摘要

目标跟踪一直以来都是计算机视觉研究的热点问题，在国防军事、智能交通和生活安防等众多领域得到了广泛的应用。近年来，视觉注意机制（显著性检测）的研究取得了实质性的进展和突破。将显著性检测引入到目标跟踪中，可以使目标处理过程更接近于人类认知机制，并提高算法的准确性。因此，将显著性检测应用到目标跟踪中具有重要的意义和应用价值。本文在均值漂移跟踪算法的框架下，通过模拟人类视觉特性，提出了基于（视觉）显著性检测的目标跟踪算法，研究工作主要包括以下两个方面。

1）提出一种利用目标先验信息的时空显著性检测（模型）算法。时间显著性检测方面，引入时域运动特征，利用光流法检测运动矢量，并对其矢量场进行滤波处理。利用运动熵值对运动矢量的显著性进行分析，建立能够表征场景中运动情况的时间显著图。空间显著性检测方面，本文模拟人类视觉特性的“记忆性”，利用初始目标区域的颜色先验信息改进自底向上的（认知视觉注意）显著性检测模型，使得生成的空间显著图能够更好地适应目标跟踪任务。最后，时空信息融合阶段，根据人眼对不同场景的主观感知，提出自适应加权的融合方法，通过设置自适应权重得到时空显著图。

2）针对传统均值漂移跟踪算法采用单一颜色特征导致跟踪鲁棒性不强的问题，提出一种基于视觉显著性特征的目标跟踪算法。首先，利用本文提出的显著性检测模型计算时空显著图，将颜色特征和视觉显著性特征融合，用以描述目标模型，并根据相似系数的大小自适应地调节转移向量融合权值，避免采用单一颜色特征跟踪时易发生的漂移现象。其次，提出核函数带宽的自适应调整策略，以适应目标的尺度变化，并且引入卡尔曼滤波器预测目标的位置，实现了复杂背景下的准确目标跟踪。实验结果表明，本文依据目标先验信息构建的显著性检测算法，检测效果较好，将其应用于均值漂移跟踪框架中，能够有效弥补仅采用颜色特征的不足，解决了目标跟踪过程中候选目标信息描述单一的问题，提高跟踪的鲁棒性和准确性。

关键词：目标跟踪，显著性检测，特征融合，均值漂移

1. 绪论

1.1课题研究背景及意义

计算机视觉是融合了图像处理、计算机图形学、模式识别、人工智能、心理学和数学等领域的一门交叉性很强的学科。计算机视觉研究的目的是利用计算机技术来处理序列图像，以完成对图像中目标的检测、识别、跟踪等任务。近些年来，计算机视觉的研究重点已经逐渐从对静态图像的形态学操作过渡到对动态图像的关联分析，己经成为广大国内外学者研究的热点问题。目标跟踪作为计算机视觉领域的核心研究方向，己经有了几十年的发展历史。目标跟踪就是在序列图片中将目标与背景持续分离的过程［5］，一般利用相关特征为目标建立模板，采用一定的跟踪策略，在图片当中搜索到与目标模板最相似的位置作为跟踪结果。其研究成果在人机交互［１］、视频监控［２］、智能交通［３］、机器人视觉定位导航［４］等多个方面都有着广泛的应用。因而研究目标跟踪技术有着重大的实用价值和广阔的发展前景。

目标跟踪的应用主要有以下几大方面：

1）智能安防监控系统：智能视频监控技术是计算机视觉与人工智能的一个重要研究领域，该技术的目标是用计算机视觉的方法,在不需要人为干预的情况下,通过对摄像机拍录的视频序列进行自动分析来对被监控场景中的变化进行定位、识别和跟踪,并在此基础上分析和判断有关目标的行为,从而既能完成日常管理,又能对异常情况及时做出反应。

2）机器人技术：机器人系统中对外界数字图像信息的处理和分析是机器人进行后续行为分析的基础，机器人跟踪系统主要包括对行为人的跟踪、特定路径的跟踪和固定物体目标的跟踪三个方面。1994年美国斯坦福大学研制的Flakey移动机器人成功完成了对办公区域内行为人的识别与跟踪，随着现在视觉认知模型的发展和计算机技术的提高，越来越多的学术科研人员将人类视觉注意机制应用于机器人视觉系统，使得其能够快速、实时和智能的处理复杂多变的庞大外界信息，并从中找出最相关的部分并做出相应的反应。

3）军事领域：近年来人们将视觉跟踪技术应用于各种安全防卫系统、导弹的精确定位、作战侦查和武器的精确定位控制等方面。在军事作战和环境考察过程中，存在一些极度危险和人类无法亲自到达的地方，此时就需要智能机器人来完成探索任务。在军事领域中的另一热点研究为武器的精确制导，武器中装备能够精确定位目标的系统组件，通过视频传感技术采集窗口信息，并应用视频目标跟踪技术准确定位攻击目标，在这个过程中跟踪的准确性和稳定性就显得尤为重要。

