

中国计算机学会学术著作丛书

计算机视觉

——算法与系统原理

高 文 陈熙霖

清华大学出版社

广西科学技术出版社

(京)新登字 158 号

(桂)新登字 06 号

内 容 简 介

计算机视觉研究的主要目的是让计算机能够利用图象和图象序列来识别和认知三维世界,其最终目标是让计算机具有“视觉”功能,以满足社会对计算机高级应用的需求。根据研究内容可以把计算机视觉划分为计算理论、表达与算法和系统实现三个层次。本书系统地介绍了计算机视觉的计算理论和算法的原理,给出了主要算法的证明,介绍了计算机视觉系统的构成原理。为了便于读者将理论与实现进行对照和进行自己的应用系统设计,本书还介绍了到目前为止国际上比较著名的计算机视觉系统。

本书可作为计算机科学、人工智能、图象处理与模式识别、人机交互、智能机器人、信息处理以及认知科学等有关专业的大学高年级学生和研究生的教学和自学教材使用,也可供以上领域中的科研工作者使用。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

计算机视觉:算法与系统原理/高文,陈熙霖著. —北京:清华大学出版社,1998

(中国计算机学会学术著作丛书)

ISBN 7-302-03232-7

I. 计… II. ①高… ②陈… III. ①计算机视觉-算法理论 ②计算机视觉-计算方法
IV. TP302.7

中国版本图书馆 CIP 数据核字(98)第 35768 号

出版者:清华大学出版社(北京清华大学校内,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者:北京市密云胶印厂

发行者:新华书店总店北京发行所

开 本:787×1092 1/16 印张:19.5 字数:460 千字

版 次:1999 年 2 月第 1 版 1999 年 2 月第 1 次印刷

书 号:ISBN 7-302-03232-7/TP·1729

印 数:0001~2000

定 价:29.00 元

清华大学出版社 广西科学技术出版社
计算机学术著作出版基金

评 审 委 员 会

主 任 委 员 张效祥

副主任委员 唐泽圣 汪成为

委 员	王鼎兴	杨芙清	李三立	施伯乐	徐家福
	夏培肃	董韫美	黄 健	焦金生	

出版说明

近年来,随着微电子和计算机技术渗透到各个技术领域,人类正在步入一个技术迅猛发展的新时期。这个新时期的主要标志是计算机和信息处理的广泛应用。计算机在改造传统产业,实现管理自动化,促进新兴产业的发展等方面都起着重要作用,它在现代化建设中的战略地位愈来愈明显。计算机科学与其它学科的交叉又产生了许多新学科,推动着科学技术向更广阔的领域发展,正在对人类社会产生深远的影响。

科学技术是第一生产力。计算机科学技术是我国高科技领域的一个重要方面。为了推动我国计算机科学及产业的发展,促进学术交流,使科研成果尽快转化为生产力,清华大学出版社与广西科学技术出版社联合设立了“计算机学术著作基金”,旨在支持和鼓励科技人员,撰写高水平的学术著作,以反映和推广我国在这一领域的最新成果。

计算机学术著作出版基金资助出版的著作范围包括:有重要理论价值或重要应用价值的学术专著;计算机学科前沿探索的论著;推动计算机技术及产业发展的专著;与计算机有关的交叉学科的论著;有较大应用价值的工具书;世界名著的优秀翻译作品。凡经作者本人申请,计算机学术著作出版基金评审委员会评审通过的著作,将由该基金资助出版,出版社将努力做好出版工作。

基金还支持两社列选的国家高科技重点图书和国家教委重点图书规划中计算机学科领域的学术著作的出版。为了做好选题工作,出版社特邀请“中国计算机学会”、“中国中文信息学会”帮助做好组织有关学术著作丛书的列选工作。

