



華夏英才基金學術文庫

高隽 谢昭 著

图像理解理论与方法



科学出版社

www.sciencep.com

华夏英才基金学术文库

图像理解理论与方法

高 隽 谢 昭 著

TP3
G145

科学出版社

北 京

内 容 简 介

“图像理解”是近年来计算机科学的热点研究领域,本书对图像理解的前沿理论与方法进行了详细论述。主要内容包括分类判别模型、生成模型、图像信息表示与特征提取、场景中的目标识别、场景中目标之间的关系、场景描述与理解、场景中的句法语义、图像理解开发环境和图像数据集等。本书紧跟上述内容的国内外发展现状和最新成果,阐述作者对图像理解理论方法的理解和认识。

本书可以作为计算机科学与技术、信息与通信工程、电子科学与技术等专业的研究生、高年级本科生教材,同时可作为从事图像理解、计算机视觉、机器学习等相关专业研究人员的参考书。

图书在版编目(CIP)数据

图像理解理论与方法/高隽,谢昭著. —北京:科学出版社,2009
(华夏英才基金学术文库)
ISBN 978-7-03-025757-4

I. 图… II. ①高…②谢… III. 图像处理—研究 IV. TP391.41

中国版本图书馆 CIP 数据核字(2009)第 180769 号

责任编辑:王志欣 任 静 于宏丽 / 责任校对:赵燕珍
责任印制:赵 博 / 封面设计:陈 敬

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

*

2009 年 10 月第 一 版 开本:B5(720×1000)

2009 年 10 月第一次印刷 印张:34 1/2

印数:1—3 000 字数:683 000

定价:68.00 元

(如有印装质量问题,我社负责调换(双青))

前 言

图像理解是近年来的热点研究领域,它是一门覆盖范围很广的综合性交叉学科,涉及图像工程、计算机视觉、人工智能和认知学等诸多学科领域,在目标识别及解释和图像与视频检索等方面有着广阔的应用前景。

本书的总体编写思路:给出图像理解的定义及框架,并将其与图像工程、人工智能的相关方法有机结合起来,逐步阐述图像理解的理论方法;以图像理解的研究内容为导向,详细介绍图像理解的方法,以及图像理解系统必不可少的开发环境和图像数据机。基于这一编写思路,本书介绍了数据/知识驱动方式不同的判别模型和生成模型,以此作为整个图像理解任务的理论基础;接着从图像理解的研究内容入手,分别详述了场景中目标识别、场景中目标之间的关系、场景描述与理解和图像句法语义四个子任务,并介绍了相应的学习算法,给出了经典、实用的应用实例。具体内容分布如下:

第1章给出了图像理解的定义以及结构框架,介绍了图像理解的研究内容,阐述了图像理解与图像工程、计算机视觉、人工智能和认知学的联系与区别及其应用;第2章介绍无结构特征组织方式的分类判别模型,有 Boosting、支持向量机以及协同神经网络;第3章介绍半结构化的生成模型,主要有无向图模型中的随机场模型和星群模型,以及有向图模型中的认知图模型、pLSA 模型和 LDA 模型;第4章从图像的处理技术入手,介绍了图像特征表示、提取和评价三个方面内容,并给出部分应用实例加以分析,重点介绍了特征不变性描述子 SIFT;第5章介绍场景中的目标识别,主要分为图像分割、目标识别以及广义目标识别三个方面,介绍了以第2、3章为理论基础的各种方法,给出仿真应用实例或经典实例;第6章介绍场景中目标之间的关系,首先简单介绍表达场景中目标之间关系的有力载体(与或图和解析图),然后介绍关联描述的视觉三层(低层、中层和高层)词汇,接着详细介绍目标之间的各种关联和结构,最后给出体现目标之间关系的经典应用实例;第7章介绍场景的描述与理解,首先介绍图像理解中场景分类的相关概念和特点;然后给出场景分类的方法(低层场景建模和高层语义建模),最后重点介绍了基于视皮层组织的场景 Gist 特征的场景全局感知分类,以及基于混合高斯模型的场景分类,并在高斯概率模型进行场景分类研究的基础上对场景分类的约束机制进行初步探讨;第8章介绍场景中的句法语义,首先简单介绍了句法语言,然后重点介绍了上下文有关随机句法的学习和推理过程,最后给出基于统计句法策略的实现新型语义关系的构建;第9章依次介绍了三种图像理解开发环境平台 IUE、OpenCV 和

