



视频跟踪： 理论与实践

(英) Emilio Maggio 著
Andrea Cavallaro

马东玺 张勇 译
曹玉君 李欣

WILEY



西安电子科技大学出版社
<http://www.xduph.com>

视频跟踪 实践

(英) Emilio Maggio 著
Andrea Cavallaro

马东玺 张 勇 译
曹玉君 李 欣



西安电子科技大学出版社

图书在版编目(CIP)数据

视频跟踪:理论与实践/(英)埃米利奥·马乔,(英)安德烈·卡瓦拉罗著;马东玺等译.

—西安:西安电子科技大学出版社,2017.5

ISBN 978-7-5606-4459-2

I. ① 视… II. ① 埃… ② 安… ③ 马 III. ① 视频系统—监视控制 IV. ① TN94

中国版本图书馆 CIP 数据核字(2017)第 081300 号

Video tracking: theory and practice

Emilio Maggio, Andrea Cavallaro

ISBN:978-0-470-74964-7(cloth)

Copyright © 2011 by John Wiley & Sons, Ltd. All rights reserved.

All Rights Reserved. This translation published under license. Authorized translation from the English language edition, published by John Wiley & Sons. No part of this book may be reproduced in any form without the written permission of the original copyrights holder copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

本书中文简体中文字版专有翻译出版权由 John Wiley & Sons, Ltd. 公司授予西安电子科技大学出版社。未经许可,不得以任何手段和形式复制或抄袭本书内容。

本书封底贴有 Wiley 防伪标签,无标签者不得销售。

陕西省版权局著作权合同登记图字:25-2016-0291 号

策划编辑 刘小莉

责任编辑 雷鸿俊

出版发行 西安电子科技大学出版社(西安市太白南路2号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfxb001@163.com

经 销 新华书店

印刷单位 虎彩印艺股份有限公司

版 次 2017年5月第1版 2017年5月第1次印刷

开 本 787毫米×1092毫米 1/16 印张 15

字 数 259千字

印 数 1~2000册

定 价 45.00元

ISBN 978-7-5606-4459-2/TN

XDUP 4751001-1

*** 如有印装问题可调换 ***

内 容 简 介

本书涵盖了近 30 年视频跟踪技术的发展,介绍了视频目标检测跟踪及其应用的最新研究成果和动态,全面系统地介绍了当今世界视频跟踪技术的最新研究进展,重点介绍了预处理技术、图像目标检测、视频目标检测、均值移位跟踪方法、基于粒子滤波的跟踪方法等方面的技术和应用。本书内容全面、新颖,写作风格上深入浅出、思路清晰、通俗易懂。为了让读者更好地掌握算法,书中还提供了一些重要的框图和简单算法的实现代码。

本书不仅具有极高的学术价值,对工程实践也具有指导作用。

译 序

图像识别与跟踪是人工智能领域的一个重要分支。基于图像识别跟踪技术，我们有了无人驾驶汽车、无人机、机器人，为我们带来方便；有了先进的安防设施、精确制导武器，为我们带来安全；有了医疗图像的自动判读和处理，为我们带来健康；还有目前风生水起的增强现实(AR)技术，打通了虚拟与现实的边界，为我们带来快乐。通过图像跟踪技术，我们可以制作逼真的动画，让无人机穿梭树林、精确悬停……正是由于这些技术的发展，我们的世界才更加精彩，更加有趣。

译者长期从事武器装备科研，与视频跟踪技术始终难舍难分。译者所从事的工作中，从导弹导引头到无人机侦察吊舱，从武器站的光电火控到排爆机器人抓取规划，均离不开视频跟踪技术。

在科研工作中，译者感觉缺少一本关于视频跟踪技术的针对性的参考书籍，直到看到英国 Vicon 公司的 Emilio Maggio 博士和伦敦玛丽女王大学的 Andrea Cavallaro 教授合著的《视频跟踪：理论与实践》一书。两位作者分别是具有丰富实践经验的一线公司的计算机视觉科学家和著名大学的教授，在目标跟踪和分类、贝叶斯滤波、图像稀疏、视频编码以及多传感器系统的多模内容分析等方面具有丰富的实践经验和扎实的理论基础。他们合作的这本书涵盖了近 30 年视频跟踪技术的发展，介绍了视频目标检测跟踪及其应用的最新研究成果和动态，全面系统地介绍了当今世界视频跟踪技术的最新研究进展，重点介绍了预处理技术、图像目标检测、视频目标检测、均值移位跟踪方法、基于粒子滤波的跟踪方法等方面的技术和应用。该书内容全面、新颖，写作风格上深入浅出、思路清晰、通俗易懂。为了让读者更好地掌握算法，本书中还提供了一些重要的框图和简单算法的实现代码。本书不仅具有极高的学术价值，对工程实践也具有指导作用。

