

M A C H I N E

V I S I O N

# 机器视觉

张广军 编著



科学出版社  
[www.sciencep.com](http://www.sciencep.com)

## 内 容 简 介

本书系统地介绍了机器视觉的基础理论、方法及关键技术与实用算法,并给出了应用实例。本书涉及机器视觉中的主要研究方向,主要内容包括空间几何变换与摄像机模型、视觉图像特征信息提取、摄像机标定、双目立体视觉、结构光三维视觉、其他三维视觉技术、多传感器三维视觉、运动视觉分析及小型构件内表面三维形貌视觉检测和天文导航两个应用实例。

本书可作为信息处理、计算机、机器人、人工智能、自动化及仪器仪表等专业的高年级本科生和研究生的教材,也可供从事相关专业和机器视觉研究工作的技术人员参考。

### 图书在版编目(CIP)数据

机器视觉/张广军编著. —北京:科学出版社,2005

ISBN 7-03-014717-0

I. 机… II. 张… III. 计算机视觉-研究生-教材 IV. TP302.7

中国版本图书馆CIP数据核字(2005)第001611号

责任编辑:马长芳 潘继敏 / 责任校对:赵桂芬  
责任印制:钱玉芬 / 封面设计:耕者创作室

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>

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

\*

2005年6月第 一 版 开本: B5 (720×1000)

2005年6月第一次印刷 印张: 17 3/4

印数: 1—3 000 字数: 344 000

定价: 32.00 元

(如有印装质量问题, 我社负责调换〈环伟〉)

## 前 言

随着信号处理理论和电子、计算机等相关技术的发展，机器视觉得到了迅速发展，并以其非接触、实时、可视化好、自动化和智能性高等特点在国民经济、科学研究及国防建设等领域得到广泛的应用，受到各国各行业的高度重视。

机器视觉涉及光学、光电子学、图像处理、模式识别、信号处理、人工智能及计算机技术等诸多学科领域，内容广泛。本书按照机器视觉所涉及的内容和主要研究方向来安排，较系统地介绍了机器视觉的基础理论、方法及关键技术与实用算法，并给出了应用实例。与当前同类书相比，本书既有较为系统的机器视觉理论与算法，又有一定篇幅的机器视觉关键技术及工程应用内容，并融入了最新技术和科研成果。全书共分 11 章，第 1 章为引论，第 2 章介绍了空间几何变换与摄像机模型，第 3 章介绍了视觉图像特征信息提取，第 4 章介绍了摄像机标定，第 5 章介绍了双目立体视觉，第 6 章介绍了结构光三维视觉，第 7 章介绍了其他三维视觉技术，第 8 章介绍了多传感器三维视觉，第 9 章介绍了运动视觉分析，第 10 章介绍了小型构件内表面三维形貌视觉检测，第 11 章介绍了天文导航。

本书是在作者十多年从事机器视觉研究和研究生教学，并参考和引用了国内外相关文献和研究成果的基础上编著而成的，同时也是作者所在研究组十多年机器视觉研究工作的总结，具有较好的基础性、系统性、先进性和实用性。

近十年来，本书作者在机器视觉研究工作中获得了多项研究基金和研究计划资助，其中有国家杰出青年科学基金、国家自然科学基金、民用航天重大研究项目、教育部跨世纪优秀人才培养计划、北京市科技新星计划、航空科学基金等十余项项目的资助。作者对就这些研究基金和研究计划给以资助的国家自然科学基金委员会、国防科工委科技质量司、教育部科技司、北京市科委、航空科学基金办公室等有关部门表示感谢。

在本书编著过程中，得到了各方面的大力支持和专家、同事的帮助。作者要特别感谢天津大学叶声华院士对作者多年的关心和帮助。作者也十分感谢研究组的周富强、魏振忠、江洁、贺俊吉、魏新国、郝学涛、孙军华等同事和同学们，本书介绍的许多工作是由他们具体完成的。

