

分类号\_\_\_\_\_

密级\_\_\_\_\_

UDC \_\_\_\_\_

编号\_\_\_\_\_

# 中国科学院研究生院

## 博士学位论文

视觉目标自适应跟踪算法研究

韩振军

指导教师 焦建彬 教授

中国科学院研究生院

申请学位级别 工学博士 学科专业名称 计算机科学与技术

论文提交日期 2011年10月 论文答辩日期 2011年10月

培养单位 中国科学院研究生院

学位授予单位 中国科学院研究生院

答辩委员会主席 \_\_\_\_\_

# 独创性声明

本人郑重声明：所提交的学位论文，是本人在指导教师的指导下，独立进行研究工作所取得的研究成果。尽我所知，文中除特别标注和致谢的地方外，学位论文中不包含其他人或集体已经发表或撰写过的研究成果，也不包含为获得中国科学院研究生院或其它教育机构的学位或证书所使用过的材料。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。

签名：\_\_\_\_\_日期：\_\_\_\_\_

## 关于学位论文使用授权的说明

本人完全了解中国科学院研究生院有关保管、使用学位论文的规定，其中包括：①学校有权保管、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤学校可以公布学位论文的全部或部分内容。

（涉密的学位论文在解密后应遵守此规定）

签名：\_\_\_\_\_导师签名：\_\_\_\_\_日期：\_\_\_\_\_

## 摘要

视觉跟踪是一种针对摄像机所获取的图像序列进行目标连续定位的技术。它融合了计算机视觉、图像处理、模式识别、人工智能以及自动控制等众多领域中的先进技术，通过对摄像头捕获的图像序列进行分析，计算出目标在每一帧图像中的坐标位置，将图像序列中的相关运动目标关联起来，最终获得目标的运动参数与运动轨迹。

与传统雷达跟踪系统相比，视觉跟踪主要采用比较廉价的设备，具有更高的性价比。它采用被动式工作模式，工作时不向外辐射无线电波，不易被电子侦察设备发现，具有一定的隐蔽性和抗电子干扰能力。同时在视觉目标跟踪系统中，人们能够直接从视频监视器上看到目标图像，具有更好的直观性。

尽管视觉目标跟踪技术获得了长足的发展，其可靠性与稳定性并不是十分令人满意。成像过程中三维信息的丢失，目标本身的运动及形变等因素，会造成图像序列中目标的尺度、角度、形态变化与遮挡问题，给视觉目标跟踪算法带来了巨大的挑战，造成了跟踪算法的不稳定与不可靠。针对上述问题，本文提出了视觉目标自适应跟踪算法，旨在解决背景变化、跟踪目标形变以及局部遮挡等问题。本文的主要研究内容和贡献如下：

1、提出了一种面向视觉目标跟踪的综合特征表示。基于颜色直方图（Histogram of Color, HC）和梯度方向直方图（Histogram of Oriented Gradient, HOG）的综合直方图 HOGC（Histogram of Oriented Gradient and Color）融合了颜色、局部轮廓以及梯度方向等特征。基于 SIFT（Scale Invariant Feature Transform）特征点的主方向方法，进一步对综合特征进行了面向视觉目标跟踪的改进，在一定程度上解决了 HOG 对跟踪目标旋转敏感的问题。

2、基于综合特征表示，提出了基于自适应综合特征表示的视觉目标跟踪算法。基于前景（跟踪目标）/背景的特征信息以及相关变化，将目标综合特征的评估嵌入到传统的滤波框架中，保证特征权重的变化是连续的过程，即具有时间连续性。基于滤波框架的特征评估有效地将滤波算法从对目标运动状态的建模扩展到了对目标特征权重的评估，拓展了传统滤波框架的应用范畴，降低了跟踪前景/背景的变化对跟踪效果的影响。

3、基于自适应综合特征表示，提出了基于稀疏自适应综合特征表示的视觉

目标跟踪算法。在跟踪的初始化过程中建立在线训练样本集合，用以对跟踪特征的实时稀疏选择，从而得到跟踪目标具有前景/背景区分性的稀疏特征表示。同时，基于滤波框架对目标的稀疏特征表示进行有效地自适应的评估，在保证特征描述具有区分性的基础上，进一步保证了特征的环境自适应性。

4、基于稀疏自适应特征表示，提出了一种基于分层自适应稀疏表示与稀疏重构的目标跟踪新框架。在跟踪的过程中，基于跟踪目标的稀疏自适应特征表示以及基于搜索区域建立的实时重构样本集合，获得目标在样本空间的自适应重构系数，从而统计跟踪目标在样本空间的概率密度分布，实现对目标的可靠跟踪。实验表明，所提出的跟踪框架既能较好的适应前景/背景的变化，又能有效的处理跟踪目标的局部遮挡等问题。

**关键词：**视觉目标跟踪、特征评估、特征选择、滤波、稀疏表示、稀疏重构

## **Adaptive Visual Object Tracking**

**Han Zhen-jun** (Computer Application)

Directed by: **Jiao Jian-bin** (Professor)

Moving objects tracking is one of the most important branches in the computer vision, which combines advanced technologies and research achievements in image processing, pattern recognition, artificial intelligence, automatic control and other relative fields. It has widely applied in video surveillance, robot navigation, video transmission, video retrieval, medical image analysis, meteorological analysis, and other fields. Therefore, this subject has important theoretical significance and wide practical value.

However, due to the object appearance variation, object occlusion and dynamic tracking background, tracking failures often happens. In this paper, we propose adaptive visual object tracking methods/framework in order to solve these tracking problems. The contributions of this paper are summarized as follows.

**1) A combined feature set for adaptive object tracking** The basic idea of using the combined feature set is that an object can often be well characterized by both the color (HC) and the contour (HOG) representation or either of them. The combined feature set is the evolvement of color, edge orientation histograms and Scale Invariant Feature Transform (SIFT) descriptors. We propose an approach to reduce the orientation sensitivity of the HOG features by calculating the dominant orientation of the object.

**2) A novel feature evaluation approach of temporal consistency** In each frame, the proposed evaluation approach adjusts the confidence of each feature based on the discriminative abilities of the feature bins in current video frame and the confidence of each feature in the previous video frames. The evaluation is formulated in the traditional filter frameworks so that the discriminative abilities of current frame and the temporal consistency from previous video frames are well comprised. The proposed feature evaluation approach extends the function of traditional filter framework from modeling motion states to modeling feature evaluation problem. To our knowledge, this is the first research of such function extension.

**3) A novel adaptive sparse feature selection approach** Based on the first  $M$  tracking frames, we construct an online training sample set for sparse feature selection by calculating the L1-norm minimization. The selected sparse representation can distinguish the object and its background well; what's more, the proposed evaluation

approach in filter frame is used to evaluate the sparse representation.

**4) A new adaptive tracking framework based on hierarchical sparse representation and sparse Reconstruction** Based on the predefined searching window, we construct a reconstruction sample set. With the adaptive sparse representation of the tracked object, we calculate its adaptive instantaneous reconstruction coefficient vector. Finally, we track the object based on the sparse reconstruction vector.

**Key Words:** Visual Object Tracking, Feature Evaluation, Feature Selection, Filter, Sparse Representation, Sparse Reconstruction

## 目录

摘要.....	I
目录.....	V
图目录.....	VIII
表目录.....	X
第一章 绪论.....	1
1.1 课题来源、背景和意义.....	1
1.1.1 课题来源.....	1
1.1.2 课题背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 跟踪目标的特征表示.....	4
1.2.2 跟踪目标匹配.....	4
1.2.3 运动轨迹预测.....	5
1.3 视觉跟踪的主要分类和方法.....	5
1.3.1 视觉跟踪分类.....	5
1.3.2 视觉跟踪的主要方法.....	6
1.4 本文的研究内容.....	9
1.5 本文的组织结构.....	10
第二章 视觉目标自适应跟踪算法基础.....	12
2.1 基于前景背景联合建模的视觉目标跟踪算法简介.....	12
2.2 综合特征的提取.....	14
2.2.1 跟踪目标区域的定义.....	15
2.2.2 颜色直方图提取.....	15
2.2.3 梯度方向直方图的提取.....	16
2.3 综合特征的改进.....	17
2.4 基于综合特征视觉目标跟踪的实验结果分析与比较.....	17

2.4.1 本文目标跟踪实验环境配置.....	17
2.4.2 实验结果分析与比较.....	18
2.5 小结.....	21
第三章 基于自适应综合特征的视觉目标跟踪算法.....	22
3.1 基于滤波框架的综合特征评估.....	22
3.1.1 目标区域与背景区域定义.....	22
3.1.2 特征判别力的定义.....	22
3.1.3 特征权重的计算.....	23
3.2 基于 Kalman 滤波综合特征评估的视觉目标跟踪算法.....	24
3.2.1 Kalman 滤波的基本原理.....	24
3.2.2 Kalman 滤波算法在传统目标跟踪中的应用.....	27
3.2.3 基于 Kalman 滤波的特征评估.....	27
3.2.4 基于 Kalman 滤波特征评估的视觉目标跟踪.....	29
3.2.5 实验结果比较与算法性能分析.....	29
3.2.6 跟踪算法在随动平台系统中的应用.....	34
3.3 基于粒子滤波的综合特征评估的视觉目标跟踪算法.....	35
3.3.1 粒子滤波算法的基本原理.....	35
3.3.2 粒子滤波算法跟踪.....	36
3.3.3 基于粒子滤波的特征评估.....	37
3.3.4 基于粒子滤波特征评估的视觉目标跟踪.....	39
3.3.5 实验结果比较与算法性能分析.....	39
3.4 基于 Kalman 与粒子滤波的综合特征评估的视觉目标跟踪算法性能比较.....	44
3.5 小结.....	46
第四章 基于自适应稀疏表示的视觉目标跟踪算法.....	47
4.1 稀疏表示简单介绍.....	48
4.1.1 L1-norm 研究综述.....	48
4.1.2 稀疏表示基本原理.....	48
4.2 基于自适应稀疏表示的视觉目标跟踪.....	51
4.2.1 在线训练样本集合的构造及更新.....	51



4.2.2 稀疏特征的选择.....	51
4.2.3 自适应稀疏表示的计算.....	52
4.2.4 基于自适应稀疏特征的目标跟踪.....	53
4.3 实验结果比较与算法性能分析.....	54
4.3.1 基于自适应稀疏表示目标跟踪算法的有效性验证.....	54
4.3.2 算法性能比较.....	56
4.4 小结.....	57
第五章 基于分层自适应稀疏表示与稀疏重构的视觉目标跟踪框架.....	58
5.1 基于分层自适应稀疏特征与稀疏重构的视觉目标跟踪.....	58
5.1.1 目标、背景以及搜索区域的定义.....	59
5.1.2 重构样本集合的构造.....	60
5.1.3 基于重构样本集合的自适应稀疏重构.....	60
5.1.4 搜索区域的有效性验证.....	62
5.1.5 基于分层自适应稀疏表示与稀疏重构的目标跟踪.....	63
5.1.6 实验结果比较与算法性能分析.....	64
5.2 基于自适应稀疏重构回归分析的多目标跟踪.....	70
5.2.1 稀疏表示在模式分类领域内的典型应用.....	70
5.2.2 基于自适应稀疏重构回归分析的多目标跟踪.....	71
5.3 小结.....	75
总结与展望.....	76
参考文献.....	79
发表文章目录.....	86
致谢.....	89

## 图目录

图 1-1 本文研究内容框图.....	10
图 2-1 目标区域定义.....	14
图 2-2 HOG 特征提取示意图.....	16
图 2-3 利用 HOG 主方向方法对目标进行旋转示意图.....	17
图 2-4 基于 VIVID-redteam 测试视频的目标跟踪结果.....	19
图 2-5 基于 VIVID-egtest01 测试视频的目标跟踪结果.....	20
图 2-6 三种目标跟踪算法 DER 结果对比图.....	21
图 3-1 跟踪目标及其背景区域定义.....	22
图 3-2 基于滤波框架特征评估框图.....	24
图 3-3 基于 Kalman 滤波的特征权重调整示意图.....	29
图 3-4 基于 VIVID-redteam 测试视频的目标跟踪结果对比图.....	30
图 3-5 基于 VIVID-egtest01 测试视频的目标跟踪结果对比图.....	32
图 3-6 四种跟踪算法的 DER 比较结果图.....	33
图 3-7 基于 Kalman 滤波特征权重评估的视觉目标跟踪算法的跟踪结果图.....	34
图 3-8 基于 Kalman 滤波特征权重评估的视觉目标跟踪算法在基于 PTZ 摄像头系统的应用.....	35
图 3-9 基于粒子滤波的特征权重评估示意图.....	38
图 3-10 基于 VIVID-egtest01 测试视频的目标跟踪结果对比图.....	40
图 3-11 存在遮挡和相似目标情况下的跟踪结果对比图.....	41
图 3-12 四种跟踪算法的 DER 对比结果图.....	42
图 3-13 基于粒子滤波框架特征权重评估跟踪算法的跟踪实验结果.....	43
图 3-14 存在遮挡和相似目标的跟踪结果对比图.....	44
图 3-15 存在遮挡和相似目标的跟踪结果对比图.....	45
图 3-16 五种跟踪算法的 DER 结果比较图.....	45
图 4-1 基于自适应稀疏表示的目标跟踪算法流程图.....	47
图 4-2 HOGC 特征与稀疏 HOGC 特征对比图.....	52
图 4-3 四段测试视频的跟踪结果实例.....	56
图 4-4 三种跟踪算法的 DER 比较结果图.....	57

---

图 5-1 模板匹配与基于样本重构的目标跟踪算法对比图.....	58
图 5-2 基于分层自适应稀疏特征与稀疏重构的视觉目标跟踪算法流程图.....	59
图 5-3 跟踪目标、背景及其搜索区域定义.....	59
图 5-4 重构样本集合的构造.....	60
图 5-5 稀疏重构向量示意图.....	62
图 5-6 搜索区域的有效性验证.....	63
图 5-7 三段测试视频跟踪结果实例.....	66
图 5-8 图像退化模糊情况下的目标跟踪结果实例.....	67
图 5-9 图像退化模糊情况下四种跟踪算法 DER 对比结果图.....	67
图 5-10 局部遮挡情况下目标跟踪结果实例.....	69
图 5-11 遮挡情况下三种跟踪算法的 DER 对比结果图.....	70
图 5-12 多目标跟踪过程中的两类跟踪目标实例.....	72
图 5-13 基于 AdaSRA 的结果分析.....	73
图 5-14 三种跟踪目标所对应的重构样本集合中部分样本实例.....	74
图 5-15 基于 AdaSRA 与 SRC 的分类结果比较.....	74
图 5-16 基于稀疏重构分析 (AdaSRA) 的多目标跟踪结果实例.....	75

## 表目录

表 2-1 定量化跟踪算法比较测试视频列表.....	17
表 5-1 算法跟踪效率对比.....	70

## 第一章 绪论

### 1.1 课题来源、背景和意义

#### 1.1.1 课题来源

本课题来源于国家重点基础研究发展计划(973 计划)、国家自然科学基金以及中国科学院“百人计划”资助课题。作者在论文工作期间,参与了以下课题:

1、“基于多源数据的飞行器进近威胁目标检测跟踪及行为预测”,国家自然科学基金重点项目(课题编号:61039003),2011.01-2014.12,在研;

2、“飞行器威胁目标识别与图像鲁棒匹配理论与方法”,国家973计划子课题(课题编号:2010CB731804-2),2010.01-2014.12,在研;

3、“复杂环境下动态目标检测及跟踪技术研究”,中国科学院“百人计划”择优支持项目(项目编号:99T3009EA2),2009.01-2011.12,在研;

4、“视频中的人体目标检测和跟踪关键技术研究”,国家自然科学基金面上项目(课题编号:60872143),在研;

5、“复杂环境下新型多目标动态检测识别追踪算法的研究”,国家自然科学基金面上项目(课题编号:60672147),2007.01-2009.12,已结题;

6、“弹药库智能监控系统”,南京军区徐州某部,已结题。

#### 1.1.2 课题背景及意义

基于视觉的目标跟踪技术(以下简称为视觉跟踪或者目标跟踪)作为计算机视觉领域的核心研究课题之一,其主要目的是模仿生理视觉系统的运动感知功能,通过对摄像头捕获到的图像序列进行分析,计算出运动目标在每一帧图像中的位置;然后,根据运动目标相关的特征值,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,从而得到各个运动目标完整的运动轨迹。简单来说,目标跟踪技术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,进而为平台自动控制、视频序列分析和理解等提供运动信息和数据。

与传统雷达跟踪系统相比,视觉跟踪主要采用比较廉价的设备,具有更高的性价比。它采用被动式工作模式,工作时不向外辐射无线电波,不易被电子侦察

设备发现，具有一定的隐蔽性和抗电子干扰能力。同时在视觉目标跟踪系统中，人们能够直接从视频监视器上看到目标图像，具有更好的直观性。

经过近 50 多年的研究和发展，视觉跟踪技术已在军事制导、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广泛应用。例如，视频监控技术已在大多社区、大型公共场所及重要设施使用；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。此外，视觉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、远距离测量、零部件质量检测等许多领域也有应用。

然而，就目前来讲，一般意义上的视觉跟踪技术还未成熟。要开发出真正可靠、实用的视觉跟踪应用系统还需要开发更为鲁棒的核心算法。在运动目标通过摄像机的检测区域过程中，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三个方面：1、运动目标本身的变化（尺度、旋转、形状等）；2、运动目标被遮挡；3、跟踪环境动态变化（光照、图像退化模糊等）。这些情况的存在使得对运动目标的跟踪变得更加困难。现有的目标跟踪算法虽然能够在一定程度上完成对运动目标的跟踪，但这些算法还存在着诸多问题：1、目前的跟踪算法大多是基于某一种特征集合对目标进行描述，其主要缺陷是对目标描述不完备；2、所提取的特征描述无法更好的区分目标与跟踪背景，因此当背景与目标比较相似或者发生较大变化时，跟踪算法往往会失效；3、由于很难长时间对运动轨迹进行准确预测，因此当遮挡频繁发生时，跟踪算法同样会失效。在现有研究中，长时间复杂动态背景中的鲁棒跟踪是关键问题，也是难点问题。欲解决此问题需要解决：1、如何更好更完备的提取目标的特征集合，从而使得对目标的描述更加准确；2、如何使目标的特征描述能够更好的区分前景/背景以及自适应他们的变化，从而使跟踪算法能够具有环境自适应性；3、如何更好的定义目标的遮挡模型，从而使跟踪算法能够更好的处理局部遮挡情况下的目标跟踪。

## 1.2 国内外研究现状

国外对于视觉跟踪理论研究起步较早[1-21]。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了可观的成果。早在 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)。近几年，自适应跟踪和智能跟踪的思想被相继提出，DARPA 成立了自动目标识别工作组 ATRWG (Automatic Target Recognition Work Group) 进行目标智能识别和跟踪等相关研究。

同时，国外有很多高级别的学术会议和期刊也包含了视觉跟踪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。

随着图像处理和分析技术、计算机视觉技术以及各种图像传感器技术的飞速发展，国内一些高校和科研院所已经逐渐开展这方面的研究[22-34]。如：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、精彩片段分析的研究；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 and 智能轮椅视觉导航系统；清华大学开发的适用于自然环境中的视觉侦查系统具有运动检测、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进行了深入的研究等。

对于目标进行跟踪，可以获取目标的移动速度、移动方位甚至更高的语义特征。近年来研究人员开发了大量新的跟踪算法[1-34]，这些算法大致可以分为三个主要的研究方向：1、跟踪目标的特征表示；2、跟踪目标的鲁棒匹配；3、跟

踪目标的轨迹预测。

### 1.2.1 跟踪目标的特征表示

目标描述和特征的选择密切相关。选择适当的特征在目标跟踪中具有重要的作用。通常，好的特征应该具有唯一性，因此可以很容易地将目标从特征空间中区分出来。例如对于基于轮廓的目标表示，边缘通常被用来作为相应的特征。目标跟踪过程中常用的特征如下[1]：

1、颜色 目标的颜色主要由两个物理因素决定，一个是光源的功率谱分布，一个是目标的表面反射性质。在图像处理领域，RGB 空间是最常用的颜色空间，但是 RGB 在颜色感知上是不均匀的。 $L^*u^*v^*$ 和  $L^*a^*b^*$ 是感知均匀的颜色空间，而 HSV 是近似均匀的颜色空间，但是这些空间对噪声很敏感。

2、边缘 与颜色特征相比，边缘的一个重要性质是对光照变化不敏感。目前最为流行的边缘检测方法是 Canny 算子[35]。一篇有关边缘检测算法的参见文献[36]。

3、光流 光流定义了像素的密集位移向量场，其通过假设相邻图像帧的对应像素之间的亮度恒定不变而计算得到[37-39]。光流常被用来表示基于运动的分割与跟踪的特征。一篇有关光流算法的综述参见文献[40]。

4、纹理 纹理是局部区域亮度变化性质（如，平滑性、规则性等）的一种描述。和颜色相比，纹理需要一个预处理过程来产生纹理描述子。已有的纹理描述子有很多种，如灰度共生矩阵、Laws 纹理描述子、小波、易控的金字塔等[41-43]。与边缘特征相似，纹理特征对于光照的变化也不敏感。

在各种特征中，颜色在跟踪领域是使用得最为广泛的特征。但是，颜色对于光照的变化很敏感，因此，如果应用环境下光照变化较大，应该采用其他特征（如，边缘、纹理等）进行跟踪。同时，基于各种特征的组合特征来更加准确完备的描述目标还有待进一步的研究。

### 1.2.2 跟踪目标匹配

当给定跟踪目标以及其特征描述，搜索算法利用匹配策略在新的图像帧中快速准确的找到对应的跟踪目标。目前，Mean-Shift 算法[10, 101] 被较为广泛地使用，其利用颜色直方图来描述目标，并通过此直方图对目标可能出现的搜索区域



中的每一个像素点赋予一个概率值，从而产生出一幅概率图像，进而实现对目标进行快速定位。**Mean-Shift** 算法是一种非参数的概率密度梯度估计方法，采用了零阶、一阶及二阶矩来估计目标的尺度和方向。

### 1.2.3 运动轨迹预测

运动模型主要是利用跟踪目标的历史运动信息来校正目标当前的运动模型以及预测目标未来的运动位置。当其动态模型和观察模型是线性的且服从高斯分布时，可以采用卡尔曼滤波器（**Kalman Filter**）[20, 44]，且能够得到最优的轨迹预测结果。如果其动态模型是非线性的并且仍然服从高斯分布，可以通过 **Taylor** 展开将其进行线性化，从而得到扩展卡尔曼滤波器（**Extended Kalman Filter**）。当其动态模型和观察模型是非线性的并且非高斯分布时，这种情况在视觉目标跟踪中是普遍存在的，此时，粒子滤波（又称为序列蒙特卡洛滤波）[7, 21]为问题的解决提供了一种非常有效的手段，其基本思想是通过一个粒子（也称为样本）集合来近似后验概率，每一个粒子有一个权重表示它的重要性，然后通过动态模型对粒子集合中的每一个粒子向下一个时刻进行传播，最终采用观察模型对粒子进行权重评价。

## 1.3 视觉跟踪的主要分类和方法

### 1.3.1 视觉跟踪分类

视频目标跟踪算法依据不同的分类标准可以进行不同的分类。

1、被跟踪目标的数目 根据被跟踪目标的数目可以将跟踪算法分为单目标跟踪算法和多目标跟踪算法[45-47]两类。相比于单目标跟踪，多目标跟踪具有更多的难点，如：需要解决当前测量和已有轨迹之间的对应、目标之间的相互遮挡及目标之间的分裂与合并等。

2、被跟踪目标的类型 根据被跟踪目标的类型可以将跟踪算法分为刚性目标跟踪和非刚性目标跟踪。刚性目标跟踪的典型代表是车辆跟踪。这类目标可以通过建立简单的三维模型对其进行描述。非刚性目标跟踪的典型代表是人体的跟踪。这类目标可以通过将其分解为若干刚性目标及其之间的相互关联来进行建模。

3、摄像机的数目 根据摄像机的数目可以将跟踪算法分为单目摄像机跟踪和

多目摄像机跟踪[48-54]。由于单个摄像机的视野有限,无法覆盖场景的整个区域,所以采用多个摄像机的主要优点是扩大了摄像机的视野。另一方面,由于多个摄像机的使用使得部分深度信息的恢复成为可能,这为多目标跟踪中出现的遮挡问题的解决提供了有效的辅助。

4、摄像机是否运动 根据摄像机是否运动可以将跟踪算法分为静止背景下的跟踪算法和运动背景下的跟踪算法。在摄像机静止的情况下,可以对背景进行建模然后采用背景减除[55-58]进行运动目标的检测,进而采用数据关联技术对检测到的目标进行跟踪。在摄像机运动的情况下,如果是针对特定目标的跟踪,可以通过事先训练得到的检测器来检测特定目标,进而对其进行跟踪。

5、传感器的类型 根据传感器的类型可以将跟踪算法分为可见光图像跟踪和可见光谱以外的图像跟踪,如红外图像跟踪[59]等。相比于可见光图像,红外图像可以提供全天候的信息。这里需要指出的是目标跟踪在雷达[60],声纳[61]等领域已经具有很长时间的研究历史,计算机视觉领域中的一些目标跟踪算法正是从这些领域借鉴过来的。

6、跟踪的速度 根据跟踪速度的要求可以将跟踪算法分为实时跟踪和非实时跟踪。实时跟踪对跟踪算法的速度要求很高,主要用在需要系统做出快速反应的场合。如在视频监控中当一个人在一辆汽车附近有异常举动时,监控系统要迅速做出反映。非实时跟踪主要用在视频编辑等领域,如特技制作等。这时对于算法的实时性要求不高,反而对精度有更高的要求。

### 1.3.2 视觉跟踪的主要方法

视觉跟踪的算法大致分为确定性方法和随机方法两大类。典型的确定性跟踪方法在对感兴趣目标进行跟踪时,将目标先验知识,如表观、颜色分布、轮廓信息等用于目标模板的建立。然后根据事先设定的相似性度量函数,在当前视频帧局部区域内,通过搜索或者迭代算法找到与目标模板或者目标表观最为相似的区域。随机跟踪方法利用状态空间对当前跟踪系统的运动进行模型化,该方法引入了概率统计的思想,将不确定性观察(如概率密度函数)与不同的状态相联系,从而不再假设运动系统输入与输出的完全确定性。

#### 1.3.2.1 确定性目标跟踪方法

视觉目标跟踪算法根据跟踪目标的表示方法和相似性度量方法可以大致分

为：基于特征匹配的跟踪算法[62]、基于区域统计匹配的跟踪算法[32]、基于模型匹配的跟踪算法[63]和基于 Mean-Shift 的跟踪算法[10]。

**基于特征匹配的跟踪方法**不考虑运动目标的整体特征，即不关心具体的运动目标，而只通过其特征（如 SIFT 特征点等）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，因此可以利用其完成目标的整个跟踪过程。无论是刚体运动目标还是非刚体运动目标，利用基于特征匹配的方法进行目标跟踪时主要包括特征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，并且在后续的视频帧提取相应的特征；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时，首先把需要跟踪的每一个人用一个矩形框封闭起来，在封闭框内进行跟踪特征的提取并建立特征模板，然后在后续的视频帧中提取相应的特征并与特征模板进行匹配，从而通过找到最佳的匹配位置完成跟踪。除了用单一的特征来实现跟踪外，还可以将多个特征信息进行综合作为跟踪特征，来提高跟踪的稳定性。另外，如果运动目标简单，可以将整个目标作为特征模板来进行跟踪，这种方法也被称为模板匹配。

**基于区域统计匹配的跟踪方法**是把图像中运动目标连通区域的共有特征信息作为跟踪特征的一种方法。在图像中常用的共有特征信息如颜色特征、纹理特征等。这种方法不需要在视频序列中找到完全相同的特征信息，而是通过计算候选区域与原始目标之间的区域统计特征的相关性来确定跟踪目标的最佳位置。例如，[64]利用小区域特征进行室内单目标人体的跟踪。该方法将人体看作由头、躯干、四肢等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。同时，该方法还常常利用滤波技术来估计和预测区域的几何形状和运动速度。

**基于模型匹配的跟踪方法**是通过建立模型来表示目标，也就是对运动目标进行建模，然后在图像序列中匹配这个建立好的模型来实现目标的跟踪。然而，在实际应用中跟踪的目标大部分都是非刚体的，其形状在不断发生变化。即使是刚体目标，由于拍摄视角及拍摄距离的不断变化也会导致其形态不停地发生变化，很难得到准确的几何模型。因此，一些学者提出通过变形轮廓模板来进行目标跟踪。目前有两种较为流行的可变形模型：自由式的可变形模型和参数可变的模型。

前者主要通过满足一些简单的约束条件(连续性、平滑性等)来跟踪任意形状的运动目标；后者通过使用参数公式或变形公式来描述目标的形状。

**基于 Mean-Shift 的目标跟踪算法**采用加权的灰度或颜色直方图来描述跟踪的目标，并通过梯度下降的方式快速迭代的查找跟踪目标。Mean-Shift 最早是由 Fukunaga 等人于 1975 年提出来的，随着 Mean-Shift 理论的发展，目前所说的 Mean-Shift 算法一般是指一个迭代的过程，即先算出当前点的偏移均值，移动该点到其偏移均值，然后以此点为新的起始继续移动，直到满足一定的条件结束。Mean-Shift 算法的形式简洁，收敛条件宽松。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。由于无需对概率密度分布进行估计，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。Mean-Shift 算法用于视觉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，作为一个无参数概率密度估计算法，很容易和其他算法结合使用；然后，采用加权直方图模型，对部分遮挡、目标旋转、形变和背景运动不敏感。但是，其也有不足之处，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，因此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，跟踪效果往往不好；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。

### 1.3.2.2 随机性目标跟踪方法

随机性目标跟踪常用的预测算法有 Kalman 滤波[20]以及粒子滤波[7,21]等算法。利用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后，可以在相对较小的区域内进行搜索，完成对目标的跟踪过程。当目标被遮挡时，利用滤波器对目标的运动轨迹进行可靠预测，可以方便地在特定区域内搜索目标，等待目标的重新出现。一般意义上，定义滤波器的系统状态为  $x_k$ ，其是一个四维向量  $(s_x, s_y, v_x, v_y)^T$ ，各分量分别表示运动目标在坐标轴上的位置和在坐标轴方向上的速度。然后在基于滤波算法的跟踪过程中，初始化目标的状态参数（如：速度、位置等），即使在不准确的情况下，经过迭代运算，上述滤波算法仍然可以准确地估计出目标的真实运动状态参数，具有预测的稳定性。滤波方法

的优点是：1、采用递归滤波的方法，可以将任意一点作为初始状态开始递归；2、计算量小，可实时计算；3、预测具有无偏、稳定和最优等特点。

#### 1.4 本文的研究内容

本文主要提出了基于综合颜色直方图和梯度方向直方图的目标跟踪特征，并在传统滤波算法的框架内，对目标跟踪的综合特征进行选择 and 自适应的评估，旨在获取目标紧致有效的特征描述，使其在目标及环境发生变化等复杂环境中有效提高目标特征区分能力以及表达能力，同时通过重构样本集合对跟踪目标的线性重构模拟目标的局部遮挡模型，使其在被遮挡的情况下仍然具有比较好的鲁棒性。主要研究内容如下：

1、提出了一种面向视觉目标跟踪的综合特征表示。基于颜色直方图（Histogram of Color, HC）和梯度方向直方图（Histogram of Oriented Gradient, HOG）的综合直方图 HOGC（Histogram of Oriented Gradient and Color）融合了颜色、局部轮廓以及梯度方向等特征。同时，基于 SIFT（Scale Invariant Feature Transform）特征点的主方向方法，进一步对综合特征进行了面向视觉目标跟踪的改进，在一定程度上解决了 HOG 对跟踪目标旋转敏感的缺点。

2、基于综合特征表示，提出了基于自适应综合特征表示的视觉目标跟踪算法。基于前景（跟踪目标）/背景的特征信息以及相关变化，将目标综合特征的评估嵌入到传统的滤波框架中，保证特征权重的变化是连续的过程，即具有时间连续性。基于滤波框架的特征评估有效地将滤波算法从对目标运动状态的建模扩展到了对目标特征权重的评估，拓展了传统滤波框架的应用范畴，同时，在一定程度上解决了跟踪前景/背景在跟踪过程中的变化对跟踪效果的影响。

