



“十三五”科学技术专著丛书

# 视觉跟踪中的 马氏链蒙特卡洛方法

周修庄 鲁继文 著

(MCMC Sampling for Visual Tracking)



北京邮电大学出版社  
www.buptpress.com



“十三五”科学技术专著丛书

# 视觉跟踪中的马氏链蒙特卡洛方法

(MCMC Sampling for Visual Tracking)

周修庄 鲁继文 著



北京邮电大学出版社  
[www. buptpress. com](http://www.buptpress.com)

## 内 容 简 介

本书基于统计机器学习的理论工具——马氏链蒙特卡洛(MCMC)采样理论,系统地阐述了目标运动不确定性条件下的视觉目标跟踪问题、方法和技术,并结合多种实际应用场景给出算法实现框架和实验分析。全书共4章,第1章介绍了视觉目标跟踪的意义和基本方法。第2章介绍了MCMC采样、自适应MCMC采样的基本理论、方法、实现技术及其在计算机视觉中的应用。第3章介绍了视觉目标跟踪中的序贯粒子采样方法和序贯随机逼近蒙特卡洛采样方法。第4章介绍了面向运动突变视觉目标跟踪的自适应MCMC采样方法。本书是视觉目标跟踪与MCMC采样方面的专著,反映了作者近年来在这一领域的主要研究成果。

本书可作为大专院校及科研院所图像处理、计算机视觉和智能视频分析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。

### 图书在版编目(CIP)数据

视觉跟踪中的马氏链蒙特卡洛方法 / 周修庄, 鲁继文著. -- 北京: 北京邮电大学出版社, 2018. 6

ISBN 978-7-5635-5388-4

I. ①视… II. ①周… ②鲁 III. ①蒙特卡罗法 IV. ①0242. 28

中国版本图书馆CIP数据核字(2018)第023471号

---

书 名: 视觉跟踪中的马氏链蒙特卡洛方法

著作责任者: 周修庄 鲁继文 著

责任编辑: 满志文 穆菁菁

出版发行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路10号(邮编:100876)

发行部: 电话: 010-62282185 传真: 010-62283578

E-mail: publish@bupt.edu.cn

经 销: 各地新华书店

印 刷: 北京九州迅驰传媒文化有限公司

开 本: 720 mm×1 000 mm 1/16

印 张: 7

字 数: 120千字

版 次: 2018年6月第1版 2018年6月第1次印刷

---

ISBN 978-7-5635-5388-4

定 价: 25.00元

• 如有印装质量问题,请与北京邮电大学出版社发行部联系 •

# 前 言

视觉跟踪是计算机视觉研究的一个基本和重要环节。近年来,随着相关学科领域理论和技术的发展,视觉跟踪研究在理论方法和实现技术方面得到了进一步的深入和发展,并且在智能交通、人机交互、视频监控和军事等领域得到广泛的应用。在实际应用中,视觉跟踪方法通常需要处理一些复杂的情况,包括:如何建模目标的运动变化,以有效地捕获目标运动的不确定性;如何设计有效的外观模型以适应外观的显著变化;如何设计稳健高效的视觉跟踪算法以满足实际的目标跟踪任务等等。运动极大不确定性条件下的目标跟踪是视觉跟踪的一个难点,基于运动连续性假设的传统贝叶斯滤波跟踪方法在解决这一问题时具有很大的局限性。本书系统地阐述自适应马氏链蒙特卡洛采样的理论方法和实现技术,以及在目标运动极大不确定性条件下的视觉跟踪方法。

本书共分4章,主要内容包括:视觉跟踪的意义、应用和主要方法概述;马氏链蒙特卡洛采样和自适应马氏链蒙特卡洛采样方法;序贯随机逼近蒙特卡洛跟踪方法;运动突变的视觉跟踪方法。

第1章简要介绍视觉跟踪的意义、应用和主要方法概述。简要分析视觉跟踪的理论框架和基本技术,包括目标表示、外观建模、状态预测与估计等技术环节及其代表性算法。

第2章介绍马氏链蒙特卡洛采样方法及其在计算机视觉中的应用。首先介绍蒙特卡洛采样的基本理论、方法和实现技术。在此基础上介绍两种自适应马氏链蒙特卡洛采样方法及其算法实现,并给出它们在目标轮廓的多边形近似中的应用。

