

分类号\_\_\_\_\_

密级\_\_\_\_\_

UDC \_\_\_\_\_

编号\_\_\_\_\_

# 中国科学院研究生院

## 硕士学位论文

基于样本空间目标重构的视频跟踪算法研究

李 莉

指导教师\_\_\_\_\_焦建彬 教授 博导\_\_\_\_\_

\_\_\_\_\_中国科学院研究生院\_\_\_\_\_

申请学位级别工学硕士 学科专业名称计算机应用技术

论文提交日期2011年4月 论文答辩日期2011年5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

视频图像中的运动目标跟踪作为计算机视觉领域的核心研究课题之一，经过近五十多年的发展得到了广泛而深入的研究。它融合了图像处理、模式识别、人工智能、自动控制等相关领域的研究成果，在视频监控、机器人导航、视频传输、视频检索、医学图像分析、气象分析等许多领域有着广泛的应用。本文以此为研究内容，具有重要的理论意义和广阔的应用价值。

视频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，所以我们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重构得到。本文提出了基于样本空间目标重构的跟踪算法，即从搜索区域的样本集中获取目标重构系数的跟踪算法。首先，在视频上一帧的跟踪目标周围建立搜索区域，在搜索区域内按照一定的规则构建样本集；然后，在获得的样本集的基础上，通过使用一类支持向量机（One-Class SVM）和稀疏表示（Sparse Representation）对样本集中的样本进行评估与选择，选择出具有代表性的样本重构目标，从而实现目标的跟踪。因为在样本空间的目标重构中，我们选择几个包含目标某一部分的候选样本作为重构目标的基，因此基于样本空间目标重构的跟踪算法具有较强的鲁棒性，在算法介绍的最后我们对提出的基于 One-Class SVM 和稀疏表示的目标重构跟踪算法进行了实验验证。

（1）基于 One-Class SVM 目标重构的跟踪算法。一类支持向量机应用于只包含一类样本的情况，在高维特征空间中构造一个尽可能多的包含样本的尽可能小的超球体，从而描述一类样本数据概率分布的二值模型。在目标跟踪下，我们将样本集作为一类样本，使用一类支持向量机进行模型的在线训练，从而对样本集中的样本进行评估与选择，最终重构目标以实现目标跟踪。

（2）基于稀疏表示目标重构的跟踪算法。稀疏表示使目标重构系数集中于尽可能少的样本上，这些样本揭示着跟踪目标的主要特征和内在结构，从而实现目标跟踪的鲁棒性。

**关键词：**目标跟踪，目标重构，一类支持向量机，稀疏表示

## Abstract

Li Li(Computer Application)

Directed by: Jiao Jian-bin(Professor)

Moving object 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.

Object tracking essentially lies in finding the object in searching area in video frame, so we could make a reasonable assumption that the tracking object can be obtained by reconstruction samples in this searching area. In this paper, we propose a new visual object tracking method based on object reconstruction of sample space. First of all, a searching box is built around tracking target in the pre-frame of video, and the sample set can be obtained in this searching box by some rule; then we make an evaluate and choice to each sample in sample set using One-Class SVM or Sparse Representation, and select the representative samples to reconstruct the target in new frame. These representative samples from searching box must be including some part of the target, so when reconstructing the new target using these samples, we have the robust tracking from object or background variation. At the end of the algorithm, we make experiments to verify the tracking result for proposed One-Class SVM and Sparse Representation tracking method.

(1) Object tracking algorithm based on object reconstruction via One-Class SVM. Different from SVM, One-Class SVM is usually used in these cases that samples are almost from one class, and One-Class SVM can construct as small hyper-sphere as possible that contains as many samples as possible in high dimensions of feature space, which describes a binary model of probability distribution of the samples. In the object tracking of our paper, the sample set can be seen from one class, and we can obtain the training model by One-Class SVM to evaluate each sample in sample set, then reconstruct the object for tracking.

(2) Object tracking algorithm based on object reconstruction via Sparse Representation. Sparse representation concentrates the signal energy in several elements, which can reflect main features and internal structure of the signal. In the

object tracking algorithm, we use several samples to describe the tracking object, so as to achieve the tracking robustness.

**KEY WORDS: Object Tracking, Object Reconstruction, One-Class SVM, Sparse Representation**

## 目 录

摘 要 .....	I
目 录 .....	IV
图目录 .....	VI
表目录 .....	VII
<b>第一章 绪论</b> .....	1
<b>1.1 课题背景和研究意义</b> .....	1
<b>1.2 视频跟踪的国内外研究现状</b> .....	2
<b>1.3 视频跟踪的主要研究内容和分类</b> .....	5
<b>1.3.1 视频跟踪的研究内容和难点</b> .....	5
<b>1.3.2 视频跟踪的研究分类</b> .....	6
<b>1.4 本文的研究内容</b> .....	7
<b>1.5 本文的组织结构</b> .....	8
<b>第二章 基于样本空间目标重构的视频跟踪算法研究</b> .....	11
<b>2.1 视频目标跟踪的主要算法</b> .....	11
<b>2.1.1 基于目标表达和相似性度量的目标跟踪算法</b> .....	11
<b>2.1.2 基于滤波预测的目标跟踪算法</b> .....	15
<b>2.1.3 基于优化匹配搜索的运动目标跟踪算法</b> .....	17
<b>2.2 基于目标重构的跟踪算法简要介绍</b> .....	20
<b>第三章 基于ONE-CLASS SVM目标重构的跟踪算法</b> .....	22
<b>3.1 ONE-CLASS SVM的基本理论</b> .....	22
<b>3.1.1 ONE-CLASS SVM的提出</b> .....	22
<b>3.1.2 ONE-CLASS SVM的基本理论</b> .....	25
<b>3.2 基于ONE-CLASS SVM的样本重构目标跟踪算法</b> .....	28
<b>3.2.1 基于搜索区域跟踪目标的样本集构造</b> .....	29
<b>3.2.2 ONE-CLASS SVM对样本集中样本再选择</b> .....	30
<b>3.2.3 基于ONE-CLASS SVM目标重构的跟踪算法</b> .....	32
<b>3.3 基于ONE-CLASS SVM目标重构的跟踪实验</b> .....	33
<b>3.3.1 实验步骤</b> .....	33
<b>3.3.2 视频测试集合及跟踪算法评估</b> .....	34
<b>3.3.3 跟踪实验结果展示及实验分析</b> .....	34



<b>第四章 基于稀疏表示目标重构的跟踪算法</b> .....	39
<b>4.1 稀疏表示的基本理论</b> .....	39
<b>4.1.1 稀疏表示的概念</b> .....	39
<b>4.1.2 信号的稀疏表示</b> .....	39
<b>4.1.3 稀疏表示的求解方法</b> .....	41
<b>4.1.4 稀疏表示在图像/视频处理领域的应用</b> .....	44
<b>4.2 基于稀疏表示目标重构的跟踪算法</b> .....	45
<b>4.2.1 目标跟踪的样本集构造</b> .....	45
<b>4.2.2 基于稀疏表示的目标重构</b> .....	45
<b>4.2.3 基于稀疏表示目标重构的跟踪</b> .....	47
<b>4.3 基于稀疏表示的目标重构的跟踪实验与算法评估</b> .....	47
<b>4.3.1 实验步骤</b> .....	47
<b>4.3.2 跟踪实验结果展示及实验分析</b> .....	48
总结与展望 .....	51
参考文献 .....	53
发表文章目录 .....	60
附录 1: 主要符号对照表 .....	61
致 谢 .....	62

## 图目录

图 1-1 目标跟踪示意图 .....	5
图 1-2 运动目标的跟踪框图 .....	7
图 1-3 本文的目标跟踪示意图 .....	7
图 2-1 Mean Shift示意图 .....	18
图 3-1 SVM最优分类超平面 .....	22
图 3-2 One-Class SVM的二维示意图 .....	25
图 3-3 One-Class SVM的超球面分类 .....	26
图 3-4 (a) 搜索区域 (b) 一个样本 (c)样本集合中部分样本示例 .....	29
图 3-5 HOG特征提取 .....	30
图 3-6 跟踪目标与样本集中的 100 个样本的相似度折线图 .....	31
图 3-7 OC-SVM超球面，小黑点表示相关样本，小圆圈表示超球面的球心 .....	32
图 3-8 使用基于One-Class SVM的目标跟踪的两个试验实例 .....	35
图 3-9 我们的算法与经典Mean-Shift算法的两个试验跟踪效果比较 .....	35
图 3-10 三种跟踪算法的统计实验结果图 .....	37
图 4-1 稀疏表示模型 .....	40
图 4-2 (a) 跟踪目标 (b) 搜索区域 (c)稀疏重构系数 .....	46
图 4-3 两个跟踪实例 .....	48
图 4-4 两个跟踪实例 .....	49
图 4-5 五种跟踪方法平均DER .....	50

## 表目录

表 3-1 基于One-Class SVM目标重构的跟踪算法.....	33
表 3-2 用于计算DER的视频文件列表 .....	36
表 4-1 基于稀疏表示目标重构的跟踪算法 .....	47



## 第一章 绪论

### 1.1 课题背景和研究意义

基于视频的运动目标跟踪技术（视频跟踪或目标跟踪）是计算机视觉领域的核心研究课题之一，越来越吸引着国家科研机构和研究人员的注意。目标跟踪是依据人类生理视觉系统发展起来的，人类的视觉系统能够通过眼睛捕捉现实的事件序列，并将这些信息转化为信号传输到大脑的视觉处理中枢，通过智能运动分析和运动感知功能，检测、识别出目标，判断目标的位置及随时间的运动变化，预测估计目标的变化趋势。计算机视觉领域的目标跟踪模拟这个过程，它使用摄像头捕捉图像序列，通过一系列的计算和分析，得出运动目标在每一帧图像中的二维坐标位置，然后，将图像序列中同一运动目标关联起来，得到目标的完整运动轨迹。简单的说，目标跟踪技术就是在视频连续帧中，找到下一帧图像中目标的确切位置，为视频序列分析提供基础运动信息和数据。

视频跟踪技术融合了图像处理、模式识别、人工智能、自动控制等众多领域中的先进技术和核心思想。与传统的雷达系统相比，视频跟踪系统采用摄像机或者其它光学系统等比较廉价的设备，并且采用被动式工作模式，工作时不向外辐射无线电波，不易被敌方的电子侦察设备发现，因此具有更高的性价比、一定的隐蔽性和抗电子干扰的能力。“同时，视频跟踪系统，可以从视频监控器上直接看到运动目标图像，获得关于运动目标更丰富的信息，因此具有很好的直观性和可靠性。

视频跟踪技术经过近多年的深入研究和发 展，已经广泛的应用于军事制导、飞机导航、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。可见光跟踪和红外跟踪技术已应用于导弹制导、火炮控制、空间飞行体和靶场测量。视频监控技术已在学校、医院、工作场所、仓库、军事基地等社区、公共场所、以及重点区域使用，实时监控视频内的内容，检测跟踪重点目标，及时作出响应；车辆实时检测、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；实时自动跟踪视频下细胞的损伤修复、细胞迁徙的过程，为医学上探讨各种疾病用药对细胞

损伤修复的影响提供有力的工具。另外，视频跟踪技术在视频会议、视频分析、视频检索、医学图像分析、零部件质量检测等许多领域也有大量应用。

目前，视频跟踪技术不断进步和发展，提出了大量的目标跟踪算法，但是还远远没有成熟，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很大的变化，这些变化主要来自于四个方面：

- 1) 运动目标的尺度伸缩发生变化；
- 2) 运动目标姿态和形状发生变化；
- 3) 运动目标被其他目标或者背景物体部分或者完全遮挡；
- 4) 环境中光线亮度发生变化。

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，但这些算法大多是基于目标某一种特征集合的跟踪，对目标特征描述不够完备，并且无法自动适应目标和环境的大变化，当背景发生较大变化时，往往会失效。因此，需要我们开发出鲁棒的、适应目标变化的目标跟踪算法。

作者在论文工作期间，参与了以下课题：

- 1) “复杂环境下动态目标跟踪技术”中科院“百人计划”择优支持项目（项目编号：99T3009EA2）；
- 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）；
- 3) “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）。

## 1.2 视频跟踪的国内外研究现状

视频跟踪在国外起步比较早，理论研究和系统设计方面都受到很多政府机构和学者的密切关注。在视频目标跟踪的理论研究方面，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，多次资助有关公司进行相关算法研究；美国国防部、美国陆军 SBIR 项目；美国海军水下战事中心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。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 进行目标智能识别和跟踪等相关研究。国外设计视频目标跟踪的原型系统也很多，主要包括：美国 DARPA 设立的 VSAM 项目；Maryland 大学的实时视频监控 W4；欧洲委员会自主的 ADVISOR 项目；日本也开展了 CDVP 计划等。

国内外的众多研究机构对视频目标跟踪也展开了深入的研究。如法国国家实验室 (INRIA)、牛津大学动态视觉组、MIT 媒体实验室、CMU 机器人研究所、马里兰大学帕克学院、DCS of Maryland, CSD of U. Southern California, NEC Research, CSC of U. Reading, CS of UC Berkeley 等。许多国际学术会议和期刊对于这一领域的最新研究成果进行了专门的讨论，如：Computer Vision and Pattern Recognition(CVPR)、International Conference on Computer Vision(ICCV)、European Conference on Computer Vision(ECCV)、Asian Conference on Computer Vision(ACCV)、International Conference on Pattern Recognition(ICPR)、International Journal of Computer Vision(IJCV)、IEEE Transaction on Pattern Analysis and Machine Intelligence(PAMI)、IEEE Transaction on Image Processing、Pattern Recognition、Image and Vision Computing 等。

随着图像处理技术和图像传感器技术的飞速发展，在国内的中科院自动化所，清华大学，哈尔滨工业大学，西安交通大学，浙江大学，上海交通大学等一些高校和科研院所也逐渐开展视频目标跟踪的研究，并且已经取得了喜人的成绩。如：中科院计算所 JDL 实验室对于体育视频中的目标分割、检测、跟踪、和精彩片断分析进行了深入的研究[9][13]；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视觉分析系统、交通行为事件分析系统、交通场景监控系统、和智能轮椅视觉导航系统[16]；中国科学院自动化研究所模式识别国家重点实验室谭铁牛研究员从 20 世纪 90 年代初就开始研究基于三维模型的智能视觉监控系统，领导视觉监控小组，取得了一系列的研究成果；清华大学开发了自然环境中运动检测、跟踪和全景图生成等功能的视觉侦查系统；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深入研究了自适应巡航控制中的车辆跟踪；上海交通大学与上海市交警总队合作开发的 SUATS 系统。

近年来，在视频目标跟踪的研究中科研人员开发了大量新算法[1-31]，同时传统的跟踪方法也不断得到改进[16]。这些算法可以分为两大类：确定性方法

和随机方法。确定性方法通常将跟踪问题简化为一个最优化问题，如何定义好这个最优化函数是确定性方法的关键。随机跟踪方法的基本思想是，已知当前时刻以及之前时刻的所有参数度量，对目标当前状态进行估计。

基于光流场的目标跟踪是一种典型的确定性跟踪方法。对目标进行跟踪时，根据预先设定好的相似性度量函数，在当前视频帧的某个区域内，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。在这种类型的算法中，误差平方和（Sum of Squared Difference, SSD）方法是最广泛使用的与目标模板匹配相似性度量函数，SSD方法虽然能够很好的反映目标的细节，但是当目标的外观发生较大变化时，SSD方法会失败。Mean-Shift目标跟踪方法，以先前目标位置为初始位置，利用无参数概率密度梯度，搜索与目标模板颜色直方图最类似的窗口作为当前目标位置。Mean Shift等优化算法被广泛用于寻找局部最优区域[17-22]。但是它对带宽的选择和跟踪过程初始化的苛刻要求限制了应用。Bouttefroy等人使用投影Kalman滤波的Mean Shift改进了传统Mean Shift的不足[22]。在目标表示方面，将外观、颜色分布、轮廓信息等先验知识用于建立目标模板，从而使跟踪更为鲁棒、有效。其中基于表观的方法可以统计估计目标在特征空间或核空间内的分布规律[17-20]。

随机跟踪方法采用预测滤波方法，模型化跟踪目标的运动状态，估计目标运动的状态方程。当运动模型为线性或高斯分布并且观察模型也为线性时，采用Kalman滤波方法可以获得最优。当状态空间为非线性、非高斯时，目前没有找到能够准确解析描述的模型，经常采用马尔科夫滤波进行目标跟踪，其中最常用的是粒子滤波的方法。粒子滤波，也称时序蒙特卡罗[20][27]，通过使用状态空间的加权随机采样表示目标状态的概率分布，计算最大后验概率获得最优目标状态，它是目前应用最为广泛的算法[22]。随机跟踪方法不再假设跟踪系统输入与输出的对应确定性，融合了概率统计的思想，把如概率密度函数的不确定性观察和不同状态进行联系。随机跟踪方法更适用于大多数的视觉目标跟踪，因为在跟踪过程中，目标以及环境相互之间存在遮挡、不断变化等很多不确定性的因素，通过使用随机性的跟踪方法达到了更加鲁棒的效果。

除此之外，最近几年出现的一种新的目标跟踪方法，方法中跟踪被看作是一个跟踪目标和背景作为两个类别的分类问题，通过使用SVM、Adaboost等常用的高效分类方法在训练集上进行训练得出模型，在每一帧中利用分类器分离目标和背景，从而达到跟踪目标的目的。基于分类的目标跟踪方法避开了目标



状态估计的复杂过程，通过分类达到很高的跟踪精度。然而这种方法需要大量的正例反例样本的学习构造分类器，因为跟踪环境的不同样本的选择具有特殊性，这个成为了使用的一个难点，同时需要设计分类器和对样本进行训练，不能满足跟踪实时的需要。

### 1.3 视频跟踪的主要研究内容和分类

#### 1.3.1 视频跟踪的研究内容和难点

在视频运动目标跟踪中，我们通过使用摄像头 CCD 捕捉视频图像序列，将现实世界中 3D 的立体物体简化为图像中的平面 2D 的物体，接着对跟踪目标进行初始化，初始化时可以采用目标检测的方法检测目标的初始位置，或者通过手动标注表明我们需要跟踪的目标，最后才开始我们真正的跟踪过程。这个过程如下图所示。