VXL,详细介绍其起源、类谱系、任务库、执行界面和接口及其应用实例;第10章介绍了图像理解环境下的图像数据集,以不同的图像理解任务为出发点,依次介绍了传统图像集和融入视觉知识的新型图像集。

本书是在三项国家自然科学基金项目(60375011、60575028、60875012)、教育部“新世纪优秀人才支持计划”(NCET-04-0560)、安徽省高等学校科技创新团队(2005TD04)、安徽省优秀青年科技基金(04042044),以及安徽省优秀青年科技基金择优滚动支持项目(08040106901)等资助计划项目研究工作的基础上完成的。在此,特向国家自然科学基金委员会、教育部、安徽省自然科学基金委员会、科技厅和教育厅表示衷心感谢!

本书是合肥工业大学图像信息处理研究室创新团队集体智慧的结晶,骆祥峰、胡良梅、谢昭、邵静、钱乐乐、徐小红、赵莹等博士,鲍捷、陈定国、董火明等硕士,以及吴克伟、冯文刚、张骏、苏菱、黄帅、朱士蓉、孙永宣、张利利、杨胜利、丁津津、宋皓、吴良海等在读博士/硕士生为本书的出版做了大量的工作,在此向直接或间接参与本书编写工作的图像信息处理研究室全体同仁表示衷心感谢!

1996年本人在德国Stuttgart大学听了Levi教授所上的“图像理解”课,2008年初又有幸得到了Levi教授具体指导。德国Stuttgart大学博士、现Heilbronn大学的Ott教授是本人长期科研合作伙伴,为图像理解做了大量的相关工作。在此特向Levi教授、Ott教授表示衷心的感谢!

由于本人水平有限,加上图像理解学科本身在不断地丰富和发展,书中不妥之处在所难免,恳请广大读者批评指正。

高 隽

2009年5月

目 录

前言

第 1 章 绪论	1
1.1 图像理解的基本概念	1
1.1.1 图像理解与图像工程	2
1.1.2 图像理解与计算机视觉	4
1.1.3 图像理解与人工智能	6
1.1.4 图像理解与认知学	9
1.2 图像理解的研究内容	11
1.2.1 场景中目标识别	12
1.2.2 场景中目标之间的关系	16
1.2.3 场景描述与理解	20
1.2.4 图像语义描述推理	22
1.3 图像理解的研究方法	25
1.3.1 判别分类方法	25
1.3.2 生成模型方法	25
1.3.3 句法语义分析方法	26
1.4 图像理解的应用	26
1.4.1 遥感图像解释	27
1.4.2 目标识别和解释	28
1.4.3 基于内容的图像和视频检索	29
参考文献	31
第 2 章 分类判别模型	34
2.1 引言	34
2.2 Boosting 分类方法	36
2.2.1 Boosting 产生与发展	36
2.2.2 Boosting 基本思想	37
2.2.3 Boosting 分类模型	38
2.2.4 方法总结	46
2.3 SVM 分类方法	47
2.3.1 统计学习理论	47