本书参考和引用的参考文献与研究成果已在文中列出或说明，感兴趣的读者可直接查阅。

机器视觉内容十分广泛，涉及诸多学科领域。由于作者水平有限，经验不足，书中不妥之处在所难免，敬请读者批评指正。

张广军

2004年8月

# 目 录

## 前言

<b>第 1 章 引论</b> .....	1
1.1 机器视觉的发展及系统构成 .....	1
1.2 Marr 的视觉理论框架.....	5
1.3 机器视觉的应用领域及面临的问题 .....	8
1.4 本书各章内容简介.....	11
思考与练习题 .....	12
参考文献 .....	12
<b>第 2 章 空间几何变换与摄像机模型</b> .....	14
2.1 空间几何变换.....	14
2.2 几何变换的不变量.....	18
2.3 欧氏空间的刚体变换.....	21
2.4 摄像机透视投影模型.....	24
2.5 摄像机透视投影近似模型.....	27
思考与练习题 .....	32
参考文献 .....	32
<b>第 3 章 视觉图像特征信息提取</b> .....	33
3.1 图像边缘与图像平滑.....	33
3.2 一阶微分边缘检测算子.....	37
3.3 二阶微分边缘检测算子.....	42
3.4 子像素级边缘检测.....	45
3.5 角点探测器.....	50
3.6 形状特征分析.....	53
3.7 椭圆孔图像中心的提取.....	56
3.8 空间椭圆中心图像位置的提取.....	58
3.9 给定形状曲线的检测——Hough 变换 .....	63
思考与练习题 .....	67
参考文献 .....	67
<b>第 4 章 摄像机标定</b> .....	69
4.1 非线性优化方法.....	69

4.2	基于3D立体靶标的摄像机标定	76
4.3	基于径向约束的摄像机标定	79
4.4	基于2D平面靶标的摄像机标定	84
4.5	机器人手眼定标	89
4.6	摄像机自标定技术	92
4.7	基于交比不变的摄像机畸变系数标定	94
	思考与练习题	97
	参考文献	97
<b>第5章</b>	<b>双目立体视觉</b>	99
5.1	双目立体视觉原理	99
5.2	双目立体视觉的精度分析	102
5.3	双目立体视觉的系统结构	105
5.4	双目立体视觉中的极线几何	108
5.5	双目立体视觉中的对应点匹配	112
5.6	双目立体视觉系统标定	122
	思考与练习题	125
	参考文献	125
<b>第6章</b>	<b>结构光三维视觉</b>	127
6.1	结构光三维视觉原理	127
6.2	结构光光模式投射系统	134
6.3	结构光三维视觉常规标定方法	139
6.4	基于交比不变的标定方法	143
6.5	基于自由移动平面靶标的现场标定方法	147
6.6	基于神经网络的标定方法	150
	思考与练习题	155
	参考文献	155
<b>第7章</b>	<b>其他三维视觉技术</b>	156
7.1	光度立体视觉	156
7.2	由纹理恢复形状	163
7.3	激光测距法	167
7.4	莫尔阴影与散焦测距	173
	思考与练习题	178
	参考文献	178
<b>第8章</b>	<b>多传感器三维视觉</b>	180
8.1	多传感器三维视觉系统	180

---

8.2 全局标定方法 .....	186
8.3 全局标定系统 .....	193
8.4 全局标定精度分析与评价 .....	200
思考与练习题 .....	206
参考文献 .....	206
<b>第9章 运动视觉分析</b> .....	<b>207</b>
9.1 图像运动特征提取 .....	207
9.2 基于光流的运动分析 .....	210
9.3 光流计算 .....	214
9.4 基于二维点匹配的运动分析 .....	218
9.5 基于二维直线匹配的运动计算 .....	221
9.6 基于三维点匹配的运动估计 .....	224
思考与练习题 .....	227
参考文献 .....	228
<b>第10章 应用实例 I——小型构件内表面三维形貌视觉检测</b> .....	<b>229</b>
10.1 概述 .....	229
10.2 系统硬件与软件构成 .....	233
10.3 图像特征分析与提取 .....	239
10.4 系统标定与测量应用 .....	244
思考与练习题 .....	249
参考文献 .....	249
<b>第11章 应用实例 II——天文导航</b> .....	<b>250</b>
11.1 概述 .....	250
11.2 CMOS 图像传感器驱动及图像预处理 .....	254
11.3 基于径向和环向特征的星图识别 .....	262
11.4 改进的三角形算法星图识别 .....	267
思考与练习题 .....	273
参考文献 .....	273

