

DUOSHEXIANGJI
XIETONG GUANZHU MUBIAO
JIANCE GENGZONG JISHU



多摄像机协同关注目标 检测跟踪技术

◎ 颜志国 等编著

◎ 公安部行业标准《视频监控系统多摄像机协同技术基本要求》课题组



东南大学出版社
SOUTHEAST UNIVERSITY PRESS

公安部行业标准《视频监控系统多摄像机协同技术基本要求》课题组

多摄像机协同关注目标 检测跟踪技术

颜志国 梅 林 徐 峥 蔡 烜 编著
段慧仙 孙明霞 葛 昊 时勇杰



东南大学出版社
SOUTHEAST UNIVERSITY PRESS

• 南京 •

内 容 提 要

本书以视频监控中关注人员目标的精准跟踪识别为抓手,以多摄像机联动协同跟踪技术为主线,以视觉基因组为内核,以视频结构化描述技术体系为筋骨,以视频感知融合主动安防机制的现实应用为脉络,通过枪球联动、空地联动、一体化指挥、大场景全景监控、摄像机语音调度、VR可视化应用等热点应用的系统阐述,形成了对当前视频监控应用热点技术图景式的描绘,从技术细节到宏观应用的分层叠进的探索,使得读者对视频监控网络中的多摄像机协同技术及其融合应用的发展脉络有了一个由表及里的立体直观认识。

本书适宜阅读对象为对多摄像机协同智能视频监控系统有所了解的安防从业技术人员,也可为具备一定的视频结构化分析领域基础知识的科研人员和高年级本科生及研究生提供多摄像机协同跟踪方向较为全面的技术参考。

图书在版编目(CIP)数据

多摄像机协同关注目标检测跟踪技术/颜志国等编
著. —南京:东南大学出版社,2017.6

ISBN 978-7-5641-7207-7

I. ①多… II. ①颜… III. ①视频系统—监控系统—目标跟踪—研究 IV. ①TN948.65

中国版本图书馆 CIP 数据核字(2017)第 126923 号

出版发行:东南大学出版社

社 址:南京市四牌楼 2 号 邮编:210096

出 版 人:江建中

网 址:<http://www.seupress.com>

电子邮箱:press@seupress.com

经 销:全国各地新华书店

印 刷:江苏凤凰数码印务有限公司

开 本:787 mm×1092 mm 1/16

印 张:14

字 数:350 千字

版 次:2017 年 6 月第 1 版

印 次:2017 年 6 月第 1 次印刷

书 号:ISBN 978-7-5641-7207-7

定 价:39.00 元

本社图书若有印装质量问题,请直接与营销部联系。电话(传真):025-83791830

前言

Preface

针对典型监控场所宽视野场景中行人目标的检测跟踪要求,我们对基于多摄像机协同策略的关注目标检测和跟踪机制进行了研究,从系统组成上分析比较了模拟架构、数字架构和高清网络架构的各自优缺点,并以双摄像协同中的枪球联动机制为例,对具有重合视野的双机联合标定方法进行了探讨,进一步研究了基于 DPM 的行人目标检测和基于人脸检测的关注目标聚焦机制。在实际应用中需要考虑算法实时性的要求,我们也简要介绍了算法实时性方面的提升策略,一方面论述了轻量化算法设计(包括前端智能化等方面的考量,如基于 DSP 或者安卓的视频结构化描述的轻量化);另外一方面则从并行计算的角度,论述 GPU 计算在这方面的实现。

针对监控系统关注人员人脸图像快速检索的需求,我们对人员脸部信息的结构化描述技术进行了研究,将整个监控视频从行人检测、联动跟踪、人脸检测与语义描述全部过程统一在视频结构化描述架构下的高效搜索框架内。值得一提的是,本书注意到深度学习机制在人脸检测方面的最新进展,以安防场景下人脸验证这一实际应用为实例,较为系统地介绍了深度学习框架下的人脸验证功能的实现。

此外,考虑到物联网架构下视频监控深度应用日新月异快速发展的现状,我们也简要介绍了毫米波雷达和摄像机联动、多传感器融合与城区场景激光三维建模、无人机的低空视频监控和地面摄像机网络联动等新颖应用,并对近年来热门应用如 VR 眼镜和现场摄像头之间的视频传输、基于语音控制的摄像机网络调度也进行了概要性的介绍。

本书第 1 章由孙明霞编写,第 2 章(第 1 节和第 2 节)、第 3 章、第 7 章(第 3 节和第 4 节)、第 8 章(第 2 节和第 3 节)及第 10 章由颜志国编写,第 4 章由时勇杰编写,第 5 章由葛昊编写,第 2 章(第 3 节和第 4 节)及第 9 章由蔡烜编写,第 6 章由段慧仙编写,第 7 章(第 1 节和第 2 节)及第 8 章(第 1 节)由徐峥编写。全书总体架构、技术框架由梅林设计,全书由颜志国统稿。孙明霞负责全书勘误和部分图表绘制工作,李滇博提供了深度学习人脸验证方面的工作内容,在此一并致谢。本书由上海市经信委“下一代视频监控网络技术研究及应用服务示范建设”项目专项资金支持出版。

公安部第三研究所物联网技术研发中心

多摄像机协同技术研究组

2017 年 1 月

目录 Contents

1 绪论	1
1.1 视频监控系统的发展	1
1.1.1 传统视频的不足	1
1.1.2 智能视频监控的应用	2
1.1.3 大场景多摄像机协同监控	5
1.2 视频协同跟踪研究现状	15
1.3 视频侦查领域的应用需求	18
2 运动目标检测与跟踪	22
2.1 运动目标检测	23
2.1.1 混合高斯建模	23
2.1.2 光流法	23
2.1.3 监督学习法	24
2.2 行人检测算法	25
2.2.1 头肩模型	27
2.2.2 DPM 模型	29
2.2.3 行人检测算法评价准则	30
2.3 跟踪策略分类	31
2.3.1 点跟踪法	32
2.3.2 核跟踪法	36
2.3.3 剪影跟踪法	39
2.4 常用人体跟踪算法	42
2.4.1 卡尔曼滤波法	43
2.4.2 MeanShift 算法	43
3 全景生成及枪球联动	46
3.1 全景生成技术	47
3.1.1 静态全景拼接	47
3.1.2 动态全景拼接	49

