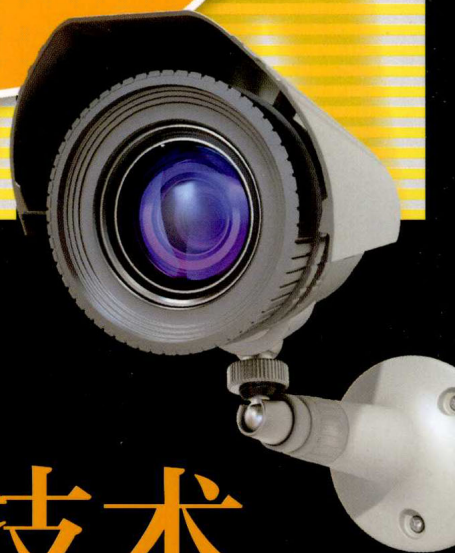


Image TRACKING

图像目标

鑫 徐立中 著

跟踪技术



人民邮电出版社
POSTS & TELECOM PRESS



图像目标 跟踪技术



王鑫 著

人民邮电出版社
北京

图书在版编目 (C I P) 数据

图像目标跟踪技术 / 王鑫, 徐立中著. — 北京 :
人民邮电出版社, 2012.12
ISBN 978-7-115-28897-4

I. ①图… II. ①王… ②徐… III. ①图象通信—目
标跟踪 IV. ①TN919.8②TN953

中国版本图书馆CIP数据核字(2012)第155048号

内 容 提 要

本书系统阐述了图像目标跟踪的有关概念、原理和方法,共分9章,第1章介绍图像目标跟踪的意义、应用及分类,第2章介绍非线性优化序贯拟蒙特卡洛滤波,第3章介绍融合背景信息的序贯拟蒙特卡洛滤波目标跟踪,第4章讨论基于概率图模型的粒子滤波多目标跟踪,第5章介绍基于序贯拟蒙特卡洛滤波的多摄像机目标跟踪,第6章介绍基于信息融合技术的目标跟踪,第7章讨论受机械参数影响的多摄像机深度估计,第8章介绍基于自适应多信息融合的均值漂移红外目标跟踪,第9章介绍融合均值漂移和粒子滤波优点的实时目标跟踪。本书是图像目标跟踪方面的专著,反映作者近年来在这一领域的主要研究成果。

本书内容新颖,理论联系实际,可作为大专院校及科研院所图像处理、计算机视觉和视频处理等领域的高年级本科生、研究生的教学和参考用书,也可供相关领域的教师、科研人员及工程技术人员作参考。

图像目标跟踪技术

-
- ◆ 著 王 鑫 徐立中
责任编辑 易东山
执行编辑 肇 丽
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
大厂聚鑫印刷有限责任公司印刷
 - ◆ 开本: 700×1000 1/16
印张: 11.75 2012年12月第1版
字数: 221千字 2012年12月河北第1次印刷

ISBN 978-7-115-28897-4

定价: 39.00 元

读者服务热线: (010)67119329 印装质量热线: (010)67129223
反盗版热线: (010)67171154

前 言

近年来,随着信息技术的飞速发展,图像目标跟踪技术成为计算机视觉领域中的一个重要课题,这是因为:首先,计算机处理能力不断提高,存储成本大幅降低,使得以视频速率或近似视频速率采集存储图像序列成为可能;其次,图像目标跟踪技术极为广阔的市场应用前景也是推动此项研究的主要动力。图像目标跟踪技术涉及视频监控、目标识别和分类、军事目标探测、机器人视觉导航、智能交通、人机交互和虚拟现实等诸多军事和民用领域,对其深入研究具有重要的理论意义和应用价值。

本书将系统阐述图像目标跟踪的有关概念、原理和方法。在内容上既选择了有代表性的图像目标跟踪经典内容,又结合作者近年来有关图像目标跟踪关键技术的研究与应用实践,选取了一些新的研究成果,具有一定的广度、深度和新颖性。

本书共分9章,主要包括:图像目标跟踪的意义、应用及分类,非线性优化序贯拟蒙特卡洛滤波技术,融合背景信息的序贯拟蒙特卡洛滤波目标跟踪技术,基于概率图模型的粒子滤波多目标跟踪技术,基于序贯拟蒙特卡洛滤波的多摄像机目标跟踪技术,基于信息融合的目标跟踪技术,受机械参数影响的多摄像机深度估计,基于自适应多信息融合的均值漂移红外目标跟踪技术,融合均值漂移和粒子滤波优点的实时目标跟踪技术。

