

视觉信息学教程

数字图像理解与智能技术

——基于MATLAB和VC++实现

◎主 编: 孙 明

◎副主编: 孙 红



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

视觉信息学教程

数字图像理解与智能技术

——基于 MATLAB 和 VC++ 实现

主 编 孙 明

副主编 孙 红

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书主要介绍图像理解与智能技术的基本原理和典型应用研究实例。第1章为数字图像理解与智能技术引论,第2~10章分别讨论了小波变换、图像融合、图像识别、神经网络、遗传算法、数字水印、视频处理、光谱成像、遥感图像处理基础,附录介绍了Matlab和VC++接口技术的基本知识。从第2章开始每章都包含可自由调用的Matlab程序,大部分章提供了VC++源程序。如有需要,请到华信教育资源网(www.hxedu.com.cn)下载。

本书可作为高等院校信息类学科数字图像理解与智能技术及其相关课程的研究生教材,也适合视觉信息学相关专业人员以及Matlab和VC++程序设计的中高级读者参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

数字图像理解与智能技术:基于MATLAB和VC++实现/孙明主编. —北京:电子工业出版社,2015.7
ISBN 978-7-121-26251-7

I. ①数… II. ①孙… III. ①数字图像处理—Matlab 软件 ②C 语言—程序设计 ③计算机视觉
IV. ①TN911.73 ②TP312 ③TP302.7

中国版本图书馆CIP数据核字(2015)第122874号

责任编辑:董亚峰 特约编辑:王 纲 文字编辑:吴长莘

印 刷:北京京科印刷有限公司

装 订:三河市华成印务有限公司

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开 本:787×1 092 1/16 印张:18.75 字数:456千字

版 次:2015年7月第1版

印 次:2015年7月第1次印刷

定 价:49.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888。

质量投诉请发邮件至 zits@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

编委会成员

主 编：孙 明 博士

副 主 编：孙 红 博士

其他成员：位耀光 博士

刘春红 博士

李振波 博士

安 冬 博士

前言

感谢您选择本书，为了帮助您更好地学习本书的知识，请仔细阅读下面的内容。

读者对象

本书适用于大专院校的教学和科研人员，以及采用视觉信息学进行研究或产品开发的人员。编者所提出的视觉信息学集成了多学科的综合知识。通过本书不仅可以系统全面地学习这门综合性科学和技术，也可以掌握用 Matlab 和 VC++ 等常用编程语言编制视觉信息学程序的基本技能，加深对本门学科的理解。

知识背景

20 世纪 60 年代，数字图像处理的一些算法只能在大型计算机上运行；在 70 年代，也只能在小型计算机上运行；而到了 80 年代，已能在普通的个人计算机上运行了。个人计算机的普及，使得数字图像处理技术的应用获得了爆炸式的增长，引起了来自工业、农业、军事、医学、交通、航天等不同领域的研究人员的关注。因此学习和掌握这门科学显得格外重要。我国主要高等院校已经把数字图像处理这门课程作为信息与信号处理、通信与电子系统、模式识别、机器人视觉、机电一体化、电子工程、信息工程、计算机科学与技术、遥感与军事侦察、农业工程和生物医学工程等专业和领域的本科生或者研究生课程。为了对各种与视觉信息、图像及其交叉科学相关的技术进行综合研究和集成应用，我们有必要构建一个新的统一框架——视觉信息学，它是一门与信息技术相结合，系统地研究各种图像科学的理论、技术和应用的新兴交叉学科。

本书特色

视觉信息学是一门实践性很强的科学，只有将理论与编程实践有机结合起来才能收到良好效果。但是以往的著作往往极力采用数学公式以突出理论性，过于注重概念和原理而忽略了实践环节，使得读者因不擅长公式推导等导致半途而废的情况并不少见。本书针对以往模式进行了改革，使读者在学习理论的同时，通过编程实践演练，可以透彻直观地理解和掌握数字图像理解与智能技术，既实用易懂、深入浅出，又系统全面，并凸显了该领域的最新成果等特色。例如，用图文并茂的形式和应用实例代替烦琐的数学推导过程，提供各种算法的 Matlab 及 VC++ 程序示例，并对每条语句都加以注释，使读者能通过上机实

