硕 士 研 究 生 ：李志远
指 导 教 师 ：黄凤岗 教授
学 科 、 专 业 ：计算机软件与理论
学位论文主审人 ：李智慧 副教授

哈尔滨工程大学

2009 年 1 月

分类号：_____ 密级：_____

UDC：_____ 编号：_____

工学硕士学位论文

基于交换采样粒子滤波的目标跟踪算法研究

硕士研究生：李志远

指导教师：黄凤岗 教授

学位级别：工学硕士

学科、专业：计算机软件与理论

所在单位：计算机科学与技术学院

论文提交日期：2009年1月

论文答辩日期：2009年2月

学位授予单位：哈尔滨工程大学

Classified Index
U.D.C.

A Dissertation for the Degree of M.Eng

The Algorithm Research of Tracking Based on Switching Particle Filters

Candidate: Li Zhiyuan
Supervisor: Prof. Huang Fenggang
Academic Degree Applied for: Master of Engineering
Speciality: Computer Software and Theory
Date of Submission: January, 2009
Date of Oral Examination: February, 2009
University: Harbin Engineering University

哈尔滨工程大学
学位论文原创性声明

本人郑重声明：本论文的所有工作，是在导师的指导下，由作者本人独立完成的。有关观点、方法、数据和文献的引用已在文中指出，并与参考文献相对应。除文中已注明引用的内容外，本论文不包含任何其他个人或集体已经公开发表的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者（签字）：李志远

日期：2009年2月25日

哈尔滨工程大学
学位论文授权使用声明

本人完全了解学校保护知识产权的有关规定，即研究生在校攻读学位期间论文工作的知识产权属于哈尔滨工程大学。哈尔滨工程大学有权保留并向国家有关部门或机构送交论文的复印件。本人允许哈尔滨工程大学将论文的部分或全部内容编入有关数据库进行检索，可采用影印、缩印或扫描等复制手段保存和汇编本学位论文，可以公布论文的全部内容。同时本人保证毕业后结合学位论文研究课题再撰写的论文一律注明作者第一署名单位为哈尔滨工程大学。涉密学位论文待解密后适用本声明。

本论文（□在授予学位后即可　□在授予学位 12 个月后　□解密后）由哈尔滨工程大学送交有关部门进行保存、汇编等。

作者（签字）：李志远　　导师（签字）：董风南

日期：2009年2月25日　　2009年2月25日

摘 要

视频目标跟踪融合了图像处理、模式识别、人工智能、自动控制、计算机应用、生理学、物理学、数学等相关领域的先进技术和研究成果，具有可视性、抗干扰、成本低等优点，已被广泛地应用到军事监控、交通管制、机器智能、医疗诊断等方面。随着各种复杂环境应用领域的不断需求，除了引入各种新的技术外，如何提高现有跟踪算法的精度以及鲁棒性也是当前目标跟踪领域研究的重点。

本文主要研究视频序列中运动目标跟踪的两个方面。一是运动目标的特征表述，二是运动目标跟踪的粒子滤波框架。

1、针对目前单一特征不能很好地表征目标的问题，本文提出了基于颜色直方图和梯度方向直方图融合的综合直方图来表征运动目标。由于颜色直方图对目标的旋转、形变不敏感，梯度方向直方图对于目标的光照变化不敏感，因此将两者的优点结合起来，形成互补，以提高目标特征的鲁棒性，能够应对复杂背景下的运动目标跟踪。为后面的粒子滤波算法的应用提供了坚实的基础。

2、针对粒子滤波有时无法得到正确的后验概率分布的情况。本文提出了一种新的粒子滤波方法——交换采样粒子滤波。本文将两个互补的采样方法结合起来，Condensation 算法和辅助粒子滤波(APF)。Condensation 更注重状态转移概率的预测，鲁棒性强。而 APF 更侧重于当前状态所得到的观察，精确性高。本文将两者的优点结合起来，在跟踪中根据当前估计状态的信任度在线的交换采样，这样将鲁棒性和精确性结合起来之后，使得算法在有混淆、干扰、遮挡的情况下仍然能很好地跟踪目标。

关键词： 目标跟踪；颜色直方图；梯度直方图；粒子滤波；辅助粒子滤波

Abstract

Moving object tracking in video assembles advanced technologies in many fields such as image processing、pattern recognition、artificial intelligence、automatic control computer application、physiology、physics and mathematics, et al. It has many excellent advantages of visual scene, anti-interference and cost-effective, and has been widely applied to military surveillance, traffic control, machine intelligence, medical diagnosis, and so on. With the continuous demand of complex environment applications, besides introducing new technology, how to improve accuracy and robust of the existing tracking algorithm is the focus of the current work.

The research in this thesis focuses on two aspects of the video tracking. One is the feature description of the object. The other is the framework based visual tracking, particle filter.

Firstly, In order to deal with the robust feature description of object, this paper proposes a new histogram-HOGC.HOGC combines the color histogram and grade histogram. The two histograms can complement to each other. The color histogram is unaware to rotation while the grade one is unaware to color. So HOGC is a robust feature of object in visual tracking.

Secondly, this thesis proposes a new particle filtering scheme, called a switching particle filter, which allows robust and accurate visual tracking under typical circumstances of real-time visual tracking. This scheme switches two complementary sampling algorithms, Condensation and Auxiliary Particle Filter, in an on-line fashion based on the variance of the current estimated target state. The accuracy and robustness of the switching scheme were evaluated using real visual tracking experiments.

Key words: Visual tracking; Color histogram; Grade histogram; Condensation; Auxiliary Particle Filter

目 录

第 1 章 绪论……1
1.1 研究背景……1
1.2 视觉目标跟踪研究概述……2
1.2.1 视觉跟踪算法的分类……2
1.2.2 处理视觉跟踪问题的两个思路……4
1.2.3 视觉跟踪中的难点……7
1.3 本文的研究工作及创新点……9
1.4 本文的组织结构……9
第 2 章 目标表观的 HOGC 特征提取……11
2.1 引言……11
2.2 目标表观的建模和提取……12
2.2.1 基于颜色分布的方法……12
2.2.2 基于模板的方法……14
2.2.3 基于轮廓的方法……15
2.2.4 基于子空间的方法……17
2.3 HOGC 特征……18
2.3.1 梯度方向直方图……19
2.3.2 颜色直方图……20
2.3.3 综合直方图（HOGC）……20
2.3.4 综合直方图的相似性度量……20
2.4 实验结果……21
2.5 本章小结……23
第 3 章 序列蒙特卡洛滤波算法……25
3.1 引言……25
3.2 贝叶斯时序滤波的框架……26

3.3 序列蒙特卡洛滤波核心思想 28
3.4 算法退化问题 31
3.5 提议分布问题 31
3.6 粒子重采样技术 33
3.7 粒子分布的一些方法 34
3.8 本章小结 35
第 4 章 交换采样粒子滤波技术 36
4.1 引言 36
4.2 Condensation 算法 37
4.2.1 Condensation 在目标跟踪中的应用 37
4.3 Auxiliary Particle Filter 算法 40
4.4 基于 PF 和 APF 的交换采样方法 41
4.4.1 硬性转换方法 42
4.4.2 软性转换方法 43
4.5 实验结果及分析 44
4.5.1 鲁棒性分析 45
4.5.2 精确性分析 47
4.5.3 整体跟踪情况分析 49
4.5.4 讨论 52
4.6 本章小结 53
结 论 54
参考文献 56
攻读硕士学位期间发表的论文和取得的研究成果 61
致 谢 62

第1章 绪论

1.1 研究背景

作为视觉的一项基本功能，人类视觉系统可以敏锐地感知所处场景中的物体运动。除非在很困难的情况下，如辨识乒乓球的运动轨迹，对生理视觉系统而言，运动感知已经成为一种自然而然的无意识视觉行为。然而对机器来说，迄今为止，赋予其感知运动的能力仍然是一项极大的挑战和尚未解决的问题，并作为视觉跟踪领域的主要研究目标而具有广泛的商业应用前景。

随着硬件成本的降低和制造工艺的改进，摄像机等成像设备的性价比快速提高，并迅速普及到生产生活中的方方面面，在安全保障等方面发挥着越来越重要的作用。但是，当前的视频监控系统只能简单地记录事情发生的映像。自“911”事件以来，恐怖主义的猖獗使国家安全危机凸显，赋予视觉监控系统能够理解所拍摄场景的能力成为日益迫切的需求。在2000年，美国DARPA 启动了HumanID 项目，其目标是试图通过视觉方法远距离地辨认人的身份信息，而步态等运动信息在其中扮演了重要角色。除视觉监控之外，视觉跟踪技术还可以在基于视频的运动分析等领域发挥重要作用，如可以采用跟踪技术分析运动员的技术动作与理想动作之间的差别。另外运动分析技术还可以用来指导真实感运动的合成技术，从而为虚拟现实技术的广泛应用提供技术支持。

视觉跟踪技术还经常作为很多系统的重要组成部分。在视频分析领域，系统输入是一组在时间轴上具有关联的图像集合。如果计算机能够将图像集合分解为事件的集合，从而得到视频的高层描述和语义特征，将对视频检索等领域产生重要影响。跟踪技术可以通过寻找在时间轴上相同物体的对应，不仅将图像集合分解为以物体为中心的对象集合，同时能够得到物体的运动信息，从而为场景语义的提取提供了重要支持。近些年提出的视频压缩领域的国际标准MPEG-4 和MPEG-7 将基于对象的编码和描述列为视频处理的研究方向。而采用跟踪技术的运动分层技术可以对这一策略提供有效支持。

另外，在机器人自动导航、智能人机交互、视频分割、军事目标定位等领域，都迫切需要成熟稳定的视觉跟踪技术。除了上述应用方面的需求之外，研究视觉跟踪技术对理解人类视觉的机制和探讨人工智能的实现手段也具有重要的指导意义。

1.2 视觉目标跟踪研究概述

视觉跟踪问题是随着计算机技术的发展而逐渐成为研究热点的．二十世纪80年代以前，由于计算机技术的限制，对图像的处理与分析主要以静态图像[1]为主，在动态图像序列的分析中，对运动目标的跟踪带有很强的静态图像分析[2]的特点。二十世纪80年代初光流法(Optical flow)[3][4]被提出之后，动态图像序列分析进入了一个研究的高潮。其中对光流法的研究热潮从其产生一直持续到了二十世纪90年代中期，有关光流法的综述性文章可以参见文献[5]。直到最近，还有文章对L-K方法[6]进行了回顾。但是，光流法所需的运算量即使在现阶段对计算机而言也是太大，很难满足实时性要求，同时由于采用假设的局限性使得光流法对噪声特别敏感，很容易产生错误的结果，这些缺点造成光流法与实际使用之间还存在着很大的距离。从二十世纪80年代之后，还出现了其他众多的视觉跟踪算法，下边将介绍其中常用的一些方法。为了清楚地对这些方法进行说明，先对具体的视觉跟踪方法进行了分类介绍，然后介绍了处理视觉跟踪问题的两种思路。

1.2.1 视觉跟踪算法的分类

从 20 世纪 80 年代起到目前为止这近 25 年里，出现了众多的视觉跟踪算法。1988 年，Aggarwal 和 Nanadhakumar 对运动图像分析算法进行了总结，将算法分为两类，一类是基于光流法的分析，另一类是基于特征点的分析。此后在视觉跟踪领域中，又出现了许多新的方法，将视觉跟踪方法分为四类，分别是基于区域的跟踪、基于特征的跟踪、基于变形模板的跟踪和基于模型的跟踪，这种分类方法概括了目前大多数视觉跟踪算法，因此下面用这种分类方法对视觉跟踪算法进行介绍。

1)基于区域的跟踪(Region-based tracking)。基于区域的跟踪算法基本思想

是：首先得到包含目标的模板(Template)，该模板通过图像分割获得或是预先人为确定，模板通常为略大于目标的矩形，也可为不规则形状；然后在序列图像中，运用相关算法跟踪目标，对灰度图像可以采用基于纹理和特征的相关，对彩色图像还可利用基于颜色的相关。最常用的相关准则是平方和准则SSD(the sum of squared differences)，如下式

$$C=\sum(i-j)^2 \tag{1-1}$$

式中，i和j分别是图像I和J图像的像素。该算法还可和多种预测算法结合使用，如线性预测、二次曲线预测、卡尔曼预测等，以估计每幅图像中目标的位置。这种算法的优点在于当目标未被遮挡时，跟踪精度非常高，跟踪非常稳定。但其缺点首先是费时，当搜索区域较大时情况尤其严重；其次，算法要求目标变形不大，且不能有太大遮挡，否则相关精度下降会造成目标的丢失。有关基于区域的跟踪算法可参见文献[7,8]。近年来，对基于区域的跟踪方法关注较多的是如何处理模板变化时的情况，这种变化是由运动目标姿态变化引起的，如果能正确预测目标的姿态变化，则可实现稳定跟踪。

2)基于特征的跟踪(Feature-based tracking)。基于特征的跟踪算法一般也采用相关算法。与基于区域的跟踪算法的不同之处在于，后者使用目标整体作为相关时的对象，而前者使用目标的某个或某些局部特征作为相关时的对象．这种算法的优点在于即使目标的某一部分被遮挡，只要还有一部分特征可以被看到，就可以完成跟踪任务，另外，这种方法与卡尔曼滤波器联合使用，也具有很好的跟踪效果。这种算法的难点是：对某个运动目标，如何确定它的唯一特征集？这也是一个模式识别问题。若采用特征过多，系统效率将降低，且容易产生错误。在特征提取时，一般采用Canny算子获得目标的边缘特征 ，而采用SUSAN算子获得目标的角点信息。有关基于特征的跟踪算法还可参见文献[9,10]。

3)基于变形模板的跟踪(Deformable-template-based tracking)。变形模板是纹理或边缘可以按一定限制条件变形的面板或曲线。目标模板为一条手划(Hand-Drawn)的封闭曲线，该曲线通过方向及方向的变形逐渐与图像中的真实目标相适应，从而检索或跟踪复杂背景中的目标。在视觉跟踪过程中，更为常用的变形模板是由Kass在1987年提出的主动轮廓模型(Active contour

models，ACM) 又称为Snake模型，Kass用于唇动的跟踪。Snake模型非常适合可变形目标的跟踪，如对运动细胞的跟踪。这种模型与卡尔曼滤波相结合能够更好地进行跟踪。Vieren等提出了一种基于Snake模型的多目标跟踪算法， 思路比较新颖。但Snake模型比较适合单目标的跟踪，对于多目标的跟踪更多地是采用基于水平集(Level Set)方法的主动轮廓模型，有关Snake算法可参见文献[11]。

4)基于模型的跟踪(Model-based tracking)。对人体进行跟踪时，通常有三种形式的模型，即线图模型、2D模型和3D模型。但是在实际的视觉跟踪算法中，更多的是采用运动目标的3D模型进行跟踪，尤其对于刚体如汽车的跟踪。其基本思想是：首先由先验知识获得目标的三维结构模型和运动模型，然后根据实际的图像序列，确定出目标的三维模型参数，进而确定出目标的瞬时运动参数。这种方法的优点是，可以精确地分析目标的三维运动轨迹，即使在运动目标姿态变化的情况下，也能够可靠地跟踪。但是其缺点在于，运动分析的精度取决于几何模型的精度，在现实生活中要获得所有运动目标的精确几何模型是非常困难的。这就限制了基于模型的跟踪算法的使用，同时，基于3D模型的跟踪算法往往需要大量的运算时间，很难实现实时的运动目标跟踪。有关基于模型的跟踪算法可以参见文献[12,13]。

1.2.2 处理视觉跟踪问题的两个思路

对于视觉跟踪问题的处理，总体上讲有两种思路，一种称之为自底向上(Bottom-up)的处理方法；另一种称之为自顶向下(Top-down)的处理方法。自底向上的处理方法又称之为数据驱动(Data-driven)的方法，这种方法不依赖于先验知识，直接从图像序列中获得目标的运动信息并进行跟踪。自顶向下的处理方法又称之为模型驱动(Model-driven)的方法，这种方法一般依赖于所构建的模型或先验知识，在图像序列中进行匹配运算或求解后验概率，采用匹配运算时，如果相似距离最为接近则认为跟踪上运动目标；求解后验概率时，选择最大后验概率所对应的状态向量作为运动目标的当前状态。下边对这两种思路做进一步的介绍。

1)自底向上 自底向上实现视觉跟踪的思路以Divad Marr的视觉计算理论最具代表性[14]。Marr认为视觉过程的主要任务是从2D图像中定量地恢复出图

像所反映的场景中3D物体的形状和空间位置。为此，Marr将视觉过程分为三个阶段：早期、中期和后期。早期视觉由输入图像获得要素图，包括图像中强度变化剧烈处的位置、几何分布和组织结构；中期视觉实现由输入图像和要素图获得的2.5D图，即在以观察者为中心的坐标系里可见表面的发祥、深度及不连续轮廓等等；后期视觉由输入图像、要素图以及2.5D图获得的物体得3D表示。

与此相似，在自底向上的视觉跟踪过程中，跟踪的目的是获得场景中运动目标的位置、速度、加速度、运动轨迹等信息，这相当于视觉过程中的后期阶段，为得到这些信息，在获得图像序列并对图像序列进行预处理之后，首先直接从图像序列中检测运动目标，这是早期阶段；检测到运动目标之后，提取出运动目标并进行识别，以判定是否跟踪，这相当于中期阶段；最后对该目标进行跟踪并获得运动目标的相关运动信息。上述过程按处理顺序共分为四步，第一步为图像预处理，一般对所获得的序列图像进行消噪或增强，提高图像质量以方便后续处理；第二步为目标检测，在摄像机不动情况下一般采用帧差法或背景差法 进行变化检测以获取运动目标，此时得到的运动目标是一个斑点(Blob)；第三步为目标分类，要完成两个任务，首先提取出检测到的运动目标，其次对该目标进行识别；最后在目标跟踪阶段获得运动目标的相关运动信息。这种方法的优点在于能快速检测并跟踪运动目标，便于工程实现，目前已有大量实际系统利用该思路解决视觉跟踪问题，该方法在摄像机静止情况下效果非常好，但如果摄像机是运动的，则实现运动目标检测与跟踪就比较困难，因此在很大程度上局限了该方法的使用范围。