# 第 1 章 引 论

人类是通过眼睛和大脑来获取、处理与理解视觉信息的。周围环境中的物体在可见光照射下，在人眼的视网膜上形成图像，由感光细胞将其转换成神经脉冲信号，并经神经纤维传入大脑皮层进行处理与理解。所以说，视觉不仅指对光信号的感受，还包括对视觉信息的获取、传输、处理与理解的全过程。

随着信号处理理论和计算机技术的发展，人们试图用摄像机获取环境图像并将其转换成数字信号，用计算机实现对视觉信息处理的全过程，这样就形成了一门新兴的学科——计算机视觉。计算机视觉的研究目标是使计算机具有通过一幅或多幅图像认知周围环境信息的能力。这使计算机不仅能模拟人眼的功能，而且更重要的是使计算机完成人眼所不能胜任的工作。机器视觉则是建立在计算机视觉理论上，偏重于计算机视觉技术工程化。与计算机视觉研究的视觉模式识别、视觉理解等内容不同，机器视觉重点在于感知环境中物体的形状、位置、姿态、运动等几何信息。

本章首先讨论机器视觉的发展及系统构成，然后介绍 Marr 的视觉理论框架，最后对机器视觉的应用领域与面临的问题，以及本书各章内容进行介绍。

## 1.1 机器视觉的发展及系统构成

视觉是人类观察世界和认知世界的重要手段。据统计，人类从外部世界获得的信息约有 80% 是由视觉获取的。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，同时又体现了人类视觉功能的重要性。随着信息技术的发展，给计算机、机器人或其他智能机器赋予人类视觉功能，是人类多年以来的梦想。虽然目前还不能够使计算机、机器人或其他智能机器也具有像人类等生物那样高效、灵活和通用的视觉，但自 20 世纪 50 年代以来视觉理论和技术得到了迅速发展，这使得人类的梦想正在逐步实现。

### 1.1.1 机器视觉的发展

计算机视觉是用计算机实现人的视觉功能——对客观世界的三维场景的感知、识别和理解。计算机视觉是在 20 世纪 50 年代从统计模式识别开始的，当时的工作主要集中在二维图像分析、识别和理解上，如光学字符识别、工件表面、



显微图片和航空照片的分析和解释等。20 世纪 60 年代, Roberts 将环境限制在所谓的“积木世界”, 即周围的物体都是由多面体组成的, 需要识别的物体可以用简单的点、直线、平面的组合表示。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。到 70 年代, 已经出现了一些视觉应用系统 (Guzman 1969, Mackworth 1973)。

1973 年, 英国的 Marr 教授应邀在麻省理工学院 (MIT) 的人工智能实验室创建并领导一个以博士生为主体的研究小组, 从事视觉理论方面的研究。1977 年, Marr 提出了不同于“积木世界”分析方法的计算视觉理论——Marr 视觉理论, 该理论在 20 世纪 80 年代成为计算机视觉研究领域中的一个十分重要的理论框架 (Marr 1982)。到了 80 年代中期, 计算机视觉获得了迅速发展, 主动视觉理论框架、基于感知特征群的物体识别理论框架等新概念、新方法、新理论不断涌现。而到 90 年代, 计算机视觉在工业环境中得到广泛应用 (Mouaddib 1997, Zhang 1997, Tech 1995), 同时基于多视几何的视觉理论得到迅速发展 (Xu 1996)。

在上面的讨论中, 计算机视觉 (computer vision) 和机器视觉 (machine vision) 两个术语是不加以区分的, 在很多文献中也是如此, 但其实这两个术语是既有区别又有联系的。计算机视觉是采用图像处理、模式识别、人工智能技术相结合的手段, 着重于一幅或多幅图像的计算机分析。图像可以由单个或者多个传感器获取, 也可以是单个传感器在不同时刻获取的图像序列。分析是对目标物体的识别, 确定目标物体的位置和姿态, 对三维景物进行符号描述和解释。在计算机视觉研究中, 经常使用几何模型、复杂的知识表达, 采用基于模型的匹配和搜索技术, 搜索的策略常使用自底向上、自顶向下、分层和启发式控制策略。机器视觉则偏重于计算机视觉技术工程化, 能够自动获取和分析特定的图像, 以控制相应的行为。具体地说, 计算机视觉为机器视觉提供图像和景物分析的理论及算法基础, 机器视觉为计算机视觉的实现提供传感器模型、系统构造和实现手段。因此可以认为, 一个机器视觉系统就是一个能自动获取一幅或多幅目标物体图像, 对所获取图像的各种特征量进行处理、分析和测量, 并对测量结果做出定性分析和定量解释, 从而得到有关目标物体的某种认识并做出相应决策的系统。机器视觉系统的功能包括: 物体定位、特征检测、缺陷判断、目标识别、计数和运动跟踪。