践加深对数字图像理解与智能技术方面算法的学习和掌握,而且对提供的程序稍加改动就可以组成自己的处理软件。因此本书既有学术价值,又有实用价值。理论和实践的有机结合是本书的一大特色,可使本课程的学习变得生动有趣。另外,本书对所出现的关键词语都加上了英文对照,为读者阅读图像处理相关的英文资料提供帮助。

教师在向学生传授理论的同时,可用本书中所提供的 Matlab 和 VC++程序讲解处理算法程序,示范处理结果,使学生在实践中透彻理解图像理解的算法,一改传统的单一理论灌输的教学模式。科研人员在项目开发中可以复制和改写上述程序,达到自己的处理目标。

读者在阅读本书之前,需要了解数字图像处理与分析的相关内容。对于这部分内容,请读者参见编者所编著的《数字图像处理与分析基础——Matlab 和 VC++实现》。

本书主编孙明博士有着日本留学经历和多年的教学与科研经验,具有广泛的国际联系,能够跟踪图像理解与智能技术及其应用的国际学术前沿,在国内外学术期刊上发表了诸多相关方面的学术论文。

全书内容共 10 章,其中第 1、6 章由孙明博士编写,第 3、5、7 章由位耀光博士编写,第 4 章由安冬博士编写,第 10 章由刘春红博士编写,第 8 章由李振波博士编写,第 2、9 章由孙红博士编写,附录由李嘉伟硕士编写。全书由孙明博士和孙红博士统编。

本书的编著得到了许多人士的大力支持和帮助。首先要感谢中国农业大学信息与电气工程学院的领导和电子工程系的教授们,感谢他们所给予的热忱帮助。还要感谢中国农业大学工学院教授陈兵旗博士。主编孙明博士曾与陈兵旗博士一起于 2004 年 3 月在清华大学出版社合作出版了《Visual C++实用图像处理》一书,并分别作为主编和副主编于 2008 年 2 月在中国农业大学出版社和清华大学出版社出版了“十一五”国家级规划教材《实用图像处理与分析》,于 2004 年 2 月在中国农业大学出版社出版了农业部“十二五”规划教材《实用图像处理与分析(第 2 版)》,本书参照和借鉴了上述书籍的部分内容。本书在编写过程中,还参阅了许多学术论文和著作。

本书大部分章节后都附有完整的相关 Matlab 和 VC++程序,如有需要,请到华信教育资源网(www.hxedu.com.cn)下载。

本书得到了北京市共建项目专项资助,在此表示诚挚的感谢。

编者热切希望各位读者对本书提出宝贵意见。

编 者

2014 年 10 月于中国农业大学

sunming@cau.edu.cn

目 录

第1章 数字图像理解与智能技术引论	1
1.1 数字图像理解概述	1
1.2 图像理解的特点	4
1.3 图像理解的研究内容	4
1.3.1 目标识别	5
1.3.2 场景理解	6
1.4 图像理解的发展及其在农业中的应用	7
1.4.1 图像理解的发展	7
1.4.2 图像理解在农业中的应用	8
1.5 图像理解与智能技术的系统构成	11
1.5.1 采像	12
1.5.2 分析	26
1.5.3 理解	39
参考文献	46
第2章 小波变换	47
2.1 小波变换概述	47
2.2 小波与小波变换	48
2.3 离散小波变换	50
2.4 小波族	53
2.5 图像处理中的小波变换	57
2.5.1 基于小波变换的图像分解与重构	57
2.5.2 基于小波变换的图像增强	59
2.5.3 基于小波变换的图像平滑	61
2.5.4 基于小波变换的图像压缩	62
2.5.5 基于小波变换的图像合成	64
参考文献	66

