

物联网感知技术与应用

智能全景视频感知

中

汤一平 著



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>

物联网感知技术与应用

——智能全景视频感知（中）

汤一平 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书是一部比较全面和系统研究全景视觉理论、全景视觉感知技术及应用的学术专著。上册介绍了智能视频感知的基本原理和技术,并提出了各种智能全景感知实际工程应用的解决方案。中册围绕实际应用案例来探讨“人与环境交互”的智能感知的实现方案,尝试通过感+知+智能+应用背景,有层次、分步骤、由浅入深地介绍全景智能感知技术方面的基础知识和专业知识,主要围绕智能交通、安防、机器人等涉及国计民生的重要领域,给出智能视频感知框架及具体的实现方法,涉及内容偏重于物联网时代的信息消费领域。书中内容紧贴物联网时代,从应用需求到实现案例,阐述实现智能感知的工学方法,以艺术与科学结合的方法来探讨计算机视觉方面的科学问题,从智能感知在不同领域中的应用启发读者的创新思想,以综合平衡提高读者的理论水平和实战能力,强调学而致用。

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

物联网感知技术与应用:智能全景视频感知. 中/汤一平著. —北京:电子工业出版社, 2015.2
ISBN 978-7-121-25585-4

I. ①物… II. ①汤… III. ①互联网络—应用—视频系统—监视控制—研究 ②智能技术—应用—视频系统—监视控制—研究 IV. ①TN94②TP277

中国版本图书馆 CIP 数据核字(2015)第 038076 号

策划编辑: 万子芬

责任编辑: 万子芬

印 刷: 三河市鑫金马印装有限公司

装 订: 三河市鑫金马印装有限公司

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: 29.5 字数: 755 千字

版 次: 2015 年 2 月第 1 版

印 次: 2015 年 2 月第 1 次印刷

印 数: 2 500 册 定价: 98.00 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 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前言

物联网时代用一句话进行概括，就是要将物理空间与信息空间有效地进行融合，实现智慧空间。几千年前，中国伟大的思想家孔子就提出来要“知天”、“畏天”，要求人们认识自然界，自觉地利用自然界为人类社会谋福祉；同时要求人们对自然界有所敬畏，规避各种自然灾害，自觉尽到保护自然的责任，最终形成了中国传统思想的核心价值理念之一，即“天人合一”思想。“天人合一”用现代科技语言来描述就是智慧地球或者智慧空间，物联网时代实现智慧空间的目的是为人类社会谋福祉。

智能全景视频分析技术作为实现智慧空间的一项重要核心技术，其中全景视频获取技术将成为一种可以从中提取大量关键信息和有价值数据的信息源，进而使具备智能视频分析功能的视频类设备升级成为一种特殊的、重要的物联网感知终端设备，为各种行业应用（智能交通、商业智能、防灾减灾、安全生产、智能安防、安全监护等）提供强大的元数据支撑。物理空间环境中大量包含趋势性、经验性、不确定性、随机性和模糊性的信息，要让环境能自主“说话”，即实现智慧空间的一种最佳选择是应用智能全景视频分析技术，这也是物联网最有价值、最具有竞争力的意义所在。

笔者从 21 世纪初开始关注智能视频分析技术，当时通过谷歌搜索引擎用“智能视频分析”关键词进行检索，几乎没有中文的智能视频分析的检索结果，但是今天的检索结果有 158 万多条。的确，最近几年智能视频分析技术得到了众多科技工作者的普遍关注，尤其是进入 21 世纪以来，计算机视觉技术得到了飞速的发展，每年全世界范围内有几万篇论文发表，大量的智能视频分析技术产品进入市场。特别是最近，随着物联网技术概念的提出，物联网时代将面临着大量的趋势性、经验性、不确定性、随机性和模糊性信息的获取、加工、传输，这样基于智能视频感知技术的物联网技术就自然被推到了风口浪尖上。事实上从国外物联网的发展规划来看，自 2020 年起将在物联网技术中全面导入智能视频分析技术。

“工欲善其事，必先利其器”。“测量技术是信息技术的重要组成部分，是信息技术的源头。”科学技术是第一生产力，而现代仪器设备和现代检测手段是第一生产力的重要条件。视觉是人类观察世界和认知世界最重要的手段。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，这既表明视觉的信息量大，人类对视觉信息有较高的利用率，同时也体现了视觉功能的重要性。

计算机视觉的最大优点是与被观测的对象无接触，在观测者与被观测者之间不会产生任何损伤，这是其他感知方式无法比拟的；另外计算机视觉能不知疲倦、始终如一地进行观测与计算，可广泛地应用于国民经济、科学研究及国防建设等各个领域。计算机视觉的

终极目标就是达到人类视觉的能力，接收视觉信息，感知场景中的运动对象，同时能够对运动对象进行智能化语义分析与理解，最终得出运动对象的行为判断，让物理环境能自主“说话”。从第一、二、三次工业产业革命的角度来说，虽然在第三次产业革命中出现的自动化技术和计算机技术拓展了人类的计算能力，但这三次工业产业革命的效果还只是延伸了人类躯体的各种能力。智能感知技术，尤其是智能视频感知技术的出现将能从更高层面融合人类文明和前三次工业产业革命的成果，让人类社会能“知天”，让“人天合一”的技术为人类社会谋福祉。

