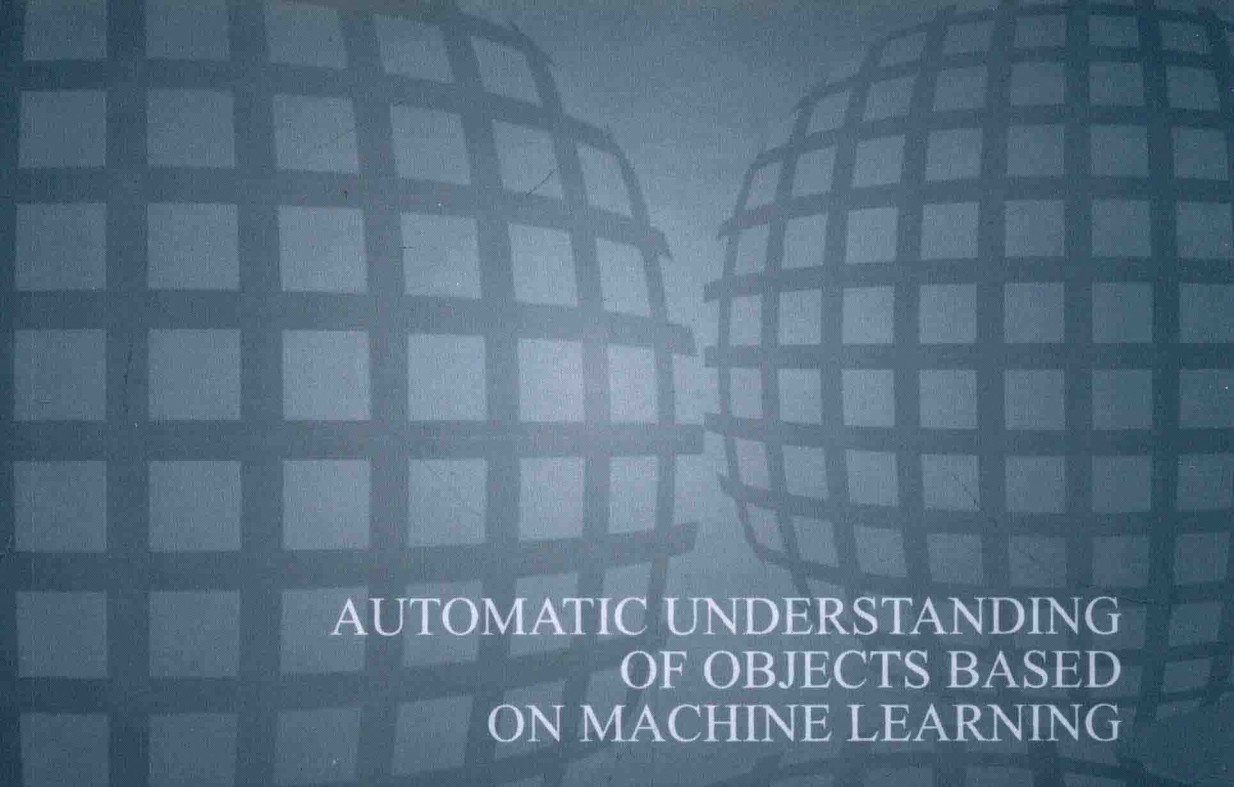




AUTOMATIC UNDERSTANDING
OF OBJECTS BASED
ON MACHINE LEARNING

基于机器学习的 物体自动理解技术

■ 刘贞报 著



科学出版社

基于机器学习的物体自动 理解技术

刘贞报 著

科学出版社

北京

内 容 简 介

本书用丰富的图示和实验,将物体的自动理解技术和机器学习的理论相结合,以实现物体的自动理解技术为主线,以机器学习的理论作为主要方法,结合实例逐步深入地介绍了机器学习相关理论。主要内容包括特征提取、三维网格分割、三维场景重建、三维模型功能性分析等,涵盖了目前常用的主流的各种学习方法。

