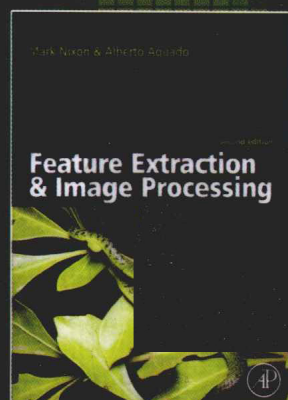


国外电子与通信教材系列

特征提取与 图像处理 (第二版)

Feature Extraction and Image Processing
Second Edition



[英] Mark S. Nixon 著
Alberto S. Aguado

实英 杨高波 译
李仁发 审校



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

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

国外电子与通信教材系列

特征提取与图像处理

(第二版)

Feature Extraction and Image Processing
Second Edition

電子工業出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

图像处理和计算机视觉是当前热门的研究课题。本书以一种简洁明了的方式为计算机视觉和图像处理领域的初学者和准专业人士提供了一个图像处理和计算机视觉技术的基础指南。读者可以基于文中给出的代码以及网站的 MATLAB 和 Mathcad 文件开发自己工作中所需的技术方法。

本书不仅关注高层次特征提取技术,还涉及与图像获取、采样理论、点运算和低层次特征提取相关的问题和技术,构成了一个明确连贯的知识体系。本书在第一版的基础上,既新增加了用于低层次特征提取的曲率方法(SIFT 和显著性算子)和相位一致性、几何主动轮廓、形态学等,还升级了图像平滑处理(各向异性扩散)、形状骨架化、形状描述(矩)等内容。

本书可作为高等学校电子工程、计算机科学、计算机工程等专业本科生的教材,也可以作为图像、视频信号处理,模式识别和计算机视觉研究方向的博士、硕士研究生,以及相关专业的科研工作者的参考用书。

Feature Extraction and Image Processing, Second Edition

Mark S. Nixon and Alberto S. Aguado.

ISBN: 978-0-12-372538-7

Copyright © 2008 by Elsevier. All rights reserved.

Authorized Simplified Chinese translation edition published by the Proprietor.

ISBN: 978-981-272-159-4

Copyright © 2010 by Elsevier(Singapore)Pte Ltd. All rights reserved.

Printed in China by Publishing House of Electronics Industry under special arrangement with Elsevier(Singapore)Pte Ltd.

This edition is authorized for sale in Mainland of China. Unauthorized export of this edition is a violation of Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书简体中文字版专有出版权由 Elsevier (Singapore) Pte Ltd. 授予电子工业出版社,仅限在中国大陆出版发行。未经许可之出口,视为违反著作权法,将受法律之制裁。

本书封底贴有 Elsevier 公司防伪标签,无标签者不得销售。

版权贸易合同登记号 图字:01-2008-3438

图书在版编目(CIP)数据

特征提取与图像处理/(英)尼克松(Nixon, M. S.), (英)阿瓜多(Aguado, A. S.)著;李实英,杨高波译.

2版. —北京:电子工业出版社,2010.10

(国外电子与通信教材系列)

书名原文:Feature Extraction and Image Processing, Second Edition

ISBN 978-7-121-11878-4

I. ①特… II. ①尼… ②阿… ③李… ④杨… III. ①数字图像处理—高等学校—教材 IV. ①TN911.73

中国版本图书馆 CIP 数据核字(2010)第 184579 号

策划编辑:马 岚

责任编辑:李秦华

印 刷:

装 订:北京季蜂印刷有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本:787×1092 1/16 印张:21 字数:538 千字

印 次:2010 年 10 月第 1 次印刷

定 价:35.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlts@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

译者序

百闻不如一见。人类在认知和理解周围环境的过程中,70%以上的信息都是从眼睛获取的。这样的视觉信息在计算机视觉和模式识别等领域体现于对图像(包括视频)的分析和理解。例如,目标跟踪和异常行为识别与分析等应用的基础内容和步骤就是图像处理 and 特征提取。

本书是由英国萨里(Surrey)大学 Alberto S. Aguado 教授和南安普顿大学的 Mark S. Nixon 教授在第一版的基础上,于2007年最新改版之作。该书出版后,受到了读者的广泛好评,于2008年6月进行了重印。从总体上看,本书具有以下几个特色。

首先,本书的内容重点突出。本书在介绍计算机视觉的基础概念和基本的图像处理操作后,重点讨论了低级的特征提取及边缘检测、基于形状匹配的特征提取、基于 Snake 模型等的灵活特征提取及对象描述、纹理描述、分割及分类等内容。也就是说,它突出了计算机视觉的主要问题——特征提取,以清晰、简洁的语言,阐述了图像处理和计算机视觉中的基础理论与技术。

其次,本书的适用面广。它既适合于电子工程、计算机科学、计算机工程等专业的本科生作为教材使用,也适合于图像、视频信号处理,模式识别和计算机视觉方向的博士、硕士研究生使用,还可以作为相关专业的科研工作者参考用书。

再次,提供了部分有参考价值的代码和伪代码。特别是,提供的一些 C 语言或 MATLAB 语言的源代码,便于工程应用时参考使用,尤其适合于高校的博士、硕士研究生进行算法研究时参考使用。

当然,特征提取是图像模式识别和计算机视觉领域的研究热点,各种新算法和技术层出不穷。本书只提供了迄今为止本领域技术水平的基础内容。尽管如此,译者仍然认为,本书是一本不可多得的好教材。

全书由李实英和杨高波分工翻译。其中,第1章到第4章和第6章到第8章由李实英翻译,前言、第5章和附录部分(第9章到第12章)由杨高波翻译。全书由李仁发教授审校。

另外,赵鹏玉、肖志鹏、牛志国、尚学智、米超、张然、赵林祥、陈日起、蒋琴、陈宁、吴潇和向仁华等人参与了部分工作。

把一种语言表达转换成另一种是一件困难的事情。看似很直白的一个词虽然理解词义,要换种语言表达时往往煞费苦心。译者力求忠实地传达书中所介绍的技术手法,保持原作者的行文风格。此外,根据作者提供的勘误表,修正了原稿中的一些错误。然而,由于时间仓促以及译者水平和经验的不足,译文中难免存在许多不当之处,恳请读者提出宝贵的意见和批评。