2)自顶向下 自顶向下的跟踪思路以Bar-Shalom的目标跟踪与关联思想最具代表性。这一思路将目标跟踪问题转换为在贝叶斯(Bayesian)理论框架下，已知目标状态的先验概率，在获得新的量测(Measurement)后不断求解目标状态的最大后验概率的过程。也就是说，在贝叶斯理论框架下，将视觉跟踪问题看作是“最优猜测”或者是一种“推理”过程，通常采用状态空间法(State space approach)实现视觉跟踪。首先确定目标的状态向量$\{X_k\}_{k=0,1,2,\cdots}$，其中k表示离散时间序列，状态是跟踪系统的输出，对应的状态方程如下：

$$X_{k+1} = F_k X_k + V_k \tag{1-2}$$

其次是目标的观测向量$\{Z_k\}_{k=1,2,\cdots}$，同样k表示离散时间序列，观测是指从图像中提取的各种特征，对应的观测方程如下：

$$Z = H_k X_k + W_k \tag{1-3}$$

在(1-2)和(1-3)式中，$\{V_k\}_{k=0,1,2,\cdots}$和$\{W_k\}_{k=1,2,\cdots}$表示噪声序列，通常假定它们是独立同分布的(Independent and identically distributed，记为i.i.d)。在确定了状态和观测之后，跟踪的任务是：a)根据k时刻前的观测来预测$k+1$时刻的状态；b)收到$k+1$时刻的观测后对预测做出修正。

假设在k时刻，有状态的先验概率$p(X_k|Z_{1:k})$，其中$Z_{1:k}=\{Z_1,Z_2,\cdots,Z_k\}$表示一直到当前时刻的所有观测量，该先验概率对于$k+1$时刻是先验，而对于$k$时刻是后验。则跟踪过程由两步组成。

预测：

$$p(X_{k+1}|Z_{1:k}) \propto p(X_{k+1}|Z_k)\, p(X_k|Z_{1:k}) \tag{1-4}$$

更新：

$$p(X_{k+1}|Z_{1:k+1}) \propto p(Z_{k+1}|X_{k+1})\, p(X_{k+1}|Z_{1:k}) \tag{1-5}$$

这样就得到了$k+1$时刻的后验概率$p(X_{k+1}|Z_{1:k+1})$。其中状态转移概率$p(X_{k+1}|X_k)$是建模系统的动态特性。当系统噪声为高斯分布(Gaussian)且F_k和H_k为线性关系(Linear)时，可以利用卡尔曼滤波(Kalman filter) 求解后验概率，此时$p(X_k|Z_{1:k})$和$p(X_{k+1}|Z_{1:k+1})$的分布都必须是高斯分布。当F_k和H_k为非线性关系(Nonlinear)时，可以利用扩展卡尔曼滤波(Extended kalman filter，EKF)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)。如果状态空间是离散的而且是由有限个状态组成的，则可以使用隐马尔可夫模型(Hidden Markov Models,HMM)进行跟踪。

但是，经典卡尔曼滤波只能处理线性、高斯、单模态(Unimodal)的情况。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(Multimodal)的。为此，一种称之为序列蒙特卡罗的方法(Sequential monte carlo methods,SMC) 被提出用于解决这类情况。该方法在刚提出之时并未受到很大关注，直到1996年，由Isard和Blake提出Condensation算法[15]之后才引起广泛注意。目前这一类方法被称为粒子滤波(Particle filter)，是当前视觉跟踪方法中的研究热点[16,17,18,19]，重要国际期刊专刊(IEEE Transactions on Signal Processing,50(2),2002)对此进行了讨论。UKF思想也可被粒子滤波方法所借用，这种情况下粒子滤波被称为UPF(unscented particle filter)。

自顶向下思路利用先验知识对跟踪问题建立模型，然后利用实际图像序列验证模型的正确性，这种方法具有坚实的数学理论基础，有很多数学工具可以使用，因此一直是理论界研究视觉跟踪问题的主流方法。

1.2.3 视觉跟踪中的难点

从控制的观点来看，视觉跟踪问题所面临的主要难点可以归结为对视觉跟踪算法在三个方面的要求，即对算法的鲁棒性、准确性和快速性要求。

1)鲁棒性(Robustness)。所谓鲁棒性是指视觉跟踪算法能够在各种环境条件下实现对运动目标持续稳定的跟踪。影响视觉跟踪算法鲁棒性的最主要原因在于

a)被跟踪运动目标的姿态改变；

b)运动目标所处环境的光照变化；

c)部分遮挡引起的运动目标不规则变形和全部遮挡引起的运动目标的暂时消失。

对运动目标进行三维重构能够有效解决其姿态改变所引起的目标丢失，基于3D模型的跟踪算法可在很大程度上解决这一问题，但基于3D模型的跟踪算法所存在的局限性严重限制了它在实际系统中的应用。当运动目标所处环境的光照发生改变时，采用图像灰度信息或色彩信息作为跟踪基础的视觉跟踪算法一般都会失效，而基于图像特征的方法往往不受光照改变的影响，如利用运动目标的边缘信息能有效避免光照变化对运动目标的影响，但在复杂环境中要将运动目标的边缘和周围其它目标边缘区分开来是非常困难的。遮

挡问题是视觉跟踪算法中又一难点问题，利用单摄像机解决遮挡问题也一直是视觉跟踪领域中的热点，而利用多摄像机可以在很大程度上解决这一问题，但正如前所述，多摄像机的应用又会引入新的难题。总之，对视觉跟踪算法的鲁棒性要求是视觉跟踪问题研究的重点，目前，该问题还远远没有得到很好的解决。

2)准确性(Accuracy)。在视觉跟踪研究中，准确性包括两个方面，一是指对运动目标检测的准确性，另一个是指对运动目标分割的准确性。对运动目标检测准确性的目的是尽可能降低运动目标检测的虚警率和漏警率，从而提高对真实运动目标的检测概率，这一点在视频监视系统中要求比较高。由于实际复杂环境中存在大量噪声，很难实现各种环境条件下对运动目标地准确检测，往往只能在虚警率和漏警率之间寻求一个折中(Trade off)。对运动目标分割的准确性主要应用于图像压缩编码中，该问题实际上是图像分割(Image segmentation)问题，至今已经出现了上千种各种类型的分割算法，但由于尚无通用的分割理论，目前并没有一种适合于所有图像的通用分割算法。在视觉跟踪算法中，对图像序列中运动目标的分割，更多的是利用目标的运动信息实现对目标的提取，如利用光流法实现对图像序列中运动目标的分割。而主动轮廓模型也是实现对运动目标提取的常用方法。一旦能够实现对运动目标的准确检测和准确分割，那么也就能够准确获得运动目标的相关运动信息。

3)快速性(Speed)。一个实用的视觉跟踪系统必须能够实现对运动目标的实时跟踪，这就要求视觉跟踪算法必须具有快速性。但是，视觉跟踪算法处理的对象是包含巨大数据量的图像，这些算法往往需要大量的运算时间，很难达到实时处理的要求。通常，简单算法能够实现实时跟踪，但跟踪精度却很差；复杂算法具有很高的跟踪精度，实时性却很差。一种通用的减小视觉跟踪算法运算量的方法是利用金字塔分解或小波变换将图像分层处理；另一种思路是设计专用硬件实现实时的视觉跟踪。上述讨论仅仅只是从控制角度简要地给出了视觉跟踪技术中的几个主要难点问题。视觉跟踪技术所面临的更为一般的难题是：如何从图像序列中实现有效的图像分割和图像对应。这里的图像分割是指将被跟踪目标从图像或图像序列中提取出来，而图像对应是指将被跟踪目标在图像序列中始终能够准确的对应起来，无论该目标是否变形、被遮挡等等。图像分割问题目前尚未得到有效解决，而图像对应问题

则是与模式识别紧密相连的人工智能难题。

1.3 本文的研究工作及创新点

本文通过研究视频运动目标跟踪的两个主要方面来改进运动目标跟踪算法，一是，运动目标特征的表述。二是，改进粒子滤波。

第一个是运动目标特征的表述。根据以前的运动目标跟踪采用的单个特征表述方法的不足，本文采用了多特征来刻画物体，增强目标刻画的鲁棒性。颜色直方图在跟踪中有很好的应用，但是在光线变化，背景变化等情况下刻画能力较弱，因为它只表现了物体的颜色信息而没有表现出物体的结构特征信息。图像的梯度有力的刻画了物体的结构特征信息，但是对旋转较敏感。因此本文想到将两者结合起来形成优势互补，来增强对物体特征的刻画，为后续的算法应用奠定了基础。

另一个是粒子滤波算法，本文针对传统粒子滤波算法在无法预料的遮挡和干扰中等复杂背景下不能很好的跟踪到目标这个问题，提出了新的方法—交换采样。本文集中研究了 Condensation 和 Auxiliary Particle Filter 在运动目标跟踪中的运用。Condensation 算法强调状态转换的预测多一点，鲁棒性强一些。APF 强调观察多一点，精确性强一些。因此本文将两者的长处结合起来形成互补，并根据一个简单的原则来进行在线交换采样。

1.4 本文的组织结构

第一章绪论，主要论述了视频跟踪算法的研究背景和意义，分析了国内外的研究现状以及视频跟踪算法中存在的难题，列出了本文的主要目的和研究内容以及本文的创新点。

第二章目标表观的 HOGC 特征提取，主要综述了当前目标表观提取这一方面比较典型的几种特征，详细阐述了这些特征表观的优缺点及各种改进。接着提出了自己的特征表观—HOGC(Histogram of Oriented Grads & Color)，讲述了 HOGC 这种特征提取的方法以及用它来做跟踪的实验结果。

第三章序列蒙特卡洛滤波算法，在这章里系统地阐述了序列蒙特卡洛滤

波算法的贝叶斯理论框架、算法的核心和该算法最重要的三个方面退化问题、提议分布和采样方法。

第四章交换采样粒子滤波技术。本文首先分别介绍了 Condensation 算法和 Auxiliary Particle Filter 算法在跟踪中的运用。接下来，详细介绍了通过计算当前估计状态的信任度来在线选择采样方法的思想，并通过一系列实验的结果和分析说明了该方法在目标跟踪中达到很好的效果。

第 2 章 目标表观的 HOGC 特征提取

2.1 引言

视觉跟踪系统主要由四项关键技术支撑：目标表观的建模，目标表观的提取，数据关联方法和滤波方法。需要指出的是，跟踪问题在雷达，声纳等军事领域具有更长远的研究历史和更丰富的研究成果，视觉跟踪领域中滤波框架和数据关联等理论方法大多借鉴自上述领域中。相对于数据关联和滤波框架更多地根植于自动控制和信号处理领域，目标表观的建模和提取则与计算机视觉研究中的图像表示、图像匹配等技术有更密切的关联，但由于目标运动的时间关联特性，又使其与静态图像匹配等领域的问题不同，形成单独的研究子领域。

粒子滤波（Particle Filter）算法由于处理非线性、非高斯问题的突出能力，在图像跟踪领域得到了广泛的应用。目标表观对于 PF 跟踪算法的跟踪性能起着决定性的作用，因此构建区分性强的视觉特征显得格外重要。选择区分性强、稳定性好的视觉特征，跟踪算法才能够有效地区分目标，从而精确地跟踪目标。文献中常用的视觉特征包括颜色特征、边缘特征等。尽管这些跟踪算法取得了较好的跟踪效果，但对于那些复杂的跟踪场景（例如光照变化、姿态改变，目标发生非平面旋转等），依靠单一的视觉特征往往是不充分和不稳定的，很难取得良好的跟踪性能。针对单一视觉特征跟踪的局限性，近来，许多学者提出了利用多视觉信息融合跟踪目标的算法[20]，这种算法利用了信息之间的互补性和冗余性，从而显著地提高跟踪算法的性能。当一种视觉信息在某些跟踪场景失效时，可以利用另外一种互补的信息实现跟踪。对于特定的场景，信息有效性是时变的、不确定的，不同信息对于不同场景会有不同的适应性，而利用多个信息跟踪目标，则会弥补单一信息不可靠的缺点，因此多信息融合跟踪算法是一种较为合理的跟踪算法[21]。

总体上说，目标表观的建模和提取是视觉跟踪研究中的独特问题，也是主要的算法性能决定因素。而滤波算法则提供了解决跟踪问题的理论框架。

2.2 目标表观的建模和提取

视觉跟踪中，目标表观表现为某图像区域中像素的集合。目标表观受自身运动(如旋转、形变、平移等)，成像参数变化(如观察角度和距离等)和外部自然条件变化(如光照等)等因素的影响，并随着上述因素的变化而改变。并且随着目标的运动，目标所处的背景也不断变化。所以，视觉跟踪问题主要处理的是时变信号的分类问题：即将时变的前景(目标)从时变的背景中正确地，连续地分离出来。

如何排除种种不确定和变化因素的影响，在“未来”的场景映像中推断出目标在时间轴上的对应关系，最主要的倚重因素就是正确地建立目标表观的刻画模型，从而对目标和背景的分类提供根本的保证。表观的提取过程则主要指在目标表观模型的基础上找到相邻帧间目标表观的对应，提取方法与所采用的目标模型是密切相关的。

本小节按照目标表观的建模及其提取策略的不同对前人的工作进行分类总结。

2.2.1 基于颜色分布的方法

图像区域颜色分布具有对旋转、小幅度仿射变换等变化和图像噪声较鲁棒的优点，并且计算简单，符合实时处理的计算要求。但是，颜色分布在对目标表观的表面颜色进行统计的同时，丢失了其颜色分布的几何特性，无法根据图像区域的颜色直方图重建图像区域，从而使完全不同的图像区域可能具有相同的颜色分布，所以，其对图像区域的刻画能力是偏弱的，只能编码目标相关的部分图像信息。

但颜色分布仍然是视觉跟踪中应用最广泛和最成功的目标刻画方法之一。通过将人的头部用椭圆近似，Birchfield 提出了采用直方图结合椭圆圆周梯度进行头部跟踪的方法。作者认为，颜色直方图刻画了椭圆内部的纹理信息，椭圆圆周梯度则表示了椭圆轮廓的信息，根据集合论的观点，二者在目标刻画上具有互补性，即当一种表示方法失效时，另一种方法往往能起到矫

正的作用。该方法计算简单，满足实时跟踪的要求。所采用的颜色空间为RGB空间的变形，三个颜色分量分别为色度分量G-R 和B-G，光照强度分量(R+G+B)/3。

相对于RGB 颜色空间，Bradski认为HSV 色彩空间中的H分量(色调分量)能够更好地区分人的肤色和其它自然色。算法将在H分量上肤色隶属度的分布用直方图概率密度估计的形式求取，然后对整幅图像中每个像素的肤色隶属度进行计算，最后采用Camshift 算法找到一个与人脸形状相似的具有最大概率可能的图像区域作为人脸跟踪结果。

为了弥补直方图的刻画能力不足的弱点，Comaniciu 采用了空间加权直方图。Enpanolkov 核函数使靠近图像区域中心的像素具有进行较大的权值从而部分编码了像素颜色分布的空间信息，使加权直方图相对于标准直方图具有了更好的目标定位精度[22]。该算法根据直方图是密度分布估计的一种，采用Bhttacharyya 距离来度量模型和图像区域直方图之间的相似度，并引入mean-shift 梯度上升算法来迭代匹配模型到正确的图像区域。不同于局部穷举搜索的方式，基于梯度上升的mean-shift方法降低了搜索的计算复杂性。为了满足mean-shift 算法的应用条件，算法要求相邻帧间的目标表观具有重叠的区域，该缺陷可以采用基于滤波的预测框架加以弥补。

与上述方法采用单一色彩空间计算直方图的方法不同，自适应地选择合适的色彩空间的方法首先在文献[23]的工作中得到了体现。在该工作中，作者注意到没有任何色彩空间在所有人脸跟踪任务中总是优于其它色彩空间的。在算法实现中，预定义了多个候选色彩空间，并以对被跟踪目标和背景的区分度作为选择色彩空间的标准，从而通过对背景的评估，自适应地从多个色彩空间中选择一个构建面部直方图来保证跟踪算法的鲁棒性。

自适应地直方图特征选择思想在文献[24]的工作中被用来进行一般图像区域的跟踪。在算法中，作者通过对RGB 颜色空间中的三个色彩分量进行加权 $w_1R+w_2G+w_3B$ 得到不同包含直方图的特征空间，其中，$\{w_i|i=1,2,3\}\in[1,5]$，共有49 种不同的特征空间可供选择。与文献[23]中的自

适应选择标准类似，算法定义了似然比来度量特定直方图对前景和背景的区分能力，作为特征空间选择的标准。

2.2.2 基于模板的方法

模板可以有效编码目标表观的信息，而且通常仅依靠较少的初始图像数量就可以进行构造。但是由于模板通常带有全局性的信息，并以像素为基本单位，难以有效地得到更新来反映目标表观的实时变化，所以基于模板的跟踪方法更适合于短时跟踪任务，而在长时序的跟踪任务中，基于模板的方法容易产生目标“漂移”现象[25,26]。

Frey 对传统的模板方法做了改进。对目标区域的每一像素，算法根据一段时间内的目标表观求取其变化范围。该模板对噪声具有更好的鲁棒性，但是增加了在训练集合上对图像像素的灰度变化进行预学习的过程。在随后的工作中Jojic 针对场景中存在多物体的情况，建立了层次性的基于模板的目标模型。由于该生成式模型中包含了叠加性高斯噪声和阿尔法通道，具有处理多目标深度信息和透明性遮挡等情况的能力。模型的学习采用了效率优于最大后验概率方法的迭代算法。在迭代过程中，在不同运动层间传递关于灰度的概率分布信息，在层内水平地传递多目标的标识信息。但在算法中，运动层的个数和目标物体的数目都是假设预知，从而限制了算法的实用性。