第3章介绍序贯随机逼近蒙特卡洛跟踪方法。通过分析贝叶斯滤波框架下的粒子滤波和基于马氏链蒙特卡洛采样的粒子滤波,引入能够有效克服粒子退化问题的序贯随机逼近蒙特卡洛采样方法。

第4章介绍运动突变的视觉跟踪方法。针对目标快速运动、视频镜头切换、低帧率视频等运动极大不确定性的视觉跟踪问题,介绍一种将目标分布自适应和提议分布自适应相结合的强自适应马氏链蒙特卡洛采样方法。

本书第1章至第4章由周修庄、鲁继文编写;全书由周修庄、鲁继文统稿。



本书是作者近年来研究工作的基础上编写而成。成书之际由衷地感谢本书作者周修庄的博士导师陆耀教授,感谢他多年的悉心培养和指导。同时,向本书参考文献的所有作者,以及为本书出版付出辛勤劳动的编辑老师们表示感谢。

限于作者的水平,本书难免有缺点和不足之处,恳请批评指正。

周修庄于北京邮电大学

2018年1月20日

# 作者简介

1. 周修庄,男,博士,北京邮电大学自动化系副教授,博士生导师。2011年毕业于北京理工大学计算机学院计算机应用技术专业,获得工学博士学位。现任职于北京邮电大学自动化学院。研究领域为计算机视觉、模式识别与机器学习,主要研究兴趣包括:人脸分析、视觉跟踪、度量学习,深度学习等。主持参与了国家自然科学基金、教育部博士点基金、北京市自然科学基金等多个国家级和省部级项目的研发工作。在IEEE TPAMI/TIP/TIFS和IEEE CVPR/ACM MM等人工智能领域国际权威期刊和顶级会议发表论文40多篇,2篇入选ESI全球高被引论文。担任国际期刊《Neurocomputing》编委,ICME/IJCB/VCIP等多个国际学术会议的Special Session Chair。中国计算机学会计算机视觉专委会委员、中国自动化学会模式识别与机器智能专委会委员、中国图形图像学会视觉大数据专委会委员。

2. 鲁继文,男,清华大学自动化系副教授,博士生导师,IEEE高级会员,中组部青年千人计划入选者(2015年)。主要研究方向为计算机视觉、模式识别和机器学习。具体研究内容包括深度学习、度量学习、人脸识别和视频分析等。2012年获得新加坡南洋理工大学电气与电子工程学院博士学位。2011年3月至2015年10月在美国伊利诺伊大学香槟分校新加坡高等研究院工作,担任研究科学家,2015年11月加入清华大学工作至今。近年来在领域国际权威期刊和顶级会议上发表论文180余篇,其中IEEE Transactions论文51篇(含PAMI论文8篇),ICCV/CVPR/ECCV/NIPS论文27篇,ESI高被引论文和热点论文6篇,SCI他引1200余次,Google Scholar引用4200余次,获国际会议最佳论文奖或最佳论文提名奖3次。担任IEEE信号处理学会多媒体信号处理技术委员会委员,IEEE信号处理学会信息取证与安全技术委员会委员,IEEE电路与系统学会多媒体系统与应用技术委员会委员,Pattern Recognition、Journal of Visual Communications and Image Representation等国际期刊编委,中国工程院院刊《Engineering》青年通讯专家,ACCV/ICIP/ICME等20多个国际会议的领域主席或专题主席。

# 目 录

第 1 章 绪论 .....	1
1.1 视觉跟踪概述 .....	1
1.2 国内外研究现状及发展趋势 .....	4
1.2.1 目标表示、特征选择与外观建模 .....	4
1.2.2 视觉跟踪方法 .....	7
第 2 章 MCMC 和自适应 MCMC .....	19
2.1 引言 .....	19
2.2 Monte Carlo 采样 .....	20
2.2.1 Monte Carlo 积分 .....	20
2.2.2 Monte Carlo 采样 .....	21
2.3 MCMC 采样 .....	23
2.3.1 Metropolis-Hastings 采样 .....	25
2.3.2 Gibbs 采样 .....	27
2.3.3 MCMC 算法分析 .....	28
2.4 自适应 MCMC 采样 .....	31
2.4.1 提议分布自适应 .....	32
2.4.2 目标分布自适应 .....	35
2.5 MCMC 采样方法在计算机视觉中的应用 .....	39
2.5.1 目标轮廓的多边形近似 .....	39
2.5.2 基于 MC 优化的多边形近似算法 .....	41
2.6 小结 .....	48
第 3 章 序贯随机逼近蒙特卡洛跟踪方法 .....	49
3.1 引言 .....	49
3.2 相关工作 .....	50