第3章 图像融合	67
3.1 图像融合概述	67
3.1.1 图像融合的应用	68
3.1.2 图像融合的方法	69
3.2 图像融合的关键技术	70
3.3 典型图像融合算法	71
3.4 塔式分解法	76
3.4.1 非采样 Contourlet 算法的特点	77
3.4.2 非采样 Contourlet 变换原理	78
3.4.3 图像融合规则	81
3.5 基于视觉特性的图像融合	86
3.5.1 人眼视觉特性的相关概念	86
3.5.2 基于视觉识别特性的图像融合算法原理	88
3.5.3 多聚焦图像融合算法	89
3.6 图像融合性能评价	92
3.6.1 图像融合方法的基本要求	92
3.6.2 图像融合效果评定方法	93
3.7 应用研究实例	95
参考文献	95
第4章 图像识别	97
4.1 图像识别概述	97
4.2 模板匹配方法	99
4.3 统计模式识别	100
4.4 支持向量机	103
4.5 仿生模式识别	106
4.5.1 仿生模式识别理论	106
4.5.2 高维空间几何理论	107
4.5.3 基于仿生模式识别的人脸识别	109
4.6 应用研究实例	114
参考文献	114
第5章 神经网络	115
5.1 神经网络概述	115
5.2 生物学基础	116
5.3 人工神经元	117
5.4 神经网络的特点	119
5.5 BP 神经网络	120

5.5.1 BP 神经网络概述	120
5.5.2 BP 神经网络的训练学习	121
5.5.3 BP 神经网络的 Matlab 示例	124
5.6 应用研究实例	126
参考文献	127
第 6 章 遗传算法	128
6.1 遗传算法概述	128
6.2 生物学基础	130
6.3 简单遗传算法	131
6.3.1 遗传表达	131
6.3.2 遗传算子	132
6.4 遗传参数	134
6.4.1 交叉概率 P_c 和变异概率 P_m	134
6.4.2 其他参数	135
6.4.3 遗传参数的确定	135
6.5 适应度函数	135
6.5.1 目标函数映射为适应度函数	136
6.5.2 适应度函数的尺度变换	136
6.5.3 适应度函数设计对 GA 的影响	137
6.6 模式定理	138
6.6.1 模式的几何解释	140
6.6.2 GA 的操作对模式的影响	141
6.7 遗传算法在模板匹配中的应用	143
6.7.1 问题的设定	143
6.7.2 GA 的应用方法	145
6.7.3 简单 GA 的 Matlab 程序示例	146
6.7.4 程序执行过程	164
6.8 应用研究实例	166
参考文献	169
第 7 章 数字水印	170
7.1 数字水印概述	170
7.2 基于 DCT 域的鲁棒水印	172
7.3 基于空间域的脆弱水印	179
7.4 基于 DWT 域的脆弱水印	185
7.5 应用研究实例	192
参考文献	193

第 8 章 视频处理	194
8.1 视频处理概述	194
8.2 视频生成与成像模型	194
8.2.1 视频生成简介	194
8.2.2 视频成像模型	195
8.3 视频编码	199
8.3.1 视频编码概述	199
8.3.2 视频编码标准	199
8.4 运动目标跟踪	202
8.4.1 运动目标检测算法	203
8.4.2 运动目标跟踪算法	204
8.5 应用研究实例	213
8.5.1 视频跟踪在昆虫运动分析和行为识别中的应用	213
8.5.2 机器视觉在草地蝗虫识别中的应用	215
8.5.3 远程农作物视频采集系统	216
参考文献	217
第 9 章 光谱成像	218
9.1 光谱成像概述	218
9.1.1 光谱成像原理	218
9.1.2 光谱成像方式	220
9.1.3 光谱成像数据及描述模型	224
9.1.4 光谱成像检测技术的发展趋势	228
9.2 光谱图像处理	230
9.2.1 光谱图像预处理	230
9.2.2 光谱图像特征的选取与分类方法	232
9.2.3 光谱图像分类中的模式识别方法	234
9.3 应用研究实例	244
9.3.1 利用多时相 Landsat 高光谱图像监测冬小麦和苜蓿种植面积	244
9.3.2 基于光谱图像的作物长势监测	247
参考文献	249
第 10 章 遥感图像处理基础	250
10.1 遥感图像处理概述	250
10.2 遥感图像目视解译与判读	251
10.2.1 判读要素	251
10.2.2 判读标志	252
10.2.3 判读方法	252

