摘　要

视频监控系统以其直观、方便、信息内容丰富而成为现代安防系统发展的主流，被广泛应用于国防建设、交通管制、智能保安等需要实时监控的场所。现有的大多数视频监控系统仍依赖于监控人员的现场操作，造成了人力资源的浪费，也影响了整个工作系统的效率。因此，研究智能视频监控系统中的关键技术并提高智能视频监控系统的性能具有重要的意义。

运动目标检测与跟踪是视频监控系统中一个重要组成部分，对计算机视觉的其他研究领域有着重要的推动作用。因此如何实现对运动目标实时、稳定和有效地检测与跟踪，是一个需要关注和研究的重要问题。本文在目前该领域研究成果的基础上，系统研究了智能视频监控中人体目标的检测，分割和跟踪方面的理论和方法。

在运动目标检测算法中，详细介绍了几种常用的目标前景检测方法，并对它们的检测性能进行了评价。在运动人体目标的定位与分割方面，以人体的头部研究为出发点，针对人体头部运动信息的轮廓近似圆形的特征，结合 Freeman 链码和 RANSAC 算法，进行多圆检测来实现多目标头部的定位，进而较快地确定人体目标的准确位置。

在目标跟踪算法中，针对目前 Mean Shift 跟踪算法存在的问题，本文提出了采用目标的颜色信息、纹理信息和运动信息的改进 Kalman 和 Mean Shift 跟踪算法，跟踪效果得到较大改进。针对单个目标运动时姿势的显著变化，本文在机器学习理论知识的基础上，提出了一种基于 RGB 直方图特征、LBP 直方图特征和 PPBTF 直方图特征的半监督学习对单目标进行跟踪的方法，使跟踪效果更具鲁棒性。

关键词：运动目标检测；头部定位；Mean Shift 跟踪；Kalman 滤波；Tri-tracking 算法

Research on Moving Object Detection and Tracking Algorithms based on Video Image

Abstract

Video surveillance system(VSS) is the main trend of modern defence system because of its visibility, convenience and abundance in content and has been widely used in many fields where a real-time surveillance is needed, such as national defense, traffic control, the intelligent public security and so on. Nowadays, VSS still depends on manual operation, which wastes resources and affects the efficiency, so studying the typical algorithms used in video surveillance and designing an intelligent video surveillance system is very important.

Moving object detection and tracking are important parts of video monitor system and play important roles to other topics' progress in computer vision. So how to detect and track object steadily, real-timely and effectively, becomes an important problem that needs to be paid attention and researched. The paper studied the key technologies of the field based on the current research achievements and mainly studied about the technologies on human object detection, segmentation and tracking.

During object detection, several object foreground detection algorithms widely used are introduced, and their performances are analyzed. During object localization and segmentation, due to head contours similar to circles seen from the video sequences captured from vertical angular camera, the paper combines Freeman chain code and Random Sample Consensus (RANSAC) algorithm to estimate circle parameters and then localizes the human object quickly and accurately.

During object tracking, in the light of current Mean Shift tracking algorithm's drawbacks, Kalman filter and Mean Shift tracking algorithm integrating with color, texture and motion information is proposed and its tracking performance has been improved than before. Aiming at the posture's significant change of a single object, a novel tracking algorthim based on RGB histogram feature, LBP (Local Binary Pattern) histogram feature and PPBTF (Pixel-Pattern-Based Texture Feature) histogram feature, using the semi-supervised learning is proposed on the basis of machine learning theory.

Key Words: Moving Object Detection; Head Localization; Mean Shift Tracking; Kalman Filter; Tri-tracking algorithm

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目：基于视频图像的运动目标检测与跟踪算法研究

作者签名：张瑞婧　　日期：2008 年 12 月 日

大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目：基于视频图像的运动目标检测与跟踪算法研究

作者签名：张瑞婧　　日期：2008 年 12 月 ___ 日

导师签名：卢湖川　　日期：2008 年 12 月 ___ 日

1 绪论

1.1 引言

随着视频分析技术及人工智能技术的发展，从图像处理与模式识别发展起来的计算机视觉在近三十年得到了突飞猛进的发展。尽管目前计算机视觉的理论发展仍然不够成熟，但已经得到了广泛的应用，在智能视频监控系统、机器人视觉导航、医学辅导诊断、工业机器人视觉系统、地图绘制，物理的三维重建与识别、智能人机接口等领域得到广泛的发展，其中智能化的视频监控系统是近年来计算机视觉领域的一个备受关注的前沿课题。

从图像处理与模式识别发展起来的计算机视觉研究方向主要是如何利用二维投影图像恢复三维景物世界。计算机视觉使用的理论方法主要是基于几何、概率、运动学与三维重构的视觉计算理论，它的理论基础包括射影几何学、刚体运动力学、概率与随机过程、图像处理、人工智能等。计算机视觉要达到的最终目的是实现计算机对三维景物世界的理解，即实现人类视觉系统的某些功能。

视频图像是对客观事物形象、生动的描述，是直观而具体的信息表达形式，是人类最重要的信息载体。人类主要通过视觉感知外界信息，据统计，人类对外界信息的感知有 80%以上是通过视觉得到的。在今天的信息社会，随着网络、通信和微电子技术的快速发展和人民物质生活水平的提高，视频监控以其直观方便和内容丰富等特点日益受到人们的青睐，监控产品也正经历着从最初的模拟化走向数字化、网络化、智能化的革命。

视频监控是实施安全监控的重要技术手段，它是计算机视觉领域一个新兴的应用方向和备受关注的前沿课题。它涉及信号与视频处理、通信和计算机视觉等多个学科的研究领域。视频通信、处理和理解是视频监测技术的三大核心技术。尽管成像设备、视频压缩、通信以及数据存储等方面的技术发展迅速并且日趋成熟，监控系统功能日益强大，但是视频内容的分析和理解工作目前仍然主要依靠人工完成，工作量繁重，因此计算机视觉和应用研究者提出新一代监控——视频监控的概念[1]，监控在不需要人为干预情况下，利用计算机视觉和视频分析的方法对摄像机拍摄的图像序列进行自动分析，实现对动态场景中目标的定位、识别和跟踪，并在此基础上分析和判断目标的行为，从而既能完成日常管理又能在异常情况发生时及时做出反应。计算机视频监控系统不仅符合信息产业的未来发展趋势，而且代表了监控行业的未来发展方向，蕴含着巨大的商机和经济效益，受到了学术界、产业界和管理部门的高度重视。

1.2 视频监控相关研究

计算机视频监控是指通过摄像头采集监控区域的图像，然后将视频信号通过相应的传输网络同轴电缆、光纤、无线或以太网等，传到指定的监控中心或是监控点，进行存储、显示、分析。具体分析过程是利用计算机视觉和图像处理方法对图像序列进行运动检测、运动目标分类、运动目标跟踪以及监视场景中目标行为的理解与描述。运动检测、目标分类和目标跟踪属于视觉中的低级处理部分，行为理解与描述则属于视觉中的高级处理部分。运动检测、目标分类与跟踪是视频监控中研究最多的三个问题，而行为理解与描述则是近年被广泛关注的研究热点。它是指对人的运动模式进行分析和识别，并用自然语言等加以描述。下面对视频监控系统有关发展状况及主要理论做简要概述。

1.2.1 国内外研究现状

智能视频监控是多学科交叉的前沿研究领域，并且具有广泛的应用前景和庞大的市场需求，有很多科研人员、科研机构及企业单位多年来从事这方面的研究、开发，并且取得了很多优秀的成果。

1997 年，美国国防高级研究项目署(DARPA)设立了以卡内基梅隆大学为首联合十几家高等院校和研究机构参加的视频监控重大项目 VSAM(Video Surveillance and Monitoring)[2,3]，主要研究对战场及普通民用场景进行监控的自动视频理解技术。2000 年，DARPA 又资助了 HID(Human Identification at a Distance)远距离人脸识别项目；由 Steve J. Maybank 和谭铁牛组织的 IEEE 视觉监控专题讨论会(VS，IEEE International Workshop on Visual Surveillance)也已经成功地举办了三届，收录了大量视觉监控领域内的最新研究成果。佛罗里达中央大学(University of Central Florida)的 KNIGHT 智能监控系统[3,4]利用计算机视觉技术能检测出监视区域目标的变化，并能标注重要事件和描述人的行为，系统对光照变化有较好的鲁棒性。英国雷丁大学[5]已开展了对车辆和行人的跟踪及其交互作用识别的相关研究。

计算机视觉领域中的权威期刊“国际计算机视觉期刊”(IJCV，International Journal of Computer Vision)和“IEEE 模式分析和机器智能汇刊”(PAMI，IEEE Transaction on Pattern Analysis and Machine Intelligence)相继在 2000 年 6 月和 2000 年 8 月出版了关于视觉监控的专刊。MaryLand 大学的实时视觉监控系统 W4[6]不仅能够定位人和分割出人的身体部分，而且通过建立外观模型来实现多人的跟踪，可以检测和跟踪室外环境中的人并对他们之间简单的交互进行监控。IBM 的智能监控系统[7]不仅能够自动监视某个现场，还能够监视数据，执行基于事件的检索，通过标准的网络设施进行实时报警，并且能提取出某项活动的长期的统计模型。

国内有一些视频监控方面的产品，如黄金眼、行者猫王等，应用于交通控制，监狱管理等方面。另外，国内产品还有数字硬盘录像系统(DVR)，将监控区域内有运动对象出现的情况录制下来，以备查询，该系统只是简单的检测出有无运动对象，而没有对运动对象做任何分析。

由于国内的研究起步较晚，技术还不够完善，开发出的产品距离智能化还有一定差距，在实际的应用中，受到很多限制，还有待于进一步的完善。

1.2.2 主要任务

人运动的视觉分析主要是针对包含人的运动图像序列进行分析处理，它通常涉及到运动检测、目标分类、人的跟踪及行为理解与描述几个过程，其一般性处理框架如图 1.1 所示。其中，运动检测、目标分类、人体跟踪属于视觉中的低级和中级处理部分(Low-level and Intermediate-level Vision)，而行为理解和描述则属于高级处理(High-level Vision)。当然，它们之间也可能存在交叉。下面从几个方面对视频处理主要任务进行探讨。

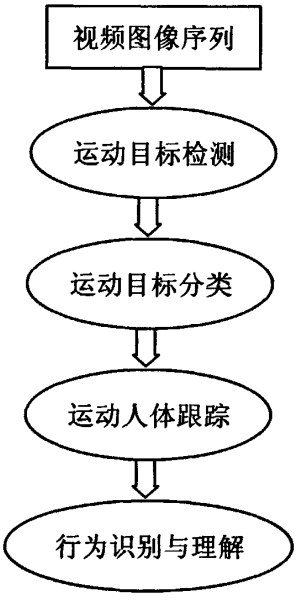


图 1.1 人运动分析的一般处理框架

Fig. 1.1 General framework of human motion analysis

(1)运动目标检测

运动检测是指从视频流中提取出目标，一般是确定目标的区域或颜色等。它必须从连续的视频流或图像序列中提取目标。目标检测的目的是从序列图像中将感兴趣的区域(一般为运动目标区域)从背景图像中抽取出来。运动区域的有效分割对于目标分类、跟踪和行为理解等后期处理是非常重要的，因为以后的处理过程仅仅考虑图像中对应于运动区域的像素。然而，由于背景动态变化，如天气、光照、影子及混乱干扰等的影响，使得运动检测成为一项相当困难的工作。目前几种常用的方法有：背景减除[5]、时间差分[11,12]、光流[13]、扩展的 EM 算法、能量运动检测、基于数学形态学的场景变化检测等。其中背景减除和时间差分均属于求差法，即都采用当前图像与参考图像进行相减的方法来完成运动目标的检测。

(2)运动目标分类

目标分类的目的是识别运动目标所属的类别，不同的运动区域可能对应于不同的运动目标，比如交通道路上监控摄像机所捕捉的序列中可能包含行人、车辆及其他诸如飞鸟、流云、摇动的树枝等运动物体，为了便于进一步对行人进行跟踪和行为分析，运动目标的正确分类是完全必要的。目前的分类方法主要有：基于形状信息的分类、基于运动特征的分类以及时间共生矩阵进行分层分类的方法等。

(3)运动目标跟踪

目标跟踪是计算机视觉领域中一个极具挑战性的课题，除视频监控外，其还被广泛地应用在人机交互、视频压缩、医学图像处理等领域中。所谓运动目标跟踪则指对目标进行连续的跟踪以确定其运动轨迹。它等价于在连续的图像帧间创建基于位置、速度、形状、纹理、色彩等有关特征的对应匹配问题。目标跟踪的关键在于得到图像检测中提取的静态目标与被跟踪运动目标的对应关系。

从运动检测得到的一般是人的投影，要进行跟踪首先要给需要跟踪的对象建立一个模型。对象模型可以是整个人体，这时候形状、颜色、位置、速度、步态等都是可以利用的信息；也可以是人体的一部分如上臂、头部或手掌等，这时需要对这些部分单独进行建模。之后，将运动检测到的投影匹配到这个模型上去。一旦匹配工作完成，那么就得到了最终有用的人体的信息了。常用的数学工具有卡尔曼滤波[14](Kalman Filtering)、Condensation算法[15]及动态贝叶斯网络[16]等。

(4)行为识别与理解

运动检测、目标分类与人的跟踪是人运动分析中研究较多的三个问题，而人的行为理解与描述是近年来被广泛关注的研究热点。人体行为识别与理解是指对人的行为模式进行分析和识别，并用自然语音加以描述，这种技术包含从视频序列中抽取相关的视觉信息并用一种合适的方法进行表达，然后解释这些视觉信息以实现识别和学习人的行

为。通常，人体行为识别是在成功实现图像序列中的人体检测和跟踪、完成特征提取的基础上进行的，属于更高一层的视觉任务，是当前计算机视觉研究领域备受关注并最具挑战的研究方向之一。

从模式识别的角度来看，行为识别可以看作为一个基于时变特征数据之上的分类问题，在获取人体目标的特征数据后，识别也就成为一个将未知类型数据序列与已知运动数据序列相匹配的过程。目前人行为识别的研究主要有头部运动、手势识别、下肢运动等，方法主要有模板匹配方法与状态空间方法两种。采用模板匹配技术的行为识别方法首先将图像序列转换为一组静态形状模式，然后在识别过程中和预先存储的行为标本相比较。基于状态空间模型方法是定义每个静态姿势作为一个状态，这些状态之间通过某种概率联系起来。任何运动序列可以看作为这些静态姿势的不同状态之间的一次遍历过程，在这些遍历期间计算联合概率，其最大值被选择作为分类行为的标准。目前，状态空间模型已经被广泛地应用于时间序列的预测、估计和检测。

1.2.3 技术难点

智能视频源于计算机视觉技术,其技术实现主要存在于以下几个方面问题：

(1)视频监控系统的鲁棒性要求很高,即要求自动、连续工作。而从实际情况来看,视频监控系统的应用环境往往相当复杂,环境的不断变化给计算机视觉技术应用带来了更高的要求。

(2)由于运动目标的多样性,如何对多种目标进行运动分析、分类甄别,尤其是对非刚性目标的运动分析尚存在一定难度。

(3)对于大范围的视频监控需要大量的摄像机协同工作,在多台摄像机之间对运动目标进行连续一致的视觉分析现在实现起来十分困难。

(4)如何建立有效的视频数据库,实现视频监控系统中海量视频信息的存储、检索和查询,目前还处于研究起步阶段。以上这些问题表明,视频监控系统的智能化还有相当长的路要走。

1.2.4 应用现状与发展趋势

视频监控系统是多媒体技术、计算机网络、工业控制和人工智能等技术的综合运用的产物，它正向着音视频的数字化、系统的网络化和管理的智能化方向不断发展。从视频监控技术的发展历程来看，视频监控系统在过去的二十多年里大致经历了三代。第一代模拟视频监控系统，第二代数字视频监控系统，第三代分布式视频监控系统[17]。

第一代视频监控系统是以模拟信号、图像的处理和传输为基础的，多路模拟摄像机产生的模拟信号通过同轴电缆传输到监控室，然后通过预置好的顺序轮流显示，监控人员通过监视器来判断监视场景的情况。

第二代监控系统主要依赖于混合模数或全数字的视频传输和处理方法，采用 H.263，MPEG 等多媒体数字压缩技术将视频图像完全数字化，节省了带宽资源。在视频监控中可以利用视频分析算法，让监控者只注意感兴趣的事物从而实现自动报警。

第三代监控系统利用低价位高性能的计算机网络、移动网络和固定的多媒体通信网络传输监控信号。视频信号在前端进行自动分析处理，然后将有价值的信息通过无线或有线网络传输到监控中心，实现自动视频监控。第三代的研究热点主要放在鲁棒性的图像传输，彩色图像的处理，基于事件的和基于模式的序列图像的理解上，目前对这些技术的应用已经获得了很多有价值的结果。与此同时，由于无线和有线的多媒体数字通信的发展，特别是随着超宽带网络的接入，第三代监控系统将在不同的领域中大范围使用。

随着人运动分析研究和其它相关技术的发展，下述几个方面已经成为未来智能视频监控的发展趋势：

1) 音频与视觉相结合的多模态接口

人的相互交流主要是依据语言，过去的许多工作是语音理解，但语音识别受距离和环境噪声的限制，尤其在机场等高噪声环境，将会严重影响语音识别的性能。人的可视化描述与语音解释一样重要，研究者们正逐渐将语音与视觉信息集成起来以产生更加自然的高级接口。目前音频和视频的信号处理相对独立，如何更好地集成音频和视频信息用于多模态用户接口是一个严峻的挑战。

2) 人的运动分析与生物特征识别相结合

在智能房间的门禁系统、军事安全基地的视觉监控系统、高级人机交互等应用中，人的运动分析与生物特征识别相结合的研究日益显得重要。在人机交互中不仅需要机器知道人是否存在、人的位置和行为，而且还需要利用特征识别技术来识别与其交流的人是谁。远距离的身份识别已经越来越重要，比如，2000 年，DARPA 赞助的重大项目——HID(Human Identification at a Distance)计划。由于人的步态具有易于感知、非侵犯性、难于伪装等优点，近年来已引起了计算机视觉研究者浓厚的研究兴趣。

3) 人的运动分析向行为理解与描述高层处理的转变

人的行为理解是需要引起高度注意并且是最具挑战的研究方向，因为观察人的最终目标就是分析和理解人的个人行为、人与人之间及人与其它目标的交互行为等。目前人的运动理解还是集中于人的跟踪、标准姿势识别、简单行为识别等问题，如人的一组最通常的行为(跑、蹲、站、跳、爬、指等)的定义和分类。近年来利用机器学习工具构建

人行为的统计模型的研究有了一定的进展，但行为识别仍旧处于初级阶段，连续特征的典型匹配过程中常引入人运动模型的简化约束条件来减少歧义性，而这些限制与一般的图像条件却是不吻合的，因此行为理解的难点仍是在于特征选择和机器学习。另外，如何借助于先进的视觉算法和人工智能等领域的成果，将现有的简单的行为识别与语义理解推广到更为复杂场景下的自然语言描述，是将计算机视觉低、中层次的处理推向高层抽象思维的关键问题。

1.3 本文的主要工作与章节安排

1.3.1 本文的主要工作

本文要研究的问题是如何从视频图像序列中提取出运动目标，并对其进行跟踪处理。视频监控根据摄像头的放置位置，大致分为两类：一类是摄像头随运动目标移动，运动对象始终保持在图像的中心位置；另一类是摄像头固定只对视场内的对象进行监控和跟踪。根据摄像头的数目可分为多摄像头协同监控和单摄像头监控。本文要研究的是垂直固定单摄像头的情况，这是多摄像头协同监控的基础。本文主要对运动区域的检测、运动人体目标的分割定位和运动人体目标的跟踪这三个方面进行了研究。

本文工作内容及创新点简要介绍如下：

(1)运动目标检测：深入研究了当前各种目标检测算法，详细介绍了帧间差分法、基于梯度建模前景提取法以及利用帧间二阶差分(SODP)与边缘检测相结合进行运动目标分割方法的基本原理，并进行了相关的实验性能分析。在帧间差分法的基础上，本文采用了一种基于均值统计的自适应阈值方法来进行运动区域分割[18]，较好地把人体的运动区域提取出来。

(2)运动目标定位：以人体头部为出发点对人体目标的定位进行了研究，在摄像头的角度为垂直的条件下，根据提取出的头部运动信息轮廓具有近圆形这一关键特征，本文采用 Freeman 链码和 RANSAC 算法相结合的方法，进行人体头部的检测识别，较快地完成了多目标的定位。

(3)多目标跟踪：研究了 Mean Shift 算法的基本原理，介绍了 Mean Shift 算法在目标跟踪中的应用。结合当前 Mean Shift 跟踪算法的优点，针对目标颜色相似与快速运动目标之间可能跟踪失败的情况，本文提出了一种融合目标颜色信息、纹理信息和运动信息(Kalman 滤波)等特征的目标跟踪方法，较好的完成了对人体目标的跟踪处理。

(4)单目标跟踪：对机器学习的发展进行了论述，对机器学习中的半监督学习进行了探讨，然后介绍了一种新的描述目标特征的 PPBTF 特征，对支持向量机给予了研究。针对目前跟踪算法建立模型描述目标的缺陷，利用半监督学习理论，提出了一种新的

Tri-tracking 单目标跟踪算法，较好地解决了当运动目标外观显著变化时跟踪失败的问题。

1.3.2 论文章节安排

本论文分为五章，各章内容安排如下：

第一章为绪论，阐述了智能视频监控的研究意义和国内外研究现状，讨论了计算机智能视频监控的主要任务、技术难点和应用现状及未来发展趋势，最后给出本文的主要工作和论文结构。

第二章对计算机智能视频监控中运动目标检测关键技术进行了研究。介绍了帧间差分法、基于梯度建模前景提取法以及利用帧间二阶差分(SODP)与边缘检测相结合进行运动目标分割方法的基本原理，并进行了相关的实验性能分析。本文在帧间差分法的基础上，采用了基于均值统计的自适应阈值方法[18]，从而判断出人体目标的运动区域。

第三章针对计算机智能视频监控中主要任务之一目标定位进行了研究。介绍了头部检测时采用的 Freeman 链码和 RANSAC 算法，结合两者对运动区域内人体目标头部进行定位。