在工作之余，我们将该书翻译成中文，供读者学习参考。由于译者水平有限，疏漏之处可能在所难免，还请读者批评指正。

感谢国家自然科学基金(51305455)的支持。感谢国防科技大学 ATR 实验室的孙刚博士校稿。感谢军械技术研究所张勇高工,国防科技大学曹玉君讲师、李欣工程师,第四军医大学的张超博士,大家共同努力,让本书得以成形。感谢西安电子科技大学出版社雷鸿俊编辑和刘小莉编辑的辛勤工作,让本书顺利出版。

译者

2017年1月

于石家庄

序

我很荣幸为 Andrea Cavallaro 教授、Emilio Maggio 博士的著作《视频跟踪：理论与实践》作序。该书内容全面、及时，写作极为用心，在视频跟踪领域具有重要的参考价值。说它内容全面，因为它涵盖了发展于 20 世纪 60 年代初期的主要应用于航天活动中的点目标跟踪以及用于计算机视觉和图像处理的扩展目标跟踪理论和方法。说它及时，因为它涵盖了近 15 年来大量的视频跟踪方面的文献，为学习视频跟踪技术的发展提供了一站式的资源。此书深入浅出，思路清晰，能够帮助学生、研究人员以及工程师们很快地了解在视频跟踪这一重要领域已经做了什么，还有什么需要去做。很多计算机科学家和电子工程师都热衷于计算机视觉的研究，但是他们对随机过程、检测和估计理论的熟悉程度有很大差异。该书充分考虑到这一点，即使没有全部学过随机过程研究生课程的计算机科学家们也很容易接受，同时对于了解相关理论但不太了解应用的电子工程师而言也能够有足够兴趣。

最初的视频跟踪主要研究的是用于军事领域的红外点目标跟踪。早期的 α - β 跟踪器很快被神奇的卡尔曼滤波器(离散和连续)及其变体所取代。毫不夸张地说，现在大多数的视频跟踪系统都是在卡尔曼滤波的基础上建立的。基于卡尔曼滤波的跟踪器的讨论，在很多书籍中都有涉及，如 Anderson、Moore、Bar-Shalom、Gelb、Blackman、Popli 等人的著作。Jaswinski 所著的一本经典的跟踪书籍，描述了跟踪器设计的基础知识。线性卡尔曼滤波有效地解决了含有高斯噪声的线性跟踪问题，但是却很难解决非线性或者非高斯跟踪问题。而扩展卡尔曼滤波器、迭代扩展卡尔曼滤波器和非线性连续跟踪器的出现为解决非线性/非高斯跟踪开辟了新的途径。

从简单的点目标跟踪发展到人脸、人体和车辆等扩展目标跟踪，给视频跟踪领域带来了新的挑战，例如几何参数(姿态、清晰度)和光度参数(光照、景象)变化带来的复杂性。在设计跟踪器时需明确在这些变量中哪些是变化的参数，哪些是不变的参数。基于蒙特卡罗马尔可夫链技术结合几何和光度的变化可以设计先进的跟踪器，现在比较普遍的粒子滤波跟踪器就是其中一员。近来，很多文献中都涉及了粒子滤波视频跟踪的理论和应用的研究，这些内容将

在此书的后几章进行描述。

本书第1章介绍了视频跟踪理论和鲁棒视频跟踪器的设计要求，给出了章节概要、相关的数学符号，以及单目标和多目标跟踪器的相关公式，并阐述了视频跟踪研究具有的重要的应用价值。第2章介绍了视频跟踪在娱乐、医疗、监控和机器人上的应用，详细说明了无人平台目标跟踪新技术的发展。第3章总结了影响视频跟踪的一些因素，如亮度、色彩、梯度、兴趣区域和目标模型，分析了视频跟踪器对以上因素的鲁棒性，同时本章还介绍了图像预处理以及背景差的算法。

本书第4章总结了跟踪模型的形状、变形和外观，并且给出了处理这些表象变化的方法。在卡尔曼滤波的设计中定义状态和量测公式或者在粒子滤波中导出概率模型时，跟踪目标的外观和移动等表象是十分重要的。第5章主要讲述了此书的最重要部分即单目标的跟踪算法，而且对卡尔曼滤波和粒子滤波的理论与实现描述得深入浅出、通俗易懂，很受读者欢迎。