2.3.2	SVM 模型	48
2.3.3	方法总结	59
2.4	协同学与协同神经网络	60
2.4.1	协同学简介	60
2.4.2	协同模式识别方法	66
2.4.3	方法总结	78
2.5	总结	78
	参考文献	79
第 3 章	生成模型	85
3.1	引言	85
3.1.1	图论中的无向图与有向图	86
3.1.2	图像理解中的标记问题	87
3.2	无向图模型	90
3.2.1	无向图简介	90
3.2.2	随机场模型	94
3.2.3	星群模型	125
3.2.4	小结	137
3.3	有向图模型	137
3.3.1	有向图简介	139
3.3.2	认知图模型	141
3.3.3	pLSA 模型	148
3.3.4	LDA 模型	153
3.3.5	小结	158
3.4	总结	160
	参考文献	160
第 4 章	图像信息表示与特征提取	169
4.1	引言	169
4.2	图像信息表示	170
4.2.1	图像数据结构	170
4.2.2	知识表示	173
4.2.3	数据与知识的融合	173
4.3	图像特征提取	175
4.3.1	基本图像特征提取	176
4.3.2	常用图像特征提取	186
4.3.3	方法小结	208

4.4 图像特征表达	209
4.4.1 直方图表达	209
4.4.2 区域特征表达	210
4.4.3 边缘特征表达	216
4.4.4 基于包的表达	218
4.4.5 方法小结	221
4.5 图像特征评价	221
4.5.1 检测算子评价	221
4.5.2 特征描述子评价	223
4.5.3 方法小结	225
4.6 总结	225
参考文献	225
第5章 场景中的目标识别	233
5.1 引言	233
5.2 图像分割	233
5.2.1 基于 SVM 的图像分割	234
5.2.2 基于取样的图像分割	238
5.2.3 全互连结构的图像分割	248
5.2.4 MRF+pLSA 区域分割标记	265
5.2.5 基于产生式规则的图像分割	272
5.3 目标识别	287
5.3.1 基于认知图的目标形状识别	288
5.3.2 基于协同神经网络的生物特征识别	300
5.3.3 基于 Boosting 的目标识别	310
5.3.4 基于 SVM 的目标识别	322
5.4 广义目标识别	335
5.4.1 Boosting 多值分类的目标检测识别	335
5.4.2 视觉注意机制引导的协同目标识别	346
5.4.3 pLSA 的视觉目标分类	352
5.4.4 pLSA 下的无向图广义目标识别	354
5.5 总结	361
参考文献	362
第6章 场景中目标之间的关系	368
6.1 引言	368
6.2 与或图 and 解析图	368

6.3 视觉词汇	370
6.3.1 视觉词汇表达	370
6.3.2 低层图像基元	371
6.3.3 中层图基元对	373
6.3.4 高层目标部分	374
6.4 关联和结构	376
6.4.1 关联	376
6.4.2 结构	383
6.5 目标间关系的视觉应用	385
6.5.1 星群模型的部分关联分析	385
6.5.2 场景-目标关联的目标识别	390
6.6 总结	396
参考文献	396
第7章 场景描述与理解	399
7.1 引言	399
7.2 场景分类	400
7.2.1 场景分类的概念	400
7.2.2 场景分类的特点	400
7.2.3 场景的视觉感知层次	402
7.2.4 场景分类的方法	403
7.3 场景理解的视觉应用	411
7.3.1 基于 Gist 特征的场景全局感知分类	411
7.3.2 基于高斯统计概率模型的场景分类	424
7.3.3 图像理解的场景分析约束机制	430
7.4 总结	432
参考文献	432
第8章 场景中的句法语义	436
8.1 引言	436
8.2 句法语言	437
8.2.1 句法重用和歧义结构	437
8.2.2 语义词汇表达	440
8.2.3 WordNet 词汇网	440
8.3 基于统计的句法分析	445
8.3.1 句法公式	446
8.3.2 随机句法	447