3.2	枪球联动系统简介	50
3.3	枪球联动系统结构	54
3.3.1	模拟架构枪球联动	55
3.3.2	数字高清架构枪球联动	56
3.3.3	网络高清架构枪球联动	58
3.4	协同控制策略	58
3.4.1	目标静止时的协同控制	58
3.4.2	目标运动时的协同控制	59
3.4.3	运动状态转换时的协同控制	59
3.5	“一枪多球”及“多枪一球”	60
4	人脸检测技术	63
4.1	人脸检测技术的发展	63
4.2	常规人脸检测方法	64
4.2.1	肤色建模	65
4.2.2	Haar 和 LBP 特征	68
4.2.3	分类器级联	70
4.2.4	基于肤色建模与 Haar 特征的人脸检测	74
4.3	深度学习框架下的人脸检测	76
4.3.1	FacenessNet 人脸检测算法	76
4.3.2	SeetaFace 人脸检测算法	77
4.3.3	Faster R-CNN 人脸检测算法	78
4.4	人脸定位与放大	79
4.5	人脸检测加速技术	81
5	深度学习机制下的目标识别	82
5.1	深度学习发展历程	82
5.2	深度学习主流框架介绍	83
5.2.1	Caffe	84
5.2.2	Torch	105
5.2.3	TensorFlow	105
5.3	深度学习在人脸检测识别方面的最新进展	107
5.3.1	科研院所方面的工作	107
5.3.2	产业方面的工作	107
5.4	我们的工作	108

5.4.1	深度学习人脸验证	108
5.4.2	ImageNet 方面的工作	114
5.4.3	TRECVID 方面的工作	115
6	摄像机标定与图像特征匹配方法研究	116
6.1	透视摄像机标定方法	117
6.1.1	传统的摄像机标定方法	118
6.1.2	摄像机自标定方法	127
6.2	全景摄像机标定方法	130
6.2.1	多摄像机联动	131
6.2.2	鱼眼摄像机标定方法	137
6.2.3	全向摄像机标定方法	139
6.3	图像特征匹配方法	147
6.3.1	特征点匹配	148
6.3.2	特征直线匹配	150
6.3.3	重复性纹理图像匹配	151
7	人员目标结构化描述及检索	154
7.1	视频结构化描述技术	154
7.2	视频结构化描述和语义网	156
7.3	人员信息结构化描述	158
7.3.1	人员面部信息结构化描述	158
7.3.2	人员行为模式结构化描述	167
7.4	个性化检索引擎	168
8	监控大数据与主动安防体系	170
8.1	视频监控进入大数据时代	170
8.1.1	视频大数据的应用难点	172
8.1.2	基于大数据框架的下一代视频监控系统	174
8.1.3	基于 VSD 的监控视频大数据分析	176
8.2	主动视觉感知融合安防体系	180
8.2.1	知识图谱	182
8.2.2	视觉基因组	187
8.2.3	基于主动视觉融合的安防体系	195
8.3	语音识别及 VR 眼镜等新技术的应用	199

9 视频一体化指挥应用	201
9.1 发展现状	201
9.2 立体防控技术解决方案	205
9.2.1 立体防控技术	205
9.2.2 大型活动立体防控平台	206
9.2.3 大型活动立体防控平台应用	207
9.3 未来工作	211
10 总结与展望	212
10.1 本书总结	212
10.2 领域发展展望	212
参考文献	214

1 绪 论

1.1 视频监控系统的的发展

1.1.1 传统视频的不足

视频监控是利用视频数据来对场景进行监控,并对异常事件和人员行为进行联动报警。世界进入 21 世纪后,由于宗教、地缘政治等因素,重大恐怖事件频发,全世界在震惊之余越来越重视安全防护,通过科技手段来解决遇到的安全问题是国际范围内学术界以及企业界关注的重点,也是安防产品升级换代的现实驱动力。视频监控系统由于其应用中的直观性、便捷性在安防领域起着十分重要的作用,作为人员监控、事件感知的重要手段已经在机场、火车站、公交站以及地铁站等公共场所广泛应用。在民用市场的安全保卫方面,视频监控系统也起到了有效的作用,被应用在大型仓储、高档社区、连锁超市等场合。相对其他监控手段,视频监控更为直观,监控范围的拓展也可以通过更换宽视野镜头或者球机的巡视功能来实现,相关异动的检测、查看非常迅捷,现场态势一目了然。视频监控系统能够在犯罪事件发生之后提供证据,成为警方判案的重要依据,通过图像分析技术提炼关键线索,协助警方搜寻犯罪嫌疑人。

传统的视频监控系统中有很大的缺陷,比如当海量的视频信息输入时没有相对应的高效检索系统,通常都是通过人工根据所需进行查看和处理,存在着“找不到、找太久、成本高”等应用困境,空有海量资源而不能有效发挥效能,只有很少一部分视频能够发挥作用。采集到的大量的视频数据长期闲置,过期后被完全删除,其中有效的信息完全丢失,这是对数据资源的一种极大浪费。人工处理视频数据,耗时耗力,已经不适应于当前社会的发展。仔细分析,现在的视频监控系统主要有以下缺陷:首先,传统的视频监控系统,往往是把在不同的场景中采集到的画面传输到监控大厅的视频墙上,一个视频墙可以显示十几个不同的画面,有时或许更多。这样,监控人员不可能同时严密观察每一个画面,只能是不定时巡查图像画面。这样就没有办法保证异动情况被及时发现。重要企事业单位一般都是多增派人手三班倒,通过人工不间断地监视重点场合的实时状况,以防错漏重要异动。其次,传统的视频监控系统为了得到较为理想的监控效果,要求监控人员时刻紧盯画面,这往往是繁重而枯燥的工作,由于人体的生理特性,疲倦的时候,人工的疏忽不可避免。最后,为了