此外,本书有4个附录,分别是“第9章”、“第10章”、“第11章”和“第12章”。为了与原文版和互联网上本书作者提供的“教辅”网页保持一致,在本书翻译过程中未做结构调整。

本书在翻译过程中,得到了许多人士的帮助。对此,译者深表感谢。

译者

2010年2月

前 言

为什么写这本书

毫无疑问, 我们被问了很多次: 究竟为什么写一本关于计算机视觉的新书? 普遍性的问题是: 既然在书店中已有很多关于计算机视觉的好书, 正如后面即将引用的那些, 因此为什么再写一本呢? 这个问题的部分答案是任何教材都是此前已经存在的材料的快照 (Snapshot)。计算机视觉, 一门处理在计算机存储图像的艺术, 已见证了高水平研究人员进行了相当多的研究, 并且其研究在近些年来将持续增长。这意味着, 许多新的技术已经被发展, 而且许多最新的方法需要反映到教材中。

但是, 本书并不是最新研究, 而是计算机视觉领域的部分快速发展, 留下了一些仍然只包含少量细节的空间。根据研究的特性, 一个人不能出版那些看上去不是推进知识则是更像填补历史空白的材料。这就是为什么一本新的教材仍然能够对计算机视觉有所贡献。

最后, 技术本身继续发展。这意味着有新的硬件、新的编程语言、新的编程环境出现。特别是计算机视觉, 技术的进步意味着计算能力、内存等现在变得相对便宜。可以肯定, 现在比起计算机视觉刚成为一个研究领域时, 要便宜得多。作者之一在这里注意到, 用来写本书的便携式计算机比他读大学整个学生阶段所使用的台式机已具有更多的内存、更快的速度、更大的磁盘空间和更好的图形能力。然而, 他本人并不年轻。近年来由于技术进步所带来的一个更有利的变化是数学编程系统 (mathematical programming system) 的进展。这些允许我们将注意力集中在数学本身, 而不是实现的细节。有几个更“尖端”的工具软件, 包括本书所选用的 Mathcad 和 MATLAB 是最流行的。我们已经在教学和科学研究中使用, 而且我们愿意确认它们具有相当多的好处。在研究工作中, 它们帮助我们更快地发展新的技术和评价它的最终实现, 对于教学工作, 现代便携式计算机和数学系统的能力在教学中联合展示给学生, 不仅包括这些技术是如何实现的, 而且包括如何以及为什么它们与传统的教学材料以明确的关系一起工作。

基于上述原因, 写作了本书。有大量的材料, 本来可以包括进来, 但我们选择了忽略。如果它们是你所喜欢的工作, 我们对此抱歉。由于计算机视觉所涉及的领域相当广, 本书仅限于计算机视觉领域的图像处理与特征提取。原因在于, 二者不但是我们研究工作的重点, 也是其他的一些同类已出版的教材所忽视的。然而, 它是应用计算机视觉的主要目标之一。我们已确定本书的目标是澄清一些它的起源和发展, 同时揭示采用数字系统的实现。正因为如此, 我们抱定上述目标撰写了本教材。

为什么再版

有许多原因, 我们更新并再版本书。首先, 尽管它是电子稿件, 第一版的许多内容在出版前是重新输入的。它引起了一些现在已经修正的错误。其次, 本领域持续发展和进步: 我们增加了一些新技术, 从写本书的第一版或者这些新技术出现后它们正受到广泛重视。一些领域的进展相对较快, 它们都反映在本书的主要改进中。再次, 得到了一些新的学术成就, 特别是以新教材的形式, 我们在新版中尽可能多地包括这些新的知识。MATLAB 和 Mathcad 仍然是

本书使用的计算工具,并增加了一个新的使用 Java 的演示站点。最后,我们保持了原始的格式。改变格式,甚至是重新对文本的格式进行排版是诱人的,但我们仍然选择不改变格式。除了勘误和澄清之外,本版与原版相比的主要变化之处如下:

- 第 1 章:升级了人眼操作(eye operation),摄像机技术和软件,升级和扩充了网络材料和文献。
- 第 2 章:很少变化(这是标准材料),除了增加反走样的一个好例子外。
- 第 3 章:引入了图像平滑的各向异性扩散,力场算子(force field operator)和数学形态学。
- 第 4 章:扩充了频域的概念和差分操作符;引入了相位一致性(phase congruency),现代曲率运算符和尺度不变特征变换(SIFT)。
- 第 5 章:强调了在遮挡、噪声条件下的特征提取以及一些运动特征的实用属性。
- 第 6 章:引入了几何自主轮廓(geometric active contour)和水平集(level set)的方法;引入了骨架化,扩充了主动形状模型(active shape model)。
- 第 7 章:扩充了关于矩,特别是 Zernike 矩的材料,引入了基于矩的重建。
- 第 8 章:澄清了一些基于特征的识别的细节。
- 附录:扩充了摄像机模型的细节和主成分分析(principle components analysis)。

如前所述,增加了一个新的基于 Java 的演示站点,网址为 http://www.ecs.soton.ac.uk/~msn/book/new_demo/。它包含了即将描述的一些技术和基于计算机视觉的生物特征如何工作的实例。这个网页将会继续被更新。