8.3.3 上下文有关随机句法	449
8.3.4 随机句法与或图	450
8.3.5 句法学习与推理	456
8.4 基于统计句法的视觉应用	464
8.4.1 人造场景解析	465
8.4.2 人体外观建模与推理	468
8.4.3 目标类别推理识别	469
8.5 总结	470
参考文献	471
第9章 图像理解开发环境	474
9.1 引言	474
9.2 图像理解环境	474
9.2.1 IUE 起源	474
9.2.2 IUE 类谱系	475
9.2.3 IUE 任务库	475
9.2.4 IUE 执行界面和接口	477
9.3 OpenCV	477
9.3.1 OpenCV 起源	477
9.3.2 OpenCV 类谱系	478
9.3.3 OpenCV 任务库	479
9.3.4 OpenCV 执行界面和接口	481
9.3.5 OpenCV 应用实例	481
9.4 VXL	489
9.4.1 VXL 起源	489
9.4.2 VXL 类谱系	490
9.4.3 VXL 任务库	490
9.4.4 VXL 执行界面和接口	491
9.4.5 VXL 应用实例	491
9.5 总结	491
参考文献	492
第10章 图像数据集	494
10.1 引言	494
10.2 传统图像集	494
10.2.1 一般目标识别图像集	495
10.2.2 图像检索图像集	509

10.2.3 手势识别图像集	512
10.2.4 数字识别图像集	519
10.2.5 PASCAL 图像集	523
10.3 融合视觉知识的图像集	525
10.3.1 图像集中的视觉知识	526
10.3.2 LabelMe 图像集	527
10.3.3 LotusHill 图像集	534
10.4 总结	538
参考文献	539

第 1 章 绪 论

1.1 图像理解的基本概念

图像理解(image understanding, IU)就是对图像的语义解释。它是以图像为对象,知识为核心,研究图像中有什么目标、目标之间的相互关系、图像是什么场景以及如何应用场景的一门科学。图像理解的输入是数据,输出是知识,属图像研究领域的高层内容。它将中低层的数据处理分析(目标识别)与高层的知识表达推理(场景描述与理解)有效结合,实现数据分析形成知识推理,知识反馈用以指导数据分析获取与分析。由此可见,图像理解的过程具有并行性、层次性与反馈性的特点。

图像理解包括从下到上的数据驱动和从上到下的知识驱动,是渐进的理解过程。其内容主要包括两方面:一是场景中的目标识别,二是场景描述与理解。目标识别是为了对场景进行更好的解释,是场景描述与理解的基础,具有主动性;场景描述与理解为目标识别提供先验信息,可以指导目标识别。因此,图像理解的主要内容也具有反馈特性。图 1-1 显示了图像理解两大研究内容及其相互间的关系。

场景中的目标识别侧重对局部区域的理解,场景描述与理解侧重对图像景物的整体分析,两者相互影响、共同作用,其中目标识别是图像理解中最重要的研究内容,就是为了确定图像中有什么物体或目标以及它们出现的位置,长期以来对目标识别做了较为广泛和深入

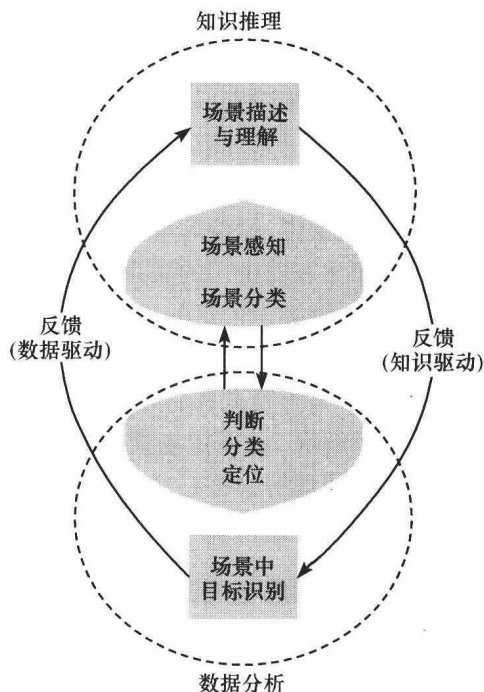


图 1-1 图像理解两大内容及其相互间的关系

的研究。在现实世界中,目标的出现都不是相互孤立的,与其周围环境、场景有着密切的联系,其研究思路主要分为两大策略:一是将不同尺度的窗口遍历滑动整个图像区域,或采用不同的视觉选择策略提取场景中的有效区域,并采用统计、匹配、

