

分类号 TP18

密级

UDC

编号

中国科学院研究生院 博士学位论文

复杂场景下目标的跟踪与姿态估计

张笑钦

指导教师 胡卫明 研究员

中国科学院自动化研究所

申请学位级别 工学博士 学科专业名称 模式识别与智能系统

论文提交日期 2010 年 3 月 论文答辩日期 2010 年 4 月

培养单位 中国科学院自动化研究所

学位授予单位 中国科学院研究生院

答辩委员会主席

Object tracking and pose estimation under complex scenes

Dissertation Submitted to

Institute of Automation, Chinese Academic of Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Engineering

by

Zhang Xiaoqin

Pattern Recognition and Intelligent System

Dissertation Supervisor: Prof. Wei-ming Hu

独创性声明

本人声明所递交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名:_____日 期: _____

关于论文使用授权的说明

本人完全了解中国科学院自动化研究所有关保留、使用学位论文的规定，即：中国科学院自动化研究所有权保留送交论文的复印件，允许论文被查阅和借阅；可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签 名:_____导师签名:_____日 期: _____

摘要

目标跟踪在目标运动的视觉分析中占有重要的地位，属于视觉的中层部分。利用对目标的跟踪，可以方便地获得目标的运动参数，为后续的高层的行为理解和识别奠定了基础。目标跟踪虽然是计算机视觉领域一个重要的研究方向和研究热点，但是目前仍然有很多理论与技术问题尚待解决，特别是跟踪过程中噪声干扰、剧烈运动、光照变化、遮挡等复杂问题。目标姿态估计在目标运动的视觉分析中属于高层部分，它是根据图像信息推断出目标物体（主要针对人体，手势）的三维方位指向。姿态估计在智能视觉监控、人机交互和视频运动分析等很多领域都有着重要的应用。

本文的工作以复杂场景下鲁棒的目标跟踪与姿态估计为目标，分别对以下三个问题进行了深入的探讨和分析：（1）基于样本分布的表现建模；（2）跟踪过程中样本的退化；（3）姿态估计中高维状态空间的解析。大量的实验表明我们方法的有效性和鲁棒性。论文的主要工作和贡献如下：

- ① 提出一种基于图嵌入判别学习的框架来学习出正样本的模式空间，以及正负样本之间的判别空间。基于这两个模式空间的学习，我们构建起鲁棒的目标表现模型。与传统基于子空间学习方法相比，该方法有效地嵌入了样本的分布特性，能够得到更符合实际的目标模式子空间，因此对噪声、局部遮挡更为鲁棒。
- ② 提出两种结合确定性跟踪框架和随机性跟踪框架的方法：（1）核贝叶斯跟踪框架；（2）基于奇异值分解的卡尔曼粒子滤波框架。（1）首先将核方法应用于当前帧，获得对运动方向以及偏移量的估计，再将这些信息作为启发式先验信息加入到贝叶斯滤波框架中，从而有效结合确定性方法与随机性方法的优点，并且该框架是一种通用的跟踪框架，任何形式的表现模型都可以在此框架下应用。与基于核方法的跟踪器（确定性跟踪方法）以及基于贝叶斯方法的跟踪器（随机性跟踪方法）相比，该跟踪框架有效地提高了跟踪的效率和精确度。（2）首先对前一帧跟踪样本的协方差矩阵进行奇异值分解，再利用 Unscented 变化（一种确定性的变化）获得预测样本，然后用卡尔曼滤波器对预测样本进行滤波，为最终的粒子滤波提供运动估计分布。与粒子滤波和基于 Unscented 的卡尔曼粒子滤波相比，我们的方法更为鲁棒。
- ③ 提出了一种基于序列粒子群优化的跟踪算法框架。该框架首先通过初始化一系列随机粒子样本，然后对这些粒子集进行时序上的传播以及优化迭代，最后对进化的粒子进行收敛性判断。我们从理论上证明了，序列粒子群优化框架实质上是一个基于分层重要性采样的粒子滤波过程，使得采样结果逼近了直接采样于最佳重要性分布的结果。我们还将序列粒子群优化的概念扩展到多种群上，将多目标跟踪中目标物体的交互和遮挡用种群间的竞争和排斥来模拟，实现了有效的多目标跟踪。与其他两种多目标跟踪的算法相比，我们的算法在 PETS2004 数据库上获得了最佳的实验结果。
- ④ 提出一种新颖的姿态估计算法，该方法从以下三个方面出发来解决姿态估计所面临的维数灾难问题：（1）采用树形结构来对人体拓扑结构进行建模并进行解析；

(2) 优化过程采用两种模式：扩散和跳跃；(3) 提出一种数据驱动的马尔可夫链蒙特卡洛方法来进行优化。该方法可以有效地将人体各部分的拓扑关系加入到优化过程，并通过表观数据的驱动来压缩搜索空间，从而大大提高了收敛速度（10 倍左右）和精度。

关键词： 图嵌入学习，粒子群优化，核贝叶斯，卡尔曼滤波，粒子滤波，目标跟踪，姿态估计

Abstract

Object tracking plays an important role in visual analysis and understanding of human motion, it is the intermediate-level vision part. Object tracking is to detect, locate and track moving objects in video sequences captured by cameras. Through object tracking, we obtain the motion parameters and the trajectories of the object, which are the foundation of the high-level activity understanding and recognition. Recent years have witness great advance in tracking literature, it still remains a great challenging task due to clutter background, dynamical motion, appearance variations, and especially the occlusion problems. Pose estimation, a high-level vision task, is the task of determining the location, orientation and scale of each object. It is important for many vision understanding applications, e.g. visual interactive gaming, immersive virtual reality, visual surveillance, content-based image retrieval, etc.

This thesis focuses on robust object tracking and pose estimation under complex scenes. Specifically, we mainly discuss the following three sub-topics: (1) graph embedding learning based appearance modeling; (2) sample impoverishment in tracking process; (3) high dimensional state space parsing in pose estimation. The main contributions of our work are summarized as follows:

- ① We propose a graph embedding framework to simultaneously learn the subspace of the target and its local discriminative structure against the background. Then a robust appearance model is constructed based on the learnt subspaces. Experimental results demonstrate that, compared with two state-of-art methods, the proposed tracking algorithm is more effective, especially in dynamically changing and clutter scenes.
- ② We propose two tracking algorithms, which combines of deterministic and stochastic tracking frameworks: (1) kernel Bayesian tracking framework; (2) singular value decomposition (SVD) based Kalman particle filter framework. (1) First, the kernel method is applied to the current frame to obtain the offset of the object state, and then these information are incorporated in Bayesian filtering framework as a heuristic priori. This algorithm can effectively combine the advantages of deterministic and stochastic tracking frameworks. And this algorithm is general to any form of appearance models. Experimental results demonstrate that, compared with Bayesian and kernel based tracking frameworks, the proposed algorithm is more efficient and effective. (2) Firstly, a set of sigma samples are generated by SVD of the eigen-covariance matrix with corresponding weights, and then are filtered by a kalman filter, finally the weighted mean and covariance are further calculated to form a better proposal distribution for particle filter. Compared with the particle filter and the unscented particle filter, the proposed algorithm is more robust in tracking performance.
- ③ We propose a sequential particle swarm optimization based tracking framework. Unlike independent particles in the particle filter, the particles in PSO

interact locally with one another and with their environment, and eventually cluster in the regions where the optima are located. Meanwhile, we show theoretically that, in a Bayesian inference view, the PSO iterations are essentially a swarm-intelligence guided multi-layer importance sampling strategy which incorporates the new observations into a sampling stage, and thus avoids the sample impoverishment problem suffered by the particle filter. We also extend the particle swarm to the species concept, and propose a species competition and repulsion model to handle the occlusion problems in multiple object tracking. Compared with two other algorithms, our algorithm achieves the best results on the PETS2004 dataset.

- ④ We contribute to human pose estimation in three aspects. First, we parse the tree structure state space into a lexicographic order according to the image observations and body topology, and the optimization process is conducted in this order. Second, we design two efficient Markov Chain dynamics to effectively explore the complex solution space. Third, an effective data-driven Markov Chain Monte Carlo (DDMCMC) framework is used to solve the optimization problem. Our algorithm realizes a much more efficient exploration than the sampling based search and exhaustive search, and thus achieves a tremendous speed-up.

Key Words: Graph embedding learning, particle swarm optimization, kernel Bayesian, Kalman filter, particle filter, object tracking, pose estimation

目 录

第一章 绪 论	1
1.1 研究背景、目的和意义	1
1.2 研究思路和研究内容	2
1.2.1 目标跟踪	2
1.2.1.1 表观模型	2
1.2.1.2 跟踪框架	3
1.2.2 姿态估计	3
1.3 论文的结构安排	3
第二章 研究现状概述	7
2.1 引言	7
2.2 运动目标跟踪	7
2.2.1 表观模型	8
2.2.2 跟踪框架	9
2.3 姿态估计	11
第三章 基于图嵌入判别学习的目标跟踪	15
3.1 基于图嵌入的学习框架	15
3.1.1 图嵌入	15
3.1.2 图嵌入的降维学习	16
3.2 基于图嵌入学习的目标跟踪	18
3.2.1 问题描述	18
3.2.2 基于样本分布的图嵌入	19
3.2.2.1 图结构	19
3.2.2.2 子空间学习	20
3.2.3 跟踪算法	22
3.2.3.1 算法概述	22
3.2.3.2 基于 SSD 的运动预测	23
3.2.3.3 基于图嵌入的表观模型	24
3.2.3.4 负样本的启发式选择	24
3.2.3.5 在线更新	25
3.2.3.6 算法总结	25
3.3 实验结果	26

第四章 确定性搜索和随机性搜索相结合的跟踪框架	31
4.1 核贝叶斯框架	32
4.1.1 核方法	32
4.1.2 贝叶斯方法	33
4.1.3 核贝叶斯方法	34
4.1.4 讨论和分析	35
4.2 核贝叶斯跟踪算法	36
4.2.1 基于空间约束混合高斯的表观模型	36
4.2.1.1 表观模型	37
4.2.1.2 参数估计	38
4.2.2 核贝叶斯跟踪框架	39
4.2.3 表观模型的选择性更新	39
4.2.4 实验结果	40
4.2.4.1 单目标跟踪	41
4.2.4.2 多目标跟踪	42
4.3 基于奇异值分解的卡尔曼粒子滤波算法	43
4.3.1 基于 Unscented 卡尔曼粒子滤波算法	44
4.3.1.1 Unscented 卡尔曼滤波器	44
4.3.1.2 基于 Unscented 卡尔曼粒子滤波算法	45
4.3.2 基于奇异值分解的卡尔曼粒子滤波算法	46
4.3.3 基于增量子空间学习的表观模型	47
4.3.4 实验结果	48
第五章 基于序列粒子群优化的跟踪方法	51
5.1 基于序列粒子群优化算法的单目标跟踪	51
5.1.1 粒子群优化算法	51
5.1.1.1 来源和背景	51
5.1.1.2 粒子群优化算法	53
5.1.1.3 算法分析	55
5.1.2 序列粒子群优化算法	55
5.1.2.1 动机	55
5.1.2.2 序列 PSO 框架	56
5.1.3 算法分析	59
5.1.3.1 粒子滤波	59
5.1.3.2 多层采样过程	61
5.1.4 跟踪算法	62
5.1.5 实验结果	62
5.1.5.1 序列 PSO vs 粒子滤波和基于 Unscented 卡尔曼的粒子滤波	63

5.1.5.2	其他场景的测试结果	65
5.2	基于多种群粒子群优化算法的多目标跟踪	67
5.2.1	动机	67
5.2.1.1	从生物群体智能角度来看单目标跟踪	67
5.2.1.2	如何扩展到多目标跟踪	67
5.2.2	多目标跟踪算法	68
5.2.2.1	问题描述	68
5.2.2.2	竞争和排斥模型	69
5.2.2.3	基于高斯退火的粒子群优化	71
5.2.2.4	表观模型的选择性更新	72
5.2.2.5	算法总结	73
5.2.3	算法理论分析	74
5.2.3.1	最优重要性建议分布	74
5.2.3.2	采样策略分析	76
5.2.4	实验结果	78
5.2.4.1	PSO vs AGPSO	78
5.2.4.2	多目标跟踪	80
5.2.4.3	讨论与总结	84
第六章	基于树模型解析的姿态估计算法	85
6.1	算法概述	85
6.1.1	三层人体树模型	85
6.1.2	问题描述	86
6.1.3	优化框架	86
6.2	后验分布	87
6.2.1	图像表观似然	87
6.2.1.1	前景区域似然	87
6.2.1.2	边缘图似然	89
6.2.2	人体拓扑约束	89
6.3	姿态估计的优化框架	91
6.3.1	初始化	91
6.3.2	局部优化	92
6.3.2.1	基于语法的树模型解析	92
6.3.2.2	优化模式选择	94
6.3.2.3	马尔可夫链蒙特卡罗优化过程	95
6.3.2.4	层次重要性传播	96
6.4	实验结果	96
6.4.1	自拍图片上的姿态估计性能	96
6.4.2	网络图片上的测试结果	97

6.4.3	定量分析	98
6.4.4	语法分析	100
6.4.5	噪声输入下的性能	101
6.4.6	参数对算法性能的影响	103
第七章	结束语	105
7.1	工作总结	105
7.2	未来展望	107
	参考文献	109
	攻读博士学位期间发表的论文	i
	致谢	v

插图目录

3-1	(a) PCA: 最小化重构误差的解空间, (b) LDA 投影方向 (虚线) 和我们算法的投影方向 (实线)	18
3-2	图嵌入结构示意图	19
3-3	算法流程图	23
3-4	产生样本粒子的过程。左边: 前一帧粒子状态; 中间: 未做运动预测产生的粒子; 右边: 根据运动预测产生的粒子	24
3-5	我们算法的跟踪结果 (白色: 未做运动估计, 绿色: 运动估计)	27
3-6	算法所需的粒子数 (白色: 未做运动估计, 绿色: 运动估计)	27
3-7	跟踪结果 (白色: 增量子空间, 绿色: 我们的方法)	28
3-8	跟踪的均方误差曲线 (红色: 增量子空间, 绿色: 我们的方法)	28
3-9	局部遮挡 (白色: 基于 LDA 的跟踪算法, 绿色: 我们的算法)	29
3-10	跟踪的均方误差曲线 (红色: 基于 LDA 的跟踪算法, 绿色: 我们的算法)	29
3-11	判别能力 (红色: 基于 LDA 的跟踪算法, 绿色: 我们的算法)	30
4-1	核函数示意图 (左边: Epanechnikov 核, 右边: 常数核)	33
4-2	输入图像的三个窗口 (左边) 分别对应状态空间中的初始状态, 观测似然分布 (右边) 的局部极大值和全局极大值	36
4-3	基于核贝叶斯的跟踪算法流程	36
4-4	基于二维高斯分布的空间约束	37
4-5	跟踪结果	41
4-6	左边: 算法运行时间, 右边: 跟踪结果的均方误差 (红色: 核方法, 绿色: 贝叶斯方法, 蓝色: 核贝叶斯方法)	42
4-7	复杂环境下的跟踪结果	43
4-8	不同场景下的实验结果	43
4-9	多目标跟踪结果	43
4-10	算法流程图	46
4-11	跟踪结果 (第一行: 传统的粒子滤波算法, 第二行: 基于 Unscented 卡尔曼粒子滤波算法, 第三行: 我们的算法)	48
5-1	粒子 i 第 n 次迭代的示意图	53
5-2	序列 PSO 算法流程	56
5-3	收敛准则	59
5-4	重要性采样的例子 (左边: 直接从 $p(x_t x_{t-1})$ 采样; 右边: 经过 PSO 迭代后)	61
5-5	跟踪结果	64
5-6	第 19 帧图像的跟踪过程	66
5-7	更多的实验结果	66

5-8	两个目标遮挡的例子	69
5-9	将重叠区域的图像观测投影到目标对应部分的子空间上	69
5-10	状态估计的一次运行结果	79
5-11	第一行：跟踪结果（蓝色：PSO，红色：AGPSO）；第二行： 均方根误差（左边），算法运行时间（右边）	80
5-12	例子 1 的跟踪结果，包含帧 204, 209, 214, 219, 224（第一行：在 遮挡过程中不对表观模型进行更新；第二行：我们的选择性更新 策略）	81
5-13	例子 2 的跟踪结果，包含帧 422,455,465,480,488,501,518	82
5-14	例子 2 中所恢复的遮挡关系。	83
5-15	购物中心场景中多目标跟踪，包含帧 218,248,272,298,359,406。	84
6-1	（a）人体结构图（数字为人体部分标号）；（b）三层人体树模 型；（c）人体部分状态示意图	86
6-2	算法流程图	87
6-3	（a）原图；（b）前景图；（c）产生的人体姿态（彩色矩形框）	88
6-4	（a）原图；（b）边缘图；（c）边缘距离映射图；（d）合成的 人体姿态（彩色矩形框）	88
6-5	黑色粗线上的像素用于计算边缘似然	89
6-6	第 i 个人体部分的约束示意图	90
6-7	(a) 人体前景区域；(b) 前景距离映射；（c）躯干方向和前景主 轴方向	91
6-8	鉴别坏躯干的例子：（a）肩部以上前景区域过大；（b）头与前 景匹配不好	92
6-9	局部优化流程图	92
6-10	（a）重要性建议分布计算示意图；（b）语法一不适合的一个例子	93
6-11	跳跃模式将激活选定人体部分作为根节点的子树上的优化过程	95
6-12	自拍图像的测试结果（第一行：室内环境，第二行：室外环境）	97
6-13	网络图片的测试结果（第一行：瑜伽，第二行：太极拳）	98
6-14	均方根误差收敛曲线	99
6-15	迭代次数为 50 次时的均方根误差的直方图	99
6-16	不同语法对算法性能的影响	100
6-17	边缘噪声下的姿态估计结果	102
6-18	前景噪声下的姿态估计结果	102

表格目录

4-1	表观模型的选择性更新	40
4-2	定量试验结果	49
5-1	定量比较实验	65
5-2	状态估计的实验结果	79
5-3	定量比较结果	83
6-1	算法平均运行时间	98
6-2	不同参数下的算法性能	103
6-3	不同 k, w 下的算法性能	103

第一章 绪 论

1.1 研究背景、目的和意义

运动目标的视觉分析是计算机视觉领域中备受关注的前沿方向，是计算机视觉领域中最活跃的研究主题之一，它是指利用计算机视觉技术从图像或视频序列中检测、跟踪运动目标并对其行为进行理解与描述 [1,2,3,4]。它涉及到模式识别、图像处理、计算机视觉、人工智能、统计分析等多学科知识，是一个具有挑战性的跨学科的研究方向。

在运动目标的视觉分析中，目标的跟踪处理占有重要的地位，属于视觉中的中级处理部分，它的目的是通过对摄像机所拍摄的视频图像序列进行处理，检测、定位和跟踪其中运动的物体。一方面它依赖于底层的运动检测，特征提取等处理，另一方面为高层的行为理解和识别奠定了基础。通过对目标的跟踪，可以方便地获得目标的运动参数和轨迹，从而使后续的行为理解直接建立在对参数的分析上，而不需要对大量的原始图像序列进行分析。运动目标视觉分析中的另外一个重要的组成部分是目标的姿态估计，所谓姿态估计问题就是：根据图像信息推断出目标物体（主要针对人体，手势）的三维方位指向，它在目标运动的视觉分析中属于高层部分，可以直接作为语义输出，例如人机交互系统，也可以为更高层的视觉任务（如行为识别）提供辅助信息。

目标跟踪和姿态估计除了具有重要的理论研究价值以外，还具有巨大的应用前景和潜在的经济价值。

- 智能监控

在高速发展的现代社会，安全问题变得日益重要。视觉监控系统普遍的应用于很多大型场所，比如交通领域、银行、机场、超市等。但是传统的视觉监控系统都是一种事后监控，对各种破坏安全性的行为起不到制止的作用，不能满足社会发展的需要。因此，智能化的视频监控系统成为时代的要求和研究的热点。其最终目的是实现对运动目标行为的实时识别，从而对异常行为作出及时的判断和预警。而运动目标的行为识别与分析等高层问题的解决要依赖于底层处理，即运动目标的跟踪，另外，在进行异常情况判断时，需要对目标三维方位进行判断，即目标的姿态估计。

- 运动分析

目标跟踪和姿态估计另一个重要的应用就是的运动分析（以人为例子）。分割图像中的人体部分并在图像序列中跟踪感兴趣的关节运动，对于建立人体的几何模型、解释人体的运动行为机制，从而提高运动性能方面有着积极的推动作用，这可以应用于体育运动、舞蹈等训练中。此外，在大量的体育视频中进行基于内容的快速搜索也是以目标的跟踪和姿态估计为基础的。

- 人机交互

在高级用户接口领域中，视觉信息（如手势、头部姿态、表情等）作为语音和自然语言理解的有效补充可以完成更加智能的人机交互。而这些都是以跟踪和姿态估计为基础的，鲁棒性的跟踪和姿态估计结果有利于后续的视觉信息的识别与分析。

- 虚拟现实

目前许多电脑游戏中人的形体、运动和行为交互的设计逼真性实际得益于物理空间中人的运动分析。此外人的运动分析在视频会议、人物动画、虚拟工作室等其他虚拟现实场合也有着相当广泛的应用。而人的跟踪和姿态估计作为人的运动分析的基础，已经成为众多研究者关注的焦点。

1.2 研究思路和研究内容

本文的研究内容主要包含两个方面：（1）目标跟踪；（2）姿态估计。下面我们将从这两个方面出发来简述一下本文的研究思路和研究内容。

1.2.1 目标跟踪

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。众所周知，目标跟踪的算法大体上要涉及两个关键性问题：（1）表观模型；（2）跟踪框架。

1.2.1.1 表观模型

表观模型就是如何对目标物体进行有效的表达，并且进行实时的更新。因此，如何构建一个好的表观模型对目标跟踪起着至关重要的作用。特别是在时序数据流中，目标的表观是随着时间而变化的，同时还可能受到各种因素的干

扰。此外，还要满足实际应用的低计算复杂度的要求。这样就给我们提出一个严峻的挑战，那就是如何构建一个低复杂度的、鲁棒的、并且自适应的表观模型，这对目标跟踪、姿态估计以及行为理解是非常重要的。正是基于此，我们研究的主要内容是如何构建一个具有判别性的目标表观模型，使得该模型对光照变化、噪音、部分遮挡、形变等复杂因素都比较鲁棒。

1.2.1.2 跟踪框架

跟踪框架就是如何对候选区域与目标模板进行有效地匹配。总的来说，目前存在的跟踪算法的理论框架大致可以分为确定性跟踪框架和随机性跟踪框架两大类。确定性跟踪框架容易陷入局部极值，而随机性跟踪框架也存在着样本退化和计算量大等问题。针对现有跟踪框架各自的缺点，我们研究的主要内容是如何建立一个有效的跟踪框架，避免陷入局部极值以及样本退化等主要问题。

1.2.2 姿态估计

现有的姿态估计算法（主要针对人的姿态）通常是将目标物体各部分的状态连接起来，并通过求解联合空间的优化问题来获得目标的姿态。该类方法面临下列困难：（1）联合状态向量无法描述人体各部分之间的拓扑约束，（2）联合状态向量维数过高（通常维数为 50+），使得优化问题无法收敛。因此，我们的研究内容从下面两个方面着手去解决上述问题：（1）利用观测信息和人体拓扑约束来对高维空间进行解析，从而获得一系列有序的低维空间；（2）采用一种基于数据驱动搜索方法来缩小搜索空间。

1.3 论文的结构安排

论文各个章节的内容安排如下：

第二章回顾了运动目标跟踪和姿态估计的研究现状。具体的说，它重点介绍了三个方面的工作：（1）常用的表观模型；（2）现有的跟踪框架；（3）常用的姿态估计方法。

第三章提出了一个基于图嵌入判别学习的表观模型，并将其应用到移动摄像机下的目标跟踪。该算法将目标跟踪视为一个前景与背景的分类问题。首先通过初始化获得正样本以及一系列负样本，并且构建起正样本、负样本以及正

负样本关系的图结构，然后通过图嵌入理论学习出正样本的模式空间，以及正负样本之间的判别空间。基于这两个模式空间的学习，我们构建起鲁棒的表观模型，并通过该表观模型获得样本的置信概率，样本的最大后验概率作为跟踪结果。最后通过基于子空间的启发式策略选择正样本和负样本，并增量地更新嵌入图的结构。另外针对粒子滤波器缺乏对目标运动有效的估计，我们提出了级联的粒子滤波框架，即在粒子滤波器的前端级联一个多尺度的模版匹配模型来提供对目标运动的有效估计。实验结果表明，与基于增量子空间方法和基于 fisher 判别的方法相比，该表观模型有效地提高了跟踪的效率和精确度，并且可以有效地处理局部遮挡以及高噪声情况下的跟踪。

第四章讨论了两种结合确定性跟踪框架和随机性跟踪框架的方法：（1）核贝叶斯跟踪框架；（2）基于奇异值分解的卡尔曼粒子滤波框架。（1）首先将核方法应用于当前帧，获得对运动方向以及其偏移量的估计，再将这些信息作为启发式先验信息加入到贝叶斯滤波框架中，从而有效结合确定性方法与随机性方法的优点，并且该框架是一种通用的跟踪框架，任何形式的表观模型都可以在此框架下应用。本文采用了基于空间约束的混合高斯模型来描述目标物体，该模型将目标的空间分布信息加入到混合高斯模型中，从而有效的解决了复杂背景以及局部遮挡所引发的困难。与基于核方法的跟踪器（确定性跟踪方法）以及基于贝叶斯方法的跟踪器（随机性跟踪方法）相比，该跟踪框架有效地提高了跟踪的效率和精确度。（2）首先对前一帧跟踪样本的协方差矩阵进行奇异值分解，再利用 Unscented 变化（一种确定性的变化）获得预测样本，然后用卡尔曼滤波器对预测样本进行滤波，为最终的粒子滤波提供重要性采样分布。与粒子滤波和基于 Unscented 卡尔曼粒子滤波相比，我们的方法更为鲁棒。

第五章在第四章的基础之上，更加深入的讨论了粒子滤波算法的样本退化问题以及其理论原因。针对这一问题，本文将跟踪问题用鸟群捕食过程来模拟，提出了一种基于序列粒子群优化的跟踪算法框架。该算法首先通过初始化一系列随机粒子样本，然后对这些粒子集进行时序上的传播以及优化迭代，最后对进化的粒子进行收敛性判断。我们从理论上证明了，序列粒子群优化框架实质上是一个基于分步重要性采样的粒子滤波过程，该分步采样过程包含两个步骤：第一步是粗采样，以状态转移分步作为重要性分布，直接对其进行采样；第二步是细采样，通过粒子群优化迭代发生进化，该过程实际上是一个隐采样过程。通过这样的两步重要性采样策略，我们的采样结果逼近了直接采

样于最佳重要性分布的结果。我们还将序列粒子群优化的概念扩展到多种群上，将多目标跟踪中目标物体的交互和遮挡用种群间的竞争和排斥来模拟，实现了有效的多目标跟踪。与其他两种多目标跟踪的算法相比，我们的算法在 PETS2004 数据库上获得了最佳的实验结果。

第六章提出了一种基于树模型解析的姿态估计算法，该算法的主要贡献如下：（1）利用图像语法对人体结构的树模型进行解析，从而减小了搜索空间；（2）优化过程采用两种模式：扩散和跳跃；（3）利用数据驱动的马尔可夫链蒙特卡洛方法来进行优化。该方法可以有效地将人体各部分的拓扑关系加入到优化过程，并通过表观数据的驱动来压缩搜索空间，从而大大提高了收敛速度（10 倍左右）和精度。实验结果表明，基于树模型解析的人体姿态估计算法在室内和室外场景下，人体部分遮挡以及各种姿态下都非常鲁棒。

第七章对全文进行了总结，并展望了进一步的研究工作。

第二章 研究现状概述

2.1 引言

由于目标运动的视觉分析有着广泛应用前景，国内外很多研究机构已经开展了大量相关项目的研究。例如，1997 年美国国防高级研究项目署 (Defense Advanced Research Projects Agency) 设立了视觉监控重大项目 VSAM (Visual Surveillance and Monitoring)，主要研究用于战场及普通民用场景监控的自动视频理解技术；英国的雷丁大学 (University of Reading) 开展了对车辆和行人的跟踪及其交互作用识别的相关研究 [5]；英国利兹大学 [6] 也同时开展了基于模型的行人监控系统的研究。在德国的法兰克福理工学院，Nagel 等人 [7] 从事了交通监控系统的研究。IEEE 从 1998 年起资助了国际视觉监控系统研讨会。在硬件设备方面，国外许多公司设计了各种专用设备，例如 Action cameral [8]，Smart cameral [9]，Omni-directional cameral [10] 等等。国内的中科院自动化所模式识别国家重点实验室 (National Laboratory of Pattern Recognition) 开发的 Vstar 演示系统，对交通场景中的车辆进行实时的跟踪与异常行为检测。当前，国际上一些权威期刊如 IEEE TPAMI、IJCV、CVIU、IEEE TCSVT、IVC 和重要的学术会议如 ICCV、CVPR、ECCV、ACCV、ICIP、ICPR、IWVS 等将人的运动分析研究作为主题内容之一，为该领域的研究人员提供了更多的交流机会。复杂场景下的目标运动分析受到了世界上许多重要研究机构的高度重视，这从另一方面说明了它的重要性。以下我们将从目标跟踪和姿态估计两个角度来回顾一下目前的研究现状。

2.2 运动目标跟踪

跟踪问题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。大体上，目标跟踪的算法都要涉及两个关键性问题：

- 表观模型——如何对目标进行建模、并且实时更新；
- 跟踪框架——如何对候选区域与目标模板进行有效地匹配。

以下我们将从这两个方面分析国内外的研究现状。

2.2.1 表观模型

目标区域的颜色直方图 [11,12] 通常被用来刻画目标的表观特性。由于颜色直方图提取简便并且对目标尺度、旋转以及非刚性形变较为鲁棒,从而被研究人员广泛地使用。但是,颜色直方图存在一个潜在的问题,那就是它完全忽略了目标表观颜色的空间分布信息。这样导致的结果就是当两个目标的颜色组成成分非常相似但颜色空间分布相差比较大的时候,颜色直方图无法区分这两个目标。为了克服颜色直方图丢失空间信息的缺陷,基于核密度估计的表观模型应运而生 [13,14],该模型的优点是通过运用一些空间加权的核来刻画目标的颜色空间分布信息。不幸的是,这种类型的模型有一个共同的毛病,那就是计算和存储复杂度非常高,以至于不能被广泛使用。近年来,混合高斯表观模型 [15,16,17,18,19,20,21] 非常流行并且被广泛地应用于计算机视觉和模式识别领域。基于混合高斯的表观模型利用混合加权的高斯来动态地学习像素颜色的时间统计特性。然而,这种模型只是独立地处理每个像素,而没有考虑像素间的内在的联系,从而使模型对噪音非常敏感。文献 [22] 提出基于混合高斯的融合空间和颜色的自适应表观模型,该模型能够获取丰富的局部空间和颜色分布信息。但是,该模型也存在一个较大的缺点,那就是像素的全局时空统计特性不能够被该模型有效地刻画,从而导致对目标表观的全局性的变化以及噪音非常敏感。另外,基于条件随机场的表观模型 [23] 通过马尔可夫随机场来建模邻近像素之间的内在关系,但是这样一来,模型的训练代价非常巨大,同时也不能考虑像素的全局分布信息。近年来,基于在线子空间学习的方法 [24] 被用来描述像素的全局统计信息。但该类方法首先要将图像展成向量,目标表观的空间信息几乎完全丢失,这样使得模型对目标表观的全局性变化以及噪音非常敏感。

最近,很多人使用基于张量的子空间方法来建模目标表观的时空统计特性,这样可使目标的时空分布冗余能够极大地被减少,从而能构建一个紧密的目标表观模型。但是目前的基于张量的目标表达方法大多数是离线操作的,他们不能在线地更新,从而导致了很大的运算和存储代价。这里我们简要地回顾一下最近相关文献中基于张量的目标建模方法。在文献 [25] 中,2DPCA (二维主成分分析) 被提出来表征一幅图像。在文献 [26] 中,2DLDA (二维线性判别

分析方法) 被提出, 该方法直接在对图像矩阵操作的基础上进行分类, 而不需要把图像展成向量。文献 [27] 提出一种 GLRAM 算法, 该算法可以构建一个低阶的模型来逼近一个序列矩阵。文献 [28] 提出一种 GPCA (广义主成份分析) 的降维算法, 该算法直接对图像集构建一个低阶的矩阵表示。文献 [29] 提出一种 R 阶张量分解方法, 该方法能够较好地减少张量的时空冗余。文献 [30] 开发了一种张量的判别分析方法, 该方法将传统的判别分析方法扩展到了张量空间。在文献 [31,32] 中, 图像集不同模态的奇异值分解被用来构建人脸图像的紧致表达。文献 [33] 提出了一种叫张量子空间分析的学习算法, 该算法通过学习一个低阶的张量子空间来刻画张量空间的内在局部几何结构。文献 [34] 给出了针对一般基于张量子空间学习的一个收敛的解。在文献 [35] 中, 动态和流张量分析被提出来挖掘高阶的数据流。同样在文献 [36] 中, 基于窗的张量分析方法被用来建模实时的数据流。综观以上基于张量的方法, 它们存在一个共同的问题, 那就是它们不能在线地自适应地更新基于张量的特征空间 (均值和特征基), 因此, 无法将这种有效的建模方法应用于时序数据, 例如, 视频流数据, 从而无法将该方法有效地应用到视觉跟踪领域。

上述的目标跟踪方法所使用的表观模型通常只考虑到正样本 (目标物体) 模式, 例如通过模板、直方图、混高斯模型或者张量表示等方法, 很少考虑正样本周围的负样本 (背景) 的模式以及正负样本之间的局部模式等, 因此产生的表观模型往往不是最具判别性的, 进而导致在实际应用中对光照、遮挡以及形变等复杂因素不够鲁棒。**如何设计一种表观模型, 能够有效地抓住目标样本的分布信息以及目标样本与其近邻区域中背景样本的相对分布信息, 是设计鲁棒跟踪器的关键所在。**

2.2.2 跟踪框架

对于跟踪的框架而言, 从优化的角度来说, 目前存在的跟踪算法的理论框架大致可以分为两类: 一种称之为确定性 (deterministic) 跟踪框架, 另一种称之为随机性 (stochastic) 跟踪框架。下面我们将进一步介绍这两种跟踪框架。

- **确定性跟踪框架:** 典型的确定性跟踪框架根据事先设定的相似度损失函数 (Similarity Cost Function), 在当前图像帧内的局部区域中通过梯度优化的迭代方法找到与目标模板最为相似的区域。较为广泛使用的损失函数是误差平方和函数 (Sum of Squared Difference, SSD) [37] 和基于核直方

图 [38, 39, 40] 的损失函数。其中, SSD 损失函数是基于模板与当前图像的像素灰度差的平方和, 并通过迭代优化获得目标的运动参数; 基于核直方图的损失函数是衡量两个核直方图之间的相似度, 并采用均值漂移 (Mean Shift) 来实现对帧间目标偏移量的估计。目前, 更为鲁棒的损失函数不断被提出, 比如基于混合高斯的 [41]、基于子空间学习 [42] 的损失函数。同时, 诸如模拟退火 [43], 遗传算法 [44] 等其它优化方法也被用来搜索与目标模板最为相似的局部最优区域。总的来说, 确定性跟踪框架一般计算量较小, 算法收敛快, 效率较高, 但容易陷入局部极小值, 导致跟踪的结果不够精确。随着跟踪误差的积累, 最后可能导致跟踪失败。

- 随机性跟踪框架: 随机性跟踪框架将跟踪问题看作概率推理 (Probability inference) 问题, 目标是通过已知物体的先验知识, 并结合观测结果来估计物体的状态, 即转化为最大化后验概率问题。技术上的困难在于, 如何得到关于目标状态的后验概率的准确描述。

跟踪过程由预测和校正两步组成:

