5.11 人们的做法

两位神经生理学家David Hubel和Thorsten Wiesel[5.13，5.14]在大脑中，特别是在视觉皮层中插了一些电极——先是猫[⊖]然后是猴子。他们在记录神经细胞的兴奋信息时，给这些动物以各种视觉刺激。他们观察到了一些吸引人的结果：首先，某些细胞仅在观察到某种图案时才出现兴奋。比如，有种特殊的细胞仅当它在观察到一条在某个特定角度从亮到暗的边缘时才兴奋。有证据表明，他们测量的每个细胞都可以从邻域得到输入，这就是所谓的“感受野”。有不同类型的感受野，它们可能都连到同一种光线检测器，它们的组合可以实现边缘检测和其他一些操作的功能。Jones和Palmer[5.17]仔细地绘制了这些感受野函数，确定这些感受野函数可以用Gabor函数精确地表示，这些Gabor函数具有式（5-37）的形式:

$$G(x,y)=\frac{1}{2\pi\sigma\beta}\exp\left(-\pi\left(\frac{x^2}{\sigma^2}+\frac{y^2}{\beta^2}\right)\right)\exp(i[\xi x+\nu y]) \tag{5-37}$$

第一个指数项是个二维的高斯，它的等亮度线为以x、y轴为长短轴的椭圆。（如果恰好碰到关于x、y轴倾斜的感受野，就需要旋转坐标系统来使这个等式成立。）第二个（复数）指数项代表了一个平面波。式（5-37）假定原点在高斯中心。图5-15画的是一个Gabor滤波器。

当式（5-37）的参数能在活的生物中测量到的时候，[5.23]观测到了关于这些参数值的如下有趣的现象:

- 椭圆的长短轴之比β/α为2:1。
- 平面波有沿椭圆短轴传播的倾向。
- 频率响应的半幅带宽大约为沿最优方向的1到1.5个倍频。

⊖ 我们本来想在这里讲一只死猫的笑话，像“在试验过程中的一只猫死了，但它的行为却是不会改变的”，但出版商告诉我们这将冒犯一些人，所以我们不得不将他们取消了。

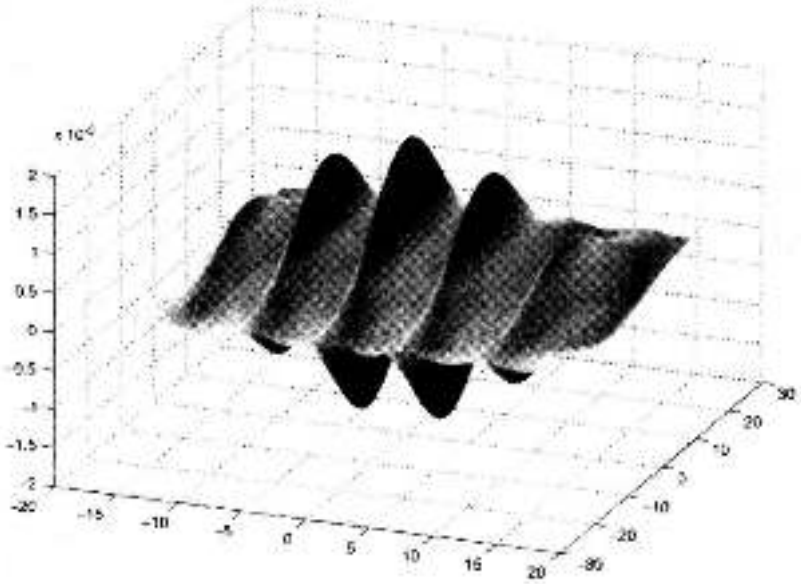


图5-15 Gabor滤波器，注意：它的正、负响应与本章前面我们所推导的那些十分相似

那么在我们的大脑中有Gabor滤波器吗？或者是类似于Gabor的小波发生器？我们并不了解你的大脑，但是Young[5.41]已经发现了哺乳动物视网膜对刺激反应的特点，并且观察到，那些能够用Gabor滤波器或小波进行建模的感受野，也一样可以用所谓的“偏移高斯的差分”的核算子进行描述，这些核算子实质上是带有高斯偏移的LOG算子。图5-16描述了一个Gabor和四阶高斯导数的一段波形，可以看到它们之间有很明显的差别，最主要的一个是Gabor的波纹可以持续增加，而四阶导数则仅有3个极值。然而，由于神经学实验的精度有限，它们是相同的。问题只不过是由于数据的度量不够精确，以致于有许多曲线可以拟合它。

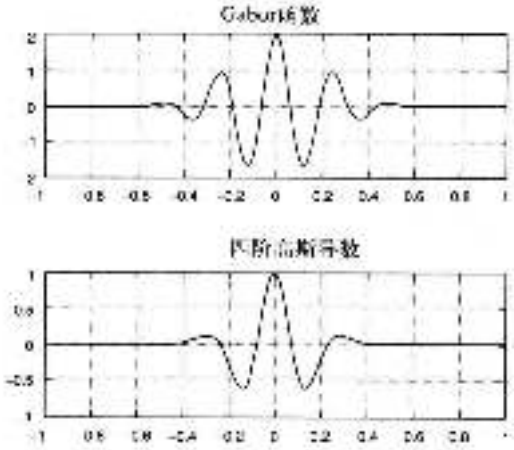


图5-16 高斯滤波器的一段和高斯四阶导数相似片断的比较

问题的实质在于，对于大脑的工作原理我们还没有任何线索，而且事实上我们对于视网膜的了解也不是很多。因此就出现了有两三个数学模型能对感受野的行为进行建模的情况。

5.12 结论

在这章中并没有显式地利用一致性，然而在作业10.1中，将可以看到一个利用一致性进行边缘检测的例子。在那个问题中，将要求你设计一个算法，该算法利用了相邻的边缘点具有平行的梯度方向这个信息，也就是说，如果像素A与像素B相邻，那么在A点处的梯度方向与B点处的梯度方向平行（或几乎平行），这就增加了两个像素属于同一条边的可信度。

在这一章，我们讨论了一些推导核算子的方法，当将这些核算子应用到图像中去的时候，将对各种类型的边缘产生很强的反应。

- 应用了导数的定义。
- 通过最小化误差平方和，采用解析函数拟合曲面。
- 将子图像转换为向量，并将这些向量对一组特定的基向量做投影，而这组基向量则刻画了类似于边缘的特征。
- 利用了核算子的线性性质，通过达到交换模糊和差分的顺序来构造核算子，它们是一些特殊的模糊核算子的导数。我们采用约束优化的方法和拉格朗日乘子法来解决这个问题。

5.13 术语

基向量	basis vector
卷积	convolution
相关	correlation
Gabor滤波器	Gabor filter
图像梯度	image gradient
内积	inner product
核算子	kernel operator
拉格朗日乘子	Lagrange multiplier
字典序	lexicographic
线性算子	linear operator
对数	LOG
投影	projection
金字塔	pyramid
四叉树	quad tree
尺度空间	scale space
误差平方和	sum-squared error

作业 5.1

前一节讲了如何通过拟合一个平面来估计一阶导数，显然这种方法对于二阶导数不起作用，因为一个平面的二阶导数处处为零。但采用双二次函数，用同样的方法：

$$f(x,y) = ax^2 + by^2 + cx + dy + e$$

那么令$[a\ \ b\ \ c\ \ d\ \ e]^{\mathrm{T}}=A$

$$[x^2\quad y^2\quad x\quad y\quad 1]^{\mathrm{T}} = x$$

求用来估计 $\frac{\partial^2 f}{\partial x^2}$ 的3×3的核算子。

作业 5.2

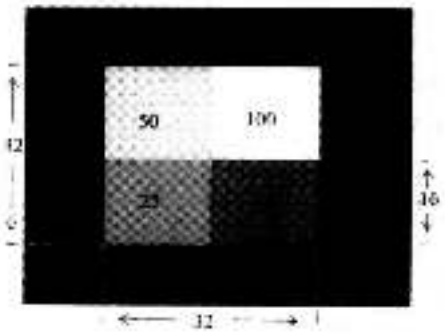
这是上个作业的另一部分！（这比上面那题简单多了——易如反掌），采用同样的方法，利用平面方程，求一个5×5的核算子，来估计其中点位置的 $\frac{\partial f}{\partial x}$。

作业 5.3

92/93 判所图5-4中的u_1和u_2实际上是否正交[1]。如果不正交，推荐一种修改方案或其他方法，使得我们所有人都可以把它用作基函数。

作业 5.4

1) 编写一个程序，产生如下所示64×64的图像。图像应该包含均匀的亮度区域，其尺寸和亮度如图所示。并将它存成一个文件，同时取名为“SYNTH1”。



2) 编写一个能读取SYNTH1的程序，并应用以下模糊核算子，将结果写入文件“BLUR1”。

1/10

1	1	1
1	2	1
1	1	1

3) 将方差为$\sigma^2=9$的高斯随机噪声加到BLUR1，并将结果写入文件“BLUR1.V1”。

作业 5.5

编一段程序，将下面两个核算子（参考关于苏贝尔算子的文献）作用于图像SYNTH1，BLUR1和BLUR1.V1。

$h_x=$

−1	0	1
−2	0	2
−1	0	1

$h_y=$

−1	−2	−1
0	0	0
1	2	1

94 为达到这个目的，执行如下操作：

1) 将h_x作用于输入图像，并将结果在内存中存成临时文件（记住数可以是负的）。

2) 将h_y作用于输入图像，并将结果在内存中存成另一个矩阵。

3) 计算第三个矩阵，使得该矩阵中每个点都是刚才存的两个矩阵相应点的平方和。最后对每个点开方，并保存结果。

4) 检查得到的值。大致上高的值预示着有边缘存在。选择一个阈值并计算一幅新的图像，使图像在边缘强度高于阈值的地方为1，其余为0。

5) 对加了噪声的模糊图像进行1)~4)同样的操作。

6) 写一篇报告。并打印所有三幅二值输出图像。是否有边缘点丢失了？是否有点被人为地加进来了？是不是有些边太粗了？讨论结果对于噪声、模糊和阈值选取的敏感性。

要彻底，这是一个研究性的课程，不仅要正确地达到作业的最低要求，也需要创新性并探索新的想法。

作业 5.6

在作业5.2中，推导了一个5 × 5的核算子。对于$\partial/\partial x$和$\partial/\partial y$，利用这个核算子重复作业5.5。

作业 5.7

1) 验证我们在式（5-30）中所做的推导。求一个3 × 3的核算子，使之能实现式（5-30）中高斯导数的垂直边缘算子。令σ=1和σ=2，并求出这两个核算子。重复如上过程求5 × 5的核算子。讨论选取不同的σ以及核算子的大小所产生的影响。假设核算子可以包含实数（浮点）。 95

2) 假定核算子只能包含整数，求能近似产生同样效果的核算子。

作业 5.8

在5.7节，讨论了对于求离散高斯核算子有用的参数。求σ的值，使之在$x=-1$和$x=1$处二阶导数达到最大。此时$\sigma=1/\sqrt{3}$吗？

作业 5.9

用多项式拟合的方法来估计$\partial^2 f/\partial y^2$。下面哪个多项式最合适：

a) $f=ax^2+by+cxy$　　c) $f=ax^3+by^3+cxy$

b) $f=ax^2+by^2+cxy+d$　　d) $f=ax+by+c$

作业 5.10

用下面的表达式拟合一个3 × 3邻域内像素点的值：$f(x, y)=ax^2+bx+cy+d$。从这个拟合求出能估计关于x的二阶导数的核算子。

作业 5.11

用函数$f=ax^2+by^2+cxy$来求一个能估计$\partial^2 f/\partial y^2$的核算子。下面哪个是产生的核算子？（注意：下面的结果没有包括尺度因子。因此，下面最佳值将是正确答案的倍数。）

(a)

6	4	0
4	6	0
0	0	4

(c)

2	6	2
−4	0	−4
2	6	2

(e)

6	4	0
4	6	0
0	0	1

(b)

1	−2	1
3	0	3
1	−2	1

(d)

1	0	−1
−2	0	2
1	0	−1

(f)

1	2	1
−2	0	−2
−1	−2	−1

96

作业 5.12

假设估计$\partial^2 f/\partial x\partial y$的核算子为:

2	−1	0
−1	0	1
0	1	−2

(用这个核算子来估计这个导数可能并不正确，但它却并不影响结果。假定它是正确的。)那么用

4	1	0
1	0	1
0	1	4

来估计$(\partial^2 f/\partial x\partial y)^2$正确吗？对你的结果进行解释。

专题5A 边缘检测器

边缘检测的过程并不仅仅是对梯度简单地取个阈值，我们需要更精确地知道边缘的位置，而不只是对梯度简单地取个阈值后加以检测。在这个领域里有两种著名的方法："Canny边缘检测器"[5.6]和"小平面模型"[4.18]。这里我们只讨论Canny边缘检测器。

5A.1 Canny边缘检测器

边缘检测算法首先要求出每个点的梯度值。Canny采用了2 × 2而不是像我们所用的3 × 3的算子，但这并不影响这方法的原理。一旦我们有了两个偏导数的估计值，就可以利用从(5.22)到(5.24)的公式计算出梯度的大小和方向，得到两个图像，$M(x, y)$和$THETA\ (x, y)$。现在就可以很容易判断哪些像素点位置具有大的梯度值。然而这并不够，因为我们需要细化幅度矩阵，仅留下最大值，并求得一个新的图像$N(x, y)$。这个过程称之为非最大值抑制(NMS)[⊖]。

实现NMS有很多种方法，然而主要的思想如下：首先，初始化$N(x, y)$使其等于$M(x, y)$。然后对于每个点(x, y)，在梯度方向和反方向各找一个像素点。如果$M(x, y)$(问题中的点)不是3个点中最大的，则将$N(x, y)$置0，否则N的值不变。

97

⊖ 抑制非最大值有时候用复数形式(nonmaxima suppression)来写。表达式意思比较含糊，它可能包含压制每个不是最大值的点，或压制所有不是极大值的点。我们选择采用单数形式(nonmaximum suppression)。

NMS之后，我们就得到了正确定位的边缘，而且仅为单像素宽。然而这些新边缘仍然存在我们以前所指出的那些问题——因噪声（错误命中）而产生额外的边缘点，或者因为模糊或噪声（错误丢失）丢失边缘点。使用双阈值的方法可以进行改进。用两个阈值τ_1和τ_2，其中τ_2比τ_1要大得多。将这两个阈值作用到图像$N(x, y)$产生两幅二值边缘图，分别定义为T_1和T_2。因为T_1由低阈值产生，它比起T_2来有更多的误检率。于是T_2中的点被看作为真实边缘的一部分。将T_2中相连的点复制输出到一幅边缘图像，找到每条边的端点，并在T_1中寻找它的延长部分，直到与T_2中另外一个边缘点相连为止，否则认为T_1中没有与之相连的边。

在[5.6]中Canny也阐述了一些灵活的近似方法，可以明显加快速度。

5A.2 改进边缘检测

Tagare和deFigueiredo[5.34]描述了边缘检测的过程，如下:

1) 用一个滤波器对输入图像进行卷积，平滑地对输入图像进行微分，使得在边缘位置和边缘位置附近产生较大的值。输出的$g(x)$是微分了的阶跃边界和过滤处理了的噪声之和。

2) 用一种决策机制将滤波器输出明显高于由噪声引起的区域隔离出来。

3) 用一种机制识别在隔离区域中的$g(x)$导数的过零点，并指定其为边缘位置。

高斯低通滤波器作用之后可以（在亚像素分辨率级）精确地找到（二阶）导数的过零点，从而找到边缘的精确位置，但这仅当边缘是直的情况才有效[5.38]。如果边缘是弯曲的，则会引入误差。比如，梯度方向的二阶导数（SDGD）和拉普拉斯算子在估计边缘位置时都将产生误差，但有趣的是，它们的方向是相反的，受其启发Verbeek和van Vliet[5.38]提出了一种算子，这种算子是两者的组合。

为了更好地找到实际边缘的位置，在这节中引用和描述的所有方法都是在边缘法线方向做信号处理[5.18]。Taratorin和Sideman[5.35]提出了一种利用已知的图像属性，如它是正的和有限支持的，来估计导数。Iverson和Zucker[5.15]通过加入逻辑/布尔推理来改进Canny方法的结果。这比起对导数取简单的阈值来提高了边缘检测的效果，但其结果没有动态轮廓模型(见第9章)和优化方法（第6章）好。此外，关于将信号处理技术应用于边缘检测的[5.36]文章还有很多。

98

通过查阅关于生物图像系统的文献可以发现，从20世纪60年代Hubel和Wiesel所做的先驱性工作以来，所有的方法都认为，生物系统分析图像是采用局部度量来量化方向、尺度和运动信息的。根据这种观点，假定我们想问这样的问题:“在这点的θ方向是否有一条边？”那么我们该如何构造一个核算子，使之在那个方向对边缘有很强的敏感性呢？一种直接的方法[5.37]就是构造两个高斯一阶导数核算子的加权和G_x和G_y，采用如下的加权形式:

$$G_\theta = G_x \cos\theta + G_y \sin\theta \tag{5-38}$$

不幸的是，除非采用相当大的核算子，否则采用这种方法得到的核算子在方向选择性方面更差。如果我们希望在尺度上进行微分，问题将变得更糟糕，因为为了最小化计算时间，尺度空间表示经常做粗略的计算。为此，Perona[5.31]提出了一种方法来解决这些问题。

5A.3 从边缘点推理线段

在选择了最好的算子来估计导数，选择了最佳的阈值，以及对边缘位置做了最佳估计之后，我们只得到了一组点（像素），其中一部分被认为可能是边缘的一部分而加了标记。如果

这些点是毗邻的，就可以从一个点"走"到另一个点，最后环绕出一个区域。有一些表达方法，如"链码"，可以使这个过程变得简单。然而，点是不可能像我们所希望的那种方式连接的，一些点会因为模糊、噪声或部分遮挡而缺失。有许多方法可以解决这个问题，包括松弛标号和参数变换，这两种方法都将在本书中讨论。另外，有一些组合方法，譬如Deng和Iyengar[5.11]的工作就综合了松弛法和贝叶斯方法，还有一些其他方法[5.29]，这里因为篇幅有限就不展开讨论了。

5A.4 空域频率表示

小波非常重要，但是完全探讨这个领域超出了本书的范围，因此这里我们仅提出一个相当基本的描述，并提供了一些参考文献，例如Castleman[4.6]中就有一章关于小波的，值得读一下。

5A.4.1 使用小波的原因

观察图5-17中的图像，显然，在图像不同的地方，空域频率是不一样的。傅里叶变换没有同时获得表示频率和位置这种直觉需要的功能。图像的傅里叶变换是一个二维矩阵，表示整个图像在每个空域频率的能量值。显然，因为傅里叶变换是可逆的，它得到了所有频率和空间的信息，但是没有明显的办法来回答这个问题：在每个位置，局部的空域频率是什么？

99

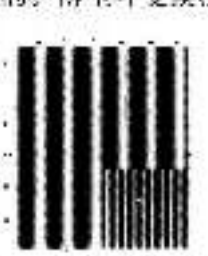


图5-17 空域频率急剧变化的一幅图像

小波方法在表示上增加自由度。虽然傅里叶变换是完备且可逆的，在原理上只需要利用一个简单的二维矩阵来刻画图像；但是，根据5.8节描述的空间/频率的思想，我们采用一个三维（或更高维）的数据结构取而代之。在这个意义上，空间/频率的表示是冗余的（过完备），因此需要比傅里叶变换大得多的存储空间。

5A.4.2 基小波和小波变换

我们定义一个基小波[⊖]$\psi(x,y)$为满足某种准则的任意一个含两个空间变量x和y的函数。但这里我们不关注这些准则，我们基本上是设计一个关于原点对称且几乎为有限支持的函数。"几乎有限支持"，是指函数值随着自变量偏离中心而迅速就减为0（通过允许的法则定义的一种方法）。一个一维基小波（如图5-18所示），即

$$\psi(x)=\frac{2}{\sqrt{3}\sqrt[4]{\pi}}(1-x^2)\exp\left(-\frac{x^2}{2}\right) \tag{5-39}$$

从这个小波，通过平移和尺度缩放可以产生出一组相似函数的集合。也就是说，通过如下方式定义ψ的一组平移和尺度变化族（一维）：

$$\psi_{a,b}(x)=\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \tag{5-40}$$

通过对函数f与一组可能的a、b值的小波计算内积，可以得到f的小波变换：

$$W_f(a,b)=\int_{-\infty}^{\infty}f(x)\psi_{a,b}(x)\mathrm{d}x \tag{5-41}$$

⊖ 这小波经常称母小波。

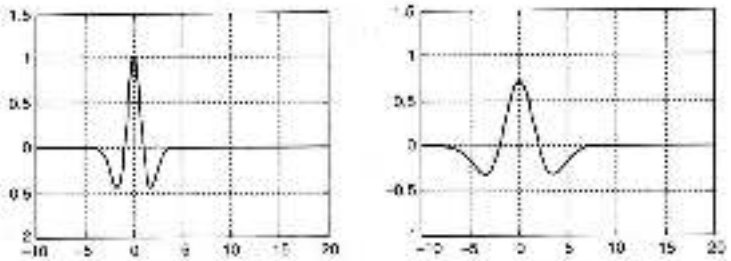


图5-18 一个基小波，以及一个尺度变化所产生的小波

可以看出这个变换是尺度和平移的函数。这个原理在二维上也同样成立，此时内积为：

$$W_f(a,b_x,b_y)=\int_{-\infty}^{\infty}f(x,y)\psi_{a,b_x,b_y}(x,y)\mathrm{d}x\mathrm{d}y \tag{5-42}$$

图5-19描述了在a取不同值时的截面图。显然，这个处理过程产生了一个尺度空间的表示。

图5-19 图像原图以及原图与三个不同的二维小波内积的结果(从小波变换截取的三个层面)

Lee[5.23]调查了神经生理学的资料，并推导出了如下形式的母小波：

$$\psi(x,y)=\frac{1}{\sqrt{2\pi}}\exp\left(-\left(\frac{4x^2+y^2}{8}\right)\right)\left[\exp(i\kappa x)-\exp\left(-\frac{\kappa^2}{2}\right)\right] \tag{5-43}$$

其中κ是一个常数，它的值依赖于带宽的假设，但近似等于3.5。对“母小波”进行缩放和平移可以产生一组滤波器（Frei-Chen基集使用同样的方法），它们表示一个图像的邻域在特征上的相似程度，并完全可以从中重构出图像。

5A.5 术语

Canny边缘检测器	Canny edge detector
非最大值抑制	Nonmaximum suppression
小波	Wavelet

作业5.A1

在高尺度层上，只有_____物体可以看见。（填空）

100
101

作业5.A2

如下表达式的估计值是什么？$(f\otimes h_1)^2+(f\otimes h_2)^2=E$，其中核算子$h_i$定义为：

0.05	0.08	0.05
−0.136	−0.225	−0.136
0.05	0.08	0.05

通过转置得到h_2。

选择最佳答案:

a) 关于x的一阶导数（x为水平方向）

b) 关于x的二阶导数

c) 拉普拉斯算子

d) 关于y的二阶导数

e) 二次变分

作业5.A3

采用高斯微分的思想来推导一个核算子。一维高斯（零均值）需要有什么样的变量，使得其一阶导数在$x=\pm 1$处有极值?

作业5.A4

与拉普拉斯算子相比，二次变分的优点是什么?

作业5.A5

令$E=(f-Hg)^{\mathrm{T}}(f-Hg)$。利用线性算子和矩阵形式的等价性，采用核算子的形式写出E的表达式。

作业5.A6

这个作业的目的是让你复习一遍构造图像金字塔的过程，这样你就可以对这种数据结构潜在的用处有更全面的了解，并且可将其应用到编码和传输领域。沿着这个方向，你可以获得关于图像编码一般领域一些额外的理解。编码并不是本书学习的主要目的，但在你的研究生涯中，你肯定会经常碰到做编码的人。拥有一些对最基本概念的大致了解，将被证明是非常有用的。

102

1) 下载图像“asterix512.ifs”，并验证它是否真的是512 × 512的（利用它本身的信息）。如果它不是512 × 512的，首先写个程序将它填充到那么大。

2) 编写一个程序，输入为二维的$n \times n$的图像，输出图像数据格式一样，而大小为$n/2 \times n/2$。程序取名为“ShrinkByTwo”，调用方式为:

```
ShrinkByTwo inimg outimg
```

该程序不能简单地对像素做间隔采样。输出图像的每个像素应该为其在输入图像相应的四个像素的平均值。注意，输出图像要求与输入图像具有相同的数据结构，这利用*ifscreate*很容易实现，只需将它第一个参数设成定义输入图像类型的字符串即可。例如: ifscreate (in->ifsdt, len, IFS_CR_ALL, 0)。用你的程序新建asterix256、asterix128、asterix64和

asterix32，不用处理低于32 × 32的图像。上交你的程序和打印的图像。

3) 写一个缩放图像的子程序，它应该有如下的调用格式:

```
ZoomByTwo(inimg, outimg)
IFSIMG inimg, outimg;
```

调用的程序应该可以对图像进行新建、读取等操作。子程序用一个缩放了的inimg对outimg进行简单的填充，你可以采用任何算法对丢失了的像素进行填充。(我们推荐将输入像素的平均来替代那些丢失了的像素。)

在处理之前，我们需要考虑一个“金字塔编码器”。当你运行ShrinkByTwo时，输出的图像集是asterix512的金字塔表示形式。对于一个金字塔编码器，目的是采用金字塔表示形式在信道上传输尽可能少的信息。主导思想是这样的: 首先，传输所有的asterix32，然后在传输端和接受端同时运行ZoomByTwo来产生一个asterix32的缩放图像，如下: 103

```
ZoomByTwo(a32, a64prime)
```

当我们从asterix64中创建出asterix32时，丢掉了一部分信息，这是很难进行恢复的，因此a64prime与asterix64是不一样的。但如果ZoomByTwo（在图像编码的文献中称为预测器）非常好，那么a64prime和asterix64的差别就很小（也就是值很小，但它仍然是64 × 64的)。为此，计算出a64prime和asterix64的差分图diff64。如果预测得非常完美，那么差分结果图像就应该是值都是0的64 × 64的图像，而这可以以一种灵活的方法进行编码（例如变长码)，并以非常低的比特数进行传输。现在让我们将diff64传输给接收端。通过将diff64叠加到由接收端产生的a64prime，我们就能纠正由预测器所产生的误差，就可以在接收端得到一个正确的asterix64，但我们只传输了diff64。我们很聪明，是吧？现在，用同样的方法，可以从asterix64中通过传输diff128得到asterix128，等等。

4) 新建上面所述的图像: diff64、diff128、diff256和diff512，计算传输每一幅图像所需要近似的比特（bit）数。为了做这些计算: 求diff64的标准差，接着取底为2的对数，这就是图像编码每个像素所需要的平均比特数。假设你要直接传输asterix512，那就需要512 × 512 × 8个比特数（假定图像是8比特的——你最好验证一下)。现在，将这个结果与你的金字塔编码器相比较，这里的金字塔编码器是通过将传输的每幅差分图像所有每个像素的比特数相加而得到的。你的编码器是不是运行得很好？在报告中对这个问题进行展开讨论。

参考文献

[5.1] V. Anh, J. Shi, and H. Tsai, “Scaling Theorems for Zero Crossings of Bandlimited Signals,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3). 1996.

[5.2] J. Babaud, A. Witkin, M. Baudin, and R. Duda, “Uniqueness of the Gaussian Kernel for Scale-space Filtering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(1), 1986.

[5.3] G. Bilbro and W. Snyder, “Optimization of Functions with Many Minima,” *IEEE Transactions on Systems, Man, and Cybernetics*, **21**(4), July/August, 1991. 104

[5.4] K. Boyer and S. Sarkar, “On the Localization Performance Measure and Optimal Edge Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[5.5] P. Burt and E. Adelson, “The Laplacian Pyramid as a Compact Image Code,” *Computer*

Vision, Graphics, and Image Processing, **16**, pp. 20–51, 1981.

[5.6] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(6), 1986.

[5.7] W. Chojnacki, M. Brooks, A. van der Hengel, and D. Gawley, "On the Fitting of Surfaces to Data with Covariances," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11), 2000.

[5.8] J. Crowley, *A Representation for Visual Information*, Ph.D. Thesis, Carnegie-Mellon University, 1981.

[5.9] J. Daugman, "Two-dimensional Spectral Analysis of Cortical Receptive Fields," *Vision Research*, **20**, pp. 847–856, 1980.

[5.10] J. Daugman, "Uncertainly Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-dimensional Visual Cortical Filters," *Journal of the Optical Society of America*, **2**(7), 1985.

[5.11] W. Deng and S. Iyengar, "A New Probabilistic Scheme and Its Application to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[5.12] W. Frei and C. Chen, "Fast Boundary Detection: A Generalization and a New Algorithm," *IEEE Transactions on Computers*, **26**(2), 1977.

[5.13] D. Hubel and T. Wiesel, "Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex," *Journal of Physiology (London)*, **160**, pp. 106–154, 1962.

[5.14] D. Hubel and T. Wiesel, 'Functional Architecture of Macaque Monkey Visual Cortex," *Proceedings of the Royal Society of London, B*, **198**, pp. 1–59, 1977.

[5.15] L. Iverson and S. Zucker, "Logical/Linear Operators for Image Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(10), 1995.

[5.16] P. Jackway and M. Deriche, "Scale-space Properties of the Multiscale Morphological Dilation–Erosion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(1), 1996.

[5.17] J. Jones and L. Palmer, "An Evaluation of the Two-dimensional Gabor Filter Model of Simple Receptive Fields in the Cat Striate Cortex," *Journal of Neurophysiology*, **58**, pp. 1233–1258, 1987.

[5.18] E. Joseph and T. Pavlidis, "Bar Code Waveform Recognition using Peak Locations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), 1994.

[5.19] M. Kelly, in *Machine Intelligence*, volume 6, University of Edinburgh Press, 1971.

[5.20] M. Kisworo, S. Venkatesh, and G. West, "Modeling Edges at Subpixel Accuracy using the Local Energy Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[5.21] A. Klinger, "Pattern and Search Statistics," in *Optimizing Methods in Statistics*, New York, Academic Press, 1971.

[5.22] P. Kube and P. Perona, "Scale-space Properties of Quadratic Feature Detectors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(10), 1996.

[5.23] T. Lee, "Image Representation Using 2-D Gabor Wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(10), 1996.

[5.24] Y. Leung, J. Zhang, and Z. Xu, "Clustering by Scale-space Filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(12), 2000.

[5.25] T. Lindeberg, "Scale-space for Discrete Signals," *IEEE Transactions on Pattern Anal-*

ysis and Machine Intelligence, **12**(3), 1990.

[5.26] T. Lindeberg, "Scale-space Theory, A Basic Tool for Analysing Structures at Different Scales," *Journal of Applied Statistics*, **21**(2), 1994.

[5.27] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proceedings of the Royal Society of London, B*, **207**, pp. 187–217, 1980.

[5.28] D. Marr and T. Poggio, "A Computational Theory of Human Stereo Vision," *Proceedings of the Royal Society of London, B*, **204**, pp. 301–328, 1979.

[5.29] R. Nelson, "Finding Line Segments by Stick Growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[5.30] E. Pauwels, L. Van Gool, P. Fiddelaers, and T. Moons, "An Extended Class of Scale-invariant and Recursive Scale Space Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(7), 1995.

[5.31] P. Perona, "Deformable Kernels for Early Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), 1995.

[5.32] P. Perona and J. Malik, "Scale-space and Edge Detection using Anisotropic Diffusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(7), pp. 629–639, 1990.

[5.33] W. Pratt, *Digital Image Processing*, Chichester, John Wiley and Sons, 1978.

[5.34] H. Tagare and R. deFigueiredo, "Reply to 'On the Localization Performance Measure and Optimal Edge Detection'," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[5.35] A. Taratorin and S. Sideman, "Constrained Regularized Differentiation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[5.36] F. van der Heijden, "Edge and Line Feature Extraction Based on Covariance Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[5.37] M. Van Horn, W. Snyder, and D. Herrington, "A Radial Filtering Scheme Applied to Intracoronary Ultrasound Images," *Computers in Cardiology*, September, 1993.

[5.38] P. Verbeek and L. van Vliet, "On the Location Error of Curved Edges in Low-pass Filtered 2-D and 3-D Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

[5.39] I. Weiss, "High-order Differentiation Filters that Work," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

[5.40] M. Werman and Z. Geyzel, "Fitting a Second Degree Curve in the Presence of Error," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 1995.

[5.41] R. Young, "The Gaussian Derivative Model for Spatial Vision: I. Retinal Mechanisms," *Spatial Vision*, **2**, pp. 273–293, 1987.

105 ~ 106

第6章 图像松弛：复原与特征抽取

To change, and to change for the better are two different things

德国谚语

本章讨论的技术旨在去除噪声与抑制退化，以便为图像分割抽取出更加清晰的特征。本章描述并比较了后验的图像复原技术与迭代式图像特征抽取技术。图像复原方法从退化图像中重构出原始图像[6.3]，而图像特征抽取方法则是从噪声图像中抽取特征，例如边缘。然而分析表明二者执行的是同样的基本运算：图像松弛。在本章“专题”中，对两种图像特征抽取方法，逐渐非凸（graduated nonconvexity, GNC）和可变传导率扩散（variable conductance diffusion, VCD），以及一种被称为均值场退火（mean field annealing, MFA）的复原/特征抽取方法进行了比较。它们之间的等价性体现了能量最小化方法与空间分析方法之间的关系，也体现了它们各自的温度参数与尺度参数之间的关系。本章的结尾讨论了图像特征抽取方法可遵循的通用原则。

6.1 松弛

“松弛”一词最初用来描述一组迭代数值方法，用于对联立非线性方程组求解（参见[6.18]）。Rosenfeld和Kak[6.64]将这一词的含义延伸为一系列的迭代分类方法，依据的是二者之间的相似性。在这里，我们给出“松弛”一词的通用定义，其中既包括了以上这些方法，也包括了一些更新的技术，后者同时也是本章所讨论的重点内容。

定义

一个松弛过程是一个多步算法，它具备以下性质：1）每一步的输出量与输入量具有相同的形式，以便算法可迭代运行；2）算法收敛至一个有界的结果。一些研究者还提出了第三个要求：对任一元素（在这里，是任一像素）的操作仅与该元素的某个有限“邻域”内的像素的状态相关，这里的“邻域”是明确定义的。依据这些标准，本章讨论的所有算法都是松弛过程。

6.2 复原

在一个图像复原问题中，假设一幅理想图像f，经退化后生成观测图像g。图像退化的一般模型是一个畸变算子与随机噪声相迭加，记畸变算子为D，则

$$g = D(f) + n \tag{6-1}$$

其中$g=[g_1, \cdots, g_N]^T$是对图像g的列向量描述，g_i表示第i个像素。对f与n的定义与之相似。那么，复原问题就是已知观测图像g，已知关于畸变（一般称作“模糊”）的某种知识和关于噪声的统计量，对f进行最佳估计的问题。

复原经常被看作是一个逆问题。即，从输入到输出是一个过程（在这里是模糊）。而我们

只能度量到输出量，并希望能够从中推断出输入量。

逆问题与病态

一个问题$g=D(f)$，如果

- 对任一f，都存在一个解g。
- 解g是唯一的。
- f连续变化时，解g随之连续变化。

则$g=D(f)$被认为是“可解的”。

一个问题，如果不具备以上条件，则被认为是“病态的”。正常情况下，病态是由问题的病态条件引起的。一个数学问题的条件由输出量对输入量变化的敏感度来度量。对一个良态条件问题而言，输入量的微小变化对输出量的影响不大；而对一个病态条件问题，输入量的很小变化也会引起输出量发生巨大的变化。

条件数（condition number）是对一个问题的条件的度量。条件数的一般定义如式（6-2）。条件数越大，则问题越是病态条件的（ill-conditioned）：

$$\text{条件数} \approx \frac{\text{输出量的变化}}{\text{输入量的变化}} \tag{6-2}$$

一个线性系统$Ax=b$的条件数由矩阵A的条件数决定。A的条件数K的定义如下：

$$K = \|A\| \|A^{-1}\| \tag{6-3}$$

108 其中$\|.\|$代表2次范式。K的取值范围是$[1,\infty)$。当$K>>1$时，这个线性系统是病态条件的。

在式（6-1）中，假设已知一个模糊过程。看起来似乎可以消除它所产生的影响。下面来看看为何情况可能并非如此。

举个例子，设有一个非常简单的图像，几乎是能够列举出来的最简单的图像——一幅3×3的图像，并用字母a，…，i为每一个像素命名。现在，假设这个图像经历了一个线性模糊过程，每个像素值都替换成其邻域内所有像素的值的平均（假设采用了4邻域定义）。在边缘或拐角处的像素的邻域像素比较少。如果新的模糊图像的值为ga，…，gi，则它们与原始值之间关系可以用表6-1所示的线性方程组来表示。

表6-1 得到的线性方程组

$ga=$		$0.5b+$		$0.5d$					
$gb=$	$0.33a+$		$0.33c+$		$0.33e$				
$gc=$		$0.5b+$				$0.5f$			
$gd=$	$0.33a+$				$0.33e+$		$0.33g$		
$ge=$		$0.25b+$		$0.25d+$		$0.25f+$		$0.25h$	
$gf=$			$0.33c+$		$0.33e+$				$0.33i$
$gg=$				$0.5d+$				$0.5h$	
$gh=$					$0.33e+$		$0.33g+$		$0.33i$
$gi=$						$0.5f+$		$0.5h$	

记作 $G=[ga\ gb\ gc\ gd\ ge\ gf\ gg\ gh\ gi]^{\mathrm{T}}$, $F=[a\ b\ c\ d\ e\ f\ g\ h\ i]^{\mathrm{T}}$.

a	b	c
d	e	f
g	h	i

$$H=\begin{bmatrix} 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0.33 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0.33 & 0 & 0 & 0 & 0.33 & 0 & 0.33 & 0 & 0 \\ 0 & 0.25 & 0 & 0.25 & 0 & 0.25 & 0 & 0.25 & 0 \\ 0 & 0 & 0.33 & 0 & 0.33 & 0 & 0 & 0 & 0.33 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$

于是，可以将模糊过程表示为$G=HF$，并通过$F=H^{-1}G$解出模糊前的值。看起来这似乎很简单。如果有畸变过程的模型（在这里，这个模型就是矩阵H），则只需对它求逆，然后再作乘法。现在来看看为何这样做可能会有麻烦。首先，数值计算H的逆。但是矩阵求逆程序告诉我们这个矩阵是奇异的，可能无法运行。那么，是我们选择的模型不好吗？ 109

事实证明，设计出一个非奇异的数值例子，是很困难的。当然，不是不可能，但的确是困难的。关键就在于：即使畸变矩阵是非奇异的，这个问题仍然可能是“病态条件的”。

我们观测到ga，…，gi，并采用矩阵乘法来确定a，…，i。如果H非奇异，那么可以非常完美地求解。然而，如工程师们所知的，噪声事实上总是存在的，因此实际上观测到的不是ga，而是$ga+\varepsilon$，其中ε是由于噪声而产生的扰动。如果这个系统（即畸变矩阵）是病态条件的（而它确实是），则ga的微小变化可能会导致a，…，i的估计值的巨大偏差。因此，即使系统是线性的，简单的矩阵求逆法亦不可行。

另一个或许更为简单的病态条件的例子[6.36]如下：考虑一个由模糊A描述的线性系统，以及一幅未知的图像f和观测图像g，它们满足：

$$g=Af$$

$$A=\begin{bmatrix}1 & 1\\ 1 & 1.01\end{bmatrix}\quad f=\begin{bmatrix}f_1\\ f_2\end{bmatrix}\quad g=\begin{bmatrix}1\\ 1\end{bmatrix} \tag{6-4}$$

矩阵A的条件数是402.0075，比1要大得多。这个系统的解为$f_1=1, f_2=0$。现在，假设观测图像g被噪声污染，$g=[1\quad 1.01]^{\mathrm{T}}$。则其解为$f_1=0, f_2=1$。观测数据的极小变化却使得结果发生了巨大的变化。

我们可以采用很多种方式来着手解决这类病态的复原问题。这些方法源自共同的体系：正则化理论（regularization theory）。一般而言，任何正则化方法都试图去分析一个与原始病态问题具有相近解的相关的可解问题 [6.57]。

最容易想到的方法就是估计出一个使均方误差的期望最小的图像。即找到一个未知图像f，使得如下表达式的值最小：

$$E=\sum_i (g_i-(f_i\otimes h))^2 \tag{6-5}$$

这里对图像中的所有像素求和，且应用一个与模糊h相应的核算子来表示畸变。由于这个

问题仍然是病态条件的，因此简单地最小化E并不可行。如果对噪声作一些假设，可能会令问题有一点儿改善。如果畸变是线性移不变系统，且噪声是静态的，则维恩纳滤波器（Wiener filter）可以给出这一判据下的最优解(见[6.28] 中的详细论述)。

110

6.3 最大后验方法

本节将介绍一些数学知识，以便进行更深入的讨论。

6.3.1 贝叶斯准则

贝叶斯（Bayes）准则同三个概率密度函数有关系：先验概率密度$p(f)$、条件概率密度$p(g|f)$、后验条件概率密度$p(f|g)$。

定义$p(f)$为某个特定图像f出现的*先验概率密度*。(如果将亮度值看作是连续的，则需要采用概率密度，而不是概率。无论采用哪一种用法都对下面的推导结果没有影响。) 即，在任何测量发生之前，图像f 出现的概率。举一个有关离散的先验概率的例子。假设有一个工厂生产法兰盘和垫圈，所生产的法兰盘个数是垫圈的9倍。法兰盘和垫圈落到传送带上的时间是随机的。但由于我们具备先验的知识，即车间生产的法兰盘数是垫圈个数的9倍，因此我们知道如果在某个任意时刻察看传送带，看到法兰盘的可能性总是比看到垫圈的可能性要大得多。因此，摄像头看到法兰盘的先验概率为0.9，而看到垫圈的先验概率为0.1。

定义$p(g|f)$是图像f 经过某种污染而得到观测图像g的*条件概率密度*。概率密度函数可以用几种可能的方式描述。一种是简单地列出变量的每个可能值发生的次数。这时，所列的表的长度表示发生的次数。这样的列表被称作是变量的*直方图*（histogram)。直方图经归一化后，就可以非常有效地表达概率密度函数。不幸的是，对图像而言，很难将其概率密度函数用一个直方图来表达。一个概率密度函数也可以采用某种解析函数(例如，高斯函数)，以参数化的方式来描述。

定义$p(f|g)$为观测图像g果真是图像f退化后所得图像的后验条件概率（密度)。$p(f|g)$正是我们要计算的值，它可以作为决策准则，或者更准确地说，作为*判别函数*。接下来将解释我们的决策准则。

已知从某个未知图像所得到的观测图像g，对每一个可能的f_i 值都计算$p(f_i|g)$。那么，如果f_i 对所有的$i \neq j$，都满足$p(f_i|g)$大于$p(f_j|g)$，则判定f_i 就是所求的未知图像。当我们基于$p(f_i|g)$的结果做出分类决策，则采用的正是**最大后验图像处理算法**。

111

通过贝叶斯准则将所定义的三个概率函数联系起来，得：

$$p(f|g)=\frac{p(g|f)p(f)}{\text{某数}} \tag{6-6}$$

$$\text{某数}=p(g) \tag{6-7}$$

式（6-6）中用“某数”代表这个条件概率密度的分母。用“某数”一词是为了引起大家对这样一个事实的注意：这个数代表了g发生的概率，与未退化的原始图像无关。由于这个数与f 无关，且对所有可能的f都一样，因此与判别最有可能属于哪一类无关。实际上，它是一个归一化常数，它保证了$p(f|g)$具备一个概率值必需的特性，即$p(f|g)$值介于0和1之间，且对所有可能图像，其总和为1（即，观测图像至少应属于这里所考虑的类之一)。

已知观测图像g，期望找到未知的原始图像f，使得后验概率密度$p(f|g)$最大。把整个图像的

概率与每个像素的概率联系起来。采用贝叶斯准则，可得以下比例关系：

$$p(f_i \mid g_i) \propto p(g_i \mid f_i) p(f_i) \tag{6-8}$$

其中i是像素的索引，$f=[f_1, f_2, \cdots, f_N]^{\mathrm{T}}$。

为了便于讨论，可以忽略这个比例式中的归一化常数，而最大化式（6-8）中的右式。

现在，假设除了噪声之外，图像无其他畸变，并假设相邻像素上的噪声具有统计上的独立性，则可写出：

$$p(g \mid f) = \prod_i p(g_i \mid f_i) \tag{6-9}$$

由于观测图像g_i与未知的原始图像像素f_i之间的唯一差异即是噪声，因此如果假定一个高斯噪声模型，可以用噪声密度来替代每个像素的条件密度，从而得：

$$p(g \mid f) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n_i^2}{2\sigma^2}\right) \tag{6-10}$$

将乘积运算移入指数运算中，可得[6.8，6.27，6.37，6.38，6.80]：

$$p(g \mid f) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left(\frac{-\sum_i (f_i - g_i)^2}{2\sigma^2}\right) \tag{6-11}$$

112

许多研究者都认为，一幅图像可以用一个适当的二或三维的马尔科夫场来表达[6.5，6.33，6.47，6.56]，则一个已知像素值发生的先验概率可用Gibbs分布来表示[6.6，6.27，6.45]。

$$p(f_i) \propto \exp\left(\frac{-\sum_{j \in \aleph_i} V_{ij}}{T}\right) \tag{6-12}$$

上式中求和运算是在像素i的邻域$\aleph_i$上计算的。回顾一下第4章，一个像素集合A关于像素集合B的流域（aura）是B中所有与A中点相邻像素的集合，其中对“相邻”的概念，由与特定问题相关的定义决定。这里的概念是类似的，但只是考虑了单个像素的邻域，而不是一个集合的邻域。正如流域的定义，邻域的定义也可以是与特定问题相关的，在这个意义上说图像中两个像素是相邻的，并不意味着二者必需是相邻的（adjacent）或者甚至是“紧贴的”。但事实上，在所有实际应用中，一个特定像素的邻域指的就是相邻的那些像素。T是一个可调整的宽度参数，而V是势函数（potential function），它通常是关于邻域内像素的函数。

再次用一个乘积来描述整个图像的先验概率：

$$p(f) = \prod_i p(f_i) \tag{6-13}$$

将式（6-11）和（6-12）代入式（6-8）中，并消除常数项[⊖]，取自然对数，改变正负号，从而使最大化概率问题转变成最小化一个目标函数的问题。

$$H(f, g) = \left(\sum_i \frac{(f_i - g_i)^2}{2\sigma^2}\right) + \sum_i \sum_{j \in \aleph_i} V_{ij} \tag{6-14}$$

⊖ 这些项并不影响最小值点的位置。式（6-14）中保留了σ，是为了用权重表示噪声项和先验项的相对重要性。

可以称式（6-14）中的第一项为“噪声项”[6.27]，而称第二项为“先验项”，可给出以下形式：

$$H(f,g)=H_n(f,g)+H_p(f) \tag{6-15}$$

113 观察“噪声”这一项，发现它在$f=g$时最小，这意味着复原图像f应与观测图像g相像。

可以选择不同的先验（正则化）项来突出期望图像具备的某种性质，例如，亮度值限定在给定的若干值 [6.70]；或者在边缘之外的区域，亮度平稳变化[6.9，6.16]；或者最常见的，亮度在局部区域内是常量而在边界处不连续[6.16，6.38]。

6.3.2 题外话：逆问题中的问题

目前为止，我们一直在描述一个逆问题。即，一个畸变过程已知（目前为止仅包括了噪声，但也可以包括模糊），结果图像也已知，它是对某个未知图像进行畸变运算后得到的。此时对未知图像进行估计的问题就是逆问题。这类问题的解决方法通常是选取并最小化一个函数（如式（6-14）），这个函数是两项之和，分别取决于观测数据与未知图像。正如本书所应用的，图像可以用像素点的集合来表示，而这些像素点的值是需要通过求解获得的；图像也可以用一个函数表示（同[6.82]比较）。如果f是一个函数，至少在概念上是一个函数，则f在采样点之间亦可赋值。这种情况下，可以把求解f的过程看作是插值（interpolation）。

6.3.3 用于边缘保留型平滑的目标函数

在6.3.1节，我们简略地讨论了先验项，而选择什么样的先验项取决于我们期望未知图像具有什么样的性质。最常用的是，图像的亮度在局部区域内是常量而在边界不连续[6.16，6.38]。为了最小化公式（6-14）描述的目标函数，用一个惩罚函数表示势函数V。

令Λ代表对像素i周围区域内亮度变化的某种数值测量。我们要做的是构造一个惩罚函数，它遵循所谓的Besag建议 [6.16，6.38]“允许图像亮度平稳地变化”。Besag建议，V_i 应该随着Λ_i的绝对值“严格递增”，且如果允许 “偶然的突变”，V_i 应“迅速达到一个最大值”。下面来探讨一下他为何提出了这样的建议。

采用二次形 $V_i=b\Lambda_i^2$ （其中$b>0$）是非常有吸引力的，因为它简单。它以非线性方式来惩罚噪声。也就是说，Λ_i是对像素i处梯度之强度的某种度量，所以梯度值越大，V_i 值越大。最小化过程选择的解必然是V_i 值很小的图像。因此，最有可能找到的是不发生突变的图像；而没有突变就意味着没有尖锐的边缘。这就意味着存在许多应用中都无法接受的模糊边缘。它仅适用于“平稳变化的真实景象”[6.6]。保持尖锐边缘的关键很简单：惩罚噪声引起的亮度变化——亮度变化越大，惩罚越大——但在某些点，不再进行惩罚（这样边缘上就没有附加的惩罚），如图6-1所示。问题在于（见6A.1节），这个函数是不可微的，因此不能采用梯度下降方法。

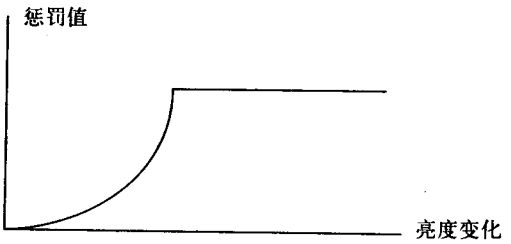


图6-1 噪声越大，惩罚应越强。假设局部亮度变化是由噪声引起的，则亮度变化越大惩罚越大。但是，局部亮度变化也可能是边缘引起的，则无需惩罚（否则，边缘会模糊）。因此令惩罚函数有一个上限

114

相反，如果选择的是倒置的高斯函数（inverted Gaussian），则有：

$$H_p(f) = -\frac{b}{\sqrt{2\pi}\tau}\left(\sum_i \exp\left(-\frac{\Lambda_i^2}{2\tau^2}\right)\right) \tag{6-16}$$

式（6-16）中的常量是无关紧要的。写上它们仅仅是为了使之看起来更像一个高斯函数。τ是一个软阈值，它表达的是关于表面粗糙度的先验知识。Λ用这种形式表示使得这个知识表达得更清楚。我们希望随着算法的推进，空间导数Λ_i变得几乎在处处都很小。关于这一概念，下一节将进行更加详细的介绍。

联立式（6-15）与（6-16），可以得到一个目标函数，如果最小化这个目标函数，则可获得一个与给定数据（在均方差意义上）相像的复原图像，同时图像的每个区域内亮度均匀，区域之间由陡峭边缘隔离。可以采用MFA（均值场退火）方法来最小化这个目标函数。

6.4 均值场退火

MFA是对复杂函数求最小值的一种技术，这类复杂函数一般具有多个最小值。统计力学中，均值场近似法可对一个粒子集合的能量状态进行连续表达。同时，MFA又近似于一种被称作“模拟退火”（SA）的随机算法。已证实SA即使对非凸问题[6.27]，亦可在概率上收敛至全局最优点。由于SA收敛需要的时间很长，令人无法接受，因而引出了许多加速技术[6.43]。MFA就是其中的一种。

自1989年引入MFA技术后[6.8，6.11]，此方法在很多领域都得到了应用，包括局部均匀(locally homogeneous)图像[6.37，6.38]和局部平滑图像[6.8，6.9]的复原、图像分割[6.69，6.70]、运动分析[6.1]以及传感器融合[6.7]等等。MFA方法结合了模拟退火[6.48]，以及统计力学中均值场近似[6.17]这两个概念。MFA的早期工作沿用了模拟退火的思路，并和统计力学相类比来进行验证。然而，随后[6.13]证实了这样的类比并不是必需的——尽管它富有洞察力且完全正确，事实上MFA完全可以从信息论的角度推导出来。 115

“MFA”这个词的用法可能会令人困惑，因为这个缩写最初是指推导一种算法的技术。但是这个词已逐渐演变成代表一类图像复原算法。这些算法都是将图像复原问题看作一个最小化问题，并采用一个特定的方法（即均值场近似方法）来求解这个最小化。所得到的求解方法结合了梯度下降法和“退火”方法。在整个算法过程中，被称作“温度”的控制参数会呈现一个单调性的变化过程。

另一个用于去除图像噪声的常用方法，称为“逐渐非凸”（graduated non-convexity，GNC）[6.16]，它具有许多同MFA相类似的地方，GNC也采用一个下降方法但少了一个控制参数。它们之间非常相像，事实上GNC可以从MFA推导出来[6.12]。

许多与MFA相关的工作都采用了MFA方法来推导类似的复原算法。不过，MFA也可以看作是一种图像特征抽取方法，并与另一种被称为“图像扩散（image diffusion）”的松弛方法等价。

对MFA的形式化推导感兴趣的读者可以参考上文中提到的文献。这里，我们对同样的概念，给出一种在教学上更具吸引力的解释，即指出MFA是*延拓方法*（continuation method）[6.2]的一种特殊形式。

延拓方法是一种实现同伦（homotopy）的算法，同伦即从一个（超）曲面到另一个（超）曲面的连续变形过程。在式（6-14）所描述的问题中，采用MFA将一个凸N-曲面扭曲为一个(通常)具有多个局部最小值的曲面。为了便于解释，考虑一下式（6-15）的最简单的一种可能

情况，它尽管简单但仍可以表示像素之间的相互作用:

$$H(f_1,f_2)=(f_1-g_1)^2+(f_2-g_2)^2-\frac{1}{\tau}\exp\left(-\frac{(f_1-f_2)^2}{\tau^2}\right) \tag{6-17}$$

这样得到了一幅（并不十分有趣的）仅由两个像素点f_1和f_2构成的图像。这两个像素经噪声污染后，可以得到观测量g_1和g_2。这里选择的先验项，会倾向于满足$f_1=f_2$的解。将MFA方法应用到式（6-17）所示函数中，其主要结果是用$\tau+T$来替换τ，其中T是一个初始值“较大的”的参数（在文献中被称作“温度”），T值会逐步减小到零。对式（6-17）完成这些替换后，可以得到一个新的目标函数:

116

$$H_T(f_1,f_2)=(f_1-g_1)^2+(f_2-g_2)^2-\frac{1}{\tau+T}\exp\left(-\frac{(f_1-f_2)^2}{(\tau+T)^2}\right) \tag{6-18}$$

将MFA替换后的结果记为H_T（也为了把它与T之间的相关性显示得更清楚）。式（6-18）中的H_T对向量f的两个元素求导，可得:

$$\frac{\partial}{\partial f}H_T(f_1,f_2)=\begin{bmatrix}2(f_1-g_1)+\frac{1}{\tau+T}\exp\left(-\frac{(f_1-f_2)^2}{(\tau+T)^2}\right)\left(\frac{2(f_1-f_2)}{(\tau+T)^2}\right)\\ 2(f_2-g_2)+\frac{1}{\tau+T}\exp\left(-\frac{(f_1-f_2)^2}{(\tau+T)^2}\right)\left(\frac{2(f_2-f_1)}{(\tau+T)^2}\right)\end{bmatrix} \tag{6-19}$$

对MFA方法的描述如下:

1）令$T=T_{\text{initial}}$（一个与问题相关的参数）。

2）使用式（6-19），执行梯度下降或其他的某种最小化技术，找到使得H_T最小的f。

3）减小T。

4）如果$T>T_{\text{final}}$，则跳至2）。

在这里，可应用的最简单的梯度下降方案是迭代计算:

$$\begin{bmatrix}f_1\\f_2\end{bmatrix}^{k+1}\Leftarrow\begin{bmatrix}f_1\\f_2\end{bmatrix}^{k}-\alpha\frac{\partial}{\partial f}H_T(f_1^k,f_2^k) \tag{6-20}$$

其中α是某个小的标量。为了了解整个过程是如何工作的，首先考虑T值很大的情况:

$$\min_f H_T(f_1,f_2)=\min_f((f_1-g_1)^2+(f_2-g_2)^2),\ 当\ (T\to\infty) \tag{6-21}$$

这是个凸函数（一个抛物面），并在$f_i=g_i$时最小。现在，为梯度下降设定一个初始起点。然后令T值稍微小一些，重复下降步骤，找到一个新解，然后再一次迭代，令T越来越小，直到T相比于τ可忽略。

即使简单化处理了这个问题，使图像仅包含两个像素，式（6-18）所描述的关于f_1、f_2的二维空间，仍然难以想像。为了捕捉到同伦问题的本质，把问题进一步简化至一维的情况。在图6-2中，举例说明了这样一个一维函数:

117

$$(f-k)^2-\frac{1}{\tau+T}\exp\left(-\frac{(f-l)^2}{(\tau+T)^2}\right)$$

其中k和l是标量。当T值较大时（$T+\tau=10$），曲线是完全凸的，随着T逐步减小，曲线逐步显示出其真实的形状。在每一次迭代中，都要跟踪最小值，并用箭头指示，箭头终止于全局最小值。

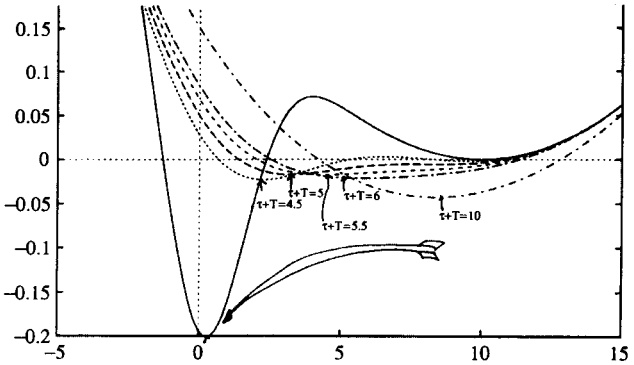


图6-2 一个函数的连续变形过程，用于寻找非凸函数的全局最小值，其初始函数是凸函数

6.4.1 选择先验项

回顾一下MFA能量函数中各项的含义。首先，考虑噪声项，并写出噪声项的通用表达式:

$$\sum_i ((D(f))_i - g_i)^2 \tag{6-22}$$

其中$(D(f))_i$表示图像f在像素i邻域内发生的某种畸变。所要寻找的图像f，要使这一项最小，则找到的结果图像f在畸变后与观测图像g（在误差平方和的意义上）最接近。现在，来看看先验项。

将先验项写成更通用的表达式:

$$-\frac{1}{\tau}\left[\sum_i \exp\left(-\left(\frac{(R(f))_i^2}{\tau^2}\right)\right)\right] \tag{6-23}$$

其中$(R(f))_i$这一项表示（未知的）图像f中像素i的某个函数，用τ替代公式（6-18）中的$\tau+T$。什么样的图像能够最小化这一项？让我们看看这种最小化又意味着什么。

首先，观察先验项最前面的负号。有了这个负号，最小化这个函数，就是找到使得指数最大的图像。那么什么样的图像会使指数取最大值？现在来看看指数的自变量。看到了负号，还看到两个平方项，因此它们总是正的。这样，指数的自变量始终是负的。那么，什么样的负自变量会使指数最大？答案是零（实际上是负零）。因此，为了使指数取值最大，应选择使得$R(f)$为零的图像f。 118

从这里能得到什么样的结论？对任意的函数$R(f)$，使得$R(f)$为零的f也就是先验项所寻找的f。这种分析给我们提供了许多的设计自由度。我们可以根据所期望的解的类型，选择函数$R(f)$。现在来看看一些范例。

范例

分段常量图像

考虑如下先验项:

$$R^2(f) = \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 \tag{6-24}$$

为了使这一项为零，两个偏微分均需为零。仅有一类曲面满足这一条件，它们在任一方

向上都不发生变化——也就是平坦的，但并非处处平坦。为何其解是分段常量，而不是常量？要知道需要求解最小值的函数是先验项和噪声项的和。先验项找到的是常量解，而噪声项寻找的却是忠实于观测量的解。因此，对这一问题求得的最优解是在每一分段上都是平坦的，图6-3给出了一维情况下的示意图。函数$R(f)$仅在f经历陡峭边缘点时取非零值。为了更清楚地看到会导出什么结果，考虑将问题扩展为连续函数。如果x是连续的，则式（6-23）中的求和变为求积分。积分的自变量仅在一些很少的有限的点不是零（被称作一个零测集），它们相比于积分的其余部分来说是微不足道的。

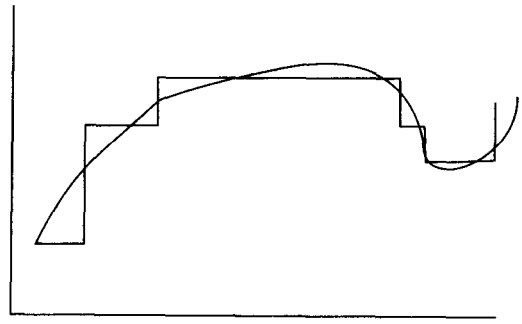


图6-3 用分段常量解对一个曲面的拟合情况。导数几乎处处为零，导数非零的点仅存在于阶跃处

119

范例

分段平面图像

现在来看看另一个例子。考虑:

$$R^2(f)=\left(\frac{\partial^2 f}{\partial x^2}\right)^2+\left(\frac{\partial^2 f}{\partial y^2}\right)^2+2\left(\frac{\partial^2 f}{\partial x\partial y}\right)^2 \tag{6-25}$$

这代表什么呢？什么样的函数，其二阶导数均为零？答案是：一个平面。因而，如果$R(f)$采用这样的表达式，就会得到一幅平面图像，同时为了保留对数据的保真度，可进一步得到一幅分段平面图像。另一个可替代算子是拉普拉斯算子，它同样也是基于二阶导数的:

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

以上两个公式我们在第5章中都提到过。

你也许会问你的指导老师，“将一个亮度图像分割为多个线性（即上文中的平面）的面片等同于假设实际曲面是个平面吗？”可能你会得到这样的答案：“是的，如果不考虑照明、反射系数，以及反射率的变化”。你会接着说，“但真实的曲面并非都是平面。”事实上答案是两方面的：首先，把所有的曲面都看作是平面而得到的观测，是没有意义的且无用的，你只需考虑一个足够小的区域即可。更准确地说，将图像分割为平面片是否有效取决于具体应用。对[6.14，6.74]的例子，可以通过分段常量分割方法来消除噪声并把每个面片看作是平面，从而获得改进的光流估计[6.41]或者双目估计[6.72]。有关对图像进行平面近似的一些基础理论可参考[6.62]。

有趣的是，Yi和Chelberg [6.83]的观测报告认为计算这样的二阶先验项比起一阶先验项，需要多得多的计算量，而且可能会使一阶先验项总是大致不变（在迭代中）。但是，我们在试验中并没有发现二阶先验会带来如此严重的计算亏损，相反，它们为重构过程提供了更多的灵活性。

对这样一个解，我们举一个一维的例子，见图6-4。为一个图像建立分段平面模型的思想，近年来从Elder及其同事的工作中[6.22，6.23，6.24]得到了更多的支持，他们提出“一幅图像的

边缘表示是可逆的，逆运算得到的图像能较好的近似原图像”。他们通过假设一幅图像，除了边缘之外的点均满足拉普拉斯方程 $\nabla^2 f(x,y)=0$，从而得到了这个著名的论断。

综上所述，你可以选择你所期望的任意函数作为指数的自变量，从而能够通过设置自变量为零而得到期望的图像。关于先验模型的一些更常规的性质，Li[6.51]中作了陈述，先验项作为一个局部的图像梯度Λ的函数，应满足1）一阶导数连续；2）是偶函数($h(\Lambda)=h(-\Lambda)$)；3）是正函数$h(\Lambda)>0$；4）自变量为正时$\Lambda>0$，其导数为负$h'(\Lambda)<0$；5）收敛至一个常量C，$\lim_{\Lambda\to\infty}|\Lambda h(\Lambda)|=C$。

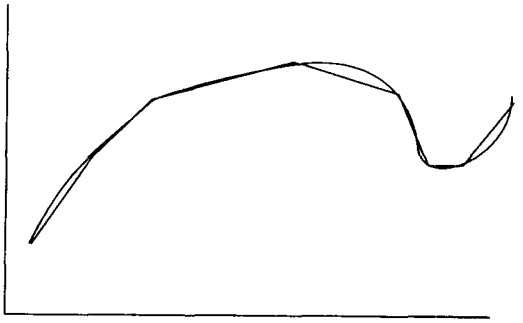


图6-4 对图6-3中数据的分段线性（即上文中的分段平面）解。显然，分段线性解相比分段常量解，能更精确地保留对数据的保真度

120

6.4.2 退火：避免局部最小值

MFA具备的另一种能力，是可避免陷入大多数局部最小值，这也使得它区别于其他MAP方法。

其特点完全归功于式（6-23）中的参数τ。以“大的”τ为起点。执行梯度下降。随着下降的迭代式推进，“缓慢”减小τ。直至τ达到一个“适当的”最小值时，终止算法。现在，让我们看看以上句子中所引用的这些词的含义是什么？

我们希望构造出这样一个函数，在相邻像素之间值相等的数量最大时，函数取值最小。对这个数量的计数方法就是delta函数，它在两个像素之间的差异为零时取值为1：

$$\delta(f_i-f_j)=\begin{cases}1 & 当\ f_i=f_j\\ 0 & 其他\end{cases} \tag{6-26}$$

因此相邻像素之间值相等的次数可表达为：

$$\sum_i\sum_{j\in\aleph_i}\delta(f_i-f_j) \tag{6-27}$$

我们需要这样一个函数，当上式中的二重求和（double summation）取值最大时函数值最小，最简单的方法就是引入一个负号：

$$H_{\mathrm{p}}=-\sum_i\sum_{j\in\aleph_i}\delta(f_i-f_j) \tag{6-28}$$

由于delta函数不可导，因此用起来并不方便，而我们需要用梯度下降来解决这个问题。关于这个表达式还有另一个问题：如果图像是连续赋值的（或甚至采用浮点表达方法），那么f_i等于f_j的含义是什么？它们之间应该有多接近才能认为是相等的？$|f_i-f_j|<0.01$如何？这足够小吗？那么0.001呢？可以了？那么你是接受了？这样看来，我们同意两点之间的差异超过0.001时会对误差计算有贡献。可是当两点间差异是0.000 999时又怎么样呢？此时只能认为这两点没有任何贡献。那么这样做有意义吗？

121

上文中引发的问题与一个度量是某个特定值的概率描述中遇到的问题非常类似。

例如，取值为整6.000 000（0的个数可任意）的概率只能为零，对此我们诉诸于一种不同的方法来表达似然概念，即采用概率密度。同样地，在这个问题上，我们遵循同样的原则。

不采用Kroneker delta函数，而是用一个连续的可微的函数来替代delta函数，而它们所表达的直觉知识是一样的。

$$H_p = -\sum_i \sum_{j \in N_i} \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(f_i - f_j)^2}{2\tau^2}\right) \tag{6-29}$$

退火的概念也适用于这个式子：以较大的τ为起点，逐步减小τ直至为零为止。这里常量的均方根并没有特别的含义，但它可用来保证这个函数以一种适当的方式保持其有界性。

关于为何退火过程能够避免局部最小值，可在其他地方[6.8，6.11，6.12]找到详细介绍，与模拟退火方法相比较的结果可参见[6.27]。

τ的初始值

我们为f选择的初始值就是g。思考一下当初始值τ_{initial}取值很大时会发生什么事情？一个很大的τ会导致指数的自变量接近于零，而指数本身的值则接近1。但这个值本身又被τ除，因此当τ很大时，先验值与$1/\tau$同一个数量级；因此如果τ是个大数，则先验项相对于噪声项而言则无关紧要了。我们可以通过选择一个较大的τ，比方说，比分子的平均值的两倍还要大，来保证τ初始值是“大的”。

$$\tau_{\text{initial}} = 2 < R(f) > \tag{6-30}$$

τ的递减

MFA方法以模拟退火方法的数学背景为基础。在模拟退火中，全局最小值可通过如下的一个对数退火进度来逼近：

$$\tau^K = \frac{1}{\ln K} \tag{6-31}$$

其中K是迭代步数。在这个进度下，τ递减的速度极端缓慢，缓慢到不切实际。相对地，可以选择一个如下的进度：

122

$$\tau^K = 0.99\tau^{K-1} \tag{6-32}$$

这一进度在许多应用中都得到了令人满意的效果，且τ递减的速度要比对数进度快得多。

6.4.3 如何对一个含有核算子的函数求导

在以上讨论中，假定噪声是造成畸变的唯一来源，而将模糊的影响排除在外。在这一节，对上文推导的结果进行扩展，在观测图像g的形成过程中将模糊包含进去，这样噪声项可写为式（6-33）。这一节里更重要的一个内容是，如何对一个含有核算子（或者说一个模糊算子）的函数求导：

$$\sum_i ((f \otimes h)_i - g_i)^2 \tag{6-33}$$

举一个一维的例子，假设f_i是原始（未知的）图像中的一个像素，g_i是观测图像中的一个像素，而h为5个像素构成的水平模糊核，如图6-5所示。现在，我们来详细解释在梯度下降中如何对式（6-33）的噪声项求导。首先，写出含有一个像素(f_4)的所有项，观测图像(g_4)是

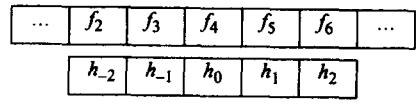


图6-5 一幅一维图像，以及由五个像素构成的一维核

在该像素f_4上引入噪声项H_4后所得的。

$$
\begin{aligned}
E_4 &= ((f\otimes h)_2 - g_2)^2 + ((f\otimes h)_3 - g_3)^2 + ((f\otimes h)_4 - g_4)^2 \\
&\quad + ((f\otimes h)_5 - g_5)^2 + ((f\otimes h)_6 - g_6)^2 \\
&= (f_0h_{-2} + f_1h_{-1} + f_2h_0 + f_3h_1 + f_4h_2 - g_2)^2 \\
&\quad + (f_1h_{-2} + f_2h_{-1} + f_3h_0 + f_4h_1 + f_5h_2 - g_3)^2 \\
&\quad + (f_2h_{-2} + f_3h_{-1} + f_4h_0 + f_5h_1 + f_6h_2 - g_4)^2 \\
&\quad + (f_3h_{-2} + f_4h_{-1} + f_5h_0 + f_6h_1 + f_7h_2 - g_5)^2 \\
&\quad + (f_4h_{-2} + f_5h_{-1} + f_6h_0 + f_7h_1 + f_8h_2 - g_6)^2
\end{aligned}
\tag{6-34}
$$

其中$(f\otimes h)_i$表示把核h应用到图像f中，h的原点（在这里是中心）落在像素f_i上。H_4关于像素f_4的导数的推导过程如式（6-35），进一步归纳可得式（6-36）：

$$
\begin{aligned}
\frac{\partial H_4}{\partial f_4} &= 2((f\otimes h)_2 - g_2)h_2 + 2((f\otimes h)_3 - g_3)h_1 + 2((f\otimes h)_4 - g_4)h_0 \\
&\quad + 2((f\otimes h)_5 - g_5)h_{-1} + 2((f\otimes h)_6 - g_6)h_{-2}
\end{aligned}
\tag{6-35}
$$

$$
\frac{\partial H_n}{\partial f_i} = ((f\otimes h) - g)\otimes h_{rev})_i \tag{6-36}
$$

123

其中$h_{rev} = h_{-2}, h_{-1}, h_0, h_1, h_2$。$((f\otimes h) - g)$对所有点都进行了计算。有关$h_{rev}$的用法，图6-6中给出了更清楚的图解，其中设$n_i = ((f\otimes h) - g)_i$。

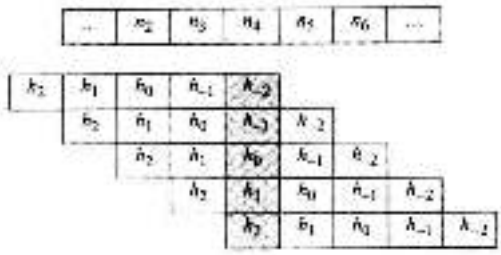


图6-6 噪声项导数中的逆序核

二维图像中逆序核的应用与一维情况相同。式（6-37）给出了一个3×3核函数（h）以及相应的逆序核h_{rev}：

$$
h = \begin{bmatrix} h_{-1,-1} & h_{-1,0} & h_{-1,1} \\ h_{0,-1} & h_{0,0} & h_{0,1} \\ h_{1,-1} & h_{1,0} & h_{1,1} \end{bmatrix} \qquad h_{rev} = \begin{bmatrix} h_{1,1} & h_{1,0} & h_{1,-1} \\ h_{0,1} & h_{0,0} & h_{0,-1} \\ h_{-1,1} & h_{-1,0} & h_{-1,-1} \end{bmatrix}
\tag{6-37}
$$

在讨论先验项之前，首先总结一下引入核算子后求导的通用形式，令$R(f\otimes h)$是某个可导函数，函数关于f的导数为：

$$
\frac{\partial}{\partial f} R(f\otimes h) = (R'(f\otimes h))\otimes h_{rev} \tag{6-38}
$$

其中$R'(t) = \frac{\partial}{\partial t} R(t)$。

在[6.63]中，以一种特别有趣的方式使用了含有核算子的函数的导数，其中每隔一个像素

才有一个度量。此时，问题不仅可以看作是图像估计，还可看作是插值问题[6.78]。有关基于MFA的缺失数据估计算法的详细实验结果以及性能评估可参见[6.63]。

除了噪声项，先验项也可以包含核算子。因为先验的能量函数对邻域运算进行建模，因此可以用一个核算子来表达，核的选择依赖于图像本身的性质。这里所举的例子中，选择了如下先验项:

$$H_{\mathrm{p}} = -\sum_i \exp(-(f \otimes r)^2) \tag{6-39}$$

$$\frac{\partial H_{\mathrm{p}}}{\partial f} = [2(f \otimes r)\exp(-(f \otimes r)^2)] \otimes r_{\mathrm{rev}} \tag{6-40}$$

实际上，先验导数$\partial H_{\mathrm{p}} / \partial f$本身就是一幅图像，这个图像是这样得到的：将$r$应用到$f$，然后与指数相乘(逐个像素)而产生另一幅图像，然后再把r的逆序应用到这幅图像中。

124

6.4.4 实际考虑：边缘保留型的平滑

这一章迄今为止讨论的所有的内容都可归于两个公式。现在，选择一个目标函数，令它保持保真度（复原图像与观测图像类似），且又是分段线性的:

$$H(f) = \sum_i \frac{(f_i - g_i)^2}{2\sigma^2} - \frac{\beta}{\tau}\sum_i \exp\left(-\frac{(R(f))_i^2}{\tau^2}\right) \tag{6-41}$$

其中$(R(f))_i$是二次变分式（6-25）在像素i上的值。当然，在特定应用中，可能会需要不同的先验项。为了完成梯度下降，必须求出关于f的导数。当发现指数的自变量的分子在x和y方向上都有变化时，问题变得更为复杂。这里有两种方案:

- 由于R是三项之和，且三项之和的指数等于三项指数的乘积，因此可利用导数的乘积规则构建一个更加复杂的表达式。
- 也可以这样说，“与其把总和放入指数的自变量中，不如把指数直接相加。”

当然，它们并不是等价的表达式。不过，最小化它们中任何一个都可以得到同样的结果：一幅分段线性的图像。由于第二个更容易实现，作为工程师，我们选择第二个选项。大家都知道如何求导，因此下面直接给出算法。

噪声项的导数是极小的。每一步迭代，只需通过$dnoise_i=(f_i - g_i)/\sigma^2$更新像素$i$的值。

确定先验项的导数还需要更多一点的工作：根据式（6-40），先验项的导数是:

$$\frac{\beta}{\tau}\left(\left(\frac{f \otimes \Delta}{\tau^2}\right)\exp\left(-\frac{(f \otimes \Delta)^2}{2\tau^2}\right)\right) \otimes \Delta_{\mathrm{rev}}$$

定义三个核算子，用以估计二次变分中的三个二阶偏导数:

$$\Delta_{xx} = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Delta_{yy} = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\Delta_{xy} = 2\begin{bmatrix} -0.25 & 0 & 0.25 \\ 0 & 0 & 0 \\ 0.25 & 0 & -0.25 \end{bmatrix}$$

125

注意这三个核是对称的，因此有$\Delta=\Delta_{\mathrm{rev}}$。

计算这三幅图像，它们的第i个像素值分别为:

$$r_{ixx} = (\Delta_{xx} \otimes f)_i, \quad r_{iyy} = (\Delta_{yy} \otimes f)_i, \quad r_{ixy} = (\Delta_{xy} \otimes f)_i$$

构造一个新的图像s_{xx}，其元素为

$$\frac{1}{\tau^3} r_{ixx} \exp\left(-\frac{r_{ixx}^2}{2\tau^2}\right)$$

同理构造图像s_{yy}和s_{xy}。

为保证梯度下降，计算像素i上由先验项引起的变化:

$$dprior_i = \beta((\Delta_{xx} \otimes s_{xx})_i + (\Delta_{yy} \otimes s_{yy})_i + (\Delta_{xy} \otimes s_{xy})_i)$$

梯度下降规则中指定，对f的每个元素都作更新，即$f_i \Leftarrow f_i - \alpha d_i$，其中$d_i = dnoise_i + dprior_i$。

学习系数（learning coefficient）α应为 $\alpha = \gamma\sigma\sqrt{\tau} / RMS(d_i)$，其中$\gamma$是一个很小的无量纲数，比如0.04；$RMS(d)$是梯度$d$的均方根范数；$\sigma$可定义为图像中的噪声变化（注意，在合成图像中这并不是一个好的定义）。这里的α在每一步迭代后都更新一次。

系数β与α在同一量级，一般来讲令$\beta=\sigma$就比较合适。

算法实现中，令退火过程经历τ的多个数量级，噪声的消除结果类似于图6-10至6-13所示结果。

6.5 结论

很容易看出，在图像优化的一些应用中存在着某种一致性。例如，本书作者之一在其撰写的一篇论文[6.7]中，利用这种一致性来实现融合，所融合的是关于同一物体的一幅（有噪声的）距离图像与一幅（有噪声的）亮度图像。由于二者均为同一物体的图像，距离图像（取决于几何学）和亮度图像（取决于反射系数，并因此也取决于几何学），因此二者必然是一致的。

在这一章中，普遍应用了优化方法，几乎可以把本章的标题改为“图像优化”了。本章通过建立一个目标函数——一个关于观测图像与（未知的）真实图像的函数，来寻找使得目标函数最小的（未知）图像。这里引入了两个词，一个噪声项——与度量相关，和一个先验项——仅与真实图像相关。然后，求解使目标函数达到最小的图像，从而找到“真实”图像。有多种最小化技术可以选择。在这一章，我们采用结合了退火的梯度下降法，当然也可以采用其他更复杂的和更快的技术，例如共轭梯度法。 126

6.6 术语

各向异性的扩散	anisotropic diffusion (见6A.2)
退火	annealing
贝叶斯准则	Bayes' rule
GNC方法	graduated nonconvexity(见6A.1)
逆问题	inverse problem
最大后验概率算法	MAP algorithm
松弛	relaxation
复原	restoration

作业6.1

式（6-34）举例说明了如何将核扩展为一个和式（sum），从而计算一个含有核的表达式的偏导数。采用这种方法，证明从式（6-39）可推导出式（6-40）。在证明中，请使用一个一维问题，以及一个3×1的核（记核的元素为h_{-1}、h_0和h_1）。

作业6.2

在图像angio.ifs上实现式（6-65），也可以使用指导老师指定的其他图像。请在各种不同运行时间和不同参数配置条件下，做多次实验。

作业6.3

在式（6-25）中，采用了二次变分来表达先验项。一个非常类似的先验项是拉普拉斯算子。二者有何差异？是否存在这样的图像特征，它使得拉普拉斯算子最小，而未能使二次变分最小？反之如何？

127

作业6.4

下列表达式中哪一个描述的是一个拉普拉斯算子？

(a) $\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}$ (c) $\left(\frac{\partial^2 f}{\partial x^2}\right)^2+\left(\frac{\partial^2}{\partial y^2}\right)^2$ (e) $\sqrt{\left(\frac{\partial^2 f}{\partial x^2}\right)^2+\left(\frac{\partial^2}{\partial y^2}\right)^2}$

(b) $\frac{\partial f}{\partial x}+\frac{\partial f}{\partial y}$ (d) $\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial x}\right)^2$ (f) $\sqrt{\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial x}\right)^2}$

作业6.5

扩散方程的一种表达式是 $df/dt = h_x \otimes (c(h_x \otimes f)) + h_y \otimes (c(h_y \otimes f))$，其中$h_x$和$h_y$分别为$x$和$y$方向上的一阶导数的估计值。这意味着必须要使用四个核来计算这个结果。不过，简单的代数运算就可以把它改写为 $df/dt = c(h_{xx} \otimes f + h_{yy} \otimes f)$，这样只需要两个核。对算法进行这样的简化是正确的吗？如果不正确，请解释为什么不正确，以及在什么情况下这样做是正确的？

作业6.6

考虑如下的图像哈密尔顿（Hamiltonian）

$$H(f)=\left(\sum_i\left(\frac{f_i-g_i}{\sigma^2}\right)^2\right)-\sum_i\frac{1}{\tau}\exp\left(-\frac{(h\otimes f)^2}{\tau^2}\right)\equiv H_n(f)+H_p(f)$$

其中⊗代表使用了一个核算子，图像中的像素按字典顺序标记，索引下标用i来代表，核h为：

$$\begin{bmatrix}-1 & 2 & -1\\ -2 & 4 & -2\\ -1 & 2 & -1\end{bmatrix}$$

令$G_p(f_k)$代表H_p关于像素k的偏导数，即$G_p(f_k)=\partial/\partial f_k H_p(f)$。写出$G_p(f_k)$的表达式。请采用核符号。

作业6.7

延续作业6.6中的问题，仅需考虑先验项。写出一个方程，来描述在简单梯度下降算法的一步迭代中，图像中像素k上的亮度变化。记这个梯度为$G_p(f_k)$，请在答案中使用这个符号。 128

作业6.8

延续作业6.7中的问题，对所推导出的$G_p(f_k)$表达式进行替换，从而扩展这个微分方程（假设亮度只在x方向上发生变化）。这是一种扩散方程吗？请讨论。（提示：在使用核的地方用适当的导数来替代。）

作业6.9

在一个扩散问题中，所要扩散的是一个向量，而不是亮度。对一个扩散方程，用适当的向量来替代这些项，并写出新的微分方程。（提示：如果记这个向量为$[a, b]^T$,可能更容易进行推导。）

作业6.10

扩散的运行时间以某种方式与模糊相关。这就是为什么一些人将这类的扩散称作“尺度空间”的原因。请讨论这一词的用法。

专题6A 替代算法与等价算法

6A.1 GNC: 一种可去除噪声的替代算法

正如前文所描述的，MFA方法通过最小化一个目标函数从而找到一幅具有尖锐边缘的图像。逐渐非凸（GNC）方法与此相同，但它所采用的目标函数，明确地针对边缘的存在。

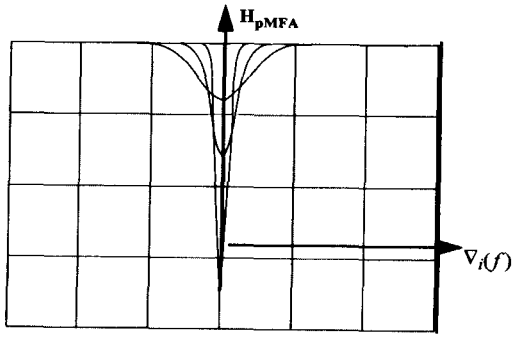


图6-7 MFA在不同T值下的先验能量。T值越小，峰越尖锐

我们考虑如下情况，由先验知识指定，在除了阶跃（step）不连续点之外的区域，图像亮度均匀。Blake和Zisserman[6.16]称这种情况为“弱隔膜”（weak membrane），而相类似的

MFA范例被称作[6.12]“分段均匀”。在目标函数(比较图6-7和图6-9)和复原图像(图6-10至图6-13)中均可以看出这两者之间的相似性。还有一些其他的表达式[6.12]，它们提出MFA问题的方式甚至与GNC更相似，[6.25]中首次提到这种相似性。GNC方法应用到“弱隔膜”中的最小化问题如下:

129

$$\min_{f,l} H_{\mathrm{GNC}} \tag{6-42}$$

其中

$$H_{\mathrm{GNC}} = H_n + S + P, \quad S = \lambda^2 \sum_i |\nabla_i(f)|^2 (1 - l_i) \quad P = \alpha \sum_i l_i \tag{6-43}$$

且符号 $\nabla_i(f)$ 可解释为“图像在点i上的梯度”。这里，$l_i \in \{0,1\}$ 表示f在第i个像素上的不连续性，如果$l_1=1$，则说明点i处的像素被认为是边缘点。同样地，f_i 表示第i个像素的亮度。[6.16]中已显示，最小化H_{GNC}可以简化为如下只含有连续变量的问题:

$$\min_f \left(H_n + \sum_i v(\nabla_i(f)) \right) \tag{6-44}$$

在式（6-43）与式（6-44）中，$|\nabla(\cdot)|$可代表任一算子，例如 $(\partial f/\partial x)^2 + (\partial f/\partial y)^2$，这类算子会返回一个对图像的局部“边缘”的标量度量。式（6-44）中的v函数是图6-8所示的“截顶抛物线”（clipped parabola）。

由于式（6-44）所定义的函数通常是非凸的，即函数可能具有多个最小值，因此无法通过梯度下降等技术实现最小化。而GNC方法采用如下分段平滑函数来逼近v:

130

$$v^*(t) = \begin{cases} \lambda^2 t^2 & (|t| < q) \\ \alpha - c^*(|t| - r)^2/2 & (q \leqslant |t| < r) \\ \alpha & (|t| \geqslant r) \end{cases} \tag{6-45}$$

其中c^*是一个标量常数，$c=c^*/p$，

$$r^2 = \alpha\left(\frac{2}{c^*} + \frac{1}{\lambda^2}\right), \quad q = \frac{\alpha}{\lambda^2 r} \tag{6-46}$$

式（6-45）与(6-46)已给出了这个算法。将参数p从1减至0，可平稳地改变v^*直至它精确等于v。这样生成的一个先验能量族，见图6-9所示。

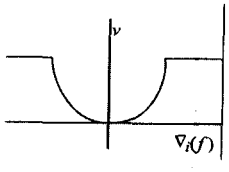


图6-8 GNC算法的先验能量

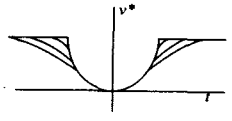


图6-9 对图6-8中能量的平滑近似。p越小，近似结果越接近理想的先验能量。这里（GNC）中的t等价于MFA中的边缘梯度大小∇

整个过程的第一步是最小化一个凸函数，这样会得到一个唯一的最小值。然后从这个最小值出发，随着p从1减至0，连续地跟踪局部最小值。

6A.2 传导率可变的扩散

VCD（variable conductance diffusion: 传导率可变的扩散）方法[6.31，6.59，6.62]是一种非常有效的图像特征抽取方法，它只允许模糊效应在边缘之外的区域发生。“边缘”一词可以宽泛地指任意的令人感兴趣的图像特征。例如Whitaker[6.79]对图像的梯度进行运算,而不是对图像本身，并且仅对脊（梯度方向尖锐变化的区域）之外的区域进行平滑。由于原始图像中的多数信息都丢失了，因此这样的运算肯定不能被复原。不过，对于从灰度图像中抽取出物体的中心轴而言，VCD方法是一个非常鲁棒的方法。

VCD的运算模拟以下扩散方程

$$\frac{\partial f_i}{\partial t} = \nabla \cdot (c_i \cdot \nabla_i f) \tag{6-47}$$

其中t是时间，$\nabla_i f$ 代表f 在像素i上的空间梯度。扩散方程对某种量（最常用的例子是热量）在一个传导率（如，热传导率）为c 的材料中的流动现象进行建模。

如果c_i 是常量，且与像素下标i无关，则偏微分方程（6-47）有一个解等价于与高斯函数作卷积运算，这个高斯函数的变量与c以及扩散的运行时间相关。特别地，令f 是一个关于空间和时间的函数，用一个特定的偏微分方程（PDE）来表达。如果能写出f 的如下表达式:

$$f(x,t) = \int_{-\infty}^{\infty} G(x,x',t) f(x',0) \mathrm{d}x' \tag{6-48}$$

131

那么，我们说 $G(x,x',t)$ 是PDE的格林函数（Green's function）。在各向同性的扩散这样的特殊情况下，可以对其一维的形式进行如下描述:

定理

高斯函数是以下PDE的格林函数:

$$\frac{\partial f}{\partial t} = c \frac{\partial^2 f}{\partial x^2} \tag{6-49}$$

证明

证明可以通过写出如下高斯函数来完成:

$$G(x,x',t) = \frac{1}{\sigma} \exp\left(-\frac{(x-x')^2}{2\sigma^2}\right)$$

其中σ可以证明是关于时间的函数（我们省略了 $1/\sqrt{2\pi}$ ，因为它在PDE的两边均有出现，可以相互抵消）。

将式（6-48）代入式（6-49）中，左式是一个积分关于t的偏导数，其中σ是关于t 的一个函数。求这个偏导数，使左式等于:

$$\left[\frac{(x-x')^2}{\sigma^4} - \frac{1}{\sigma^2}\right] \frac{\partial \sigma}{\partial t} \tag{6-50}$$

同样地，我们可以求关于x的二阶偏导数，得右式:

$$\frac{c}{\sigma}\left[\frac{(x-x')^2}{\sigma^4} - \frac{1}{\sigma^2}\right] \tag{6-51}$$

联立式（6-50）与（6-51），使二者相等可得等式:

$$\frac{\partial \sigma}{\partial t} = \frac{c}{\sigma} \tag{6-52}$$

其解为:

$$\sigma^2 = 2ct \tag{6-53}$$

证明完毕。

在使用VCD方法的情况下，传导率变为一个关于空间坐标的函数，此时用参数i表示空间坐标。特别要指明的是，它变成了图像局部亮度本身的一个特性。传导率c_i 被看作是空间的局部压缩因子。

为了只在边缘之外进行平滑，那么如果i是边缘像素（也就是说，如果所选择的图像特性具有局部非均匀性），则令c_i 很小。如果c_i 很小，（在类似的热传导情况中）只有很少的热流动（即空间被拉伸），（类似地）在图像中，就几乎没有平滑发生。另一方面，如果c_i 较大，则在像素i的附近会发生较大程度的平滑（即空间被压缩）。那么，正如我们已讨论的MFA和GNC的形式，VCD方法在经迭代后也产生一个几乎分段均匀的结果。

132

6A.3 面向边缘的各向异性扩散

如我们在式（6-48）中看到的，高斯函数是扩散方程的格林函数。即，在一幅图像上运行一个扩散过程会生成与高斯函数卷积相同的结果，其中高斯函数的变量取决于扩散运行时间的长短。常量—传导率扩散方程为:

$$f_t = c(f_{xx} + f_{yy}) \tag{6-54}$$

如果图像中有一条边缘，我们希望在边缘两侧都去除噪声，但又不使边缘模糊。此时，在与边缘相切的方向上做扩散是有意义的。假定一幅二维图像中一条边缘的法向量与切向量为:

$$N = \frac{[f_x f_y]^{\mathrm{T}}}{\sqrt{f_x^2 + f_y^2}} \qquad T = \frac{[-f_y, f_x]^{\mathrm{T}}}{\sqrt{f_x^2 + f_y^2}}$$

现在考虑，f在N和T方向上求二阶偏导数: f_{NN}和f_{TT}。

由于拉普拉斯算子是旋转不变量，我们可以在新的坐标系中写出扩散PDE（式（6-54）），新的公式为 $f_t = c(f_{NN} + f_{TT})$。

推导关于偏导数之间的关系如下:

$$\begin{aligned} f_{NN} &= (f_x^2 f_{xx} + 2 f_x f_y f_{xy} + f_y^2 f_{yy})/(f_x^2 + f_y^2) \\ f_{TT} &= (f_y^2 f_{xx} - 2 f_x f_y f_{xy} + f_x^2 f_{yy})/(f_x^2 + f_y^2) \end{aligned} \tag{6-55}$$

将上式代入式（6-54）中并减去法向流，最终能得到这样一个PDE，它只在沿着边缘的方向上进行平滑，而对边缘的垂直方向上则不进行平滑:

$$f_t = (f_y^2 f_{xx} - 2 f_x f_y f_{xy} + f_x^2 f_{yy})/(f_x^2 + f_y^2) \tag{6-56}$$

上式称为各向异性的扩散。

6A.4 对图像松弛算子的一个通用描述

6A.4.1 MFA和GNC

在这两种情况下，均有一个能量函数以递增的方式惩罚图像中梯度的发生。在采用GNC

的情况下，先验项保留了它的原始形状，而“退火”过程（即p的减小）可以实现对预定先验形状的连续的逐步逼近。在采用MFA的情况下，先验形状本身就会改变，但其会保留一个不变区域，当T减小时这个区域也会变得越来越窄。由此，分段常量的MFA与GNC的弱隔膜达到同样的结果并不奇怪。 133

这两类技术的等价性将在下一节进行论证。[6.12]中给出了形式化的证明，请读者参考文中的深入分析。接下来描述的试验在该论文中也有介绍，这里写出来是为了辅助读者理解算法的作用过程。

试验中两种算法复原同样的图像，且使用多种信噪比，将MFA和GNC各自应用到同一噪声图像中，并分别调整参数使其达到最优的图像复原结果。每一个算法都完成了不同参数值下的数百次运行，我们发现对每一幅噪声图像，两种算法都存在各自的某种参数组合，使得复原图像的品质相当。

图6-10至图6-13描绘出了复原图像达到的品质，对比了原始图像（图6-10），SNR=2时的退化图像（图6-11）、MFA方法复原图像（图6-12）以及GNC复原图像（图6-13）。

图6-10 原始图像

图6-11 退化图像

图6-12 MFA复原图像

图6-13 GNC复原图像

GNC算法的代码可在[6.16]中找到，其中梯度采用SOR(successive over-relaxation，连续的过松弛)来实现。我们发现噪声严重的情况下（SNR<3），同样采用SOR时，MFA的运行速度约比GNC的速度要快10倍。而对清晰一些的图像，如SNR>=4时，GNC的运行速度更快。

6A.4.2 MFA与VCD—等价算法

在完成如下比较之前，首先需要详尽地介绍空间导数的一个特性。一幅图像的亮度关于距离的导数可写为：

$$\frac{\partial f}{\partial x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x} \tag{6-57}$$

然而，在一个采样图像（比如，所有的数字图像）中，求极限的过程是没有任何意义的，因为正如Romeny所指出的[6.73]，没有人能在比一个像素更小的尺度上取不同的值。相反，为了估计一个点上的导数，必须对该点的某个邻域完成运算，如何估计导数是个很古老的论题，在这里我们不再更多的描述，但我们要指出，大多数分析结果已经推出如下的论断：这一类估计（也包含对高阶导数的估计），在一维情况下，通常通过一个核算子计算，而在n维情况下是通过一个有n个算子的阵列的欧几里德范数来计算。对这个求导过程，我们只考虑了x方向上的导数。这里将用几个段落的篇幅来概括描述这个过程。首先，将先验项改写为： 134

$$H_x(f)=\frac{-b}{\sqrt{2\pi}T}\sum_i \exp\left(-\frac{(f\otimes r)_i^2}{2T^2}\right) \tag{6-58}$$

符号$(f\otimes r)_i$代表将一个核r应用到图像f上后，在点i上得到的结果，所选择的核可用于强调与

问题相关的图像特性。这个通式已用于对分段常量[6.38]图像与分段线性图像[6.8，6.9]去除噪声。

在以下推导过程中，只考虑了先验项。

为了实现梯度下降，需要计算这个导数，因此有:

$$\frac{\partial H_{\mathrm{p}}}{\partial f_i}=\frac{b}{\sqrt{2\pi}T}\left[\left(\frac{(f\otimes r)}{T^2}\exp\left(-\frac{(f\otimes r)^2}{2T^2}\right)\right)\otimes r_{\mathrm{rev}}\right]_i \tag{6-59}$$

其中r_{rev}代表核r的镜像图像。

现在，假定像素i上的亮度变化量Λ_i，是图像梯度的大小，且几乎处处很小。众所周知[6.53]，采用高斯导数作为一个核是针对噪声图像估计其导数的极好的方式。记G_x为一个高斯函数关于x的导数，用$(G_x,G_y)^{\mathrm{T}}=\nabla G$替代公式（6-59）中的$r$；并将$f\otimes\nabla G$写成$\nabla f$，可得结果

$$\frac{\partial H_p}{\partial f}=-\kappa(\nabla((\nabla f)\exp(-(\nabla f)^2))) \tag{6-60}$$

在上式中，将常量放在一起用κ表达，并为了看得更清晰，设退火控制参数T为1。然后使用了如下关系式，即对于一阶导数核，有$f\otimes h=-(f\otimes h_{\mathrm{rev}})$。关于$T$所产生的影响将在下一节进行讨论。

最后，我们考虑在梯度下降算法中如何使用$\frac{\partial}{\partial f}H(f)$。在梯度下降的最简单的实现中，$f$是由以下式子（同式（6-20）作比较）更新的:

$$f_i^{k+1}=f_i^k-\alpha\frac{\partial H}{\partial f_i} \tag{6-61}$$

其中f^k表示在第k步循环后f的值，而α是某个小的常量（或者，在更复杂的算法中，是H的黑塞的一个函数）。改写式（6-61），得:

$$\frac{\partial H}{\partial f_i}=\frac{f_i^{k+1}-f_i^k}{\alpha} \tag{6-62}$$

我们注意到式（6-62）的左式代表的是f在第k步迭代与第$k+1$步迭代之间的变化，而实际上它与f的导数的形式具有很强的类同之处。如果定义第k步迭代在时刻t进行计算，而第$k+1$步迭代在$t+\Delta t$时刻进行计算，则可以使这个相似性显现出来。（在相类似的上下文中，t有时被看作“卷积参数”。）由于t是人工引入的参数，没有任何有意义的物理单位，可以用任一方便的比例常数来衡量，则可得

135

$$\frac{\partial H}{\partial f_i}=\frac{f_i(t+\Delta t)-f_i(t)}{\Delta t}=\frac{\partial f_i}{\partial t} \tag{6-63}$$

其中，已将常量α重新定义为Δt，使之看起来像一个关于时间的导数。将这个（重）定义代入到式（6-60）中，并简单地改变符号，又令两次迭代之间的时间间隔明确化，则MFA先验项的导数最终可重写为:

$$\frac{\partial f_i}{\partial t}=-\kappa(\nabla((\nabla f)\exp(-(\nabla_T f)^2))) \tag{6-64}$$

其中$\nabla_T f$代表梯度值的缩放比例为T。

写出扩散公式

$$\frac{\partial f_i}{\partial t} = -\kappa(\nabla(c\nabla f)) \tag{6-65}$$

其中，如果传导率c由指数项来替代，则可观察到式（6-64）精确等于（VCD）[6.31，6.59，6.62]中采用的扩散公式的形式。常量κ简单地将公式所包含的所有常量合并成一个量。

通过式（6-65），我们说明了MFA与VCD之间的等价性，其中MFA的运算中没有包含噪声项式（6-11）。这个等价性为关于图像分析的两大类思维形式的联合作了准备：第一类（最优化）方法考虑的是一幅图像应该具有的特性。因此它会构造一个优化问题，使其解为所期望的图像。也可将这种方法称之为*复原类*。第二类（过程）方法，则更关心如何确定可应用的适当的空间分析法（spatial analysis）。自适应滤波，扩散，模板匹配，等等，都更关心过程本身，而并不关心过程对图像的某种假定的“能量函数”做什么样的操作。这一节的结果显示出这两类方法不只是在原理上等价，至少在保留边缘的图像平滑问题这一特殊形式上，它们是*精确*相等的。

以上等价性中，对MFA方法的讨论仅仅考虑了目标函数的先验项。增加噪声项，可使一个图像特征抽取算法转换为一个受约束的复原算法。

Nordström [6.59]也观察到了扩散技术与正则化（优化）方法之间的相似性。他认为：“各向异性的扩散方法（Whitaker[6.79]称之为VCD）并不打算找到任何形式的最优解。”这并不十分准确。或许换一种说法，说“这个技术的开发者并没有打算把它当作一个最小化技术用”会更好一些。之后，Nordström又接着说：“可以将正则化方法与各向异性的扩散方法统一起来，尽管二者最初看起来非常不同。”接着，他非常完美而精确地定义了一个成本函数，其作用正是一个各向异性的扩散，与这里描述的推导方式相类似。Nordström又认为，有必要增加一个“稳定成本”以“限制图像函数可能估计值的空间”。而当读者得知稳定成本有如下形式时，就不会感到奇怪：

$$\sum_i (f_i - g_i)^2 \tag{6-66}$$

136

因为式（6-11）中已说明了它是关于高斯噪声对无模糊成像系统所产生影响的一个量测。因此，可以认为*有偏的各向异性的扩散*（biased anisotropic diffusion，BAD）[6.59]是对一幅图像的最大后验复原。根据这种观察结果，可以认为VCD/BAD的研究者们，在具备有关附加的噪声生成过程的信息时，可以考虑采用不同形式的稳定成本。

6A.5 与神经元网络之间的关系

Hopfield神经元网络与最优化问题之间的关系是众所周知的。给定一个Ising类型的目标，可直接找到一个递归神经元网络，使其稳定状态是目标函数的最小值（见[6.39]和[6.40]）。

这类递归网络的最简单的形式，是对二值变量进行操作，用一个神经元代表一个变量。在这种用法中，一个“神经元”就是一个乘积之和（sum-of-products）算子，可生成输入量的加权和。接着，对这个和再进行一次单调非线性化处理，通常是某种限幅函数，例如sigmoid函数。根据以上定义，可以将这些算子描述为一个二层网络，其表达式如式（6-59），每层都仅局部连通，如图6-14所示（也可见[6.13]）。

从实现的角度来看，局部连通性是很重要的，因为正是这一点使并行的、实时的硬件架构变得可行。

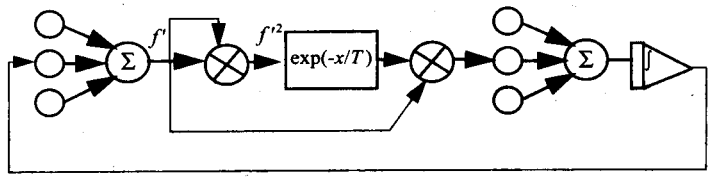


图6-14 设想用一个局部连通的递归网络，来完成这里所描述的图像松弛网络是直截了当的

6A.6 结论

图像的优化问题与扩散问题之间存在着等价性。其他研究者也作了类似的观测。另外，除了Nordström[6.59]之外，Geiger和Yuille也得出了一个相似的结论[6.26]，即能量函数要求明确的线过程（line processes）。由于事实已经证明[6.38]线过程并不是必需的。因此现在对他们的结论可以给出更一般的解释。对一个图像优化问题（尤其是复原问题），人们会通过定义一个准则函数，并应用某种最小化方案来搜索一个全局（或至少是一个好的）最小值。因而，一个图像复原问题可以看作是以下目标的结合：1）保留原始图像中的信息，从而使生成的结果图像与原始图像（或者是对原始图像应用某个算子得到的结果）在某个方面是相似的；2）生成的结果图像具有某些特性，例如，除边界之外的区域是局部平滑的。如果丢掉第一个目标，则复原问题就会变成图像特征（iconic feature）抽取问题。Wu 和 Doerschuk [6.81]针对这项工作进行了富有吸引力的扩展。

137

最后，回顾一下[6.19]中的内容，其中演示了MFA算子可以用一个二层的且局部连通的递归神经元网络来运算。从这篇文章出发，有人可能会推出这样的结论，认为GNC与VCD也同样可以用简单直接的神经元网络来实现。

根据以上结果，针对特征抽取算法的设计，我们推测出如下的指导原则。

- 松弛是个中心概念。一个松弛算法应具有以下特征：
  1）它必须是迭代式的。即，算法每个循环的输出量与输入量具有同样的格式，以便可将算法本身的输出量，再次应用到算法中。
  2）它必须收敛[⊖]。
- 松弛算法在本质上必须是局部的。即，在任一时刻，任一像素的变化仅与此像素的局部邻域相关。依从这个原则，可使全局性相互作用随着时间（迭代步数）和空间平稳地发生，并通过Gibbs/Markov场的等价性为分析算子提供一个理论基础。
- 扩散与最优化之间的等价性有利于理解这两种形式的算法的性能。对所设计的那些扩散算法而言，将它们看作是一个最优化松弛(optimizing relaxation)是很有帮助的：因为所有的松弛都会最小化某样东西，而观察扩散步骤（diffusion step）的积分（尽管这一过程经常是很难处理的），能够发现在一种特殊技术的开发/应用中究竟是哪个性质真正被最小化了。对于所设计的那些最优化方法，可以看作是一个依照局部不均匀程度的空间形变，且在其后紧随一个平均化过程。这种思路可以帮助理解最优化方法的时空效应。

⊖ 一些作者对松弛的定义还包括了第三个要求："局域性"。我们在另一个圆点标注的项中单独列出了这一点。

- 在所设计的特征抽取算法中，即便没有残差被明确地最小化，将算法看作是一个复原是也是有益的。人们更容易理解这一算法对图像产生的精确效果。
- 空间分析算法中的尺度变化，与退火算法中的温度控制是紧密相关的，二者的能力也是相关的。
- 最后，在以上提到的所有算法中，非线性运算（指数）是成功的绝对要素。Kolmogorov定理 [6.49]证明了充分性，他指出一个线性运算再加上后续的适当的非线性化，可以计算任一种映像（mappings）。我们主张这样的非线性化不仅是充分的而且是必要的。这个事实很可能对近来许多神经元网络的成功应用有着极其重要的贡献。

参考文献

[6.1] I. Abdelqader, S. Rajala, W. Snyder, and G. Bilbro, "Energy Minimization Approach to Motion Estimation using Mean Field Annealing," *Signal Processing*, July 1992.