第四章引入了描述运动目标的纹理特征，介绍了预测运动目标轨迹的 Kalman 滤波器，给出了 Mean Shift 理论及其在跟踪中的应用，然后提出了一种融合目标颜色信息、纹理信息和运动信息的改进的 Kalman 和 Mean Shift 跟踪算法，实现对运动目标的良好跟踪。

第五章分析了传统跟踪算法的现状，介绍了机器学习理论，针对建立模型进行跟踪的不足之处，本文利用机器学习的方法，提出了一种半监督学习跟踪算法——Tri-tracking 跟踪算法，对姿势显著变化的单个目标进行鲁棒性跟踪，并对提出的算法做了大量实验，同时给出了定性的分析。

最后对本论文进行了总结，指出了不足之处，并对今后工作进行了展望。

2 运动目标检测技术研究

2.1 引言

运动目标检测是应用视觉研究领域的一个重要课题，在军事和工业等领域如军事目标跟踪、交通自动导航、视频信号传输和机器人视觉等领域应用广泛。目前，视频信号的智能化处理需求日益增加，正确地从视频流中提取运动目标、判断运动方向是许多智能视频系统的基础部分。运动目标检测是智能视频监控系统中非常关键的一步，它的目的就是提取监控场景中的运动目标，为运动物体的识别跟踪提供必备条件。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。然而由于天气、光照、影子及混乱干扰等的影响，使得运动检测成为一项相当困难的工作。

根据视频序列图像中摄像机和场景之间是否运动将目标的运动划分为四种模式

(1)摄像机静止——目标静止。这是静态场景，对其处理方法就是静态图像中的处理方法。

(2)摄像机静止——目标运动。这是一类非常重要的动态场景，对其处理一般包括运动目标检测、目标特性估计等，主要用于预警、监视、目标跟踪等场合。

(3)摄像机运动——目标静止。这主要用于机器人视觉导航、电子地图的自动生成以及三维场景理解等。

(4)摄像机运动——目标运动，这是运动目标的检测和跟踪最复杂的一种情况，但也最普通的情况。目前关于这方面的研究还比较少，理论还没有成熟。

本文的目标检测算法研究是在垂直固定摄像头的前提下，综合考虑了室内和室外的监控算法，在下面的章节中将对常用的前景检测方法进行介绍并给出实际的检测结果和性能分析。

2.2 静态背景下的运动目标检测

2.2.1 基于背景建模的方法

基于背景建模的方法是目前运动目标检测中最常用的一种方法，它的基本思想是输入图像与背景图像进行比较，通过判定灰度等特征的变化，或用直方图等统计信息的变化来判断异常情况的发生和分割运动目标。简单常用的方式为：直接抽取视频序列中某一幅图像，或计算多幅图像的平均值作为背景。它一般能够提供最完全的特征数据，而对于动态场景的变化，如光照和外来无关事件的干扰等特别敏感，最简单的背景模型是

时间平均图像，这种方法有一些问题并且需要在没有前景物体时的一段训练周期。训练周期后背景物体的运动或是训练过程中前景物体将会认为是用旧的前景物体。此外，这种方法还不能处理场景中的梯度的亮度变化。大部分研究人员目前都致力于开发不同的背景模型和自适应算法，以期减少动态场景变化对于运动分割的影响。Haritaoglu[5]等利用最小、最大强度值和最大时间差分值为场景中每个像素进行统计建模，并且进行周期性的背景更新；McKenna[19]等利用像素色彩和梯度信息相结合的自适应背景模型来解决影子和不可靠色彩线索对于分割的影响；Karmann 与 Brandt[20]和 Kilger[21]采用基于卡尔曼滤波(Kalman Filtering)的自适应背景模型以适应天气和光照的时间变化；Stauffer 与 Grimson[22]利用自适应的混合高斯背景模型(即对每个像素利用混合高斯分布建模)，并且利用在线估计来更新模型，从而可靠地处理了光照变化、背景混乱运动的干扰等影响。

(a)背景差分法

背景差法首先对背景建立模型，选取背景中的一幅或几幅图像的平均作为背景图像，然后把以后的序列图像和背景图像相减，进行背景消除。若所得到的像素值大于某一阈值，则判定监视场景中有运动物体，从而得到运动目标。用公式表示如下：

$$\mathrm{d}=|I_L(x,y,i)-B_L(x,y)| \tag{2.1}$$

$$ID_L(x,y,i)=\begin{cases} d & d\geq\tau \\ 0 & d<\tau \end{cases} \tag{2.2}$$

式中：ID_L是背景帧差图；B_L是背景的亮度分量；i表示帧数($i=1,2,\cdots,N$)；τ为阈值。

(b)背景模型的建立

按照所处理背景的自身特性，背景模型可分为单模态和多模态两种。前者在每个背景点上的颜色分布比较集中，可以用单体概率分布模型来描述即只有一个模态，后者的分布则比较分散，需要多个分布模型来共同描述具有多个模态。最常用的描述背景点颜色分布的概率分布是高斯分布(正态分布)，下面就单模态和多模态两种情况下的背景模型分别加以说明和讨论。

以下用$\eta(x,\mu,\Sigma)$来表示均值为μ、协方差为Σ的高斯分布的概率密度函数。

1. 单高斯分布背景模型

单高斯分布背景模型适用于单模态背景情形，它为每个图象点的颜色分布建立了高斯分布表示的模型$\eta(x,\mu,\Sigma)$，其中下标 t 表示时间。设图象点的当前颜色度量为X_t，若$\eta(x,\mu_{t-1},\Sigma_{t-1})\leq T_p$(这里$T_p$为概率阈值)，则该点被判定为前景点，否则为背景点(这时又称为X_t与$\eta(x,\mu_{t-1},\Sigma_{t-1})$相匹配)。在实际应用中，可以用等价的阈值替代概率阈值。

如记$d_t = X_t - \mu_t$，以σ_t表示均方差，则常根据d_t/σ_t的取值设置前景检测阈值 T：如设$d_t/\sigma_t > T$，则该点被判定为前景点，否则为背景点。单高斯更新快慢的常数称为更新率α，则该点高斯分布参数的更新可表示为

$$\mu_t = (1-\alpha)\mu_{t-1} + \alpha X_t \tag{2.3}$$

$$\sigma_t^2 = (1-\alpha)\sigma_{t-1}^2 + \alpha(X_t - \mu_{t-1})^2 \tag{2.4}$$

2. 多高斯分布背景模型

多模态背景的情形则需要多个分布来共同描述一个图像点上的颜色分布。Stauffer 等提出了一种自适应混合高斯模型，对每个图像点采用了多个高斯的混合表示。设用来描述每个点颜色分布的高斯分布共有 K 个，分别记为$\eta(x, \mu_{t,j}, \Sigma_{t,j})$，$i = 1,2,\cdots,K$。各高斯分布分别具有不同的权值$\omega_{t,j}$($\sum_i \omega_{t,j} = 1$)，它们总是按照权值从高到低的次序排序。在检测前景点时，按照权值从大到小的次序将X_t与各高斯分布逐一匹配，若没有表示背景分布的高斯分布与X_t匹配，则判定该点为前景点，否则为背景点。

多高斯分布背景模型的更新较为复杂，因为它不但要更新高斯分布自身的参数，还要更新各分布的权重等。若检测时没有找到任何高斯分布与其匹配，则将权值最小的一个高斯分布取出，并引入一个新的均值为当前值的高斯分布，赋予较小的权值和较大的方差，然后对所有高斯分布重新进行权值归一化处理。若第m个高斯分布与之匹配，则对第i个高斯分布的权值更新如下：

$$\omega_{t,j} = \begin{cases} (1-\beta)\omega_{t-1,j} + \beta & i = m \\ (1-\beta)\omega_{t-1,j} & \text{其它} \end{cases} \tag{2.5}$$

其中β是另一个表示背景更新快慢的常数——权值更新率。以上公式表明只有与X_t匹配的高斯分布的权值才得到提高，其他分布的权值都被降低。另外，未匹配的高斯分布的均值μ和均方差σ保持不变，相匹配的高斯分布的参数也按照式(2.3)和式(2.4)进行更新。在更新完高斯分布的参数和各分布权值后，还要对各个分布重新进行排序。

2.2.2 基于光流场的方法

光流[23]是空间运动物体对观测面上的像素点运动产生的瞬时速度场，包含了物体表面结构和动态行为的重要信息，一般情况下，光流有相机运动、场景中目标运动、或是两者的共同运动产生。光流的计算方法大致可分为三类：基于匹配的、频域的或梯度的方法。

(1)基于匹配的光流计算方法包括基于特征和基于区域的两种。基于特征的方法不

断地对目标主要特征进行定位和跟踪，对大目标的运动和亮度变化具有鲁棒性。存在的问题是光流通常很稀疏，而且特征提取和精确匹配也非常困难。基于区域的方法先对类似的区域进行定位，然后通过相似区域的位移计算流量。这种方法在视频编码中得到广泛的应用，然而计算的光流仍不稠密。

(2)基于频域的方法利用速度可调的滤波组输出频域或相位信息。虽然能获得很高精度的初始光流估计，但往往涉及复杂的计算。另外，进行可靠评价也非常困难。

(3)基于梯度的方法利用图像序列的时空微分计算 2D 速度场(光流)。由于计算简单和较好的实验结果，基于梯度的方法得到了广泛研究。虽然很多基于梯度的光流估计方法取得了较好的效果，但是在计算光流时涉及到可调参数的人工选取、可靠性评价因子的选择困难，以及预处理对光流计算结果的影响，在应用光流对目标进行实时监测与自动跟踪时仍然存在很多问题。

总的说来，光流法的优点是能够检测独立运动的对象，不需要预先知道场景的任何信息，并且可用于摄像机运动的情况，但是由于噪声、多光源、阴影、透明性和遮挡性等原因，使得计算出的光流场分布不是十分可靠和精确；而且多数光流法计算复杂、耗时多，除非有特殊的硬件支持，否则很难实现实时检测[24]。日本大阪大学的 Ryuzo Okada [25]等人对此方法作了深入的研究，并已研制出比较成熟的系统。借助于多个数字信号处理器，这些系统都实现了实时目标检测和跟踪，处理速度可以达到 15 帧/s。

2.2.3 基于时间差分的方法

时间差分方法是将前后两帧或三帧图像相减，将差值大于某一阈值的部分判为运动对象。例如 Lipton[11] 等利用两帧差分方法从实际视频图像中检测出运动目标，进而用于目标的分类与跟踪；一个改进的方法是利用三帧差分代替两帧差分，如 VSAM[2,3]开发了一种自适应背景减除与三帧差分相结合的混合算法，它能够快速有效地从背景中检测出运动目标。时域差分法的优点是鲁棒性较好，能够适应各种动态环境，其缺点是只能提取出边界点，不能提取出对象的完整区域。另外，当运动对象速度缓慢时，则可能检测不到，而运动速度较快时，将把部分背景也检测为运动对象。

2.3 常用检测方法及其性能分析

分析图像序列，对运动目标进行检测是图像处理系统的重要内容，通过运动检测，可以得到目标在视频图像中的位置、方向、大小等信息。如何有效快速地检测出运动目标，是其他后续目标分类和目标跟踪的基础，因此，运动检测对后续目标跟踪的正确完成具有极其重要的作用。本文对常用的相邻帧间差方法、基于梯度建模提取前景法和利

用帧间二阶差分(SODP)与边缘检测相结合进行运动目标分割方法进行了详细介绍，并利用相关视频图像对几种方法进行了性能分析比较。

2.3.1 相邻帧间差方法

相邻帧差法又称为图像序列差分法、帧间差法，当监控场景中出现运动目标时，帧与帧图像之间会出现比较明显的差别，两帧相减，得到两帧图像亮度差的绝对值，判断它是否大于阈值来分析视频或图像序列的运动特性，确定图像序列中有无运动的目标，对图像序列逐帧进行差分，相当于对图像序列进行时域上的高通滤波。当前有很多研究者在帧间差上进行了改进[2]，这里本文从最基本的相邻帧差法进行介绍和分析。

其表达公式如下：

$$G_{i,i-1}(x,y)=\left|I_i(x,y)-I_{i-1}(x,y)\right| \tag{2.6}$$

$$B_{i,i-1}(x,y)=\begin{cases}1, \text{ if } G_{i,i-1}(x,y)>th\\ 0, \text{ if } G_{i,i-1}(x,y)<th\end{cases} \tag{2.7}$$

式中：$G_{i,i-1}(x,y)$为像素点(x,y)处相邻帧的像素差值，$I_i(x,y)$和$I_{i-1}(x,y)$分别为第i帧和第$i-1$帧在像素点(x,y)处亮度分量，i表示帧数($i=1,2,...,N$)，N为序列总帧数，th为阈值。

这种方法的优点是算法简单，程序设计复杂度低，对光线和场景变化不太敏感，但是不能提取出目标对象的完整区域，只能提取出边界运动信息，运动目标的实体不能很好提取，容易出现空洞效应。

2.3.2 基于梯度的前景检测法

该方法[26]利用图像的梯度信息建立背景模型，首先利用 Canny 边缘检测从视频中提取出物体的边缘(边缘特征优于其他类型特征的两个方面：1.对突然的光照变化更具有鲁棒性；2.整个处理过程有效，因为边缘信息存储仅需要较少计算量的二进制形式，相对于 256 级像素灰度值)。

然后利用 N 帧图像建立基于梯度的背景模型，其中，输入帧的梯度信息利用 Canny 边缘检测来提取。边缘图像(边缘像素用 1 表示，非边缘像素用 0 表示)作为背景模型的输入。N帧图像中每个像素的N个最近值存储在 2-bin 直方图中，直方图构建过程如下式所示。

$$h(pixel(x,y)=r)=n_r \tag{2.8}$$

其中n_r表示像素值$r=\{0,1\}$的个数，$pixel(x,y)=0$代表非边缘像素，$pixel(x,y)=1$代表边缘像素，N是最近视频帧数，$n_0+n_1=N$，本实验中N取 50 帧，即用 50 帧建立背

景模型。如果像素 $pixel(x,y)$ 满足条件 $n_r < kN$，则该像素被认为前景像素，否则为背景像素。k 为预先定义的经验值。

经过上面提取的边缘图像有大部分非连通边缘像素和单像素噪声，直接在图像上利用形态学滤波在目标中会包含很多噪声，导致一个块中包含多个物体。在边缘区域分组前从噪声中分割出前景目标，把图像分割成40×40像素的若干小矩形区域，如果该区域像素密度高于某个阈值，认为该区域为前景区域。目标区域被标记后，在每个区域中的像素利用形态学闭操作进行连通，从而能较好地检测出运动目标。该前景运动目标检测方法的流程图如图 2.1 所示：

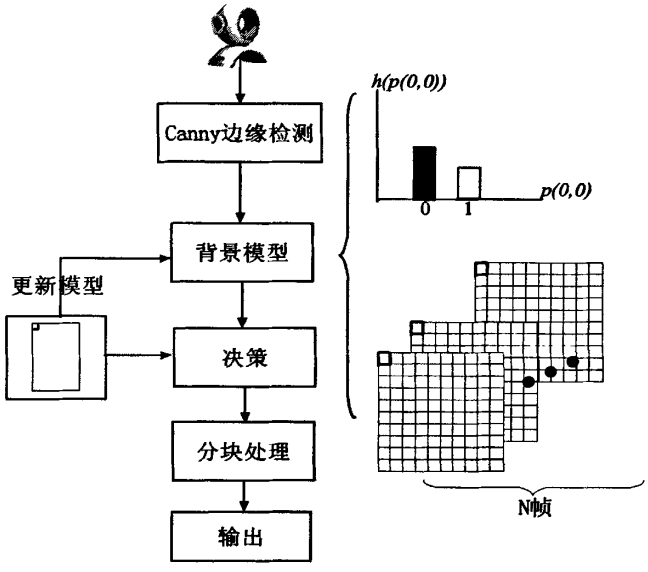


图 2.1　前景运动目标检测方法流程图

Fig.2.1　The flow chart of foreground object detection

该检测方法利用二进制边缘像素的 2-bin 直方图来模拟背景，检测效果比较理想，因为需要存储一定空间保存历史帧以保持前景物体存在，比普通背景模型用的存储空间稍大。

2.3.3　基于帧间二阶差分的前景检测法

该方法[27]利用序列图像中运动目标时空一致性，将帧间二阶差分(SODP)与边缘检测相结合进行运动目标分割，提高了运动目标检测的准确性。该方法把视频流看作在时间上有严格先后顺序、相邻帧间有强相关性的一组静态图像组成的图像序列，对其的分

割结果不仅要求帧内静止图像具有分割的合理性，而且还要保持帧间运动物体分割的延续性。结合运动检测图像和运动目标空间梯度图像进行分割，取得运动目标。

利用二阶差分图像(SODP)即拉普拉斯图像，改变原有差分图像方式，进行运动目标分割。设序列图像中相邻三帧图像分别为 $F_{i,j}(t_{n-1})$，$F_{i,j}(t_n)$，$F_{i,j}(t_{n+1})$，θ 为阈值，其二阶差分二值图像为

$$L_{i,j}(t_n)=\begin{cases}255, & |F_{i,j}(t_{n+1})-2F_{i,j}(t_n)+F_{i,j}(t_{n-1})|\geq\theta\\ 0, & |F_{i,j}(t_{n+1})-2F_{i,j}(t_n)+F_{i,j}(t_{n-1})|<\theta\end{cases} \tag{2.9}$$

由于分割利用了三帧图像信息，提高了检测的质量与精度，当采样帧率为 18 帧时，该方法仍可获得较满意运动图像时间梯度图像。在二阶差分处理中，因涉及到相邻三帧图像，会产生一帧时间滞。因系统采样间隔非常小，一帧滞后不会对实时处理产生较大影响. 该方法的缺点为阈值的确定不能自适应化，它随着视频的不同而变化，当目标与背景颜色相似或选择的阈值不合理时检测的目标不完整。

2.3.4 实验结果与分析

本文对以上三种方法用两个不同的视频图像序列进行了性能测试，由于室内环境中影子的干扰影响较大，检测的效果对比更有代表性，所以本文选取了一组室内带阴影的视频图像序列和室内多人环境的视频图像序列分别进行测试分析。

为了形象的说明三种方法的检测效果，图 2.2(a)中显示的检测图为未进行任何处理的原视频图像。从图 2.2(a)中，可以看出，对于室外环境下阴影的影响，相邻帧间差方法受影响不大，能很好地检测出运动目标边缘信息，算法简单并且能够实现实时的检测。基于梯度背景建模的方法对阴影也有一定的抑制作用，所以检测效果也较理想。但由于需要存储 N 帧图像，所以算法空间复杂度稍高，但计算量不太大；基于二阶差分的前景检测法，也能够清晰地检测出运动目标的轮廓，但该方法因涉及到相邻三帧图像，会产生一帧时间滞后，且由于当目标与背景颜色相似时，检测的目标不够完整。

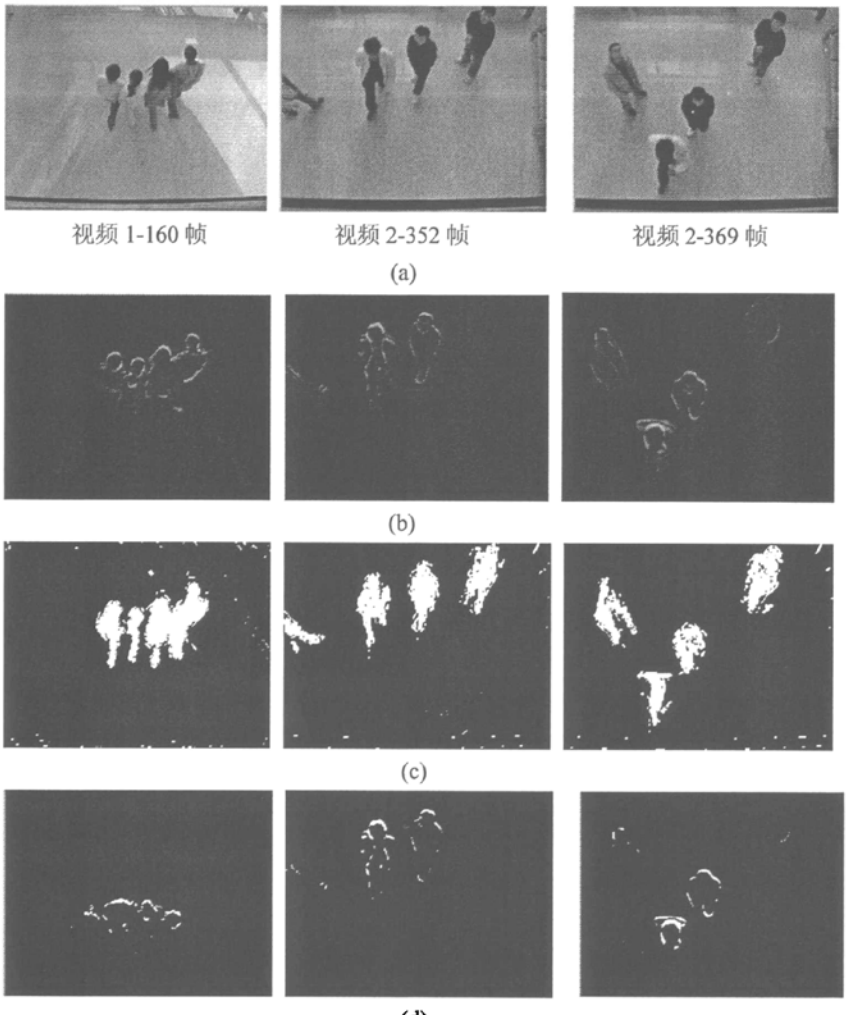


图 2.2　三种算法的检测效果；

(a)原始连续视频图像序列；(b)相邻帧间差方法；(c)基于梯度建模的方法；(d)基于帧间二阶差分的前景检测法

Fig 2.2　The detection results of the three algorithms.

