

# 第一章 引论

人类在征服自然、改造自然和推动社会进步的过程中，面临着自身能力、能量的局限性，因而发明和创造了许多机器来辅助或代替人类完成任务。智能机器，包括智能机器人，是这种机器最理想的形式，也是人类科学研究中所面临的最大挑战之一。智能机器是指这样一种系统，它能模拟人类的功能，能感知外部世界并有效地解决人所能解决问题。人类感知外部世界主要是通过视觉、触觉、听觉和嗅觉等感觉器官，其中约 80% 的信息是由视觉获取的。因此，对于智能机器来说，赋予机器以人类视觉功能对发展智能机器是及其重要的，也由此形成了一门新的学科——机器视觉(也称计算机视觉或图像分析与理解等)。机器视觉的发展不仅将大大推动智能系统的发展，也将拓宽计算机与各种智能机器的研究范围和应用领域。

机器视觉是研究用计算机来模拟生物外显或宏观视觉功能的科学和技术。机器视觉系统的首要目标是用图像创建或恢复现实世界模型，然后认知现实世界。机器视觉系统获取的场景图像一般是灰度图像，即三维场景在二维平面上的投影。因此，场景三维信息只能通过灰度图像或灰度图像序列来恢复处理，这种恢复需要进行多点对一点的映射逆变换。在信息恢复过程中，还需要有关场景知识和投影几何知识。

机器视觉是一个相当新且发展十分迅速的研究领域，并成为计算机科学的重要研究领域之一。机器视觉是在 20 世纪 50 年代从统计模式识别开始的[1]，当时的工作主要集中在二维图像分析和识别上，如光学字符识别，工件表面、显微图片和航空图片的分析和解释等。60 年代，Roberts(1965)通过计算机程序从数字图像中提取出诸如立方体、楔形体、棱柱体等多面体的三维结构，并对物体形状及物体的空间关系进行描述[Roberts 1965]。Roberts 的研究工作开创了以理解三维场景为目的的三维机器视觉的研究。Roberts 对积木世界的创造性研究给人们以极大的启发，许多人相信，一旦由白色积木玩具组成的三维世界可以被理解，则可以推广到理解更复杂的三维场景。于是，人们对积木世界进行了深入的研究，研究的范围从边缘、角点等特征提取，到线条、平面、曲面等几何要素分析，一直到图像明暗、纹理、运动以及成像几何等，并建立了各种数据结构和推理规则。到了 70 年代，已经出现了一些视觉应用系统[Guzman 1969, Mackworth 1973,]。

70 年代中期，麻省理工学院(MIT)人工智能(AI)实验室正式开设“机器视觉”(Machine Vision)课程，由国际著名学者 B. K. P. Horn 教授讲授。同时，MIT AI 实验室吸引了国际上许多知名学者参与机器视觉的理论、算法、系统设计的研究，David Marr 教授就是其中的一位。他于 1973 年应邀在 MIT AI 实验室领导一个以博士生为主体的研究小组，1977 年提出了不同于“积木世界”分析方法的计算视觉理论(computational vision)，该理论在 80 年代成为机器视觉研究领域中的一个十分重要的理论框架。

可以说，对机器视觉的全球性研究热潮是从 20 世纪 80 年代开始的，到了 80 年代中期，机器视觉获得了蓬勃发展，新概念、新方法、新理论不断涌现，比如，基于感知特征群的物体识别理论框架，主动视觉理论框架，视觉集成理论框架等。

到目前为止，机器视觉仍然是一个非常活跃的研究领域。许多会议论文集都反应了该领域的最新进展，比如，International Conference on Computer Vision and Pattern Recognition(CVPR); International Conference on Computer Vision(ICCV); International Conference on Pattern Recognition(ICPR); International Conference on Robotics and Automation(ICRA); Workshop on Computer Vision, and numerous conferences of SPIE。还有许多学术期刊也包含了这一领域的最新研究成果，如，IEEE Transaction on Pattern Analysis and Machine Intelligence(PAMI); Computer Vision, Graphics, and Image Processing(CVGIP); IEEE Transaction on Image Processing; IEEE

Transaction on Systems, Man, and Cybernetics(SMC); Machine Vision and Applications; International Journal on Computer Vision(IJCV); Image and Vision Computing; and Pattern Recognition. 每年还出版许多研究专集、学术著作、技术报告, 举行专题讨论会等. 所有这些都是研究机器视觉及其应用的很好信息来源.

## 1. 2 Marr 的视觉计算理论

Marr 的视觉计算理论[Marr1982]立足于计算机科学, 系统地概括了心理生理学、神经生理学等方面业已取得的所有重要成果, 是视觉研究中迄今为止最为完善的视觉理论. Marr 建立的视觉计算理论, 使计算机视觉研究有了一个比较明确的体系, 并大大推动了计算机视觉研究的发展. 人们普遍认为, 计算机视觉这门学科的形成与 Marr 的视觉理论有着密切的关系. 事实上, 尽管 20 世纪 70 年代初期就有人使用计算机视觉这个名词[Binford, 1971], 但正是 Marr 70 年代末建立的视觉理论促使计算机视觉这一名词的流行. 下面简要地介绍 Marr 的视觉理论的基本思想及其理论框架.