Rucklidge提出了一种有效率地定位灰度模式的方法。算法基于块匹配策略，采用平方差之和(Sum of Squared Differences, SSD)作为图像匹配度准则。算法将特定模式匹配到图像区域的函数变换空间划分为若干子空间。在每个子空间中，考虑所有可能的变换形式，对变换后的特定模式和图像区域之间的差值根据经验设定阈值，从而忽略那些最佳匹配值亦大于阈值的子空间来降低计算量。对最佳匹配值低于阈值的子空间则进行进一步划分，并采用深度优先的搜索策略进行求解。

Olson 等人认为广泛使用的模板匹配准则SSD对所有像素等同看待，所以对噪声等外部影响敏感，并且该准则不随目标表观和表观的偏离情况呈线性关系。算法将模板匹配过程纳入到基于极大似然估计的概率求解框架中。为达到亚像素的匹配精度和处理遮挡等诸多非确定因素，算法中采用高斯分

布来拟合模型表观间相似性分布的峰值区域，并将高斯分布的峰值作为匹配的位置，高斯分布的方差则给出了估计结果的信度。在相同条件下的实验表明，该算法优于基于SSD匹配准则的方法。

Morris 分析了运动范围约束对跟踪具有关节运动特性的运动物体的作用，认为当运动与摄像机的视角同方向时，会造成约束矩阵的奇异。如果矩阵奇异，算法的收敛性变差甚至会丢失被跟踪目标。作者针对该问题提出了二维尺度棱镜模型(2D Scaled Prismatic Model, SPM)。相对于三维运动模型，该模型更少发生矩阵奇异问题，并且不需要预先学习特定的三维运动规律。算法还可以对发生矩阵奇异的情况进行判断从而在发生奇异时通过借助三维模型进行跟踪。

2.2.3 基于轮廓的方法

轮廓是目标刻画中广泛使用的图像特征，轮廓点通常包含目标的重要信息。基于轮廓的方法一般求解策略是通过迭代方法趋近于问题相关的能量函数的局部最优解。基于轮廓的方法通常对搜索的起始位置比较敏感，所以初始位置的选择是基于轮廓方法性能的重要依赖。在视觉目标跟踪中，算法通常将前帧的搜索结果作为后帧的起始搜索位置。

Snake 算法是视觉跟踪中最著名的轮廓模型之一，其采用样条曲线连成的封闭轮廓来逼近目标物体的形状[11]。算法中的能量函数包含三个子项：控制轮廓光滑程度的内力，控制初始位移的外力和控制边缘、特征点吸引程度的图像力。通过梯度下降来迭代最小化上面三个能量项的加权和，最终收敛到能量一维分布中的某局部极小作为算法的输出，其中三个能量项的权重往往需要根据具体应用问题进行调整。

Leymarie采用Snake算法进行细胞轮廓的跟踪。细胞轮廓的初始位置在参考帧中由人工确定。并采用前帧中的估计结果作为后帧中的初始搜索位置。当相邻帧间的细胞位移和形变较小，能观察到算法具有较好的跟踪性能。与原始Snake 算法中的能量项不同，作者采用轮廓上的分段能量平均最小值作为优化标准，并认为相对于求取[11]中定义的目标函数最小，新的能量函数在避免搜索中的振荡和定位到最优解的频率上都有提高。

Paragios 根据水平集(Level Set)理论提出了测地主动轮廓模型(Geodesic

active contour)[27]。在该算法中，根据水平集理论递归地改变预定义轮廓的位置和形状来拟合目标的轮廓，达到同时跟踪和检测目标的目的。

上述的基于轮廓的目标跟踪方法主要采用经验定义的能量函数，并试图利用通用的能量函数解决不同的跟踪问题。在文献[28,29]的工作中，学习的机制被引入到主动轮廓模型中，虽然得到模型的过程需要繁琐的学习训练，并且学到的模型是应用相关的，但这种针对特定问题的模型表现出了比通用模型优秀的定位和跟踪特性。

主动形状模型(Active Shape Model, ASM) 和主动表面模型(Active Appearance Model, AAM)都是基于点分布模型(Point Distribution Model, PDM)的算法[28,29]。在PDM 中，特定类别物体(比如人脸、人手)的轮廓形状通过若干关键的特征点进行定义，这些特征点的坐标串接构成描述目标的原始形状向量。对训练集中的所有形状向量进行对齐操作后，对他们进行PCA分析建模，保留的主成分形成最终的形状模型，形状模型的参数个数反映了形状的主要可变化模式。ASM 搜索首先通过局部纹理模型匹配得到各个特征点的更佳的位置，经过相似变换对齐后，通过统计形状模型对其进行约束，而后再进行局部纹理模型的匹配，形成一个迭代过程，以期形状模型最终匹配到输入的形状模式上去。在ASM 中，仅使用了特征点局部纹理特征作为启发式信息，没有使用全局的纹理约束，实践中发现ASM 很容易陷入局部极小。

而在AAM中，则采用了形状和纹理二者融合的统计约束，即所谓的统计表观模型。AAM 搜索借鉴了基于合成的分析技术(Analysis-By-Synthesis, ABS)的思想，通过模型参数的优化调整使得模型能够不断逼近实际输入模式。模型参数的更新则放弃了ASM 中的局部纹理搜索过程，仅使用一个线性预测模型根据当前模型和输入模式之间的差别来预测和更新模型参数。AAM 尽管利用了全局纹理，但却抛弃了局部纹理匹配过程，因此会在一定程度上降低关键特征点配准的精度，而且其线性预测模型也有较大的局限性，在初始位置偏离目标位置过大时，则很难收敛到正确位置。

除了在轮廓模型本身的改进，Isard 将主动轮廓模型融入到CONDENSATION(Conditional Density Propagation)算法中[15]。在算法实现中，轮廓形状的后验分布采用一组离散的粒子及其相应的权重表示。这种非参数的离散表示方法可以求解任何形式的后验分布。而前述算法中主要采用基于

梯度下降的策略，相当于力图求解目标轮廓形状的后验概率分布的期望。当后验分布呈尖峰状分布的时候，期望具有较好的解性质，否则，很容易造成被跟踪目标的丢失。而在文献[15]中，采用粒子模拟后验概率分布的方法不仅解决了由于暂时的误差所造成的误差传递问题，同时也在一定程度上解决了主动轮廓模型对其搜索位置要求较苛刻的问题。

2.2.4 基于子空间的方法

从本质上说，跟踪问题处理的是非静态信号，前景目标和背景都在随着时间的变化和目标的运动而改变。虽然现在有众多的方法能在短时间隔内和可控环境中很好地跟踪目标，但是当目标表观本身发生剧烈变化或者环境(比如光照)发生不可预知改变时，往往会造成跟踪中的“漂移现象”，造成跟踪任务的失败。在基于子空间建模的方法中，由于其能够充分地利用历史数据或者训练集得到目标物体表观变化的子空间，从而提供了完整描述前景目标变化的能力，减少因为目标的表观变化所产生“漂移现象”的可能。

Black 提出了基于视角的子空间表示方法并将两不同视角的子空间匹配过程归结为求解优化问题[30]。该工作的主要贡献有两点：首先，对匹配中具有较大误差的特征项进行降权，提高了误差范数的鲁棒性，从而避免了等权使用特征系数时对噪声和奇异点敏感的缺点，比如发生部分遮挡等情况。其次，算法中泛化了光流计算中的灰度恒定假设，提出了子空间一致性假设。该假设认为，目标的视角图像，如果找到正确的参数变换来调整图像，可以用相同的特征基组进行重建，并且可以保证重建图像和原图像具有相同的图像灰度分布。在该假设的基础上，通过求取目标视角图像和子空间重构图像之间的形变参数，做到视角无关的目标跟踪。在其后续工作中，作者进一步采用混合模型构造子空间来建模目标表观的变化。

Torre 在文献[30]的基础上，将基于子空间的表示方法应用到人体跟踪任务中[31]。根据当前需处理的帧数据和训练集中图像的相似性，选择训练集合中的相似度高的一个子集来构造所跟踪人体的特定子空间。算法中采用肤色进行面部区域的分割，并采用了卡尔曼滤波算法来估计仿射运动模型的参数。

Hager 和Belhumuer 提出了基于参数化运动模型的模板匹配方法。通过梯度下降策略来求解运动模型的参数从而将固定的模板匹配到正确的图像位

置。由于是应用在视觉跟踪中，为了避免每帧都要计算雅可比矩阵，该工作将矩阵分为由模板的亮度梯度乘以其变换的空间微分的常量项和针对运动模型的参数进行微分的可变项，从而大大简化了计算复杂度。该方法在模板的基础上利用子空间方法处理光照问题，通过对涵盖各种光照情况的训练图像的学习得到光照描述的子空间。算法中队部分遮挡问题也采用M估计算子(M-estimator)进行了简单地处理。

Jepson 等人提出了一种融合稳定部分、噪声部分和帧间变化部分的目标模型[32]。该模型采用小波基响应作为图像特征，采用期望最大化(Expectation-Maximization)算法估计混合模型中三部分的权重并且采用稳定部分进行仿射运动的估计。在该模型能够处理目标表观变化和光照变化的同时，作者指出在背景变化平缓的情况下，该模型也可能同时估计了背景的稳定部分，从而有可能造成被跟踪中常见的“漂移”问题。

虽然基于子空间的方法能够更好地刻画被跟踪目标的表观变化，但是由于需要事先的训练，从而阻止了在一些条件不满足场合的应用。Ross等人提出了渐进获取子空间基的方法，并且子空间基随着表观数据的不断增加而得到完善。在相邻帧间，算法采用类似粒子滤波的方法来对仿射运动的参数进行采样并求取最大后验估计。该算法可以处理较大的光照、姿态和尺度变化。但在当前作者给出的实验中，都是在比较简单的背景下进行的，如何能够正确完整地分割目标物体从而正确地更新子空间，在复杂背景中，仍是一个比较困难的问题。

2.3 HOGC 特征

正如前面所说各种目标表观的提取都有各自的特点。目前使用较多的是颜色直方图，比较经典的就是 Mean-shift 算法[22]。传统的颜色直方图是图像区域的一个有效的特征表达，近些年出现了一些经典的基于颜色直方图的跟踪系统，直方图计算简单并能够满足实时处理，同时对运动目标的旋转和尺度伸缩不敏感。但是由于颜色直方图信息隐藏了图像的空间位置信息，并且易受光照等外界环境因素的影响，导致了传统的直方图跟踪算法跟踪精度不高和效果不稳定。针对上述问题，有学者提出了梯度方向直方图的概念并且用于人

体检测，有学者提出改进的直方图方法并用于运动目标跟踪[33]，还有学者提出了直方图相似度的改进方法。本文受 SIFT 特征提取和匹配的启发[34,35]，设计了一种滤波预测框架下的基于综合直方图的运动目标跟踪算法(HOGC, Histograms of Oriented Grads & Color)，融合了梯度方向直方图(HoG, Histograms of Oriented Grads)和颜色直方图(HoC, Histograms of Colors)。然后采用改进的综合直方图表示跟踪目标模板和候选目标，来解决运动目标在复杂背景环境下的变化，如遮挡和形变等场景下的实时鲁棒跟踪。

通过运动目标在不同场景下的颜色直方图和梯度方向直方图匹配结果的比较和误差分析，发现颜色直方图和梯度方向直方图具有更好的匹配精度。因此本文中综合直方图采用的是 48 维颜色直方图和 72 维的梯度方向直方图，所提出的综合直方图很容易扩展到其他的颜色空间进行比较和分析。

2.3.1 梯度方向直方图

根据 SIFT 特征点提取的思想，定义灰度图像块的梯度方向直方图(HOG)表达。采用灰度空间，将物体缩放到固定尺度上(32*32)，通过固定模板计算得到物体的梯度方向，然后统计成 72 维直方图。梯度直方图保留了物体局部结构信息，但对于旋转敏感。

在计算过程中，以图像的中心领域采样，并用直方图统计领域像素的梯度方向，直方图的峰值则代表了该图像的方向。

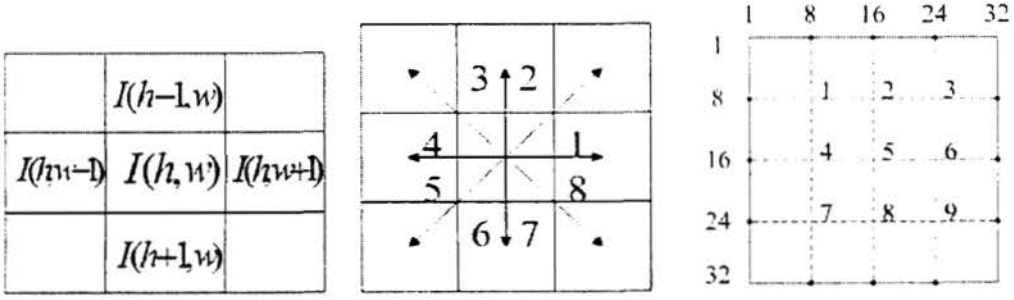


图 2.1 HOG 计算模板

定义 72 维的梯度方向直方图，以 0.25π 为单位将梯度空间均匀地分成 8 个 bin。固定尺度上共有 16 个区域，每 4 个相邻区域构成的正方形为一个窗口，共 9 个窗口（见图 2.1）。对于每一个窗口区域，使用(3*3)梯度模板

$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$依次计算像素上的梯度，并统计落在每个 bin 上的像素个数。

$$\begin{aligned} &I = G(\sigma,0) * I_0 \\ &dy = I(h+1,w) - I(h-1,w) \\ &dx = I(h,w+1) - I(h,w-1) \\ &ori(h,w) = atan2(dy,dx) \quad ori \in [0,2\pi] \end{aligned} \tag{2-1}$$

上式是计算一个像素在特定尺度下一个像素点方向的过程。$G(\sigma,0)$表示一个高斯函数，I_0表示原图像，I表示经过平滑后的图像。$ori(h,w)$表示像素梯度方向（目标轮廓方向）。对于目标区域的所有像素累计，可以获得 72（9 个窗口，每个窗口 8 个 bin）维 HOG 特征。此特征因为只保留方向特征，不保留强度特征，对于亮度、光照等变化具有适应性。

2.3.2 颜色直方图

类似于 Mean-shift 中的直方图，但是我们所使用的颜色直方图没有加权。颜色直方图为三个分量(R,G,B)，每个分量分成 16 个 bin,依次统计图像块里落在每个 bin 里的像素个数，共 48 维 RGB 直方图。

2.3.3 综合直方图（HOGC）

融合了颜色直方图和梯度直方图的 120 维的综合直方图。前 48 维是颜色直方图，后 72 维是梯度直方图。该综合直方图是两者直方图的加权组合。根据先验经验，我们偏向于颜色直方图，也就是颜色直方图的权重高于梯度直方图。这种直方图计算简单。

2.3.4 综合直方图的相似性度量

本实验中的跟踪是手动初始化。在初始化选定目标之后，运动目标模板的直方图$p(u)$就被统计出来了。在粒子滤波框架下的跟踪过程中，粒子的权值来源于当前状态下的观察，在本实验中就是该粒子状态下得到下的观察

$q(u)$和模板$p(u)$之间的相似度。因此我们需要一个相似性度量法则。比较流行的是巴特查理亚系数(Bhattacharyya coefficient)：

$$\rho[p,q]=\int\sqrt{p(u)q(u)}du \tag{2-2}$$

考虑到我们所使用的直方图的离散形式 $p=\left\{p^{(u)}\right\}u=1,2,3,\ldots m$ 和 $q=\left\{q^{(u)}\right\}u=1,2,3,\ldots m$，巴特查理亚系数可以定义为：

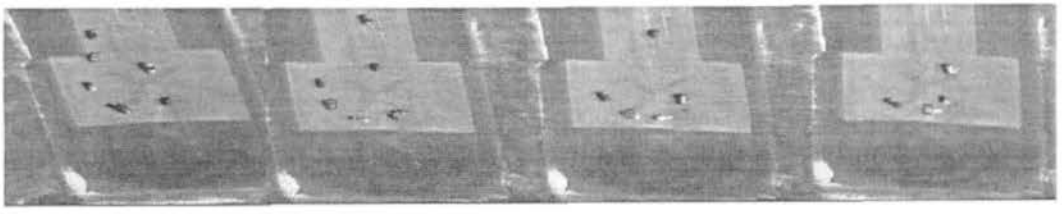
$$\rho[p,q]=\sum_{u=1}^{m}\sqrt{p(u)q(u)} \tag{2-3}$$

ρ越大，$p(u)$和$q(u)$的相似度越大。对于两个归一化的直方图，如果$\rho=1$，标识了最好匹配。按照两者距离的思想我们可以定义巴特查理亚距离：

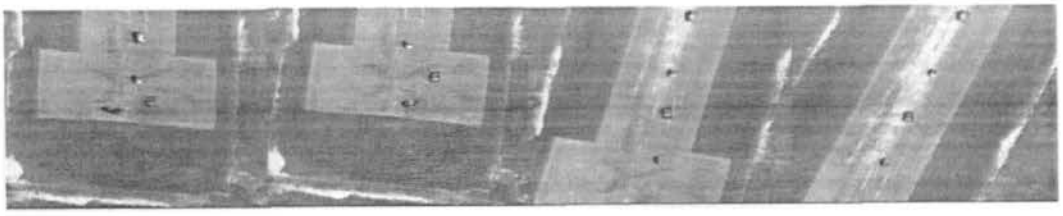
$$d=\sqrt{1-\rho[p,q]} \tag{2-4}$$

2.4 实验结果

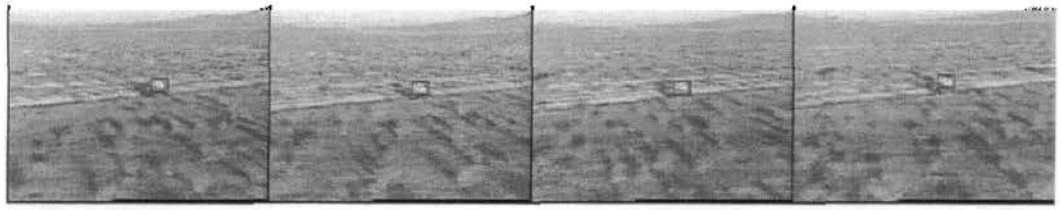
为了验证我们提出的综合直方图在跟踪中的效果。我们在那些有复杂背景的视频做了实验。所有实验在 CPU 3.0 GHZ、1G 内存和独立显卡的 PC 机上进行，实验结果如图 2.2 视频-1、视频-2、视频-3 所示：