(a)original video image sequence; (b)two-frame differencing detection results; (c)gradient-based model detection results; (d)SODP based detection results

以上三种前景检测算法虽然都能够针对不同情况，检测出运动目标。但是它们都存在一个共同要解决的问题，如何从前景图像中确定运动信息，提取出目标运动区域的参数(运动目标的位置等)。

经过大量的实验分析和总结，在帧间差分法和双向投影的基础上，本文采用了自适应阈值选取的方法[18]来确定前景运动区域。图 2.3 为利用实验室拍摄的室内和室外视频图像序列的运动区域检测效果图(蓝色框表示检测的运动区域)，分别显示了对单个目标和多个人体目标运动区域的检测效果。

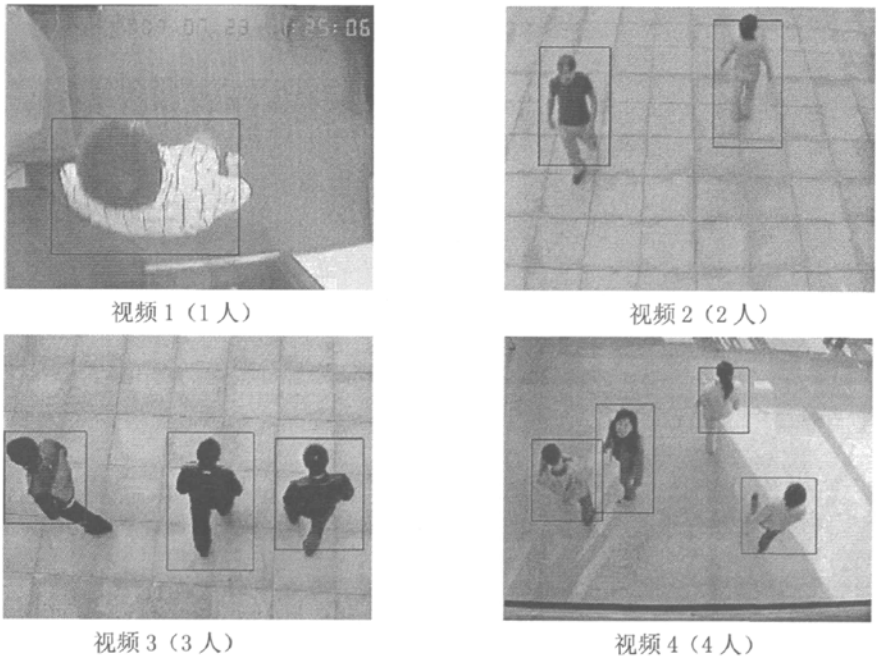


图 2.3　运动区域检测效果

Fig. 2.3　Moving Region Detection results

2.4　小结

运动目标检测是计算机智能视频监控的一个基础环节，是所有后续工作的基础，本章首先给出了几类运动目标检测方法的综述，然后详细的介绍了帧间差方法、基于梯度的前景检测方法和基于二阶差分的前景检测法，给出了相应的检测效果图，并对检测性能进行了分析。最后采用一种基于均值统计的自适应阈值方法来确定目标的运动区域，增强了算法的鲁棒性，比较好的完成了对视频图像中人体运动信息的提取。

3 多人体目标定位和分割

3.1 引言

头部检测(Head Detection)是指在输入图像中确定所有人体头部位置与大小。头部检测系统的输入是可能包含人体头部的图像，输出是关于图像中是否存在头部以及头部的数目、位置、尺度、姿态等信息的参数化描述。头部检测由于其在计算机视觉应用领域的重要性及相关工程项目中的广泛需求，头部检测技术已经普遍应用于人脸识别、头部跟踪、头部定位、表情识别、姿势估计等热点问题中[28]。它可以用于智能视觉研究领域，如乘客检测[29]，行人流量检测跟踪等[30]等。

本文针对固定垂直摄像头下的视频图像序列，由于头部运动信息边缘轮廓近似圆形(圆弧)的特征，因此本文头部检测定位问题也就转化为检测图像运动区域中圆的位置和大小问题。通常圆检测问题可以认为估计圆参数问题，而复杂背景会出现大量异常数据(outliers)，因此实时地检测圆在图像处理中是个挑战性的问题。

圆的识别方法主要可以分为 2 类：一类是基于 Hough 变换(Hough transformation，HT)的圆识别方法；另一类是基于圆的几何特征的识别方法。

第一类方法中，FCHT(Fast Circle Hough Transform)[31]把参数空间分解为几个较少维的几个子空间，它引入了几何特征和边缘的梯度向量来构造通过几何中心的直线。FHT(Fast Hough Transform)[32]用来定位几何中心，但在实时的环境中精确的求边缘梯度很困难，并且速度很慢不能满足实时检测要求。

第二类方法是搜索边缘图像中的像素点组成的链路，利用圆形的几何特征进行识别，文献[33]提出的 SRAS(stepwise recovery arc segmentation)算法将弧线分段，利用每段弧线的中垂线交点求取对应圆的圆心，由于存在误差，因此一般情况下得到的是一个范围，最后用迭代的方法进行优化，但是当圆形边缘有局部变形时，效果并不理想，甚至完全失效。

近年来，Freeman 链码因其能用较少数据来存储较多信息，而被广泛地应用到模式识别、图像表示等领域，Chan 等[34]提出了一种基于 Freeman 准则的直线检测算法，该算法通过跟踪线段子元，并根据两相邻线段子元间的偏转角度判断是否需要进行子线段合并。在通常的图像检测方法中，把边缘特征看作孤立的点，而链码把特征点当作轮廓(一个像素宽的连续线)，即利用链码可以把特征点看作多个不相连的特征轮廓。

在上一章，已经确定出了运动人体的运动区域，但是并未确定每个运动区域内人的具体数目以及每个人的位置，这是目标定位要解决的问题。本文由于在垂直摄像头下，

头部运动信息轮廓具有近圆形的特征，利用 Freeman 链码和 RANSAC 算法相结合的方法[35]来估计图像中圆的位置和大小，首先利用链码检测图像中的轮廓，然后对每个独立的轮廓，通过 RANSAC 算法估计候选圆的参数，也即求出图像中的圆，进而确定人体位置。下面对进行轮廓特征提取的 Freeman 链码和 RANSAC 算法给予详细论述。

3.2 利用 Freeman 链码提取轮廓

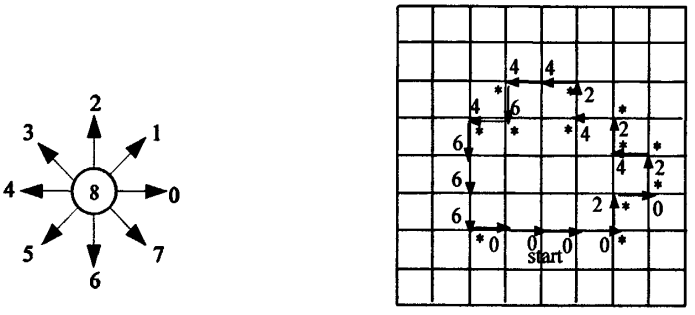
大多数已经存在的方法，例如[31]，把边缘图像的所有特征点看作一个整体，这样不得不处理大量的特征点集，因此需要大量的内存，同时计算量也很大。本文把被检测物体的特征作为连通约束，它强调目标的边缘点应该是连续的。根据该约束，大量的特征点集可以分为若干小的子集(轮廓)。因此检测圆的问题被简化为在每个轮廓内检测目标圆的问题。

Freeman 链码[36]指相邻两像素连线的八种可能方向值 $c_i \in \{0,1,2,3,4,5,6,7\}$，即链码 c 是 $\{0,1,2,3,4,5,6,7\}$ 元素的有限集，链码中的第 i 个码字由 c_i 标示，属于第 i 个码值的向量 $\mathrm{V_i}$ 由下式给出，也就是说每个链码值 c_i 都有一个向量 V_i 与之对应：

$$\mathrm{V}_3=(-1,1) \quad \mathrm{V}_2=(0,1) \quad \mathrm{V}_1=(1,1)$$
$$\mathrm{V}_4=(-1,0) \quad \mathrm{V}_8=(0,0) \quad \mathrm{V}_0=(1,0)$$
$$\mathrm{V}_5=(-1,-1) \quad \mathrm{V}_6=(0,-1) \quad \mathrm{V}_7=(1,-1)$$

如图 3.1，链码用像素点八邻域绝对方向进行描述，其中图 3.1(b)中的每个方向数字代表当前像素相对上一个邻域像素(即父像素)的位置变化。数字化的二值离散曲线(有限的八连通点集)可以用链码来表示 $\{a_i\}^n$，n 表示曲线上点集的个数。

为了进行有效的轮廓提取，利用 Freeman 链码检测轮廓之前，先利用 Canny 算子进行边缘检测。利用 Freeman 链码寻找目标轮廓的过程为，首先将检测过程中遇到的第一个点作为一条链码的起点(例如图 3.1(b)中的 start 点)，然后顺序扫描该点周围的八个点有没有边缘点，每遇到一个边缘点，链码的长度增加一，同时将该点设置为非边缘点以免被重复检测，然后继续扫描该点周围的八个点，如此循环形成链码，直到某个点周围没有边缘点为止。利用该链码求解过程，可以得出如图 3.1(b)链码方向数字，即 00202424246466600。在对运动区域内目标边缘进行处理时，为了降低拟合模型时的需要的特征点数，利用轮廓的拐点(方向数字相对于上个方向数字变化的点)作为拟合模型的特征点，其中图 3.1(b)中标示为“*”的像素点为轮廓的拐点。这样大大降低了模型的采样点，有效的提高了采样速度，同时节省了存储空间的大小。为了进一步降低采样点数，因为水平和垂直方向的拐点较少，特征点数下一步计算贡献很小，这部分拐点在进行 RANSAC 算法之前滤去不做计算。



(a) (b)

图 3.1 链码描述；(a)绝对方向；(b)链码描述，“*”表示拐点

Fig 3.1 Chain code representation;

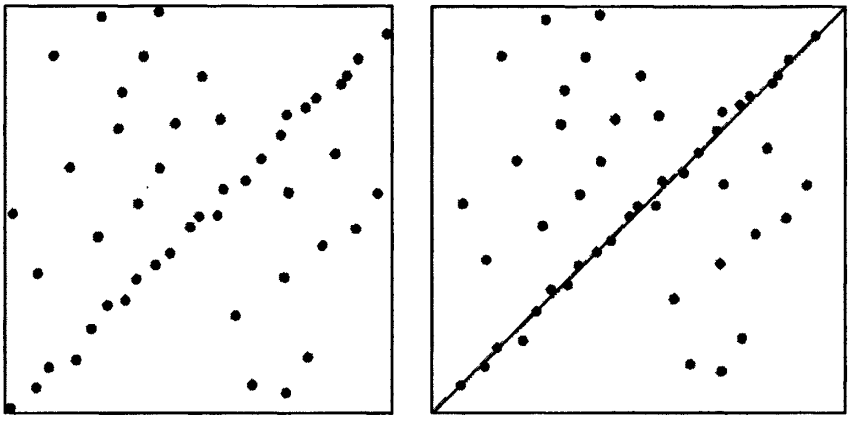
(a) Absolute direction; (b) Chain code representation, '*'denotes the inflexion

3.3 RANSAC 算法

3.3.1 RANSAC 基本思想

由 Fishler 和 Bolles[37]提出的 Random Sample Consensus(RANSAC)算法，对错误率超过 50%的数据仍然能够处理，是最有效的 Robust 估计算法之一，在计算机视觉领域得到了广泛的应用：例如基础矩阵估计[38]、特征匹配[39]，运动模型选择[40]等，同时，根据随机抽样的思想，衍生了很多鲁棒算法。

在模型参数估计中，为了消除异常数据(outliers)对估计的影响，最直观的想法就是从原始数据找出一组只包含正常数据(inliers)的数据抽样来进行参数估计，要找到这样的数据抽样，需要搜索原始数据所有可能的组合，这样计算量会很庞大。RANSAC 算法：认为只需要搜索 M 组抽样，只要 M 足够大，就可以在一定置信概率下保证这 M 组抽样中至少有一组抽样不包含异常数据；然后用这 M 组抽样分别估计模型参数，根据一定的评选标准，找出最优模型参数，则可以认为最优模型参数对应的抽样数据不包含异常数据；最后，利用找出的最优模型参数作为假设模型，根据一定规则对其他数据进行筛选，找出和该模型参数符合的所有数据，并用这些数据估计最终模型参数。图 3.2 为用 RANSAC 算法拟合直线的结果，图 3.2(b)中浅蓝色直线为拟合后的结果，直线周围附近的蓝色点为正常数据，离直线较远的红色点为异常数据。



(a) (b)

图 3.2 拟合直线图：(a)带有异常数据的直线；(b)拟合后的直线（浅蓝色）

Fig. 3.2 The map of fitted line; (a) Line with outliers; (b) Fitted line (light blue)

3.3.2 RANSAC 算法步骤

(1)根据置信概率 P 和数据错误率 ε，由(3.1)式计算需要选择的最小抽样数量 M

$$1-(1-(1-\varepsilon)^m)^M = P \tag{3.1}$$

其中，ε 为数据错误率(异常数据在总的原始数据中所占的比例)，m 为拟合该模型参数需要的最小数据量，P 为置信概率。

(2)随机抽取原始数据组成一个抽样，每组抽样的样本数为估计模型参数需要的最小数据量，计算该抽样对应的模型参数。

(3)用所有原始数据来检验模型参数(称为全数据检验)，获得支持每个模型参数的正常数据数量；重复第 2，3 步，直到完成 M 组抽样的处理。

(4)根据正常数据数量和误差的方差选择最优模型参数。

(5)找出最优模型参数对应的所有正常数据，并用这些数据计算最终的模型参数。

3.3.3 估计圆参数的 RANSAC 算法步骤

RANSAC 算法是一种在很多外部噪声点存在的情况下，进行鲁棒性拟合圆参数的算法，算法的步骤如下：

(1)从连续的轮廓线 m 个数据点上随机采样三个不在同一条直线的点。

(2)利用三点估计候选圆的参数(半径坐标和圆心坐标)，在给定的容许误差(tolerance)下，从 m 个数据中找到轮廓上匹配该模型参数的点个数 K。

(3)如果K足够大，认为该次抽样计算的圆参数有效，然后对轮廓上适合该模型的所有点进行最小二乘法(least-squares)配准，找到拟合该圆参数的最佳配准，比较该配准与通过三点计算的圆参数距离，当满足一定阈值条件时，认为该圆参数找到，成功跳出。否则重复 1-3 步。

(4)如果重复的次数大于最大迭代次数L，失败退出。

利用 RANSAC 算法在整个图像的边缘轮廓中拟合圆的例子如图 3.3 所示。图中红色数据点为异常数据(outliers，可以认为图像中的噪声和干扰)，蓝色数据点为正常数据(inliers，对拟合圆参数有贡献的数据点)，绿色圆为利用随机抽样三点拟合计算的圆，粉红色圆为利用 RANSAC 算法迭代后的最终圆参数，该圆上面和最靠近该圆附近处含有最多的正常数据。由此可以看出，利用 RANSAC 算法在图像轮廓中识别圆，可以得到较好的检测结果，较 Hough 变换检测圆效率要高，因为 Hough 变换需要对图像中所有边缘轮廓点计算圆参数，进而利用欧式距离判断这些圆参数之间的距离决定最终圆参数。所以，利用 RANSAC 算法估计圆参数时，如果需要迭代次数较少，检测效率会得到很大程度提高，避免了大量无效采样带来的时间消耗问题。

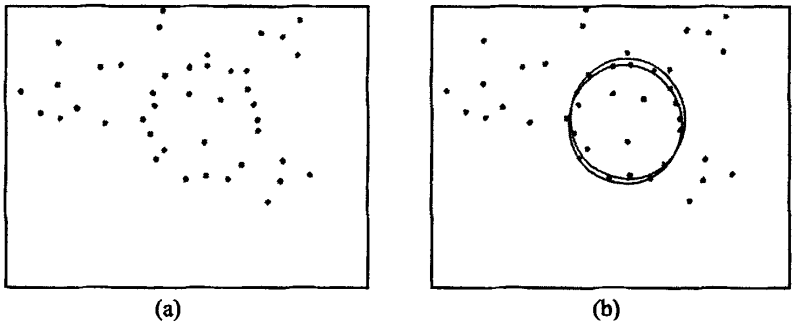


图 3.3　拟合圆图：(a)带有异常数据的圆；(b)拟合后的圆（粉红色）

Fig 3.3　The map of fitted circle; (a) Circle with outliers; (b) Fitted circle (pink)

假定在候选轮廓上有足够的圆特征点，则圆模型拟合失败的概率p_{fail}(没有找到合适的模型退出该算法的概率)由(3.2)式给出

$$p_{fail}=(1-p_g^n)^L \tag{3.2}$$

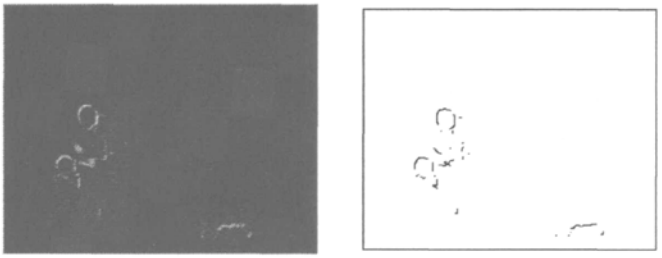
其中p_g是随机选择的数据点能够拟合一个较好模型的概率。

最大迭代次数L为

$$L=\frac{\log(p_{fail})}{\log(1-p_g^n)} \tag{3.3}$$

3.4 本文基于 Freeman 链码和 RANSAC 的圆检测算法

为了提高检测的速度和消除身体其它运动信息骨架的影响，缩小检测圆的空间开销，在前一章中，我们已经检测出目标的运动区域，需要进一步对运动区域进行预处理和消除图像中肩膀和衣服下摆对检测可能造成的影响，本文采用[18]方法进行预处理和消除干扰，有效地降低了无效数据点数，图像预处理后的结果如图 3.3 所示。



(a) (b)

图 3.3 预处理前后图像比较：（a）处理前帧间差灰度图像；(b)处理后图像

Fig. 3.3 Preprocessing result comparison

(a) Gray differencing image (b) Image after processing

若在当前帧图像中分割出 M 个矩形运动区域，记为 $W=(W_1,W_2,..W_M)$ 区域窗口序列，本文利用 Freeman 链码和 RANSAC 算法相结合拟合圆参数(检测圆)的具体算法步骤如下：

(1)采用 8 领域搜索方法，对区域 W_1 采用 Freeman 链码检测出轮廓曲线，并且标记各个曲线的位置，统计各个曲线的长度(链码长度)，保留满足条件的轮廓曲线，加入线段集 $D=(L_1,L_2,...,L_n)$。

(2)选取轮廓曲线段 $L_i(1\le i\le n)$，随机采样不在同一直线上的 d_1,d_2,d_3 三点，估计圆参数(半径和圆心坐标)，在给定的阈值(容许误差)下，在轮廓曲线段内找到适合该模型的参数个数，并把该个数表示为 K。

(3)如果该个数 K 足够大，认为该参数确定的圆为候选圆，然后对适合该配准的图像上轮廓点做最小二乘法配准，找到另一个合适的圆参数。在该参数与 RANSAC 计算的圆参数满足一定阈值时，认为找到了合适的圆参数，若 i=1，该参数为第一个圆参数，加入 C 中；若 i>=2 判断该圆参数 C_i 与其他圆参数 $C=(C_1,C_2,..,C_N)(N\le i)$ 的相似度。若与其它圆相似，C_i 不为新出现的圆参数，若不相似 C_i 为新圆，加入 C，重复(2)、(3)进

行下一曲线段 L_{i+1} 检测。若进行最小二乘法不能使拟合的圆参数与从 $L_i(1 \le i \le n)$ 计算的圆参数配准，则仍从 $L_i(1 \le i \le n)$ 开始，重复(2)、(3)。

(4)若 i=n ,则当前区域 R_1 内圆检测结束，同时保存单元集 C。对当前图像中运动区域窗口序列 W 中所有其它区域的边缘图像重复(1)、(2)、(3)步操作进行处理。

(5)若当前图像中所有投影区域检测结束，输出单元集 C 中所有的圆参数，同时进行下一帧图像的处理，重复(1)-(5)操作。

3.5 实验结果与分析

本文对不同光照和场合下的图像序列进行了测试，结果表明该方法在人体遮挡较严重的情况下，若目标的头部没有被其他运动目标长时间的遮挡，可以准确的对头部进行检测识别。

3.5.1 检测结果

在实验中，视频图像序列中各个参数取值分别为 $K=30$，$L=22$，$p_g=70\%$，$p_{fail}=1\%$，图 3.4 为四组视频图像序列的头部圆检测定位结果，蓝框为初步检测到的运动区域，红色的圆代表在蓝框运动区域内检测到的头部位置和大小。