第1章简要介绍图像目标跟踪的意义、应用及分类。简要分析单摄像机视觉跟踪和多摄像机视觉跟踪中不确定性问题研究的热点和难点,并介绍了红外图像目标跟踪的意义和应用,以及智能视频监控的背景、意义和一些典型的系统。

第2章介绍基于贝叶斯框架的跟踪问题描述,给出贝叶斯滤波的蒙特卡洛实现和拟蒙特卡洛实现,并针对序贯拟蒙特卡洛滤波算法计算复杂度高,粒子表示能力弱等缺点,设计非线性优化序贯拟蒙特卡洛滤波目标跟踪算法。

第3章简要分析单摄像机视觉跟踪中目标表观和背景数据的不确定性,并将序贯拟蒙特卡洛滤波作为跟踪框架应用于单摄像机视觉跟踪。

第4章针对单摄像机多目标跟踪中的众多不确定性问题,首先介绍概率图模型理论,在此基础上,介绍基于概率图模型的粒子滤波多目标跟踪算法。

第5章首先介绍基于贝叶斯理论的多摄像机目标跟踪建模,及多摄像机之间的单应约束原理和计算方法,然后介绍基于序贯拟蒙特卡洛滤波的多摄像机目标

跟踪算法。

第6章首先介绍监控视频信息融合的概念、结构和方法,然后针对单一摄像机介绍一种基于目标颜色和边缘方向特征随机融合的粒子滤波跟踪方法,最后介绍一种基于多摄像机数据融合新策略的多摄像机视频跟踪融合系统。

第7章首先简要介绍立体视觉模型,然后针对双目立体视觉可能出现的机械参数的不确定性进行分析,并介绍5种不确定性单独出现情况下的深度估计算法及多个不确定性同时出现时的深度估计算法。

第8章首先介绍均值漂移的基本原理及其在目标跟踪中的应用,然后针对复杂背景下红外目标跟踪问题,介绍一种改进的基于自适应多信息融合的均值漂移红外目标跟踪算法。

第9章首先针对彩色目标跟踪问题,介绍一种融合均值漂移和粒子滤波优点的实时彩色目标跟踪算法。此外,为了能够快速和准确地跟踪红外运动目标,介绍一种融合均值漂移和粒子滤波优点的红外目标跟踪算法。

本书第1章、第4章、第6章、第8章、第9章由王鑫编写;第2章、第3章、第5章、第7章由徐立中、丁晓峰编写;全书由王鑫、徐立中统稿。

本书是作者在近年来研究工作的基础上写作而成的。成书之际由衷地感谢作者的导师唐振民教授,感谢他多年来对作者的培养和悉心指导,有幸在他领导的研究室中参加科研工作并受到锻炼,使作者受益一生。

衷心感谢徐立中教授提供了优良的博士后研究工作环境,奠定了本书的写作基础。

在研究和写作过程中,课题组刘超颖、吴正军等提供了本书的部分素材,在此向他们表示衷心的感谢。

向所有的参考文献作者及为本书出版付出辛勤劳动的同志们表示感谢。

限于作者的水平,书中难免有缺点和不完善之处,恳请批评指正。

王鑫于河海大学

2012年3月14日

目 录

第 1 章 绪论	1
1.1 图像目标跟踪的意义和应用	1
1.2 单摄像机目标跟踪	3
1.2.1 目标表示模型	4
1.2.2 目标动态模型	6
1.2.3 目标状态估计模型	6
1.3 多摄像机目标跟踪	8
1.3.1 目标匹配	9
1.3.2 摄像机标定及拓扑关系估计	10
1.3.3 数据关联	12
1.4 红外图像中目标的跟踪	13
1.5 智能视频监控系统	15
1.5.1 智能视频监控的背景和意义	15
1.5.2 智能视频监控系统	16
参考文献	18
第 2 章 非线性优化序贯拟蒙特卡洛滤波	24
2.1 引言	24
2.2 基于贝叶斯框架的跟踪问题描述	25
2.2.1 贝叶斯滤波的蒙特卡洛实现	27
2.2.2 贝叶斯滤波的拟蒙特卡洛实现	29
2.3 非线性优化序贯拟蒙特卡洛滤波	32
2.3.1 信赖域方法	32
2.3.2 基于信赖域的序贯拟蒙特卡洛滤波算法	33
2.4 实验与分析	35
2.4.1 非线性动态模型	35
2.4.2 二维点目标跟踪中的应用	38
参考文献	39