本书与支持

本书的每一章都介绍了涉及到图像处理与计算机视觉中特征提取的特定信息包。每个信息包都是在原稿的基础上,参考了更新的材料得到的。理所当然,在实现(采用 MATLAB 或 Mathcad)之前通常有理论上的进展。我们提供了绝大多数所描述的主要技术的实现,并应用它们处理一组选定的图像。尽管工作的重点在分析医学图像或者生物信息(一门通过行为或心理特征识别人科学,类似于人脸识别),这些技术都具有通用性,可以移植到其他的应用领域。

读者可以在本书的支持网站, <http://www.ecs.soton.ac.uk/~msn/book/>, 找到大量的进一步支持信息。首先,可以得到工作表(与本书配套的 MATLAB 和 Mathcad 实现),以方便学习。另外,还有课堂讲授用的电子教案。通过投影仪,它的字体更大,包括更多的交互式演示。此外,还有习题(含答案)和演示站点。站点将会尽可能地保持更新,因为它包含其他材料的链接,例如关于技巧和应用的站点,可以得到的软件和在线文献。最后,任何勘误将会及时报告。错误的存在,是我们的遗憾和责任。如果读者找到了我们不知道的错误的,请直接用网站上的 <mailto:> 给我们发邮件,我们将非常感谢。

在本书中包含了一定数量的数学内容,目标读者是电气/电子工程、软件工程、计算机科学或者数学、物理专业的四年制大学本科的三、四年级学生和硕士研究生。他们具备相应的数学分析能力。计算机视觉可以被认为是应用数学的一个分支,尽管它并不真正应用到应用数学的一些领域。实际上,涉及到的数学主要有微分和几何,尽管它可能比传统的课程讲授的内容更细一些。可以肯定,并不是这里所有的材料在南安普敦的大学课程中都被详细覆盖到。

本书从对计算机视觉的硬件、软件和已有材料(established material)的综述开始,并参考了

已发展的更为复杂的视觉系统:人类视觉系统。尽管处理本质的确切细节仍然有待深化,仍然有相当大范围的硬件和软件允许我们给予计算机系统具备捕获、处理和进行推理的能力,实现“观察”(sight)的能力。第1章还提供了关于本主题材料的综合性书目,不仅包括教材,还有可用的软件和其他材料。毫无疑问,它们将会发生变化,因此有必要访问站点以得到最新的信息。建议的期刊(Journal)参考资料是那些在当地的大学图像馆或网络上易于找到的,特别是IEEE专辑(IEEE Transaction)。由于价格相对低廉,质量很高,它们通常被订阅。

第2章涉及计算机视觉使用到的基本信号处理理论。它引入了傅里叶变换,允许以一种新的方式,从频率内容来观察一个信号。它还允许我们得到图像最小化,可同时保持信息并从频率的角度分析内容,并将有助于后续的计算机视觉算法的加速。遗憾的是,它会涉及到少量的(数学)等式,但是它是观察数据和信号的一种新方式,并已被证明是一种颇有成效的学习方式。

第3章是基本的图像处理技术,首先是图像的点处理,即考虑原始图像的每一个像素,图像的每个像素被映射到一个新值。然后考虑像素集。不仅是让图像的视觉、特别是人眼视觉效果更好的常见处理,也包括如何处理经常遇到的不同类型图像噪声的问题。这些技术都采用Mathcad和MATLAB以算法形式实现,以帮助明确地显示这些等式是如何起作用的。此外,还将介绍一些现代的方式,以去除噪声得到“干净”的图像,以及采用形状的概念,而不是映射过程处理图像。

第4章是低级的特征提取,它们是从整幅图像而不是一个特定区域(distinct region)的层次,描述图像内容的技术。最重要的过程之一是边缘检测。实际上,边缘检测将图像缩减为一种漫画式骨格化的形式,但没有漫画家的夸张表现。边缘检测的主要技术都进行了详细的介绍,以及对它们实现的描述。我们可以得到的其他图像属性包括曲率(curvature)和移动量(movement),它们也涵盖在本章中。

这些边缘、曲率或者运动需要按照某种方式被组织到一起,以便我们能够从一幅图像找到形状。形状提取的第一方法涉及到分析低级信息与已知的目标形状模板的匹配。因为这可能是一个计算上特别累赘的过程,我们拟发展一种在保持其优化性能的同时,能改善其计算性能的技术,也就是我们所熟知的Hough变换。长久以来,Hough变换是计算机视觉领域学者的一个热门目标,意在澄清它的基础、改善速度以及提供它的准确性和鲁棒性。实际上,利用Hough变换,我们可以估计控制形状(包括直线、椭圆基于未知形状)的外观的参数。

形状提取的一些应用除了要求控制外观参数的确定以外,还要求变形(deform)或者弯曲(flex)的能力以匹配图像模板。正因为如此,紧接着基于匹配的形状提取是第6章灵活的形状分析。这是一个已经见到相当多进展的主题,特别是引入Snake(主动轮廓,active contour)的概念以后。水平集(level set)方法为形状提取带来了新的动力。它们寻求通过分析局部属性来为一幅图像匹配形状。而且,我们将会看到如何利用骨架描述形状,尽管具有可以借助对称性(symmetry)缓解的实际困难,以及涉及形状外观统计的全局约束如何被用来指导最终的形状提取。

至此,我们仍然没有考虑那些可以被用来描述一幅图像中找到的形状的方法。我们将会发现有两种主要的方法涉及到描述某形状的周长和面积。一些周长描述方法,如傅里叶描述(Fourier descriptor),依赖于分析它们频率内容的傅里叶变换理论。面积描述的主要方法,如统计矩(statistical moment),也是一种访问频率成分的方式,但与傅里叶分析具有明显不同的本质。一种优势在于:对描述能力的理解可以通过重构来实现,而重构应该还原到原始形状。

最后一章是纹理分析和一些模式分类的介绍。纹理描述了在没有已知分析描述下的模式,

是计算机视觉和图像处理领域相当多研究工作的目标。将它放在这里，更多是作为一种前面介绍的傅里叶变换和面积描述符的铺垫。此外，还有一些入门级的材料，关于如何对已知数据分类这些模式。

