

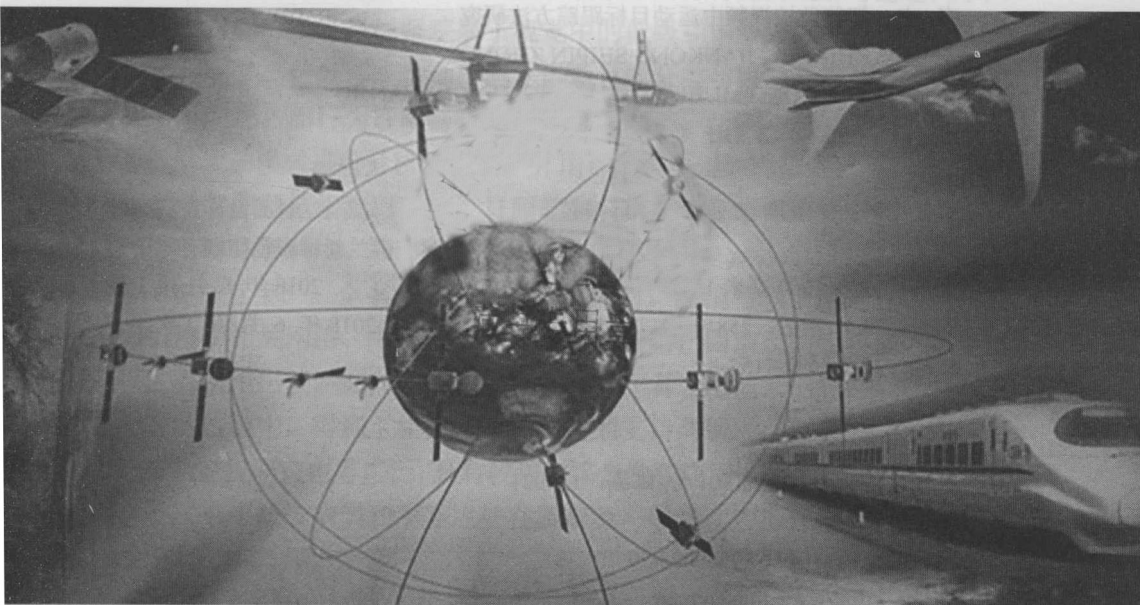
智能监控视频中 运动目标跟踪方法研究

冯莹莹 郭常山 ○ 著

吉林大学出版社

智能监控视频中 运动目标跟踪方法研究

冯莹莹 郭常山 著



吉林大学出版社
JILIN UNIVERSITY PRESS

图书在版编目 (CIP) 数据

智能监控视频中运动目标跟踪方法研究 / 冯莹莹, 郭常山著. -- 长春: 吉林大学出版社,
2017.10

ISBN 978-7-5692-1152-8

I. ①智… II. ①冯… ②郭… III. ①视频系统—监视控制—研究②音乐教育—教育模
式—研究—高等学校 IV. ①TN94

中国版本图书馆 CIP 数据核字 (2017) 第 265175 号

书 名: 智能监控视频中运动目标跟踪方法研究
ZHINENG JIANKONG SHIPIN ZHONG
YUNDONG MUBIAO GENZONG FANGFA YANJIU
作 者: 冯莹莹 郭常山 著

责任编辑: 陈颂琴 责任校对: 邢国春
吉林大学出版社出版、发行
开本: 710×1000 毫米 1/16
印张: 13.5 字数: 258 千字
ISBN 978-7-5692-1152-8

封面设计: 程娃娃
廊坊市兰新雅彩色印刷有限公司
2018 年 6 月第 1 版
2018 年 6 月第 1 次印刷
定价: 50.00 元

版权所有 翻印必究
社址: 长春市明德路 501 号 邮编: 130021

发行部电话: 0431-89580028/29
网址: <http://www.jlup.com.cn>
E-mail: jlup@mail.jlu.edu.cn

前言

智能视频监控是计算机视觉领域一个新兴的应用方向和备受关注的前沿课题。伴随网络技术和数字视频技术的飞速发展, 监控技术正向着智能化、网络化方向不断前进。因此计算机视觉和应用研究者适时提出了新一代监控—智能视频监控的概念。智能视频监控在不需要人为干预情况下, 利用计算机视觉和视频分析的方法对摄像机拍摄的图像序列进行自动分析, 实现对动态场景中目标的定位、识别和跟踪, 并在此基础上分析和判断目标的行为, 从而既能完成日常管理工作又能在异常情况发生时及时做出反应。智能视频监控系统不仅符合信息产业的未来发展趋势, 而且代表了监控行业的未来发展方向, 蕴藏着巨大的商机和经济效益, 受到学术界、产业界和管理部门的高度重视。