不留下死角,传统的视频监控系统需要架设大量的摄像机,这些摄像机的视场的设计往往都是固定的,不能保证都集中到目标上,摄像机之间的联动不足,摄像机点位配置缺乏科学规划,视频信息无法充分利用。

总之,在传统监控系统中,计算机只是储存照相机或摄像头获得的图像或视频数据,需要人眼去识别场景中的物体和事件。这些系统有的是人在控制中心实时监视,需要消耗大量的人工,另外由于人的疲劳容易出现疏漏;有的是将视频储存起来,以备事件发生后查询,起不到警报的作用。随着计算机视觉技术的发展,越来越多的视频监控系统及应用期望计算机能够对图像或视频数据自动进行分析和处理,及时对一些目标和事件进行检测,并能够根据语义知识库进行一定程度推演,得到一些结论或者预警,从而在某种程度上实现人类视觉系统和人脑的某些初级研判功能,或者说实现智能监控。

1.1.2 智能视频监控的应用

近十年来,计算机视觉技术飞速发展,出现了智能视频监控系统,逐步解决了传统视频监控系统中存在的典型问题。智能视频监控系统是在计算机视觉技术的基础上,利用模式识别和机器学习等人工智能方法,对采集到的视频进行实时处理或离线批量分析,来达到目标检测以及事件特性分析的目的。通过上述处理过程进行目标的物理特征和表观特征提取(比如,行人的高、矮、胖、瘦,性别,种族特征,步态,衣着等),更进一步来理解目标行为,对异常的行为进行事前预警和实时报警。智能视频监控系统将大大减少人员的投入,将人从传统的耗时耗力的视频查验工作中解脱出来,且能极大地提高监控效果。智能监控系统的应用前景十分广泛,逐渐成为计算机视觉研究领域的热点。智能视频监控极大地解放了人力,为整个社会监控节约了人力资源,特别是目前在中国正逐步跨入老年社会,经济结构正由劳动力密集型向技术密集型转型,智能监控技术的发展更显得尤为重要,具有鲜明的时代意义。以上技术进步释放的人力资源体现了视频监控技术的成本优势,这点反过来又会推动智能视频监控系统的普及,打造真正的智能城市和安全城市。

人工智能技术在近十年中取得了飞速的发展,但是,还无法满足实际应用复杂场景下的多层次多方位的个性化需求。究其原因,主要是在实际的应用中,场景环境差异很大,并且对实时性、准确性和稳定性的要求都很高。从这些需求看来,智能监控系统的研究是具有很大的现实意义和社会、经济价值的。

智能视频监控在技术实现上主要包括三个层次,从底层特征分析到语义描述逐级递进,高一层次在低一层次的基础上建立。第一层实现目标检测与跟踪,是对全画面的像素级的操作。该层在连续帧中将目标从背景中分离并标记。例如在视频中检测出所有的行人和车辆并跟踪。第二层对关注目标的特征提取,实现身份识别。比如人脸识别、车牌识别等,又如行人身高、年龄、性别、种族等特性分析;对机动车检索系统来说,车辆颜色、种类、品牌等特征的识别与检索也是在这一层实现的。第三层是根据前面两层的结构化信息来理解、预测关注目标的行为,并在知识图谱的架构内,实现视频目标、事件的智能解析,支撑视觉基因组数据的构建。在实际应用角度上,第一层和第二层的应用比较多,第三层的应用还处于初级阶段,实际应用中的例子比较少。例如,现在视频摘要技术的应用较为广泛,是指将同一背景下的不同目标进行拼接,把几个小时长的视频数据进行浓缩,最终成为

包含检测目标的几分钟的视频。该应用所需的基础技术就是目标检测与跟踪。视频摘要技术目前也从被动式摘要技术发展到主动式摘要。简单地说,被动式摘要技术主要是过滤掉视频中的空白画面或者静止画面,保留运动画面;主动式摘要技术则可根据预定义的规则对视频内容进行浓缩,预定义的规则包括目标特征和事件等,如“黄色上衣的行人”“车辆闯红灯”。视频检索技术同视频摘要紧密相连,通过输入结构化的信息能从监控视频中将对应的视频片段查找出来,如搜索特定颜色种类的车辆等。从这种角度理解,也可以认为主动式视频摘要就是视频检索的一种应用,该应用的基础就是第二个层次。比较常见的智能视频分析应用方面,比如行人检测(越界、徘徊),车辆检测(犯规检测、车牌号)等,这些是在第二层的基础上的应用。在第二层提取物理特征要求图像清晰,若拍摄到的图片是模糊的,人脸检测及识别、车辆识别等算法的精度会大幅降低。

智能视频监控系统要求计算机系统能够感知场景中的物体和事件,识别出什么时刻什么物体在什么位置干什么。智能视频监控系统可以应用的领域很广泛,目前主要集中在治安事件监控、场所安防监控、交通智能监控、银行智能监控、工业智能监控等。智能视频监控系统在不同的应用领域中对监控结果要求也不同:有的为了及时发现异常行为发出警报,需要实时获得车辆或行人的运动轨迹,比如路边交通监控系统和银行监控系统;有的为了统计交通流量或识别障碍物和道路标志,需要检测识别场景中的行人、车辆甚至道路标志,比如路边车载交通监控系统和自动驾驶系统;另外,为了更好地监控场景,智能视频监控系统往往还需要获得场景的一些信息,比如场景深度信息、路面所处位置等。在设备方面,根据监控场景的摄像头数量,视频监控系统可以分为单目摄像头监控系统 and 多目摄像头监控系统;摄像头根据焦距视角的参数固定与否可分为固定焦距静止镜头和参数可调的主动镜头。