本书结构紧凑,内容逐步深入,同时包含大量实验进行讲解说明。适合相关领域的本科生、研究生以及工程技术人员阅读。

图书在版编目(CIP)数据

基于机器学习的物体自动理解技术/刘贞报著. —北京:科学出版社,2016.12
ISBN 978-7-03-051073-0

I. ①基… II. ①刘… III. ①机器学习—研究 IV. ①TP181

中国版本图书馆 CIP 数据核字 (2016) 第 307414 号

责任编辑:赵敬伟/责任校对:钟 洋

责任印制:张 伟/封面设计:耕者工作室

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>

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016年12月第 一 版 开本:720×1000 B5

2016年12月第一次印刷 印张:15 3/8

字数:300 000

定价:98.00 元

(如有印装质量问题,我社负责调换)

前 言

物体自动理解是室内机器人、无人驾驶汽车、无人飞行器面临的一项复杂课题。理解的任务在于,利用搭载的深度传感器、多视角相机、机载雷达等,主动识别场景中的物体,认识物体与物体之间的结构关系,明确场景中人作为活动对象与物体之间的交互作用。理解的实现方式为,将室内场景、外部道路周边环境、空间环境自动转换为物体标注,解析物体关系,建立人与物体的交互关系,重建三维环境。正确的场景感知可以实现自主认知和适应环境,有利于室内机器人室内定位和移动路径规划;有利于无人驾驶汽车实现在复杂道路上的安全自动驾驶;有利于无人飞行器实现空中搜救与环境监测等无人值守的全自动任务。

机器学习也是近年来的研究热点,在机器学习领域,各种新型的算法层出不穷。本书将物体自动理解与机器学习相结合,从多个方面对基于机器学习的物体自动理解技术进行了详细的介绍。

本书由以下几部分构成:

第一部分叙述了物体自动理解技术的概念和研究意义,详细阐述了本书的研究价值。

第二部分介绍了多种前沿的、有效的特征提取方法。针对图像的场景解析,本书提出了一种全新的深度学习框架,能够有效地进行结构化学习。设计了两套构建高层结构化信息的三维信息学习框架:基于环特征的学习框架和基于词包编码的特征学习框架。同时,提出了基于深度置信网络的三维形状高层特征提取方法,成功将深度学习方法应用于三维形状数据的理解中。

第三部分在分析研究主流三维网格分割算法的基础上,提出了两个新的基于多标准的三维网格评价度量手段,提供了一种更符合人类认知的评价方法,解决了网格分割标准不唯一的问题。同时,提出了一种基于凹陷区域探测和启发式快速行进二分类方法的三维模型自动分割算法,利用该方法提取三维模型的标准姿态,实现了对非刚性三维模型的检索。

第四部分在三维物体识别的基础上,进行了三维场景的重建。与传统重建技术有着本质区别,本书提出的三维重建方法不是简单地将深度相机获得的点云数据进行配准得到场景重建,而是对场景中的所有物体进行识别并与数据库中的模型进行匹配,实现了三维场景的语义重建。本书利用随机回归森林法从数据库中匹配出物体的最相似三维模型,并根据深度信息变换到场景中,得到单幅图像的模型重建,利用改进的迭代最近点方法进行数据配准,得到多幅图像之间的坐标变换,从

而将单幅图像的重建结果融合到一个完整的场景。

第五部分对三维模型的功能性分析技术进行了介绍。功能性检测是最近几年才被提出来的三维模型理解方法，本书将以基于人体骨架的三维模型功能性分析为主，对该类方法进行详细的论述。通过将特定的模型与待检测的物体模型互相交互来判断某物体是否具有某项功能，进而找到该物体的功能位置；同时，采用基于舒适性的姿态调整方法生成人体模型，通过舒适程度的大小来判断当前人体模型的姿态合理性。