3、基于自适应综合特征表示，进一步提出了基于稀疏自适应综合特征表示的视觉目标跟踪算法。在跟踪的初始化过程中建立在线训练样本集合，用以对跟踪特征的实时稀疏选择，从而得到跟踪目标具有前景/背景区分性的稀疏特征表示。同时，基于滤波框架对目标的稀疏特征表示进行有效地自适应的评估，在保证特征描述具有区分性的基础上，进一步保证特征的环境自适应性，从而能够更好地解决前景/背景在跟踪过程中的变化对跟踪效果的影响。

4、基于稀疏自适应特征表示，提出了基于分层自适应稀疏表示与稀疏重构的目标跟踪新框架。在跟踪的过程中，基于跟踪目标的稀疏自适应特征表示以及

基于搜索区域建立的实时重构样本集合，获得目标在样本空间的自适应重构系数，从而统计跟踪目标在样本空间的概率密度分布，最终完成对目标的跟踪。所提出的跟踪框架既能较好的解决前景/背景的变化，同时又能较好的解决跟踪目标的局部遮挡等问题对跟踪效果的影响。本文的主要研究内容总结如图 1-1 所示：

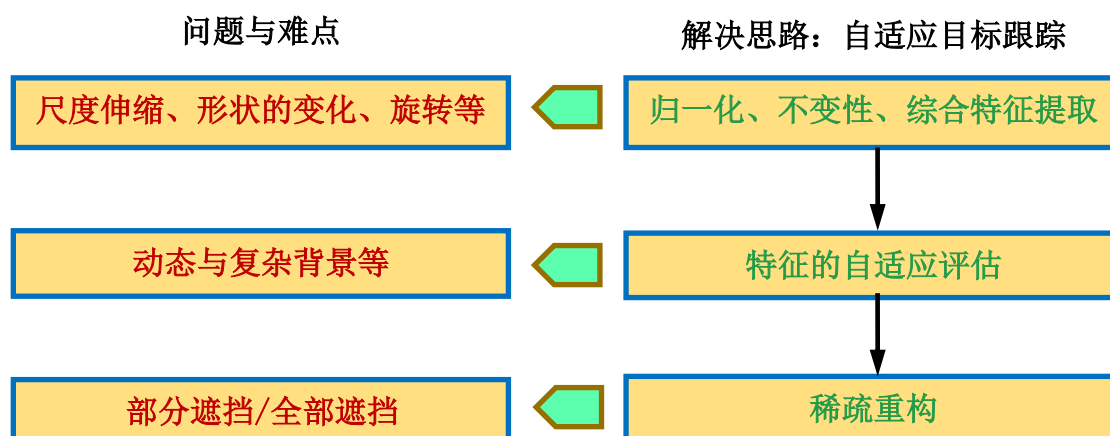


图 1-1 本文研究内容框图

## 1.5 本文的组织结构

第一章，绪论。主要论述了视觉目标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觉跟踪算法中存在的难题，给出了本文的主要研究目的和研究内容，总结了本文的主要贡献。

第二章，视觉目标自适应跟踪算法基础。主要对综合特征的提取和改进进行详细的描述。

第三章，基于自适应综合特征的视觉目标跟踪算法。详细介绍了基于综合特征集和滤波框架对特征权重进行时间连续性评估的算法。分别在 Kalman 滤波框架和粒子滤波框架中进行实验验证，给出了该跟踪算法在各种复杂场景下的运动目标跟踪实验，从而对算法在运动目标尺度伸缩变化、前景/背景动态变化等情况下的鲁棒性跟踪进行验证，同时将该算法与传统的跟踪算法的性能进行了比较并对比较结果进行了分析。

第四章，基于自适应稀疏特征表示的视觉目标跟踪算法。详细介绍了基于自适应综合特征和压缩感知理论对特征进行稀疏选择，使得跟踪目标的紧致表示具有很好的前景/背景区分性。同时在滤波框架中对选择的特征进行自适应的评估，

进一步保证跟踪目标的自适应稀疏表示具有时间连续性。最后，给出了该跟踪算法在各种复杂场景下的运动目标跟踪实验，从而对算法在运动目标尺度伸缩变化、形变、前景/背景动态变化等情况下的鲁棒性跟踪进行了验证，同时将该算法与传统的跟踪算法的性能进行了比较与分析。

第五章，基于分层自适应稀疏表示与稀疏重构的视觉目标跟踪新框架。详细介绍了基于稀疏自适应综合特征和稀疏重构对目标进行跟踪的过程。稀疏自适应特征表示使得跟踪目标和重构样本集合具有很好的区分性，同时稀疏重构保证跟踪目标在部分被遮挡的情况下仍然能够从重构样本集合中获取最紧致的重构样本子集，使得跟踪即能够有效解决前景/背景的变化，同时也能够解决跟踪目标的局部遮挡等难题。最后，给出了该跟踪算法在各种复杂场景下的运动目标跟踪实验，从而对算法在运动目标尺度伸缩变化、形变、前景/背景动态变化、部分遮挡等情况下的鲁棒性跟踪进行了验证，同时将该算法与传统的跟踪算法的性能进行了比较与分析。

最后，论文对现有工作进行了总结并对未来工作的进行了展望。对如何进一步提高跟踪算法的鲁棒性、准确性和实时性等研究热点和难点进行了探讨。

## 第二章 视觉目标自适应跟踪算法基础

以 Kalman 滤波器为代表, 传统的目标跟踪研究侧重于将目标简化成一个理想的坐标点, 进而对目标运动特性进行建模。而近年来, 随着人们对目标跟踪精度及目标表观数据表示的多样化, 人们更多地将视觉跟踪作为模式识别中一个分类问题进行研究[11, 23, 32], 即如何将跟踪目标区域(前景)与跟踪背景(背景)进行更好的区分, 使得属于前景或者背景的像素更容易进行聚类, 从而提高跟踪的精度和鲁棒性。将目标跟踪看成是前景背景分类问题的思想为目标跟踪开辟了一条新的途径, 带来了更广阔的应用前景。从理论上讲, 任何模式识别中的特征选择和分类器构建方法都可以进行应用。但是, 对于目标跟踪而言, 在线分类与实时性要求较高, 所以传统的模式分类中复杂的特征选择和分类方法难以实用。

### 2.1 基于前景背景联合建模的视觉目标跟踪算法简介

Collins 和 Liu[11]提出了一种基于前景背景联合建模的跟踪算法, 即一种在线选择具有前景和背景区分性的跟踪特征的方法, 该算法主要是基于前景与背景特征相似度的变化而进行特征权重的调整。Collins[11]等针对颜色特征对目标的视角、目标之间的部分遮挡以及目标形状的鲁棒性提出了一种基于颜色的目标特征描述的新方法。他们所提出的跟踪特征集合是颜色特征  $R$ 、 $G$  和  $B$  的线性组合, 定义如下:

$$F_1 = \{w_1R + w_2G + w_3B \mid w_* \in [-2, -1, 0, 1, 2]\} \quad (2-1)$$

其中, 每个颜色分量参数  $w_*$  的取值范围为 $[-2, 2]$ 之间的整数, 因此原则上便会得到 125 种基于三种颜色分量的组合特征。消除特征之间的冗余, 例如三个颜色分量的参数相等等情况之后, 最终得到 49 种组合特征, 从而构成了他们所定义的颜色特征集合。然后, 将每种颜色特征都归一化到 0-255 之间, 将得到的颜色统计结果离散化为直方图, 这样就得到特征集合中每个颜色特征的直方图描述形式。这种特征集合有如下三个优点: 1、计算简单, 由于特征都是基于视频帧中每个像素颜色分量的统计, 所以比较直观、快速; 2、这种特征集合类似于三维颜色空间在一维空间中的投影, 基本上覆盖了原有的颜色特征空间; 3、相比于原有的颜色特征定义, 这种特征集合还包含了比原有颜色特征更多的描述形式,



例如： $R-B$ 、 $2G-R-B$ 等颜色分量的组合。

当提取得到前景和背景的跟踪特征之后，Collins 等人首先计算了每种特征在前景与背景之间的相似性，定义如下：

$$L(i) = \log \frac{\max\{p(i), \delta\}}{\max\{q(i), \delta\}} \quad (2-2)$$

其中， $p(i)$  定义为目标中的第  $i$  个特征，而  $q(i)$  为背景中的第  $i$  个特征。 $L(i)$  为目标与背景基于第  $i$  个特征的  $\log$  似然比，包含了前景与背景在某一个特征之间的相似度信息。然后，通过统计第  $i$  个特征在视频帧中得到的每一个  $L(i)$ ，并计算其方差便得到了目标与背景在第  $i$  个特征之间的判别力，公式表达如下：

$$VR(L; p, q) = \frac{\text{var}(L; (p+q)/2)}{[\text{var}(L; p) + \text{var}(L; q)]} \quad (2-3)$$

其中， $\text{var}$  定义为某一特征相似度的方差。

Liang 等人[23]提出了一种类似的方法，其特征的判别力是通过贝叶斯误差率定义的，即通过目标和前景颜色特征直方图的交集进行计算得到的。Yin 和 Collins[65]进一步扩展了[11]的工作，鉴于背景可能具有多模态分布的特点，以至于一种特征可以很好地区分前景和其中的一种模态，而对另外的模态无效，他们提出对背景进行分割，对每一组目标和背景选择不同的特征并且生成一幅权重图像。最后，将这些权重图像组合并结合运动信息和均值移动算法对目标进行跟踪。Avidan[66]采用 AdaBoost 算法在线地训练得到前景和背景分类器，并用其对下一帧图像进行分类得到一幅权重图像。同样地，均值移动算法被用来对目标进行定位。此外，通过剔除过时的弱分类器并补充进新的弱分类器可以对强分类器进行更新，从而适应前景和背景的表现发生变化的情况。Grabner 和 Bischof[67]提出了采用在线的 Boosting 算法进行分类器构建的方法并将其应用到目标跟踪中。Wang 等[68]提出了一种采用粒子滤波进行目标跟踪和特征选择的方法，他们采用的特征是类似于 Haar 小波的特征，该特征在人脸检测领域得到了成功的应用。其优点是可以进行快速的计算并且具有很强的表示能力。通过粒子滤波中的背景粒子采样得到背景信息，这样可以采用 AdaBoost 建立一个强分类器，以其作为粒子的观察模型。该分类器在跟踪过程中还会动态进行更新。

文献中已有的算法大都采用单一的特征来描述（颜色或者纹理）跟踪目标，并且他们在特征评估方面都有一定的局限性（如：无法保证特征选择/评估具有时间连续性）。因此，在他们研究的基础上，本章后续部分将主要针对跟踪目标的更完备描述进行研究，即面向跟踪目标的综合直方图的提取和改进。综合直方图将目标的颜色和局部轮廓特征有效地融合起来，其对目标描述的更加准确，且包含更多目标的信息量，为本文后续的特征在线选择/评估等研究内容提供了基础。

## 2.2 综合特征的提取

在过去的几年，颜色直方图（Histogram of Color, HC）被大量的应用于目标跟踪，并且取得了不错的效果。但是大量的实验表明，当目标和背景颜色比较相似的时候，基于颜色的跟踪算法往往不能取得很好的效果。近几年，梯度方向直方图（Histogram of Oriented Gradient, HOG）开始被广泛地用来进行目标的检测和跟踪，并且取得了很好的效果。HOG 提取的是目标的边缘和轮廓特征，而这些特征对目标的颜色变化不敏感。Dalal 等人[69][102]已经通过实验验证了固定尺度的 HOG 和 SIFT 对目标的描述能力几乎一样强。但是，HOG 同样有一些缺点，比如在目标跟踪方面 HOG 的性能要差于 HC，并且当目标出现大面积的平滑区域时，HOG 的性能会很快的降低。虽然每种单独的特征都有一些缺陷，但 HC 和 HOG 两种特征之间的互补性使得我们有理由把它们融合在一起。

基于 HC 和 HOG 两种特征在目标描述方面的互补性，以及受到 SIFT 特征提取、匹配以及 HOG 特征在人体检测中的启发，本文提出了一种融合了颜色和梯度方向信息的直方图，称之为综合直方图 (Histograms of Orientated Gradient and Color, HOGC)。

## 2.2.1 跟踪目标区域的定义

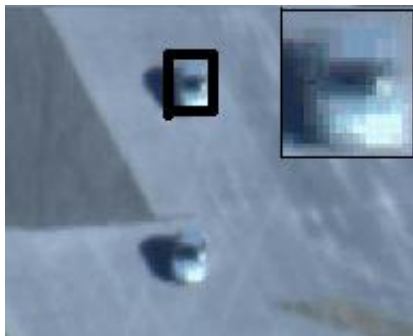


图 2-1 目标区域定义

在本文的研究背景下，跟踪目标可以定义为任意尺寸和形状的兴趣物体。例如，水面上的船只、天上的飞机、地面上的汽车和行人等。要对感兴趣的目标进行跟踪，首先需要给出目标的区域。常见的基本几何形状目标的区域定义通常包括矩形框和椭圆。本文采用矩形框的形式描述跟踪目标区域。如图 2-1 所示，本文中的目标区域定义为  $h \times w$  的矩形框。目标的特征  $\{F_i(x)\}, i = 0, 1, \dots, N$  从以  $x$  位置为中心的目标区域的像素中提取。

## 2.2.2 颜色直方图提取

在 RGB 颜色空间中，每个颜色分量（R、G 和 B）被均匀量化为一个 16 维的颜色直方图，三个颜色分量总计形成了一个 48 维的颜色直方图。以 R 颜色分量为例，采用窗宽为  $h_1$  的剖面函数  $k(\square): [0, \infty) \rightarrow R$  进行平滑后的颜色直方图为：

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

其中， $I(X_i)$  是视频帧中  $X_i$  位置处像素的颜色值（R、G 或者 B）。其中， $y_0$  是目标区域的中心位置； $n$  是目标区域像素的数目； $\hat{q}_b(y_0)$  是平滑后直方图分量  $b$  的值； $\delta(\cdot)$  是 Kronecker Delta 函数； $m$  是颜色直方图级的个数（ $m=16$ ）； $X_i$  是第  $i$  个像素点的位置且  $X_i = [h_i, w_i]$ ，常数  $C_1$  由约束条件  $\sum_{b_j=1}^M \hat{q}_{b_j}(y_0) = 1$  得到，即

$$C_1 = \frac{1}{\sum_{i=1}^n k \left( \left\| \frac{y_0 - X_i}{h_1} \right\|^2 \right) (1 + d(X_i))} \quad (2-5)$$

### 2.2.3 梯度方向直方图的提取

当给定特定的图像大小和特征尺度，梯度直方图可以很好地表达出跟踪目标的局部轮廓。此特征因为只保留方向特征，不保留强度特征，因此对于亮度、光照等变化具有一定的适应性。它与颜色直方图可以形成目标表示上的互补性。

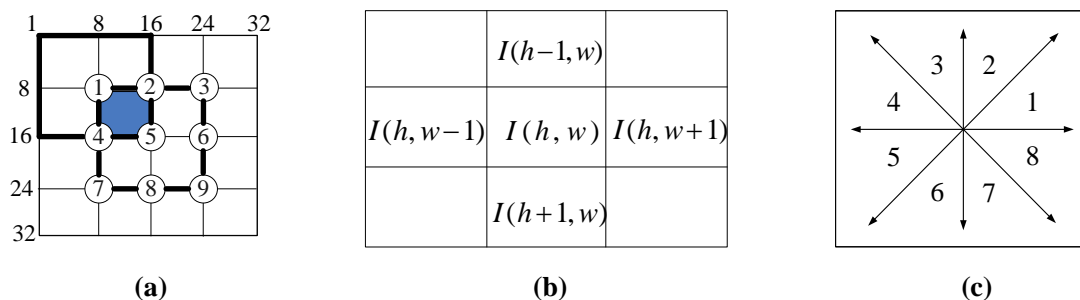


图 2-2 HOG 特征提取示意图。(a)图像窗口的划分，(b)计算梯度方向的模板，(c)8 个方向投影区间。

HOG 特征提取的详细过程：将跟踪目标转化为灰度图像；将跟踪目标窗口归一化到固定大小的图像窗口：32x32 像素；然后把图像划分为 16 个 8x8 的小区域（cell），并且每个 2x2 的相邻小区域组成一个图像块（block，标号为从①到⑨），因此每个相邻的图像块有 2 个小区域彼此重合；本文将每个图像块的梯度方向统计投影到一个 8 维的梯度方向直方图中（图 2-2c 所示）。从图 2-2a 中可以看到，由于每个图像窗口包含 9 个图像块，因此每个目标得到 72 维的梯度方向直方图来描述其局部轮廓。每个像素的梯度方向计算如下（如图 2-2b 所示）：

$$\begin{aligned} I &= G(\sigma, 0) * I_0 \\ dy &= I(h+1, w) - I(h-1, w) \\ dx &= I(h, w+1) - I(h, w-1) \\ ori(h, w) &= a \tan 2(dy, dx) \quad ori \in [0, 2\pi] \end{aligned} \quad (2-6)$$

其中，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， $I$  表示经过平滑滤波后的图像。 $ori(h, w)$  表示

像素梯度方向（目标轮廓方向）。对目标区域/背景区域的所有像素进行累计，可以获得 72（9 个图像块，每个块 8 个方向，如图 2-2a 和 c 所示）维 HOG 特征。

### 2.3 综合特征的改进

为了使得 HOG 特征能够适应于目标在视频序列中的旋转运动，本文采用了 SIFT 算法中的主方向参考思想调整 HOG 的特征。即在计算的过程中，在提取图像的 HOG 特征之前，在图像窗口中心的邻域范围内进行采样，提取该邻域内像素的粗的梯度方向直方图，其峰值则代表了该图像窗口的主方向（如图 2-3a1 与 a2 所示）。在跟踪的过程中，提取候选目标图像窗口的粗的梯度方向直方图（如图 2-3b1 与 b2 所示），将其所对应的主方向都归一化到初始跟踪目标的主方向上（如图 2-3c1 与 c2 所示），这样提取的 HOG 特征在一定程度上与目标的旋转运动无关。

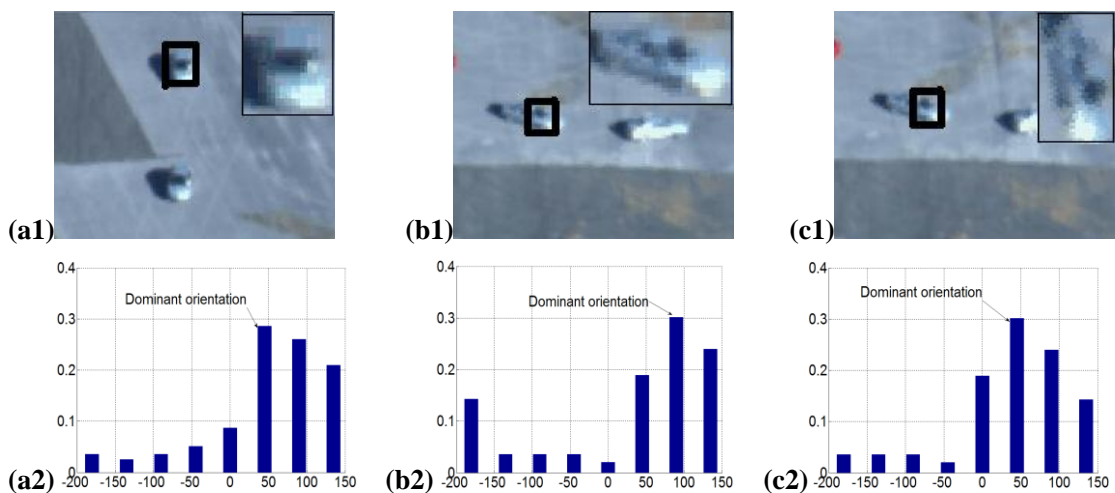


图 2-3 利用 HOG 主方向对目标进行旋转的示意图

## 2.4 基于综合特征视觉目标跟踪的实验结果分析与比较

### 2.4.1 本文目标跟踪实验环境配置

本文的实验测试视频均选自于 VIVID [70], CAVIAR [71] 和 SDL [72] 这三个公共测试集合。测试跟踪视频主要包括运动的人体和车辆，同时测试视频包括了运动目标的遮挡、尺度变化、旋转、形变以及动态背景等跟踪过程中经常遇到的难点问题。

表 2-1 跟踪算法测试视频列表

Video test set	Video name
VIVID Tracking video set	redteam
	egtest01
	egtest02
	egtest04
SDL Tracking video set	xiangshan_0032
	xiangshan_0043
CARVIA Tracking video set	Browse1
	Fight_Chase
	OneStopMoveEnter1cor
	EnterExitCrossingPaths2front

为了量化的对比每种跟踪算法，本文从三个公共测试集中挑选了 10 段测试视频（如表 2-1 所示）进行比较，同时定义了量化参数，即跟踪结果与手工标定目标实际位置之间的相对错误率（Relative Displacement Error Rate, DER），如下：

$$DER = \frac{d(\overline{O}_i(x, y, s), \overline{GT}(x, y, s))}{\sqrt{AREA_{GT}}} = \frac{\sqrt{(x_o - x_{GT})^2 + (y_o - y_{GT})^2}}{\sqrt{wid_{GT} \times hei_{GT}}} \quad (2-7)$$

其中， $\overline{GT}(x, y, s)$  表示手工标定的目标真实位置， $\overline{O}_i(x, y, s)$  表示跟踪算法获取的目标位置， $AREA_{GT}$  表示手工标定目标区域  $\overline{GT}(x, y, s)$  的面积。从 DER 的定义可以分析得到，DER 的均值越小，跟踪算法的相对错误率越小，表明跟踪算法性能越好，同时 DER 的方差越小，跟踪算法相对错误率的变化值越小，表明跟踪算法的稳定性越好。

#### 2.4.2 实验结果分析与比较

本节对基于 HOGC, Color [73] 和 SIFT [74] 特征的跟踪算法的性能进行比较。图 2-4 中所示的第一段测试视频来自于 VIVID 数据集，跟踪目标为一辆运动模式相对简单的吉普汽车，跟踪背景变化相对稳定。同时，由于跟踪目标与背景比较明显的特征差别（目标主要为白色和红色，背景主要为灰色与绿色），三种特征均取得了相对较好的跟踪结果，其中基于 SIFT 特征点的跟踪算法由于背

景中灌木丛产生的 SIFT 特征点的影响，导致了跟踪相对不稳定（图 2-4b 中的第 300 视频帧所示）。

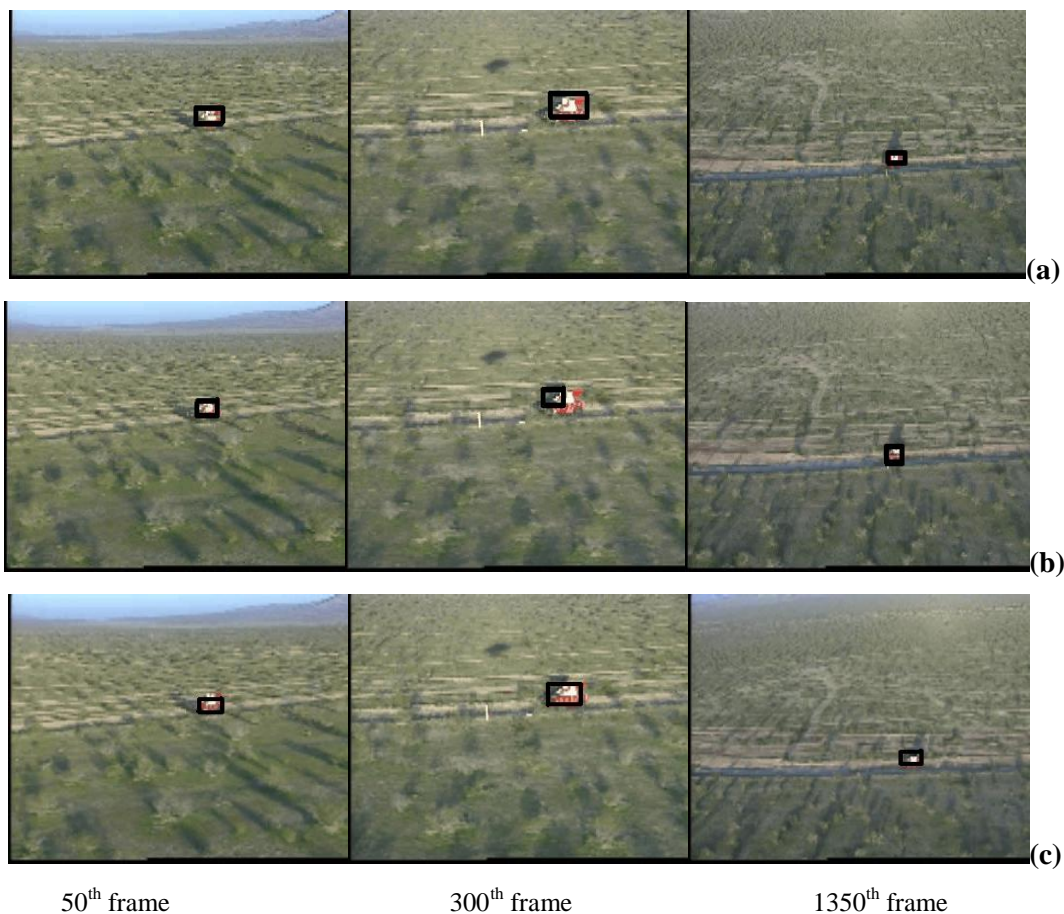


图 2-4 基于 VIVID-redteam 测试视频的目标跟踪结果。(a)基于颜色特征的目标跟踪结果，(b)基于 SIFT 特征的目标跟踪结果，和(c)基于 HOGC 的目标跟踪结果。

图 2-5 中所示的第二段跟踪测试视频仍然来自于 VIVID 数据测试集合，测试视频中的跟踪目标为一辆在比较相似背景中运动的车辆。首先，跟踪目标做了一个 180° 的调头，然后加速直行，直到超过另外 3 辆车。在整个跟踪过程中，跟踪目标具有尺度变化，旋转，同时背景中有一辆与跟踪目标几乎完全相似的车辆，因此导致这段测试视频具有一定的挑战性。如图 2-5a 所示，基于颜色的跟踪算法由于跟踪目标与背景颜色特征的相似性，在第 200 帧时产生了跟踪不稳定，在第 700 帧时产生了跟踪错误。由于背景相对平滑（背景产生较少的 SIFT 特征点干扰），基于 SIFT 特征跟踪算法能够比较好的描述跟踪目标，取得了比较好的跟踪结果（如图 2-5b 所示）。基于 HOGC 的跟踪算法由于同时包括跟踪目标的颜色和局部轮廓信息，因此也取得了令人满意的跟踪结果(如图 2-5c 所示)。

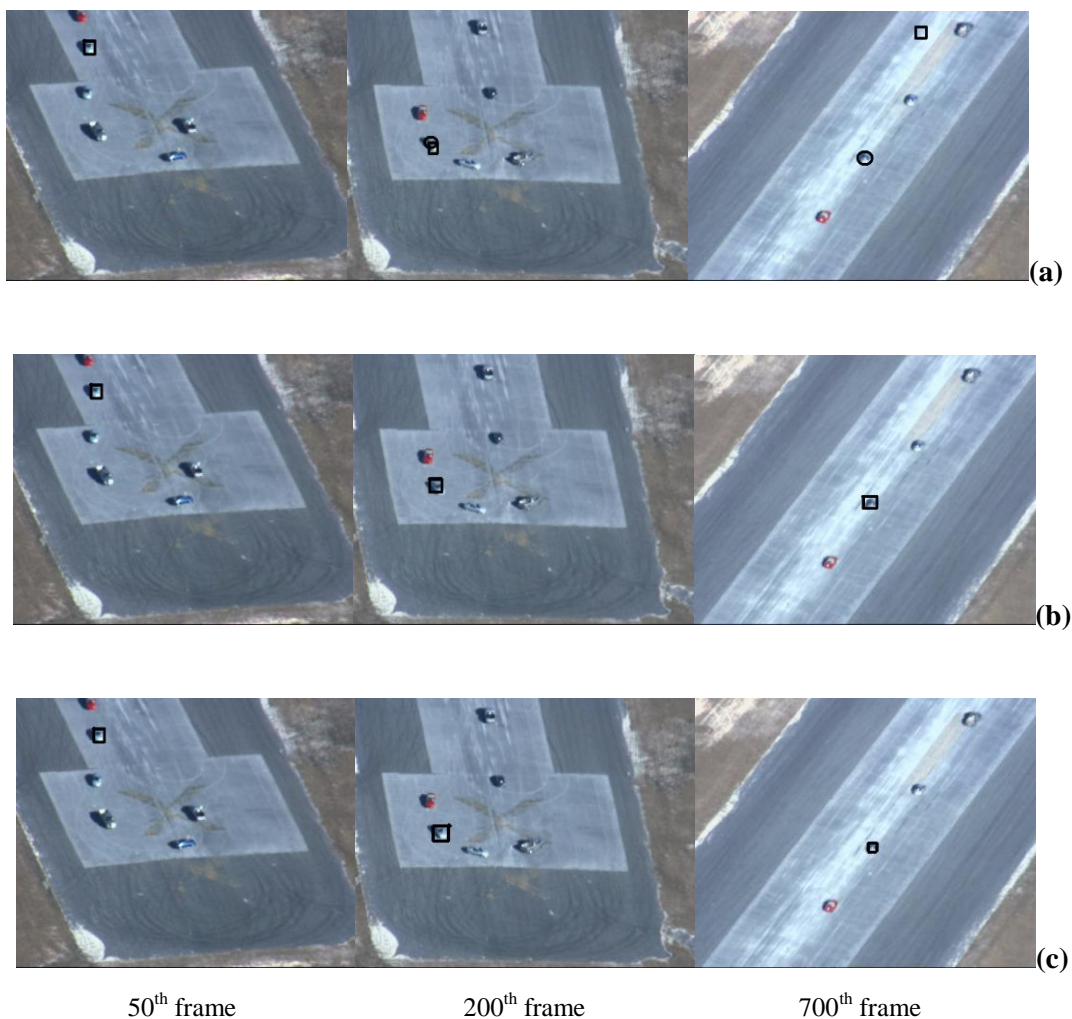


图 2-5 基于 VIVID-egtest01 测试视频的目标跟踪结果，如果跟踪错误发生时，黑色的椭圆表示手工标定的目标实际位置。(a) 基于颜色特征的目标跟踪结果，(b) 基于 SIFT 特征的目标跟踪结果，(c) 基于 HOGC 的目标跟踪结果。

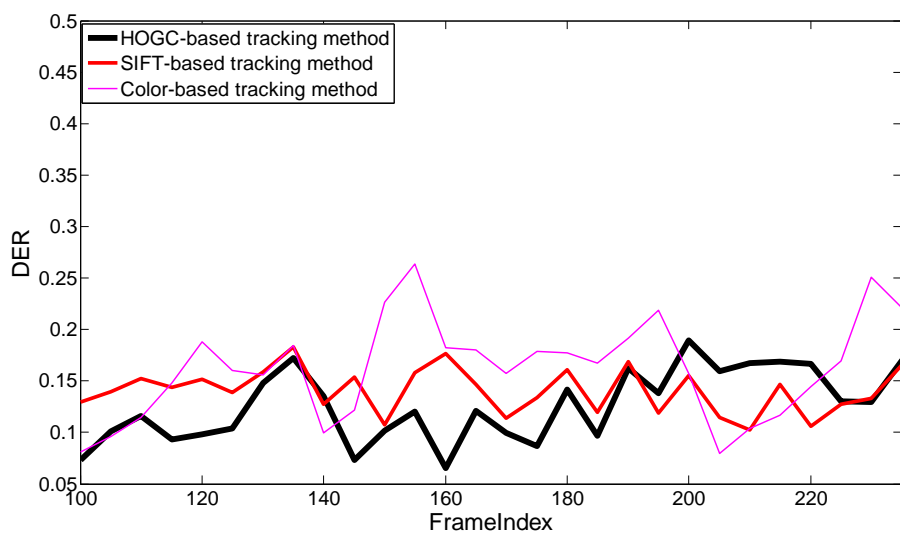


图 2-6 三种目标跟踪算法 DER 结果对比



在定量对比实验中，本文采用表 2-1 中所列出的 10 段测试视频（从 100 帧到 250 帧）的平均 DER 值进行比较。如图 2-6 所示，基于 HOGC 的跟踪算法和基于 SIFT 的跟踪算法都具有相对较小的 DER 值（大约 0.05 到 0.2），而基于颜色特征的跟踪算法具有相对较大的 DER 值（大约 0.1 到 0.25）。DER 的统计值表明，HOGC 和 SIFT 对目标的描述相比较颜色特征更好。但是 SIFT 特征的提取具有非常大的计算量，往往难以进行实时的提取，因此，不适用于实时性要求较高的应用系统。