第3章 融合背景信息的序贯拟蒙特卡洛滤波目标跟踪	41
3.1 引言	41
3.2 协方差特征	43
3.2.1 基于积分图的协方差特征快速计算方法	44
3.2.2 协方差特征的基本性质	45
3.2.3 基于协方差特征的目标表观和背景模型	46
3.3 融合背景信息的序贯拟蒙特卡洛滤波目标跟踪算法	47
3.3.1 融合背景信息的粒子权重估计	48
3.3.2 目标匹配方法和表观模型更新策略	49
3.4 实验与分析	50
3.4.1 实验1	50
3.4.2 实验2	51
3.4.3 实验3	53
参考文献	54
第4章 基于概率图模型的粒子滤波多目标跟踪	57
4.1 引言	57
4.2 概率图模型	59
4.2.1 图论的基本概念	59
4.2.2 无向概率图模型	60
4.3 基于概率图模型的粒子滤波多目标跟踪算法	61
4.3.1 多目标概率图模型建模	61
4.3.2 概率图模型推理	62
4.3.3 目标空间特征建模	64
4.3.4 基于概率图模型的粒子滤波算法	65
4.4 实验与分析	66
参考文献	70
第5章 基于序贯拟蒙特卡洛滤波的多摄像机目标跟踪	72
5.1 引言	72
5.2 基于贝叶斯理论的多摄像机目标跟踪建模	73
5.2.1 系统描述	73
5.2.2 后验概率推导	74
5.3 多摄像机单应约束原理	75
5.3.1 单应约束矩阵计算原理	75

5.3.2 基于 RANSAC 的单应矩阵计算方法	76
5.4 基于序贯拟蒙特卡洛滤波的多摄像机目标跟踪算法	77
5.4.1 目标表观建模	77
5.4.2 似然性	77
5.4.3 自适应重要性采样	78
5.5 实验与分析	80
5.5.1 实验 1	80
5.5.2 实验 2	83
5.5.3 实验 3	84
参考文献	86
第 6 章 基于信息融合技术的目标跟踪	88
6.1 引言	88
6.2 监控视频信息融合	88
6.2.1 信息融合基本概念	89
6.2.2 视频信息融合结构	89
6.2.3 视频信息融合方法	90
6.3 基于多信息融合的粒子滤波目标跟踪	92
6.3.1 基于贝叶斯估计的目标跟踪框架	92
6.3.2 粒子滤波理论	93
6.3.3 基于多信息融合的粒子滤波目标跟踪算法	96
6.3.4 提出算法的具体步骤	101
6.3.5 实验与分析	102
6.4 多摄像机融合跟踪	106
6.4.1 多摄像机融合系统	106
6.4.2 多摄像机数据融合	107
6.4.3 系统设计与仿真	109
参考文献	114
第 7 章 受机械参数影响的多摄像机深度估计	116
7.1 引言	116
7.2 立体视觉模型	118
7.3 不同机械参数误差下的深度估计模型	119
7.3.1 摄像机畸变	120
7.3.2 摄像机平移	121

