

学校代码: 10475

学 号: 104753040912

河南大学研究生硕士学位论文

视频运动目标跟踪算法研究

The Research on Tracking Algorithm of Video Moving

Objects

专 业 名 称: 计算机应用

专 业 代 码: 081203

研 究 方 向: 目标跟踪

年 级: 二〇〇四级

研 究 生 姓 名: 常向魁

导师姓名、职称: 刘先省 教授、焦建彬 教授

完 成 日 期: 二零零七年五月

论 文 主 题 词: 目标跟踪/Mean-Shift/综合直方图

关于学位论文独立完成和内容创新的声明

本人向河南大学提出硕士学位申请。本人郑重声明：所呈交的学位论文是本人在导师的指导下独立完成的，对所研究的课题有新的见解。据我所知，除文中特别加以说明、标注和致谢的地方外，论文中不包括其他人已经发表或撰写过的研究成果，也不包括其他人为获得任何教育、科研机构的学位或证书而使用过的材料。与我一同工作的同事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。

学位申请人（学位论文作者）签名：_____

20 年 月 日

关于学位论文著作权使用授权书

本人经河南大学审核批准授予硕士学位。作为学位论文的作者，本人完全了解并同意河南大学有关保留、使用学位论文的要求，即河南大学有权向国家图书馆、科研信息机构、数据收集机构和本校图书馆等提供学位论文（纸质文本和电子文本）以供公众检索、查阅。本人授权河南大学出于宣扬、展览学校学术发展和进行学术交流等目的，可以采取影印、缩印、扫描和拷贝等复制手段保存、汇编学位论文（纸质文本和电子文本）。

（涉及保密内容的学位论文在解密后适用本授权书）

学位获得者（学位论文作者）签名：_____

20 年 月 日

学位论文指导教师签名：_____

摘 要

动态场景视频中的运动目标检测与跟踪算法是计算机视觉领域中基础的、富有挑战性的课题之一，在现代工业、国防和空间技术等领域有着广阔的应用前景。展开研究具有重要的实践意义和理论价值。

但是迄今为止这个课题仍然存在许多问题没有解决，特别是当背景较复杂时间题将变得更加困难。其中主要困难来自于目标的尺度伸缩、遮挡、背景复杂性和非刚体形变等。本论文研究的目的是将目标表示、分层思想、多分辨率分析、预测判断以及推理等人类视觉常用的方法引入到视觉跟踪研究中，寻求有效的鲁棒实时性动态跟踪算法。

本文首先提出了一种基于SIFT(Scale invariant feature transformation)特征匹配和Kalman滤波的运动目标跟踪算法。该算法首先提取SIFT特征点表示运动目标，然后基于欧式距离和近邻搜索进行特征匹配，并且对匹配结果进行聚类分析以去除噪声干扰和误匹配。最后，在Kalman滤波框架内，结合特征点匹配结果和滤波预测进行目标跟踪。大量实验结果表明，此算法中用于目标量测的SFIT特征具有尺度伸缩不变性及旋转不变性，对部分形变和亮度变化不敏感。特征点匹配聚类和Kalman滤波框架的结合能够有效解决跟踪中的运动目标尺度伸缩和部分遮挡问题。

上述算法虽然具有良好的跟踪效果和精度，但是因为SIFT特征检测和匹配计算量很大，影响了算法的速度。在于PC机系统中进行视频动态目标跟踪难于达到实时跟踪效果。因此在随后的研究中对特征表达和匹配搜索算法进行了改进。

改进算法基于传统的 Kalman 滤波和 Mean-Shift 优化框架，和现有算法的不同之处在于算法融合了色度直方图和梯度方向直方图，形成了一种新的综合直方图特征，以解决运动目标的旋转和形变；然后构建了运动目标图像区域的金字塔，以实现多尺度的目标跟踪处理；最后算法采用 Kalman 滤波预测耦合 Mean-Shift 算法，在尺度、位移空间内进行优化匹配搜索，确定最佳候选目标区域的位置信息。大量实验结果表明，本文提出的综合直方图特征能够很好的适应伴随尺度伸缩、旋转和形变的运动目标跟踪，可以解决传统直方图跟踪算法的稳定性差和精度不高的问题，滤波和匹配优化框架下的跟踪算法具有比传统算法更高的精度和速度。同时在于 PC 机系统中进行视频动态目标可以达到实时效果。

本文的研究为视频分析、智能安全监控应用、人体行为分析与合成等后续工作，提供了保证。

关键词：运动目标跟踪；尺度伸缩和遮挡；Kalman 滤波；Mean-Shift 算法

Abstract

Moving objects detection and tracking in dynamic scenes are the fundamental and challenge research topics in computer vision area at present. They have great research and application values in modern industry, military, aviation and spaceflight area.

While to now robust object tracking is still an open problem, especially when the background is complex. The main obstacles in the research are the scale of objects, occlusion, background complexity and non-rigid deformation. In this research our goal is find a real-time and robust object tracking algorithm by employing the novel object feature representation, layer idea, multi-scale analysis, filter and prediction.

An effective moving object tracking algorithm is proposed by combining Kalman filtering method and template feature points matching. Firstly, SIFT (FULL NAME) feature points are extracted to represent the moving object, which are then matched with these feature points on the object template to obtain motion vectors of the objects. Secondly the motion vectors of the object will be processed by a clustering algorithm to reduce noise and mis-matching. Finally, the tracking algorithm is developed by Kalman filtering method on feature matching result. Experiments show that the SIFT features are invariant both to scale and rotation. It is insensitive to object deformation and illumination variation. The feature point clustering algorithm together with Kalman filtering is robust to scale variances and partially object occlusion.

Despite of the precision and robustness of the above mentioned algorithm, its efficiency is very low since the feature extraction of SIFT has a very large computational cost. In a PC platform, it is quite difficult to obtain a real-time performance. That is the very reason why we proposed the following modification.

To obtain both precise tracking result and tracking efficiency, an improved moving object tracking algorithm is presented using the multi-scale Histograms of Oriented Grads & Colors (HoGC) based on the framework of traditional Kalman Filter and Mean-Shift optimization. Firstly, based on the fusion of oriented grads and color histogram a new HoGC is proposed, which is invariant to moving object's rotation and Deformation. Secondly, HoGC pyramid is constructed to characterize the multi-scale objects more robustly. Then, by coupling kernel-based Mean-Shift algorithm with

Kalman Filter, HoGC matching is optimized in the scale and position spaces and the position of the candidate object is identified. Finally, experiments on representative video sequences are accomplished. Experiments validate that the proposed multi-scale HoGC is robust to represent the object and is invariant to scale and deformation, and the proposed tracking algorithm not only can improve the reliability and accuracy which traditional histogram-based tracking algorithm cannot, but also can improve the real-time performance which scale-invariant tracking algorithm fails. The improved tracking algorithm is better than the traditional algorithm in accuracy and speed.

Our work can be applied in video analysis, intelligent surveillance, activity analysis and synthetic(??).

Key words: Moving object tracking, scale variances & occlusion, Kalman filter, Mean-Shift

目 录

摘 要.....	I
ABSTRACT.....	III
第 1 章 绪论.....	1
1.1 课题来源、背景和意义.....	1
1.1.1 课题来源.....	1
1.2.1 课题背景及意义.....	1
1.2 国内外的研究发展动态.....	2
1.3 视频跟踪算法的主要研究内容和分类.....	4
1.4 本文的主要贡献.....	6
1.5 本文的组织结构.....	7
第 2 章 运动目标跟踪算法基础.....	9
2.1 引言.....	9
2.2 运动目标跟踪算法的分类.....	10
2.2.1 基于特征的跟踪算法.....	10
2.2.2 基于区域统计特性的跟踪算法.....	11
2.2.3 基于模型的跟踪算法.....	12
2.3 滤波预测算法综述.....	12
2.3.1 最优线性均方误差估计算法.....	12
2.3.2 扩展 Kalman 滤波算法.....	15
2.3.3 粒子滤波算法.....	17
2.3.4 Kalman 滤波的应用.....	18
2.3.5 算法分析.....	18
2.4 相似性度量方法.....	19
2.4.1 欧式距离 (Euclidean Distance).....	19
2.4.2 加权距离 (Weighted Distance).....	19
2.4.3 巴特查理亚系数 (Bhattacharyya).....	20
2.4.4 Hausdorff 距离.....	20
2.4.5 改进的 Hausdorff 距离.....	20
2.5 匹配搜索的优化算法.....	21

2.5.1 Mean-Shift 理论基础.....	21
2.5.2 算法分析.....	24
2.6 本章小结.....	24
第 3 章 基于 SIFT 特征匹配的运动目标跟踪算法.....	25
3.1 概述.....	25
3.2 SIFT 特征点提取和相似性度量.....	26
3.2.1 SIFT 特征点介绍.....	26
3.2.2 SIFT 特征点提取和表达.....	27
3.2.3 SIFT 特征点相似性度量.....	29
3.3 最大模糊熵的高斯聚类算法.....	30
3.4 基于 SIFT 特征匹配和 KALMAN 滤波的运动目标跟踪算法研究.....	31
3.4.1 概述.....	31
3.4.2 改进的 Kalman 滤波算法.....	32
3.5 实验结果和算法分析.....	33
3.5.1 尺度伸缩和形变的非刚体运动目标的跟踪实验.....	33
3.5.2 部分及完全遮挡情况下的非刚体运动目标的跟踪实验.....	36
3.5.3 光线亮度变化及部分遮挡情况下的刚体运动目标的跟踪实验.....	37
3.5.4 跟踪算法误差分析.....	39
3.6 本章小结.....	40
第 4 章 基于综合直方图的运动目标跟踪算法.....	41
4.1 引言.....	41
4.2 综合颜色和梯度方向直方图.....	42
4.2.1 梯度方向直方图.....	42
4.2.2 色度直方图.....	44
4.3 MEAN-SHIFT 算法.....	45
4.4 基于多尺度综合直方图的运动目标跟踪算法.....	46
4.5 跟踪算法比较和性能分析.....	47
4.5.1 运动目标尺度伸缩变化情况下的跟踪实验.....	47
4.5.2 运动目标遮挡情况下的跟踪实验.....	49
4.5.3 运动目标形变情况下的跟踪实验.....	50
4.5.4 跟踪算法的定量比较与误差分析.....	52
4.6 本章小结.....	53

总结与展望.....	55
致 谢.....	57
参考文献.....	59
攻读硕士学位期间完成的论文及参加的科研项目	64

第1章 绪论

1.1 课题来源、背景和意义

1.1.1 课题来源

本课题来源于国家自然科学基金（60272024）、河南省高校杰出科研人才创新工程项目（2003KYCX003）、河南省高校创新人才培养工程（2004-2009）及河南省自然科学基金（0411010400）资助课题。

1.2.1 课题背景及意义

计算机视觉已经成为人工智能领域最热门的课题之一，与专家系统、自然语言理解一起成为人工智能最活跃的三大领域。计算机视觉起源于人工智能的一个梦想，让计算机能够像我们人类一样进行视、听、说和思考，辅助人类对环境进行感知、解释和理解。随着计算机技术的进一步发展，有理由相信人工智能的梦想离现实越来越近。

视频运动目标跟踪技术（以下简称为：视频跟踪技术）作为计算机视觉领域的核心研究课题之一，主要目的是模仿生理视觉系统的运动感知功能，通过对摄像头获得的图像序列进行分析，计算出运动目标在每一帧图像上的二维坐标位置，并根据不同的特征值，将图像序列中连续帧间的同一运动目标关联起来，得到每帧图像中目标的运动参数以及相邻帧图像间运动目标的对应关系，得到各个运动目标完整的运动轨迹，也即在连续的视频序列中建立运动目标的对应关系，简单来说，就是在下一帧图像中找到目标的确切位置，并反馈给跟踪系统进行跟踪，为视频序列分析和理解提供重要的依据和基础。

视频跟踪技术，融合了图像处理、模式识别、人工智能、自动控制以及计算机等众多领域中的先进技术和核心思想。视频跟踪系统，与传统雷达系统相比，采用摄像机和光学系统等比较廉价的设备，并且是被动式工作模式，工作时不向外辐射无线电波，不易被敌方的电子侦察设备发现，因此具有更高的性价比、一定

的隐蔽性和抗电子干扰能力。“眼见为实”的视频跟踪系统，从视频监视器上能直接看到运动目标图像，可以获得关于运动目标更丰富的信息，因此具有很好的直观性和可靠性。

视频跟踪技术，经过近 50 多年的深入研究和发 展，在军事制导、视觉导航、安全监控、智能交通、医疗诊断以及气象分析等方面有着广泛的应用和发展前景。在现代化军事国防和民用方面都有着重要的研究价值和应用前景，如军事方面，主要采用电视跟踪和红外跟踪技术，具体应用如导弹制导、火炮控制、空间飞行体和靶场测量等。民用方面，如视频跟踪通过摄像机监控动态场景，已被广泛地应用在社会生活的各个方面，可应用于社区、大型公共场所及重要设施的安全监控；智能交通系统中，进行车辆的实时检测和跟踪，可以实时监控车流量、车速、车流密度、交通事故、违章逃逸车辆等交通状况，用于实时的智能交通调度。另外，视频跟踪技术在视频会议、视频分析、视频检索、基于视频的运动分析和合成、基于运动信息的身份识别、图像检索、水文观测、港口管理、医学图像分析、以及远距离测量、零部件质量检测等许多领域也是大有作为。一般意义上的视频跟踪技术还远未成熟，要开发出真正鲁棒、实用的视觉跟踪应用系统还需要更为鲁棒的核心算法设计和研究。

1.2 国内外的研究发展动态

国外对于视频跟踪理论研究起步较早。美国自然科学基金委员会及美国军方对复杂环境下多目标的动态检测、识别及跟踪的算法的研究及其应用非常重视。美国自然科学基金多次资助有关公司进行相关算法研究，美国国防部 (Department of Defense, DoD)、美国国防部高级研究项目署 (Defense Advanced Research Project Agency, DARPA)、美国陆军夜视及电子探测设备委员会 (Night Vision and Electronic Sensors Directorate, NVESD)、美国陆军 SBIR (Small Business Innovation Research) 项目，美国海军水下战事中心 (Navy Underwater Warfare Center, NUWC) 等部门也先后投巨资资助该类算法的开发及其相关应用研究并取得可观的成果。在 20 世纪 50 年代初期，GAC 公司就为美国海军研制开发了自动地形识别跟踪系统 (Automatic Terrain Recognition and Navigation (System), ATRAN)。80 年代初期，自适应跟踪和智能跟踪的思想相继提出，美国国防部 DARPA 和 JSG&CC 联合发起和成立了自动目标识别工作组 ATRWG。许多国际学术会议和期刊也包含了这一

领域的最新研究成果，如：Int. Conf. on Computer Vision and Pattern Recognition (CVPR)、Int. Conf. on Computer Vision (ICCV)、European Conf. on Computer Vision (ECCV)、Asian Conf. on Computer Vision (ACCV)、Int. Conf. on Pattern Recognition (ICPR)、Int. Journal of Computer Vision (IJCV)、IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)、IEEE Trans. on Image Processing、Pattern Recognition、Image and Vision Computing 等。

国外的许多高校及科研机构对动态目标检测问题进行了深入的研究。美国卡耐基·梅隆大学的 Alan J. Lipton 等提出了改进的帧相减算法[1]，美国马里兰大学 Larry S. Davis 领导的研究小组对基于背景模型的动态目标检测算法进行了长期研究[2]，麻省理工学院的 Chris Stauffer 等人对自适应背景模型进行了深入研究[3]，德国 Daimler-Benz 研究中心的 D. M. Gavrilu 等对实时算法进行了研究[4]，韩国 Sogang 大学对基于熵计算的动态目标检测算法进行了研究[5]。

随着图像处理和分析技术、计算机技术和各种图像传感器技术的飞速发展，国内一些高校和科研所已经逐渐开展了这方面的研究，目前已经取得了喜人的成绩。如：中科院自动化所模式识别国家重点实验室图像和视频分析研究组开发了人体运动的视觉分析系统、交通行为事件分析系统、交通场景监控系统和智能轮椅视觉导航系统；中科院计算所先进人机通信技术联合实验室对于体育视频中的目标分割和跟踪、精彩片断分析的研究；清华大学的野外自然环境中的视觉侦查系统，该系统具有运动检测、跟踪和全景图生成等功能；西安交通大学的人工智能与机器人研究所使用光流和帧差两种算法对自适应巡航控制中的车辆跟踪的研究。微软亚洲研究院对基于熵计算的动态目标检测算法进行了研究[6]。

当运动目标通过摄像机的检测区域时，运动目标的图像可能发生明显的变化，这些变化主要来自于四个方面：运动目标的尺度伸缩变化、运动目标被其他物体或者背景的部分及完全遮挡、运动目标姿态和形状的变化、光线亮度的变化等。由于这些情况的存在使的对运动目标的检测与跟踪变得更加困难。现有的运动目标跟踪算法，虽然能够完成对运动目标的可靠跟踪，但是大多是基于特征或者相关跟踪，要么是跟踪的精度不高，要么就是跟踪精度提高了，特征检测和匹配的计算复杂度也变大了，难以满足实时处理的要求。因此鲁棒的、实时的视频跟踪系统还需要解决下列几个难题：