### 2.5 小结

本章详细的描述了综合特征的提取以及改进。所提取的综合特征融合了目标的颜色和梯度方向信息，能够较完备的描述目标的颜色、局部轮廓、梯度方向等特征。

本章对提出的综合特征在若干段具有挑战性的公开视频序列上进行了定性的目标跟踪测试，并且与领域内经典的跟踪特征进行了定量化的跟踪结果比较。实验结果表明，在整体性能上，本章提出的综合特征要优于几个领域内经典的跟踪特征。

### 第三章 基于自适应综合特征的视觉目标跟踪算法

基于前景背景联合建模的跟踪算法,将视觉跟踪作为模式识别中一个分类问题进行研究,能够有效的将前景/背景的像素进行聚类,从而提高跟踪的精度和鲁棒性。然而,为了使得目标特征对前景/背景的变化具有适应性,即如何进行更好的保持特征变化的时间连续性以及稳定性还有待深入的研究。

#### 3.1 基于滤波框架的综合特征评估

由于颜色特征和梯度方向特征对于物体描述具有很好的互补性,基于上一章中所提取的综合特征,本章提出了通过在跟踪过程中在滤波框架中实时环境自适应的更新目标特征的权重。

##### 3.1.1 目标区域与背景区域定义

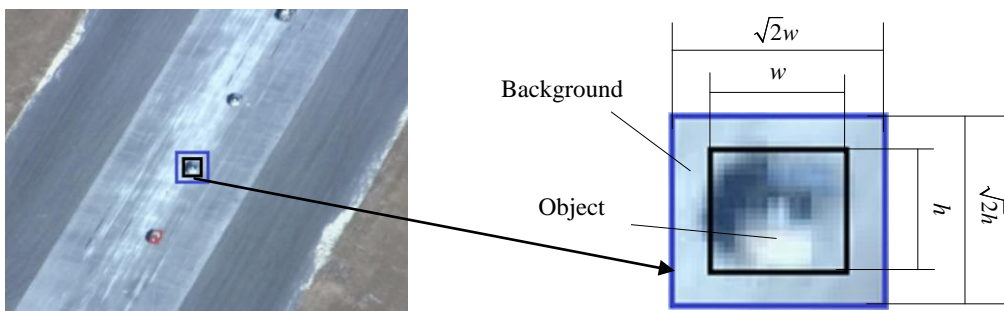


图 3-1 跟踪目标及其背景区域定义

如图 3-1 所示,目标区域定义为  $h \times w$  的矩形框,背景区域定义为  $\sqrt{2}h \times \sqrt{2}w$  的矩形框与目标区域之间的环形区域。目标的特征  $\{F_i(x)\}, i = 0, 1, \dots, N$  从以  $x$  位置为中心的目标区域的像素中提取,背景的特征  $\{B_i(x)\}, i = 0, 1, \dots, N$  从以  $x$  位置为中心的背景区域的像素中提取。

##### 3.1.2 特征判别力的定义

在本章中,依照 Collins [11]中定义的特征区分前景/背景的能力去定义特征的判别力,如公式 3-1 和 3-2 所示:

$$\boxed{S}_t^i = \max(0, \min(1, \log \frac{\max(F_t^i(x, y), \delta)}{\max(B_t^i(x, y), \delta)})), i = 1 \dots N \quad (3-1)$$

$$S_t^i = \frac{\boxed{S}_t^i}{\sum_{i=1}^N \boxed{S}_t^i}, i = 1 \dots \Lambda \quad (3-2)$$

其中， $F_t^i(x, y)$ 和 $B_t^i(x, y)$ 分别表示为 $t$ 时刻前景/背景综合特征中的第 $i$ 维特征。根据经验， $\delta$ 被赋值为0.005，用来避免除数为0。直观地，如果第 $i$ 维特征在前景占有更重要的比重，则 $\log \frac{\max(F_t^i(x, y), \delta)}{\max(B_t^i(x, y), \delta)}$ 的所得结果为正值，反之所得结果为负数或者0。因此， $\boxed{S}_t^i$ 用来描述第 $i$ 维特征的判别能力（第 $i$ 维特征在前景/背景中所占比重），即第 $i$ 维特征区分前景/背景的能力。 $\max()$ 和 $\min()$ 用来保证 $\boxed{S}_t^i$ 的值在(0.0, 1.0)区间范围。公式3-2用来归一化 $\boxed{S}_t^i$ 。

### 3.1.3 特征权重的计算

基于公式3-1和公式3-2，能够计算得到 $S_t^i$ 用来描述第 $i$ 维特征的判别力。但是，由于视频中难以避免的噪声影响，因此不能直接定义特征判别力为特征的权重。考虑到视频变化是个连续的过程，因此在跟踪过程中目标特征的变化应该同样具有时间连续性，所以在本章采用滤波框架对特征的权重进行评估。

假设特征的变化过程是一个一阶马尔科夫过程， $w_t(i)$ 定义为 $t$ 时刻第 $i$ 维特征的权重，因此 $w_t(i)$ 不仅取决于 $t$ 时刻第 $i$ 维特征的判别力 $S_t^i$ ，同时取决于 $t-1$ 时刻第 $j$ 维特征的权重 $w_{t-1}(j)$ ，如公式3-3所示：

$$w_t(i) = f_{t,t-1}(w_{t-1}(j), S_t^i) + u_t \quad (3-3)$$

其中， $f_{t,t-1}$ 表示为一个滤波过程，当 $i=j$ 时，其表示为一个Kalman滤波，当 $i \neq j$ 时，其表示为一个粒子滤波（本章的后续章节会详细叙述此部分内容）。 $u_t$ 表示为高斯噪声。在每一次特征权重更新的迭代过程中，公式3-4用来对特征的权重进行归一化。

$$w_t(i) = \frac{w_t(i)}{\sum_{i=1}^N w_t(i)}, i=1 \dots N \quad (3-4)$$

根据公式(3.3)以及(3.4)中权重的定义，基于滤波框架的特征评估的框图如下所示：

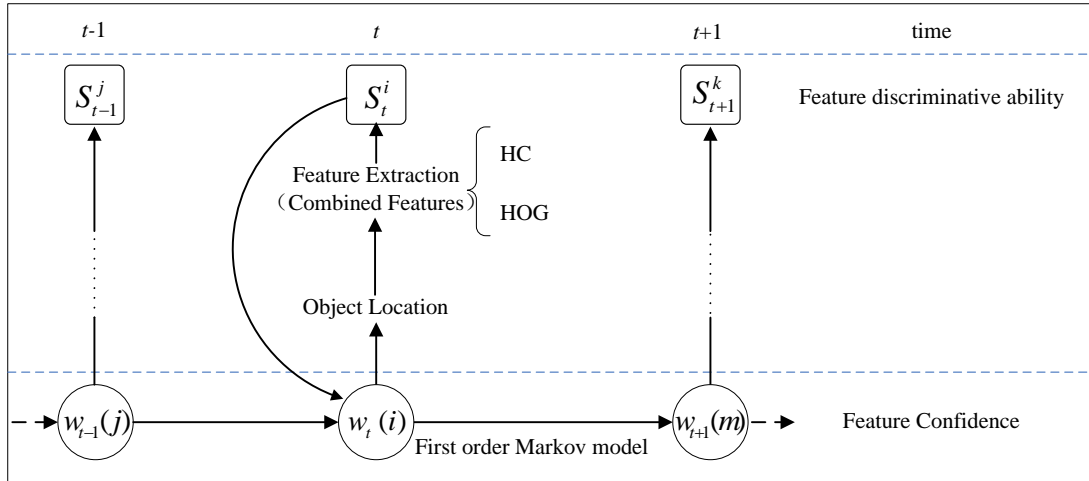


图 3-2 基于滤波框架特征评估框图

## 3.2 基于 Kalman 滤波综合特征评估的视觉目标跟踪算法

### 3.2.1 Kalman 滤波的基本原理

Kalman 滤波方法[44]是 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。他们把状态空间的概念引入到随机估计理论中，把信号处理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。

Kalman 滤波方法的优点是：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，Kalman 滤波方法在视觉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目标的位置，以此作为量测算法匹配搜索的起点，从而提高运算速度。其基本原理如下：

系统的状态方程：

$$X_k = A \times X_{k-1} + B \times u_{k-1} + w_{k-1} \quad (3-5)$$

系统的观测方程：

$$z_k = H \times X_k + v_k \quad (3-6)$$

其中， $w_k$  和  $v_k$  分别为独立不相关的状态噪声和观测噪声，并且分布满足如下公式：

$$\begin{aligned} p(w) &\square N(0, Q) \\ p(v) &\square N(0, R) \end{aligned} \quad (3-7)$$

状态预测方程（先验估计）：

$$\hat{X}(k|k-1) = A \times \hat{X}(k-1|k-1) + B \times u_{k-1} \quad (3-8)$$

经观测值修正过的后验估计：

$$\hat{X}(k|k) = \hat{X}(k|k-1) + K_k(z_k - H\hat{X}(k|k-1)) \quad (3-9)$$

其中， $\hat{X}(k|k-1)$  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； $\hat{X}(k|k)$  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； $z_k - H\hat{X}(k|k-1)$  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：

$$\begin{aligned} \hat{X}(k|k-1) &= X(k) - \hat{X}(k|k-1) \\ \hat{X}(k|k) &= X(k) - \hat{X}(k|k) \end{aligned} \quad (3-10)$$

先验估计误差和后验估计误差的协方差分别为：

$$P(k|k-1) = E\left[\hat{X}(k|k-1)\hat{X}^T(k|k-1)\right] \quad (3-11)$$

$$\begin{aligned}
 P(k|k) &= E\left[\bar{\mathbf{X}}(k|k)\bar{\mathbf{X}}^T(k|k)\right] \\
 &= E\left[(X(k) - \bar{\mathbf{X}}(k|k))(X(k) - \bar{\mathbf{X}}(k|k))^T\right] \\
 &= E[(X(k) - \bar{\mathbf{X}}(k|k-1) - K(z_k - H\bar{\mathbf{X}}(k|k-1))) \\
 &\quad (X(k) - \bar{\mathbf{X}}(k|k-1) - K(z_k - H\bar{\mathbf{X}}(k|k-1)))^T] \\
 &= P(k|k-1) - P(k|k-1)H^TK^T - KHP(k|k-1)^T \\
 &\quad + K(HP(k|k-1)H^T + R)K^T
 \end{aligned} \tag{3-12}$$

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：

$$K_k = P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1}. \tag{3-13}$$

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，则

$$\begin{aligned}
 \lim_{R \rightarrow 0} \bar{\mathbf{X}}(k|k) &= \bar{\mathbf{X}}(k|k-1) + H^{-1}(z_k - H\bar{\mathbf{X}}(k|k-1)) = z_k \\
 \lim_{P(k|k-1) \rightarrow 0} \bar{\mathbf{X}}(k|k) &= \bar{\mathbf{X}}(k|k-1) + 0 \times (z_k - H\bar{\mathbf{X}}(k|k-1)) \\
 &= \bar{\mathbf{X}}(k|k-1)
 \end{aligned} \tag{3-14}$$

**Kalman** 滤波使得每一时刻  $k$  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：

$$\begin{aligned}
 \bar{\mathbf{X}}(k|k-1) &= A\bar{\mathbf{X}}(k-1|k-1) + Bu_{k-1} \\
 P(k|k-1) &= AP(k-1|k-1)A^T + Q
 \end{aligned} \tag{3-15}$$

修正部分：

$$\begin{aligned}
 K_k &= P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1} \\
 \bar{\mathbf{X}}(k|k) &= \bar{\mathbf{X}}(k|k-1) + K_k(z_k - H\bar{\mathbf{X}}(k|k-1)). \\
 P(k|k) &= (I - K_kH)P(k|k-1)
 \end{aligned} \tag{3-16}$$

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因子  $\gamma$  ( $\gamma > 1$ )，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：

$$P(k|k-1) = \gamma A(k-1)P(k-1|k-1)A^T(k-1) + Q(k-1). \quad (3-17)$$

### 3.2.2 Kalman 滤波算法在传统目标跟踪中的应用

Kalman 滤波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[20]。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，因此可以假设目标在此时间间隔内是匀速运动，所以匀速运动模型在一般情况下可以很好地反映目标的运动趋势。

定义 Kalman 滤波的系统状态为  $x_k$ ，其是一个四维向量  $(s_x, s_y, v_x, v_y)^T$ ，分别代表运动目标在坐标轴上的位置以及沿着坐标轴方向的运动速度。根据运动目标在单位时间间隔内是匀速运动的假设，定义状态转移矩阵  $A$  和观测矩阵  $H$ ：

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \quad (3-18)$$

### 3.2.3 基于 Kalman 滤波的特征评估

提取的综合特征 HOGC 在跟踪的过程中将会根据背景和目标本身的变化被实时的自适应评估。不失一般性，在此做以下两个假设：1、特征权重和特征判别力都是在 (0.0,1.0) 之间的浮点数；2、权重调整之后，背景/前景判别力大的特征的权重也应该大，反之亦然。

在跟踪过程中，跟踪目标和背景的表现时刻发生着变化。在近年的目标跟踪研究中，为了使得特征对于背景的改变具有适应性，一些学者提出在跟踪过程中实时更新目标特征的权重来评估特征有效性。在本节中，提出了一种全新的基于 Kalman 滤波的特征权重调整策略。假定目标特征的变化在跟踪过程中满足线

性、高斯的假设，Kalman 滤波中的状态定义为特征的权重，Kalman 滤波中的观察定义为特征的判别力。则我们所提出的 Kalman 滤波的状态方程和观察方程可表示如下：

$$\begin{cases} \begin{pmatrix} w_{t+1} \\ \Delta w_{t+1} \end{pmatrix} = \begin{pmatrix} I_{N \times N} & I_{N \times N} \\ 0 & I_{N \times N} \end{pmatrix} \begin{pmatrix} w_t \\ \Delta w_t \end{pmatrix} + u_t \\ (S_t) = (I_{N \times N} \ 0) \begin{pmatrix} w_t \\ \Delta w_t \end{pmatrix} + v_t \end{cases} \quad (3-19)$$

其中， $w_t = \{w_t(1), w_t(2), \dots, w_t(N)\}$  表示第  $t$  视频帧中特征集所对应的特征权重集合， $w_t(i)$  表示第  $t$  视频帧中第  $i$  维特征所对应的特征权重，并且  $\Delta w_t = w_t - w_{t-1}$ ； $S_t = \{S_t^1, S_t^2, \dots, S_t^N\}$  定义为第  $t$  视频帧中特征集所对应的特征判别力集合， $S_t^i$  是第  $t$  视频帧中第  $i$  个特征所对应的特征判别力。 $u_t$  和  $v_t$  均为高斯白噪声。

### 基于 Kalman 滤波的特征权重评估算法

**1. 初始化 ( $t=0$ )**. 在这个阶段，首先初始化目标的特征权重为：

$$(w_0(i), \Delta w_0(i))^T = \left(\frac{1}{N}, 0\right)^T;$$

**2. 预测 ( $t>0$ )**. 对于特征集合中的每一个特征，进行如下操作：

- (a) 利用 Kalman 滤波预测该特征的先验权重；
- (b) 利用先验的加权 HOGC 特征来指导目标跟踪，找到目标所在位置。

**3. 修正 ( $t>0$ )**. 当搜索到目标在下一帧的最佳位置以后，进行如下操作：

- (a) 提取目标在下一帧中的 HOGC 特征；
- (b) 计算 HOGC 特征的判别力；
- (c) 通过 Kalman 滤波来修正 HOGC 特征的权重，得到当前帧中后验特征权重。

**4.  $t = t+1$ ;** 跳转到步骤 2 或者结束循环。

上面描述的利用 Kalman 滤波调整特征权重的过程如下图所示：



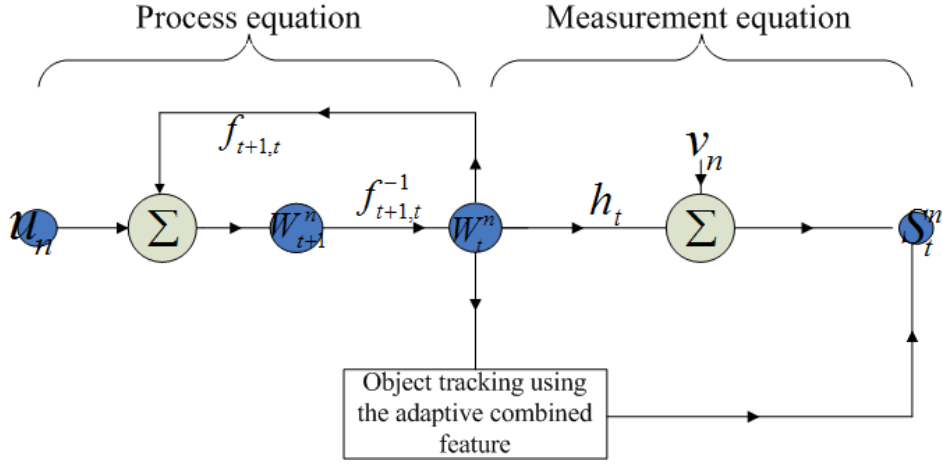


图 3-3 基于 Kalman 滤波的特征权重调整示意图

### 3.2.4 基于 Kalman 滤波特征评估的视觉目标跟踪

目标跟踪是对跟踪目标在搜索区域  $\Omega_t$  内进行搜索的过程。跟踪的目的为通过最大化  $p((x, y)_t | \Omega_t, (x, y)_{t-1})$  在  $\Omega_t$  内找到与跟踪目标最匹配的跟踪结果  $(x, y)_t$ 。通过贝叶斯推理过程来求解  $p((x, y)_t | \Omega_t, (x, y)_{t-1})$  的最优解，如下所示：

$$\begin{aligned}
 & \max_{(x, y)_t} p((x, y)_t | \Omega_t, (x, y)_{t-1}) \\
 & = \max_{(x, y)_t} \int p(F_t^i((x, y)_t^c) | (x, y)_t^c, F_{t-1}^i) p((x, y)_t^c | \Omega_t) dc \quad , \quad (3-20) \\
 & = \max_{(x, y)_t} \int \left( \sum_{i=1}^N (p(F_t^i) p(F_t^i((x, y)_t^c), F_{t-1}^i)) \right) p((x, y)_t^c | \Omega_t) dc
 \end{aligned}$$

其中， $(x, y)_t^c$  定义为  $\Omega_t$  中第  $c$  个候选目标的位置， $p(F_t^i((x, y)_t^c), F_{t-1}^i)$  表示为  $F_t^i((x, y)_t^c), F_{t-1}^i$  两个特征的相似度， $p(F_t^i)$  表示为  $t$  时刻第  $i$  维特征的置信度，此处不失一般性将其视为  $t$  时刻第  $i$  维特征的权重  $w_t(i)$ ，因此公式(3-20)可以描述如下：

$$\begin{aligned}
 & \max_{(x, y)_t} p((x, y)_t | \Omega_t, (x, y)_{t-1}) \\
 & = \max_{(x, y)_t} \int \left( \sum_{i=1}^N (w_t(i) p(F_t^i((x, y)_t^c), F_{t-1}^i)) \right) p((x, y)_t^c | \Omega_t) dc \quad . \quad (3-21)
 \end{aligned}$$

### 3.2.5 实验结果比较与算法性能分析

本节对所提出的基于 Kalman 滤波框架特征评估的视觉目标跟踪算法进行了

实验验证，并将其与三种经典的算法进行了比较。所选择的经典跟踪算法包括：基于特征选择的跟踪算法(Variance ratio feature shift based method)[11]、基于 SIFT 特征点的跟踪算法 (SIFT-Based method) [74]和传统的 Kalman 滤波跟踪算法 (traditional Kalman filter method) [20]。实验测试视频来自于 VIVID 测试视频集。以下示例中跟踪目标是在动态背景中的一辆小吉普车，视频是通过直升飞机从空中拍摄所得。

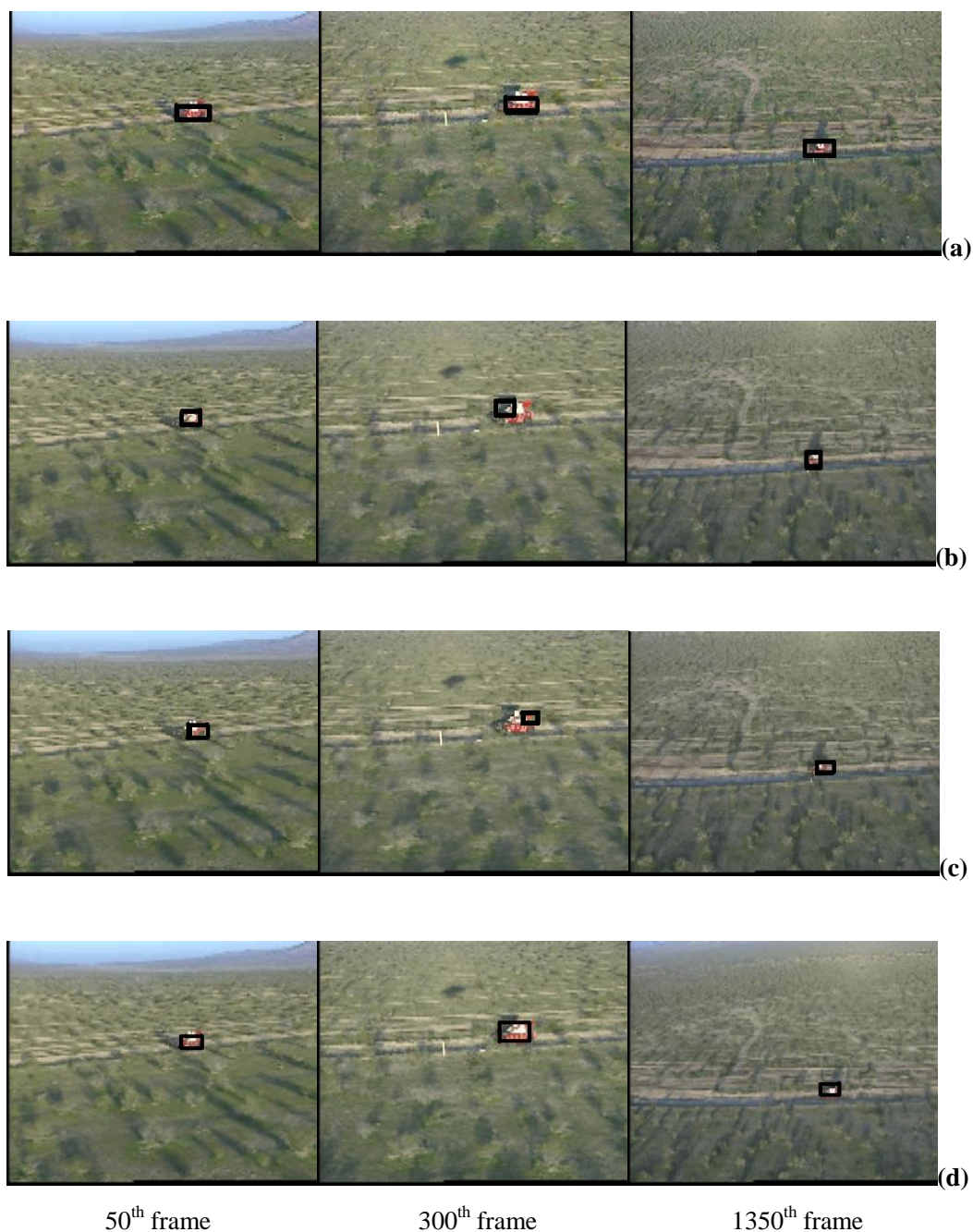
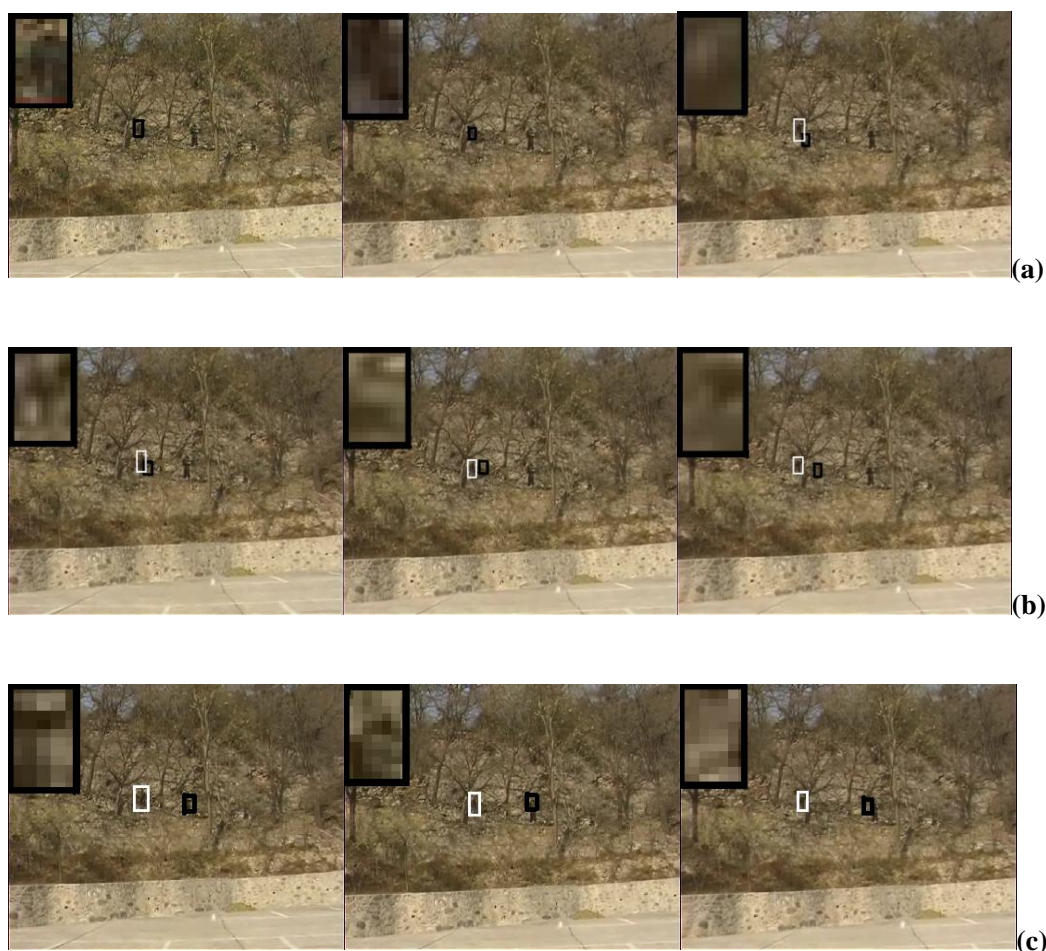


图 3-4 基于 VIVID-redteam 测试视频的目标跟踪结果对比图。(a) Variance ratio feature shift based method 的跟踪结果, (b) SIFT-based method 的跟踪结果, (c) traditional Kalman filter method 的跟踪结果, (d)本节所提出算法的跟踪结果。

在上图所示的跟踪实验结果中，由于跟踪目标的背景的变化、目标本身表观的变化以及相互之间的运动模式都相对比较简单，四种跟踪算法都取得了比较好的实验结果。与其他三种算法相比，SIFT-based method 算法取得的结果相对较差，这是因为虽然 SIFT 特征点对目标有很强的描述能力，但是由于背景中的灌木丛在目标特征点的计算过程中同样会产生大量的 SIFT 特征点，从而导致有很多背景的噪声点掺杂在跟踪目标 SIFT 特征点的描述中，使得对跟踪目标描述的准确度降低，从而导致跟踪实验结果不稳定。Variance ratio feature shift based method 和 traditional Kalman Filter method 在这段测试视频的跟踪结果都比较好，其中 Variance ratio feature shift based method 算法通过特征选择可以使对跟踪目标的描述自适应背景的变化，traditional Kalman Filter method 通过预测和修正目标的运动模型可以很好地估计目标在视频帧中的位置。尽管四种跟踪算法都取得了比较好的实验结果，从实验结果图中还是可以看出本节所提出的跟踪算法取得的实验结果最为稳定。



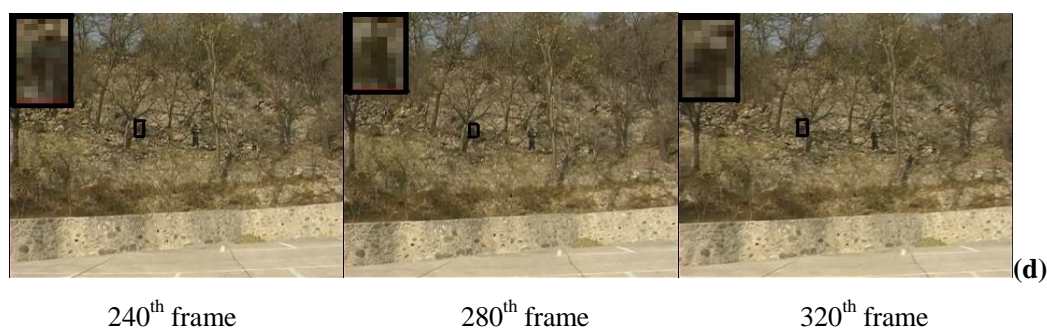


图 3-5 基于 VIVID-egtest01 测试视频的目标跟踪结果对比图，跟踪结果用黑框表示，如果产生跟踪错误，真实的目标位置用白框表示。(a) Variance ratio feature shift based method 的跟踪结果，(b) SIFT-based tracking method 的跟踪结果，(c) traditional Kalman filter method 的跟踪结果，(d) 本节所提出方法的跟踪结果。

第二组实验比较结果如上图所示，仍然对上述提到的四种跟踪算法进行比较。此视频是 SDL 测试视频集，跟踪目标是一个尺度很小的人体，且背景对跟踪目标的干扰很大。不仅背景的颜色和目标的比较类似，而且背景中树的结构与跟踪目标的结构也极其相似，这些都给此运动目标的跟踪带来极大的困难。为了更清晰的呈现跟踪结果，将跟踪结果放大显示在每个图片的左上角。从各个跟踪算法的结果可以清楚地看到，本节所提出的跟踪算法在复杂的背景下仍能取得比较稳定的跟踪结果。Traditional Kalman filter method (图 3-5c)在跟踪的初始阶段便丢掉了跟踪目标，同样，SIFT-based method (图 3-5b)同样在跟踪的初始阶段产生了跟踪错误。相比上述两种跟踪算法，基于特征选择的 Variance ratio feature shift based method 算法 (图 3-5a)取得了相对比较好的跟踪结果，尽管部分视频帧产生了跟踪错误（如 320<sup>th</sup> 视频帧），但是其在大部分的视频帧中都能比较好地对目标进行跟踪。其中跟踪错误的产生在一定程度上是由于 Collins 等所提到的特征选择算法没有考虑特征变化的时间连续性，而是直接通过特征判别力及其相关的变化对特征进行排序和筛选，这就使得背景噪声很容易对跟踪目标的特征描述产生干扰，从而导致部分跟踪结果不正确。

为了定量地评价本节所描述的跟踪算法，DER 被用来量化比较上述四种跟踪算法。在比较实验中，共计算了 10 段视频（表 2-1 所示）的平均 DER 来评价上述四种跟踪算法。DER 的比较结果如下图所示。从图中可以看出，本节所提出的跟踪算法的 DER 不仅比其他三种跟踪算法的都要小，而且 DER 的变化也比较平坦，表明了该算法的稳定性。