[6.2] E. Allgower and K. Georg, *Numerical Continuation Methods*, Berlin, Springer-Verlag, 1990. 138

[6.3] H. Andrews and B. Hunt, *Digital Image Restoration*, Englewood Cliffs, NJ, Prentice-Hall, 1977.

[6.4] D. Baker and J. Aggarwal, "Geometry Guided Segmentation of Outdoor Scenes," *SPIE Applications of Artificial Intelligence*, **VI**, pp. 576–583, 1988.

[6.5] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society, B*, **36**, pp. 192–326, 1974.

[6.6] J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society, B*, **48** (3), 1986.

[6.7] G. Bilbro and W. Snyder, "Fusion of Range and Luminance Data," *IEEE Symposium on Intelligent Control*, Arlington, August, 1988.

[6.8] G. Bilbro and W. Snyder, "Range Image Restoration using Mean Field Annealing," In *Advances in Neural Network Information Processing Systems*, San Mateo, CA, Morgan-Kaufmann, 1989.

[6.9] G. Bilbro and W. Snyder, "Mean Field Annealing, an Application to Image Noise Removal," *Journal of Neural Network Computing*, Fall, 1990.

[6.10] G. Bilbro and W. Snyder, "Optimization of Functions with Many Minima," *IEEE Transactions on Systems, Man, and Cybernetics*, **21**(4), July/August, 1991.

[6.11] G. Bilbro, R. Mann, T. Miller, W. Snyder, D. Van den Bout and M. White, "Optimization by Mean Field Annealing," In *Advances in Neural Information Processing Systems*, San Mateo, CA, Morgan-Kauffman, 1989.

[6.12] G. Bilbro, W. Snyder, S. Garnier, and J. Gault, "Mean Field Annealing: a Formalism for Constructing GNC-like Algorithms," *IEEE Transactions on Neural Networks*, **3**(1) pp. 131–138, 1992.

[6.13] G. Bilbro, W. Snyder, and R. Mann, "Mean Field Approximation Minimizes Relative Entropy," *Journal of the Optical Society of America, A*, **8**(2), February 1991.

[6.14] M. Black and A. Jepson, "Estimating Optical Flow in Segmented Images Using Variable-order Parametric Models with Local Deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(10), 1996.

[6.15] A. Blake, "Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**(1), 1989.

[6.16] A. Blake and A. Zisserman, *Visual Reconstruction*, Cambridge, MA, MIT Press, 1987.

[6.17] E. Brezin, D. LeGuillon, and J. Zinn-Justin, "Field Theoretical Approaches to Critical Phenomena," *Phase Transitions and Critical Phenomena*, volume 6, eds. C. Domb and M. Green, New York, Academic Press, 1976.

[6.18] R. Burden, J. Faires, and A. Reynolds, *Numerical Analysis*, Boston, Prindle, 1981.

[6.19] H. Chang and M. Fitzpatrick, "Geometrical Image Transformation to Compensate for MRI Distortions," *SPIE Medical Imaging IV*, **1233**, pp. 116–127, February, 1990.

[6.20] H. Derin and H. Elliot, "Modeling and Segmentation of Noisy and Textured Images using Gibbs Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**, pp. 39–55, 1987.

[6.21] H. Ehricke, "Problems and Approaches for Tissue Segmentation in 3D-MR Imaging," *SPIE Medical Imaging IV: Image Processing*, **1233**, pp. 128–137, February, 1990.

[6.22] J. Elder, "Are Edges Incomplete?" *International Journal of Computer Vision*, **34**(2), 1999.

[6.23] J. Elder and R. Goldberg, "Image Editing in the Contour Domain," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3), 2001.

[6.24] J. Elder and S. Zucker, "Scale Space Localization, Blur, and Contour-based Image Coding," *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 1996.

[6.25] D. Geiger and F. Girosi, "Parallel and Deterministic Algorithms for MRFS: Surface Reconstruction and Integration," *AI Memo*, No 1114, Cambridge, MA, MIT, 1989.

[6.26] D. Geiger and A. Yuille, "A Common Framework for Image Segmentation by Energy Functions and Nonlinear Diffusion," *MIT AI Lab Report*, Cambridge, MA, 1989.

[6.27] D. Geman and S. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(6), November, 1984.

[6.28] R. Gonzalez and P. Wintz, *Digital Image Processing*, 2nd edn, Reading, MA, Addison-Wesley, 1987.

[6.29] A. Gray, J. Kay, and D. Titterington, "An Empirical Study of the Simulation of Various Models used for Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[6.30] B. Groshong, G. Bilbro, and W. Snyder, "Restoration of Eddy Current Images by Constrained Gradient Descent," *Journal of Nondestructive Evaluation*, December, 1991.

[6.31] S. Grossberg, "Neural Dynamics of Brightness Perception: Features, Boundaries, Diffusion, and Resonance," *Perception and Psychophysics*, **36**(5), pp. 428–456, 1984.

[6.32] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, New Haven, CT, Yale University Press, 1923.

[6.33] J. Hammersley and P. Clifford, "Markov Field on Finite Graphs and Lattices," unpublished.

[6.34] F. Hansen and H. Elliot, "Image Segmentation using Simple Markov Field Models," *Computer Graphics and Image Processing*, **20**, pp. 101–132, 1982.

[6.35] R. Haralick and G. Shapiro, "Image Segmentation Techniques," *Computer Vision, Graphics, and Image Processing*, **29**, pp. 100–132, 1985.

[6.36] E. Hensel, *Inverse Theory and Applications for Engineers*, Englewood Cliffs, NJ, Prentice-Hall, 1991.

[6.37] H. Hiriyannaiah, *Signal Reconstruction using Mean Field Annealing*. Ph.D. Thesis, North Carolina State University, Raleigh, NC, 1990.

[6.38] H. Hiriyannaiah, G. Bilbro, W. Snyder, and R. Mann, "Restoration of Locally Homogeneous Images using Mean Field Annealing," *Journal of the Optical Society of America A*, **6**, pp. 1901–1912, December, 1989.

[6.39] J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons," *Proceedings of the National Academy of Science* USA, **81**, pp. 3058–3092.

[6.40] J. Hopfield, and D. Tank, "Neural Computations of Decisions in Optimization Problems," *Biological Cybernetics*, **52**, pp. 141–152, 1985.

[6.41] M. Irani, B. Rousso, and S. Peleg, "Recovery of Ego-motion Using Region Alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), 1997.

[6.42] A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, New York, IEEE Press, 1988.

[6.43] S. Kapoor, P. Mundkur, and U. Desai, "Depth and Image Recovery using a MRF Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(11), 1994.

[6.44] I. Kapouleas and C. Kulikowski, "A Model-based System for Interpretation of MR Human Brain Scans," *Proceedings of the SPIE, Medical Imaging* II, vol. 914, February, 1988.

[6.45] R. Kashyap and R. Chellappa, "Estimation and Choice of Neighbors in Spatial-interaction Model of Images," *IEEE Transactions on Information Theory*, **29**, pp. 60–72, January, 1983.

[6.46] M. Kelly, In *Machine Intelligence*, vol 6, Edinburgh, University of Edinburgh Press, 1971.

[6.47] R. Kindermann and J. Snell, *Markov Random Fields and Their Applications*, Providence, RI, American Mathematical Society, 1980.

[6.48] S. Kirkpatrick, Gelatt C, and Vecchi M, "Optimization by Simulated Annealing," *Science*, **220**, pp. 671–668, 1983.

[6.49] A. Kolmogorov, "On the Representation of Continuous Functions of One Variable by Superposition of Continuous Functions of One Variable and Addition," *AMS Translation*, **2**, pp. 55–59, 1957.

[6.50] R. Lee, and R. Leahy, "Multi-spectral Tissue Classification of MR Images Using Sensor Fusion Approaches," *SPIE Medical Imaging IV: Image Processing*, **1233**, pp. 149–157, February, 1990.

[6.51] S. Li, "On Discontinuity-adaptive Smoothness Priors in Computer Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(6), 1995.

[6.52] R. Malik and T. Whangbo, "Angle Densities and Recognition of 3D Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), 1997.

[6.53] D. Marr, *Vision*, San Francisco, CA, Freeman, 1982.

[6.54] J. Marroquin, *Probabilistic Solution to Inverse Problems*, Doctoral Dissertation, MIT, 1985.

[6.55] P. Morris, in *Nuclear Magnetic Resonance Imaging in Medicine and Biology*, Oxford, Clarendon Press, 1986.

[6.56] J. Moussouris, "Gibbs and Markov Systems with Constraints," *Journal of Statistical*

Physics, (10), pp. 11–33, 1974.

[6.57] M. Nashed, "Aspects of Generalized Inverses in Analysis and Regularization," in *Generalized Inverses and Applications*, ed. by M. Nashed, New York, Academic Press, 1976.

[6.58] T. Nelson, "Propagation Characteristics of a Fractal Network: Applications to the His-Purkinje Conduction System," *SPIE Medical Imaging IV: Image Processing*, **1233**, pp. 23–32, February, 1990.

[6.59] N. Nordström, "Biased Anisotropic Diffusion – A Unified Regularization and Diffusion Approach to Edge Detection," *Image and Vision Computing*, **8**(4), pp. 318–327, 1990.

[6.60] T. Pavlidis, *Structural Pattern Recognition*, Berlin, Springer-Verlag, 1977.

[6.61] A. Pentland, "Interpolation using Wavelet Bases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[6.62] P. Perona and J. Malik, "Scale-space and Edge Detection using Anisotropic Diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, pp. 629–639, July, 1990.

[6.63] H. Qi, *A High-resolution, Large-area Digital Imaging System*, Ph.D. Thesis, North Carolina State University, 1999.

[6.64] A. Rosenfeld and A. Kak, *Digital Picture Processing*, 2nd edn, Vol. 2, New York, Academic Press, 1982.

[6.65] Samet, *The Design and Analysis of Spacial Data Structures*, Reading, MA, Addison-Wesley, 1989.

[6.66] P. Santago, K. Link, W. Snyder, J. Worley, S. Rajala, and Y. Han, "Restoration of Cardiac Magnetic Resonance Images," *Symposium on Computer Based Medical Systems*, Chapel Hill, NC, June 3–6, 1990.

[6.67] S. Shemlon and S. Dunn, "Rule-based Interpretation with Models of Expected Structure," *SPIE Medical Imaging IV*, **1233**, pp. 33–44, February, 1990.

[6.68] W. Snyder, G. Bilbro, A. Logenthiran, and S. Rajala, "Optimal Thresholding – A New Approach," *Pattern Recognition Letters*, **11**(12), December, 1990.

[6.69] W. Snyder, P. Santago, A. Logenthiran, K. Link, G. Bilbro, and S. Rajala, "Segmentation of Magnetic Resonance Images using Mean Field Annealing," *XII International Conference on Information Processing in Medical Imaging*, Kent, England, July 7–11, 1991.

[6.70] W. Snyder, A. Logenthiran, P. Santago, K. Link, G. Bilbro, and S. Rajala, "Segmentation of Magnetic Resonance Images using Mean Field Annealing," *Image and Vision Computing*, **10**(6), pp. 361–368, 1992.

[6.71] C. Soukoulis, K. Levin, and G. Grest, "Irreversibility and Metastability in Spin-glasses. I. Ising Model," *Physical Review B*, **28**(3), pp. 1495–1509, 1983.

[6.72] B. Super and W. Klarquist, "Patch-based Stereo in a General Binocular Viewing Geometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), 1997.

[6.73] B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, "Scale Space: Its Natural Operators and Differential Invariants," *XII International Conference on Information Processing in Medical Imaging*, Kent, England, July 7–11, 1991.

[6.74] P. Torr, R. Szeliski, and P. Anandan, "An Integrated Bayesian Approach to Layer Extraction from Image Sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3), 2001.

[6.75] J. van Laarhoven and E. Aarts, *Simulated Annealing: Theory and Applications*, Norwell, MA, Reidel, 1988.

[6.76] M. Vannier, *et al.*, "Multispectral Analysis of Magnetic Resonance Images," *Radiology*, **154**, pp. 221–224, January, 1985.

[6.77] D. Van den Bout and T. Miller, "Graph Partitioning using Annealed Neural Networks," *IEEE Transactions on Neural Networks*, **1**(2), pp. 192–203, 1990.

[6.78] C. Wang, W. Snyder, and G. Bilbro, "Optimal Interpolation of Images," *Neural Networks for Computing Conference*, Snowbird, UT, April, 1995.

[6.79] R. Whitaker, "Geometry-limited Diffusion in the Characterization of Geometric Patches in Images," TR91-039, Dept. of Computer Science, UNC, Chapel Hill, NC, 1991.

[6.80] G. Wolberg and T. Pavlidis, "Restoration of Binary Images Using Stochastic Relaxation With Annealing," *Pattern Recognition Letters*, **3**(6), pp. 375–388, December, 1985.

[6.81] C. Wu and P. Doerschuk, "Cluster Expansions for the Deterministic Computation of Bayesian Estimators Based on Markov Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(3), 1975.

[6.82] M. Yaou and W. Chang, "Fast Surface Interpolation Using Multiresolution Wavelet Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

[6.83] J. Yi and D. Chelberg, "Discontinuity-preserving and Viewpoint Invariant Reconstruction of Visible Surfaces Using a First-order Regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(6), 1995.

139 ~ 143

第7章　数学形态学

A man's discourse is like to a rich Persian carpet, the beautiful figures and patterns of which can be shown only by spreading and extending it out; when is is contracted and folder up, they are obscured and lost

希腊哲学家Plutarch

后缀“-ology”意指“××学”，因此很显然，“morphology”指的是关于形态的研究学科，要解答的是如下的重要问题：“它们为何只在夜晚出来，然后朝着明亮处飞行？”以及“为何微波灭虫器只能烤到那些无害的动物，却单单不能对付‘蚊子’？”还有——停住！这里说的是形态学（MORPH-ology），研究的是形态（SHAPE），而不是蛾子（moth）！再来一次……

7.1　二值形态学

考虑亮度值仅取二值的图像，并以此为起点，这很重要，要记住！仅取二值！首先讨论两个算子，接下来，在理解了它们的工作原理后，我们将说明如何将它们应用于二值图像。最后，作为对二值形态学的一种扩展，对灰度形态学运算及其相应算子进行描述。

7.1.1　膨胀

首先，给出直观定义：对一幅（二值）图像的膨胀结果是将该图像的所有前景区域都稍稍扩大之后的图像。

现在，给出形式化表示：考虑两幅图像，f_A和f_B，且令A和B为有序对集，分别由f_A和f_B中各自的前景像素坐标构成。

考虑f_A中的一个像素，以及B中与其相应的元素（有序对），称之为$b\in B$。通过将A中的每一个有序对与有序对b进行相加，可生成一个新的集合。让我们看一个小例子。

对图像$A=\{(2,8),(3,6),(4,4),(5,6),(6,4),(7,6),(8,8)\}$，如果与有序对$(-1,1)$相加，则可得集合$A_{(-1,1)}=\{(1,9),(2,7),(3,5),(4,7),(5,5),(6,7),(7,9)\}$。相应图像见图7-1，我们希望读者会注意到$A_{(-1,1)}$只是$A$的一个平移，对这个概念有了牢固的认识之后，再考虑一下如果建构一个集合，由A关于B中每个有序对的平移构成，会发生什么事情？记该集合为$\{A_b, b\in B\}$，即，b是B中的一个有序对。

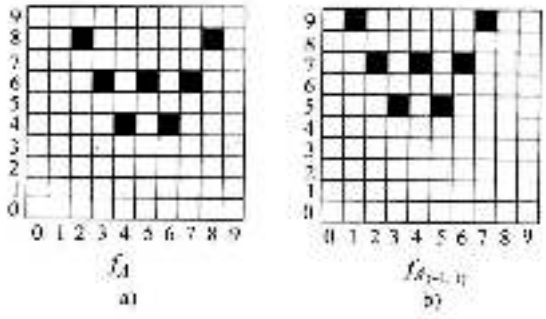
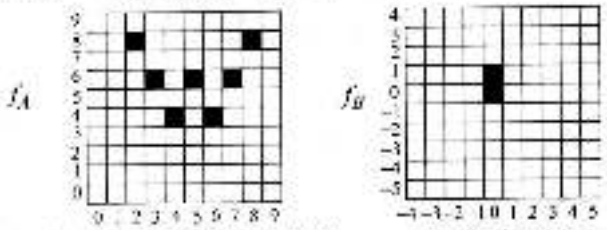


图7-1　膨胀例子a）初始二值图像 b）关于$B=\{(-1,1)\}$膨胀后的二值图像

形式上，定义A关于B的膨胀运算为 $A\oplus B=\{a+b\mid(a\in A,b\in B)\}$，该运算等同于对所有$A$的平移结果取并集，

$$A\oplus B=\bigcup_{b\in B}A_b \tag{7-1}$$

对图像的膨胀，采用了同一个标记符号：$f_A\oplus f_B$。这里给出另一个例子，



$A=\{(2,8),(3,6),(4,4),(5,6),(6,4),(7,6),(8,8)\}$　　$B=\{(0,0),(0,1)\}$

基于以上定义，

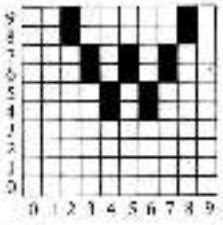
$$A\oplus B=A_{(0,0)}\cup A_{(0,1)} \tag{7-2}$$

145

$$=\{(2,8),(2,9),(3,6)(3,7),(4,4),(4,5),(5,6),(5,7),(6,4),(6,5),(7,6),(7,7),(8,8),(8,9)\} \tag{7-3}$$

且

$f_A\oplus f_B=$

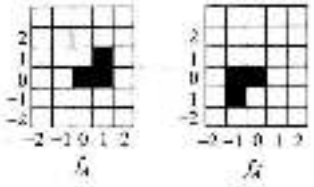


用$\#A$表示集合A的元素个数。在这个例子中，$\#A=7$，且$\#(A\oplus B)=14$。这种情况的发生有一定的巧合。这种巧合是因为$A(0,0)$和$A(0,1)$之间不重叠，换一种方式来说就是：$A_{(0,0)}\cap A_{(0,1)}=\varnothing$。

在更通常的情况下，则并非如此。通常

$$\#(A\oplus B)\leqslant\#A\cdot\#B \tag{7-4}$$

为了进一步讨论，需要定义一些符号：假设x是一个有序对，则1）一个集合A关于x的平移[⊖]记为A_x，2）记A的反对称集为 $\hat{A}=\{(-x,-y)\mid(x,y)\in A)\}$，3）记$A$的余集为$A_c$。关于反对称集的一个例子如下：



⊖ 这里无需局限于二维空间，只要x和A取自同一个空间，没是走点等，更一般地，如果A和B是同一个空间x内的集合，且$x\in x$，则A关于x的平移表示为$A_x=\{y\mid$对某个$a\in A,\ a=a+x\}$。

在这个例子中，$A=\{(0,0),(1,0),(1,1)\}$，而A的反对称集为$\{(0,0),(-1,0),(-1,-1)\}$。

7.1.2 腐蚀

现在，定义（在一定程度上可以看作是）膨胀的逆运算——腐蚀（Erosion）。

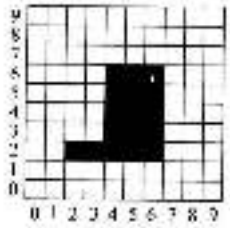
$$A\ominus B=\{a\mid(a+b)\in A \quad 对于所有\ (a\in A,b\in B)\} \tag{7-5}$$

它可写成如下用平移表示的式子

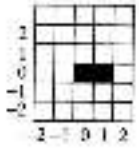
$$A\ominus B=\bigcap_{b\in B}A_b \tag{7-6}$$

146

需要注意两件事情：对第二个集合B，取反对称集；采用了交集符号，再看一个例子。



f_A

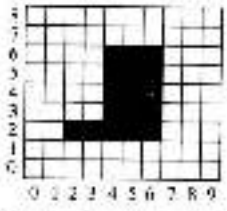


f_B

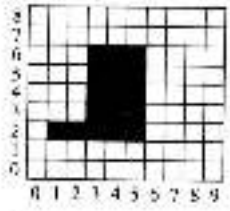
$B=\{(0,0),(1,0)\}$

$\tilde{B}=\{(0,0),(-1,0)\}$

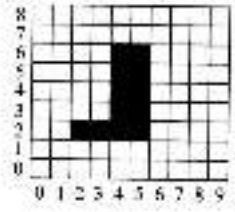
相比于列出所有17个元素的冗长的工作，只画出



$f_{A_{-1,0}}$



$f_{A_{-1,-1}}$



$f_{A_{-1,0}}\cap f_{A_{-1,-1}}$

至此，已定义了膨胀和腐蚀，读者会发现通常（对所有实际应用）其中的一幅图像比另一幅图像要“小”；即在上述例子中

$$\#A>>\#B \tag{7-7}$$

在这种情况下，称小一点的图像f_B为“结构元素”。

147

7.1.3 膨胀与腐蚀的性质

- 交换律：膨胀满足交换律，即可以交换两幅图像的次序而进行膨胀，不会改变结果

$$A \oplus B = B \oplus A \tag{7-8}$$

- 结合律：膨胀满足结合律，即可以对多幅图像之间进行的膨胀运算按任意次序组合，不改变结果

$$(A \oplus B) \oplus C = A \oplus (B \oplus C) \tag{7-9}$$

- 分配性：对于一幅图像与另一幅由两幅图像并集生成的图像进行膨胀，可以先膨胀后取并集。换句话说，可以把膨胀运算分布到括号内的所有项

$$A \oplus (B \cup C) = (A \oplus B) \cup (A \oplus C) \tag{7-10}$$

- 增长性：如果 $A \subseteq B$，则对任意的结构元素K，有

$$A \oplus K \subseteq B \oplus K \tag{7-11}$$

当满足以上性质时，我们说该算子是“增长的”。

一个例证：膨胀是增长的

令集合A由元素A_i组成，$A=\{A_1, A_2, \cdots, A_n\}$，以同样方式定义$B$。此外，假定 $B \subseteq A$。现在，假设A和B均关于同一个结构元素K膨胀。取K的一个元素，记为K_1，令A的每一个元素关于K_1膨胀，$A \oplus K_1 = \{A_1 + K_1, A_2 + K_1, \cdots, A_n + K_1\}$，并以同样方式膨胀$B$。由于$B$中的每个元素也是$A$中的元素，因此$B \oplus K_1$中每个元素也属于$A \oplus K_1$。由于对$K$中任意元素的观测结果均满足以上关系，因此所有$K$的元素也都满足。现在考虑对应用结构元素$K$中的两个元素对$A$膨胀后的结果取并集：$A_{12} = (A \oplus K_1) \cup (A \oplus K_2)$。由于 $B \oplus K_1 \subseteq A \oplus K_1$ 且 $B \oplus K_2 \subseteq A \oplus K_2$，我们从集合论中可知如果$R_1 \cup S_1$ 且 $R_2 \cup S_2$，则 $R_1 \cup R_2 \subseteq S_1 \cup S_2$，由此证明完毕。

- 外延性（extensive property）与抗外延性（anti-extensive property）：如果我们称一个算子是“外延的”，意思是将该算子应用到一个集合A，可生成一个包含A的结果。当结构元素包含了原点（即元素(0, 0)）时，膨胀是外延的：

$$A \oplus K \supseteq A \tag{7-12}$$

如大家所猜测的，腐蚀也有某种外延性：即，腐蚀是“抗外延的”。如果$o \in B$，则 $A \Theta B \subseteq A$，其中o代表位于原点的元素。

148

- 对偶性（duality）：对偶性与DeMorgan定律相似，它将集合的补运算、膨胀运算以及腐蚀运算联系起来。

$$\begin{aligned}(A \Theta B)^c &= A^c \oplus \tilde{B} \\ (A \oplus B)^c &= A^c \Theta \tilde{B}\end{aligned} \tag{7-13}$$

其中上标c表示集合的补运算。

- 腐蚀的其他性质。

$$A \Theta (B \oplus C) = (A \Theta B) \Theta C \tag{7-14}$$

$$(A \cup B) \Theta C \supseteq (A \Theta C) \cup (B \Theta C) \tag{7-15}$$

$$A \Theta (B \cap C) \supseteq (A \Theta B) \cup (A \Theta C) \tag{7-16}$$

$$A \Theta (B \cup C) = (A \Theta B) \cap (A \Theta C) \tag{7-17}$$

警告：不能撤销已完成的形态学运算。例如，对$A = B \Theta C$，令两边均关于C膨胀，得$A \oplus C = (B \Theta C) \oplus C$，如果膨胀与腐蚀的确互为逆，则右式应该等于$B$。但事实上，右式是$B$关于$C$的开运算，而不是等于$B$那样简单。

7.1.4 开运算与闭运算

f_A 关于一个结构元素f_B的开运算（opening）可写为

$$f_A o f_B = (f_A \ominus f_B) \oplus f_B \tag{7-18}$$

且（大家可能猜到）f_A 关于一个结构元素f_B的闭运算（closing）可写为

$$f_A \cdot f_B = (f_A \oplus f_B) \ominus f_B \tag{7-19}$$

一个应用

那么所有这一切都是为了什么？让我们看一个例子：检验印制电路板。这里有一幅印制电路板的图，在通过波动焊接机时，两个轨迹线被一根粘在板上的头发短路了。我们采用开运算来检验这个短路。

149

首先，用一个小的结构元素腐蚀该图像。选择一个比所感兴趣的特征（轨迹线）小一些，但比瑕疵大一些的结构元素。则腐蚀后看起来如下图：

现在，再使用同样的结构元素膨胀回来，

此时，可以非常惊奇地发现瑕疵不见了。为了达到检验的目的，可以从原始图像中减去开运算结果，则所得图像将仅包含瑕疵。此外，这些运算还可以由硬件实现，以异常快的速度运行。

另一种对开运算的思考方式

f_A 关于f_B 的开运算，就是精确地选择了f_A 中，那些以如下方式与f_B 相"匹配"的点：取结构元素f_B，并将它放置在f_A 上，使得f_B 的每个前景像素都能够覆盖f_A 中的某个前景像素。如果能够找到一个满足以上条件的位置放置结构元素，则f_A 中每个被覆盖的像素都在开运算结果中。现在，在四周移动（只准平移）f_B，并找到每个这样的点——每个可以放置结构元素的位置，使其完全在图像内，被覆盖的像素集合即决定了f_A 关于f_B 的开运算的结果。这里有一个一维开运算的例子，就是从这样的角度来观察的（摘自Haralick和Shapiro[7.12]）。令A是开区间

$$A = (3.1, 7.4) \cup (11.5, 11.6) \cup (18.9, 19.8) \tag{7-20}$$

并使用结构元素$K=(-1, 1)$进行开运算。则

150

$$AoK = (3.1, 7.4) \tag{7-21}$$

这个例子首先表明了形态学的概念可以扩展到连续域。(但是，请暂时记住，是在分辨率上连续，而不是在亮度值上连续，亮度值仍然是二值的。我们很快会修正这一点。)第二，它表明了这样一个事实：开运算精确地保留了“足够大的”对象的几何形状，而完全删除了小一些的对象。在这个意义上，开运算类似于均值滤波器的功能，后者使每个像素都被其邻域亮度的均值所替代。

7.1.5 开运算与闭运算的性质

开运算与闭运算的一些性质罗列如下（应该做好准备计算其中的某些性质）：

- 对偶性： $(AoK)^c = A^c \bullet \tilde{K}$ 。

对偶性证明 注意该证明是如何完成的。我们希望读者能够仔细地加以证明。

1） $(AoK)^c = [(A\Theta K)\oplus K]^c$ 开运算的定义

2） $= (A\Theta K)^c \Theta \tilde{K}$ 膨胀的补集

3） $= (A^c \oplus K)\Theta \tilde{K}$ 腐蚀的补集

4） $= A^c \bullet \tilde{K}$ 闭运算的定义。

- 等幂性（idempotency）：开运算和闭运算是等幂的。即同一运算的循环执行不会产生更进一步的影响：

$$AoK = (AoK)oK$$
$$(A \bullet K) = (A \bullet K) \bullet K$$

- 闭运算是外延的： $A \bullet K \supseteq A$ 。
- 开运算是抗外延的： $AoK \subseteq A$ 。
- 关于f_K膨胀后的图像在关于f_K进行开运算后保持不变。即$f_A \oplus f_K = (f_A \oplus f_K)of_K$。

证明

1） $A \bullet K \supseteq A$ 这是因为闭运算是外延的

2） $(A \bullet K)\oplus K \supseteq A \oplus K$ 因为膨胀是增长的

3） $((A\oplus K)\Theta K)\oplus K \supseteq A \oplus K$ 闭运算的定义

4） $(A\oplus K)oK \supseteq A \oplus K$ 开运算的定义

5）对任意B， $BoK \subseteq B$ 开运算是抗外延的

6）从而， $(A\oplus K)oK \subseteq A \oplus K$ ，用$A\oplus K$替代B

151

7） $(A\oplus K)oK = A \oplus K$ 这是因为$A\oplus K$既大于等于又小于等于$(A\oplus K)oK$，因而只能是相等

快速计算开运算与闭运算

任何平移不变的递增运算(Ψ)，例如开运算和闭运算，可通过如下形式的并集运算来实现

$$\Psi(A) = \bigcup_i A\Theta K_i \tag{7-22}$$

其中$K=\{K_1, K_2, \cdots\}$是某个结构元素集合，被称为该运算的一个基集（basis set）。腐蚀可以并行计算，而并集运算可采用查找表方法快速运行。请参见[7.17]中更详细的描述。

7.2 灰度形态学

到目前为止，均假定所处理的图像为二值图像。现在，放开这个要求并允许f_A可取一个有限范围($fmin_A \leqslant f_A \leqslant fmax_A$)内的连续的值。在这种扩展下，由于并集算子（union operator）没有定义，从而无法再对膨胀等运算进行简单的定义。为了定义灰度图像的形态学运算，需要首先定义一个新的概念，"本影"（umbra）。

二维灰度图像f_A的本影$U(f_A)$是所有满足$0<U \leqslant f_A(x, y)$的有序三元组(x, y, U)的集合。如果把f_A看作是连续值，则$U(f_A)$是一个无限集合。为了使形态学运算变得可行，假定f_A量化为M个值，

$$\#U(f_A) \leqslant M \cdot \#A \tag{7-23}$$

为了说明本影的概念，令f_A是一维图像，（这里，将一维函数解释为某一维始终为零的二维函数。现在给出一个例子，可以很容易地将它扩展为二维函数）：

$$A=\{(0,0),(1,0),(2,0),(3,0),(4,0),(5,0),(6,0)\}$$

其相应坐标的像素值为

$$f_A(x,0)=[1,2,3,1,2,3,3]$$

注意新的标记：由于f_A可取不同的值，这取决于考虑的是A中的哪一个元素，因此采用了函数符号。画出f_A可得图7-2。

在图7-2中，黑粗线代表f_A，本影是f_A下面的阴影区域，这幅图中，黑粗线位于本影顶部，可将本影写为一个有序三元组的集合：

$$\begin{aligned}U(f_A)=\{&(0,0,1),(1,0,1),(1,0,2),(2,0,1),(2,0,2),(2,0,3),(3,0,1),\\&(4,0,1),(4,0,2),(5,0,1),(5,0,2),(5,0,3),(6,0,1),(6,0,2),(6,0,3)\}\end{aligned}$$

152

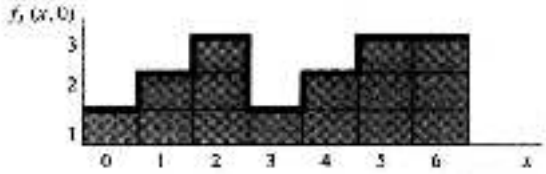


图7-2 对本影的举例说明

这里有一个奇特之处：尽管灰度图像不再是二值的（从而不能用集合隶属关系来表示），但本影仍然具备那些二值图像所具有的性质。

因此可以定义一幅灰度图像f_A关于灰度结构元素f_B的膨胀为

$$f_A(x,y)\oplus f_B(x,y)=\mathrm{TOP}(U(f_A)\oplus U(f_B)) \tag{7-24}$$

腐蚀也可同样方式定义。此外，灰度开运算和闭运算可根据灰度膨胀和腐蚀来定义。

推广这种概念至二维图像，则本影变成三维的，是如下三元组的集合

$$U(f(x,y))=\{(x,y,z)\mid(z\leqslant f(x,y))\} \tag{7-25}$$

那么，灰度膨胀和腐蚀可简洁地写为

$$g(x,y)\oplus h(x,y)=\{(x,y,z)\mid(z\leqslant \max(g(x-x_1,y-y_1)+h(x_1,y_1)))\}\ \forall(x_1,y_1)$$

$$g(x,y)\ominus h(x,y)=\{(x,y,z)\mid(z\leqslant \min(g(x-x_1,y-y_1)-h(x_1,y_1)))\}\ \forall(x_1,y_1)$$

对于$(x_1, y_1) \in \Omega \subset Z \times Z$，其中$\Omega$代表可能的像素位置的集合，这里假定像素位置取正整数。

7.3 距离变换

形态学运算的一个非常重要的应用是它引出了距离变换（The distance transform，DT）。有多种定义距离变换的方式，我们在第7A.4节中给出了一种，但这里将介绍另外一种定义，比先前的要简单。DT对图像的一种表示法，DT中的每个像素都包含了从某个特征到输入图像中相应像素之间的距离。最常见的情况下，采用某个边缘作为特征。在本节，将区域R的边界作为特征，并用点集∂R来表示。图7-3举例说明了该区域内部的距离变换。

153

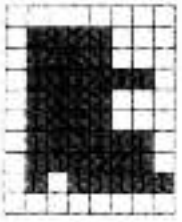


图7-3 区域内部的像素用阴影表示。假定边界恰好位于区域外部。图中显示了在计算中采用4连通邻域定义而得到的区域的距离变换

形式上，距离变换可描述为

$$DT(x) = \min_{y \in \partial R} \|x - y\| \tag{7-26}$$

其中x和y是坐标的两个向量。这个变换是如下微分方程的一个解

$$\|\nabla DT(x)\| = 1 \tag{7-27}$$

其初始条件为$DT(x)=0, x \in \partial R$。

我们可以采用传统形态学运算来计算距离变换：假设期望获得与对象外部边缘之间的距离，采用某种"适合的"结构元素重复腐蚀图像。每消失一个像素，记录下像素消失时的迭代步数。将该迭代步数储存到相应像素中作为距离变换。非常简单，对吗？这种定义并不能完全给出欧氏距离。Huang和Mitchell [7.15]给出了如何采用灰度形态学获得欧氏距离的变换。Breu等[7.6]给出了如何在线性时间内计算欧氏距离的变换。在这种迭代式腐蚀中，一个有趣的问题是"这些点到哪里去了？"可以证明，对一个严格凸的结构元素，这些点沿着与腐蚀的区域边界相垂直的方向消失了[7.43]。

7.3.1 采用掩模计算距离变换

计算距离变换可以通过迭代应用一个如图7-4所示的掩模（mask）来实现。

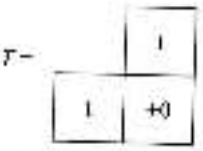


图7-4 用于计算4连通邻域距离变换的掩模。该掩模的原点用一个加号标记出来

在第m步迭代中，采用如下方程来更新距离变换

$$D^m(x, y) = \min_{(k,l) \in \Omega} (D^{m-1}(x+k, y+l) + T(k, l)) \tag{7-28}$$

现在给出更详细的解释：首先，距离变换$D(x, y)$在每个非边缘点上都初始化为一个特殊符号"无穷大"，$D^0(x, y) = \infty, \forall (x, y) \notin \partial R$。而对于任意边缘点，则取零$D^0(x, y) = 0, \forall (x, y) \in \partial R$。然后，从图像的左上角开始应用掩模，将掩模的原点置于图像中的像素（1，1）上，并应用公式（7-8）来计算（1，1）上新的距离变换的值。在图7-5所示例子中，距离变换用一个空白方块所表示的无限大来代表，而边缘则用零来代表。将图7-4所示掩模应用于阴影区域，则计算公式（7-28）的结果可得到箭头所指示像素的距离变换值min(1+0, 1+∞)。

经由自顶向下、自左向右的一次遍历后，翻转掩模（在两个方向上），再次自底向上、自

右向左地应用掩模。 154

在每一个像素上重复这一过程，这样遍历图像中的所有像素，然后再重复迭代，直至距离变换（DT）中的所有像素都标记为一个有限值。采用与图7-4所示掩模不同的掩模，可生成距离变换的其他形式。特别地，图7-6可生成切削映射。如果除以二，则切削映射生成的距离变换，可相当不错地近似欧氏距离。

7.3.2 Voronoi图

在本书的后面部分，将考虑区域之间的连接状态，以及相邻区域之间的关系。我们也会偶尔需要考虑并不真正相接触的两个区域之间的关系。对此，Voronoi图的概念是非常有用的，在本节介绍这个概念，是因为它与距离变换之间具有某些相似特性。

考虑图7-7所示图。在该图中，白色圆圈中的若干区域有灰色块。对任意区域i，该区域的Voronoi域是一些点的集合，这些点与该区域中点的距离要比与其他区域中点的距离更近：

$$V_i = \{x \mid d(x,P_i) < d(x,P_j), \forall (j \neq i)\} \tag{7-29}$$

其中P_i代表区域i中的任意点。

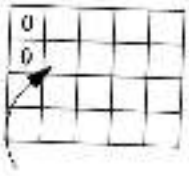


图7-5 计算某个特定点上的距离变换值，结果为1

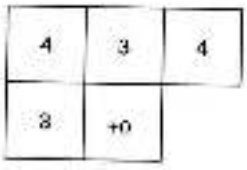


图7-6 一个可生成切削映射（使用模板进行距离变换的方法）的掩模

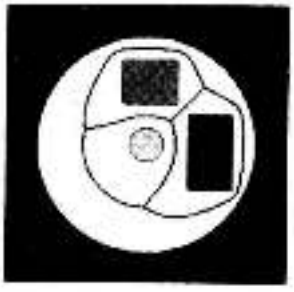


图7-7 若干区域以及每到的Voronoi图 155

同其中的两个区域等距离的点的集合称为Voronoi图，并在图中用黑线表示。这等于是说，Voronoi图中的点不能属于任意一个区域的Voronoi域。

7.4 结论

本章着眼于一种特殊的处理区域形状的方法。形态学运算对二值图像是尤其有用的，但也可以应用到灰度图像中。和本书中的大多数章不同的是，本章没有使用最优化方法或者一致性。

7.5 术语

闭运算	closing
膨胀	dilation
距离变换	distance transform
腐蚀	erosion
外延的	extensive

增长的	increasing
开运算	opening
本影	umbra
Voronoi图	Voronoi diagram

作业7.1

在第7.1.3节中，说明了膨胀满足交换律是由于加法满足交换律。腐蚀运算也包含了加法，只是需要对两幅图像中的一幅取其对称集。那么，腐蚀满足交换律吗？请证明是正确的或是错误的。

作业7.2

在第7.3节，提供了一个掩模，并说明了应用该掩模可生成一个距离变换，而该距离变换"相当不错地近似"从内点到最近邻边缘点的欧氏距离。那么它是的确很不错吗？请找出一个例子，使得应用该掩模而生成的值，不同于到最近邻边缘点的欧氏距离。

156

作业7.3

考虑一个面积为500像素，边界为120像素的区域。请找出每个内部像素到边界的距离变换，并使用欧氏距离。(注意：位于边界上的像素不认为是在区域之内，至少在这个问题上是这样。) 计算复杂度是多少？注意：可能你所提出的算法，优于得出如下答案的算法，因此如果你的算法不能得出如下结果，请给出解释说明。

a) 60000

b) 120000

c) 45600

d) 91200

作业7.4

一个技巧性问题：为了解答作业7.3中的问题，需要计算多少个平方根才能确定这个距离变换？记住，是欧氏距离。

作业7.5

请证明（或者证明为错）二值图像在经由一个核K腐蚀后，再经K做闭运算，图像保持不变。即，证明 $A\Theta K=(A\Theta K)\bullet K$ 。

作业7.6

请证明如果结构元素不包含原点，则膨胀不一定是外延的。

作业7.7

请证明膨胀是增长的。

作业7.8

令C是仅含单个黑色像素的一类二值图像。给定此类图像中的某个特定图像，令此黑色像素位于(i_0, j_0)。

157

请设计一个算子，采用包含元素{(0, 0)}的核，做腐蚀以及膨胀运算，即，由腐蚀与膨胀，以及结构元素（可能只需要一个结构元素，也可能需要多个）组成的集合，使得该结构元素应用于C的一个元素时，输出图像中的黑色像素移动到(i_0+2, j_0+1)（忽略边界）。

作业7.9

如下式子中哪一个是正确的？（应仅通过式子来推论，而不必做任何证明。）

$$\text{(a)}\ (A\Theta B)\Theta C = A\Theta(B\Theta C) \tag{7-30}$$

$$\text{(b)}\ (A\Theta B)\Theta C = A\Theta(B\oplus C) \tag{7-31}$$

作业7.10

采用作业5.5以及作业5.6中生成的阈值化后的图像。选择一个合适的结构元素，然后应用开运算消除噪声。

专题7A 形态学

7A.1 有效地计算腐蚀、膨胀

公式（7-14）（或公式（7-31））是正确的，这正是我们希望大家在做作业7.9时能够推断出的结论。也就是说关于一个大的结构元素（比方说K）的腐蚀，可以分段为两个相续的腐蚀，首先关于B，然后再关于C，只要我们能够找到B和C，使得$K=B\oplus C$。有时这种情况称为腐蚀的“链规则”。它对硬件实现有着巨大的影响。

假设有特定的硬件，能够以帧速率完成结构元素为3 × 3的腐蚀运算。但在一些应用中，却需要采用一个特定的4 × 4结构元素进行腐蚀。链规则说的是，此时假如能够（以某种方式）找到两个3 × 3结构元素，使得关于它们的膨胀也就是所需要的关于4 × 4结构元素的膨胀，则通过这个特定的硬件对输入图像遍历两遍，就能得到同样的结果。但是如何找到B和C呢？为了说明应该如何完成这一点，现在考虑一个非常简单的分解方法，即分解为一组仅包含两个元素的结构元素，这两个元素包括原点和另外某一个点。即期望找到$H_1, H_2, \cdots, H_N$，使得$A\oplus H=(\cdots[(A\oplus H_1)\oplus H_2]\cdots)$，且$H_i=(0, p_i)$。可采用如下方法来寻找$p$ [7.12]：在H中的所有点对中搜索一对点p_1和p_2，使H在关于这两个点的差的开运算下是不变量$H=Ho\{0, p_1-p_2\}$。如果找到这样的两个点，则$H_i=(0, p_1-p_2)$，我们用H相减而得$H'=H\Theta H_i$。递归地完成这一过程。如果无法找到这样的两个点，则试图搜索四个点p_1, p_2, p_3, p_4，使之满足 $H=Ho\{p_1-p_2, p_3-p_4\}$，依次类推。

158

Matheron[7.21]证明了任意大的形态学运算都能够转化为一组腐蚀的并集运算，或者根据对偶性，转化为一组膨胀的交集运算。如何选择一组“基集”，以便在这种方式下，使用给定结构元素完成给定运算，一直以来都是一个相当重要的研究课题[7.18，7, 20]。

分解为一组3×3的结构元素

下面，将解释说明Park和Chin的方法[7.23]，并采用他们的例子。这里也将对算法进行介绍。读者可以参考原论文以了解相关理论。本节给出的例子来自[7.23]。

为了理解结构元素的分解方法，需要首先了解第9.5节描述的关于链码的知识。链码，是对沿着逆时针方向围绕一个区域行走的一种表示方法，这种方法采用了一个数字序列，序列中所有数字都在0和7之间，代表每一步的行走方向（见图7-8）。在这项工作中，稍稍扩展了链码的概念，使得不仅包括了八个主方向上单个像素的跨步，还包括了沿着一些特定类型的凹的边界片断的跨越。

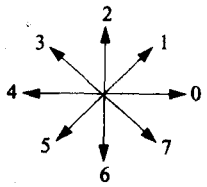


图7-8 八个方向，可在边界处沿着其中一个方向从一个像素走到另一个像素

现在将结构元素分解为一组3×3元素。首先，观察发现可填充一个3×3区域的凹的边界片断类型仅有28种。图7-9中列举出了所有类型并为它们命名。

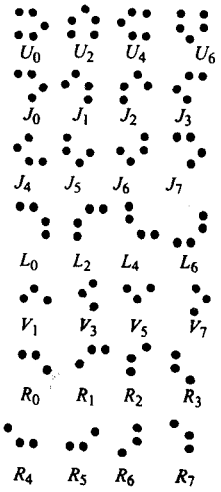


图7-9 3×3图像中可能出现的凹边界。数字代表第一个链码的方向

现在定义一类相当普通的结构元素，是单连通的（没有空洞），且边界可表示为某种形式：一个含有凹的边界片断和链码主方向的正则表达式。例如，正则表达式$L_0 2^2 4^2$代表了一条这样生成的曲线：从L_0片断出发，接下来沿着“2”方向前进两步（上标代表重复次数），最后沿着“4”方向前进两步，如图7-10所示。

能够采用该算法来分解的一类结构元素就是所有单连通结构元素的集合，这些结构元素的边界可以写为如下形式（在这种定义下，下标的顺序是很重要的）

$$U_0^{S_{U0}} J_0^{S_{J0}} L_0^{S_{L0}} R_0^{S_{R0}} 0^{S_0} J_1^{S_{J1}} V_1^{S_{V1}} R_1^{S_{R1}} 1^{S_1} \cdots J_7^{S_{J7}} V_7^{S_{V7}} R_7^{S_{R7}} 7^{S_7} \tag{7-32}$$

其中任何一个上标都可以是零。例如，$V_1 1^2 2 R_4^2 4 6^3 R_7$是该集合中的一员，而$J_7 1 J_2$则不是。

定义

称图像A是图像S的一个因子（factor），当且仅当能够把S写为关于A的膨胀，即$S=A\oplus B$。称因子A是一个素因子（prime factor），当且仅当A不能分解为除了它自身与单像素图像之外的任意形式。在表7-1中，列出了所有从R_0出发的素因子。素因子并不要求满足公式（7-32）的形式。在表7-2中，介绍了其他素因子，但只列出了它们的链码表示。

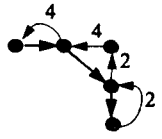


图7-10 图例表示如何用$L_0 2^2 4^2$来描述一个形状。L_0用粗线画出

现在，用一个例子来说明结构元素的分解方法。现在分解如图7-11所示的结构元素，其链码是 $S = L_0 0^3 1 2^4 R_4 4^3 R_6 6^2$ 。将该边界上的凹的部分记为$v_1=L_0$, $v_2=R_4$, $v_3=R_6$。而凸的部分记为$d_1=0$, $d_2=1$, $d_3=2$, $d_4=4$, $d_5=6$。

表7-1 从R_0出发的素因子

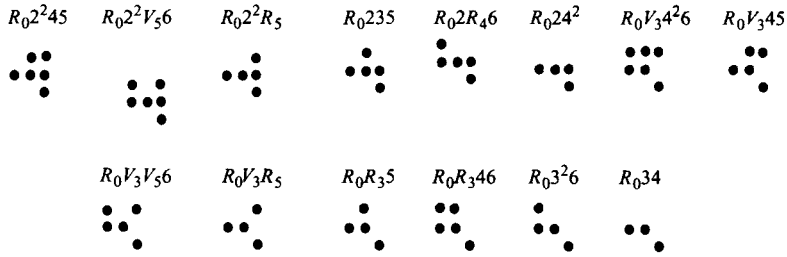


表7-2 其他从U_0, J_0, I_0, V_1, R_0和R_1出发的素因子。为了获得i=2, 4, 6时的V_i, J_i 等，给每个链码加i。而i=3, 5, 7时，则给每个链码加(i－1)

U_0	$U_0 0^2 2^2 4^2$	$U_0 0^2 234$	$U_0 0124^2$	$U_0 0134$	
J_0	$J_0 02^2 4^2$	$J_0 0234$	$J_0 124^2$	$J_0 134$	
J_1	$J_1 2^2 4^2 6$	$J_1 2^2 45$	$J_1 2346$	$J_1 235$	
L_0	$L_0 2^2 4^2$	$L_0 234$			
V_1	$V_1 2^2 4^2 V_7$	$V_1 2^2 4R_6$	$V_1 2^2 V_5 6^2$	$V_1 2^2 V_5 V_7$	$V_1 2^2 R_5 6$
V_1	$V_1 234V_7$	$V_1 23R_6$	$V_1 2R_4 6^2$	$V_1 2R_4 V_7$	$V_1 24^2 6$
V_1	$V_1 V_3 4^2 6^2$	$V_1 V_3 4^2 V_7$	$V_1 V_3 456$	$V_1 V_3 4R_6$	$V_1 V_3 V_5 6^2$
V_1	$V_1 V_3 R_5 6$	$V_1 V_3 5^2$	$V_1 R_3 46^2$	$V_1 R_3 4V_7$	$V_1 R_3 56$
V_1	$V_1 2^2 5^2$	$V_1 245$	$V_1 V_3 V_5 V_7$	$V_1 R_3 R_6$	
V_1	$V_1 3^2 6^2$	$V_1 3^2 V_7$	$V_1 346$	$V_1 35$	
R_0	$R_0 2^2 45$	$R_0 2^2 V_5 6$	$R_0 2^2 R_5$	$R_0 235$	$R_0 2R_4 6$
R_0	$R_0 242$	$R_0 V_3 4^2 6$	$R_0 V_3 45$	$R_0 V_3 V_5 6$	$R_0 V_3 R_5$
R_0	$R_0 R35$	$R_0 R346$	$R_0 3^2 6$	$R_0 34$	
R_1	$R_1 24^2 V_7$	$R_1 24R_6$	$R_1 2V_5 6^2$	$R_1 2V_5 V_7$	$R_1 2R_5 6$
R_1	$R_1 25^2$	$R_1 346^2$	$R_1 34V_7$	$R_1 356$	$R_1 3R_6$
R_1	$R_1 R_4 6^2$	$R_1 R_4 V_7$	$R_1 4^2 6$	$R_1 45$	

首先，确定所有包含L_0、R_4和R_6，并与该图像相容的素因子，如图7-12所示。为了更清楚地理解它们是如何与图像相容的，考虑片断$R_4 6^2 01$，观察到这与图7-11中右上角的R_4片断相配。

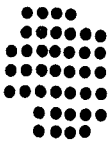


图7-11 一个待分解为一组3×3元素的结构元素

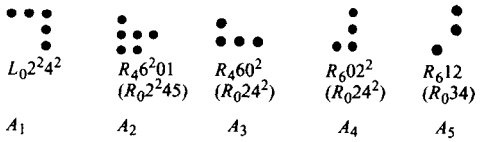


图7-12 与图7-11中的边界片断相匹配的素因子，括号内的编码表示旋转后所得等价边界从R_0出发的编码

这一过程的下一步是建构一个矩阵Θ，其中Θ_{ij}代表v_i在A_j中发生的次数。在这个例子中，

$$\Theta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

从中可以看出，R_4在A_2和A_3中各发生了一次，但没有在A_1、A_4和A_5中发生。再下一步，建构一个矩阵Ω，其中Ω_{ij}是对d_i在A_j中发生的次数的计数。这个例子中，

$$\Omega=\begin{bmatrix}0&1&2&1&0\\0&1&0&0&1\\2&0&0&2&1\\2&0&0&0&0\\0&2&1&0&0\end{bmatrix}$$

再定义两个向量，Y代表v_i在初始边界发生的次数，而Z则代表d_i在初始边界发生的次数。在这个例子中，$Y=[1\ 1\ 1]^T$且$Z=[3\ 1\ 4\ 3\ 1]^T$。满足如下式子的向量X

$$\begin{aligned}\Theta X&=Y\\ \Omega X&\leq Z\end{aligned}\tag{7-33}$$

是该分解的一个解。在这个例子中，$X=[10110]^T$满足以上两个方程。注意$\Theta X=[30421]^T$小于等于Z，也就是ΘX的每个元素都要小于等于Z中相应元素。从而，可以经由关于A_1的膨胀一次，关于A_2的膨胀零次，A_3一次，A_4一次，以及A_5零次，来分解边界，这一过程可以表示为$S=A_1\oplus A_3\oplus A_4\oplus B$。

现在只剩下一件事情，就是确定核B。办法是考虑ΘX与Z之间的差。$Z-\Theta X=[01011]^T$。因此，所求的核的边界可通过序列d_2, d_4, d_5来描述，每一个仅重复一次。这一序列是146，如图7-13所示。这样，得出了一个结构元素序列，每个大小都是3×3，且依次应用该序列所得结果与应用图7-11中的核所得的结果相同。

图7-13 序列146所描述的结构元素

7A.2 形态学采样定理

现在，来谈谈采样定理，非常类似于著名的香农（Shannon）采样定理。不同的是我们将从一幅已经采样过的图像开始，并用一个数字网格（digital grid）来表示该图像，然后考虑如何将它（进一步）采样为更小的网格。

首先定义采样网格（sampling grid）。一个所谓的采样网格，是一幅这样的图像，其前景由原始图像上每一个要被采样的点构成。任何一种见过的网格都可以，只是采样网格还必须满足

160~161

$$S\oplus S=S\tag{7-34}$$

以及

$$S=\tilde{S}\tag{7-35}$$

公式（7-34）表明了这样一个性质，即这个特定的S“在膨胀下是闭合的。”一种常见的网格是对每三个点进行一次采样，如图7-14所示。

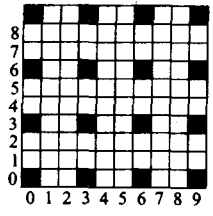


图7-14 对每三个点进行一次采样的表示图

那么现在，采样就意味着读取并记录与网格所有黑色像素位置对应的图像值。现在假设K是某个结构元素。这里提出一个相当简单的重构（reconstruction）算法，基本思路是：用指定的采样网格对图像进行采样，然后，用某个结构元素K进行膨胀。那么问题是：在什么条件下，采样图像关于结构元素K的膨胀结果仍为原始图像？对此，

Florencio和Schafer[7.7]指出需要满足如下性质：首先，采样网格自身关于结构元素膨胀所得结果，必须是整个空间：

$$S \oplus K = \xi \tag{7-36}$$

其次，K关于S中所有点的平移构成空间的一个分区

$$\forall (x,y) \in S, x \neq y), K_x \cap K_y = \varnothing \tag{7-37}$$

并且K必须包含原点。

这里有一个结构元素，可令上文给出的采样网格满足这三个条件：

事实上，它就是一个中心像素（原点）及其8邻域，这并不十分有趣。但它确满足这三个条件。

如果S和K满足这些性质，则定理描述如下：令F为某图像，令P是对其$A \subseteq S$满足$F_A = A \oplus K$的图像F的集合，且令Q是满足$F_A = F_A \circ K = F_A \cdot K$的图像$F_A$的集合，则以下两部分成立：

部分A　按S中的点，对$F \in P$进行采样，对完全重构出F是充分且必要的。

部分B　按S中的点，对$F \in Q$进行采样，对重构出具有边界误差的F是必要的，而对重构出最大误差为$r(K)$的F是充分的，其中$r(K)$是包含了K的最小环的半径。

关于这个定理，还可以讨论很多：首先，部分A意味着什么？你能用集合符号将它描述出来吗？你应该可以做到，但你可能会认为"如果它所意味的正是我所想的，那太微不足道了。"好的，它的含义是：如果取S中的某些点的集合，并关于核K进行膨胀，能够生成F，则我们需要记住的是S中的那些点是必须的。是的，在某种意义下，这的确很微不足道，不是吗？但它意味深长。这一点，在讨论尼奎斯特（Nyquist）率时会有所体会。 [162]

现在，来理解一下部分B：豪斯多夫（Houssdorf）距离是对两幅（子）图像之间差异的度量。给定两个点集$S = \{s_1, s_2, \cdots, s_n\}$和$T = \{t_1, t_2, \cdots, t_m\}$，豪斯多夫距离的定义是

$$d_H(S,T) = \max(h(S,T), h(T,S)) \tag{7-38}$$

其中$h(S,T) = \max_{s \in S} \min_{t \in T} \|s-t\|$。即，豪斯多夫距离是一个集合中的元素与另一个集合中元素的最小距离的最大值。

部分B说的是，如果F在关于K的开运算下是闭合的，即，如果F在关于K做开运算后再关于K做闭运算，F并没有发生改变，则F在S中点上进行采样，对几乎精确地表示出F是充分的。这里"几乎精确"的含义是：采样点的集合关于K膨胀后，几乎完全等价于初始的F。

现在来看一个例子：图7-15给出了原始图像F_A，以及采用S采样后所得图像。

由于已知我们只能获得第0、3、6、9行或列的数据，因此在对图像降低分辨率时，可以丢掉其他的行和列，从而得到一幅小一些的图像。这样，在图7-16的左侧，得到了一个对原始图像间隔采样后的小图像。现在通过膨胀将它放大回来：将K放在每个采样点上，能够得到图7-16所示的重构结果。它看起来并不是非常像原始的未被采样过的图像，对不对？（不论"精确重构"是如何定义的？）这是一个艰难的问题，以上定理声称能够通过在每三个点上采样一次来重构一个信号，这不是同香农定理相矛盾吗？香农不是说过必须总是每隔一点采样一次吗？（事实上，香农定理是针对模拟信号而阐述的。该如何将香农定理应用到这一情况？耐奎斯特率究竟是什么？）我们的采样定理并没有说可以重构出任意的信号，它说的是

某种关于频率的东西，不是吗？（请说，是的。）事实上，采样定理基本上说的是，一个在某个特定频率（与网格间距相同）下的采样网格不能够用于存储其变化多于半个网格间距的信息。在形态学采样中，定理变得复杂是由于这样一个事实，我们不仅可以选择一个网格，还可以选择一个核。定理的部分B给出了一些约束信息，除非图像预先经由K进行"预滤波"（prefiltered），否则K无法令它复原。对此，Haralick[7.12，p.252]则以如下方式叙述："形态学采样定理不能实现这样的重构，它要求其定位精度要高于重构结构元素K的外切圆的半径。"

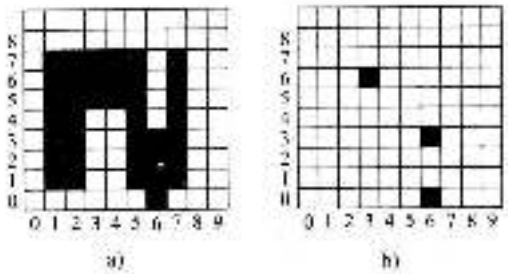


图7-15 a）原始图像 b）采用图7-14中的S对原始图像采样后的结果

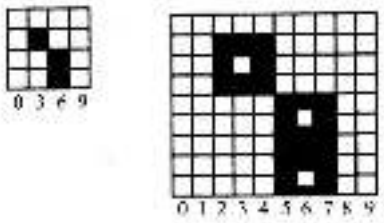


图7-16 间隙采样后的图像与通过膨胀所得重构图像。注意图7-15中的原始图像——该图像属于P或Q吗？

163

7A.3 选择结构元素

通常，我们选择原点在中心的，尺寸较小且对称的结构元素。Schonfeld[7.34]给出了一些建议，但一般情况下，这些建议也仅可作为指南。

7A.4 边缘及曲面上的缝隙弥合[⊖]

与本书的惯例不同，本节中的算法描述要稍微详细一点。一名机器人视觉学科的研究生有必要学会如何写出一篇好的期刊论文，而这一章的写法可看作是对这样一篇论文的各个组成部分的示范。它包括引言、相关工作、描述并解释算法的主体部分，以及解释并说明算法与其他已发表算法之间性能比较的结果部分。因此人家不仅应该阅读本章的技术内容，而且还应该注意文章的组织结构。

7A.4.1 引言

在本节，将应用距离变换、连通分量（见第8章），以及类似于二值形态学运算的概念，解决边缘缝隙弥合（其他策略可参见[7.16]）问题。在二维情况下，边缘是一条曲线，而在三维情况下，则是一个曲面。我们知道，任何边缘算子都必然会遇到失败的情况，此时会导致生成额外的边缘或边缘上的缝隙。如果边缘上有了缝隙，则连通分量标号程序会失败，会将内点和外点标成相同的号，因此，必须要找到一个能够校正这种边缘检测错误的方法。我们把这些方法与形态学运算联系起来。在[7.10]中，提出了一种所谓的基于距离变换的弥合方法，并在二维和三维下均实现了这一方法。这里将这一新方法与二值形态学（掩模腐蚀方法[7.1]）和迭代并行细化方法[7.44]、三维并行细化技术[7.42]和三维形态学方法等进行了比较。在所有情况下，该方法都表现出了更优的性能，因为它能够更好地保留尖锐拐角的形状。

164

⊖ 本节中的图片来自参考文献[7.36]。

一个题外话：连通分量标号

第8章介绍了一种所谓的“连通分量标号”（connected component labeling，CCL）运算。这里，很快就要使用连通分量的概念。如果还没有阅读该部分，请首先阅读并理解第8.3节，然后再回到本节。

问题定义

给定一幅边缘图像。由于噪声、模糊或其他误差，真实图像经边缘检测后，边缘或曲面上有时候会有“缝隙”——由于边缘检测器的反馈不够强而无法得到肯定的判断。（请参见Pratt[7.25，第17节]中对边缘检测误差的精彩讨论。）有多种形态学运算可用来弥合这样的缝隙。这里给出的算法是所谓的“DT驱动的弥合方法”（DT-driven closing）

7A.4.2 距离变换

我们已经知道，距离变换$D(x, y)$是对从点x, y到最近边缘点（nearest edge point）的距离的某种度量[7.28，7.29，7.31，7.32]。我们采用了一种特别的距离变换，称之为切削映射（使用模板进行距离变换的方法），记为$C(x, y)$，并称该映射的值为“距离变换（DT）距离。”

同样，可以把距离变换的概念扩展至三维空间。变换$D(x, y, z)$同理可参数化为$g(x, y, z)$，且包含从点$<x, y, z>$到最近边缘的DT距离。

距离变换可用于辅助度量区域的特性：通过距离变换，可以得到在每一点上，与任何一条边缘都不相交的最大的核的尺寸[7.35]。

一个点(x, y)的k近邻可在二维或三维空间内定义为：

$$N_k(x, y)=\{\langle x_{11}, y_{11}\rangle, \langle x_{12}, y_{12}\rangle, \cdots, \langle x_{km}, y_{km}\rangle\}$$
$$=\{\langle x_{ni}, y_{ni}\rangle \mid \text{距离变换距离}(\langle x_{ni}, y_{ni}\rangle, \langle x, y\rangle)=k\}$$

在二维情况下，$m=8k$，$k\geq 1$，而在三维情况下，$m=(2k+1)^3-(2k-1)^3$。

在我们的早期工作[7.10]中，凡是对应用了基于公式（7-28）的掩模所得的距离变换，都称作“切削映射[7.2]”，但是在其他文献中[7.4，7.5，7.31]，术语“切削”（chamfer）仅限于基于图7-17中的第二个掩模的距离变换。不过二者都能满足将在下一节进行描述的边缘弥合方法。

165

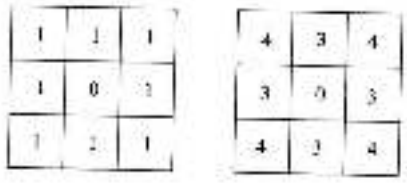


图7-17 对距离的两种定义：二者均可用于计算距离变换。位于右侧的可生成切削映射，非常接近欧氏距离

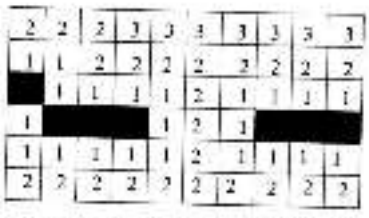


图7-18 边缘缝隙附近的距离变换

为了建构一个完整的距离变换，迭代计算公式（7-28），直至两步迭代之间不再发生变化。这就是Bister等[7.3]所采用的方法。但是，对于这里讨论的应用，我们需要设定某种关于最大缝隙长度的先验知识，从而可以定义一个值K_{max}来反映这项知识。K_{max}将代表一个足够大的距离，以致于使得距离如此远的像素决不可能是边缘的一部分。一般情况下我们采用值$K_{max}\leq 4$，此时能够弥合六个像素大小的缝隙。图7-18给出了一个关于边缘缝隙附近的距离变换的例子。

可采用相类似的方式生成三维距离变换。我们发现生成距离变换时的计算复杂度与图像中边缘点个数以及K_{max}的大小成正比。三维情况下，K_{max}增大时，核的大小$\aleph_k$会大幅度增大。

7A.4.3 基于连通分量的分割

到某一条边缘的距离变换的距离低于K_{max}的像素被标注为“TBA”（to be assigned，待赋值），其含义是，它们可能属于某一条边缘。余下的像素分割为互不交叉的区域。实现这种分割的一种简单直接的方式，是对未标注为TBA的像素进行连通分量分析。不过，也可以采用更加复杂的协同处理过程，例如第8章和[7.14]中所描述的处理方法。无论采用何种方法，都可生成一幅图像，比如将在第8章中讨论的标号图像，其中$L(x, y)=j$表示“$\langle x, y\rangle$ 处的像素属于区域j。” L中距离边缘较近的像素，标注为边缘的潜在部分：

166

$$\{\mathrm{DT}(x,y) < K_{max}\} \rightarrow \{L(x,y) = \mathrm{TBA}\}$$

Bister等[7.3]对距离变换中的局部最大值进行辨识，从每个最大值都得到一个潜在的区域。但是，他们注意到，“由于距离变换对于噪声非常敏感，导致区域边界很不规则，因此一个凹处（cavity）（区域）可以包含多个十分接近的局部最优。为了排除假的最大值，采用了一个滤波器来融合（merge）这些最优值，使得高度之和比它们之间的几何距离要大得多。”我们推测该滤波器的作用与我们选择的K_{max}相当。只是我们没有去搜索一个最优值，而是采用了连通分量（第8章）。这两种方法均可在区域中确定一个范围，这个范围能够鲁棒地表征该区域。

7A.4.4 重标号未赋值点

算法的最后一步是重标号（relabeling）待赋值（TBA）点：采用距离变换$\mathrm{DT}(x, y)$和标号图像$L(x, y)$，生成一个新的标号图像$L'(x,y)$。从$\mathrm{DT}(x, y) \geqslant K_{max}$的点（与边缘的距离为$K_{max}$或者更远的点的集合）开始进行，每个点都要重标号，并用该点附近最恰当的邻近点的标号来赋值。对有疑问的像素，进行迭代式“腐蚀”。在每一步迭代i中，标号图像$L(x, y)$中只有与$\mathrm{DT}(x, y)=K_{max}-i$的像素相对应的待赋值像素，才会在每一次遍历中进行重标号。

TBA像素仅在有较强的论据（将在下一节中定义）证明该像素应赋值为某个区域时，才进行重标号。如果无法确定是否再赋值，则该像素仍表示为TBA并推迟到下一次遍历时再做决定。当$L(x, y)$中所有与距离变换距离为k的像素相对应的TBA像素被重新赋值为一个合理的图像区域后，或者在迭代过程中不再发生变化时，k减1，并考虑对下一个TBA集合进行重标号。当所有边缘像素（$k=0$）都被赋值为合理的图像区域时，重标号过程结束。如果要求区域之间的分辨率为单像素分辨率，则应采用更复杂的区域规范化方法[7.8]。

缝隙弥合的关键在于，用什么策略选择最合理的区域，为当前像素$L(x, y)$重新赋值。在二维空间内，为了达到这一点，应检验该像素周围的八个邻域。在三维空间，则应检验该体素的26个邻域。我们称之为搜索像素或体素的“最佳邻域”。

像素四周的邻近像素属于如下三种可能的类中的某一类，这取决于在当前一步迭代中，是如何对这些邻近像素进行标号的：

1）对应于标号图像$L(x, y)$中的某个对象或某个对象的某个区域的标号。

2）对应于图像背景的标号。

3）TBA标号。

若要将当前像素重标号至前两类中的某一类，可以首先简单地计数有多少个邻近像素属于这些类，然后选择计数结果最大的类作为“最优”类，从而对当前像素进行重标号。这种

策略在所有邻域都是TBA时会失败，或是由于在重标号算法中引入的约束（见下一段），不希望将该像素的重新赋值与看上去是它的最优邻域相同。 167

在重标号算法中的一个约束，是针对选择背景区域作为最佳邻域的。相比于属于前景区域的邻域像素，我们以更加严格的方式来考虑属于背景的邻域像素。背景像素必须与当前考虑的像素直接相连时，才会选取背景值为最佳邻域。这一点可以避免造成有时候会在一个封闭边界内找到孤立的背景像素这样一种不合宜的结果。（请参见[7.30]中更多关于这种连通性的反常性（paradox）的讨论。）

当k=0时，为了进行重标号而搜索的TBA像素，或者是实际边缘或者是噪声像素。根据定义，图像中的一条边缘，代表一个对象区域与图像中另外某个区域之间的分界。这另一个区域可能是另一个对象区域，也可能是背景。选择一个边缘像素并重标号为图像区域的一部分时，没有考虑计数结果。只在边缘像素的周围全部是背景的情况下，选择背景为最佳邻域。

重标号算法　三维数据

```
/* 在这个函数例子中，我们省略了若干的细节，例如变量定义。不过，这个例子还是抓住了算法的本质 */
{
/* 在所有体素上循环执行 */
  for(frame = 0; frame < numberframes; frame++)
  for(row = 0; row < numberrows; row++)
  for(col = 0; col < numbercols; col++) {
    if(DT[frame][row][col] >= Kmax)
      Lprime[frame][row][col] = L[frame][row][col];
    else if(DT[frame][row][col] == 0)
      Lprime[frame][row][col] = EDGE;
    else Lprime[frame][row][col] = TBA;
    }
  for(k = Kmax-1; k > 0;k--){
    changing = TRUE;
    while(changing == TRUE) {
      changing = FALSE;
      for(frame = 0; frame < numberframes; frame++)
      for(row = 0; row < numberrows; row++)
      for(col = 0; col < numbercols; col++){
        if(((L[frame][row][col] == TBA)
          || (L[frame][row][col] == EDGE))
          &&(DT[frame][row][col] >= k)) {
          Lprime[frame][row][col] =
Best26Neighbor(frame,row,col);
          changing = TRUE;
          }
else Lprime[frame][row][col] =
L[frame][row][col];
    } /* 结束对帧、行、列的循环 */
     copyarray(Ltemp, Lprime); /* 将整个数组Lprime 复制到Ltemp */
     copyarray(Lprime, L);
    copyarray(L, Ltemp);
    } /* 结束while */
```

168

```
  } /*结束for k */
} /*结束relabel */
/*-----------------------------------------------------------------------*/
/* 在该函数中，p和n是含有行和列的数据结构 */
/* 一个体元素的坐标 */
int Best26Neighbor(p)
  {
  while (n = neighbor(p)) != NULL)
    {
    if (L(n) != EDGE)
      {
      if(L(n) != BACKGROUND) Card[L(n)]++;
      else if(faceconnected(n,p)) Card[BACKGROUND]++;
      }
    }
    if ((maximum(Card) == BACKGROUND) && (L(n) == EDGE))
      return NextMax(Card);
      else return maximum(Card);
} /*结束Best26Neighbor */
```

在这个算法中，"Card"就是一个数组，它维护的是对一个特定的标号与所感兴趣的体元素之间相邻次数的计数结果（cardinality）。

7A.4.5 例子

本节将为上文所描述的方法提供竞争形态学策略。

紧密相邻的区域

其个边缘检测程序的输出结果如图7-19所示。注意边缘上有一个很典型的缝隙，其大小大于两个区域之间的间隔。将距离变换应用至该边缘图像，距离变换以及连通分量标号的结果如图7-20所示。而图7-21显示的是由基于距离变换的弥合算法得到的分割结果。

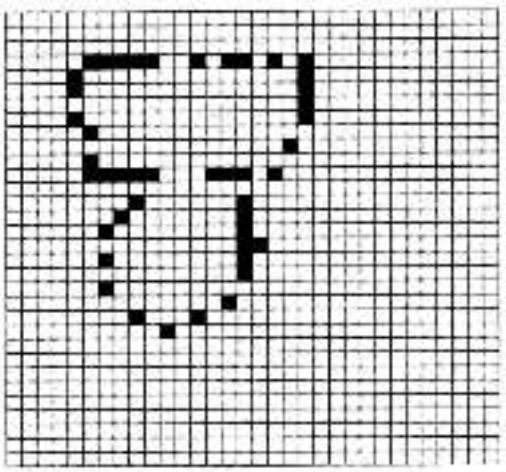


图7-19 边界上有较大缝隙的两个区域。摘自[7.36]，已获得使用许可

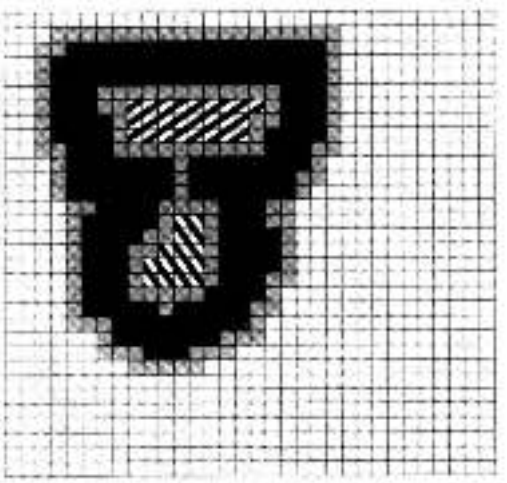


图7-20 距离变换以及对图7-19进行连通分量标号后所得结果。摘自[7.36]，已获得使用许可

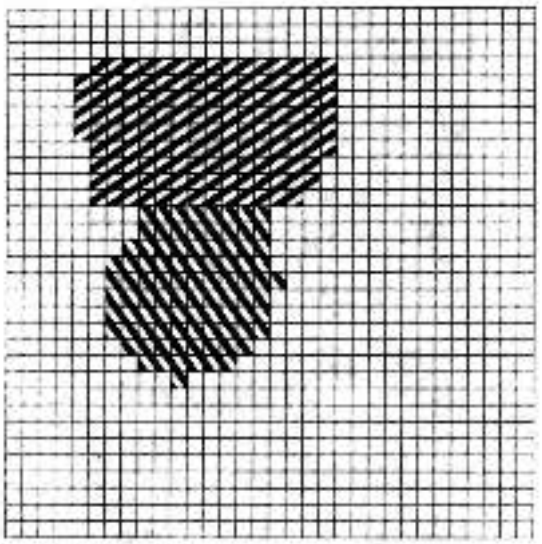
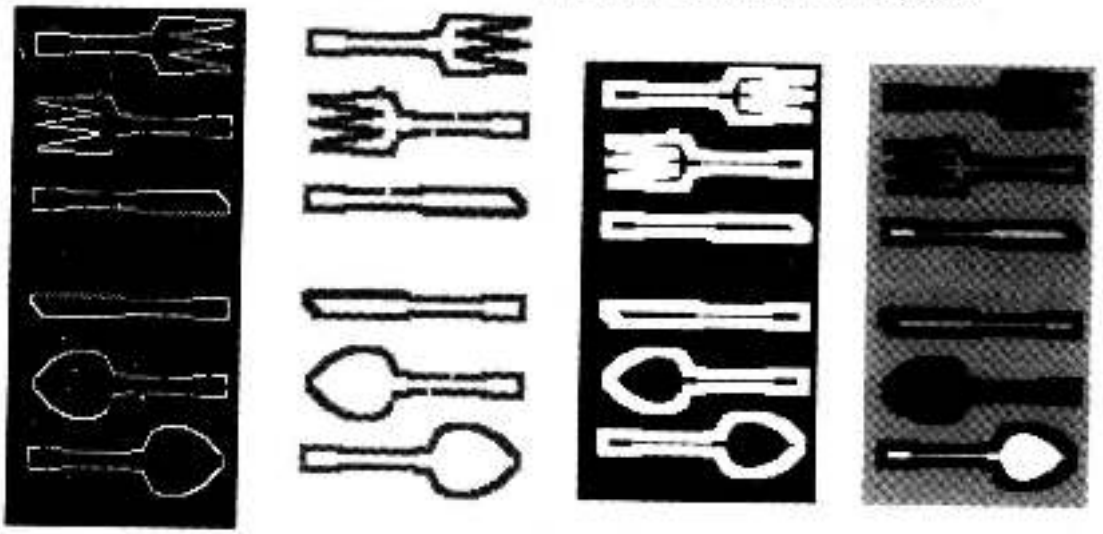


图7-21 应用区域驱动的扫描算法所得分割结果。注意这里正确地划分了区域，划分精度渐增到了单位像素。摘自[7.36]，已获得使用许可

其他的二维图像

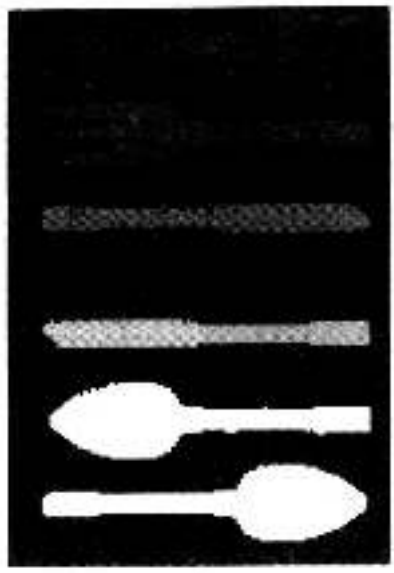
对一幅包含若干餐具对象（刀叉勺汤匙）的二值图像进行处理。图7-22~图7-24给出了对该输入图像处理后得到的结果，按次序分别表示：a）原始图像；b）距离变换本身；c）"加粗后的边缘图像"，包括所有满足$C(x,y)<K_{max}$从而在初始时重标号为TBA的像素；d）对加粗后的边缘图像进行标号所得图像；e）基于距离变换与标号图像所得的重标号图像。 169



a) b) c) e)

图7-22 a）原始的餐具图，边缘上有缝隙 b）原始餐具图像的距离变换。摘自[7.36]，已获得使用许可

图7-23 c）加粗后的边缘图像。 d）对膨胀后所得边缘的标号图像。摘自[7.36]，已获得使用许可



e)

图7-24 e) 重标号后的餐具图像。摘自[7.36]，已获得使用许可

7A.4.6 与已发表的二维方法之间的比较

为了评估DT驱动边缘弥合方法对二维图像的处理性能，将该方法与另外两类二值细化方法——迭代式腐蚀以及迭代式细化，进行了比较。这里选用一幅题为“餐具”的位图作为以上几类算法的输入。图7-22中给出了该图关于5 × 5的核膨胀后所得结果，该运算旨在对图中所有缝隙进行封闭。基本思路是：首先使用膨胀运算以弥合边界缝隙；然后使用细化技术使得“粗的”边界变为“细的”边界。比较结果将在下一节进行描述。

需要注意的是，尽管细化过程和骨架化过程所完成的运算非常类似，但是细化能够保留边缘的连通性，而骨架化则不能。根据形态学运算，可以对骨架进行如下定义。

结构元素A的第k阶同位homothetic，是将当前所关心的算子（这里是膨胀）应用到A（k次——译者注）后所得结果，记为kA。即，采用结构元素A对{0}进行膨胀，接着再次采用A对所得结果进行膨胀，直至完成k次这样的膨胀为止。

170
↓
172

有一种常规的形态学骨架化方法，是将图像分解为多个骨架子集S_i，例如对图像X，有

$$S_i = (X \ominus iA) \setminus [(X \ominus iA) \circ (A)] \tag{7-39}$$

其中，$X \setminus Y$代表所有属于X而不属于Y的元素，此时可由如下公式来构造骨架（skeleton）：

$$Skeleton = \bigcup_{i=1}^{N} S_i \oplus iA \tag{7-40}$$

上述各个子集中包含了关于尺寸、方向和连通性的信息。最小骨架具有能够精确重构出原始图像的性质，但它并不必然保留路径或曲面的连通性[7.20]。另一种可选的形态学骨架化方法，是形态学形状分解方法（morphological shape decomposition, MSD），这里没有予以考虑。Reinhardt和Higgins[7.27]对MSD和形态学骨架化方法进行了比较，他们的结论认为MSD的性能稍微高一些。尽管形态学骨架化能够用于许多应用（例如图像编码）且被广泛地研究，但相比可保留连通性的2维或3维细化方法，仍不具备直接可比性。由于在边缘/曲面缝隙弥合的

应用中，我们坚持要求保留连通性，因此下面仅仅考虑了具备这项特性的后一类方法。

形态学二维细化方法

Arcelli等[7.1]采用一个掩模序列来实现细化。原始图像经由八个3×3掩模顺序进行腐蚀，其中每一步所得结果图像都作为下一步迭代的输入，直至所有可能的像素都被腐蚀过。图7-25[7.1，7.26]中所示每个掩模中，标注为黑色的位置代表边缘像素，而白色代表背景像素，斜线区域则代表计算中不用考虑进去的像素。如果一个掩模的中心点落在像素(k, l)上，而所有被掩模覆盖的图像像素中，边缘以及背景像素的分布模式，与相应掩模的模式相同，则称该掩模在坐标(k, l)上"匹配"该图像。如果一个掩模在一个特定边缘点上匹配一幅图像，则该点处的图像边缘重新设置为背景。按如下次序应用掩模：A1、B1、A2、B2、A3、B3、A4、B4。采用该算法对图7-23进行细化后的结果见图7-25。请注意该算法使一些尖锐顶点（如叉子齿）的形状发生了变形。 173

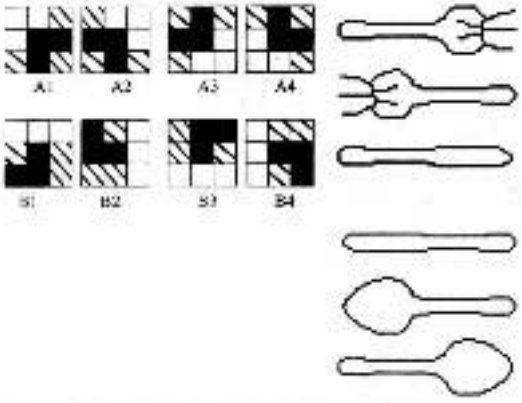


图7-25 3×3掩模（左图）以及餐具图像腐蚀细化后的结果（右图）。摘自[7.36]，已获得使用许可

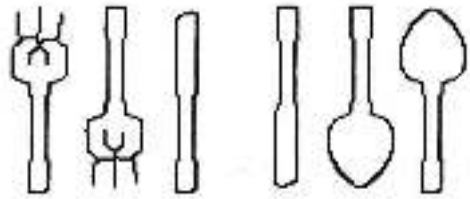


图7-26 迭代式细化后的结果。摘自[7.36]，已获得使用许可

迭代式二维细化

Zhang和Suen的二值图像细化算法[7.44]对图像进行迭代遍历，以确定是否有轮廓点可以删除。所谓轮廓点，是这样一个边缘像素，其8邻域中至少有一个像素是背景像素。每一步迭代都包含两次遍历，且在每一步迭代中，确定是否删除一个像素取决于像素邻域中边缘点的数量、一个序列中该像素四周发生0-1转换的次数以及描述邻域中背景构成的两个点集。膨胀后的餐具图像（图7-23）经由迭代式细化后的结果如图7-26所示。有趣的是，该方法与Arcelli等的方法非常相似。 174

7A.4.7 三维图像

将DT驱动的弥合算法应用于一幅三维图像，所得结果如图7-27~图7-29所示。图中是一个内部隔板断裂的方盒。结果图像显示的次序与显示二维例子结果的次序相同（请参考图7-22~图7-24的含义——译者注）。请注意，尽管x, y方向上的缝隙都很大，但算法成功地弥合了这些缝隙，且在最终的重标号图像中仍保持了边缘锐利。

三维算法也在合成的虚拟椭球体上（图7-30）进行了测试。该图因有意的欠采样而生成了较大的缝隙。令K_{max}值为3，应用DT驱动弥合算法，所得结果如图7-30所示。

a) b)

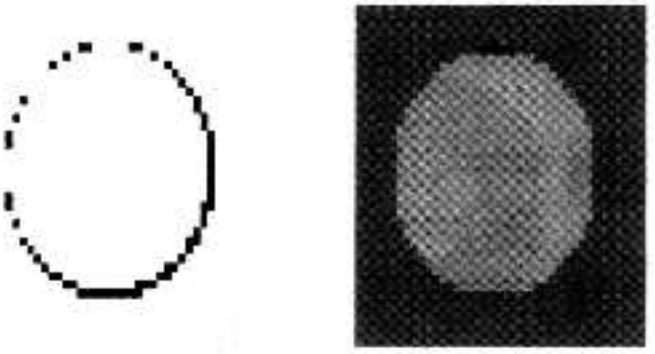
图7-27 a) 原始的断裂方盒图像，第20帧 b) 断裂方盒图像的距离变换

c) d)

图7-28 c) 加粗后的边缘图像 d) 对膨胀后所得结果的细化图像

e)

图7-29 e) 重标注后的断裂方盒图像



a) b)

图7-30 有缝隙的椭球体a)以及重标注后的椭球体b)

7A.4.8 与已发表的三维方法之间的比较

其他关于三维细化与骨架化方法的研究[7.11，7.19，7.41，7.42]主要是以计算每个三维实体的欧拉连通数N为基础[7.13]，$N=V-E+F$，这里V，E和F分别代表顶点个数、边缘个数以及对象的面片个数。当一个实体不包含隧道（tunnels）或者空穴（cavity）时，$N=2$。每有一个隧道或洞（hole）穿过对象时，N减2；对象中每有一个空穴，N加2。在细化过程中，必须要保持连通数不变，以保留原始对象的拓扑结构[7.26，7.33]。

三维细化

我们对DT驱动的细化方法与Tsao和Fu[7.42]所描述的拓扑三维细化算法进行了比较，后者能够保留路径和曲面的连通性，如果能够保留曲面连通性，则路径连通性或连通数也能够保留。将该算法应用于一个3×3×3立方体的中心体元素，在六步子迭代——N（orth），S（outh），E（ast），W（est），U（p）以及D（own）中的每一步迭代中，如果相应N，S，E，W，U或D邻域为零，则将它划分为一个边界点。如果删除一个体元素既不改变3×3×3窗口中剩余体元素的连通性，又不改变两个"校验平面"内的连通性，则该体元素可删除。采用Tsao和Fu的算法对加粗后的边缘图像——图7-28c细化后所得结果如图7-31所示。

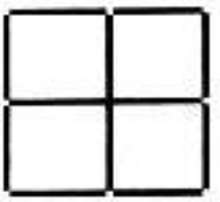
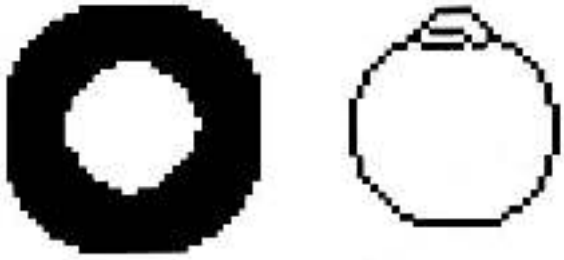


图7-31 对膨胀后的方盒经由Tsao和Fu方法细化后所得结果

对图7-30中的椭球体，仍采用5×5×5的核进行膨胀以弥合缝隙（图7-32a），然后应用Tsao和Fu的算法细化后的结果如图7-32b所示。图中顶部由于受了邻近平面的量化影响，而生成了额外的线条。

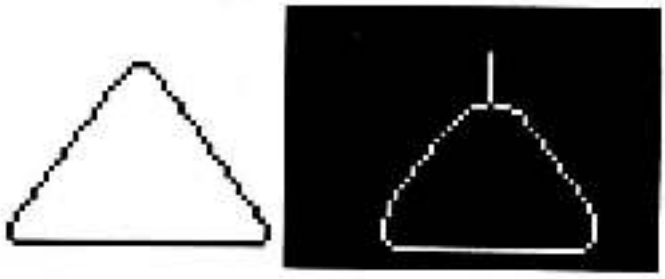


a) b)

图7-32 a) 膨胀后的椭球体 b) 采用Tsao和Fu算法细化后的椭球体。摘自[7.36]，已获得使用许可

7A.4.9 保留几何形状

在DT驱动弥合算法的性能中最重要的一点，大概就是能够保留曲面几何形状的能力，尤其是顶点附近的几何形状。该算法对二维图像的处理性能已在图7-24中给出了非常好的范例。为了演示该算法是如何处理三维顶点几何形状的，我们合成了一个圆锥，并采用一种三维边缘检测器抽取出它的表面。 176



a) b)

图7-33 a) 对圆锥完成基于DT的弥合方法处理后所得穿过圆锥顶点的一个横截面
b) 采用Tsao-Fu算法细化后的结果。摘自[7.36]，已获得使用许可

然后采用DT驱动弥合算法来弥合表面上的缝隙。采用传统膨胀算法对同样的圆锥曲面进行缝隙弥合，然后采用Tsao-Fu算法来细化，所得结果如图7-33所示。由于膨胀处理结果替代了原始边缘信息，因此采用Tsao-Fu算法对膨胀后的图像进行细化时，对初始表面没有任何"记忆"。当然，二者都不能完美地处理几何形状。然而，由于DT驱动弥合算法能够通过DT来保留关于原始的，未膨胀前的几何形状的"记忆"，因此该算法能更好地在弥合缝隙后恢复出原始几何形状。更多有关细节以及计算速度的描述见[7.38]。

7A.4.10 为什么写这一节

为什么我们在这一节以如此冗长的篇幅介绍DT驱动边缘弥合算法？有若干个原因：第一，实现任何一种实际技术都要求对多种相关技术有一定的了解，这是非常重要的。你们肯定会用得到连通分量、边缘检测、形态学膨胀等知识。第二，由于本书主要面向研究生，我们希望学生能够从这里了解到应该如何去写一篇期刊论文。本章在许多方面与[7.38]非常相似（且包含了很多从该文献摘录的图），请注意格式：介绍问题，描述算法，引证文献，以及最重要的：对新的技术与现有技术进行比较。只要遵循以上的简单规则，那么你也可以发表论文了！

7A.5 术语

切削映射	chamfer map
素因子	prime factor
采样	sampling

作业7.A1

在第7A.1.1节，给出了分解结构元素的一个例子。请证明这个分解所给出的结果与对图像{(0,0)}关于原始结构元素进行膨胀后的结果相同。

177

作业7.A2

细化算法的输出结果与距离变换的最大值之间的主要区别是什么？从下列选项中选择最佳答案。

a) 细化算法保留连通性。距离变换的最大值不必然是连通的。

b) 距离变换的最大值是唯一的，而细化算法生成的结果不唯一。

c) 距离变换保留连通性。细化算法不保留。

d) 细化算法能够生成亮度对称轴。距离变换则不能。

作业7.A3

采用距离变换来计算一幅图像的中轴线。具体图像将由教师在课堂中给出。

参考文献

[7.1] C. Arcelli, L. Cordella, and S. Levialdi, "Parallel Thinning of Binary Pictures," *Electronics Letters*, **11**, pp. 148–149, 1975.

