

信息科学技术学术著作丛书

视频图像运动目标 检测技术

李艾华 崔智高 姜柯 苏延召 著



科学出版社

信息科学技术学术著作丛书

视频图像运动目标检测技术

李艾华 崔智高 姜 柯 苏延召 著

科学出版社

北 京

内 容 简 介

运动目标检测是智能视频监控的一个重要而基础的研究内容。本书围绕运动目标检测技术进行讨论,主要包括运动目标检测的介绍(第1章)、静态背景下的运动目标检测技术(第2、3章)、动态背景下的运动目标检测技术(第4章)、运动阴影消除技术(第5、6章),以及运动目标检测和阴影消除的并行加速(第7章),介绍了相关方法的研究背景、理论基础和算法描述,并给出了相应的实验结果。

本书可作为模式识别、图像处理和机器视觉等领域高年级本科生、研究生的教材,也可作为相关领域的教师、科研人员,以及从事安防和视频监控行业的工程技术人员的参考学习用书。

图书在版编目(CIP)数据

视频图像运动目标检测技术/李艾华等著. —北京:科学出版社,2016
(信息科学技术学术著作丛书)

ISBN 978-7-03-050618-4

I. 视… II. 李… III. 视频-图像处理-应用-运动目标检测-研究
IV. TP72

中国版本图书馆 CIP 数据核字(2016)第 271838 号

责任编辑:魏英杰 / 责任校对:桂伟利

责任印制:张 伟 / 封面设计:陈 敬

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

*

2017 年 1 月第 一 版 开本:720×1000 1/16

2017 年 1 月第一次印刷 印张:8 3/4

字数:178 000

定价:80.00 元

(如有印装质量问题,我社负责调换)

《信息科学技术学术著作丛书》序

21 世纪是信息科学技术发生深刻变革的时代,一场以网络科学、高性能计算和仿真、智能科学、计算思维为特征的信息科学革命正在兴起。信息科学技术正在逐步融入各个应用领域并与生物、纳米、认知等交织在一起,悄然改变着我们的生活方式。信息科学技术已经成为人类社会进步过程中发展最快、交叉渗透性最强、应用面最广的关键技术。

如何进一步推动我国信息科学技术的研究与发展;如何将信息技术发展的新理论、新方法与研究成果转化为社会发展的新动力;如何抓住信息技术深刻发展变革的机遇,提升我国自主创新和可持续发展的能力?这些问题的解答都离不开我国科技工作者和工程技术人员的求索和艰辛付出。为这些科技工作者和工程技术人员提供一个良好的出版环境和平台,将这些科技成就迅速转化为智力成果,将对我国信息科学技术的发展起到重要的推动作用。

《信息科学技术学术著作丛书》是科学出版社在广泛征求专家意见的基础上,经过长期考察、反复论证之后组织出版的。这套丛书旨在传播网络科学和未来网络技术,微电子、光电子和量子信息技术、超级计算机、软件和信息存储技术,数据知识化和基于知识处理的未来信息服务业,低成本信息化和用信息技术提升传统产业,智能与认知科学、生物信息学、社会信息学等前沿交叉科学,信息科学基础理论,信息安全等几个未来信息科学技术重点发展领域的优秀科研成果。丛书力争起点高、内容新、导向性强,具有一定的原创性;体现出科学出版社“高层次、高质量、高水平”的特色和“严肃、严密、严格”的优良作风。

希望这套丛书的出版,能为我国信息科学技术的发展、创新和突破带来一些启迪和帮助。同时,欢迎广大读者提出好的建议,以促进和完善丛书的出版工作。

中国工程院院士
原中国科学院计算技术研究所所长



前 言

随着世界范围内安全意识的提高,基于视觉的智能监控技术受到了广泛重视,并且逐渐应用到各种军事和民用场合,因此智能视频监控的研究具有重要意义。智能视频监控是指以监控摄像机作为前端采集设备,利用计算机视觉的相关技术,对采集的视频序列进行智能分析,如运动目标的检测与跟踪、目标的分类与识别、行为分析和异常检测等,从而辅助管理人员获取监控场景中的有用信息,并根据具体需要发出预警信号或实施相关决策。