附录部分包括一个 Mathcad 和 MATLAB 工作表单的缩减版。其他部分包括与本书密切相关的材料，例如照相机模型、坐标几何，最小均方法和主成分分析方法。它们都只定位于一些简单的介绍，考虑到与本书内容密切相关才加到附录中来。其他的相关材料，特别是在线材料，在正文中被引用。

通过这种方式，本书覆盖了图像处理与计算机视觉的特征提取所涵盖的所有主要领域。然而，与在本书中介绍到的内容相比，仍然有相当多的材料没有包括进来。例如，有大量的三维计算机视觉和二维信号处理的材料，在本书中仅仅被提及。那些本书特别注意没有包括进来的内容包括颜色、三维处理和图像编码，因为涵盖所有内容将会导致一本无人买得起和拿得起的怪书！因此，我们承认本书只是一个“快照”，但是我们希望它被认为是为这样一个引人入胜和富有回报的领域打开了另外一扇窗。

致谢

非常感谢我们的同事，特别是 Steve Gunn 教授和 John Carter 博士。此外，我们还非常感谢南安普敦 (Southampton) 大学电子与计算机科学学院 Mark 教授领导的信息、信号、图像与系统 (ISIS) 研究组、计算机视觉领域那些过去和现在的研究人员。这些人员包括 Alberto 和 Steve, Hani Muammar 博士, Xiaoguang Jia 教授, Yan Qiu Chen 教授, Adrian Evans 博士, Colin Davies 博士, David Cunado 博士, Jason Nash 博士, Ping Huang 博士, Liang Ng 博士, Hugh Lewis 博士, David Benn 博士, Douglas Bradshaw 博士, David Hurley 博士, John Manslow 博士, Andy Tatem 博士, Chew Yam 博士, James Hayfron-Acquah 博士, Yalin Zheng 博士, Jeff Foster 博士, Peter Gething Myerscough 博士, David Wagg 博士, Ahmad Al-Mazeed 博士, Jang-Hee Yoo 博士, Nick Spencer 博士, Mike Jewell 博士, David Wagg 博士, Alex Bazin 博士, Hidayah, Rahmalan, Xin Liu, Imed Bouchrika, Banafshe Arbab-Zavar, Dan Thorpe, Cem Direkoglu (最后两位特别提供了主动轮廓的材料), John Evans (提供了河马的照片) 和 Jamie Hutton, Ben Dowling, Sina Samangooei (Java 演示网站)。我们还感谢其他已毕业的南安普敦大学电子工程、信息工程的本科生，计算机工程的本科生和硕士生，软件工程硕士生以及计算机科学专业的本科生，他们指出了本书的早期错误、一些需要澄清的部分以及志愿提供了本书的一些其他材料。除了南安普敦大学外，我们仍然对两个版本的编辑以及 Daniel Cremers 教授, Timor Kadir 博士和 Tim Cootes 教授心存感激，因为他们对本书的关注与改进以及允许使用他们的图像。上述各位，我们特别感谢！

最后的话

我们自己也已经从撰写本书以及第二版收获颇多。正如我们所知，以前的学生已经获益，也对本书有所贡献。但是，我们仍然希望，它将激励人们投入到这项引人入胜的，以及具有回报的学科，对于工作者来说它已被证明是快乐与灵感的源泉。

Mark S. Nixon

南安普敦 (Southampton) 大学

Alberto S. Aguado

萨里 (Surrey) 大学

目 录

第 1 章 绪论	1
1.1 概述	1
1.2 人类视觉和计算机视觉	1
1.3 人类视觉系统	3
1.4 计算机视觉系统	7
1.5 数学系统	12
1.6 相关文献资料	20
1.7 小结	23
1.8 参考文献	23
第 2 章 图像、采样和频域处理	26
2.1 概述	26
2.2 图像形成	26
2.3 傅里叶变换	29
2.4 采样标准	33
2.5 离散傅里叶变换 (discrete Fourier transform)	37
2.6 傅里叶变换的其他特性	43
2.7 傅里叶以外的其他变换	46
2.8 频域特性的应用	50
2.9 更多阅读资料	52
2.10 参考文献	52
第 3 章 基本图像处理运算	54
3.1 概述	54
3.2 直方图	54
3.3 点算子	55
3.4 群运算	64
3.5 其他统计算子	71
3.6 数学形态学	81
3.7 更多阅读资料	89
3.8 参考文献	89
第 4 章 低层次特征提取 (包括边缘检测)	92
4.1 概述	92
4.2 一阶边缘检测算子	93
4.3 二阶边缘检测算子	109
4.4 其他边缘检测算子	115

4.5	边缘检测算子的比较	115
4.6	关于边缘检测的更多阅读资料	116
4.7	相位一致性	117
4.8	定位特征提取	122
4.9	描述图像运动	134
4.10	小结	143
4.11	参考文献	144
第5章	形状匹配的特征提取	147
5.1	概述	147
5.2	阈值处理和差分	148
5.3	模板匹配	149
5.4	霍夫变换	158
5.5	广义霍夫变换	179
5.6	霍夫变换的其他扩展	191
5.7	更多阅读资料	193
5.8	参考文献	193
第6章	弹性形状提取(蛇模型及其他方法)	196
6.1	概述	196
6.2	可变形模板	196
6.3	主动轮廓(蛇模型)	198
6.4	形状骨架化	216
6.5	弹性形状模型:主动形状和主动外观	221
6.6	更多阅读资料	223
6.7	参考文献	224
第7章	目标描述	227
7.1	概述	227
7.2	边界描述	227
7.3	区域描述符	252
7.4	更多阅读资料	264
7.5	参考文献	264
第8章	纹理描述、分割和分类基础	267
8.1	概述	267
8.2	什么是纹理?	267
8.3	纹理描述	269
8.4	分类	274
8.5	分割处理	278
8.6	更多阅读资料	279
8.7	参考文献	280