多特征融合技术能够提高跟踪性能，一直以来备受学者的关注。融合运动和物体的强度特征有助于在复杂背景、混乱环境下跟踪模糊的目标。第6章提出了融合方法，针对目标的形状、运动、行为等其他特征，设计了鲁棒跟踪器。当实现多目标跟踪时，使它们关联起来便成为一个至关重要的问题。早在20多年前出现的数据关联概率的方法，随着现代图论技术的发展越来越受到人们的关注，第7章对该方法进行了论述。因为当多目标跟踪或者跟踪一个多特征的运动目标时，一些特征将会消失，取而代之的将会是一些新的特征，所以在本章中提出了如何处理特征的“新生和死亡”，而这对长时间跟踪问题是非常重要的。本章详细介绍了此难点问题。

20世纪90年代早期，计算机视觉的研究者们将目标检测、识别、跟踪技术和上下文结合起来。第8章讨论了上下文在提高视频跟踪性能中的作用，并且提出了如何把确定目标可能存在位置的上下文信息融合在视频跟踪器中。跟踪器的设计实用方面之一是能为跟踪系统的工作提供性能界限。许多组织研讨会意识到用指标和通用数据库评价跟踪器的性能是一个繁琐但很重要的问题。虽然离散理论分析已经应用在评价跟踪器性能上，但是大部分现有的经验方法都是基于地面实况数据的。第9章介绍了跟踪器评估的发展历程和应用。

最后，本书简要地介绍了未来需要解决的问题，对比了此书论述的几种跟踪器的性能，以及对深度阅读此书的一些建议。

我很喜欢这本书，因为在本书中涵盖了近 30 年视频跟踪技术的发展，其中不仅有大多数研究员创造的科学学术价值，而且还强调科学、技巧和技术在视频跟踪领域的应用，充分考虑了图像处理和计算机视觉阅读群体的背景和需求，所以此书可以帮助学生、研究人员和工程师们入门并深入了解视频跟踪这一重要领域。

Rama Chellappa
马里兰州帕克大学

作者简介

Emilio Maggio 是动作捕捉市场的领导者——英国 Vicon 公司的计算机视觉科学家，他的研究方向是目标跟踪和分类、贝叶斯滤波、图像稀疏以及视频编码等。在 2003 年和 2007 年，他分别在瑞士联邦技术学院(EPFL)的信号处理研究所和三菱研究实验室(MERL)做访问学者，在 2005 年和 2007 年曾先后两次获得了 IEEE ICASSP 优秀论文奖，并曾在 2002 年赢得过 IEEE 计算机学会国际设计竞赛。

Andrea Cavallaro 是伦敦玛丽女王大学多媒体信号处理的教授。他的研究方向是目标跟踪和多传感器系统的多模内容分析。他获得了 2004 年美国电信实验室研究奖学金；2007 年获得了皇家学院的工程教学奖；2005 年、2007 和 2009 年三篇论文分别在 IEEE ICASSP 获奖；2009 获得了 IEEE AVSS 优秀论文奖。他是《IEEE 信号处理杂志》、《IEEE 多媒体汇刊》、《IEEE 信号处理汇刊》的副编辑。

主要译者简介：

马东玺，生于 1980 年，男，博士，陆军某研究所工程师，本硕博毕业于国防科技大学，长期从事光电装备科研工作。获军内科技进步一等奖 1 项，二等奖 1 项，三等奖 3 项。主持国家自然科学基金 1 项，军内科研多项。

前 言

视频跟踪就是在图像序列中估算目标位置随时间变化的技术。本书首次为视频跟踪这个新兴的领域提供了全面的知识构架，介绍了基本算法的发展及其应用，介绍并讨论、验证了最新的视频跟踪算法。

本书首先介绍基本问题的定义，介绍了以往和现在视频跟踪的主要应用，列举了目前比较流行的视频跟踪器，例如基于相关度和梯度下降最小的跟踪器。书中应用实例，图文并茂地说明了运用确定性方法设计跟踪器的优势和局限性。本书提出了更加准确有效的视频跟踪算法，并讨论了基于贝叶斯递归结构近代算法在真实情景跟踪中的应用。由于把基本的数学模型转化为有效真实模型时方案的选择是非常重要的，所以接下来本书讨论了该方案的设计选择和实现。为了让读者更好地掌握算法，本书中还提供了一些重要的框图和简单算法的实现代码。