热诚希望得到广大计算机界同仁的支持和帮助。

清华大学出版社
广西科学技术出版社 计算机学术著作出版基金办公室

1992年4月

序 言

计算机是当代发展最为迅猛的科学技术,其应用几乎已深入到人类社会活动和生活的一切领域,大大提高了社会生产力,引起了经济结构、社会结构和生活方式的深刻变化和变革,是最为活跃的生产力之一。计算机本身在国际范围内已成为年产值达 2500 亿美元的巨大产业,国际竞争异常剧烈,预计到本世纪末将发展为世界第一大产业。计算机科技具有极大的综合性质,与众多科学技术相交叉而反过来又渗入更多的科学技术,促进它们的发展。计算机科技内容十分丰富,学科分支生长尤为迅速,日新月异,层出不穷。因此在我国计算机科技尚比较落后的情况下,加强计算机科技的传播实为当务之急。

中国计算机学会一直把出版图书刊物作为学术活动的重要内容之一。我国计算机专家学者通过科学实践,做出了大量成果,积累了丰富的经验与学识。他们有撰写著作的很大积极性,但相当时期以来计算机学术著作由于印数不多,出版往往遇到不少困难,专业性越强越有深度的著作,出版难度越大。最近清华大学出版社与广西科学技术出版社为促进我国计算机科学技术及产业的发展,推动计算机科技著作的出版工作,特设立“计算机学术著作出版基金”,以支持我国计算机科技工作者撰写高水平的学术著作,并将资助出版的著作列为中国计算机学会的学术著作丛书。我们十分重视这件事,并已把它列为学会本届理事会的工作要点之一。我们希望这一系列丛书能对传播学术成果、交流学术思想、促进科技转化为生产力起到良好作用,能对我国计算机科技发展具有有益的导向意义,也希望我国广大学会会员和计算机科技工作者,包括海外工作和学习的神州学人们能积极投稿,出好这一系列丛书。

中国计算机学会

1992 年 4 月 20 日

前 言

1. 计算机视觉研究的目的与意义

人类研究计算机的目的，是为了提高社会生产力水平，提高生活质量，把人从单调复杂甚至危险的工作中解脱出来。今天的计算机在计算速度上已经远远超过了人，然而在很多方面，特别是在与人类智能活动有关的方面例如在视觉功能、听觉功能、嗅觉功能、自然语言理解功能等方面，还不如人。这种现状，无法满足一些高级应用的要求。例如，我们希望计算机能够及早地发现路上的可疑情况并提醒汽车驾驶员以避免发生事故，我们更希望计算机能够帮助我们进行自动驾驶，目前的技术还不足以满足诸如此类高级应用的要求，还需要更多的人工智能的理论研究成果和系统实现的经验。在人工智能领域，尤其是在计算机视觉领域，已有哪些成果，其原理是什么，还有哪些问题需要研究等等，是本书要介绍的主要内容。

什么是人工智能呢？人工智能，是由人类设计并在计算机环境下实现的模拟或再现人的某些智能行为的技术。一般认为，人类的智能活动可以分为两类：感知行为与思维活动。模拟感知行为的人工智能研究的一些例子包括语音识别、话者识别等与人类的听觉功能有关的“计算机听觉”，物体三维表面的形状知识、距离、速度感知等与人类的视觉有关的“计算机视觉”，等等。模拟思维活动的人工智能研究的例子包括符号推理、模糊推理、定理证明等与人类的思维过程有关的“计算机思维”，等等。

从图象处理与模式识别发展起来的计算机视觉研究的对象之一是如何利用二维投影图象恢复三维景物世界。计算机视觉使用的理论方法主要是基于几何、概率和运动学计算与三维重构的视觉计算理论，它的基础包括射影几何学、刚体运动力学、概率论与随机过程、图象处理、人工智能等理论。计算机视觉要达到的基本目的至少有以下几个：

- (1) 根据一幅或多幅二维投影图象计算出观察点到目标物体的距离；
- (2) 根据一幅或多幅二维投影图象计算出目标物体的运动参数；
- (3) 根据一幅或多幅二维投影图象计算出目标物体的表面物理特性；
- (4) 根据多幅二维投影图象恢复出更大空间区域的投影图象。