近年来，视觉显著性检测作为计算机视觉领域的热门方向受到广泛关注。所谓显著性，是指一个物体从其周围环境突显出来的特性。人类视觉系统具有高效的数据筛选能力，能够迅速将注意力集中到显著目标上。视觉显著性检测试图模拟视觉系统的这一能力对图像信息进行筛选处理。与一般的图像或视频处理方法不同，显著性检测模型是基于已经被证实的人类视觉对于外界处理的机制建立起来的图像和视频的处理框架。它更加符合人类视觉对于图像和视频的感知，而传统的计算机视觉处理方法大多由任务驱动，由所要完成的任务来决定处理的过程。在图像和视频的处理与分析的应用中，视觉显著性的计算往往作为一个预处理的过程，是图像与视频分割、目标检测以及目标跟踪等任务的重要基础，它为解决这些问题提供了一条新的思路。不仅如此，显著性检测在目标识别、图像视频的自适应压缩、视频检索等领域也有重要应用。高效而准确地对图像和视频的显著区域提取对后续的处理和应用都有着积极的影响。将视觉显著性检测模型引入目标跟踪系统中，通过模拟人类特有的视觉选择性注意机制，可以为目标的搜索提供先验知识，势必会极大提高跟踪的鲁棒性，是一个值得尝试的方向。

1.2国内外研究现状

1.2.1 目标跟踪技术

目标跟踪作为计算机视觉领域最具挑战性的研究课题之一，受到国内外许多研究者的广泛关注。近年来许多学术会议将其作为研究讨论热点，其中包括机器视觉与模式识别领域的著名国际会议ICCV(International Conference on Computer Vision)和CVPR(International Conference on Computer Vision and Pattern Recognition)。在国外许多大学和研究机构一直在进行目标跟踪的研究，如：卡内基·梅隆大学的 Digital Mapping 实验室，康奈尔大学的机器人与计算机视觉实验室等。在国内也有许多团队在做目标跟踪方面的工作，如西安交通大学人工智能与机器人研究所，香港中文大学，上海交通大学图像处理与模式识别研究所等。中国科学院自动化研究所下属的模式识别国家重点实验室，已经在智能交通监控、人脸识别与跟踪、异常行为检测方面取得了巨大研究成果。

视觉跟踪算法众多且各自具有优缺点，广大学者针对不同的跟踪算法进行优化改进，并取得了卓越的效果。视觉跟踪方法可以分为生成式跟踪方法(Generative tracking methods)和判别式跟踪方法(Discriminative tracking methods)两大类。

1）生成式跟踪方法通过学习目标的知识建立目标模型，并将跟踪看作寻找与目标模板最相似的候选区域的过程，其中最具代表性的就是均值漂移(Mean Shift)算法、粒子滤波(Particle Filter)算法和卡尔曼滤波(Kalman Filter)算法。Mean Shift算法是基于密度梯度上升的一种非参数方法，它通过迭代来寻找跟踪目标的位置，从而实现对目标的有效跟踪，它是核函数统计迭代算法应用到目标跟踪的代表。Particle Filter算法是将贝叶斯滤波应用到目标跟踪上的代表，它将目标跟踪过程看作是目标状态的概率密度函数在时间上的传播过程。它的核心思想来源于蒙特卡洛方法(Monte Carlo methods)，将目标状态的后验概率密度通过一组带有权重的离散粒子集来表示，并利用从该后验概率中随机抽取的离散粒子来表示目标状态的分布情况，从而准确定位目标位置，实现对目标的有效跟踪。Kalman Filter算法在目标跟踪中运用滤波原理，通过建立目标的运动模型预测估计目标位置。该算法为一种最优滤波方法，但是其要求系统噪声统计模型为高斯分布。 近些年涌现了许多基于以上算法的改进算法，使其得到不断的完善。例如GR Bradski 在Mean Shift算法的基础上提出的连续自适应Mean Shift算法，即CamShift(Continuously Adaptive Mean Shift)算法，使得跟踪窗口可以自适应地根据目标尺度的变化而变化。该方法提出之后迅速成为目标跟踪领域的研究热点，为许多研究人员开辟了新思路。

