

1

绪论

本章对全书内容和结构进行概括介绍,其内容安排如下:

1.1 节首先介绍图像工程的一些概念和定义,然后列举两个相关文献综述的一些统计数据,借以介绍和概括图像工程近年的发展情况。

1.2 节概括介绍图像理解的研究内容和在图像工程中的位置,讨论图像理解与计算机视觉以及与相关学科(如机器视觉/机器人视觉、模式识别、人工智能、计算机图形学等)的联系和区别,并介绍图像理解的一些应用领域。

1.3 节列出本书主要内容、结构和安排,并对各章重要内容给予概括介绍。

1.1 图像工程的发展

1. 基本概念和定义

图像是用各种观测系统以不同形式和手段观测客观世界而获得的,可以直接或间接作用于人眼,并进而产生视知觉的实体[章 1996a]。人的视觉系统是具有这样能力的典型系统。视觉是人类观察世界、认知世界的重要功能手段。人类从外界获得的信息约有 75% 来自视觉系统,这既说明视觉信息量巨大,也表明人类对视觉信息有较高的利用率。正因为如此,人们还制造了许多利用各种辐射对客观场景成像的系统,以利用视觉信息观察世界。图像是表达视觉信息的一种物理形式。数字图像采集的最终结果常是某种能量的样本阵列,所以常用矩阵或数组来表示,其中每个元素的坐标对应场景点的位置,而元素的值对应场景点的某个物理量。

人们用各种技术方式和手段对图像进行加工,以获得需要的信

息。从广义上,图像技术可看作是各种图像加工技术的总称。它包括利用计算机和其他电子设备进行和完成的一系列工作,例如图像的采集、获取、编码、存储和传输,图像的合成和产生,图像的显示、绘制和输出,图像的变换、增强、恢复(复原)和重建,图像的分割,特征的提取和测量,目标的检测、表达和描述,序列图像的校正,图像数据库的建立、索引、查询和抽取,图像的分类、表示和识别,3-D 景物的重建复原,图像模型的建立,图像信息的融合,图像知识的利用和匹配,图像和场景的解释和理解,以及基于它们的推理、判断、决策和行为规划(如何推断出应实现的目标及构造实现目标的操作序列)等。另外,图像技术还可包括为完成上述功能而进行的硬件和系统设计及制作等方面的技术。许多具体技术已在本套书上册和中册进行了介绍。

对图像技术的综合研究和集成应用可在图像工程这个整体框架下进行[章 1996a]。众所周知,工程是指将自然科学的原理应用到工业部门而形成的各学科的总称。图像工程学科则是利用数学、光学等基础科学的原理,结合电子技术、计算机技术及在图像应用中积累的技术经验而发展起来的一个对整个图像领域进行研究应用的新学科。事实上,图像技术多年来的发展和积累为图像工程学科的建立打下了坚实的基础,而各类图像应用也对图像工程学科的建立提出了迫切的需要[章 2000a],[Zhang 2002a]。

图像工程的内容非常丰富,应用也非常广泛,根据抽象程度、研究方法、操作对象和数据量等的不同可分为三个层次(见本套书上册图 1.2.1):图像处理、图像分析和图像理解。图像处理是比较低层的操作,它主要在图像像素级上进行处理,处理的数据量非常大。图像分析则处于中层,分割和特征提取把原来以像素描述的图像转变成比较简洁的非图形式的描述。图像理解主要指高层的操作,基本上根据较抽象的描述进行解析、判断、决策(符号运算),其处理过程和方法与人类的思维推理有许多类似之处。在这里,随着抽象程度的提高,数据量是逐渐减少的。具体说来,原始图像数据经过一系列的处理过程,逐步转化为更有组织和用途的信息。在这个过程中,语义不断引入,操作对象也逐步发生变化。另外,高层操作对低层操作有指导作用,能提高低层操作的效能,完成复杂的任务。

概括地说,图像工程是既有联系又有区别的图像处理、图像分析及图像理解三者的有机结合,并包括对它们的工程应用。从概念上讲,图像工程既能较好地许多相近学科兼蓄并容,也更强调图像技术的应用,所以这里选用图像工程来概括整个图像领域的研究和应用,也使图像处理、图像分析及图像理解三者的关系更紧密。