- 1). 运动目标图像的三维重建；三维空间的运动目标到二维图像本身丢失了重

要的深度信息；

- 2). 图像的外界影响；现实中图像成像模型受到物体自身的材质、外界光源、摄像机的视角和距离、透明度和遮挡等诸多因素的影响，并且像素值与此多因素之间的依赖关系是非线性的，像素值的变化往往掩盖了图像变化的真正原因；
- 3). 运动目标的部分及完全遮挡；背景对目标的遮挡以及目标之间遮挡是视频跟踪系统中一个难题，遮挡是造成目标的图像表达突然变化的重要原因之一，并且这种变化具有突然性和不连续性，从而容易引起跟踪算法的失效；
- 4). 运动目标的尺度伸缩变化和形变；在跟踪的过程中目标的运动，引起与摄像机的距离和视角发生变化，会造成目标图像的尺度伸缩变化和形变，这些变化会造成目标图像表达的非线性变化，从而导致跟踪的失效；
- 5). 场景光线亮度变化；同一目标在不同的光线亮度条件下具有明显差异的目标图像，光线亮度的影响是很多室外实际应用系统中阻碍性能提升的瓶颈因素；
- 6). 目标模板的更新；由于运动目标在跟踪的过程中伴随着尺度变化、形变以及场景光线亮度的变化，需要定义合理的模板更新策略，避免运动目标的漂移；
- 7). 目标复杂的运动模式；跟踪算法主要采用滤波预测算法，在运动目标可能出现的区域内进行局部搜索，当目标具有复杂的运动模式时，如目标的运动速度或者方向突然发生改变，运动目标不在的预测的搜索区域中，造成简单滤波预测算法容易失效，进而造成目标丢失；
- 8). 实际应用中复杂的背景；复杂背景是影响跟踪算法精度和鲁棒性的最重要因素之一。实际应用中的背景是混乱复杂的，并且是动态变化的；另外，出于跟踪系统实时处理的考虑，要求目标模板表达所采用的图像特征及其模型必须简单，简单的目标模板可能造成所选择的图像特征与背景中其他干扰物体相似而造成跟踪算法的失效。

1.3 视频跟踪算法的主要研究内容和分类

视频运动目标跟踪算法仍然是计算机视觉领域的研究热点问题之一，实际应用当中，复杂的动态背景、运动目标的尺度伸缩、目标遮挡、光线亮度信息的变化、背景中物体的变动以及运动目标的阴影等问题，解决上述场景中的难题是该领域所面临的一个巨大挑战，如何提高运动目标跟踪算法的实时性、鲁棒性和精确性，是视频跟踪算法研究的热点和难点。

视频跟踪算法根据运动目标和背景的相对运动，可以分为：动态背景的运动目标跟踪算法，动态背景的静止目标跟踪算法，静止背景的运动目标跟踪算法，静止背景的静止目标跟踪算法，以前三类运动目标跟踪算法为研究重点，因此本文提到的跟踪算法没有特别说明都指的是运动目标跟踪算法。视频跟踪算法根据跟踪目标的个数，大致可以分为：单目标跟踪算法和多目标跟踪算法，本文主要研究的是单目标跟踪算法，但跟踪框架的设计很容易扩展的多目标跟踪的情况。视频跟踪算法根据跟踪目标的表达方法和相似性度量方法，大致可以分为：基于轮廓的跟踪算法、基于特征的跟踪算法、基于区域统计特性的跟踪算法和基于模型的跟踪算法。视频跟踪算法根据跟踪目标的属性，可以分为：刚体目标跟踪算法和非刚体目标的跟踪算法。

目前已有的视频跟踪算法有的是针对刚体目标的跟踪、有的是针对非刚体目标的跟踪、有的是针对提高跟踪匹配的准确性而提出的，这类算法的主要特点是通过选取稳定性和可区分性好的特征来提高目标的搜索匹配速度和匹配的正确程度；有的算法是针对缩小目标搜索范围提出的，这类算法主要特点是通过滤波方法预测目标下一时刻可能出现的位置，通过缩小目标的搜索范围来缩短目标搜索时间。

本文的主要研究旨在提出一种鲁棒实时的跟踪系统，能够解决运动目标的尺度伸缩、部分及完全遮挡、运动目标的形变以及光线亮度变化等视频跟踪中的开放

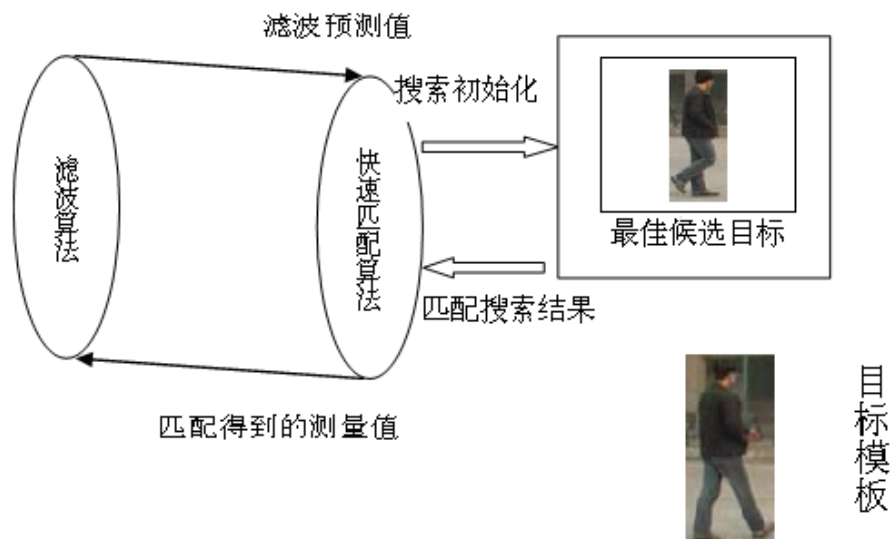


图 1-1 滤波算法和快速匹配算法耦合的原理示意图

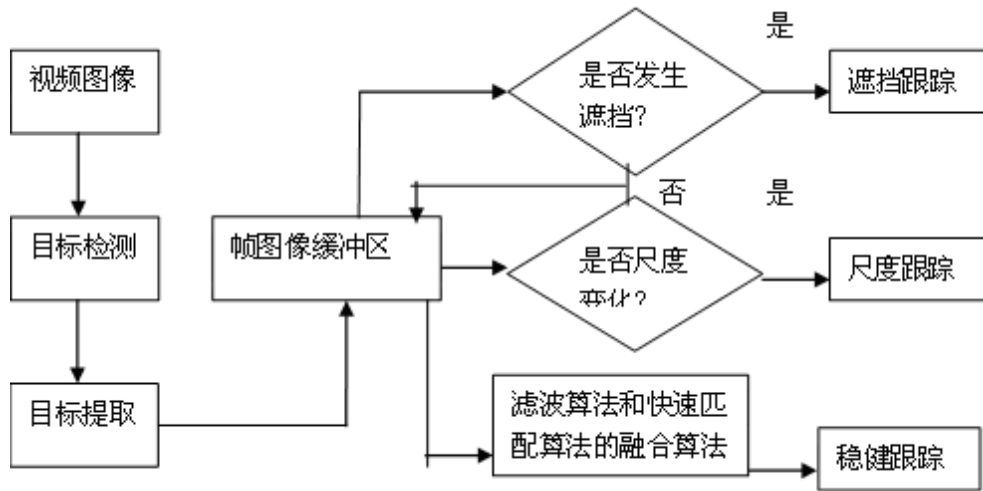


图 1-2 本文所提出视频跟踪系统的功能框图

性难题，跟踪系统功能框图如图 1-2 所示。本文的主要研究内容，在 Kalman 滤波预测的框架内（如图 1-1 所示），研究了基于特征的跟踪算法和基于区域统计特性的跟踪算法，分别提出了“基于 SIFT 特征匹配和 Kalman 滤波的运动目标跟踪算法”和“基于综合直方图的运动目标跟踪算法”，滤波预测算法为匹配搜索算法提供初始化的起点，并对环境噪声具有一定的抑制作用，匹配搜索算法为滤波预测算法的修正更新提供量测值。

1.4 本文的主要贡献

本文综合采用多尺度分析、分层思想、预测判断以及推理等人类视觉常用的方法引入到本课题的研究当中，针对运动目标跟踪的尺度变化、部分及完全遮挡、形变和旋转等复杂场景下的传统跟踪算法精度不高、实时处理不佳的情况，寻求一种鲁棒的实时处理的运动目标跟踪算法。通过对动态目标跟踪系统、基于 SIFT 特征匹配的动态目标跟踪算法、基于综合直方图的动态目标跟踪算法、复杂场景下跟踪实验进行研究和分析，并且取得了一些有意义的成果。

本文的主要贡献如下：

- 1). 基于 SIFT 特征匹配和 Kalman 滤波预测的动态目标跟踪算法，可以有效地解决动态目标的尺度伸缩变化、遮挡和形变等情况下的鲁棒跟踪。本文通过计算多尺度图像的极值点检测作为特征点，并且采用其邻域的梯度方向直方向

量表示该特征点；由于同一运动目标的特征点应该表现出运动的一致性，采用聚类算法对相邻帧图像之间匹配的特征点运动进行聚类分析，有效地滤出噪声、相似目标等干扰点；最后在 Kalman 滤波预测的框架下实现跟踪算法，提高了跟踪算法的跟踪精度，同时也降低了算法的计算量。

2). 基于综合直方图的动态目标跟踪算法。基于 SIFT 特征跟踪算法虽然很好的跟踪精度，但是由于特征点检测和向量匹配计算复杂，很难达到跟踪系统实时处理的要求；基于传统直方图的跟踪算法，虽然计算简单并且能够满足跟踪系统的实时处理，但是由于传统直方图的跟踪算法丢失了目标图像的空间结构信息，造成了跟踪算法精度不高和稳定性差。以上两种跟踪算法的启发，本文提出了一种新的综合直方图，融合了目标图像的色度直方图和色度图的梯度方向直方图，用多尺度综合直方图来表示运动目标；在 Kalman 滤波预测和 Mean-Shift 算法耦合的框架下，在尺度和位移分量上优化匹配搜索的过程，提高了基于直方图跟踪算法的精度和稳定性，近似达到了基于 SIFT 特征算法的精度，同时基本上可以满足跟踪系统实时处理的要求。

3). 建立了视频中运动目标跟踪算法的测试集合和算法比较的数据，针对视频中运动目标的尺度伸缩变化、部分及完全遮挡、形变、动态背景等各种复杂场景，分别建立测试视频集合和运动目标的手工框定的空间位置信息，便于进行算法的误差分析和性能比较。

1.5 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，总结了本文的主要贡献。

第二章运动目标跟踪算法基础，主要综述了视频跟踪算法的滤波预测和匹配量测两部分内容，论述了滤波预测算法的 Kalman 滤波、扩展 Kalman 滤波以及粒子滤波算法，论述了四种基于不同的目标表达的量测方法以及相应的相似性度量方法。

第三章基于 SIFT 特征匹配的运动目标跟踪算法，主要论述了 SIFT 特征点检测、表达方法和相似性度量；运动一致性的最大模糊熵的高斯聚类算法；该算法在各种复杂场景下的运动目标跟踪实验与性能分析，可以实现对运动目标在尺度

伸缩变化、部分及完全遮挡和部分形变等情况下的鲁棒跟踪，但时尚未能解决跟踪系统要求的实时处理。

第四章基于综合直方图的运动目标跟踪算法，提出了新的综合直方图并且用于运动目标的表达，综合了颜色直方图和梯地方向直方图；在 Kalman 滤波预测和 Mean-Shift 算法耦合的框架内，在尺度、位移分量上优化匹配搜索过程；该算法在各种复杂场景下的运动目标跟踪实验，并且与基于 SIFT 的跟踪算法和基于传统直方图的跟踪算法进行了性能分析，本文所提出的跟踪算法是在跟踪精度和实时处理要求之间的折衷算法，基本上可以实现复杂场景下的鲁棒实时跟踪。

最后，对本文工作的总结和展望。如何提高跟踪算法的鲁棒性、精度和实时性，一直是视频跟踪系统的研究热点和难点。

第2章 运动目标跟踪算法基础

运动目标跟踪算法仍然是计算机视觉领域的研究热点问题之一。当运动目标通过摄像机的检测区域时，运动目标的图像可能发生明显的变化，这些变化主要来自于以下几个方面：运动目标的尺度伸缩变化、运动目标被其他物体或者背景部分及完全遮挡、运动目标姿态和形状的变化、光线亮度的变化等。由于这些情况的存在使的对运动目标的检测与跟踪变得更加困难。实际应用当中，如何提高运动目标跟踪算法的实时性、鲁棒性和准确性，解决上述场景中的难题是运动目标跟踪算法研究的热点，同时也是该领域面临的一个巨大挑战。

2.1 引言

一个鲁棒的、精度高的实时的运动目标跟踪系统主要包括相辅相成的耦合的两部分：预测部分和量测部分（如图 1-1 所示），模板匹配是一个很好的量测方法，滤波预测为下一时刻最佳候选目标匹配提供一个搜索的初始化位置，提高了跟踪的精度也降低了不必要的计算，同时匹配搜索的结果获得了当前时刻运动目标的精确的状态参数信息，为下一步滤波预测提供了修正的量测值。

量测的基本原理是利用运动目标的模板图像，在图像序列中按一定规则进行搜索，寻找最佳匹配位置，其中匹配所采用的搜索策略是个关键问题。如果采用遍历搜索整帧图像，运算量非常大，而且也没有必要。如果能够提前预测运动目标可能出现的位置，就可以在较小的候选邻域内进行匹配搜索，运动目标匹配量测和滤波预测是计算最佳候选目标位置的很好方法。一般来说，两帧图像之间的时间间隔比较小（大约为 40ms），并且目标的运动都是连续变化的，因此可以利用已知的量测信息为目标在下一帧图像中的位置进行预测，然后以预测位置为中心的区域附近进行搜索，这样做相对于整帧图像的遍历搜索计算量小了很多，降低了其他相似目标的干扰的匹配误差，提高了跟踪的精度。

本章主要介绍视频跟踪算法涉及的滤波预测算法、运动目标匹配量测及相似性度量的方法。常见的滤波预测算法有 Kalman 滤波、扩展 Kalman 滤波和粒子滤波

方法等。依据运动目标的表达和匹配量测方法，运动目标跟踪算法可以分为四类：基于边缘的方法、基于特征的方法、基于区域统计特性的方法和基于模型的方法。跟踪算法精度很大程度上取决于对运动目标的定义表达和匹配量测。

2.2 运动目标跟踪算法的分类

视频跟踪技术通常有两种算法思想：通过识别目标来进行跟踪和通过运动目标检测来进行跟踪。前一种算法思想通过识别每一帧的图像中跟踪目标，来确定该运动目标的位置和实现跟踪；该算法思想包含了目标识别和目标匹配两部分。而后一种跟踪算法思想通过检测和发现运动目标，并确定运动目标的位置进行跟踪，该方法不需要考虑目标的形状、尺寸，可以检测任何物体。

视频跟踪算法根据跟踪目标的表达方法和相似性度量方法，可以大致分为：基于轮廓的跟踪算法、基于特征的跟踪算法、基于区域统计特性的跟踪算法和基于模型的跟踪算法。基于轮廓的跟踪算法通过对目标的分割和边缘提取，实现对运动目标的跟踪，但是当目标被部分及全部遮挡时，跟踪算法受到很大的限制。基于模型的跟踪算法，采用目标的物理知识建立模型，而不仅仅是目标的图像信息，因此基于 3D 模型跟踪算法是完美的，但是在无先验知识的实际跟踪系统中很难得到目标的模型。基于特征表达的跟踪算法，通过提取具有尺度伸缩、旋转等不变性的特征点（如 SIFT、KLT、Harris Corners、SUSAN Corners 等），解决尺度伸缩和部分遮挡等问题，SIFT 是 D. Lowe 提出的一种具有尺度和旋转不变性特征点提取算法 [23]，是基于图像金字塔的多尺度极值点检测和梯度方向表示的方法，在图像配准和跟踪方面具有很高的匹配精度和鲁棒性，但是特征自身的检测和提取往往计算复杂难以满足跟踪系统实时处理的要求。基于区域统计特性的跟踪算法，利用的是目标图像的全局描述信息，很难精确地定位目标的位置，同时当背景或者其他目标的区域统计特性相似的时候，跟踪算法也会失效（如基于颜色或者灰度直方图的跟踪算法）。

2.2.1 基于特征的跟踪算法

基于特征匹配的跟踪方法不考虑运动目标的整体特征，即不关心目标是什么，只通过目标物体的一些特征来进行跟踪。由于图像采样时间间隔通常很小，可以认为这些特征在运动形式上是平滑的，因此可以完成目标的整体跟踪过程。

无论是刚体运动目标还是非刚体运动目标,基于特征匹配的方法进行目标跟踪时包括特征的提取和特征匹配两个过程。在特征提取中要选择适当的跟踪特征,并且在下一帧图像中提取特征;在特征匹配中将提取的当前帧图像中目标的特征与特征模板相比较,根据比较的结果来确定目标,从而实现目标的跟踪。例如使用特征点对人体进行运动跟踪[[33]。该方法把需要跟踪的每一个人用一个矩形框封闭起来,封闭框的质心被选择为目标的跟踪特征。除了用单一的特征来实现跟踪外,还可以采用多个特征信息综合在一起作为跟踪特征,来提高跟踪的效果。利用多个目标特征联合进行跟踪的算法,如 Stan Birchfield 提出了将目标的颜色直方图模型和灰度梯度模型联合的方法,实现对人头部的实时跟踪[’ 4]; H.Chen 和 T. Liu 联合目标的颜色概率分布和边缘密度信息完成对非刚体目标的跟踪[is}

如果运动目标物体简单,可以将整个目标作为特征来进行跟踪,这种方法也被称为模板匹配。

2.2.2 基于区域统计特性的跟踪算法

基于区域匹配的跟踪方法是把图像中运动目标的连通区域的共有特征信息作为跟踪检测值的一种方法。在连续的图像中有多种区域信息,例如颜色特征、纹理特征等等。这种方法不需要在视频序列的图像中找到完全相同的特征信息,通过计算区域的原始目标之间的相关性来确定跟踪目标的位置。基于区域的跟踪方法我们可以选取整个区域的单一的特征来实现跟踪,但是实际跟踪过程中单一的特征不太好选择,所以一般采用运动目标的多个特征来进行跟踪。例如, Wren} 34】利用小区域特征进行室内单人的跟踪。该方法将人体看作由头、躯干、四肢等身体部分所对应的小区域块的联合体,通过分别跟踪各个小区域块最终完成对整个人体的跟踪。