目 录

前言

第一部分 绪 论

第 1 章 什么是物体自动理解技术	3
1.1 物体自动理解技术概念	3
1.2 研究意义	3

第二部分 特 征 提 取

第 2 章 三维模型概述	7
2.1 三维模型的起源与发展	7
2.2 三维模型的获取与存储	8
2.3 国内外三维模型特征提取的研究现状	9
第 3 章 基于二维图像的特征提取	11
3.1 深度学习的出现和发展	11
3.2 深度学习原理概述	14
3.3 基本的前馈神经网络模型	18
3.4 深度学习模型训练中的优化方法和常用技巧	22
3.5 内嵌推理的深度网络框架	25
3.6 特征学习层	26
3.7 结构学习层	33
3.8 特征融合层	35
3.9 实验分析	36
第 4 章 基于深度学习的三维模型特征提取	41
4.1 三维模型的底层特征提取	41
4.2 模型的中层特征提取	45
4.3 三维模型的高层特征提取	51
4.4 实验验证与分析	57
第 5 章 基于深度置信网络的三维模型特性提取	70
5.1 相关算法理论	70

5.2	算法框架	74
5.3	数据预处理和视角图像的生成	76
5.4	特征提取	77
5.5	实验验证与结果分析	79
5.6	小结	84

第三部分 三维网格分割与模型检索

第 6 章	视觉分割技术概述	89
6.1	分割背景和意义	89
6.2	图形分割的国内外研究现状	90
第 7 章	二维图像分割算法研究	93
7.1	颜色空间及分析	93
7.2	纹理特征及分析	94
7.3	常用彩色图像分割方法	95
7.4	小结	99
第 8 章	三维网格分割研究	100
8.1	网格分割概述	101
8.2	基于有监督学习的三维模型分割	102
8.3	基于谱嵌入的三维模型共分割	106
8.4	基于用户交互的快捷分割	109
8.5	基于凸度估计和快速行进方法的三维模型自动分割算法	111
第 9 章	多标准三维物体分割评价	122
9.1	相似性汉明距离	122
9.2	熵增	127
9.3	自适应熵增	130
9.4	评价实验	131
9.5	其他评价方法	138
第 10 章	非刚性单位模型检索应用	141
10.1	三维形状分类方法简介	141
10.2	形状检索定义	141
10.3	三维形状检索研究现状	141
10.4	常用检索评价方法	143
10.5	非刚性三维模型检索	143

第四部分 三维场景重建

第 11 章 基于三维场景分割的三维重建	155
11.1 图像预处理	155
11.2 基于边缘检测的 RGBD 图像过分割	160
11.3 地面和墙的测定及对分割结果的处理	164
11.4 基于随机回归森林的场景区域识别	165
11.5 目标提取	168
11.6 场景更新	169
11.7 实验结果	173
11.8 小结	174

第五部分 三维模型功能性分析

第 12 章 功能性分析相关理论	177
12.1 功能性分析的引入	177
12.2 基于舒适性的人体骨架模型姿态设计	179
12.3 研究现状	180
第 13 章 基于人体骨架的三维模型功能性分析	184
13.1 三维模型功能性检测及应用	184
13.2 基于舒适性的人体模型生成	198
13.3 功能性分析实验验证	202
13.4 相关讨论	222
参考文献	226

第一部分 绪 论

第1章 什么是物体自动理解技术

1.1 物体自动理解技术概念

本书介绍了一些前沿的基于机器学习的物体自动理解技术。所谓物体自动理解,广义上说,即在任意环境下识别任意物体。从计算机的角度来说,就是建立在对图像理解的基础上,通过对获取的图像数据进行处理和分析,让计算机能自动地把图像数据中的物体进行分类。物体自动理解是一个颇具挑战性的问题,是智能化信息处理的关键技术,相关研究成果已经应用于不同的领域,主要包括图像检索、视频监控、智能家居、机器人领域的目标寻找、避障等,具有重要的理论研究意义和工程应用价值。

