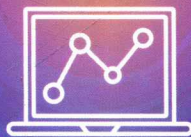




工业和信息化部“十二五”规划教材

# 计算机视觉



## Computer Vision

刘绍辉 姜 峰 / 编著

 中国工信出版集团

 电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>

工业和信息化部“十二五”规划教材

# 计算机视觉

刘绍辉 姜 峰 编著

电子工业出版社

**Publishing House of Electronics Industry**

北京·BEIJING

## 内 容 简 介

计算机视觉是目前研究最为活跃的领域之一，很多新的技术和方法在计算机视觉中得到了成功的应用。本书以计算机视觉相关技术和模型为主线，讨论当前这个领域的传统技术和方法。全书共分8章。第1章介绍了人类视觉系统及其计算模型、JND模型和显著性模型。第2章介绍了图像的基本表示，以及底层特征，包括边缘、角点、几何形状的检测。第3章介绍了基本的色度学及颜色模型，并对图像形成过程进行了详细介绍。随后介绍了齐次坐标及坐标变换相关的知识，最后介绍了相机位置、方向和姿态估计。第4章介绍了从图像序列中估计2D和3D运动。第5章介绍了差分运动分析及基于核函数的视觉跟踪。第6章介绍了蒙特卡罗(Monte Carlo)运动分析。第7章介绍了铰链运动分析及人体姿态估计。第8章介绍了多目标跟踪算法。

计算机视觉相关技术在实际生活中有着广泛的应用，并在最近几年取得巨大进展，尤其是深度学习技术的发展使更多人对计算机视觉领域产生了兴趣。本书可供计算机科学与技术、软件工程、多媒体处理和信号处理等领域中关注计算机视觉、图像处理、模式识别及其应用的工程技术人员和科研教学人员阅读，也可以作为研究生和大学高年级学生学习的教材和参考书。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

### 图书在版编目(CIP)数据

计算机视觉 / 刘绍辉, 姜峰编著. —北京: 电子工业出版社, 2019.3

ISBN 978-7-121-35548-6

I. ①计… II. ①刘… ②姜… III. ①计算机视觉 IV. ①TP302.7

中国版本图书馆CIP数据核字(2018)第259596号

策划编辑: 章海涛

责任编辑: 章海涛 文字编辑: 孟宇 特约编辑: 张帆

印刷: 北京捷迅佳彩印刷有限公司

装订: 北京捷迅佳彩印刷有限公司

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: 14.75 字数: 378千字

版次: 2019年3月第1版

印次: 2019年3月第1次印刷

定价: 52.00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, 88258888。

质量投诉请发邮件至 [zltz@phei.com.cn](mailto:zltz@phei.com.cn)，盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式：192910558 (QQ群)。

# 前 言

自马尔提出计算机视觉理论以来,计算机视觉技术作为与现实生活紧密联系的重要研究内容,引起了国内外学术界的普遍关注。随着计算能力和技术的发展,其实际应用越发促进了计算机视觉技术的进一步发展。从最开始的实时人脸识别到目标检测与跟踪,从传统的手工提取特征到基于深度学习方式的提取特征,计算机视觉技术正从理论研究及在可控环境中的应用逐步向更广泛的普及应用扩展。

本书叙述了计算机视觉相关的一些基本理论和技术,主要包括人类视觉系统的建模、JND模型和显著性模型、图像的形成过程及相关的坐标变换、图像的底层特征提取与检测、图像中物体运动与关联分析等。由于篇幅有限,本书对深度学习内容介绍较少,后续版本会增加深度学习的内容。本书内容及材料主要来自自己公开发表的文献、书籍、网上博客,以及作者所在单位的硕士生、博士生的相关工作。本书对相关的基本原理有较详细的介绍,具有较好的实用性,对本领域研究人员和科技工程人员均有较好的参考价值。

本书第1、2、3章由刘绍辉撰写,其余各章由姜峰撰写。全书由刘绍辉、姜峰统稿。在本书的撰写和校稿过程中,哈尔滨工业大学计算机科学与技术学院智能接口与人机交互技术研究中心的博士生和硕士生做了大量工作,在此一并表示感谢。特别感谢电子工业出版社的章海涛主任,正是在他的鼓励和帮助下才得以完成全书的撰写。

