

神经反应及其 在图像识别中的应用

胡政发 翁 竞 著



科学出版社

神经反应及其在图像识别中的应用

胡政发 翁 兢 著

科 学 出 版 社

北 京

版权所有,侵权必究

举报电话:010-64030229,010-64034315,13501151303

内 容 简 介

图像识别是模式识别的重要分支,在计算机视觉领域有着广泛的应用.近几年来,分层学习算法得到了高度关注,成为机器学习和图像识别领域研究的热点之一.结合国内外最新的研究成果,本书系统地介绍基于层次结构模型的神经反应及其各种推广形式的学习算法,并将其应用到图像识别中.首先对人类视觉特性和机器视觉特征进行对比与概述,接下来重点介绍经典神经反应、局部神经反应、软稀疏编码神经反应以及特征传导神经反应等.最后,对神经反应中的一个关键因素,即模板选择方法进行一些积极探索.本书还提供大量的仿真实验,对各种不同的算法进行多角度、全方位的测试和验证.

本书对神经反应相关理论及其在图像识别中的应用进行全面、系统、详尽的阐述,可作为从事机器学习和图像识别等相关理论与方法研究的科研工作者和研究生的参考书.

图书在版编目(CIP)数据

神经反应及其在图像识别中的应用/胡政发,翁兢著. —北京:科学出版社, 2017.6

ISBN 978-7-03-053636-5

I. ①神… II. ①胡… ②翁… III. ①结构模型(系统工程)-研究
IV. ①N945.12

中国版本图书馆 CIP 数据核字(2017)第 129987 号

责任编辑:吉正霞 王 晶/责任校对:董艳辉

责任印制:彭 超/封面设计:苏 波

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

武汉市首壹印务有限公司印刷

科学出版社发行 各地新华书店经销

*

开本:B5(720×1000)

2017 年 7 月第 一 版 印张:11

2017 年 7 月第一次印刷 字数:218 000

定价:58.00 元

(如有印装质量问题,我社负责调换)

前 言

本书试图探索一个基本问题:人类视觉超凡脱俗的识别能力的基础是什么?

随着计算机技术的迅猛发展和网络资源的无所不在,人们对数字图像识别的需求越来越频繁地出现在如图像检索、视频监控、机器视觉、导航以及无人驾驶技术等实际工程和日常生活中.数字图像识别的一个基本方法是依据图像的像素值提取特征并利用这些特征来区分图像.这样,高维的图像被压缩并表达为低维的特征向量.这种采用特征向量而不是用原始像素值表示图像的方法,其明显好处是可以显著地简化图像表示,但最重要的作用是利用特征向量可以加强图像的低级特征和高级语义之间的关联.因此,开发能够有效表达图像或感兴趣的目标对象的本质属性的特征提取方法是十分重要的.

在过去的数十年里,研究者已经开发出了大量的数字图像特征提取方法,如主成分分析(principal component analysis, PCA)、线性判别分析(linear discriminant analysis, LDA)、尺度不变特征变换(scale invariant feature transform, SIFT)、方向梯度直方图(histogram of oriented gradients, HOG)、小波(wavelet)、特征包(bag of features, BoF)、空间金字塔匹配(spatial pyramid matching, SPM)、卷积神经网络(convolutional neural network, CNN)、类视觉皮层机制分层模型(hierarchical model of cortex-like mechanisms, HMCM)、神经反应(neural response, NR)、局部神经反应(local neural response, LNR)以及基于对数极坐标变换的导出核(log-polar transform-based derived kernel, LPTDK)等.在这些方法中,由于简单,主成分分析和线性判别分析成为最常用的全局方法.然而,在实际中许多数据集包含不能通过主成分分析和线性判别分析发现的重要的非线性结构.为了解决这个问题,非线性流形学习技术被开发出来,如局部线性嵌入(locally linear embedding, LLE)、等距特征映射(isometric feature mapping, IFM)以及拉普拉斯特征映射(Laplacian eigenmaps, LEM)等,这些非线性的方法能够成功地获得非线性流形的内在特征.然而,这些方法没有为不在训练集中的新的样本点

提供映射,因而影响了其推广能力.虽然已经有一些方法尝试处理这个问题,但它们对如观察条件变化、遮挡和噪声等很敏感.尺度不变特征变换和特征包是具有代表性的局部方法,其中尺度不变特征变换可以容忍一定程度的光照变化、透视失真,对图像变换、图像遮挡具有强大的鲁棒性.特征包及其改进方法提取关键点或显著区域等局部特征,实验结果表明特征包可以在一定程度上改进分类和识别效果.不幸的是,确定图像的关键点和显著区域往往依赖于复杂的图像分割技术.然而,到目前为止图像分割技术仍然是图像处理中的难点问题之一,因此在一定程度上限制了这些方法在实际中的应用.