### 1.1.2 机器视觉系统构成

机器视觉系统一般以计算机为中心, 主要由视觉传感器、高速图像采集系统及专用图像处理系统等模块构成, 如图 1.1 所示。

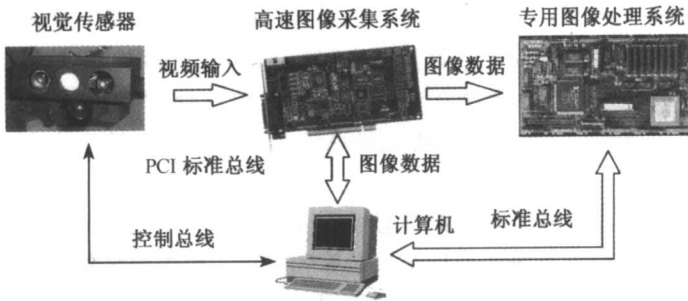


图 1.1 机器视觉系统基本组成模块

视觉传感器是整个机器视觉系统信息的直接来源，主要由一个或者两个图像传感器组成，有时还要配以光投射器及其他辅助设备。它的主要功能是获取足够的机器视觉系统要处理的最原始图像。图像传感器可以使用激光扫描器、线阵和面阵 CCD 摄像机或者 TV 摄像机，也可以是最新出现的数字摄像机等。尤其是线阵和面阵 CCD 摄像机，它们在计算机视觉的发展和应用中起着至关重要的作用。随着半导体集成技术和超大规模微细加工技术的发展，面阵 CCD 摄像机不仅商品化，而且具有高分辨率和高工作速度。另外，它所具有的二维特性、高灵敏度、可靠性好、几何畸变小、无图像滞后和图像漂移等优点使其成为计算机视觉中非常适合的图像传感器。光投射器可以作为普通照明光源、半导体激光器或者红外激光器等，它的功能主要是参与形成被分析的物体图像的特征。其他辅助设备为传感器提供电源和控制接口等功能。

进入 20 世纪 90 年代，为满足对小型化、低功耗和低成本成像系统消费需求的增长，出现了几种新的固体图像传感技术，其中最引人注目且最有发展潜力的是采用标准 CMOS 半导体工艺生产的图像传感器，即 CMOS 图像传感器。可以预计，CMOS 图像传感器以其独特的优点在计算机视觉系统中将具有广泛的应用前景。

高速图像采集系统是由专用视频解码器、图像缓冲器和控制接口电路组成的。它的主要功能是实时地将视觉传感器获取的模拟视频信号转换为数字图像信号，并将图像直接传送给计算机进行显示和处理，或者将数字图像传送给专用图像处理系统进行视觉信号的实时前端处理。随着专用视频解码器芯片和现场可编程逻辑门阵列（FPGA）芯片的出现，现在的大多数高速图像采集系统由少数几个芯片就可以完成。图像采集系统与计算机的接口采用工业标准总线，如 ISA 总线、VME 总线或者 PCI 总线等，使得图像采集系统到计算机的实时图像数据传输成为可能。

专用图像处理系统是计算机的辅助处理器，主要采用专用集成芯片（ASIC）、数字信号处理器（DSP）或者 FPGA 等设计的全硬件处理器。它可以

实时高速完成各种低级图像处理算法，减轻计算机的处理负荷，提高整个视觉系统的速度。专用图像处理系统与计算机之间的通信可以采用标准总线接口、串行通信总线接口或者网络通信等方式。随着各种硬件处理系统的出现，如基于FPGA的超级计算机和实时低级图像处理系统等，为机器视觉系统实时实现提供有利的条件。