这两年国内摄像机终端厂家在成像技术、应用标准统一规范等方面有了长足的进展,主要体现在三个方面:(1)H. 265 编码标准及 SVAC 标准的推广应用。H. 265 编码标准将高清视频流传输的网络带宽压力极大缓解,同等的画面质量仅需要 H. 264 编码标准带宽的一半;比如,200 万像素高清网络相机的视频流,基本上 2 MB 的带宽即可满足主码流传输要求。SVAC 标准是《安全防范监控数字视音频编解码技术标准》(Surveillance Video and Audio Coding)的简称,从安防监控行业独特要求出发,支持感兴趣(Region Of Interest, ROI)区域可变质量编码和可伸缩视频编码,在带宽及存储有限的情况下提供更符合实际需要的高质量视频,着力解决视频监控图像互联共享及便捷维护问题。此外,28181 协议(公安部组织通过的国标 GB/T 28181—2011)的推出,也为解决城市视频监控报警联网系统的深度应用提供了应用规范,统一了信息传输、交换、控制等方面的基本要求和安全性要求。但是,GB/T 28181 标准对网络系统的安全解决方案还不够完善,在 QoS 设计方面存在缺失,这些,都是在实际应用中需要进一步解决的问题。(2)微光条件乃至星光条件下成像质量大幅提升。海康、大华、天地伟业等主流安防摄像机终端厂家纷纷推出的星光摄像机,使得弱光场景下的视频监控内容的可分析性大大增强,将视频智能分析推向“全天候”;基于以上产品,在一些夜晚照明不足的城郊和乡镇地区,先后推出了“雪亮工程”,有效地解决了城市边远地区和乡镇夜晚监控失效的问题,有效遏制了夜晚违法犯罪活动的发生。由此可见,安防行业的每一次技术进步都能够快速促进产品升级换代,技术成果转化的效能最直接地反映在社会公共安全保障上,社会效益明显。(3)像素动态瞬时分配技术。以格林深

瞳的人眼摄像机 Foveacam 为代表,号称有接近 10 亿像素的人眼辨识力的人眼摄像机正式走上台前。像素动态瞬时分配是该类相机的核心技术,能够瞬间把局部画面有效像素提升百倍,整体的画面甚至能够达到数亿的等效像素。格林深瞳称 Foveacam 能够瞬时分配有效像素达 2 亿,在范围 50 m 之内进行人脸识别,在 100 m 内能看清全身特征,视场角也达到了 70° ,抓拍覆盖面积很大。

夜视场景应用中,带有红外夜视功能的摄像机可以监控数十米到 100 m 的监控范围。但是画面为灰度图像,缺乏监控目标的纹理细节,用于粗略的目标侦测和事件检测尚可,用于目标智能分析,则困难较大。星光摄像机和人眼摄像机的出现,为宽视野远距离场景和夜视场景下的视频监控系统智能分析提供了图像采集端的硬件保障。图 1-1 为普通摄像机夜景照片(左)与星光摄像机实际效果画面(右),图 1-2 为人眼摄像机及其实际应用效果。



图 1-1 普通摄像机夜景照片(左)与星光摄像机画面(右)



图 1-2 人眼摄像机(左)与实际应用效果(右)

(图片来源: <http://www.deepglint.com/fovea>)

由于成本原因,很多实际应用的视频监控系统是采用单目静止摄像头来对场景进行监控。但是,在大场景监控应用中,基于多摄像机的全景交互监控和关注目标协同跟踪已经成为应用的热点,也是产业界新产品迭出的最新领域(比如公安部第三研究所推出的枪球联动产品,大华海康的最新枪球联动产品)。多摄像机协同监控系统是在单一摄像头监控基础上,通过多摄像机之间的协同和联动,实现视频监控系统功能的拓展和监控场景的扩大,是目前视频监控系统的主流应用方向,即多摄像机协同监控系统。多摄像机协同监控系统是以目标表示和定位、滤波和数据关联等理论为基础,还涉及多摄像机之间的标定、联

动跟踪机制的设计,算法实现更加复杂。通过多视角信息对同一客观场景进行不同的表达,其间存在着必然的相关特性,在目标跟踪的环节中,要考虑各个视角的信息交互。不同摄像头信息之间的交互和融合使得算法的复杂度加大,但优点在于目标跟踪的适用性和鲁棒性得到了提高。多摄像头系统开阔了监控视野范围,可以较好地弥补由于遮挡而造成的跟踪失败;通过信息的连通交互来缩小搜寻目标的范围,极大地提高了实时性,并且通过信息融合减少了跟踪误差。多摄像机协同监控系统在复杂开放场景的应用还存在着多目标准确连续跟踪、任务分配机制优化、精准协同标定等诸多难点,不断激发着学术界的研究热情。

1.1.3 大场景多摄像机协同监控

随着视频监控技术的快速发展,对大视场视频监控的需求不断扩大。如对整座大型机场、跨江大桥、整幢高层住宅等的监控,单个摄像机无法达到如此大范围的监控,则需要架设多个监控设备。

目前,安防领域中主要有枪型摄像机(枪机)和球形摄像机(球机)两种。在大视场监控环境下,使用枪机的缺点是场景较多时要不断来回切换,易导致监控人员精神疲劳,注意力只能集中于有限的画面上,重要的信息可能会遗漏掉。球机监控是在监控场景中通过转动球机进行监控,其不足之处在于视野有限,只能监控一个场景。在通过监控设备进行目标跟踪时,若采用枪机,则目标容易出现在多个画面中。另外,监控画面的不断切换也会影响到监控者捕捉视频画面信息,枪机的安装朝向固定,监控场景固定,目标跟踪上很容易跟丢。若采用球机进行目标跟踪,则在跟踪过程中,由于监控视野集中在跟踪目标(“点”)附近,导致容易漏掉感兴趣区域(“面”)上的其他重要信息。综上,单纯的枪机或者球机在大场景中的应用中存在顾此失彼,“点”“面”无法兼顾的不足。

一般的视频监控场景,摄像机同目标的距离在 10 m 以上。这样在监控行人时成像小,难以看清行人特征。这种情况在球机监控中,我们可以通过手动调整云台放大光学变焦来获得高分辨率的行人图像。不过,这样的操作只有在行人静止不动时有效,若行人在运动行走过程中,往往放大后,目标已经走出了摄像机的视野,此外,由于网络延迟等原因,往往操作还存在滞后,因而依靠操作员手动对目标的跟踪放大是一件困难的事情。