理解的任务在于利用搭载的深度传感器、多视角相机、机载雷达等,主动识别场景中的物体,认识物体与物体之间的结构关系,明确场景中人作为活动对象与物体之间的交互作用。理解的实现方式为,将室内场景、外部道路周边环境、空间环境自动转换为物体标注,解析物体类别和关系,建立人与物体的交互关系,重建三维环境。正确的场景物体感知,可以实现自主认知和适应环境,有利于室内机器人室内定位和移动路径规划,有利于无人驾驶汽车实现复杂道路上的安全自动驾驶,有利于无人飞行器实现空中搜救与环境监测等无人值守的全自动任务。因此,根据不同应用需求,目前场景物体理解的工作可以划分为两类,室外和室内。室内场景与室外环境截然不同,效果无法移植,麻省理工学院 Quattoni 教授曾在研究中指出:许多物体识别模型在室外场景表现良好,但并不适用于室内^[1]。相对于室外场景,室内场景通常包含更多照明变化、杂乱布置、遮挡、大小不同的物体。另外,室外场景元素(如地面、天空、大海、绿地等)无法变化外观,然而室内场景物体可变性较强。

本书将围绕室内复杂环境中的物体自动理解技术,以三维模型作为切入点,通过三维模型的特征提取、三维网格语义分割、三维模型的检索、三维场景的重建以及三维模型功能性检测等多方面对该技术进行了详细的论述。

1.2 研究意义

随着智能时代的到来,机器人、无人机、自动驾驶汽车、智能手机等便携设备

在民用和军事领域的应用越来越广泛,通过这些智能设备对环境做出环境感知,尤其是场景的解析,能够帮助其实现更高层次的功能。例如,实现无人机或者机器人的自主运行,自动寻着目标等;对于便携式设备可以实现智能导航、虚拟现实、智能战场等。如何利用这些设备对周围环境做出感知和认识是目前的研究热点之一,而这一切都源于场景识别、场景解析、物体识别、地图构建和定位 (SLAM)、场景的三维重建等技术的成熟。

在民用领域,自动驾驶汽车和机器人是主要的应用对象,如图 1-1 所示,目前美国的谷歌公司,德国的奔驰、宝马等知名公司已基本实现汽车的自动驾驶,但其缺点是使用较为复杂的传感器和计算设备,导致系统价格高、全天候运行能力差等。为克服这些问题,目前主要的研究方向是如何减少传感器的使用量,并提高其对环境的感知能力。例如,高精度、高实时性的行人、车辆、红绿灯、指示牌等的识别、判断等。预期到 2020 年左右,自动驾驶系统将成熟并实现商业化运作。

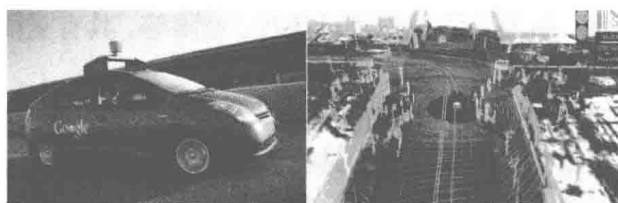


图 1-1 谷歌公司的自动驾驶汽车和其场景感知的示意图 (阅读彩图请扫封底二维码)

对于军事领域,应用较多的一类智能设备是无人机 (unmanned aerial vehicle, UAV),如图 1-2 所示。目前无人机系统的智能程度和自主水平还比较低,无人机的控制方式主要以操作员遥控和预编程控制为主,这种控制方式无法应对高度不确定的环境变化,不能有效处理各种突发威胁。针对目前的技术水平,美国国防部《2009—2034 财年无人系统综合路线图》规划指出:无人机系统自主能力和鲁棒性的提高,能够改进对战场的感知,提高目标定位的速度和精度,增强生命力,扩大任务的灵活性,计划到 2015 年无人机系统将实现感知-规避能力,到 2034 年实现在线态势感知,具有完全自主能力^[2]。因此场景的感知已成为目前学术界和工业界比较重要的一个研究方向。