本书分为上、中、下三册。上册偏重于介绍智能视频感知和全方位视觉传感器的理论；中册主要围绕智能全景视频分析技术的各种应用问题进行详细的讲解，重点是智能交通领域的智能视频分析应用；下册详细探讨智能全景视频分析技术在特殊行业中的典型应用、3D视觉和基于视觉的各种新型传感器的开发及应用的典型案例。本书的中册共分5章，这5章更偏重于物联网时代的信息消费领域。具体分述如下。

第1章主要介绍智能视频分析的关键核心技术，对视频图像序列中的背景建模，前景对象的检测、识别、跟踪，以及对其行为进行理解和描述这几个步骤的主要关键技术与实现方法进行了比较全面的分析和阐述。本章是智能全景视频分析技术的重要理论基础，可以作为高年级本科生或研究生计算机视觉课程的必修内容。

第2章基于全景计算机视觉的智能交通技术，主要介绍全景智能视频感知技术在智能交通领域中的各种应用。本章以物联网时代的ITS（智能交通）及相关的视觉检测技术为主线，系统全面地分析和阐述了物联网时代的ITS技术路线，其中包括交通安全运行环境的视觉检测、交通对象的视觉检测、交通事故的视觉检测、交通状态的视觉检测和停车位视觉检测等。

第3章智能视频分析技术在内河航道智能监控中的应用，它实际上也是一种智能交通领域的应用研究。本章全面系统地介绍了一种基于计算机视觉的内河航道智能监控系统，在实现方法上从概要设计、详细设计、检测算法实现和实验验证等方面进行了详尽的阐述；在实现功能上从船体目标跟踪、船体载重量识别、船舶身份识别、船舶异常行为识别和河道船舶流量统计等方面进行了全方位的研究论证。

第4章基于全景视觉的移动机器人技术，它是全方位视觉在机器人领域中的一个重要应用。移动机器人的一个重要需求是以观察者为中心的全景视觉（视觉=视频分析+坐标+测量）。如何实时感知移动机器人周围的环境，如何检测移动机器人在行走过程中的障碍物，如何获得移动机器人的自身状态，如何使移动机器人具有鲁棒的自主导航能力和如何通过全景视频图像处理、分析和理解来实现移动机器人的目标识别、建图、自定位是本章探讨的问题。

第5章基于智能全景视频感知的节能与安全辅助技术，给出了智能全景视频感知在电梯及节能领域的一些应用范例。对于群控电梯的智能调度问题，通过智能全景视频分析技术获取实时的需求，提高群控电梯的使用效率。通过本章的阅读，使得读者开拓视野，充分理解物联网时代的智能全景视频感知的应用前景并进一步把握其核心技术。

本书的内容适合于从事物联网和智能视频分析领域研究和实际工程项目的广大读者。书中从全景视频图像的获取、处理、传输和理解发布的过程出发，以各种应用案例系统讨论全景视频数据如何获取，如何将全景视频数据加工成视频信息，如何将视频信息加工成知识，然后如何用这些知识自动进行决策，为从事这方面学习和研究的学生、科研工作者

及项目开发人员提供帮助。书中的很多内容都是笔者多年来从事智能视频科学研究的成果,其中很多实例就是来自于笔者所申请的国家知识产权局国家发明专利中的内容,因此本书不仅可以作为专业技术人员的实用参考资料,也可以作为发明创新的启蒙教材,更适用作为从事物联网技术或计算机视觉的高年级本科生和研究生的教材。

作为从事物联网技术的高年级本科生或研究生计算机视觉课程的教材使用时,建议将本书作为数字图像处理和计算机视觉的高级课程或者后续课程。笔者曾经用本书中的几个章节作为研究生智能视频分析课程教材,上课效果及学员反映都很好;同时本书也可以作为从事视频监控企业的培训教材,用于提升员工智能视频方面的理论水平和实战能力。相信通过本书的学习,能拓宽读者的视野,从实战中真正把握智能视频分析技术的精华。笔者相信,基于视觉的智能感知领域覆盖面非常大,读者只要从工程实际问题出发,通过模仿本书的案例,按照书中介绍的工学方法,是完全可以进行各种创新活动的,期待看到读者的各种发明创造。