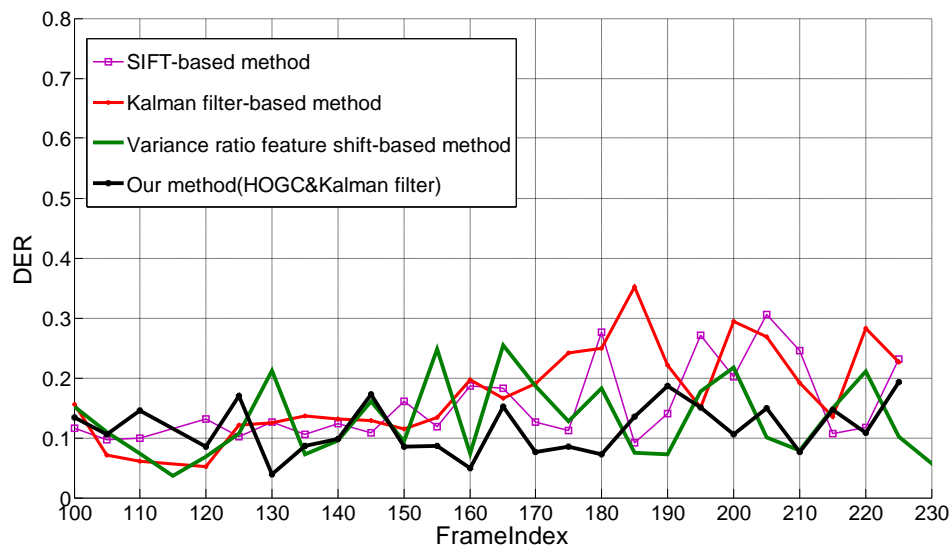
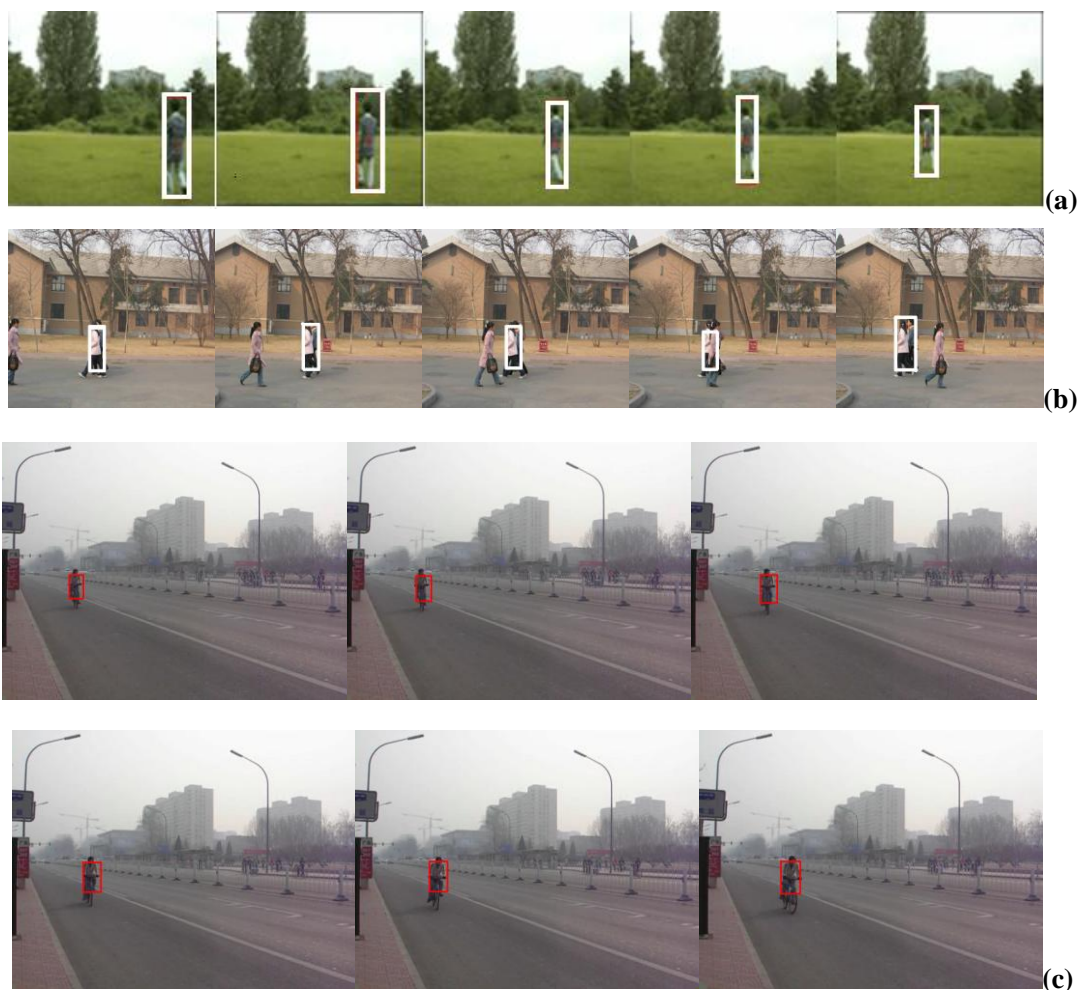


图 3-6 四种跟踪算法的 DER 比较

基于 Kalman 滤波特征权重评估视觉目标跟踪算法在其他测试视频上的跟踪结果如下图所示:



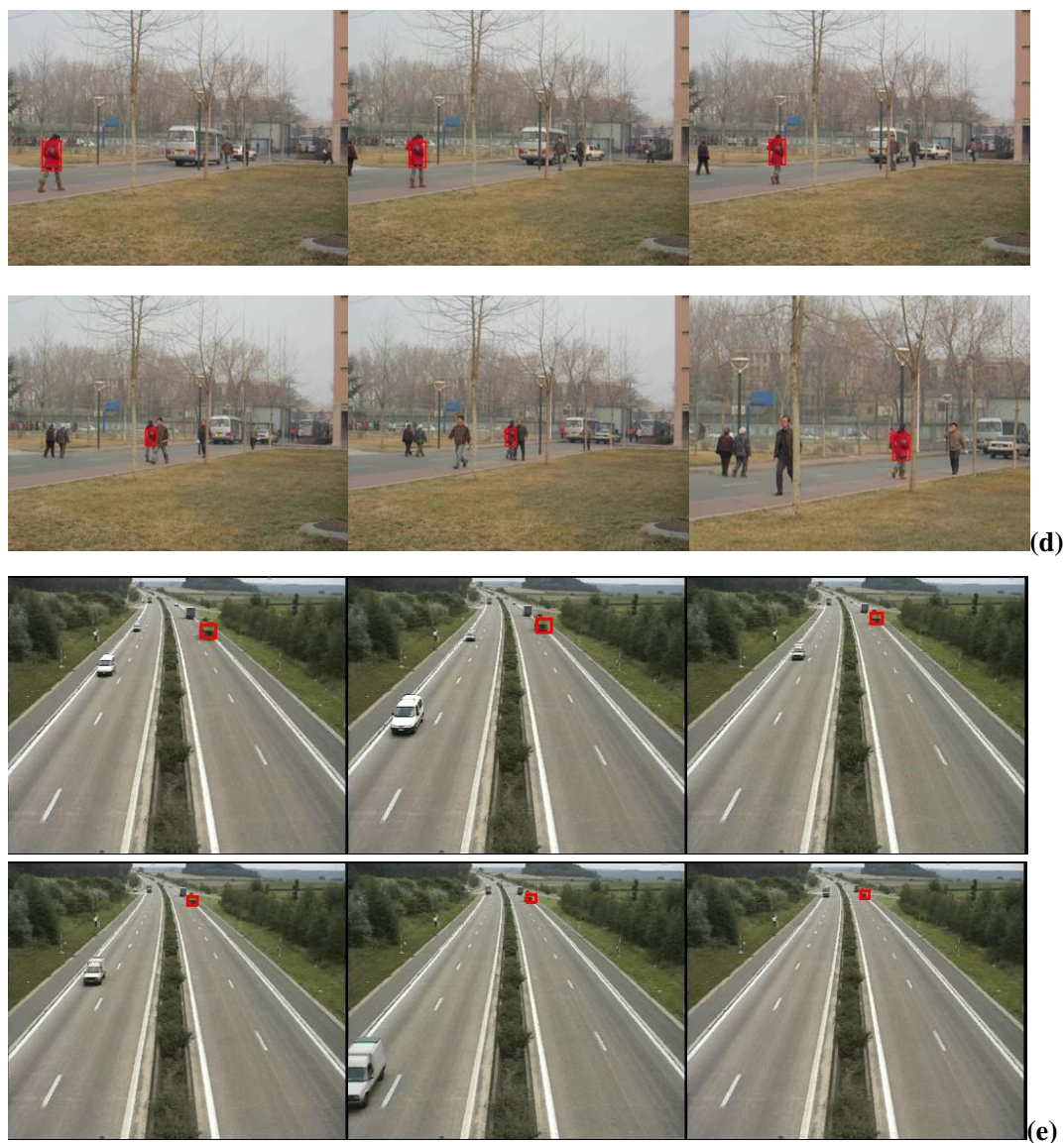


图 3-7 基于 Kalman 滤波特征权重评估的视觉目标跟踪算法的跟踪结果

### 3.2.6 跟踪算法在随动平台系统中的应用

进一步，将本节所提到的跟踪算法嵌入到一个实际的应用测试系统中，其中包括一个 PTZ 和一个视频服务器，从应用的角度分析所提出的跟踪算法在实际应用中的效果。首先通过背景建模[76-78]得到运动的目标区域并对其进行初始化，然后利用 PTZ 随动地对目标进行跟踪以使得 PTZ 能够在目标的整个运动过程中都能观察到被跟踪的目标。系统结构图和跟踪结果如下所示：

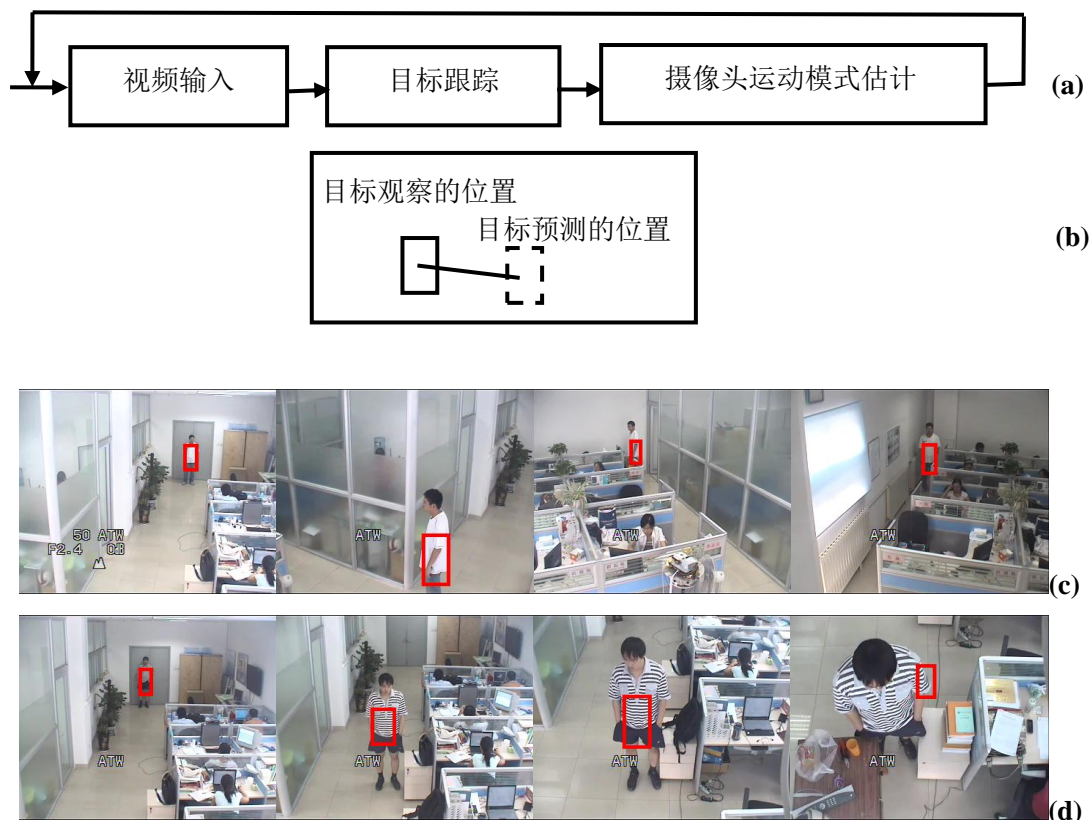


图 3-8 基于 Kalman 滤波特征权重评估视觉目标跟踪算法在基于 PTZ 摄像头系统的应用。(a) 基于 PTZ 的目标跟踪流程图, (b)跟踪目标示意图, (c)正确的跟踪结果和(d)错误的跟踪结果。

### 3.3 基于粒子滤波的综合特征评估的视觉目标跟踪算法

Kalman 滤波通常基于线性和高斯的假设，因此在一些情况下不能很好的预测和修正目标的特征权重，本节又提出了一种基于非线性、非高斯假设的特征权重评估方法，即基于粒子滤波的特征权重评估。

#### 3.3.1 粒子滤波算法的基本原理

粒子滤波 (Particle Filter)，亦称 Bootstrap 滤波、Condensation 算法、序列蒙特卡罗(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。近年来引入视觉跟踪研究领域的序列蒙特卡罗滤波算法与 Kalman 滤波算法一样根植于贝叶斯时序滤波框架。在某种意义上，Kalman 滤波算法可以看作序列蒙特卡罗滤波算法的特例。序列蒙特卡罗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 Kalman 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。

其基本原理如下：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时序滤波器的实现，该算法的基本思想是通过一组带有相关权重的随机样本，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，其中  $N_s$  是采样点数， $\omega_k^i$  是  $k$  时刻第  $i$  个采样点的权重，并且满足  $\sum_i \omega_k^i = 1$ ，则  $k$  时刻后验概率分布  $p(x_k | z_{1:k})$  可近似表示为：

$$p(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^{N_s} \omega_k^i \delta(x_{0:k} - x_{0:k}^i) \quad (3-22)$$

上式中  $\omega_k^i$  可以通过重要性采样 (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  $q(x_k | z_{1:k})$  获得，则第  $i$  个粒子的归一化权重为

$$\omega_k^i \propto \frac{p(x_{0:k}^i | z_{1:k})}{q(x_{0:k}^i | z_{1:k})} \quad (3-23)$$

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

$$\begin{aligned} p(x_{0:k} | z_{1:k}) &= \frac{p(z_k | x_k) p(x_k | x_{k-1})}{p(z_k | z_{1:k-1})} p(x_{0:k-1} | z_{1:k-1}) \\ &\propto p(z_k | x_k) p(x_k | x_{k-1}) p(x_{0:k-1} | z_{1:k-1}) \end{aligned} \quad (3-24)$$

### 3.3.2 粒子滤波算法跟踪

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，利用粒子滤波可以对目标的运动轨迹进行可靠预测[7]，方便在特定区域内搜索目标，等待目标的重新出现。

粒子滤波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，因此我们可以利用粒子滤波器对目标的运动模型进行很好地估计，速度和速度的变化率足以反映目标的运动趋势。通常定义粒子滤波器的系统状态为  $x_k$ ，并且是一个六维向量  $(s_x, s_y, v_x, v_y, \nabla v_x, \nabla v_y)^T$ ，分别代表运动目标在坐标轴上的位置、沿着坐标轴方向的运动速度以及速度的变化率。

粒子滤波算法中，即使在初始化目标的状态参数（如：位置）不准确的情况下，经过迭代运算，仍然可以准确地估计出目标的真实运动状态参数（如：位置），因此，具有预测的稳定性。

### 3.3.3 基于粒子滤波的特征评估

首先将目标的 HOGC 特征集合分为 12 个特征子集，其中包括 3 个颜色子集（R、G 和 B）和 9 个 HOG 子集。然后为每个特征子集定义一个粒子滤波，用来独立地调整目标特征子集中特征的权重。具体定义如下：在  $t$  时刻， $\{p_t(j)\} = \{(j^i, w_t(j^i))\}, j=1,2,\dots,12$ ，代表每个特征子集的 12 个粒子滤波，其中  $j$  表示 R、G、B 颜色子集或者一个 HOG 子集， $j^i$  表示第  $j$  个特征子集（粒子滤波）的第  $i$  维特征， $w_t(j^i)$  表示第  $j^i$  特征所对应的特征权重。

#### 基于粒子滤波的特征权重评估算法

1. 初始化第  $j$  个特征子集（粒子滤波）中每个特征（Particle）相同的权重，其中  $M$  为该特征子集中特征的个数；
2. 对于第  $j$  个特征子集  $\{p_t(j)\}, j=1,2,\dots,12$ ，我们进行如下的操作：

2.1 计算 每个 Particle 的累计权重，如下：

$$\begin{cases} c_t^0 = 0 \\ c_t^i = c_t^{i-1} + w_t(j^i) \end{cases} \quad i=1, 2, \dots, M.$$

得到新的集合  $Set_t(j) = (j^i, w_t(j^i), c_t^i | i=1,2,\dots,M)$ ；

2.2 重采样 从  $Set_t(j)$  重采样选择  $k$  个 Particles, 对于每个重采样得到的粒子

$i'$  按照如下的方法进行选择:

- (a) 从 0 到 1 的均匀分布中, 产生一个随机数  $r, r \in [0,1]$ ;
- (b) 找到最小的  $i$  使得  $c_i^i \geq r$ ;
- (c) 然后  $i' = i$ ;

2.3 预测 通过  $p(i|i')$  预测重采样得到的 Particle  $\{i' | i' = 1, 2, \dots, M\}$  的权重。在

实验中  $p(i|i')$  被定义为  $p(i|i') \propto N(0, \delta)$ , 其中  $\delta$  被设定为  $2M$ ;

2.4 更新 通过  $p(S_{t+1}^i | i)$  来更新第  $i$  个特征 (Particle) 的权重  $w_{t+1}(j^i)$ ; 在实验

中,  $p(S_{t+1}^i | i)$  定义为相应特征的判别力;

2.5 归一化  $\sum_i w_{t+1}(j^i) = 1$ 。

- 3. 在跟踪过程中, 搜索目标的最适位置;
- 4.  $t = t+1$ ; 跳转到步骤 2 或者结束循环。

以某一 HOG 特征子集为例, 基于粒子滤波的特征权重更新过程如下图所示:

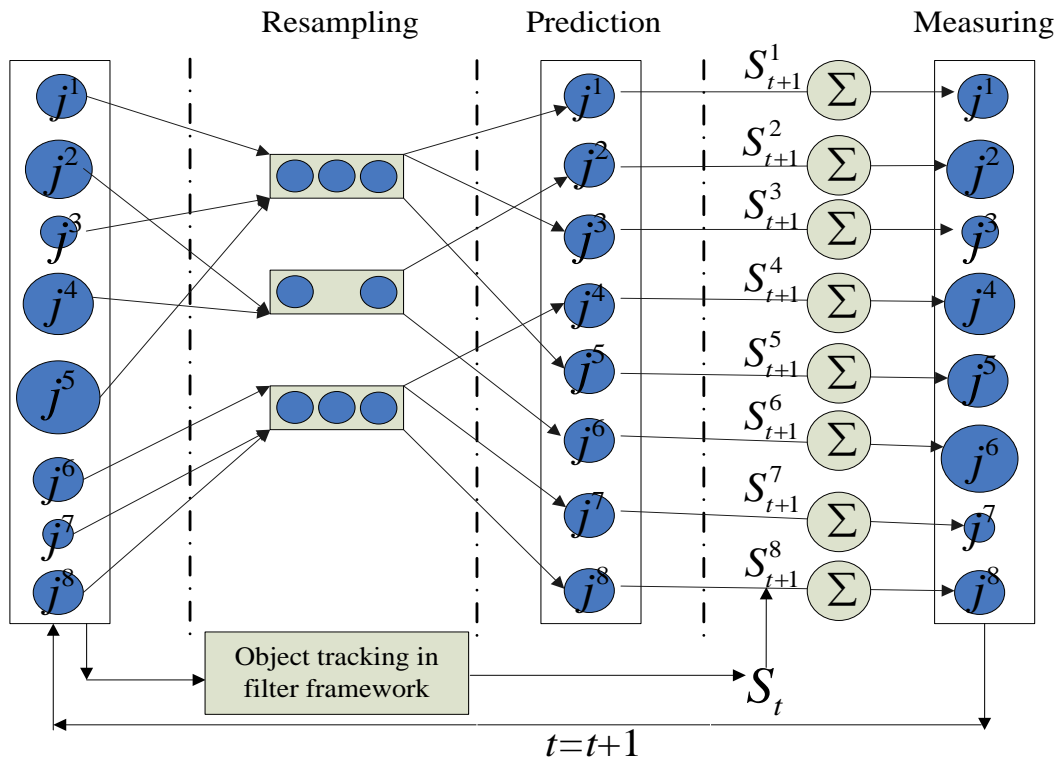


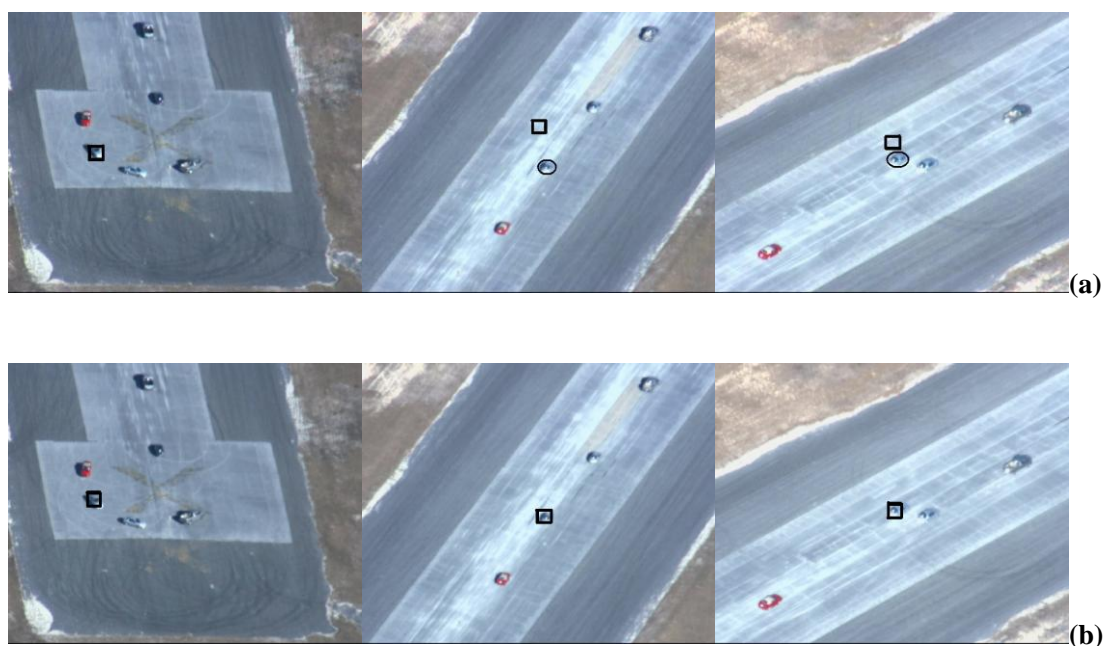
图 3-9 基于粒子滤波的特征权重评估示意图

### 3.3.4 基于粒子滤波特征评估的视觉目标跟踪

同样,基于 3.2.4 节中定义的公式(3-20)与(3-21)对搜索区域 $\Omega$ 进行目标跟踪,获取最终的跟踪结果。

### 3.3.5 实验结果比较与算法性能分析

第一段测试视频来自于 VIVID 测试数据集,其中跟踪目标为一辆和背景颜色极为相似的车辆,视频同样是由直升机从空中拍摄所得。比较的算法包括本节提出基于粒子滤波特征评估的视觉目标跟踪算法以及另外三种传统的跟踪算法:基于特征选择的跟踪方法 (Variance ratio feature shift based method) [11]、基于 SIFT 特征点的跟踪方法 (SIFT-Based method) [74]和传统的粒子滤波跟踪方法 (traditional Particle Filter method) [7]。实验结果如下图所示,其中 Variance ratio feature shift based method 在第 700 和 1350 帧中产生了跟踪错误,由于 Variance ratio feature shift based method 只采用了颜色特征对目标进行跟踪,所以当有背景颜色对其干扰时,其往往会产生较差的跟踪结果;由于背景比较的平滑,所以不会产生大量的 SIFT 噪声点对目标的描述进行干扰,因此 SIFT-Based method 取得了比较满意的跟踪结果;同时,由于跟踪目标与背景之间的相对运动比较简单,所以 traditional Particle Filter method 能够得到很好的预测和修正目标的运动模型,准确地得到目标在视频帧中的位置。



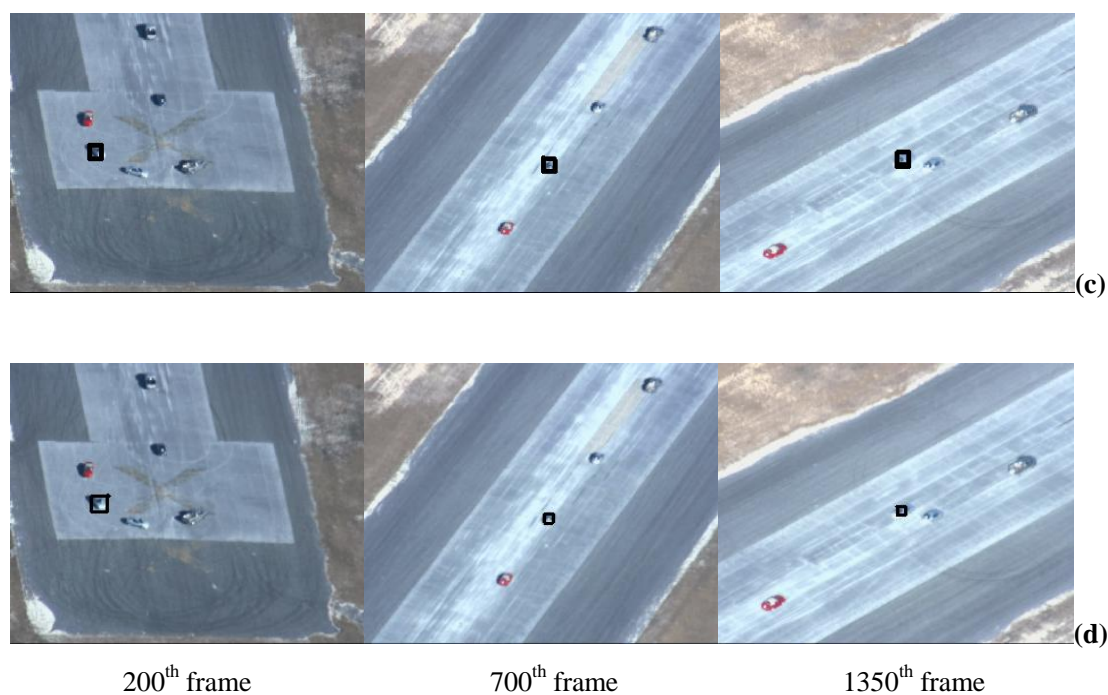


图 3-10 基于 VIVID-egtest01 测试视频的目标跟踪结果对比图。跟踪结果用黑框表示，真实结果用黑色的椭圆表示。(a) Variance ratio feature shift based method 的跟踪结果，(b) SIFT-based tracking method 的跟踪结果，(c) traditional Particle filter method 的跟踪结果和(d) 我们所提出跟踪算法的跟踪结果。

在下图所示的实验结果比较中，跟踪目标为人体，其中跟踪目标与其他目标之间存在着遮挡且目标本身也具有尺度变化。从实验结果可以看出，所提出的跟踪算法仍然取得了最好的跟踪结果。traditional Particle Filter method 方法也取得了较为令人满意的跟踪结果，但是基于 SIFT 的跟踪算法(图 3-11b)产生了一些跟踪错误，这是由于在背景中存在着相似的人体，对跟踪目标 SIFT 特征点的提取会产生很大的干扰，从而使得跟踪结果不是特别稳定。Variance ratio feature shift based method 虽然也会根据背景的变化来自动进行特征选择，但是由于自适应特征选择机制的缺陷使得当背景存在噪声干扰时，不能很好地跟踪目标，因此在第 829 帧中产生了跟踪错误。





图 3-11 存在遮挡和相似目标情况下的跟踪结果对比图。(a)Variance ratio feature shift based method 的跟踪结果, (b) SIFT-based method 的跟踪结果, (c) traditional Particle filter method 的跟踪结果和(d)本节所提出算法的跟踪结果。

下面仍然沿用上章中所定义的 DER 来定量地比较本章中所提到的四种跟踪算法的整体性能, 测试视频仍然选用为上章中的 10 段视频, 对比结果如下:

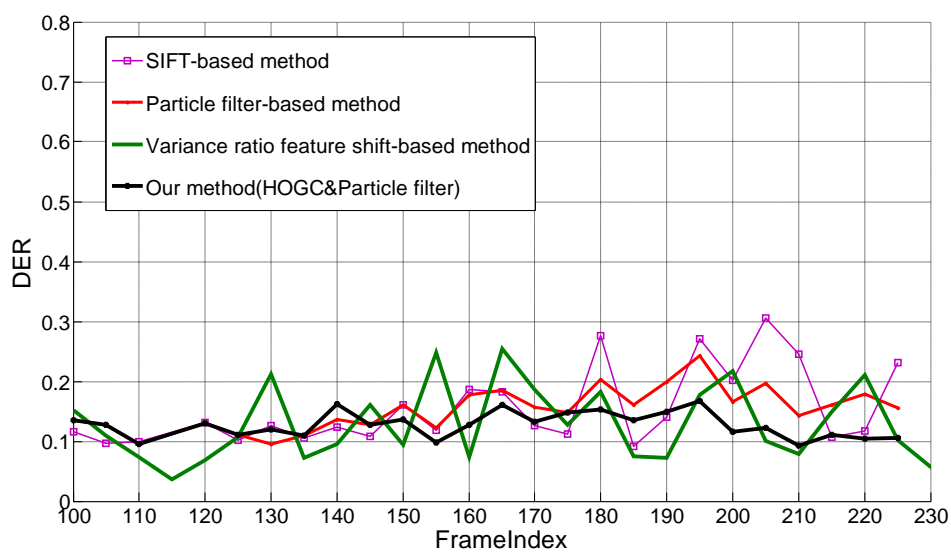
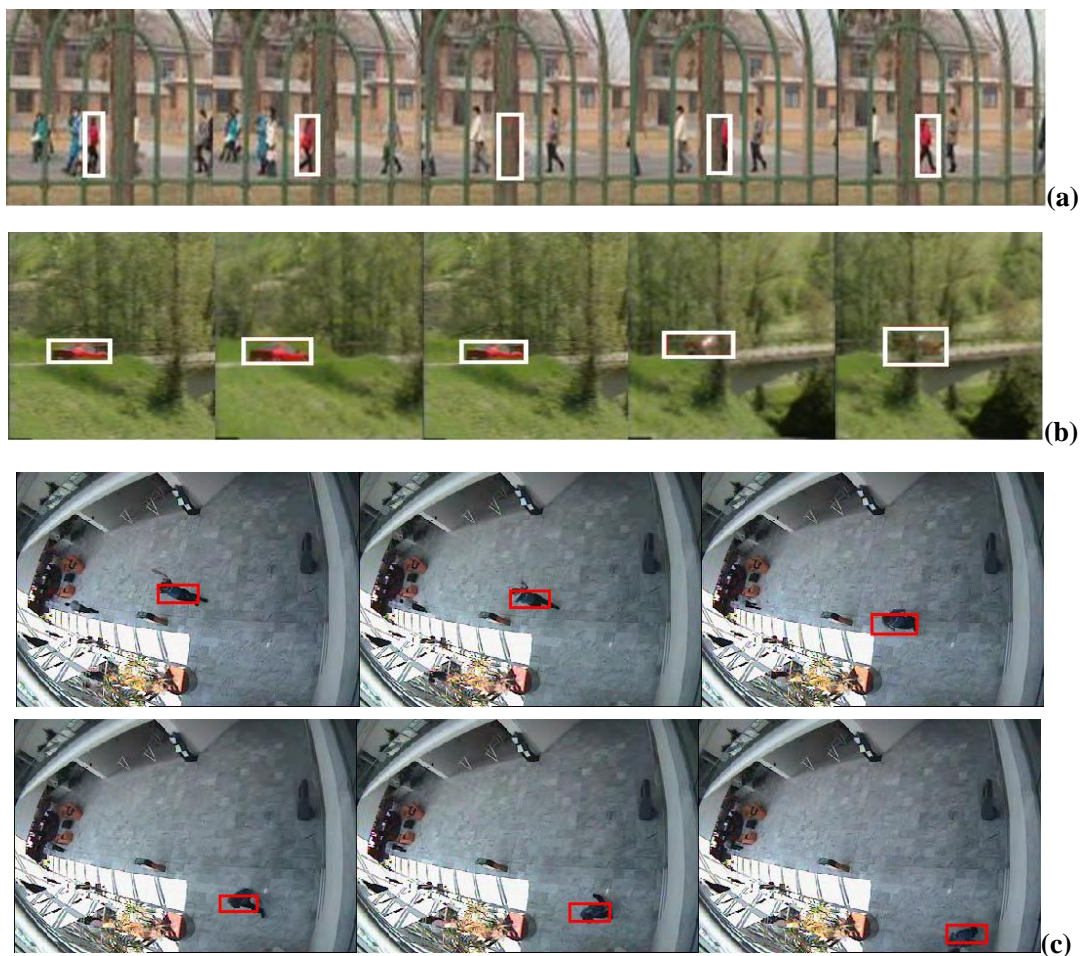