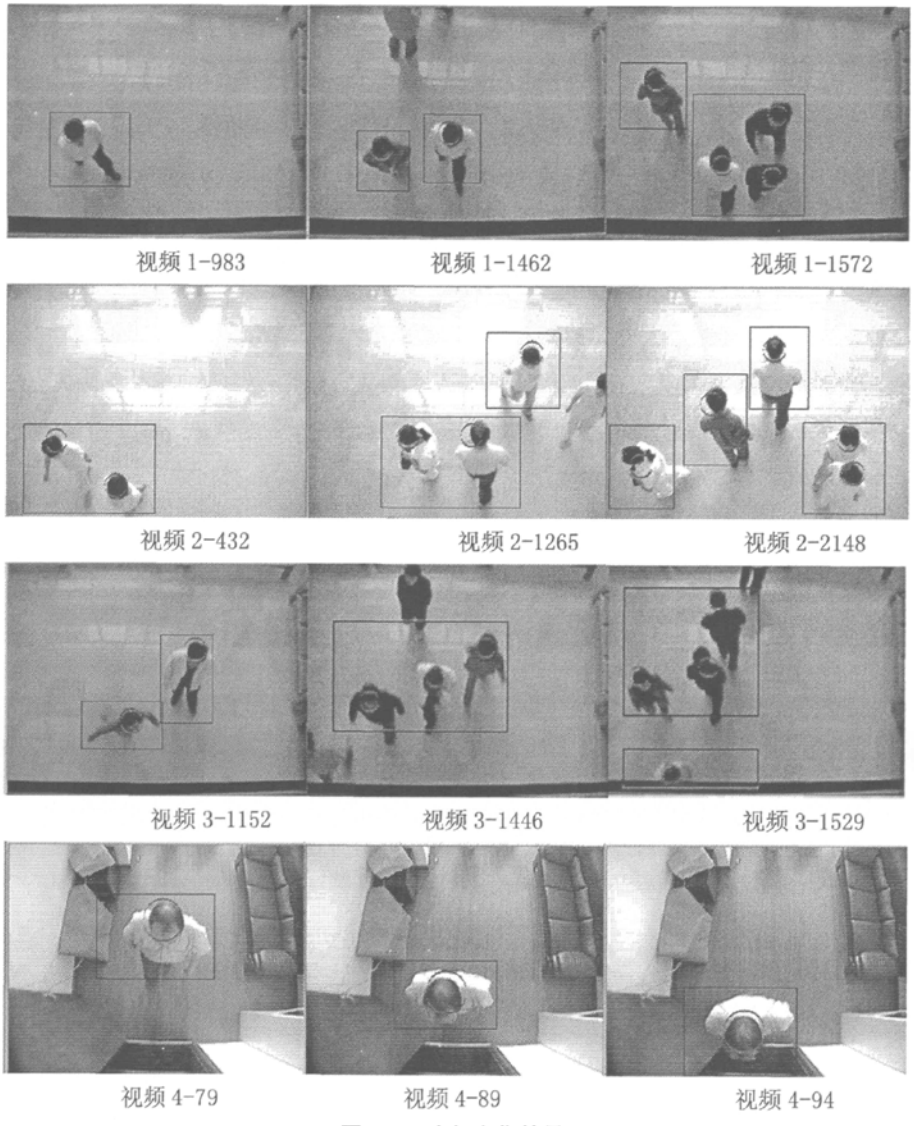


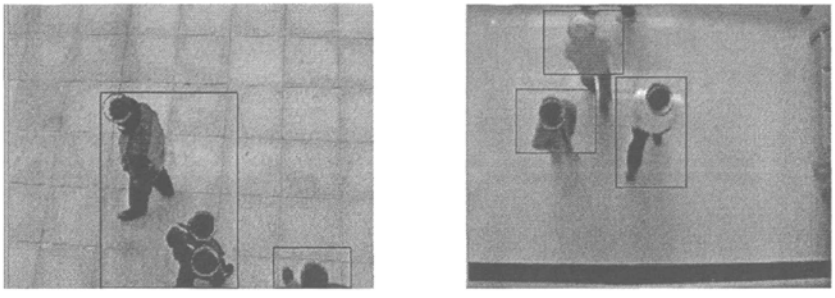
图 3.4　头部定位结果

Fig. 3.4　The results of head localization

3.5.2　实验分析

本文在 PC 机（2.4G，256M）上进行了实验，对不同视频图像(10000 帧)进行了测试，结果表明该算法对身体运动边缘的干扰有一定的抑制作用，具有较高的检测速度和正确率。当人体的头部尺寸相对较小时，检测和定位的速度能达到每秒 18 帧，可以达

到实时的效果。由于 RANSAC 算法本身参数较多，现在参数基本依据大量实验来定，参数不能自适应化，这是将来研究的一个方向。



(a) (b)

图 3.5 误检情况示意图

Fig. 3.5 Illustration of error detection

当人体其他部分的运动边缘轮廓较清晰且近似圆形时，如肩膀与头部轮廓具有很大的相似性时，人体的肩部被误判为头部，会造成了误检如图 3.5(a)所示，当头部颜色信息与周围背景较相似时，帧间差会失去头部轮廓，造成漏检，如图 3.5(b)。为了减小这种误检率的发生，可以对图像采取别的处理办法消除干扰，同时加入别的特征提取目标头部，这也是本文今后的努力方向。

3.6 小结

运动目标的定位与分割是计算机智能视频监控的一个基础环节，也是计算机视觉的一个主要部分和研究热点。本章在对 Freeman 链码和 RANSAC 算法理论知识研究的基础上，针对每个运动区域内人体头部运动信息的边缘轮廓具有近圆形(圆弧)的关键特征，采用了融合链码和 RANSAC 算法的一种头部检测定位方法，该算法能较快速的对多个人体头部进行定位，从而较好地定位出多人体目标，达到了比较好的检测效果。

4 Kalman滤波器和Mean Shift相关理论及其在跟踪中的应用

4.1 引言

运动目标跟踪是近期视觉领域内一个备受关注的课题，它是在视频图像序列的每一帧图像中确定出我们感兴趣的运动目标的位置，来实现目标的跟踪。在机器视觉研究领域里，随着技术不断发展，自动目标跟踪越来越受到研究者的重视，具有广阔的应用前景。运动目标的跟踪在医学研究、视频监控、交通流量观测监控等很多领域都有重要的实用价值。跟踪的难点在于如何快速而准确的在每一帧图像中实现目标定位。

目标跟踪技术在二战之前就已经在军事上得到应用，1937年世界上出现第一部跟踪雷达站 SCR-28。之后，随着科技的进步，各种跟踪系统的相继出现与不断完善，跟踪理论和方法在各国学者的努力下也获得了很大的发展。1955年 Wax 提出了多目标跟踪的基本概念，目标跟踪研究有了一定的发展。但直到 70 年代，卡尔曼滤波理论[41]被成功地应用在目标跟踪领域[42]之后，目标跟踪技术才真正引起人们的普遍关注和极大兴趣。近二十年来，研究学者在此领域进行了大量的研究，同时随着其它新技术的出现，比如扩展卡尔曼滤波(Extended Kalman Filter)、粒子滤波(Particle Filter)、均值迁移(Mean Shift)、多模型、多速率处理等技术，结合这些技术提出了许多方法，取得了长足的进步。

就跟踪的方法而言，目标的跟踪主要包括基于运动(Motion-based)和基于模型(Model-based)这两种方法。美国卡内基一梅隆大学研究开发的“实时视频中动目标识别分类与跟踪系统”(Moving Target Classification and Tracking from Real-time Video)就采用了基于模型和运动相结合的方法对视频中的目标实时监控与跟踪，并识别两种目标人和汽车[43]。

本章的研究内容是典型视频监控系统中的多运动目标跟踪，是建立在第三章的多人体目标定位基础上完成的，首先对描述运动目标的纹理特征进行了介绍，然后重点对 Mean Shift 理论及其在跟踪中的应用进行了论述，最后给出了一种融合目标颜色、纹理和运动信息的改进的 Kalman 和 Mean Shift 跟踪方法。

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

4.2 纹理特征

局部二进制模式 LBP(Local Binary Pattern)是近年来提出的一种基于灰度图像有效的纹理描述方法，它是通过比较图像中每个像素与其邻域内像素灰度值的大小，并利用二进制模式表示的比较结果来描述图像的纹理。近年来 LBP 纹理在人脸识别、表情识别以及背景建模等计算机视觉应用领域中表现出良好的性能，但在运动目标跟踪方面的应用刚刚起步，Quang 等[44]提出了利用图像灰度值和 LBP 对单色热度视频进行跟踪的算法，对单色的视频图像序列达到较好跟踪效果。王永忠等[45]提出基于 LBP 纹理特征的红外成像目标跟踪方法，将 LBP 纹理特征集成到核跟踪方法中，构造目标及候选目标的特征模型，利用 Mean Shift 方法实现基于纹理特征的红外成像目标的跟踪。宁纪锋和吴成柯提出用 $LBP_{8,1}^{ri}$ 纹理模型中表示边界和角的 5 种基本模式表示目标，并将之成功嵌入 Mean Shift 算法进行目标跟踪。纹理特征不仅能充分表达图像的纹理，而且算法具有复杂度低、计算速度快和一定程度上不受光照变化的影响等良好特性，越来越受到目标跟踪研究领域研究者的重视。

4.2.1 基本局部二进制模式

LBP 算子是由 Ojala 等人于 1996 年提出的[47]，最初被称为基本的 LBP 算子。对图像中的每个像素，通过计算以其为中心的 3*3 邻域内各像素和中心像素的大小关系，把像素的灰度值转化为一个八位二进制序列。对 i 处灰度值为 a_i 的像素，它的 LBP 值为

$$LBP(x_i)=\sum_{j=1}^{8} g(a_j-a_i)\times 2^{(j-1)} \tag{4.1}$$

a_j 为 x_i 八邻域像素集中索引为 j ($j=\{1,2,...,8\}$)的像素灰度值，$g(a_j-a_i)$ 是一个函数定义为

$$g(a_j-a_i)=\begin{cases}1 & if \quad a_j-a_i\geq 0\\ 0 & otherwise\end{cases} \tag{4.2}$$

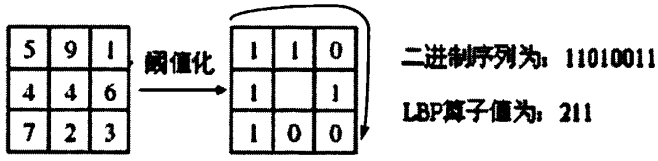


图 4.1 基本 LBP 算子

Fig. 4.1 Basic LBP operator

具体计算过程如图 4.1 所示，对于图像的任意一点 I_c，其 LBP 特征计算为，以 I_c 为中心，取与 I_c 相邻的 8 个点，按照顺时针的方向记为 I_0，I_1，…，I_7；以 I_c 点的象素值为阈值，如果 I_i 点的象素值小于 I_c ，则 I_i 被二值化为 0，否则为 1；将序列看成一个 8 位二进制数，将该二进制数转化为十进制就可得到 I_c 点处 LBP 算子的值。

4.2.2 扩展局部二进制模模式

基本的 LBP 算子只局限在 3*3 的邻域内，对于较大图像大尺度的结构不能很好的提取需要的纹理特征，因此研究者们对 LBP 算子进行了扩展[48]。新的 LBP 算子 $LBP_{P,R}$ 可以计算不同半径邻域大小和不同像素点数的特征值，其中 P 表示周围像素点个数，R 表示邻域半径，同时把原来的方形邻域扩展到了圆形，图 4.2.给出了四种扩展后的 LBP 例子。其中，R 可以是小数，对于没有落到整数位置的点，根据轨道内离其最近的两个整数位置的像素灰度值，利用双线性插值的方法计算它的灰度值。

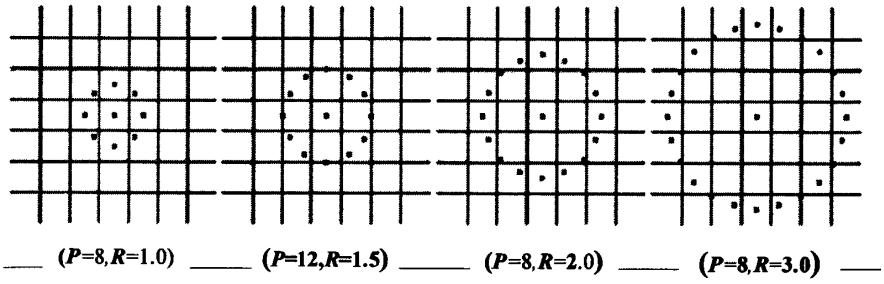


图 4.2 扩展的 LBP 图例

Fig. 4.2 Extended LBP operators

$LBP_{P,R}$有2^P个值，也就是说图像共有2^P种二进制模式，然而实际研究中发现，所有模式表达信息的重要程度是不同的，统计研究表明，一幅图像中少数模式特别集中，达到总模式的百分之九十左右的比例，Ojala等人定义这种模式为Uniform模式，如果一个二进制序列看成一个圈时，0-1以及1-0的变化出现的次数总和不超过两次，那么这个序列就是Uniform模式(如图4.3所示)，比如，00000000、00011110、11111111。在使用LBP表达图像纹理时，通常只关心Uniform模式，而将所有其他的模式归到同一类中。

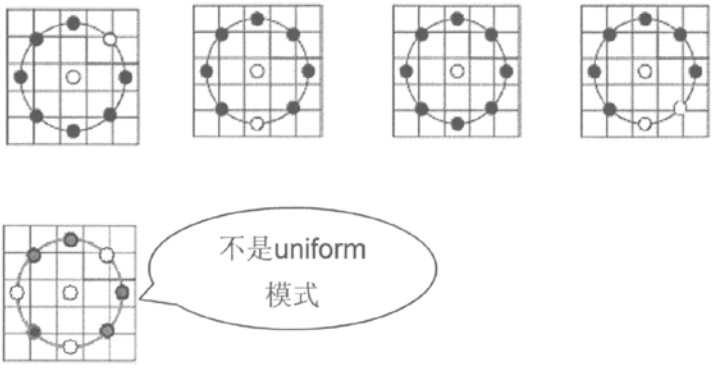


图 4.3　Uniform LBP 模式

Fig. 4.3　Uniform LBP patterns

在表情识别中，最常用的是把 LBP 的统计柱状图作为表情图像的特征向量。为了考虑特征的位置信息，把图像分成若干个小区域，在每个小区域里进行直方图统计，即统计该区域内属于某一模式的数量，最后再把所有区域的直方图依次连接到一起作为特征向量接受下一级的处理[49]，如图 4.4 所示。

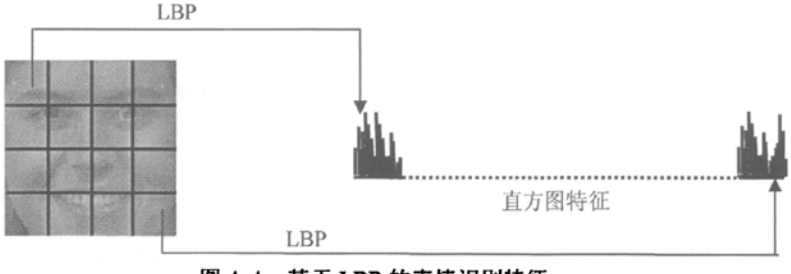


图 4.4　基于 LBP 的表情识别特征

Fig. 4.4　Facial feature based on LBP for expression recognition

LBP 算子利用了周围点与该点的关系对该点进行量化。量化后可以有效地消除光照对图像的影响，只要光照的变化不足以改变两个点象素值之间的大小关系，那么 LBP 算子的值不会发生变化，所以一定程度上，基于 LBP 的识别算法解决了光照变化的问题，但是当图像光照变化不均匀时，各像素间的大小关系被破坏，对应的 LBP 模式也就发生了变化。

4.3 Kalman 滤波器

卡尔曼滤波器是一种在时域内采用递归滤波的方法对系统状态进行最小均方误差估计的方法，具有计算量小，可实时处理的特点，利用卡尔曼滤波器实现对目标轨迹的估计和预测是非常有效的。下面给出 Kalman 滤波器模型及其在本文中参数设置情况。

4.3.1 Kalman 滤波器建模

场景中的目标在每一帧图像中的位置构成了目标的运动轨迹，卡尔曼滤波器引入的目的就是根据以往目标的位置信息预测当前帧中目标的可能位置。因此卡尔曼滤波器中的状态变量和观测值均为目标的位置信息，更准确地说是被跟踪目标中心坐标的相关信息。

在第三章中，我们检测到了目标的头部位置，根据它可以定位运动人体的区域，我们用矩形框描述运动目标，其参数分别为中心坐标(x_c, y_c)，宽度为 $width$，高度为 $height$，在跟踪的过程，假定人在场景中的运动是线性运动，因此我们把系统看作一个线性动态系统，我们只需要利用最基本的 Kalman 滤波器来描述运动模型。

卡尔曼滤波器算法通过状态方程和观测方程来描述一个动态系统。设线性系统的状态方程和观测方程分别为式(4.3)和式(4.4)

$$\boldsymbol{x}_k = \boldsymbol{A} \cdot \boldsymbol{x}_{k-1} + \boldsymbol{B} \cdot \boldsymbol{u}_k + \boldsymbol{w}_k \tag{4.3}$$

$$\boldsymbol{z}_k = \boldsymbol{H} \cdot \boldsymbol{x}_k + \boldsymbol{v}_k \tag{4.4}$$

这里，$\boldsymbol{x}_k$ 是 t_k 时刻的 $n\times1$ 维的状态向量，$\boldsymbol{z}_k$ 是 t_k 时刻的 $m\times1$ 维的观测向量，$\boldsymbol{A}$ 是 t_{k-1} 时刻到 t_k 时刻的状态转移矩阵，为 $n\times n$ 维，$\boldsymbol{B}$ 为系统控制矩阵，$\boldsymbol{u}_k$ 为系统的控制量，本系统没有系统控制量，故二者均为 0，故式(4.3)可变为：

$$\boldsymbol{x}_k = \boldsymbol{A} \cdot \boldsymbol{x}_{k-1} + \boldsymbol{w}_k \tag{4.5}$$

$\boldsymbol{H}$ 是系统的观测矩阵，为 $m\times n$，$\boldsymbol{w}_k$ 是 t_k 时刻状态的随机干扰噪声向量，为 $n\times1$ 维，$\boldsymbol{v}_k$ 是 t_k 时刻的观测噪声向量，为 $m\times1$ 维，$\boldsymbol{w}_k$ 和 $\boldsymbol{v}_k$ 通常假设为互相独立的零均值的高斯白噪声向量，它们的协方差矩阵分别为：

$$E[w_k w_i^T]=\begin{cases}Q_k & i=k\\ 0 & i\neq k\end{cases} \tag{4.6}$$

$$E[v_k v_i^T]=\begin{cases}R_k & i=k\\ 0 & i\neq k\end{cases} \tag{4.7}$$

$$E[w_k v_i^T]=0 \quad \text{对所有的} k \text{和} i \tag{4.8}$$

4.3.2 Kalman 滤波器各参数设置

在本文系统中，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，由于相邻两帧之间的间隔较短，目标的运动状态变化较小，所以假设目标在单位时间内为匀速运动。

定义 Kalman 滤波器系统状态 x_k 是一个四维向量 $x_k=[x(k),y(k),x'(k),y'(k)]^T$，其中，$x(k)$ 和 $y(k)$ 分别是目标中心在 x,y 轴上的坐标分量，$x'(k)$ 和 $y'(k)$ 分别是目标在 x,y 轴方向上的速度。观测向量为 $z_k=[x_c(k),y_c(k)]^T$，其中 $x_c(k)$、$y_c(k)$ 分别表示当前帧中观测到的目标中心在 x,y 轴上的坐标信息。

由于假设目标是在单位时间间隔内作匀速直线运动，所以状态转移矩阵 A 定义为：

$$A=\begin{bmatrix}1,0,\Delta T,0\\ 0,1,0,\Delta T\\ 0,0,1,0\\ 0,0,0,1\end{bmatrix}$$

其中：$\Delta T=T_k-T_{k-1}$，即前后两帧时间差。

由系统状态和观测状态的关系可知，观测矩阵 H 设置为

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

4.4 Mean Shift 理论及其在跟踪中的应用

Mean Shift 的框架最早是由 FuKunaga 和 Hostetler[50]于 1975 年提出的，是一种无参算法，它沿着概率密度梯度的上升方向，寻找分布的峰值。1995 年 Cheng. Y 用它来解决聚类分析问题[51]。直到 1998 年，Bradsk 将 Mean Shift 算法用于人脸的跟踪才使得它的优势在目标跟踪领域体现出来。它利用梯度优化方法来减少特征搜索匹配的时间，实现快速的目标定位。Isard 和 Blake 于 1998 年，提出将基于颜色分布的 Mean Shift 算法和粒子滤波器应用于目标跟踪也取得了非常好的效果[52]。Dorin Comaniciu 于 1999 年发

表了“Mean shift Analysis and Application”[53]。之后连续发表了多篇文章使 Mean Shift 算法在目标跟踪领域得到了更为广泛的应用。

Comaniciu 等在文章中证明了 Mean Shift 算法在满足一定条件下，一定可以收敛到最近的一个概率密度函数的稳态点，因此 Mean Shift 算法可以用来监测概率密度函数中存在的模态，并在文献[54]中，Comaniciu 等主要讨论了 Mean Shift 在跟踪中的应用。在本章后面的部分，将详细介绍 Mean Shift 理论及其在跟踪中的应用，并给出了一种融合目标颜色信息和纹理信息到 Mean Shift 算法模型中，同时结合 Kalman 滤波器运动信息的跟踪算法。

4.4.1 Mean Shift 搜索法

定义 d 维空间 R^d 上 n 个样本点 x_i，$i=1,...,n$，$K(x)$ 表示多元核函数，窗口的半径为 h，则在点 x 上的密度估计可表示为：

$$\hat{f}(x)=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right) \tag{4.9}$$

通过最小化该估计值与真实密度值得平均全局误差，得到核密度估计函数 Epanechnikov 如下：

$$K_{E(x)}=\begin{cases}\frac{1}{2}c_d^{-1}(d+2)(1-\|x\|^2) & if\ \|x\|<1\\ 0 & otherwise\end{cases} \tag{4.10}$$

其中 c_d 是 d 维超球的体积。

此外，另有一个通用的核密度估计函数是正态多元高斯函数：

$$K_{N(x)}=(2\pi)^{-d/2}\exp(-\frac{1}{2}\|x\|^2) \tag{4.11}$$

这里定义 K 的一个剖面函数 $k:[0,\infty)\to R$，使得：

$$K(x)=k\left(\|x\|^2\right) \tag{4.12}$$

则公式(4.10)的 Epanechnikov 函数可写为：

$$K_{E(x)}=\begin{cases}\frac{1}{2}c_d^{-1}(d+2)(1-x) & if\ \|x\|<1\\ 0 & otherwise\end{cases} \tag{4.13}$$

则公式(4.11)的正态多元高斯函数可写为：

$$K_{N(x)}=(2\pi)^{-d/2}\exp(-\frac{1}{2}x^2) \tag{4.14}$$

同样，在点 x 上的密度估计(4.9)式可以写为

$$\hat{f}(x)=\frac{1}{nh^d}\sum_{i=1}^{n}k\left\|\frac{x-x_i}{h}\right\|^2 \tag{4.15}$$

我们定义： $g(x)=-k'(x)$ (4.16)

假设除了有限点 k 对于所有的 $x:[0,\infty)$ 存在导数，那么核函数 G 可以定义为如下：

$$G(x)=cg(\|x\|^2) \tag{4.17}$$

其中 C 为标准化常量，在 x 点的密度估计根据(4.9)式计算得到：

$$\hat{f}(x)=\frac{C}{nh^d}\sum_{i=1}^{n}g\left\|\frac{x-x_i}{h}\right\|^2 \tag{4.18}$$