10.2.4 遥感判读的基本技术	253
10.3 遥感图像的数字表达	254
10.4 遥感图像的存储	254
10.4.1 HDF	255
10.4.2 BSQ	258
10.4.3 BIP	258
10.4.4 BIL	258
10.5 遥感图像特征的统计分析	258
10.5.1 遥感图像的特征	258
10.5.2 成像方式分类	259
10.6 常用遥感图像处理软件	260
10.6.1 ENVI	260
10.6.2 ERDAS IMAGINE	261
10.6.3 Titan Image	262
10.7 高光谱遥感	264
10.7.1 高光谱遥感的特点	265
10.7.2 高光谱遥感的优势	265
10.7.3 高光谱遥感的应用领域	265
10.8 应用研究实例	267
10.8.1 正交子空间投影目标探测法	267
10.8.2 实验结果与分析	270
参考文献	273
附录 Matlab 和 VC++接口技术	274
1. Matlab 和 VC++接口技术概述	274
2. Matlab 和 VC++常用接口技术比较	274
3. MATCOM C++数学库的使用	277
4. 程序编写过程示例	279

第1章

数字图像理解与智能技术引论

1.1 数字图像理解概述

人类通过眼、耳、鼻、舌等感官接受信息，其中约 83%的信息来自我们的眼睛，即视觉。民间谚语“百闻不如一见”，是说费尽口舌还不如一幅画一目了然；单凭名字不能回忆起某人时，不妨翻看一下他/她的照片；仔细阅读某种机器设备的使用说明书之前，不妨先浏览一下其说明图表等，这些都说明了视觉信息的重要性。而在心理学研究领域中，研究人员发现人类获取味觉和听觉信息也与视觉信息紧密相关。可见，视觉是人类获取信息的最重要途径。因此，研究视觉信息的获取、处理等相关学科至关重要。

目前，视觉信息研究蓬勃发展，与之相关的研究学科日益增多，形成了相对完整的研究体系。以下对相关学科的概念和术语进行简单介绍。

1. 信息学 (informatics)

信息学是研究信息的获取、处理、传递和利用的规律性的一门新兴学科。信息学是以信息为研究对象，以计算机等技术为研究工具，以扩展人类的信息功能为主要目标的一门综合性学科。

2. 图像科学 (image science)

图像是人类视觉的基础。图像科学在广义上是与各种图像有关的科学的总称。目前人们研究的是数字图像，主要应用的是计算机图像处理技术。这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，如图像的采集和获取，图像的编码、存储和传输，图像的合成和产生，图像的显示和输出，图像的变换、增强、恢复（复原）和重建，图像的分割，图像目标的检测、表达和描述，图像特征的提取和测量，序列图像的校正，3D 景物的重建复原，图像数据库的建立、索引和抽取，图像的分类、表示和识别，图像模型的建立和匹

配, 图像和场景的解释和理解以及基于它们的判断决策和行为规划等。另外, 图像科学还包括为完成上述功能而进行的软、硬件设计及制作等方面的技术。

3. 数字图像 (digital image)

把连续的图像在坐标空间和性质空间都进行离散化, 变成供计算机进行处理的图像, 称为数字图像, 简称图像。图像中每个基本单元叫做图像元素, 简称像素 (pixel)。

4. 视觉信息学 (visual informatics)