### 1. 2. 1 三个层次

Marr 认为, 视觉是一个信息处理系统, 对此系统研究应分为三个层次: 计算理论层次, 表示(representation)与算法层次, 硬件实现层次, 如表 1-1 所示.

表 1-1

计算理论	表示和算法	硬件实现
计算的目的是什么? 为什么这一计算是合适的? 执行计算的策略是什么?	如何实现这个计算理论? 输入、输出的表示是什么? 表示与表示之间的变换是什么?	在物理上如何实现 这些表示和算法?

按照 Marr 的理论, 计算视觉理论要回答视觉系统的计算目的和策略是什么, 或视觉系统的输入和输出是什么, 如何由系统的输入求出系统的输出. 在这个层次上, 信息系统的特征是将一种信息(输入)映射为另一种信息(输出). 比如, 系统输入是二维灰度图像, 输出则是三维物体的形状、位置和姿态, 视觉系统的任务就是如何建立输入输出之间的关系和约束, 如何由二维灰度图像恢复物体的三维信息. 表示与算法层次是要进一步回答如何表示输入和输出信息, 如何实现计算理论所对应的功能的算法, 以及如何由一种表示变换成另一种表示, 比如创建数据结构和符号. 一般来说, 不同的输入、输出和计算理论, 对应不同的表示, 而同一种输入、输出或计算理论可能对应若干种表示. 在解决了理论问题和表示问题后, 最后一个层次是解决用硬件实现上述表示和算法的问题, 比如计算机体系结构及具体的计算装置及其细节. 从信息处理的观点来看, 至关重要的乃是最高层次, 即计算理论层次. 这是因为构成知觉的计算本质, 取决于解决计算问题本身, 而不取决于用来解决计算问题的特殊硬件. 换句话说, 通过正确理解待解决问题的本质, 将有助于理解并创造算法. 如果考虑解决问题的机制和物理实现, 则对理解算法往往无济于事.

上述三个层次之间存在着逻辑的因果关系, 但它们之间的联系不是十分紧密, 因此, 某些现象只能在其中一个或两个层次上进行解释. 比如神经解剖学原则上与第三层次即物理实现联

系在一起。突触机制、动作电位、抑制性相互作用都在第三个层次上。心理物理学与第二层次(即表示与算法)有着更直接的联系。更一般地说,不同的现象必须在不同的层次上进行解释,这会有助于人们把握正确的研究方向。例如,人们常说,人脑完全不同于计算机,因为前者是并行加工的,后者是串行的。对于这个问题,应该这样回答:并行加工和串行加工是在算法这个层次上的区别,而不是根本性的区别,因为任何一个并行的计算程序都可以写成串行的程序。因此,这种并行与串行的区别并不支持这种观点,即人脑的运行与计算机的运算是不同的,因而人脑所完成的任务是不可能通过编制程序用计算机来完成。

## 1. 2. 2 视觉表示框架

视觉过程划分为三个阶段,如表 1-2 所示。第一阶段(也称为早期阶段)是将输入的原始图像进行处理,抽取图像中诸如角点、边缘、纹理、线条、边界等基本特征,这些特征的集合称为基元图(primitive sketch);第二阶段(中期阶段)是指在以观测者为中心的坐标系中,由输入图像和基元图恢复场景可见部分的深度、法线方向、轮廓等,这些信息的包含了深度信息,但不是真正的物体三维表示,因此,称为二维半图(2. 5 dimensional sketch);在以物体为中心的坐标系中,由输入图像、基元图、二维半图来恢复、表示和识别三维物体的过程称为视觉的第三阶段(后期阶段)。

表 1-2 由图像恢复形状信息的表示框架

名 称	目 的	基 元
图像	光强表示	图像中每一点的强度值
基元图	表示二维图像中的重要信息,主要是图像中的强度变化位置及其几何分布和组织结构	零交叉,斑点,端点和不连续点,边缘片断,有效线段,组合群,曲线组织,边界
2. 5 维图	在以观测者为中心的坐标系中,表示可见表面的方向、深度值和不连续的轮廓	局部表面朝向(“针”基元) 离观测者的距离 深度上的不连续点 表面朝向的不连续点
3 维模型表示	在以物体为中心的坐标系中,用由体积基元和面积基元构成的模块化多层次表示,描述形状及其空间组织形式。	分层次组成若干三维模型,每个三维模型都是在几个轴线空间的基础上构成的,所有体积基元或面积形状基元都附着在轴线上。

Marr 理论是计算机视觉研究领域的划时代成就,但该理论不是十分完善的,许多方面还有争议。比如,该理论所建立的视觉处理框架基本上是自下而上,没有反馈。还有,该理论没有足够地重视知识的应用。尽管如此,Marr 理论给了我们研究计算机视觉许多珍贵的哲学思想和研究方法,同时也给计算机视觉研究领域创造了许多研究起点。