把密度估计的梯度作为密度梯度的估计，可以得到：

$$\begin{aligned}\hat{\nabla}f_{K(x)}=\nabla\hat{f}_{K(x)}&=\frac{2}{nh^{d+2}}\sum_{i=1}^{n}(x-x_i)k'\left(\left\|\frac{x-x_i}{h}\right\|^2\right)\\&=\frac{2}{nh^{d+2}}\sum_{i=1}^{n}(x_i-x)g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)\\&=\frac{2}{nh^{d+2}}\left[\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n}x_ig\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}-x\right]\end{aligned} \tag{4.19}$$

其中假定 $\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$ 为非零值。

上式中最后的括号中包含了均值平移矢量：

$$M_{h,G(x)}=\frac{\sum_{i=1}^{n}x_ig\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}-x=m_{h,G(x)}-x \tag{4.20}$$

其中 $m_{h,G(x)}$ 为采样均值。

根据(4.18)和(4.20)式可以简化(4.19)式为：

$$\hat{\nabla}f_{K(x)}=\hat{\nabla}f_{G(x)}\frac{2/C}{h^2}M_{h,G(x)} \tag{4.21}$$

$$M_{h,G(x)}=\frac{h^2\times\hat{\nabla}f_{K(x)}}{2/C\times\hat{\nabla}f_{G(x)}} \tag{4.22}$$

上式表明核函数G中包含的 Mean Shift 向量是根据核函数K计算得到的归一化的密度梯度估计值。

所谓的 Mean Shift 算法就是连续不断地向采样均值移动位置。均值平移过程可以定义为计算均值向量$M_{h,G(x)}$和根据$M_{h,G(x)}$来变化核函数G的中心的过程。假设$\{y_j\}_{j=1,2,...}$表示核函数G中心位置的连续位置序列，其中y_{j+1}是y_j加权均值，在核函数G上得到，而y_j是核函数的初始中心。

$$y_{j+1}=\frac{\sum_{i=1}^{n}x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)},j=1,2,... \tag{4.23}$$

根据核函数K，上式中的点上密度估计值是：

$$f_K=\{f_K(j)\}_{j=1,2,...}=\{f_K(y_j)\}_{j=1,2,...} \tag{4.24}$$

上两式都具有收敛性[55]。

假设一采样集 $S=\{s_i:s_i\in R^n\}$和一个核函数K，则用核函数表示x点的采样均值为

$$m(x)=\frac{\sum_i s_i K(s_i-x)}{\sum_i K(s_i-x)} \tag{4.25}$$

Fukunaga 和 Hostetler[50]用平坦的核函数K将均值平移定义为差值$m(x)-x$，Cheng, Y[51]将均值偏移的定义一般化，其中$m(x)$被定义为：

$$m(x)=\frac{\sum_i s_i K(s_i-x)w(s_i)}{\sum_i K(s_i-x)w(s_i)} \tag{4.26}$$

式中$w(s_i)$表示搜索窗口。

通过多次迭代$x\leftarrow m(x)$的过程成为 Mean Shift 算法。如果核函数K是高斯函数，$\hat{P}(x)=C\sum_i K(x-s_i)w(s_i)$是用核函数表示的密度估计，那么

$$\frac{\nabla\hat{P}(x)}{\hat{P}(x)}=m(x)-x \tag{4.27}$$

可以看出均值偏移 $m(x)-x$ 在密度估计 $\hat{P}(x)$ 的梯度方向上，因此连续的迭代 $x \leftarrow m(x)$ 将会收敛于密度的局部极大值，即满足 $m(x)=x$ 的固定点，也就是说均值偏移是一个最陡上升的过程。

4.4.2 Mean Shift 算法的搜索过程

图 4.5 是 Mean Shift 算法的流程。

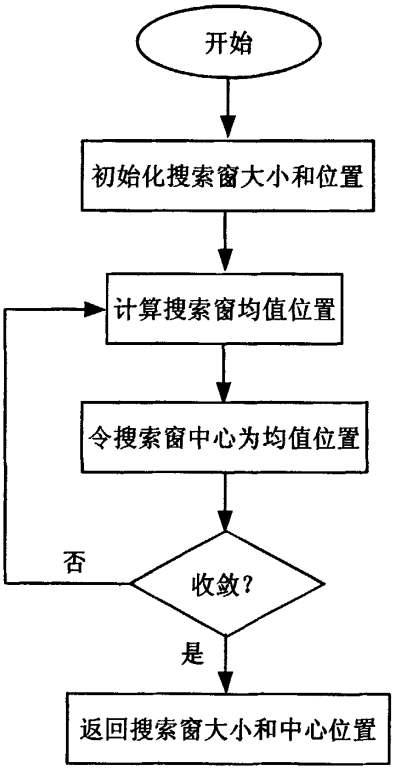


图 4.5 Mean Shift 算法流程

Fig. 4.5 The flow chart of Mean Shift algorithm

算法描述如下：

(1)选择搜索窗的大小；

(2)选择搜索窗的初始中心位置；

(3)计算搜索窗的均值位置；

(4)将搜索到的均值位置设置为搜索窗口的中心位置；

(5)重复 3，4 直到收敛(或者直到均值位置的移动小于预设值)；

4.4.3 Mean Shift 算法在跟踪中的应用

Mean Shift 算法是一种半自动跟踪方法，在起始跟踪帧，通过手动或检测方法确定搜索窗口来选择运动目标。计算核函数加权下的搜索窗口的直方图分布，用同样的方法计算当前帧对应窗口的直方图分布，以两个分布的相似性最大为原则，使搜索窗口沿密度增加最大的方向，移动到目标的真实位置。

本节内容参见[55]，这里根据前面检测到的人体位置和大小，确定了被跟踪目标的初始区域，区域的大小等于核函数的尺度，即核函数的作用区域。对初始目标区域内所有的像素点，计算特征空间中每个特征值的概率，称为目标模型。以后的每帧图像中存在的目标候选区域中对特征空间每个特征值的计算成为候选模型。下面介绍目标模型和候选模型的建立及其 Mean Shift 跟踪过程。

(1)初始目标模型的建立

通过第三章中对连续视频图像序列的分析处理，得到某一目标的位置和大小，也即确定了搜索窗口大小和位置。以一定间隔的颜色值为单位，将取值为颜色特征空间量化为多个特征 $u=1,...,\mathrm{m}$。同时，为了消除目标模板尺寸不同造成的影响，现假设所有的目标模板首先进行归一化处理，经过归一化后的目标模板长度分别为 h_x 和 h_y，即为核函数的尺度，并且区域的中心坐标为(0,0)，定义 b：$\mathrm{R}^2 \to \{1..\mathrm{m}\}$，为图像 x_i^* 处的索引函数，即 $\mathrm{b}(x_i^*)$ 表示取出 x_i^* 处像素对应的特征值。那么在初始帧，包含目标的搜索窗口中，第 u 个特征的概率可以通过下式计算得到[55]

$$\hat{q}_u = C\sum_{i=1}^{n} k(\|x_i^*\|^2)\delta\left[b(x_i^*)-u\right] \qquad u=1,...,\mathrm{m} \tag{4.28}$$

$\hat{q}=\{\hat{q}_u\}_{u=1,...,m}$ 组成目标模型直方图(目标模型的概率直方图)，其中 $\delta(\mathrm{x})$ 是 Delta 函数，$k(x)$ 为核函数。C 为一个标准化常量系数，使得

$$\sum_{u=1}^{m}\hat{q}_u = 1$$

因此

$$C = \frac{1}{\sum_{i=1}^{n} k\left(\|x_i^*\|^2\right)}$$

(2)候选模型的建立

设$\{x_i\}i=1,...,n$是候选目标区域的归一化像素位置，y_0是当前帧搜索窗口的中心像素坐标，利用核宽为h的剖面函数$k(x)$。类似式(4.28)，当前帧中搜索窗口的特征值u的概率为：

$$\hat{p}_u(y)=C_h\sum_{i=1}^{n}k(\left\|\frac{y_0-x_i}{h}\right\|^2)\delta\left[b(x_i)-u\right]\quad u=1,...,m \tag{4.29}$$

$\hat{p}=\{\hat{p}_u(y)\}_{u=1,...,m}$组成候选模型直方图(候选模型的概率直方图)，其它参数含义如式(4.28)所述，此处归一化常数C_h为：

$$C_h=\frac{1}{\sum_{i=1}^{n}k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)}$$

这里需要注意的是常数C_h，因为不依赖于y_0，y_0也是$\{x_i\}i=1,...,n$的其中之一，所以常数C_h可以在给定核剖面函数k(x)和其尺度h后提前计算，这里尺度h的选择被定义为候选目标的尺度。在实际运算的过程中，也就是目标的尺寸。

(3)相似函数度量

相似性函数描述目标模型与当前帧模型之间的相似程度，在理想情况下两个模型的概率分布是完全一样的。这种函数有很多，比如：Bhattacharyya 系数，Fisher linear discrimant、直方图交集及 Kullback 散度等。Bhattacharyya 系数是一种散度型测量，其直接的几何意义是两个向量间角度的 cosine 值，Comaniciu 在文献中说明了在 Mean Shift 算法中 Bhattacharyya 系数是优于其他相似性函数的一种选择。对于目标模型为m值的直方图，设目标的离散密度函数估计为

$$\hat{q}=\{\hat{q}_u\}_{u=1,...,m}$$

在目标y处的候选目标的密度函数估计：

$$\hat{p}=\{\hat{p}_u(y)\}_{u=1,...,m}$$

这里定义目标模板和候选模板这两个离散分布的距离为：

$$d(y)=\sqrt{1-\hat{\rho}(y)} \tag{4.30}$$

其中

$$\hat{\rho}(y)\equiv\rho[p(y),q]=\sum_{u=1}^{m}\sqrt{\hat{p}_u(y)\hat{q}_u} \tag{4.31}$$

式(4.31)为目标模板分布与候选模板分布的 Bhatacharyya 系数，其值在 0~1 之间，$\hat{\rho}(y)$的值越大，表示两个模型越相似。使目标模型和候选模型匹配程度最大，需要$d(y)$最小，又由式(4.30)可见，距离函数$d(y)$的最小化与 Bhatacharyya 系数$\hat{\rho}(y)$的最大化是等价的，因此求当前帧中目标位置转化为求 Bhatacharyya 最大值的问题。

(4) 目标定位过程

定位过程，也即求$d(y)$最小化的过程是从前一帧的目标模板位置y_0处开始，在当前帧中搜索与之匹配的新的目标位置，因此，首先计算当前帧y_0处的候选目标的概率密度$\{p_u(y_0)\}_{u=1,\ldots,m}$，对 Bhattacharyya 系数$\rho(y)$在$y_0$处用 Taylor 公式展开，考虑在通常情况下，两帧之间的时间间隔很短，可以保证候选目标与初始目标模板之间没有剧烈的变化，所以，式(4.31)在$\hat{p}_u(\hat{y}_0)$点泰勒展开可得

$$\rho[p(y),q]\approx\frac{1}{2}\sum_{u=1}^{m}\sqrt{p(y_0)q_u}+\frac{1}{2}\sum_{u=1}^{m}p_u(y)\sqrt{\frac{q_u}{p_u(y_0)}} \tag{4.32}$$

把式(4.29)带入式(4.32)，整理可得，

$$\rho[p(y),q]\approx\frac{1}{2}\sum_{u=1}^{m}\sqrt{p(y_0)q_u}+\frac{C_h}{2}\sum_{i=1}^{n}w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \tag{4.33}$$

其中

$$\mathrm{w}_i=\sum_{u=1}^{m}\sqrt{\frac{q_u}{p_u(y_0)}}\delta[b(x_i)-u] \tag{4.34}$$

由于式(4.32)中的第一项是与y无关的，所以只需要对第二项进行最大化处理即可。第二项是在当前帧中利用核剖面函数$k(x)$和图像像素的加权值计算得到的概率密度估计。令

$$f=\frac{C_h}{2}\sum_{i=1}^{n}w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \tag{4.35}$$

当核函数k和g满足式(4.16)时，从y_0处递归计算出新的目标位置y_1

$$y_1 \equiv \frac{\sum_{i=1}^{m} x_i \omega_i g\left(\left\| \frac{y_0 - x_i}{h} \right\|^2\right)}{\sum_{i=1}^{m} \omega_i g\left(\left\| \frac{y_0 - x_i}{h} \right\|^2\right)} \tag{4.36}$$

Mean Shift 算法以 y_0 为起点，以两个模型相似比最大的方向移动，具体迭代过程如下：

(1)初始化目标模板 q_u；

(2)在当前帧中 y_0 处，计算 $\{p_u(y_0)\}_{u=1,\ldots,m}$，由公式(4.33)计算目标模型和候选模型的相似比系数；

(3)计算权值 w_i；

(4)由公式（4.36）寻找下一个位置 y_1，计算 y_1 处的候选目标模型，重新计算 $\rho(y_1)$

(5)若 $\rho(y_1) < \rho(y_0)$，$y_1 \leftarrow 1/2(y_1 + y_0)$，若 $\|y_1 - y_0\| < e$，停止计算；否则 $y_0 \leftarrow y_1$ 转到第二步计算。

Mean Shift 搜索当前目标位置的过程概括为：首先在跟踪区域内，根据先前帧位置计算目标模型，当前帧建立初始候选模型，然后利用加权的 Mean Shift 迭代不断改变候选模型位置，最后找到目标中心位置，从而确定当前帧中目标的位置。

4.5 融合色彩、纹理和运动信息的跟踪算法

为了描述一个目标，可以采用一个或更多的特征空间来估计非参概率密度函数。特征空间的理想选择应该是一种能够明显地区分目标和周围背景的特征，同时它也应该对噪声和图像混乱具有一定的鲁棒性，最常用的特征空间是 RGB 颜色空间。然而，还有其它特征供选择，例如亮度和对比度等。在 Mean Shift 跟踪算法中，目标表示方法对跟踪性能有着重要影响。Comaniciu 等[55]利用 RGB 值作为特征，它把空间加权的颜色直方图(对应 Mean Shift 中的目标模型和候选模型)作为相似函数的输入，通过 Mean Shift 迭代使相似函数最大达到跟踪的目的。Comaniciu 等对颜色空间中的每维特征进行量化，构建16×16×16 bins 颜色直方图。实际上该直方图是一个三维的立体图如图 4.6 所示。其中，每一维分别对应于色彩图像空间的 Red、Green 和 Blue 通道。Collins 等[6]和 Liang 等[9]利用在线特征选择机制，从式(4.37)RGB 种子特征集 F_1 中选择最优 RGB 特征组合，来提高跟踪性能。

$$F1 = \{w_1 R + w_2 G + w_3 B \mid w_i \in [-2, -1, 0, 1, 2]\} \tag{4.37}$$

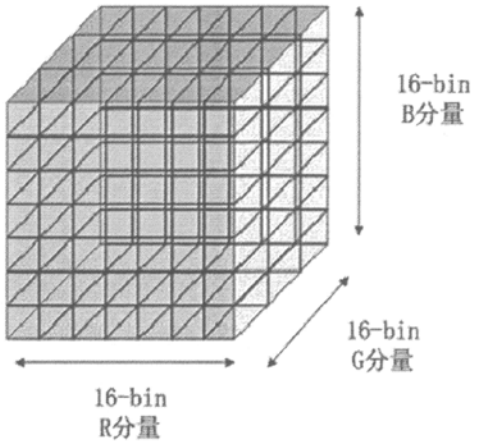


图 4.6 16×16×16-bin 直方图的 3D 视图表示

Fig.4.6 3D view of the 16×16×16-bin histogram

图4.7 不同目标及其相应的纹理表示；

（a）目标1；（b）目标1纹理图；（c）目标2；（d）目标2纹理图

Fig. 4.7 Different objects and their corresponding texture representation

(a) Object 1; (b) The texture map of object 1; (c) Object 2; (d) The texture map of object 2

仅利用RGB颜色空间，当物体之间的颜色分布极其相似时，跟踪很容易出现混乱，同时，RGB颜色空间易受噪声的影响。最近，基于像素灰度值的LBP纹理作为一个纹理原被认为目标的纹理描述符。由于基于核的Mean Shift算法，在跟踪中利用的特征空间中不局限于颜色空间。如图4.7所示，不同的物体具有不同的LBP纹理，可以把描述目标特征的纹理信息加入到Mean Shift，来跟踪目标。Quang等人[44]，考虑图像亮度值(灰度值)作为图像平面的实值函数，LBP作为被跟踪目标的局部不变纹理原，基于此观点，构建图像灰度值和LBP纹理特征值的2维 $k\times k$ 直方图描述(目标跟踪器的目标模型和候选模型)，然后利用基于核的Mean Shift算法进行目标跟踪。该跟踪方法主要用于跟踪单色(灰度)热视频图像序列中的简单运动目标，当视频图像序列对比度较低，且跟踪目标数

较少时，有较好的跟踪效果。此外，该算法对光照和背景纹理的变化具有很好的鲁棒性，由于跟踪算法建立在二维特征空间而不是三维的RGB空间上，所以降低了计算复杂度。但当目标外观比较复杂并且处理的图像序列为彩色图像时，不能很好地进行跟踪。

本文在Quang等人[44]跟踪算法的启发下，针对他们在跟踪彩色视频图像且图像对比度较大，目标是较复杂人体目标时，跟踪不理想的缺点，结合Comaniciu等[55]利用RGB值作为特征的Mean Shift算法，对该算法进行了改进。本文结合RGB颜色信息和纹理信息来描述目标，进而利用Mean Shift跟踪算法准确地跟踪运动目标。首先在第三章目标检测位置的基础上，利用以前帧的位置信息用Kalman滤波器预测当前帧目标的大致位置，然后利用RGB颜色信息和纹理信息建立目标模型和候选模型，即在描述目标时同时在彩色空间和灰度图像空间考虑它们的RGB信息和纹理信息，利用Mean Shift算法进行跟踪，较好地解决了目标快速运动和目标外观复杂时，Quang跟踪算法失败的问题。在利用Kalman滤波器预测的以每个目标位置为中心的矩形跟踪区域内，建立目标模型，对每个区域的每个像素，把RGB彩色图像的每个通道颜色值和灰度图像对应的LBP值分别量化成 k 个bin，形成 $k\times k\times k\times k$ bins的四维直方图特征向量。

$$\hat{q}_u = C\sum_{i=1}^{n} k(\|x_i^*\|^2)\delta\left[b(x_i^*)-u\right] \qquad u=1,...,\mathrm{m}$$

其中，在目标区域同个像素位置处的各个通道量化值和LBP值作为直方图的每个bin的索引。m为 $k\times k\times k\times k$，也即 k^4，其中前面的 $k\times k\times k$ 分别指RGB颜色通道中R、G和B中的量化值，最后的 k 指灰度图像对应的LBP量化值，候选模型的建立过程类似于式(4.29)。

然后利用Bhattacharyya系数来度量目标模型和候选模型的相似度，最后利用Mean Shift迭代在每个目标的四维特征空间中使Bhattacharyya系数最大值处定位目标。

融合颜色信息、纹理信息和目标运动信息的跟踪过程为，在前一帧利用第三章头部定位方法，根据人身体一定的比例关系，得到每个人体的大致位置，在该位置附近利用Kalman 滤波器预测当前帧位置，然后利用 Mean Shift 在预测位置附近给出当前帧准确的目标位置，该过程重复直到运动目标走出视频区域或视频序列结束。

4.6 实验结果与分析

4.6.1 跟踪结果

在实验中，本文对不同监控场景下的几个视频进行了跟踪测试，本文在 PC 机(2.4GHZ，256M)上进行了实验，该跟踪系统是在 Windows XP 系统下用 VC6.0 编程实现的，采用的视频图像序列的大小为320×240 。在本文跟踪算法中，由第三章目标定位

位置，利用 Kalman 滤波器预测当前位置，采用 RGB 颜色空间和 4.3 节的基本 LBP 值，然后分别对它们进行量化，形成8×8×8×8 bins 直方图的目标模型和候选模型，然后利用基于核的 Mean Shift 在四维特征空间上搜索确定目标位置，进行准确跟踪。在室内单人运动目标情况下，我们与 Quang[41]跟踪方法进行比较，Quang 利用灰度值和 LBP 值构造16×16的 2 维直方图。我们采用矩形框来框住目标，其中红色和黑色圆为利用第三章圆检测方法确定的人体头部位置，蓝色框为 Kalman 预测结果，绿色框为 Mean Shift 最后跟踪结果，黑色框为利用 Quang 方法跟踪结果。二者的比较结果如图 4.8(该视频长度为 15000 帧)。

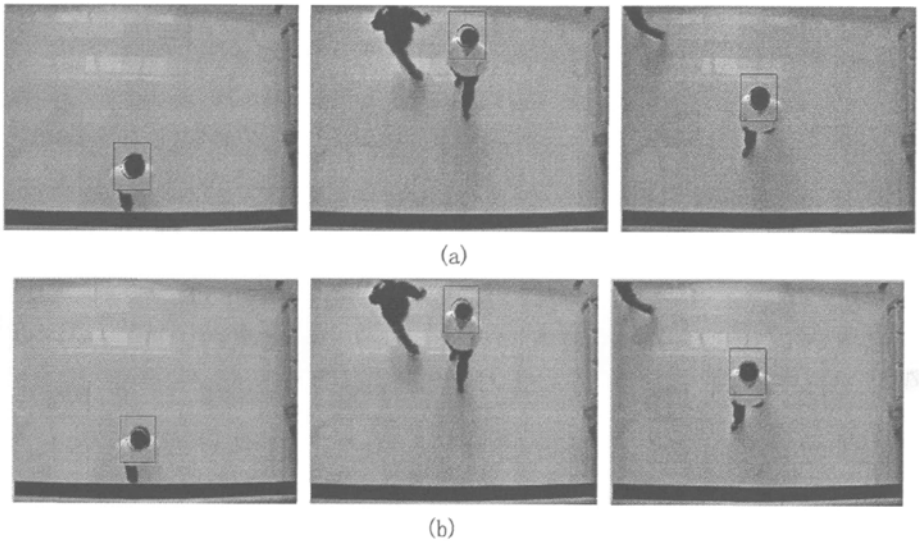


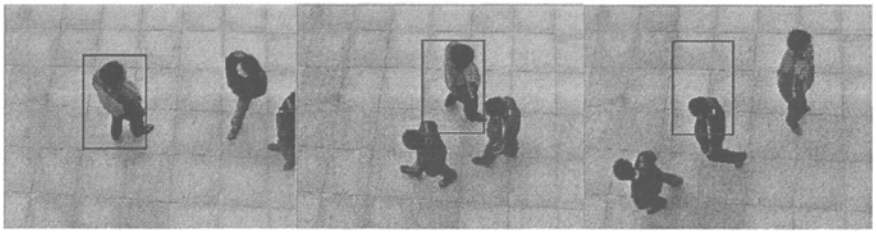
图 4.8　视频 1 检测与跟踪效果

（a）LBP和灰度值的跟踪结果；（b）结合色彩、纹理和运动信息的跟踪结果图

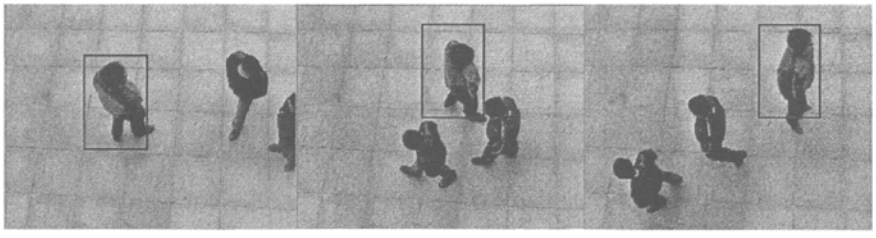
Fig. 4.8　video1 detection and tracking results