限于作者水平,书中不足之处在所难免,敬请广大读者批评指正。

作 者

2018年9月

于哈尔滨工业大学

# 目 录

第 1 章 人类视觉系统及其建模 .....	1
1.1 人类视觉系统概述 .....	1
1.2 人眼基本视觉特性 .....	4
1.2.1 色彩空间 .....	4
1.2.2 多通道特性 .....	5
1.2.3 亮度自适应 .....	5
1.2.4 对比度敏感度函数 .....	5
1.2.5 视觉掩蔽效应 .....	6
1.3 立体视觉的形成过程 .....	7
1.3.1 立体视觉系统的生理特性 .....	7
1.3.2 立体视觉系统的心理特性 .....	9
1.3.3 立体视觉生理特性与心理特性的关系 .....	9
1.4 JND 模型与显著性模型 .....	10
1.4.1 JND 模型 .....	10
1.4.2 显著性模型 .....	18
1.5 本章小结 .....	19
参考文献 .....	19
第 2 章 图像边缘、角点检测 .....	20
2.1 图像 .....	20
2.2 灰度位图图像和彩色位图图像 .....	21
2.3 GIF 图像格式和 JPEG 压缩 .....	24
2.3.1 GIF 图像 .....	25
2.3.2 JPEG 图像 .....	25
2.4 图像边缘及其检测 .....	31
2.4.1 边缘类型 .....	32
2.4.2 边缘检测的三个阶段 .....	33
2.4.3 滤波操作及双边滤波器 .....	33
2.4.4 差分操作 .....	36
2.4.5 边缘检测操作 .....	37
2.4.6 非极大值抑制操作 .....	37
2.4.7 几种典型的边缘检测算子 .....	38

2.5	图像角点检测	41
2.5.1	图像角点检测基本原理及 Harris 角点检测	41
2.5.2	FAST 角点检测	44
2.6	形状检测	44
2.6.1	标准 Hough 变换及圆形 Hough 变换	44
2.6.2	广义 Hough 变换	46
2.6.3	三种常见 Hough 变换的区别	47
2.7	直线段检测	47
2.8	本章小结	49
	参考文献	50
<b>第 3 章</b>	<b>图像形成与相机几何</b>	<b>51</b>
3.1	引言	51
3.1.1	色度学	51
3.1.2	三色视觉原理	52
3.2	图像的形成	53
3.2.1	取景透镜	54
3.2.2	抗混叠滤波器	57
3.2.3	红外截止滤波器和保护玻璃	59
3.2.4	图像传感器	59
3.3	齐次坐标	64
3.4	小孔成像	65
3.5	图像坐标系、相机坐标系和世界坐标系	67
3.5.1	图像坐标系	67
3.5.2	相机坐标系	68
3.5.3	世界坐标系	68
3.6	坐标平移、伸缩和旋转	68
3.6.1	坐标平移	68
3.6.2	坐标伸缩	69
3.6.3	坐标旋转	69
3.6.4	绕任意轴旋转的矩阵表示	71
3.6.5	罗德里格斯公式	74
3.6.6	四元数	75
3.6.7	旋转矩阵与旋转角的关系	76
3.6.8	矩阵与运动的对应关系	77
3.6.9	世界坐标系到图像坐标系的变换	78
3.6.10	透视相机模型	80
3.7	相机标定	80

3.7.1	相机畸变	83
3.7.2	相机标定	84
3.8	相机位置和方向	84
3.9	姿态估计	86
3.10	本章小结	88
	习题 3	88
	参考文献	88
<b>第 4 章</b>	<b>从图像序列中估计 2D 和 3D 运动</b>	<b>89</b>
4.1	运动场与光流场	89
4.2	光流计算	91
4.2.1	光流方程	91
4.2.2	经典光流算法	92
4.2.3	光流算法的改进	94
4.3	光流技术的研究难点及策略	100
4.3.1	鲁棒的光流估计	101
4.3.2	压缩感知及基于稀疏模型的光流估计	103
4.3.3	光流分布信息的统计研究	105
4.4	3D 运动恢复	107
4.4.1	透视投影成像的几何模型	107
4.4.2	3D 运动场	109
4.4.3	3D 运动场到 2D 速度场的投影转换模型	111
4.4.4	2D 速度场的局部拟合及 3D 运动参数估计	112
4.5	本章小结	114
	参考文献	115
<b>第 5 章</b>	<b>差分运动分析及基于核函数的视觉跟踪</b>	<b>121</b>
5.1	基于差分方法的光流计算	121
5.1.1	光流通用模型	121
5.1.2	模型数据项	122
5.1.3	HS 模型的计算方法	127
5.2	视觉跟踪概述	129
5.2.1	视觉跟踪的分类	130
5.2.2	视频目标跟踪算法的组成	131
5.3	核函数跟踪算法	131
5.3.1	核函数跟踪算法的问题及发展	132
5.3.2	目标及候选图像区域表达	132
5.3.3	相似性测度	133
5.3.4	概率密度估计	134