(a) 美国全球鹰无人机



(b) 美军的微型侦查无人机



(c) 本研究室制作的四旋翼微小型无人机

图 1-2 三种无人机

第二部分 特征提取

第2章 三维模型概述

本书主要论述基于消费级深度传感器的室内场景物体自动理解技术，即物体自动理解是以三维数据作为数据源的。因此，我们首先对其数据源进行简单的介绍。

2.1 三维模型的起源与发展

随着科技水平的不断提高，现有的数据类型已经不能满足人们的需要，三维模型的出现，极大地丰富了可供选择的数据形式。三维模型能够保留物体丰富的曲面、颜色和纹理等信息，目前已经被广泛应用于图形领域、虚拟现实、多媒体、计算机视觉、娱乐、设计和制造等。

在建筑设计领域，通常会在施工之前，用三维软件仿制出建筑的构造图，进行受力分析和外观修正，确保建筑设计的合理和美观；在考古领域，三维模型技术被用来对损坏的文物进行修补；在娱乐方面，3D 电影早已进入普通人的日常生活中，带来了前所未有的观影感受，而 3D 游戏的产生，可以给玩家带来更加绚丽的操作体验。三维模型诸如此类的应用不胜枚举，由此看来，三维模型已经完全融入到我们的生活中，成为不可或缺的一部分。

三维模型领域发展最迅速的当属 3D 打印。3D 打印技术出现很早，最早能够追溯至 20 世纪末期，这项技术的基本原理是将粉末状的材料作为打印对象，按照分层打印的方式来生成物体。最初的 3D 打印技术只能用于一些简单模型或模具的制造，图 2-1(a) 中展示了一个 3D 打印的汽车模型外壳。但经过最近几年的发展，3D 打印技术有了很大的提高，可以进行许多复杂产品的生产制造，图 2-1(b) 中

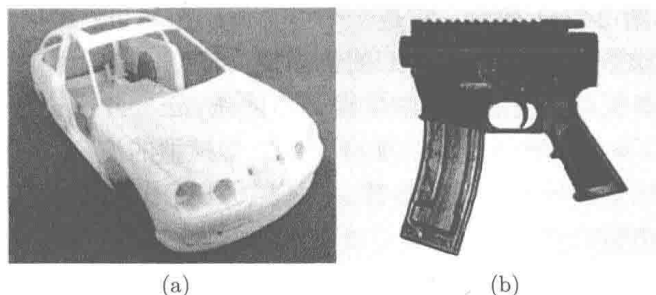


图 2-1 3D 打印产品

展示了一款由 3D 打印技术制造的机枪。3D 打印技术的进一步发展,会使得三维模型的制造变得非常容易。

在三维模型广泛应用的背后,是与三维模型相关的许多技术的突破性进展,越来越多的三维模型被更加容易地设计和制造出来。

2.2 三维模型的获取与存储

目前,人们获得三维模型的方法越来越多样化,许多公开的模式库,如 Google 3D warehouse 等三维模型数据库,包含的模式也越来越丰富。另外,三维模型扫描技术已经愈加成熟,所使用的设备也愈加易于上手使用。还有各种三维模型的设计软件,其功能也变得愈加强大,使用界面也愈加简单易用。