相关等方法判断指定(窗口)区域的类别或背景(Furht, 1995; Kelly et al., 1994);二是通过统计目标间出现的可能性,构建目标间的共生关系,对目标的表示、位置以及尺度进行鲁棒估计,完成具有上下文关系的目标位置及类别信息的强预测(边肇祺等, 2000)。前者主要以自底向上的数据驱动(bottom-up data-driven)为主,后者主要以自顶向下的模型驱动(top-down model-driven)为主。

场景描述与理解是近两年来的研究热点。2006年在麻省理工学院首次召开了场景理解研讨会(scene understanding symposium, SUnS),会上明确了场景描述与理解,特别是场景分类将会是图像理解一个新的有前途的研究方向。目前对场景描述与理解的研究主要分为两个方向:一是集中于视觉心理学和生理学的研究,主要研究快速场景感知的心理和生理机理(Guiroga et al., 2005; Evans et al., 2005; Oliva et al., 2001);二是研究场景分类的计算模型,通过建立简单的统计计算模型实现几类场景(室内/室外等)的判断(Oliva et al., 2001; Li et al., 2001)或是场景空间中的语义分析(Torralba et al., 2008)。

尽管图像理解的两方面内容已备受研究学者和专家的关注,但仍有很多亟待解决的问题(Porway et al., 2008; Zhang et al., 2003; Li et al., 2001; Shi et al., 2000),例如,如何用场景信息指导目标的选择注意;如何构建场景的先验信息来指导场景中的目标识别;如何将目标识别结果构建场景描述,形成可进一步进行场景中目标识别与场景理解的先验信息等。这些都为图像理解的发展开辟了新的方向和研究思路。

从本质上说,图像理解作为一门基础应用学科,与图像工程、计算机视觉、人工智能、认知学有着密切的联系,但它们侧重不同的研究内容、采用不同的研究方法。图像理解强调知识信息的处理,而传统的研究注重视觉信息的表达和处理;图像理解借用图像工程中各种图像处理技术,以及人工智能中各种智能信息处理方法,借鉴认知学中人类视觉和心理学的研究基础,区别于计算机视觉的3D世界的景物分析,着重解决2D图像场景中的目标识别、关联分析、场景描述和句法语义问题。简言之,图像理解的核心任务是对数据分析形成知识推理,知识反馈指导数据获取,将中低层的数据处理分析与高层的知识表达推理有效地结合,实现渐进理解。

1.1.1 图像理解与图像工程

图像工程(章毓晋, 2007; 2000)综合了各种图像技术,是集图像研究和应用为一体的整体框架,包括利用计算机和其他电子设备完成一系列工作,例如,图像采集、获取、编码、存储和传输,图像合成和生成,图像水印的嵌入和提取,图像的显示和输出,图像变换、增强、恢复(复原)和重建,目标检测、跟踪、表达和描述,目标特征提取和测量,序列图像校正,3D景物重建复原,图像数据库的建立、索引和抽取,图像分类、表示和识别,图像模型的建立和匹配,场景描述与理解以及相应的判断

决策和行为规划等。

图像工程内容丰富,覆盖范围广,根据研究内容和研究方法等的不同可分为三个层次:图像处理、图像分析和图像理解,三者的抽象程度和数据量成反比。图 1-2 对图像工程和图像理解的联系和区别进行了很好的诠释(章毓晋,2007)。