本书第1章首先介绍了视频跟踪问题，并把此问题划分为五个主要的逻辑任务。然后提供了视频跟踪中问题的描述，最后讨论了视频跟踪的难点。第2章介绍了一些当代视频跟踪的应用，涉及的主要应用领域有媒体制作、医学数据处理、监视系统、智能商业、机器人、远程协作、互动游戏和艺术。第3章进一步综述了图像获取过程并指出了描述目标特点的相关术语。第4章讨论了多种形状近似策略和外观建模技术。第5章对定位算法进行了分类，并且比较了单假设和多假设策略。第6章讨论了多特征融合目标跟踪，从跟踪级和特征级两个方面说明了融合技术的优点与缺点，并且提出了在融合中如何量化特征可靠性的优先级。第7章把本书前面几章的内容延伸到数量可变的跟踪。为了更好地说明该方法，本章重点叙述了多假设数据关联算法在视频监控中的应用，并讨论和评估了一个基于有限集统计的多目标视频跟踪器。以该跟踪器为例，第8章讨论了如何构造场景模型以改善视频跟踪器的性能，并指出了如何运用主动式和交互式的策略对图像中感兴趣的部分进行学习。第9章介绍了评估视频跟踪器及其效果的协议，并为读者提供了跟踪器全面的工作指标和大量

的评价数据集。最后，在附录中讨论了本书所选择的跟踪算法的对比结果，在后记中简要介绍了视频跟踪的当前发展方向和未来的挑战，并为读者提供了深入阅读书籍的列表。

本书主要针对的群体有研究生、研究员和工作者。而工作者主要从事多种视觉的解释应用、智能环境、行为建模、机器人、视频标注等方面的研究，或者为监控、运动获取、虚拟现实和医学图像分析的程序开发人员。

我们创建了本书的支持网站 www.videotracking.org 为读者提供本书的辅助材料。该网站为视频跟踪研究者提供了全面的软件算法列表，为教师提供了与本书知识点相关的课堂讲解演示文稿。

Emilio Maggio
Andrea Cavallaro
于英国伦敦

致 谢

我们非常感谢为本书出版做出贡献的每一个人。本书经过了长达十年之久的构思、试验和教学才得以出版。一直以来我都非常感激我的同事、学生和朋友们，和他们一起工作是我们宝贵的经历。特别地，要对 Murtaza Taj 和 Elisa Piccardo 给予诚恳的感谢。特别感谢 Sushil Bhattacharjee 对本书初稿具有启发性的讨论和改进建议。感谢 Samuel Pachoud、Timothy Popkin、Nikola Spriljan、Toni Zgaljic 和 Huiyu Zhou 提供了视频动态测试，并对许多示例给予了重要的支持。同时感谢那些为视频跟踪评估提供了大量参考数据的同事，这些数据可在 www.spevi.org 网站获得。感谢英国物理工程科学研究委员会 (EPSRC)、欧委会的 MOTINAS (EP/D033772/1)、APIDIS (ICT - 216023) 工程对本书工作的资助。同样感谢 Rama Chellappa 为本书作序。特别感谢 Nicky Skinner、Alex King、John Wiley & Sons 公司的 Clarissa Lim 和 Aptar 公司的 Shalini Sharma 对本书编写和出版的大力支持，并对他们积极主动的合作表示由衷的感谢。最后，我们真诚地感谢 Silvia 和 Maria 对我们的鼓励以及自始至终的支持。

本书中使用符号的意义

(w, h) : 目标宽度和高度

E_0 : 单目标量测空间

E_s : 单目标状态空间

E_1 : 像空间

I_k : k 时刻像

$M(k)$: k 时刻目标个数

$N(k)$: k 时刻量测个数

X_k : 多目标状态

Z_k : 多目标量测

(u, v) : 目标质心

X_k : k 时刻有效的轨迹序列

Z_k : k 时候由分配给有效轨迹的量测序列

x : 形成目标轨迹的状态集合

$D_{k|k}(x)$: k 时刻假设概率密度

$F(E)$: E 中元素的所有有限子集集合

$f_{k|k-1}$: 状态传递概率密度函数

k : 时间(帧)