2）判别式跟踪方法也称为基于检测的跟踪(Tracking by Detection)，将跟踪问题看作目标与背景的分类问题，是一种在线识别学习算法，通过寻找训练分类器的决策边界，有效分离目标与背景样本。典型方法有Adaboost算法、K近邻(K-Nearest Neighbor，KNN）分类算法、支持向量机(Support Vector Machine，SVM)等。Adaboost算法是判别式跟踪中广泛应用的算法，它将目标跟踪过程看作简单的目标和背景的二分类问题，在对目标进行跟踪时，Adaboost算法通过对一组训练集合中的每个样本赋予不同的分布概率，从而得到不同的训练集合，并且对每一个训练集合进行训练从而产生若干个不同的弱分类器，最后将这些弱分类器通过不同的权值结合起来，就组成了最终的强分类器。K近邻分类算法的决策分类只与其最近邻的几个样本有关，如果一个样本的最近邻样本中的大多数样本是属于某一个类别的，那么判别该样本也是属于这个类别的，其中选择的最近邻样本是已经正确分类的。SVM是一种监督式学习的方法，它将低维空间向量映射到一个更高维的空间中，然后在这个高维空间中建立一个最大间隔超平面，该最大间隔超平面使得两个不同类别的数据间隔最大，并通过这个最大间隔超平面来对目标与背景进行有效的分类。判别式跟踪方法的有效性主要依赖于特征的选择和分类器的有效更新，跟踪器的识别能力与特征空间的选择有直接的关系，所选特征区分背景与目标的能力越强，跟踪分类效果越好，反之亦然。

1.2.2 显著性检测技术

人类视觉系统具有从大量复杂的视觉信息中迅速识别目标的特点，视觉显著性检测能够模拟视觉系统的这一能力对图像信息进行筛选处理，在目标检测、图像分割、视频压缩等方面具有重要的意义，如今已经成为计算机视觉领域的一个新的热点问题。现有的显著性检测模型大体可以分为三类。第一类是基于底层视觉特征的视觉注意模型，目前应用最为广泛的是 Itti 等[20]提出的认知视觉注意模型。该模型提取场景中的特征显著图并采用线性合并的方式整合为最终显著图，并以赢者取全和返回抑制相结合的方式引导视觉注意焦点的选择和转移。第二类是基于纯数学计算的视觉注意模型。比如 Hou 等[22]提出的基于频域的谱残差法(SR)，对图像进行二维傅里叶变换后得到频域的相位谱和幅度谱，认为频谱域上的统计奇异对应图像的异常区域，因此该区域的物体显著性高。四元相位谱傅里叶变换法(PQFT)[23]是基于频域的注意模型，利用图像的相位谱保留各个频率成分的位置信息，并在颜色、亮度等静态初级视觉特征基础上，增加了运动信息通道。第三类是将前两种相结合的方法，最具代表性的是 Schölkopf 等[24]提出的基于图论的算法(GBVS)，底层特征采用场景中颜色、亮度、方向特征，在显著图的生成部分提出引入马尔科夫链的方法来计算显著性。

视频由具有较强相关性的图像序列组成，研究表明，视频条件下动态语义特征的显著性远高于与静态特征的显著性[34]，因此视频的显著性检测模型往往需要加入运动信息。Itti在认知视觉注意模型的框架下引入了运动特征和帧间闪烁[13]，使得改进后的模型能够检测运动信息。Guo[14]等人则将谱残差方法扩展到视频中，对视频帧分别提取运动、红绿对比、蓝橙对比色以及亮度特征，然后通过四元傅里叶变换得到各特征的相位谱，最后将各通道融合成最终的显著图。Lu[15]等人在他们的显著性模型中使用了底层特征如颜色、纹理和运动以及一些认知特征。Cheng[16]等人也在他们的静态图像的显著模型中加入了运动信息，通过分析水平方向和垂直方向的像素运动，计算显著图。Bioman[17]等人提出了视频中的时空维度上不规则性的检测方法，该方法通过视频块的2维和3维纹理与训练数据进行比较，检测得到视频中的不规则的运动。Meur[18]等提出了基于视觉注意机制的时空模型，该方法的时间显著图是通过对仿射分量的分析得到的。Kienzle[19]通过对眼动数据的学习，构建了基于时空兴趣点的检测器对输入信号分别在时间和空间域内滤波，从而检测得到显著目标。

1.3目标跟踪的挑战和难点

运动目标跟踪技术虽然经历了几十年的发展，但各种算法大多都有其局限性，需要满足一定的使用条件，到目前为止，提髙跟踪的鲁棒性依然是运动目标跟踪中的核屯、课题。影响跟踪鲁棒性的因素非常多，化如图像从Ｈ维空间投影到二维空间中发生的信息丢失、图像自身的噪声[9]等。但总的看来，图片中目标与背景的复杂性是影响跟踪鲁捧性的两个最主要因素，详细描述如下：

1）目标自身不确定性

主要包括在跟踪中出现的目标自遮挡和姿态变化。自遮挡，是指目标在跟踪过程中出现的水平旋转现象，当目标发生自遮挡后，目标表观在图片中会发生很大变化，导致与模板差异较大，给跟踪带来难度；姿态变化，是指非刚体目标出现的表观形变等变化，也会导致目标区域与其模板有一定差异，若目标外观模型和模板更新机制不能适应这种变化，就会出现＂漂移＂现象，因此跟踪算法的外观模型和模板更新机制的设计是技术关键。