[7.2] H.G. Barrow, "Parametric Correspondence and Chamfer Matching," *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, August, 1977, pp. 659–663.

[7.3] M. Bister, Y. Taeymans, and J. Cornelis, "Automated Segmentation of Cardiac MR Images," *Computers in Cardiology*, Washington, DC, IEEE Computer Society Press, pp. 215–218, 1989.

[7.4] G. Borgefors, "Distance Transformations in Arbitrary Dimensions," *Computer Graphics, Vision, and Image Processing*, **27**, pp. 321–345, 1984.

[7.5] G. Borgefors, "Distance Transformations in Digital Images," *Computer Graphics, Vision, and Image Processing*, **34**, pp. 344–371, 1986.

[7.6] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, "Linear Time Euclidian Distance Transform Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), 1995.

[7.7] D. Florencio and R. Schafer, "Homotopy and Critical Morphological Sampling," *Proceedings of the SPIE*, **2308**, June, 1994.

[7.8] J.D. Foley, A. vanDam, S.K. Feiner, and J.F. Hughes, *Computer Graphics: Principles and Practice*, Reading, MA, Addison-Wesley, pp. 91–99, 1990.

[7.9] W. Gong and G. Bertrand, "A Simple Parallel 3D Thinning Algorithm," 10th *Inter-*

national Conference on Pattern Recognition, June, 1990.

[7.10] B. R. Groshong, and W. E. Snyder, "Using Chamfer Maps to Segment Images," Technical Report CCSP-WP-86/11, Center for Communications and Signal Processing, North Carolina State University, Raleigh, NC, USA, December, 1986.

[7.11] K. Hafford and K. Preston Jr., "Three-dimensional Skeletonization of Elongated Solids," *Computer Vision, Graphics, and Image Processing*, **27**, pp. 78–91, 1984.

[7.12] R. Haralick and L. Shapiro, *Computer and Robot Vision*, Volume 1, New York, Addison-Wesley, 1992.

[7.13] D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*, New York, Chelsea, 1952.

[7.14] H. Hiriyannaiah, G. Bilbro, and W. Snyder, "Restoration of Locally Homogeneous Images using Mean Field Annealing," *Journal of the Optical Society of America, A*, **6**(12), pp. 1901–1912, 1989.

[7.15] C. Huang and O. Mitchell, "A Euclidian Distance Transform using Grayscale Morphology Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[7.16] X. Jiang, "An Adaptive Contour Closure Algorithm and its Experimental Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11), 2000.

[7.17] R. Jones and I. Svalbe, "Morphological Filtering as Template Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[7.18] R. Jones and I. Svalbe, "Algorithms for the Decomposition of Gray-scale Morphological Operations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), 1994.

[7.19] S. Lobregt, P. Verbeek, and F. Groen, "Three-Dimensional Skeletonization: Principle and Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2**(1), pp. 75–77, 1980.

[7.20] P. Maragos and R. Schafer, "Morphological Skeleton Representation and Coding of Binary Images," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **34**, pp. 1228–1244, 1986.

[7.21] G. Matheron, *Random Sets and Integral Geometry*, New York, Wiley, 1975.

[7.22] H. Park and R. Chin, "Optimal Decomposition of Convex Morphological Structuring Elements for 4-connected Parallel Array Processors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), 1994.

[7.23] H. Park and R. Chin, "Decomposition of Arbitrarily Shaped Morphological Structuring Elements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[7.24] I. Pitas and A. Venetsanopoulos, "Morphological Shape Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(1), 1990.

[7.25] Pratt, W. K., *Digital Image Processing*, New York, Wiley, 1978.

[7.26] K. Preston, and M. Duff, *Modern Cellular Automata Theory and Applications*, New York, Plenum Press, 1984.

[7.27] J. Reinhardt and W. Higgins, "Comparison Between the Morphological Skeleton and Morphological Shape Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(9), 1996.

[7.28] A. Rosenfeld, "Distance Functions on Digital Images," *Pattern Recognition*, **1**, pp. 33–61, 1968.

[7.29] A. Rosenfeld, *Picture Processing by Computer*, New York, Academic Press, 1969.

[7.30] A. Rosenfeld, "Connectivity in Digital Pictures," *Journal of the Association for Computing Machinery*, **17**(1), pp. 146–160, 1970.

[7.31] A. Rosenfeld and A. Kak, *Digital Picture Processing*, Volume 2, New York, Academic Press, 1982.

[7.32] A. Rosenfeld and J. Pfaltz, "Sequential Operations in Digital Picture Processing," *Journal of the Association for Computing Machinery*, **13**, pp. 471–494, 1968.

[7.33] P. Saha and B. Chaudhuri, "Detection of 3-D Simple Points for Topology Preserving Transformations with Application to Thinning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), 1994.

[7.34] D. Schonfeld, "Optimal Structuring Elements for the Morphological Pattern Restoration of Binary Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), 1994.

[7.35] W. Snyder and G. Bilbro, "Segmentation of Three Dimensional Range Images," CCSP-TR-84/7, North Carolina State University, November, 1984.

[7.36] W. Snyder and A. Cowart, "An Iterative Approach to Region Growing Using Associative Memories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5**(6), pp. 638–651, 1985.

[7.37] W. Snyder and C. Savage, "Content-Addressable Read-Write Memories for Image Analysis," *IEEE Transactions on Computers*, **31**(10), pp. 963–968, 1982.

[7.38] W. Snyder, M. Hsiao, K. Boone, T. Hudacko, and B. Groshong, "Closing Gaps in Edges and Surfaces," *Image and Vision Computing*, October, 1992.

[7.39] J. Srihari, K. Uhurpa, and M. Yau, "Understanding the Bin of Parts," IEEE Conference on Decision Control, pp. 44–49, December, 1979.

[7.40] H. Tagare, F. Vos, C. Jaffe, and J. Duncan, "Arrangement: A Spatial Relation Between Parts for Evaluating Similarity of Tomographic Section," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(9), 1995.

[7.41] J. Toriwaki and Y. Tatsuhiro, "Topological Properties and Topology- preserving Transformation of a Three-dimensional Binary Picture," *Proceedings of the Sixth International Pattern Recognition Conference*, pp. 414–419, Munich, 1982.

[7.42] Y. Tsao and K. Fu, "A Parallel Thinning Algorithm for 3-D Pictures," *Computer Graphics and Image Processing*, **17**, pp. 315–331, 1981.

[7.43] R. van den Boomgaard and A. Smeulders, "The Morphological Structure of Images: The Differential Equations of Morphological Scale-space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), 1994.

[7.44] T. Zhang and C. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns," *Communications of the Association for Computing Machinery*, **27**(3), pp. 236–239, 1984.

178
~
180

第8章 分　割

Galia est omnes diviso in partes tres.

Julius Caesar

分割是从背景中将目标分离出来的过程，它是所有后续过程，例如形状分析、目标识别等的基本模块。本章首先讨论若干常用分割算法，包括基于阈值的算法、基于区域的算法（或者叫连通分量分析算法）、基于边缘的算法、基于曲面的算法。在本章“专题”中，还介绍了一些最新的分割算法。

8.1 划分图像

在许多机器视觉的应用中，场景中可能出现的物体集合是非常有限的。例如，如果摄像机正在察看一个传送带，传送带上可能只出现一种类型的零件。此时的视觉任务就可能是确定零件的位置和方向。在其他应用中，所察看到的零件可能是若干零件类型中的一种，此时视觉任务不仅要识别类型还要定位每个零件。最后，摄像机也可用于检验零件，以便进行质量管理。

图8-1　具有两个前景区域的图像

在本节中假定零件非常简单，从而用单目摄像机提供的二维投影即可表现出其特征。此外，还假定形状本身足以刻画这些目标的特征，即不需要颜色或者亮度变化信息。首先考虑将图像划分为几个连通区域的情况。

图8-2　对图8-1所示图像的一种分割及标号结果

对图像的一种分割结果是将它划分为几个连通区域，其中每个区域在某种意义上都是均匀的，且由一个唯一的标号来标识。例如，图8-2是一幅“标号图像”（label image），区域1被确定为背景，尽管区域4也是背景，但由于它与区域1之间不连通，因此被标注为一个单独的区域。

“均匀”一词值得讨论。它的含义可以是所有像素的亮度相同，但是对于大多数实际应用而言，这个标准过强。它的含义也可以是所有像素的亮度接近于某个代表（平均）值。更加形式化的描述[8.80]是：如果一个区域的亮度值分布服从于一个特定的概率分布，则说这个区域是均匀的（也可参考[8.44]中的分析）。在距离图像[8.25]中，我们可以用一个方程来描述曲面，此时如果一个区域可以用结合该方程的某个概率退化模型来描述，则可以说这个区域是均匀的。例如，如果一幅距离图像的一个区域中所有点，除了偏离点（deviation）之外均位于同一个平面，而偏离点又可以由一个特定的高斯分布来描述，此时可以说该区域是均匀的。

实现分割的方式有很多种。基于阈值的技术可以保证将图像分割为闭合区域，因为这种技术仅仅将所有高于（或者低于，取决于具体问题）某个特定阈值的像素指定为同一个区域。基于边缘的技术假定区域之间可由一个相邻的边缘强度高的区域来隔离。基于区域的方法从单元（如，均匀的）区域开始，对它们进行分裂或归并（split or merge）。于是，又有多种混

合方法，其中包括分水岭（watershed）[8.5]方法。分水岭方法通常对图像的梯度进行运算；分割过程就是在图像中进行洪泛（flooding），在洪泛过程中，矗立的区域边界（高边缘强度区域）会阻止从不同种子点（seed point）出发的水混合在一起。传统的“区域增长”方法其实是分水岭方法的一种变种[8.1]。

在进一步讨论有关分割的论题与技术之前，需要了解在处理数字图像时会遇到的一些有关几何学和拓扑的有趣和意料之外的问题。在第4.5节中，我们讨论了有关连通性的自相矛盾的论点。我们发现一个物体可以有一个闭合边界但仍可以具有连通的内界与外界。还有另一个例子，就是如何估计一个区域的周长的问题，或者说是一个如何按它的采样表示估计一条线的长度的问题。这个问题在图像分割中有直接的应用，但仍不清楚如何去进行估计[8.31]。由于本章下面还要继续讨论许许多多有关连通性的话题，以上这些问题都是很可能会遇到的。

8.2 阈值化分割

在一些应用中，区域的具体灰度值并不重要，此时就可以通过选择亮度的阈值，将一幅图像分割为“对象”和“背景”。定义亮度高于阈值的任意一个区域为对象，而低于阈值的区域为背景。

选择阈值的方式有很多种，从极简单到非常复杂。随着方法复杂性的增加，性能会随之提高，但付出的代价是计算复杂度也会同时增加。

182

需要注意的最重要的因素大概是阈值化的局部特性，即很难有单个阈值适用于整个场景。而一般都是对象与背景之间的局部对比度，包含了重要信息。由于摄像机的敏感度受了如图8-3所示的抛物线畸变和/或格晕（vignetting）的影响，从图像中心到边缘逐步下降，因此试图建立一个全局阈值通常是无益的。这类效果的显著例子可从直线栅格图像中看出，“均匀的”白光在其表面上发生了巨大的变化。

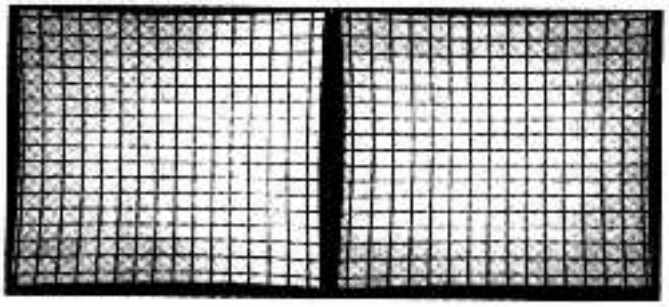


图8-3 两个探测器组成的一幅直线栅格图像。光源是均匀的，但是，图像显示出了两种不同的畸变：辐射度（亮度）畸变（左图中右侧中心位置的亮度较高，而右图中中心位置的亮度较高）与几何畸变（直线以典型的正畸变形式弯曲）

类似抛物线畸变与格晕的效果是可预测的且是易于校正的。事实上，现有的（off-the-shelf）硬件刚好可用于这样的应用中。不过，对于非均匀的环境光照，例如从窗口射入的阳光在一天时间内会发生根本的变化，所产生的影响是更加难以进行预测并校正的。

由于单个阈值不能提供足够的性能，因此必须选择局部阈值。最常用的方法就是所谓的分块阈值化（block thresholding）方法，这种方法将图像划分为矩形块并为不同的矩形块采用

不同的阈值。512 × 512图像中采用的典型的矩形块大小为32 × 32或者64 × 64。首先对每个矩形块进行分析以选择阈值；然后使用分析所得阈值对图像中的矩形块进行阈值化。在更加复杂的（但更慢）的分块阈值化方法中，对每个分块分析所得阈值，仅用于该分块中心处的单个像素。然后方块移动一个像素，再重复整个过程。 183

选择阈值

选择阈值时，可采用的最简单的策略是对分块中的亮度取平均值，并选择$i_{avg}+\Delta i$作为阈值，其中Δi 是某个小的增量，例如在256个灰度级别中取5。这样一个选取阈值的简单方案却可以给出好得令人惊讶的结果。

然而，当简单一些的方案失败时，不得不采用更复杂的技术，例如基于直方图分析的阈值选取法。在描述这个方法之前，应先定义直方图。

一幅图像$f(x, y)$的*直方图*$h(i)$是一个定义在规定亮度值范围上的函数。在一个典型的成像系统中，亮度取值范围在0X00（黑）和0XFF（白）之间。一个表示每个灰度级别在图像中出现的次数的曲线图，称作一幅图像的直方图。图8-4所示直方图所对应的图像,描述的是在白色传送带上传输一些黑色零件的场景。

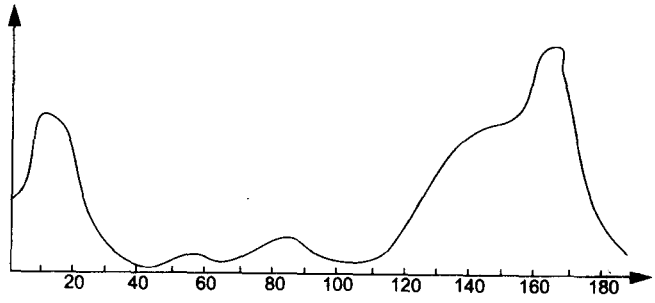


图8-4　一幅双峰图像的直方图，图像中包含了非常多很亮的像素（亮度在169左右）和很暗的像素（亮度在11左右）

注意，在图8-4中有两个明显的峰，一个处于灰度级别为11的位置上，几乎是纯黑色，另一个在灰度级别为169的位置，亮白色。除噪声像素之外，图像中的每个点都分别属于这两个区域中的一个。那么，在这两个峰值区域之间的任何一个点都可以是一个适当的阈值。

直方图很少如图8-4所示的这样“好”，因此通常需要一些附加的处理（参考文献[8.14, 8.51, 8.70]中解释并试验比较了若干种这样的方法）。下一节将描述一种寻找最优阈值的更为复杂的方法。

多高斯拟合（Fitting a sum of Gaussians）

参考文献[8.7]中描述了一种可以找到一个多变量函数的全局最小值的方法，即使该函数有多个局部极小值。这是一种称为*树退火*（tree annealing, TA）的方法，可按以下方式[8.62]应用于直方图分析和阈值化问题中。 184

给定一个双峰直方图，$h(f)$，其中f 表示亮度值，采用两个高斯的和来拟合直方图是找到最优阈值的一种标准方法[8.14]:

$$h(x)=\frac{A_1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right]+\frac{A_2}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right] \tag{8-1}$$

如果对$h(f)$进行了恰当的归一化，就可以实现分别调整两个高斯分量的归一化，使它们每一个在256个离散灰度级别上的总和都为一（而不是在区间上的连续积分为一），且满足附加的约束条件$A_1+A_2=1$。使用这个约束条件，可将待估计参数的个数从六个减少至五个；不过，我们通过实验发现,采用六个参数的TA方法可以更准确地解决问题，并且无需在每一步迭代中都重新调整归一化。传统的下降方法在面对双高斯问题时，经常在一个次优局部最小值停止，对三高斯问题则更不可靠。TA可以很轻易地处理这两种情况。采用三高斯拟合一幅图像的直方图的结果如图8-5所示。

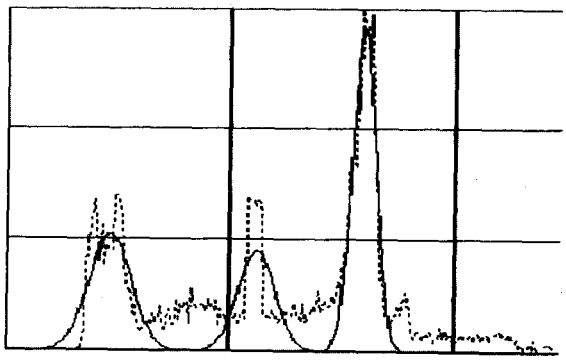


图8-5 采用三高斯对直方图进行最小均方误差（MSE）拟合

无论采用什么算法，基于直方图的阈值化方法的原则都是一样的：寻找直方图中的峰值点，然后在峰值中间选择阈值。

在许多工业环境中，光照可以很好地控制。在这样的控制下，最优阈值是不随时间变化的常量，可在系统建立过程中以交互方式选取。不过，通常对图像的不同区域会采用不同的阈值。

8.3 连通分量分析

现在假定，已选取了一个适当的阈值，且图像已划分为纯黑区域和纯白区域，如图8-1所示。但是若要生成如图8-2所示分割图像，则需要进行连通性分析。即，如果一个像素的亮度值高于阈值，且与区域i中的一个像素相邻，则认为该像素在区域i中。由于区域可能是弯曲的或者是叉形的，分析过程不会像从顶端出发然后逐步向下标注相邻像素那样简单，而是需要一种更为复杂的技术。

185

“连通分量”一词源自图论。考虑图8-6所示的图。

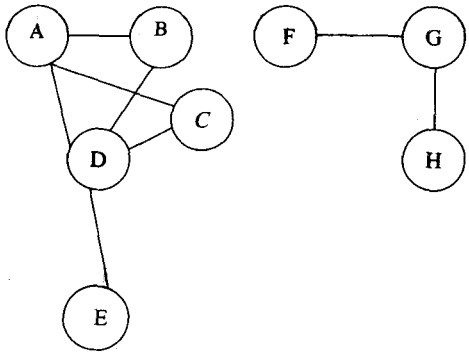


图8-6 由两个连通成分组成的图

图中有八个顶点、八条边以及两个连通分量。即从A到D、从B到E、从F到H等，都存在一条沿着边的路径可到达。但从A到F却没有可行的路径。在图像分析中，将前景像素看作是图中的顶点，并把“相邻”特性表达为图中的边。在这个定义下，可以看出图8-7由两个连通分量组成。第12章将更详细地介绍这些图。

图8-7　一幅由两个连通分量组成的图像

为了使用连通分量标号（connected component labeling，CCL）这一概念，必须也定义一个“标号图像”，它是一种用符号对图形进行表示的方法，与原始图像同构，但每个像素都包含了它所属分量的序号。

生成标号图像的一种算法称为“区域增长”。它使用标号存储器，并与原始图像的帧缓存相对应，就像图8-7与图8-8相对应一样。在这种描述下，称“黑色”像素为对象，而“白色”像素为背景。

图8-8　与图8-7相应的标号图像

初始时，标号存储器L中的每个单元都设为零。我们称这个图像的帧缓存为f。因此对某个标记序号为N的标注操作可写为$L(x, y) \leftarrow N$。

8.3.1　递归式区域增长算法

该算法采用了一个下推堆栈来实现区域增长。下推堆栈可临时保留在区域中有关像素的坐标。

186

1）找到一个未标号（即$L(x, y)=0$）的黑色像素，为这个区域选取新的标号序号，用N表示。如果所有的像素都已被标号，算法停止。

2）$L(x, y) \leftarrow N$。

3）如果$f(x-1, y)$是黑色的，且$L(x-1, y)=0$，将坐标$(x-1, y)$压入堆栈。
　如果$f(x+1, y)$是黑色的，且$L(x+1, y)=0$，将$(x+1, y)$压入堆栈。
　如果$f(x, y-1)$是黑色的，且$L(x, y-1)=0$，将$(x, y-1)$压入堆栈。
　如果$f(x, y+1)$是黑色的，且$L(x, y+1)=0$，将$(x, y+1)$压入堆栈。

4）从堆栈中取出一个值作为新的(x, y)。

5）如果堆栈为空，跳转至步骤1，否则跳转至步骤2）。

标号操作的结果可得到一组连通区域，每个区域都分配有一个唯一的标号序号。对任一给定像素，如果想要找到其所属区域，计算机只需访问该像素在L存储器的相应位置，并读出区域序号即可。

范例：应用区域增长算法

图8-9给出了一个4×7的像素组。假定<x, y>的初始值为<2, 4>，应用“增长”算法，并给出每执行完一次步骤3）后的堆栈和L的内容。令N的初始值为1。

解

第1轮　刚执行完步骤3）后的情况。算法查看前景像素<2, 4>，以及它的4邻域，并检测到它的4邻域中只有一个像素<3, 4>属于前景。于是，将这一像素的坐标压入堆栈顶部。

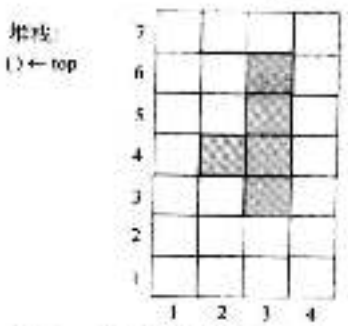


图8-9　一幅例图阈值化后的结果

187

堆栈:

$\boxed{\langle 3,4\rangle}$ ← top

$$L=\begin{array}{c}7\\6\\5\\4\\3\\2\\1\end{array}\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\begin{array}{c}\\\\\\\\\\\\\\1\ \ 2\ \ 3\ \ 4\end{array}$$

第2轮 从堆栈顶部移开像素<3, 4>，并将L图像中的对应位置标记为1，查看它的邻域，在4邻域中找到了两个像素<3, 3>和<3, 5>；将二者都压入堆栈。

堆栈:

$\boxed{\langle 3,5\rangle}$ ← top

$\boxed{\langle 3,3\rangle}$

$$L=\begin{array}{c}7\\6\\5\\4\\3\\2\\1\end{array}\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&1&1&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\begin{array}{c}\\\\\\\\\\\\\\1\ \ 2\ \ 3\ \ 4\end{array}$$

第3轮 堆栈顶部为<3, 5>。从堆栈中移开这一像素，并将L图像中的对应位置标记为1。查看它的所有邻接点，并在4邻域中检测到一个像素<3, 6>。于是，将这一像素的坐标压入堆栈顶部。

堆栈:

$\boxed{\langle 3,6\rangle}$ ← top

$\boxed{\langle 3,3\rangle}$

$$L=\begin{array}{c}7\\6\\5\\4\\3\\2\\1\end{array}\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&1&0\\0&1&1&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\begin{array}{c}\\\\\\\\\\\\\\1\ \ 2\ \ 3\ \ 4\end{array}$$

188

第4轮 再次“出”栈，这一次去除的是<3, 6>，并将L图像中的对应位置标记为1。查看这个像素，在4邻域中未发现没有标号过的像素。

堆栈:

$\boxed{\langle 3,3\rangle}$ ← top

$$L=\begin{array}{c}7\\6\\5\\4\\3\\2\\1\end{array}\begin{bmatrix}0&0&0&0\\0&0&1&0\\0&0&1&0\\0&1&1&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\begin{array}{c}\\\\\\\\\\\\\\1\ \ 2\ \ 3\ \ 4\end{array}$$

第5轮 再次出栈，移开<3, 3>，并将L图像中的对应位置标记为1。查看这个像素，在4邻域中未发现没有标记过的像素。

堆栈:

() ← top

$$L=\begin{array}{c}7\\6\\5\\4\\3\\2\\1\end{array}\begin{bmatrix}0&0&0&0\\0&0&1&0\\0&0&1&0\\0&1&1&0\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\begin{array}{c}\\\\\\\\\\\\\\1\ \ 2\ \ 3\ \ 4\end{array}$$

第6轮　再次出栈，返回值为“堆栈为空”，则算法结束，而所有的黑色像素均已标号。

这个区域增长算法仅仅是执行连通分量分析的若干策略之一。还有其他的策略，比上文描述的方法在速度上还要快，其中包括一些可按光栅扫描的速度运行的策略[8.6]。现在就来看看一种这样的技术。

8.3.2 迭代式连通分量分析方法

由于区域增长方法的结果总是能够得到闭合区域，因此更适用于其他的基于边缘检测或者线拟合的技术。关于基本的区域增长方法的许多变种及其应用可参见[8.10，8.16]。尽管区域增长方法被证明是场景分析中的主要方法，但是它的应用会迅速成为计算瓶颈，尤其是对高分辨率图像。这促使人们考虑有关区域分割的另一些更快更加硬件化的方法。

这一节将介绍递归式区域增长的一种替代算法。这个算法在功能上等同于递归式区域增长方法，同样返回一组标注好的像素，且满足相邻性以及相似性等标准。 189

尽管这个算法在功能上与传统的区域增长方法相同，但在概念上以及可能的实现方法上都有根本的区别。

设计这个算法的目的是寻找一种方式，可达到与区域增长同样的结果，但同时期望能够通过单向遍历数据来实现。采用内容可寻址的存储器的概念，可以达到这样的结果。这个存储器可以是硬件的一个物理单元，也可以是模拟软件中一种查找表驱动的访问方法。

该算法以图像中像素之间的等价关系的概念为基础。等价性的定义如下：两个像素a和b被定义为是等价的（记为$R(a, b)$），如果它们都属于图像中的同一区域。在这样的定义下，这一关系表现出了自反性（$R(a, a)$），对称性（$R(a, b)\Rightarrow R(b, a)$），以及传递性（$R(a, b)\text{AND}R(b, c)\Rightarrow R(a, c)$），因此可称之为“等价关系”。

传递特性使一个区域内的所有像素可仅依据局部相邻特性来确定。在算法中，每个像素按自左向右，自顶向下的光栅扫描方式与其每个相邻像素进行比较。这些比较操作的结果是为像素分配一个区域标号。对简单的二值图像中的像素，可按光栅扫描次序标号。区域标号过程简单直接：首先查看左侧像素，将它看作是邻近像素并采用它的标号。再查看位于上方的像素，将它看作邻近像素。如果二者均已标号且标号不同，则有一个问题。图8-10给出了这样比较的结果可能会产生的情形，此时在标记为问号的像素处可发现等价关系$R(1, 2)$。

```
111  22
111  22
111  22
111  22
111 11?
```

图8-10　标号分配中的歧义性

在[8.60]提出的系统中，使用硬件为像素分配区域标号，并维护一个等价关系表。图8-12中显示该硬件位于图像存储器与一台主机之间。功能上，该硬件对主机是透明的。在图8-10所示例子中，主机可感知到所有的属于区域1的像素（这里等价关系中次序较低的区域标号优先）。

关于硬件操作流程的描述见图8-11。为了理解这个流程框图，引入如下的符号：

- $f(x, y)$是像素(x, y)的灰度值，存储于图像存储器中。
- $(x, y)_i$是像素(x, y)的第i个邻接点。
- $f_i(x, y)$是像素(x, y)的第i个邻接点的灰度值。
- $L(x, y)$是与像素(x, y)相应的区域标号，存储于图像存储器中。
- $L_i(x, y)$是与像素(x, y)的第i个邻接点相应的区域标号。
- $K(i)$是等价存储器中第i个元素的内容。这个存储器是内容可寻址存储器。 190
- $K^*(i)=K(j)$意味着如下序列：

- 读取$K(j)$。
- 以i为搜索线索来搜索K（也就是确定所有的l，使得$K(l)=i$）。
- 将$K(j)$写入搜索得到的所有肯定响应者中。

- T是阈值。
- p是区序号最高的区域标号（初始值为1）。
- N是邻域中的像素个数（4或8连通）。

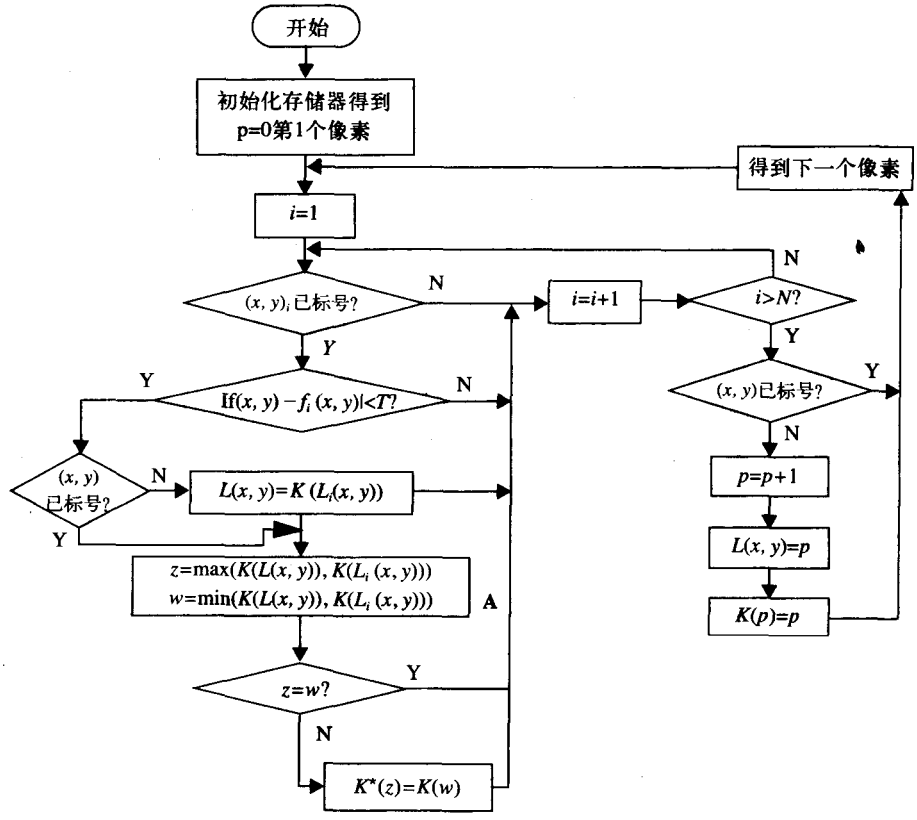


图8-11 算法的流程框图（摘自[8.60]）

图8-13显示的是，该算法在从Milgram等[8.41]任意选出的某个区域上运行的结果。这里注意，关系式$|f(x, y)-f_i(x, y)|<T$用来检验两个像素是否相似。还有其他可用的相似性度量，包括局部一阶和/或二阶统计量。如果两个像素符合这个标准且二者相邻，则它们在同一区域内。根据定义，如果两个像素在同一区域内，则有$R(a, b)$。即$\{(\langle x,y\rangle,\langle x',y'\rangle$相邻$)\wedge|f(x,y)-f_i(x,y)|<T\}\leftrightarrow R(\langle x,y\rangle,\langle x',y'\rangle)$

191

如果仅有R的传递性，而不考虑相邻特性，并不能推断出$R(\langle x,y\rangle,\langle x',y'\rangle)\rightarrow|f(x,y)-f_i(x,y)|<T$。

随着区域分割的实时（即与光栅扫描同步）进行，有两个操作必须执行。第一，每个在考虑中的像素的区域标号序号必须载入L存储器，第二，必须用已发现的所有等价关系来更新存储器。例如，如果区域4实际上与区域2相同，则$K(2)$和$K(4)$的内容均为2（次序更低的区域标记优先）。因此，当主机访问L存储器中的像素(x, y)时，界面/处理器根据存储器解释$L(x, y)$

并向主机返回$K(L(x, y))$。例如，图8-13中的像素(11, 11)将返回成属于区域1，这是因为$K(L(11, 11))=K(4)=1$。

在此类程序中常常会遇到的困难是链接问题。即如果$R(2, 4)$和$R(3, 4)$已经确定，则$R(2, 3)$也必然可以推断出来。然而，在图像处理结束后要求机器搜索出所有此类可能情况，会使得在扫描期间执行区域分割的最初目标受挫。这里，算法通过确保$K(2)=K(3)=K(4)=2$，来避免链接。不过，这仅仅是将链接问题转移给了扫描和标号过程而已。

算法流程图8-11中的模块A解决的是链接问题。一旦检测到等价关系，K存储器中所有包含较大区域标签序号的位置就要重新载入更小的区域标号序号。对于传统的随机访问存储器而言，实时完成这一步是不太可行的，不过它正属于内容可寻址存储器的一项功能，关于后者将在下一节进行讨论。

执行体系结构

用于执行该算法的体系结构见图8-12。我们打算将该硬件作为现有基于计算机的图像处理系统中的专用处理器。这个体系结构主要包括四部分：图像存储器(f)，区域标号存储器(L)，等价存储器(K)，以及一个界面/处理器。图像的灰度值存储于图像存储器中。一般情况下，f存储器包含512 × 512个字节。为每个像素分配的区域标号存储于区域标号存储器中。但是，L存储器的内容还包括所有作为媒介的区域标号，等价关系正是根据它们确定的。L存储器的内容通过采用简单直接的表查找硬件，根据等价存储器的内容进行解释。L存储器的大小与用来表达每个像素的区域标号（包括媒介区域标号）所需的比特长度直接相关。 192

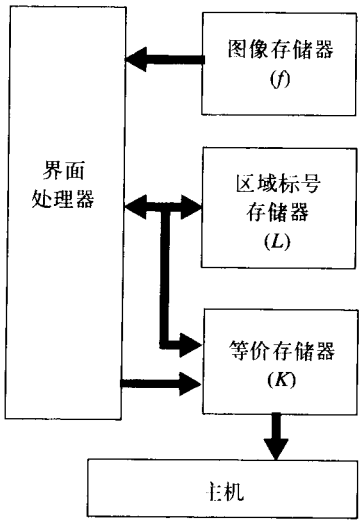


图8-12　一个区域标号系统的体系结构

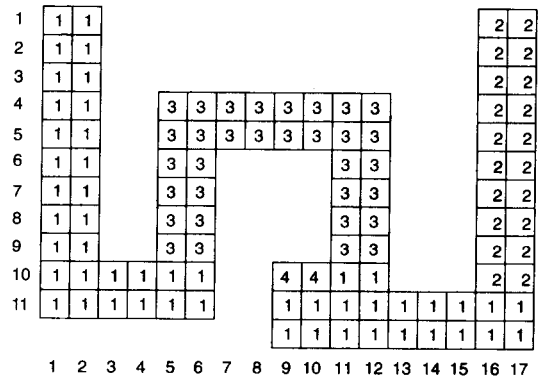


图8-13　一个比较困难的标号问题的例子（摘自[8.60]）

图8-13举例说明了一个较困难的标号问题，表8-1解释了对这个例子的标号过程。

f存储器和L存储器均为传统的随机访问存储器。但是，等价存储器有两种操作模式。它可以是一个传统的RAM（随机访问存储器）模式，输入地址与区域标号表相对应，而输出数据是等价关系表。它也可以在副存储器模式下，用于更新等价关系表。这一模式下，有两种操作在双相时钟下可同时发生。 193

表8-1 标号算法进行过程中等价存储器的内容

像素		K存储器地址			
x	y	1	2	3	4
1	1	1			
16	1	1	2		
5	4	1	2	3	
5	10	1	2	1	
9	10	1	2	1	4
11	10	1	2	1	1
16	11	1	1	1	1

第一相位：内容与数据总线的内容相匹配的所有存储器单元，均设置为允许触发（flip-flop）（见图8-14）。

第二相位：所有允许触发的存储器单元都设置为可读取数据总线内容。

通过这项操作，可在扫描过程中有效地并行更新等价表。

在扫描结束时，通过一个搜索算法来更新等价关系表，也是可以的。但是，在扫描期间并行地更新可以极大地减少表中的等价关系个数（因为总是使用最低的序号），从而可以减少K存储器所需的比特数。内容可寻址存储器具有这样的特性：存储器单元可以依据其内容被访问或加载[8.45，8.69，8.77]。

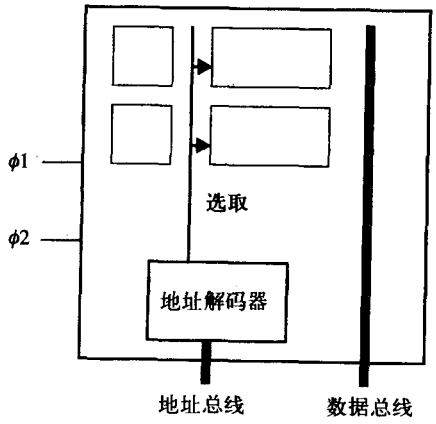


图8-14 K存储器的组织结构

在设计一个令人满意的存储器时至关重要的参数，同时也是令系统能够实时操作的重要因素，是存储器的大小。如果能够较好地权衡存储器大小与访问速度之间的矛盾，则会得到一个接近实时的系统。有关存储器大小的问题在[8.60]中作了深入的讨论，其中还对引入了真实图像的模拟进行了描述。

体系结构的最后一部分是界面/处理器。这一单元的主要目的是执行本节以及流程框图8-11中介绍的算法。另外，它必须能够1）对视频信号输入进行处理，将其灰度值存入存储器中，以及2）依据K存储器来解释L存储器。

仿真

将算法应用至一个512 × 512的文本图像，图像在分割之前已进行了阈值化。有两个参数是我们感兴趣的：1）单元区域（这些区域的标号存储于L中）的个数，它影响L和K的字宽以及K的长度；2）算法所觉察的区域个数，它决定了从图像中收集到有用信息之前主机必须执行进一步处理的量。

仿真结果总结如下：

- 4邻域连通。912个基本区域，138个觉察区域。
- 8邻域连通。883个基本区域，109个觉察区域。

从这些结果可以看出一幅图像需要512 × 512 × 10比特L存储器以及一个1024 × 10比特存储器。

本节讨论的是对电视扫描数据进行实时的图像分析操作。前文已表明，按这种方式能够设计出硬件，来执行区域增长操作。使用等价关系来分割一幅输入图像的概念是这个算法的基础。此外，采用内容可寻址的读/写存储器促进了这样的等价关系处理过程的实时实现。 195

这些概念是通过考虑将来可能的硬件结构而发展出来的；然而，算法并不妨碍它在数字计算机上的实现。这里描述的程序用来仿真这个方法的有效性，并被采用于在图像分割中标号区域。在达到同样的性能时，它的操作速度，即使是在仿真中，仍然优于早期的区域增长算法。

8.3.3 标号图像的一种替代方法

在本章，我们主要采用了标号图像作为辨识像素所属区域的手段，另一些研究者[8.73]采用了一种基于树的方法。即建立一幅图，对给定的父亲，假定其所有儿子属于同一个区域。因而，这幅图中的层级与尺度级别相对应。即一个“父节点”相比子节点处于更高一层（更模糊）级别的尺度。如果1）一个子节点与父节点在几何上相邻2）二者亮度值相似，则定义该子节点与其父节点相关。更多相关细节可参见[8.73]。

8.4 曲线分割

有时区域的边界已知，而我们对以某种方式描述这个边界感兴趣，这种描述应该或者是适合于表征整个边界或者适合于表征个别片断。为此，有许多实现方法 [8.17，8.42，8.52]，但几乎所有这些方法都需要识别边界上的特异点[8.15，8.26，8.55，8.79]，然后再描述这些“特异”点之间的曲线。显然，如何定义特异性 [8.18，8.23，8.48]对算法性能是非常重要的。

首先，我们知道曲线就是在二维空间中弯曲的一维函数。即曲线可以用单个参数的参数化表示，这个参数通常是弧长。然而，采用弧长的参数化表示对仿射变换并不是不变的[9.1]。为此，Rivlin和Weiss [8.49]设计了不采用任何参数的曲线不变量。曲线上一点的$x-y$坐标可写为关于弧长s的一维函数，$\psi(s)=[x(s), y(s)]$，s也就是沿着曲线前进的距离。如果曲率作为弧长的函数是已知的，则可以任意建构一条平滑曲线，在一个刚体运动条件下是等价的[8.76]。当然，我们处理的曲线不可能是一条以弧长为参数的平滑曲线，因为我们只能得到对这样的曲线进行采样的结果。而且，真实弧长与曲线通过的像素个数之间的关系并不像想像中那样简单[8.31]。事实上，即使对于无噪声的曲线，数字化过程也必然会引入误差[8.76]。有趣的是，附加的噪声，可能会令我们更准确地估计曲线（至少是直线的）所在的位置 [8.37]。

曲线在点s处的速率为

$$\dot{\psi}(s)=\sqrt{\left(\frac{\partial x}{\partial s}\right)^2+\left(\frac{\partial y}{\partial s}\right)^2}$$ 196

曲线在点s处的外法向为

$$n_\psi(s)=\frac{\left(\frac{\partial}{\partial s}y(s)-\frac{\partial}{\partial s}x(s)\right)}{\dot{\psi}(s)}$$

假设曲线是闭合的，则内部（INSIDE）和外部（OUTSIDE）的概念是有意义的。给定平面上不在当前曲线上的一点$x=[x_i, y_i]$，令ψ_x代表曲线上与x距离最近的点（在这一点上，弧长定义为s_x）。如果$[x-\psi_x]n_\psi(s_x)<0$，则称x属于内部，否则称x属于外部。

在执行曲线演变时（见第9.8节），有一种方式[8.54]可以使闭合区域的面积始终保持不变。

还有一种方法不要求必须找到特异点。Chen [8.13]仅仅是将一个方向选择滤波器应用于所有可能的方向，如果两个曲线片段具有非常不同的方向，则滤波器的响应结果会具有多个峰值，而峰值的位置可用于辨识这两个片断。这个方法似乎很适合于由X或T型交叉点线（见第10章中关于交叉类型的讨论）构成的图像。

Rosen和West [8.50]提出了一个稍稍不同的寻找特异点的策略。他们利用任何一种看起来可能适合的函数（椭圆或直线），来拟合数据点序列。拟合结果最差的数据点被认为是特异点，然后将曲线分割为两个片段，接着对每一片断再递归地重复拟合过程。

8.5 主动轮廓线

主动轮廓线（蛇形线）的概念最初用于解决这样一个现实问题：任何边缘检测算法都会对某些图像失效，这是因为在图像的某些区域中，边缘完全不存在。例如，图8-15是在核医学中拍摄的一个人的心脏图像。放射性药物被送入循环系统，图像中的每一点反映的是该点上的辐射强度。某一点上的亮度值，也就是在与成像平面正交的方向上，对该像素所覆盖的范围内血的总量的观测量。心室中的血量因此可以通过对整个心室区域内亮度的总和来计算。当然，这要求正确地分割出心室的边界，但是如下的事实又带来了一个难题，即心室左上角处根本没有亮度差异。这是因为从其他位于心室之后的源头（较高或较低静脉腔，等等）出来的放射线模糊了边界上的亮度差异。因此，需要这样一个技术，它可以搭接这些（边界片断之间）相当大的缺陷——这些缺陷太大，大到无法采用第7章所描述的闭运算来接合。

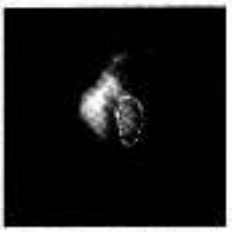


图8-15 核医学中拍摄的左心室图像

[197]

根据主动轮廓线的原理，首先要初始化一个轮廓，初始轮廓可以是用户定义的也可以是自动生成的，然后，移动边界，直至许多/大多数轮廓线与图像的边缘点相匹配。由于轮廓在搜索边界点时的移动过程，会使观察者联想起蛇的行为，因此这些边界通常被称作"蛇形线"。

推导蛇形线算法时，可遵循两个原则：能量最小化原则和偏微分方程原则。

8.5.1 能量最小化原则

边界的移动会减弱能量：$E=E_i+E_e$，其中内能（E_i）表征曲线本身，外能（E_e）表征轮廓线当前所在位置附近的图像区域。

内能量测的是沿着曲线方向弯曲的程度，以及曲线有多长等等。

内能的具体形式取决于具体应用，常用形式如下：

$$E_i=\sum \alpha\left|X_i-X_1\right|+\beta\left|X_{i+1}-2X_i-X_{i-1}\right|$$

其中，$X_i=[x_i\ y_i]^T$是轮廓线上的点。最小化第一项，会使曲线上的点相互靠拢。最小化第二项，会使曲线几乎不弯曲。第二项可能会产生一个消极影响，即最小化这一项的结果会使轮廓线退化至单个点。因此，在许多应用中还引入了一个"扩展"项，以使整个曲线增大。

外能量测的是轮廓边界穿越的区域是否为边缘。这里也有许多函数可以采用，我们最喜

欢采用的是

$$E_E = \sum \exp(-\|\nabla f(X_i)\|)$$

对二维空间内的最小化问题，可以通过动态规划来解决[9.19]。但是，除了度量位于边界上的局部边缘性，也可利用内部与外部之间的平均对比度差异[9.55]的方法，但该方法仅在外部区域相对均匀时是有意义的。

MAP(最大后验概率)方法完全适用于这个问题，且可以使用模拟退火（SA）[9.78]。但是，搜索邻域时会遇到问题。正如我们已经讨论的，SA可以保证从一组状态中，找到全局最小的状态。但是，这组状态必须按SA要求的次序进行采样，算法才能够起作用。在参考文献[9.78]中，用一个现有轮廓线作为初始状态，在每次迭代中，只对那些距离当前轮廓一个像素范围之内的轮廓进行采样，并从中选取一个最小值。最终轮廓是采样所得一组轮廓中最优的那个，但并不一定是所感兴趣的整个区域中最优的那一个。 198

所选择的能量函数的形式应当对尺度、平移以及旋转是不变量，这一点是非常重要的。为此，可采用的一个办法是使用两条蛇形线，并赋予适当的权重，一个在假定的边界之外往里收缩，另一个初始时在边界之内然后逐步扩大[9.25]。

8.5.2 偏微分方程方法

一条轮廓线C的移动可以描述为$dC_i/dt=s(X_i)N(X_i)$，其中$N(X_i)$是轮廓线上第i个点的法向量，C_i是轮廓线上一点，s是速度。左式代表轮廓线上第i个边界点的移动。右式描述的则是边界沿着法向量的移动。点C_i的“移动”实际上包括了该点x，y坐标的变化。

速度s，是两个函数的乘积

$$s(x,y) = s_I(x,y)s_E(x,y) \tag{8-2}$$

其中 $s_I(x,y) = \pm 1 - \varepsilon\kappa(x,y), s_E(x,y) = 1/(1+\Delta(x,y))$ 。$\Delta(x, y)$是对图像中点x，y处的“边缘”的观测量，$\kappa(x,y)$则是对点x，y处的轮廓线曲率的观测量。

Manhaeghe等采用Kohonen映射[9.47]得到了一个类似主动轮廓线的结果；也可参见[9.88]。这个方法的一个好处是，计算具有局部性。只需查看当前边界上的一点，并考虑该点应如何移动。然后选择一个候选位置并确定移动到该位置时能量是增加的还是减少的（如果使用了能量最小化方法）。

但是，仅考虑边界点的移动会引入一些问题。首先是从边界点准确计算出曲率很困难。正如大家所了解的，任一基于导数的运算对噪声都极其敏感。由于曲率计算引入了二阶导数，因此效果甚至更差。另一个问题是并不真正存在一种有效的方法，能使边界划分为几个相互分离的部分。接下来要描述的水平集方法试图解决这些困难。

还记得距离变换吗？在第7章，从距离变换引出一个函数$DT(x, y)$，像素点在边界上时函数取值为零，随着像素点远离边界，函数值增大。现在，考虑距离变换的一种新形式，在相关轮廓的外部，它的值与原距离变换函数相同（记住，轮廓线是闭合的，因此在这里内部和外部的概念是有意义的。）在轮廓内部，新的函数（我们称之为*度量函数*）是距离函数的负值。 199

$$\psi(x,y) = \begin{cases} DT(x,y) & ,(x,y)\text{在轮廓外部} \\ -DT(x,y) & ,(x,y)\text{在轮廓内部} \end{cases} \tag{8-3}$$

值得一提的是对于轮廓上的点，度量函数等于零。满足$\psi(x, y)=C$的点集被称作ψ的C水平集，而我们特别感兴趣的是零水平集。

现在，我们要修改度量函数ψ。对每个点(x, y)，我们计算一个新的值$\psi(x, y)$。有多种修改这些点的方式，下面会谈到其中的一部分，但要记住，修改后感兴趣的轮廓仍然是度量函数为零的点的集。首先进行初始化，使它等于距离变换，但从这里开始，不能再将它认为是距离变换（尽管它仍然保留了距离变换的某些特性）。相反，只需认为它是另外一个“亮度值”，一个关于x和y的函数。

什么是度量函数的梯度？当它是亮度值时，我们是知道如何计算它的梯度的。这里没有任何区别。将一个水平集看作是等亮度线，就可知其梯度是正交于等亮度线的，由此，给定梯度向量时，我们该如何得到法向量呢？还记得怎么计算梯度吗？

$$G(x,y)=\left[\frac{\partial}{\partial x}\psi(x,y),\frac{\partial}{\partial y}\psi(x,y)\right]^{\mathrm{T}} \tag{8-4}$$

那么，法向量恰好（必然地）是梯度归一化后的值。

$$\boldsymbol{n}(x,y)=\frac{\boldsymbol{G}(x,y)}{|\boldsymbol{G}(x,y)|} \tag{8-5}$$

因此，我们可以将轮廓的法向量与度量函数的梯度联系起来。也可以将轮廓在法向上的移动与一个函数（被称作速度函数）联系起来[9.46]，后者描述度量函数值以多快的速度变化，并可得到对这个变化的一个微分表达式：

$$\frac{\partial}{\partial t}\psi(x,y)=s(\kappa)|\nabla\psi(x,y)| \tag{8-6}$$

其中，s是个与问题相关的速度函数，以曲率与图像边缘作为它的参数，与公式（8-2）中的定义类似。

有多种方式修改度量函数的方式。例如，可以采用一个看起来像梯度下降的一种形式

$$\psi^{n+1}(x,y)=\psi^{n}(x,y)-\alpha s(x,y)|\nabla\psi(x,y)| \tag{8-7}$$

或者是看起来像微分方程的形式

200

$$\frac{\psi^{n+1}(x,y)-\psi^{n}(x,y)}{\Delta t}=s(x,y)|\nabla\psi(x,y)| \tag{8-8}$$

其中s引入了有关图像中亮度变化的某个量，又引入了等亮度线在x，y上的曲率（二维）。当然，如果你一定要使用零水平集的二维曲率，则需要将曲率与函数ψ联系起来，后者幸好不是特别难以对付。由于已计算出了法向量，可以将曲率与ψ在法向上的变化量联系起来，因此有：

$$\kappa=\frac{\psi_{xx}\psi_x^2-2\psi_x\psi_y\psi_{xy}+\psi_{yy}\psi_y^2}{(\psi_x^2+\psi_y^2)^{3/2}} \tag{8-9}$$

这里去掉了函数标记符号。

在整个算法过程中，度量函数按照类似公式（8-7）的规则演变。随着它的演变，会在不同点取零值，而这些点正是定义了轮廓的演变过程。

在实现这样一个算法时，一个必须要考虑的有趣的细节是，轮廓有可能会穿过它自身。

例如，考虑图8-16所示轮廓片断。在图中当前轮廓包含了一个尖锐的凹点,图中还画出了一些典型的法向量。沿着位置最低的点上的法向量，前进一个单位的长度，会使新的轮廓点落在轮廓内部。处理这个问题有一个很简单的启发式方法——已标注为内点的点不再被认为处于轮廓之外。

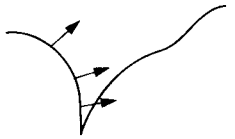


图8-16　轮廓线内有一个非常尖锐的折点。在折点附近，沿着法向移动一个单元，会使新的轮廓线进入原轮廓线内部

对自适应轮廓线采用水平集的想法最初由Sethian[9.66，9.67]提出。Malladi等[9.46]观察到仅考虑一组靠近当前轮廓线的点有很多优势，从而扩展了这个想法。Taubin 和Ronfard[9.84]隐含地使用了水平集的概念来拟合分段线性曲线。Kimmel等[9.40]则论证了水平集可以用在其他事情上，如寻找最短曲面路径。

并非所有采用可变形轮廓方法的算法都遵循第8.5节中描述的策略。例如，Lai和Chin[8.33]描述了另一种方法，将轮廓点看作是随机变量序列，从而可以用一个马尔科夫过程来描述，并采用MAP策略来优化。尽管这里不能讨论这些算法，但是读者可以从本章末尾列出的参考文献中找到相应文章。

8.6　曲面分割

在距离图像中，（一般情况下）含有许多曲面。通常采用两种策略进行曲面分割。第一种策略，只搜索弯曲速度不是很大的曲面。采用这个原则得到的算法可搜索到平滑解，且该算法沿着高曲率曲面的线进行区域分割。采用这个原则的一个例子已在第6章中进行了讨论，在那里，我们描述了这样一个算法，它在为数据点搜索最佳分段线性拟合时，可去除噪声。这样的拟合等价于用一组平面拟合一个曲面。平面之间接合的点或者产生“屋脊”边缘，或者产生“阶跃”边缘，这取决于视点。如果采用了类似MFA的退火算法，对更一般的曲面，可得到好的分割结果，但不能从头至尾执行这一种算法，否则会得到一个平面解[8.6]。第二种关于距离图像的分割策略是为这个曲面假定一些方程，例如二次曲面（普通的二阶曲面，第8.6.1中有定义）,然后令所有满足这个方程且相互邻近的点都属于同一个曲面。这种策略将分割问题与拟合问题混合起来，因为直到得出某种分割[8.7，8.59]结果时，才能知道在估计曲面参数时应使用哪些点。在下一节，我们将更详细地讨论这两种策略。

201

8.6.1　曲面描述

我们经常会需要对数据进行曲面拟合。Taubin等[8.68]仔细研究了对数据进行曲面拟合的问题，并首先观察到多项式曲面是非常富有吸引力的，但是这样的多项式应该是偶数阶的。隐式方程明显地更具吸引力，但更难以拟合。例如考虑第4章提到的二阶形式。显式表达式是

$$z = ax^2 + by^2 + cxy + dx + ey + f \tag{8-10}$$

而隐式表达式是

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0 \tag{8-11}$$

公式（8-11）的表达式称为一个二次曲面（quadric），它是描述所有二阶曲面（圆锥、圆球、平面、椭球等等）的通用形式。第5章中已经学习了如何通过最小化均方误差，用一个显式函数进行数据拟合。不幸的是，显式函数形式并不适用于z为高阶项的情况。可以采用二次形式

解决公式（8-11）中的z项，然后得到一个显式表达式。但是现在右式有一个平方根，因此也就不可能采用线性方法解系数向量。

我们可以采用隐式表达式，首先定义 $f(x,y,z)=ax^2+by^2+cz^2+dxy+exz+fyz+gx+hy+iz+j$ 并得到如下观察结果：如果点$[x_i, y_i, z_i]^T$在参数向量$[a, b, c, d, e, f, g, h, i, j]^T$描述的曲面上，则$f(x_i, y_i, z_i)$应精确等于零。我们定义一个函数的水平集为一组点集$[x_i, y_i, z_i]^T$，这些点关于某个标量$L$有$f(x_i, y_i, z_i)=L$。这样，我们可以找到最小化 $E=\sum_i(f(x_i,y_i,z_i))^2$（也称作从点$(x_i, y_i, z_i)$到曲面的代数距离）的系数。在某些情况下，这种方法可以得到很好的结果，但这并不是我们真正想要的；真正想要的是使得 $\sum_i d([x_i,y_i,z_i],f(x,y,z))$ 最小的解，其中d是某种距离度量，例如从该点到曲面的欧氏距离（被称为到曲面的几何距离[8.66]）。这类问题又一次证明了是难以用代数方法处理的。（关于实现过程中的细节可参见[8.67]和[17.37]。）尽管基于代数距离的方法在大多数时候工作得相当好，但它们也必然会有失败的情况。不论采用哪种距离度量方式，它都应该具有如下性质[17.37]：1）只要实际的（欧式，几何）距离为零，量测结果也应为零（代数距离满足这一点）；2）在采样点上，实际距离与量测距离关于参数的导数相等。

202

当然，无论你选择了哪种表达方式（多项式是最常用的），始终希望它具有仿射不变性[8.27]。

8.6.2 椭圆及椭球拟合

前文已经介绍过了代数距离的概念，尽管本章讨论的是曲面拟合，但可以首先谈谈更为简单一些的情况，即用椭圆来拟合曲线数据，然后将它扩展到三维的情况，也就是椭球拟合。关于圆锥曲线的通用方程可用来描述一个椭圆：

$$ax^2+bxy+cy^2+dx+ey+f=0 \tag{8-12}$$

这个隐式方程不只描述了椭圆，还描述了直线、双曲线、抛物线以及圆。同时，为了保证结果曲线是个椭圆，我们必须保证它满足

$$b^2-4ac<0 \tag{8-13}$$

为满足这个约束条件，必然会引出一个非线性优化问题。相反，如果我们找到了系数$a-f$，使得下式取最小值

$$\sum_i(ax_i^2+bx_iy_i+cy_i^2+dx_i+ey_i+f)^2 \tag{8-14}$$

我们会得到这样一个解，它试图用双曲线弧来拟合低曲率区域，而不是采用椭圆。同样的问题在试图采用椭球去拟合距离数据时也会遇到。具体细节参见Wang等[8.74]、Rose和West[8.50]和Fitzgibbon等[8.19]。

在执行这一类拟合时，一个很重要的问题是，需要弄清何时某个点完全变成了“外点”。也就是说，该点受到了大噪声的影响，使得实际上本该属于当前曲面的点变成了外点，或者是它实际上属于另一个被遮挡了的曲面。Darrell和Petland[13.9]研究了这个问题的有关细节，并指出了采用“M估计”可得到极好的分割结果。Cabrera和Meers[8.11]采用了一种称为“自举法”的迭代式算法，该算法可消除椭圆拟合中受外点影响引起的偏差。

203

如何用一个函数来拟合数据，还取决于附加在数据上的噪声或者说污染的性质。如果噪声是加性的、零均值高斯分布（我们几乎总是会假设这样一个函数），则最小垂直距离（MMSE）或最小法向距离（我们称之为特征向量线拟合）方法很有效。如果噪声不是高斯分布的，则应采用其他方法。例如，核医学图像主要是被计数（泊松，Poisson）噪声污染。泊松噪声与高斯噪声之间的区别有两点：前者始终为非负，且与信号相关。对远离零的点泊松噪声可以用加性高斯噪声来合理建模，使其方差等于信号的方差。其他传感器会引入其他类型的噪声。Stewart[8.64]考虑了内点和外点两种情况，但假定有些坏数据随机分布在传感器的动态区域内，即噪声不是加性的。

给定一个分割结果，是否应该合并两个相邻区域？如果是相邻的且在某种噪声度量条件下满足同样的方程，它们应该合并[8.8，8.29，8.34，8.56]。其他关于曲面拟合的相关论文包括[8.4，8.75，8.78]。

实现分割时，采用什么曲面度量方式作为基础，是遇到的另一个问题。对于这个问题，曲率显得尤其引人瞩目，这是因为曲率观测量是视点不变的。但是，“曲率估计对量化噪声非常敏感”[8.71]。

8.7 评估分割质量

至此，我们看到有很多种分割算法及其变种。但是哪一种最好呢？有谁知道？我们需要一个算法可以评估分割的质量。但是哪一个是最好的评估算法呢？我们还需要一个算法来评估质量评估算法，但是哪一个...（救命啊！）。

现有若干个方法可对分割质量进行评估。由于分割的一个结果是得到边缘，因此可以通过测量边缘位置，来间接地推断出分割质量。Pratt[5.33]提供了这样一个算法。

Bilbro和Snyder[8.6]首先去除噪声，然后对去噪后结果拟合曲面。他们仅仅考虑了去噪质量，这可以很简单地测试：从原始图像中减去分割后的干净图像。你所看到的应当仅有噪声。如果去噪算法生成的是一幅含有特征的图像，则它一定是除了噪声外还去除了其他的东西。

评估一个亮度图像的分割质量的确很困难，这是因为，关于哪种答案是正确的,不同的观测人员会得到不同的结论。不过，对于距离图像，确定“真值”会稍微简单一些，因为可以用物理方法测量曲面。 204

Hoover等[8.22]提出了如下形式：将人工分割的图像作为真值（ground-truth，GT），与一幅机器分割（machine-segmented，MS）的图像的质量作比较。令M和G各自代表MS以及GT图像；$M_i(i=1,\cdots,m)$表示M中的一个区域；令$G_j\ (j=1,\cdots,m)$表示G中的一个区域。$|R|$代表区域R中的像素个数。令O_{ij}是同时属于MS图像的区域i与GT图像的区域j的像素个数。最后，令T为一个阈值，$0.5<T\leqslant 1.0$。

有五种不同的分割结果，定义如下：

1）当$O_{ij}\geqslant T|M|$且$O_{ij}\geqslant T|G|$时，认为得到了正确的分类。

2）当GT图像中的一个区域在MS图像中分裂为若干区域时，称为过分割（oversegmentation）。形式化描述如下：给定GT中的一个区域（i）以及若干MS中的区域$(j_1, j_2, \cdots, j_n)$过分割符合如下情形：

a) MS的每个区域中，至少100T个百分比的像素实际上属于GT中的区域i，$(O_{i,j_l}\geqslant T|j_l|,\forall l)$。

b) 至少100T个百分比的像素实际上属于区域i的像素被标记为属于区域$j_1, j_2, \cdots, j_n$的

并集，（$\sum_{l=1}^{n} O_{i,j_l} > T|i|$）。

3）当GT图像中不同区域内的像素被确定为属于MS图像中的同一区域，则称为*欠分割*（undersegmentation）。这个定义与过分割的定义是相互对应的，如果把两幅图像的位置互换，就变成过分割了。

4）当GT图像中的一个区域既没有正确地分割，又不是过分割或欠分割中的一部分，则称为*错分类*（missed-classification）。

5）*噪声分类*（noise classification）的定义与错分类的类似，除了所关注的区域属于MS图像。

如果是距离图像，可通过计算GT区域以及相应MS区域的法向量，计算这两个向量之间的大角的绝对值来进一步比较两种正确的分割结果。

根据这些定义，我们可以通过对正确的或者错误的分割结果进行计数，并得到关于总的角度误差的观测量，来评估分割质量。通过画出T与观测量之间的关系图，并比较这些图，可确定有关分割性能的一个观测结果。

Hoover等[8.22]采用了以上方法，对四种不同的距离图像分割算法，实现了完整的质量评估。

8.8 结论

本章使用了一致性概念来辨识一个区域的分量。在我们研究的第一个例子中，如果所有的像素都具有同样的亮度，则它们被定义为属于同一个区域。在第8.6.1节的例子中，所有满足同一个曲面方程的点被定义为属于同一个区域。

205

第8.2.1节中，我们采用了一种优化方法（最小均方误差）来寻找最优阈值。在8.5节，我们采用主动轮廓线原理，并通过指定一个针对特定问题的目标函数，并最小化该函数，来获得一个闭合边界。任何适当的最小化技术都可以采用。在8A.5节中，我们会再次看到用于最大后验方法中的函数最小化过程（基于梯度下降法，并带有退火过程），用于寻找使得特定的目标函数最小的图像，得到分割结果。

8.9 术语

主动视觉	active contour
代数距离	algebraic distance
连通分量	connected component
显式	explicit form
几何距离	geometric distance
直方图	histogram
均匀的	homogeneous
隐式	implicit form
标号图像	label image
法向	normal direction
过分割	oversegmentation
二次曲面	quadric

区域增长	region growing
显着点	salient point
分割	segmentation
蛇形线算法	snake
曲线速率	speed of a curve
阈值化	thresholding
欠分割	undersegmentation

作业8.1

Haralick和Shapiro[4.18]中第2.3.5节描述了一种标号算法，与本文介绍的标号方法有一些类似的地方。请对二者之间的差异进行比较。考虑以下因素：1）实现一个单处理器（uniprocessor）的简易性（代码的简易性）；2）在单处理器上的执行速度；3）并行化实现的可能性。 206

表8-2　查找表

位置	0	1	2	3	4	5	6	7	8	9
内容	0	0	2	2	4	4	2	4	2	2

作业8.2

在执行连通分量标号方案的过程中，在察看某一个点时，查找表如表8-2所示。

现在，我们发现了如下等价关系：9 = 7。请在表8-2中的空白行，写出处理完该等价关系后查找表的内容。

专题8A　分割

到目前为止，本章考虑的问题是如何将图像划分为相互之间在亮度或者深度上有某种区别的区域。不过，对于可用来表征像素，或者像素邻近区域的任意特征，前面所描述的算法都是适用的。由于这个原因，这里将讨论一些其他可采用的度量方式，包括纹理、颜色以及运动。本节也将提到其他的分割方法，如基于边缘的分割方法。

8A.1　纹理分割

在第4A.2.2节中，对纹理进行了讨论。如果能够对纹理的概念进行量化——为了区分两种不同的纹理，分别赋予它们两个不同的值——那么纹理也可以作为特征，在分割算法中使用。此时，不再把具有相似的亮度定义为相邻，而是用具有相似的纹理定义。还可以将颜色与纹理结合起来[8.46，8.65]。

一些研究人员已注意到实际上有两种根本不同的纹理类型，一种在某种意义上是“确定的”，而另一种则在某种意义上是“随机”的，理论上可以用马尔科夫随机场来建模[8.2]。Liu和Picard[8.38]以及其他人[8.20，8.21，8.58]发现傅里叶变换的峰值有助于理解两种纹理特征之间是怎样区分的。 207

表8-3 在分形维度的计算中用的二等分

n	0	1	2	n
ε	1	1/2	1/4	$1/2^n$
M_ε	1	2	4	2^n

表8-4 分形维度

n	0	1	2	n
ε	1	1/2	1/4	$1/2^n$
M_ε	1	4	16	2^{2n}

分形维度

分形维度（Fractal dimension）是关于一个形状在不同尺度下量测自相似性时的一种度量。这种度量方式可以提供一个关于前景区域中（可假定为感兴趣的目标）点的空间分布的度量。为了了解分形维度的用法，可以设想一个二维空间内的点集S，并定义S的分形维度为

$$\dim(S) = \lim_{\varepsilon \to 0} \frac{\log M_\varepsilon}{\log(1/\varepsilon)} \tag{8-15}$$

其中，M_ε是覆盖S所需$\varepsilon \times \varepsilon$方格的个数。现在举一些例子来看看它们的分形维度，首先从单个点开始。很显然，一个方格即可以完全覆盖一个点，与方格的大小无关。因此，在公式（8-15）中，M_ε总是等于1，而分母的极限为无穷大。因此，我们得到一个点的分形维度为零。现在，考虑一条单位长度直线。很明显，这样的一条直线应由一个1×1的方格来覆盖。但是，它也可以用两个方格来覆盖，每个方格大小为$1/2 \times 1/2$，或者4个方格，每个方格大小为$1/4 \times 1/4$。每次方格大小减半，所需要的方格个数则加倍。我们将这个过程按参数n制成表格，n等于已完成的二等分次数（见表8-3）。

根据表8-3，我们可以按下式估计$\dim(S)$

$$\dim(S) = \lim_{n \to \infty} \frac{\log 2^n}{\log(1/(1/2))^n} = 1 \tag{8-16}$$

最后，令S为一个方形区域，为了不失一般性，令其边长为一。我们可以用单个1×1的方格来覆盖它，或者4个$1/2 \times 1/2$方格，或者16个$1/4 \times 1/4$方格，等等。

从表8-4可以看出，其分形维度为

208

$$\dim(S) = \lim_{n \to \infty} \frac{\log 2^{2n}}{\log(1/(1/2))^n} = \lim_{n \to \infty} \frac{\log 2^{2n}}{\log 2^n} = \lim_{n \to \infty} \frac{2n}{n} = 2 \tag{8-17}$$

因此，最后我们可以得出一个直观上非常吸引人的结果：单个点是零维物体，一条直线是一维物体，而方形是二维的。至少，这里得出的关于这些简单形状的结果，与直觉知识是一致的。

给定一幅图像，关于如何将公式（8-15）应用到离散的像素域，我们必须给出某种想法。这是因为我们很显然不可能使有界方格的大小比ε=1还要小。求解方程（8-15）中的M_ε，我们发现对一个小的ε

$$\log M_\varepsilon = \left(\log\left(\frac{1}{\varepsilon}\right)\right) \dim(S) \tag{8-18}$$

方程可简写为

$$\log M_\varepsilon = -(\log \varepsilon)\dim(S) \tag{8-19}$$

因此，对一个小的ε，M_ε的对数与ε的对数之间满足线性关系，其斜率为S的维度。从而通过考虑我们拥有的ε的两个最小值（1和2）来估计出S的维度，并使用如下公式计算出该斜率

$$\dim(S) = -\frac{\log M_1 - \log M_2}{\log 1 - \log 2} = \frac{\log M_1 - \log M_2}{\log 2} \tag{8-20}$$

这样可生成一个简单算法：找到M_1，即覆盖对象所需的1×1方格的个数，这也就是用像素来量测目标时的面积大小，找到M_2，即覆盖同一个对象所需的2×2方格的个数，并取各自对数之间的差。

考虑如下例子。如图8-17中的左图，其前景区域面积为36，可以被9个2×2方形覆盖因此，它的分形维度为$(\log 36 - \log 9)/\log 2 = 2$。右图中的前景区域与左图中的前景区域面积大小相同，但需要12个2×2方形来覆盖，因此其分形维度为$(\log 36 - \log 12)/\log 2 = 1.58$。

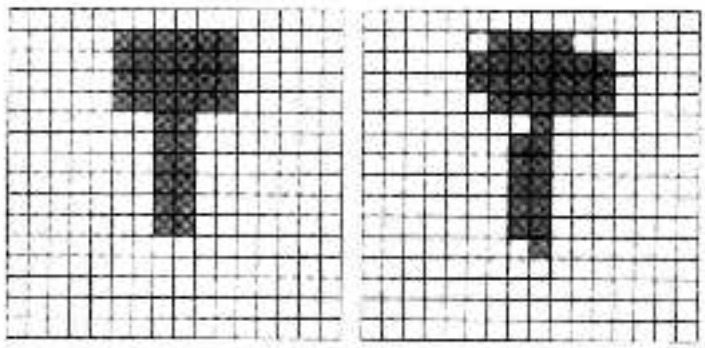


图8-17　左图的分形维度为2.0，右图的分形维度为1.58。右图中，用黑色线框出了覆盖前景区域的2×2方形

分形维度的概念可以扩展至灰度图像，也可以扩展至以这种度量方式抽取的灰度图像特征[8.12]。 [209]

8A.2　使用边缘的图像分割

实现分割的一种方式是利用连通分量算法，但需要假定有一些特殊的点不与任何点相连，边缘点就是这种点的例子[8.75]。

实用边缘分割时通常遇到的问题是边缘检测，正如大家所熟知的，边缘检测时既容易产生额外的边缘又容易丢失部分边缘，从而造成边缘缺口。Jacobs[8.24]针对这个问题提出了一种很有趣的方法，他首先根据边缘定义了一个“可接受区域”，要求区域是凸的且区域边界上的边缘集合大部分是真实测量的，而不是推断出的。换句话说，任意两个点集之间，我们都可以假定边缘是存在的。但是，如果某个特定的点集是连通的，只有少许缺口，且这些缺口加起来比起区域周长要小，则我们可以相信这样一个区域，也就是显著群(salient group)，是真实存在的。Jacobs提出了一些非常精巧的启发式规则，简化了对可能区域的组合搜索过程，使得这一过程相当容易控制。

第11章中描述的Hough变换，提供了另外一种识别边缘片断的方法，可以处理部分边缘缺失的情况。

8A.3 运动分割

如果能够确定所有具有相同运动特性的连通像素，则可以应用前面讨论过的连通分量方法。尽管有些论文非常强调运动分割算法——Patras等[8.47]采用了分水岭方法和一个多步分割方法——但是对任一运动分割算法来说，都是以找到有效的检测和表达运动的方法为基础的，这是非常重要的。

有大量的研究工作致力于解决如何表达图像运动特征的问题。对此，我们将在第9A.3节进行更详细的讨论。

8A.4 颜色分割

正如纹理变化可用于实现分割，颜色变化也可用于分割。基于颜色变化的分割，包括提出一个最优化问题然后最小化一个目标函数。于是，使得目标函数最小的图像就是分割结果。Liu和Yang [8.36]使用模拟退火方法来搜索一个好的颜色分割结果。聚类是另一种方法[8.72]，关于聚类方法将在第15章中进行简略描述。

8A.5 使用MAP方法的分割

1991年，Snyder等[8.63]发现有一种非常简单直接的方式，只要已知每个类的亮度的均值或者还有方差，就可以将一个复原问题转化为一个分类问题[8.39，8.63]。2000年，研究者使用变分方法也推出了同样的结论[8.53]。为了将一幅图像表面“分类”为不同的类，即，不同的区域，我们可以使用与解决图像复原问题时相同的MAP方法。

210

为了使用MAP方法，只需修改先验项，以便将期望亮度包括进去。举个例子，假设我们具有先验知识，认为一幅图像应该在除了阶跃不连续处之外都是平滑的；另外，每个像素值都只能取三种亮度中的一个（比如，脑脊髓液（黑色）、脑白质以及脑灰质）。则在第i个像素处的亮度值等于其邻域亮度值，或者等于k_1、k_2或k_3时，如下先验项取最大。

$$H(f)=\sum_{i}\sum_{j\in\mathcal{N}_i}\exp(-(f_i-f_j)^2(f_i-k_1)^2(f_i-k_2)^2(f_i-k_3)^2) \tag{8-21}$$

关于进一步的讨论，可参考Zhu和Yuille[8.80]，该文中表明几个相似的算法可结合起来。此外，他们还举例说明了多种图像表达方式之间的关系。

8A.6 人如何完成分割

有关分割，还有一个问题，我们很遗憾没有更多时间在本书中进行描述。这个问题就是人是如何完成分割的。即什么是“正确的”分割？例如，Koenderink和van Doorn [8.28]观察到人在进行分割时会倾向于将三维物体的投影理解为多个椭圆的集合。有关内容可参见 [8.57]。

参考文献

[8.1] R. Adams and L. Bischof, “Seeded Region Growing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), 1994.

[8.2] P. Andrey and P. Tarroux, “Unsupervised Segmentation of Markov Random Field Modeled Textured Images using Selectionist Relaxation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(3), 1998.

[8.3] K.E. Batcher, "STARAN Parallel Processor System Hardware," *Proceedings of AFIPS National Computer Conference*, vol 43, pp. 405–410, 1974.

[8.4] J. Berkmann and T. Caelli, "Computation of Surface Geometry and Segmentation using Covariance Techniques," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), 1994.

[8.5] S. Beucher, "Watersheds of Functions and Picture Segmentation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Paris, May, 1982.

[8.6] G. Bilbro and W. Snyder, "Range Image Restoration using Mean Field Annealing," In *Advances in Neural Network Information Processing Systems*, San Mateo, CA, Morgan-Kaufmann, 1989.

[8.7] G. Bilbro and W. Snyder, "Optimization of Functions with Many Minima," *IEEE Transactions on Systems, Man, and Cybernetics*, **21**(4), July/August, 1991.

[8.8] G. Blais and M. Levine, "Registering Multiview Range Data to Create 3D Computer Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[8.9] K. Boyer, M. Mirza, and G. Ganguly, "The Robust Sequential Estimator: A General Approach and its Application to Surface Organization in Range Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), 1994.

[8.10] C.R. Brice and C.L. Fennema, "Scene Analysis using Regions," *Artificial Intelligence*, **1**, pp. 205–226, Fall, 1970.

[8.11] J. Cabrera and P. Meer, "Unbiased Estimation of Ellipses by Bootstrapping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(7), 1996.

[8.12] B. Chaudhuri and N. Sarkar, "Texture Segmentation using Fractal Dimension," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[8.13] J. Chen, Y. Sato, and S. Tamura, "Orientation Space Filtering for Multiple Orientation Line Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), 2000.

[8.14] C. Chow and T. Keneko, "Automatic Detection of the Left Ventricle from Cineangiograms," *Computers and Biomedical Research*, **5**, pp. 388–410, 1972.

[8.15] T. Davis, "Fast Decomposition of Digital Curves into Polygons using the Haar Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(8), August, 1999.

[8.16] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, New York, Wiley, 1973.

[8.17] M. Fishler and R. Bolles, "Perceptual Organization and Curve Partitioning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(1), 1986.

[8.18] M. Fishler and H. Wolf, "Locating Perceptually Salient Points on Planar Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(2), 1994.

[8.19] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct Least Square Fitting of Ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(5), May, 1999.

[8.20] J. Francos, "Orthogonal Decomposition of 2-D Random Fields and Their Applications in 2-D Spectral Estimation," *Signal Processing and its Applications*, ed. N.K. Bose and C.R. Rao, Amsterdam, North Holland, 1993.

[8.21] J. Francos, Z. Meiri, and B. Porat, "A Unified Texture Model Based on a 2-D Wold Like Decomposition," *IEEE Transactions on Signal Processing*, **41**, pp. 2665–2678, August, 1993.

[8.22] A. Hoover, G. Jean-Baptiste, X. Jiang, P. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D. Eggbert, A. Fitzgibbon, and R. Fisher, "An Experimental Comparison of Range Image Segmentation Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(7), 1996.

[8.23] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(11), 1998.

[8.24] D. Jacobs, "Robust and Efficient Detection of Salient Convex Groups," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(1), 1996.

[8.25] K. Kanatani, "Statistical Bias of Conic Fitting and Renormalization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), 1994.

[8.26] N. Katzir, M. Lindenbaum, and M. Porat, "Curve Segmentation under Partial Occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[8.27] D. Keren, "Using Symbolic Computation to Find Algebraic Invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), 1994.

[8.28] J. Koenderink and A. Van Doorn, "The Shape of Smooth Objects and the Way Contours End," *Perception*, **11**, pp. 129–137, 1982.

[8.29] K. Köster and M. Spann, "MIR: An Approach to Robust Clustering – Application to Range Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), 2000.

[8.30] L. Krakauer, "Computer Analysis of Visual Properties of Curved Objects," Project MAC TR-82, 1971.

[8.31] S. Kulkarni, S. Mitter, T. Richardson, and J. Tsitsiklis, "Local vs. Global Computation of Length of Digitized Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

[8.32] S. Kumar, S. Han, D. Goldgof, and K. Bowyer, "On Recovering Hyperquadrics from Range Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[8.33] K. Lai and R. Chin, "Deformable Contours: Modeling and Extraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), pp. 1084–1090, 1995.

[8.34] S. LaValle and S. Hutchinson, "A Bayesian Segmentation Methodology for Parametric Image Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 1995.

[8.35] K. Lee, P. Meer, and R. Park, "Robust Adaptive Segmentation of Range Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(2), 1998.

[8.36] J. Liu and Y. Yang, "Multiresolution Color Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

[8.37] X. Liu and R. Ehrich, "Subpixel Edge Location in Binary Images using Dithering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(6), 1995.

[8.38] F. Liu and R. Picard, "Periodicity, Directionality, and Randomness: Wold Features for Image Modeling and Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(7), 1996.

[8.39] A. Logenthiran, W. Snyder, and P. Santago, "MAP Segmentation of Magnetic Resonance Images using Mean Field Annealing," *SPIE Symposium on Electronic Imaging, Science and Technology*, February, 1991.

[8.40] A. Matheny and D. Goldgof, "The Use of Three- and Four-dimensional Surface Har-

monics for Rigid and Nonrigid Shape Recovery and Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(10), 1995.

[8.41] D.L. Milgram, A. Rosenfeld, T. Willett, and G. Tisdale, "Algorithms and Hardware Technology for Image Recognition," *Final Report to U.S. Army Night Vision Lab*, March 31, 1978.

[8.42] F. Mokhtarian and A. Mackworth, "Scale-based Description and Recognition of Planar Curves and Two-dimensional Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(1), 1986.

[8.43] K. Mori, M. Kidode, H. Shinoda, and H. Asada, "Design of Local Parallel Pattern Processor for Image Processing," *Proceedings AFIPS National Computer Conference*, vol 47, pp. 1025–1031 June, 1978.

[8.44] W. Ng and C. Lee, "Comment on Using the Uniformity Measure for Performance Measure in Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(9), 1996.

[8.45] B. Parhami, "Associative Memories and Processors: An Overview and Selected Bibliography," *Proceedings of the IEEE*, **61**, pp. 772–730, June, 1973.

[8.46] D. Panjwani and G. Healey, "Markov Random Field Models for Unsupervised Segmentation of Textured Color Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(10), 1995.

[8.47] I. Patras, E. Hendriks, and R. Lagendijk, "Video Segmentation by MAP Labeling of Watershed Segments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3), 2001.

[8.48] A. Pikaz and I. Dinstein, "Using Simple Decomposition for Smoothing and Feature Point Detection of Noisy Digital Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), 1994.

[8.49] E. Rivlin and I. Weiss, "Local Invariants for Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(3), 1995.

[8.50] P. Rosen and G. West, "Nonparametric Segmentation of Curves into Various Representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(12), 1995.

[8.51] A. Rosenfeld and A. Kak, *Digital Picture Processing*, 2nd edition, New York, Academic Press, 1997.

[8.52] G. Roth and M. Levine, "Geometric Primitive Extraction using a Genetic Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

[8.53] C. Samson, L. Blanc-Fèraud, G. Aubert, and J. Zerubia, "A Variational Model for Image Classification and Restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), 2000.

[8.54] G. Sapiro and A. Tannenbaum, "Area and Length Preserving Geometric Invariant Scale Spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[8.55] H. Sheu and W. Hu, "Multiprimitive Segmentation of Planar Curves – a Two-level Breakpoint Classification and Tuning Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(8), 1999.