3.3 方法介绍 .....	52
3.3.1 贝叶斯滤波 .....	52
3.3.2 粒子滤波 .....	53
3.3.3 基于 MCMC 采样的粒子滤波 .....	56
3.3.4 序贯随机逼近蒙特卡洛跟踪方法 .....	57
3.4 实验结果 .....	66
3.5 小结 .....	69
<b>第 4 章 运动突变的视觉跟踪方法 .....</b>	<b>71</b>
4.1 引言 .....	71
4.2 相关工作 .....	72
4.3 方法介绍 .....	73
4.3.1 基本思想 .....	73
4.3.2 密度网格 .....	75
4.3.3 提议分布自适应 .....	76
4.3.4 强自适应的 MCMC 采样 .....	77
4.3.5 实现细节 .....	79
4.4 实验结果 .....	80
4.4.1 提议分布自适应的影响 .....	80
4.4.2 定性比较 .....	81
4.4.3 定量比较 .....	85
4.4.4 讨论 .....	87
4.5 小结 .....	88
<b>结论 .....</b>	<b>89</b>
<b>参考文献 .....</b>	<b>91</b>



# 第 1 章

## 绪 论

### ◆ 1.1 视觉跟踪概述 ◆

据研究,人类通过感官感知外部世界的信息超过 70%来自视觉。计算机视觉是在对人类视觉特性研究的基础上,利用计算机成像装置替代人类视觉器官作为信息输入手段,由计算机代替人脑完成对输入信息的处理和分析,使得计算机能够像人一样感知和理解外部世界,甚至完成人类难以完成的特定任务。计算机视觉是一门交叉性学科,涉及计算机科学、统计学、认知心理学、物理学、信号处理和分析,等等。近年来,随着各相关学科技术的发展,计算机视觉研究取得了令人瞩目的成就,在各个领域得到了广泛的应用。计算机视觉的研究起始于 20 世纪中叶。这个时期的研究工作主要是针对简单的二维图像进行识别和分析。在 20 世纪 70 年代,麻省理工学院人工智能实验室的 David Marr 教授提出了计算机视觉计算理论,该理论成为计算机视觉研究领域中第一个具有完整体系的理论框架,是计算机视觉研究领域一个里程碑式的工作,它极大地促进了计算机视觉理论的研究和发展<sup>[1]</sup>。

广义上,视频目标跟踪(视觉跟踪)是指对视觉场景中的运动目标进行检测、识别和跟踪,获得目标的运动、状态以及外观特性等参数,进而对跟踪的运动目标作进一步的处理与分析,实现对运动目标的更高层次的理解。视觉跟踪是计算机视觉领域中的一个研究方向。由于它具有极为广阔的应用前景,因而近年来已成为计算机视觉领域中备受瞩目的研究热点。视觉跟踪在军事国防、人机交互、智能监控、机器人导航、智能交通、生物医学工程等领域具有广阔的应用前景<sup>[2]</sup>。

在军事国防领域,视觉跟踪技术已应用于战场和基地侦查、导弹制导、地面和空中目标攻击等系统中,包括:机载系统对地面目标的监视以及地对空系统对空中目标的跟踪等;跟踪和分析试验飞机和导弹的轨迹和姿态;等等。在人机交互方面,传统的输入方式已不能很好地满足人与计算机的更为复杂的交流需要。视觉跟踪技术是计算机识别和理解人的各种姿态和动作、实现更为自然的人机交流的关键环节之一。在智能监控领域,系统需要对摄像机采集到的视频数据进行实时处理和分析,实现自动的目标检测、识别和跟踪。在此基础上对目标的运动轨迹进行分析。目标的检测、识别、跟踪和行为分析都是计算机视觉领域的主要研究课题,而有效的视觉跟踪技术是智能监控系统中核心的技术环节。智能机器人导航主要通过视觉信息识别周围场景中的物体,实现自主运动。在机器人手眼应用中,视觉跟踪技术可用于跟踪机器人所拍摄的场景中的物体、计算其运动轨迹,并根据需要选取最佳姿势完成特定的机械操作。智能交通系统是目前交通运输领域的研究热点。车辆的实时检测、跟踪和行人行为分析是智能交通系统的关键技术。系统通过对视频序列图像的自动分析,完成车辆和行人的检测与跟踪,进一步对车辆和行人的行为给出语义描述。其中需要解决的关键问题就是对车辆和行人的分割与跟踪。在生物医学工程领域,视觉跟踪技术在超声图像和显微图像的自动分析中有广阔的应用前景。例如,显微图像序列中的杂质与噪声严重影响了对图像中运动粒子(细胞)的提取和分析,视觉跟踪技术利用图像序列中目标运动和外观变化的连续性信息可以实现对感兴趣目标的跟踪与运动分析。

