

基于视频信息的目标 检测与跟踪

焦建彬、叶齐祥、韩振军、李策 编著

内容简介

本书系统介绍了基于视频信息的目标检测与跟踪的基本问题及其相关处理技术，主要内容涉及计算机视觉的目标检测与跟踪理论、算法和典型应用实例。本书共 6 章，包括绪论、目标表示、目标检测方法、目标检测典型应用、目标跟踪方法、目标跟踪典型问题。本书由浅入深，着重于经典内容和最新进展的结合，并辅以较多的应用范例，可供从事计算机视觉、图像处理、模式识别研究的科技人员及高等院校相关专业的师生参考。

前 言

基于视频信息的目标检测与跟踪是计算机视觉领域的一个备受关注新兴应用方向，是智能监控、人机交互、移动机器人视觉导航、工业机器人手眼系统等应用的基础和关键技术。伴随着网络技术和数字视频技术的飞速发展，计算机视觉正向着智能化、网络化方向不断前进。计算机视觉和应用研究者适时提出了一些基于视频信息的目标检测与跟踪技术，利用计算机视觉和模式识别的方法，对摄像设备捕获的图像序列进行分析，实现对动态场景中目标的定位、识别和跟踪。

本书叙述了基于视频信息的目标检测与跟踪的基本理论和相关的应用技术，主要研究视觉目标表示、目标检测方法及典型应用、目标跟踪方法及典型应用等。本书力图对视觉目标检测和跟踪方法进行分类，总结常规的有效算法，同时将基本理论与实践应用相结合进行具体的实例分析，保证了从应用角度将基本理论有效地串联起来，从而推动基于视频信息的目标检测与跟踪技术的发展。

本书是作者所在的课题组多年来在基于视频信息的目标检测与跟踪方面所做的研究工作总结。内容材料来源于所主持项目研究过程中采集的实验数据、提出的新方法、已在国内外公开发表的文献等。本书涉及了目标检测和跟踪的关键技术和最新动向，同时重视应用实例，既具有前沿性与先进性，又具有很好的实用性，对本领域研究人员和科技工程人员均具有很大的参考价值。

本书第一章由焦建彬、李策撰写，第二章由叶齐祥撰写，第三章由叶齐祥、梁吉祥撰写，第四章由叶齐祥、高文撰写，第五章由韩振军、高山撰写，第六章由韩振军、李策撰写。全书由焦建彬、李策统稿。

在本书的撰写和校稿过程中，中国科学院大学模式识别与智能系统实验室的彭艺、陈孝罡、梁吉祥、武博、武利军等做了大量工作，在此一并表示感谢。本书的完成特别感谢徐冉博士、王建宇博士、梁大为博士、陈杰博士提供的帮助。

因为时间仓促，书中不足之处在所难免，敬请广大读者批评指正。

焦建彬

2014 年 4 月

中国科学院大学

目 录

第 1 章 绪论.....	1
1.1 引言	1
1.1.1 计算机视觉研究的特点	1
1.1.2 计算机视觉与相关领域的关系	2
1.1.3 计算机视觉研究的发展	3
1.1.4 基于视频信息的目标检测与跟踪的研究意义.....	5
1.2 基于视频信息的目标检测与跟踪的研究内容	6
1.2.1 视觉目标检测的研究内容和分类.....	7
1.2.2 视觉目标跟踪的研究内容和分类.....	7
1.3 基于视频信息的目标检测与跟踪的研究现状	10
1.3.1 基于视频信息的目标检测与跟踪的研究难题.....	10
1.3.2 研究现状与应用前景	11
1.4 参考文献	13
第 2 章 目标表示.....	17
2.1 颜色特征	17
2.1.1 颜色空间	17
2.1.2 颜色直方图特征	18
2.2 纹理特征	20
2.3 形状特征	22
2.3.1 Harr-Like 特征.....	22
2.3.2 HOG 特征.....	24
2.3.3 SIFT 特征.....	25
2.3.4 MSO 特征.....	29
2.4 特征融合	30
2.5 特征降维及选择	31
2.5.1 特征降维方法综述	31
2.5.2 特征选择方法综述	32
2.5.3 稀疏特征选择	38

2.6 参考文献	40
第 3 章 目标检测方法	43
3.1 运动目标检测	43
3.1.1 基本概念	43
3.1.2 帧差法	44
3.1.3 背景减除法	45
3.1.4 光流法	52
3.2 图像匹配方法	54
3.2.1 基于像素灰度的匹配方法	55
3.2.2 基于特征提取的匹配方法	56
3.2.3 基于变换域的匹配方法	61
3.2.4 基于投影的匹配方法	64
3.3 人工神经网络	65
3.3.1 神经网络历史	65
3.3.2 感知器	66
3.3.3 多层神经网络	68
3.4 支撑向量机	72
3.4.1 线性支撑向量机	72
3.4.2 非线性支撑向量机	74
3.4.3 经验风险与结构风险	77
3.5 Boosting 方法	79
3.5.1 Adaboost 算法	80
3.5.2 Logiboost 算法	82
3.6 参考文献	83
第 4 章 目标检测典型应用	87
4.1 人脸检测	87
4.1.1 人脸检测数据库	89
4.1.2 人脸检测的一般计算模型	90
4.1.3 人脸检测的常用特征	92
4.1.4 实例分析	94

4.2	人体检测	97
4.2.1	人体检测数据库	97
4.2.2	人体检测常用特征	98
4.2.3	实例分析	100
4.3	车辆检测	107
4.3.1	车辆检测数据库	108
4.3.2	车辆检测的一般方法	108
4.3.3	实例分析	110
4.4	字符类目标检测	114
4.4.1	字符类目标检测数据库	119
4.4.2	字符检测常用特征	119
4.4.3	特征提取	119
4.4.4	特征选择	124
4.4.5	字符检测分类器及方法	126
4.5	参考文献	128
第 5 章	目标跟踪方法	131
5.1	跟踪目标描述	131
5.1.1	跟踪目标的特征表示	131
5.1.2	跟踪目标的表示方法	131
5.2	确定性目标跟踪方法	135
5.2.1	跟踪目标模板提取	135
5.2.2	跟踪目标模板度量	139
5.3	非确定性目标跟踪方法	140
5.3.1	基于 Kalman 滤波器的视觉目标跟踪	142
5.3.2	基于粒子滤波器的视觉目标跟踪	148
5.4	多目标跟踪	159
5.4.1	“最近邻”法	163
5.4.2	概率数据关联滤波器 (PDAF)	163
5.4.3	联合概率数据关联滤波器 (JPDAF)	163
5.4.4	多假设跟踪 (MHT)	165
5.4.5	多维 S-D 分配算法	168

5.4.6 基于稀疏重构分析的多目标跟踪（课题组研究内容五）	171
5.5 参考文献	177
第 6 章 目标跟踪典型问题	185
6.1 快速运动目标跟踪	186
6.1.1 基于积分图的运动目标特征快速提取	186
6.1.2 Mean shift 跟踪	187
6.2 变化及复杂背景下的目标跟踪	191
6.2.1 特征自适应选择和评估	191
6.2.2 稀疏自适应特征表示	193
6.2.3 实例分析	198
6.3 遮挡情况下的目标跟踪	199
6.3.1 基于轨迹预测运动目标跟踪	199
6.3.2 基于目标稀疏重构的目标跟踪	201
6.3.3 实例分析	204
6.4 参考文献	205
后 记	207

第 1 章 绪论

1.1 引言

让机器人具有自动感知能力是人类多年以来的梦想。自信号处理理论和计算机出现以后，人们就开始尝试采用摄像机获得环境的数字图像信息，用计算机实现对数字图像信息处理的过程。这样就形成了一门新兴的学科—计算机视觉^[1]。

计算机视觉的研究目标是使机器不仅能感知三维环境中物体的几何信息，如它的形状、位置、姿态、运动等，而且能对它们进行描述、存贮、识别和理解。计算机视觉是一门交叉性很强的学科，涉及计算机心理学、生理学、物理学、信号处理和应用数学等。近二十年随着各学科的发展和计算机技术的发展，计算机视觉取得了蓬勃的发展，并在不同领域取得了广泛应用：如移动机器人^[2]、视觉导航^[3]、医学辅助诊断^[4]、工业机器人的手眼系统^[5]、地图绘制、物体三维形状分析与识别^[6]、卫星照片解释、智能人机接口^[7]以及监控系统等。

1.1.1 计算机视觉研究的特点

计算机视觉是多学科的交叉与结合，涉及的主要相关学科见图 1.1^[8]。视觉是一个涉及生理心理的复杂过程，不仅仅和眼睛有关，还和大脑的推理、学习有关。研究计算机视觉的目的是要实现对人类视觉模拟的延伸。要用形式化的方法研究视觉问题除了涉及生理、心理之外还需要光学以及相应的数学理论支持。此外计算机是目前实现人工智能的主要工具之一。

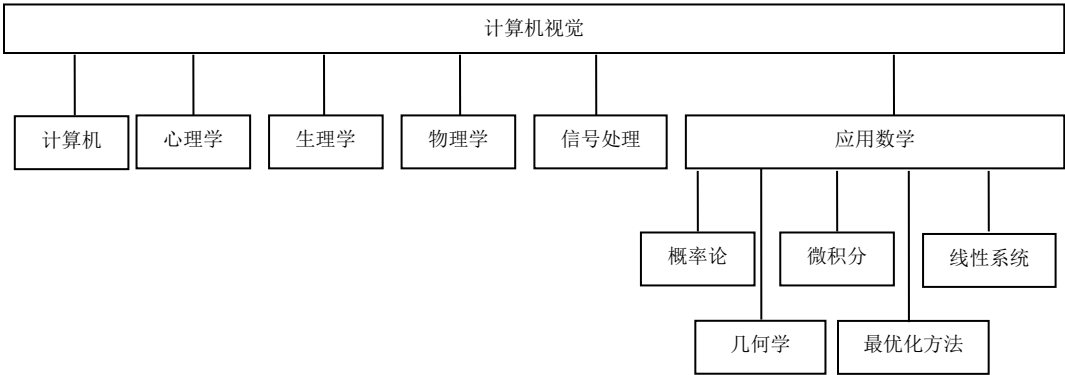


图 1.1 计算机视觉涉及的相关学科

其次，在计算机视觉研究领域中大数问题都是病态的，存在着诸多不确定性因素。由于视觉问题是成像过程的逆过程，成像是从三维向二维投影的过程，因而在这个过程中会丢失深度信息，同时诸如光照、材料特性、朝向、距离等信息都反映成唯一的测量值—灰度。要从这唯一的测量值中恢复上述一个或几个特征参数是一个病态的过程^[9]。不仅如此，成像中由于大气扰动、镜头因素、传感器噪声以及量化噪声等的干扰都会产生失真，这些干扰多具有随机性。对于这些病态问题，通常的处理方法之一是通过引入约束构成正则化问题，转化为优化问题；另一类常见处理方法是利用 Gibbs 分布^[10]将问题转化为对参数的估计问题。

计算机视觉系统同时也是一个知识管理系统。在视觉过程中对一幅图像的理解需要大量的关于领域的知识。因此，视觉系统中的知识管理是一个另外一重要问题。

1.1.2 计算机视觉与相关领域的关系

计算机视觉处理的原始信息又多是图像。因此，这一领域与图像处理有着十分密切的联系。从广义上讲，图像处理包括处理和识别，在这个意义上计算机视觉应当是图像处理的一部分。从狭义上看，图像处理的目的是对输入的原始图像通过处理得到改善的新图像，这种改善指的是增强某些人们关心的部分。在这一意义下，图像处理的内容涉及图像的变换、分割、增强以及压缩等内容。这里的处理是纯数学意义上的处理，一般而言不包括任何形式的反馈。换句话说，这其中不需要任何形式的智能行为。

对于实际应用的视觉系统而言，识别是极其重要的。因此，计算机视觉与模式识别领域有着天然的联系。模式识别研究的是分类问题，其输入应当是被分类对象的特征属性，输出是被分类对象的所属类别。在模式识别领域中，统计方法和句法方法是两种主流的方法。统计模式识别是基于概率和统计理论发展起来的，句法方法则是建立在形式语言理论的基础之上的。相对于上述两种较为成熟的方法而言，近年来发展起来的基于神经网络的方法则要年轻得多，这种方法不论是在生理、心理背景还是数学基础上都有待进一步完善。

与计算机视觉研究目标正好相反，计算机图形学的研究目的在于如何从描述生成逼真的计算机图形，从某种意义上讲这两门学科正好是互逆的。在图形学中所处理的都是确定性的问题，

因此更多的是实时性和逼真度的问题。对于动画等非实时应用而言，逼真度显得尤为重要，在虚拟现实，为了得到实时响应，往往以牺牲逼真度为代价换取处理速度。

处理视觉问题时，巨大的计算量是不容忽视的问题，特别是在处理实时计算时更是如此。解决这一问题可以从两方面入手，一是从算法上改进，降低算法复杂度，其次就是从计算机上入手，提高其性能。性能可以随着通用计算机能力的提高而改进，但是针对特殊问题如特定的数据结构、公用的算法，在系统结构上加以改进也是重要的措施之一，这就要求从计算机体系结构上加以支持。因此，专用计算机体系结构的研究也是计算机视觉中值得考虑的问题之一。

1.1.3 计算机视觉研究的发展

视觉研究在 Roberts 之前都是基于二维的，而且多数是采用模式识别的方法完成分类工作。Roberts 用程序成功地对三维积木世界进行解释^[11]。在其后类似的研究中，Gumzina 在视觉处理研究中引入符号化处理和启发式方法。以后 Huffman、Clowes 以及 Waltz 等人对积木世界进行了研究并分别解决了由线段解释景物和处理阴影等问题。积木世界的研究对视觉研究的发展起了促进作用，但对于稍微复杂的景物便难以奏效。

20 世纪 70 年代中期，以 Marr、Barrow 和 Tenenbaum 等人为代表的一些研究者提出了一整套视觉计算的理论来描述视觉过程^[12]，其核心是从图像恢复物体的三维形状。在视觉研究的理论上，以 Marr 的理论影响最为深远。其理论强调表示的重要性以及从不同层次上去研究信息处理问题，在计算理论和算法实现上又特别强调计算理论的重要性。这些理论多数建立在对人类感知三维信息分析的基础上，尽管这些方法在数学上是可行的，但由于各种干扰的存在以及逆成像的问题，使得问题本身是病态的，也即是问题的解不唯一，或不连续依赖于初始条件。

进入 80 年代中后期，随着移动式机器人研究的发展，人们大量运用空间几何的方法以及物理知识来研究视觉，其目标主要是完成对道路障碍的识别处理。这一时期引入主动视觉的研究方法，并采用了距离传感器并引入融合技术等。由于这种研究方法可直接取得深度图像或通过移动获取深度图，因而使很多病态问题变成良态的。此外，在视觉的研究中重视了对定性视觉、有目的视觉的研究。

上述这些理论和方法的研究有力地促进了相关应用的发展。这些应用领域包括对图像（特别是航空与卫星图像）的解释、精确制导、移动机器人视觉导航、医学辅助诊断、工业机器人的手眼系统、地图绘制物体三维形状分析与识别以及智能人机接口等。

早期进行数字图像处理研究的目的之一，就是要通过采用数学技术提高照片的质量，辅助进行航空与卫星图像的判读与分类。由于需要判读的图像数量巨大，于是希望有自动的视觉系统进行判读解释，在这样的背景下，产生了许多判读系统，如 ACRONYM 和 SPAM 等。

精确制导一直是武器系统中的一个研究热点。常用的精确制导方式包括激光制导、电视制导和图像制导，在导弹系统中常常将惯性制导与图像制导相结合，利用图像进行精确的末端制导。由于可见光图像光照条件等的影响，因此，在图像制导中常采用多谱图像或者根据不同的光照条件采用不同的图像。典型的采用地图匹配技术的导弹有美国的战斧、潘兴 II 等。

对于移动机器人而言，路径规划是极其重要的。规划的前提是对环境的了解。当然某些场合可以采用遥视、遥控技术，但是当机器人处在几十万公里甚至是以光年计的太空中时，遥视和遥控将越来越不现实。即使是在地面上，当由于传输媒介的原因而通讯不畅时，自主导航对于移动机器人将十分重要。因此，随着移动式机器人的发展，产生了一些专门用于道路分析和避障的视觉系统。典型的系统如卡内基梅隆大学研制的自主行驶汽车 Navlab 上的视觉系统。该系统采用包括光学摄像机、激光测距仪以及超声在内的多种传感器获取道路信息，利用多传感器融合和基于聚类的方式完成室外道路的跟踪和识别。此外在美国麻省理工大学人工智能实验室 (MIT-AI-Lab) 研制的 Polly 机器人中引入了所谓的轻度视觉的概念，利用这一方法大大加快了道路识别和避障的速度。尽管目前移动机器人的视觉系统研究已取得了很大的进步，但实际上移动机器人都还处于实验阶段，以美国发射到火星的探路者号为例，目前采用的仍然是遥控和遥视方法。

目前的医学存在两方面的问题，一是医生的训练无法科学化，二是诊断的结果加入了较多的人为因素。现代医疗仪器的使用在一定程度上解决了这两方面的问题，但是仪器除了直接提供数据外，几乎不能进行任何自动分析工作，很多仪器如 X 光机、CT、B 超、各种显微诊断设备等提供的都是图像，这样诊断的结论又再一次依赖判读者的经验。利用计算机视觉技术可以减少人为因素的影响。现在医疗图像的辅助诊断系统已经在早期疾病诊断方面起到非常重要的作用，同时医学图像处理近年来也成为计算机视觉领域中的重要研究方向。

工业机器人的手眼系统是计算机视觉应用最为成功的实例之一，由于工业现场的诸多因素如光照条件、成像方向均是可控的，因此，使得问题大为简化，有利于构成实际的系统。目前市场上已有相当数量的手眼系统。

计算机视觉的另一个重要应用是用于构成智能人机接口。现在计算机与人的交流还是机械式的，计算机无法识别用户的真实身份，除了键盘、鼠标，其他输入手段还不成熟。利用计算机视觉技术可以使计算机检测到用户是否存在，鉴别用户身份及识别用户的体势。此外这种人机交互方式还可以推广到一些需要人和机器交互的场合，如入口安全控制、过境人员的验放等。

上述这些理论和应用的研究，使得计算机视觉在过去三十年中成为人工智能研究中最活跃的一部分。通过这些研究，人们对这一领域中的困难和问题也有了更深入的认识。总之，随着对定量研究的重视，新的描述方式、求解手段的探索，以及距离传感器等设备的使用，计算机视觉的研究必将会迎来一个更加繁荣的时代。

1.1.4 基于视频信息的目标检测与跟踪的研究意义

1.1.4.1 视觉目标检测的研究意义

图像或视频中的目标检测，意在基于目标表观和轮廓区域等信息，准确地对图像中感兴趣的目标进行定位，即将目标的定位和识别合二为一。它的基本任务是从图像序列中检测出运动信息，简化图像处理过程，得到所需的运动矢量，从而能够识别与跟踪物体。于运动目标的正确检测与分割影响着运动目标能否被正确跟踪与分类，视觉目标检测是数字图像处理技术的一个主要部分，成为了视频监控系统研究中的一项重要研究课题。

虽然在近年来的图像处理和模式识别领域中，人脸、车牌等其他目标检测方法已经日趋成熟，但是在复杂环境下可靠的目标检测算法还有待进一步研究，其原因在于：首先，一些目标是一个非刚性的、多姿态的、多角度的物体，例如人体目标；其次，含有目标的图像背景一般都复杂多变；再次，目标很容易被其他目标或者其他物体遮挡。因此，通过运用机器学习与模式识别中的相关知识，使计算机能够自动、准确地检测目标，实现鲁棒、快速地目标自动提取和检测显得极为重要，与此相关的问题逐渐也成为模式识别、图像处理领域内的研究热点。此外，开展此方面的研究对目标的模式表达及检测方法等领域内核心问题的解决具有重要的理论意义。

1.1.4.2 视觉目标跟踪的研究意义

视觉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，进而为平台随动控制、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。

经过近五十年的研究和发展的，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，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。

1.2 基于视频信息的目标检测与跟踪的研究内容

随着对定量研究的重视，新的描述方式、求解手段的探索，以及距离传感器等设备的使用，

计算机视觉的研究必将会迎来一个更加繁荣的时期。

1.2.1 视觉目标检测的研究内容和分类

从应用的角度，目标检测算法可以简单地分为两类：基于静止背景视频的检测和基于单帧静止图片的检测。

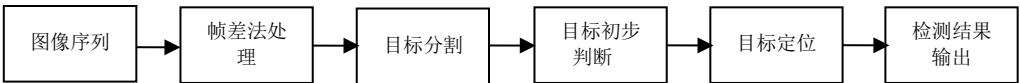


图 1.2 目标检测处理流程

如图 1.2，基于静止背景视频中的的检测算法主要是结合运动信息来识别“动目标”。目前代表性的基于动态视频帧的检测算法有背景差分法^[13]、帧差法^[14]等。这些基于动态视频帧的检测算法在复杂、动态背景下，效果不是很好，原因在于这些算法一般只适用于静止的背景。而且，在静止的背景中，前后帧之间光线的轻微变化及树木的摇摆等多种干扰信息，都极易被判别为目标。由于视频中背景的复杂化与动态化，目前基于动态视频帧的检测算法的发展受到了一定的约束，因此在应用方面会受到一些影响。在最近的一些研究中，研究者们更关注基于单帧静止图片的检测。虽然在静止图片的检测中，目标的定位相对繁琐一些，然而，一旦在单帧图片上目标检测的研究得到突破，结合北京建模，将会在实际应用中取得很好的效果。

基于单帧静止图片的检测方法中，根据目标构建模式的划分不同，可分为整体模型检测与部分模型检测方法。整体模型^[15,16]是基于整个训练样本窗口，提取目标的整体或者部分信息作为目标特征，然后采用某种分类准则获得判别函数。部分模型是指将目标进行分块，例如，按着图片的大致比例将目标人体分为头部区域，躯干区域^[17]等或者基于目标的自然形态进行分块^[18]。无论是整体模型检测还是部分模型检测，都涉及两部分内容，一是特征表述（特征描述子），它是指从图像中提取出表示目标的特征向量，该特征应该尽量对光照、背景、表观等因素的变化不敏感。二是分类器的构建，它主要是使用前面所提取目标的某种特征，按照某种学习准则，获取分类函数的过程。特征表述和分类器，主要是让计算机意识到什么样的模式是属于目标的，什么样的模式是属于背景的。

1.2.2 视觉目标跟踪的研究内容和分类

当检测到图像中的目标存在，精确检测出目标的位置，根据前面结果建立的目标模板，进

行特征匹配，跟踪目标。目标跟踪处理流程见图 1.3。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。

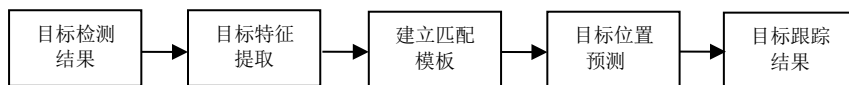


图 1.3 目标跟踪处理流程

1.2.2.1 研究内容

近年来研究人员开发了大量新的跟踪算法，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. 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，可以很容易地将目标从背景中区分出来。目标跟踪过程中常用的特征如下^[19]：

(1)颜色特征 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比较敏感。

(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^[20]。

(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^[21]。光流常被用来表示基于运动的分割与跟踪的特征。

(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^[22]。与边缘特征相似，纹理特征对于光照的变化也不敏感。

面对复杂背景及光照条件变化，如何基于特征组合来更加准确完备的描述目标还有待进一

步的研究。

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^[23]被人们广泛使用的匹配方法。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。Mean-Shift 通过对目标可能出现的搜索区域中的每一个象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匹配。

3. 运动模型

运动模型主要是利用跟踪目标的运动信息来校正目标当前的运动参数, 并预测目标未来的运动趋势。当模型是线性的或者服从高斯分布时, 可以采用卡尔曼滤波器 (Kalman Filter)^[24]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, 粒子滤波 (又称为序列蒙特卡洛滤波)^[25]为问题的解决提供了一种非常有效的手段。粒子滤波的基本思想是通过一个粒子 (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, 然后通过运动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。

1.2.2.2 研究分类

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. 目标数目

根据目标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^[26]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. 目标类型

根据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近似刚性区域及其之间的相互关联来进行建模。

3. 摄像机数目

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。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扩大摄像机的视野。另一方面,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解决办法。

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/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^[27]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分类模型来检测目标,进而对其进行跟踪。

5. 传感器类型

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^[28]等。这里需要指出的是目标跟踪在雷达,声纳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中借鉴而来。

6. 跟踪速度

根据跟踪速度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。这种应用对于算法的实时性要求不高,而对精度有更高的要求。

1.3 基于视频信息的目标检测与跟踪的研究现状

1.3.1 基于视频信息的目标检测与跟踪的研究难题

视觉目标检测和跟踪仍有下述的关键技术有待解决和突破,也是今后研究的难点问题。

1.目标的表达方式不统一。目前大多数目标检测与跟踪算法都是基于监督学习方法,即通过现有的训练样本,让计算机学习到目标统一的表达模式。然而,实际生活中目标姿态的多样

性、视角变化、衣服样式等使得训练样本的模式并不是很统一。因此,即使再好的分类学习方法也难于建立统一的表达方式。

2. 在检测与跟踪过程中,目标姿态、视角变化、服饰等表现因素与训练过程中学习得到的目标模型有差别,另外目标与摄像机的距离的远近也会在一定程度上改变目标目标的特征,很大程度上的影响着检测的精度。这些问题的产生,对目标特征描述子以及分类器的设计提出了更高的要求。总而言之,目标的特征表示与分类算法的设计,在很大程度上决定了目标检测与跟踪算法的检测精度。

3. 目标的部分及完全遮挡。目标有可能受到背景或者其他目标目标的遮挡。部分遮挡造成目标的部分特征缺失,而这种特征缺失是随机性的,计算机很难捕捉到这种缺失信息,对此只能依赖未被遮挡的特征信息。这种信息随机性的缺失,对特征的表述及分类方法提出了更高的要求。对于目标的部分遮挡,一些算法提出使用图像分割或深度信息等手段解决,但是暂时还没有很好的算法可以完全的解决这一问题。

4. 背景模式与光照亮度的影响。实际应用中的背景比较复杂,存在各式各样、形态各异的物体,例如:路边树木、栏杆等,这些信息和学习到的目标竖直模式比较相似,此外,背景光线亮度变化等这些干扰信息,如同一目标在不同的光线亮度条件下会产生明显不同的图像模式。因此,背景模式的变化与光照亮度的影响是实际应用系统中阻碍性能提升的瓶颈。

1.3.2 研究现状与应用前景

近些年来,随着计算机软硬件性能的不断发展,各种面向复杂应用背景的基于视频信息的目标检测与跟踪的系统也随之大量涌现。正是如此大的应用前景,基于视频信息的目标检测与跟踪引起许多国家的高度重视,并投入大量资金和科技人员进行了广泛研究。

国外对于视觉跟踪理论研究起步较早^[23-26]。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。

当前, 国际上的有很多高级别的学术会议和期刊如: Computer Vision and Pattern Recognition (CVPR)、International Conference on Computer Vision (ICCV)、European Conf. on Computer Vision (ECCV)、Asian Conf. 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 等都将基于视频信息的目标检测与跟踪作为主题内容之一, 包含了视觉跟踪领域的最新研究成果, 为该领域研究人员提供了广泛的交流机会。

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开展这方面的研究^[27-28]。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、精彩片断分析的研究; 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、交通场景监控系统和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。

目标检测与跟踪算法在近几年比较热门, 层出不穷。目前现有的目标检测与跟踪算法, 虽然能达到一定的检测性能, 但是因为目标姿态多样性、遮挡等问题, 加上检测算法对实时性、鲁棒性、准确性等要求, 使得现有的检测和跟踪算法仍有待改进。例如, 一些算法误检率或者漏检率比较高; 一些算法虽然在检测精度上有了很好的提高, 但是检测时间过长。如何在复杂背景下, 兼备速度与精度从而提高目标检测算法的实时性、准确性、鲁棒性是目前目标检测的难点。

近年来, 各行各业对基于视频信息的目标检测与跟踪的需求不断升温。基于视频信息的目标检测与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。在国防领域,鲁棒的人体目标检测可用于辅助电子探测设备自动识别危险人体目标,帮助特种部队及空降兵来探测周边区域的情况并保卫军营的安全,防止人为的偷袭和破坏等活动;在民用领域,该技术可用于重点建筑(博物馆、体育场馆、重要桥梁等)及重要地点(核电站、水电站、地铁站、银行等)的安全保卫工作;随着城市规模的扩大,住宅小区逐渐成为安全防范的重点。结合红外/近红外等视频设备,该技术可强化小区安全监控的能力、减少犯罪,从而增强居民的安全感。在智能交通领域,可以及时监控交通事故并可以在十字路口、新建高速路口等进行自动行人检测和安全预警。在汽车辅助驾驶系统中,可以通过车载视频,实时分析路况和检测行人变化,从而警示驾驶员路况信息。

1.4 参考文献

- [1] Sage K, Young S. Security Applications of computer vision. *IEEE Transactions on Aerospace and Electronic Systems*, 1999, 14 (4): 19~29.
- [2] Takahashi H, Morisawa M, Ohnishi K. Mobility of a mobile robot. *The 8th IEEE International Workshop on Advanced Motion Control*, 2004: 253~257.
- [3] Gaspar J, Winters N, Santos-Victor J. Vision-based navigation and environmental representations with an omnidirectional camera. *Robotics and Automation*, 2000, 16(6): 890~898.
- [4] Brown R A, Zhu H M, Mitchell J R. Distributed vector processing of a new local multiscale fourier transform for medical imaging applications. *Medical Imaging*, 2005, 24(5): 689~691.
- [5] Aoki T. A tracking control to multiple objects for plural hand-eye systems. *The 6th International Workshop on Advanced Motion Control*, 2000: 1~6.
- [6] Li T, Wu C K, Tsui H T, Liu S G. Algorithm for 3D reconstruction with both visible and missing data. *Electronics Letters*, 2003, 39(23): 1640~1642.
- [7] Guedj R. *Human-machine interaction and digital signal processing*. 1982.

- [8] 高文, 陈熙霖, 计算机视觉—算法与系统厦理 2000, 北京:清华大学出版社.
- [9] Fung J, Mann S. Computer vision signal processing on graphics processing units. 2004, 93-6(5).
- [10] Chan M, Levitan E, Herman G T. Image-modeling Gibbs distributions for Bayesian restoration. 1994: 7~12.
- [11] Roberts L. Machine perception of three electron-optical information processing. Cambridge, MA: MIT Press, 1965.
- [12] Marr D 著. 姚国正等译. 视觉计算理论. 北京: 科学出版社, 1988.
- [13] HagaT, Sumi K, Yagi Y. Human detection in outdoor scene using spatio-temporal motion analysis. IEEE International Conference on Pattern Recognition, 2004: 331~334.
- [14] Elzein H. A motion and shape-based pedestrian detection algorithm. IEEE Intelligent Vehicles Symposium, 2003:500~504.
- [15] Mu Y, Yan S, Liu Y, Huang T, Zhou B. Discriminative Local Binary Patterns for Human Detection in Personal Album, IEEE Conference on Computer Vision and Pattern Recognition, 2008, 23(28): 1~8.
- [16] Wang X Y, Han T X, Yan S C. An HOG-LBP Human Detector with Partial Occlusion Handling, IEEE International Conference on Computer Vision, Kyoto, 2009.
- [17] Vinay D, Neumann J, Ramesh V, Davis L S. Bilattice-Based Logical Reasoning for Human Detection, IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [18] Andriluka M, Roth S, Schiele B. Pictorial Structures Revisited: People Detection and Articulated Pose Estimation, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1014~1021.
- [19] Yilmaz A, Javed O, and Shah M, Object Tracking: A Survey, ACM Computing Surveys, 2006.
- [20] Canny J. A computational approach to edge detection, IEEE Transaction Pattern Analysis and Machine Intelligence, 1986: 679~698.
- [21] Szeliski R, Coughlan J. Spline-based image registration. International Journal of Computer Vision, 1997: 199~218.

- [22] Haar A. Zur Theorie der orthogonalen Funktionensysteme. Mathematische Annalen, 1910: 331~371.
- [23] Collins R. Mean-shift blob tracking through scale space. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp:234-240, 2003.
- [24] Cuevas E, Zaldivar D, Rojas R. Kalman filter for vision tracking, Technical Report B, Fachbereich Mathematikund Informatik, Freie Universität Berlin, 2005.
- [25] 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.
- [26] Cox I J, Hingorani S L. An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. IEEE Transaction Pattern Analysis and Machine Intelligence, 1996: 138~150.
- [27] Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects. IEEE Transaction Pattern Analysis and Machine Intelligence, 2006: 657~662.
- [28] Yilmaz A, Shafique K, Shah M. Target tracking in airborne forward looking infrared imagery, Image and Vision Computing, 2003: 623~635.

第 2 章 目标表示

2.1 颜色特征

在图像的形状、颜色、纹理等特征中，颜色特征是最显著、最可靠、最稳定的视觉特征。彩色图像中含有丰富的信息（颜色与图像中所包含的物体或场景的相关性很高），人们对一幅图像的印象，往往从图像中颜色的空间分布开始，因此颜色特征被广泛应用于图像处理和模式识别领域。此外，其相对于几何特征而言，颜色特征对图像中子对象的大小与方向的变化都不敏感，具有相当强的鲁棒性。

2.1.1 颜色空间

颜色空间也称彩色模型(又称彩色空间或彩色系统)。现在采用的大多数颜色模型都是面向硬件或面向应用的。颜色空间从提出到现在已经有上百种，大部分只是专用于某一领域。

颜色空间有许多种，常用有 RGB、CMY、HSV、HIS。采用红、绿和蓝 3 种基色来匹配所有颜色的模型称为 RGB 颜色空间。国际照明委员会将 3 种单色光的波长分别定义为红色(700nm)、绿色(546.1nm)和蓝色(453.8nm)。根据三基色原理，任意一种颜色可以由下式匹配：

$$C = rR + gG + bB \tag{2-1}$$

式中的系数 r、g、b 分别为三基色红、绿和蓝的比例系数。RGB 颜色模型通常用于彩色阴极射线管和彩色光栅图形显示器。各个原色的光能叠加在一起产生复合色。RGB 是依据人眼识别的颜色定义出的空间，可表示大部分颜色。但在科学研究一般不采用 RGB 颜色空间，因为颜色细节难以进行数字化的调整，RGB 是最通用的面向硬件的彩色模型。

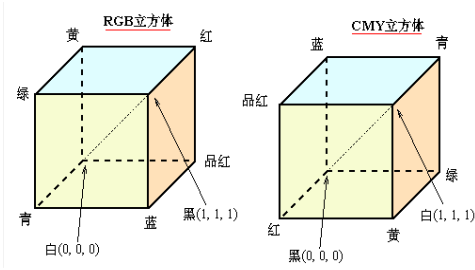


图 2.1 RGB 和 CMY 立方体示意图

CMY 是工业印刷采用的颜色空间。具体应用如打印机：一般采用四色墨盒，即 CMY 加黑色墨盒。与 RGB 类似，CMY 代表 3 种印刷油墨的颜色：青色 Cyan、洋红色 Magenta、黄色 Yellow。从理论上来说，只需要 CMY 三种油墨就足够了，它们三个加在一起就应该得到黑色。但是由于目前制造工艺还不能造出高纯度的油墨，三者相加的结果实际是一种暗红色，因此还需要加入一种专门的黑墨来调和。图 2.1 是 RGB 和 CMY 立方体的示意图。

HSV、HSI 两个颜色空间都是为了更好地数字化处理颜色而提出来的。现有许多种 HSX 颜色空间，其中的 X 可能是 V，也可能是 I，依据具体用途 X 含义不同。H 是色调，S 是饱和度。

Lab 颜色空间用于计算机色调调整与彩色校正。它独立于设备的彩色模型。Lab 颜色是以一个亮度分量 L 及两个颜色分量 a 和 b 来表示颜色的。其中 L 的取值范围是 0-100，a 分量代表由绿色到红色的光谱变化，而 b 分量代表由蓝色到黄色的光谱变化，a 和 b 的取值范围均为-120-120。Lab 模式所包含的颜色范围最广，能够包含所有的 RGB 和 CMY 模式中的颜色。CMY 模式所包含的颜色最少，有些在屏幕上看到的颜色在印刷品上却无法实现。

2.1.2 颜色直方图特征

基于颜色的特征提取，研究者已经提出了多种方案，譬如颜色直方图、主色调、颜色矩(Color Moments)、颜色集(Color sets)、聚类、扫描线投影等，其中颜色直方图是目前广泛使用的特征。颜色直方图描述的是图像中每个亮度值的像素的数量分布。颜色直方图广泛应用于计算机视觉中，比如在处理遥感图像时，每个波段的颜色直方图能提供关于原始图像质量的信息，如对比度的强弱，是否为多峰值等。通过标记帧间显著的边缘和颜色的变化，直方图被用于检测视频中场景的变化。边缘、色彩和角等直方图构成了通用特征类型。色彩和边缘的直方图序列还可以用来识别网络视频是否被复制等^[1]。颜色直方图中的数值都是统计得到的，描述了该图像中关于颜色的数量特征，可以反映图像颜色的统计分布和基本色调；直方图只包含了该图像中某一颜色值出现的频数，而丢失了像素的空间位置信息；任一幅图像都能唯一的给出一幅与它对应的直方图，但不同的图像可能有相同的颜色分布，从而就具有相同的直方图，因此直方图与图像是一对多的关系；如将图像划分为若干个子区域，所有子区域的直方图之和就是全图的直方图；一般情况下，由于图像上的背景和前景物体颜色分布明显不同，从而在直方图上会出现

双峰特性，但背景和前景颜色较为接近的图像不具有这个特性。

颜色直方图是在目标检测中被广泛采用的颜色特征。它所描述的是不同色彩在整幅图像中所占的比例，而并不关心每种色彩所处的空间位置，即无法描述图像中的对象或物体。颜色直方图特别适于描述那些难以进行分割的图像。颜色特征的表达涉及若干问题。首先，我们需要选择合适的颜色空间来描述颜色特征；其次，我们要采用一定的量化方法将颜色特征表达为向量的形式；最后，还要定义一种相似度（距离）标准用来衡量图像之间在颜色上的相似性。

颜色直方图可以基于不同的颜色空间和坐标系。最常用的颜色空间是 RGB 颜色空间，原因在于大部分的数字图像都是用这种颜色空间表达的。然而，RGB 空间结构并不符合人眼对颜色相似性的主观判断。因此，有人提出了基于 HSV 空间、Luv 空间和 Lab 空间的颜色直方图，因为它们更接近于人们对颜色的主观感觉。其中 HSV 空间是直方图最常用的颜色空间。它的三个分量分别代表色彩（Hue）、饱和度（Saturation）和值（Value）。

直方图是对数据的统计，将统计值组织到一系列事先定义好的 bin 中。Bin 中的数值是从数据中计算出的特征的统计量，这些数据可以是梯度、方向、色彩或其它任何特征。计算颜色直方图需要将颜色空间划分成若干个小的颜色区间，每个小区间成为直方图的一个 bin。这个过程称为颜色量化。然后，通过计算颜色落在每个小区间内的像素数量可以得到颜色直方图。颜色量化有许多方法，例如向量量化、聚类方法或者神经网络方法。最为常用的做法是将颜色空间的各个分量（维度）均匀地进行划分。相比之下，聚类算法则会考虑到图像颜色特征在整个空间中的分布情况，从而避免出现某些 bin 中的像素数量非常稀疏的情况，使量化更为有效。另外，如果图像是 RGB 格式而直方图是基于 HSV 空间的，我们可以预先建立从 RGB 空间到 HSV 空间的查找表（look-up table），从而加快直方图的计算过程。上述的颜色量化方法会产生一定的问题。设想两幅图像的颜色直方图几乎相同，只是互相错开了一个 bin，这时如果我们采用 L1 距离或者欧拉距离计算两者的相似度，会得到很小的相似度值。为了克服这个缺陷，需要考虑到相似但不相同的颜色之间的相似度。一种方法是采用二次式距离。另一种方法是对颜色直方图事先进行平滑过滤，即每个 bin 中的像素对于相邻的几个 bin 也有贡献。这样，相似但不相同颜色之间的相似度对直方图的相似度有所贡献。

选择合适的颜色小区间（即直方图的 bin）数目和颜色量化方法与具体应用的性能和效率要求有关。一般来说，颜色小区间的数目越多，直方图对颜色的分辨能力就越强。然而，bin 的数目很大的颜色直方图不但会增加计算负担，也不利于在大型图像库中建立索引。而且对于

某些应用来说，使用非常精细的颜色空间划分方法不一定能够提高检索效果，特别是对于不能容忍对相关图像错漏的那些应用。另一种有效减少直方图 bin 的数目的办法是只选用那些数值最大（即像素数目最多）的 bin 来构造图像特征，因为这些表示主要颜色的 bin 能够表达图像中大部分像素的颜色。实验证明这种方法并不会降低颜色直方图的检索效果。事实上，由于忽略了那些数值较小的 bin，颜色直方图对噪声的敏感程度降低了，有时会使检索效果更好。

2.2 纹理特征

局部二元模式（Local Binary Pattern, LBP）是 Ojala 等^[1]提出的一种有效的纹理描述方法，该方法通过 LBP 算子来提取灰度图像中局部相邻区域的纹理特征。Ojala 等给出的 LBP 算子形式是一个固定大小为 3x3 的矩形块，包含了中心点和周围八个邻域的共九个灰度值，计算 LBP 特征值时，将四周邻域的 8 个灰度值与中心像素点的灰度值相比较，大于等于中心灰度值的子块表示为 1，否则表示为 0，然后为各个邻域像素点赋予不同的权值，计算得到的十进制数值作为该 3x3 矩形块的特征值，最后计算整幅图像中每个独立特征值的出现概率，并以直方图的形式作为该图像的纹理特征的描述。LBP 算子的特征计算过程如图 2.2。

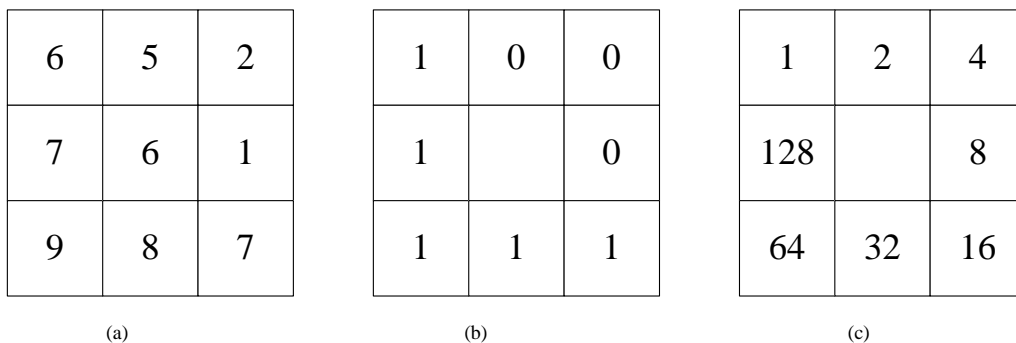


图 2.2 LBP 特征计算示意图

图中，(a)是当前算子覆盖区域的灰度值分布图，中心点的灰度值为 6。根据 8 个邻域像素点灰度值和中心点灰度值的大小比较，得到该区域的二值图(b)，然后按图(c)所示为不同邻域像素点赋予不同的权值，即该区域的二进制特征串为:10101001，其对应的十进制数为 169，此即为当前矩形块的 LBP 特征值。图 2 显示了一副人脸图像采用原始 LBP 算子提取的 LBP 特征图及其纹理特征的直方图。

由于所提出的 LBP 算子为固定大小的矩形块，无法提取大尺度结构纹特征，ojala 等对原始 LBP 算子进行了扩展，其基本思想：灰度图像的纹理特征通过灰度的分布来体现。假设 是图像的一个区域，则该区域的纹理特征 F_L 可以通过该区域的灰度分布 G 来描述，即：

$$F_L = G(g_c, g_0, g_1, \dots, g_{P-1}) \quad (2-2)$$

其中： g 为区域中心点 c 的灰度值， $g_i (i=0,1,\dots,P-1)$ 为 P 个以中心点 c 为圆心，以 R 为半径的圆周上的等距离分布的邻域像素点的灰度值，不同的 P ， R 的取值，对应了不同尺度的 LBP 算子，如图 2.3。

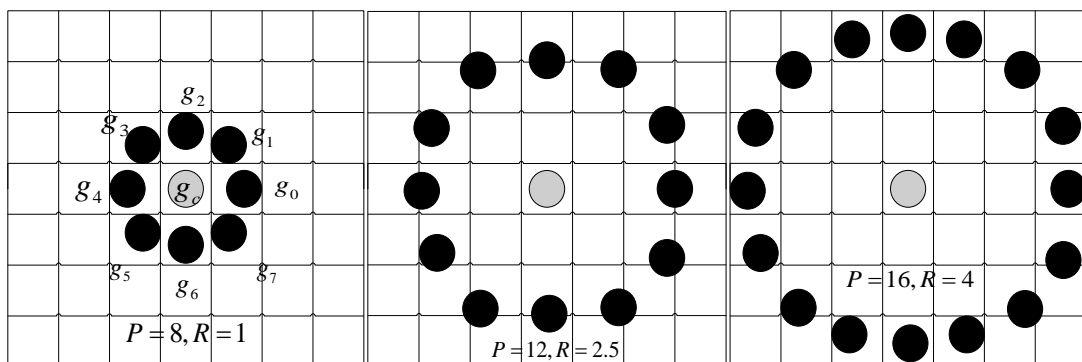


图 2.3 不同 P ， R 值的 LBP 算子形式

通常情况下，当前区域所有像素点的灰度值同时增加或减少一个灰度值常数，对区域的纹理结构没有影响，因此，该区域的纹理特征 可以通过区域灰度分布 的另外一种形式来表示：

$$F_L = G(g_c, s_0, s_1, \dots, s_{P-1}) \quad (2-3)$$

其中： $s_i = g_i - g_c (i=0,1,\dots,P-1)$ 表示第 i 个邻域像素点和中心像素点的灰度差值，根据 s_i 数值的正负，按式(2-4)可以得到二元化的区域纹理特征：

$$s_i = \begin{cases} 1, & g_i - g_c \geq 0 \\ 0, & g_i - g_c < 0 \end{cases} \quad (2-4)$$

假设邻域像素点和中心像素点的灰度差值独立于中心像，则 F_L 又可通过式(2-5)来近似表示：

$$F_L = G(g_c) \cdot G(s_0, s_1, \dots, s_{P-1}) \quad (2-5)$$

上式中， $G(g_c)$ 仅反映了区域的灰度(亮度)信息，与纹理特征无关，则又有：

$$F_L = G(s_0, s_1, \dots, s_{P-1}) \quad (2-6)$$

为了得到该区域的十进制 LBP 特征值，可以为每一个邻域像素点分配一个权值，则该 LBP 特征值即可通过式(2-7)来计算:

$$LBP_{P,R} = \sum_{i=0}^{P-1} s_i \times 2^i \quad (2-7)$$

由以上分析可知，利用扩展的 LBP 算子求取一幅图像 LBP 特征的过程，就是利用该 LBP 算子作为模板，逐点扫描计算 LBP 特征值的过程:首先将 LBP 算子在图像上逐点扫描，并对当前中心点的像素灰度值和其邻域点像素灰度值的大小进行比较，如果邻域像素的灰度值大于等于当前中心点的像素灰度值，则将该邻域像素的灰度值置为 1，否则置为 0，然后根据顺时针方向读出每个邻域的二进制值，将该二进制值转换成十进制数，即得当前矩形块的 LBP 特征值。最后，对扫描区域的全部 LBP 特征值进行统计，以直方图的形式来描述该区域的纹理特征。

2.3 形状特征

2.3.1 Harr-Like 特征

Harr-Like 特征是由 Haar 小波演变而来的，Haar 小波的形式见图 2.4，而在灰度图像上，可以表现成图 2.5 的“Harr-Like 特征”形式。

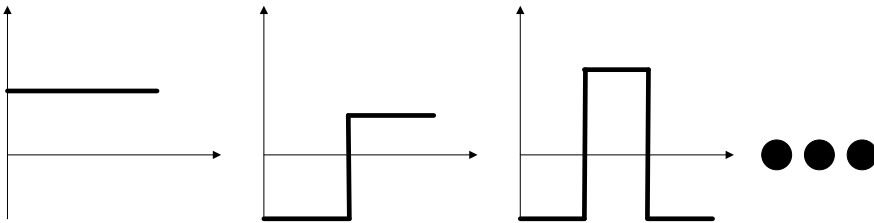


图 2.4 Haar 小波

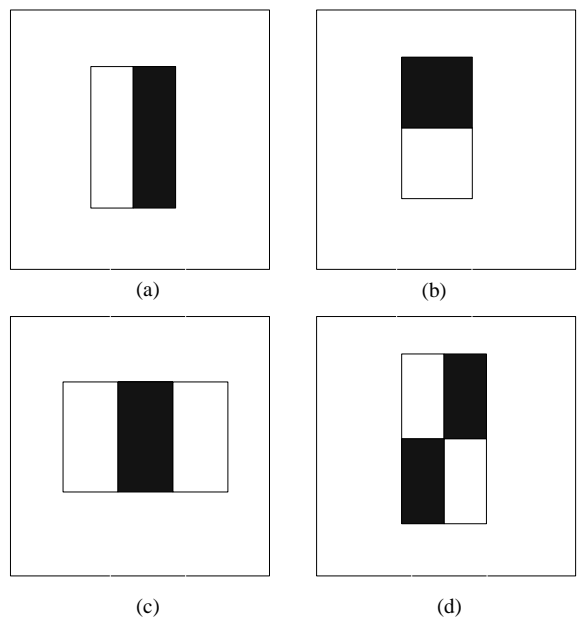


图 2.5 Haar-Like 特征

特征的值等于灰色矩形框中所有像素的颜色值和减去白色矩形框中的所有像素的颜色值。

(A)和(B)为两个矩形的特征,(C)为三个矩形的特征,(D)为四个矩形的特征。

对于由 2 个矩形所组成的特征的值，可以通过对两个矩形区域的所有像素分别求和然后再相减来实现。这些相邻的矩形区域都有相同的大小、形状、以及相同的排列方式。对于由 3 个矩形组成的特征的值是通过使用中间矩形的 2 倍减去两边矩形的和。最后一种是由 4 个矩形组成的本书可以按照对角线的两个矩形相加然后再求差来得到这种特征的值。

后来，Viola 等人又对上面基本的 Haar 特征进行了拓展，便有了现在的“拓展的 Haar-Like 特征”，见图 2.6。



图 2.6 拓展的 Haar-Like 特征

在给定的一个标准的 24*24 分辨率的图片中，那些矩形特征所组成的完备集中的元素的个数非常巨大，超过了 180000 个。

2.3.2 HOG 特征

HOG 是梯度方向直方图（Histograms of Oriented Gradients）的缩写，由 Dalal 和 Triggs 于 2005 年提出，并提出了基于 HOG 特征的人体目标检测算法。下面我们就来介绍一下 HOG 特征。

与 SIFT 特征类似，HOG 特征通过提取局部区域的边缘或梯度的分布，来表征局部区域内目标的边缘或梯度结构，进而表征目标的形状。由于是在局部区域统计求得，HOG 特征对小的形变和配准误差有较强的鲁棒性。

而对于 HOG 特征而言，Dala 是将 64×128 的训练样本按照 8×8 个像素的小块（cell）进行划分，这样就形成了 $8 \times 16 = 128$ 个 cell。然后再将每相邻的 4 个 cell（田字形结构的 4 个 cell）划分为 1 组（block），通过滑动 block 可以得到多组田字形局部区域特征，每个 block 一次滑动 8 个像素，由此对于一个 64×128 的训练样本便具有 $7 \times 15 = 105$ 个 block。首先，本书对于每个 cell 都按照 SIFT 的方法，将其中的所有像素的梯度方向进行投影，形成每个 cell 各自的梯度方向直方图。

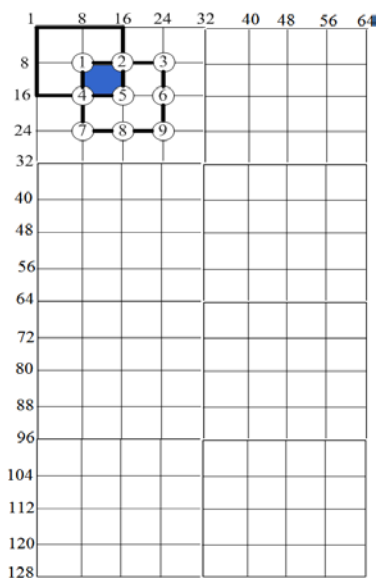


图 2.7 Dalal 的 HOG 特征块

但这里的方向 bins 的数量被设定为 9，而不是 SIFT 中的 8，即每 20 度一个 bin，0-180 与 180-360 的方向采用对等角相等的方法进行归类划分。然后，再将每个 block 中的 4 个 cell 的梯

度直方图的数据串联起来，由于每个 cell 的梯度直方图为一个 9 维的向量，所以每个 Block 便是一个 36 维的向量。再将所有的 block（对于 64x128 像素的样本共 105 个 block）依次串联起来，便形成了对每个训练图像的 $36 \times 105 = 3780$ 维的特征。

2.3.3 SIFT 特征

SIFT（Scale Invariance Feature Transform）特征不只具有尺度不变性，即使改变旋转角度，图像亮度或拍摄视角，仍然能够得到好的检测效果。