[8.56] H. Shum, K. Ikeuchi, and R. Reddy, "Principal Component Analysis with Missing Data and Its Application to Polyhedral Object Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(9), 1995.

[8.57] K. Siddiqi and B. Kimia, "Parts of Visual Form: Computational Aspects," *IEEE Trans-*

actions on Pattern Analysis and Machine Intelligence, **17**(3), 1995.

[8.58] R. Sriram, J. Francos, and W. Pearlman, "Texture Coding Using a Wold Decomposition Model," *Proc. International Conference on Pattern Recognition*, Jerusalem, October, 1994.

[8.59] W. Snyder and G. Bilbro, "Segmentation of Range Images," *International Conference on Robotics and Automation*, St. Louis, March, 1985.

[8.60] W. Snyder and A. Cowart, "An Iterative Approach to Region Growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5**(3), 1983.

[8.61] W.E. Snyder and C.D. Savage, "Content-Addressable Read-Write Memories for Image Analysis," *IEEE Transactions on Computers*, **31**(10), pp. 963–967, 1982.

[8.62] W. Snyder, G. Bilbro, A. Logenthiran, and S. Rajala, "Optimal Thresholding, A New Approach," *Pattern Recognition Letters*, **11**(11), 1990.

[8.63] W. Snyder, P. Santago, A. Logenthiran, K. Link, G. Bilbro, and S. Rajala, "Segmentation of Magnetic Resonance Images using Mean Field Annealing," *XII International Conference on Information Processing in Medical Imaging*, Kent, England, July 7–11, 1991.

[8.64] C. Stewart, "MINIPRAN: A New Robust Estimator for Computer Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(10), 1995.

[8.65] P. Suen and G. Healey, "The Analysis and Recognition of Real-world Textures in Three Dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), 2000.

[8.66] S. Sullivan, L. Sandford, and J. Ponce, "Using Geometric Distance Fits for 3-D Object Modeling and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(12), 1994.

[8.67] G. Taubin, "Nonplanar Curve and Surface Estimation in 3-space," *IEEE Robotics and Automation Conference*, Philadelphia, May, 1988.

[8.68] G. Taubin, F. Cukierman, S. Sullivan, J. Ponce, and D. Kriegman, "Parameterized Families of Polynomials for Bounded Algebraic Curve and Surface Fitting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), 1994.

[8.69] K.J. Thurber, *Large-Scale Computer Architecture: Parallel and Associative Processors*, Rochelle Park, NJ, Hayden, 1976.

[8.70] O. Trier and A. Jain, "Goal-directed Evaluation of Binarization Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(12), 1995.

[8.71] E. Trucco and R. Fisher, "Experiments in Curvature-based Segmentation of Range Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 1995.

[8.72] T. Uchiyama and M. Arbib, "Color Image Segmentation Using Competitive Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(12), 1994.

[8.73] K. Vincken, A. Koster, and M. Viergever, "Probabilistic Multiscale Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(2), 1997.

[8.74] R. Wang, A. Hanson, and E. Riseman, "Fast Extraction of Ellipses," *Ninth International Conference on Pattern Recognition*, Rome, 1988.

[8.75] M. Wani and B. Batchelor, "Edge-region-based Segmentation of Range Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), 1994.

[8.76] M. Worring and A. Smeulders, "Digitized Circular Arcs: Characterization and Parameter Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(6), 1995.

[8.77] S. Yau and H. Jung, "Associative Processor Architecture – A Survey," *Computer*

Surveys, 9, pp. 3–26, March, 1977.

[8.78] X. Yu, T. Bui, and A. Kryzak, "Robust Estimation for Range Image Segmentation and Reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[8.79] P. Zhu and P. Chirlian, "On Critical Point Detection of Digital Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[8.80] S. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(9), 1996.

211 ~ 215

第9章 形　状

Space tells matter how to move, and matter tells space to get bent.

Douglas Adams

在本章中，假定我们已经对图像进行了成功的分割，接下来探讨一下如何描述分割得到的区域。我们首先考虑二维区域，假定区域内的像素取值为1，区域外的像素取值为0，并假定每次只处理一个区域。这些假定是合理的，因为在学习连通部件标号（connected component labeling）时，我们已经学会了实现这些假定的方法。

在对图像进行分割的过程中，计算机对区域内的每个像素进行区域增长操作（region growing operation）。在这一过程中，计算机可以很容易地跟踪区域的面积。面积是一个物体区别于其他物体的众多特征之一。例如，在图像中，连杆通常比活塞占据更大的面积（或者说，覆盖了更多的黑色像素）。这样，通过度量区域的面积，就可以识别物体的类别。

9.1 线性变换

本章我们要讨论的话题之一是区域对线性变换的不变性。假定区域内的每个像素都用其二维坐标x, y构成的二维向量来表示，而变换就作用在这些二维向量上。首先，让我们来看一下如下的正交变换:

$$\boldsymbol{R}_z = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

它作用于区域内像素的原坐标生成新的坐标。例如，

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \boldsymbol{R}_z \begin{bmatrix} x \\ y \end{bmatrix}$$

其中$\boldsymbol{R}_z$定义如上，表示绕z轴的旋转。给定一个区域s，我们可以很容易地构造一个矩阵$\boldsymbol{S}$，$\boldsymbol{S}$中的每一列对应s中某个像素的坐标x, y。例如，给定区域$s = \{(1,2), (3,4), (1,3), (2,3)\}$，对应的$\boldsymbol{S}$矩阵就是

$$\boldsymbol{S} = \begin{bmatrix} 1 & 3 & 1 & 2 \\ 2 & 4 & 3 & 3 \end{bmatrix}$$

通过矩阵乘法我们可以把正交变换作用于整个区域，例如作如下的旋转变换:

$$\boldsymbol{S}' = \boldsymbol{R}_z \boldsymbol{S} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 & 2 \\ 2 & 4 & 3 & 3 \end{bmatrix}$$

用这种形式表示旋转变换很方便，但如何表示平移变换呢？为了表示平移变换，我们通过增加一行和一列来增广变换矩阵，除右下角元素为1外，其他新增元素均为0。增广后的新矩阵如下:

$$R_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

同时我们在每个点的坐标后增加一个1，使得二维坐标(x, y)变为

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

现在，我们就可以用一个矩阵把旋转和平移结合起来，这种矩阵称为齐次变换矩阵（homogeneous transformation matrix）。通过改变矩阵第三列的元素来描述平移，我们就可以做到这一点了。例如，如果一个点绕原点旋转θ，沿x轴和y轴分别平移dx和dy，就可以用如下的矩阵乘法来表示:

$$x' = \begin{bmatrix} \cos\theta & -\sin\theta & dx \\ \sin\theta & \cos\theta & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{9-1}$$

到目前为止，我们可以用一个矩阵乘法来描述图像平面内的旋转（即绕z轴的旋转）和平面内的平移。上述变换同属于一类变换，这类变换被称为“相似变换”（similarity transformation）。相似变换的特征是只改变物体的位置，不改变物体的形状。

217

那么我们又如何表示平面外的旋转呢？为了回答这个问题，我们需要先定义仿射变换（affine transformation）。仿射变换把一个二维向量$\boldsymbol{x} = [x, y]^{\mathrm{T}}$变换成另一个二维向量$\boldsymbol{x}' = [x', y']^{\mathrm{T}}$：

$$\boldsymbol{x}' = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \tag{9-2}$$

其中$\boldsymbol{b}$也是一个二维向量。这一变换看上去与上文提到的相似变换差不多，但是这里不要求$\boldsymbol{A}$是正交矩阵，只要$\boldsymbol{A}$是非奇异矩阵即可。（本节$\boldsymbol{R}_z$的非增广定义是$\boldsymbol{A}$的一个特例，它是正交矩阵。——译者注）仿射变换会改变区域的形状。如图9-1所示，仿射变换可能会产生剪切形变。

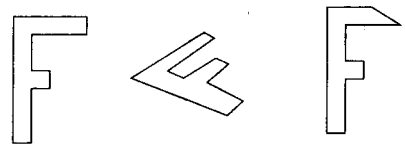


图9-1 仿射变换可以改变坐标轴的比例。如果各坐标轴的比例变换不同，图像会发生剪切形变

可能你已经意识到了，平面物体在图像平面外的转动相当于对物体进行仿射变换。这给了我们一个考察平面外转动的视角，尽管这一视角还非常有限。如果一个物体几乎位于同一平面上，平面外旋转的角度又不大，也就是说没有出现遮挡，那么我们可以用二维的仿射变换来表示这一三维运动。例如，图9-2给出了几幅飞机的图片，它们互为彼此的仿射变换。

图9-2 互为仿射变换的飞机图片（引自[9.1]）

经过平移校正后，仿射变换中的二维矩阵可以分解成各种变换的组合——旋转、缩放和剪切：

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \alpha & 0 \\ 0 & \delta \end{bmatrix}\begin{bmatrix} 1 & \beta \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \tag{9-3}$$

那么，我们如何应用这些变换的概念呢？我们可以通过逆变换对变换进行校正，并对准物体，这将有助于对物体形状的分析。例如，我们通过校正平移变换可以把物体的重心移到原点处，通过校正旋转变换可以使物体的主轴与坐标轴重合。 218

通过特定的线性变换可以找到物体的主轴，这种线性变换可以把物体（或物体边界）的协方差矩阵变换成单位阵。这种变换涉及白化变换（whitening transformation）和K-L变换。不幸的是，一旦对区域进行了这样的变换，区域内点与点之间的欧氏距离就被改变了。

在上一段中出现了“距离”这个词。通常人们会认为“距离”就是指“欧氏距离”。但在本书中，“距离”一词会多次出现并有几种不同的含意。因此，我们有必要对这一概念进行更严格的定义。欧氏距离只是度量（metric）的一种，只要满足如下条件就可以称为度量：

- $d(a,a)=0 \quad \forall a$
- $d(a,b)=d(b,a) \quad \forall(a,b)$
- $d(a,b)+d(b,c)\geq d(a,c) \quad \forall(a,b,c)$

在后面的章节中，我们将有机会考察几种不同的度量。

9.2 基于协方差矩阵的变换方法

考察图9-3中点的分布。每个点的位置都可以用一对有序数对(x_1, x_2)来描述，但只用x_1或x_2都不足以描述点的位置。

现在让我们来看一下图9-4。在这幅图中显示了两条新的坐标轴y_1和y_2。同样的，有序数对(y_1, y_2)可以精确地描述点的位置。但在大多数情况下，y_2的值都接近于0（与y_1的值相比）。这样，如果我们丢掉y_2而仅用标量y_1来描述点的位置，损失的信息就比较少。我们的目标是如何找到最优的y_1轴和y_2轴。

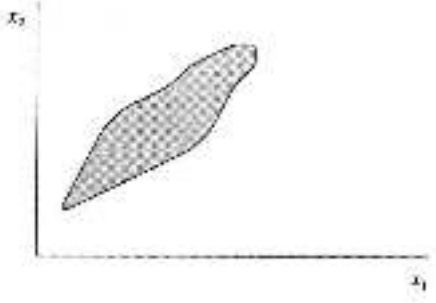


图9-3　一个近似于椭圆的区域

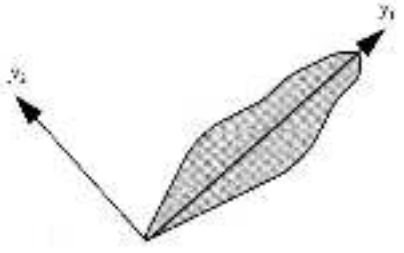


图9-4　旋转原坐标轴得到的新坐标系，新坐标系中一个单标就可以很好地描述数据 219

9.2.1 K-L展开的推导

设x是d维空间的随机向量，我们可以用一组基向量来描述x，即用如下方式表示x：

$$x=\sum_{i=1}^{d} y_i b_i \tag{9-4}$$

这里的向量$\boldsymbol{b}_i$是确定的，更通俗地说，可以是预先指定的。因为任何一个随机向量$\boldsymbol{x}$都可以用同一组的d个向量$\boldsymbol{b}_i$ $(i=1,\ldots,d)$来表示，因此我们说向量集$\boldsymbol{b}_i$张成了$\boldsymbol{x}$所在的空间，并把这组向量称为$\boldsymbol{x}$的一组基向量。为了进一步利用基向量，我们要求[⊖]:

1）基向量$\boldsymbol{b}_i$之间线性无关;

2）基向量$\boldsymbol{b}_i$之间两两正交（并且基向量为单位向量——译者注），即

$$\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{b}_j=\begin{cases}1 & (i=j)\\ 0 & (i\neq j)\end{cases} \tag{9-5}$$

这样一来，我们就可以得到$\boldsymbol{x}$在$\boldsymbol{b}_i$上的投影y_i了。投影的定义如下:

$$y_i=\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{x}(i=1,\cdots,d) \tag{9-6}$$

我们令

$$\boldsymbol{y}=[y_1,\cdots,y_d]^{\mathrm{T}} \tag{9-7}$$

在这里，我们称y_i为$\boldsymbol{x}$在基向量$\boldsymbol{b}_i$上的投影。

如果我们想丢掉$\boldsymbol{y}$向量的一些分量，仅保留$m(m<d)$个分量，并用这m个分量（被称作主分量）来表示$\boldsymbol{x}$，尽管这种表示是有误差的。假定我们用投影法计算出$\boldsymbol{y}$的前m个分量，其余的分量取常量，就生成了一个对$\boldsymbol{x}$的估计:

220

$$\hat{\boldsymbol{x}}=\sum_{i=1}^{m}y_i\boldsymbol{b}_i+\sum_{i=m+1}^{d}\alpha_i\boldsymbol{b}_i \tag{9-8}$$

在公式（9-8）中，由于我们用常量α_i取代了$\boldsymbol{y}$的一些分量而引入了如下的误差:

$$\begin{aligned}\Delta\boldsymbol{x}&=\boldsymbol{x}-\hat{\boldsymbol{x}}\\&=\boldsymbol{x}-\left[\sum_{i=1}^{m}y_i\boldsymbol{b}_i+\sum_{i=m+1}^{d}\alpha_i\boldsymbol{b}_i\right]\\&=\sum_{i=m+1}^{d}[y_i-\alpha_i]\boldsymbol{b}_i\end{aligned} \tag{9-9}$$

既然我们把$\boldsymbol{x}$看作随机向量，那么误差$\Delta\boldsymbol{x}$也是随机向量。我们可以用$\Delta\boldsymbol{x}$的模的期望来评价这种表示的近似效果:

$$\begin{aligned}\varepsilon^2(m)&=E\left\{\sum_{i=m+1}^{d}\sum_{j=m+1}^{d}(y_i-\alpha_i)\boldsymbol{b}_i^{\mathrm{T}}(y_j-\alpha_j)\boldsymbol{b}_j\right\}\\&=E\left\{\sum\sum(y_i-\alpha_i)(y_j-\alpha_j)\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{b}_j\right\}\end{aligned} \tag{9-10}$$

我们注意到y_i是标量，同时利用公式（9-5），可以得到:

$$\varepsilon^2(m)=\sum_{i=m+1}^{d}E\left\{(y_i-\alpha_i)^2\right\} \tag{9-11}$$

为找到α_i的最优值，我们求ε^2相对于α_i的极小值:

$$\frac{\partial\varepsilon^2}{\partial\alpha_i}=\frac{\partial}{\partial\alpha_i}E\{(y_i-\alpha_i)^2\}=-2(E\{y_i\}-a_i)=0 \tag{9-12}$$

⊖ 通常情况下并不要求基向量正交，只要不平行即可，但在这里，我们要求基向量正交。

得到结果如下:

$$\alpha_i = E\{y_i\} = \boldsymbol{b}_i^{\mathrm{T}} E(\boldsymbol{x}) \tag{9-13}$$

因此，用$\boldsymbol{y}$的一些分量的期望值来代替相应分量的实际值，可以得到最优的近似效果——在这一点上，数学推导和我们的直觉是吻合的。

把公式（9-13）代入公式（9-11），得到:

$$\varepsilon^2(m) = \sum_{i=m+1}^{d} E[(y_i - E\{y_i\})^2] \tag{9-14}$$

把公式（9-6）代入公式（9-14），得到:

221

$$\begin{aligned}\varepsilon^2(m) &= \sum_{i=m+1}^{d} E[(\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{x} - E\{\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{x}\})^2] \\ &= \sum_i E[(\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{x} - E\{\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{x}\})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}_i - E\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}_i\})] \\ &= \sum_i E[\boldsymbol{b}_i^{\mathrm{T}}(\boldsymbol{x} - E\{\boldsymbol{x}\})(\boldsymbol{x}^{\mathrm{T}} - E\{\boldsymbol{x}^{\mathrm{T}}\})\boldsymbol{b}_i]\end{aligned} \tag{9-15}$$

公式（9-15）可以改写成:

$$\varepsilon^2(m) = \sum_i \boldsymbol{b}_i^{\mathrm{T}} E[(\boldsymbol{x} - E\{\boldsymbol{x}\})(\boldsymbol{x} - E\{\boldsymbol{x}\})^{\mathrm{T}}]\boldsymbol{b}_i \tag{9-16}$$

我们看到公式（9-16）中在$\boldsymbol{b}$之间的部分是$\boldsymbol{x}$的协方差矩阵:

$$\varepsilon^2(m) = \sum_{i=m+1}^{d} \boldsymbol{b}_i^{\mathrm{T}} \boldsymbol{K}_x \boldsymbol{b}_i \tag{9-17}$$

可以证明，使公式（9-17）取得最小值的向量$\boldsymbol{b}_i$满足:

$$\boldsymbol{K}_x \boldsymbol{b}_i = \lambda_i \boldsymbol{b}_i \tag{9-18}$$

也就是说，最优的基向量就是$\boldsymbol{K}_x$的特征向量。在后文中，我们将结合直线拟合的例子证明这一点。

$\boldsymbol{y}$的协方差矩阵很容易与$\boldsymbol{K}_x$建立联系:

$$\boldsymbol{K}_y = E[(\boldsymbol{y} - E\{\boldsymbol{y}\})(\boldsymbol{y} - E\{\boldsymbol{y}\})^{\mathrm{T}}] \tag{9-19}$$

$$= \boldsymbol{B}^{\mathrm{T}} \boldsymbol{K}_x \boldsymbol{B} \tag{9-20}$$

这里$\boldsymbol{B}$矩阵的列向量由基向量$\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_d$组成。

进一步的，若$\boldsymbol{B}$由$\boldsymbol{K}_x$的特征向量组成，则$\boldsymbol{B}$将把$\boldsymbol{K}_x$变成对角阵:

$$\boldsymbol{K}_y = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix} \tag{9-21}$$

把公式（9-21）代入公式（9-17）得到:

$$\varepsilon^2(m) = \sum_{i=m+1}^{d} \boldsymbol{b}_i^{\mathrm{T}} \lambda_i \boldsymbol{b}_i \tag{9-22}$$

由λ_i为标量有:

$$\varepsilon^2(m)=\sum_{i=m+1}^{d}\lambda_i b_i^{\mathrm{T}} b_i \tag{9-23}$$

又由b_i正交得:

222

$$\varepsilon^2(m)=\sum_{i=m+1}^{d}\lambda_i \tag{9-24}$$

这样，我们就可以用一个m维向量y来表示一个d维向量x（$m<d$）:

$$y_i=b_i^{\mathrm{T}}x \tag{9-25}$$

其中b_i是x的协方差矩阵的特征向量。

随机向量在其协方差矩阵的基向量上展开被称为“K-L展开”。

9.2.2 K-L展开的性质

不失一般性，我们把特征向量b_i按照其对应的特征值的大小排序，即，给特征向量标明下标以使其对应的特征值满足下式:

$$\lambda_1\geqslant\lambda_2\geqslant\lambda_3\geqslant\cdots\lambda_d \tag{9-26}$$

然后，我们把与λ_1相对应的b_1称作“主特征向量”。

用超椭球描述数据

如果我们认为x对应的数据点分布在一个超椭球上，那么它的主轴将通过数据的重心，且主轴的方向与K_x的最大特征值对应的特征向量一致，如图9-5所示。这样，K-L变换用椭圆拟合二维数据，椭球拟合三维数据，超椭球拟合更高维的数据。

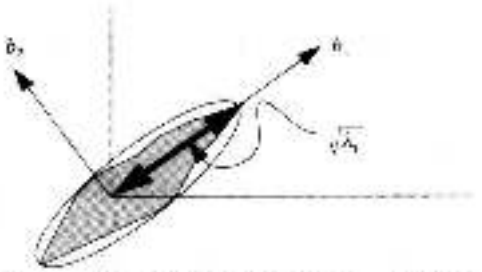


图9-5 一个协方差矩阵可以想像成代表一个超椭球，它的特征向量与椭球各主轴的朝向一致，它的特征向量的平方根对应各主轴的长度

用直线拟合数据

考虑随机向量x的一个样本集:

223

$$\{x_i\}(i=1,\cdots n) \tag{9-27}$$

我们想找到一条最能代表这些数据的直线。首先，我们把原点移到数据集的重心，然后，我们用垂线方向的单位向量n来表示所求的直线。这样，对于每一个数据点x_i，从数据点x_i到最优拟合直线间的垂直距离就等于那一点在垂线方向n上的投影，我们用$d_i(n)$来表示这个距离:

$$d_i^2(n)=(n^{\mathrm{T}}x_i)^2 \tag{9-28}$$

为找到最优的拟合直线，我们求垂直距离平方和的最小值:

$$\varepsilon^2=\sum_{i=1}^{n}d_i^2(n)=\left(\sum_{i=1}^{n}(n^{\mathrm{T}}x_i)^2\right)=\sum_{i=1}^{n}(n^{\mathrm{T}}x_i)(x_i^{\mathrm{T}}n)=n^{\mathrm{T}}\left[\sum_{i=1}^{n}x_ix_i^{\mathrm{T}}\right]n \tag{9-29}$$

在满足n是单位向量的约束下，我们求上式的最小值:

$$n^{\mathrm{T}}n=1 \tag{9-30}$$

我们用拉格朗日乘子来求解约束条件下的最小化问题:

$$\boldsymbol{n}^{\mathrm{T}}\left(\sum_i \boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\right)\boldsymbol{n}-\lambda(\boldsymbol{n}^{\mathrm{T}}\boldsymbol{n}-1) \tag{9-31}$$

定义 $S=\sum_i \boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}$，对$\boldsymbol{n}$求偏微分:

$$\frac{\partial}{\partial \boldsymbol{n}}(\boldsymbol{n}^{\mathrm{T}}S\boldsymbol{n}-\lambda(\boldsymbol{n}^{\mathrm{T}}\boldsymbol{n}-1)) \tag{9-32}$$

对二次型$\boldsymbol{n}^{\mathrm{T}}S\boldsymbol{n}$求偏微分，我们得到$2S\boldsymbol{n}$，令偏微分等于零得到:

$$2S\boldsymbol{n}-2\lambda\boldsymbol{n}=0 \tag{9-33}$$

这与我们前面提到的特征值问题相同。因此我们得到如下结论:

最优拟合直线经过数据集的均值点，并会落在数据集的协方差矩阵的主特征向量的方向上。

到现在为止，我们已经见到了两种寻找最优拟合直线的方法: 一是5.3节中提到的最小二乘法。这种方法适用于直线而不适用于区域，它求数据点到直线的最小垂直距离。本节中提到的方法则求公式（9-29）所示的垂线投影距离的最小值。此外，还有其他方法。例如，[9.53]中找到一种分段线性的表示方法可以保留任意指定阶的矩。 224

我们在很多场合都会用到数据拟合。例如，O'Gorman在[9.54]中不仅研究了拟合直线边缘的问题，而且研究拟合点集、直线、带直线边的区域的问题。这样可以达到分像素级（subpixel）的精度。

接下来，我们将考察一些简单的特征，这些特征可以用来描述区域的形状（要了解更多的特征，请参考[9.2]）。

9.3 简单特征

在本节中，对于图像分割得到的区域，我们要介绍一些描述区域形状的特征。其中的许多特征在图像分割阶段就可以计算出来。例如，由于连通部件标号程序必须遍历区域的每一个像素，因此可以很容易地跟踪区域的面积。

下面罗列了一些同样易于计算的简单特征:

- *平均灰度值*——在黑白“剪影”（silhouette）式的图片中，这很容易计算。
- *最大灰度值*——直接计算即可。
- *最小灰度值*——直接计算即可。
- *面积*（A）——区域内像素的数目。
- *周长*（P）——有几种不同的定义。最简单的一种就是区域内所有与其他区域像素相邻的像素的数目。
- *直径*（D）——直径描述最长弦，即区域边界上相距最远的两个像素之间的距离[9.68, 9.71]。在下一节中，我们将讨论计算这一参数的方法。
- *细长度*（也称紧致度）[⊖]（T）——有两种定义: $T_a=(P^2/A)-4\pi$ 衡量的是周长的平方与面积的比; $T_b=D/A$衡量的是直径与面积的比。图9-6比较了示例区域上两种不同的度量方法。
- *重心*（CG）——区域内N个点的重心坐标可以写成:

⊖ 一些作者[9.69]不愿把紧致度的数学定义与这一定义混为一谈，因此把这一度量称为*等周度量*(isoperimetric measure)。

$$m_x = \frac{1}{N}\sum x, \qquad m_y = \frac{1}{N}\sum y$$

但我们更习惯写成向量形式:

225

$$\boldsymbol{m} = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{9-34}$$

- X-Y纵横比（参见图9-7）——纵横比是区域的长方形边框的长宽比。这是很容易计算的。
- 最小纵横比（参见图9-8）——也是一种长宽比，但是要找到包围区域的最小长方形计算量就大多了。

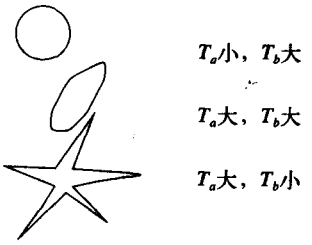


图9-6 不同的紧致度定义下几种区域的度量结果。由于相同面积的区域中圆的周长最短，所以圆的T_a值最小，而同样面积的星形周长就长得多

最小纵横比的计算是很困难的，因为这是一个极值点的搜索问题。但如果我们把区域视为点在椭圆上的分布，就可以得到一个很好的近似值。正如我们在图9-5中讨论的那样，数据点的协方差矩阵的特征值反映了数据在两条正交轴（长轴和短轴）上的分布，两个特征值的比值就是对最小纵横比的很好近似。

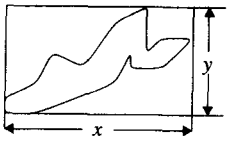


图9-7 y/x是纵横比的一种，即区域的长方形边框的水平边与垂直边的比值

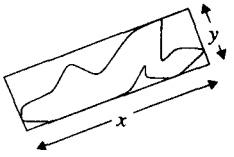


图9-8 y/x是最小纵横比

- 孔洞的数目——是区域中孔洞的数目，这是一个描述能力很强同时又相对容易计算的特征。
- 三角相似度——设P_1, P_2, P_3是区域边界上的三个点，设$d(P_i, P_j)$是两点间的欧氏距离，则$S= d(P_1, P_2)+ d(P_2, P_3)+d(P_3, P_1)$就是三角形的周长。二维向量

$$\left[\frac{d(P_1,P_2)}{S}, \frac{d(P_2,P_3)}{S}\right] \tag{9-35}$$

即三角形中两条边的边长与周长的比定义为三角相似度。这一特征具有旋转、平移和缩放不变性。

- 对称性——在二维空间中，如果一个区域绕某条直线旋转180度后与原区域重合，我们就说这个区域是镜面对称的。这条直线被称为对称轴。如果一个区域绕一点（通常是重心）旋转$2\pi/n$角度后与原区域重合，我们就说这个区域是n阶旋转对称的。这里有两个问题：一是确定对称轴的位置；二是回答这样的问题："它有多对称？"1995年以前，在计算机视觉的应用中，大多数论文在分析区域的对称性时都把它作为一个判定性问题：对称或不对称。Zabrodsky等人在[9.97]中则提出了一个"对称距离"的度量来衡量一个区域有多对称。

226

计算直径

前文中提到的直径若与其他特征结合起来，并用恰当的方法计算，是一种鲁棒的形状度量方法。因此，我们专用一小节来介绍直径的计算过程。

设想如下的情况：假定你在模拟计算机寻找图像中具有特定形状的区域，这样的区域应当是细长的，并有特定的朝向。你将如何定量描述诸如“细”、“长”等模糊特征呢？又将如何确定区域的朝向呢？毕竟区域不过是由区域增长算法输出的一系列点组成。当然在文献中可以找到许多描述区域形状的方法。其中的一个方法就是利用区域的“极点”（extremes）的位置或区域的“直径”。所谓“极点”，就是区域中的两个点A、B，它们之间的距离$d(A, B)$大于或等于区域中任意两个点的距离。

这种计算在需要定量描述图像中区域的形状时可以找到许多应用。例如，组装线上对零件自动分类，光学字符识别等等。此外，在机器人应用中也会用到这种方法，例如机器人为了拾起物体需要先识别区域的朝向。

设一个二维图像平面的区域由点集R构成，问题就变成了寻找$A, B \in R$，使得$\forall p_1, p_2 \in R$，都有$d(A, B) \geqslant d(p_1, p_2)$。如果集合$R$比较小，比如说只有10－20个点，最直接的方法就是计算集合中每一点与其他各点的距离并进行比较。但是，如果随着R的规模变大，比较操作的次数将增长到$n(n-1)/2$次，即达到$O(n^2)$的量级。

本节中，我们将提到一种解决这个问题的技巧，即利用特征值分析的方法找到区域主轴的最优估计（在一般的最小平方意义上），并把该轴上分布在最远端的两个点定义为极点。在9A.1节中还会提到另一种适用于非凸区域的方法，这种方法是启发式的，实现起来有些复杂，但速度很快并保证收敛。

主轴方法

这种方法首先找到区域的主轴以及区域边界上距离主轴最近的两个点。在这种方法中，小的偏离——例如图9-9中的毛刺——将被忽略，尽管这些地方可能包含一个极点。

227

图9-9　带有毛刺的区域

第一步是计算主轴。这里运用了平方误差最小化的技巧。很重要的一点是误差最小化的过程必须与坐标轴的选择无关，因此我们采用了本章中前面提到过的特征向量直线拟合方法。如果区域内的点到一条直线的垂直距离的平方和最小，我们把该直线定义为主轴的最优表示。

假定区域R由点集$R = \{(x_i, y_i) \mid i = 1, \cdots, n\}$表示。设点$(x_i, y_i)$用向量$v_i$表示，点$v_i$到主轴的距离用$d_i$表示。这样，区域的主轴就是使下式取得最小值的直线：

$$d^2 = \sum_{i=1}^{n} d_i^2 \tag{9-36}$$

很容易证明主轴必然经过区域的重心，因此只要找到主轴的倾角即可。因为主轴经过重心，所以我们可以把重心作为坐标系的原点。那么问题变成了：给定n个均值为零的数据点，找到过原点的直线使d^2最小。

这与我们前面描述的特征向量问题是同一问题。因此我们可以通过如下步骤找到主轴：

1）每个点的坐标减去重心的坐标使坐标原点移到重心；

2）找出新坐标系下点分布矩阵的主特征向量；

3）主轴就是通过重心，并与主特征向量方向平行的直线。

找到了主轴后，就可以计算边界上每个点在主轴上的投影。极点就是主轴两侧投影最长的点（极点并不总有唯一解）。这种方法得到了一种最小均方误差意义下对区域形状的一种精确表示。它找到的并不一定是相距最远的两点。在很多应用中，这样的近似正是我们所需要的。然而，有时我们可能会遇到带有毛刺的区域（如图9-9所示），而且毛刺是不能被忽略的。此时，我们要用一种算法找到真正的极点（参见9A.1节）。

228

凸包（convex hull）

如果我们用绷紧的皮带绕区域的边界，皮带围成的区域就称为凸包（参见图9-10）。

一个区域与它的凸包的相差部分被称为凸补（convex discrepancy）。凸包的快速算法参见Shamos[9.68]，计算凸包的并行算法参见[9.30]。

凸包可以有$O(n \log n)$次，因此，寻找凸包有另一种简单的特征：凸补，参见图9-10。

图9-10 区域的凸包，阴影部分是凸补

9.4 矩

形状的矩是很容易计算的，而且我们将会看到，矩对相似变换具有不变性。

区域的$p+q$阶矩定义如下：

$$m_{pq} = \sum x^p y^q f(x,y) \tag{9-37}$$

我们假定区域的灰度是均一的，且区域内的像素灰度值为1，区域外像素的灰度值为0，那么区域的面积就是m_{00}，重心坐标如下：

$$m_x = \frac{m_{10}}{m_{00}} \qquad m_y = \frac{m_{01}}{m_{00}} \tag{9-38}$$

从矩可以派生出一系列类似的度量（中心矩），这些度量对平移变换具有不变性：

$$\mu_{pq} = \sum (x-m_x)^p (y-m_y)^q f(x,y) \tag{9-39}$$

把旋转和缩放考虑在内，可以派生出更多类似的度量，而且通过选择高阶的矩，并将它们组合起来，可以定义任意多种特征。由中心矩可以进一步定义规范化中心矩（normalized central moments）：

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \qquad \gamma = \frac{p+q}{2}+1$$

最后，不变矩（invariant moments）[9.21]具有平移、旋转、和比例变换不变性，这就是说即使图像经过了平移、旋转或缩放，我们仍得到相同的矩[⊖]。这些不变矩列在表9-1中。

229

它们最初由Hu在[9.33]中提出，此后Ruthe等人在[9.62]中又提出了对仿射变换具有不变性的矩。

尽管矩的特性很吸引人，但基于矩的方法仍存在问题，它们中的相当一部分对量化和采样非常敏感[9.45]（参考作业9.9）。

⊖ Gonzalez 和 Wintz 在[9.21]中将缩放称为"比例变换"，但我们使用至"比例"一词意义相有不同，因此我们更愿意使用"缩放"一词。

表9-1 不变矩

$$\varphi_1 = \eta_{20} - \eta_{02}$$
$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2$$
$$\varphi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + (3\eta_{21} - \eta_{03})(\eta_{03} + \eta_{21})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2]$$
$$\varphi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$\varphi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2]$$

图像匹配中有一类称为“规范化方法”（method of normalization）的通用方法，矩方法实际上是其中的一种。这类方法的基本原理是首先对区域内的所有点进行变换（通常是线性变换）以便把区域纳入一种规范的框架。最简单的变换就是所有点的坐标都减去重心的坐标，这样坐标原点就移到了区域的重心。在更一般的情形中，这类变换可能是仿射变换，包括平移、旋转和剪切。然后在变换域内进行匹配，因为在变换域内同一类物体（例如三角形）看起来都相同。

如果要计算灰度图像的矩，还需要对算法进行改进，此时公式（9-37）中的f值就不是阈值化后的结果了。所有的不变性理论仍然成立，不过正如Gruber和Hsu在[9.24]中指出的那样，噪声对矩计算污染的程度与数据有关。

一旦程序抽取了一些特征，这些特征总会有用武之地，或者是匹配两个观察值，或者把观察值与模型进行匹配。在13.2节中将介绍匹配中一些简单特征的用法。

9.5 链码

链码是一种描述区域边界的特征。在链码中，我们用数字序列表示按逆时针方向沿边界前进时，每一步的方向。 230

如果使用八个方向，所有的链码都在0～7之间，如果使用四个方向，所有的链码都在0～3之间。在表示边界时，链码指示了每一步的方向。链码的八个主方向和四个主方向的定义如图9-11所示。这样每个区域的边界就可以用一串数字来表示。一种更紧致的表示方法是在某个方向重复出现时使用上标。例如，0012112 776660可以表示成$0^2121^2 27^2 6^3 0$，它所表示的边界如图9-12所示。用符号序列表示边界的能力在“语法模式识别”（syntactic pattern recognition）领域扮演了重要的角色，并在机器视觉的文献中频繁出现，包括本书的其他地方。

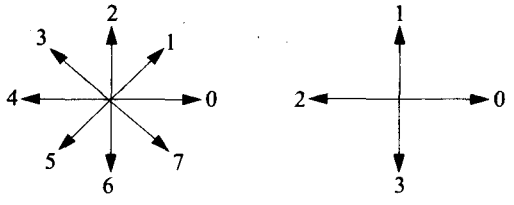


图9-11 沿边界前进时下一个像素与当前像素的八种方向关系（8邻域）和四种方向关系（4邻域）

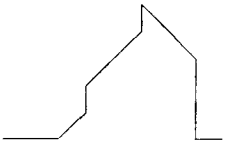


图9-12 由链码$0^2121^2 27^2 6^3 0$表示的边界片段

9.6 傅里叶描述子

傅里叶描述子是另一种描述区域边界的特征。给定区域的边界，我们假定[⊖]区域位于一个复数平面上，每一点的x坐标代表在实数轴上的投影，y坐标代表在虚数轴上的投影，因此边界上的每个点都可以看作一个复数。遍历边界就得到一个轮转的复数序列。对这个序列进行傅里叶变换得到另一个复数序列，可以证明这个序列具有表9-2所示的一系列不变性。

下面就是一个这样的例子，尽管有些过于简化。

假定我们有两个边界的傅里叶描述子如下:

$f_1 = 0.7, 0.505, 0.304, 0.211, ...$

$f_2 = 0.87, 0.505, 0.304, 0.211, ...$

可以看出这两个序列只有第一项直流分量不同，因此它们代表的两个边界之间只差一步平移变换。

231

表9-2 在图像域和变换中同一种变化的对应关系

图像域	变换域
尺寸的变化	乘以一个常数
绕原点旋转角度Φ	相位平移
平移	直流分量的变化

这个例子过于简化了，因为真实的序列是一系列复数，而不像演示中那样只有实部，但原理上是一样的。

在实际使用傅里叶描述子时应考虑的问题

我们表示边界移动的方式是至关重要的。仅用四邻域链码得到的结果很差。用八邻域链码可以减少40%～80%的错误，但仍不如使用亚像素插值得到的结果好。此外，还有其他的复杂问题，包括一些常用的边界参数（例如弧长）不具有仿射变换不变性[9.1, 9.96]。[9.39]中的实验比较了具有仿射变换不变性的傅里叶描述子和自动递归方法。要了解更多关于傅里叶描述子的内容，请参考[9.1]。

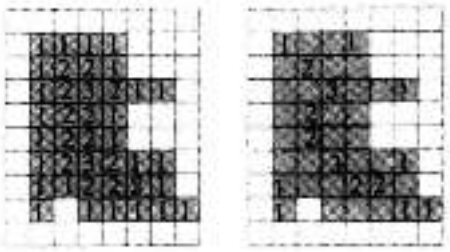
9.7 中轴

在二维空间中，一个区域的中轴定义为“最大圆”的圆心的轨迹。最大圆是指以区域内给定点为圆心且区域内能够容纳的最大的圆。让我们说得更准确些（我们都需要学习如何用数学工具使我们的语言更严谨）。给定区域Ω中的一点(x, y)，绕该点画半径为R的圆，让R尽可能大但要满足：1）圆周上的任何一点都不能超出区域的范围；2）圆周与区域边界至少有两个交点。可以证明中轴上的每一点都是距离变换DT(distance transform)的一个局部极大值点。如果一个点的DT值为k，且它的所有相邻点的DT值都不超过k，则称该点取得了DT值的局部极大值。图9-13a与图7-3相同，这个DT变换的局部极大值点如图9-13b所示。

我们也可以把中轴想像成静电场的电压。如果边界刚好是直线或者三维空间中的平面[9.21]，这种想法更容易实现。在[9.16]中可以找到其他的计算均值的高效方法。

232

⊖ 注意，本段内的数据类型，很容易忽视。



a)　　b)

图9-13　用四邻域法计算的区域的DT值a) 用形态学方法抽取的骨架；由DT值的局部极大值点组成b)

中轴与脊的联系

中轴的定义仅适用于二值图像，即仅适用于区域可以由边界表示的图像，而脊是为灰度图像定义的。我们可以用如下方法把这两种特征联系起来：用不同的尺度表示图像，也就是说，把二值图像模糊化。这样的过程可以把二值图像变成灰度图像，然后再来找脊。首先在尺度最大（最模糊）的地方找脊；你可以把找到的脊作为中轴的初值。现在降到稍低的尺度，看看出现了哪些新的脊，把这些新的脊加入到你的估计中去。在不同的尺度空间重复这一过程。要想了解这个方法的原理，请参考Pizer等人的[9.59]。

9.8　变形模板

还记得第8.5节中讲过的主动轮廓（snake）吗？如果把区域想像成被一条蛇围成的，而不是把区域本身想像成蛇，那么区域就可以变形了，因此发明了“变形模板”（deformable template）。利用模板可以变形的思想可以跟踪物体[9.102]。此外，变形模板在图像数据库的存取中也很有用。例如，Bimbo和Pala通过比较图像和“图标索引”来检索图像[13.5]。这种由用户提供的“图标索引”也是变形模板的一种。最优的匹配模板可以写成：

$$\phi(s) = \tau(s) + \theta(s) \tag{9-40}$$

其中s是弧长（规范化的），$\tau(s)$是数据库中存储的模板，$\theta(s)$是使特定模板与读取的图像中的边界点序列匹配所需要的形变。重点在于$\theta(s)$是初始模板与变形后模板之间的差。数据库中与模板最匹配的图像就是使“初始模板与变形后模板间的差”最小的图像，使下式取最小值可以找到最优匹配图像：

$$E = \int_0^1 \left\{ \alpha \left[\left(\frac{d\theta_x}{ds} \right)^2 + \left(\frac{d\theta_y}{ds} \right)^2 \right] + \beta \left[\frac{d^2\theta_x}{ds^2} + \frac{d^2\theta_y}{ds^2} \right] - I_G(\varphi(s)) \right\} ds \tag{9-41}$$

上式中的第一项表示初始模板需要经过怎样的变形才能与物体相匹配，第二项表示模板变形所消耗的能量，因此这是一个变形模板问题。这个优化问题可以用数值方法求解[13.5]。

变形模板的一个变种就是“几何流”（geometric flow）的思想。几何流的方法通过改变初始曲线的形式使之更适合于物体识别或模板匹配。几何流中的“几何”一词是指，流完全由曲线的几何形状决定。Pauwels等人在[9.58]中用这一准则回答了如下的问题：“能否用函数优化的方法刻画曲线的几何形状，使其在强化显著特征的同时又减少噪声？”

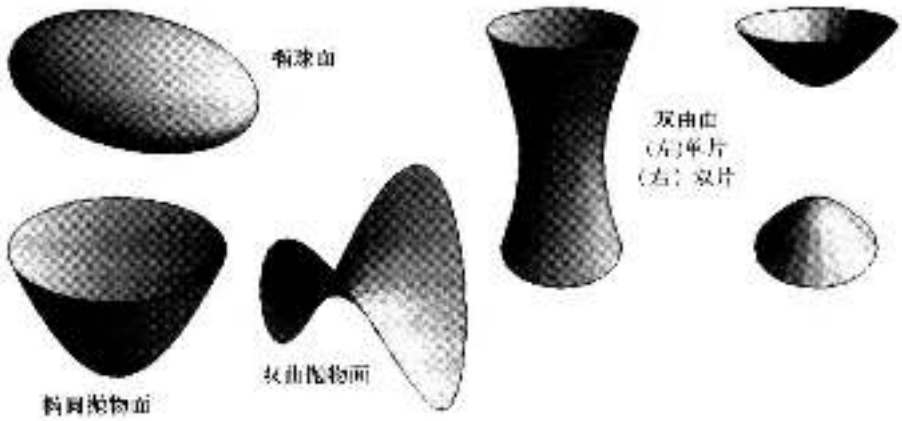
233

9.9 二次曲面

用二次代数方程描述的曲面被称为二次曲面。二次曲面的通用形式如下：

$$ax^2 - by^2 + cz^2 + 2fyz + 2gzx + 2hxy + px + qy + rz + d = 0 \tag{9-42}$$

这一形式的方程描述了所有的二次曲面，其中的几种如图9-14所示：



234 图9-14 二次方程可以描述很多类型的曲面[9.103]（使用该图片得到了CRC出版社的许可）

如果二次曲面以原点为中心，而且主轴恰好与坐标轴平行，那么相应的二次方程的形式就比较特别。例如，这样的椭球面的方程如下：

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{9-43}$$

但是，如果二次曲面的轴不与坐标轴平行，那么相应的方程就只能是公式（9-42）的一般形式。

从深度数据或其他表面数据可以计算出二次曲面方程的系数。第8.6.1节就介绍了这样的方法。系数确定后，二次曲面的类型可以用如下方法判定[⊖]。

如果方程中有常数项d，则各项同时除以这个常数，得到常数项为1的二次方程：

$$ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy + px + qy + rz + 1 = 0 \tag{9-44}$$

上面的方程可以改写成：

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} a & h & g & p \\ h & b & f & q \\ g & f & c & r \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = 0 \tag{9-45}$$

考察左上角的3×3子阵：

$$E = \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix}$$

⊖ 作者感谢G. L. Bilbro关该方法提供的公式。

计算出它的三个特征值λ_1、λ_2和λ_3并求出所有非零特征值的倒数$r_1=1/\lambda_1$，$r_2=1/\lambda_2$，$r_3=1/\lambda_3$。如果方程表示的是真实的曲面，那么至少有一个特征值的倒数为正。如果只有一个倒数为正，表示的是双片双曲面；如果只有两个倒数为正，表示的是单片双曲面；如果三个倒数都为正，表示的是椭球面，并且r_1、r_2、r_3的平方根就是椭球的主轴。另外，双曲面的焦距由这些r的模决定。 235

9.10 表面的谐波函数表示

这种方法用确定的谐波基函数的线性组合来表示表面。谐波函数是拉普拉斯方程的解:

$$\nabla^2\psi(x,y,z)=0 \tag{9-46}$$

在笛卡儿坐标系中可以写成:

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}=0 \tag{9-47}$$

但是大部分谐波函数表示的工作都没有采用笛卡儿坐标系，而采用了球面坐标系（关于其他表示形式的讨论参见Matheny和Goldgof[8.40]）。任何可以写成$r=r(\theta,\varphi)$形式的连续函数都可以表示成球面谐波函数的线性组合。在球面坐标系中，拉普拉斯方程如下:

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2}=0 \tag{9-48}$$

我们只对ψ函数能分解成单变量函数乘积的情况求解。此时，ψ函数可以分解成:

$$\psi(r,\theta,\varphi)=R(r)\Theta(\theta)\Phi(\varphi) \tag{9-49}$$

在这一条件的限制下，偏微分方程能够分解成三个普通的微分方程，且解的形式如下:

$$P_l^m\cos\theta\sin m\varphi \tag{9-50}$$

其中参数l被称为“度”，m是一个比l小的整数，P是一个勒让德多项式（Legendre polynomial）。

这样任何函数都可以表示成:

$$r(\theta,\varphi)=\sum_{l=0}^{L}\left\{U_l^0P_l\cos\theta+\sum_{m=0}^{l}[U_l^mP_l^m\cos\theta\cos m\varphi+V_l^mP_l^m\cos\theta\sin m\varphi]\right\} \tag{9-51}$$

其中的系数可以通过数据拟合得到。

9.11 超二次曲面

在8.6节我们讨论了用表面函数分割深度图像的问题，并介绍了如何根据数据拟合函数。在本节中，我们将描述如何用超二次曲面（superquadrics和hyperquadrics两种超二次曲面是不同的，但在中文中不易区分，故以下在必要时用原英文名加以区分——译者注），拟合深度数据。 236

超二次曲面（superquadric）是如下方程对应的表面:

$$\left|\frac{x}{a}\right|^{\gamma_1}+\left|\frac{y}{b}\right|^{\gamma_2}+\left|\frac{z}{c}\right|^{\gamma_3}=1 \tag{9-52}$$

而超二次曲面(hyperquadric)的方程如下:

$$\sum_{i=1}^{N}\left|A_i x + B_i y + C_i z + D_i\right|^{\gamma_i} = 1 \tag{9-53}$$

上面的方程简写成$F(x, y, z) = 1$。

Kumar等人在[8.32]中提出了一种用hyperquadric拟合深度数据的方法。它的操作过程如下：首先要意识到使下式取得最小值的参数，就是一个拟合得相当好的函数，因为在表面F的每一点上，函数F的值都应当是1.0。

$$\mathrm{EOF} = \sum_{i=1}^{N}(1 - F(x_i, y_i, z_i))^2 \tag{9-54}$$

要注意的是这使得点到表面的代数距离（参见8.6.1节）取得最小值。如果点落在表面上则代数距离（同时欧氏距离也）为零，否则点到表面的欧氏距离和代数距离之间不存在简单的对应关系。Kumar等人还发现："这个函数是有偏差的，特别是遇到拉长的物体时。"在代数距离的所有应用中都不可避免地存在这个问题。

要想得到好一些的拟合，在拟合函数是hyperquadric时，可以使用下面的方法。假定我们对表面有一个初始估计，这个估计不太坏，但也许不是最佳的。设这个估计是由一组参数A、B、C和D定义的函数$F(x, y, z)$。对一个特定点(x_i, y_i, z_i)，把坐标值代入F得到$w_i = F(x_i, y_i, z_i)$。如果这点落在表面上，则w应当为1。现在考察一下由$F(x, y, z) = w_i$定义的表面。到这一表面的垂直距离可以用梯度方向的距离d_i来近似。

$$F(x, y, z) = F(x_i, y_i, z_i) + d_i\left\|\nabla F(x_i, y_i, z_i)\right\| \tag{9-55}$$

从上式得：

$$d_i = \frac{1 - F(x_i, y_i, z_i)}{\left\|\nabla F(x_i, y_i, z_i)\right\|} \tag{9-56}$$

这里的d_i是我们真正想要最小化的。这是一个经过点(x_i, y_i, z_i)，与我们要确定的表面大致平行的表面，到要确定的表面之间的距离。这些还只是估计值，我们需要不断地迭代修正这些估计直至得到真正的解。为做到这一点，把平方误差用d_i项改写成：

237

$$E = \sum_{i=1}^{N}\frac{(1 - F(x_i, y_i, z_i))^2}{\left\|\nabla F(x_i, y_i, z_i)\right\|^2} \tag{9-57}$$

（分母的平方是译者修改的。——译者注）

并按照下面的步骤最小化目标函数：

1）首先，选定初始估计，很可能是最小化公式（9-54）的数值解。这个估计并不太差，但由于是有偏估计，所以可以拟合得更好。

2）对每一个数据点，计算$w_i = 1/\|\nabla F(x_i, y_i, z_i)\|^2$。

3）最小化$\sum_{i=1}^{N}(w_i(1 - F(x_i, y_i, z_i)))^2$。

4）如果解已经足够好，结束；否则转到2）。

可以发现这里所演示的是一种拟合隐函数的通用想法，而不仅局限于超二次曲面(hyperquadric)。对于过数据点的曲线的一个垂直于其梯度方向的函数进行最小化，能够解决很多甚至大部分显函数拟合的问题。

Dickinson等在[13.11]中把Superquadric表示方法和外观图(aspect graphs)的概念结合起来了。

9.12 广义柱体

圆柱体可以描述成一个圆沿着一条垂直于圆面的直线平移时所经过的区域。如果设想一下那条直线是可以弯曲的任意一条空间曲线，并可以用弧长参数s表示,这条曲线就变成了s的向量函数$x(s), y(s), z(s)$。然后，允许圆的半径随着在曲线上的位置而变化，即$R = R(s)$。那么，你对广义柱体就有一些认识了（参见文献[9.4, 9.21,9.23, 9.74, 9.75, 9.98]）。但是广义柱体(generalized cylinder)的概念比上面描述的还要广泛。沿着曲线移动的物体不一定是圆，可以是任何二维形状。

如果我们用广义柱体去匹配区域，我们可以用曲线的参数方程和圆的半径方程作为特征描述区域的形状。然而，用广义柱体匹配区域时会遇到重大的挑战。本书中对广义柱体不做进一步的讨论。读者可以从文献中找到很多有意思的论文，其中的一部分列在上一段中。

9.13 结论

本章中定义了一些可以定量描述区域形状的特征。有些特征，例如矩，是很容易得到的度量。另一些特征，例如直径和凸补，需要开发相当复杂的算法以避免非常耗时的计算。还有一些工作致力于形状描述子的自动学习。 238

对视觉皮层信号处理的研究表明图像在视觉皮层的表示经过了对数极坐标变换(log-polar transformation)，因此具有不变性[9.63, 9.90]。还有一些变换[9.87]可以为计算机应用提供等价的甚至是更优的表示。

在本章中，我们再一次看到了一致性原理的应用，如果一个像素集中的每个点都满足同一个方程我们就说这些像素满足一致性，不论满足的方程是二次曲面还是广义圆柱体。

最优化方法在本章中出现了好几次:

- 在9.2.2节中演示了一个检测集的最优拟合直线的推导，这里的最优是指在垂线上的投影之和最小。为求解这一问题，我们要借助拉格朗日乘子法求解约束条件下的最小化问题。
- 在9.8节中我们利用积分平方误差最小化检测模板的形变。
- 在9A.2节中，我们将遇到一个非方阵的求“逆”问题。当然，这样的矩阵无法求逆，但可以推导出一个“伪逆”（pseudo-inverse）。我们还将证明伪逆实际上是使平方误差最小化的算法。

9.14 术语

仿射变换	affine transform
纵横比	aspect ratio
基向量	basis vector
重心	center of gravity
链码	chain code
紧致性	compactness
凸包	convex hull
凸补	convex discrepancy

变形模板	deformable template
直径	diameter
傅里叶描述子	Fourier descriptor
广义柱体	generalized cylinder
齐次变换矩阵	homogeneous transformation matrix
不变矩	invariant moment
K-L变换	K-L transform
线性变换	linear transformation
中轴	medial axis
度量	metric
矩	moment
正交变换	orthogonal transformation
主分量	principal component
相似变换	similarity transform
细度	thinness

239

专题9A 形状的描述

9A.1 求非凸区域的直径

一个估计直径的快速算法如下:

1）在边界上任选一点，设为P_0;

2）在边界上找到距离P_0最远的点，设为P_1。

3）计算P_0和P_1的中点M。

4）在边界上找到距离M最远的点P_2。

5）在边界上找到距离P_2最远的点P_3。

6）如果$d(P_2, P_3) < d(P_0, P_1)$，$P_0$和$P_1$之间的距离就是极大值，结束；否则，令$P_0 \leftarrow P_2$, $P_1 \leftarrow P_3$，跳到（3）。

这一方法很容易编程实现，而且收敛得很快。在这一点上我们很走运。不幸的是，这个算法不保证收敛到全局的极值点，尽管有很大的可能性。

上文中的算法经过扩展提供了一种保证收敛的策略。另外，这个扩展的算法还提供了一种迅速缩小搜索空间的机制。

首先，定义一个线性搜索函数$M(P_i, R)$，这个函数的返回值$P_{i+1} = M(P_i, R)$满足$\forall x \in R, d(P_i, x) < d(P_{i+1}, P_i)$。

任选一点$P_1 \in R$，找到$P_2 = M(P_1, R - P_1)$。然后，找到$P_3 = M(P_2, R - \{P_1, P_2\})$。$d(P_1, P_2)$和$d(P_2, P_3)$之间的关系只有下面三种可能，如图9-15至图9-17所示:

情况1: $d(P_1, P_2) > d(P_2, P_3)$

情况2: $d(P_1, P_2) = d(P_2, P_3)$

情况3: $d(P_1, P_2) < d(P_2, P_3)$

在第3种情况中，我们计算一个新的P_3，设为P_3'，使其满足$d(P_2, P_3') = d(P_2, P_3)$，且$P_1$,

P_2和 P_3' 在同一条直线上。这样，最多经过两次线性搜索，我们就得到了一个覆盖了R中所有点的对称的透镜形区域。 240

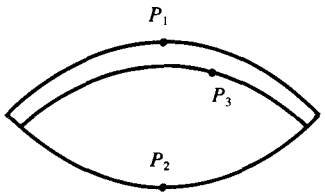


图9-15　情况1

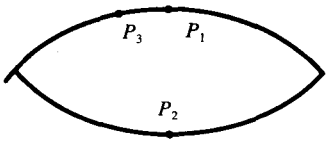


图9-16　情况2

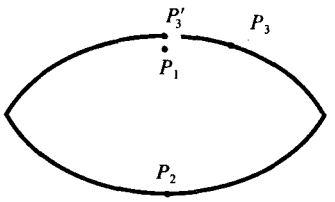


图9-17　情况3

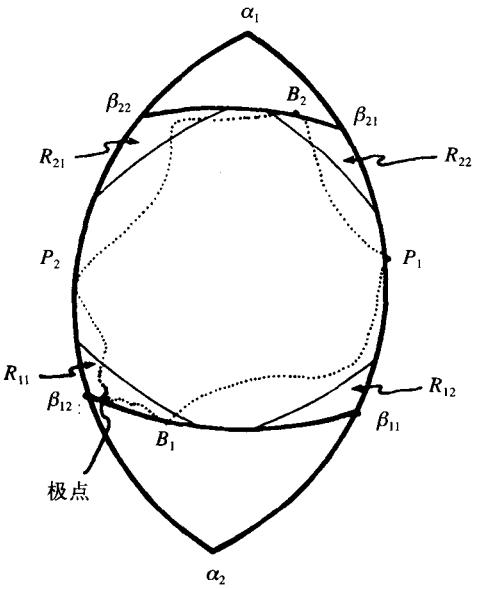


图9-18　在本图中，B_2是一个极点，另一个极点是R_{11}中的一个元素（根据[9.71]重新绘制）

我们的启发式搜索策略认定“最好”的搜索方向在垂直于P_1P_2的方向上（情况3中垂直于P_2P_3'）。

计算图9-18所示的顶点α_1和α_2。然后找到$B_1=M(\alpha_1,\{R-\{P_1,P_2\}\})$。如果$d(\alpha_1,B_1)\leqslant d(P_1,P_2)$，停止，否则把$R-\{P_1,P_2\}$分成两个互斥的区域$R_1$和$R_2$，其中 $R_1=\{x\in R-\{P_1,P_2\}\,\|\,d(x,\alpha_1)>d(P_1,P_2)\}$（译者注：原文中为 $d(x,P_1)>d(P_1,P_2)$。找到$B_2=M(\alpha_2,R-R_1)$和 $R_2=\{x\in R-\{P_1,P_2\}-R_1\,\|\,d(x,a_2)>d(P_1,P_2)\}$。

注意：1）一般说来，$R_1\cup R_2\subset R$。但是，对 $R-R_1\cup R_2$ 中的点我们不感兴趣。2）如果$R_1=\Phi$或$R_2=\Phi$，我们可以停止搜索，且P_1P_2就是直径。

如果R_1和R_2都非空，我们把它们定义为“对跖区域”（antipodal region）。对跖区域的含义如下：如果存在比$d(P_1,P_2)$长的直径，那么必然有一个端点落在R_1中，另一个端点落在R_2中。这样可以把搜索空间划分成两部分。 241

以α_1为圆心，$d(\alpha_1,B_1)$为半径作弧，弧与透镜形区域的交点标记为β_{11}和β_{12}。同理可得β_{21}和β_{22}。注意 $d(\beta_{21},\beta_{12})=d(\beta_{11},\beta_{22})$，这就是直径的上界。在数字图像中，得到上界后可以早些停止算法，因为如果我们知道了某个候选直径和上界，并且二者之差小于1.414（像素区域

的对角线长度），就没有必要继续搜索了。

计算 $r = \mathrm{MAX}(d(B_1, B_2), d(P_1, P_2))$，用这一直径作弧。以$\beta_{21}$为圆心，$r$为半径作弧可以把$R$分成两个区域$R_{11}$和$R_1 - R_{11}$。同样的，以$\beta_{22}$为圆心，$r$为半径作弧可以找到 $R_{12} \subset R_1$。注意 $R_1 - \{R_{11} \cup R_{12}\}$ 中不包含我们感兴趣的点。同理可得R_{21}和R_{22}。如果 $R_{21} \neq \Phi$（译者注：原文中为$R_{21} = \Phi$）且 $R_{11} \cap R_{12} = \Phi$，那么$R_{11}$和$R_{12}$就是一对对跖区域。$R_{21}$和$R_{22}$的情况类似。在任何情况下，$R_{11} \cap R_{12}$和 $R_{21} \cap R_{22}$都是对跖区域。这些对跖区域是我们下一阶段的搜索空间。

到此为止我们要么已经确定了极点的位置，要么得到了其他有用的结果，特别是：

1）直径的上界已经得到。

2）一个相当大的区域内的所有点都已经从极点候选中排除。

3）其余点被分割成对跖区域。

此时，如果剩余点的数目少于K（通常情况都如此），最合适的方法就是计算凸包。（K的最优值是区域拓扑学研究的问题。我们经验是$K = 50$似乎是个不错的选择。）这步计算可以利用以下观察结果：

- 如果对跖区域中有任何一个集合为空，则另一个集合可以从候选中排除。
- 极点必然位于对跖区域的凸包的两侧的位置，这样又可以排除更多的点，因为我们可以计算更小的区域的凸包。

另一方面，如果剩余的点很多，就必须递归地调用算法，把成对的对跖区域作为搜索区域，选择那些靠近，β_{21}、β_{12}、β_{11}和β_{22}的点作为新的起始点。

对于有N个点的区域R，盲目的穷举搜索要进行$O(n^2)$次距离计算和比较。我们的算法也是穷尽整个空间的（因此保证收敛），但由于考虑了区域的整体形状而变得更加智能。通过不断排除不可能位于直径端点的点，初始的搜索空间R被划分成较小的互斥的子空间。

尽管每次递归调用算法后子空间的数目都会增加两个，但每次都可以从搜索空间中排除很多点。因此，搜索空间迅速地收缩。收缩的速度取决于区域R的形状。

这个方法来自几何方面的考虑，因此它收敛的速度与区域的几何形状有密切的关系。所以很难精确地计算这个算法的复杂度。仅做预处理就要进行$O(4n)$次计算，还要加上计算剩余点的凸包的时间，即$O(k \log k)$（其中k表示预处理后剩余点的数目）。

242

当然在最坏的情况下，几乎没有点被排除，最终计算n个点的凸包的时间是$O(n \log n)$。事实上，在这种情况下，这个方法比直接计算凸包还慢，因为程序更加复杂。

但是，由于分支点的数目很多，算法在几乎所有的区域都会很快退出，迅速收敛。

9A.2 从图像推测三维形状

人们写了很多从各种线索提取三维形状的论文：从轮廓[9.42, 9.43, 9.44, 9.49, 9.101]；从镜面反射体的图像[9.64]；从三个正交投影（X射线投影）[9.81]；利用物体一般有正交性[9.22]或对称性的假定[9.18]。最后，在所有这些算法中，还必须重视可见性的问题[9.83]。

人们可能会认为深度图像已经包含了完整的三维形状描述，但是你无法从一幅图像中看到整个表面[9.31]。因此一个很困难的问题就是如何把几幅深度图像结合起来形成三维物体的描述[9.76, 9.93]。

尽管你可能已经在一定程度上成功地分割了表面，但是这些分割几乎从来都不是完全正确的。你可能认为边缘交点的方程可以直接找到，因为你可能认为毕竟你已经有了表面方程，

而正是表面的交集决定了边缘，因此只要计算交集就可以了！但是事情从不是那么简单。当你计算顶点的时候问题出现了——顶点是边缘的交点，而这里的边缘又是三面体或多面体各面的交点。你推出的方程从不会交于一点。Hoover等人在[9.31]中重点讨论了这个问题，并扩展了不可见表面的解决方法。

提取三维形状的另一个重要问题是选择什么样的表示，包括从透视关系推测三维形状、从阴影推测三维形状、从纹理推测三维形状等。

9A.2.1 从透视关系推测三维形状

我们把从透视关系推测三维形状的问题形式化描述如下：一个实际上位于三维空间中$[x_i, y_i, z_i]$的点被观察到落在图像的r_i行和c_i列。观察坐标和真实坐标之间的关系如下：

$$\begin{bmatrix} r_i \\ c_i \\ 1 \end{bmatrix} = \begin{bmatrix} k_u f & 0 & u_0 & 0 \\ 0 & k_v f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \tag{9-58}$$

其中k_u、k_v、u_0、v_0和f是照相机的投影参数，R是一个3×3的旋转矩阵，T是一个3×1的平移向量。假定我们知道每一点的实际三维坐标和相应的二维观察值，我们应当能推出变换矩阵和照相机参数。这个问题有若干种变形[9.95]：已知n对对应点时的n点透视问题（PnP）；已知n对对应直线时的n线透视问题（PnL）；已知n对对应角时的n角透视问题（PnA）。P3P[9.17]、P3L[9.14]和PnA[9.95]问题的解析解已经得到。在[9.26]中可以找到照相机未经校准条件下（但假定已经解决了准确对应的问题）的线性方法的详细解释。 243

对于更加困难条件下的对应问题，人们也做了一些工作[9.34]。在这种情况下，不仅照相机没有校准，而且照相机的拍摄角度导致外极线假设都未必成立。此时，立体视觉匹配问题变成了一个搜索最优匹配对的问题，要同时用上辐射能测量信息和几何信息来缩小搜索空间。

9A.2.2 从影调推测三维形状

从影调推测三维形状最初是由Horn提出的，他指出一些关于光线如何产生、反射和被观察到的知识可以提高机器视觉系统的性能。考察图9-19，并假定已经知道：

- 光源的角度
- 观察者的角度
- 测量得到的像素的亮度
- 光散射的定律
- 表面反射率

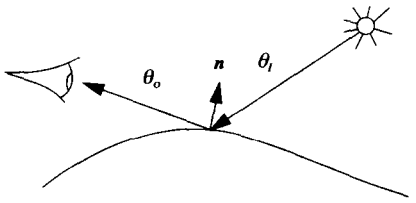


图9-19　光线与表面法向成一定的入射角，被反射/散射到其他方向

你能得出表面的法线方向吗？（如果你得到了每一点的法线方向，又如何确定表面呢？）

通过解微分方程可以找到答案。首先我们把表面法线向量写成$\boldsymbol{n} = \boldsymbol{r}/|\boldsymbol{r}|$，其中方向向量

$$\boldsymbol{r} = \left[\frac{\partial z}{\partial x}, \ \frac{\partial z}{\partial y}, \ 1 \right]^{\mathrm{T}}$$

在大多数从影调推测三维形状的文献中，偏微分简写成 $p \equiv \dfrac{\partial z}{\partial x}, q \equiv \dfrac{\partial z}{\partial y}$ 。

尽管我们用了“亮度”这个术语，但它实际上没有严格的物理定义。按照Horn[9.32]，我们把“辐照度”（irradiance）定义为落到表面单位面积上的能量，度量单位是瓦/平方米。然后我们可以定义“辐射率”（radiance）为单位立体角内单位透视面积上的能量。辐射率依赖于透视面积意味着观察角度在决定景物的“亮度”时扮演着重要的角色。

244

通常表面的反射模型是已知的或可度量的。例如，观察到的亮度与观察角度无关，而取决于入射光的角度。例如，反射亮度与入射光的亮度有如下关系：

$$R(x,y)=aI(x,y)\cos(\theta_I) \tag{9-59}$$

这样，如果我们知道入射光的亮度I，表面反射率a（表面材质）和反射光的亮度R，我们就能解出θ_I，并由此推出表面法向，进而求出表面。公式（9-59）中的反射函数称为朗伯模型（Lambertian model）。注意朗伯模型中不涉及观察角度。另一个我们熟悉的反射函数是镜面模型：

$$R(x,y)=aI(x,y)\delta(\theta_I-\theta_O) \tag{9-60}$$

它描述了镜面反射——仅当观察角度与入射角度相等时才能观察到反射光。当然大部分表面尽管“闪闪发亮”，却不是完全的镜面反射体，因此一个也许更实际的混合表面模型是：

$$R(x,y)=aI(x,y)\cos^4(\theta_I-\theta_O) \tag{9-61}$$

尽管要利用反射函数，需要先对照相机进行辐射能测量校准，但这个要求本身并不是主要的困难。为了考察这个问题的复杂度[9.51]，让我们把公式（9-59）中的入射角展开为入射光方向向量与法线向量的点积（原文中为观察向量和法线向量的点积——译者注）：

$$R(x,y)=aI(x,y)\cos(I\cdot N)=aI(x,y)\left(I_x\frac{\partial z}{\partial x}+I_y\frac{\partial z}{\partial y}+I_zN_z\right) \tag{9-62}$$

假定我们知道入射光（原文中为观察的角度——译者注）的角度（实际上我们最多只能知道大致的方向）、表面反射率和入射光强，我们仍然需要解一个偏微分方程才能确定表面方程z。

许多论文和Horn的经典教科书[9.32]都致力于求解从影调推测三维形状问题的各种特例。在最近的一篇论文[9.100]中Zhang等人综述了到1999年为止这个领域中的工作。（别忘了，公式（9-62）也是一种特例——它假定亮度与观察方向无关。）下面，我们将讨论另外一种特例——光度测量立体视觉（photometric stereo）。

光度测量立体视觉

在很多情况下，认为表面的反射光强与表面法向和入射光方向之间夹角的余弦成正比是合理的：

$$I(x,y)=r_o(\boldsymbol{N}_i\cdot\boldsymbol{n}) \tag{9-63}$$

其中N_i是光源i方向的单位向量。如果我们足够幸运有一个具有朗伯反射表面的物体，即满足这个方程，同时表面具有相同的反射率r_0，与入射光线无关，我们就可以利用不同角度光照下的多幅图像确定表面法线方向[9.35, 9.94]。假定我们用三个不同的光源照射一个特定点（每次用一个光源），并分别测量该点对应像素的亮度。在该像素点，我们得到一个由三个观察值组成的向量：

245

$$\boldsymbol{I}=[I_1,I_2\ I_3]^{\mathrm{T}} \tag{9-64}$$

我们知道每个光源的方向。假定这些方向由从表面点到光源的单位向量表示，并记为N_1、

N_2和N_3。把这三个向量写在一个矩阵中，每个向量占一行:

$$N=\begin{bmatrix}N_1\\N_2\\N_3\end{bmatrix}=\begin{bmatrix}n_{11}&n_{12}&n_{13}\\n_{21}&n_{22}&n_{23}\\n_{31}&n_{32}&n_{33}\end{bmatrix}\tag{9-65}$$

现在我们把公式（9-63）写成矩阵形式:

$$\boldsymbol{I}=r_0\boldsymbol{N}\boldsymbol{n}\tag{9-66}$$

因为$\boldsymbol{N}$已知，且$\boldsymbol{n}$是单位向量，我们得到

$$r_0=|\boldsymbol{N}^{-1}\boldsymbol{I}|\tag{9-67}$$

一旦我们得到r_0了，就可以解出$\boldsymbol{n}$了:

$$\boldsymbol{n}=\frac{1}{r_0}\boldsymbol{N}^{-1}\boldsymbol{I}\tag{9-68}$$

注意这一光度测量立体视觉的推导中，假定对每个角度的表面反射率（albedo）都相同。在下面的小节中，我们将演示把从影调推测三维形状和光度测量立体视觉结合起来的一个应用，那时将不需要做这样的假设。例如，镜面反射提供了一个特殊条件——观察角度与入射角刚好相等。这使得特别的技术可以得到应用[9.56]。

但如果我们用三个以上光源呢？这给了我们一个绝好的机会讨论一个重要的问题：超定系统和伪逆。

如果我们实际用了三个以上的光源，我们希望能够去掉一些噪声和（或）测量误差的影响。假定我们有k个光源，那么公式（9-66）就可以重写成$\boldsymbol{I}_{k\times1}=\boldsymbol{N}_{k\times3}\boldsymbol{n}_{3\times1}$，其中下标用来突出矩阵的维度，为了说明问题清楚$r_0$被去掉了。

现在，我们不能简单地乘以$\boldsymbol{N}$的逆了，因为$\boldsymbol{N}$不是方阵了。就像我们以前多次做过的那样，我们改为构造一个最小化问题：我们要找到一个表面法线向量$\boldsymbol{n}$使得测量值$\boldsymbol{I}$和$\boldsymbol{N}$与$\boldsymbol{n}$之间的点积之差的平方和最小。当然，如果公式（9-63）处处都严格满足，我们就不需要做最小化了。反过来，如果公式（9-63）真的处处满足，也没必要进行三次以上的测量了。我们认为测量是不完美的，因此多测几次是有好处的。定义目标函数E引入寻找最优解的目标:

$$E=\sum_{i=1}^{k}(I_i-\boldsymbol{N}_i^{\mathrm{T}}\boldsymbol{n})^2=(\boldsymbol{I}-\boldsymbol{N}\boldsymbol{n})^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{N}\boldsymbol{n})\tag{9-69}$$

246

展开乘积得到:

$$E=\boldsymbol{I}^{\mathrm{T}}\boldsymbol{I}-2\boldsymbol{n}^{\mathrm{T}}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{I}+\boldsymbol{n}^{\mathrm{T}}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{n}\tag{9-70}$$

我们希望找到能够最小化平方误差E的表面法向$\boldsymbol{n}$，因此我们求E对$\boldsymbol{n}$的偏微分:

$$\nabla_n E=-2\boldsymbol{N}^{\mathrm{T}}\boldsymbol{I}+2\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{n}\tag{9-71}$$

令梯度为零，我们得到:

$$\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{n}=\boldsymbol{N}^{\mathrm{T}}\boldsymbol{I}\tag{9-72}$$

即

$$\boldsymbol{n}=(\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N})^{-1}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{I}\tag{9-73}$$

（在你还没意识到的时候，伪逆就出现了。）这可是绕了一大圈。让我们看看有没有简单

一些的方法:

回到公式（9-66），为说明方便r_0又被去掉了，两边同时乘以N^T:

$$N^T I = N^T N n \tag{9-74}$$

两边同时乘以$(N^T N)^{-1}$，我们得到了与公式（9-73）相同的结果。

那么我们为什么那么麻烦——从公式（9-69）到公式（9-72）的工作似乎是个浪费，其实并非如此。我们如此做的目的是为了证明，乘以伪逆$(N^T N)^{-1} N^T$可以得到超定线性系统的最小均方误差估计。这是一个重要的结果：不仅在光度测量立体视觉中重要，在其他许多应用中也如此。

在两个光源条件下用影调推测三维形状

现在我们将举例说明在只有两个光源的条件下如何用影调推测三维形状。我们在这里不深入讨论物理细节，但是扫描电子显微镜提供了一个很好的应用实例。在这种显微镜中会生成两幅图像——一幅来自二次电子发射(SE)，另一幅来自反向散射电子(BSE)。反射率函数很容易度量。

我们并不去构造显微镜几何结构的精确模型，只是在显微镜前面放一个球并成像来测量反射率函数R_{SE}和R_{BSE}。假定表面反射率已知（这是一个重要的假设），在图像SE和图像BSE中，测得的亮度都是p和q的函数。例如，在图9-20中，特定的亮度值可以用点的轨迹来表示。尽管在每幅图像中，都有无穷多种可能的p、q值，但只可能有两个(p, q)对能够同时满足两幅图像中测量的亮度值。因此我们通过定义一个目标函数并使目标函数最小化来求解表面$z(x, y)$。为完成这个任务，我们定义一个函数$\rho((p_1, q_1), (p_2, q_2))$来表示两个表面法线方向之间的差异（它可以很简单，例如可以用两个向量夹角的余弦$\rho((p_1, q_1), (p_2, q_2)) = p_1 p_2 + q_1 q_2$）。然后，
247 我们考察第i个(p, q)对所定义的法线方向与垂直于表面$z(x, y)$的向量之间的差异，并把这个差异写成:

$$d_i(x,y) = \rho\left(\left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right), (p_i, q_i)\right)$$

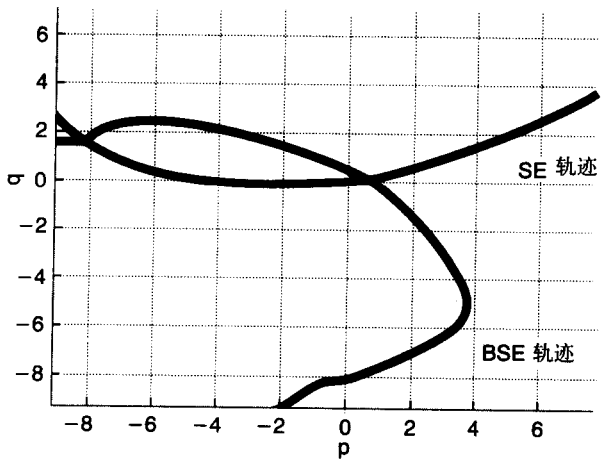


图9-20 在SE图像和BSE图像中的一个亮度测量值对由p，q的轨迹决定（作者感谢B. Karacali提供本图）

最后，假定图9-20中的两条曲线相交于m个点（我们把m作为x，y的函数是为了提醒读者，所有这些都是对同一个x，y点）。我们定义目标函数如下:

$$E=\sum_{x,y}\left(\sum_{i}^{n(x,y)}(d_i(x,y))^{-1}\right)^{-1}+\lambda R \tag{9-75}$$

允许R作为一个正则项，例如分段线性。

当表面不只发生反射时，上面给出的简单模型会发生变化。例如，在红外波长范围内（或许还有一些其他的波段），从表面测量到的能量是反射能量和放射能量（也称为"黑体辐射"）的混合，其中放射能量是由于物体表面的温度产生的[9.48]。

广义柱体（GC）的概念在9.12节中已经讨论过了。如果把广义柱体和从影调推测三维形状的方法结合起来，可以改进广义柱体模型的功能[9.23]。

9A.2.3　结构光

在第4.2.2节已经介绍了结构光的基本概念。关键是通过控制光照消去立体视觉中的一个或多个未知量。让我们看一个具体的例子来了解这种方法是如何工作的。

我们要解决的问题是机器人视觉中的一个应用：机器人要拾起一个闪光的金属涡轮叶片，并把它放到机器上等待进一步处理。为定位叶片，一束激光通过圆柱形透镜投射到物体上产生一条狭长的水平光带。图像生成装置的几何构造如图9-21所示[9.57]。

248

如果没有叶片，激光产生的光带通过载体的反射将在图像中形成一条水平线。叶片的存在导致光带的垂直位移，垂直位移的线数与角度差成正比，由此可以求出角度Φ。已知两个角度和照相机与投影仪之间的距离h，计算距离z就很容易了：

$$z=\frac{h\tan\theta}{\tan\theta+\tan\varphi} \tag{9-76}$$

尽管这个关系相对简单，但事实证明建立一个z与行位移之间关系的查找表则更加简便而准确。

在实际应用中遇到的一个问题是由涡轮叶片的镜面反射引起的：光斑的亮度比图像中其他像素的亮度高几个数量级。这个问题的解决方法是让激光束通过一个偏振光过滤器。通过在透镜前放置一个这样的过滤器，镜面反射产生的光斑的亮度被削弱了很多。

在这个利用结构光的例子中，一次只投射了一条光带，因此不会出现混淆不同投影仪产生的光带的问题。但是，在更一般的情况下，可能会使用多个投影仪，此时，需要用一些方法保证这些光带不会被混淆[9.7, 9.50]。

9A.2.4　从纹理推测形状

纹理，或者更恰当的说纹理在空间的变化，可以用来刻画三维形状，如图9-22所示。在[9.82]中可以找到很漂亮的显示如何从纹理推测形状的图片集。当然，为了从纹理推测形状，必须先抽取纹理基元，如我们在4A.2.2节中讨论的那样。

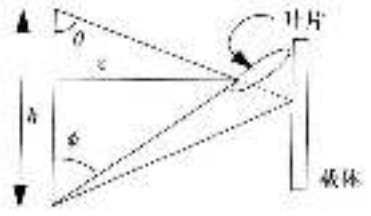


图9-21　叶片的存在导致水平光带在图像中垂直方向上产生位移

图9-22　纹理的变化可以提供关于形状的信息（引自[9.13]）© 2005 Artists Rights Society (ARS), New York / ADAGP, Paris

从纹理推测形状的想法在恢复三维运动的工作中也能够找到应用[9.80]，还可以参考[9.77]。

9A.2.5 从聚焦推测形状

很显然我们可以从聚焦推测深度。但是，想把算法扩展到能从聚焦得到鲁棒的形状是很困难的。主要的问题是如何精确地判定每个像素是否在焦点上[9.52, 9.79]。

249

9A.3 运动分析与跟踪

运动分析可以看作是两个不同的问题。第一种情况是摄像机运动而物体静止。此时提取摄像机的运动是个挑战。第二种情况中，摄像机是静止的，物体在运动。最后，还有二者的混合——摄像机和物体都在运动。运动分析与立体视觉有很多共同的问题。例如，在立体视觉中，对应是个主要问题，而在运动分析中，对应关系建立在随时间变化的场景之间，而不是建立在因摄像机位置不同的场景之间。

运动分析的方法之一被称为“光流”。考察同一物体的两幅图像，假定第二帧中的图像与第一帧相同，只不过发生了位移:

$$f_1(x+\delta)=f_2(x) \tag{9-77}$$

把左式展开成泰勒序列:

$$f_1(x)+\delta f_1'(x)+\cdots=f_2(x) \tag{9-78}$$

如果我们把这个序列截断只保留两项并求解δ，可得

$$\delta=\frac{f_2(x)-f_1(x)}{f_1'(x)} \tag{9-79}$$

这里有一个严重的问题。如果梯度为零怎么办？梯度为零意味着在图像的该点处没有任何信息。想像你通过望远镜观察一辆正在经过的牵引式半拖车 (a semi-tractor-trailer truck)，望远镜的视野仅容你看到卡车上一块很小的面积，例如几平方英寸。当挡板经过时，你得到信息，知道发生了运动。然而，当拖车的顶端经过时，在一段相当长的时间内你看不到任何变化。

在计算和应用光流时遇到的另一个麻烦是在二维空间中，公式9-79变成了微分方程（参见[9.10]）。为了解决这些问题，研究光流的人们尝试了各种方法试图把局部度量得到的信息组合起来推测全局知识，例如利用聚类识别共同运动的点集[9.41]。正如第4章讨论过的那样，你可以匹配点或边界。例如，Quan[9.61]匹配二次曲线；Taylor和Kriegman[9.85]匹配线段，Zhang[9.99]也如此；Smith和Nandha Kumar[4.35]匹配纹理。[4.35, 9.3]中讨论了运动分割。

在二维空间中，公式（9-77）中的δ变成了一个向量。光流算法生成了一个视差场 (disparity field)，这是一个向量场，它把每个像素和一个向量联系起来[9.20]。[10.19, 17.13, 17.14]中探讨了光流法的有效实现问题，[17.60]中还包括了变形体。

移动的摄像机的图像平面上某一点处的光流为:

250

$$u_i=\frac{1}{z_i}A_iT+B_i\Omega \tag{9-80}$$

其中矩阵

$$\boldsymbol{A}_i=\begin{bmatrix}-f & 0 & x_i\\ 0 & -f & y_i\end{bmatrix} \quad 和 \quad \boldsymbol{B}_i=\begin{bmatrix}x_iy_i & -(1+x_i^2) & y_i\\ (1+y_i^2) & -x_iy_i & -x_i\end{bmatrix} \tag{9-81}$$

完全由摄像机决定。$\boldsymbol{T}$是摄像机的平移速度，Ω是摄像机的旋转速度。z_i 是在摄像机坐标（x_i, y_i）处图像点的深度。有几位研究者致力于解决确定$\boldsymbol{T}$和Ω的问题[9.36, 9.37]。Earnshaw和Blostein[9.15]比较了这些方法并介绍了一种新的变种。

运动可以从拖影（smear）中估计出来。Chen等[9.9]观察到："心理生理学研究表明人类的视觉系统会整合120毫秒内的视网膜图像。由于这样的整合，运动拖影在所难免。据报告当人类视觉系统看到一幅由于运动产生模糊的图像时，在短时间内（至多20毫秒）感知到的拖影会随着观察时间的累积而增加。但在更长的时间内，观察到的图像反而变得清晰了。我们推测人类视觉系统在对图像进行一种去模糊或锐化的操作。"这一观察催生了"从拖影恢复运动"算法[9.9]。

最近，在运动分析的文献中可以观察到人们关注的焦点逐渐从分析图像或摄像机的运动转移到了标记正在发生的动作，即关注点从"怎样运动"转移到了"发生了什么"。

从运动推测形状

在从运动推测形状的研究领域中通常假定对应的问题已经解决了：即某些点已经被识别出来了，并在每一帧中都找到了对应点。在这里，我们将描述由Kanade与其合作者提出的从运动推测形状的方法[9.60, 9.86]。这里的描述建立在一个不真实的假设基础上：物体正交（垂直）投影到图像平面上。我们引入这个假设是因为在这种情况下容易理解。不过，如果你要实现这一算法，请参考[9.60]看对透视投影算法需做哪些改动。还可以参考Soatto和Perona[9.72, 9.73]全面了解近期的研究。

在下面的推导中，物体是静止的，只有摄像机在运动。空间坐标为$\boldsymbol{s}_p$的点p被投影到第f帧图像的坐标（u_{fp}, v_{fp}）处。摄像机在移动，在每一帧，摄像机位于$\boldsymbol{t}_f$处，它的朝向由三个向量（$\boldsymbol{i}_f$, $\boldsymbol{j}_f$, $\boldsymbol{k}_f$）描述。如果我们已知$\boldsymbol{s}_p$，就可以得到p的摄像机坐标，公式如下：

$$u_{fp} = \boldsymbol{i}_f^{\mathrm{T}}(\boldsymbol{s}_p - \boldsymbol{t}_f) \quad v_{fp} = \boldsymbol{j}_f^{\mathrm{T}}(\boldsymbol{s}_p - \boldsymbol{t}_f) \tag{9-82}$$

定义

$$x_f = -\boldsymbol{t}_f^{\mathrm{T}}\boldsymbol{i}_f \qquad y_f = -\boldsymbol{t}_f^{\mathrm{T}}\boldsymbol{j}_f \tag{9-83}$$

我们可以把公式（9-82）重新写成

$$u_{fp} = \boldsymbol{i}_f^{\mathrm{T}}\boldsymbol{s}_p + x_f \quad v_{fp} = \boldsymbol{j}_f^{\mathrm{T}}\boldsymbol{s}_p + y_f \tag{9-84}$$

251

现在我们有一个逆问题：已知图像坐标，我们必须从中求出摄像机的空间位置和物体点的空间坐标。

把图像点的所有观察值组成一个矩阵：

$$\boldsymbol{W} = \begin{bmatrix} u_{11} & \cdots & u_{1P} \\ \cdots & \cdots & \cdots \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1P} \\ \cdots & \cdots & \cdots \\ v_{F1} & \cdots & v_{FP} \end{bmatrix} \tag{9-85}$$

假定有F帧和P个点。我们发现矩阵的每一列是同一个点在不同帧图像中的坐标的列表，而每一行是所有点在同一帧中的坐标的列表。然后，我们定义一个矩阵$\boldsymbol{M}$，它是一个$2F \times 3$的矩阵，它的行是$\boldsymbol{i}_f$ 向量和$\boldsymbol{j}_f$ 向量，再定义一个S矩阵，它是一个$3 \times P$维的"形状矩阵"，它的列是$\boldsymbol{s}_p$向量。最后，定义一个由x_f 和y_f 组成的$2F$维的移动向量$\boldsymbol{T}$。有了这些定义，我们可以把所

有点的公式（9-84）重写成

$$W = MS + T\mathbf{1}_P \tag{9-86}$$

其中$\mathbf{1}_P$是一个长度为P的全1向量。（注意T与全1的行向量的乘积是一个外积，构造出一个$2F \times P$的矩阵。）

如果我们把物体的重心移到原点（为什么不呢？原点的位置可以任意选择），得到

$$C = \frac{1}{P}\sum_{p=1}^{P} s_p = 0 \tag{9-87}$$

这样设置的原点位置可以使我们立即解出T，观察到由于S的行和为零，W的任意一行的和就只是PT，而T的任意一行就可以由W的相应行除以P得到。现在，从W中减去T得到一个新的矩阵$\hat{W}$满足：

$$\hat{W} = MS \tag{9-88}$$

用奇异值分解，我们可以找到一种最合适的$\hat{W}$的分解，记作$\hat{W} = \hat{M}\hat{S}$。不幸的是，这不一定是我们想要的M和S，因为我们在$\hat{M}$和$\hat{S}$之间插入任意矩阵的乘积AA^{-1}都不会改变二者的乘积。因此我们要搜索一个矩阵A使得

$$M = \hat{M}A \quad S = A^{-1}\hat{S} \tag{9-89}$$

为了找到A，我们利用M的行向量是摄像机的方向向量因此它们彼此正交的条件。有了这些额外的约束，A可以确定了，这样我们就知道所有P个点的三维空间坐标和每一帧时摄像机的角度了。

252

9A.4 术语

幅照度	irradiance
光流	optic flow
透视	perspective
光度测量立体视觉	photometric stereo
伪逆	pseudo-inverse
反射率	reflectivity
从影调推测三维形状	shape from shading
结构光	structured illumination

作业9.1

对9.3节中描述的每个特征，判定该特征对以下变换是否具有不变性：1）在视平面内的旋转；2）在视平面内的平移；3）平面外的旋转（如果物体位于一个平面上则是仿射变换）；4）缩放。

作业9.2

设两点之间的欧氏距离由$d(P_1, P_2)$表示（在下一题中你也许会用到它）。设计一个单调的度量$R(P_1, P_2)$把所有的距离映像到0和1之间。即，如果$d(P_1, P_2) = \infty$，则$R(P_1, P_2) = 1$；如果$d(P_1, P_2) = 0$，则$R(P_1, P_2) = 0$。

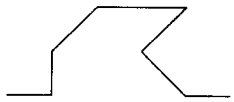
对你设计的度量，说明你将如何证明它是一个正规的度量。只要设定问题即可，如果你真的证明出来了可以得到加分。

作业9.3

一个区域边界上的五个点如下：(1, 1)，(2, 1)，(2, 2)，(2, 4)，(3, 2)。用特征向量方法将这个点集拟合成一条直线，从而找到这个区域的主轴。找到主轴后，估计这个区域的纵横比。

作业9.4

写出下图的链码。



253

作业9.5

讨论下面的假定：设P_1和P_2是区域的两个极点并决定了区域的直径，那么P_1和P_2一定位于区域的边界上。

作业9.6

从公式（9-19）推出公式（9-20）。

作业9.7

对称性的亮度轴（intensity axis of symmetry）与中轴之间有何区别？

作业9.8

在表9-1中，证明不变矩φ_1对缩放变换不变。

作业9.9

你的老师将指定一幅只含有一个区域的图像，该区域内所有像素的亮度均为单位值，背景亮度均为零。

1）计算前景区域的七个矩。

2）将前景区域绕其重心分别旋转十度、二十度和四十度，计算结果图像的不变矩。你得到了什么结论？

作业9.10

证明公式（9-35）对下面的变换具有不变性：1）平移；2）旋转；3）放缩。

作业9.11

图9-12的标题正确吗？

作业9.12

254 两个轮廓A和B已经被测量出来并进行了编码。然后，傅里叶描述子也计算出来了。描述子如表9-3所示。

表9-3 傅里叶描述子的复数值

物体A	物体B
5.00+*i*0.00	5.83+*i*1.80
4.2+*i*1.87	3.69+*i*2.57
3.86+*i*1.00	3.48+*i*2.00
2.95+*i*2.05	2.30+*i*2.77
3.19+*i*1.47	2.70+*i*2.24

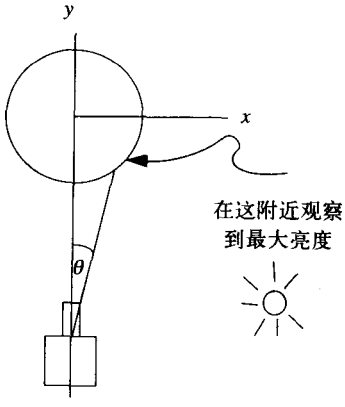
这两个物体有可能代表彼此的相似变换。（相似变换等价于刚体运动，只有平移和旋转。）它们可能是彼此的仿射变换吗？（仿射变换是一种线性变换，不仅包括刚体运动，还可能包括坐标轴的比例变换。如果两个坐标轴（在二维空间中）的比例变换相同，就是缩放。如果二者的比例变换不同，就是剪切。）

如果你判定这两套描述子表示的是相同的形状，可能经过了变换，那么描述并证明从A到B经过了什么类型的变换。如果你认为它们表示的不是相同的形状，说明原因。

作业9.13

一个单位半径、高度为10的圆柱体垂直放置在原点处，已知圆柱体的表面为朗伯反射体。即，反射光的亮度与观察角度无关，只与入射角有关，并遵循如下的关系$f = aI\cos\theta_i$，其中a是表面反射率，I是光源的强度。这个圆柱体的表面反射率是个常数。

摄像机位于$x = 0$，$y = -2$，$z = 2$的位置上，摄像机的光轴指向原点。



光源距离原点4个单位，并已知为朝各方向均匀发光的点光源。在圆柱体的图像中最亮的
255 点位于29度角处。光源在哪里？为方便起见坐标系统的简图显示在插图中。（注意简图不是按

比例的，甚至不保证是正确的。）

参考文献

[9.1] K. Arbter, W. Snyder, H. Burkhardt, and G. Hirzinger, "Application of Affine-invariant Fourier Descriptors to Recognition of 3-D Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(7), pp. 640–647, 1990.