- ① 预测: 根据前 $i-1$ 帧的观测值 (用 $y_0 \dots y_{i-1}$ 表示) 预测第 i 帧目标的状态 (用 x_i 表示), 即有如下表达式:

$$\begin{aligned}
 & p(x_i | y_0 \dots y_{i-1}) \\
 &= \int p(x_i, x_{i-1} | y_0 \dots y_{i-1}) dx_{i-1} \\
 &= \int p(x_i | x_{i-1}, y_0 \dots y_{i-1}) p(x_{i-1} | y_0 \dots y_{i-1}) dx_{i-1} \\
 &= \int p(x_i | x_{i-1}) p(x_{i-1} | y_0 \dots y_{i-1}) dx_{i-1} \\
 &\propto p(x_i | x_{i-1}) p(x_{i-1} | y_0 \dots y_{i-1}) \tag{2-1}
 \end{aligned}$$

- ② 校正: 结合第 i 帧的观测值修正状态的预测值, 因此 i 时刻的后验概率 $p(x_i | y_0 \dots y_i)$ 由下式表达:

$$\begin{aligned}
 & p(x_i | y_0 \dots y_i) \\
 &= \frac{p(y_i | x_i, y_0 \dots y_{i-1}) p(x_i | y_0 \dots y_{i-1}) p(y_0 \dots y_{i-1})}{p(y_0 \dots y_i)} \\
 &= \frac{p(y_i | x_i, y_0 \dots y_{i-1}) p(x_i | y_0 \dots y_{i-1})}{\int p(y_i, x_i | y_0 \dots y_{i-1}) dx_i} \\
 &= \frac{p(y_i | x_i, y_0 \dots y_{i-1}) p(x_i | y_0 \dots y_{i-1})}{\int p(y_i | x_i, y_0 \dots y_{i-1}) p(x_i | y_0 \dots y_{i-1}) dx_i}
 \end{aligned}$$

$$\propto p(y_i|x_i)p(x_i|y_0 \dots y_{i-1}) \quad (2-2)$$

上述推导过程中的假设条件请参考 [45]。其中有两个重要的先验模型：一个是预测过程中的 $p(x_i|x_{i-1})$ ，它是用来描述系统动态特性的运动模型；另一个是校正过程中的观测模型 $p(y_i|x_i)$ ，用来描述观测值与模型的相似程度。根据运动模型和观测模型的不同，适用的跟踪滤波器也不同。当两者都服从高斯分布，为线性关系时，可以利用卡尔曼滤波 (Kalman filter) [46] 来获得后验概率的解析表达。当两者为非线性关系时，可以利用扩展卡尔曼滤波 (Extended kalman filter, EKF) [46] 求解后验概率，此时后验概率仍然满足高斯分布。一种替代 EKF 的方法是 UKF(Unscented Kalman Filter, UKF) [47] [48]。如果状态空间是离散的而且是由有限个状态组成的，则可以使用隐马尔科夫模型 [49] 进行跟踪。但是，在实际的视觉跟踪过程中，后验概率的分布往往是非线性、非高斯、多模态的，此时一种称之为序贯蒙特卡罗的方法 (Sequential monte carlo methods, SMC) [50] 被提出用于解决这类情况，利用重要性采样来逼近后验分布。该方法在刚出之时并未引起视觉领域的很大关注，直到 1996 年，由 Isard 和 Blake 提出 Condensation [51] 算法之后，才引起广泛注意。目前这一类方法又被称为粒子滤波 (Particle filter) [52]，是当前视觉跟踪方法中的研究热点。另外，UKF 思想也可被粒子滤波方法借用，这种情况下粒子滤波被称为 UPF(Unscented particle filter) [53]。

与确定性跟踪框架相比，随机性跟踪框架更为鲁棒，但是计算量过大，而且随着目标个数的增长呈指数增长。另外，由于没有好的机制选择重要性分布，粒子滤波器存在着严重的样本退化问题。**确定性跟踪框架和随机性跟踪框架有着各自的优点和缺点，若将确定性跟踪框架和随机性跟踪框架这两种思路结合起来使用，就有可能克服各自弱点，实现更为有效的跟踪。**

2.3 姿态估计

姿态估计问题就是：确定目标物体（本文主要针对人体）的三维方位指向问题。姿态估计在智能视觉监控、人机交互和视频运动分析等很多领域都有重要应用。由于姿态估计在计算机视觉领域的巨大学术价值和应用价值，近年来

有很多工作 [54,55,56,57,58] 尝试去解决姿态估计以及其相关领域的一些问题。在文献 [59] 中，作者认为大多数现有的姿态估计算法粗略的可以分为以下四类。

- 基于例子的方法：基于例子的姿态匹配算法将一系列典型姿态的特征做为例子保存起来，然后对新来的测试样本提取同类型特征，并用 k 近邻分类器对其进行分类，将最相似的例子做为测试样本的姿态。在 [60] 中，预先定义的姿态的两维轮廓图被作为例子，分类过程采用形状上下文特征 [61] 以及链式运动模型进行匹配。文献 [62] 提出了一种新的算法来学习一系列 hashing 函数，在姿态估计过程中利用这些函数来对姿态例子建立索引。在文献 [63] 中，作者提出一种基于运动例子的姿态估计算法。在他们的工作中，一系列基于光流特征的运动例子被提取，并将 Gibbs 采样和梯度法用于姿态匹配。文献 [64] 将人体姿态用梯度方向直方图来描述，在此基础上开展了对基于例子的姿态估计算法进行评价的工作。在文献 [65] 中，作者采用二维和三维 haarlets 特征来描述人的姿态，而姿态识别是通过学习 haarlets 特征的最大判别性函数来实现的。文献 [66] 提出一种基于例子的三维人体模型跟踪算法，其中每一个例子都对应于一个视角下的人体视觉信息和三维骨架。上述基于例子的姿态估计算法的主要缺点是：由于人体姿态分布于高维空间中，基于例子的算法如果要进行精确匹配的话需要的例子数量将会很大，这就带来了巨大的计算量。为了克服这个问题，文献 [67] 将例子根据其相似关系设计成树形结构，当测试姿态与例子进行匹配时，采用一种由粗到细的搜索模式，这一策略大大的降低了例子匹配过程中的复杂度。
- 自上而下的方法：这一类方法直接对整个人体模型产生候选姿态，并与图像进行匹配。由于我们可以设计包含任意参数的从姿态投到图像的函数，因此投影自上而下的方法往往可以跟图像信息进行精确的匹配，而自上而下的方法又可以叫做姿态合成——分析方法。在文献 [68] 中，姿态估计可以看成是一个迭代分解的过程，在每一次迭代中建立更好的匹配特征。[69] 通过少量的图像帧学习特定姿态的判别性模型，并用于该类姿态的检测和跟踪。在文献 [70] 中，著名的图案结构方法 (pictorial structures) 被用于人体姿态估计。文献 [71] 提出一种多层次的框架来结

合随机优化, 过滤, 以及局部优化, 并用于人体运动捕捉。在文献 [72] 中, 作者利用人体不同部分之间的关节约束来对三维模型与图像之间损失函数进行优化。总的来说, 姿态状态过高的维数是自上而下方法的瓶颈所在, 因此一些研究人员通过对状态空间进行分解来解决上述问题。在文献 [73] 中, 作者采用分层搜索的策略, 首先从躯干和头部, 再到其他低级别的肢体。它对状态进行了分解, 从而提高了搜索的效率。然而, 这种搜索方式在输入图像初始化做的不好的情况下, 往往得不到令人满意的结果。文献 [74] 提出一种方法来逐步减小人体部分的搜索空间, 提高了姿态估计的成功率。

- 自下而上的方法: 这类方法首先对人体各部分进行检测或者分割, 然后依据一定的规律或者人体约束将它们组合成人体的姿态。文献 [75] 提出一种自下而上的人体解析方法, 在每一层解析过程中, 通过预先定义的形状来找人体的各个部分。在文献 [76] 中, 首先检测人体各部分, 然后采用动态规划算法将它们拼凑起来。类似的, 文献 [77] 采用二次整数规划来对人体部分进行整合。在实际应用中, 由于人体部分检测器会产生许多误报, 这将导致不准确的估计或需要大量的时间去检查是否为误报。因此, 文献 [78] 采用图像分割技术来建立四肢和躯干检测器。在文献 [79] 中, 作者结合边缘方向, 局部运动和颜色信息, 并通过对联合特征进行聚类来对人体各部分进行检测。[80] 采用稀疏运动模型对运动特征属于一个特定的人体部分的概率进行评价。当发生遮挡时, 被遮挡部分的检测器将会失败, 而单一树模型难以解决这个问题, 从而导致不正确的估计。文献 [81] 采用多树模型结构, 他们的模型可以通过融合多个模型之间的信息来解决遮挡问题。文献 [82] 对遮挡信息进行编码, 并对此建立遮挡不敏感的图像局部观测似然函数。他们的模型可以通过置信传递算法进行求解。在文献 [83] 中, 作者将 [82] 的工作扩展到多视角情况, 并将各个视角下的运动约束统一到一个框架下。
- 自上而下和自下而上结合的方法: 自下而上的方法具有不需要初始化的优点, 因此可作为自上而下方法的初始化, 从而结合两类方法的优点。在文献 [84] 中, 一种基于底层特征的数据驱动马尔科夫链蒙特卡洛方法被用于姿态估计。文献 [85] 提出一种结合确定性和随机性搜索的策略, 其中边缘

和肤色信息被用来引导搜索过程。在文献 [86] 中，作者利用观测数据生成的建议映射（proposal maps）来产生候选姿态，然后利用自上而下的方法进行验证。[87] 通过图像观测和关节约束来检测人体部分，然后在分层框架下对姿态进行估计。[88] 将人体各部分的图像信息，空间约束，运动信息融到一个基于图的产生式模型中，然后通过求最大后验估计来获得人体各部分的位置。[89] 在多视角下融合图像分割信息，并用于姿态估计。

第三章 基于图嵌入判别学习的目标跟踪

复杂场景下目标的运动跟踪是近些年来计算机视觉领域里的前沿研究方向之一，也是该领域中的难点之一。该任务最大的挑战与难点在于如何对目标进行有效的表达，而目标表达大多数是通过目标表观进行建模来实现的。因此，如何构建一个好的表观模型对目标识别起着至关重要的作用。特别是在时序数据流中，目标的表观是随着时间而变化的，同时还可能受到各种因素的干扰。此外，还要满足实际应用的低计算复杂度的要求。这样就给我们提出一个严峻的挑战，那就是如何构建一个低复杂度的、鲁棒的、并且自适应的表观模型，这对目标的运动跟踪是非常重要的。

现有的跟踪方法大多只着眼于前景目标的匹配，而丢弃了背景信息对目标跟踪的帮助。针对这一问题，本文提出了一个基于图嵌入判别学习的表观模型，该模型将目标跟踪视为一个前景与背景的分类问题。首先通过初始化获得正样本以及一系列负样本，并且构建起正样本、负样本以及正负样本关系的图结构，然后通过图嵌入理论学习出正样本的模式空间，以及正负样本之间的判别空间。基于这两个模式空间学习，我们构建起鲁棒的观测模型，并通过该观测模型获得样本的置信概率，样本的最大后验估计即为所跟踪的状态。最后通过基于子空间的启发式策略来选择正样本和负样本，并增量地更新嵌入图的结构。在本章中，我们首先介绍图嵌入学习框架，然后提出基于图嵌入判别学习的表观模型以及它在跟踪中的应用，最后给出了实验结果。

3.1 基于图嵌入的学习框架

图嵌入是一种基于特定约束的图结构设计方法。Yan 等 [90] 提出了一个基于图嵌入的降维学习框架，将主成分分析 (principal component analysis)，线性判别分析 (linear discriminant analysis)，核主成分分析 (Kernel PCA) 等纳入到一个统一的数学形式下，并指出该框架可以依据用户需求设计不同的约束条件，从而获得不同的降维算法。

3.1.1 图嵌入

令 $x_i \in \mathbb{R}^d (i = 1, 2, \dots, n)$ 为 d 维输入样本， $y_i \in \{1, 2, \dots, C\}$ 为对应数据类别的标号， n_c 为属于第 c 类的样本个数，且满足 $\sum_{c=1}^C n_c = n$ 。基于已有的样

本，我们可以构建一个无向图 $G = \{\{x_i\}_{i=1}^n, W\}$ ，其中 $\{x_i\}_{i=1}^n$ 为图的节点， W 为相似度矩阵，它的元素 w_{ij} 表示节点 x_i 与节点 x_j 之间的相似度。在此基础上，我们定义如下的拉普拉斯矩阵：

$$L = D - W \quad (3-1)$$

其中 D 为对角矩阵，它的对角元素定义如下：

$$d_{ii} = \sum_{j \neq i} w_{ij} \quad (3-2)$$

在本文中，基于无向图 G 的图嵌入可以描述为以下的问题：给定各节点之间的权重矩阵和节点类别标号，如何寻找一种样本节点的最优的低维表示。图保留准则（graph preserving criterion）通常被用作衡量上述图嵌入问题的标准。

$$\begin{aligned} Z^* &= \arg \min_{Z^T B Z = I} \sum_{i,j} \|z_i - z_j\|^2 w_{ij} \\ &= \arg \min_{Z^T B Z = I} 2tr(Z^T L Z) \end{aligned} \quad (3-3)$$

其中 z_i 是 x_i 的低维表示， Z 是 $\{z_i\}_{i=1}^n$ 的矩阵形式， B 为约束矩阵。

3.1.2 图嵌入的降维学习

在 Yan 等 [90] 的工作中，它提出了一个基于图嵌入的降维学习框架，该学习框架将经典的降维算法，如主成分分析，线性判别分析以及核主成分分析（Kernel PCA）等，纳入到一个统一的数学形式下，并指出在该框架下用户可以依据自己的需求设计不同的约束条件，从而获得不同的降维算法。在本文我们仅以线性降维为例子来说明基于图嵌入的降维学习框架，至于非线性的降维算法如何在图嵌入框架下表示，请参考 [90]。

根据公式（3-3），在线性降维的条件下， z_i 可以表示为 $z_i = P^T x_i$ ，因此公式（3-3）中的优化函数可以改写为如下的形式，

$$P^* = \arg \min_{P^T X B X^T P = I} tr(P^T X L X^T P) \quad (3-4)$$

其中 X 是样本的矩阵形式 $X = (x_1 | x_2 | \cdots | x_n)$ 。

下面我们来看如何将主成分分析和线性判别分析用图嵌入形式来表示。首先我们来看主成分分析（PCA），PCA 的目标就是能够找到一个较好的低维子空间，这样使得在误差平方和最小的准则下能够最好地描述原始数据。也就是说，主成份分析（PCA）旨在寻找一个最优的线性变换能够将原始高维空间的

样本投影到能够较好地表达数据内在属性的低维线性子空间上。具体地说，对于 n 个 d 维的样本 x_1, x_2, \dots, x_n ，PCA 试图寻找一个变换矩阵 $P \in \mathcal{R}^{d \times k} (k < d)$ 将原始样本 x_1, x_2, \dots, x_n 分别投影到 k 个主成份上，使得投影后的样本在这 k 个主成份上的离散度的和最大。令原始样本的协方差矩阵定义如下，

$$\begin{aligned} S &= \sum_i (x_i - \mu)(x_i - \mu)^T \\ &= X(I - \frac{1}{n}ee^T)X^T \end{aligned} \quad (3-5)$$

PCA 首先通过对协方差矩阵进行特征分解，然后将获得的特征向量按其对应的特征值大小排列，取前 k 个特征向量即可获得投影矩阵 P

PCA 可以用如下的图嵌入形式来表示，

$$\begin{aligned} P^* &= \arg \min_{P^T P = I} -tr(P^T S P) \\ &= \arg \min_{P^T P = I} -tr(P^T X(I - \frac{1}{n}ee^T)X^T P) \end{aligned} \quad (3-6)$$

其中图结构和约束条件为 $\{w_{ij} = 1/n, i \neq j; B = I\}$ 。

线性判别分析（LDA 或者 FDA），是以提高样本在子空间中的可分类性为目标，寻找一组基向量，在这些基向量张成的子空间中，不同类别的训练样本能有最小的类内离散度，最大的类间离散度。定义如下的样本类内散度矩阵和类间散度矩阵。

$$S^{(w)} = \sum_{c=1}^C \sum_{i:y_i=c} (x_i - \mu_c)(x_i - \mu_c)^T = X(I - \frac{1}{n_c} \sum_{c=1}^{n_c} e^c e^{cT})X^T \quad (3-7)$$

$$S^{(b)} = \sum_{c=1}^C n_c(\mu_c - \mu)(\mu_c - \mu)^T = S - S^{(w)} \quad (3-8)$$

其中 $\sum_{i:y_i=c}$ 表示对属于 c 类的样本 x_i 求和， μ_c, μ 分别为属于 c 类的样本的均值和所有样本的均值， e^c 是一个 n 维向量满足 $e^c(i) = 1$ ，如果 $y_i = c$ ，否则 $e^c(i) = 0$ 。根据上述定义，LDA 可以用如下的图嵌入形式来表示，

$$P^* = \arg \min_P tr(\frac{P^T S^{(w)} P}{P^T S^{(b)} P}) = \arg \min_P tr(\frac{P^T S^{(w)} P}{P^T S P}) \quad (3-9)$$

LDA 的图结构为 $\{w_{ij} = \delta_{y_i, y_j} / n_{y_i}\}$ ，其中 δ_{y_i, y_j} 为冲击函数，当 $y_i = y_j$ 时， $\delta_{y_i, y_j} = 1$ ，否则 $\delta_{y_i, y_j} = 0$ 。约束条件为 $\{B = I - \frac{1}{n}ee^T\}$ 。

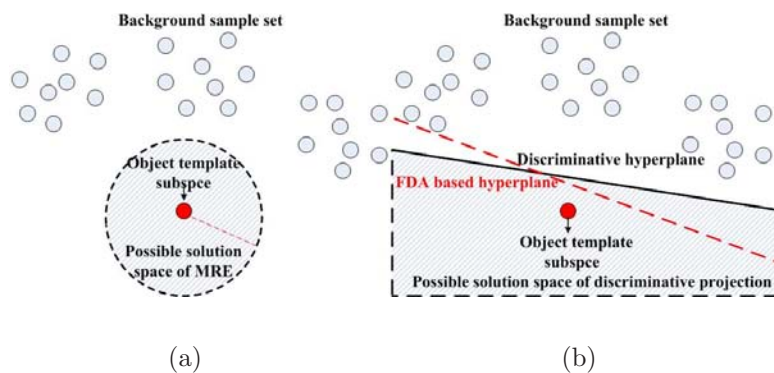


图 3-1 (a) PCA: 最小化重构误差的解空间, (b) LDA 投影方向 (虚线) 和我们算法的投影方向 (实线)

3.2 基于图嵌入学习的目标跟踪

3.2.1 问题描述

在介绍我们的工作之前,我们先来回顾一下两种基于子空间学习的跟踪算法。Lim 等 [91] 提出了基于增量子空间的目标跟踪算法,该算法采用 SKL 变化 (Sequential Karhunen-Loeve) [93] 来有效地学习目标物体和光线的变化。但是由于该算法只考虑了目标物体的样本,而没有利用背景的信息。如图 3-1(a) 所示,基于前景目标子空间的重构误差只提供了相对狭小的解空间,因此当目标表观变化超出了这个解空间的范围,跟踪就会失败。基于线性判别子空间的跟踪算法 [92] 将背景信息作为负样本引入到子空间的构建过程中,从而在一定程度上提高了跟踪性能。线性判别分析只有在样本符合高斯分布且具有相同方差的情况下才能获得最佳效果,然而在跟踪过程中背景样本的分布是凌乱和多模 (Multimodal) 的 (如图 3-1 所示),因此基于线性判别分析的子空间投影方向并不能得到理想的效果。所以跟踪中主要面临的问题:在复杂背景和噪声情况下,如何同时有效地学习目标模式的子空间以及目标与背景之间的判别空间。

3.2.2 基于样本分布的图嵌入

在文献 [94] 中, 作者指出了 LDA 在样本多模形式下无法取得理想的效果, 是因为在定义类内散度矩阵和类间散度矩阵的时候用了全局连接。由于 LDA 准则可表示为, 给类内散度一个特定的约束, 然后最大化类间散度。当样本分布是多模的情况下, 类内散度非常大, 使得对类内散度的约束放的非常宽, 因此对类间的最大化能力将相应的减小。一种好的解决方法就在定义类内散度矩阵和类间散度矩阵的时候采用局部连接, 这样可以有效抓住样本的多模信息。在下面的部分, 我们将提出一种新的图嵌入结构, 而这个图结构可以很好的反映跟踪过程中样本的分布信息。

3.2.2.1 图结构

假定我们收集了一系列正样本和负样本, 分别对应跟踪过程中的目标样本和背景样本。同样采用 §3.1.1 中的符号表示, 样本点表示为 $\{x_i\}_{i=1}^n$, 每个样本点 x_i 的类标表示为 $y_i \in \{1, 2\}$ 。在此基础上, 我们定义如下的图结构,

- 前景/目标图 $\{G^f, W^f\}$:

如图 3-2(a) 所示, 由于前景样本分布相对紧致, 我们对前景样本建立全连接图 $\{w_{ij}^f = 1/n_f\}$ (实际上就是 §3.1.2 中的 PCA 图), 其中 n_f 是前景样本的个数。

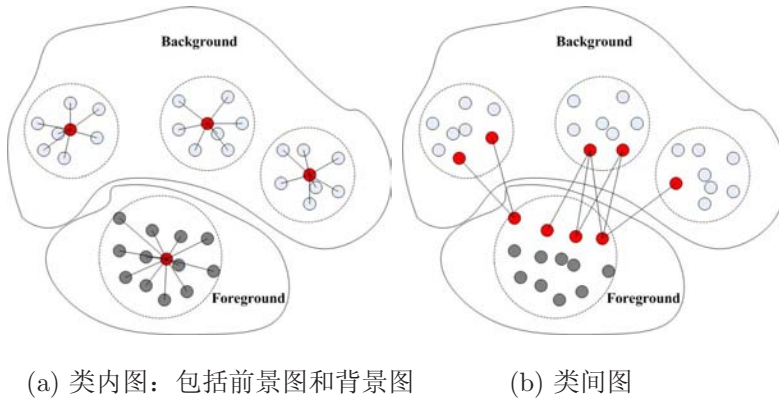


图 3-2 图嵌入结构示意图

- 背景图 $\{G^b, W^b\}$:

如图 3-2(a) 所示, 对于样本 x_i 和样本 x_j , 如果 x_j 是 x_i 的 k 近邻, 我们对这两个样本建立连接, 该连接的权重采用基于尺度的局部距离测度方法 [95], 具体形式如下,

$$w_{ij}^b = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j}\right) \quad (3-10)$$

其中 σ_i 代表样本 x_i 的 k 近邻的尺度, 通常定义如下,

$$\sigma_i = \|x_i - x_i^{(k)}\| \quad (3-11)$$

这里 $x_i^{(k)}$ 是样本 x_i 的 k 近邻, 在 [95] 中, 作者证明了 $k = 7$ 可以处理各类尺度的数据, 因此在我们的工作也采用这个设置。如果 x_j 不是 x_i 的 k 近邻, 则 $w_{ij}^b = 0$ 。

- 类间图 $\{G', W'\}$:

对于类间图 G' , 我们只考虑类别标号 $y_i \neq y_j$ 的样本对 x_i 和 x_j , 同样的, 对于这样的 x_i 和 x_j , 如果 x_j 是 x_i 的 k 近邻, 我们对这两个样本建立连接, 连接权重同样采用基于尺度的局部距离测度方法 (见公式 (3-10)), 如图 3-2(b) 所示, 最大化该类间距离与 SVM 中最大化类间间隔 (margin) 的思想比较相似, 因此在最大化类间距离方面的效果要好于类间全连接。

3.2.2.2 子空间学习

给定图结构后, 目标模式子空间和目标与背景之间的判别子空间可以通过以下步骤获得,

- 正如前景/目标图定义所述, 目标图实际上就是 PCA 图, 所以目标模式子空间 P 可以通过求样本协方差矩阵的特征结构获得。

- 目标与背景之间的判别子空间 V , 可以通过求解下列的优化问题获得,

$$\text{Maximize } J(V) = \sum_{i,j} \|V^T x_i - V^T x_j\|^2 w'_{ij}$$

$$\text{subject to } \sum_{i,j} \|V^T x_i - V^T x_j\|^2 w_{ij} = 1$$

其中 W 定义如下,

$$W = \begin{pmatrix} W^f & 0 \\ 0 & W^b \end{pmatrix}$$

上述优化问题可以转化为下列求解广义特征向量的问题,

$$X(D' - W')X^T v = \lambda X(D - W)X^T v \quad (3-12)$$

具体证明如下,

证明: 因为 $\|A\|^2 = \text{tr}(AA^T)$, 所以

$$\begin{aligned} J &= \sum_{i,j} \text{tr}\{(V^T x_i - V^T x_j)(V^T x_i - V^T x_j)^T\} w'_{ij} \\ &= \sum_{i,j} \text{tr}\{V^T (x_i - x_j)(x_i - x_j)^T V\} w'_{ij} \end{aligned}$$

矩阵的迹是线性操作符且 w'_{ij} 为标量, 所以将求和符号和 w'_{ij} 放入到迹操作里面, 对结果不会造成影响,

$$\begin{aligned} J &= \text{tr}\{V^T \sum_{i,j} ((x_i - x_j) w'_{ij} (x_i - x_j)^T) V\} \\ &= \text{tr}\{V^T (2XD'X^T - 2XW'X^T) V\} \\ &= 2\text{tr}\{V^T (X(D' - W')X^T) V\} \end{aligned}$$

所以优化问题可以重新表示为,

$$\text{Maximize } J(V) = 2\text{tr}\{V^T (X(D' - W')X^T) V\}$$

$$\text{subject to } 2\text{tr}\{V^T (X(D - W)X^T) V\} = 1$$

优化问题的拉格朗日形式如下，

$$L = 2\text{tr}\{V^T(X(D' - W')X^T)V\} + \lambda\{1 - 2\text{tr}\{V^T(X(D - W)X^T)V\}\}$$

令 $V = [v_1, \dots, v_l]$, 这样我们可以得到,

$$\frac{\partial L}{\partial v} = 4X(D' - W')X^T v - 4\lambda X(D - W)X^T v$$

所以, 该优化问题的求解过程可以转化为求解下列广义特征向量的问题,

$$X(D' - W')X^T v = \lambda X(D - W)X^T v$$

因此, 判别子空间 V 可以表示为最大的 l 个特征值对应的特征向量, $V = [v_1, v_2, \dots, v_l]$ 。

正如图 3-1(b) 所示, 在样本多模形式下, 我们的方法可以取得比 LDA 更具判别性的投影方向。事实上, 即使在样本分布为高斯的情况下, 我们的方法也可以达到与 LDA 相当的性能。

3.2.3 跟踪算法

3.2.3.1 算法概述

粒子滤波提供了一个有效的跟踪框架 [45]。因此, 我们将图嵌入判别学习所得到的表观模型加入粒子滤波框架下, 应用于视觉跟踪。主要的跟踪算法流程见图 3-3。首先, 我们采用基于平方差总和 (SSD: sum of squared differences) [96] 的迭代算法来预测当前帧目标的运动, 为粒子滤波产生粒子的过程提供了方向信息。另外, 运动预测的残差可以用来控制粒子个数。当粒子产生后, 每一个粒子都要用图嵌入判别学习所得到的表观模型进行评价, 拥有最大评价概率的粒子将作为跟踪结果, 也将添加到目标样本中, 同时我们还需要选择一些背景样本, 用于更新图嵌入表观模型。

下面各部分将对该跟踪算法进行详细地描述, 并在最后部分对该算法进行总结。

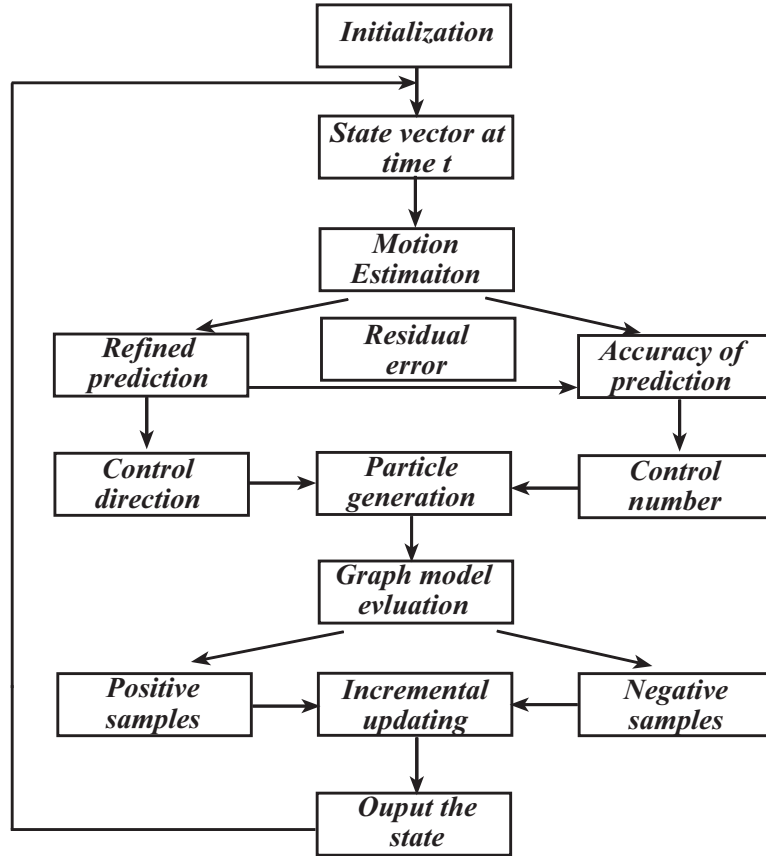


图 3-3 算法流程图

3.2.3.2 基于 SSD 的运动预测

假设目标在 $t-1$ 时刻已经被正确跟踪，且对应目标的状态表示为 s_{t-1} 。对于第 t 帧图像，我们首先对其做 SSD 迭代，并将迭代收敛后所得到的状态作为预测状态 \hat{s}_t ，为了加速迭代过程，我们采取多分辨率 SSD 来做分层运动估计，该算法收敛之后，我们将预测的信息整合到一个一阶状态转移模型中，

$$s_t = \hat{s}_t + \epsilon_t \quad (3-13)$$

其中 ϵ_t 是高斯噪声，它的方差由预测状态 \hat{s}_t 所对应的残差决定。

如图 3-4 所示，最左边为上一帧的状态，中间列是未做运动预测的粒子产生的效果，最右边是根据运动预测产生的粒子。可以看出，与未做运动预测的粒子产生模型相比，我们的方法生成的粒子更加有效，因为它们紧紧围绕在目

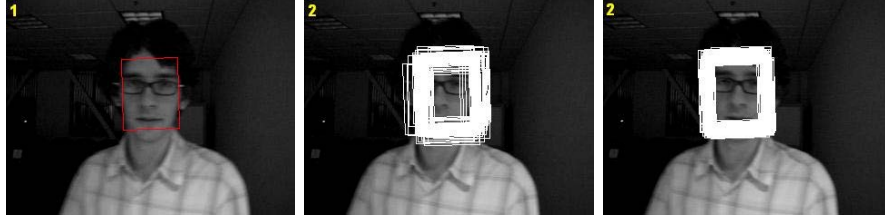


图 3-4 产生样本粒子的过程。左边：前一帧粒子状态；中间：未做运动预测产生的粒子；右边：根据运动预测产生的粒子

标物体周围，因此采用较少的粒子就可以获得令人满意的跟踪结果。

3.2.3.3 基于图嵌入的表观模型

表观模型是粒子滤波跟踪中的一个非常重要的组成部分，在这一章节，我们提出一个基于图嵌入的表观模型，该模型同时考虑了目标模式子空间以及目标与背景之间的判别子空间。

令 P 和 V 分别代表标模式子空间以及目标与背景之间的判别子空间，则表观模型可以定义如下，

$$p(o_i|z^+, z^-, P, V) \propto \exp(-\|o_i - PP^T o_i\| + \alpha(\|z^- - V^T o_i\| - \|z^+ - V^T o_i\|)) \quad (3-14)$$

其中 o_i 表示候选区域的图像观测， z^+ 和 z^- 分别代表判别空间中的目标样本和背景样本的中心， α 为权重因子。在公式 (3-14) 中，右边第一项为候选观测 o_i 在目标子空间上的重建误差，用来表示候选观测与目标子空间的相似性；而右边第二部分表示候选观测在判别子空间中的相对位置，这是一项判别性约束，类似于一项作用力，将候选观测向正样本中心推同时又将它拉离负样本中心。

3.2.3.4 负样本的启发式选择

负样本在判别学习过程中起到很重要的作用。如果负样本离目标子空间太远，则最大化类间距离并不能分开目标与其近邻的负样本。另一方面，如果负样本太接近目标子空间，而且可能发生部分重叠，这样得到的判别子空间会使得正样本分布偏离其真正的位置。

因此，我们提出一种启发式的负样本选择策略。在介绍该策略之前，我们先给出该策略采用的一些符号的定义。公式 (3-14) 中，所有候选粒子的重构误差与判别性约束分别表示为 $\{\pi_i^r, \pi_i^d\}_{i=1}^N$ ，其中 N 是粒子个数。这两个量分别代表了候选粒子与目标子空间的距离，以及在判别子空间中的相对位置。根据 $\{\pi_i^r, \pi_i^d\}_{i=1}^N$ 的分布，我们定义两个对应的阈值 $\{T^r, T^d\}$ 。负样本的启发式选择策略如下

- 如果 $(\pi_i^r > T^r) \& (\pi_i^d < T^d)$: 表示该样本离目标子空间太远;
- 如果 $(\pi_i^r < T^r) \& (\pi_i^d > T^d)$: 表示该样本离目标子空间太近;
- 如果 $(\pi_i^r < T^r) \& (\pi_i^d < T^d)$: 表示该样本离目标子空间较近，但在判别子空间中靠近背景样本群。因此，它被选中作为负样本。

3.2.3.5 在线更新

在大多数跟踪应用中，跟踪器需同时处理目标物体和环境的变化，因此，有必要对图嵌入结构和 SSD 模板进行在线更新。为了使图模型更多的依赖最近的观测，我们假定过去的的数据逐步被遗忘，新的信息逐步添加到图结构中。假设跟踪 k 帧后，我们获得 k 个目标样本和 m 个背景样本。首先，我们需要有效地更新目标子空间以及它的结构图。本文采取文献 [91] 的策略，采用 SKL 的方法 [93] 来在线更新样本子空间。由于背景样本比较混乱，因此需要长期保存每一时刻的背景样本，并且考虑到时间效率问题，背景样本采用批次替换的策略。此外，SSD 的模板也在第 $[k/2]$ 帧时刻更新，这就意味着两个不同的模型可以交替更新，以便不会同时占有计算资源。

3.2.3.6 算法总结

基于图嵌入的跟踪算法总结如下：

Algorithm 1 基于图嵌入的目标跟踪

输入: 给定第 t 时刻的目标状态 s_t 和图嵌入学习得到的子空间信息 $\{P, V, z^+, z^-\}$;

1. 将多分辨率 SSD 迭代应用到第 $t+1$ 帧获得对目标状态的预测值 \hat{s}_{t+1} ;

$$\hat{s}_{t+1} = SSD(I_{t+1}, s_t)$$

其中 I_{t+1} 为第 $t+1$ 帧图像;

2. 计算预测状态 \hat{s}_{t+1} 的残差, 然后采用一阶转移模型产生粒子;

$$s_{t+1}^{(n)} = \hat{s}_{t+1} + \epsilon_{t+1}, n = 1 \cdots N$$

3. 根据图嵌入表观模型对所有粒子进行评价;

$$\pi^{(n)} = p(o_{t+1} | s_{t+1}^{(n)}, z^+, z^-, P, V), n = 1 \cdots N$$

保留每个样本的重构误差和判别性约束值;

5. 将最大后验 (MAP, Maximum A-Posteriori) 的粒子作为目标样本;

$$s_{t+1} = \arg \max_{s_{t+1}^{(n)}} p(s_{t+1}^{(n)} | o_{1:t+1}) \approx \arg \max_{s_{t+1}^{(n)}} \pi^{(n)}$$

6. 根据 3.2.3.4 定义的策略选择负样本;

7. 更新图嵌入模型或者 SSD 模板;