SIFT 特征的提取大致分为以下几个步骤：

1. 构建尺度空间，通过生成尺度空间来创建原始图像的多层表示以保证尺度不变性。

2. LoG（Laplacian of Guassian）近似寻找感兴趣特征点，使用 LoG 能够很好的找到感兴趣特征点（interesting point），但是需要大量的计算量，所以通过计算图像 DoG（Difference of Gaussian）的极值点来近似。

3. 除去不好的特征点，边界和低亮度区域是不好的特征点，除去它们以使得算法有效和鲁棒，可以通过 Harris Corner 检测算子实现。

4. 为特征点赋值一个方向，为每个特征点计算得到一个主方向，依照这个方向做进一步的计算，步骤有效的消除方向的影响，使 SIFT 特征具有旋转不变性。

5. 生成 SIFT 特征，通过对方向收集区域进行梯度方向统计得到 SIFT 特征的描述子。

尺度空间的构造是用第一幅图像逐渐生成一组模糊后的图像，然后将原始图像的尺寸缩小一半，再逐渐生成下一组模糊后的图像以此类推。然后对尺度空间的图像做 DoG 操作得到 DoG 图像，DoG 的操作方式见图 2.8。

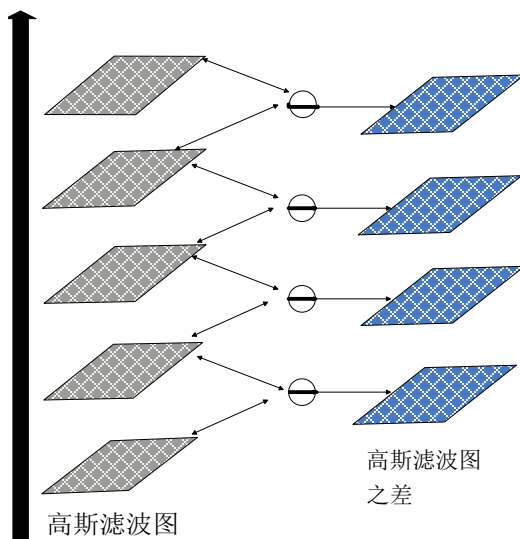


图 2.8 DoG 操作示意图

通过 DoG 图像可以找到好的特征点，分为两个子步骤：

1. 在 DoG 图像中找到极大或极小像素点。
2. 找到子像素级的极大极小值点。

第一步是粗糙地找到极大极小值像素点，这很简单，我们可以扫描每个像素并且检测所有的邻接像素点，邻接像素点不仅包括当前图像中的邻接像素，而且包括上一层和下一层图像中的邻接像素。在 DoG 图像中找极大极小像素点见图 2.9。

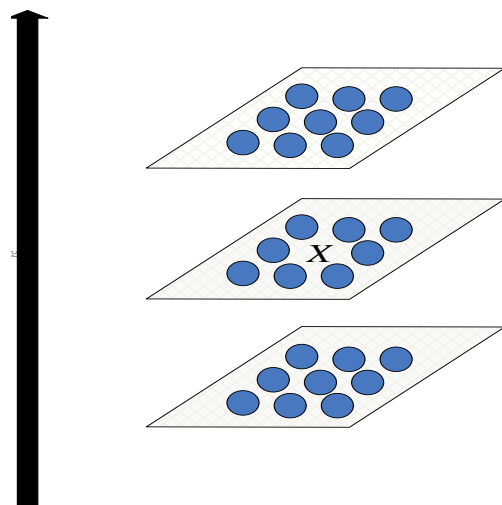


图 2.9 DoG 图像极大极小像素点示意图

X 标记当前像素点，绿色的圈标记邻接像素点，用这个方式，最多检测 26 个像素点。**X** 被标记为特征点，如果它是所有邻接像素点的最大值或最小值点。通常对于非极大或极小值点不需要遍历所有 26 个邻接像素点，少数的几个检测就足够抛弃它。注意最高层和最底层的尺度是不需要检测的，经过上一个操作，我们从每组图像中得到 4 幅 DoG 图像，只需要对中间的两幅 DoG 图像进行极大极小值像素点进行检测。这步结束后，所标记的点就是近似的极大极小值点，之所以说是“近似的”是因为极大极小值点都不会巧好在像素点的位置上，它一般位于像素的中间，但是我们无法直接访问像素间的数据，所以我们需要通过插值得到子像素的位置。红色的叉标记图像中的像素，但是真正的极值点是绿色的位于像素间的点。使用可用的像素数据，可以生成子像素的值，它通过近似的特征点附近图像的 Taylor 展开来计算。通过对 x 求偏导并将结果置为零，我们可以简单地计算出方程的极值，从而得到子像素特征点的位置，这些子像素点可以增加匹配和算法稳定性的机率。

若子像素点与近似特征点间的偏移量大于 0.5，则按照偏移近似特征点的方向相应改变（移动）特征点，然后再把该点当作近似特征点，重复该操作，直到子像素点与近似特征点间的偏移量小于等于 0.5。

分配特征点的方向：在每个特征点周围计算图像的梯度方向和大小，我们可以得到最显著的方向，并且将该方向赋给该特征点。后面的操作都相对该方向进行计算，确保了旋转不变性。

在特征点附近，创建一个方向收集区域来控制该特征点的影响范围，方向收集区域的大小依赖于它所在图像的尺度，尺度越大，收集区域越大。在方向收集区域中每个像素点的梯度大小和方向用公式 (2-8) 和 (2-9) 计算，从而得到另外两幅图，分别是梯度的幅值图 and 方向图。

$$m(x, y) = \sqrt{(I(x+1, y) - I(x, y))^2 + (I(x, y+1) - I(x, y))^2} \quad (2-8)$$

$$\phi(x, y) = \arctan\left((I(x, y+1) - I(x, y)) / (I(x+1, y) - I(x, y))\right) \quad (2-9)$$

我们用一个直方图来统计方向收集区域中像素的平均方向，在直方图中，将 360 度的方向分成 36 个 bins，每个 bin 包含 10 度。假设方向收集区域中某个像素点的梯度方向是 18.75 度，把它放入 10-19 度 bin 中，并且加入到 bin 中的量与该像素点的梯度大小成正比。

一旦对方向收集区域中的每个像素点都执行了这个操作后，直方图在某个 bin 上出现最高

峰值（见图 2.11）。图中，直方图的最高峰值在 20-29 度，所以特征点的方向被赋值为 3（第三个 bin）。并且所有大于最高峰 80% 的峰值也被转化为一个新的特征点，这个新的特征点和原来的特征点一样拥有相同的位置和尺度，但是新特征点的方向是另一个峰值。所以方向将一个特征点分为多个特征点。

生成 SIFT 特征描述子：这是 SIFT 算法的最后一步，现在我们已经得到了拥有尺度不变性和旋转不变性的特征点，接下来要为每个特征点创建一个唯一标识它的“指纹”，SIFT 算法作者将它称为 SIFT 描述子(descriptor)。所生成的 SIFT 描述子既要能让相同场景中图像的特征点能够正确匹配，而且还要让不同场景中图像的特征点能够正确区分。

为了得到这样的 SIFT 描述子，我们将特征点周围 16×16 的窗口分解为 16 个 4×4 的子窗口，分解的过程见图 2.10。在每个 4×4 的子窗口中，计算出梯度的大小和方向，并用一个 8 个 bin 的直方图来统计子窗口的平均方向。

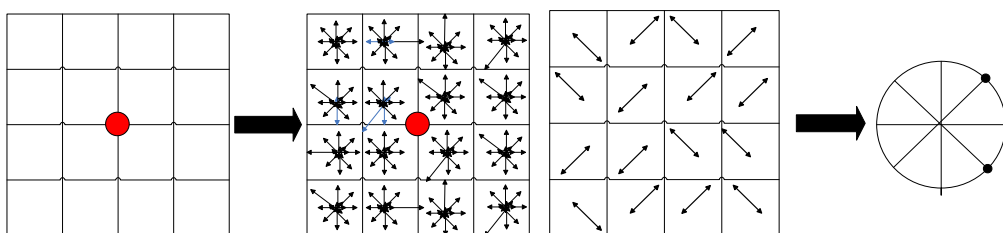


图 2.10 特征点周围的窗口分解与为每个子窗口创建的 8 bin 的直方图

梯度方向在 0-44 度范围的像素点被放到第一个 bin 中，45-89 度范围的像素点被放到下一个 bin 中，依此类推。同样加入到 bin 中的量依赖于该像素点梯度的大小。与之前不同的是，加入的量不仅与像素点的梯度大小相关，而且还依赖离特征点的距离，这样远离特征点的像素点会加入较少的量到直方图中。这通过一个高斯加权函数来实现，这个函数生成一个加权值（像一个二维的钟形曲线），用它乘以 16×16 的窗口中每个像素点的梯度大小，得到加权后的梯度大小，距离特征点越远，要加入直方图的像素点的梯度大小越小。

这样每个 4×4 的子窗口都对应一个 8 bin 的直方图，且直方图中加入的值是像素的用高斯加权后的梯度大小，而特征点周围 16×16 的窗口中包含 16 个 4×4 的子窗口，共有 $16 \times 8 = 128$ 个数，然后将这 128 个数组成的向量进行单位化，单位化后的 128 维向量就是 SIFT 的描述子。

2.3.4 MSO 特征

在固定尺度上的单个特征的计算中，每个特征区域都是一个 $m*m$ 的正方形，对于这个正方形特征的特征值的计算，我们可以分为两步：

1. 计算正方形区域的毛梯度方向（因为我们这里是在大尺度上求解梯度，相对于原始定义的梯度而言更加粗糙，因此这里将其命名为“毛梯度”）。

2. 将梯度方向映射成固定的特征值编码。

首先，我们将正方形分为左右两个等份，计算左右两个小区域的所有像素的灰度值之和，分别记为 SUM_R 和 SUM_L ，得出水平灰度差 dx ；再将正方形分为上下两等份，计算出竖直灰度差 dy 。这样这个正方形在形状上便体现了田字形的特征。

区域灰度值的累加和 SUM_A 的计算公式如下：

$$SUM_A = \sum_{x1 < x < x2, y1 < y < y2} PixelValue(x, y) \quad (2-10)$$

其中， $PixelValue(x, y)$ 表示坐标为 (x, y) 像素点的灰度值，这是一种比较直观的求法，然而我们可以参考积分图的求解，这会在很大程度上降低整幅图像的运算复杂度。这样便可以求出正方形特征水平灰度差 dx 和竖直灰度差 dy 。

$$\begin{aligned} dx &= SUM_R - SUM_L \\ dy &= SUM_D - SUM_U \end{aligned} \quad (2-11)$$

其中， SUM_R 、 SUM_L 、 SUM_D 和 SUM_U 分别表示为正方形特征区域的右半部、左半部、下半部和上半部的灰度值的累加。

接着，我们先将 $0 \sim 360$ 度等分为 18 个区间，并保证对顶角区域属于同一区间。这样我们就可以根据所求出来的 dx 和 dy 计算出这个矩形区域内的“大致”梯度方向 Ori_Rect ：

$$Ori_Rect = F(\arctan(dy / dx)) \quad (2-12)$$

其中，函数 F 为将角度转换为数值 $0, 1, 2, \dots, 8$ 的函数，如：当 $\arctan(dy / dx)$ 的值在 $1 \sim 20$ 或 $181 \sim 200$ 之间的时候，则 Ori_Rect 的值为 0；当 $\arctan(dy / dx)$ 的值在 $21 \sim 40$ 或 $201 \sim 220$ 之间的时候，则 Ori_Rect 的值为 1。

2.4 特征融合

特征融合属于信息融合的范畴，所谓信息融合是指对来自多个传感器的多源信息或者同一传感器的不同特征信息进行多级别，多方面，多层次的处理和综合，从而获得更为丰富，精确，可靠的有用信息，信息融合系统可以最大限度地获得对目标和场景的完整信息描述。

本节中提到的特征融合是指视觉目标特征的融合，待融合的视觉目标特征取自单一视觉传感器。特征融合是从图像的原始信息中提取各种特征信息进行综合分析及融合处理，不仅增加了特征信息的检测精度，且提高了目标检测的性能。

特征融合可分为状态信息融合和特性融合。目标状态信息融合主要应用于多传感器目标跟踪领域，用于实现参数相关的状态矢量估计。目标特性融合是将同一目标的不同特征经过某些变换综合在一起得到一种复合特征，即本节中的特征融合。目标特性融合过程用到了模式识别中的一些相关技术，且融合前需要对不同特征进行配准，对特征矢量进行分类和综合。特征融合中的配准要求不像像素级融合那么高，可通过特征的集合变换得以实现。特征融合过程示意图如图 2.11 所示。

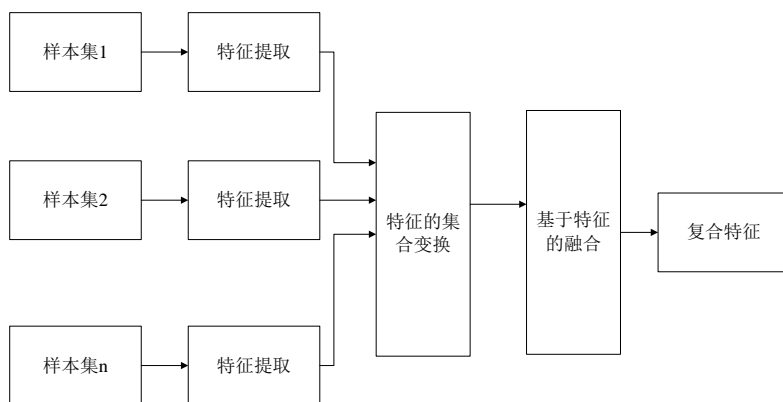


图 2.11 特征融合过程示意图

2.5 特征降维及选择

特征降维 (Feature Dimension Reduction) 是一个从初始高维特征集合中选出低维特征集合, 以便根据一定的评估准则最优化缩小特征空间的过程, 通常作为机器学习的预处理步骤。特征降维自 20 世纪 70 年代以来就获得了广泛的研究。近几年以来, 在许多应用 (如基因染色体组工程、文本分类、图像检索、消费者关系管理) 中, 数据的实例数目和特征数目都急剧增加, 这种数据的海量性使得大量机器学习算法在可测量性和学习性能方面产生严重问题。例如, 高维数据即具有成百上千特征的数据集, 会包含大量的无关信息和冗余信息, 这些信息可能极大地降低学习算法的性能。因此, 当面临高维数据时, 特征降维对于机器学习任务显得十分必要。大量研究实践证明, 特征降维能够有效地消除无关和冗余特征, 提高挖掘任务的效率, 改善预测精确性等学习性能, 增强学习结果的易理解性。然而, 数据在数量和维度上的剧增趋势也对特征降维算法提出了更加严峻的挑战。

2.5.1 特征降维方法综述

通常, 高维特征集合存在以下几方面问题: 大量的特征; 许多与给定任务无关的特征, 即存在许多与类别仅有微弱相关度的特征; 许多对于给定任务冗余的特征, 如特征相互之间存在强烈的相关度; 噪声数据。特征降维是一种降低特征维度从而提高给定任务效率的方法, 常见的有一些特征抽取的降维方式。

特征抽取也被称为特征重参数化。由于自然语言中存在大量的多义词、同义词现象, 特征集无法生成一个最优的特征空间对数据内容进行描述。特征抽取通过将原始特征空间进行变换, 重新生成一个维数更小、各维之间更独立的特征空间。可以按照主成分分析、独立成分分析等方法对特征抽取。

1. 主成分分析

主成分分析法是研究如何以最少的信息丢失将众多原有变量浓缩成少数几个因子 (主成分), 使这些主成分在一定的程度上复现原有变量所携带的信息的多元统计分析方法。具有以下四个特点: (1) 主成分个数远远少于原有变量的个数; (2) 主成分能够反映原有变量的绝大部分信息; (3) 主成分之间应该互不相关; (4) 主成分具有命名解释性。

基本原理:

设 X_1, X_2, \dots, X_p 为原始变量, F_1, F_2, \dots, F_m 为 m 个主成分因子 $F_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$, 其

使方差越大, 表示 F_1 包含的信息越多, 故称 F_1 为第一主成分。

$$\begin{cases} F_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ F_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ \dots\dots\dots \\ F_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{mp}X_p \end{cases} \quad (2-13)$$

F_i 与 F_j 互不相关, $Cov(F_i, F_j) = 0$, F_1 是 X_1, X_2, \dots, X_p 的一切线性组合中方差最大的, 即 F_m 是与 F_1, F_2, \dots, F_{m-1} 都不相关的 X_1, X_2, \dots, X_p 的所有线性组合中方差最大者。 F_1, F_2, \dots, F_m ($m \leq p$) 为构造的新变量指标, 即原变量指标的第一、第二、...、第 m 个主成分。

2. 独立成分分析

独立成分分析的目的是将观察到的数据进行某种线性分解, 使其分解成统计独立的成分。最早提出独立成分分析概念的 Jutten, 他们给出的一种简单描述是, 独立成分分析是从线性混合信号里恢复出一些基本的源信号的方法。

为了给出严格定义, 这里需要使用一个隐藏的统计变量模型 $x = As$, 其中的统计模型称为独立成分分析。它表示被观察到的数据是如何由独立成分混合而产生的, 独立成分是隐藏的变量, 意味着它不能被观察到, 而且混合矩阵也被假设为未知的, 所有能观察到的仅仅是随机向量 x , 必须顾及出 A 和 s , 而且必须在尽量少得假设条件下完成它。因此, 独立成分可以由 $s = Wx$ 计算出, 其中 W 是 A 得逆。

主成分分析和独立成分分析都是多用途的统计方法, 它们在语音信号分离、生物医学信号处理、金融数据分析、图像消噪、远程通信、人脸识别等方面的应用成果充分显示了重要的应用价值。

2.5.2 特征选择方法综述

随着文本分类、组合化学和生物信息学等领域数据大规模的增长, 数据中包含成百上千维的特征, 大部分数据间存在冗余性或者不相关性。由于数据的维度过高, 导致所需要的训练样本数目呈指数型增长, 即产生模式识别领域中谈及的“维数灾难”问题。因此对数据进行降维是一个非常迫切和具有挑战性的研究课题。

数据 (简称为特征) 降维常用的两类方法是特征变换 (特征提取) 和特征选择。特征变换是指将原有特征通过某种线性变换 (主成分分析, PCA)^[1] 或者非线性的变换 (低秩矩阵)^[2], 得到一种低维的新特征。特征的可理解性比较重要, 但特征变换使得特征的理解性变差, 因为

即使很简单的线性组合也能构造出新特征。相反地，特征选择是从原始特征集中，按着某种标准选择出特征子集，并未改变原始特征的信息，增加了特征的可理解性。

2.5.2.1 特征选择的定义

特征选择最早的研究始于上个世纪 60 年代，当时的范围主要涉及统计学和信号处理等领域，所面临的特征维数并不高，维数灾难现象不是很明显。随着 90 年代大规模数据的出现和机器学习领域的发展，特征选择问题受到学者们的广泛关注。

目前，特征选择还没有比较统一的数学定义，因为特征选择需要和具体的应用背景和研究问题相结合。针对不同的问题如分类^[3, 4-7]、回归^[8-9, 10]等，特征选择可笼统地定义为：在满足一定的准则下，寻找原始特征集中“最好”的某一特征或者某一特征子集。从其定义上，可以看出特征选择的本质是对原始特征进行排序或者加权，最重要的特征排在前几位（权重很大）。特征选择的目的是主要有三个方面：一是改进分类器的预测性能；二是对特征的生成过程有一个更好的认识；三是减少存储空间的使用和计算费用，提高预测速度。

特征选择算法的实现过程一般需要四个要素^[11-13]，它们分别是：原始样本集、搜索策略、评价函数和终止条件。特征选择算法首先需要输入原始样本，特征维数等信息，然后根据某种搜索策略，寻找候选特征子集，并判断该特征子集是否能使评价函数的性能有所改进，最后与终止条件进行比较，判断是否获得最优解，从而执行下一步的方案。特征选择算法的流程图如图 2.12 所示。

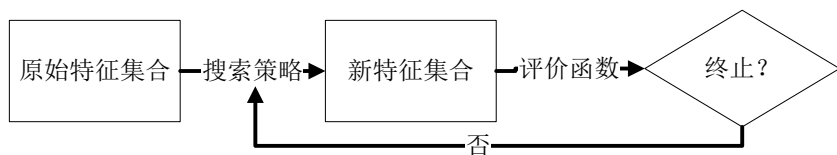


图 2.12 特征选择流程图

下面重点介绍下特征选择算法的搜索策略和评价函数这两个基本要素。

1. 搜索策略

根据不同的搜索方向，可以得到不同的搜索策略。搜索方向主要有^[14]：前向搜索、后向搜索、双向搜索和随机搜索等。前向搜索是指在当前特征子集的基础上，增加一个或多个新的特征；后向搜索是指删除当前特征集中的一个或多个特征；双向搜索是指先删除一些特征，再增加若干个新的特征；随机搜索是指依据概率分布，随机的选择一个或者多个特征。根据上述的几种搜索方向，特征选择的搜索策略大致分为以下三大类^[15-18]：

(1) 全局最优搜索策略：它是指从全局的角度，选择最优的特征子集。典型的代表方法有枚举法和分支界定法，其中枚举法是从当前候选子集出发，采用逐一举例法访问搜索空间中的每个状态，以找到最优子集。理论上，给定一个原始特征集和评价函数，枚举法可以找到最优的特征子集，然而这种方式计算费用比较高。另一种典型的全局搜索法是分支界定法^[19]，它利用评价函数的单调性，以树状结构的形式对搜索空间进行划分。树中的每个节点对应一个候选子集，若某节点的评价函数值小于当前候选集的值，则将该节点及其子节点集删除，再考虑其他节点，这样迭代的求解直至获得最优解为止。这种方法比枚举方法效率要高一些。

(2) 序列搜索策略：顺序地向当前特征子集中添加或者删除某些特征，这样迭代地搜索满足终止条件的特征子集。该策略的典型代表方法有：顺序前向搜索，顺序后向搜索等。某些特征一旦被添加到当前特征子集，或者从子集中被删除后，不再考虑对该特征进行任何操作，因此序列搜索策略容易陷入局部极值。为解决此问题，提出了双向的增 m 减 n 搜索法，即搜索方向不再是单向的，可以根据评价函数灵活地增减，然而增 m 减 n 搜索法的缺点在于 m 和 n 的大小难以确定。为克服此缺点，Pudil 等提出了顺序浮动搜索算法^[20]。该算法可以浮动地改变 m 和 n 的值，减少了不必要的回溯并在需要时增加回溯的深度。后来，Somo1 等人提出了自适应浮动搜索算法^[21]，根据当前特征子集大小来控制搜索空间的大小，自适应的浮动搜索减小了算法陷入局部最优的可能性。

(3) 启发式搜索策略：指在搜索特征子集时，使用启发式算法选择特征，并以一定的概率选择特征进入候选特征子集。典型代表算法有遗传搜索^[22]、模拟退火搜索等。遗传搜索在特征选择算法中应用很广，因为它速度比较快，不易陷入局部最优解等^[23]，但是理论上，它不能保证找到最优解。模拟退火搜索算法因为它的计算量较大，初始温度以及迭代次数等参数的值难以确定等因素，使得它的应用不是很多。从文献^[24-25]的实验结果可见，在上述所提搜索算法中，自适应浮动搜索和遗传搜索是性能较好的两种搜索算法。

2. 评价函数

评价函数，是指能够区分特征“好”、“坏”的度量函数，依照评价函数的某种优化准则在原始的特征集中挑选特征。因此，评价函数决定保留哪些特征，抛弃哪些特征。常见的优化准则如下：

- (1) 从原始特征集中找到某个特征子集，使得评价函数的值达到最优；
- (2) 在评价函数值大于某一个给定的阈值条件下，寻找一个最小特征子集；
- (3) 从原始特征集中寻找一个最小特征子集，并且同时使得评价函数的值尽可能最优。

上述三种准则体现了不同的优化目的，总体说来评价函数的定义是特征选择的重要因素。下面简单介绍下几种常见的评价函数。

(1) 距离度量函数

距离度量函数利用统计模式识别中一些常见的距离等度量方式作为评判标准。从是否引入随机因素的角度，距离度量函数大体可分为确定距离和概率距离。确定距离通常是指欧式空间中的距离，如欧氏距离和马氏距离等。概率距离是指引入概率形式来衡量距离等度量方式。

表 2.1 距离度量函数

距离类别	距离名称	公式
确定距离	欧氏距离	$d(x_i, x_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$
	马氏距离	$d(x_i, x_j) = (x_i - x_j)^T C^{-1} (x_i - x_j)$
	Chebychev 距离	$d(x_i, x_j) = x_{ik} - x_{jk} $
概率距离	Bhattacharyya 距离	$d = -\ln \int_x p(x Y_i)^{1-s} p(x Y_j)^s dx$
	Kolmogorov 距离	$d = \int_x (p(x Y_i) - p(x Y_j)) dx$

(2) 一致性度量函数

一致性是指：给定两个样本，若它们的特征值均相同，且其类别标号相同，则称它们是一致的，否则是不一致的^[25]。样本集合的不一致性是指该数据集中不一致的样本数与样本总数之间的比例。一致性度量利用样本集合的不一致性反映特征的重要性程度^[26]。如果去除某一特征后，数据集的不一致性明显增大，那么就认为该特征很重要；否则认为它不重要。一致性度量函数的优点是能获得一个较小的特征子集，但它对噪声数据敏感，且只适合离散特征。

(3) 相关性度量函数

相关性度量^[7, 27]利用类别与特征之间的统计相关性度量特征的重要程度，即如果已知两个变量是统计相关的，那么就可以利用其中一个变量的值来估计预测另一个变量的值。目前，常用的统计相关系数，如 t 检验、Pearson 相关系数、Fisher 分数^[28]等用来表达特征相对于类别可分离性的重要性程度。

表 2.2 相关性度量函数

相关度量函数	公式
Pearson 相关系数	$r = \text{cov}(X_i, Y) / \sqrt{\text{var}(X_i) \text{var}(Y)}$
Fisher 分数	$Fisher(x_k) = \sum_i n_i (m_k^i - m_k)^2 / \sum_i n_i (\sigma_k^i)^2$

(4) 信息度量函数

信息度量^[29-32]主要利用信息熵、信息增益、互信息等手段量化特征的不确定性程度，以判

定其包含的信息含量。信息度量的优点在于它是一种无参数的、非线性度量，且它不需要预先知道样本的分布。由于上述优点，信息度量函数在特征选择算法中得到广泛关注。表 2.3 给出一些常见的信息度量函数的表示形式。

表 2.3 信息度量函数

信息度量名称	公式
信息熵	$H(x_i) = \sum_k p(x_{ik}) \log_2 p(x_{ik})$
互信息	$MI(x_i, y) = \sum_{x_i} \sum_y p(x_i, y) \log_2 [p(x_i, y) / p(x_i)p(y)]$
信息增益	$IG(X) = -\sum_y p(y) \log_2 p(y) - \sum_x [p(x) \sum_y p(y x) \log_2 p(y x)]$

(5) 误差度量函数

特征选择的目的是期望选择后的特征能够提高学习算法的性能，学习算法的性能一般由误差函数来体现。针对分类问题而言，分类误差是衡量特征选择算法优劣的标准之一。目前，有许多特征选择算法直接采用分类误差衡量特征的重要性，如 Stoppiglia^[33]等人提出使用均方误差准则作为分类误差函数，通过前向搜索逐步添加能够使误差函数改变最多的特征。Huang等^[34]使用遗传搜索与互信息熵结合的封装式算法获取特征子集，并能明显提高分类模型的性能。Neumann等^[36]提出新颖的特征选择方法，其融合核函数约束与 l_p 范数正则项构建分类目标函数，取得较好的分类性能。Schapire等人^[36]提出 Boosting 算法进行特征选择方法，其采用重采样前向搜索策略和误差最小为评价函数。

2.5.2.2 特征选择的分类

根据特征选择与学习算法的结合方式，特征选择算法大体上可以分为三类：过滤式(Filter)、封装式(Wrapper)、嵌入式(Embedded)算法。

1. 过滤式特征选择

过滤式特征选择[30]的评估标准独立于学习算法，直接对数据集本身进行处理，未考虑和学习算法有关的信息。通常是选择与评价函数相关度大的特征或者特征子集，一般认为相关度较大的特征或者特征子集会以高概率提高学习算法的精度。过滤式特征选择的评估方法很多，如距离度量、信息增益、相关性以及不一致性等等(这即是前面评价函数中所介绍的度量准则)。过滤式特征选择因为运行效率较高而适用于大规模数据集，但 Kohavi^[20]等人指出因为其脱离了学习算法的任何信息，因此找到的特征不一定能改进学习算法的性能。

2. 封装式特征选择

封装式特征选择由 John 等人提出^[21]。该算法的核心思想是：如果特征选择不与学习算法相结合（如 Filter 特征选择），那么在特征选择之后，仍然会产生和学习算法无关的冗余信息。最重要的是，不同学习算法偏好不同的特征子集，因此他们认为特征选择过程不能脱离学习算法，应该以学习算法性能作为特征评价的标准。他们把学习算法看成是一个黑盒子，特征选择的标准是算法的分类性能（针对分类问题而言），一般通过交叉验证的方法看分类器的性能。然而，封装式特征算法并未涉及到学习算法的具体细节，比如说目标函数的形式等，它的处理方式只看当前候选特征子集对算法性能的改进，没有对学习算法有任何限制。因此，封装式特征方法比较简单、通用。

3. 嵌入式特征选择

嵌入式特征选择是指：特征选择过程嵌入到学习算法里，作为学习算法的某个组成部分。嵌入式特征选择算法很早就有，比如在决策树的训练中，如 Breiman 的 CART 算法和 Quinlan 提出的决策树与 bagging 和 boosting 相结合的算法等。这些算法在每一个节点选择分类能力最强的特征，然后基于选中的特征进行子空间分割，继续此过程，直到满足终止条件，可见决策树生成的过程也就是特征选择的过程。后来，一些嵌入式特征选择方法通过估计目标函数值的改变，指导变量在特征空间的移动。Rakotomamonjy 等人对线性 SVM 的目标函数，采用有限差分方式观察目标函数值的改变，从而指导向量的特征选择。Guyon 等人提出使用保持对偶变量值不变的情况下，估计目标函数有限差分的变化，实现对核函数 SVM 的特征进行选择。

2.5.3 稀疏特征选择

范数是某种距离度量。1 范数是 l_p 范数中的一种，常见的范数还有 2 范数，0 范数等。某一向量的 1 范数是指向量中每一分量的绝对值求和。对于 1 范数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，但是其一直未受到重视。在实际应用中 2 范数一直都受到学者的青睐，原因在于 1 范数的连续但非光滑性质使得其求解算法比较复杂。直到上节所提到的稀疏表示问题的产生以及压缩感知理论在信号处理领域中的出现，1 范数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 0 范数去重构信号，但由于 0 范数的求解是 NP-hard 的问题，所以该理论使用 1 范数来代替 0 范数，而且理论上已经证明 1 范数可以作为 0 范数的一个近似估计。

此外，在模式识别领域，很多学者应用 1 范数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 1 范数在模式识别领域应用的文献，要数 John Wright 和 Yi Ma 等人提出的 SRC (Sparse Representation Classification) 算法。该算法使用 1 范数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 1 范数最小化。SRC 算法使用 1 范数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目标的识别。

$$\begin{aligned} \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (2-14)$$

假设有 k 类不同样本，同一个目标在不同环境、不同光照下的图像构成一个集合 A_i ，所有目标构成训练样本集合 $A = [A_1, A_2 \cdots A_k] \in R^{m \times n}$ ，上述优化模型可等价写成：

$$\begin{aligned} \tilde{x} = \arg \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (2-15)$$

考虑到噪声与误差的影响，那么 (2-15) 式重写成：

$$\begin{aligned} \tilde{x} = \arg \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & \|Ax - y\|_2 \leq \varepsilon \end{aligned} \quad (2-16)$$

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

$$\text{class}(y) = \arg \min_i r_i(y) \quad (2-17)$$

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。受到上述研究的启发，本书借鉴 SVM、Adaboost 和凹函数支持向量机等分类算法，研究基于了 1 范数最小化的分类问题，结合 1 范数的稀疏性设计检测模型，得到权重向量以及一个人体训练样本的加权特征向量（即权重与人体样本向量的相应维度进行乘积，得到一个同样维度的向量，称为加权特征向量）。当设定一个阈值时，加权特征向量中多数都是小于这个给定的阈值，其余的少数是主要的成份，这一稀疏表示过程可以视为特征选择的过程。实际上，稀疏表示的特征选择的作用是去除冗余信息，尽可能地减少冗余信息将有用的信息淹没的可能性，即可以消除遮挡和噪声对目标的影响。

2.6 参考文献

- [1] Jolliffe I T. Principal Component Analysis, Sprinberg Publishers, 2002
- [2] Candes E, Li X D, Ma Y, Wright J. Robust principal component analysis. Journal of the ACM, 2011:3.
- [3] Kira K, Rendell L. A practical approach to feature selection. The 9th International Conference on Machine Learning, 1992: 249~256.
- [4] Mitra P, Murthy C A, Pal S K. Unsupervised feature selection using feature similarity, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(3): 301~312,.
- [5] Hall M A. Correlation-based feature subset selection for machine learning. Hamilton, NewZealand University of Waikato, 1999.
- [6] Wu S, Flach P A. Feature selection with labelled and unlabelled data. ECML/PKDD'02 workshop on Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning, 2002, 156~167,.
- [7] Dy J G, Brodley C E. Feature selection for unsupervised learning. Journal of Machine Learning Research, 2004, 5: 845~889,.
- [8] Kononenko I. Estimation attributes: analysis and extensions of RELIEF. The European Conference on Machine Learning, 1994, 171~182.
- [9] Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learning, 2003, 53:23~69.
- [10] Tibshirani R. Regression selection and shrinkage via the lasso. Technical Report, Stanford University.
- [11] 刘华文. 基于信息熵的特征选择算法研究. 吉林大学博士论文. 2010.
- [12] 张丽新. 高维数据的特征选择及基于特征选择的集成学习研究. 清华大学博士论文. 2004.

- [13] 苏映雪. 特征选择算法研究. 国防大学硕士论文. 2006.
- [14] Molina L C, Belanche L, Nebot A. Feature selection algorithms: a survey and experimental evaluation number. Barcelona, 2002.
- [15] Liu H, Motoda H. Feature selection for knowledge discovery and data mining. Boston, Kluwer Academic Publishers, 1998.
- [16] Liu H, Yu L. Toward, integrating eeature selection algorithms for classification and clustering. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(4):491~502.
- [17] Somol P, Novovicova J, Pudil P. Notes on the evolution of feature selection methodology. Kybernetika, 2007, 43(5): 713~730.
- [18] Somol P, Pudil P, Kittler J. Fast branch and bound algorithms for optimal feature selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(7):900~912.
- [19] Kohavi R, John G H. Wrappers for feature subset selection. Artificial Intelligence Journal, Special Issue on relevance, 1997, 97(1-2): 273~324.
- [20] John G, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem. The Eleventh International Conference on Machine Learning, 1994:121~129.
- [21] Casillas J, Cordon O. Genetic feature selection in a fuzzy rule-based classification system learning process for high dimensional problems. Information Sciences, 2001, 136(1-4): 135~157.
- [22] Yang J, Vasant H. Feature subset selection using a genetic algorithm. IEEE Intelligent Systems, 1998, 13:44~49.
- [23] Jain A K, Zongker D. Feature selection: evaluation, application, and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(2):153~158.
- [24] Kudo M, Jack S. Comparison of algorithms that select features for pattern classifiers. Pattern Recoginiton, 2000, 33:25~41.
- [25] Dash M, Liu H. Consistency-based search in feature selection. Artificial Intelligence, 2003, 151(1-2):155~176.

- [26] Arauzo-Azofra A, Benitez J M, Castro J L. Consistency measures for feature selection. *Journal of Intelligent Information System*, 2008, 30: 273~292.
- [27] Devijver P A, Kittler J. *Pattern recognition—a statistical approach*. London, Prentice Hall, 1992.
- [28] Hall M A. *Correlation-based feature subset selection for machine learning*. Hamilton, NewZealand University of Waikato, 1999.
- [29] Zhang D, Chen S, Zhou Z-H. Constraint score: a new filter method for feature selection with pairwise constraints. *Pattern Recognition*, 2008, 41(5): 1440~1451.
- [30] Forman G. An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, 2003, 3: 1289~1305.
- [31] Hall M A, Holmes G. Benchmarking, attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering*, 2003, 15(3): 1041~4347.
- [32] Yang Y, Pedersen J O. A comparative study on feature selection in text categorization. *The 14th International Conference on Machine Learning*, 1997: 412-420.
- [33] Stoppiglia H, Dreyfus G, Dubois R, Oussar Y. Ranking a random feature for variable and feature selection. *Journal of Machine Learning Research*, 2003, 3: 1399~1414.
- [34] Huang J, Cai Y, Xu X. A hybrid genetic algorithm for feature selection wrapper based on mutual information. *Pattern Recognition Letters*, 2007, 28: 1825~1844.
- [35] Neumann J, Schnorr C, Steidl G.. Combined SVM-based feature selection and classification. *Machine Learning*, 2005, 61: 129~150.
- [36] Schapire R E, Freund Y, Bartlett Y. Boosting the margin: a new explanation for the effectiveness of voting methods. *Annals of Statistics*, 1998, 26(5):1651~1686.

第3章 目标检测方法

3.1 运动目标检测

3.1.1 基本概念

运动目标检测是指在序列图像中检测出变化区域并将运动目标从背景图像中提取出来。通常情况下，目标分类、跟踪和行为理解等后处理过程仅仅考虑图像中对应于运动目标的像素区域，因此运动目标的正确检测与分割对于后期处理非常重要。然而，由于场景的动态变化，如天气、光照、阴影及杂乱背景干扰等的影响，使得运动目标的检测与分割变得相当困难。根据摄像头是否保持静止，运动检测分为静态背景和运动背景两类。运动目标检测一直是视频序列分析的重点和难点，而大多数视频监控系统是摄像头固定的，因此静态背景下的基于运动信息的运动目标检测算法受到广泛关注，常用的方法有帧差法、光流法、背景减除法等。首先介绍三个基本概念。

1. 时间序列图像

为了观察物体的运动，人们将摄像机放置在某一位置上，对运动物体进行观测，所拍摄到的一系列图像称为时间序列图像或者运动图像。

2. 运动物体特征

任何一个运动物体都有其自身的特征，例如一些尖锐点，边缘直线，边缘曲线等，它们分别称为特征点、特征直线特征曲线等。物体在空间运动时，只要是在观察者的可视范围内，物体上的特征就均可在视频图像上反映出来。也就是说，可以通过运动物体的特征来观察分析物体的运动。

3. 投影图像

将某一时刻拍摄到的图像称为投影图像，投影图像可以理解为某一时刻运动物体在图像平面上的投影，显然，视频序列就是一系列的投影图像。在不同时刻拍摄到的投影图像上，运动

物体的投影坐标是不同的。运动分析与估计的研究内容就是测量计算投影坐标在图像平面上的变化，并用来分析运动物体的结构及估计物体的运动参数。

3.1.2 帧差法

帧差法，也称序列差分法，相邻帧差法，是最为常用的运动目标检测和分割方法之一。在摄像机视场中，相邻帧之间背景区域大致不变，可以相互消除。当出现运动目标时，两帧在运动目标出现区域会有较明显的差别，在图像序列相邻两帧或三帧间采用基于像素的时间差分通过闭值化来提取出图像中的运动区域。首先，将相邻帧图像对应像素值相减得到差分图像，然后对差分图像二值化，在环境亮度变化不大的情况下，如果对应像素值变化小于事先确定的阈值时，可以认为此处为背景像素，如果图像区域的像素值变化很大，可以认为这是由于图像中运动物体引起的，将这些区域标记为前景像素，利用标记的像素区域可以确定运动目标在图像中的位置。

帧差法的形式化描述如下，只需用当前帧 $I_n(x, y)$ 的前一帧或前 k 帧 $I_{n-k}(x, y)$ 替换背景图像即可，如公式 (3-1) (3-2) 所示。

$$D_n(x, y) = |I_n(x, y) - I_{n-k}(x, y)| \quad (3-1)$$

$$BW_n(x, y) = \begin{cases} 1, & \text{if } (D_n(x, y) \geq T) \\ 0, & \text{if } (D_n(x, y) < T) \end{cases} \quad (3-2)$$

这里 $k \in (1, 2, 3, \dots)$ 是预先设定的帧间间隔，当序列图像中目标运动缓慢时可以设置间隔 k 较大，一般设为 1 时，即为严格意义上的相邻帧间差分。按照帧差法检测原理，在做差的两帧图像的背景区域，背景会相互消除，而在目标存在的区域，实际进行了两种情况的差分：目标与背景的差分和目标与目标不同位置的差分，原理的示意图见图 3.1。

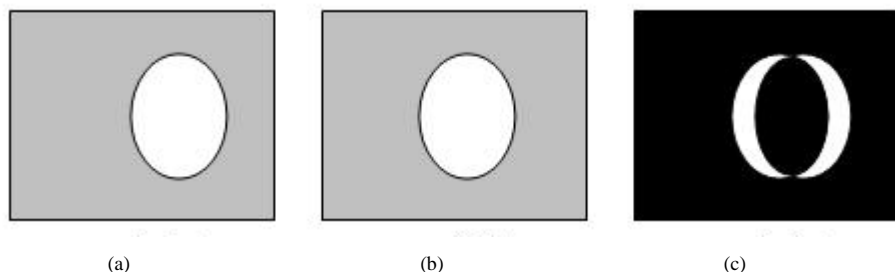


图 3.1 帧间差分原理示意图

可见，帧差法往往不能检测到对象的完整区域，而且区域范围往往大于目标自身，是两帧图像中目标区域的并集。当目标在两帧中运动非常缓慢，甚至可以看成静止时，两帧中目标位置几乎相同，帧差法将得不到目标，此时需要将帧间间隔 k 设置较大；当目标在两帧中运动较快时，两帧中目标重叠区域很小，算法得到的目标区域将很大，甚至得到两个相互分离的目标，这种情况需要缩小 k 的值；一般情况下，目标速度适中时，算法得到的是目标的边缘信息。

帧差法使用时，往往与背景差分相结合或使用三帧差分来获得好的效果，在此不再累述。

帧差法的优点在于相邻两帧间的时间间隔非常短，所以，背景部分往往变化不大，差分运算时能相互消去，从而能避免背景的某些变化而引起的误差，使算法对场景的光线不太敏感，适应光照等变化较频繁的环境因素，具有较好的稳定性。用前一帧图像作为当前帧的背景模型具有较好的实时性，其背景不积累，且更新速度快、算法简单、计算量小。

算法的不足在于对环境噪声较为敏感，阈值的选择相当关键，选择过低不足以抑制图像中的噪声，过高则忽略了图像中有用的变化。对于比较大的、颜色一致的运动目标，有可能在目标内部产生空洞，无法完整地提取运动目标。

3.1.3 背景减除法

背景减除法是一种有效的运动对象检测算法，由于背景减除简单，而且摄像机在很多情况下是固定的，因此在视频安全应用领域应用较广。其基本思想是利用背景的参数模型来近似背景图像的像素值，将当前帧与背景图像进行差分比较来实现对运动区域的检测，其中区别较大的像素区域被认为是运动区域，而区别较小的像素区域被认为是背景区域。

实现背景减除，我们必须先建立背景模型，且背景图像必须是随着光照或外部环境的变化

而实时更新的，因此背景减除的关键是背景建模及其更新。针对如何建立对于不同场景的动态变化均具有自适应性的背景模型，减少动态场景变化对运动分割的影响，研究人员已提出了许多背景建模算法，但总的来讲可以概括为非回归递推和回归递推两类。非回归背景建模算法是动态的利用从某一时刻开始到当前一段时间内存储的新近观测数据作为样本来进行背景建模。非回归背景建模方法有最简单的帧间差分、中值滤波方法、Toyama 等利用缓存的样本像素来估计背景模型的线性滤波器、Elg~al 等提出的利用一段时间的历史数据来计算背景像素密度的非参数模型等。回归算法在背景估计中无需维持保存背景估计帧的缓冲区，它们是通过回归的方式基于输入的每一帧图像来更新某个时刻的背景模型。这类方法包括广泛应用的线性卡尔曼滤波法、Stauffer 与 Grimson 提出的混合高斯模型^[1]等。

用 $B(x, y)$ 代表背景灰度图像， $C(x, y)$ 代表当前帧图像， $D(x, y)$ 表示差分图像， $BW(x, y)$ 表示二值图像，则背景差分过程可表示为：

$$D(x, y) = |C(x, y) - B(x, y)| \quad (3-3)$$

$$BW(x, y) = \begin{cases} 1, & \text{若 } D(x, y) \geq T \\ 0, & \text{若 } D(x, y) < T \end{cases} \quad (3-4)$$

算法步骤：首先选取无目标情况下摄像机采集到的一帧或几帧图像的均值作为背景图像；然后把采集到的当前帧和背景图像做差分，消除相同的背景区域；接着选择合适的阈值，对差分结果进行二值化，检测出有变化的区域；最后，对疑似目标区域进行判定，得到运动目标。原理示意图 3.2。

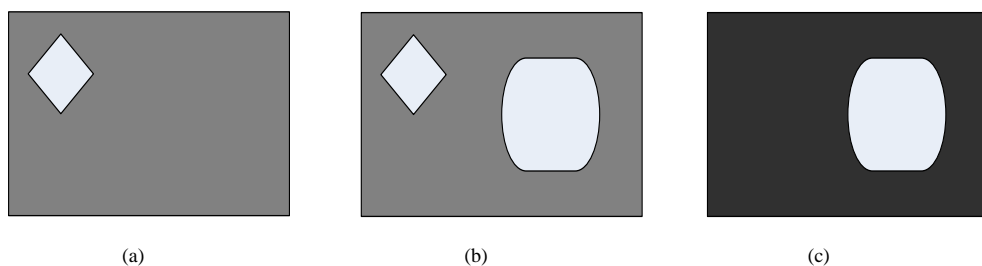


图 3.2 背景减除原理示意图

背景减除优缺点：背景差分法在实时视觉跟踪系统中应用非常广泛，是目标检测最常用的算法之一。它是利用当前图像与背景图像的差分来检测出运动区域的一种技术。它一般能够提供最完整的特征数据，其原理和算法设计简单；根据实际情况确定阈值进行处理，所得结果直接反映了运动目标的位置、大小、形状等信息，能够得到比较精确的运动目标信息。但是，该方法受光线、天气等外界条件变化的影响较大，对背景的依赖性较高。由于背景的建立时间和当前检测图像往往有一定时间间隔，当环境光线等发生变化时，当前图像的背景部分往往和开始建立的背景模型发生很大变动，从而使大量的背景被认为是前景目标，影响了算法的检测效果。另外，算法需要人为设定二值化阈值。此阈值依赖具体情况而定，没有统一的设定方法，而阈值的选择，对于检测又至关重要，影响系统的自动化和环境适应能力。

这里主要介绍两种主要的背景建模算法：单高斯背景建模算法和混合高斯背景建模算法。高斯模型是概率论与统计学中的一个统计模型，又称为正态分布。无论是在自然界还是在现实情况下有很多大数据量在统计意义上都服从该分布，例如全国的人口年龄等。此分布具有良好特性，在实际中有广泛应用。因此，科研人员将其引入到运动目标检测当中，这也是背景减除方法的一个转折点。

3.1.3.1 单高斯模型背景建模算法

单高斯模型的本质是统计建模方法，算法假设背景对应坐标的每个像素值都服从该分布。开始时需要采用一定数量的视频帧对背景进行初始化，当背景收敛后，可以利用此模型对运动目标进行检测。如果对应坐标的像素值与背景模型期望的偏差较大，则认为是运动目标，否则认为是背景。因为算法为统计模型方法，外界环境如光线等会随着时间发生改变，所以要对模型进行实时的更新。算法流程见图 3.3。

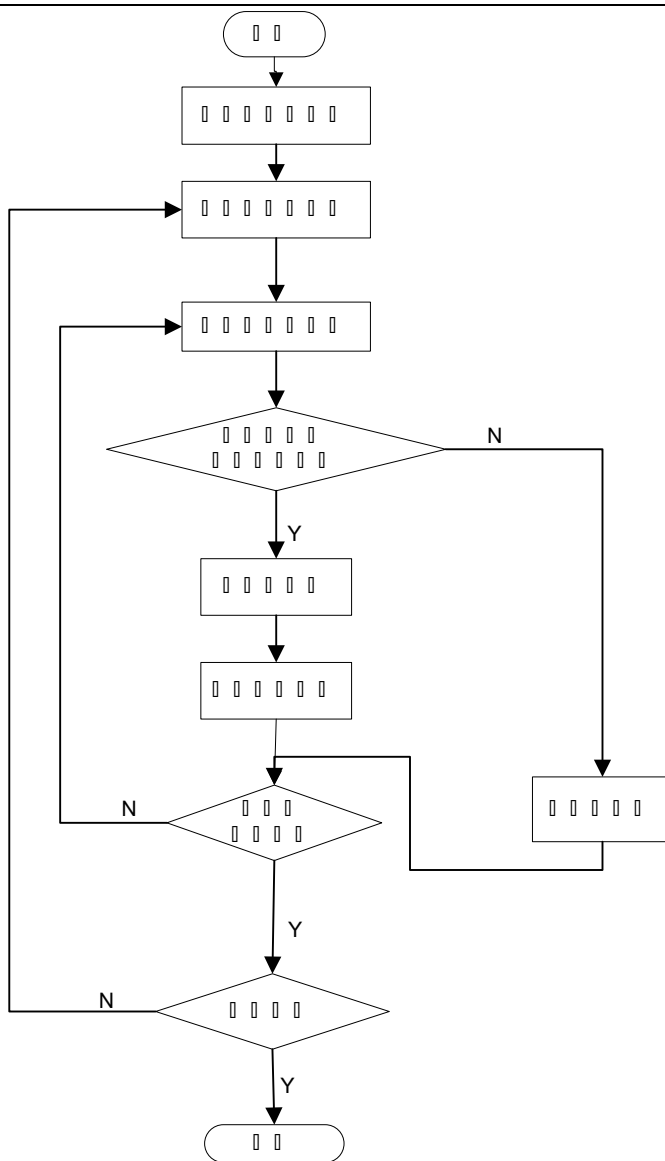


图 3.3 单高斯背景建模算法流程图

建立高斯模型的第一步是对模型进行初始化，即确定模型的两个参数期望(μ)和方差(σ^2)。算法要求对模型初始化时不能有运动目标存在，我们在视频中连续取 K 帧，通过下式对期望和方差进行初始化：

$$u_{(x,y)} = \frac{1}{K} \sum_{i=1}^K u_{(x,y)}(i) \quad (3-5)$$

$$\sigma^2_{(x,y)} = \frac{1}{K} \sum_{i=1}^K (u_{(x,y)}(i) - u_{(x,y)})^2 \quad (3-6)$$

公式表示的意义为： $u_{(x,y)}$ 是高斯背景模型的均值，为连续 K 帧图像对应像素值的均值。 $\sigma^2_{(x,y)}$ 表示连续 K 帧图像对应像素值与均值像素值之差的平方的加权和。

模型收敛以后，可以通过此模型对运动物体进行检测，将当前帧的对应坐标像素值与背景模型对应坐标像素值进行比较。公式如下：

$$|I(x, y) - \mu(x, y)| < M * \sigma(x, y) \quad (3-7)$$

其中 $I(x, y)$ 表示当前图像帧坐标 (x, y) 处的图像的像素值。 $\mu(x, y), \sigma(x, y)$ 意义是背景模型在此坐标处的期望和标准差。 M 为阈值系数，阈值根据不同情况大小不同。通常采用标准差的三倍作为阈值，即 $M = 3$ 。如果当前帧图像对应坐标处与背景模型期望差的绝对值大于 3 倍的标准差则认为运动前景，否则认为是背景。

由于光照等不可抗拒因素，背景是在不断变化的，因此高斯模型要对背景进行实时的更新。对高斯模型背景进行更新，实际上是对模型背景对应像素点的参数进行更新 $(\mu(x, y), \sigma(x, y))$ 。更新方法为，如果通过比较认为此像素点为前景则模型的参数不变，如果是背景点则按公式如下对模型参数进行更新。

$$\mu(x, y) = (1 - \alpha) * \mu(x, y) + \alpha * I(x, y) \quad (3-8)$$

$$\sigma^2(x, y) = (1 - \alpha) * \sigma^2(x, y) + \alpha * (I(x, y) - \mu(x, y))^2 \quad (3-9)$$

其中 α 为更新的权重。一般通过实验选择，通常选择 0.03。

经各种实验结果显示，单高斯模型对简单环境（环境变化不大）检测效果很好，但是对于复杂环境，尤其是光线变化较大、有风导致树枝摇晃或有水波纹导致光线分布不均的情况下会有大量的噪声存在，检测运动目标准确率降低，造成判断错误。并且为后续运动目标分析带来很大的影响。由于其存在固有的局限性，无法在复杂的外部环境中使用。因此在此基础上提出了混合高斯模型。

3.1.3.2 混合高斯模型背景建模算法

混合高斯模型来源于单高斯模型，单高斯模型通过一个高斯模型对背景进行建模。在简单背景下由于背景对应像素值不会发生大的变化所以效果良好。但是在复杂环境下例如有树枝摇摆的情况下，背景一会儿为树枝的像素点，一会儿为背景的像素点的值用一个高斯模型无法表示出来，如果采用多个高斯分布对背景进行建模则可以解决这个问题，成功将摇摆的树枝等噪声融入到背景当中去，这样便可以检测出我们真正需要的运动物体，为后续的目标跟踪和分析提供良好的基础。因此单高斯模型只适用于单模态形式的背景，对于复杂的外部环境需要采用混合高斯模型。