## 1. 3 机器视觉的应用

机器视觉技术正广泛地应用于各个方面,从医学图象到遥感图像,从工业检测到文件处理,从毫微米技术到多媒体数据库,不一而足。可以说,需要人类视觉的场合几乎都需要机器视觉。应该指出的是,许多人类视觉无法感知的场合,如精确定量感知、危险场景感知、不可见物体感知等,机器视觉更突显其优越性。下面是一些机器视觉的典型应用。

### (1) 零件识别与定位

由于工业环境的结构、照明等因素可以得到严格的控制，因此，机器视觉在工业生产和装配中得到了成功的应用。图 1. 1 是一个具有简单视觉的工业机器人系统示意图，其视觉系统由一个摄像机和相关的视觉信息处理系统组成。摄像机位于零件传输带上方，对于不同的零件，可以选择不同颜色的传输带，比如，明亮的物体，选择黑色传输带，暗色的零件，选择白色的背景，这样有利于视觉系统将零件从传输带上分离出来，并进行识别和定位，识别的目的是为机器人提供是否操作或进行何种操作的信息，定位的目的是导引机器人手爪实时准确地夹取零件。

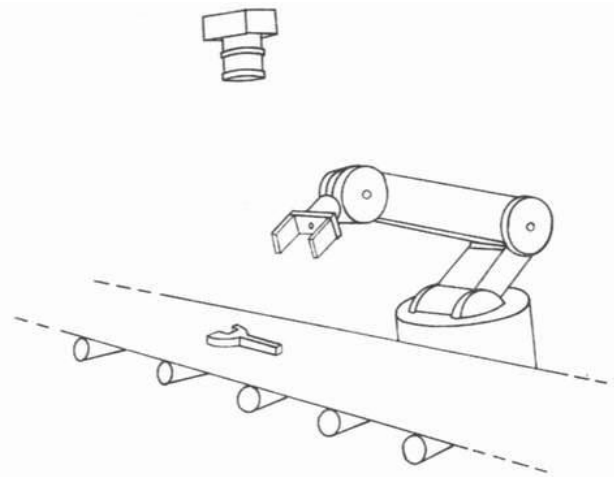


图 1. 1 用于生产线上具有简单视觉系统的工业机器人系统示意图

### (2) 产品检验

机器视觉在工业领域中另一个成功的应用是产品检验。目前已经用于产品外形检验、表面缺陷检验，比如，滑块及滑槽的外形检验以及装配后的位置检验，以决定它们能否装配在一起，并且准确无误地完成装配任务；发动机内壁麻点、刻痕等缺陷检查，以决定产品的质量。通过 X 射线照相或超声探测获取物体内部的图像，可以实现内部缺陷检验，如钢梁内部裂纹和气孔等缺陷检验。

### (3) 移动机器人导航

我们来看一下图 1. 2 所示的两组图像，每一组图像称为一个立体对 (stereo pair)，是由移动机器人上的两个摄像机同步获取的，表示某一时刻关于场景的不同视点的两幅图像。机器人利用立体对可以恢复周围环境的三维信息。移动机器人可以利用场景的三维信息识别目标、识别道路、判断障碍物等，实现道路规划、自主导航，与周围环境自主交互作用等。将立体图像对和运动信息组合起来，可以构成满足特定任务分辨率要求的场景深度图。这种技术对无人汽车、无人飞机、无人战车等自主系统的自动驾驶十分有用。比如，著名的美国 Sojourner 和 Rocky7 等系列火星探测移动机器人都使用了立体视觉导航系统。

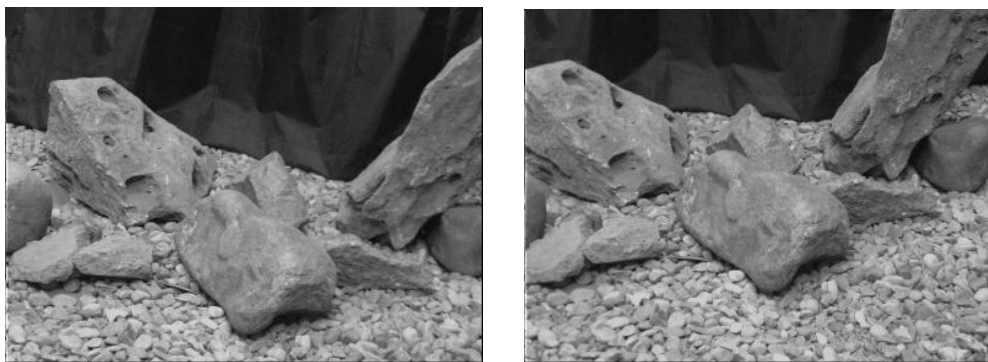


图 1. 2 由移动机器人立体视觉系统获取的立体图像对，可用来重建场景三维信息

#### (4) 遥感图像分析

目前的遥感图像包括三种：航空摄影图像、气象卫星图像、资源卫星图像。这些图像的共同特点是在高空对地表或地层进行远距离成像，但三种图像的成像机理完全不同。航空图像可以用普通的视频摄像机来获取，分析方法也同普通的图像分析一样。卫星图像的获取和应用随着成像机理不同而变化很大，气象卫星使用红外成像传感系统可以获取不同云层的图像，即云图，由此分析某一地区的气象状况；海洋卫星使用合成孔径雷达获取海洋、浅滩图像，由此重构海洋波浪三维表面图；资源卫星装备有多光谱探测器（multiple spectral sensor, MSS），可以获取地表相应点的多个光谱段的反射特性，如红外、可见光、紫外等，多光谱图像被广泛地用于找矿、森林、农作物调查、自然灾害测报、资源和生态环境检测等。