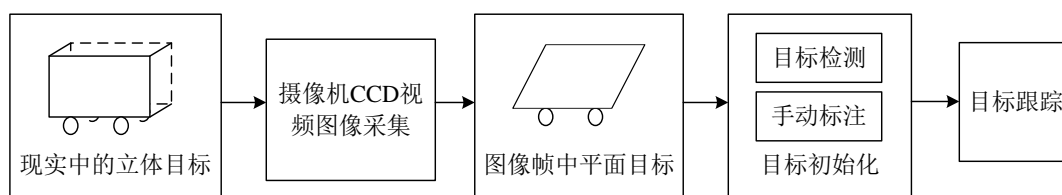


图 1-1 目标跟踪示意图

视频目标跟踪通过一系列的计算和分析，得出运动目标在每一帧图像中的确切位置。在整个目标跟踪的视频采集、目标初始化、和跟踪过程中，存在一些重点难题：

- 1) 目标图像的三维重建。三维空间的运动目标变换为二维图像，损失了很多重要的深度信息，如果能够重建三维信息，目标跟踪的效果会更加准确；
- 2) 摄像机的成像影响。在摄像机 CCD 进行采集过程中，获得的视频图像受到目标的材质纹理、外界光源、摄像机的感知功能、空气透明度等因素的影响，并且图像像素与这些因素之间的依赖关系是非线性的；
- 3) 目标的初始化影响。在目标跟踪进行之前，需要通过目标检测算法、或者手动标注来获得目标的初始化位置和内容，因此目标跟踪的效果与目标检测算法的有效性、或者手动标注的准确性存在很大的相关性，当目标检测出现错误、手动标注中的背景区域过多，都会严重影响跟踪的效果；
- 4) 运动目标的尺度伸缩变化和形变。在跟踪的过程中，目标的旋转、偏移

等运动变化、摄像机焦距的变化，都会引起造成摄像机采集到的目标图像的尺度伸缩和目标图像表达的变化，从而容易引起跟踪算法的失效；

5) 目标的部分及完全遮挡。背景对目标的遮挡、以及目标之间的相互遮挡是视频目标跟踪中的一个重大难题，遮挡会造成目标图像表达突然发生严重变化，这种变化难以预测和估计，具有突发性和不连续性，容易引起跟踪算法的失效；

6) 场景光线亮度变化。同一目标在不同的光线亮度条件下会产生有明显差异的目标图像，不同光线对于同样的摄像机采集到的目标图像也存在明显差异，这个影响是很多室外应用系统中阻碍性能提升的瓶颈；

7) 运动目标复杂的运动模式。跟踪算法采用滤波的方法进行目标运动的预测，然后在运动目标可能出现的区域内进行搜索，当目标运动比较复杂，如目标的运动速度或者方向突然发生较大的改变，运动目标不在预测的搜索区域内，这时候简单的滤波预测算法会失效，从而无法跟踪目标。

### 1.3.2 视频跟踪的研究分类

目标跟踪的算法非常多，按照不同的标准有很多不同的分类，常常让人难以明白是如何的跟踪算法，通过详细的调查和细分，目标跟踪的分类标准有：

1) 根据运动目标和背景的相对运动关系，目标跟踪算法可以分为：静止背景的静止目标跟踪算法、静止背景的运动目标跟踪算法、动态背景的静止目标跟踪算法、动态背景的运动目标跟踪算法。本文提到的跟踪算法均为基于运动目标或者运动背景的跟踪算法。

2) 根据跟踪目标的个数不同，视频跟踪算法可以分为单目标跟踪算法和多目标跟踪算法。本文主要研究单目标跟踪算法，但基于单目标的跟踪方法很容易扩展到多目标跟踪中。

3) 根据跟踪目标属性的不同，视频跟踪算法可以分为：刚体目标跟踪算法和非刚体目标跟踪算法。

4) 根据跟踪目标的表达方法和相似性度量方法不同，视频跟踪算法可以分为基于轮廓的跟踪算法、基于特征的跟踪算法、基于区域统计特性的跟踪算法和基于模型的跟踪算法等。

现在大部分的算法是基于动态背景或者动态目标、单目标、刚体与非刚体一体的研究，这些算法的跟踪框图如下图所示。这些跟踪算法中，部分是针对提高跟踪匹配的准确性而提出的，这类算法的主要是通过选取稳定的、可区分

性好的特征，从而提高目标的搜索速度和匹配的正确度；部分算法是针对缩小目标搜索范围提出的，这类算法主要是通过滤波方法预测目标下一时刻可能出现的位置，从而缩小目标的搜索范围来缩短目标搜索时间。

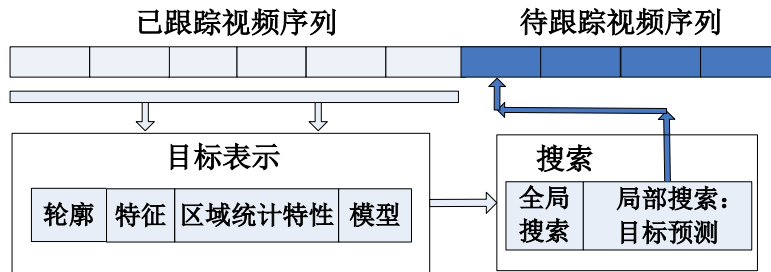


图 2-2 运动目标的跟踪框图

现有的视频目标跟踪算法，在一定程度上完成了目标的跟踪，但是都存在跟踪精度不高、特征提取匹配的计算复杂度太大等问题，难以满足现实对跟踪算法的要求。如何提高目标跟踪算法的实时性、鲁棒性和精确性，是目前视频跟踪算法研究的热点和难点。为了开发一种鲁棒的目标跟踪算法，解决运动目标的尺度伸缩、部分及完全遮挡、运动目标背景的形变以及光线亮度变化等问题，达到对目标长时间的鲁棒性跟踪，本文提出的算法采用了基于综合颜色直方图和梯度方向直方图的综合统计特征，并研究了一类支持向量机和稀疏表示，提出了基于这两种方法的目标跟踪算法，对目标的尺度伸缩、遮挡以及环境的光线变化等都具有很好的鲁棒性。本文跟踪算法的研究框图如下：

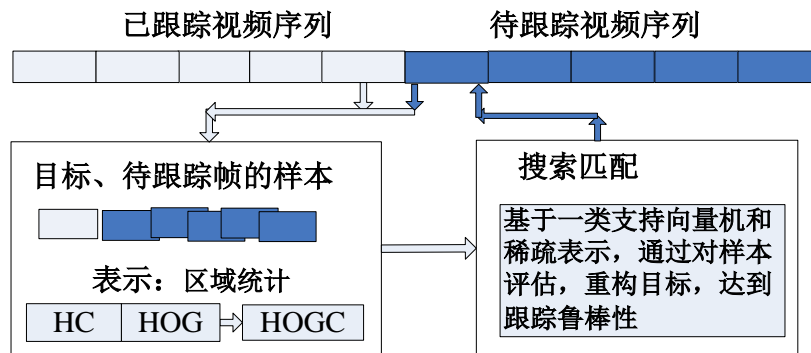


图 3-3 本文的目标跟踪示意图

## 1.4 本文的研究内容

在对目标跟踪算法和应用研究的基础上，本文主要提出了基于样本空间目标重构的跟踪算法，该算法通过分类器、稀疏表示等方法对样本集中的样本进

行选择和评估，使用样本的评价系数来重构目标，从而达到目标跟踪。在目标尺度伸缩、局部遮挡以及环境光线变化等复杂环境中，目标重构的跟踪方法仍然具有较好的鲁棒性。基于样本空间目标重构的跟踪算法中，我们着重研究两种具体算法：基于 One-Class SVM 目标重构的跟踪算法、和基于稀疏表示目标重构的跟踪算法。

- 基于 One-Class SVM 目标重构的跟踪算法：一类支持向量机通过对一种类型的样本进行回归，获得样本集中的样本的选择和评估系数，从而重构目标进行跟踪。基于这个思路，我们提出了基于 One-Class SVM 目标重构的跟踪新方法。算法中，首先通过计算选择搜索区域，并在搜索区域内按照一定的规则构建样本集，接着使用巴氏系数进行相似度度量在样本集上进行样本再选择，最后我们使用一类支持向量机对选择出的样本进行评估，得到样本的重构系数，重构目标从而进行跟踪。这种目标重构的跟踪方法，因为使用具有代表性的样本重构目标，从而保证了物体因为旋转等变化的跟踪效果，提高了跟踪的稳定性和鲁棒性。通过大量的实验验证了该方法的有效性。
- 基于稀疏表示目标重构的跟踪算法：由于视频搜索区域（样本集构造区域）往往要比跟踪目标大很多，导致使用样本集进行样本重构的系数会非常的稀疏，背景区域内的样本对应的重构系数趋向于 0，因此我们提出了即基于稀疏表示获取样本稀疏系数来重构目标的跟踪方法。这种方法不仅仅保留了一类支持向量机对于样本评估的优势，而且稀疏表示能够更好的对所有样本进行评估，从而选择出最有代表性的几个样本来重构目标，使目标跟踪更加鲁棒。本文通过大量实验验证了算法的有效性。

## 1.5 本文的组织结构

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

第二章，基于样本空间目标重构的跟踪。主要综述了视频跟踪算法的主要研究内容和方法，同时也对基于样本空间目标重构的跟踪算法的基本思想进行简单介绍。

第三章，基于 One-Class SVM 目标重构的跟踪算法。主要论述了 One-Class SVM 基本知识和用于目标跟踪的原理方法，详细介绍了 One-Class SVM 目标重构的跟踪算法。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、部分及完全遮挡和部分形变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，基于稀疏表示目标重构的跟踪算法。主要论述了稀疏表示在处理信号、图像等领域的稀疏优势，以及求解方法。详细介绍了基于稀疏表示获取目标重构系数，从而在视频中跟踪目标的方法。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、部分及完全遮挡和部分形变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。

最后是对现有工作的总结和对未来工作的展望，以及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的探讨。



## 第二章 基于样本空间目标重构的视频跟踪算法研究

视频跟踪是计算机视觉领域的热点研究问题之一。目前，针对如何提高运动目标跟踪算法的实时性、鲁棒性和准确性等方面的研究逐步成为视觉跟踪研究的热点。本章将主要介绍视频跟踪的一些传统方法，并对基于样本空间目标重构的跟踪算法思想进行简单介绍。在接下来的第三章和第四章，我们将详细介绍基于 One-Class SVM 和稀疏表示的运动目标跟踪方法。

### 2.1 视频目标跟踪的主要算法

经过二十多年的发展，视频跟踪得到了充分的发展，提出的算法越来越多，它们的思想通常有两种：通过识别目标来进行跟踪和通过运动目标检测来进行跟踪。前一种算法思想是通过识别每一帧的图像中跟踪目标，来确定该运动目标的位置，从而实现跟踪，该算法思想包含了目标识别和目标匹配两部分；而后一种跟踪算法思想通过检测，发现运动目标并确定运动目标的位置进行跟踪，该方法不需要考虑目标的形状、尺度，可以检测任何目标。

#### 2.1.1 基于目标表达和相似性度量的目标跟踪算法

根据跟踪目标的表达和相似性度量，目标跟踪算法可以分为：基于轮廓的跟踪算法[67]、基于特征的跟踪算法[30]、基于区域统计特征的跟踪算法[32][33]和基于模型匹配的跟踪算法[17][66]。

基于轮廓的跟踪算法通过对目标的分割和边缘提取实现对运动目标的跟踪。然而，当目标被部分及全部遮挡时，跟踪算法将受到很大的限制。基于模型的跟踪算法采用目标的物理知识而不仅仅是目标的图像信息建立模型，因此基于 3D 模型跟踪算法是比较完善的。该方法的局限在于无先验知识的实际跟踪系统中很难得到目标的模型。基于特征表达的跟踪算法通过提取具有尺度伸缩、旋转等不变性的特征点（如 SIFT、KLT、Harris Corners、SUSAN Corners 等），解决尺度伸缩和部分遮挡等问题。SIFT[30]特征点主要是通过图像金字塔的多尺度极值点检测得到的，用梯度方向进行描述，因此这种特征点具有尺度和旋转不变性等优点，在图像配准和跟踪方面具有很高的匹配精度和鲁棒性。

然而,特征自身的检测和提取往往计算复杂难以满足跟踪系统实时处理的要求。基于区域统计特性的跟踪算法利用的是目标图像的全局描述信息,很难精确定位目标的位置。同时,当背景或者其他目标的区域统计特性相似的时候,跟踪算法也会失效(如基于颜色或者灰度直方图的跟踪算法)。

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。

基于区域统计特征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,[68]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,基于区域统计特征的跟踪方法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。

在目标跟踪中，特征搜索匹配就是在所得到的目标和跟踪目标之间进行特征相似度度量。一般使用向量的形式来表示特征，相似度度量常就是计算两个向量之间的距离。在图像处理技术中，常用的相似性度量方法有欧氏距离、街区距离、棋盘距离、加权距离、巴特查理亚系数、Hausdorff 距离等。其中，欧氏距离由于其简便性，是使用最广泛的相似度衡量；Hausdorff 距离通常用于衡量特征集之间的相似性；直方图之间的相似性通常采用巴特查理亚系数。

1) 欧氏距离。已经二维平面上的两个点  $P_1, P_2$ ，坐标分别为  $P_1(x_1, y_1), P_2(x_2, y_2)$ ， $P_1, P_2$  之间的欧式距离为坐标向量的平方差之和的开方，公式如下所示：

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2-1)$$

2) 加权距离。已经二维平面上的两个点  $P_1, P_2$ ，坐标分别为  $P_1(x_1, y_1), P_2(x_2, y_2)$ ， $P_1, P_2$  之间的加权距离为坐标向量的每一项的绝对距离的加权和，公式如下所示：

$$d(P_1, P_2) = \begin{cases} \omega_0 |x_1 - x_2| + \omega_1 |y_1 - y_2| & \text{if } |x_1 - x_2| > |y_1 - y_2| \\ \omega_0 |y_1 - y_2| + \omega_1 |x_1 - x_2| & \text{others} \end{cases} \quad (2-2)$$

其中， $\omega_i > 0, i = 1, 2$ 。街区距离、棋盘距离是加权距离特殊情况，当  $\omega_0 = 1, \omega_1 = 1$  时，加权距离就是街区距离，即坐标向量每一项的绝对距离之和；当  $\omega_0 = 1, \omega_1 = 0$  时，加权距离就是棋盘距离，即坐标向量第一项的绝对距离；当  $\omega_0 = 1, \omega_1 = \sqrt{2} - 1$  时，加权距离就变成类欧式距离。

3) 巴特查理亚系数，简称 Bhattacharyya 系数或巴氏系数。已经两个向量点  $P_1, P_2$ ，坐标分别为  $P_1(x_1, x_2, \dots, x_m), P_2(y_1, y_2, \dots, y_m)$ ， $P_1, P_2$  之间的巴氏系数为坐标向量的每一项的乘积的开方，几何意义是两个  $m$  维向量  $P_1(x_1, x_2, \dots, x_m), P_2(y_1, y_2, \dots, y_m)$  夹角的余弦。公式如下所示：

$$d(P_1, P_2) = \sum_{i=1}^m \sqrt{x_i \cdot y_i} \quad (2-3)$$

4) Hausdorff 距离 (Hausdorff Distance, HD)。HD 距离是有限像素点集合之间的一种距离，度量的是两个特征点集之间的相似度。已知两组有限像素点集合  $A = \{a_1, a_2, \dots, a_p\}$  和  $B = \{b_1, b_2, \dots, b_q\}$ ，则 HD 距离定义如下：

$$D(A, B) = \max\{d(A, B), d(B, A)\} \quad (2-4)$$

其中， $d(A, B) = \max_{a_i \in A} \min_{b_j \in B} \|a_i - b_j\|$ ，是点集  $A = \{a_1, a_2, \dots, a_p\}$  到点集  $B = \{b_1, b_2, \dots, b_q\}$  的有向 HD，表示点集  $A$  中所有点到点集  $B$  的距离  $d_B(a_i)$  的最大值。定义  $d_B(a_i) = \min_{b_j \in B} \|a_i - b_j\|$ ，表示点  $a_i$  到集合  $B$  的距离，是点  $a_i$  到集合  $B$  中每一点距离的最小值。这里  $\|\cdot\|$  表示某种范式，根据使用特征的不同，HD 使用不同的范式形式，灰度特征的 HD 主要使用均方误差函数、绝对差累加和函数、归一化互相关函数等；彩色特征的 HD 使用 Bhattacharyya 系数等。

5) 改进的 Hausdorff 距离。HD 的改进算法主要是对有向 HD 的定义进行修正，如部分 HD 距离和基于平均距离值的 MHD。

部分 HD 是指先求出点集  $A$  中所有点到点集  $B$  的距离，然后将这些距离由小到大排序，其中序号  $k$  的距离即为  $h_k(A, B)$ 。部分 HD 对目标存在严重遮挡或图像退化时可以获得较好效果。部分 HD 的有向 HD 定义如下：

$$d_k(A, B) = K_{a_i \in A}^{th} d_B(a_i) \quad (2-5)$$

基于平均距离值的 MHD，是指先求出点集  $A$  中所有点到点集  $B$  的距离，然后选择这些距离的平均值，MHD 的公式表示如下：

$$d_{MHD}(A, B) = \frac{1}{N_A} \sum_{a_i \in A} d_B(a_i) \quad (2-6)$$

部分 HD 和平均距离值的 MHD 的融合距离，将上面两种方法融合在一起，它的公式表示如下：

$$d_{LTS}(A, B) = \frac{1}{k \times N_A} d_B(a_i)_{(m)} \quad (2-7)$$

### 2.1.2 基于滤波预测的目标跟踪算法

使用预测滤波方法对目标进行跟踪是常用的跟踪方法，包括 Kalman 滤波、扩展 Kalman 滤波、和粒子滤波等。当运动模型为线性或高斯分布并且观察模型也为线性时，采用 Kalman 滤波方法可以获得最优。当状态空间为非线性、非高斯时，目前没有找到能够准确解析描述的模型，经常采用马尔科夫滤波进行目标跟踪，其中最常用的是粒子滤波的方法。粒子滤波，也称时序蒙特卡罗 [20][27]，通过使用状态空间的加权随机采样表示目标状态的概率分布，通过计算最大后验概率获得最优目标状态。这里将我们主要介绍 Kalman 滤波算法。