视觉信息学是 20 世纪 90 年代初被提出的一个学科概念, 并于 2009 年 11 月在马来西亚首都吉隆坡召开了第一届视觉信息学会议^[1]。视觉信息学不仅包括信息学和计算机科学, 而且包括工程信息学、虚拟技术与系统、医学和健康信息学及其教育等其他相关领域。它综合运用数学、光学、计算机科学等基本原理及各种工具, 对图像科学中的视觉信息进行有效提取、存储、传递、处理、分析和解释, 它是将图像科学发展过程中出现的各种新理论、新方法、新算法、新手段、新设备等进行综合研究和集成应用的一个新的统一框架。最近, 它已成为计算机视觉、信息可视化、实时图像处理、医学图像处理、虚拟现实、增强现实、表达视觉信息学、3D 图形学、多媒体融合、视觉数据挖掘、视觉本体、视觉文化与服务等多学科交叉的领域。视觉信息学的内容非常丰富, 覆盖范围广, 根据抽象程度和研究方法等的不同, 主要可分为三个层次, 即图像处理、图像分析和图像理解。

5. 图像处理 (image processing)

图像处理着重强调在图像之间进行的变换。输入和输出的都是图像, 包括图像采集和获取、图像增强、图像平滑、图像锐化、频率变换、几何变换、图像编码和图像复原等。

6. 图像分析 (image analysis)

图像分析指主要针对图像中感兴趣的目标进行检测和测量, 以获得它们的客观信息, 从而建立对图像的描述。输入的是图像, 输出的是特征。包括图像分割、边缘检测、形态学图像处理、特征选择和纹理分析等。

7. 图像理解 (image understanding)

图像理解的重点是在图像分析的基础上进一步研究图像中各个目标的性质和它们之间的相互联系, 并得出对图像内容含义的理解以及对原来客观场景的解释, 从而指导和规划行动。输入的是数据, 输出的是知识, 可以看成利用计算机实现人的视觉功能, 以获取对客观世界中二维场景的感知、识别和解释的知识。包括图像识别、图像融合, 以及利用小波变换、神经网络、遗传算法、数字水印、光谱分析、X 射线、视频、遥感等新理论、新工具、新技术的视觉认知、知识表达和更深层理解。

在上述研究学科和体系中, 本书重点阐述图像理解与智能技术的要点知识, 介绍图像理解基础、图像理解研究的主要内容、图像理解技术分支及应用的实现方法。

对数字图像理解 (简称图像理解) 的研究始于 20 世纪 50 年代初, 研究初期以计算机视觉为载体。它是一门多学科交叉的综合学科, 其研究内容十分丰富, 与图像处理与分析、计算机视觉、智能技术、认知学有着密切的联系, 虽然它们侧重于不同的研究内容、采用不同的研究方法, 但是这些学科都是图像理解研究的基础和依据^[2]。

计算机视觉(又称机器视觉)简单地说,就是研究用计算机来模拟人类(或灵长类动物)视觉的一门科学,由图像数据来产生视野环境内有用符号描述的过程,强调用计算机实现人类(或灵长类动物)的视觉功能,对客观世界的三维场景进行感知、识别和解释。图像数据包括静态图像和动态图像(视频),同时又包括二维图像和立体图像。计算机视觉强调利用计算机实现人类的视觉功能,主要研究三维到二维以及二维到三维的问题,其主要对象是数据;而图像理解则通过对图像的处理和分析,得到对场景的描述和语义解释,其输入是数据,而输出是知识。

智能技术以“知识”为核心,通过模式识别、机器学习等方法解决知识的表达、推理和学习的问题;而图像理解关注的是图像内容,如场景中的目标以及场景的描述等。因此,图像理解与智能技术密切相关。简单地说智能技术就是研究用计算机模拟人脑,模拟人类智能,以期有效地达到某种预期的目的。传统的智能技术以人类的知识为对象,在已有知识系统的基础上对知识结构关系、语义网络、通用匹配、推断决策、产生式系统、问题求解、规划控制反馈和经验学习进行研究,其实现过程表现为研究知识的表示、存储、推理以及获取新的知识。图像理解既是对计算机视觉研究的延伸和拓展,又是智能技术的研究新领域,渗透并贯穿于智能技术的研究进程。从智能技术角度出发研究图像理解问题的关键在于通过知识与图像数据之间的转换解释场景。