混合高斯模型背景建模的基本原理

采用混合高斯算法对背景进行建模是由美国麻省理工大学的两位研究人员提出的。单高斯模型对背景的每个像素采用一个高斯分布来描述，而混合高斯模型的算法是对背景的每个像素采用多个高斯分布联合描述。我们假设采用 K 个高斯分布来表示背景的一个像素值， K 的大小一般为 3 到 5 之间，具体要根据实际的情况来确定。 K 的取值大那么对背景的颜色变化处理能力就强，但是所需要的计算量也随之提高，如果 K 过大那前景也会认为是背景。因此要合理的选择其值。在模型收敛的情况下，设在 t 时刻图像帧某点像素的颜色值用 X_t 表示，则背景的概率密度可以用公式 (3-10) 表示：

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (3-10)$$

上述公式是一个 3 维的高斯函数，公式的意义为： $\omega_{i,t}$ 为权重系数，表示此时刻第 i 个高斯分布

在所有分布中所占的权重，并且有 $\sum_{i=1}^K \omega_{i,t} = 1$ ， $\eta(X_t, \mu_{i,t}, \Sigma_{i,t})$ 表示第 i 个高斯分布，期望是 $\mu_{i,t}$ ，

协方差矩阵是 $\Sigma_{i,t}$ 。其中 $\eta(X_t, \mu_{i,t}, \Sigma_{i,t})$ 采用公式 (3-11) 计算：

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (3-11)$$

$i = 1, 2, \dots, K$

此公式的本质依然为联合高斯分布函数，公式中的 n 表示维数。在实际情况下，视频帧要分成一张一张的图像进行处理，图像一般都是采用三个通道即 R 、 G 、 B 来表示一个像素点的值。三个通道通常认为是相互独立的，并且有相同的方差。因此 $\Sigma_{k,t} = \sigma_k^2 I$ ，其中 I 表示一个三维的单位矩阵， σ_k^2 表示方差。

混合高斯模型背景建模的学习方法

高斯分布是由期望和方差两个参数共同决定的，混合高斯模型与之原理相似，需要确定 K 个高斯分布的两个参数。这两个参数将直接影响背景模型的准确性以及稳定性。

由 Stauffer 和 grimson 提出的 GMM(混合高斯模型)。在背景初始化完毕后需要检测当前帧的每一个像素值，将当前帧像素值与背景模型的 K 个高斯分布进行匹配比较。假设当前帧某点像素值为 X_t 。如果满足公式 (3-12)，则说明此点为背景点，否则为前景。

$$|X_t - \mu| < \beta \sigma \quad (3-12)$$

其中 β 为判定参数，一般在 2 到 3 之间。

如果当前帧的像素值 X_t 与背景模型中的 K 个高斯模型都匹配不上，则需要进行替换，用一个新的高斯分布来代替背景模型中 K 个高斯分布权重最小的那个高斯模型。新的高斯模型的期望为当前的像素值，同时需要调整新高斯分布的权重和方差，一般权重较小，方差大。

混合高斯模型的权重调整方式如公式 (3-13)：

$$\omega_{k,t} = (1 - \alpha) \omega_{k,t-1} + \alpha M_{k,t} \quad (3-13)$$

式中 α 是一个常数称为学习速率，如果 t 时刻下当前帧的像素值 X_t 跟背景模型的第 i 个高斯分布匹配，则 $M_{i,t}$ 为 1，此时第 i 个高斯分布在背景模型中的权重增大。否则 $M_{i,t}$ 为 0，高斯分布在整个模型中的权重减小。

与当前帧像素值不匹配的高斯模型的参数不变，与之匹配的高斯模型的参数按下列公式进行在线更新。

$$\mu_t = (1 - \rho) \mu_{t-1} + \rho X_t \quad (3-14)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) \quad (3-15)$$

$$\rho = \alpha \eta(X_t | \mu_k, \sigma_k) \quad (3-16)$$

前面两个公式意义和单高斯模型相同， α 是学习率，为一定的常数。

第三个公式直接影响前两个公式的学习速度，通过实验表明，方程的学习速率比较小，模型的收敛速度慢并且准确度比较低。很多相关文献对此进行了改进。

$$\rho = \alpha / \omega_{k,t} \quad (3-17)$$

此参数公式将学习速率与模型的权重结合，当一开始对模型初始化的时候由于高斯分布的权重较小，学习速率较大，这样可以加快模型的收敛速度，在模型稳定以后背景的权重比较大更新速度放慢，致使背景比较稳定。

这种方法可以使模型快速收敛到当前的均值，但是会致使方差也偏小容易造成漏检，使得大量背景点判断为前景点，造成大量的噪声。

混合高斯模型的前景决策方法

混合高斯模型采用多个高斯分布来表示背景中的一个像素值，假设有 k 个高斯分布组成，为了提高算法速度，我们总是将匹配概率大的高斯模型先进行比较，所以要先将 K 个高斯模型进行排序确定优先级。算法假设图像的背景所占的像素个数比较大，所以根据 ω / σ 对其进行排序，根据其值大小确定优先级。模型将 K 个高斯分布的前 B 个分布作为背景，其余视为前景。

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_k > T \right) \quad (3-18)$$

T 为选定的阈值，选定阈值的方法一般是经试验决定的经验值。采用这种方法是因为一般情况下背景出现的频率要高于前景出现的频率。 T 值的实际意义即为在监控场景中背景出现的频率。

3.1.4 光流法

光流的概念最早由 Gibson 提出，Gibson 认为在空间中物体的运动可以由运动场表示，而由运动场的变化而带来的光线的变化像流水一样源源不断的流过眼睛的视网膜，而让我们察觉到了运动的物体，“光流”的名字便来源于此，1981 年 horn 和 Schunck 将光流的概念引入到计算

机视觉的研究当中，将图像的灰度变化同二维速度场联系起来，Horn 和 Schunck 认为客观物体在真实世界中的相对运动是连续变化的，所以运动过程投影到摄像机传感器二维平面上的图像也应该是连续变化的，图像上的光流场就反映了像素点在图像平面上运动的速度矢量。而在图像序列中，物体的运动体现为图像平面上像素值灰度的变化。对此我们就能给出这样的假设：对于客观世界中的同一个点，投影到图像上的一个点，该点的瞬时灰度值是不变的。

基于上述假设，我们可以给出这样的式子：

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (3-19)$$

上式中， $I(x, y, t)$ 表示在 t 时刻处于图像平面上的 (x, y) 位置的像素点的灰度值， $I(x + dx, y + dy, t + dt)$ 表示 $t + dt$ 时刻该点运动到了 $(x + dx, y + dy)$ 位置的灰度值。 dx 、 dy 、 dt 分别表示 x 方向上的偏移量、 y 方向上的偏移量和时间上的偏移量。由于我们假设其灰度值连续变化，所以 dt 趋近于 0 时我们认为其灰度值不变。

上述公式在 dt 趋近于 0 时由泰勒级数展开并忽略二阶无穷小，可以得到：

$$\frac{\partial I(x, y, t)}{\partial t} = \frac{\partial I}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial I}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial I}{\partial t} \quad (3-20)$$

因为被跟踪的像素点在 t 时刻没有发生改变，因此：

$$\frac{\partial I(x, y, t)}{\partial t} = 0 \quad (3-21)$$

所以上式可整理为：

$$-\frac{\partial I}{\partial t} = \frac{\partial I}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial I}{\partial y} \frac{\partial y}{\partial t} \quad (3-22)$$

其中 $\frac{\partial x}{\partial t}$ 可以理解为 t 时刻 x 方向上的变化率，我们用 u 代替， $\frac{\partial y}{\partial t}$ 可以理解为 t 时刻 y 方向上的变化率，我们用 v 代替，则上式最终写为：

$$-\frac{\partial I}{\partial t} = \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v \quad (3-23)$$

在实际应用中，我们可以通过其他约束条件与光流约束方程求解出 u 、 v 。设 $w = (u, v)$ ，

则像素点 (x, y) 的光流可以表示为 $w(x, y)$ 。在理想情况下，被监控场景中的背景应该是不变化的，因此，背景像素点的光流应该为 0，而不为 0 的点就应该是检测到的前景运动目标。

基于光流场计算的方法是通过计算图像的光流场，并根据光流场的分布特征进行聚类，从而提取出运动目标所在的区域。基于光流场计算的方法能在运动背景的情况下完成目标检测和目标跟踪任务。基于光流场计算的运动目标检测方法的检测精度高，适合做精确分析。与基于背景差分和基于帧间差分的运动目标检测方法相比，基于光流场的运动目标检测方法能获得更多的目标运动信息。但是，虽然光流场计算的方法能够检测出运动目标，却无法获得运动目标的准确轮廓；光流场的计算十分复杂，在没有特殊硬件支持的情况下无法做到实时检测，光流场非常容易受到噪声的干扰，在采集设备精度受限的情况下，基于光流场计算的运动目标方法很难做到鲁棒的监控。这些缺点使得光流场的运动目标检测方法离实际的应用还存在一定的距离。

3.2 图像匹配方法

图像匹配指的是将来自不同传感器、不同时间或者不同视角的两幅或多幅图像中相同场景通过一定的变换空间实现像素层的位置匹配的过程。它作为模式识别和数字图像处理的一种基本手段，在遥感遥测、医学影像以及计算机视觉各个方面都有广泛的应用。由于单一获取的图像提供的信息不能满足现代应用的需要，应用图像匹配技术将多种来源的图像中相同的信息匹配起来，则可以获得信息量更大的图像内容，如图像拼接；利用图像匹配技术可以辅助检测目标的变换，如医学诊断；在跟踪领域，用于对人或车辆的识别、跟踪。总之，图像匹配是实现上述各类问题的关键问题，直接影响到精度和有效性。

图像匹配的数学定义一般表示为：如果将图像表示成一个二维序列 $I_1(x, y)$, $I_2(x, y)$ ，其中 $I_1(x, y)$, $I_2(x, y)$ 分别表示两幅需要配准的图像在点 (x, y) 处的灰度值。图像 I_1 , I_2 之间配准关系的数学表达式如下： $I_2(x, y) = g(I_1(f(x, y)))$ 。其中 I_1 , I_2 分别代表参考图像和待配准影像， f 代表二维几何空间变换函数， g 为一维灰度变换函数。从以上图像配准关系数学表达式

可以看出, 图像配准包括两方面含义, 一方面是实现几何空间上的一一对应关系, 这一点可以在对函数 f 的求解上得以体现; 另一方面是对应像素之间灰度上的一致性, 这一点可以在对函数 g 的求解上得以体现。因此, 图像配准的主要目的可以看作是寻找配准图像间空间与灰度的最佳变换关系, 在考虑畸变的前提下, 实现两幅图像的最佳匹配。

确定两幅影像足够数量的匹配控制点是图像配准技术的核心部分, 也是广大研究者研究的重点。许多研究者把图像配准技术分成以下四个功能块:

1、特征空间, 是指从图像中提取出的用于匹配的信息集合。图像特征在图像配准中有着非常重要的意义, 它决定了这幅影像使用什么样的方法进行配准。常见的特征如灰度值、边缘、曲线、曲面等; 拐角、交叉线、高曲率的点等显著的视觉特征; 不变矩以及高层语义的描述算子等统计特征。

2、搜索空间, 指在偏移图像特征和参考图像特征之间建立对应关系的所有变换集合。图像的几何变换可以分为三类: 全局的、局部的和位移场形式的。全局的变换通常基于矩阵代数理论, 用一个参数矩阵来表述整个图像的变换。局部变换指不同的位置具有不同的变化参数模型。位移场方法, 也称为光流场法, 使用一个(连续的)函数优化机制, 为图像中的每一点计算出一个独立的位移量, 并使用某种规整化机制进行约束。

3、搜索策略, 用于提高搜索效率的策略, 在搜索空间大、精度要求高时, 最佳匹配方法往往非常耗时, 因此采用相应的策略是必要的。控制点结合最小二乘准则是一个很通用的变换参数确定方法, 它通常用在基于特征的匹配方法上。此外, 常用的搜索策略还有: 松弛法、分层匹配法、牛顿法、共轭梯度法、最速下降法等。

4、相似度测量。用来评估两幅图像之间的特征数据的匹配程度, 如相关系数等。

根据匹配基元的不同, 现有的图像匹配方法可分为四类: 基于像素灰度的匹配方法、基于特征提取的匹配方法、基于变换域的匹配方法和基于投影的匹配方法。

3.2.1 基于像素灰度的匹配方法

基于像素灰度的图像匹配是指利用图像的灰度信息对两幅或者多幅图像进行匹配, 是一类最简单的匹配算法。该算法利用图像本身具有的灰度的一些统计信息来度量图像的相似程度。

这里简单介绍模板图像的具体步骤。首先, 我们定义模板图像为 $T(x, y)$, 定义待匹配的目标图

像为 I ，沿图像中的所有点移动模板并在每个位置计算相似度量 S 。 S 是一个相似度函数，该函数的参数包括模板中各点的灰度值 $T(x, y)$ 以及待匹配区域移到图像当前位置时区域中的灰度值 $f(x+u, y+v)$ ，根据这些已经得到的灰度值计算一个标量值作为相似度量。采用这个方法，在变换空间中每个点都会得到一个相似度量：

$$S(x, y) = S\{t(x, y), f(x+u, y+v); (u, v) \in I\} \quad (3-24)$$

最简单的相似性度量方法是计算模板图像与待匹配的目标图像之间差值的绝对值总和 (SAD) 或者差值的平方和 (SSD)。在模板图像和待匹配的目标图像是完全相同的情况下，计算后所得到的相似度量应该为 0。区别越大，值越大。在光照情况保持不变的情况下，这种相似性度量方法的结果准确度比较高。但是在光照发生变化的情况下，图像中的灰度值已经不再相等，采用该方法结果会发生偏差。归一化互相关系数 (NCC) 相似性度量不随任何线性的光照变化而变化。归一化互相关系数的绝对值越接近 0 就表示模板与图像越不一致。

该类算法实现简单，在灰度及几何畸变不大的情况下有较好的估计精度和鲁棒性，抗噪性强。但应用范围较窄，不能直接用于校正图像的非线性形变，在最优变换的搜索过程中往往需要巨大的运算量。

3.2.2 基于特征提取的匹配方法

基于特征提取的图像匹配算法需要提取图像的特征信息，比如图像边缘，图像角点，图像纹理等，然后将图像的匹配过程转变为特征之间的匹配过程。这种方法有很多分支，共同点是要对所提取的初级信息进行处理，然后进行匹配。大致可分为：基于角点的图像匹配算法和基于边缘的图像匹配算法。

3.2.2.1 基于角点的图像匹配算法

直接基于图像灰度的角点检测算法选取局部范围内灰度和梯度变化剧烈的极大点，主要是通过计算曲率及梯度来检测角点。其稳定性较基于边缘的角点提取算法要好，可以给出更为准确的角点定位。目前在图像匹配中，常见的基于角点提取有 Moravec 算子，SUSAN 算子，Harris

算子等。其中较著名的有：Smith 和 Brady 提出的一种 SUSAN(Smallest Univalve Segment Assimilating Nucleus)探测算子。

SUSAN 算子是由英国牛津大学的 SMITH S M, BRADYJM 于 1997 年首先提出的，是 Smallest Univalve Segment Assimilating Nucleus 的缩写，即同化核分割最小值。它主要用来计算图像中的角点特征。SUSAN 算法的特点是：

- 1、对角点检测比对边缘检测的效果要好，使用图像特征提取。
- 2、无需梯度运算，保证了算法的效率。
- 3、具有积分特征，这样使得 SUSAN 算法在抗噪和计算速度方面有较大改进。

将位于圆形窗口模板中心等待检测的像素点称为“核”。核的领域被划分为两个区域：亮度值等于（相似于）核亮度的区域即核值相似区（即 USAN 区域）和亮度不相似于核灰度的区域。USAN 区域包含了图像结构中大量的信息，是 SUSAN 算法的准则，USAN 的几种典型形状见图 3.3。图形窗口模板滑动到不同位置时 USAN 区域的面积大小是不同的，如当核位于直线边缘上时，USAN 区域面积接近最大值的一半，当模板核靠近角点时，USAN 面积接近最大值四分之一。在一幅图像中搜索图像角点或边缘点，就是搜索 USAN 最小的点，即搜索最小化同化分割相同值。

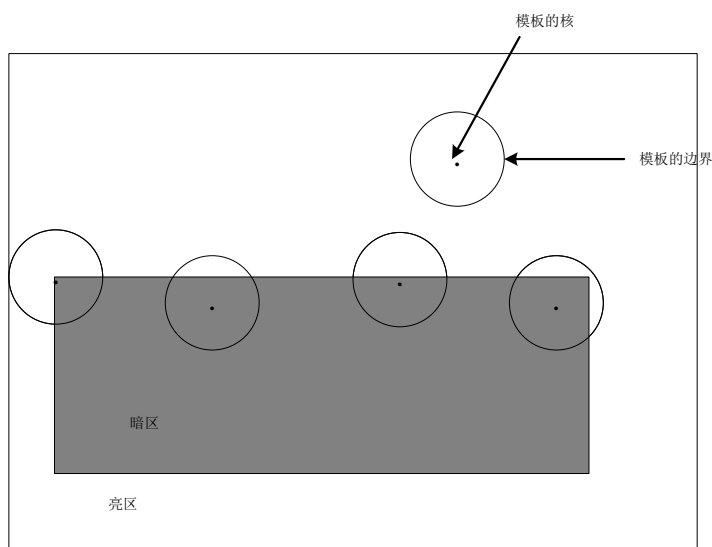


图 3.4 SUSAN 模板

SUSAN 模板在图像上滑动，在每个位置上，比较模板内各图像像素的亮度与模板核的亮度：

$$C(\vec{r}, \vec{r}_0) = \begin{cases} 1 & \text{if } |I(\vec{r}) - I(\vec{r}_0)| \leq T \\ 0 & \text{if } |I(\vec{r}) - I(\vec{r}_0)| > T \end{cases} \quad (3-25)$$

其中 \vec{r}_0 为模板核在图像中的坐标， \vec{r} 为模板区域内其它点的坐标。 $I(\vec{r})$ 为 \vec{r} 点的亮度值， $I(\vec{r}_0)$ 为点 \vec{r}_0 点的亮度值。T 为表示亮度差值的阈值，控制角点提取数量。 $C(\vec{r}, \vec{r}_0)$ 为亮度值比较结果。

为了取得更加稳定可靠的结果，可以将公式（3-25）改写为：

$$C(\vec{r}, \vec{r}_0) = \begin{cases} \exp\{-[\frac{I(r) - I(r_0)}{t}]^6\} & \text{if } |I(\vec{r}) - I(\vec{r}_0)| \leq T \\ 0 & \text{if } |I(\vec{r}) - I(\vec{r}_0)| > T \end{cases} \quad (3-26)$$

其中 t 为参数，一般设为 75。

模板内所有点与核亮度值比较的和（即区域的 SUSAN 值）为：

$$n(\vec{r}_0) = \sum_{\vec{r}} C(\vec{r}, \vec{r}_0) \quad (3-27)$$

将 $n(\vec{r}_0)$ 与一给定阈值 g （称为几何阈值）比较，得到 SUSAN 算子对图像的边缘响应：

$$R(\vec{r}_0) = \begin{cases} g - n(\vec{r}_0) & \text{if } n(\vec{r}_0) < g \\ 0 & \text{otherwise} \end{cases} \quad (3-28)$$

假设模板能取得的最大 n 值为 n_{\max} 。为了消除噪声的影响，通常用阈值 g 控制角点的提取质量，通常取为 $3 \times n_{\max} / 4$ 。这样就可以确定边缘的位置。

当区域边界模糊时，SUSAN 算法会产生虚假的角点，为了消除虚假角点的影响，通常使用式计算重心：

$$\bar{\vec{r}}(\vec{r}_0) = \frac{\sum_{\vec{r}} \vec{r} c(\vec{r}, \vec{r}_0)}{\sum_{\vec{r}} c(\vec{r}, \vec{r}_0)} \quad (3-29)$$

然后求出核到重心的距离，对应正确角点，其重心距离核较远，通过该距离可以消除虚假

的角点。最后使用非最大抑制(No Max Suppression)方法,即通过将一个边缘点作为 3×3 模板的中心,在它的八邻域范围内的点进行比较,亮度最大保留,这样就可以找出角点。

3.2.2.2 基于边缘的图像匹配算法

基于边缘特征点的图像匹配,首先通过边缘检测寻找边缘特征点,然后将特征进行相似性度量,最终达到定位的目的。边缘检测的算法很多,主要分为以下几类:

1、微分检测边缘算法。一般是对边缘灰度曲线求一阶或二阶导数再取极值,如微分算子 Roberts、Sobel、Prewitt、Laplacian 等。

2、先滤波后检测边缘算法,如 LoG、Canny 等。

3、基于曲面拟合的边缘检测算法。

4、现代边缘检测算法,如神经网络边缘检测及小波分析法等。

其中, Sobel 和 Prewitt 算子对图像检测出来的边缘比较稳定。伪边缘比较少,但检测出的边缘比较宽,定位精度不高,对噪声较敏感; Roberts 和 Laplacian 算子的边缘定位精度较高,但对噪声也比较敏感; LoG 算子中的 σ 正比于低通滤波器的宽度, σ 越大,平滑作用越显著,去噪越好,但图像的细节也损失越大,边缘定位精度也就越低,因此要根据具体问题对噪声水平和边缘定位精度的要求适当选取 σ , 自适应能力不强; Canny 算子在理论上和实践中都有很好的边缘检测效果,去噪能力强,适用性广泛。下面主要介绍 Canny 算子。

Canny 算子自从被提出以来,就成为边缘检测中广泛应用的算法之一。其基本思想是:首先用高斯滤波器平滑图像,然后使用差分模板近似计算梯度值和方向,再在梯度方向上确定局部极大模点,用双门限法去除噪声并保留边缘点,最后进行插值将边缘点连接起来以形成连贯的曲线。具体算法如下:

1、高斯滤波平滑图像

设二维高斯滤波函数为:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (3-30)$$

其梯度向量为:

$$\nabla G = \begin{bmatrix} \partial G / \partial x \\ \partial G / \partial y \end{bmatrix} \quad (3-31)$$

用分解的方法将梯度矢量 ∇G 的两个滤波卷积模板分解为两个一维的行列滤波器，以提高速度：

$$\frac{\partial G}{\partial x} = kx \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{y^2}{2\sigma^2}\right) = h_1(x)h_2(y) \quad (3-32)$$

$$\frac{\partial G}{\partial y} = ky \exp\left(-\frac{y^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) = h_1(y)h_2(x) \quad (3-33)$$

其中：

$$\begin{aligned} h_1(x) &= \sqrt{k}x \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad h_2(y) = \sqrt{k} \exp\left(-\frac{y^2}{2\sigma^2}\right) \\ h_1(y) &= \sqrt{k}y \exp\left(-\frac{y^2}{2\sigma^2}\right), \quad h_2(x) = \sqrt{k} \exp\left(-\frac{x^2}{2\sigma^2}\right) \end{aligned} \quad (3-34)$$

将这两个卷积模板分别与图像 $f[x, y]$ 进行卷积运算，得到输出图像 $I[x, y]$ ：

$$\begin{aligned} E_x[x, y] &= \frac{\partial G}{\partial x} f[x, y] \\ I[x, y] &= \frac{\partial G}{\partial y} E_x[x, y] \end{aligned} \quad (3-35)$$

其中 k 为常数， σ 为高斯滤波器分布参数，它控制着平滑强弱度。对于 σ 小的滤波器，虽然定位精度高，但图像平滑效果弱； σ 大的情况则相反。

2、计算梯度的幅值和方向

Canny 算法采用 2×2 邻域一阶偏导的有限差分来计算平滑后的数据数组 $I[i, j]$ 的梯度幅值和梯度方向。其中 x 和 y 方向偏导数的两个数组 $P_x[i, j]$ 和 $P_y[i, j]$ 分别为：

$$\begin{aligned} P_x[i, j] &= (I[i, j+1] - I[i, j] + I[i+1, j+1] - I[i+1, j]) / 2 \\ P_y[i, j] &= (I[i, j] - I[i+1, j] + I[i, j+1] - I[i+1, j+1]) / 2 \end{aligned} \quad (3-36)$$

像素的梯度幅值和梯度方向用直角坐标到极坐标的坐标转化公式来计算，梯度幅值为：

$$M[i, j] = \sqrt{p_x[i, j]^2 + p_y[i, j]^2} \quad (3-37)$$

梯度方向为:

$$\theta[i, j] = \arctan(P_y[i, j] / P_x[i, j]) \quad (3-38)$$

3、对梯度幅值进行非极大值抑制

为了获得较高的定位精度，需细化梯度幅值图像 $M[i, j]$ 中的宽屋脊带，保留幅值局部变化最大的点，这就是非极大值抑制。在这一过程中，Canny 算法使用包含 8 方向的 3×3 邻域对梯度幅值数组 $M[i, j]$ 中所有像素沿梯度方向进行插值。在每一像素点上，将中心点像素 $m[i, j]$ 与沿梯度方向的两个插值结果进行比较， $\xi[i, j]$ 为像素邻域中心处沿着梯度方向的扇区区域，非极大值抑制在此区域进行。如果邻域中心点的幅值 $m[i, j]$ 低于梯度方向上两个插值的结果，则认为其不是边缘，并将 $m[i, j]$ 对应的边缘标志位置为 0；反之认为 $m[i, j]$ 是边缘，将 $m[i, j]$ 对应的边缘标志位置为 1。通过非极大值抑制把 $M[i, j]$ 宽屋脊带细化为一个像素宽，同时还保留了屋脊的梯度幅值。非极大值抑制的过程用数学公式表示为：

$$N[i, j] = NMS(M[i, j], \xi[i, j]) \quad (3-39)$$

4、用双阈值算法检测和连接边缘

双阈值算法是对经过非极大值抑制的图像 $N[i, j]$ 分别使用高、低两个阈值 $th1$ 和 $th2$ 获得边缘图像 $T_h[i, j]$ 和 $T_l[i, j]$ 。由于图像 $T_h[i, j]$ 是由高阈值得到，因此它不包含伪边缘，但其边缘闭合性也较低。因此双阈值算法要在 $T_h[i, j]$ 中把边缘连接成轮廓，就需要在达到轮廓端点时，在低阈值得到的边缘图像 $T_l[i, j]$ 的 8 邻域位置寻找可以连接到轮廓上的边缘，利用递归跟踪的算法不断在 $T_l[i, j]$ 中寻找边缘，直到将 $T_h[i, j]$ 中所有的不连续边缘都连接起来为止。

基于特征的匹配方法能够对整个图像的各种分析转化为对图像特征的分析，从而大大减少了图像处理过程的运算量，对灰度变化、图像变形以及遮挡等都有较好的适应能力。但是此方法没有统一的模型，而且所提取的特征各不相同，因此每种方法都有各自的适用领域。

3.2.3 基于变换域的匹配方法

基于变换域信息的图像配准方法包括：基于傅立叶、小波变换和 Warsh 变换等。其中常用的是傅立叶变换图像配准方法，该方法主要有以下一些优点：图像的平移、旋转、仿射等变换

在傅立叶变换频域中都能反映出来，同时傅立叶变换域的方法对抵抗噪声具有一定的鲁棒性。傅氏变换由于有成熟的快速算法和易于硬件实现，因此，傅氏变换是图像配准中常用的方法之一。但它也有相当的局限性，如傅氏变换方法只能用来配准灰度属性有线性正相关的图像，图像之间必须是严格满足定义好的变换关系等。

3.2.3.1 基于傅氏变换的图像配准方法

基于傅氏变换方法的相位相关方法，是通过相位关系来反映偏移量，而灰度变化在频域中主要影响幅值，对相位影响很小，所以相位相关法具有很高的稳定性。

1975 年，相位相关(Phase Correlation)的概念被应用到图像配准领域中，很好地解决了仅存在平移的图像之间的配准。相位相关方法的主要依据是傅立叶的平移性质。考虑两幅图像 f_1 和 f_2 存在 (d_x, d_y) 的平移，即：

$$f_2(x, y) = f_1(x - dx, y - dy) \quad (3-40)$$

对其进行傅氏变换，反映到频域上具有以下形式：

$$F_2(w_x, w_y) = F_1(w_x, w_y) e^{-j2\pi(w_x d_x + w_y d_y)} \quad (3-41)$$

上式说明，两幅具有平移量的图像变换到频域中有相同的幅值，但有一个相位差，而这个相位差与图像间的平移量 (d_x, d_y) 有直接的关系。根据平移定理，可知以上相位差等于两幅图像的互功率谱的相位，即：

$$\frac{F_1(w_x, w_y) F_2^*(w_x, w_y)}{|F_1(w_x, w_y) F_2^*(w_x, w_y)|} = e^{j2\pi(w_x d_x + w_y d_y)} \quad (3-42)$$

其中*表示复共轭，式的右边部分为一个虚指数，其傅立叶反变换在平移量 (d_x, d_y) 处为一个单位冲击函数，指示出了两图像间的平移位置。

De Castro 等将相位方法应用到存在旋转变换的图像之间的配准，当图像 f_1 和图像 f_2 间除平移量 (d_x, d_y) 外还存在角度为 θ_0 的旋转变换时，图像间的关系为：

$$f_2(x, y) = f_1(x \cos \theta_0 + y \sin \theta_0 - dx, -x \sin \theta_0 + y \cos \theta_0 - dy) \quad (3-43)$$

由傅立叶变换的平移和旋转性质，图像 f_1 和图像 f_2 的傅立叶变换间的关系为：

$$F_2(w_x, w_y) = e^{-j2\pi(w_x d_x + w_y d_y)} \times F_1(w_x \cos \theta_0 + w_y \sin \theta_0, -w_x \sin \theta_0 + w_y \cos \theta_0) \quad (3-44)$$

由上式可以看到平移信息仅存在于图像频谱的幅角中，若令 M_1 和 M_2 表示 F_1 和 F_2 的模，对上式两边取模可得：

$$M_2(w_x, w_y) = M_1(w_x \cos \theta_0 + w_y \sin \theta_0, -w_x \sin \theta_0 + w_y \cos \theta_0) \quad (3-45)$$

所以， M_2 和 M_1 间仅相差一个旋转角度。利用极坐标的性质，可将旋转参数转化为平移参数，从而利用相位相关法可以求得：

$$M_2(\rho, \theta) = M_1(\rho, \theta - \theta_0) \quad (3-46)$$

如果图像间还存在比例缩放因子 s ，则极坐标下其傅里叶变换频谱的模之间的关系变为：

$$M_2(\rho, \theta) = M_1(\rho / s, \theta - \theta_0) \quad (3-47)$$

令 $u = \log \rho$, $v = \log s$ ，则上式变换为：

$$M_2(u, \theta) = M_1(u - v, \theta - \theta_0) \quad (3-48)$$

可见，上述图像间也是相差一个平移量，同样可由相位相关法求得。

3.2.3.2 基于小波的图像配准方法

20 世纪 80 年代，自提出小波变换以来，小波变换一度成为图像处理领域的研究热点，于是也发展了基于小波变换的各种图像配准算法。文献^[2]中，阐述了利用小波分解方法做图像配准的合理性，结合小波分解和相关系数法，利用子空间搜索和统计特征匹配代替穷尽搜索法，在并行计算机上实现对遥感图像的多分辨率配准；文献^[3]中，首先利用小波分解得到低频图像信息，同时结合交互信息法得到初始参数。然后，在下一层的计算中采用前期的参数作为初始值，以减少计算负担，不断细化，直至产生最优配准参数的集合。

3.2.4 基于投影的匹配方法

投影匹配算法是把二维的图像灰度值投影变换成一维的数据，再在一维数据的基础上进行匹配运算，通过减少数据的维数来达到提高匹配速度的目的，从而有效提高匹配的速度。

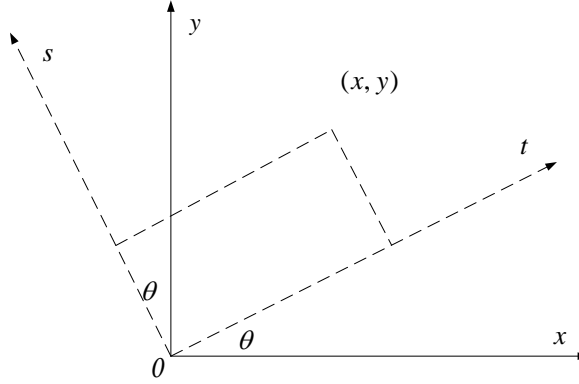


图 3.5 投影坐标与原坐标间的对应关系示意图

投影坐标 (t, s) 与原坐标 (x, y) 间的对应关系见图 3.4, 其中 $f(x, y)$ 为图像函数, s 为投影方向, t 为其垂直方向, 则 $f(x, y)$ 沿着 s 的投影定义为:

$$p(t, \theta) = \int f(t \cos \theta - s \sin \theta, t \sin \theta + s \cos \theta) ds \quad (3-49)$$

当 θ 固定时, $p(t, \theta)$ 为 t 的函数, 是一个一维波形。当 θ 在 $0 \sim 2\pi$ 间变化时, 可以得到 $f(x, y)$ 在不同方向上的投影, 特别地在 x, y 轴上的投影为:

$$\begin{cases} p_x = \int f(x, y) dy \\ p_y = \int f(x, y) dx \end{cases} \quad (3-50)$$

根据图像灰度的空间特性可以利用一维投影数据的差分字符串进行图像匹配的快速运算。考虑到两幅匹配图像相邻像素的灰度值的相对大小关系应该是相同的, 如果某搜索窗口与参考图像匹配, 那么相应像素之间的灰度大小关系一定与参考图像相一致。反之, 如果不满足这个条件就可以认为与参考图像不匹配, 而且灰度值之间的大小关系不随灰度值的线性变换改变而改变, 这一特征对灰度的线性变换有“免疫性”。

基于以上考虑,对数据一维投影后,对数据串进行差分处理:在相邻的两个数据中,用后面的减前面的,结果为正,记为 1,结果为负,记为 0。这样,一个长度为 N 的一维数据可以得到 $N-1$ 维的 0~1 字符串,这个字符串就构成了表征模板或者搜索窗口的特征。这样在模板匹配过程中,只需要将模板的 0~1 字符串和搜索窗口中的字符串进行匹配运算,就可以获得配准的结果。这种方法在保证匹配结果的前提下,可以提高匹配的速度。

另外,为了能在输入图像中检测出已经旋转的模板,一些文献中引入圆投影方法。圆投影将圆形窗口中的二维图像函数变换为一个以一系列同心圆的半径为自变量的一维函数。二维图像的圆投影不仅具有旋转不变性,而且还大幅度降低了匹配的计算复杂度,是在输入图像中检测目标物体的一种快速有效的方法。

3.3 人工神经网络

3.3.1 神经网络历史

人工神经网络(Artificial Neural Network, 简称 ANN),意在模拟人类大脑或者动物神经基本特性而建立的数学模型。人工神经网络数学模型的提出,最早可追溯到 1943 年由 McCulloch 和 Pitts 所提出的 MP 模型,其给出了基本神经元的数学描述和神经网络构建方法,并能用于实现一般的布尔运算。1957 年, Rosenblatt 提出的感知机(Perceptron)模型,实现了简单的线性分类器功能,使得这一阶段关于神经网络的研究出现了高潮;但在 1969 年, Minsky 和 Papert 在著作《Perceptron》中,指出了感知机模型的缺陷,即无法解决高阶谓词问题,使得对于人工神经网络的研究陷入了一个较长的停滞期。

在 1982 年,人工神经网络迎来了又一次的研究进展, Hopfield 提出了一种反馈神经网络模型——Hopfield 神经网络模型,其在神经网络之上引入了“能量函数”的概念,运用非线性动力学的研究方法研究人工神经网络的特性。而后 Hopfield 和他的同事将最初的模型由离散型推广为连续型,并提出了获得 NP 问题近似最优解的方法,极大的推动了神经网络的发展。1985 年, Hinton 等在 Hopfield 神经网络的基础上提出了 Boltzmann 机模型,引入了随机机制,保证系统能够收敛到全局稳定点。1986 年, Rumelhart 等提出了多层神经网络的学习算法——“误差反向传播算法”,即 BP 算法,解决了多层神经网络的学习问题。在此基础上,许多其他形式的神

神经网络也被逐渐提出，如径向基函数网络和函数链网络等。近年来，深层神经网络（Deep Neural Network）的提出并成功的应用在多任务学习中，使得人工神经网络又再次得到了发展。

3.3.2 感知器

人工神经网络的基本构成是以感知器（Perceptron）为基础单元的，见图 3.5。单感知器网络也称为单层人工神经网络。

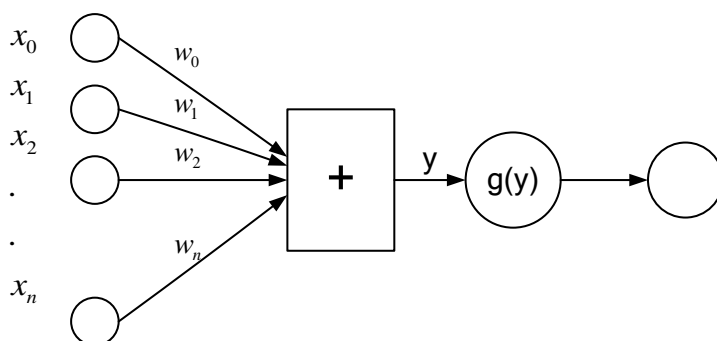


图 3.6 人工神经网络的构成示意图

感知器以实数向量为输入，通过计算输入向量的带权组合，判别结果是否大于给定阈值，若是，则输出 1；否则，输出-1。计算公式：

$$g(y) = \begin{cases} 1 & \text{if } \sum_{i=0}^n w_i x_i > threshold \\ -1 & otherwise \end{cases} \quad (3-51)$$

对于单感知器的学习，就是要找到一个权值向量，使得对于给定的训练样本，能正确的进行判别，即输出对应的+1 或者-1。感知器的学习重要的是要保证其收敛性，常用的学习法则为两种：感知器法则和 delta 法则。

感知器法则

感知器法则的训练过程从随机设定的权值开始，迭代的对每个训练样本应用该感知器，若错分了样本，则修改感知器的权值，直到对于给定的训练样本，全部正确分类为止。权值的修改法则为：

$$w_i \leftarrow w_i + \Delta w_i \quad (3-52)$$

其中:

$$\Delta w_i = \alpha(y_i - o_i)x_i \quad (3-53)$$

y_i 是第 i 个样本的标准输出, o_i 是感知器的输出, α 代表学习速率。学习速率的作用是调整权值修改的幅度, 若设过大, 则算法不易收敛, 若过小, 则需要运算量较大。在应用中, 常设一较小的数值。Minsky 和 Papert 在 1969 年证明了, 在训练样本线性可分的条件下, 能通过有限次的使用感知器法则后, 能得到对所有训练样本都能正确分类的权向量。如果样本不是线性可分的, 以上过程并不保证能最终收敛。

Delta 法则

对于线性可分的样本, 使用感知器法则就能得到对应的判别向量, 当样本不是线性可分时, 使用感知器法则不能够收敛, 在此情况下, 可求一个分类器的最佳近似, 这时候可采用 delta 法则。

对于给定样本集, 要寻找权向量 w^* , 使得输出 o_i 与期望 y_i 尽可能接近。对于给定的样本集, 首先定义训练误差为:

$$E = \frac{1}{2} \sum_k (y_k - o_k)^2 \quad (3-54)$$

其中 y_k 是各个样本的期望输出, $o_k = f(w^* \cdot x_k)$ 是第 k 个样本对应的输出。训练感知器即求分类向量, 使训练误差最小。为使 E 最小, delta 法则从任意初始权向量开始, 使用梯度下降方法对权向量进行更新, 直到误差最小化。 计算 E 的梯度:

$$\nabla E = \frac{\partial E}{\partial w} = \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n} \right] \quad (3-55)$$

$$\begin{aligned}
 \frac{\partial E}{\partial w_i} &= \frac{1}{2} \sum_k \frac{\partial (y_k - o_k)^2}{\partial w_i} = \sum_k (y_k - o_k) \frac{\partial}{\partial w_i} (y_k - o_k) \\
 &= \sum_k (y_k - o_k) \frac{\partial}{\partial w_i} (y_k - \vec{w} \cdot \vec{x}_k) \\
 &= \sum_k (y_k - o_k) (-x_{ik})
 \end{aligned} \tag{3-56}$$

至此，就可以由该梯度方向 $\frac{\partial E}{\partial w_i}$ 对权值进行更新。与感知器法则一样，Delta 法则训练过程中对权值的更新为：

$$w_i \leftarrow w_i + \Delta w_i \tag{3-57}$$

其中：

$$\begin{aligned}
 \Delta w_i &= -\alpha \frac{\partial E}{\partial w_i} \\
 &= \alpha \sum_k (y_k - o_k) x_{ik}
 \end{aligned} \tag{3-58}$$

α 代表学习速率，注意当 α 过大时，则算法容易超过最优值，而不会收敛，因此 α 的选择对于算法十分重要，一般选择一足够小的速率。

3.3.3 多层神经网络

单层感知器或者单层神经网络其实质是数据空间中的一个线性判别平面，为达到非线性化判别的目的，我们可以采用多层感知器联接的方式，前层网络的输出作为后层网络的输入，典型的多层神经网络的结构见图 3.7。

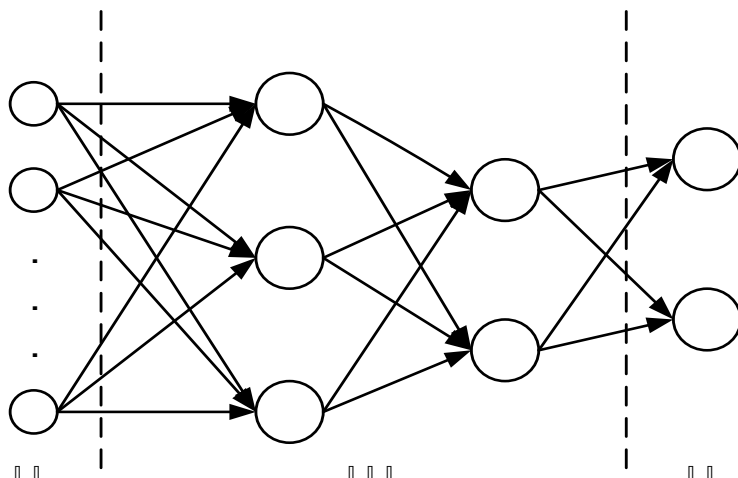


图 3.7 多层神经网络

多层神经网络中，信号只能由低层向高层传播，其中除输入和输出层之外的网络层都称为隐藏层，隐藏层在网络中不接受外界信号，也不对外界进行输出。多层神经网络可以实现非线性映射，如上图中的网络其输出函数为：

$$f(x) = f_3(f_2(f_1(x \cdot w_1)w_2)w_3) \quad (3-59)$$

可见，引入多层结构之后，可以使神经网络实现更复杂的判别和决策，但这同时也提高了学习的复杂度。前小节所介绍的线性感知器单元却并不适合于构建多层网络，由于其为不可微单元，训练时候会十分复杂，因此，一般采用 sigmoid 单元作为基本单元来构建多层神经网络。

Sigmoid 单元的输出为：

$$f(y) = \frac{1}{1 + e^{-y}} = \frac{1}{1 + e^{\sum w_i \cdot x_i}} \quad (3-60)$$

$$f'(y) = -\frac{1}{(1 + e^{-y})^2}(-e^{-y}) = f(y)(1 - f(y)) \quad (3-61)$$

后面可以看到，sigmoid 单元导数具有的性质对于采用梯度下降方法学习神经网络权值会带来极大的方便。对于多层神经网络的学习，反向传播算法（Back Propagation，BP）是最常用也是最简单的方法之一，BP 算法包含两个步骤：前向传播和反向传播，算法通过不断的迭代重复这两步，直到收敛。前向传播阶段依据训练样本计算实际的输出和测量误差，利用该输出误

差在后向传播阶段计算输出层的前导层的输出误差，继续利用所得的输出误差计算更前一层的输出误差，直到输入接受层，可见，反向传播阶段就是测量误差向后逐层传递的过程。

对每一个训练样本 k ，定义其训练误差为：

$$E_k = \frac{1}{2} \sum_{p \in Op} (y_{kp} - o_{kp})^2 \quad (3-62)$$

其中 Op 代表的是输出单元， y_{kp} 和 o_{kp} 分别代表 p 单元的目标值和实际输出值。令 w_{pi} 代表 p 单元第 i 个输入的权值，则在训练过程中，权值的更新为：

$$w_{pi} \leftarrow w_{pi} + \Delta w_{pi} \quad (3-63)$$

$$\Delta w_{pi} = -\alpha \frac{\partial E_k}{\partial w_{pi}} \quad (3-64)$$

接下来，分别针对输出层和隐藏层介绍其权值更新的计算方式。

输出层权值更新

输出层权值更新见图 3.8，权值 w_{pi} 仅能通过输出节点 p 作用于网络，令 $o_p = f(y_p)$ ，则：

$$\begin{aligned} \frac{\partial E_k}{\partial w_{pi}} &= \frac{\partial E_k}{\partial o_p} \frac{\partial o_p}{\partial w_{pi}} \\ &= \frac{1}{2} \frac{\partial \sum_{p \in Op} (y_{kp} - o_{kp})^2}{\partial o_p} \frac{\partial o_p}{\partial w_{pi}} \\ &= -(y_{kp} - o_{kp}) o'_{kp} x_{pi} \end{aligned} \quad (3-65)$$

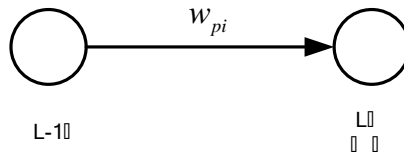


图 3.8 输出层权值更新

由前面 sigmoid 函数性质可知 $o'_p = o_p(1 - o_p)$ ，得到输出单元的权值更新为：

$$\Delta w_{pi} = -\alpha \frac{\partial E_k}{\partial w_{pi}} = -\alpha (y_{kp} - o_{kp}) o_{kp} (1 - o_{kp}) x_{pi} \quad (3-66)$$

隐藏层权值更新

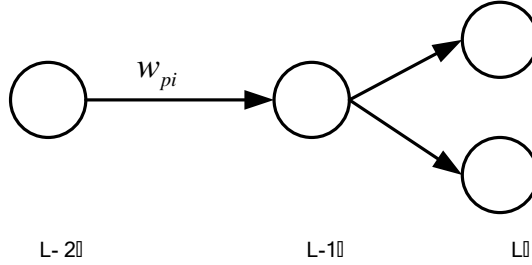


图 3.9 隐藏层权值更新

对于多层网络中的隐藏层单元来说，其输出作为下一层的输入，最终影响到输出误差 E_k ，见图 3.9。由网络的联结结构，对于节点单元 p ，所有 p 向前传播所经历的节点单元所造成的测量误差都应在更新 w_{pi} 时进行考虑。

$$\begin{aligned} \frac{\partial E_k}{\partial w_{pi}} &= \frac{\partial E_k}{\partial net_p} \frac{\partial net_p}{\partial w_{pi}} \\ &= \frac{\partial net_p}{\partial w_{pi}} \sum_M \frac{\partial E_k}{\partial net_m} \frac{\partial net_m}{\partial net_p} \end{aligned} \quad (3-67)$$

其中 $net_p = \sum w_i \cdot x_i$ ， net_m 代表 p 前向传播所经历的单元。

令 $\delta_m = -\frac{\partial E_k}{\partial net_m} = -(y_{km} - o_{km}) o_{km} (1 - o_{km})$ ，则：

$$\begin{aligned} \frac{\partial E_k}{\partial w_{pi}} &= x_{pi} \sum_M -\delta_m \frac{\partial net_m}{\partial net_p} \\ &= x_{pi} \sum_M -\delta_m \frac{\partial net_m}{\partial o_p} \frac{\partial o_p}{\partial net_p} \\ &= x_{pi} \sum_M -\delta_m w_{mp} o_p (1 - o_p) \end{aligned} \quad (3-68)$$

权值的更新为：

$$\Delta w_{pi} = -\alpha \frac{\partial E_k}{\partial w_{pi}} = -\alpha x_{pi} \sum_M -\delta_m w_{mp} o_p (1 - o_p) \quad (3-69)$$

可以看出，多层网络的权值训练时相当耗费计算的一个过程，在应用中，一般不使用超过四层的神经网络。尽管会带来计算复杂度上的增长，但是隐藏层中往往包含着许多十分有意义的信息，因此，多层网络的隐藏层设计也是一个研究的热点。

3.4 支撑向量机

支持向量机（Support Vector Machine, SVM）最早由 Vapnik 等人提出，是一种基于最大化边界原则的学习方法^[4-7]，后来 Vapnik 等人^[8]证明最大边界原则是为了最小化结构风险（测试误差的上界），这使得 SVM 学习方法具有深厚的理论根基。目前关于支持向量机的研究较多，它已广泛被用于分类、识别等问题中，受到机器学习、模式识别领域中学者的青睐。下面，本节主要就支持向量机的分类性能，依次介绍支持向量机的理论基础、线性支持向量机及非线性支持向量机。

3.4.1 线性支撑向量机

首先先看一个简单的二类分类问题，见图 3.10，假设在二维空间中有两类样本点，分别用圆点和三角形点表示两类样本点，分类问题意在如何将两类样本点分开。线性支持向量机的原理是基于边界最大化原则，该原则的核心思想是找到最优线性平面（若在高位空间中，称为线性超平面），该平面要尽量地将两类样本分隔开，并且使得两类样本集中最近样本的间隔最大。

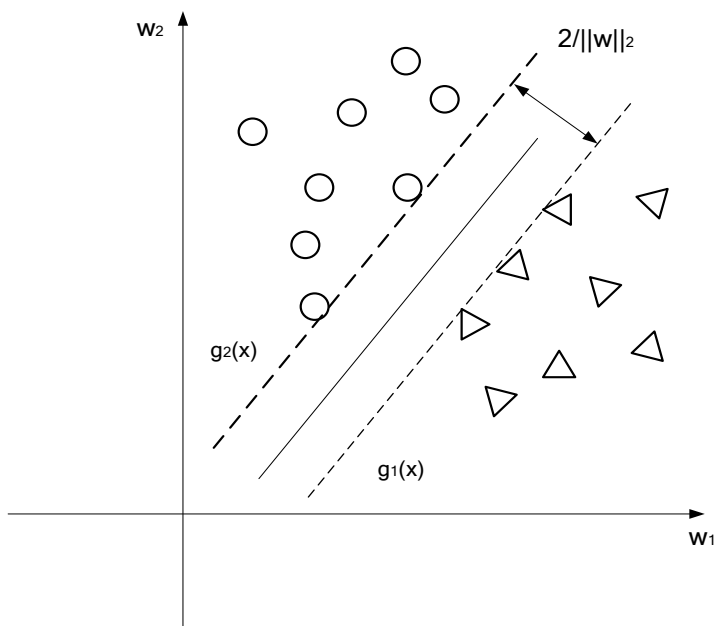


图 3.10 最大间隔原理

简单起见，先考虑线性可分的问题，设两类样本 $\{x_i, y_i\}$, $i=1, \dots, N$ ，其中 x_i 为第 i 个样本的特征向量， y_i 为它的类别标号，设线性超平面的方程形式为： $g(x) = w^T \cdot x_i + b$ 。 w 是线性超平面的法向量， b 是超平面的阈值。对线性可分的正反例样本，拟构造两个平行的线性超平面，通过调整线性超平面方程的阈值，可以要求对正例样本有 $g_1(x) = w^T \cdot x_i + b \geq 1$ ，对反例样本有 $g_2(x) = w^T \cdot x_i + b \leq -1$ 。这样两条平行超平面之间的“间隔”为 $2/\|w\|_2$ 。要使“间隔”最大，即要使 $\|w\|_2$ 最小，并要求正反例样本尽量地被正确分类，建立优化模型如公式 (3-70) 所示：

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i [(w^T \cdot x_i) + b] - 1 \geq 0 \quad i = 1, \dots, N \end{aligned} \quad (3-70)$$

模型 (3-70) 是一个凸二次规划，是连续可微的。因此，可以通过求解其对偶规划得到法向量 w 和阈值 b 的解析式。在实际中，样本的分布比较复杂且容易受到噪声的影响，使得其分布不是严格线性可分的，此时模型 (3-70) 是无解的，即不存在一条线性超平面将正反例样本

全部分对，因此需要对每个样本加入误差扰动项，改进后的优化模型如公式（3-71）所示。

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i [(w^T \cdot x_i) + b] \geq 1 - \xi_i, \\ \xi_i \geq 0, \end{cases} \quad i = 1, \dots, N \end{aligned} \quad (3-71)$$

在式(3-71)中： ξ_i 表示第 i 个样本被误分的程度。 C 是平衡误分程度与边界之间的惩罚因子。为方便求解，将上述凸规划模型转换为其对偶规划。可以通过引入 Lagrange 函数对原变量求偏导的方式，将原模型转换成对偶规划。针对模型（3-71），它的 Lagrange 函数如下：

$$L(w, b, \xi, \alpha_i, t_i) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^N \xi_i (C - \alpha_i - t_i) - \sum_{i=1}^N \alpha_i y_i (w^T \cdot x_i + b) + \sum_{i=1}^N \alpha_i \quad (3-72)$$

其中 α_i, t_i 为 Lagrange 系数。根据 Wolfe 对偶定理，公式(3-72)分别对 w, α_i, b 求偏导之后，然后将原变量用对偶变量进行表示，回代到优化模型中，就可以把上述问题转换为一个原模型的对偶问题，如式（3-73）所示。