非线性的状态方程和观测方程可以表示为：

$$\begin{aligned} X_k &= A \times X_{k-1} + B \times u_{k-1} + w_{k-1} \\ z_k &= H \times X_k + v_k \end{aligned} \quad (2-8)$$

$w$  和  $v$  分别为独立不相关的状态噪声和观测噪声，分布满足：

$$\begin{aligned} p(w) &\sim N(0, Q) \\ p(v) &\sim N(0, R) \end{aligned} \quad (2-9)$$

状态预测方程（先验估计）：

$$\hat{X}(k|k-1) = A \times \hat{X}(k-1|k-1) + B \times u_{k-1} \quad (2-10)$$

经观测值修正过的后验估计：

$$\hat{X}(k|k) = \hat{X}(k|k-1) + K_k (z_k - H \hat{X}(k|k-1)) \quad (2-11)$$

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：

$$\begin{aligned} \tilde{X}(k|k-1) &= X(k) - \hat{X}(k|k-1) \\ \tilde{X}(k|k) &= X(k) - \hat{X}(k|k) \end{aligned} \quad (2-12)$$

先验估计误差和后验估计误差的协方差分别为：

$$P(k|k-1) = E \left[ \tilde{X}(k|k-1) \tilde{X}^T(k|k-1) \right] \quad (2-13)$$

$$\begin{aligned}
 P(k|k) &= E\left[\widetilde{X}(k|k)\widetilde{X}^T(k|k)\right] \\
 &= E\left[(X(k) - \widehat{X}(k|k))(X(k) - \widehat{X}(k|k))^T\right] \\
 &= E[(X(k) - \widehat{X}(k|k-1) - K(z_k - H\widehat{X}(k|k-1))) \\
 &\quad (X(k) - \widehat{X}(k|k-1) - K(z_k - H\widehat{X}(k|k-1)))^T] \\
 &= P(k|k-1) - P(k|k-1)H^TK^T - KHP(k|k-1)^T \\
 &\quad + K(HP(k|k-1)H^T + R)K^T
 \end{aligned} \tag{2-14}$$

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

$$K_k = P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1} \tag{2-15}$$

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

$$\begin{aligned}
 \lim_{R \rightarrow 0} \widehat{X}(k|k) &= \widehat{X}(k|k-1) + H^{-1}(z_k - H\widehat{X}(k|k-1)) = z_k \\
 \lim_{P(k|k-1) \rightarrow 0} \widehat{X}(k|k) &= \widehat{X}(k|k-1) + 0 \times (z_k - H\widehat{X}(k|k-1)) \\
 &= \widehat{X}(k|k-1)
 \end{aligned} \tag{2-16}$$

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

预测部分：

$$\begin{aligned}
 \widehat{X}(k|k-1) &= A\widehat{X}(k-1|k-1) + Bu_{k-1} \\
 P(k|k-1) &= AP(k-1|k-1)A^T + Q
 \end{aligned} \tag{2-17}$$

修正部分：

$$\begin{aligned}
 K_k &= P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1} \\
 \widehat{X}(k|k) &= \widehat{X}(k|k-1) + K_k(z_k - H\widehat{X}(k|k-1)) \\
 P(k|k) &= (I - K_kH)P(k|k-1)
 \end{aligned} \tag{2-18}$$

滤波增益  $K_k$  是预测与量测的权衡。当量测精确时应该增大 **Kalman** 增益，

反之则减少 Kalman 增益。针对传统 Kalman 滤波过多依赖于历史量测，一段时间后会导导致 Kalman 增益减小而失去滤波的作用，因此改进的算法增添了消渐因子  $\gamma$  ( $\gamma > 1$ )，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：

$$P(k|k-1) = \gamma A(k-1)P(k-1|k-1)A^T(k-1) + Q(k-1) \quad (2-19)$$

### 2.1.3 基于优化匹配搜索的运动目标跟踪算法

Mean Shift 最早是由 Fukunaga 等人于 1975 年提出来的，其最初含义是偏移的均值向量。但随着 Mean Shift 理论的发展，Mean Shift 的含义也逐步地发生着变化，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。

#### 1) 基本 Mean Shift

给定  $d$  维空间  $R^d$  中的  $n$  个样本点  $x_i$ ,  $i=1, \dots, n$ ，在  $x$  点的 Mean Shift 向量的基本形式定义为：

$$M_h(x) \equiv \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad (2-20)$$

其中， $S_h$  是一个半径为  $h$  的高维球形区域，满足以下关系的  $y$  点的集合，

$$S_h(x) \equiv \left\{ y : (y-x)^T (y-x) \leq h^2 \right\} \quad (2-21)$$

$k$  表示在这  $n$  个样本点  $x_i$  中，有  $k$  个点落入区域  $S_h$  中。

我们可以看到  $(x_i - x)$  是样本点  $x_i$  相对于点  $x$  的偏移向量，公式 2-20 定义的 Mean Shift 向量  $M_h(x)$  就是对落入区域  $S_h$  中的  $k$  个样本点相对于点  $x$  的偏移向量求和然后再平均。直观上讲，如果样本点  $x_i$  从一个概率密度函数中采样得到，由于非零的概率密度梯度指向概率密度增加最大的方向，因此从平均上来说， $S_h$  区域内的样本点更多的落在沿着概率密度梯度的方向。因此，对应的 Mean Shift 向量  $M_h(x)$  应该指向概率密度梯度的方向。

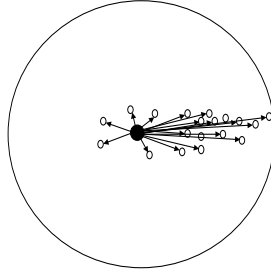


图 2-1 Mean Shift 示意图

如上图所示，大圆圈所圈定的范围就是  $S_h$ ，小圆圈代表落入  $S_h$  区域内的样本点  $x_i \in S_h$ ，黑点就是 Mean Shift 的基准点  $x$ ，箭头表示样本点相对于基准点  $x$  的偏移向量，因此我们可以得到，平均的偏移向量  $M_h(x)$  会指向样本分布最多的区域，也就是概率密度函数的梯度方向。

## 2) 扩展 Mean Shift

从公式 2-20 我们可以看出，只要落入  $S_h$  区域内的采样点，无论其离  $x$  远近，对最终的  $M_h(x)$  计算的贡献是一样的。然而，一般来说离  $x$  越近的采样点对估计  $x$  周围的统计特性越有效，因此我们引进核函数的概念，即在计算  $M_h(x)$  时可以考虑距离的影响，认为在所有的样本点  $x_i$  中，重要性并不一样，因此我们对每个样本都引入一个权重系数。

因此，基本的 Mean Shift 形式扩展如下：

$$M(x) \equiv \frac{\sum_{i=1}^n G_H(x_i - x) w(x_i) (x_i - x)}{\sum_{i=1}^n G_H(x_i - x) w(x_i)} \quad (2-22)$$

其中： $G_H(x_i - x) = |H|^{-1/2} G(H^{-1/2}(x_i - x))$ ； $G(x)$  是一个单位核函数； $H$  是一个正定的对称  $d \times d$  矩阵，我们一般称之为带宽矩阵； $w(x_i) \geq 0$  是一个赋给采样点  $x_i$  的权重。

在实际应用的过程中，带宽矩阵  $H$  一般被限定为一个对角矩阵  $H = \text{diag}[h_1^2, \dots, h_d^2]$ ，甚至可以更简单的被定义为正比于单位矩阵，即  $H = h^2 I$ 。

由于后一形式只需要确定一个系数  $h$ ，因此常常被采用，因此公式 2-22 又可以被改写为：

$$M_h(x) \equiv \frac{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i) (x_i - x)}{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i)} \quad (2-23)$$

可以看到，如果对所有的采样点  $x_i$  满足：

$$(1) \quad w(x_i) = 1;$$

$$(2) \quad G(x) = \begin{cases} 1 & \text{if } \|x\| < 1 \\ 0 & \text{if } \|x\| \geq 1 \end{cases}.$$

则公式 2-23 完全退化为公式 2-20，也就是说扩展的 Mean Shift 在某些情况下会退化为最基本的 Mean Shift。

### 3) Mean Shift 在目标跟踪中的应用

我们采用一个目标的灰度或颜色分布来描述这个目标，假设目标中心位于  $x_0$ ，则该目标可以表示为：

$$\hat{q}_u = C \sum_{i=1}^n k \left( \left\| \frac{x_i^s - x_0}{h} \right\|^2 \right) \delta [b(x_i^s) - u] \quad (2-24)$$

加权的灰度或颜色直方图，使距离目标中心的像素点具有较大的权值，具有抵抗部分遮挡和边缘噪声影响的优点。

候选的位于  $y$  的目标可以描述为：

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k \left( \left\| \frac{x_i^s - y}{h} \right\|^2 \right) \delta [b(x_i^s) - u] \quad (2-25)$$

因此目标跟踪可以简化为寻找最优的  $y$ ，使得  $\hat{p}_u(y)$  与  $\hat{q}_u$  最相似。

$\hat{p}_u(y)$  与  $\hat{q}_u$  的相似性用 Bhattacharya 系数  $\hat{\rho}(y)$  来度量，即

$$\hat{\rho}(y) \equiv \rho[p(y), q] = \sum_{u=1}^m \sqrt{p_u(y) \hat{q}_u} \quad (2-26)$$

公式 2-26 在  $\hat{p}_u(\hat{y}_0)$  点泰勒展开可得，

$$\rho[p(y), q] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p(y_0)q_u} + \frac{1}{2} \sum_{u=1}^m p_u(y) \sqrt{\frac{q_u}{p_u(y_0)}} \quad (2-27)$$

把公式 2-25 带入公式 2-27, 整理可得:

$$\rho[p(y), q] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p(y_0)q_u} + \frac{C_h}{2} \sum_{i=1}^n w_i k \left( \left\| \frac{y - x_i}{h} \right\|^2 \right) \quad (2-28)$$

其中,

$$w_i = \sum_{u=1}^m \delta[b(x_i) - u] \sqrt{\frac{q_u}{p_u(y_0)}} \quad (2-29)$$

对公式 2-29 右边的第二项, 可以利用 Mean Shift 算法对其进行最优化。

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 算法用于视频跟踪有以下几个优势: 首先, 算法计算量不大, 在目标区域已知的情况下完全可以做到实时跟踪; 其次, 作为一个无参数概率密度估计算法, 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。

## 2.2 基于目标重构的跟踪算法简要介绍

目前, 对于目标跟踪的研究越来越多的集中于跟踪的鲁棒性上。经过二十多年的发展, 视频跟踪得到了充分的发展, 提出的算法越来越多, 总结起来, 我们可以把这些算法归结于三个方面: 特征表达、运动预测、目标匹配与重构。基于特征表达的目标跟踪算法的基本思想是通过寻找完备表达目标的方法达到高效的跟踪。由于不同的特征适应于不同的场景, 基于多特征融合的跟踪算法获得了较基于单特征表达目标跟踪算法更鲁棒的跟踪效果。近年来, 基于自适



应性的多特征融合与选择的目标跟踪算法，通过实时的评估与选择特征，在跟踪实验中取得了较好的鲁棒性。基于运动预测的跟踪算法，主要通过使用滤波预测的方法，如 Kalman 或者粒子滤波，对目标运动状态进行预测，缩小目标搜索范围，使得目标跟踪在速度上和准确率上得到了较大的提升。基于目标重构的跟踪算法基本思想是通过样本集获取样本的重构系数，即使用样本选择与分类的方法（分类器、稀疏表示等）从样本集中获取重构系数，从而有效的对目标的状态（特征表达及位置）进行重构，最终完成对目标的跟踪，此方法可取得较好的跟踪效果，因此在最近几年受到越来越多的关注，亦是本文研究的主要内容。

由于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重构得到。本文主要通过对样本集中样本进行选择 and 评估，获取样本的重构系数来重构目标从而进行目标跟踪。由于样本集中的样本包含有目标的部分细节或者整个目标，因此可以看作为一个类别的样本，符合一类支持向量机的一类要求。基于 OC-SVM 的目标跟踪，将基于分类的目标跟踪和一类样本两者结合起来，使用一类支持向量机对样本集中的样本进行训练回归，得到这些样本的评价系数，从而重构出跟踪目标的特征表示，继而进行目标跟踪。

在基于 OC-SVM 的目标跟踪的基础上，由于跟踪过程中在视频帧中搜索区域往往要比跟踪目标本身大很多，这样便会导致样本重构的系数会非常的稀疏，因为搜索区域中背景区域内的样本（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，只有部分反例样本和部分正例样本（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）。基于此，本文又提出了基于稀疏样本重构的目标跟踪算法，即将样本集中的样本对目标进行稀疏表达，获取稀疏系数，得到重构目标最紧致样本子集，从而对目标进行跟踪。它的基本思路可以使用下式表示：

$$A\psi \approx F \Leftrightarrow \arg \min \|\psi\|, \text{ subject to } \|A\psi - F\|_2 \leq \varepsilon \quad (2-30)$$

其中， $F$  表示跟踪目标的特征表示， $\psi = \{\psi^k, k=1, \dots, K\}$  是对应样本集  $A$  中样本所对应的重构（稀疏）系数向量， $\psi^k$  就是第  $K$  个样本的重构系数， $\varepsilon$  为重构误差阈值。通过上式可以得到目标的稀疏重构表达，在解决物体局部遮挡的问题中具有一定的优势，使得跟踪算法更加稳定、鲁棒。

### 第三章 基于ONE-CLASS SVM目标重构的跟踪算法

SVM (Support Vector Machine) 方法是近年来热门的机器学习算法, 因为它较好的数据处理效果及对样本的宽松要求, 广泛的应用于视频目标的检测、跟踪与识别中。其中, One-Class SVM 方法于 2001 年由 Scholkopf et al.[34]在支持向量机 (SVM) 的基础上提出, 但是它与 SVM 有很大的不同, 主要体现在: 其不要求大量的正例反例样本, 且主要应用只包含一类样本信息中。One-Class SVM 的基本思想是在特征空间中估计一块体积最小的部分, 包含尽可能多的相关样本, 同时保持较大的分类间隔。正是因为 One-Class SVM 对样本的特殊要求, 它可以应用到目标跟踪算法中的目标重构中, 通过对样本集中的样本进行评价得出重构系数来重构目标, 从而完成跟踪, 并取得了较好的跟踪结果。

#### 3.1 ONE-CLASS SVM的基本理论

##### 3.1.1 ONE-CLASS SVM的提出

基于统计学习的 VC 维理论和结构化风险最小化原理的支持向量机(SVM)是统计学系理论中较年轻的内容, 它能够在有限的样本信息的前提下, 在模型的复杂性和学习能力之间衡量, 以获得最好的推广能力[34]。

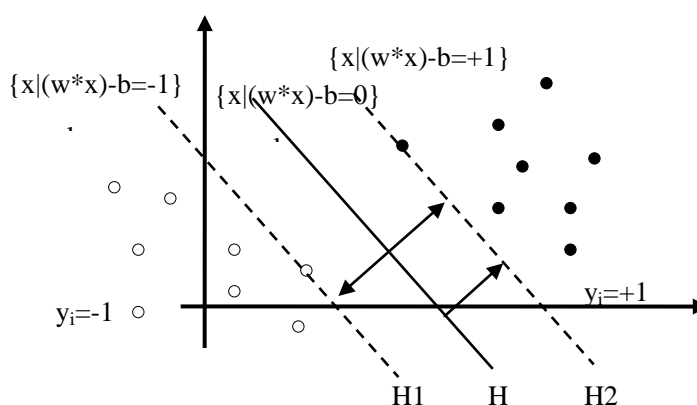


图 3-1 SVM 最优分类超平面

SVM 的基本思想是:通过非线性变换  $\Phi(i)$  将输入空间映射到一个高维特征空间中,将输入空间中线性不可分的数据集映射后变换为线性可分,然后在特征空间中求取最大的间隔超平面  $f(x) = w^T \Phi(x) + b$ , 其中  $w$ ,  $b$  分别是这个超平面的权值和和阈值。

空心和实心点分别代表正例、反例样本,  $H$  为分隔超平面,  $H_1$  和  $H_2$  为两个距离最远且平行于分隔超平面的超平面,  $H_1$  和  $H_2$  之间的距离叫做最大分类间隔。

设训练样本集为  $\{(x_1, c_1), (x_2, c_2), \dots, (x_n, c_n)\}$ ,  $i = 1, 2, \dots, n$ ,  $x \in R^d$ ,  $c \in \{+1, -1\}$ , 其中  $c_i$  为 1 或 -1, 分别表示数据点属于正例、反例。 $x_i$  是表示一个样本的  $n$  维向量, 为了防止变量随机影响分类结果, 每个元素都被缩放到区间  $[0, 1]$  或  $[-1, 1]$ 。设分隔超平面的数学表示为  $w \cdot x - b = 0$ ,  $w$  是垂直于分隔超平面的向量,  $b$  是间隔距离的位移。与最优分隔超平面平行, 且距离支持向量最近的超平面, 可由方程表示为:

$$\begin{cases} w \cdot x - b = +1 \\ w \cdot x - b = -1 \end{cases} \quad (3-1)$$