另外,人类视觉感知系统具有通过少量实例学习就能很好地进行推广的能力,堪称为自然界中一个无与伦比的图像特征描述子.因此,构造一套可以与人类视觉相媲美的机器视觉系统,就成为机器视觉研究领域的一个宏伟目标.

神经科学和生物解剖学上的研究表明,灵长类动物的这种视觉感知和辨识能力是其视觉皮层的分层结构所赋予的.因此,最近几年来基于视觉皮层的分层学习算法得到了更多的关注,如类视觉皮层机制分层模型、卷积神经网络、神经反应和局部神经反应方法等.层次结构模型搭建了一个深度学习的框架,将全局方法和局部方法有机结合起来,使得提取的特征能够在辨识能力和鲁棒性之间达到一种相对理想的均衡.特别地,类视觉皮层机制分层模型严格遵循人类视觉皮层的组织结构,应用于人脸识别、场景分类和手写字识别等领域,并表现出令人鼓舞的识别效果. Smale 和 Poggio 等在麻省理工学院提出的神经反应模型提供了一个有效的分层特征提取方法,该方法模拟人类视觉系统分层处理信息的过程建立数学模型.在这一模型中,神经反应是分级的,而导出核是定义在神经反应基础上的一个相似性度量,由此可以递归地定义图像的一种非常接近视觉感知的特征描述子.理论分析和实验结果均表明,基于分层结构的神经反应及其改进方法在图像特征表示中表现良好,在不同的应用中该模型具有进一步改进和提高的潜力.

例如,在神经反应模型的理论框架下, Li 等提出了一种称为局部神经反应的分层特征提取模型并应用于图像识别.该方法在各层采用局部编码的方式以提取比神经反应更具辨识能力和鲁棒性的特征.然而,局部神经反应算法在局部编码过程中采用的是一种硬分配模式,即在对同一层的所有子块进行编码时,利用的模板个数是相同的.这必将导致一方面在对某些图像块的编码中会

丢失一些重要模板的信息,另一方面在对某些图像块的编码中又会包含不必要的甚至起干扰作用的一些模板的信息.同时,局部神经反应算法在对每个图像块的特征表示中引入了二次优化问题,这将导致沉重的计算负担,使得该方法难以满足某些图像识别任务中的实时性要求.

本书从对人类视觉系统特性和机器视觉系统结构的对比分析入手,系统地介绍基于层次结构模型的神经反应及其各种推广形式的深度学习算法,并将其应用到图像识别中.其中既有深入、严谨的数学上的理论分析,又包含大量的实证研究,并对各种算法的优缺点进行深度剖析,为机器学习和图像识别等相关理论与方法的研究提供一个全面、系统、详尽的参考.

本书共六章.各章后面都附有详细的参考文献,以方便读者查阅.

第1章对照人类视觉系统,阐述基本的机器视觉系统是如何运作的,对机器视觉和人类视觉进行概述.讲述人类视觉的主要组成元素,并介绍人类的眼睛和视觉神经系统是如何起作用的.此外还介绍在机器视觉系统模拟人类视觉系统的过程中的另一个关键的系统——数学系统,它包括对图像所做的一系列数学运算以及将这些运算转化为机器指令的相关程序和软件,这也是本书所做理论分析和实证研究的基础.

第2章介绍经典神经反应及其数学表示.2008年,Smale、Poggio等基于视觉皮层的神经科学提出了神经反应这一概念,给出了一个自然的图像表示.神经反应的内积称为导出核,它给出了图像间相似性的一个度量.基于一个分层的体系结构,他们为人类视觉皮层分级处理信息的过程建立数学模型,递归地定义了神经反应和相关的导出核.神经反应和导出核可用于多种应用领域,如图像、文本字符和基因组数据的分类等.本章将介绍这一开创性工作的核心思想,这也是本书后继各章所涉及工作的研究基础.