计算机是整个机器视觉系统的核心，它除了控制整个系统的各个模块的正常运行外，还承担着视觉系统的最后结果运算和输出。由图像采集系统输出的数字图像可以直接传送到计算机，由计算机采用纯软件方式完成所有的图像处理和其他运算。如果纯软件处理能够满足视觉系统的要求，专用硬件处理系统就不出现在机器视觉系统中。这样，一个实用机器视觉系统的结构、性能、处理时间和价格等都可以根据具体应用而定，因此比较灵活。

针对有些机器视觉系统，还需配有相应的工件传输和定位系统，以使待监测的工件通过特定的传送系统安放到预定的空间内，必要的时候，加以定位限制。

为适应现代工业发展的需要，在各种小型机、微型机，特别是在功能强大的IBM-PC机上开发各种专用微型视觉组件变得更为重要。越来越多的公司投入大量人力物力研究视觉组件产品。单就美国而言，早在1983年底就有一百多家公司跻身于计算机视觉系统的市场，而到现在近二十年的发展，投入到这个领域的公司不计其数。随着微处理器和超大规模集成电路技术日益成熟，从而能生产出更小、更先进、更灵活可靠耐用的视觉组件产品，并使它们走出实验室进入实际工作现场。

在二维视觉处理中，随着机器视觉的飞速发展，已从二值视觉系统发展为灰度视觉系统，并达到实用。二值视觉系统仅通过像素0到1或由1到0的变化提取图像边缘点，它需要高对比度图像。灰度视觉系统具有检测复杂场景的能力，如复杂工件识别和表面特征（纹理、阴影、模式等）分析。采用一定的算法，系统精度受照明变化的影响很小。灰度是图像辐射度或亮度的量化测量。该信息是通过视频A/D转换器存储在帧存体中获得的，灰度分辨率随机器视觉系统的不同而不同，但数值通常是2的乘方：4、16、64和256。灰度分辨率将确定视觉系统检测区域亮度值的最小变化。灰度分辨率结合“子像素”能力在机器视觉系统中起着重要作用。

在三维视觉信息获取上，近年来也取得了巨大的进步，由于实现思想和条件不同，产生了响应的诸多方法。例如，根据照明方式可分为主动测距法和被动测距法。前者需要利用特别的光源所提供的结构信息，而后者获取深度信息是在自然光下完成的。被动测距法适合于受环境限制和需保密的场合。而主动测距法可应用的领域非常广泛，且具有测量精度高、抗干扰性能好和实时性强等优点。总之，三维视觉的引入进一步扩大了机器视觉的应用领域。

此外, 机器视觉系统的界面是开放性的, 用户可根据应用需要进行计算机编程, 以改善系统的功能。为进一步使机器视觉系统与工业自动化现场相适应, 由菜单驱动系统, 用鼠标或光笔在屏幕上选择功能项即可, 易于操作, 无需专门培训操作员。

机器视觉系统具有高度的智能和普遍的适应性, 随着它的不断完善, 已完全能用于工业现场, 满足现代生产过程的要求。

## 1.2 Marr 的视觉理论框架

20 世纪 80 年代初, Marr (1982) 首次从信息处理的角度综合了图像处理、心理物理学、神经生理学及临床神经病学等方面已取得的重要研究成果, 提出了第一个较为完善的视觉系统框架, 使计算机视觉研究有了一个比较明确的体系。虽然这个理论还需要通过研究不断改进和完善, 但 Marr 的视觉计算理论是首次提出的阐述视觉机理的系统理论, 并且对人类视觉和计算机视觉的研究都产生了深远的推动作用。下面简要介绍 Marr 视觉理论的基本思想及理论框架 (Marr 1982, 吴健康 1993)。

### 1.2.1 视觉系统研究的三个层次

Marr 从信息处理系统的角度出发, 认为对视觉系统的研究应分为三个层次, 即计算理论层次、表达与算法层次和硬件实现层次。

计算理论层次要回答视觉系统的计算目的与计算策略是什么, 或视觉系统的输入输出是什么, 如何由系统的输入求出系统的输出。在这个层次上, 视觉系统输入是二维图像, 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研究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。表达与算法层次是要进一步回答如何表达输入和输出信息, 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。一般来说, 不同的表达方式, 完成同一计算的算法会不同, 但 Marr 算法与表达是比计算理论低一层次的问题, 不同的表达与算法, 在计算理论层次上可以是相同的。最后一个硬件层次, 是解决用硬件实现上述表达和算法的问题, 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。