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i x_j - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^N \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C, \end{cases} \quad i = 1, \dots, N. \end{aligned} \quad (3-73)$$

通过求解对偶规划 α_i ，再根据对偶变量 α_i 与原变量 w, b 之间的关系，可以求解到法向量 w 和阈值 b 。最终的线性分类超平面如下：

$$f(x) = w^T \cdot x_i + b = \sum_{i=1}^N \alpha_i y_i (x_i \cdot x) + b \quad (3-74)$$

3.4.2 非线性支撑向量机

若样本分布是非线性的，虽然上述的线性支持向量机算法可以得到一个线性超平面，但是

强行使用线性分类器训练非线性分布的样本，得到的识别率和分类率会大大降低。因此，需要设计一个能够处理非线性情况的分类器。

SVM 理论认为：对于低维空间中不是线性可分的样本，可以将样本映射到一个高维空间中，在高维空间中样本可以认为是近似线性可分的，仍然可以使用线性超平面去分类。但是一般样本的特征向量本身已经处于高维空间中，再对样本进行升维会极大增加运算费用。**SVM** 理论巧妙的采用了“核函数”（**Kernel function**）来解决这一问题。核函数是样本间的某种内积形式，在高维空间中它反映的是样本的相似程度。在线性 **SVM** 的对偶规划中，存在样本内积的二次项形式，核函数是一种内积形式的度量，所以对偶规划中的样本内积可以合理地由某种核函数来代替，这样即使对样本升维，在高维空间中的内积计算量没有比原空间中优化模型的计算量多很多。

从上述介绍可知，核函数之所以受到极大的欢迎是因为其内积性质既度量了样本的相似性，又减少了计算费用。下面从数学上正式引入核函数的表达形式。首先，定义核函数[3]之前，需要引入矩阵算子和半正定性等概念。

$X = \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_m}\}$ 是一个有限点集，令 $K = (K(\overline{x_i}, \overline{x_j}))$ 是一个 $m \times m$ 的矩阵。

定义 4.1 （矩阵算子）：对一个 m 维的向量 $u = (u_1, u_2, \dots, u_m)^T \in R^m$ ， $Ku \in R^m$ ，并且其分量由下式表示：

$$[Ku]_i = \sum_{j=1}^m K(\overline{x_i}, \overline{x_j}) u_j \quad i = 1, 2, \dots, m \quad (3-75)$$

则称 $K = (K(\overline{x_i}, \overline{x_j}))$ 为一个矩阵算子。

定义 4.2 （半正定性）：如果对于任意的一个 m 维向量 $u = (u_1, u_2, \dots, u_m)^T \in R^m$ ，和一个矩阵算子 K 若有：

$$u^T Ku = \sum_{i,j=1}^m K(\overline{x_i}, \overline{x_j}) u_i u_j \geq 0 \quad (3-76)$$

则称矩阵算子 K 是半正定的。

定义 4.3 （核函数）：若一个矩阵算子 K 是半正定的，则存在从 $X = \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_m}\}$ 到

R^m 的映射 ϕ ，使得 $K(\bar{x}_i, \bar{x}_j) = (\phi(\bar{x}_i) \cdot \phi(\bar{x}_j))$ ， $i, j = 1, 2, \dots, m$ ，其中 (\cdot) 是 Hilbert 空间 R^m 上的内积， $K(\cdot, \cdot)$ 是被称为一个核函数。

从上述定义可知，核函数其实是在 Hilbert 空间上的一种内积运算，因此选取不同的映射 ϕ ，然后在其上定义内积运算即可形成核函数。目前常见的核函数有以下几种：

$$\text{多项式核函数: } K(x, \bar{x}) = ((x \cdot \bar{x}) + c)^d \quad \text{其中 } c \geq 0 \quad (3-77)$$

$$\text{径向基核函数: } K(x, \bar{x}) = \exp(-\sigma \|x - \bar{x}\|^2) \quad \text{其中 } c \geq 0 \quad (3-78)$$

$$\text{傅立叶核函数: } K(x, \bar{x}) = \frac{1 - q^2}{2(1 - 2q \cos(x - \bar{x}) + q^2)} \quad 0 < q < 1 \quad (3-79)$$

$$\text{Sigmoid 核函数: } K(x, \bar{x}) = \tan(c(x \cdot \bar{x}) + d) \quad \text{其中 } c \geq 0, d < 0 \quad (3-80)$$

在引入核函数之后，相应地，将 SVM 的线性优化模型（3-73）式中样本的内积 $x_i \cdot x_j$ 替换为核函数的形式，便可得到非线性支持向量机在样本升维之后的对偶规划，这个对偶规划和原来的样本单独升维到高维空间，再使用 Lagrange 函数求解出来的对偶规划一致。同理，通过求解对偶规划得到 SVM 的对偶变量，从而获得在高维空间中的线性超平面的法向量和阈值。非线性支持向量机的优化模型如下：

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^N \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C, \quad i = 1, \dots, N. \end{cases} \end{aligned} \quad (3-81)$$

通过优化求解对偶变量 α_i 能够得到高维空间中，线性超平面的法向量 w 和 b 。最后，非线性支持向量机的判别函数如下：

$$f(x) = w^T \cdot x_i + b = \sum_{i=1}^N a_i y_i K(x_i, x) + b \quad (3-82)$$

3.4.3 经验风险与结构风险

考虑一个分类问题，如何建立一个决策函数 f （这里称决策函数为“假设”），使得其能够更好地处理分类问题和具有更好的推广能力。下面本节引入衡量一个决策函数推广能力的标准。这个标准，学者们很自然地会联想到分类误差。但是是否在训练样本集上的分类误差率越低越好呢？伴随着这些问题，下面需要引进几个概念[3]，它们分别是损失函数、结构风险、经验风险、VC 维和学习算法在概率意义下近似正确（PAC）等。

设样本集 $X \subset R^n$ ， $Y = \{-1, 1\}$ ， (x, y) 是随机向量，其中 $x \in X, y \in Y$ ，记函数 $P(\bar{x}, \bar{y}) = P(x \leq \bar{x}, y \leq \bar{y})$ 为 $X \times Y$ 上的概率分布。损失函数是评价预测准确程度的一种度量，根据某个假设的结果定义损失函数。

定义 4.4（损失函数）：记 $(x, y, f(x)) \in X \times Y \times Y$ 是一个三元组，这里 $f(x)$ 是一个预测值，若映射 $c: X \times Y \times Y \rightarrow [0, \infty)$ 对任意的 $x \in X, y \in Y$ ，都有 $c(x, y, y) = 0$ ，则称 c 是一个损失函数。一种比较常用的损失函数是 0-1 损失函数，其定义如下：

$$c(x, y, f(x)) = \begin{cases} 0 & \text{如果 } y = f(x) \\ 1 & \text{其他情况} \end{cases} \quad (3-83)$$

定义 4.5（结构风险）：设 $P(x, y)$ 为 $X \times Y$ 上的概率分布， $c: X \times Y \times Y \rightarrow [0, \infty)$ 为给定的损失函数，那么假设 $f(x)$ 关于概率分布 $P(x, y)$ 的结构风险是指：

$$R[f] = E[c(x, y, f(x))] = \int_{X \times Y} c(x, y, f(x)) dP(x, y) \quad (3-84)$$

定义 4.6（学习算法在概率意义下近似正确，PAC 原则）：如果对任意给定实数 $\delta (0 < \delta < 1)$ 和 $\varepsilon (0 < \varepsilon < 1)$ ，对 $X \times Y$ 上的任意概率分布 $P(x, y)$ ，存在 l_0 ，使得对于长度大于等于 l_0 的按分布 $P(x, y)$ 选取的独立同分布的训练集 $T = \{(x_1, y_1), \dots, (x_l, y_l)\}$ ， $l \geq l_0$ ，由学习算法得到的 $f(x)$ 若满足 $P\{R[f] \leq \varepsilon\} \geq 1 - \delta$ 成立，则称算法在概率意义下近似正确。

因此可以看出，若使学习算法能够在概率意义下近似正确，那么则寻找一个决策函数让结

构风险比较小。如果已知损失函数和样本的概率密度分布 $P(x, y)$ ，当然可以通过最小化结构风险求得决策函数。但是在现实应用中，样本的分布非常复杂，它们的概率分布通常是未知的，已知的只是训练集 T ，因此想通过最小化结构风险这个标准来构造决策函数是不可能的，所以必须通过别的方式去评价决策函数的好坏。接下来，本节引入经验风险的概念。

定义 4.7 (经验风险): 设任意给定训练集 $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$ ，并且给定损失函数 c ，所谓决策函数 $f(x)$ 的经验风险是指：

$$R_{emp}[f] = \frac{1}{l} \sum_{i=1}^l c(x_i, y_i, f(x_i)) \quad (3-85)$$

所以衡量一个学习算法好坏的直观想法是，对决策函数集进行一定范围的限制，然后在这个范围内寻求使得经验风险最小的假设 $f(x)$ 。那么是不是一味的追求经验风险最小会得到最优的 $f(x)$ 使得它对于样本的结构风险也最小呢？也即会使得假设函数 $f(x)$ 对于测试样本效果非常好呢？这个答案是否定的。因为结构风险不是单调的。那么结构风险与经验风险之间的关系。

首先介绍 VC 维^[5]的概念：VC 维是统计学习理论中比较重要的一个概念，VC 维最初由 Vapnik 和 Chervonenkis 提出的，意在评判一种学习算法（决策函数）的表达能力。假设集 F 是由决策函数组成的集合，VC 维是建立在点集被 F 打散的基础上。下面引入一个点集被 F “打散”的概念。

定义 4.8 ($N(F, Z_m)$): 设 F 是一个假设集，它的定义域是 $X \subset R^n$ ，值域是 $\{-1, 1\}$ 。令 $Z_m = \{x_1, \dots, x_m\}$ 是由 X 中 m 个点组成的集合。考虑当决策函数 f 取遍 F 中所有可能的假设，产生一个 m 维向量 $(f(x_1), \dots, f(x_m))$ 。记 $N(F, Z_m)$ 为上述 m 维向量中不同向量的个数。

定义 4.9 (Z_m 被 F 打散): 设 F 是一个假设集，如果 $N(F, Z_m) = 2^m$ ，则称 Z_m 被 F 打散。

定义 4.10 (VC 维): 定义假设集 F 的 VC 维为: $VC(F) = \max\{m : N(F, Z_m) = 2^m\}$ ，如果 $\{m : N(F, Z_m) = 2^m\}$ 是一个无限集时，定义 $VC(F) = \infty$ 。

举个简单的例子来说，一个 2 维平面上，对于任意不共线的 3 个点，一条直线的 VC 维是 3。

推广到一般情况，对于一个线性分类面 $y = w^T x + \theta$ ，其中 $w \in R^n$ 是分类面的法向量， θ 是分类面的常数项（阈值），则线性分类面的 VC 维 $= \dim(w) + 1 = n + 1$ 。

下面的定理证明了 VC 维与结构风险（测试误差）之间存在如下关系。

定理：令 F 的 VC 维 h ，若 $N > h$ ，且 $h(\ln(2N/h) + 1) + \ln(4/\delta) \geq 0.25$ ，则对于任意概率分布 $P(x, y)$ 和任意的 $\delta \in (0, 1]$ ， F 中的任意假设 f 都可以使得下列不等式至少以 $1 - \delta$ 的概率成立。

$$R[f] \leq R_{emp}[f] + \sqrt{\frac{h(\log(2N/h) + 1) - \log(\eta/4)}{N}} \quad (3-86)$$

从上述定理可以得知，结构风险 $R[f]$ 的上界是经验风险 $R_{emp}[f]$ 与 VC 维之和的某个函数，其中不等式 (3.11) 右边的两项之和也被称为结构风险。而此处的经验风险即对应着训练误差，结构风险对应着测试误差。当经验风险（训练误差）一定时，最小化 VC 维，会导致结构风险变小，因此会使得测试误差的上界变小，这样有利于决策函数的推广能力。

由于 VC 维较难定量刻画，后来 Vapnik 等人证明最大化边界 $1/\|w\|_2$ 也可以间接实现 VC 维最小化，从而使得最小化结构风险的上界。自此，支持向量机的理论逐渐完善，在实际应用中受到广泛欢迎。

3.5 Boosting 方法

Boosting 由 Freund 和 Schapire 于上个世纪 90 年代提出，该算法是一种提升算法，意在将多个“弱学习”算法组合提升为一个“强学习”算法。在 Boosting^[9]法中，首先根据已有的训练样本集设计一个分类器，然后依次地添加多个分量分类器，最后形成一个总体的分类器，它对训练样本集的准确率能够任意的高。

针对两类样本问题，举例说明 Boosting 算法是如何产生分量分类器的。首先，从大小为 n 的原始样本集 D 中随机抽取 n_1 个样本点，组成样本集 D_1 。根据样本集 D_1 ，训练得到第一个分类

器记为 g_1 。现在继续构造第二个样本集 D_2 ，该样本集应该根据分类器 g_1 得到的最富信息的样本点组成。更明确地说， D_2 中一半的样本应该能被 g_1 正确分类，而另一半则是被 g_1 错分。具体的构造方式如下：取 D 中剩余的样本点，并逐个地送入 g_1 进行分类，保留被 g_1 正确分类的正例样本，保留通过 g_1 但被其错分的反例样本，这样得到的集合 D_2 便是根据 g_1 得到的最富信息的样本构成。同理构造第三个训练样本集 D_3 时，其构造方式类似于 D_2 的构造，只不过这次挑选样本点的时候，要用 g_1 和 g_2 进行分类，若样本通过 g_1 和 g_2 的判别，即被认为是“正例”（包括虚假正例）的反例样本，那么就把该样本加入集合 D_3 ，再用 D_3 训练第三个分类器 g_3 。

Boosting 算法中的典型代表算法分为两大类，Majority Boosting 和 Adaboost 算法。Adaboost 在目标的检测和识别中应用很广泛，后来的 Logitboost 算法是在 Adaboost 算法基础上的一种改进，本节将分别介绍 Adaboost 和 Logitboost 算法的原理。

3.5.1 Adaboost 算法

Adaboost^[10-13] 算法是 Boosting 系列算法的一种，意在将弱学习算法提升为强学习算法。Adaboost 算法中的弱学习算法也称为弱分类器，强学习算法称为强分类器。它采用加权投票机制，将一些有投票权的弱分类器线性组合起来形成强分类器。如何确定哪些弱分类器有投票权呢？Adaboost 采用贪婪的迭代方式，每次迭代选择一个最好的弱分类器，最后将所有迭代中选择出的弱分类器进行线性加权组合，形成强分类器。

在每次迭代中，Adaboost 对每个训练样本赋予一个权重，这样在每次迭代中所有样本的权重形成一套概率分布。每次迭代中选择分类误差最小的弱分类器参与到强分类器的构建中，并调整每个训练样本的权重。权重调整的原则是更加重视被误分的样本，被误分的样本获得较大的权重，被正确分类的样本权重减少。这样，随着迭代次数的增加，权重的增加，算法会更加关注难以训练的样本上。最后，根据每次迭代选择出的弱分类器的加权组合，从而形成强分类器，并且弱分类器的权重为其在训练集上的权重。

一个弱分类器对应一个特征，在选择哪些弱分类器形成强分类器的同时，也即完成了特征

选择的功能,因此 Adaboost 算法可以同时实现特征选择与分类器的设计。一个弱分类器 $h_j(x)$ 对应一个特征 f_j , 一个阈值 θ_j , 还有一个用来指示不等式符号方向的函数 p_j , 其表示形式如下:

$$h_j(x) = \begin{cases} 1 & \text{if } (p_j f_j(x) < p_j \theta_j) \\ 0 & \text{otherwise} \end{cases} \quad (3-87)$$

对于每一个特征,相应的弱分类器学习得到一个最佳的分类函数,使得训练样本的错误分类数达到最小。换句话说,在当前权重分布的情况下,每个弱分类器需要训练一个最佳的阈值 θ_j , 使得这个弱分类器对所有训练样本的分类误差最低。阈值的确定方式如下:

对于第 j 个特征 f_j , 计算所有训练样本关于该特征的特征值,并将其排序,通过观察特征值的分布,可以为这个分类器确定一个最优的阈值。具体来说,对排好序的表中每个特征值,分别计算下面四个值:

- 1) 全部正例样本的权重之和 W^+ ;
- 2) 全部反例样本的权重之和 W^- ;
- 3) 在此特征值之前的正例样本的权重之和 S^+ ;
- 4) 在此特征值之前的反例样本的权重之和 S^- ;

这样,当选取当前特征值和它前面的一个特征值之间的实数作为阈值时,这个阈值所带来的分类误差为:

$$e = \min (S^+ + (W^- - S^-), S^- + (W^+ - S^+)) \quad (3-88)$$

于是,通过对所有排好序的特征值扫描一遍,选择分类误差最小的那个阈值作为弱分类器的最佳阈值。

表 3.1 Adaboost 算法流程

- 给一组训练样本 $(x_1, y_1), \dots, (x_n, y_n)$, 其中, y_i 为样本 x_i 的类别标号

- 分别将正例样本和反例的权重初始化为 $\omega_{1,i} = \frac{1}{2m}, \frac{1}{2l}$,

这里的 m 和 l 分别为反例样本和正例样本的数量

- For $t=1, \dots, T$:

1、归一化权重,

$$\omega_{t,i} \leftarrow \frac{\omega_{t,i}}{\sum_{j=1}^n \omega_{t,j}}$$

2、对于每一个特征 j , 训练一个弱分类器 h_j 都要严格的遵守使用一个特征的原则.

计算每个弱分类器的误差:

$$\varepsilon_j = \sum_i \omega_i |h_j(x_i) - y_i|.$$

3、选择本次迭代中误差 ε_t 最小的那个弱分类器.

4、更新权重:

$$\omega_{t+1,i} = \omega_{t,i} \beta_t^{1-e_i}$$

其中, 如果样本被正确分类, 则 $e_i = 0$;

否则, $e_i = 1$, 并且 $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.

- 最终的强分类器为:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & otherwise \end{cases}$$

$$\alpha_t = \log \frac{1}{\beta_t}.$$

其中

3.5.2 Logitboost 算法

Logitboost 方法^[14]是 Friedman 等人于 2000 年提出的, 该算法是在 Adaboost 基础上进行的

改进，其基本思想与 Adaboost 算法相似，都是基于现有样本数据集构建一系列的“弱分类器”，反复调用该“弱分类器”，通过计算每个样本的相应值和权重拟合弱分类器，经过多轮循环，最后采用组合的方法将各轮的弱分类器构成强分类器，从而得到较高精度的预测模型。其与 Adaboost 算法的区别主要在于，Adaboost 算法的权重更新函数是指数型，Logitboost 算法是采用的负对数似然更新函数。相对于优化指数型的 Adaboost 算法，Adaboost 算法更具有鲁棒性，可以显著地降低训练误差的同时提高泛化能力，并改善 Adaboost 算法存在的过拟合以及对噪声敏感等问题。

Logitboost 算法的具体过程如下：首先考虑二分类问题，样本 x 属于类别 1 的概率记为

$$p(x) = e^{F(x)} / (e^{F(x)} + e^{-F(x)}), \text{强分类器 } F(x) = \sum_{k=1}^K f_k(x) \text{ 是弱分类器的 } f_k(x) \text{ 组合, 其中 } f_k(x) \text{ 是}$$

使用牛顿迭代法最小化负的对数似然函数求得。Logitboost 算法的求解 $f_k(x)$ 的核心是使用最小

二乘回归拟合样本点加权响应的值。算法的具体步骤如下： $p(x_i) = 1/2$ 。

1、开始分配样本权重 $w_i = 1/N, i = 1, 2, \dots, N$ ，概率密度，初始的分类函数 $F(x) = 0$

2、循环迭代 $m = 1, 2, \dots, M$

(1) 计算所有样本点的响应值和权重：

$$\begin{aligned} z_i &= \frac{y_i - p(x_i)}{p(x_i)(1 - p(x_i))} \\ w_i &= p(x_i)(1 - p(x_i)) \end{aligned} \quad (3-89)$$

(2) 使用加权的样本点响应值拟合弱分类器 $f_m(x)$ 。

(3) 更新判别函数 $F(x) = 0$ 和概率密度 $p(x_i) = 1/2$ 。

3. 得到最后的强分类器 $F(x) = \sum_{m=1}^M f_m(x)$ 。

3.6 参考文献

- [37] Stauffer C and Grimson W. Adaptive background mixture models for real-time tracking, in Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 1999: 246~252,
- [38] Moigne J L, Campbell W J, and Crompt R F. An automated parallel image registration technique based on the correlation of wavelet features. Proc. of IEEE Trans. on Geoscience and Remote Sensing, 2002(40): 1849~1864.
- [39] X Fan, Rhody H and Saber E. Automatic registration of multisensor airborne imagery. Proc.in Applied Imagery and Pattern Recognition Workshop, 2005 (34): 6.
- [40] Papageorgiou C and Poggio T. A trainable system for object detection, International Journal of Computer Vision, 2000(38): 15~33.
- [41] Yang Y. Pedersen J O. A comparative study on feature selection in text categorization, in Proc of the 14th International Conference on Machine Learning, 1997: 412~420.
- [42] Stoppiglia H, Dreyfus G, Dubois R and Oussar Y. Ranking a random feature for variable and feature selection. Journal of Machine Learning Research, 2003(3): 1399~1414.
- [43] Tibshirani R. Regression selection and shrinkage via the lasso. Royal Statistical Society, 1996: 267~288.
- [44] Ron K and George H. J. Wrappers for feature subset selection. Artificial Intelligence journal, Special Issue on relevance, 1997(97): 273~324.
- [45] Pati Y C, Rezaifar R and Krishnaprasad P S. Orthogonal, Matching pursuit: Recursive function approximation with applications to wavelet decomposition. Proc. of IEEE 27th Annual Asilomar Conference on Signals, Systems and Computers. Pacific Grove, 1993: 40~44.
- [46] Collins M, Schapire R.E and Singer Y. Logistic Regression, AdaBoost and Bregman Distances, Machine Learning, 2002(48): 253~285.
- [47] Shimizu H and Poggio T. Direction Estimation of Pedestrian from Multiple Still Images, Proceedings of the IEEE Intelligent Vehicles Symposium, 2004: 596~600.

- [48] Liu Y Z, Shan S G, Zhang W C, Chen X L and Gao W. Granularity-tunable gradients partition (GGP) descriptors for human detection. Proc in IEEE International Conference on Computer Vision and Pattern Recognition, 2009: 1255~1262.
- [49] George J, Robert K and Pfleger K. Irrelevant features and the subset selection problem. In: Cohen W W, Hirsh H, Eds. The Eleventh International Conference on Machine Learning. 1994: 121~129.
- [50] Friedman J, Hastie T and Tibshirani R. Additive logistic regression: A statistical view of boosting. Ann. Statist, 2000(28): 337~407.

第 4 章 目标检测典型应用

人类所感知的外界信息 80% 以上来自于视觉。人类似乎天生具有这样一种能力，就是从纷繁复杂的客观世界中，迅速而准确地找到他(她)所感兴趣的物体，例如街道、人脸、行人、汽车、文字、建筑物以及其它日常生活中的各种物体等。在视觉感知领域，这种能力被概括成为物体检测。物体检测是人类视觉中的基本步骤和基本功能，为人类了解周围的环境和景物提供了至关重要的前提。

让计算机或者机器人具有包括物体检测在内的视觉感知能力是人类多年以来的一个科学梦想。为了达到这个目标，有两种途径。第一种途径是仿生学方法，尽管科学家们从生理学、心理学、神经学等各方面对人类视觉系统进行了观察和探索，并且对视觉的物理和化学性质以及人类视觉系统的组织已经积累了很多有价值的知识，但是视觉过程的机理至今还是一个谜。第二种途径是工程方法，主要思路就是利用计算机来模拟人类的视觉功能。科研工作者用各种计算理论和工程方法在该领域内进行了大量的研究，他们沿着这个方向所取得的研究进展(包括各种计算机视觉技术的原理、算法和系统等)，促进了物体检测领域和计算机视觉学科的发展。

基于图像处理和模式识别技术的目标检测在 21 世纪随着各项新兴技术的发展得到了不断完善，这为以图像理解为基础的目标检测进入实际应用提供的良好的契机。目标检测成功的在一些场景中得以实际应用，目前的主流应用包括：人脸检测，人体检测，车辆检测，以及字符类目标检测。

4.1 人脸检测

过去的几十年内，人脸检测由于其基本的挑战和潜在应用已经在计算机视觉领域得到了广泛的研究，特别是最近基于安全因素的考虑。一些相关的应用主要包括：视频监控，人机交互，人脸识别和人脸数据库的管理等。

迄今为止，机器视觉技术的发展已经历了一个较为漫长的旅程。它最初只是应用于生产流水线，简单的模拟人的视觉来执行一些繁琐、重复性的任务。经过研究者们不懈努力，今天

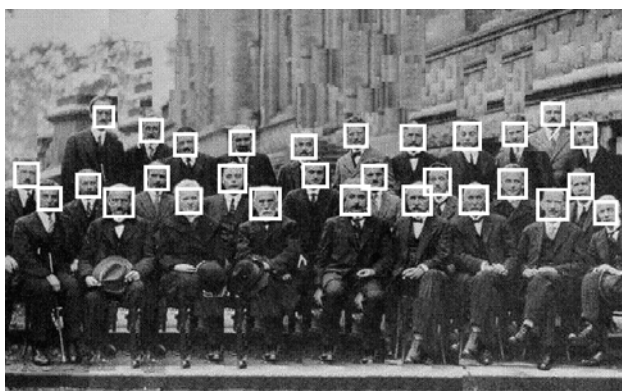
已经可以利用它实现更有效、友好、自由的人机交互界面，并且随着人脸研究的进一步发展，更可以让计算机实现通过观察一个人以及这个人的表情变化等做出相应的反映。近几年电子商务的蓬勃发展对安全性提出了更高的要求，这也推进了人脸研究的发展和应用，最终将促使人类进入了一个全新的人机世界。

计算机视觉和模式识别领域的人脸研究主要包括人脸识别和人脸检测的研究。最初人脸研究主要集中在人脸识别领域，而且早期的人脸识别算法都是在认为已经得到了一个正面人脸或者人脸很容易获得的前提下进行的。但是随着人脸应用范围的不断扩大和开发实际系统需求的不断提高，这种假设下的研究不再满足需求。人脸检测开始作为独立的研究内容发展起来。目前，国内外的文献中所涉及的人脸检测算法已经有很多种，许多重要的国际会议和期刊也都涉及到人脸检测问题的研究。人脸检测也开始广泛应用到全新的人机界面、基于内容的检索、数字视频处理、视觉监测等许多领域。

人脸检测是一个复杂的具有挑战性的模式检测问题，其主要的难点有两方面，一方面是由于人脸内在的变化所引起：(1)人脸具有相当复杂的细节变化，不同的外貌如脸型、肤色等，不同的表情如眼、嘴的开与闭等；(2)人脸的遮挡，如眼镜、头发和头部饰物以及其他外部物体等。另一方面是由于外在条件变化所引起：(1)由于成像角度的不同造成人脸的多姿态，如平面内旋转、深度旋转以及上下旋转等，其中深度旋转影响较大；(2)光照的影响，如图像中的亮度、对比度的变化和阴影等；(3)图像的成像条件，如摄像设备的焦距、成像距离，图像获得的途径等等。这些困难都为解决人脸检测问题造成了难度。



(a)



(b)

图 4.1 人脸检测的示例

4.1.1 人脸检测数据库

人脸图像数据库对于算法的研究、训练和测试是不可或缺的。在算法训练阶段，所采用的人脸库的规模大小、光照条件、表情、姿势等在很大程度上影响着算法的精度和鲁棒性。在算法测试时所用到的人脸数据库的规模同样决定了实验设计的合理性和结果的有效性。以下介绍一些在人脸识别领域内常用的人脸数据库。

FERET 人脸数据库

1993 年美国国防部高级研究项目署和美国陆军研究实验室成立了 FERET(Face Recognition Technology)项目组，建立了 FERET 人脸数据库[1]，用以评价人脸识别算法的性能。该数据库包含了 14051 幅多姿态、不同光照条件下的人脸图像，是人脸识别领域应用最广泛的人脸数据库之一。

PIE 人脸数据库

在 2000 年的 10 至 12 月期间，该数据库由卡耐基梅隆大学(Carneige Mellon University)建立了 PIE(Pose, Illumination and Expression)人脸数据库[2]，库内包含了 68 位志愿者的共 41368 幅人脸图像。其中每人包括 13 种姿势，43 种光照条件以及 4 种表情下的图片。

AR 人脸数据库

该数据库是由普渡大学(Purdue University)的计算机视觉中心于 1998 年创建[3]，包括 126 人在不同光照、表情、遮挡和老化的彩色图像近 4000 幅。

AT&T 人脸数据库

该数据库由英国剑桥大学 AT&T 实验室创建^[4]，其中包含 40 人共 400 张面部图像，部分志愿者的图像包括了姿态、表情和面部饰物的变化。该人脸库在人脸识别研究的早期经常被人们采用，但由于变化模式较少，多数系统的识别率均可达到 90%以上。

XM2VTS 多模态生物特征数据库

该数据库包括 295 人在 4 个月时间内 4 次录制的人脸和语音数据，每次采集都包括 2 个头部分旋转视频片断和 6 种不同声音和视频片断，其中包括了高质量和彩色图像和 3D 模型^[5]。

MIT 人脸数据库

该数据库由麻省理工大学多媒体实验室(MIT Media Laboratory)创建于 1989 年, 包含 16 位志愿者的 2592 幅不同姿势、光照和大小的人脸图片^[6]。

Yale 人脸数据库

该数据库由耶鲁大学计算机视觉与控制中心创建, 包含 15 位志愿者共 165 幅灰度人脸图像, 每人各有 11 幅图像, 包含光照、表情、姿态以及脸部饰物的变化^[7]。

Yale 人脸数据库 B

该数据库包含了 10 个人的 5760 幅人脸图像, 包括 9 种姿势, 64 种不同光照下采集的正面人脸图像, 其中的姿势和光照变化的图像都是严格控制的条件下采集的, 主要用于光照和姿态问题的建模与分析^[8]。

BioID 人脸库

该数据库包含 23 位志愿者的 1521 张正面灰度图像, 图像中的人大部分为欧美人, 每张照片的分辨率为 384286。人脸图像在不同季节、地点被采集, 光线、背景、人脸大小均有较大变化^[9]。

UMIST 人脸数据库

该数据库由曼彻斯特大学创建, 包括 20 位志愿者的共 564 张正面和侧面照片。该数据库考虑了人种、性别、外表等的变化^[10]。

4.1.2 人脸检测的一般计算模型

通常所说的人脸检测是基于光学图像定位出现的人脸具体位置和大小简称。光学人脸图像(以下简称人脸图像)是外界光源(包括太阳光、室内人造光源和其他物体表面反射而来的光线)发出的光线照射在人脸上, 经人脸表面反射后传播到摄像机传感器的光线强度的度量^[11]。

一般来说, 一幅人脸图像是一个 3D 的人头或者是一幅 2D 人脸图像在一定光照条件下的 2D 投影。人脸图像的变化可由许多因素导致, 具体来说, 这一成像过程实际上涉及到三大类关键要素:

1、人脸内部属性

包括人脸组件(眼睛, 鼻子和嘴)的基本形状, 人脸表面的反射属性(如反射系数等, 通常简称为纹理[Texture])、人脸 3D 形状(表面法向量方向)、以及人脸表情、胡须等属性的变

化。

2、 外部成像条件

包括光源（位置和强度等）、其他物体（比如眼镜、帽子）或者人体其他部件（比如头发）对人脸的遮挡，头部姿态等。

3、 摄像机成像参数

摄像机成像参数包括摄像机位置（视点）、摄像机的焦距、光圈、快门速度，成像设备的畸变等内外部参数。其中第一个要素是人脸成像的内因，第二、三个要素是人脸成像的外因。因此，光学人脸图像的成像过程可以简单地形式化为：

$$I = f(F; L; C) \quad (4-1)$$

其中，函数 f 表示成像函数， F 、 L 和 C 分别表示人脸内部属性、外部成像条件和摄像机成像参数这三类要素， I 为生成的人脸图像。如果进一步假设：人脸皮肤的反射属性满足朗伯（Lambertian）模型，人脸为凸表面结构，光源为无穷远处的单色点光源，上述成像公式可以进一步改写为：

$$I(x, y) = f_c(\rho_{(x,y,z)} \vec{n}_{(x,y,z)} \cdot \vec{s}_{(x,y,z)}) \quad (4-2)$$

其中， (x, y, z) 表示人脸表面的一点 P 的三维坐标； $\rho_{(x,y,z)}$ 表示 P 点的表面反射率； $\vec{n}_{(x,y,z)}$ 表示 P 点的表面法向量方向； $\vec{s}_{(x,y,z)}$ 表示光源的方向和强度； f_c 则表示摄像机的成像函数；而 $I(x, y)$ 则为摄像机最终输出的对应 P 点的图像像素的强度。

不难理解，在上述成像过程中，人脸表面 3D 结构及其反射属性才是人脸相对稳定的本质属性；而人脸表情变化、有无胡须等属于人脸内部属性，它们并不改变人脸区别于非人脸的本质属性。光源等外部成像条件以及摄像机参数等外部因素就更不能作为人脸检测所依赖的特征属性。

综上所述，在理想情况下，要描述人脸样本的不变性，基于统计学习的方法就应该从人脸图像表现中统计出人脸稳定不变的本质属性，馈入到后端的判别分类器中进行判断。设输入图像为 I ，这一过程可以形式化为以下三个步骤：

1、 属性分离。分离人脸本质属性要素与光源条件 s^* 、摄像参数 c^* 等外部参数要素：

$$(\rho_{(x,y,z)}^*, \vec{n}_{(x,y,z)}^*, s^*, c^*) = f_c^{-1}(I(x, y)) \quad (4-3)$$

2、 **特征提取**。从人脸属性要素中提取能够体现人脸的特征：

$$F^* = T(\rho_{(x,y,z)}^*, \vec{n}_{(x,y,z)}^*) \quad (4-4)$$

其中，T 表示特征提取过程。

3、 **分类判别**。根据图像中输入子窗口 w_i 提取的特征，判断该子窗口的模式：

$$y = g(F^*, w_i) \quad (4-5)$$

其中 $g(\cdot)$ 表示统计学习得到的分类器， $y \in \{+1, -1\}$ ，其中+1 表示输入的子窗口 w_i 为人脸，否者为非人脸。

4. 1. 3 人脸检测的常用特征

4. 1. 3. 1 基于肤色特征的检测

过去若干年中对于肤色检测问题的研究有很多，这为人脸检测中使用肤色特征打下了基础。如 Michael J. Jones 等人基于大量的样本图像对肤色检测问题进行了统计分析，比较了用直方图模型和混合高斯模型进行肤色检测的情况；B. Martinkauppi 对 Skin locus 的理论和方法在不同光照条件下的肤色分布问题作了细致的分析，他指出在特定的摄像机条件下，各种光照条件下的肤色分布可以用两个二次或者多次的多项式来完全描述，这为解决肤色检测的光照问题提供了一些依据。



图 4. 2 肤色检测

(a)黄色光照条件下的人脸图像；(b) 从图(a)中检测到的肤色区域（白色部分）；(c)光照补偿后的图像；(d) 从图(c)中检测到的肤色区域

肤色用于人脸检测时，可以采用不同的建模方法，主要有高斯模型、高斯混合模型以及非参数估计等。利用高斯模型和混合高斯模型在不同颜色空间中建立肤色模型来进行人脸检测。skin locus 建立肤色模型，提取彩色图像中的人脸面部区域实现人脸检测，该方法能够处理多种光照的情况，但是算法需要在固定摄像机参数前提下才有效。Dorin Comaniciu 使用非参数的核函数概率密度估计法来建立肤色模型，并使用 mean shift 方法进行局部搜索实现了人脸的检测和跟踪。该算法提高了人脸的检测速度，对于遮挡和光照也有一定的鲁棒性。但是这种方法和其他方法的可结合性不是很高，用于人脸检测时，处理复杂背景和多个人脸时存在困难。为了解决光照问题，可以针对不同光照进行补偿，然后再检测图像中的肤色区域，见图 4.2。这样解决了彩色图像中偏光、背景复杂和多个人脸的检测的问题，而且对图像中人脸的色彩、位置、尺度、旋转、姿态和表情等具有不敏感性。使用肤色和形状信息来定位人脸和提取面部特征的方法，首先在 HSV 颜色空间中进行颜色分割以确定肤色区域，然后在粗糙的分辨率下进行区域增长以连接各肤色区域。对于每个连通区域，用几何矩的方法拟合出最佳的椭圆，并将那些与椭圆相近的连通区域作为候选的人脸。接下来，在这些候选的区域中进行面部特征的搜索以确认它是否为人脸。

4.1.3.2 基于边缘特征的检测

利用图像的边缘特征检测人脸时，其计算量相对要小的多，可以用来提高检测速度。大多数使用边缘特征的算法都是基于人脸的边缘轮廓相似于椭圆的特性，如提取边缘特征，利用建立的椭圆模板进行椭圆匹配来检测人脸。也有人对椭圆模板进行了改进，采用一个椭圆环模型来代替椭圆，并加入了边缘的方向特征，实现了简单背景的人脸检测。但是利用椭圆特性进行计算时，椭圆的判断又增加了整个检测过程的计算量，抵消了边缘特征数量少的一些优势，阻碍了速度的提高。

B. Fröba 采用了基于 EOM (Edge-Orientation Matching) 的方法，利用边缘强度信息建立一个 EOM 模板，在检测时进行模板匹配，他建立的一个模板（见图 4.3）。虽然该算法在复杂背景下误检率比较高，但是与其他的特征相融合后获得了很好的效果。

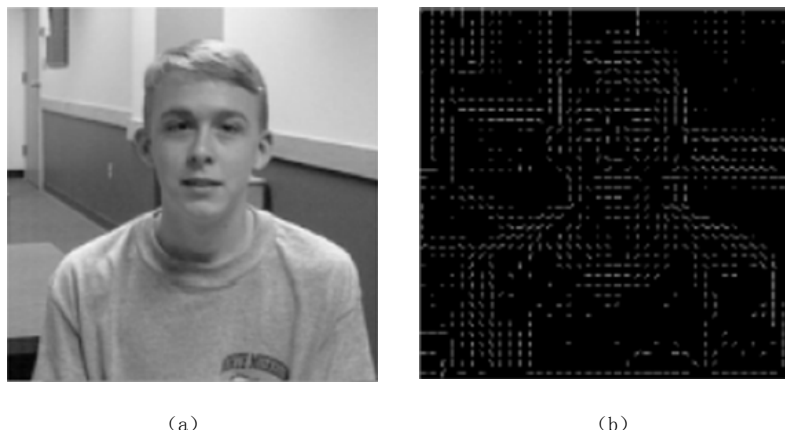


图 4.3 图像边缘方向的向量场示例

一般说来，人脸识别的过程主要包括三个方面：预处理、特征提取和识别分类器的设计。其中，人脸特征的提取是对人脸的另一种表述，即在低维空间内对原来的图像，也就是原来的高维空间中的人脸进行描述并建模，目的就是将这种用于描述的语言从图像中提取出来。所以，特征的提取是人脸识别的关键，提取出的有效特征不仅可以更好地为后续的分类器设计服务，也可以提高整个识别的精确度和识别率。但是由于人脸模型是一种非刚体，其中包括表情等的变化，外界中有视角、光照等的影响，这些变化使人脸图像的类内离散度远大于类间离散度，以至于很难有效地提取人脸的特征向量。因此，人脸特征提取也是人脸识别的主要难点所在，尤其是对于工作在不同光照条件下的人脸识别系统。所以，注重鲁棒性同时兼顾效率的人脸特征提取技术一直是研究的热点。

4.1.4 实例分析

图 4.4 中的人脸检测系统级联了两个人脸检测器，第一个是采用 AdaBoost 算法学习得到的检测器，第二个是采用 SVM 算法学习得到的检测器。这里将 AdaBoost 算法作为第一级主要是因为该算法计算速度快，而且该算法可以达到和其他算法相当的性能；采用 SVM 算法作为第二级是因为在采用 AdaBoost 算法学习的过程中，最后总有一些人脸和非人脸模式难以区分，而且其检测的结果中存在一些与人脸模式并不象的子窗口。实验发现，那些 AdaBoost 算法难以区分的人脸和非人脸模式对于 SVM 来说却相对容易区分（这是因为这些样本可能位于人脸和非人脸样本的边界，属于难以区分的样本。但是 SVM 却正是利用边界样本（支持向量）来实施分类，因此在这里采用 SVM 正是对 AdaBoost 分类器一个合理的补充。），而且采用 SVM 作为第二级也可

以去除 AdaBoost 误判断为人脸模式的非人脸子窗口。因此,在这种意义上来说,级联 AdaBoost 和 SVM 组成的分类器可以提高人脸检测系统的性能。

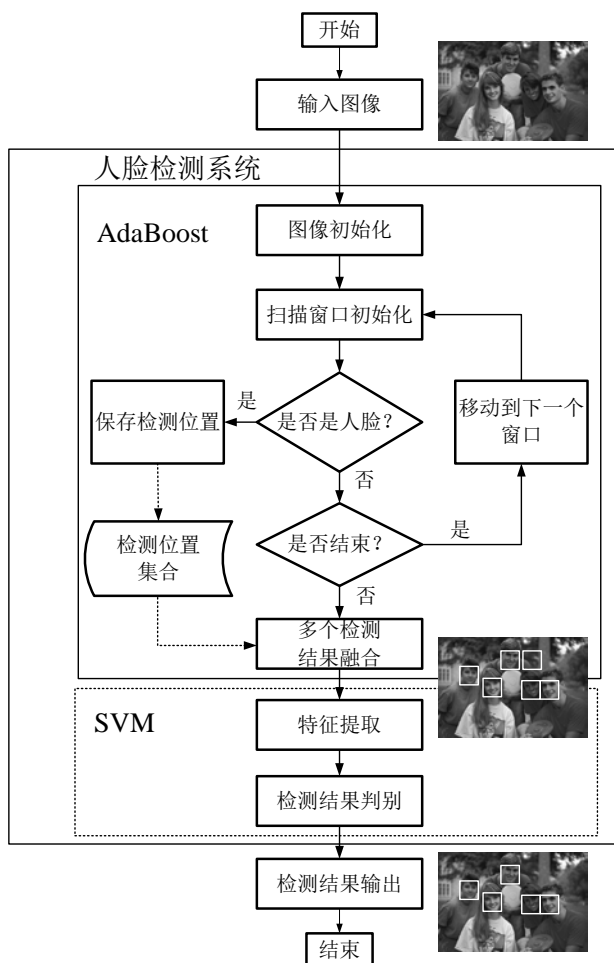


图 4.4 人脸检测系统框图

如图 4.4 所示,人脸检测器工作方式:首先将输入的图像初始化,主要是计算输入图像的积分图,其中需要对图像进行多次缩放,最终缩放得到的图像大小和训练的人脸样本相等(20×20),并对所有缩放的图像计算积分图。其次对输入检测器的子窗口进行初始化,主要包括对每个输入子窗口的方差归一化(这样可以在一定的程度上减弱光照的影响)和输入子窗口的特征值计算。然后将这些子窗口的特征值送入 Adaboost 分类器进行计算以判断该窗口是否为人脸窗口,并对于那些 Adaboost 判断为人脸的子窗口保留其大小和位置信息,在 Adaboost 扫描完一幅图像后,对所有判断为人脸的子窗口进行合并。接着,对于合并后得到的窗口提取特

征并送入 SVM 进行验证，以排除那些不可能是人脸模式的字窗口。最后输出 SVM 验证后的人脸子窗口的参数（包括该窗口的大小和位置信息）作为人脸检测系统的输出。图 4.5 给出了上述流程下的人脸检测的结果。

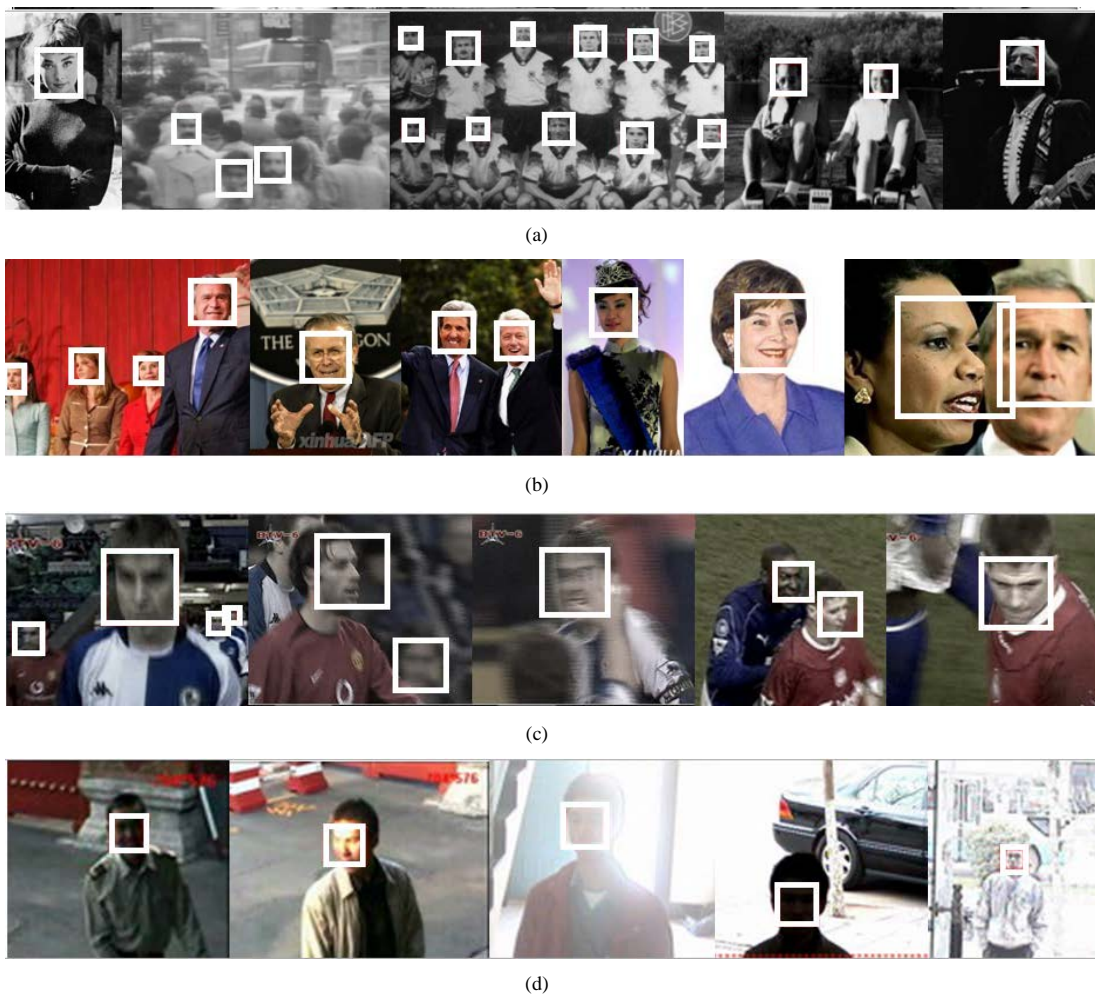


图 4.5 一些检测到的输入图像中的人脸的结果。

图像主要来自于(a) MIT+CMU 正面人脸测试集。(b) 网络上下载的图像。(c) 体育视频。(d) 视频监控

4.2 人体检测

在国防领域，可靠的人体目标检测可以应用于电子探测设备自动识别危险目标，帮助特种部队及空降兵来探测周边区域的情况并保卫军营的安全，防止偷袭和破坏等方面；在民用领域，该技术可用于重点建筑（核电站、水电站、水库、银行、重要桥梁等）及重要地点（国家重点保护建筑、博物馆、体育场馆、地铁站、监狱等）的安全保卫工作；在智能交通领域，可以及时监控交通事故并可以在十字路口、新建高速路端等进行自动行人检测和安全预警。随着城市规模的扩大，住宅小区逐渐成为安全防范的重点，结合红外/近红外等视频设备，该技术可增强小区安全监控、减少犯罪，从而增强居民的安全感。

4.2.1 人体检测数据库

表 4.1 列出了现有的一些主要行人检测数据库。可以看出根据图像获取装置的不同，所有的数据库大致可以分为两类：静态图像和视频。INRIA 数据库自 2005 年发布以来，一直备受关注，推动了行人检测技术的快速发展。中国科学院研究生院模式识别与智能系统实验室（Pri-SDL）也发布了自己的行人数据库，供研究者训练和测试。PASCAL VOC 2007 数据库是近年来目标检测领域最具影响力和挑战性的数据库，最优秀的行人检测算法在这个数据库上的成绩也不理想，说明了行人这一类可形变的目标检测还有很大的研究空间。INRIA 数据库中行人目标的分辨率相对较高，而通过车辆上的移动摄像机拍摄的数据库目标相对较小，行人高度一般处在 50-100 个像素之间，从一个侧面反映了低分辨率情况下行人检测的迫切需求，尤其是针对移动行人检测平台的应用。

表 4.1 行人检测公开数据集

数据集	训练集	测试集	注解
INRIA	2416(64*128)	1132(64*128)	1) 彩色图片
People Dataset ^[12]	/ 1218(含标定)	/ 453(含标定)	2) 背景复杂
Caltach	250,000 帧, 350,000 bounding boxes		1) 驾驶员视角
Pedestrian Dataset ^[13]	2300 行人（标定）		2) 视频帧序列

TUD-Brussels Pedestrian Dataset ^[14]	508 幅 640*480 的图片，共 1326 个行人	多尺度、多视角的行人
VOC2007 dataset ^[15]	正样本 4096 /负样本 3003	4528 行人
TUD Multi-view Pedestrians Dataset ^[16]	4732 幅彩色图片（含标定数据）	250 幅彩色图片
SDL Pedestrian Dataset ^[17]	正面 1000 幅 / 侧面 3050 幅 / 多视角 7550 幅	140 幅彩色图片 /258 幅多 视角图片

- 1) 彩色图像
- 2) 行人检测最具挑战和影响
- 1) 图片尺寸不一
- 2) 含部位标定
- 3) 共 8 个视角
- 包含正面/侧面/多视角
64*128 像素（训练）

4.2.2 人体检测常用特征

在人体检测的常用特征除了第二章中介绍的 Haar 特征和 HOG 特征之外，在这些广泛使用的特征上提出的新特征也得到了不错的效果。这里我们主要介绍 v-HOG 特征和 MSO 特征，它们都是典型特征上的扩展。

4.2.2.1 v-HOG 特征

v-HOG(variable-size HOG)特征^[18]是基于 HOG 特征演变出来的一种变尺度块状特征。HOG 特征是固定位置、固定尺度块上的梯度方向直方图特征，v-HOG 特征是由变尺度、非固定位置块上的梯度方向直方图组成的。我们根据特征选择与分类器的结果，挑选出哪些 v-HOG 块可以用来表示人体目标。v-HOG 块的长宽比一般为 1:1、1:2、2:1，块的大小从 12x12 变化到 64x128 的像素区域，所有块仍然像 HOG 特征那样是由 2x2 个 cell 组成，每个 cell 里面统计的是一个 9 维的梯度方向直方图，每个块中 cell 的大小随着块的大小而改变，但是无论 cell 大小如何变化，仍然是往 9 个方向进行投影，因此每个 cell 得到 9 维梯度直方图的向量，相应地每个 v-HOG 特征块的特征向量是 4x9=36 维。图 4.6 显示了 v-HOG 特征的提取过程。

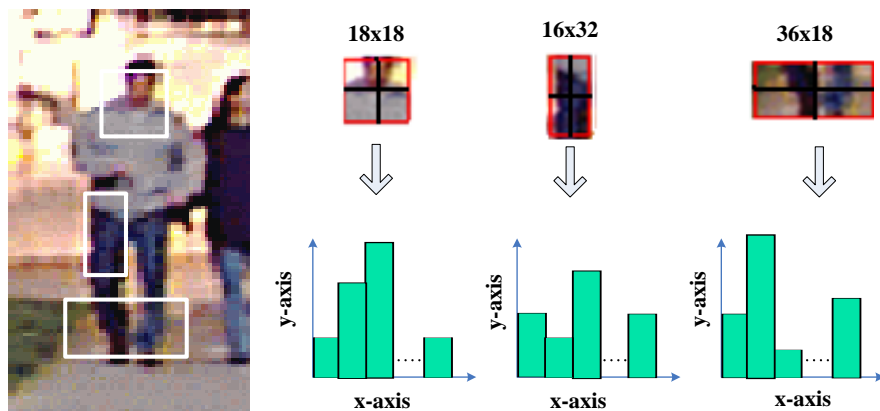


图 4.6 v-HOG 特征

4.2.2.2 MSO 特征

MSO (Multi-scale orientation) 特征^[19]是受到 Haar-like 特征与 HOG 特征的启发,融合了两种特征的优点,能够更好地捕捉人体模式的变化和提高了计算效率。MSO 特征分为单方向、多方向两种形式。单方向 MSO 特征由尺度不同、位置各异的块组成,它以图像的颜色特征为基础,通过对特征块像素的颜色值进行累加计算,得到块的方向特征用以描述块的整体方向。多方向 MSO 特征结合了 HOG 的思想,对块内所有像素的梯度方向进行统计得到一个 8 维特征向量。多方向 MSO 特征可以很好地把握图像的整体信息,而且由于其块的维数小于 HOG 特征块的维数,所以具有很快的运算速度。