可以用八个字来概括本书的内容:“视觉艺术、认知科学”。本书中很大一部分内容涉及人的行为识别、理解及人类对各种事物的认知,其中也包括人类的第六感。视觉艺术的表现形式多种多样,不只是局限于线条、颜色、整体布局等平面的艺术形式,还包括三维的、动态的、虚拟的、现实增强的实用与观赏并存的艺术形式,其中视觉还包括坐标系、测量和解释三方面的内容。当我们观察一些不规则的图形或是一个全景图像时,只有被看作一种由清晰的方向、一定的大小及各种几何形状和色彩等要素组成的结构图式时,才能真正被知觉;目前的视频分析技术尝试用计算机视觉的方式来解释视频图像,我们要认识到视频与视觉是有很大差距的;当今的视频分析技术总体水平约相当于3岁孩童的认知水平。认知科学作为一门新兴的科学,是研究人、动物和机器智能的本质和规律的科学,是多学科相互交叉、渗透与融合的产物,研究内容包括知觉、学习、记忆、推理、语言理解、知识获得、注意、情感、意识和动作控制等高级心理现象,是探索人类的智力如何由物质产生以及人脑信息处理的过程。

想使本书的内容面面俱到是不太现实的。本书涉及物理学、高等数学、数值计算、计算机软件、数字图像处理、机器视觉、计算机视觉、模式识别和人工智能等方面的内容,这些方面的具体内容已有专门的相关教材和参考书籍。由于本书直接面向解决智能全景视频分析方面的一些科学和实际工程应用问题,相信已经对数字图像处理和计算机视觉的一些基础知识比较了解的读者都能够看懂;对于一些尚未接触过数字图像处理和计算机视觉的读者,建议在读本书时配一本计算机视觉、机器视觉及模式识别方面的有关书籍,或者随时通过搜索引擎检索关键词来获得相关知识。

本书也是一本物联网技术的科普读物,尤其是对关注物联网技术的年轻学子来说,可以开阔视野,激发对这门“人天合一”技术的学习热情。本书还可以作为智能视频分析工程技术人员的自学教材和手册,中册是对上册第4章中所介绍的各行各业智能视频分析解决方案更为详细的阐述,是提高从事智能视频分析工程技术人员业务水平的有效途径。希望本书的出版能为我国物联网技术的发展、为普及智能视频分析技术尽微薄之力。

关于具体实现和编程语言问题,虽然本书所介绍的案例中绝大部分代码是用Java语言开发实现的,采用了基于Java语言的中间件技术,但是本书的介绍不依赖任何编程语言,而是使用了通用的算法符号和流程。对于已经学会编程的读者来说,实现这些算法并非十

分困难，有些可以直接调用 OpenCV、JavaCV 和 PCL 所提供的函数库；这一点从笔者所指导的学生身上已经得到证明，其中不少算法都是在本科毕业生和硕士研究生的毕业论文中实现的，借此也对学生们在本书中所做的工作表示感谢。

在本书的中册执笔过程中发现智能全景视频分析技术还存在很多未知科学问题，需要进一步探索，加上近年来计算机视觉领域新的研究成果不断涌现，书中不少内容还是笔者及团队几年前完成的研究成果，在此抛砖引玉。书中的一些研究得到了国家基金（基金号：61070134，项目名称：主动三维立体全景视觉传感技术研究；基金号：61379078，项目名称：基于物联网技术的生物式临震预测关键技术研究）的资助，本书的出版得到了浙江省重中之重学科“信息处理与自动化技术”和浙江工业大学 2014 年度专著与研究生教材出版基金的资助，在写作过程中得到了研究所实验室研究生们的帮助，在此一并表示感谢！本书（中册）成稿之际，同时还要感谢家人和朋友对笔者著书期间的理解和支持。由于时间仓促和学识水平有限，不足之处望读者不吝赐教。

最后，让我们怀着“天人合一”为人类社会谋福祉的理想，践行“人天合一”的技术路线，自觉地利用自然界为人类社会谋福祉，自觉地为保护我们人类赖以生存的环境尽到自己的一份责任。

汤一平 (typ@zjut.edu.cn)

于 2014 年 10 月

第 1 章 智能视频分析的关键核心技术	1
1.1 目标对象特征选择与运动表征	5
1.1.1 目标对象的特征选择	5
1.1.2 目标对象的运动表征	6
1.2 行为识别	8
1.2.1 基于模板匹配方法	8
1.2.2 基于状态转移图模型方法	9
1.3 高层行为与场景理解	13
1.4 行为理解中存在的问题与发展趋势	17
1.5 基于全景视觉的前景目标对象分割、定位及简单跟踪的实现	18
1.5.1 混合高斯模型	19
1.5.2 混合高斯模型的算法	20
1.5.3 基于混合高斯模型的 EM 算法	21
1.5.4 基于核密度估计的背景差法技术	22
1.5.5 基于 DSP 的全景视频多目标实时检测及跟踪的实现	24
1.5.6 阴影消除算法	32
1.6 基于全景视觉的多目标对象跟踪的实现	46
1.6.1 多目标跟踪方法概述	47
1.6.2 基于全景视觉的多目标跟踪技术	51
1.7 全景视频图像质量自诊断算法	59
1.7.1 视频亮度异常检测算法	60
1.7.2 视频图像清晰度异常检测算法	62
1.7.3 视频图像偏色检测算法	62
1.7.4 视频图像冻结检测算法	63
1.7.5 视频图像缺失检测算法	64
1.7.6 视频图像干扰检测算法	65
本章结束语	66
参考文献	67
第 2 章 基于全景视觉的智能交通技术	77
2.1 交通安全运行环境的监测	85
2.1.1 高速公路隧道交通事故特点	86