2）目标所处背景的复杂性

主要包括场景光照变化、目标与背景的相似性、目标被静止的障碍物遮挡等情况。颜色特征是在目标检测与跟踪中经常使用的特征，其对部分遮挡和姿态变化抵抗性较强，但是对光照变化非常敏感，当场景发生光照变化时，会对目标表观颜色造成很大改变，导致模板的不准确；目标跟踪实际是目标与背景的持续分离问题，当目标周围出现相似背景的干扰的时候，会降低这种区分性，可能会造成跟踪结果转移到该相似背景上，出现误跟踪现象；遮挡是跟踪中最常见，也是最难处理的问题之一，如何对遮挡情况进行正确判断以及充分利用目标未被遮挡部分的信息进行跟踪是解决送类问题的关键。

图1.1为实际跟踪过程中遇到的一些挑战，分别为光照变化、目标被静止障碍物遮挡、姿态变化、目标自遮挡等，好的鲁棒性跟踪算法要能应对其中的一两种挑战。

1.4论文的主要研究内容

1.4.1论文的主要内容

本文的研究建立在视觉跟踪算法的基础之上，针对目标跟踪技术的不同环节和存在的技术难题进行了改进和优化，并在VisualStudio2013仿真平台上进行了验证和分析。本文尝试将显著性检测应用到目标跟踪当中，提出了基于视觉显著性特征的目标跟踪算法。首先，对自底向上和自顶向下的视觉显著性检测模型，以及运动显著性的引入进行了研究；在此基础上，针对目标跟踪问题构建利用目标先验信息的时空显著性检测模型；最后，利用提出的显著性检测模型对CamShift跟踪算法进行改进，以视觉显著性特征为注意引导实现目标跟踪。本文的研究内容主要包括以下三个部分。

1）构建空域静态视觉显著性模型。在对“自底向上”显著性算法分析的基础上，通过其在目标跟踪方面这一具体应用，引入基于目标的自顶向下的注意，提出将目标先验信息融入显著性检测算法当中。人眼在跟踪物体时，总能凭借对物体的记忆，通过将注意力转移到具备物体特征的区域来保持对该物体的跟踪。本文构建的空域显著性模型正是以这种现象为依据，通过在基于颜色对比度的显著性检测算法中加入目标先验信息，使具备目标特征的区域在显著图中显著度较高，更符合人类视觉的感知过程。

2）构建时空域融合的视觉显著性模型。运动特征在目标跟踪中占有主导地位，因此在空域静态显著性模型的基础上，引入时域运动特征通道，利用光流法检测运动场，利用运动熵值计算时间显著性，建立起时空域融合的视觉显著性模型。选取 i LIDS database of AVSS 2007 conference 标准库、PETS2006数据库以及实拍行人视频进行实验仿真。结果表明，与现有的多个视觉注意模型相比，本文提出的模型在行人检测中具有较高的准确率。

3）提出基于视觉显著性的行人跟踪算法。首先，采用本文提出的视觉显著性模型生成显著图，利用显著性特征对CamShift跟踪算法中的反向投影图进行融合加权，充分发挥视觉显著性特征对于跟踪过程中出现的光照变化和姿态变化的不变性，从而提高跟踪鲁棒性。最后，在CamShift跟踪框架下，以显著性特征为注意引导，实现目标跟踪。

本文对均值漂移跟踪算法中的反向投影图进行显著性加权，使得跟踪的结果不仅仅在颜色信息上与目标是相似的，而且还具有较高的显著性。实验表明，本文算法在行人跟踪中取得了良好的效果。

1.4.2论文各章节结构安排

本文主要分为五个章节，各章节的内容主要有：

第一章主要介绍本设计的研究背景和意义、研究现状、挑战和难点、本文的主要工作以及本文的结构安排。

第二章介绍信道化技术的基本理论与信号频谱检测方法。整体阐述信道化技术，介绍当前成熟的均匀信道化技术，滤波器组技术同信道化技术二者的关联引出动态信道化技术的概念。阐述了基于能量检测的频谱检测方法，分析了影响改方法性能的窗函数、信号重叠率等因素。并做出了不同窗函数、信号重叠率下检测性能的对比，并给出了相应的信号频谱的仿真验证结果。

第三章给出了一种均匀信道化技术的改进方法与一种动态信道化的实现方法。利用频率响应屏蔽技术设计窄过渡带带宽的优势，提出了一种均匀信道化技术的改进方法，该方法在相同性能下可以有效解决硬件资源。利用调制滤波器组的重构特性，本文给出了基于调制滤波器组的非均匀信道化结构，并以该结构为基础给出了动态信道化的实现方案。并对动态信道化进行了仿真验证。

第四章从工程应用的角度上研究动态信道化方法的实现。介绍新一代SCA架构的软件无线电平台的特点和整体系统结构，针对平台特点，给出了动态信道化方法的在平台上的软硬件结合的实现方案，并重点介绍了硬件FPGA的信号处理流程与上位机软件重构的程序流程。