图像理解作为人类知识获取的重要来源,需要人类视觉生理学和认知心理学作为其理论支持,图像理解和认知学研究的共同目的是实现人类视觉认知的能力。认知学为图像理解提供人类视觉的基本原理和认知过程,但由于认知学研究方法和结论的局限性,认知学仅仅作作为图像理解的外延存在并发展着。由此,图像理解是一门研究图像中目标和场景的自然科学,任何人类知识和认知过程必须在计算机中实现,才能借助于计算机的高效性和便捷性来实现图像理解任务。

综上所述,图像理解与计算机视觉的联系在于研究具有与人类水平相当的视觉能力。图像理解和计算机视觉的重要研究目标之一就是将该研究作为探索和认识人脑视觉工作机理的手段,以进一步深入掌握和理解人脑视觉机能。反过来,对人脑视觉认知过程的充分理解也促进了图像理解和计算机视觉的深入研究和发[3]。因此,为了有效地建立图像理解系统,必须充分研究人的视觉系统以及人利用知识进行图像理解的过程。

由图像理解的层次结构可以看出,图像理解过程是通过图像中的低层操作、中层操作和高层操作的大量数据信息与知识信息相互交织,共同作用,形成具有并行性、层次性以及反馈特性的动力学系统[2]。事实上,图像理解系统采用的层次结构以及其中的许多重要原理和方法与视觉感知系统的结构与功能有着紧密的联系,它与视觉感知系统所采用的层次化感知机理是高度一致的,因此对视觉通道结构和功能的理解及其数学模型的建立是扩充和发展现有图像理解方法的一种重要手段和根本出发点,真正从高等生物视觉系统的神经生理属性出发所建立的图像理解模型更能体现图像理解方法的本质,对其进行研究具有重要的理论意义和应用价值。

随着神经生理学、认知学、计算神经学等学科的蓬勃发展,大量生理解剖学和心理学的事实为人们深入理解生物视觉系统提供了丰富的实验基础[4]。研究表明,生物大脑中的视觉信息处理系统是由庞大的初级系统、次级系统以及许多更高级系统按照层次结构构成

的,各个系统内部及其之间通过大量神经元的突触连接方式进行交互作用,形成对视觉目标和场景的有效描述。低层的初级系统从图像中抽取点、线、面等基本数据特征,实现图像处理的过程;中层的次级系统对初级系统处理的结果进行简化抽象,以形成可以与目标表示相匹配的结构形式;高层的高级系统则根据低层和中层的处理结果,利用视皮层的反馈推理机制产生并试验可供解释的有关假设,形成对目标和场景的二维或三维结构的复杂描述。视觉感知系统实际上就构成了一个完整的图像理解系统,两者之间的层次结构和功能关系具有较好的对应性。因此,一方面需要从视觉生理学和心理学的角度理解视觉系统的机能;另一方面则要借助于数学建模和计算机仿真的方法,建立能够实现以上机能的图像理解系统。

综上所述,图像理解以图像为对象,以知识为核心,研究图像中有什么、它们之间的相互关系、图像是什么场景以及如何应用场景的一门学科。图像理解根植于计算机视觉,并在其理论基础上不断发展完善,它通过融合智能技术和认知学领域的知识,形成了与计算机视觉、智能技术和认知学密切关联的学科领域,特别是与计算机视觉在很多研究内容上交叉重合。事实上,这两个术语也经常混合使用,使得它们在概念上和实用上都没有绝对的界限,有时可认为它们虽各有侧重,但互为补充。在很多场合,将它们看成不同专业背景的人习惯使用不同的术语。本书对它们也不做刻意区分。