从20世纪80年代至今近30年的时间里,视觉跟踪的研究进展迅速,出现了众多的视觉跟踪方法。然而,由于视觉场景的高度复杂性,包括:目标运动不确定性、阴影、光照变化、背景噪声、目标干扰、遮挡、目标变形等,使得实现一个稳健、准确和快速的视觉跟踪算法依然非常困难<sup>[3]</sup>。传统的视觉跟踪技术一般都基于目标运动和外观变化较小的光滑性假设,因而在某些复杂场景中很难实现有效的目标跟踪。这些复杂的视觉场景包括如下几个方面。

### 1. 突变的动力学特性

随着视觉跟踪技术的应用领域的不断扩大,需要跟踪的运动目标的动力学特征也呈现出复杂多样化。在实际应用的复杂场景中,运动目标经常做不连续运动,包括自身不确定性的动力学特征和在外力情况下的突变运动<sup>[13]</sup>。例如,网球运动员的突然变速启动、乒乓球受球拍撞击后的变向运动。这种极大的运动不确定性破坏了传统跟踪算法的运动连续性假设,从而容易导致跟踪失败。

## 2. 低帧率视频

为节省存储代价,降低对网络带宽的过度依赖,在视频监控和互联网实时应用领域,对低帧率视频序列中目标的在线实时跟踪显得尤为重要。与此同时,低帧率往往导致目标运动的不连续和外观的显著变化。尽管基于随机采样的视觉跟踪算法,例如,粒子滤波<sup>[4]</sup>和基于马氏链蒙特卡洛的粒子滤波<sup>[5]</sup>等方法,能够对目标状态的不确定性进行有效的建模并维持多模假设,然而,目标的突变运动和外观的显著变化使得目前的大多数跟踪算法难以实现稳健、准确、快速的目标跟踪。

## 3. 多视角场景中的镜头切换

在多视角多摄像机的视觉场景中,由于摄像机镜头的切换导致目标在空间位置、尺度和外观方面的不连续变化。这种不连续变化很容易导致传统的跟踪算法失效。现有的大部分基于随机采样的概率跟踪算法为了实现精确的跟踪任务,需要使用大量的粒子在庞大的状态空间中逼近复杂的多模分布。因而,其计算复杂度非常高,难以满足实际应用中的跟踪需要。

## 4. 目标外观的显著变化

运动的极大不确定性,可能导致目标外观的显著变化。当前视觉跟踪领域的一个研究热点是基于外观模型(Appearance Model)的自适应跟踪。传统的视觉跟踪方法采用了静态外观模型,这种模型或者被人为地定义,或者是基于第一帧图像的训练样本,但是这些方法在所跟踪的目标发生外观变化时就容易导致跟踪失败。在复杂的现实场景中,一个自适应的外观模型是获得稳健的跟踪性能的关键。近年来,尽管基于生成模型(Generative Model)和判别模型(Discriminative Model)的视觉跟踪算法在目标外观发生变化时能够比较有效地实现自适应的跟踪,但是当目标外观发生显著变化时,如何更加有效地实现稳健的视觉跟踪至今仍是一项富有挑战性的研究课题。

作为计算机视觉领域的经典的、核心的研究内容之一,视觉跟踪技术融合了图像处理与理解、模式识别、机器学习以及计算机软硬件等众多学科领域的方法与技术。计算机视觉研究具有较高的科学价值和广阔的应用前景,而视觉跟踪是计算机视觉研究领域的基础和关键环节之一,是该领域中更高层次(如行为分析、理解等)研究的基础。综上所述,对于视觉跟踪技术的研究,来源于计算机视觉应用领域的实际需求。为了解决这一课题,需要进一步探索更为有效的视觉跟踪方法和技术。因此,本课题的研究具有重要的学术和应用价值。