计算机视觉要达到的最终目的是实现利用计算机对于三维景物世界的理解，即实现人的视觉系统的某些功能。什么是计算机视觉研究的本质问题？如果用一句话来概括，就是利用二维投影图象来重构三维物体的可视部分。

人类智能系统是一个完整的功能系统，是一个整体。它不能被分割成毫不相干的几个子系统来单独运行。这与目前人工智能通常采用的习惯出发点是根本不同的。人

工智能目前处于比较初级的研究阶段，所研制和开发的对象都是独立运行的功能子系统，例如，听觉子系统、视觉子系统、触觉子系统、推理子系统等等。一般认为，只有当这些单独的子系统的研究都比较彻底，理论体系比较成熟，人的智能活动的内部机理全部搞清楚之后，才有可能在较高层次上研究真正意义下的人工智能系统。然而，这种观点的最大问题是人工智能系统的工作机理与人类智能系统的工作机理完全不同。人类智能系统中的感知行为是多通道协同工作的。换句话说，每一个通道负责获取一部分信息，这一部分信息对于完成环境或事件的理解也许是不完全的，但所有信息通道得到的信息总和(加上历史信息和背景信息)却是完全的。人工智能研究中的子系统总是被希望做成一个完美的系统，一个不需要和其他通道合作就可以独立工作的完整系统。这样就要求该通道所提供的信息必须是完全的，而这与我们日常生活中的信息背景是不一样的，是对信息源提出的过分苛刻的要求。为了达到这样的要求，要么制造一些与现实环境相差甚大的理想数据，要么施加种种约束条件。但不管怎样，这样的系统一定是不很实用的和脆弱的。作者希望读者现在就建立一个观念：不要认为我们下面将要介绍的计算机视觉系统是孤立的和万能的，而是在不同的应用背景中它应该与不同的感知通道相配合，例如用在人类通信中可以与语音通道相配合，用在发现和跟踪目标中可以与激光和超声波等技术相配合，等等。

为了达到计算机视觉的目的，有两种技术途径可以考虑。第一种是仿生学方法，即从分析人类视觉的过程入手，利用大自然提供给我们的最好参考系——人类视觉系统，建立起视觉过程的计算模型，然后用计算机系统实现之。第二种是工程方法，即脱离人类视觉系统框框的约束，利用一切可行的和实用的技术手段实现视觉功能。此方法的一般做法是，将人类视觉系统作为一个黑盒子对待，实现时只关心对于某种输入，视觉系统将给出什么样的输出。在大量统计实验的基础上，得到一个较为令人满意的脉冲响应函数，然后利用工程的方法实现。这两种方法从理论上都是可以使用的，但面临的困难是，人类视觉系统对应于某种输入的输出到底是什么，这是无法直接测得的。因为不像在动物实验时可以做到的那样，我们不能在人的脑细胞或者视网膜或者视神经之上插入电极进行刺激响应的测试，因而无法利用生理学实验的方法得到任何实际的脉冲响应对。再者，即使可以用动物(例如狗)的视觉系统做上述实验，然而对于什么是视觉输入的真正信号，什么是噪音，实际上很难区分。而且由于人的智能活动是一个多功能系统综合作用的结果，即使是得到了一个输入输出对，也很难肯定它是仅由当前的输入视觉刺激所产生的响应，而不是一个与历史状态综合作用的结果。由于仿生学方法的进展较缓，在本书中，我们讨论的内容大多数属于工程方法。

不难理解，计算机视觉的研究具有双重意义。其一，是为了满足人工智能应用的需要，即用计算机实现人工的视觉系统的需要。这些成果可以安装在计算机和各种机器人上，使计算机和机器人能够具有“看”的能力。其二，视觉计算模型的研究结果反过来对于我们进一步认识和研究人类视觉系统本身的机理，甚至人脑的机理，也同样具有相当大的参考意义。