若训练数据样本线性可分, 则存在两个超平面, 其间没有样本点且距离最大, 通过几何计算可知两个超平面之间距离为  $2/\|w\|$ , 最小化  $\|w\|^2$  才可以得到最大的距离。此外, 为了保证超平面间隔之间没有样本数据点, 要确保所有的样本满足如下条件:

$$w \cdot x - b \geq 1 \text{ 或者 } w \cdot x - b \leq -1 \quad (3-2)$$

$$\text{即 } c_i(w \cdot x_i - b) - 1 \geq 0, 1 \leq i \leq n \quad (3-3)$$

经过转换后, 寻求最优分隔超平面的问题就变为在  $c_i(w \cdot x_i - b) - 1 \geq 0, 1 \leq i \leq n$  的约束条件下最小化  $\|w\|$  的问题, 也就是说转化为一个二次规划最优化的问题:

$$\min \|w\|^2 / 2, \text{ s.t. } c_i(w \cdot x_i - b) - 1 \geq 0, 1 \leq i \leq n \quad (3-4)$$

这样, 我们就可以得到最优分类面或最大间隔分类面, 其中, 在分类面  $H_1$  和  $H_2$  上的训练样本点称为支持向量(Support Vector, SV)。

若样本数据点是线性不可分的, 则可以在公式 3-4 的条件

$c_i(w \cdot x_i - b) - 1 \geq 0, 1 \leq i \leq n$  中增加一个松弛项  $\xi_i, i=1, 2, \dots, n$ ，同时在求解中加入

松弛变量，要求  $\|w\|^2 / 2 + c \sum_{i=1}^n \xi_i$  的最小值，这样公式 3-4 可以转化为：

$$\min \|w\|^2 / 2 + c \sum_{i=1}^n \xi_i, \text{ s.t. } c_i(w \cdot x_i - b) - 1 + \xi_i \geq 0, 1 \leq i \leq n \quad (3-5)$$

折衷考虑最小误差样本和最大分类间隔，则可以得到最优分类面，其中  $c > 0$  是一个常数，对控制样本分类正确度惩罚的程度。

求解上面的二次规划，我们可以得到最优分类超平面函数为：

$$f(x) = \text{sgn}\{w^* \cdot x - b^*\} = \text{sgn}\left\{\sum_{i=1}^n \partial_i y_i (x_i \cdot x) - b^*\right\} \quad (3-6)$$

该公式中仅包含待分类样本与训练样本中的支持向量机的内积运算，因此，特征空间中的最优线性分类问题则转化为求解空间中的内积运算。

对于非线性问题，可以通过核函数  $K(x, y) = \Phi(x) \cdot \Phi(y)$  将非线性变化转化为某个高维空间中的线性问题。由于在原空间和转换的高维空间中只用到了内积运算，因此对应的分类超平面的函数变为：

$$f(x) = \text{sgn}\left\{\sum_{i=1}^n \partial_i y_i K(x_i \cdot x) - b^*\right\} \quad (3-7)$$

总之，SVM 专门针对在有限样本信息情况下，在模型的复杂性和学习能力之间寻求最优解，从而获得最好的学习能力。与 SVM 用于两类的分类问题不同，One-Class SVM 主要是应用于密度估计和聚类。One-Class 概念在 1993 年 Moya 和 1997 年 Ritter 都有涉及过，而被引入到机器学习概念中则是在 1995 年由 Bishop 引入。One-Class SVM 的基本思想是训练样本只有一类，标签为 1 或者 -1，即用一种训练样本训练分类器，再用分类器去测试标签为 1 和 -1 的测试样本。

Scholkopf[34]等学者提出了两种办法来解决一类分类问题：

1) 通过核函数将数据映射到特征空间中，然后将这些数据同原点以最大的间隔分开。实际上，这就是一个 two-class 分类问题，负样本中的唯一元素就是原点，而所有的训练数据都是正样本；

2) 通过核函数将数据映射到特征空间中，寻找到一个超球面，尽可能包含所有的训练数据，并且球体越小越好；

这两种解决方法分别对应于 One-Class SVM 的普通形式和球面形式，两种

形式是完全统一的，但是针对于不同的应用场景和样本数据。在本文中，我们使用 One-Class SVM 球面形式进行视频运动目标的跟踪。

### 3.1.2 ONE-CLASS SVM的基本理论

One-Class SVM 的基本思想就是假设某一类样本数据在特征空间中具有一定的概率分布，通过构建描述样本概率分布的二值模型，判断待测试的样本在特征空间是否服从该模型分布，即该待测数据属于该类或不属于该类。Schoelkopf 等提出的 One-Class SVM 巧妙的利用原点作为负类的代表，通过最大化原点和目标数据间的最小几何间隔  $\rho/\|w\|$ ，来寻找最优超平面  $w \cdot x - \rho = 0$

(其中  $w$  是超平面的法向量， $\rho$  是原点到超平面函数间隔)，使超平面尽量远离原点，从而最小化大部分目标数据所在的正半空间。如图 3-2 所示：

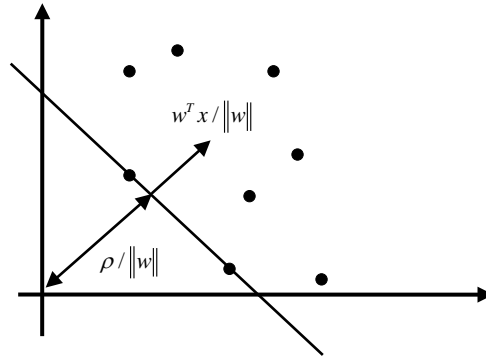


图 3-2 One-Class SVM 的二维示意图

考虑数据集  $\{x_1, x_2, \dots, x_n\}, x_i \in \mathbb{R}^d, i = 1, \dots, n$ ，记  $\Phi$  是一个从输入空间到某个内积空间的特征映射，且有核函数  $K(x, y) = \Phi(x) \cdot \Phi(y)$ ，并引入松弛变量  $\xi_i, i = 1, 2, \dots, n$ 。为了将数据与原点尽量分开，使  $\rho/\|w\|$  最大化，One-Class SVM 通过求解下面的优化问题实现：

$$\min \frac{1}{2} \|\omega\|^2 + \frac{1}{\nu n} \sum_i \xi_i - \rho, \quad s.t. \quad (\omega \cdot \Phi(x_i)) \geq \rho - \xi_i, \xi_i \geq 0 \quad (3-8)$$

其中， $\nu$  是百分比估计，是边界支持向量的上界，是全部支持向量个数的下界。其相应的拉格朗日函数为：

$$L(w, \rho, \xi, \delta, \beta) = \frac{1}{2} \|\omega\|^2 + \frac{1}{\nu n} \sum_i \xi_i - \rho - \sum_i \delta_i (w \cdot \Phi(x_i) - \rho + \xi_i) - \sum_i \beta_i \xi_i \quad (3-9)$$

其中， $\delta, \beta \geq 0$  是拉格朗日乘子。通过对拉格朗日函数  $L(w, \rho, \xi, \delta, \beta)$  关于  $w, \rho, \xi$

求偏导，可以到的：

- 1) 系数  $\partial_i$  满足约束  $\sum_i \partial_i = 1, 0 \leq \partial_i = \frac{1}{vn} - \beta_i \leq \frac{1}{vn}$
- 2) 法向量  $w$  是训练样本集中各向量的线性组合： $w = \sum_i \partial_i \Phi(x_i)$

将上式代入到原始的拉格朗日函数（式）得到其对偶形式：

$$\begin{aligned} L(w, \rho, b) &= \frac{1}{2} \|\omega\|^2 + \frac{1}{vn} \sum_i \xi_i - \rho - \sum_i \partial_i (w \cdot \Phi(x_i) - \rho + \xi_i) - \sum_i \beta_i \xi_i \\ &= \frac{1}{2} \sum_{i,j} \partial_i \partial_j k(x_i, x_j) - \sum_{i,j} \partial_i \partial_j k(x_i, x_j) \\ &= \frac{1}{2} \sum_{i,j} \partial_i \partial_j k(x_i, x_j) \end{aligned}$$

于是得到 One-Class SVM 原始优化问题（公式 3-8）的对偶问题：

$$\min \frac{1}{2} \sum_{i,j} \partial_i \partial_j k(x_i, x_j), \quad s.t. \quad \sum_i \partial_i = 1, 0 \leq \partial_i \leq \frac{1}{vn}, i = 1, 2, \dots, n \quad (3-10)$$

通过经典的二次规划软件包或者序列最小优化算法，求解该二次拉格朗日乘子  $\partial_i$ ， $\partial_i > 0$  对应的样本点  $x_i$  为支持向量 SV，法向量  $w$  可以通过支持向量得到：

$$w = \sum_i \partial_i \Phi(x_i) = \sum_{i \in SV} \partial_i \Phi(x_i) \quad (3-11)$$

决策函数通过下式得到：

$$f(x) = \text{sgn}(w \cdot \Phi(x) - \rho) = \text{sgn}(\sum_i \partial_i k(x_i, x) - \rho) = \text{sgn}(\sum_{i \in SV} \partial_i k(x_i, x) - \rho) \quad (3-12)$$

其中  $\rho$  可以通过某个支持向量  $x_i$  求得：

$$\rho = w \cdot \Phi(x) = \sum_i \partial_i k(x_i, x) = \sum_{i \in SV} \partial_i k(x_i, x) \quad (3-13)$$

前面的理论是将 one-class 问题转化为 two-class 分类问题来解决，然后寻求一个能够将样本分开的由支持向量所支撑的最优超平面。下面我们寻找一个能够包含某类全部（或几乎全部）样本在内的由支持向量所支撑的最紧致超球面。

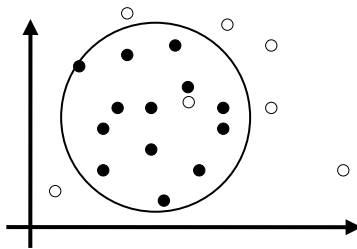


图 3-3 One-Class SVM 的超球面分类

假设某单类样本集  $\{x_1, x_2, \dots, x_n\}$ ，其中， $x_i \in \mathbb{R}^d, i=1, \dots, n$  为样本特征向量， $n$  为样本数。希望找到一个几乎包含所有样本点  $x_i$  的体积最小的球。该球由球心  $c$  和最小半径  $R$  来描述。与 SVM 类似，其最初的带有约束条件的优化问题为：

$$\min R^2 + \frac{1}{vn} \sum_i \xi_i, \quad \text{s.t.} \quad \|\Phi(x_i) - c\|^2 \leq R^2 + \xi_i, \xi_i \geq 0 \quad (3-14)$$

其中， $\xi_i$  是松弛变量， $v \in [0, 1]$  控制着球的体积与被拒绝在该球之外的样本数目之间的平衡关系。 $v$  越小意味着球的体积越大，较少的样本被拒绝在该球之外； $v$  越大意味着球的体积越小，较多的样本被拒绝在该球之外。根据 KKT 条件，引入拉格朗日系数  $\partial, \beta$ ，其拉格朗日函数为：

$L(R, c, \xi, \partial, \beta) = R^2 + \frac{1}{vn} \sum_i \xi_i - \sum_i \beta_i \xi_i - \sum_i \partial_i (R^2 + \xi_i - ((x_i \cdot x_i) - 2(c \cdot x_i) + (c \cdot c)))$ 。其中  $\partial_i, \beta_i \geq 0$  为拉格朗日乘子。由式对  $R, c, \xi$  的偏微分为 0，可导出如下约束条件：

$$\frac{\partial L}{\partial R} = 2R(1 - \sum_i \partial_i) = 0 \Rightarrow \sum_i \partial_i = 1$$

$$\frac{\partial L}{\partial c} = -\sum_i 2\partial_i (x_i - c) = 0 \Rightarrow c = \sum_i \partial_i x_i$$

$$\frac{\partial L}{\partial \xi_i} = \frac{1}{vn} - \partial_i - \beta_i = 0 \Rightarrow 0 \leq \partial_i \leq \frac{1}{vn}$$

由此，式 3-14 可转换成如下相对简单的对偶问题：

$$\min \sum_i \partial_i (x_i \cdot x_i) - \sum_{i,j} \partial_i \partial_j (x_i \cdot x_j), \quad \text{s.t.} \quad \sum_i \partial_i = 1, 0 \leq \partial_i \leq \frac{1}{vn}, i=1, 2, \dots, n \quad (3-15)$$

其中， $\partial_i$  是 Lagrange 乘积因子，只有一部分是非零的，每个非零的 Lagrange 乘积因子  $\partial_i$  对应的点就是通过训练支持向量机得到支持向量 (SV)。决策函数为：

$$f(x) = R^2 - \sum_{i,j \in SV} \partial_i \partial_j (x_i \cdot x_j) + 2 \sum_{i \in SV} \partial_i (x_i \cdot x) - (x \cdot x) \quad (3-16)$$

同样，大多实际问题在原始空间中并不是球形分布的。引入一个映射  $\Phi$  来解决非线性二次规划问题。 $\Phi$  将原始空间的样本数据映射高维特征空间，通过在特征空间进行线性建模来实现原始空间的非线性建模问题。实际的算法是不需知道映射  $\Phi$ ，而是通过满足 Mercer 的核函数来实现。所以式可重写为：

$$\min \sum_i \partial_i K(x_i \cdot x_i) - \sum_{i,j} \partial_i \partial_j K(x_i \cdot x_j), \quad \text{s.t.} \quad \sum_i \partial_i = 1, 0 \leq \partial_i \leq \frac{1}{vn}, i=1, 2, \dots, n \quad (3-17)$$

使用一类支持向量机对目标样本集进行训练，解决上述的二次规划问题。  
高维特征空间中的超球心为：

$$c = \sum_i \partial_i x_i \quad (3-18)$$

高维特征空间中的超球半径为：

$$R^2 = \sum_{i,j \in SV} \partial_i \partial_j K(x_i, x_j) + 2 \sum_{i \in SV} \partial_i K(x_i, x) - K(x, x) \quad (3-19)$$

高维特征空间中的超球面函数为：

$$f(x) = R^2 - \sum_{i,j \in SV} \partial_i \partial_j K(x_i, x_j) + 2 \sum_{i \in SV} \partial_i K(x_i, x) - K(x, x) \quad (3-20)$$

只要一种运算满足 Mercer 条件，它就可以作为这里的核函数使用，核函数  $K(x, x)$  的选择是构造支持向量机的关键所在，常用的核函数有

- 1) 线性核函数:  $K(x, y) = x \cdot y$ ;
- 2) 多项式核函数:  $K(x, y) = (x \cdot y + \theta)^q, q = 1, 2, \dots, \theta, q$  为调整参数,  $q$  为多项式次数;

- 3) 高斯径向基核函数:  $K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$ ;

- 4) Sigmoid 核函数:  $K(x, y) = \tanh[b(x \cdot y) - c]$ 。

与 SVM 相比较，One-Class SVM 使用一类的数据，从密度估计和聚类的角度出发，构建一个远离原点的超平面或者超球面。这个超球面对于数据分布的反应，是使用支撑向量的近似估计，虽然只是一定程度上反应了数据的情况，但是给我们提供了一些重要的信息（如球心）可以用于我们的跟踪领域，这点完全不同于以往的 SVM 在跟踪上应用[35-40]。

### 3.2 基于ONE-CLASS SVM的样本重构目标跟踪算法

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，基于 One-Class SVM 的样本重构目标跟踪算法的基本思路是将目标跟踪视为一类问题，通过使用 One-Class SVM 在样本空间形成的紧致的超球面，对所获取的样本集合中的样本进行选择评价，然后使用样本的重构系数进行目标重构，从而进行目标跟踪。在跟踪的过程中，样本集是通过视频帧的搜索区域动态建立的。基于 One-Class SVM 的样本重构目标跟踪算法需要着重研究两个方面的



问题：1) 样本集合的构造；2) 基于 One-Class SVM 的样本选择评估。

### 3.2.1 基于搜索区域跟踪目标的样本集构造

为了对目标进行跟踪, 首先需要在跟踪视频帧中缩小目标可能出现的范围, 因此我们将意义的样本集合起来形成样本集, 作为目标的候选样本或者通过这些样本构建目标。

根据运动的连续性和视频时间间隔的短暂性, 运动目标在连续的视频帧中处于相对连续的位置上。因此在新视频帧中获取目标的样本集, 是在上一帧的目标物体周围来获得。在当前的视频帧中, 我们在前一帧搜索结果的周围按照一定比例获取一个大小为  $W \times H$  的长方形框, 样本集可以通过搜索框中抽取图像区域获得。为了得到精确的搜索框, 我们可以通过基于常量运动模型的卡尔曼滤波来得到。

样本集是作为候选目标的集合, 样本集中的样本可以在搜索框中得到, 是搜索框的子框, 我们把它表示为  $r = (x, y, s, \alpha)$  ( $0 < x < W$ ,  $0 < y < H$ ,  $s > 0$ ,  $0^\circ \leq \alpha \leq 360^\circ$ ), 其中  $(x, y)$ ,  $s$ ,  $\alpha$  分别表示搜索框的位置, 尺度, 旋转角度。在搜索框中我们可以获得无数个这样的样本。所以我们按照下面的规则选择有意义的样本:

- 1)  $(x, y)$  分别在水平、垂直方向上, 分别以整数步长  $m$ 、 $n$  进行递增;
- 2) 将  $s$  固定到  $[0.8, 1.2]$  之间, 暂时忽略目标尺度的过大过小的变化;
- 3)  $\alpha$  设置为 0, 暂时不考虑目标旋转的情况。

通过这些限定, 我们可以构造一个具有  $K$  个样本的样本集  $\{\vec{S}_i^j, i=1 \dots K\}$ 。

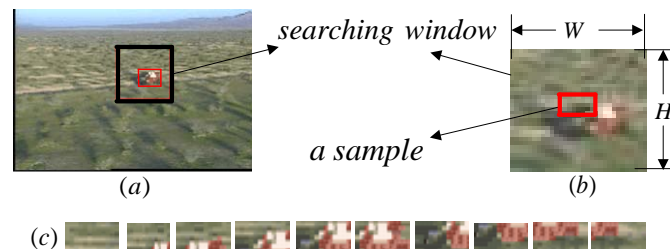


图 3-4 (a) 搜索区域 (b) 一个样本 (c) 样本集合中部分样本示例 (与目标小车相关或者不相关)

为了表示目标和样本, 我们使用包含 120 维的特征向量的综合直方图 HOGC, 包括 48 维的颜色直方图 HC 和 72 维的梯度直方图 HOG。具体定义如

下：

颜色直方图：颜色直方图能够直观的反映物体的表面特征，我们定义 48 维的颜色直方图(HC)。考虑采用 RGB 颜色空间，对于每一个颜色分量，均匀量化到 16 维。