2.1.2	高速公路隧道交通安全保障体系	87
2.1.3	高速公路隧道交通安全保障体系的建立	88
2.1.4	高速公路隧道交通安全保障系统的运作模式	89
2.1.5	高速公路隧道交通安全保障系统的关键技术	89
2.1.6	基于机器视觉的高速公路隧道路况能见度检测技术	90
2.1.7	基于机器视觉的道路障碍物检测技术	93
2.1.8	基于机器视觉的隧道火灾检测技术	94
2.1.9	基于机器视觉的铁路路口安全检测技术	99
2.1.10	基于机器视觉的路面状态检测技术	109
2.1.11	基于全景计算机视觉的轨道交通安全检测技术	113
2.2	交通违法行为的检测	117
2.2.1	电子警察的基本功能	119
2.2.2	违章行为识别	120
2.2.3	闯红灯行为的视觉检测	121
2.2.4	违章车辆的身份识别	127
2.2.5	“闯绿灯”行为的视觉检测	138
2.2.6	违章超速行驶行为的视觉检测	139
2.2.7	违章逆向行驶行为的视觉检测	143
2.2.8	违章变道行驶行为的视觉检测	146
2.2.9	违章停车行为的视觉检测	147
2.3	动态道路交通状态的检测	157
2.4	静态交通状况的视觉检测	175
2.4.1	停车位状态检测的研究现状	176
2.4.2	基于采样点的停车位图像特征模型分析	178
2.4.3	停车位图像特征提取	181
2.4.4	基于采样点的背景建模技术	182
2.4.5	基于采样点的干扰滤波算法	182
2.4.6	车辆遮挡对停车位视觉检测的影响	184
2.4.7	停车位状态视觉检测	186
2.5	驾驶安全辅助技术	188
2.5.1	车辆周围全景影像获取技术	192
2.5.2	车辆行驶过程中关注性区域的生成	194
2.5.3	车辆偏离预警与保持	202
2.5.4	关注区域内的障碍物视觉检测	213
2.6	基于全景视觉的驾驶疲劳检测	215
2.6.1	驾驶员的行为特征技术及研究现状	215
2.6.2	基于全景视觉的驾驶疲劳检测系统的概念设计	218
2.6.3	基于全景视觉的驾驶疲劳检测系统的模块设计与实现	219
2.7	基于全景视觉的智能倒车辅助装置	235
2.8	基于计算机视觉的油门当刹车的防止技术	242

2.9 基于全景视觉的重大交通事故的快速测绘技术	245
2.9.1 BODVS 交通事故现场测绘平台的设计	248
2.9.2 BODVS 的测量及测量精度研究	251
2.9.3 基于 BODVS 测量技术的事故现场图绘制及实验研究	259
2.10 道路岔口交通状态的视觉检测和交通信号灯控制	264
2.11 基于全景视觉检测的交通信息云服务	272
本章结束语	278
参考文献	279
第3章 智能视频分析技术在内河航道智能监控中的应用	283
3.1 水路智能交通领域国内外研究现状	285
3.1.1 水路交通流参数检测的研究现状	285
3.1.2 内河船舶实际载重吨位实时检测研究现状	286
3.1.3 基于计算机视觉的内河航道智能监控系统框架	287
3.2 基于计算机视觉的内河航道智能监控系统的概要设计	289
3.2.1 系统硬件设计	289
3.2.2 系统软件设计	293
3.3 基于计算机视觉的内河航道智能监控系统的详细设计与实现	295
3.3.1 低层视觉: 船舶运动目标检测与分割	295
3.3.2 中层视觉: 船舶跟踪与交通流检测	304
3.3.3 高层视觉: 实际载重吨位估算与异常预警	312
3.4 船舶的身份识别	320
本章结束语	327
参考文献	328
第4章 基于全景视觉的移动机器人技术	331
4.1 基于全方位视觉移动机器人的自定位和地图创建技术	338
4.1.1 基于全方位视觉移动机器人的 SLAM 研究现状	342
4.1.2 基于全方位视觉的移动机器人拓扑导航	344
4.2 基于全方位视觉的移动机器人的避障	363
4.2.1 基于 AODVS 的障碍物测量原理	364
4.2.2 基于 AODVS 的全方位避障策略	366
4.2.3 移动机器人的视觉检测工作流程及相应算法	367
4.2.4 实验研究	369
本章结束语	371
参考文献	371
第5章 基于智能全景视频感知的节能与安全辅助技术	380
5.1 基于智能全景视频感知的自动扶梯节能与安全辅助技术	380
5.1.1 乘客人体运动对象与活动台阶对象的检测与分割	382