在实际监控应用中,往往需要获得清晰的视频图像,通过目标位置检测,对运动目标进行自动跟踪,要保证跟踪目标始终在镜头中央出现。通过使用变焦 PTZ 摄像机,解决大场景下个体目标近景跟踪问题,即 PTZ 跟踪技术。该技术利用 PTZ 摄像机和传统跟踪算法,再结合 PTZ 摄像机控制算法,可以在一定程度上解决该问题。PTZ 摄像机作为 PTZ 跟踪的核心硬件,解决了固定摄像机视角无法改变的问题。PTZ 摄像机有三个主要的可控参数:Pan(左右)、Tilt(上下)和 Zoom(缩放),这也是其名称得来的缘由。Pan 表示的是摄像机云台左右平扫操作;Tilt 表示的是摄像机云台上下俯仰操作;Zoom 表示的是摄像机镜头变倍、变焦控制。PTZ 跟踪的过程是,当目标对象出现后,Pan 和 Tilt 两个电机可以自动控制进行运动,并且根据场景中出现的目标对象的大小远近,驱动控制 Zoom 摄像机。PTZ 摄像机的拍摄角度和焦距可以根据目标运动特性进行主动改变,属于主动视觉系统,可以拍摄到环境中的不同区域;由于焦距可以主动改变,于是可以获得不同分辨率的目标图像

信息。近年来,摄像机的制作工艺逐渐成熟,PTZ 摄像机也跟着提升了控制精度,在云台旋转俯仰速度方面和变倍速度方面,都有了大幅度提升,同时降低了生产成本,这样 PTZ 摄像机在实际监控系统中广泛应用起来。目前市面上的“高速球机”是一种典型的高度集成 PTZ 摄像机,这种球机由光学变焦高清摄像机、控制云台和防护罩等组成,水平键控速度和垂直键控速度都能达到 $200^{\circ}/\text{s}$ 以上,变倍速度也能达到 2 s 以下,能够满足实时控制响应的要求,具有使用简单、工作稳定、成像清晰以及控制精准的优异特点,广泛应用在各种监控领域。

使用 PTZ 自动跟踪系统能够保证运动目标始终出现的视野的中部位置,在进行光学变焦之后也不会出现上面提到的目标移出监控视野范围的情况。这样就能够得到监控目标的高分辨率图像。当自动跟踪到感兴趣的目标(比如行人)之后,摄像头变焦拉近拍摄目标的高清晰的图像,在这样高分辨率的图像上进行目标特征提取将极大地提高特征的区分度。在这样的平台上使用人脸识别算法,高质量的图像保障了人脸识别算法的识别精度,能够真正应用到视频监控领域的实时布控场景,实现基于视频流的重点人员检测识别并实时报警。更进一步,智能视频监控领域中的复杂事件监测涉及的异常行为检测、人流检测、遗留物检测等,都要进行行人的识别和跟踪,这也是智能监控产品需要解决的首要问题。

在国外,很多研究学者将 PTZ 摄像机同固定摄像机组合成视频监控系统。通过主从控制进行协同工作,全局信息由固定摄像机获得,并进行目标跟踪,得到跟踪结果后再控制 PTZ 摄像机,以便得到高清的目标图像信息。这种方法对摄像机标定有很大程度的依赖。还有研究人员采用两个 PTZ 摄像机,组成双目视觉系统,这样能够三维重建现实场景,可以获得更多的目标信息。

上面提到的固定式摄像机一般为枪型摄像机,PTZ 摄像机一般为球机或者半球机。这种固定式摄像机和 PTZ 摄像机协同的工作机制可以称为枪球联动。枪球联动系统是双摄像头监控系统中的一种,摄像头采用一个广角枪机(焦距固定)和一个 PTZ 球机,其中广角枪机负责宽视野内的关注目标检测,PTZ 球机(又称主动摄像机)对关注目标进行聚焦放大和持续跟踪。

通过在每个监控点设置枪机和球机设备,通过两者之间的协同工作,一方面可以获得完整的、大视野的、边界画面清晰的全景图像;另外一方面,可以获取目标的特写图片并在全景画面中保持对当下目标的跟踪,见图 1-3 和图 1-4。当在全景图像上检测到用户设置的感兴趣区域中的运动目标或者检测到用户进入到某个球机的监控范围,即可控制球机设备跟踪全景中感兴趣区域中的运动目标并进行人脸、车牌等识别,灵活性高。在每个监控点设置枪机和球机,在初始设置时要求枪机监控范围的设置要覆盖球机的监控范围。