## ◆ 1.2 国内外研究现状及发展趋势 ◆

视觉跟踪在计算机视觉的相关领域得到了广泛的研究和应用。在智能视频监控领域,伯克利的视觉研究组在 20 世纪 90 年代开展了初步的研究工作,提出了多个有效的、能够初步处理遮挡问题的车辆跟踪算法。同时期,在智能机器人研究领域,卡耐基梅隆大学机器人研究所提出了智能控制和视觉技术相结合的视觉跟踪方法。在视频分析领域,麻省理工学院媒体实验室提出了体育视频中多目标跟踪方法。在智能人机交互领域,伊利诺伊大学香槟分校在目标识别、行为分析和视觉跟踪方面开展了一些卓有成效的工作。近年来,澳大利亚格拉斯理工大学的计算机图形与视觉研究所,以及苏黎世联邦高等工业大学的视觉研究组在视觉跟踪领域开展了基于在线学习外观模型的研究工作,并取得了一些具有重要影响的成果。此外,在国内,一些高校和研究机构近年来也在视觉跟踪领域陆续开展了这方面的研究工作,这些研究机构包括中科院自动化所、中科院计算所、清华大学、北京大学、西安交通大学、哈尔滨工业大学和华中科技大学等。

近年来,视觉跟踪问题已成为计算机视觉领域的研究热点之一,一些重要的国际期刊和国际学术会议,例如:IEEE Transactions on PAMI, IJCV, CVPR, IC-CV, ECCV 等,发表了大量有关视觉跟踪方面的研究论文。其中一些学术会议就视觉跟踪问题组织了专门的工作组进行技术讨论。如前所述,视觉跟踪之所以引起如此广泛的关注,主要是由于它能够广泛应用于社会生产生活的诸多领域。

视觉跟踪的研究内容涉及目标表示、特征选择、外观建模、目标检测识别、滤波技术、最优化方法和随机采样方法等众多相关领域的理论方法和实现技术,具有较高的研究难度和挑战性。本节首先对视觉跟踪的基本方法和技术进行简要的分类与总结,然后对视觉跟踪领域的主要研究工作进行回顾与评述。

### 1.2.1 目标表示、特征选择与外观建模

在视觉场景中,被跟踪目标的形状和外观建模是视觉跟踪的基础与关键。表示模型的优劣对跟踪算法的性能有重要影响。目标的形状表示方法可以大致归纳为如下几种<sup>[6]</sup>。

(1) 使用单个点<sup>[7]</sup>或一组点<sup>[8]</sup>表示目标。当被跟踪目标在图像中占有较小区域的时候,这种表示方法虽然简单,但通常也比较有效。

(2) 使用矩形框或椭圆<sup>[9]</sup>表示目标。这种表示方法在视觉跟踪研究中最常见,其特点是比较适合目标的仿射变换与射影变换,可以用来近似地表示刚性目标与非刚性目标。

(3) 使用轮廓<sup>[10]</sup>表示目标。轮廓通常应用在复杂非刚性目标的视觉跟踪,在这种情况下简单的几何形状难以有效地表示目标的轮廓。

(4) 基于部件或骨架的目标表示<sup>[11]</sup>。具有关节特性的目标可以认为它是由关节或部件链接在一起的,这种表示方法便于结合特定的运动模型,对具有关节体的运动目标进行精确的跟踪。此外,骨架模型可以用来表示刚性和非刚性目标,它通常通过图像变换从目标轮廓内部区域中提取而得到。

给定目标表示方式选择恰当的图像特征来刻画被跟踪目标,在视觉跟踪中具有重要的作用。特征选择通常是面向问题的,“好”的特征通常应具备较好的区分性或判别力。特征的选择通常依赖于特定的目标表示方法。例如,颜色用来作为基于直方图的外观表示的特征。而对于基于轮廓的目标表示方法,边缘通常用来作为相应的特征。通常,跟踪算法使用若干特征的组合。视觉跟踪中通常采用的基本的特征如下<sup>[6]</sup>。