5.1.2	基于全景智能视频感知的自动扶梯节能设计与实现	395
5.1.3	基于全景智能视频感知的自动扶梯安全运行检测	399
5.2	基于计算机视觉的电梯轿厢内异常行为检测	409
5.2.1	电梯轿厢内人体前景对象的提取	412
5.2.2	电梯轿厢内人数统计	415
5.2.3	电梯轿厢内人体运动特征获取	419
5.2.4	电梯轿厢内异常行为检测	426
5.2.5	基于智能视频感知的电梯安全辅助系统架构	437
5.3	基于智能全景视频感知的群控电梯智能调度技术	440
5.3.1	轿厢内乘客数的检测	441
5.3.2	楼层中等待上行或下行乘客数的检测	442
5.3.3	基于智能全景视频感知的群控电梯智能调度	443
5.4	基于智能全景视频感知的集中空调节能技术	448
5.4.1	基于智能全景视频感知的集中空调节能系统设计	451
5.4.2	基于全景视觉的室内人数统计	452
	本章结束语	460
	参考文献	460

第 1 章

智能视频分析的关键核心技术

本套书上册第 2 章中已经简单地介绍了视频分析技术的一些关键核心技术，这里将更详细地研究和探讨如何将这些关键核心技术应用于各行各业的解决方案中。

视频分析技术范围很广，只要从视频中通过各种运算加工处理、提取有用信息或者知识的技术都可以称为视频分析技术。由于都是属于对视频内容的“分析、识别和理解”，因而目前已经比较成熟并已形成产品且在实际项目中得到应用的技术，如车牌识别技术、视频检索技术、视频人脸检测等，都属于视频分析技术。因为它们都属于对视频中有用信息的提取，如提取车牌号码，提取视频中的文字或特定图形等。严格意义上讲，这些技术都只是视频分析技术中比较简单的内容，只不过由于这些技术已得到较好地研究和应用，有了一些专门的称谓，才单独提出来。目前，通常所说的视频分析技术一般特指对视频中目标对象行为的分析、提取和识别，它常用于安防监控领域，其所指代的范围比字面含义已大幅缩小。

视频分析技术由于在智能监控、高级人机交互、虚拟现实、现实增强和基于内容的视频检索与解说等方面有着广泛的应用前景和潜在的经济价值，激发了广大科研工作者及相关商家的浓厚兴趣，已成为计算机视觉领域中备受关注的前沿方向之一。目前，在世界范围内已经开展了大量的视觉分析研究项目。例如，1997 年美国国防高级研究项目署(Defense Advanced Research Projects Agency, DARPA)设立了视觉监控 VSAM (Visual Surveillance And Monitoring) 重大项目，主要研究用于战场及普通民用场景监控的自动视频理解技术；实时视觉监控系统不仅能够定位和分割出人的身体部分并实现多人跟踪，而且能检测人是否携带物体等简单行为；1999 年欧盟 Framework 5 程序委员会设立重大项目 ADVISOR 系统，项目旨在开发一个公共交通（如地铁、机场）的安全管理系统，它覆盖了人群和个人的行为模式分析、人机交互等研究；2000 年，DARPA 又资助了 HID (Human Identification at a Distance) 计划，其任务是开发多模式监控技术，以实现远距离情况下人的检测、分类和识别，从而增强国防、民用等场所抵御恐怖袭击的能力；英国雷丁大学 (University of Reading) 先后开展了 REASON (Robust Methods for Monitoring and Understanding People in Public Spaces)、ETISE (Evaluation of Video Scene Understanding)、ISCAPS (Integrated

Surveillance of Crowded Areas for Public Security) 等项目, 以研究基于视觉的单人行为、人与人之间的交互行为识别和场景理解等; 此外还有能够对人进行跟踪与行为理解的实时系统 Pfinder (Person Finder), 由欧盟和奥地利科学基金会共同资助的大型视频监控技术研究项目 AVITRACK (Aircraft Surroundings, Categorized Vehicles & Individuals Tracking for Apron's Activity Model Interpretation & Check) 等。

最近几年, 国内对智能视频感知技术非常重视。科技部通过重大专项多年连续支持智能视频分析技术的相关项目, 国家基金委的立项中智能视频分析相关研究也得到了重点支持。同时, 国内商家也看到了该技术极大的市场机遇, 斥资投入智能视频分析的相关产品研发, 如海康、大华和华三等上市企业。总之, 目前整个社会对该领域的技术人员需求也比较大。

目前, 国内有些大学和研究机构, 如中国科学院自动化所模式识别国家重点实验室、亚洲微软研究院、北京大学视觉与听觉信息处理国家重点实验室、清华大学、武汉大学、西安交通大学、上海交通大学、浙江大学以及浙江工业大学等对智能视频分析的关键技术进行了较为深入的研究。