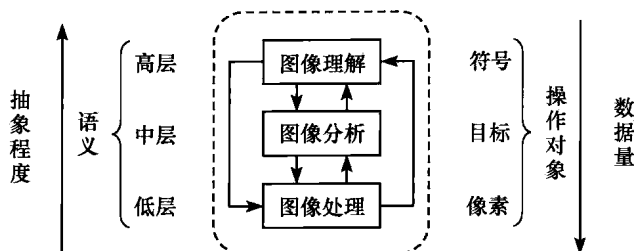


图 1-2 图像工程的三层次示意图

图像处理属于低层操作,抽象度低,数据量大,主要在图像像素级上进行处理,强调图像的变换及其之间的相互转换。广义上图像处理泛指各种图像技术,但狭义上主要是指对图像进行各种变换以改善图像的视觉效果,为其后的图像分析和理解奠定基础,或对图像进行压缩编码以减少图像的存储空间或传输时间。

图像分析属于中层操作,抽象度提高,数据量减小,主要指对图像中感兴趣的目标进行检测,实现图像分割和特征提取,然后用更简洁的数据形式描述图像。这里的数据可以是对目标特征测量的结果,或是基于测量的符号表示,其涵盖了图像中目标的特点和固有属性。

图像理解属于高层操作,抽象度高,数据量小,主要是对从描述抽象出来的符号进行运算,进一步研究图像中各目标的性质和它们之间的相互联系,并通过对图像内容含义的理解得出原来客观场景的解释,从而指导和规划主观行为,其处理过程和方法与人类的思维推理可以有许多类似之处。

位于高层的图像理解本身也具有层次特性,其低层数据是视觉感知获取的图像信息,以 Marr 视觉理论为核心的计算机视觉研究方法为主;高层数据是知识信息,以传统人工智能为理论依据,二者的交集正是图像理解的核心。图 1-3 显示了图像理解的广义内涵,实际上图像分析过程就是融合人的经验知识从原始图像生成数据的过程,属于图像理解的低层处理阶段,而图像理解的高层处理则以知识语义分析推理为主要内容,数据分析形成知识推理,知识反馈能指导数据获取与分析。

综上所述,图像处理是从图像到图像的过程,图像分析是从图像到数据的过程,而图像理解则是从数据符号到知识语义的过程。图像工程的这三个层次,宏观上来看也就是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更有用的信息,在这个过程中,不断引入语义,使操作对象发生变化,压缩数据量;同时,高层操作指导低层操作,从而提高了低层操作的效能。

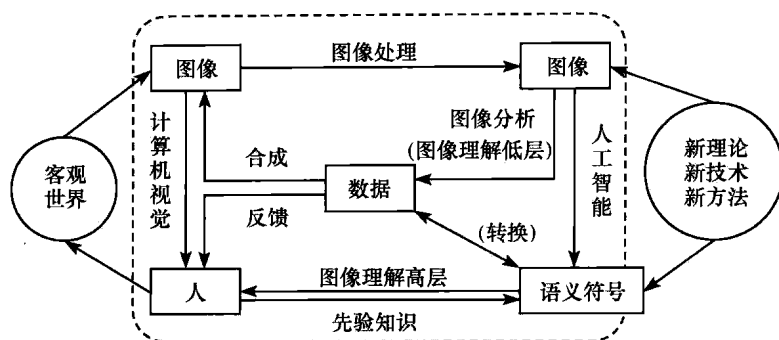


图 1-3 图像理解与图像工程等学科的关系

1.1.2 图像理解与计算机视觉

人类的很多研究以延伸人类能力为目的,早期的工作是在体力上的延伸,计算机的发明将其拓展至脑力和感知能力的延伸,计算机视觉(computer vision, CV)就是采用计算机模拟人类的视觉感知能力。早期计算机视觉也经常被称为图像理解,是指研究完成一项任务所需的视觉信息以及如何从图像中获取这些信息的研究领域。