第9章	附录1:工作表实例	283
9.1	第3章的 Mathcad 工作表实例	283
9.2	第4章的 MATLAB 工作表实例	285
第10章	附录2:摄像机几何基础	288
10.1	图像几何	288
10.2	透视摄像机	288
10.3	透视摄像机模型	289
10.4	仿射摄像机	295
10.5	弱透视模型	299
10.6	摄像机模型实例	300
10.7	讨论	308
10.8	参考文献	308
第11章	附录3:最小二乘分析	309
11.1	最小二乘标准	309
11.2	最小二乘曲线拟合	310
第12章	附录4:主成分分析	313
12.1	介绍	313
12.2	数据	313
12.3	协方差	314
12.4	协方差矩阵	316
12.5	数据变换	316
12.6	逆变换	317
12.7	特征值问题	318
12.8	求解特征值问题	318
12.9	PCA 方法小结	319
12.10	实例	319
12.11	参考文献	324

第1章 绪 论

1.1 概述

我们将通过了解人类视觉系统来研究视觉是什么、如何使计算机理解图像数据,以及如何处理图像。本章的主要内容如表 1.1 所示;以后每一章的开头都有一个这样的内容概述。内容概述里不标注参考(引用)文献,它们在文中有标注,并集中列出在每一章的末尾。

表 1.1 第 1 章内容概述

主要话题	子 话 题	要 点
人类视觉系统	眼睛的工作机制,视觉信息处理和视觉错觉	视力,眼球晶体,视网膜,图像,颜色,单色,处理,脑,视觉错觉
计算机视觉系统	电子图像的成像原理,视频输入,计算机视频信息处理	像素,视频标准,摄像机技术,像素技术,性能效果,专业摄像机,视频变换,计算机语言,处理工具包。处理方法的演示
数学系统	处理图像的数学工具, MATLAB 和 Mathcad 系统介绍	结果可视化,用法介绍,工作表实例
参考文献	图像处理、计算机视觉和特征提取相关的教材和其他资料	期刊,教材,网站及本书的网站

1.2 人类视觉和计算机视觉

计算机系统所处理的图像是通过电子摄像机获取的,正如人类视觉系统中由大脑处理眼睛获取的图像一样。计算机视觉对于电子工程师、计算机科学家和许多其他学者都是非常有价值的一个研究领域,并逐渐呈现出广阔的商业前景。目前有许多视觉系统用于常规工业,比如利用摄像机检测机械零部件的尺寸、食品质量检测,就连用于天文科学的图像也得益于计算机视觉技术。应用计算机视觉的法医鉴定(forensic studies)和生物测定学(biometrics, 是用于识别人的方法)主要包括自动人脸识别和虹膜纹理识别。这些研究由生物学家和心理学家共同进行,他们不断研究人类视觉系统的机制,研究我们是如何看见和识别目标(和人)的。

图 1.1 给出了几帧(计算机)图像;这些图像由一组点或图像元素(通常简称像素, pixel)构成,作为一系列数字存储在计算机内。例如要从图 1.1(a)识别人的脸部,需要能够分析出各个部位的形状,如鼻子、眼睛和眉毛的形状,得到一些量度来描述继而识别人脸。[图 1.1(a)是图像处理领域中最常用的图像之一。它被称为 Lena 图像,是 1972 年 Playboy 杂志刊登的 Lena Sjöblom 的一张照片]。图 1.1(b)是颈动脉(carotid artery, 位于颈侧附近,提供血液给大脑和脸部)的超声波图像(ultrasound image),是颈动脉的横截面图(cross-section)。图像的上部靠近皮肤,下部在颈内侧。这帧图像是经由组织(tissue)的超声辐射(ultrasound radiation)反射合成得到的,来源于一项对颈动脉进行三维(3D)建模以辅助血管手术(vascular surgery)的研究。需要注意的是,这帧图像有很多噪声,使(椭圆形)颈动脉的形状看起来很不清晰。遥感图像(remotely sensed image)通常利用纹理内容来进行分析。如图 1.1(c)中交叉路口和不同类型的植被,它们的纹理在感觉上有很大不同。最后,图 1.1(d)是人体中腹部横截面磁共振成

像(Magnetic Resonance Imaging, MRI)所得到的图像。胸部位于图像的上端,肺和血管处于较暗区域,内部器官和脂肪位于灰色区域。由于可以提供高质量图像,MRI 图像目前已经被广泛应用于医学领域。

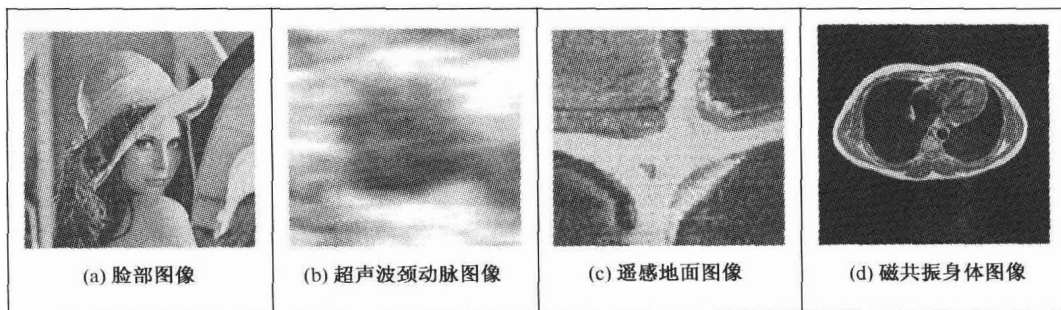


图 1.1 不同来源的真实图像

获取图像的方法有许多。在医学研究中,MRI 适合于获取软组织的图像,但不显示骨骼结构[图 1.1(d)中看不到脊柱];与软组织相反,骨骼结构的图像可以利用计算断层摄影术(computed tomography, CT)来获取,该技术适合于骨骼成像。遥感图像可以通过红外(热)传感器(infrared sensor)或合成孔径雷达(synthetic-aperture radar)获取,而不是利用摄像机,如图 1.1(c)所示。空间信息可以由二维传感元件阵列来提供,包括声呐阵列(sonar array)等。可以说医学研究领域里获取空间数据的方法比其他领域更加多样化。不过,计算机视觉技术可以用来分析各种类型的数据,并非只处理摄像机获取的图像。