5.3.5	均值漂移算法	137
5.3.6	误差平方和测度及牛顿迭代	141
5.3.7	多极值点问题	142
5.4	本章小结	143
	参考文献	143
<b>第6章</b>	<b>蒙特卡罗运动分析</b>	<b>146</b>
6.1	跟踪问题的形式化表示	146
6.1.1	Markov 性假设	146
6.1.2	三个基本模型	146
6.1.3	推测过程	147
6.2	卡尔曼滤波与广义卡尔曼滤波	147
6.2.1	状态空间表示法和参数估计	147
6.2.2	标准卡尔曼滤波	148
6.2.3	广义卡尔曼滤波	149
6.3	特征表示与提取	150
6.3.1	颜色直方图	150
6.3.2	形状特征	151
6.3.3	尺度不变轮廓特征 的表示	151
6.4	目标跟踪方法评价指标	153
6.5	序列 Monte Carlo 方法研究	154
6.5.1	Monte Carlo 方法	155
6.5.2	重要性采样	157
6.5.3	序列重要性采样	159
6.5.4	退化问题	161
6.5.5	粒子滤波的改进方法	165
6.5.6	粒子滤波的收敛性	171
6.6	本章小结	175
	参考文献	176
<b>第7章</b>	<b>铰链运动分析及人体姿态估计</b>	<b>180</b>
7.1	人体模型及观测似然函数	180
7.1.1	人体模型	180
7.1.2	观测似然函数	181
7.2	模拟退火的粒子滤波算法	181
7.2.1	粒子滤波算法	181
7.2.2	模拟退火算法	182
7.2.3	模拟退火的粒子滤波	183
7.3	非参数置信传播算法	183



7.4	人体运动估计	184
7.4.1	条件随机场理论	185
7.4.2	人体与外观模型	190
7.5	本章小结	192
	参考文献	192
<b>第 8 章</b>	<b>多目标跟踪算法</b>	<b>194</b>
8.1	多目标跟踪概述	194
8.2	数据关联算法	197
8.2.1	最近邻域法	197
8.2.2	概率数据关联算法	197
8.2.3	联合概率数据互联算法	199
8.2.4	其他的经典算法	201
8.3	基于图模型的多目标跟踪算法	202
8.3.1	概率图模型分析	203
8.3.2	概率图模型推理方法	213
8.4	本章小结	224
	参考文献	225

## 人类视觉系统及其建模

### 1.1 人类视觉系统概述

计算机视觉的研究对象之一是如何利用二维投影图像恢复三维景物世界,其基本目的可以归结为从单幅或多幅二维投影图像(或视频序列)计算出观察点和目标对象之间的空间位置关系及目标对象的物理属性。例如,包括目标对象与观察点的距离(即深度信息)、目标对象的运动特性和表面物理特性等。计算机视觉最终的目标就是实现计算机对于客观世界的理解,从而使得计算机可实现人类视觉系统的某些功能。

人类视觉系统(Human Visual System, HVS)是一个非常复杂的系统,至今还没有被完全地理解,而且大多数的人类视觉系统视觉特性不是凭直觉获知的。人类视觉系统是人们理解和认知自然世界的关键工具,是揭示大脑秘密的一个突破点。所以,了解人类视觉系统的生理结构,分析信息在大脑中的传递过程和形成处理过程是研究计算机视觉的必要前提。深入研究人类视觉系统对视觉刺激的处理机制,合理地利用这种人眼视觉特性,才有可能在更高层次上研究真正意义的计算机视觉算法和系统。

正因为如此,才使得脑科学成为目前研究界的热点。在大脑接收来自外部世界的大量信息中,绝大部分是通过视觉系统进行加工处理的。视觉信息加工机制是心理学、神经科学、计算机科学等学科研究的重大课题之一。人类的视觉系统是目前已知的功能最完备、机制最复杂的信息加工系统之一,对它的研究无疑对了解人类自身信息加工的能力和机理,对理解大脑中所表达的信息和外部物质世界的关系等问题具有重大的意义。此外,计算机科学的发展突飞猛进,创造具有人类大脑那样的智能计算机是研究者最大的期望之一。为使这一期望成为现实,首先就要使计算机具有大脑那样处理大量复杂信息(特别是视觉信息)的能力。这个重大课题虽然经过计算机科学家们几十年的努力但仍未得到解决,其中最重要的原因之一就是迄今为止我们仍然没有透彻了解人类视觉系统的工作机制。

不过,随着脑科学的蓬勃发展,人类对自身视觉系统的研究逐步深入,从初级视觉皮层到高级视觉区域,从视觉感知到高级视觉机理等,都取得了许多重要的研究成果。

现代脑科学和神经科学的发展使得我们得以更进一步地了解人类视觉系统的构造以及功能。视网膜引出的视觉神经的传递路径及外侧膝状体、视觉皮层的很多工作机理已经逐渐为人们所了解。初级视觉皮层中神经细胞的感受及相应的神经编码机理也已经被揭示出来。视觉系统提供给我们的信号是经过多级处理的,在处理的过程中,丢失了相当多的信息,剩下的信息在尺度和强度上与视网膜神经细胞感受到的信息有着显著的不同。

视觉生理学研究已经表明，视觉信息处理过程包括 4 个方面：光学处理、视网膜处理、外侧膝状体（LGN）处理和视觉皮层处理，其原理图如图 1.1 所示。

光学处理是通过眼睛完成的，眼睛的感光系统如图 1.2 所示。它的主要功能相当于“相机”。相应地，巩膜类似于球形相机的保护壳和暗箱，把眼球整个包围起来。同时角膜在集中照明上扮演着重要的角色。虹膜被认为是一个孔径，控制着瞳孔的大小，瞳孔负责调节视网膜上的亮度，同时也影响着系统的焦距。镜头就像晶状体，视网膜相当于胶卷。最后，光线集中在视网膜上，形成一个清晰的物体图像。现代相机的各个功能部件都可以在人眼中找到相对应的部分，但人眼远比相机科学、灵活。例如，人眼可以靠直接调整透镜的曲率来调整焦距，也可以通过眼球外侧的 6 块肌肉的运动来控制眼球的视线指向，产生双目视差以形成深度知觉。