1.2 图像理解的特点

人或其他高等生物的视觉过程要解决的两个最基本的问题是“在视场中有什么及其在哪儿(what-where)”,其复杂的信息过程包括了感知和信息处理两部分。图像理解也有类似的两部分,即成像和信息处理及分析,因此贯穿图像理解的过程具有以下几个鲜明的特点。

- ① 层次性:分阶段信息处理带来了信息的多层表示。
- ② 描述性:对图像的解释以某种形式的描述实现。
- ③ 知识性:图像的正确解释离不开知识的引导。

具体而言,图像理解具有鲜明的层次性,具有低层、中层和高层描述的层状结构。二维像素点阵图像通常作为低层的输入,低层输出通常是以像素为单位检测出的图像特征;中层输出是在低层描述的编组、抽象后形成的符号描述,减少了数据量,提高了描述质量,更接近图像的本质;高层理解主要通过以中层符号描述为基本单元的、反映景物与目标特性的模型和服务于解释的知识库,完成解释图像的任务。

1.3 图像理解的研究内容

从研究的广泛性看,图像理解的处理信息分为视觉数据信息和人类知识信息两部分,前者侧重于原始获取的数据信息以何种结构存储在计算机中,后者侧重于知识的表述如何指导计算机的理解过程,两部分表示相辅相成。图像理解中对视觉信息和知识信息的研究

过程就是进行信息表示、处理和分析的过程。图 1-1 显示了基于图像理解实现对输入图像中的目标识别及场景理解的示例。因而，图像理解的研究内容主要包括两个方面：一是目标识别，二是场景理解。

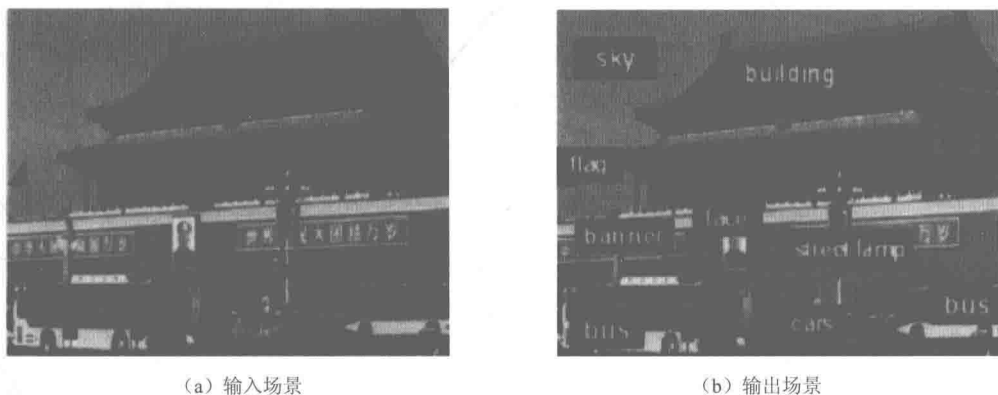


图 1-1 图像输入与基于图像理解的部分输出^[2]

1.3.1 目标识别

图像理解的基本任务之一是让计算机识别判断场景中有什么物体及其在哪儿，解决“what-where”问题。通常，场景中的目标可视为具有较高显著度并符合局部感知一致性的区域，该过程也是计算机对场景中的物体进行特征分析和概念理解的过程。目标识别的过程主要包括目标证实（verification）、目标检测（detection）、目标鉴别（identification）及目标识别（recognition）。其中，目标证实是判断图像中指定区域是否为目标，目标检测是定位目标的准确位置，目标鉴别是对目标类别进行实例细分，目标识别是对图像中的目标区域进行识别。

人能够快速识别出将近一万种目标物体^[5]且不受各种客观条件的影响，具有极佳的鲁棒性，对新目标的识别仅需很少的信息指导和实例分析；对计算机而言，虽然低廉的数字图像存储设备能将客观的目标进行有效的存储，但对其进行分析却非常困难，目标的表示仅仅基于像素级，缺乏人处理知识特有的高层表示和分析，对目标识别的研究涉及目标变化的若干方面，主要包括以下几点。