7.3.3	摄像机旋转	122
7.3.4	摄像机上下倾斜	123
7.3.5	摄像机左右偏转	124
7.3.6	多个误差同时发生	126
7.4	实验与分析	128
7.4.1	特征点匹配	128
7.4.2	参数 d 、 α 和 β 的估计	129
7.4.3	参数 γ 的估计	129
7.4.4	实验结果	131
7.4.5	讨论分析	133
	参考文献	134
第 8 章	基于自适应多信息融合的均值漂移红外目标跟踪	137
8.1	引言	137
8.2	Mean Shift 理论	138
8.2.1	Mean Shift 向量	138
8.2.2	基于核函数直方图的均值漂移目标跟踪算法	140
8.2.3	基于概率分布图的均值漂移目标跟踪算法	143
8.3	基于自适应多信息融合的均值漂移红外目标跟踪算法	145
8.3.1	新的红外人体目标模型	145
8.3.2	提出的目标跟踪算法	148
8.4	实验与分析	154
8.4.1	提出的算法和基于灰度信息均值漂移算法的比较	154
8.4.2	自适应多信息融合算法和固定权值多信息融合算法比较	157
8.4.3	遮挡问题的处理	159
8.4.4	改进算法的效率分析	161
	参考文献	161
第 9 章	融合均值漂移和粒子滤波优点的实时目标跟踪	163
9.1	引言	163
9.2	融合均值漂移和粒子滤波优点的实时彩色目标跟踪	163
9.2.1	改进的融合颜色和运动信息的目标模型	164
9.2.2	改进的均值漂移算法	166
9.2.3	本节算法的具体实现	167
9.2.4	实验与分析	168

9.3 融合均值漂移和粒子滤波优点的实时红外目标跟踪.....	172
9.3.1 改进的融合灰度和运动信息的目标模型	173
9.3.2 改进的均值漂移算法	174
9.3.3 本节算法的具体实现	174
9.3.4 实验与分析.....	174
参考文献.....	176

第1章 绪 论

1.1 图像目标跟踪的意义和应用

在现代社会中，人的生命和财产安全一直是人类所关注的焦点。由于视频序列可以给我们提供大量有价值的信息，利用视频进行监视已经成为一种有效保护人的生命和财产的工具。越来越多的闭路电视（Closed-Circuit Television, CCTV）系统^[1]被越来越广泛地安装在与我们日常生活息息相关的场所，如社区和重要建筑物的内部（大型购物场所、休闲广场等）、城市公共汽车站（飞机场、火车站、汽车站、公交站台等）等。

但是传统的CCTV系统主要是依赖人工操作处理从各个摄像机传送到监视中心屏幕上的图像。实际生活中，大多数大规模CCTV系统中，摄像机的个数与监视中心操作人员的比例在4:1~30:1之间，研究表明，人工操作员最多能同时有效地处理来自4个摄像机的视频图像^[2]。此外，异常情况总是在某一较长时间的正常情况之后发生的，而这一时段又是监视人员的监视疲劳期，这在一定程度上又使得操作人员及时成功发现异常情况的可能性降低。

随着计算机视觉、模式识别、控制工程、人工智能等理论及相关技术的发展，视频监控系统已经从第一代的模拟CCTV系统经过第二代的半自动视频监控系统发展到了现在的第三代广域自动智能视频监控（Intelligent Video Surveillance, IVS）^[3]系统。第二代的半自动视频监控系统在传统CCTV系统上结合了计算机视觉技术，可以处理来自摄像机的图像和信号，使系统主动检测某些异常事件，而区别于第一代被动的录制，这种系统所涉及的大都是集中式处理。第三代广域自动视频监控系统旨在建立更广范围内实时、高效的自动智能监视系统，由于分布式计算、嵌入式计算和传感器网络的发展，分布式多智能摄像机监控系统成为了第三代广域自动视频监控系统中的主流方向^[4]。

目前，智能视频监控的研究主要集中在视觉跟踪以及后续的视频内容理解等方面。其中，视觉跟踪是指对图像序列中的运动目标进行检测、提取、识别和跟踪，获得运动目标的运动参数，如位置、速度、加速度等，以及运动轨迹，从而为后续的各种高级处理，如目标行为分析、目标识别等建立基础^[5]。世界上许多国家与视觉相关的研究团队针对视频监控中的目标检测和跟踪等方面做出了大量

的工作,并在众多著名国际期刊(IEEE Transactions on Pattern Analysis and Machine Intelligence; IEEE Transactions on Image Processing; IEEE Transactions on Circuits and Systems for Video Technology; International Journal of Computer Vision; Computer Vision and Image Understanding; Image and Vision Computing; Pattern Recognition; Pattern Recognition Letters; Machine Vision and Applications; IET Computer Vision; IET Image Processing; IET Signal Processing 等)和重要国际会议(ICCv; ECCV; CVPR; BMCV; VS-PETS; ICDSC 等)上发表了相关的论文。