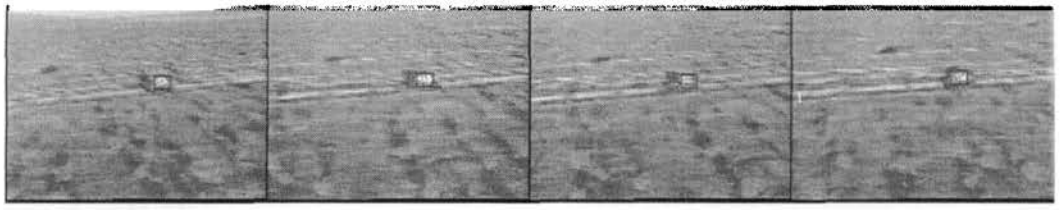
视频-1　100 帧　　200 帧　　250 帧　　300 帧



350 帧　　400 帧　　500 帧　　600 帧



视频-2　70 帧　　90 帧　　120 帧　　150 帧



180 帧　　220 帧　　250 帧　　270 帧

视频-3　50 帧　　100 帧　　150 帧　　200 帧

250 帧　　300 帧　　350 帧　　400 帧

图 2.2 HOGC 跟踪算法在复杂背景下的实验结果

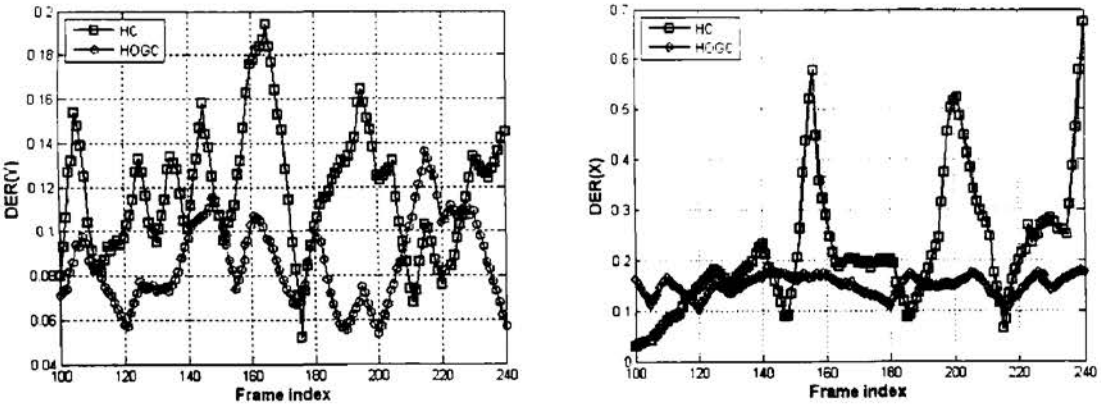


图 2.3 HOGC 跟踪算法与 HC 算法在跟踪误差上的比较

先做这个实验的定性分析，在视频-1 中，跟踪的难点在于车转向掉头的那一系列帧中。在这一段视频中，车转向有光照发生变化和转向后车的表观发生了变化，因此给跟踪带来了困难。从跟踪效果来看 HOGC 跟踪算法在车转向时和转向后仍然能够很好的跟踪上车。而单纯的直方图跟踪在车转向的时候会发生丢失。从视频-1 可以看出，HOGC 的优点，在车转向时，颜色直方图受到较大影响，但是梯度直方图有对光照不敏感的优点，因此还可以继续跟踪上。等车已经掉头后，车的形态已发生了变化，这时颜色直方图就发挥了它对目标变形不敏感的优点。视频-2 是一辆车在草原中行进，这个视频的特点就是目标比较弱小，并且跟踪中目标也发生了形变，这个视频的跟踪结果 HOGC 和 HC 的特征都可以跟踪上，但是 HOGC 的准确度要高一些，图 2.3 是它的定量分析。视频-3 的特点是目标跟背景的颜色是相似的，目标是一个人骑着摩托车在公路上，目标表观信息不强，在这个视频中，HOGC 跟踪算法可以很好的跟踪目标，而单一直方图在开始可以跟踪上，但到了 300 帧以后就会发生目标丢失。

为了验证 HOGC 特征的跟踪算法的鲁棒性和精确性，本文将传统的单一直方图算法(HC Tracking)和本文提出的综合直方图算法(HOGC Tracking)在视频-2 中进行了比较，结果如图 2.3 所示。本文将跟踪得到的目标位置与人工划定的目标真实位置做误差分析。无论在 X 轴的坐标上和在 Y 轴的坐标上，HOGC 的精度都要比 HC 的精度要高，并且也要稳定一些。

2.5 本章小结

本章详细介绍了当前目标表观的几种典型的类别。这些基于颜色、轮廓和模板的目标表观方法都有各自的优缺点。颜色对于相似背景的区分力弱，轮廓对于形变区分力弱，模板对于模板更新是个难点。因此使用单一的特征表观已经不能很好的应用到复杂背景下的目标跟踪。于是本文采用了最近流行多线索融合跟踪的方法。将几个典型的目标表观融合起来，取长补短形成一个综合特征来表征目标。本章就详细介绍了我们所用到的目标表观特征—HOGC，并将其用到粒子滤波的框架中进行了跟踪实验。实验表明 HOGC 特征在复杂背景下跟踪的鲁棒性效果要比单纯的颜色特征要好。同时，由于在

粒子滤波框架下，目标模板和粒子状态的候选模板之间要进行相似性度量以确定该粒子的权值，所以也介绍了在跟踪中我们如何采用巴特查理亚系数来确定粒子的权值。

第 3 章 序列蒙特卡洛滤波算法

3.1 引言

在视觉跟踪问题中，由于目标自身的运动，背景的变化以及成像条件的复杂多样性等诸多不确定性因素的存在，采用简单的匹配搜索策略已经无法处理日益复杂的视觉跟踪问题。从自动控制领域引入的卡尔曼滤波算法首先为视觉跟踪问题提供了完整的求解理论框架。由于卡尔曼滤波算法基于贝叶斯理论，其能够处理各种非确定性因素和融合多通道信息。卡尔曼滤波算法目前仍是视觉跟踪研究中应用最广泛的算法，但由于算法基于线性系统和高斯分布假设，随着视觉跟踪问题研究的深入，卡尔曼滤波算法在诸多复杂视觉跟踪问题上越发力不从心。

近年来引入视觉跟踪研究领域的序列蒙特卡洛滤波算法与卡尔曼滤波算法一样根植于贝叶斯时序滤波框架。在某种意义上，卡尔曼滤波算法可以看作序列蒙特卡洛滤波算法的特例。序列蒙特卡洛滤波算法由于下述优点而得到了越来越多研究人员的重视：

1) 相对于卡尔曼滤波要求后验分布满足高斯形式，算法可以求解后验分布为任意函数形式的视觉跟踪问题；
2) 算法在理论上能够解决带有任意非线性特性的复杂视觉跟踪问题；
3) 算法能够在时间轴上传递条件分布的变化，从而系统地处理不确定性问题；
4) 当粒子数目足够多时，算法总能够得到收敛的跟踪结果；
5) 由于在贝叶斯的框架下进行推断，算法能够有效融合多通道信息。

由于本论文以序列蒙特卡洛滤波算法为理论框架来系统地研究以人体跟踪为例的视觉跟踪问题。本章依据文献[36]简要介绍序列蒙特卡洛滤波算法，从而为阐述提出的相关算法奠定基础。

3.2 贝叶斯时序滤波的框架

视觉跟踪的主要研究目标是如何通过输入的图像时间序列产生对目标物体运动信息的感知。考虑视觉跟踪问题中需要处理各种不确定性因素的发生，本节从基于概率统计和贝叶斯框架的角度来定义，阐述和分析视觉跟踪问题。从贝叶斯统计推断的角度，算法的理想输出是关于物体运动状态的后验概率分布。如何在存在噪声和诸多不确定因素的目标表观基础上，准确而有效地表示和求取目标运动状态的后验概率分布是跟踪算法面临的主要任务。用 X_t 表示目标物体在时刻 t 的运动状态，用 Z_t 表示时刻 t 的目标表观。用 $X_{1:t}$ 表示状态序列 $\{X_1,\cdots,X_t\}$，类似有 $Z_{1:t}$。

在视觉跟踪研究中，主要需要分析和解决如下三个基本问题：

1) 目标运动状态的预测：假设已经知道 $p\left(X_{1:t-1}\middle|Z_{1:t-1}\right)$，如何预测 X_t？要解决这一问题，需要定义 $p\left(X_t\middle|Z_{1:t-1}\right)$。

2) 数据关联问题：在时刻 t 提取的所有表观，只有部分与目标运动状态有关联，需要通过定义 $p\left(Z_t\middle|Z_{1:t-1}\right)$ 分辨与目标相关的表观，从而依据正确的表观求取目标的运动状态。

3) 状态校验问题：在得到 t 时刻表观 Z_t 后，需要根据最新表观调整后验概率得到 $p\left(X_t\middle|Z_{1:t-1}\right)$ 解的更新估计。

由于上面的求解后验概率问题的复杂性，首先进行如下的条件独立性假设：

1) 目标的运动状态只与其相邻状态条件相关，即

$$p\left(X_t\middle|X_{0:t-1}\right)=P(X_t|X_{t-1}) \tag{3-1}$$

2) 目标的当前表观唯一条件依赖于目标的当前运动状态，即当 X_t 给定，

Z_t 与其他时刻的表观条件无关，表示为

$$p\left(Z_{t:t-m}\middle|X_t\right)=p\left(Z_t\middle|X_t\right)p\left(Z_{t-1:t-m}\middle|X_t\right) \tag{3-2}$$

在解决实际问题中积累的经验表明，上述问题简化的假设是合理的，而且极大地简化了视觉跟踪问题的求解过程。根据上述假设，对状态预测问题，可以导出

$$\begin{aligned}
&p\left(X_t\middle|Z_{1:t-1}\right)\\
&=\sum_{X_{t-1}}p\left(X_t,X_{t-1}\middle|Z_{1:t-1}\right)\\
&=\sum_{X_{t-1}}p\left(X_t\middle|X_{t-1},Z_{t-1}\right)p\left(X_{t-1}\middle|Z_{1:t-1}\right)\\
&=\sum_{X_{t-1}}P(X_t|X_{t-1})p\left(X_{t-1}\middle|Z_{1:t-1}\right)
\end{aligned} \tag{3-3}$$

对状态校验问题

$$\begin{aligned}
&p(X_t|Z_{1:t})=\frac{p\left(X_t,Z_{1:t}\right)}{p\left(Z_{1:t}\right)}\\
&=\frac{p\left(Z_t\middle|X_t,Z_{1:t-1}\right)p\left(X_t\middle|Z_{1:t-1}\right)p\left(Z_{1:t-1}\right)}{p\left(Z_{1:t}\right)}\\
&=p\left(Z_t\middle|X_t\right)p\left(X_t\middle|Z_{1:t-1}\right)\frac{p\left(Z_{1:t-1}\right)}{p(Z_{1:t})}\\
&=\frac{p\left(Z_t\middle|X_t\right)p\left(X_t\middle|Z_{1:t-1}\right)}{\sum_{X_t}p\left(Z_t\middle|X_t\right)p\left(X_t\middle|Z_{1:t-1}\right)}
\end{aligned} \tag{3-4}$$

假设在初始时刻，有 $p\left(X_0\middle|Z_0\right)=p\left(X_0\right)$，则 $p\left(X_t\middle|Z_{1:t}\right)$ 可以通过预测和状态更新的两步迭代运算得到。设 $t-1$ 时刻的目标状态后验概率分布为 $p\left(X_{t-1}\middle|Z_{1:t-1}\right)$。在预测阶段，采用公式(3-3)到目标在 t 时刻的先验概率分布

$p(X_t|Z_{1:t-1})$。在t时刻，提取目标表观Z_t，根据公式(3-4)得到目标状态的后验概率$p(X_t|Z_{1:t})$。

贝叶斯时序滤波框架是依据预测校验步骤，在贝叶斯框架下进行迭代后验概率估计的一类滤波器的总称，公式(3-3)和(3-4)构成了贝叶斯滤波器的核心步骤。通过在时间轴上递归地传递后验概率分布，可以得到贝叶斯规则下的最优序列解。贝叶斯滤波器包括广泛使用的卡尔曼滤波器，扩展卡尔曼滤波器，序列蒙特卡洛滤波器等。

3.3 序列蒙特卡洛滤波核心思想

作为蒙特卡洛方法(Monte Carlo)的一种，序列蒙特卡洛滤波器在很多研究领域中都有应用，如bootstrap filtering , CONDENSATION, particle filtering, interacting particle approximation 和survival of fittest 都是序列蒙特卡洛滤波算法在不同领域中的名称。序列蒙特卡洛滤波器是通过蒙特卡洛模拟对贝叶斯滤时序滤波器的实现，其核心思想是采用对特定概率分布的一组随机采样和其相应权重的集合来表示待求后验分布，并根据这些采样计算后验分布的某种测度。根据蒙特卡洛特性，当采样数据趋于无穷多，算法可以无穷逼近后验概率的真实分布，从而可以得到无限接近最优解的估测结果。

设$\{X_{0:t}^i, w_t^i | i=1,\cdots,N_s\}$是对后验分布$p(X_{0:t}|Z_{1:t})$的随机采样集合，其中$\{X_{0:t}^i | i=1,\ldots,N_s\}$是采样点集合，这里称为粒子集合，$\{w_t^i | i=1,\ldots,N_s\}$是与采样点相关联的权重集合，并满足$\sum w_t^i = 1$。于是，后验分布$p(X_{0:t}|Z_{1:t})$可以用该粒子集合离散地近似为

$$p(X_{0:t}|Z_{1:t}) \approx \sum_{i=1}^{N_s} w_t^i \delta(X_{0:t} - X_{0:t}^i) \tag{3-5}$$

公式(3-5)中的权重集合$\left\{w_t^i \middle| i=1,\ldots,N_s\right\}$依据重要性采样原理确定[37][38]：假设$p(x)\propto\pi(x)$是一个难以进行采样的分布例如无法得到$p(x)$的解析解或者只了解$p(x)$的部分信息等情况。用$X^i\sim q(X), i=1,\ldots,N_s$表示从分布$q(\bullet)$得到的采样，称$q(\bullet)$为提议分布(proposal distribution)或者重要性密度(importance density)。于是，$p(\bullet)$可表示为

$$p(X)\approx\sum_{i=1}^{N_s}w^i\delta\left(X-X^i\right) \tag{3-6}$$

其中，

$$w^i\propto\frac{\pi\left(X^i\right)}{q\left(X^i\right)} \tag{3-7}$$

是第i个采样的权重。

于是，如果粒子$X_{0:t}^i$是从提议分布$q\left(X_{0:k}\middle|Z_{1:k-1}\right)$的采样，则在公式(3-7)中的权重可以写做：

$$w_t^i\propto\frac{p\left(X_{0:t}^i\middle|Z_{1:t}\right)}{q\left(X_{0:t}^i\middle|Z_{1:t}\right)} \tag{3-8}$$

考虑视觉目标跟踪问题，在时刻t，已经有后验分布$p\left(X_{0:t-1}\middle|Z_{1:t-1}\right)$的粒子集合逼近，现在需要更新粒子集合来逼近$p\left(X_{0:t}\middle|Z_{1:t}\right)$。如果提议分布可以分解为如下的形式：

$$q\left(X_{0:t}\middle|Z_{1:t}\right)=q\left(X_t\middle|X_{0:t-1},Z_{1:t}\right)q\left(X_{0:t-1}\middle|Z_{1:t-1}\right) \tag{3-9}$$

则可以得到，$X_t^i\sim q\left(X_t\middle|X_{0:t},Z_{1:t}\right)$在此基础上融合$\left\{X_t^i,x_{0:t-1}^i\sim q\left(X_{0:t-1}\middle|Z_{1:t-1}\right)\right\}$从而得到$t$时刻的采样$X_{0:t}\sim q\left(X_{0:t}\middle|Z_{1:t}\right)$。

注意到有

$$p\left(X_{0:t}\middle|Z_{1:t}\right)=\frac{p\left(Z_t\middle|X_{0:t-1},Z_{1:t-1}\right)p\left(X_{0:t}\middle|Z_{1:t-1}\right)}{p\left(Z_t\middle|Z_{1:t-1}\right)}$$

$$=\frac{p\left(Z_t\middle|X_{0:t},Z_{1:t-1}\right)p\left(X_t\middle|X_{0:t-1},Z_{1:t-1}\right)}{p\left(Z_t\middle|Z_{1:t-1}\right)}\times p\left(X_{0:t-1}\middle|Z_{1:t-1}\right) \tag{3-10}$$

$$\frac{p\left(Z_t\middle|X_{0:t},Z_{1:t-1}\right)p\left(X_t\middle|X_{0:t-1},Z_{1:t-1}\right)}{p\left(Z_t\middle|Z_{1:t-1}\right)}\times p\left(X_{0:t-1}\middle|Z_{1:t-1}\right)$$

$$\propto p\left(Z_t\middle|X_t\right)p\left(X_t\middle|X_{t-1}\right)\times p\left(X_{0:t-1}\middle|Z_{1:t-1}\right) \tag{3-11}$$