基于区域的跟踪与基于区域的分割过程联系较为紧密,所以也可以利用分割同跟踪相结合的方法来进行跟踪。例如 Francois} 3s} 提出的区域跟踪方法依赖于以前的检测来区分运动目标,然后跟踪目标。运用滤波跟踪技术精确估计了区域几何形状和速度。Jorge} 36} 等人提出的区域跟踪算法不仅利用了分割结果来给跟踪提供信息,同时也能利用跟踪所提供的信息改善分割效果,把连续帧的目标匹配起来达到跟踪目标的目的。

2.2.3 基于模型的跟踪算法

基于模型匹配的跟踪方法是用模型来表示目标物体，也就是建模，然后在图像序列中跟踪这个建好的模型来实现目标的跟踪。对于刚体目标来说，其运动状态变换主要是平移、旋转等，可以利用该方法实现目标跟踪。但是实际应用中跟踪的物体往往是非刚体的，目标确切的几何模型不容易得到，因此可以利用变形轮廓模板来匹配到目标来进行跟踪。目前有两种可变形模型：一种是自由式的可变形模型，主要满足一些简单的约束条件(连续性、平滑性等)用来跟踪任意形状的运动目标。另一种是参数可变的模型。这种方法用一个参数公式，或一个变形公式来描述目标物体的形状。

在现在目标跟踪方法中，有一种算法建立在鲁棒统计与概率分析的基础之上的 Mean Shift 算法。这种算法通过非参数估计，沿着图像梯度方法查找运动目标的概率分布，从而在视频图像中跟踪目标。该算法要选择上一次确定的目标位置结果作为当前帧图像中目标的初始位置。本论文就采用 Mean Shift 算法来实现对运动目标的跟踪。

2.3 滤波预测算法综述

运动目标跟踪常用的预测算法有 Kalman 滤波、扩展 Kalman 滤波以及粒子滤波等滤波算法，在这里我们主要研究了经典的线性系统均方误差最小估计的 Kalman 滤波和非线性的扩展 Kalman 滤波算法。

2.3.1 最优线性均方误差估计算法

卡尔曼滤波方法是美国著名学者 Kalman 和 Bucy 于 1960 年提出的一种递推最优线性均方误差估计的方法，他把状态空间的概念引入到随机估计理论中，把信号过程视为白噪声作用下的一个线性系统的输出，用状态方程来描述这种输入-输出关系，估计过程中利用系统状态方程、观测方程和白噪声激励(系统噪声和观测噪声)的统计特性形成滤波算法，由于所用的信息都是时域内的量，所以不但可以对平稳的一维的随机过程进行估计，也可以对非平稳的多维随即过程进行估计。

卡尔曼滤波方法其优点是：采用线性递归滤波的方法，可以任意一点作为初始化开始递归；计算量小，可实时计算；预测具有无偏、稳定和最优的特点。它的

提出为现代控制理论的发展做出了巨大的贡献，首先被应用于航天领域，后来在机器人导航、控制、传感器数据融合以及很多军事领域都得到了广泛的应用。目前，卡尔曼滤波方法在视频跟踪算法中也得到了非常广泛的应用，主要用来预测目标下一时刻的位置，以此作为量测算法匹配搜索的初始化起点，从而提高运算速度。

系统的状态方程：

$$X_k = A \times X_{k-1} + B \times u_{k-1} + w_{k-1} \quad (2-1)$$

观测方程：

$$z_k = H \times X_k + v_k \quad (2-2)$$

w 和 v 分别为独立不相关的状态噪声和观测噪声，并且分布满足，

$$p(w) \square N(0, Q)$$

$$p(v) \square N(0, R)$$

状态预测方程（先验估计）：

$$X(k|k-1) = A \times X(k-1|k-1) + B \times u_{k-1} \quad (2-3)$$

经观测值修正过的后验估计：

$$X(k|k) = X(k|k-1) + K_k(z_k - H X(k|k-1)) \quad (2-4)$$

其中， $X(k|k-1)$ 是 k 时刻的先验估计，即由 $k-1$ 时刻的状态向量预测得到 k 时刻的状态向量；

$X(k|k)$ 是 k 时刻的后验估计，即得到 k 时刻的观测值 z_k 后，经过观测值修正后的状态向量；

$z_k - H X(k|k-1)$ 称为差值，是观测估计和实际观测之间的误差；

K_k 是 k 时刻的滤波增益，用来最小化 k 时刻的后验估计误差协方差。

那么 k 时刻的先验估计误差和后验估计误差分别为：

$$X(k|k-1) = X(k) - X(k|k)$$

$$X(k|k) = X(k) - X(k|k) \quad (2-5)$$

先验估计误差和后验估计误差的协方差分别为：

$$P(k|k-1) = E \left[X(k|k-1) X^T(k|k-1) \right] \quad (2-6)$$

$$\begin{aligned}
P(k|k) &= E \left[X(k|k) X^T(k|k) \right] \\
&= E \left[(X(k) - X(k|k)) (X(k) - X(k|k))^T \right] \\
&= E \left[(X(k) - X(k|k-1) - K(z_k - HX(k|k-1))) \right. \\
&\quad \left. (X(k) - X(k|k-1) - K(z_k - HX(k|k-1)))^T \right] \\
&= P(k|k-1) - P(k|k-1)H^T K^T - KHP(k|k-1)^T \\
&\quad + K(HP(k|k-1)H^T + R)K^T
\end{aligned} \tag{2-7}$$

得到后验估计误差协方差 $P(k|k)$ 关于滤波增益 K_k 的表达式，为了最小化后验估计误差协方差，经 $P(k|k)$ 对 K_k 求偏导，令其结果等于零，解出滤波增益 K_k 的表达式：

$$K_k = P(k|k-1)H^T (HP(k|k-1)H^T + R)^{-1} \tag{2-8}$$

考虑两个极端的情况，滤波增益是由先验估计误差协方差 $P(k|k-1)$ 和观测噪声协方差 R 共同决定的。

$$\begin{aligned}
\lim_{R \rightarrow 0} X(k|k) &= X(k|k-1) + H^{-1}(z_k - HX(k|k-1)) = z_k \\
\lim_{P(k|k-1) \rightarrow 0} X(k|k) &= X(k|k-1) + 0 \times (z_k - HX(k|k-1)) \\
&= X(k|k-1)
\end{aligned} \tag{2-9}$$

Kalman 滤波使每一时刻 k 的目标状态向量的后验估计值的误差协方差最小，由预测和修正两步骤组成；预测部分包括状态预测和先验估计误差协方差预测；修正部分包括滤波增益的计算和利用滤波增益对状态预测值和后验估计误差协方差进行修正，计算得到状态向量的后验估计和最小化的后验估计误差协方差，具体如下：

预测部分：

$$\begin{aligned}
X(k|k-1) &= AX(k-1|k-1) + Bu_{k-1} \\
P(k|k-1) &= AP(k-1|k-1)A^T + Q
\end{aligned} \tag{2-10}$$

修正部分：

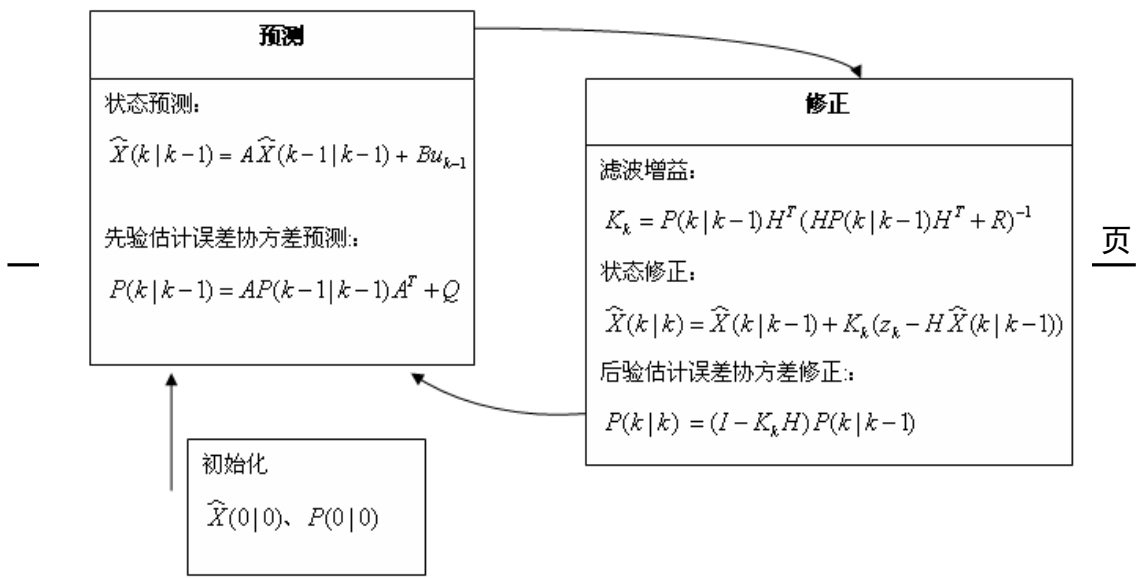


图 2-1 标准 Kalman 滤波算法递归的两个步骤

$$\begin{aligned}
 K_k &= P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1} \\
 X(k|k) &= X(k|k-1) + K_k(z_k - HX(k|k-1)) \\
 P(k|k) &= (I - K_kH)P(k|k-1)
 \end{aligned}
 \tag{2-11}$$

滤波增益 K_k 是预测与量测的权衡。当量测精确时应该增大 Kalman 增益，反之则减少 Kalman 增益。针对传统 Kalman 滤波过多依赖于历史量测，一段时间后会导致 Kalman 增益减小而失去滤波的作用，因此改进的算法增添了消渐因子 γ ($\gamma > 1$)，降低历史量测的影响，增加新近量测数据在滤波中的作用。改进 Kalman 滤波的协方差预测方程为：

$$P(k|k-1) = \gamma A(k-1)P(k-1|k-1)A^T(k-1) + Q(k-1) \tag{2-12}$$

2.3.2 扩展 Kalman 滤波算法

标准的 Kalman 滤波算法解决了线性过程中的状态估计问题，实际应用中状态方程、观测方程一般是非线性的，扩展 Kalman 滤波就是为了解决这类系统的状态估计问题。

非线性的状态方程和观测方程可以表示为：

$$\begin{aligned}
 x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}) \\
 z_k &= h(x_k, v_k)
 \end{aligned}
 \tag{2-13}$$

忽略了状态噪声和观测噪声就得到状态和观测值的近似估计，即：

$$\begin{aligned} x(k|k-1) &= f(x(k-1|k-1), u_{k-1}, 0) \\ \hat{z}(k|k-1) &= h(x(k|k-1), 0) \end{aligned} \quad (2-14)$$

为了估计非线性过程, 必须对公式(2-13)进行线性化, 分别计算函数 $f()$ 对 x 和 w 求偏导, 然后得到雅克比矩阵 A 和 W ; 对函数 $h()$ 分别对 x 和 v 求偏导, 并且得到雅克比矩阵 H 和 V ; 线性化过程为:

$$\begin{aligned} x_k &= x(k|k-1) + A(x_{k-1} - x(k-1|k-1)) + Ww_{k-1} \\ z_k &= \hat{z}(k|k-1) + H(x_k - x(k|k-1)) + Vv_k \end{aligned} \quad (2-15)$$

$$\begin{aligned} A_{[i,j]} &= \frac{\partial f_{[i]}}{\partial x_{[j]}}(x(k-1|k-1), u_{k-1}, 0) \\ W_{[i,j]} &= \frac{\partial f_{[i]}}{\partial w_{[j]}}(x(k-1|k-1), u_k, 0) \\ H_{[i,j]} &= \frac{\partial h_{[i]}}{\partial x_{[j]}}(x(k|k-1), 0) \\ V_{[i,j]} &= \frac{\partial h_{[i]}}{\partial v_{[j]}}(x(k|k-1), 0) \end{aligned} \quad (2-16)$$

线性化后的状态方程和观测方程在后验估计误差协方差最小的条件下计算扩展 Kalman 滤波的两个步骤, 预测和修正:

预测部分:

状态预测:

$$x(k|k-1) = f(x(k-1|k-1), u_{k-1}, 0) \quad (2-17)$$

先验估计误差协方差预测:

$$P(k|k-1) = A_k P(k-1|k-1) A_k^T + W_k Q_{k-1} W_k^T \quad (2-18)$$

修正部分:

滤波增益:

$$K_k = P(k|k-1) H_k^T (H_k^T P(k|k-1) H_k^T + V_k R_k V_k^T)^{-1} \quad (2-19)$$

修正过的状态向量的后验估计:

$$x(k|k) = x(k|k-1) + K_k (z_k - h(x(k|k-1), 0)) \quad (2-20)$$

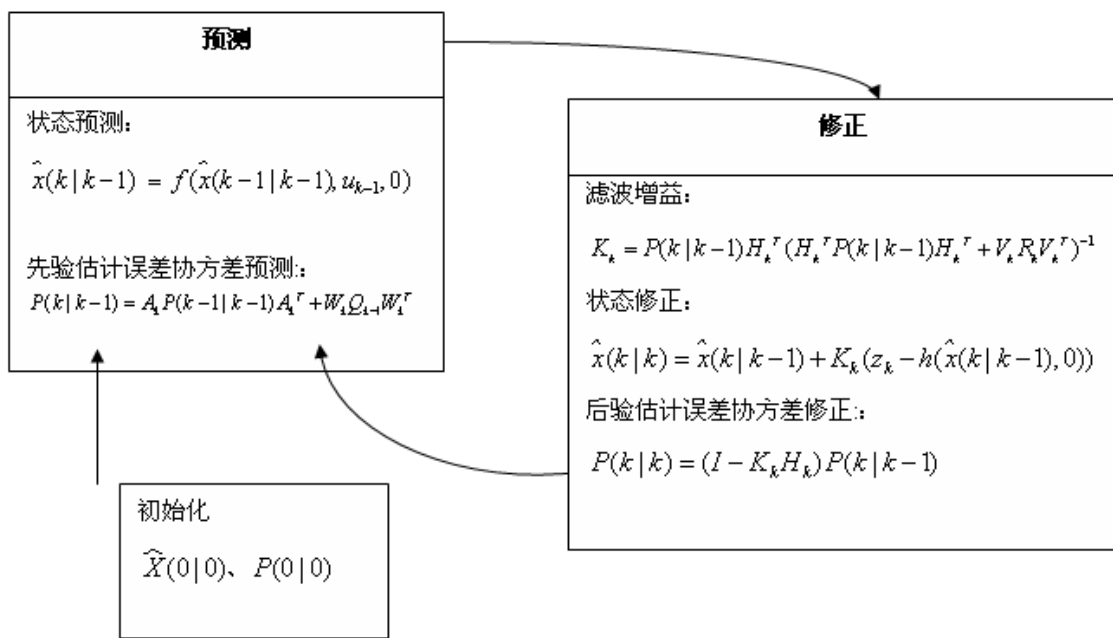


图 2-2 扩展 Kalman 滤波算法递归的两个步骤

修正后的后验估计误差协方差:

$$P(k|k) = (I - K_k H_k) P(k|k-1) \quad (2-21)$$

2.3.3 粒子滤波算法

粒子滤波 (Particle Filter), 亦称 Bootstrap 滤波、Condensation 算法、序贯蒙特卡罗 (Sequential Monte Carlo) 滤波等, 是一种目前研究较多的非线性非高斯环境下的状态估计算法。较之于 Kalman 滤波方法, 粒子滤波算法可以认为是一种条件更弱的次优目标跟踪算法。该算法的基本思想是通过一组带有相关权值的随机样本, 以及基于这些样本的估计来表示状态的后验概率分布 $p(x_k | z_{1:k})$ 。令 $\{x_{0:k}^i, \omega_k^i\}_{i=1}^{N_s}$ 表示后验概率分布 $p(x_k | z_{1:k})$ 的随机观测值, N_s 是采样点数, ω_k^i 是 k 时刻第 i 个采样点的权值, 并且满足 $\sum_i \omega_k^i = 1$, 则 k 时刻后验概率分布 $p(x_k | z_{1:k})$ 近似为:

$$p(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^{N_s} \omega_k^i \delta(x_{0:k} - x_{0:k}^i) \quad (2-22)$$

上式中 ω_k^i 可以通过重要采样 (Importance Sampling) 方法得到。若采样点集合 $x_{0:k}^i$ 可以由某重要度函数 $q(x_k | z_{1:k})$ 取得, 则第 i 个粒子的归一化权值为

$$\omega_k^i \propto \frac{p(x_{0:k}^i | z_{1:k})}{q(x_{0:k}^i | z_{1:k})} \quad (2-23)$$

根据 Markov 过程基本概念，后验概率分布 $p(x_k | z_{1:k})$ 亦可以表示为

$$\begin{aligned} p(x_{0:k} | z_{1:k}) &= \frac{p(z_k | x_k) p(x_k | x_{k-1})}{p(z_k | z_{1:k-1})} p(x_{0:k-1} | z_{1:k-1}) \\ &\propto p(z_k | x_k) p(x_k | x_{k-1}) p(x_{0:k-1} | z_{1:k-1}) \end{aligned} \quad (2-24)$$