(a) Image intensity with LBP; (b) Tracking result with color, texture and motion information

在室外环境多运动目标情况下，本文与仅利用 RGB 颜色信息的 Mean Shift 跟踪算法进行了针对单个目标的跟踪实验比较，由于只跟踪单个目标，本实验不需要进行前期的目标检测工作，针对要跟踪的首次出现的目标手动初始化它，然后用两个跟踪算法分别进行跟踪，跟踪效果如图 4.9 所示。



(a)



(b)

图 4.9　视频 2 跟踪结果图比较(从左到右依次为 5104 帧，5182 帧和 5231 帧)：(a) 利用 RGB 信息跟踪方法结果；(b) 本文改进的方法跟踪结果

Fig. 4.9　The comparison of tracking result with Video 2(frame 5104，from 5182 and frame 5231)：(a) Tracking result of tracking method with RGB information；(b) Tracking result of our improved method

在室内和室外不同的环境下，本文对另三个存在多个运动目标的视频进行了跟踪实验，跟踪结果如图 4.10 所示。

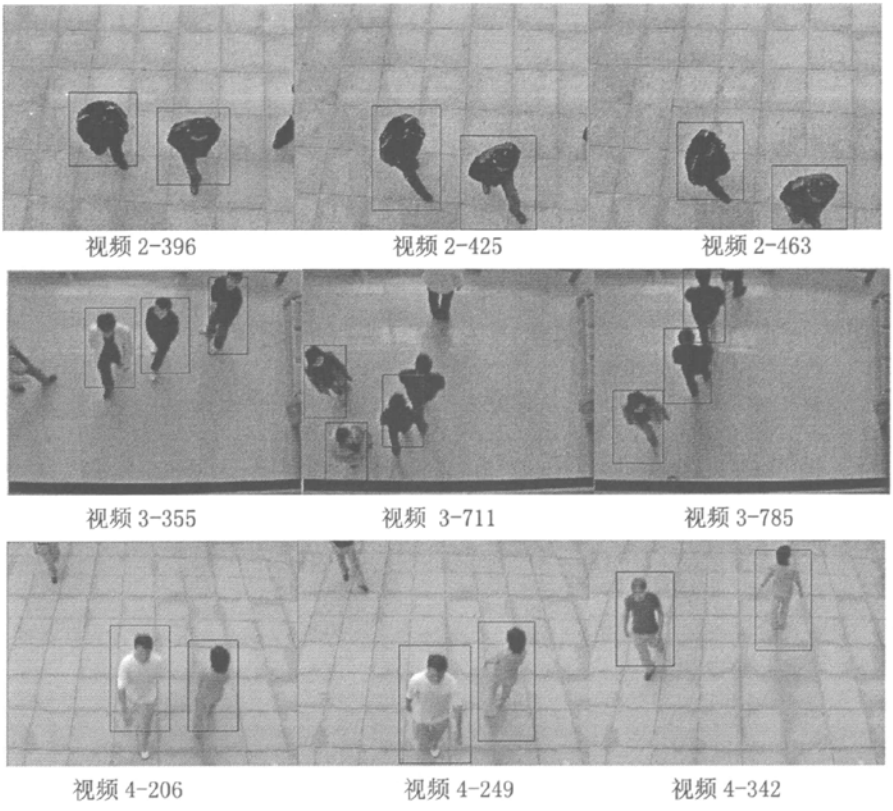


图 4.10 视频 2、视频 3 与视频 4 跟踪结果图

Fig.4.10 Tracking results of Video 2 , Video 3 and Video 4

4.6.2 跟踪性能分析

在对如图 4.8 室内单个人体目标进行跟踪时，由于考虑了目标的颜色、纹理和运动信息，所以可以准确地跟踪目标。本文的结果比 Quang 方法结果有了一定程度改进。在如图 4.9 室外环境多运动目标环境跟踪单个目标时，当两个目标距离较近且颜色相似时，我们的跟踪方法由于融入了纹理和运动信息，所以从帧 5182 帧两目标靠近，到目标之间逐渐远离(帧 5231)时，我们的跟踪算法都能准确跟踪目标。而 Comaniciu 等[55]方法把 RGB 颜色信息嵌入到 Mean Shift 中进行跟踪，但两目标特征极其相似，所以目标从擦肩开始核函数会漂移到另一个目标，造成跟踪失败。

由于 Mean Shift 算法用直方图作为目标的模式，若仅利用颜色信息，那么目标之间则容易混淆，若利用灰度和纹理信息，当图像对比度大时，跟踪效果不理想。但本文中加入结合颜色、纹理和运动信息，从图 4.10 视频 3 中可以看到，有 3 人他们在颜色分布

上具有很大的相似性，但由于很好地结合了三个不同特征信息，有效地抑制了这种情况的发生。由于目标有时会存在距离很近且目标间很相似的情况，这时会出现有的目标跟踪不上的情况，如图 4.10 视频 3 第 711 帧，这也是本文今后需要改进的地方。

同时本文用 Comaniciu 等[57]的 Mean Shift 结合 Kalman 滤波器跟踪算法，Quang 等[44]的灰度和 LBP 值跟踪算法，及本文中融合颜色、纹理和运动信息的跟踪算法针对相同视频作了跟踪测试，从 Mean Shift 的快速收敛次数、目标运动的快速性以及多目标运动的非线性这三个方面进行了分析，得到表 4.1。

表 4.1 算法性能分析

Tab. 4.1 The performance analysis of algorithm

	Comaniciu 等	Quang 等	本文方法
利用的特征	运动和颜色信息	灰度值和纹理信息	运动、颜色和纹理信息
迭代次数	6	6	4
目标快速运动	较好	一般	较好
非线性运动	不好	较好	较好

从表 4.1 中我们可以看出，本文中采用的跟踪算法，对目标的快速跟踪和非线性运动的跟踪处理方面都有了较大的改善。对于拍摄的不同视频和不同的目标，Mean Shift 的迭代次数和目标的运动速度直接相关，运动速度快的目标相对的搜索迭代次数也较多，运动速度较慢的目标迭代次数较少。

4.7 小结

本章首先引入了一种新的在目标跟踪中用于描述目标的特征-纹理特征，介绍了 Kalman 滤波器跟踪，然后重点讲述了 Mean Shift 理论及其在目标跟踪中的应用，最后把颜色信息和纹理信息融合到 Mean Shift 算法中，给出了一种融合目标色彩、纹理和运动信息的改进的 Kalman 和 Mean Shift 跟踪算法。

5 Tri-tracking 跟踪算法

5.1 引言

传统的跟踪算法通过建立通用模型来描述目标的外观，例如第四章中通过 Mean Shift 建立目标模型和候选模型来找到目标位置。由于外观模型不断变化，为了适应这种变化，目标模型需要不断更新。然而，物体外观模型的变化是非线性的，很难找到一个合理的模型，目标模型的不正确更新也会导致漂移问题[58]。而问题的本质是，一般的跟踪器仅依赖于前景，完全忽略了背景，而背景是导致漂移的根部原因。

最近，有些代表性的跟踪方法把跟踪看作为目标和背景的两类分类问题，不是建立复杂的模型来描述目标，而是找到决策来区分目标和背景，当目标外观变化时，该方法只需要更新决策而不用更新目标外观模型。Collins 等[56]首次把跟踪看作两类分类问题，利用先前帧学习的分类器对当前帧进行分类。Avidan 的集成分类器[59]也把跟踪问题看作目标和背景的分类问题。它通过在线训练多个弱分类器，利用 Adaboost 把多个弱分类器结合成一个强分类器(集成分类器)，利用该强分类器区分每个像素属于目标或背景，形成一个置信图，最后利用 Mean Shift 在置信图上找到顶点，即找到目标的当前帧位置。同时，为了适应目标外观的变化，用不断学习的弱分类器来更新集成分类器。协同跟踪方法(Co-tracking)[60]借助两个分类器通过为标定的数据互相学习来训练分类器，运用两个独立的特征——颜色直方图和 HOG (方向梯度直方图)来描述目标在线训练两个分类器，然后分类器协同地对标定的数据进行分类，同时用这个新标定的数据更新分类器，但该算法当目标的运动姿势变化较大时，存在跟踪失败的问题。本文在研究该算法的基础上，对其进行了改进，给出了一种新的跟踪单个运动目标的算法，把跟踪运动目标问题也看作目标和背景的分类问题。利用三个不同的特征——颜色直方图特征、LBP 直方图特征和 PPBTF 直方图特征，训练三个 SVM 分类器，利用机器学习的方法对目标外观的变化不断学习和更新，并随着时间推移不断获得目标和背景新的外观信息，达到了较好的跟踪效果。

下面首先对机器学习理论以及协同训练算法进行探讨，然后对该算法中训练分类器的 PPBTF 直方图特征以及 SVM 分类器做了简单介绍，最后给出 Tri-tracking 算法的流程，并对该算法做了大量实验，同时分析了算法的跟踪性能。

5.2 机器学习

机器学习不仅是人工智能的一个核心研究领域，而且已成为整个计算机领域中最活跃、应用潜力最明显的领域之一，它扮演着日益重要的角色。机器学习是伴随着人工智

能理论和技术的发展而产生一个重要领域，是智能系统不断积累经验以改善系统性能的过程。20 世纪 80 年代，机器学习蓬勃发展起来，逐渐形成为人工智能研究的一个主流研究领域。遗传算法的研究，人工神经网络的兴趣，都给了机器学习领域带来了强有力的生命力，并对人工智能的其他研究与应用领域产生了很大的影响。机器学习的系统模型如图 5.1。其中，环境向系统提供学习信息，学习元对这些信息进行整理、分析、归纳和类比，生成新的知识元或改进知识库的组织结构，执行元以学习后得到的新知识库为基础，执行一系列任务，并将执行结果报告学习元，以完成对新知识的评价，指导进一步的学习工作。

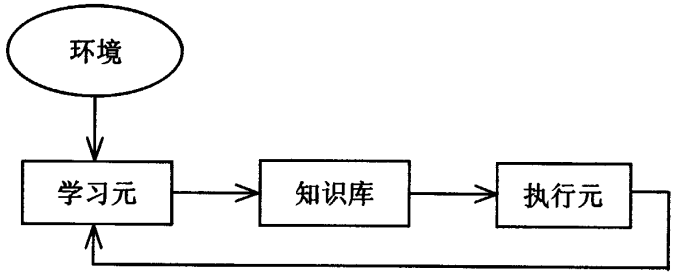


图 5.1 机器学习系统模型框图

Fig. 5.1 The flow chart of machine learning model

机器学习分为监督学习、半监督学习和无监督学习，准确来说半监督学习还是监督学习的一种。有无监督的学习主要区别是有无标记的数据，而半监督学习的主要就是利用无标记的数据从而达到最终监督学习的目标，或者会自动标记。

5.2.1 半监督学习

传统的监督学习算法需要利用大量有标记的(labeled)样本进行学习，从而建立模型预测未见数据的标记，在分类问题中标记就是示例的类别。随着信息技术的飞速发展，收集大量未标记的(unlabeled)样本已相当容易，而获取大量有标记的示例则相对较为困难，因为获得这些标记可能需要耗费大量的人力物力。例如在计算机辅助医学图像分析中，可以从医院获得大量的医学图像作为训练例，但如果要求医学专家把这些图像中的病灶都标识出来，则往往是不现实的。事实上，在真实世界问题中通常存在大量的未标记示例，但有标记示例则比较少，尤其是在一些在线应用中这一问题更加突出。如何利用大量的未标记样本来改善学习性能成为当前机器学习研究中备受关注的问题。

目前，利用未标记示例的主流学习技术主要有三大类，即半监督学习(semi-supervised learning)、直推学习(transductive learning)和主动学习(active learning)。这三类技术都是试图利用大量的未标记示例来辅助对少量有标记示例的学习，但它们的基本思想却有显著的不同。在半监督学习[61]中，学习器试图自行利用未标记示例，即整个学习过程不需人工干预，仅基于学习器自身对未标记示例进行利用。直推学习与半监督学习的相似之处是它也是由学习器自行利用未标记示例，但不同的是，直推学习假定未标记示例就是测试例，即学习的目的就是在这些未标记示例上取得最佳泛化能力。主动学习和前面两类技术不同，它假设学习器对环境有一定的控制能力，可以“主动地”向学习器之外的某个“神谕”进行查询来获得训练例的标记。因此，在主动学习中，学习器自行挑选出一些未标记示例并通过“神谕”查询获得这些示例的标记，然后再将这些有标记示例作为训练例来进行常规的监督学习，而其技术难点则在于如何使用尽可能少的查询来获得强泛化能力。

一般认为，半监督学习的研究始于 B. Shahshahani 和 D.Landgrebe 的工作[62]，但为标记示例的价值实际上早在上世纪 80 年代末就已经被一些研究者意识到了。D.J. Miller 和 H.S. Uyar [63]认为，半监督学习的研究起步相对较晚，可能是因为在当时的主流机器学习技术(例如前馈神经网络)中考虑未标记示例相对比较困难。随着统计学习技术的不断发展，以及利用未标记示例这一需求的日渐强烈，半监督学习近年来逐渐成为一个研究热点，国内外开展了大量研究，取得了很多研究成果。

半监督学习的基本设置是给定一个来自某未知分布的有标记示例集 $L=\{(x_1,y_1),(x_2,y_2),...,(x_{|L|},y_{|L|})\}$ 以及一个未标记示例集 $U=\{x_1',x_2',...,x_{|U|}'\}$，期望数学函数，$f:X->Y$ 可以准确地对示例 x 预测其标记 y。这里 x_i，$x_j' \in X$ 均为 d 维向量，$y_i \in Y$ 为示例 x_i 的标记，$|L|$ 和 $|U|$ 分别为 L 和 U 的大小，即它们所包含的示例数。

根据半监督学习算法的工作方式，可以大致将现有的很多半监督学习算法分为三大类。第一类算法以生成式模型为分类器，将未标记示例属于每个类别的概率视为一组缺失参数，然后采用 EM 算法来进行标记估计和模型参数估计，其代表有[63]。此类算法可以看成是在少量有标记示例周围进行聚类，是早期直接采用聚类假设的做法。第二类算法是基于图正则化框架的半监督学习算法，其代表有[64]。此类算法直接或间接地利用了流形假设，它们通常先根据训练例及某种相似度度量建立一个图，图中结点对应了(有标记或未标记)示例，边为示例间的相似度，然后，定义所需优化的目标函数并使用决策函数在图上的光滑性作为正则化项来求取最优模型参数。第三类算法是协同训练(co-training)算法。此类算法隐含地利用了聚类假设或流形假设，它们使用两个或多个学

习器，在学习过程中，这些学习器挑选若干个置信度高的未标记示例进行相互标记，从而使得模型得以更新。

半监督学习提供一个总体框架对没有足够标记数据的不同类型的物体训练分类器。目前半监督学习的例子主要有 EM(Expectation-Maximization)算法[65]，协同训练(co-training)算法[66]，tri-training 算法[67]和直推支持向量机[68]。

5.2.2 协同训练算法

最初的协同训练算法(或称为标准协同训练算法)是 A. Blum 和 T. Mitchell[66]在 1998 年提出的，该算法提出后，很多研究者对其进行了研究并取得了很多进展，使得协同训练成为半监督学习中最重要的风范(paradigm)之一，而不再只是一个算法。他们假设数据集有两个充分冗余(sufficient and redundant)的视图(view)，即两个满足下述条件的属性集：第一，每个属性集都足以描述该问题，也就是说，如果训练例足够，在每个属性集上都足以学得一个强学习器；第二，在给定标记时，每个属性集都条件独立于另一个属性集。A. Blum 和 T. Mitchell 认为，充分冗余视图这一要求在不少任务中是可满足的。A. Blum 和 T. Mitchell 的算法在两个视图上利用有标记示例分别训练出一个分类器，然后，在协同训练过程中，每个分类器从未标记示例中挑选出若干标记置信度(即对示例赋予正确标记的置信度)较高的示例进行标记，并把标记后的示例加入另一个分类器的有标记训练集中，以便对方利用这些新标记的示例进行更新。协同训练过程不断迭代进行，直到达到某个停止条件。

S. Goldman 和 Y. Zhou[69]提出了一种不需要充分冗余视图的协同训练算法。他们使用不同的决策树算法，从同一个属性集上训练出两个不同的分类器，每个分类器都可以把示例空间划分为若干个等价类。在协同训练过程中，每个分类器通过统计技术来估计标记置信度，并且把标记置信度最高的示例进行标记后提交给另一个分类器作为有标记训练例，以便对方进行更新。该过程反复进行，直到达到某个停止条件。在预测阶段，该算法先估计两个分类器对未见示例的标记置信度，然后选择置信度高的分类器进行预测。此后，他们又对该算法进行了扩展，使其能够使用多个不同种类的分类器。

为了进一步放松协同训练的约束条件，Zhou 等[67]提出了一种既不要求充分冗余视图、也不要求使用不同类型分类器的 tri-training 算法。该算法的一个显著特点是使用了三个分类器，不仅可以简便地处理标记置信度估计问题以及对未见示例的预测问题，还可以利用集成学习(ensemble learning)[70]来提高泛化能力。该算法首先对有标记示例集进行可重复取样(bootstrap sampling)以获得三个有标记训练集，然后从每个训练集产生一个分类器。在协同训练过程中，各分类器所获得的新标记示例都由其余两个分类器协作

提供，具体来说，如果两个分类器对同一个未标记示例的预测相同，则该示例就被认为具有较高的标记置信度，并在标记后被加入第三个分类器的有标记训练集。在对未见示例进行预测时，tri-training 算法不再象以往算法那样挑选一个分类器来使用，而是使用集成学习中经常用到的投票法来将三个分类器组成一个集成来实现对未见示例的预测。

近年来，利用机器学习半监督框架进行跟踪的研究越来越多，Avidan[59]把目标跟踪问题看作前景和背景分类问题，利用集成学习方法，在线训练多个弱分类器，然后利用 Adaboost 把这些弱分类器结合成强分类器，利用该分类器对新一帧的每个像素进行分类，形成置信图，然后利用 Mean Shift 算法找到置信图的顶点，也即当前帧的目标位置。王震宇等[8]基于 SVM 和 AdaBoost 的红外目标跟踪算法，也把目标跟踪问题转化为目标和背景的两类分类问题，然后根据每一帧的正负样本训练 SVM 作为分量分类器，并通过恰当的参数调整策略，利用 AdaBoost 算法把这些分量分类器组合成一个总体分类器；接着利用该总体分类器来区分下一帧中的目标和背景，并得到置信图；最后通过 Mean Shift 算法找到置信图的峰值，得到目标的新位置。Tang 等的协同跟踪算法[60]利用两个特征——颜色直方图和 HOG(方向梯度直方图)描述目标和背景，训练两个 SVM，根据每个分类器进行的错误率，求得每个分类器的权重，利用两个权重得出一个新的 SVM 分类器，在新的一帧中利用搜索窗方法，形成置信图，并利用 Mean Shift 算法找到目标的位置，然后利用该位置信息形成新的抽样来更新 SVM 分类器，从而对外观变化的目标进行跟踪，但该算法当目标的运动姿势变化较大时，存在跟踪失败的问题。本文在该算法和 tri-training 算法基础上，提出了一种新的半监督跟踪算法，下面详细介绍该跟踪算法。

5.3 Tri-tracking 跟踪算法

5.3.1 PPBTF 直方图特征

PPBTF(Pixel-Pattern-Based Texture Feature)是一种基于像素模式的纹理特征，首先通过统计的方法定义几种待处理图像的纹理模式，然后计算原始图像中每个像素对应的模式，用相应模式的类别标号来表示，这样就把原始图像转化为了模式图，再构造纹理特征。该特征是日本立命馆大学智能图像系统研究室陈延伟等人 2003 年提出的[71]，目前已经成功应用在纹理图像分割方面。对 PPBTF 算法的研究表明，通过恰当选择若干个模板，PPBTF 特征能充分描述图像整体和局部的纹理信息，是一种基于外观的特征提取方法，这是其在图像处理、模式识别领域的一个重要应用之处。

(1)基于模式图的纹理建模

为了去除图像冗余信息和噪声，更好的刻画对判别很重要的纹理信息，引入了模式图的概念，用 M 个模板 $\{w_i\}$ 将原始图像转化成模式图，突出图像中特征较明显的点、线和区域，其中每个模板代表一类模式。设灰度图像为 I，像素坐标为 (x,y)，则 z_i 是像素 (x,y) $S\times S$ 邻域和第 i 个模板的内积，即

$$z_i = b \bullet W_i \tag{5.1}$$

在模式图 P 中，定义像素 (x,y) 的值用 k 来表示，其中 $z_k = \max(z_1, z_2, ...z_M)$，$k$ 即为和该邻域匹配的模板索引，表示原始灰度图像中像素所属的模式类别。