#### (5) 医学图像分析

目前医学图像已经广泛用于医学诊断，成像方法包括传统的 X 射线成像、计算机层析（computed tomography, CT）成像、核磁共振成像（magnetic resonance imaging, MRI）、超声成像等。机器视觉在医学图像诊断方面有两方面的应用，一是对图像进行增强、标记、染色等处理来帮助医生诊断疾病，并协助医生对感兴趣的区域进行定量测量和比较；二是利用专家知识系统对图像（或是一段时期内的一系列图像）进行自动分析和解释，给出诊断结果。

#### (6) 安全鉴别、监视与跟踪

用机器视觉系统可以实现停车场监视、车辆识别、车牌号识别、探测并跟踪“可疑”目标；根据面孔、眼底、指纹等特征识别特定人。目前人们正在研究一种面部运动参数的提取和描述，以分析人的表情及内心活动。

#### (7) 国防系统

机器视觉在国防系统中的作用越来越重要，一个理由是满足自主操作的需要，另一个理由是分析大量先进成像传感器的输出。显而易见，在国防系统中迅速作出反应是极其重要的，这就需要在人工尽可能少的干预下作出各种决策，尤其是与图像和视觉方法有关的各种技术，比如，图像制导与目标识别等。

#### (8) 其它

机器视觉已经用于各种球类运动分析、人体测量，食品，农业、心理学、电视电影制作、美术模型、远程教育，多媒体教学等场合。

## 1. 4 机器视觉研究内容与面临的困难

机器视觉研究可以分为如下五大研究内容：

- 输入设备
- 低层视觉

- 中层视觉
- 高层视觉
- 体系结构

输入设备(input device)包括成像设备和数字化设备。成像设备是指通过光学摄像机或红外、激光、超声、X 射线对周围场景或物体进行探测成像,得到关于场景或物体的二维或三维数字化图像。获取数字化图像是机器视觉系统的最基本的功能。目前用于视觉研究的大多数输入设备是商品化的产品,如,CCD 黑白或彩色摄像机,数字扫描仪,超声成像探测仪,CT 成像设备等。但这些商品化的输入设备远远不能满足实际的需要,因此,仍有许多研究人员在研究各种性能先进的成像系统,如,红外成像系统,激光成像系统,还有所谓的计算成像系统(computational imaging),即每一个像素元(或若干像素元)对应一个简单的处理器,这样可以适应复杂场景动态变化的场合[Brajovic 1996]。

低层视觉(low level)主要是对输入的原始图像进行处理。这一过程借用了大量的图像处理技术和算法,如图像滤波、图像增强、边缘检测等,以便从图像中抽取诸如角点、边缘、线条、边界以及色彩等关于场景的基本特征;这一过程还包含了各种图像变换(如校正)、图像纹理检测、图像运动检测等。

中层视觉(middle level)的主要任务是恢复场景的深度、表面法线方向、轮廓等有关场景的 2.5 维信息,实现的途径有立体视觉(stereo vision)、测距成像(rangefinder)运动估计(motion estimation)、明暗特征、纹理特征等所谓的从 X 恢复形状的估计方法。系统标定、系统成像模型等研究内容一般也是在这个层次上进行的。

高层视觉(high level)的任务是在以物体为中心的坐标系中,在原始输入图像、图像基本特征、2.5 维图的基础上,恢复物体的完整三维图,建立物体三维描述,识别三维物体并确定物体的位置和方向。另外,主动视觉(active vision)涵盖了上述各个层次的研究内容。

值得指出,低层、中层和高层机器视觉基本上与 Marr 视觉的三个阶段相对应。

体系结构(system architecture)这一术语最通常的含义系指在高度抽象的层次上,根据系统模型而不是根据实现设计的具体例子来研究系统的结构。为了说明这一点,可以考虑建筑设计中某一时期的建筑风格(如清朝时期)和根据这一风格设计出来的具体建筑之间的区别。体系结构研究涉及一系列相关的课题:并行结构、分层结构、信息流结构、拓扑结构以及从设计到实现的途径。

人们对上述几个研究内容进行了卓有成效的研究,研究出大量的技术和算法,并且在各个领域中得到广泛的应用。不过,机器视觉技术仍处于十分不成熟的阶段,其发展远远落后于人们所寄予的发展水平。

对于人类视觉来说,识别和理解周围场景是一件非常容易的事,但对于机器来说,却是一件很困难的事。主要困难体现在如下几方面:

- (1) 图像多义性: 三维场景被投影为二维图像,深度和不可见部分的信息被丢失,因而会出现不同形状的三维物体投影在图像平面上产生相同图像的问题,如图 1.3 所示。另外,在不同角度获取同一物体的图像会有很大的差异。
- (2) 环境因素影响: 场景中的诸多因素,包括照明、物体形状、表面颜色、摄像机以及空间关系变化都会对投影的图像有影响,因此,当任何一个因素发生变化时,都会对图像产生影响。
- (3) 知识引导: 同样的图像在不同的知识引导下,将会产生不同的识别结果。图 1.4—1.5 所示的图像,在不同的约束或知识引导下,具有不同的识别结果。图 1.4(b)可能表示一个少女前视远方的侧面图像,也可能表示一个老妇人凝视下前方的正侧面图像。图 1.5 包含有几个的立方体,不同的知识引导可能产生不同的空间关系,也就具有不同意义。
- (4) 大量数据: 灰度图像,彩色图像,深度图像的信息量十分巨大,比如分辨率为  $512 \times 512$  的灰度图像的数据量为 256K,同样分辨率的彩色图像的数据量是 768K。如果处理的是图像序列,则数据量更大。巨大的数据量需要很大的存储空间,同时不易实现快速

处理.

为了解决视觉所面临的问题, 研究人员不断寻求新的途径和手段, 比如, 主动视觉(active vision), 面向任务的视觉(task-oriented vision), 基于知识、基于模型的视觉, 以及多传感融合和集成视觉等方法, 其中人们越来越重视对知识的应用. 我们会看到, 机器视觉系统的最大特征是, 在视觉的各个阶段, 系统尽可能地进行自动运算. 为此, 系统需要使用各种知识, 包括特征模型、成像过程、物体模型和物体间的关系. 如果机器视觉系统不用这些知识, 则其应用的范围及其功能将十分有限. 因此, 视觉系统应该使用那些可以被明确表示的知识, 以使系统具有更高的适应性和鲁棒性. 合理地使用知识不仅可以有效地提高系统的适应性和鲁棒性, 而且可以求解机器视觉中较难的问题.

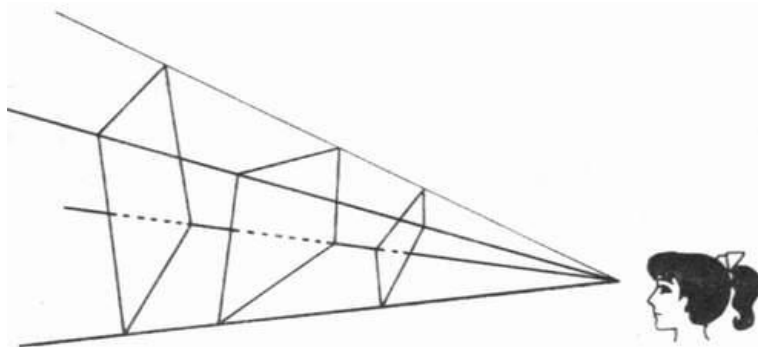


图 1. 3 不同形状的三维物体投影在图像平面上产生相同图像



图 1. 4 一组人脸侧面图像. 注意, 中间图像既与左边图像相似, 又与右边图像相似[Hochberg 1964].



图 1.5 几个立方体构成的多义性图像[选自 Combs 的作品]

## 1.5 机器视觉与其它学科领域的关系

与机器视觉有关的学科有许多。本节主要讨论一些与机器视觉密切相关的领域。关于机器视觉与其它学科的关系，我们不作详尽的讨论。

图像处理是一个发展比较成熟的领域。图像处理技术通常是把一幅图像变换成另外一幅图像，也就是说，图像处理系统的输入是图像，输出仍然是图像，信息恢复任务则留给人来完成。图像处理包括图像增强、图像压缩和模糊校正与非聚焦图像等课题。机器视觉系统把图像作为输入，产生的输出为另一种形式，比如图像中物体轮廓的表示。因此，机器视觉的重点是在人的最小干预下，由计算机自动恢复场景信息。图像处理算法在机器视觉系统的早期阶段起着很大的作用，它们通常被用来增强特定信息并抑制噪声。

计算机图形学是通过几何基元，如线、圆和自由曲面，来生成图像，它在可视化（Visualization）和虚拟现实（Virtual Reality）中起着很重要的作用。机器视觉正好是解决相反的问题，即从图像中估计几何基元和其它特征。因此，计算机图形学属于图像综合，机器视觉属于图像分析。这两个领域在其发展的早期阶段是没有什么联系的，但是近十几年来发展的越来越相近了。机器视觉使用了计算机图形学中的曲线和曲面表示方法以及其它的一些技术，而计算机图形学也使用机器视觉技术，以便在计算机中建立逼真的图像模型。可视化和虚拟现实把这两个领域紧密地联系在一起。