输出: MAP 粒子: s_{t+1} 。

3.3 实验结果

在我们的实验中, 目标状态用仿射变化参数来表示: $s = (t_x, t_y, a_1, a_2, a_3, a_4)$, 其中 $\{t_x, t_y\}$ 表示平移量, $\{a_1, a_2, a_3, a_4\}$ 是变形参数。每个状态粒子所对应的图像通过仿射变换校正为 20×20 的图像块, 因此图像特征可以用一个 400 维的向量来表示。所有的实验都是在双 CPU 奔腾 IV 3.4GHz

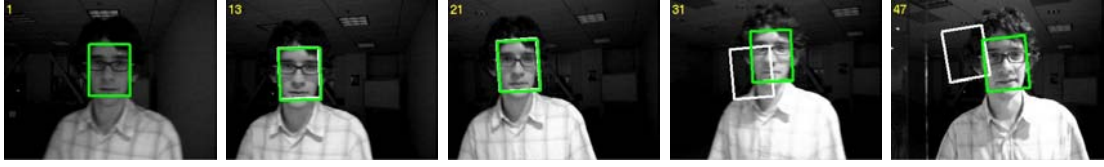


图 3-5 我们算法的跟踪结果（白色: 未做运动估计, 绿色: 运动估计）。

处理器与 512M 内存的 PC 机上运行。

首先来测试我们的算法跟踪快速移动目标的能力。为了说明基于 SSD 运动估计的重要性，我们对 David 视频序列进行隔帧采样，形成一个快速运动的测试序列。在该实验中，我们对有运动估计和未做运动估计的算法进行比较，在未做运动估计的算法中参数设置如下： $\{N = 600, var(\epsilon) = [5^2, 5^2, 0.01^2, 0.02^2, 0.002^2, 0.001^2]\}$ 。如图 3-5 所示，很明显无运动估计的算法在第 31 帧失败，因为它无法捕捉目标的快速移动。而有运动估计的算法可以实现更好的性能。图 3-6 显示两种方法所拥有的实际的粒子数，结果表明运动预测提供了一个更好的初始状态，因此需要较少的粒子即可获得满意的跟踪结果，该算法需要的平均粒子数为 206.9，相比于无运动估计的算法，在这种情况下实际上节省了近 400 个粒子。

第二部分是我们的跟踪算法与增量子空间（ISL: incremental subspace

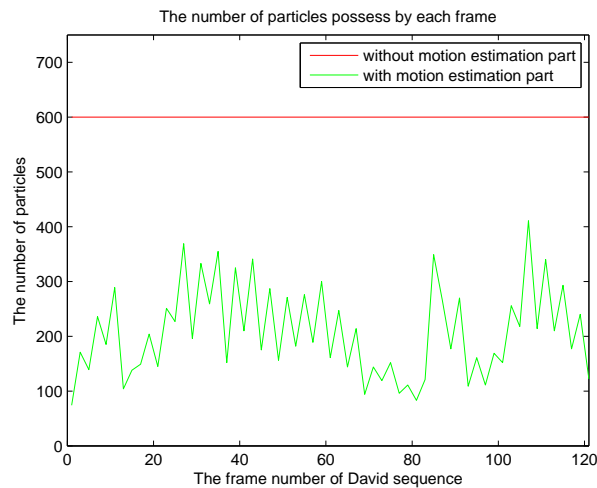


图 3-6 算法所需的粒子数（白色: 未做运动估计, 绿色: 运动估计）

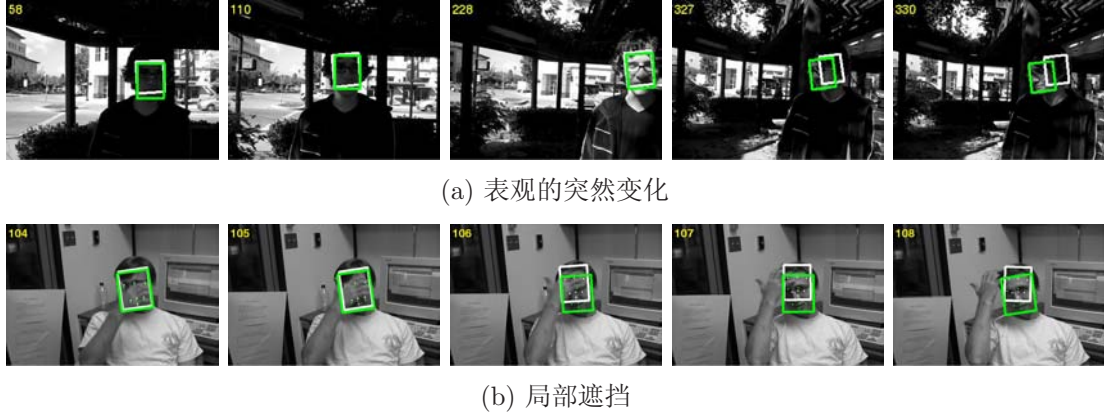


图 3-7 跟踪结果（白色: 增量子空间, 绿色: 我们的方法）。

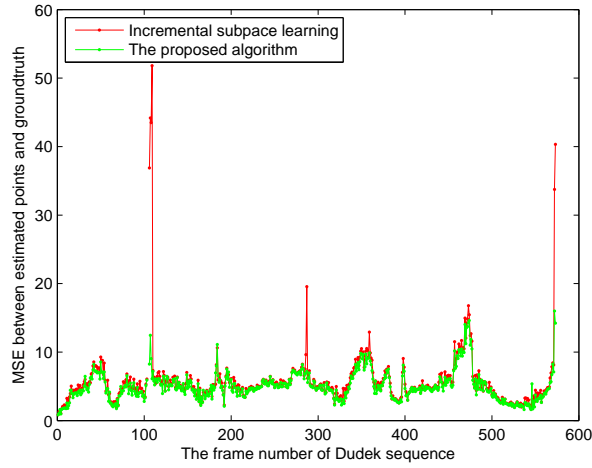


图 3-8 跟踪的均方误差曲线（红色: 增量子空间, 绿色: 我们的方法）

learning) 算法 [91] 在处理目标表观突然变化和局部遮挡方面的比较实验。如图 3-7 (a) 所示, ISL 算法能够有效的适应目标物体外观缓慢的变化, 但是当目标表观发生突然变化的话, ISL 算法将会失败。产生这种现象的原因是因为子空间的重构误差提供了一个非常小的解空间 (见图 3-1(a)), 而表观的突然变化将会超出该解空间, 因此导致跟踪失败。相反的, 我们的算法可以有效地处理外观和光线的突然变化, 原因是我们融合了背景样本和目标样本的信息从而获得一个更大的解空间来适应上述的表观变化。为了进一步说明我们算法的优越性, 我们将这两种算法在有 Groundtruth 的视频序列—— Dudek 序列上做测试, 并计算跟踪结果与 Groundtruth 之间的均方误差 (MSE, mean

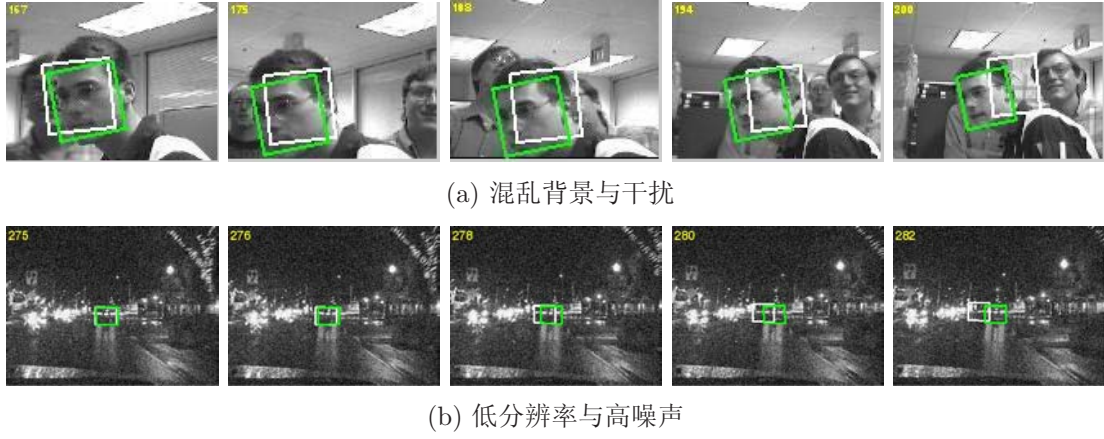


图 3-9 局部遮挡 (白色: 基于 LDA 的跟踪算法, 绿色: 我们的算法)

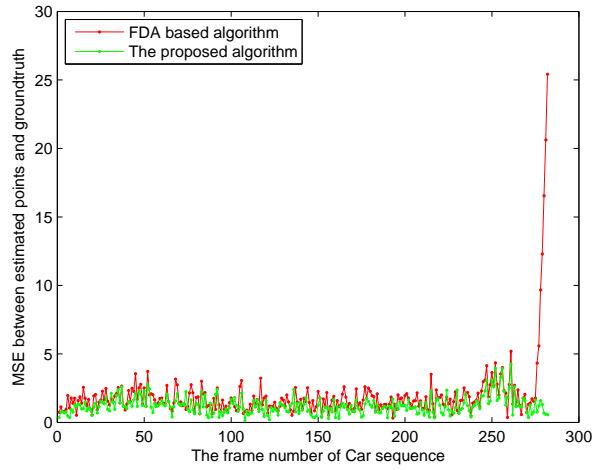


图 3-10 跟踪的均方误差曲线 (红色: 基于 LDA 的跟踪算法, 绿色: 我们的算法)

square error)。图 3-8 的结果表明, 我们的方法实现了更精确的跟踪定位, 特别是在 105 帧图像附近当目标发生局部遮挡的时候。我们算法的平均均方误差是 4.8194, 而 ISL 的均方误差是 5.7386。图 3-7 (b) 显示了两种算法局部遮挡下的跟踪结果, 很显然, 判别性约束能将我们的跟踪器约束在 groundtruth 的位置附近, 使它不被遮挡物带走。

第三部分对我们的算法与基于 LDA 的跟踪算法在判别性方面进行了比较。如图 3-9 所示, 我们将这两种方法在两组视频序列分别进行测试。第一组视频序列具有混乱的背景, 且背景中包含与目标物体外观相似的对象。第二组视频序列是一组成像质量比较差, 具有高噪声的序列。从跟踪的结果可以看出 (见

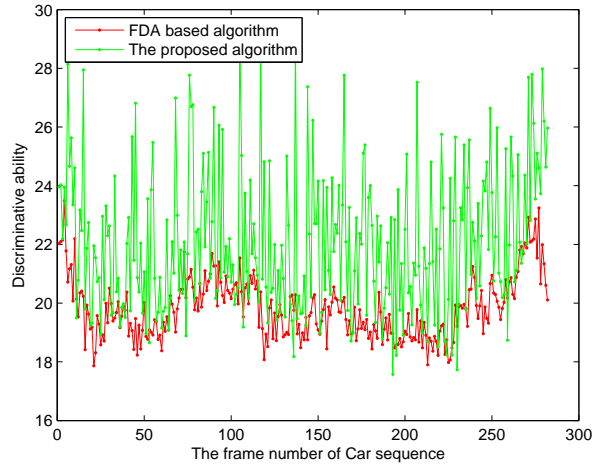


图 3-11 判别能力 (红色: 基于 LDA 的跟踪算法, 绿色: 我们的算法)

图 3-9)，基于 LDA 的跟踪算法，逐步偏离了 groundtruth 位置，最终完全丢失了待跟踪的目标，而我们的算法在两段视频序列上都取得了令人满意的结果。为了分析上述两种算法表现不同的原因，我们对它们的跟踪精度和判别性能力进行了研究，其中 t 时刻算法的判别能力定义如下，

$$Discr(t) = N_n^{-1} \sum_{i=1}^{N_n} \|z^+ - z_i^-\|^2 \quad (3-15)$$

其中 z_i^- 是第 i 个负样本在判别空间中的投影， N_n 是负样本的个数。图 3-10 和图 3-11 显示了述两种算法在第二组序列上的均方误差以及判别能力。从这两幅图我们可以看出，我们的方法实现了更准确的定位，而且比 LDA 更具有判别力。由于 LDA 准则是在类内散度特定的约束下来最大化类间散度。所以如果背景样本分布非常凌乱，类内散度将非常大，这将使得对类内散度的约束放的非常宽，从而对类间的最大能力将相应的减小。而我们的算法是基于 k 近邻的局部连接，因此类内散度的约束较小。另外，我们把样本的分布嵌入到图结构中，也在一定程度上加强了判别能力。

第四章 确定性搜索和随机性搜索相结合的跟踪框架

总的来说，现有的跟踪算法的框架从优化角度来看大致可以分为确定性搜索和随机性搜索两大类。典型的确定性跟踪框架根据事先设定的相似度损失函数（Similarity Cost Function），在当前图像帧内的局部区域通过梯度优化的迭代方法找到与目标标准模板最为相似的区域。确定性跟踪框架一般计算量较小，效率较高，但容易陷入局部极小值，导致跟踪的结果不够精确。随着跟踪误差的积累，最后可能导致跟踪失败。随机性跟踪框架利用状态空间对当前跟踪系统进行模型化，通过随机产生粒子来逼近状态的后验分布。粒子滤波（Particle Filter），也被称为序贯蒙特卡罗（sequential Monte Carlo）方法，是目前最为广泛使用的随机性方法。与确定性跟踪框架相比，随机性跟踪框架更为鲁棒，但是计算量过大，而且复杂度随着目标状态的增长呈指数增长。另外，由于没有好的机制选择重要性分布，粒子滤波器存在着严重的样本退化问题。

本章讨论了两种结合确定性跟踪框架和随机性跟踪框架的方法：（1）核贝叶斯跟踪框架；（2）基于奇异值分解的卡尔曼——粒子滤波框架。在工作（1）中，我们首先将核方法应用于当前帧，获得对运动方向以及偏移量的估计，再将这些信息作为启发式先验信息加入到贝叶斯滤波框架中，从而有效结合确定性方法与随机性方法的优点，并且该框架是一种通用的跟踪框架，任何形式的表观模型都可以在此框架下应用。我们采用了一种基于空间约束的混合高斯模型来描述目标物体，该模型将目标的空间分布信息加入到混合高斯模型中，从而更加有效的解决了复杂背景以及局部遮挡所引发的困难。与基于核方法的跟踪器（确定性跟踪方法）以及基于贝叶斯方法的跟踪器（随机性跟踪方法）相比，该跟踪框架有效地提高了跟踪的效率和精确度。在工作（2）中，我们首先对前一帧跟踪样本的协方差矩阵进行奇异值分解，再利用 Unscented 变化

（一种确定性的变化）来获得预测样本，然后用卡尔曼滤波器对预测样本进行滤波，为最终的粒子滤波提供运动估计分布。与粒子滤波和基于 Unscented 卡尔曼——粒子滤波相比，我们的方法更为鲁棒。

4.1 核贝叶斯框架

在介绍核贝叶斯跟踪框架前，我们先来回顾一下两种典型的确定性跟踪框架和随机性跟踪框架：核方法与贝叶斯方法。

4.1.1 核方法

基于核的跟踪算法首先分别对目标图像模板和候选区域建立核密度分布的估计，然后通过最小化两个核密度分布的距离来实现目标的跟踪。最有名的核跟踪算法是均值漂移（mean shift），该算法最早由 Fukunaga 等人提出 [97]，是一种非参数化的对密度函数的梯度进行估计并进行模式搜索的方法。令 x 为 n 维空间 X 中的一个样本点， $A \subset X$ 为样本 x 的邻域点的集合， $K(\cdot)$ 为一核函数， $w : x \rightarrow (0, +\infty)$ 为权重函数，均值漂移向量定义如下：

$$ms = \frac{\sum_a K(a-x)w(a)a}{\sum_a K(a-x)w(a)} - x, a \in A, x \in X \quad (4-1)$$

均值漂移向量指向下列侧影函数的梯度方向：

$$q(x) = \sum_a H(a-x)w(a), a \in A \quad (4-2)$$

其中 $H(\cdot)$ 为核函数 $K(\cdot)$ 的侧影核。侧影核通常采用 Epanechnikov 函数：

$$\mathcal{H}(x) = \begin{cases} 1 - \|x\|, & \|x\| < 1 \\ 0, & \|x\| > 1 \end{cases} \quad (4-3)$$

对应的，核函数 $K(\cdot)$ 为常数核：

$$\mathcal{K}(x) = \begin{cases} 1, & \|x\| < 1 \\ 0, & \|x\| > 1 \end{cases} \quad (4-4)$$

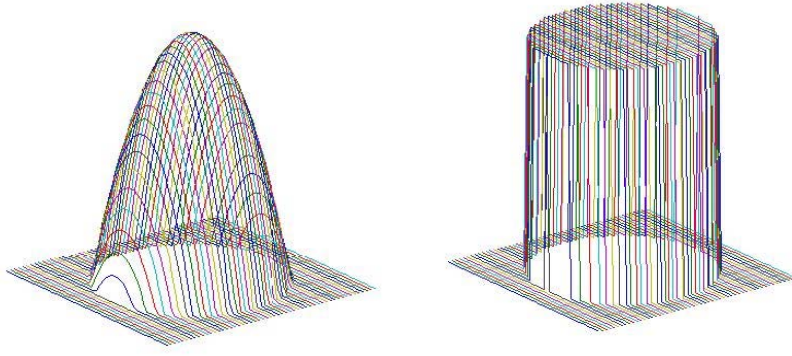


图 4-1 核函数示意图（左边: Epanechnikov 核, 右边: 常数核）

Epanechnikov 核与常数核如图 4-1 所示。

均值漂移算法通过重复以下步骤来搜索 x 点附近的局部极大值:

- ① 根据公式 (4-1) 计算 x 点的偏移向量 ms ;
- ② 如果 ms 小于一定的阈值, 则表示迭代收敛; 否则更新 x 点的位置 $x + ms \rightarrow x$ 。

在基于均值漂移的跟踪算法中, 目标模板与候选区域用核颜色直方图表示, 直方图之间的相似性采用 Bhattacharyya 系数来度量, 并通过均值漂移迭代达到收敛。

4.1.2 贝叶斯方法

另一种很常用的方法是将跟踪过程看成一个贝叶斯推理过程, 通过一系列观测 $o_{1:t}$ 来推理出目标的运动状态 s_t , 即估计目标状态的后验概率分布 $p(s_t|o_{1:t})$ 。在贝叶斯推理的框架中, 通常采用状态空间对当前跟踪系统的运动进行模型化,

$$\text{state transition model} : s_t = f_t(s_{t-1}, \epsilon_t) \leftrightarrow p(s_t|s_{t-1}) \quad (4-5)$$

$$\text{observation model} : o_t = h_t(s_t, \nu_t) \leftrightarrow p(o_t|s_t) \quad (4-6)$$

其中 s_t, o_t 分别代表系统（目标）状态和观测， ϵ_t, ν_t 分别表示系统噪声和观测噪声， $f_t(\cdot, \cdot)$ 和 $h_t(\cdot, \cdot)$ 分别表示状态转移模型和观测模型。这两个模型分别由状态转移分布 $p(s_t|s_{t-1})$ 和观测分布 $p(o_t|s_t)$ 决定。

在贝叶斯推理过程中，后验分布可以表示为：

$$p(s_t|o_{1:t}) \propto p(o_t|s_t)p(s_t|o_{1:t-1}) \quad (4-7)$$

其中 $p(s_t|o_{1:t-1})$ 表示对目标状态的预测，它可以通过将上一时刻的后验分布在时间轴上传播得到，

$$p(s_t|o_{1:t-1}) = \int p(s_t|S_{t-1})p(s_{t-1}|o_{1:t-1})ds_{t-1} \quad (4-8)$$

当状态转移模型和观测模型为非线性和非高斯时，上述积分形式没有解析解，因此我们需要采用数值逼近的方法来求解。粒子滤波是一种被广泛应用的基于重要性采样的后验分布逼近算法。它的主要思想是利用一堆采样点 $\{s_t^i\}_{i=1}^N$ 来逼近目标状态的后验分布，即 $p(s_t|o_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta(s_t - s_t^i)$ ，其中 $\delta(\cdot)$ 为冲击函数。但是在实际应用中，我们无法直接对后验分布进行采样，因此粒子滤波通过重要性采样来解决上述问题。定义一个方便采样的分布，称作重要性分布 $q(\cdot)$ ，然后从重要性分布中采样 N 个样本，并对每个样本进行评价，从而获得粒子样本的权重，

$$w_t^i \propto \frac{p(o_t|s_t^i)p(s_t^i|s_{t-1}^i)}{q(s_t^i|s_{t-1}^i, o_{1:t})} \quad (4-9)$$

最后后验分布可以表示为

$$p(s_t|o_{1:t}) = \sum_{i=1}^N w_t^i \delta(s_t - s_t^i) \quad (4-10)$$

在跟踪应用中，最大后验估计或者样本均值都可以作为跟踪结果的输出。

4.1.3 核贝叶斯方法

核方法简单易行且计算量小，但是容易陷入局部极小值，而贝叶斯方法通过引进随机采样来提高跟踪的鲁棒性，但是随机采样需要大量的样本，这也增

加了巨大的计算负担。因此，我们提出核贝叶斯方法来结合两者的优点，并克服各自的缺点。

在贝叶斯跟踪方法中，通常将状态转移分布作为重要性分布来采样。将 $q(\cdot) = p(s_t | s_{t-1})$ 代入到公式 (5-4) 中，样本粒子的权重可简化为 $w_t^i = p(o_t | s_t^i)$ 。如何选择状态转移分布（模型）对跟踪性能有着直接影响。现有跟踪算法的状态转移模型大致可以分为两类：简单随机游走和基于学习的回归模型。前一类方法没有利用任何运动信息，因此需要产生大量的粒子且游走所覆盖的区域要足够大，这样才能保证粒子能够捕捉到目标物体。同时该算法非常耗时，当目标状态维数增加时，会陷入维数灾难。后一类方法通过学习来获得目标物体的运动规律，但是该方法容易导致过拟合，使得这类状态转移模型只对特定视频有效。

由于均值偏移向量提供目标的运动方向信息，因此我们将均值偏移嵌入到贝叶斯框架下，为贝叶斯推理提供启发式先验信息。首先我们对当前观测做均值漂移迭代来获得目标运动的方向和偏移量，并记录迭代结果与目标模板的相似度，然后把运动方向、偏移量融入到一阶状态转移模型中，而相似度则用于控制产生粒子的个数及其覆盖的范围，最后通过贝叶斯推理对后验分布进行估计。在这一方式下，我们可以将核方法和贝叶斯方法结合到一个统一的框架中。

4.1.4 讨论和分析

在这一部分，我们从优化的角度来分析核贝叶斯方法为何能结合两种方法的优点。如图 4-2 所示，这幅输入图像上有三个窗口，分别对应状态空间中的初始状态，观测似然分布的局部极大值和全局极大值。假如我们直接从初始状态产生粒子，那么需要大量粒子以及很大的覆盖范围才能找到全局极大值。但是如果利用均值漂移事先移动到局部极大值的位置上，然后在此基础上可以利用少量的粒子就可以到达全局极大值。在核贝叶斯框架下，核方法用确定

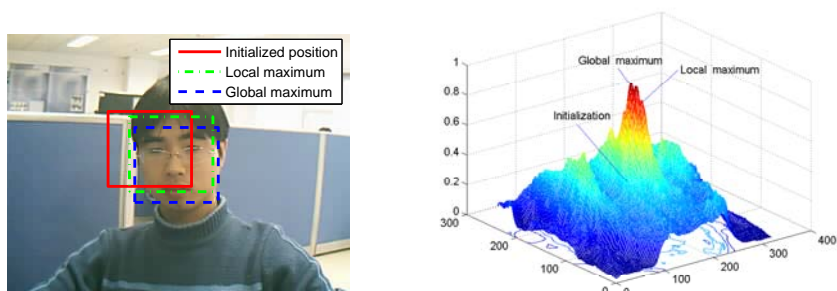


图 4-2 输入图像的三个窗口（左边）分别对应状态空间中的初始状态，观测似然分布（右边）的局部极大值和全局极大值

性搜索提供了更好的初始位置和先验信息，贝叶斯方法则利用随机采样来搜索全局最优值。

4.2 核贝叶斯跟踪算法

基于核贝叶斯的跟踪算法流程见图 4-3。首先利用核方法来预测更好的初始状态和预测位置的精确度，并将这些先验信息融入到一阶状态转移模型中，然后利用该模型产生粒子，并通过一个基于空间约束混合高斯的表观模型对每个粒子进行评价。最后，将最大后验估计作为跟踪结果输出，并对表观模型进行更新。

4.2.1 基于空间约束混合高斯的表观模型

我们采用基于空间约束混合高斯模型来对目标物体的表观进行建模，该模型的参数通过在线期望最大化（Expectation Maximization）算法来估计。

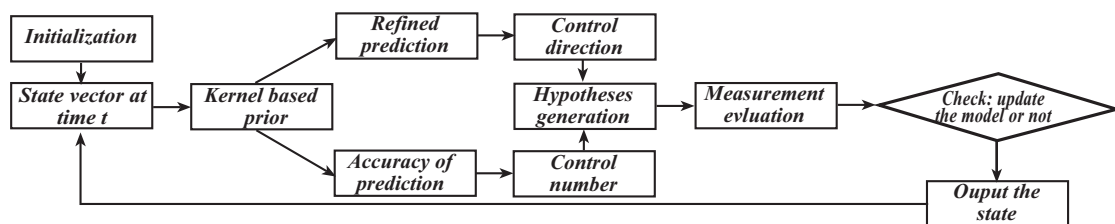


图 4-3 基于核贝叶斯的跟踪算法流程

4.2.1.1 表观模型

首先我们来看文献中 [20] 的表观模型，他们的表观模型由三个组成部分 S, W, F ，其中 S 代表跟踪过程中模型的稳定成分， W 用于描述两帧间的差异， F 为初始化时的固定模板。固定模板 F 用于防止表观模型因误差积累而失真。然而，这个表观模型中各个像素是独立的，并且每个像素的重要性都是一样的，忽略了目标物体的空间分布信息。因而该模型在以下两种情况下可能会导致跟踪失败：1) 背景中有与目标物体相似的干扰；2) 局部遮挡。在我们的工作中，我们将空间约束加入到该表观模型中，该空间约束为一个二维的高斯分布，如图 4-4 所示，这个高斯分布的均值为窗口的中心，协方差为窗口两个方向的尺度。该空间约束认为距离窗口中心点越近的像素越重要，这个假设与跟踪中目标物体的实际情况比较吻合，距离中心点较远像素经常是背景，另外局部遮挡也往往都先在目标物体边缘部分发生。

根据上述空间约束，我们的表观模型的似然函数定义如下：

$$p(o_t|s_t) = \prod_{j=1}^d \left\{ N(x(j); x_c, \Sigma_c) * \sum_{i=s,w,f} \pi_{i,t}(j) N(o_t(j); \mu_{i,t}(j), \sigma_{i,t}^2(j)) \right\} \quad (4-11)$$

其中 $N(x; \mu, \sigma^2)$ 表示高斯分布，

$$N(x; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \quad (4-12)$$

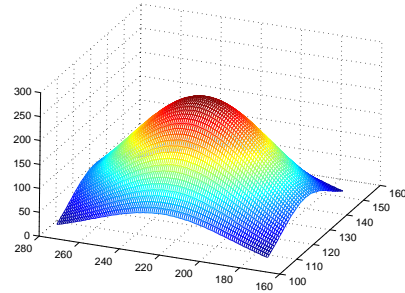


图 4-4 基于二维高斯分布的空间约束

$\{\pi_{i,t}, \mu_{i,t}, \sigma_{i,t}, i = s, w, f\}$ 表示混合高斯各组成部分的权重, d 为候选区域中的像素个数, x_c 和 Σ_c 为空间约束高斯分布的均值和协方差矩阵。

4.2.1.2 参数估计

在跟踪过程, 目标物体的表观会发生变化, 因此, 需要对表观模型的参数进行实时地更新。为了使模型参数更多地反映与当前时刻较接近的表观变化, 我们采用指数衰减的方式对不同时时刻的图像观测进行遗忘处理, 这样表观模型将更多地关注近期内的变化。另外, 为了避免保留所有的图像观测数据, 我们采用在线期望最大化算法来对模型参数进行估计和更新。

① 在 E 步骤中, 混合高斯中各组成成分的后验概率可以计算如下,

$$m_{i,t}(j) \propto \pi_{i,t}(j)N(o_{i,t}(j); \mu_{i,t}(j), \sigma_{i,t}^2(j)) \quad (4-13)$$

② 各组成成分的权重更新如下,

$$\pi_{i,t+1}(j) = \alpha m_{i,t}(j) + (1 - \alpha)\pi_{i,t}(j); i = s, w, f \quad (4-14)$$

另外, 候选区域的矩特征统计量 $\{M_{k,t+1}; k = 1, 2\}$ 更新如下,

$$M_{k,t+1}(j) = \alpha o_t^k(j)m_{s,t}(j) + (1 - \alpha)M_{k,t}(j); k = 1, 2 \quad (4-15)$$

③ 在 M 步骤中, 各组成成分的均值和方差更新如下,

$$\mu_{s,t+1}(j) = \frac{M_{1,t+1}(j)}{\pi_{s,t+1}(j)}, \sigma_{s,t+1}^2 = \frac{M_{2,t+1}(j)}{\pi_{s,t+1}(j)} - \mu_{s,t+1}^2(j)$$

$$\mu_{w,t+1}(j) = o_t(j), \sigma_{w,t+1}^2(j) = \sigma_{w,1}^2(j)$$

$$\mu_{f,t+1}(j) = \mu_{f,1}(j), \sigma_{f,t+1}^2(j) = \sigma_{f,1}^2(j)$$

4.2.2 核贝叶斯跟踪框架

如 §4.1.3 中所述, 我们将均值漂移嵌入到贝叶斯推理框架下, 为状态转移模型提供启发式先验, 因此能够在一定程度上克服贝叶斯框架中样本退化的问题。

假设 $t-1$ 时刻已经得到正确的跟踪结果, 且跟踪结果的状态表示为 s_{t-1} 。对于第 t 帧图像, 我们首先对其做均值漂移迭代, 但是如何将我们的表观模型表示成均值漂移可以迭代的形式呢? 文献 [98] 证明了只有对称的核函数才能在均值漂移框架下迭代, 因此我们定义如下加权函数,

$$\omega(x) = N(x, x_c, \Sigma_c) \sum_{i=w,s,f} \pi_{i,t}(j) N(o_t(x); \mu_{i,t}(x), \sigma_{i,t}^2(x)) \quad (4-16)$$

另外由于我们采用常数核, 所以均值偏移迭代可以表示为,

$$\hat{x}_t = \frac{\sum_{x_i=1}^d w(x_i) x_i}{\sum_{x_i=1}^d w(x_i)}, x_i \in \text{candidate}. \quad (4-17)$$

我们把该迭代收敛的状态作为预测状态 \hat{s}_t , 并将该预测状态加入到一阶状态转移模型中,

$$s_t = \hat{s}_{t-1} + \text{Affine}(\hat{x}_t - x_{t-1}) + \epsilon_t \quad (4-18)$$

其中 Affine 表示仿射变换, 同时, 该预测位置的精度可以用过表观似然函数来评价, 并用于控制 ϵ_t 和样本个数。基于产生的粒子, 后续步骤采用经典的贝叶斯推理算法来获得跟踪结果。

4.2.3 表观模型的选择性更新

在实际中, 对每帧图像的表观模型都采用 §4.2.1.2 中的方法进行参数更新有时候效果并不好。例如, 当一部分背景被错误的引入到目标模板中, 或者部分目标区域被遮挡。因此, 我们需要设计一种选择性更新的策略来避免对目标模型的错误更新。

表 4-1 表观模型的选择性更新

1:	if ($\pi_a > T_a$)
2:	if ($\pi_{sw} > T_{sw}$)&&($\pi_f > T_f$)
3:	更新目标的表观模型;
4:	else if ($\pi_{sw} > T_{sw}$)&&($\pi_f \leq T_f$)
5:	只更新目标表观模型中的 SW 成分;
6:	else if ($\pi_{sw} \leq T_{sw}$)&&($\pi_f > T_f$)
7:	只更新目标表观模型中的 F 成分;
8:	else if ($\pi_{sw} \leq T_{sw}$)&&($\pi_f \leq T_f$)
9:	不更新;
10:	end if
11:	end if

当每帧跟踪结束时，我们获得该帧的最大后验估计作为跟踪结果。首先，我们对该状态对应的观测分别用表观模型， SW 组合成分， F 成分进行评价，得到三个评价值，分别为 π_a, π_{sw}, π_f 。同时，我们定义三个对应的阈值 $\{T_a, T_{sw}, T_f\}$ ，则选择性更新的策略如表 4-1 所示。

其中 SW 用于表示目标表观的变化， F 用于防止目标表观模型的漂移 (drifting)。从表 4-1 可以看出，我们的选择性更新的策略既考虑了目标表观的变化，又考虑了如何防止目标表观的漂移。

4.2.4 实验结果

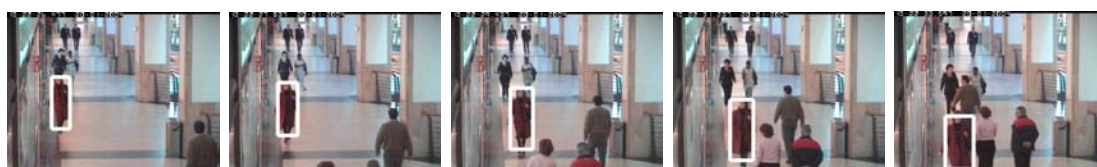
在我们的实验中，目标状态用仿射变化参数来表示： $s = (t_x, t_y, a_1, a_2, a_3, a_4)$ ，其中 $\{t_x, t_y\}$ 表示平移量， $\{a_1, a_2, a_3, a_4\}$ 是变形参数。每个状态粒子所对应的图像通过仿射变换校正为 30×15 维的图像块，因此图像特征可以用一个 450 的向量来表示。所有的实验都是在双 CPU 奔腾 IV 3.2GHz

处理器与 512M 内存的 PC 机上运行。

4.2.4.1 单目标跟踪

在这一部分，我们设计了三个实验来验证核贝叶斯跟踪算法的性能。

第一个实验将核贝叶斯跟踪算法分别与贝叶斯跟踪算法以及核跟踪算法在跟踪精度和效率方面进行比较。如图 4-5 所示，第一行是基于核贝叶斯的跟踪算法，从中可以看出我们的算法能非常精确地对目标进行跟踪。第二行是贝叶斯算法的跟踪结果，该算法取得了与我们算法类似的跟踪精度。第三行是核方法的跟踪结果，由于核方法容易陷入局部极值，因而在跟踪定位方面不够精确，另外核方法无法处理目标尺度的变化。为了进一步比较三种算法的性能，我们对它们的跟踪精度和算法运行时间进行了定量的分析，其中跟踪精度采用跟踪结果与 groundtruth 之间的均方误差来衡量。如图 4-6 所示，左边为算法运行时间，右边是跟踪精度。可以看出，核方法效率最高但是定位精度最差。相反的，贝叶斯方法由于产生了大量的粒子，获得了更为精确的跟踪定位结果。我们将核贝叶斯算法和贝叶斯算法进行具体比较，贝叶斯算法的平均运



(a) 核贝叶斯跟踪算法



(b) 贝叶斯跟踪算法



(c) 核跟踪算法

图 4-5 跟踪结果

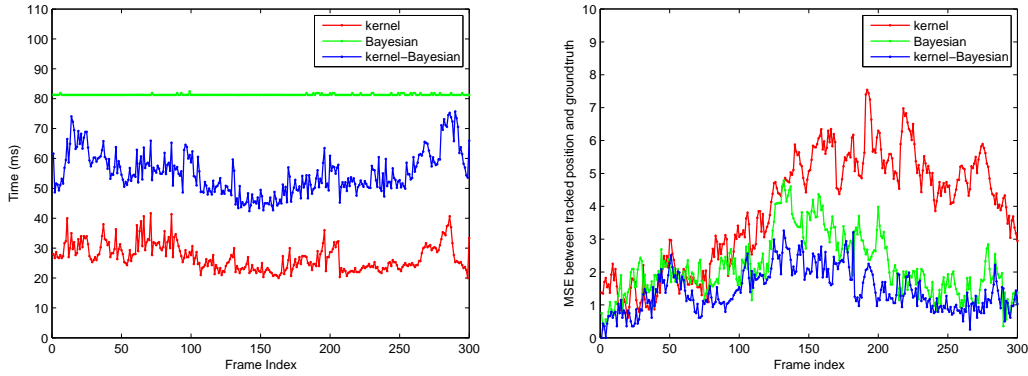


图 4-6 左边：算法运行时间，右边：跟踪结果的均方误差（红色：核方法，绿色：贝叶斯方法，蓝色：核贝叶斯方法）

行时间为 81 毫秒，平均均方误差为 8.6521。而核贝叶斯算法的平均运行时间为 55 毫秒，平均均方误差为 5.8012。在核贝叶斯框架下，由于核方法为状态转移模型提供了启发式先验，有效地克服了贝叶斯方法中的样本退化问题，因此获得了最佳的跟踪性能。