### 1. 目标的颜色

在计算机视觉领域,RGB颜色空间是最为常用的,此外还有 Lab 颜色空间和 HSV 颜色空间也较为常用。这三种颜色空间的差别在于:RGB 对人眼的色彩感知上是非均匀分布的,各个分量之间并不独立;相比 RGB 空间,Lab 颜色空间是感知均匀的,而 HSV 颜色空间是感知近似均匀的。总而言之,这几种颜色空间在视觉跟踪领域的应用都在一定程度上依赖于特定的问题和任务,并不存在一种颜色空间对所有问题是普遍适用的。

### 2. 目标的边缘

在图像处理领域,通常利用目标在边界处的强度的显著变化来检测目标的边缘<sup>[12]</sup>。边缘特征一般对场景光照变化不敏感,这个特性对于视觉跟踪来说非常重要。常用的边缘检测算子包括 Canny 算子、Sobel 算子和 Laplacian 算子等。

### 3. 目标的纹理

一般认为,纹理是图像局部亮度变化特性的一种描述,这种特性包括规则性或平滑性<sup>[1]</sup>。与颜色特征不同,纹理特征对于场景光照变化不敏感。常用的纹理描述有多种,包括:灰度共生矩阵、Laws 纹理描述子和小波等。

### 4. 光流场

光流特征在计算机视觉以及图像处理中有着广泛的应用,它与运动检测和运

动估计密切相关。在物理中,运动可以用运动场描述;而在一个图像平面上,目标的运动往往是通过图像序列中不同像素点的灰度强度分布的变化来体现的。空间中的运动场变换到图像平面上就成为了光流场。计算光流的经典方法包括:Lucas-Kanade 方法<sup>[14]</sup>、Horn-Schunck 方法<sup>[15]</sup>等。

此外,在视觉跟踪领域常用的目标特征还包括:梯度方向直方图(Histogram of Oriented Gradient, HOG)<sup>[16]</sup>,局部二元模式(Local Binary Pattern, LBP)<sup>[17]</sup>,尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)<sup>[18]</sup>等。一般情况下,用于视觉跟踪的特征的选择是依赖于特定跟踪任务的,不同的跟踪问题需要人为地选择合适的特征或特征组合。近年来,自动的特征选择在视觉跟踪领域得到了广泛的研究<sup>[19,20]</sup>。基于 AdaBoost 方法的在线特征选择是此类方法的典型代表。

基于目标表示方式和选定的特征(集合),用于视觉跟踪的目标外观模型包括如下几种。

### 1. 模板

模板(Template)是视觉跟踪中最为常用的外观表示方法。它一般使用简单的几何形状(如椭圆、矩形等)来表示<sup>[21]</sup>。特定设计的模板,例如基于分块表示的模板<sup>[22]</sup>,可以同时包含目标的空间和外观(颜色等特征)信息。

### 2. 概率分布

目标外观的概率分布是另一种较为常用的外观表示方式。参数化的概率分布表示可以采用单个高斯分布<sup>[23]</sup>或混合高斯分布<sup>[24]</sup>;而非参数的表示方式主要是核密度估计<sup>[25,9]</sup>。

### 3. 主动外观模型

在主动外观模型(Active Appearance Model, AAM)<sup>[11]</sup>中,目标的形状由一组特征点来描述,这组特征点位于目标边界或目标区域内部。每一个特征点对应于一个外观特征(颜色、边缘、纹理等)向量。AAM 主要特点在于它同时把目标的形状和外观特征模型化。

### 4. 多视外观模型

在复杂的场景中,当运动目标的外观变化较大且遮挡频繁发生时,采用多摄像机的多视角(multi-view)跟踪是个不错的选择。与此相对应的多视外观模型,用来模型化目标的不同视角。多视外观模型通常考虑从不同视角的样本中生成一个子空间<sup>[26]</sup>,或者考虑从不同视角的样本中训练得到一个判别分类器<sup>[27]</sup>。

## 1.2.2 视觉跟踪方法

前面的章节已经对目标表示和外观建模做了简要的归类 and 介绍。概括地说,视觉跟踪问题可以分解为:目标表示、外观建模和跟踪框架(方法)。在视觉跟踪研究领域,外观建模通常分为:前景建模和前景背景联合建模。与此相对应的外观模型可以归结为两大类:原生模型和判别模型。原生模型试图构建一个紧凑的外观模型来描述可能的目标外观变化。判别模型不同于原生模型,旨在将目标外观与背景区分开来。大多数判别外观模型是通过机器学习的方法学习一个两类分类器(Binary Classifier)来区分跟踪的目标和背景。至于跟踪框架(方法),大致可以归类为:确定性的跟踪方法和基于采样的随机跟踪方法两种。在视觉跟踪应用中,跟踪方法通常与外观建模方式有关。在这一节里,我们对视觉跟踪的主要方法做一个简要的回顾与总结。