模式识别主要用于识别各种符号、图画等平面图形。模式一般指一类事物区别于其它事物所具有的共同特征。模式识别方法主要有统计方法和句法方法两种，统计方法是指从模式抽取一组特征值，并以划分特征空间的方法来识别每一个模式。句法方法是指利用一组简单的子模式（模式基元）通过文法规则来描述复杂的模式。模式识别方法是机器视觉识别物体的重要基础之一。机器视觉识别物体还经常需要其它的技术。我们将在物体识别部分简要地讨论统计模式识别的主要内容。



人工智能 (artificial intelligent, AI) 涉及到智能系统的设计和智能计算的研究。在经过图像处理和图像特征提取过程后, 接下来要用人工智能方法对场景特征进行表示, 并分析和理解场景。人工智能有三个过程: 感知、认知和行动。感知把反应现实世界的信息转换成信号, 并表示成符号, 认知是对符号进行各种操作, 行动则把符号转换成影响周围环境的信号。人工智能的许多技术在机器视觉的各个方面起着重要作用。事实上, 机器视觉通常被视为人工智能的一个分支。

人工神经网络 (artificial neural networks, ANNs) 是一种信息处理系统, 它是由大量简单的处理单元 (称为神经元) 通过具有强度的连接 (connection) 相互联系起来, 实现并行分布式处理 (parallel distribution processing, PDP)。人工神经网络的最大特点是可以通过改变连接强度来调整系统, 使之适应复杂的环境, 实现类似人的学习、归纳和分类等功能。人工神经网络已经在许多工程技术领域得到了广泛的应用。神经网络作为一种方法和机制将用于解决机器视觉中的许多问题。

神经物理学与认知科学长期将人类视觉作为主要的研究对象。机器视觉中已有的许多方法与人类视觉极为相似。目前, 许多机器视觉研究者对研究人类视觉计算模型比研究机器视觉系统更感兴趣, 希望机器视觉更加自然化, 更加接近生物视觉。我们在第二章介绍人类视觉的一些研究成果, 使读者在研究机器视觉时或多或少能从生物视觉机理中得到启发。不过, 我们的重点在于研究机器视觉系统, 因此, 不讨论机器视觉与神经物理学或认知科学的联系。

## 1. 6 成象几何基础

一般的成象系统通常将三维场景变换成二维灰度或彩色图像, 这种变换可以用一个从三维空间到二维空间的映射来表示:

$$\begin{aligned} f: R^3 &\rightarrow R^2 \\ (x, y, z) &\rightarrow (x', y') \end{aligned} \quad (1. 1)$$

如果考虑时变三维场景, 则上述变换是四维空间到三维空间的变换, 如果再考虑某一波段或某几个波段的光谱, 则上式的维数将增加到五维或更高维。这里我们只考虑三维空间到二维空间的两种常用映射: 透视投影变换和正交投影变换。

### 1. 6. 1 透视投影

透视投影 (perspective projection) 是最常用的成像模型, 可以用针孔 (pinhole) 成像模型来近似表示。透视投影成像模型的特点是所有来自场景的光线均通过一个投影中心, 它对应于透镜的中心, 经过投影中心且垂直于图像平面 (成像平面) 的直线称为投影轴或光轴, 如图 1. 6 所示。其中  $oxyz$  是固定在摄象机上的直角坐标系, 遵循右手法则, 其原点  $o$  位于投影中心,  $z$  轴与投影轴重合并指向场景,  $x$  轴、 $y$  轴与图像平面的坐标轴  $x'$  和  $y'$  平行,  $xy$  平面与图像平面的距离为  $f$ , 一般称为摄象机的焦距。在实际摄像机中, 图像平面位于投影中心后面距离为  $f$  的位置 (图 1. 6), 其投影图像是倒立的。为了避免这种倒立图像, 假定图像平面位于投影中心的前面, 如图 1. 7 所示。场景中一点  $(x, y, z)$  在图像平面上的投影位置  $(x', y')$  可以通过计算点  $(x, y, z)$  的视线 (空间点与投影中心的连线) 与图像平面的相交位置得到。

点  $(x, y, z)$  的视线、点  $(x, y, z)$  到  $z$  轴的垂直线段以及  $z$  轴形成了一个三角形。图像平面上点  $(x', y')$  的视线、点  $(x', y')$  到  $z$  轴垂直线段以及  $z$  轴形成了另一个三角形。这两个三角形是相