1972年,与图像技术密切相关的第一份国际期刊——计算机图形学与图像处理(*Computer Graphics and Image Processing*, CGIP)诞生。随着计算机视觉的发展,1983年该刊改名为计算机视觉、图形学与图像处理(*Computer Vision, Graphics and Image Processing*, CVGIP)。1991年 CVGIP 又分成两个领域:一是图模型与图像处理(graph models and image processing, GMIP);二是计算机视觉和图像理解(computer vision and image understanding, CVIU)。这标志着,随着图像处理技术研究的深入,图像理解和计算机视觉从图形与图像中脱离出来,各自成为一门独立的学科。

计算机视觉主要强调用计算机实现人类的视觉功能。视觉是人类观察世界、认知世界的重要功能手段,人类从外界获取的信息约有 75% 来自视觉系统,人类视觉过程可以看成是从感觉到知觉的过程,即感受到的是对 3D 世界的 2D 投影的二维图像,以及由 2D 图像认知 3D 世界的内容和含义。从狭义上讲,视觉的最终目的是要对场景做出对观察者有意义的解释和描述;从广义上讲,还包括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的行为规划。因此,计算机视觉依据人脑视觉的工作机理,强调用计算机实现人类的视觉功能,对客观世界的 3D 场景进行感知、识别和理解。

计算机视觉中需要解决两大任务。第一项任务是 3D 场景在 2D 图像中的表达。20 世纪 70 年代,美国麻省理工学院的 Marr 教授提出了人类的视觉计算理

论,奠定了计算机视觉的基础。由于视觉过程是一种复杂的信息处理过程,其目的是要把握外部世界各种有用的情况,并将其表达出来,这种任务具有鲜明的层次特点,视觉信息处理必须在若干不同的层次上加以理解,而各个层次基本上是独立的,Marr 期望通过这种层次方式解决 3D 场景如何在 2D 图像中表达的问题。原始图像作为一种 2D 表达,可以提供各种物体的轮廓信息,但只利用轮廓等 2D 基元表达的场景和物体并不能保证得到其唯一的解释,从而引起视错觉。

如图 1-4 所示,如果观察者将注意力集中于图(a)右上方三线相交处,则会解释为图(b),即认为成像的立方体见图(c);如果集中于图(a)左下方三线相交处,则会解释为图(d),即认为成像的立方体见图(e)(艾海舟,2003)。这是因为图(a)给人以(一部分)3D 物体(立方体)的线索,但当人借助经验知识试图从中恢复 3D 深度时,由于所采用方法的不同可得出两种不同的解释。虽然在 2D 图像中存在边缘等物体表面信息的几何描述,但由于视觉任务的复杂性,需要规定一系列的步骤进行改进,Marr 理论通过分析 3D 场景如何在 2D 图像中表达,得到了解决或理解一个计算机视觉的特定问题的途径。依据 Marr 理论,如图 1-5 所示,图像视觉信息能够形成可见信息的三级表达结构,从而解决 Necker 等描述的由大脑意识混乱引起的方体错觉问题,为完成计算机视觉中的第一项任务奠定了理论基础。

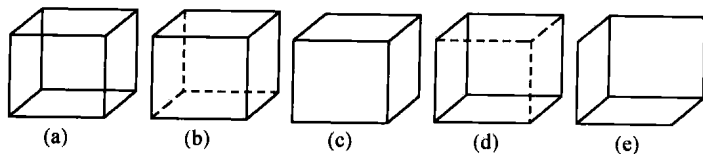


图 1-4 Necker 方体和两种可能的解释

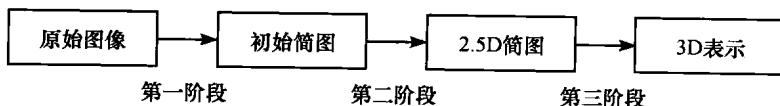


图 1-5 Marr 视觉理论三个阶段

计算机视觉的第二项关键任务是如何将 2D 图像转化为 3D 场景的描述。基于图像技术的 2D 图像几何描述实质上仍然不是以物体为中心的,这就有必要去除表面信息对位置的依赖性,将其转化为以物体为中心的描述。