图像工程是一门系统地研究各种图像理论、技术和应用的新的交叉学科。从它的研究方法来看,它与数学、物理学、生物学、生理学(特别是神经生理学)、心理学、电子学、计算机科学等许多学科可以相互借鉴;从它的研究范围来看,它与模式识别、计算机视觉、计算机图形学等多个专业又互相交叉。另外,图像工程的研究进展与人工智能、神经网络、遗传算法、模糊逻辑等理论和技术都有密切的联系,它的发展应用与生物医学、

材料、遥感、通信、交通管理、军事侦察、文档处理和工业自动化等许多领域也是不可分割的。

图像工程是全面系统研究图像理论方法,阐述图像技术原理,推广图像技术应用及总结生产实践经验的新学科。结合本册书内容,图像工程的主要构成可用图 1.1.1 所示的整体框架来表示,其中虚线框内为图像工程的基本模块。这里要用到各种图像技术以帮助人们从场景获得信息。首先要进行的就是利用各种方式从场景获得图像。接下来对图像的低层处理主要是为了改善图像的视觉效果或在保持视觉效果的基础上减少图像的数据量,处理的结果主要是给用户观看的。对图像的中层分析主要是对图像里感兴趣的目标进行检测、提取和测量。分析的结果能为用户提供描述图像目标特点和性质的数据。最后对图像的高层理解是要通过对图像里各目标的性质和它们之间相互关系的研究,了解把握图像内容并解释原来的客观场景。理解的结果能为用户提供客观世界的信息,从而可以指导和规划行动。这些从低层到高层所用的图像技术都得到了包括人工智能、神经网络、遗传算法、模糊逻辑、图像代数等新理论、新工具、新技术的有力支持。为完成这些工作还要采取合适的策略进行控制。

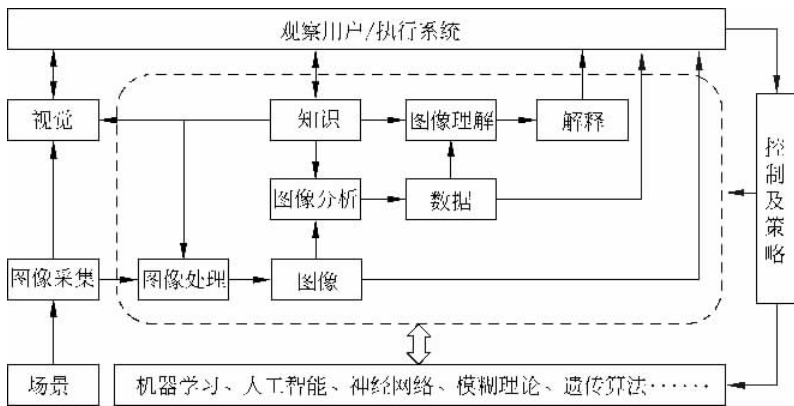


图 1.1.1 图像工程整体框架

本套书上册对低层图像处理的基本原理和技术进行了详细的介绍,本套书中册对中层图像分析的基本原理和技术进行了详细的介绍,本册书的内容主要将包括高层图像理解中有关的视觉理论(视感觉和视知觉),(在处理和分析基础上)对 3-D 客观场景的重建、表达、解释等理解技术,上述过程中的相关知识及应用,以及为完成这些工作所采用的控制和策略等。对高层理解的研究现已成为图像技术研究发展的一个重点,对不同层次图像技术的综合应用推动了图像事业的快速发展。

2. 一个已结束的图像技术综述系列

在 20 世纪最后约 30 年的期间内曾连续发表了一个由 30 篇论文组成的关于图像技

术方面的综述系列。这个系列中综述论文的题目直到 1986 年,均称为“图像处理”,而从 1987 年开始,称为“图像分析和计算机视觉”。该综述系列已于 2000 年由作者结束。作者声称,由于在线信息获取的进展,该综述系列已不必要再延续下去[Rosenfeld 2000]。其 30 年综述的总体情况,包括各综述论文题目、所综述的文献年份和数量、刊登各综述论文的刊物名称和综述论文的发表年份概括在表 1.1.1 中[Zhang 2002a]。