在第二个实验中，我们将基于空间约束混合高斯的表观模型（SMOG: spatial-constraint mixture of Gaussians）与不加空间约束的混合高斯表观模型（MOG: mixture of Gaussians）进行比较。如图 4-7 所示，可以看出基于 SMOG 的表观获得了更令人满意的跟踪结果。原因是当背景中出现与目标表观相似的物体时，相比于 MOG，SMOG 抓住了目标的空间分布信息，比 MOG 表观模型更具判别性，因此能更好区分目标与背景。

最后一个实验用于测试我们提出的选择性更新策略。如图 4-8 所示，我们的更新策略能够很好地适应光线的变化以及目标表观的变化，并且目标表观模型没发生漂移现象。

4.2.4.2 多目标跟踪

在这个部分中，我们将核贝叶斯算法扩展到多目标跟踪的任务。如图 4-9，尽管存在局部遮挡，该算法依旧可以很好地跟踪序列中的三个目标。与贝叶斯方法相比，由于我们的算法占用的计算资源更少，因此具有更大的潜力去实时



(a) SMOG 的跟踪性能



(b) MOG 的跟踪性能

图 4-7 复杂环境下的跟踪结果



(a) 光线变化



(b) 局部遮挡

图 4-8 不同场景下的实验结果

地跟踪更多的目标。

4.3 基于奇异值分解的卡尔曼粒子滤波算法

另一种广受关注的结合确定性搜索和随机性搜索的方法是基于 Unscented 卡尔曼粒子滤波算法，该算法首先利用 Unscented 变化（一种确定性的变化）获得一批 sigma 样本，然后采用卡尔曼算法对这些样本进行滤波并获得滤波后 sigma 样本的均值和协方差矩阵，最后根据上面获得的均值和协方差矩阵为

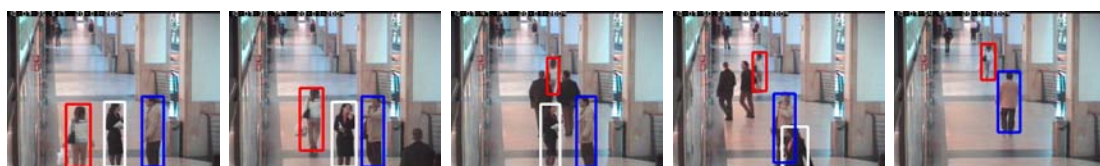


图 4-9 多目标跟踪结果

粒子滤波提供重要性分布。但是在利用 Unscented 变化产生 sigma 样本的过程中需要对样本协方差矩阵求平方根，由于协方差矩阵的随机性，在求解平方根过程经常遇到病态问题，从而影响了跟踪的性能。因此，在本章中我提出了基于奇异值分解的卡尔粒子滤波算法，在该算法中我们通过对协方差矩阵做奇异值分解（SVD: singular value decomposition）来获得 sigma 样本，实验结果证明我们的算法无论在计算稳定度还是跟踪的鲁棒性方面都优于传统的基于 Unscented 卡尔曼粒子滤波算法。

4.3.1 基于 Unscented 卡尔曼粒子滤波算法

4.3.1.1 Unscented 卡尔曼滤波器

在介绍基于 Unscented 卡尔曼的粒子滤波算法之前，我们首先来介绍一下 Unscented 卡尔曼滤波器（UKF: unscented Kalman filter），相比于扩展卡尔曼滤波器（EKF: extended Kalman filter）只能估计非高斯数据的一阶统计信息，Unscented 卡尔曼滤波器能精确地估计非高斯数据的二阶统计量。

在实现过程中，UKF 采用扩展状态空间，即 $s_{t-1}^a = [s_{t-1}^T \epsilon_{t-1}^T \nu_{t-1}^T]$ ，其中 $s_{t-1}, \epsilon_{t-1}, \nu_{t-1}$ 分别为 $t-1$ 时刻目标的状态，系统噪声和观测噪声。 s_{t-1}^a 的维数和协方差矩阵分别表示为 $N_a = N_x + N_\epsilon + N_\nu$ 和 P_{t-1}^a 。这里同样采用公式（5-19）和（5-20）中的状态空间模型，于是 Unscented 卡尔曼滤波器算法可以表述如下：

- ① 根据 Unscented 变化产生 $2N_a$ 个 sigma 样本以及各自样本的权重

$$\begin{aligned} \mathcal{S}_{0,t-1}^{(i)a} &= \bar{s}_{t-1}^{(i)a} \\ \mathcal{S}_{j,t-1}^{(i)a} &= [\bar{s}_{t-1}^{(i)a} \bar{s}_{t-1}^{(i)a} \pm \sqrt{(n_a + \lambda) P_{t-1,j}^a}] \\ W_0^{(m)} &= \frac{\lambda}{N_a + \lambda}, \quad W_0^{(c)} = \frac{\lambda}{N_a + \lambda} + (1 - \alpha^2 + \beta) \\ W_j^{(m)} &= W_j^{(c)} = \frac{1}{2(N_a + \lambda)}, \quad \lambda = \alpha^2(N_a + \kappa) - N_a \\ j &= 1, \dots, 2N_a \end{aligned}$$

② sigma 样本根据状态空间模型进行演化:

$$\begin{aligned}\mathcal{S}_{j,t|t-1}^{(i)s} &= f(\mathcal{S}_{j,t-1}^{(i)s}, \mathcal{S}_{j,t-1}^{(i)\epsilon}), \bar{s}_{t|t-1}^{(i)} = \sum_{j=0}^{2N_a} W_j^{(m)} \mathcal{S}_{j,t|t-1}^{(i)s} \\ P_{t|t-1}^{(i)} &= \sum_{j=0}^{2N_a} W_j^{(c)} [\mathcal{S}_{j,t|t-1}^{(i)s} - \bar{s}_{t|t-1}^{(i)}][\mathcal{S}_{j,t|t-1}^{(i)s} - \bar{s}_{t|t-1}^{(i)}]^T \\ \mathcal{O}_{j,t|t-1}^{(i)} &= h(\mathcal{S}_{j,t-1}^{(i)s}, \mathcal{S}_{j,t-1}^{(i)\nu}), \bar{o}_{t|t-1}^{(i)} = \sum_{j=0}^{2N_a} W_j^{(m)} \mathcal{O}_{j,t|t-1}^{(i)}\end{aligned}$$

③ 对观测进行验证:

$$\begin{aligned}P_{o_t, o_t} &= \sum_{j=0}^{2N_a} W_j^{(c)} [\mathcal{O}_{j,t|t-1}^{(i)} - \bar{o}_{t|t-1}^{(i)}][\mathcal{O}_{j,t|t-1}^{(i)} - \bar{o}_{t|t-1}^{(i)}]^T \\ P_{s_t, o_t} &= \sum_{j=0}^{2N_a} W_j^{(c)} [\mathcal{S}_{j,t|t-1}^{(i)s} - \bar{s}_{t|t-1}^{(i)}][\mathcal{O}_{j,t|t-1}^{(i)} - \bar{o}_{t|t-1}^{(i)}]^T \\ K_t &= P_{s_t, o_t} P_{o_t, o_t}^{-1}, \bar{s}_t^{(i)} = \bar{s}_{t|t-1}^{(i)} + K_t(o_t - \bar{o}_{t|t-1}^{(i)}) \\ \bar{P}_t^{(i)} &= P_{t|t-1}^{(i)} - K_t P_{o_t, o_t} K_t^T\end{aligned}$$

4.3.1.2 基于 Unscented 卡尔曼粒子滤波算法

根据上述 Unscented 卡尔曼滤波器所滤波后所得到的样本均值和协方差矩阵 $\bar{s}_t^{(i)}, \bar{P}_t^{(i)}$, 我们定义了如下重要性分布函数 $q(s_t^i | s_{t-1}^i, o_{1:t}) = \mathcal{N}(\bar{s}_t^{(i)}, \bar{P}_t^{(i)})$ 。基于上述重要性分布函数, 我们在上述 Unscented 卡尔曼滤波器算法后面加上如下步骤即可得到基于 Unscented 卡尔曼粒子滤波算法:

① 从如下重要性分布中产生粒子,

$$s_t^i \sim N(\bar{s}_t^{(i)}, \bar{P}_t^{(i)})$$

② 利用表观模型对粒子进行评价

$$w_t^i = p(o_t^i | s_t^i)$$

③ 输出最大后验估计或者样本均值作为跟踪结果。

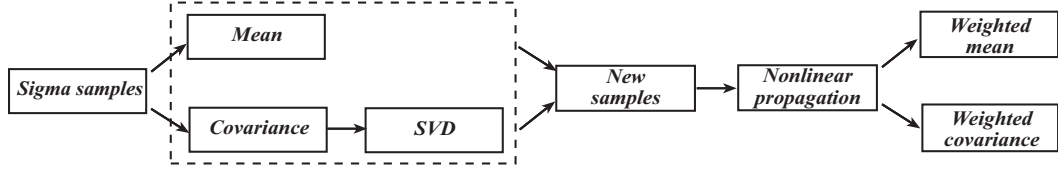


图 4-10 算法流程图

4.3.2 基于奇异值分解的卡尔曼粒子滤波算法

在 §4.3.1.1 我们曾经说过，在利用 Unscented 变化产生 sigma 样本的过程中需要对样本协方差矩阵求平方根，由于协方差矩阵的随机性，在求解平方根过程中常遇到病态问题，从而影响了跟踪的性能。因此，在本章中我们提出了基于奇异值分解的卡尔粒子滤波算法，如算法流程图 4-10 所示，在该算法中我们首先通过对协方差矩阵做奇异值分解（SVD: singular value decomposition）来获得 sigma 样本，然后经过卡尔曼滤波获得 sigma 样本的均值和协方差矩阵，为后续的粒子滤波提供重要性采样分布，最后由粒子滤波输出跟踪结果。

利用 SVD 分解产生 sigma 样本的具体过程如下：

- ① 计算 sigma 样本协方差矩阵的 SVD 分解，

$$P_{t-1}^a = U_{t-1} S_{t-1} V_{t-1}^T$$

- ② 产生新的 sigma 样本:

$$\mathcal{X}_{0,t-1}^{(i)a} = \bar{x}_{t-1}^{(i)a}$$

$$\mathcal{X}_{j,t-1}^{(i)a} = [\bar{x}_{t-1}^{(i)a} \bar{x}_{t-1}^{(i)a} \pm \rho U_{j,t-1} \sqrt{s_{j,t-1}}]$$

其中 $U_{j,t-1}, s_{j,t-1}$ 分别为第 jth 特征向量和特征值， ρ 是尺度参数。

通过 SVD 分解来产生 sigma 样本的主要原因是：sigma 样本协方差矩阵所包含的信息可以被其特征值和特征向量充分描述，并且 SVD 分解比 Cholesky 分解在数值计算方面要更为鲁棒。

4.3.3 基于增量子空间学习的表现模型

在该工作中，我们采用基于增量子空间的表观模型，该模型通过逐步学习目标表观的低维特征空间表示来适应跟踪过程中目标表观的变化。

给定前 t 帧目标表观的数据矩阵 $A = \{I_1, \dots, I_t\}$ 以及该数据矩阵的 SVD 分解 $A = U\Sigma V^T$ ，其中 I_i 是第 i 帧目标表观的数据。在又经过 k 帧的跟踪后，我们获得了 k 个新的目标表观的数据 $E = \{I_{t+1}, \dots, I_{t+k}\}$ ，对于这样一个新的数据矩阵 $A' = (A|E)$ ，我们引入 R-SVD 算法 [99] 来对它进行低复杂度的 SVD 分解。该算法在已有的奇异值分解和新加入的行或列数据的基础上，可以有效地对新矩阵进行奇异值分解。具体算法如下：

① 首先对 E 进行 QR 分解，得到 E 的正交基 \tilde{E} 。另外，令 $U' = (U|\tilde{E})$ 。

② 令 $V' = \begin{pmatrix} V & 0 \\ 0 & I_k \end{pmatrix}$ 其中， I_k 是 $k \times k$ 单位矩阵。这样以来，我们就有

$$\begin{aligned} \Sigma' = U'^T A' V' &= \begin{pmatrix} U^T \\ \tilde{E}^T \end{pmatrix} (A|E) \begin{pmatrix} V & 0 \\ 0 & I_k \end{pmatrix} \\ &= \begin{pmatrix} U^T A V & U^T E \\ \tilde{E}^T A V & \tilde{E}^T E \end{pmatrix} = \begin{pmatrix} \Sigma & U^T E \\ 0 & \tilde{E}^T E \end{pmatrix} \end{aligned}$$

③ 对 Σ' 进行奇异值分解，即： $\Sigma' = \tilde{U}\tilde{\Sigma}\tilde{V}^T$ ，因此，数据矩阵 A' 的奇异值分解 $A' = U'\Sigma'V'^T$ 就可以被表示为：

$$A' = U'(\tilde{U}\tilde{\Sigma}\tilde{V}^T)V'^T = (U'\tilde{U})\tilde{\Sigma}(\tilde{V}^T)V'^T = \hat{U}\hat{\Sigma}\hat{V}^T$$

在获得目标表观的子空间矩阵 \hat{U} 和候选区域观测数据 o_t ，我们就可以计算该候选区域观测数据 o_t 在目标子空间 \hat{U} 中的重构误差，

$$RE = \|o_t - \hat{U}\hat{U}_t^T\|^2 \quad (4-19)$$

根据上述重构误差，基于增量子空间学习的表观模型可以定义如下，

$$p(o_t|s_t) = \exp(-RE) \quad (4-20)$$

4.3.4 实验结果

在我们的实验中，目标状态用仿射变化参数来表示： $s = (t_x, t_y, a_1, a_2, a_3, a_4)$ ，其中 $\{t_x, t_y\}$ 表示平移量， $\{a_1, a_2, a_3, a_4\}$ 是变形参数。每个状态粒子所对应的图像通过仿射变换校正为 20×20 的图像块，因此图像特征可以用一个 400 维的向量来表示。所有的实验都是在双 CPU 奔腾 IV 3.4GHz 处理器与 512M 内存的 PC 机上运行。

为了验证我们算法的有效性，我们将该算法分别与传统的粒子滤波算法和基于 Unscented 卡尔曼粒子滤波算法进行了比较。另外，为了做定量的实验比较，我们对 David 序列进行了人工标注，以此作为基准来衡量跟踪的精确度。在该实验中，我们对三种算法中的相同的参数进行了统一的设置，具体值如下： $\{N = 200, \text{var}(\epsilon) = [5^2, 5^2, 0.01^2, 0.02^2, 0.002^2, 0.001^2]\}$ ，分别对应粒子个数和系统噪声方差。如图 4-11 所示，传统的粒子滤波算法在第 31 帧偏离了真正

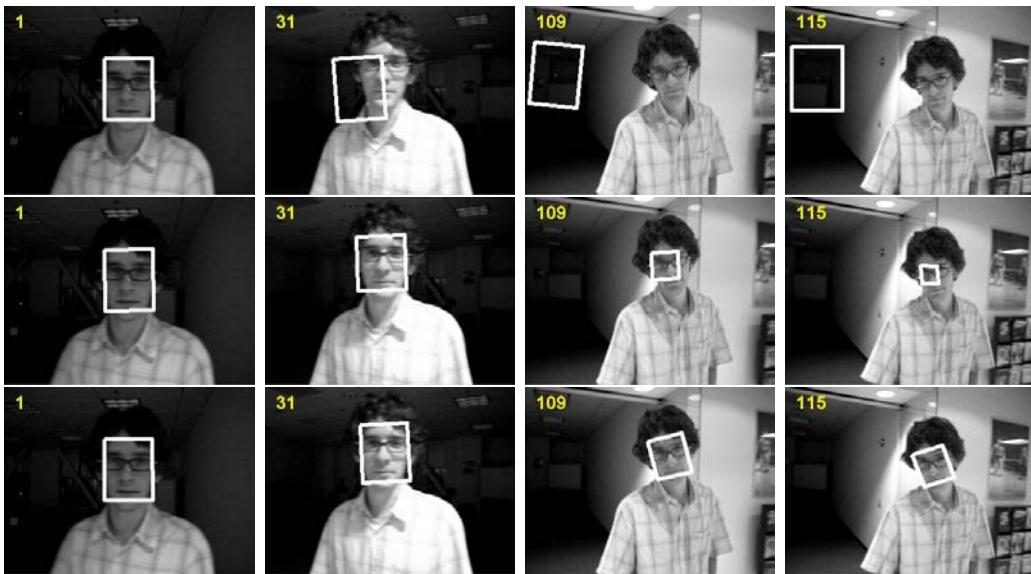


图 4-11 跟踪结果 (第一行：传统的粒子滤波算法，第二行：基于 Unscented 卡尔曼粒子滤波算法，第三行：我们的算法)

跟踪方法	成功被跟踪的帧数	均方误差
传统的粒子滤波算法	16/61	26.9481
基于 Unscented 卡尔曼粒子滤波算法	61/61	7.1875
我们的算法	61/61	3.9868

表 4-2 定量试验结果

的目标，原因是因为直接从状态转移分布采样的粒子无法捕捉到目标的运动。当目标具有快速和任意运动时，产生的粒子无法覆盖表观似然函数峰值区域，导致绝大部分粒子的权重都非常低，从而导致跟踪失败，这就是样本退化问题。更多的粒子并扩大噪声方差能在一定程度上提高粒子滤波的性能，但这种策略将引入更多噪声和计算量。图 4-11 的第二行显示了基于 Unscented 卡尔曼粒子滤波算法的结果，从中我们可以看出虽然该跟踪算法成功的捕捉到了目标的运动，但由于 Cholesky 分解有时存在病态问题，从而导致不准确的跟踪定位和目标尺度。相比之下，我们的方法实现了更精确的结果，这是因为协方差矩阵的信息被特征向量和特征值充分地描述，并且 SVD 在数值计算方面比 Cholesky 分解更加鲁棒。与此同时，我们利用人工标记的 groundtruth 对三种算法进行定量评价，并记录以下两个指标：成功被跟踪的帧数，跟踪定位与 groundtruth 之间的均方误差，结果见表 4-2，很显然我们的算法取得了最佳的跟踪结果。

第五章 基于序列粒子群优化的跟踪方法

本章将在上一章的基础之上，更加深入的讨论了粒子滤波算法的样本退化问题以及其理论原因。针对这一问题，本文将跟踪问题用鸟群捕食过程来模拟，提出了一种基于序列粒子群优化的跟踪算法框架。该算法首先通过初始化一系列随机粒子样本，然后对这些粒子集进行时序上的传播以及优化迭代，最后对进化的粒子进行收敛性判断。我们从理论上证明了，序列粒子群优化框架实质上是一个基于分步重要性采样的粒子滤波过程，该分步采样过程包含两个步骤：第一步是粗采样，以状态转移分步作为重要性分布，直接对其进行采样；第二步是细采样，通过粒子群优化迭代发生进化，该过程实际上是一个隐采样过程。通过这样的两步重要性采样策略，我们的采样结果逼近了直接采样于最佳重要性分布的结果。我们还将序列粒子群优化的概念扩展到多种群上，将多目标跟踪中目标物体的交互和遮挡用种群间的竞争和排斥来模拟，实现了有效的多目标跟踪。与其他两种多目标跟踪的算法相比，我们的算法在 PETS2004 数据库上获得了最佳的实验结果。

5.1 基于序列粒子群优化算法的单目标跟踪

在介绍序列粒子群优化算法之前，我们首先来介绍一下传统的粒子群优化算法的来源以及它的具体形式 [100]。

5.1.1 粒子群优化算法

5.1.1.1 来源和背景

为了说明粒子群优化算法的发展和形成背景，我们首先介绍一下早期的简单模型，即 Boid (Bird-oid) 模型。这个模型是为了模拟鸟群的行为而设计的，它也是粒子群优化算法的直接来源。

一个最简单的 Boid 模型是这样的：每一个鸟的个体用直角坐标系上的点表示，随机地给它们赋一个初始速度和初始位置，程序运行的每一步都按照“最

近邻速度匹配”规则，使某个个体的最近邻点的速度变得与它一样，如此迭代计算下去，很快就会使得所有点的速度变得一样。因为这个模拟太简单而且远离真实情况，于是在速度项中增加了一个随机变量，即在迭代的每一步，除了满足“最近邻速度匹配”之外，每一步速度还要添加一个随机变化的量，这样使得整个模拟看起来更为真实。

基于上述模型，Heppner 设计了一个更为复杂的“谷地模型”来模拟鸟群的觅食行为 [101]。假设在平面上存在一个“谷地”，即食物所在地，鸟群开始时随机地分散在平面上，为了寻觅食物所在地，它们按照如下规则运动：

首先假设谷地的位置坐标为 (x_f, y_f) ，单个鸟的位置和速度坐标分别为 (x, y) 和 (v_x, v_y) ，用当前位置到谷地的距离：

$$s = \sqrt{(x - x_f)^2 + (y - y_f)^2} \quad (5-1)$$

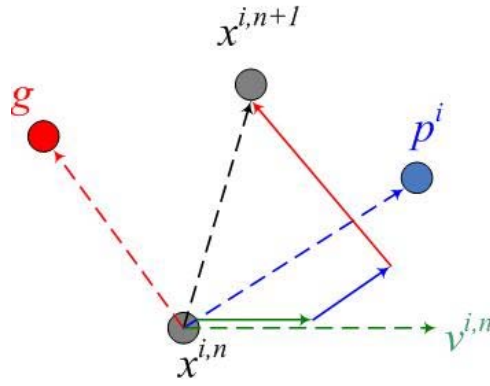
来衡量当前位置和速度的“好坏程度”，离谷地的距离越近，则越“好”，反之越“坏”。假设每一个鸟具有记忆能力，能够记住曾经达到的最好位置，记作 (p_x, p_y) ，并记 a 为系统规定的速度调节常数， $rand$ 为一个 $[0,1]$ 间的随机数，设定速度项按照下述规则变化：

$$v_x = \begin{cases} v_x - rand * a, & x > p_x \\ v_x + rand * a, & x < p_x \end{cases} \quad (5-2)$$

另外假设群体之间可以以某种方式通讯，每个个体能够知道并记住到当前为止整个群体的最好位置，记为 (g_x, g_y) ，记 b 为系统规定的速度调节常数，则速度项在经过以上调整后，还必须按照下述规则变化：

$$v_x = \begin{cases} v_x - rand * b, & x > g_x \\ v_x + rand * b, & x < g_x \end{cases} \quad (5-3)$$

在计算机上模拟的结果显示：当 a/b 较大时，所有的个体很快地聚集到“谷地”上；反之，粒子缓慢地摇摆着聚集到“谷地”的四周。通过这个简单的模拟，发现群体能很快地找到一个简单函数（5-1）的最优点。


 图 5-1 粒子 i 第 n 次迭代的示意图

5.1.1.2 粒子群优化算法

受上述模型启发，Kennedy 和 Eberhart 设计出了一种演化优化算法 [102]，并通过不断的试验，最后得到算法的固定形式，下面我们将给出粒子群优化算法（PSO, particle swarm optimization）的完整的形式化表述。为了与后续部分的算法一致，我们对粒子群优化算法中的符号进行重新定义。设粒子群体规模为 N ，其中每个粒子为 d 维空间的状态向量，且初始状态定义为 $x^{i,0}$ ，每个向量有一个对应的适应值 $f(x^{i,0})$ 和速度向量 $v^{i,0}$ 。另外，粒子的个体最优状态（即该粒子经历过的最优状态）记为 p^i ，群体最优状态（即该粒子群中任意个体经历过的最优状态）记为 g 。不失一般性，以最大化适应值函数为例，在最初版本的粒子群优化算法中，粒子的状态速度和状态的第 n 次迭代公式分别为（见图 5-1），

$$v^{i,n+1} = v^{i,n} + \varphi_1 u_1 (p^i - x^{i,n}) + \varphi_2 u_2 (g - x^{i,n}) \quad (5-4)$$

$$x^{i,n+1} = x^{i,n} + v^{i,n+1} \quad (5-5)$$

其中 φ_1, φ_2 为加速度常数， u_1, u_2 为 $[0,1]$ 间的均匀随机数。由于初始版本在优化问题中应用时效果并不太好，所以初始算法提出不久之后就出现了一种改进

算法 [103], 在速度迭代公式中引入了惯性权重 w , 速度迭代公式变为:

$$v^{i,n+1} = w^n v^{i,n} + \varphi_1 u_1(p^i - x^{i,n}) + \varphi_2 u_2(g - x^{i,n}) \quad (5-6)$$

$$x^{i,n+1} = x^{i,n} + v^{i,n+1} \quad (5-7)$$

其中 w^n 为惯性权重因子, 一般情况下它是迭代次数 n 的单调递减函数。例如, 给定一个最大权重 w_{max} 和最小权重 w_{min} , w^n 可以通过下列方式计算得到,

$$w^{n+1} = w^n - dw, \quad dw = (w_{max} - w_{min})/T \quad (5-8)$$

虽然该改进算法与初始版本相比复杂程度并没有太大的增加, 但是性能却得到了质的提升, 因而被广泛使用。一般的, 将该改进算法称为标准粒子群优化算法, 而将初始版本的算法称为原始粒子群优化算法。

通过分析 PSO 算法的收敛行为, Clerc 提出了一种带收缩因子的 PSO 算法 [104], 该收缩因子保证了收敛性并提高了收敛速度。此时的速度迭代公式为,

$$v^{i,n+1} = \mathcal{X}(v^{i,n} + \varphi_1 u_1(p^i - x^{i,n}) + \varphi_2 u_2(g - x^{i,n})) \quad (5-9)$$

$$x^{i,n+1} = x^{i,n} + v^{i,n+1} \quad (5-10)$$

其中 \mathcal{X} 为收缩参数, 它的目的是使得粒子的速度满足: $\|v^{i,n}\| \leq v^{max}$ 。从本质上来说, 迭代公式 (5-6) 和 (5-9) 并无本质区别, 只要适当选取参数, 二者完全相同。

在完成一次迭代之后, 我们对每个粒子用预先定义的适应值函数 $f(\cdot)$ 进行评价, 基于评价的结果我们对粒子的个体最优状态和群体最优状态进行更新,

$$p^i = \begin{cases} x^{i,n+1}, & \text{if } f(x^{i,n+1}) < f(p^i) \\ p^i, & \text{else} \end{cases} \quad (5-11)$$

$$g = \arg \max_{p^i} f(p^i) \quad (5-12)$$

所有的粒子通过以上的迭代方式寻找适应值函数 $f(\cdot)$ 的最优值直到收敛。

5.1.1.3 算法分析

粒子迭代的示意图见图 5-1，在每一次粒子迭代中，粒子的状态主要由三部分决定：惯性速度，个体最优状态，群体最优状态。从社会学的角度来看速度迭代公式，其中第一部分为粒子先前速度的影响，表示粒子对当前自身运动状态的信任，依据自身的速度进行惯性运动；第二部分取决于粒子当前位置与自身最优位置之间的距离，为“认知（Cognition）”部分，表示粒子本身的思考，即粒子的运动来源于自己经验的部分；第三部分取决于粒子当前位置与群体中全局最优位置之间的距离，为“社会（Social）”部分，表示粒子间的信息共享与相互合作，即粒子的运动来源于群体中其他粒子经验的部分，它通过认知模拟了较好同伴的运动。因此，粒子群优化算法的好处是，一方面，单个粒子计算和评价简单，另一方面，具备群体粒子共享信息的优越性能。

5.1.2 序列粒子群优化算法

5.1.2.1 动机

这一部分，我们从随机优化的角度去来分析为什么粒子群优化算法能应用到目标跟踪任务中，并且取得良好的效果。

从本质上讲，视频目标跟踪等价于在连续的图像帧间搜索一特定感兴趣的区域。具体的，让我们将从下面角度去审视一下跟踪问题：在这里，我们采用鸟群觅食过程来模拟跟踪问题，其中目标用食物来模拟，样本粒子用鸟群模拟。整个跟踪过程可以表示为，假定图像空间（状态空间）中随机产生一组粒子（鸟群），没有一个粒子（鸟）知道目标（食物）在哪里，但是每个粒子（鸟）可以通过自己的适应值来判断目标（食物）距离自己多远，在这种情况下，如何有效地利用全部粒子（鸟群）所获得的信息，并用于快速而有效的找

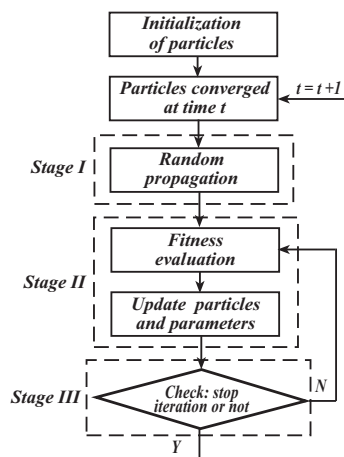


图 5-2 序列 PSO 算法流程

到目标（食物）——即是跟踪所面临的问题。基于群体智能的演化计算技术——粒子群优化算法，为解决上述问题提供了一种可行的思路。

然而，在跟踪应用中，数据通常是一个时间序列，因此跟踪任务基本上是一个动态优化问题。与传统优化问题不同的是，在这种情况下，适应值函数同时是目标状态和时间的函数。为了有效地解决上述动态优化的问题，我们需要回答这些问题：1) 如何利用两帧间目标状态的时序连续性，b) 如何保持优化过程中粒子的多样性。

5.1.2.2 序列 PSO 框架

根据上述讨论，我们提出了一种基于序列粒子群优化的框架。如图 5-2 所示，该框架主要包含三个部分：时序上的随机传播，自适应的粒子群迭代，收敛条件。

随机传播： 当 PSO 应用于时序优化问题时，主要的问题在于，由于上一个时刻优化算法的收敛，导致粒子群的多样性丢失。因此，我们需要对上一时刻的粒子进行随机传播，来增强粒子的多样性。

设计一种有效的随机传播算法需要有目标运动的先验知识，但是对目标运动的预测是非常困难的，特别是对于移动摄像机条件下拍摄的视频，观测到的

运动是摄像机和目标的联合运动，因此我们无法分辨哪个是摄像机运动哪个是目标运动。在本节中，我们采用简单的高斯分布对粒子集合进行随机传播，而该高斯分布的均值和方差定义如下：

给定 t 时刻粒子群集合的个体最优状态 $\{p_t^i\}_{i=1}^N$ ，那么，从 t 时刻向 $t+1$ 时刻的随机传播策略如下：

$$x_{t+1}^{i,0} \sim \mathcal{N}(p_t^i, \Sigma) \quad (5-13)$$

其中 p_t^i 为高斯分布的均值， Σ 为高斯分布的协方差矩阵，它是一个对角矩阵，其对角元素由预测速度 v_t^{pred} 所决定，

$$v_t^{pred} = g_{t-1} - g_{t-2} \quad (5-14)$$

与此同时，粒子群的初始速度 $v_{t+1}^{i,0}$ 也将从以下的均匀分布 $\mathcal{U}(0, v_t^{pred})$ 中采样得到。

在我们的随机传播过程中，不需要像粒子滤波一样采用“重采样”过程来避免样本退化，这是由于个体最优状态 p_t^i 提供了一个非常紧致和有效的传播样本集，因而不需要“重采样”来去掉不好的粒子，复制好的粒子。另外，由于个体最优状态 p_t^i 集并不是像重采样那样单一的复制好的粒子，因此在保留粒子多样性方面具有更好的性能。尽管我们的随机传播模型非常简单，但它足以为后续过程提供一个不错的初始值。

自适应 PSO： 在 §5.1.1.2 的讨论中，我们发现传统的粒子群优化算法需要人为的设置多个参数：加速度参数 φ_1, φ_2 ，最大限制速度 v_t^{max} 或者惯性权重因子 w^n 。如果缺乏有效合理的机制来控制这些参数的话，粒子群优化算法将存在发散的危险。为了克服这一不足之处，我们提出了参数自适应的粒子群优化算法，其中加速度参数 φ_1, φ_2 可以设置如下，

$$\varphi_1 = 2f(p^i)/(f(p^i) + f(g)) \quad (5-15)$$

$$\varphi_2 = 2f(g)/(f(p^i) + f(g)) \quad (5-16)$$

相对于文献 [102] 将加速度参数设置为常数 2，我们的方法更为合理。公式 (5-15) 和 (5-16) 表明，对“认知”部分或者“社会”部分的信赖程度应该由他们的适应值决定，而不是一视同仁。另外，在跟踪应用中，最大速度限制提供了合理的约束，以防止粒子随意的发散运动。在传统算法中，最大速度通常设置为一个预定义的常量。然而，这在跟踪中往往是不合理的，因为目标物体的运动是随机的，运动幅度往往也是不一样的。因此，在预测速度 v_t^{pred} 的基础上，我们提出了下列选择最大速度 v_t^{max} 的方法，

$$v_t^{max} = 1.2 * v_t^{pred} \quad (5-17)$$

基于上式定义的最大速度 v_t^{max} ，则 \mathcal{X} 可以定义如下，

$$\mathcal{X} = \begin{cases} \|v_t^{max}\|/\|v_t^{i,n+1}\|, & \text{if } \|v_t^{i,n+1}\| > \|v_t^{max}\| \\ 1, & \text{else} \end{cases} \quad (5-18)$$

这样，最大速度 v_t^{max} 就可以根据预测速度 v_t^{pred} 改变而改变，从而在一定程度上可以适应目标运动的变化，并且它是 v_t^{pred} 的 1.2 倍，因此可以提供给粒子一定的加速度。

收敛准则： 跟踪的目的是尽快的找到待跟踪的目标，因此它的任务与传统优化算法的任务并不完全一样，它没有必要让所有粒子都收敛到 groundtruth 上去。因此，在本文中，收敛准则定义如下：

$f(g_t) > Th$ ，其中 Th 为预先定义的阈值，并且个体最优状态 $\{p_t^i\}_{i=1}^N$ 都落在群体最优状态 g_t 的邻域中，如图 5-3 所示，或者到达最大迭代次数。

根据这一收敛准则，目标物体可以被有效地搜索和定位，并且收敛后的体最优状态 $\{p_t^i\}_{i=1}^N$ 为下一个时刻的优化提供了很好的初始粒子集，从而将跟踪过程中的时序信息有效的加入到粒子群优化的框架中。另外，个体最优状态集 $\{p_t^i\}_{i=1}^N$ 还保留了样本的多样性，同时也避免了样本退化问题。

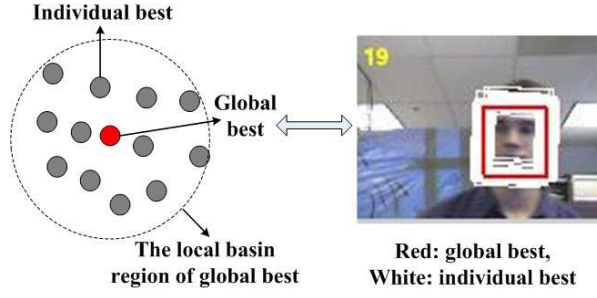


图 5-3 收敛准则

5.1.3 算法分析

在这一部分，我们从贝叶斯推理的角度对序列粒子群算法进行分析，发现序列粒子群优化算法从本质上来说就是一个基于多层重要性采样的粒子滤波算法。多层的采样过程将最新的图像观测加入到状态转移分布中，从而有效的逼近了最佳的重要性分布 $p(x_t|x_{t-1}^i, o_t)$ [105]。下面我们首先简要回顾一下标准粒子滤波算法以及它所面临的问题，然后在粒子滤波框架下分析为什么序列粒子群优化算法能解决这些问题。