① 多方位变化（multi-aspect variation）：不同场景中目标的多方位变化。

② 多视角变化（multi-viewpoint variation）：同一目标物体在不同视角下的变化，包括目标的平面内变换（平移、平面旋转、放缩、倾斜等）和平面外变换（深度旋转等）。

③ 光照（illumination）：目标在不同的光照环境下呈现出不同的强度变化，像素强度与光源强度呈线性变化，与光源位置呈非线性变化，光照产生的阴影区域给目标识别带来了很大的困难。

④ 背景干扰（background clutter）：实际场景中复杂背景的干扰涉及识别方法的实用性和通用性问题。

⑤ 遮挡（occlusion）：目标中的部分区域被其他目标或目标的其他区域遮挡引起的识别困难。

⑥类内变化 (intra-class variation): 目标几何形状的变化和内部区域的变化都会极大影响同一类别内的目标识别的判断结果。

目前对前三种变化的研究较为深入, 主要包括对汽车^[6]、数字^[7]、人脸^[8,9]及指纹^[10]等有限类别生物特征的识别。现有方法已经可以有效地降低甚至消除这些变化对识别造成的影响。而图像理解的基本任务就是在真实世界的环境下对广泛意义上的目标进行识别和检测, 虽然各种目标千差万别, 但极可能属于同一认知概念的划分。因此, 后三种情况出现的问题是对图像理解中目标识别研究的极大挑战。

目标识别通常包含学习和分类两个阶段, 训练样本通过提取特征构建特征库, 再通过概率推理模型或分类器模型进行学习, 为待识别目标提供识别决策, 最后完成识别任务, 图 1-2 显示了图像理解中目标识别的基本过程。识别结果的正确与否与从图像中得到的特征数据及其融合知识的分类建模密切相关。虽然从视觉信息学的角度看, 特征选择仅属于中层操作的图像分析, 但由于图像理解目标识别的广义特性, 使得特征选择产生了新的方向, 即类内共性分析, 因此特征选择可作为图像理解的低层数据获取阶段, 也属于图像理解的研究范畴^[11]。同时从图 1-2 可以看出, 目标识别中的模型或分类方法是对输入数据信息的决策判断, 模型或方法的判断依据为目标识别提供了必要的知识信息, 既需要关于待识目标的特殊信息, 也需要关于目标从属类别的较高层次的一般知识。

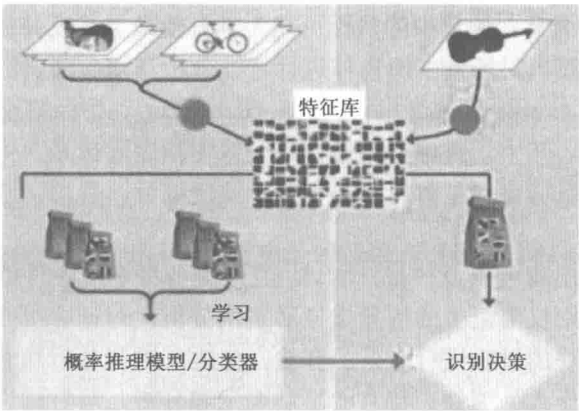


图 1-2 目标识别的基本过程^[2]

1.3.2 场景理解

场景理解研究主要包括两个内容: 一是集中于视觉心理学和生理学方面, 主要研究快速场景感知的心理和生理机理; 二是场景分类的计算建模, 通过建立简单的统计计算模型实现几类场景 (室内、室外等) 的判断或场景中的空间语义分析。对图像理解中场景分类的研究刚刚起步, 2006 年在 MIT 首次召开了场景理解研讨会 (Scene Understanding Symposium, SUnS), 明确了场景分类将会是图像理解一个新的研究热点。

本书不讨论这方面内容。