表 1.1.1 已结束综述系列的概况

序号	综述论文题目	引用年	数量	刊登综述论文的刊物	发表年
1	Picture Processing by Computer	—1969 年前	408	ACM Computing Surveys	1969
2	Progress in Picture Processing: 1969—1971	1969—1971	580		1972
3	Picture Processing: 19xx (1972—1986)	1972	350	Computer Graphics and Image Processing (CGIP)	1973
4		1973	245		1974
5		1974	341		1975
6		1975	354		1976
7		1976	461		1977
8		1977	609		1978
9		1978	819		1979
10		1979	700		1980
11		1980	897		1981
12		1981	982		1982
13		1982	1185	Computer Vision, Graphics and Image Processing (CVGIP)	1983
14		1983	1138		1984
15		1984	1252		1985
16		1985	1063		1986
17		1986	1436		1987
18	Image Analysis and Computer Vision: 19xx (1987—1999)	1987	1412	CVGIP: Image Understanding	1988
19		1988	1635		1989
20		1989	1187		1990
21		1990	1611	CVGIP: Image Understanding	1991
22		1991	1178		1992
23		1992	1897		1993
24		1993	1281	Computer Vision and Image Understanding (CVIU)	1994
25		1994	1911		1995
26		1995	1561		1996
27		1996	2148		1997
28		1997	1691		1998
29		1998	2268		1999
30		1999	1693		2000

这个综述系列 30 年里所引用的共 34293 篇文献分别来自 40 多个刊物(大部分选自美国或国际刊物)和 10 几个大型国际会议。由表 1.1.1 可见,除了开始的两篇综述论文,其余当年发表的综述论文都提供了上一年有关文献的清单。这个系列的前两篇综述论文发表在“ACM Computing Surveys”上,其后的 28 篇都发表在“Computer Graphics and Image Processing (CGIP)”和其更名后的刊物上。

例 1.1.1 CGIP 刊物的名称变化

1972 年,一份与图像技术关系密切的国际期刊创刊,它当时的名称是 CGIP(计算机图形学与图像处理)。1983 年它改名为 CVGIP(计算机视觉、图形学与图像处理)。1991 年它分成两个版本,分别称为 CVGIP-GMIP(图形模型与图像处理)和 CVGIP-IU(图像理解)。1995 年它正式变成了两个独立的期刊,前者命名为 GMIP(图形模型与图像处理),后者命名为 CVIU(计算机视觉与图像理解)。□

3. 新的图像工程综述系列

从 1996 年起,开始了一个新的关于图像工程的综述系列[章 1996a],[章 1996b],[章 1997a],[章 1998a],[章 1999a],[章 2000a],[章 2001a],[章 2002a],[章 2003a],[章 2004a],[章 2005],[章 2006a]。与上面已结束的那个综述系列不同,该综述系列的特点不仅是不仅对选取的文献进行了分类,而且对它们还进行了统计、比较和分析,所以除有助于文献检索外,还有助于确定图像工程研究方向和制订工作决策。

下面仅对该综述系列前 10 年的情况进行概括。在 10 年中,综述从 15 种有关图像工程的重要中文期刊(共 941 期)上发表的 18819 篇学术研究和应用技术文献中选取出了 4032 篇属于图像工程领域的文献。这 10 年来每年的文献总数、选取总数和选取率可见表 1.1.2。文献选取率反映了图像工程在各刊所覆盖的专业范围中的相对重要性。回顾 10 年来关于选取率的统计数据,1995 年的选取率仅为约 1/7,1996 年的选取率约为 1/6,其后直到 2002 年一直在 1/5 左右徘徊。2003 年的选取率达到约 1/4,2004 年的选取率首次超过了 1/4。从选取数量上说,2004 年选取的图像工程文献数量已是 1995 年的 4 倍多(文献总数仅为 2 倍多)。这是图像工程方面的研究成果和投稿数历年都在不断增加的结果,也是图像工程学科蓬勃发展的明证。