在学术方面, 一些国际权威期刊如 PAMI (IEEE Transactions on Pattern Analysis and Machine Intelligence)、IJCV (International Journal on Computer Vision)、PR (Pattern Recognition)、CVIU (Computer Vision and Image Understanding)、IVC (Image and Vision Computing) 和重要的学术会议如 CVPR (IEEE Computer Society Conference on Computer Vision and Pattern Recognition)、ICCV (International Conference on Computer Vision)、ECV (European Conference on Computer Vision) 和 ICPR (International Conference on Pattern Recognition) 等将智能化视频监控技术研究, 尤其是人运动视觉分析作为主题内容之一, 这些期刊为该领域的研究人员提供了更多的交流机会。在全世界范围内每年都有上万篇相关研究论文发表。智能监控涉及人的行为理解和描述, 当前人的行为理解与描述受到了广泛关注, 发表在上述权威期刊和重要学术会议上的关于行为理解研究的论文正逐年增加。关于行为理解研究的会议文章从 2002 年起急剧增加, 在 2005 和 2006 年每年发表的会议论文均保持在 70 篇以上, 近年来更有快速增长之势; 期刊文章从 2003 年也开始大幅增加, 尤其是现在 1 个月的文章数量就超过 2005 年全年的数量。行为理解已经成为了未来计算机视觉研究的前瞻性方向之一。

在知识产权方面, 读者只要将“智能视频分析”、“智能视频监控”、“计算机视觉”等关键词输入 SOOPAT 等专利检索引擎中就能检索到上万篇相关专利文献, 这些文献都是我们学习和研究的宝贵资源。

在本书的上册第 2 章中我们曾经介绍了 W4 监控系统, 该系统主要包括四个因素: When (何时)、Where (何地)、Who (谁) and What (干什么), 在不同的监控需求下可以利用不同的因素或者综合利用若干个因素, 有些场景下利用身份 (Who) 就可以监控, 有些场景下利用位置 (Where) 就可以监控, 但是在对人的身份和位置都不能限制的公共场所, 就必须进行行为的分析识别和理解, 即对 “What” 进行研究。

视频分析技术的处理大致上可以分为视频图像序列中检测、识别、跟踪, 并对其行为进行理解和描述这几个步骤, 如图 1.1 所示。这个过程可分为低层视觉模块 (Low-Level Vision)、数据融合模块 (Intermediate-Level Vision) 和高层视觉模块 (High-Level Vision)。

其中，低层和中层视觉模块主要包括背景建模、前景对象的检测、目标对象跟踪等运动分析方法，以及主要解决多摄像机数据融合处理问题；高层视觉模块包括目标的识别和有关于运动信息的语义理解与描述等。

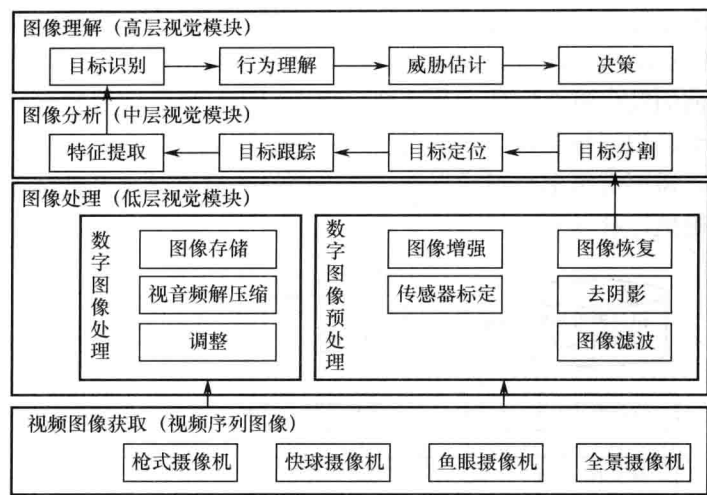


图 1.1 智能视频分析系统一般性处理框架

智能视频分析的研究内容主要包括运动目标对象的检测、分类、跟踪以及对监控场景中目标对象行为的理解和描述。其中，各种数字图像的预处理技术、目标对象的分割、定位、跟踪及分类均属于视觉中的低级和中级处理部分，而行为理解和描述则属于高级视觉处理。感兴趣目标对象的检测、跟踪和分类是智能视频分析中研究较多的三个问题，而行为理解与描述则是近年来备受关注的研究热点，它要对视频序列图像中出现的感兴趣目标对象的运动模式进行自动分析、识别与理解，最后用自然语言等方式加以描述，让机器代替人来监视场景及场景中的目标对象，得到对场景及目标安全性的评价，并采取相对应的决策和措施。

智能视频分析的主要理论基础是计算机视觉理论，David Marr 在 20 世纪 70 年代末提出了计算视觉三表象理论，首先将复杂的视觉过程变成了一个可计算的信息处理过程。图 1.2 为 Marr 视觉的三表象简图。他认为低层视觉从原始图像获得一些基本的图像特征，构成所谓的“要素图”；中层视觉在以观察者为中心的坐标系中构成对环境的维描述，即部分、不完整地进行描述；高层视觉从 2.5 维描述得到以物体为中心的坐标系中完整的三维描述。