图 3-12 四种跟踪算法的 DER 对比结果图

基于粒子滤波特征评估的视觉目标跟踪算法在其他测试视频上的跟踪结果如下所示：



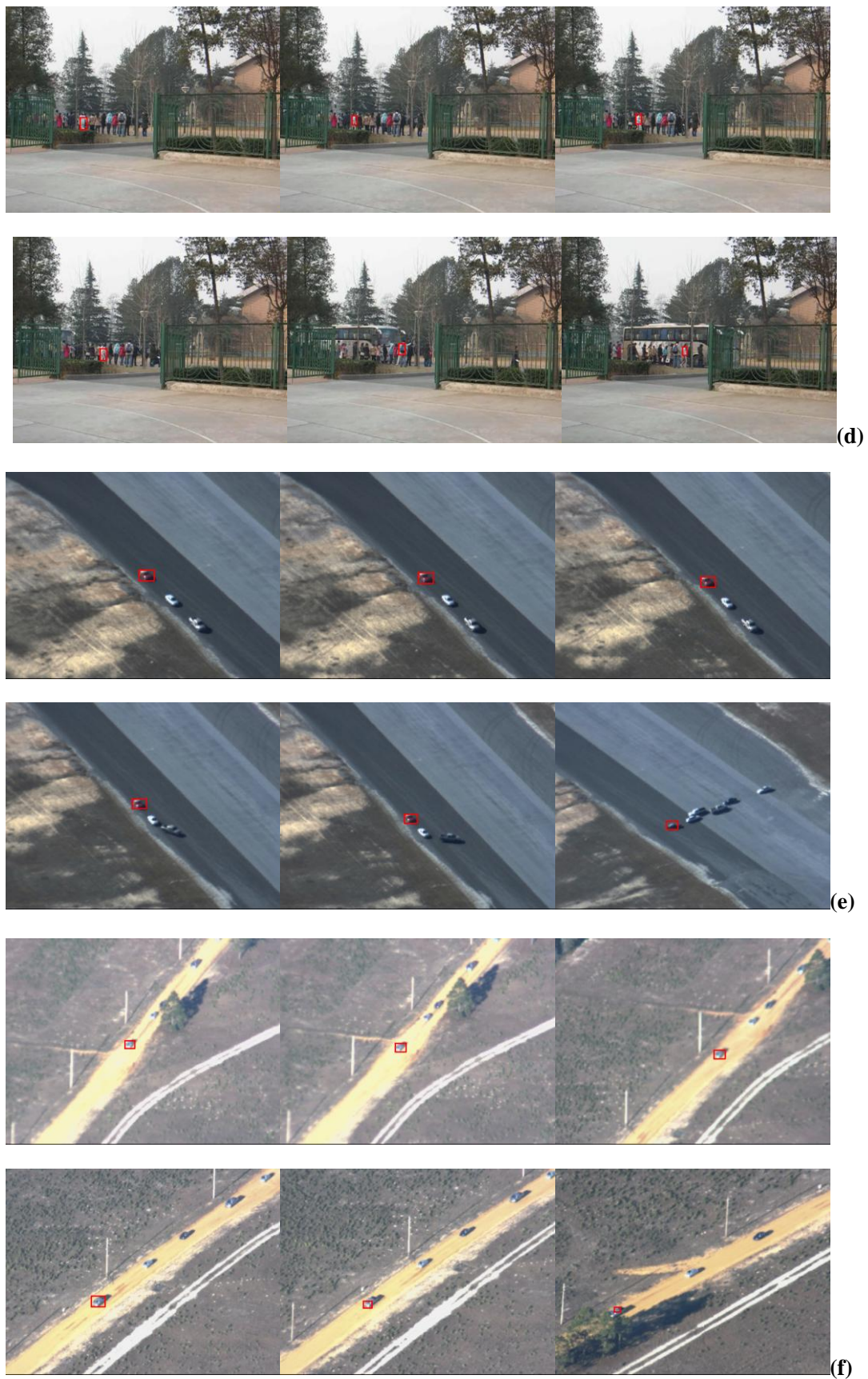


图 3-13 基于粒子滤波框架特征权重评估跟踪算法的跟踪实验结果

3.4 基于 Kalman 与粒子滤波的综合特征评估的视觉目标跟踪算法性能比较

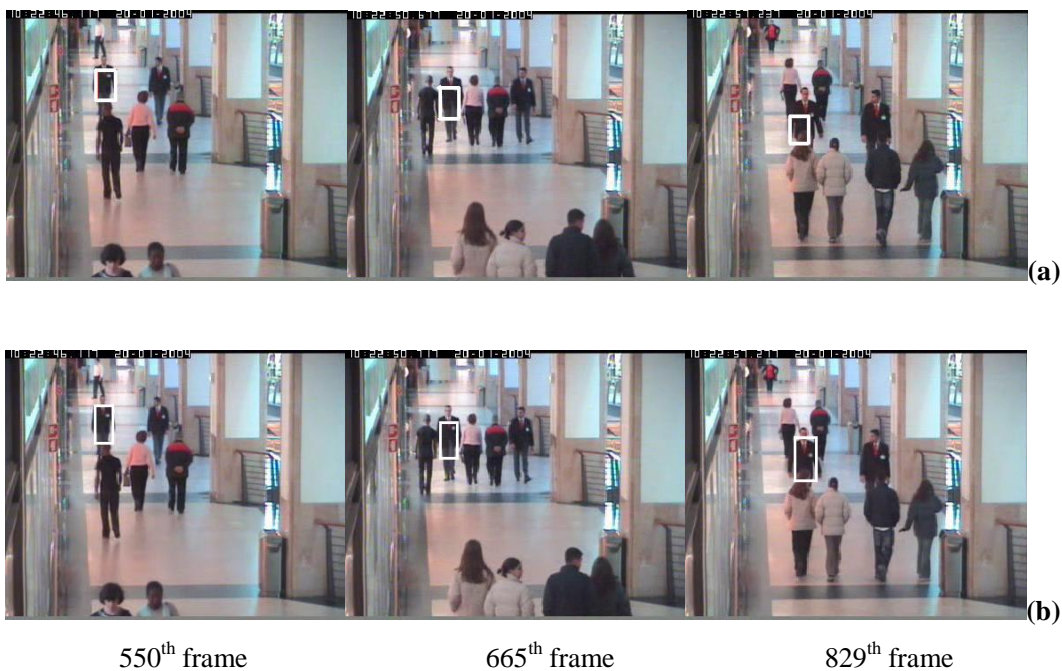
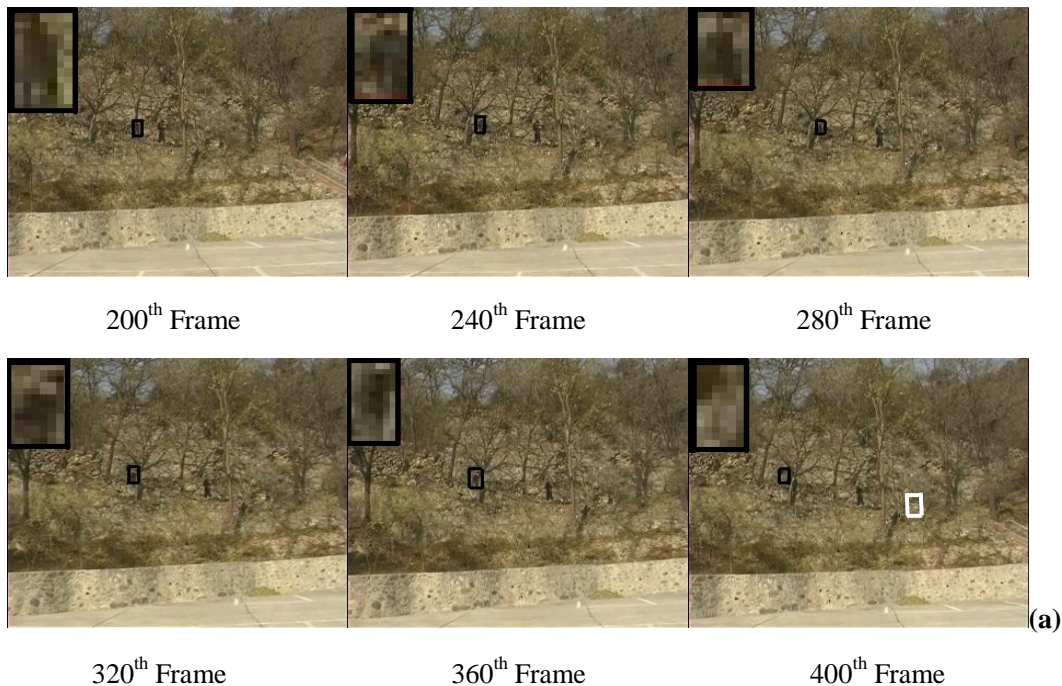


图 3-14 存在遮挡和相似目标的跟踪结果对比图。(a) 基于 Kalman 滤波特征评估的视觉目标跟踪算法跟踪结果，(b) 基于粒子滤波特征评估的视觉目标跟踪算法跟踪结果。





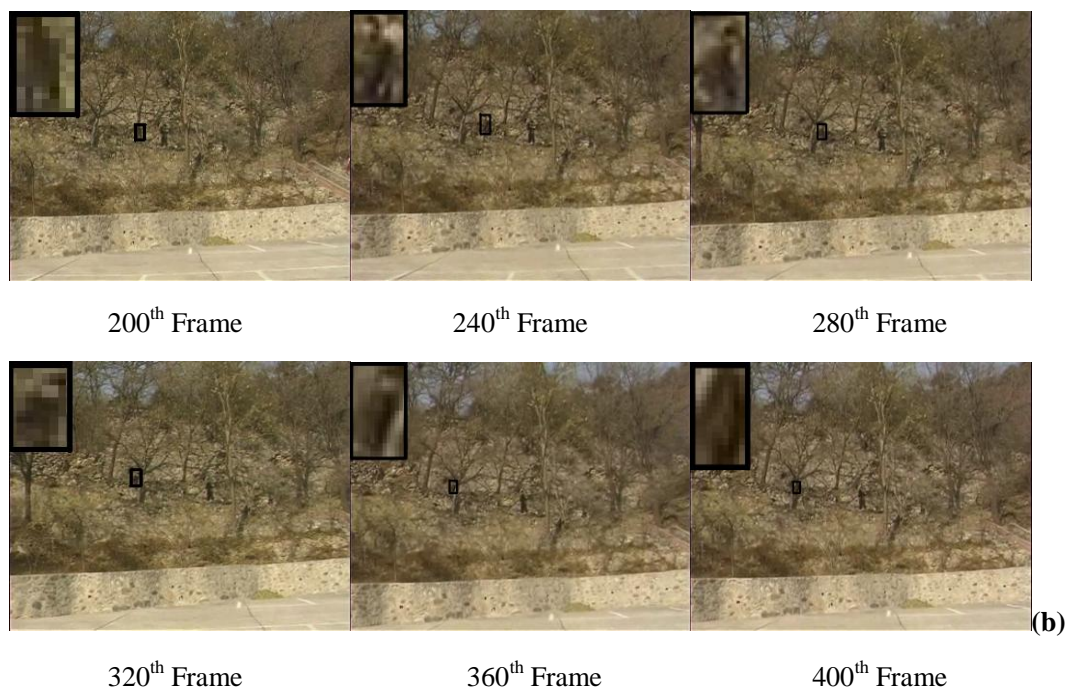


图 3-15 存在遮挡和相似目标的跟踪结果对比图。(a) 基于 Kalman 滤波特征评估的视觉目标跟踪算法跟踪结果，(b) 基于粒子滤波特征评估的视觉目标跟踪算法跟踪结果

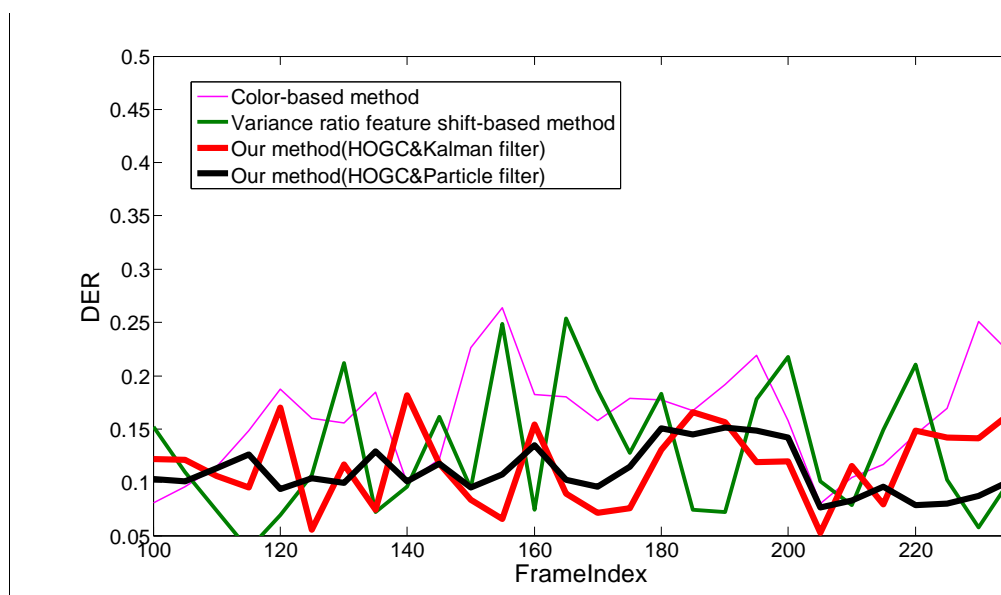


图 3-16 五种跟踪算法（包括基于 Kalman 滤波特征评估的视觉目标跟踪算法与基于粒子滤波特征评估的视觉目标跟踪算法）的 DER 比较

从以上的实验结果的分析可以得出，基于粒子滤波特征评估的视觉目标跟踪算法比基于 Kalman 滤波特征评估的视觉目标跟踪算法具有更好的稳定性，验证了目标特征的变化不完全局限于高斯或线性形式的假设。

### 3.5 小结

本章提出了基于时间连续性的特征权重评估方法。在每帧视频中，通过当前视频帧中的跟踪目标特征区分目标本身与相关背景的能力以及先前视频帧中跟踪目标特征的权重对特征的有效性进行重新评估。

本章分别在 Kalman 滤波框架和粒子滤波框架中进行实验验证，给出了该跟踪算法在各种复杂场景下的运动目标跟踪实验，从而对算法在运动目标尺度伸缩变化、前景/背景动态变化等情况下的鲁棒性跟踪进行验证，同时将该算法与传统经典的基于特征选择的跟踪算法进行了比较并对比较结果进行了定量的分析。实验结果表明，在整体性能上，本章提出的基于时间连续性的特征权重评估方法要优于领域内经典的基于特征选择的跟踪算法。

## 第四章 基于自适应稀疏表示的视觉目标跟踪算法

*“Many could be better than all”*

最新关于人类视觉系统（Human Vision System, HVS）的研究表明[79]，人类的视觉神经元细胞中只有一小部分对外界环境的刺激，例如颜色、纹理、形状以及尺度变化等，比较活跃。考虑到人类具有数以亿计的视觉神经元细胞，因此可以合理的推断活跃的神经元细胞及其稀疏。基于此最新研究，计算机视觉领域内的相关学者提出了目标的稀疏表示，并将此成功的应用到了人脸识别方面[80]。同时，其他领域的相关研究也表明[81-83]：基于稀疏性原则，利用样本子集比样本全集能更有效的描述一个物体。

受到目标稀疏表示的启发，在本章中，提出了一种基于目标自适应稀疏表示的跟踪算法。首先，建立一个动态更新的在线训练样本集合；其次，基于样本集合，对跟踪目标进行实时的稀疏特征选择，即稀疏表示；最后，基于稀疏表示，对目标进行跟踪，同时在滤波框架中调整目标的稀疏特征的权重，获得具有时间连续性的自适应稀疏表示。基于自适应稀疏表示的目标跟踪算法流程图，如下所示：

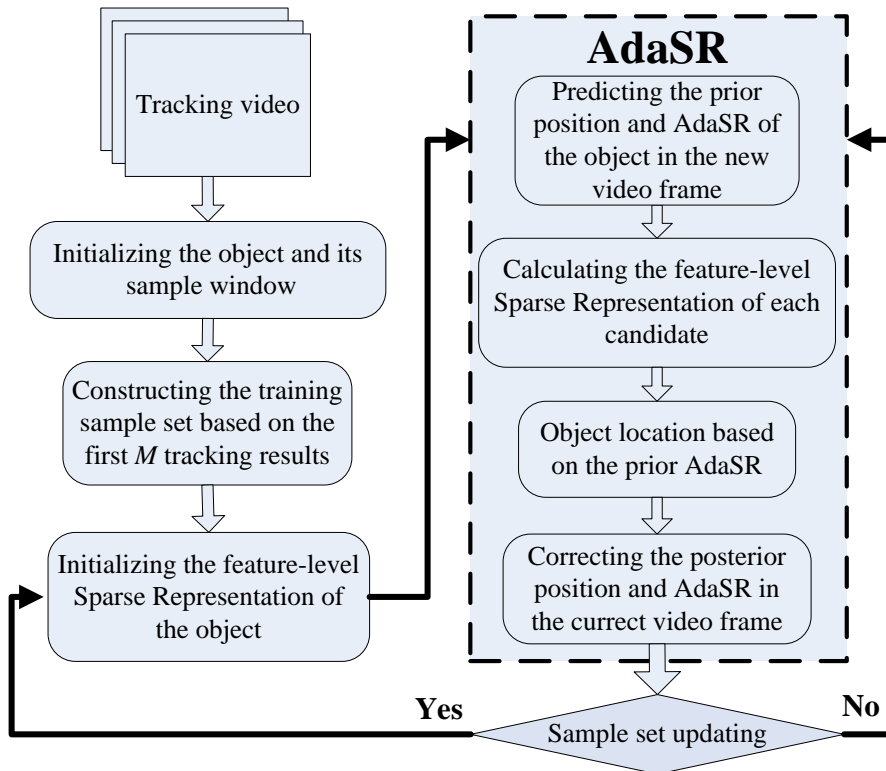


图 4-1 基于自适应稀疏表示的目标跟踪算法流程图

自适应稀疏表示 (Adaptive Sparse Representation, AdaSR) 的优点在于, 基于 L1-norm 最小化求解得到的特征稀疏系数能够有效的区分前景及背景, 同时基于滤波框架的特征评估, 又能够保证特征权重变化是一个连续的过程。因此, AdaSR 既具有特征区分性, 又能保证特征的时序一致性。

## 4.1 稀疏表示简单介绍

### 4.1.1 L1-norm 研究综述

范数是一种距离度量。L1-norm 是  $l_p$  中的一种, 常见的范数还有 L2-norm, L0-norm 等。某一向量的 L1-norm 是指向量中每一分量的绝对值求和。在实际应用中, L2-norm 一直都受到学者的青睐。针对 L1-norm 的研究早在范数理论创建的时候, 许多数学爱好者就对其性质进行了分析[84], 但是由于 L1-norm 连续但非光滑的性质使其求解算法比较复杂, 其一直未受到重视。直到压缩感知理论在信号处理领域中的出现并普及, L1-norm 终于因其优良的性质而得到研究者的青睐。压缩感知理论中最早考虑使用 L0-norm 去重构信号, 但由于 L0-norm 的求解是 NP 难度问题, 所以实际求解中使用 L1-norm 来近似 L0-norm。

### 4.1.2 稀疏表示基本原理

早在 1993-1997 年[85-87], 一些基于稀疏表示的研究成果已经相继发表。首先是 Chen 提出使用 L1-norm 最小化方法建立优化模型解决回归问题, 并使用基追逐法去求解。后来非常著名的 Lasso[85]算法也发表在统计学期刊上。Lasso 算法是针对回归问题, 建立最小二乘误差和 L1-norm 正则化的约束模型, 并提出使用贪婪算法去求解。但是, 在当时这个模型的真正意义并没有引起很多人的注意, 大多数人也没有意识到 L1-norm 和稀疏性之间的联系。

随着信号处理领域压缩感知理论[88-91]的提出, 稀疏表示受到学者们越来越多的重视。以往指导信号采样的理论基础一直是著名的 Nyquist 采样定理[88], 它指出只有当采样速率达到信号频率的两倍以上时, 才能由采样信号精确重建原始信号。但是, 随着信号的频率变得越来越高, 根据 Nyquist 采样定理, 那么相应的信号采样速率必须越来越高。现在, 针对信号传输速度的高要求和存储空间有限等问题, 常见的解决方案是对信号压缩。以那么高的速率对信号采样, 然后

再对信号实施压缩变换，这样极大的浪费了资源，因为高速率采样后的数据在压缩过程中只保留很少的信息，抛弃了很多冗余信息。从这个角度考虑，研究者们思考是否存在另一种基于信息采样的理论框架，使得采样的同时能实现信号的压缩，并且只需远少于 Nyquist 采样定理所要求的采样数目就可精确或近似精确重建原始信号，于是诞生了压缩感知采样理论。

压缩感知(compressed sensing)或压缩采样(compressive sampling)的某些抽象结论源于 Kashin 创立的范函分析和逼近论[92]，后来由 Candès、Tao、Romberg、Donoho 等人系统化地创立。该理论否认了 Nyquist 采样定理，也即信号采样速率不再取决于信号频率，并且指出采样速率在很大程度上取决于两个要素，信号本身的稀疏性和非相干性。换句话说，该理论主要论述：当信号在某个变换域是稀疏的或可压缩的，可以利用与变换矩阵非相干的测量矩阵将变换系数线性投影为低维观测向量，同时这种投影保持了信号重建时所需的信息，通过进一步求解稀疏优化问题就能够从低维观测向量精确地或高概率精确地重建原始信号。

压缩感知理论主要涉及变换域中的过完备字典、满足非相干性的测量矩阵和快速鲁棒的信号重建三个问题。这三个问题的核心都是信号是否可以进行稀疏表示，如果能进行稀疏表示，应该使用何种模型去表示，包括模型的解法等一系列问题。压缩感知最初的稀疏性重构模型，使用的是 L0-norm。某一个向量  $x$  的 L0-norm  $\|x\|_0$  指向量  $x$  中不为 0 的元素的个数。

$$\begin{aligned} \min_x \quad & \|x\|_0 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (4-1)$$

上述的 L0-norm 优化模型中， $A$  为观测矩阵， $x$  为原始信号经过稀疏变换后的表示系数， $y$  是接受端对信号进行重构之后的结果。关于模型(4-1)，其解的存在以及是否唯一，是值得考究的。在 2003 年，Donoho & Elad [93]等学者证明，如果矩阵  $A$  满足： $\sigma(A) \geq 2\|x\|_0$ ，那么 L0-norm 最优化问题具有唯一解。 $\sigma(A)$  指矩阵  $A$  中最小的线性相关的列向量集所含的向量个数。然而，虽然证明出模型的解是存在唯一的，但是上述优化模型的求解过程是 NP 难问题。

随着对稀疏表示问题的深入研究，在 2006 年 Tao 和 Candes[94-95]合作证明了在限制等距性质(Restricted Isometry Property, RIP)条件下[92]，L0-norm 优化

问题与模型(4-2)中 L1-norm 优化问题具有相同的解,其中 RIP 条件指稀疏信号在观测矩阵作用下必须保持几何性质相一致。

$$\begin{aligned} \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (4-2)$$

在 RIP 性质提出之后,稀疏表示终于在理论上和实践中得到统一。后来很多研究学者也注意到稀疏表示的优点,将其用于模式识别和计算机视觉等领域中。目前稀疏表示的实现主要有三种方式:直接优化法,迭代收缩方法和  $l_p$  范数正则化方法[96]。

**直接优化法:**指直接使用 L0-norm 最小化作为稀疏性追求的目标函数,优化模型与(4-1)优化模型类似。这类优化问题的求解一般都使用贪婪算法去求解,典型的代表算法有匹配追逐法[87]和正交匹配追逐法[97]等。

**迭代收缩优化方法:**指所提到的著名 LASSO 方法,该方法可以通过不断松弛模型中的  $\varepsilon$  因子,来实现迭代收缩求解。

$$\begin{aligned} \min_x \quad & \|Ax - y_2\|_2 \\ \text{s.t.} \quad & |x_1| \leq \varepsilon \end{aligned} \quad (4-3)$$

**$l_p$  范数正则化优化方法:**目前研究比较热门,  $l_p$  范数,当  $p=1$  时,即为所谈到的 L1-norm。因为 Candes 等人理论上证明 L1-norm 的稀疏性,所以现在关于 L1-norm 的研究比较多。当  $0 < p < 1$  时,有些研究者表明此时的  $l_p$  范数比 L1-norm 更加有利于实现稀疏性[118]。目前  $l_p$  范数正则化的常见求解方法是使用基追逐[98]方法,软阈值方法(常见的有坐标梯度法和 Bregman 迭代法)等。

近几年来,稀疏表示(Sparse Representation)是信号处理、机器学习和模式识别等领域中的一个热点研究问题。从近几年的模式识别与计算机视觉领域的各种会议期刊文献中,便可看出稀疏表示这一问题是属于非常前沿的研究问题,并且 ICCV'09 与 CVPR'10 中分别开设了一系列的专题报告。下面本节围绕稀疏表示对自适应目标特征表示进行紧致的特征选择,从而实现目标跟踪。

## 4.2 基于自适应稀疏表示的视觉目标跟踪

### 4.2.1 在线训练样本集合的构造及更新

1、在线训练样本集初始化：在样本集的初始化构建过程中，利用前  $M$  帧的跟踪结果及其背景，分别构造样本集合中的正例  $\ell_j^+$  与反例  $\ell_j^-$ ，通过 HOGC 特征的提取，获得构建  $2M$  个训练样本  $training\_set = \{(\ell_j^+, \ell_j^-)\}, j=1\dots M$ 。

2、在线训练样本集实时更新：在跟踪过程中，对训练样本序列进行实时更新，即将最新的跟踪正例和反例样本替换样本集合中样本。在本节中，随机从样本集中挑选一组样本  $(\ell_j^+, \ell_j^-)$  对其进行替换。通过样本集合中个别样本的逐步更新过程，可以分析得到由于每次只更新  $1/M$  的样本序列，即使是一次较差甚至是错误的样本更新过程对整个样本集合的影响也是非常小的，这样就能有效避免了跟踪过程中经常产生的模板漂移问题[103]，保证了跟踪的稳定性。

### 4.2.2 稀疏特征的选择

基于在线训练样本集，通过 L1-norm 最小化的方法对目标特征进行稀疏选择，这样既保证了对特征稀疏选择，又保证了选择的结果具有很好的前景/背景区分力。这个过程可以优化建模如下：

$$\begin{aligned} & \min \|S\|_1 \\ \text{s.t.} & \begin{cases} S^T \ell_j^+ \geq \alpha \\ S^T \ell_j^- \leq -\alpha \end{cases}, \end{aligned} \quad (4-4)$$

其中， $\|\bullet\|_1$  代表 L1-norm；约束条件保证训练样本能够被正确分类； $S \in R^N$  是特征的前景背景区分能力； $\ell_j^+ \in R^N$  是训练样本集的第  $j$  个正例样本； $\ell_j^- \in R^N$  是训练样本集的第  $j$  个反例样本； $\alpha$  保证了正例样本和反例样本之间的距离最少为  $2\alpha$ 。

由于这个优化模型较难直接解决，需要引入向量来有效地解决这个问题。定义  $\eta \in R^N$ ， $\mu \in R^N$  且  $\eta \geq 0, \mu \geq 0$ 。同时令  $\eta^i = (S^i)_+$ ， $\mu^i = (-S^i)_+$ ， $i=1,2,\dots,N$ ， $i$  表示特征集合的维数， $(\bullet)_+$  被定义为取正运算符  $(S^i)_+ = \max\{0, S^i\}$ 。然后便得到

$S = \eta - \mu$  以及  $\|S\|_1 = I_N^T \eta + I_N^T \mu$ ,  $I_N = [1, 1, \dots, 1]^T$ 。公式 (4-4) 便可被重新写为下面的凸优化模型:

$$\begin{aligned} & \min I_N^T \eta + I_N^T \mu \\ & s.t. \begin{cases} (\eta - \mu)^T \ell_j^+ \geq \alpha \\ (\eta - \mu)^T \ell_j^- \leq -\alpha \\ \eta \geq 0 \\ \mu \geq 0 \end{cases} \end{aligned} \quad (4-5)$$

经过稀疏特征选择之后的特征如下图所示:

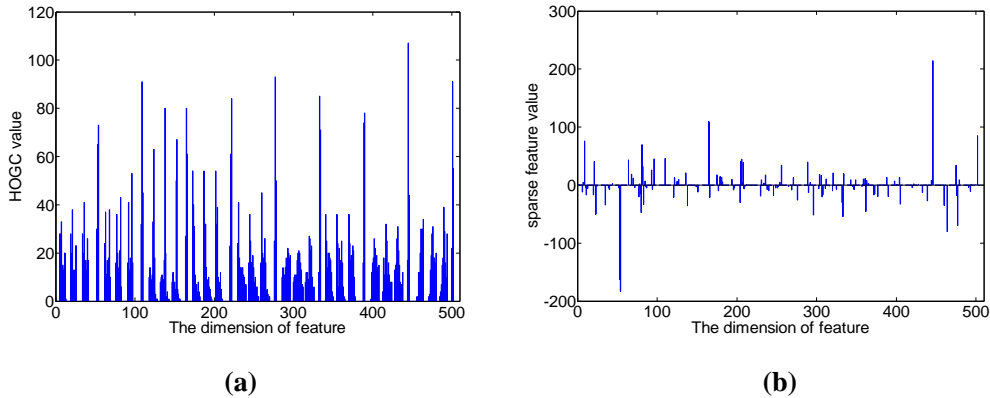


图 4-2 HOGC 特征与稀疏 HOGC 特征对比图。(a) HOGC 特征, (b) 稀疏特征选择之后的 HOGC 特征。

#### 4.2.3 自适应稀疏表示的计算

为了能够使得稀疏表示能够满足时间连续性, 对其在滤波框架中进行连续的评估。Kalman 滤波的状态定义为: 目标的稀疏表示及其位置; Kalman 滤波的观察定义为: 跟踪结果的稀疏表示及其跟踪结果。Kalman 滤波的状态预测以及观测方程如下所示:



$$\begin{cases} \begin{pmatrix} \overline{w}_{t+1} \\ \Delta \overline{w}_{t+1} \\ \overline{Pos}_{t+1} \\ \Delta \overline{Pos}_{t+1} \end{pmatrix} = \begin{pmatrix} I_{N \times N} & I_{N \times N} & 0 & 0 \\ 0 & I_{N \times N} & 0 & 0 \\ 0 & 0 & I_{L \times L} & I_{L \times L} \\ 0 & 0 & 0 & I_{L \times L} \end{pmatrix} \begin{pmatrix} w_t \\ \Delta w_t \\ Pos_t \\ \Delta Pos_t \end{pmatrix} + u_t \\ \\ \begin{pmatrix} S_t \\ mPos_t \end{pmatrix} = \begin{pmatrix} I_{N \times N} & 0 & 0 & 0 \\ 0 & 0 & I_{L \times L} & 0 \end{pmatrix} \begin{pmatrix} w_t \\ \Delta w_t \\ Pos_t \\ \Delta Pos_t \end{pmatrix} + v_t \end{cases} \quad (4-6)$$