似三角形，因此得到透视投影方程：

$$\frac{x'}{x} = \frac{y'}{y} = \frac{f}{z} \quad (1.2)$$

点  $(x, y, z)$  在图像平面中的位置由下式给出：

$$x' = \frac{f}{z} x \quad y' = \frac{f}{z} y \quad (1.3)$$

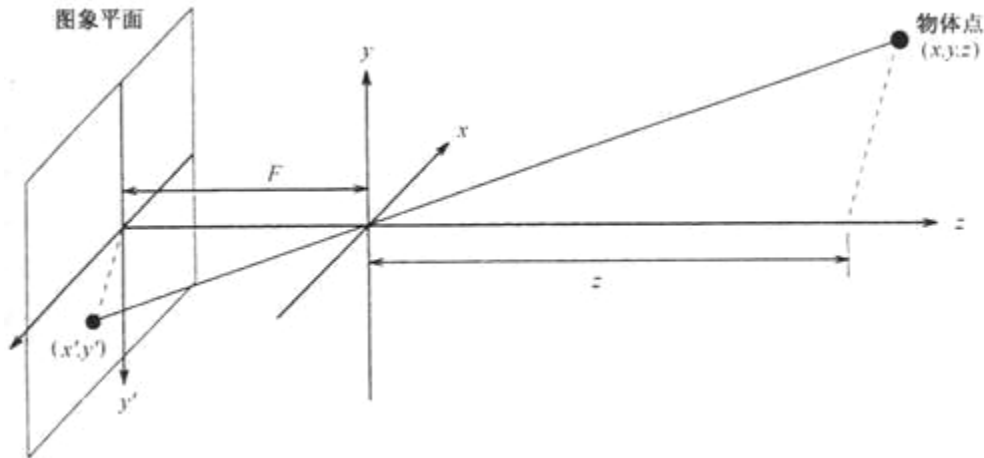


图 1. 6 透视投影倒立成像几何示意图

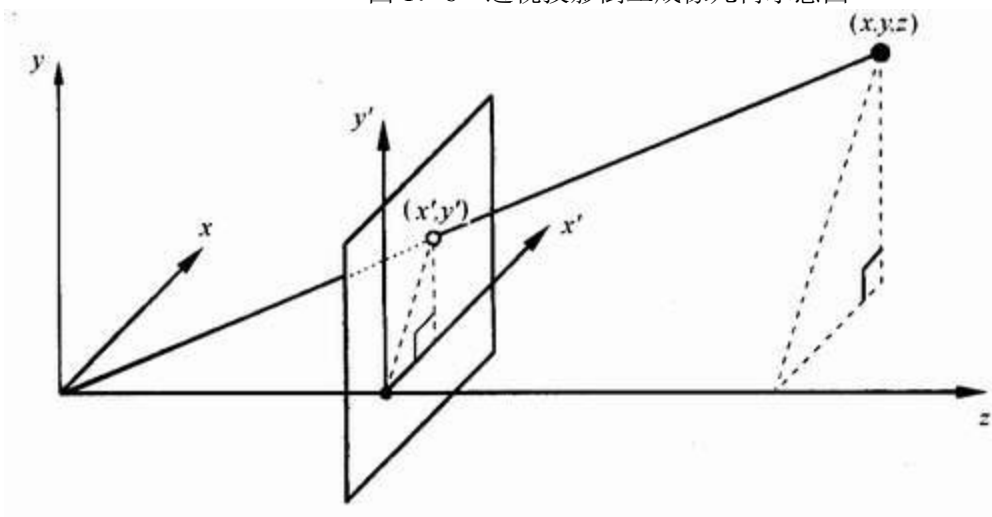


图 1. 7 透视投影几何示意图

## 1. 6. 2 正交投影

正交投影 (orthogonal projection) 指用平行于光轴的光将场景投射到图像平面上，因此也称为平行投影 (parallel projection)，如图 1. 8 所示。正交投影是透视投影的一个特例，当透视投

影模型的焦距  $f$  很大且物体距投影中心很远时，透视投影就可以用正交投影来近似。此时，投影方程为：

$$\begin{aligned} x' &= x \\ y' &= y \end{aligned} \quad (1.4)$$

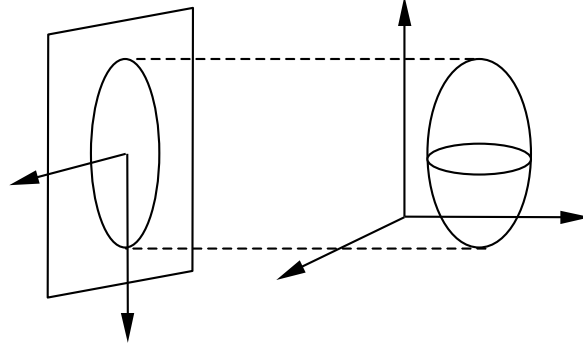


图 1.8 正交投影几何示意图

### 1.6.3 视觉系统坐标系

在计算机视觉系统中涉及到以下几种坐标：

- 像素坐标：表示图像阵列中图像像素的位置；
- 图像平面坐标：表示场景点在图像平面上的投影；
- 摄像机坐标：即以观察者为中心的坐标，将场景点表示成以观察者为中心的数据形式。
- 场景坐标：也称作绝对坐标(或世界坐标)，用于表示场景点的绝对坐标；

像素坐标也称为图像阵列坐标或网格坐标。图像处理通常是在图像阵列的像素坐标系中进行，其中图像的行数和列数  $[i, j]$  对应于图像网格的整数坐标，即像素的坐标是整数值，像素  $[0,0]$  位于图像的左上角， $i$  指向下方， $j$  指向右方。这种图像像素坐标系，正好对应计算机程序里的阵列语法。为了提高图像处理的精度，需要引入子像素(sub-pixel)概念。所谓子像素是指像素与像素之间的点，其坐标用小数表示。使用摄像机几何假设条件可以将像素点坐标转化成图像平面坐标。