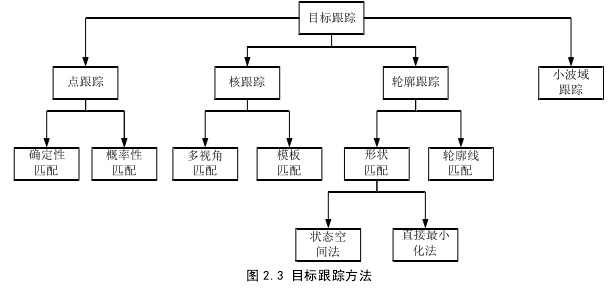
第五章系统的实测与分析。给出了所实现的系统所应用的测试工具、测试环境、测试方法与实际的测试结果。并对测试的结果的误差进行了有效的分析。

最后对本文的整体研究思路进行了总结，并在自己研究设计的基础上对信道化的研究前景进行了展望，以及为下一步的具体工作提出了几点建议，同时对研究生生涯中给予我重大帮助与支持的以及同学和朋友表达我衷心的致谢。

第二章 目标跟踪算法与视觉显著性检测模型

2.1引言

2.2目标跟踪算法



（1）基于点的目标跟踪：

在连续帧中检测的目标用点来表示，点之间的相关性可以用包含目标位置和运动状态信息来表示，该方法需要一个判决机制来对提取的点进行分类、比较。

（2）基于核的目标跟踪：

核函数来表示目标的形状或外观信息，通过计算相邻帧中核的移动来跟踪目标，通过一定的判决准则使概率密度函数移动到极值处，从而获得目标的定位，这种移动通常以一定的参数变化的形式，例如平移、旋转和仿射变换等。

（3）基于轮廓的跟踪：

通过估计每一帧中的目标区域来进行跟踪，它需要对轮廓内的目标信息进行编码，通常以颜色密度或形状模型的形式，那么跟踪的过程即为形状匹配或轮廓变换的过程，且都需要时域的目标分割算法。

（4）小波域目标跟踪：

通过小波变换后的目标含有丰富信息，可以同时利用空间域和频率域的特征来进行匹配和跟踪。它对微弱照明变化、对照物变化和空间变换不敏感，该算法在场景中存在大量噪声、照明变换和目标运动较快的情况下仍然具有较好的跟踪效果。

2.3.1 基于点的目标跟踪

基于点的目标跟踪对旋转、尺度变化和仿射变化具有鲁棒性，对相邻帧中用点来表示的目标的匹配过程却非常复杂，尤其是在目标遮挡、误检测、目标突然进入和走出摄像机等情况下。总之，点匹配方法可分为两大类：确定性方法和统计性（概率性）方法。确定性方法使用定性运动启发式来约束匹配问题，统计性方法则是按照概率估计来进行目标的匹配。

（1）确定性点匹配方法

确定性点匹配方法是使用一系列运动约束来建立第 t-1 帧中的目标到第 t 帧中各个目标之间的关联关系，并通过最小化匹配代价得到最优匹配。最小化匹配代价为一个组合最优的问题，它可以通过最优分配方法来解决，最优分配方法能够很好地解决所有潜在的关联对中一对一的匹配问题。匹配代价通常考虑下列约束之间的关系：

（a）近邻效应：假设相邻两帧的目标定位不会发生显著地变化；

（b）最大速率：定义一个目标速率的上限并且对目标与其周围邻域匹配的有一定的概率限定；

（c）小速率变化（平滑操作）：假定目标的速度和方向不会急剧变化；

（d）常规运动：假定目标在一个小范围邻域内的速率是一样的，这种约束适用于多点表示的目标；

（e）刚体：假定三维世界里的目标是刚体的，因此目标上任意两点之间的距离保持不变；

（f）邻近均匀性：是对邻近效应和小速率变化约束的集合。 必须指出，这些约束不仅适用于确定性方法，而且也适用于统计性方法。Shafique 和Shah[14]提出一种基于多帧方法来计算目标的时域速度和位置的联接性，把匹配问题转化为图论分割问题。多帧匹配需要查找到对每个点的唯一最优路径0{ ,..., }hiP =x x （上标表示帧数），第 k 帧中的点生成定向图，它把节点分为正、负两类，从而变为二值图，‘+’节点表示有直接相连边的节点，‘-’节点表示间接相连边的节点，然后使用贪婪算法来对目标进行估计匹配，最后实现匹配跟踪。近几年，利用 S

IFT 算法来提取点特征并进行匹配跟踪也是一个热门研究。文献[15]对感兴趣区域提取SIFT 点特征，并在相邻帧中进行匹配，同时对变化参数进行估计。Bay 等[16]加速了特征点的选取，使用 SURF 算法，它的速度为SIFT 算法的 3-4 倍。Mondragon 等[17]使用 SIFT 和RANSAC 算法来提取感兴趣的特征点和估计仿射变换，并对特征点进行筛选，丢弃不属于目标的特征点。文献[18]对检测到的目标提取感兴趣的SIFT 特征点来进行识别跟踪。Zhao 等[19]利用特征点分析来恢复出仿射参数，并通过仿射参数估计出两帧之间尺度的变化。此外，点特征与其它特征相结合的方法，如颜色特征等[20]，也是近年来研究的热点。