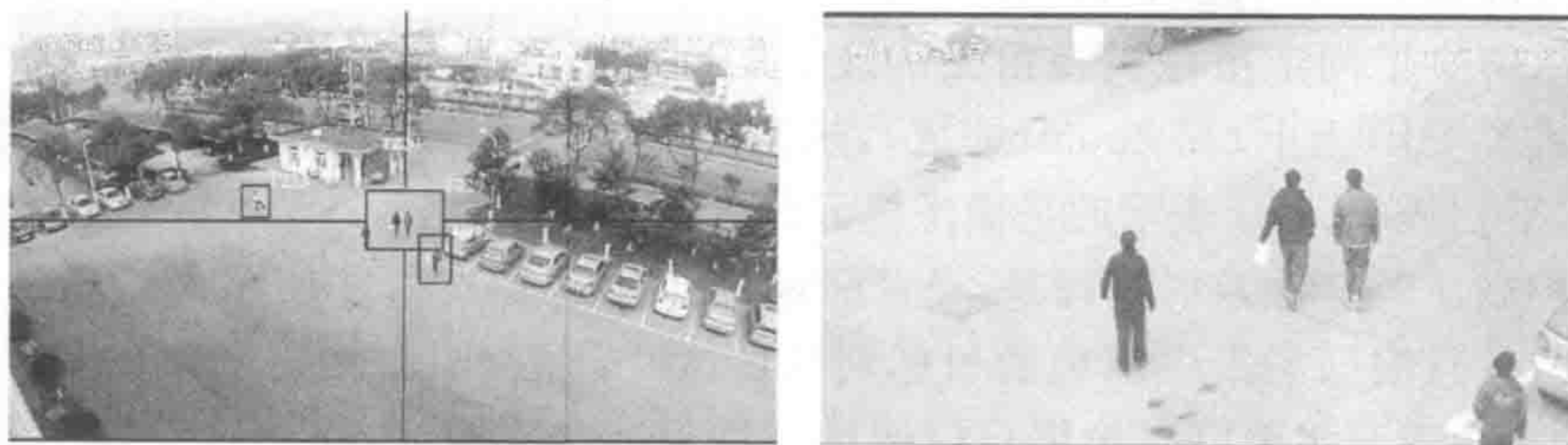


图 1-3 枪机(左图)和球机(右图)联动监控



图 1-4 夜视场景下的枪机和球机联动

市面上常见的枪球联动产品一般采用将枪机和球机放置在同一竖直面上,优点是视野的中心区域重合度高,中心区域两个摄像机画面映射的精度高,见图 1-5。

图 1-5 是枪球联动组合部署时候的上下部署和左右部署的对照图。

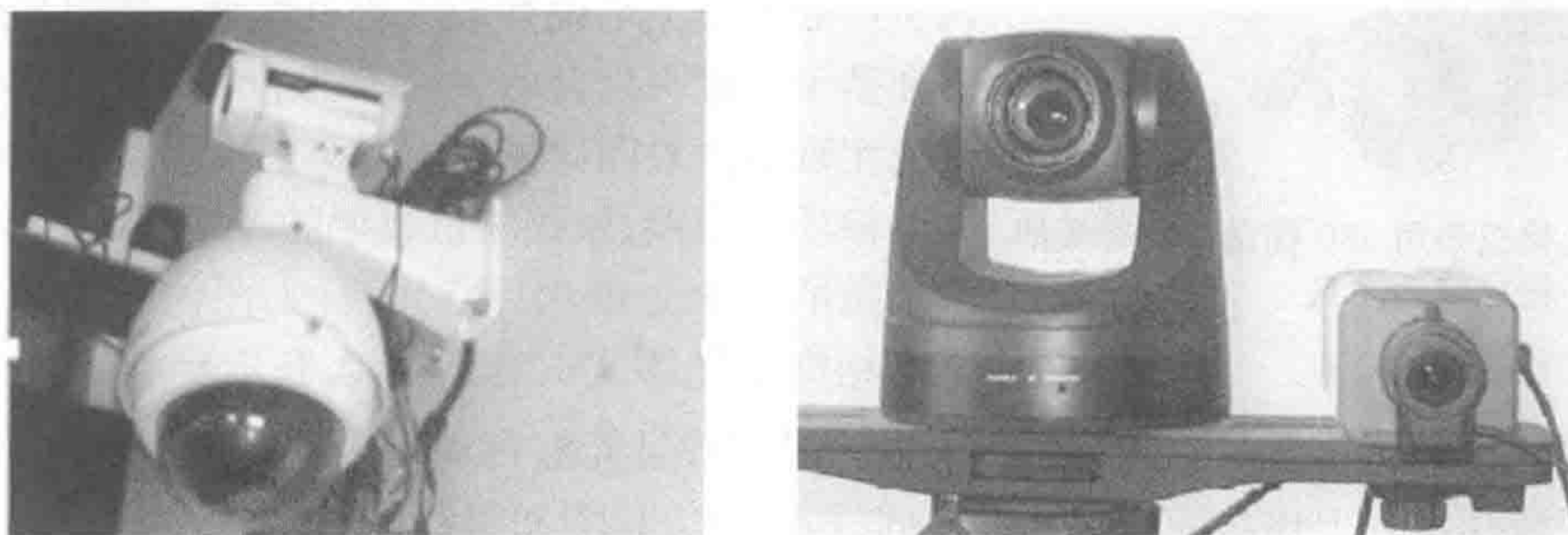


图 1-5 枪球联动产品两种形态

可以根据监控场景的实际情况进行枪机设备的部署。如果对宽大正面都有监控要求而单枪机视野不能完全覆盖,这种情况下可以采用枪机集群,通过多枪机图像的拼接,生成所需的全景图像,也可以采用鱼眼摄像机进行全景监控,见图 1-6。