在智能视频监控的相关研究中,运动目标检测是一项基础性、关键性的研究内容。特别是近二十年来,由于计算机和摄像机性价比的大幅提高,以及目标检测在视频监控、人机交互、视频编码、医疗诊断等领域不断提高的广泛应用价值,运动目标检测逐渐成为图像处理和机器视觉领域的研究热点。然而,由于监控场景的复杂性,采集的视频序列一般具有动态背景(树叶晃动、摄像机运动)、光照变化、运动阴影、图像噪声等干扰,因此运动目标检测是图像处理和机器视觉领域中一个公认的具有挑战性的问题,其研究需要图像处理、模式识别、矩阵理论、概率论和随机过程等多学科的知识。

本书紧紧围绕运动目标检测技术展开讨论。第1章绪论,介绍运动目标检测技术的基本概念、分类和研究现状。全书剩余6章,从内容来看可以分为四个部分。第一部分是静态背景下的运动目标检测,包括第2、3章,分别介绍基于隶属度竞争的空间混合高斯模型和基于随机策略的码本模型;第二部分是针对动态背景下运动目标检测的第4章,介绍基于多组单应约束和马尔可夫随机场的运动目标检测;第三部分是运动阴影去除,包括第5、6章,介绍基于多特征融合和直方图反投影的运动阴影去除,以及基于全局纹理和统计推断的运动阴影去除;第四部分包括第7章,主要研究CUDA架构下运动目标检测和运动阴

影去除的 GPU 加速策略。

本书由李艾华教授审定大纲,撰写第 1 章,并对全书进行统稿、修改和定稿;崔智高执笔第 2~4 章;苏延召和姜柯分别执笔第 5 章和第 6、7 章。本书的出版得到火箭军工程大学机关、机电教研室的支持和帮助,在此一并表示感谢。另外,感谢作者所在研究组的徐斌、姚良、李辉、蔡艳平、金广智、李庆辉、马鹏程、方浩、王华、袁梦等同事和同学,他们为本书提供了很多有价值的素材,并协助完成了审读、校对和修改等工作。

视频图像运动目标检测是尚在发展中的新技术,同时限于作者水平,难免存在不妥之处,恳请读者指正。

作 者

2016 年 5 月于西安

目 录

《信息科学技术学术著作丛书》序

前言

第 1 章 绪论	1
1.1 智能视频监控技术	1
1.2 运动目标检测技术	5
1.2.1 静态背景下的运动目标检测技术	6
1.2.2 动态背景下的运动目标检测技术	8
1.2.3 运动阴影消除技术	11
1.3 本书内容安排	12
参考文献	14
第 2 章 采用改进混合高斯模型的运动目标检测	19
2.1 引言	19
2.2 混合高斯模型原理	20
2.3 混合高斯背景模型的性能提升	23
2.3.1 基于偏差均值的匹配判断	23
2.3.2 基于非线性学习速率的权重更新	24
2.3.3 低权重模型的移除	25
2.3.4 模型等权值初始化方法	26
2.3.5 实验结果及分析	26
2.4 基于隶属度竞争的空间混合高斯模型	29
2.4.1 隶属度的计算	30
2.4.2 前景检测	31
2.4.3 模型更新	33
2.4.4 实验结果及分析	34
参考文献	38
第 3 章 基于随机码本的运动目标检测	40
3.1 引言	40