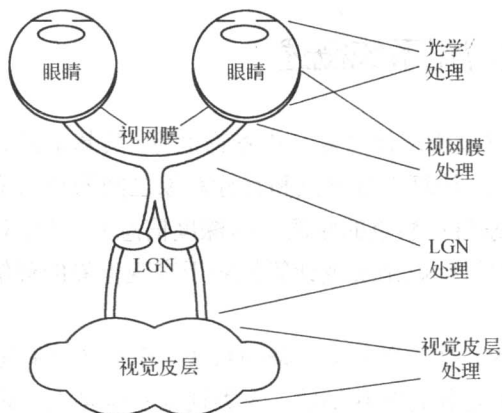


图 1.1 视觉信息处理原理图

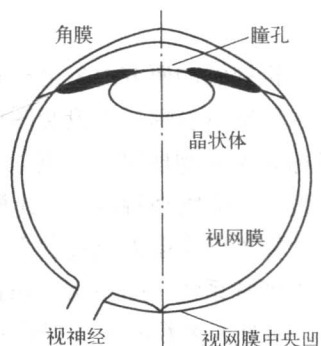


图 1.2 眼睛的感光系统

研究显示，视网膜主要由三种神经细胞构成，即感光细胞、双极细胞和神经节细胞，它们负责光、电转换和信息传输。感光细胞主要包括视锥细胞和视杆细胞，视锥细胞主要在强光下辨别强光信息，称为明视觉；而视杆细胞主要对低照度的景物较敏感，称为暗视觉。人类的明视觉和暗视觉的特性不同，表现在人眼对明、暗视觉下最敏感的可见光波长不同。双极细胞的作用是负责联络视细胞，即可以使多个视细胞相互联系。神经节细胞位于最内层，专门负责传导。这种细胞与细胞之间的联系，以及视觉信号通过眼睛最后在大脑中形成影像并理解影像内容的处理流程就成为计算机视觉的一个重要参考标准。神经网络，包括现在流行的深度学习技术，都是对这种流程进行模拟的尝试，并在近几年中，取得了很好的效果，如深度学习在图像分类、目标检测与跟踪、行为分析与理解等方面都获得了空前的成功。

通过视网膜，光信号被编码为电压脉冲，再以调频形式传递给 LGN。LGN 作为信号从视网膜到视觉皮层的传输站，同时对控制信息数量起着重要作用。最后，视觉皮层实现了对物体的识别、感知与理解的过程。

总体来说，外界物体在视网膜成像时，实际过程是：光线这个刺激因素被视网膜的感光细胞（视杆细胞和视锥细胞）转变为电信号，后者经视网膜内双极细胞传到神经节细胞形成神经冲动，即视觉信息，视觉信息再经视神经传向大脑。双极细胞可以看成视觉传导通路的第 1 级神经元，神经节细胞是第 2 级神经元，很多神经节细胞发出的神经纤维可以组成较粗大的视神经。LGN 是视觉信息的中转站，视辐射可由 LGN 中含有的第 3 级神经元组成，最后将这些神