合成(synthesized)图像便于用来评价(evaluating)不同方法,了解它们如何实现并找出它们各自性能的局限。图 1.2 给出两帧合成图像。图 1.2(a)是用数学方法(mathematically)定义的同心圆图像。该图像显示的是一个理想的情形:这些圆都被完全正确定义,其亮度级(brightness level)为常量。这类合成图像不仅可以用来评价一些方法是否能够正确检测出形状的边缘(边缘)及其形状本身,还可以用来描述形状。图 1.2(b)由几帧不同的真实图像数据合成,各图像数据区域的边界由程序精确设定。图像数据取自有名的纹理库,即 Brodatz 纹理相簿(Brodatz album of texture)。纹理库里的这些图像数据都是通过扫描并被存储为计算机图像的。该合成图像可以用来分析计算机视觉算法如何有效地识别不同的纹理区域。

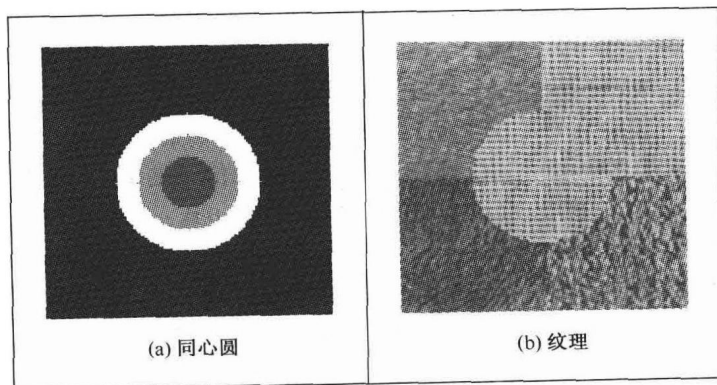


图 1.2 两帧合成图像

本章将对照人类视觉系统,介绍基本的计算机视觉系统是如何运作的,它涵盖人类视觉的主要组成元素,并介绍我们的眼睛是如何起作用(以及如何被骗)的。从计算机视觉方面来看,这一章主要介绍的是用于图像分析的硬件和软件,简单介绍 Mathcad 和 MATLAB 这两个软件工具,本书将用它们来实现计算机视觉算法。另外,我们还将提供一些其他资料的线索,便于更加详细地了解本章中所提到的内容。

1.3 人类视觉系统

人类视觉是一个感觉到视觉刺激(visual stimuli)并对其做出反应的复杂系统。它已经进化了几百万年,主要用以自卫和生存。直观上,计算机视觉和人类视觉具有相同的功能。两者的目标都是解读一维以上的空间数据。然而,尽管计算机视觉和人类视觉在功能上相似,我们不能指望计算机视觉系统完全复制人类眼睛的功能。一方面我们还没有完全理解眼睛和大脑视觉系统的运行机制,这一点在本节中将会提到。因此,我们还无法设计一个系统来完全复制人类视觉系统的功能。其实,在开发计算机视觉技术时人类眼睛的有些特征是有用的,有些功能却并不受欢迎。尽管如此,将来仍然可以看到计算机视觉技术在一定程度上能拥有人类视觉系统的功能,在有些方面甚至能比后者更优越。

我们考虑一下这种情形,把两只手各伸出一个手指放在面前,试图估测它们之间的距离。这有些难度,而且可以肯定这样的估测不会很准确。现在把那两个手指慢慢接近彼此,即使它们之间的距离非常小,我们仍然可以判断它们是分开的。所以说人类视觉善于区分相对(relative)距离,而不善于区分绝对(absolute)距离。计算机视觉正好相反:它很善于估测绝对差异,但是对相对差异估测精度很低。图像中像素的个数体现计算机视觉系统的精度,这一点将在下一章中提到。我们先了解一下人类视觉系统是如何运行的。

人类视觉中,光感元(sensing element)是眼睛。通过眼睛,图像由视神经(optic nerve)传到大脑进行进一步处理。视神经不具备足够的带宽(bandwidth)来传输眼睛获取的这些信息。因而,图像在经视神经传送前需要进行预处理。人类视觉系统可以用三个部分来建模:

- 眼睛:这是一个物理模型,由于它的许多功能可以用病理学(pathology)来确定。
- 处理系统:这是一个实验模型,因为它的功能可以建模,但不能精确断定。
- 脑分析:这是一个心理模型,因为我们不能直接接近脑处理过程或者为其建模,只能通过实验和推理来判断其运转状态。

1.3.1 眼睛

眼睛的功能是成像;图 1.3 是眼睛的横截面图。视力(vision)要求眼睛具有对目标物体有选择性聚焦的能力,这一能力是通过眼球晶体(lens)周围的睫状肌(ciliary muscle)来实现的。正是因为这些肌肉变得松弛,老年人的眼睛才会失去短距离聚焦的能力。虹膜(iris)或瞳孔(pupil)就像摄影机的光圈,控制进入眼睛的光量。虹膜是一个非常脆弱的系统,需要角膜(cornea)或巩膜(sclera)的保护。角膜的内侧是脉络膜(choroid),有血管提供营养,而且是不透明的,便于减少光量。视网膜(retina)在眼球内部,光投射到上面形成图像。通过这一系统,肌肉使眼球转动并且改变晶体的形状,以便在中央凹(fovea,即焦点)形成图像,因为这里聚集了大多数传感元。而盲点(blind point)是视神经的起点,那里没有传感元。