智能视频监控利用计算机视觉和图像处理的方法对图像序列进行运动检测、运动目标分类、运动目标跟踪, 以及对监视场景中目标行为的理解与描述。其中, 运动检测、目标分类、目标跟踪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处理。运动检测、运动目标分类与跟踪是视频监控中研究较多的三个问题; 而行为理解与描述则是近年来被广泛关注的研究热点, 它是指对目标的运动模式进行分析和识别, 并用自然语言等加以描述。

本书由阜阳师范学院冯莹莹, 日照职业技术学院郭常山撰写, 具体分工如下: 冯莹莹编写第 1、2、3、5、7、8 章节 (共 18 万字), 郭常山编写第 4、6、9 章节 (共 7 万字), 全书由郭常山统稿并校对。在编写本书的过程中, 我们查阅和引用了相关网络、书籍以及期刊等相关资料, 因涉及内容较多, 在这里不一一注明引用出处。谨向本书所引用资料的作者表示诚挚的感谢。此外, 本书在编写的过程中, 得到了相关专家和同行的支持与帮助, 在此一并致谢。由于智能视频监控研究内容广泛, 具有较强的综合性和应用性, 加之编者水平有限, 时间仓促, 书中缺点错误和不妥之处在所难免, 敬请读者批评指正, 以便今后进一步修改, 使之日臻完善。

第一节 行为建模	94
第二节 行为识别	96

目 录

第一章 概述	1
第一节 智能监控视频概述	2
第二节 智能监控视频的研究内容	6
第三节 智能监控视频的研究现状与应用前景	10
第四节 目标跟踪的分类	12
第五节 目标跟踪的应用	13
第二章 计算机运动视觉相关理论	41
第一节 摄像机的标定	42
第二节 双目立体视觉	47
第三节 运动视觉	55
第四节 场景理解	57
第三章 运动目标检测技术	63
第一节 运动目标检测概述	64
第二节 视频监控中的背景建模	74
第三节 ROI 面积缩减车辆检测搜索算法	79
第四章 运动目标行为理解技术	86
第一节 行为理解的特征选择与运动特征	88
第二节 场景分析	92
第三节 行为建模	94
第四节 行为识别	96

第五章 目标定位	104
第一节 概述	104
第二节 坐标系及坐标转换	106
第三节 定位误差的度量	111
第四节 空间定位方法及误差分析	112
第六章 运动目标跟踪技术	116
第一节 目标跟踪的分类	117
第二节 目标跟踪方法	118
第三节 粒子滤波器	120
第四节 多视角目标跟踪	124
第七章 目标跟踪基本方法	129
第一节 概述	129
第二节 坐标系的选择	131
第三节 数据预处理	133
第四节 目标运动模型	135
第五节 量测模型	138
第六节 基本的目标跟踪滤波方法	140
第八章 基于多摄像头协同的运动目标跟踪方法	143
第一节 多摄像头协同的运动目标跟踪相关技术	146
第二节 多摄像头协同的运动目标跟踪算法框架	149
第三节 系统的设计与实现	156
第四节 系统效果与评估	159
第九章 基于多传感器数据融合的目标跟踪方法	167
第一节 概述	167
第二节 时间与空间配准	175
第三节 数据关联	180
第四节 航迹及其融合	189

总结 201

参考文献 203

第一章 概述

计算机视觉是计算机科学与人工智能的一个重要分支，近 30 年来其研究水平有了突飞猛进的进展，而且这种发展势头还将保持一个相当长的时期。原因很清楚百闻不如一见，视觉信息在人类活动所涉及的各种信息中所占比重最大，由于其独特的空间特性和结构特性，不能为其他任何信息所代替。

从图像处理与模式识别发展起来的计算机视觉研究方法主要是研究如何利用二维投影图像恢复三维景物世界，计算机视觉使用的理论方法主要是基于几何、概率、运动学与三维重构的视觉计算理论，它的理论基础包括射影几何学、刚体运动学、概率与随机过程、图像处理、人工智能等。计算机视觉要达到的最终目的是实现计算机对三维景物世界的理解，即实现人类视觉系统的某些功能。