其中,  $w_t = \{w_t(i) | i = 1, N\}$  表示  $t$  图像帧中目标的后验稀疏综合特征, 并且  $\Delta w_t = w_t - w_{t-1}$ 。  $S_t$  表示  $t$  图像帧中稀疏系数向量。  $Pos_t$  表示  $t$  图像帧中目标的后验位置, 并且  $\Delta Pos_t = Pos_t - Pos_{t-1}$ 。  $mPos_t$  表示  $t$  图像帧中跟踪结果的位置。  $u_t$  和  $v_t$  为高斯噪声。  $I_{N \times N}$  和  $I_{L \times L}$  为单位矩阵。

#### 4.2.4 基于自适应稀疏特征的目标跟踪

目标跟踪是对跟踪目标在搜索区域  $\Omega_t$  内进行搜索的过程。跟踪的目的为在  $\Omega_t$  内找到与跟踪目标最匹配的跟踪结果  $(x, y)_t$  通过最大化  $p((x, y)_t | \Omega_t, (x, y)_{t-1})$ 。我们通过贝叶斯推理过程来求  $p((x, y)_t | \Omega_t, (x, y)_{t-1})$  的最优解, 如下所示:

$$\begin{aligned} & \max_{(x, y)_t} p((x, y)_t | \Omega_t, (x, y)_{t-1}) \\ & = \max_{(x, y)_t} \int p(F_t((x, y)_t) | (x, y)_{t-1}, p(x, y) | \Omega_t) dc \quad (4-7) \\ & = \max_{(x, y)_t} \int \left( \sum_{i=1}^N (p(F_t^i((x, y)_t) | F_{t-1}^i)) p(x, y) | \Omega_t) dc \right) \end{aligned}$$

其中,  $(x, y)_t^c$  定义为  $\Omega_t$  中第  $c$  个候选目标的位置,  $p(F_t^i((x, y)_t^c), F_{t-1}^i)$  表示为  $F_t^i((x, y)_t^c), F_{t-1}^i$  这两个特征的相似度,  $p(F_t^i)$  表示为  $t$  时刻第  $i$  维特征的置信度, 此处将其视为  $t$  时刻第  $i$  维特征的稀疏系数  $w_t(i)$ , 因此公式(4-7)可以如下描述:

$$\begin{aligned} & \max_{(x, y)_t} p((x, y)_t | \Omega_t, (x, y)_{t-1}) \\ & = \max_{(x, y)_t} \int \left( \sum_{i=1}^N (w_t(i) p(F_t^i((x, y)_t^c), F_{t-1}^i)) \right) p((x, y)_t^c | \Omega_t) dc \quad (4-8) \end{aligned}$$

### 4.3 实验结果比较与算法性能分析

#### 4.3.1 基于自适应稀疏表示目标跟踪算法的有效性验证

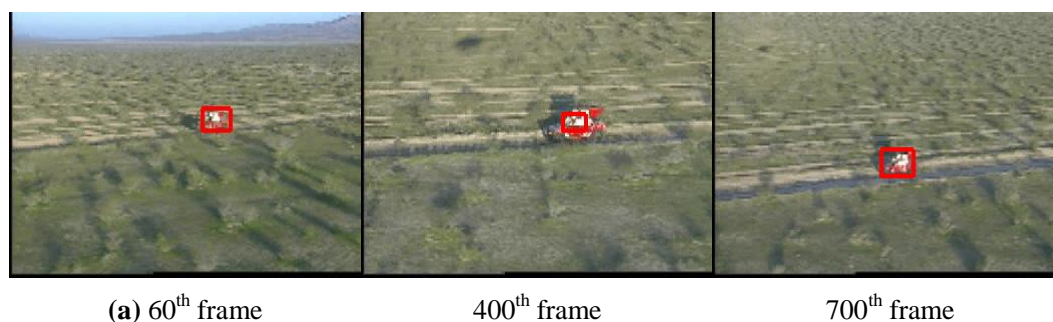
第一段测试视频来自于 VIVID 测试数据集，跟踪目标为一辆运动模式相对简单的吉普汽车，跟踪背景变化相对稳定。同时，由于跟踪目标与背景比较明显的特征差别（目标主要为白色和红色，背景主要为灰色与绿色），此段测试视频的主要难点在于跟踪目标的尺度变化较大。

第二段跟踪测试视频仍然来自于 VIVID 数据测试集合，测试视频中的跟踪目标为一辆在比较相似背景中运动的车辆。首先，跟踪目标做了一个 180° 的调头（第 50, 280 和 450 帧视频），然后加速直行（第 700 和 1000 帧视频），直到超过另外 3 辆车（第 1360, 1470 和 1670 帧视频）。在整个跟踪过程中，跟踪目标具有尺度变化，旋转，同时，在第 1360 帧的背景中有一辆与跟踪目标几乎完全相似的车辆，因此导致这段测试视频具有一定的挑战性。同时，跟踪目标与背景的颜色相似性对跟踪的稳定性也具有较大的影响。

第三段测试视频来自于 CARVIA 数据测试集合，此测试视频最大的挑战主要来自于跟踪目标与其他相似目标的部分遮挡问题（第 600 和 829 帧视频）以及跟踪目标由小到大的尺度变换。

第四段测试视频来自于 SDL 数据测试集合，此视频的主要挑战是跟踪目标频繁的被背景中的其他人体或者树遮挡（第 350, 400, 500 和 810 帧视频），并且这些遮挡物在形状以及颜色等外观表现方面和跟踪目标比较相似（第 600 和 810 帧视频）。

跟踪目标自适应的稀疏表示选择特征集合中最具有前景/背景区分特性的特子集来描述跟踪目标，这些子集中的特征在具有区分性的基础上，同时由于在滤波框架中的滤波调整，其也能够较好的自适应于前景/背景的特征变化。图 4-3 中的目标跟踪结果表明了跟踪算法的有效性。





1000<sup>th</sup> frame

1300<sup>th</sup> frame

1800<sup>th</sup> frame



(b) 50<sup>th</sup> frame

280<sup>th</sup> frame

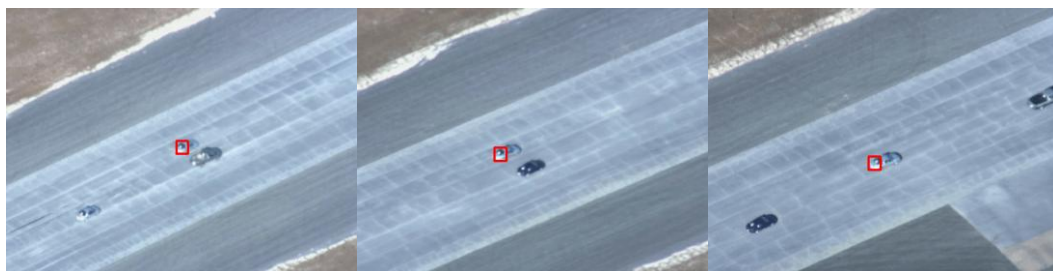
450<sup>th</sup> frame



700<sup>th</sup> frame

1000<sup>th</sup> frame

1360<sup>th</sup> frame



1470<sup>th</sup> frame

1670<sup>th</sup> frame

1800<sup>th</sup> frame



(c) 250<sup>th</sup> frame

350<sup>th</sup> frame

400<sup>th</sup> frame

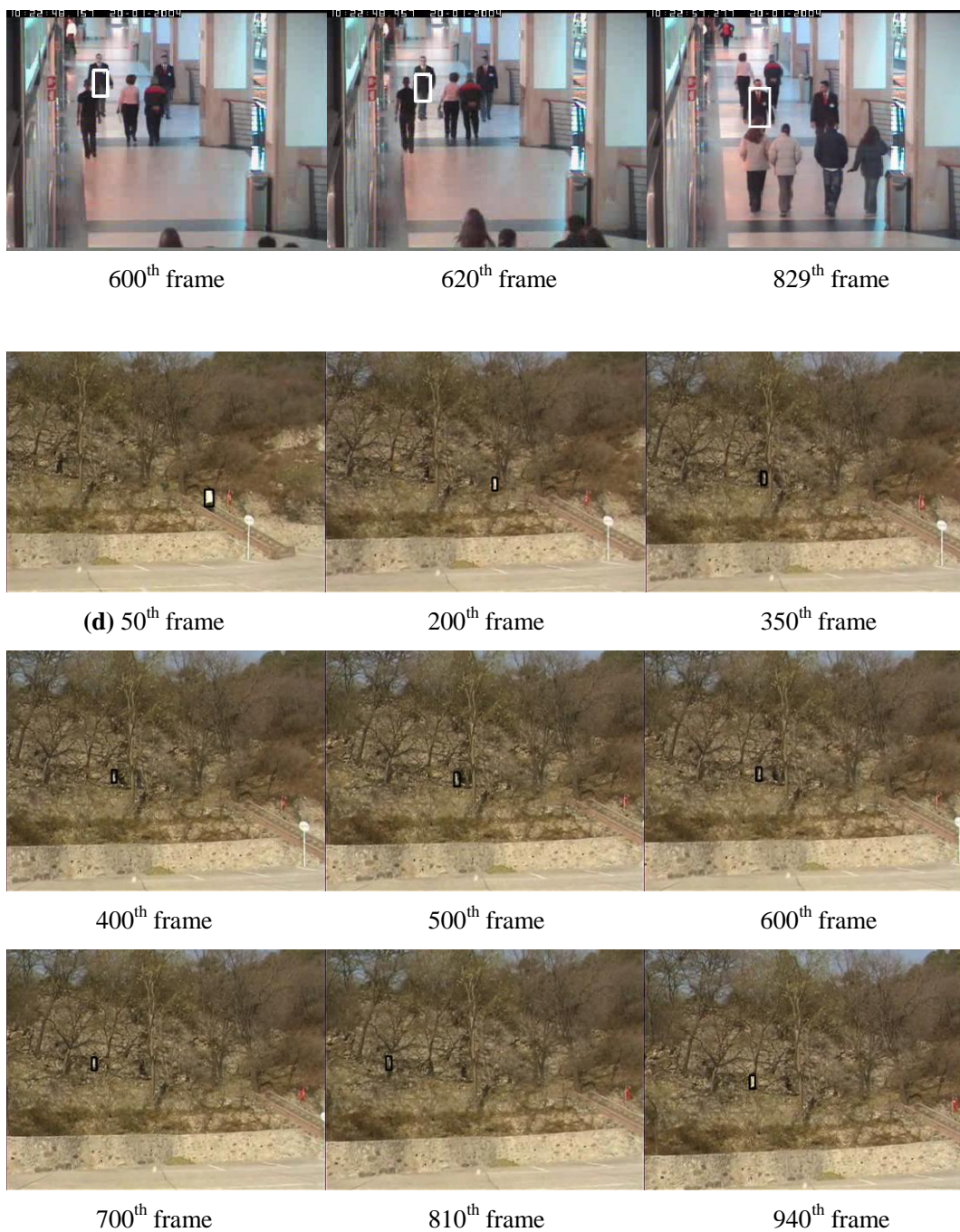


图 4-3 四段测试视频的跟踪结果实例

### 4.3.2 算法性能比较

进一步，为了更加直观地体现基于自适应稀疏表示目标跟踪算法的高性能，本节中我们与经典的特征选择/评估算法(Variance ratio feature shift [11], and Peak difference feature shift tracking methods [11])进行比较。图 4-4 所示，基于自适应稀疏表示目标跟踪算法的统计 DER 的均值与方差要明显小于经典的特征选择/

评估算法，表明自适应稀疏表示的特征选择与评估方法对于目标的表示描述更好，能够更好的区分前景/背景以及适应前景/背景的变化。

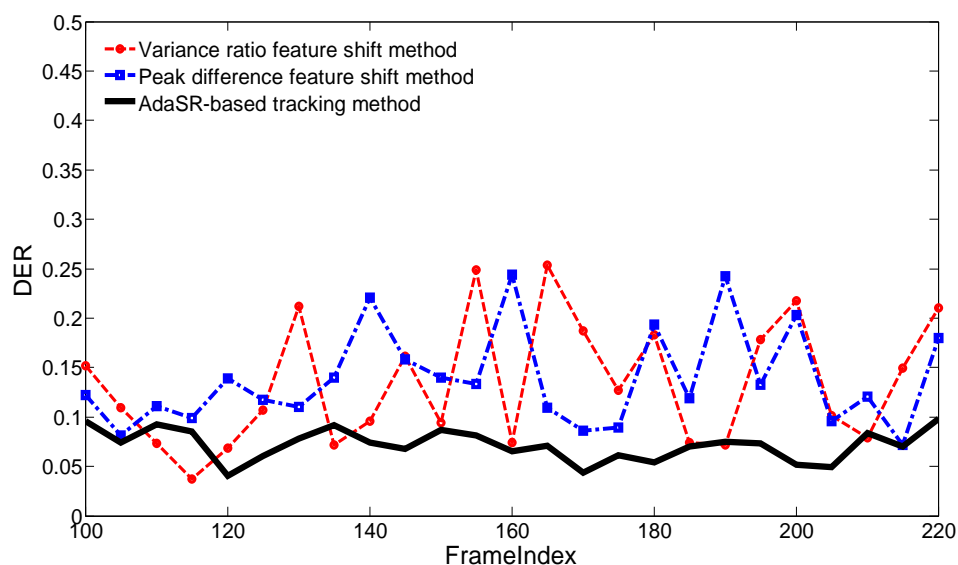


图 4-4 三种跟踪算法的 DER 比较

#### 4.4 小结

本章在自适应综合特征表示的基础上，基于压缩感知理论在计算机视觉领域内的最新研究成果，提出了基于自适应稀疏特征表示的视觉目标跟踪算法。稀疏特征表示使得跟踪目标的表示既紧致又具有很好的前景/背景区分性。同时在滤波框架中对选择的特征进行自适应的评估，进一步保证跟踪目标的自适应稀疏表示具有时间连续性。

本章给出了该跟踪算法在各种复杂场景下的运动目标跟踪实验，从而对算法在运动目标尺度伸缩变化、形变、前景/背景动态变化等情况下的鲁棒性跟踪进行了验证，同时将该算法与传统的跟踪算法的性能进行了比较与分析，验证了所提跟踪算法在综合性能上更加高效。

## 第五章 基于分层自适应稀疏表示与稀疏重构的视觉目标跟踪框架

*“Many could be better than one”*

从感知角度出发,目标跟踪的本质在于在视频帧的搜索区域中找到相应的跟踪目标,所以可以合理的假设跟踪目标可以通过搜索区域内的跟踪样本进行线性重构得到。同时,在跟踪过程中,视频帧中的搜索区域往往要比跟踪目标大很多,这样便会导致样本线性重构的系数非常稀疏,此时,搜索区域中的背景内的样本(重构反例样本)所对应的系数往往会趋向于0。即使在目标被遮挡的情况下,只有部分重构反例样本和部分重构正例样本(样本来自于目标区域,比如目标的一部分或者全部)所对应的系数有效(非零)。相比较于传统的在特征空间内对整个目标进行描述的跟踪方法(分块描述其实是将跟踪目标的每个分块看成是一个单独的跟踪子目标),基于线性重构的目标跟踪算法优势在于在样本空间内对目标进行重构,将基于样本空间的重构系数看做是目标基于样本空间的特征,此特征直接描述目标与各个样本的重构程度。在目标发生遮挡时,基于样本空间的特征具有更好的描述性能,因此能够得到更好的跟踪性能(如图5-1所示)。

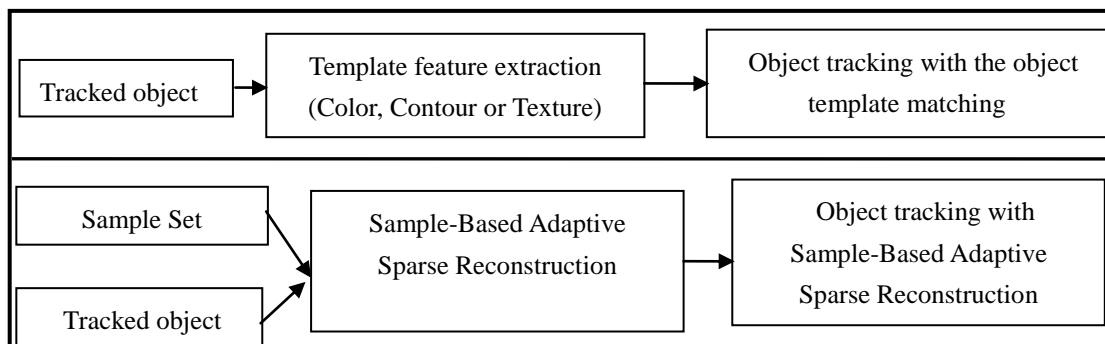


图 5-1 模板匹配与基于样本重构的目标跟踪算法对比图(图中第一行为基于目标模板匹配的跟踪算法流程,第二行为基于样本重构的目标跟踪算法流程)

### 5.1 基于分层自适应稀疏特征与稀疏重构的视觉目标跟踪

本节将目标的自适应稀疏表示与稀疏重构结合起来,提出了一种新的目标跟踪框架,旨在对目标的表示以及搜索等跟踪中的关键问题从稀疏的角度进行统一的描述和分析。首先,在跟踪的过程中根据搜索区域动态的建立重构样本集;其次,通过提取目标以及重构样本集中样本的自适应稀疏表示,建立自适应的稀疏

重构模型；最后，根据基于样本空间的重构系数，获得目标在样本空间内的分布模型，从而获得目标的跟踪结果。基于分层自适应稀疏特征与稀疏重构的视觉目标跟踪算法（Adaptive Visual Object Tracking via Hierarchical Sparse Representation and Reconstruction, HSRR）流程图如下所示：

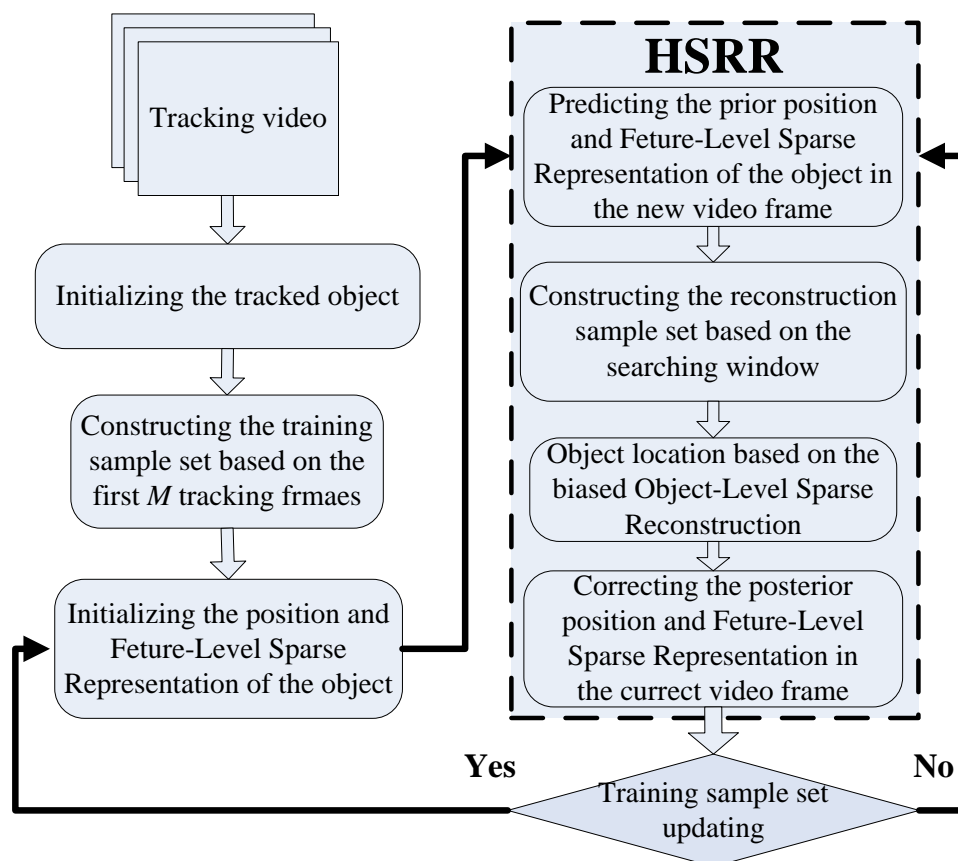


图 5-2 基于分层自适应稀疏特征与稀疏重构的视觉目标跟踪算法流程图

### 5.1.1 目标、背景以及搜索区域的定义



图 5-3 跟踪目标（黑色）、背景（蓝色与黑色之间）及其搜索区域定义（红色）

如上图所示，目标区域定义为 $h \times w$ 的矩形框，背景区域定义为 $\sqrt{2}h \times \sqrt{2}w$ 的矩形框与目标区域之间的环形区域，搜索区域为 $W \times H$ 的矩形区域，其中 $W = m \times w$ 且 $H = m \times h$ ， $m=3$ 或者4。

### 5.1.2 重构样本集合的构造

基于跟踪搜索区域，重构样本集合被动态的建立（如图 5-4 所示）。样本集合中的样本定义为搜索区域内的一个子窗口，描述为 $r = (x, y, s)$ ，其中 $0 < x < W$ ， $0 < y < H$ 且 $s > 0$ 。基于此定义，样本集合的构造过程如下：

- 1、样本的位置 $(x, y)$ 沿着水平和垂直方向行进，每次移动步长为 $d$ 个像素（在本章算法中 $d = 4$ ）；
- 2、样本的尺度 $s$ 为从 $\{0.8, 0.9, 1.0, 1.1, 1.2\}$ 集合内均匀采样得到，当样本的位置确定之后。

通过如上所定义的构造过程，可以得到具有 $K$ 个样本的重构样本集合 $Reconstruction\_set = A = \{a^k, k=1 \dots K\}$ ，其中样本集合中的大部分样本来至于搜索区域内跟目标无关的背景区域，其余样本都至少与目标有部分重叠（如图 5-4c 所示）。

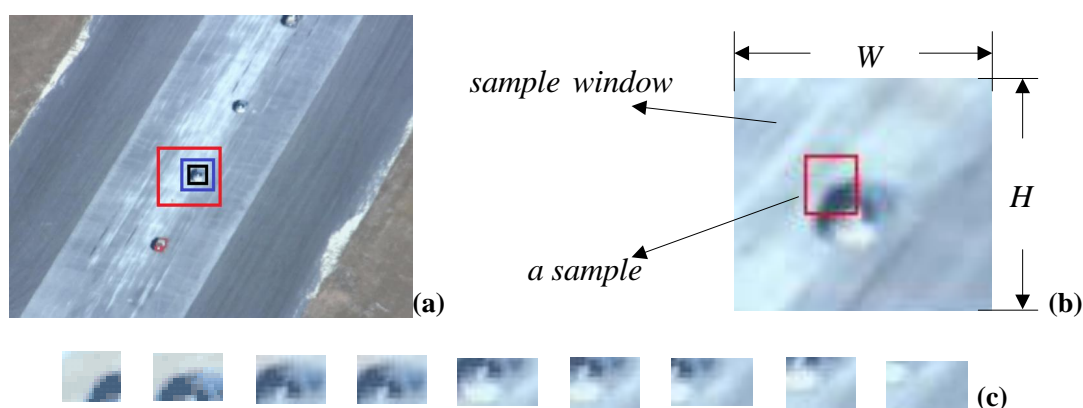


图 5-4 重构样本集合的构造。(a)搜索区域（样本区域），(b)一个样本，(c)重构样本集合中的样例。

### 5.1.3 基于重构样本集合的自适应稀疏重构

当获取重构样本集合 $A$ 之后，线性重构模型可以定义如下：



$$A\psi \approx F, \quad (5-1)$$

其中,  $F$  为跟踪目标的 HOGC 特征集合,  $\psi = \{\psi^k, k=1, \dots, K\}$  基于  $A$  和  $F$  的线性重构系数向量,  $\psi^k$  为第  $k$  个样本所对应的重构系数。在跟踪过程中, 考虑到前景/背景的区分性以及特征变化, 因此对目标及样本集合中的样本提取其自适应的稀疏表示, 线性重构模型可以修改为如下自适应模型:

$$g(A)\psi \approx g(F), \quad (5-2)$$

其中  $g(*)$  表示为自适应稀疏特征选择过程 (详细过程已在第四章中介绍), 其具体描述如下:

$$\begin{aligned} g(*) &= w_i F = (w_i(1) \times F(1), w_i(2) \times F(2), \dots, w_i(N) \times F(N)) & \text{if } * \text{ is a vector} \\ g(*) &= (w_i a^1, w_i a^2, \dots, w_i a^K) & \text{if } * \text{ is a matrix} \end{aligned} \quad (5-3)$$

在实际的跟踪过程中, 视频帧中的搜索区域往往要比跟踪目标大很多, 这样便会导致样本自适应线性重构的系数会非常的稀疏, 因为重构样本集合中有大量的反例样本 (来自于搜索区域内和目标无关的区域), 导致重构系数向量中大量的系数为 0 (重构反例样本所对应的系数往往会趋向于 0)。即使在目标被局部遮挡的情况下, 仍然只会有部分重构反例样本和部分重构正例样本 (样本来自于目标区域, 例如目标的一部分) 所对应的系数有效 (非零), 此时大部分系数仍然为 0, 整个向量仍然非常稀疏。假设在  $\psi = \{\psi^k, k=1 \dots K\}$  中有  $r$  个稀疏为非 0, 可以合理的推理得到  $r \ll K$ , 此时, 定义样本集合能够  $r$ -阶稀疏重构跟踪目标。原则上, 非 0 系数的个数一般描述为  $\|\psi\|_0$ , 但是最小化  $\|\psi\|_0$  是一个 NP 难度问题。基于压缩感知的最新研究表明, 在限制等距性质 (Restricted Isometry Property, RIP) 条件下,  $\|\psi\|_0$  与  $\|\psi\|_1$  具有相同的解。基于样本集合  $A$ , 自适应稀疏线性重构模型可以定义如下:

$$\begin{aligned} \min & \|\psi\|_1 \\ \text{s.t.} & g(A)\psi \approx g(F) \end{aligned}, \quad (5-4)$$

假设  $g(\phi) = g(A)\xi$ , 并且  $\xi = (\eta^T, \mu^T)^T$ ,  $\eta \in \mathbb{R}^N$ ,  $\mu \in \mathbb{R}^N$ ,  $\psi = \eta - \mu$ ,  $\eta \geq 0, \mu \geq 0$ 。然后可以得到  $g(A)(\eta - \mu) = g(F)$ ,  $(g(A), -g(A))(\eta^T, \mu^T)^T = g(F)$  或者  $g(\phi)\xi = g(F)$ ,  $\xi \geq 0$ 。因此最小化  $\|\psi\|_1$  等价

于最小化  $I_N^T \xi$ ,  $I_N = [1, 1, 1, \dots, 1]^T$  是一个  $N$  维的单位向量。通过上述变换, 公式(5-4)可以等价的转换为一个标准的线性规划模型, 如下所示:

$$\begin{aligned} & \min I_N^T \xi, \\ & s.t. \begin{cases} g(\phi) \xi = g(F) \\ \xi \geq 0 \end{cases} \end{aligned} \quad (5-5)$$

通过求解公式 5-5, 可以计算得到  $r$ -阶的自适应稀疏重构向量( $r \ll K$ )。如图 5-5 所示, 给出了一次求解得到的自适应稀疏重构向量的实例, 其中重构样本集合包含 100 个重构样本, 而其中非 0 的系数所对应的有效样本大约为 10 个左右, 显示出结果的稀疏性。稀疏重构的特性能够保证跟踪目标被最相似的样本优先紧致重构, 也就是说, 稀疏重构过程能够选择得到与跟踪目标最为相近的重构样本子集。

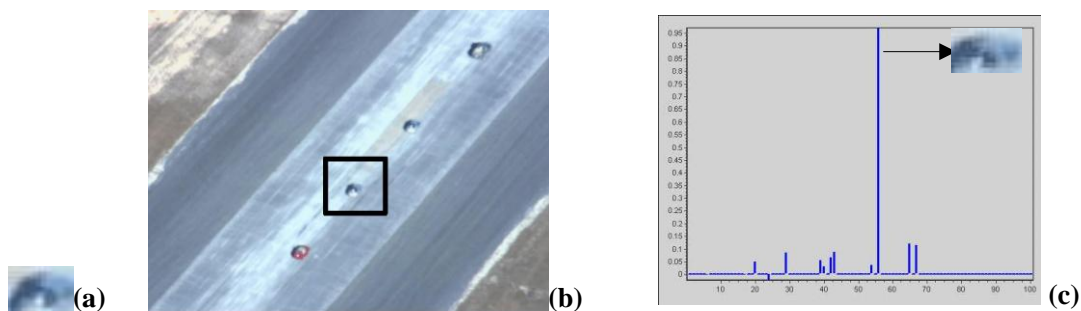


图 5-5 稀疏重构向量示意图。(a) 跟踪目标, (b) 搜索区域, (c) 稀疏重构向量。

#### 5.1.4 搜索区域的有效性验证

搜索区域的有效性验证是目标跟踪中的另一个非常重要的研究问题。传统的验证方法主要通过计算跟踪结果与初始化跟踪目标之间的残差来判断搜索区域是否有效。在[80]的研究中表明, 在搜索区域的有效性验证方面, 基于稀疏系数的稀疏性统计结果比残差具有更好的效果。因此, 在本章所提出的跟踪算法中, 利用基于稀疏系数向量的稀疏性统计结果来判断搜索区域是否有效。假设  $t$  时刻的稀疏系数向量为  $\psi_t$ , 其稀疏性比率 (Sparsity Ratio,  $SR$ ) 定义如下:

$$SR(\psi_t) = \frac{\sum_{k=1}^K \delta(\psi_t^k)}{K}, \quad (5-6)$$

$$\delta(\psi_t^k) = \begin{cases} 1, & \text{if } |\psi_t^k| > 0.1 \\ 0, & \text{otherwise} \end{cases}. \quad (5-7)$$

其中，通过  $SR$  的定义可以，当  $SR(\psi_t) \approx 1$  时，整个集合中的每一个样本基本上都对跟踪目标的重构起到作用，往往这样的情况只发生在搜索区域里不包含跟踪目标或者包含非常多和跟踪目标非常相似的伪目标。因此通过定义的阈值  $\tau$ ，如果  $SR(\psi_t) \leq \tau$  则我们认为搜索区域是有效的，否则我们认为其无效，搜索区域的验证流程如图 5-6 所示：

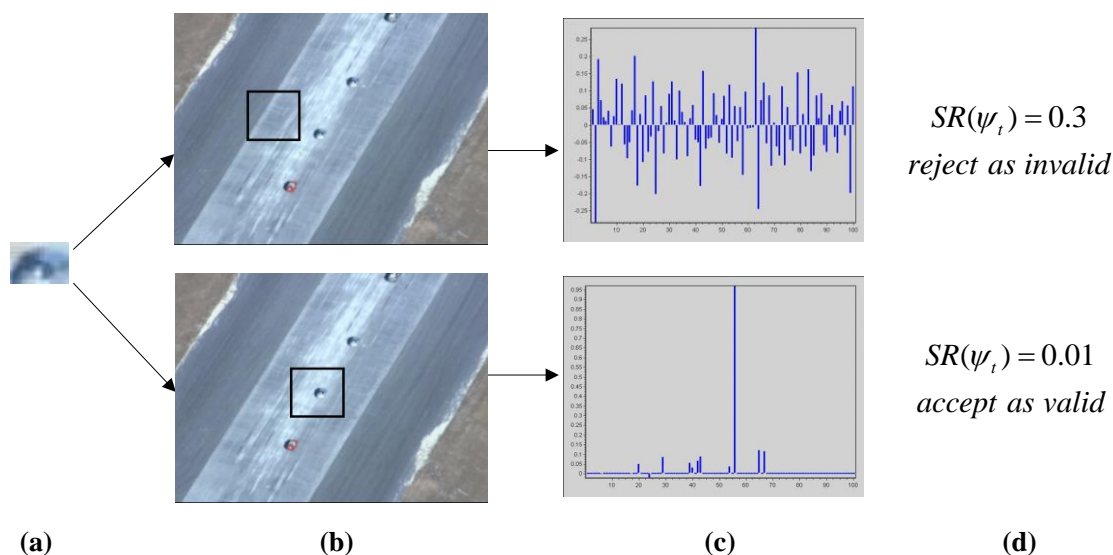


图 5-6 搜索区域的有效性验证。(a) 跟踪目标，(b)  $t$  时刻的搜索区域，(c) 稀疏重构系数向量，(d) 搜索区域有效性结果。

### 5.1.5 基于分层自适应稀疏表示与稀疏重构的目标跟踪

当搜索区域被验证为有效后，跟踪过程可以看作为基于目标的自适应稀疏表示在搜索区域内寻找跟踪目标的最佳匹配结果。由于搜索区域被描述为了重构样本集合  $Reconstruction\_set = A = \{ \hat{a}, k=1 \dots K \}$ ，因此跟踪可以等价的被视作基于自适应稀疏表示的重构样本集合  $g(A)$  与自适应稀疏表示的跟踪目标  $g(F)$  寻找最佳匹配的过程。由于稀疏重构的特性能够保证跟踪目标被和其最相似的样本优先紧致重构，也就是说，稀疏重构过程能够选择得到与跟踪目标最为相近的重