2. 本书的写作过程

从 1991 年的秋季学期，本书的第一作者开始给计算机科学系的研究生和本科生开设“计算机视觉”课程。当时，很想找一本合适的中文版教材，但没能找到，最后使用了一本日文的著作作为教材的蓝本，再加上一些相关的英文论文。因为学生中第一外语学日语的很少，所以当时作者是一边翻译一边讲课的。从那时起作者就计划写一本计算机视觉方面的教材。后来，从 1992 年的秋季开始，作者先后得到了清华大学智能系统国家开放实验室和国家自然科学基金委的支持，从事计算机视觉模型方面的研究，并在此方向上开始培养硕士生和博士生，教材和研究生参考书的双重需求使作者下决心要尽早动笔。本书的实际写作是 1992 年年底开始的。由于当时科研和其他业务工作的任务很重，作者花了一年时间完成了现在您所看到的本书的前三章及第五章的大部分内容，而且这主要是在寒假和暑假中完成的。1994 年和 1995 年，由于承担了一些国家计划的工作和出国进行客座研究，曾使得写作工作一度停止下来。后来，在众多师长与同仁的鼓励和清华大学出版社几位编辑的关心下，我们两位作者又用了两年的时间，总算使这本书脱稿。

计算机视觉是一个发展中的学科方向，其理论体系还不完备，新理论、新算法、新应用还在不断地涌现。目前，每年计算机视觉方面的文章有数千篇，有关的专业国际会议有十几个。因此，现在要想写出一本内容全面的计算机视觉教材和专业参考书是很困难的。尽管我们知道现在动笔很可能是挂一漏万，但是为了教学和科研的需要，也实在不能等到理论体系完备了以后再动笔。

3. 致谢

在本书出版之际，我们向支持本研究的清华大学智能系统国家开放实验室和国家自然科学基金委表示感谢，没有他们的支持就不会有本书的问世。另外，很多人为本书的写作作出了贡献。本书第三章中的第 1 节，第 3 节，第 4 节和第五章部分内容的草稿由黄庆明博士执笔，第七章的大部分内容的草稿由战德臣博士执笔。尹宝才博士、张晶博士、赵德彬博士以及刘岩副教授等二十几位同事和学生都曾参加过有关内容的研究和讨论，他们对本书的写作起到了很大的帮助作用，在此表示衷心的感谢。

高文 陈熙霖

1998 年 4 月于北京

目 录

第一章 概论	1
1.1 人类视觉系统的构成与视觉机理	2
1.1.1 眼睛	3
1.1.2 视觉神经系统	4
1.1.3 视觉机理假说	6
1.1.4 视觉中的一些心理和生理特征与现象	8
1.1.5 视知觉对深度的感知	13
1.2 计算机视觉研究的特点	15
1.3 计算机视觉与相关领域的关系	16
1.4 计算机视觉的发展	17
第二章 计算机视觉中的空间关系	20
2.1 基本概念	20
2.2 成象模型与视觉坐标系	21
2.3 齐次坐标与 N 矢量	23
2.4 平面对偶原理	27
2.5 直射变换、对射变换与标准极变换	28
2.6 平移运动	37
2.6.1 N 速度与轨迹	37
2.6.2 平移运动的出现点与从平移恢复形状	38
2.7 旋转运动	41
2.7.1 旋转变换	41
2.7.2 旋转矩阵的计算	42
2.8 一般刚体运动	45
2.8.1 一般刚体运动的坐标变换	46
2.8.2 平面投影图象的变换	48
2.8.3 变换参数的计算	52
2.8.4 面参数的计算	54
2.8.5 一般刚体运动分析	56
第三章 立体视觉	59
3.1 一般性原理	59
3.2 内极点、内极线与内极平面	60
3.2.1 一般情况	60