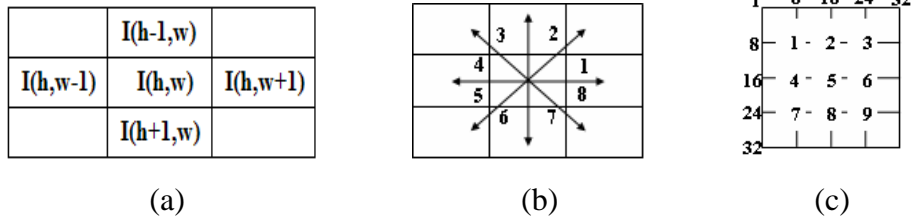


图 三-5 HOG 特征提取，(a)计算梯度方向模板，(b)8 个方向投影，(c)图像窗口中的 9 个块。

梯度方向直方图：当给定特定的图像大小和特征尺度，梯度直方图可以反映物体的局部轮廓，能够很好的区分背景和跟踪目标，而且对于物体的旋转也很敏感，因此我们考虑使用 72 维的 HOG 特征。HOG 特征提取的详细过程为：将样本图像转化为灰度图像；将图像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，并且每个 2x2 的小区域组成一个图像块，因此每个相邻的图像块有 2 个小区域彼此重合；本文中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。因此每个目标得到 72 的梯度方向直方图来描述目标的局部轮廓。每个像素的梯度方向计算如下：

$$\begin{aligned}
 I &= G(\sigma,0) * I_0 \\
 dy &= I(h+1, w) - I(h-1, w) \\
 dx &= I(h, w+1) - I(h, w-1) \\
 ori(h, w) &= atan2(dy, dx) \quad ori \in [0, 2\pi]
 \end{aligned}
 \tag{3-21}$$

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 3-5b 和 3-5c 所示）维 HOG 特征。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。

### 3.2.2 ONE-CLASS SVM对样本集中样本再选择

样本集中的样本数据有的与目标部分相关，有的与目标完全没有相关性。如果将这些完全不相关的样本数据作为训练数据，会大大降低 One-Class SVM

的训练效果，因此在开始使用 One-Class SVM 前，我们需要对样本集数据进行处理，清除掉与样本数据无关或者相关性较小的样本，保证数据的一类性。

度量样本与目标的相关性，需要使用它们特征之间相似性来表示，其中跟踪目标的特征是未知的，一般使用上一帧的目标特征近似为跟踪目标的特征。特征相似度的度量方法有很多，我们使用较为直观简单的巴氏距离（Bhattacharyya distance）计算，它的公式如下：

$$Bhat(F, A_i^j) = \sum_{j=1}^M \sqrt{F(j)A_i^j(j)} \quad (3-22)$$

其中， $F, A_i^j$  分别表示上一帧目标和样本集中第  $i$  个样本的特征向量表示， $M$  是向量的维数（这里应该是 120）。下图为某一时刻的目标与样本集中样本的相似度曲线图 3-6：

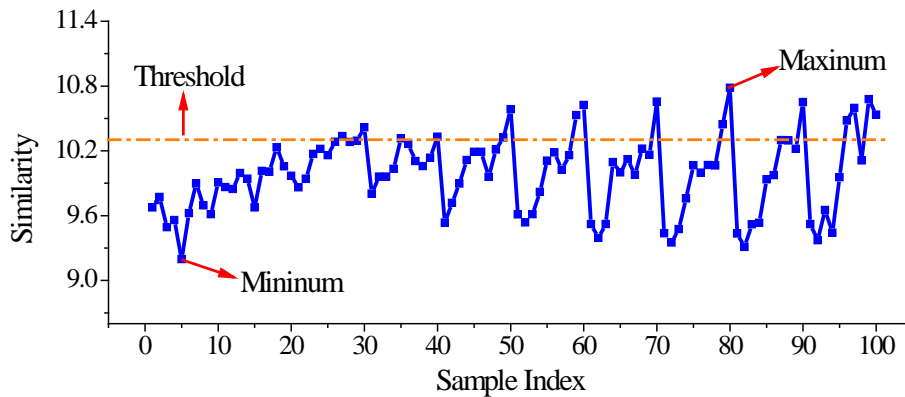


图 3-6 跟踪目标与样本集中的 100 个样本的相似度折线图。同时图中标明相似度最大值、最小值及阈值。样本集中大于阈值的样本可以用来作为 One-Class SVM 的输入。

从图中我们看出，目标与样本的相似度主要分散在一定的区间内。相似度越低，表明样本与目标不相关或者相关性很小；相似度越高，表明样本包含了一部分目标或者全部目标。样本作为 One-Class SVM 的训练基础，我们需要选择与目标相关性强的样本。我们使用自适应阈值的方法，选择所有样本与目标样本相关性最大和最小之间的一定比率值作为阈值，挑选出 One-Class SVM 输入的跟踪样本  $\{E_i^j, i=1 \dots L\}$ ，自适应的阈值公式定义为：

$$Threshold = \varphi * \max + (1 - \varphi) * \min \quad (3-23)$$

其中， $\varphi \in [0,1]$  表示相关性最大值与最小值的比率。经过大量的视频数据的实

验表明,  $\varphi = 0.2$  or 20% 二八分规则是最好的选择。这样, 我们就可以很方便的选择出与目标相关性较强的样本, 作为 One-Class SVM 的训练样本输入:

$$\begin{cases} Bhat \geq Threshold; & sample \in \{examples\} \\ Bhat < Threshold; & sample \notin \{examples\} \end{cases} \quad (3-24)$$

### 3.2.3 基于ONE-CLASS SVM目标重构的跟踪算法

选择完成 One-Class SVM 的样本输入, 我们就可以使用 One-Class SVM 对样本进行回归, 获得跟踪目标的特征向量。

在本章的前几节中提到, 不同于 SVM 的二分类的分类问题, One-Class SVM 是针对仅仅包含一种类型的数据, 从密度估计的思想出发, 构建一个将数据尽可能的包围的尽可能小的超球面。这个超球面可以看作是这些样本数据分布的反映, 它使用超球面的支撑向量的近似估计样本分布, 虽然这个数据密度估计只能较小程度上反映数据的情况, 但是它也给我们提供了一些重要的信息, 这个超球面对于每一个样本的评价系数可以作为我们重构目标的重构系数, 从而进行跟踪。

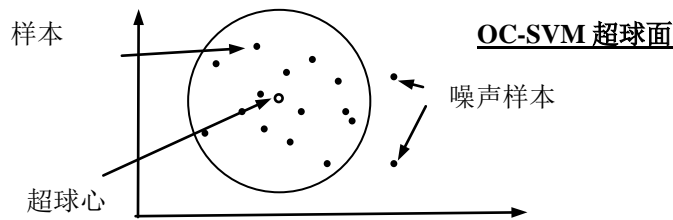


图 3-7 OC-SVM 超球面, 小黑点表示相关样本, 小圆圈表示超球面的球心

经过 One-Class SVM 样本选择之后样本子集  $\{E_i^i, i=1...L\}$ , 及对应的特征表示  $A_i = \{A_i^i, i=1...L\}$ 。根据 One-Class SVM 的定义, 超球面  $(R, c)$  可以通过以下的二次规划来解决:

$$\min R^2 + \frac{1}{vL} \sum_i \varepsilon_i, \quad s.t. \quad \|A_i^i - c\|^2 \leq R^2 + \varepsilon_i, \varepsilon_i \geq 0, i=1, 2, \dots, L \quad (3-26)$$

其中,  $R$  是超球面的半径,  $c$  是超球面的中心,  $v$  是超球面内部的样本占所有样本的比例, 也就是正例样本相对于所有相关样本的占有率, 当  $v$  取值不同时, 产生不同的超球面  $(R, c)$ 。

求解公式 3-25 中的二次规划时, 需要转换成它的对偶形式, 公式如下:

$$\min \sum_i \partial_i K(A_t^i \cdot A_t^i) - \sum_{i,j} \partial_i \partial_j K(A_t^i \cdot A_t^j), \text{ s.t. } \sum_i \partial_i = 1, 0 \leq \partial_i \leq \frac{1}{vn}, i = 1, 2, \dots, L \quad (3-27)$$

其中,  $\partial_i$  是对应于  $A_t^i$  的 Lagrange 乘积因子。通过上面的公式计算得到的  $\partial_i$  中只有一部分是非零的, 我们将每个  $\partial_i$  作为对应的样本  $E_t^i$  的重构向量来重构目标, 公式如下:

$$F \approx \sum_i \partial_i E_t^i \quad (3-28)$$

同时我们通过样本位置来重构目标的位置, 公式如下:

$$pos(F) \approx \sum_i \partial_i pos(E_t^i) \quad (3-29)$$

### 3.3 基于 ONE-CLASS SVM 目标重构的跟踪实验

本节总结了本章所提出的基于 One-Class SVM 目标重构的目标跟踪算法, 并且在数据集上进行实验, 分析实验结果, 将我们的算法与经典的跟踪算法进行实验对比。实验数据表明本章提出的目标跟踪算法在应对部分遮挡, 物体旋转等挑战问题上具有较好的效果。

#### 3.3.1 实验步骤

基于 One-Class SVM 目标重构的目标跟踪算法的基本步骤, 如下表所示:

表 3-1 基于 One-Class SVM 目标重构的跟踪算法

1. **初始化 ( $t=0$ ).** 首先, 初始化目标的位置和尺度  $r = (x, y, s)$ ;
2. **构造样本集 ( $t>0$ ).** 在上一帧的目标周围, 按照一定比例建立样本搜索区域, 进行如下操作构建样本集:
  - (a) 分别在水平、垂直方向上, 以步长为  $m, n$  进行递增;
  - (b) 将尺度固定到  $[0.8, 1.2]$  之间;
3. **获取目标重构稀疏 ( $t>0$ ).** 对样本集的数据进行选择, 具体操作如下:
  - (a) 在样本集中选择适合 One-Class SVM 的相关样本  $\{\overline{E}_t^i, i = 1 \dots L\}$ ;
  - (b) 将  $\{\overline{E}_t^i, i = 1 \dots L\}$  在特征空间上基于 One-Class SVM 进行训练, 获得目标重构系数;
4. **目标定位 ( $t>0$ ).** 根据目标重构得到目标在视频帧的最佳位置。
5.  **$t = t + 1$ ;** 跳转到步骤 2 或者结束循环。

### 3.3.2 视频测试集合及跟踪算法评估

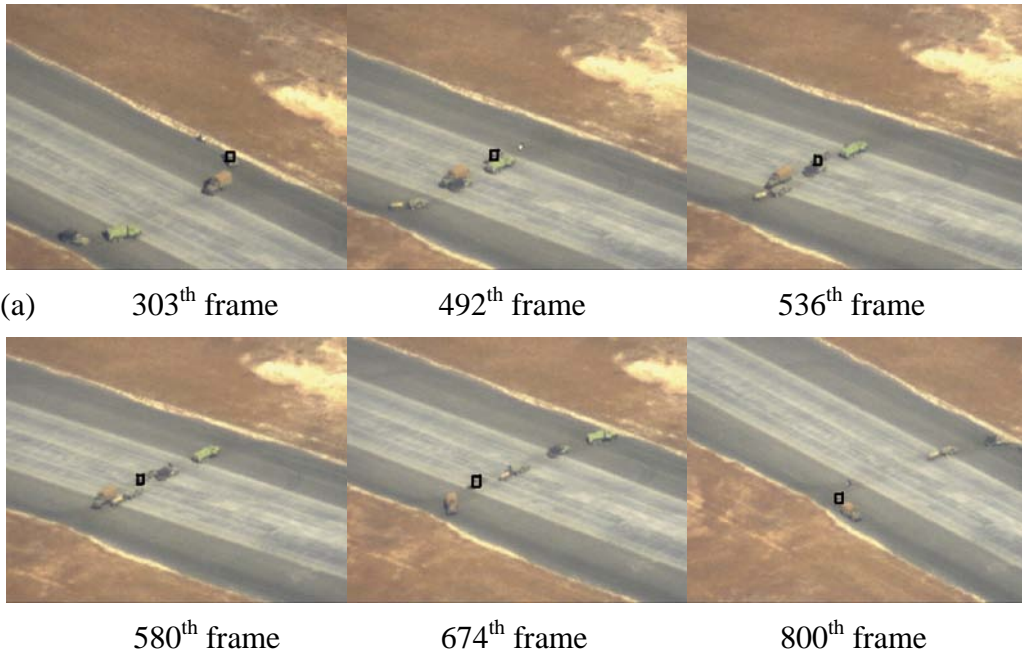
本章所使用的测试视频集合，主要来自于目前比较流行的公共测试集和实验室我们自己的跟踪测试集，主要包括：VIVID data set [69]、CAVIAR data set [70] 和 SDL data set [71]。这些测试视频针对运动背景和前景的变化，其中背景和目标的相对变化主要包括了跟踪目标之间的相互遮挡、目标的尺度变化、目标的旋转、背景光照变化以及相似背景等。在测试视频中不仅有对人体的跟踪，也有对运动车辆的跟踪。

为了更加形象并且量化地评价本章所描述的跟踪算法，我们对跟踪的视频进行了人工标注，然后分别对跟踪算法的跟踪结果与手工标注的结果进行比较。跟踪的效果采用跟踪错误率（displacement error rates, DER)来衡量，其定义如下：

$$DER = \frac{\text{Displacement error between tracked object position and groundtruth}}{\text{Size of the object}}. \quad (3-30)$$

从 DER 的定义可知：跟踪算法的性能越好，其 DER 便会越小；跟踪算法越稳定，其 DER 的在图像上表现的越平稳，这样便可以更加直观地比较跟踪算法的性能。

### 3.3.3 跟踪实验结果展示及实验分析





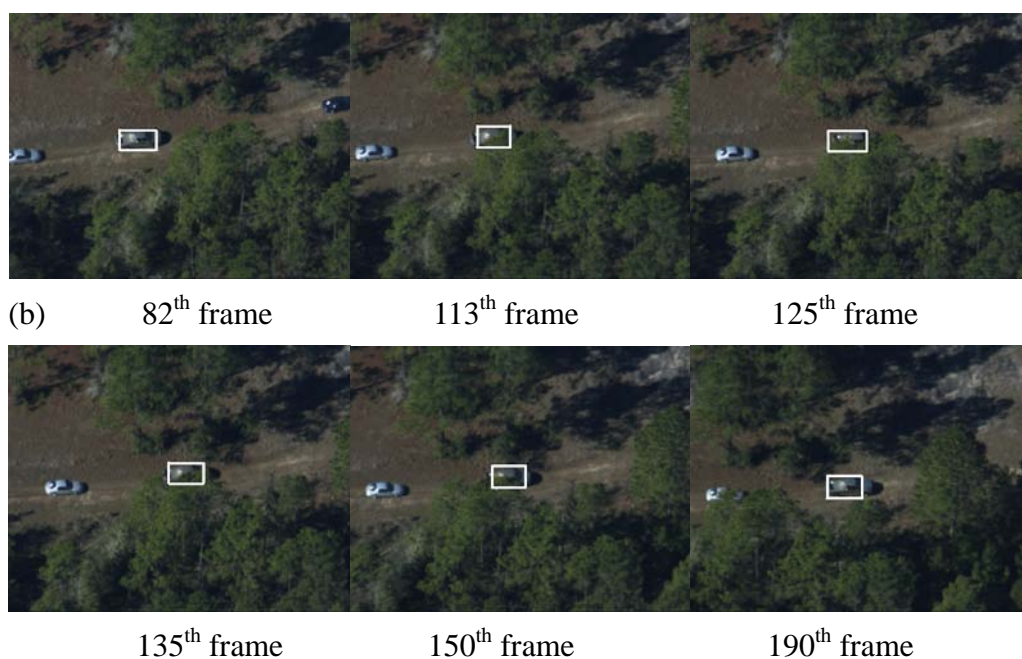


图 3-8 使用基于 One-Class SVM 的目标跟踪的两个试验实例

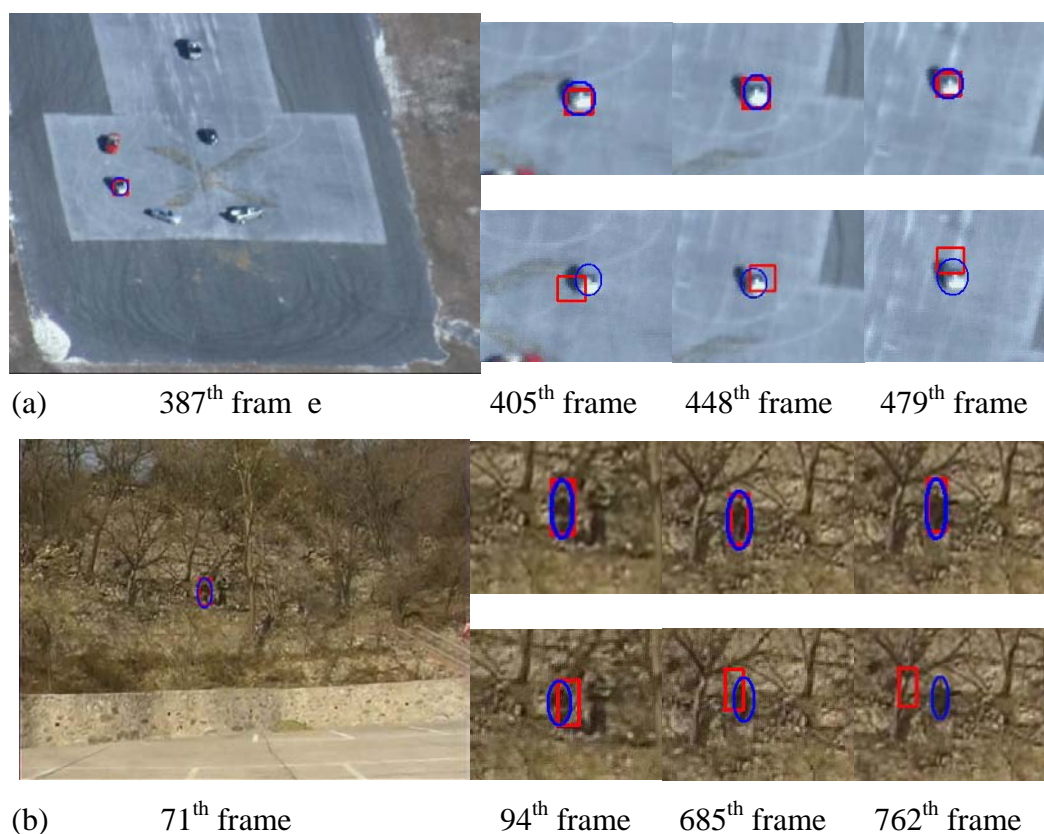


图 3-9 我们的算法与经典 Mean-Shift 算法的两个试验跟踪效果比较

下面使用基于 One-Class SVM 的方法进行跟踪实验效果展示。首先，我们通过两组实验展示我们方法的有效性，尤其在物体旋转、物体发生遮挡时，依

然保持着良好的跟踪效果。接着，我们用图示的方法，将基于 One-Class SVM 的跟踪方法与经典的 Mean-shift 跟踪方法进行比较。最后，将我们的方法与 Collins[73]、SIFT-Based[72]的跟踪方法进行比较，使 DER 的跟踪评价标准，得到实验对比曲线，可以明显的看出我们的方法的优越性。