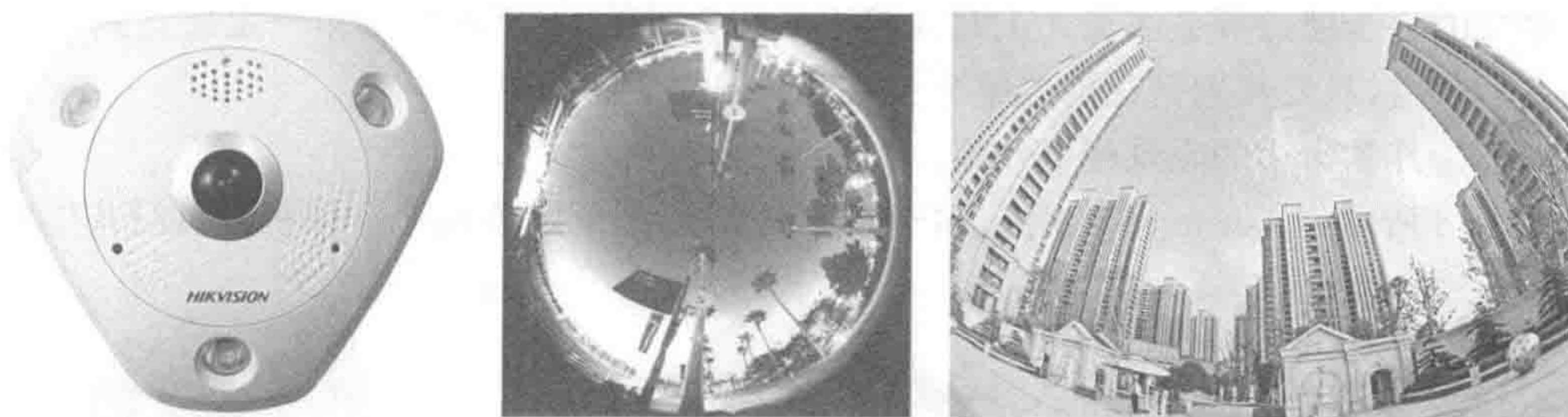


图 1-6 鱼眼摄像机及其成像效果

全景相机可以采用鱼镜头头或者通过多个枪机进行拼接来实现,在形态上有一体集成式的和分散式的。广角镜头在同样的距离内能够拍摄到较宽一些的物品景象,并且不会变形。鱼镜头头的视角更为广阔,能够达到甚至超过 180° 。但是,鱼镜头头拍摄出来的景象

周边会变形得很厉害。由于鱼眼镜头边缘畸变严重,对常规的基于形状、轮廓特征的目标检测算法的应用是一大挑战。室外监控实际应用中,从智能分析系统的后续对接方面考虑,更多的是采用广角镜头(视角不超过 180°)而不是鱼眼镜头。广角镜头哪怕是桶形畸变再明显,都好过鱼眼镜头。图 1-7 所示是采用广角镜头的枪球联动产品。



图 1-7 枪机采用 180° 视角广角镜头的枪球联动产品

图 1-7 中的枪机镜头焦距为 $f=1.8\text{ mm}$, 传感器为 $1/2.5''$, 视场角为 180° 的 500 万像素短焦镜头。球机镜头为 $f=4.7\sim 94\text{ mm}$, 视场角为 $55.4^\circ\sim 2.9^\circ$, 传感器为 $1/2.8''$ 的 200 万像素 CMOS。

进一步来讲,选择全景相机(“全景”并非指 360° 的视野拼接图,而指的是根据我们实际应用的需要,摄像机镜头选用符合一定要求的广角镜头,视野能够完全覆盖我们关注的较为宽大的正面区域,后文我们即以“全景相机”指代这种按需配置广角镜头的枪机或者枪机阵列)和跟踪相机相结合的 PTZ 跟踪系统。全景相机从产品形态上来说有采用超短焦镜头的单摄像机形态和基于多镜头组合的多摄像机组合形态。基于单摄像机实现的全景图像虽然成像范围大但是拍摄到的行人图像质量很差;跟踪相机成像的范围小,但能够进行高倍数的光学变焦,拍摄到的行人图像较为清晰。在全景相机中能够实时跟踪行人获取到行人的运动轨迹。

跟踪相机通过云台对感兴趣的行人进行跟踪并且能够自动对焦,拍摄到清晰的行人视频图像。成功实现上述功能涉及两方面的技术:行人检测和跟踪技术。

我们一般多采用常规镜头拼接方式生成全景图像。一体集成式的在结构上一般采用水平等角度排列,在成像焦段各个子相机的视野具有大小相等的重叠区域。分散式的安装相对自由,一般也要求相机高度相同,在同一水平面上成圆周排列,各个子相机具有重叠区域,各重叠区域的大小不一定相同,后期通过图像拼接算法实现全景图像的生成。这里,我们所说的全景画面,更为通常的是指几个相互之间具有重叠视野的水平布置枪机成像画面拼接生成的完全覆盖所关注宽大正面区域的宽视角图像,不要求视场角一定要达到 180° 的正前方全覆盖,只要覆盖到正前方整个关注区域的大扇面即可。