3.2.2 平行立体视中的内极线约束.....	63
3.3 立体视中的配准.....	65
3.3.1 双目配准.....	65
3.3.2 三目配准.....	75
3.4 三维表面重建.....	78
第四章 运动分析.....	83
4.1 运动场与光流.....	83
4.1.1 运动场.....	83
4.1.2 光流.....	84
4.2 基于梯度的光流计算.....	84
4.2.1 基本等式的引入.....	85
4.2.2 光流计算与附加约束.....	85
4.3 基于区域匹配的光流计算方法.....	92
4.3.1 基于保守信息的初始速度场估计.....	92
4.3.2 利用邻域信息的求精.....	93
4.3.3 采用 Kalman 滤波器的速度场估计.....	94
4.4 光流计算的其他方法.....	96
4.5 从运动恢复结构.....	98
4.5.1 空间运动与对应视平面的运动场.....	98
4.5.2 平面对象的光流.....	100
第五章 基于其他线索的三维信息恢复.....	108
5.1 从阴影恢复三维形状.....	108
5.1.1 成象过程的数学模型.....	108
5.1.2 反射图方法.....	110
5.1.3 球状点假设方法.....	111
5.2 从纹理恢复三维形状.....	121
5.2.1 基于纹理梯度的方法.....	122
5.2.2 基于收敛线的方法.....	126
5.2.3 基于正规化纹理属性图的方法.....	126
5.2.4 由形变估计表面方向.....	128
5.3 融合阴影与纹理的立体视觉.....	131
5.3.1 光度立体技术.....	131
5.3.2 纹理立体技术.....	131
第六章 不确定性视觉计算.....	133
6.1 视觉逆问题及其不确定性.....	134
6.1.1 成象过程与视觉计算.....	134

6.1.2 逆问题与不确定性.....	135
6.2 不确定性的描述与求解	136
6.2.1 有限晶格系统, MRF 与 Gibbs 分布	136
6.2.2 估计构造	138
6.2.3 问题求解	142
6.2.4 参数估计	143
6.3 基于不确定性方法的边缘检测.....	146
6.3.1 模型的建立	146
6.3.2 估计的求解	149
6.4 基于不确定性方法的立体视差计算.....	150
6.4.1 构造估计问题	150
6.4.2 WTA 网络.....	151
6.5 小结.....	153
第七章 高层表示	154
7.1 二维图象的表示	154
7.1.1 边界表示	154
7.1.2 区域表示	156
7.2 三维物体的几何模型	156
7.2.1 物体的结构表示	157
7.2.2 物体的多视特征表示.....	164
7.3 空间物体的关系模型	166
7.3.1 语义网的概念	166
7.3.2 属性超图	170
7.3.3 语义网的操作	172
7.3.4 其他知识表达方法介绍.....	181
第八章 主动视觉与融合	186
8.1 主动视觉.....	186
8.1.1 从阴影恢复形状	187
8.1.2 从运动恢复结构	193
8.1.3 主动跟踪	197
8.1.4 通过变形模板提取面部特征.....	198
8.2 视觉计算与融合技术	203
8.2.1 融合处理中需考虑的问题.....	204
8.2.2 不同级融合的一个例子.....	206
8.2.3 融合的一般方法	207
8.2.4 物体识别与信息融合.....	214

第九章 计算机视觉系统	219
9.1 引言	219
9.2 三维物体识别	219
9.2.1 三维物体识别概念	219
9.2.2 物体识别系统的组成及特性	220
9.2.3 物体识别的模型组织与特性	221
9.3 视觉系统中的若干问题	226
9.3.1 视觉系统中的控制	226
9.3.2 视觉系统中的推理与规划	227
9.3.3 视觉系统中的学习	233
9.4 三维计算机视觉系统	234
9.4.1 ACRONYM——基于模型驱动的视觉系统	234
9.4.2 INGEN——用于一般形状物体识别的视觉系统	237
9.5 通用图象理解环境	243
9.5.1 KBVision——基于知识的图象理解环境	243
9.5.2 IUE——通用图象理解环境	251
参考文献	254
附录一 算法证明	264
1.1 算法 2.3 的证明	264
1.2 算法 2.4 的计算过程	271
附录二 IUE 中类的谱系图	275
2.1 IUE 基本类谱系图	275
2.2 IUE 图象类谱系图	278
2.3 IUE 空间对象类谱系图	279
2.4 IUE 图象特征类谱系图	286
2.5 IUE 坐标系与变换类谱系图	288
2.6 IUE 空间索引类谱系图	290
2.7 IUE 传感器类谱系图	291
2.8 IUE 传感器模型类谱系图	293
2.9 IUE 射线类谱系图	294
2.10 IUE 镜头类谱系图	295
2.11 IUE 滤波器类谱系图	295
2.12 IUE 景物类谱系图	297
2.13 IUE 统计类谱系图	298