第3章在经典神经反应的基础上,给出一种新的分层特征提取的方法并应用于图像识别.该方法的核心思想是通过将输入的图像进行局部编码以及最大化操作的交替作用,保持对不同类别的图像的辨识性和对图像变换的不变性,从而有效地提取图像的特征.这种特征称为局部神经反应.局部编码是在局部线性流形上进行的,可以提取图像块的显著特征,并生成一个稀疏的度量矩阵,再在其上进行最大化操作.通过最大化操作,将平移不变性引入模型之中.同时还将证实该特征表示对其他变换的不变性,如旋转和缩放,都可以通过这个模

型诱发. 大量的实验结果表明, 这种方法与经典神经反应相比, 在对图像的识别精度上有较明显的提升, 与目前先进的算法对比, 该方法仍具有一定的优势, 特别是该方法对图像的局部失真和噪声污染都是鲁棒的.

第4章给出一种基于软稀疏编码的分层特征提取方法并用于图像识别. 该方法的核心思想是在神经反应的框架下, 通过建立一个模拟人类视觉感知系统的分层结构, 在对图像的底层视觉特征进行深度学习的同时利用软稀疏编码得到图像更加有效的特征表示. 和神经反应方法类似, 首先在图像的空间域上建立一个分层结构并构造相应的模板集. 然后从底层开始, 通过对图像块与模板之间的相似度进行软稀疏编码, 得到图像块的一个稀疏编码矩阵, 在此矩阵上进行最大化联合操作得出图像子块的特征表示. 在接下来的各层, 递归地交替实行这种软稀疏编码和最大化联合操作, 并最终将图像表达为一个称为软稀疏编码神经反应(SSCNR)的特征向量. 实验结果显示, 与目前存在的类似方法相比, 基于软稀疏编码神经反应的图像识别方法表现出较好的识别效率和效果.

在神经反应的基础之上, 第5章给出一种新的图像特征表示方法以用于基于内容的图像检索(content-based image retrieval, CBIR). 该方法的核心思想是通过建立一个模拟人类视觉感知系统的分层结构, 在对图像的底层视觉特征进行深度学习的同时将图像的语义成分引入特征表示中, 从而构造出特征传导神经反应(feature conduction neural response, FCNR)这一新的图像特征表示. 具体来讲, 首先在图像的局部区域上提取底层视觉特征, 并对图像的局部特征表示建立一个分层结构, 然后在这一分层结构的底层子块上利用局部特征构造神经反应, 通过神经反应归一化的内积将这些特征逐层传导到高级子块, 最终将图像表达为一个称为特征传导神经反应的特征向量. 在此基础上定义图像相似性度量, 利用这种相似性度量, 可以构建基于特征传导神经反应的图像检索方法. 在两个通用图像数据集上的实验结果显示, 无论是与经典的神经反应算法还是与某些专门用于基于内容的图像检索的其他图像特征描述子相比, 基于特征传导神经反应的图像检索方法都表现出较好的检索效率和效果.

第6章专门考虑神经反应算法中的一个关键因素, 即模板的构造问题. 模板是联系现实世界和数学模型的纽带. 与人类的视觉系统相比, 模板就相当于视觉系统中的神经元. 这些神经元储存了对现实世界认知的一些最基本的信息, 当这些神经元受到某种刺激时, 那些储存了相关信息或知识的神经元就会

变得活跃而作出积极响应.相反,那些与某种刺激无关的神经元就会显得无能为力,对认知起不到明显作用.图像的神经反应,其实质是图像对模板的刺激而产生的模板对图像的某种响应.从数学的观点看,模板是神经反应的基石.把图像(或图像块)表示成神经反应,就是通过模板将其表示成一个向量.从这个意义上看,模板集就相当于图像空间的一组基,而神经反应就是图像在这组基下的坐标.在该章构造模板的过程中重点考虑了这些问题,给出了“自下而上”和“自上而下”两种不同的方式来选择模板,并将按这些方式选择的模板应用到前面各章相应的部分实验中.实验结果显示,与随机选择模板的方式相比,这两种模板选择方法,都在一定程度上提高了图像的识别精度和识别效率.

本书是在湖北省教育厅重点科研项目(D20131803)、湖北汽车工业学院博士科研基金(BK201209)、湖北汽车工业学院青年科研基金(X2012XQ09)的研究课题资助下,对近年研究成果的系统总结和提升.在本书写作过程中,湖北大学的李落清、华中科技大学的李红提供了大量参考资料,并对有关问题给予了精心指导;同时,湖北汽车工业学院的罗时军及正在华中科技大学攻读博士学位的肖海霞对本书给予重要支持并提出了一些富有启发性的意见和建议.在此,感谢所有课题组成员的指导和帮助!