3.2 码本背景模型原理	41
3.2.1 码本结构定义	41
3.2.2 码字匹配	41
3.2.3 码本训练	43
3.2.4 前景检测	44
3.2.5 模型更新	44
3.2.6 码本模型不足	45
3.3 视觉背景提取模型原理	45
3.4 基于随机码本的运动目标检测	47
3.4.1 基于 YUV 空间的码本模型	48
3.4.2 基于空间邻域像素点的码本训练	48
3.4.3 基于随机思想的码本更新	50
3.5 实验结果及分析	51
3.5.1 “鬼影”抑制实验	52
3.5.2 综合性能对比实验	53
参考文献	54
第4章 采用多组单应约束的运动目标检测	56
4.1 引言	56
4.2 基于多组单应约束的前背景轨迹分离	58
4.2.1 邻近帧背景运动的单应模型	58
4.2.2 前背景轨迹分离方法	59
4.3 基于马尔可夫随机场的前背景像素标记	62
4.3.1 能量函数的构造	62
4.3.2 惩罚函数的设计	63
4.4 实验结果及分析	65
4.4.1 轨迹分离实验	66
4.4.2 像素标记实验	68
参考文献	72
第5章 基于多特征融合和直方图反投影的运动阴影去除	75
5.1 引言	75
5.2 阴影产生机理及其光照模型	76

5.2.1 阴影产生机理	76
5.2.2 光照模型	78
5.3 阴影检测常用特征和多特征融合	78
5.3.1 阴影检测常用特征	78
5.3.2 阴影检测常用特征对比	80
5.3.3 多特征融合方式	81
5.4 基于多特征融合和直方图反投影的阴影检测	82
5.4.1 运动阴影初选	83
5.4.2 多特征联合直方图	84
5.4.3 直方图反投影	86
5.4.4 自适应阈值分割	87
5.5 实验结果及分析	88
5.5.1 评价指标与测试数据集	88
5.5.2 实验结果	89
参考文献	93
第 6 章 基于全局纹理和统计推断的运动阴影去除	97
6.1 引言	97
6.2 基于 YUV 分量变化比率的阴影检测	98
6.2.1 基于 YUV 色彩空间的阴影检测模型	98
6.2.2 阴影检测模型的可行性分析	98
6.3 阴影检测阈值的自适应估计	100
6.3.1 抽样样本及其推断	101
6.3.2 全局边缘纹理构造	102
6.3.3 基于抽样推断的阈值区间估计	102
6.4 实验结果及分析	104
6.4.1 不同边缘检测算子对阈值估计的影响	104
6.4.2 手动阈值与估计阈值的对比	104
6.4.3 阴影检测性能对比	104
参考文献	108
第 7 章 采用 GPU 并行加速的运动目标检测与阴影去除	110
7.1 引言	110

7.2	CUDA 并行计算架构	111
7.2.1	CUDA 编程模型	112
7.2.2	多处理器结构及内部资源	115
7.2.3	多级存储器模型	117
7.2.4	CUDA 程序性能优化策略	119
7.3	参数自整定的并行粒度划分算法	120
7.3.1	不同硬件之间的资源和性能差异	120
7.3.2	并行粒度划分参数对计算性能的影响	121
7.3.3	并行粒度划分参数自整定	122
7.3.4	图像畸变校正的并行加速实验	124
7.4	随机码本模型和阴影检测算法的 GPU 加速实验	125
7.4.1	随机码本模型的 GPU 加速	125
7.4.2	阴影检测算法的 GPU 加速	127
	参考文献	128

第 1 章 绪 论

1.1 智能视频监控技术

智能视频分析技术是计算机视觉的一个重要分支,其通过将场景中背景和目标分离,进而分析和追踪在摄像机场景内出现的目标,并对目标进行分类识别、行为分析和异常报警。智能视频监控(intelligent video surveillance, IVS)则是利用智能视频分析技术对视频监控数据进行处理、分析和理解,协助视频监控系统实施控制和跟踪,并对监控目标进行抽象描述和行为分析,从而提高视频监控系统的智能化、自动化水平。智能视频监控系统能够及时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并提供有用信息,从而更加有效地协助管理人员处理危机,最大限度地降低误报和漏报等现象。