视觉跟踪在以下多个方面有实际的应用需求。

(1) 视频监控 (Video Surveillance): 跟踪人或车辆从而检测非正常行为,并对特定的目标进行识别。

(2) 视频编辑 (Video Editing): 在视频录像中给一个移动目标或人增加一个图形内容。

(3) 分析体育运动事件 (Analysis of Sport Events): 提取运动员在体育比赛中的位置信息,通过这些数据,运动专家可以分析运动员的表现。

(4) 跟踪实验室动物 (Tracking of Laboratory Animals): 学习自然界多智能主体的相互作用。

(5) 人机交互 (Human-Computer Interfaces): 用于智能环境从而协助人的日常任务。

(6) 认知系统 (Cognitive System): 通过跟踪来学习不同目标在其环境中的动态性质。

然而,视觉跟踪系统又是一个包含大量不确定性信息的系统,如固有可视数据的不确定性、目标动态特性的不确定性以及摄像机参数的不确定性等。因此,从视觉跟踪中存在的不确定性问题入手,对视觉跟踪开展研究具有重要的意义。

按照不同的情况,如摄像机是否运动、场景中运动目标的数目、场景中运动目标的类型、摄像机的数目等,可以将视觉跟踪问题分为很多类型^[6]。

1. 固定摄像机与运动摄像机

根据摄像机是否运动可将跟踪问题分为基于固定摄像机的目标跟踪和基于运动摄像机的目标跟踪两类。大多数的视频监控系统都是采用固定摄像机对关注的场景进行监视,此时背景是静止的,目标是移动的。这种情况下通常使用背景差法对运动目标进行检测。但是,在很多复杂情况下,摄像机是运动的,其运动的形式可以分为两种:一种是摄像机的支架固定,摄像机进行旋转、俯仰以及缩放;另一种是摄像机装在某个移动的载体上,如移动的车辆、飞行器。在这些情况下,背景和前景都是运动的,对运动目标进行检测并跟踪需要对背景运动进行估计或匹配。如果摄像机的运动方式被限定在垂直于光轴的平面上,可以采用图像拼接的方法将背景拼在一起,然后按照摄像机静止时的跟踪方法进行处理。如果摄像

机是沿着光轴运动的, 可以考虑使用光流法 (Optical Flow) [7]。

2. 单目标与多目标

根据场景中运动目标数目的多少, 视频跟踪问题可分为单目标跟踪和多目标跟踪。单目标跟踪虽然只跟踪一个目标, 但在复杂背景下仍然不是一个容易处理的问题。由于场景中噪声的影响, 往往不能准确地检测到真正的运动目标, 即使检测到了运动目标, 由于阴影或光照的影响, 提取出来的运动目标也可能不完整, 而多目标跟踪比单目标跟踪又要困难得多。在多目标跟踪过程中, 通常采用一个跟踪器跟踪一个目标, 它除了涉及单目标跟踪的问题, 还必须考虑到多个目标在场景中的互相遮挡、合并、分离等情况, 这是多目标跟踪问题的难点[8]。

3. 刚体与非刚体

根据场景中运动目标类型的不同, 可将视觉跟踪问题分为刚体 (Rigid) 和非刚体 (Non-rigid) 的跟踪。所谓刚体是指具有刚性结构、不易变形的物体, 例如车辆、飞机等人造物体。这些物体的共同特点是结构比较规范, 能够用几何模型描述, 对这类运动目标的跟踪常采用基于模型的跟踪方法。而非刚体是指外形能够变化的物体, 如人、动物、细胞等, 对这类目标进行跟踪常采用基于变形模板 (Deformable Template) 的方法。

4. 可见光图像与红外图像

根据所使用的传感器种类的不同, 可将视觉跟踪问题分为由摄像机获得的可见光图像和由红外传感器获得的红外图像的跟踪等, 这两种类型的跟踪分别在不同的情况下使用。由于红外图像中物体的亮度主要取决于物体的温度和辐射的热量, 与光线条件无关, 因此可使用红外图像在夜间进行目标跟踪, 而白天使用可见光图像进行目标跟踪。如果将两者所获得的图像进行融合处理, 可以获得比单一传感器更多的信息, 从而提高目标检测和跟踪的能力。对于红外图像的目标检测可采用基于模板匹配的方法、基于亮度阈值的方法, 或者采用某种模式识别或机器学习方法对感兴趣区域 (Region of Interest, ROI) 进行分类以确定检测区域。