经纤维投射到视觉皮层中。视觉传导神经通路如图 1.3 所示。

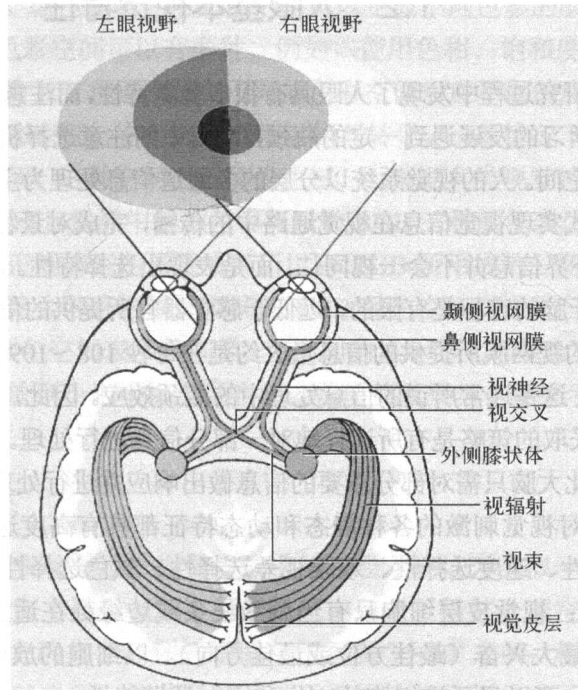


图 1.3 视觉传导神经通路

视觉皮层中 17 区被称为第一视区 (V1) 或纹状皮层。它接受外侧膝状体的直接输入, 因此也被称为初级视觉皮层。对视觉皮层的功能研究大多数是在这一级皮层进行的。这是大脑皮层处理视觉信息的起点, 从初级视觉皮层开始, 视觉信息通过多个通道把视觉信息传入更高级的皮层进行处理。V1 区主要包括两类神经细胞: 具有朝向选择性的简单细胞和对位置不敏感的复杂细胞。作为整个视觉皮层的底层部分, V1 区的神经细胞的功能是, 将底层视路的信息转换成初级视觉信息表示, 并且直接向高级视觉皮层输送视觉信息。现代计算机视觉和视觉信息的表示都是基于人类这种视觉特性而进行构建的, 两者都是对人类视觉系统的模拟。如在计算机视觉中, 采用 RGB 三基色来表示自然界的各种颜色, 实际上, 其基本原理就来自于感光细胞中的三类细胞: 红敏细胞、绿敏细胞和蓝敏细胞。并且绿敏细胞的数量是最多的, 因此在任何颜色空间的转换中, 亮度信号中绿色分量也就是 G 分量对应的系数是最大的。

基于对人类视觉系统生理上的不断理解和深入研究, 所获得的新的认知原理都可能应用在计算机视觉的各个方面, 从而使得新型计算机视觉技术更加适应人类的认知习惯, 更加符合人类的感知特性。例如, 在 3D 视觉中, 3D 电影和电视近年来不断刷新人类的认知, 但大家普遍感觉戴 3D 眼镜来观看 3D 电影并不是一种很好的体验, 尤其是长时间观看 3D 电影, 可能会对人类的视觉造成损伤。这表明目前的 3D 电影虽然也是通过模拟人类视觉系统来进行研究并设计的产品, 但可能对于人类与此有关的一些视觉特性并没有研究透彻。因此促使研究者更加深入地研究这个问题, 目的是能设计出更符合人类视觉特性的 3D 电影和电视。可以预见, 随着技术的进步, 计算机视觉系统的设计将会越来越更真实地模拟人类的视觉功能, 并在实际应用中更加体现出超越人类的性能。

## 1.2 人眼基本视觉特性

在人类对人脑的研究过程中发现了人眼具有很多视觉特性,而注意选择机制是其中尤为重要的特性,近年来深度学习的发展遇到一定的瓶颈,而其中的注意选择机制则异军突起,为深度学习打开了一个广阔的空间。人的视觉系统以分层的多通道信息处理为主要特征,它以在时间域和空间域逐级整合的方式实现视觉信息在视觉通路中的传播,完成对景物的概念化抽象。然而,在处理过程中,人脑对外界信息并不会一视同仁,而是表现出选择特性。这有两方面的原因:第一,可用资源的限制,由于脑的容积是有限的,远低于感觉器官所提供的信息总量,这在视觉系统尤为重要(据估计,人的视网膜所提供的信息量大约是在每秒 108~109 位,而大脑皮层细胞的总数仅为 108~109 个),这是通常所说的信息处理中的瓶颈效应。因此,要实时地处理全部信息是不可能的,视觉系统采取的策略是有所选择地对一部分信息进行处理。第二,由于外界环境信息并不全部都重要,因此大脑只需对部分重要的信息做出响应并进行处理即可。

视觉皮层神经元对视觉刺激的各种静态和动态特征都具有高度选择性,包括方位/方向选择性、空间频率选择性、速度选择性、双眼视差选择性、颜色选择性。

**方位/方向选择性:**视觉皮层细胞只有当刺激线条或边缘处在适宜的方位角并按一定的方向移动时,才表现出最大兴奋(最佳方位或最佳方向)。以细胞的放电频率相对于刺激方位和运动方向做成直方图,可以显示该细胞的方位和方向调谐特性。

**空间频率选择性:**正弦波调制的光栅是视觉实验中经常使用的刺激图形。用这种刺激图形的主要优点是便于对视觉反应的时空特性进行定量的数学分析。每个视觉皮层细胞都有一定的空间频率调谐。在同一皮层区内,不同细胞也有不同的空间频率选择性。

**速度选择性:**视觉皮层细胞对移动图形的反应比对静止的闪烁图形要强得多。每个皮层细胞不仅对运动的方向有选择性,而且要求一定的运动速度。只有当刺激图形在适宜的方向上以一定速度移动时,细胞反应才达到最大,这个速度称为该细胞的最佳速度。当移动速度高于或低于最佳速度时,反应都会减小。

**双眼视差选择性:**与外侧膝状体细胞不同,大部分视觉皮层细胞接受双眼输入。因此,每个细胞在左、右视网膜上都有一个感受野,这一对感受野在视网膜上的位置差(相对于注视点)称为“视差”。若左、右感受野与注视点的距离差为零,则表示该点正好在注视平面上;若两个感受野都向额侧偏离,则表示该细胞的调谐距离(最佳距离)比注视点远;若两个感受野向鼻侧偏离,则意味着该细胞的调谐距离比注视点近。