[9.2] D. Ballard and C. Brown, *Computer Vision*, Englewood Cliffs, NJ, Prentice-Hall, 1982.

[9.3] M. Bichsel, "Segmenting Simply Connected Moving Objects in a Static Scene," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), pp. 1138–1142, 1994.

[9.4] T. Binford, "Visual Perception by Computer," *IEEE Conference on Systems and Control*, Miami, December, 1971.

[9.5] A. Bobick and J. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3), pp. 257–267, 2001.

[9.6] A. Califano and R. Mohan, "Multidimensional Indexing for Recognizing Visual Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), pp. 373–392, 1994.

[9.7] D. Caspi, N. Kiryati, and J. Shamir, "Range Imaging with Adaptive Color Structured Light," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(5), pp. 470–480, 1998.

[9.8] C. Chen, T. Huang, and M. Arrott, "Modeling, Analysis, and Visualization of Left Ventricle Shape and Motion by Hierarchical Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), pp. 342–356, 1994.

[9.9] W. Chen, N. Nandhakumar, and W. Martin, "Image Motion Estimation from Motion Smear – A New Computational Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), pp. 412–425, 1996.

[9.10] A. Chhabra and T. Grogan, "On Poisson Solvers and Semi-direct Methods for Computing Area Based Optic Flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), pp. 1133–1138, 1994.

[9.11] K. Cho and S. Dunn, "Learning Shape Classes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), pp. 882–888, 1994.

[9.12] J. Chuang, C. Tsai, and M. Ko, "Skeletonization of Three-dimensional Object using Generalized Potential Field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11), pp. 1241–1251, 2000.

[9.13] M. Clerc, "Texture Gradient," http://www.masterworksfineart.com/inventory/vas_originals.htm.

[9.14] M. Dhome, M. Richetin, J. Lapreste, and G. Rives, "Determination of the Attitude of 3-D Objects from a Single Perspective View," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(12), pp. 1265–1278, 1989.

[9.15] A. Earnshaw and S. Blostein, "The Performance of Camera Translation Direction Estimators from Optical Flow: Analysis, Comparison, and Theoretical Limits," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(9), pp. 927–932, 1996.

[9.16] A. Ferreira and S. Ubeda, "Computing the Medial Axis Transform in Parallel with

Eight Scan Operations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(3), pp. 277–282, 1999.

[9.17] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography," *Communications of the Association for Computing Machinery*, **24**(6), 1981.

[9.18] P. Flynn, "3-D Object Recognition with Symmetric Models: Symmetry Extraction and Encoding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), pp. 814–818, 1994.

[9.19] D. Geiger, A. Gupta, L. Costa, and J. Vlontzos, "Dynamic Programming for Detecting, Tracking, and Matching Deformable Contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(3), pp. 294–302, 1995.

[9.20] S. Ghosal and P. Vanek, "A Fast Scalable Algorithm for Discontinuous Optical Flow Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(2), pp. 181–194, 1996.

[9.21] R. Gonzalez and P. Wintz, *Digital Image Processing*, Reading, MA, Addison-Wesley, 1977.

[9.22] A. Gross, "Toward Object-based Heuristics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), pp. 794–802, 1994.

[9.23] A. Gross and T. Boult, "Recovery of SHGCs from a Single Intensity View," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(2), pp. 161–180, 1996.

[9.24] M. Gruber and K. Hsu, "Moment-based Image Normalization with High Noise-Tolerance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(2), pp. 136–139, 1997.

[9.25] S. Gunn and M. Nixon, "A Robust Snake Implementation: A Dual Active Contour," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), pp. 63–68, 1997.

[9.26] R. Hartley, "Projective Reconstruction and Invariants from Multiple Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), pp. 1036–1041, 1994.

[9.27] P. Havaldar and G. Medioni, "Full Volumetric Descriptions from Three Intensity Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(5), pp. 540–545, 1998.

[9.28] G. Healey and R. Kondepudy, "Radiometric CCD Camera Calibration and Noise Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), pp. 267–276, 1994.

[9.29] D. Heeger and A. Jepson, "Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation," *International Journal of Computer Vision*, **7**(2), pp. 95–117, 1992.

[9.30] D. Helman and J. JáJá, "Efficient Image Processing Algorithms on the Scan Line Array Processor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), pp. 47–56, 1995.

[9.31] A. Hoover, D. Goldgof, and K. Bowyer, "Extracting a Valid Boundary Representation from a Segmented Range Image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(9), pp. 920–924, 1995.

[9.32] B.K.P. Horn, *Robot Vision*, Cambridge, MA, MIT Press, 1986.

[9.33] M. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on*

Information Theory, **8**, pp. 179–187, 1962.

[9.34] X. Hu and N. Ahuja, "Matching Point Features with Ordered Geometric, Rigidity, and Disparity Constraints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), pp. 1041–1049, 1994.

[9.35] Y. Iwahori, R. Woodham, and A. Bagheri, "Principal Components Analysis and Neural Network Implementation of Photometric Stereo," *Proceedings IEEE Conference on Physics-Based Modeling in Computer Vision*, June, 1995, pp. 117–125, 1995.

[9.36] A. Jepson and D. Heeger, "Linear Subspace Methods for Recovering Translational Direction," In *Spatial Vision in Humans and Robots*, ed. L. Harris and M. Jenkin, Cambridge, Cambridge University Press, 1993.

[9.37] K. Kanatani, "Unbiased Estimation and Statistical Analysis of 3-D Rigid Motion from Two Views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(1), pp. 37–50, 1993.

[9.38] K. Kanatani, "Comments on 'Symmetry as a Continuous Feature'," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), pp. 246–247, 1997.

[9.39] H. Kauppinen T. Seppänen, and M. Pietikäinen, "An Experimental Comparison of Autoregressive and Fourier-based Descriptors in 2D Shape Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), pp. 201–207, February, 1995.

[9.40] R. Kimmel, A. Amir, and A. Bruckstein, "Finding Shortest Paths on Surfaces Using Level Sets Propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(6), pp. 635–640, 1995.

[9.41] D. Kottke and Y. Sun, "Motion Estimation via Cluster Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), pp. 1128–1132, 1994.

[9.42] A. Laurentini, "The Visual Hull Concept for Silhouette-based Image Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(2), pp. 150–162, 1994.

[9.43] A. Laurentini, "How Far 3D Shapes can be Understood from 2D Silhouettes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), pp. 188–195, 1995.

[9.44] S. Lavallée and R. Szeliski, "Recovering the Position and Orientation for Free-form Objects from Image Contours Using 3D Distance Maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(4), pp. 378–390, 1995.

[9.45] S. Liao and M. Pawlak, "On Image Analysis by Moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), pp. 254–266, 1996.

[9.46] R. Malladi, J. Sethian, and B. Vemuri, "Shape Modeling with Front Propagation: A Level Set Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), pp. 158–175, 1995.

[9.47] C. Manhaeghe, I. Lemahieu, D. Vogelaers, and F. Colardyn, "Automatic Initial Estimation of the Left Ventricular Myocardial Midwall in Emission Tomograms using Kohonen Maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), pp. 259–266, 1994.

[9.48] J. Michel, N. Nandhakumar, and V. Velten, "Thermophysical Algebraic Invariants from Infrared Imagery for Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), pp. 41–51, 1997.

[9.49] F. Mokhtarian, "Silhouette-based Isolated Object Recognition through Curvature

Scale Space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), pp. 539–544, 1995.

[9.50] R. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissano, "Structured Light using Pseudorandom Codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(3), pp. 322–327, 1998.

[9.51] S. Nayar and R. Bolle, "Reflectance Based Object Recognition," *International Journal of Computer Vision*, **17**(3), pp. 219–240, 1996.

[9.52] S. Nayar and Y. Nakagawa, "Shape from Focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), pp. 824–831, 1994.

[9.53] T. Nguyen and B. Oommen, "Moment-preserving Piecewise Linear Approximations of Signals and Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), pp. 84–91, 1997.

[9.54] L. O'Gorman, "Subpixel Precision of Straight-edged Shapes for Registration and Measurement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(7), 1996.

[9.55] F. O'Sullivan and M. Qian, "A Regularized Contrast Statistic for Object Boundary Estimation – Implementation and Statistical Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), pp. 561–570, 1994.

[9.56] M. Oren and S. Nayar, "A Theory of Specular Surface Geometry," *International Journal of Computer Vision*, **24**(2), pp. 105–124, 1996.

[9.57] N. Page, W. Snyder, and S. Rajala, "Turbine Blade Image Processing System." In *Advanced Software in Robotics*, ed. A. Danthine, Amsterdam, North-Holland, 1984.

[9.58] E. Pauwels, P. Fiddelaers, and L. Van Gool, "Enhancement of Planar Shape Through Optimization of Functionals for Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(12), 1995.

[9.59] S. Pizer, C. Burbeck, J. Coggins, D. Fritsch, and B. Morse, "Object Shape Before Boundary Shape: Scale Space Medial Axis," *Journal of Mathematical Imaging and Vision*, **4**, pp. 303–313, 1994.

[9.60] C. Poelman and T. Kanade, "A Paraperspective Factorization Method for Shape and Motion Recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), pp. 206–218, 1997.

[9.61] L. Quan, "Conic Reconstruction and Correspondence from Two Views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(2), 1996.

[9.62] I. Rothe, H. Süsse, and K. Voss, "The Method of Normalization to Determine Invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[9.63] G. Sandini and V. Tagliasco, "An Anthropomorphic Retina-like Structure for Scene Analysis," *Computer Graphics and Image Processing*, **14**, pp. 365–372, 1980.

[9.64] H. Schultz, "Retrieving Shape Information from Multiple Image of a Specular Surface," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(2), pp. 195–201, 1994.

[9.65] E. Schwartz, "Computational Anatomy and Functional Architecture of Striate Cortex, Spatial Mapping Approach to Perceptual Coding," *Vision Research*, **20**, pp. 645–669, 1980.

[9.66] J. Sethian, "Curvature and Evolution of Fronts," *Communications in Mathematical Physics*, **101**, pp. 487–499, 1985.

[9.67] J. Sethian, "Numerical Algorithms for Propagating Interfaces: Hamilton–Jacobi Equations and Conservation Laws," *Journal of Differential Geometry*, **31**, pp. 131–161, 1990.

[9.68] M. Shamos, "Geometric Complexity," *7th Annual ACM Symposium on Theory of Computing*, May, 1975, Albuquerque, NM, pp. 224–233, 1975.

[9.69] D. Sinclair and A. Blake, "Isoperimetric Normalization of Planar Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), pp. 769–777, 1994.

[9.70] S. Smith and J. Brady, "ASSET-2: Real-time Motion Segmentation and Shape Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[9.71] W. Snyder and I. Tang, "Finding the Extrema of a Region," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2**, pp. 266–269, 1980.

[9.72] S. Soatto and P. Perona, "Reducing 'Structure from Motion': A General Framework for Dynamic Vision.1. Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(9), pp. 933–942, 1998.

[9.73] S. Soatto and P. Perona, "Reducing 'Structure from Motion': A General Framework for Dynamic Vision. 2. Implementation and Experimental Assessment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(9), pp. 943–960, 1998.

[9.74] B. Soroka, "Generalized Cylinders from Parallel Slices," *Proceedings of the Conference on Pattern Recognition and Image Processing*, 1979.

[9.75] B. Soroka and R. Bajcsy, "Generalized Cylinders from Serial Sections," *3rd International Joint Conference on Pattern Recognition*, November, Coronado, CA, 1976.

[9.76] M. Soucy and D. Laurendeau, "A General Surface Approach to the Integration of a Set of Range Views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(4), pp. 344–358, 1995.

[9.77] J. Stone and S. Isard, "Adaptive Scale Filtering: A General Method for Obtaining Shape from Texture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(7), pp. 713–718, 1995.

[9.78] G. Storvik, "A Bayesian Approach to Dynamic Contours through Stochastic Sampling and Simulated Annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), pp. 976–986, 1994.

[9.79] M. Subbarao and T. Choi, "Accurate Recovery of Three-dimensional Shape from Image Focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(3), pp. 266–274, 1995.

[9.80] S. Sull and N. Ahuja, "Integrated 3-D Analysis and Analysis-Guided Synthesis of Flight Image Sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), pp. 357–372, 1994.

[9.81] Y. Sun, I. Liu, and J. Grady, "Reconstruction of 3-D Tree-like Structures from Three Mutually Orthogonal Projections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), pp. 241–248, 1994.

[9.82] B. Super, and A. Bovik, "Shape from Texture Using Local Spectral Moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(4), pp. 333–343, 1995.

[9.83] K. Tarabanis, R. Tsai, and A. Kaul, "Computing Occlusion-free Viewpoints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), pp. 279–292, 1996.

[9.84] G. Taubin and R. Ronfard, "Implicit Simplicial Models for Adaptive Curve Reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), pp. 321–325, 1996.

[9.85] C. Taylor and D. Kriegman, "Structure and Motion from Line Segments in Multiple Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), pp. 1021–1032, 1995.

[9.86] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," *International Journal of Computer Vision*, **9**(2), pp. 137–154, 1992.

[9.87] F. Tong and Z. Li, "Reciprocal-wedge Transform for Space-variant Sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), pp. 500–511, 1995.

[9.88] B. Vemuri and Y. Guo, "Snake Pedals: Compact and Versatile Geometric Models with Physics-based Control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), pp. 445–459, 2000.

[9.89] K. Voss and H. Suesse, "Invariant Fitting of Planar Objects by Primitives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), pp. 80–84, 1997.

[9.90] C. Weiman and G. Chaikin, "Logarithmic Spiral Grids for Image Processing and Display," *Computer Graphics and Image Processing*, **11**, pp. 197–226, 1979.

[9.91] D. Weinshall and C. Tomasi, "Linear and Incremental Acquisition of Invariant Shape Models from Image Sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), pp. 512–517, 1995.

[9.92] M. Werman and D. Weinshall, "Similarity and Affine Invariant Distances Between 2D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), pp. 810–814, 1995.

[9.93] P. Whaite and F. Ferrie, "Autonomous Exploration: Driven by Uncertainty," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), pp. 193–205, 1997.

[9.94] R. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Engineering*, **19**, pp. 139–144, 1980.

[9.95] Y. Wu, S. Iyengar, R. Jain, and S. Bose, "A New Generalized Framework for Finding Object Orientation using Perspective Trihedral Angle Constraint," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), pp. 961–975, 1994.

[9.96] R. Yip, P. Tam, and D. Leung, "Application of Elliptic Fourier Descriptors to Symmetry Detection under Parallel Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), pp. 277–286, 1994.

[9.97] H. Zabrodsky, S. Peleg, and D. Avnir, "Symmetry as a Continuous Feature," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(12), pp. 1154–1166, 1995.

[9.98] M. Zerroug and R. Nevatia, "Three Dimensional Descriptions based on the Analysis of the Invariant and Quasi-invariant Properties of Some Curved-axis Generalized Cylinders," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), pp. 237–253, 1996.

[9.99] Z. Zhang, "Estimating Motion and Structure from Correspondences of Line Segments between Two Perspective Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(12), pp. 1129–1139, 1995.

[9.100] R. Zhang, P. Tsai, J. Cryer, and M. Shah, "Shape-from-shading: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(8), pp. 690–706, 1999.

[9.101] J. Zheng, "Acquiring 3-D Models from Sequences of Contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(2), pp. 163–178, 1994.

[9.102] Y. Zhong, A. Jain, and M. Dubuisson-Jolly, "Object Tracking using Deformable Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), pp. 544–549, 2000.

[9.103] D. Zwillinger, CRC Standard Mathematical Tables and Formulae, 31st edn., CRC Press, 2003.

256 ~ 262

第10章 一致性标号

On axis, as the planets run,
Yet make at once their circle round the sun;
So two consistent motions act the soul;
And one regards itself and one the whole.

Alexander Pope

在计算机视觉中一个最具有挑战性的问题是“局部/全局性推理问题”。计算机视觉中遇到的问题就像“盲人与大象”这个著名的寓言所描述的那样，要从一组局部度量来推断景物的全局性性质。换句话说，在第8章与第9章分别讨论了分割与形状分析的问题之后，下一个问题是，如何用从分割与形状分析方法得到的物体局部信息，来解释整个景物，这些景物是由多个物体组成的。要解决局部到全局推理问题，需要使用一致性概念。

10.1 一致性

首先要说明一下所使用的符号，一组物体表示为$\{x_1, x_2, \cdots, x_k\}$，物体的标号集是$\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$。为了方便暂且假设景物中每个物体只可能有一个标号，不同标号之间彼此无关，以及每个物体都有一个标号。将赋予标号的物体用有序对（x_i, λ_j）表示，该例子是说第i个物体被赋予第j种标号。

下面用对线条图中的物体标号问题作为一致性标号的一个例子。从机器视觉研究的最初阶段开始人们就对线条图分析问题感兴趣，这有三方面的原因：首先,人们能够通过观察线条图轻易地对其作出解释;其次,心理学实验[10.1，10.6，10.10]已经令人信服地证实，正是那些亮度变化最明显的地方为物体的形状提供了最多的信息，并且将边缘转换成线条，也是相对较容易做到的；第三,线条图极大降低了图像中的数据量（但信息并不因此而明显下降），并且人们认为学会如何处理线条图分析很可能会使分析算法运行较快。线条图分析的基础性工作大多是在20世纪60年代末期到20世纪70年代完成的[10.5，10.8]，但是仍然需要继续发展[10.17]。

在本章讨论中，线条图的每条线可以赋予三种标号：凸的（是指在三维中的一条边是指向观察者的，类似于一张桌子的角）、凹的（三维中的一条边是指向离开观察者的方向，例如一个房间的墙与地板形成的交点）或遮挡边（这条边出现是由于一个表面被另一表面部分遮挡造成的）。例如考察图10-1所示的一个例子。在该图中由凸边形成的线用“+”符号表示，凹边则用“－”号表示，遮挡边则用箭头符号表示。箭头的方向用来表示该边两侧的遮挡关系，它的定义是如果你沿这个箭头所指方向移动，则在你的左边是被遮挡的表面，该图并没有对所有边都加以标号，这是故意这样做的，目的在于表示,在给定某种物体、其线条图以及对这些线的三种定义的条件下，线条图分析的任务是要让计算机学会像人们很容易做到的那样，对景物中图形的线条自动做出解释。

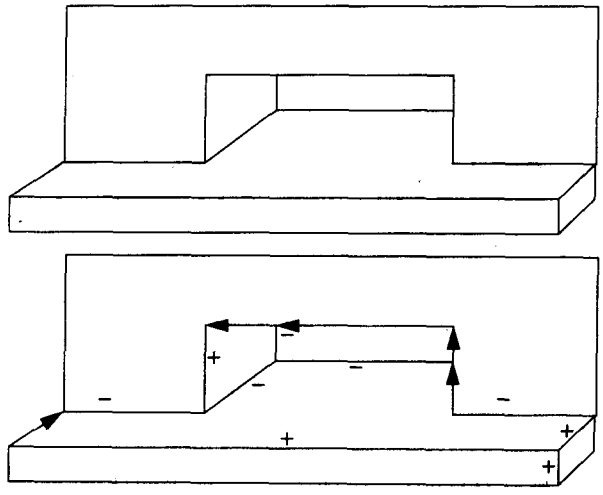


图10-1 用来表示凸、凹与遮挡边的线条图以及对该幅图的标号

在讨论此类问题时，需要说明不幸的歧义性问题，它是因自然语言引入的。标号（Labeling）这个术语可以有两种意义，它可以是指赋予一个物体、一对物体或整个景物的标号本身，但也可以指赋标号这个过程（译者注：最后一句在原文没有，是译者根据文意加上的）。但是利用上下文可以明白该术语在所讨论上下文中的含义。为了实现为景物赋予标号的目的，必须考虑一致性（consistency）这个概念。对景物中两个物体同时赋予标号，要考虑所赋标号之间的兼容性（compatibility），这里用一种函数$r(i, \lambda, j, \lambda')$表示。如果两个标号能并存（相互兼容）则该函数定义为1，如它们不能并存，则函数定义为-1。例如$r(i, +, i, -)=-1$。因为物体i不能同时标为凹的与凸的（参见图10-2）。如果$\sum_i \sum_{i \neq j} r(i, \lambda, j, \lambda')=n(n-1)$，那么称这幅图像的标号是完备的（complete）与一致的（consistent），也就是说，图中物体的所有标号都是可允许存在的，在本章中，会用到兼容函数的若干种不同的定义版本。

图10-2 一条线不可能既能是凸的又是凹的

264

下面进一步深入讨论线条图标号的细节，尽管要标号的是线条图中的线条，实际上还要与顶点一起考虑才有意义。一个顶点是几条线交汇之处。如果每条线可以有4种标号（凹、凸、箭头入与箭头出），则一个由三条线交汇的顶点就会有4^3种标号法。然而并不是所有这些组合在物理上是可能出现的。在图10-3到图10-5中标出“Y”“弯头”与“箭头”顶点在物理上可能成立的所有种类，而使用这些信息可以有各种不同的方法。一种方法是用深度优先搜索，其相应算法为:

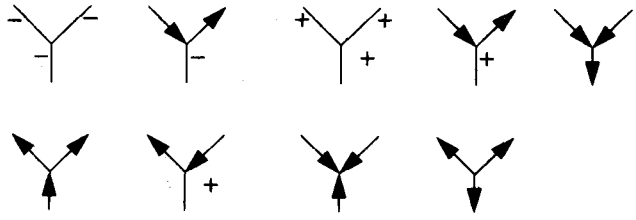


图10-3 在物理上能够成立的所有Y形类接点

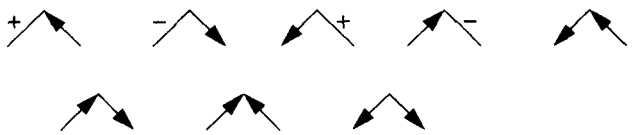


图10-4 在物理上能够成立的所有弯头(ELL)类接点

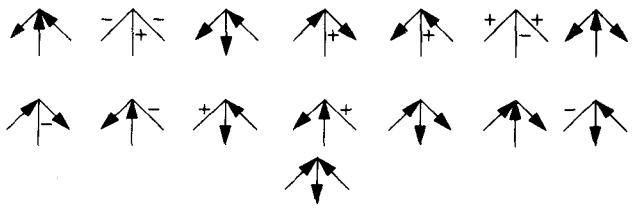


图10-5 在物理上能够成立的所有箭头类接点

1) 选择一个起始顶点（称其为顶点1），然后对到达该顶点的所有线条用物理上可成立的方式标号。

2) 选择顶点1的一个相邻顶点（称其为顶点2）,对到达该顶点的所有线条用物理上可成立的方式标号，并且所标的号要与先前的标号一致，也就是从顶点1到顶点2的线只能有一个标号。

3) 如果不可能有一致的标号，则返回。

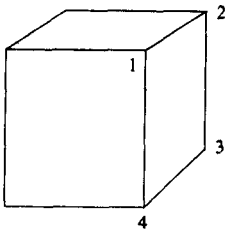


图10-6 用来给出标号一致性的物体

对图10-6中所示的三维物体的标号过程在图10-7中说明，该过程从为顶点1有关的线条选择一种可能的标号开始。给定了该种标号后对顶点2的线可以选择任何一种标号，但条件是在顶点1与2之间的连线的符号必须为“+”号，如图10-7中第2步所示。对图所示顶点2可能的标号种类中，假设选择了左边的那一种，那么对顶点3的选择必须与顶点1与2两者选择都一致。至此为了确定对顶点4的标号，则必须为顶点3假设一个“正确”的标号。假设仍取图中左边那一种。这样一来，为顶点4有关线条选择的标号必须与（假设正确的）顶点3与顶点1的标号相一致。由于到达顶点4的线中已有两条线的标号已经确定，因此对另一条线只有一种能维持一致性的标号选择，即将第3条线确定为指向该顶点的箭头标号。

265

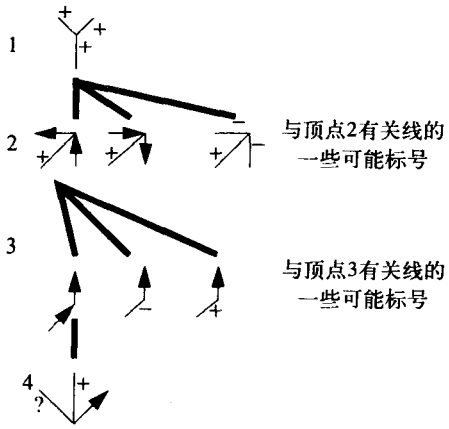


图10-7 图10-6中物体的可能标号

假如到了为顶点4标号并发现没有可能的标号，则说明前面的某种假设不成立。这就需要返回前面对顶点3进行重新标号，如果所有顶点3的标号都不能满足对顶点4进行一致性标号，则需要返回到更前面，对顶点2重新标号。依此类推，直到找到所有顶点的一致性标号，如果没有找到，则该物体没有一致性标号。

由于人们早期开始研究的是线条标号，许多研究人员随后提出了一些增强措施，例如Parodi 和Piccioli[10.13]表明，如果消隐点能够确定，以及有一点的三维坐标已知，则所有标号点的三维坐标都可以找到。

10.2 松弛法标号

在讨论了一致性标号原理之后，可以把这种概念加以推广，使得一个具体的物体可以具有不止一种标号的可能性。为了做到这一点，定义一个变量$p_i(\lambda_j)$，并用它表示物体i具有标号λ_j。为此要讨论某些术语以便继续讨论下去。可能读者会听到把$p_i(\lambda_j)$解释成物体i具有标号λ_j的概率的说法。尽管有时是有这种说法，但是这个术语是不正确的。的确p具有与概率一致的范围$0 \leq p_i(\lambda) \leq 1$，以及它的积分值为1，$\sum_j p_i(\lambda_j)=1$，但这里并没有随机过程，其实$p_i(\lambda_j)$是我们对做出如此标号是否做得对的一种可信度。随着对问题的了解逐渐加深，以及一致性的约束逐渐增强，希望这些p的值能从最初假设的在0到1之间的连续值，逐渐变成不是0就是1。

266

10.2.1 利用一致性来修正标号

线性松弛法

使用一致性的概念的第一种方法，是建立一个线性系统来综合考虑初始概率与一致性。把物体i具有λ标号以及j物体具有λ'标号之间的兼容性定义成$r_L(i, \lambda, j, \lambda')$（其中下标$L$表示该兼容性是在线形松弛法中使用的），并且要求$0 \leq r_L \leq 1$以及

$$\sum_{\lambda} r_L(i,\lambda,j,\lambda') = 1 \quad \text{对所有 } i,j,\lambda' \tag{10-1}$$

线性松弛过程是按下式迭代地修改各种标号赋值权重的过程

$$p_i(\lambda) = \sum_j \sum_{\lambda'} r_L(i,\lambda,j,\lambda') p_j(\lambda') \tag{10-2}$$

很容易看出来式（10-2）是一致性矩阵与加权向量的乘积。参考文献[10.15]中已经证明重复地应用这个式子，则加权向量会收敛到特征值为1对应的特征向量。在某些条件下，一致性矩阵是可以通过学习得到的[10.14]，但不幸的是这并不能给我们以我们所要的信息。那么怎样能做到迭代结果独立于初始条件呢？

非线性松弛法

在非线性松弛法中，仍然要使用一致性原则来修改标号的赋值过程。与线性松弛法相同，同样需要开发一个算法把标号赋值的可信度能根据度量的结果收敛至0或1。首先修改规则要有一个新的定义。在算法的第$k+1$步迭代使用

267

$$p_i^{k+1}(\lambda) = \frac{p_i^k(\lambda)[1+q_i^k(\lambda)]}{\sum_j p_i^k(\lambda_j)[1+q_i^k(\lambda_j)]} \tag{10-3}$$

式中的分母的作用只是使p_i的值保持在0~1的范围内。式中的$q_i(\lambda)$项表示$p_i(\lambda_j)$与所有其

他物体的标号之间兼容性的一种度量。尽管p的值是严格为正的，而q的值却可正可负。如果$q_i(\lambda)$为负值，则表示p_i用λ标号与其他大部分标号不相容。

$$q_i^n(\lambda)=\sum_j C_{ij}\left[\sum_{\lambda'} r_n(i,\lambda,j,\lambda')p_j^n(\lambda')\right] \tag{10-4}$$

式中仍有一个兼容性函数$r(\cdot)$，这在本章前面已见过，但加了一个下标N是表示，在非线性松弛过程中使用。其间的不同在于r的值不仅限于-1与$+1$，而可取其间任何值。如果两个标号赋值之间是完全一致的，它们的兼容性应为1，如果它们的标号赋值是完全不一致的，兼容性就应为-1。如果对某一个物体的标号赋值并不影响其他物体，兼容性则为0。接下来考虑式（10-4）的含义。

从式（10-3）已知，对物体i标号λ的可信度的改变取决于$q_i^n(\lambda)$。而式（10-4）表明它是i物体相应标号，与其他所有当前标号赋值的兼容程度之和。请注意式子中兼容性与其他标号正确性的可信度相乘。这也就是说，如果我们对物体j的标号的可信度很低，那么就不用管它与物体i的标号之间的兼容程度。式中引入C_{ij}是为了方便，它只是对物体i受物体j的影响程度进行加权，而不管它们之间是什么标号。如果我们预先就知道物体i与j彼此间没有影响，那么该系数值就可以为零。C_{ij}是一个选择项，如果将其的作用归并到兼容性函数中去，该系数也就不会出现了。

10.2.2 标号问题示例

在本节结合一些例子讨论兼容函数r的选择方法。

模型匹配问题

假设你已经将距离图像分割成平面面块，现在希望确定一组模型中哪一个与所观察的物体匹配得最好。假设分割过程产生的面块是平面的，由于图像是距离图像，因此可以计算这些面块在三维空间中的朝向。于是要解决的问题变成在模型（或一组模型）中寻找与图像中一组平面块匹配得最好的一组平面表面块。此时图像中的物体是面块，模型中的区域（的标号）是要赋给这些面块的标号。下面是一种定义标号兼容性的方法。设想分别将图像面块A赋予模型区域1，图像块B赋予2时涉及兼容性的一些方面，看看面块A是否与B毗邻，以及区域1与区域2是否毗邻？这里有如图10-8所示的4种可能性。如果图像中两个区域是相邻的，而相应两个区域在模型中也相邻的，那么可以按如下式定义两组标号的兼容性。 268

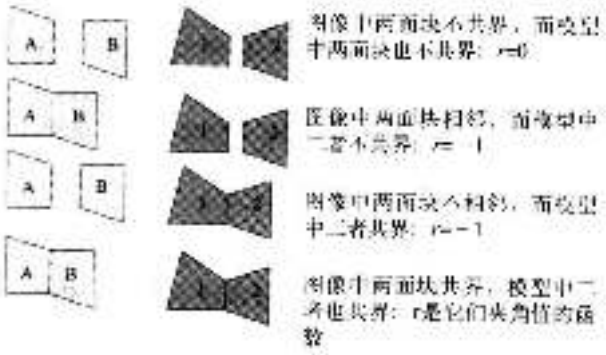


图10-8 对两个面块以及相应两个模型之间进行标号而定义的兼容性有4种可能的情况

$$r_N(A,\lambda_1,B,\lambda_2)=\cos(\theta_{AB}-\theta_{1,2}) \tag{10-5}$$

其中θ_{AB}表示面块A与B之间的夹角（由于这是距离图像，因而其夹角是可以度量的）。

这里要强调的一点是兼容函数的定义完全取决于要解决的问题。而剩下的标号松弛过程就可按式（10-3）与（10-4）定义的方式进行简单的迭代。

另一个例子：跟踪运动目标

假设要跟踪的物体是正在移动的4轮汽车，而摄像机只能记录轮胎的位置（这是一个古怪的摄像机）。因此我们的目标是要确定图像中的哪个轮胎与下一幅图中哪个轮胎对应。图10-9表示了这种情况，其中第n帧图像的轮廓的位置用空心圆表示，而第n+1帧图像中轮廓位置则用实心圆表示。在这个应用中，标号任务可表示成：在n帧图像中的物体（轮子）要用n+1帧图像中的标号（也是轮子）来赋值。那么要用怎样的兼容函数呢？为了找到答案，让我们设想某些不正确的标号赋值。例如：图10-10用箭头表示从物体到标号的关系。你认为这种标号关系是正常的吗？

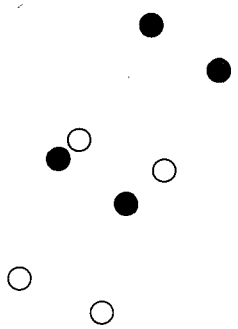


图10-9 第1帧的物体用空心圆表示，而第2帧的物体用实心圆表示，要解决的问题是要找到空心圆与实心圆之间最一致的标号赋值

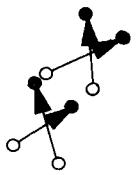


图10-10 一种可能是不正确的标号赋值关系

如果图10-10所示情况是正确的，那么前左轮胎就会转换成前右轮胎，后面轮胎也有类似情况。只有出现汽车翻车，这种解释才合理，而显然我们并不希望这种情况发生。而图10-11所示的解释则显然更加合理，在图中表示标号关系的箭头是接近平行的。这表明可以用箭头之间角度的余弦来定义兼容函数。

269

$$r_N(i,m,j,n)=\cos(\theta(i,m)-\theta(j,p)) \tag{10-6}$$

其中i与j分别是第n帧中的轮子，m与p是第n+1帧中的轮子。而式（10-6）是度量当i与j轮分别对应m与p时的兼容性。尽管这个概念已在本书前几年的非正式版本中已提到，但它是W_u[10.19]于1995年正式发布的。

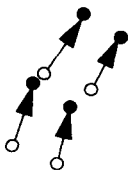


图10-11 一个更加合理的标号赋值关系

10.3 结论

这一章专门讨论一致性问题。我们希望学生们能接受这样一个观点：对来自不同信息源的信息进行融合的最好方法是搜索符合一致性的标号赋值。

求最优解的方法将在下一节中正式使用，下一节建立一个优化问题，然后使用共轭梯度

的数值优化技术来求解。共轭梯度技术不在本章讨论，读者可参阅数学方法的标准教科书。这种技术在许多方面与梯度下降法相似，但是运算快得多。

学者们正在继续从事改进一致性标号概念的工作，其中包括松弛标号[10.4]。松弛法或相似的算法已在许多不同的应用中使用，诸如字符识别[10.12]，边缘提取[10.15]。其基本理论可阅读[10.9，10.16]。

10.4 术语

兼容性	compatibility
下凹的边	concave edge
一致的	consistent
凸起的边	convex edge
标号	labeling
线性松弛法	linear relaxation
局部/全局推理	local/global inference
非线性松弛法	nonlinear relaxation
被遮挡的	occluded
松弛标号	relaxation labeling

270

作业10.1

在本章中，你们已经看到设计兼容性函数的两个例子，下面有机会自行设计兼容性函数。要讨论的问题是：你已用一种边缘检测器在图像中检测出边缘，对图像中的每个像素，进行了梯度计算，得到了它的幅度以及梯度方向。边缘检测中必然有部分被噪声污染，以及模糊化了的，试用松弛标号方法来确定“真实”的边缘像素。

提示：一个“真实”的边缘像素的梯度向量应该与相邻的边缘像素具有相同的方向，采用这种概念设计一个兼容性函数。描写一下如何使用这个概念，可以使用伪码，或文字，或流程图，或三者都用，不用编写实际的软件。

专题 10A 二维线条图的三维解释

从上述讨论中可以看出线条图的解释是一个难题。单个线条图只表示了一个三维物体的某一种视图，因而是带歧义性的。这种歧义性可以通过使用一组存储起来的模型来解决。这种方法需要对图像中可能出现的物体有先验知识，而对没有存储模型的新物体就很难给出比较合理的解释。

Marill[10.11]提出一种解释线条图的另一种方法，它不需要模型，他在对二维线条图作三维解释时只使用启发式规则。使用这种规则可以认为，如果在某种三维解释中，一些线条之间的推算角度明显大于其他线条之间推算的角度时，这种解释的正确性较差。换一种更明确的说法是，在所有与二维线条图具有一致性的三维模型中，优先考虑的解释是角度值的标准差(SDA)为最小的那一个，标准差在式（10-7）中定义，其中θ角在图10-12中说明。

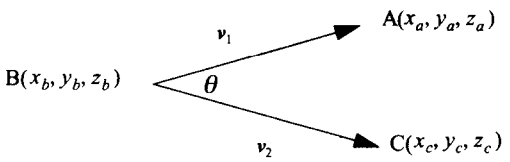


图10-12 线与角度

271

$$SDA = \sqrt{n\sum_{\theta}\theta^2 - \left(\sum_{\theta}\theta\right)^2} \tag{10-7}$$

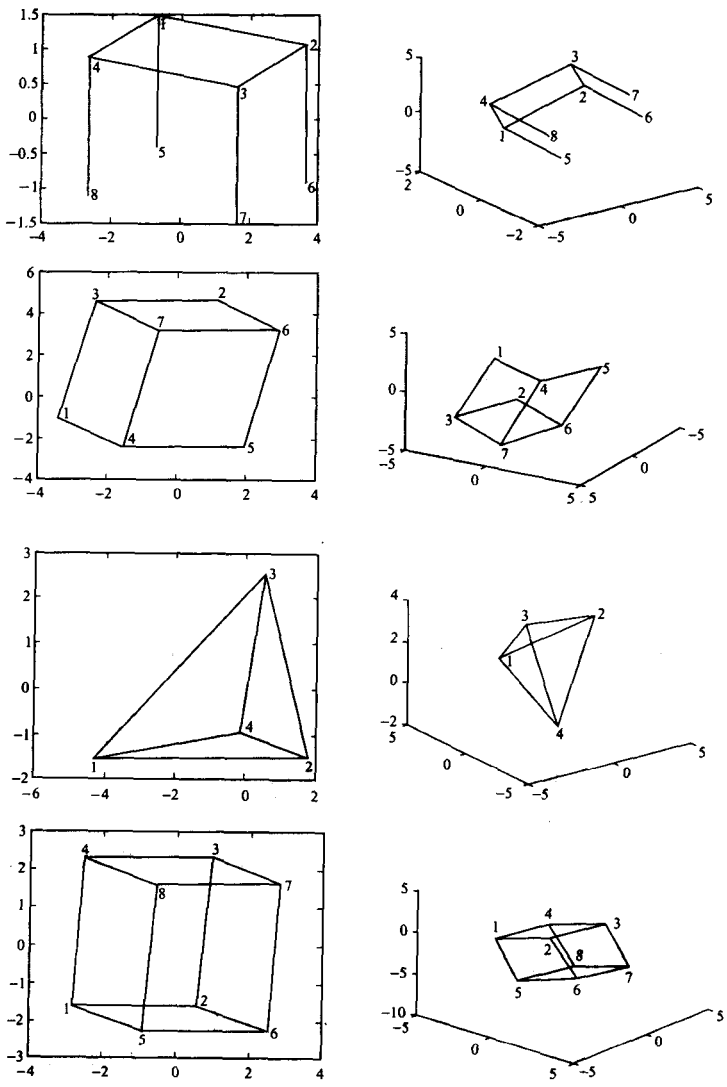


图10-13 左：二维线条图，右：使用仿真法做出的三维解释

这个算法可以对相当多种类不同的线条图做出解释，并且尽管没有明显的模型，但看起来经常与人们做出的解释相一致。

为了使问题简化，将目标函数SDA求平方，并简称为S。我们要为二维图像中的找到第三个坐标值（z_i），使得目标函数S达最小值

272

$$S = n\sum_{\theta}\theta^2 - \left(\sum_{\theta}\theta\right)^2 \tag{10-8}$$

S对z_i的偏导数如式（10-9）所示

$$\frac{\partial S}{\partial z_i} = 2n\sum_{\theta}\left(\theta\frac{\partial\theta}{\partial z_i}\right) - 2\sum_{\theta}\theta\sum_{\theta}\frac{\partial\theta}{\partial z_i} \tag{10-9}$$

其中由两个向量v_1与v_2可以用下式计算

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{v_1 v_2}{\|v_1\|\|v_2\|}\right)\\ &= \cos^{-1}\left(\frac{(x_a-x_b)(x_c-x_b)+(y_a-y_b)(y_c-y_b)+(z_a-z_b)(z_c-z_b)}{\sqrt{(x_a-x_b)^2+(y_a-y_b)^2+(z_a-z_b)^2}\sqrt{(x_c-x_b)^2+(y_c-y_b)^2+(z_c-z_b)^2}}\right)\end{aligned} \tag{10-10}$$

某些二维线条图的三维解释显示在图10-13中，优化问题是用共轭梯度法求解的。

对Marill提出的算法，Wang[10.18]在几个方面做了改进，其中大部分的计算复杂度都不大，包括采用线段幅值的标准差（DSM）作为目标函数[10.3]，以及采用梯度下降法来求解最小化问题[10.2]。

参考文献

[10.1] F. Attneave, "Some Informational Aspects of Visual Perception," *Psychology Review*, **61**(3), 1954.

[10.2] L. Baird and P. S. Wang, "3D Object Perception Using Gradient Descent," *International Journal of Mathematical Imaging and Vision*, **5**, pp. 111–117, 1995.

[10.3] E. W. Brown and P. S. Wang, "Why We See Three-Dimensional Objects: Another Approach," *http://www.ccs.neu.edu/home/feneric/msdsm.html*.

[10.4] W. Christmas, J. Kittler, and M. Petrou, "Structural Matching in Computer Vision using Probabilistic Relaxation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[10.5] M. Clowes, "On Seeing Things," *Artificial Intelligence*, **2**(1), 1971.

[10.6] J. Elder and S. Zucker, "Evidence for Boundary-specific Grouping," *Vision Research*, **38**(1), 1998.

[10.7] D. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, New York, Basic Books, Inc., 1979.

[10.8] D. Huffman, "Impossible Objects as Nonsense Sentences," in *Machine Intelligence*, vol. 6, ed. B. Meltzer and D. Michie, Edinburgh University Press, 1971.

[10.9] R. Hummel and S. Zucker, "On the Foundations of Relaxation Labeling Processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5**(3), 1983.

[10.10] J. Koenderink, "What Does the Occluding Contour Tell Us About Solid Shape?" *Perception*, **13**, pp. 321–330, 1984.

[10.11] T. Marill, "Emulating the Human Interpretation of Line-drawings as Three-dimensional Objects," *International Journal of Computer Vision*, **6**(2), pp. 147–161, 1991. 273

[10.12] J. Ohya, A. Shio, and S. Akamatsu, "Recognizing Characters in Scene Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(2), 1994.

[10.13] P. Parodi and G. Piccioli, "3D Shape Reconstruction by Using Vanishing Points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(2), 1996.

[10.14] M. Pelillo and M. Refice, "Learning Compatibility Coefficients for Relaxation Labeling Processes," *IEEE Transactions on Pattern Analysis and Machine Intel-*

ligence, **16**(9), 1994.

[10.15] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene Labeling Based on the Relaxation Principle," *IEEE Transactions on Systems, Man, and Cybernetics*, **6**, pp. 420–433, 1976.

[10.16] P. Sastry and M. Thathachar, "Analysis of Stochastic Automata Algorithm for Relaxation Labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[10.17] M. Shpitalni and H. Lipson, "Identification of Faces in a 2D Line Drawing Projection of a Wireframe Object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(10), 1996.

[10.18] P. S. Wang, "3D Line Drawing Image Analysis and Recognition," *Progress in Image Analysis and Processing*, **111**, pp. 14–21, 1994.

[10.19] Q. Wu, "A Correlation-relaxation-labeling Framework for Computing Optical Flow – Template Matching from a New Perspective," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(9), 1995.

274

第11章 参数变换

Suppose I was on the other side of the glass, wouldn't the orange still be in my right hand?

Lewis Carroll

本章讨论参数变换，它是对局部到全局推理问题的另一种求解方法。在这种方法中假设从图像中搜索的物体可以用一种数学表达式描述，从而可用一组参数表示。例如一条直线可以表示成以下斜率-截距形式:

$$y=ax+b \tag{11-1}$$

其中a与b是描述该条线的参数。这种方法可描述如下：给定一组点（或其他特征），它们全都服从相同的方程式，要求找到这个方程式的参数。从某种意义上讲，这与一组点拟合一条曲线的问题相同，但随着讨论进行下去，你会发现这种参数变换方法能发现多条曲线，而不需要先验地知道哪个点属于哪条曲线。本章讨论从检测直线这一特定情况开始。

11.1 Hough变换

假设要解决的问题是在如图11-1所示的图像中检测直线。如果图像中只有一条直线，则可用直线拟合来确定曲线（此时为直线——译者注）的参数。但该图像中有两条直线段。如果我们能够先对其进行划分，然后就可对每一段分别进行拟合。这的确可看成一个分割问题，但是这是一个将边界划分成边界段的问题，而不是将图像划分区域的问题。在本章要学习如何实现这个目标的方法。

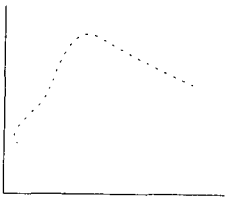


图11-1 一幅从边缘检测器输出的图像，人们立即能察觉到这段边缘曲线大致是由两条线直线段组成的

首先先证明一个带原理性的定理。

定义

在d维空间给定一个点，以及在该空间定义一条曲线的参数表示式，则这个点的参数变换是一条曲线，它把这个点作为一个常数，而参数则是相应的变量。例如，从式（11-1）得到一个参数变换 275

$$b=y-xa \tag{11-2}$$

它本身就是在二维空间〈a, b〉中的一条直线。如果该点为x=3，y=5，则相应的参数变换为$b=5-3a$。

定理

如果有几个点在二维空间中共线，则对应每个点的参数变换如采用$b=y-xa$形式，会相交于〈a, b〉空间的同一点。

证:

设n个点$\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$都满足以下同一个方程式

$$y=a_0x+b_0 \tag{11-3}$$

考虑其中的两个点（x_i, y_i）与（x_j, y_j），它们的参数变换是曲线（在此是直线）

$$\begin{aligned} y_i &= x_i a + b \\ y_j &= x_j a + b \end{aligned} \tag{11-4}$$

将其写成以a与b作为独立变量的形式

$$\begin{aligned} b &= y_i - x_j a \\ b &= y_j - x_j a \end{aligned} \tag{11-5}$$

这两条曲线在a, b的交点如图11-2所示。

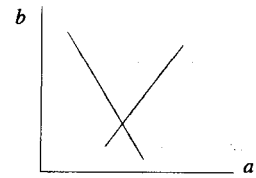


图11-2 参数空间中的两条线的交点

对式（11-5）的两个方程求联立解得到

$$y_j - y_i = (x_j - x_i)a \tag{11-6}$$

因而得到 $a = \dfrac{y_j - y_i}{x_j - x_i}$

将a代入式（11-5）中得到b

$$b = y_i - x_i \frac{y_j - y_i}{x_j - x_i} \tag{11-7}$$

于是得到两条曲线相交处的a与b的值。但是从式（11-3）中可知，所有点的（x, y）值都满足同一条曲线。将其代入式（11-7），就得到

$$b = (a_0 x_i + b_0) - x_i \frac{(a_0 x_j + b_0) - (a_0 x_i + b_0)}{x_j - x_i} \tag{11-8}$$

简化后可得

276

$$b = (a_0 x_i + b_0) - x_i a_0 = b_0 \tag{11-9}$$

同样有

$$a = \frac{y_j - y_i}{x_j - x_i} = \frac{(a_0 x_j + b_0) - (a_0 x_i + b_0)}{x_j - x_i} = a_0 \tag{11-10}$$

于是对用参数a_0与b_0表示的直线上的任两点，它们的参数变换交于$a = a_0$与$b = b_0$。因为任何两点的变换都交在这一点，所有点的变换都交于一个公共点。证毕。

重新回顾一下这里的一个概念：图像中的一个点在参数空间中产生一条曲线（可能是直线）。如果图像中的一些点都在一条直线上，则在参数空间中相应的曲线会交汇在一个公共点上。

懂得这一点了吗？那么进入下一问题的讨论。

11.1.1 垂直线带来的问题

将上述方法用到垂直线会怎么？哇！参数a会变成无穷大。这样可不好，看来需要另一种直线方程式，它如下式表示

$$\rho = x\cos\theta + y\sin\theta \tag{11-11}$$

选择ρ与θ的一组值，并将其固定，那么满足式（11-11）的一组点就是一条直线，这种方程式的几何解释如图11-3所示。

用这种方式表示直线有许多好处，与使用斜率不同，ρ与θ参数值都是有界的。ρ的值必然

不会大于图像的最大对角线长度，而θ则不会超过2π。任何一个角度的直线不会有奇异性。

使用这种参数表达式解决了可能出现无穷大斜率的问题，另一个问题是计算交点的问题。

11.1.2 如何找到交点——累加器数组

采用找到所有曲线的所有交点，然后确定其中哪些是相互接近的方法是不现实的。我们采用累加器数组的概念。为了构造一个累加器数组，我们构造一幅图像，并设其有360列与512行。该幅图像的每个像素值都初始化为零。这幅特殊图像的每一个像素就作为累加器。图11-4表示在一个很小的累加器数组中画出三条直线的情况，它是使用下述算法生成的：

对边缘图像的每个点x_i与y_i：

1) 对θ的所有值计算ρ值。

2) 根据(ρ, θ)的计算值在累加器数组内相应的点(ρ, θ)上的值加1。

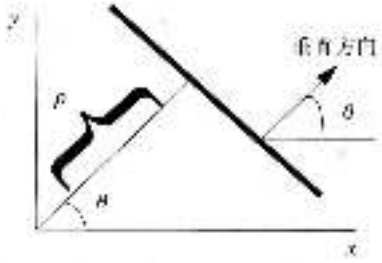


图11-3 一条直线的ρ和θ表达式中，ρ是直线到坐标系原点的垂直距离，而θ是直线梯度方向与x轴之间的夹角

图11-4 三条线用累加器数组标出来，三条线本身也画在其中，每当一条线穿过某累加器，该累加器中的值加1

这个算法使得对应于交点的累加器多重增值。于是累加器数组中的峰值对应于多重交点，从而获得了相应的参数选择。

图11-5显示了一幅有两条直线的图像，以及相应的Hough变换，其中每个像素的亮度是累加器中的值。

图11-5 a) 一幅有两条直线段的图像，其中的两条直线的斜率与截距是明显不同的，但线上的点的位置被很强的噪声所扰动。b) 相应的Hough变换

11.2 减少计算复杂度

参数变换的计算复杂度会相当高，Hough变换也因此被喻为“咖啡变换”，因为你可以命令计算机执行Hough变换，并在它完成之前喝下一整杯的咖啡。例如要计算Hough变换的一幅图像尺寸为512×512，并且希望角度分辨率为1度。因此累加器数组的尺寸为$(512\times2\sqrt{2})\times360$，每次迭代要对每个$\theta$值计算$P$值，于是对图像中的每个边缘点要计算与给360个累加器增值。

使用梯度信息

如果注意到图像中的边缘点通常是某些梯度算子的输出，因此如果知道梯度的幅值与方向，则可以利用这些信息来降低计算的复杂度，这是降低计算复杂度的一种方法。为了说明这一点，请看图11-3，如果我们知道一个点的梯度，我们就知道了边缘的方向，也就知道了θ值。于是只需要计算ρ值一次，而不是360次，并且只需对一个累加器增值，一个360：1的加速，真棒!

当然，这种方法也有实际问题。首先大多数梯度运算检测出的梯度方向不是很准确，因此要增值的累加器单元的位置不十分准确。对此有一种简单的解决办法。做法是不仅仅在一个点（一个单元）增值，而是在一个邻域的单元都增值。例如可以对所计算出的单元增值2，而对该单元的邻域单元增值1，类似于高斯函数。

累加器单元的增值过程中也可以使用梯度的幅值。在前面提到的方法中曾建议对梯度图像先进行阈值化，并在图像的边缘点对应的累加器中增加1。另一种方法是用梯度的幅值进行增值，或用与幅值成正比的值增值。这需要使用浮点表示式的累加器数组，但这并不成为问题。

总之，可以调整在图像空间的计算量（计算梯度的幅值与方向）与参数空间的计算量，以求计算明显加速。在本章后续讨论中还会看到这种方法。其他种启发式措施[11.13]也能产生加速效果。

279

11.3 检测圆

可以将Hough变换推广到包含圆弧线段的检测。在使用维数较低的累加器数据条件下，这种技术比较实用。

11.3.1 用非共线三个像素表示一个圆的定位

设一幅图像的后选边缘像素的亮度值存在数组T中，整数T_{rc}表示在第r行，c列的边缘像素的幅值。给定T中三个非共线的点，本节将讨论如何确定过该三点一个圆的圆心与半径的原理，而在11.3.2节中将其集成为一个参数变换，用来检测圆弧线段。

假设任何两根弦的垂直平分线，（图11-6中的B01与B12），相交于某个有限远点C，很容易证明C就是包含P0、P1与P2三点的圆的圆心。如令L0C是P0到C的线段，L1C是P1到C的线段，以及L2C是P2到C的线段。由于C处在垂直平分线上，P0、P1与P2对C点等距。因此L0C的长度R等于L2C的长度。这三个点对同

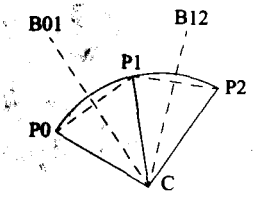


图11-6 包含P0、P1与P2三点的圆的半径

一点C等距离，从而满足以C为圆心，半径为R的圆的定义。

如果B01与B02相互平行，此时它们将不会交在有限远点，则线L01与L12有相同的斜率，P0、P1与P2也因而共线。然而任何三个非共线的点都可以处在同一圆上，如上所述它的圆心可通过这些点的连线的垂直平分线的交点确定，而半径可用圆心到三点中任一点的距离计算。包含P0、P1与P2的圆可用下式定义在XY平面

$$(x-h)^2+(y-k)^2=R^2 \tag{11-12}$$

其中$C=(h,k)$是圆的圆心，半径为R。

那么在上述讨论基础上如何用参数变换技术来找到有关圆的相应参数呢？

11.3.2 用参数变换检测圆

式（11-12）表明了一个圆的方程式，其中x和y是变量，对应圆周上每一点的坐标，而h、k与R则是常数。与上一节求直线时的方法相似，可以把该式写成另一种形式，并交换参数与变量在式中所扮演的角色 280

$$(h-x_i)^2+(k-y_i)^2=R^2 \tag{11-13}$$

在（h, k）空间中，该式描述怎样的几何形状呢？可以猜测它是一个圆。如果R已知，则图像空间中每个点（x_i, y_i）在参数空间中就会产生一条曲线，如果所有这些点在图像空间中属于同一个圆，那么参数空间的这些曲线将会交在何处呢？读者至今应该能回答这个问题。

如果R也未知呢？问题是同样性质的，但是此时不能只改变让h在量值范围内变化来计算k值，而应该让h与k两个量都在其量值范围内变化，并计算R的值，此时的参数空间就变成三维的。允许两个变量变化计算第三个，定义了三维空间的一个表面。这种表面是什么类型的呢（椭圆、双曲函数、锥面、抛物面、平面）？

11.3.3 检测圆中利用梯度信息降低计算量

假设我们知道图像中只有一个圆，或一个圆的一部分，那么怎样才能用最小的计算复杂度检测圆心呢？这里有一个窍门：在每个边缘点计算梯度向量。累加器与该图像是同态的（即图像的每个像素位置对应一个累加器——译者注）。对每个边缘点，沿着它的梯度向量移动，对处在行进中路过的每一累加器位置（也就是对应位置的图像像素处）增值。与前述情况相同，累加器值达到最大的累加器位置，就是该圆的圆心。图11-7显示了这种情况。

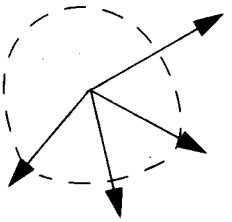


图11-7 各个梯度向量的延线趋向于交于同一点

那么如果圆的半径已知又会怎样呢？如果能从梯度向量得到方向信息，又从已知半径得到距离信息，那么就能知道圆心的位置（至多只有一个处在那一边的歧义性）。说得更清楚一些，如果假设在（x_i, y_i）处的点是半径为R的圆上一点，并且该点梯度的幅度为M，方向为θ，就如图11-8所示，那么圆心的位置应为

$$\begin{aligned} x_0 &= x_i - R\cos\theta \\ y_0 &= y_i - R\sin\theta \end{aligned} \tag{11-14}$$

如前所述，累加器数组增值的方法应该是一个累加器及其邻域，而不仅仅是该累加器。并且累加器所增之值与M成正比是有好处的。

281

11.4 广义Hough变换

迄今为止讨论的情况都假设为，要检测的形状能够用一个解析函数及一组参数来表示。其实这种由数据项进行“投票”确定形状的概念，可以推广到更广义的形态中去。首先不妨假设有一个任意形状的区域，其朝向、形状与放缩比例已知。那么首要的问题是要找到一种表示方法，使得用类似Hough变换的方法检测出该区域。下面是方法之一[11.2]。

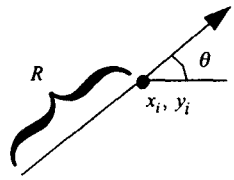


图11-8 假设点（x_i, y_i）在半径为R的圆周上，梯度沿箭头所指方向。如果已知圆比背景要暗，则可沿梯度向量相反方向长度为R处找到圆心。如果对比度与上述情况相反，则要沿梯度方向的正方向找圆心。如果圆与背景的亮暗对比关系未知，则在两个方向上的累加器位置增值

首先，对该形状区域定义一个参考点。选择参考点的方法可以是任意的，但一般用重心比较方便。称该参考点为O。对该形体的任一轮廓边缘点P_i计算其梯度向量，以及该点到O的向量$\vec{OP_i}$。将梯度方向的分布范围量化成几个值，并建立一个有n行的表。对该区域边缘点P_j计算梯度方向，如其值为G_i（i=1, …, n），则在第i行上增加一个新的列，将$\vec{OP_j}$填入。于是在轮廓上尽管可能有多个点具有相同的梯度方向，但它们（由于$\vec{OP_j}$不同）仍然记录在表中同一行中的不同列中。图11-9表示了一个形状，而其中三个轮廓点的有关数据在R表中的记录展示在表11-1中。

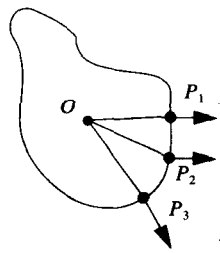


图11-9 图中P_1与P_2轮廓点的梯度有相同的方向，在R表中，它们列在同一行内，P_3的数据显示在R表中的第二行

表11-1 包含图11-9中三个轮廓点数据的R表

梯度方向（度）	轮廓点到参考点的向量		
0	−1, −0.1	−1,0.5	…
300	−0.6, 1.1	空	…

为了使用这样一种形状表示方法实现形状区域的匹配与定位，可以使用以下算法：

1) 生成一个累加器数组，用来存储参考点的后选位置，置该数组初始值为零。

2) 对每个边缘点P_i做：

2.1）计算梯度方向，并在R表中找到相应行。

2.2）对该行的每个项j：

a) 将在R表中所存储的轮廓点位置向量$T[i, j]$与P_i相加，求得形状参考点后选位置：$A=T[i, j]+P_i$。

282

b) 对由A确定的累加器增值。

11.5 结论

在本章中以Hough变换及其推广形式的方法为例，介绍了另一种分析一致性的方法，它们

都用累加器数组。用累加器数组的方法检测一致性很方便，因为体现出一致性的“事件”都会在同一累加器单位中增值，或至少可在相近的累加器单元中增值。通过使用累加器数组累计（对被检测对象）候选假设来检测形状的方法，还可以提高抵抗度量中噪声干扰的能力，因为不一致的解一般不会在全局性（一致的）解的累加器单元中增值，而由噪声产生的假设往往会分散在不同的累加器中。

11.6 术语

累加器数组	accumulator array
广义Hough变换	generalized Hough transform
Hough变换	Hough transform
参数变换	parametric transform

专题11A 参数变换

11A.1 检测抛物线

下面介绍Wechsler与Sklansky[11.12]研究出的，在图像中检测抛物线问题的一种方法。

抛物线是由一些点组成的轨迹，这些点到某一定点（称为焦点）的距离等于它们到某一特定线（称为准线）的距离，如图11-10所示。

$$x^2 = (x-2a)^2 + y^2 \tag{11-15}$$

或

$$y^2 = 4a(x-a) \tag{11-16}$$

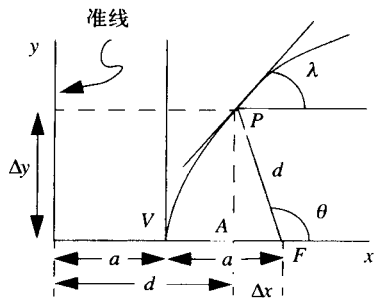


图11-10 一条抛物线

对式（11-16）求对x的导数:

$$\beta = \tan\lambda = \frac{\mathrm{d}y}{\mathrm{d}x} = \sqrt{\frac{a}{x-a}} \tag{11-17}$$

并求出其中的a

$$a = \frac{x\beta^2}{1+\beta^2} \tag{11-18}$$

在图11-10中使用直角三角形PAF来确定

$$\tan\theta = -\tan(\pi-\theta) \tag{11-19}$$

与

$$\tan\theta = \frac{y}{x-2a} = \sqrt{\frac{4a(x-a)}{x-2a}} \tag{11-20}$$

用式（11-17）计算的a代入得

$$\tan\theta = \frac{2\tan\lambda}{1-\tan^2\lambda} = \tan 2\lambda \tag{11-21}$$

而式（11-21）的解是

$$\theta = 2\lambda \tag{11-22}$$

再回到图11-10中得到

$$\Delta x = d\cos(\pi-\theta) = d\cos\theta \tag{11-23}$$

以及

$$\Delta y = -d\sin(\pi-\theta) = -d\sin\theta \tag{11-24}$$

$$x_F = x_p + \Delta x \tag{11-25}$$

$$y_F = y_p + \Delta y \tag{11-26}$$

x_F与y_F可以用作累加器数组的参数。

从式（11-17）到（11-26）的推导提供了使用累加器数组检测抛物线的方法。然而这种方法是很受限制的，其中包含了以下假设:

- 在视场中只有一条抛物线。
- 抛物线对称于一条水平线。

式（11-16）假设焦点在$x=a$处，上述方法只在视场中只有一条抛物线，以及坐标原点位置随意的情况下适用。然而在更加一般性的情况中应该关注原点的位置问题。

为了克服上述困难，Wechsler与Sklansky[11.12]推导了下述方法。

首先，继续认为抛物线对称于一条水平线，但原点则是某个任意点〈x_0, y_0〉，此时抛物线方程变成（参阅图11-11）。

284

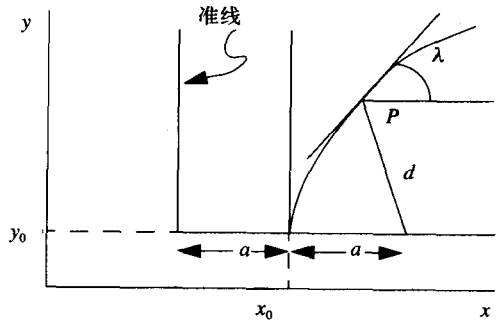


图11-11 任意点的抛物线

$$(y-y_0)^2 = 4a(x-x_0) \tag{11-27}$$

与式（11-17）一样对其求导得到

$$\beta = \tan\lambda = \frac{dy}{dx} = \sqrt{\frac{a}{x-x_0}} \tag{11-28}$$

和

$$a = \beta^2(x - x_0) \tag{11-29}$$

将式（11-29）代入式（11-27）

$$(y - y_0)^2 = 4\beta^2(x - x_0)^2 \tag{11-30}$$

对两边求平方根得到

$$(y - y_0) = 2\beta(x - x_0) \tag{11-31}$$

式（11-31）描述了(x_0, y_0)参数空间中一条直线。于是通过局部求导，式（11-27）中的三参数问题降至更容易操作的二维问题。

显然除了圆、抛物线之外的其他形状也可用这种方法检测到[11.3]。

11A.2 检测峰值

使用这种方法显然希望累加器数组的峰值能落在真实的参数值处或其附近，但一般往往由于原始数据中的噪声而发生偏离。解决这个问题有几种不同的方法，包括一些复杂的方法[11.4, 11.11]与一些简单的方法。一般说来这是一个聚类问题[11.4]，所以能对数据进行聚类的方法，对峰值检测也适用。一些简单的技术经验能解决问题，例如可以用在15.2.2节中描述的k均值方法。 285

聚类是在数据中检测内含类别的过程。在本章中的应用是找到累加器峰值的最佳估计，然而聚类不仅仅只限于检测分布的模式，它有更广泛的应用。例如McLean与Kotturi[11.9]使用聚类技术在图像中检测消隐点[11.5]。聚类将在第15章中详细讨论。

11A.3 高斯图

“高斯图”（Gauss Map）映射为表示距离图像的运算提供了有效的途径。它在原理上是一种参数变换，其概念是十分简单的，首先将一个球的表面分割成由许多面片镶嵌而成的形式，面片尺寸可根据需要选择，而它就决定了这种图的分辨率，每个面片附有一个计数器（累加器）。对距离图像，在其每个像素计算表面的法线，并且为具有相同表面法线的高斯图单元增值，这实质上是法线向量的直方图，而它可以用来对距离图像中物体的方向进行识别。

由于曲率表示法线方向改变的速率，高斯图可以与曲率联系起来。高斯图的可逆性以及在旋转与移位条件下的不变性，在[11.7]中有更详细的讨论。这种图也可用到对图像中鉴别消隐点[11.8]。

11A.4 立体视觉中的参数一致性

在4.2.2节中曾提到对应问题使得立体视觉难以提供有关世界的三维信息。本节再深入地分析一下这个问题，并基于累加器数组原理提供一个部分解决该问题的方法。

下面考虑一个双摄像机立体视觉问题。回顾图4-9可以知道视差是两个对应像素之间的像素距离，z是到对应这两个像素的景物点的距离，而B是基线（两个摄像机之间的距离）。

$$d = \frac{BF}{z} \tag{11-32}$$

其中F是两个摄像机中任何一个的焦距（假设两者的焦距相等）。当然困难的问题是确定（在两幅图像中）哪两个点对应三维空间中的同一点。设想在最左边的图像中抽取出一个小的窗口，并用平方差之和（SSD）沿一水平线在另一幅图像中的小窗口内进行模板匹配。我们

可以用视差或距离倒数来构筑目标函数。可以发现使用距离倒数是更方便的，并且发现典型的情况是匹配函数具有多个最小值，如图11-12所示。如果使用第3个或第4个摄像机取图像，而它们到第1个摄像机的基线距离各不相同，也会得到相似的带凹陷的曲线。然而所有的曲线在对应正确视差的同一点具有最小值。这就体现出一种一致性！因此可以定义一个新的函数（称为以距离倒数计算的SSD），即从多重基线对的摄像机图像中提取这些曲线并求和，则该函数对应于正确答案之处呈现锐利的最小值。Okutomi与Kanade[11.10]已经证实，这个函数在正确匹配的位置上呈现明显的最小值，并且随着基线对数目的增加，度量的不确定性随之下降。

286

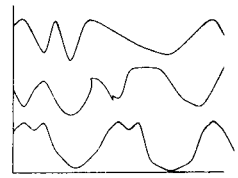


图11-12 立体视觉对中第一幅图像中抽取的一个模板，与第二幅图像进行匹配的结果，是到达匹配点距离倒数的一个函数

11A.5 结论

参数变换的基本概念是搜索一致性！这是其关键。在某种意义上，具有一致性的许多点会为累加器的相同单元中提供累计值——某种意义上称其为“投票”（vote）。所期望的是，噪声的影响在投票过程中会因分布不集中而排除掉，从而可以得到具有一致性的解。

在计算层析技术（CT）中，度量到的信号是沿从x射线源发出到x射线检测器的射线的线积分。执行积分运算的这条线可以用ρ与θ参数表示成直线

$$R(\rho,\theta)=\int(\rho-(x(s)\cos\theta+y(s)\sin\theta))\mathrm{d}s \tag{11-33}$$

仔细观察式（11-33）就可得出结论：Hough变换可以在形式上用同位素变换（Radon transform）表示。除了具体应用不同这一点外，它们是同样的变换。

除了使用这些变换来鉴别特定形状之外，Leavers[11.6]指出通过观察在参数空间的分布，而不仅仅在参数空间找峰值，还可以确定若干区域的凸包与一些形状参数。

Aghajan与Kailath[11.1]曾提出一种取代Hough变换的有趣方法——使用波前传播（wavefront propagation），其原理是把每个像素想像成无线电发射器，所发出的信号被位于每行末端的接收器检测到，他们指出使用关于信号处理到达方向的数学（mathematics of direction of arrival signal processing）可以检测直线，而计算复杂度要远比传统的Hough变换低。用波前传播的概念来解释人脑如何检测直线的一种机制有一定道理。

11A.6 术语

高斯图	gauss map
抛物线	parabola
同位素变换	radon transform
平方误差的总和	sum of squared differences，SSD

作业11.1

电路板“导孔”的手册是一组从电路板导引穿入的导线的图像。这些孔大体上是圆的深

色孔。试用参数变换方法检索这些孔。这是一个程序作业，要求编写程序加以实现，并处理尽可能多的图像。如果该方法有时失效，讨论其原因。 287

作业11.2

试用广义Hough变换来表示一个物体与检测图像中的该物体。该物体是一个很精确的正方形，中心在坐标原点，边长2个单元，但你事先并不知这些参数，你只得到5个点：(0, 1)，(1, 0)，(1, 0.5)，(−1, 0) 与 (0, −1)。画出R表并填上相应项，以备广义Hough变换检测该物体使用。(表11-2有4行，这只是示意，你不一定都填满，如果需要更多行，也可以添加。)

表11-2 R表

?	$p1(x, y)$	$p2(x, y)$

作业11.3

设$P_1=[x_1, y_1]=[3, 0]$，$P_2=[x_2, y_2]=[2.39, 1.42]$是同一圆盘上两点的近似。我们事先并不知道圆盘内部是暗的还是亮的。P_1与P_2点处的梯度分别为$5\angle 0$与$4.5\angle\frac{\pi}{4}$（用极坐标表示）。

试用Hough变换来估计圆盘的中心位置与半径，并确定圆盘比背景暗些还是亮些。

参考文献

[11.1] H. Aghajan and T. Kailath, "SLIDE: Subspace-based Line Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), 1994.

[11.2] D. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, **13**(2), 1981.

[11.3] N. Bennett, R. Burridge, and N. Saito, "A Method to Detect and Characterize Ellipses Using the Hough Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(7), 1999.

[11.4] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[11.5] T. Hofmann and J. Buhmann, "Pairwise Data Clustering by Deterministic Annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), 1997.

[11.6] V. Leavers, "Use of the Two-dimensional Radon Transform to Generate a Taxonomy of Shape for the Characterization of Abrasive Powder Particles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(12), 2000.

[11.7] P. Liang and C. Taubes, "Orientation-based Differential Geometric Representations for Computer Vision Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(3), 1994.

[11.8] E. Lutton, H. Maître, and J. Lopez-Krahe, "Contribution to the Determination of Vanishing points using the Hough Transform," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, **16**(4), 1994.

[11.9] G. McLean and D. Kotturi, "Vanishing Point Detection by Line Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[11.10] M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(4), 1993.

[11.11] J. Princen, J. Illingworth, and J. Kittler, "Hypothesis Testing: A Framework for Analyzing and Optimizing Hough Transform Performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[11.12] H. Wechsler and J. Sklansky, "Finding the Rib Cage in Chest Radiographs," *Pattern Recognition*, **9**, pp. 21–30, 1977.

288 ~ 289

[11.13] Ylä-Jääski and N. Kiryati, "Adaptive Termination of Voting in the Probabilistic Circular Hough Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

第12章 图和图论概念

Functions are born of functions, and in turn, give birth or death to other. Forms emarge from forms and other arise or decend from these.

L. Sullivan

第8章中的连通子图已经涉及到了图论中的术语。我们在前面章节里用到“连通子图”这个术语是将每个像素看作图中的顶点，同时由每个顶点引出4条、6条或者8条边连接到其他顶点，即认为原图像中的像素是四连通、六连通（如果是六边形的像素）或者八连通的。尽管引用了这个概念，我们当时并没有建立完整的集合论或其他数据结构。这将是本章的主要内容。本章讨论的图匹配技术在第13章中有广泛的用途。

12.1 图

图是一个关系数据结构。图由若干数据元素组成，这些数据元素称为顶点（vertices）或者结点（node），顶点之间的联系称为边。

图可以用集合来描述。所有顶点的组合为一个简单集合，所有边的组合为一个有序点对的集合。例如，令$G=\langle V, E\rangle$表示一个图，其中$V=\{a,b,c,d,e,f\}$是顶点集，$E=\{(a,b),(b,c),(a,c),(b,e),(d,f)\}$是边集。图也可以用示意图表示。上面这个例子的示意图表示如图12-1所示，该图有两个连通子图。

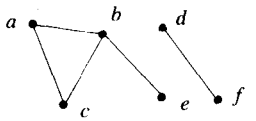


图12-1 一个图，包含6个顶点、5条边、2个连通子图

通过这个例子，我们能够进一步描述边这个概念。图一般是有方向的，也就是说边所描述的点与点的关系是可以有方向的。例如，考虑“above”（上下）和“adjacent to”（相邻）这两种关系，显然，ABOVE$(a, b)\neq$ABOVE(b, a)。所以ABOVE关系可以用有向图表示，而关系ADJACENT_TO则没有方向性。有向图的示意图可以通过在边一端加上箭头来表示。如果我们用包括两个方向的有序点对的描述方式重新描述上面的例子，原图表示为：$G=\langle V, E\rangle$，其中，$E=\{(a,b),(b,a),(b,c),(c,b),(a,c),(c,a),(b,e),(e,b),(d,f),(f,d)\}$。如果$\forall(a,b\in V)[(a,b)\in E\Leftrightarrow(b,a)\in E]$，则定义为无向图；反之，该图是有向图（在某些特殊的情况下，也用到局部有向这个较罕见的术语）。

290

12.2 图的性质

在这一节里，我们定义一些在后续关于图的讨论中将要用到的术语：

- 结点的度（degree）是指进入这个结点的边的数目。
- 结点v_0和v_l之间的路径（path）是指一个结点序列$v_0, v_1, \cdots, v_l$，它满足：对于任意的i，v_i和v_{i+1}之间存在一条边。
- 图连通（connected）是指图中任意两个结点之间存在一条路径。
- 完全子图（clique）是指任意两个结点之间都有一条边的子图。

• 树（tree）是指没有回路的图。在加速马尔科夫模型的应用[12.8]中将涉及这个概念。

12.3 图结构的实现

在计算机中实现存储图结构的第一种数据结构就是如图12-2所示的链表。这一数据结构包括两部分数据:

• 结点（nodes）——由两个指针组成（地址）。

• 元素（atoms）——数据。

我们用某1位来表示元素和结点，该位往往是一个计算机字中最大的一位。某些结点右半包含一个0，用来表示链表的表尾。在图12-2中这些结点用右半部画叉的符号表示。链表同样可以存储计算机指令，这为自动编程提供了强有力的机制。这是程序语言LISP的基础。

链表的概念不仅引入到更现代的程序语言中，而且使其更加通用。例如一种如图12-3所示的结构既包含数据域也包含指针域，这个指针域既可以指向同类指针，也可以指向不同类的指针。下面C语言的定义描述了如图12-3所示的数据结构。

```
struct patch
{
int area;
int perimeter;
struct *patch;
struct *patch;
}
```

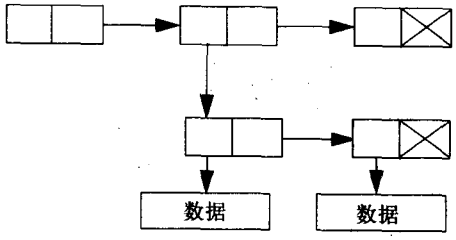


图12-2 在链表中，每个结点包括两个指针。右部含0的指针表示表尾

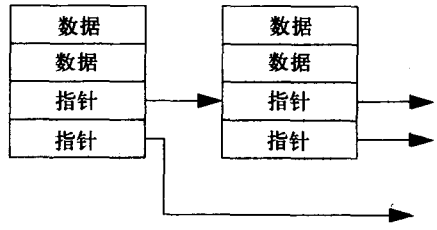


图12-3 更一般的数据结构既包括数据域也包括指针域

291

C语言会将*patch识别为指向另一个结构类型patch的指针。

12.4 区域邻接图

在模型匹配中，我们将用到区域邻接图（region adjacency graph，RAG）来确定图像划分（segmented image）中的区域是否能够匹配上三维模型中的面。图12-4表示一个六面体模型。

图12-4所示模型的RAG表示如图12-5所示。图12-6给出了另外一个例子。

现在的问题是：对于给定一个观测结果及由此得到的RAG表示，和一个模型的集合及其对应的RAG图，要求哪个模型与观测结果最为匹配？我们将在后面详细讨论这个匹配问题。

还有其他的图表示方法，也是可行且有用的，例如，结构实体几何学（constructive solid geometry，CSG）领域使用一组基元的变换表示自动另件制造系统的输入。其中基元是指类

似球体、柱体的物体。参考文献[13.8]提出了使用CSG模型表示的场景与模型的匹配方法以及RAG模型表示。

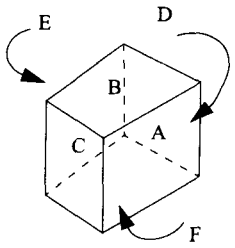


图12-4 六面体

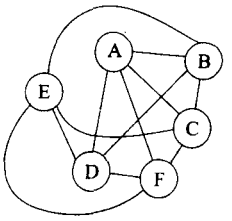


图12-5 图12-4所示三维物体的RAG表示

292

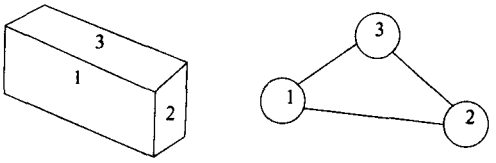


图12-6 图像及其RAG表示

场景图

首先，为了叙述方便，在讨论模型的时候，我们使用“区域”表示可以辨认的表面、面等。当我们讨论分割算法的输出时，观测结果中的“区域”一般用“面片”（patch）来表示。

模型的RAG表示并没有包含很多信息，仅仅包括区域集合及其邻接关系。我们可以通过在RAG中增加信息得到一张新图，这就是我们所说的*场景图*（scene graph）。在场景图中，结点有若干属性，比如，其对应片的面积、颜色、反射率等等，如图12-7所示。

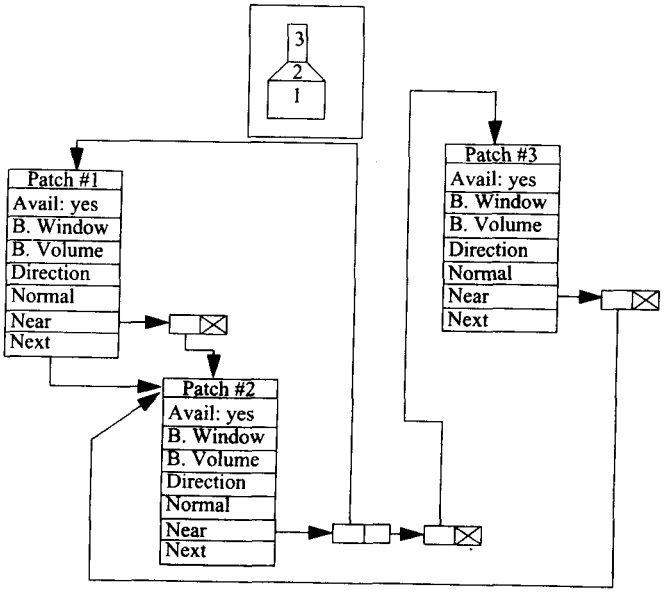


图12-7 被分隔成三个片的物体的图像的场景图

更进一步讲，场景图也可以用来表示多重衡量的关系。在一个简单的RAG中，边仅用来表示结点之间的邻接（ADJACENT_TO）关系，而我们可以同样用边表示仅大于（JUST_LARGER_THAN）关系,从而能够具有一种机制，根据图中的面片大小或者其他属性选择图中的路径。对于某些特殊的应用而言，它能够提供便利。

通常，场景图有以下特点:

293 • 面片链表根据基势（cardinality，片中像素的个数）排序。

• 每个结点都有一个指向相邻面片的指针。

12.5 图匹配的应用：子图同构问题

在计算机科学中，如令n表示问题的数据规模，则根据程序运行的时间与n的函数关系，所有的问题被分为如下几类:

• 多项式时间复杂度问题：时间耗费可以用n的多项式表示，如$t \propto n^3$。

• 指数时间复杂度问题：时间耗费可以用指数表示，如$t \propto e^n$。

• NP难解（np-hard）问题：NP表示非多项式级，而实际是指数级的意思。可以证明NP难解问题的时间复杂度是指数级的，也就是说，可以证明不存在一种多项式时间复杂度的算法能够解决此类问题。

• NP完全(np-complete)问题：NP完全问题是指目前已知的算法中不存在多项式时间复杂度的算法能够求解的问题，但是不能证明不存在这样的算法。任何能够用多项式级时间复杂度解决一个NP完全问题的算法都将是戏剧性的突破，因为所有的NP完全问题是等
294 价的，也就是说，能够解决一个NP完全问题的算法，就能解决其他问题。

子图同构问题是解决场景图与RAG图匹配的一种方法。其原理很简单,我们得到了一个场景图,并假设分割器正常工作(在实际图像中是一个理想化的假设)，此外，对于每个模型，都有一个RAG图。我们是通过寻找与某模型的子图匹配的场景图来实现匹配的。然而，构造出一个模型的图的所有子图的时间复杂度是指数级的，这是一个NP完全问题。使用启发式搜索能够极大地提高算法性能[12.4]。

我们将在13章详细讨论这个问题。

12.6 外观图

假设一个多面体以其重心为原点。考虑一个面向原点的摄像机拍摄的图片，摄像机位于三维空间中球坐标为[ρ, θ, φ]的位置，其中θ表示沿x轴的旋转角度，φ表示沿y轴的旋转角度，如图12-8所示。

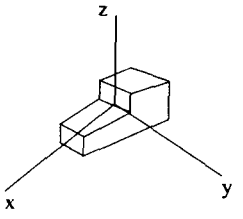


图12-8 令物体的重心在坐标系的原点

如果ρ是常数，则摄像机可能的位置的轨迹是以原点为圆心的一个球面。这里，我们仅考虑这种情况。因此，可以认为摄像机在球面上运动。如图12-9所示，两个不同拍摄位置的拍摄结果截然不同。但是除了少数特殊情况以外，如果假设摄像机的位置只有微小的变动，拍摄结果也只会产生较小的差别。

然而有时当摄像机只有微小移动时，视图却可能发生某些质的改变——物体的某些表面会出现或者完全消失。也就是说，视图的拓扑结构变了。

我们定义两个视点V_1和V_2在外观上是等价的，记做$V_1 \sim V_2$。条件是当且仅当摄像机沿一条

从V_1和V_2的路径,经过一串微小的移动后，所拍摄的图的拓扑结构保持不变。显然，这个等价具有对称性、自反性和传递性。因此，这是等价关系，并将球面上所有的视点划分成等价类——视点空间等价划分（viewpoint space partition），记为VSP。等价划分中的每个元素称为视点区域（viewing region）。图12-10显示了图12-5中的四面体模型的VSP。外观图（aspect graph）是VSP图的对偶。

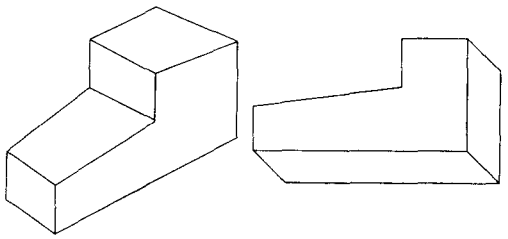


图12-9 两个不同观测位置产生的视图完全不同

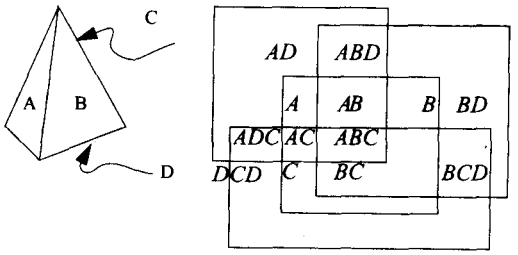


图12-10 VSP的每个划分是根据不同视角所能观测到的物体表面决定的

295

我们可以通过构造标号结构图（labeled image structure graph，LISG）来计算外观图[12.2]，标号结构图是一种带标号的图（augmented graph），其中每个结点对应线条图中的一个顶点，每条弧对应被结点分成的线段。我们给这些弧增加一些标记，+、-和→，如第10章那样用来对应表示凸、凹和遮挡的含义。参考文献[12.2]中提到的分割球算法使得每个分割中的所有点都有同构的标号结构图。

对于使用正投影的任意多面体（可能是非凸的多面体），该算法的（多项式级）计算复杂度很高。对于n面体，最坏情况下的时间复杂度为$O(n^8)$。

外观图最初是由Koenderink和Van Doorn[12.3]提出的，继而被很多人继续研究。Bowyer和Dyer[12.1]在1990前对此做了很好的研究。很多近期的研究都集中在研究采样数据这种十分麻烦的问题上[12.7]。

12.7 结论

图的概念贯穿机器视觉的始终。参考文献[12.6]介绍了更多的场景结构图以及用到了场景结构图的贝叶斯网。正如我们现在所知道的，在10.1节中使用的用于标记线条图的搜索算法实际上搜索了一棵解释树。 296

12.8 术语

外观图	aspect graphs
完全子图	clique
连通	connected
度	degree
边	edge
同构	isomorphic
结点	node
NP完全问题	NP-complete

路径	path
区域邻接图	region adjacency graph, RAG
场景图	scene graph
树	tree
顶点	vertex

参考文献

[12.1] K. Bowyer and C. Dyer, "Aspect Graphs: An Introduction and Survey of Recent Results," *International Journal of Imaging Systems and Technology*, **2**, pp. 315–328, 1990.

[12.2] Z. Gigus and J. Malik, "Computing the Aspect Graph for Line Drawings of Polyhedral Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(2), 1990.

[12.3] J. Koenderink and A. Van Doorn, "The Internal Representation of Solid Shape with Respect to Vision," *Biological Cybernetics*, **32**, pp. 211–216, 1979.

[12.4] B. Messmer and H. Bunke, "A New Algorithm for Error-tolerant Subgraph Isomorphism Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(5), 1998.

[12.5] H. Plantinga and C. Dyer, "Visibility, Occlusion, and the Aspect Graph," *International Journal of Computer Vision*, **5**(2), pp. 137–160, 1990.

[12.6] S. Sarkar and P. Soundararajan, "Supervised Learning of Large Perceptual Organization: Graph of Spectral Partitioning and Learning Automata," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), 2000.

[12.7] I. Shimshoni and J. Ponce, "Finite-resolution Aspect Graphs of Polyhedral Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(4), 1997.

[12.8] C. Wu and P. Doerschuk, "Tree Approximations to Markov Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(4), 1995.

297 ~ 298

第13章 图像匹配

One of these things is not like the other.