智能视频监控技术是在视频监控技术的基础上发展起来的。随着技术的发展,目前视频监控技术已经历了三个发展阶段。

(1) 第一代视频监控系统

第一代视频监控系统是以磁带录像机(video cassette recorders, VCR)为代表的传统 CCTV 系统,主要由模拟摄像机、专用电缆、视频切换矩阵、模拟监视器、模拟录像设备和盒式录像带等构成,如图 1-1 所示。第一代系统存在许多明显的缺点,如维护工作繁琐、无法进行远程访问、无法与其他安防系统(如门禁、周界防护等)有效集成、录像质量随着时间推移下降等。尽管如此,在数字技术和网络技术并不发达的年代,用户仍然非常广泛的采用这类系统。

(2) 第二代视频监控系统

20 世纪 90 年代中期,出现了以数字录像机(digital video recorder, DVR)为代表的第二代视频监控系统,如图 1-2(a)所示。DVR 可以将模拟视频信号进行数字化,并存储在硬盘而不是盒式录像带上。数字

化存储大大提高了用户对录像信息的处理能力,而且可以实现移动侦测功能,此外对于报警事件及报警信息的搜索也更为简单。进入 21 世纪,随着网络技术的发展,DVR 系统又进一步发展成为具有网络功能的 NVR(network video recorder)系统,如图 1-2(b)所示。与 DVR 系统相比,NVR 系统不但实现了视频信息的数字化存储,而且实现了视频信息的数字化传播,即 NVR 可以直接接入 IP 网络中,从而使存储的视频信息通过网络方便地共享。总体而言,DVR 和 NVR 系统均属于第二代,是部分数字化的视频监视系统。

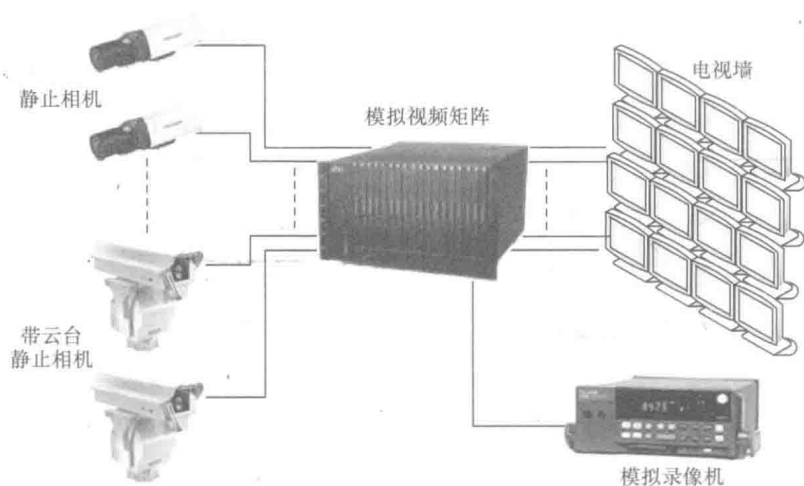
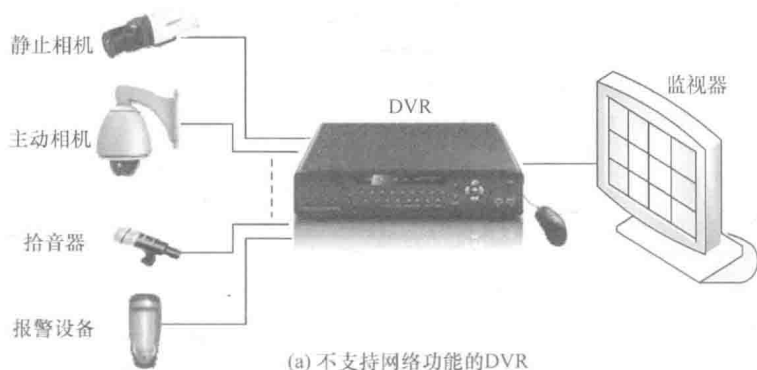