由于作者水平有限,本书难免有不足之处,诚请读者和专家批评指正.

作 者

2017年3月

目 录

第 1 章 人类视觉与机器视觉	1
1.1 人类视觉系统	2
1.1.1 眼睛	3
1.1.2 神经系统	6
1.1.3 处理系统	7
1.2 计算机视觉系统	8
1.2.1 摄像机	10
1.2.2 计算机	12
1.2.3 图像处理	14
1.2.4 研究现状分析	15
1.3 数学系统	16
1.3.1 MATLAB 软件	17
1.3.2 图像的数学表示及运算	19
1.4 小结	25
参考文献	26
第 2 章 经典神经反应及其数学表示	29
2.1 主要目的和基本思想	30
2.2 神经反应与导出核	31
2.2.1 预备知识	31
2.2.2 神经反应与导出核的构建	35
2.2.3 神经反应的归一化	38
2.3 神经反应的不变性	39
2.4 神经反应算法的复杂性	43
2.5 实验及分析	45
2.5.1 MNIST 手写数字图像数据库上的实验	45

2.5.2 手写汉字图像数据库	50
2.5.3 Yale 人脸图像数据库	54
2.6 小结	56
参考文献	57
 第 3 章 局部神经反应及其对图像的特征表示	59
3.1 局部神经反应方法产生的背景	60
3.2 局部神经反应算法	62
3.2.1 符号和说明	62
3.2.2 方法的提出	63
3.2.3 局部神经反应和神经反应的区别	70
3.3 局部神经反应方法的算法分析	72
3.3.1 局部神经反应方法的不变性	72
3.3.2 关于范围压缩问题	76
3.4 实验结果与分析	78
3.4.1 MNIST 数据库上的实验	78
3.4.2 在 Yale 人脸数据库上的实验	79
3.4.3 Caltech 人脸数据库上的实验	80
3.5 小结	82
参考文献	83
 第 4 章 软稀疏编码神经反应及其应用	88
4.1 研究背景和主要工作	89
4.2 准备工作及神经反应的矩阵表示	93
4.2.1 相关记号	93
4.2.2 神经反应的矩阵表示	94
4.3 软稀疏编码神经反应	96
4.3.1 软稀疏编码的动机	96
4.3.2 软稀疏编码方法	98

4.4 实验结果及分析	101
4.4.1 在 MNIST 数据库上的实验	101
4.4.2 在 Caltech 人脸数据库上的实验	103
4.4.3 在 Corel 数据库上的实验	104
4.5 小结	106
参考文献	107
第 5 章 特征传导神经反应及其在图像检索中的应用	110
5.1 基于内容的图像检索	111
5.2 特征传导神经反应	114
5.2.1 相关概念与记号	114
5.2.2 构建特征传导神经反应	116
5.2.3 特征传导神经反应的算法复杂性	118
5.3 基于特征传导神经反应的基于内容的图像检索方法	120
5.3.1 局部底层特征提取	121
5.3.2 相似性度量	122
5.3.3 关联性反馈	123
5.4 实验结果及分析	124
5.4.1 在 Caltech 人脸数据库上的实验	124
5.4.2 在 Corel-1000 数据库上的实验	128
5.5 小结	133
参考文献	134
第 6 章 神经反应中的模板选择方法	137
6.1 “自下而上”的模板选择方法	139
6.1.1 初始模板集的确定	139
6.1.2 T_u 的确定	139
6.1.3 T_v 的确定	141
6.2 “自上而下”的模板选择方法	144

6.2.1	创建初始模板集	145
6.2.2	二级模板集 T_v 的构建	146
6.2.3	一级模板集 T_u 的构建	149
6.3	实验结果	152
6.3.1	在 MNIST 上的模板选择实验	153
6.3.2	在 Caltech 101 上的模板选择实验	155
6.3.3	在 Corel 上的模板选择实验	158
6.4	小结	160
	参考文献	160

第 1 章 人类视觉与机器视觉

人类视觉系统堪称一个完美的图像处理系统,这一系统在一定的认识学习和知识积累的基础上能够准确地获取图像的高级语义特征并准确快速地识别图像.因此,构建模拟和逼近人类视觉系统的机器视觉系统一直是图像处理这一研究领域的宏伟目标^[1].