Sesame Street

这一章将考虑与匹配有关的问题。匹配是指将观察到的图像与模型匹配，以及图像之间的匹配。我们将考虑图像表达的匹配以及图论表达方式的匹配。

匹配的结果表示对景物的一种解释，也可以说它将两种表达方式对应起来，具有如下特征：

- 两种表达可以是同一种类型的。例如，相关计算将观察到的图像与模板进行匹配。与此相似，子图同构将一个区域邻接图与模型图的子图进行匹配。
- 两种表达也可以是不同形式的。例如一幅图像与描述某件事的一段话匹配。在这类应用中大多是关于用一个方程式与一些数据匹配的问题，在这种情况下，用“拟合”（fitting）这个词似乎更恰当。

在这一章中我们将讨论除了拟合之外的各种匹配问题，而拟合已在本书的前些章中讨论过了。

13.1 图像表达的匹配

13.1.1 模板匹配

模板是图像（或其子图）的一种表达方式，它本身仍然是一幅图像。一般情况下，模板在目标图像（target image，指含检索目标的图像——译者注）中来回移动，直到找到使某个匹配函数值达到最大的位置为止。最常用的函数是式（13-1）表示的平方误差

$$SE(x,y)=\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}(f(x-\alpha,y-\beta)-T(\alpha,\beta))^2 \tag{13-1}$$

（其中，假设模板的尺寸是$N\times N$），它对模板（T）与图像（f）在(x, y)坐标处图像之间匹配的程度提供了度量。如果将平方计算展开并进行求和计算，就可以得到

$$\begin{aligned}SE(x,y)=&\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}f^2(x-\alpha,y-\beta)-2\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}f(x-\alpha,y-\beta)T(\alpha,\beta)\\&+\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}T^2(\alpha,\beta)\end{aligned} \tag{13-2}$$

让我们看一看其中这些项：第一项是图像中所关注点（邻域）亮度的平方和。它与图像与模板之间的匹配与否无关（尽管它的确与图像有关）。第三项只是模板中元素平方之和，是一个常数，与模板处在图像中什么位置无关。显然，第二项是匹配与否的关键，而该项就是相关计算（correlation）。

在使用最佳准则进行匹配的过程中，常假设匹配的质量可以用一组参数$a=\{a_1, a_2, \cdots, a_n\}$描述，这些参数也可以是像素本身。我们定义一个品质函数（merit function）$M(a, f(x))$对模板与局部图像之间的匹配质量进行评估。匹配过程就是要确定使M值达最大的参数a。a往往就是确定模板所在位置的XY坐标，这是很典型的情况。

如果M对a是单调变化的，则可通过下式使M达到最大:

$$M_{a_j} = \frac{\partial M}{\partial a_j} = 0 \quad 对\ j = 1, \cdots, n \tag{13-3}$$

如果M不是单调变化的，那么寻找使该偏导数为零的点的过程就可能停留在局部极大值。此外，我们也曾提到，式（13-3）也可能没有解析解，此时，可以使用爬坡法（hill climbing）。

$$a_j^k = a_j^{k-1} + cM_{a_j} \tag{13-4}$$

一般说来，这种策略对模板匹配是一种困难的方法，原因是，由于$\frac{\partial M}{\partial \alpha}$通常包含多重极小，以及平坦的峰而很难利用。然而一旦碰巧接近到匹配点，则利用爬坡方法就可能获得一个最佳匹配。峰值的宽度可以用某种技巧（heuristics）加以减小，譬如[13.7]中只用纯相位信息进行匹配（phase only matching）。

如果假设对图像的分割结果是正确的，那么匹配就可以在二维的二值形状（binary shape）之间进行，这些形状可以先用一种合适于它们的仿射（或相似）变换进行变换，然后对变换后的区域进行模板匹配。一般说来，可以通过将一些不可能的匹配方案屏蔽掉的方法降低模板匹配的复杂度[13.37]。正如在11A.4节中已看到的，图像之间的匹配是立体视觉系统中极其重要的环节[13.31]。

299

在模板匹配系统中一个基本的问题是对模板尺寸的选择问题。如果模板选择得过小，就难以顾及图像的变化，以致无法做出准确的估计。如果模板太大，则窗口中覆盖的面积太大，就无法实现匹配。因此有一些自适应的算法[13.27]。除此之外，在匹配时要考虑对怎样的目标函数实现最小化。平方误差与信噪比并不一定总是最好的选择。

13.1.2 点匹配

也可以把图像看作是点的集合，点之间的距离已知（这可以看作是下一节讨论的带弹簧模板问题的一种情况，模板本身就很简单）。作为一个例子，在合成孔径雷达（Synthetic Aperture Radar，SAR）图像中识别目标时，这类数据就会出现。此时，解决问题的方法是假设物体的三维模型已知，搜索一个从三维到二维的变换，它能最好地解释观察的图像[13.53]；也可参考[13. 3]。

在立体视觉中进行匹配可以利用边与角点的特征以及外极线约束，也可使用概率模型[13.47]。

13.1.3 线段匹配

短的弧与一组长弧之间的匹配问题也在参考文献中提出[13.3，13.39，13.47]，但相应的三维问题至今仍未引起太多人注意，原因是抽取三维曲线太困难；[13.22]提供了一种方法。

13.1.4 特征图像

特征图像（eigenimage）技术已成为物体鉴别与识别这一类问题的一个有效方法[13.49，13.50]。在这些问题中一个未知物体的图像与数据库（或一个训练集）中的已知物体相比较，如果找到一个很相近的匹配，则该未知物体就被鉴别或识别出来。这种比较当然可以用像素对像素的方式做，但是，如果图像尺寸很大时，以及数据库中图像的数目很大时，这显然是很费时间的。

300

特征图像技术起源于主分量分析（principal component analysis，PCA），这是一种流行的降维方法。一种类型的PCA与K-L变换在9.2.1节中详细地描述。PCA用一组正交基向量来描述数据，这些基向量是由数据生成的协方差矩阵的特征向量组成。将数据投影到占主导地位的特征向量（对应大的特征值）上，可在信息损失最小条件下将原数据集的维数降低。

与此相似，在特征图像方法中，每幅图像也是用一组占主导地位的主分量（特征图像）的线性组合表示，因此匹配转而在线性组合的系数（或对特征图像投影的权重）基础上进行，这大大加速了匹配过程。由于投影保留了大部分能量，因而最大程度地反映了数据库的变化量。下面讨论计算特征图像的细节。

令$f_1, f_2, \cdots, f_p$表示数据库中已知物体的一组图像。为了不失一般性，假设这些图像具有相同尺寸$m \times n$，执行以下步骤可以得到特征图像。

第1步 对每幅图像f_i，减去平均图像A，$A = \frac{1}{p}\sum_{i=1}^{p} f_i$生成相应的字典顺序式表达（即将图像逐行顺序排列起来成向量），从而得到一向量集$\{I_1, \cdots, I_p\}$，$I_i = f_i - A$，每个向量是$mn \times 1$维。

第2步 计算这向量集的协方差矩阵

$$C = \frac{1}{p-1}\sum_{i=1}^{p} I_i I_i^{\mathrm{T}} \tag{13-5}$$

其中C是一个$mn \times mn$矩阵。

第3步 使用特征值分解技术得到矩阵C的特征向量与特征值

$$C = E\Lambda E^{\mathrm{T}}$$

其中E是一个$mn \times mn$矩阵，其每个列向量是C的特征向量（或称C的一幅特征图像），Λ是一个对角矩阵

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ 0 & \cdots & 0 \\ 0 & \cdots & \lambda_{mn} \end{bmatrix}$$

主对角元素是C的特征值，令$\lambda_1 > \lambda_2 > \cdots > \lambda_{mn}$。

第4步 假设在所有mn个特征值中，前k个值要比其余的大很多，也就是$\sum_{j=1}^{k}\lambda_j / \sum_{j=1}^{mn}\lambda_j = \eta$很接近于1。因此可以用前$k$个特征值来构造原图，而不至于丢失太多信息。希望能做到$k << mn$。

301

第5步 为了进行比较，计算每幅图像f_i（减去平均值的图像——译者注）对所选的特征图像的投影系数

$$W_i = I_i^T \times [E_1 \cdots E_k] \tag{13-6}$$

其中E_j是第j个特征图像，W_i是一个$p \times k$的矩阵，包含原图到每幅特征图像的投影系数。则比较过程可以很容易按以下方式进行：给出在待测试图像中未知物体的图像I_{test}，计算其投影系数

$$W_{test} = I_{test}^T \times [E_1 \cdots E_k]$$

将W_{test}与数据库中所有W_i进行比较（欧氏距离可能是最简单的情况）；距离最近者被选作匹配结果。

下面显示将特征图像方法应用到人脸识别的例子[13.51]。假设在数据库中有三幅图像（Lena、Einstein 以及一座钟），而未知图像是Monalisa（蒙娜丽莎）。按照上述第1到第3步，得到64×64特征图像。只使用占主导分量中的两个，因为这两个特征值之和与所有特征值之和之比已接近于1。图13-1表示了所有原图以及由这两幅特征图像合成的结果。按第5步方式，计算所有4幅原图对这两幅特征图像的投影系数，它们也列在图13-1。根据最简单的欧氏距离计算方法，与Monalisa最匹配的是Einstein（爱因斯坦），是否感到惊奇，其实不然，从计算机的"眼睛"看来，这两幅图像的相似程度的确高于Monalisa与Lena之间的相似程度。

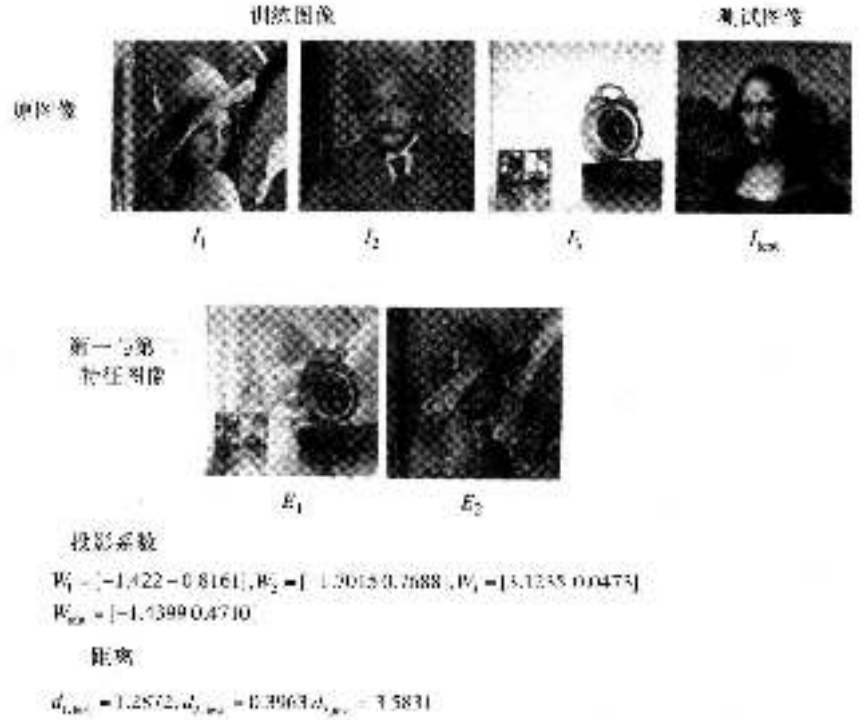


图13-1 利用有三幅图像（Lena、Einstein 与钟）的数据库以及一幅测试图像（Monalisa）为例，演示特征图像方法。这些图像尺寸有所调整，以便于演示

降低计算复杂度

尽管特征图像方法在图像匹配方面很有潜在应用，但是从上述计算过程来看，最费时间的步骤是获取特征系统（指特征值与特征向量——译者注），当图像的尺寸较大时，计算协方

差矩阵（是$mn \times mn$矩阵）会占用大量计算资源，或者就根本无法实现。读者可参考[13.34，13.35]了解更加有效的计算方法。

我们通过一个例子来说明减少计算量的方法。假设每幅图像只有三个像素，并且数据集中只有两幅图像，它们分别为

$$f_1 = [1 \quad 2 \quad 3]^{\mathrm{T}}$$
$$f_2 = [5 \quad 8 \quad 9]^{\mathrm{T}}$$

那么其均值$A=[3\ 5\ 6]^{\mathrm{T}}$，以及

$$I_1 = [-2 \quad -3 \quad -3]^{\mathrm{T}}$$
$$I_2 = [2 \quad 3 \quad 3]^{\mathrm{T}}$$

构造一个矩阵$I=[I_1, \cdots, I_p]$，其中第i列是图像I_i并考虑乘积$S=II^{\mathrm{T}}$。在这个例子中

$$I = \begin{bmatrix} -2 & 2 \\ -3 & 3 \\ -3 & 3 \end{bmatrix}, \quad S = \begin{bmatrix} 8 & 12 & 12 \\ 12 & 18 & 18 \\ 12 & 18 & 18 \end{bmatrix}$$

可以看到如果$p<mn$，那么S就是散射矩阵。它与协方差矩阵除了在一个相乘比例因子外是一样的。如果图像尺寸是256×256，那么S的尺寸是很大的，是$256^2 \times 256^2$。然而，如果说数据集中只有5幅图像，则I是$256^2 \times 5$，而$I^{\mathrm{T}}I$是5×5。如果μ_i是$I^{\mathrm{T}}I$特征向量中的一个，则有 302～303

$$I^{\mathrm{T}} I \mu_i = \lambda_i \mu_i \tag{13-7}$$

以下是数学问题：用I乘（13-7）式两边，得到$I\ I^{\mathrm{T}}I\mu_i=I\lambda_i\ \mu_i=\lambda_i\ I\mu_i$，注意，$I\mu_i$是$I^{\mathrm{T}}I$的一个特征向量。由于$I^{\mathrm{T}}I$的尺寸要远小于$II^{\mathrm{T}}$，则在决定特征向量的过程中计算复杂度也相应极大降低。这样一来，如果e_i是S的特征向量，则可以用下式得到

$$e_i = I\mu_i = \sum_{k=1}^{p} \mu_{ik} I_k \tag{13-8}$$

13.2 匹配简单特征

在第9章中讨论过使用简单特征的最直接的方法，是在模式分类器中使用它们。为了做到这一点，要为模型与（待匹配）物体提取统计表示式，并用这些表示式进行匹配。其策略如下：

- 确定用何种度量方法来描述形状。例如，可以建立一个有8个特征的系统，其中7个是不变矩，另一个是长宽比。最好的特征组合是依赖于具体应用的，而最佳地选择特征集的方法超出本书范围（参考[14.4，14.11，18.30]，这些只是许多统计方法教科书中的一些例子）。将这8个特征组织成向量$\boldsymbol{x}=[x_1, x_2, \cdots, x_8]^{\mathrm{T}}$
- 用一组作为例子的图像（称为“训练集”），在这些图像中抽取出相应特征，并用此来描述“模型”物体。以下继续讨论有8个特征的例子，我们可以收集斧子（axe）的几幅图像，对每个斧子的特征进行度量，然后用这些斧子特征的平均值来描述斧子的“模型”，$\mu_{\mathrm{axe}} = \frac{1}{n}\sum_{x \in \mathrm{axe}} \boldsymbol{x}$。手斧（hatchet）也可以用一组手斧的样本图像建立模型。
- 按上所述，对一个给定的未知区域用一个特征向量$\boldsymbol{x}$表示，那么形状匹配就是要找到一个模型，它在某种意义上与所观察到的区域“最接近”。一般“接近”的最简单定义是

使用欧氏距离:

$$d(\text{modelaxe}, \text{observation}) = \sqrt{(x-\mu_{\text{axe}})^{\text{T}}(x-\mu_{\text{axe}})}$$
$$d(\text{modelhatchet}, \text{observation}) = \sqrt{(x-\mu_{\text{hatchet}})^{\text{T}}(x-\mu_{\text{hatchet}})}$$

而决策由距离最小者确定。

这个例子的一些计算还是过于简朴的，有比使用均值更好的建模方式，使用其他的距离度量也可能会比欧氏距离的效果更好。统计模式识别领域的一些概念会在第14章更详细地讨论。

304

13.3 图匹配

在这一节我们讨论的问题基本上仍是基于图的图像表示匹配问题。但是在图的结点中存的数据包括图像或模板。

回顾一下N尺度的完全子图（clique）是尺寸为N的全连通子图。我们仍然想使用诸如平均方误差或13.1.1节中提到的相关作为匹配度量，来度量匹配的“良好程度”。

在一个单纯基于图的表示中，一个匹配度量可以是子图同构。但是子图同构并不真正允许相近但并不完美的匹配。大多数机器视觉的专家认为它的变通性太差。

图匹配问题也可以用退了火的神经元网络（annealed Neural network）实现[13.8]。

在前面已看到，松弛标号也能为某种类型的图匹配提供机制。在10.2.2节看到的例子中能够用模型图的子集与场景图的子图进行匹配（例子中的两个表面）。也有些变型方法，例如Gold与Rangarajan [13.14]描述了一种图匹配的变型方法，他们使用的是一种非线性优化方法。他们宣称，这种方法比松弛标号法运行得快得多，也准确得多。

下面讨论另两种方法——*关联图*（association graph）与*弹性负荷模板*（spring-loaded template）。这些方法能用于混合表示方法产生匹配，也就是说，它基本上是基于图的方法，但包含有图像的信息。

13.3.1 关联图

关联图体现了一种比同构的限制要少而收敛得更快的方法。它会收敛到一个符合一致性的解，但不一定最优（当然这取决于具体应用中使用的最优准则）。

305

这种方法将模型的一组结点与从图像中（抽取出）的一组结点匹配。

定义

这里一幅图定义为$G=<V, P, R>$，其中V表示一组结点，P表示一组用于结点的单元谓词，而R表示结点之间的二值关系。

谓词（predicate）是一种只取True 与False两值之一的语句。例如令χ表示距离图像中的一个区域，那么CYLINDRICAL（x）是一个谓词，它是真（true）或假（false）取决于x内的所有像素是否都处在一个柱形表面上。

一个二值关系描述一对结点所具有的属性。它可以表示成一组有序对$R=\{(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)\}$。在大多数应用中这种顺序是很重要的。也可以把关系想像成谓词，因为对任何给定结点对，如(a_k, b_k)，可以是R集合中的一个元素，也可以不是。但是在以下讨论中用关系这个词似乎描述起来更方便。

给定两幅图，$G_1=\langle V_1,P,R\rangle$ 与$G_2=\langle V_2,P,R\rangle$，可构筑一个关联图。

- 对两幅图中每个结点 $v_1\in V_1,v_2\in V_2$ 如果v_1与v_2具有同样属性，则在G中建立一个结点，标为(v_1, v_2);
- 如果$r\in R$以及 $r(v_1,v_1')\Leftrightarrow r(v_2,v_2')$，则将$(v_1, v_2)$与 (v_1',v_2') 连起来。

则G_1与G_2之间的最佳匹配是G的最大完全子图。

与机器视觉中任何其他技术一样，需要问："这种方法有多好？"，要想回答这个问题时又引起一些其他问题。

问题1 最大完全子图是否是最佳匹配？最大完全子图是一致性匹配的最大集，它是不是确实是最佳匹配？

问题2 计算复杂性，与子图同构问题一样，寻找最大完全子图的问题是NP完全问题，也就是说还没有用少于指数时间解此问题的算法。

使用关联图将场景与模型进行匹配的一个例子

在图13-2中表示了一个观察中出现一个分割错误，过分割的情况，其中区域B与C实际上都是同一区域的一部分。由于某种度量或算法错误，被标成两个分开的区域。在该例中所用的单元谓词是标号spherical（圆形的），cylindrical（柱形的）与planar（平面的）。区域A与1是圆形的，而B，C，D，2与3是柱形的。只有具有相同谓词的结点才能作为匹配的候选。所以只有A与1匹配，所有侯选匹配作为结点构造一幅图，就得到了图13-3所示的关联图的结点。 306

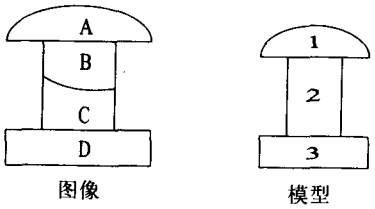


图13-2 距离摄像观察到一个场景，并将其分割成一段段满足同一方程的分段，但是有一个分割误差

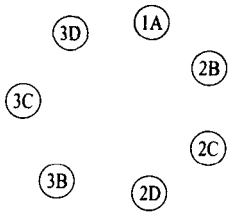


图13-3 候选匹配

问题的挑战性在于：鉴别出哪些是符合一致性的——确定r_A（1，A，2，B），其中兼容函数r的含义与第10章中的相同，而下标A只是表示正在使用一个关联图。一般说来，确定哪些是不一致的（Not consistent）要容易些，并且这是一个与问题有关的决定。在这里定义两个标号具有一致性，条件是只要它们并不对应于同一区域（即一个区域不能有两种不同标号）。对这个例子来说一些一致性可表示成

$$r_A\ (1,A,2,B)=1$$
$$r_A\ (2,B,2,C)=-1$$
$$r_A\ (2,B,3,B)=-1$$

其中第二行的含义是图像中的B区域以及C区域不能同时对应于模型中的2区域，在两个例子中对不一致性的定义都是根据分割算法工作正确为前提的。然而应该允许分割结果有错误。此时关联图中新的边就应该加入，因为有新的关系体现了一致性。例如r_A（3，C，3，D）=1，因为我们认为两片可以是同一区域的一部分（分割算法可以由于过分割导致错误），但是r_A（2，D，3，D）=－1仍然成立，因为我们仍然认为分割算法不会将区域合并起来（由欠分割

导致的失误）。允许过分割就得到了图13-4所示的关联图。

要注意另一种类型的不一致性会阻止一些边的生成：3D与3B是不能连结起来的，因为B与D并不共享边界。这也就是说，我们认为分割的失误如果是由过分割引起的，那么分割器不会在这两个之间引入一个完全新的值。这里必须强调，制定这些规则完全取决于问题。

如果制定这些规则，匹配就是直截了当的事，只要找到所有最大的完全子图就行了。最大完全子图并不是唯一的，因为很有可能有几个尺寸相同的完全子图。

在这个例子中，至少有两个最大完全子图，其中两个为：{(1, *A*) (2, *B*) (2, *C*) (3, *D*)}和 307 {(1, *A*) (3, *B*) (2, *C*) (2, *D*)}。

13.3.2 弹性模板

另一种用于局部/全局问题的方法是“弹簧和模板策略”[13. 13]。这是一种混合模型匹配方法，其中包含图结构的匹配与模板匹配。这种模型是一组刚性“模板”，用“弹簧”连起来，这种连接描述了对模型要加多少变形才能使其与图像匹配。图13-5用一幅简单的人脸模型来说明这种概念。

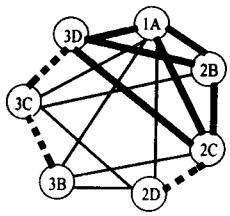


图13-4 细实线表示假设分割器没有错误时能出现的边，虚线表示允许过分割错误的情况，粗线表示最大完全子图

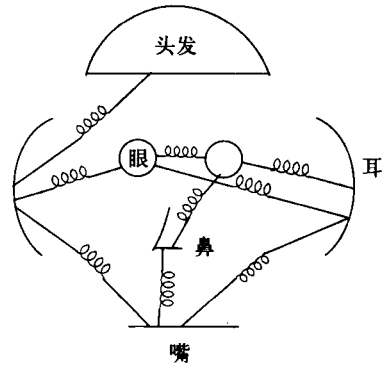


图13-5 人脸的一个弹性加模板模型

其中一些特定的特征，如眼睛等采用图像（iconic）表示方法，使用模板匹配方法进行匹配。然而为了匹配整张脸，模板最佳匹配的位量之间的距离也被记录下来。（例如模板匹配后眼图像与鼻图像之间的距离。——译者注）匹配则是基于如下总成本函数的最小化

$$\begin{aligned}\text{Cost} = &\sum_{d\in\text{templates}} \text{TemplateCost}(d,F(d)) \\ &+ \sum_{d,e\in\text{ref}\times\text{ref}} \text{SpringCost}(F(d),F(e)) \\ &+ \sum_{c\in(R_{\text{missing}}\times R_{\text{missing}})} \text{MissingCost}(c)\end{aligned} \tag{13-9}$$

在式（13-9）中d是模板，$F(d)$是图像中应用该模板的点。因而TemplateCost是一个函数，它表示当一个模板用到它的最佳匹配点时，模板与图像匹配的程度。（匹配程度越高，该项函数值越小。——译者注）SpringCost则是度量为了使这些特定模板在这些匹配位置上匹配，模

型所需的变形（弹簧的拉伸）。还要提到的最后一点是，并不一定要求每个模板都要找到其位置，譬如在某些图像中左眼可能看不到，此时需要用为丢失的物体付出代价（Cost既可译代价，也可译成本。——译者注）。所有这些成本都是根据经验确定的，但是一旦它们被确定下来，那么给定图像与任何给定模型的匹配好坏就容易确定了。

这个问题并不只是弹簧加模板才有。一般常用的方法是对成本计算进行规范化，譬如使用如下方式

$$\text{Cost} = \frac{\text{SpringCost(unary and binary)}}{\text{Total number of springs}} + \frac{\text{constant}}{\text{Total number of references matched}}$$

13.4 结论

关联图直接使用一致性标号的概念与体系。使用图结构的优点在于，搜索最大完全子图可以求助于大量现成的执行这种搜索的软件，在计算复杂性限定条件下尽可能地快。而弹簧加模板的想法既度量一致性，又度量偏离一致性的程度。（其实式（13-9）度量的是匹配不一致性，方位匹配的不一致性，以及无法匹配体现的不一致性，用成本表示。—译者注）弹簧加模板的概念也说明了人们怎样做才能构造一个合适的目标函数，以及如果人们不注意对目标函数的分析，问题就很容易产生——如果我们将匹配质量求总和，许多事件的一个好的匹配（将数目众多的小数值加起来）可能会大于（并因而要差于）一个只对少量事件匹配的差匹配（对数量不多但数值较大者求和）。

在本章一开始我们就曾指出正规的优化方法，如梯度下降法或“爬山”法，很难用到图像匹配上，这是因为搜索空间布满局部极小值。但是，如果我们的算法的初始值足够接近问题的解，这种技术还是能工作的。我们使用平方差的总和（SSD），有时也称平方误差总和，作为目标函数。

特征图像是原图像的低维表示。选择这种投影能使原数据与投影后的数据之间的误差最小。

13.5 术语

关联图	association graph
完全子图	clique
对应	correspondence
变形模板	deformable template
特征图像	eigenimage
爬山方法	hill - climbing
匹配测度	matching metric
主分量分析	principal component analysis, PCA
模板	template

308 ~ 309

作业13.1

在这一章中，我们指出找最大完全子图的问题是NP完全的。这究竟意味什么？假设你有一个10个结点的关联图，用20根边连接，你需要进行多少次测试才能找到所有完全子图（为了确定哪一个完全子图是最大的，你必须做这件事）？请从图论书中找到完全子图检测方法。

作业13.2

在13.3.1节有一个例子是关于关联图的，它容忍分割时出现误差，这幅关联图的结果是两

个最大完全子图，它们意味着对景物的不同解释。用语言描述这两种解释。

作业13.3

在本章的参考文献中，对Olson[13.36]有一个不完整的引用。首先，找到这篇文章的拷贝。你可以使用一个搜索引擎、Web网、图书馆，或任何你想用的资源。在那篇文章中，作者以一种不同的方式进行模板匹配；使用一个二值（边缘）图像与一个相似的模板，他并没有问“该模板与这一点附近的图像匹配吗？”而是问“在这一点，它离最近的边缘点有多远？”

他是如何高效地执行这种搜索操作的？

310 一旦他得知到最近边缘点的距离，他是如何利用这种信息来计算匹配度量的质量的？

作业13.4

在一个图像匹配问题中，我们有两种类型的物体，狮子（lion）和羚羊（antelope）（每个物体只有一个像素大小）。

- 一个场景可能只包含狮子与羚羊。
- 狮子成群觅食，所以你如果看到一只狮子，会至少看见另一只狮子，通常距离有5个像素。
- 羚羊是尽可能接近地靠在一起。
- 除了某些极少出现以及（对羚羊）不愉快的事件发生，狮子与羚羊之间隔的很远。

我们希望用松弛标号方法来解这个赋标号问题。除了一致性函数$r(a, \lambda_1, b, \lambda_2)$之外，所有的公式都可在书中找到，$a$与$b$是图像中感兴趣的点，$\lambda$表示或是“羚羊”或是“狮子”的标号。为此问题设计一个r函数，也就是如何计算以下函数：

1）$r(a, lion, b, antelope)$

2）$r(a, lion, b, lion)$

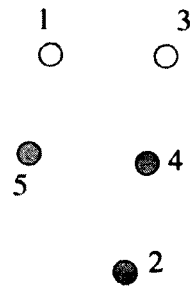
3）$r(a, antelope, b, lion)$

4）$r(a, antelope, b, antelope)$

作业13.5

你认为弹簧加模板的概念可以应用到作业13.4中吗？讨论一下。

作业13.6



继续考虑狮子与羚羊的问题，你从另一角度观察场景：草图没有严格按照比例，但是为了方便于你，我们已将每对动物之间的距离列在表中（见表13-1）。

狮子是黄色（在草图中用灰色孔表示）或棕色（用黑色孔表示，在图中实际上并没有——译者注）。羚羊是白色的（用白色孔表示）或黄的。希望用关联图方法求解该问题；由于这种技术不如非线性松弛法强有力，你请教一位植物学家，是的，的确是一位植物学家！这超出了他的管理范围。他给予你一些改进的信息：狮子永远不会彼此接近到距离在3像素之

311 内，而羚羊之间距离则永远不会超过3像素。

对此问题画一幅关联图（在关联图中结点用成对的符号表示，例如1L表示“将结点1解释

成狮子”)。最大完全子图的结点用圆圈圈出。

表13-1 成对动物之间距离

对	距离（任意单位）
1,2	5.5
1,3	2
1,4	3
1,5	2
2,3	2
2,4	3
2,5	4
3,4	2
3,5	3.8
4,5	3.4

专题13A 匹配

13A.1 弹簧加模板回顾

回顾一下在4.2.2 节讨论过的对应问题，它可以重新表示为“给定在两幅图像中的一组特征，鉴别出在图像1中哪个特征与图像2中的哪一个特征对应。”模型匹配的问题与对应问题之间没有太大差别，除了在对应问题中两幅图像都受到噪声影响。

为了求解对应问题，可以使用基于一致性的原理。在这一节中用一个例子来叙述这种原理，这个例子也将在13A.2中使用。第一步是鉴别一些相对容易区分的特征点。接着算法要利用这些点之间的关系。在一个区域边界上曲率改变符号的点满足这个要求，就如图13-6所示。

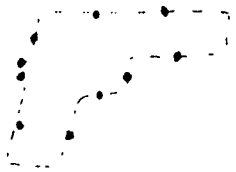


图13-6 一个物体的边界，其上一些曲率改变符号的点被标出

312

这个例子的推导最早是由Shapiro与Brady描述的[13.46]，他们使用了如下的特征向量方法。

与弹簧加模板的原型表示方法一样，在一堆特征点集中找出一个集与某一特定集最佳匹配。令d_{ij}表示特征点x_i与x_j之间的欧氏距离，并构筑一个权重矩阵

$$H=[H_{ij}],\qquad H_{ij}=\exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right) \tag{13-10}$$

矩阵H用标准方法对角化成三个矩阵的相乘

$$H=E\Lambda E^{\mathrm{T}} \tag{13-11}$$

其中E是一个矩阵，以H的特征向量为其列向量，Λ是一个对角矩阵，其对角元素是H的特征值。假定E和Λ的行与列是经过排序的，从而使得特征值沿对角以降序排列。把E的每一行考虑为一个特征向量，用F_i表示，则有

$$E=\begin{bmatrix}F_1\\ \cdots\\ F_m\end{bmatrix}$$

假设有两幅图像，f_1与f_2，并且设f_1中有m个特征点，而f_2中有n个特征点，并且假设$m<n$。那么如果将每个特征点集看成是独立的，则对f_1图像有$H_1 = E_1\Lambda_1E_1^T$，对图像f_2有$H_2 = E_2\Lambda_2E_2^T$。由于每幅图像的特征点数目不同，矩阵H_1与H_2有不同数目的特征值。因此只选最主要的k个特征进行比较。

一个重要的事情是，进行匹配的特征向量的方向要一致，但改变其符号并不影响正交与归一性。我们选择E_1为参考，然后将E_2的轴的朝向调整至这样一种方向，使得两组特征向量对得最齐；可参考[13.46]了解细节。在将轴对准以后，一个表征图像1与图像2之间匹配程度的向量Z定义为

$$Z_{ij} = (F_{i1} - F_{j2})^T(F_{i1} - F_{j2}) \tag{13-12}$$

则最佳匹配就可由Z的元素表示，这些元素在它们的行与列中分别是最小的。我们将在下一节再介绍这个例子。

Sclaroff 与Pentland[13.44]采用了另一种弹簧加模板的表示方法：首先，用一种对采样与参数化误差鲁棒的方法计算整个形状的描述。然后使用整个形状的描述找到一个有效表达该形状的坐标系统。将这种做法用到图像与模型中，对确定作为基准的方向（cardinal direction）是很直接的。

W_u [10.19]在计算光流的问题中使用松弛标号法来检测一致性模板的匹配。

变形模板的概念可以与图表达方式相结合，匹配形状相似但又不完全相同的物体（例如手的x射线图像）[13.1]。变形模型的想法可以看作MAP方法的扩展，可以参考[13.26]，它对此有一简单扼要的描述。与此类似的方法在自动目标识别（ATR）中也被应用到目标跟踪中[13.12]。

313

13A.2 人工神经元网络用于物体识别

我们已经讨论过对被观察的事物进行度量得到一组特征后，模式识别技术为我们提供了决策“它是什么？”的方法。开发分类器有许多方法，使用人工神经元网络原理的方法是其中最为成功的一种方法。神经元网络以物体的特征作为输入，在其输出端做决策。它们是基于人们对每个神经元是如何计算的机理的了解，经过数学抽象的原理工作的。

能够执行匹配计算的神经元网络有两种类型，前馈（feedforward）和递归(recurrent)网络。

13A.2.1 前馈神经元网络

在前馈神经元网络中，每个计算单元（以后称之为“神经元”）有数量较大的输入端和单个输出端。尽管在文献中已开发出许多种神经元模型，但单个神经元最常用的计算方式仍然是 $y = S(\sum_i w_i x_i)$，其中S是S形函数（sigmoid），x_i 是输入，w_i 是权重，用来修改到神经元的各个输入量的重要性。图13-7显示了一个神经元的结构与命名惯例。

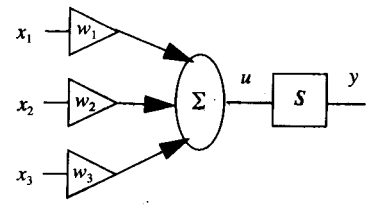


图13-7 由单个神经元执行的计算。每个输入（x_i）都乘以一个权重w_i，其结果加在一起产生一个信号u，它经过一个类似于S形非线性函数（S）产生神经元的输出y

几乎所有的任意函数都可以用图13-8所示的分层神经元网络来计算。神经元网络设计的主要问题是确定各个权重值，使得输入能得到适当的输出。在前馈网络中，有关模型的信息是嵌入于权重值中的，当多半以形状特征方式

表达的图像信息输入到该网络时，如果这些特征向量与模型相匹配，网络就会回答“是”。确定这些权重值的主要方法是通过解梯度下降问题找到这些权重，梯度下降法多半是通过减少神经元网络实际输出与理想输出之间差异的均方值来实现。这样一种算法用所熟知的梯度下降规则实现

$$w_{ij}(t+\Delta t)=w_{ij}(t)-c_2\frac{\partial}{\partial w_{ij}}MSE$$

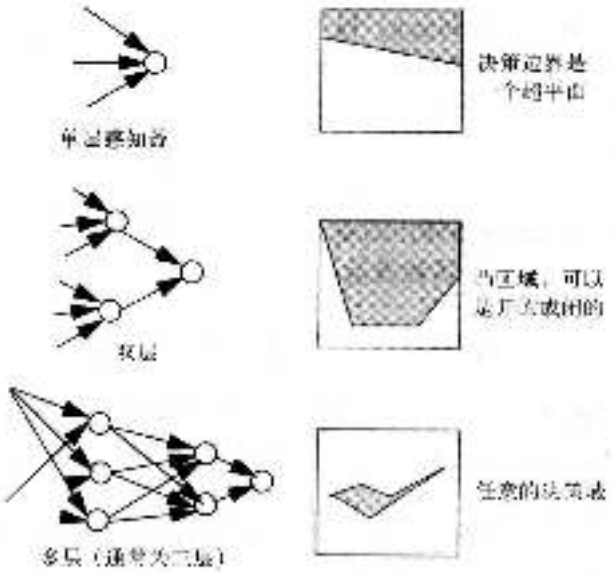


图13-8 前馈神经元网络的类型，以及它们能实现的决策域

使用图13-9所示的三层神经元网络，梯度下降规则可以很容易地用导数的传递规则来实现。Hussain和Kabuka[13.24]演示了用神经元网络进行字符识别的结果。

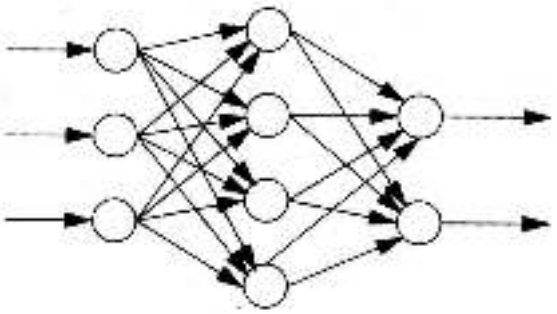


图13-9 一个有三个输入端与两个输出端的前馈神经元网络。每个圆圈表示一个神经元，权重没有明显地表示，而是存在于连接中

314
~
315

13A.2.2 递归神经元网络

一个递归神经元网络（NN）是在运行时将输出回馈到输入端的网络，如图13-10所示。使用先前用过的相同表示，在网络处于稳态时神经元的输出满足

$$v_i = S(y_i) = S\left(\sum_{j=1}^{d} w_{ij} v_j - I_i\right) \tag{13-13}$$

神经元的这种性能模型只是在稳态时成立。这是因为神经元的输出取决于它的输入，而这些输入值又是（它以及别的神经元的）输出，它们又取决于……（对人类来说是迭代，对占卜来说是预言——L.P.Deutsch）但是当事情在变化中时，这种表示式是极端不合适的。此时需要有关于系统动力学的模型。有许多种模型可以用，读者可以参考[13.15，13.20，13. 23]作进一步的了解。这里我们只考虑一个简单模型，其中从求和器输出的变化率与输入有关，并可用一阶微分方程表示。

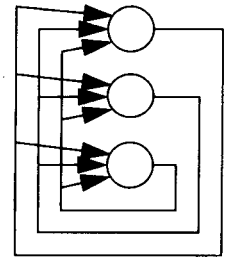


图13-10 一个有三个神经元的递归神经元网络，权重没有画在上面，但每个神经元的每个输入都有相应4权重

$$\frac{\mathrm{d}}{\mathrm{d}t} y_i(t) = -\beta y_i(t) + \sum_{j=1}^{n} w_{ij} S(u_j) - I_i \tag{13-14}$$

其中y_i与以前一样，是神经元的输出，w_i是权重，而I_i是从外部世界来到每个神经元的输入（在图中没有表示）。于是输出的改变与当前的状态、从所有其他神经元来的输入以及外部输入成正比。

在运算过程中，一个特定的输入送到一个递归神经元网络中，然后神经元网络就开始运行。每个神经元应该收敛到一个特定的状态。

霍普尔德曾描述过这个模型[13.23]，也有其他人描述过这个模型。在Hopfield的模型中，递归网络是用阻容运算放大器实现的，而变化常数β由电容与电阻模型确定。

我们暂且将（13-14）式放在一边，考虑以下一个欲使其最小化的目标函数

$$E = -\frac{1}{2}\sum_i \sum_j w_{ij} v_i v_j + \sum_i \beta \int_0^{v_i} S^{-1}(v)\mathrm{d}v + \sum_i I_i v_i \tag{13-15}$$

如果希望找到使该函数最小化的各个v的值，则需要求E对各个v值的偏导数，如此做就得到

$$\frac{\partial E}{\partial v_i} = -\sum_j w_{ij} v_j + \beta S^{-1}(v_i) + I_i \tag{13-16}$$

从该式可以看出E与变量v的偏导数与Hopfield模型的动力学模型具有相同的形式或者可写成

$$\frac{\partial E}{\partial v_i} = -\frac{\mathrm{d}u_i}{\mathrm{d}t} \tag{13-17}$$

设想一下式（13-17）描述的系统的稳态。当该网络的变化结束时（所有随时间的导数都为零），所有能量函数的偏导数也都为零，因此我们得到的是极值。其实式（13-15）中恼人的积分是可以忽略不顾的，这一点是很容易证明的，因此一个Hopfield神经元网络能找到一组变量v_i，它使式（13-15）（没有积分项）描述的目标函数达最小值。我们通过一个例子说明这样一种网络在匹配中的用途。

316

使用与前一节相同的特征集，边界曲率为零的点，我们给每个特征点赋予一个局部性度量，

这里用到（两个）相邻点的向量之间的夹角，如图13-11所示。我们将用这个度量以及一个更加带全局性的特征——特征点之间的距离，来解决对应问题。

假设图像1（可以把它想像成模型，如果你希望这样做的话）有n个特征点，图像2有m个特征点。我们定义一个神经元矩阵，其中有n列与m行。在第i行j列的神经元的值在0与1之间，它取决于第1幅图像中第i个特征点与第2幅图像的第j个特征点之间匹配的程度。

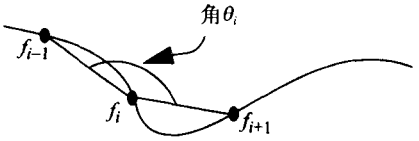


图13-11 点i到两个相邻特征点向量之间的夹角是该特征点的一个局部度量

匹配的过程可以表示成使下式最小化的问题

$$E=-\frac{A}{2}\sum_i\sum_j\sum_k\sum_l C_{ijkl}V_{ik}V_{jl}+\frac{q}{2}\left(\sum_i\sum_k\sum_{k\neq l}V_{ik}V_{il}+\sum_k\sum_l\sum_{i\neq j}V_{ik}V_{jk}\right) \tag{13-18}$$

其中第1项检验ik与jl之间匹配的兼容性。而后两项包含在内是促使匹配的唯一性（一幅图像的某个特征都只能与另一幅图像的一个特征匹配——译者注）。选择这种形式的函数是允许特征之间可能有遮挡关系。兼容系数是这三项之和

$$C_{ijkl}=\omega_1\Gamma(\theta_i,\theta_k)+\omega_2\Gamma(\theta_j,\theta_l)+\omega_3\Gamma(r_{ij},r_{kl}) \tag{13-19}$$

其中

$$\Gamma(a,b)=\begin{cases}1 & (|a-b|<T)\\ -1 & \text{其他}\end{cases} \tag{13-19a}$$

其中T是阈值；θ_i 是对特征点i的局部度量，如图13-11所示；而r_{ij} 是对特征点之间关系度量的相似度的度量。例如，如果i 与j点之间的距离与k与l点之间的距离相同，则标号ik与jl 是一致的。

式（13-18）经过适当处理，可以使它表示成如式（13-15）的形式，从而可以用神经元网络加以最小化。[13.28，13.54]有更多的细节。在图13-12中显示的是一支手枪的轮廓，其中部分被一锤子遮挡。使用这个原理的神经元网络可以从这幅图像中鉴别出这两个物体。

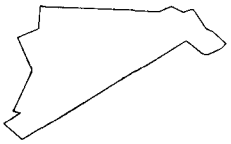


图13-12 一支枪的影子轮廓，其中有部分被锤子遮挡

317

13A.3 图像索引

到目前为止，我们讨论了在模型数据库中搜索一个与所观察到景物匹配程度最小的图像匹配技术，但我们并没有涉及“搜索”过程本身。搜索当然可以采用试探所有模型的办法，但是太耗费时间，以致于不可能采用，尤其当模型数据库中的模型数量极大时更是如此。在自动目标识别这类应用中，要求搜索大数据库，且速度要快[13.45]。因此要求更好的方法。一种替代的策略是索引（indexing）（有时称为图像散列，image hashing），它在[9.6]中分析。在索引的方法中从图像中抽取出一组参数，显然这些参数需要对尽可能多的图像变换保持其不变性，并且性能要鲁棒[13.1]。所得到的参数向量就用作一个查找表的索引，这个查找表包含查找模型的信息。在查找表查找的结果返回一张候选模型的名单，这些候选模型对该参数向量具有一致性。为了说明它为什么能有效，考虑以下的算法。

首先从观察边界处的局部区域开始，并力图将局部区域与线、圆弧以及最小、最大曲率等特征描述量相匹配。假设对未被遮挡的一个物体的分割是正确的，我们从一幅边缘图像开

始，其中边缘并不要求是连续的。在边缘上某点$[x_0, y_0]$周围，我们用一种采样策略[⊖]对其邻域进行采样，这种采样对图像放缩具有不变性。对该点与其相邻两个点生成所有可能的组合，并产生一种与在[9.37]中描述相似的不变参数向量。这个参数向量用来对局部形状的数据库进行索引。对每个所选择的项，在表中索引出相应的特征例证（feature instance）（例如是哪个模型在图像中体现的方位。——译者注），在所有三种特征（指线、圆弧及曲率等——译者注）都已考虑之后，获得索引率最高的特征例证被选出。

此时物体的边界就可以用一个特征例证序列表示，而索引方法可以在高一层次（物体模型层次）重复，此时使用一个关于物体模型的查找表，它用特征例证的几何以及是否出现作为索引。

还有相当多的索引[13.5，13.32]；在[13.48]中对此有很好的综述。某些索引方法对空间的需求在[13.25]中分析。

随着数据库变大，必须在整个数字库的环境中考虑图像索引问题。我们建议读者参考*IEEE Transactions on Pattern Analysis and Machine Intelligence* (Aug.1996)的一个专集，它讨论这个问题。

13A.4 匹配几何不变量

我们从简单地找到一组不变的数值开始，这种方法是要在三维模型中找到5个点，并从它们计算出某些特性，能够以某种不变量方式唯一表征这些点。然后在图像中找出5个点并确定它们与哪个模型最匹配。

在三维模型中选择5个点$\{X_1, X_2, X_3, X_4, X_5\}$，其中至少4个点是非共面的。由于5个点不可能线性独立，可将其中之一写成其余点的线性组合。选择x_5用这种方法表示，使用齐次坐标（9.1节）得

318

$$X_5 = aX_1 + bX_2 + cX_3 + dX_4 \tag{13-20}$$

利用点矩阵的行列式对刚体运动是不变的这一现象[⊜]，并且将由5个点中之4个构造的行列式，并用被弃用的点的下标作为下标，例如

$$M_1 = |X_2\ X_3\ X_4\ X_5| \tag{13-21}$$

将X_5的线性相关表示（式13-20）代入（13-21）式中X_5的每个场合，得到

$$\begin{aligned} M_1 = a|X_2\ X_3\ X_4\ X_1| + b|X_2\ X_3\ X_4\ X_2| \\ + c|X_2\ X_3\ X_4\ X_3| + d|X_2\ X_3\ X_4\ X_4| \end{aligned} \tag{13-22}$$

由于有两列相同的矩阵的行列式是零，则有

$$M_1 = a|X_2\ X_3\ X_4\ X_1| \tag{13-23}$$

由于行列式中任两列互换只需将行列式的符号互换，因而又可进一步简化得

$$M_1 = (-a)|X_1\ X_3\ X_4\ X_2| = a|X_1\ X_3\ X_2\ X_4| = (-a)|X_1\ X_2\ X_3\ X_4| \tag{13-24}$$

所以有

$$M_1 = -aM_5 \tag{13-25}$$

⊖ 为了避免因引入许多细节而使索引算法的叙述过于零乱，请原谅我们省略了某些细节，它们都在参考文献的文章中。

⊜ 事实上，线性形式的绝对不变量总是行列式幂的比[13.19]。

与此相似

$$\begin{aligned} M_2 &= bM_5 \\ M_3 &= -cM_5 \\ M_4 &= dM_5 \end{aligned} \tag{13-26}$$

由此可以写成相应系数的表示式

$$a = -\frac{M_1}{M_5} \quad b = \frac{M_2}{M_5} \quad c = -\frac{M_3}{M_5} \quad d = \frac{M_4}{M_5} \tag{13-27}$$

在二维情况，这5个点投影到一组三维向量（也使用齐次坐标），得

$$x_5 = ax_1 + bx_2 + cx_3 + dx_4 \tag{13-28}$$

可以用舍弃其中两个索引的方式构造3×3矩阵，并用被舍弃的索引号表示

$$m_{12} = |x_3 \ x_4 \ x_5| \tag{13-29}$$

319

此时，可以将x从表示式中去掉，只保留相应的索引号，使得表示式简化，因此M_{12}又可以写成

$$m_{12} = |3 \quad 4 \quad 5| \tag{13-30}$$

与前面相似，可以通过代数运算将行列式与系数联系起来，例如

$$\begin{aligned} m_{12} &= a|3 \ 4 \ 1| + b|3 \ 4 \ 2| \\ &= a|1 \ 3 \ 4| + b|2 \ 3 \ 4| \\ &= am_{25} + bm_{15} \end{aligned} \tag{13-31}$$

以及

$$\begin{aligned} m_{13} &= am_{35} - cm_{15} \\ m_{14} &= am_{45} + dm_{15} \end{aligned} \tag{13-32}$$

因此我们得到了用M_i表示系数的形式，将这些关系写到方程式中，就得到

$$\begin{aligned} M_5m_{12} + M_1m_{25} - M_2m_{15} &= 0 \\ M_5m_{13} + M_1m_{35} - M_3m_{15} &= 0 \\ M_5m_{14} + M_1m_{45} - M_4m_{15} &= 0 \end{aligned} \tag{13-33}$$

这些关系对三维与二维运动都是不变的，但是一个乘法因子会以同样的方式影响所有的M_i。由于使用比例可以消除这种影响，因此定义三维不变量为

$$I_1 = \frac{M_1}{M_5} \qquad I_2 = \frac{M_2}{M_5} \qquad I_3 = \frac{M_3}{M_5} \tag{13-34}$$

以及二维不变量为

$$i_{12} = \frac{m_{12}}{m_{15}} \quad i_{13} = \frac{m_{13}}{m_{15}} \quad i_{25} = \frac{m_{25}}{m_{15}} \quad i_{35} = \frac{m_{35}}{m_{15}} \tag{13-35}$$

这些分母都不为零，因为它们是矩阵的行列式，而我们知道这些矩阵是非奇异的。观察式（13-33）并将其第一行除以M_5:

$$\frac{M_5}{M_5}m_{12} + \frac{M_1}{M_5}m_{25} - \frac{M_2}{M_5}m_{15} = 0 \tag{13-36}$$

它又被简化为

$$m_{12} + I_1 m_{25} - I_2 m_{15} = 0 \tag{13-37}$$

同样又可除以m_{15}得到两个独立的方程式

$$\begin{aligned} i_{12} + I_1 i_{25} - I_2 &= 0 \\ i_{13} + I_1 i_{35} - I_3 &= 0 \end{aligned} \tag{13-38}$$

320 因此如果我们有二维不变量，我们对三维不变量就有两个方程式。不幸的是，式（13-38）这两个方程并不确定这三维不变量。这两个方程确定的仍然是三维空间，I，的一条空间直线。

如何使用这种做法呢？对一个给定物体的三维模型，以及5个点，其中4个是非共面的，我们可以得到I_1、I_2与I_3，它表示在三维空间中的一个点。为了执行识别，首先从二维图像中抽取（若干个）5个一组的特征点组，并用它们构造二维不变量。每5个点给出在I_1、I_2与I_3空间的两个方程式，也就是在三维不变空间中一条直线。如果一个在二维图像中的5点组合是三维中同一个5点组合的投影，则所得到的直线会穿过代表该模型的一个点。如果有这些点的其他投影，我们可以得到一条不同的直线，但它仍然穿过这个模型点。

将该方法在实际图像中实现，会比上述情况略为复杂一些，因为必须考虑射影几何而不能假设是正交投影。另一个使问题复杂化的因素是要确定选择5点组合的合适方式，并且要有一种方法对付线“几乎”（即不能准确）穿过这个点的情况。Weiss 和Ray[13.52]讨论过这些问题。

13A.5 结论

13A.5.1 用哪个模型

到目前为止，我们已经讨论了相当多种类的物体表示方法，当然这不是在文献能找到的表示方法的全部。其他模型中包括变形模型的变种[13.10,13.11]，尤其是在距离图像中使用的模型[13.21]。

请考虑图13-13所示的情况：应该将该图形与一个圆匹配呢，还是一个六面体？显然对这个问题没有一个简单的答案。如果你有与问题有关的先验知识，譬如：你经常讨论的是圆形物体，你可能会选择圆形模型，它显然比多面体要简单些。最小描述长度（minimum description length ,MDL）的想法对此会有帮助。MDL策略宣称对一给定图像的最佳表示，可以由表示式的编码及残差的综合长度最小化确定。有趣的是可以证明[13.9，13.30]一个MAP表示式（最大后验概率）与MDL表示式是等价的，此时先验知识确实表示了（观察到的）信号。

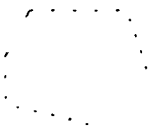


图13-13 由一个边缘检测器得到的一组点，它可能来自一个圆或一个多面体

Schweitzer[13.43]使用MDL策略开发计算光流的算法，Lanterman[13.29]用它表征ATR应用中的红外图像——“如果有几种描述与所观察的现象兼容，我们选择消耗最少的”[13.29]。

Rissanen [13.41]建议物体与模型匹配的质量可以表示为

$$L(x,\theta) = -\log_2 P(x \mid \theta) + L(\theta) \tag{13-39}$$

其中x是所观察的物体，θ是用参数向量表示的模型，P（$x|\theta$）是给定模型条件下能得到这些度量的条件概率，L（θ)表示这个模型所需的比特数，则条件概率的对数是数据如何匹配模型

的一种度量。我们因此可以在使用更复杂模型的更精确匹配与使用较简单模型的稍差匹配之间加以折衷[13.6]。 321

毕竟机器视觉是不能只靠一个程序、一个算法或一组数学概念解决问题的。归根结底解决问题的关键取决于构造系统时，能否把一堆专门技术集中在一起的能力。现在仍然无法判断如何做到这一点，遗憾的是只有很少的文章已承担起这个艰巨任务。例如Grosso与Tistarelli [13.18]将立体视觉与运动相结合。Bilbro 与Snyder[13.4]融合照明与距离以改善距离成像的质量，Pankanti与Jain[13.38]将立体视觉、影调与松弛标号融合。Zhu 与Yuille [8.80]将MDL方法中与主动轮廓（active contour），与区域增长（region growing）等技术融合在一起统一地处理图像分割问题。Gong 与Kulikowski [13. 16]使用一个规划策略，主要用在医学应用领域。

13A.5.2 匹配中的一致性与优化

在13A.1与13A.2节所讨论的问题中，第一步都是检测一些相对特殊的特征点。然后算法利用这些点之间的关系，依靠一致性来发现最优匹配。

在讨论递归神经元网络时，显示了这种网络能达到一种稳态，它事实上是执行式（13-15）目标函数的最小化。

13A.6 术语

特征向量	eigenvector
前馈神经网络	feedforward neural net
几何不变量	geometric invariant
图像索引	image indexing
递归神经网络	recurrent neural net

参考文献

[13.1] Y. Amit and A. Kong, "Graphical Templates for Model Registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), 1996.

[13.2] K. Astrom, "Fundamental Limitations on Projective Invariants of Planar Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[13.3] B. Bhanu and O. Faugeras, "Shape Matching of Two Dimensional Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(2), 1984.

[13.4] G. Bilbro and W. Snyder, "Fusion of Range and Luminance Data," *IEEE Symposium on Intelligent Control*, Arlington, August, 1988.

[13.5] A. Bimbo and P. Pala, "Visual Image Retrieval by Elastic Matching of User Sketches," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(2), 1997.

[13.6] J. Canning, "A Minimum Description Length Model for Recognizing Objects with Variable Appearances (The VAPOR Model)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), 1994.

[13.7] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric Phase-only Matched Filtering of Fourier–Mellin Transforms for Image Reconstruction and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(12), 1994.

[13.8] T. Chen and W. Lin, "A Neural Network Approach to CSG-based 3-D Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 322

1994.

[13.9] T. Darrell and A. Pentland, "Cooperative Robust Estimation Using Layers of Support," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), 1995.

[13.10] D. DeCarlo and D. Metaxas, "Blended Deformable Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[13.11] S. Dickinson, D. Metaxas, and A. Pentland, "The Role of Model-based Segmentation in the Recovery of Volumetric Parts From Range Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), 1997.

[13.12] M. Dubuisson Jolly, S. Lakshmanan, and A. Jain, "Vehicle Segmentation and Classification using Deformable Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), 1996.

[13.13] M. Fischler and R. Elschlager, "The Representation and Matching of Pictorial Structures," *IEEE Transactions on Computers*, **22**(1), 1973.

[13.14] S. Gold and A. Rangarajan, "A Graduated Assignment Algorithm for Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[13.15] R. Golden, *Mathematical Methods for Neural Network Analysis and Design*, Cambridge, MA, MIT Press, 1996.

[13.16] L. Gong and C. Kulikowski, "Composition of Image Analysis Processes Through Object-centered Hierarchical Planning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(10), 1995.

[13.17] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, and N. Otsu, "Face Recognition System Using Local Autocorrelation and Multiscale Integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(10), 1996.

[13.18] E. Grosso and M. Tistarelli, "Active/Dynamic Stereo Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(9), 1995.

[13.19] G. Gurevich, *Foundations of the Theory of Algebraic Invariants*, Transl. Raddock and Spencer, Groningen, The Netherlands, Nordcliff Ltd, 1964.

[13.20] S. Haykin, *Neural Networks, A Comprehensive Foundation*, Englewood Cliff, NJ, Prentice-Hall, 1999.

[13.21] M. Hebert, K. Ikeuchi, and H. Delingette, "Spherical Representation for Recognition of Free-form Surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(7), 1995.

[13.22] D. Heisterkamp and P. Bhattachaya, "Matching of 3D Polygonal Arcs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), 1997.

[13.23] J. Hopfield, "Neural Networks and Physical System with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Science*, 79, pp. 2554–2558, 1982.

[13.24] B. Hussain and M. Kabuka, "A Novel Feature Recognition Neural Network and its Application to Character Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 1994.

[13.25] D. Jacobs, "The Space Requirements of Indexing Under Perspective Projections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), 1996.

[13.26] A. Jain, Y. Zhong, and S. Lakshmanan, "Object Matching Using Deformable Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), 1996.

[13.27] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive

Window: Theory and Experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

[13.28] J. Kim, S. Yoon, and K. Sohn, "A Robust Boundary-based Object Recognition in Occlusion Environment by Hybrid Hopfield Neural Networks," *Pattern Recognition*, **29**(12), 1996.

[13.29] A. Lanterman, "Minimum Description Length Understanding of Infrared Scenes," *SPIE Automatic Target Recognition VIII*, **3371**, April 1998.

[13.30] Y. Leclerc, "Constructing Simple Stable Descriptions for Image Partitioning," *International Journal of Computer Vision*, **3**, pp. 73–102, 1989.

[13.31] M. Lew, T. Huang, and K. Wong, "Learning and Feature Selection in Stereo Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

[13.32] S. Li, K. Chan, and C. Wang, "Performance Evaluation of the Nearest Feature Line Method in Image Classification and Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11), 2000.

[13.33] T. Mitchell, *Machine Learning*, New York, McGraw-Hill, 1997.

[13.34] H. Murakami and B. Kumar, "Efficient Calculation of Primary Images from a Set of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **4**(5), 1982.

[13.35] H. Murase and M. Lindenbaum, "Partial Eigenvalue Decomposition of Large Images Using the Spatial Temporal Adaptive Method," *IEEE Transactions on Image Processing*, **4**(5), 1995.

[13.36] C. Olson, "Maximum Likelihood Template Matching," Completing this citation is a homework assignment (Assignment 13.3).

[13.37] D. Paglieroni, G. Ford, and E. Tsujimoto, "The Position-orientation Masking Approach to Parametric Search for Template Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

[13.38] S. Pankanti and A. Jain, "Integrating Vision Modules: Stereo, Shading, Grouping, and Line Labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(9), 1995.

[13.39] B. Parsi, A. Margalit, and A. Rosenfeld, "Matching General Polygonal Arcs," *Computer Vision, Graphics, and Image Processing. Image Understanding*, **53**(2), pp. 227–234, March, 1991.

[13.40] K. Rao and J. Ben-Arie, "Optimal Edge Detection Using Expansion Matching and Restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(12), 1994.

[13.41] J. Rissanen, "A Universal Prior for Integers and Estimation by Minimum Description Length," *Annals of Statistics*, **11**(2), pp. 416–431, 1983.

[13.42] J. Schwartz and M. Sharir, "Identification of Partially Obscured Objects in Two and Three Dimensions by Matching Noisy Characteristic Curves," *International Journal of Robotics Research*, **6**(2), 1987.

[13.43] H. Schweitzer, "Occam Algorithms for Computing Visual Motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[13.44] S. Sclaroff and A. Pentland, "Model Matching for Correspondence and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(6), 1995.

[13.45] K. Sengupta and K. Boyer, "Organizing Large Structural Modelbases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(4), 1995.

[13.46] L. Shapiro and J. M. Brady, "Feature-based Correspondence: an Eigenvector

Approach," *Image and Vision Computing*, **10**(5), 1992.

[13.47] X. Shen and P. Palmer, "Uncertainty Propagation and Matching of Junctions as Feature Groupings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(12), 2000.

[13.48] A. Smeulders, M. Worring, S. Santini, G. Gupta, and R. Jain, "Content-based Image Retrieval at the End of the Early Years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(12), 2000.

[13.49] D. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(8), 1996.

[13.50] M. Turk, and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, **3**(1), pp. 71–86, 1991.

[13.51] X. Wang and H. Qi, "Face Recognition Using Optimal Non-orthogonal Wavelet Basis Evaluated by Information Complexity," *International Conference on Pattern Recognition*, vol. 1, pp. 164–167, Quebec, Canada, August, 2002.

[13.52] I. Weiss and M. Ray, "Model-based Recognition of 3D Objects from Single Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(2), 2001.

[13.53] M. Yang and J. Lee, "Object Identification from Multiple Images Based on Point Matching under a General Transformation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

323 ~ 325

[13.54] S. Yoon, *A New Multiresolution Approximation Approach to Object Recognition*, Ph.D. Thesis, North Carolina State University, 1995.

第14章 统计模式识别

Statistics are used much like a drunk uses a lamppost; for support, not illumination.

Vin. Scully

单讲统计模式识别这一学科本身，可能就需要好几本书（实际情况也的确如此）。因此，这一章的内容没有涵盖统计模式识别问题的各项细节。但是，学习机器视觉需要了解统计模式识别的相关知识，这样才能便于阅读有关文献，并方便理解其他机器视觉问题。所以，本章主要对统计方法进行简单介绍。如果要对机器视觉进行深入地研究，只靠这一章的讲解是远远不够的，应该完整地学习一门统计模式识别课程。关于这方面的教材，推荐如下几部：Duda和Hart[14.3]的原版教材，包含统计模式识别分类和机器视觉两方面的内容，但是在新版[14.4]中主要限于分类方面。更早的还有Fukanaga的教科书[14.6]，其中仍然包含大量的有用信息，我们还推荐[14.11]，因为它可读性强。

14.1 分类器设计

回顾13.2节所讲的例子。在那个例子中，通过计算斧子或者手斧的样本均值，利用统计的方法推导出它们的模型。我们称这些集合为“训练集”。在13.2节中，把用特征向量表示的未知对象与这两个模型进行比较，并将未知对象归类为与它非常相似的模型类别，其中“非常相似”是指未知对象的特征向量与该模型之间的欧几里德距离非常小。本章将说明，这种“最近均值”决策规则实际上是类别分布为高斯概率密度时最大似然的简化情况。本章将进一步说明，利用其他决策规则而不是最近均值，分类效果会更好，而且计算效率也更高。

326

下面我们开始讨论几种决策规则。

14.1.1 决策规则的选择

从第1章图1-2可以看到，对图像进行度量，而度量过程的输出是图像特征。这些特征是描述图像特性的一组数，在分类中要用到这些特征。

分类器设计包括寻找一种实现决策规则的方法。确定决策规则的*形式*（form）和描述规则特征的*参数*（parameter）的方法有若干种选择，下面简单介绍这些决策规则。

线性决策规则

图14-1显示两种不同工业零件经过很多次度量后得出来的结果。面积特征和长度特征的度量结果在图上用符号来表示（这种类型的图称为“散点图”）。“x”点代表一类，表示法兰；“o”点代表另一类，表示垫圈。另外，图上还画出一条*线性决策分界*（decision boundary）。线性决策规则是：如果度量结果落在决策分界的左边，就确定该未知零件是法兰，否则就是垫圈。线性决策规则比较吸引人，因为用*线性*

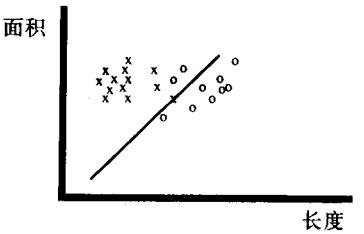


图14-1 线性决策分界

机（linear machine）就可以实现。这些线性机可能有很大的并行度，因此运行速度很快。从图中可以看出，这两类不是线性可分（linearly separable）的，也就是说，不存在一条直线能够把这两类完全分开。最好的分界线就是线性分类器设计过程的输出结果。在14A.2节中，我们介绍支持向量机，它是线性机的一种新类型。

14.1.2 最大似然分类器

6.2节介绍图像存储时，已经接触过最大似然（maximum likelihood）这一术语。在此将说明该类数学知识在分类器设计中的应用。

327

用最大似然方法进行分类器设计时，用统计表示描述具有某个度量值集合的样本属于某个类别的概率。也就是说，已知被测样本的某个度量值（向量）$\boldsymbol{x}$，要估计它属于类别w_i的条件概率$P(w_i|\boldsymbol{x})$。对每个w_i都计算出这样的概率，于是决策规则就是：如果对于所有的j来说，都有$P(w_i|\boldsymbol{x})> P(w_j|\boldsymbol{x})$，那么该样本所属的类别就是$i$。

必须注意，术语“最大似然”不一定意味着是最佳性能。典型的最大似然算法需要假设概率密度函数的形式，该假设也许对于同一类的许多数据点来说是无效的。设计最大似然分类器时，不仅需要为每个类别选择概率密度函数的形式（比如高斯密度函数），而且也要选择密度函数的参数（例如均值和方差）。

14.1.3 分类器学习算法

训练集是从每一类中选择出来的样本集，用它们统计地描述这些类别。所使用的确定模式分类器参数的方法，与我们手头上是否有这样的训练集有关。

监督学习

如果每个类别都有相应的训练集，并且可以从这些训练集中推导出类别的统计特征表示，那么这个过程便称为“监督学习”。监督（supervised）意味着训练集中每个数据点都根据所属类别进行了独立标记。每类的统计特征可以是它的均值、方差或其他统计度量，也可用其他参数来表示。图14-1显示监督采样的数据分布情况，其中“x”点属于一类，而“o”点属于另一类。前面13.2节中区分斧子和手斧的例子，就是一个监督学习问题，因为前提条件是已经得到两个类别的训练集。

无监督学习

图14-2表示无监督学习过程可能出现的结果。在这幅图中，没有对样本进行类别标记，但从图中可以很明显地看到两个不同的聚类（cluster）。举个例子，在一块麦地的卫星多谱图像上，如果种植着两种不同类型的小麦，统计时就会出现两个聚类。从这个例子看，聚类过程似乎非常简单，只需看一眼，就会看到两个明显的聚类。但是一定要记住，并不是所有的度量向量都是二维的，对于多维数据，人们处理起来就非常困难。而无监督学习或者聚类算法能够自动识别出这两个（或多个）聚类，并把样本点与（在某种意义上）最接近的聚类联系起来。

328

随着讨论的深入，你会学会如何确定为了实现分类算法所需的度量值。聚类算法在其他机器视觉算法中也同样有用，例如确定Hough变换累加数组中的峰值位置。这在第15章中会进行更详细的讨论。

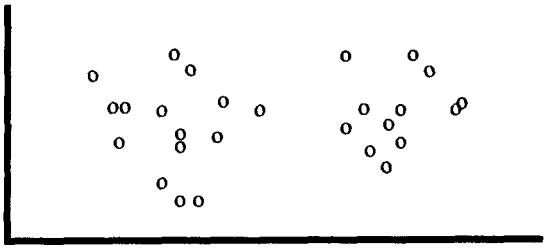


图14-2 无监督学习，只有一个样本度量集，但这些数据呈现出自然聚类现象

14.2 贝叶斯规则和最大似然分类器

本节根据训练集统计特性的有关假设进行分类器设计。如果这些假设对于所用训练集来说是有效的，那么分类器的分类性能将是最佳的。我们也会研究分类器的性能，包括错误率等。

细心的读者会注意到，这里对统计学概念的描述和第6章中所讲的非常类似。这种类似不仅正确而且也是预先安排的。在第6章中，用统计方法寻找使某种特性最小化的一幅图像。而在本章中，寻找一种决策，决定一个样本到底属于哪个类别。

14.2.1 贝叶斯规则

定义$P(w_i)$为类别w_i发生的*先验概率*（priori probability），也就是说，在没有进行任何度量之前w_i发生的概率。举例来说，有一个制造法兰和垫圈的工厂，生产的法兰和垫圈的比例是9：1。法兰和垫圈以随机的顺序从流水线上下来。根据工厂生产的法兰数量是垫圈的九倍这一先验知识，我们知道无论何时看流水线，看到法兰的几率都要远远大于看到垫圈的几率。因此法兰的先验概率是0.9，而垫圈的先验概率是0.1。 329

用$p(x|w_i)$表示属于类别w_i的样本x出现的*条件概率密度*。对于某个类别w_i，$p(x|w_i)$应被看成是x的函数。假设有一个生产斧子和手斧的工厂，我们想找出如图14-3所示的斧子和手斧长度的概率密度。从图中可以看到，斧的可能长度是30英寸，而手斧的可能长度是12英寸。当然，每个斧子、手斧的长度有差异。

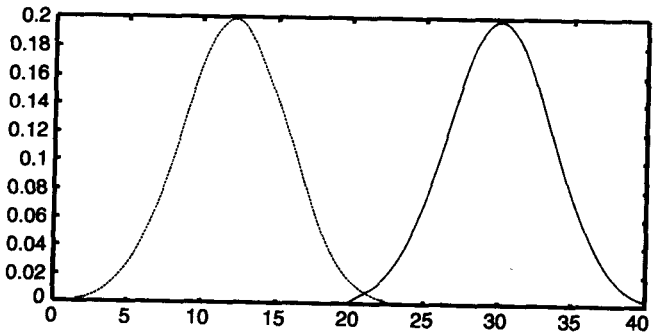


图14-3 对两类对象的相同度量会产生不同的均值。但由于存在度量噪声或者实际偏差，这些度量值会发生重叠

概率密度函数可以用好几种方法表示。一种方法是，对于变量的所有可能取值，统计每个值发生的次数，制成简单的表格，上例中的变量就是长度。这种表格被称作变量的*直方图*

(histogram)。经过适当的规范化处理，直方图就可用来表示概率密度函数，当然要求变量只有有限种可能的取值。也可以用参数化解析函数（如高斯函数）表示概率密度。

最后，$P(w_i|\boldsymbol{x})$ 定义为后验条件概率，它表示已知度量$\boldsymbol{x}$的条件下，被测样本属于w_i 的条件概率。$P(w_i|\boldsymbol{x})$就是我们所要的，我们用它作为决策规则，或者更确切地说，作为判别函数(discrimination function)。决策规则描述如下：对于未知样本的度量$\boldsymbol{x}$，计算每个类别的$P(w_i|\boldsymbol{x})$。也就是对每一个可能的i值，计算其$P(w_i|\boldsymbol{x})$。如果对于所有的$j \neq i$，$P(w_i|\boldsymbol{x})$都比$P(w_j|\boldsymbol{x})$大，那么该未知样本应归于i类。基于$P(w_i|\boldsymbol{x})$进行分类决策时，用到了最大似然分类器。

330

根据贝叶斯规则，所定义的这三个函数的关系为：

$$P(w_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|w_i)P(w_i)}{Something} \tag{14-1}$$

$$Something = p(\boldsymbol{x}) = \sum_{j=1}^{c} p(\boldsymbol{x}|w_j)P(w_j) \tag{14-2}$$

在式（14-1）中，用“*something*”表示条件概率密度中的分母。用“*something*”主要是提醒读者注意这样的事实：这个数值表示该$\boldsymbol{x}$值发生的概率密度与观测类别无关。既然这个数值与类别无关，那么它对于所有类别都是相同的，所以在判定类别时它不能提供任何帮助。它只是一个规范化常量，用来保证数值$P(w_i|\boldsymbol{x})$具有希望的概率特性。$P(w_i|\boldsymbol{x})$的值应介于0和1之间，并且对所有类别$P(w_i|\boldsymbol{x})$之和为1（被观测样本至少属于所考虑类别中的某一类）。

从某种意义上说，式（14-1）解决了模式识别问题。它告诉我们在已知等式右边每一项的前提下如何去做决策。下一节介绍如何确定式中的每一项。

14.2.2 参数模式分类器

上一节内容表明，如果具备以下两方面的知识就可以利用最大似然分类器进行决策：所考虑类别的先验概率，以及每个类别的条件概率密度。到此为止，我们还没有讨论采用哪种形式的概率密度函数。在图14-3中，密度函数采用大家熟悉的“钟形曲线”，这种形状的曲线可用数据表或者解析函数来描述，对于标量值x，钟形曲线解析函数具有如下形式：

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{14-3}$$

其中$p(x)$是众所周知的一元高斯（规范化）密度函数，“一元”表示x是一个标量，并且是一个单变量。均值μ和标准差σ是参数向量中的元素。

$$\theta_i = \begin{bmatrix} \mu_i \\ \sigma_i \end{bmatrix} \tag{14-4}$$

这两个数可以完全确定高斯密度函数下属于类别i的变量x的条件概率密度。

331

如果度量不是单变量，而是有多个，那么$\boldsymbol{x}$就是一个向量，并且$p(\boldsymbol{x}|w_i)$的形式为多元高斯密度函数。

$$p(\boldsymbol{x}|w_i) = \frac{1}{(2\pi)^{d/2}|K_i|^{1/2}}\exp[-(1/2)(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}K_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)] \tag{14-5}$$

在式（14-5）中，d是向量$\boldsymbol{x}$的维数，$\boldsymbol{\mu}_i$ 是表示随机度量向量平均值的均值向量，K_i 是$d \times d$的协方差矩阵。

在一元高斯密度函数情况下，用两个数值来表示度量值的条件概率密度。在多元条件下，用d维向量和$d \times d$的矩阵来表示。有了这些参数，把它们代入式（14-5）中，然后再代入式（14-1）中，便可以算出未知目标最有可能属于哪一类。不幸的是，在多数情况下，均值和协方差并没有给定，需要根据训练集估计。

对式（14-5）中的右边取对数，就可以去掉指数。因为对数函数是单调的，就可以推导出一种参数表示，包括均值向量$\boldsymbol{x}$，以及类别i的统计值μ_i 和k_i。求该表示的极大值，与求原始概率的极大值等价。这时，分类问题就变得非常简单：只是把$\boldsymbol{x}$代入具有不同均值和协方差的每个公式中。如果有c个类别，就能得到c个不同的函数，这些函数称为判别函数。所有的判别函数都具有相同的形式，但有着不一样的参数。最大判别函数所对应的类别就确定为$\boldsymbol{x}$所属的类别。

14.2.3 概率密度估计

因为随机变量的实际分布密度经常具有高斯（正态）函数的形式，而且高斯函数使用起来也非常方便，所以我们常常用高斯函数描述分布密度，对它特殊对待。

要用监督学习方法设计模式分类器，需要先根据样本训练集估计密度参数。参数集用向量θ表示。

一元高斯情况

假设用X_i 表示n_i 个元素的集合，其中的每个元素表示对属于类别i的一个样本的某项度量值。由于在此只考虑一元情况，X_i 的元素是标量。下一小节中将讨论向量值元素，以及多元密度情况。

假设X_i 中的样本，除了类别i之外，没有给出任何有关其他类别的参数的信息。这样就可以对每个类别进行单独分析。

332

假定样本是相互独立的，整个集合X_i 产生的概率由下式确定：

$$p(X_i) = \prod_{k=1}^{n_i} p(x_{ik}) \tag{14-6}$$

由于这个概率本质上依赖于参数θ_i，为了更清楚地表示这种依赖关系，把式（14-6）重写如下：

$$p(X_i \mid \theta_i) = \prod_{k=1}^{n_i} p(x_{ik} \mid \theta_i) \tag{14-7}$$

从而θ_i 的最大似然估计定义为：使$p(X_i \mid \theta_i)$取最大值的θ_i 值。式（14-7）描述了任何特定训练集发生的可能性，其中概率分布通过参数向量θ_i 来描述。因为采用的是高斯密度函数，将式（14-7）改写为：

$$p(X_i \mid \theta_i) = \prod_{k=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_{ik}-\mu_i)^2}{2\sigma_i^2}\right) \tag{14-8}$$

可以看到一个重要的现象：使$p(X_i \mid \theta_i)$最大化的θ_i 值同样也使得$\ln[p(X_i \mid \theta_i)]$达到最大值。这一点毫无疑问，因为自然对数函数本身是单调上升函数。因此可以选择寻找使概率密度或者使其对数最大化的参数向量θ_i 。可以看出，对数形式用起来非常方便。对等式右边取对数可得如下形式：

$$\ln(p(X_i \mid \theta_i)) = \sum_{k=1}^{n_i} \ln\left(\frac{1}{\sqrt{2\pi}\sigma_i}\right) - \sum_{k=1}^{n_i} \frac{1}{2}\left(\frac{x_{ik}-\mu_i}{\sigma_i}\right)^2 \tag{14-9}$$

14.2.4 均值估计

用 $\hat{\mu}$ 表示μ的估计值。为了找出使式（14-9）为最大值的μ值，式两边对μ求导并让结果等于0。

$$\sum_{k=1}^{n_i} \frac{x_{ik}-\hat{\mu}_i}{\sigma_i} = 0 \tag{14-10}$$

上式可以简化为:

$$\sum_{k=1}^{n_i} x_{ik} - \sum_{k=1}^{n_i} \hat{\mu}_i = 0 \tag{14-11}$$

$$\hat{\mu}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} x_{ik} \tag{14-12}$$

333 $\hat{\mu}_i$ 称为样本均值，并且样本均值等于平均值，这也符合直观情况。

14.2.5 方差估计

推导过程与前面一样，但是这次把均值和方差作为未知量。在下文中，为了简化符号表示，不再写表示类别的下标i。先考虑一元情况，然后再考虑二维参数向量θ，

$$\theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix} \tag{14-13}$$

简化式（14-9），则概率的对数形式为:

$$L = -\sum \ln(\sqrt{2\pi}) - \sum \ln\sigma - \frac{1}{2}\sum\left(\frac{x_k-\mu}{\sigma}\right)^2 \tag{14-14}$$

利用$\partial L/\partial\mu=0$，可得

$$\sum_{k=1}^{n} \frac{1}{\sigma^2}(x_k - \hat{\mu}) = 0 \tag{14-15}$$

利用$\partial L/\partial\sigma=0$，可得

$$-\sum_{k=1}^{n} \frac{1}{\hat{\sigma}} + \sum_{k=1}^{n} \frac{(x_k-\hat{\mu})^2}{\hat{\sigma}^3} = 0 \tag{14-16}$$

式（14-15）两边同乘以 $\hat{\sigma}^2$，则可简化为:

$$\sum_{k=1}^{n} x_k = \hat{\mu}\sum_{k=1}^{n} 1 \tag{14-17}$$

以及和前面一样的结果:

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{14-18}$$

同样，对式（14-16）进行简化，可得:

$$\frac{n}{\hat{\sigma}}=\frac{1}{\hat{\sigma}^3}\sum_{k=1}^{n}(x_k-\hat{\mu})^2 \tag{14-19}$$

从而得出

$$\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\hat{\mu})^2 \tag{14-20}$$

可以看出，正态分布密度函数的最佳参数估计就是我们熟悉的样本均值和样本方差。 334

多元高斯密度参数

对于多元规范高斯密度函数的参数向量，推导过程与前一小节类似，可以得到这些参数的最大似然估计为:

$$\hat{\mu}_i=\frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{x}_{ik} \tag{14-21}$$

$$K_i=\frac{1}{n_i}\sum_{k=1}^{n_i}(\boldsymbol{x}_{ik}-\hat{\mu}_i)(\boldsymbol{x}_{ik}-\hat{\mu}_i)^{\mathrm{T}} \tag{14-22}$$

对于多元情况得到的结果必然与一元情况类似，即高斯分布的均值和方差，在最大似然意义上的最佳估计就是样本均值和样本（协）方差。

现在看看在这一章中我们要记住些什么？为了求出一组参数的最大似然估计，需要已知一个训练集，假设集合中的元素是相互独立的（如果可以的话），并且把整个训练集发生的概率写成乘积。两边取对数并求导，然后让它等于0，就生成联立方程。求解这些方程后，就得到了参数的最佳估计。这个方法不光是对高斯分布，对其他任意的概率分布都是可行的。但在某些情况下，联立方程的求解过程可能会十分复杂。

最后，除了最大似然方法外，还有另外一些方法也可以求得参数。由于篇幅和时间的限制，在此我们就不讨论那么多了。

14.2.6 似然比

我们希望做出使正确可能性最大的决策。为此，选择使后验概率达到最大的类别，这个概率也就是在已知样本度量值的情况下做出正确决策的概率。

也就是说，选择i使得$P(w_i|\boldsymbol{x})$达到最大。为了做到这一点，回想贝叶斯规则，对于类别1，我们有:

$$P(w_1|\boldsymbol{x})=\frac{p(\boldsymbol{x}|w_1)P(w_1)}{p(\boldsymbol{x})} \tag{14-23}$$

如上所述，无论$\boldsymbol{x}$属于类别1还是类别2，$p(\boldsymbol{x})$都是一样的。由于分母不受分类决策的影响，进行决策时就可以将其忽略。

在只有两类的情况下，如果$P(w_1|\boldsymbol{x})>P(w_2|\boldsymbol{x})$，就选择类别1。代入贝叶斯规则，如果下式成立就选择类别1。

$$p(\boldsymbol{x}|w_1)P(w_1)>p(\boldsymbol{x}|w_2)P(w_2) \tag{14-24}$$ 335

也就是

$$\frac{p(x|w_1)}{p(x|w_2)} > \frac{P(w_2)}{P(w_1)} \tag{14-25}$$

公式左边的表达式就称为似然比（likelihood ratio）。式（14-25）中包含的关系，反映了似然比和先验概率之间的权衡关系。如果它是错误的，我们就选择类别2。注意这种形式的决策，是根据使正确决策的概率最大化的原则，利用了类别的度量知识和先验概率知识推导得到的。我们也可以用其他指标，例如，可以选择用条件风险最小化替代概率最大化。

14.3 决策区间和错误概率

任何决策规则的结果都是将特征空间分成c个决策空间$\Omega_1, \Omega_2, \cdots, \Omega_c$。假设我们定义了一组判别函数$g_i(x)$，如果对所有的$j \neq i$，都有$g_i(x) > g_j(x)$，那么$x \in \Omega_i$并且选中类别$w_i$。当$\Omega_i$与$\Omega_j$相邻时，决策分界方程为$g_i(x) = g_j(x)$。在只有两类的情况下，可以算出决策区间上的错误概率为：

$$P(error) = P(x \in \Omega_2, w_1) + P(x \in \Omega_1, w_2) \tag{14-26}$$

也就是说，当x实际属于w_2但是落在区间Ω_1中，或者x实际属于w_1但是落在区间Ω_2中时，就会发生错误。由于这些事件不可能同时为真，则有

$$\begin{aligned} P(error) &= P(x \in \Omega_2 | w_1)P(w_1) + P(x \in \Omega_1 | w_2)P(w_2) \\ &= \int_{\Omega_2} p(x|w_1)P(w_1)\mathrm{d}x + \int_{\Omega_1} p(x|w_2)P(w_2)\mathrm{d}x \end{aligned} \tag{14-27}$$

我们也用$P(error|w_2)$表示w_2为真实状态时做出错误决策的概率。图14-4显示两类别的后验概率密度、决策分解的推导过程以及错误概率。

总的来说，如果$p(x|w_1)P(w_1) > p(x|w_2)P(w_2)$，就应该做出$x$属于区间$\Omega_1$的决策。这样就使较小的项对误差积分有贡献，这正好与贝叶斯决策规则相符合。

336

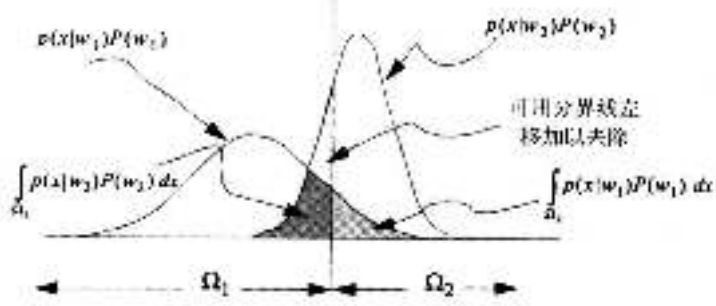


图14-4 两类别高斯形式的后验概率密度函数、决策分界以及错误概率

在多类别情况下，因为计算错误的项的项数要比正确的项多，所以计算正确的概率更加简单。

$$\begin{aligned} P(correct) &= \sum_{i=1}^{c} P(x \in \Omega_i, w_i) \\ &= \sum_{i=1}^{c} p(x \in \Omega_i | w_i)P(w_i) = \sum_{i=1}^{c} \int_{\Omega_i} p(x|w_i)P(w_i)\mathrm{d}x \end{aligned} \tag{14-28}$$

不管特征空间如何划分，这个结果都是有效的。贝叶斯分类器通过选择使积分最大的区间，而使这个概率最大化。

14.4 条件风险

设$W=(w_1, w_2,\cdots, w_c)$是可能类别的有限集合，且$A=\{\alpha_1, \alpha_2, \cdots, \alpha_a\}$是可能的决策集合。如果样本落在类别$w_i$中，就会发生决策$\alpha_i$。一般地，如果决策不正确就会带来一些损失。先假设我们有某种方法定量描述这些损失。也就是说，可以用C_{ij}表示当样本实际属于类别w_j但决策为α_i所带来的损失。

要继续这一节的学习，需要复习一下关于随机变量的期望值的定义。你应该已经知道这个定义，只需要想一想。假设随机变量x可以有四个不同的取值0、1、2和3。如果我们知道这些取值的概率，假设它们是

$$p(x=0)=0,\ p(x=1)=0,\ p(x=2)=1.0,\ p(x=3)=0$$

你希望观察到的x值是什么呢？很明显对吧？当然总是看到$x=2$。现在看另一种稍复杂的情况， 337

$$p(x=0)=0,\ p(x=1)=0,\ p(x=2)=0.5,\ p(x=3)=0.5$$

x的期望值是2.5，你同意吗？也就是说，在一半的时间里$x=2$，而在另一半时间里$x=3$。

现在来总结期望值的概念。假设随机变量有很多可能的取值，每个取值都对应着一个概率。如果可能的取值数目是有限的，那么期望值具有如下求和的形式。

$$<x>=\sum xP(x) \tag{14-29}$$

更一般的情况是，如果x是连续的，可以用密度函数来代替概率，然后将加法变成积分。

假设我们观测到$\boldsymbol{x}$并打算采取决策α_i。如果实际状态是w_j，将会产生损失C_{ij}。因为$P(w_j|\boldsymbol{x})$是w_j为真的概率，与α_i相关的期望损失是:

$$r_i=\sum_{j=1}^{c} C_{ij}P(w_j|\boldsymbol{x}) \tag{14-30}$$

期望损失称为风险（risk）。我们将r_i表示为$r(\alpha_i|\boldsymbol{x})$，这样可以更清楚的表示这是一个条件风险。

我们希望通过选择最佳α_i使得总风险达到最小，将总风险表示为r。函数$\alpha(\boldsymbol{x})$就是一个决策规则，它可以告诉我们采用什么决策使r最小化。

对于每一个$\boldsymbol{x}$来说，假设决策函数$\alpha(\boldsymbol{x})$的值分别为$\alpha_1, \alpha_2, \cdots, \alpha_a$。这样，总风险便与决策规则联系起来。

因为r_i或者$r(\alpha_i|\boldsymbol{x})$是条件风险，总风险为

$$r=\int r(\alpha(\boldsymbol{x})|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{14-31}$$

对于$\boldsymbol{x}$的每一个值，如果我们选择$\alpha(\boldsymbol{x})$使得$r(\alpha(\boldsymbol{x})|\boldsymbol{x})$尽可能的小，那么总风险就最小。为了使总风险$r$最小，计算最小的$r_i$及相应的决策$\alpha_i$。所得到的总风险就称为贝叶斯风险。

考虑风险的似然比

这里详细介绍两类情况的推导过程。如果你弄明白了两类情况，多类情况的推导过程就

变得简单了。当然，要记住，似然比只适应于两类情况。

338 对于两类情况，让我们更详细地看看前面最小风险准则的推导过程。

如果只有两类，就有四种概率：两种对应正确的决策，另外两种对应错误的决策。因此总风险是：

$$
\begin{aligned}
r = &\int_{\Omega_1} C_{11} P(w_1 \mid x)\mathrm{d}x \\
&+ \int_{\Omega_1} C_{12} P(w_2 \mid x)\mathrm{d} \\
&+ \int_{\Omega_2} C_{21} P(w_1 \mid x)\mathrm{d} \\
&+ \int_{\Omega_2} C_{22} P(w_2 \mid x)\mathrm{d}
\end{aligned}
\tag{14-32}
$$

上式也就是，x属于类别1，并处在决策规则判定x属于类别1的区间上，积分计算得到的风险，加上在决策规则认为应属类别1的x区间上，x属于类别2积分计算得到的风险，再加上在决策规则认为应属类别2的x区间上，x属于类别1积分计算得到的风险，再加上在决策规则认为应属类别2的x区间上，x属于类别2积分计算得到的风险。

对上式整理后得到：

$$
\begin{aligned}
r = &\int_{\Omega_1} [C_{11} P(w_1 \mid x) + C_{12} P(w_2 \mid x)\mathrm{d}x \\
&+ \int_{\Omega_2} [C_{21} P(w_1 \mid x) + C_{22} P(w_2 \mid x)]\mathrm{d}
\end{aligned}
\tag{14-33}
$$

由于任意密度函数，如q，在区间Ω_2上的积分，可以改写为1减去q在区间Ω_1上的积分，因此所有积分都可以在区间Ω_1上进行，于是得到：

$$
\begin{aligned}
r = &C_{11} \int_{\Omega_1} P(w_1 \mid x)\mathrm{d}x + C_{12} \int_{\Omega_1} P(w_2 \mid x)\mathrm{d}x \\
&+ 1 - C_{21} \int_{\Omega_1} P(w_1 \mid x)\mathrm{d}x + 1 - C_{22} \int_{\Omega_1} P(w_2 \mid x)\mathrm{d}x
\end{aligned}
\tag{14-34}
$$

整理后得出：

$$
r = 2 + \int_{\Omega_1} ((C_{11} - C_{21}) P(w_1 \mid x) + (C_{12} - C_{22}) P(w_2 \mid x)\mathrm{d}x \tag{14-35}
$$

我们的目的是让r最小（记住，这是全部四个决策所带来的风险），也就是说，决策规则确实决定了决策区间。对于这种情况，因为只有两个决策区间，任何时候不是类别1就是类别2，所以我们要做的就是确定区间Ω_1。为了做到这一点，首先需要做一个假设。假设做正确决策
339 代价总是比做错误决策的代价小，那么（$C_{11} - C_{21}$）<0，等等。

如何选择积分的上下限而使积分尽量小呢？（记住，确定积分上下限其实就是确定类别1的区间分界线。）简单的方法是，选择使被积函数处处为负的决策区间。这样做就产生了选择区间1所需的条件，即选择Ω_1使下式成立：

$$
(C_{11} - C_{21}) P(w_1 \mid x) + (C_{12} - C_{22}) P(w_2 \mid x) < 0 \tag{14-36}
$$

用条件密度和先验概率的乘积代替后验概率，可以得到:

$$(C_{11}-C_{21})p(x\mid w_1)P(w_1)<(C_{22}-C_{12})p(x\mid w_2)P(w_2) \tag{14-37}$$

经过适当的整理，可以形成如下决策规则。如果下式成立就选择类别1，否则就选择类别2。

$$\frac{p(x\mid w_1)}{p(x\mid w_2)}>\frac{(C_{12}-C_{22})P(w_2)}{(C_{21}-C_{11})P(w_1)} \tag{14-38}$$

上述表达式称为“似然比检验”。

考虑对称损失函数:

$$C_{ij}=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases} \tag{14-39}$$

这就是说，所有的错误代价都是一样的，并且如果决策正确的话就不付出代价。现在重写出条件风险，做出决策i的代价是:

$$r_i=\sum_{j=1}^{c}C_{ij}P(w_j\mid \boldsymbol{x})=\sum_{j\neq i}P(w_j\mid \boldsymbol{x})=1-P(w_i\mid \boldsymbol{x}) \tag{14-40}$$

这样，为了使平均错误概率最小，我们选择了使后验概率$P(w_i\mid \boldsymbol{x})$最大的$i$。也就是说，为了使代价最小，如果对于所有的$i\neq j$，都有$P(w_i\mid \boldsymbol{x})>P(w_j\mid \boldsymbol{x})$，则确定是$w_i$，在简单的最大似然分类器设计时我们已经看到了这一点。因此可以看出，最大似然分类器结合对称代价函数使贝叶斯风险最小。

14.5 二次函数分类器

考虑两类情况下一般的多元高斯分类器。像作业14.1那样取对数，可推出基于似然比的决策规则。如果下式成立，就选择类别1，否则选择类别2。 340

$$\ln\frac{|K_1|}{|K_2|}+(\boldsymbol{x}-\mu_1)^{\mathrm{T}}K_1^{-1}(\boldsymbol{x}-\mu_1)-(\boldsymbol{x}-\mu_2)^{\mathrm{T}}K_2^{-1}(\boldsymbol{x}-\mu_2)<Threshold \tag{14-41}$$

其中 $Threshold=-2\ln\left\{\frac{P(w_2)(C_{12}-C_{22})}{P(w_1)(C_{21}-C_{11})}\right\}$

如果定义

$$\begin{aligned}&A=K_1^{-1}-K_2^{-1},\boldsymbol{b}=2(K_2^{-1}\mu_2-K_1^{-1}\mu_1)\\&c=\mu_1^{\mathrm{T}}K_1^{-1}\mu_1-\mu_2^{\mathrm{T}}K_2^{-1}\mu_2+\ln\frac{|K_1|}{|K_2|}\end{aligned} \tag{14-42}$$

就可以把式（14-41）的左边用下式表示

$$g(\boldsymbol{x})\equiv\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}+c \tag{14-43}$$

则决策规则变成：如果$g(\boldsymbol{x})<T$，则选择类别1。从这个公式可以很清楚地看到，高斯参数分类器被称为二次函数分类器的原因。

让我们来看看这个决策规则的含义。对于测度$(\boldsymbol{x}-\mu_1)^{\mathrm{T}}K_1^{-1}(\boldsymbol{x}-\mu_1)$，其中涉及度量值$\boldsymbol{x}$和以均值向量和协方差矩阵为参数的一个类别。这个度量称为马哈拉诺比斯（Mahalanobis）距离的平方（平方是译者加的——译者注）。

首先看看协方差矩阵为单位阵的情况。这时马哈拉诺比斯距离的平方简化为$(\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}_1)$，即先求度量值和均值之差，这是一个向量，然后求该向量的内积，也就是该向量幅值的平方。这个结果表示的是什么呢？当然是度量值与该类均值的欧几里德距离的平方。如果先验概率是相同的，并且采用对称代价函数，则*Threshold*（阈值）的值为0，那么决策规则就简化为：如果下式成立，则选择类别1，否则选择类别2。

$$(\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}_1)-(\boldsymbol{x}-\boldsymbol{\mu}_2)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}_2)<0 \tag{14-44}$$

如果度量值更接近类别1的均值，那么上式左边小于 0 。因此我们称这种简单的分类器为最近均值分类器，或者是最近均值决策规则。

下面使这个规则变得稍微复杂一些。不再假设协方差矩阵是单位阵，而是假设它们是相等的（$K_1=K_2\equiv K$）。这时，式（14-42）中的A矩阵变成了0，由于运算式中不再含二次项，得到的是线性分类器。

我们可以选择忽略协方差矩阵行列式的比值这一项，或者更恰当一些，将该数值包含在阈值T中。于是得到最小距离决策规则，但这时的距离不是欧几里德距离。我们称这种分类器为最小马哈拉诺比斯距离分类器。

341

还有另外一种特殊情况。如果协方差矩阵不仅相等，而且是对角矩阵，会出现什么情况呢？这时马哈拉诺比斯距离具有特殊的形式。利用一个三维度量向量来说明这一点，设均值为0。

$$[x_1,x_2,x_3]\begin{bmatrix}\frac{1}{\sigma_{11}} & 0 & 0\\ 0 & \frac{1}{\sigma_{22}} & 0\\ 0 & 0 & \frac{1}{\sigma_{33}}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

对上式展开可得

$$\frac{x_1^2}{\sigma_{11}}+\frac{x_2^2}{\sigma_{33}}+\frac{x_3^2}{\sigma_{33}}$$

这是一个椭球体方程式，以原点为中心，轴线沿坐标轴方向（或者，在均值不为0的情况下，以均值为中心）。更一般的情况是，协方差矩阵不是对角阵，那么唯一的结果就是椭球体发生旋转。因此表示一个点到一个类别的马哈拉诺比斯距离的方程，得到一个椭球体。

下面是一种更有趣的情况。假设协方差矩阵是相等的，对角的，并且与单位阵成正比$K_i=\sigma^2 I$。那么类别i的判别函数具有下列形式：

$$g_i(\boldsymbol{x})=\frac{2\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{x}}{\sigma^2}-\frac{|\boldsymbol{\mu}_i|^2}{\sigma^2}+2\ln P(w_i) \tag{14-45}$$

进一步假设所有均值的幅值都是相等的。也就是说，所有均值都落在以原点为中心的超球体上。那么，就不再需要考虑公式（14-45）中的第二项，判别函数化简为：

$$g_i(\boldsymbol{x})=\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{x}=\sum_{k=1}^{d}\mu_{ik}x_k \tag{14-46}$$