**颜色选择性:**同视网膜和外侧膝状体神经元一样,皮层细胞也具有颜色选择性。与皮层下的单颞颥式感受野不同,视觉皮层细胞的颜色感受野具有双颞颥式结构。例如,对于 R-G(红—绿)型感受野来说,其颜色结构可能有两种形式。感受野中心可能被绿敏视锥细胞的输入兴奋,同时被红敏视锥细胞输入抑制,或者相反。外周对颜色的反应性质正好与中心相反。因此,双颞颥式感受野通过中心的颜色颞颥能分辨红色和绿色,通过中心与外周之间的相互作用能使红—绿对比的边缘得到增强。对于 B-Y(蓝—黄)型感受野,情况也一样。

### 1.2.1 色彩空间

色彩是指人眼能根据光的不同频率而产生的不同感受。人眼能够识别色彩的原因是因为有

能够吸收光的不同波长范围的三种视锥细胞，而这三种视锥细胞能辨别红色、绿色和蓝色这三种颜色。我们知道把红色、绿色和蓝色搭配在一起可以生成不同色彩的颜色，这就是一个色彩空间，如 RGB。另外，色彩空间可以有多种，例如，使用色相、饱和度与明度来呈现一个色彩空间，这种方法被命名为 HSI 色彩空间。不同的色彩空间对应不同的应用，因此我们可以根据需要选择使用不同的色彩空间，它们之间也可以相互转化。一般情况下，色彩空间覆盖了自然界中绝大部分的颜色空间，这种空间一般称为色域。而根据人类视觉特性设计出来的色彩空间（如 RGB 空间）是，假定自然界中的颜色都可以被 RGB 三种颜色表达出来。而实际情况并不是这样，因为色域不是正规的三角形区域，所以由 RGB 所形成的三角形的色彩空间只是包括了自然界色域空间中的绝大部分区域。不同的 RGB 颜色空间对应了自然界色域空间中的不同区域。这样，采用不同颜色空间表示的图像，在不同的显示器上可能会呈现出不一样的效果，例如，苹果手机上的图像放到小米手机上观察，会发现其效果与在苹果手机上的效果不一样。具体可参见 3.1 节的内容。

### 1.2.2 多通道特性

视觉生理学和心理学实验显示，视觉皮层中的神经元被认为类似一个有方向的带通滤波器，能够在不同频率和方向上进行分解。视觉系统中包含了能够处理空间频率的单元，也称通道。例如，人的黑白视觉的几个倍频的通道存在于  $30^\circ \sim 60^\circ$  之间；相似的关于人眼的彩色视觉通道存在于  $60^\circ \sim 130^\circ$  之间。这些通道相当于把原始信号划分成子带后再进行处理。因此，在常见的计算机视觉处理中，采用数学变换来处理图像，形成不同频率成分的系数或子带，并进行后续的处理。典型的变换如傅里叶变换，离散余弦变换和离散小波变换，都可以对输入的图像进行频谱的划分处理，从而模拟视觉系统中的这种多通道特性，对不同通道的信号进行不同的处理，如数据压缩就是对高频分量进行粗糙的量化处理。

### 1.2.3 亮度自适应

人的视觉系统对光的适应范围是很宽的，大致范围约为  $10^{-2} \sim 10^6 \text{cd/m}^2$ 。在背景照明不变的情况下，人的视觉的感光范围很窄，它可以根据光的强度来适当调节。当人眼适应某个环境亮度后，人的视觉会产生一个变动，随之调节到一个较小的范围，这就是亮度适应现象。

当人的视觉在适应背景照明不变的情况下，人眼能感知到对黑白色彩的范围缩小了。因此当图像重现时，即使图像重现的亮度与原本的实际景物的亮度不同，也能够保持重现图像和原本的实际图像之间亮度的相对比值，人们就能感觉到同样的真实感觉。

在连续背景下对噪声监测阈值的衡量表明，人类视觉和听觉都具有一定的掩蔽效应，也就是其察觉的亮度或响度对其周围的亮度或响度有掩蔽的效果。在人类视觉模型中，这种阈值衡量是一个非线性函数并依赖于局部的图像特征。一般而言，背景越亮或者越暗，人眼对该区域的敏感程度就越低，这就是亮度掩蔽特性。

### 1.2.4 对比度敏感度函数

对比度是一种度量亮度相对变化的量，大致可认为，对比度正比于激励信号的相对量度幅度。通常激励信号的颜色、时间频率、空间频率都与人眼对比度的敏感度有关，而对比度敏感度函数（CSF）是定量描述这种关系的。当时域频率为零时，人类视觉系统的空间对比度敏感度

函数被定义为调制转移函数。调制转移函数的研究指出，人类视觉系统对静止图像的空间频率响应表现为带通特性。因此，空间对比度敏感度可以使用带通滤波器来模拟。归一化后的对比度敏感度函数如图 1.4 所示。人眼的对比度敏感度不仅存在空间中，而且存在时域上，因此也可以通过简单的时域滤波来实现。