2.3.4 Kalman 滤波的应用

利用 Kalman 滤波器来估计目标运动，当系统对目标的运动位置和速度有了可靠估计，可以在相对较小的区域内进行搜索，完成对目标的模板匹配，而且当目标被遮挡时，利用 Kalman 滤波器对目标的运动轨迹进行可靠预测，可以方便在特定区域内搜索目标，等待目标的重新出现。

Kalman 滤波器对于提高跟踪系统的处理速度和性能有着极其重要的作用。假设目标的运动状态参数为某一时刻目标的位置和速度。在跟踪过程中，由于相邻两帧图像的时间间隔比较短，目标在这么短的时间间隔内状态变化比较小，可以假设目标在单位时间间隔内是匀速运动，所以速度足以反映目标的运动趋势。

定义 Kalman 滤波器的系统状态为 x_k ，并且是一个四维向量 $(s_x, s_y, v_x, v_y)^T$ ，分别代表运动目标在坐标轴上的坐标和在坐标轴方向的速度。根据运动目标在单位时间间隔内是匀速运动的，定义状态转移矩阵 A 和观测矩阵 H ：

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ H &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{aligned} \quad (2-25)$$

Kalman 滤波算法中，初始化目标的状态参数（如：速度），即使在不准确的情况下，经过有限次的迭代滤波运算，Kalman 滤波仍然可以准确地估计出目标的真实运动状态参数（如：位置），具有预测的稳定性。

2.3.5 算法分析

基于权值计算的方式，可以将自适应算法分闭环梯度算法和开环算法。目前，

闭环算法越来越受到重视，其主要原因是这些算法的收敛速度快，而且只需要有限输入数据就能准确地描述外部环境特性。目前收敛较快的自适应算法应为 SMI 和 G-S 正交化解算法，因而成为人们首选的算法。而在这两类算法中，G-S 正交化解算法比 SMI 算法具有更好的数值特性和并行处理特性，且不需要显示计算协方差阵。闭环梯度算法实现较为简单，性能也比较可靠，不需要存储大量数据。但主要缺点是收敛于最佳权值的时间取决于数据特征值分布，在某些干扰分布情况下（当特征值差异很大时），算法收敛速度非常慢。

2.4 相似性度量方法

70年代末期，R.O.Schmidt提出了多重信号分类法（MUSIC）用于测向，具有测向精度高、分辨率高、能对多个来波信号同时测向及对瞬时短信号的测向等。此后，基于特征分解类的超分辨空间谱估计技术得到了广泛的重视和研究，提出了许多算法，如最小内积法（MN）、MUSIC根法、旋转不变技术的参数估法ESPRIT等，以及用于对相干信号源测向的空间平滑法等。在理想情况下，它们具有相对较少的计算量，而且具有传统方法无法比拟的方向超分辨能力^[46]。

2.4.1 欧式距离（Euclidean Distance）

MUSIC算法是利用接收数据的协方差矩阵分离出信号子空间和噪声子空间，利用信号方向向量与噪声子空间的正交性来构成空间扫描谱，实现信号的参数

$$d[(x_1, y_1), (x_2, y_2)] = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2-26)$$

2.4.2 加权距离（Weighted Distance）

$$d[(x_1, y_1), (x_2, y_2)] = \begin{cases} \omega_0 |x_1 - x_2| + \omega_1 |y_1 - y_2| & \text{if } |x_1 - x_2| > |y_1 - y_2| \\ \omega_0 |y_1 - y_2| + \omega_1 |x_1 - x_2| & \text{else others} \end{cases} \quad (2-27)$$

其中， $\omega_i > 0, i=1,2$ 。当 $\omega_0=1, \omega_1=1$ 时，加权距离就是街区距离（City-Block Distance）；当 $\omega_0=1, \omega_1=0$ 时，加权距离就变成了棋盘距离（Chessboard

Distance)；当 $\omega_0=1, \omega_1=\sqrt{2}-1$ 时，加权距离就变成了类欧式距离 (Quasi-Euclidean Distance)。

2.4.3 巴特查理亚系数 (Bhattacharyya)

Bhattacharyya 系数，其定义为：

$$\rho(y) \equiv \rho[p(y), q(y_0)] = \sum_{b=1}^m \sqrt{p_b(y)q_b(y_0)} \quad (2-28)$$

其几何意义为两个 m 维向量 $(\sqrt{p_1}, \dots, \sqrt{p_m})$ 和 $(\sqrt{q_1}, \dots, \sqrt{q_m})$ 夹角的余弦。

2.4.4 Hausdorff 距离

Hausdorff 距离 (Hausdorff Distance, HD)，描述了两个特征点集之间的相似程度的一种度量，即有限像素点集合之间的一种距离定义形式。

已知两组有限像素点集合 $A = \{a_1, a_2, \dots, a_p\}$ 和 $B = \{b_1, b_2, \dots, b_q\}$ ，则 HD 距离定义如下：

$$H(A, B) = \max\{h(A, B), h(B, A)\} \quad (2-29)$$

其中

$h(A, B) = \max_{a_i \in A} \min_{b_j \in B} \|b_j - a_i\|$ ，称为 A 到 B 的有向 HD，表示点集 A 中所有点到点集 B 的距离 $d_B(a_i)$ 的最大值，并且 $d_B(a_i) = \min_{b_j \in B} \|a_i - b_j\|$ ，表示点到集合的距离，是改点到集合 B 中每一点的距离的最小值。

2.4.5 改进的 Hausdorff 距离

HD 的改进算法都是对有向 HD 的定义做了修正。

部分 HD 先求出点集 A 中所有点到点集 B 的距离，然后将这些距离由小到大排序，其中序号 k 的距离即为 $h_k(A, B)$ ；部分 HD 的有向 HD 定义如下：

$$h_k(A, B) = K_{a_i \in A}^{th} d_B(a_i) \quad (2-30)$$

基于平均距离值的 MHD, 有向 MHD 的改进如下:

$$h_{MHD}(A, B) = \frac{1}{N_A} \sum_{a_i \in A} d_B(a_i) \quad (2-31)$$

部分 HD 和平均距离值的 MHD 的融合距离定义如下:

$$h_{LTS}(A, B) = \frac{1}{k \times N_A} d_B(a_i)_{(m)} \quad (2-32)$$

2.5 匹配搜索的优化算法

贝叶斯方法是基于统计理论的一种经典方法, 适合于有关参数估计问题。最大似然 (ML) 估计方法就是贝叶斯估计方法中的一种特例, 是在已知白噪声情况下的贝叶斯最优估计。在 ML 算法中, 观测所得信号的似然函数被定义为含有未知参数的条件概率密度函数, 目的是选定未知参数以使得该似然函数尽可能大。通过最大化似然函数求出的解都被认为是未知参数的一个估计^[1]。

子空间拟合 (SF) 的概念是 20 世纪 90 年代之后提出的, 其基本思想在于构造这样一个事实, 即阵列流型矩阵与阵列接收数据的子空间之间存在一个拟合关系。SF 算法与 ML 算法的相同点在于: 这个拟合关系也是一个最大或最小化的似然函数求解问题, 都可通过最大或最小化似然函数求出未知参数的一个估计^[1]。

2.5.1 Mean-Shift 理论基础

给定 n 个 d 维空间 R^d 中的点集 $\{x_i\}_{i=1,2,\dots,n}$, 利用核函数 $K(x)$ 、核函数窗宽 h , 在点 x 处估算的概率密度为:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (2-33)$$

较常用的核函数是 Epanechnikov 核和高斯核, 其表达式分别为:

$$\text{Epanechnikov 核: } K_E(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1-\|x\|^2), & \text{if } \|x\| < 1 \\ 0, & \text{if } \|x\| \geq 1 \end{cases} \quad (2-34)$$

高斯核:
$$K_N(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}\|x\|^2) \quad (2-35)$$

上式(2-34)中 c_d 是 d 维球体的体积, 例如, 通常取 $c_1 = 2, c_2 = \pi, c_3 = 4\pi/3$ 等。

定义核函数 K 的剖面 (Profile) 函数为 k , 并且满足 $k: [0, \infty) \rightarrow R$, 即 $K(x) = k(\|x\|^2)$ 。则上述两种核函数的剖面函数分别为:

$$k_E(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1-x^2), & \text{if } x < 1 \\ 0, & \text{if } x \geq 1 \end{cases} \quad (2-36)$$

$$k_N(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}x^2) \quad (2-37)$$

此时公式(2-33)用剖面函数表示为:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \quad (2-38)$$

核函数直方图:

已知中心坐标为 y_0 目标模板图像 $\{x_i\}_{i=1,2,\dots,u}$, 模板图像的核函数直方图表示 $\{q_b(y_0)\}_{b=1,2,\dots,m}$ 为:

$$q_b(y_0) = C_1 \sum_{i=1}^n k\left(\left\|\frac{y_0-x_i}{h_1}\right\|^2\right) \delta[I(x_i)-b], \quad b=1,2,\dots,m \quad (2-39)$$

其中 $q_b(y_0)$ 是分量 b 的值, m 是直方图级的个数。 $I(x_i)$ 是 x_i 处像素的量化值, h_1 是

核函数窗宽。常数 C_1 由约束条件 $\sum_{b=1}^m q_b(y_0) = 1$ 得到, 即
$$C_1 = \frac{1}{\sum_{i=1}^n k\left(\left\|\frac{y_0-x_i}{h_1}\right\|^2\right)}$$

同理, 对于候选目标区域, 像素集合 $\{x_i\}, i=1,2,\dots,N$, 该图像区域的中心位置

预测位置为 y ，则候选目标的核函数直方图模型 $\{p_b(y)\}, b=1, 2, \dots, m$ 为：

$$p_b(y) = C_2 \sum_{i=1}^N k\left(\left\|\frac{y-x_i}{h_2}\right\|^2\right) \delta[I(x_i)-b], \quad b=1, 2, \dots, m \quad (2-40)$$

当前帧图像中，候选目标区域与目标模板之间的相似性采用 Bhattacharyya 系数度量，由定义公式 (2-28) 知其 Bhattacharyya 系数为：

$$\rho(y) \equiv \rho[p(y), q(y_0)] = \sum_{b=1}^m \sqrt{p_b(y)q_b(y_0)} \quad (2-41)$$

由于目标运动的连续性，Mean-Shift 算法在当前帧图像中搜索的位置是从 Kalman 的预测位置 y_0 开始的。对公式 (4-6) 在 y_0 处泰勒展开并省略高次项得到：

$$\rho[p(y), q(y_0)] \approx \frac{1}{2} \sum_{b=1}^m \sqrt{p_b(y_0)q_b(y_0)} + \frac{1}{2} \sum_{b=1}^m p_b(y) \sqrt{\frac{q_b(y_0)}{p_b(y_0)}} \quad (2-42)$$

这种近似处理候选目标在预测位置 y_0 变化不大时是有效的。将公式 (4-5) 代入上式 (2-42) 得到：

$$\rho[p(y), q(y_0)] \approx \frac{1}{2} \sum_{b=1}^m \sqrt{p_b(y_0)q_b(y_0)} + \frac{C_2}{2} \sum_{i=1}^d \omega_i k\left(\left\|\frac{y-x_i}{h_2}\right\|^2\right) \quad (2-43)$$

其中，

$$\omega_i = \sum_{b=1}^m \sqrt{\frac{q_b(y_0)}{p_b(y_0)}} \delta[I(x_i)-b] \quad (2-44)$$

公式 (4-7) 中的第一子项在 y_0 和 y_0 已知时是定值与 y 无关，因此第二子项是在当前帧位置 y 处利用 ω_i 加权的核剖面函数 $k(x)$ 估算的概率密度。概率密度的极值问题可以利用 Mean-Shift 理论求得。

将 $\rho(y)$ 对 y 求偏导，并且使其等于零：

$$\frac{\partial \rho}{\partial y} = 0 + \frac{C_h}{2} \sum_{i=1}^{N_h} \omega_i g\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \times 2 \times \frac{y-x_i}{h^2} = 0 \quad (2-28) (2-45)$$

其中， $g(x) = -dk/dx$ 表示核函数剖面函数的负梯度方向。

得到当 $\rho(y)$ 取最大值时的 y_1 为:

$$y_1 = \frac{\sum_{i=1}^{N_h} w_i x_i}{\sum_{i=1}^{N_h} w_i} \quad (2-46)$$

2.5.2 算法分析

Mean-Shift 算法的形式非常简洁优美, 收敛条件也很宽松。只要确定了跟踪目标, 整个跟踪过程就不需额外的参数输入。与通常的优化方法不同, Mean-Shift 方法每一步都能使概率密度估计 $f(x)$ 获得增长, 具有极佳的稳定性。由于无需对密度分布 $f(x)$ 进行估计, 就可以直接沿着梯度方向搜索局部最大值, 大大减少了运算量, 所以这种算法具有良好的实时性, 因而基于 Mean-Shift 优化算法代表了视频跟踪算法的发展水平, 得到了计算机视觉界的广泛关注。

Mean-Shift 算法用于视频跟踪有以下几个优势: 首先, 算法计算量不大, 在目标区域已知的情况下完全可以做到实时跟踪; 其次, 作为一个无参数密度估计算法, 很容易作为一个模块和别的算法集成; 然后, 采用核函数直方图建模, 对部分遮挡、目标旋转、形变和背景运动不敏感。但是, 它也有不足之处, 比如: 缺乏必要的模板更新算法, 整个跟踪过程中窗宽的大小保持不变, 当目标有尺度变化时, 可能跟踪失败; 颜色直方图是一种比较通用的目标特征描述子, 当背景和目标的颜色直方图相似时, 跟踪算法效果不好; 另外, Mean-Shift 算法是局部最优的优化算法, 当出现多个局部峰值时, 算法可能会收敛于局部最优值而造成跟踪算法失效。

2.6 本章小结

本章主要介绍了空间谱估计中的波束形成法、线性预测法、部分空间法、参数法, 并对每种算法进行了展开讨论分析比较, 为后面的迭代算法和多重信号分类算法用于均匀线阵和均匀圆阵奠定了基础。

第3章 基于 SIFT 特征匹配的运动目标跟踪算法

目标跟踪在制导、视觉导航、安全监控等方面有着广泛的应用^[1]。随着数字化技术的普及，视频中的运动目标跟踪已经成为计算机视觉领域中的研究热点之一，具有重要的应用价值。视频目标跟踪问题本身汇集了图像处理、模式识别、人工智能、自动控制技术等众多领域的技术，具有很好的理论研究意义。

3.1 概述

视频中的运动目标跟踪是指通过计算运动目标在每帧图像中的位置、速度等状态信息及状态信息在图像帧之间的对应和匹配关系来刻画目标的运动轨迹和运动趋势（如图 3-1 所示）。一般来说，目标跟踪问题包含“量测”和“预测”两个部分，但有时跟踪算法也会只包括“量测”而不包含“预测”部分。“量测”可以看作目标模板和候选区域的匹配过程，“预测”则是对于目标运动参数的估计。目标运动“预测”已经有很多方法和框架，其中 Kalman 滤波是一个简单有效的传统方法^[29]。基于概率框架的“粒子滤波”方法也得到研究人员的重视^{[23][24][25]}。

尽管近年来目标跟踪技术有了长足的发展，但在实际应用中还存在以下主要问题^[1]：

- 1) 背景对目标的遮挡和目标的自遮挡；
- 2) 运动目标在跟踪过程中的尺度伸缩变化和形变。

解决目标遮挡问题方法大致分为两类^[40]：Merge-Split（MS）方法和 Straight-Through（ST）方法。MS 方法通过一系列创建、删除、合并和分裂算子，对遮挡情况下的目标区域进行合并或者分裂运算，实现运动目标的鲁棒跟踪；ST 方法则通过直接提取对遮挡不敏感的特征对目标区域进行跟踪，不需要进行复杂的合并和分裂运算。两类方法都依赖对目标遮挡情况的检测。针对尺度伸缩问题，有学者从“量测”的角度进行考虑提出了基于灰度直方图匹配的跟踪算法，利用直方图反映灰度级的分布特性，来解决运动目标尺度等比例伸缩变化时的鲁棒跟踪。但是由于直方图不能反映目标的结构信息，因此当场景与目标的灰度信息一致时就会造成跟踪算法失效。跟踪系统中非刚体目标在发生形变和旋转情况下的

鲁棒跟踪是视频中运动目标跟踪领域的难点和热点问题之一^{[1][34]}。

本文的研究目的是针对视频跟踪中的典型应用场景，解决视频跟踪中的运动目标部分遮挡和尺度伸缩问题，取得尽可能高的跟踪精度，因此，设计了一种基于 Kalman 滤波预测和 SIFT (Scale-Invariant Feature Transform) 特征点匹配量测的目标跟踪算法。算法分为 Kalman 滤波预测、SIFT 特征点提取、特征点匹配和运动一致性聚类、噪声干扰点剔除和目标模板更新等步骤。

本章的组织结构如下：第 1 节介绍视频中运动目标跟踪的问题和研究现状，第 2 节详细 SIFT 特征点检测提取和表达，详细叙述基于 SIFT 特征匹配的运动目标跟踪算法，第 3 节是实验结果分析和对比，第 4 节是结论和展望。