与基于灰度图像构造的特征相比，基于模式图构建的特征优势相当明显。第一，在模式图中，图像的边缘线、特殊点等的纹理更加突出了，由此构建的特征更具有判别能力；第二，模式图中的像素值取值范围变小了，这不仅节省了系统资源，更重要的是由其构建的特征和灰度值已经没有直接关系了，因此不再受灰度值的影响，提高了算法的鲁棒性。

(2) 构建特征矢量

在文[72]构建特征图时，假设模式图中模式数为 M，即用 M 个模板 $\{w_i\}$ 建模，模式图 P 中像素的取值范围则为[1,M]，对任意像素求它周围特征窗内的特征。而本文构建特征窗时，把图像目标区域和背景区域分为若干子区域在模式图上求每个特征个数，即求每个子区域的特征直方图。如第四章扩展二进制模式把 LBP 的统计柱状图作为目标的特征向量方法一样，本文把目标区域和背景区域分别被划分为若干小区域 $R_0, R_1, ..., R_m$，在每个子区域 R_j 直方图可以表示为

$$H_{j,l} = \sum_{(x,y)\in R_j} h_l(x,y) \quad j = 0,...,m, l = 1,...,M \tag{5.2}$$

其中，h 是本文定义的一个二元函数，即

$$h_l(x,y) = \begin{cases} 1 & if \quad P(x,y) = l, \\ 0 & otherwise, \end{cases} \tag{5.3}$$

在表情识别中，最常用的是把 LBP 的统计柱状图作为表情图像的特征向量。为了考虑特征的位置信息，把图像分成若干个小区域，在每个小区域里进行直方图统计，即统计该区域内属于某一模式的数量，最后再把所有区域的直方图依次连接到一起作为特征向量接受下一级的处理[49]，如图 4.4 所示。

在利用 PPBTF 描述运动目标和周围背景时，类似于第四章 4.2 节将统计表情图像分成每个小块，求每小块区域的 LBP 直方图统计，然后把所有区域直方图连接到一起作为特征向量的方法一样。本文把运动目标分成若干子区域，从每个子区域 R_j 提取的

PPBTF 直方图连接成一个空间增强的直方图特征向量作为目标和背景的描述符，如图 5.2 所示。

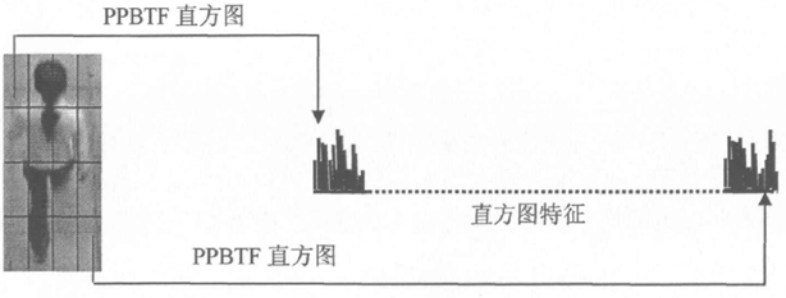


图 5.2 基于 PPBTF 的目标识别特征

Fig. 5.2 Object recognition feature based on PPBTF histogram

5.3.2 支持向量机(SVM)

支持向量机(SVM)是一种机器学习方法，它是在统计学习理论的基础上发展而来的，最早由 Vapnik 等人于 1992 年在计算机理论大会上提出，其主要内容在 1992-1995 年间才基本完成，目前仍处在不断发展阶段，并且被广泛应用到模式识别的各个领域。传统的统计学习方法中采用的经验风险最小化准则(empirical risk minimization，ERM)，虽然可以使训练误差最小化，但并不能最小化学习过程的泛化误差。

它的训练方法是用少数的支持向量来代表整个样本集，其基本思想是在样本空间或特征空间，构造出最优超平面，使得超平面与不同类样本集之间的距离最大，从而达到最大的泛化能力。支持向量机结构简单，并且具有全局最优性和较好的泛化能力。

SVM 是从线性可分条件下的最优分界面发展而来的[72]，其在模式识别应用中的核心思想是得到一个超平面作为最优分类平面(Optimal Hyperplane)，使得正负样本之间有最大的分类间隔。为此要解决一个受限二次规划问题，得到最优分类器。考虑图 5.3 所示的二维两类线性可分情况，圆圈和方块分别表示两类的训练样本，中间的把两类正确分开的直线为最优分类线，所谓最优分类线就是要求分类线不但能将两类完全正确地分开，而且要保证使两类的分类间隙最大。推广到高维空间，最优分类线就成为最优分类面。

设线性可分的样本集为 (x_i, y_i)，$i=1,\cdots,n$，$x\in R^d$，$y\in\{+1,-1\}$ 是类别标号。d 维空间中线性判别函数的一般形式为 $g(x)=w\cdot x+b\sqrt{a^2+b^2}\lim\limits_{x\to\infty}$，分类面方程为：

$$w \cdot x + b = 0 \tag{5.4}$$

将判别函数归一化，使两类所有样本都满足$|g(x)| \geq 1$，即使离分类面最近的样本的$|g(x)|=1$，这样分类间隔就等于$2/\|w\|$，因此使分类间隔最大等价于使$\|w\|$最小；而要求分类面对所有样本都正确分类，就是要求它满足：

$$y_i[(w \cdot x_i) + b] - 1 \geq 0, \quad i = 1, 2, \cdots, n. \tag{5.5}$$

因此，满足上述条件并且使$\|w\|$最小的分类面就是最优分类面。过两类样本中离分类面最近的点且平行于最优分类面的超平面即下图中的$w \cdot x + b = 1$和$w \cdot x + b = -1$两个超平面上的训练样本点称为支持向量。

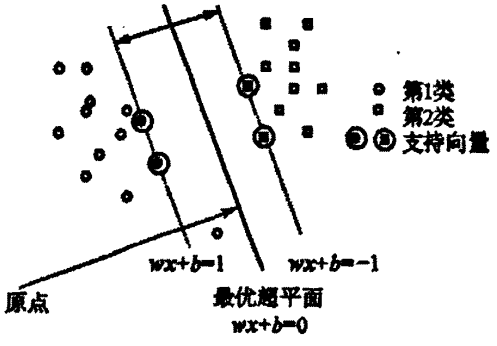


图 5.3 最优分类超平面示例

Fig. 5.3 The optimal hyperplane

根据上面的介绍，最优分类面求解问题可以表示成如下的约束优化问题，即在条件(5.5)的约束下，求函数

$$\phi(w) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w \cdot w) \tag{5.6}$$

的最小值。为此，可以定义如下的 Lagrange 函数：

$$L(w, b, \alpha) = \frac{1}{2}(w \cdot w) - \sum_{i=1}^{n} \alpha_i \{y_i[(w \cdot x_i) + b] - 1\} \tag{5.7}$$

其中，$\alpha_i > 0$为 Lagrange 系数，现在的问题是对w和b求 Lagrange 函数的极小值。

最终得到的最优分类函数为

$$f(x) = \mathrm{sgn}\{(w^* \cdot x) + b^*\} = \mathrm{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*\} \tag{5.8}$$

sgn()为符号函数。由于非支持向量对应的α_i均为 0，因此,式(5.8)中的求和实际上只对支持向量进行。而b^*是分类的域值，可以由任意一个支持向量用式(5.5)进行求解。

5.3.3 Tri-tracking 跟踪算法

本文把跟踪看作前景和背景二类分类问题，在开始的 N 帧，首先通过第四章融合目标空间色彩信息、纹理信息和运动信息的 Kalman/Mean shift 跟踪器的跟踪结果初始化 Tri-tracking 跟踪器，Tri-tracking 跟踪框架把新来的帧作为未标定的数据，对三个独立直方图特征颜色、LBP 和 PPBTF 建立三个 SVM，用标定的帧训练 SVM 分类器，每个分类器分别获得一个权重，然后为每个分类器建立一个置信图，结合每个分类器的权重，形成最终置信图，通过置信图找到当前帧目标位置，然后利用 Tri-tracking 框架获得新的抽样并更新 SVM，进一步求得每个分类器的权重，并从 SVM 分类器移去旧的抽样，该过程不断重复，直到目标从运动区域消失为止。本文采用的跟踪算法流程如图 5.4 所示：

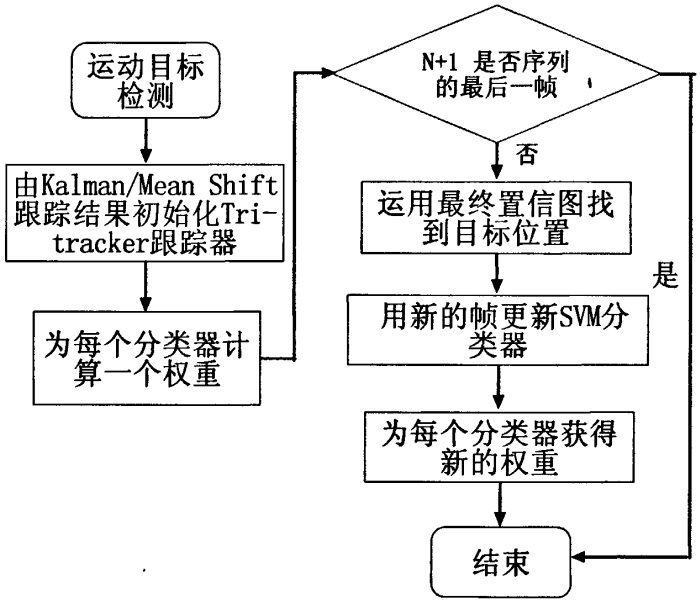


图 5.4 本文跟踪流程框图

Fig. 5.4 Flow chart of the tracking scheme

(1)跟踪器的初始化

对头 N 帧视频图像序列，首先利用第三章的头部定位方法确定头部位置，根据头部与身体其他部位的比例关系，确定目标区域，然后利用 Kalman/Mean Shift 跟踪器给

出精确的目标位置。利用这 N 个标定的帧，通过计算目标区域的特征向量获得正样本，计算周围背景区域的特征向量作为负样本。正样本和负样本对应的区域如图 5.5 所示，其中本文中实蓝色矩形框对应目标，实蓝框外面和虚点划线蓝框内对应背景区域。

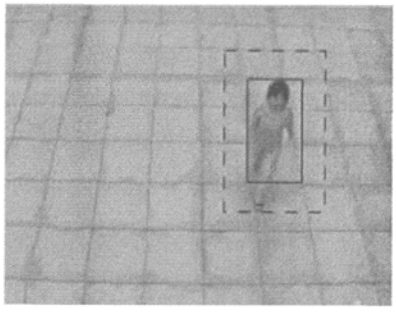


图 5.5 目标和背景表达

Fig. 5.5 Object and background representation

(2)计算分类器权重

通过第一步由 Kalman/Mean Shift 跟踪器获得精确的标定帧后，我们基于 SVM 错误率来计算每个分类器的权重。利用已标定的正负样本，计算分类器的错误率 γ 如下式：

$$\gamma=((N-Y_+)+(N+Y_-))/2$$

$$Y_+=\sum_{i=1}^{N} sign(S(V_{i+})) \quad Y_-=\sum_{j=1}^{N} sign(S(V_{j-})) \tag{5.9}$$

其中 S()为训练的 SVM 分类器，V_{i+} 表示第 i 个正样本，V_{j-} 代表第 j 负样本， N 表示正样本或负样本的数目（也即已标定帧的数目）。根据用颜色直方图，LBP 直方图和 PPBTF 直方图特征计算得到三个 SVM 错误率，三个分类器的权重 W_{LBP}，W_{color} 和 W_{PPBTF} 分别为

$$W_{LBP}=1-(\gamma_{LBP}+\tau)/(\gamma_{LBP}+\gamma_{color}+\gamma_{\mathrm{PPBTF}}+\tau) \tag{5.10}$$

$$W_{color}=1-(\gamma_{color}+\tau)/(\gamma_{LBP}+\gamma_{color}+\gamma_{\mathrm{PPBTF}}+\tau) \tag{5.11}$$

$$W_{PPBTF}=1-(\gamma_{\mathrm{PPBTF}}+\tau)/(\gamma_{LBP}+\gamma_{color}+\gamma_{\mathrm{PPBTF}}+\tau) \tag{5.12}$$

其中 τ 为一个很小的正数来避免当 γ_{LBP}，γ_{color} 和 γ_{PPBTF} 都为零情况的发生。

(3)目标定位

在求出三个分类器权重的情况下，利用每个特征向量（颜色，LBP 和 PPBTF），每个 SVM 分类器会为未标定当前帧产生一个置信图，最终的置信图通过对由每个 SVM 决定位置的分类结果投票建立，最终置信图的产生过程如图 5.6 所示。获得最终置信图后，

我们利用 Mean Shift 算法在当前帧中定位目标，也即找到目标的位置，目标跟踪结果如图 5.6(a)。

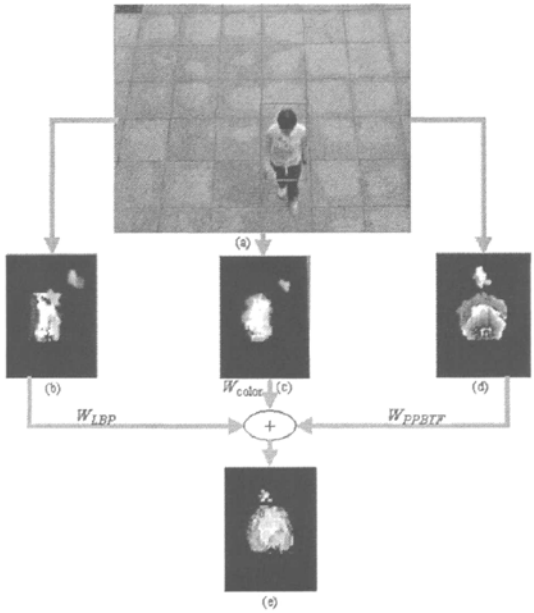


图 5.6　由两或三个特征形成的置信图得到最终置信图过程；（a）是原始视频帧；（b）、（c）和（d）分别为由 LBP 特征，RGB 特征和 PPBTF 特征训练的 SVM 得到的置信图

Fig. 5.6　The process of combining two or three confidence maps from corresponding features into the final confidence map; (a)original video frame and tracking result; (b), (c) and (d) are confidence maps from corresponding SVMs trained by LBP histogram, RGB histogram and PPBTF histogram, respectively.

其中(a)是原始视频帧；(b),(c),(d)分别为由 LBP 特征，RGB 特征和 PPBTF 特征训练的 SVM 得到的置信图。原始视频帧中橘红色框代表对应最后置信图峰值位置，也即跟踪结果。

(4) 更新 SVM 分类器

为了适应目标运动时外观的变化，需要不断地更新 SVM 分类器，以获得较好的跟踪效果。在本文中，利用新的被标定的帧来代替 N 帧中最早旧的帧，也即不断加入新的一帧移去最旧的帧来适应不断变化的外观，确保 SVM 检验集中抽样都能被正确标定。用新的标定帧更新 SVM 分类器后，三个分类器权重利用式(5.10)，(5.11)和(5.12)求得，以便求下一帧的目标位置。

5.4 实验结果与分析

5.4.1 实验结果

本文在 PC 机(2.4GHZ，256M)上进行了实验，该跟踪系统是在 windows XP 系统下用 VC6.0 编程实现的。采用固定摄像头捕获的视频图像序列，图像大小为320×240，在计算中，我们采用粉红色矩形框来框住目标。

LBP 直方图为利用 4.3 节扩展的 LBP 即 $LBP_{P,R}^{riu}$ 算子(其中 $P=8$，$R=1$，riu 代表旋转不变的 Uniform)标定图像求得的 59 维直方图，RGB 颜色直方图为在 $\{R,G,B\}$ 三个空间均匀量化后得到的 64(4×4×4)维直方图。为了计算 PPBTF 直方图，目标和背景区域分别被分为 50 个子区域，其中目标区域内每个子区域的宽和高为 $m_1 \times n_1$，背景区域每个子区域的宽和高 $m_2 \times n_2$。PPBTF 中模式图的 M 为 8，因此由 50 个子区域连接成的 PPBTF 直方图是 400(50×8)特征向量。

第一个实验为实验室人员在校园内自拍的只含单运动目标的视频序列，其中目标区域分块时每个子区域宽高($m_1=10, n_1=8$)，背景区域每个子区域宽高($m_2=20, n_2=9$)。从试验中看到该目标在行走时不断变化姿势，但该算法可以很好地对其跟踪。对应两个视频图像的跟踪结果和置信图如图 5.7 所示。

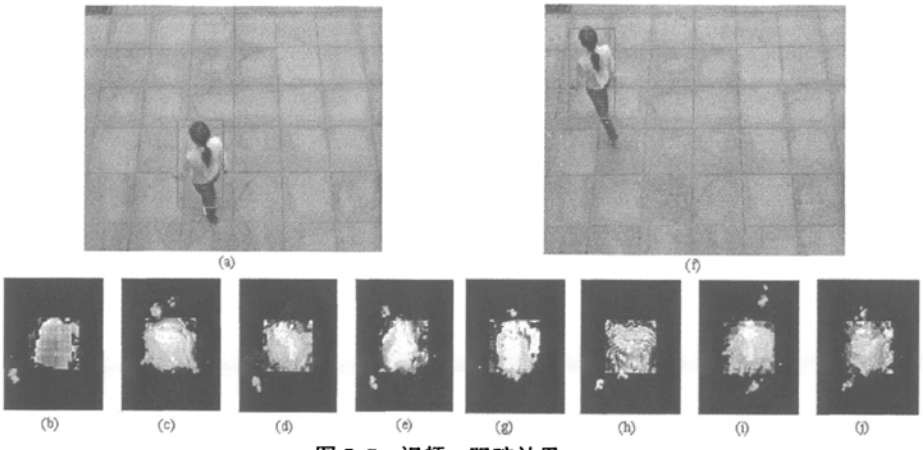


图 5.7 视频一跟踪效果

Fig. 5.7 Tracking results of Video 1

本文与协同跟踪算法[60]和 Comaniciu 等[55]仅利用 RGB 值作为特征的 Mean Shift 跟踪算法在两个自拍的视频上做了比较实验。两者的跟踪结果如图 5.8 和图 5.9 所示。

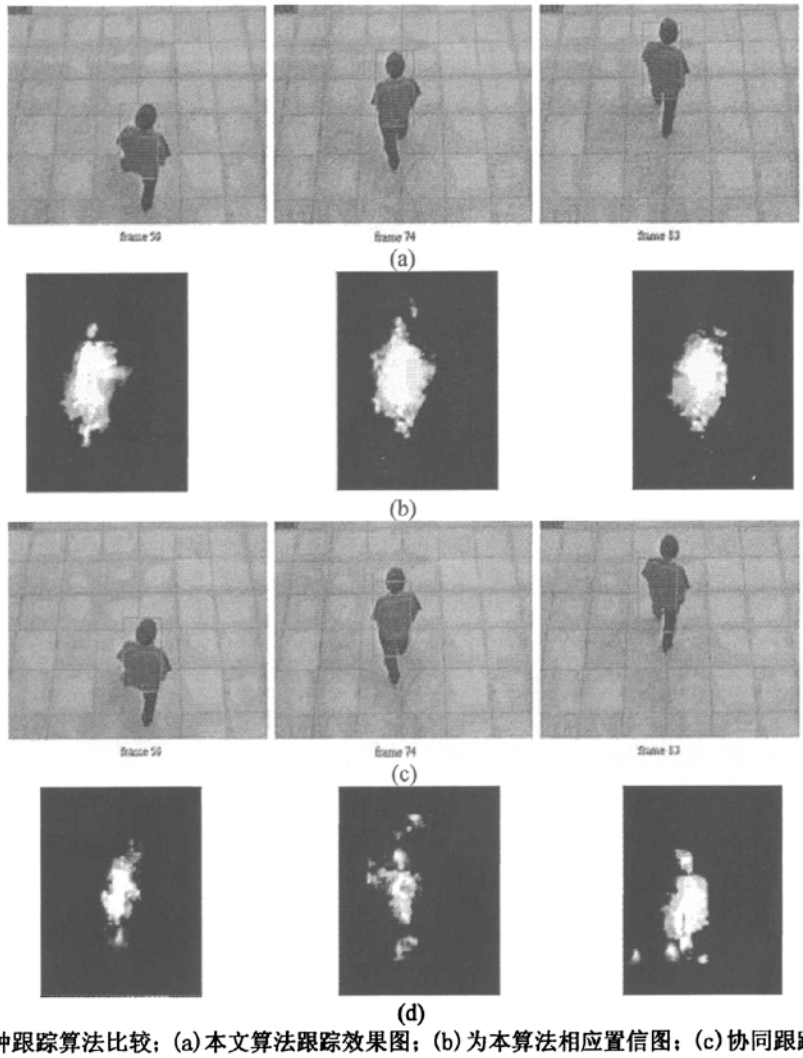
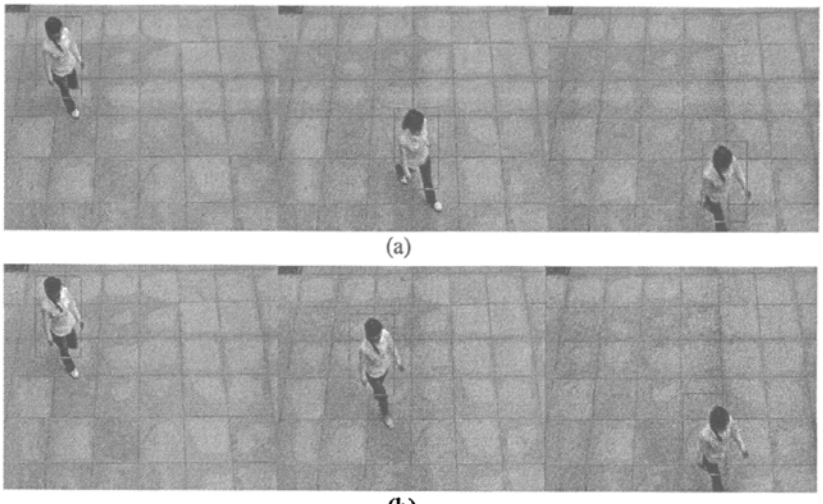


图 5.8　两种跟踪算法比较：(a)本文算法跟踪效果图；(b)为本算法相应置信图；(c)协同跟踪效果；(d)协同跟踪算法置信图

Fig. 5.8　The comparison of two tracking algorithms; (a) The tracking result with our method; (b) Confidence map with our method; (c)The tracking result using Co-tracking method; (d) Confidence map with Co-tracking method