（2）统计性匹配

目标跟踪系统中含有大量的噪声，而且目标的运动也受随机扰动的影响，统计性匹配方法考虑到目标模型的健壮性，通过一定的参数和目标估计过程中模型的不确定性来解决这个问题。统计性匹配使用状态空间来对目标特征如位置、速度和加速度等进行建模。首先，通过一定的检测机制来获得目标在图片中的位置，然后通过对特征点的状态空间进行估计来进行目标跟踪。考虑场景中的一个移动目标，用tX （t=1,2,…）来表示目标的状态信息，状态预测方程如下所示：

其中，tW （t=1,2,…）为白噪声，tf 为状态转移的函数。

测量值和待估参数之间的关系通过测量方程可以得出： ( , )t t t tZ =h X N （2.2）

其中tN 是白噪声，它独立于tW ，th 为观测测量的函数。

跟踪的目的是利用当前时刻所有测量值和先验概率密度函数1,..., 1( / )t tp X Z−相结合，使用递归贝叶斯滤波对状态量tX 作出最优估计。

统计性匹配最常见的方法为粒子滤波（Particle Filter）算法[21]和卡尔曼滤波（Kalman Filter）算法[22]。粒子滤波算法使用序列蒙特卡洛和修正的贝叶斯算法来计算状态概率，它在高斯噪声和非高斯噪声下效果都很好，而卡尔曼滤波在高斯白噪声下能够取得较好的效果。后续章节我们将详细介绍这两种方法。

2.3.2 基于核的目标跟踪

基于核的目标跟踪通过计算以核函数表示的候选目标模板来进行跟踪，通常目标是以简单的几何形状如椭圆、矩形或圆等框定出来。通常，目标模板以一定的参数模型表示，它对一定的变换如旋转、保角映射和仿射变换等具有鲁棒性。核跟踪是当今特别流行的一种跟踪方法，因为它对不确定的空间信息具有鲁棒性并且具有广泛的收敛性。它按照模板模型和外观模型进一步进行分类。

模板模型使用一定的相似测量函数来对目标模板和目标候选模型进行匹配。差值平方和算法（Difference Square Sum Algorithm,简写为 SD）、归一化互相关算法和巴氏系数是比较常用的相似测量方法[23]。由于目标模板比较简单，所以得到了广泛的应用。VSAM 系统使用归一化互相关算法来对目标进行跟踪[24]。W4 算法中，也使用互相关函数来进行人体跟踪。文献[25]使用SSD 相似函数来进行跟踪，均值漂移（Mean Shift）算法[26]使用巴氏系数作为测量似性的准则。

其中，Mean Shift 算法也是目前比较流行的算法之一，首先它使用 Epanechnilov 核加权的颜色直方图来表示目标，其中离目标近的像素点权值大，反之，亦然。然后使用 Mean Shift迭代使巴氏系数到达最大值，进而估计出当前帧中目标的定位。尽管计算起来简单方便，但是当目标颜色和背景颜色相似或者目标之间的颜色相似时，该算法失效。另外，Mean Shift易陷入局部最大值，因此文献[27]使用多带宽方法来使 Mean Shift 迭代从任何起始点都能到达密度函数最大的全局模式。对于遮挡问题，Adam 等[28]提出了基于块的跟踪，效果改善的同时也大大增加了计算量和存储空间。文献[29]对 Mean Shift 算法进行改进，目标外观模型发生变化时，它使用目标检测机制来快速更新目标模板，为提高鲁棒性，目标被分为几个块，然后用 Mean Shift 算法来找到每块最匹配的区域，核函数融合了背景信息、边界信息和颜色直方图多种特征，并解决了尺度变化的问题，最后综合这些信息来定位目标。但是，此文并没有指出如何获得目标的最优估计，而且内部特征的选择也会导致跟踪漂移和一些有效目标块的权重下降。文献[30]融合了空间信息到 Mean Shift 算法中，并提出一种新的融合算法来降低计算量，但是，在目标遮挡情况下，容易丢失目标。文献[31]使用基于多种模型的 Mean Shift算法，目标表示中使用多个连通区域来包含目标的空间信息，当目标外观模型发生变化时，通过贝叶斯概率准则来自动进行模型转换。