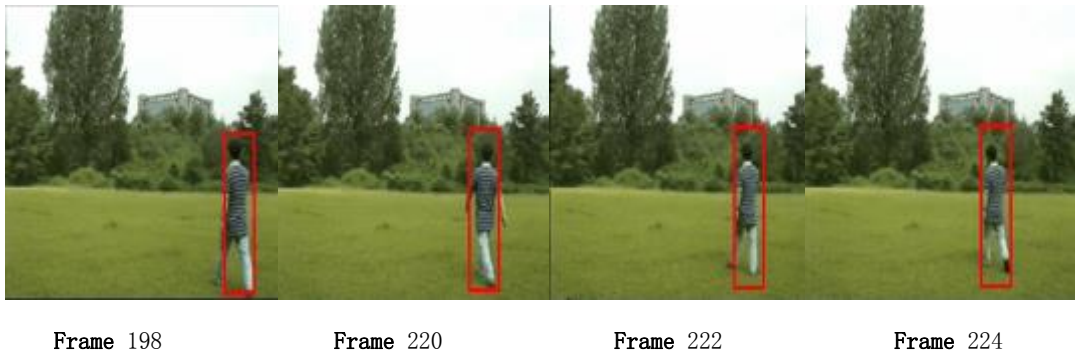


图 3-1 视频图像中的运动目标跟踪示意

3.2 SIFT 特征点提取和相似性度量

图像处理中通常提取边缘点或者角点作为特征来进行目标识别、跟踪和模式分析^{[1][10][11][16][23]}。角点是图像中颜色信息变化显著边缘的交点或者图像的强度信息在水平和垂直方向均变化显著的点。目前常用的角点提取算法有：Harris Corners^[11]、KLT Corners^{[9][10]}、SIFT Corners^{[13][14][15]}和 SUSAN Corners 等。

3.2.1 SIFT 特征点介绍

SIFT 是 D. Lowe 提出的一种具有尺度和旋转不变性特征点提取算法^{[13][14][15]}，是一种鲁棒性很高的局部特征表示方法，本质上基于高斯差分图像金字塔的多尺度极值点检测和梯度方向表达的方法，具有很高的匹配精度和鲁棒性，主要用于图像配准、基于内容的图像检索、人脸跟踪等领域。SIFT 具有尺度伸缩不变性和旋转不变性，并且在一定程度的仿射变换、不同视角、环境噪声、光线亮度变化等

情况下具有很高的匹配精度^{[13][14]}。SIFT 角点具有以下优点：1) 在空域和频域都有很好的定位效果和精度，降低了目标遮挡、复杂场景以及噪声的影响，这对于运动目标跟踪尤其重要；2) 具有很高的匹配效率，特征点有很好的可区分性，即便是保留很少特征点的图像也有很高的正确匹配概率；3) 基于多尺度图像金字塔的结构，特征点的计算量大大降低。

3.2.2 SIFT 特征点提取和表达

SIFT 本质上是由多尺度高斯图像的极值点检测和梯度方向直方图表达两部分组成。计算方法描述如下^{[13][14]}。

将原始图像 $I(x, y)$ 经过高斯算子和 DoG (Difference of Gauss) 算子计算分别得到多尺度高斯金字塔图像 $L(x, y, \sigma)$ 和 DoG 图像 $D(x, y, \sigma)$ ：

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3-1}$$

$$\begin{aligned} D(x, y, \sigma) &= L(x, y, k_i \sigma) - L(x, y, \sigma) \\ &= G(x, y, k_i \sigma) * I(x, y) - G(x, y, \sigma) * I(x, y) \\ &= (G(x, y, k_i \sigma) - G(x, y, \sigma)) * I(x, y) \end{aligned} \tag{3-2}$$

其中， $G(x, y, \sigma)$ 为高斯算子： $G(x, y, \sigma) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\frac{x^2 + y^2}{2\sigma^2})$ ；DoG 图

像 $D(x, y, \sigma)$ 是尺度空间中尺度 s_i 上相邻高斯滤波图像 $L(x, y, \sigma)$ 和 $L(x, y, k_i \sigma)$ 的差

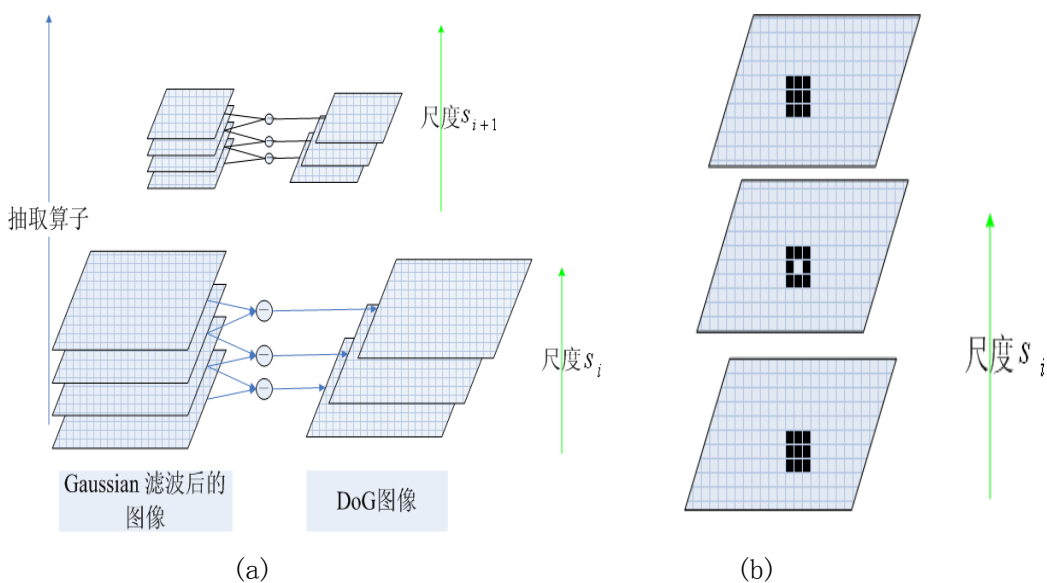


图 3-2 (a) 图像在尺度空间上的 Gaussian 滤波后的图像和 DoG 图像
(b) 计算尺度空间上 DoG 图像的 SIFT 特征点

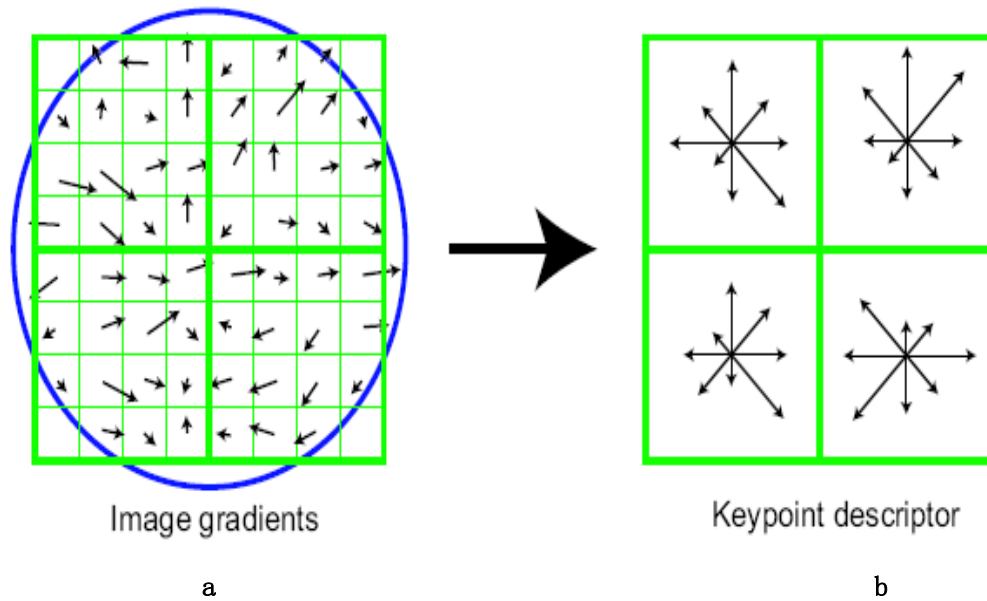


图 3-3 SIFT 特征点的梯度方向直方图表示

a 2×2 邻域子块的梯度方向计算

b 2×2 邻域子块的八位梯度方向直方图

图像（如图 3-2a 所示）， k_i 是一组尺度为 s_i 的高斯图像的高斯核的系数。

计算尺度 s_i 上 DoG 图像 $D(x, y, k_i \sigma)$ 的像素点 (x, y) 在当前尺度图像中的八邻域和相邻 DoG 图像 $D(x, y, k_{i-1} \sigma)$ 和 $D(x, y, k_{i+1} \sigma)$ 的 9×2 邻域中是否局部极值点（如图 3-2 b 所示），如果是极值点则把该像素点作为该图像尺度 s_i 上的候选特征点。

每个特征点 h_i 由其相邻的 4×4 子块的梯度方向直方图 $F(h_i)$ 、位置信息 $L(h_i)$ 、尺度信息 $S(h_i)$ 和主方向 $O(h_i)$ 组成，其中，每个子块的大小是 8 像素 \times 8 像素，梯度方向直方图 $F(h_i)$ 的每个分量的值是特征点 h_i 的邻域内所有像素点在该方向分量上的梯度幅值的和，梯度方向直方图的峰值所对应的方向设置为该候选特征点的主方向 $O(h_i)$ 。特征点的梯度方向直方图分量是相对于特征点主方向的相对量，当运动目标发生旋转时，目标的 SIFT 特征点的方向直方图分量是相对不变的，因此，SIFT 特征点具有一定的旋转不变性。

SIFT 特征点的计算步骤（SIFT 提取效果如图 3-5 所示）：

1). 对输入图像 $I(x, y)$ ，在尺度 s_i 上用一组 m 个高斯核

$\{(0, \sigma_1^2), (0, \sigma_2^2), \dots, (0, \sigma_m^2)\}$ 组成高斯算子进行滤波，由公式 (3-1) 得到一组

滤波后的高斯图像 $L(x, y, \sigma_i)$ ；其中 $m=4$ ； $\sigma_{i+1}^2 = k\sigma_i^2$ ， $k=2$ ；

- 2). 利用 DoG 算子由公式 (3-2) 计算尺度 s_i 上该组高斯图像 $L(x, y, \sigma_i)$ 相邻图像的 DoG 图像 $D(x, y, \sigma_i)$ ， $1 \leq i \leq m-1$ ；
- 3). 对步骤 1) 得到的一组高斯图像 $L(x, y, \sigma_i)$ 进行抽取，在多尺度空间中得到新尺度 s_{i+1} 上的一组高斯图像 $L(x, y, \sigma_{i+1})$ ，重复步骤 2) 计算尺度 s_{i+1} 上的一组 DoG 图像；
- 4). 计算尺度空间中每个尺度 s_i 上 DoG 图像的特征点，其中特征点是尺度 s_i 上 $m-1$ 个 DoG 图像中 26 邻域内的极值点（如图 3-2b 所示）；
- 5). 每个特征点 h_i 由 $\{F(h_i), L(h_i), S(h_i), O(h_i)\}$ 组成，分别表示特征点的 4×4 个窗口的方向直方图、位置信息、尺度信息和主方向。

3.2.3 SIFT 特征点相似性度量

候选目标的特征点 h_i 和目标模板特征点 T_j 的相似性度量：

$$dist = \sqrt{\sum_{m=1}^n [F(h_i)_m - F(T_j)_m]^2} \quad (3-3)$$

其中， $dist$ — h_i 和 T_j 的 n 阶方向直方图的欧式距离， n 取值 128， i, j 分别表示候选目标和模板的特征点标号。

由于跟踪过程中运动目标的变化具有连续性和渐进性，特征点匹配的相似性度量还可以参考特征点的位置、尺度等信息，其中尺度信息还是目标模板更新策略的重要依据，目标模板更新策略是下一步项目研究的重点之一^[28]。利用特征点方向直方图的欧式距离作为相似性度量，在跟踪实验中已经得到较好的跟踪效果，在实验分析部分有详细的论述。

依据特征点相似性度量，在每一帧图像中计算目标模板特征点的最佳候选匹配特征点，这些候选匹配特征点 h_i 包含方向直方图 $F(h_i)$ 、位置 $L(h_i)$ 、尺度 $S(h_i)$ 和主方向 $O(h_i)$ 等信息，同一目标的特征点具有运动信息的一致性，因此利用特征点

运动信息的一致性聚类算法，得到运动目标更为鲁棒的状态信息。

3.3 最大模糊熵的高斯聚类算法

采用最大模糊熵的高斯聚类算法^[30]对特征点运动信息 $x_{t,k}$ 聚类，其中 $x_{t,k} = (s, v_x, v_y)$ 由 t 时刻第 k 个待聚类特征点与前一帧匹配特征点的尺度比值 s 和在 x 轴、 y 轴方向上的运动分量 v_x 、 v_y 构成。

损失函数 L 定义为：

$$L = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik} d_{ik}^2 \quad (3-4)$$

其中， c 为类的数目，即待跟踪目标的个数； n 为待聚类运动向量的个数；

μ_{ik} — 第 k 个待聚类运动向量属于第 i 类的隶属度函数，即第 k 个特征点的运动向量属于第 i 个目标隶属度函数，定义为：

$$\mu_{ik} = \frac{\exp(-\frac{d_{ik}^2}{2\sigma^2})}{\sum_{j=1}^c \exp(-\frac{d_{jk}^2}{2\sigma^2})} \quad i = 1, 2, \dots, c; k = 1, 2, \dots, n; \sum_{i=1}^c \mu_{ik} = 1 \quad (3-5)$$

d_{ik}^2 — 特征点运动信息 $x_{t,k}$ 到第 i 类的聚类中心 C_i 的距离， $d_{ik}^2 = \|x_{t,k} - C_i\|^2$ ；

第 i 类的聚类中心 C_i 为：

$$C_i = \frac{\sum_{k=1}^n \mu_{ik} x_{t,k}}{\sum_{k=1}^n \mu_{ik}}, \quad (1 \leq i \leq c) \quad (3-6)$$

最大模糊熵的高斯聚类步骤如下：

- 1). 确定需要聚类的数目 c （即被跟踪目标的个数）和由 Kalman 滤波的预测值初始化每个聚类的中心 $C_i(t), (1 \leq i \leq c)$ ；
- 2). 确定算法的参数：“钟型”隶属度函数的方差 $\sigma = 1.5$ 、最大迭代的次数 T 、

判断相邻次迭代的隶属度趋于一致的阈值 $\varepsilon = 0.001$;

- 3). 用当前的聚类中心 $C_i(t)$, 根据公式 (3-5)。计算各个待聚类特征点的隶属度 $\mu_{ik}(t+1)$;
- 4). 用当前的隶属度函数, 根据公式 (3-6) 更新每个新类的聚类中心 $C_i(t+1)$;
- 5). 若 $\max |\mu_{ik}(t) - \mu_{ik}(t+1)| < \varepsilon$ 或者迭代次数 $t > T$, 算法停止; 否则, $t = t+1$ 重复步骤 3)——5)。

经过聚类以后, 选取包含最多特征的那个类代表整个运动物体, 并且使用此类的类中心 $C_i(s, v_x, v_y)$ 代表物体的运动信息。通过对运动目标特征点的运动信息和特征信息进行聚类分析, 可以去除干扰和噪声的误匹配, 达到更精确的跟踪效果。

3.4 基于 SIFT 特征匹配和 Kalman 滤波的运动目标跟踪算法研究

3.4.1 概述

基于 Kalman 滤波的运动目标跟踪系统主要包括两部分: 预测和量测。预测为下一时刻目标检测提供一个量测的初始位置, 既提高了跟踪的准确性又减少了计算量; 量测的目的是获取当前时刻目标的精确状态信息, 为下一步预测提供修正的基础。由于视频相邻两帧的时间间隔很短, 而且目标运动是连续变化的, 因此可以利用历史的量测信息对目标在下一帧图像中的位置进行预测, 然后在以预测位置为中心的预测邻域内进行搜索, 降低目标匹配的计算量和噪声干扰造成的误差, 提高跟踪的精度和算法的速度。

本文设计的基于 Kalman 滤波预测和模板特征匹配量测的目标跟踪算法 (如图 3-4 所示), 该算法在改进的 Kalman 滤波的预测框架下, 提取具有众多不变性优点的 SIFT 特征, 利用改进的特征点匹配算法计算候选目标的精确位置, 同时对特征点运动向量进行聚类分析有效地剔除噪声干扰等错误匹配, 采用交互的模板更新策略对目标模板进行有效地更新。公开测试序列上的大量实验表明, 该算法能够有效地解决视频中运动目标跟踪的部分遮挡问题、尺度伸缩问题、场景光线亮度变化问题和非刚体目标的旋转问题。

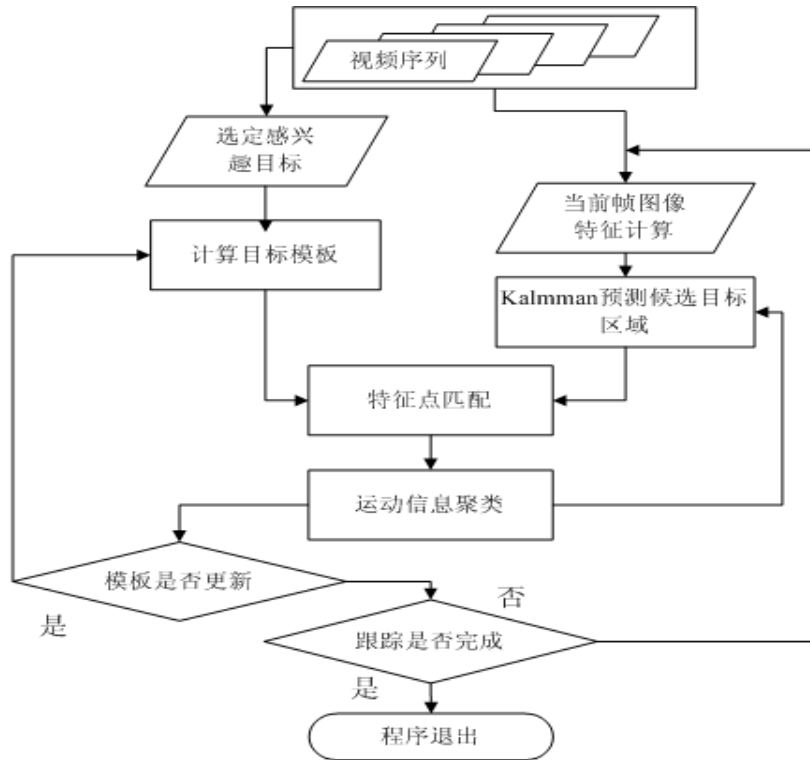


图 3-4 基于 SIFT 特征匹配和 Kalman 滤波跟踪算法的框图

3.4.2 改进的 Kalman 滤波算法

Kalman 滤波器是对动态系统的状态序列进行线性最小均方误差估计的算法，其优点是：采用线性递归滤波的方法，可以任意一点作为初始化开始递归；计算量小，可实时计算；预测具有无偏、稳定和最优。Kalman 系统的状态方程为：

$$X_k = A \times X_{k-1} + w_{k-1} \quad (3-7)$$

系统的观测方程为：

$$z_k = H \times X_k + v_k \quad (3-8)$$

k 时刻的后验估计误差的协方差为：

$$P(k|k) = P(k|k-1) - P(k|k-1)H^T K^T - KHP(k|k-1)^T + K(HP(k|k-1)H^T + R)K^T \quad (3-9)$$