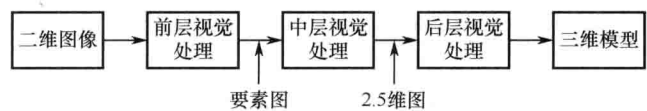


图 1.2 Marr 视觉的三表象简图

我们注意到 Marr 视觉理论表述中的一些关键词（以观察者为中心、以物体为中心和维描述），视觉主要需要几何、测量和解析三个方面的技术支撑，随着从低层视觉到高层视觉的进化，几何和测量的技术成分更显重要。目前智能视频分析技术主要关注视频图像的解析，所以遇到了很多瓶颈。本书作者认为，仅仅关注视频图像的解析是不够的，完美的智能视频分析技术肯定涉及几何和测量等方面的相关技术，毫无疑问需要利用人类的先验知

识来提高系统的性能。

行为理解（智能视频监控的最高层面）研究一般有特征选择与运动表征、行为识别、高层行为与场景理解等几个基本过程，如图 1.3 所示。特征选择与运动表征是在对目标进行检测、分类和跟踪等低层和中层处理的基础上，从目标的运动信息中提取目标图像特征并用来表征目标的运动状态；行为识别则是将从输入序列中提取的运动特征与参考序列进行匹配，判断当前的动作符合哪种行为模型；高层行为与场景理解结合行为发生的场景信息和相关领域知识，识别复杂行为，从而实现了对事件和场景的理解。其中场景理解是一项极其艰巨的任务，需要有很多知识的支撑，这方面往往需要通过人机交互等方式使得计算机能更快速、准确地理解场景。

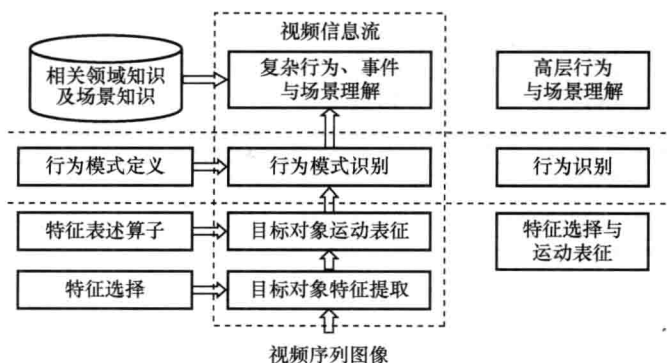


图 1.3 行为理解的一般性处理框架

一般的智能视频分析系统都以目标对象运动跟踪为基础，对目标对象进行分类、识别、分析并理解其行为等，从而得出对监控场景及目标安全性的判断。下面就对智能视频分析中的关键技术和研究热点问题简单回顾。

（1）快速准确的目标对象运动检测技术、跟踪及实时鲁棒的目标对象分类与识别技术是智能视频分析的基础。在这套书上册第 2 章中介绍了一种实时高效鲁棒的目标对象运动检测与跟踪技术，即 MHoEI（Motion History or Energy Images）算法，该算法较好地结合了目标对象的时空特征，采用运动历史和运动能量相结合的检测方法实现了对目标对象的提取与跟踪。

（2）探索如何根据目标对象的运动来实时控制摄像机的运动，实现对目标对象的主动跟踪。目前一般采用的技术是在线调节摄像机的参数，包括内部参数和外部参数来扩大视场范围，但是这种解决方案需要在线标定摄像机的内外参数，目前采用这种解决方案实现在线跟踪监控还十分困难。本书作者的解决方案是采用 Where 视觉与 What 视觉相结合的方法，即采用单视点的全景摄像机实现以观察者为中心 Where 视觉，采用高速快球以 Where 视觉为线索控制高速快球的各种参数，实现对目标对象的主动跟踪。

（3）多摄像机协作监控技术。单个摄像机的视场范围有限，要监控大范围的动态场景就需要多个摄像机。目前的难点是多摄像机协作、标定及数据融合问题。由于本书主要推荐采用全景视频感知方法，每个全景视觉传感器都只有一个视点，两个相邻的全景视觉传感器的监控视场上有一个交集，只要对全景视觉传感器的成像区域进行网格化处理，就能方便地实现同一监控目标对象在不同全景视觉传感器中的映射。

(4) 运动分析及行为理解技术。对目标对象的整体运动和局部运动进行跟踪、分析,获得目标对象的轮廓、姿态、运动轨迹、动作等信息,从而进一步理解目标对象的行为。这方面的技术涉及相关领域的知识及场景知识,需要具体情况具体分析,本书将从一般的智能视频监控、智能交通、老人监护等不同的应用角度详细解说运动分析及行为理解技术。

(5) 异常事件的检测、报警与未来事件的预测技术。智能视频分析系统的最终目的是为了解释监控场景中所发生的事件,根据系统设定的要求对异常事件进行报警,并能根据目标对象所处的状态对将要发生的事件进行预测。事件是行为的进一步综合,是对象与对象之间作用的结果,因此需要利用先验知识或者机器学习方法建立事件模型,然后再利用事件模型来分析、识别事件。可以利用贝叶斯估计、D-S 证据理论、模糊理论等来预测未来某事件发生的可能性或不确定性。