第一章 概 论

任何一门学科、一个理论、一种方法的产生与发展，都是由人类试图认识自然、认识人类世界、认识自我和改造自然、改造世界、改造自我这个根本要求所推动的。而人类这种要求的水平，是不能脱离当时所处时代的科学技术水平的。

人工智能，是由人类设计并在计算机环境下实现的模拟或再现人的某些智能行为的技术。可以相信，人类的智能行为至少包括两类：感知行为与思维活动。模拟感知行为属于人工智能研究的范畴包括语音识别、话者识别等与人类的听觉功能有关的“计算机听觉”，以及印刷与手写文字识别、图象模式识别和三维表面形状知识与距离、速度感知等与人类的视觉有关的“计算机视觉”，等等。人工智能研究中的模拟思维活动研究包括符号推理、模糊推理、定理证明等与人类的思维过程有关的“计算机思维”，等等。

本书所涉及的范畴，是人工智能中计算机视觉的有关部分。

任何一个学科领域都有自己明确的研究方向。本书首先向读者说明的问题是：计算机视觉研究的对象是什么，使用什么样的理论方法，所要达到的目的是什么，可能应用到哪些方面。

计算机视觉研究的对象之一是如何利用二维投影图象恢复三维景物世界。计算机视觉使用的理论方法主要是基于几何、概率和运动学计算与三维重构的计算视觉理论，它的基础包括射影几何学、刚体运动力学、概率论与随机过程、图象处理、人工智能等理论。计算机视觉要达到的基本目的至少有三个：

1. 根据一幅或多幅二维投影图象计算出观察点到目标物体的距离；
2. 根据一幅或多幅二维投影图象计算出目标物体的运动参数；
3. 根据一幅或多幅二维投影图象计算出目标物体的表面物理特性。

要达到的最终目的是实现对于三维景物世界的理解，即实现人的视觉系统的某些功能。什么是计算机视觉研究的本质问题？如果用一句话来概括就是，利用二维投影图象来重构三维物体的可视部分。

人类智能系统是一个完整的功能系统，是一个整体。它不能被分割成毫不相干的几个子系统来单独运行。这与目前人工智能所通常采用的习惯出发点从根本上是不同的。人工智能目前处于比较初级的研究阶段，所研制和开发的对象都是独立运行的功能子系统，例如听觉子系统，视觉子系统，触觉子系统，推理子系统等等。一般认为，只有当这些单独的子系统的研究都比较彻底，理论体系比较成熟，人的智能活动的内部机理全部搞清楚之后，才有可能在较高层次上研究真正意义下的人工智能系统。然而，这种观点的最大问题是与人类智能系统的工作机理完全不同。人类智能系统中的感知行为是多通道协同工作的。换句话说，每一个通道负责获取一部分信息，这一部分信息对于完成

环境或事件的理解也许是不完全的。但所有信息通道得到的信息总和(加上历史信息和背景信息)却是完全的。人工智能研究中的子系统,总是被希望做成一个完美的系统,一个不需要和其他通道合作就可以独立工作的完整系统,否则它就没有办法演示,就会不被人相信。这样就要求该通道所提供的信息必须是完全的,而这与我们日常生活中的信息背景是不一样的,是对信息源提出的过分苛刻的要求。为了达到这样的要求,要么制造一些与现实环境相差甚大的理想数据,要么施加种种约束条件。但不管怎样,这样的系统一定是不很实用的和脆弱的。这一点,作者希望读者现在就建立一个观念:不要认为我们下面将要介绍的计算机视觉系统是孤立的和万能的,当用在不同的应用背景中它应该与不同的感知通道相配合,例如用在人类通信中应与语音通道相配合,用在发现和跟踪目标中应与激光和超声波等技术相配合,等等。