图 1-1 第一代视频监控系统



(a) 不支持网络功能的DVR

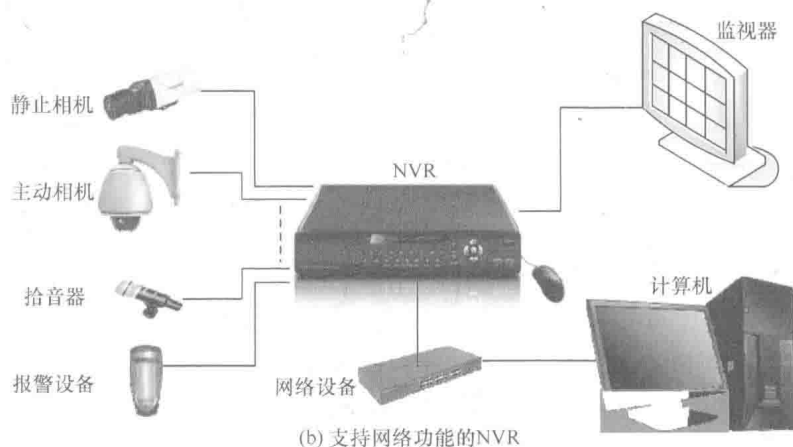


图 1-2 第二代视频监控系统

(3) 第三代视频监控系统

第三代视频监控系统指的是目前正在蓬勃发展的网络化视频监控系统,又称为 IP 监控系统,如图 1-3 所示。网络化视频监控系统从一开始就是针对网络环境设计的,它可以使用户通过网络中任何一台电脑观看、录制和管理实时视频信息。第三代视频监控系统是基于标准的 TCP/IP 协议、完全数字化的系统,能够通过网络进行传输,布控区域大大超过了前两代系统;采用开放式架构,可与门禁、报警、巡更、语音、MIS(management information system)等系统进行无缝集成;灵活性强,可以实现任意组合的预览和回放。

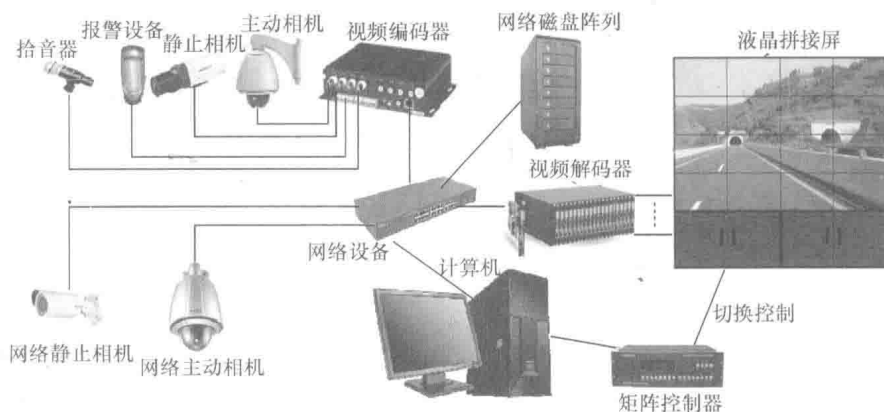


图 1-3 第三代视频监控系统

普通视频监控系统面对获取的图像无法自动判断其内容的含义,即难以判断摄像机所面对的场景到底“发生”了什么事情。该问题的解决往往依靠人工完成,即值班人员利用双眼“监控”成百上千的摄像机屏幕,对存储的视频数据进行回放“检索”,对可疑信息进行人工对比和分析。