为了达到计算机视觉的目的，通常有两种技术途径可以利用。第一种是仿生学方法，即从分析人类视觉入手，利用大自然提供给我们的最好参考系——人类视觉系统，建立视觉过程的计算模型，然后利用计算机系统给予以实现。第二种是工程方法，即脱离人类视觉系统做文章，利用一切可行的和实用的技术手段实现视觉功能。由于仿生学方法的多重障碍，现在计算机视觉的研究主要采用工程方法。

虽然目前计算机视觉理论研究多，方法多，但已经得到了广泛的应用，如计算机人机交互、自动导航、医学、军事图像处理与可视化、生产自动化、工业检测与监控等领域。我们相信，随着认知神经科学、视觉计算理论等基础科学的发展，随着计算机性能的不断提高，以及各种应用场合对机器视觉的需求不断增长，计算机视觉将得到更广泛的应用，并将成为 21 世纪高科技产品中重要组成部分。

智能视频监控是计算机视觉领域一个新兴的应用方向和备受关注的的前沿课题。伴随网络技术和数字视频技术的飞速发展，监控技术正向着智能化、网络化方向不断前进。监控系统功能日益强大，但是依然需要工作人员不间断地分析监视场景内的活动，日夜值守，工作繁重。智能视频监控在不需



第一章 概述

计算机视觉是计算机科学与人工智能的一个重要分支,近 30 年来其研究水平有了突飞猛进的进展,而且这种发展势头还将保持一个相当长的时期。原因很清楚百闻不如一见,视觉信息在人类活动所涉及的各种信息中所占比重最大,由于其独特的空间特性和结构特性,不能为其他任何信息所代替。

从图像处理与模式识别发展起来的计算机视觉研究方法主要是研究如何利用二维投影图像恢复三维景物世界,计算机视觉使用的理论方法主要是基于几何、概率、运动学与三维重构的视觉计算理论,它的理论基础包括射影几何学、刚体运动力学、概率与随机过程、图像处理、人工智能等。计算机视觉要达到的最终目的是实现计算机对三维景物世界的理解,即实现人类视觉系统的某些功能。

为了达到计算机视觉的目的,通常有两种技术途径可以利用。第一种是仿生学方法,即从分析人类视觉入手,利用大自然提供给我们的最好参考系——人类视觉系统,建立视觉过程的计算模型,然后利用计算机系统给予以实现。第二种是工程方法,即脱离人类视觉系统框框的约束,利用一切可行的和实用的技术手段实现视觉功能。由于仿生学方法的进展缓慢,现在计算机视觉的研究大多使用工程方法。

虽然目前计算机视觉理论的发展不够成熟,但已经得到了广泛的应用,如计算机人机交互、自动导航、医学、三维场景建模与可视化、生产自动化、工业检测与监控等领域。我们相信,随着认知神经科学、视觉计算理论等基础科学的发展,随着计算机性能的不断提高,以及各种应用场合对机器视觉的需求不断增长,计算机视觉将得到更广泛的应用,并将成为 21 世纪高科技产品中重要组成部分。

智能视频监控是计算机视觉领域一个新兴的应用方向和备受关注的前沿课题。伴随网络技术和数字视频技术的飞速发展,监控技术正向着智能化、网络化方向不断前进。监控系统功能日益强大,但是依然需要工作人员不间断地分析监视场景内的活动,日夜值守,工作繁重。智能视频监控在不需



人为干预情况下，利用计算机视觉和视频分析的方法对摄像机拍摄的图像序列进行自动分析，实现对动态场景中目标的定位、识别和跟踪，并在此基础上分析和判断目标的行为，从而既能完成日常管理又能在异常情况发生时及时做出反应。

数据处理用于对采集到的视频信息进行调整、压缩和存盘，以方便该信息在网。

第一节 智能监控视频概述

一、智能视频监控的发展

智能视频监控是在传统的视频监控基础上发展起来的。传统的视频监控系统投入实际应用已经数十年，其技术发展也随之经历了三个时代。

第一代：模拟时代。在 20 世纪 90 年代初，视频监控主要是以模拟式磁带录像机为代表，其缺点是无法进行远程访问、无法与其他安防系统如门禁、边界防护等有效集成，信息存储方式给检索和查询带来诸多不便。