5.1.3.1 粒子滤波

粒子滤波将跟踪过程可以看成是一个贝叶斯推理过程，通过一系列观测 $o_{1:t}$ 来推理出目标的运动状态 x_t ，即估计目标状态的后验概率分布 $p(x_t|o_{1:t})$ 。在贝叶斯推理的框架中，通常采用状态空间对当前跟踪系统的运动进行模型化，

$$\text{state transition model} : x_t = f_t(x_{t-1}, \epsilon_t) \leftrightarrow p(x_t|x_{t-1}) \quad (5-19)$$

$$\text{observation model} : o_t = h_t(x_t, \nu_t) \leftrightarrow p(o_t|x_t) \quad (5-20)$$

其中 x_t, o_t 分别代表系统（目标）状态和观测， ϵ_t, ν_t 分别表示系统噪声和观测噪声， $f_t(.,.)$ 和 $h_t(.,.)$ 分别表示状态转移模型和观测模型。这两个模型分别由状态转移分布 $p(x_t|x_{t-1})$ 和观测分布 $p(o_t|x_t)$ 决定。

在贝叶斯推理过程中，后验分布可以表示为：

$$p(x_t|o_{1:t}) \propto p(o_t|x_t)p(x_t|o_{1:t-1}) \quad (5-21)$$

其中 $p(x_t|o_{1:t-1})$ 表示对目标状态的预测，它可以通过将上一时刻的后验分布在时间轴上传播得到，

$$p(x_t|x_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|o_{1:t-1})dx_{t-1} \quad (5-22)$$

当状态转移模型和观测模型为非线性和非高斯时，上述积分形式没有解析解，因此我们需要采用数值逼近的方法来求解。粒子滤波是一种被广泛应用的基于重要性采样的后验分布逼近算法。它的主要思想是利用一堆采样点 $\{x_t^i\}_{i=1}^N$ 来逼近目标状态的后验分布，即 $p(x_t|o_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta(x_t - x_t^i)$ ，其中 $\delta(\cdot)$ 为冲击函数。但是在实际应用中，我们无法直接对后验分布进行采样，因此粒子滤波通过重要性采样来解决上述问题。首先定义一个方便采样的分布，称作重要性分布 $q(\cdot)$ ，然后从重要性分布中采样 N 个样本，并对每个样本进行评价，从而获得粒子样本的权重，

$$w_t^i \propto \frac{p(o_t|x_t^i)p(x_t^i|x_{t-1}^i)}{q(x_t|x_{t-1}^i, o_{1:t})} \quad (5-23)$$

最后后验分布可以表示为

$$p(x_t|o_{1:t}) = \sum_{i=1}^N w_t^i \delta(x_t - x_t^i) \quad (5-24)$$

在跟踪应用中，最大后验估计或者样本均值都可以作为跟踪结果的输出。

在粒子滤波算法中，重要性分布 $q(\cdot)$ 的选择是非常关键的，因为它关系到能否从后验分布的重要部分获得粒子的问题。在跟踪过程中，通常选择 $p(x_t|x_{t-1})$ 作为重要新分布函数，将 $q(\cdot) = p(x_t|x_{t-1})$ 代入到公式 (5-4) 中，样本粒子的权重可简化为 $w_t^i = p(o_t|x_t^i)$ 。当状态转移分布位于观测分布的尾部的时候，如图 5-4 左边所示，所获得的粒子权重绝大多数都非常小，导致样本退化问题。在文献 [105] 中，Doucet 等证明了在最小化重要性权重方差的意义下， $p(x_t|x_{t-1}^i, o_t)$ 为最佳的重要性采样分布，但在实际情况中，我们无法获得该分布的解析形式，因此，解决因重要性采样引起的样本退化问题的关键在于，如何将当前帧的观测信息 o_t 融入到状态转移分布中，从而逼近最佳的采样分布。

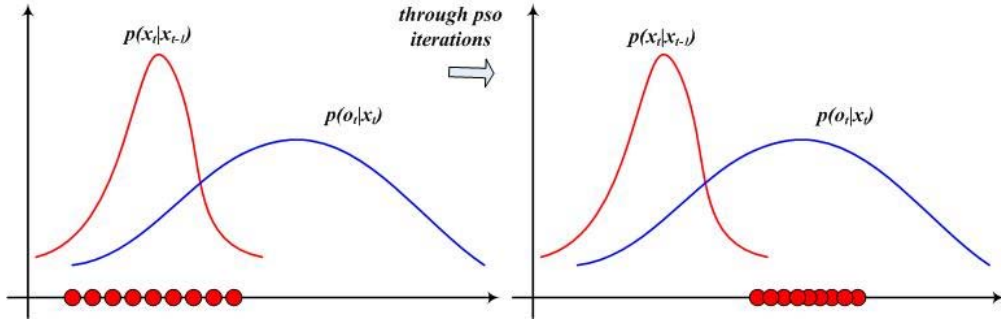


图 5-4 重要性采样的例子（左边：直接从 $p(x_t|x_{t-1})$ 采样；右边：经过 PSO 迭代后）

5.1.3.2 多层采样过程

根据 §5.1.2 的讨论以及贝叶斯推理过程，我们可以看到，序列 PSO 算法从本质上来看是由 PSO 迭代过程和粒子滤波过程来组成，但传统的粒子滤波算法不一样的是：传统的粒子滤波算法直接从状态转移分布进行采样得到粒子，而序列 PSO 算法则需要对采样出来的粒子进行进化。从贝叶斯推理的角度来看，序列 PSO 算法所采用的策略可以看成多层采样过程，因而逼近直接从最优分布 $p(x_t|x_{t-1}^i, o_t)$ 采样的结果。

具体来看序列 PSO 算法中的粒子进化过程，首先是粗采样部分：通过对状态转移分布采样得到初始粒子，

$$x_{t+1}^{i,0} \sim \mathcal{N}(p_t^i, \Sigma) \quad (5-25)$$

然后细采样部分：利用 PSO 算法对粒子进行迭代，根据公式 (5-9) 和 (5-10)，我们发现其实每一次迭代过程都可以看成以下的采样形式，

$$x_t^{i,n+1} \sim \mathcal{U}(0, U) \quad (5-26)$$

$$U = \min(v^{i,n} + \varphi_1(p^i - x^{i,n}) + \varphi_2(g - x^{i,n}), v^{max})$$

其中 $\mathcal{U}(\cdot, \cdot)$ 为均匀分布，尽管采样分布为简单的均匀分布，但是每一次采样均能把当前帧的观测通过个体最优状态和群体最优状态融入到状态转移分布产生的粒子中去。通过不停的迭代，即采样，我们可以将当前帧最有用的观测

信息 o_t 融入到 $p(x_t|x_{t-1})$ 所产生的粒子中去，从而有效的逼近了最佳采样分布 $p(x_t|x_{t-1}^i, o_t)$ 所产生的结果。具体的采样过程见算法 2。

Algorithm 2 多层采样过程

1. **输入：** 由状态转移分布产生的初始粒子 $X_t = \{x_t^{i,0}\}_{i=1}^N$ 。

2. **for** $n = 0 : T$ **do**

3. 采样过程，

$$x_t^{i,n+1} \sim \mathcal{U}(0, U)$$

4. 通过对每个粒子所对应的观测进行评价，

$$f(x_t^{i,n+1}) = p(o_t^{i,n+1}|x_t^{i,n+1})$$

5. 更新采样分布的参数

$$p_t^i = \begin{cases} x_t^{i,n+1}, & \text{if } f(x_t^{i,n+1}) > f(p_t^i) \\ p_t^i, & \text{else} \end{cases}$$

$$g_t = \arg \max_{p_t^i} f(p_t^i)$$

6. 收敛标准: if satisfied, **break**;

7. **end for**

8. **输出：** $X_t = \{x_t^{i,n}\}_{i=1}^N$

如图 5-4 所示，当状态转移分布位于观测分布的尾部的时候，所获得的粒子权重绝大多数都非常小，导致样本退化问题。相反的，通过多层采样的方法，粒子向观测似然重要的区域移动，最终停留在它的峰值附近，从而有效的改善了采样的结果。

5.1.4 跟踪算法

在该跟踪算法中，我们采用第四章核贝叶斯跟踪算法中的目标状态参数表示和基于空间约束混合高斯的表观模型，具体算法见算法 3，

5.1.5 实验结果

在实验中，我们采用与第四章核贝叶斯相同的参数设置以及表观模型，下面我们将给出具体的实验结果。

Algorithm 3 基于序列粒子群优化的跟踪算法

输入: 给定 t 时刻的个体最优状态集 $\{p_t^i\}_{i=1}^N$;

1. 利用状态转移分布对个体最优状态集进行随机传播, 以增强粒子的多样性,

$$x_{t+1}^{i,0} \sim \mathcal{N}(p_t^i, \Sigma)$$

其中 Σ 为状态参数的协方差矩阵, 它的对角元分别为仿射运动参数的方差 $\sigma_x^2, \sigma_y^2, \sigma_\theta^2, \sigma_s^2, \sigma_\alpha^2, \sigma_\phi^2$ 。

2. 对粒子群采用 PSO 迭代进行进化,

$$v_t^{i,n+1} = \mathcal{X}(v_t^{i,n} + \varphi_1 u_1(p_t^i - x_t^{i,n}) + \varphi_2 u_2(g_t - x_t^{i,n}))$$

$$x_t^{i,n+1} = x_t^{i,n} + v_t^{i,n+1}$$

3. 利用基于空间约束混合高斯的表现模型对每个粒子的适应值进行评价,

$$f(x_{t+1}^{i,n+1}) = p(y_{t+1}^{i,n+1} | x_{t+1}^{i,n+1}), i = 1 \cdots N, n = 0 \cdots T$$

4. 根据适应值评价结果更新个体最优状态 $\{p_{t+1}^i\}_{i=1}^N$ 和群体最优状态 g_{t+1} ,

$$p^i = \begin{cases} x^{i,n+1}, & \text{if } f(x^{i,n+1}) > f(p^i) \\ p^i, & \text{else} \end{cases}, g = \arg \max_{p^i} f(p^i)$$

5. 收敛判断: 若满足收敛条件, 则继续; 否则的话跳到第 2 步。

输出: 群体最优状态 g_{t+1} ;

5.1.5.1 序列 PSO vs 粒子滤波和基于 Unscented 卡尔曼的粒子滤波

首先, 我们在一段标注过 groundtruth 的视频序列上分别对序列粒子群优化算法 (SPSO, sequential PSO), 粒子滤波 (PF) 以及基于 Unscented 卡尔曼的粒子滤波 (UPF) 进行测试。然后, 对所得实验结果的原因进行讨论和分析。

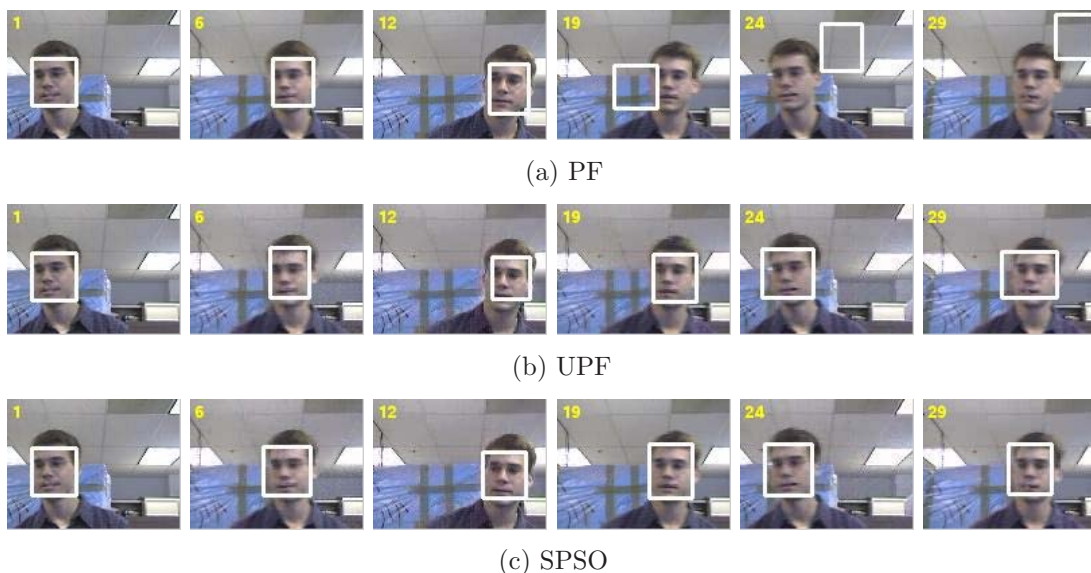


图 5-5 跟踪结果

这个部分中的测试视频序列¹包含快速且无规则的运动，并且还伴随有摄像机的运动，尽管该视频内容相对简单，但是它包含的快速运动对测试算法采样性能非常有效。在我们的算法实现过程中，PF 和 UPF 中的粒子个数，状态转移分布的协方差矩阵分别设置如下： $\{N = 600, \Sigma = \text{diag}(8^2, 8^2, 0.02^2, 0.02^2, 0.002^2, 0.002^2)\}$ 。为了将算法在同一条件下进行比较，在 SPSO 中，我们采用同样的协方差矩阵，另外，粒子个数设为 60 个，最大迭代次数为 10 次。从图 5-5 (a) 可以看出，由于 PF 无法捕捉到目标的快速运动，基于 PF 的跟踪器在第 19 帧图像中跟踪失败，增大状态转移分布的协方差矩阵的值并且产生更多粒子可能在一定程度上改善跟踪结果。但是这样会引入很大的计算量，而且当目标状态维数增加时，可能陷入维数灾难问题。图 5-5 (b) 显示基于 UPF 的跟踪器在该视频序列上的跟踪结果，从该行结果我们看到 UPF 可以捕捉到目标的运动，但是定位精度不够准确。相反的，如图 5-5 (c) 所示，我们的方法从未丢失待跟踪的目标，并且取得了很好的定位精度。另外，我们还对三种算法的结果在以下两个方面做了定量分析：成功被跟踪的帧数，跟踪定位与 groundtruth 之间的均方误差，具体结果见表 5-1，很

¹这个序列可以从下列 URL 上获得 <http://vision.stanford.edu/birch/headtracker/seq/>。

明显可以看出 PF 在第 19 帧跟踪失败，而 UPF 和 SPSO 在整个视频序列中都对目标进行成功地跟踪，但是 SPSO 的跟踪定位精度要好于 UPF。

通过对上述实验结果进行讨论和分析，我们发现 PF 算法的不理想结果是由于样本粒子的退化造成的。为了能说明这一结论，我们来看 PF 在第 19 帧的具体跟踪过程，如图 5-6 所示，利用状态转移分布采样得到的样本粒子无法覆盖到目标真正的运动，即大部分样本粒子都无法到达观测似然的重要区域，使得粒子的权重很小，从而导致跟踪失败。对于 UPF 而言，它能在一定程度上改善跟踪性能，原因它是将基于 Unscented 变换的卡尔曼滤波引入到粒子滤波器中，为粒子滤波器提供重要性采样分布。该算法首先通过 Unscented 变换产生一系列 sigma 样本，然后对这些样本用卡尔曼进行滤波，滤波后得到的样本均值和方差为重要性分布提供了参数。然而，Unscented 变换只能估计到非高斯数据的二阶统计量，因此所提供的重要性分布并不能与实际运动完全符合，从而在一定程度上降低了定位精度。另外，产生 sigma 样本和对其进行滤波都非常耗时。而在基于 SPSO 的跟踪框架中，我们通过样本粒子中的局部最优和全局最优将当前帧的观测信息引入到状态转移分布中，从而改善了采样结果，避免了样本退化问题。图 5-6 的第二行显示了 SPSO 框架的采样过程，从中我们可以明显的看出 PSO 迭代可以将粒子吸引到目标位置上去，并且精确地围绕在目标周围。

5.1.5.2 其他场景的测试结果

为了进一步测试 SPSO 算法的性能，我们将该算法在另外三组不同的视频

Tracking Framework	Frames Tracked	MSE of Position (by pixels)
<i>PF</i>	18/31	17.069
<i>UPF</i>	31/31	6.975
<i>SPSO</i>	31/31	4.172

表 5-1 定量比较实验

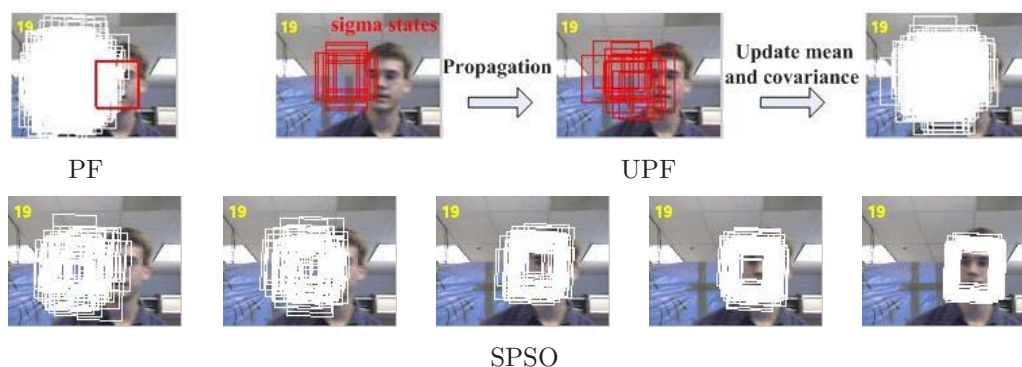


图 5-6 第 19 帧图像的跟踪过程

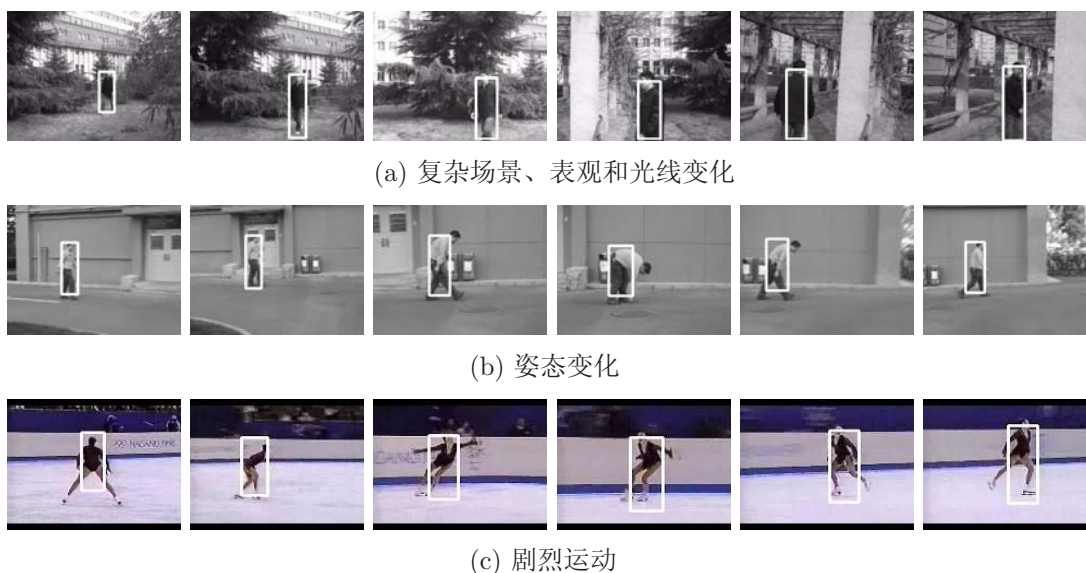


图 5-7 更多的实验结果

序列上进行测试。第一个视频序列为一个男人在大草坪上步行，该视频序列包含混乱的背景以及表观和光照的变化。第二个视频序列也是个关于行人行走的例子，这个序列包含人体姿态的巨大变化（弯下腰到地面，然后再站起来）。这两个视频序列都是在室外场景下，利用移动摄像机拍摄的。第三个视频序列是一个花样滑冰比赛，其中包含花样滑冰运动员和摄像机的剧烈运动。

从图 5-7（a）的实验结果可以看出我们表观模型的在线更新策略可以很好的解决第一个视频序列中表观和光线的变化，并且 SPSO 框架对复杂混乱背景下的跟踪提供了一个有效的解决方案，因为该算法具备更好的能力来寻找全局最优。图 5-7（b）为我们算法在第二个视频序列上的跟踪结果，该结果表明了我们的算法在跟踪姿态变化方面的有效性。最后图 5-7（c）给出在第三个视频

序列上的测试结果，该序列虽然非常具有挑战性，但是由于我们的算法具有非常好的采样性能，因此能够捕捉到花样滑冰选手的快速运动。

5.2 基于多种群粒子群优化算法的多目标跟踪

5.2.1 动机

5.2.1.1 从生物群体智能角度来看单目标跟踪

在介绍算法动机前，我们首先做以下三个模拟假设：1) 将目标真实状态及在图像中所覆盖的区域用食物所在地来模拟；2) 粒子集合用鸟群来模拟；3) 候选图像的观测函数用鸟的适应值，即该鸟寻找食物的能力来模拟。基于上述三个假设，我们可以利用鸟群觅食过程来模拟跟踪问题，整个跟踪过程可以表示为，假定在图像空间（状态空间）中随机产生一组粒子（鸟群），没有一个粒子（鸟）知道目标（食物）在哪里，但是每个粒子（鸟）可以通过自己的适应值来判断目标（食物）距离自己多远，在这种情况下，如何有效地利用全部粒子（鸟群）所获得的信息，并用于快速而有效的找到目标（食物）——即是跟踪所面临的问题。基于群体智能的演化计算技术——粒子群优化算法，为解决上述问题提供了一种可行的思路。

5.2.1.2 如何扩展到多目标跟踪

当多个目标相互分开且未发生交互或者遮挡，上述单目标的跟踪框架很容易就可以扩展到多目标跟踪的任务：对每个目标都分配一个跟踪器，然后利用序列粒子群优化算法框架分别进行跟踪。但是当多个目标相互靠近并发生遮挡时，基于序列粒子群优化的单目标跟踪算法将会失败。如 §5.2.1.1 所述，我们将目标真实状态以及它所覆盖的区域用食物所在地来模拟，因此当两个目标间发生遮挡，它们所在区域发生重叠，意味着该食物所在地被两群鸟共同争夺。这样的话，从生物角度来说，不可避免地会发生种群间的竞争和排斥，并且更强的种群将有更大的概率获得竞争的胜利。因此，多目标跟踪的关键就在于如何将生物种群间的竞争和排斥用图像信息来表示和模拟。

5.2.2 多目标跟踪算法

在该算法中，我们采用与 4.3 中相同的目标状态表示方法以及目标表观模型。

如图 5-8 所示，当目标间发生遮挡时，不同目标所对应的图像区域发生了重叠，使得图像区域信息变得不再可靠。为了有效地克服遮挡所带来的困难，我们提出基于多种群粒子群优化算法来进行多目标跟踪，并且将目标间的遮挡用种群间的竞争和排斥来模拟。下面我们将从以下四个部分对算法进行具体介绍：1) 问题描述；2) 竞争与排斥模型；3) 基于高斯退火的粒子群优化；4) 表观模型的选择性更新。最后对该算法进行总结。

5.2.2.1 问题描述

首先我们对该算法中所采用的符号进行定义，目标状态用 $\mathcal{X} = \{x_{t,k}^{i,n}, i = 1, \dots, N, k = 1, \dots, M\}$ 来表示，所对应的图像观测用 $\mathcal{O} = \{o_{t,k}^{i,n}, i = 1, \dots, N, k = 1, \dots, M\}$ 来表示，其中 N 为粒子个数， M 为目标的个数。然后多目标跟踪的问题通过最大化联合似然分布函数来获得，具体形式可以描述如下，

$$\mathcal{X}^* = \arg \max_{\mathcal{X}} p(\mathcal{O}|\mathcal{X}) \quad (5-27)$$

当没有发生遮挡时，上述优化问题可以简化为最大化目标个体似然分布函数（为了简化符号，我们这里不写上标 i, n ），

$$x_{t,k}^* = \arg \max_{x_{t,k}} p(o_{t,k}|x_{t,k}), k = 1, \dots, M \quad (5-28)$$

当目标间发生遮挡时，我们将目标所对应的图像观测区域分为两个部分：1) 非重叠部分；2) 重叠部分。如图 5-8 所示，假设遮挡发生于目标 k_1, k_2 间，则对于这两个目标的跟踪问题可以描述为

$$x_{t,k_1}^* = \arg \max_{x_{t,k_1}} p(\tilde{o}_{t,k_1}|x_{t,k_1})p(\hat{o}_{t,k_1}|x_{t,k_1}, x_{t,k_2}) \quad (5-29)$$

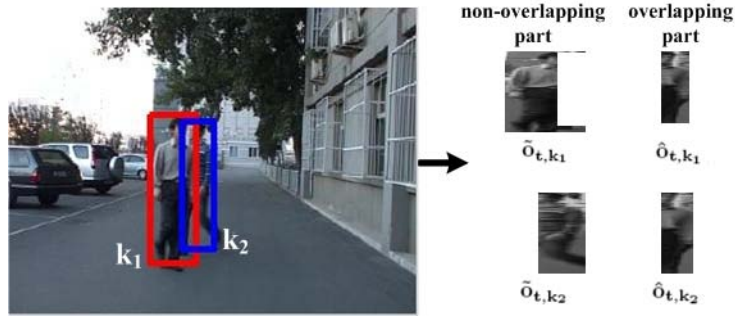


图 5-8 两个目标遮挡的例子

$$x_{t,k_2}^* = \arg \max_{x_{t,k_2}} p(\tilde{o}_{t,k_2} | x_{t,k_2}) p(\hat{o}_{t,k_2} | x_{t,k_2}, x_{t,k_1}) \quad (5-30)$$

其中 $p(\hat{o}_{t,k_1} | x_{t,k_1}, x_{t,k_2})$ 和 $p(\hat{o}_{t,k_2} | x_{t,k_2}, x_{t,k_1})$ 分别为对应目标在重叠区域上的交互似然。公式 (5-29) 和 (5-30) 相互进行交叉迭代直到收敛。三个目标或更多目标间发生遮挡也可以进行类似的描述。

5.2.2.2 竞争和排斥模型

竞争模型: 如图 5-8 所示, 当目标间发生遮挡时, 在这种情况下, 两个目标所对应的种群将对重叠区域进行竞争。于是我们面临的问题是如何有效的对这种竞争进行建模。

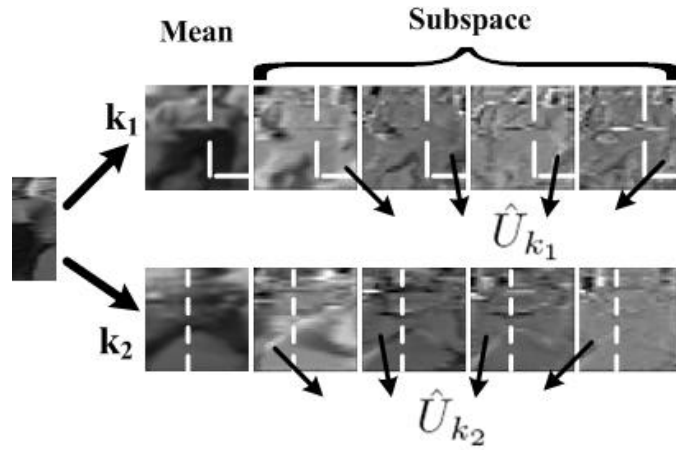


图 5-9 将重叠区域的图像观测投影到目标对应部分的子空间上

为了有效的回答上述问题，我们需要将竞争现象与图像信息进行关联，在提出我们的模型前，我们首先来讨论两个相关的工作 [106,107]。尽管他们的算法都是在观测层面上对目标间的交互和遮挡进行建模，但是具体模型与建模的方式却完全不同。在文献 [106] 中，对于发生遮挡的两个目标，它们的遮挡建模是基于整个目标所覆盖的区域上，而不是目标之间的重叠区域上。这在实际跟踪中往往是不太合理的，因为种群间的竞争一般只发生在重叠区域，如果对整个目标区域进行建模的话，竞争模型的有效性将被非重叠区域所影响，因而不能够精确地描述不同的目标对重叠区域的争夺。与 [106] 不同的，文献 [107] 只在重叠区域上对目标间的遮挡进行建模，但是它们的模型是基于像素级别，因此在两个目标物体的颜色相近的情况下，往往不是很鲁棒。在本文中，我们将目标间的重叠区域看成一个整体，并将该部分的图像观测投影到各个目标对应于区域的子空间上（见图 5-9），并将投影后的重构误差通过指数变换转换成对应种群的适应值，用于衡量该种群对该重叠区域资源的争夺强度，

$$power^k = p(\hat{o}_{t,k}|x_{t,k}) = \exp(-||\hat{o}_{t,k} - \hat{U}_k \hat{U}_k^T \hat{o}_{t,k}||^2) \quad (5-31)$$

其中 \hat{U}_k 为重叠区域在目标 k 上所对应的子空间。根据公式 (5-31)，对于目标 k_1 而言，它与目标 k_2 之间的交互似然函数 $p(\hat{o}_{t,k_1}|x_{t,k_1}, x_{t,k_2})$ 可以定义如下¹，

$$p(\hat{o}_{t,k_1}|x_{t,k_1}, x_{t,k_2}) = \frac{power^{k_1}}{\sum_{i=1,2} power^{k_i}} \quad (5-32)$$

有了交互似然函数的定义，在遮挡情况下，我们用它来表示每个目标所对应的种群的竞争力。在这个定义下，具有更高竞争力的种群更有可能赢得公共区域，这个对应的跟踪结果就是该目标更有可能是处于前面的物体，遮挡了其他目标物体。这个结论将会在实验部分得到验证，见 §5.2.4.2。

排斥模型：众所周知，多目标跟踪算法在遮挡情况下会导致错误聚合问题（coalescence），即其中被遮挡物体的跟踪器丢失了与它相关联的目标，并被

¹这里我们同样假定遮挡发生在目标 k_1 和 k_2 之间。

遮挡它的目标所吸引，导致两个跟踪器聚合到一个目标上去。在实际的生物世界中却往往不会发生这种情况，一个竞争力更强的种群将会对竞争力弱的种群进行排斥，不会与它们分享食物或者资源。为了解决错误聚合的问题，我们提出了排斥模型，在该模型中，种群的排斥是通过它们之间的排斥力来实现的。当目标 k_1 和 k_2 之间发生遮挡时，目标 k_2 对目标 k_1 的排斥力定义如下，

$$F_{k_2, k_1}^{\rightarrow} = p(\hat{o}_{t, k_2} | x_{t, k_1}, x_{t, k_2}) V_{k_1} \quad (5-33)$$

其中 V_{k_1} 是目标物体 k_1 的速度矢量，尺度参数 $p(\hat{o}_{t, k_2} | x_{t, k_1}, x_{t, k_2})$ 由目标物体 k_2 的竞争强度所决定，代表着目标 k_2 对目标 k_1 的排斥强度。当某个目标竞争力越强时，它对其他物体的排斥力也将会越大。

我们将目标间排斥力引入到粒子进化的迭代公式中（见 §5.2.2.3），用于防止某种群的粒子被其他种群所吸引，从而有效的克服了遮挡下错误聚合问题。

5.2.2.3 基于高斯退火的粒子群优化

在 §5.1.1.2 的讨论中，我们发现传统的粒子群优化算法需要人为的设置多个参数：加速度参数 φ_1, φ_2 ，最大限制速度 v_t^{max} 或者惯性权重因子 w^n 。如果缺乏有效合理的机制来控制这些参数的话，粒子群优化算法将存在发散的危险。因此在这个部分中，我们提出了基于高斯退火的粒子群优化算法，其中粒子状态和速度更新如下，

$$v^{i, n+1} = |r_1|(p^i - x^{i, n}) + |r_2|(g - x^{i, n}) + \epsilon \quad (5-34)$$

$$x^{i, n+1} = x^{i, n} + v^{i, n+1} \quad (5-35)$$

其中 $|r_1|$ 和 $|r_2|$ 为高斯随机数 $\mathcal{N}(0, 1)$ 的绝对值， ϵ 为零均值的高斯白噪声，用于避免算法陷入局部极小值，它的协方差矩阵以一种自适应退火的方式 [108] 被定义，

$$\Sigma_\epsilon = \Sigma e^{-cn} \quad (5-36)$$

其中 Σ 为状态转移分布的协方差矩阵, c 为退火常数, n 为迭代次数。从以上定义可以看出, 协方差矩阵的元素随着迭代次数的增加而迅速减小, 因此可以加快收敛速度。

当目标 k_1 和 k_2 之间在时刻 t 发生遮挡时, 我们将上述定义的排斥力加入到粒子进化过程中, 因此粒子状态和速度的更新方程定义如下,

$$v_{t,k_1}^{i,n+1} = |r_1|(p_{t,k_1}^i - x_{t,k_1}^{i,n}) + |r_2|(g_{t,k_1} - x_{t,k_1}^{i,n}) + |r_3|F_{\overrightarrow{k_2,k_1}} + \epsilon \quad (5-37)$$

$$x_{t,k_1}^{i,n+1} = x_{t,k_1}^{i,n} + v_{t,k_1}^{i,n+1} \quad (5-38)$$

同样的, r_3 为高斯随机数 $\mathcal{N}(0,1)$ 的绝对值, 右边式子第三项表示目标 k_2 对目标 k_1 的排斥效应。

下面我们对上述两部分进行总结, 在上述两个部分中, 我们对目标间的竞争现象在图像观测层面进行建模, 同时将他们竞争的结果排斥效应在目标状态空间层次建模, 然后将建立好的模型融入到粒子进化过程中。实验结果 (见 §5.2.4.2) 证明, 我们的模型是合理和有效的。

5.2.2.4 表观模型的选择性更新

传统的多目标跟踪算法中, 在遮挡发生时, 都会停止对目标表观模型的更新。但是如果目标的表观或者光照在遮挡过程中发生改变的话, 跟踪系统可能会在遮挡之后丢失待跟踪的目标。为了解决遮挡情况下目标表观的更新问题, 我们提出了一种选择性更新的策略: 1) 如图 5-8 所示, 对于非重叠部分, 我们采用增量子空间方法 [91] 对其进行更新; 2) 对于重叠部分, 我们首先计算该区域中每个像素在其对应子空间上的重构误差,

$$R = \|\hat{o}_{t,k} - \hat{U}_k \hat{U}_k^T \hat{o}_{t,k}\|^2 \quad (5-39)$$

如果该区域中像素的重构误差小于一定的阈值的话, 我们就对其进行更新, 否则的话就保留它原来的像素值。

5.2.2.5 算法总结

为了使算法更加简洁易懂，我们还是采用两个目标物体之间的遮挡作为例子，而三个目标或者更多目标之间发生遮挡的算法也可以进行类似的描述。另外，该算法只描述了如何处理发生遮挡下的多目标跟踪，对于没有发生遮挡的跟踪器，可以分别用没有排斥力项的单目标序列粒子群优化算法来实现。下面我们将给出基于多种群粒子群优化的多目标跟踪算法：

- ① 初始化： $t = 0$ ，待跟踪的多个目标的状态手动初始化为对应种群的群体最优状态 $\{g_{t,k}\}_{k=1}^M$ ，每个种群对应的个体最优状态 $\{p_{t,k}^i\}_{i=1}^N$ 设置为与群体最优状态相同。
- ② **for** $t = 1, 2, \dots$ **do**
- ③ 根据上一时刻种群的群体最优状态 $\{g_{t-1,k}\}_{k=1}^M$ 来判断是否有遮挡发生，例如，在状态 g_{t-1,k_1} 和 g_{t-1,k_2} 间检测到遮挡。
- ④ 利用状态转移分布对每个种群个体最优状态集进行随机传播，以增强粒子的多样性，

$$x_{t,k}^{i,0} \sim \mathcal{N}(p_{t-1,k}^i, \Sigma_k)$$

其中 Σ_k 对应第 k 个目标的状态转移分布的协方差矩阵。

- ⑤ **for** $n = 1, 2, \dots, T$ **do**
- ⑥ 对目标 k_1 的粒子群采用 PSO 迭代进行进化。

$$v_{t,k_1}^{i,n+1} = |r_1|(p_{t,k_1}^i - x_{t,k_1}^{i,n}) + |r_2|(g_{t,k_1} - x_{t,k_1}^{i,n}) + |r_3|F_{k_2,k_1} + \epsilon$$

$$x_{t,k_1}^{i,n+1} = x_{t,k_1}^{i,n} + v_{t,k_1}^{i,n+1}$$
- ⑦ 对新进化粒子的适应值利用表观模型和交互模型来进行评价。

$$f(x_{t,k_1}^{i,n+1}) = p(\tilde{o}_{t,k_1} | x_{t,k_1}^{i,n+1})p(\hat{o}_{t,k_1} | x_{t,k_1}^{i,n+1}, g_{t,k_2})$$