MSO 特征使用 8 个尺度扫描训练图像,分别得到在 8 个尺度下的正方形特征块,见图 4.7(a)。在某个尺度上单方向 MSO 特征的特征值计算,可分为两步:首先,计算正方形区域的梯度方向;其次,将梯度方向映射成固定的特征值编码。将正方形区域的划分成水平和竖直两个方向的四个子区域,分别计算两种划分下的水平灰度差 dx 和竖直灰度差 dy 。水平灰度差 dx 的计算,首先需要将正方形分为左右两个等份,分别计算两个小区域的所有像素的灰度值之和,得出水平灰度差;再将正方形分为上下两等份,用同样的方法计算出竖直灰度差,见图 4.7(b)。

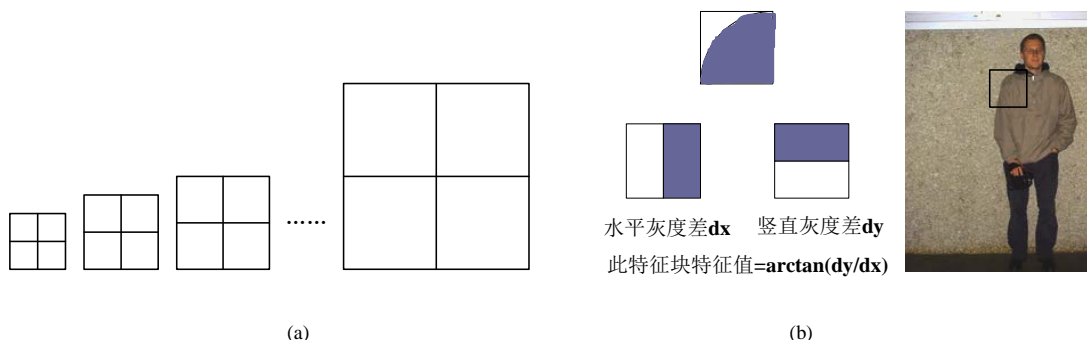


图 4.7 MSO 特征

(a) 多尺度的特征窗口及其划分; (b) 单方向 MSO 特征的提取

然后, 根据所求出的 dx 和 dy 计算出这个特征区域内的梯度方向 Ori_Rect :

$$Ori_Rect = F(\arctan(dy/dx)) \quad (4-6)$$

其中, $F(\cdot)$ 为一个离散的映射, 取值在非负整数集合 $\{0, 1, \dots, 8\}$ 中。它等价于将范围在 $0^\circ \sim 360^\circ$ 度的每个梯度方向 $\arctan(dy/dx)$ 分别向 9 个方向进行投影。当 $\arctan(dy/dx)$ 的值在 $0^\circ \sim 20^\circ$ 或 $180^\circ \sim 200^\circ$ 之间时, $F(\arctan(dy/dx))$ 的值为 1; 当 $\arctan(dy/dx)$ 的值在 $21^\circ \sim 40^\circ$ 或 $201^\circ \sim 220^\circ$ 之间时, $F(\arctan(dy/dx))$ 的值为 2, 依此类推, 可以得到所有块的梯度方向。

相对于单方向 MSO 特征, 多方向 MSO 特征主要是对所选的特征区域, 使用一个 8 维的直方图作为此区域的特征值。首先利用水平和垂直方向的 Prewitt 算子 $[-1, 0, +1]$ 分别计算特征区域内像素点 (x, y) 的水平和竖直灰度差 dx 和 dy , 然后计算区域中每个像素位置的梯度方向, 最后将所有像素点的梯度方向投影到 8 个方向上形成直方图向量。

4.2.3 实例分析

基于分类的目标检测框架是为了对图像中的单个区域进行分类而建构起来的。这种目标检测框架可以分成两个部分: 训练和检测。训练的目的是为了创建出一个二值分类器, 用于区分图像中的某个区域 (或窗口) 是否是所检测的目标。而检测的任务是对测试图像中的检测区域

进行密集的扫描，并利用分类器对扫描结果进行决策。最后，将所有决策融合成最终的检测结果。训练和检测都包括三个步骤，如图 4.8 所示。这就构成了一个目标检测的通用框架。最后的检测结果取决于分类器的准确度和可依赖度，同时也取决于最后结果的融合情况。

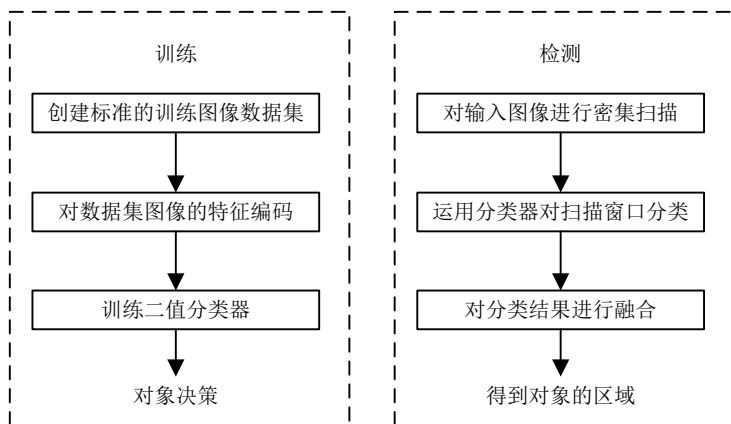


图 4.8 检测通用框架

训练的第一步是创建训练样本集。训练的正样本是目标分布在正中的图像窗口。而负样本是从不含任何正例目标的图像进行随机抽样生成的。分类器就是从这样的样本集合中训练得到的。理想的情况下，每个正样本应该只包含一个正例目标，正样本的尺寸应该略大于检测窗口的尺寸。

这种简单的窗口框架具有很多的优点。它可以兼容传统的分类器进行检测，并且使传统分类器更加灵活，在窗口的位置和尺寸改变时，不影响分类器的效果，我们称之为位置不变性和尺寸不变性（虽然还有一些其他的转换也不影响分类器的效果，但是姿态、视角和光照等因素不在此范围之列）。其次，它的优点还在于分类器能够工作在相关的坐标系中（如：与当前窗口中心相关的特征位置），这使得一些相关的刚性特征集能够被使用。在另一方面，这也意味着分类器是运行在大量的窗口之上的，这需要花费大量的计算资源，同时也使得整体的检测结果对正样本的错误率非常的敏感。实际上，一个 640×480 像素的图像中，包含了 10,000 个窗口，所以正样本错误率的有效范围通常是单位窗口 10^{-4} 以下。

基于 SVM 的人体检测算法的简单流程见图 4.9。主要包含六个步骤：颜色空间标准化，梯度计算，空间和方向上的梯度统计，对比度标准化，特征向量生成以及基于 SVM 的分类器的训

练。其中前五个步骤的作用是从输入图像中提取 HOG 特征，这个过程不仅可以用于分类器的训练，也可以用于分类器的检测，是整个算法的核心部分。下面详细介绍每一步的算法原理以及整个流程的算法实现。

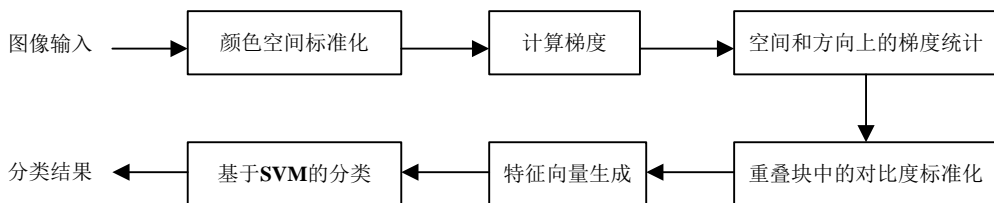


图 4.9 基于 HOG 特征和 SVM 的人体检测流程

颜色空间标准化 (Gamma 标准化)

我们评估了图像像素点的几种不同表示方法，包括灰度信息、RGB 和 LAB 颜色空间信息，并选择性地进行了幂处理。当可用时，颜色信息会有所帮助。比如，在人体检测中，RGB 和 LAB 颜色空间信息对结果的影响大致相同，而使用灰度信息时可以将每万个单位窗口的错误正检测率 (10^{-4} FPPW, False Positives Per Window) 降低 1.5%。

平方根和 log 压缩是两种不同方式的 Gamma 标准化。这两种方式都需在颜色空间的每个通道上进行。通过 log 压缩可以判断出图像的信息形式形成是一个乘法过程，所以可以断定从反曝光率到曝光率，log 因素对照明影响的变化是很缓慢的。同样 CCD 检测器的光子噪音与光照的平方根成正比，所以平方根标准化能够有效抑止噪音的影响。对大多数目标类来说，平方根压缩对检测效果的提高都能够有一个很小的保证。这个保证之所以小是因为后面的特征归一化也能够达到相似的效果。对人体检测器来说，平方根压缩能够在很小的程度上（每万个单位窗口的错误正检测率 (10^{-4} FPPW) 降低 1%）提高检测效果。然而 log 压缩的作用力太强了而使得检测效果降低了 2% (10^{-4} FPPW)。经验证明，平方根压缩对于目标包括自行车、摩托车、汽车、公交车，以及人的表现更加优越。而对于包括动物在内一些以颜色变化为主的目标，如猫，狗，马等，未经标准化的 RGB 信息要更适合一些，然而对于牛和羊来说，平方根压缩依旧要更加优秀。这也许就如同一个典型的图像背景和目标本身的问题一样，难以区分。

梯度的计算

我们采用一种离散的平滑模板对图像进行可选的高斯平滑并评价它们的区别。对于颜色图像（RGB 空间或 LAB 空间）来说，我们首先分别计算每个颜色通道的梯度，然后选出最规范的一个作为像素的梯度向量。在 Dalal 等人的论文中，对各种尺度的平滑包括 $\sigma = 0$ 进行了评估。计算梯度时的平滑会严重破坏检测的效果。对于高斯导数，将 $\sigma = 0$ 逐渐增加到 $\sigma = 2$ 将会使错误率提高大约一倍，从 11% 到 20% (10^{-4} FPPW)。导数模板的测试包括了多个一维的点导数模板、

3×3 的 Sobel 模板和 2×2 的对角矩阵模板 $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ 和 $\begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}$ （这两个是最简洁的中心二维导数模板）。其中，简单的模板 $(-1, 0, 1)$ 得到的效果是最佳的。采用诸如立方修正、3×3 的 Sobel 模板或者 2×2 的对角矩阵模板的平滑方式会系统地降低效果。不同类型导数运算的比较如表 4.2 所示。

表 4.2 不同梯度模板对检测结果的影响

面具类型	一维中心	一维非中心	一维立方修正	2×2 对角	3×3Sobel
操作符	$[-1, 0, 1]$	$[-1, 1]$	$[1, -8, 8, -1]$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ $\begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$
错误率 (10^{-4} FPPW)	11%	12.5%	12%	12.5%	14%

总体说来，检测器的效果对于梯度计算的方式是非常敏感的，而最简单的一维中心的 $(-1, 0, 1)$ 模板在 $\sigma = 0$ 情况下取得的效果最佳。使用其他形式的平滑或者更大更复杂的模板都会损坏最终的检测效果。造成这种结果的最可能的原因是图像中的归整信息对目标检测是非常重要的：从根本来说，图像是基于边缘的，而平滑会降低边缘信息的对比度，从而减少图像中的信号信息。一个有用的推论就是最佳的图像梯度应该能够快速且简单地计算得到。

空间和方向上的梯度统计

这一步是描述符的基本非线性化。每个像素为梯度方向贡献一个加权的值，这个方向是指梯度元素中心的方向。接下来将一个局部空间内所有像素点的梯度加权值累积到所有的梯度方向上。这个局部空间被称之为“单元格”。这个单元格可以是矩形，也可以是圆形。而梯度方向

就平均地分布在单元格的空间里，对于无符号的梯度值分布空间为 $0\sim 180$ ，而对于有符号的梯度值分布空间为 $0\sim 360$ 。为了避免混淆，梯度权值将在方向和位置上三角插值到相邻地两个方向中心上。权值是像素的梯度量级的一个函数，无论是量级本身，量级的平方，量级的平方根，还是量级的省略形式都能够反应像素上的边缘信息。根据 Dalal 等人论文的测试结果，采用量级本身得到的检测效果最佳，使用量级的平方根会轻微降低检测结果，而使用二值的边缘权值表示会严重降低效果（约为 5% 个单位 10^{-4} FPPW）。

对于所有的目标分类，好的效果很大程度上取决于优良的方向编码，而空间梯度方向上的插值可以相对粗糙一些。实验证明，关于人体数据集，对于无符号梯度值，即空间分布在 0 和 180 之间的，增加空间中梯度方向的数量可以大大提高分类效果，而方向数量大于 9 时，效果的改变不再明显了。而对于有符号的梯度值，即使将方向数量翻倍，分类效果也会降低。可以推断的是，对于人体目标来说，因为衣服和背景颜色的变化范围太广，使得梯度的符号变得无足轻重了。但在一些其他目标检测器中，梯度值的符号却扮演着很重要的角色，如对于行人、汽车、自行车、摩托车等目标的检测。

重叠块中的特征标准化

由于图片中局部的曝光率和前景—背景的对比度的多样化，梯度值的变化范围也非常广。因此，有效的局部对比标准化对于好的运行效果至关重要。标准化的方式有很多种，大多数的原理都是将一组单元格放到一个更大的空间块中，然后在每个空间块中分别进行标准化。实际上，块基本上都是互相重叠的，所以每一个单元格的标量会在不同的块中计算多次，并且以不同的值出现在最终的特征向量中。这样虽然看上去很累赘，但是包含重叠块的信息能够大大地改善分类结果，这对于好的标准化很重要。

特征空间中块的选择

在单元格大小和块大小不同的情况下，检测的错误率（单位为： 10^{-4} FPPW）会有所变化。对于人体目标检测，块的大小为 3×3 个单元格，单元格的大小为 6×6 个像素时，检测效果是最好的，错误率约为 10% (10^{-4} FPPW) 左右。块的大小为 2×2 个单元格，单元格大小为 8×8 个像素时，也相差无几。 $6\sim 8$ 个像素宽的单元格， $2\sim 3$ 个单元格宽的块，其错误率都在最低的一个平面上。块的尺寸太大时标准化的作用被削弱了从而导致错误率上升，而如果块的尺寸太小时，有用的信息反而会被过滤掉。

在将梯度累加到相邻的方向之前，通过对空间窗口进行高斯处理，从而降低块边缘像素的权值，可以改进检测效果。当采用 $\sigma = 0.5 \times$ 块宽的高斯处理时，检测效果可以提高 1% (10^{-4} FPPW)。

与此同时，我们也可以采用不同的块对特征进行编码，这种不同包括块尺寸的不同和块中包含的单元格的尺寸的不同。如果在特征中只使用 3×3 和 2×2 的单元格（单元格尺寸为 $\eta = 8 \times 8$ ）的话，效果的改善很微小，而使用多种不同尺寸的单元格（ $\eta = 8 \times 8$, $\zeta = 2 \times 2$ 和 $\eta = 4 \times 4$, $\zeta = 3 \times 3$ ）能够提高约 3% (10^{-4} FPPW) 的检测效果。然而， $\eta = 8 \times 8$, $\eta = 8 \times 8$, $\zeta = 2 \times 2$ 和 $\eta = 4 \times 4$, $\zeta = 4 \times 4$ 在块尺寸为 16×16 的情况下，对结果的影响却不那么明显。所以，多种块的编码方式应该同时考虑单元格和块尺寸的大小。当然，这样也会有所代价，与之对应的特征向量也会因此而增加，所以它更适用于一些特征选择机制，如 Adaboost，以避免对过大的特征向量进行编码。 $\zeta = 2 \times 2$ 和 $\eta = 4 \times 4$, $\zeta = 4 \times 4$ 在块尺寸为 16×16 的情况下，对结果的影响却不那么明显。所以，多种块的编码方式应该同时考虑单元格和块尺寸的大小。当然，这样也会有所代价，与之对应的特征向量也会因此而增加，所以它更适用于一些特征选择机制，如 Adaboost，以避免对过大的特征向量进行编码。

图 4.10 给出了实验中采用的训练集合，包括正例图片（上一行）2416 幅和反例图片（下一行）约 3000 幅，样本尺寸为 64×128 像素。

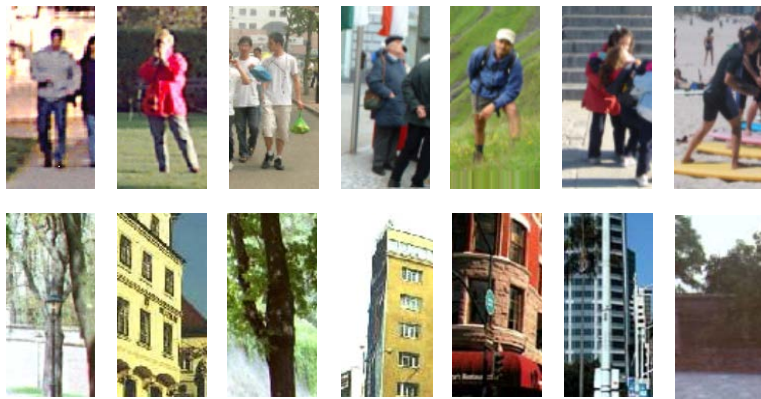
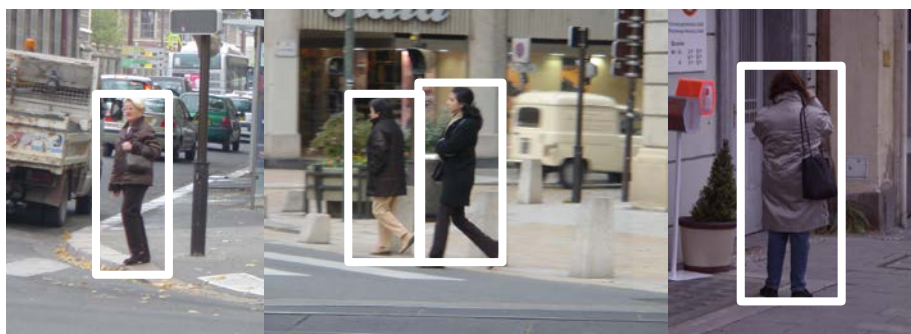


图 4.10 训练中的正例和反例样本

实验效果



(a)

(b)

(c)



(d)

(e)



(f)

(g)

图 4.11 部分检测结果

图 4.11 中列举了一些具有代表性的图片，包含了背景复杂、姿态多变的人体样本。通过选择合适的特征描述和分类器，基本解决了直立人体的多视角和多姿态的检测问题。尽管如此，

在某些视角的人体检测方面仍然不够完美，还存在着一定问题。在图 4.11 中，当人体目标比较多的时候，或者当人与人之间有比较多的相互遮挡时，就会出现一些错误的窗口。这说明目前的一些特征描述和分类算法还并不可以彻底解决人体检测中的视角问题，算法还有待进一步的提高。尤其对于一些人体形变特别大的目标，比如：蹲着、趴着人体目标往往被分类器漏检。而这些也是目前人体检测研究的热点和难点。

4.3 车辆检测

车辆交通事故给司机和行人带来了越来越多的危害，统计事故表明司机面临的主要危险来自路面的其他车辆。因此，有必要开发智能交通系统（Intelligent Transportation Systems，简称 ITS）来解决这些问题，减少交通事故的发生。智能交通系统是二十世纪六十年代提出，在九十年代发展起来的，它是本世纪交通发展的方向，是新时代下科技和社会发展推动下的产物。目前，世界各国根据本国国情制定了相应的交通发展计划，确定了本国的交通发展重心。美国将投资 6000 亿美元用于 ITS 的研究和开发利用，欧共体也将投资 1000 亿美元制定庞大的 ITS 未来五年发展计划，亚洲国家如日本、韩国、新加坡等也将大力发展 ITS 技术，我国也加入了研究和开发利用 ITS 的行列，已赶上发达国家 ITS 研究和开发水平，建立我们自己的智能交通产业。

近年来，作为智能交通系统的一部分，车辆辅助驾驶系统受到了研究者的很大关注，这些系统致力于提醒驾驶者驾驶的周围环境。在这样的系统中，有效的车辆检测，特别是路面车辆检测，是很重要的一个研究内容，它为智能交通系统提供了必要的数据来源。

在车辆检测中使用的传感器有两种类型：主动式和被动式。

激光测距仪、毫米波雷达、电磁感应线圈等为主动式传感器，使用主动式传感器进行车辆检测，算法实现简单，性能较好，但是仍然存在一些不足，如分辨率较低，传感器之间互有干扰，成本昂贵，以及可能带来的环境问题等。

CCD 摄像机以及目前在军事、民事领域广泛使用的红外热象仪则属于被动式传感器。CCD 摄像机价格便宜，并且能够提供丰富的视频资料信息，通过采用计算机视觉技术对视频资料进行分析，在车辆行驶路线发生改变（如转弯）时，可以实现更有效的检测与跟踪，而且丰富的视频信息可以用来进行相关的应用，如道路检测、交通标识识别，以及行人、障碍物的检测与识

别等，但是最初的性能不稳定。随着计算机硬件和软件技术、图像处理技术、计算机智能技术的发展，解决了许多以往道路交通监视控制中和图像处理技术的许多难题，克服了存储量小、计算速度慢和模型泛化能力的局限性，提高了处理速度和检测精度，逐步满足了实际系统实时性、安全性、可靠性的要求。因此，被动式传感器得到了广泛的应用，基于计算机视觉的车辆检测方法也成为人们研究的重点。

4.3.1 车辆检测数据库

MIT-CBCL 车辆检测数据库

该数据集是由麻省理工学院 MIT-CBCL 实验室建立的，包含了不同的形状、颜色、视角，并且有局部遮挡和光照变化的图片。车辆的视角分为正视角和后视角，每幅图片由原始图片归一化到 128x128 像素大小，且车辆位置居中，共 516 幅车辆图片。

参考：<http://cbcl.mit.edu/software-datasets/index.html>

Pri-SDL 车辆检测数据库

该数据集是由中国科学院研究生院工程教育学院模式识别与智能系统实验室 Pri-SDL 建立的，包含了 1528 幅训练集图片，164 幅测试集图片。车辆训练集图片视角同样包括正视角和后视角。

参考：<http://coe.gucas.ac.cn/SDL-HomePage/resource.asp>

4.3.2 车辆检测的一般方法

基于计算机视觉的车辆检测系统包括两个基本步骤：假设产生阶段和假设验证阶段。在假设产生阶段，图像中可能存在车辆的区域被假设出来；在假设验证阶段，采取一些测试措施来验证图像中的车辆区域^[20]。

假设产生的方法大体分为三类：基于知识的方法，基于立体视觉的方法和基于运动信息的方法。基于知识的方法利用先验知识如对称性、颜色、阴影、角点和水平垂直边缘、纹理、车灯等来确定车辆在图像中的初始位置。该方法简单、直观，易于编程实现，但需要估计多个经验阈值，例如车辆长宽的经验比值、车辆边缘的最小长度、车辆阴影与道路的灰度差异阈值等。经验阈值的准确与否，直接关系着系统性能的优劣。基于立体视觉的方法建立在视差或频差理论基础上，运用两个或多个摄像机对同一景物从不同位置成像获得立体像对，通过各种算法匹

配出相应像点，从而恢复深度（距离）信息，并使用这些信息利用逆投影变换^[21]来估计车辆和其他障碍物在图像中的位置。但是这种方法要求对摄像机进行正确标定，这是非常困难的，因为不可避免地要受到车辆运动或天气等的影响。基于运动的方法主要利用序列图像之间存在的大量相关信息进行车辆的初始假设。主流算法有帧差法和光流法。

假设验证的方法主要有两类：基于模板的方法和基于外观的方法。基于模板的方法首先建立车辆类在不同情况下的各种模板，然后计算输入图像和模板之间的相关性，通过对相关性的分析来完成验证。基于外观的方法从大量的训练数据中学习车辆类的特性，首先，每一个训练样本都用一组局部或全局特征来表示，然后，通过训练一个分类器来学习到车辆和非车辆的决策边界，最后，利用该决策边界来对一个新输入的图像进行判断。对于训练数据，通常需要大量不同外观的车辆作为正例样本和许多不同的非车辆图像作为反例样本。

特征提取方法和分类器的选择是基于外观的检测方法的核心问题。PCA（Principal Component Analysis，简称 PCA）^[22]，局部方向编码（Local Orientation Coding，简称 LOC）^[23]，Haar 小波特征^[24]，梯度方向直方图特征（Histograms of Oriented Gradients，简称 HOG）^[25]都已经被用于特征表述。常用的基于统计的分类器包括神经网络（Neural Network）^[22-24]，支持向量机（Support Vector Machines，简称 SVMs）和 Adaboost 分类器等^[24]。

在目标检测过程中，为目标物体选择一个适当的表述方法的最终目的就是这个表述方法可以产生较大的类间散度，同时最小化类内散度^[24]。由于同一物体内部也存在一定的差异性，因此，为了对一个物体类的结构进行编码，必须在一个特定的分辨率上来识别物体的特征，在这个特定的分辨率上，物体内部具有一定的一致性，同时可以抑制噪声。在某种分辨率下无法发现的特性在另一种分辨率下将很容易被发现，由此有人提出了多分辨率理论，而小波变化就是多分辨率特性的一种实现，因此，越来越多的研究者开始通过小波变换对图像进行分析和表述。

Haar 小波是小波的一种，通过将图像进行 Haar 小波变换，将图像由像素空间变化为小波系数空间，产生一系列不同尺度上的小波系数，从而形成一个完备的特征字典，形成对图像的描述。它使用不同尺度的小波特征对图像进行了多分辨率的描述，获取了图像不同层次上的细节特征。字典中的特征集可以反映图像局部区域在几个不同方向上的灰度差。Papageorgiou 等在行人检测中使用两个尺度上的小波函数对图像进行 Haar 小波变换，并在所有的小波系数中选择能够表现类内一致性的系数作为简单的特征。后来，Viola 等人又对基本的 Haar 特征进行了拓展，实现了对图像更精细的描述，该特征在人脸检测上取得了好的效果。

Gabor 小波是一种被高斯函数调制的带有特殊频率和方向的正弦波。它可以在捕获图像空间频率结构的同时保留空间关系信息，因此，它适用于提取方向依赖性较强的模式信息。最初，它只用于一维的信号分解。自从 20 世纪 80 年代开始，由于生物学方面发现二维 Gabor 滤波器和视觉外皮神经细胞之间的相似性，大量的研究者开始在计算机视觉应用中使用 Gabor 滤波器作为空间和卷积滤波器。Gabor 滤波器可以成功地应用于不同的图像分析领域包括边缘检测，图像编码，纹理分析，手写数字识别，人脸识别，车辆检测和图像检索。尽管 Gabor 滤波器在模式识别系统中已广泛应用，但是它的主要应用领域是特征提取。它能够同时获取空间和频率信息，优于传统信号表达只能获取空间或频率信息的一种。Sun 等使用 Gabor 滤波器的幅值响应作为特征用于车辆检测。也有研究者结合遗传算法(Genetic Algorithms)来选择适当的 Gabor 滤波器，然后利用幅值响应的矩做为特征，提高了检测性能。

直方图描述子已经被广泛地用于表述、分析和识别图像，因为它计算简便有效，并且对噪声和图像局部变化具有一定的鲁棒性。Dalal 提出使用梯度方向直方图描述子用于人体检测。然而，由于一定程度上的不变性，HOG 特征对于一般的目标识别不可能同样有好的效果。同时，直方图描述子丢失了物体的结构信息，对于很多应用来说，这种描述是不够精确的。因此，为了在不变性和敏感性两个方面达到一个平衡，多分辨率的直方图被提出来。通过将图像与高斯滤波器、Gabor 滤波器等进行卷积，从而可以对图像的结构信息进行编码。

4.3.3 实例分析

遵循车辆检测的一般方法，本书也采用假设产生和假设验证两个步骤来进行车辆检测。在假设产生阶段，需要一种快速且有效的方法定位出车辆的候选区域，因此，在此阶段，我们需要采用计算相对简便的特征和较快的分类器。通过对 Haar-Like 特征进行改进，结合图像的梯度特征，我们采用多尺度方向特征 (Multi-Scale Orientations, MSO) 来对车辆进行表述，并利用 Adaboost 级联分类器进行车辆候选区域的检测。

在假设验证阶段，需要更有效，表述更精确的特征和更准确有效的分类器进行车辆区域的最后判别。通过对目前国内外所使用特征的分析，我们提出了一个新的特征集，多尺度方向直方图 (Histograms of Multi-Scale Orientations, H-MSO)，来用于车辆的表述，并采用 SVMs 分类器来进行分类判别。多尺度方向特征通过将图像与不同尺度和方向的 Gabor 滤波器卷积得

到，这个特征集合将空间直方图和多尺度方向结合了起来。它是从 HOG 特征演化而来，我们用 Gabor 方向响应代替了梯度方向，从而获得了多尺度的方向，这使得我们的算法在路面车辆检测中更加鲁棒。

检测算法的框架如图 4.12 所示。首先，所有的正例和反例的训练样本都用 MSO 特征集来描述，并对所有的样本经过 Adaboost 分类器进行训练，形成初始的分类模型，来完成车辆假设的产生过程。然后，所有的正例和反例的训练样本都用 H-MSO 特征集来描述，并对所有的样本用 SVMs 分类器进行训练，形成最后的分类模型。在车辆检测过程中，逐行逐列地扫描固定大小的图像块，然后将图像块依次用 Adaboost 分类器和 SVMs 分类器进行判别，来最终完成图像中车辆区域的检测。

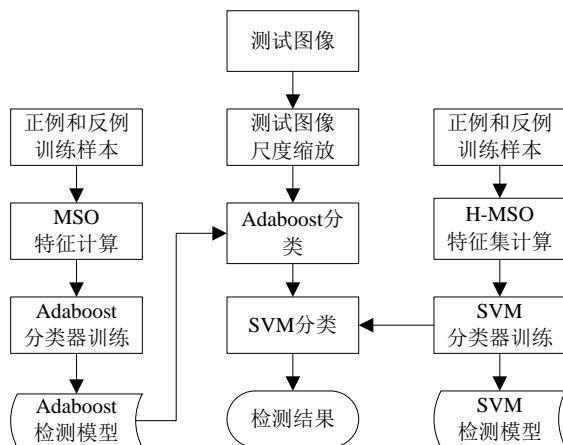


图 4.12 辆检测算法框架

实验数据

对于车辆检测来说，实验数据也起到了很重要的作用。训练数据的选择直接影响到分类器模型的好坏，而测试数据则用来评估算法的性能。

训练数据集

由于本书研究的路面车辆主要是正视的车辆，因此，在训练过程中，我们也主要使用正视的车辆作为正例样本。我们一共收集了 1500 多幅正视的车辆图像作为正例样本和多于 5000 幅的非车辆图像反例样本（如图 4.13 所示）。所有的训练样本都是手工进行收集和标定的。为了确保数据的多样性，我们分别在不同的时间和条件下收集了大量的样本。为了保证训练样本的

一致性，训练集中的每一个样本都被归一化到 64×64 的大小。

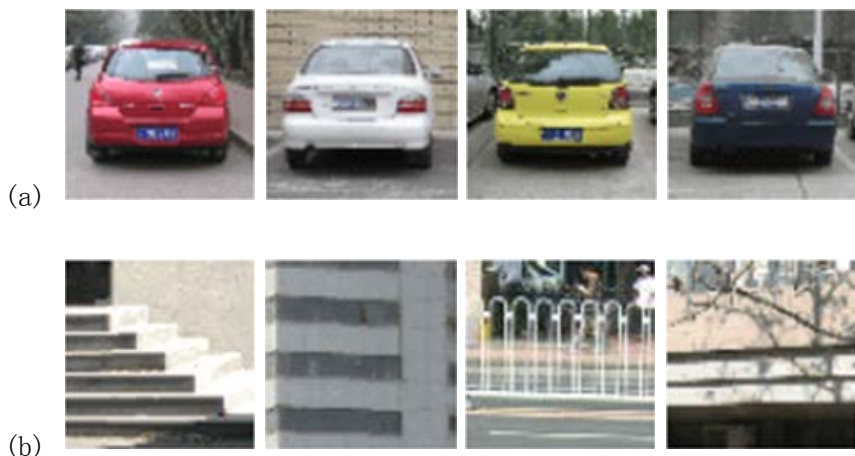
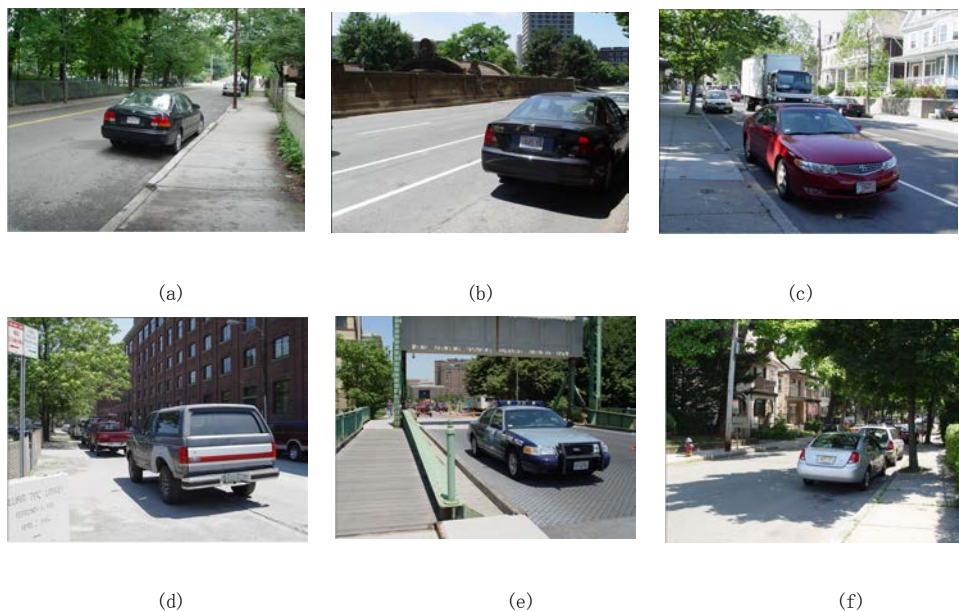


图 4.13 训练样本集 (a) 正例样本 (b) 反例样本

测试数据用于评估算法的性能。我们一共使用了 684 幅图片来定量地评估本书所提出方法的有效性，这些图片是从麻省理工学院 (MIT)，CBCL 实验室下载的测试集。这些图片包含了不同的形状、颜色、视角，并且有局部遮挡和光照变化的影响等。我们所使用的一些测试图片如图 4.14 所示。





(g)



(h)



(i)

图 4.14 测试图片示例

图 4.15 显示了用本书的算法进行假设验证过程的结果,可以看到在假设产生阶段的基础上,验证阶段又过滤掉了大量的误检区域。(f) 图表明本算法对一定程度的光照变化具有鲁棒性。(g) 图表明算法可以检测到不同类型的车辆。在 (d) 图中,存在一定视角变化的车辆和不同形状的车辆都被检测到了。(a) 图、(b) 图、(c) 图、(e) 图和 (l) 图等表明了本算法对视角变化具有一定的鲁棒性。并且在 (d) 图所处的复杂背景下,没有错误的检测发生。(i) 图说明算法对部分遮挡具有一定的鲁棒性。所有的检测结果可以表明,本书使用的算法在各种复杂的背景下,对正视车辆都具有很好的效果,并且对于视角的轻微变化具有一定的鲁棒性,对于轮廓与轿车类似的其他车型,也可以检测到。



(a)



(b)



(c)



(d)



(e)



(f)

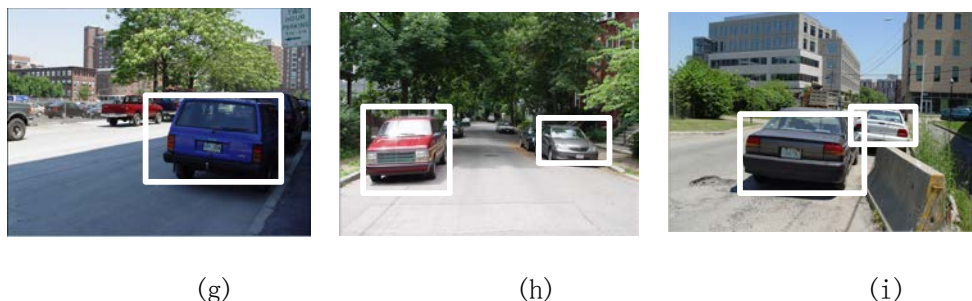


图 4.15 测试图片车辆检测结果

为了验证本书提出的算法是否能够满足实时的路面车辆的检测，我们对在路面环境下拍摄的视频中的图像进行了检测，检测结果如图 4.16 所示：



图 4.16 视频中的路面车辆检测结果

4.4 字符类目标检测

随着数字化浪潮的兴起，网页上和数据库中的数字化视频和图片正在呈现爆炸式增长。传统的通过手工标注和索引视频的方法已经不能满足人们的需求，亟需研究新的方法和手段管理和检索视频和图片。这就导致了近年来基于内容的图片和视频检索技术的兴起。对于基于内容的视频检索来说，文字是一种重要的对象，因为它包含了直接的高层语义信息。例如，网页图片中的文字对于网站内容分析具有重要作用，数字化的书籍和杂志上的封面文字有助于从数据库中检索这些资源。新闻视频的标题文字往往包含正在播报的新闻时间、地点或者人物信息。体育视频中的记分牌等往往包含了比赛分数、运动员信息或者精彩事件的出现。与其他图像特征甚至音频特征相比较，文字信息是人工叠加到图片中或者印刷到对象上，这些文字能够直接

从特定的角度反映图像内容，同时不需要复杂的运算。因此，近年来，不少研究人员开始关注图片和视频帧中的文字检测和识别技术研究。



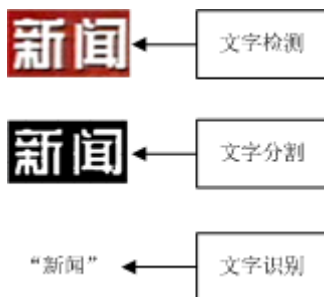
(a) 体育视频中的文字检测



(b) 新闻视频中的文字检测



(c) 自然场景图像中的文字检测



(d) 文字检测、分割和识别过程

图 4.17 文字检测系统示意图

“文字检测”是对文字检测系统的一个统称。完整的系统应该包含“文字检测”、“文字识别”两个部分。为了方便阐述文字检测系统，图 4.17 中给出了一个示意。“文字检测”是在一幅图像中找到文字行的位置，并且准确确定出其外轮廓。“文字识别”是将检测到的文字行中的内容读取出来，读取可以使用基于灰度特征的 OCR（光学字符识别）软件或者基于二值特征的 OCR 软件。当然，也有研究者针对图像中的文字字符识别研究专门的识别算法。

在图 4.17，给出了文字检测系统的示例。其中，图 4.17(a)是一幅体育视频中包含比赛信息的图像；图 4.17(b)是一幅新闻视频中包含新闻播报内容的图像；图 4.17(c)是一幅自然场景中包含了警示牌文字信息的图像；在图 4.17(d)中，作者用箭头指示了对于一个检测到的文字行，通过文字分割算法分割得到二值图像，并且使用 OCR 软件进行识别。

在文字检测系统中，需要深入研究的主要是如何从图像中找到文字的精确位置，也就是定

位问题。其中，与传统的模式识别问题一样，特征提取和选择是解决问题的关键。结合具体特征和样本特点进行模型训练也是需要解决的问题。同时，在文字行检测过程中，还需要研究图像区域分割和布局分析的问题。

文字检测的目的是从一幅图像中找到只包含文字信息的区域，这个区域可以被突出显示给用户看或者输入到光学字符识别软件中识别。如果想识别文字内容，准确的文字定位是必不可少的步骤。在某些情况下，仅仅检测到文字行也是有意义的。例如，如果在新闻视频中检测到有文字行的出现，那么就有助于新闻视频的结构分析，有助于精确的分割新闻条目^[26]。当然，作为一个完整的系统，仍然需要研究文字的识别的工作，文字识别相关内容在本书中将不予涉及，请读者自行查阅相关文献。图 4.18 是文字检测、识别的一个系统框图。

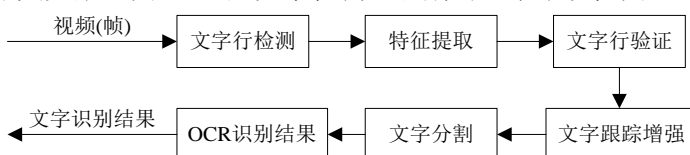


图 4.18 文字检测和识别框图

在过去的研究中，研究人员已经提出了许多文字检测方法，但是快速并且鲁棒的文字检测技术，尤其是适应复杂背景的检测方法仍然需要深入的研究。这里“鲁棒性”可以归结为以下内容：

- 1) 对于字体大小、文字颜色、方向、背景甚至文字语言的变化能够保证较好的检测性能（较高查全率和较低错误率）；
- 2) 稳定的检测性能，不需要经常的手工调整参数或者重新选择训练样本，更换分类模型等，从而给实际应用带来方便。

在实际当中，开发一种快速而且鲁棒的文字检测算法并非易事，因为在其中确实存在着许多困难，包括：

- 1) 文字可能被镶嵌到复杂背景中；
- 2) 难于提取有效的底层特征用于文字/非文字的分类，尤其是区分文字和一些很像文字的

目标，如树叶、窗帘等纹理；

- 3) 文字模式因字体的不同，字体颜色的差别，甚至是语言的不同，而发生较大的变化；
- 4) 在图片，尤其是视频帧中，文字质量会因为图像编（解）码而降低，有些低分辨率的文字甚至用肉眼都无法分辨。

在一些的检测算法研究中，考虑到以上困难和问题，较多的采用了文字的以下属性来研究文字检测算法：

- 1) 密集灰度变化；
- 2) 文字和其背景的对比度；
- 3) 文字行的结构信息；
- 4) 文字区域的纹理特性。

本节介绍的方法，没有完全通过监督学习的方法^{[27][28]}将图片块分类成文字块/非文字块，而是在由粗到精的检测框架下，通过在不同的检测阶段使用不同的检测特征来实现检测算法。在粗检测阶段，通过文字属性 1 和 2，我们可以认为所有的文字区域都包含密集的灰度变化以及和背景的对比度，从而利用这两个属性粗略的定位文字候选区域。通过属性 3（结构信息）可以将候选文字区域分割成为候选文字行。在精确分类的过程中，利用属性 4（纹理属性）判别文字和非文字模式。因为在候选当中，非文字和文字一样具有强的灰度变化，仅仅使用灰度和结构信息难于区分。在此过程中，我们提取了三类纹理特征和一种结构特征用于分类，其中纹理特征包括小波矩特征、小波直方图特征、小波共生特征，结构特征是穿越线直方图特征。这里使用了前向搜索的特征选择方法找到有效的特征组合，并且使用了 SVM 方法进行分类，图 4.19 是算法的框图。

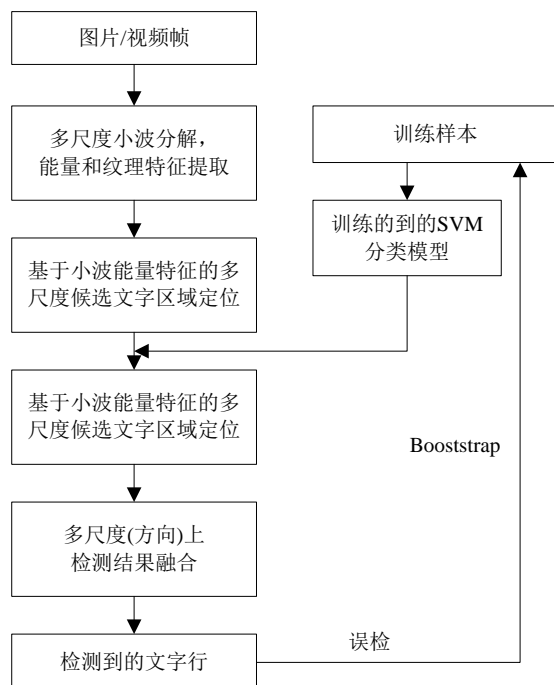


图 4.19 文字检测算法框图

算法的优势在于如下几个方面：

1、**快速检测** 提出的检测方法是在由粗到精的框架下，这样就避免了对图像块进行一块挨一块的分类，大大减少了计算复杂度，提高了检测速度。同时，特征选取程序也降低了特征维数，提高了分类效率。

2、**多尺度检测** 文字的模式会随着文字大小的变化而变化，在本书中，作者使用多尺度的小波特征来检测字体大小不同的文字，而不需要对原图像做缩放。相比于只是提取了小波系数的均值、二阶和三阶、中心矩特征，而且对于所有的文字都是使用所有前三个小波分解尺度上检测文字而言，这里提出的方法，使用对应尺度内的小波特征检测大小一定的文字。

3、**特征融合的方法** 在以前的方法中，研究们简单的使用传统纹理特征来判别纹理和非纹理模式。这里我们融合了 3 种纹理特征和一种结构特征进行分类，并且对于提取的特征，还使用了特征选择的方法寻找优化的特征融合。

4、**鲁棒性的检测方法** 此处提出的方法能够检测尺度和颜色不同的文字，通过旋转区域增长模版，还能够检测不同方向的文字。精确分类程序降低了检测的错误率，从而使得提出的检

测方法在复杂背景下也能够工作良好。更为重要的是，精确分类的特征是在一整行文字上去提取的，这些特征比从一小块图像中提取的特征更为鲁棒，因为从一小块图像中提取的特征可能不能够包含足够的纹理和结构特征进行分类。这些都保证了算法的鲁棒性。

4.4.1 字符类目标检测数据库

(待完善)

4.4.2 字符检测常用特征

4.4.3 特征提取

文字可以看作一种纹理模式。但是，通过研究我们发现其纹理属性，如规则性、方向性等是很差的。与其说文字是一种纹理，不如说它包含了一些图像的突变信号，这些突变信号在固定尺度内具有一定宽度，并且这些突变信号在特定方向上形成了一个矩形结构，我们称之为“文字行”。于是，简单的一种纹理特征很难反映文字行的纹理特性，因为纹理特征往往依赖于纹理的几个基本统计量，如规则性、方向性、相关性和熵等。这里给出了将纹理和结构特征进行融合的方法。三种小波域上的纹理特征主要是用来反映文字区域内的突变信号的分布和属性，一种基于梯度的结构特征统计沿文字行方向的整体突变信号的周期性特点。梯度信号在原始灰度图像上提取，小波域的纹理特征将在包含当前候选文字像素最多尺度(suitable scale)上提取。

文字中的字符笔画宽度会随着文字大小和字体的变化而变化。我们认为宽的文字笔画是宽度较大的阶跃信号，窄的文字是宽度较小的阶跃信号或者是脉冲信号。根据小波分解的理论，小波是由一组带通滤波器构成，不同的带通滤波器对于宽度不同的信号响应不同，但是由较宽的带通滤波器对宽信号滤波的结果应该和较窄的滤波器对较窄的信号的滤波结果相似。也就是说，字体较大的文字在较深的小波分解尺度上的结果和字体较小的文字在较浅的小波分解尺度上的结果经过规整能够相似。因此，文字在合适的尺度内提取特征将能够把字体大小不同的文字模式统一起来。同时，更为重要是大字体的文字将在较深的分解尺度具有较强的小波响应，而小字体的文字在较浅的分解尺度具有较大的响应。这将有利于区分文字模式和非文字模式。如果一个文字行在尺度 n 上检测到的候选像素最多，我们认为尺度 n 就是其合适的尺度，而所有的纹理特征将从这个尺度内提取。表4.3给出了提取的特征的描述。小波域的特征是已经被研究者们提出的特征并且在纹理分类上得到了应用，穿越线直方图特征则是作者提出的反映字符沿着文字行方向的周期性的信特征。

表 4.3 特征描述

特征集	特征描述	特征数目	选用的特征数目
小波矩特征	均值、二阶和三阶中心	9	6
小波直方图特征	小波能量直方图 and 方向直方图	20	10
小波共生特征	能量、熵、和谐性和相关性	180	16
穿越线直方图特征	穿越线直方图	16	9
总数		225	41

4.4.3.1 小波矩特征

文字和非文字具有不同的空间灰度分布，并且形成了不同的空间灰度变化。使用小波系数均值和中心矩特征来反映这种变化，而这些特征是小波域最常用的纹理特征，它们反映了小波系数的大小和分布均匀情况，这些特征在文字块的检测方面已经被证明了是十分有效的。给定一个尺寸为 $M \times N$ 的文字行 T ，其均值 (m)，二阶中心矩 (μ_2) 和三阶中心矩 (μ_3) 可以用如下公式计算：

$$m(T) = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} T(i, j), \quad (4-7)$$

$$\mu_2(T) = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (T(i, j) - m(T))^2, \quad (4-8)$$

$$\mu_3(T) = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (T(i, j) - m(T))^3, \quad (4-9)$$

其中 $T(i, j)$ 表示像素 (i, j) 的小波系数，上式中的特征都是在小波分解的高频子带 (HL, LH 和 HH 子带) 上提取。通过在一个固定的尺度提取 3 个小波子带上的 3 维特征，总共得到 9 维小波矩特征。

4.4.3.2 小波直方图特征

直方图是一种一阶的能够反映信号分布的特征。在本书中，使用了能量直方图 ($WEH(i), i=0, \dots, 15.$) 和方向直方图 ($WDH(i), i=0, 1, 2, 3.$) 两个统计量反映文字区域的能量分布及其突变信号的方向。为了计算 $WEH(i), i=0, \dots, 15.$ ，小波能量首先需要进行量化，本

书采用均匀量化的方法得到 16 级离散的能量值。其中离散值得计算方法如下式所示：

$$WE_q = WE \cdot 16 / (WE_{\max} - WE_{\min}), \quad (4-10)$$

式中 WE 表示一个像素的小波能量， WE_{\max} 和 WE_{\min} 分别表示一幅图片中小波能量的最大和最小值。根据最大值和最小值，可以确定小波能量的分布范围，并且使用式（2-8）实现均匀量化。根据量化结果，小波能量直方图 $WEH(i)$ 通过计算能量等于 i 的像素数目占所有像素的数目的百分比得到。对于一个文字行来说，直方图 $WEH(i)$ 的前几个和后几个位的数值应该比较大，而中间部分比较小（如图 4.20 所示）。这正是由于文字和其背景之间的对比度造成的。而对于一个非文字模式，直方图 $WEH(i)$ 则不一定具有这样的特点。因此我们说直方图 $WEH(i)$ 具有一定的区分性。

$WDH(i)$ 包含了水平（ $i=1$ ），竖直（ $i=2$ ），对角线（ $i=3$ ）和无方向（ $i=0$ ）四个位。一个候选文字像素的方向定义为其水平、竖直和对角线方向系数的最大值。也就是说一个像素在尺度 n 上是水平（竖直、对角线）方向的如果 D_{n1} （ D_{n2} ， D_{n3} ）是 $\{D_{nk}\}_{k=1,2,3}$ 中最大的。所有非文字像素都被认为是没有方向的。 $WDH(i)$ 而在方向 i 上的数值 是在这个方向上的像素数目的百分比。一般文字区域的灰度变化方向都是比较零乱的， $WDH(i)$ 可以用来区分那些具有较好方向性的纹理（见 4.20 右图）。



图 4.20 图像的小波直方图

4.4.3.3 小波共生特征

直方图特征是信号的一阶统计，它只能反映信号的总体分布，不能反映信号的局部特性，尤其是相互之间的关系。当一阶统计不足以区分所有文字/非文字模式时，二阶统计能够通过反映相邻像素之间的相互关系提高判别能力。共生特征就是这样能够反映信号之间相互关系的统

计特征。共生矩阵 $C(d, \theta)$ 是一个方阵，元素 (i, j) 的数值定义为小波系数 $D_{nk} = i$ 和小波系数 $D_{nk} = j$ 在方向 θ 和距离 d 同时发生的概率。随着 θ 和 d 的改变，同一幅图像可以得到多个共生矩阵。

通过统计共生矩阵的系数，我们得到了小波域内的共生特征，这些共生特征包括了能量、熵、惯性矩、局部和谐性和相关性，它们计算方法如下：

$$\begin{aligned} \text{Energy: } E(d, \theta) &= \sum_{i,j} C^2(d, \theta), \\ \text{Entropy: } H(d, \theta) &= \sum_{i,j} C(d, \theta) \cdot \log C(d, \theta), \\ \text{Inertia: } I(d, \theta) &= \sum_{i,j} (i - j)^2 C(d, \theta), \end{aligned} \quad (4-11)$$

$$\text{Local homogeneity: } L(d, \theta) = \sum_{i,j} \frac{1}{1 + (i - j)^2} C(d, \theta),$$

$$\text{Correlation: } C(d, \theta) = \frac{\sum_{i,j} (i - \mu_x)(j - \mu_y) C(d, \theta)}{\sigma_x \sigma_y},$$

其中 μ_x , μ_y 和 σ_x , σ_y 表示共生矩阵 $C(d, \theta)$ 的均值和方差。 θ 是共生矩阵的方向。在实验中作者选用了 0 度、45 度、90 度和 135 度四个方向。 d 是共生距离，实验中选用 1、3、5 个像素，其实选用更大距离的共生已经没有意义，在一个固定尺度内，文字笔画的宽度一般在几个像素的范围内。通过在 3 个小波域内 (HL, LH 和 HH) 计算 12 个共生矩阵上的统计特征，我们总共得到了 180 维的共生特征 (3 小波子带 \times 12 共生矩阵 \times 5 种特征)。

4.4.3.4 穿越直方图特征

以上提取的特征只是考虑文字信号的整体分布特征，并没有考虑字符沿着文字行方向的周期性特点。但是，这种“字符-空格-字符-空格”的分布却是一个重要的特征，能够区分文字行和其他杂乱无章的对象。在信号处理的领域，信号周期性的特点可以通过“空域-频域”的变换反映出来。但是，文字行的长度是有限的，有时候只有几个字符，“空域-频域”的变换很难反