第二代：数字时代。在 20 世纪 90 年代中期，得益于数字视频压缩编码技术的发展，数字式视频录像机开始出现。DVR 的使用让用户可以将模拟的视频信号数字化，并存储在电脑硬盘而不是盒式录像带上。数字化的存储大大提高了用户对视频信息的处理能力。此外对于报警事件，以及事前事后报警信息的搜索也变得异常简单。

第三代：网络时代。进入 21 世纪后，随着网络技术的发展，DVR 系统又进一步发展成为网络数字视频录像机系统。与 DVR 系统相比，系统不但实现了视频信息的数字化存储，还实现了视频档案信息的数字化传播，即 NVR 可以直接接入到网络中，从而使存储下来的视频信息可以通过网络方便地进行共享。进一步，网络化视频监视系统，又称为 IP 监视系统（IPVS）开始出现。网络化视频监视系统从一开始就是针对在网络环境下使用而设计的，因此它克服了 DVR/NVR 无法通过网络获取视频信息的缺点，用户可以通过网络中的任何一台电脑来观看、录制和管理实时的视频信息。第三代视频监控系统是完全数字化的系统，它基于标准的 TCP/IP 协议，能够通过局域网、无线网和互联网传输，布控区域大大超过了前两代系统；它采用开放式架构，可与门禁、报警、巡更、语音、管理信息等系统无缝集成；基于嵌入式技术，性能稳定，无须专人管理；灵活性大大提高，监控场景可以实现任意组合，任意调用。



视频监控系统与设备虽然在功能和性能上得到了极大的提高,但是仍然受到了一些固有因素的限制,从而导致整个系统在安全性和实用性方面仍然没有达到人们期望。具体的制约因素如下:

(一) 人类自身的弱点

在很多情况下,人类并非是一个可以完全信赖的观察者,无论是观看实时视频流还是观看录像回放,由于自身生理上的弱点,人类经常无法察觉安全威胁,从而导致漏报现象的发生。

(二) 监控时间

除了一些规模较小的监控应用之外,很少有视频监控系统会按照 1:1 的比例为监控摄像机配置监视器。这意味着对于那些机场、港湾等大型的视频监控系统来讲,各个监控点并非每时每刻都处于监控当中。

(三) 误报和漏报

误报和漏报是目前视频监视系统中最常见的两大问题。漏报可能会导致非常危险的后果发生,而误报会浪费人力物力,并且这两种问题都会大大降低人们对监控系统的信任,从而降低监控系统的应用价值。

(四) 数据分析困难

报警发生后对录像数据进行分析通常是安全人员必须要做的工作之一,而误报和漏报现象则进一步加重了进行数据分析的工作负担。另外,安全人员经常被要求找出与报警事件相关的录像资料,找到肇事者、确定事故责任或评估该事件的安全威胁程度。但由于传统视频监控系统缺乏智能因素,录像数据无法被有效地分类存储更不用说其他智能分析,最多只能打上时间标签,因此数据分析工作变得极其耗时,并且很难获得全部的相关信息,而经常发生的误报漏报现象使得无用数据进一步增加,有用数据经常缺失,从而给数据分析工作带来了更大的困难。

(五) 响应时间长

由于安全威胁的响应速度关系到一个安全系统的整体性能。传统的视频监控系统通常都由安全工作人员对安全威胁做出响应和处理,这对于处理实时响应要求较低的安全威胁来说可能足够。但是很多情况下,在威胁发生时,需要安全系统的多个功能部分,甚至多个安全相关的部门在最短的时间内协调配合,共同处理危机。这时候,监控系统的响应速度将直接关系到用户的人身安全或财产的损失情况。

为了解决上述导致视频监控效率低下的问题,人们尝试把计算机视觉中的相关技术引入到视频监控中,从而发展起来的新型视频监控技术——智能视频监控,智能视频监控也称自动视频监控,它在视频监控中起着核心的作用,