表 1.1.2 图像工程综述系列中各年度的选取率

年 度	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
文献总数	997	1205	1438	1477	2048	2117	2297	2426	2341	2473
选取总数	147	212	280	306	388	464	481	545	577	632
选取率(%)	14.74	17.59	19.47	20.72	18.95	21.92	20.94	22.46	24.65	25.60

注: □为例题结束符,全书同。

综述对选取的图像工程文献根据其主要内容分别归入了图像处理、图像分析、图像理解、技术应用和综述 5 个大类,然后进一步分入了 21 个专业小类(分类表见本套书中册表 1.1.1)。这里将图像处理、图像分析、图像理解和技术应用这 4 大类文献数量 10 年来在统计文献中所占百分比的变化情况画在图 1.1.2 中(综述评论类总数量较小,统计意义不明显,没有列出)。由图可见,各年各类文献量的比例变化基本上比较稳定,特别是近五年来各大类文献数量的比例保持了固定的次序(曲线互不交叉)。总体来说,图像处理和图像分析的文献数量一直领先(前 5 年持平,后 5 年图像处理更多),技术应用文献基本保持在第 3 位,图像理解文献一直相对较少。由此可见,一方面对图像理解的研究还需加强,另一方面图像理解的研究还大有可为。

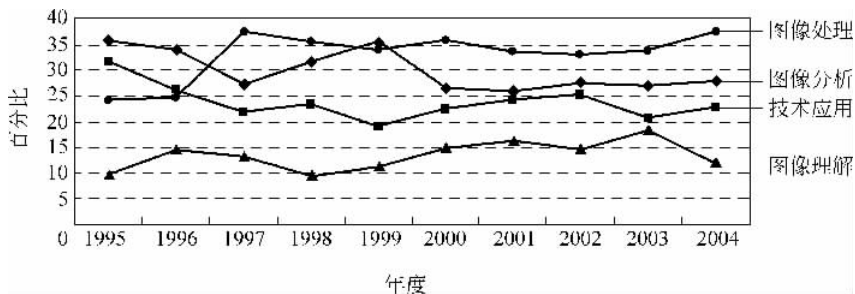


图 1.1.2 4 大类文献数量 10 年来在统计文献中所占百分比的变化情况

1.2 图像理解概述

1.2.1 图像理解

如前所述,图像工程从低到高的三个层次组成一个有机的联合体,它们在从图像获取信息方面都起着重要的作用。但目前对利用高层技术的研究和应用相对比较少[章 1997a],[章 1999b],[章 2000b],[章 2000c],[章 2002c],[章 2005],[章 2006a],[Zhang 2002a],[Zhang 2008]。

人类视觉过程可看作是一个复杂的从感觉(感受到的是对 3-D 世界之 2-D 投影得到的图像)到知觉(由 2-D 图像认知 3-D 世界内容和含义)的过程。视觉的最终目的从狭义上说是要对场景作出对观察者有意义的解释和描述,从广义上讲,还有基于这些解释和描述并根据周围环境和观察者的意愿制定出行为规划。这实际上也就是图像理解的目标。

图像理解作为图像工程的高层,其重点是在图像分析的基础上,结合人工智能和认知理论,进一步研究图像中各目标的性质和它们之间的相互联系,并理解图像内容的含义以及解释对应的客观场景,从而指导和规划行动。如果说图像分析主要是以观察者为中心研究客观世界(主要研究可观察到的事物),那么图像理解在一定程度上是以客观世界为

中心,借助知识、经验等来把握整个客观世界(包括没有直接观察到的事物)。

图像理解关心的是如何根据图像作出对场景的描述和判断,它要用计算机系统来帮助解释图像的含义,从而实现利用图像信息解释客观世界。它要确定为完成某个视觉任务需要通过图像采集从客观世界获取哪些信息,需要通过图像处理和分析从图像中提取哪些信息,以及利用哪些信息继续加工以获得需要的解释。它要研究理解能力的数学模型,并通过对数学模型的程序化,实现理解能力的计算机模拟。