5. 单摄像机与多摄像机

根据摄像机的数目, 视觉跟踪可以分为单摄像机视觉跟踪和多摄像机视觉跟踪。下面两节将重点分析单摄像机视觉跟踪和多摄像机视觉跟踪中存在的不确定性问题, 并介绍单摄像机视觉跟踪和多摄像机视觉跟踪在国内外的研究现状。

1.2 单摄像机目标跟踪

单摄像机目标跟踪中的一个突出问题是固有的可视数据的不确定性和跟踪目标动态特征的不确定性。为了解决这些不确定性, 首先必须对目标的可视化

数据进行顽健的描述,其次要针对目标的动态特征进行正确的假设,最后对目标的状态进行预测。因此,一个典型的单摄像机视觉跟踪系统可以分为以下 3 部分:目标表示模型;目标动态模型;目标状态估计模型。目标表示模型用于捕获跟踪目标的可视属性,动态模型用来描述目标的动态特征,目标状态估计模型用于连续地估计目标的状态。针对具体的跟踪问题,将这 3 个部分相互协作、合理结合,构造出准确、顽健的跟踪器,从而实现单摄像机目标跟踪。

1.2.1 目标表示模型

目标表示模型是将图像中感兴趣的目标通过某种方法提取出来^[9]。基于目标表示模型中包含的信息,可以将目标表示模型分为四类:基于点特征的模型;基于形状;基于表观的模型;基于运动的模型。

1. 基于点特征的模型 (Point-based Model)

基于点特征的模型通过检测目标上的感兴趣点,将这些感兴趣点组成的集合作为描述目标的依据。一般地,目标上的感兴趣点是目标上具有多个方向奇异性的点,例如:Kanade-Lucas-Tomasi (KLT)^[10]特征点、Harris 角点^[11]、SIFT (Scale-invariant Feature Transform) 特征点^[12]、SURF (Speed Up Robust Feature) 特征点^[13]等。一般特征点分布在整个目标上,即使当目标有部分被遮挡时,仍然可以跟踪到其他的特征点。基于点特征的跟踪在对于刚体时有较好的效果,当用点特征来描述非刚体时,由于非刚体在运动过程中容易形变,从而使目标上的点特征不时地减少、增加,从而增加了跟踪的难度。

2. 基于形状;的模型 (Shape-based Model)

基于形状;的模型是通过对目标形状的描述来对目标建模。早期的目标形状模型是通过利用可变形线 (Deformable Line) 和主动轮廓线模型 (Active Contour Model) 来反复匹配目标的边缘特征,从而建立目标的形状模型。这种建模方法的主要缺点是对噪声比较敏感,不能处理多个目标之间发生遮挡的情况,而且模型把背景中出现的伪特征当成目标的边缘,从而建立错误的形状模型。当跟踪一类拥有相似形状的目标时,可以采用点分布模型 (Point Distribution Model) 来描述目标^[14],为了保证轮廓的平滑和目标的形状描述的准确性,需要大量的离散点,导致系统方程维数过高。Masouri^[15]将水平集用于目标跟踪领域。Parogis 和 Deriche^[16]利用水平集进行纹理分割以及运动目标分割和跟踪。

当跟踪的目标较小且形状变化较快时,非传统的形状可能更适合。例如,在体育运动视频的处理中,Perez 和 Kovacic^[17]利用 14 个二进制 Walsh 函数核来编码目标的形状,并通过相似性寻找目标在下一时刻的位置。Dalal 和 Triggs^[18]利用有向梯度直方图 (Histogram of Oriented Gradient, HOG) 来表示行人形状。Lu 等^[19]利用 HOG 方法来检测和跟踪冰球运动中的运动员。

3. 基于表观的模型 (Appearance-based Model)