机器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^[2,3].目前有许多机器视觉系统用于常规工业,如利用摄像机检查机器零部件的尺寸、食品质量检测等,就连用于天文、军事、航空航天图像识别也得益于计算机视觉技术.应用机器视觉的法医鉴定和生物测定学主要包括自动人脸识别与虹膜纹理识别.这些研究往往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及其机制,探索人类是如何看见并且识别目标的^[4-10].

那么,人类视觉系统的机制是什么?

机器视觉系统的机制又是什么?

如何利用机器视觉系统模拟人类视觉系统来理解图像数据以及处理图像?

本章将对照人类视觉系统,阐述基本的机器视觉系统是如何运作的,它涵盖了人类视觉的主要组成元素,并介绍人类的眼睛和视觉神经系统是如何起作用的.本章还将介绍在机器视觉系统模拟人类视觉系统过程中的另一个关键系统——数学系统,它包括对图像所做的一系列数学运算以及将这些运算转化为机器指令的相关程序和软件,这也是本书要研究的核心问题.

1.1 人类视觉系统

人类视觉系统是一个感受到视觉刺激(visual stimuli)并对其作出反应的复杂系统.它已经进化了几百万年,主要用于自卫和生存.理想状态下,机器视觉应该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,我们不能指望机器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,只能近似地模拟这种机制.目前,机器视觉技术严格模拟人类视觉系统的功能只是机器视觉研究领域的宏愿.另一方面,在开发机器视觉技术时人类视觉的有些功能是有用的,有些功能却并不受欢迎.尽管如此,将来依然可以看到机器视觉技术在一定程度上能够拥有人类视觉系统的功能,在某些方面甚至比后者更优越.

我们考虑如下一个心理学中常用的案例^[11].图 1.1(a)中的两条线段是相连的且绝对长度相等,但我们的视觉系统告诉我们这两条线段是相连的但竖线较长,而且我们也很难说清它们到底有多长.图 1.1(b)中的两个图形是分开的且大小完全一样,我们的视觉系统能够清楚地发现它们是分开的,但感觉上面的图形要大于下面的图形.这个示例说明,人类视觉善于区分相对距离,而不善于区分绝对距离,有时候甚至会产生严重的错觉.机器视觉正好相反:它很善于估计绝对差异,但对相对差异的辨识能力较低.构建机器视觉系统的核心问题就是在绝对差异和相对差异的辨识能力上达到某种均衡,以满足机器视觉的实际需要.

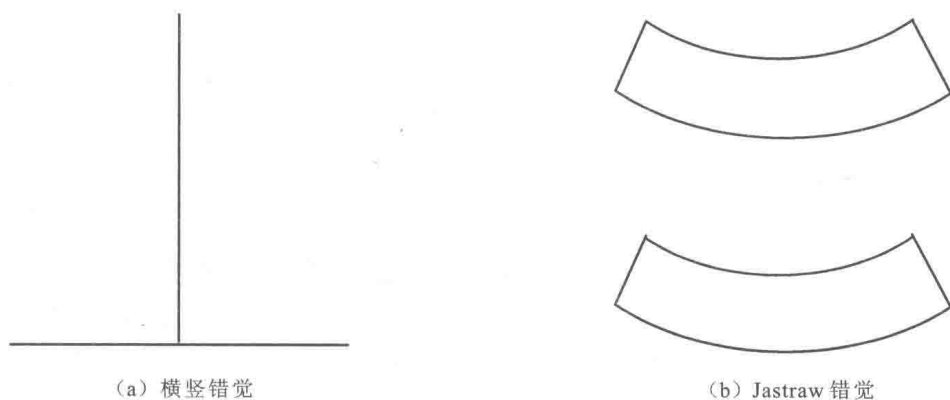


图 1.1 人类视觉系统的错觉

为此,先了解一下人类视觉系统的运行机制是必要的.在人类视觉中,光感元(sensing element)是眼睛.通过眼睛,图像由视神经(optic nerve)传送到大脑进行进一步处理.视神经不具备足够的带宽(bandwidth)来传输眼睛所获取的这些信息.因而,图像在视神经传输前需要进行预处理.人类的视觉系统可以用三个部分来建模.

- (1) 眼睛:这是一个物理模型,因为它的许多功能可以用病理学来确定.
- (2) 神经系统:这是一个实验模型,因为它的功能可以建模,但不能精确判定.
- (3) 处理系统: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,只能通过实验或推理来判断其运行状态.