第一段视频数据来自 VIVID 数据集，这段视频非常有挑战性，摄像机从高空拍摄，两组相似的车辆相向行驶，直到相遇后错开。由于两组车辆颜色相似，并且车辆在相遇时会出现相互遮挡，一般情况下跟踪误差比较大。图 3-8a 中展示了我们方法的跟踪结果，表明我们的方法具在物体旋转、微小目标、和相似的背景下，有较强的鲁棒性。第二段视频数据来自于 CAVIAR 数据集，在这段视频中车辆发生了很大的遮挡（113<sup>th</sup>, 125<sup>th</sup>, 135<sup>th</sup>, 150<sup>th</sup>），图 3-8b 中显示了我们方法对于目标的遮挡具有稳定的跟踪效果。

表 3-2 用于计算 DER 的视频文件列表

视频测试集 (Video test set)	视频名称 (Video name)
VIVID Tracking video set	Redteam.avi
	egtest01.avi
	egtest02.avi
	egtest05.avi
SDL Tracking video set	xiangshan_0032.avi
	xiangshan_0043.avi
CARVIA Tracking video set	Browse2.avi
	Browse_WhileWaiting1_output.avi
	OneStopMoveEnter1cor.avi
	EnterExitCrossingPaths2front.avi

下面我们用图示的方法，将基于 One-Class SVM 的跟踪方法与经典的 Mean-shift 跟踪方法进行比较。如图 3-9 所示，左边的大图片是跟踪场景，右侧较小的上面一行是使用我们的方法的跟踪结果，下面一行是使用经典 mean-shift 方法的跟踪结果。红色长方形是跟踪实验的结果，蓝色椭圆是手动标示的跟踪对象的实际位置。图 3-9a 中在视频数据来自于 VIVID 数据集，摄像机从高空拍摄车辆进行 360 度转弯，在这段视频中车辆的表面发生了很大的变化（479<sup>th</sup>、405<sup>th</sup>、448<sup>th</sup>、479<sup>th</sup>），图 3-9a 中显示了，相对于 mean-shift 算法，我



们方法对于车辆变化具有稳定的跟踪效果。图 3-9b 中的视频数据来自于 SDL 数据集，这个具有很高的挑战性，视频中存在着严重的物体遮挡 (405<sup>th</sup>, 448<sup>th</sup>, 479<sup>th</sup>)，并且目标与背景具有相似颜色的对象(一个很小的人，有一些小的树木很相似)。我们的方法可以正确的跟踪物体对象。这段视频的跟踪结果表明，该方法能有效地处理部分遮挡,外观变化和相似的颜色和外形上具有很好的效果。

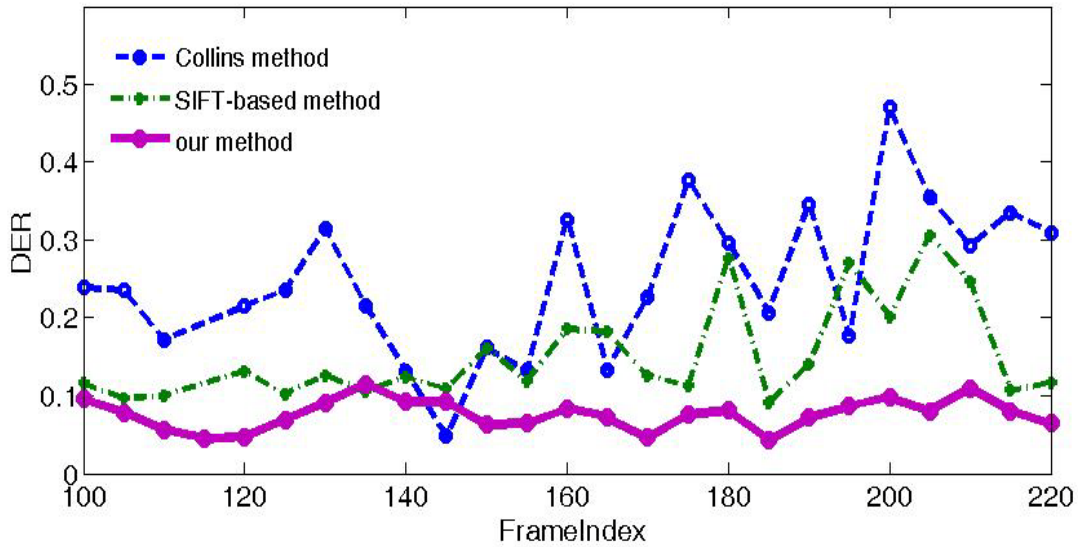


图 3-10 三种跟踪算法的统计实验结果图

最后,我们使用来自三个数据集的 10 段视频数据(视频数据如表 3-2 所示),将本章所提到的基于 One-Class SVM 的跟踪算法与目前比较流行的其他两种跟踪算法进行比较,其中包括:基于特征选择的跟踪算法(Collins method)[72]、和基于 SIFT 特征点的跟踪算法(SIFT-Based method) [73]。并且使用前面提到的跟踪错误率 DER 评价指标,三个方法的结果曲线图如图 3-11 所示。从这个图中可以看出,相对于其他两种方法,我们的方法获得了较好的效果。



## 第四章 基于稀疏表示目标重构的跟踪算法

### 4.1 稀疏表示的基本理论

稀疏表示 (Sparse Representation) 起源于压缩感知理论 (Compressed Sensing), 由斯坦福大学的 David Donoho, 加州理工学院的 Emmanuel Candes 以及加州大学洛杉矶分校的陶喆轩等人提出的压缩感知理论[41-45], 是对奈奎斯特采样定理的突破, 理论中信号可以以低于奈奎斯特采样频率的频率进行采样, 并且仍然能够精确重建原始信号。近年来, 随着压缩感知理论的发展, 吸引了信号处理、图像处理、模式识别等领域工作者的关注, 在很多相关的问题上得到了广泛的研究和应用[46-50]。

#### 4.1.1 稀疏表示的概念

广义上, 稀疏是指集合的大部分元素为 0, 其中集合可以是向量、也可以是矩阵等等。信号的稀疏表示, 一般是指信号在一个过完备的基上的表示使稀疏的。信号的线性组合表示为:

$$x = D\alpha \quad (4-1)$$

其中,  $x$  表示信号向量,  $D$  表示过完备的基, 那么  $\alpha$  就是  $x$  在这组基  $D$  上的线性展开的系数向量。 $x$  在基  $D$  的稀疏性, 就是指系数向量  $D$  的非零项很少。

信号的稀疏表示有一定的抵抗噪声的能力。由于稀疏表示对信号的表达简洁而准确, 它已经广泛的应用于图像处理与计算机视频领域, 如图像恢复、图像压缩、图像去噪及模式分类等, 并取得了令人瞩目的成果。

#### 4.1.2 信号的稀疏表示

稀疏表示就是利用过完备原子库, 将信号表示成这些少数原子的线性组合的过程。稀疏表示使信号能量只集中于较少的原子上, 非零系数的原子就揭示了信号的主要特征和内在结构。

给定一个集合  $D = \{g_k, k = 1, 2, \dots, K\}$ , 其中元素  $g_k \in R^N$  是单位矢量, 原子库中原子个数为  $K$ ,  $K \gg N$ , 称集合  $D$  为原子库, 它是过完备的或冗余的, 其

元素称为原子。对于任意给定长度为  $N$  的信号  $x$ ，在  $D$  中找到一个由  $m$  个原子组成的线性组合：

$$x \doteq x_m = \sum_{k=0}^{m-1} \partial_k g_k \quad (4-2)$$

其中  $\partial_k$  为展开系数，由于  $m$  远小于空间的维数  $N$ ，因此这种表示信号的方法被称作信号的稀疏表示。由于原子库的冗余性 ( $K \gg N$ )，原子矢量  $g_k$  不再是线性无关的。因此上式有多个不同的线性组合，不同的线性组合中，原子和原子个数也不尽相同。

稀疏表示算法的目标是将信号分解成少数原子的线性组合，它模型如下图所示：

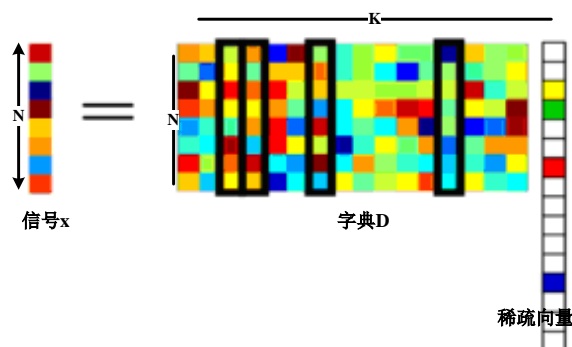


图 4-1 稀疏表示模型

其中，字典的每一列为一个原子， $\alpha$  是信号分解产生的稀疏向量，即为信号  $x$  的表示系数。稀疏分解的问题就是在满足逼近误差条件的前提下，如何从各种可能的组合中，选出  $\alpha$  最为稀疏的一个，或者说  $m$  取值为最小的一个。即

$$\arg \min_{\alpha} \|\alpha\|_0, \quad \text{s.t. } x = D\alpha \quad (4-3)$$

其中  $\|\cdot\|_0$  为  $l^0$  范数，表示系数向量非零项的个数。

信号的稀疏表示问题最终转化为式和式的最优化问题。当  $D$  是  $H$  空间的一个正交基时，要得到最稀疏表示非常容易，然而对于一个冗余字典来说，求解稀疏表示最优化问题将是一个 NP-hard 问题。为了得到稀疏表示最优化问题的解，我们采用了一些逼近的求解方法。

2004 年，在 Donoho 和 Huo[54]的研究基础上，Elad 和 Bruckstein[55]证明了下述引人注目的定理。

定理 1  $D$  是一个相干系数为  $\mu$  的原子库, 如果一个信号  $x \in R^n$  可以表示为:

$$x = \sum_{k=0}^{m-1} \partial_k g_k \quad (4-4)$$

并且  $\|\partial\| < 1/\mu$ , 则式 4-4 就是  $x$  在  $D$  中的唯一

这一结果说明, 如果  $x$  在  $D$  中有非常稀疏的扩展, 那么通过 BP 算法, 这个稀疏扩展就可以精确重构。

定理 2 如果信号  $x$  在原子库  $D$  中有一个稀疏表示, 并且满足

$$\|\partial\| < (\sqrt{2} - 0.5)/\mu \quad (4-5)$$

则最小化问题  $l^1$  有一个唯一解, 它也是最小化  $l^0$  的解。

这个结果意味着, 我么可以通过 BP 算法求解较为简单的  $l^1$  问题, 取代原始的  $l^0$  问题。

在足够稀疏的情况下, 用  $l^1$  范数作为惩罚项来代替  $l^0$  范数作为惩罚项, 两者得到的解是几乎等价的。因此, 原先优化问题中的  $l^0$  范数的惩罚项, 可以用  $l^1$  范数的惩罚项来代替, 而且系数向量的稀疏性仍然有理论保证。这样, 就回避了最小化  $l^0$  范数这个 NP 完全问题, 而只需要解一个最小化  $l^1$  范数的问题:

$$\arg \min_{\alpha} \|\alpha\|_1, \quad \text{s.t. } x = D\alpha \quad (4-6)$$

最小化  $l^1$  被称为基追踪(BP)[53][59], 从而将一个 NP 完全问题转换成一个非 NP 完全问题是较为简单的问法解决。

### 4.1.3 稀疏表示的求解方法

式 4-2 中  $\arg \min_{\alpha} \|\alpha\|_0$  是针对  $l^0$  范数最小的重构算法, 在实际应用情况下, 因为有噪声存在难以完全重构物体, 因此我们将式 4-2 加入噪声项, 变形为下式:

$$\arg \min_{\alpha} \|\alpha\|_0, \quad \text{s.t. } \|x - D\alpha\|^2 < \varepsilon \quad (4-7)$$

很多稀疏表示的求解方法都是基于式 4-7 提出的, 这些算法都为贪婪算法, 主要有匹配追踪方法(Matching Pursuit, MP)[76], 正交匹配追踪方法(Orthogonal Matching Pursuit, OMP)[56], 正则化正交匹配追踪方法(Regularized Orthogonal

Matching Pursuit, ROMP)[57], 最优正交匹配追踪方法(Optimized Orthogonal Matching Pursuit, OOMP)[58]等。

式 4-2 的  $l^0$  范式是一个 NP 难的问题, 根据 Donoho 教授[52]证明结果, 可以使用  $l^1$  范数最小化的重构算法, 将  $l^0$  范式的 NP 难问题转化为求解线性规划最优化问题。这种方法主要有基追踪方法(Basis Pursuit, BP)[53], 梯度投影稀疏重构方法(Gradient Projection for sparse Reconstruction, OPSR)[60]等。下面我们对其中几种比较典型的方法进行简单介绍。

■ 匹配追踪算法(MP)

匹配追踪算法(Matching Pursuit, MP)是比较早提出的稀疏表示的求解方法, 1993 年 Mallat 和 Zhang 在小波分析的基础上, 利用平方可积空间中基本函数的平移、调制和尺度变换构造具有时频特性的过完备时频词典, 提出了用于信号时频表示的 MP 方法。MP 算法的基本思想是迭代的进行贪婪算法, 首先从词典中选择最贴近待分解信号的原子用于信号表示, 同时从待分解信号中去掉其在该原子上的投影, 获得残差信号作为新的待分解信号, 不断地迭代重复这个过程, 一直到残差信号的能量小于某个给定的阈值。

首先, 从过完备词典中选择出与待分解信号最为匹配的原子  $g_{r_0}$  :

$$\langle x, g_{r_0} \rangle = \sup \langle x, g_r \rangle \quad (4-8)$$

这时信号可以分解为最佳原子  $g_{r_0}$  上的分量和剩余残差信号两部分, 即:

$$x = \langle x, g_{r_0} \rangle g_{r_0} + R^1 x \quad (4-9)$$

其中,  $R^1 x$  是用最匹配原子在原信号进行匹配后的剩余残差信号, 这个残差信号作为新一轮的待分解信号, 不断进行这个迭代分解:

$$R^i x = \langle R^i x, g_{r_i} \rangle g_{r_i} + R^{i+1} x \quad (4-10)$$

其中  $g_{r_i}$  是词典中选择出来的最贴近残差信号的原子, 选择方法如下:

$$\langle R^i x, g_{r_i} \rangle \sup \langle R^i x, g_i \rangle \quad (4-11)$$

不断的迭代上面的过程, 直到残差信号的能量小于某个给定的阈值  $T$  :

$$|R^i x| < T \quad (4-12)$$

这时假设进行了  $n$  次迭代, 信号  $x$  可以由分解后的信号来表示为:

$$x = \sum_{i=0}^{n-1} \langle R^i x, g_{r_i} \rangle g_{r_i} + R^n x \quad (4-13)$$

我们需要用少数的原子来近似估计信号，如下式所示：

$$x \approx \sum_{i=0}^{n-1} \langle R^i x, g_{r_i} \rangle g_{r_i} \quad (4-14)$$

上面是使用匹配追踪的方法进行信号的稀疏表示算法。MP 算法原理比较简单，容易理解，算法的计算复杂度也是所有稀疏表达求解算法中最低的，是目前信号稀疏表示最常用的求解方法。但有个很明显的缺点，MP 在已选定的原子组成的子空间上，信号的扩展不是最好的。

#### ■ 正交匹配追踪法(OMP)

正交匹配追踪算法(Orthogonal Matching pursuit, OMP)，是在 MP 算法基础上的一种改进算法。它与 MP 都是从过完备字典中找出与待分解信号最为匹配的原子，不同之处在于，它将所选定的原子利用 Gram Schmidt 正交化方法进行正交化处理，然后将信号在这些正交原子构成的空间上投影，从而可以得到信号在已选定原子上的分量和剩余残差分量，最后使用 MP 相同的方法分解残差分量。

给定字典  $D = \{g_r, r = 1, 2, \dots, N\}$ ，且  $\|g_r\|^2 = 1$ 。利用正交匹配追踪对信号  $x$  进行分解的过程如下：

$$x = \langle R^0 x, g_{r_0} \rangle g_{r_0} + R^1 x \quad (4-15)$$

其中  $g_{r_0}$  为字典中最接近信号的原子，即使残差能量最小的原子，这点与 MP 算法相似，公式表示为：

$$\langle x, g_{r_0} \rangle = \sup \langle x, g_r \rangle \quad (4-16)$$

经过  $i$  次的迭代循环，与式 4-16 类似， $g_{r_i}$  为字典中选择出来的最贴近残差信号的原子，选择方法如下：

$$\langle R^i x, g_{r_i} \rangle \text{MP} \quad \langle R^i x, g_i \rangle \quad (4-17)$$