这里我们从行为理解的一般性处理框架出发,对目前行为理解的研究现状、存在问题及发展趋势等作一个概要性的回顾与介绍。

1.1 目标对象特征选择与运动表征

在智能视频分析中,由于视角的影响,同一种行为可能有不同的投影轨迹,而不同的行为可能有相同的投影轨迹。此外,光照变化、遮挡等现实环境因素也都给行为理解带来了很大困难。所以从原始视频数据中提取何种低层特征(如形状信息,包括侧影、轮廓;运动信息,包括位置、速度、方向等)来描述目标对象的运动,如何选择充分有效的描述方法来表征目标对象的运动状态,以及在时间分辨率和维数上减少计算量成为行为理解的关键问题。

1.1.1 目标对象的特征选择

表 1.1 为 2000—2007 年间 400 多篇关于行为理解研究的权威期刊与重要国际会议文章中使用特征的统计分析表。由表 1.1 可见,目前行为理解研究所采用的特征主要有如下四种:外观形状特征、运动特征、时空特征、形状特征与运动特征两者的混合特征。其中,形状特征和运动特征是两种最常用、最主要的特征,其比例均保持在 30%以上。另外,在行为理解研究中时空特征开始得到越来越广泛的使用。

表 1.1 行为理解使用特征统计分析表(%)

年 份	外观形状特征		运 动 特 征	混 合 特 征	时 空 特 征
	外观	模型			
2006—2007	36	17	30	4	13
2004—2005	38	1	44	11	6
2002—2003	39	8	42	4	7
2000—2001	43	7	36	10	4

基于外观形状特征^[1~9]方法先利用人体几何结构、轮廓或姿态以及区域颜色等信息来估计运动目标每一时刻的静止姿态,然后用这些静止姿态序列来描述人体运动。外观形状

特征容易获得且比较稳定，对纹理变化不敏感。基于运动特征方法^[10~14]则没有利用有关人体结构的任何形状信息，而是直接从图像序列中提取出目标运动特性（时分）信息（如光流信息、运动方向、轨迹、位置、速度等）来表征运动状态。心理学研究成果表明，运动特征具有静态形状特征无法比拟的优势，即使是在远距离、恶劣的能见度下和最小运动系统 MLD (Moving Light Display)^[15]中，人能够根据运动模式轻松地区别和识别目标的运动类型，而在此种情况下形状特征往往变得模糊不清。但目前的计算机视觉方法很难准确地提取运动特征，所以，为了充分利用形状特征和运动特征各自的优势，研究人员采用两者的混合特征来^[16~18]识别人的行为。此外，时空特征^[19~24]是近年来行为理解研究中使用比较多的特征。时空特征通过采用二维形状在一段时间内形成的空时立体或差分图像来描述行为，集合了形状和运动特征两者的优点。由于时空特征考虑了运动信息，也有学者将其归类于运动特征。

1.1.2 目标对象的运动表征

目标对象，比如人的运动表征是指从人的运动序列中提取合适的特征数据来描述人的运动状态，这是人行为理解中至关重要的一步。依据所用的特征，运动表征方法可分成基于外观形状特征的和基于运动特征的。

1. 基于外观形状特征的运动表征方法

依据是否利用人体形状模型的先验知识，基于外观形状特征的运动表征方法又可分为基于模型和基于外观特征（或基于视觉特征）两类。

基于模型方法的核心思想是首先建立描述空间中人体形状模型；然后利用目标的形态特征和关节位置等信息，在首帧将模型与输入图像序列进行匹配，在后续帧的跟踪过程中，定义损失函数来描述输入图像数据与模型之间的近似关系，并采用人体运动参数限制条件和最优策略（如最小平方^[25]）或随机采样策略（如粒子滤波^[26~27]）等方法将损失函数最小化来求得当前的运动姿态。常用的模型结构可以分为矩形框、棍棒形状^[28]、二维轮廓^[5]和 3D 立体模型^[25]。基于模型方法集合了先验模型知识和当前输入，适用于复杂行为的行为理解，但通常要求首先选择模型，初始化较难、计算量大，而且由于人运动的自由度高，容易产生局部最小值，很难找到全局最优且鲁棒的模型参数。此外，由于在后续的跟踪中存在累计误差，难以分析和跟踪长序列运动。

基于外观方法并不试图恢复人体运动的结构模型，而是直接从图像中提取物体的二维空间信息，如高度、宽度、侧影轮廓和颜色等，并用这些特征直接或通过与已经训练好的模型进行匹配来获得对当前运动的描述。在这类方法中，形状分析理论常用来描述运动目标的形状变化特征（有关形状分析请参考文献[29]）。常用的形状分析方法有质心-边界距离^[1, 3]、傅里叶描述符^[8]、形状上下文^[9]和统计形状分析^[30, 2, 31, 32]等。其中，质心-边界距离方法用一维函数对形状的质心到形状边界点的距离进行编码，这种描述符简单但不能精确地编码非凸起形状，如质心与一个边界点之间的距离连线可能与另一个边界点相交，这会造成细节的丢失。

Poppe 等人^[8]提出用傅里叶描述算子恢复人姿态，其思想是用固定的边界采样点对人体二维侧影轮廓进行采样，并对采样点进行傅里叶变换，利用变换后傅里叶系数的低频部分