限于目前计算机系统的能力和图像理解技术的水平,上述许多任务还无法完全自动实现。在多数情况下,“系统”完成较低层的工作,而人需要接着系统完成剩下的较高层的工作,如图 1.2.1 所示。

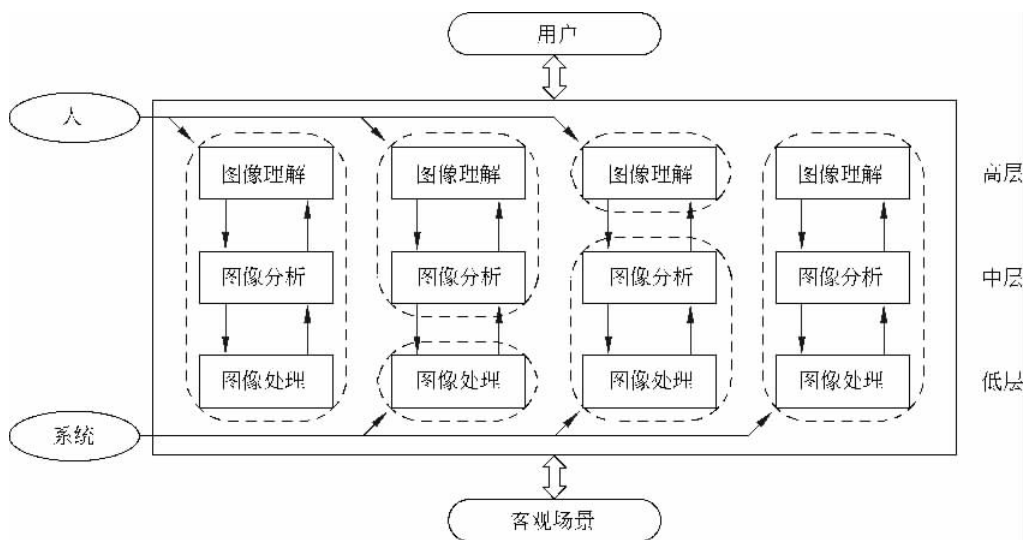


图 1.2.1 系统和用户覆盖不同的层次

图像理解是要用户通过图像认识客观世界。如果完全没有系统,人就必须执行所有各层工作。如果系统只有低层能力,那么人就需要在系统处理的基础上完成中层和高层的工作。如果系统有了低层和中层的能力,那么用户只需要在系统处理的基础上完成高层的工作就可以了。而如果系统有了从低层直到高层的能力,用户很容易就可作出决策。

随着计算机科学和技术的发展,计算机系统正在逐步实现对图像的处理、分析和理解功能,减少人的介入。目前研究和发展的瓶颈主要在理解层次。

1.2.2 计算机视觉

计算机视觉是指用计算机实现人的视觉功能,即对客观世界的三维场景的感知、识别和理解。视觉研究的原始目的是去理解有关场景的图像,辨识和定位其中的目标,确定它

们的结构、空间排列分布以及目标间的相互关系等。计算机视觉的研究目标是根据感知到的图像对实际的目标和场景作出有意义的判断[Shapiro 2001]。

计算机视觉的研究方法主要有两种:一种是仿生学的方法,参照人类视觉系统的结构原理,建立相应的处理模块完成类似的功能和工作;另一种是工程的方法,从分析人类视觉过程的功能着手,并不去刻意模拟人类视觉系统内部结构,而仅考虑系统的输入和输出,并采用任何现有的可行的手段实现系统功能。本书主要讨论第二种方法。