每次原子选定之后，采用 Gram-Schmidt 算法对  $g_{r_i}$  进行正交化处理。即，

$$u_i = g_{r_i} - \sum_{p=0}^{i-1} \langle g_{r_i}, u_p \rangle u_p / \|u_p\|^2 \quad (4-18)$$

此时，残差  $R^i x$  投影在  $u_i$  上，而不是  $g_{k_i}$ ：

$$R^i x = \left\langle R^i x, u_i \right\rangle u_i / \|u_i\|^2 + R^{i+1} x \quad (4-19)$$

信号  $x$  进行了  $n$  次迭代， $x$  可以由分解后的信号来表示为：

$$x = \sum_{i=1}^{n-1} \left\langle R^i x, u_i \right\rangle u_i / \|u_i\|^2 + R^n x \quad (4-20)$$

上面的过程就是 OMP 算法的基本步骤，其中信号残差和前面每一步中从原子库中选择的原子进行正交，在有限维空间中，OMP 能在有限迭代后收敛。OMP 算法可以作为逼近方法找到全局最优解，而且收敛速度比 MP 算法更快。

#### ■ 基追踪(BP)

1999 年 Donoho 提出了基追踪(Basis Pursuit, BP)方法，它是信号稀疏表示领域的一种新方法。它寻求从过完备字典集中得到信号的最稀疏的表示，即用尽可能少的基精确地表示原信号，从而获得信号的内在本质特性。基追踪方法采用表示系统的范数作为信号稀疏性的度量，通过最小化  $l^1$  将信号稀疏表示问题定义为一类有约束的极值问题，进而转化成线性规划问题进行求解。BP 算法的优点是重建精度高，需要的重建原子个数少，但是计算复杂度相对较高。

#### 4.1.4 稀疏表示在图像/视频处理领域的应用

稀疏表示在理论上的不断突破，尤其是 Donoho 教授提出的用  $l^1$  范数代替  $l^0$  范数的理论保障[54]，使得稀疏表示技术被迅速应用到图像处理等各个领域。这里我们简要介绍几个稀疏表示在图像处理领域的应用。

在 2008 年 Julien Mairal 等[61]提出了基于稀疏表示的图像重建，他们使用稀疏表示对图像进行去噪，同时对较小的区域进行修复，取得良好的结果。2008 年 John Wright 等[62]提出了基于稀疏表示的人脸识别方法，他们将训练集中的人脸作为信号的原子，对测试样本在这组原子上进行展开。在展开的表达式中，我们加入最小化  $l^1$  的限制，从而得到一个稀疏的系数向量来表示这个人脸，然后基于这个稀疏的系数向量对人脸进行识别。该算法在人脸有部分遮挡的时候有很大优势，可以自动找出被遮挡的部分，并且对该部分进行重建。在 2008 年 Jianchao Yang 等[63]提出了基于稀疏表示的超分辨率重建。他们在图像信号具有稀疏性的基础上，在高分辨率图像和低分辨率图像中建立两组对应的基。根据高低分辨率之间的对应性，从而他们实现了超分辨率重建。而且，给定两组对应的基之后，算法不再需要当前待处理图片以外的信息，即可以用一张图片就完成超分辨率重建。在图像处理的其他很多方面，如图像融合、图像目标



检测、图像恢复等领域也不断发展，提出了令人瞩目的基于稀疏表示的算法 [64-65]。

## 4.2 基于稀疏表示目标重构的跟踪算法

稀疏表示在信号领域的巨大发展和它对噪声的良好特性，使得稀疏表示广泛的应用于图像处理各个领域，本章我们将稀疏表示应用于视频目标跟踪中，同样取得了良好的效果。

### 4.2.1 目标跟踪的样本集构造

基于稀疏表示的目标跟踪中的样本集构造，同基于 One-Class SVM 的跟踪算法是一致的。目标样本集是基于样本窗口构建的，集合中的每个样本是样本窗口中的小窗口。假设样本框口为  $W \times H$ ，我们分别在水平、垂直方向按照一定的间隔获取样本，最后可以得到目标样本集  $\{s^k, k = 1 \dots K\}$ 。样本集应该在跟踪每一帧目标进行跟踪，但是为了加快跟踪速度，我们采取每  $M$  ( $M=5$ ) 帧在样本集中随机选择一个样本作为目标来更新样本集。样本集的更新保证了样本表面发生变化或者目标发生遮挡时的跟踪效果，从而预防模板漂移，使跟踪更好的达到自适应。我们使用 HOGC 的特征来表示跟踪目标和样本，于是我们可以得到特征集  $A = \{a^k, k = 1 \dots K\}$ 。

在跟踪过程中，视频帧中的搜索区域往往要比跟踪物体大很多，这样便会导致样本线性重构的系数会非常的稀疏，因为背景区域内的样本（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，只有部分反例样本和部分正例样本（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）。

### 4.2.2 基于稀疏表示的目标重构

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，如下面的公式：

$$A\psi \approx F \quad (4-21)$$

其中， $F$  是跟踪目标的特征表示， $\psi = \{\psi^k, k = 1, \dots, K\}$  是对应  $A$  中样本的系数向

量， $\psi^k$  就是第  $K$  个样本的特征系数。

在实际实验中， $\psi = \{\psi^k, k=1, \dots, K\}$  中只有  $r$  个是非零的系数 ( $r \ll K$ )。在数学中，我们将这样的矩阵称为样本的  $r$  稀疏表示。非零系数的个数可以用  $\|\psi\|_0$  来表示，所以可以通过最小化  $\|\psi\|_0$  获得目标的稀疏表示。这个  $l^0$  的问题是 NP-hard 的问题。根据压缩感知的最新理论， $l^1$  范式最小化 可以用来解决稀疏表示问题：

$$\arg \min \|\psi\|_1, \text{ subject to } A\psi = F \quad (4-22)$$

其中， $\|\cdot\|_1$  表示  $l^1$  范式。在实际应用中，图像中一般包含有噪声，使用样本表示目标不一定能够非常精确，因此我们修改上式，引入  $\varepsilon > 0$  来表示目标：

$$\arg \min \|\psi\|_1, \text{ subject to } \|A\psi - F\|_2 \leq \varepsilon \quad (4-23)$$

其中， $\|\cdot\|_2$  为  $l^2$  范式。这个模型可以通过二次规划方法在多项式的时间内解决。

通过求解上式，我们可以得到  $r$  稀疏的系数向量。图 4-2 是一个稀疏向量的实例，我们选取了 100 个样本的样本集重构跟踪目标。样本集样本经过  $l^1$ -norm 最小化，从图中可以看出，有 10 个左右的系数为非零，通过这 10 个样本重构的目标，重构错误率较低。

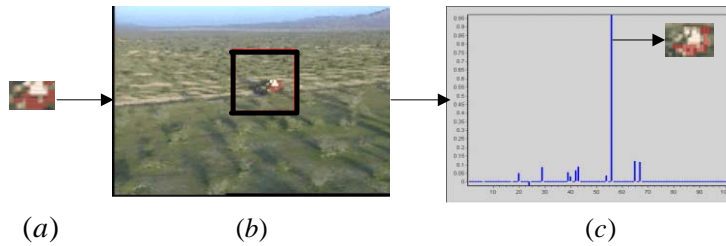


图 4-2 (a) 跟踪目标 (b) 搜索区域 (c)稀疏重构系数

通过样本集  $A$  和计算得到的稀疏系数向量  $\psi$ , 目标就可以通过重构来得到，重构公式为：

$$F \approx A\psi = \sum_{k=1}^K A^k \psi^k \quad (4-24)$$

基于  $l^1$  范式最小化的稀疏表示，保证了目标重构是很紧密的, 也就是说, 样本

集中样本代表着目标的某一部分或者整个目标，我们从样本集中挑选出最有代表的样本来重构目标，这样可以有效的自适应目标外观变化和局部遮挡。

### 4.2.3 基于稀疏表示目标重构的跟踪

当目标重构完成，我们需要在视频帧中定位目标的位置，算法中使用 Kalman 匀速运动模型滤波和穷举搜索，通过最小化重构目标和搜索框  $\Omega_t$  的候选集之间的差异来完成，公式如下：

$$\text{Min}_{(x,y) \in \Omega_t} (\|F - F(C_t(x,y))\|_1) = \text{Min}_{(x,y) \in \Omega_t} \left( \left\| \sum_{k=1}^K A^k \psi^k - F(C_t(x,y)) \right\|_1 \right) \quad (4-25)$$

其中  $C_t(x,y)$  是搜索框中候选集  $\Omega_t$  中样本的位置  $(x,y)$ ， $F(C_t(x,y))$  表示候选样本  $C_t(x,y)$  的特征向量， $F$  是重构的目标的特征向量。

## 4.3 基于稀疏表示的目标重构的跟踪实验与算法评估

本节总结了本章所提出的基于稀疏表示目标重构的跟踪算法，并且在数据集上进行实验，分析实验结果，并且将我们的算法与经典的跟踪算法进行实验对比，用实验数据表明基于稀疏表示的目标跟踪算法在应对部分遮挡，物体旋转等挑战性问题具有较好的效果。

### 4.3.1 实验步骤

表 4-1 基于稀疏表示目标重构的跟踪算法

1. **初始化 ( $t=0$ )**. 初始化跟踪目标，并且计算其特征  $F$ ；
  - 1.1 构建目标的样本集  $S$ ，计算它们的特征集  $A$ ；
  - 1.2 使用  $A$  计算目标的稀疏系数  $\psi$ ；
  - 1.3 基于样本  $S$  和稀疏系数  $\psi$  重构目标；
2. **物体跟踪 ( $t>0$ )**. 在新的视频帧里：
  - 2.1 在搜索区域  $\Omega_t$  里，使用最小化重构目标和候选目标，搜索目标及其位置；
  - 2.2 在下一帧中预测区域  $\Omega_{t+1}$ ；
3.  **$t=t+1$** . 在新的时刻，如果  $\Omega_{t+1}$  更新，进入第四步；如果不更新，返回第二步；
4. **样本集更新 ( $t>0$ )**.
  - 4.1 在样本集中随机选择一个样本，用他代替上一帧的跟踪结果；
  - 4.2 跳转到步骤1.2。

基于稀疏表示目标重构的跟踪算法的基本步骤，如上表所示。

### 4.3.2 跟踪实验结果展示及实验分析

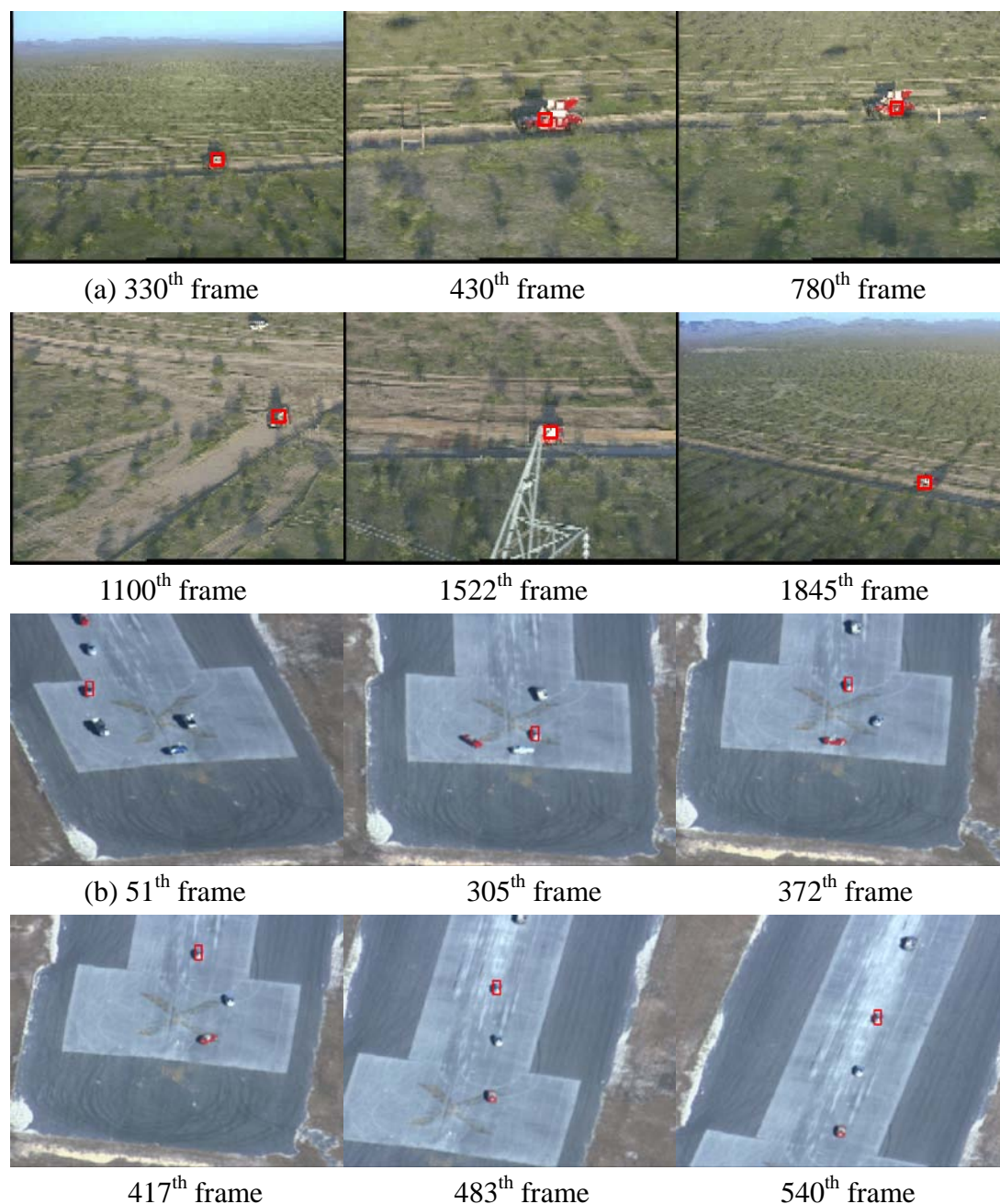


图 4-3 两个跟踪实例

图 4-3 中的数据来自于 CMU 数据集。图 4-3a 中的视频段中摄像机拍摄小车的角度不断改变，并且摄像机在不断的调整焦距使得小车的尺度也在发生变化。从图中可以看出，我们基于稀疏表示目标重构的跟踪方法成功的跟踪了目标和环境的变化。图 4-3b 中的车辆行驶在颜色相似的公路上，以圆形的进行

360 度得转弯然后直行。车辆在旋转过程中外表发生了严重的变化(305<sup>th</sup> 和 372<sup>th</sup> 帧), 基于稀疏表示的跟踪方法成功进行跟踪, 可以有效地应对目标外观变化。

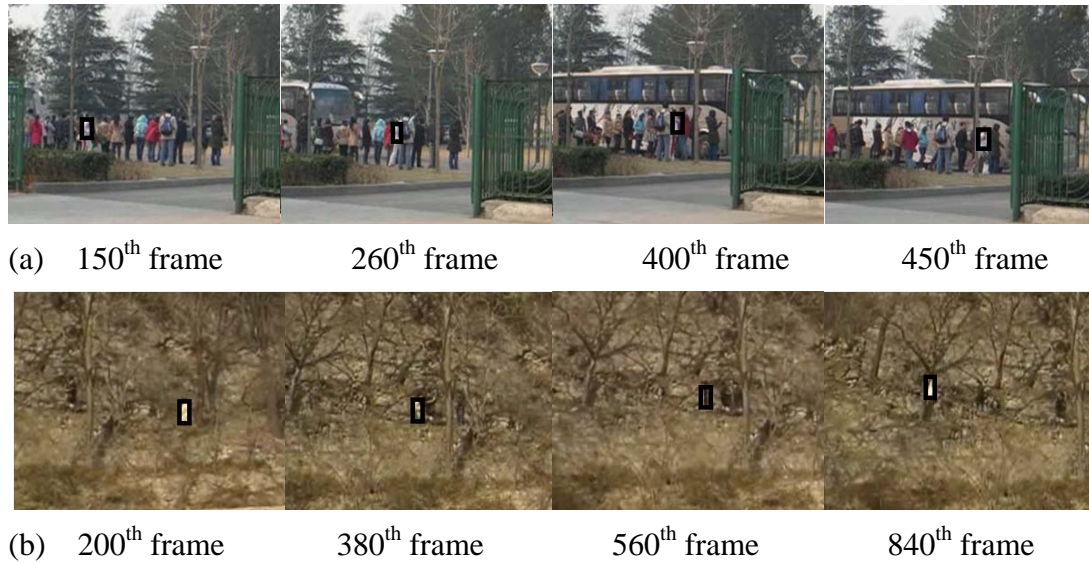


图 4-4 两个跟踪实例

图 4-4 中的视频段来自于 SDL 数据集, 视频中出现了严重的目标遮挡, 并且遮挡物与目标之间非常相似, 属于比较难的跟踪视频。图 4-4a 中的视频段的目标跟踪面临的主要挑战在于目标人被别人频繁的部分遮挡。当目标被遮挡(260<sup>th</sup> 和 400<sup>th</sup> 帧)时,使用初始化整体对象模板进行跟踪时会比较困难。我们的基于稀疏表示的跟踪方法, 通过使用相似样本的子集来重构目标,成功的跟踪部分遮挡的物体。图 4-4b 中的视频段也是很有挑战性的。其中不仅有严重遮挡(380<sup>th</sup> 和 840<sup>th</sup> 帧),而且目标有很大的外观变化(560<sup>th</sup> 帧)。我们的方法可以正确的追踪目。这两段视频的跟踪结果表明, 本节提出的方法可以有效地处理部分遮挡和目标外观变化。

我们将我们基于稀疏表达的方法同其他四个有代表性的跟踪方法进行比较: 基于特征选的 Collins 跟踪方法[73], 基于 SIFT 跟踪方法[72], 基于 Kalman 滤波的跟踪方法[75], 基于粒子滤波的方法[74]。在试验中, 同基于 One-Class SVM 的目标跟踪一样, 我们使用十个视频段进行实验, 相对错误率 DER 来评估我们的跟踪算法。在实验评估中有各种因素影响跟踪效果: 不同的观察点, 光照变化, 目标改变, 部分遮挡。我们在试验中使用相同的初始化, 实验结果图如图 4-5 所示。