区分以上三个不同层次，对于深刻理解计算机视觉与生物视觉系统以及它们的关系都是有益的，例如，人的视觉系统与目前的计算机视觉系统在“硬件实现”层次上是完全不同的，前者是极为复杂的神经网络，而后者是目前使用的计算机，但它们可能在计算理论层次上完成相同的功能。

视觉系统研究的三个层次可归纳为表 1.1。

表 1.1 视觉系统研究的三个层次的含义

要素	名称	含义和所解决的问题
1	计算理论	什么是计算目的，为什么要这样计算
2	表达和算法	怎样实现计算理论，什么是输入输出表达，用什么算法实现表达间的转换
3	硬件实现	怎样在物理上实现表达和算法，什么是计算结构的具体细节

### 1.2.2 视觉信息处理的三个阶段

Marr 从视觉计算理论出发，将系统分为自下而上的三个阶段，即视觉信息从最初的原始数据（二维图像数据）到最终对三维环境的表达经历了三个阶段的处理，如图 1.2 及表 1.2 所示。第一阶段（也称为早期阶段）构成所谓“要素图”或“基元图”（primary sketch），基元图由二维图像中的边缘点、直线段、曲线、顶点、纹理等基本几何元素或特征组成；第二阶段（中期阶段），Marr 称为对环境的 2.5 维描述，2.5 维描述是一种形象的说法，意即部分的、不完整的三维信息描述，用“计算”的语言来讲，就是重建三维物体在观察者为中心的坐标系下的三维形状与位置。当人眼或摄像机观察周围环境物体时，观察者对三维物体最初是以自身的坐标系来描述的，另外，我们只能观察到物体的一部分（另一部分是物体的背面或被其他物体遮挡的部分）。这样，重建的结果是以观察者坐标系下描述的部分三维物体形状，称为 2.5 维描述。这一阶段中存在许多并行的相对独立的模块，如立体视觉、运动分析、由灰度恢复表面形状等不同处理单元。事实上，从各种不同角度去观察物体，观察到的形状都是不完整的，不能设想，人脑中存有同一物体从所有可能的观察角度看到的物体形象，以用来与所谓的物体的 2.5 维描述进行匹配与比较。因此，2.5 维描述必须进一步处理以得到物体的完整三维描述，而且必须是物体本身某一固定坐标系下的描述，这一阶段称为第三阶段（后期阶段）。视觉信息处理的三个阶段如表 1.2 所示。

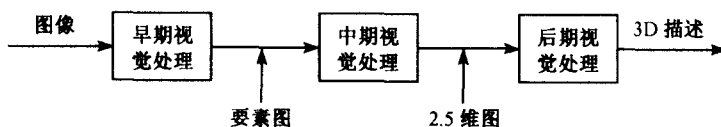


图 1.2 Marr 框架的视觉三阶段

表 1.2 由图像恢复形状信息的表达框架

名 称	目 的	基 元
图像	亮度表示	图像中每一点的亮度值
基元图	表示二维图像中的重要信息, 主要是图像中的亮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, 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, 积极推动了这一领域的研究, 多年来对图像理解和计算机视觉的研究发展起了重要的作用。但 Marr 理论也有其不足之处, 其中有 4 个有关整体框架 (图 1.2) 的问题 (吴立德 1993):

- (1) 框架中输入是被动的, 给什么图像, 系统就处理什么图像。
- (2) 框架中加工目的不变, 总是恢复场景中物体的位置和形状等。
- (3) 框架缺乏或者说未足够重视高层知识的指导作用。
- (4) 整个框架中信息加工过程基本自下而上, 单向流动, 没有反馈。