### 1. 原生外观模型

#### 1) 基于目标模板

目标模板既可反映目标的空间信息,也可包含目标特征(颜色、纹理等)信息。在基于模板的视觉跟踪研究方面,较早期的工作是1981年Lucas和Kanade<sup>[14]</sup>的图像配准方法。这一开创性研究工作的基本思路是:将首帧图像的目标区域(矩形)设定为模板,在后续的图像序列帧中搜寻与设定模板匹配度最高的区域。采用的匹配准则是目标模板和候选矩形区域对应像素灰度值差的平方和(Sum of Squared Differences, SSD)。在文献[28]中,只能获得平移运动参数的Lucas-Kanade方法被扩展到可以获得仿射变换参数,并应用于特征点跟踪。在这一研究工作中,作者还提出了在评估特征点质量的方法:计算当前帧特征点与首帧特征点的不相似度,当不相似度大于一定阈值时认为该特征点不再可靠进而被丢弃。Shi和Tomasi<sup>[28]</sup>认为,一个“好”的特征点的选择取决于这一特征点是否能给跟踪算法带来更好的跟踪性能。此外,Baker和Matthews<sup>[29]</sup>对经典的Lucas-Kanade方法做了全面的回顾,并对该框架下的各种算法进行了较为全面的评述。

视觉跟踪过程中使用静态不变的外观模板是不可行的,这是因为目标外观在视觉场景中受各种内部或外在因素影响而不断变化。因而,为了获得稳健的跟踪性能,模板更新是一个必须考虑的重要问题。一种非常简单的更新策略是使用当前跟踪的结果对目标模板进行更新。这一更新策略的问题在于跟踪过程中误差累积将最终导致目标漂移(Drifting)并导致跟踪失败。针对这一问题,Matthews等人<sup>[30]</sup>给出了一种解决模板更新问题的方案:在跟踪过程中保留首帧的模板,用当



前帧的跟踪结果对目标模板进行更新,并和首帧模板进行对齐。在文献[31]中,作者提出用一个参数模型来建模目标的运动,用一个子空间来建模场景中目标光照的变化,用稳健估计方法来建模目标的遮挡。此外,Nguyen 和 Smeulders<sup>[32]</sup>针对遮挡情况下的模板匹配问题进行了深入的研究,并提出对目标模板中的每一像素分别调用卡尔曼滤波器进行更新。该方法对于场景中目标光照变化具有较好的稳健性。在文献[33]中,Jepson 等人提出了一种可以处理光照变化和外观变形的跟踪方法。它融合目标稳定部分、帧间差别部分和噪声影响部分的 WSL 三分量模型。该方法使用小波作为目标特征,采用基于在线的 EM 算法更新模型参数。在这一模型中,目标的仿射运动估计是通过目标稳定分量进行。该模型的主要缺陷在于稳定分量部分,当背景变化较缓慢时,部分背景也将被估计成目标的一部分。同样基于在线估计参数的方式,Han 和 Davis<sup>[34]</sup>提出使用高斯混合模型对目标区域内的像素进行在线建模。在该模型中,作者通过自适应带宽的 Mean Shift 算法在线自动估计各个模型参数。使用单一目标模板的跟踪方法在实际跟踪任务中应用广泛,但在目标姿态变化较大时往往导致跟踪失败。