- ⑧ 根据适应值评价结果更新个体最优状态和全体最优状态。
- ⑨ 对于目标物体 k_2 采用类似的迭代过程（对于没有发生遮挡的跟踪器，可以采用没有排斥力项的迭代过程来实现）。
- ⑩ 收敛性判断： $f(g_{t,k_i}) > Th$ ，且 $f(g_{t,k_i}), i = 1, 2$ 与上次迭代结果比变化小于一定的阈值。
- ⑪ 如果目标 k_i 满足收敛条件，
停止对该目标的迭代；
- ⑫ **end for**
- ⑬ 更新对应目标的表观模型。
- ⑭ 输出 t 时刻的跟踪结果： $\{g_{t,k}\}_{k=1}^M$
- ⑮ **end for**

5.2.3 算法理论分析

在这部分中，我们首先推导出在遮挡情况下目标物体 k_1 的最优重要性建议分布，然后表明 §5.2.2.5 中的算法从本质上来说是通过多层采样来逼近最优重要性建议分布采样的结果。

5.2.3.1 最优重要性建议分布

在文献 [105] 中，Doucet 等证明了在最小化重要性权重方差的意义下， $p(x_t|x_{t-1}^i, o_t)$ 为最佳的重要性建议分布。然而，在实际非线性和非高斯的情况下，我们不可能直接对 $p(x_t|x_{t-1}^i, o_t)$ 进行采样，原因如下：1) 我们无法得到 $p(x_t|x_{t-1}^i, o_t)$ 的解析形式，因此无法对它进行采样；2) 在这个分布基础上，我们无法对粒子进行评价 $p(o_t|x_{t-1}^i) = \int p(o_t|x_t)p(x_t|x_{t-1}^i)dx_t$ 。

在多目标跟踪过程中，如果目标 k_1 与其他目标之间没有发生遮挡，明显的，根据 Doucet 的理论，目标物体 k_1 的最佳重要性建议分布为 $p(x_{t,k_1}|x_{t-1,k_1}^i, o_t)$ 。但是，如果目标 k_1 与其他目标之间发生遮挡，例如，目标 k_1 与目标 k_2 之间发生了遮挡，那么对于目标 k_1 ，它的最优重要性建议分布又是什么呢？

定理 1. 如果遮挡发生于目标 k_1 与目标 k_2 之间，在最小化重要性权重方差的意义下， $q(\cdot)=p(x_{t,k_1}|x_{t-1,k_1}^i, o_t, g_{t,k_2})$ 为目标 k_1 的最佳重要性建议分布。

证明：根据粒子滤波算法的定义 [45]，目标 k_1 的第 i 个粒子的权重计算如下，

$$w_t^i = w_{t-1}^i \frac{p(o_t|x_{t,k_1}^i, g_{t,k_2})p(x_{t,k_1}^i|x_{t-1,k_1}^i)}{q(\cdot)} \quad (5-40)$$

因此，权重 w_t^i 的方差可以计算如下：

$$\text{var}_{q(\cdot)}(w_t^i) = E((w_t^i)^2) - E^2(w_t^i)$$

其中 $E(\cdot)$ 为期望函数，根据期望的定义，上述式子中的期望可计算如下，

$$\begin{aligned} E((w_t^i)^2) &= (w_{t-1}^i)^2 \int \frac{(p(o_t|x_{t,k_1}^i, g_{t,k_2})p(x_{t,k_1}^i|x_{t-1,k_1}^i))^2}{q(\cdot)} dx_{t,k_1} \\ E(w_t^i) &= \int p(o_t|x_{t,k_1}^i, g_{t,k_2})p(x_{t,k_1}^i|x_{t-1,k_1}^i) dx_{t,k_1} \\ &= \int p(o_t|x_{t,k_1}^i, g_{t,k_2})p(x_{t,k_1}^i|x_{t-1,k_1}^i, g_{t,k_2}) dx_{t,k_1} \\ &= p(o_t|x_{t-1,k_1}^i, g_{t,k_2}) \end{aligned}$$

因此可得到，

$$\text{var}_{q(\cdot)}(w_t^i) = (w_{t-1}^i)^2 \left[\int \frac{(p(o_t|x_{t,k_1}^i, g_{t,k_2})p(x_{t,k_1}^i|x_{t-1,k_1}^i))^2}{q(\cdot)} dx_{t,k_1} - p^2(o_t|x_{t-1,k_1}^i, g_{t,k_2}) \right]$$

当方差为零时，我们可以得到 $q(\cdot) = p(x_{t,k_1}|x_{t-1,k_1}^i, o_t, g_{t,k_2})$ 。

然而，在实际情况中，我们也无法直接对 $p(x_{t,k_1}|x_{t-1,k_1}^i, o_t, g_{t,k_2})$ 进行采样，原因与上述无法对 $p(x_t|x_{t-1}^i, o_t)$ 进行采样类似。而逼近该分

布 $p(x_{t,k_1}|x_{t-1,k_1}^i, o_t, g_{t,k_2})$ 的采样结果比逼近 $p(x_t|x_{t-1}^i, o_t)$ 的结果更加困难，它除了要将当前帧的观测信息 o_t 融入到状态转移分布，还需要获得当前帧目标 k_2 的最优状态 g_{t,k_2} ，而目标 k_2 在 t 时刻的最优状态 g_{t,k_2} 实际上也是未知的，也是需要通过算法去求得的。如何通过交叉迭代使得粒子逼近最佳重要性建议分布是解决多目标跟踪的关键所在。

5.2.3.2 采样策略分析

我们通过对 §5.2.2.5 中的算法进行分析，发现该算法可以通过以下两个步骤来逼近最佳重要性建议分布的采样结果：1) 通过分级采样策略将当前帧的观测 o_t 引入到状态转移分布 $p(x_t|x_{t-1})$ 中来；2) 通过交叉迭代将目标 k_2 的最优状态 g_{t,k_2} 融入到粒子进化过程。

首先来看分级采样过程，该部分分为两个步骤，第一个步骤是粗采样部分，通过对状态转移分布采样得到初始粒子，

$$x_{t,k}^{i,0} \sim \mathcal{N}(p_{t-1,k}^i, \Sigma_k) \quad (5-41)$$

第二个步骤是细采样步骤，初始粒子群通过粒子群优化进行迭代，并对进化后的粒子利用对应的图像观测进行评价。我们来看具体一次迭代过程，

$$v_{t,k_1}^{i,n+1} = |r_1|(p_{t,k_1}^i - x_{t,k_1}^{i,n}) + |r_2|(g_{t,k_1} - x_{t,k_1}^{i,n}) + |r_3|F_{\overrightarrow{k_2,k_1}} + \epsilon \quad (5-42)$$

$$x_{t,k_1}^{i,n+1} = x_{t,k_1}^{i,n} + v_{t,k_1}^{i,n+1} \quad (5-43)$$

其中 r_1 , r_2 和 r_3 是高斯随机数 $\mathcal{N}(0,1)$, ϵ 是一个零均值的高斯噪声，它的协方差矩阵为 Σ_ϵ 。假设目标状态 $x_t \in \mathbb{R}^d$ 是一个 d 维向量，则向量 $|r_1|(p_{t,k_1}^i - x_{t,k_1}^{i,n})$ 中的第 l 个分量满足下列分布：

$$|r_1|(p_{t,k_1}^i - x_{t,k_1}^{i,n})_l \sim \begin{cases} 2\mathcal{N}(0, (p_{t,k_1}^i - x_{t,k_1}^{i,n})_l^2) [0, +\infty), & \text{if } (p_{t,k_1}^i - x_{t,k_1}^{i,n})_l \geq 0 \\ 2\mathcal{N}(0, (p_{t,k_1}^i - x_{t,k_1}^{i,n})_l^2) (-\infty, 0), & \text{else} \end{cases}$$

其中 $l = 1, \dots, d$ ，这样的话，向量 $|r_1|(p_{t,k_1}^i - x_{t,k_1}^{i,n})$ 满足下列分布

$$|r_1|(p_{t,k_1}^i - x_{t,k_1}^{i,n}) \sim R_1 = 2\mathcal{N}(0, \Sigma_1), \quad \Sigma_1 = \begin{pmatrix} (p_{t,k_1}^i - x_{t,k_1}^{i,n})_1^2 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & (p_{t,k_1}^i - x_{t,k_1}^{i,n})_d^2 \end{pmatrix}$$

根据实际迭代中向量 $(p_{t,k_1}^i - x_{t,k_1}^{i,n})$ 各个分量的符号，可以决定该分布的定义域，大于零对应于 $[0, +\infty)$ ，小于零对应于 $(-\infty, 0)$ 。

同理可得，

$$|r_2|(g_{t,k_1} - x_{t,k_1}^{i,n}) \sim R_2 = 2\mathcal{N}(0, \Sigma_2), \quad \Sigma_2 = \begin{pmatrix} (g_{t,k_1} - x_{t,k_1}^{i,n})_1^2 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & (g_{t,k_1} - x_{t,k_1}^{i,n})_d^2 \end{pmatrix}$$

$$|r_3|F_{\overrightarrow{k_2, k_1}} \sim R_3 = 2\mathcal{N}(0, \Sigma_3), \quad \Sigma_3 = \begin{pmatrix} (F_{\overrightarrow{k_2, k_1}})_1^2 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & (F_{\overrightarrow{k_2, k_1}})_d^2 \end{pmatrix}$$

另外，噪声 ϵ 满足以下分布 $\epsilon \sim R_4 = \mathcal{N}(0, \Sigma_\epsilon)$ ，这样的话，我们就可以获得隐含在迭代公式 (5-42) 和 (5-43) 中的采样分布： $R = R_1 * R_2 * R_3 * R_4$ 加上 $x_{t,k_1}^{i,n}$ 平移量。尽管我们无法得到该分布的解析形式，但是通过分析可以知道，如果用当前帧与粒子对应观测的评价，来对个体最优状态和全体最优状态进行更新，并将进化后的最优状态用于更新采样分布 R ，那么当前帧的观测信息 o_t 将逐步加入到粒子的进化过程中。

接下来我们分析一下该算法如何把目标 k_2 的最优状态 g_{t,k_2} 引入到粒子进化过程中。在算法 5.2.2.5 中，粒子的评价函数为

$$f(x_{t,k_1}^{i,n+1}) = p(\tilde{o}_{t,k_1} | x_{t,k_1}^{i,n+1}) p(\hat{o}_{t,k_1} | x_{t,k_1}^{i,n+1}, g_{t,k_2})$$

从中我们可以看出，每次粒子迭代过程中都考虑到了当前 g_{t,k_2} 的影响，而 g_{t,k_2} 的具体状态值是通过目标 k_1 和目标 k_2 的似然函数 (5-29) 和 (5-30) 交叉迭代获得的。

综上所述，我们的算法可以有效的将当前帧的观测信息 o_t 和相互遮挡的目标 k_2 的最优状态融入到当前待跟踪目标 k_1 的采样过程，从而有效的逼近了最佳的采样结果。

5.2.4 实验结果

在我们的实验中，目标状态用仿射变化参数来表示： $s = (t_x, t_y, a_1, a_2, a_3, a_4)$ ，其中 $\{t_x, t_y\}$ 表示平移量， $\{a_1, a_2, a_3, a_4\}$ 是变形参数。每个状态粒子所对应的图像通过仿射变换校正为 20×20 的图像块，因此图像特征可以用一个 400 维的向量来表示。所有的实验都是在双 CPU 奔腾 IV 3.4GHz 处理器与 512M 内存的 PC 机上运行。

5.2.4.1 PSO vs AGPSO

状态估计：首先，我们将基于高斯退火粒子群优化框架（AGPSO）与常规的粒子群优化算法（PSO）分别应用到非线性系统状态估计的任务。为了统一比较标准，本文采用文献 [53] 中的非线性系统，该系统被很多学者广泛应用。

该系统的状态转移模型为

$$x_t = 1 + \sin(w\pi(t-1)) + \phi_1 x_{t-1} + v_{t-1}, \quad x_t \in \mathbb{R} \quad (5-44)$$

其中 v_{t-1} 是伽玛分布 $\mathcal{Ga}(3, 2)$ 的随机数，用来模拟系统噪声； $w = 4e - 2$ 和 $\phi_1 = 0.5$ 为尺度参数。非静态观测模型为，

$$y_t = \begin{cases} \phi_2 x_t^2 + n_t, & t \leq 30 \\ \phi_3 x_t - 2 + n_t, & t > 30 \end{cases} \quad (5-45)$$

其中 $\phi_2 = 0.2, \phi_3 = 0.5$ ，观测噪声 n_t 是服从 $\mathcal{N}(0, 0.00001)$ 分布的高斯随机数。该状态估计的任务是，给定带有噪声的观测 y_t ，利用不同滤波器对系统状态 x_t 进行估计。

PSO 框架和 APSO 框架下所需的参数设置如下： $\Sigma = 0.8, c = 2, \varphi_1 = \varphi_2 = 1, w_{max} = 0.8, w_{min} = 0.1, T = 20$ 。图 5-10 显示了两种框架在该非线性系

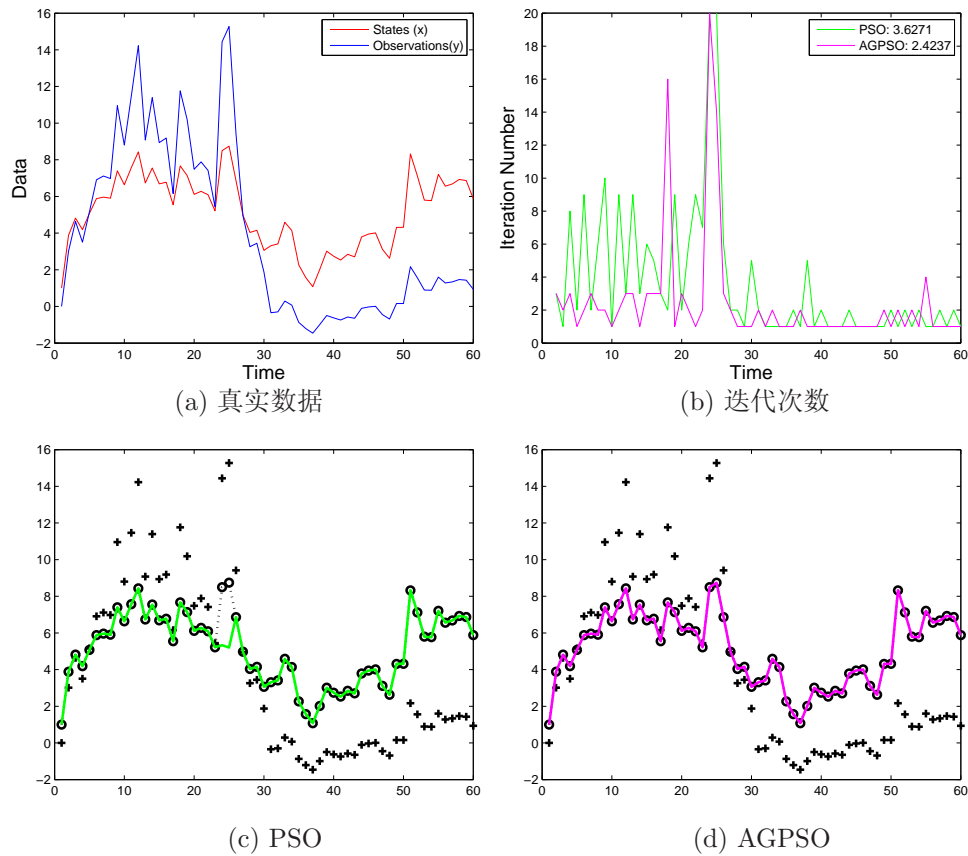


图 5-10 状态估计的一次运行结果

Algorithm	MSE mean	MSE var	Time(s)
PSO	0.13019	0.044086	10.2087
AGPSO	0.060502	0.06852	6.8005

表 5-2 状态估计的实验结果

统上一次运行的结果。从图 5-10 (c) 可以看出, PSO 框架在第 24 和 25 时刻估计错误, 因此这两个时刻的观测被噪声污染得比较严重。而 AGPSO 在整个过程中都可以取得不错的估计结果, 原因是因为它引入了一项基于退火的噪声因子, 避免了陷入局部极值 (见图 5-10 (d))。同时, 我们还比较了两个框架的收敛速度, 如图 5-10 (b) 所示, 其中 AGPSO 只需要平均 2.4327 次迭代就可以达到收敛, 而 PSO 需要平均 3.6271 次。由于该非线性系统带有随机噪声, 因此就一次运行而言, 它的结果带有一定的随机性, 因此对该实验在随机

初始化条件下重复 100 次，表 5-2 对两种框架的下列三个方面进行了总结：均方误差的均值和方差，平均运行时间。可以明显看出，AGPSO 框架的估计精度要远好于 PSO 框架，另外，它也比 PSO 收敛的更快。

单目标跟踪：在这一部分，我们将上述两个框架分别应用到单目标跟踪任务上。其中粒子数和状态转移分布的协方差矩阵分别设置为 $\{N = 200, \Sigma = \text{diag}(8^2, 8^2, 0.02^2, 0.02^2, 0.002^2, 0.002^2)\}$ ，另外，两种框架采用相同表观模型。图 5-11 的第一行显示了显示了两个框架在快速运动序列上的测试结果，其中蓝色框为 PSO 的跟踪结果，红色框为 AGPSO 的跟踪结果。图 5-11 的第二行为跟踪结果的均方根误差（左边）和算法平均运行时间（右边），从中我们可以看出 AGPSO 取得了比 PSO 稍好的跟踪定位精度，并且取得了比 PSO 更快的收敛速度。

上述实验结果产生的原因是：每一次迭代过程中粒子的惯性速度携带了很少有用的信息，因此公式（5-4）中的惯性速度这一部分实际上对状态估计或者跟踪并没有太多帮助；而 AGPSO 框架中的噪声项首先增强了粒子的多样性，使得算法能跳出局部极值，其自适应退火效应使得算法能够很快收敛。

5.2.4.2 多目标跟踪

在这个部分中，我们将基于多种群粒子群优化的多目标跟踪算法应用到三

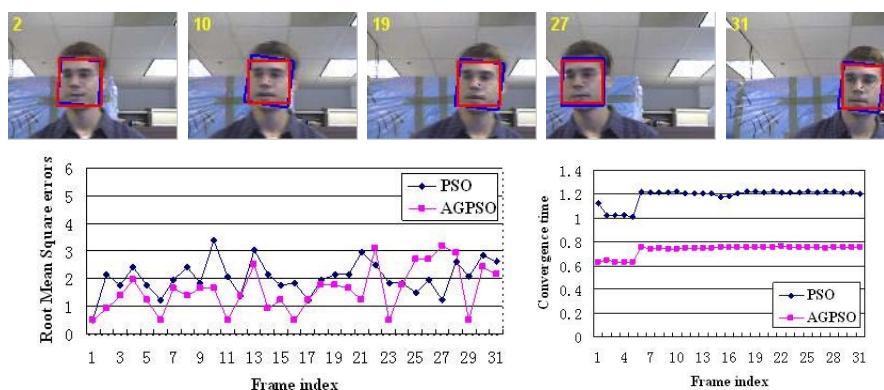


图 5-11 第一行：跟踪结果（蓝色：PSO，红色：AGPSO）；第二行：均方根误差（左边），算法运行时间（右边）

个视频序列上，并对实验结果进行分析和总结。

例子 1：例子 1 中的视频包含两个行人，并且发生严重遮挡。在这个视频中，我们将测试下面两种表观模型的更新策略对跟踪结果的影响：1) 在遮挡情况下停止对表观模型的更新；2) 我们的选择性更新策略。如图 5-12 所示，我们可以看到，没有更新的表观模型在第 211 帧丢失了被遮挡的行人，并且再也无法恢复。原因是被遮挡的行人在遮挡过程中姿态发生了缓慢的变化，等到他从遮挡中恢复出来之后，他与原先的子空间模型的像素对应关系不再准确，因此导致跟踪失败。相比之下，我们的选择性更新策略有效跟踪两个行人直至他们分开，并且保留正确的身份。这是因为该策略对遮挡过程中的表观变化进行了选择性更新，利用遮挡过程中所能利用的一切视觉信息来帮助跟踪器对被遮挡行人在遮挡结束后的重新定位。

例子 2：在这个例子中，我们的测试视频序列来自于国际上公开的监控测试数据库 PETS2004，该数据库可以从下面这个链接获得：<http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>。为了进一步说明我们方法的优越性，将我们的算法与 Qu 的算法 [106]，Yang 的算法 [107] 进行了比较，因为他们的工作与我们的算法有一定相似之处，并且他们的算法是在两个跟踪领域中经典的算法框架下设计的：粒子滤波 [51]，均值漂移 [38]。在我们的实现过程中，Qu 的

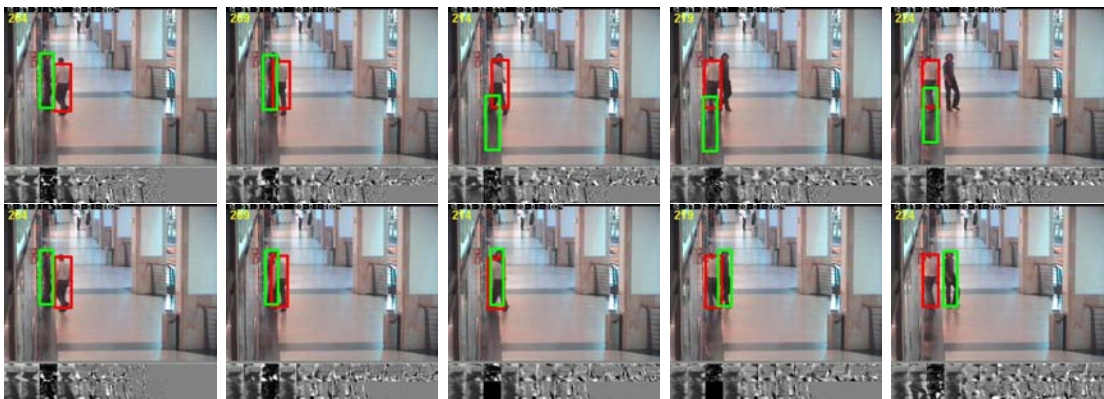


图 5-12 例子 1 的跟踪结果，包含帧 204, 209, 214, 219, 224（第一行：在遮挡过程中不对表观模型进行更新；第二行：我们的选择性更新策略）

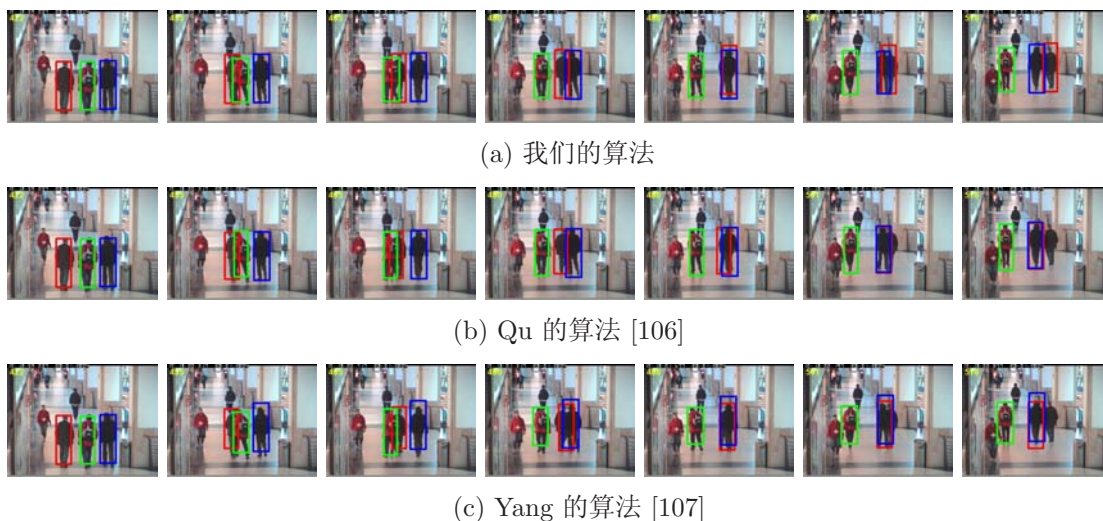


图 5-13 例子 2 的跟踪结果，包含帧 422,455,465,480,488,501,518

算法采用与我们算法一样的表观模型和更新策略。图 5-13 列举了整个遮挡过程中几帧重要的跟踪结果，其中行人 A 用红色框来表示，行人 B 用绿色框来表示，行人 C 用蓝色框来表示。从中我们可以看到，我们的算法可以很好地处理不同目标之间的交互和遮挡。而当行人 A 被行人 C 遮挡时，并且行人 C 的表观与行人 A 的表观较为相似，在这种情况下 Qu 和 Yang 的工作都对行人 A 跟踪失败，原因是由于它们对遮挡情况的建模不够有效和鲁棒。另外，我们的 AGPSO 框架在寻找全局最优状态方面比粒子滤波算法和均值漂移算法更强，因此可以有效跳出局部极值。图 5-14 显示了例子 2 中各个目标之间的遮挡关系，其中 x 轴表示例子 2 的帧索引号， y 表示遮挡关系。从图 5-14 可以看出，我们的方法可以通过种群的竞争能力正确推导出目标之间的遮挡关系，而该实验也验证了 §5.2.2.2 的假设：在遮挡情况下，具有更高竞争力的种群更有可能赢得公共区域，用跟踪术语来讲的话，即该目标更有可能是处于前面的物体，遮挡了其他对象目标。

为了进一步验证我们提出的方法的性能，将我们的算法与 Qu 和 Yang 的工作在以下两个方面做了定量比较实验：成功跟踪的帧数（跟踪成功率），跟踪定位结果与 groundtruth 之间的均方根误差（由于该数据中公布 groundtruth

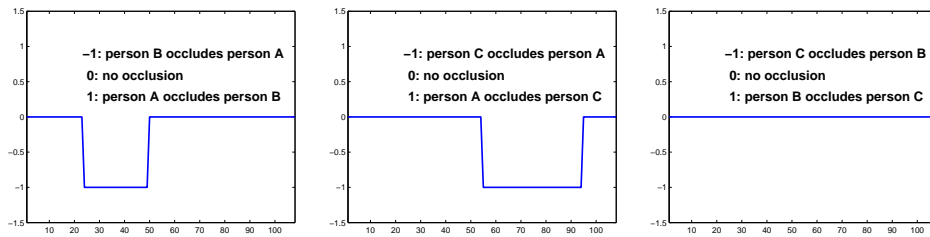


图 5-14 例子 2 中所恢复的遮挡关系。

方法		Yang 的算法	Qu 的算法	我们的算法
成功被跟踪的帧数	行人 A	80/101	80/108	108/108
	行人 B	108/108	108/108	108/108
	行人 C	108/108	108/108	108/108
中心位置误差 (像素)	行人 A	12.9768	11.5537	3.6145
	行人 B	5.4128	4.8482	3.3087
	行人 C	15.2104	2.6483	2.6262

表 5-3 定量比较结果

并不是很紧致的矩形框，而是包含许多背景像素。以这样的 groundtruth 作为初始化不适合基于子空间跟踪的算法，因此，只有目标 groundtruth 状态中的中心位置用于定位精度的评估）。表 5-3 列出了三种算法在上述两个定量指标上的结果。从表中可以看出，Qu 和 Yang 的算法在 501 帧时发生了错误聚合问题，即行人 A 的跟踪器被行人 C 错误的吸引。而我们的方法由于利用了种群的竞争和排斥模型，可以有效地防止错误聚合的问题，成功地跟踪整个序列。此外，我们的方法达的定位精度也远好于其他两种算法。

例子 3： 第三个例子中的视频序列同样来自于 PETS2004 的数据库。这是个非常具有挑战性的视频序列，其中包含 5 个行人行走、站立交谈、遮挡。图 5-15 显示了从 200 帧到 410 帧的跟踪结果。从中可以看出，我们的算法很好的跟踪了所有目标，特别是在帧 277-340 中三个行人之间发生了遮挡。从这个实验可以说明我们的种群竞争和排斥模型可以有效地解决两个以上的目标遮挡



图 5-15 购物中心场景中多目标跟踪，包含帧 218,248,272,298,359,406。

问题。

5.2.4.3 讨论与总结

在这一部分，我们对产生上述实验结果的原因进行讨论和总结。首先，种群间的竞争和排斥机制为多目标跟踪中目标之间的交互和遮挡提供了一种可行的解决方案。第二，AGPSO 框架是一种有效的全局最优的搜索框架，特别是在高维状态空间中。第三，我们精心设计的选择性更新策略，可以有效地适应外观的变化，同时防止表观模型的漂移。

第六章 基于树模型解析的姿态估计算法

传统的姿态估计算法通常是将物体各部分的状态连接起来，并通过求解联合空间的优化问题来获得目标的姿态。该类方法面临下列困难：（1）联合状态向量无法描述目标各部分之间的拓扑约束；（2）联合状态向量维数过高，使得优化问题无法收敛。在这一章中，我们提出了一种基于树模型解析的姿态估计算法。该算法从以下三个方面出发来解决上述问题：（1）利用图像语法对人体结构的树模型进行解析，从而减小了搜索空间；（2）优化过程采用两种模式：扩散和跳跃；（3）利用数据驱动的马尔可夫链蒙特卡洛方法来进行优化。通过以上三个方面，该方法可以有效地将人体各部分之间的拓扑关系加入到优化过程中，并通过表观数据的驱动来压缩搜索空间，从而大大提高了收敛速度（10倍左右）和精度。实验结果表明，基于树模型解析的人体姿态估计算法在室内和室外场景下，人体部分遮挡以及各种姿态情况下都非常鲁棒。

6.1 算法概述

在这一部分，我们首先对基于树模型解析的姿态估计算法做一个简要的概述，该算法主要包含以下三个部分：人体模型、问题描述和优化框架。下面将就这三部分一一的展开介绍。

6.1.1 三层人体树模型

如图 6-1 所示，人体各部分机构采用三层的树结构模型来表示，它包含 10 个人体部分：1) 根节点：躯干部分；2) 第二层节点：头部，左/右胳膊上肢，左/右腿部上肢；3) 第三层节点：左/右胳膊下肢，左/右腿部下肢。因此，人体的姿态可以表示为 $\mathcal{X} = \{X_1, X_2, \dots, X_{10}\}$ ，其中 X_i 表示第 i 个人体部分（见图 6-1）。具体来说，每个人体部分 X_i 都可以用一个矩形来表示，其参数可以描述为 $X_i = \{x_i, y_i, \theta_i, l_i, w_i\}$ （见图 6-1 (c)），其中 x_i, y_i 和 θ_i 表示该人体部分相对于其母节点的坐标位置和旋转参数， l_i, w_i 表示该人体部分的长和宽

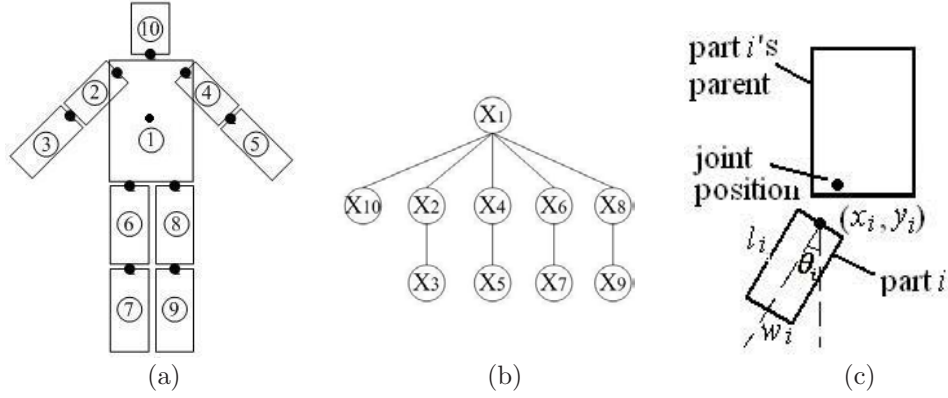


图 6-1 (a) 人体结构图（数字为人体部分标号）； (b) 三层人体树模型； (c) 人体部分状态示意图

（根节点由于没有母节点，因此它的参数 x_1 , y_1 , θ_1 是图像坐标内的绝对值，并且 (x_1, y_1) 表示躯干部分的中心位置）。这种三层人体树模型可以有效地反映人体各部分之间的运动学约束，此外，这种结构可以使用分级搜索策略来加速优化的过程。

6.1.2 问题描述

记图像表观为 I ，则姿态估计问题可以表示为如下后验分布的贝叶斯推理问题 [110]：

$$P(\mathcal{X}|I) \propto P(I|\mathcal{X})P(\mathcal{X}) \quad (6-1)$$

其中 $P(I|\mathcal{X})$ 是给定人体姿态 \mathcal{X} 时的图像表观似然， $P(\mathcal{X})$ 是包含人体约束的先验分布。该问题的一种简单常见的解是最大后验估计（MAP: maximum a posteriori estimate），MAP 估计可以表示如下，

$$\mathcal{X}_{MAP} = \arg \max_{\mathcal{X}} P(\mathcal{X}|I) \quad (6-2)$$

6.1.3 优化框架

姿态估计的优化框架如图 6-2 所示，算法输入包括人体部分前景图（通过背景减除获得）和人体部分边缘图（通过图像分割获得）。在每次迭代中，人

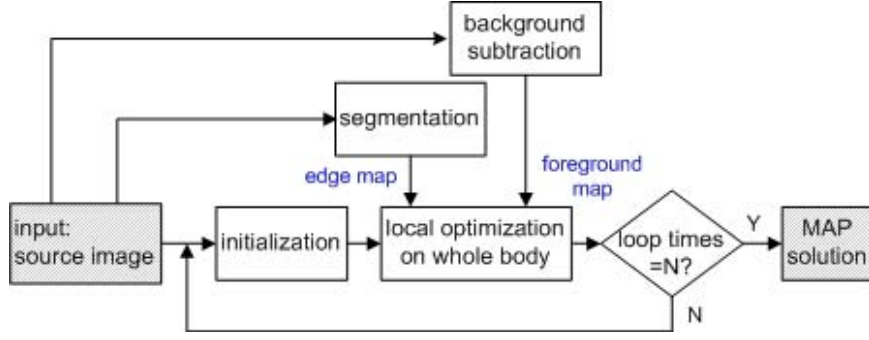


图 6-2 算法流程图

体首先通过先验分布和图像建议分布进行初始化，然后在数据驱动马尔可夫链蒙特卡罗方法下对整个人体进行优化以达到局部极值。上述过程被重复迭代 N 次，最大后验估计被保留作为最终姿态估计的结果。

6.2 后验分布

在这个部分，我们来具体介绍后验分布的计算方法。

6.2.1 图像表观似然

对于公式 (6-1) 中的表观似然，我们同时考虑前景区域似然和边缘图似然，其计算公式如下：

$$\begin{aligned}
 P(I|\mathcal{X}) &= P(I_f \cup I_e|\mathcal{X}) \\
 &= P_f(I|\mathcal{X})P_e(I|\mathcal{X})
 \end{aligned} \tag{6-3}$$

从一系列姿态估计的实验，我们发现边缘图似然 $P_e(I|\mathcal{X})$ 并不像前景区域似然 $P_f(I|\mathcal{X})$ 那么可靠，所以我们对边缘图似然 $P_e(I|\mathcal{X})$ 加入一个衰减因子 $\alpha (\alpha < 1)$ 来限制它对后验分布的影响。因此，图像表观似然可以改写成如下形式：

$$P(I|\mathcal{X}) = P_f(I|\mathcal{X})P_e(I|\mathcal{X})^\alpha \tag{6-4}$$

6.2.1.1 前景区域似然

前景区域似然是用来衡量人体与前景的匹配程度。根据当前人体状态 \mathcal{X} ，

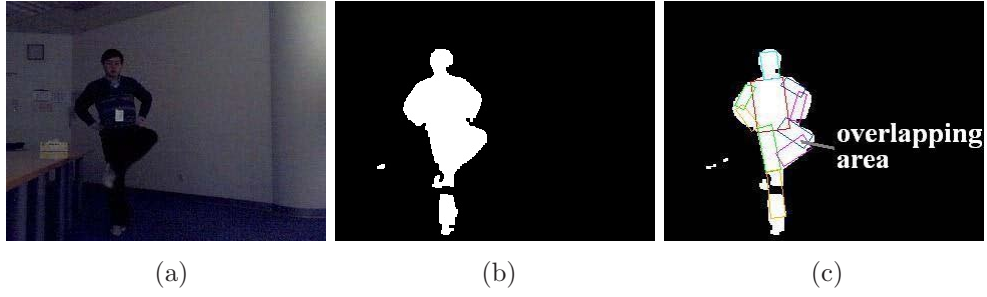


图 6-3 (a) 原图; (b) 前景图; (c) 产生的人体姿态 (彩色矩形框)