(a)

(b)

图 5.9　与文献[55]跟踪算法比较；（a）本文跟踪结果；（b）Comaniciu 算法跟踪结果

Fig. 5.9　The comparison between our method and tracking algorithm in the paper[55]

(a)The tracking result with our method; (b)The tracking result with tracking algorithm in the paper[55]

5.4.2　实验性能分析

图 5.8 中运动目标的位移较大，并且显著地变化身体姿势，本文算法可以很好地跟踪，但利用协同跟踪算法从第 83 帧开始不能准确地定位目标，逐渐远离目标中心，原因是利用协同跟踪算法产生的置信图在人体目标运动区域范围内很模糊，通过 Mean Shift 算法不能准确定位它，该跟踪器最终导致跟踪失败，然而我们的方法可以稳定可靠地跟踪该目标。图 5.9 中，运动目标速度较快且步幅较大，Comaniciu 等[55]跟踪算法从中间视频帧开始逐渐偏离目标中心，到最后一帧离目标中心已较远，因为它只利用 RGB 颜色特征作为 Mean Shift 的跟踪特征，在利用相似函数度量目标模板和候选模板时，不满足泰勒公式展开的条件，所以核函数会逐渐漂离目标中心，而目标模板又没有得到更新，致使跟踪不稳定，造成跟踪失败，而我们的方法根据目标的外观变化及时更新 SVM 分类器，从而跟踪器能实时地捕捉到这个变化信息，对目标进行鲁棒性跟踪。

本文也对自拍的室内视频进行了跟踪测试，从图 5.10 两个视频跟踪效果图来看，两个视频的背景都相对较复杂，对跟踪造成一定干扰，对一般跟踪算法影响较大，但本文算法也能进行鲁棒性跟踪。

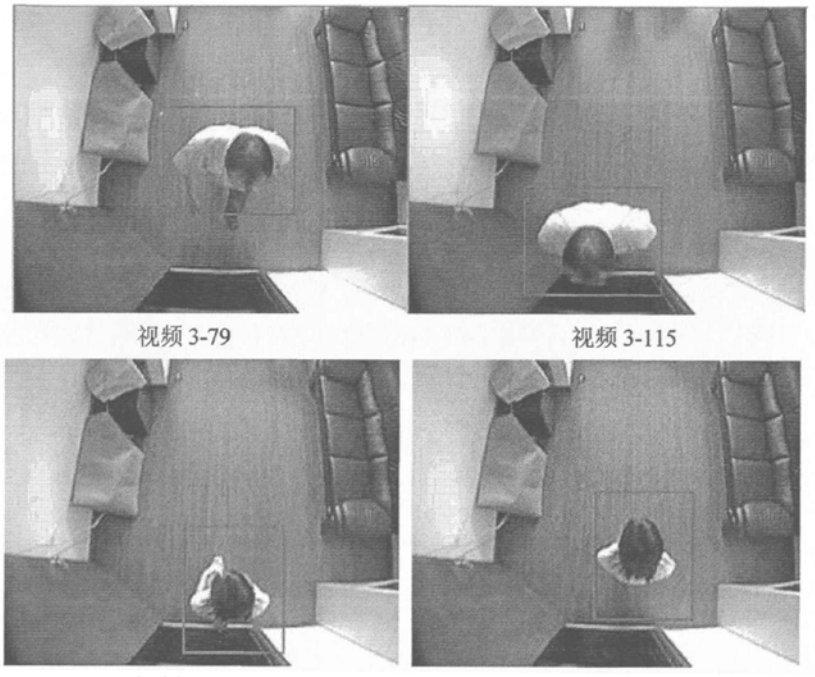


图 5.10 人体目标跟踪图

Fig. 5.10 Human Objects tracking results

5.5 小结

首先分析了建立模型描述目标特征的优缺点，针对传统跟踪算法在外观变化时，存在跟踪漂移的问题，引入了利用半监督学习框架进行学习和跟踪目标的方法，接着介绍了 Tri-tracking 跟踪算法中训练 SVM 的 PPBTF 直方图特征，并对支持向量机 SVM 理论知识进行论述，然后本文对协同跟踪算法进行了改进，把运动目标跟踪问题也看作目标和背景的分类问题，提出了一种基于三个不同的特征向量——颜色直方图、LBP 直方图和 PPBTF 直方图，跟踪单个运动目标的算法，可以对外观不断显著变化的运动目标进行鲁棒性跟踪。由于该算法利用机器学习的方法对目标外观的变化不断学习和更新，随着时间推移不断获得目标和背景信息，达到了较好的跟踪效果。

结　论

智能视频监控是目前国内外计算机视觉研究领域中的热点问题之一。智能视频监控系统不仅符合信息产业的未来发展趋势，而且代表了监控行业的未来发展方向。作为一个有着广泛应用背景的研究领域，视频监控具有很强的任务依赖性，一个特定的应用场合和需求直接决定了视频监视所涉及到的不同的算法实现，因此在实际中，往往是针对不同的应用假设，选用不同的算法。根据这种研究现状，本文在导师的指导下，研究了图像处理技术中的运动目标检测、定位和跟踪技术，分析了算法中存在的问题，并在分析问题的基础上，选择了有效的方案和算法来提高系统性能，最终在 VC6.0 环境下实现了一个具有对人体目标进行检测与跟踪的仿真系统。

本文的研究内容与成果包括：

1) 在运动目标检测方面，研究了较为成熟的目标运动区域检测方法，并对帧间差法、基于梯度的前景检测以及基于帧间二阶差分的前景检测法进行了理论阐述和实验说明，在帧间差和双向投影的基础上，采用了统计均值的自适应方法，较好的把人体的运动区域检测出来。

2) 在多目标定位方面，在垂直固定摄像头的基础上，由于头部运动边缘轮廓具有近圆形的特征，本文采用 Freeman 链码和 RANSAC 算法相结合的方法检测头部，实现人体头部的定位，从而完成目标的定位，较好的解决了多人遮挡问题。

3) 在多人体目标跟踪方面，针对当前 Mean Shift 跟踪算法在跟踪快速、目标与周围背景颜色分布相似、图像对比度大时跟踪易失败的问题，本文对其进行改进，提出了一种融合目标颜色、纹理和运动信息的改进的 Kalman 和 Mean Shift 跟踪算法，对目标的运动建模，利用卡尔曼滤波器根据以往的目标位置信息预测在当前帧的可能位置，然后利用 Mean Shift 在该位置的邻域内找到目标的真实位置，达到了较好的跟踪效果。

4) 在单目标跟踪方面，本文研究了机器学习中的半监督学习方法，针对一般跟踪算法建立模型不能很好处理外观显著变化的问题，提出了一种新的利用半监督学习框架基于颜色直方图特征、LBP 直方图特征和 PPBTF 直方图特征的跟踪算法——Tri-tracking 跟踪算法，针对姿势不断变化的单目标进行鲁棒性地跟踪。

本文在视频序列目标检测、定位和跟踪方面做了一些工作，也取得了一些研究成果，但该领域所涉及的研究内容和应用背景十分广泛，并且时间紧迫，需要考虑的因素有很多，本系统距一个完善的智能视频监视系统还有很大距离。因此，在以后的工作和学习中需要对有关问题做进一步研究和探索：

1) 结合链码和 RANSAC 算法的头部检测算法，对人体的头部检测定位具有良好的

效果，但是 RANSAC 参数选择比较复杂，参数目前不能自适应化，如果选择不当，拟合的圆可能不是最优的，也即定位的头部位置，不太准确。另外，当人体其他部位的轮廓相似于头部运动轮廓时，会造成一定的误检。今后工作的重点是对 RANSAC 算法进行改进，提高该算法鲁棒性同时结合其他方法，排除身体其他部分的干扰，以便能很好的应用在各种监视场合。

2) Mean Shift 算法以核函数加权下的直方图来描述目标，由于加入了纹理信息，当背景和目标的颜色分布较相似时，算法跟踪效果有了很大程度提高，但是当目标颜色分布相似并且目标间距离较近时，即使纹理信息也可能不能很好区分目标，跟踪会出现失败的情况，可以考虑引入更鲁棒性的特征，改善跟踪效果，作者由于时间有限，没有进行相应的工作，希望今后能够有所投入。

3) Tri-tracking 算法的一个缺陷是分类器更新问题，本文每帧更新时选择把新的一帧加到训练集而移除时间最早的帧作为更新策略，但人外观变化规律并非呈线性变化，应该根据外观变化选择更新策略，另一个问题，目前该算法主要针对单目标进行跟踪，以后研究的重点是扩展到多个分散目标的跟踪。由于时间有限，作者在这一方面没有进行具体深入研究，这也是一个以后努力的方向。

参 考 文 献

[1] 谭铁牛. 智能视频监控技术概述[C]. 第一届全国智能视频监控学术会议. 北京, 2002.

[2] Collins R, Lipton A. A system for video surveillance and monitoring VASM final report[M]. CMU-RI-TR-00-12, Robotic Institute Carnegie Mellon University, 2000.

[3] Collins R, Lipton A, Kanade T. Introduction to the special section on video surveillance[C]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2000, 22(8):745-746.

[4] Javed O, Zeeshan R, Alatas O et al. Knight M: A real time surveillance system for multiple overlapping and non-overlapping cameras[J]. The fourth International Conference on Multimedia and Expo, Baltimore, Maryland, 2003.

[5] Haritaoglu I, Harwood D, Davis L. W4:real-time surveillance of people and their activities[C]. IEEE Trans Pattern Analysis and Machine Intelligence, 2000, 22(8):809-830.

[6] Remagnino P, Tan T, Baker K. Multi-agent visual surveillance of dynamic scenes[J]. Image and Vision Computing, 1998, 16(8):529-532.

[7] Hampapur A. S3-R1: The IBM Smart Surveillance System Release 1[J]. IBM T.J Watson Research Center, 2005.

[8] 王震宇, 张可黛, 吴毅等. 基于SVM和AdaBoost的红外目标跟踪[J]. 中国图象图形学报, 2007, 12 (11):2052-2057.

[9] Liang D, Huang Q, Jiang S et al. Mean-Shift Blob Tracking with Adaptive Feature Selection and Scale Adaptation[C]. IEEE International Conference on Image Processing, 2007, 3: 369-372.

[10] 王亮, 胡卫名, 谭铁牛. 人运动的视觉分析综述[J]. 计算机学报, 2002, 3: 225-237

[11] Lipton A, Fujiyoshi H, Patil R. Moving target classification and tracking from real-time video[C]. IEEE Workshop on Applications of Computer Vision, Princeton, NJ, 1998: 8-14.

[12] Anderson C, Bert P, Vander W G. Change detection and tracking using pyramids transformation techniques[C]. Conference on Intelligent Robots and Computer Vision, Cambridge, MA, 1985, 579:72-78.

[13] Meyer D, Denzler J, Niemann H. Model based extraction of articulated objects in image sequences for gait analysis[C]. IEEE International Conference on Image Processing, Santa Barbara, California, 1997:78-81.

[14] Welch G, Bishop G. An introduction to the Kalman filter[M]. In: http://www.cs.unc.edu, UNC-ChapelHill, TR95-041, 2000.

[15] Isard M and Blake A. Condensation-conditional density propagation for visual

tracking[J]. International Journal of Computer Vision, 1998, 29(1):5-28.

[16] Pavlović V, Rehg J, Cham T J et al. A dynamic Bayesian network approach to figure tracking using learned dynamic models[C]. In: Proc IEEE International Conference on Computer Vision, Corfu, Greece, 1999:94-101.

[17] 田径杯. 视频监控系统的若干关键问题的研究和实现:(优秀硕士学位论文).北京:北京邮电大学, 2007.

[18] 卢湖川,张继霞,张明修. 基于Hough变换头部检测与跟踪的方法研究[J].系统仿真学报, 2008, 20, (8):2127-2132

[19] McKenna S J. Tracking groups of people[J].Computer Vision and Image Understanding, 2000, 80(1):42-56.

[20] Karmann K, Brandt A. Moving object recognition using an adaptive background memory[J]. in Capellini, editor, Time-varying Image Processing and Moving Object Recognition. Elsevier, 1990:297-307.

[21] Kilger M. A shadow handler in a video-based real-time traffic monitoring system [C]. IEEE Workshop on Applications of Computer Vision, Palm Springs, CA, 1992:1060-1066.

[22] Stauffer C, Grimson W. Adaptive background mixture models for real-time tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, 1999, 2:246-252.

[23] Barron J, Fleet D, Beauchemin S. Performance of optical flow techniques[J]. International Journal of Computer Vision, 1994, 12(1):42-77.

[24] Xiong Y. and Shafer S. Monent and hypergeometric filters for high precision computation of focus stereo and optical flow[J]. International Journal of computer vision, 1997, 22(1):25-29.

[25] Mae Y, Shirai Y, Miura J et al. Object tracking in cluttered background based on optical flow and edges[C]. 13th International Conferenceon Pattern Recognition, 1996:196-200.

[26] Kiratiratanapruk K, Dubey P, Siddhichai S. A gradient-based foreground detection technique for object tracking in a traffic monitoring System[J]. Proc. of AVSS, 2005:377-381.

[27] 高成英,刘宁,罗笑南. 基于序列图像的实时人流检测与识别算法研究[J]. 计算机研究与发展, 2005, 42 (3) :431-437

[28] 高枝宝. 基于视频的行人流量检测研究:(硕士学位论文).成都:四川大学,2006.

[29] Trivedi M, Cheng S Y, Childers E etal. Occupant posture analysis with stereo and thermal infrared video: algorithms and experimental evaluation[J]. IEEE Trans. Veh. Technol, 2004, 53(6):1968-1712.

[30] Zhao L, Thorpe E. Stereo and neural network-based pedestrian detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2000, 1(3):148-154.

[31] Guil N, Zapata E.L. Lower Order Circle and Ellipse Hough Transform[J].Pattern Recognition, 1997, 30(10):1729-1744.

[32] Li H, Lavin M A, Le Master R.J. Fast Hough Transform: A Hierarchical Approach[J]. CVGIP ,1986, 36(2-3):139-161.

[33] Dov D, Liu W. Stepwise Recovery of Arc Segmentation in Complex Line Environments [J].International Journal on Document Analysis and Recognition, 1998, 1(1):62-71.

[34] Chan T S, Yip R K K.Line Detection Algorithm[C].In:Proceedings of the 13th International Conference on Pattern Recognition,Vienna,1996:126-130.

[35] Cai W, Yu Q, Wang H. A fast contour-based approach to circle and ellipse detection[C].Proceedings of the fifth World Congress on Intelligent Control and Automation, 2004, 5:4689-4690.

[36] 裘镇宇,危辉. 基于Freeman链码的边缘跟踪算法及直线段检测[J].微型电脑应用, 2008, 4(1):17-21.

[37] Fischler M. A, Bolles R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography[J]. CACM, 1981, 24(6):381-395.

[38] Chen Z, Wu C.A. Linear Algorithm with High Accuracy for Estimating Fundamental Matrix[J]. Journal of Software, 2002, 13(4):840-845.

[39] Rousseeuw P.J. Robust Regression and Outlier Detection[M]. New York: Wiley, 1987.

[40] Torr P.H.S, Murray D.W. Outlier Detection and Motion Segmentation[J]. In SPIE 93, Boston, MA,USA, 1993:432-443.

[41] Kalman R E. A new approach to linear filtering and prediction problems[J]. Transactions of the ASME Journal of Basic Engineering, 1960, 82(1):35-45.

[42] Tenney R, Hebbert R, Sandell N J. A tracking filter for maneuvering sources[J].IEEE Trans.On Automation Control, 1977, 22(2):246-251.

[43] Tan T N, Sullivan G D, Baker K D. Model-based localization and recognition of road vehicles[C].International Journal of Computer Vision, 1998,(1):5-25.

[44] Nguyen Q A, Robles-Kelly A, Shen C. Enhanced Kernel-based Tracking For Monochromatic and Thermographic Video[C]. In IEEE International Conference on Advanced Video and Signal Based Surveillance, 2006.

[45] 王永忠,赵春晖,梁彦等. 一种基于纹理特征的红外成像目标跟踪方法[J].光子学报, 2007, 36(11):2163-2167.

[46] 宁纪锋,吴成柯. 一种基于纹理模型的 Mean Shift 目标跟踪算法[J].模式识别与人工智能, 2007, 20(5):612-618.

[47] Ojala T, Pietikinen M, Harwood D. A comparative study of texture measures with classification based on featured distribution[J]. Pattern Recognition, 1996, 29(1):51-59.

[48] Ojala T, Pietikinen M, Menp T. Multiresolution grayscale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7):971-987.

[49] Shan C, Gong S, McOwan P W. Recognizing Facial Expressions at Low Resolution[C]. IEEE Conference on Advanced Video and Signal Based Surveillance, 2005:330-335.

[50] Fukunage K, Hostetler L.D. The estimation of the gradient of a density function with application in pattern recognition[J]. IEEE Trans. Information Theory, 1975, 21(1):32-40.

[51] Cheng Y. Mean Shift mode seeking and clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(8):790-799.

[52] Isard M, Blake A. Condensation-conditional density propogation for visual tracking [J]. International Journal of Computer Vision, 1998, 29(1):55-28.

[53] Comaniciu D, Meer P. Mean shift analysis and applications[C], IEEE Int' l Conf. Computer Vision, Kerkyra, Greece, 1999:1179-1203.

[54] Comaniciu D, Ramesh V, Meer P. Kernal-based object tracking[J]. IEEE Transactions. On Pattern Analysis and Machine Intelligence, 2003, 25(2):564-577.

[55] Peter M. Real-Time Tracking of Non-Rigid Objects using Mean Shift[C]. IEEE CVPR, 2000.

[56] Collins R, Liu Y, Leordeanu M. Online selection of discriminative tracking features[J]. PAMI, 2005, 27(10):1631-1643.

[57] Comaniciu D, Ramesh V. Mean Shift and Optimal Prediction for Efficient Object Tracking[J]. International Conference on Image Processing, 2000:70-73.

[58] Isard M, Blake A. Condensation-conditional density propagation for visual tracking[J]. International Journal of Computer Vision, 1998, 29(1):5-28.

[59] Avidan S. Ensemble tracking[J]. PAMI, 2007, 29(2):261-271.

[60] Tang F, Brennan S, Zhao Q et al. Co-Tracking Using Semi-Supervised Support Vector Machines[C]. IEEE Conference on Computer Vision (ICCV), 2007.

[61] Chapelle O, Schölkopf B, Zien A. Semi-Supervised Learning[M]. Cambridge, MA: MIT Press, 2006.

[62] Shahshahani B, Landgrebe D. The Effect of Unlabeled Samples in Reducing the. Small Sample Size Problem and Mitigating the. Hughes Phenomenon[J]. IEEE Transactions on Geoscience and Remote Sensing, 1994, 32(5):1087-1095.

[63] Miller D J, Uyar H S. A mixture of experts classifier with learning based on both labelled and unlabelled data[M]. In: M. Mozer, M. I. Jordan, T. Petsche, eds. Advances in Neural Information Processing Systems 9, Cambridge, MA: MIT Press, 1997:571-577.

[64] Zhu X, Ghahramani Z, Lafferty J. Semi-supervised learning using Gaussian fields and harmonic functions[C]. In: Proceedings of the 20th International Conference on Machine Learning (ICML' 03), Washington, DC, 2003:912-919.

[65] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the em algorithm[J]. Journal of the Royal Statistical Society, Series B, 1977, 5: 1-38.

[66] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training[C]. In: Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT' 98), Wisconsin, MI, 1998:92-100.

[67] Zhou Z, Li M. Tri-training: Exploiting unlabeled data using three classifiers[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(11):1529-1541.

[68] Joachims T. Transductive inference for text classification using support vector machines[C]. In ICML, 1999: 200-209.

[69] Goldman S, Zhou Y. Enhancing supervised learning with unlabeled data[C]. In: Proceedings of the 17th International Conference on Machine Learning, 2006:5-10.

[70] Dietterich T. G. Ensemble methods in machine learning[C]. In: Proceedings of the 1st International Workshop on Multiple Classifier Systems (MCS' 00), Cagliari, Italy, LNCS 1867, 2000:1-15.

[71] Zeng X Y, Chen Y W, Nakao Z et al. Texture representations based on pattern maps[J]. Signal Processing, 2004, 3:589-599.

[72] 边肇祺,张学工等. 模式识别(第二版)[M]. 北京:清华大学出版社, 2000.

致　谢

论文经过两年的悉心准备，终于落稿。这期间，得到了很多人的大力帮助，在此特别提出感谢。

首先，感谢我的导师卢湖川副教授。卢老师为我提供了一个学习和研究的和谐环境，让我能够不断的更新知识和接触最前沿的研究课题。当我在研究中碰到难题时，老师总是用他渊博的学识和敏锐的眼光为我做细心的指导。老师的支持和鼓励，也是我不断进取的动力。我有今天的成绩离不开卢老师的悉心指导和谆谆教诲，在此由衷地感谢卢老师！

其次要感谢和我一起奋斗的教研室同学，尤其是贾春华同学，还要感谢我们同届的其他几位同学，在我遇到问题时他们也给我许多指点和帮助，同时也非常感谢教研室其他师弟师妹和已经毕业的师兄师姐对我课题工作中给予的帮助，在论文准备过程中，他们都给予了我不少帮助，提出一些宝贵意见。

另外非常感谢孔祥维教授为我们拍摄视频提供摄像仪器，同时也非常感谢郭成安老师，在科研期间提供的投影仪，才能很好地把科研进展情况及时向导师和教研室其他同学汇报，使自己科研水平得到很大程度的提高。此外，由于受日本 Revolsystem 株式会社与导师国际合作项目的启发，论文才得以改进和提高，在此也对之表示谢意！

最后，非常感谢我的家人和我的男朋友，是他们的支持和鼓励让我没有后顾之忧，能够安心的完成毕业论文。也是他们的关怀，让我有信心克服困难，不断进取，最终顺利的完成学业。

向在过去所有日子里，关心、爱护和帮助过我的人们致以最衷心的感谢！

由于本人能力有限，论文中难免出现疏漏之处，敬请各位老师和同学批评指正，本人不胜感激。