智能视频监控技术就是在这种背景下出现并不断发展完善的,主要用于解决人工值守可能出现的异常信息的遗漏问题,并期望实现多点位、全天候不间断的自动监控。正如 Maryland 大学的智能视频监控系统 W^[4] 所描述的那样,传统视频监控为人们解决了监控中能不能看见的问题,而智能视频监控则要回答监控目标的出现时间(when)、空间位置(where)、身份(who)和行为(what)等问题^[1]。围绕这四个问题及其延伸出来的其他任务,研究人员进行大量的理论研究和系统设计,并在工程实践中解决了一系列关键技术难题,设计出了一大批优秀的智能视频监控系统。

在早期的监控系统中,Pfinder(person finder)^[2]也许是最广为人知的一个。Pfinder 是实现人体跟踪和行为分析的一种实时系统,该系统通过颜色和形状的多分类统计模型获取头部和四肢的二维描述,可以对室内人员的行为进行实时监视与判定。

VSAM(visual surveillance and monitoring)系统^[3]是由美国卡内基梅隆大学戴维 SARNOFF 研究中心研制的智能视频监控系统,其目标是利用视频理解、网络通信,以及多传感器融合等技术实现战场环境的监控,主要功能包括融合多种类型的传感器,对监控区域进行全方位昼夜监控;分析、预测人的异常行为,并根据预测结果进行自动提示和报警。

英国爱丁堡大学负责的 BEHAVE: computed-assisted prescreening of video streams for unusual activities 项目^[4]可以对视频序列中的异常行为进行检测,并去除包含正常行为的视频片段,从而只针对有意义的部分进行专门的分析。其目标是检测、理解、区分不同的交互方式,以及分析人群的正常行为与异常行为等。

美国 Vidient 公司推出的 Smart Catch 系统^[5]通过使用先进的机器

学习、多物体跟踪和行为推理技术,提供包括周边入侵、人群聚集,以及物品滞留等十余种事件的准确监测功能,准确率高达95%以上。该系统主要应用于机场安防监控,如美国旧金山国际机场、圣地亚哥国际机场、德国法兰克福国际机场等。

法国 VisioWave 公司设计的 VisioWave 视频监控系统^[6]除了具备一定的运动检测功能之外,还拥有高效的视频压缩、智能化的网络结构和灵活的存储归档机制。该系统先后赢得巴黎、纽约、伦敦三个城市地铁的智能视频安防项目,其中纽约地铁项目摄像机的数量达到了25 000多个。

韩国三星公司研制的哨兵机器人 SGR-A1^[7]将智能视频分析与军用机器人技术进行有机结合,通过可见光摄像机和红外热像仪获取场景图像并识别监控场景内的潜在威胁,同时向总部发出警告,控制中心操作员在确认闯入者的敌对身份以后可以根据战场情况命令机器人向闯入者开火。

国内在视频监控方面的研究相对滞后,至今未开发出一套类似于VSAM的大型智能视频监控系统,只在军事和商业领域有小规模的应用。在国内研究机构中,研究较为深入的是中国科学院自动化研究所下属的模式识别实验室,研究范围包括人脸检测与跟踪^[8]、智能交通^[9]、多摄像机联合跟踪^[10]、异常行为检测^[11]等。该实验室开发的 Vs-star(visual surveillance star)交通监控原型系统^[12]主要以研究为目的,部分技术已付诸实际应用。此外,上海交通大学、北京航空航天大学、南京理工大学、中国科学院计算技术研究所等高校和研究机构也进行了这方面的研究。

1.2 运动目标检测技术

智能视频监控系统的一般结构框架如图1-4所示,主要包括运动目标检测模块、目标特征描述模块、运动目标跟踪模块、目标行为分析及异常检测模块。在上述四个主要模块中,运动目标检测的主要目的是将运动目标从序列图像中分离出来^[13],是后续目标描述、目标分类,以