图像平面坐标定义如下：假定摄像机主点，即光轴与图像平面的交点，位于图像的中心点，若图像阵列为  $m \times n$ ，则图像平面坐标的中心坐标为：

$$\hat{c}_x = \frac{m-1}{2} \quad \hat{c}_y = \frac{n-1}{2} \quad (1.5)$$

其中  $(\hat{c}_x, \hat{c}_y)$  表示中心坐标的估计值；图像平面坐标  $x'$  轴方向是列数标号增加的方向，但  $y'$  轴方向指向行数标号增加方向的反向。假定图像阵列的行列间距是相等的，则像素坐标  $[i, j]$  到图像坐标  $(x', y')$  的变换公式为：

$$x' = j - \frac{m-1}{2} \quad y' = -\left(i - \frac{n-1}{2}\right) \quad (1.6)$$

如果图像阵列的行列间距不相等，并设列间距为  $s_x$ ，行间距为  $s_y$ ，则像素坐标到图像坐标的变换式为：

$$x' = s_x \left( j - \frac{m-1}{2} \right) \quad y' = -s_y \left( i - \frac{n-1}{2} \right) \quad (1.7)$$

为了简化运算，在机器视觉应用中普遍要求摄像机的像素为正方形像素。如果使用了非正方形像素摄像机，则必须考虑非正方形像素对测量的影响。若有影响，则必须在测量之前将其坐标转化为标准的图像平面坐标。图像坐标可以是摄像机误差(如透镜畸变)修正之后或未修正的真实图像坐标。

在本书中，场景坐标或绝对坐标用  $(X, Y, Z)$  表示，摄像机坐标用  $(x, y, z)$  表示。摄像机坐标系也称为以观察者为中心的坐标系。在场景中也许有多个摄像机，每个摄像机都有其自己的坐标系。例如，在双目立体系统中，就有左摄像机坐标和右摄像机坐标，以及表示深度测量的立体坐标。这里请读者注意，由于在许多情况下，图像平面坐标系是摄像机坐标系的一部分，因此，为了书写简单，在不混淆的情况下，通常使用摄像机  $x-y$  平面坐标代替图像平面  $x'-y'$  坐标。

## 1.7 本书内容向导

本书各章内容基本上是按照由低级到高级，由灰度到彩色，由二维到三维，由分析到表示和理解这种顺序安排的。大致说来，除了第二章是有关生理视觉的介绍外，本书的前半部分基本上限于二维空间，后半部分把范围则扩展到了三维空间。关于二值图像的第三章介绍了机器视觉领域所用的基本术语和概念，那里所讨论的技术可用于视觉系统的所有方面。第四章介绍了图像区域检测技术，同时也讨论了区域表示方法。所有的视觉系统都离不开区域检测或分析技术，因此这一章讨论的内容是视觉的基本内容。第五章介绍了图像滤波方法，论述了图像增强技术和一些其它的滤波技术。图像滤波的主要目的是突显用于视觉识别的特征，去除噪声等，因此是实现机器视觉的基础。边缘检测技术也是许多机器视觉系统中最重要、最基本的步骤，第六章介绍了几种边缘检测技术并比较了它们的性能以及在应用中注意的问题。接下来一章介绍了轮廓表示方法，边缘是局部的，只有把它们组成有意义的物体并表示出来才能使用它们。纹理在诸如表面检测、场景分类、表面姿态和形状分类等许多机器视觉的任务中起着重要作用，这些内容将在第八章讨论。辐射测量学将在第九章讨论。在许多应用中，色彩是图像的重要组成部分，我们将在第十章讨论。第十一章介绍主动方法和被动方法恢复图像深度信息的技术。这种技术把我们从二维图像带到了三维图像。第十二章讨论了摄像机标定技术，为了从图像中获取三维信息，必须知道摄像机的位置和姿态及其自身的一些参数。这一章讨论了各种用于摄像机标定的技术。第十三章讨论了空间中曲线和曲面的表示和它们的一些性质，并讨论了内插和逼近技术。第十四章介绍了图像二维运动估计，包括图像变化检测、基于运动特性的图像分割和物体跟踪技术。关于从运动恢复结构的方法等有关三维运动分析和估计是最近几年的研究热点，这一部分内容在第十五章讨论。物体识别是视觉系统的一个最普遍的应用。我们将在第十六章讨论一些物体识别的基本方面。还有一些最新的信息处理方法，如神经网络、分数维方法、子波方法，主要从应用的角度进行了讨论，这些内容分布在有关章节中。

我们认为练习是课程中一个十分重要的环节。本书各章都给出了一些概念题和计算题来检验读者对书中内容的理解，另外还有一定数量的计算机练习题，以增加读者的实际。我们把这门课看成是一门理论与实验相结合的课程。请读者注意，完成一个简单的视觉系统设计练习题对于真正理解机器视觉系统非常重要。