将(3-10)和(3-8)代入(3-7)，可以得到权重的更新方程：

$$w_t^i\propto\frac{p\left(Z_t\middle|X_t^i\right)p\left(X_t^i\middle|X_{t-1}^i\right)p\left(X_{0:t-1}^i\middle|Z_{1:t-1}\right)}{q\left(X_t^i\middle|X_{0:t-1}^i,Z_{1:t}\right)q\left(X_{0:t-1}^i\middle|Z_{1:t-1}\right)}$$

$$=w_{t-1}^i\frac{p\left(Z_t\middle|X_t^i\right)p\left(X_t^i\middle|X_{t-1}^i\right)}{q\left(X_t^i\middle|X_{0:t-1}^i,Z_{1:t}\right)} \tag{3-12}$$

更进一步，如果有$q\left(X_t^i\middle|X_{0:t-1}^i,Z_{1:t}\right)=q\left(X_t^i\middle|X_{t-1}^i,Z_t\right)$，则提议分布仅仅依赖于$X_{t-1}$和$Z_t$。如果只关心$p\left(X_t\middle|Z_{1:t}\right)$形式的后验分布，该假设极大简化了滤波过程，在后面的叙述中，如果没有特别说明，都以此为假设。于是，权重更新方程简化为

$$w_t^i\propto w_{t-1}^i\frac{p\left(Z_t\middle|X_t^i\right)p\left(X_t^i\middle|X_{t-1}^i\right)}{q\left(X_t^i\middle|X_{t-1}^i,Z_t\right)} \tag{3-13}$$

可以用

$$p\left(X_t\middle|Z_{1:t}\right)\approx\sum_{i=1}^{N_s}w_t^i\delta\left(X_t-X_t^i\right) \tag{3-14}$$

模拟关心的后验分布。通过不断接受表观数据来递归地传递粒子和更新其相应的权重，序列蒙特卡洛滤波不断更新后验分布。

3.4 算法退化问题

在视觉跟踪过程中，通常经过一段时间的循环运行之后，会出现某个粒子有接近于1的权重，而其余所有粒子的权重都非常小，从而意味着绝大部分计算量都对最终结果只有很小的贡献，即粒子群发生退化问题。有工作证明[39]，粒子集合的方差只能随着时间增加，所以粒子退化问题几乎是不可避免的。

为解决粒子退化问题，首先定义粒子集合退化程度的测量[40]

$$N_{eff}=\frac{N_s}{1+Var\left(w_t^{*i}\right)} \tag{3-15}$$

公式(3-15)中，$w_t^{*i}=p\left(X_t^i\middle|Z_{1:t}\right)\Big/q\left(X_t^i\middle|X_{t-1}^i,Z_{1:t}\right)$指粒子的真实“权重”。由于其无法得到，将度量近似为

$$\hat{N}_{eff}=\frac{1}{\sum_{i=1}^{N_s}\left(w_t^i\right)^2} \tag{3-16}$$

其中，w_t^i是由公式(3-13)求得的归一化权值。由公式(3-15)，有$N_{eff}\le N_s$。越小的N_{eff}说明粒子集合退化越严重。根据公式(3-15)，可以通过增大N_s的方法避免退化问题，但同时所带来的计算量增长影响了该策略的实用性。

在实际的视觉跟踪问题中，主要采用以下两种策略：

a) 选择合适的提议分布；

b) 对粒子集合进行重新采样。

3.5 提议分布问题

在X_{t-1}^i和Z_t确定的情况下，能够控制公式(3-16)中权重方差增加的最优提议分布应满足如下的形式[39]：

$$q\left(X_t\middle|X_{t-1}^i,Z_t\right)_{opt}=p\left(X_t\middle|X_{t-1}^i,Z_t\right) \tag{3-17}$$

$$w_t^i \propto w_{t-1}^i \frac{p\left(Z_t \middle| X_t^i\right) p\left(X_t^i \middle| X_{t-1}^i\right)}{q\left(X_t^i \middle| X_{t-1}^i, Z_t\right)} \tag{3-18}$$

将(3-18)带入(3-13)有

$$w_t^i \propto w_{t-1}^i p\left(Z_t \middle| X_{t-1}^i\right)$$

$$= w_{t-1}^i \int p\left(Z_t \middle| X_t'\right) p\left(X_t' \middle| X_{t-1}^i\right) dX_t' \tag{3-19}$$

于是可以得到， 无论如何从 $q\left(X_t \middle| X_{t-1}^i, Z_t\right)_{opt}$ 中采样， 只要 X_{t-1}^i 确定，则 $w_t^i \propto w_{t-1}^i$，即有 $Var\left(w_t^i\right) = Var\left(w_{t-1}^i\right)$。也就是说，最优的提议分布也仅仅能使粒子集合的方差不增大。

但是， 找到最优提议分布存在两个困难， 需要从未知的分布 $p\left(X_t \middle| X_{t-1}^i, Z_t\right)$ 中进行采样和求取积分 $\int p\left(Z_t \middle| X_t'\right) p\left(X_t' \middle| X_{t-1}^i\right) dX_t'$。最优提议分布一般只有在两种特殊情况下是可得的。一是当 X_t^i 的取值属于一个有限集合，则积分问题变为求和问题。另一种情况是当状态表观关联模型 $p\left(Z \middle| X\right)$ 是线性变换的时候，则 $p\left(X_t \middle| X_{t-1}^i, Z_t\right)$ 满足高斯分布，即

$$X_t = \mathrm{f}_t\left(X_{t-1}\right) + \mathrm{V}_{t-1} \tag{3-20}$$

$$Z_t = \mathrm{H}_t X_t + \mathrm{n}_t \tag{3-21}$$

其中， $\mathrm{f}_t : R^{n_x} \to R^{n_x}$ 是非线性映射， $\mathrm{H}_t \in R^{n_z \times n_x}$ 是观测矩阵。

$$\mathrm{V}_{t-1} \sim N\left(\mathrm{V}_{t-1}, \mathrm{o}_{n_u \times 1}, \mathrm{Q}_{t-1}\right) \tag{3-22}$$

$$\mathrm{n}_t \sim N\left(\mathrm{n}_t, \mathrm{o}_{n_n \times 1}, \mathrm{R}_t\right) \tag{3-23}$$

定义

$$\sum_t^{-1} = Q_{t-1}^{-1} + H_t^T R_t^{-1} H_t \tag{3-24}$$

$$m_t = \sum_t \left(Q_{t-1}^{-1} f_t(X_{t-1}) + H_t^T R_t^{-1} Z_t \right) \tag{3-25}$$

可以推导出：

$$p\left(X_t \middle| X_{t-1}, Z_t\right) = N\left(X_t; m_t, \sum_t\right) \tag{3-26}$$

$$p\left(Z_t \middle| X_{t-1}\right) = N\left(Z_t; H_t f_t(X_{t-1}), Q_{t-1} + H_t R_t H_t^T\right) \tag{3-27}$$

对其他的各种情况，通常无法得到解析解。然而，通过局部线性化获得次优解是可能的。比如用高斯模型逼近分布 $p\left(X_t \middle| X_{t-1}, Z_t\right)$ 或者采用unscented transform来估测对高斯模型的近似程度[41]。通常，采用这样的提议分布，可以在取得相同性能的情况下大大减少所需粒子的数量。

需要指出的是，由于直观和容易实现，一种常用的提议分布为状态转移模型，即

$$p\left(X_t \middle| X_{t-1}^i, Z_t\right) = p\left(X_t \middle| X_{t-1}^i\right) \tag{3-28}$$

将(3-28)代入(3-18)，可以得到

$$w_t^i \propto w_{t-1}^i p\left(Z_t \middle| X_t^i\right) \tag{3-29}$$

3.6 粒子重采样技术

另一种避免粒子集合退化的方法是在退化现象发生的情况下采用重采样技术更新粒子集合(比如，当 $\hat{N}_{eff}$ 小于某一指定阈值)。重采样的基本思想是去除权值较小的粒子从而使权值较大的粒子获得进化或者生存的机会。其基本方法是通过对 $p\left(X_t \middle| Z_{1:t}\right)$ 在离散点的近似表示形式

$p(X_t|Z_{1:t}) \approx \sum_{i=1}^{N_s} w_t^i \delta(X_t - X_t^i)$ 进行 Ns 次有放回的独立采样生成新的粒子集合 $\{X_t^{i^*} | i = 1,\ldots,N_s\}$ 并且满 $\Pr(x_t^{i^*} = x_t^j) = w_t^j$。既然新粒子集合是从 $p(X_t|Z_{1:t})$ 的独立采样，可以重置所有粒子的权值为 $w_t^{i^*} = 1/N_s$。

然而，重采样算法在解决粒子集合退化问题的同时，也带来了新的问题。首先，重采样限制了算法的并行性；另外，如果一个粒子具有很高的权值，则其将被采样多次，使新粒子集合存在相同粒子的多个克隆版本，从而影响粒子集合的多样化问题。该问题被称为“采样贫瘠”(sample impoverishment)，尤其在过程噪声较小的情况下比较严重。第三个问题是当粒子多样性受到限制的时候，给予粒子的估测将发生退化。解决该问题的一种方法是假设粒子的状态变化是已由前向滤波器确定，然后在采用后向滤波器重新计算粒子的权值。另一种方法是采用基于马尔科夫链的蒙特卡洛模拟算法(Markov Chain Monte Carlo, MCMC)。

近些年来，针对采样贫瘠问题提出了一些解决技术。一种方法是 resample-move 算法[42]。另一种广泛应用的策略是基于正则化的方法[43]。

3.7 粒子分布的一些方法

跟踪过程中更常见的情况是无法获得最优提议分布。例如当采用状态转移模型 $p(X_i|X_{i-1})$ 相对于状态观测关联模型 $p(Z_k|X_k)$ 具有平坦得多的分布形式，势必造成大多数的粒子具有较低的权值分布。部分学者提出了能使粒子运动到“适当区域”的方法。Clapp 和Oudjane 都是在先验分布和相似性分布之间引入中间分布[44,45]。粒子根据在这些中间分布上的权值进行重采样，从而“驱赶”粒子运动到特征空间中正确的区域。当呈尖峰状分布 $p(Z_k|X_k)$ 但能够分解为诸多较平坦分布的时候，分区采样是有效的解决粒子权值过小

问题的办法[46]。

3.8 本章小结

在本章中，系统地介绍了序列蒙特卡洛滤波算法。序列蒙特卡洛滤波算法虽然具有在解决复杂视觉跟踪问题上的通用性，然而，这种通用性也限定了算法从根本上说只是一个理论框架，即只给出了解决视觉跟踪问题的总体策略。如何有效地应用序列蒙特卡洛滤波算法解决具体的视觉目标跟踪问题，仍然有很多应用相关的问题需要探讨。序列蒙特卡洛滤波算法是一个复杂的贝叶斯理论问题。目前对此研究的很多，也有很多变种的粒子滤波及各种改进的方法。如何让它更有效地应用到目标跟踪，是本文所要讨论的问题。

第 4 章 交换采样粒子滤波技术

4.1 引言

如第三章所言，基于序列蒙特卡洛滤波方法的目标跟踪得到越来越多的研究。同时在复杂的环境跟踪下取得了很好的效果。但粒子滤波也存在很多问题，在跟踪中也遇到了很多困难，使得需要对粒子滤波进行改进。于是出现了很多变种的粒子滤波如辅助粒子滤波、核粒子滤波、隐马尔科夫-粒子滤波等。影响粒子滤波在跟踪中的效果的主要因素在于对粒子的重采样，如何更好的采样才能更精确的模拟状态的后验概率分布。因此本文对粒子滤波研究的重点也在于粒子重采样。

为了更好地模拟状态的后验概率分布，本文采用了 PF 和 APF 在线交换采样的方法。本文的粒子滤波算法借用了 Michael Isard 的 Condensation 算法。Condensation 算法是粒子滤波(PF)算法的一个应用实例，在目标跟踪中有很成功的运用。PF 是一种非线性非高斯方法，它不需要像 kalman 滤波方法样要求系统是高斯线性的。因此 PF 有很广泛的应用，并且用在运动目标跟踪中可以很好地应对复杂背景下的跟踪。但是在运动目标跟踪中，粒子滤波有时也不能很好的来正确模拟后验概率分布，因此有时也会有跟丢的情况发生。最近提出了很多对传统粒子滤波的改进，产生了很多粒子滤波的变种。辅助粒子滤波 APF 就是其中一种。相对于 PF 更多地侧重于对于目标状态转换的预测，APF 则偏重于当前的观察。APF 是一种依靠最近观察值设计提议性分布的方法。它借用了一个辅助变量来近似后验概率密度。APF 通过考虑粒子的一步预测似然，结合最新观察值，一步预测似然大的粒子进行状态转移。它促使粒子更多地向高相似区域分布。本文将两者结合起来，以当前估计状态的信任度为依据，当信任度高时算法采用 Condensation 采样，而当信任度低时，算法采用 APF 采样，来引导粒子更好的分布。通过结合这两种算法的优点，本文的算法力求加强跟踪的鲁棒性和准确性。

本章主要介绍了 Condensation 算法和 APF 算法，并介绍了在线交换采样

粒子滤波技术的思想。最后给出了实验结果和分析。

4.2 Condensation 算法

本文所采用的粒子滤波算法是 Michael Isard 和 Andrew Blake 于 1997 年 3 月发表的“CONDENSATION—Conditional density propagation for visual tracking”经典论文所用的 Condensation 算法[15]。

4.2.1 Condensation 在目标跟踪中的应用

给定一个观察 Z_k，目标就是估计状态 X_k。我们不在动态系统的本性这一点上做假设。基于 MMSE 的最优估计由下列条件平均数给出：

$$\widehat{X_k} = E\left[X_k \middle| Z_0^k\right] \tag{4-1}$$

Z_0^k 是到 0 到 k 时间的一系列观察。估计这个期望值需要一个后验概率 $p\left(X_k \middle| Z_0^k\right)$。有了这个后验概率我们就可以得到 MMSE 的估计。通常想知道后验概率需涉及到贝叶斯方法。我们可以通过下面的这些式子来递归估计。

$$p\left(X_k \middle| Z_0^k\right) = \frac{p\left(X_k \middle| Z_0^{k-1}\right) p\left(Z_k \middle| X_k\right)}{p\left(Z_k \middle| Z_0^{k-1}\right)} \tag{4-2}$$

其中，

$$p\left(X_k \middle| Z_0^{k-1}\right) = \int p\left(X_k \middle| X_{k-1}\right) p\left(Z_k \middle| X_k\right) dX_{k-1} \tag{4-3}$$

归一化常量 $p\left(Z_k \middle| Z_0^{k-1}\right)$

$$p\left(Z_k \middle| Z_0^{k-1}\right) = \int p\left(X_k \middle| Z_0^{k-1}\right) p\left(Z_k \middle| X_k\right) dX_k \tag{4-4}$$

这个回归式说明了当前状态是先验概率和最近观察的函数。状态空间又是根据状态转移概率 $p\left(X_k \middle| X_{k-1}\right)$ 和观察概率密度 $p\left(Z_k \middle| X_k\right)$ 得到的。特别地，

$p(X_k|X_{k-1})$由下面的状态更新方程决定的

$$X_{k+1}=\mathrm{f}(k,X_k)+w_k \tag{4-5}$$

举个例子，给定一个附加的高斯噪声模型$p(w_k)=\mathrm{N}(0,\mathrm{R}^v)$，则

$p(X_k|X_{k-1})=\mathrm{N}(\mathrm{F}(X_{k-1}),\mathrm{R}^v)$。类似地，$p(Z_k|X_k)$由下面的观察方程得到

$$Z_k=\mathrm{h}(k,X_k)+V_k \tag{4-6}$$

理论上，知道这些概率密度和初始条件$p(X_0|Z_0)=\dfrac{p(Z_0|X_0)p(X_0)}{p(Z_0)}$就可以对所有的$k$推出$p(X_k|Z_0^k)$。但不幸的是，对于式(4-3)和式(4-4)的多维积分在大多数系统中是难处理的。通常使用的方法是应用蒙特卡洛采样技术将积分转变成有限集。

粒子滤波被用来跟踪混乱中的目标和随着时间变化的感兴趣区域，特别是在非高斯或多维空间下。基本方法是构造一个基于采样样本来代表整个概率密度。然后根据一系列状态预测和校正，得到后验概率。每一个状态更新的预测根据式(4-5)进行。

多假设状态变量X_k，每个假设即一个粒子，每个粒子绑定了一个该粒子的权重，也就是该粒子对状态的贡献。对状态的计算是由这些粒子的求和得到的。粒子滤波是一个递归算法$\mathrm{f}(k,X_k)$，在预测阶段，包含进了附加的随意噪声w_k为了更好的模拟变量变化的效果。然后，每个粒子的权重根据当前得到的观察进行更新，这个是更新阶段。在重采样的阶段，那些权值较小的粒子将会被剔除。更具体的，在本文中，状态即为物体的中心坐标$X_k=[x_k,y_k]$。在时间系列k，状态是由M个采样点（即：粒子）组成的集合所代表的$S_k^i=\left[X_k^i,b_k^i\right]:i=1,2,...,M$，$i$代表粒子数，每个粒子由该粒子所采样的状态$X_k$和该粒子的权值$b_k$，即该粒子对最终状态估计的贡献组成。粒子权值满足$\sum_{i=1}^{M}b_k^i=1$。