可以有效提高视频监控的效率。同时它也是网络化视频监控领域最前沿的应用模式之一。

智能视频监控技术主要包括对视频图像序列自动地进行运动对象的提取、描述、跟踪、识别和行为分析等方面的内容。如果把摄像机看作人的眼睛，而智能视频系统或设备则可以看作人的大脑。智能视频监控技术就是借助计算机强大的数据处理功能，对视频画面中的海量数据进行高速分析，过滤掉监控者不关心的信息，仅仅为监控者提供有用的关键信息。智能视频监控以数字化、网络化视频监控为基础，但又有别于一般的网络化视频监控，它是一种更高端的视频监控应用。智能视频监控系统能够识别不同的物体，发现监控画面中的异常情况，并能够以最快和最佳的方式发出警报和提供有用信息，从而能够更加有效地协助安全人员处理危机，并最大限度地降低误报和漏报现象。其最终目的就是要使计算机能够分析、描述和理解视频画面中的内容。智能视频监控涉及计算机视觉、图像视频处理和人工智能领域中的众多核心技术，是一个非常具有挑战性的高难问题。

二、智能视频监控中的关键问题

智能视频监控的目标是让机器代替人来监视场景及场景中的目标，从而得到场景及目标安全性的评价。

智能视频监控的主要理论基础是计算机视觉理论，MITAI 实验室在 20 世纪 70 年代末提出了他的计算视觉表象理论，第一次把复杂、神秘的视觉过程变成一个可计算的信息处理过程，为机器视觉提供第一个较为完善的理论框架——视觉计算理论，它认为机器视觉的具体目标应该根据一个景物的一到几幅图像定量地、精确地决定场景中物体的形状、位置、物理特性，对景物进行 3D 重建，即将从图像推得形状信息的过程分成低层视觉、中层视觉和高层视觉三个表象阶段。

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 2.5 维描述，即部分地、不完整地描述，高层视觉从描述中得到以物体为中心的坐标系中完整的三维描述。

Marr 将视觉信息处理分为计算理论、表达和算法及硬件实现三个层次。计算理论层次回答计算的目的；策略表达和算法层次给出系统各部分的信息表达和实现系统各部分目标的具体算法；硬件层次则回答如何用硬件来实现以上算法。Marr 视觉理论总是从底层向高层推理，但实际上，人类在观察周围世界的时候，总是利用丰富的先验知识，这就启发人们恰当地引入先验知识，



提高系统的性能。目前智能视频监控系统在视觉的三个层次都不同程度地存在很多的问题,一方面要尽量解决这些问题,另一方面要应用先验的视觉知识来指导或者回避某些暂时无法克服的难题,只有这样才能逐步逼近完美的智能视频监控。

针对 W4 视频监控系统, Ismail Haritaoglu 和 Dvaid Hawrood 总结了视频监控需要解决的问题,即 W4 问题。所谓 W4 指的是:他们是谁?他们在干什么?何地?何时?

由于受理论及技术的限制,目前只能实现较低智能层次的视频监控。一般智能视频监控系统都是以运动跟踪为基础,对运动目标进行分类、识别、运动分析、行为理解等,从而得出对场景及目标安全性的判断。目前主要存在以下问题有待解决,这些问题也是目前计算机视频、视频监控的研究热点。

快速准确的运动检测、跟踪技术及实时、鲁棒的目标分类、识别技术。这些技术是智能视频监控的基础,关于这方面的研究开展得最早,但仍然存在很多问题,特别是多目标分割及运动跟踪中存在的遮挡等问题严重影响系统的性能。

基于移动摄像机的视频监控技术,即将现有参数固定的静态摄像机改进为参数可在线调节的动态摄像机,包括摄像机外部参数调整及内部参数调整,这样扩大监视范围,可以实现对运动目标的主动跟踪。如何根据目标的运动来实时控制摄像机的运动,并对摄像机的内外参数进行在线标定是非常重要的问题。

多摄像机协作监控技术。单个摄像机的视野有限,要监控大范围的动态场景就需要多个摄像机,此外,多个摄像机也有利于解决遮挡问题。其技术难点是多摄像机协作、多摄像机的定标及数据融合问题。

运动分析及行为理解技术对目标的整体运动和局部运动进行跟踪、分析,获得目标的轮廓、姿态、行为、路径等信息,从而进一步理解目标的行为。关键是要用机器学习方法,对序列图像进行自组织、自学习的训练,建立行为模式,然后再利用该模式来分析、识别、理解场景中发生的行为。