得到后验估计误差协方差 $P(k|k)$ 关于滤波增益 K_k 的表达式，为了最小化后验

估计误差协方差, 经 $P(k|k)$ 对 K_k 求偏导, 令其结果等于零, 解出滤波增益 K_k 的表达式为:

$$K_k = P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1} \quad (3-10)$$

滤波增益 K_k 是预测与量测的权衡。当量测精确时应该增大 Kalman 增益, 反之则减少 Kalman 增益。针对传统 Kalman 滤波过多依赖于历史量测, 一段时间后会导导致 Kalman 增益减小而失去滤波的作用, 因此改进的算法增添了消渐因子 γ ($\gamma > 1$), 降低历史量测的影响, 增加新近量测数据在滤波中的作用。改进 Kalman 滤波的协方差预测方程为:

$$P(k|k-1) = \gamma A(k-1)P(k-1|k-1)A^T(k-1) + Q(k-1) \quad (3-11)$$

3.5 实验结果和算法分析

SIFT 特征点具有尺度伸缩和旋转不变性, 以及在不同视角、光线亮度变化、环境噪声等情况下具有很高的匹配率。对本文提出的基于 SIFT 特征匹配和 Kalman 滤波的运动目标跟踪算法, 在运动目标伴随旋转、形变及显著尺度变化、部分及完全遮挡和光线亮度变化等复杂场景的视频序列上进行实验, 实验结果表明, 本文所提出的算法是有效的鲁棒跟踪算法。跟踪算法是用 Visual Studio 6.0 和 OpenCV 软件包开发的, 所有实验在 CPU 3.0 GHz、1 G 内存和独立显卡的 PC 机上进行的, 实验 3.5.2 采用的实验视频是分辨率为 320×240 、帧率为 25 fps 和 AVI 的公开的测试序列, 其余的实验视频时 SDL 实验室自己设计的复杂场景的 AVI 视频数据, 分辨率均为 640×480 和帧率为 30 fps。

3.5.1 尺度伸缩和形变的非刚体运动目标的跟踪实验

本实验采用的测试视频待跟踪的目标是非刚体的人体, 整个跟踪的过程运动目标不断发生形变和旋转以及显著的尺度缩小。首先计算感兴趣运动目标模板的 SIFT 特征点, 然后在 Kalman 滤波预测的框架下, 计算每帧图像中最佳候选目标区域的 SIFT 特征点, 根据 SIFT 特征点匹配和特征点运动信息聚类分析得到最佳候选目标的特征点 (如图 3-5、图 3-6 和图 3-7 所示), 计算最佳候选目标区域和最近候选的质心坐标 (图 3-8、图 3-9 和图 3-10 所示), 跟踪结果图是大约二三十



Frame 230

Frame 250

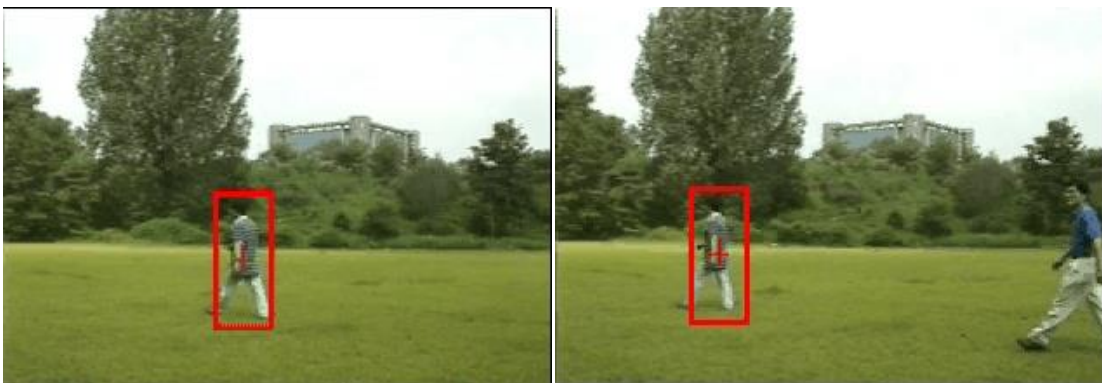
图 3-8 运动目标尺度和形变情况下的基于 SIFT 特征点的跟踪结果



Frame 270

Frame 290

图 3-9 运动目标尺度和形变情况下的基于 SIFT 特征点的跟踪结果



Frame 320

Frame 350

图 3-10 运动目标尺度和形变情况下的基于 SIFT 特征点的跟踪结果



Frame 10

Frame 20

图 3-12 运动目标部分遮挡和尺度伸缩情况下的基于 SIFT 特征点的跟踪结果



Frame 30

Frame 40

图 3-11 运动目标部分遮挡和尺度伸缩情况下的基于 SIFT 特征点的跟踪结果

3.5.2 部分及完全遮挡情况下的非刚体运动目标的跟踪实验

本实验分别采用栅栏前的行人作为跟踪目标，人体在前进的过程中时常发生部分或者完全遮挡，并且行人也伴随着一定的形变和尺度变化。由于本文跟踪算法中提取的 SIFT 特征点具有尺度不变性和唯一可区分性，即便只有少数的特征点也可以保持很好的匹配效果，Kalman 滤波算法当运动目标完全遮挡时候预测算法占了更大的权重，实验结果表明该算法具有尺度伸缩不变性和抗遮挡的特性，跟踪结果图 3-12、图 3-11、图 3-13 和图 3-14 所示的是测试序列第 10—68 帧图

像中发生遮挡前、遮挡中和遮挡后的抓拍跟踪效果图像，如第 10 帧图像为运动目标遮挡前的跟踪效果，第 20 帧图像为部分遮挡中的跟踪效果，第 30 帧和 40 帧为部分遮挡后的跟踪效果；第 50 帧、65 帧、66 帧和 68 帧图像分别表示运动目标完全遮挡前、遮挡时和遮挡后的跟踪效果。

3.5.3 光线亮度变化及部分遮挡情况下的刚体运动目标的跟踪实验

场景光线亮度信息变化往往是传统的基于颜色信息的跟踪算法失效（如灰度直方图匹配和像素的相关匹配等算法），该实验采用亮度信息突然变亮的小轿车作为



Frame 50

Frame 64

图 3-13 运动目标完全遮挡和尺度伸缩情况下的基于 SIFT 特征点的跟踪结果



Frame 66

Frame 68

图 3-14 运动目标完全遮挡和尺度伸缩情况下的基于 SIFT 特征点的跟踪结果

跟踪的运动目标（如图 3-16 所示），其中第 89 帧运动目标图像突然变亮，运动目标在跟踪过程中还伴随有一定旋转和形变，以及由于背景造成的部分遮挡，由于 SIFT 特征反映了目标的整个图像区域的梯度变化信息，光线亮度的变化并没有太多影响到目标图像的特征点提取和匹配，因此在场景光线亮度信息显著变化的情况下，本文的算法也能够实现较好的跟踪效果，图 3-15、图 3-16 和图 3-17 所示的是跟踪过程中抓拍的跟踪效果，其中第 89 帧图像发生了图像的场景光线变亮，第 87 帧、91 帧和 93 帧图像显示的是发生部分遮挡情况的跟踪效果。



Frame 73

Frame 83

图 3-15 部分遮挡情况下的刚体运动目标的跟踪结果



Frame 87

Frame 89

图 3-16 部分遮挡和光线亮度变化情况下的刚体运动目标的跟踪结果



图 3-17 部分遮挡情况下的刚体运动目标跟踪结果

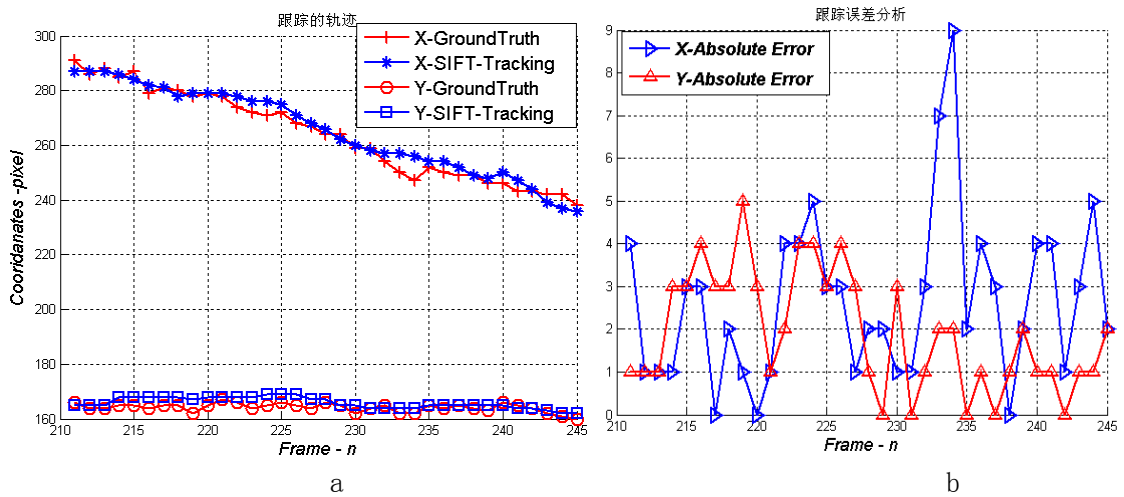


图 3-18 实验 3.5.1 采用的视频序列的误差分析

a 跟踪轨迹与真实轨迹比较 b 跟踪算法在水平和垂直方向的误差

3.5.4 跟踪算法误差分析

采用实验 3.5.1 的跟踪结果，与人工标定运动目标的真实位置做误差分析，分别得到水平和垂直方向的跟踪轨迹分析和跟踪的标准误差分布，得到实验 3.5.1 测试视频中第 212—245 帧运动目标的质心在水平和垂直方向上的跟踪轨迹和真实坐标（如图 3-18a 所示），水平和垂直方向上的误差分析（如图 3-18b 所示），该实验的跟踪效果水平方向的平均误差是 2.6286 个像素、垂直方向上的平均误差为

Tracking Algorithm	AXIS	Mean Error	Max Error	Min Error
SIFT	X	2.6286	9	0
	Y	1.9143	5	0

图 3-19 SIFT 跟踪算法的平均、最大和最小误差
X 表示水平方向的误差，Y 表示垂直方向的误差

1.9143 个像素（如图 3-19 所示）。

3.6 本章小结

本章融合 Kalman 滤波方法和 SIFT 特征匹配量测方法, 提出了一种有效的运动目标跟踪算法。该算法首先提取 SIFT 特征点表示运动目标, 然后基于欧式距离和近邻搜索进行特征匹配, 并且对匹配结果进行聚类分析以去除噪声干扰和误匹配。最后, 在 Kalman 滤波框架内, 结合特征点匹配结果和滤波预测进行目标跟踪。大量实验结果表明, 本文中用于目标量测的 SFIT 特征具有尺度伸缩不变性及旋转不变性, 对部分形变和亮度变化不敏感。特征点匹配聚类和 Kalman 滤波框架的结合能够有效解决跟踪中的运动目标尺度伸缩和部分遮挡问题。

基于 SIFT 特征匹配和 Kalman 滤波的运动目标跟踪算法, 属于基于特征的跟踪算法研究的一个尝试, 算法跟踪结果的精度较高, 水平和垂直方向上的跟踪误差不超过 3 个像素点; 但是由于运动目标的 SIFT 特征点本质上是多尺度高斯差分图像的极值点检测以及 128 维梯度方向直方图表达和匹配计算量很大, 目前的计算机处理的能力尚未能满足跟踪系统实时处理的要求。

第4章 基于综合直方图的运动目标跟踪算法

4.1 引言

通过实验分析发现,基于轮廓的跟踪算法的稳定性受到部分或者全部遮挡情况的严重制约^{[26][40]};基于模型的跟踪算法,在无先验知识的实际跟踪系统中很难得到目标的模型^[27];基于区域统计特性的跟踪算法(如基于传统颜色直方图的跟踪算法),利用的是目标图像的全局描述信息,很难精确地定位目标的位置,跟踪效果的稳定性和精度不高^{[38][62]};基于特征表达的跟踪算法,通过提取具有尺度伸缩、旋转等不变性的显著特征信息(如 SIFT、KLT、Harris Corners、SUSAN 等),可以解决尺度伸缩和部分遮挡等问题,具有很好的跟踪精度,但是尚未能满足实时处理的要求^{[7]—[16]}。

传统的颜色直方图自身是图像区域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,利用直方图计算简单和能够满足实时处理的要求,并且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。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^{[17][18]},有学者提出改进的直方图方法用于运动目标跟踪^{[20][21]},有学者提出了直方图相似度的改进方法^[22]。

在目标“量测”方面,如果运动目标不涉及旋转、形变和尺度伸缩等变化,简单的关联匹配方法既可以取得较好后果。否则,在平移、旋转、缩放等复杂运动模式下,匹配模式就指数级增长,从而使得跟踪速度急剧下降。为了提高运动目标和模板图像之间的匹配速度,研究人员提出了基于最速梯度下降优化思想的 Mean-Shift 算法^{[2]—[6]}。给定跟踪目标合理表达,Mean-Shift 算法在许多跟踪问题中取得了良好效果。但是,Mean-Shift 算法仍然是局部最优化算法,当目标区域出现类似目标的物体,或者产生部分遮挡的情况下,因为出现了多个极值点,算法的效果会受到很大影响。

本文受 SIFT 特征提取和匹配的启发,本文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), 融合了颜色直方图和梯地方向直方图, 采用多尺度综合直方图来表示跟踪目标模板和候选目标, 主要用来解决动态背景下运动目标的尺度变化、部分遮挡和形变等困难场景下运动目标的实时鲁棒跟踪。本文的跟踪算法相比于现有的跟踪算法, 主要贡献如下:

- 1). 提出了用于目标跟踪的综合直方图特征;
- 2). 将综合直方图特征应用在 Kalman 和 Mean-Shift 算法框架内进行目标跟踪。

本文组织结构如下: 第 1 节介绍视频中运动目标跟踪的问题和研究现状, 第 2 节详细介绍跟踪算法, 第 3 节是实验结果分析和算法对比, 第 4 节是结论和未来的工作展望。

4.2 综合颜色和梯度方向直方图

首先, 提出了运动目标跟踪的系统框架和综合直方图的目标表示方法, 综合直方图融合了梯度直方图和颜色直方图各自的优点, 是一种更为鲁棒的目标表达; 然后构建多尺度综合直方图的金字塔; 最后在基于 Kalman 滤波估计预测的框架下, 采用 Mean-Shift 算法进行搜索优化, 利用 Kalman 预测降低 Mean-Shift 搜索过程中相似物体的干扰, 提高了跟踪精度, 同时降低了计算量。

像素点梯度方向表示了其邻域的空间关系, 具有更精确的匹配效果, 并且具有一定的旋转不变性。色彩空间 HSV 中, 色度 (Hue) 是决定颜色本质的基本特性, 是由物体反射光线中占优势的波长来决定的。通过运动目标不同姿态下的颜色直方图和梯度直方图匹配比较和误差分析 (如图 4-2 所示), 发现色度直方图和色度图的梯度方向直方图具有更好的匹配精度。因此本文算法中综合直方图采用的是 16 位色度直方图和色度图 8 位的梯度方向直方图, 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。

4.2.1 梯度方向直方图

梯度方向直方图, 先计算目标图像的色度图像, 然后在色度图内计算所有像素点 $I(h, w)$ 的梯度 $d(h, w)$ 和方向 $ori(h, w)$, 梯度方向包含了领域内的空间位置信息 (如图 4-1b 所示), 其梯度和方向计算公式为:

$$\begin{aligned}
 dy &= I(h+1, w) - I(h-1, w) \\
 dx &= I(h, w+1) - I(h, w-1) \\
 d(h, w) &= \sqrt{(dx)^2 + (dy)^2} \\
 ori(h, w) &= a \tan 2(dy, dx) \quad ori \in [-\pi, \pi]
 \end{aligned}
 \tag{4-1}$$