## 2) 基于目标特征的概率分布

这种目标外观建模方法通常采用目标内部的特征(颜色、边缘等)的概率分布来描述目标。这种建模方法的主要特点在于它忽略了目标区域的空域信息,因而能用于处理目标的非刚性形变。这方面的早期研究工作包括: Birchfield<sup>[35]</sup>采用一个椭圆表示人脸并使用颜色直方图对人脸进行跟踪。同样基于颜色直方图表示,Comaniciu 等人<sup>[36]</sup>结合核密度估计方法,在计算颜色直方图时使用核函数对直方图进行加权使得越靠近目标中心的像素权重越大,从而在一定程度上避免了嘈杂背景的干扰。此外,作者还提出了处理目标尺度变化的方法,它通过对当前跟踪得到的尺度进行一定范围的伸缩(例如,5%,10%等),然后分别计算相似度,相似度最大的作为最终的目标尺度。Comaniciu 等人的方法为目标(位置和尺度)跟踪提供了一种新的思路。该方法的一个缺陷是基于核函数的颜色直方图并不能很好地描述目标的尺度,往往在匹配时小尺度比大尺度具有更高的相似度,从而造成尺度跟踪漂移现象。之后,Collins<sup>[37]</sup>提出了一个基于尺度空间的处理方法。该方法采用一个尺度核函数,在空间位移和尺度变化上交替代调用 Mean Shift 算法,从而获得目标的空间位置和尺度。

在使用颜色直方图方面,Mean Shift 等确定性跟踪方法在多模复杂场景下往往易于陷入局部模式从而导致跟踪失败。针对这一问题,在文献[38]和[39]中提出了结合随机跟踪方法和颜色直方图的处理方案。在文献[38]中,Pérez 等人提出



在粒子滤波框架下基于颜色直方图的跟踪方法,该方法对复杂多模场景具有一定的稳健性。当使用大量粒子时,对颜色直方图的计算是个很大的计算负担。为了减少计算代价,Porikli<sup>[40]</sup>提出使用积分直方图。与此类似,Adam等人<sup>[41]</sup>提出使用基于分块目标表示的颜色直方图,其中每一个子块都用颜色直方图来表示。基于这种外观模型的跟踪方法对局部遮挡具有较好的稳健性。此外,文献<sup>[42]</sup>和<sup>[46]</sup>还给出了处理更多运动参数(包括位移、尺度、旋转方向等)的解决方案。这些方案大多基于椭圆表示的颜色直方图模型,采用核函数进行跟踪。

对于非刚性形变的目标跟踪,基于目标特征的概率分布的跟踪方法能够获得较好的跟踪性能,具有一定的稳健性。不过,当场景中具有多个相似特征的目标,或在嘈杂的背景下,这种基于概率分布的外观模型极易导致跟踪漂移和失败,其根本原因在于这一外观建模方式忽略了目标的空间分布信息。

### 3) 基于子空间

基于子空间的跟踪方法通常是通过训练样本学习(在线或离线)构造(通常通过主成分分析)一个子空间模型。该方法的主要优点在于它同时兼顾了目标的光照变化与姿态变化,这使得它能够处理较为复杂场景下的视觉跟踪任务。

基于离线构造子空间方面,Black和Jepson<sup>[26]</sup>给出了多视角子空间构造方法。值得注意的是,该方法通过深入分析光流计算中的灰度强度不变假设,提出了子空间的一致性假设。作者认为,多视角目标观测可以采用相同的特征基(集)进行重建,使得重建图像与原始图像具有一致的灰度强度分布。基于这一假设可以实现不依赖于特定视角的视觉跟踪任务。

确实,基于离线的构造方法尽管也能在一定程度上描述和适应目标的外观变化,但是基于离线的方式在复杂的场景中往往由于累积的跟踪误差而导致跟踪漂移。针对这一问题,Ross等人<sup>[47]</sup>提出了基于在线学习渐进构造子空间基的算法。在该算法中,子空间基是随着顺序获得的外观样本的逐渐增加而不断提升其准确性的。作者的实验表明,该算法在处理显著的光照、姿态和尺度变化等方面具有较强的稳健性。此外,在Log欧氏黎曼度量之下,Li等人<sup>[48]</sup>提出了一种基于SLAM模型的在线跟踪算法来处理目标姿势、尺度自适应变化的情况。Wu等人<sup>[49]</sup>基于渐进的协方差张量学习提出了一种姿势和外观自适应的实时跟踪算法。Li等人<sup>[50]</sup>通过自适应地更新样本均值和特征基,渐进地学习一个低阶张量特征空间表示来建模目标的外观模型,提出了一种在线的张量子空间学习算法来实现有效的目标跟踪。Yang等人<sup>[54]</sup>提出一种基于子空间的外观自适应策略,它基于三个约束,即负样本数据、自底向上的成对正负样本数据以及自适应的目标动力学特征。