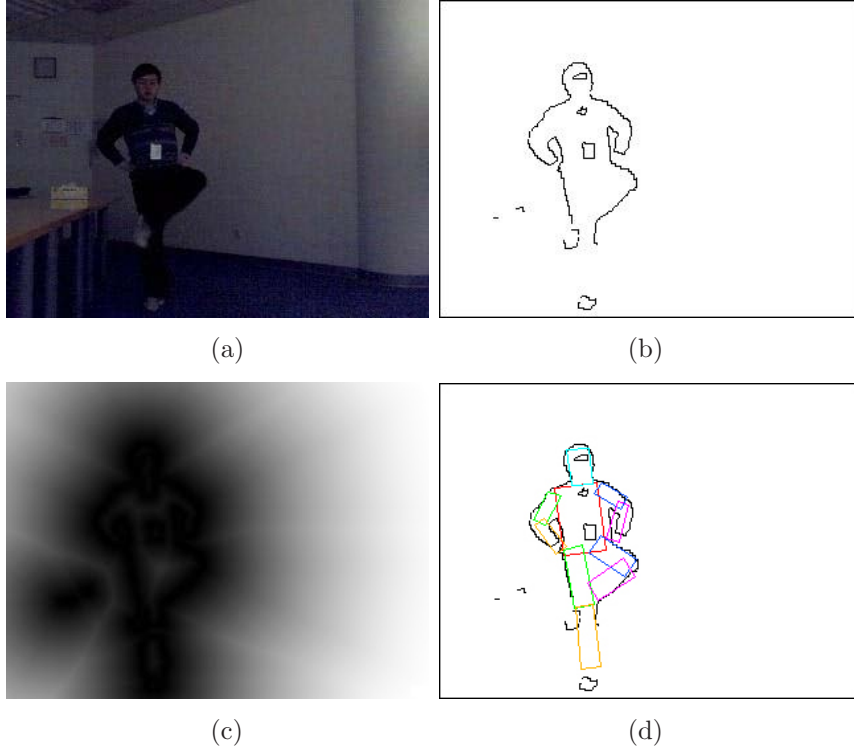


图 6-4 (a) 原图; (b) 边缘图; (c) 边缘距离映射图; (d) 合成的人体姿态 (彩色矩形框)

生成一个人体姿态并与前景图进行比较, 如图 6-3 所示。令 S_i 表示人体各部分与前景区域交集的面积, S_u 表示二者并集的面积, S_o 表示人体各部分之间交叠区域的面积, 基于上述的定义, 前景区域似然可以表示为

$$P_f(I|\mathcal{X}) \propto \exp\left(\frac{S_i - \beta S_o}{S_u}\right) \quad (6-5)$$

其中 $\beta(\beta = 0.5)$ 是对人体各部分之间交叠区域的惩罚因子。当由 \mathcal{X} 产生的人体姿态与前景区域符合得最好时 (也就是人体姿态尽可能地覆盖前景, 同时尽

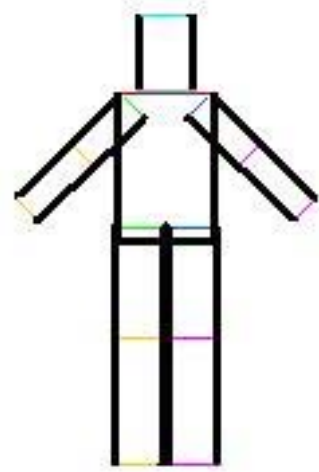


图 6-5 黑色粗线上的像素用于计算边缘似然

量少地覆盖背景以及自我重叠），前景区域似然达到最大值。

6.2.1.2 边缘图似然

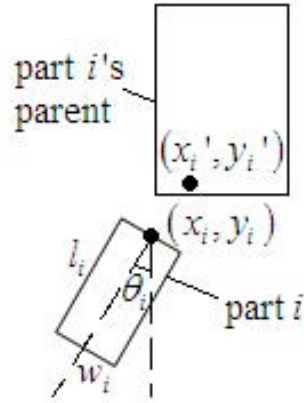
边缘图似然是用来衡量人体姿态与图像边缘的符合程度。首先对前景区域进行膨胀，然后根据膨胀的结果估计人体所在的大致区域，对这个区域内的图像进行分割，得到边缘图，如图 6-4 (b) 所示。通过对边缘图进行距离变换，可以得到边缘距离映射，如图 6-4 (c) 所示。实际上，我们并不是计算产生的人体姿态上的所有边缘点，由于噪声的污染，人体姿态上的很多点边缘并不明显，所以我们对边缘强度较大的点进行计算，如图 6-5 所示。对于产生的人体姿态，图 6-5 中的黑色粗线边框上的每一个点 (u, v) ，记 $d(u, v)$ 为该点在边缘距离映射上的值（见图 6-4 (c)）， E_p 为人体姿态黑色边线上的点的集合， T 为 E_p 中的总点数，边缘图似然可以表示为

$$P_e(I|\mathcal{X}) \propto \exp\left(-\frac{\sum_{(u,v) \in E_p} d(u, v)}{T}\right) \quad (6-6)$$

当身体的边线与图像边缘匹配得最好时，边缘图似然达到最大值。

6.2.2 人体拓扑约束

公式 (6-1) 中的先验分布包含了人体拓扑约束，包括人体各部分之间的空间约束和长度关系（在这里，我们采用躯干区域的长度 l_{torso} 作为参考长度）。

图 6-6 第 i 个人体部分的约束示意图

如图 6-6 所示，对于第 i 个人体部分 $X_i = \{x_i, y_i, \theta_i, l_i, w_i\}$ ，记 (x'_i, y'_i) 为第 i 个人体部分与它母节点之间的链接处的坐标，因此第 i 个人体部分的约束可以表示为，

$$C_i = N(x_i, x'_i, \sigma_{x_i}^2) N(y_i, y'_i, \sigma_{y_i}^2) N(l_i, \mu_{l_i} l_{torso}, \sigma_{l_i}^2 l_{torso}^2) \\ N(w_i, \mu_{w_i} l_{torso}, \sigma_{w_i}^2 l_{torso}^2) M(\theta_i, \mu_{\theta_i}, m) \quad (6-7)$$

其中 $N(\cdot)$ 为高斯分布，在公式 (6-7) 中，它的参数（如均值和方差）都是从训练样本中学习得到。其中角度的分布 M 是 von Mises 分布

$$M(\theta, \mu, m) \propto e^{m \cos(\theta - \mu)} \quad (6-8)$$

其中 m 是一个较小的常数，也可以从训练集中学习得到。

在公式 (6-7) 中， x_i ， y_i 和 θ_i 的约束项用于度量该人体部分与其母节点的相对位置； l_i 和 w_i 的约束项用于度量该人体部分尺度与躯干的相对关系。对于躯干部分，由于它没有母节点， x_i ， y_i 和 θ_i 的约束项都设置成 1。因此，整个人体拓扑约束可以表示为

$$P(\mathcal{X}) \propto \exp\left(\sum_i \omega_i C_i\right) \quad (6-9)$$

其中 ω_i 为第 i 个人体部分的权重，且满足 $\sum_i \omega_i = 1$ 。

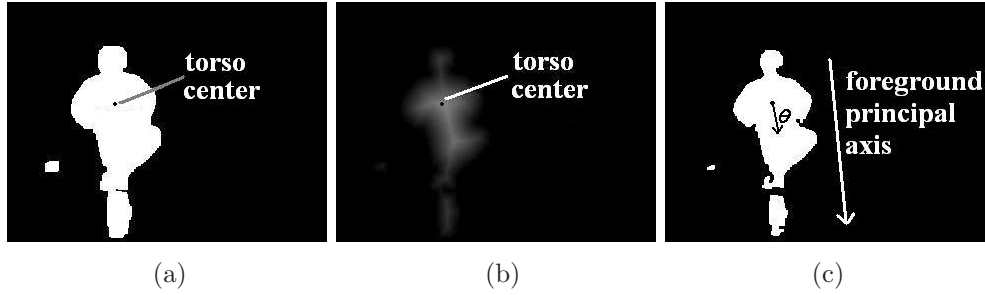


图 6-7 (a) 人体前景区域；(b) 前景距离映射；(c) 躯干方向和前景主轴方向

6.3 姿态估计的优化框架

6.3.1 初始化

由于人体被描述为树形结构，首先应该对根节点（即躯干）进行初始化。我们可以从腐蚀膨胀后的前景区域预测出了人体身高的估计 l_{body} ，假设 $l_{torso} \sim N(\mu_{l_{torso}} l_{body}, \sigma_{l_{torso}}^2 l_{body}^2)$ （其中 $\mu_{l_{torso}}$ 和 $\sigma_{l_{torso}}$ 可以从训练样本学习得到），根据这个分布我们可以直接采样获得 l_{torso} 。在获得 l_{torso} 的初始值后， w_{torso} 可以利用 §6.2.2 中提到的先验人体约束来获得。躯干的其他状态 x_{torso} ， y_{torso} 和 θ_{torso} ，可以通过从图像表观估计其建议性分布（proposal distribution），然后从对应的建议性分布中采样获得它们的初始值，具体过程如下所述。

首先对人体前景区域进行距离变换得到前景距离映射，如图 6-7 所示。如下假设通常是合理的：一个点在前景距离映射上的值越大，它越有可能是躯干的中心。基于这个假设，我们可以获得躯干的中心。另外，对于 θ_{torso} ，它的建议性分布是以前景主轴方向为均值的高斯函数，如图 6-7 所示。

然而，很多时候，这样抽取的初始躯干位置并不好。幸好，通过检验其与前景区域的符合程度，我们可以轻易地判断出许多“坏”躯干。这一点对于降低无谓的计算量，提高算法性能是十分重要的。判断“坏”躯干的标准是：1) 肩部以上前景的面积大于某值（见图 6-8 (a)）；2) 头或躯干中背景的比例大于一定阈值（见图 6-8 (b)）。如果躯干例子被判定为“坏”的，则需重新进

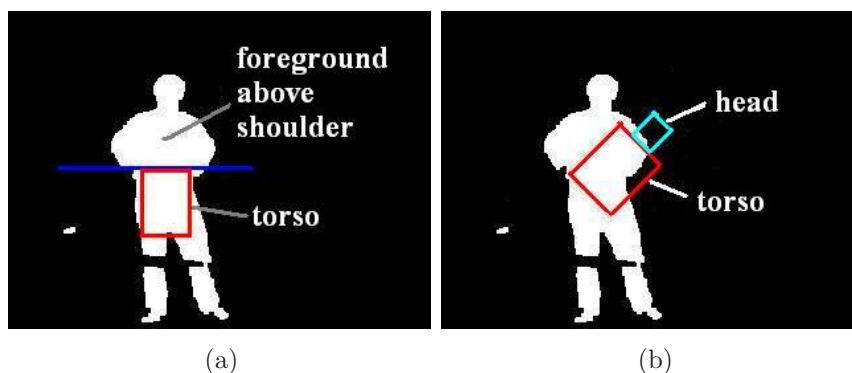


图 6-8 鉴别坏躯干的例子：（a）肩部以上前景区域过大；（b）头与前景匹配不好

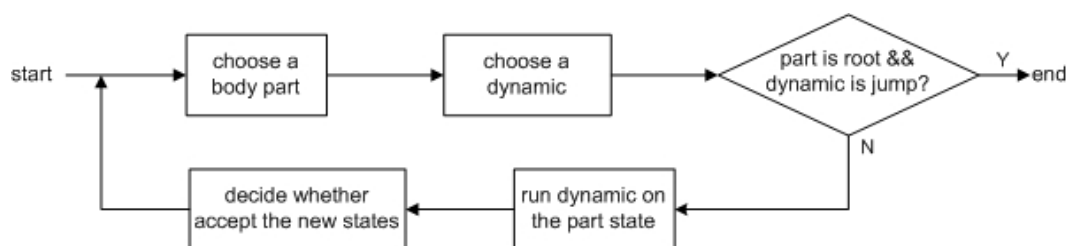


图 6-9 局部优化流程图

行初始化。躯干初始化完成后，其它人体部分的状态可以根据 §6.2.2 中定义的先验分布采样获得。

6.3.2 局部优化

局部优化是在树形结构的状态空间中寻找局部极值的过程。如图 6-9 所示，局部优化过程主要由四个部分组成：1）解析树模型状态空间，然后选择待优化的人体部分状态；2）选择优化模式；3）采用 Metropolis Hasting 方法来决定是否接收新的状态；4）同时，一种概率传递机制被用来引导下次优化中哪个人体部分将被选中。

6.3.2.1 基于语法的树模型解析

在传统的姿态估计算法中，树模型通常用来对人体拓扑结构进行模拟，但是树模型中人体姿态的高维表示是传统人体姿态估计算法的瓶颈所在。因此，在本文中我们提出了树解析语法来对高维树模型进行解析，从而将高维人体模型解析成一系列有序的状态向量，并在该次序上进行优化，从而避免了维数灾

难问题。

语法一:与图像信息吻合越差的人体部分具有优先被优化的权力。

当选择人体部分进行优化时，我们期望选择与前景匹配最差的人体部分。基于这种考虑，我们设计了重要性建议分布，它可以有效地衡量人体部分和前景的匹配程度。首先根据人体状态生成的身体各部分区域并与前景进行比较，然后计算重要性建议分布，该计算过程中将考虑如下几个方面：1) 人体部分区域内的背景面积 S_{bgIn} ；2) 未被覆盖的前景面积 S_{fgOut} （仅考虑虚线框内的区域，如图 6-10 (a) 所示）；3) 与其他人体部分之间重叠的区域 S_o 。根据以上的分析，第 i 个人体部分的重要性建议分布 $P_{impt}(i)$ 可以通过如下公式计算得到，

$$P_{impt}(i) \propto \frac{S_{bgIn}}{S_i} + w \frac{S_{fgOut}(S_i + S_o)}{S_i^2} \quad (6-10)$$

其中 S_i 是第 i 个人体部分的面积， w 是一个权重系数。这个公式意味着一个内部包含大片背景或者重叠区域，又或者周围有大片未被覆盖前景的人体部分，具有较大的重要性建议分布。

可是，仅考虑图像表现是不够的，图 6-10 (b) 给出了一个例子，根据语法一下肢将先被优化，但是由于下肢（子节点）受到上肢（母节点）的关节约束，无法到达正确的位置，因此在这种情况下仅根据语法一无法获得令人满意的结果，在这里上肢应该先于下肢进行优化尽管下肢具有更大的重要性建议分

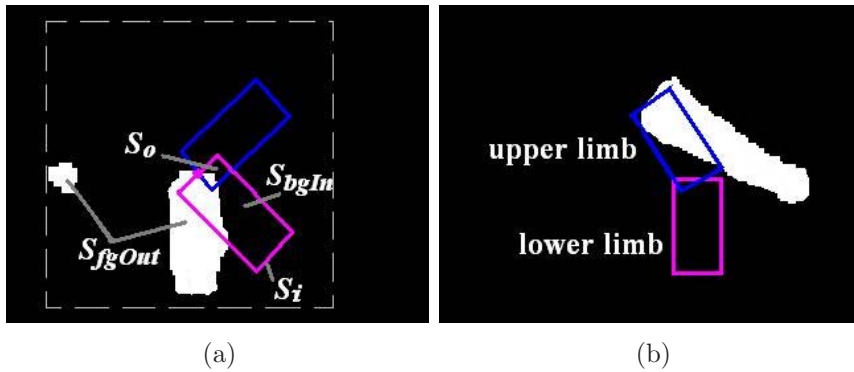


图 6-10 (a) 重要性建议分布计算示意图；(b) 语法一不适合的一个例子

布。因此，我们定义第二个语法，

语法二:人体各部分的优先权从根节点依次下降，同一层上节点的优先权相同。

根据语法二，我们为每个人体部分设计了优先权 $P_{Lprior}(i)$ ，这些优先权包含了人体模型的结构信息，并且将会随着优化过程的演化而改变（见 §6.3.2.4）。因此，可以将人体各部分的重要性表示为

$$P_{priority}(i) \propto P_{impt}(i)P_{Lprior}(i) \quad (6-11)$$

从公式（6-11）我们可以看出，该公式有效地结合了图像观测和人体的拓扑结构来选取最合适的人体部分进行优化。

6.3.2.2 优化模式选择

在选定待优化的人体部分后，我们设计了两种马尔可夫链动态模式：扩散和跳跃。跳跃对应于状态空间的较大变化（可能涉及多个状态）。在这种情况下，首先重新采样选中人体部分的状态，然后对该人体部分的子树（以选中的人体部分为根节点）进行局部优化。这种动态模式使得优化过程可以跳出局部极大值，跳到邻域的其它局部极大值。扩散对应于选定状态的微小变化，它是对选定状态局部细微的搜索操作。为了能对不同的优化模式进行选择，我们定义了如下语法，

语法三: 如果 ① 选定人体部分的重要性建议分布小于给定的阈值，或者 ② 在以选定人体部分作为根节点的子树上达到局部极大值¹，则选择跳跃模式；否则的话选择扩散模式。

当扩散模式被选中的情况下，我们对选定状态 ϕ （ ϕ 为 $\{x_i, y_i, \theta_i, l_i, w_i\}$ 中的一个成分）进行如下更新：

$$\phi' = \phi \pm \lambda \pm \epsilon \quad (6-12)$$

¹当选中的人体部分为根节点时，在这种情况下，优化过程到达收敛。

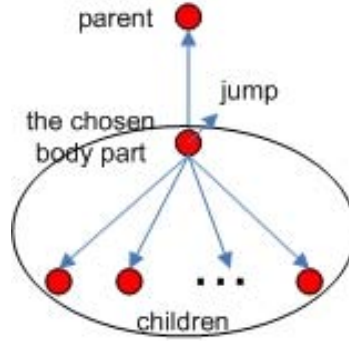


图 6-11 跳跃模式将激活选定人体部分作为根节点的子树上的优化过程

其中 λ 是步长, ϵ 是高斯扰动。其中符号的选择标准是: 选择使后验概率增大的符号。这个更新操作重复执行, 直到后验概率减小或超出范围。这两种优化模式是互补的, 并且可以被语法三很好地协调组织起来, 因此可以高效地搜索复杂的状态空间。

如果跳跃模式被选择, 我们首先对选定的人体部分状态根据 §6.2.2 中定义的先验进行重采样, 然后选择以选定人体部分作为根节点的子树, 并对它开展新的优化过程 (见图 6-11)。这两种优化模式相结合既可以引导一种快速搜索 (扩散) 又可以跳出局部极值 (跳跃)。

6.3.2.3 马尔可夫链蒙特卡罗优化过程

在这里, 我们采用一种马尔可夫链蒙特卡罗优化——Metropolis Hasting 方法来决定是否接受由上述优化模式产生的新人体状态, 其中接受新状态 \mathcal{X}' 替代当前状态 \mathcal{X} 的概率为

$$P(\mathcal{X} \rightarrow \mathcal{X}') = \min \left\{ 1, \frac{P(\mathcal{X}'|I)P(\mathcal{X}|\mathcal{X}')}{P(\mathcal{X}|I)P(\mathcal{X}'|\mathcal{X})} \right\} \quad (6-13)$$

为了简便, 假设 $P(\mathcal{X}|\mathcal{X}') = P(\mathcal{X}'|\mathcal{X})$, 并且为了避免坏结果过于容易被接受, 对后验概率的计算加上一个指数项 $k(k > 1)$, 则公式变为

$$P(\mathcal{X} \rightarrow \mathcal{X}') = \min \left\{ 1, \frac{P(\mathcal{X}'|I)^k}{P(\mathcal{X}|I)^k} \right\} \quad (6-14)$$

6.3.2.4 层次重要性传播

语法四：当选定的人体部分被优化过后，它的层次重要性将传递给它的母节点和子节点。

该语法基于这样的先验知识：如果选定的人体部分被优化过后，它的层次重要性将传递给它的近邻节点，这样的话，它的近邻节点在下一轮优化过程中将有更大的可能性被选择优化。

6.4 实验结果

我们的算法是通过 C++ 编程语言实现的，并且在奔腾 IV2.8GHz 以及 512M 内存的 PC 机上运行的。为了测试该算法的有效性，分别采用自己拍摄的图片以及网上下载的图片来对所提出的算法进行测试，并且这些测试图片集包含了各种姿态、光线变化、复杂背景以及局部遮挡。另外，我们对实验结果分别做了定性和定量的分析，用于验证我们算法的贡献。在实验中采用的参数分别设置为： $\alpha = 0.6$ ， $\beta = 0.5$ ， $w = 0.2$ ， $m = 0.01$ ， $\lambda = 2, 2, 5, 5, 4$ （分别对应人体各个部分状态的 5 个分量）。

6.4.1 自拍图片上的姿态估计性能

在这个部分中，我们用两组自己拍摄的图像序列来对算法进行测试。第一组图像序列是在室内光线条件比较差的环境中拍摄的，它包含了 380 幅图像，图像大小为 320*240（见图 6-12 的第一行和第二行）。第二组图像序列是在室外复杂背景中拍摄的，它包含了 300 幅图像，图像大小为 400*240（见图 6-12 的第三行和第四行）。如图 6-12 的前两行所示，尽管存在着图像成像质量差，姿态复杂等问题，我们的算法在这组图像序列上取得了非常好的结果。另外，从这一行第 118 帧的结果可以看出，我们的算法可以有效的处理局部遮挡问题，这主要是因为我们的算法在搜索过程中加入了人体的拓扑约束。后两行是我们算法在第二组图像集上的测试结果，可以看出该算法可以鲁棒的处理背景

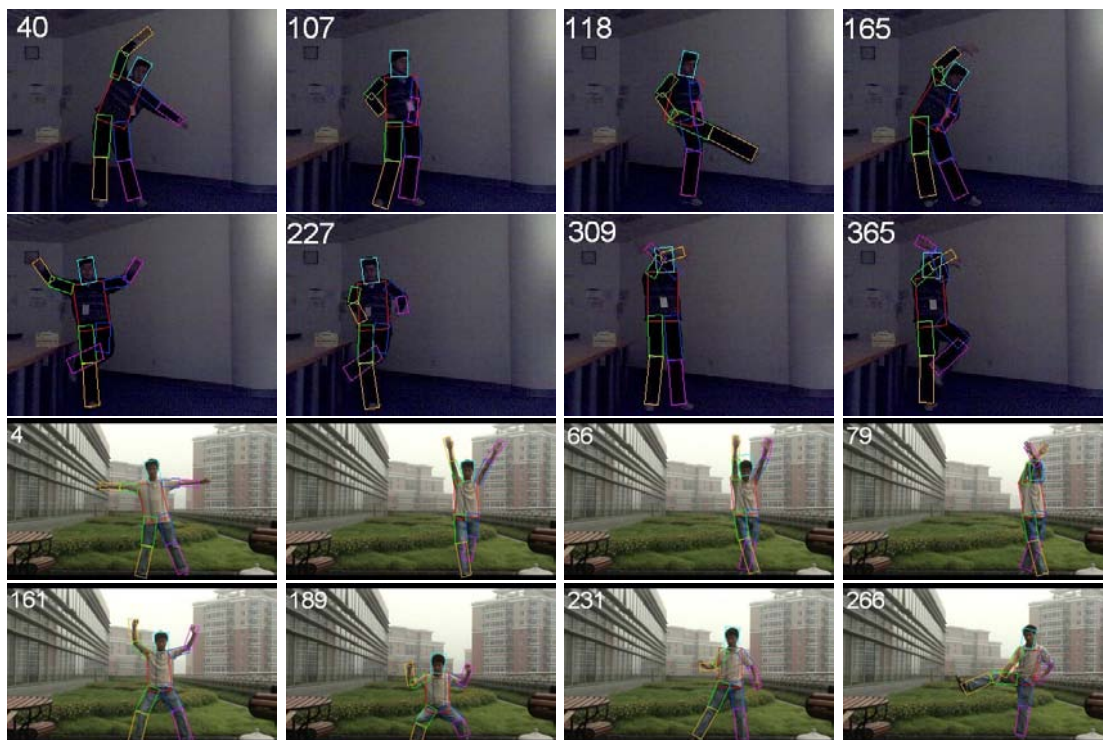


图 6-12 自拍图像的测试结果（第一行：室内环境，第二行：室外环境）

杂乱以及姿态复杂的情况。

6.4.2 网络图片上的测试结果

在该部分中，我们将算法用网络上下载的两组图像进行测试。第一组是一组瑜伽图像序列（见图 6-13 的第一行和第二行），第二组图像是太极拳教程序列（见图 6-13 的第三行和第四行）。这两组图像包含了许多非常复杂且不常见的姿态，因此对这两个序列进行姿态估计是非常具备挑战性的任务。如图 6-13 所示，我们的算法在这两组序列上都取得了令人满意的结果。令人意外的是，第三行和第四行的第二列图像中，人体的上肢被身体完全遮挡，但是我们的算法依旧获得了可以接受的结果。这是因为图像数据驱动和人体关节约束共同作用的结果。



图 6-13 网络图片的测试结果（第一行：瑜伽，第二行：太极拳）

Methods	Number of iteration	Average running time (by minute)
our algorithm	50	1
Gang's work	6	2-3
Zhang's work	—	5

表 6-1 算法平均运行时间

6.4.3 定量分析

为了对算法性能进行定量分析，我们对所有的测试图像进行了手工标定。基于标定的 groundtruth，计算我们的算法在所有图像序列上的均方根误差，除了最大后验估计之外，我们还记录了姿态估计的 Rank5 和 Rank10 的结果。其中 Rank5/Rank10 结果表示后验概率排名前 5/前 10 的结果中均方根误差最小的估计。首先我们来看姿态估计的均方根误差随迭代次数的变化，

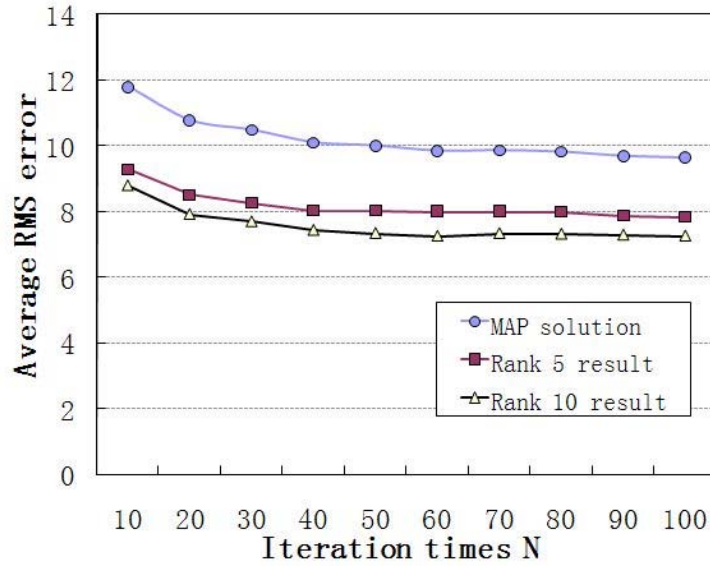


图 6-14 均方根误差收敛曲线

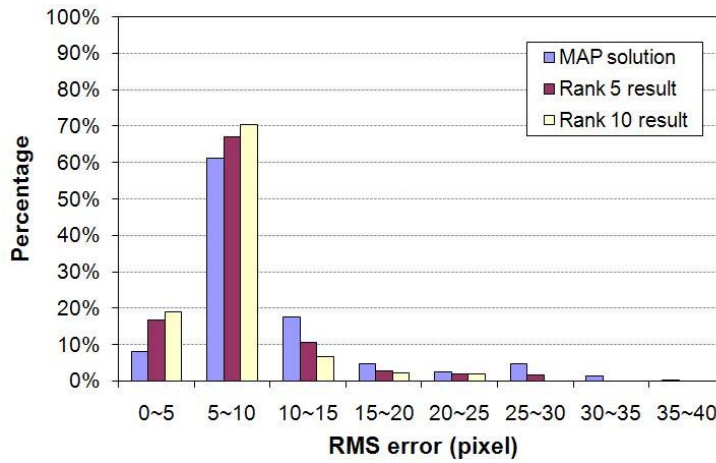


图 6-15 迭代次数为 50 次时的均方根误差的直方图

如图 6-14 所示，可以看出该算法的平均均方根误差随机迭代次数快速减小，然后在 50 次迭代后趋于平稳。另外，在第 50 次迭代后，我们对最大后验估计，Rank5/Rank10 估计进行了直方图统计，情况如图 6-15 所示，横坐标为均方根误差，纵坐标为出现在对应均方根误差上的比例。可以看出，70% 的最大后验估计落到 0-10 像素的误差范围内（对于 Rank10 估计，这个值为 90%）。三种估计的平均均方根误差分别为 10.02，8.04，7.34，图 6-15 显示出我们的算法在各类图像上的稳定性。与此同时，我们将该算法与 Gang 等的工作 [84]

和 Zhang 等 [109] 的工作在平均运行时间上进行了比较, 比较结果见表 6-1。可以看出, 我们算法的平均运行时间为 1 分钟, 效率远高于另外两种方法。另外, 我们还将我们的算法与基于蒙特卡罗的算法进行比较, 结果表明我们的策略比蒙特卡罗的搜索在效率上提高了近 10 倍。

6.4.4 语法分析

在这一部分中, 我们首先设计以下四种算法: 1) 基于蒙特卡罗的搜索方法 (MCSS: Monte Carlo sampling based searching method); 2) MCSS + 语法一 (树解析模型只基于语法一); 3) MCSS + 语法一 + 语法二 (树解析模型基于语法一和语法二); 4) 我们的算法。然后对这几种算法在自拍数据上进行测试, 实验结果如图 6-16 所示, 其中 MCSS + 语法一取得了与 MCSS 差不多的估计精度, 而且加入语法二和语法四分别使得算法性能得到大幅度的提升。

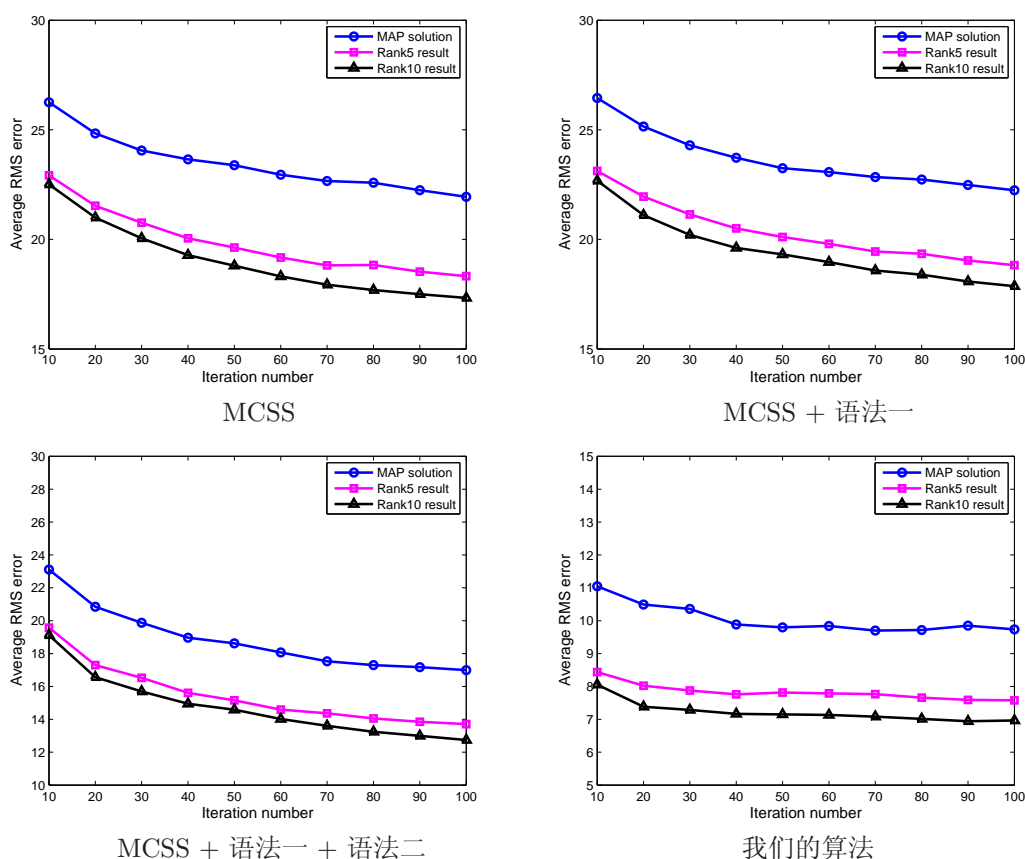


图 6-16 不同语法对算法性能的影响

下面我们来仔细地分析一下几种语法对算法性能的影响以及其原因，首先来看图 6-16 的第一行结果，我们发现算法 MCSS 与算法 MCSS + 语法一的性能几乎没有差别，其原因是由于背景减除和图像分割所产生的噪声（见 §6.4.5），使得语法一中的基于数据驱动的重要性建议分布往往不够鲁棒，另外，语法一并没有考虑人体间的拓扑约束，因此仅通过语法一来对树模型进行解析并不合理，从而不能提高估计性能。接着我们来看图 6-16 的第二行结果，可以看出算法 MCSS + 语法一 + 语法二在估计精度上面比算法 MCSS 和算法 MCSS + 语法一有了很大程度上的提高，这是因为语法二考虑了人体间的拓扑关系，并且语法一和语法二之间可以进行互补，弥补了各自的缺陷，从而为解析树模型提供了有效的信息。然而，如果树解析算法仅仅依据语法一和语法二，也会在某些情况下陷入麻烦，例如，当选定人体部分的前景特征由于噪声原因显示为背景或者该人体部分被其他人体部分遮挡，在这些情况下，算法将在该人体部分上陷入死循环，因为它的综合优先权 $P_{priority}$ 远高于其他的人体部分。正是由于上述原因，算法 MCSS + 语法一 + 语法二的收敛速度比较慢。最后来看我们的算法的性能，我们的算法不仅仅在估计精度上比算法 MCSS + 语法一 + 语法二有了大幅度的提升，而且算法收敛速度也要远远快于其他三种算法。从图 6-16 可以看出，误差曲线在 50 次迭代后开始趋于平稳，而且它们几种算法需要将近 500 次迭代才能收敛。这是因为语法四设计了合理的优先概率传播机制，从而协助语法一和语法二完成算法的快速收敛。

6.4.5 噪声输入下的性能

下面的部分给出了一些噪声情况下的估计例子和分析。如图 6-17 所示，当输入的边缘不够精确时，我们的算法也能取得不错的估计结果。原因是由于我们对边缘似然加入了衰减因子，使得估计结果对边缘似然具有更大的容错能力。另外，我们精心设计的初始化过程也在一定程度上提高算法对噪声的抵抗能力。