其中， dy 表示像素点 $I(h, w)$ 垂直方向相邻点色度值的差， dx 表示像素点 $I(h, w)$ 水

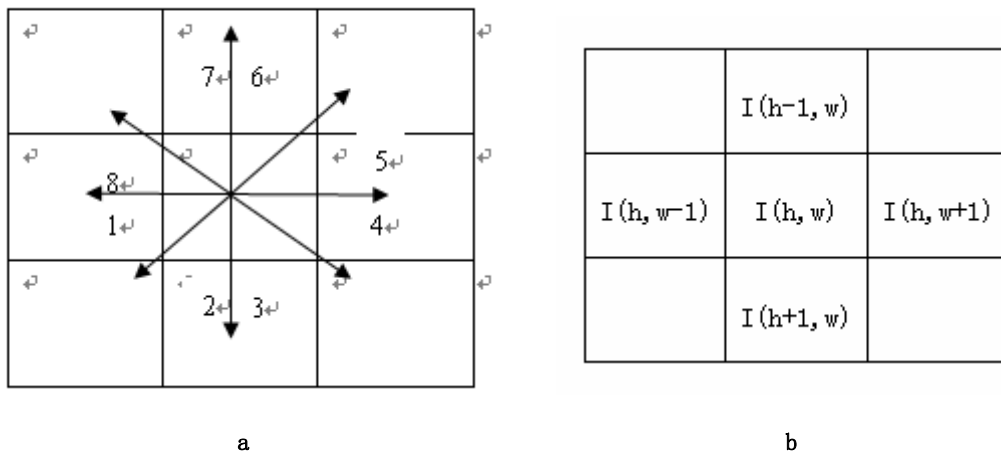


Fig. 1 a. Eight sub regions of gradient's orientation
 b. Calculation of oriented grads in hue image

图 4-1 梯度方向空间的 8 个区域和色度梯度方向的计算

a 梯度方向空间的 8 个区域 b 像素点 $I(h, w)$ 的空间信息

平方向相邻点色度值的差， $ori(h, w)$ 的大小范围为 $[-\pi, \pi]$ 。

定义 8 位的梯度方向直方图，以 0.25π 为单位将梯度方向空间均匀地分成 8 个区域（如图 4-1a 所示），以像素点的方向作为分量和以梯度作为量化值，构建运动目标的 8 位的梯度方向直方图；为了降低目标区域图像边界噪声的影响，采用窗宽为 h_1 的剖面函数 $k(\square): [0, \infty) \rightarrow R$ 进行平滑后的梯度直方图为：

$$q_b(y_0) = C_1 \sum_{i=1}^n k\left(\left\|\frac{y_0 - X_i}{h_1}\right\|^2\right) d(X_i) \delta[o(X_i) - b_j], \quad \text{if } m \leq j \leq M \tag{4-2}$$

其中， y_0 是目标模板的中心位置； n 是目标模板像素的数目； $q_b(y_0)$ 是平滑后直方图分量 b 的值； $\delta(\cdot)$ 是 Kronecker Delta 函数； m 是色度直方图级的个数 ($m=16$)， M 是综合直方图级的个数 ($M=24$)； n 是目标区域像素点的数目， X_i

是第 i 个像素点的位置, $X_i = [h_i, w_i]$, $d(X_i)$ 是像素 X_i 处的梯度, $o(X_i)$ 是梯度方向空间 8 区域量化后的像素 X_i 处的梯度方向; 常数 C_1 由约束条件 $\sum_{b_j=1}^M q_{b_j}(y_0) = 1$ 得到, 即

$$C_1 = \frac{1}{\sum_{i=1}^n k\left(\left\|\frac{y_0 - X_i}{h_1}\right\|^2\right)(1 + d(X_i))} \quad (4-3)$$

4.2.2 色度直方图

定义 16 位的色度直方图, 以像素点的色度值作为分量, 构建运动目标的 16 为色度直方图, 同理采用窗宽为 h_1 的剖面函数 $k(\square): [0, \infty) \rightarrow R$ 进行平滑后的色度直

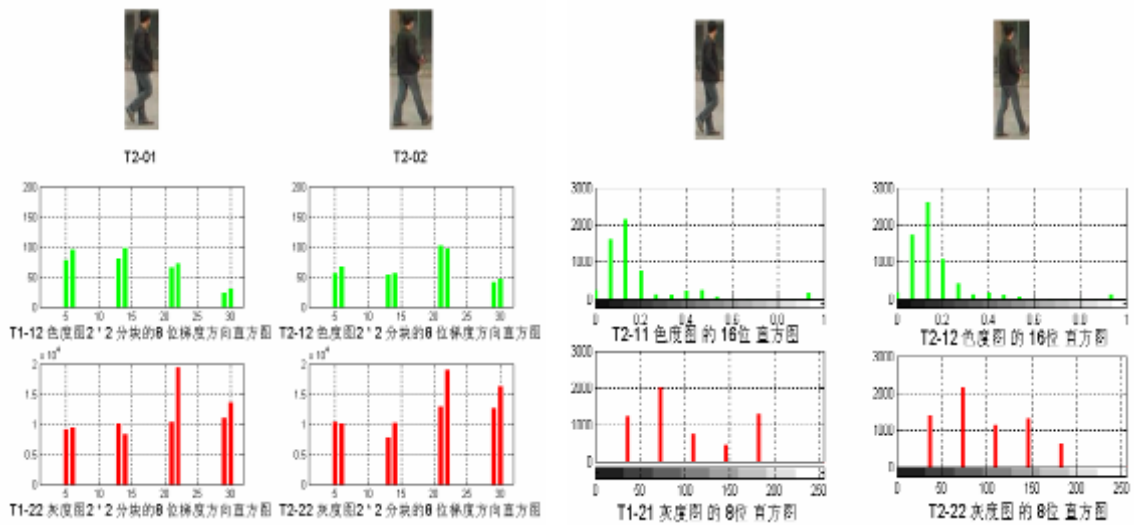


图 4-2 目标图像的梯度方向直方图和颜色直方图

a 色度图和灰度图 2*2 分块的 8 位梯度方向直方图

b 色度图的 16 位直方图和灰度图 8 位直方图

方图为:

$$q_b(y_0) = C_1 \sum_{i=1}^n k\left(\left\|\frac{y_0 - X_i}{h_1}\right\|^2\right) \delta[I(X_i) - b_j], \quad \text{if } 1 \leq j \leq m \quad (4-4)$$

其中, $I(X_i)$ 是 16 位色度级量化后的色度图 X_i 处像素的色度值。

图像金字塔是图像处理和多尺度分析中一个重要的数据结构, 小尺度反映图像的细节信息, 大尺度反映图像的概要信息。多尺度综合直方图采用 s 尺度的图像金字塔的数据结构, 在每一个尺度上计算图像的综合直方图, 多尺度图像的计算考虑到计算量采用简单的均值滤波, 也可以扩展到其他的平滑滤波的算法。基于多尺度的综合直方图跟踪算法, 通过实验分析发现近似达到了与 SIFT 跟踪的跟踪精度, 跟踪算法的比较和误差分析在第 4.5.4 部分有详细的论述(如图 4-16 和图 4-17 所示); 并且由于直方图计算简单, 基本上可以满足实时处理的要求。

4.3 Mean-Shift 算法

Mean-Shift 算法是一种计算局部最优的搜索算法 [1] [2] [6], 通过计算候选目标与目标模板之间相似度的概率密度分布, 然后利用概率密度梯度下降的方向来获取匹配搜索的最佳路径, 加速运动目标的定位和降低搜索的时间。本文在 Kalman 滤波预测估计框架下利用 Mean-Shift 算法优化跟踪算法, 提高算法的实时性和跟踪精度。

对于中心位置为 y 的候选目标区域图像的综合直方图, 经过窗宽为 h_1 的剖面函数 $k(\square): [0, \infty) \rightarrow \mathbf{R}$ 进行平滑后的模型 $\{p_b(y)\}$, $j=1, 2, \dots, m, m+1, \dots, M$ 为:

$$p_b(y) = C_2 \sum_{i=1}^N k\left(\left\|\frac{y - X_i}{h_2}\right\|^2\right) \delta[I(X_i) - b], \quad b=1, 2, \dots, m \quad (4-5)$$

候选目标区域与目标模板之间的相似性度量采用 Bhattacharyya 系数, 其定义为:

$$\rho(y) \equiv \rho[p(y), q(y_0)] = \sum_{b=1}^m \sqrt{p_b(y)q_b(y_0)} \quad (4-6)$$

其几何意义为两个 m 维向量 $(\sqrt{p_1}, \dots, \sqrt{p_m})$ 和 $(\sqrt{q_1}, \dots, \sqrt{q_m})$ 夹角的余弦。

由于目标运动的连续性, Mean-Shift 算法在当前帧图像中搜索的位置是从 Kalman 的预测位置 y_0 开始的。对公式 (4-6) 在 y_0 处泰勒展开并省略高次项得到:

$$\rho[p(y), q(y_0)] \approx \frac{1}{2} \sum_{b=1}^m \sqrt{p_b(y_0)q_b(y_0)} + \frac{C_2}{2} \sum_{i=1}^d \omega_i k\left(\left\|\frac{y - X_i}{h_2}\right\|^2\right) \quad (4-7)$$

$$\text{其中, } \omega_i = \sum_{b=1}^m \sqrt{\frac{q_b(y_0)}{p_b(y_0)}} \delta[I(X_i) - b]$$

为了最佳候选目标的中心位置 y_1 , 公式(4-7)中 $\rho(y)$ 对 y 求偏导, 并且 $\frac{\partial \rho}{\partial y} = 0$

时的 y_1 为:

$$y_1 = \frac{\sum_{i=1}^{N_h} w_i X_i}{\sum_{i=1}^{N_h} w_i} \quad (4-8)$$

4.4 基于多尺度综合直方图的运动目标跟踪算法

本文提出的基于多尺度综合直方图的目标跟踪算法, 是在 Kalman 滤波的预测估计框架下, 采用多尺度综合直方图的目标表示, 然后通过 Mean-Shift 算法进行搜索优化, 较之于传统的直方图跟踪算法提高了跟踪的精度, 较之于 SIFT 跟踪算法达到了近似的跟踪精度和满足了实时处理的要求。Kalman 滤波对目标的历史位置信息进行滤波, 得到目标的运动参数估计, 减小了候选目标匹配的搜索范围, 同时也降低了综合直方图相似物体的干扰。

第 k 帧 Mean-Shift 算法描述如下:

- 1). 由 Kalman 滤波计算运动目标第 k 帧的预测位置 X_k , 为 Mean-Shift 算法提供了搜索的初始化位置 $y_0 = X_k$;
- 2). 计算候选目标区域的多尺度综合直方图和剖面函数平滑后的模型 $\{p_b(y_0)\} \quad b=1, 2, \dots, n$, 计算剖面函数平滑后的多尺度综合直方图尺度 s 上候选目标与目标模板的巴氏距离 $\rho_s[p(y_0), q]$;
- 3). 计算权重 $\{\omega_i\} \quad i=1, 2, \dots, n$ 和下一个新的位置 y_1 ;
- 4). 重新计算目标模型 $\{p_b(y_1)\} \quad b=1, 2, \dots, n$; 以及其与目标模板的距离 $\rho_s[p(y_1), q]$;
- 5). While $\rho_s[p(y_1), q] < \rho_s[p(y_0), q]$ do

$$y_1 = \frac{1}{2}(y_0 + y_1)$$

再次计算 $\rho_s[p(y_1), q]$;

- 6). 如果 $\|y_1 - y_0\| < \varepsilon$, 迭代停止; 否则 $y_0 = y_1$, 返回到第 2) 步;
- 7). 当 $\rho_s[p(y_1), q]$ 取最大值时的 y_1 作为最佳候选目标区域的中心位置, 并且把 y_1 作为量测值代入到 Kalman 滤波, 返回到第 1) 步; 直至跟踪视频结束。

4.5 跟踪算法比较和性能分析

为了验证本文跟踪算法尺度变化、部分及完全遮挡和动态背景等复杂场景下的性能, 分别设计了针对尺度变化、部分和完全遮挡、静态和变化场景等情况下的非刚体目标的跟踪实验 4.5.1 (如图 4-3、图 4-4 和图 4-5 所示)、实验 4.5.2 (如图 4-6 至图 4-10 所示) 和实验 4.5.3 (如所示), 跟踪算法是用 Visual Studio 6.0 和 OpenCV 软件包开发的, 所有实验在 CPU 3.0 GHz、1 G 内存和独立显卡的 PC 机上进行的, 所有实验采用 SDL 实验室拍摄的分辨率为 640×480 , 帧率为 30 fps 的 AVI 视频。最后实验 4.5.4 是对上述实验 4.5.2 进行算法比较和跟踪误差分析; 实验的结果表明本文的跟踪算法比传统直方图跟踪算法有更稳定的效果, 可以达到与 SIFT 跟踪近似的跟踪精度, 并且基本上可以满足跟踪系统实时处理的要求。

4.5.1 运动目标尺度伸缩变化情况下的跟踪实验

该实验采用具有显著尺度伸缩变化的行人作为跟踪目标, 图 4-3 至图 4-5 所示的是该测试序列第 50—417 帧大约每 80 帧的抓拍的跟踪效果图, 在第 50 帧时选择感兴趣的运动目标, 跟踪算法一直锁定该运动目标, 此过程运动目标的尺度发生了严重的缩放, 到第 417 帧时最佳候选目标的大小仅为第 50 帧时的一半; 并且运动目标在第 263 帧和第 347 帧时还发生了部分遮挡。该实验表明, 基于多尺度综合直方图的跟踪算法有效地解决运动目标的尺度缩放变化。



Frame 50

Frame 130

图 4-3 尺度伸缩变化情况下的跟踪实验大尺度情况下的效果



Frame 180

Frame 263

图 4-4 尺度伸缩变化情况下的跟踪实验中尺度情况下的效果



Frame 347

Frame 417

图 4-5 尺度伸缩变化情况下的跟踪实验小尺度情况下的效果

4.5.2 运动目标遮挡情况下的跟踪实验



Frame 88

Frame 94

图 4-6 运动目标完全遮挡前的跟踪效果



Frame 98

Frame 100

图 4-7 运动目标完全遮挡时的跟踪效果



Frame 102

Frame 110

图 4-8 运动目标完全遮挡后的跟踪效果

错误!未找到引用源。和错误!未找到引用源。所示的是同一个测试序列的两节，

运动目标在跟踪的过程中均发生了部分甚至是完全遮挡，其中，第 64 帧和第 121 帧是运动目标发生遮挡前的状态，第 66 帧和第 123 帧是几乎完全遮挡时的运动目标状态，第 67 帧和第 122 帧是部分遮挡时运动目标的状态，第 68 帧和第 124 帧是遮挡后的仍然保持鲁棒跟踪的结果图像。该实验表明，基于 Kalman 滤波预测的跟踪算法在运动目标发生遮挡时，有效地预测了运动目标的状态，得到了鲁棒的跟踪效果。

4.5.3 运动目标形变情况下的跟踪实验



Frame 271

Frame 286

图 4-9 光线亮度变暗和运动目标部分遮挡前的跟踪效果



Frame 290

Frame 294

图 4-10 光线亮度变暗和运动目标部分遮挡时的跟踪效果



Frame 60



Frame 80

图 4-11 静止背景下非刚体目标严重形变的跟踪效果



Frame 100



Frame 120

图 4-12 静止背景下非刚体目标严重形变的跟踪效果



Frame 140



Frame 160

图 4-13 静止背景下非刚体目标严重形变的跟踪效果

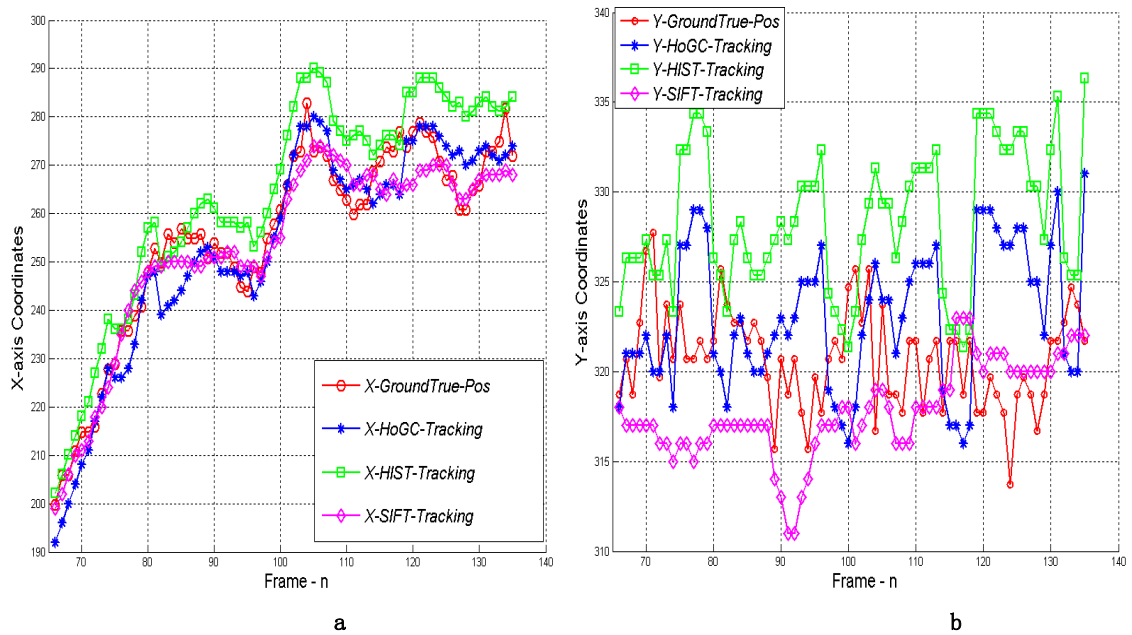


Fig. 16 Comparison of proposed HoGC Tracking Algorithm with Traditional Histogram Tracking Algorithm and SIFT Tracking
 图 4-16 HoGC 跟踪算法与传统直方图和 SIFT 跟踪算法的跟踪轨迹比较

Tracking Algorithm	AXIS	Mean Error	Max Error	Min Error
HIST	X	6.6032	20.8289	0.1711
	Y	9.6387	27.2409	0.2409
HoGC	X	5.5840	16.4326	0.4326
	Y	7.7348	24.6656	0.3344
SIFT	X	4.2262	16.5674	0.4326
	Y	3.5615	11.6656	0.3344