正如图 4-5 所示, 我们的跟踪方法的 DER 在 0.04 和 0.1 之间, 低于其它的四方法。这些比较实验显示提出的跟踪算法比其它方法具有良好的表现。但



是，基于稀疏表示的跟踪方法的计算时间要相对长一些，这个跟踪时间与样本集的更新和目标的稀疏系数的重新计算有关。在我们的试验中，平均每 10 帧进行一次样本集的更新，我们在 Pentium IV CPU (2.4G)的计算机上，大约每秒钟跟踪 15 帧，这样的速度几乎同 Kalman 滤波和粒子滤波的跟踪方法一样。

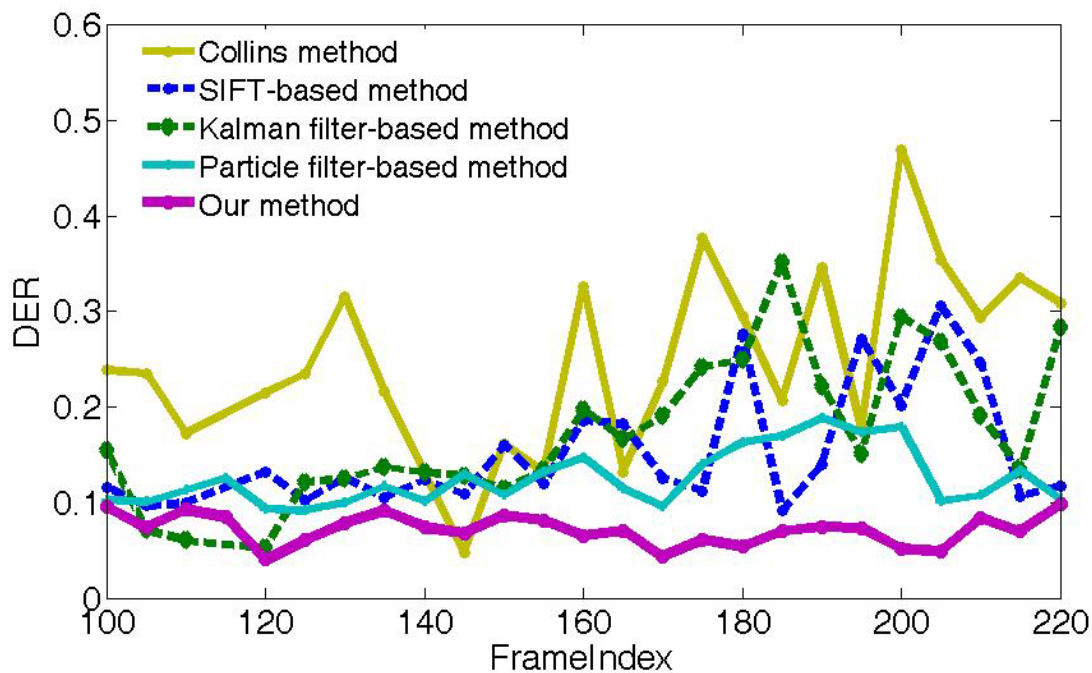


图 4-5 五种跟踪方法平均 DER

## 总结与展望

本文通过从跟踪样本集获取重构目标的样本从而实现目标跟踪。由于目标和周边环境的各种变化，光线对视频成像的影响，尤其是目标的局部遮挡，能够持续的对目标进行跟踪并且精确的确定目标的位置，是视频跟踪对跟踪方法的三大考验。

本文首先简述了目标跟踪的研究意义、国内外研究现状和目前已存在的研究方法，并认真分析了从跟踪样本集合构造以及目标重构的思路和方法；提出了基于 **One-Class SVM** 目标重构的跟踪算法，同时在对跟踪结果以及重构系数深入研究的基础上，发现由于视频帧中的搜索区域（样本集合构造区域）往往要比跟踪目标大很多，这样便会导致样本重构的系数会非常的稀疏，因为背景区域内的样本（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，只有部分反例样本和部分正例样本（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），进一步提出了基于稀疏表示目标重构的跟踪算法。基于所提出的两种算法，我们在公共视频测试集和 **SDL** 视频测试集合进行实验验证，并且将实验结果与经典算法进行比较。实验结果表明，基于 **One-Class SVM** 和稀疏表示目标重构的跟踪算法在目标旋转、背景变化、目标部分遮挡等情况下，可以取得比较鲁棒的实验效果。

本文所提出的两种基于目标重构的跟踪算法虽然取得了一定的成功，但是仍然存在不足。首先在目标大尺度的变化情况下，目标跟踪时需要持续的估计运动目标的边界，才能精确的进行目标跟踪。另一个不足的地方是没有加入运动预测，对目标长时间的遮挡不够鲁棒。在将来的工作中，我们拟加入目标的运动预测估计，研究当目标发生遮挡情况下运动目标的跟踪算法。同时，将对跟踪目标的大尺度变化问题进行深入地研究，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仍然能对其进行鲁棒的跟踪。





## 参考文献

- [1] C. Stauffer, and W. Grimson. Adaptive background mixture models for real-time tracking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1999: 246-252.
- [2] 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, 1998: 214-219.
- [3] 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 Transactions on Robotics and Automation, 1993: 14-35.
- [4] S. Avidan. Ensemble tracking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005: 494-501.
- [5] J. Shi, and C. Tomasi. Good features to track. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1994: 593-600.
- [6] M. Isard, and A. Blake. CONDENSATION-conditional density propagation for visual tracking. International Journal on Computer Vision, 1998:5-28.
- [7] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003: 564-577.
- [8] S. Baker, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. International Journal on Computer Vision, 2004: 221-255.
- [9] D.W. Liang, Q.M. Huang, W. Gao, and H.X. Yao. Online selection of discriminative features using Bayes error rate for visual tracking. 7th Pacific-Rim Conference on Multimedia, 2006: 547-555.
- [10] D. Chen, and J. Yang. Robust object tracking via online spatial bias appearance model learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007: 2157-2169.
- [11] R. Collins, and Y. Liu. Online selection of discriminative tracking features. Proceedings of 9th IEEE International Conference on Computer Vision, 2003: 346-352.

- [12] H. Nguyen, and A. Smeulders. Tracking aspects of the foreground against the background. Proceedings of European Conference on Computer Vision, 2004: 446-456.
- [13] 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, 2005: 1037-1042.
- [14] J.Q. Wang, and Y.S. Yagi. Integrating color and shape-texture features for adaptive real-time object Tracking. IEEE Transactions on Image Processing, 2008: 235-240.
- [15] H. Bohyung, and L. Davis. Object tracking by adaptive feature extraction. IEEE International Conference on Image Processing, 2004: 1501-1504.
- [16] Weiming Hu, Dan Xie and Tieniu Tan. A hierarchical self-organizing approach for learning the patterns of motion trajectories. IEEE Transactions on Neural Networks, 2004: 135-144.
- [17] 李美花. 三维人体跟踪. 大连理工大学硕士学位论文, 2004.
- [18] T. Yang, Q. Pan, J. Li, and S.Z. Li. Real-time multiple objects tracking with occlusion handling in dynamic scenes. IEEE Conference on Computer Vision and Pattern Recognition, 2005: 970-975.
- [19] 彭宁嵩, 杨杰, 刘志, 张风超. Mean-Shift 跟踪算法中核函数窗宽的自动选取. 软件学报, 2005: 1542-1550.
- [20] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. IEEE Transaction on Pattern Analysis and Machine Learning, 2003: 564-577.
- [21] A. Elgammal, R. Duraiswami, and L. Davis. Probabilistic tracking in joint feature-spatial spaces. IEEE Conference on Computer Vision and Pattern Recognition, 2003: 781-788.
- [22] R. Collins. Mean-shift blob tracking through scale space. IEEE Conference on Computer Vision and Pattern Recognition, 2003: 234-240.
- [23] M. Isard, and A. Blake. Condensation - conditional density propagation for visual tracking. International Journal of Computer Vision, 1998: 5-28.
- [24] T.X. Han, V. Ramesh, Y. Zhu, and T.S. Huang. On optimizing template matching via performance characterization. International Conference on

- Computer Vision, 2005: 182-189.
- [25] B.Han, Y. Zhu, D. Comaniciu, and L. Davis. Kernel-based Bayesian filtering for object tracking. IEEE Conference on Computer Vision and Pattern Recognition, 2005: 227-234.
- [26] Y. Huang, and I Essa. Tracking multiple objects through occlusions. IEEE Conference on Computer Vision and Pattern Recognition, 2005: 1051-1058.
- [27] Y. Wu. Robust visual tracking by integrating multiple cues based on co-inference learning. International Journal of Computer Vision, 2004: 55-71.
- [28] Z. Khan, T. Balch, and F. Dellaert. A rao-blackwellized particle filter for eigentracking. IEEE Conference on Computer Vision and Pattern Recognition, 2004: 980-986.
- [29] P. Perez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. European Conference on Computer Vision, 2002: 661-675.
- [30] David G. Lowe. Distinctive image features from scale-invariant key points. International Journal of Computer Vision, 2004: 91-110.
- [31] K. Okuma, A. Taleghani, N. de Freitas etc. A boosted particle filter: multi target detection and tracking. European Conference on Computer Vision, 2004: 28-39.
- [32] Z.J. Han, Q.X. Ye, and J.B. Jiao. Online Feature Evaluation for Object Tracking using Kalman Filter. International Conference on Pattern Recognition, 2008.
- [33] Z. Han, Q. Ye, Y. Liu, et al. Feature evaluation by particle filter for adaptive object tracking. Proceedings of International Society for Optical Engineering, 2009.
- [34] B. Scholkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. Neural computation, 2001:1443-1471.
- [35] S. Avidan. Support Vector Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004: 1064-1072.
- [36] G.Y. Zhu, D.W. Liang, Y. Liu, Q.M. Huang, and W. Gao. Improving Particle Filter with Support Vector Regression for Efficient Visual Tracking. IEEE International Conference on Image Processing, 2005: 1501-1504.
- [37] H.D. Dai, M. Chen, and S.L. Zhou. Support Vector Machine Based Adaptive

- Kalman Filtering for Target Tracking. *General System and Control System*, 2007: 114-118.
- [38] F. Tang, S. Brennan, Q. Zhao, and H. Tao. Co-tracking Using Semi-Supervised Support Vector Machines. *IEEE International Conference on Computer Vision*, 2007: 992-999.
- [39] J. Zhang, D.W. Chen, and M. Tang. Combining Discriminative and Descriptive Models for Tracking. *Asian Conference on Computer Vision*, 2009.
- [40] N. Kabaoglu. Target Tracking Using Particle Filters with Support Vector Regression. *IEEE Transactions on Vehicular Technology*, 2009: 2569-2573.
- [41] E. Candes, J. Romberg and T. Tao. Robust uncertainty principles: exact signals reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 2006: 489-509.
- [42] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 2006: 1289-1306.
- [43] R. Baraniuk. A lecture on compressive sensing. *IEEE Signal Processing Magazine*, 2007: 118-121.
- [44] R. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 2007: 918-925.
- [45] D. Donoho, and Y. Tsaig. Extensions of compressed sensing. *Signal Processing*, 2006: 533-548.
- [46] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk. An architecture for compressive imaging. *IEEE International Conference on Image Processing*, 2006: 1273-1276.
- [47] J. Ma. Single-pixel remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 2009: 199-203.
- [48] S.F. Cotter, and B.D. Rao. Sparse channel estimation via matching pursuit with application to equalization. *IEEE Transactions on Communications*, 2002: 374-377.
- [49] G. Taubock, and F. Hlawatsch. A compressed sensing technique for OFDM channel estimation in mobile environments: exploiting channel sparsity for reducing pilots. *IEEE International Conference on Acoustics, Speech and Signal*

- Processing, 2008: 2885-2888.
- [50] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling, *IEEE Signal Processing Magazine*, 2008: 83-91.
- [51] B.A. Olshausen, and D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996: 607-609.
- [52] D.L. Donoho. For most large underdetermined systems of equations, the minimal  $l_1$ -norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*, 2006: 907-934.
- [53] S. Chen, and D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *Society for Industry and Applied Mathematics Journal on Scientific Computing*, 1999: 33-61.
- [54] D. Donoho, and X. Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Transactions on Information Theory*, 2001: 2845-2862.
- [55] M. Elad, and A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 2002: 2558-2567.
- [56] J. Tropp, and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2008: 4655-666.
- [57] D. Needell, and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching Pursuit. *IEEE Journal of Selected Topics in Signal Processing*, 2010: 310-316.
- [58] L. Rebollo-eira, and D. Lowe. Optimized Orthogonal Matching Pursuit Approach. *IEEE Signal Processing Letters*, 2002: 137-140.
- [59] S.S. Chen, D.L. Donohoand, and M.A. Saunders. Basis Pursuit, 1994 Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers, 1994: 41-44.
- [60] M. Figueiredo, R. Nowak, and S. Wright. Gradient Projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 2007: 586-597.

- [61] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transaction on Image Processing*, 2008: 53-69.
- [62] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2009: 210-227.
- [63] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches, *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [64] T. Wan, N. Canagarajah, and A. Achim. Compressive image fusion. *IEEE International Conference on Image Processing*, 2008.
- [65] K. Egiazarian, A. Foi, and V. Katkovnik. Compressed sensing image reconstruction via recursive spatially adaptive filtering. *IEEE International Conference on Image Processing*, 2007.
- [66] G. Hager, and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1998: 1025-1039.
- [67] A. Yilmaz, X. Li, and B. Shah. Contour based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2004: 1531-1536.
- [68] N. Pérez de la Blanca, J.M. Fuertes, and M. Lucena. Deformable Object Matching Based on Multi-scale Local Histograms. *Articulated Motion and Deformable Objects*, 2004: 154-162.
- [69] R. Collins, X. Zhou, S.K. Teh. An open source tracking testbed and evaluation web site. *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2005.
- [70] CAVIAR Test Case Scenarios at: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR>.
- [71] SDL Data Set at: <http://coe.gucas.ac.cn/SDL-HomePage>.
- [72] Y.J. Li, J.F. Yang, R.B. Wu, F.X. Gong. Efficient Object Tracking Based on Local Invariant Features. *International Symposium on Communications and Information Technologies*, 2006: 697-700.
- [73] R. Collins, Y. Liu, M. Leordeanu. Online Selection of Discriminative Tracking

- Features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005: 1631-1643.
- [74] E. Cuevas, D. Zaldivar, and R. Rojas. Particle filter for vision tracking. Technical Report B, Fachbereich Mathematikund Informatik, Freie Universität Berlin, 2005.
- [75] E. Cuevas, D. Zaldivar, and R. Rojas. Kalman filter for vision tracking. Technical Report B, Fachbereich Mathematikund Informatik, Freie Universität Berlin, 2005.
- [76] S.G. Mallat, Z. Zhang. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, 1993: 3397-3415.

## 发表文章目录

### 已录取文章目录

- Li Li, Zhenjun Han, Jianbin Jiao, Qixiang Ye, “Visual Object Tracking via Sparse Reconstruction”, Proc. of IEEE International Conference on Internet Multimedia Computing and Service, 2010: 151-154.
- Li Li, Zhenjun Han, Qixiang Ye, Jianbin Jiao, “Visual Object Tracking via One-Class SVM”, ACCV2010 Workshop on Visual Surveillance, 2010.



## 附录 1：主要符号对照表

SVM	支持向量机 (Support Vector Machine)
OC-SVM	一类支持向量机 (One Class Support Vector Machine)
SV	支持向量 (Support Vector)
HC	颜色直方图 (Histogram of Color)
HOG	梯度方向直方图 (Histogram of Oriented Gradient)
BP	基追踪 (Basis Pursuit)
MP	匹配追踪 (Matching Pursuit)
OMP	正交匹配追踪算法 (Orthogonal Matching Pursuit)
ROMP	正则化正交匹配追踪 (Regularized Orthogonal Matching Pursuit)
OOMP	最优正交匹配追踪 (Optimized Orthogonal Matching Pursuit)
OPSR	梯度投影稀疏重构 (Gradient Projection for sparse Reconstruction)

## 致 谢

在攻读硕士学位三年的学习生活中，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，由衷地感谢这三年来曾经给予我无数帮助的老师、同学、朋友和家人。

首先要感谢我的导师焦建彬教授，本课题的研究工作是在焦建彬老师的悉心指导下完成的。在我攻读硕士学位期间我所获得的每一步成功，都是和焦老师的指导和鼓励分不开的。三年来，焦老师对我的学习和工作给予了悉心的指导，给我的研究提供了良好的环境和机会，而且教会我做人，使我受益终生。焦老师严谨的治学态度和平易近人的学者风范深深地影响着我。在此我要对焦老师表示衷心的感谢，他的言传身教将使我受益终生。

其次，感谢叶齐祥老师和韩振军师兄，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。他们渊博的专业知识，精益求精的动手能力，和对我的谆谆教导，使我从他们身上学到了不少可贵的东西，让我受益匪浅。

感谢参加开题及中期答辩的各位老师和专家们，他们丰富的经验和无私的工作对论文方向和研究进度的把握和指点给整个研究工作带来了巨大的帮助。

感谢同实验室的韩振军、刘志花、徐冉等同学，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 702 班的全体同学，忘不了他们对我的热心帮助以及一起学习生活中结下的深厚友谊，特别是好友张霖、张红侠、廖景行、梁莘燊。三年的时间里，我们互相关心、互相帮助，共同面对成长中的喜悦和困惑，共同走过的日子是我最难忘的回忆，在即将要毕业之际，我想对你们说的就是，这一路上有你们相伴很精彩。

特别要感谢我的父母和姐姐李惠，这么多年以来他们一直给我最无私的关怀，无微不至的关心和爱护。亲人的爱是最博大的爱，感谢我的父母为我所做的一切，愿他们能够为我而骄傲，祝福他们健康长寿。他们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。

最后，感谢参加论文评审和答辩的各位老师。