映出文字行的周期性特点。因此，我们提出了一种在空间域直接使用穿越线捕捉文字周期性特点的特征——穿越直方图特征（CCH）。

CCH 是水平方向上的穿越线和 GPM 的相交次数的直方图统计。计算过程中，为了得到简化的空间信号分布，我们首先提取了文字的梯度图(图 4.22(b))，并且通过竖直方向所有像素将梯度投影到一维信号得到梯度投影(称为 Gradient projection map, GPM, 如图 4.21(c)所示)。对 GPM 进行高斯平滑后，可以观察到，在 GPM 上有近似的周期性和规则性存在。这是在水平方向上画一些穿越线(图 4.22(d))，然后统计每一个穿越线和 GPM 的交汇次数（图 4.21(d)中包含了 4 根示意的穿越线）。假设有 N 个穿越线，第 K 个穿越线的穿越次数是 $CC(k)$ ， $k = 1, 2, \dots, N$ ，那么穿越线直方图可以通过下式计算：

$$CCH'(k) = \frac{CC(k)}{\sum_{i=1}^N CC(i)}. \quad (4-12)$$

例如，如果梯度投影图的最大值是 300，我们使用 $N = 300$ 条线扫描 GPM，得到 300 个直方图位 $CCH'(k)$ ， $k = 1, 2, \dots, 300$ ，通过均匀的划分直方图的位 $CCH'(k)$ ，并且将其累加：

$$CCH(i) = \frac{1}{16} \sum_{k=i \cdot \frac{N}{16}}^{(i+1) \cdot \frac{N}{16}} CCH'(k), i = 1, 2, \dots, 16., \quad (4-13)$$

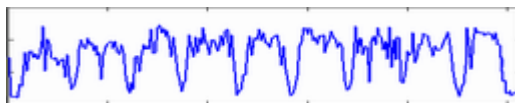
从而就得到了 16 位的穿越直方图，上式中， $\left(i \cdot \frac{N}{16}, (i+1) \cdot \frac{N}{16}\right)$ 表示一个无重叠的窗口，窗口内的直方图位数累加起来形成直方图的一维。



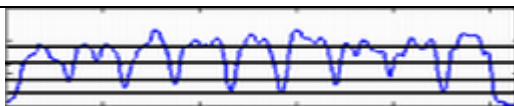
(a)



(b)



(c)



(d)

图 4.21 图像的穿越线直方图

经过以上分析和特征提取过程可以看出，穿越直方图特征反映的是水平方向上信号被穿越的一个整体统计，它能够粗略的反映出文字行中字符的周期性特点，同时实验证明它对于周期性比较差的文字行也不是很敏感。

4.4.4 特征选择

基于上面的特征提取方法，对于整个文字行，我们一共提取了 225 位特征。尽管所有的特征都能够用于区分文字和非文字模式，但是，其中的一些特征的区别性可能更好，而一些较差。当然，有些特征可能和别的特征组合起来分类性能更好，一些则会变差，在有些时候使用特征集合的子集反而会取得更好的效果，因此需要从特征集合选取有效的特征组合提高分类性能。更为重要的是，在有限训练样本集合的情况下，降低特征的维数有利于提高分类模型的性能。

在本书中，我们没有使用 PCA 或者 LDA 的降维方法，而是使用前向搜索的方法进行特征选择，因为提取的特征已经具有中层的含义，而不是原始的灰度或者结构信息，不需要对特征进行变换。当然，使用特征选择的方法选择的特征是次优的，而且会保留一些相关性，不过这种方法能够降低分类时候的复杂度。

前向特征选取过程中，特征的评价是基于 SVM 分类精度进行的，其具体的算法可以描绘如表 4.4。

表 4.4 前向特征选取算法

特征集合 F 首先被划分为选用特征集合 F_s 和未选用特征集合 F_u ，接着使用如下程序逐个的选择特征：

- 1) 设置 $F_s = \phi$ ， $F_u = F$ ；
- 2) 将集合 F_u 中的特征设定为未测试过；
- 3) 从 F_u 中任意选择一个未测试的特征 f ，并且标记它为已测试；

4) 将 f 放入到集合 F_S 中, 形成临时特征集合 \tilde{F}_S ;

5) 评估 \tilde{F}_S 的分类性能;

在此过程中, 使用 1000 个文字样本和 2500 个非文字样本组成样本集合进行训练和分类。样本被均匀的分成 10 份, 一份用来训练, 其余用来测试。训练和测试不断循环直到所有的样本都曾做过样本和测试样本。平均的分类精度就是当前特征集合的分类精度, 分类精度 (Accuracy) 计算如下:

$$Accuracy = \frac{\text{正确分类的样本数目}}{\text{总的样本数目}}. \quad (4-14)$$

6) 如果仍然有 F_U , 转到 3);

7) 从所有未测试特征中寻找特征 \hat{f} , 当将其加入特征集合 \tilde{F}_S 时, 能取得最高的分类精度, 即:

$$\hat{f} = \arg \max Accuracy(\hat{F}_S) \quad (4-15)$$

并且将 \hat{f} 从 F_U 中移动到 F_S 中;

8) 如果 F_U 中仍然有未测试的特征, 转到 2), 如果 F_U 所有特征都已经处理, 特征选择结束。

图 4.22 是特征选择过程中的分类精度曲线。从曲线上可以看出随着选用的特征数目的增加, 分类精度首先有很大的增加, 但是当选用的特征数目超过 41 维时, 分类精度又会下降, 于是选用 41 维特征用于分类。可以看出, 在选用的特征中, 几乎所有的小波矩特征都被选用了, 这表明这些特征的区分性最好。大约一半的穿越线直方图被选用, 这表明它们的区分能力是中等的。有较少的共生特征被选用, 这表明其区分能力是比较差的, 至少和前面的特征组合起来以后区分能力是比较差的。

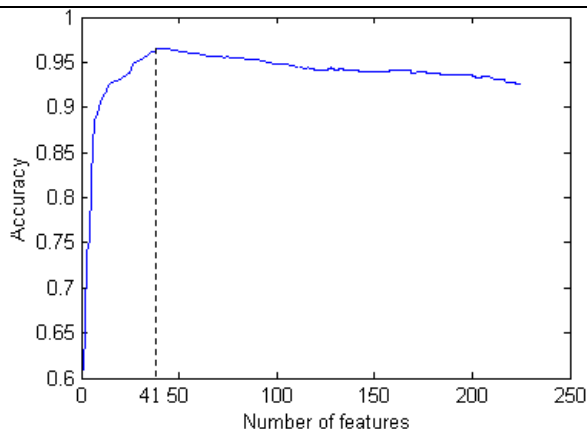


图 4.22 特征选择：最好的分类结果在特征数目是 41 维时

4.4.5 字符检测分类器及方法

我们选用了 SVM 做为分类器，和其他分类器相比，如：人工神经网络、决策树等，SVM 更易于训练，需要更少的训练样本。更为重要的是，SVM 在粗糙的训练集合上仍然能够训练得到推广性较好的分类模型，保证结构风险最小化。考虑到文字模式的分散性本书中选用了 SVM 为分类器。

SVM 在近年的许多模式分类问题中取得了很好的效果，尤其是在是二类分类问题^{[30][31]}中成果显著。传统的统计分类方法使用经验风险最小化进行训练分类模型，在推广性方面缺少数学上的证明。SVM 使用结构风险最小化准则寻找优化的分开两个类别的超平面。在本节中，我们通过分类性能比较，为 SVM 选择了多项式核函数。

我们准备了一个包含 3200 个文字样本和 8000 个非文字样本的训练集合用于训练 SVM 分类模型。图 4.23 给出了一些文字（正例）和非文字（反例）样本的例子。尽管可以比较容易的标注正例样本，但是反例样本确是难于寻找，尤其是有代表性的反例样本，因为非文字的模式多种多样，不一而足。于是，在得到一个初始的分类模型以后，使用了“bootstrap”的过程）进行重复训练模型，提高分类性能。在重复训练的过程中，一些容易被误检为文字的样本需要作为反例进行训练。从图 4.24 中的例子可以看出，在粗定位的过程中，算法产生了许多错误的候选。在经过精确分类以后，许多错误的候选被去除了，这说明了精确分类过程在降低误检率方面是十分有效的。



图 4.23 训练样本例子，(a) 是正例样本，(b) 是反例样本

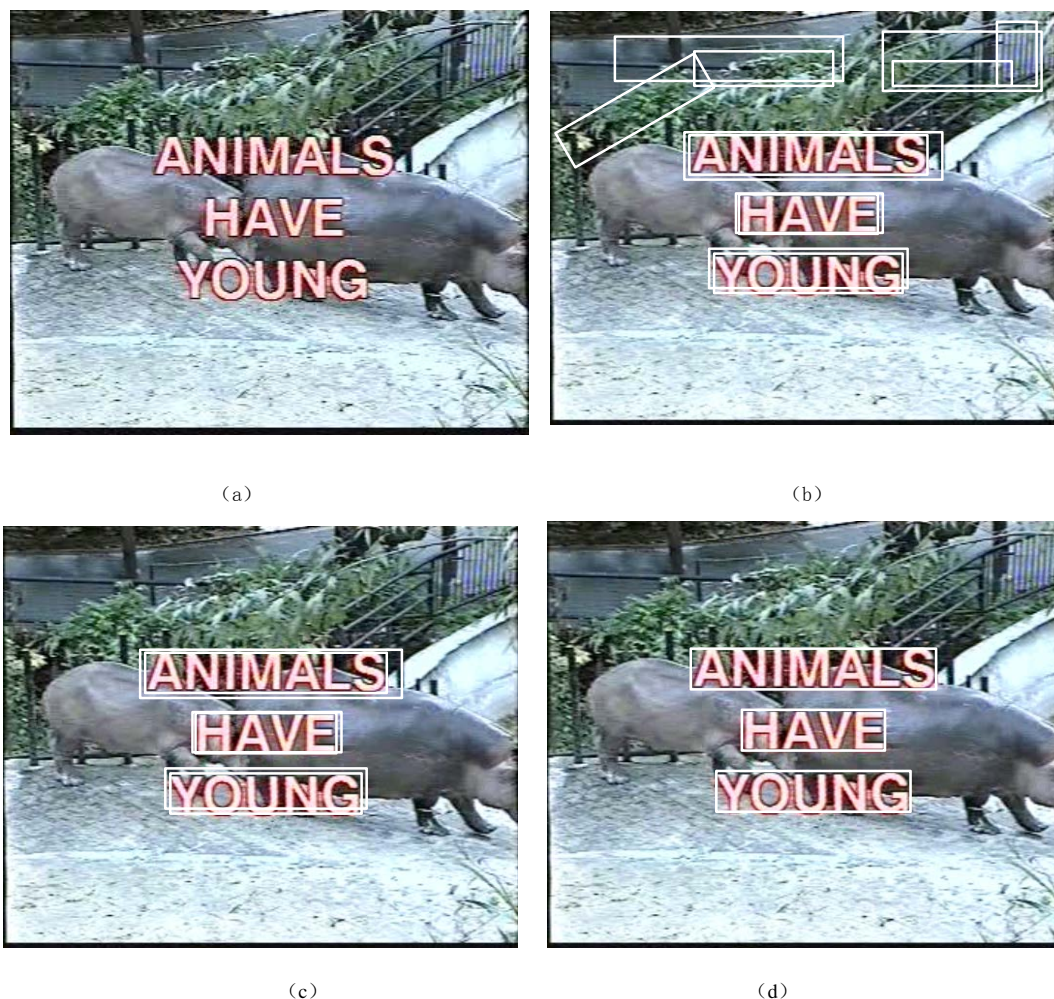


图 4.24 文字检测结果的例子，(a) 是原始图像，(b) 是粗定位的结果，(c) 是精确分类后的结果，(d) 是经过多尺度（方向）合并后的最终检测结果

有时一个文字行或者文字行的部分会被在多个方向或者尺度上检测到，这样检测结果会在

原始的图像内发生重叠（如图 4.24(b) 所示）。当两个候选文字行的重叠部分超过一半于原文字行的面积时，我们认为其中的一个候选应该被删除，按照比较直观的做法，是文字概率较小的文字行被删除。这个概率可以通过 SVM 的分类过程粗略估算得到，具体方法可以参照。同时，文字行被重叠的次数也可以参考。图 4.24(d) 是经过多尺度检测结果合并的最终结果。

4.5 参考文献

- [1] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
- [2] http://www.ri.cmu.edu/projects/project_418.html
- [3] http://cobweb.ecn.purdue.edu/aleix/aleix_face_DB.html
- [4] <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [5] <http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/>
- [6] <ftp://whitechapel.media.mit.edu/pub/images/>
- [7] <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>
- [8] <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>
- [9] <http://www.bioid.com/downloads/facedb/index.php>
- [10] <http://images.ee.umist.ac.uk/danny/databse.html>
- [11] 山世光. 人脸识别中若干关键问题的研究, 中国科学院计算技术研究所, 博士学位论文, 2004 年 7 月
- [12] <http://pascal.inrialpes.fr/data/human/>
- [13] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: A Benchmark[C]. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 304 – 311, 2009.
- [14] C. Wojek, S. Walk, and B. Schiele. Multi-cue Onboard Pedestrian Detection, [C]. Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 794-801, 2009.
- [15] <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/index.html>.

- [16] <http://www.mis.tu-darmstadt.de/tud-brussels>
- [17] <http://coe.gucas.ac.cn/SDL-HomePage/resource.asp>
- [18] Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, Fast Human Detection Using a Cascade of Histograms of Oriented Gradients, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, pp.1491-1498, 2006.
- [19] Qixiang Ye, Jianbin Jiao, Baochang Zhang, Fast pedestrian detection with multi-scale orientation features and two-stage classifiers, IEEE International Conference on Image Processing, pp:881-884, 2010
- [20] Z. Sun, B. George, and M. Ronald. On-Road Vehicle Detection Using Evolutionary Gabor Filter Optimization. IEEE Transactions on Intelligent Transportation Systems, June 2005(6): 125-137
- [21] A. Broggi, M. Bertozzi, A. Fascicli, C. Guanrino Lo Bianco and A. Piazzzi. Visual Perception of Obstacle and Vehicle for Platooning. IEEE Transactions on Intelligent Transportation Systems, 2000(1): 164-176.
- [22] N. Matthews, P. An, D. Charnley and C. Harris. Vehicle Detection and Recognition in Greyscale Imagery. Control Eng. Pract, 1996(4): 473-479.
- [23] C. Georick, N. Detlev, and M. Werner. Artificial Neural Net-works in Real-time Car Detection and Tracking Applications. Pattern Recognition Letters, 1996(17): 335-343.
- [24] C. Papageorgiou, and T. Poggio. A Trainable System for Object Detection. International Journal of Computer Vision, 2000(38): 15-33.
- [25] D. Navneet, and T. Bill. Histograms of Oriented Gradients for Human Detection. Computer Vision and Pattern Recognition, June 2005(11): 886-893.
- [26] H. Wang, "Automatic character location and segmentation in color scene images," International Conference on Image Analysis and Processing 2001.
- [27] R. Lienhart and A. Wernicke, "Localizing and segmenting text in images and videos," IEEE Trans. Circuits and Systems for Video Technology, Vol.12, pp.256-268, 2002.
- [28] H. Li, D. Doermann, and Omid Kia, "Automatic text detection and tracking in digital video," IEEE Trans. on Image Processing, Vol.9, pp.147-156, 2000.

- [29] X. Tang, X. B. Gao, J. Liu and H. Zhang, "Spatial-temporal approach for video caption detection and recognition," IEEE Trans. Neural Networks, 13 (2002) 961-971.
- [30] K. Sung and T. Poggio, "Example-based learning for view-based human face detection," Mass, Inst. Technol., Cambridge, MA, A.I. Memo 1521, 1994.
- [31] B. Heisele, T. Serre, S. Mukherjee and T. Poggio, "Feature reduction and hierarchy of classifiers for fast object detection in video images," Int. Conf. on Computer Vision and Pattern Recognition, 2001, pp.18-24.

第5章 目标跟踪方法

5.1 跟踪目标描述

5.1.1 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、**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。

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。光流常被用来表示基于运动的分割与跟踪的特征。

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。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。

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。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，声纳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。

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。

5.1.2 跟踪目标的表示方法

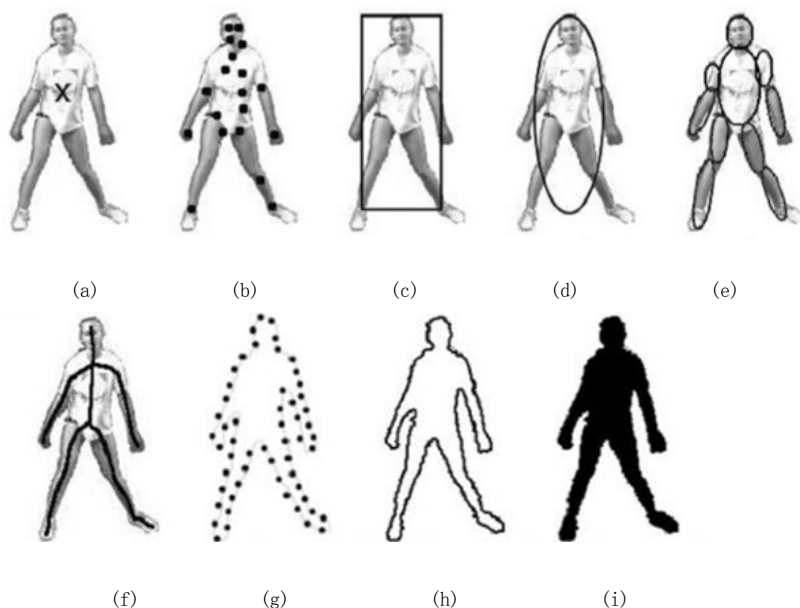


图 5.1 对象的表示方法

- (a) 对象的质心；(b) 对象上的多个点；(c) 矩形片；(d) 椭圆形片；
 (e) 基于部件的多个片；(f) 对象的骨架；(g) 轮廓上的控制点；
 (h) 对象的整个轮廓；(i) 对象的廓影

确定性目标跟踪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。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。

被跟踪对象可以定义为任意的想要跟踪的感兴趣的東西，例如，水面上的船只、天上的飞机、地面上的汽车和行人等。要对感兴趣的对象进行跟踪，首先需要给出对象的表示。通常，被跟踪对象可以用它们的形状和表观来表示。这一小节我们给出在对象跟踪领域常用的对象表示方法。首先我们给出对象的形状表示方法，然后给出对象的表观表示方法。对象的形状表示

方法有：

1、**点** 对象可以用一个点来表示，也就是对象的质心，如图 5.1 (a) 所示。通常这种表示方法适用于被跟踪对象在图像中占有较小的区域的情况。对象也可以用一组点来表示，如图 5.1 (b) 所示。

2、**基本几何形状** 对象可以用矩形框(如图 5.1 (c) 所示)或椭圆(如图 5.1 (d) 所示)来表示。这种情况下，通常采用的对象运动模型包括平移、仿射变换或摄影变换。尽管这种表示方法主要适用于简单的刚性对象，它们仍然可以用来表示非刚性对象。

3、**轮廓和廓影(silhouette)** 轮廓(如图 5.1 (g) (h) 所示)定义了对象的边界，而轮廓以内的区域被称为廓影(如图 5.1 (i) 所示)。轮廓和廓影的表示方法适用于跟踪复杂的非刚性的对象形状^[错误！未找到引用源。]。

4、**关节形状模型** 有关节的对象由通过关节连接在一起的部件组成。例如，人体就是由头、躯干、腿、胳膊、手和脚组成，并通过关节将它们连接在一起的。人体不同部件之间的运动符合一定的运动学模型，例如，关节的角度等。为了表示有关节的对象，我们可以将它的每一个组成部件用圆柱体或椭圆来表示(如图 5.1 (e) 所示)。

5、**骨架模型** 对象的骨架(如图 5.1 (f) 所示)可以通过应用某种变换从对象的廓影中提取出来。这种表示方法通常用来做物体识别。骨架模型既可以用来表示有关节的对象，也可以用来表示刚性的对象。

除了形状表示之外，对象还可以用表观模型来表示。形状和表观也可以联合使用来对对象进行跟踪。在对象跟踪领域，一些常用的表观表示方法如下：

1、**概率密度** 对象表观的概率密度可以是参数的，也可以是非参数的。参数的表示包括高斯和混合高斯两种类型。非参数的表示包括核密度估计和直方图。对象表观特征(颜色、纹理等)的概率密度可以从形状模型(椭圆、轮廓等)指定的图像区域计算得到。

2、**模板** 模板通常由简单的几何形状(如，矩形等)或廓影组成。模板的优点是它同时包含了对对象的空间和表观信息。但是，通常模板只包含单一视点的对象的表观信息。因此它们只适用于在跟踪过程中对象的姿态变化不大的情况。

3、**主动表观模型** 主动表观模型同时模型化了对象的形状和表观信息。通常情况下，对象的形状由一些特征点定义。和基于轮廓的表示方法类似，特征点通常位于对象的边界上，当然，

也可以位于对象区域上。对于每一个特征点，有一个表观向量与其对应，这个表观向量可以从颜色、纹理或梯度产生的。主动表观模型的参数需要进行训练得到。通常采用的训练方法是主成分分析法。

4、**多视表观模型** 这种模型模型化了对象的不同视角。一种表示对象不同视角的方法是从这些视角的样本中产生一个子空间。主成分分析方法仍然可以应用到这种情况。另外一种表示对象的不同视角的方法是从这些样本中训练得到一个分类器。例如，可以采用支持向量机和光流结合进行对象跟踪。多视表观模型的一个主要局限性是需要考虑所有不同视角的表观。

通常，被跟踪对象的表示方法是和采用的跟踪算法紧密相连的，并且和具体的应用领域相关。如果被跟踪对象在图像中占有的区域很小，那么采用点表示是适当的。如果被跟踪对象的形状能够很好地由矩形或椭圆近似，那么采用基本几何形状来表示是可行的。如果被跟踪对象具有复杂的形状，例如，人体跟踪，那么采用轮廓及廓影来表示被跟踪对象是再恰当不过的了。

5.2 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^[1]、基于区域统计匹配的跟踪算法^[2]、基于模型匹配的跟踪算法^[3]和基于Mean-Shift的跟踪算法^[4]。

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，文献^[5]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。

5.2.1 跟踪目标模板提取

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^[6,7]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，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)。

5.2.1.1 跟踪目标区域的定义



图 5.2 目标区域定义

在本书的研究背景下，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。本书采用矩形框的形式描述跟踪目标区域。目标区域定义为 $h \times w$ 的矩形框见图 5.2，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。

5.2.1.2 颜色直方图的提取

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 $k(\cdot): [0, \infty) \rightarrow R$ 进行平滑后的颜色直方图为：

$$\hat{q}_b(y_0) = C_1 \sum_{i=1}^n k\left(\left\|\frac{y_0 - X_i}{h_1}\right\|^2\right) \delta[I(X_i) - b_j], \quad \text{if } 1 \leq j \leq m \quad (5-1)$$

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：

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

5.2.1.3 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。

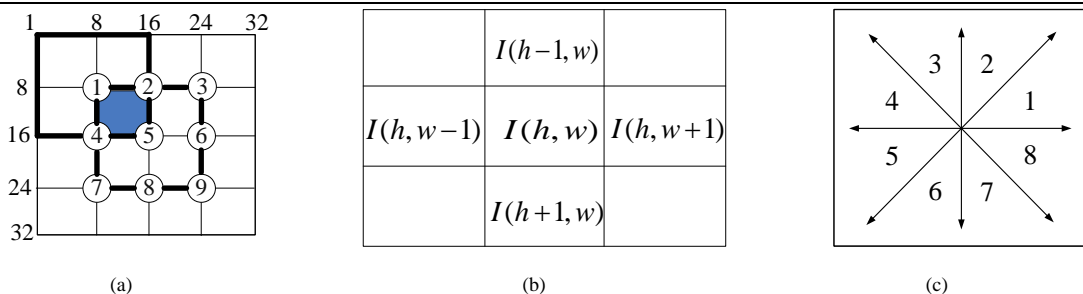


图 5.3 HOG 特征提取示意图。(a) 图像窗口的划分, (b) 计算梯度方向的模板, (c) 8 个方向投影区间。

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: 32×32 像素; 然后把图像划分为 16 个 8×8 的小区域(cell), 并且每个 2×2 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; 本书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(见图 5.3 (c))。从图 5.3 (a)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(见图 5.3 (b)):

$$\begin{aligned}
 I &= G(\sigma, 0) * I_0 \\
 dy &= I(h+1, w) - I(h-1, w) \\
 dx &= I(h, w+1) - I(h, w-1) \\
 ori(h, w) &= \text{atan2}(dy, dx) \quad ori \in [0, 2\pi]
 \end{aligned} \tag{5-3}$$

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, 见图 5.3 (a)和图 5.3 (c))维 HOG 特征。

5.2.1.4 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, 本节采用了 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(见图 5.4(a1)与图 5.4(a2))。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(见图 5.4 (b1)与图 5.4 (b2)), 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(见图 5.4 (c1)与图 5.4 (c2)), 这样提取的 HOG 特征在一定程度上与目标的旋转运动无关。

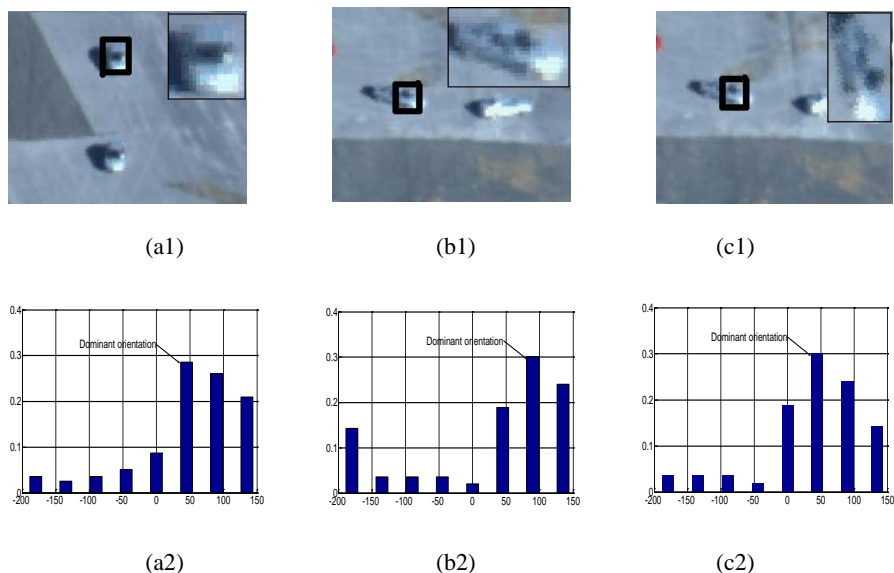


图 5.4 利用 HOG 主方向对目标进行旋转的示意图

5.2.2 跟踪目标模板度量

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:

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (5-4)$$

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:

$$d(P_1, P_2) = \begin{cases} \omega_0 |x_1 - x_2| + \omega_1 |y_1 - y_2| & \text{if } |x_1 - x_2| > |y_1 - y_2| \\ \omega_0 |y_1 - y_2| + \omega_1 |x_1 - x_2| & \text{others} \end{cases} \quad (5-5)$$

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 $d(A, B) = \max_{a_i \in A} \min_{b_j \in B} \|a_i - b_j\|$; 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。已知两个向量点 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:

$$d(P_1, P_2) = \sum_{i=1}^m \sqrt{x_i \cdot y_i} \quad (5-6)$$

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:

$$D(A, B) = \max\{d(A, B), d(B, A)\} \quad (5-7)$$

其中, $d(A, B)$ 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：

$$d_k(A, B) = K_{a_i \in A}^{th} d_B(a_i) \quad (5-8)$$

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：

$$d_{MHD}(A, B) = \frac{1}{N_A} \sum_{a_i \in A} d_B(a_i) \quad (5-9)$$

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：

$$d_{LTS}(A, B) = \frac{1}{k \times N_A} d_B(a_i)_{(m)} \quad (5-10)$$

5.3 非确定性目标跟踪方法

随机性目标跟踪常用的预测算法有 Kalman 滤波以及粒子滤波等算法。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。

5.3.1 基于 Kalman 滤波器的视觉目标跟踪

5.3.1.1 Kalman 滤波的背景介绍

滤波是从被干扰信号中确定出有效信号的一种方法。最早的滤波方法是 1809 年 Gauss 提出的最小二乘法，这是一种基于最小方差的最优滤波方法，由于它不需要信号的统计特性，目前仍在广泛使用。20 世纪 40 年代，Wiener 为了火力控制上的需要提出了一种线性最小方差滤波，这种方法充分利用了信号测量值的统计特性，但是仅适用于平稳随机信号。1960 年，Kalman 用状态方程来描述随机线性系统，推广了 Wiener 滤波模型，提出了一种更为实用的滤波方法，这种方法同时考虑了信号与测量值的基本统计特性，而且由于采用了状态空间的概念，使得它能够估计多维非平稳随机信号。这种滤波方法另一个突出特点是滤波过程采用递推计算，不需要存储历史数据，利用数字计算机，能够实时的给出信号的估计，而且对于状态维数较大的系统，也能够在计算机上实现，这就是 Kalman 滤波。卡尔曼滤波是在线性无偏最小方差估计原理下推导得出的一种递推滤波方法，它引入了状态空间的概念，借助系统的状态估值和当前时刻的观测值递推新的状态估值。

5.3.1.2 Kalman 滤波器基本原理及其扩展

Kalman 滤波器是一种对动态系统的状态序列，求线性最小方差误差估计的算法，利用动态的状态方程和观测方程来描述系统。它的基本思想是：首先建立描述随机动态变量随时间变化的先验模型，然后在对随机变量进行实时观测的情况下，利用 KF (Kalman filter, KF) 方程组实时获得目标状态基于全局信息的最优估计值。它不仅能方便地通过协方差矩阵对估计精度进行度量，而且能通过残差向量的变化，判断假设的目标运动模型与实际的目标运动特性是否符合。它采用递归滤波的方法进行计算，具有计算量小和能实时计算的特点，因此在目标跟踪系统中得到了很普遍的应用。

Kalman 滤波的基本算法如下：对于状态空间 $x \in \mathfrak{R}^n$ ，Kalman 滤波可以对它的值进行最优估计。加入一个离散时间控制过程的系统，此系统可通过一个线性随机微分方程描述，我们称

为状态方程:

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad (5-11)$$

再加上系统的测量值:

$$z_k = Hx_k + v_k \quad (5-12)$$

随机变量 w_k 和 v_k 分别代表过程噪声和测量噪声。假设它们相互独立, 且服从公式一下的统计特性:

$$\begin{aligned} E[w_k] &= 0, & E[w_k w_l^T] &= Q_k \delta_{kl} \\ E[v_k] &= 0, & E[v_k v_l^T] &= R_k \delta_{kl} \\ E[W_k V_l^T] &= 0 \end{aligned} \quad (5-13)$$

其中 Q_k 是系统过程噪声 w_k 的 $p \times p$ 维对称非负定方差矩阵, R_k 是系统观测噪声 v_k 的 $m \times m$ 维对称正定方差矩阵, δ_{kl} Kronecker- δ 函数, 在实际的应用中, 过程噪声协方差矩阵 Q_k 和观测噪声协方差矩阵 R_k 应该是不断改变的, 在此我们假设它们为常数, 分别是 Q 和 R 。定义为 $\hat{x}_k^- \in \mathfrak{R}^n$ 根据状态转移方程前 $k-1$ 步的先验知识得出的先验状态估计, 定义 $\hat{x}_k \in \mathfrak{R}^n$ 为第 k 步结合测量值 z_k 得出的后验状态估计, 它们的误差分别为公式 (5-14) 和公式 (5-15):

$$e_k^- \equiv x_k - \hat{x}_k^- \quad (5-14)$$

$$e_k \equiv x_k - \hat{x}_k \quad (5-15)$$

那么先验估计误差的协方差为公式:

$$P_k^- = E[e_k^- e_k^{-T}] \quad (5-16)$$

而后验估计误差的协方差为公式:

$$P_k = E[e_k e_k^T] \quad (5-17)$$

为了得到卡尔曼滤波的式子，我们首先找到一个计算后验状态估计 \hat{x}_k 的等式，要求计算 \hat{x}_k 的公式是一个先验估计 \hat{x}_k^- 和一个加权差分的线性组合，这个加权差分是一个真实测量值 z_k 和测量预测值 $H\hat{x}_k^-$ 之间的加权差分。

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \quad (5-18)$$

其中，差分式 $(z_k - H\hat{x}_k^-)$ 称为新息，也称为残差。残差表示预测值和真实测量值之间的偏差。如果残差为零，意味着这两者之间没有误差。公式(5-18)就是系统状态更新方程，其中 K 为卡尔曼滤波的增益矩阵，也可以看作残差的权重。这里的关键就是求出 K ，它可以通过最小化后验的估计误差协方差 P_k 来获得，将式(5-18)代入公式(5-15)求出 e_k ，然后把式 e_k 代入式(3.12)，计算 P_k 并对其求关于 K 的导数，令该导数为零，即可得出：

$$\begin{aligned} K_k &= P_k^- H^T (H P_k^- H^T + R)^{-1} \\ &= \frac{P_k^- H^T}{H P_k^- H^T + R} \end{aligned} \quad (5-19)$$

由公式(5-19)可以看出，当测量误差的协方差 R 趋向零时，Kalman 滤波增益 K 越来越大，在计算 \hat{x}_k 时，也对残差赋予更高的权重，有：

$$\lim_{R_k \rightarrow 0} K_k = H^{-1} \quad (5-20)$$

另一方面，当先验估计误差协方差 P_k^- 趋于零时，Kalman 滤波增益 K 越来越小，在计算 \hat{x}_k 时，将更加信任先验估计 \hat{x}_k^- ，有：

$$\lim_{P_k^- \rightarrow 0} K_k = 0 \quad (5-21)$$

综合公式(5-18)至公式(5-21)可以看出，当测量误差协方差 R 趋于零时，真正的测量值 z_k 是越来越可靠的，而预测值 $H\hat{x}_k^-$ 的作用越来越小。另一方面，当先验估计误差协方差 P_k^- 趋于零，真实测量值 z_k 的作用越来越小，而最终结果越来越倾向于预测值 $H\hat{x}_k^-$ 。

卡尔曼滤波利用反馈控制系统估计运动状态：滤波器估计某一时间的状态，并获得该状态

的预测值。也就是说，卡尔曼滤波分为预测和更新两个步骤。预测是利用当前的状态和误差协方差估计得到先验估计，为下一步的时间状态所用；而更新是反馈部分，将新的观测值和先验估计一起考虑，从而获得后验估计。通过上述论证，可以得到完整的卡尔曼滤波公式，如下：

状态一步预测：

$$\hat{x}_k^- = A\hat{x}_k + Bu_{k-1} \quad (5-22)$$

一步预测误差方差阵：

$$P_k^- = AP_{k-1}A^T + Q \quad (5-23)$$

滤波增益矩阵：

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad (5-24)$$

状态修正：

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \quad (5-25)$$

估计误差方差阵：

$$P_k = (I - K_k H)P_k^- \quad (5-26)$$

公式(5-22)至公式(5-26)为随机线性离散系统中 Kalman 滤波基本方程。在假定初值 \hat{x}_0 和 P_0 后，根据 k 时刻的观测值 z_k 可递推地计算出 k 时刻的状态估计 \hat{x}_k ($k=1,2,\dots$)。

从 Kalman 滤波在使用系统信息和贯穿信息的先后次序，可以看出，在一个滤波周期内 Kalman 滤波具有两个信息更新过程：时间更新过程和观测更新过程公式(3.17)给出了根据 $k-1$ 时刻的状态估计 k 时刻状态的方法，公式(5-22)定量描述了这种预测的效果。在这两个式子的计算中，仅使用了状态一步转移矩阵、噪声输入阵、过程噪声方差阵等与系统的动态特性有关的信息。从时间的推移过程，可以看出，这两个式子将时间从 $k-1$ 时刻推进到 k 时刻，详细给出了 Kalman 滤波的时间更新过程。公式(5-23)、(5-24)以及(5-25)用于计算时间更新值的修正量，这个修正量由四个值确定，分别是时间更新的质量优劣 P_k^- 、观测信息的质量优劣 R 、观测与

状态的关系 H 以及具体的观测信息 z_k 。综上所述,所有方程都围绕着正确合理地利用观测 z_k 这个目的,因此这个过程描述了 Kalman 滤波的观测更新过程 Kalman 滤波算法见图 5.5。从图 5.5 中可以看出, Kalman 滤波具有增益计算。回路和滤波计算回路这两个计算回路。其中,增益计算回路可独立计算,而滤波计算回路则依赖于增益计算回路。

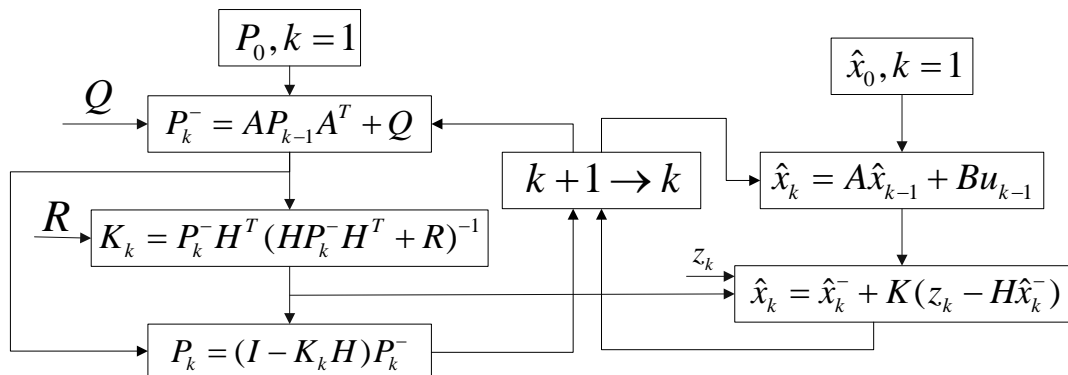


图 5.5 Kalman 滤波算法框图

Kalman 滤波具有以下几个性质:

Kalman 滤波算法把估计信号视为在高斯白噪声的作用下随机线性系统的输出,而且由状态方程和输出方程在时间域内给出它的输入输出关系,所以这种滤波方法不但适用于平稳序列的滤波,更加适用于平稳马尔可夫序列、非平稳马尔可夫序列,以及高斯—马尔可夫序列的滤波。

Kalman 滤波的基本方程是一种递推形式,其计算过程是通过不断地“预测-修正”,不需要存储大量的数据,只要观测到了新的数据,可立即计算出新的滤波值,所以非常便于实时处理。

Kalman 滤波器的增益矩阵与观测值无关,可预先离线算出它,可减少实时在线的计算量;对于滤波器增益矩阵 K_k 的计算,要求计算 $(HP_k^- H^T + R)^{-1}$, 矩阵的阶数只取决于观测方程的维数 m (m 通常很小),所以上面的求逆运算比较简单;在求解 Kalman 增益矩阵时,可同时计算出滤波器的精度指标 P_k , P_k 对角线上的元素即为滤波误差向量各分量的方差。

增益矩阵 K_k 与初始方差阵 P_0 系统噪声方差阵 Q 以及观测噪声方差阵 R 之间关系如下: 由 Kalman 滤波的基本方程(5-22)和(5-23),可以得出, P_0 , Q , R 同乘以一个常数时,增益矩阵 K_k 的值不变。在一维的情况下,当测量噪声协方差矩阵增大时,增益矩阵 K_k 反而变小。说明,测量噪声越大,增益取的越小,减弱了测量噪声对估计值的影响,导致预测值在最后的結果中占得比重加大。

当 P_{k-1} 或者 Q 矩阵两者之一变小, 或两者同时变小时, P_k^- 和 $K_k P$ 变小表示估计值较好, Q 变小表示状态转移的随机波动减小, 新的测量值减弱了状态估计值的矫正影响, 所以增益矩阵 K_k 应当变小。通过上面的分析, 可知: 增益矩阵 K_k 与 Q 成正比, 与 R 成反比。即: 当 R 越大时, 测量噪声越大, 所以 K_k 要变小, 以使测量值在最后估计结果中所占的比重较小; 当 Q 比较大则状态噪声较大, 此时预测值被严重干扰, 所以 K_k 值应该变大, 以保证预测值在最后估计结果中所占的比重较小。总而言之, 状态噪声越大, 则状态噪声协方差矩阵 Q 的值也越大, 更有利于捕捉目标机动状态, 滤波收敛速度快, 但得到的滤波结果的精度较差; 状态噪声越小, 状态噪声协方差矩阵 Q 的值越小, 得到的滤波结果的精度较好, 但是滤波的收敛速度慢, 而且如果目标进行大机动状态运动, 跟踪丢失的可能性较大。

5.3.1.3 Kalman 滤波器在跟踪领域中的应用

Kalman 滤波器对于提高跟踪系统的处理速度和性能有着极其重要的作用。假设目标的运动状态参数为某一时刻目标的位置和速度。在跟踪过程中, 由于相邻两帧图像的时间间隔比较短, 目标在这么短的时间间隔内状态变化比较小, 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 $(s_x, s_y, v_x, v_y)^T$, 分别代表运动目标在坐标轴上的位置和在坐标轴方向的速度。根据运动目标在单位时间间隔内是匀速运动的, 定义状态转移矩阵 A 和观测矩阵 H :

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5-27)$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

5.3.2 基于粒子滤波器的视觉目标跟踪

5.3.2.1 粒子滤波的背景介绍

近年来随着计算机的计算成本的不断下降以及计算能力的不断提高,以非参数化的 Monte-Carlo 模拟为特色的粒子滤波(Particle filtering)方法逐渐成为非线性、非高斯随机系统估计问题的研究热点和有效方法。EKF 对系统的非线性模型做线性近似后再进行线性变换,而粒子滤波直接利用系统非线性模型做非线性变换,这就体现了一种先进的思想,即非线性估计算法更接近系统的非线性本质。传统的非线性随机系统估计方法大都是基于高斯假设的,对于非高斯分布的非线性随机系统其估计性能较差。在实际情况中非线性、非高斯随机系统估计问题具有更为普遍的意义。粒子滤波采用样本形式而不是函数形式,对非线性随机系统状态的概率分布进行描述。它不需要对状态变量的概率分布作过多的约束,是目前较适合于非线性、非高斯随机系统估计问题的滤波方法。由于粒子滤波在处理非线性、非高斯随机系统估计问题方面具有独到的优势,国外已经开始将其应用于视频与图像处理、导航与定位、目标跟踪、无线通信、状态监测和故障诊断等众多领域中。

由于粒子滤波是近年来出现的新方法,其算法本身还不成熟,因此仍有大量的问题亟待解决。例如,直接影响粒子滤波性能的重要性函数选取问题、粒子多样性匮乏问题、粒子退化问题、粒子滤波算法的收敛性问题、与其他非线性滤波方法的结合使用问题、粒子滤波算法的硬件实现问题、拓展粒子滤波新的应用领域问题等。目前,粒子滤波算法在国外发展很快,并取得了许多研究成果,而国内相关研究则刚刚起步。若能对上述问题进行深入研究,必将促进粒子滤波方法及其应用的发展,在非线性、非高斯系统的数据分析和处理领域中紧跟世界发展趋势。

估计问题一般分为三类:利用当前和过去的观测值来估计未知量的当前值,称为滤波;利用过去的观测值来估计未知量的将来值,称为预测或外推;利用过去的观测值来估计未知量的过去值,称为平滑或内插。最优估计是指按照某种最优准则来估计未知量的值。通常的估计准则和方法有:最小二乘估计、极大似然估计、极大验后估计、最小方差估计、线性最小方差估计等。

随着计算机计算能力的快速增长和计算成本的不断降低,粒子滤波(Particle Filter, PF)逐渐

成为当前研究非线性、非高斯随机系统状态估计问题的一个热点和有效方法。粒子滤波使用了大量随机样本，采用蒙特卡罗仿真(Monte Carlo Simulation)来完成递推贝叶斯滤波(Recursive Bayesian Filter)过程，其核心是使用一组具有相应权值的随机样本（粒子）来表示状态的后验分布。该方法的基本思路是选取一个重要性概率密度并从中进行随机抽样，得到一些带有相应权值的随机样本后，在状态观测的基础上调节权值的大小和粒子的位置，再使用这些样本来逼近状态后验分布，最后通过这组样本的加权求和作为状态的估计值。粒子滤波不受系统模型的线性和高斯假设约束，采用样本形式而不是函数形式对状态概率密度进行描述，使其不需要对状态变量的概率分布作过多的约束，适用于任意非线性非高斯动态系统，是目前最适合于非线性、非高斯系统状态的滤波方法。

早在 20 世纪 50 年代，一种通过离散随机样本逼近概率分布的蒙特卡罗方法(Monte Carlo)——序贯重要性采样(Sequential Importance Sampling, SIS)算法就出现了。SIS 算法由预测和更新两部分组成，是一种基于贝叶斯理论的时域递归滤波实现方法。该算法的基本思想是从一个重要性概率密度中抽取一个带有权值的样本集合来代表所研究的系统状态的后验概率密度函数，并利用这些样本及其权值对该系统状态进行估计。该算法随后被应用到物理和自动控制等领域。但由于 SIS 对于当时计算机的计算能力而言的计算复杂性太高，并且算法本身存在其固有的样本退化问题，因此该算法在相当长一段时间内未能引起人们广泛的关注，其算法本身没有取得较大进展。随着现代计算技术的迅速发展，现代计算技术为基于 Monte Carlo 模拟的滤波方法提供了有力支持，SIS 算法再次引起人们的关注。1993 年 Gordon 等为了克服 SIS 算法中的样本退化问题，将重采样(resampling)步骤引入 SIS 算法，并由此产生了基本粒子滤波算法——序列重要性重采样算法(sequential importance resampling, SIR)。样本（又称为粒子）退化是指经过若干次迭代后，大部分粒子的权值极小以至于可以忽略不计，粒子权值的方差随时间逐渐增大，大量运算时间浪费在这些对系统状态估计几乎不起作用的粒子权值更新上。SIR 算法在 SIS 算法的基础上增加了重采样步骤，由预测、更新和重采样三部分组成。

SIR 算法的重采样在一定程度上缓解了粒子退化问题，但是它的重采样在原有粒子集合组成的离散分布上依粒子权值大小，采取随机选取复制的方式进行，即采用以较大概率复制权值较高粒子的方法。这样就使原有粒子集合中有很多粒子由于权值太小没有“后代”，而少数权值

较大的粒子则有很多相同的“后代”，导致重采样得到的粒子集合由大量重复粒子构成，使粒子集合失去了多样性，从而限制了粒子并行运行的机会。这一现象又称为粒子多样性匮乏问题。粒子多样性匮乏将导致滤波精度的下降，并且使基于粒子路径多样性的系统状态平滑操作性能下降。1998年Liu等人提出一种称为有效样本数量的指标来度量粒子退化的程度。解决粒子退化问题的另一个方法是选择更接近系统状态真实概率密度的重要性概率密度，但往往又难于从这样的重要性概率密度中进行抽样。此后针对粒子退化问题和粒子多样性匮乏问题的研究主要集中在：(1)优化重采样，减小粒子多样性匮乏；(2)产生既接近系统状态真实概率密度又便于从中抽样的重要性概率密度。

作为近年来出现的新方法，粒子滤波算法本身还不成熟，仍有大量需要研究改进的方面，主要体现在以下几点：

1、重要性密度函数的选取方法有待进一步研究。重要性密度函数的选取直接影响粒子滤波性能的高低。目前大多数改进算法的重要性密度函数选取方法对通用问题缺乏一般借鉴意义。如何有效利用系统信息和最新观测值来选取合理的重要性密度函数是一个重要的研究方向。

2、重采样方法有待进一步改进。目前大多数重采样改进算法仅在特定条件下对粒子滤波性能有所改善，缺乏较强的普遍适用性。虽然粒子滤波器的重采样缓解了粒子退化程度，但使粒子失去了多样性，导致平滑操作性能下降，如何设计有效的粒子平滑算法值得进一步研究。

3、粒子滤波算法的计算量较大。通常的重采样方法存在计算量较大的问题。粒子滤波的计算量还与粒子数目和状态维数都直接相关。如何在满足估计精度的前提下减少粒子滤波的计算量仍有待于今后深入的研究。此外，在实际应用中如何通过硬件方法实现粒子滤波的并行处理具有重要的现实意义。

4、粒子滤波的收敛性问题尚需进一步深入研究。目前有关粒子滤波收敛性方面的研究成果已经为粒子滤波的应用提供了一定的理论依据，但是重要的一致性收敛证明仍依赖于对动态模型作出较强的假设，而这些假设条件在很多实际问题中是不成立的。若能解决上述问题，则能在数学基础上对粒子滤波提供更有力的支持，并能为粒子退化及多样性匮乏现象的抑制提供有益的指导。

5.3.2.2 几种典型的粒子滤波改进方法

辅助粒子滤波：SIR 算法虽然容易抽样、简单易求，但由于没有考虑到系统状态的最新观测值，其抽样具有一定的盲目性，从而限制了滤波性能的进一步提高。1999 年 Pitty 和 Shephard 提出辅助粒子滤波（Auxiliary Particle Filter,APF）算法，该算法以 SIS 算法为基础，通过引入一个辅助变量 $u_k(i)$ 得到一个考虑了系统状态的最新观测值 y_k 的重要性概率密度如下：

$$q(x_k | x_{0:k-1}(i), y_{1:k}) = w_{k-1}(i) p(y_k | u_k(i)) p(x_k | x_{k-1}(i)) \quad (5-28)$$

其中，辅助变量 $u_k(i)$ 代表在给定 $x_{k-1} = x_{k-1}(i)$ 下 x_k 的某些特征，通常是 x_k 的均值：

$$u_k(i) = E[x_k | x_{k-1}(i)] \quad (5-29)$$

也可以是 $p(x_k | x_{k-1}(i))$ 中的一个抽样值即 $u_k(i) \sim p(x_k | x_{k-1}(i))$ 。而利用上述重要性概率密度抽样得到的粒子集合表示为 $\{x_k(i), j^i\}_{i=1}^N$ ，其中 j^i 代表 $k-1$ 时刻粒子的序号， $x_k(i)$ 代表由粒子 $x_{k-1}(j^i)$ 经过系统状态转移方程得到的粒子。 $x_k(i)$ 的权值 $w_k^*(i)$ 如下：

$$w_k^*(i) = \frac{p(y_k | x_k(i))}{p(y_k | u_k(j^i))} \quad (5-30)$$

辅助粒子滤波算法执行步骤如下：

- 1、计算每个粒子的辅助变量 $u_k(i)$ ： $u_k(i) = E[x_k | x_{k-1}(i)]$ ， $i=1,2,\dots,N$ ；
- 2、计算每个粒子的辅助变量的权值 $\pi_k^*(i)$ ： $\pi_k^*(i) = w_{k-1}(i) p(y_k | u_k(i))$ ，并计算

其归一化权值 $\pi_k(i)$ 如下：

$$\pi_k(i) = \frac{\pi_k^*(i)}{\sum_{i=1}^N \pi_k^*(i)} \quad (5-31)$$

其中， $i=1,2,\dots,N$ ；

- 3、据上述归一化的辅助变量权值对 $k-1$ 时刻的粒子集合 $\{x_{k-1}(i)\}_{i=1}^N$ 进行重采样，得到粒子

集合 $\{x_{k-1}(j^i)\}_{i=1}^N$;

4、用系统状态转移概率密度得到 k 时刻的粒子集合, 即 $x_k(i) \sim p(x_k | x_{k-1}(j^i))$, $i=1,2,\dots,N$;

5、用 3 式计算 $x_k(i)$ 的权值 $w_k^*(i)$ 并利用计算其归一化权值 $w_k(i)$;

6、估计:

$$\hat{x}_k = \sum_{i=1}^N w_k(i) x_k(i) \quad (5-32)$$

由辅助粒子滤波算法的执行步骤可见, 该算法利用辅助变量 $u_k(i)$ 作为系统状态的一步预测值, 用 $p(y_k | u_k(i))$ 计算出各一步预测值的似然值, 再通过重采样选择出预测似然值较大的粒子进行粒子状态转移, 其重要性抽样过程结合了最新观测值 y_k , 从而得到的粒子接近于真实状态的可能性更高。实际应用中, 当系统状态转移噪声较小时 APF 算法好于 SIR 算法。与 SIR 算法不同, 为了考虑系统状态的最新观测值 y_k , 辅助粒子滤波算法中的重采样发生在上一时刻。

正则化粒子滤波: SIR 算法的重采样在一定程度上缓解了粒子退化问题, 但同时又引入了一个新问题即粒子多样性匮乏。SIR 算法的重采样在原有粒子集合组成的离散分布上, 依粒子权值大小采用随机选取复制的方式进行。这样, 原有粒子集合中很多粒子由于权值太小没有“后代”, 而少数几个权值较高的粒子则有很多相同的“后代”, 重采样以后的粒子集合由大量重复的粒子构成, 使粒子集合失去了多样性, 从而限制了粒子并行运行的机会。这一现象又称为粒子多样性匮乏问题。Musso 等针对该问题提出了正则化粒子滤波 (regularized particle filter, RPF) 算法。该算法在一个系统状态的连续近似分布上进行重采样, 以克服粒子多样性匮乏问题。RPF 算法和 SIR 算法除了重采样步骤不同外, 其它步骤一致。SIR 算法的重采样是在公式(5-28)所示的系统状态离散近似分布上进行重采样, 而 RPF 算法的重采样是在如下所示的系统状态连续近似分布上进行重采样:

$$p(x_k | y_{1:k}) \approx \hat{p}(x_k | y_{1:k}) = \sum_{i=1}^N w_k(i) K_h(x_k - x_k(i)) \quad (5-33)$$