图 4-17 HoGC 跟踪算法与传统直方图和 SIFT 跟踪算法的误差分析
 X 表示水平方向的误差，Y 表示垂直方向的误差

4.6 本章小结

本文在传统的 Kalman 滤波和 Mean-Shift 优化框架下提出了一种新的视频运动目标跟踪算法。首先，算法融合了色度直方图和梯度方向直方图，形成了一种新的综合直方图特征，解决运动目标的旋转和形变；然后构建了运动目标图像区域

的金字塔，以实现多尺度的目标跟踪处理；最后算法采用 Kalman 滤波预测耦合 Mean-Shift 算法，在尺度、位移空间内进行优化匹配搜索，确定最佳候选目标区域的位置信息。大量实验结果表明，本文采用的综合直方图特征能够很好的适应伴随尺度伸缩、旋转和形变的运动目标跟踪，可以解决传统直方图跟踪算法的稳定性差和精度不高的问题，滤波和匹配优化框架下的跟踪算法具有比传统算法更高的精度和速度。

本文提出了滤波预测框架下基于多尺度综合直方图的跟踪算法，采用色度和梯度方向综合直方图表示运动目标，利用经典的 Mean-Shift 算法优化搜索过程。实验结果表明，本文能够算法能够实现对尺度伸缩、形变和遮挡情况下的运动目标进行跟踪，跟踪效果的精度和稳定性明显高于基于传统直方图的跟踪算法，可以近似达到基于 SIFT 跟踪算法的精度，同时可以解决基于 SIFT 特征跟踪算法未能达到的实时处理要求。

运动目标在跟踪的过程中伴随着尺度、形状等变化，长时间跟踪鲁棒性的跟踪还需要适时的更新目标模板。因此，更新策略是下一步的研究重点。

总结与展望

1. 总结

本文简要介绍了几种基础的空间谱估计算法原理，分析比较了线形阵列和圆形阵列的综合方向图，并对线阵和圆阵的特点作了详细的分析。在线性阵列基础上，分析了不同的空间谱估计算法的应用范围和性能，并仿真分析了这些算法在线阵和圆阵上的应用。

本文对线性阵列、圆形阵列天线方向图特性进行了比较深入和系统的研究。研究了空间谱估计中的幂迭代 (MPI) 算法、多重信号分类 (MUSIC) 算法。MPI 算法中针对圆形阵列所固有的缺陷，并综合线阵的优点，利用模式转换，将圆形阵列转换为虚拟的线阵，虚拟线阵同时综合了圆阵和线阵的优点，可以扫描 360 度的范围，副瓣比主瓣的相对高度低 13dB。并在模式转换的基础上应用 MPI 算法通过分析比较可以得出虽然阵列总的输出功率降低^[5]，但由于形成了虚拟线阵，将均匀圆阵的非线性阵列流型线性化，降低了在非信号方向的能量输出，使得阵列在信号方向的能量输出明显增加。提高了阵列对信号的分辨能力。在 MUSIC 算法中，分别讨论了经典 MUSIC 算法、求根 MUSIC 算法、基于波束空间形成的 MUSIC 算法、基于解相干的 MUSIC 算法，并将各种算法用于线性阵列，做出了仿真图形进行分析比较。通过分析可以得出 MUSIC 算法在特定的条件下具有很高的分辨力、估计精度及稳定性。

2. 展望

本文虽然对幂迭代和多重信号算法研究做了大量的工作，并取得了一些认识，但空间谱估计的理论与技术已日益成熟，需要（或值得）研究的方向还有很多，由于本人的能力和时间有限，有些问题研究不够深刻，还有许多问题尚未触及。下面结合自适应数字波束形成的现状与我们的工作体会对下一步研究走向做一个介绍：

融合先进的滤波预测算法、匹配算法以及数字图像处理和计算机视觉领域的先进技术，视频图像中运动目标的检测和跟踪进一步发展可以解决下列问题：

- 1) 在复杂的场景中(如光线变化，混乱的背景)，目标的高精度和鲁棒跟踪。
- 2) 尺度变化、目标与背景以及目标自身的遮挡难题、仿射变化等等视频跟踪

中的开放性难题可以得到很好的解决。

- 3) 精确跟踪上的后续处理, 比如: 运动的分析和合成、虚拟现实、基于运动信息的身份识别等等。
 - 4) 民用方面的安全监控、交通管制; 军用方面的导弹制导、巡航导航等等领域有着更为广泛的应用。
 - 5) 先进的计算机视觉等自动化技术为人民更安全、智能、舒适的生活提供技术支持和保障。
-

致 谢

在攻读硕士学位的三年学习生活中，我经历了坎坷，付出了努力，也得到了很大的收获。在毕业论文完成之际，由衷的感谢这三年来曾经给予我帮助的老师、同学、朋友和家人。

首先，感谢我的导师刘先省教授和焦建彬教授，感谢他将我领入这个领域，以无比的耐心在三年来对我不断的点拨和培养。刘老师学识渊博，作风严谨，不仅在学习上给予了我很大的帮助，在为人处事和对生活的态度上也给了我很多的建议。特别是他对科学孜孜以求，精益求精的精神，和对工作勤勤恳恳，严谨认真的态度将使我终生受益。感谢他在繁忙中对我的悉心教诲和真切关怀，并在我的硕士论文完成的过程中给予了很多的指导。为我提供了极大帮助。他的人格魅力更是我敬佩

感谢河南大学计算机与信息工程学院和中科院研究生院工程教育学院，使我有机会到中国科学院读书。在中科院研究生院学习的这两年中，我学到了很多相对来说较为领先的知识和技术，为我日常的学习工作和硕士论文的写作打下了良好的基础。

在这三年里，我庆幸生活在一个团结友爱的集体中，感谢 SDL 实验室的叶齐祥博士，他渊博的专业知识和精益求精的动手能力，在理论学习和项目开发过程中给我提供了耐心的引导和帮助；感谢数据融合实验室的胡振涛博士、赵淑红、赵要军、车颖涛等，我们共同讨论、共同学习，他们为我的论文写作提出了很好的建议。

我要感谢支持我的父母和弟弟、妹妹，他们无私的爱和无条件的支持一直指引我奋发拼搏、积极向上，他们对我的爱我将终生铭记。

最后，感谢我的爱人，他在我求学的过程中一直支持我、鼓励我，是我能够毫无后顾之忧的完成自己的学业。

感谢本论文参考文献中的所有作者，没有他们的工作做铺垫，我的工作是无法开展的！

最后，感谢参加论文评审和答辩的各位老师对本文的认真评阅。
常向魁

二零零七年六月

参考文献

- [1]. Jianyu Wang. Sequential Monte Carlo Filtering Based Visual Tracking. Dissertation for the Doctoral Degree in Engineering. Harbin Institute of Technology, February, 2006.
 - [2]. D. Comaniciu, V. Ramesh. Real-time Tracking of Non-rigid Objects Using Mean Shift. Proceedings of IEEE Conf. on on Computer Vision and Pattern Recognition (CVPR2000), 2000.
 - [3]. D.Comaniciu, V.Ramesh, P.Meer. The Variable Bandwidth Mean-Shift and Data-Driven Scale Selection. ICCV'01. VOI: 438-445. 2001.
 - [4]. D.Comaniciu, P.Meer. Mean Shift Analysis and Applications. ICCV'99. 1997.
 - [5]. US patent Pending. Real Time Tracking of Non-Rigid Objects using Mean Shift.
 - [6]. V. P. Namboodiri, A. ghorawat, S. Chaudhuri. Improved Kernel-Based Object Tracking Under Occluded Scenarios. In the Int. Conf. on ICVGIP, pp. 504-515, 2006.
 - [7]. J. Lee, S. Kim, D. Kim, J. Shin, J. Paik. Feature Fusion-Based Multiple People Tracking. In Int. Conf. on PCM, pp. 843-853, 2005.
 - [8]. N. Saunier, T. Sayed. A feature-based tracking algorithm for vehicles in intersections. In Proceedings of the 3rd Canadian Conf. on Computer and Robot Vision (CRV'06), 2006.
 - [9]. J. Shi, C. Tomasi. Good Features to Track. In Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 593-600, 1994.
 - [10].C. Tomasi, T. Kanade. Detection and Tracking of Points Features. Technical Report CMU-CS-91-132, Carnegie Mellon University, Pittsburg, PA, April 1991.
 - [11].C. Harris, M. Stephens. A Combined Corner and Edge Detector. Proc. of the Fourth Alvey Vision Conference, Manchester, pp. 147-151. 1998.
 - [12].W. Du, J. Piater. Tracking by Cluster Analysis of Feature Points and Multiple Particle Filter. Int. Conf. on Advanced Video and Signal based Surveillance, pp. 165-170, 2005.
 - [13].D. Lowe. Object recognition from local scale-invariant features. In Int. Conf. on Computer Vision, pp. 1150-1157, 1999.
 - [14].D. Lowe. Distinctive image features from scale-invariant keypoints. In Journal of Computer Vision, 60(2):91-110, 2004.
-

- [15].A.E. Abdel-Hakim, A.A Farag. CSIFT: A SIFT Descriptor with Color Invariant Characteristics. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR2006), pp. 1978-1983, 2006.
- [16].Adam Stanski, Olaf Hellwich. Spiders as Robust Point Descriptors. In Int. Conf. on DAGM, pp. 262-268, 2005.
- [17].Qiang Zhu, Shai Avidan, Mei-Chen Yeh, and Kwang-Ting Cheng. Human Detection Using a Cascade of Histograms of Oriented Gradients. Int. Conf. on Computer Vision and Pattern Recognition (CVPR'06), pp. 1491-1498, 2006.
- [18].Amit Adam, Ehud Riviliin, Ilan Shimshoni. Robust Fragments-based Tracking using the Integral Histogram. Int. Conf. on Computer Vision and Pattern Recognition (CVPR'06), pp. 798-805, 2006.
- [19].Feng Zhao, Qingming Huang, Wen Gao. Image Matching by Multiscale Oriented Corner Correlation. In Int. Conf. on ACCV, pp. 928-937, 2006.
- [20]. Navnnet Dalal, Bill Triggs. Histograms of Oriented Gradients for Human Detection. Int. Conf. on Computer Vision and Pattern Recognition (CVPR'05), pp. 886-893, 2005.
- [21].Stanley T.Birchfield, Sriram Rangarajan. Spatiograms versus Histograms for Region-Based Tracking. Int. Conf. on Computer Vision and Pattern Recognition (CVPR'05), pp. 1157-1162, 2005.
- [22].Haibin Ling, Kazunori Okada. Diffusion Distance of Histogram Comparison. Int. Conf. on Computer Vision and Pattern Recognition (CVRP'06), pp. 246-253, 2006.
- [23].Guangyu Zhu, Dawei Liang, Yang Liu, Qingming Huang, Wen Gao. Improving Particle Filter with Support Vector Regression for Efficient Visual Tracking. In the Int. Conf. on Image Processing (ICIP2005), pp. 422-425, 2005.
- [24].Yun Lei, Xiaoqing Ding, Shengjin Wang. AdaBoost Tracker Embedded in Adaptive Particle Filtering. 2006.
- [25].K.Smith, D.G.Perez, J.M.Odobez. Using Particles to Track Varying Numbers of Interacting People. CVPR'05. VOI: 962-969. 2005.
- [26].Ganesh Sumdaramoorthi, Jeremy D.Jackson, Anthony Yezzi and Andrea C.Mennucci.Tracking with Sobolev Active Contours.Int. Conf. On Computer Vision and Pattern Recognition (CVPR'06), pp.674-680, 2006.
-

-
- [27].Hwasup Lim, Vlad I.Morariu, Octavia I.Camps, and Mario Sznaiar. Dynamic Appearance Modeling for Human Tracking. Int. Conf. On Computer Vision and Pattern Recognition (CVPR'06), pp.751-757, 2006.
- [28].I.Matthews, T. Ishikawa, S.Baker. The Template Update Problem. IEEE Transactions on PAMI,pp. 810-815, 2004.
- [29]. Arthur E.C. Pece, the Kalman-EM Contour Tracker. <http://www.diku.dk/users/aecp>. 2003.
- [30].Zhiwei Zhao, Xueqin Li, R.W. Gunderson. A novel fuzzy entropy clustering algorithm. In Proceedings of the 3rd IEEE Conf. on Fuzzy Systems, pp. 636-641, 1994.
- [31].Yizheng Cai, Nando de Freitas, and James J. Little. Robust Visual Tracking for Multiple Targets. ECCV. 2006.
- [32].Junxian Wang, George Bebis, Ronald Miller. Robust Video-Based Surveillance by Integrating Target Detection with Tracking. CVPR. 2006.
- [33].David Ross, Jongwoo Lim and Ming-Hsuan Yang. Adaptive Probabilistic Visual Tracking with Incremental Subspace Update. ECCV. pp. 470--482, 2004.
- [34].N. Pérez de la Blanca, J.M. Fuertes, M. Lucena. Deformable Object Matching Based on Multi-scale Local Histograms. AMDO 2004, pp.154-162, 2004.
- [35].Tony Lindeberg. Feature Detection with Automatic Scale Selection. Int. J. of Computer Version.1998.
- [36].Bastian Leibe, Bernt Schiele. Scale-Invariant Object Categorization using a Scale-Adaptive Mean-Shift Search. DAGM'04 Pattern Recongition Symposium. 2004.
- [37].B.Leibe, B.Schiele. Scale-Invariant Object Categorization using a Scale-Adaptive Mean-Shift Search. DAGM'04, 2004.
- [38].N.P.Blanca, J.M.Fuertes, M.Lucena. Deformable Object Matching Based on Multi-Scale Local Histograms. AMDO 2004. pp. 154-162. 2004.
- [39].Krystian Mikolajczyk, Cordelia Schmid. Scale & Affine Invariant Interest Point Detectors. ICCV. 2004.
- [40].Bo Wu, Ram Nevatia. Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection. CVPR. 2006.
- [41].Serge Belongie, Jitendra Malik, Jan Puzicha. Shape Matching and Object Recongition Using Shape Contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002
-

- [42]. Krystian Mikolajczyk, Cordelia Schmid. A performance Evaluation of Local Descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [43]. Krystian Mikolajczyk, Cordelia Schmid. An Affine Invariant Interest Point Detector. ECCV. pp.128-142. 2002
- [44]. Daniel Freedman. Effective Tracking through Tree-Search. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003
- [45]. Changjiang Yang, Ramani Duraiswami, Larry Davis. Fast Multiple Object Tracking via a Hierarchical Particle Filter. ICCV.2005.
- [46]. Tao Zhang, Daniel Freedman. Tracking Objects Using Density Matching and Shape Priors. ICCV. 2003.
- [47]. Ying-Tung Hsiao, Cheng-Long Chuang, Yen-Ling Lu, Joe-Air Jiang. Robust multiple objects tracking using image segmentation and trajectory estimation scheme in video frames. Image and Vision Computing. 24(2006) 1123-1136. 2006.
- [48]. Lius M. Fuentes, Sergio A. Velastin. People tracking in surveillance applications. Image and Vision Computing. 24(2006) 1165-1171.
- [49]. B.W, R.Nevatia. Detection of Multiple Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. ICCV'05. VOI I: 90-97. 2005.
- [50]. T.Zhao, R.Nevatia. Tracking Multiple Humans in Crowded Environment. CVPR'04. VOI II: 406-413. 2004.
- [51]. P.Viola, M.Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. CVPR'01. VOI: 511-518. 2001.
- [52]. P.Viola, M.Jones, D.Snow. Detecting pedestrians using patterns of motion and appearance. ICCV'03. Pp.734-741. 2003.
- [53]. L.Davis, V.Philomin, R.Duraiswami. Tracking Humans from a Moving Platform. ICPR'00. VOI: 171-178. 2000.
- [54]. S.S.Intille, J.W.Davis, A.F.Bobick. Real-Time Closed-World Tracking. CVPR'97. Pp.697-703. 1997.
- [55]. S.Huwer, H.Niemann. Adaptive Change Detection for Real-Time Surveillance Applications. Proceedings of the IEEE WorkShop on Visual Surveillance. Pp.37-43. 2000.
- [56]. R.Rosales, S.Claroff. Improved Tracking of Multiple Humans with Trajectory Prediction
-

and Occlusion Modeling. CVPR'98. 1998.

[57]. J. Davis, V. Sharma. Robust background subtraction for Person Detection in Thermal Imagery. CVPR'04. 2004.

[58]. J. Davis, M. Keck. A two stage Template Approach to person Detection in Thermal Imagery. CVPR'05. 2005.

[59]. D. Hall, J. Nascimento, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet. Comparison of Target Detection Algorithms using Adaptive Background Models. 2005.

[60]. 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.

[61]. 薛雪, 刘泽平, 丁艳. 区域增长的图像跟踪算法的研究. 光学技术. 2005.

[62]. 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.

[63]. 王建宇. 基于序列蒙特卡洛滤波算法的视觉目标跟踪. 博士学位论文. 2006.

[64]. <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>.

攻读硕士学位期间完成的论文及参加的科研项目

完成的论文：

[1]. 常向魁, 焦建彬, 刘先省, 叶齐祥. 基于 SIFT 特征匹配和 Kalman 滤波的运动目标跟踪算法. 中国图象图形学报. (二审中).

[2]. 常向魁, 叶齐祥, 刘先省, 焦建彬, 韩振军. 基于综合色度和梯度方向直方图的运动目标跟踪算法.

参加的科研项目：

[1]. 国家自然科学基金项目, “复杂环境下新型多目标动态监测、识别及跟踪算法研究” (60672147)。

[2]. 中国科学院百人计划资助项目, “复杂环境下动态目标跟踪技术”。

[3]. 2005 年, 河南省杰出科研人才资助项目。