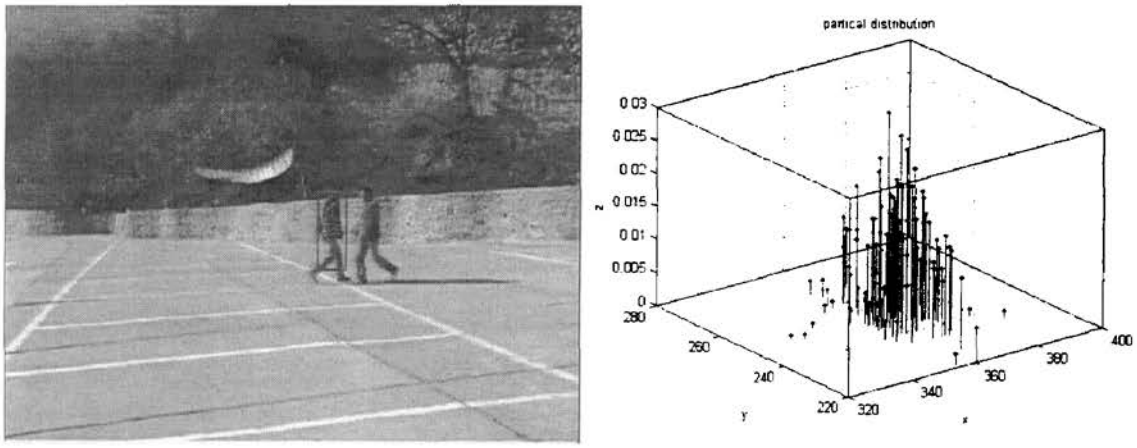


图 4.1 粒子采样分布图

Condensation 算法

迭代：

从 $t-1$ 时刻的粒子集合 $\left\{s_{t-1}^{(n)},\pi_{t-1}^{(n)},c_{t-1}^{(n)},n=1,\ldots,N\right\}$ 构造出新的粒子集合

$\left\{s_t^{(n)},\pi_t^{(n)},c_t^{(n)}\right\},n=1,\ldots,N$。

按照下面步骤产生 N 中的第 n^{th} 个粒子：

1. **选择：**按照下面步骤选择 $s_t^{\prime(n)}$：

(a) 产生均匀分布的随机数 $\gamma\in[0,1]$。

(b) 采用二分查找法找出最小的 j 使得 $c_{t-1}^{(j)}\geq\gamma$。

(c) 赋值 $s_t^{\prime(n)}=s_{t-1}^{(j)}$。

2. **预测：**由 $p\left(X_t\middle|X_{t-1}=s_t^{\prime(n)}\right)$ 得到预测状态 $s_t^{(n)}$.例如，根据状态转移方程 $s_t^{(n)}=As_t^{\prime(n)}+Bw_t^{(n)}$,其中 $w_t^{(n)}$ 是噪声变量，且 BB^T 是噪声协方差。

3. **量测：**根据得到的观察值计算粒子的权值 $Z_t:\pi_t^{(n)}=p\left(Z_t\middle|X_t=S_t^{(n)}\right)$，

并归一化 $\sum_n\pi_t^{(n)}=1$ 按照下式累计粒子权值的分布 $\left\{s_t^{(n)},\pi_t^{(n)},c_t^{(n)}\right\}$ 其中

$$c_t^{(0)}=0$$

$$c_t^{(n)} = c_t^{(n-1)} + \pi_t^{(n)} \left(n = 1, \ldots, N\right)$$

一旦粒子的状态预测出来，粒子的权值计算完毕，就可以根据下式计算t时刻最终预测出来的状态：$\varepsilon\left[f\left(X_t\right)\right] = \sum_{n=1}^{N} \pi_t^{(n)} f\left(s_t^{(n)}\right)$。

4.3 Auxiliary Particle Filter 算法

Michael Pitt 在 1997 年 10 月的一篇论文[47] "Filtering via simulation: Auxiliary particle filter"提出了对标准粒子滤波的改进，即 APF。

APF 的核心就是，它所用到的重要性概率密度是依赖于最近的观察。它对粒子滤波近似后验概率进行了扩展。

$$\hat{p}\left(X_{t+1} \middle| Z_{1:t+1}\right) \propto \sum_{n=1}^{N} w_t^n p\left(Z_{t+1} \middle| \mu_{t+1}^n\right) p\left(X_{t+1} \middle| X_t^n\right) \tag{4-7}$$

其中，μ_{t+1}^n是由状态转换模型$p\left(X_{t+1} \middle| X_t^n\right)$得到。算法从混合概率中进行$m = 1 \ldots N$次采样。得到的粒子其权重按下式计算：

$$w_{t+1}^m \propto \frac{p\left(Z_{t+1} \middle| X_{t+1}^m\right)}{p\left(Z_{t+1} \middle| \mu_{t+1}^{n_m}\right)} \tag{4-8}$$

其中$\mu_{t+1}^{n_m}$是在第m次采样后从$p\left(X_{t+1} \middle| X_t^{n_m}\right)$得到的值。APF 就是从依赖于当前观察的重要性概率密度来采样，然后根据这个采样的粒子来模拟后验概率密度。下面给出了 APF 的算法。

APF 算法

对于每个时刻 $n = 1, 2, \cdots$

1．对$i = 1, \cdots, N_p$,计算 $\mu_n^{(i)}$ 如，$\mu_n^{(i)} = \mathrm{E}\left[p\left(\mathrm{x}_n \middle| \mathrm{x}_{n-1}^{(i)}\right)\right]$

2．对$i = 1, \cdots, N_p$,计算第一次粒子的权重 $W_n^{(i)} = W_{n-1}^{(i)} p\left(\mathrm{y}_n \middle| \mu_n^{(i)}\right)$

并且归一化 $\tilde{W}_n^{(i)} = \dfrac{W_n^{(i)}}{\sum_{j=1}^{N_p} W_n^{(j)}}$.

3．采用 Condensation 算法中的重采样方法得到新的粒子状态$\left\{\mathrm{x}_n^{(i)}, \xi^i\right\}_{i=1}^{N_p}$

4．对$i = 1, \cdots, N_p$,有新的$\mathrm{x}_n^{(i)} \sim p\left(\mathrm{x}_n \middle| \mathrm{x}_{n-1}^{(i)}, \xi^i\right)$,计算第二次粒子的权值根据

$$w_{t+1}^m \propto \frac{p\left(Z_{t+1} \middle| X_{t+1}^m\right)}{p\left(Z_{t+1} \middle| \mu_{t+1}^{n_m}\right)}。$$

4.4 基于 PF 和 APF 的交换采样方法

在这一节中，本文提出了一种新的粒子滤波采样策略，在线交换采样，它能够在实时视频跟踪中达到鲁棒性和准确性。它使用了相互互补的两个采样策略，PF 和 APF，依据状态估计信任度在线的选择采样。

我们集中于两个粒子滤波算法：Condensation 和 Auxiliary Particle Filter。Condensation 算法是在实时跟踪中应用比较成功的粒子滤波，成功的应用有 RoboCup 机器人视觉[48], 会议室的人脸跟踪[49]等。Condensation 也有它固有的缺点即外部数据问题。在先验概率分布和后验概率分布之间巨大的区别会引起对后验概率分布不精确的估计。而在实际的目标跟踪中外部问题通常是由于无法预料的遮挡、干扰和缺乏运动目标足够的信息所造成的。因此 APF 借助于当前的观察信息用来解决这个外部问题。如后面所说，这两个粒子滤波器在精确性和鲁棒性上是互补的。本文用一个相对简单的原则来交换使用两种粒子滤波来采样。

对于估计当前目标状态，Condensation 强调状态转换的预测比 APF 多一点，而 APF 则强调观察要多一点。强调观察的 APF 使得跟踪精确些，相反，Condensation 中的状态转换预测使得跟踪鲁棒性强一些，特别是在遮挡和干扰的情况下。为了提取两者的长处，我们提出了一种基于状态估计信任度的交换采样策略。考虑到计算机资源的条件，我们采用在线的交换采样策略。

本文提出了两种不同的交换采样方法：硬性转换方法和软性转换方法。硬性转换方法是一种突变的转换策略。软性转换方法则是一种平滑过渡性的转换策略。

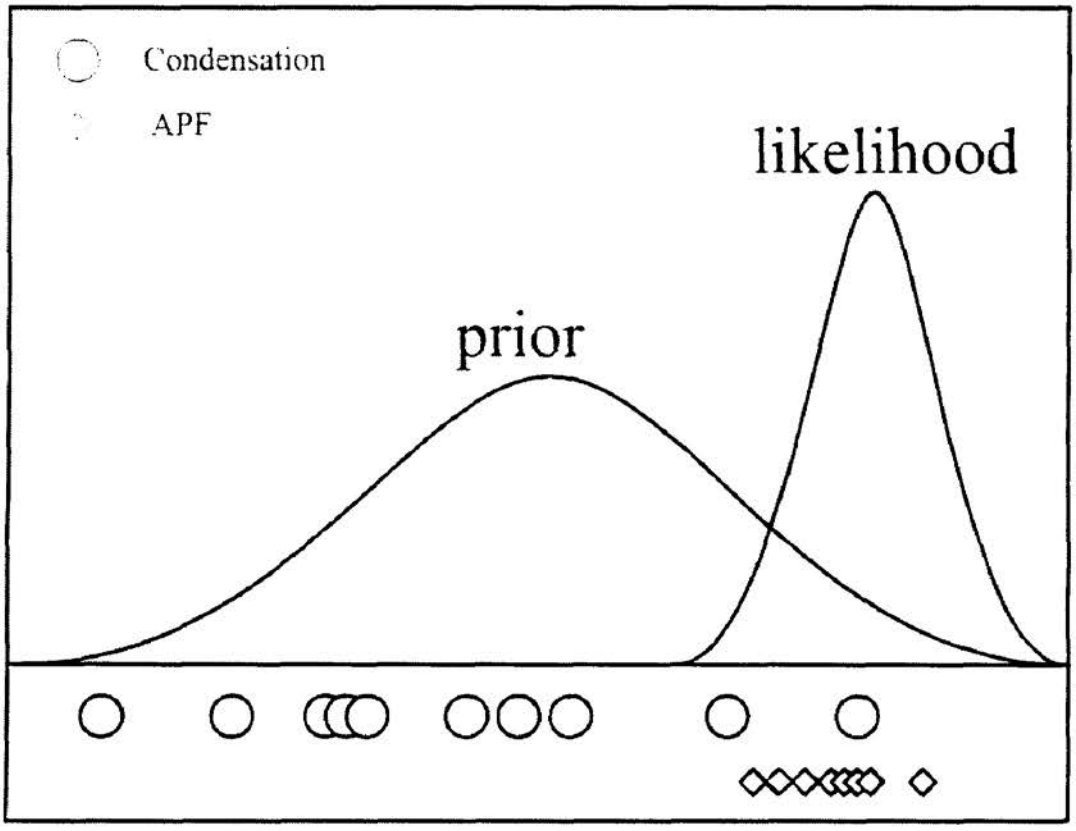


图 4.2 粒子从提议分布采样

4.4.1 硬性转换方法

本文根据一个简单的原则，目前被估计状态的的信任度，来完成从一种采样转换到另一种采样。

当前被估计状态的信任度被定义为

$$V_{est}(t)=\left(\sigma_{est,1}^{2},\ldots,\sigma_{est,N_s}^{2}\right)^{T} \tag{4-9}$$

$$=\mathrm{diag}\left\{\frac{\sum_{n} w_t^{(n)}\left(\hat{\mathrm{x}}_t-\mathrm{x}_t^{(n)}\right)\left(\hat{\mathrm{x}}_t-\mathrm{x}_t^{(n)}\right)^{T}}{N-1}\right\} \tag{4-10}$$

其中

$$\hat{x}_t = \sum_{n=1}^{N} w_t^{(n)} x_t^{(n)} \tag{4-11}$$

$\mathrm{diag}(M)$标识了组成M矩阵对角线的向量。如果$x_i\,(i=1,\ldots,N_x)$被简单的假设为相互独立的，则当前被估计的协方差为 0. 如果状态向量的元素独立性是不成立的，则$V_{est}(t)$一定会扩展成一个方差-协方差矩阵。然后设定一个阈值$\gamma=\{\gamma_1,\ldots,\gamma_{N_x}\}^T$。转换策略可以定义为如下：

如果 $(\exists i)\sigma_{est,i} > \gamma_i$，则使用 APF 采样，否则使用 Condensation 采样。

向量γ是一个阈值，它依靠计算当前被估计状态的信任度来判断是否跟踪成功。向量$V_{est}(t)$越大，估计信任度越低，说明现在的粒子分布情况趋于错误分布，这时要根据观察让 APF 引导粒子到高相似区域去，反之亦然。

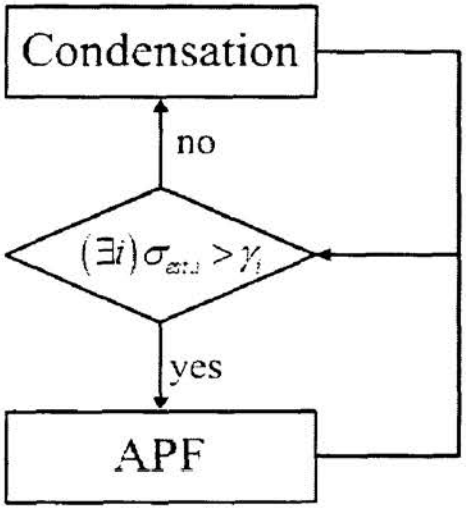


图 4.3 硬性转换方法

4.4.2 软性转换方法

软性转换方法和硬性转化方法类似，只不过它根据信任度水平来控制两种采样方式对分别所分配粒子的使用比例。两种采样方式所采样的粒子数目N_{cond}和N_{apf}定义如下：

$$N_{cond} = N - N_{apf}\,,\quad N_{apf} = \max_i N_{apf,i} \tag{4-12}$$

$$\begin{cases} N_{apf,i}=0 & \left(\sigma_{est,i}<\gamma_{\min,i}\right) \\ N_{apf,i}=\dfrac{N\left(\sigma_{est,i}-\gamma_{\min,i}\right)}{\gamma_{\max,i}-\gamma_{\min,i}} & \left(\gamma_{\min,i}\le\sigma_{est,i}\le\gamma_{\max,i}\right) \\ N_{apf,i}=N & \left(\sigma_{est,i}>\gamma_{\max,i}\right) \end{cases} \tag{4-13}$$

它的意思是越大的信任度，由 Condensation 采样的粒子越多，反之亦然，越小的信任度则由 APF 采样的粒子越多。

这两种转换策略动态地利用了 Condensation 和 APF 的优点。其基本思想是如果跟踪被认为是成功的则更强调于先验，如果不被认为是成功的则强调观察。

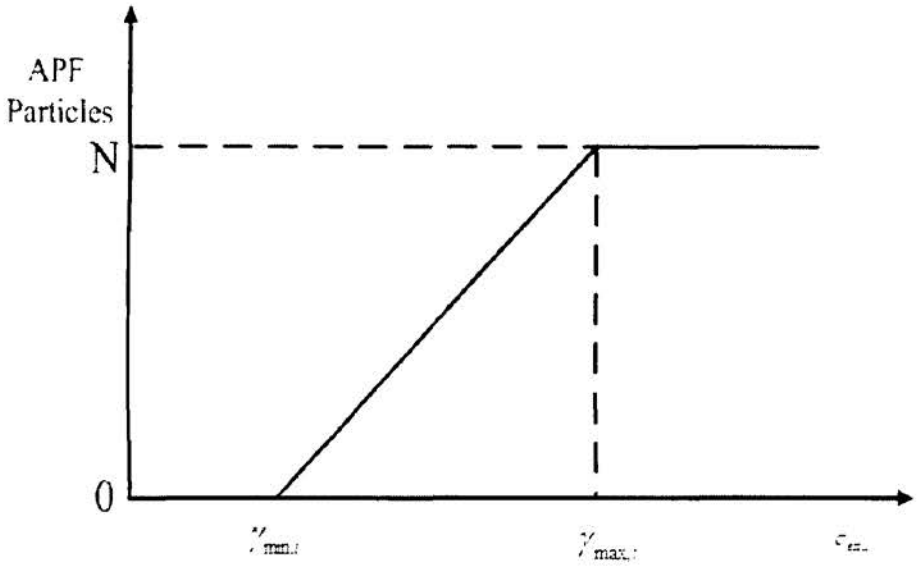


图 4.4 软性转换方法

4.5 实验结果及分析

本文做了大量的实验，这些实验主要集中在遮挡、干扰、前景背景混淆、摄像机晃动等这几个跟踪中的难点。本文所用的实验数据包括 CMU 的公共测试集视频和自己采集的视频。本文对实验结果做了定量分析和定性分析。定量分析主要是鲁棒性和精确性分析，并且这两个分析是针对硬性转化方法和软性转换方法进行的。最后的整体跟踪情况分析是做的定性分析。图 4.6 和图 4.8 的跟踪误差分析，纵坐标采取了水平方向和垂直方向误差相加除以

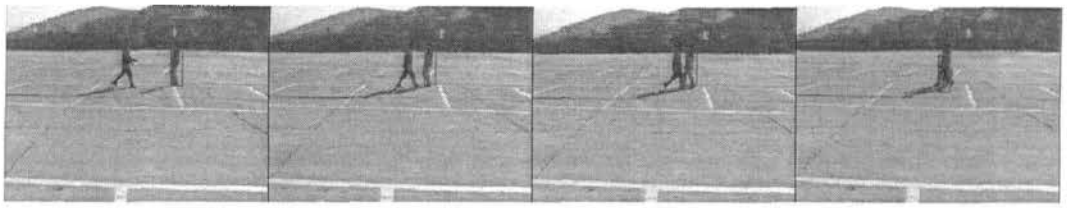
2，并相对于目标本身大小来作为衡量标准的。横坐标为对应的帧号。

4.5.1 鲁棒性分析

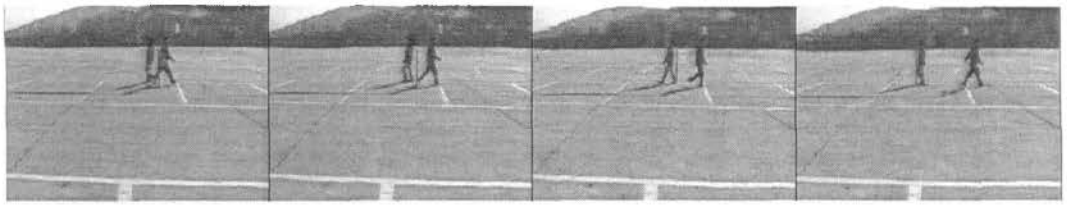
本文的鲁棒性分析主要针对于遮挡情况下，跟踪能否有效的跟踪到目标，不同于其他处理遮挡的算法[50]，本文没有做专门的遮挡处理。根据实际经验本文定义的鲁棒性为：

$$if\left(\hat{x}_i - s_t\right) > \left(width/2, height/2\right), sum_{false} = sum_{false} + 1 \tag{4-14}$$