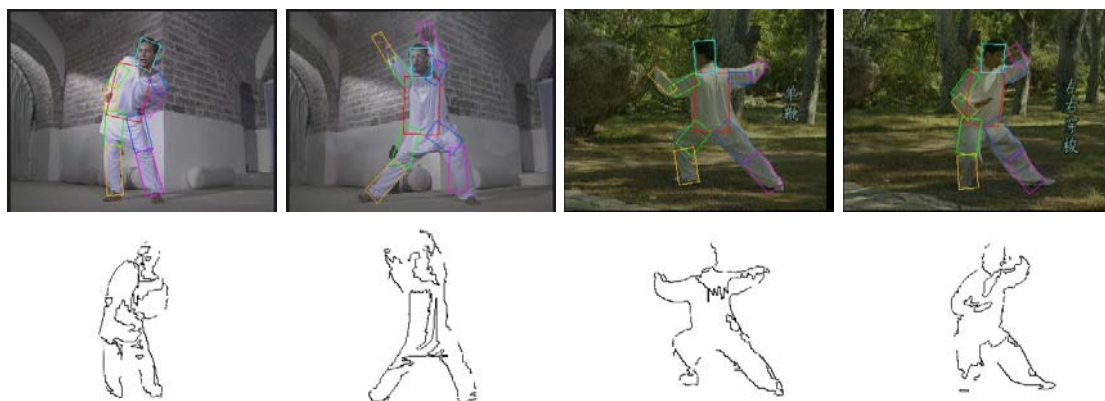


图 6-17 边缘噪声下的姿态估计结果

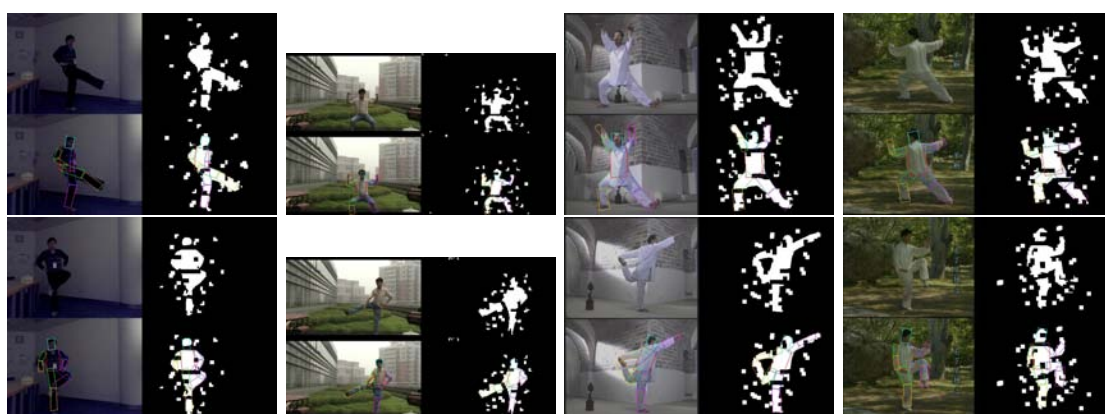


图 6-18 前景噪声下的姿态估计结果

如图 6-18 所示，在噪声前景的前提下，我们的算法同样取得了令人满意的结果。主要原因如下：

- ① 该算法首先要对前景输入进行预处理，从而可以将一些小的噪声去掉；
- ② 基于跳跃的搜索机制可以跳出一些局部极大值（通常由前景噪声引起的）；
- ③ 我们将人体的拓扑约束加入到了算法的优化过程中，因此，即使某个人体部分的特征被噪声严重影响或者被其他人体部分遮挡了，也可以通过其母节点或者子节点的约束来获得一个合理的估计。

Parameter	Group 1	Group 2	Group 3	Group 4	Group 5
α	0.6	0.66	0.50	0.57	0.52
β	0.5	0.57	0.46	0.45	0.42
m	0.01	0.011	0.009	0.011	0.008
w	0.2	0.19	0.20	0.19	0.23
λ	4	4.0	4.8	3.6	4.2
Average RMS error	10.01	9.51	10.62	10.05	10.33
Average running time	64	60	59	63	70

表 6-2 不同参数下的算法性能

6.4.6 参数对算法性能的影响

如 §6.4 所述，我们的算法中主要有以下参数： α ， β ， w ， m ， λ ，并且在本文的实验中设置如下： $\alpha = 0.6$ ， $\beta = 0.5$ ， $w = 0.2$ ， $m = 0.01$ ， $\lambda = 2, 2, 5, 5, 4$ （分别对应人体各个部分状态的 5 个分量）。

由于参数较多，如果对上述参数挨个进行分析，将会非常繁琐，而且也没有这样的必要。因此，我们采样如下策略来对参数进行分析：将上述所有参数在其 20% 的范围内进行扰动，然后对新参数下的算法性能进行测试，包括姿态估计的均方根误差以及算法的平均运行时间。上述过程重复进行 100 次，然后将参数对算法性能的影响进行统计。表 6-2 显示了其中 5 组参数下的算法性能。根据 100 组不同参数下的算法性能，我们可以得出如下结论：当参数

	Average RMS error	Average running time
$m=0.01, w=0.2$	10.01	64
$m=0.1, w=0.4$	10.84	67

表 6-3 不同 k, w 下的算法性能

在 20% 范围内变化时，算法性能并没有发生显著变化，变化幅度仅在 10% 以内（包括姿态估计的均方根误差和算法的平均运行时间）。另外，对于一些非核心的参数，即使变化幅度超过 20%，也不会对算法性能产生显著影响，如表 6-3 所示，当参数 m 从 0.01 变为 0.1，参数 w 从 0.2 变为 0.4，算法性能的变化幅度依旧在 10% 以内。从上述实验可以看出，我们算法对参数变化非常鲁棒。

第七章 结束语

7.1 工作总结

目标跟踪在目标运动的视觉分析中占有重要的地位，属于视觉的中层部分。利用目标的跟踪，可以方便地获得目标的运动参数，为后续高层的行为理解和识别奠定了基础。目标跟踪虽然是计算机视觉领域一个重要的研究方向和研究热点，但是目前仍然有很多理论与技术问题有待解决，特别是跟踪过程中噪声干扰、剧烈运动、光照变化、遮挡等复杂问题的解决。目标的姿态估计在目标运动的视觉分析中属于高层部分，它是根据图像信息推断出目标物体（主要针对人体，手势）的三维方位指向。姿态估计在智能视觉监控、人机交互和视频运动分析等很多领域都有着重要的应用。

本文的工作以复杂场景下鲁棒的目标跟踪与姿态估计为目标，分别在下面三个问题上进行了深入的探讨和分析：（1）基于样本分布的表观建模；（2）跟踪过程中样本的退化；（3）姿态估计中高维状态空间的解析。大量的实验表明我们方法的有效性和鲁棒性。论文的主要工作和贡献如下：

- ① 提出了一种基于图嵌入判别学习的框架来学习出正样本的模式空间，以及正负样本之间的判别空间。基于这两个模式空间的学习，我们构建起鲁棒的跟踪观测模型。与传统基于子空间学习的方法相比，该方法有效地嵌入了样本的分布特性，能够得到更符合实际的目标模式子空间，因此对噪声、局部遮挡更为鲁棒。
- ② 提出两种结合确定性跟踪框架和随机性跟踪框架的方法：（1）核贝叶斯跟踪框架；（2）基于奇异值分解的卡尔曼粒子滤波框架。（1）首先将核方法应用于当前帧，获得对运动方向以及偏移量的估计，再将这些信息作为启发式先验信息加入到贝叶斯滤波框架中，从而有效结合确定性方法与

随机性方法的优点，并且该框架是一种通用的跟踪框架，任何形式的表观模型都可以在此框架下应用。与基于核方法的跟踪器（确定性跟踪方法）以及基于贝叶斯方法的跟踪器（随机性跟踪方法）相比，该跟踪框架有效地提高了跟踪的效率和精确度。（2）首先对前一帧跟踪样本的协方差矩阵进行奇异值分解，再利用 Unscented 变化（一种确定性的变化）获得预测样本，然后用卡尔曼滤波器对预测样本进行滤波，为最终的粒子滤波提供运动估计分布。与粒子滤波和基于 Unscented 的卡尔曼粒子滤波相比，我们的方法更为鲁棒。

③ 提出了一种基于序列粒子群优化的跟踪算法框架。该框架首先通过初始化一系列随机粒子样本，然后对这些粒子集进行时序上的传播以及优化迭代，最后对进化的粒子进行收敛性判断。我们从理论上证明了，序列粒子群优化框架实质上是一个基于分步重要性采样的粒子滤波过程，使得采样结果逼近了直接采样于最佳重要性分布的结果。我们还将序列粒子群优化的概念扩展到多种群上，将多目标跟踪中目标物体的交互和遮挡用种群间的竞争和排斥来模拟，实现了有效的多目标跟踪。与其他两种多目标跟踪的算法相比，我们的算法在 PETS2004 数据库上获得了最佳的实验结果。

④ 我们提出一种新颖的姿态估计算法，该方法从以下三个方面出发来解决姿态估计所面临的维数灾难问题：1）采用树形结构来对人体拓扑结构进行建模，并通过图像语法对其进行解析；2）优化过程采用两种模式：扩散和跳跃；3）提出一种基于数据驱动的马尔可夫链蒙特卡洛方法来进行优化。该方法可以有效地将人体各部分的拓扑关系加入到优化过程中，并通过表观数据的驱动来压缩搜索空间，从而大大提高了收敛速度（10 倍左右）和精度。

7.2 未来展望

本文主要着眼于复杂场景下鲁棒的目标跟踪与姿态估计，探讨了不同表观模型和跟踪框架下的跟踪算法和姿态估计中高维状态空间的解析问题，取得了一些研究成果，但这些方法或多或少存在一些局限性，因此仍有许多工作需要进一步研究：

- ① 我们的跟踪框架都是基于二维表观模型的，而没有考虑目标的三维信息。

如果目标在较长时间内受到严重的遮挡，那么我们的跟踪框架很难进行目标的鲁棒跟踪。因此，充分利用目标的三维信息来进行目标表观建模是我们未来的工作。比如，我们可以采用多摄像机相互协同配合的策略来捕获目标的三维信息。当然也可以对目标所在的场景进行摄像机标定，从而得到目标的三维形状信息。

- ② 在姿态估计中，由于状态空间维数过高并且受噪声影响，从而导致了极具挑战性的问题：图像观测似然函数的最优值与实际的真实姿态并不符合。

导致这种现象的本质原因还是因为所采用表观模型不够鲁棒。因此，除了前景信息和边缘信息之外，我们需要将其他更多的信息，例如运动信息，形状信息加入到模型中，使得模型更加具有判别性，从而使得图像观测似然函数的最优值与 groundtruth 具有更好的一致性。

经过了几十年的发展，运动目标跟踪和姿态估计已经取得了很大发展和诸多成果，目前已在多个领域展开了初步的应用。建立一个稳健的实时跟踪和姿态估计系统所需要解决的问题是复杂的、困难的。但是，我们相信，随着计算机视觉、模式识别相关技术的发展，这些问题可以被逐步地解决。

参 考 文 献

- [1] D.M. Gavrilu, “The Visual Analysis of Human Movement: A Survey”, CVIU, vol.73, no.1, pp.82-98, 1999.
- [2] T.B. Moeslund and E. Granum, “A Survey of Computer Vision-Based Human Motion Capture”, CVIU, vol.81, no.3, pp.231-268, 2001.
- [3] J.K. Aggarwal and Q. Cai, “Human Motion Analysis: A Review”, CVIU, vol.73, no.3, pp.428-440, 1999.
- [4] L. Wang, W. Hu and T. Tan, “Recent Development of Human Motion Analysis”, Pattern Recognition, vol.36, no.3, pp.585-601, 2003.
- [5] W.M. Hu, T.N. Tan, L. Wang and S.J. Maybank, “A Survey on Visual Surveillance of Object Motion and Behaviors”, IEEE Trans.on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol.34, no.3, pp.334-352, 2004.
- [6] A. Baumberg and D.C. Hogg, “Learning Deformable Models for Tracking the Human Body, In Motion-Based Recognition”, Kluwer Academic, pp.39-60, 1996.
- [7] M. Haag and H.H. Nagel, “Combination of Edge Element and Optical Flow Estimates for 3D Model-Based Vehicle Tracking in Traffic Image Sequences”, IJCV, pp.295-319, 1999.
- [8] EVI D30/30 Specification. From <http://www.sony.com>, 2002.

- [9] S.E. Kemeny, R. Panicacci, B. Pain, L. Matthies and E.R. Fossum, “Multiresolution Image Sensor”, IEEE Trans.on Circuits and Systems for Video Technology, vol.7, no.4, pp.575-583, 1997.
- [10] Y. Onoe, N. Yokoya, K. Yamazawa and H. Takemura, “Visual Surveillance and Monitoring System Using an Omnidirectional Video Camera”, ICPR, 1998.
- [11] K. Nummiaroa, E. Koller-Meierb, and L.V. Gool, “An adaptive color-based particle filter,” Image and Vision Computing, 21:pp.99-110, 2003.
- [12] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, “Color-based probabilistic tracking,” in Proc. ECCV, pp.661-675, 2002.
- [13] A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, “Background and foreground modeling using nonparametric kernel density estimation for visual surveillance,” Proceedings of IEEE, Iss. 7, Vol. 90, pp.1151-1163, 2002.
- [14] C. Yang, R. Duraiswami, and L.S. Davis, “Efficient mean-shift tracking via a new similarity measure,” in Proc. CVPR, pp.176-183, 2005.
- [15] S. J. McKenna, Y. Rajab, and S. Gong, “Tracking colour objects using adaptive mixture models,” Image and Vision Computing, 17:pp.225-231, 1999.
- [16] B. Han, and L. Davis, “On-line density-based appearance modeling for object tracking,” in Proc. ICCV, pp.1492-1499, 2005.
- [17] Y. Wu and T. S. Huang, “Robust visual tracking by integrating multiple cues based on co-inference learning,” IJCV, vol. 58, no. 1, pp.55-71, 2004.

- [18] S. Khan and M. Shah, “Tracking people in presence of occlusion,” in Proc. ACCV, pp.263-266, 2000.
- [19] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, “Robust Online Appearance Models for Visual Tracking,” in Proc. CVPR’01, Vol. 1, pp.415-422, 2001.
- [20] S. K. Zhou, R. Chellappa, and B. Moghaddam, “Visual Tracking and Recognition Using Appearance-Adaptive Models in Particle Filters,” IEEE Trans. on Image Processing, Vol. 13, pp.1491-1506 , November 2004.
- [21] C. Stauffer, and W.E.L. Grimson, “Adaptive Background Mixture Models for Real-Time Tracking,” in Proc.CVPR’99, Vol. 2, 1999.
- [22] H. Wang, D. Suter, and K. Schindler, “Effective appearance model and similarity measure for particle filtering and visual tracking,” in Proc. ECCV, pp. 606-618, 2006.
- [23] Y. Wang, K. Loe, and J. Wu, “A Dynamic Conditional Random Field Model for Foreground and Shadow Segmentation ,” IEEE Trans. PAMI., Vol. 28, Iss. 2, pp.279-289, 2006.
- [24] Y. Li, “On Incremental and Robust Subspace Learning,” Pattern Recognition, Vol. 37, Iss. 7, pp.1509-1518, 2004.
- [25] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, “Two-dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition,” in IEEE Trans. PAMI., Vol. 26, Iss. 1, pp.131-137, Jan. 2004.
- [26] J. Ye, R. Janardan, and Q. Li, “Two-Dimensional Linear Discriminant Analysis,” NIPS’04, pp.1569-1576, MIT Press,2004.

- [27] J. Ye, “Generalized low rank approximations of matrices,” ICML’04, July 2004.
- [28] J. Ye, R. Janardan, and Q. Li, “GPCA: An Efficient Dimension Reduction Scheme for Image Compression and Retrieval,” ACM KDD’04, pp.354-363, August 2004.
- [29] H. Wang and N. Ahuja, “Rank-R Approximation of Tensors Using Image-as-matrix Representation ,” in Proc. CVPR’05, Vol. 2, pp.346- 353, 2005.
- [30] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang and H. Zhang, “Discriminant analysis with tensor representation,” in Proc. CVPR’05, Vol. 1, pp.526-532, June 2005.
- [31] M. A. O. Vasilescu and D. Terzopoulos, “Multilinear Subspace Analysis of Image Ensembles,” in Proc. CVPR’03, Vol. 2, pp.93-99, June 2003.
- [32] M. A. O. Vasilescu and D. Terzopoulos, “Multilinear Subspace Analysis of Image Ensembles: TensorFaces,” in Proc. ECCV’02, pp.447- 460, May 2002.
- [33] X. He, D. Cai and P. Niyogi, “Tensor Subspace Analysis,” NIPS’05, Dec. 2005.
- [34] H. Wang, S. Yan, T. Huang and X. Tang, “A Convergent Solution to Tensor Subspace Learning,” in Proc. IJCAI’07, 2007.
- [35] J. Sun, D. Tao and C. Faloutsos, “Beyond Streams and Graphs: Dynamic Tensor Analysis,” ACM KDD’06, Aug. 2006.
- [36] J. Sun, S. Papadimitriou and P. S. Yu, “Window-based Tensor Analysis on High-dimensional and Multi-aspect Streams,” in Proc.ICDM’06, Dec. 2006.

- [37] G. D. Hager, P. N. Hager, “Efficient region tracking with parametric models of geometry and illumination”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(10): 1025-1039, 1998.
- [38] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-based object tracking”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5): 234-240, 2003.
- [39] A. Elgammal, R. Duraiswami, and L. “Davis. Probabilistic tracking in joint feature-spatial spaces”, In Proc. IEEE Conf. Computer Vision and Pattern Recognition, volume I, pages 781–788, Madison, WI, 2003
- [40] R. T. Collins, “Mean-shift blob tracking through scale space”, Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, ,Wisconsin,USA,2003,2:234-240.
- [41] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, “Robust online appearance models for visual tracking”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(10): 1296-1311, 2003.
- [42] J. Lim, D. Ross, R. S. Lin, and M. H. Yang, “Incremental learning for visual tracking”, In Advances in Neural Information Processing Systems, pp.793-800, 2004, The MIT Press.
- [43] C. Shen, M. J. Brooks, A. van den Hengel, “Fast Global Kernel Density Mode Seeking: Applications to Localization and Tracking”, vol. 16, no. 5, pp. 1457-1469, 2007.
- [44] J.Rocha, A.Mir, “Articulated Object Tracking via a Genetic Algorithm”, Proceedings of the Third International Workshop on Energy Minimization

- Methods in Computer Vision and Pattern Recognition Pages: 134 - 152
,2001
- [45] S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking", IEEE Trans. on Signal Processing, 50(2), pp. 174-188, 2002.
- [46] Y. Bar-Shalom and T.E. Fortmann, "Tracking and Data Association", Academic Press, New York, 1988.
- [47] S. Julier, D. Fleet and E. Maraghi, "A New Extension of the Kalman Filter to Nonlinear System", Proceedings of SPIE, pp.182-193, 1997.
- [48] P. Li, T. Zhang, B. Ma, "Unscented Kalman Filter for Visual Curve Tracking", Image and Vision Computing, vol.22, no.21, pp.157-164, 2004.
- [49] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, vol.77, no.2, pp.257-285, 1989.
- [50] D. Salmond, N. Gordon and A. Smith, "A Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation", IEEE Proceedings on Radar, Sonar and Navigation, vol.140, no.2, pp.107-113, 1993.
- [51] M. Isard and A. Blake, "Condensation-Conditional Density Propagation for Visual Tracking", IJCV, vol.29, no.1, pp.5-28, 1998.
- [52] C. Hue, J. Cader, P. Perez, "Tracking Multiple Objects with Particle Filtering", IEEE Transactions on Aerospace and Electronic Systems, vol.38, no.3, pp.791-812, 2002.

- [53] R. Merwe, A. Doucet, N. Freitas, E. Wan, “The Unscented Particle Filter”, Technical Report, Engineering Department, Cambridge University, 2000.
- [54] A. Mittal, L. Zhao and L. Davis, “Human body pose estimation using silhouette shape analysis,” Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 263–270, 2003.
- [55] J. C. Niebles and L. Fei-Fei, “A hierarchical model model of shape and appearance for human action classification,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.
- [56] N. R. Howe, M. E. Leventon, W. T. Freeman, “Bayesian Reconstruction of 3D Human Motion from Single-Camera Video,” Proceedings of Advances in Neural Information Processing Systems, 1999.
- [57] T. J. Roberts, S. J. Mckenna, and I. W. Ricketts, “Human Pose Estimation Using Partial Configurations and Probabilistic Regions,” International Journal of Computer Vision, vol. 73, no. 3, pp. 285-306, 2007.
- [58] X. Lan and D. P. Huttenlocher, “Beyond Trees: Common Factor Models for 2D Human Pose Recovery,” Proceedings of IEEE International Conference on Computer Vision, vol. 1, pp. 470-477, 2005.
- [59] R. Poppe, “Vision-based human motion analysis: An overview,” Computer Vision and Image Understanding, vol.108, no. (1-2), pp. 4-18, 2007.
- [60] G. Mori and J. Malik, “Estimating Human Body Configurations Using Shape Context Matching,” Proceedings of European Conference on Computer Vision, vol. 3, pp. 666-680, 2002.

- [61] S. Belongie, J. Malik, and J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 4, pp. 509-522, 2002.
- [62] G. Shakhnarovich, P. Viola, and T. Darrell, “Fast Pose Estimation with Parameter Sensitive Hashing,” *Proceedings of IEEE International Conference on Computer Vision*, pp. 750-757, 2003.
- [63] A. Fathi and G. Mori, “Human pose estimation using motion exemplars”, *Proceedings of IEEE Conference on Computer Vision*, 2007.
- [64] R. Poppe, “Evaluating example-based pose estimation: Experiments on the HumanEva sets”, In *Proceedings of the Workshop on Evaluation of Articulated Human Motion and Pose Estimation*, 2007.
- [65] M. Van den Bergh, E. Meier and L. Van Gool, “Real-time body pose recognition using 2D or 3D haarlets”, *International Journal of Computer Vision*, vol. 83, no. 1, pp. 72-84, 2009.
- [66] E. Ong, A. S. Micilotta, R. Bowden and A. Hilton, “Viewpoint invariant exemplar-based 3D human tracking”, *Computer Vision and Image Understanding*, vol. 104, no. (2-3), pp. 178-189, 2006.
- [67] D. Gavrilu, “A Bayesian, exemplar-based approach to hierarchical shape matching”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1408-1421, 2007.
- [68] D. Ramanan, “Learning to Parse Images of Articulated Bodies,” *Proceedings of Advances in Neural Information Processing Systems*, 2006.

- [69] D. Ramanan, D. Forsyth, and A. Zisserman, “Strike a Pose: Tracking People by Finding Stylized Poses,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 271-278, 2005.
- [70] P. F. Felzenszwalb, D. P. Huttenlocher, “Pictorial Structures for Object Recognition,” *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55-79, 2005.
- [71] J. Gall, B. Rosenhahn, T. Brox and H. Seidel, “Optimization and filtering for human motion capture: A Multi-Layer Framework”, *International Journal of Computer Vision*, vol. 87, no. (1-2), pp. 75-92, 2009.
- [72] Q. Delamarre and O. Faugeras, “3D articulated models and multiview tracking with physical forces”, *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 328-357, 2001.
- [73] D. Gavrilu and L. Davis, “Tracking of humans in action: A 3D model-based approach”, In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 73-80, 1996.
- [74] V. Ferrari, M. Jimenez and A. Zisserman, “Progressive search space reduction for human pose estimation”, In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.
- [75] P. Srinivasan, J. Shi, “Bottom-up Recognition and Parsing of the Human Body”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

- [76] R. Ronfard, C. Schmid, and B. Triggs, “Learning to Parse Pictures of People,” Proceedings of European Conference on Computer Vision, pp. 700-714, 2002.
- [77] X. Ren, A. C. Berg, and J. Malik, “Recovering Human Body Configurations Using Pairwise Constraints between Parts,” Proceedings of IEEE International Conference on Computer Vision, vol. 1, pp. 824-831, 2005.
- [78] G. Mori, X. Ren, A. Efros, and J. Malik, “Recovering Human Body Configurations: Combining Segmentation and Recognition,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 326-333, 2004.
- [79] P. Kuo, D. Makris, N. Megherbi and J. Nebel, “Integration of local image cues for probabilistic 2D pose recovery”, In Proceedings of the International Symposium on Advances in Visual Computing, pp. 214-223, 2008.
- [80] B. Daubney, D. Gibson and N. Campbell, “Real-time pose estimation of articulated objects using low-level motion”, In Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.
- [81] Y. Wang and G. Mori, “Multiple tree models for occlusion and spatial constraints in human pose estimation”, In Proceedings of the European Conference on Computer Vision, pp. 710-724, 2008.
- [82] L. Sigal and M. J. Black, “Measure Locally, Reason Globally: Occlusion-Sensitive Articulated Pose Estimation,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2041-2048, 2006.

- [83] A. Gupta, A. Mittal and L. Davis, “Constraint integration for efficient multiview pose estimation with self-occlusions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 493-506, 2008.
- [84] G. Hua, M. Yang, Y. Wu, “Learning to Estimate Human Pose with Data Driven Belief Propagation,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 747-754, 2005.
- [85] J. Zhang, J. Luo, R. Collins, and Y. Liu, “Body Localization in Still Images Using Hierarchical Models and Hybrid Search,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1536-1543, 2006.
- [86] M. Lee and I. Cohen, “A Model-Based Approach for Estimating Human 3D Poses in Static Images,” IEEE transactions on pattern analysis and machine intelligence, vol. 28, no. 6, pp. 905-916, 2006.
- [87] R. Navaratnam, A. Thayananthan, P. Torr and R. Cipolla, “Hierarchical part-based human body pose estimation”, In Proceedings of the British Machine Vision Conference, 2005.
- [88] G. Hua and Y. Wu, “A decentralized probabilistic approach to articulated body tracking”, Computer Vision and Image Understanding, vol. 108, no. 3, pp. 272-283, 2007.
- [89] P. Kohli, J. Rihan, M. Bray and P. Torr, “Simultaneous segmentation and pose estimation of humans using dynamic graph cuts”, International Journal of Computer Vision, vol. 79, no. 3, pp. 285-298, 2009.

- [90] S. Yan, D. Xu, B. Zhang and H. Zhang, “Graph Embedding: A General Framework for Dimensionality Reduction”, In Proceeding of International Conference on Computer Vision and Pattern Recognition, vol.2, pp.830-837, 2005.
- [91] D. Ross, J. Lim, R. S. Lin, and M. H. Yang, “Incremental learning for visual tracking”, International Journal of Computer Vision, vol. 77, no. 1, pp. 125-141, 2008.
- [92] R. S. Lin, D. Ross, J. Lim, and M. H. Yang, “Adaptive discriminative generative model and its applications”, In Advances in Neural Information Processing Systems, pp.801-808, 2004, The MIT Press.
- [93] A. Levy and M. Lindenbaum, “Sequential Karhunen-Loeve Basis Extraction and its Application to Images”, IEEE Transactions on Image Processing, 9(8): 1371-1374, 2000.
- [94] M. Sugiyama, “Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction”, Proceedings of International Conference on Machine Learning, pp.905-912, 2006.
- [95] Z. Manor and P. Perona, “Self-tuning spectral clustering”, In Advances in Neural Information Processing Systems, pp.1601-1608, 2004, The MIT Press.
- [96] G. D. Hager, P. N. Hager, “Efficient region tracking with parametric models of geometry and illumination”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(10): 1025-1039, 1998.

- [97] K. Fukunaga, L. Hostetler, “The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition”, IEEE Transactions on Information Theory, 21(1): 32-40, 1975.
- [98] V. Parameswaran, V. Ramesh, I. Zoghlami, “Tunable Kernels for Tracking”, In Proceeding of International Conference on Computer Vision and Pattern Recognition, pp. 2179-2186, 2006.
- [99] G. H. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, 1996.
- [100] 江明, “微粒群优化算法的随机收敛性分析及其应用研究”, 清华大学工学博士学位论文, 2007.
- [101] F. Heppner and U. Grenander, “A stochastic nonlinear model for coordinated bird flocks”, In: The ubiquity of Chaos. AAAS publications, 1990.
- [102] J. Kennedy, and R. C. Eberhart, “Particle swarm optimization”, in Proc. IEEE Int’l Conf. on Neural Networks, pp. 1942-1948, 1995.
- [103] Y. Shi and R. C. Eberhart, “A modified particle swarm optimizer”, In: Proceedings of the IEEE Congress on Evolutionary Computation, pp. 69-73, 1998.
- [104] M. Clerc, and J. Kennedy, “The particle swarm-explosion, stability, and convergence in a multidimensional complex space”, IEEE Trans. on Evolutionary Computation, 6(1): 58-73, 2002.
- [105] A. Doucet, S. Godsill, and C. Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering”, Statistics and Computing, 10(3): 197-208, 2000.

- [106] W. Qu, D. Schonfeld, and M. Mohamed, “Real-time distributed multi-object tracking using multiple interactive trackers and a magnetic-inertia potential model”, IEEE Trans.on Multimedia, vol. 9, no. 3, pp. 511-519, 2007.
- [107] M. Yang, T. Yu, and Y. Wu, “Game-Theoretic Multiple Target Tracking”, In: Proc. of International Conference on Computer Vision, pp. 1-8, 2007.
- [108] L. Ingber, “Simulated Annealing: Practice Versus Theory”, Journal of Mathematical and Computer Modeling 18(11), 29-57, 1993.
- [109] J. Zhang, J. Luo, R. Collins, and Y. Liu, “Body Localization in Still Images Using Hierarchical Models and Hybrid Search,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1536-1543, 2006.
- [110] T. Zhao, R. Nevatia, “Bayesian Human Segmentation in Crowded Situations,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 459-466, 2003.

攻读博士期间完成的论文和专利申请

图书章节

1. Qingdi Wei, **Xiaoqin Zhang**, Weiming Hu, “Action Recognition”, in Machine Learning for Human Motion Analysis: Theory and Practice, 2009.

杂志论文

1. **Xiaoqin Zhang**, Weiming Hu, Steve Maybank, “Sequential Particle Swarm Optimization for Visual Tracking”, submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence.
2. **Xiaoqin Zhang**, Changcheng Li, Weiming Hu, Xiaofeng Tong and Steve Maybank, “Efficient Human Pose Estimation and Tracking via Parsing a Tree Structure Based Human Model”, submitted to International Journal of Computer Vision.
3. **Xiaoqin Zhang**, Xiaoqin Zhang, Weiming Hu, Steve Maybank, “Multi-Object Tracking via Species Based Particle Swarm Optimization”, submitted to IEEE Transactions on Circuits and Systems for Video Technology.

会议论文

1. **Xiaoqin Zhang**, Changcheng Li, Xiaofeng Tong, and Weiming Hu, “Efficient Human Pose Estimation via Parsing a Tree Structure Based Human Model”, IEEE International Conference on Computer Vision (ICCV), 2009.

2. **Xiaoqin Zhang**, Weiming Hu, Steve Maybank, and Wei Qu, “Multi-Object Tracking via Species Based Particle Swarm Optimization”, International Workshop on Visual Surveillance conjugated with International Conference on Computer Vision (**VS**), 2009
3. **Xiaoqin Zhang**, Weiming Hu, Steve Maybank, “A Smarter Particle Filter”, Asia Conference on Computer Vision (**ACCV**), 2009.
4. **Xiaoqin Zhang**, Xianglin Zeng, Weiming Hu, Wanqing Li, “Graph Cut Based Coherence Measure Construction for Scene Segmentation”, Pacific-Rim Conference on Multimedia, (**PCM**), 2009.
5. Yu Kong, Weiming Hu, **Xiaoqin Zhang**, Hanzi Wang, Yunde Jia, “Learning Group Activity in Soccer Videos from Local Motion ”, Asia Conference on Computer Vision (**ACCV**), 2009.
6. Wei Li, **Xiaoqin Zhang**, and Weiming Hu, “Contour Tracking With Abrupt Motion”, IEEE International Conference on Image Processing (**ICIP**) , 2009.
7. **Xiaoqin Zhang**, Weiming Hu, Steve Maybank, Xi Li, “Sequential Particle Swarm Optimization for Visual Tracking”, IEEE International Conference on Computer Vision and Pattern Recognition (**CVPR**), 2008
8. Xi Li, Weiming Hu, Zhongfei Zhang, **Xiaoqin Zhang**, Guan Luo, “Visual Tracking Via Incremental Log-Euclidean Riemannian Subspace Learning”, IEEE International Conference on Computer Vision and Pattern Recognition (**CVPR**), 2008.
9. **Xiaoqin Zhang**, Weiming Hu, “SVD Based Kalman Particle Filter for

- Robust Visual Tracking”, International Conference on Pattern Recognition (**ICPR**), 2008.
10. Yu Kong, **Xiaoqin Zhang**, Qingdi Wei, Weiming Hu, Yunde Jia, “Group Action Recognition in Soccer Videos”, International Conference on Pattern Recognition (**ICPR**), 2008.
 11. Xi Li, Weiming Hu, Zhongfei Zhang, **Xiaoqin Zhang**, “Robust visual tracking Based on An Effective Appearance Model”, European Conference on Computer Vision (**ECCV**), 2008.
 12. **Xiaoqin Zhang**, Weiming Hu, Steve Maybank, and Xi Li, “Graph based discriminative learning for robust and efficient object tracking”, IEEE International Conference on Computer Vision (**ICCV**) **Oral presentation, acceptance rate is 3.9%** , 2007.
 13. Xi Li, Weiming Hu, Zhongfei Zhang, **Xiaoqin Zhang**, and Guan Luo, “Robust Visual Tracking Based on Incremental Tensor Subspace Learning”, IEEE International Conference on Computer Vision (**ICCV**) , 2007.
 14. **Xiaoqin Zhang**, Weiming Hu, Guan Luo and Steve Maybank, “Kernel-Bayesian Framework for Object Tracking”, Asia Conference on Computer Vision (**ACCV**), Part I, LNCS 4843, pp.821-831, 2007.
 15. **Xiaoqin Zhang**, Zhiyong Liu, Hong Qiao, “A Robust Multiple Cues Fusion based Bayesian Tracker”, IEEE International Conference on Robotics and Automation (**ICRA**), pp.4614-4619, 2007.
 16. Qingdi Wei, Weiming Hu, **Xiaoqin Zhang** and Guan Luo, “Dominant Sets-based Action Recognition Using Image Sequence”, IEEE International Con-

ference on image processing (**ICIP**), 2007.

17. **Xiaoqin Zhang**, Hong Qiao, Zhiyong Liu, “Multi-Information Fusion for Scale Selection in Robot Tracking”, IEEE International Conference on Intelligent Robots and Systems (**IROS**), pp.2828-2832, 2006

专利申请

1. 胡卫明, **张笑钦**, 罗文寒, “一种基于序列粒子群的跟踪方法”, 专利申请号: 201010117757.0
2. 胡卫明, 王燕国, **张笑钦**, “基于在线提升算法的网络入侵检测方法”, 专利申请号: 200710304223.7

致 谢

衷心感谢我的导师胡卫明研究员。在我攻读博士学位期间，他在我的学习、生活和科研方面给予了无微不至的关怀和孜孜不倦的教诲。从一进课题组开始，胡老师对学术研究的执着追求和对工作的热情，以及随和的处世态度就深深鼓舞和影响了，使我终身受益。每一篇论文，无论是会议论文还是博士论文，胡老师都会给予悉心的指导和认真的修改。从胡老师身上，我不仅学到了理论知识和实践经验，更重要的是很多做人、做事的道理。

感谢课题组的罗冠和吴偶两位老师。他们在工作上和生活上同样给予了我很多的指导和帮助。与他们一起工作我感到非常的荣幸。

感谢模式识别国家重点实验室生物特征信息处理研究团队的各位老师和同学，是他们共同创造了一个良好的学习氛围。感谢综合办的连国臻、赵微、马丽霞三位老师，她们的辛勤工作是我们科研工作得以顺利进行的可靠保障。感谢研究生处和科技处的邸凌、李磊、毛磊、卜树云、曹娟、王莉、石秋生和韩伟等老师为我们创造更好的科研和生活环境所付出的劳动。

感谢胡老师带领的研究团队中一起工作学习过的各位同学：陈颖、周雪、左海强、朱鹏飞、李玺、魏青帝、李莉、朱明亮、李华北、王燕国、原春锋、谢年华、曾祥林、贺主、孔屿、李威、高君、胡瑞光、刘洋、史信楚、罗文寒等。在与他们的讨论中，得到了很多有意义的启发和灵感。感谢课题组的徐秋艳秘书，我们的生活、学习和科研离不开她的管理和后勤工作。这是一个团结温暖的大家庭，与大家在一起度过了我人生中难忘的五年时光。这一段美好的日子将永远值得回忆。

感谢几年来帮助过我的学长、同学和朋友们：王永庆、常亮、郑碎武、白鹏、雷震、廖胜才、何召锋、董晶等，从他们身上学到了很多可贵的东西。

感谢自动化所的所有老师们和领导这五年来在生活和学习上对我的关心。

致 谢

最后，要特别感谢我最亲爱的父母，在我情绪最低落的时候，给予我极大的支持和鼓励。更要感谢老婆一直以来对我的理解和信任。他们的爱给了我克服困难的勇气和积极进取的热情。

2010 年 4 月