构样本子集，即样本的重构系数越大，表明样本与跟踪目标在样本空间上越相似（如图 5-5c 所示）。因此，为了得到  $\overline{O}_t(x, y, s)$ ，将重构样本集合以及其所对应的重构稀疏系数向量视为跟踪目标在样本空间内的概率分布， $t$  时刻的最终跟踪结果定义为该时刻概率分布的均值，如下所示：

$$\overline{O}_t(x, y, s) = E(\overline{a}_t^k(x, y, s)) = \sum_{k=1}^K \overline{a}_t^k(x, y, s) \times \psi_t^k. \quad (5-8)$$

其中， $\overline{O}_t(x, y, s)$  定义为  $t$  时刻的跟踪结果，其位置为  $(x, y)$  且尺度为  $s$ 。通过此迭代过程，完成对视频中目标的跟踪。

### 5.1.6 实验结果比较与算法性能分析

#### 5.1.6.1 基于 HSRR 目标跟踪框架的有效性验证

第一段测试视频来自于 VIVID 测试数据集，跟踪目标为一辆运动模式相对简单的吉普汽车，跟踪背景变化相对稳定。同时，由于跟踪目标与背景比较明显的特征差别（目标主要为白色和红色，背景主要为灰色与绿色），此段测试视频的主要难点在于跟踪目标的尺度变化较大。

第二段跟踪测试视频仍然来自于 VIVID 数据测试集合，测试视频中的跟踪目标为一辆在比较相似背景中运动的车辆。首先，跟踪目标做了一个 180° 的调头，然后加速直行，直到超过另外 3 辆车。在整个跟踪过程中，跟踪目标具有尺度变化，旋转，同时，在第 1350 帧的背景中有一辆与跟踪目标几乎完全相似的车辆，因此导致这段测试视频具有一定的挑战性。同时，跟踪目标与背景的颜色相似性对跟踪的稳定性也具有较大的影响。

第三段测试视频来自于 CARVIA 数据测试集合，此测试视频中的跟踪目标一个穿过走廊的人体。由于环境光照的影响，跟踪目标整体几乎没有颜色信息，呈现为灰色。同时，在走廊中有很多和跟踪目标相似的伪目标（例如走廊中的柱子以及商场中的模特等）。

由于本章所提的跟踪算法，即能够很好的区分前景/背景以及自适应其变化（自适应稀疏表示），又能够通过样本空间对跟踪目标进行描述（稀疏重构）。因此，如图 5-7 中的实验结果所示，此跟踪算法对跟踪过程中经常遇到的目标变化、环境变化以及部分遮挡具有较好的鲁棒性。



(a) 60<sup>th</sup> frame

400<sup>th</sup> frame

700<sup>th</sup> frame



1000<sup>th</sup> frame

1300<sup>th</sup> frame

1800<sup>th</sup> frame



(b) 60<sup>th</sup> frame

360<sup>th</sup> frame

700<sup>th</sup> frame



1000<sup>th</sup> frame

1350<sup>th</sup> frame

1600<sup>th</sup> frame



(c) 330<sup>th</sup> frame

335<sup>th</sup> frame

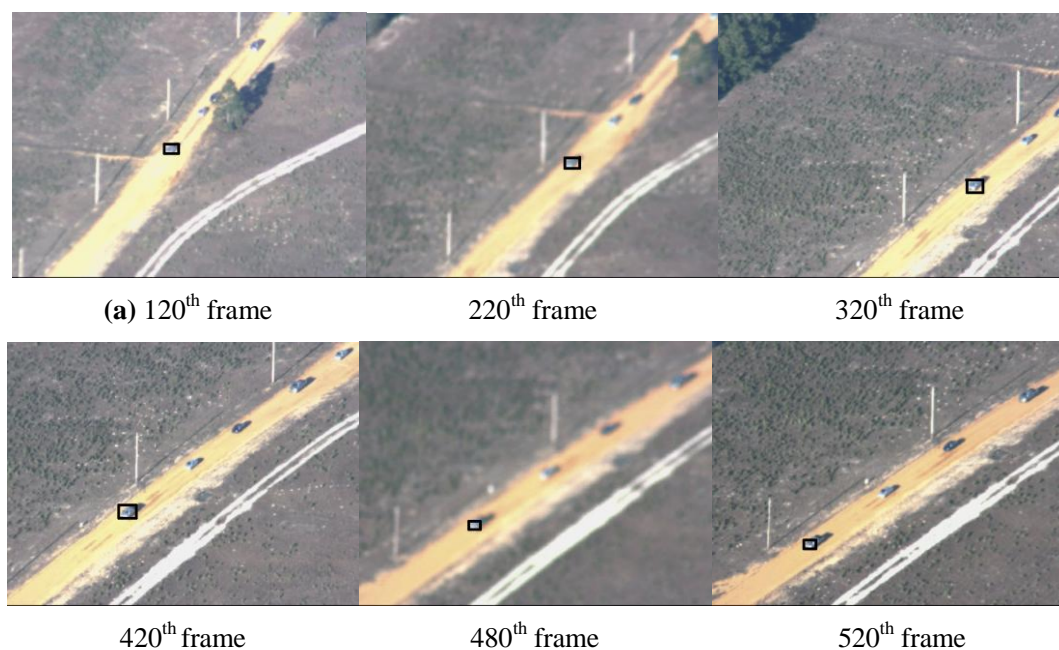
340<sup>th</sup> frame



图 5-7 跟踪结果示例

### 5.1.6.2 图像模糊退化情况下的算法性能比较

本章所提的跟踪算法与[100]的主要不同在于目标的自适应稀疏特征表示以及重构样本集合是根据当前视频帧动态建立的。因此相比较于[100]，本文的算法能够更好的保证最新的跟踪信息体现在跟踪过程，以及滤波框架中的特征评估又可以很好的抑制视频中噪声的干扰，从而能够获取更鲁棒的跟踪性能。图 5-8 所示的测试视频来自于 VIVID 数据测试集合，跟踪目标为一辆车，其沿着道路直线行驶，在跟踪的过程中，视频帧会发生严重的退化变模糊(第 220 和 480 视频帧)。在图 5-8 中，我们比较了本章所提的算法以及[100]，结果显示所提算法能够很好的跟踪目标，而[100]中所提算法在目标严重退化是发生跟踪错误（第 420, 490 以及 520 帧）。



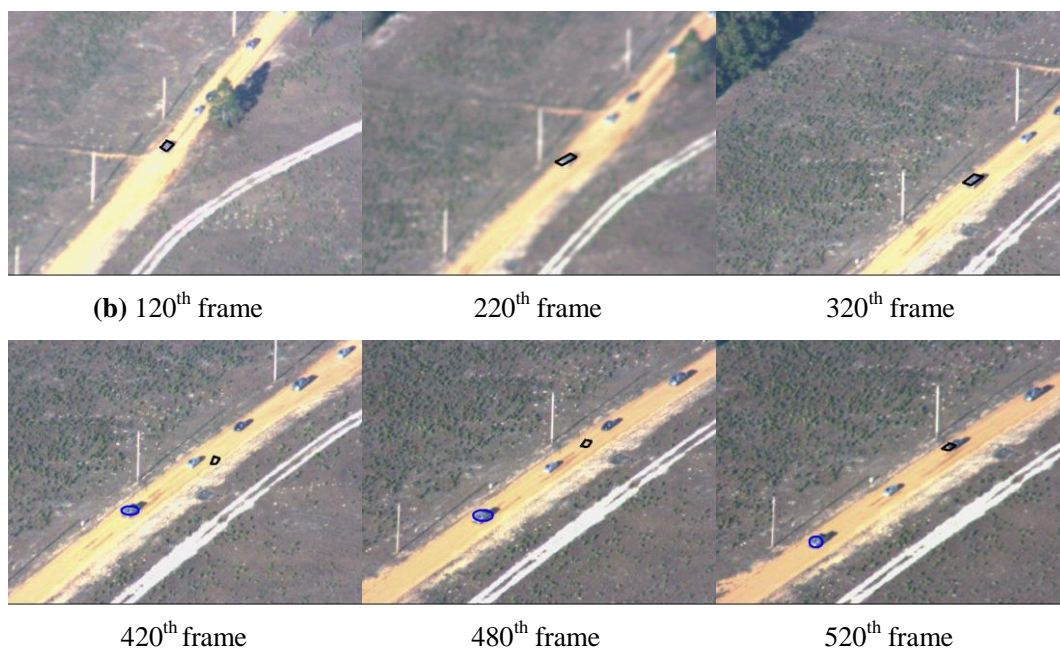


图 5-8 图像退化模糊情况下的目标跟踪结果。跟踪结果用黑色的矩形框表示，若有跟踪错误发生时，目标的实际位置通过蓝色椭圆框标定。(a) 本章所提算法跟踪结果，(b) 文献[100]中所提算法跟踪结果。

为了更加直观地体现本章所提跟踪算法的高性能，本节中与基于 HOGC, Color [73] 和 SIFT [74]特征的跟踪算法的性能进行比较。图 5-9 所示，基于分层稀疏表示与稀疏重构的目标跟踪算法的 DER 的均值与方差要明显小于其他三种跟踪算法的结果。

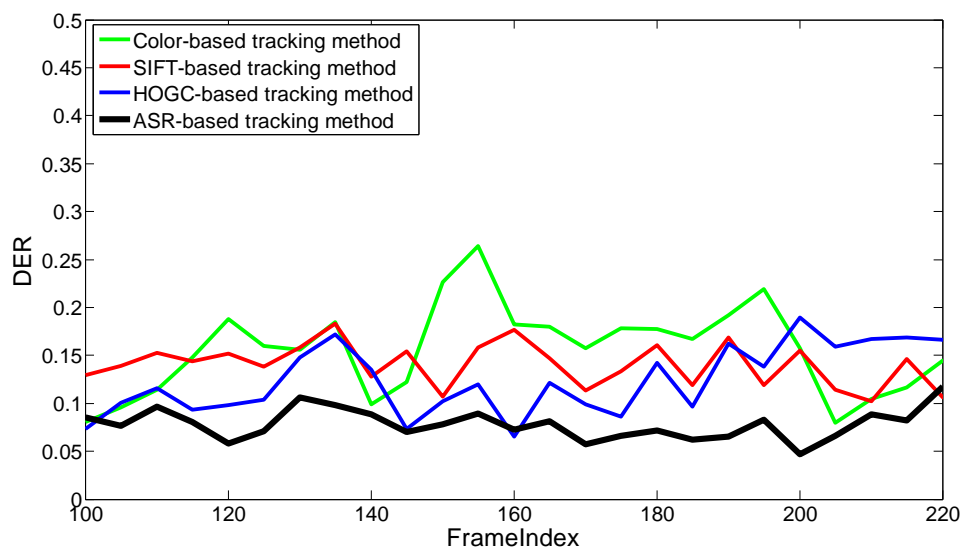
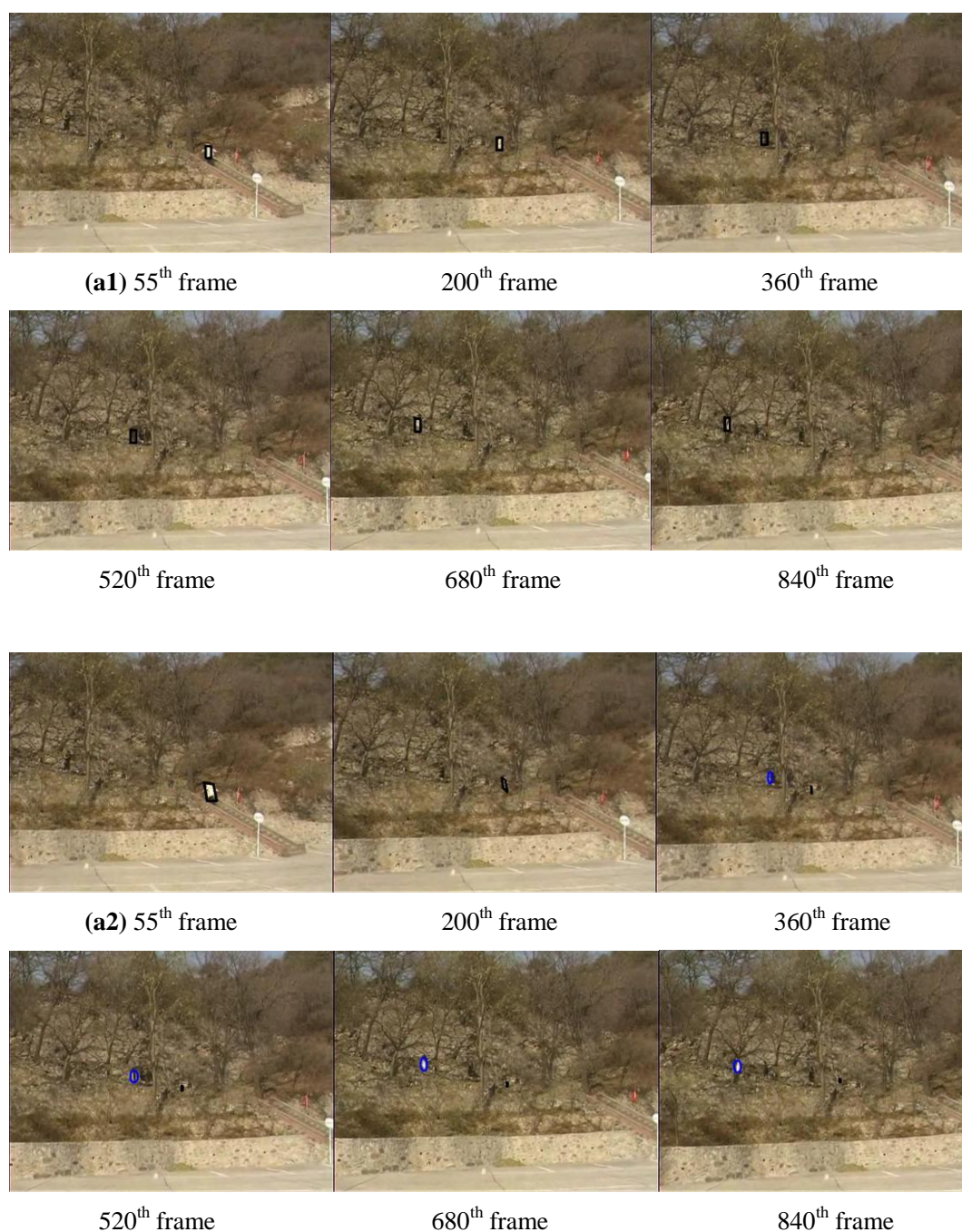


图 5-9 图像退化模糊情况下四种跟踪算法 DER 对比

### 5.1.6.3 局部遮挡情况下的算法性能比较

第一段测试视频来自于 SDL 数据测试集合，此视频的主要挑战是跟踪目标频繁的被其他人体或者树遮挡，并且这些遮挡物在形状以及颜色等外观表现方面和跟踪目标比较相似。第二段跟踪视频同样来自于 SDL 数据测试集合，此视频主要挑战仍然在于跟踪过程中目标被频繁的遮挡。在图 5-10 中，我们比较了本章所提的算法以及[100]，结果显示所提算法能够很好的跟踪目标，而[100]中所提算法在目标严重局部遮挡时发生跟踪错误。





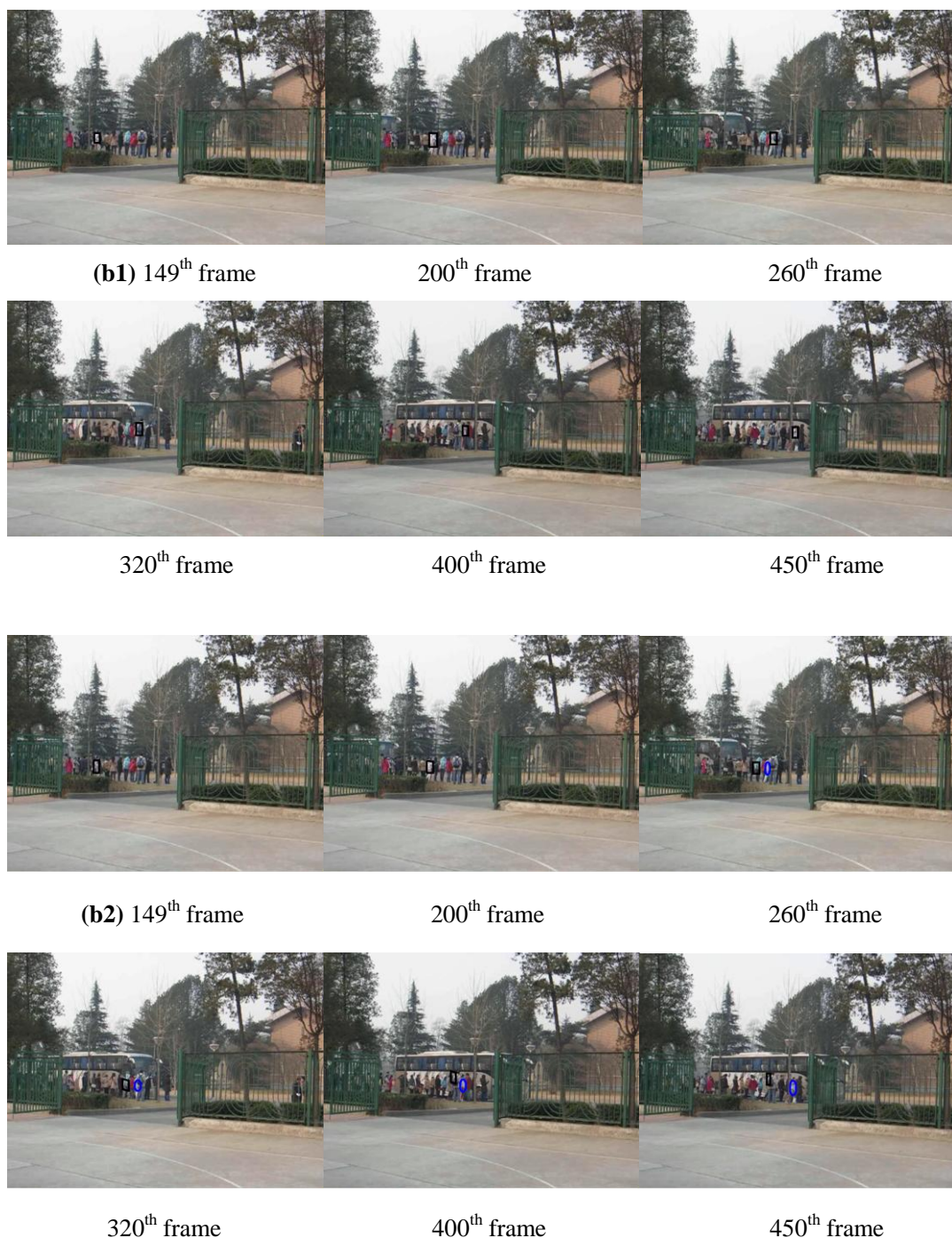


图 5-10 局部遮挡情况下目标跟踪结果实例。跟踪结果用黑色的矩形框表示，若有跟踪错误发生时，目标的实际位置通过蓝色椭圆框标定。(a) 本章所提算法跟踪结果，(b) 文献[100]中所提算法跟踪结果。

在定量对比实验中，比较了本章所提出的跟踪算法与经典的 Kalman Filter based tracking method [20]以及 Particle Filter based tracking method [7]。定量对比结果如图 5-11 所示，分层稀疏表示与稀疏重构跟踪算法取得了更小的 DER（大约 0.04 到 0.1）。

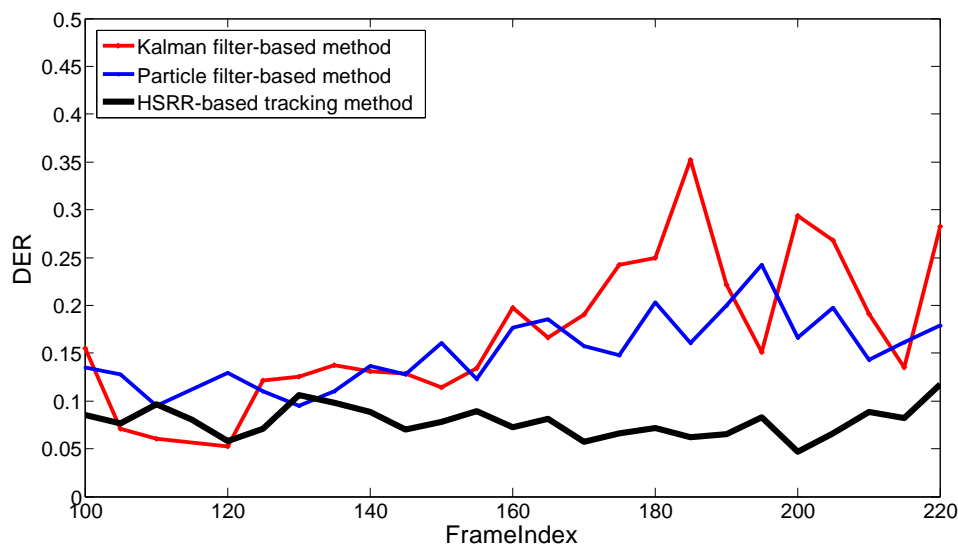


图 5-11 遮挡情况下三种跟踪算法的 DER 对比结果图

跟踪效率是视觉目标跟踪的另一个重要的评价指标。本节中，对本章所提算法（基于 HOGC 和改进的 HOGC）与[100]进行对比，其结果如表 5-1 所示。实验平台为普通台式计算机（Core(TM)2 Duo CPU (2.53GHz) and 3GB memory）。

表 5-1 算法跟踪效率对比

Tracking method	Seconds / 100 video frames
X. Mei and H. b. Ling [100]	>300
Our approach (HSRR) with HOGC	<b>5.76</b>

## 5.2 基于自适应稀疏重构回归分析的多目标跟踪

对于多目标跟踪，一方面要精确的跟踪每个目标（鲁棒的单目标跟踪算法），另一方面要确定跟踪的是哪个目标（分类算法）。

### 5.2.1 稀疏表示在模式分类领域内的典型应用

在模式识别领域，很多学者应用 L1-norm 的稀疏特性去判别分类。这里面比较典型的应用有图像重建、人脸识别、图像对齐等。早期的关于 L1-norm 在模式识别领域应用的文献，要数 John Wright 和 Yi Ma[80]等人提出的 SRC (Sparse

Representation Classification)算法。该算法使用 L1-norm 最小化的稀疏表示识别人脸。该问题是一个多类回归问题，约束方程假设一张测试人脸样本  $y$  可以由训练样本集  $A$  进行线性的稀疏表示，目标函数是表示系数  $x$  的 L1-norm 最小化。SRC 算法使用 L1-norm 的目的是希望这种线性表示尽可能的稀疏，最后根据重构误差最小来实现对人脸的识别：

$$\begin{aligned} \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (5-9)$$

假设有  $k$  类不同人的脸，同一个人的不同表情、不同光照下的图像构成一个集合  $A_i$ ，所有人的脸构成训练样本集合  $A = [A_1, A_2 \cdots A_k] \in R^{m \times n}$ ，上述优化模型可等价写成：

$$\begin{aligned} \tilde{x} = \arg \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (5-10)$$

考虑到噪声与误差的影响，那么 (5-10) 式可以重写成：

$$\begin{aligned} \tilde{x} = \arg \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & \|Ax - y\|_2 \leq \varepsilon \end{aligned} \quad (5-11)$$

最后计算残差  $r_i(y) = \|y - A\mathcal{D}_i(\tilde{x}_1)\|_2$ ，残差最小的标号则是测试样本  $y$  的类别。

$$\text{class}(y) = \arg \min_i r_i(y) \quad (5-12)$$

该算法最巧妙的地方就是通过稀疏表示，可以消除遮挡和噪声对人脸图像的影响，详细的关于算法的实验部分本节不再介绍，请参见[80]。从压缩感知到人脸识别问题，这些问题的本质思想，都是通过稀疏表示来实现对物体的识别，属于回归问题。本文受到上述研究的启发，研究基于 L1-norm 最小化的分类问题，本文研究的算法与 Yi Ma [80] 等人所提出的回归模型以及压缩感知中重构部分的想法不同。

## 5.2.2 基于自适应稀疏重构回归分析的多目标跟踪

基于自适应稀疏重构回归分析 (Adaptive Sparse Reconstruction Analysis,

AdaSRA), 提出了一种新颖的方法对每个跟踪目标进行准确分类, 从而实现多目标跟踪。在多目标跟踪过程中, 建一个归一化的重构样本集合, 即将每个跟踪目标的样本结合串联在一起构成一个新的样本集合, 如公式(5-13)所示:

$$B = \cup\{A_j\} = \{a_1^1, a_1^2, \dots, a_1^K, a_2^1, a_2^2, \dots, a_2^K, \dots, a_J^1, a_J^2, \dots, a_J^K\}, \quad j=1 \dots J, \quad (5-13)$$

其中,  $J$  表示跟踪目标的数目,  $A_j$  表示第  $j$  个跟踪目标的重构样本集合。如果没有目标所对应的重构样本集合中有  $K$  个重构样本, 那么归一化的重构样本集合总共包含有  $K \times J$  个重构样本。

### 5.2.2.1 可行性分析

在跟踪的初始化时, 对于每一个跟踪目标, 计算其基于归一化重构样本集合的归一化重构系数  $\psi_{0,j}$ , 将此稀疏向量定义为该目标的参考重构系数向量。同时, 考虑到每个跟踪目标不可避免的全会发生变化, 我们同样在滤波框架中对每个目标对应的参考重构系数向量  $\psi_{0,j}$  进行自适应的评估。在  $t$  时刻, 当获取一个未知的跟踪结果  $F_{unknown}$  以及其所对应的归一化重构系数  $\psi_{t,unknown}$  时, 我们通过计算  $\psi_{t,unknown}$  与  $\psi_{0,j}$  的相似度来判断跟踪结果的分类, 即跟踪的是哪个目标。分类准则如下所示:

$$\text{Classify}(F_{unknown}) = \underset{j}{\text{argmin}} r_j(F_{unknown}) \quad \text{where } r_j(F_{unknown}) = \|\psi_{t,unknown} - \psi_{0,j}\|_2. \quad (5-14)$$

如图 5-12 和 5-13 中所示的是两类跟踪目标 (每类跟踪目标具有两个在不同时刻的跟踪结果) 及其所对应的重构系数向量。图 5-12a 和 b 来自于测试视频 VIVID-redteam, 图 5-12c 和 d 来自于测试视频 VIVID-egtest01。 $\|\cdot\|_2$  表示 2 范式, 图 5-13 显示每组稀疏重构向量之间的差, 其中最小的两组差用分别用红色字体表示。从结果可以看出, AdaSRA 能够对跟踪目标进行准确分类。

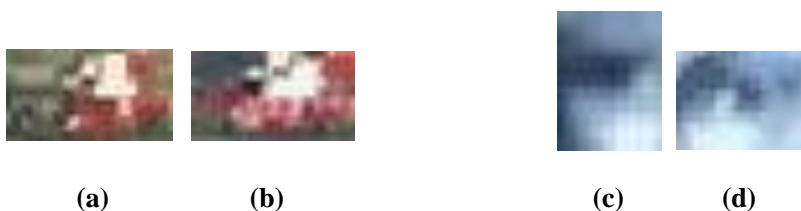


图5-12 多目标跟踪过程中的两类跟踪目标实例。(a) 和 (b) 来自于同一个跟踪目标, 且 (c) 和 (d) 来自于另一个跟踪目标。

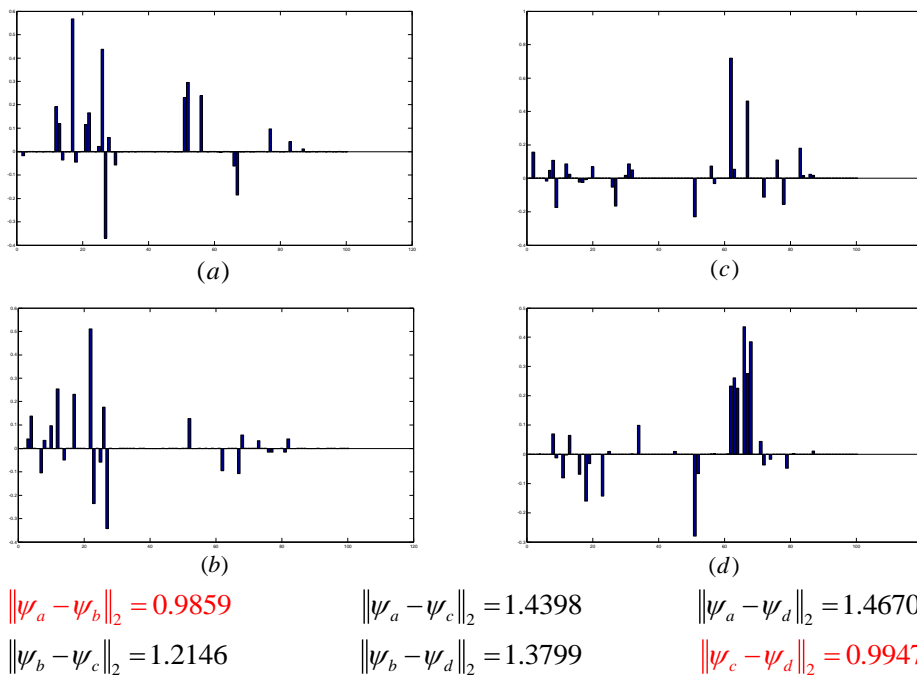


图5-13 AdaSRA的结果分析。(a)–(d)分别为图5-12 (a)–(d)分别对应的归一化重构稀疏系数。

### 5.2.2.2 有效性验证

公式 5-14 所定义的分类准则与 SRC[80]不同之处主要体现在，SRC 每次只利用某一类样本所对应的重构系数信息去比较，最后分类结果为具有最小残差所对应的类型，而 AdaSRA 利用的是重构系数向量中的全部信息进行比较。虽然 SRC 在人脸识别上具有非常成功的应用，简单分析可以得到，由于人脸样本数据量大且经过大量的线下的预处理，这些对于实时的在线目标跟踪来讲，几乎是不可能的。

在对比分析中，我们定义了三个跟踪目标，及其所对应的部分重构样本集合如图 5-14 所示。我们从三个跟踪目标中挑选出一个跟踪目标（如图 5-15a 所示，其属于第一类跟踪目标）。在对比实验一中，每个重构样本集合由 50 个样本构成（归一化重构样本集合有  $50 \times 3 = 150$  个样本），基于 AdaSRA 与 SRC 的  $r_j(F_t)$  的结果如图 5-15b 所示，SRC 将跟踪目标错误的分类为第二类，而 AdaSRA 能够正确的将目标分类。实验二，我们将每个重构样本集合的样本数目增加为 100（归一化样本集合的样本数为 300），此时基于 SRC 与 AdaSRA 计算得到的  $r_j(F_t)$  的结果如图 5-15c 所示，SRC 的计算结果虽然有所改进，但是其仍将跟踪目标错误的分类为第二类，而 AdaSRA 仍然能够正确的将目标分类。