我们称该式为内积分类器。

14.6 最小最大规则

有时候先验概率是未知的或变动的。这时，若按某一种假设的先验概率设计风险分类器，则固定的决策规则将不会使实际的风险最小，因此我们采用最小最大规则并且试图使最大可能的风险最小化。（由于原文这段话表达得不够准确，翻译时做了些必要的改动。——译者注）假设有c个类别，总期望风险为:

$$r=\sum_i P(w_i)\sum_j \int_{\Omega_i} C_{ij}p(\boldsymbol{x}\mid w_j)\mathrm{d}\boldsymbol{x} \tag{14-47}$$

342

也就是，某个自然状态的概率，乘以该自然状态的所有可能做出的决策结果，以及与这些决策结果相关的代价。

为了更清楚地看懂这一点，考虑两类情况，并设$i=1$。可以看到，对$P(w_i)$来说r是线性的，（因为式（14-47）最右边求和式固定。——译者注）则在$P(w_i)$的两个极值上，即$P(w_1)=0$或$P(w_1)=1$，r取得最大值。如果设$C_{11}=C_{22}=0$，那么r的最大值为

$$\int_{\Omega_1} C_{12}p(x\mid w_2)\mathrm{d}x \tag{14-48}$$

或者为

$$\int_{\Omega_2} C_{21}p(x\mid w_1)\mathrm{d}x \tag{14-49}$$

因为$\Omega_1\cup\Omega_2$是完备空间，那么在式（14-51）成立时，式（14-50）取得最小值。

$$\max\left(\int_{\Omega_1} C_{12}p(x\mid w_2)\mathrm{d}x,\int_{\Omega_2} C_{21}p(x\mid w_1)\mathrm{d}x\right) \tag{14-50}$$

$$\int_{\Omega_1} C_{12}p(x\mid w_2)\mathrm{d}x=\int_{\Omega_2} C_{21}p(x\mid w_1)\mathrm{d}x \tag{14-51}$$

如果$C_{21}=C_{12}$，根据最小最大规则，通过选择Ω_1和Ω_2使两种错误的概率相等。也就是说，我们已经选择了一个条件，在没有类别的先验信息情况下，将使最大风险最小化。

14.7 最近邻方法

在前面的几节中，都假设存在概率密度模型，常常是高斯分布。如果我们只有一个训练集，而且这些数据可能符合高斯分布，也可能不符合（或者是其他参数模型分布）。可能就要采用简单的启发式方法，称为“最近邻规则”，即将未知目标的类别确定为训练集中最近邻的类别。

这一节，把最近邻规则进行扩展，并说明通过扩展最近邻规则进行分布密度估计，这与最大似然分类器是等价的。

343

该方法使用环绕未知目标的一个体积V[⊖]。我们只是简单地统计其中各类别出现的点数，那么各类别的条件密度通过下式进行估计

$$p(\boldsymbol{x}\mid\omega_m)=\frac{k_m}{n_m V} \tag{14-52}$$

其中k_m表示体积V内属于类别m的样本个数，体积V以$\boldsymbol{x}$为中心，n_m是训练集中属于类别m的总的样本个数。

⊖ 当然，在三维以上的空间里，这是一个超体积，为了简单，我们将继续使用“体积”这一词，只要明白没有维数限制就行了。

利用常量体积存在一个问题。对于密集的体积区域（训练集的很多点离得很近），内含大量的点，得到的分布密度估计会过于平滑；而对于分布稀疏的区域，虽然是同样大小的体积，却会导致不充分的估计结果。一种简单的解决方法是，让体积与数据挂钩。例如，从n个样本中估计$p(\boldsymbol{x})$，可以以$\boldsymbol{x}$为中心划一块空间，然后使它慢慢增大直到包含k_n 个样本为止，其中k_n是n的某个确定的函数。如果$\boldsymbol{x}$附近的样本密度很高，那么划出来的体积会很小，分布密度估计的分辨率就较好。如果密度小的话，区域就会增大，只要光滑就行。Duda等[14.4]指出，$k_n=\sqrt{n}$ 提供一种合理的k_n表达方式。

可以对k最近邻 (k-NN) 规则稍作扩展，以便把该策略直接用于分类。给定c个训练集，把所有训练集中的所有样本点组合成为含n个点的数据集，即

$$n=\sum_{i=1}^{c} n_i \tag{14-53}$$

其中n_i是训练集i中的样本数。

对于给定的x点，在该点处我们希望确定出统计量，可以找到只包含混合集中k个点的球体V。如果在该球体中，有k_m 个点属于类别w_m，那么像前面一样，通过下式估计概率密度:

$$p(\boldsymbol{x}\mid\omega_m)=\frac{k_m}{n_m V} \tag{14-54}$$

以及

$$P(\omega_m)=\frac{n_m}{n} \tag{14-55}$$

$$p(\boldsymbol{x})=\frac{k}{nV} \tag{14-56}$$

对式（14-54）~（14-56）应用贝叶斯规则，就会得到

344

$$P(\omega_m\mid\boldsymbol{x})=\frac{k_m}{k} \tag{14-57}$$

这条规则告诉我们，对于k个样本，应该考虑未知特征向量的邻域。在邻域内，如果有较多的样本属于类别i而不是其他类别，那么就将未知向量划归为类别i，这样我们就得到了k最近邻分类规则。

应当注意，在k最近邻规则中，我们并没有准确定义“最近”的计算方法。一般认为欧几里德距离是距离的最佳度量方式，但是也可以用其他方法。

根据作者对大量工业数据集的分类经验，发现最近邻算法的分类效果非常好。

k最近邻分类策略在实际中用时存在的主要缺点是，所有的数据都要存储下来。这是一个非常大的存储负担，特别是与参数方法做比较时，因为参数方法只需要几个点。k最近邻方法的计算负担也非常重，为了找到k个最近邻，必须算出未知量到其他所有近邻点的距离。有人利用启发式的方法使这个过程的运算速度显著提高，可以参考相关文献。例如，读读Hand[15.7]中描述的压缩近邻规则（condensed-nearest-neighbor rule）。

14.8 结论

在本章对统计模式识别的简要介绍中，我们已经看到统计方法是如何在决策过程中得到

应用的。我们也看到了优化方法在求解问题时发挥重要作用。利用最大似然法对概率密度进行估计，其中的似然性是概率的衍生物。对于高斯分布来说，最大似然就简化成了平方和误差（sum-squared error）。

我们已经学习了如何找出使总风险最小的决策区间，即使是在不同决策产生不同代价的时候，也可以找到使被积函数为负的积分区间的上下限。甚至在风险不可计算的情况下，也有能使最大风险最小化的方法。

分类经常被认为是一个求“最小距离”的过程。也就是说，要做出使某种距离最小化的决策，在本章中已经看到好几个这样的例子。

14.9 术语

贝叶斯规则	Bayes'rule
聚类	cluster
条件密度	conditional density
决策分界	decision boundary
决策规则	decision rule
判别函数	discriminant function
特征向量	feature vector
似然比	likelihood ratio
线性机	linear machine
线性可分	linearly separable
最大似然	maximum likelihood
最小最大	minimax
多元，多变量	multivariate
先验概率	prior probability
二次函数分类器	quadratic classifier
风险	risk
监督学习	supervised learning
训练集	training set
一元，单变量	univariate
无监督的学习	unsupervised learning

345

作业14.1

假设类别1和类别2都很好地服从高斯密度分布，且参数如下：类别1的均值为0，方差为1。类别2的均值为3，方差为4。把高斯表达式代入式（14-24），推导出选中类别1时x的取值范围公式。需要先对先验概率做出一个合理的假设（通常选择相等的概率）。

提示：替换之后，对两边同时取自然对数。

作业14.2

在一个一维问题中，类别1的条件密度服从均值为0、方差为2的高斯分布，类别2的条件

密度服从均值为3、方差为1的高斯分布。即，

$$p(x \mid w_1) = \frac{1}{\sqrt{2}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x}{\sqrt{2}}\right)^2\right)$$

$$p(x \mid w_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-3)^2\right)$$

1）在相同坐标系下画出这两个密度函数。

346

2）求出似然比是多少？

3）假设$P(w_1)=P(w_2)=0.5$，$C_{11}=C_{22}=0$，$C_{12}=1$，$C_{21}=\sqrt{3}$。采用贝叶斯决策规则的积分形式，求错误概率。

作业14.3

在一个一维问题中，对于特征x，两类别的条件概率密度是

$$p(x \mid w_1) = \begin{cases} \exp(-(x-r)) & x \geqslant r \\ 0 & \text{其他} \end{cases}$$

$$p(x \mid w_2) = \begin{cases} \exp(x-3) & x < 3 \\ 0 & \text{其他} \end{cases}$$

其中$P(w_1)=P(w_2)=0.5$。

1）假设$r<3$，请画出概率密度的示意图。确定使错误概率最小的决策规则，并说明对于x轴上的一点，这个决策规则意味着什么？

2）求出使$P(error \mid w_2)$最小的r值。

专题14A 统计模式识别

14A.1 用统计方法匹配特征向量

就像上面提到的那样，要专门讨论统计模式识别，怎么也得用整本书的篇幅，实际上已经有很多这方面的书籍。我们通过很简单的例子[13.17]，简单介绍该学科的内容。讨论的问题是人脸识别。首先采集只包含脸部的各种图像（这样就避免了分割问题），要求每个人都穿黑色衣服，背后是墙也是黑色的。这样就得到一些分辨率相对较低的图像，图像大小为180 × 120。然后用如图14-5所示的特征提取器扫描图像。每个特征提取器对每个像素点的邻域进行运算，这与核算子的运算方式一样，不过这种运算返回的是与核内黑色像素对应的图像像素的积，而不是积之和。首先我们观察到，参与这种运算的每个核，返回的是在某个方向上局部范围内图像的自相关。Φ_{ij} 表示对像素j的邻域应用核i的结果。然后按下式相加，生成含25个元素的向量，该向量在某种意义上表示原始图像。

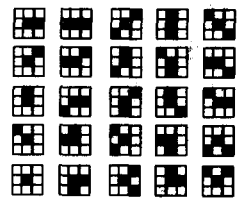


图14-5 25个核，用于从图像中抽取由25个元素组成的特征向量

$$x_i = \sum_{j=1}^{n} \Phi_{ij} \tag{14-58}$$

这样，对于每幅图像，我们都得到由25个元素组成的向量。现在的问题是，如何根据该25维向量做出合适的决策。第一步要将维数减少到某种程度，以便处理起来比25维的更加简单。 347

我们要找到一种降低维数的方法，能将维数从d维减少到$c-1$维，因为我们希望将数据分成c个类别。（由于某些原因，我们必须知道c，在这个例子中c表示对于不同人的人脸类别个数。）接下来的策略是文献中称作“费希尔（Fisher）线性判别”的扩展方法。

假设我们有c个不同的类别，以及一个训练集，每个类别的样本个数为X_i。于是，这是一个监督学习问题。定义类内散布矩阵为:

$$S_W = \sum_{i=1}^{c} S_i \tag{14-59}$$

其中

$$S_i = \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^{\mathrm{T}} \tag{14-60}$$

μ_i 是类别i的均值。于是，S_i 就度量了每类偏离下列平均值的程度。

$$\mu_i = \frac{1}{n_i} \sum_{x \in X_i} x \tag{14-61}$$

定义类间散布矩阵为:

$$S_B = \sum_{i=1}^{c} n_i (\mu_i - \mu)(\mu_i - \mu)^{\mathrm{T}} \tag{14-62}$$

其中，μ是所有训练集中所有点的均值，n_i 是类别i中的样本数量。为了明白上式的含义，请参见图14-6。类间散布表示每类均值和样本总均值间的距离之和。S_B 最大化使各类均值相分离，并远离总均值。

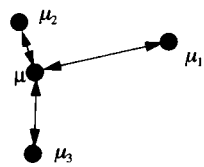


图14-6 类间散布度量类均值与总均值间的总距离

这种做法的思想是找出每个数据向量$\boldsymbol{x}$在向量$\boldsymbol{y}$上的投影，

$$\boldsymbol{y} = W\boldsymbol{x} \tag{14-63}$$

首先，y的维数要比x的维数少。其次，投影之后要能够更好地将各类分开。

从d维空间投影到$c-1$维空间，通过下列$c-1$维线性判别函数实现。

$$y_i = \boldsymbol{w}_i^{\mathrm{T}} \boldsymbol{x} \tag{14-64}$$

如果把y_i 看成是向量的组成部分，$\boldsymbol{w}_i$ 是矩阵W的列向量，那么就可以用下列矩阵方程描述所有的判别函数:

$$\boldsymbol{y} = W^{\mathrm{T}} \boldsymbol{x} \tag{14-65}$$

现在定义准则函数，该函数是W的函数，并且度量类间散布与类内散布之比。我们希望S_B 与S_W之比达到最大，或者使 $S_W^{-1}S_B$ 的某种度量最大。$S_W^{-1}S_B$ 的迹，等于 $S_W^{-1}S_B$ 在主成分方向上的散差（spread）之和。在只有两类的情况下，可以很清楚看出其中的含义。 348

$$J = \mathrm{tr} S_W^{-1} S_B = \frac{n_1 n_2}{n_1 + n_2} \mathrm{tr} S_W^{-1} (\mu_1 - \mu_2)(\mu_1 - \mu_2)^{\mathrm{T}} = \frac{n_1 n_2}{n_1 + n_2} D^2 \tag{14-66}$$

其中D^2是两类间的马哈拉诺比斯距离的平方。

我们先不使矩阵的迹最大化，而是利用下列行列式重新定义J:

$$J(W)=\frac{\left|W^{\mathrm{T}}S_BW\right|}{\left|W^{\mathrm{T}}S_WW\right|} \tag{14-67}$$

因为该行列式的值等于特征值之积，因此也等于主方向上的散差之积。与费希尔线性判别方法的情况相同，通过特征值分析法可以求出该方程的解，而最优W的列向量是与最大特征值对应的特征向量。

$$S_B\boldsymbol{w}_i=\lambda_iS_W\boldsymbol{w}_i \tag{14-68}$$

其中特征值是下列特征方程的根:

$$\left|S_B-\lambda_iS_W\right|=0 \tag{14-69}$$

然后再求解如下方程就得到向量$\boldsymbol{w}_i$。

$$(S_B-\lambda_iS_W)\boldsymbol{w}_i=0 \tag{14-70}$$

这种扩展的费希尔线性判别方法，广泛存在于各种计算机软件包中。经常用它选择原始变量的子集，而不是通过变换使维数降低。通过仔细察看所得到的特征值，并按变量对应的相对大小（正或负）标记（译者注：应按特征值的相对大小（正或负）标记）。较大的特征值意味着对应的变量为类间可分性做出的贡献较大，就和我们在$K-L$变换中学过的一样。

我们减少了25维向量的维数（这里假设类别数目小于25）。在这个低维空间中进行分类器设计，怎么进行呢？这一点很容易做到。利用公式（14-5），或者公式（14-5）中右边的对数形式（请自己进行推导！），推导出判别函数，做到当样本划归为正确的类别时函数值最大。

14A.2 支持向量机

基于支持向量概念的模式分类器相对来说是比较新的。它们由Vapnik[14.12]第一次提出来，以结构风险最小化为基础。在这里提出支持向量分类方法，是因为这类方法看起来比其他很多模式分类方法具有更好的性能。为了说明该类方法，要求使分类间隔最大的优化问题，我们还给出最简单的支持向量方法的推导过程。

349

14A.2.1 线性可分条件下的支持向量机推导

在这一节，假设训练集中的样本可以用线性平面分开。在下一小节中，对这一假设进行讨论，并且说明如何保证这一假设条件。

跟前面一样，想办法用一个超平面将特征空间划分为两部分，其中训练集中的样本是线性可分的。定义类别1中与超平面距离最近的点之间的距离为d_1，同样定义d_2是类别2中与超平面距离最近的点之间的距离，定义分类间距（margin）等于d_1+d_2。我们要找到一个使分类间距最大的超平面（如图14-7和14-8中所示）。

用超平面划分，则可表示成：对于样本点$\boldsymbol{x}$，将这个样本投影到某个单位向量φ，如果$\varphi^{\mathrm{T}}\boldsymbol{x}-q>0$，其中$q$为常量，则按此决策规则确定属于类别1。

设$\boldsymbol{x}_1$是类别1中的一点，同样，$\boldsymbol{x}_2$是类别2中的一点。因为要寻找离决策线最近的点，因此希望选择$\boldsymbol{x}_1$和$\boldsymbol{x}_2$以使它们在φ上的投影尽可能靠近。如图14-9所示，寻找点$\boldsymbol{x}_1$和$\boldsymbol{x}_2$以使下式中的正值ρ最小:

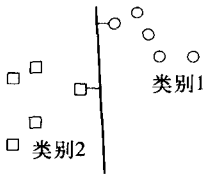


图14-7 较差的分界面（本例二维情况是一条线），分类间距较小

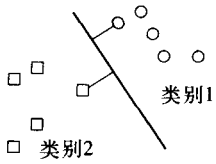


图14-8 较好的分界面，分类间距较大

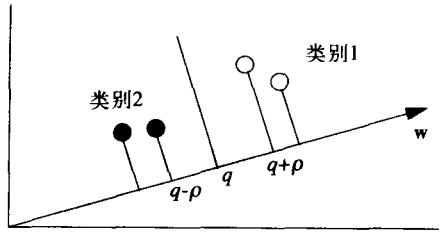


图14-9 标记1和2的两类别的点投影到同一直线w上。样本分界线垂直于w，要从训练集中找出最接近分界线的点

$$\varphi^{\mathrm{T}}\boldsymbol{x}_i = q+\rho \qquad \varphi^{\mathrm{T}}\boldsymbol{x}_2 = q-\rho \tag{14-71}$$

定义I_i是属于类别i的训练样本集。对于I_1中的任意点，都有$\varphi^{\mathrm{T}}\boldsymbol{x}-q>\rho$；对于$I_2$中的任意点，都有$\varphi^{\mathrm{T}}\boldsymbol{x}-q<\rho$我们需要做如下的工作：1）找到一对点，其中每个类别中一个点[⊖]，并且两个点尽可能靠近。我们称这些点向量为支持向量。2）找到一个向量，使支持向量到该向量的投影相距最远。我们通过如下的方法来解决这个问题。

注意φ是一个单位向量。这样它就等于在与它相同方向上某个向量除以自身幅值的结果。我们要找的这个向量，带有一定的特性，这将在下面介绍。设$\boldsymbol{x}_1$表示I_1中的任意点，不必是支持向量；同样设$\boldsymbol{x}_2$表示I_2中的任意点。那么，

$$\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right)^{\mathrm{T}}\boldsymbol{x}_1 - q \geqslant \rho \qquad \left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right)^{\mathrm{T}}\boldsymbol{x}_2 - q \leqslant \rho \tag{14-72}$$

350

可以推出：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_1 - q\|\boldsymbol{w}\| \geqslant \rho\|\boldsymbol{w}\| \qquad \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_2 - q\|\boldsymbol{w}\| \leqslant \rho\|\boldsymbol{w}\| \tag{14-73}$$

定义$b=-q\|\boldsymbol{w}\|$，我们对$\boldsymbol{w}$进行限制，要求它的幅值具有如下特性：

$$\|\boldsymbol{w}\| = 1/\rho \tag{14-74}$$

则下面的两个式子可以描述类别1或类别2中的任意点具有的性质：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_1+b \geqslant 1 \qquad \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_2+b \leqslant 1 \tag{14-75}$$

由于希望找出使公式（14-74）中的类间距ρ最大的直线，可以看出也就是找出幅值最小的投影向量$\boldsymbol{w}$，于是得到 $\boldsymbol{w} = \arg\min\left(\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}\right)$。不幸的是，零向量也能使之达到最小，因此需要添加其他约束条件来避免这种无效解。

设y_i是点x_i的标记，标记定义为：

$$y_i = \begin{cases} 1 & x_i \in I_1 \\ -1 & x_i \in I_2 \end{cases} \tag{14-76}$$

因此不论$\boldsymbol{x}_i$属于哪个类别，表达式$y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)$的值总是大于1或者是等于1。于是我们就得到一个约束条件，而最小化问题变为：找出使$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$最小化的$\boldsymbol{w}$，以使$y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b) \geqslant 1$。

⊖ 每类可能有不只一个支持向量，因为有可能存在两点到超平面有相同的距离。

实现这一点，可以通过建立如下的约束优化问题:

$$L(w,b,\lambda)=\frac{1}{2}w^{\mathrm{T}}w-\sum_{i=1}^{l}\lambda_i(y_i(w^{\mathrm{T}}x_i+b)-1) \tag{14-77}$$

其中l表示训练集中的样本数量。

式中的拉格朗日乘子将全部为正。对w求偏导，

$$\frac{\partial L}{\partial w}=w-\sum\lambda_i x_i y_i \tag{14-78}$$

并且令上式为0，可得:

$$w=\sum\lambda_i x_i y_i \tag{14-79}$$

同样，对b求偏导，可得:

$$\frac{\partial L}{\partial b}=-\sum\lambda_i y_i=0 \tag{14-80}$$

有了这两个结果，关于L的公式可以简化为:

$$L=\frac{1}{2}\sum_i\sum_j\lambda_i y_i\lambda_j y_j x_i^{\mathrm{T}}x_j-\sum_i\sum_j\lambda_i y_i\lambda_j y_j x_i^{\mathrm{T}}x_j-b\sum_i\lambda_i y_i+\sum_i\lambda_i \tag{14-81}$$

351 其中第一项除了系数1/2外，和第二项是相同的。第三项等于0。如果定义矩阵A如下:

$$A=[y_i y_j x_i^{\mathrm{T}}x_j] \tag{14-82}$$

从而得到L的矩阵形式，

$$L=-\frac{1}{2}\Lambda^{\mathrm{T}}A\Lambda+1^{\mathrm{T}}\Lambda \tag{14-83}$$

其中1表示元素全为1的向量。

寻找使L最小化的拉格朗日乘子向量是一个二次优化问题。有若干个数学运算包都具有这种运算功能。一旦我们得到了拉格朗日算子集，就可以利用式（14-79）求出最佳投影向量，它需要求训练集中所有元素的总和。为了求出b，要用到库恩－塔克（Kuhn-Tucker）[14.5]条件:

$$\lambda_i(y_i(w^{\mathrm{T}}x_i+b)-1)=0 \quad \forall i \tag{14-84}$$

原则上，公式（14-84）可以用任意i求解b，但是数学上最好使用平均值。

同样，我们也注意到A的维数与训练集中的样本数目是相等的。因此，在建立SVM之前，如果不对训练集进行“过滤”的话，计算起来将会十分复杂。

14A.2.2 非线性支持向量机

在非线性支持向量机（support vector machine，SVM）中，不是直接处理实际样本，而是采用非线性变换产生一个更高维的向量$y_i=\vartheta(x_i)$。例如，如果$x=[x_1, x_2]^{\mathrm{T}}$是二维的，$y$可能是六维向量:

$$y_i=[x_1^2\ x_2^2\ x_1\ x_2\ x_1\ x_2\ 1]^{\mathrm{T}} \tag{14-85}$$

奇怪的是，维数的提高并不破坏分类器的性能。“自由度”扩展怎么能够提高训练集的精确度和推广能力呢？关于这个问题，Burges[14.2]曾经说过:

一般，把数据映射到高维数的特征空间，预示着支持向量机的推广性能非常差。毕竟，所有超平面集合$\{w, b\}$由$\dim(H)+1$个参数表示。多数包含上亿个甚至无限多参数的模式识别系统，连开始的那道门都通不过。那么支持向量机是怎么做得这么好呢？有人或许会争论说，在给定解的形式这一前提下，最多有$l+1$个可调参数（此处l指训练样本的个数）。但是这似乎是在逃避问题。必须要做点什么，使超平面满足类间距最大的要求，这样才能扭转不利局面……举出一种有力的证据来支持这种主张。

式（14-85）表示的扩展形式增加了空间的维数，并提高了在更高维空间类别线性可分的可能性（关于其原因已经超出了简要介绍的范围）。它同样也提供了一种根据度量值进行非线性信息综合的机制。式（14-85）的多项式形式，只是扩展度量向量维数的一种方法。观察式（14-83）时，另一种更有趣的综合方法就会从你的脑海出现，并且可以看到，计算最优分界面时，不需要知道这些向量本身，而只需计算所有内积的标量结果。因此，我们不需要将每个向量都映像到高维空间，然后再计算向量内积，只要事先算出其内积结果就行。 352

核与内积

在$\boldsymbol{y}_i=\vartheta(\boldsymbol{x}_i)$，$\vartheta:(\Re^d\rightarrow\Re^m)$，$m>d$定义的更高维空间，寻找最佳分界面。然后，关于$A$中元素的等式变为新向量$A=[(y_i\ y_j\ \vartheta^{\mathrm{T}}(\boldsymbol{x}_i)\ \vartheta(\boldsymbol{x}_j))]$的内积的函数。为了表示方便起见（从而使结果更清楚），我们定义一个核算子，$K(\boldsymbol{x}_i, \boldsymbol{x}_j)$，其中考虑了非线性变换$\vartheta$以及内积。现在的问题不是“我应该用什么样的非线性算子呢”，而是另一个不同的问题：“对于一个特定的核，它可以表示一个非线性算子和一个内积的组合吗？”。令人惊奇的是，答案竟然是肯定的，在特定条件下是可能的。这些条件称为Mercer条件：给定一个带两向量自变量的核函数$K(\boldsymbol{a}, \boldsymbol{b})$，如果对于具有有限能量（即，$\int(g(\boldsymbol{x}))^2d\boldsymbol{x}$是有限的）的任意$g(\boldsymbol{x})$，都有$\int K(\boldsymbol{a}, \boldsymbol{b})\ g(\boldsymbol{a})\ g(\boldsymbol{b})\ d\boldsymbol{a}\ d\boldsymbol{b}\geqslant 0$，那么就存在一个映像$\vartheta$和$K$的下列形式的分解

$$K(\boldsymbol{a},\boldsymbol{b})=\sum_i\vartheta_i(\boldsymbol{a})\vartheta_i(\boldsymbol{b}) \tag{14-86}$$

在公式（14-86）中，下标i表示向量函数ϑ的第i个元素。因此，该表达式表示一个内积。注意到Mercer条件只是简单地规定：如果K满足这些条件，那么K可被分解成函数ϑ的两个实例的内积。这就行了，我们不必知道更多的。实际上即使向量ϑ的维数可能是无限的，那也不成问题。

我们知道有一个核函数符合Mercer条件，并且在支持向量机文献中经常被用到，这就是径向基函数。

$$K(\boldsymbol{a},\boldsymbol{b})=\exp\left(-\frac{(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{a}-\boldsymbol{b})}{2\sigma^2}\right) \tag{14-87}$$

在文献中，支持向量机被用于解决各种问题，例如人脸识别[14.10]，以及乳腺癌检查[14.1]。在这些文献的前期研究[14.7]和比较分析中，学者们用经验验证了支持向量机比古典分类工具如神经网络和最近邻规则[14.9,14.13,14.14]的效果都要好。有趣的一次是，针对超光谱数据，将支持向量机与另一个分类器做了对比。比较结果证明，对于多谱数据（将原始超光谱数据过滤后得到），用支持向量机分类器在性能上明显优于基于原始数据的分类器[14.8]。 353

14A.3 结论

统计学方法提供了基于度量值的决策工具。如果那些度量值有足够大的差别，可以简单

地利用阈值就能分开，那么就不需要更复杂的统计方法。但是,搜集一堆度量进行简单的比较，在通常情况下并不足以用来做出决策。

从14A.1节看到的例子，仅仅是机器视觉的一个具体应用，我们用它介绍了统计模式识别这一学科的某些方面。如果想了解有关统计方法的知识，本书介绍的内容远远不够，不能让你掌握需要的全部知识。因此你应该修一门完整的统计模式识别课程。希望通过这一章的学习，能够促使你选修一门统计模式识别课程。

在14A.1节，我们推导出一个目标函数，如果它取极大值，就可以使不同类别的数据的投影值尽可能分开。它又将最大化问题转变成了特征值问题。

SVM找出能使类间距最大化的决策分界，此处的类间距指的是，不同类样本与决策分界面相距最近的点之间的距离，支持向量机的推导需要用到带拉格朗日乘子的受约束最优化问题。

14A.4 术语

类间散布	between-class scatter
费希尔线性判别	Fisher's linear discriminant
类间距	margin
Mercer条件	Mercer's condition
支持向量	support vector
类内散布	within-class scatter

参考文献

[14.1] P. S. Bradley, U. M. Fayyad, and O. L. Mangasarian, "Mathematical Programming for Data Mining: Formulations and Challenges," *INFORMS Journal on Computing*, **11**(3), pp. 217–238, 1999.

[14.2] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, Dordrecht, Kluwer, 1998.

[14.3] R. Duda and P. Hart, *Pattern Recognition and Scene Analysis*, New York, Wiley, 1973.

[14.4] R. Duda, P. Hart, and D. Stork, *Pattern Classification, Second Edition*, New York, Wiley, 2001.

[14.5] R. Fletcher, *Practical Methods of Optimization*, New York, Wiley, 1987.

[14.6] K. Fukanaga, *Introduction to Statistical Pattern Recognition*, New York, Academic Press, 1972.

[14.7] B. Karacali and H. Krim, "Fast Minimization of Structural Risk Using the Nearest Neighbor Rule," *IEEE Transactions on Neural Networks*, **14**(1), Jan. 2003.

[14.8] B. Karacali and W. Snyder, "On-the-fly Multispectral Automatic Target Recognition," *Combat Identification Systems Conference*, Colorado Springs, June, 2002.

[14.9] D. Li, S. M. R. Azimi, and D. J. Dobeck, "Comparison of Different Neural Network Classification Paradigms for Underwater Target Discrimination," *Proceedings of SPIE, Detection and Remediation Technologies for Mines and Minelike Targets* V, **4038**, pp. 334–345, 2000.

[14.10] E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines: An Appli-

cation to Face Detection," *Proceedings of Computer Vision and Pattern Recognition*, Puerto Rico, June, 1997.

[14.11] C. Therrien, *Decision, Estimation, and Classification*, New York, Wiley, 1989.

[14.12] V. Vapnik, *The Nature of Statistical Learning Theory*, Berlin, Springer, 1995.

[14.13] M. H. Yang and B. Moghaddam, "Gender Classification Using Support Vector Machines," *Proceedings of IEEE International Conference on Image Processing*, Vancouver, BC, Canada, September, 2000, vol. 2, pp. 471–474, 2000.

[14.14] Y. Yang and X. Liu, "Re-examination of Text Categorization Methods," *Proceedings of the 1999 22nd International Conference on Research and Development in Information Retrieval (SIGIR'99)*, Berkeley, CA, pp. 42–49, 1999.

354
~
355

第15章 聚　类

Woes cluster; rare are solitary woes.

Edward Young

在本章，我们研究第14章曾经涉及的一个问题，即一个训练集仅仅只包括一些点，而这些点没有任何标记来表明它们属于哪一类。在前一章，我们只对这个领域作了简要的介绍，并建议读者参考其他著作[14.4,15.7]，以获得更加全面的认识。这里略去了一个非常重要的研究领域，就是生物学模型在聚类方面的应用[15.4,15.5,15.6]，强烈建议读者看看这些参考资料。

我们在更一般意义上讨论聚类问题，但会重点介绍聚类的一个特殊应用实例，也就是识别Hough变换数组的峰值。

考虑人造卫星模式分类问题。人造卫星俯瞰地球，绕地球运动。从地球表面上的每个被观察点，采集从该点地面发出或反射的一些光线的度量数据。一般情况下，从某个观察点采集7种不同的度量值，每个度量值对应不同的光谱带。因此所得到的“图像”的每个像素，都将是一个7维向量，向量中的元素可能表示远红外、近红外、蓝光、绿光等光谱强度。假设我们对训练集进行标记，指明了小麦、谷物、草、树木等对应的像素区域。有了这些训练集，似乎能够设计出一个分类器。事实上确实可以。到目前为止，这个问题还属于监督学习问题。然而再想想被称为“树木”的这个类别，该类包含常青类树木以及叶子已落的树木，随着季节的变化，它们会产生截然不同的光谱效果。这种情况说明存在一个参数分类器不容易解决的模式分类问题。当然我们可以用非参数的方法，但是参数方法更加吸引人。取代非参数分类器的一种方法是，考虑用某种方法确定子类的存在，把训练集中的点归类到正确的子类中去，然后将每个子类用参数形式表述。图15-1显示的是二维空间问题，可以很清楚地看到，同一个训练集中存在两个类别。在下面的讨论中，将称这些子类为“聚类”。每个聚类都可以用二维高斯函数精确表示。但整个度量空间很明显是双模式（bimodel）的。

图15-1　有两个子类的训练集合

356

对我们人类来说，这样的聚类在二维空间中很容易做到可视化，但在超过三维的空间中就不容易做到可视化了。

15.1　聚类之间的距离

定义并显示点与点之间的距离度量相对比较容易，包括欧几里德距离$d(a, b)=|a-b|$，城市街区距离$\sum_i |a_i - b_i|$。这种思想可以很容易推广到用来度量点与聚类之间的距离。一种度量方法是点x与聚类均值μ之间的欧几里德距离(或城市街区距离)。另一种度量方法，考虑了聚类的分布情况，也就是大家熟悉的马哈拉诺比斯距离（参见14.5节）。点x和聚类A之间的马哈拉诺比斯距离（的平方——译者注）如下。

$$d_{\text{mah}}(x, A) = (x-\mu_A)^{\mathrm{T}} K_A^{-1} (x-\mu_A) \tag{15-1}$$

其中K_A^{-1}是聚类A的协方差的逆。

我们还没有考虑两聚类间的距离度量方法，而讨论聚类算法，需要有这样的度量方法。

如何定义两聚类之间的距离就不那么显而易见了（从下面的讨论中就可以看到），因为每个聚类都有可能包含很多点。最简单的方法是利用两聚类样本均值之间的距离（也称为质心距离）

$$d_{\text{mean}}(A,B)=|\mu_A-\mu_B| \tag{15-2}$$

也可以考虑把点与聚类之间的马哈拉诺比斯距离推广到聚类之间，定义为

$$d_{\text{fisher}}(A,B)=\frac{|\mu_A-\mu_B|}{\sigma_A^2+\sigma_B^2} \tag{15-3}$$

先把两聚类映射到两均值的连线上，然后求出上式分母上的两个σ值。式（15-3）中的参数是映射后数据的样本均值和样本方差。

关于如何定义两个聚类之间的距离，可以提供一种更为正式的描述方式，首先说明这种距离应具有的特性，要求它满足下列条件：

$$\begin{aligned}&d(A,B)\geqslant 0\\&d(A,B)=0\text{，如果}A=B\\&d(A,B)=d(B,A)\end{aligned} \tag{15-4}$$

357

满足这些条件的度量方式有很多。例如，可以在整个样本空间对密度求积分，得到散度（divergence），

$$d_{\text{div}}(A,B)=\int[p(\boldsymbol{x}\mid A)-p(\boldsymbol{x}\mid B)]\ln\frac{p(\boldsymbol{x}\mid A)}{p(\boldsymbol{x}\mid B)}\mathrm{d}\boldsymbol{x} \tag{15-5}$$

对于多元高斯函数，公式变为

$$((\mu_1-\mu_2)^{\mathrm{T}}(K_1^{-1}+K_2^{-1})[\mu_1-\mu_2]+\mathrm{tr}(K_1^{-1}K_2+K_2^{-1}K_1-2I))/2 \tag{15-6}$$

当两个协方差矩阵相等时，上式可以简化为

$$\Delta^2=(\mu_1-\mu_2)^{\mathrm{T}}K^{-1}(\mu_1-\mu_2) \tag{15-7}$$

Chernoff 距离是

$$d_{\text{ch}}(A,B)=-\ln\int(p(\boldsymbol{x}\mid A))^{1-s}(p(\boldsymbol{x}\mid B))^{s}\mathrm{d}\boldsymbol{x} \tag{15-8}$$

对于多元高斯函数，公式变为

$$\begin{aligned}&\frac{1}{2}s(1-s)(\mu_A-\mu_B)^{\mathrm{T}}[(1-s)K_A+sK_B]^{-1}(\mu_A-\mu_B)\\&+\frac{1}{2}\left(\ln\frac{|(1-s)K_A+sK_B|}{|K_A|^{(1-s)}|K_B|^{s}}\right)\end{aligned} \tag{15-9}$$

当s=1/2时，Chernoff距离变为巴塔恰里亚（Bhattacharyya）距离：

$$\frac{1}{8}(\mu_A-\mu_B)^{\mathrm{T}}\left(\frac{K_A+K_B}{2}\right)^{-1}(\mu_A-\mu_B)+\frac{1}{2}\ln\frac{\left|\frac{1}{2}(K_A+K_B)\right|}{|K_A|^{1/2}|K_B|^{1/2}} \tag{15-10}$$

在文献中最常用的聚类距离是最近邻距离：

$$d_{\min}(A,B)=\min_{a,b}d(a,b)\quad\text{对 }a\in A,b\in B \tag{15-11}$$

也就是说，在一个来自聚类A，一个来自聚类B的所有点对中，选择两个最靠近的点，并将这

两点间的距离定义为两聚类间的最近邻距离。

同样，可以定义最远邻距离为:

$$d_{\max}(A,B)=\max_{a\in A,b\in B}d(a,b) \tag{15-12}$$

以上每种定义给出了标量形式的度量方式，在某种意义上表示了两个聚类的分开程度。

无论何时，在对具有向量值的量进行距离度量时，应该注意到坐标比例的变化会引起结果发生改变。例如，考虑图15-2所示的点集。另一个说明聚类影响的例子，是对向量$[a,b]^{\mathrm{T}}$的分类问题，其中a表示居住人口数，b表示大学数目。现在，考虑向量x和向量y之间的距离，358

$$d(x,y)=\sqrt{(x_a-y_a)^2+(x_b-y_b)^2}\approx x_a-y_a \tag{15-13}$$

在这种情况下，第二个特征不必考虑，因为和第一个特征的影响相比，它的影响是微不足道的。解决这种问题的一种通用的标准方法是，将每个特征除于它的标准差。

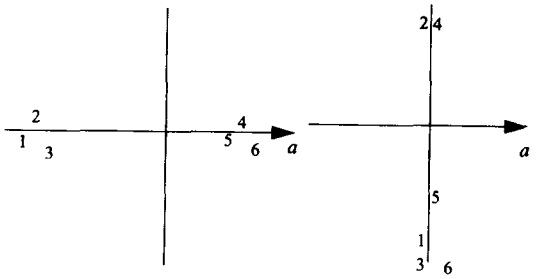


图15-2　坐标轴的比例变化对聚类结果有明显的影响

15.2　聚类算法

这里只介绍两种聚类算法，合并聚类（agglomerative clustering）和k均值聚类，并把聚类问题表示成最优化问题。文献中还有其他一些算法，但在实际应用中用得比较少。

15.2.1　合并聚类

在合并聚类中，首先把训练集中的每个数据点都看作一个单独的聚类，如果有N个数据点，那么一开始就有N个聚类。接下来，进行迭代：将两个最近的聚类合并。

“合并”的含义是：1）找到两个最近的聚类（把每个聚类当作一个集合）；2）建立一个新的集合，由两个聚类合并而成；3）去掉原来的两个聚类，如图15-3所示。

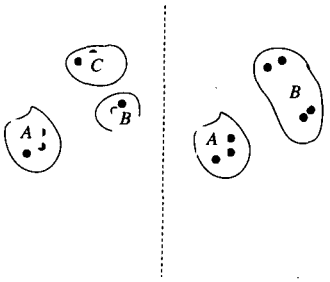


图15-3　在聚类迭代之前（左边），数据点已经被分为3个聚类。聚类B与C之间的距离要比A与B或者A与C之间的距离近（以某种方式度量），因此合并B与C，并且重新命名一个新的聚类B

依此类推直到只剩下c个聚类。假设c的值事先知道。

开始的时候，每个聚类只包含一个单点，聚类之间的距离就是点之间的距离。然而合并过程开始后，就要用前面介绍的聚类之间的距离度量方法。

359

合并聚类的特点

用这种算法得到的聚类，很大程度上依赖于所用的距离度量方法。例如，利用d_{min}进行距离运算，得到的聚类可以用如图15-4中的连线所示，于是得到这个图的最小生成树（minimum spanning tree，MST），继续这个过程直到只有一个聚类为止。如果想得到三个聚类，只需将图中两条最长的边割断。人们从中可以意识到，如果形式化的图论运算来源于聚类，那么逆命题也很可能成立。关于图的任何知识，都可以帮助我们设计聚类算法。特别是，下面这种算法可以快速构造一个图的最小生成树。

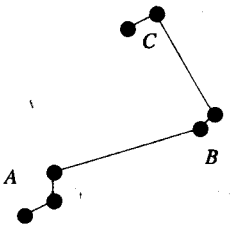


图15-4 通过线段表示每次聚类选择的距离，从而得到最小生成树

将操作y =FIND(x)定义为返回包含x的集合的名字。与此相似UNION(A, B, C)创造一个新集$c=A \cup B$然后删去A集和B集。

算法：求最小生成树

1）计算图中所有的边（在聚类中就是求所有点对之间的距离，不是所有可能的聚类，而仅仅是点）。

2）按边的长度进行排序（如果有N个点，就有N^2条边）。

3）从最短的边开始，对于结点u和v之间的每条边，进行以下运算：

3.1) A=FIND(u)

3.2) B=FIND(v)

360

3.3) IF($A \neq B$) THEN C=UNION(A, B, C),并去掉集合A和B

一般在第3.3步中，对每一个集合都建立一个整数索引。去掉A和B的时候并不把它们的索引也去掉，而是利用A或B的索引（值小的那个）作为新集合C的索引。

像参考文献中[15.1,15.8]所讨论的那样，存在一些并行算法,能够在恒定时间内执行寄存器UNION－FIND运算。并行算法假定存在一个进行FIND运算的查找表,然后通过并行硬件，按如下方式实现UNION运算：对点u和v执行一次UNION－FIND运算。

1）寄存器1<=查找(u)

2）寄存器2<=查找(v)

3）If XOR(寄存器1，寄存器2)，*rewrite*(minimum(u, v), maximum(u, v))，其中*rewrite*是一种并行运算。

rewrite(x, y)将把查找表中所有含y的位置置为x。

现在我们知道了合并聚类算法的工作原理，也明白了使用d_{min}距离的合并聚类算法将会产生点集的最小生成树，而且这种树也可以

361

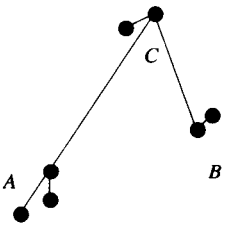


图15-5 用d_{max}聚类，并用线段表示每次聚类选择的距离,得到的聚类结果与利用d_{min}的结果非常相似

利用图论或者并行算法得到。

看看d_{min}是怎样对同一组数据聚类是很有意思的，其结果如图15-5所示。然而，如图15-6至15-8所显示的那样，情况并不总是如此。特别地，d_{min}倾向于选择细长的聚类，而d_{max}将选择大体上圆形的聚类。同学们经常会对最大和最小的说法感到迷惑，因此让我们对所做的工作再多说几句。在合并聚类算法中，我们一直在合并两个最邻近的聚类，而定义聚类之间的距离时用的是最大距离。d_{max}是用于度量聚类之间的距离，也就是这些聚类中那些相互距离具有最大值的点之间的距离。

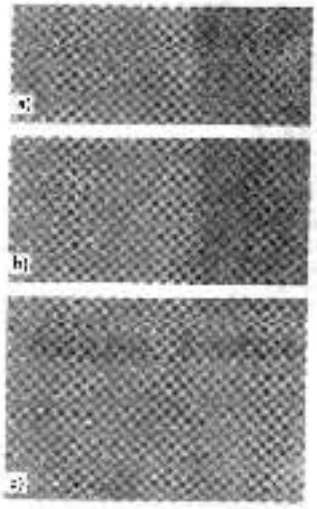


图15-6　三例二维聚类问题（来自文献[14.31]，使用经过允许

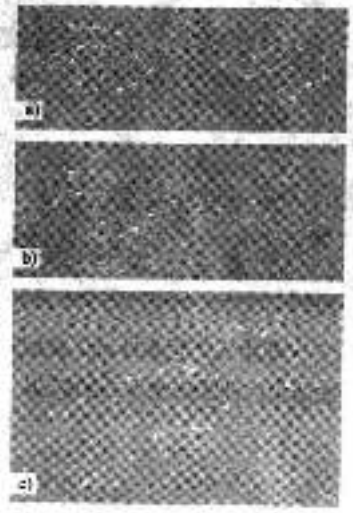


图15-7　对图15-6中的例子使用d_{min}的结果，使用经过允许

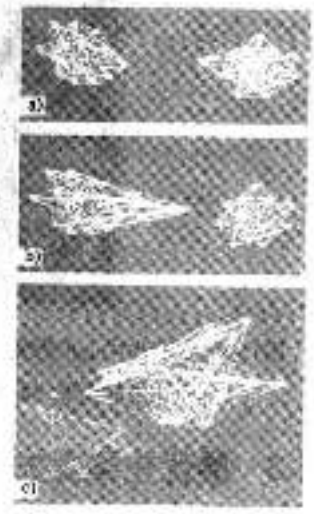


图15-8　对图15-6中的例子使用d_{max}的结果，使用经过允许

15.2.2　k均值聚类

当然还有许多其他聚类方法，其中k均值法可能是最流行的，具体描述如下：

算法：k均值聚类

第一步：以任意的方式，将样本分为几个聚类，或者，如果你不想任意划分，选择一个聚类中心的任意集合，然后将所有样本都划分到最邻近的聚类。如何选择聚类中心与问题有关。例如，如果对彩色空间中的点进行聚类，其中的三个维是红、绿、蓝，选择的聚类中心可以均匀分布在3维空间中，或者使聚类中心沿着点(0,0,0)到最红点、最绿点、最蓝点的连线。

第二步：计算每个聚类的均值。

第三步：重新分配每个样本，使它们属于具有最近均值的聚类。

第四步：如果此次迭代没有发生变化，就退出；否则转到第二步。

图15-9显示的是，利用k均值法对Hough累加器数组中峰值的识别结果。所用的累加器数组与图11-5中的一样，累加器数组中的每个位置，其数值等于该位置处样本点的个数。一开始选择的聚类中心，比较分离，并且远离实际中心。k均值算法的最简单实施方法，不适合这种应用，因为累加器数组中有很多位置只包含一个点，所有这些位置共同作用，会使算出的

均值位置远离我们要找的峰值位置。解决方法是将数值小的位置忽略。在本例中，任何不够三个点的位置都被忽略。其他启发式求解方法也是可以的（参见作业15.4）。

362 ISODATA算法[15.2]对k均值算法进行了扩展。ISODATA算法允许选择合适的聚类数目，并且在规定最大和最小的聚类规模上更加灵活。

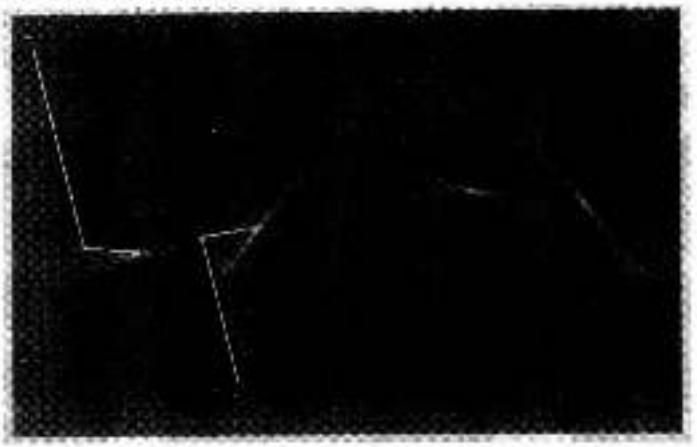


图15-9 两聚类中心的跟踪路径。一开始它们远离Hough累加器数组中的最终位置。线段长度表示从当前聚类中心到下次迭代估计的聚类中心的运动

15.3 最优化聚类方法

试试看，能否通过定义并求解一个最优化问题，从而实现严密、正规的聚类运算。

我们想找到最好的聚类过程，也就是说，将某个样本点分配给某个聚类，从某种意义上说是一种最优的分配。在解决这个问题之前，首先要寻找一种判别分配性能优劣的标量度量方式。类内散布矩阵提供了一种合理的度量方式。

$$S_w = \sum_{i=1}^{c}\sum_{x\in X_i}(x-\mu_i)(x-\mu_i)^t \tag{15-14}$$

其中μ_i是聚类X_i的均值（质心）。S_w提供一种很好的度量方式，它表示训练集样本点与均值之间的平均偏差，然而，为了求最小值，需要用一种关于S_w的标量度量方法，如矩阵的迹或行列式的值。以下S_w的迹

$$\mathrm{Tr}(S_w) = \sum_{i=1}^{c}\sum_{x\in X_i}\mathrm{Tr}((x-\mu_i)(x-\mu_i)^t) = \sum_{i=1}^{c}\sum_{x\in X_i}(x-\mu_i)^T(x-\mu_i) \tag{15-15}$$

是各点与均值的偏差的平方和。这种迹准则在聚类算法中使用的主要缺点是，如果变量的尺度改变，就会产生不同的结果。

当坐标轴比例变化时，S_w的行列式的值是恒定的，但使用行列式时要假设所有的聚类有
363 大致相同的形状。

一旦我们选择了最优化指标，无论是$\mathrm{Tr}(S_w)$、$\det(S_w)$、还是其他指标，为了找到最佳聚类效果，都需要搜索样本点所有可能的聚类空间。因此，还需要考虑优化的方法问题。本节我们只介绍“分支定界”这种方法，但另一种更重要的模拟退火算法，已在2.3.3节讨论过。

15.3.1 分支定界法

假设我们要让一个标量准则函数最小化，把该准则用J表示（为什么这些准则总是用J或H表示？你知道吗？也就是Hamiltonian和Jacobian的首字母）。下面通过一个例子说明这种优化方法。假设要把I、J、K、L四件事分成两个聚类，即聚类1和聚类2。先定义一些标记。

1112表示把I、J、K归到聚类1，L归到聚类2；112X表示把I、J归到聚类1，K归到聚类2，把L归到哪个聚类还不确定。

假设条件

分支定界技术要能够工作需要做些假设：将一个点增加到一个聚类中，总是会使代价增加，这与用什么样的决策无关。也就是说，假设已经算出了112X的代价，即J (112X)=a。当我们对L作决策时，代价就会增加，即J(1121)>a，以及J (1122)>a。

有了这个假定，就可以定义分支定界搜索算法了。既然没有更好的搜索方法，就开始按顺序估计各种可能性，依次尝试1XXX、11XX、111X、1111、1112、112X等。按这个方法，如果得出J (12XX)大于J (1112)的结果，那么再评估12XX的子分支就没有意义了，因为，在某个聚类中增加一个点，只能使准则函数代价增大，我们只能得到比已有值更高的结果。这就是分支定界法的本质。这种方法按顺序对列出的所有可能进行搜索，但是要记住找到的最低值和是以上面的假设为基础对搜索树剪枝。

15.3.2 向量量化

在“向量量化”研究领域，给计算机提供n个向量组成的集合，计算机要对这些向量进行自然分类。换种说法，计算机要找到由c个参考向量组成的集合，这c个参考向量是n个向量的最佳表示。如果觉得这听起来像是聚类，那你就猜对了，它确实就是聚类，所以我们这么称呼它。

364

15.3.3 赢者全得算法

赢者全得策略（winner-take-all，也称竞争学习）最初是由那些为认知过程(称之为“概括”)建立模型感兴趣的研究者提出的，因此每个参考向量或聚类中心，用模拟神经元的一种数学结构来表示。对我们来说，没有必要深入了解神经元的生理学模型。在这里我们用“聚类中心”这个名词，而在有的文献中可能用的是“神经元”。

每个聚类中心ω_i都有一个相关的权向量，$\omega_i=[\omega_{ij}]^T$，$j=1,\cdots,d$。注意ω向量描述d维空间中的一个位置。在进行聚类时，输入向量表示为$[v_1, v_2,\cdots, v_d]^T$。定义赢家是（在某种意义上）离输入向量最近的聚类中心。也就是

$$d(\omega_i, v_j) \leqslant d(\omega_k, v_j) \quad \forall(k \neq i) \tag{15-16}$$

假设聚类中心α是赢家。然后，通过下式调整α的权值，

$$\omega_{\alpha j} \Leftarrow \omega_{\alpha j} + \varepsilon(v_j - \omega_{\alpha j}) \tag{15-17}$$

其中标量ε称之为“学习参数”，该参数的取值一般很小，数量级为0.01。输入数据以随机顺序重复向算法提供。对整个数据集轮一遍称作一代（epoch）。经过几代之后，聚类中心将运动到数据中准确代表聚类的位置。

通常，总是只选择出一个聚类。显然这不是我们所希望的，为了允许有时也能选择其他的聚类中心，增加一种称为“孤独性”（loneliness）的参数。在每一代中，没有赢得任何点的聚类中心，稍稍提高其孤独性的值。聚类中心的选择，很大程度上取决于聚类中心之间的偏离程度，也与孤独性的值有点关系。

Kohonen特征映射

Kohonen特征映射是一种聚类算法，它是对赢者全得算法的扩展。在这个算法中，假设在每对聚类中心之间都存在与问题有关的*拓扑距离*（topological distance）。那么，当对获胜的聚类中心更新时，该拓扑距离意义上的相邻点也跟着更新。主要的聚类，ω_i（获胜者），更新方式如式（15-17）所示。其他聚类按如下方式更新，

365

$$\omega_j^{k+1} = \omega_j^k + F(\eta, d_{ij})(v - \omega_j^k) \tag{15-18}$$

其中，v表示此次迭代中算法的输入数据；F是d_{ij}的某个不增标量函数，d_{ij}是聚类i和j的距离的一种度量；η是该距离的最大值。这个算法很容易编程实现，并且能够收敛到很好的聚类结果。

15.4 结论

我们已经注意到，不同形式的聚类算法对聚类结果有很大影响。已经有人做了一些工作[15.3]，试图减少对算法形式的依赖性，但该领域仍然没有出现新的思想。

我们把聚类方法看作是为了确定一致性的一类方法（例如确定Hough变换中的峰值），但不是利用一致性解决其他问题的方法。

聚类方法总的来说依赖于优化方法。在15.3节中，用使散布矩阵的迹最小化的方法，以获得良好的聚类。

当必须计算点在聚类之间转换的成本时，用分支定界法来加快解决组合问题。

尽管没有用神经网络方面的术语，15.3.3节中的赢者全得策略使用式（15-17）（很容易让人想到梯度下降法），检测最好的聚类中心。

15.5 术语

合并聚类	agglomerative clustering
巴塔恰里亚距离	Bhattacharyya distance
分支定界	branch and bound
Chernoff距离	Chernoff distance
聚类	cluster
竞争学习	competitive learning
距离	distance
欧几里德距离	Euclidean distance
最远邻距离	furthest neighbor distance
k均值	k-means
Kohonen映射	Kohonen map
马哈拉诺比斯距离	Mahalanobis distance
最小生成树	minimum spanning tree

最近邻距离	nearest neighbor distance
合并查找	union-find

作业15.1

证明对于协方差相等的情况，巴塔恰里亚距离将与式（15-7）表示的距离相同。

作业15.2

要把下面的点分成两个聚类:
[0，0]，[0，1]，[0，2]，[0，3]，[0，4]，[0，5]，[0，7]，[0，8]。
画出这些点并标出最小生成树。
用d_{min}找出并确定这两个聚类。如果愿意可以画出图来。
用d_{max}找出并确定这两个聚类。
在这道题中，可以用其他距离度量方法吗？讨论一下。

作业15.3

在你的图像目录里有三幅图像，分别为

facered.ifs
faceblue.ifs

以及（你能猜出来吗？）

facegreen.ifs

一幅彩色图像包含红、蓝、绿三种成分。每个像素可以用8位红、8位蓝和8位绿的数值表示。所以，在这幅图中可能存在有2^{24}种颜色。不幸的是，你的工作站（可能的）只用8位数表示颜色，对应总共256种颜色。你的任务是（你应该选择接受该任务），找到一个办法，使你的工作站可以显示图像中的所有颜色。

方法：用一些聚类的算法。找到能够代表颜色空间的128个聚类，并将所有的点都归到其中某个类别。然后建立一个文件，包含下列数据:

```
brightness_value  red  green  blue
Example
1  214  9  3
```

366 ~ 367

意思是，如果一个像素亮度值为1，那么它就应该以红214、绿9、蓝3的效果显示在屏幕上。这样的像素点看起来像是纯红。于是，每个聚类中心就用一种颜色来表示。在上面的例子中，聚类中心1是几乎纯红的一点。现在，绘制出一幅图像，其中每个像素的亮度值等于最近聚类的类别编号。

作业15.4

在图15-9中，显示的是一个Hough累加器。在CDROM中的hough.ifs，包含同样的累加器数组。通过聚类算法提出一种寻找峰值的新方法。（不要只是简单地找出最亮点）。建议用累加器数值的平方对每个点进行加权，做一些与值平方的指数有关的事。

参考文献

[15.1] R. Anderson and H. Woll, "Wait-free Parallel Algorithms for the Union-Find Problem," *Proc. 22nd ACM Symposium on Theory of Computing*, pp. 370–380, New York, ACM Press, 1991.

[15.2] G. Ball, "Data Analysis in the Social Sciences: What about the Details?" *Proc. AFIPS Fall Joint Computer Conference*, Washington, DC, Spartan Books, 1965.

[15.3] G. Beni and X. Liu, "A Least Biased Fuzzy Clustering Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

[15.4] G. Carpenter and S. Grossberg, "A Massively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine," *Computer Vision, Graphics, and Image Processing*, **37**, pp. 54–115, 1987.

[15.5] G. Carpenter and S. Grossberg, "ART-2: Stable Self-organization of Stable Category Recognition Codes for Analog Input Patterns," *Applied Optics*, **26**(23), p. 4919, 1987.

[15.6] G. Carpenter and S. Grossberg, "ART-3: Hierarchical Search using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures," *Neural Networks*, **3**, pp. 129–152, 1990.

[15.7] D. Hand, *Discrimination and Classification*, New York, Wiley, 1989.

[15.8] W. Snyder and C. Savage, "Content-Addressable Read–Write Memories for Image Analysis," *IEEE Transactions on Computers*, **31**(10), pp. 963–967, 1982.

368

第16章 句法模式识别

Ours is the age of substitution:
Instead of language, we have jargon;
Instead of principles, slogans;
and instead of genuine ideas, bright suggestions.
Eric Bentley

本章讨论另一种完全不同的模式识别方法，它以类比于理解语言的方法为基础。目前这类方法不是很常用，主要是由于它们对噪声和畸变十分敏感。但对于某些特殊的应用场合，这种方法就比较适合用。建议读者对这方面的内容多了解一些，从而认识到其潜在的应用价值。

考虑用链码表示的边界线段。链码中的每个环节都是一个字符，是0~7之间的一个整数，所以边界线段就用一个字符串来表示。句法模式识别的工作基础是，该字符串与形式语言描述的字符串具有类似之处。

16.1 术语

为了方便进一步的讨论，需要先对一些术语进行定义。这些定义关系到对字符串所进行的分析，就像语言分析中对字符串进行分析一样。

*终止符*就是一个单词，如“马”，“土豚”，“教授”，“长跑”，“等级”。终止符也可以是一个线段、图像的一部分或者其他特征。我们一般用小写形式表示终止符。多数情况下，终止符用单个字符表示，如“a”或“o”，但是在英文单词例子中，终止符一般是单词，而不是字母。

*非终止符*是描述文法结构的字符，如“名词”，“动词”，“动词短语”，“副词短语”等等。我们一般用大写字母如“A”、或大写字符串如“动词短语”的缩写“VP”，来表示非终止符。 369

文法，指终止符集合、非终止符集合、以及产生字符串集合的规则集合。句法模式识别的关键技术成分，就在于对于每种文法，都存在一种机器（例如，有限状态机，下推自动机，图灵机），该机器能识别由该种文法产生的语言。

重写规则，或者称为*产生式*，表示允许的替代规则，式中包括一个箭头。箭头左边的字符串可以被右边的字符串代替。

为了更全面地理解句法模式识别，可以参考[16.5]。

本节我们讨论“传统”的形式语言（像chomsky[16.2]中定义的那样）。但是，当前的研究仍在继续，包括随机文法方面[16.3]的研究兴趣，其中重写规则有相应的概率。如果要了解关于形式语言理论的更全面信息，可以参考[16.7]。

下面是关于英语句子有限集的一个文法例子（参见表16-1）:

终止符：{马，男孩，教授，长跑，绿色，那一，快速的，睡觉}

起始符：S

非终止符：{S，VP，NP，N，V，ADV，ART，ADJ}

表16-1 简单文法中的产生式，起始符是S

S	>	NP VP
VP	>	VP ADV
VP	>	V
NP	>	ADJ N
NP	>	ADJ NP
NP	>	ART N
N	>	马
N	>	教授
V	>	长跑
V	>	睡觉
ADV	>	快速的
ADJ	>	绿色
ART	>	the

370 在文法应用中，对任何产生式都可以按任意顺序，使用任意次数。推理是对起始符用产生式获得的一个实例，例如:

S>NP VP>ART N VP>ART N V ADV> 该教授很快睡着了。

表示该文法中的合理推理。

关于文法有几点值得提一提。首先是文法可以产生无限多的字符串。在产生式的左右两边有相同的非终止符，例如NP>ADJ NP,可以很清楚的表示出这种无限性。文法可能产生的所有字符串的集合，称为该文法产生的语言。

对于每种文法，都有对应的自动机来识别由该种文法产生的语言。另外，建立识别机的规则是比较简单直接的。关于识别，我们指的是：如果向机器提供一个用该语言描述的字符串，那么机器就会终止，并且显示“是的，这个字符串可以由该文法产生”。这是一个关键点。如果我们能发明一种文法，可以产生出一组特殊的字符串，那么我们也能够建立一个识别机，而且对我们人类来说，发明字符串生成器要比发明识别机更容易。

16.2 文法类型

所有可能的文法可以分成四大类，这主要取决于对产生式类型的限制条件。

16.2.1 0型文法

在0型文法中，任何重写规则都是允许的。产生式左边可以包含任意的终止符和非终止符，例如下面的产生式是允许的:

abAaBc>abAaCCc

重申一下该式的意思：如果在推理过程中，出现了字符串abAaBc，就可以用abAaCCc来代替。例如aardvabAaBcark可以用aardvabAaCCcark代替。

对于任何0型文法，都存在一个图灵机，可用来识别由该文法产生的语言。

16.2.2 1型文法[⊖]

在1型文法中，除了字符串不允许缩短之外，任何重写规则都是允许的。产生式左边可以

⊖ 1型文法有时候也称为“上下文有关”，但我们不使用这一短语，因为有的人会感到迷惑。

包含任意的终止符和非终止符。因此下式是不允许的，因为结果字符串比原来的短， 371

abAaBc>aabCC

但是下式是允许的，

bAaBc>aabbCC

对于任意的1型文法，都存在一个LR自动机，可用来识别由该文法产生的语言。

16.2.3 2型文法[⊖]

在2型文法中，产生式左边只可以包含一个非终止符。例如下式就是不允许的，因为左边的字符数超过一个。

Bc>aabCC

但是下式是允许的。

B>aabbCC

对于任意的2型文法，都存在一个下推机（pushdown automaton），可用来识别由该文法产生的语言。下面我们介绍一个2型文法实例，你会觉得很有意思。

实例：2型文法

终止符：{0，1}

非终止符：{S}

起始字符：S

表16-2 2型文法实例的产生式

S	>	0S1
S	>	01

该文法会产生什么样的语言呢？能明显看出来吗？结果是一串0后面跟着数目相同的一串1，表示为0^n1^n。

本章只详细讨论2型文法和3型文法。 372

16.2.4 3型文法

在3型文法中，产生式左边只包含一个非终止符。右边部分只包含像“a”或“aA”形式的字符串，也就是一个终止符，或者一个终止符后面跟着一个非终止符。

例如，B>aCCb是不允许的，因为右边字符串不属于允许的形式。但是B>aC是允许的。

16.3 用文法结构进行形状识别

本节我们通过几个例子，说明句法模式识别在形状识别方面的应用。

16.3.1 3型文法

对于任何3型文法，都存在一个有限状态机（finite state machine，FSM），可用来识别由该文法产生的语言。有限状态机系统，其存在的状态数是有限的[⊜]，用大写字母来表示；其输入字符数也是有限的，用小写字母或数字表示。有限状态机的运算通过形如$\delta(A, a)=B$的转

⊖ 2型文法有时候也称为“上下文无关”文法。

⊜ “有限状态机系统存在有限个状态”。虽然这么说，我们不敢相信。

化规则集来控制。也就是说，当机器处于状态A并接收输入a时，就转化为状态B。机器的输出仅仅依赖于它所处的状态。当处于“接受”状态时，机器就会产生一个输出。

从文法产生识别机需要两步，首先，建立一种有点奇怪的状态机，称为不确定有限状态机（nondeterministic finite state machine）。然后，把该状态机转换为我们实际要建立的状态机，对描述规则ECG[⊖]的文法进行这种操作。首先，谈一点心脏的跳动情况。

正常心跳曲线如图16-1所示，曲线由几个波组成：P波是“去极化”电信号，引起两心房发生收缩，心房是指心脏中较小的腔室。心房发生跳动后，ECG返回到零电位（isoelectric line），用i表示，经过一段较短的时间，允许心脏中较大的腔室即心室充满血液，然后心室去极化，产生了QRS信号（用R表示）。对于健康心脏，信号重新返回到零电位，直到心室重新极化产生T波为止，用t表示。于是进入另一个零电位周期，直到P波出现，用P表示。因此，健康心脏产生的符号序列类似为pititipititipititi（永远重复，或者持续90年左右）。图16-2到16-5表示心脏出问题的几种情况。

473

当然，这个问题简化得有点过度，但是它为我们进行研究提供一个有用的实例。表16-3表示产生正常心跳的文法。

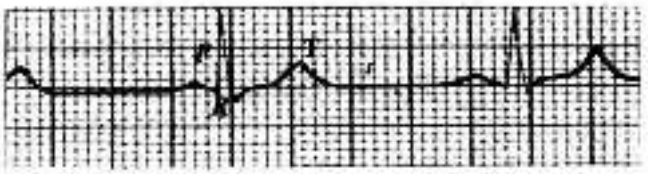


图16-1 正常心跳

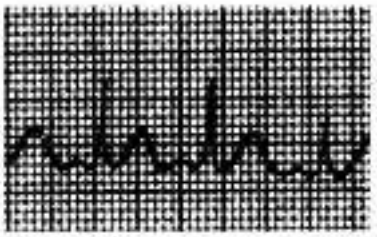


图16-2 窦性心律失常：在T和P之间的间隔时间不稳，pititip

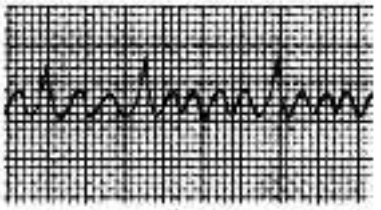


图16-3 心房颤动：出现未受控制的P波，pipipipipipipi

图16-4 房室传导延滞：P与R之间出现延迟，piititipiititip

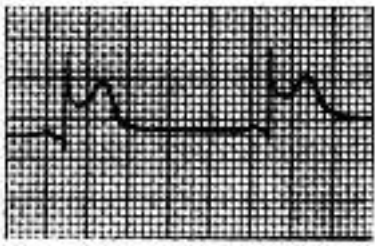


图16-5 心肌梗塞：R、T之间信号没有回到零电位，pititi

为了建立我们所要的机器，称为M，需要遵循以下的步骤：文法中的每个非终止符成为机器

⊖ ECG是指心电图，我们避免了“EKG”来自心电图的德文缩写。

的一个状态。另外还要有一个附加状态，记为Q。然后，对于每个形如$A>aB$的产生式，构造形如$\delta(A, a)=B$的状态变化形式。对每个形如$A>a$的产生式，构造形如$\delta(A, a)=Q$的状态变化形式。最后，如果a是任意的输入符号，那么$\delta(Q, a)=\varphi$，其中φ是表示“空”的符号。

表16-3 产生正常心电图的3型文法

S	>	pA
A	>	iC
C	>	rD
D	>	iE
E	>	tF
F	>	iG
G	>	i
G	>	iH
H	>	i
H	>	iS

机器的状态变化描述如表16-4所示，该机器识别由该文法所产生的语言。

表16-4 识别ECG的不确定FSM

$\delta(S, p)=A$	$\delta(A, i)=C$	$\delta(C, r)=D$
$\delta(D, i)=E$	$\delta(E, t)=F$	$\delta(F, i)=G$
$\delta(G, i)=\{H, Q\}$	$\delta(H, i)=\{S, Q\}$	$\delta(Q, i)=\varphi$

你明白为什么称为“不确定”吗？这种机器在输入为i的情况下，从H变为Q，在输入同样为i的情况下又可以从H变为S。意思并不是说它有时候到Q，有时候却到S，而是说它两种情况都到，这种现象在实际物理机器上是不可能发生的。

为了将这个结果转化为可以实现的情况，按如下方式构造机器M'。

374～375

新机器的状态是原机器状态的所有可能的子集，包括φ（但不是所有的状态都必须用到）。在这个例子中，有2^9个这样的状态。这些状态用原始状态名字的列表加方括号表示。如果变化形式的左边与状态集合运算值有关，那么新状态就会成为原始机器曾经发生过的状态的并集（哇，足够棘手的吧？）。新机器的接受状态，是与Q有关的任何状态，或者曾经是原机器中的任意接受状态。新机器的接受状态是包含原始机器接受状态的所有状态。这一过程产生出如表16-5和图16-6所示的物理上可实现的机器。在这个例子中，尽管新机器存在2^9种状态，但只用到其中的几个状态。

表16-5 确定性有限状态机

$\delta([S], p)=[A]$	$\delta([A], i)=[C]$	$\delta([C], r)=[D]$
$\delta([D], i)=[E]$	$\delta([E], t)=[F]$	$\delta([F], i)=[G]$
$\delta([G], i)=[H, Q]$	$\delta([H], i)=[S, Q]$	$\delta([Q], i)=\varphi$
$\delta([H, Q], i)=[S]$	$\delta([S, Q], i)=A$	

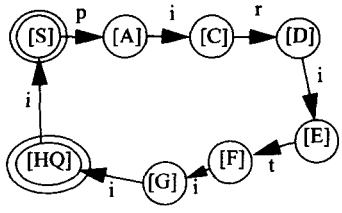
这样，我们就设计出了一个心率识别机。希望你已经观察到，我们只列举了“正常”情况、或者输入是所预期的情况会发生的事。现在，轻松一下，对状态图16-6做些修改，让它包括病态情况。例如，可以增加$\delta(D, t)=Y$，这会产生一个报警状态，表明病人可能患有心肌梗塞（或者主动脉夹层，或者其他严重疾病）。

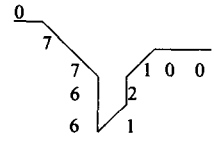
接下来是关于3型文法的另一个例子。该例子中使用了链码。但是，首先介绍另一种描述规则语言的方法：正则表达式。

给出终止符集合T，正则表达式是由T中连续元素以及符号“*”（表示重复）构成的字符串，用圆括号注明运算的顺序，用逗号表示逻辑“或”运算。

本节中，我们将用到终止符{0, 1, 2, 3, 4, 5, 6, 7}（链码的元素）。

图16-7显示的是，由(0, 7)(0, 7)*(7, 6)(7, 6)(61, 72)(1, 2)(1, 2)0(0, 1)*产生的语言的一个元素。图16-8所示的有限状态机可识别由该正则表达式产生的所有字符串。





376

图16-6 识别正常ECG的确定性有限状态机。圆圈内表示接受状态

图16-7 与链码0776612100对应的边界线段

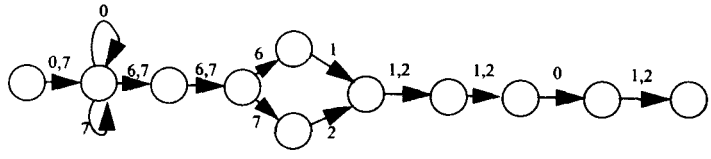


图16-8 不确定FSM的状态图。不确定FSM识别上面规则表达式产生的字符串。被逗号隔开的两个数字，表示其中任何一个都会引起相应的转化。任何其他输入引起的转化，将会产生错误的状态，图中没有显示出来（其中数据经过修改。——译者注）

在这个例子中，显示了如何用链码来表示边界线段，以及用相同的机器如何识别出可能无限多种类似的边界线段。

许多图像表示方法用的是字符串，而不是链码，包括诸如曲线编码[16.4，16.8]。

16.3.2 2型文法

尽管3型文法实现起来很简单，像一个简单的有限状态机，只用双稳态触发器和组合逻辑就可以建立，但它的通用性不够，还不足以解决许多问题。在有的应用中，利用其他类型的文法可能更加合适。本小节中，介绍利用2型文法进行形状识别的两个实例。

染色体识别

下面的例子，摘自Gonzale2 和Thomason[16.6]的著作，最开始时来源于Ledley等[16.9]的著作，它说明了上下文无关文法在染色体类型识别中的应用。

该文法中的终止符是边界线段，用a、b、c、d、e表示，如图16-9所示。在识别框架中，可以称这些为边界基元（boundary primitive）。一条染色体由a-e组成的符号序列来描述。注意除了符号d（和b——译者注）存在两种方式外，其他符号都是有方向性的。

通过该文法可以识别出两种类型的染色体，端着丝粒（telocentric）染色体和亚中着丝粒（submedian）染色体，就像图16-10中显示的那样。每种都可用一个边界线段序列来描述。下

面的文法（如表16-6所示）将产生这两种染色体中的任意一种。

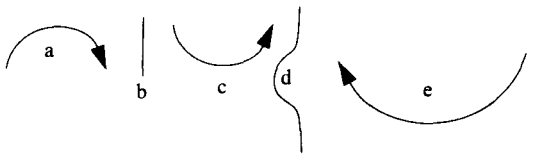


图16-9 句法模式识别中用到的基元边界线段。线段的大小及方向是很重要的

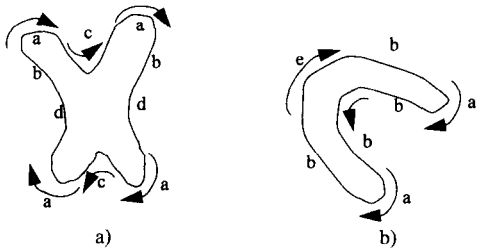


图16-10 a）亚中着丝粒染色体 b）端着丝粒染色体（根据[16.6]重画）

377

表16-6 生成染色体图像的产生式

亚中着丝粒			端着丝粒		
S	>	S1	S>	>	S2
S1	>	AA	S2	>	BA
A	>	CA	A	>	AC
A	>	DE	A	>	FD
B	>	bB	B	>	Bb
C	>	Cb	C	>	bC
D	>	Db	D	>	bD
E	>	cD	F	>	Dc
B	>	e	C	>	b
C	>	d	D	>	a

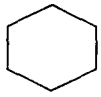
这些产生式并不是随意发明出来的。包括起始符S的前两个产生式，控制产生染色体图像的类型，S1表示亚中着丝粒染色体，S2表示端着丝粒染色体。另外，其他符号表示染色体边界的成分。也就是说，A会引起双臂产生，B会导致底部产生，C使侧面产生，D使胳膊产生，E产生右侧，F产生左侧。

378

形状文法

最后一个例子来自参考文献[9.2]，用形状文法[15.57]产生并且识别纹理。在形状文法中，不管是终止符集合V_T，还是非终止符集合V_N，都是形状的集合，并且受到$V_T \cap V_N=\phi$的限制。

在这个例子中，终止符集合只包含着一个元素，即六边形:



同样，非终止符也只包含一个元素，即一个点: ·。

产生式描述了如何按与终止符的关系扩展非终止符，以产生其他的形状。例如，一个点与六边形的关系，用点正对着六边形的某一边表示，那么在该点所在的位置可以用产生式复制六边形。很多这样的规则就构成了文法，图16-11中显示出两条规则。

Olstad和Torp[16.10]利用句法模式识别，扩展了主动轮廓方法的性能，最近有一篇文章利用这些方法进行成像分析，识别遮挡情况。

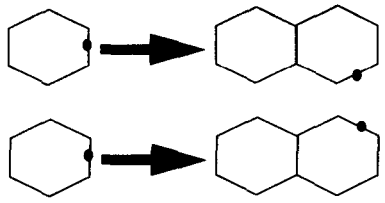


图16-11 用来产生六边形纹理的两条产生式（根据[9.2]重画）

2型语言识别机

2型语言利用下推自动机进行识别。下推自动机是一种有限状态机，只是在下压堆栈的形式上增强为无限额存储器。这是一个后进先出的存储器。在该存储器内保存信息的操作称为压入(PUSH)，读出栈顶信息的操作称为弹出(POP)。注意弹出操作要做两件事情：读出存储的信息，并改变存储单元（栈顶地址——译者注），以便下一个弹出操作能够返回栈顶下面的那个值。

为了实现下推自动机，我们把栈顶符号加到状态变换规则中去。也就是$\delta(A, i, j)=(C, q)$，意思就是“如果状态为A，输入为i，栈顶符号为j，那么状态就转变为C，并且将元素q压入堆栈”。利用这种自动机，就可以识别与16.2.3节例子所示类似的语言，其中1的个数必须与0的个数相等。原理很简单：每当我们看到一个0，就向堆栈中压入一个0，堆栈状态保持不变。当我们看到第一个1时，改变状态并弹出堆栈。随后，每当看到一个1，就弹出堆栈。如果看到另一个0，就进入错误状态。如果没有进入错误状态，则在堆栈为空时，1的个数与0的个数就应该相等，就转变为接受状态。

379

染色体识别和ECG识别两个例子，很好地说明了句法模式识别研究者所关心的主要问题。这两个系统都假设存在一个识别机，能够识别像T波这样的基元。隐含假设了这种基元预处理器是十分简单的，或许就是一个抗噪声的模板匹配器。在实际中，这一点很难做到，而且可能要求所设计的文法本身具有一定的噪声容忍度。读者可以查找参考文献[16.5,16.6]，以便更详细地了解句法模式识别方法。

16.4 结论

除了对终止符进行分类可能涉及优化方法或者一致性问题外，句法模式识别与优化方法或者一致性无关。需要用低层处理算法从图像中抽取这些特征。这其实是句法模式识别的主要缺点。因为该识别方法依赖其他算法提供输入，在以下两种情况下可能会引起识别失败。由于存在噪声、模糊、遮挡以及其他无法预料的变化，符号识别机会因此而失效。另一方面，只是由于目标的特征可能与文法所设计的结果有点出入。它可能看起来与文法所设计的结果相似，但是没有一种简单的方法可以将“相似性”综合到一种文法中去。

16.5 术语

推理	derivation
有限状态机	finite state machine，FSM
文法	grammar
非终止符	nonterminal *s*ymbol

基元	primitive
产生式	production
下推自动机	pushdown automaton
正则表达式	regular expression
正则文法	regular grammar
形状文法	shape grammar
终止符	terminal symbol

380

作业16.1

证明图16-10a中，表示亚中着丝粒染色体的字符串，可以由表16-6中的文法法产生。

作业16.2

早先曾声明表16-6中的文法为2型文法。证明这一声明是否正确。

参考文献

[16.1] M. Chen, A. Kundu, and J. Zhou, "Off-line Handwritten Word Recognition using a Hidden Markov Model Type Stochastic Network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[16.2] N. Chomsky, "Three Models for the Description of Language," *IRE Transactions on Information Theory*, **2**(3), 1956.

[16.3] A. Corazza, R. De Mori, R. Gretter, and G. Satta, "Optimal Probabilistic Evaluation Functions for Search Controlled by Stochastic Context-free Grammars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), 1994.

[16.4] C. Fermuller and W. Kropatsch, "A Syntactic Approach to Scale-space-based Corner Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(7), 1994.

[16.5] K.S. Fu, *Syntactic Pattern Recognition and Applications*, Englewood Cliffs, NJ, Prentice-Hall, 1982.

[16.6] R. Gonzalez and M. Thomason, *Syntactic Pattern Recognition*, Reading, MA, Addison-Wesley, 1978.

[16.7] J. Hopcroft and J. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Reading, MA, Addison-Wesley, 1979.

[16.8] W. Kropatsch, "Curve Representation in Multiple Resolution," *Pattern Recognition Letters*, **6**(3), 1987.

[16.9] R. Ledley, L. Rotolo, R. Kirsch, M. Ginsberg, and J. Wilson, "FIDAC: Film Input to Digital Automatic Computer and Associated Syntax-directed Pattern-recognition Programming System," In *Optical and Electro-optical Information Processing*, ed. J. Tippet, D. Beckowitz, L. Clapp, C. Koester, and A. Vanderburgh, Cambridge, MA, MIT Press, 1965.

[16.10] B. Olstad and A. Torp, "Encoding of a priori Information in Active Contour Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(9), 1996.

381

第17章 应　用

Example isn't another way to teach, it is the only way to teach

爱尔伯特·爱因斯坦

从天文学[17.44]到工业检测，到自动目标识别，机器视觉得到了一系列广泛的应用。要详细介绍与这些应用相关的所有内容是不太可能的。本章为读者选择的内容，不像是一篇教材正文，而更像是带注解的参考资料。我们将简要介绍几个应用实例，并且提供一些参考文献。在下一章，将选择自动目标识别这个应用领域，进行比较详细的介绍。

17.1 多谱图像分析

多光谱图像分析的策略，是把空间表示和光谱表示合二为一。多谱图像分析中每一个像素对应一个向量，该向量是各度量值的有序集合。一个明显的例子就是颜色，颜色值的向量组成元素为[r, g, b]，关于颜色处理的文献有很多。多数已发表的工作致力于对图像质量的改善，只有最近的少数论文研究颜色在识别方面的应用[17.14，17.18，17.53，17.58]。

我们学过的处理单变量图像的方法，例如用马尔科夫随机场方法去除噪声干扰，同样也可用于多谱图像[17.3]。通常，表示像素时只需用向量代替标量即可。

17.2 光学字符识别

尽管我们比较喜欢这一话题，并且存在大量致力于这方面研究的论文（在这一段，我们只引用其中的几篇参考文献[16.1，17.32，17.64]），但由于篇幅所限，我们不能详细讨论相关的方方面面。光学字符识别中的第一个问题是自动分区（zoning）[17.28]问题，即确定文本在页面的位置[17.37]。很多光学字符识别方面的论文，是将我们学过的图像技术用于某种特殊的应用场合，例如细化技术、骨架计算技术等。

17.3 自动诊断和辅助诊断

机器视觉技术在医学方面的应用（和需求）逐渐增多。Tagari等[7.40]为了了解心脏组成的拓扑结构，利用Voronoi图，提出一种基于图的表示方法，使二维图像和三维视图联系起来。利用核医学图像 [17.11]以及多幅常规x光图像[17.52]，计算射出率（ejection fraction），也就是一次心跳射出的血液量与心室总血量（尤其是左心室）之比，这些工作引起了众多研究人员的关注。

有的表达方式适合描述身体的特定部位及运动。例如谐波表示在描述心脏方面得到了广泛应用[8.40]。Gong和Kulikowski[13.16]利用规划策略（planning strategy）识别MRI图像的特征。

热成像在20世纪70年代初期十分火爆，因为人们认为，可借助热像识别诊断类似乳腺癌这样的疾病。然而，这股热潮很快就冷却下来，因为越来越清楚，这种方法的精度不够高，而且误报率太高。因此，尽管在工业和军事方面的很多领域，热成像得到了广泛应用，但在

医学方面的应用却在下降[17.27]。近年来，发表了一些论文和研究成果，对红外热成像在医学方面的应用重新进行了评价，原因在于如下三个方面：1）红外技术有了很大进展。特别是军用夜视系统的发展，使新一代红外摄像机的精度得到很大提高。2）图像处理能力更强。包括图像增强、存储和分割等方面的先进技术，在红外图像处理方面已经得到了有效应用。3）对红外图像所隐含的人体病理生理学方面的知识，我们有了更深入的认识。

17.4 检测和质量控制

人们猜想制造业的检测领域应该对机器视觉更感兴趣，因为企业利用最新技术可以获得竞争优势。然而，事实并非如此。例如，在2000年春季，美国自然科学基金会举办了一次会议，讨论如何才能使企业和大学合作的效率更高一些。超过200位机器视觉公司的总裁（大部分与制造业检测有关）都被邀请参加会议，但结果却只有不到30人出席。为什么看起来如此重要的会议主题，却只有这么少的人出席呢？

383

可能的答案是，多数机器视觉公司规模都还很小。但是也有人会问："这么多的机器视觉公司，为什么规模都这么小呢？"。我们认为该领域具有的独特性才是问题的答案所在。建立一个机器视觉公司，需要的资金投入实际上很少。只需有一台电脑，一些便宜的硬件和几个比较好的点子，就可以加入到这个行业中。如果你不是要进入硬件生产领域、或者是复杂的专业化生产领域，你的公司就可以运转起来，而不需要风险投资。在机器视觉行业，公司并不需要很大的规模。由于公司规模较小，因此受市场驱动的影响就较大，也往往看不到基础性研究在短期内会给他们带来些什么帮助。有时候他们是对的！

尽管如此，在工业检测方面还是做了一些基础性的研究。例如，坐标点对准[17,61]、自动特征抽取[17.62]、重叠零件识别[17.21]、以及视觉技术在装配方面的应用[17.43]。

一个公司，如果其制造业务的流水线生产已经有很多年了，很多产品设计可能还没有输入到CAD数据库中。逆向工程学（reverse engineering）是一个从传统设计到现代数据库的过程。逆向工程学需要读入蓝图[17.13]，也需要实际产品的CAD模型及数据库，进而需要从深度数据（range data）抽取几何基元特征[17.40]，例如球体、圆柱体、锥体等，这些特征也可以从其他坐标测量机得到。

显微镜技术是机器视觉起重要作用的另一个应用领域。例如，宫颈涂片检查(Pap smear)经常采用自动显示系统，而且统计白细胞数目也由计算机来完成。利用外荧光（epi-fluorescence）显微镜技术[14.76]跟踪管状分子，这是最近发表的研究成果。

许多工业零件存在镜面反射，通过使用多个光源可以得到它们的形状和粗糙度特征[17.17，17.56]。

无论如何，不管机器视觉的应用背景是什么，都需要建立系统、建立传感器模型、并生成假设[17.73]。

17.5 安全检测和入侵者识别

检测入侵者需要进行身份识别，要综合应用各种线索[17.6]包括人脸面部特征。关于人脸识别，研究人员做了大量的研究[17.26]，限于篇幅，我们在此就不讨论了。

384

17.6 机器人视觉

机器人视觉用到了我们学过的各种知识。首先摄像机标定[4.37]是一个重要的组成部分，

而机器人不仅需要识别，还需要导航，因此需要建立地图[17.4，17.60]。最近Sogo等[17.54]专注于生成定性的地图，其中嵌入了称作INFRONTOF的信息，而不需要标定过的位置数据。

在工业自动化方面，尤其是利用机器人处理工业零件时，很多应用场合都离不开“稳定状态”这一概念，这可以通过一个实例进行说明。如图17-1所示，是我们可能看到的一枚美分硬币的三幅图像。显而易见，如果我们看到图H或图T，很可能就认出这是一枚美分硬币。而对于图E，虽说有可能是硬币，但看起来太不像了，以致于任何“合乎情理”的机器视觉系统（包括人类本身）都会拒绝它是一枚硬币（如果不提供其他信息的话）。这样一来，一枚美分硬币只存在两种“稳定状态”。机器视觉系统的数据库中，可能存在硬币的两种模型，即PENNYH和PENNYT，并把它们各自当作单独的目标来看待，而不是针对一枚美分硬币的所有可能的视图进行处理。当然，这个“稳定状态”的概念只针对像硬币这样的物体，他们只有几种可能的外形。这种外形常常用“外观”（aspect）这个术语来描述（参见12.6节，其中比较详细地讨论了外观图，这是一种表示物体所有可能外观的数据结构）。

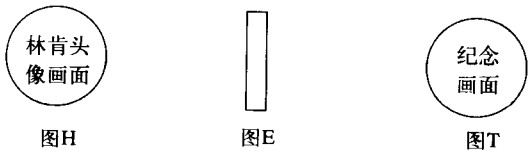


图17-1　一枚美分硬币的三种视图。中间那幅图尽管可能是硬币的一幅图像，但看起来太不像了，应该抛弃

可以对稳定性思想进行推广[17.71]，考虑视点的可能性和稳定性，包括为了获得下一个视图而确定最佳位置[9.93]，作者称这个过程为“自主探测”，这是主动视觉的一种。利用主动视觉进行跟踪的另一种方法，使用所有运动都是平面内运动的假设[17.5]。

17.6.1　机器人外科手术

对精确定位要求较高、而人类又做不到的应用场合，机器人辅助外科手术变得越来越重要。常见的应用场合就是脑外科手术方面，头部需要严格并精确固定不动[17.23，17.34，17.38]。在机器人辅助外科手术中，必须将三维医学图像（MRI或CT）与二维X光投影进行匹配。这可以表示为从二维图像估计三维目标的空间姿态[17.36]。最近机器人外科手术方面的成就，包括跳动心脏上的冠状动脉搭桥术[17.9]，胃外科手术[17.7]和胆囊外科手术[17.24]。 385

17.6.2　机器人驾驶

机器人导航包括在道路内外导航，以及识别和避障[17.74]。利用各种成像模式[17.20，17.65，17.72]以及地面毫米波雷达[17.35]，从正在行驶的车辆上能够检测道路边缘。

再说一次，机器人视觉实际上是一门系统科学。它综合了这本书其他章中学过的各种技术。例如，光流可用于分析摄像机运动，并且能够使摄像机锁定目标[17.49]。Grosso和Tistarlli[13.18]将立体视觉和运动结合起来。Zhang等[17.74]利用了机器人在地面上运动这一假设。

事实上，如果想通过设计一个项目，使学生们学到尽量多的工程知识，机器人系统可能是一个最佳课题。

对机器人视觉感兴趣的读者，可以进一步阅读IEEE *International Conference on Robotics and Automation*的论文集。

参考文献

[17.1] M. Anbar, *Quantitative Dynamic Telethermometry in Medical Diagnosis and Management*, Boca Raton, FL, CRC Press, 1994.

[17.2] M. Barzohar and D. Cooper, "Automatic Finding of Main Roads in Aerial Images by Using Geometric-stochastic Models and Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(7), 1996.

[17.3] M. Berman, "Automated Smoothing of Image and Other Regularly Spaced Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[17.4] Ö. Bozma and R. Kuc, "A Physical Model-based Analysis of Heterogeneous Environments using Sonar – ENDURA method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1994.

[17.5] K. Bradshaw, I. Reid, and D. Murray, "The Active Recovery of 3D Motion Trajectories and Their Use in Prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), 1997.

[17.6] R. Brunelli and D. Falavigna, "Person Identification Using Multiple Cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(10), 1995.

[17.7] G. Cadiere, J. Himpens, M. Vertruyen, J. Bruyns, O. Germay, G. Leman, and R. Izizaw, "Evaluation of Telesurgical (Robotic) NISSEN Fundoplication." *Surg. Endosc.*, **15**(9), pp. 918–923, 2001.

[17.8] V. Caglioti, "Uncertainty Minimization in the Localization of Polyhedral Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 1995.

[17.9] I. W. Chitwood and L. Nifong, "Minimally Invasive Videoscopic Mitral Valve Surgery: The Current Role of Surgical Robotics," *J. Card. Surg.*, **15**(1), pp. 61–75, 2000.

[17.10] I. Cox and S. Hingorani, "An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and Its Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(2), 1996.

[17.11] X. Dai, W. Snyder, G. Bilbro, R. Williams, and R. Cowan, " Left-Ventricle Boundary Detection from Nuclear Medicine Images," *Journal of Digital Imaging*, February, 1998.

[17.12] N. Diakides, ed., Special Issue on Infrared Imaging, *IEEE EMBS Magazine*, June, 2000.

[17.13] D. Dori, "Vector-based Arc Segmentation in the Machine Drawing Understanding System Environment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[17.14] F. Ennesser and G. Medioni, "Finding Waldo, or Focus of Attention Using Local Color Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[17.15] C. Fan, N. Namazi, and P. Penafiel, "A New Image Motion Estimation Algorithm Based on the EM Technique," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3), 1996.

[17.16] D. Fleet and K. Langley, "Recursive Filters for Optical Flow," *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, **17**(1), 1995.

[17.17] J. Franke and W. Snyder, "Determination of Part Pose with Unconstrained Moving Lighting," *International Conference on Robotics*, Atlanta, 1984.

[17.18] B. Funt and G. Finlayson, "Color Constant Color Indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), 1995.

[17.19] D. Geman and B. Jedynak, "An Active Testing Model for Tracking Roads in Satellite Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(1), 1996.

[17.20] F. Gibbs and B. Thomas, "The Fusion of Multiple Image Analysis Algorithms for Robot Road Following," in *Image Processing and its Applications*, IEE, Edinburgh, 1995.