异常事件的检测、报警与未来事件预测技术。视频监控系统的最终目的是为了解释监视场景中所发生的事件,根据要求对异常事件进行报警,并能根据当前目标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,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、数据理论、模糊理论等来预测未来某事件发生的可能性或不确定性。

非接触式身份识别技术。生物特征识别技术与人的运动分析的结合是视

频监控系统的一个重要问题, 脸像与步态是具有可感知性与非接触性优点的生物特征, 是目前被认为可以用于视频监控系统中的身份识别的两个主要生物特征, 也是目前智能视频监控系统能体现出其智能性的功能。一般近距离采用人脸识别, 而远距离采用步态识别, 前者的研究较为深入, 已经有较为成熟的专用人脸识别产品, 但视频监控中, 视频人脸的被动识别仍然存在很多问题, 如光线、分辨率、姿态等对识别率的影响较大。步态识别仍处于探索阶段, 其难题在于远距离、多角度步态特征的提取困难, 并且难以解决小训练集与大测试集之间的矛盾。

第二节 智能监控视频的研究内容

各种视频监控系统的结构、实现方式等均有所不同, 为了能够进一步从研究的角度直观说明智能视频监控系统的主要任务, 这里给出了智能视频监控系统的结构框图。

图像获取、数据处理、图像处理、络中传输。硬件系统获取的原始图像, 由于噪声、光照、运动等原因, 图像质量不高, 所以需要进行预处理, 以利于提取我们感兴趣的信息。预处理模块主要包括传感器标定、滤波、图像增强与恢复等。

图像分析模块主要是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。图像分析模块包括目标分割、目标定位、目标跟踪和特征提取。在这部分主要是解决“何时?”和“何地?”的问题。

图像理解是在图像分析的基础上, 进一步研究图像中各目标的性质和它们之间的相互联系, 并得出对图像内容含义的理解以及对客观场景的解释, 从而指导和规划行动。内容包括目标识别、行为理解、威胁估计和决策推理四部分。其中, 目标识别是解决“我看到了什么?”的问题; 行为理解和威胁估计用来解决“他们正在做什么?”和“他们将会做什么?”; 最后是通过决策推理解决“我应该采取什么措施?”

智能视频监控的研究内容包括运动目标检测、运动目标分类、运动目标跟踪以及对监视场景中目标行为的理解与描述。其中, 运动检测、目标分类、目标跟踪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处理。运动检测、运动目标分类与跟踪是视频监控中研究较多的三个问题; 而行为理解与描述则是近年来被广泛关注的研究热点, 它是指对目标的运动模式进行分析和识别, 并用自然语言等加以描述。

运动检测的目的是从序列图像中将变化区域从背景图像中提取出来。运



动区域的有效分割对于目标分类、跟踪和行为理解等后期处理是非常重要的,因为以后的处理过程仅仅考虑图像中对应于运动区域的像素。然而,由于背景图像的动态变化,如天气、光照、影子及混乱干扰等的影响,使得运动检测成为一项相当困难的工作。目前几种常用的方法有:背景减除、时间差分、光流、扩展的EM算法、能量运动检测、基于数学形态学的场景变化检测等。

目标分类的目的是识别运动目标所属的类别。不同的运动区域可能对应于不同的运动目标,比如交通道路上监控摄像机所捕捉的序列图像中可能包含行人、车辆及其他运动物体,为了便于进一步对行人进行跟踪和行为分析,运动目标的正确分类是完全必要的。目前常用的目标分类方法有:基于形状信息的分类、基于运动特性的分类以及时间共生矩阵进行分层分类的方法等。

传统上目标跟踪表达为一个组合:测量——状态分配值问题。目标跟踪等价于在连续的图像帧间创建基于位置、速度、形状、纹理、色彩等有关特征的对应匹配问题,常用的数学工具有卡尔曼滤波、Condensatino 算法、动态贝叶斯网络等。在这些方法中,卡尔曼滤波是基于高斯分布的状态预测方法,不能有效处理多峰模式分布情况;Condensatino 算法是以因子抽样为基础的密度传播方法,结合可学习的动态模型,能够完成鲁棒的运动跟踪。目标跟踪的方法主要有基于模型的跟踪、基于区域的跟踪、基于活动范围的跟踪和基于特征的跟踪。