基于表观的模型不对目标的物理结构、形状等属性进行直接建模,而是采用图像颜色、梯度、灰度等信息对目标进行建模。

直接利用颜色信息对目标进行建模是一类最常见的方法。如 Senior^[20]采用自适应统计模型来对目标的颜色进行建模,每一个目标通过一个矩形框来标记,利用高斯分布来建模框内像素的颜色分布。由于颜色特征受光照变化等因素影响, Jopson 等^[21]利用期望最大化算法 (Expectation Maximization, EM) 对目标外观颜色变化进行估计,在线更新模型参数。颜色直方图作为颜色特征的扩展,近年来已被成功应用于目标跟踪领域。利用颜色直方图作为目标特征的描述时,为了增强其顽健性, Comaniciu 和 Meer^[22]不仅考虑了目标本身的颜色直方图,还考虑了目标临近领域是否为目标的重要组成部分。另外, Wang 等^[23]采用混合高斯模型 (Mixture of Gaussians, MoG) 对目标的颜色分布进行建模,同时该模型还融合了空间信息。

为了在跟踪过程中获得更加顽健的目标表示模型,目标的多个特征,如颜色、纹理等常被组合在一起共同来描述目标。Li 和 Chaumette^[24]在光照变化等复杂背景下,通过将目标的形状、颜色、结构和边缘信息四者组合来对目标进行跟踪。Cootes 等^[25]提出了基于主动表观模型 (Active Appearance Model, AAM) 的目标表示方法,该模型可以看作是主动轮廓模型和主动形状模型的进化。Tuzel、Porikli 和 Meer^[26,27]提出了基于协方差的目标表观模型。协方差特征是一个对称正定的矩阵,可以用来描述图像中任意形状。随后他们又将协方差特征用于行人检测^[28],提出了基于协方差的行人检测方法,获得了比梯度直方图更好的检测效果。

4. 基于运动的模型 (Motion-based Model)

基于运动的模型主要利用目标在图像中运动的信息来获得目标的位置。典型的方法是考虑图像目标像素的外表运动,如光流法。光流法是根据连续几帧图像计算像素运动的大小和方向,利用运动场区分背景和运动对象。一般分为全局光流场和基于特征点的光流场两种。全局光流场的计算方法^[29]有 Horn-Schunck 方法、Lucas-Kanade 方法,在计算得到全局光流场后,通过比较运动目标与背景之间的运动差异实现对运动目标的分割,从而达到跟踪的目的。特征点光流法通过特征匹配来求得特征点处的流速,与全局光流场算法相比较,这种算法具有计算量小和快速灵活的特点。但是特征点光流法得到的是稀疏的光流场,因而很难提取到目标精确的形状。一般情况下,由于噪声、多光源、阴影、透明性和遮挡性等原因,计算得出的光流场分布不是十分可靠和精确,而且多数光流法计算复杂,除非有特殊的硬件支持,否则很难实现算法的实时处理。

1.2.2 目标动态模型

目标动态模型是用于描述目标的动态特性,即目标在图像中的运动方程。在跟踪中如果知道了目标的动态特性,就可以根据它推测目标在下一时刻可能出现的位置,然后再通过目标的表现模型对预测的位置进行搜索,从而确立目标下一时刻准确出现的位置。

在跟踪中常使用的目标动态模型有一阶动态模型、二阶动态模型、自适应速度模型等。为了能够描述目标可能发生的多种动态性,一些学者提出了相互作用的多个模型(Interacting Multiple Model, IMM)来描述目标的运动,此模型利用多个跟踪器来并行地跟踪目标,其中每一个跟踪器使用不同的目标动态模型。Bar-Shalom 等^[30]设计了一个评价标准来判定不同模型描述目标运动的准确性,另外,还分析了多个运动模型如何组合,以及不同跟踪器如何结合起来能够得到最好的跟踪结果。由于 IMM 采用了多个跟踪器,从而使该方法需要消耗大量的计算资源。

在一些情况下,例如体育运动视频跟踪、人机交互视频跟踪等,我们并不能准确地知道目标的动态特性,此时一般采用随机步长(Random-Walk)模型或近似匀速模型来描述目标的运动特性。

1.2.3 目标状态估计模型

目标跟踪问题可以看作是对目标状态随时间的统计估计。在视觉跟踪中,当确定了目标表现模型和目标动态模型后,通过目标状态估计模型可以连续估计目标在视频中的状态。从统计学角度,目标当前的状态被表示成目标状态空间中的一个概率密度函数(Probability Density Function, PDF),跟踪就是按照新的测量来递归估计目标的后验分布。