其中, $w_k(i)$ 是归一化后的权值。

$$K_h(x) = \frac{1}{h^n} K\left(\frac{x}{h}\right) \quad (5-34)$$

这里的 $K_h(\cdot)$ 是一个对核密度 (Kernel Density) 函数 $K(\cdot)$ 重新标度 (Rescaled) 过的核密度函数; $h>0$ 称为核带宽; n 是变量 x 的维数; $K(\cdot)$ 是一个对称的概率密度函数, 满足:

$$\int x K(x) dx = 0 \quad (5-35)$$

$$\int \|x\|^2 K(x) dx < \infty \quad (5-36)$$

选择核函数 $K(\cdot)$ 和核带宽 h 时, 应力求使真实概率密度和(5-35)式等号右端的正则化近似密度之间的平均积分平方误差 (mean integrated square error, MISE) 最小, 该平均积分平方误差定义如下:

$$MISE = E[\int [p(x_k | y_{1:k}) - \hat{p}(x_k | y_{1:k})]^2 dx_k] \quad (5-37)$$

在所有粒子归一化权值 $w_k(i)$ 均相等的情况下 (即有 $w_k(i) = 1/N, i = 1, 2, \dots, N$), 最优核密度函数 $K(\cdot)$ 是 Epanechnikov 核密度, 即

$$K = \begin{cases} \frac{n+2}{2c_n} (1 - \|x\|^2), & \text{如果 } \|x\| < 1 \\ 0, & \text{否则} \end{cases} \quad (5-38)$$

其中, c_n 是 n 维空间上单位超球体的体积。为了减少运算量, 也可使用高斯核密度。当真实概率密度 $p(x_k | y_{1:k})$ 是具有单位协方差矩阵的高斯概率密度时, 核带宽 h 的最优选择为:

$$h_{opt} = AN^{\frac{1}{n+4}} \quad (5-39)$$

其中,

$$A = \left[8c_n^{-1} (n+4) (2\sqrt{\pi})^n \right]^{\frac{1}{(n+4)}} \quad (5-40)$$

尽管只有粒子权值相等且真实概率密度为高斯密度时公式(5-37)至(5-40)是最优的, 但一般情况下仍然可以利用这些结果将 RPF 作为次优滤波器使用。RPF 算法的执行步骤如下所示:

- 1、从系统状态转移概率密度中抽取新粒子 $x_k^*(i)$: $x_k^*(i) \sim p(x_k | x_{k-1}(i))$, ($i=1, 2, \dots, N$);

2、 计算各 $x_k^*(i)$ 的权值 $w_k^*(i)$ ，再归一化上述权值得到 $w_k(i)$ ；

3、 状态估计：

$$\hat{x}_k = \sum_{i=1}^N w_k(i) x_k(i) \quad (5-41)$$

4、 计算粒子的经验方差 S_k ，求出满足 $D_k(D_k)^T = S_k$ 的 D_k ；

5、 在 $\{x_k^*(i), w_k(i)\}_{i=1}^N$ 上进行重采样得到 $\{x_k'(i)\}_{i=1}^N$ ；

6、 从 Epanechnikov 核密度中抽取出样本 $e_k(i)$ ，计算 $x_k(i) = x_k'(i) + h_{opt} D_{ke_k(i)}$ ， $x_k(i)$

即是从(6)式所示的连续近似分布上重采样得到的粒子($i=1,2,...,N$)。

RPF 算法可以有效地缓解重采样过程造成的粒子多样性匮乏问题。同 SIR 算法的重采样步骤相比，RPF 算法的重采样主要附加了 N 次从核密度中抽样的过程。因此，在算法计算量上这两种算法相差不多。在粒子多样性匮乏较为严重的情况下（例如当系统状态转移噪声较小时），RPF 算法的估计精度优于 SIR 算法。但 RPF 算法有一个理论上的缺陷，即它所得的离散近似分布不能保证是渐进地接近系统状态真实分布。

高斯粒子滤波和高斯混合粒子滤波：该算法在粒子滤波的基本框架下用单个高斯分布去近似系统状态分布，是高斯滤波器的变种。高斯滤波器是一类采用高斯概率密度去近似被估计概率密度的滤波器，例如 EKF 和它的变种。高斯滤波器的每一次递归更新都只传播所求概率密度的均值和方差，其理由是：在上述高斯假设下只有均值和方差需要被跟踪，只需要给定均值和方差即可确定所求的概率密度。高斯滤波应用于非线性问题的效果依赖于问题的非线性程度，在模型具有较强的非线性或当被估计概率密度为多峰模式时它的性能较差。高斯粒子滤波结合粒子滤波和高斯滤波的优点，使用重要性采样来近似未知变量的均值和方差。

一个 m 维高斯随机变量 X 的概率密度可以写作：

$$N(X; \mu, \Sigma) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \left(-\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu) \right) \quad (5-42)$$

其中， μ 是其均值， Σ 是其方差阵。假设在 $k-1$ 时刻时我们得到 $p(x_{k-1} | y_{1:k-1})$ 的近似如下：

$$p(x_{k-1} | y_{1:k-1}) \approx \hat{p}(x_{k-1} | y_{1:k-1}) = N(x_{k-1}; \mu_{k-1}, \Sigma_{k-1}) \quad (5-43)$$

当新的观测值到来时我们可以通过粒子滤波的预测步骤和更新步骤得到 $p(x_k | y_{1:k})$ 的近似如下:

$$p(x_k | y_{1:k}) \approx \hat{p}(x_k | y_{1:k}) = N(x_k; \mu_k, \Sigma_k) \quad (5-44)$$

大多数情况下, 上式中的均值 μ_k 和方差 Σ_k 的解析表达式得不到。GPF 算法利用粒子滤波中的粒子及其权值来求 μ_k 和 Σ_k 的估计值。GPF 算法的执行步骤如下所示:

- 1、从 $N(x_{k-1}; \mu_{k-1}, \Sigma_{k-1})$ 中抽样得到样本集合 $\{x_{k-1}(i)\}_{i=1}^N$;
- 2、分别从 $p(x_k | x_{k-1}(i))$ ($i=1, 2, \dots, N$) 中抽样, 得到样本集合 $\{x_k^*(i)\}_{i=1}^N$;
- 3、计算样本 $\{x_k^*(i)\}_{i=1}^N$ 的均值 μ_k^* 和方差 Σ_k^* 如下:

$$\mu_k^* = \frac{1}{N} \sum_{i=1}^N x_k^*(i) \quad (5-45)$$

$$\Sigma_k^* = \frac{1}{N} \sum_{i=1}^N (\mu_k^* - x_k^*(i))(\mu_k^* - x_k^*(i))^T \quad (5-46)$$

- 4、从重要性概率密度 $q(x_k | y_{1:k})$ 中抽取样本 $\{x_{k-1}(i)\}_{i=1}^N$;
- 5、得到观测值 y_k 后利用下式计算各 $x_k(i)$ 的权值 $w_k^*(i)$ 。

$$w_k^*(i) = \frac{p(y_k | x_k(i))N(x_k(i); \mu_k^*, \Sigma_k^*)}{q(x_k(i) | y_{1:k})} \quad (5-47)$$

- 6、归一化权值 $w_k^*(i)$ 得到 $w_k(i)$ 。

$$w_k(i) = \frac{w_k^*(i)}{\sum_{i=1}^N w_k^*(i)} \quad (5-48)$$

- 7、利用 $\{x_k(i), w_k(i)\}_{i=1}^N$ 计算均值 μ_k 和方差 Σ_k 如下:

$$\mu_k = \sum_{i=1}^N w_k(i) x_k(i) \quad (5-49)$$

$$\Sigma_k = \sum_{i=1}^N w_k(i)(\mu_k - x_k(i))(\mu_k - x_k(i))^T \quad (5-50)$$

从而得到 k 时刻 $p(x_k | y_{1:k})$ 的近似如下:

$$p(x_k | y_{1:k}) \approx N(x_k; \mu_k, \Sigma_k) \quad (5-51)$$

其中, 均值 μ_k 可以作为 k 时刻 x_k 的估计值。

通常, GPF 算法中重要性概率密度 $q(x_k | y_{1:k})$ 可以采用 $N(x_k; \mu_k^*, \Sigma_k^*)$ 或 $p(x_k | x_{k-1}(i))$ 。

文献[32]证明了当所求概率分布服从高斯分布时, 随粒子数目的增加 GPF 算法为渐进最优算法, 其估计值可以渐进地收敛到最小均方估计。粒子滤波算法计算量的一个重要组成部分来自于重采样过程。在高斯假设成立时 GPF 不会产生粒子退化问题, 从而 GPF 算法无重采样步骤, 使算法计算量更低。这使 GPF 在计算速度上优于前述的几种粒子滤波算法。实际应用中, GPF 算法的估计精度优于前述几种粒子滤波算法。另外, GPF 并不要求系统噪声和观测噪声是加性的, 它还可以应用于非加性噪声问题。

当所求概率分布为非高斯分布时, 采用单个高斯分布往往不能获得有效的近似效果。针对该问题, 文献[34]中将高斯混合滤波和粒子滤波结合, 提出了高斯混合粒子滤波 (Gaussian sum particle filter, GSPF) 算法。该算法使用有限个高斯密度的加权求和来近似所求的非高斯概率密度, 各高斯密度的均值和方差的估计值利用粒子滤波中的粒子及其权值来求得。

假设在 k 时刻 $p(x_{k-1} | y_{1:k-1})$ 可以近似为如下 G 个高斯密度的加权求和:

$$p(x_{k-1} | y_{1:k-1}) \approx \hat{p}(x_{k-1} | y_{1:k-1}) = \sum_{j=1}^G W_{k-1,j} N(x_{k-1}; \mu_{k-1,j}, \Sigma_{k-1,j}) \quad (5-52)$$

其中, $W_{k-1,j}$ 代表第 j 个高斯密度的权值, $\mu_{k-1,j}$ 和 $\Sigma_{k-1,j}$ 分别代表第 j 个高斯密度的均

值和方差。当新的观测值到来时, 通过粒子滤波的预测步骤和更新步骤可以得到 $p(x_k | y_{1:k})$ 的近似如下:

$$p(x_k | y_{1:k}) \approx \hat{p}(x_k | y_{1:k}) = \sum_{j=1}^G W_{k,j} N(x_k; \mu_{k,j}, \Sigma_{k,j}) \quad (5-53)$$

GSPF 算法的执行步骤如下所示：

1、分别从 G 个高斯密度 $\{N(x_{k-1}; \mu_{k-1}, \Sigma_{k-1}, j) | j=1, 2, \dots, G\}$ 中抽样，得到 G 个样

本集合 $\{(x_{k-1,j}(i))_{i=1}^N, j=1, 2, \dots, G\}$ ；

2、分别从 $p(x_k | x_{k-1,j}(i))$ ($i=1, 2, \dots, N; j=1, 2, \dots, G$) 中抽样，得到 G 个样本集合

$\{(x_{k,j}^*(i))_{i=1}^N, j=1, 2, \dots, G\}$ ；

3、分别计算步骤 2 中的 G 个样本集合的均值 $\mu_{k,j}^*$ 和方差 $\Sigma_{k,j}^*$ ；

4、从重要性概率密度 $q(x_k | y_{1:k})$ 中抽样得到 G 个样本集合 $\{(x_{k,j}(i))_{i=1}^N, j=1, 2, \dots, G\}$ ；

5、得到观测值 y_k 后利用下式计算各样本 $x_{k,j}(i)$ 的权值 $W_{k,j}^*$ 。

$$W_{k,j}^* = \frac{p(y_k | x_{k,j}(i))N(x_{k,j}(i); \mu_{k,j}^*, \Sigma_{k,j}^*)}{q(x_{k,j}(i) | y_{1:k})} \quad (5-54)$$

(6) 计算 G 个样本集合 $\{(x_{k,j}(i))_{i=1}^N, j=1, 2, \dots, G\}$ 的均值 $\mu_{k,j}$ 和方差 $\Sigma_{k,j}$ 如下：

$$\mu_{k,j} = \frac{\sum_{i=1}^N w_{k,j}^*(i) x_{k,j}(i)}{\sum_{i=1}^N w_{k,j}^*(i)} \quad (5-55)$$

$$\Sigma_{k,j} = \frac{\sum_{i=1}^N w_{k,j}^*(i) (\mu_{k,j} - x_{k,j}(i)) (\mu_{k,j} - x_{k,j}(i))^T}{\sum_{i=1}^N w_{k,j}^*(i)} \quad (5-56)$$

(7) 更新 G 个高斯密度的权值 $W_{k,j}^*$ 并计算其归一化权值 $W_{k,j}$ 如下：

$$W_{k,j}^* = W_{k-1,j}^* = \frac{\sum_{i=1}^N w_{k,j}^*(i)}{\sum_{j=1}^G \sum_{i=1}^N w_{k,j}^*(i)}, \quad j=1, 2, \dots, G \quad (5-57)$$

$$W_{k,j} = \frac{W_{k,j}^*}{\sum_{j=1}^G W_{k,j}^*} \quad (5-58)$$

从而得到 k 时刻 $p(x_k | y_{1:k})$ 的近似如下:

$$p(x_k | y_{1:k}) \approx \sum_{j=1}^G W_{k,j} N(x_k; \mu_{k,j}, \Sigma_{k,j}) \quad (5-59)$$

利用该近似密度可以计算 x_k 的估计值如下:

$$x_k = \sum_{j=1}^G w_{k,j} \mu_{k,j} \quad (5-60)$$

GSPF 算法中, 通常选择 $N(x_k; \mu_{k,j}^*, \Sigma_{k,j}^*)$ 或 $q(x_k | x_{k-1}(i))$ 作为重要性概率密度 $q(x_k | y_{1:k})$ 。

上述 GSPF 算法也存在退化现象, 即算法经过多次迭代后, 高斯密度权值 $W_{k,j}$ 的方差会随时间逐渐增大, 使得少数高斯密度权值很大而大多数高斯密度权值很小, 甚至所求概率密度被近似为一个高斯密度。针对该问题, 可以在上述 GSPF 算法的第(7)步后增加一个重采样步骤, 其与前述粒子滤波的重采样方法很类似, 只不过这里是从 $\{(\mu_{k,j}, \Sigma_{k,j}), W_{k,j}\}_{j=1}^G$ 中进行重采样, 重采样后各高斯密度的权值均等于 $1/G$ 。

5.3.2.3 粒子滤波在跟踪领域中的应用

目标跟踪基于自身位置来衡量其它运动目标的位置和速度, 广泛存在于自动控制、航空航天、通信等领域中。实际中大多数目标跟踪问题都是非线性非高斯的。例如, 当运动目标具有一定的机动性时, 对目标位置和速度的估计实质上就是一个非线性系统状态估计问题^[8]。解决目标跟踪问题的传统方法是卡尔曼滤波和扩展卡尔曼滤波等^[9,10]。但在强非线性或噪声非高斯时, 它们在目标跟踪中性能不稳定、误差较大甚至发散。1993年 Gordon 等将 SIR 粒子滤波应用于单站被动纯角度目标跟踪问题上取得了优于扩展卡尔曼滤波的跟踪效果, 从此粒子滤波成为目标跟踪领域中一个重要的研究热点。

针对传统检测前跟踪(TBD)方法(例如动态规划和最大似然估计等^[11,12,13]) 在目标具有机动时性能较差, 文献^[14,15]利用混合估计的思想提出了一种基于粒子滤波器的检测前跟踪方法, 取

得了较好的效果。文献^[16]进一步针对低信噪比下幅值波动的目标跟踪问题,利用粒子滤波得到状态的后验概率分布以及目标的出现概率,用 Bayes 似然比作为目标检测的判决准则,提出一种联合检测和跟踪算法,同时实现对目标的检测与跟踪。仿真结果表明其算法的有效性。文献^[17]对杂波中的目标跟踪问题提出了混合粒子滤波器。文献^[18]给出了闪烁噪声下的粒子滤波跟踪算法。文献^[19]对杂波中的机动目标跟踪问题提出了一组粒子滤波算法,并详细分析了粒子滤波设计参数对滤波性能的影响。文献^[20]提出了一种适用于定位、导航和跟踪的粒子滤波通用框架,其仿真结果优于扩展卡尔曼滤波。文献^[21]把粒子滤波用于移动机器人定位。文献^[22]把粒子滤波用于传感器网络下的协同跟踪。文献^[23]在移动机器人上采用粒子滤波和多传感器融合技术跟踪随机移动人体目标,当目标移动速度在人体正常合理范围内时取得了连续平滑的跟踪效果。文献^[24]研究了粒子滤波中观测似然函数对目标跟踪性能的影响。文献^[25]将粒子滤波应用于多目标跟踪与传感器管理问题。文献^[26]采用多模型粒子滤波算法进行机动目标跟踪。文献^[27]研究了多目标跟踪与数据融合问题。文献^[28]给出了基于粒子滤波的群目标跟踪算法。文献^[29]针对来自多个传感器的无序测量值提出基于粒子滤波的多目标跟踪和信息融合方法。文献^[30]把粒子滤波用于自动目标识别。文献^[31]提出语音源定位的粒子滤波算法。上述粒子滤波用于目标跟踪的研究成果表明,粒子滤波能较好地处理非线性、低信噪比、测量断续和群目标跟踪等传统难题,已经成为目标跟踪领域中的重要技术支撑工具之一。

近年来依据序列图像对感兴趣区域进行持续跟踪的问题引起人们广泛关注。计算机视觉图像跟踪领域的传统难点问题包括:在复杂背景影响下运动目标检测缺乏有效的手段;由于目标被障碍物遮挡导致测量数据出现断续现象,使传统方法出现失跟的问题;密集目标的线路交叉跟踪问题等^[32]。文献^[33]率先将粒子滤波引入计算机视觉跟踪领域。文献^[34,35]提出并系统研究了采用粒子滤波算法如何实现目标的轮廓跟踪。文献^[36]将粒子滤波应用于图像序列中线路交叉目标的跟踪。文献^[37]提出一种解决图像传感器多目标跟踪的粒子滤波算法。文献^[38]和^[39]利用听觉、视觉联合目标定位和跟踪,提出一种声音和视觉融合的粒子滤波算法。文献^[40]提出基于粒子滤波的手形跟踪识别方法。文献^[41]结合视频对象颜色特征和运动特征,在粒子滤波理论框架下提出一种基于颜色特征和运动特征信息融合的视频跟踪算法。该算法可以准确地跟踪上复杂场景中的运动目标和快速运动目标,较采用单一颜色特征的粒子滤波跟踪算法在可靠性、准确性方

面有一定的提高。文献^[42]将均值漂移算法^[43,44]和粒子滤波算法结合,提出了一种基于粒子滤波的自适应运动目标跟踪方法。该方法能够根据跟踪的过程自适应地调整参数,较好地处理图像序列中由于光线变化或遮挡所带来的影响。该方法与均值漂移方法相比,即使在复杂的情形下也能够准确地对目标进行跟踪。文献^[45]针对在一般人脸跟踪算法中用简单椭圆来描述人脸形状线索易受背景干扰的问题,以及视频目标跟踪中动态模型和观测模型的非线性非高斯特点,提出一种以颜色和形状直方图为线索的粒子滤波人脸跟踪算法。该算法在粒子滤波基本框架之下用直方图来描述人脸形状并以此作为人脸跟踪的形状线索。该跟踪方法不仅能有效地处理人脸旋转、背景中的肤色干扰和部分遮掩问题,并且能够解决因大面积遮掩等原因造成的目标丢失问题,重新捕获已丢失的目标。上述粒子滤波用于视觉图像跟踪的研究成果表明,在粒子滤波算法下一些传统的难点问题如目标检测、遮挡、交叉、失跟等得到了更好的解决。

在无线通讯中粒子滤波被广泛用于信道盲均衡、盲检测、多用户检测和单站定位跟踪等方面^[46]。文献^[47,48,49]分别将粒子滤波用于解决时不变信道、时变信道、加性高斯和非高斯信道中的盲均衡问题,验证了粒子滤波方法解决盲均衡问题的有效性。文献^[50,51]分别将粒子滤波应用在CDMA系统的多用户检测中,讨论了衰落系数已知和未知时的多用户检测。移动通信技术的不断发展使蜂窝网移动台定位跟踪问题越来越引起人们的重视^[52,53]。传统的移动台定位方式绝大多数采用多个基站同时进行测量^[54],利用单个基站进行定位跟踪的研究比较少。尽管采用多站测量的方式能提供多组信息从而得到较高精度的位置估计,但是蜂窝网络中由于功率控制的使用和多址干扰等因素,使得同时测量多个基站信号比较困难。而单站定位跟踪具有跟踪实现简单并具有较高的可定位概率。因此,研究单个基站移动台定位跟踪就十分必要^[55]。但目前单站定位跟踪存在精度较低的问题。文献^[56]采用粒子滤波进行蜂窝网移动台单站定位跟踪,大大提高了定位跟踪的精度,与扩展卡尔曼滤波相比粒子滤波收敛速度快,性能更稳健。红外目标跟踪研究在红外制导、侦察和安检等军事与民用应用中具有重要意义。传统的红外目标跟踪方法主要有:模板匹配法^[57]、光流法^[58]、卡尔曼滤波和扩展卡尔曼滤波方法^[59,60]等。模板匹配法简单易行,但匹配过程中搜索空间会随着图像大小和模板大小的增加而急剧增大,从而降低目标跟踪的效率。光流法在强噪声和强杂波图像背景下的跟踪效果不稳健。卡尔曼滤波和扩展卡尔曼滤波在红外目标模型具有较强非线性或噪声非高斯时跟踪容易发散。基于此一些研究人员将粒子滤波引入红外目标跟踪问题中。文献^[61]提出了一种基于粒子滤波的红外目标稳健跟踪新方法,

实验结果表明了该方法的有效性和稳健性。文献^[62]提出一种基于 *unscented* 粒子滤波的目标跟踪算法,有效解决了非线性环境中的红外弱小目标跟踪问题。文献^[63]研究了低信噪比复杂环境下红外小目标检测和跟踪的问题,提出基于辅助粒子滤波的检测前跟踪(TBD)算法。该算法应用于信噪比为 20dB 的真实红外图像弱小目标时获得了较好的检测和跟踪效果,算法性能优于传统的 TBD 方法。随着近年来多模制导技术的兴起,红外/毫米波双模制导技术成为一个新的研究热点^[64,65]。由毫米波雷达得到目标的距离、方位角、俯仰角测量值,再由红外传感器得到目标的方位角、俯仰角测量值,利用信息融合技术处理两组测量值,可以有效提高目标跟踪精度^[66]。由于系统模型和测量方程的非线性,一些常用线性逼近滤波方法(例如扩展卡尔曼滤波)的跟踪误差会随着时间增加,甚至导致发散^[67]。文献^[68]将粒子滤波引入红外/毫米波复合制导目标跟踪问题中。仿真结果表明,与传统的扩展卡尔曼滤波相比,在观测方程非线性情况下粒子滤波取得了较好的效果,有效提高了制导精度。上述粒子滤波用于无线通讯定位跟踪和红外目标跟踪的研究成果表明,采用粒子滤波算法使目标跟踪的精度得到了进一步的提高。

5.4 多目标跟踪

目标跟踪与数据关联包括了单(传感器)站单目标定位与跟踪、单站多目标定位与跟踪(数据关联)、多站(多平台)单目标定位与跟踪和多站(多平台)多目标定位与跟踪四部分内容,其中多站(多平台)多目标定位与跟踪是最为复杂的定位与跟踪,单站单目标、单站多目标、多站单目标为其特例,但从应用对象与其功能性能的要求却不完全相同,因此不能要求多站多目标的信息融合研究取代如上特例。实际上,对于无源多静止站多目标定位与跟踪来说,它离不开多站单目标定位与跟踪,特别离不开单站多目标数据关联。就象单站多目标数据关联离不开单站单目标定位与跟踪一样。

当传感器只能测量目标的方位信息时,此时仅仅依靠单个静止站的信息是无法获得目标的全部运动参数的;此时采取的方法往往是应用参数求解方法来实现方位角跟踪或部分运动参数跟踪(在已知目标运动速度的前提下,可以获得目标的位置)。而当多个传感器测量站可以同时提供对同一目标的方位角测量信息时,通常可以利用视线交叉定位方法来实现对目标的定位与

跟踪，当传感器异步工作、测量信息本身存在传播延时，简单的利用最小二乘方法来实现交叉定位已不可行，必须同时进行数据校准的工作。纯方位多站多目标信息融合的研究文献较少，最典型的方法是利用最大似然原理，进行目标状态估计与数据关联两步迭代。纯方位多站多目标信息融合所面临的最大困难是量测对目标的分类问题，分类算法的性能对融合算法起到了关键作用。

5.4.1 “最近邻”法

1971 年 Singer 等^[69]提出了一种具有固定记忆要求且能在多量测环境下工作的跟踪滤波器。该滤波器仅仅利用在统计意义上与被跟踪目标的预测位置(跟踪门或关联波门中心)最近的量测作为候选量测。所谓“最近”往往表示统计距离最小或者残差概率密度最大，统计距离一般采用欧氏距离：

$$d^2[z(k)] = [z(k) - \hat{z}(k|k-1)]^T \Sigma^{-1}(k) [z(k) - \hat{z}(k|k-1)] \quad (5-61)$$

其中 $\hat{z}(k|k-1)$ 表示目标的预测位置， $\Sigma(k)$ 表示 $z(k) - \hat{z}(k|k-1)$ 的协方差。

该方法的本质是一种“贪心”算法，并不能在全局意义下保持最优，该算法选择离关联波门中心最近的量测对目标航迹进行更新，而离中心最近的量测未必就是正确的目标量测，因此，“最近邻”法往往会发生误跟和丢失目标的现象。

5.4.2 概率数据关联滤波器 (PDAF)

考虑系统描述为：

$$x(k+1) = A(k)x(k) + w(k), k = 1, 2, \dots \quad (5-62)$$

$$z(k) = C(k)x(k) + v(k), k = 1, 2, \dots \quad (5-63)$$

其中, $x(k)$ 表示 k 时刻的状态向量, 在 $z(k)$ 表示 k 时刻的观测向量, $A(k)$ 表示状态转移矩阵, $C(k)$ 表示量测矩阵, $w(k)$ 和 $v(k)$ 是零均值相互独立的白色高斯过程噪声。

设 $Z(k) = \{z_j(k)\}_{j=1}^m$, 表示传感器在时刻 k 确认的量测集合; m 表示在 k 时刻确认的量测的个数; $Z^k = \{Z(j)\}_{j=1}^k$ 表示直到时刻 k 的累积量测集。

概率数据关联理论三个基本假设是：1)假量测在关联波门中服从均匀分布；2)正确量测服从正态分布；3)在杂波环境下仅有一个目标存在，并且这个目标的航迹已经形成。概率数据关联滤波算法可以简单的描述如下：

在 k 时刻有 m 个量测落入有效探测范围内，第 j 个有效量测 $z_j(k)$ 源自目标的概率为：

$$\beta_j = \frac{e_j}{b + \sum_{l=1}^m e_l}, \quad j = 1, 2, \dots, m \quad (5-64)$$

m 个量测中没有一个是来自目标的量测的概率为：

$$\beta_0 = \frac{b}{b + \sum_{l=1}^m e_l} \quad (5-65)$$

其中 $e_j = \exp\left\{-\frac{1}{2}v_j^T(k)S^{-1}(k)v_j(k)\right\}$ ， $v_j(k)$ 是第 j 个有效量测 $z_j(k)$ 的信息， $S(k)$ 为量测的协方差。

对于参数模型的概率数据关联滤波算法，上述两式中的 b 由下式给定：

$$b = \lambda \sqrt{\det[2\pi S(k)]} \frac{1 - P_D P_G}{P_D} \quad (5-66)$$

式中 λ 为杂波的空间分布密度， P_D 是检测概率， P_G 凡是源自目标的量测落入有效探测范围的概率，一般情况下 $P_G = 1$ 。

对于非参数模型的概率数据关联滤波算法， b 的表达形式与参数模型相同，不同的是，式中的 λ 由 m/V 来替换， V 是有效探测范围的面积或体积。

PDAF 算法中的目标状态估计由下式完成更新：

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)v(k) \quad (5-67)$$

其中 $\hat{x}(k|k)$ 为状态更新估计， $\hat{x}(k|k-1)$ 为状态预测值， $K(k)$ 为卡尔曼增益， $v(k)$ 表示组合信息。

$$v(k) = \sum_{j=1}^m \beta_j v_j(k) \quad (5-68)$$

$$v_j(k) = z_j(k) - C(k)\hat{x}(k | k-1) \quad (5-69)$$

目标状态更新估计的协方差由下式确定：

$$P(k | k) = \beta_0 P(k | k-1) + [1 - \beta_0] P^c(k) + \tilde{P}(k) \quad (5-70)$$

其中，

$$P^c(k) P(k | k-1) - K(k) S(k) K^T(k) \quad (5-71)$$

$$\tilde{P}(k) = K(k) \left[\sum_{j=1}^m \beta_j v_j(k) v_j^T(k) - v(k) v^T(k) \right] K^T(k) \quad (5-72)$$

PDAF 算法在杂波环境中具有较好的跟踪性能，适用于杂波环境中单个目标的跟踪。PDAF 算法是在独立一致空间分布的情况下，将所有不正确的量测建立为“随机干扰”模型，因此对于多目标情况下，近邻目标的存在会引起建模不正确，其性能急剧下降。Bar-Shalom 等^[70]在 PDAF 算法的基础上，提出了适用于跟踪多个目标的一种数据关联算法，就是著名的联合概率数据关联滤波(Joint probability Data Association Filter, JPDAF)算法。

5.4.3 联合概率数据关联滤波器 (JPDAF)

联合概率数据关联滤波(Joint probability Data Association Filter, JPDAF)算法^[71]被公认为是解决密集量测下多目标数据关联最有效的算法之一，其目标跟踪成功率在各种环境下都很高。JPDAF 算法对所有可能的目标关联解进行搜索，并在此基础上计算出最佳关联概率。

JPDAF 算法的关键在于计算每一个量测与其可能的各种源目标相关联概率。设 $\theta_i(k) = \{\theta_i(k)\}_{i=1}^{\theta_k}$ 表示在 k 时刻的所有可能的联合事件的集合， θ_k 表示 $\theta(k)$ 中元素的个数，其中：

$$\theta_i(k) = \bigcap_{j=1}^m \theta_{j_i}^i(k) \quad (5-73)$$

$\theta_{j_i}^i(k) = \{\text{在第 } i \text{ 个联合事件中由目标 } t_j \text{ 得到的量测 } j\}, j=1,2,\dots,m; t=0,1,\dots,T, t_j \text{ 是考虑事件中的}$

与量测 j 关联的目标记号。

定义确认矩阵：

$$\Omega(k) = \{\omega_{jt}(k)\}, j=1,2,\dots,m; t=0,1,\dots,T \quad (5-74)$$

二进制元素 $\omega_{jt}(k)$ 为 1 表示量测 j 处在目标 t 的确认门内, 为 0 表示量测 j 没有处在目标 t 的确认门; $t=0$ 表示“没有目标”, $\Omega(k)$ 的相应列全部是 1, 因为此时每一个量测都来自杂波或虚警。

由此, 一个联合关联事件 $\theta_i(k)$ 可以表示成以下矩阵:

$$\hat{\Omega}(\theta_i(k)) = [\hat{\omega}_{jt}^i(\theta_i(k))] \quad (5-75)$$

其中:

$$\hat{\omega}_{jt}^i(\theta_i(k)) = \begin{cases} 1, & \text{若 } \theta_{jt}^i(k) \subset \theta_i(k) \\ 0, & \text{其它} \end{cases} \quad (5-76)$$

可能事件是以下的事件: 量测只有一个源, 即

$$\sum_{t=0}^T \hat{\omega}_{jt}^i(\theta_i(k)) = 1, \quad j=1,2,\dots,m \quad (5-77)$$

源于一个目标的量测不大于 1, 即

$$\delta_t(\theta_i(k)) = \sum_{j=1}^m \hat{\omega}_{jt}^i(\theta_i(k)) \leq 1, \quad j=1,2,\dots,T \quad (5-78)$$

上式定义的 δ_t 称为目标检测指示器, 表示任一量测是否与联合事件 $\theta_i(k)$ 中的目标 t 关联。同时

定义量测关联指示器用以表示量测 j 是否与只 $\theta_i(k)$ 事件中的任一个目标关联:

$$\tau_j(\theta_i(k)) = \sum_{t=0}^T \hat{\omega}_{jt}^i(\theta_i(k)), \quad j=1,2,\dots,m \quad (5-79)$$

联合事件 $\theta_i(k)$ 中假量测的数量是:

$$\phi(\theta_i(k)) = \sum_{j=1}^m [1 - \tau_j(\theta_i(k))] \quad (5-80)$$

利用贝叶斯法则推导在 k 时刻的联合事件 $\theta_i(k)$ 的条件概率为:

$$\begin{aligned} p(\theta_i(k) | Z^k) &= p(\theta_i(k) | Z(k), Z^{k-1}) = \frac{1}{c} p[Z(k) | \theta_i(k), Z^{k-1}] p(\theta_i(k) | Z^{k-1}) \\ &= \frac{1}{c} p[Z(k) | \theta_i(k), Z^{k-1}] p(\theta_i(k)) \end{aligned} \quad (5-81)$$

其中,

$$c = \sum_{i=1}^{\theta_k} p[Z(k) | \theta_i(k), Z^{k-1}] p(\theta_i(k)) \quad (5-82)$$

假设不与任一目标关联的量测在探测确认区域 V 内服从均匀分布, 而与某个目标关联的正确量测服从高斯分布, 通过计算可以得到:

$$p(\theta_i(k) | Z^k) = \frac{\gamma(\theta_i(k))}{c'} \quad (5-83)$$

$$c' = \sum_{i=1}^{\theta_k} \gamma(\theta_i(k)) \quad (5-84)$$

对于参数 JPDAF 有:

$$\gamma(\theta_i(k)) = \lambda^{\phi(\theta_i(k))} \prod_{j=1}^m N_{t_j}[z_j(k)]^{\tau(\theta_i(k))} \prod_{i=1}^T (P_D^t)^{\delta(\theta_i(k))} (1 - P_D^t)^{1 - \delta(\theta_i(k))} \quad (5-85)$$

对于非参数 JPDAF 有:

$$\gamma(\theta_i(k)) = \frac{\phi(\theta_i(k))!}{V^{\phi(\theta_i(k))}} \prod_{j=1}^m N_{t_j}[z_j(k)]^{\tau(\theta_i(k))} \prod_{i=1}^T (P_D^t)^{\delta(\theta_i(k))} (1 - P_D^t)^{1 - \delta(\theta_i(k))} \quad (5-86)$$

其中:

$$N_{t_j}[z_j(k)] = N[z_j(k); \hat{z}_{t_j}(k | k-1); S_{t_j}(k)] = [2\pi S_{t_j}(k)]^{-1/2} \exp \left\{ -\frac{1}{2} v_j^T S_{t_j}^{-1}(k) v_j(k) \right\} \quad (5-87)$$

$$v_j(k) = z_j(k) - \hat{z}_{t_j}(k | k-1) \quad (5-88)$$

通过对所有联合事件求和, 可以得到第 j 个量测与目标关联的概率为:

$$\beta_{jt}(k) = p(\theta_{jt}(k) | Z^k) = \sum_{i=1}^{\theta_k} p(\theta_i(k) | Z^k) \omega_{jt}(\theta_i(k)) \quad (5-89)$$

目标 t 的状态更新估计由全概率公式得：

$$\hat{x}_t(k|k) = \sum_{j=0}^m \beta_{jt}(k) \hat{x}_j^t(k|k) \quad (5-90)$$

其中 $\hat{x}_j^t(k|k) = E[x^t(k) | \theta_{jt}(k), Z^k]$, $j=1,2,\dots,m$ 表示在 k 时刻第 j 个量测对目标 t 进行卡尔曼滤波所得的状态估计。 $\hat{x}_0^t(k|k)$ 表示 k 时刻没有量测源于目标的情形，这时用预报值代替：

$$\hat{x}_0^t(k|k) = \hat{x}^t(k|k-1) \quad (5-91)$$

由上可知，JPDAF 算法中对关联解的搜索实际上是一个求组合问题，其搜索过程的计算量随目标和量测的数量增长呈指数增长趋势，JPDAF 算法的计算量在目标和量测密集时呈“组合爆炸”的趋势，因此在实际工程中难于广泛应用。

5.4.4 多假设跟踪（MHT）

Reid 根据多目标跟踪问题基于 Singer 等的“全邻”最优滤波器和 Bar-Shalom 的确认矩阵的概念，提出了多假设跟踪方法(Multiple Hypothesis Tracking, MHT)。MHT 算法与 JPDAF 算法是被公认为在多目标跟踪领域最有效的两种关联方法。MHT 算法考虑量测来源于目标、杂波和新目标等各种可能的情况，构造面向量测的关联假设树，并利用贝叶斯后验概率的传递特性，对假设树的各个分枝进行概率计算，不断删除小概率假设，合并相同目标假设，以实现多目标的数据关联与跟踪。

假设的产生：

设 Ω^k 是至 k 时刻的关联假设集，由 $k-1$ 时刻的关联假设集和 Ω^{k-1} 当前量测集 $Z(k)$ 关联得到 $Z(k) \overset{\Delta}{=} \{z_i(k)\}_{i=1}^m$ 。

量测 $z_i(k)$ 和一个假设的关联有三种可能：

1. 与原有的一个假设关联，是一个已有航迹的继续；
2. 是一个新目标；
3. 是虚警。

每个目标最多能与一个现在量测关联，该量测必须落入确认区域内。

概率计算:

设关于当前量测的事件 $\theta(k)$ 包含 τ 个源于已建立航迹的量测， v 个源于新目标的量测和 ϕ

必个源于虚警或杂波的量测。对于 $i=1, 2, \dots, m$ ，定义与 $\theta(k)$ 事件有关的指示变量：

$$\tau_i = \tau_i[\theta(k)] = \begin{cases} 1, & \text{如果 } z_i(k) \text{ 来自已经确认的航迹} \\ 0, & \text{否则} \end{cases}$$

$$v_i = v_i[\theta(k)] = \begin{cases} 1, & \text{如果 } z_i(k) \text{ 是新目标} \\ 0, & \text{否则} \end{cases}$$

$$\phi_i = \phi_i[\theta(k)] = \begin{cases} 1, & \text{如果 } z_i(k) \text{ 源于虚警或杂波} \\ 0, & \text{否则} \end{cases} \quad (5-92)$$

则在 $\theta(k)$ 中已建立的航迹数为:

$$\tau = \sum_{i=1}^m \tau_i \quad (5-93)$$

在 $\theta(k)$ 中起始的新航迹数为:

$$v = \sum_{i=1}^m v_i \quad (5-94)$$

在 $\theta(k)$ 中假量测数为:

$$\phi = \sum_{i=1}^m \phi_i = m - \tau - v \quad (5-95)$$

设 $\Theta^{k,l}$ 表示 k 时刻的第 l 条航迹，在 k 时刻的联合累积事件表示为:

$$\Theta^{k,l} = \{\Theta^{k-l,s}, \theta(k)\} \quad (5-96)$$

事件 $\theta(k)$ 表示当前量测集 $z(k)$ 中的量测与目标或杂波关联的一种可能的情形， $\theta(k)$ 可以表示为:

$$\theta_k = \{\theta_T(k), \theta_N(k), \theta_F(k)\} \quad (5-97)$$

其中,

$$\theta_T(k) = \bigcap_{i:\tau_i=1} \theta_{t_i}(k) \quad (5-98)$$

表示与已确认航迹的关联, t_i 是航迹的标号, 该航迹与 $z_i(k)$ 关联;

$$\theta_N(k) = \bigcap_{i:\tau_i+v_i=0} \theta_{t_{n_i}}(k) \quad (5-99)$$

表示在 k 时刻的新目标事件, n_i 是对应由 $z_i(k)$ 起始的新目标航迹标号;

$$\theta_F(k) = \bigcap_{i:\tau_i+v_i=0} \theta_{i0}(k) \quad (5-100)$$

表示在 k 时刻的虚警事件, 0 表示 $z_i(k)$ 源于杂波或虚警;

假设如果一个量测 $z_i(k)$ 源于一条已建立的航迹, 则其服从高斯概率分布; 如果一个量测 $z_i(k)$ 源于杂波或虚警, 则其在探测范围内服从均匀分布; 如果一个量测 $z_i(k)$ 起始一条新航迹, 也假设服从均匀分布, 则应用贝叶斯法则可以得到假设 $\Theta^{k,l}$ 的概率和假设 $\Theta^{k-l,s}$ 的概率之间的递归关系:

$$\begin{aligned} p(\Theta^{k,l} | Z^k) &= p\{\theta(k), \Theta^{k-l,s} | Z^{k-l}\} \\ &= \frac{1}{c} p\{Z(k) | \theta(k), \Theta^{k-l,s}, Z^{k-l}\} p\{\theta(k) | \Theta^{k-l,s} | Z^{k-l}\} \\ &= \frac{1}{c} \frac{\phi!v!}{m!} \mu_F(\phi) \mu_N(v) \prod_t (P_D^t)^{\delta_t} (1-P_D^t)^{1-\delta_t} \prod_{i=l}^m [N_{t_i}[z_i(k)]]^{\tau_j} p(\Theta^{k-l,s} | Z^{k-l}) \end{aligned} \quad (5-101)$$

其中,

$$c = p\{Z(k) | Z^{k-l}\} \quad (5-102)$$

P_D^t 表示目标 t 的探测概率; $\mu_F(\phi)$ 和 $\mu_N(v)$ 分别表示虚警数和新航迹数的概率分配函数。然后根据计算所得的假设概率进行假设管理, 包括假设删除和假设合并等。最后得到的高概率的假设即为真目标的估计航迹。MHT 算法同样面临目标数和量测数增加时所带来的计算“组合爆炸”现象, 具有一定的局限性。

5.4.5 多维 S-D 分配算法

当传感器是无源时，只能探测到目标的方位角，此时数据关联问题变得更为复杂，人们必须将来自多个传感器的量测相互关联以得到目标的状态估计，从而导致问题的组合爆炸。JPDA 算法和 MHT 算法的实用性和可行性都遇到了困难，研究者发现使用诸如 S-D 分配的多维分配算法的数据关联方法是实用而且可行的，S-D 分配算法^[72]是数据关联问题的一种离散优化算法，需要克服的主要困难是要解决随之出现的 NP 多维分配难题。

考虑 S 个无源传感器的情况，并假定目标为静态，设传感器编号为 $S=1, 2, \dots, S$ ，传感器 S 接收到的观测数为 n_s ，各传感器接收到的观测数可能不相等，观测编号为 $i_s=0, 1, 2, \dots, n_s$ ，其中 0 为漏检，设各传感器的探测概率为 P_D 。各传感器 S 只能测量得到目标 j 的相对于传感器 S 的方位角 θ_{sj} ，目标 j 的坐标为 (x_j, y_j) ， $j=1, 2, \dots, T$ ，传感器 S 的坐标为 (x_s, y_s) ，则测量模型可以表示为：

$$z_{si} = \begin{cases} \theta_{sj} + v_{si}, & \text{测量来自目标 } j \\ w_{si}, & \text{是一个虚警} \end{cases} \quad (5-103)$$

其中 $\theta_{sj} = \arctan[(y_j - y_s)/(x_j - x_s)]$ 为高斯白噪声 $N(0, \sigma_s^2)$ ， w_{si} 是虚警，其在传感器探测范围内是一致分布的，概率密度函数为 $p_{w_{si}}(w) = \frac{1}{\phi_s}$ ， ϕ_s 是传感器 S 的有效探测范围(二维为面积，三维是体积)。

传感器 S 的测量集表示为 $Z_s = \{z_{si}\}_{i_k=0}^{n_s}$ ，包括了漏检 z_{si_0} ；同一监测范围内所有传感器的测量集表示为 $Z = \{Z_s\}_{s=1}^S$ ；某一可能的空间位置由 S 个传感器的观测(方位视线)构成，表示为 $Z_{i_1 i_2 \dots i_S}$ ，此时又称为 S 元观测，若传感器数量为 3，则空间位置 $Z_{0 i_2 i_3}$ 表示该位置由传感器 1 漏检、传感器 2 的第 i_2 个观测，传感器 3 的第 i_3 个观测构成。一个 S 元观测分配给空间位置 (x_j, y_j) 的似然函数为：

$$\Lambda(Z_{i_1 i_2 \dots i_s} | x_j, y_j) = \prod_{s=1}^s [P_{D_s} p(z_{s i_s} | x_j, y_j)]^{1-\delta_{0i_s}} [1-P_{D_s}]^{\delta_{0i_s}} \quad (5-103)$$

其中 δ_{0i_s} 为一指示函数，有

$$\delta_{0i_s} = \begin{cases} i, i_s = 0 \text{ 传感器S漏测} \\ 0, \text{ 其他情况} \end{cases} \quad (5-104)$$

$p(z_{s i_s} | x_j, y_j) = N(\theta_{s j}, \sigma_s^2)$, $i = 1, 2, 3, \dots, n_s$; $s = 1, 2, 3$ 为概率密度函数。

此时目标位置由最大化该似然函数决定，即

$$(\hat{x}_j, \hat{y}_j) = \arg \max_{x_j, y_j} (Z_{i_1 i_2 \dots i_s} | x_j, y_j) \quad (5-105)$$

当目标 $p = \emptyset$ 时，也即测量都是虚警或与此目标不相关时的似然函数为：

$$\Lambda(Z_{i_1 i_2 \dots i_s} | p = \emptyset) = \prod_{s=1}^s \left[\frac{1}{\Phi_s} \right]^{1-\delta_{0i_s}} \quad (5-106)$$

与目标 (x_j, y_j) 相关的 S 元观测的权由负的对数似然函数给出：

$$c_{i_1 i_2 \dots i_s} = -\ln \frac{\Lambda(Z_{i_1 i_2 \dots i_s} | x_j, y_j)}{\Lambda(Z_{i_1 i_2 \dots i_s} | p = \emptyset)} \quad (5-107)$$

将上式代入， S 元观测与目标备选相关的权为：

$$c_{i_1 i_2 \dots i_s} = -\sum_{s=1}^s [(1-\delta_{0i_s}) (\ln(\frac{P_{D_s} \Phi_s}{\sqrt{2\pi} \sigma_s}) - \frac{1}{2} (\frac{z_{s i_s} - \hat{\theta}_{s j}}{\sigma_s})^2) + \delta_{0i_s} \ln(1-p_{D_s})] \quad (5-108)$$

其目标是找到 S 元观测的最大似然集，即使得每个测量分配给一个目标且仅有一个目标，或判断为虚假目标，以及每个目标最多获得一个传感器一个扫描周期内的一个测量。该问题于是可以归纳为以下普适的 S -D 分配问题：

$$J^* = \min_{\rho_{s_1 s_2 \dots s_3} \in P} J(\rho) \quad (5-109)$$

$$J(\rho) = \sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \dots \sum_{i_s=0}^{n_s} c_{i_1 i_2 \dots i_s} \rho_{i_1 i_2 \dots i_s} \quad (5-110)$$

满足约束：

$$\begin{aligned}
\sum_{i_2=0}^{n_2} \cdots \sum_{i_s=0}^{n_s} \rho_{i_1 i_2 \cdots i_s} &= 1, \text{ 对所有 } i_1 = 1, 2, \cdots, n_1 \\
\sum_{i_1=0}^{n_1} \sum_{i_3=0}^{n_3} \cdots \sum_{i_s=0}^{n_s} \rho_{i_1 i_2 \cdots i_s} &= 1, \text{ 对所有 } i_2 = 1, 2, \cdots, n_2 \\
&\vdots \\
\sum_{i_1=0}^{n_1} \cdots \sum_{i_{s-1}=0}^{n_{s-1}} \rho_{i_1 i_2 \cdots i_s} &= 1, \text{ 对所有 } i_s = 1, 2, \cdots, n_s
\end{aligned}
\tag{5-111}$$

其中 $\{\rho_{i_1 i_2 \cdots i_s}\}$ 是二元相关变量, 使得若 S 元观测 $Z_{i_1 i_2 \cdots i_s}$ 是一个备选目标相关的则 $\rho_{i_1 i_2 \cdots i_s} = 1$, 否则为 0。

S -D 分配问题在 $S \geq 3$ 时是 NP-hard 问题, Poore^[73]的方法是扩展使用了分割技术把图分割为更小的不连接的图, 从而可用分支和界定算法最优地去解 S -1 维问题, 该方法用于最小测量容量的稀疏图是成功的, 但当用于不易分开的图时却需要花费大量的、不可接受的计算资源。对于更密集的图, Poore 提出针对此困难使用“价值函数”来解决。Krishna R.Pattipati 和 somnath Deb 等人提出了同时松弛所有 $(S-2)$ 维约束的拉格朗日松弛算法, 算法的关键在于必须同时更新与多个约束集合相关的拉格朗日乘子以获得更快的收敛。这些算法在传感器和目标数量较多时均不能在给定时间内获得满意解, 因此工程应用中往往采用高性能的多处理机进行并行运算以获得一定的解, 寻求简单且有效的数据关联算法成为众多研究者的努力方向。

本节概述了目标跟踪与数据关联的理论和方法, 介绍了几种具有典型意义的跟踪与数据关联算法。综观众多算法, 它们利用的信息包括了目标的属性、目标的运动参数(位置、速度、加速度、航向等)以及同一目标对应于多个传感器量测必然存在的某种相似特征等信息, 其目的就是将落入到波门内的多个量测进行分类, 以确定它们是分别属于已有航迹的续航、新目标的出现还是虚警。随着跟踪目标数量、传感器数量以及杂波密度的增加, 计算量也急剧增加, 多数算法无法在给定时间内完成计算, 从而大大限制了这些算法的实际应用。

因此在实际工程应用中, 往往要根据实际情况来确定算法的计算复杂度与算法性能之间的关系, 任何企图以某几种算法覆盖所有实际应用的想法是不现实的。本文将针对无源声音探测预警网络的实际工程应用, 研究了多运动目标的无源跟踪与数据关联的方法, 提出了系列符合实际工程应用的算法。研究表明, 所提供的算法是有效的。

5.4.6 基于稀疏重构分析的多目标跟踪

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:

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

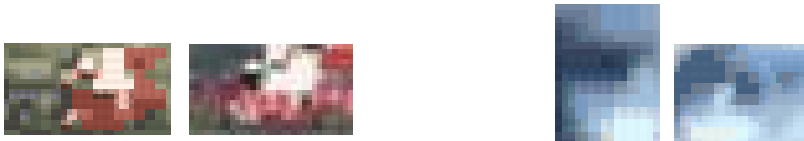
其中, J 表示跟踪目标的数目, A_j 表示第 j 个跟踪目标的重构样本集合。如果没有目标对应的重构样本集合中有 K 个重构样本, 那么归一化的重构样本集合总共包含有 $K \times J$ 个重构样本。

5.4.6.1 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:

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

如图 5.6 和 5.7 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 5.6 (a)和(b)来至于测试视频 VIVID-redteam, 图 5.6c 和 d 来至于测试视频 VIVID-egtest01。 $\|\cdot\|_2$ 表示 2 范式, 图 5.7 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。



(a) (b) (c) (d)

图5.6 多目标跟踪过程中的两类跟踪目标实例。(a) 和 (b) 来至于同一个跟踪目标，且 (c) 和 (d) 来至于另一个跟踪目标。

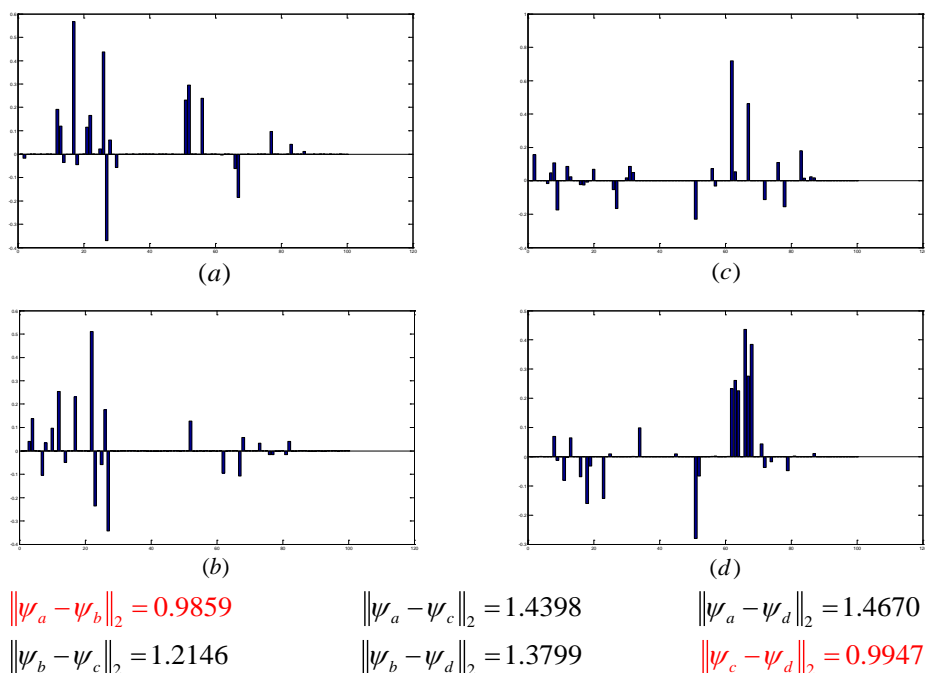


图5.7 AdaSRA的结果分析。(a)–(d)分别为图5.6 (a)–(d)分别对应的归一化重构稀疏系数。

5.4.6.2 有效性验证

公式 5-14 所定义的分类准则与 SRC 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 5.8 所示。我们从三个跟踪目标中挑选出一个跟踪目标（如图 5.9 (a) 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 5.9 (b) 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 5.9 (c) 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。

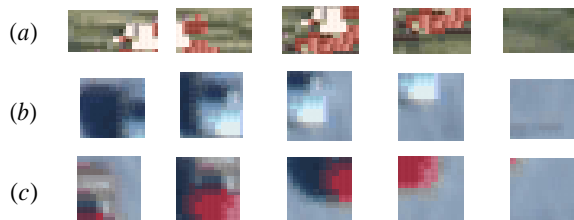


图 5.8 三种跟踪目标所对应的重构样本集合中部分样本实例。(a) 第一类目标的部分样本，(b) 第二类目标的部分样本，(c) 第三类目标的部分样本。

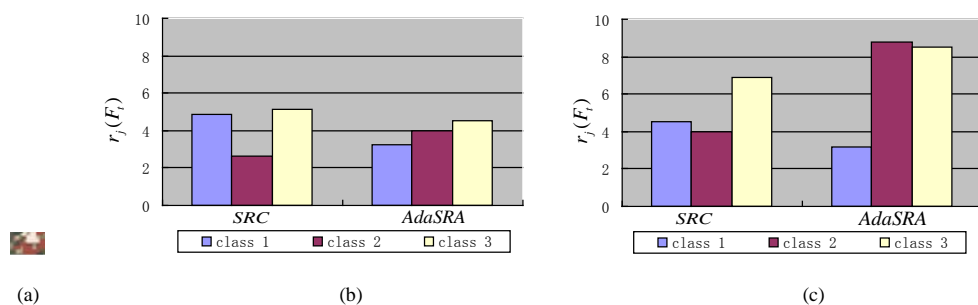


图 5.9 基于 AdaSRA 与 SRC 的分类结果比较。(a) 测试目标（第一类），(b) 当每一类重构样本数目为 50 时计算得到的分类结果，(c) 当每一类重构样本数目为 100 时计算得到的分类结果。

5.4.6.3 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，如图 5.10 所示：



(a) 150th frame

330th frame

510th frame

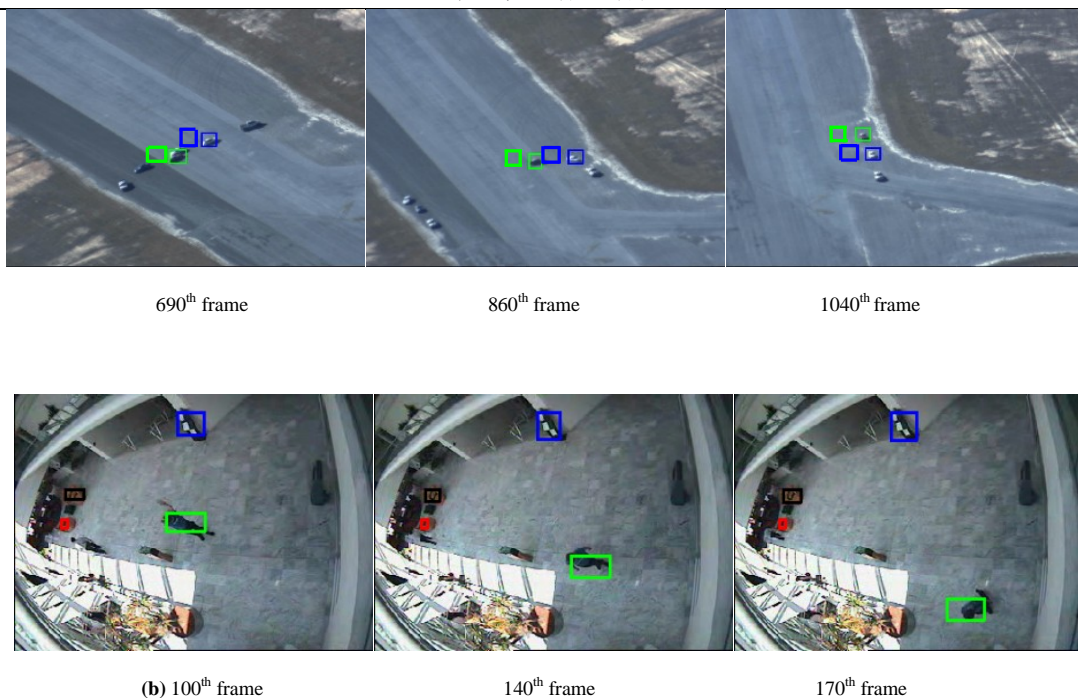


图 5.10 基于稀疏重构分析 (AdaSRA) 的多目标跟踪结果实例

5.5 参考文献

- [1] Lowe D G. Distinctive image features from scale-invariant key points, International Journal of Computer Vision, 2004:91~110.
- [2] Han Z j. Ye Q x., and Jiao J b., Online feature evaluation for object tracking using kalman filter, Proceedings of International Conference on Pattern Recognition, 2008.
- [3] Hager G, and Belhumeur P. Efficient region tracking with parametric models of geometry and illumination, IEEE Trans. Pattern Analysis and Machine Intelligence, 1998:1025–1039.
- [4] Collins R. Mean-shift blob tracking through scale space, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2003:234-240.
- [5] Pérez de la Blanca N, Fuertes J M, and Lucena M, Deformable object matching based on

- multi-scale local histograms, AMDO 2004, 2004:154-162.
- [6] Dalal N, and Triggs B. Histograms of oriented gradients for human detection, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005:886-893.
- [7] Zhu Q, Avidan S, Yeh M C, and Cheng K T. Fast human detection using a cascade of histograms of oriented gradients, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006.
- [8] Fogel E, Gavish M N. Nth order dynamics target observability from angle measurements [J]. IEEE Trans on Aerospace and Electronic Systems, 1988, 24(3):305-308.
- [9] Springarn K. Passive position location estimation using the extended Kalman filter systems [J]. IEEE Trans on Aerospace and Electronic Systems, 1987, 23(4):558-567.
- [10] Pham D T. Some Quick and Efficient Methods for Bearing-Only Target Motion Analysis [J]. IEEE Transactions on Signal Processing, 1993, 41(9):2727-2751.
- [11] Barniv Y, Kella O. Dynamic Programming Solution for Detecting Dim Moving Targets, Part E: Analysis [J]. IEEE Trans on Aerospace and Electronic Systems, 1987, 23(2):776-788.
- [12] Xia M G, He Y, Huang X D, et al. A Survey of Small Target Detection and Tracking Based on Infrared Images [J]. Fire Control and Command Control, 2003, 28(1):125-134.
- [13] Tonissen S M, Bar-Shalom Y. Maximum Likelihood Track-before-detect with Fluctuating Target Amplitude [J]. IEEE Trans on Aerospace and Electronic Systems, 1998, 34(3):796-808.
- [14] Salmond D J, Birch H. A Particle Filter for Track before detect [A]. Proc of the American Control Conf [C]. Arlington VA, 2001:3755-3760.
- [15] Rollason M, Salmond D. A Particle Filter for Track-before-detect of a Target with Unknown Amplitude [J]. IEEE Target Tracking: Algorithms and Applications, 2001, (1):16-17.
- [16] 杨小军, 潘泉, 张洪才. 基于粒子滤波和似然比的联合检测与跟踪. 控制与决策, 2005, 20 (7): 838-840.