及行为分析的基础,运动目标检测算法的好坏直接影响到这些后续应用的质量,因此运动目标检测是智能视频监控的一个重要而基础的研究内容。根据摄像机是否运动,可以分为静态背景下的运动目标检测和动态背景下的运动目标检测两类^[14,15],前者的研究相对广泛和深入,而後者的研究则相对较少。此外,由于光照等因素的影响,运动目标检测的结果中往往包含一定的阴影成分,这些阴影成分将会随运动目标一起运动,使得目标分割不彻底,从而最终导致后续目标跟踪和行为分析的失败,因此去除阴影成分在运动目标检测中有着非常重要的作用^[16]。



图 1-4 智能视频监控系统的结构框架

1.2.1 静态背景下的运动目标检测技术

静态背景下,摄像机与监控场景保持相对静止,在此类摄像机静止——目标运动的情形下,运动目标的检测较为容易,常用的方法包括帧间差分法^[17,18]、背景差分法^[19]等。帧间差分法将相邻两帧图像进行差分,并将亮度变化超出阈值的像素判为前景,该方法运算速度很快,但由于目标运动的复杂性,检测结果常会出现空洞,并且包含大量的噪声。背景差分方法利用背景模型表征背景像素,通过当前图像与背景模型的差分比较实现运动目标的检测,同时当前的检测结果又用来更新背景模型。这类方法能在一定程度上适应场景的微弱变化和噪声扰动。

原始的背景差分法中背景通常是静态单一的背景图像。随着应用范围的不断延伸,处理的场景越来越复杂,背景也从单一分布发展为多模态分布的背景,因此广义的背景差分并不是简单地将当前帧与背景图像作差分,因为背景图像可能不是唯一的。背景差分方法能在适中

的硬件配置下实现较为精确的目标分割,同时具备一定的实时性,因此在智能视频监控领域得到了广泛的应用。Toyama^[20]阐述了背景差分法应当遵循的几个基本准则。

- ① 背景差分模型不应当处理目标语义上的区别。
- ② 背景差分模型应当能检测出第一次进入场景的感兴趣目标。
- ③ 应当定义一个像素级的背景静止准则,使得满足此准则的像素被检测为背景并被忽略。
- ④ 背景模型应当能够适应背景的突然变化或者缓慢变化。
- ⑤ 背景模型应当考虑背景在不同空间尺度上的变化。

目前背景差分法中采用最多的仍然是混合高斯模型,它已经成为背景差分法中的一种标准模型,同时也是很多其他算法的基础。混合高斯模型的本质是利用多个高斯函数对像素的分布进行建模,以处理背景多峰分布情况,同时高斯函数的参数能根据时间推移自动进行更新,以适应背景的复杂变化。然而,现实场景中的背景运动方式往往是非平稳的,有些甚至是杂乱无章的,采用混合高斯模型检测得到的结果虚警率往往偏高。针对含有非平稳背景的场景中的目标检测问题,Ren等^[21]提出空间分布高斯模型(spatial distribution of Gaussians, SDG),通过同时从时间和空间的角度对背景像素的分布进行建模来判断像素的归属,这种方法要进行一个图像配准的过程,因此检测结果的好坏在很大程度上依赖于配准的精度,且配准过程也大大降低了算法的实时性。Pless等^[22]提出多个基于图像亮度分布和时空导数的像素模型,同时对这些模型进行了对比分析,指出不同的背景模型适用于不同的场景,并提出利用运行特性接收曲线(receiver operator characteristic plot, ROC)作为选择最佳背景模型的依据。Mittal等^[23]提出基于光流和颜色联合像素模型的自适应核密度估计框架,将光流作为更高维空间中的一种特征,并采用数据驱动核密度估计带宽来解决特征计算中固有的模糊性问题。Sheikh^[24]提出动态场景的贝叶斯模型,采用无参数密度估计的方法对多模态空间不确定性、空间位置和颜色之间复杂的相关性进行建模,在MAP-MRF框架下基于背景和前景模型的概率之比来确定像素的归属。