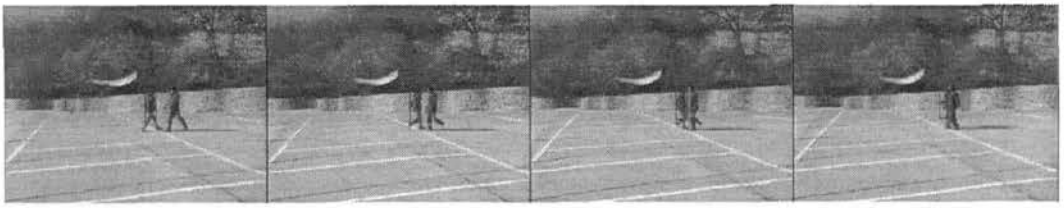
我们人工标定目标的真实位置 s_t，然后拿跟踪到的目标位置 $\hat{x}_i$ 和真实的位置做比较，然后统计跟踪中目标位置偏离真实位置达到跟踪框的 1/2 长和 1/2 宽的帧数。为了简单起见，我们只统计遮挡附近的帧的情况。我们选取典型的两个遮挡实验来做统计。



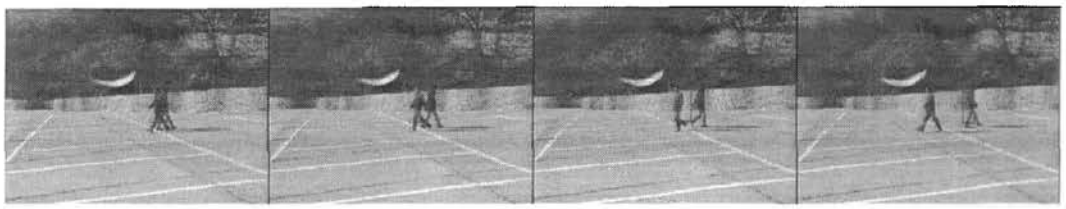
视频-4　100 帧　110 帧　114 帧　119 帧



123 帧　125 帧　128 帧　135 帧



视频-5　431 帧　434 帧　436 帧　439 帧



442 帧　　445 帧　　448 帧　　455 帧

图 4.5 测试鲁棒性的两个视频在发生遮挡时的帧图片

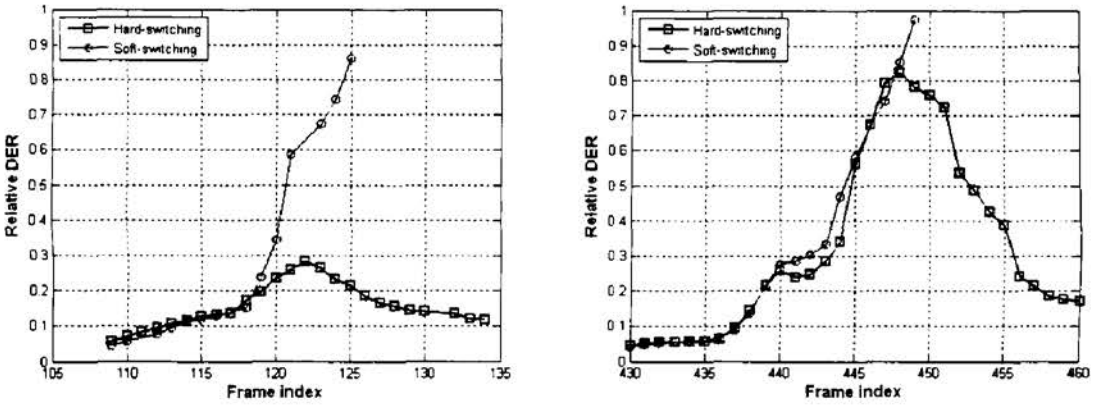


图 4.6 两种方法在视频-4 和视频-5 上的跟踪目标位置误差分析图

表 4.1 两种方法在视频-4 和视频-5 的失败帧统计表

视频	方法	总帧数	失败帧数
视频-4	硬性	24	0
	软性		11
视频-5	硬性	28	5
	软性		14

图 4.5 所示的两个视频的实验效果图都是硬性转换方法的实验结果，可以看出硬性转换方法在遮挡后，目标没有丢失。软性转换方法在这两个视频上均无法成功处理遮挡。视频-4 和视频-5 两者都是交叉遮挡，并且两个目标颜色相近。图 4.6 是误差统计，在遮挡时，软性转换方法偏离真实位置越来越远，最后丢失目标。但是硬性转换方法可以很好的误差在控制的范围内。表 4.1 表示了失败帧的帧数，可以看出硬性转换方法只有在视频-5 有失败帧，而软性转换方法在遮挡结束后就丢失了目标，后面的帧皆为失败帧。

从以上的定性和定量分析，说明在有遮挡时，硬性转换方法要比软性转

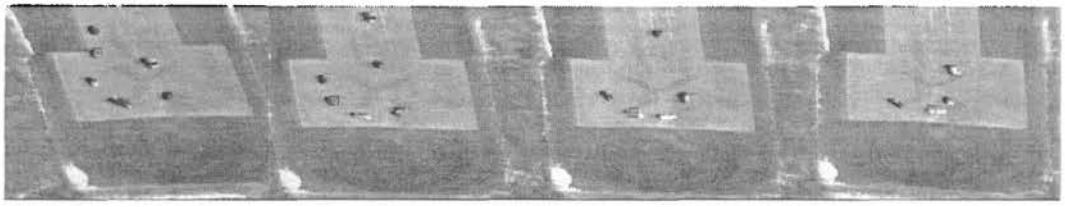
换方法要鲁棒。在遮挡时，PF 的鲁棒性要比 APF 要好，因为 PF 强调了转移概率，而 APF 强调了观察，但是在遮挡时观察都是错误的，这样 APF 会误导粒子偏向于遮挡物（也就是看得见的物体），这样误差会越来越大，以致无法校正。软性转换方法比较平滑，容易有部分粒子进入 APF 采样，因此进入遮挡的时候 APF 容易误导粒子的分布，所以会发生跟丢现象，因为它强调了可看得见的观察但它并不知道跟踪失败。再看视频-5 中的跟踪，硬性转换方法在遮挡即将结束时也有 5 帧的失败帧，但是然后纠正了过来又跟踪上了目标。这是因为在错误帧中，也有很多粒子分布在目标上，还有一部分分布到跟踪框，在跟踪失败时 APF 会让分布在目标上的粒子强调观察，致使错误可以校正过来，跟踪到目标。

4.5.2 精确性分析

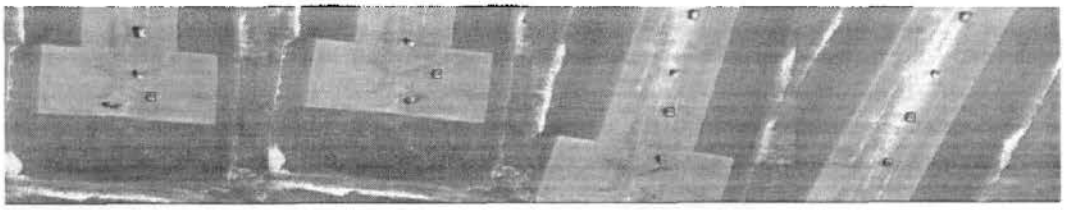
本文的精确性分析不针对遮挡的情况。因为遮挡的情况软性转换方法的鲁棒性不够好。本文针对精确性做了两个实验，通过计算一系列帧中跟踪的平均误差率来看这两种方法的精确程度。

$$\text{Tracking error} = \frac{1}{L}\sum_{i=1}^{L}\left|\hat{x}_i - s_i\right| \tag{4-15}$$

s_i 表示目标的真实位置，$\hat{x}_i$ 是跟踪得到的目标位置，L 为视频的总帧数。



视频-6　100 帧　200 帧　250 帧　300 帧

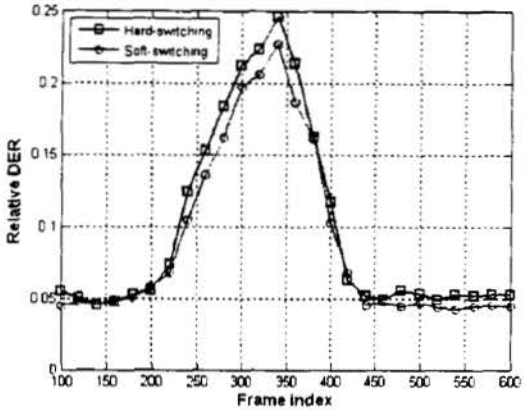


350 帧　400 帧　500 帧　600 帧

视频-7　900 帧　1100 帧　1200 帧　1300 帧

1400 帧　1500 帧　1600 帧　1700 帧

图 4.7 测试精确性的视频-6 和视频-7 的效果图



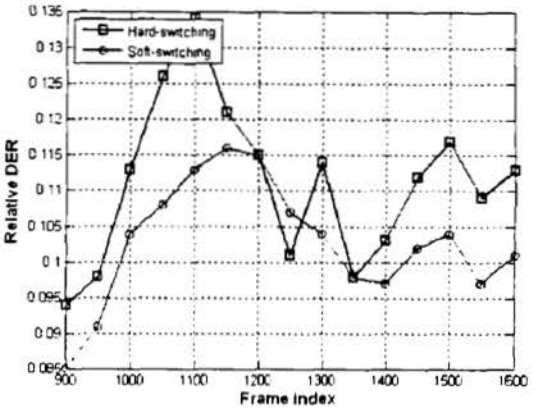


图 4.8 两种方法在视频-6 和视频-7 上的跟踪目标位置误差分析图

表 4.2 两种方法在视频-6 和视频-7 上平均误差率统计

视频	方法	总帧数	平均误差率
视频-6	硬性	600	2.3246
	软性		1.8327
视频-7	硬性	651	1.9878
	软性		1.5361

图 4.7 是软性转换方法在视频-6 和视频-7 实验中的跟踪效果图，硬性转换方法在这两个实验上的效果和软性转换方法几乎一样，在实验效果图上没有差别，只是在跟踪精度上，有稍微的差别。从图 4.8 和表 4.2 可以看出，软

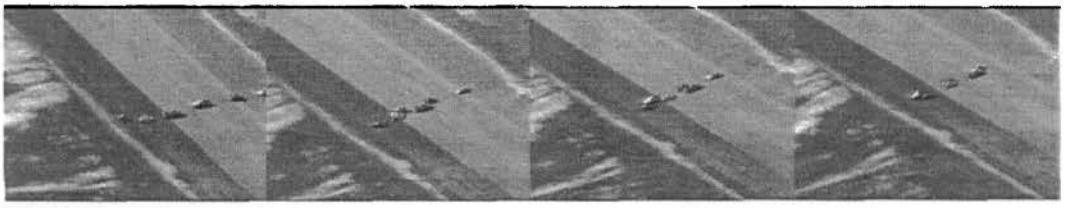
性转换方法的精确度要比硬性转换方法高一些。图 4.8 中的左图有一段精确度偏低是因为本文的实验没有考虑尺度缩放，所以在视频-6 的 250 帧到 350 帧，车掉头时只能跟踪车的尾部，因此这一段误差较大。

从以上的定性和定量分析中可以看出，在没有遮挡发生时，在精度上，软性转换方法要好于硬性转换方法。硬性转换方法是突然转换，它强调 PF 算法多一些，而软性转换方法的转换要平滑一些，也就是它比硬性转换方法更易处于过度阶段，会有一部分粒子跟踪中转向 APF 算法，因此在强调观察的跟踪中精度会高一点，因为 APF 会在高相似区域采样。同时，这也是为什么在有遮挡的情况下软性方法会失败的原因。

4.5.3 整体跟踪情况分析

视频-8 400 帧　　444 帧　　457 帧　　473 帧

476 帧　　490 帧　　510 帧　　520 帧



视频-9 440 帧　　530 帧　　560 帧　　590 帧

620 帧　　640 帧　　660 帧　　690 帧

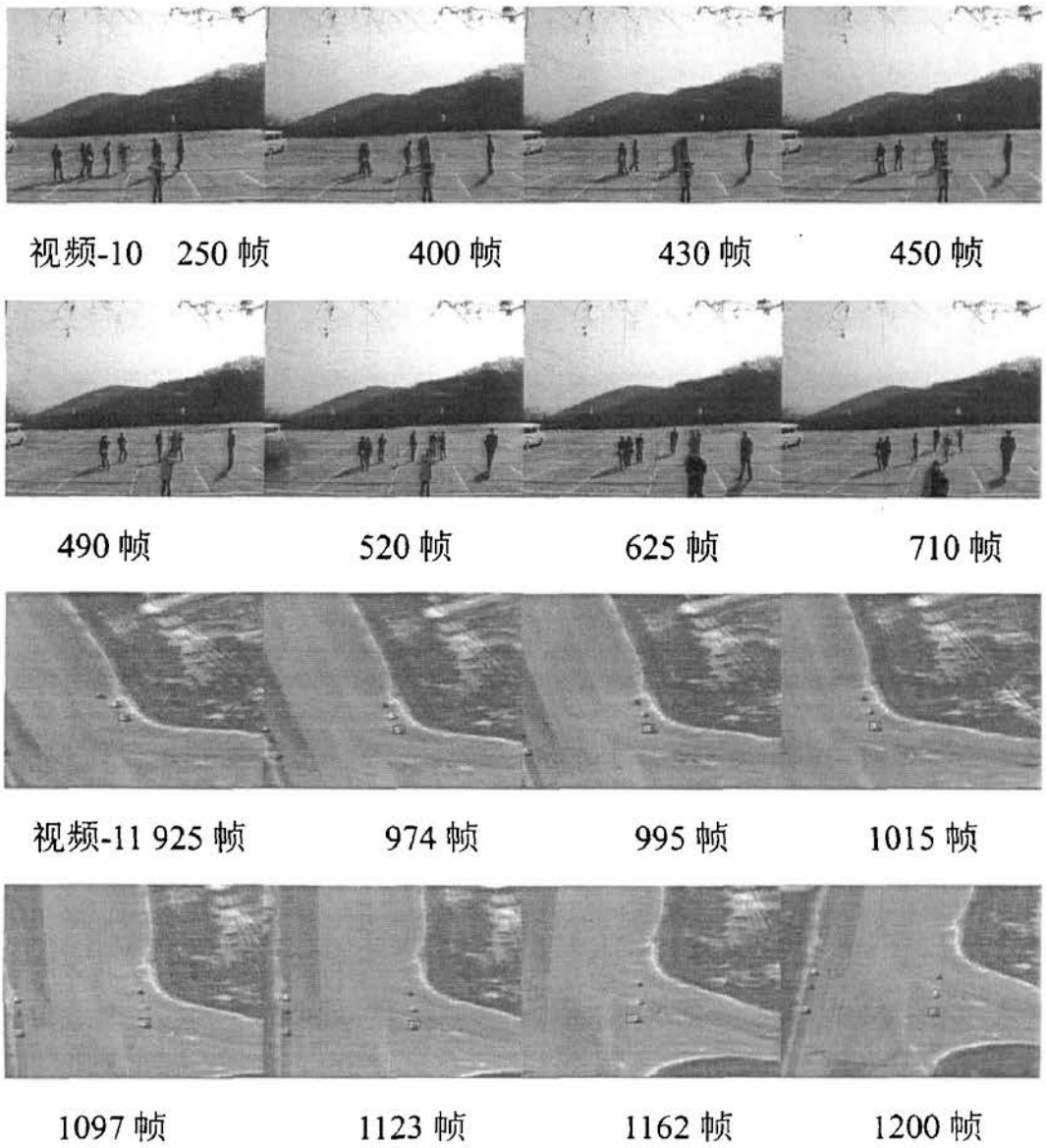


图 4.9 交换采样粒子滤波算法的效果图

图 4.9 是本文做的四个典型的难度较高的跟踪实验的效果图，以此来说明本文所提出的交换粒子滤波算法有很好的跟踪效果，由于在精度上，硬性转换方法只是比软性转换方法稍逊一筹，而在鲁棒性上硬性转换方法有很好的处理遮挡的能力，因此本文选用硬性转换方法作为代表，以上四个实验都是硬性转换方法的实验结果。这四个视频跟踪效果图分别展示了 Condensation 算法、HOGC 特征、APF 算法、粒子滤波的优点。

视频-8 是经典的公共测试集视频，这个视频有很高的难度。第 400 帧中，红、绿、蓝三种颜色框锁定的目标。在这个视频中三人有交叉遮挡和目标混淆，最难跟踪的是红框锁定的目标，它两次被遮挡，并且和蓝框目标相混淆。

从视频-8 的效果图可以看出本文的算法始终可以分别将三个目标跟踪上。在这个实验中，目标上的粒子分布都完全很好近似了高斯分布，因此计算出来的目标位置误差会小，所以能很好的跟踪到目标。由此可以看出 Condensation 算法在鲁棒性上的优点。

视频-9 是公共测试集视频，这个视频是一个具有干扰的视频。目标在 530 帧、590 帧、640 帧中分别与三辆车擦肩而过，并且第一辆和第三辆车都是同颜色的车辆。从视频-9 的效果图可以看出跟踪框的“飘移”现象没有发生，这就显示出了 HOGC 特征的优点，它的刻画目标的能力给粒子滤波算法提供了很好的基础。