上面已经提到,计算机视觉研究的最终目标是实现对三维景物世界的理解。为了达到这个目的,有两种技术方法可以考虑。第一种是仿生学方法,即从分析人类视觉的过程入手,利用大自然提供给我们的最好参考系——人类视觉系统,建立起视觉过程的计算模型,然后用计算机系统实现之。第二种是工程方法,即脱离人类视觉系统框框的约束,利用一切可行的和实用的技术手段实现视觉功能。例如使用激光扫描或超声波扫描,这是人类所无法做到的。此方法的一般做法是,将人类视觉系统作为一个黑盒子对待,实现时只关心对于某种输入,视觉系统将给出什么样的输出。在大量统计实验的基础上,得到一个较为令人满意的脉冲响应函数,然后利用工程的方法实现。这两种方法从理论上都是可以使用的,但面临的困难是,人类视觉系统对应于某种输入的输出到底是什么,这是无法直接测得的。因为不像在动物实验时可以作到的那样,我们不能在人的脑细胞或者视网膜或者视神经之上插入电极进行刺激响应的测试,因而我们无法利用生理学实验的方法得到任何实际的脉冲响应对。再者,即使我们可以用动物(例如狗)的视觉系统做上述实验,然而对于什么是视觉输入的真正信号,什么是噪音,实际上是很难区分的。而且由于人的智能活动是一个多功能系统综合作用的结果,即使是得到了一个输入输出对,也很难肯定它是仅由当前的输入视觉刺激所产生的响应,而不是一个与历史状态综合作用的结果。这也是为什么计算机视觉的研究一直特别困难的主要原因。

计算机视觉的研究具有双重意义。其一,是为了满足人工智能应用的需要,即用计算机实现人工的视觉系统的需要。这些成果可以安装在各种机器人之上,使机器人能够具有“看”的能力。其二,视觉计算模型的研究结果反过来对于我们进一步认识人类视觉系统本身也具有相当的参考意义。

人类视觉系统是计算机视觉系统的原始模型。因此,我们有必要首先介绍一下到目前为止生理学和心理学在人类视觉系统的构成和视觉机理方面的研究结果。

1.1 人类视觉系统的构成与视觉机理

人类的视觉系统是由眼睛和视觉神经系统构成的。

1.1.1 眼睛

现已清楚，人的眼睛的结构是由如图 1.1 所示的构造所组成的。眼球被直径约 24mm 的球状巩膜所包围着。巩膜的作用相当于一个保护壳和球状暗箱。球状暗箱前端的中间开口处是晶状体。晶状体相当于照相机镜头上的凸透镜片。晶状体可以在肌肉的作用下改变形状(厚薄)。这相当于照相机镜头的凸透镜的曲率发生变化，从而可以调整焦距，以使被注视物体能够刚好落在视网膜的表面上，使成象最清晰。熟悉照相的读者都知道照远景要用长焦距镜头(例如 200mm 镜头)，照中景要用中焦距镜头(例如 70mm 镜头)，照近景要用短焦距镜头(例如 35mm 镜头)。但由于一种镜头的焦距是固定的，所以除了靠换镜头进行粗调整外，还要通过拉伸镜头与底片的距离进行微调。人的眼睛要比照相机科学，是靠直接调整透镜的曲率来调整焦距。当然，如果一个人眼睛的调整机能低下或失调，人就要带眼镜了。

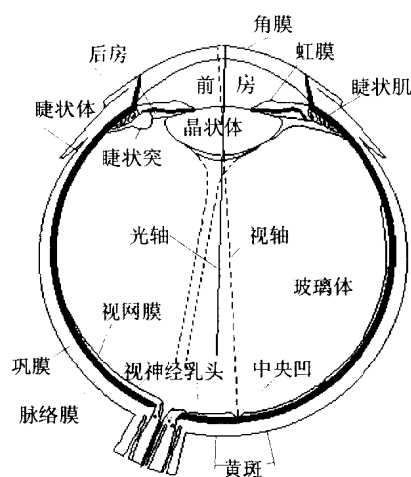


图 1.1 眼睛的构造

晶状体前面是虹膜。虹膜形成瞳孔，起到照相机光圈的作用。它可以根据外界光线的强度来调整暗箱口开放的大小以使进入视网膜的光线的量产生的刺激不弱也不强。眼睛的底部是视网膜。视网膜的作用是将光信号变换、滤波和编码成神经系统的内部表达信号(电信号)以传送给视觉神经系统和中枢神经系统。