以下是几款常用的全景相机介绍。

图 1-8 所示为 Jaunt 全景相机,用于虚拟现实应用场景,实现全立体视觉沉浸式体验。

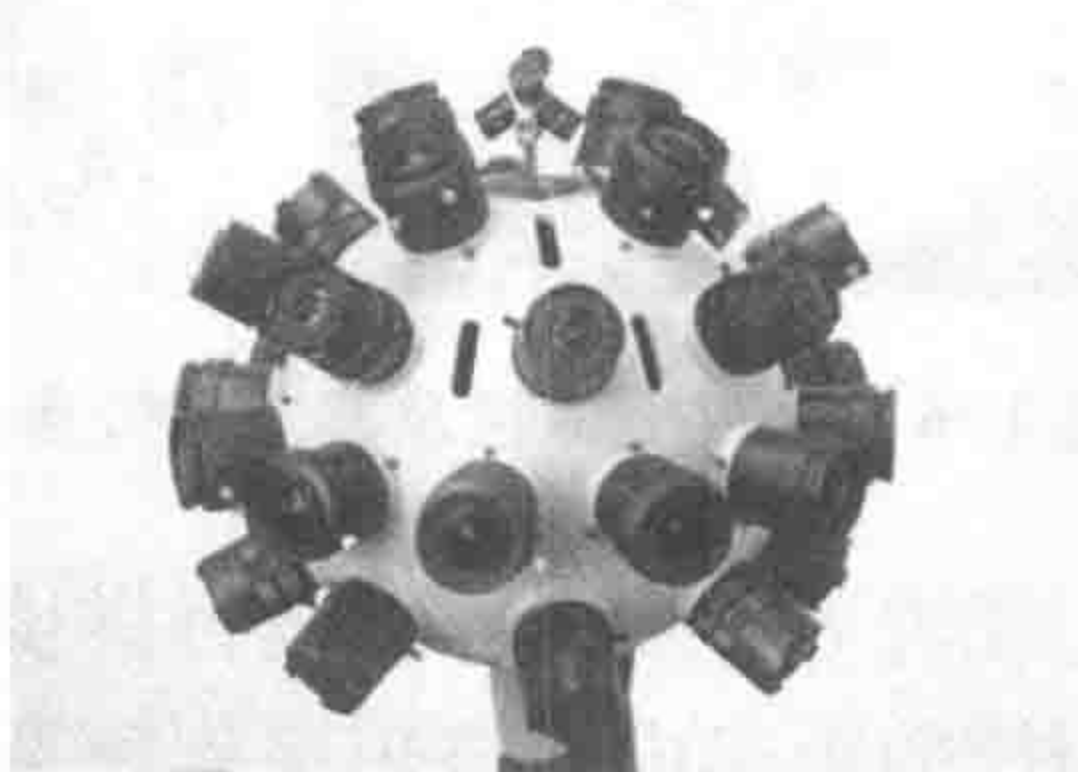


图 1-8 Jaunt 全景相机

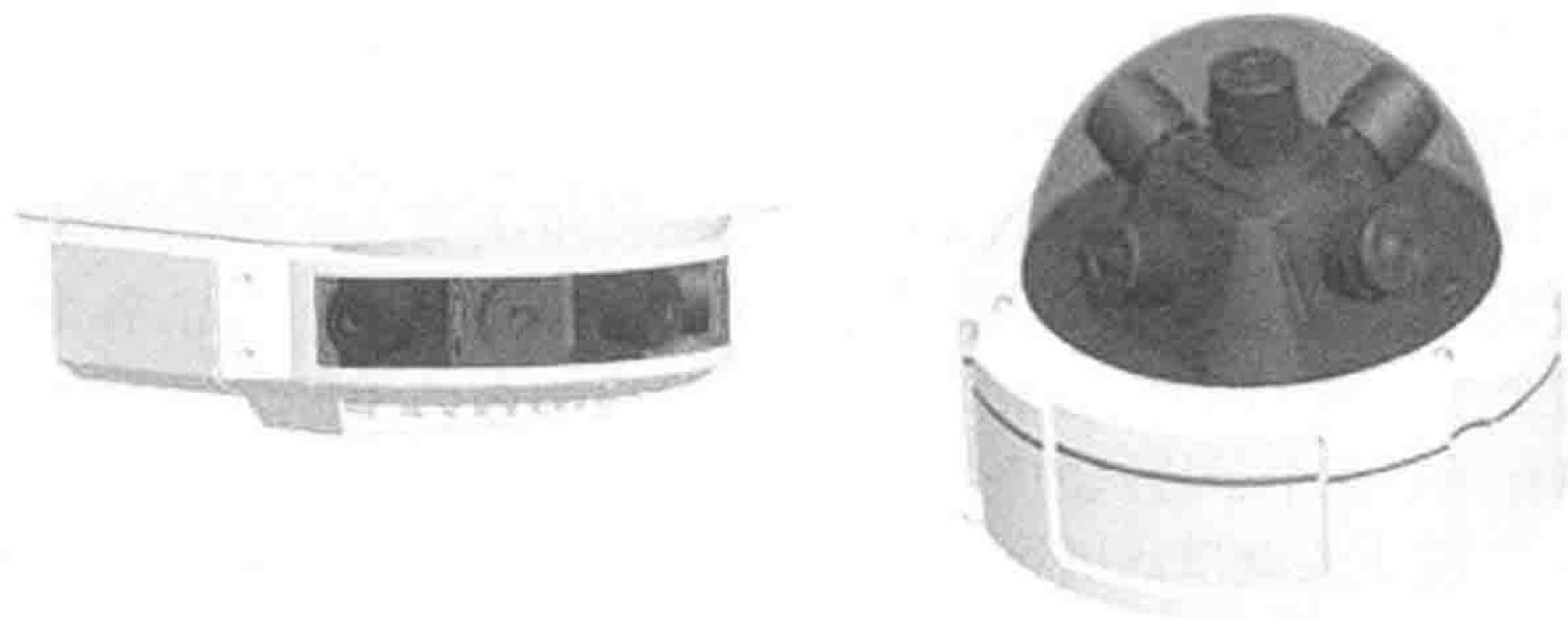


图 1-9 视频监控中的常用全景相机

图 1-9 中所示为视频监控系统常用的全景相机方案。左图是 4 个 500 万像素的枪机进行水平等角部署,生成 180° 的覆盖视角,画面等效像素为 2 000 万。右图通过 4 个 300 万像素定焦枪机等角分布形成全周覆盖监控,同时增加了一个向下的枪机,这种摄像机用于室内大厅部署,能够生成全景图片,又弥补了鱼眼相机的不足。

另外,有些采用单一超短焦距镜头的定焦摄像机,也能形成 $150^\circ \sim 180^\circ$ 的超宽视场角且具备前段实时畸变校正功能,产品形态及成像效果如图 1-10 所示。



图 1-10 180° 视场角相机及画面

PTZ 跟踪系统涉及图像处理技术、相机标定技术、视频跟踪以及自动控制技术等众多相关领域知识,有很强的专业技术综合性,其中跟踪算法和控制算法是系统的核心。跟踪算法通常是指检测跟踪的目标对象在图像中的位置信息,由控制算法通过目标对象在图像中的位置信息控制 PTZ 摄像机的高速步进电机转动,将目标锁定在图像的中央位置。

在现有的枪球联动的系统中,根据 PTZ 跟踪系统里面相机数目,分成单相机和多相机 PTZ 控制算法,对应的系统可以形象化地称为一枪一球联动系统和一枪多球联动系统,如图 1-11 中所示的三枪一球和一枪两球产品。在单相机控制算法中,通过目标在相机中的像素位置信息,计算出 PTZ 相机的两个电机旋转量参数,控制目标对象呈现在视野中央。双相机 PTZ 控制算法则是根据目标对象在固定相机中的位置信息来计算这两个旋转量,根据两个 PTZ 相机和目标的位置关系,对多目标的检测跟踪任务进行优化分配。概括来说,联动系统中的检测目标对象的位置信息的相机,同待控制角度的相机并不是同一个。

形态上说,一枪一球部署灵活机动,配置简单;一枪多球能够满足多目标实时跟踪,需要根据任务优先性设计多任务调度系统,多枪一球产品需要在全景画面拼接上满足实时和高精度要求。一枪一球和多枪一球更多应用于宽阔场所、人员目标较少的场景下,球机可