另一方面，许多算法使用多个特征进行跟踪来提高可行性和跟踪效果，文献[32]中，作者融合形状-纹理和颜色特征到 Mean Shift 跟踪框架中，使用一个定向直方图来表示形状-纹理特征。Ning 等[33]结合局部二值模式（LBP）纹理特征到 Mean Shift 颜色直方图中。文献[34]中，几何特征用于实时车辆跟踪中，纹理模式产生目标的空间结构，然而当目标有大量平滑区域时，纹理特征失效。Hu 等[35]对每个跟踪目标提取三个直方图，分别为头部，躯干和腿部的直方图，这样同时使用颜色和颜色的分布来进行匹配跟踪。Shen 等[36]推广了目标定位和跟踪的统计学习算法，相对于使用一帧图像来建立目标模板，他统计了大量样本中的目标模型，并使用基于支持向量机（Support Vector Machine，简写为SVM）的核函数来判别。支持向量机是一种监督式学习的方法，它广泛用于统计分类及回归分析中，它能够同时最小化经验误差与最大化几何边缘区。在图像区域中通过训练背景，最大化 SVM 分类值来进行跟踪，因此对于复杂背景下的跟踪效果很好。

在设计外观模型来跟踪时，跟踪器需要对目标外观特征（如姿势，照明等）的变化具有一定的鲁棒性。近年来，提出许多能够自适应外观模型的跟踪方法，取得了较好的效果。Ross等[37]采用递增子空间算法来更新外观模型，他把目标表示为一个包含外观模型主成分分量的低维子空间，并通过后继帧中跟踪的目标块进行自适应模型更新，然后使用粒子滤波和仿射变换算法来进行跟踪，尽管该算法能够在不同照明条件下对目标进行跟踪，但是子空间更新算法并没有考虑遮挡问题，且对纹理丰富的目标效果才比较好。文献[38]提出基于自适应外观模型的跟踪算法，对每个像素的亮度值进行混合高斯建模，使用序列核密度估计自动更新参数。

2.3.3 基于轮廓的跟踪

人体目标有复杂的形状，如手，头，肩膀和腿等，因而只用简单的几何形状无法有效的描述目标特征，基于轮廓的目标跟踪对目标模型进行精确描述，通过一个封闭的轮廓线精确分割出目标，然后特征化得到目标模型，可以是颜色直方图、目标边界或目标轮廓线。一般来讲，跟踪过程基于两种匹配机制：形状匹配和轮廓线匹配。形状匹配是在当前帧中搜索出与目标模板相似的形状。轮廓线匹配则使用状态空间或最小化能量函数来找出初始轮廓在当前帧中的位置。

（1）形状匹配

形状匹配跟踪类似于基于模板的匹配跟踪，通过计算候选的目标形状与目标模板形状之间的相似度来进行搜索跟踪，对刚体目标的跟踪具有较好的跟踪效果。但是，近几年研究侧于跟踪非刚体目标，如人。Paragios 等[39]使用测地线主动轮廓方法和水平集配方进行目标跟踪。Yilmaz 等[40]针对遮挡问题提出基于主动轮廓的目标跟踪。Sato 等[41]通过对速度空间进行霍夫变换产生目标轨迹，进而获得连续帧中目标的轮廓。

（2）轮廓线匹配

基于轮廓线的跟踪是一个迭代算法，一般通过状态空间模型或最小化能量函数来获得目标的轮廓线。

（a）状态空间模型：目标的状态空间是以轮廓线的形状或运动参数来表示，通过最大化

后验概率来不断更新状态，后验概率依赖于先验概率和当前观测值之间的相似性。Chen 等[42]提出基于轮廓的跟踪器，其中每个轮廓都有一个相关的隐马尔科夫模型（Hidden Markov Model，简写为 HMM），而且每个HMM 的状态通过位于轮廓线上的点集来确定。

（b）最小化能量函数模型：包括最小化图片能量或者形状能量，图片能量包括颜色和纹理能量，能量最小化是通过水平集分割算法来实现。

2.3.4 小波域的目标跟踪

上述讨论的算法很多使用空间域或时间域的特征，因而具有一定的局限性，很难处理大量噪声、照明变化、外观变化和摄像机抖动等情况。近些年来，基于小波特征跟踪技术迅速发展，由于其具有以下特征可以为目标提供了丰富和健壮的表示模型。

（a）小波变换是空间和频率的局部变换，因而能有效地从信号中提取信息；

（b）为图像的多层表示和存储提供有效的框架；

（c）具有一定的抗噪性；

（d）高频子带和小波变换表示边界信息；

（e）小波变换也是下一代压缩技术的核心。

离散小波变换（DWT）具有把图像解压缩为四个不同频域系数后仍保持空间信息的特征，高频子图像（水平系数，垂直系数和对角系数）含有丰富的细节信息。文献[43]中首先利用小波变换检测出移动目标的边缘，然后获得该目标的关键点进行匹配跟踪，该算法可以降低边界模糊和噪声干扰对跟踪结果的干扰。文献[44]首先使用帧间差分算法来找到车辆目标，然后使用基于小波的神经网络来识别移动车辆，虽然该算法使用小波变换（WT）压缩图片为低分辨率来降低计算量，但是识别过程中计算量依然很大。文献[45]采用基于 Daubechies 复小波变换的目标跟踪方法，复杂小波域采用结构相似性索引，它对微弱照明变化、对照物变化和空间变换不敏感，初始帧的目标依据 Daubechies 复小波变换进行特征矢量建模。该算法在场景中存在大量噪声、照明变换和目标运动较快的情况下仍然具有较好的跟踪效果。但是计算量大难以达到实时跟踪。所以，降低计算量以实现实时跟踪是该算法的研究方向。