聚焦需要改变晶体的形状,而不是如摄像机那样给它定位。晶体形状变化实际上是通过拉伸作用使较近的物体图像变大而使较远的物体图像变小。晶体的焦点中心距离根据其形状不同大约在 14 mm 和 17 mm 之间变化。这就是说,一个三维场景被转换成 2 mm^2 左右的一个平面。好的视力具有高视敏度(acuity,即清晰度),意味着成像区附近一定有大量传感元。

视网膜(retina)周围散布着 1 亿多个光感元。光投射到这些光感元上激励光化学反应,引起神经冲动,这些神经冲动聚集在一起成为由眼睛传递的信号。光感元分成两类:一类是视杆细胞(rod),用于黑白视觉(scotopic vision,即暗视觉);另一类是视锥细胞(cone),用于色彩视觉(photopic vision,即明视觉)。视锥细胞有 1 千万左右,分布在中央凹 5° 范围以内,其余 1 亿视杆细胞分布在视网膜周围,大多位于中央凹 20° 和 5° 之间。视敏度表现为空间分辨率(即清晰度)和亮度/色彩分辨率,在中央凹 1° 范围内最大。

视杆细胞只有一种,而视锥细胞有三种,它们分别是:

- S(短波长):感应可视光谱区的蓝色光
- M(中波长):感应绿色光
- L(长波长):感应红色光

视锥细胞的总体响应是这三种类型的视锥细胞响应的总和,涵盖整个可视光谱范围。视杆细胞对可视光谱全区的光非常敏感,具有暗视觉的单色能力(monochrome capability)。因此,当光很微弱时,利用视杆细胞的高灵敏度,图像在远离中央凹的区域形成,但没有视锥细胞的色彩视觉。需要注意的是,蓝视锥细胞数量非常少,而另两种视锥细胞数量比较多。然而,我们仍然能看到很多蓝色(尤其是随处可见的牛仔布!)。大概是人类视觉系统以某种方式弥补了蓝光感元的不足,使我们能够感知到蓝色。否则,如果水是红颜色,而不是蓝颜色,这个世界会是多么不可思议呀!视觉响应是对数型的,并依赖于从暗处到亮处的亮度适应(brightness adaptation),暗处成像在视杆细胞上,而亮处成像在视锥细胞上。

眼睛有一种先天特性,即马赫带(Mach band)效果,影响我们感知图像的方式。马赫带是如图 1.4 所示的那些条带,即紧挨着的具有不变色调的两个样条。通过给这些图像亮度级赋予不同的值,画出的亮度横截面如图 1.4(a)所示,从中可以看出,图像是由某些亮度值不变的样条形成的。而人类视觉感知到的图像是图 1.4(c)所示的横截面。这些马赫带其实并不存在,是由我们的眼睛产生的,是不同亮度区域的边界在人眼睛中的过冲(overshoot)响应(它有助于我们区分视界内的不同目标)。真正的横截面如图 1.4(b)所示。值得注意的是,人类眼睛只能区分相当少的灰度级。它总共能够辨别 32 级(相当于 5 位)灰度,虽然图 1.4(a)还可以再增加许多个亮度级。这就是为什么我们的感知更难辨别图 1.4(a)中的低亮度样条[需要注意的是,马赫带现象在图 1.2(a)的同心圆图中并没有出现,那是由于灰度值的排列不同]。这些就是我们对人类视觉初级研究的主要成就;Cornsweet(1970)为感兴趣的读者介绍了更多有关视觉感知的知识。

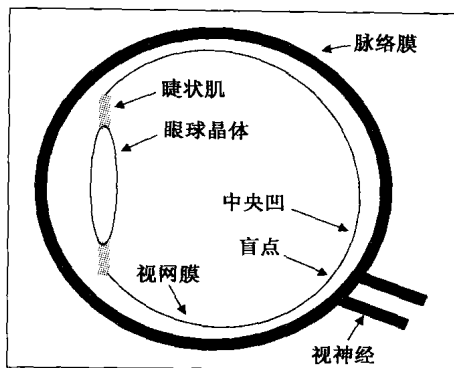


图 1.3 人类眼球结构图

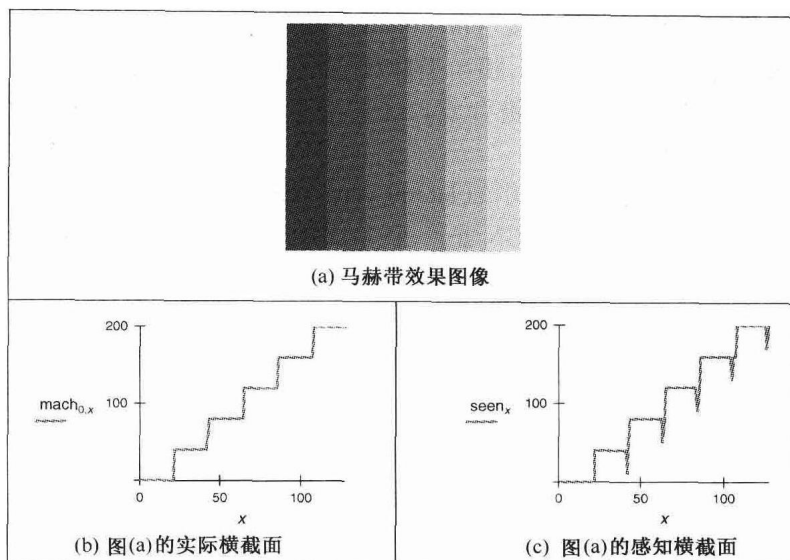


图 1.4 马赫带效果示意图

我们已经确认了与眼睛有关的两个特性，即马赫带和对非感知现象 (unsensed phenomenon) 的灵敏度，这两个特性在计算机视觉系统中很难被采用，通常也不受欢迎。这些特性集成在一起成为人类视觉。目前，人类视觉比我们想要实现的计算机视觉系统复杂得多。红外制导导弹 (infrared guided-missile) 视觉系统还很难辨别 100 m 远处的一只鸟和 10 km 远处的一架飞机。这是鸟的不幸 (或是飞机的幸运)！然而，人类视觉就能轻易做到。