1.1.1 眼 睛

眼睛的功能是成像,图 1.2 是眼睛的横截面图^[12].视力(vision)要求眼睛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(bacillary muscle)实现的.虹膜(iris)或瞳孔(pupil)就像摄影机的光圈,控制进入眼睛的光亮.虹膜是一个非常脆弱的系统,需要角膜(cornea)或巩膜(scleroses)的保护.角膜的内侧是脉络膜(choroid),由血管提供营养,而且是不透明的,便于减少光亮.视网膜(retina)在眼球内,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(forgave)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。

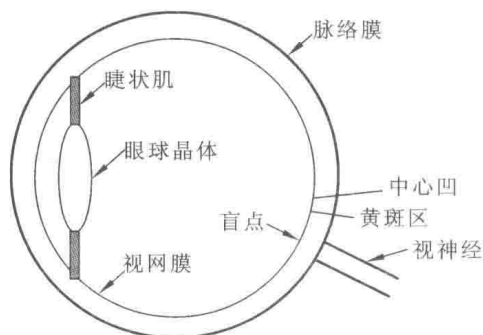


图 1.2 眼睛的横截面图

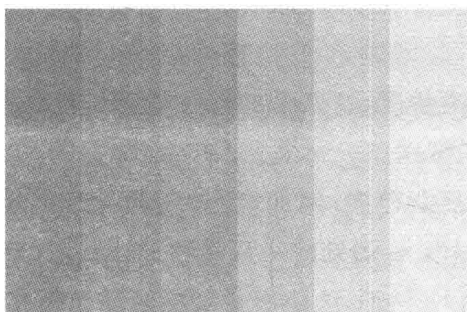
聚焦需要改变晶体的形状,而不是如摄像机那样定位.晶体形状变化实际上是通过拉伸作用使较近的物体图像变大而使较远的物体图像变小.晶体的焦点中心距离根据其形状的不同在 14~17 mm 变化.这就是说,一个三维场景被转换成一个 2 mm^2 左右的平面.好的视力具有高视敏度(acuity),意味着成像区附近有大量传感元.

视网膜周围散布着 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,用于黑白视觉;另一类是视锥细胞,用于彩色视觉.视锥细胞有 1 千万个左右,分布在中央凹 5° 范围以内,其余 1 亿个视杆细胞分布在视网膜周围,大多位于中央凹 $5^\circ\sim 20^\circ$.视敏度表现为空间分辨率和亮度/色彩分辨率,在中央凹 1° 范围内最大.

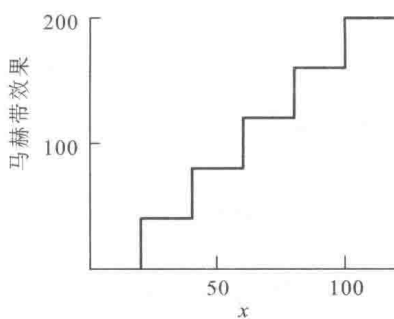
视杆细胞只有一种,而视锥细胞有三种,它们分别可以感应蓝色光、绿色光和红色光.视锥细胞的总体响应是这三种类型的视锥细胞响应的总和,涵盖整个可视光谱范围.视杆细胞对可视光谱全区的光非常敏感,具有暗视觉的单色能力.因此,当光很微弱时,利用视杆细胞的高敏感度,图像在远离中央凹的区域形成,但没有视锥细胞的彩色视觉.

眼睛有一种先天特性,即马赫带(Mach band)效果,影响人们感知图像的方式.马赫带是如图 1.3(a)所示的那些条带,即一系列紧挨着的具有不变色调的样条.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 1.3(b)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

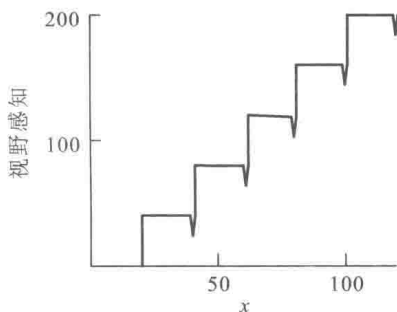
到的图像是如图 1.3(c)所示的横截面. 这些马赫带其实并不存在,是由人们的眼睛产生的,是不同亮度区域的边界在人眼睛中的过冲(overshoot)响应(它有助于人们区分视界内的不同目标). 值得注意的是,人类眼睛只能区分相当少的



(a) 马赫带效果图



(b) 图 (a) 的实际横截面



(c) 图 (a) 的感知横截面

图 1.3 马赫带效果示意图