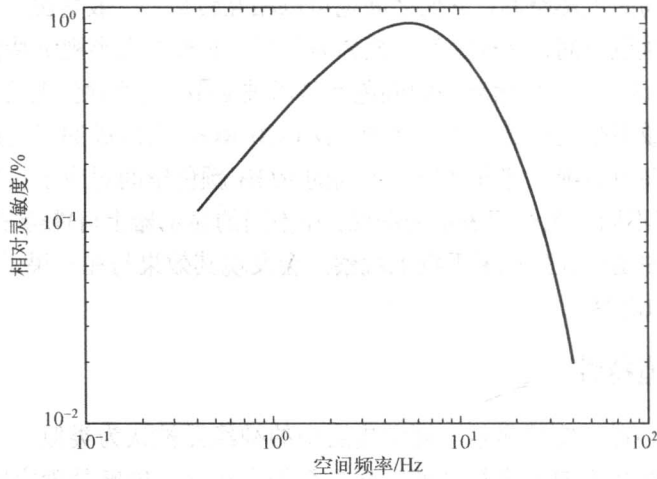


图 1.4 归一化后的对比度敏感度函数

Kelly 通过实验测量了不同频率下的对比度敏感度的数据。公式为

$$CSF(f_s, f_t) = 4\pi^2 f_s f_t e^{-4\pi(f_t + 2f_s)/45} \times \left( 6 + 7 \left| \log \frac{f_t}{3f_s} \right|^3 \right) \quad (1.1)$$

其中， $f_s$ 和 $f_t$ 分别为激励的空间频率和时间频率。Kelly 又发现，CSF 可以用两个时空分离的部分线性组合近似逼近，这样可以降低模型的计算复杂度。很多真实模型都是用非线性方式表达的，但绝大部分算法都采用这种线性近似的方式对其进行逼近，或者采用二次逼近。其本质原理就是将原来的非线性函数进行泰勒展开，根据需要取其一阶、二阶直到给定阶数来进行近似。由于这种近似具有很好的收敛性质，因此该方法是解决实际工程问题的通用方法。

### 1.2.5 视觉掩蔽效应

在对 CSF 的研究过程中，为了使问题简化，假设视觉激励信号是一个常量或单一的频率信号。在观看图像时，很多激励信号会对人眼产生作用，此时人眼对一个激励的响应不仅受激励信号本身影响，而且同一时刻也会与其他激励信号的影响有关。在一个视觉激励存在的情况下，人类视觉系统会在其他激励影响下改变当前激励的可见阈值，这种现象称为视觉掩蔽效应。

常见的视觉掩蔽效应是 19 世纪的实验心理学家 Ernst Weber 在实验过程中发现的，人们能觉察到的背景强度的增量阈值（又称为刚好可区分的差异）与背景强度的比值是一个常量，这个关系就是韦伯定律。公式为

$$\frac{\Delta I}{I} = k \quad (1.2)$$

其中， $\Delta I$  表示增量阈值； $I$  表示刺激的原始强度； $k$  表示等式左侧的比例关系为常量，不会因为  $I$  的大小而变化。比值  $\Delta I / I$  就是韦伯比(Weber Fraction)，又称为费克纳比(Fechner Fraction)。

韦伯定律指出刚好可区分的差异 ( $\Delta I$ ) 与原刺激值的大小的比例是常量。例如, 如果你在一个嘈杂的环境中, 那么你必须放大音量才能让别人听见你说话, 但是在一个非常安静的环境中你只需要耳语就足够了。类似地, 当你测量不断变化的背景刚好可区分的增量阈值  $\Delta I$  时, 该阈值  $\Delta I$  会与原始强度  $I$  的大小成正比。

### 1.3 立体视觉的形成过程

随着技术的发展, 立体图像和视频的应用也越来越广泛, 例如, 近年来的 3D 电影、电视技术逐渐普及, 这些应用促进了立体视觉的研究。立体视觉的产生大致分为三个过程。首先, 给定构成立体图像的图像对 (具有标准视差): 左图像和右图像, 在此阶段, 双眼要同时观察这两幅图像; 其次, 视觉系统会通过观察到的这两幅图像, 经过一系列的复杂处理融合成一幅图像; 最后, 视觉系统会结合心理因素将平面图像的信息转化成立体信息, 最终产生立体视觉。到目前为止, 经过研究者不断探索, 视觉生理学和视觉心理学已经获得了很多的研究发现。这也是今年 3D 电影和电视技术逐渐得到推广的原因。当然, 由于对生理学和人类视觉心理学的研究仍然在进行中, 因此, 目前的 3D 电影和电视技术可能还无法与真正的人类视觉系统相媲美, 仍有很大的改进空间。

#### 1.3.1 立体视觉系统的生理特性

生理立体视觉是由人眼的晶状体调节、运动视差、双眼会聚、双眼视差和融合图像等因素构成的立体视觉。人眼的晶状体调节是指外界图像在视网膜上成像的过程, 它是根据睫状体的收缩和放松完成这个过程。晶状体有自适应调节焦距的功能, 根据不同的远近景, 晶状体会通过改变形状来使进来的光线聚焦在视网膜上。当观看近景时, 晶状体的弧度变弯曲, 此时睫状肌的状态是收缩的; 当观看远景时, 晶状体的弯曲程度降低, 此时睫状肌处于扩张状态, 且屈光度数随之减小, 以便使来自远处的光线恰好聚焦在视网膜上。