视频-10 是实验室所采集的用于跟踪实验的视频。这个视频里的目标在 400 帧到 450 帧之间被其他两个目标所严重遮挡，无法发现目标。但当目标又走出来时，又可以成功的继续跟踪到目标。在遮挡时，粒子是分布在假目标上的，但是当目标又开始露出来时，会有一些粒子在目标上，却还有一些粒子停留在假目标上，这时强调观察的 APF 会引导粒子分布到真目标上，这样就会在下一帧会有跟多的粒子分布在真目标上，因此遮挡后仍然能跟踪上。这就显示了 APF 强调观察的优点，它可以将粒子引向高相似区域。

视频-11 是公共测试视频。这个视频的特点是摄像机突然左右晃动，致使跟踪跟着摇摆。第 974 帧向左，995 帧向右，1015 帧向左，1097 帧向右，依次类推。这个视频显示了粒子滤波的优点，因为这个视频在基于卡尔曼滤波的算法上是失败的。虽然视频突然晃动，所有粒子的坐标都无规律的发生了突变，但是只要有粒子在目标上，这些在目标上的粒子有很大的权值，在下一帧它就会引导更多的粒子在目标上分布，因此不会发生跟丢现象。而卡尔曼滤波只能预测有规律的运动，对突然的晃动，卡尔曼预测出的结果会有很大偏差，并且无法修复。

从以上 8 个视频测试的结果可以看出，交换采样的思想使得粒子滤波在跟踪中的应用有很大的提高。硬性转换方法强调了状态转移概率，鲁棒性强。软性转换方法中的平滑过渡策略在没有遮挡的情况下有较好的效果，它在平滑过度段会强调观察来修正粒子向高相似区域分布，精确性高。本文的算法在 Condensation 采样的基础上，让 APF 来辅助采样，加强精确性，有效的将鲁棒性和精确性结合起来。

4.5.4 讨论

在目标特征的表观上，本文的算法提出了融合颜色直方图和梯度直方图的综合直方图来刻画目标的特征。这两种直方图在刻画目标时的权值是 0.6 和 0.4。也就是本文的实验还是侧重于颜色信息多一点。但是跟踪时并非都是依赖于颜色直方图，因此有时的跟踪效果不是很好。所以在线的权值选择是一个很好的方法，能提高目标的刻画能力，为粒子滤波提供更好的基础。

在粒子滤波里的状态转移方程里，本文的算法没有用状态转移方程，只是加上了高斯扰动量。因为，我们无法预料目标的运动，目标的运动无规则性和突然性是普遍的，无法用一个固定的运动方程就可以描述出来。虽然使用运动方程有助于粒子滤波中的粒子分布，但在实际中，不规则的运动和突然的变化很难描述。

粒子滤波的算法核心就在于粒子的采样，如何采样能让它更好地模拟状态的后验概率分布。Condensation 算法就是采用重要性采样。APF 则是双重采样，这样使粒子更多地到高相似区，但牺牲了粒子的多样性。在遮挡时，得到的观察都是错误的，得到粒子的权值的分布也是不正确的，这时一定要维持粒子的多样性，而不是集中粒子在观察上。使得目标重新露出来时，有粒子分布在目标上，而这些在目标上的粒子会得到更高的权值，APF 会使得粒子集中在高相似区域，这样粒子就会在下一帧逐渐分布到目标上。而如果不采用 APF，则这些在目标上的粒子起到的作用有限。本文的遮挡只对短时的人与人的交叉遮挡有效，对长时间的大物体遮挡是无效的。

转移采样的阈值的选择是一个很重要的方面。本文的阈值 γ 是根据经验值来选择 $\gamma=[weight/2, height/2]$。$\gamma$ 值的选择对硬性转换方法的影响很明显。而在软性转换方法中，转换则是平滑的过度，特别是中间阶段 $(\gamma_{\min}-\gamma_{\max})$，一部分粒子使用 Condensation 算法采样，而另一部分则使用 APF 采样，这样使得粒子的多样性和集中性能够平衡，避免转换过于敏感。但在有遮挡发生时，APF 会由于错误的观察而对粒子的分布起误导，因此软性转换方法的跟踪效果不好。当 $\gamma_{\min}=\gamma_{\max}$，软性转换方法就是硬性转换方法。硬性转换方法在消耗资源方面要小于软性转换方法。在未来的研究中，可以在线的根据跟踪的情况来设定阈值，以此来找到最佳的成功跟踪和错误跟踪的分界线。

4.6 本章小结

本章提出了本文的核心算法-交换采样粒子滤波算法。本章首先介绍了一下 Condensation 算法和它在目标跟踪中的应用，以及 APF 算法理论。最后本章介绍了根据计算被估计粒子状态的信任度来在线选择采样方法的思想。本文采用了两种转换方式，一是硬性转换方法，即直接由一种采样方式转换到另一种采样方式；另一个是软性转换方法，即平滑过度，在过度阶段，一部分粒子由 Condensation 来采样，另一部分由 APF 来采样。同时本文也通过大量的实验来验证该方法的有效性。实验表明，硬性转换方法的鲁棒性要好于软性转换方法，软性转换方法的精确性要好于硬性转换方法。整体的跟踪结果看，将两种采样方式的优点结合起来，能够提高跟踪的鲁棒性和精确性。实验结果表明，该方法在复杂背景下仍然有很好的跟踪效果。

结 论

本文研究了运动目标跟踪的两个主要方面：目标表观和跟踪框架。在目标表观上，研究综合直方图（颜色直方图和梯度方向直方图的融合）的提取。在跟踪框架上，我们采用的是粒子滤波算法，本文主要研究如何采样使粒子能更好的模拟后验概率分布。本文针对于以上两个方面的改进，使得运动目标跟踪算法的鲁棒性和准确性得到加强。

基于综合颜色直方图和梯度方向直方图的运动目标跟踪算法，属于滤波预测框架下基于区域统计特性的跟踪算法。利用梯度方向直方图反映了图像的空间位置信息，同时颜色直方图反映了目标自身的感光特性，由此提出了融合颜色直方图和梯度方向直方图的综合直方图，然后采用综合直方图表示运动目标，利用经典粒子滤波算法模拟状态的后验概率分布。实验表明，使用多线索融合的目标表观能够在光照变化、目标形变、干扰和遮挡情况下进行有效跟踪，跟踪效果的精度和稳定性明显好于基于传统直方图的跟踪算法，且能够满足实时处理的需要。

基于粒子滤波的运动目标跟踪框架，是一个很有效的运动目标跟踪框架，它在复杂背景下的跟踪有很好的表现。前人也对粒子滤波做了大量的研究和实验，以及一些改进。本文主要集中研究了粒子滤波的采样问题，目的是使粒子能够更好的模拟后验概率分布。于是本文将 Condensation 算法的鲁棒性和 APF 的准确性结合起来。根据一个简单的原则，采用在线的交换采样的策略，来力求更好的模拟目标状态的后验概率的分布，以适应变化情况下的跟踪。本文也在公共测试集上做了大量实验，包括了光照变化、目标形变、前景背景混淆、视频抖动、干扰、遮挡等多种情况下的目标跟踪，实验表明在复杂背景情况下，采用交换采样的粒子滤波能够有效的进行跟踪。

本算法应用到 PTZ 旋转摄像头的监控系统中，这个系统要求只要目标出现在摄像头可以监控到的范围内，无论目标运动到哪，摄像头都要实时锁定目标，目标处于摄像头取得的视频中心，此问题是一个典型的目标跟踪问题。

运动目标跟踪作为一个独立的课题发展至今，即便融合更为先进的滤波

预测算法、运动目标表达方法以及匹配搜索算法等，依然不能很好的解决运动目标跟踪中的难点。结合本文的研究现状，依然有以下几个问题有待解决：

1、本文的研究是针对于固定尺寸的，不能进行尺度缩放。这样使得在目标尺寸变化较大的情况下，不能很好地跟踪。如何来拟合轮廓跟踪，是一个研究点

2、本文的算法在跟踪开始时，取得的目标特征的模板在跟踪的整个过程中是固定不变的。由于粒子滤波较强的鲁棒性以及颜色直方图对目标的形变不敏感，所以在发生较小的形变时，依然可以跟踪上。但是当目标发生较大的形变时，跟踪往往失效。因此，如何更新模板是本文下一步的研究方向。

3、本文提出的综合直方图只加入了颜色直方图和梯度方向直方图两部分的信息，如何把更多的特征加入到综合直方图的特征集里面，以达到更鲁棒的物体跟踪效果是本文下一步的研究方向。

4、本文的算法可以处理短时的遮挡，如人对人的遮挡。对长时间的遮挡则不能处理，长时间的遮挡完全改变了粒子的分布使粒子无法正确的模拟目标状态的后验概率分布，因此不能正确跟踪。遮挡也是跟踪中的难点问题。如何在粒子滤波下处理长时间的跟踪是本文需要研究的一个方面。

5、粒子滤波是目标跟踪算法里很好的一个框架。但是，粒子滤波本身存在很多需要解决的问题，如：合适的提议分步、粒子的重采样、粒子的衰退等。而在实际的目标跟踪中，如何采样粒子来更好的模拟后验概率分布始终是一个关键性问题。本文认为在这一方面要加强研究。

参考文献

[1] Rosenfeld A. From image analysis to computer vision: An annotated bibliography. Computer Vision and Image Understanding,2001.84(2): 298-324P

[2] Ballard D H,Brown C M .Computer Vision.Prentice-Hall,Inc,1982

[3] Lucas B D,Kanade T. An iterative image-registration technique with an application to stereo vision. In: Proceedings of the International Joint Conference on Artificial Intelligence. Vancouver, BC, Canda: William Kaufmann,1981.674-679P

[4] Horn B K, Schunck B G. Determine Optic Flow. Artifical Intelligence, 1981. 17(1-3):185-204P

[5] Beauchemin S S, Barron J L .The computation of optical flow. ACM Computing Surveys,1995.27(3): 433-467P

[6] Baker S, Matthews I. Lucas-Kanade 20 years on: A unifying framework. International Journal of Computer Vision, 2004.56(3):221-225P

[7] Meyer F, Bouthemy P. Region-based tracking using affine motion models in long image sequences. CVGIP: Image understanding,1994.60(2):119-140P

[8] Bascle B, Deriche R. Region tracking through image sequences. In: Proceedings of IEEE International Conference of Computer Vision.France 1995.302-307P

[9] Tissainayagam P, Suter D. Object tracking in image sequences using point feature. Pattern Recognition,2005.38(1):105-113P

[10] Nickels K, Hutchinson S. Estimating uncertainty in SSD-based feature tracking. Image and Vision Computing, 2002.20(1):47-58P

[11] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. International Journal of Computer Vision, 1988.1(4): 321-333P

[12] Jurie F. Tracking objects with a recognition algorithm. Pattern Recognition

Letters,1998.19(3-4):331-340P

[13] Drummond T, Cipolla R. Real-time visual tracking of complex structures. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. 24(7):932-946P

[14] Marr D. Vision: a computational investigation into the human representation and processing of visual information, Freeman W .H. and Company, San Francisco,1982

[15] Isard M ,B1ake A. CONDENSATION—Conditional density propagation for visual tracking. International Journal of Computer Vision,1998.29(1):5-28P

[16] Spengler M ,Schiele B. Towards robust multi-cue integration for visual tracking. Machine Vision and Applications,2003.14(1):50-58P

[17] Wu Y, Huang T. Robust visual tracking by integrating multiple cues based on co-inference learning. International journal of Computer Vision, 2004, 58(1):55-71P

[18] 查宇飞，毕骛彦. 一种基于粒子滤波的自适应运动目标跟踪方法. 电子与信息学报,2007.29(1):92-95 页

[19] 江宝安，卢焕章. 粒子滤波器及其在目标跟踪中的应用. 雷达科学于技术,2003.1(3):170-174 页

[20] 王永忠，梁彦. 基于多特征自适应融合的核跟踪方法.自动化学报，2008,34(4):393-399 页

[21] 朱琳，周杰，宋靖雁. 一种新的基于在线采样的多物体跟踪方法.计算机学报,2008,31(1)：151-156 页

[22] Comaniciu D, V. Ramesh, P. Meer. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. 25(5): 564-577P

[23] Stern H, B. Efros. Adaptive color space switching for face tracking in multi-colored lighting environments. Int. Conf. On Automatic Face and Gesture Recognition, 2002:236-241P

[24] 岑峰，戚飞虎. 短程线主动轮廓跟踪算法的研究—在复杂背景和非刚性运动目标跟踪中的运用. 计算机研究与发展,2003.40(2):285-288 页

[25] Collins R, Y. Liu. On-line selection of discriminative tracking features. Proceedings of International Conference on Computer Vision, 2003: 346-352P

[26] Matthews I,T. Ishikawa, S. Baker, The Template Update Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004. 26(6): 810-815P

[27] Paragios N, R. Deriche. Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000.22(3):266-280P

[28] Cootes T, Active Shape Models - Their Training and Application. Int.J. of Computer Vision and Image Understanding, 1995. 61(1):38-59P

[29] Cootes T, G. Edwards, C. Taylor. Active appearance models. European Conference on Computer Vision, 1998. 2:484-498P

[30] Black, M.J, A.D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using view-based representation. European Conference on Computer Vision, 1996:329-342P

[31] Torre, D.l, G.F, M.S.S.J. View-based adaptive affine tracking. European Conference on Computer Vision. LNCS, 1406, Springer Verlag, 1998: 828-842P

[32] Jepson,A.D, D.J.Fleet, T.F. El-Maraghi. Robust online appearance models for visual tracking. IEEE, Proc.on Computer Vision and Pattern Recognition,2001.1:415-422P

[33] 胡明昊，任明武，杨静宇. 一种基于直方图模式的运动目标实时跟踪算法.计算机应用与研究, 2004, 40(3)：71-74 页

[34] D. Lowe. Object recognition from local scale-invariant features. In Int. Conf. on Computer Vision. 1999: 1150-1157P

[35] D. Lowe. Distinctive image features from scale-invariant keypoints. In Journal of Computer Vision. 2004.60(2):91-110P

[36] Arulampalam S, et al. A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Processing,

2002.50(2):174-188P

[37] Bergman N. Recursive Bayesian Estimation: Navigation and tracking applications. PhD Thesis, Linkoping University, Sweden, 1999

[38] D. S. Lee, N. K. China, A particle algorithm for sequential Bayesian parameter estimation and model selection, IEEE Trans. Signal Processing, 2002.50(2):326-336P

[39] Doucet A. On sequential Monte Carlo Methods for Bayesian Filtering. Technical Report, University of Cambridge, UK, Department of Engineering, 1998

[40] Lin, J.S, R. Chen. Sequential Monte Carlo Methods for Dynamical Systems. Journal of the American Statistical Association, 1998.93: 1032-1044P

[41] Merwe, R.v.d,et al.The unscented particle filter. Advances in Neural Information Processing Systems, 2000,12

[42] Gilks W.R, C. Berzuini. Following a Moving Target-Monte Carlo inference for Dynamic Bayesian Models. Journal of the Royal Statistical Society, 2001.63(B):127-146P

[43] Musso C.,N. Oudiane, F. LeGland. Improving Regularised Particle Filters. In, eds A. Doucet, J. Freitas and N. Gordon, Sequential Monte Carlo Methods in Practice,Springer-Verlag, 2001

[44] Clapp T, S. Godsill. Improving strategies for Monte Carlo particle filters. Sequential Monte Carlo Methods in Practice, eds A. Doucet, J.Freitas and N. Gordon, Springer-Verlag, 2001

[45] Oudjane. N,C. Musso. Progressive Correction for Regularized Particle. Proc. 3rd Int. Conf. on Information Fusion,France,2000

[46] MacCormick, J, A. Blake. A probabilistic exclusion principle for tracking multiple objects. Proceedings of International Conference on Computer Vision,1999.(1):572-578P

[47] M.k.Pitt, N.Shephard. Filtering via simulation: Auxiliary particle filter, Journal of American Statistical Association ,1999.94(446):590-599P

[48] A. Treptow, A. Zell. Real-time object tracking for soccer robots without

color information, Robotics and Autonomous Systems 2004.41-48P

[49] F.Wallhoff, M. Zobl, G. Rigoll, I. Potucek. Face tracking in meeting room scenerios using omni directional views. In: International Conference on Pattern Recognition, 2004.933–936P

[50] 赵建伟，刘重庆. 适用于遮挡问题的目标跟踪算法. 计算机辅助设计与图形学学报, 2003,15(9),1091-1095 页

攻读硕士学位期间发表的论文和取得的研究成果

致 谢

本课题的研究工作是在导师黄凤岗教授和焦建彬教授的悉心指导下完成的。黄凤岗教授和焦建彬教授在我的学习和实验中，从理论和实验方面给予了大量的、极其有益的建议和具体的指导，并在论文的撰写和审稿中倾注了大量的心血，他们诲人不倦的精神和对我的谆谆教导，使我从恩师身上学到了不少可贵的东西，让我受益匪浅。此外，还在生活和思想上关心我、照顾我，再次表示万分的感谢！尤其是两位恩师严谨的治学态度、积极的进取精神、平易近人的工作作风将使我终身受益。

其次，特别要感谢赵靖老师给我提供了一个很好的实验环境和平台以及叶齐祥老师在算法研究上的悉心指导。感谢中科院研究生院工程教育学院实验室的彭艺、何新英、杨志辉、范焱同学，他们在我的论文工作和撰写中，给予了我极大的支持和帮助。另外，感谢跟踪组的韩振军和刘志花同学，他们两人和我经常在算法上进行探讨，共同研究，对我的研究提供了很大的帮助。

我要衷心感谢我在 GABRobins 里的恩师，注册公估师陈继清先生，他对我的研究没有直接的帮助。但是在我作为陈先生的助手的时间里，陈先生作为我的老师，带领我在社会这样一所大学里学到了很多知识，对我的成长有很大的帮助。

我还要衷心感谢我敬爱的父母、亲人，他们对我的工作和学业的热情鼓励和殷切期望，都给了我巨大的动力。同时衷心感谢所有关心和帮助过我的老师和同学。

最后感谢中国科学院研究生院工程教育学院所有给予我关心、指导的老师、同学！