$p_{k-1|k-1}$: 先验概率密度函数

$p_{k|k-1}$: 预测概率密度函数

$p_{k|k}$: 后验概率密度函数

q_k : 重要性采样函数

x_k : 目标在 k 时刻的状态

$y_{a,b}$: 元素集合(如标量和向量) $\{y_a, y_{a+1}, \dots, y_b\}$

z_k : 单目标在 k 时刻的量测

g_k : 似然函数

目 录

第 1 章 视频跟踪概述	1	3.3.2 光度颜色不变量	33
1.1 引言	1	3.3.3 梯度和导数	35
1.2 视频跟踪的设计	2	3.3.4 拉普拉斯算子	40
1.2.1 问题描述	2	3.3.5 运动	41
1.2.2 视频跟踪算法的主要组成	4	3.4 中级特征	42
1.3 问题描述	6	3.4.1 边缘	42
1.3.1 单目标跟踪	6	3.4.2 兴趣点与兴趣区域	43
1.3.2 多目标跟踪	8	3.4.3 均匀区域	48
1.3.3 定义	8	3.5 高级特征	52
1.4 交互式跟踪与自动跟踪的对比	9	3.5.1 背景模型	53
1.5 小结	10	3.5.2 目标模型	54
第 2 章 应用	11	3.6 总结	55
2.1 引言	11	参考文献	55
2.2 媒体制作和增强现实	11	第 4 章 目标描述	62
2.3 医疗应用和生物研究	14	4.1 引言	62
2.4 监控和智能商业	15	4.2 形状描述	62
2.5 机器人和无人驾驶	17	4.2.1 基本模型	62
2.6 远程协作和互动游戏	17	4.2.2 关节模型	64
2.7 装置艺术及表演	18	4.2.3 变形轮廓模型	64
2.8 小结	19	4.3 外观描述	65
参考文献	19	4.3.1 模板	66
第 3 章 特征提取	23	4.3.2 直方图	68
3.1 引言	23	4.3.3 外观变化的应对	73
3.2 从光线到有用的信息	23	4.4 总结	73
3.2.1 光线测量	23	参考文献	74
3.2.2 目标外观	25	第 5 章 定位	77
3.3 初级特征	27	5.1 引言	77
3.3.1 颜色	27	5.2 单假设定位方法	77

5.2.1 梯度跟踪器	77	7.6.4 重采样	128
5.2.2 贝叶斯跟踪和卡尔曼滤波	81	7.6.5 粒子群	132
5.3 多假设定位方法	84	7.6.6 例子	135
5.3.1 网格采样	85	7.7 小结	139
5.3.2 粒子滤波	86	参考文献	139
5.3.3 混合方法	89	第 8 章 上下文模型	143
5.4 总结	95	8.1 引言	143
参考文献	95	8.2 上下文模型的跟踪问题	143
第 6 章 融合技术	98	8.2.1 上下文信息	143
6.1 引言	98	8.2.2 上下文的影响	144
6.2 融合策略	98	8.3 新生和杂波密度估计	146
6.2.1 跟踪级融合	98	8.3.1 新生密度	146
6.2.2 测量级融合	100	8.3.2 杂波密度	151
6.3 粒子滤波中的特征融合	102	8.3.3 上下文反馈跟踪	154
6.3.1 似然融合	102	8.4 小结	155
6.3.2 多特征重采样	102	参考文献	155
6.3.3 特征可靠性	104	第 9 章 性能评估	157
6.3.4 时间平滑	107	9.1 介绍	157
6.3.5 举例	107	9.2 分析方法与经验方法	158
6.4 小结	109	9.3 地面实况	159
参考文献	109	9.4 评价分数	161
第 7 章 多目标管理	112	9.4.1 定位分数	161
7.1 引言	112	9.4.2 分类分数	163
7.2 量测验证	113	9.5 跟踪器的对比	166
7.3 数据关联	114	9.5.1 目标寿命	166
7.3.1 最近邻域法	114	9.5.2 统计学意义	167
7.3.2 图匹配	115	9.5.3 重复性	167
7.3.3 多级假设跟踪	119	9.6 评估协议	168
7.4 随机有限集的跟踪	121	9.6.1 低级协议	168
7.5 概率假设密度滤波	123	9.6.2 高层协议	171
7.6 粒子 PHD 滤波	124	9.7 数据集	175
7.6.1 动态观察模型	125	9.7.1 监控	175
7.6.2 新生和杂波模型	127	9.7.2 人机交互	179
7.6.3 重要性采样	127	9.7.3 运动分析	184
		9.8 小结	187

参考文献	187	A.3.4 计算复杂性	206
附录 A 结果对比	190	A.4 PHD 滤波	210
A.1 单色直方图与结构直方图	190	A.4.1 实验装置	210
A.1.1 实验装置	190	A.4.2 详述	210
A.1.2 详述	190	A.4.3 失效模式	213
A.2 定位算法	195	A.4.4 计算成本	214
A.2.1 实验装置	195	A.5 上下文建模	216
A.2.2 详述	196	A.5.1 实验装置	216
A.3 多特征融合	200	A.5.2 详述	216
A.3.1 实验装置	200	参考文献	219
A.3.2 可靠性分数	202	后记	221
A.3.3 自适应与非自适应跟踪	203	延伸阅读	222