如果假设目标的动态过程和测量过程是线性的、高斯的,则目标后验概率可以通过卡尔曼滤波(Kalman Filter, KF)计算得到。然而,目标跟踪大多为非线性、非高斯过程,卡尔曼滤波中的线性、高斯假设则过于严格,如果直接利用卡尔曼滤波的假设应用于视觉跟踪中,一般会导致卡尔曼滤波的性能退化。

为了解决非线性问题,有学者提出了扩展卡尔曼滤波(Extended Kalman Filter, EKF)^[31]。EKF 是采用参数化的解析形式对系统的非线性模型进行线性近似。一般利用一阶 Taylor 展开对非线性部分进行线性化。为了获得更好的近似效果,又引入更高阶的 Taylor 展开。然而,随着阶数的增加,求解的复杂度也越来越高。EKF 有两个明显的缺点:(1)在对系统模型和测量模型的线性化过程中, Taylor 展开只是在一个点上展开,并没有考虑所有的点;(2)假设先验和后验概率都是服从高斯分布的,实际模型中的非线性会导致后验概率不服从高斯分布和在

多步传播后引起多模态,这破坏了EKF中高斯假设,从而导致滤波性能退化甚至使滤波完全失效。

为了更好地处理测量模型和系统状态转换模型的非线性,一些学者将加权统计线性回归(Weighted Statistical Linear Regression)方法引入到卡尔曼滤波中,通过一组确定性采样点(Sigma点集合)来捕获系统的相关统计参数。这一类改进的卡尔曼滤波统称为Sigma点卡尔曼滤波。根据sigma点选取的不同,主要分为无迹卡尔曼滤波(Unscented Kalman Filter, UKF)^[32]和中心差分卡尔曼滤波(Central-difference Kalman Filter, CDKF)^[33]。UKF是在无迹变换(Unscented Transform, UT)的基础上,基于卡尔曼滤波框架而建立,利用确定性采样来逼近状态的后验概率密度,而不对非线性函数进行近似,避免了雅可比矩阵的计算。与EKF类似,UKF中的后验概率是通过高斯分布近似,但是这个高斯分布是通过Sigma点集合来描述的。在传播过程中,这些点通过真正的非线性系统传播,利用这些点描述的后验概率的期望与方差的准确性达到了Taylor展开三次式。相对于EKF中的近似只是准确到Taylor展开一次式,UKF有了很大的提高,而且UKF的计算复杂度比EKF小。Norgarrd等^[33]提出了中心差分卡尔曼滤波,它利用Sterling内插公式对非线性函数作多项式逼近,与EKF相比在不增加计算量的前提下,提高了滤波精度,且比UKF有稍高的理论精度,更加易于实现。

考虑到一个非高斯分布可以被足够多的高斯分布组成混合高斯分布来近似,一些学者利用高斯混合模型对多模后验概率进行建模,提出了高斯和滤波(Gaussian Sum Filter, GSF)^[34]。高斯和滤波在准确性上比利用单个高斯模型更加准确,但是它在传播过程中,混合高斯模型的每一个元都是基于EKF的,这就将EKF的缺点引入了高斯和滤波中,而且,混合高斯模型中单独高斯的个数难以确定,如果观测噪声和系统噪声都是采用高斯混合模型来近似,则随着时间的推移,后验概率中混合高斯的组成元的个数呈指数增加。

Sigma点卡尔曼滤波、高斯和滤波均是采用高斯模型或者混合高斯模型来近似后验概率,这些模型均是参数模型。为此,一些学者提出了基于非参数的模型来近似后验概率,例如基于网格细胞的离散分布来近似后验概率、利用样条函数来近似后验概率、利用步长函数来近似后验概率、利用求积分的方法来近似后验概率等,这些方法均可称为基于网格的方法。由于基于网格的方法需要指定细胞的数目,而且网格需要充分的稠密才能够较好地近似后验概率,随着状态维数的增加,计算代价会越来越大。

上述介绍的算法虽然能够处理一些非线性问题,但是它们大都是基于高斯逼近的方法,而现实世界中的目标跟踪问题,大都是非线性、非高斯问题,采用均值和方差为特征显然不能精确描述目标状态的后验概率。

直到1996年,Isard和Blake^[35]提出了条件概率传播(Condition Density