在 Marr 理论框架之下,视觉信息的三个阶段被称为低层(早期)视觉、中层(中期)视觉以及高层(后期)视觉。低层视觉的任务在于图像传感和预处理,即清楚地表示原始 2D 图像中的重要信息,如零交叉、角点、边界、线段等,它们对应图像中灰度急剧变化的位置及其几何分布和组织结构;中层视觉则由输入图像和低层视觉输出的初始简图得到 2.5D 简图,比如在以观察者为中心的坐标系中可见表面法向、大致深度及不连续轮廓等。按 Marr 理论,中层视觉处理又可归为

Shape-from-X, 此处 X 指代运动(motion)、立体(stereo)、阴影(shading)、轮廓(contour)以及纹理(texture)等;高层视觉涉及由输入图像、初始简图以及 2.5D 简图获得物体的 3D 表示,即在以物体为中心的坐标系中获得物体的几何结构和空间位置关系,表 1-1 显示了视觉可计算性问题的表达框架,Marr 理论通过这一表达框架完成 2D 图像中的 3D 场景描述(章毓晋,2000)。

表 1-1 视觉可计算性问题的表达框架

名 称	表达目的	表达形式
2D 原始图像	表达场景和物体的亮度	像素(值)
2. 1D 基元图	表达图像中灰度变化位置、物体轮廓的几何分布和组织结构	滤波器的二阶过零点、端点、边缘段、边界等
2. 5D 简图	在以观察点为中心的坐标系中,表达物体可见表面的法向、深度、边界等性质	局部表面法向、表面法向不连续点、深度、深度上不连续点
3D 模型	在以物体为中心的坐标系中,用体元或面元集合,描述形状和形状的空间组织形式	以轴线为骨架,将体元或面元附在轴线上

Marr 将视觉信息处理的两大任务,划分成三个不同的层次进行描述,即计算理论、表达和算法、实现机制,其中 Marr 理论中最关键的是表达和算法层次,该层次需要精确的给出计算是如何进行的,特别应该包括信息的表达以及处理这些表达的算法。虽然 Marr 理论为计算机视觉搭建好了理论性的框架,但是其中很多技术细节超出了计算机技术的实现能力。由于缺乏人类视觉工作机理的研究,现阶段依据 Marr 理论开展 2. 1D 基元图的研究,还缺乏统一的认识和完整的体系框架,该任务中如何从视感觉形成视知觉成为计算机视觉的关键问题。

综上所述,图像理解和计算机视觉密切相关,二者的联系在于都研究具有与人类水平相当的视觉能力。区别在于研究的侧重点不同,计算机视觉主要研究从 3D 到 2D,以及从 2D 到 3D 的问题,其主要对象是数据;图像理解没有明确的 3D 问题,它通过对 2D 图像的处理和分析,得到对场景的描述和解释,图像理解的输入是数据,输出是知识,是关于对图像的理解、对图像的语义解释。

1. 1. 3 图像理解与人工智能

人类的认知过程非常复杂,人们通过对人类智力活动奥秘的探索与记忆思维机理的研究,来开发人类智力活动的潜能、探讨用各种机器模拟人类智能的途径,使人类的智能得以物化与延伸,产生了一门学科——人工智能(artificial intelligence, AI)。对所见场景的认识和理解是人类认知的重要内容之一。随着人工智能的不断发展,逐渐形成人工智能的三大学派,即符号主义(symbolicism)、连接主义(connectionism)和行为主义(behaviorism)。也有人把人工智能分成两大类:一类是符号智能(symbolic intelligence),另一类是计算智能(computing intelligence)。符号智能是