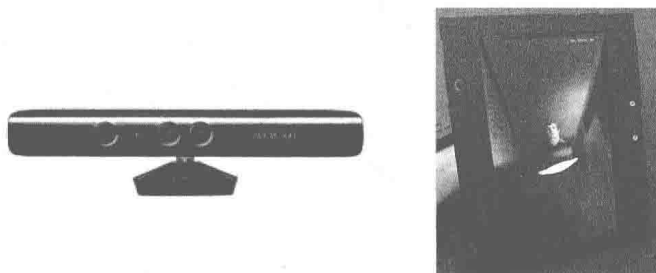
2.2.1 常见三维模型获取方式

扫描式和软件生成是目前应用较为广泛的两种三维模型获取方式。

常见的三维模型扫描设备有接触式与非接触式之分,所谓接触式测量即通过物理上的触碰来获得物体的形状信息,该扫描方法的精度极高,对于物体的某些复杂区域可以重复探测,且不受被扫描物体的颜色和外界环境的影响,但是该扫描方法的探测过程极为复杂,硬件要求较高,还有可能损坏物体,而且应用范围受到了很大限制,多为工业应用,人们常用的扫描对象(包括人体)都无法使用该方法进行扫描。非接触式测量是近年来随着多媒体处理技术的发展而出现的一类技术,这类技术利用电磁波或者声波的反射原理判断扫描探头与被扫描物体之间的距离关系,不再需要物理上的接触,然后利用三维模型的对齐技术或者陀螺仪定位系统计算扫描探头的位置变动情况,进而通过特定的算法将获得的各个部分融合成一个整体,根据扫描探头使用的原理不同,可分为光学测量、电磁测量和声学测量等几大方向。非接触式扫描技术与接触式扫描技术相比虽然扫描精度较低,但是简单易用得更多,一些接触式扫描难以处理的物体使用该方法可以较好地扫描。在具体的扫描过程中,选取的方法应视情况而定。2009 年微软发布了名为 Kinect 的首款 RGBD 设备,如图 2-2(a) 所示,它是一个 3D 体感摄影机,可以通过多种简单方便的方式得到三维数据,由于 Kinect 的价格较为低廉,在实际中可以进行大量的应用。德州大学奥斯汀分校的几个学生设计出的 LynxAcamera 立体照相机,如图 2-2(b) 所示,可以像操作普通照相机那样完成三维模型的渲染,这些捕获到的三维模型可直接在 3D 建模软件中进行操作。这些 RGBD 设备的广泛应用,使得三维数据的获取更加方便和简单。

随着计算机多媒体技术的快速发展,各种用于绘制和处理三维模型的软件,如 3Dmax, Google Sketch Up 等,也日臻完美,一些复杂的人的大脑都难以想象的物

体或者是只有几张从不同角度拍摄的某个物体的照片,这类软件也能够较为完美地生成对应的三维模型,给人们带来了空前的便利。同时,这类软件让人们在处理一些现有的三维模型,例如对模型进行拉伸变形等操作时更加方便,也能获得更好的更合理的效果。



(a) Kinect

(b) LynxAcamera

图 2-2 RGBD 设备

2.2.2 三维数据的存储

通过各个途径获得的三维模型用特定的方法进行保存,一般来说有点云模型和面模型两种保存形式。非接触式扫描模型常用的保存形式为点云模型,即只保存了物体上各个扫描点的坐标信息,将这些点按其坐标值出现在一个场景中,可以大致表现出所扫描物体的空间形态,点的密度越高,表现效果越好。面模型是将扫描所得的点云模型中的相邻点按一定的规则连成多边形,各个多边形之间按照不同的大小和连接关系,可以表现出极为复杂的形状。由于三角形的稳定性,所以三角网格的效果最好。同时由于面信息的使用,三维模型可以表现出颜色等高层次的特征,使其表现的形式更加丰富,在现实中的应用范围也更加广泛,这类模型多由三维软件直接生成,或者是由三维点云模型按照一定的计算方法重建出来的。

2.3 国内外三维模型特征提取的研究现状

无论是点云模型还是面模型,其所包含的信息都较为抽象,难以直接使用,因此,如果我们要使用三维模型,需要从三维模型中提取一些具体的信息即特征,来进行使用。

10 多年来,三维模型的研究发展极快,取得了巨大的进步,各种三维模型的研究方法被纷纷提了出来,这些方法多数都与三维模型的特征描述相关,文献 [3] 整理出一些早期三维模型研究工作中常用的特征描述符,文献 [4] 和文献 [5] 则是一些较新的研究进展。