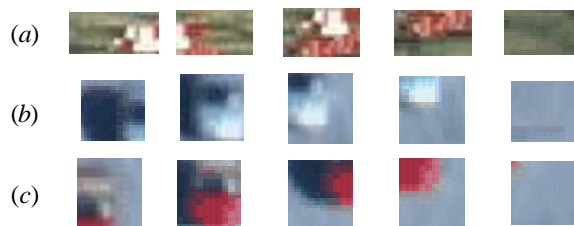


图 5-14 三种跟踪目标所对应的重构样本集合中部分样本实例。(a) 第一类目标的部分样本，(b) 第二类目标的部分样本，(c) 第三类目标的部分样本。

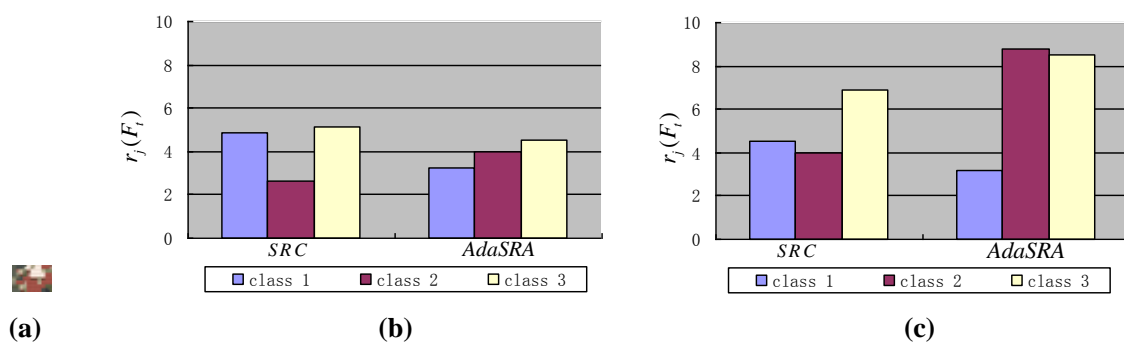
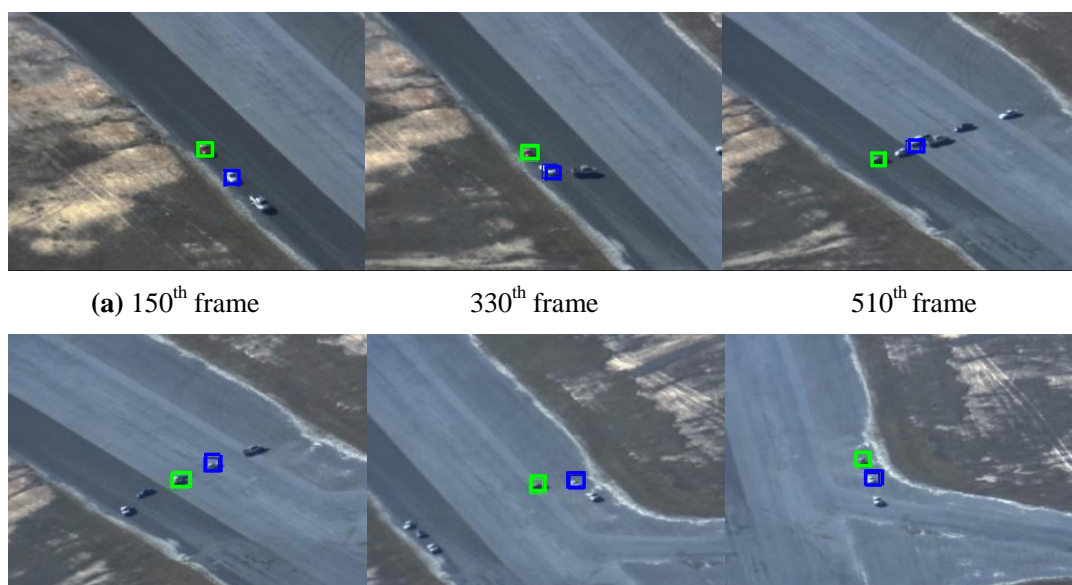


图5-15 基于AdaSRA与SRC的分类结果比较。(a) 测试目标（第一类），(b) 当每一类重构样本数目为50时计算得到的分类结果，(c) 当每一类重构样本数目为100时计算得到的分类结果。

### 5.2.2.3 基于自适应稀疏重构分析的多目标跟踪实验结果

基于 AdaSRA 的多目标跟踪实验中，第一段跟踪实验有两个跟踪目标，在跟踪的过程中会发生车辆的交错行驶；第二段跟踪视频中包括四个跟踪目标，其中既有运动的人体，也有静止的沙发与书柜。跟踪结果，如下图所示：



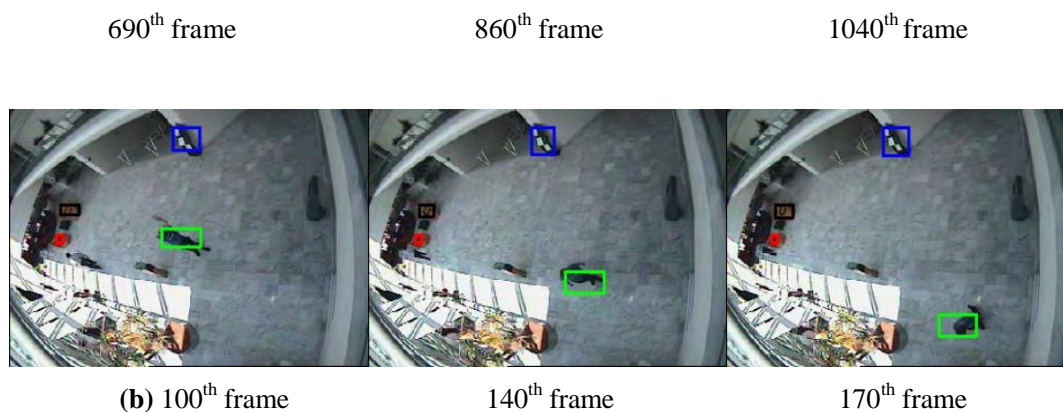


图 5-16 基于稀疏重构分析 (AdaSRA) 的多目标跟踪结果实例

### 5.3 小结

本章基于前面第二到第四章的研究内容,提出了基于分层自适应稀疏表示与稀疏重构的视觉目标跟踪框架。详细介绍了基于稀疏自适应综合特征和稀疏重构对目标进行跟踪的过程。

本章给出了该跟踪框架在各种复杂场景下的运动目标跟踪实验,从而对算法在运动目标尺度伸缩变化、形变、前景/背景动态变化、部分遮挡等情况下的鲁棒性跟踪进行了验证,同时将该算法与传统的跟踪算法的性能进行了比较与分析。

## 总结与展望

目标跟踪是计算机视觉的一个重要问题，是其他计算机视觉问题（如行为识别、轨迹分析等）的基础。经过近 50 多年的深入研究和发 展，视觉跟踪技术已在军事制导、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广泛应用。然而，到目前为止，一般意义上的视觉跟踪技术还未成熟。本文面向目标跟踪过程中目标表示的基础理论，针对前景/背景动态变化、局部遮挡等典型难点问题，提出了一种视觉目标自适应跟踪技术体系。

首先，提出了一种面向跟踪的目标综合特征的；其次，基于此综合特征分别提出了两种自适应目标跟踪算法（自适应综合特征与自适应稀疏表示视觉目标跟踪算法）；然后，基于最近的压缩感知理论，进一步提出了一种自适应目标跟踪的新框架。取得了如下的研究成果：

- 1、提出了一种面向视觉目标跟踪的综合特征表示。基于颜色直方图（Histogram of Color, HC）和梯度方向直方图（Histogram of Oriented Gradient, HOG）的综合直方图 HOGC（Histogram of Oriented Gradient and Color）融合了颜色、局部轮廓以及梯度方向等特征。同时，基于 SIFT（Scale Invariant Feature Transform）特征点的主方向方法，进一步对综合特征进行了面向视觉目标跟踪的改进，在一定程度上解决了 HOG 对跟踪目标旋转敏感的缺点。

- 2、基于综合特征表示，提出了基于自适应综合特征表示的视觉目标跟踪算法。基于前景（跟踪目标）/背景的特征信息以及相关变化，将目标综合特征的评估嵌入到传统的滤波框架中，保证特征权重的变化是连续的过程，即具有时间连续性。基于滤波框架的特征评估有效地将滤波算法从对目标运动状态的建模扩展到了对目标特征权重的评估，拓展了传统滤波框架的应用范畴，并且在一定程度上解决了跟踪前景/背景在跟踪过程中的变化对跟踪效果的影响。

- 3、基于自适应综合特征表示，进一步提出了基于稀疏自适应综合特征表示的视觉目标跟踪算法。在跟踪的初始化过程中建立在线训练样本集合，用以对跟踪特征的实时稀疏选择，从而得到跟踪目标具有前景/背景区分性的稀疏特征表示。同时，基于滤波框架对目标的稀疏特征表示进行有效地自适应的评估，在保证特征描述具有区分性的基础上，进一步保证特征的环境自适应性，从而能够更



好的解决前景/背景在跟踪过程中的变化对跟踪效果的影响。

4、基于稀疏自适应特征表示，提出了基于分层自适应稀疏表示与稀疏重构的目标跟踪新框架。在跟踪的过程中，基于跟踪目标的稀疏自适应特征表示以及基于搜索区域建立的实时重构样本集合，获得目标在样本空间的自适应重构系数，从而统计跟踪目标在样本空间的概率密度分布，最终完成对目标的跟踪。所提出的跟踪框架既能较好的解决前景/背景的变化，又能有效的处理跟踪目标的局部遮挡等问题对跟踪效果的影响。

实验结果证明，本文所提出的视觉目标自适应跟踪算法与框架均能取得较好的跟踪效果，验证了其有效性。进一步，通过与经典的跟踪算法的定量实验对比，表明我们所提出的跟踪算法与框架在前景/背景动态变化以及局部遮挡等情况下，对视觉目标跟踪的优势更加明显，验证了其有效性。同时，我们也将基于 Kalman 滤波权重评估的目标跟踪算法嵌入到 PTZ 系统中，在实际应用中对其进行验证，验证了将自适应目标跟踪算法推向实际应用系统的可能性。

本文提出的新概念和新技术主要包括：基于颜色和梯度方向的综合直方图、基于滤波框架的自适应特征表示、基于压缩感知的自适应稀疏表示以及基于分层自适应稀疏表示与稀疏重构的视觉目标跟踪新框架。其中，综合直方图可以很容易地融合其他一些特征，例如纹理特征等；其次，所提到的基于滤波框架的自适应特征描述的获取也不仅仅局限于 Kalman 或者粒子滤波算法，其他类似的滤波算法也可以在此框架中进行跟踪特征权重评估，例如扩展 Kalman 滤波等；同时，所提出的稀疏表示与稀疏重构的获取也不仅仅局限于 L1-norm minimization，其他的算法也可以取得稀疏解，例如 L<sub>1/2</sub>-norm 等。

视觉目标跟踪是一项充满了挑战的研究课题。尽管本文针对其中的几个关键问题进行了一些探索和尝试，取得了一些研究成果。但是，必须指出的是，视觉目标跟踪仍然是一个开放问题，距离问题的最终解决还有很长的路要走。由于时间和精力关系，我们没有对 Kalman 滤波或者粒子滤波的理论进行深入研究并对其进行改进以使其对目标运动的描述更加精确。因此，在某些情况下，论文所得到的跟踪结果并没有某些基于运动预测的跟踪算法好。另一个不足在于，我们提出的综合特征集在跟踪目标的尺度发生很大变化的时候，容易产生跟踪错误。在将来的工作中，我们拟提出更多的对光照和尺度变化不敏感的鲁棒特征并

将其融合到所提出的综合特征中,使得所提出的跟踪方法能够在目标具有大尺度变化的情况下具有鲁棒性。最后需要指出的是,视觉目标跟踪算法研究是受到实际需求驱动的应用型研究,需要针对应用系统中的实际问题,不断研究新技术、新算法,解决新问题。

## 参考文献

- [1] A. Yilmaz, O. Javed, and M. Shah, Object Tracking: A Survey, ACM Computing Surveys, 2006.
- [2] J. Shi, and C. Tomasi, Good features to track, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:593-600, 1994.
- [3] C. Stauffer, and W. Grimson, Adaptive background mixture models for real-time tracking, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:246-252, 1999.
- [4] G. Bradski, Real time face and object tracking as a component of a perceptual user interface, Proceedings of IEEE Workshop on Applications of Computer Vision, pp:214-219, 1998.
- [5] N. Papanikolopoulos, P. Khosla, and T. Kanade, Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision, IEEE Trans. Robotics and Automation, pp:14-35, 1993.
- [6] K. Hariharakrishnan, and D. Schonfeld, Fast object tracking using adaptive block matching, IEEE Trans. Multimedia, pp:853-859, 2005.
- [7] M. Isard, and A. Blake, CONDENSATION-conditional density propagation for visual tracking, International Journal on Computer Vision, pp:5-28, 1998.
- [8] D. Comaniciu, V. Ramesh, and P. Meer, Kernel-based object tracking, IEEE Trans. Pattern Analysis and Machine Intelligence, pp:564-577, 2003.
- [9] A. Elgammal, R. Duraiswami, and L. Davis, Probabilistic tracking in joint feature-spatial spaces, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:781-788, 2003.
- [10] R. Collins, Mean-shift blob tracking through scale space, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:234-240, 2003.
- [11] R. Collins, Y. Liu and M. Leordeanu, Online selection of discriminative tracking features, IEEE Trans. Pattern Analysis and Machine Intelligence, 2005.
- [12] H. Nguyen, and A. Smeulders, Tracking aspects of the foreground against the background, Proceedings of European Conference on Computer Vision, pp:446-456, 2004.
- [13] D. Chen, and J. Yang, Robust object tracking via online spatial bias appearance model learning, IEEE Trans. Pattern Analysis and Machine Intelligence, pp:2157-2169, 2007.
- [14] H. Bohyung, and L. Davis, Object tracking by adaptive feature extraction, Proceedings of IEEE International Conference on Image Processing, pp:1501-1504, 2004.

- [15] B. Han, Y. Zhu, D. Comaniciu, and L. Davis, Kernel-based Bayesian filtering for object tracking, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:227-234, 2005.
- [16] Y. Huang, and I. Essa, Tracking multiple objects through occlusions, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:1051-1058, 2005.
- [17] Y. Wu, Robust visual tracking by integrating multiple cues based on co-inference learning, International Journal of Computer Vision, pp:55-71, 2004.
- [18] Z. Khan, T. Balch, and F. Dellaert, A rao-blackwellized particle filter for eigentracking, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:980-986, 2004.
- [19] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, Color-based probabilistic tracking, Proceedings of European Conference on Computer Vision, pp:661-675, 2002.
- [20] E. Cuevas, D. Zaldivar, and R. Rojas, Kalman filter for vision tracking, Technical Report B, Fachbereich Mathematikund Informatik, Freie Universit ät Berlin, 2005.
- [21] E. Cuevas, D. Zaldivar, and R. Rojas, Particle filter for vision tracking, Technical Report B, Fachbereich Mathematikund Informatik, Freie Universit ät Berlin, 2005.
- [22] 王建宇. 基于序列蒙特卡洛滤波算法的视觉目标跟踪. 哈尔滨工业大学博士论文. 2006.
- [23] 梁大为. 视频运动对象跟踪技术研究. 哈尔滨工业大学博士论文. 2009.
- [24] J.Q. Wang, and Y.S. Yagi, Integrating color and shape-texture features for adaptive real-time object Tracking, IEEE Trans. Image Processing, pp:235-240, 2008.
- [25] W.m. Hu, D. Xie, and T.n. Tan, A hierarchical self-organizing approach for learning the patterns of motion trajectories, IEEE Trans. Neural Networks, pp:135-144, 2004.
- [26] 李美花. 三维人体跟踪. 大连理工大学硕士学位论文. 2004.
- [27] 彭宁嵩, 杨杰, 刘志, 张凤超. Mean-Shift跟踪算法中核函数窗宽的自动选取. 软件学报. pp:1542-1550, 2005.
- [28] 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.
- [29] 薛雷, 刘泽平, 丁艳. 区域增长的图像跟踪算法的研究. 光学技术. 2005.
- [30] 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.
- [31] 常发亮. 彩色图像分割与复杂场景下视觉目标跟踪方法研究. 山东大学博士学位论文. 2006.
- [32] Z.j. Han, Q.x. Ye, and J.b. Jiao, Online feature evaluation for object tracking using kalman filter, Proceedings of International Conference on Pattern Recognition, 2008.

- [33] Z.j. Han, Q.x. Ye, Y.m. Liu and J.b. Jiao, Feature evaluation by particle filter for adaptive object tracking, Proceedings of SPIE, 2009.
- [34] T. Yang, Q. Pan, J. Li, and S.Z. Li, Real-time multiple objects tracking with occlusion handling in dynamic scenes, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:970-975, 2005.
- [35] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Analysis and Machine Intelligence. pp:679-698, 1986.
- [36] K. Bowyer, C. Kranenburg, and S. Dougherty, Edge detector evaluation using empirical ROC curves, Computer Vision and Image Understanding, pp:77-103, 2001.
- [37] B.K.P. Horn, and B.G. Schunck, Determining optical flow, Artificial Intelligence, pp:185-203, 1981.
- [38] M.J. Black, and P. Anandan, The robust estimation of multiple motions: parametric and piecewise-smooth flow fields, Computer Vision and Image Understanding, pp:75-104, 1996.
- [39] R. Szeliski, and J. Coughlan, Spline-Based image registration, International Journal of Computer Vision, pp:199-218, 1997.
- [40] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, Performance of optical flow techniques, International Journal of Computer Vision, pp:43-77, 1994.
- [41] R.M. Haralick, K. Shanmugam, and I. Dinstein, Textural features for image classification, IEEE Trans. Systems, Man, and Cybernetics, pp:610-621.
- [42] A. Haar, Zur Theorie der orthogonalen Funktionensysteme, Mathematische Annalen, pp:331-371, 1910.
- [43] Charles K. Chui, An introduction to wavelets, Academic Press, San Diego, ISBN 0585470901, 1992.
- [44] R.G. Brown, and P.Y.C. Hwang, Introduction to random signals and applied Kalman Filtering, John Wiley & Sons, New York, 1992.
- [45] D. Reid, An algorithm for tracking multiple targets, IEEE Trans. Automatic Control, pp:843-854, 1979.
- [46] C.j. Yang, R. Duraiswami, and L. Davis, Fast multiple object tracking via a hierarchical particle filter, Proceedings of IEEE Conference on Computer Vision, pp:212-219, 2005.
- [47] I.J. Cox, and S.L. Hingorani, An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking, IEEE Trans. Pattern Analysis and Machine Intelligence, pp:138-150, 1996.
- [48] L. Lee, R. Romano, and G. Stein, Monitoring activities from multiple video streams: establishing a common coordinate frame, IEEE Trans. Pattern Analysis and Machine Intelligence, pp:758-767, 2000.
- [49] S.L. Dockstader, and A.M. Tekalp, Multiple camera tracking of interacting and occluded human motion, Proceedings of the IEEE, pp:1441-1455, 2001.

- [50] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, Tracking across multiple cameras with disjoint views, Proceedings of IEEE International Conference on Computer Vision, pp:952-957, 2003.
- [51] K. Jinman, I. Cohen, and G. Medioni, Continuous tracking within and across camera streams, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:267-272, 2003.
- [52] S. Khan, and M. Shah, Consistent labeling of tracked objects in multiple cameras with overlapping fields of view, IEEE Trans. Pattern Analysis and Machine Intelligence, pp:1355-1360, 2003.
- [53] A. Mittal, and L.S. Davis, M2Tracker: A multi-view approach to segmenting and tracking people in a cluttered scene, International Journal of Computer Vision, pp:189-203, 2003.
- [54] O. Javed, K. Shafique, and M. Shah, Appearance modeling for tracking in multiple non-overlapping cameras, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:26-33, 2005.
- [55] C. Stauffer, and W.E.L. Grimson, Adaptive background mixture models for real-time tracking, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:246-252, 1999.
- [56] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, Wallflower: principles and practice of background maintenance, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:255-261, 1999.
- [57] A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, Proceedings of the IEEE, pp:1151-1163, 2002.
- [58] M. Heikkila, and M. Pietikainen, A texture-based method for modeling the background and detecting moving objects, IEEE Trans. Pattern Analysis and Machine Intelligence, pp:657-662, 2006.
- [59] A. Yilmaz, K. Shafique, and M. Shah, Target tracking in airborne forward looking infrared imagery, Image and Vision Computing. pp:623-635, 2003.
- [60] S. Blackman, and R. Popoli, Design and analysis of modern tracking systems, Artech House, 1999.
- [61] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, Sonar tracking of multiple targets using joint probabilistic data association, IEEE Journal of Oceanic Engineering, pp:173-184, 1983.
- [62] D.G. Lowe, Distinctive image features from scale-invariant key points, International Journal of Computer Vision, pp:91-110, 2004.

- [63] G. Hager, and P. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp:1025–1039, 1998.
- [64] N. Pérez de la Blanca, J.M. Fuertes, and M. Lucena, Deformable object matching based on multi-scale local histograms, *AMDO 2004*, pp:154-162, 2004.
- [65] Z. Yin, and R. Collins, Spatial divide and conquer with motion cues for tracking through clutter, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp:570-577, 2006.
- [66] S. Avidan, Ensemble Tracking, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp:494-501, 2005.
- [67] H. Grabner, and H. Bischof, On-line boosting and vision, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp:260-267, 2006.
- [68] J. Wang, X. Chen, and W. Gao, Online selecting discriminative tracking features using particle filter, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp:1037-1042, 2005.
- [69] N. Dalal, and B. Triggs, Histograms of oriented gradients for human detection, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp:886–893, 2005.
- [70] VIVID Tracking Evaluation Web Site at: <http://www.vividevaluation.ri.cmu.edu/datasets/datasets.html>.
- [71] CAVIAR Test Case Scenarios at: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR>.
- [72] SDL data set at: <http://coe.gucas.ac.cn/SDL-HomePage/>.
- [73] T.D. Grove, K.D. Baker, and T.n. Tan, Colour based object tracking, *Proceedings of International Conference on Pattern Recognition*, 1998.
- [74] Y.J. Li, J.F. Yang, R.B. Wu, and F.X. Gong, Efficient object tracking based on local invariant features, *IEEE International Symposium on Communications and Information Technologies*, pp:697-700, 2006.
- [75] A. Yilmaz, X. Li, and B. Shah, Contour based object tracking with occlusion handling in video acquired using mobile cameras, *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp:1531-1536, 2004.
- [76] A. Monnet, A. Mittal, and N. Paragios, Background modeling and subtraction of dynamic scenes, *Proceedings of IEEE International Conference on Computer Vision*, pp:1305-1312, 2003.
- [77] T.F. Cootes, G.J. Edwards, and C.J. Taylor, Active appearance models, *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp:681-685, 2001.
- [78] A. Elgammal, D. Harwood, and L. Davis, Non-parametric model for background subtraction, *Proceedings of IEEE Frame Rate Workshop*, 1999.

- [79] T. Serre, Learning a dictionary of shape-components in visual cortex: Comparison with neurons, humans and machines, Ph.D dissertation, MIT, 2006.
- [80] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2008.
- [81] J. Rissanen, Modeling by shortest data description, *Automatica*, pp:465-471, 1978.
- [82] M. Hansen, and B. Yu, Model selection and the minimum description length principle, *Journal of the American Statistical Association*, pp:746-774, 2001.
- [83] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, and Z. Jin, KPCA Plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp:230-244, 2005.
- [84] P.S. Bradley, O.L. Mangasarian, Feature selection via concave minimization and support vector machines, *Proceedings of the Fifteenth International Conference on Machine Learning*, 1998.
- [85] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society*, pp:267-288, 1996.
- [86] J.y. Pan, B. Hu, and J.Q. Zhang, Robust and accurate object tracking under various types of occlusions, *IEEE Trans. Circuits and Systems for Video Technology*, pp:223-236, 2008.
- [87] S. Mallat, and Z. Zhang, Matching pursuit with time-frequency dictionaries, *IEEE Trans. Signal Processing*, pp:3397-3415, 1993.
- [88] 石光明, 刘丹华, 高大化, 刘哲, 林杰, 王良君, 压缩感知理论及其研究进展, *电子学报*, 2009.
- [89] D.L. Donoho, Compressed sensing, *IEEE Trans. Information Theory*, pp:1289-1306, 2006.
- [90] D. Donoho, For most large underdetermined systems of linear equations the minimal L1-norm near solution approximates the sparsest solution, *Comm. on Pure and Applied Math*, pp:797-829, 2006.
- [91] K. Huang, SAviyente, Sparse representation for signal classification, *Advances in Neural Information Processing Systems*, Proc. of the Twentieth Annual Conf. on Neural Information Processing Systems, 2007.
- [92] B.S. Kashin. Diameters of some finite-dimensional sets and classes of smooth functions, *Izv. Akad. Nauk SSSR Ser. Mat.*, pp:334-351, 1977.
- [93] D.L. Donoho, and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via L1-norm minimization, in *Proc. Nat. Aca. Sci.*, pp:2197-2202, 2003.
- [94] E.J. Candès, J. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math*, pp:1207-1223, 2006.
- [95] E.J. Candès, and T. Tao, Decoding by linear programming, *IEEE Trans. Information*



Theory, pp:4203-4215, 2005.

[96] J.y. Yang, Y.g. Peng, W.l. Xua, and Q.h. Dai, Ways to sparse representation: A comparative study, Tsinghua Science & Technology, pp:434-443, 2009.

[97] Y.C. Pati, R. Rezaiifar, P. Krishnaprasad, and S. Orthogonal, Matching pursuit: Recursive function approximation with applications to wavelet decomposition, Proceedings of IEEE 27th Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, pp:40-44, 1993.

[98] S.S. Chen, Basis pursuit, Ph.D Thesis, Stanford University, 1995.

[99] Y. Qian, G. Medioni, and I. Cohen, Multiple target tracking using spatial-temporal markov chain monte carlo data association, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:1-8, 2007.

[100] X. Mei, and H.b. Ling, Robust visual tracking using L1 minimization, Proceedings of IEEE Conference on Computer Vision, 2009.

[101] D. Comaniciu, V. Ramesh, and P. Meer, Real-time tracking of non-rigid objects using mean shift, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:142-149, 2000.

[102] Q. Zhu, S. Avidan, M.C. Yeh, and K.T. Cheng, Fast human detection using a cascade of histograms of oriented gradients, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006.

[103] I. Matthews, T. Ishikawa, and S. Baker, The template update problem, IEEE Trans. Pattern Analysis and Machine Intelligence, pp:810-815, 2004.

## 发表文章目录

### 已发表论文目录

1. **Zhenjun Han**, Jianbin Jiao<sup>✉</sup>, Baochang Zhang, Qixiang Ye, Jianzhuang Liu, “Visual Object Tracking via Sample-Based Adaptive Sparse Representation (AdaSR)”, Pattern Recognition, Elsevier, 2011, Volume 44, Issue 9, pp: 2170-2183. ( SCI 收录 )
2. **Zhenjun Han**, Qixiang Ye, Jianbin Jiao<sup>✉</sup>, “Combined feature evaluation for adaptive visual object tracking”, Computer Vision and Image Understanding, Elsevier, 2011, Volume 115, Issue 1, pp: 69-80. ( SCI 收录 )
3. **Zhenjun Han**, Jianbin Jiao<sup>✉</sup>, Qixiang Ye, “A FAST OBJECT TRACKING APPROACH BASED ON SPARSE REPRESENTATION”, Proc. of IEEE International Conference on Image Processing, 2011, pp: 1905-1908. ( EI 收录 )
4. **Zhenjun Han**, Qixiang Ye, Jianbin Jiao<sup>✉</sup>, “Feature Evaluation by Particle Filter for Adaptive Object Tracking”, Proc. of SPIE Visual Communications and Image Processing (VCIP), 2009. ( EI 收录 )
5. **Zhenjun Han**, Qixiang Ye, Jianbin Jiao<sup>✉</sup>, “Online feature evaluation for object tracking using Kalman filter”, Proc. of IEEE International Conference on Pattern Recognition (ICPR), 2008, pp: 1-4. ( EI 收录 )
6. Ce Li, **Zhenjun Han**, Qixiang Ye, and Jianbin Jiao<sup>✉</sup>, “Abnormal Behavior Detection via Sparse Reconstruction Analysis of Trajectory”, Proc. of IEEE International Conference on Image and Graphics (ICIG), 2011, pp: 807-810. ( EI 收录 )
7. Li Li, **Zhenjun Han**, Jianbin Jiao<sup>✉</sup>, Qixiang Ye, “Visual Object Tracking via Sparse Reconstruction”, Proc. of IEEE International Conference on Internet Multimedia Computing and Service, 2010, pp: 151-154. ( EI 收录 )
8. Li Li, **Zhenjun Han**, Qixiang Ye, Jianbin Jiao<sup>✉</sup>, “Visual Object Tracking via One-Class SVM”, ACCV2010 Workshop on Visual Surveillance, 2010, pp: 216-225. ( EI 收录 )

9. Zihua Liu, **Zhenjun Han**, Qixiang Ye, Jianbin Jiao<sup>✉</sup>, “A New Segment-Based Algorithm for Stereo Matching”, Proc. of IEEE International Conference on Mecharonics and Automation (ICMA), 2009, pp: 999-1003. ( EI 收录 )
10. Yang Liu, **Zhenjun Han**, Zhenkai Zhang, Jianbin Jiao<sup>✉</sup>, “Video-Based Hazardous Event Detection in the Driving Environment”, Proc. of International Symposium on Emergency Management (ISEM), 2009, pp: 736-739.
11. 张宇, **韩振军**, 焦建彬, “一种基于综合特征评估的运动目标跟踪算法”, 中国科学技术大学学报 [0253-2778], Vol. 40, No.5, 2010, pp: 491-495.
12. Bo Wu, Jixiang Liang, Qixiang Ye, **Zhenjun Han**, Jianbin Jiao<sup>✉</sup>, “Fast pedestrian detection with laser and image data fusion”, Proc. of IEEE International Conference on Image and Graphics (ICIG), 2011, pp: 605-608.

#### 已投稿论文目录

1. **Zhenjun Han**, Qixiang Ye, and Jianbin Jiao<sup>✉</sup>, “Adaptive Visual Object Tracking via Hierarchical Sparse Representation and Reconstruction”, submitted to IEEE Trans on Circuits and Systems for Video Technology.
2. Ce Li, **Zhenjun Han**, Qixiang Ye, and Jianbin Jiao<sup>✉</sup>, “Abnormal Behavior Detection via Sparse Reconstruction Analysis”, submitted to Neurocomputing, Elsevier.
3. Yifei Liu, **Zhenjun Han**, Qixiang Ye, Jianbin Jiao<sup>✉</sup> and Ce Li, “Object Tracking Based on Sparse Feature Representation”, submitted to Proc. of IEEE International Conference on Acoustics, Speech, and Singal Processing, 2012.

#### 国家发明专利

1. 焦建彬, 叶齐祥, **韩振军**, “基于 Internet 和 GRPS 的智能视频监控系统”, 中国发明专利, 批准日期: 2008 年, 申请号: 2008100851130.



## 致谢

在中国科学院研究生院攻读硕士学位与博士学位的五年半的学习生活中,我经历了诸多坎坷,也付出了艰辛的努力,同时也得到了很大的收获。在毕业论文完成之际,由衷地感谢这五年多来曾经给予我无数帮助的老师、同学、朋友和家人。

首先,本课题的研究工作是在导师焦建彬教授的悉心指导下完成的。焦建彬教授在我的学习中,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,他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,特别要感谢叶齐祥老师在我硕士与博士的学习与生活中对我的极大帮助和支持,他渊博的专业知识、自强不息的学习精神和扎实的动手能力,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。我还要特别要感谢北京航空航天大学张宝昌老师以及 NEC 中国研究院的曾炜老师和朱光宇老师等在科研上对我的指导,他们孜孜不倦的科研态度是我学习的榜样。

再次,感谢同实验室的常向魁、刘志花、徐冉、李莉、刘一飞、李策等同学,他们在我的论文工作和撰写中,给予了我巨大的支持和帮助。

我还要衷心感谢我敬爱的亲人们,尤其是我的母亲以及已故的父亲,同时特别感谢我的未婚妻崔少娟,她对论文的校对与修改给予了我许多的帮助。亲人们对我工作和学业的热情鼓励和殷切期望,都给我注入了巨大的动力,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,我将对此终生铭记。

本项目的完成离不开国内外的技术资料,在此对参考文献的作者、译者以及出版单位表示感谢!

最后,再次向在学习、工作和生活中给予过自己关心、支持与鼓励的所有老师、同学、朋友们表示最诚挚的谢意!

**韩振军**

**2011年9月**