眼睛的构造除了新陈代谢需要的一些功能部分以外，可以说其余部分都可以与现在所使用的照相机中的各功能部件找到一一对应的关系。但是，人类眼睛比照相机更灵活的地方是，照相机在对准目标时是要靠人工调节或机械驱动装置辅助人工来实现，而眼睛是靠眼球外侧的六块肌肉的运动来自动调整。这些肌肉移动眼睛使两只眼睛的视线指向某一特定位置，并产生双目的视差以利于深度知觉。值得一提的是，现在市场上一些高级的照相机已经有了自动调焦的功能，这实际上可以说我们人类在人工调焦这个问题上已经找到了合适的模型或者脉冲响应函数。

1.1.2 视觉神经系统

视觉神经系统的作用是将视网膜上得到的内部编码信号传送给大脑。已经清楚,信号的主传送途径是通过视神经送到外侧膝状体,然后再由上丘经视放线送给大脑皮层的视觉区。其主要传输路径以及所经过的主要功能块如图 1.2 所示。

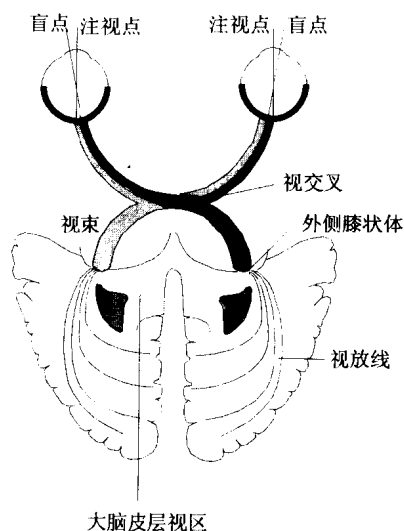


图 1.2 视觉神经系统

人类的两只眼睛的视神经为什么是分别连接在大脑的两个不同的半脑上(左半脑和右半脑)而不是连在一个半脑上,为什么眼球与视神经系统又是交叉相连而不是直线相连的呢?目前此问题我们还回答不了,但有人曾提出以下的假说。

深度知觉计算假说:人的深度知觉能力是由视差比较计算而生成的,而视差的计算是基于左、右两个半脑所得到的两眼视网膜的二维投影图象信息而实施的。这仅回答了前半个问题,至于为什么眼球与视神经系统是交叉相连而不是直线相连的问题,目前还没有特别令人满意的解释,这是一个有待于解决的问题。

本书后面几章中将会阐明:视差计算是计算机视觉中的深度知觉的主要方法。因而也可以说深度知觉计算假说是计算机视觉的生理学模型。

人类视觉系统中信息的处理、存储、传递和度量的单位是什么,由视网膜向大脑传递的信息是未被处理(未进行信息抽取)的原始信号还是处理后所得到的高效率的内部表达信号,这是一些有必要探明的问题。我们知道人的神经传导系统的信号传导速度并不快,然而人却能在 $1/5 \sim 1/10$ 秒的时间内对于眼睛所看到的危险处境(例如一块石头向你的面部飞来)做出反应,足见人的视觉系统的信息处理速度是相当的快。我们现在极想知道的问题是视觉处理是在哪一级完成的,是在眼睛本身,还是在视网膜之后,或者是在大脑的视觉区呢?

已经清楚,视网膜组织是由多层构造组成的。所有脊椎动物的视网膜构造大体上相同。图 1.3 是它的一个示意图。该组织从视细胞一侧看起,第一层是视细胞,第二层