计算机视觉的主要研究目标可归纳成两个,它们互相联系补充。第一个研究目标是建成计算机视觉系统来完成各种视觉任务。换句话说,即要使计算机能借助各种视觉传感器(如 CCD, CMOS 摄像器件等)获取场景的图像,而感知和恢复 3-D 环境中物体的几何性质、姿态结构、运动情况、相互位置等,并对客观场景进行识别、描述、解释,进而作出决断。这里主要研究的是技术机理。目前这方面的工作集中在建成各种专用系统,完成在各种实际场合提出的专门视觉任务,而从长远来说,则要建成通用的系统[Jain 1997]。第二个研究目标是把该研究作为探索人脑视觉工作机理的手段,进一步加深对人脑视觉的掌握和理解(如计算神经科学)。这里主要研究的是生物学机理。长期以来,对人脑视觉系统已从生理、心理、神经、认知等方面进行了大量的研究,但还远没有揭开视觉过程的全部奥秘,可以说对视觉机理的研究和了解还远落后于对视觉信息处理的研究和掌握。需要指出,对人脑视觉的充分理解也将促进计算机视觉的深入研究[Finkel 1994]。本书主要考虑第一个研究目标。

由上可见,计算机视觉要用计算机实现人的视觉功能,同时其研究又从人类视觉那里得到许多启发。计算机视觉方面的许多重要研究都是通过理解人类视觉系统而实现的,典型的例子如用金字塔作为一种有效的数据结构,利用局部朝向的概念,以及通过滤波技术来检测运动。另外,借助对人类视觉系统巨理解能力的研究可帮助人们开发新的图像理解和计算机视觉算法。

图像理解和计算机视觉密切相关。图像是表达视觉信息的一种物理形式,图像理解必须借助计算机,基于对视觉信息的处理和分析来进行。计算机视觉作为一门学科,与许多以图像作为主要研究对象的学科,特别是图像处理、图像分析、图像理解有着非常密切的联系和不同程度的交叉。计算机视觉主要强调用计算机实现人的视觉功能,这中间实际上需要用到图像工程三个层次的许多技术,虽然目前的研究内容主要与图像理解相结合。

在建立图像/视觉信息系统,用计算机协助人类完成各种视觉任务方面,图像理解和计算机视觉都需要用到摄影几何学、概率论与随机过程、人工智能等方面的理论。例如它们都要借助两类智能活动:感知,如感知场景中可见部分的距离、朝向、形状、运动速度、相互关系等;思维,如根据场景结构分析物体的行为,推断场景的发展变化,决定和规划主体行动等。

计算机视觉开始是作为一个人工智能问题来研究的,因此也被称为图像理解[Shah

2002]。事实上,图像理解和计算机视觉这两个名词也常混合使用。从本质上讲,它们互相联系,在很多情况下覆盖面和内容交叉重合,在概念上或实用中并没有绝然的界限。在许多场合和情况下,它们虽各有侧重,但常常是互为补充的,所以将它们看作是专业和背景不同的人习惯使用的不同术语更为恰当,在本书中也不去刻意区分它们。

1.2.3 其他相关学科

图像理解与计算机科学有密切的联系(作为图像理解基础的图像处理和图像分析也与计算机科学有密切的联系),它是随着计算机技术的发展和深入研究而获得突飞猛进发展的。除了计算机视觉,其他一些与计算机相关的学科,如机器视觉/机器人视觉、模式识别、人工智能、计算机图形学等都对图像理解的发展起到了、并将继续起到重要的影响和作用。

机器视觉(machine vision)/机器人视觉(robot vision)与计算机视觉有着千丝万缕的联系,很多情况下都是作为同义词使用的。具体地说,一般认为计算机视觉更侧重对场景分析和对图像解释的理论和算法,而机器视觉/机器人视觉则更关注图像的获取,系统的构造和算法的实现。

模式包括的范围很广,图像就是模式的一种。识别是指从客观事实中自动建立符号描述或进行逻辑推理的数学和技术,因而人们定义模式识别是对客观世界中物体和过程进行分类、描述的学科。目前,对图像模式的识别主要集中在对图像中感兴趣的内容(目标)的分类、分析和描述。图像理解和计算机视觉中使用了很多模式识别的概念和方法,但视觉信息有其特殊性和复杂性,传统的模式识别(竞争学习模型)并不能把图像理解和计算机视觉全部包括进去。