- [17] Gordon N.A. Hybrid Bootstrap Filter for Target Tracking in Clutter.IEEE Transactions on Aerospace and Electronic Systems,1997,33(1):353-358.
- [18] HU H t, Jing Z l, Li A p, et al1.An MCMC-based particlefilter for tracking target in glint noise environment[C].Proc of the 7th Int Conf on Information Fusion Stockholm, Sweden: International Society of Information Fusion,2004:922~927.
- [19] Morelande M R, Challa S. Maneuverings Target Tracking in Clutter Using Particle Filters.IEEE Transactions on Aerospace and Electronic Systems,2005,41(1):252~270.
- [20] Fredrik G, Niclas B, et all. Particle filters for positioning, navigation,and tracking.IEEE Transactions on signal processing,2002, 50(2):425~437.
- [21] Doucet A, de Freitas J F G, Gordon N J. Sequential Monte Carlo Methods in Practice. New York:Springer-Verlag, 2001:401~427.
- [22] Guo D, Wang X D. Dynamic Sensor Collaboration via Sequential Monte Carlo. IEEE Journal on Selected Areas in Communications, 2004, 22(6):1037~1046.
- [23] Prahlad V, Liu J. Improved Particle Filter in Sensor Fusion for Tracking Randomly Moving Object. IEEE Transactions on instrumentation and measurement, 2006, 55(5):1823-1832.
- [24] Jeroen L,Marcel R, Emile H. Influence of the observation likelihood function on particle filtering performance in tracking applications.Proceedings of Automatic Face and Gesture Recognition(FGR' 04),IEEE computer society, 2004:767~772.
- [25] Doucet A, Davy M. Particle filtering for multi-target tracking and sensor management[A]. The 5th Int Conf on Information Fusion[C].Annapolis, 2002:474~481.
- [26] Mcginnity S,Irwin G W.Multiple Model Bootstrap Filter for Maneuvering Target Tracking. IEEE Transactions on Aerospace and Electronic System,2000,36(3):1006~1012.
- [27] Hue C, Cadre J, Perez P. Sequential Monte Carlo methods for multiple target tracking and data fusion[J]. IEEE Trans on Signal Processing, 2002, 50(2):309~325.
- [28] Storvik G. Particle filters for state-space models with the presence of unknown static

- parameters[J]. IEEE Trans on Signal Processing, 2002,50(2):281~289.
- [29] Orton M, Marrs A. A Bayesian Approach to Multi-Target Tracking and Data Fusion with Out-of-Sequence Measurements.IEE Target Tracking:Algorithnls and Applications,2001,1:1~5.
- [30] Srivastava A, Lanterman A D, Grenander U, et al. Monte Carlo Techniques for Automated Target Recognition.Doucet A, de Freitas J F G,Gordon N J. Sequential Monte Carlo Methods in Practice.New York:Springer-Verlag,2001:533~552.
- [31] Ward D B, Lehmann E A, Williammson R C. Particle Filtering Algorithms for Tracking an Acoustic Source in a Reverberant Environment.IEEE Transactions on Speech and Audio processing, 2003,11(6):826~836.
- [32] Perez P, Vermaak J, Blake A. Data fusion for visual tracking with particles[J]. Proceeding of the IEEE, 2004, 92(3):495~513.
- [33] Isard M, Blake A. Condensation-conditional density propagation for visual tacking[J].J of Computer Vision, 1998, 29(1):5~28.
- [34] Isard M, Blake A. Condensation:Unifying low-level and high-level tracking in a stochastic framework[A].Proc of the Fifth European Conf on Computer Vision[C].Berlin,1998:893~908.
- [35] Lip H,Zhang T W. Visual contour tracking based on sequential importance sampling/ resampling algorithm[C]. Proc of the 16 th Int Conf on Pattern Recognition1 Canada:IEEE Computer Society, 2002, 2:564~568.
- [36] Brunom G S. Bayesian methods for multiaspect target tracking in image sequences[J]. IEEE Trans on Signal Processing,2004,52(7):1848~1861.
- [37] Maskell S, Rollason M,Gordon N, et all. Efficient particle filtering for multiple target tracking with application to tracking in structured images[C].Proc of SPIE on Signal and Data Processing of Small Targets.Washington: SPIEPress, 2002, 4:251~262.
- [38] Vermaa K J, Gangnet M, Blake A, et all.Sequential Monte Carlo fusion of sound and vision for

- speaker tracking[C].Proc of the Eighth IEEE Int Conf on Computer Vision, Vancouver: IEEE Press,2001,1:741~746.
- [39] Kwon J. Visual tracking via particle filtering on the affine group. Proceeding of the 2008 IEEE international conference on information and automation, seoul national university, Korea, IEEE Press, 2008:997~1002.
- [40] Shan C, Wei Y, Tan T N, et al. Real time hand tracking by combining particle filtering and mean shift[C]. Proc of the Sixth IEEE Int Conf on Automatic Face and Gesture Recognition, Seoul, Korea: IEEE Press, 2004: 669~674.
- [41] 马加庆, 韩崇昭. 一类基于信息融合的粒子滤波跟踪算法.光电工程,2007,34(4):22~25.
- [42] 查宇飞, 毕笃彦. 一种基于粒子滤波的自适应运动目标跟踪方法.电子与信息学报, 2007, 29(1): 92~95.
- [43] Comaniciu D, Ramesh V and Meer P. Real-time tracking of non-rigid objects using mean shift.IEEE Conference on Computer Vision and Pattern Recognition Hilton Head Island,South Carolina.2000,2:142~149.
- [44] Kailath T. The divergence and Bhattacharyya distance measures in signal selection.IEEE Trans.on Commun.Technol.,1967,15(1):52~60.
- [45] 高建坡,王煜坚,杨浩,吴镇扬.以颜色和形状直方图为线索的粒子滤波人脸跟踪,中国图象图形学报,2007,12(3):466~473.
- [46] Djuric P M, Kotecha J, Zhang J, et al. Particle Filtering:A review of Theory and How It Can Be Used for Solving Problems in Wireless Communications.IEEE Signal Processing Magazine, 2003:19~38.
- [47] Miguez J, Djuric P M. Blind Equalization by Sequential Importance Sampling. In Proc of IEEE ISCAS, Phoenix, 2002:845~848.
- [48] Puskas E, Andrieu C, Doucet A, etc. Particle Filtering for Demodulation in Fading Channels with Non-Gaussian Additive Noise.IEEE Trans on Communications, 2001, 49(4): 579~582.

- [49] Yang Z, Wang X. A Sequential Monte Carlo Blind Receiver for OFDM Systems in Frequency-Selective Fading Channels. *IEEE Trans on Signal Processing*, 2002, 50(2): 271~280.
- [50] Christophe A, Arnaud D, Azzedine T. Adaptive Map Multi-User Detection for Fading CDMA Channels. in *Proc.IEEE Workshop SSAP*, Pocono Manor, PA,2000:6~9.
- [51] Punskeya E, Doucet A, Fitzgerald W J. Particle Filtering for Multiuser Detection in Fading CDMA Channels. In *Proc.11th IEEE Workshop SSP*, Singapore,2001:38~41.
- [52] Reed J H, Krizman K J, Woerner B D, and Rappaport T S. An overview of the challenges and progress in meeting the e-911 requirement for location service. *IEEE Communication. Magazine*, 1998,36(4):30~37.
- [53] Zhao Y. Standardization of mobile phone positioning for 3G systems. *IEEE Communication. Magazine*, 2002,40(7):108~116.
- [54] Hellebrandt M, Scheibenbogen M. Location tracking of mobiles in cellular radio networks. *IEEE Trans.on Veh.Technol.*,1999,48(9):1558 - 1562.
- [55] Porretta M, Nepa P, et al. A novel single base station mobile location algorithm:Description and validation using deterministic propagation model. *3G Mobile communication Technologies*, 2003 4th International Conference on (Conf.Publ.No.494),Pisa Univ., Italy, 2003: 277~281.
- [56]]王丽,刘正军胡捍英.基于粒子滤波的蜂窝网移动台单站跟踪. *电子与信息学报*, 2007, 29(7):1556~1559.
- [57] Yang H Y,Zhang G L.Design and realization of a new correlation tracker algorithm[J]. *J.Infrared Millim Waves*, 2000,19(5):377~380.
- [58] Li X, Ni G Q. Optical flow computation of infrared image[J]. *Infrared and Laser Engineering*, 2002, 31(3):189~193.
- [59] Zhao J F, Wang M H. A study on missile plume tracking and localizing by means of forward looking infra-red[J].*Journal of Solid Rocket Technology*,2000,23(4):64~68.

- [60] Ling J G, Liu E Q, Yang J, et al. Approach of infrared small target motion prediction and tracking based on filter[J].J.Infrared Milim waves,2005,24(5):366~369.
- [61] 程建,周越,蔡念,杨杰.基于粒子滤波的红外目标跟踪.红外与毫米波学报,2006,25 (2): 113~117.
- [62] 康莉,谢维信,黄敬雄.基于 unscented 粒子滤波的红外弱小目标跟踪.系统工程与电子技术,2007, 29(1):1~4.
- [63] 胡洪涛,敬忠良,胡士强.基于辅助粒子滤波的红外小目标检测前跟踪算法.控制与决策, 2005, 20(11):1208~1211.
- [64] 胡适友,等.多模复合制导技术情报研究[J].飞航导弹情报研究报告文集(二), 1996:173~189.
- [65] Lin K P. FOA Pursuing Dual Millimeter-Wave PIR Sensors[J]. Aviation Wave&Space Technology, 1995,7(9):63~64.
- [66] 崔智社,曾涛,龙腾.基于数据融合提高红外/毫米波复合导引头跟踪精度的新算法[J].兵工学报, 2003,24(2):199~202.
- [67] 何友,等.雷达数据处理及应用[M].北京:电子工业出版社,2006:12~13.
- [68] 张婷,李靖,刘盼.基于粒子滤波的红外 P 毫米波复合制导目标跟踪方法.制导与引信,2006, 27(3): 19~24.
- [69] Singer R A, Stein J J. An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems, Proceedings of the 1971 IEEE Conference on Decision and Control, Miami Beach,1971:171~175.
- [70] Yaakov B. Extension of the probabilistic data association filter in multi-target tracking, Proceedings of the 5th Symp. On Nonlinear Estimation, 1974:16~21.
- [71] Fortmann T E, Bar-Shalom Y, Scheffe M. Multitarget tracking using joint probabilistic data association, Proceedings of 19th IEEE Conference on Decision and Control, Albuquerque: NM, 1980:807~812.
- [72] Yaakov B, William D B. Multitarget-Multisensor Tracking Applications and Advances,

Volume111, Artech House, 2000.

- [73]Poore A, Rijavee N. Multitarget Tracking , Multidimensional Assignment Problems, and Lagrangian Relaxation, Proc. SDI Panels on Tracking, 1991:51~74.

第6章 目标跟踪典型问题

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，进而为平台随动控制、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^[1]。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。

6.1 快速运动目标跟踪

由于现实情况的需求，目标跟踪往往要求具备实时性好的特点。然而，为了准确地获取目标的信息，目标表达的特征越来越复杂，Haar 小波、轮廓特征、纹理特征、颜色特征及其组合被越来越多地应用在了目标表达中来，因此目标匹配搜索的时效性对目标跟踪的最终算法性能有很大的影响。本节我们主要对目标的快速跟踪进行介绍与讨论。

6.1.1 基于积分图的运动目标特征快速提取

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，但是大多存在跟踪精度不高、特征提取匹配的计算复杂度太大、鲁棒性不够等问题，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。

将目标跟踪看成前景背景像素点分类问题，在上一帧采样前景和背景像素点估计每一个特征的贝叶斯错误率，采用贝叶斯错误率评价每一个特征的分类能力,并选择具有最小错误率的若干个特征构建贝叶斯分类器组合。然后，贝叶斯分类器组合被用来赋予当前视频帧的像素点属于前景像素点的概率。最后，粒子滤波被用来在概率图上对运动目标进行跟踪。一种新的同时考虑了区域和边界信息的粒子滤波观察模型被提了出来，使得粒子滤波能够很好地处理目标的尺度变化问题。此外，借助一种叫做积分图的数据结构，粒子滤波的观察模型可以被快速地计算出来。

模板匹配作为一种常用的相似度度量方法，在目标变化较少的情况下，具有很高的准确性，

因此在目标跟踪中有着广泛地应用。但是在匹配过程中候选项的特征提取需要进行大量的计算，当目标较大或者搜索区域较大时，往往不能达到实时性的要求^[2]。在已有的算法中，有的通过缩小搜索区域的方法来加快匹配的速度，有的通过改进特征提取的方法减少提取的时间，还有一种被广泛使用的方式是特征积分图的方法来减少候选项特征集的计算时间。

在目标搜索区域的每个候选项特征提取前，我们可以通过将整个搜索区域的特征积分图提取，然后通过特征积分图的加减运算得到每个候选项的特征向量。这样，搜索区域中任何位置的特征向量可以有积分图中该位置以及前面左上、左、上等四个位置的特征向量积分制通过加减运算得到，大大减少了目标特征提取的时间。如图 6.1 所示。

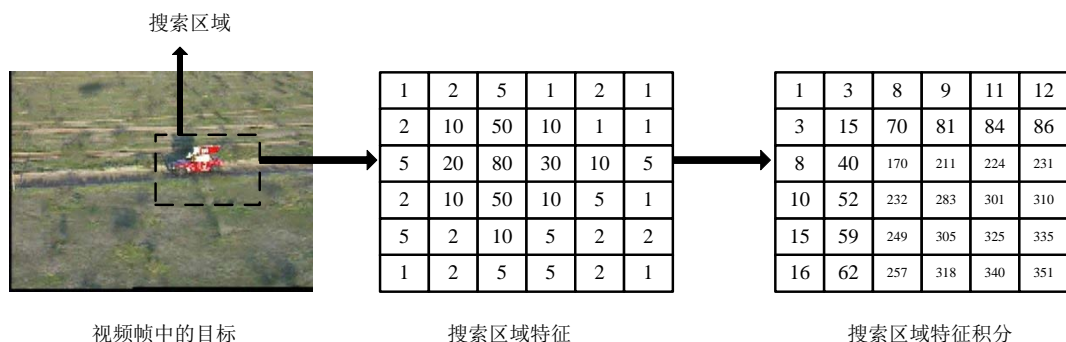


图 6.1 搜索区域积分示意图

6.1.2 Mean shift 跟踪

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^[3]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。

6.1.2.1 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:

$$M_h(x) \equiv \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad (6-1)$$

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,

$$S_h(x) \equiv \left\{ y : (y-x)^T (y-x) \leq h^2 \right\} \quad (6-2)$$

其中, 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 6-1 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。

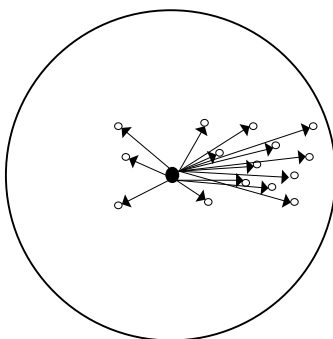


图 6.2 Mean Shift 示意图

如图 6.2 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。

6.1.2.2 扩展 Mean Shift

从公式 6-1 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:

$$M(x) \equiv \frac{\sum_{i=1}^n G_H(x_i - x)w(x_i)(x_i - x)}{\sum_{i=1}^n G_H(x_i - x)w(x_i)} \quad (6-3)$$

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 6-3 又可以被改写为:

$$M_h(x) \equiv \frac{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right)w(x_i)(x_i - x)}{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right)w(x_i)} \quad (6-4)$$

可以看到,如果对所有采样点 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 6-4 完全退化为公式 6-1,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。

6.1.2.3 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:

$$\hat{q}_u = C \sum_{i=1}^n k \left(\left\| \frac{x_i^s - x_0}{h} \right\|^2 \right) \delta [b(x_i^s) - u] \quad (6-5)$$

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 \mathcal{Y} 的目标可以描述为：

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k \left(\left\| \frac{x_i^s - y}{h} \right\|^2 \right) \delta [b(x_i^s) - u] \quad (6-6)$$

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 Bhattacharrya 系数 $\hat{\rho}(y)$ 来度量，即

$$\hat{\rho}(y) \equiv \rho[p(y), q] = \sum_{u=1}^m \sqrt{p_u(y) \hat{q}_u} \quad (6-7)$$

公式 6-7 在 $\hat{p}_u(\hat{y}_0)$ 点泰勒展开可得，

$$\rho[p(y), q] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p(y_0) q_u} + \frac{1}{2} \sum_{u=1}^m p_u(y) \sqrt{\frac{q_u}{p_u(y_0)}} \quad (6-8)$$

把公式 6-6 带入公式 6-8，整理可得：

$$\rho[p(y), q] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p(y_0) q_u} + \frac{C_h}{2} \sum_{i=1}^n w_i k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \quad (6-9)$$

其中，

$$w_i = \sum_{u=1}^m \delta [b(x_i) - u] \sqrt{\frac{q_u}{p_u(y_0)}} \quad (6-10)$$

对公式 6-10 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，在目标区域已知的前提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。

6.2 变化及复杂背景下的目标跟踪

6.2.1 特征自适应选择和评估

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。

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，即基于滤波框架的综合特征评估。

6.2.1.1 目标区域与背景区域定义

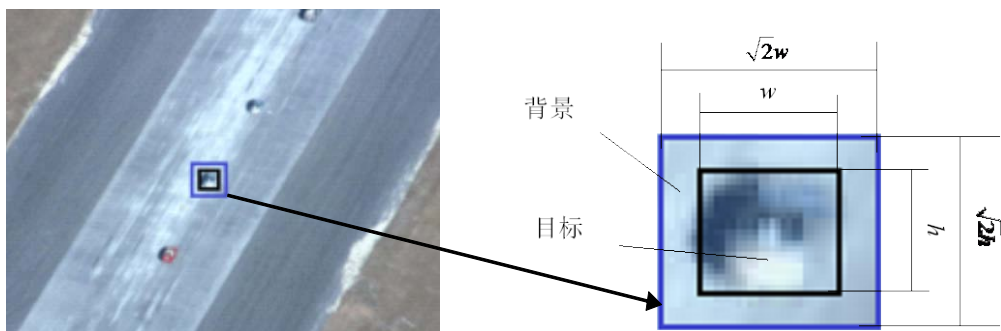


图 6.3 跟踪目标及其背景区域定义

如图 6.3 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。

6.2.1.2 特征判别力的定义

依照 Collins^[4]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 6-11 和 6-12 所示:

$$\widetilde{S}_t^i = \max(0, \min(1, \log \frac{\max(F_t^i(x, y), \delta)}{\max(B_t^i(x, y), \delta)})), i = 1 \dots N \quad (6-11)$$

$$S_t^i = \frac{\widetilde{S}_t^i}{\sum_{i=1}^N \widetilde{S}_t^i}, i = 1 \dots N \quad (6-12)$$

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, δ 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, \widetilde{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 \widetilde{S}_t^i 的值在 (0.0, 1.0) 区间范围。公式 6-12 用来归一化 \widetilde{S}_t^i 。

6.2.1.3 特征权重的计算

基于公式 6-11 和公式 6-12, 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 $w_{t-1}(j)$,

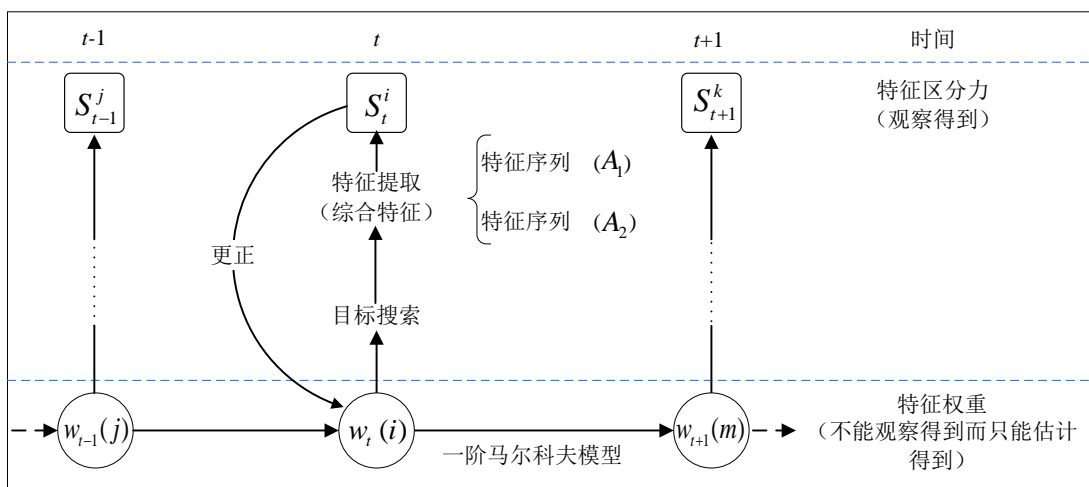
如公式 6-13 所示：

$$w_t(i) = f_{t,t-1}(w_{t-1}(j), S_t^i) + u_t \quad (6-13)$$

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 6-14 用来对特征的权重进行归一化。

$$w_t(i) = \frac{w_t(i)}{\sum_{i=1}^N w_t(i)}, i = 1 \dots N \quad (6-14)$$

根据公式 6-13 以及 6-14 中权重的定义，基于滤波框架的特征评估的框图如图 6.4 所示：



图

6.4 基于滤波框架特征评估框图

6.2.2 稀疏自适应特征表示

最新关于人类视觉系统（Human Vision System, HVS）的研究表明^[5]，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^[6]。同时，其他领域的相关研究也表明^[7-9]：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：

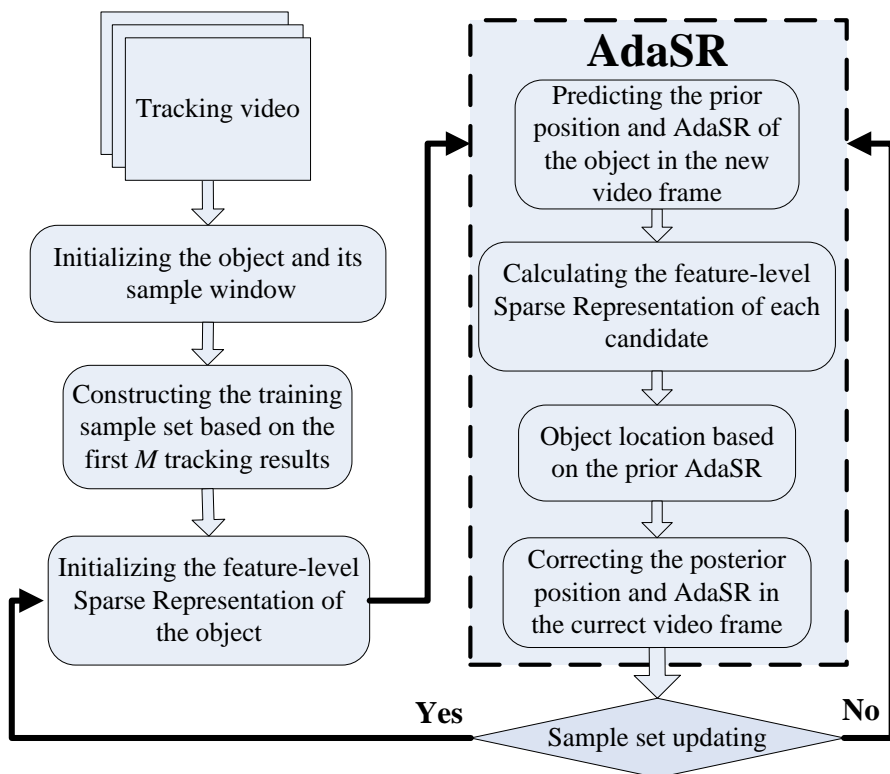


图 6.5 基于自适应稀疏表示的目标跟踪算法流程图

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。

6.2.2.1 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 ℓ_j^+ 与反例 ℓ_j^- ，通过 HOGC 特征的提取，获得构建 $2M$ 个训练样本 $training_set = \{(\ell_j^+, \ell_j^-)\}, j=1...M$ 。

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 (ℓ_j^+, ℓ_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^[10]，保证了跟踪的稳定性。

6.2.2.2 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：

$$\begin{aligned} \min \quad & \|S\|_1 \\ \text{s.t.} \quad & \begin{cases} S^T \ell_j^+ \geq \alpha \\ S^T \ell_j^- \leq -\alpha \end{cases}, \end{aligned} \quad (6-15)$$

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； α 保证了正例样本和反例样本之间的距离最少为 2α 。

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 (6-15) 便可被重新写为下面的凸优化模型：

$$\begin{aligned} \min \quad & I_N^T \eta + I_N^T \mu \\ \text{s.t.} \quad & \begin{cases} (\eta - \mu)^T \ell_j^+ \geq \alpha \\ (\eta - \mu)^T \ell_j^- \leq -\alpha \\ \eta \geq 0 \\ \mu \geq 0 \end{cases} \end{aligned} \quad (6-16)$$

经过稀疏特征选择之后的特征如下图所示：

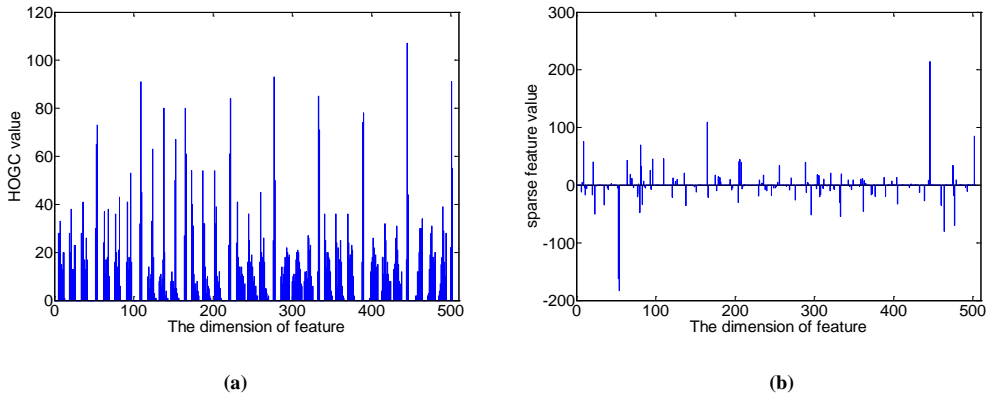


图 6.6 HOGC 特征与稀疏 HOGC 特征对比图。(a) HOGC 特征，(b) 稀疏特征选择之后的 HOGC 特征。

6.2.2.3 基于自适应稀疏特征的目标跟踪

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：

$$\begin{cases} \begin{pmatrix} \widetilde{w_{t+1}} \\ \widetilde{\Delta w_{t+1}} \\ \widetilde{Pos_{t+1}} \\ \widetilde{\Delta Pos_{t+1}} \end{pmatrix} = \begin{pmatrix} I_{N \times N} & I_{N \times N} & 0 & 0 \\ 0 & I_{N \times N} & 0 & 0 \\ 0 & 0 & I_{L \times L} & I_{L \times L} \\ 0 & 0 & 0 & I_{L \times L} \end{pmatrix} \begin{pmatrix} w_t \\ \Delta w_t \\ Pos_t \\ \Delta Pos_t \end{pmatrix} + u_t \\ \\ \begin{pmatrix} S_t \\ mPos_t \end{pmatrix} = \begin{pmatrix} I_{N \times N} & 0 & 0 & 0 \\ 0 & 0 & I_{L \times L} & 0 \end{pmatrix} \begin{pmatrix} w_t \\ \Delta w_t \\ Pos_t \\ \Delta Pos_t \end{pmatrix} + v_t \end{cases} \quad (6-17)$$

其中, $w_t = \{w_t(i), i=1 \dots N\}$ 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。

目标跟踪是对跟踪目标在搜索区域 Ω_t 内进行搜索的过程。跟踪的目的为在 Ω_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:

$$\begin{aligned} & \max_{(x, y)_t} p((x, y)_t | \Omega_t, (x, y)_{t-1}) \\ &= \max_{(x, y)_t} \int p(F_t((x, y)_t^c) | (x, y)_t^c, F_{t-1}) p((x, y)_t^c | \Omega_t) dc \quad , \quad (6-18) \\ &= \max_{(x, y)_t} \int \left(\sum_{i=1}^N \left(p(F_t^i) p(F_t^i((x, y)_t^c), F_{t-1}^i) \right) \right) p((x, y)_t^c | \Omega_t) dc \end{aligned}$$

其中, $(x, y)_t^c$ 定义为 Ω_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 6-18 可以如下描述:

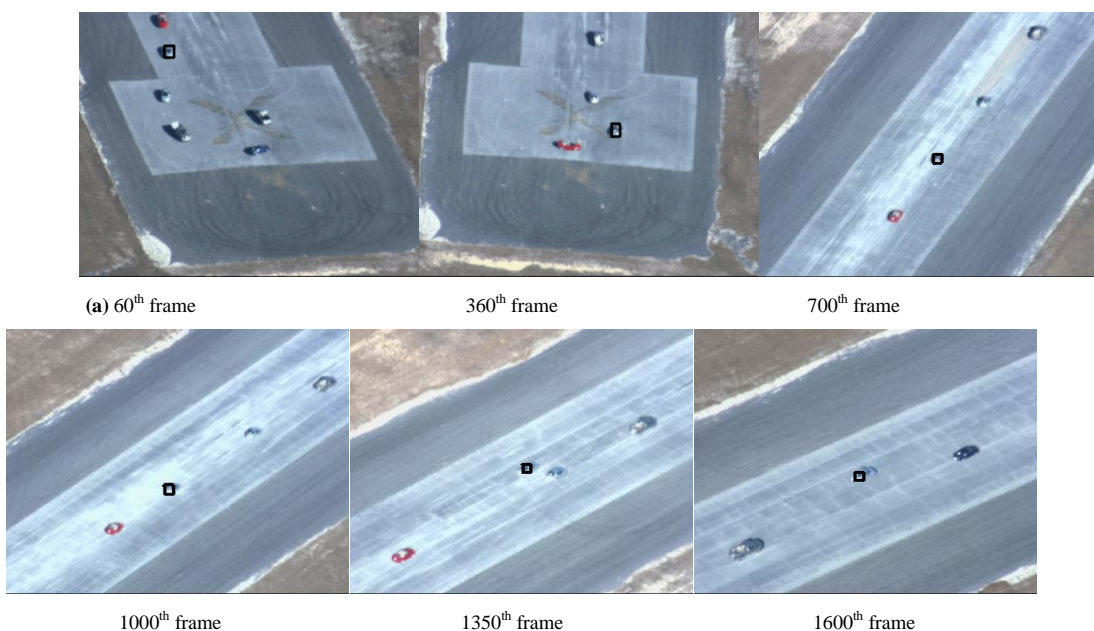
$$\begin{aligned} & \max_{(x, y)_t} p((x, y)_t | \Omega_t, (x, y)_{t-1}) \\ &= \max_{(x, y)_t} \int \left(\sum_{i=1}^N \left(w_t(i) p(F_t^i((x, y)_t^c), F_{t-1}^i) \right) \right) p((x, y)_t^c | \Omega_t) dc \quad . \quad (6-19) \end{aligned}$$

6.2.3 实例分析

第一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。

第二段测试视频来至于 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）。

跟踪目标自适应的稀疏表示选择特征集中最具前景/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 6.7 中的目标跟踪结果表明了跟踪算法的有效性。



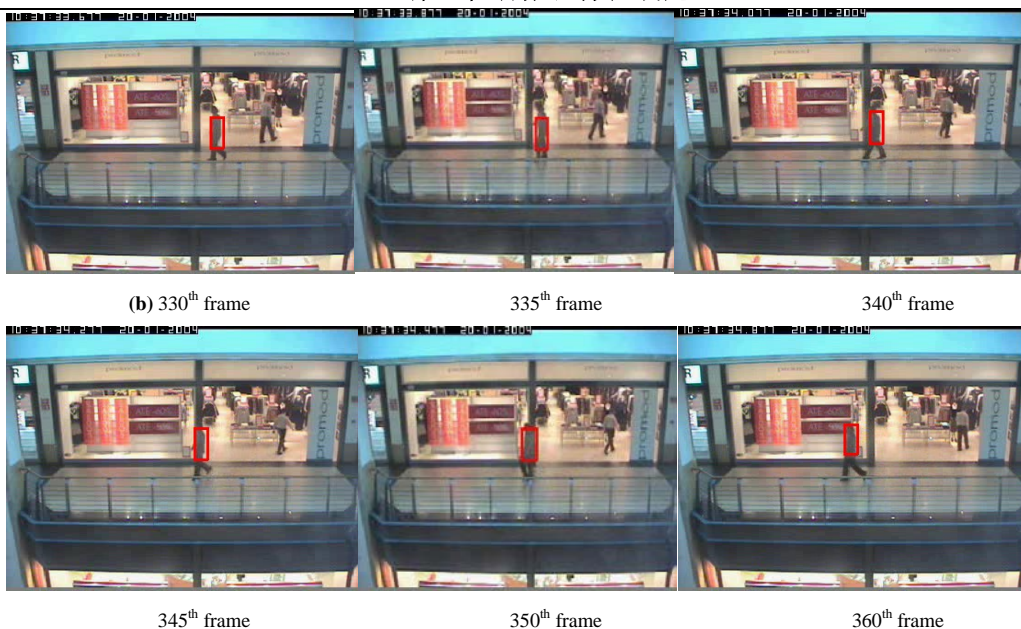


图 6.7 跟踪结果示例

6.3 遮挡情况下的目标跟踪

6.3.1 基于轨迹预测运动目标跟踪

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，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。

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。

表 6.1 基于粒子滤波的特征权重评估算法

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 $c_t^i \geq r$ ；
- (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 $p(i | i') \sim N(0, \delta)$ ，其中 δ 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图 6.8 所示：

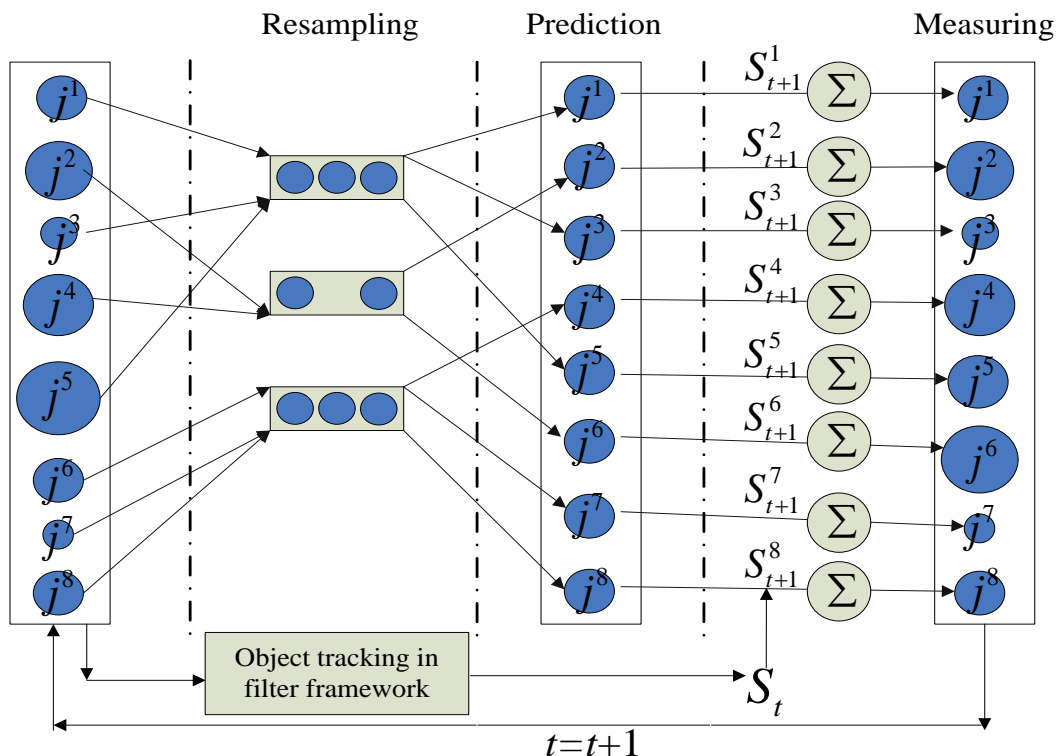


图 6.8

基于粒子滤波的特征权重评估示意图

6.3.2 基于目标稀疏重构的目标跟踪

按照 Li 的思路^[1]，基于稀疏表示的目标跟踪中的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）。

6.3.2.1 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：

$$A\psi \approx F \quad (6-20)$$

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， ψ^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，所以我们可以通过最小化 $\|\psi\|_0$ 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：

$$\arg \min \|\psi\|_1, \text{ subject to } A\psi = F \quad (6-21)$$

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：

$$\arg \min \|\psi\|_1, \text{ subject to } \|A\psi - F\|_2 \leq \varepsilon \quad (6-22)$$

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 6.9 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 rm 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。

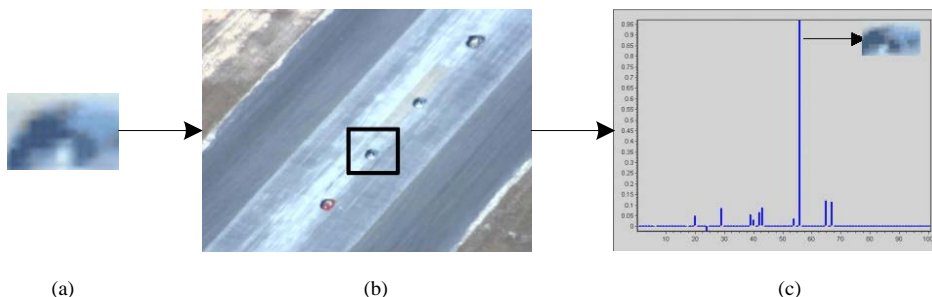


图 6.9 (a) 跟踪目标 (b) 搜索区域 (c) 稀疏重构系数

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

$$F \approx A\psi = \sum_{k=1}^K A^k \psi^k \quad (6-23)$$

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。

6.3.2.2 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 Ω_t 的候选集之间的差异来完成, 公式如下:

$$\text{Min}_{(x,y) \in \Omega_t} \left(\|F - F(C_t(x,y))\|_1 \right) = \text{Min}_{(x,y) \in \Omega_t} \left(\left\| \sum_{k=1}^K A^k \psi^k - F(C_t(x,y)) \right\|_1 \right) \quad (6-24)$$

其中 $C_t(x,y)$ 是搜索框中候选集 Ω_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。

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。

6.3.2.3 实验步骤

表 6.2 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 ψ ;
 - 1.3 基于样本 S 和稀疏系数 ψ 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 Ω_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 Ω_{t+1} ;
3. $t=t+1$. 在新的时刻, 如果 Ω_{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。

6.3.3 实例分析

第一段测试视频来至于 CARVIA 数据测试集合, 此测试视频最大的挑战主要来至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。

第二段测试视频来至于 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)。

由于本节所提的跟踪算法, 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 6.10 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。



第 6 章 目标跟踪典型问题

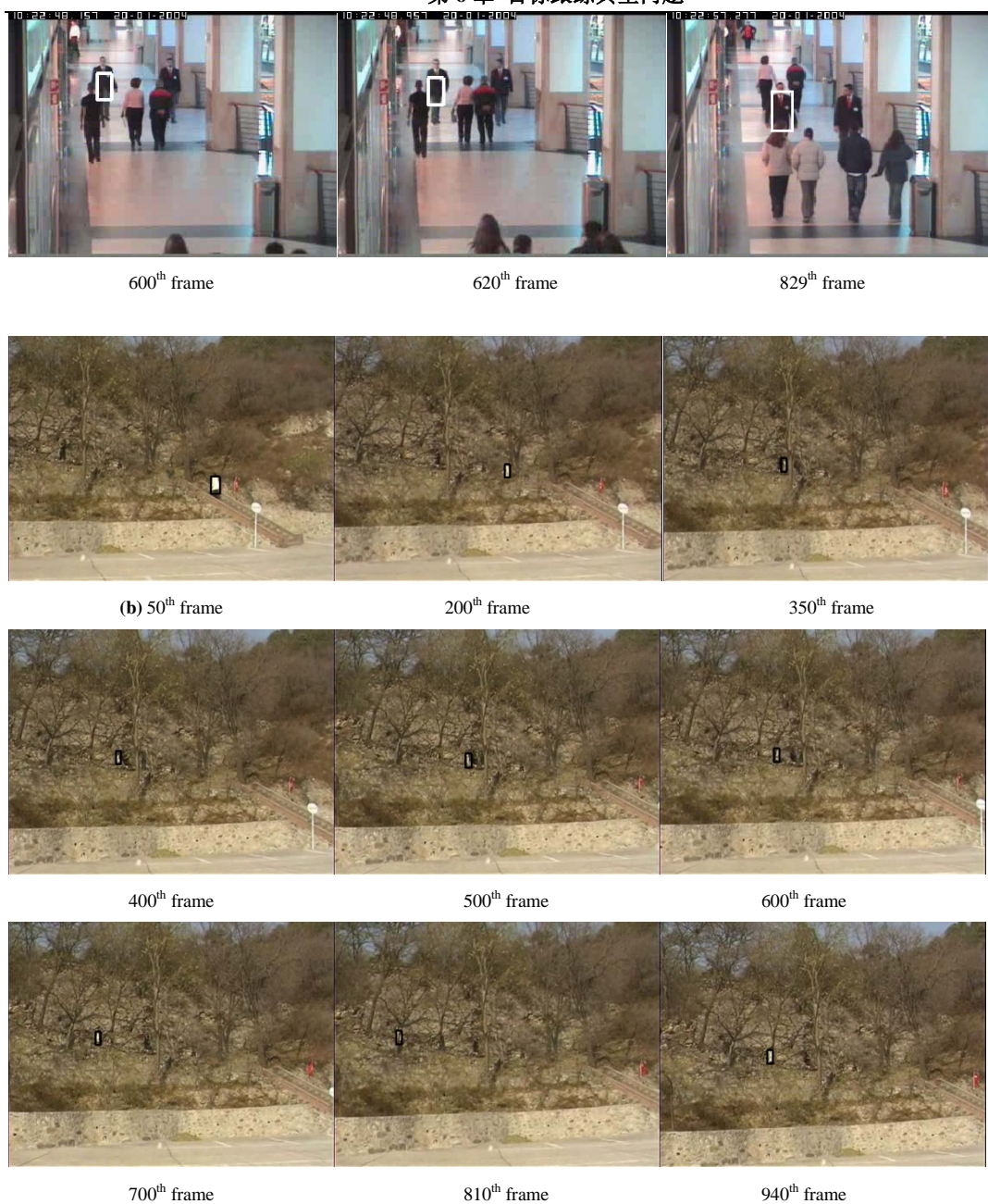


图 6.10 四段测试视频的跟踪结果实例

6.4 参考文献

- [1] 韩振军. 视觉目标自适应跟踪算法研究. 中国科学院研究生院博士学位论文, 2011: 1~2.
- [2] 刘一飞. 基于稀疏表示的快速目标跟踪算法研究. 中国科学院研究生院硕士学位论文, 2012:

31~32.

- [3] Comaniciu D, Ramesh V and Meer P. Real-time tracking of non-rigid objects using mean shift. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000: 142~149.
- [4] Collins R, Liu Y and Leordeanu M. Online selection of discriminative tracking features. IEEE Trans. Pattern Analysis and Machine Intelligence, 2005: 1631~1643.
- [5] Serre T. Learning a dictionary of shape-components in visual cortex: Comparison with neurons. humans and machines, Ph.D dissertation, MIT. 2006.
- [6] Wright J, Yang A Y, Ganesh A, Sastry S S and Ma Y. Robust face recognition via sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence, 2009: 210~227.
- [7] Rissanen J. Modeling by shortest data description. Automatica, 1978: 465~471.
- [8] Hansen M and Yu B. Model selection and the minimum description length principle. Journal of the American Statistical Association, 2001: 746~774.
- [9] Yang J, Frangi A F, Yang J Y, Zhang D and Jin Z. KPCA Plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition. IEEE Trans. Pattern Analysis and Machine Intelligence, 2005: 230~244.
- [10] Matthews I, Ishikawa T and Baker S. The template update problem. IEEE Trans. Pattern Analysis and Machine Intelligence, 2004: 810~815.
- [11] 李莉. 基于样本空间目标重构的视频跟踪算法研究. 中国科学院研究生院硕士学位论文, 2011: 43~46.

后 记

《基于视频信息的目标检测与跟踪》经过近一年的努力，终于奉献给读者了。本书意在紧密跟踪国民经济发展中急需解决和关注的技术难题，采用理论、热点、实例相结合的方法，比较全面地向读者介绍计算机视觉中目标检测与跟踪的基本理论、技术难点和最新动向，以便帮助读者进一步熟悉基于视频信息的目标检测与跟踪技术，以此推动这项技术的发展。由于时间和篇幅有限，有些内容阐述不够翔实深入，读者可参考相关参考文献做进一步了解。

本书的研究工作得到了国家自然科学基金重点项目（61039003）《基于多源数据的飞行器近威胁目标检测跟踪及行为预测》课题、国家 973 计划项目（2010CB731804-2）子课题《飞行器威胁目标识别与图像鲁棒匹配理论与方法》、国家 973 计划项目（2011CB706901）课题《事故致灾过程和事故致因理论》、中国科学院研究生院“百人计划”择优支持项目（99T3009EA2）课题《复杂环境下动态目标跟踪技术》、航天科工哈尔滨风华有限公司科技攻关项目《风机塔筒焊缝自动检测系统》和《天目 360 度智能监控系统产业化开发》，以及国家自然科学基金面上项目（60672147）课题《复杂环境下新型多目标动态检测识别追踪算法的研究》的资助，取得了专利《基于 Internet 和 GRPS 的智能视频监控系统》和《基于人脸检测的错误阅读姿势检测方法》两项。书中包含的实例和研究工作部分来自于中国科学院研究生院模式识别与智能系统实验室的博士和硕士学位论文，包括：叶齐祥的《图像和视频文字检测技术研究》；徐冉的《基于 1 范数最小化学习的人体检测算法研究》；韩振军的《视觉目标自适应跟踪算法研究》；孔凡静的《基于图像的路面车辆检测算法研究》；张宁的《图像人体目标的快速及鲁棒性检测算法研究》；李莉的《基于样本空间目标重构的视频跟踪算法研究》。在此，谨一并表示十分感谢！

作者

2014 年 4 月