运动检测、目标分类与跟踪是运动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被广泛关注的研究热点。它是指对运动模式进行分析和识别,并用自然语言等加以描述。行为理解可以简单地被认为是时变数据的分类问题,即将测试序列与预先标定的代表典型行为的参考序列进行匹配。由此可见,行为理解的关键问题是如何从学习样本中获取参考行为序列,并且学习和匹配的行为序列必须能够处理在相似的运动模式类别中空间和时间尺度上轻微的特征变化。

尽管智能视频监控研究已经取得了一定的成果,但仍有下述几个关键技术有待解决和突破,也是今后研究的难点问题。

一、运动分割

快速准确的运动分割是相当重要,并且困难的问题。这是由于动态环境中图像受到多方面的影响,比如天气变化、光照变化、背景混乱干扰、运动目标的影子、物体与环境之间或者物体与物体之间的遮挡,以及摄像机的运动等。这些都给准确有效的运动分割带来了困难。以运动目标的影子为例,它可能与被检测的目标相连,也可能与目标分离。在前者情况下,影子扭曲



了目标的形状,从而使得以后基于形状的认识方法不再可靠;在后者情况下,影子有可能被误认为是场景中的目标。尽管目前图像运动分割主要利用背景减除方法,但如何建立对于任何复杂环境的动态变化均具有自适应性的背景模型仍是相当困难的问题。一个可喜的发展是一些研究者们正利用时空统计的方法构建自适应的背景模型。对于不受限环境中的运动分割而言,这也许是个更好的选择。

二、遮挡处理

目前,大部分运动分析系统都不能很好地解决目标之间互遮挡和人体自遮挡问题,尤其是在拥挤状态下,多目标检测和跟踪问题更是难于处理。在发生遮挡时,目标只有部分是可见的,而且这个过程一般是不可训练的。此时简单依赖于背景减除进行运动分割的技术将不再可靠。为了减少遮挡或深度所带来的歧义性问题,必须开发更好的模型来处理遮挡时特征与目标部分之间的准确对应问题。另外,遮挡前后的跟踪初始化也缺少自举方法。目前比较有效的方法是利用统计方法从可获得的图像信息中进行目标姿势、位置等的预测。不过,对于解决遮挡问题最有实际意义的潜在方法应该是基于多摄像机的跟踪系统。

三、三维建模与跟踪

二维跟踪方法在早期的运动分析中被证明是很成功的,尤其对于那些不需要精确的姿势恢复或图像低分辨率的应用场合。二维跟踪有着简单快速的优点,主要的缺点是受摄像机角度的限制。而三维方法在不受限的复杂的运动判断(如人的徘徊、握手与跳舞等)、更加准确的物理空间的表达、遮挡的准确预测和处理等方面的优点是二维方法所不能比拟的;它能提供更加有意义的与身体姿势直接相关的可视化特征应用于行为识别;同时,三维恢复对于虚拟现实中的应用也是必需的。目前基于视觉的三维跟踪研究仍相当有限,三维姿势恢复的实例也很少,且大部分系统由于要求鲁棒性而引入了简化的约束条件。三维跟踪也导致了从图像中目标模型的获取、遮挡处理、参数化建模、摄像机的标定等一系列难题。以建模为例,模型通常使用许多形状参数表达。过去的一些工作几乎都假设3D模型依据先验条件而提前被指定,实际上这些形状参数应当从图像中估计出来。总之,3D建模与跟踪在未来工作中应值得更多的关注。

四、多摄像机跟踪

使用单一摄像机的三维跟踪研究很少,因为目标姿势及其运动在单一视



角下由于遮挡或深度影响而容易产生歧义现象,因此使用多摄像机进行三维姿势跟踪和恢复的优点是很明显的。同时,多摄像机的使用不仅可以扩大监视的有效范围,而且可以提供多个不同的方向视角以用于解决遮挡问题。很明显,未来的运动分析系统将极大受益于多摄像机的使用。对于多摄像机跟踪系统而言,我们需要确定在每个时刻使用哪一个摄像机或哪一幅图像。也就是说,多摄像机之间的选择和信息融合是一个比较关键的问题。