在深度认识的过程中, 一方面是来自日常生活观察和经验的累积。例如, 人们平时对看到的物体的大小和形状的认知, 对纹理和结构的认知, 对光线产生的阴影和遮挡的认知, 对物体运动情况的认知, 等等。通过了解人们日常观察的认知和经验, 这些视觉线索被人们认为是距离信息或者深度信息, 从而营造了一种深度感。另一方面, 人的双眼是分开的, 双眼瞳孔间的水平距离约为 6.5cm, 如图 1.5 所示。

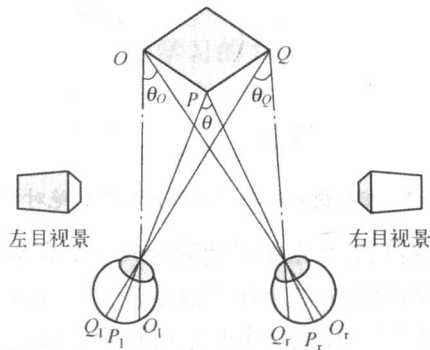


图 1.5 双目视差



当我们集中注意一个物体时，双眼就会将视线同时聚于该物体，此时物体就会在两眼视网膜上的相应位置成像。但由于双眼间的差异，两个眼睛中接收的图像会有略微的差异，这种差异就称为视差。同样一个物体，当我们只用左眼观看和只用右眼观看时，会发现物体转动一定角度并向旁边移动了一些。我们从双眼中观察到图像的差异称为双目视差，根据这种差异就会产生立体的深度感。

当观察目标物体时，眼球内转使双眼视轴交汇于注视目标，这个过程称为会聚，如图 1.6 所示。外界信号会通过大脑传递给眼肌，而眼肌会通过控制眼球会聚到目标点。适应性调节和双眼会聚的共同作用才能完成一个注视的动作，大脑会通过认知来融合稍有差异的左右眼中的图像使之具有立体感。融合则是指把两个视网膜的对应点上的物像整合成完整的符合人的印象的功能。

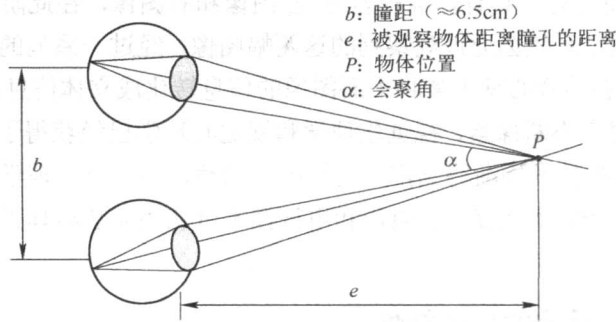


图 1.6 会聚过程

双目视差是使人眼产生立体视觉的因素，它分为相对视差和绝对视差。当在不同位置观看物体时，参考左右两眼的视网膜中央凹，绝对视差是指左右两眼上的两个投影点形成的角度。而它们的绝对视差之差就是两点之间的相对视差。如图 1.7 所示，相对视差是  $\alpha - \beta$ ，绝对视差是  $\alpha$ 、 $\beta$ 。

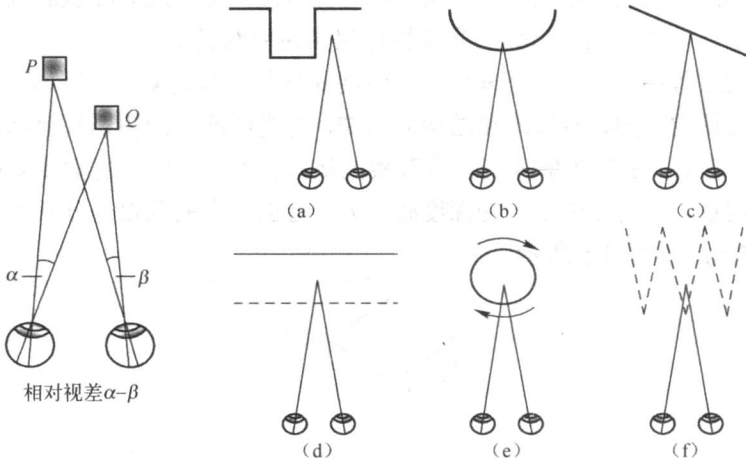


图 1.7 相对视差和不同观察目标下的绝对视差

由于当分别用左右眼去观察左右两张图时能融合成一个完整的三维图像，因此这说明视差携带了深度信息。在利用立体图对的方法来研究深度视觉问题时，Bela Julesz 在美国贝尔电话实验室进行了实验。他开始时利用计算机产生两张相同的随机点的图，并将其中一张中的一点图像的随机点水平位移一个距离，这就构造了一对具有视差的图对。然后用双眼去观看，就会