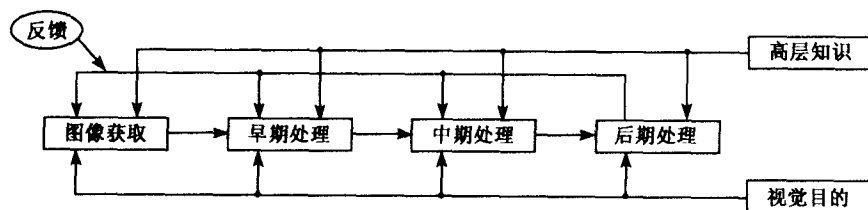


图 1.3 改进的 Marr 框架

针对上述问题, 近年来人们提出了一系列改进思路, 对应图 1.2 的框架, 可将其改进并融入新的模块得到图 1.3 的框架, 具体改进如下:

(1) 人类视觉是主动的, 会根据需要改变视角, 以帮助识别。主动视觉指视觉系统可以根据已有的分析结果和视觉的当前要求, 决定摄像机的运动以从合适的视角获取相应的图像。人类的视觉又是有选择的, 可以注视 (以较高分辨率观察感兴趣区域), 也可以对场景中某些部分视而不见。选择性视觉指视觉系统可以根据已有的分析结果和视觉的当前要求, 决定摄像机的注意点以获取相应的图像。考虑到这些因素, 在改进框架中增加了图像获取模块。该模块要根据视觉目的的选择采集方式。

(2) 人类的视觉可以根据不同的目的进行调整。有目的视觉（也称定性视觉）指视觉系统根据视觉目的进行决策，例如，是完整地恢复场景中物体的位置和形状等还是仅仅检测场景中是否有某物体存在。事实上，有相当场合只需定性结果就可以，并不需要复杂性高的定量结果。因此在改进框架中增加了视觉目的模块（Aloimonos 1992），但定性分析还缺乏完备的数学工具。

顺便指出，有一种相关的观点认为马尔关于对场景先重建后解释的思路可以简化视觉任务，但与人的视觉功能并不完全吻合。事实上重建和解释不总是串行的。

(3) 人类可在仅从图像获取部分信息的情况下完全解决视觉问题，原因是隐含地使用了各种知识。例如，借助 CAD 设计资料获取物体形状信息（使用物体模型库），可帮助解决由单幅图恢复物体形状的困难。利用高层知识可解决低层信息不足的问题，所以在改进框架中增加了高层知识模块（Huang 1993）。

(4) 人类视觉中前后处理之间是有交互作用的，尽管对这种交互作用的机理了解得还不充分，但高层知识和后期处理的反馈信息对早期处理的作用是重要的。从这个角度出发，在改进框架中增加了反馈控制流向。

最后需要指出，限于历史等因素，Marr 没有研究如何用数学方法严格地描述视觉信息的问题，虽然较充分地研究了早期视觉，但基本没有论及对视觉知识的表达、使用和基于视觉知识的识别等。近年来有许多试图建立计算机视觉理论框架的工作，其中 Grossberg 宣称建立了一个新的视觉理论：表观动态几何学（dynamic geometry of surface form and appearance）。它指出感知的表面形状是分布在多个空间尺度上多种处理动作的总结果，因此 2.5 维图并不存在，向 Marr 的理论提出了挑战（吴健康 1993）。但 Marr 的理论使得人们对视觉信息的研究有了明确的内容和较完整的基本体系，仍被看作是研究的主流。现在新提出的理论框架均包含它的基本成分，多数被看作它的补充和发展。尽管 Marr 的理论在许多方面还存在争议，但至今它仍是广大计算机视觉工作者所普遍接受的计算机视觉理论基本框架。

### 1.3 机器视觉的应用领域及面临的问题

机器视觉在国民经济、科学研究及国防建设等领域都有着广泛的应用。视觉的最大优点是与被观测的对象无接触，因此对观测与被观测者都不会产生任何损伤，这是其他感觉方式无法比拟的。另外，视觉方式所能检测的对象十分广泛，人眼观察不到的范围，机器视觉也可以观察，例如，红外线、微波、超声波等人类就观察不到，而机器视觉则可以利用这方面的敏感器件形成红外线、微波、超声波等图像。因此可以说是扩展了人类的视觉范围。另外人无法长时间地观察对

象, 机器视觉则不知疲劳, 始终如一地观测, 所以机器视觉可以广泛地用于长时间恶劣的工作环境。不过, 机器视觉技术仍处于一个完善和发展的阶段, 其发展还远远落后于人们所寄予的发展水平。