五、场景理解

如何根据场景的信息结合目标检测与跟踪,提高准确率。特别是在目标缺损等情况下,视觉注意被认为是减少视觉信息,从而更容易处理大小的一种方式,以便于人脑在有限计算资源条件下能够处理。而将人类视觉系统基础研究成果应用于计算机视觉、场景分析和目标识别一直是计算机科学家追求的目标,近几年来视觉系统选择注意机制研究成果已引起了计算机视觉界的关注,并提出了几个相关的计算模型:J.K.Tsotsos 等在被关注的位置上使用局部胜者为王和自上而下的机制来选择性地调整模型神经元;G.Deco 等基于自上而下地注意控制信号,调整图像的空间分辨率;基于目标的视觉注意,比如 Duncan 的集成竞争理论,视觉注意的自下而上和自上而下交互作用;基于目标和基于位置注意的集成;按照并扩展 Duncan 的集成竞争假设,Y.Sun 和 R.Fisher 使用“分组”,研究和实现了基于目标和基于位置视觉注意的一个通用计算框架;L.Itti 等基于连续地扫描一个显著图而引入了自下而上基于特征显著性的空间选择注意模型,但该模型忽略了自上而下的、任务依赖对注意集中的影响,不关心目标识别,而是集中于发现感兴趣的特征,并只给出了感兴趣目标的位置,而不能提供目标的轮廓。生物视觉是通过指导对选择的场景区域凝视即基于动态选择感兴趣区域(ROI)的,MartinClaus 等利用 Itti 和 Koch 提出的基于显著性模型,分析研究了在图像中选择显著性位置的注意计算机制,并提出了利用独立排序度量、Stringed it distance 度量等对感兴趣区域(ROI)进行量度。国内在计算机视觉领域研究视觉选择注意及其在目标识别中的应用还很少,才刚刚开始,也主要集中于基于显著性的目标识别方面的应用研究。

六、性能评估

一般而言,鲁棒性、准确度和速度是运动分析系统的三个基本要求。例如,系统的鲁棒性对于监控应用特别重要,这是因为它们通常被要求是自动、连续地工作,因此这些系统对于如噪声、光照、天气等因素的影响不能太敏感;



系统的准确度对于控制应用特别重要,例如基于行为或姿势识别的接口控制场合;而系统的处理速度对于那些需要实时高速的监控系统而言更是非常关键。因此,如何选择有效的工作方案来提高系统性能、降低计算代价是特别值得考虑的问题。同时,如何利用来自不同用户、不同环境、不同实验条件的大量数据,测试系统的实时性、鲁棒性亦相当重要。

第三节 智能监控视频的研究现状与应用前景

近些年来,随着计算机软硬件性能的不断发展,各种面向复杂应用背景的智能视频监控系统也随之大量涌现,智能视频监控系统在商业、国防安全和军事应用领域中的需求日益增加。正是由于智能视频监控系统具有如此大的应用前景,引起许多国家的高度重视,并投入大量资金和科技人员进行了广泛研究。

在1996年至1999年间,在美国国防高级研究计划局(DARPA)资助下,卡内基梅隆大学、戴维 SARNOFF 研究中心等几家著名研究机构合作,联合研制了视频监视与监控系统。VSAM 的目标是为未来城市和战场监控应用开发的一种自动视频理解技术,用于实现未来战争中人力监控费用昂贵、非常危险或者人力无法实现等场合下的监控。

该系统目前仍处于试用阶段,其主要功能有:

具有先进的视频分析处理器,不但能检测和识别异常对象的类型,还能分析与预测人的活动,根据运动对象行为的危害性进行自动提示和报警。

使用地理信息和三维建模技术提供可视化图形操作界面。当视频分析处理器报告了运动对象、对象类别及位置之后,操作员不仅可以使虚拟的对象(人、汽车、坦克等)在地理信息界面上进行标记,而且还能在辅助窗口观察对象的真实活动情况。

机载航空摄像机不需要经常性的人工操纵就能自动对准地面监视目标,实现对重要目标的长时间监视;自动协调多个图像传感器无缝接入,实现整个战场场景的监视。因此,VSAM 不但能进行一般性的军事安全监控,如军事基地、军械弹药库和边海防线的监控;而且能够进行局部战争战场的实时监控,如敌方军力部署及调动情况等。DARPA 在2000年又资助了远程人类识别重大项目,研究开发多模式的监控技术以实现远距离情况下人的检测、分类和识别,以增强国防、民用等场合免受恐怖袭击的保护能力。

欧盟长期研究项目(EULTR)资助比利时 Katholieke 大学的电子工程系、法国国家计算机科学和控制研究院等欧洲著名的大学和科研机构联合研究为