1.3.2 神经系统

眼睛产生的神经信号实际上是视锥细胞和视杆细胞这两种波长应变 (wavelength-dependent) 感受器细胞所传送的响应。图 1.5 所示的是把这些信号叠加起来的一个加法模型。响应通过一个对数函数进行转换，映射眼睛产生的已知响应。然后，与权重因子相乘来调节某一个特定光感元的贡献，这样可以把来自特定区域的响应结合在一起处理。权重因子可以根据是否有特定过滤作用来进行选择。比如，侧抑制 (lateral inhibition) 现象中，中心光感元的权重要比末端光感元的权重大得多，便于使中心光感元的响应在合成响应中起主导作用。如果把一半权重因子设为负数，而另一半设为正数，权重函数的微分处理结果将表现为对比检测 (即亮度变化)。

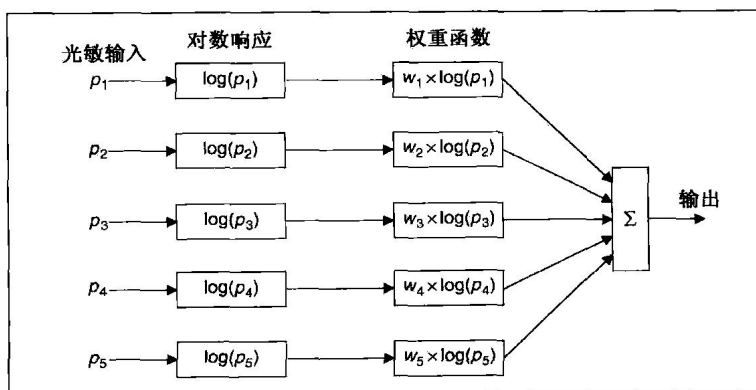


图 1.5 神经处理过程

视锥细胞的信号可以通过反射色度(chrominance)和亮度(luminance)的方法进行合成。这可以利用对数函数的差分方法来实现,相当于计算它们的对数比,得到色度。通过这种方法,光感元的信号在经由视神经传送之前被合成。这是一个实验模型,还有许多其他方法可以用来把不同的信号合成在一起。

然后,视觉信息被送回到外侧膝状核(lateral geniculate nucleus, LGN),它位于丘脑(thalamus),是视觉信息的主要处理器。LGN 是一个多层结构,包括不同类型功能各异的细胞。LGN 的细胞轴突(axon)把信息继续传给视觉皮层(visual cortex)。尽管研究表明 LGN 在传输信号编码上起到很大作用,但它的功能还远没有被了解。还有人认为它帮助视觉系统集中注意力,比如关注声音的来源。关于视网膜神经网络的更多信息可参考 Ratliff(1965);神经处理过程的另一类研究可参考 Overington(1992)。

1.3.3 处理系统

神经信号被传送到大脑的两个区域进行进一步处理。这两个区域分别是联络皮层(associative cortex)和枕叶皮层(occipital cortex),其中联络皮层负责建立目标之间的链接(link),而枕叶皮层负责处理模式(pattern)。当然,很难精确判断大脑区究竟发生了什么。到目前为止,还没有志愿者愿意让科学家对他们的脑功能进行详细研究[尽管新成像方式的发展,比如正电子发射断层扫描(positive emission tomography)或者电阻抗断层成像(electrical impedance tomography),无疑将会大有帮助]。因此,只有用心理模型来推测大脑区是如何工作的。

众所周知,人类视觉系统的功能之一是应用目标的边缘或边界。我们能够轻易地读出图 1.6(a)所示的单词;这是在已有知识基础上判断该模式很可能是一个打印的单词,而对缺损的边界进行填充来实现的。不过从这张图像我们还可以有更多的推测;可以假设有一个照明,使光照不到的区域出现了阴影。如果这个光源很亮,图像会被洗白,使眼睛填充的边界消失。所以说,除了物理响应,还有知识在起作用,包括立体几何(solid geometry)的先验知识。这种情形如图 1.6(b)所示,它既可表示为即将碰撞的三个“吃豆先生”(pacmen),也可表示为一个三角形放在三个黑色圆圈上,这两种情形都有可能。

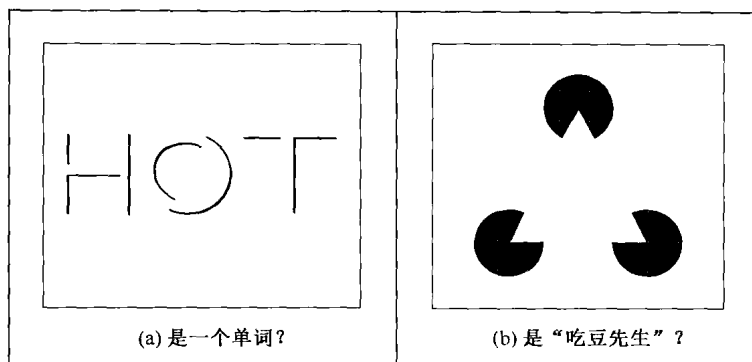


图 1.6 人类视觉如何利用目标边缘

人类视觉可以被欺骗,主要是通过人们没受过训练处理的场景。有名的措尔纳错觉(Zollner illusion)[参见图 1.7(a)]中,那些竖棒看起来都是歪斜的,但实际上都是垂直的(可以把一枝笔放在两条直线之间进行验证):那些小横条误导了我们的眼睛,使我们把垂直的竖棒看