[17.21] W. Grimson and T. Lozano-Perez, "Localizing Overlapping Parts by Searching the Interpretive Tree," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**(4), 1987.

[17.22] Z. Haddad and S. Simanca, "Filtering Image Records Using Wavelets and the Zakai Equation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[17.23] S. Hayati and M. Mirmirani, "Improving the Absolute Positioning Accuracy of Robot Manipulators," *Journal of Robotic Systems*, 2, 1985.

[17.24] J. Himpens, G. Leman, and G. Cardiere. "Telesurgical Laparoscopic Cholecystectomy," *Surg. Endosc.*, **12**, p. 1091, 1998.

[17.25] Y. Huang and C. Suen, "A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[17.26] X. Jia and M. Nixon, "Extending the Feature Vector for Automatic Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(12), 1995.

[17.27] B. F. Jones, "A Reappraisal of the Use of Infrared Thermal Image Analysis in Medicine," *IEEE Transactions on Medical Imaging*, **17**(6), pp. 1019–1027, 1998.

[17.28] J. Kanai, S. Rice, T. Nartker, and G. Nagy, "Automated Evaluation of OCR Zoning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[17.29] T. Kanungo, R. Haralick, H. Baird, W. Stuezle, and D. Madigan, "A Statistical, Nonparametric Methodology for Document Degradation Model Validation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11), 2000.

[17.30] A. Katz and P. Thrift, "Generating Image Filters for Target Recognition by Genetic Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

[17.31] J. R. Keyserlingk, P. D. Ahlgren, E. Yu, N. Belliveau, and M. Yassa, "Functional Infrared Imaging of the Breast," *IEEE Engineering in Medicine and Biology*, May/June, pp. 30–41, 2000.

[17.32] G. Kopec and P. Chou, "Document Image Decoding using Markov Source Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), 1994.

[17.33] A. Kumar, Y. Bar-Shalom, and E. Oron, "Precision Tracking Based on Segmentation with Optimal Layering for Imaging Sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 1995.

[17.34] Y. Kwoh, J. Hou, E. Jonckheere, and S. Hayati, "A Robot with Improved Absolute

Positioning Accuracy for CT Guided Stereotactic Brain Surgery," *IEEE Transactions on Biomedical Engineering*, **35**(2), 1988.

[17.35] S. Lakshmanan and D. Grimmer, "A Deformable Template Approach to Detecting Straight Edges in Radar Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[17.36] S. Lavallée and R. Szeliski, "Recovering the Position and Orientation of Free-form Objects from Image Contours Using 3D Distance Maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(4), 1995.

[17.37] K. Lee, Y. Choy, and S. Cho, "Geometric Structure Analysis of Document Images: A Knowledge-based Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11), 2000.

[17.38] P. Le Roux, H. Das, S. Esquenzai, and P. Kelly, "Robot-assisted Microsurgery; Feasibility in a Rat Microsurgical Model," *Neurosurgery*, **48**, 2001.

[17.39] E. Marchand and F. Chaurmette, "Active Vision for Complete Scene Reconstruction and Exploration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(1), 1999.

[17.40] D. Marshall, G. Lukacs, and R. Martin, "Robust Segmentation of Primitives from Range Data in the Presence of Geometric Degeneracy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3), 2001.

[17.41] N. Merlet and J. Zerubia, "New Prospects in Line Detection by Dynamic Programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[17.42] J. Michel, N. Nandhakumar, and V. Velten, "Thermophysical Algebraic Invariants from Infrared Imagery for Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), 1997.

[17.43] J. Miura and K. Ikeuchi, "Task-oriented Generation of Visual Sensing Strategies in Assembly Tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(2), 98.

[17.44] R. Molina, "On the Hierarchical Bayesian Approach to Image Restoration: Applications to Astronomical Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), 1994.

[17.45] H. Nogawa, Y. Nakajima, Y. Sato, and S. Tamura, "Acquisition of Symbolic Description from Flow Fields: A New Approach Based on Fluid Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), 1997.

[17.46] B. Parvin, C. Peng, W. Johnson, and F. Maestre, "Tracking of Tubular Molecules for Scientific Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[17.47] H. Qi, P. T. Kuruganti, and Z. Liu, "Early Detection of Breast Cancer Using Thermal Texture Maps," *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, Washington, DC, July, 2002.

[17.48] G. Ravichandran and D. Casasent, "Advanced In-plane Rotation-invariant Correlation Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[17.49] S. Reddi and G. Loizou, "Analysis of Camera Behavior During Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 1995.

[17.50] D. Ringach and Y. Baram, "A Diffusion Mechanism for Obstacle Detection from Size-Change Information," *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, **16**(1), 1994.

[17.51] J. Rocha and T. Pavlidis, "A Shape Analysis Model with Applications to a Character Recognition System," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(4), 1994.

[17.52] Y. Sato, M. Moriyama, M. Hanayama, H. Naito, and S. Tamura, "Acquiring 3D Models of Non-rigid Moving Objects from Time- and Viewpoint-invariant Images: A Step Toward Left Ventricle Recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(3), 1997.

[17.53] D. Slater and G. Healey, "The Illumination-invariant Recognition of 3D Objects Using Local Color Invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(2), 1996.

[17.54] T. Sogo, H. Ishiguro, and T. Ishida, "Acquisition and Propagation of Spatial Constraints Based on Qualitative Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3), 2001.

[17.55] W. Sohn and N. Kehtarnavaz, "Analysis of Camera Movement Errors in Vision-based Vehicle Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[17.56] F. Solomon and K. Ikeuchi, "Extracting the Shape and Roughness of Specular Lobe Objects Using Four Light Photometric Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[17.57] G. Stiny and J. Gips, *Algorithmic Aesthetics: Computer Models for Criticism and Design in the Arts*, University of California Press, 1972.

[17.58] M. Swain and D. Ballard, "Color Indexing," *International Journal of Computer Vision*, **7**(1), 1991.

[17.59] T. Syeda-Mahmood, "Indexing of Technical Line Drawing Databases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(8), 1999.

[17.60] H. Takeda, C. Facchinetti, and J. Latombe, "Planning the Motions of a Mobile Robot in a Sensory Uncertainty Field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(10), 1994.

[17.61] M. Tichem and M. Cohen, "Submicron Registration of Fudicial Marks using Machine Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(8), 1994.

[17.62] S. Trika and R. Kashyap, "Geometric Reasoning for Extraction of Manufacturing Features in Iso-oriented Polyhedrons," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(11), 1994.

[17.63] L. Tsap, D. Goldgof, and S. Sarkar, "Nonrigid Motion Analysis Based on Dynamic Refinement of Finite Element Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), 2000.

[17.64] T. Wakahara, "Shape Matching using LAT and its Application to Handwritten Numeral Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), 1994.

[17.65] R. Wallace, A. Stentz, C. Thorpe, H. Moravec, W. Whittaker, and T. Kanade, "First Results in Robot Road-Following," *Proceedings of the International Joint Conference on Artificial Intelligence*, 1985.

[17.66] C. Wang, "Collision Detection of a Moving Polygon in the Presence of Polygonal Obstacles in the Plane," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(6), 1994.

[17.67] C. Wang and W. Snyder, "MAP Transmission Image Reconstruction via Mean Field Annealing for Segmented Attenuation Correction of PET Imaging," *17th International Conference of the IEEE Engineering in Medicine and Biology Society*, Montreal, September, 1995.

[17.68] C. Wang and W. Snyder, "Frequency Characteristic Study Of Filtered-Backprojection Reconstruction And Maximum Reconstruction For PET Images," *17th International Conference of the IEEE Engineering in Medicine and Biology Society*, Montreal, September, 1995.

[17.69] C. Wang, W. Snyder, and G. Bilbro, "Performance Evaluation of Filtered Backprojection Reconstruction and Iterative Reconstruction Methods for PET Images," *Computers in Medicine and Biology*, **9**(3), 1998.

[17.70] C. Wang, W. Snyder, G. Bilbro, and P. Santago, "A Performance Evaluation of FBP and ML Algorithms for PET Imaging," *SPIE Medical Imaging*, 1996.

[17.71] D. Weinshall and W. Werman, "On View Likelihood and Stability," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(2).

[17.72] J. Weng, and S. Chen, "Vision-guided Navigation using SHOSLIF," *Neural Networks*, 1998.

[17.73] M. Wheeler and K. Ikeuchi, "Sensor Modeling, Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(3), 1995.

386 ~ 391

[17.74] Z. Zhang, R. Weiss, and A. Hanson, "Obstacle Detection Based on Qualitative and Quantitative 3D Reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(1), 1997.

第18章 自动目标识别

Luke, you've switched off your targeting computer. what's wrong?

George. Lucas

本章主要介绍机器视觉的应用[⊖]。我们选择的应用领域是自动目标识别（automatic target recognition，ATR），将要介绍在实际应用中如何利用前面学过的数学知识与算法。要强调的一点是，几乎所有的应用场合，需要的不只是某一种技术，而是融合了前面学过的大多数技术。和前面各章一样，我们主要解释一些概念，并且指导大家阅读更高深的文献。但由于本章强调的是应用，因此本章中就不再包含专题这部分内容。

自动目标识别是研究构造系统与相关技术的工程学领域的术语，其目的是为了识别具体物理目标、确定其在场景中的位置、以及计算目标所具有的特征[18.7，18.9，18.69]，经常用于军事环境。针对该研究领域的综述请参考[18.3，18.8，18.21，18.66，18.74，18.79，18.89]。本章只考虑基于图像的自动目标识别系统。因此，对术语（如杂物）的使用，限制在对成像情况有意义的范围内。

18.1 层次体系

本节定义几个自动目标识别系统中常用的术语和缩略语，先从自动目标识别分层体系的五个层次入手。

检测（detection）：确定目标是否存在于给定的场景中。

分类（classification）：该术语的原意是区分履带车和轮式车，至少在军队用语中如此。然而，自从出现了该定义后，多数自动目标识别研究都没有根据性能要求进行分类。另外，现在美国军队正在慢慢淘汰履带车，如果真是做到了这一点的话，原定义肯定会被废除。 392

识别（recognition）：在相似种类的目标中加以区分。例如，将坦克与前端装载机区分开，将吉普车与小汽车区分开，将火箭发射器和学校公共汽车区分开等等。

辨认（identification）：辨认出目标的类型，例如坦克所属的类型（是T90还是M1等）。

特征描述（characterization）：对已经辨识出来的目标进行更详细地描述。在军队用语中，该层次的任务是根据车上装载的武器数量和武器类型来描述目标的特征，例如，一辆T90坦克尾部另挂一个55加仑的油桶。

自动目标识别层次体系的每一级对目标的描述都比前一级精细，其中特征描述揭示了目标的最详细信息。

术语

另外还有一些在自动目标识别文献中经常用到的术语，我们给出如下几个定义：

目标图片（chip）：常常是包含某个目标的一幅小图片（一小块图像区域。——译者注），是

⊖ 作者十分感激Rajeev Ramanath，他对本章内容给予了很大帮助，他实际上还编写了部分内容。感谢Richard Sims和Irvine，他们对内容进行了仔细审查，并提出了重要的意见。

从场景的大图像中抽取出来的，辨别目标可能存在的目标提示算法（target cueing algorithm），就经常以目标图片作为输出结果。

检出率（detection rate）：系统正确检测到目标的比率。

分类率（classification rate）：将目标正确分类的比率，或者更一般地说，在目标已经检测到的条件下，正确识别目标的概率。

杂物（clutter）：图像中不是目标的其他物体。典型的杂物可能有树木、房屋或者其他车辆，包括图像上不是目标的任何物体。

人造杂物（cultural clutter）：指的是人造物体，如碉房等，与自然杂物相对。

误警率（false alarm rate）：一般是指所检测出的"目标"实际上并非目标的比例。但如果任务是分类而不是检测的话，就应该对这个定义修改一下。我们可以看到误警率和误警的概率是不一样的，误警率通常指每平方公里上的误警情况，参见18.3节。

393

FLIR（forward looking infrared）：指中波（3-5μm）和长波（8-14μm）谱带内形成的图像。术语"前视"（forward looking）不再具有实际意义，但是首字缩略语依然使用。

IFF：辨别敌（foe）友(friend)。

18.2 系统组成

自动目标识别系统的算法组成可以分解为：预处理、检测、分割和分类四个部分（参见图18-1）。当然某个系统的实现可能只包含其中的一个或多个部分。例如，如果在视场中只存在一个热点区域，而且不需要进行明确分割，那么一个"团状物"（blob）跟踪器就可以跟踪该图像的重心即可。

自动目标识别系统把"看到"的图像作为它的输入。成像模式是各种各样的，每种模式都有各自固有的优点，因为每种模式各自都有"看到"目标不同属性的能力。例如，战场上伪装良好的坦克，用可见光谱成像可能具有隐蔽性，但用红外谱成像时会看得很清楚，因为它的引擎在不停地转动！图18-2是不同成像模式形成的两幅图像。其中图18-2a是用普通摄像镜头捕捉到的图像，图18-2b是用FLIR镜头捕捉到的图像，注意观察坦克的引擎和振动是怎样当为"发热"而被发现的。

表18-1列出了一些常用的谱带，其中划分谱带的边界波长因人而异[18.9]。

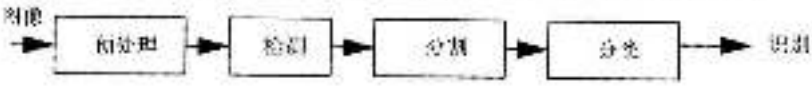


图18-1 常规自动目标识别系统的组成部分

a) b) c)

图18-2 不同的成像模式a）可见光图像 b）热红外图像（注意看左边的发热的坦克引擎）c）地面实况（摘自[18.5]，使用经过允许）

394

表18-1 常用谱带

谱带名字	波长范围	能　源
可见光(V)	0.4~0.7μm	日光
近红外线(NIR)	0.7~1.1μm	日光
短波红外线(SWIR)	1.1~2.5μm	日光
中波红外线(MWIR)	3~5μm	日光
热红外线(TIR)或长波红外线(LWIR)	8~12μm	热源
微波，雷达波RADAR	1mm~1m	热源，人造源

给出一幅场景图像（其中包括可能的目标和背景），需要检测出目标。目标检测可分为两步[18.86]。第一步，利用低层次图像处理技术从图像提取出合适的度量值。然后根据这些度量值对图像进行初步区域分割。第二步，对于分割后区域，通过高层次描述子确定目标是否存在，也可能对目标进行分类。

18.3 算法的性能评价

本节我们考虑对自动目标识别系统的性能评价问题。在下面的叙述中，如果系统的功能是将场景中的目标进行正确分类，我们使用“分类器”这个词。如果系统的指标属于自动目标识别分层体系中的其他层次，我们也会用与前面相同的术语如“误警”等。

为了用语的方便，我们将在模式识别问题的含义中定义这些术语，其中模式识别把结果分为“在（场景中）”和“不在（场景中）”两大类别。这种分类器的应用场合包括自动目标检测（敌对目标或非目标）、医学诊断（肿瘤或非肿瘤），以及数字信道检验（在一端发送‘1’或‘0’，在另一端接收）。第14章中提到了几篇参考文献，可以先回顾一下那些内容。

为了说明检测理论中的几个关键点，先举一个简单例子作为开始。考虑数字通信系统，发送符号a表示“0”，发送符号b表示“1”。

当发送“0”时我们称有假设0(H_0)；当发送“1”时，我们称有假设1(H_1)。(在自动目标识别问题中，可以用H_0表示目标不在场景中，用H_1表示目标在场景中。)。根据自然规律，系统中将会存在噪声。因此，当发送“0”、或者说是发送a时，我们将接收到$a+n$，其中n一个是噪声样本。于是有: 395

$$\begin{aligned} H_0 \quad & Z = a + n \\ H_1 \quad & Z = b + n \end{aligned} \tag{18-1}$$

其中Z表示二值假设的接收信号。

图18-3对该决策理论进行了形象化的总结。

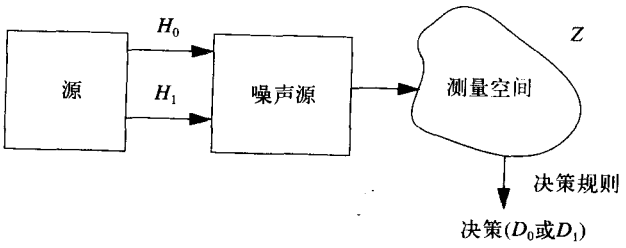


图18-3 二值决策系统

其中，D_0和D_1表示分类器做出的决策，当然也可能是错误的决策，而H_0和H_1则表示事实。我们用$P(H_0)$和$P(H_1)$表示先验概率，或在受控的实验中的真值。

在自动目标识别算法评估中，一般要用到四个术语，它们是:

正确辨别（True positive）：目标存在（例如，病人体内有肿瘤；发送了“1”），分类器也认为目标存在（我们确定病人体内有肿瘤；收到了“1”）。

正确拒识（True negative）：目标不存在（例如，病人体内没有肿瘤；发送了“0”），分类器也认为目标不存在（没有肿瘤，收到了“0”）。

漏检（False negative）：目标存在（例如，病人体内有肿瘤，发送了“1”），但分类器认为目标不存在（病人体内没有肿瘤，收到了“0”）。

误检（False positive）：目标不存在（例如，病人身体健康，发送了“0”），但分类器认为目标存在（病人体内有肿瘤，收到了“1”）。

很明显，漏检和误检这两种错误情况都是不利的。漏检可能会更糟，因为可能会丢掉危险的实际目标，或者会忽视掉一个恶性肿瘤。然而，这两种错误的危害可以用不同代价加以考虑（参见第14章）。有时上面的术语也可用其他名字代替，例如，“误警”（false positive）和“漏警（false misses）”。以这四个值为基础，可以推导出如下两个概率:

396

敏感程度（sensitivity）：正确辨别的概率，也就是正确辨别与正确辨别及漏检之和的比值，即$P(D_1|H_1)P(H_1)$。在目标检测的某个应用领域（与分类、识别和辨别不同），敏感程度指的是检测概率，用P_d表示。

确切程度（specificity）：正确拒识的概率，即正确拒识与正确拒识及误检之和的比值，即$P(D_0|H_0)P(H_0)$。

可以看出$P(D_i|H_i)P(H_i)=P(D_i,H_i)$，那么正确决策的概率为:

$$P(C)=P(D_0,H_0)+P(D_1,H_1)=P(D_0|H_0)P(H_0)+P(D_1|H_1)P(H_1)$$

18.3.1 性能表示

本节讨论如何表示和估计自动目标识别系统的性能。

一旦设计出自动目标识别系统，我们总会有一些可以调整的参数。其中一个参数可能是亮度阈值或目标的像素数阈值；在二次分类器中，决策边界就是参数；而在k最邻近分类器中（回忆14.7节的内容），数值k就是参数，等等。那么系统的性能又是如何随着这些参数的变化而变化的呢？要明白，所设计的系统如果具有100%敏感程度，而不考虑确切程度的话，这并不是最合适的。告诉看病的每一位病人“你得了肿瘤”的医生，就具有100%的敏感程度。但是，我们虽然希望系统的敏感程度能尽可能接近100%，同时也希望尽可能具有理想的确切程度。通过调整有关参数，往往可以改善某种性能，而以牺牲另一种性能为代价，就像图18-4中显示的那样。

在这幅图中，*正确辨别率*（true positive fraction）（$P(D_1|H_1)$的另一种说法）已经在纵轴上注出，而*误检率*（false positive fraction）也在横轴上注出。图中的每条曲线表示某个具体的分类器（或者是某个特定系统）的性能，其中参数在一定范围内变化。在三个系统中，弯曲最厉害、最接近左上角的那一条曲线，很明显是最优的。可以对这个分析进行量化表示，即通过计算曲线上面的面积，并选择面积最小的系统。

这样的曲线称为ROC曲线，其中ROC是通信理论中术语“受试者操作特性”（Receiver Operating Characteristic）的缩写。

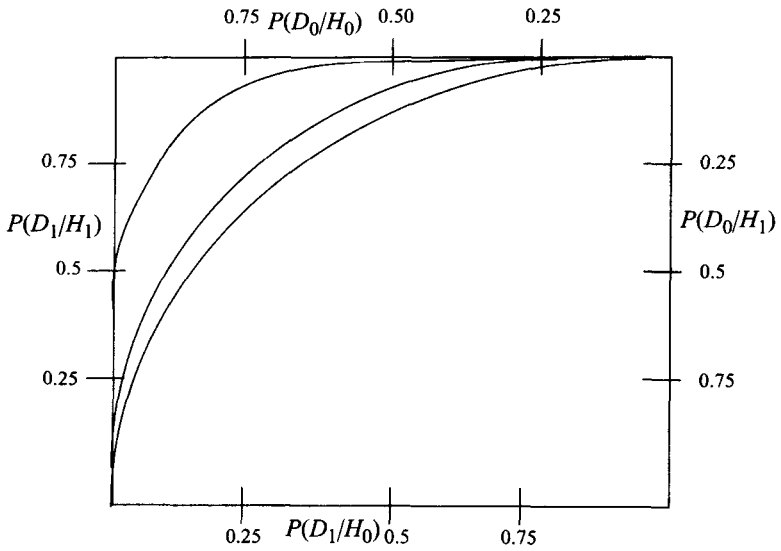


图18-4 受试者操作特性（ROC）曲线

18.3.2 从训练数据生成ROC曲线

对训练数据应用分类器，可以生成ROC曲线。首先，复习一下错误概率的定义。 397

在决策区域内，我们可以写出错误的平均概率如下:

$$\varepsilon = \int_{\Omega_0} P(H_1)p(z \mid H_1)\,\mathrm{d}z + \int_{\Omega_1} P(H_0)p(z \mid H_0)\,\mathrm{d}z \tag{18-2}$$

其中，Ω_i是我们确定类别i的度量空间上的区域。$P(H_i)$是属于类别i的样本先验概率，$p(x|H_i)$是样本属于类别i的条件下度量值（如亮度）的条件概率密度。当然，我们并不知道真正的概率密度，我们只知道根据训练集估计出的密度。另外，根据这些估计出的密度，推导出了决策区域Ω_i。

也可以简单数一下训练集中被错误分类的元素个数，通过这种方式来决定错误率。称这个错误率为*视在错误率*（apparent error rate）。不幸的是，利用视在错误率往往会导致乐观的结果，当测试不属于训练集中的数据时，它会低估系统的错误率。这是因为设计自动目标识别系统的根据，就是使训练集的错误分类样本数最小。并且，如果训练集的分布不能理想地代表数据的真实分布，分类器反映的将是训练集的特性，而不是全部样本的实际分布状况。

我们必须将视在错误率和*真实错误率*（true error rate）区分开。尽管我们没有办法确定真实错误率，但通过下面两种不同的方法，可以较好地估计出真实错误率。

将训练集与测试集分开

这个方法很容易理解。只是将原始训练集随机分为两部分，利用其中的一半数据来建立分类器。然后用另一半数据对系统进行测试。如果我们有很大的训练集（成千上万的样本实例），或者说，有10^d个样本，其中d是问题的维数，那么该方法的结果将是十分合理的。（这里的维数指的是什么？）。不幸的是，在多数问题中这么大的训练集是不可能的。 398

留一检验法

假设在训练集中有n个样本数量。从集合中先取出样本1，并且用其他$n-1$个样本来设计分类器。然后用样本1对结果进行测试。对所有的样本都像这样重复一遍。可以证明所得到的错误率，是对用所有n个样本设计的分类器的期望错误评的一个接近于无偏的估计。当然，这种方法需要设计n个分类器，这是不允许的。但是，有了这个结果，就有了可以绘出ROC曲线的数据了。

18.3.3 性能和系统评价

自动目标识别算法只能放到系统要实现的目标的环境中来评价。对只能跟踪热点的导弹（自动识别算法）用能否区分坦克和汽车的能力进行评价，这是非常愚蠢的。因此，我们只想评估自动目标识别算法的每一个“模块”(block)。但即使这样仍然存在一个问题。可以想像，系统的每部分在各自独立的环境中可能正常工作，但整个系统仍有可能满足不了总指标。因此，人们试图在评价每部分自身的性能时，仍要参照系统要达到的总目标。

性能评价是一个比较棘手的问题，因此急需一种针对自动目标识别应用特点的信息理论度量方法。就像度量传输信道的容量那样，我们很想知道：“在这类场景下工作的任一个自动目标识别算法，能期望得到的最佳性能是什么？”不幸的是，在本书的写作过程中还没有这样的度量方法，尽管已经有了一些进展[18.10，18.61]。

预处理运算对输入数据进行修正，从而能改善图像的质量，这包括去噪，去模糊，以及其他图像修正措施。例如，去噪系统可以用边缘检测方法进行评价，如比较测得边缘的数量、边缘细化程度或者边缘连续性等。总的来说，预处理步骤就是为了增加目标与背景的可区分度，因此特征空间的距离测度就可以利用了，因为检测运算就是要确定目标区域的位置。这样，就可用检测概率和误警概率来评价这一步骤。在目标被检测出来之后，通过分割运算可以提取出目标。然后我们就可以使用一些度量指标，例如错误分类的像素数、实际目标与提取目标之间的相关系数等。

399

18.4 特有的机器视觉问题

自动目标识别领域中的难点可以归纳为目标标志（signature）的易变性、误警率、分割、特征选择、由于信息不完全导致的性能退化，以及性能评价[18.9，18.71]。

18.4.1 目标信号可变性和误警率

目标的“标志”可以是几何性标志，或者是某种光谱性标志。面对“聪明的敌人”，这两种标志都很容易发生改变。在所有的影响因素中，目标表面反射、自然照射以及遮挡情况的易变性，都是十分难处理的问题。

表面反射和自然照射的易变性

所有图像接收的信号都是发射辐射和反射辐射之和。

$$f(x,y)=\xi(x,y)+\rho(x,y) \tag{18-3}$$

但是，发射能量可能比反射能量（在可见光情况下）小得多，或者大得多（在长波红外线情况下）。发射率（emissivity）是发射辐射与总辐射之比。

$$c(x,y)=\frac{g(x,y)}{f(x,y)} \tag{18-4}$$

发射辐射是随着温度增加而增加的：

$$\xi=\sigma A\exp(-T^{-1}) \tag{18-5}$$

上式中的A是表面积，单位是m^2，σ是Stefan常数，$5.67\times10^{-8}Wm^{-2}K^{-4}$。（W是瓦特——译者注）T是绝对温度K度。

因为在24小时内，物体从太阳吸收到的辐射和目标反射的辐射总量变化很大，并且物体上各部件变热和变冷的速度是不一样的，某些部件要比其他部件冷却的更快一些，因此同一物体在一天之内的对比度就可能发生逆转。图18-5显示了这个现象。 400

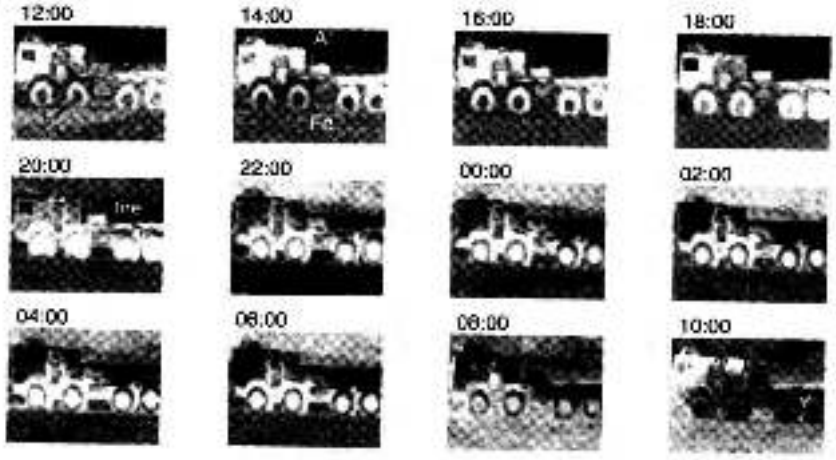


图18-5 物体在24小时内可能出现对比度逆转（图片来自美国军队夜视及电子光学研究中心，使用经过允许）

遮挡

对于工业用机器视觉系统，在制造设备设计时就考虑到要避开遮挡情况，而自动目标识别系统则与工业机器视觉不同，遮挡不仅发生在自动目标识别场景中，目标也经常会出现部分被遮挡的情况，事实上，对手将想办法将自己的装备尽可能地隐藏起来[4.13]。在后面的图18-7中，显示的就是树后隐藏着一辆汽车的图像。

所有这些变化的情况，使“应该怎样对自动目标识别系统进行比较好的训练？”这一问题变得更加突出了。将一个系统训练过度（over-tram）的情况就太容易发生了，对于训练过的数据，系统的性能会表现得非常好，但是对于未参加训练的数据，系统的性能会表现得非常糟糕，即使这些数据在人眼看来可能非常接近，问题不是让检测概率（probability of detection）变得很高，而是让检测概率高的同时，保证误警率比较低。Neyman-Pearson检验[18.53]提出一种限定误警概率条件下的使另一类误差最小化的方法（在这里是指限定误警概率条件下使检测概率达到最高——译者注）。

18.4.2 跟踪

在自动目标识别中，即使不是多数也有很多应用需要进行目标跟踪，而且，这里的跟踪

问题与民用领域的跟踪相比，约束条件更少且更具挑战性。质心跟踪（centroid tracking）是跟踪算法中最简单的类型。虽然仍有很多方法来改善其复杂性[18.39]，质心跟踪器（The centroid tracker）通常假设视场中只有一个目标，而且（目标）亮点要比背景明亮得多。如果这些假设正确，那么目标质心就是视点的中心。更复杂的运动目标跟踪经常要用到最优滤波器，例如卡尔曼-布西（Kalman-Bucy）滤波器。Haddad和Simanca[18.28]讨论了Kalman滤波方法的缺陷，并提出基于小波和Zakai方程的非线性跟踪滤波器。Amoozegar等[18.3]综述了模糊与神经网络技术在跟踪中的应用情况。

401

跟踪过程也同样可以与车辆分类过程结合起来[13.12，18.22]。

18.4.3 分割

在多数自动目标识别场合，将目标与杂物分开是最基本的问题。杂物随着场景类的变化而变化，需要采用自适应表示方法[18.38]。但是，现在关于“信号与杂物”（signal-to-clutter）还没有统一的、能够广泛接受的定义[18.61，18.68，18.78]。

一旦确定了可能的目标位置，就要求尽可能精确地从背景中提取了出来。然而，每个分割器对目标及其邻域像素都有某种先验假设。这些假设并不是对所有的场景都是正确的。我们在第8章中学过，两种常用的分割方法是边缘或边界勾画以及区域生长 [18.68]。边界检测方法以识别图像中的差异为基础，而区域生长利用的则是图像中的相似性。由于边界检测技术对噪声非常敏感，成功的边界检测常常依赖高层语义知识（semantic knowledge）。区域生长技术的抗噪能力较强，因此对语义知识的依赖不是很明显。Qi等[18.63]提出一种有效的分割方法，能够从无人驾驶飞机(UAV)拍摄的图像中分割出人造目标。方法是利用曲率信息，该信息从图像直方图得到，图像直方图用Bézier样条做了平滑处理。实验结果表明，利用增强后的直方图而不是原始图像，可以得到类似的分割结果，但是效率却提高了很多。在[18.87]中，提出一种基于图像金字塔数据结构的分割策略，从塔顶向塔底运行，分层处理图像细节。

正如在第6章中学过的一样，作为分割运算的组成部分，扩散(diffusion)和类扩散(diffusion-like)过程[18.41，18.42]提供了非常好的噪音消除步骤。

18.4.4 特征选择

研究人员利用的特征多数是几何特征、拓扑特征、光谱特征[18.7]。特征选择的主要目的，应使同一类目标的相似程度最大，使不同类目标的相似程度最小。特征选择的数学计算问题在Hand[18.30]所着的教材中进行了很好的讲解。

402

18.5 自动目标识别算法

下面，将本书前面讲过的一些问题与具体的自动目标识别应用联系起来。主要参考了自动目标识别方面的相关文献。我们将不对自动目标识别方面的相关文献做综述，因为利用统计模式识别方法的有关文献实在是太多了，统计模式识别本身就可以编写一本教科书。但是在我们引用的出版物中，将会提到这些方法。

自动目标识别应用所表现出的特性，与多数其他机器视觉应用不同，如与工业检测就不同。关键差别在于:

1）自动目标识别系统不仅必须而且必然要处理非结构化环境，也就不可能对照明、观察

角度、天气环境等进行控制。

2）目标被遮挡不只是可能的，而且是很可能的。

3）目标像素很可能只有少数。而正确分类的概率又强烈依赖于目标的像素点数。这一点可以从图18-6中看出来，图中显示的是虽然只是一个神经网络分类器的结果。但是类似的结果，尤其是在目标像素为50个左右的位置，概率曲线会发生显着变化这一点，对任何系统，包括人眼识别系统，都存在。

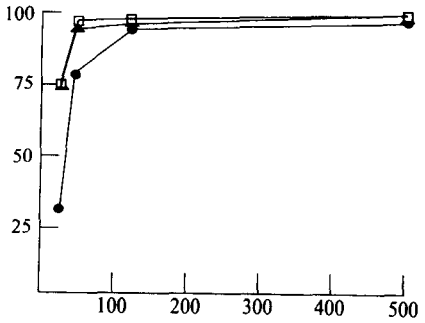


图18-6 检测概率（用圆形表示），对检测结果的分类概率（用三角形表示），以及对检测结果的识别概率（用方形表示），它们是目标像素数的函数（根据[18.84]重画得到）

403

有意思的是，第一次研究目标像素数与P_d的关系，其时间要早于对数字图像的研究，但考虑的是扫描线而不是像素。对于军事目标的红外图像，Johnson[18.36]发现，人眼识别目标时，目标在最小一维的尺寸至少需要四个像素。当然，这个结论也会随着目标复杂程度和识别任务的细节不同而变化，但是跨越3~5个像素的范围，对应方形目标上大约20个目标像素点，就可以产生相当好的效果。参见文献[18.37]。

显然，自动目标识别系统的观测器需要与目标保持尽可能远的距离，这样就造成目标像素数总是很少！事实上，观察者是不愿意靠近那里的，这样就强烈促进了机器人前视观测器的研发需要 [18.35]。

4）必须考虑三维信息。这里三维的意思是，要从多个角度观察目标的外表（参见第9章和第12章）。

5）任何自动目标识别系统必须考虑杂物和伪装物。这些杂物和伪装物是场景中的一些物体，跟目标非常相似，但又不是目标，它们出现在场景中可能出于偶然，但也可能是有意安排的。最明显的伪装物就是与目标类型相同的但却没有敌意的设备。

上面的这些要求，对自动目标识别系统设计影响很大。例如，你可能会考虑这样一个系统，利用多种大小不同的模板进行目标识别，每个模板对应目标的一种外形（aspect）。或者，简单地抽出各种大小不同的窗口，将维数降低（利用K-L变换），并将这些结果传送给分类器[18.14]。

可用不同的分类法对自动目标识别方法进行分类。一种方法是，将自动目标识别方法分成两组[18.7]，第一组是前面提到的经典模式识别法，用的是统计学技术。这种方法很常见，因为实现起来比较容易，而且计算速度也较快，只要能够有效提取出实用的特征就行。另一种方法是基于AI（人工智能）的方法，需要附加符号计算这一步。

根据自动目标识别的作用原理不同，该领域的另一些专家，将自动目标识别系统分为基

于几何（大部分为单一传感器）的系统，和基于光谱的系统[18.71]。基于几何的系统的基础中自于以下设想：如保留关于目标的尽可能多的信息，并利用智能推理，可以对包含所关注目标的场景做出比较好的解释[18.70]。这样，有了对图像生成机制这一正向过程比较好的理解，其逆问题，即根据收到的信号进行目标识别的任务，就变得更加可行了。然而，通过考虑这些模型的相互作用，对这些模型进行集成，形成紧密结合的完整模型的工作，现在还处于起步阶段。

基于几何的系统，可以再继续分成不同的类型，例如，可将基于几何的系统分成使用较大数量模板集（外观图，aspect graph）的系统，以及基于三维模型[18.77]的系统。

基于光谱的系统，通过光谱分析和光谱匹配[18.13]进行目标识别。直观上讲，用到的谱带越多，揭示的场景细节就越丰富，光谱信号中的深层细节能够提高目标的可检测性，尤其对于像素级或者亚像素级的小型目标[18.16，18.47]。

404

另一种自动目标识别系统，以下列假设为基础：来自所关注目标的传感器数据越多，系统的性能就越好。这一点在直观上是显而易见的，因为不同的传感器具有互补作用。由于单传感器的自动目标识别系统在很多方面受到限制，现在有一种向多传感器目标检测系统发展的趋势，于是出现了多传感器数据相关和融合问题，有时称为多传感器融合。但是，信息源可能是不同的传感器（传感器融合），或者是不同的算法（算法融合）[18.32]。

最后，一些研究者将自动目标识别的算法集分成基于模型的方法，基于统计的方法和基于模板的方法。下面对这三类方法进行详细讨论。

18.5.1 基于模型的技术

多数基于模型的技术都是基于几何的，他们试图回答这样一个问题："给定某个观察角度，目标看起来是什么样的呢？"[18.1，18.6，18.12，18.13]。这可能是一个有效的指导思想，因为它提供了由于目标所在位置的影响，目标的哪部分会被遮挡掉的信息，例如，从某个视点来看，可能看不到坦克的炮管。然而，如果我们有目标的三维模型，就可以生成所有可能的视图，并进行综合搜索[18.65]得到匹配结果。基于模型的技术容易与不同的数据类型结合，特别是深度（激光雷达）图像[18.87]。但是，就像机器视觉中随处可见的那样，优化问题必须得到解决，可以利用神经网络[18.29]、遗传算法[18.10]、或者其他优化方法。

通常，只能得到有关场景结构和场景几何的描述，而不是场景的物理学特性（热、光、材料性质等）。匹配也就是假设和验证模型与图像点相匹配的过程，该过程产生了从三维到二维的变换，使三维模型点与二维图像点对应起来。最佳匹配也就是对场景做出最佳解释的变换。从三维到二维的对应解基本上是透视方程的解。投影后的模型点与相应的图像点之间的误差，用来验证匹配的好坏。

这些方法非常有效，但是需要做大量的处理，并且需要大型数据库。当目标存在遮挡时[18.75]，执行效果就比较差，因为遮挡造成了"不完整信息"的出现。为了解决这个问题，要在根据被遮挡的视图恢复目标的"实际几何形状"这方面，做很多工作[18.75]。

图18-7 被树遮挡的坦克图像[18.65]，SPIE(使用经过允许)

一般来说，遮挡又分成两种类型，即相邻类型（墙房或壁龛树之间，如图18-7所示）以及分布类型（树枝之间）。

405

第一种情况处理起来比较容易，因为未被遮挡的部分为解

决问题提供了足够的信息。在[18.73]中，Sadjadi提出一种在遮挡下进行检测的方法。将分割图像的区域边界转化成链码,然后用直方图求交（下一节介绍）将链码与一组模型进行匹配。利用匹配结果作为度量系统的置信度。但是，如果置信度仍然很低的话，那么系统就假设该目标存在部分遮挡，于是将它与数据库中的遮挡模型进行匹配。

18.5.2 基于统计的技术

基于统计的技术背后的思想与第14章中的内容是相同的：1）获取特征；2）计算表征不同类别的统计度量值；3）进行决策，使某种测度最优，如最小代价、最大正确决策概率等。本节中，我们只考虑多谱度量方法。

多谱匹配

一种多光谱分析技术，使用了Swain和Ballard[18.82]针对理想彩色图像提出的直方图求交思想。这一思想很简单，就是比较两幅图像的直方图，并且确定重叠因子（数据库直方图中有多少像素与新图像直方图中的像素相匹配）。特别地，给出两幅直方图，I（来自新图像）和M（来自数据库），每一幅图像都包含n个槽(bin)，求交运算定义为 $\sum_{j=1}^{n}\min(I_j, M_j)$。其结果就是在这两幅图像中颜色相同的像素数目。为了得到重叠因子，可对这个结果进行规范化处理。显然，由于目标的颜色值受光照的影响很大，在这种情况下，简单算法并不能给出较好的匹配结果。为了克服这个难题，Funt和Finlayson[18.24]把直方图求交与“颜色恒常性”[18.23]这一概念结合起来，去掉光照条件变化的影响，实质上是将图像规范化为标准的光照条件。既然数据库中的图像也具有标准的光照条件，就可以利用上面介绍的直方图求交法，将“苹果与苹果做比较”。我们没有对直方图的维数做任何限制，因此可以将这个概念扩展到更高维（更多的传感器），于是就得到一个更加健壮的系统。

在已知目标标志与被观测标志之间，另一种光谱匹配的度量措施是，将标志看成向量，求出两向量之间的内积[18.93]。匹配的越好，两向量之间的夹角就越接近于0。换句话说，如果有两个d维量值光谱性标志，即X和Y，那么这两个量值之间的距离可以用这两个光谱之间的角度来表示 406

$$\cos\theta = \frac{X \cdot Y}{|X||Y|}$$

θ较小表示这两个光谱在数量上是相似的；同样，θ较大表明这两个光谱不同。在[18.93]中，Weisberg等用该测度进行聚类运算，将一幅图像分割成若干块兴趣区域。还有很多其他类似的度量方法也在使用[18.34，18.62]。

在[18.86]中，Trivedi在遥感领域的应用中，提出了使用相对光谱信息替代绝对信息。这增强了系统的健壮性。例如，某个目标也许在某个特定的频道下比背景亮一些，而在另一种频道下又比背景暗一些。

18.5.3 模板匹配

在自动目标识别应用领域中，模板匹配是最简单最常用的算法。当建立外观图容易时，模板匹配是最有效的方法。举例来说[18.51]，如果目标是一艘轮船，很容易找出它的重心，就可以去掉平移的自由度。另外，轮船一般偏长偏窄，其方向可以通过计算主轴来确定。将

刚体运动排除之后，只需要几次模板比较就可确定轮船的精确姿态，并进行识别。

在有的应用中目标姿态很难估计，此时可以使用一组聚焦滤波器（focused filters），其中最强烈的响应指出了目标的姿态[18.2，18.40]。

第9章中的可变形模型，很适合检测和分类已经粗定位的目标[18.44，18.76]，如飞行器的侧面轮廓[18.65]，其中优化问题是作为MAP（最大后验概率）估计问题提出来的。

卫星图像中道路识别和道路跟踪问题，在[18.26，18.48]中进行了讨论。Barzohar和Cooper[18.4]也同样提出了MAP估计问题。

18.6 Hough变换

因为很多人造物体都具有直线形的特点，Hough变换就经常出现在自动目标识别应用中。例如，它可用于[18.19]识别从太空中观察到的导弹轨迹（亚像素级）。对时间计算差分，就可以看到目标的轨迹，当然还附带很多噪声。然而，这条轨迹或多或少是一条直线，经Hough变换后会看得十分清楚。Cowart等[18.19]也考虑到用参数变换方法，这样可以跟踪机动目标。

407

从上往下看，轮船是一个直条形目标（如果轮船没有发生不幸的碰撞）。当用合成孔径雷达（Synthetic Aperture Radar，SAR）进行观察时，轮船图像由亮暗斑点组成(spots and dropouts)。利用简单的Hough变换就可以估计出[18.25，18.50]轮船的方向。令人惊奇的是，Hough变换的噪声敏感度要比主轴方法的噪声敏感度小。如果轮船正在航行，它的航迹就更长、更直，并且更容易发现[18.18]。

从地上望远镜观察，人造卫星具有直线边缘（参见图18-8），同样可用Hough变换识别出这些边缘，并抽取边缘特征[18.20]。

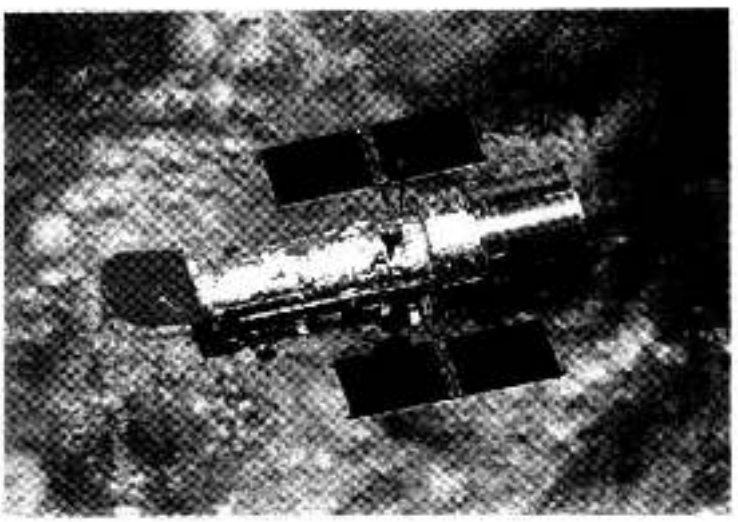


图18-8 哈勃太空望远镜（图片来自NASA）

18.7 形态学技术

在文献中看到的形态学运算类型受自动目标识别特殊需求的影响。例如，用合成孔径雷达往下看目标时，目标往往只有几个像素。因此，消除噪声就显得非常重要，可以利用形态

学中的开运算消除噪声，该运算具有旋转不变性。在文献[18.60]中，将公式（18-6）定义的传统开运算，利用公式（18-7）来代替。

$$(foB)(n,m) = \max_{i,j\in B}\{\min_{i,j\in B}\{f(n+i,m+j)\}\} \tag{18-6}$$

$$\gamma(f_o)(n,m) = \min\{\max_{i,j\in B}\{f(n+i,m+j)\}, f(n,m)\} \tag{18-7}$$

408

其中B是图18-9所示的环形结构元(structuring element)。该结构元的行为随着目标大小L和结构元半径R的相对变化而不同。基于这类方法[18.58]，形成了更明确的目标识别[18.10]或目标跟踪[18.92]的形态学算法。结果如表18-2所示。

图18-9 环形结构元和矩形目标（根据[18.60]重画）

Phan等[18.58]也在利用形态学进行形状识别。

表18-2 修改后开运算的结果（环形结构元）

$L<R$	去掉目标
$R<L<2R$	去掉一部分，保留一部分
$2R<L$	保留目标

18.8 链码

模板匹配和基于特征的方法各自都存在一定的缺陷。多数算法依赖于目标的先验知识，而且还要求在传感器视场内所用特征相对于目标的大小和方向具有不变性。基于模板匹配的方法需要大型数据库，检索时间也很长。然而，可以对轮廓进行匹配[18.72]，利用分割后目标的链码，对链码直方图进行处理（参见图18-10）。这一策略有以下两个有效特性：

1）图像域尺度变化与链码直方图域的垂直移位等价。

2）方向变化与直方图水平循环移位等价。

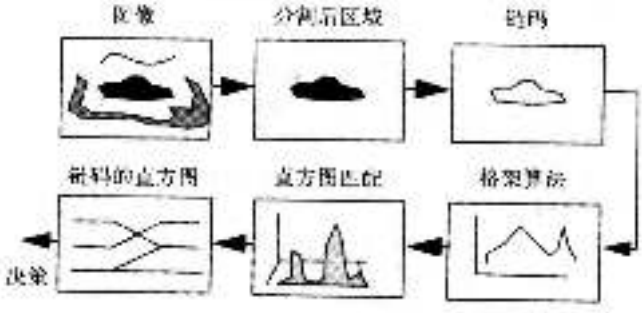


图18-10 Sadjadi提议的方法示意图（经过重画）

409

两个不同的目标可能具有相同的链码直方图。为了解决这个问题，可以用格架算法（mellis algorithm）来区分这两个目标[18.81]。变形模式的格架结构是通过把每个行向量用作“分级模式”（格架中的结点）而建立的。大量观测数据为这一格架提供统计根据，其中用到了维特比算法（参见2.4.2节）。尽管这种方法针对的只是手写体识别，但是也可以推广到任何变形模式情况。

18.9 结论

在公开发表的文献中，有很多文献对不同的识别技术做了对比。例如，Li等[18.43]通过

对神经网络、统计学和基于模型的方法进行对比，得出结论："至少对于FLIR图像，神经网络方法的识别效果要比PCA（主成分分析）和LDA（线性判别分析）方法的效果更好"。他们发现基于Hansdorff距离的方法识别效果也很好。

参考文献中的有关论文所做出的结论，常常使人觉得自动目标识别问题几乎已经被解决了。但实际上，这些系统不是没经实际军事数据进行检验，就是应用范围受特殊应用或特殊条件的限制，其"真实"性能令人怀疑。我们必须明白，这里所讨论的问题，不仅很重要，而且当敌人对这些问题有了新发明时，这些问题处理起来就更加困难。今天的"最新技术水平"并不意味着明天仍然是。

有鉴于此，作者认为要从工程学的视角（也就是从实用的角度）出发，所构建的基本的自动目标识别系统，必须满足下面的条件：

- 可供研究团体使用的作为标准的大型实际图像集合，并且包含真值（ground truth）信息。
- 有一种工具，它能对大量自动目标识别技术进行恰当的评估，并将其中最好的技术组合到一个具有"最佳"性能的系统中。

410 在考虑这些方面时，就会带出若干问题：我们如何来制定一个标准？要达到什么样的目标——是ROC曲线最佳的高效系统，还是硬件最少的轻便系统？很明显，这些性能要求不会同时得到满足，需要进行权衡处理。

然而，仍然可以看出自动目标识别系统发展的总趋势。采用多传感器系统、系统具有轻便性、很少需要人的干预而成为真正意义上的自动化系统、使用越来越多的数学工具，为系统的发展提供了可靠的基础，显而易见这些都是研究必须面对的。在这一章，作者希望写出了问题的广泛性，同时也表达了为解决该问题所取得的科学成就。

参考文献

[18.1] J. Albus, "Applications of an Efficient Algorithm for Locating 3D Models in 2D Images," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.2] K. Al-Ghoneim and B. Kumar, "Combining Focused MACE Filters for Target Detection," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.3] F. Amoozegar, A. Notash, and H. Pang, "Survey of Fuzzy Logic and Neural Network Technology for Multi-target Tracking," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.4] M. Barzohar and D. Cooper, "Automatic Finding of Main Roads in Aerial Images by Using Geometric-stochastic Models and Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(7), 1996.

[18.5] J. Beveridge, D. Panda, and T. Yachik, "November 1993 Fort Carson RSTA Data Collection Final Report," Colorado State University Technical Report, CS-94-118, August, 1994.

[18.6] J. Bevington and K. Siejko, "Ladar Sensor Modeling and Image Synthesis for ATR Algorithm Development," *Automatic Target Recognition VI, Proceedings* SPIE, **2756**, April 1996.

[18.7] B. Bhanu, "Automatic Target Recognition: State of the Art Survey," *IEEE Transactions on Aerospace and Electronic Systems*, **22**(4), 1986.

[18.8] B. Bhanu and T. Jones, "Image Understanding Research for Automatic Target Recognition," *Proceedings of the DARPA Image Understanding Workshop*, 1992.

[18.9] B. Bhanu and T. Jones, "Image Understanding Research for Automatic Target Recognition," *IEEE Aerospace and Electronics Systems Magazine*, **8**(10), 1993.

[18.10] M. Boshra and B. Bhanu, "Predicting Performance of Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(9), 2000.

[18.11] M. Bullock, D. Wang, S. Fairchild, and T. Patterson, "Automated Training of 3-D Morphology Algorithm for Object Recognition," *Automatic Target Recognition IV, Proceedings SPIE*, **2234**, April, 1994.

[18.12] J. Burrill, S. Wang, A. Barrow, M. Friedman, and M. Soffen, "Model-based Matching using Elliptical Features," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

[18.13] L. Carin, L. Felsen, and C. Tran, "Model-based Object Recognition by Wave-oriented Data Processing," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.14] L. Chan, N. Nasrabadi, and D. Torrieri, "Discriminative Eigen Targets for Automatic Target Recognition," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.15] B. Chen, S. Thomopoulos, and C. Lin, "Feature Estimation and Objects Extraction using Markov Random Field Modeling," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.16] S. S. Chiang and C. I. Chang, "Subpixel Detection for Hyperspectral Images Using Project Pursuit," *EUROPTO Conference on Image and Signal Processing for Remote Sensing V, Proceedings SPIE*, **3871**, pp. 107–115, September, 1999.

[18.17] C. Chun, D. Fleming, and E. Torok, "Polarization-sensitive Thermal Imaging," *Automatic Target Recognition IV, Proceedings SPIE*, **2234**, April, 1994.

[18.18] A. Copeland, G. Ravichandran and M. Trivedi, "Localized Radon Transform for Ship Wake Detection in SAR Imagery," *Automatic Target Recognition IV, Proceedings SPIE*, **2234**, April, 1994.

[18.19] A. Cowart, W. Snyder, and H. Ruedger, "The Detection of Unresolved Targets Using the Hough Transform," *Computer Graphics and Image Processing*, December, 1982.

[18.20] X. Du, S. Ahalt, and B. Stribling, "Estimating and Refining Orientation Vectors for Sub-components of Space Object Imagery," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.21] Dudgeon D.E. and Lacoss, R.T. "An Overview of Automatic Target Recognition," *MIT Lincoln Laboratory Journal*, **6**(1), pp. 2–10, Spring, 1993.

[18.22] D. Ernst, H. Gross, D. Stricker, and U. Thönnesen, "Improvement of Object Classification in Image Sequences," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.23] D. Forsyth, "A Novel Color Constancy Algorithm," *International Journal of Computer Vision*, **5**, 1990.

[18.24] B. Funt and G. Finlayson, "Color Constant Color Indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5), 1995.

[18.25] L. Gagnon and R. Klepko, "Hierarchical Classifier Design for Airborne SAR Images of Ships," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.26] D. Geman and B. Jedynak, "An Active Testing Model for Tracking Roads in Satellite Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(1), 1996.

[18.27] R. Ghamasaee, F. Amoozegar, H. Pang, Y. Chin, and S. Blackman, "Survey of Neu-

ral Networks as Applied to Target Tracking," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.28] Z. Haddad and S. Simanca, "Filtering Image Records Using Wavelets and the Zakai Equation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(11), 1995.

[18.29] M. Hamilton and T. Kipp, "Model-based Multi-sensor Automatic Target Identification for FLIR Fused with MMW," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.30] D. J. Hand, *Discrimination and Classification*, New York, Wiley, 1981.

[18.31] H. Hayes, C. Priebe, G. Rogers, D. Marchette, J. Solka, and R. Lorey, "Improved Texture Discrimination and Image Segmentation with Boundary Incorporation," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.32] Y. Huang and C. Suen, "A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 1995.

[18.33] R. L. Huguenin and J. L. Jones, "Intelligent Information Extraction from Reflectance Spectra: Absorption Band Positions," *Journal of Geophysical Research*, **91**(B8), pp. 9585–9598, 1986.

[18.34] D. Huttenlocher, D. Klanderman and A. Rucklige, "Comparing Images Using the Hausdorff Distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(9), 1993.

[18.35] X. Jiang and H. Bunke, "Vision Planner for an Intelligent Multisensory Vision System," *Automatic Target Recognition IV, Proceedings SPIE*, **2234**, April, 1994.

[18.36] J. Johnson, "Analysis of Image Forming Systems," *Proceedings of the Image Intensifier Symposium*, October 6–7, 1958.

[18.37] J. Johnson and W. Lawson, "Performance Modeling Methods and Problems," *Proceedings of the IRIS Specialty Group on Imaging*, January, 1974.

[18.38] A. Katz and P. Thrift, "Generating Image Filters for Target Recognition by Genetic Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(9), 1994.

[18.39] A. Kumar, Y. Bar-Shalom, and E. Oron, "Precision Tracking Based on Segmentation with Optimal Layering for Imaging Sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 1995.

[18.40] B. Kumar, A. Mahalanobis, and A. Takessian, "Optimal Tradeoff Correlation Filters with Controlled In-plane Rotation Response for Target Recognition," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.41] A. Lanterman, M. Miller and D. Snyder, "Implementation of Jump-diffusion Algorithms for Understanding FLIR Scenes," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.42] A. Lanterman, M. Miller, and D. Snyder, "Representations of Thermodynamic Variability in the Automated Understanding of FLIR Scenes," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

[18.43] B. Li, Q. Zheng, S. Der, R. Chellappa, N. Nasrabadi, L. Chan, and L. Wang, "Experimental Evaluation of Neural, Statistical and Model-based Approaches to FLIR ATR, *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.44] J. Li and C. Kuo, "Automatic Target Shape Recognition via Deformable Wavelet

Templates," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

[18.45] Lincoln Labs, "Special Issue on Automatic Target Recognition," *Lincoln Laboratory Journal*, **6**(1), Spring, 1993.

[18.46] A. Mahalanobis and D. Kelly, "High Value Target Recognition using Correlation Filters," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

[18.47] D. Manolakis, G. Shaw, and N. Keshava, "Comparative Analysis of Hyperspectral Adaptive Matched Filter Detectors," *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, Proceedings SPIE*, **4049**, pp. 2–17, 2000.

[18.48] N. Merlet and J. Zerubia, "New Prospects in Line Detection by Dynamic Programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(4), 1996.

[18.49] V. Mirelli, D. Nguyen, and N. Nasrabadi, "Target Recognition For FLIR Imagery Using Learning Vector Quantization And Multi-layer Perception," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.50] S. Musman, D. Kerr, and C. Bachmann, "Automatic Recognition of ISAR Ship Images," *IEEE Transactions on Aerospace Electronic Systems*, **32**(4), pp. 1392–1404, 1996.

[18.51] Y. Nakano, Y. Hara, J. Saito, and Y. Inasawa, "Radar Target Recognition System using 3-D Mathematical Model," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.52] E. Natonek and C. Baur, "Model based 3-D Object Recognition using Intensity and Range Images," *Automatic Target Recognition IV, Proceedings SPIE*, **2234**, April, 1994.

[18.53] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society London Series A*, **168**, pp. 268–282, 1933.

[18.54] C. Paiva, "Theater Targets Plume Edge Extraction and Hardbody Aimpoint Selection," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.55] T. Patterson, "Radiometrically Correct Sharpening of Multispectral Images using a Panchromatic Image," *Automatic Target Recognition IV, Proceedings SPIE*, **2234**, April, 1994.

[18.56] A. Pears and E. Pissaloux, "Using Hardware Assisted Geometric Hashing for High Speed Target Acquisition and Guidance," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.57] J. Perez-Jacome, "Automatic Target Recognition Systems with Emphasis on Model-Based Approaches," CSIP TR-97-01, ECE-Georgia Tech., 1996.

[18.58] Q. Pham and M. Smith, "A Morphological Multistage Algorithm for Recognition of Targets in FLIR Data," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

[18.59] Q. Pham, T. Brosnan, and M. Smith, "Sequential Digital Filters for Fast Detection of Targets in FLIR Image Data," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.60] Q. Pham, T. Brosnan, M. Smith, and R. Mersereau, "A Morphological Technique for Clutter Suppression in ATR," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.61] M. Phillips and R. Sims, "A Signal to Clutter Measure for ATR Performance Com-

parison," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.62] K. Plataniotis and A. Venetsanopoulos, *Color Image Processing and Applications*, Berlin, Springer, 2000.

[18.63] H. Qi, W. E. Snyder, and D. Marchette, "An Efficient Approach to Segmenting Man-made Targets from Unmanned Aerial Vehicle Imagery," *Optical Engineering*, **39**(5), pp. 1267–1274, 2000.

[18.64] H. Ranganath and R. Sims, "Self Partitioning Neural Networks for Target Recognition," *Automatic Target Recognition IV, Proceedings SPIE*, **2234**, April, 1994.

[18.65] K. Rao, "Combinatorics Reduction for Target Recognition in ATR Applications," *Automatic Object Recognition II, Proceedings SPIE*, **1700**, 1992.

[18.66] J. Ratches, C. Walters, R. Buser, and B. Guenther, "Aided and Automatic Target Recognition Based upon Sensory Inputs from Image Forming Systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(9), Sept. 1997.

[18.67] A. Reno, D. Gillies, and D. Booth, "Deformable Models for Object Recognition in Aerial Images," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, 1998.

[18.68] E.M. Riseman and M.A. Arbib, "Computational Techniques in Visual Segmentation of Static Scenes," *CCGIP, 7, Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.69] R. Robmann and H. Bunke, "Towards Robust Edge Extraction – a Fusion Based Approach using Greylevel and Range Images," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.70] F.A. Sadjadi, "A Model-Based Technique for Recognizing Targets by Using Millimeter Wave Radar Signatures," *International Journal of Infrared and Millimeter Waves*, **10**(3), 337–342, 1989.

[18.71] F.A. Sadjadi, "Automatic Object Recognition: Critical Issues and Current Approaches," *Proceedings SPIE*, **1471**, 1991.

[18.72] F.A. Sadjadi, "Automatic Object Recognition: Critical Issues and Current Approaches," 1991. Selected SPIE Papers on CD-ROM, Volume 6. Automatic Target Recognition SPIE: 1 PO Box 10, Bellingham, WA 98227-0010, USA.

[18.73] F.A. Sadjadi, "Automatic Object Recognition: Critical Issues and Current Approaches," 1991. Selected SPIE Papers on CD-ROM, Volume 6. Automatic Target Recognition SPIE: 1 PO Box 10, Bellingham, WA 98227-0010, USA.

[18.74] F.A. Sadjadi, Special Section on ATR, *Optical Engineering*, **31**(12), 1992.

[18.75] F.A. Sadjadi, "Application of Genetic Algorithm for Automatic Recognition of Partially Occluded Objects," *Automatic Object Recognition IV, Proceedings SPIE*, **2234**, 1994.

[18.76] R. Samy and J. Bonnet, "Robust and Incremental Active Contour Models for Objects Tracking," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.77] R. Sharma and N. Subotic, "Construction of Hybrid Templates from Collected and Simulated Data for SAR ATR Algorithms," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.78] R. Sims, "Signal to Clutter Measurement and ATR Performance," *Automatic Target Recognition VIII, Proceedings SPIE*, **3371**, April, 1998.

[18.79] R. Sims and B. Dasarathy, "Automatic Target Recognition using a Passive Multisensor Suite," Special Section on ATR, *Optical Engineering*, **31**(12), 1992.

[18.80] A. Srivastava, B. Thomasson, and R. Sims, "A Regression Model for Prediction of IR Images," *Proceedings of SPIE Aerosense ATR XI*, Orlando, 2001.

[18.81] L.B. Stotts, E.M. Winter, L.E. Hoff, and I.S. Reed, "Clutter Rejection using Multispectral Processing," *Proceedings SPIE Signal and Data Processing of Small Targets*, **1305**, pp. 2–10, 1990.

[18.82] M. Swain and D. Ballard, "Color Indexing," *International Journal of Computer Vision*, **7**(1), 1991.

[18.83] H. Tanaka, Y. Hirakawa, and S. Kaneku, "Recognition of Distorted Patterns Using the Viterbi Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **4**(1), 1982.

[18.84] W. Thoet, T. Rainey, D. Brettle, L. Stutz, and F. Weingard, "ANVIL Neural Network Program for Three-dimensional Automatic Target Recognition," *Optical Engineering*, **31**(12), December, 1992.

[18.85] M.M. Trivedi, "Detection of Objects in High Resolution Multispectral Aerial Images," *SPIE Applications of Artificial Intelligence* II, 1985.

[18.86] M. Trivedi, "Object Detection Using Their Multispectral Characteristics," *Proceedings SPIE*, **754**, 1987.

[18.87] M. M. Trivedi and J. C. Bezdek, "Low-Level Segmentation of Aerial Images with Fuzzy Clustering," *IEEE Transactions on Systems, Man, and Cybernetics*, **16**(4), 1986.

[18.88] A. Ueltschi and H. Bunke, "Model-based Recognition of Three-dimensional Objects from Incomplete Range Data," *Automatic Target Recognition V, Proceedings SPIE*, **2485**, April, 1995.

[18.89] J. Wald, D. Krig, and T. DePersia, "ATR: Problems and Possibilities for the IU Community," *Proceedings of the DARPA Image Understanding Workshop*, IUW, 255–264, San Diego, January, 1992.

[18.90] B. Wallet, D. Marchette, and J. Solka, "A Matrix Representation for Genetic Algorithms," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

[18.91] B. Wallet, D. Marchette, and J. Solka, "Using Genetic Algorithms to Search for Optimal Projections," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.92] S. Wang, G. Chen, D. Sapounas, H. Shi, and R. Peer, "Development of Gazing Algorithms for Tracking Oriented Recognition," *Automatic Target Recognition VII, Proceedings SPIE*, **3069**, April, 1997.

[18.93] A. Weisberg, M. Najarian, B. Borowski, J. Lisowski, and B. Miller, "Spectral Angle Automatic Cluster Routine (SAALT): An Unsupervised Multispectral Clustering Algorithm," *Proceedings of IEEE Aerospace Conference*, 307–317, 1999.

[18.94] D. Xue, Y. Zhu, and G. Zhu, "Recognition of Low-contrast FLIR Tank Object Based on Multiscale Fractal Character Vector," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

[18.95] C. Zhou, G. Zhang, and J. Peng, "Performance Modeling Based Adaptive Target Tracking in Multiscenario Environment," *Automatic Target Recognition VI, Proceedings SPIE*, **2756**, April, 1996.

411 ~ 416

索　引

注：索引中的页码为英文原书页码，即本书页边所标页码。

3D analysis（3维分析），261
4-connected（4连通），53
8- connected（8连通），53

A

accepting state（接受状态），373
accumulator arrays（累加器数组），278
active contours（主动轮廓），98, 322
active testing（主动测试），387, 412
active vision（主动视觉），385
affine（仿射），196
affine transformation（仿射变换），218
aircraft（飞机），407
albedo（反射率），244
algebraic distance（代数距离），202, 203, 237
algebraic invariants（代数不变量），389
algorithm fusion （算法级融合），405
algorithms, clustering（算法, 聚类），359
algorithms, performance of（算法,性能），395
angle density（角密度），141
anisotropic diffusion（各向异性传播），133
annealing（退火），121, 127,133, 206
annealing schedule（退火程序），122
annealing, tree（退火树），184
anti-extensive property（抗外延性质），149, 151
apparent error rate（视在错误率），398
arc length（弧长），196
array processor（数组处理器），258
aspect（外观），385
aspect equivalent（外观等价），296
aspect graphs（外观图），295
aspect ratio（长宽比），226, 304
assembly（装配），384
atoms（数据项），291
ATR（自动目标识别），367, 392
ATR, performance of（自动目标识别性能），395
attention, focus of（注意力中心），387
aura（流域），53
autonomous exploration（自主探测），385
axes, principal（主轴），235
axis of symmetry（对称轴），226

B

BAD（带偏置各向异性传播），137
bar code（条形码），105
basis（基），11, 59, 152
basis functions（基函数），75
basis set（基集），220
basis vectors（基向量），75, 220
Bayes' classifier（贝叶斯分类器），337
Bayes' risk（贝叶斯风险），338
Bayes' rule（贝叶斯公式），23, 329
between-class scatter（类间离散度），348
Bhattacharyya distance（巴塔恰里亚距离），358, 366
biased anisotropic diffusion（带偏置各向异性传播），137
binary（二进制），38
binary hypothesis（二值假设），396
binary morphology（二值形态学），144
biquadratic（二元二次方程），39, 93
black body radiation（黑体辐射率），248
blanking（视频消隐区），45
block thresholding（分块阈值化），183
blueprint,reading of（蓝图），384
blur（模糊），109
bootstrapping（自举），204
boundary conditions（边界条件），74
brightness（亮度），38

C

camera calibration（摄像机标定），48, 64
camera motion（摄像机运动），386
Canny edge detector（Canny边缘检测器），97
cardinality（数目），294
Cartesian（笛卡尔坐标），58
causality（因果性），87

CCD（电荷耦合器件），42,43
CCD camera calibration（CCD摄像机标定），63,257
center of gravity（重心），225
central axis（中轴），131
centroid tracking（(目标）中心跟踪），401
cepstrum（倒频谱），61
chain code（链码），159,230,369,376,377,409
chain code histogram（链码直方图），410
chaining（连锁反应），192
chamfer map（切削映像），165
chamfer-based closing（基于切削的闭运算），165
character recognition（字符识别），322,382,389
characteristic equation（特征方程），15
characterization（特征描述），393
charge-coupled device（电荷耦合器件），42
Chernoff distance（Chernoff距离），358,366
chip（目标图片），393
circles,finding（检测圆），280
circulant matrix（循环位移矩阵），74
circular arcs（圆弧），215
city block distance（街区距离），357
classification（分类），392
classification rate（分类正确率），393
classifier, inner product（内积分类器），342
classifier, linear（线性分类器），341
classifier, maximum likelihood（最大似然分类器），327, 331
classifier, quadratic（二次分类器），340
clique（完全子图），291
closing（闭运算），149
cluster（聚类），328
clustering（聚类），210,288,329,359
clustering, agglomerative（分级合并聚类），359,366
clutter（杂物），393,404
clutter, cultural（人造杂物），393
coffee transform（耗时长的变换），279
color constancy（颜色恒常性），406
color indexing（颜色索引），382
color information（彩色信息），387
color invariants（颜色不变量），382
commutative property（交换性），148
compactness（紧致性），225
competitive learning（竞争学习），215
composite（复合视频信号），44
computational complexity（计算复杂性），279
conditional probability（条件概率），8
conditional risk（条件风险），336,338
conductivity（传导率），131
cone（锥体），202
conformable（可相乘（矩阵）），12
confuser（伪装物），404
conic fitting（二次曲线拟合），212
conic section（二次曲线），203
conjugate gradient（共轭梯度），127,270,273
connected（连通的），291
connected component（连通成分），167,186,290
connected component labeling（连通成分标记），216
connected regions（连通区域），181
connectivity paradox（连通性疑题），54
consistency（一致性），92,267,286
consistent labeling（一致性标记），266
constructive solid geometry（构造立体几何），293
context-free（上下文无关），372,381
context-sensitive（上下文有关），371
continuation method（连续延拓法），116
contouring(（伪）轮廓化），47
contrast reversal（对比度反相），400
convex discrepancy（凸（包）偏差），229
convex hull（凸包），229
convolution（卷积），50,74
convolution and correlation（卷积与相关），67
correlation filters（相关滤波器），413
correspondence（对应），250
correspondence problem（对应问题），48,286,312,321
covariance（协方差），79,106,219,224,226
covariance techniques（协方差技术），211
critical point（关键点），215
curvature（曲率），60
curvature estimates（曲率估计），204
curvature-based segmentation（基于曲率的分割），204
curve code（曲线编码），377
curve evolution（曲线演变），197
curve segmentation（曲线分割），212
curves（曲线），196
cylinders, generalized（广义柱体），238

D

decision boundary（决策边界），327
decision rule（决策规则），326,330,338
decision rules, linear（线性决策规则），327
decomposition（分解），179
deconvolution（解卷积），109
deformable contour（变形轮廓），201
deformable models（变形模型），321
deformable template（变形模板），233,312,388,405
degeneracy, geometric（几何退火），388

degree（结点的边数），291
delta function（脉冲函数），121
density（密度），39
density, conditional probability（条件概率密度），331
depolarization（去极化），373
derivation（推理），370
derivative estimation（导数估计），69
derivatives（导数），67, 68
detection（检测），392
detection rate（检出率），393
diagrams, Voronoi（Voronoi图），383
diameter（直径），225, 227
differentiation（求导），123
differentiation filters（微分滤波器），106
diffusion（扩散），402
diffusion equation（扩散方程），131
digital library（数字图书馆），318
dilation（膨胀），144
dimensionality（维数），399
Dirac delta（脉冲函数），121
directrix（准线），283
discriminant function（判别函数），329, 336
disparity（视差），64, 286
distance transform（距离变换），153, 165, 179, 199, 232
distance transform, Euclidian（欧几里德距离变换），154
distance, affine-invariant（仿射不变的距离），261
distance, Mahalanobis（马哈拉诺比斯距离），341
distances,Euclidian（欧几里德距离），219
distortion（畸变），49, 108
dithering（颤振），213
divergence（散度），358
duality（对偶），149
dynamic programming（动态规划），198
dynamic range（动态范围），46

E

ECG（心电图），373
edge detection（边缘检测），76, 97
edge extension using relaxation（用松弛法进行边缘扩展），99
edge strength（边缘强度），69
edges（边），290
edges in range images, determining（确定深度图像中的边），243
edges of roads（道路边缘），386
ego-motion（自运动），141
eigenimage（特征图像），300
eigenvalues（特征值），15, 349
eigenvector（特征向量），15, 222
ejection fraction（喷出率），383
electrocardiogram（心电图），373
electron-hole（电子-空穴），43
ellipse, direct fitting（椭圆直接拟合），203
ellipse, finding（椭圆检测），288
ellipses, fitting（椭圆拟合），203
ellipses, least square fitting（椭圆的最小二乘拟合），212
ellipsoid（椭球），202, 235
ellipsoids, fitting（椭球拟合），203
ellipse, finding（椭圆检测），285
emissivity（发射率），400
epipolar line（外极线），48
equivalence memory（等价存储器），190
equivalence relation（等价关系），190
erosion（腐蚀），146
error rate（错误率），398
Euclidian distance（欧几里德距离），219, 304, 356
Euclidian distance transform（欧几里德距离变换），154
events（事件），8
explicit（显式），202
explicit representation（显式表达），39, 71
extensive property（扩展性质），148
external energy（外能量），198

F

face recognition（人脸识别），346, 384
facet model（面片模型），97
factor（因子），159
false alarm（误警），396
false alarm rate（误警率），393
false negative（漏检（假阴）），396
false positive（误报（假阳）），396
false positive fraction（误报率（假阳率）），397
fast Fourier transform（快速傅里叶变换），75, 84
feature（特征），304
feature selection（特征选择），324
feature vector（特征向量），304
FIDAC（参考文献[16.9]标题中的一个缩写），381
filters（滤波器），41
finite element models（有限元方法），390
finite state machine（有限状态机），22, 370, 373
Fisher's linear discriminant（费希尔线性判别），348
fitting（拟合），69, 73
fitting surfaces（表面拟合），202
fitting, straight lines（直线拟合），223
fitting, subpixel precision of（分像素拟合精度），225
FLIR（中长波红外图像），394

focus, of a parabola（抛物线焦点），284
focused filters（聚焦滤波器），407
formal language（形式语言），369
Fourier descriptors（傅里叶描述子），231, 262
Fourier transform（傅里叶变换），41, 83, 207
fractal dimension（分形维），207, 212
fractal dimension for gray-scale（灰度图的分形维），209
frame time（一帧时间），43
frequency response（频率响应），41
FSM （有限状态机），23
function fitting（函数拟合），69
functional（函数的），39
furthest neighbor distance（最远邻距离），358
fusion（融合），405

G

Gabor function（Gabor函数），90
Gauss curvature（高斯曲率），61
Gauss map（高斯图），286
Gaussian（高斯分布），78, 88, 279
Gaussian, mean of（高斯分布均值），333
Gaussian, multivariate（多元高斯分布），335, 358
generalized Hough transform（广义Hough变换），282
genetic algorithm（遗传算法），214, 403
genetic learning（遗传学习），388, 413
geometric flow（几何流），234
geometry, curve（几何曲线），234
geometry-based（基于几何的），404
GNC（分级非凸性），129
gradient（梯度），13, 97, 279
gradient descent（梯度下降法），16, 18, 127,206,270
gradient magnitude（梯度幅值），78
graduated nonconvexity（逐级非凸性），116, 129
grammar（文法），370
graph, directed（有向图），290
graphs, aspect（外观图），238
gray-scale（灰度），38
gray-scale morphology（灰度形态学），152, 179
Green's function（格林函数），132, 133
grid（网格），67
ground plane（地面），386
ground truth（实际真值），396

H

handwrriten word recognition（手写字识别），380
harmonic（谐波），236
Hausdrorff distance（豪斯多夫距离），410
heart（心脏），383
heat equation（热流方程），132
Hessian（黑塞矩阵），14, 135
hexagonal（六角形），57
hexagonal derivative（六近邻求导），71
hidden Markov model（隐马尔科夫模型），22, 62, 381
histogram（直方图），184,330
histogram intersection（直方图求交），406
homogeneity.predicate（均质谓词），181
homogeneous（齐次，同质），181
homogeneous transformation（齐次变换），217
homotopy（同伦），116
Hopfield neural network （Hopfield神经元网络），137
horizontal blanking（水平消隐），45
horizontal resolution（水平分辨率），46
Hough transform （Hough变换），275, 288, 356, 407
Hough transform, generalized（广义Hough变换），282
hull, visual（视野），258
human movement（人的运动），256
hyperboloid（双曲面），235
hyperellipsoid（超椭球体），223
hyperquadric（超二次曲面），237

I

iconic（图像的, 图标），38
idempotent property（等幂性），151
identification（辨别，鉴别），393
IFF（辨别敌友），394
IFS（图像文件系统），29
ill-posed（病态的），108
image formation（图像生成），43, 404
image hashing（图像散列）318
image representations（图像表示），38
imaging, infrared（红外成像的），383
implicit（隐式），202
implicit representation（隐式表达），39
increasing property（增长性），148, 151
independent（独立），8
indexing（检索），256
infrared imagery（红外图像），389
inliers（内点），204
inner product classifier（内积分类器），342
inside a curve（曲线内部），197
inside/outside（内部/外部），40
inspection of PC boards（PC板检验），150
inspection/quality control（检验/质量控制），383
integrating several range images（数个距离图像整合），243
intensity gradient（强度梯度），71

internal energy（内部能量），198
interpolation（内插），114
interpretive tree（解释树），387
intersections of edges（边的交），243
invariance（不变量），239
invariant（不变性），203
invariant moments（不变矩），304
invariant shape models（不变形状模型），261
invariants, geometric（几何不变的），318
invariants, thermophysical（热物理学不变性），389
inverse problem（逆问题），108, 114, 404
Ising type（Ising类型），137
ISODATA（ISODATA聚类算法），362
isoelectric（零电位），373
isophote（等亮度线），51, 201

J

Jacobian（雅可比），14
joint probability（联合概率），8
junction matching（结点匹配），325

K

Kalman Bucy filter（卡尔曼-布西滤波器），401
kernel（核），66
K-L transform（K-L变换），220, 404
k-means（*k*-均值聚类方法），362
k-neighbors（*k*近邻），165
Kroneker delta（delta函数），121, 122
DT distance（距离变换距离），165
Kuhn-Tucker conditions（库恩-塔克条件），352

L

label image（标号图），181
label memory（标号存储器），186
labeling（标号），264
labels（标号），263
Lagrange multiplier（拉格朗日乘子），224
language（语言），371
Laplace's equation（拉普拉斯等式），120, 236
Laplacian（拉普拉斯算子），82
Laplacian of Gaussian（拉普拉斯高斯），88
learning（学习），324
learning shape（形状学习），239
leave-one-out approach（交叉检验（留-检验法）），399
left ventricle（左心室），256
Legendre polynomial（勒让德多项式），236
level set（水平集），200, 202
lexicographic（字典式的），74
library, digital（数字图书馆），318
light sources, multiple（多光源），249
likelihood ratio（似然比），335
likelihood ratio test（似然比测试），340
likelihood, of views（视图的可能性），385
line drawing（线条图），263
line segments, motion from（从运动分析线段），250
linear（线性），40
linear classifier（线性分类器），341
linear machines（线性机），327
linear operator（线性算子），65, 74, 78
linear relaxation（线性松弛法），267
linear transformation（线性变换），12
linearly independent（线性独立），11
linearly separable（线性可分），327
linked list（链表），291
LISP（LISP语言），291
local/global（局部/全局），275
local/global inference problem（局部/全局推理问题），263
LOG（拉普拉斯高斯），88
logarithmic spiral grids（对数螺旋网络），261
log-polar transformation（对数极坐标变换），239
low-pass（低通），78, 98, 106
luminance（亮度），38
LWIR（长波红外），395

M

magnetic resonance imaging（核磁共振图像），39
Mahalanobis distance（马哈拉诺比斯距离），341, 349, 357
major axis（主轴），223
MAP（最大后验概率），121, 198, 201
MAP estimation（最大后验概率估计），407
maps, Kohonen（Kohonen映射），259
margin（类间距），349
Markov（马尔科夫模型），201
Markov chain（马尔科夫链），20
Markov model（马尔科夫模型），20
Markov model, hidden（隐马尔科夫模型），22
Markov random field（马尔科夫随机场），64, 291, 381
maximum likelihood（最大似然），327
mean curvature（平均曲率），61
mean field approximation（均质场近似），116
measurement system（度量系统），49
medial axis（中轴），53, 232
median filter（中轴滤波），151

Mercer's conditions（Mercer条件），353
metric（距离），219, 358
metric function（距离函数），200
microscopy, epi-fluorescence（外荧光显微镜），384
minimax rule（最小最大规则），342
minimization（最小化），15
minimum description length（最小描述长度），321
minimum spanning tree（最小生成树），360
minimum-squared-error（最小平方误差），239
MMSE（最小垂直距离），204
model（模型），304
model-based（基于模型的），405
modeling edges（模型边缘），105
molecules, tubular（管状分子），384
moments（矩），61, 224, 230
moments, central（中心矩），229
morphological filtering（数学形态学滤波器），179
morphological sampling（数学形态学采样），162
morphological shape decomposition（数学形态学分解），173
morphological skeleton（数学形态学骨架），171
morphology（数学形态学），144, 408, 409
motion（运动），322, 386
motion analysis（运动分析），115, 250
motion segmentation（运动分割），210, 250
MRI（核磁共振图像），39, 383
MSE estimate（最小平方误差估计），69
multiple light sources（多光源），384
multisensor fusion（多传感器融合），405
multispectral（多光谱），382
multispectral ATR（多谱自动目标识别），406
multispectral matching（多谱匹配），406
multivariate Gaussian（多元高斯分布），79
multivariate pdf（多元概率密度函数），332
multiview range data（多视点深度图），211
MWIR（中波红外），395

N

nearest feature line（最近特征线），324
nearest neighbor measure（最近邻度量），358
neighborhood（邻域），53
neighboring state（邻域状态），24
neural network（神经元网络），137, 403
neuron（神经元），137
Newton-Raphson（牛顿-拉弗森方法），17
Neyman-Pearson（Neyman-Pearson方法），401
nodes（结点），290, 291, 305
noise（噪声），49
noise estimation（噪声估计），63, 257
noise sensitivity（噪声敏感度），68
noise, counting（噪声统计），204
noise, Gaussian（高斯噪声），204
noise, non-Gaussian（非高斯噪声），204
noise, Poisson（泊松噪声），204
nonlinear relaxation（非线性松弛法），267
nonmaximum suppression（非最大值抑制），97
nonterminal symbol（非终止符），369
normal vector to an edge（边的法线向量），133
normalization（规范化），309
normalized central moments（规范化中心矩），229
NP-complete（NP完全），294
NP-hard（NP复杂度指数增长的问题），294
nuclear medicine（核医学），204, 383

O

objective function（目标函数），15, 115, 126
objects（物体），263
occlusions（遮挡），401
opening（开运算），149
optic flow（光流），120, 250, 256, 313, 321, 386, 387
optimal structuring elements（最佳结构元素），180
optimization（最优化），405
orthogonal（正交），11
orthogonal functions（正交函数），11
orthogonality（正交性），243
orthonormal（正交归一），93
orthonormal transformation（正交归一变换），13
outliers（外点），204
outside a curve（曲线外部），197
outward normal direction（外指法线方向），197
oversegmentation（过(度)分割），205, 306

P

P wave（P波），373
panoramic camera（全景摄像机），48
panoramic stereo（全景体视），48
parabola, finding（抛物线检测），283
parameters, model（模型参数），25
parametric transformation（参数变换），275
partial differential equation（偏微分方程），131
parts manufacturing（零部件制造），293
parts, overlapping（重叠零件），384
path（路径），24, 291
pattern classifier（模式分类），304
pattern recognition（模式识别），304
PC board（PC板），149

PCA（主分量分析，主成分分析），301
PDE（偏微分方程），132
peak, finding（峰值检测），285
perimeter（周长），225
perspective（透视投影），62
photometric stereo（光度立体视觉），245
piecewise-constant（分段常数），119,120,135
piecewise-linear（分段线性），135
piecewise-linear fitting（分段线性拟合），224
piecewise-planar（分段平面），120
pixels（像素），38
planar approximations（平面近似），120
plane, fitting （平面拟合），70
planning（规划），322, 383
polarization（偏振），49
polyhedron（多面体），292, 295
polynomials（多项式），39, 202
pose（位姿），48, 62
positive definite（正定），13
power spectrum（功率谱），61
predicate（谓词），306
prime factor（质元素），159
principal axis（主轴），219, 227
principal component（主分量），214, 220
principal component analysis（主分量分析，主成分分析），301
principal curvatures（主曲率），60
prior probabilities（先验概率），396
probabilistic（概率），40
probability distribution（概率分布），9
probability of error（错误概率），340
probability, prior（先验概率），329
probability, transition（转移概率），22
production（产生式），370
projection（投影），12, 75, 76, 220
projective invariants（投影不变性），322
pseudo-inverse（伪逆），239
puns, bad（双关语），383
pushdown stack（下压堆栈），186
pyramid ,（金字塔），85,402

Q

quad tree（四叉树），86
quadratic（二次函数），202
quadratic classifier（二阶分类器），340
quadratic form（二次型），13
quadratic variation（二阶变分），83
quadric（二次(隐式)函数），39, 202, 234
quality, segmentation（分割质量），205
quantization（量化），46
quantization error（量化误差），47
quantized（量化的），42

R

radar（雷达），386
radial basis function（径向基函数），353
radiometric（辐射度量的），63, 257
radius of a convolution kernel（卷积核半径），71
ramp edge（坡边），76
range（距离），39, 47
range image segmentation（深度图像分割），211, 215
range images（深度图像），201, 267
range images, merging of（深度图像的拼合），204
range images, registration of（深度图像对准），204
raster scanning（光栅扫描），43
receiver operating characteristic（受试者操作特性），397
recognition（识别），393
recognizer（识别器），371
recurrent network（递归网络），137
recursive region growing（递推区域生长），186
reflection（反射），226
reflection of a set（反对称集），146
reflectivity（反射率），245
region adjacency graph（区域邻接图），292
region alignment（区域对齐），141
region growing（区域生长），182, 322, 402
registration（对准），204
regular expression（正则表达式），376
relabeling（重标号），167
relational representations（关系表示），42
relaxation（松弛），107
relaxation labeling（松弛标号），266, 305, 313, 321
restoration（复原），108
reverse engineering（逆向工程），384
ridge seeking（脊搜索），63
ridges（脊），52
roads, finding in images（在图像中检测道路），386
ROC（接收器操作特征），397
roof edge（屋顶型边缘），76
rotation（旋转），217, 218
rotation, correction for（旋转校正），218

S

salient（特征），196
salient group（特征集），210
sample mean（采样均值），333, 334

sample variance（采样方差），334
sampled（采样的），42
sampling（采样），45
sampling grid（采样网格），161
sampling theorem（采样定理），46, 161
sampling time（采样时间），46
SAR（合成孔径雷达），408
satellite images（卫星图像），407
scale space（尺度空间），85
scale space causality（尺度空间因果关系），87
scaling（比例缩放），358
scanning（扫描），43, 192
scatter matrix（散布矩阵），70
scene graph（场景图），293
secondary electrons（二次(发射)电子），247
segmentation（分割），115, 120, 181, 322, 402
segmentation of surfaces（表面分割），201
segmentation, based on texture（基于纹理分割），207
segmentation, psychology of（按心理学的分割），211
segmentation, quality of（分割质量），204
segmentation, range（深度图分割），388
self-organizing（自组织），368
sensitivity（敏感程度），397
sensor fusion（传感器融合），115, 405
separability（可分性），350
set of measure zero（度量为零的集），119
shape from（从…分析形状），243
shape from focus（利用聚焦分析形状），249
shape from shading（从影调分析），244, 322
shape from texture（基于纹理形状分析），249
sharpness（锐利度），41
shear（剪切），218
sigmoid（S形函数），137
signal-to-clutter（信号与杂物），402
signature（标志），400
silhouettes（轮廓），243, 259, 407
similarity transformations（相似变换），217
simple points（简单点），180
simplicial models（纯形模型），261
simply connected（单连通），159
simulated annealing（模拟退火），19, 115, 122, 198
singular-valued decomposition（奇异值分析），252
skeleton（骨架），171
slope-intercept（斜率-截距形式），275
smoothing（平滑），386
snakes（蛇形线），197
Sobel（Sobel边缘检测算子），78, 94
Sobel operator（Sobel算子），68
sonar（太阳能），386
space-invariant（空间不变），50
spatial（空域），49
spatial derivatives（空域导数），67
spatial frequency（空域频率），40, 61
spatial vision（空域视觉），106
specificity（确切程度（非目标正确判别率）），397
spectral bands（谱带），394
spectral-based（基于光谱的），404
specular（镜面反射），49, 243, 260
specular reflectors（镜面反射），49, 384
speech recognition（语音识别），26
speed of a curve（曲线上点坐标随参数的速率变化），196
sphere（球），202
spherical harmonics（球形谐波），236
springs and templates（(脸的)弹性模板），308
squared residual（残差平方），
stability, of views（视图的稳定性），385
stable state（稳定状态），385
state sequence（状态序列），23
statistical mechanics（统计力学），115
statistical pattern classification（统计模式识别），305
step（阶跃(边缘)），76
stereo（立体视觉），64, 120, 322
stereo matching（体视匹配），324
stereo, patch-based（基于片的立体视觉），142
stereopsis（立体视觉），47, 286, 300, 322, 386
stick growing（线段增长），106
stochastic grammars（随机文法），370
stochastic network（随机网络），381
straight line, fitting of（直线拟合），223
striate cortex（条纹皮层），105, 260
structure from motion（从运动分析结构），64, 260
structured illumination（结构光照明），48, 248
structured lighting（结构光），48
sum of squared differences（平方差总和），286
superquadric（超二次曲线），237, 238
supervised learning（有监督学习），328
support vector machine（支持向量机），349
surface geometry（表面几何），211
surface harmonic（表面的谐波表示），236
surface normal（表面法线），244, 247
surfaces, quadric（二次曲线），234
SWIR（短波红外），395
symmetry（对称），226, 243
symmetry distance（对称距离），227
symmetry, mirror（镜面对称），226

symmetry, rotational（旋转对称），226
sync（同步脉冲），45
syntactic pattern recognition（句法模式识别），369

T

target detection（目标控制），395
template matching（模板匹配），179, 299, 313, 405
templates（模板），404
temporal templates（时间模板），256
terminal symbol（终止符），369
tessellation（棋盘形布局），57
test set（测试集），399
texture （纹理），61, 88, 261, 379
texture discrimination（纹理分类），63
texture segmentation（纹理分割），207, 212
thermal imaging（热成像），383
thermophysical invariants（热物理不变量），389
thinness（细微度），225
thinning（细化），97, 171
threshold（阈值），182
thresholding（阈值化），184
tomography（层析X射线摄影机），180
tracking（跟踪），250, 269, 401
tracking of molecules（分子跟踪），384
tracking roads （道路跟踪），407
tracking using deformable templates（利用变形模板进行跟踪），233
training set（训练集），304, 326, 328, 348, 399
transform,（K-L变换），K-L 219
transformations, orthogonal（正交变换），216
transition matrix（转移矩阵），22
transition probability（转移概率），22
translation（位移），145
translation of a set（(结构)集的位移），146
tree（树），291
tree annealing（树退火算法），184
triangle similarity（三角形相似度），226
true error rate（真实错误率），398
true negative（正确辨识非目标（真阴）），396
true positive（正确辨识目标（真阳）），396
true positive fraction（目标正确辨识率（真阳率）），397
tubular molecules（管状分子），384
Turing machine（图灵机），370

U

umbra（本影），152
uncertainty propagation（不确定性传播），325
undersegmentation（欠分割），205
uniformity（一致性），213
univariate pdf（一元概率密度函数），331
unsupervised learning（非监督学习（无监督学习）），328

V

vanishing points（消隐点），266, 286
variable conductance diffusion（传导率可变的扩散），131
variance, estimation of（方差估计），334
VCD（变传导系数扩散），131, 136
vector representations（向量表示），73
vector space（向量空间），12
vector, infinite-dimensional（无限维向量），11
vehicle tracking（车辆跟踪），389
ventricle（心室），373
ventricle, left（左心室），383
vertex, multihedral（多面体相交顶点），243
vertex, trihedral（三面相交顶点），243
vertical blanking（场回扫），45
vertical resolution（垂直分辨率），46
vertices（顶点），290
vidicons（摄像管），42
view likelihood（视图可能性），385
view stability（视图稳定性），385
viewing region（视角划分），296
viewpoint space partition（视点空间划分），296
visual potential（可能视图），296
Viterbi algorithm（维特比算法），24, 408
Voronoi diagram （Voronoi图），155, 382
VSP（视点空间划分），296

W

Waldo, finding（Waldo检测），387
watershed（分水岭），182, 210, 211
wavefront propagation（波前传播），287
wavelet（小波），101, 142, 143, 400
weak membrane （弱隔膜），129
whitening（白化），219
within-class scatter（类内散布），348
Wold features（荒原特征），213

Z

Zakai equation（Zakai方程），402
zero crossings（零交叉），88, 98
zero set（零集），39
zoom（变焦），218