### 1.3.1 机器视觉的应用领域

机器视觉应用系统将在第 10、11 章介绍, 下面只列举一些机器视觉的主要应用领域。

(1) 工业自动化生产线应用。产品检测, 工业探伤, 自动流水线生产和装配, 自动焊接, PCB 印制板检查, 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) 各类检验和监视应用。标签文字标记检查, 邮政自动化, 计算机辅助外科手术, 显微医学操作, 石油、煤矿等钻探中数据流自动监测和滤波, 在纺织、印染业进行自动分色、配色, 重要场所门廊自动巡视, 自动跟踪报警等。

(3) 视觉导航应用。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、精确制导及自动巡航捕获目标和确定距离等, 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) 图像自动解释应用。对放射图像、显微图像、医学图像、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。

(5) 人机交互应用。人脸识别、智能代理等。同时让计算机可借助人体的手势动作(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。

(6) 虚拟现实应用。飞机驾驶员训练、医学手术模拟、场景建模、战场环境表示等, 它可帮助人们超越人的生理极限, “亲临其境”, 提高工作效率。

机器视觉的应用是多方面的, 它已经取得并将继续取得越来越广泛的应用。

### 1.3.2 机器视觉面临的问题

对于人的视觉来说, 由于人的大脑和神经的高度发展其目标识别能力很强。但是人的视觉也同样存在障碍, 例如, 即使具有一双敏锐视觉和极为高度发达头脑的人一旦置于某种特殊环境(即使曾经具备一定的先验知识), 其目标识别能力也会急剧下降。事实上人们在这种环境下面对简单物体时, 仍然可以有效而简便地识别; 而在这种情况下面对复杂目标或特殊背景时, 则在视觉功能上发生障碍。两者共同的结果是导致目标识别的有效性和可靠性的大幅度下降。将人的视



觉引入机器视觉中，机器视觉也存在着这样的障碍。它主要表现在三个方面（王斌 1996）：一是如何准确、高速（实时）地识别出目标；二是如何有效地增大存储容量，以便容纳下足够细节的目标图像；三是如何有效地构造和组织出可靠的识别算法，并且顺利地实现。前两者相当于人的大脑这样的物质基础，这期待着高速的阵列处理单元，以及算法（如神经网络、分维算法、小波变换等算法）的新突破，用极少的计算量以及高度地并行性实现功能。为了便于理解，现将人的视觉与机器视觉的对比列于表 1.3、表 1.4。

表 1.3 机器视觉与人的视觉能力比较

能力	机器视觉	人的视觉
测距	能力有限	定量估计
定方向	定量计算	定量估计
运动分析	定量分析，但受限制	定量分析
检测边界区域	对噪声比较敏感	定量、定性分析
图像形状	受分割、噪声制约	高度发达
图像机构	需要专用软件，能力有限	高度发达
阴影	初级水平	高度发达
二维解释	对分割完善的目标能较好解释	高度发达
三维解释	较为低级	高度发达
总的的能力	最适合于结构环境的定量测量	最适合于复杂的、非结构化环境的定量解释

表 1.4 机器视觉与人的视觉性能标准比较

性能标准	机器视觉	人的视觉
分辨率	能力有限	定量估计
处理速度	零点几秒/每帧图像	定量估计
处理方式	串行处理，部分并行处理	每只眼睛每秒处理（实时） $10^{10}$ 空间数据
视觉功能	二维、三维立体视觉有限	自然形式三维立体视觉
感光范围	紫外、红外、可见光	可见光

另外，由于当前对人类视觉系统和机理、人脑心理和生理的研究还不够，目前人们所建立的各种视觉系统绝大多数是只适用于某一特定环境或应用场合的专用系统，而要建立一个可与人类的视觉系统相比拟的通用视觉系统是非常困难的。主要原因有以下几点：

(1) 图像对景物的约束不充分。首先是图像本身不能提供足够的信息来恢复景物，其次是当把三维景物投影成二维图像时丧失了深度信息。因此，需要附加约束才能解决从图像恢复景物时的多义性。

(2) 多种因素在图像中相互混淆。物体的外表受材料的性质空气条件、光源角度、背景光照摄像机角度和特性等因素的影响。所有这些因素都归结到一个单