人类智能主要指人类理解世界、判断事物、学习环境、规划行为、推理思维、解决问题等的的能力。人工智能则指由人类用计算机模拟、执行或再生某些与人类智能有关功能的能力和技術。视觉功能是人类智能的一种体现,所以图像理解和计算机视觉与人工智能密切相关。图像理解和计算机视觉的研究中使用了许多人工智能技术,反过来,图像理解和计算机视觉也可看作是人工智能的一个重要应用领域,需要借助人工智能的理论研究成果和系统实现经验。

计算机图形学研究如何从给定的描述生成“图像”,与计算机视觉也有密切关系。一般人们将计算机图形学称为计算机视觉的反(inverse)问题,因为视觉从 2-D 图像提取 3-D 信息,而图形学使用 3-D 模型来生成 2-D 场景。计算机图形学很多时候与图像分析联系更多。某些图形可以认为是图像分析结果的可视化,而计算机真实感景物的生成又可以认为是图像分析的逆过程[章 1996b]。另外,图形学技术在视觉系统的人机交互和建模等过程中也起很大作用。近期两相结合的一个研究热点——基于图像的绘制(image based rendering)就是一个很好的例子,一个概括介绍可见[章 2002b]。需要注意,与图像

理解和计算机视觉中存在许多不确定性相比,计算机图形学处理的多是确定性问题,是通过数学途径可以解决的问题。在许多实际应用中,人们更多关心的是在图形生成的速度和精度,即在实时性和逼真度之间取得某种妥协。

从更广泛的领域看,图像理解要用工程方法解决一些生物的问题,完成生物固有的功能,所以与生物学、生理学、心理学、神经学等学科也有着互相学习互为依赖的关系。近年图像理解研究者与视觉心理生理研究者紧密结合,已获得了一系列研究成果。图像理解属于工程应用科学,与电子学、集成电路设计、通信工程等密不可分。一方面图像理解的研究充分利用了这些学科的成果;另一方面图像理解的应用也极大地推动了这些学科的深入研究和发

1.2.4 图像理解的应用领域

近年来,图像理解已在许多领域得到广泛应用,下面是一些典型的例子:

(1) 工业视觉,如工业检测、工业探伤、自动生产流水线、邮政自动化、计算机辅助外科手术、显微医学操作,以及各种危险场合工作的机器人等。将图像和视觉技术用于生产自动化,可以加快生产速度,保证质量的一致性,还可以避免人的疲劳、注意力不集中等带来的误判。

(2) 人机交互,如人脸识别、智能代理等,让计算机可借助人的手势动作(手语)、嘴唇动作(唇读)、躯干运动(步态)、表情测定等了解人的愿望要求而执行指令,这既符合人类的交互习惯,也可增加交互方便性和临场感等。

(3) 视觉导航,如巡航导弹制导、无人驾驶飞机飞行、自动行驶车辆、移动机器人、精确制导以及智能交通的各个方面等,既可避免人的参与及由此带来的危险,也可提高精度和速度。

(4) 虚拟现实,如飞机驾驶员训练、医学手术模拟、场景建模、战场环境表示等,它可以帮助人们超越人的生理极限,产生亲临其境的感觉,提高工作效率。

(5) 图像自动解释,包括对放射图像、显微图像、遥感多波段图像、合成孔径雷达图像、航天航测图像等的自动判读理解。由于近年来技术的发展,图像的种类和数量飞速增长,图像的自动理解已成为解决信息膨胀问题的重要手段。

(6) 对人类视觉系统和机理、人脑心理和生理的研究等。

另外值得指出的是,在2003年世界机器人足球杯比赛期间,人们提出了一个大胆的目标:到2050年,将组建一只完全独立的类人机器人(人形机)足球运动队(fully autonomous humanoid robot soccer),而且它将按照世界足球联盟(FIFA)的比赛规则战胜当时的(人类)世界杯冠军队(见 www.robocup.org)。这样一个计划所需的时间(约50年)与历史上从莱特兄弟制造出第一架飞机到阿波罗飞船将人送到月球上并安全返回,以及从发明数字计算机到制造出击败人类国际象棋冠军的“深蓝”所需的时间可以比拟。