2.3视觉显著性理论及其算法介绍

2.3.1视觉显著性概述

视觉感知是人类了解世界最重要的方式之一。当我们挣开双眼，各种海量信息便蜂拥而至。进入人类视野的信息如此复杂多样，但神奇的是，人类总能迅速把注意力集中到自己感兴趣的目标上。这一切主要得益于人类视觉系统在处理外界信息时表现出来的卓越的数据筛选能力。由于人类视觉系统的资源是有限的，当面对涌入的视觉信息时，它会有选择地筛选出它所认为的重要信息，进而对这些信息进行处理，而其它信息则被视觉系统过滤掉。这种具有选择性、主动性和记忆性的心理现象被称为视觉注意机制。当前一种广泛被学界接受的说法是：人类的视觉注意力是由显著性机制驱动的。显著性是由多种视觉敏感特征引起的一种局部反差，一般表现为“信息中最为特殊的部分”或“最为突出和值得关注的部分”。

计算机视觉领域的显著性检测指的是利用图像处理的方法模拟人类视觉注意系统的机能对视场中信息的重要程度进行计算。通过采用相应的显著性检测算法对输入图像进行处理得到视觉显著图。显著图一般是一幅单通道的灰度图，在显著图中，灰度值越高的地方越显著，灰度值较高的区域被认为是显著性区域。近十多年来，随着视觉显著性研究的活跃，国内外研究人员提出了各种各样的视觉显著性检测模型，一是特征驱动的“自底向上”模型，二是任务驱动的“自顶向下”模型。其中，“自底向上”的视觉显著性检测模型是由场景底层数据驱动的，通过提取底层视觉特征来计算场景的视觉显著性。在自底向上的方式中，吸引注意的区域是那些和周围特征相比足够有判别力的区域。“自顶向下”的视觉显著性计算模型是指考虑了自顶向下因素的显著性计算模型。这些模型大多是基于具体任务，如视觉搜索、玩游戏、汽车驾驶等。由于自顶向下因素的多样性和复杂性，因此建模自顶向下的注意（计算显著性）是一项较为困难的工作。目前针对该模型的研究工作相对较少，下面对“自底向上”和“自顶向下”两种情况中一些有代表性的模型进行介绍。有关这部分的详细内容见三篇视觉显著性研究的综述性文献。

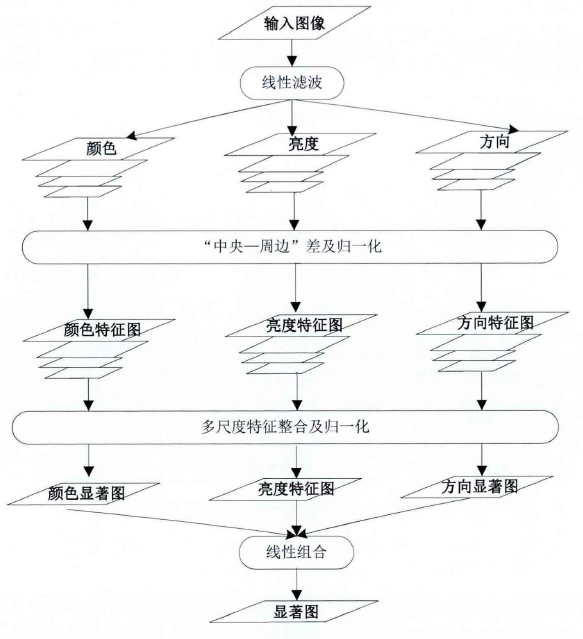
视觉显著性在人类的目标跟踪机制中起到十分重要的作用，许多专家学者将视觉显著性与目标跟踪相结合，提出了目标跟踪与视觉显著机制的假设学说[4]，即目标跟踪是通过判别目标背景区域中的显著性区域，在任务或数据驱动下完成基于目标特征调节的认知机制。目标跟踪的视觉显著性假设理解主要包含三方面的内容：(1)在跟踪过程中，视觉显著区域的贡献值大于非显著区域对目标跟踪结果的准确性贡献值；(2)跟踪结果的有效性很大程度上依赖于显著性检测过程中视觉显著特征的选择过程；(3)人类的视觉注意机制和目标跟踪机制都可由低级的生物神经机制来完成实现。该机制是视觉显著跟踪的基础性假设，并得到广大学者的论证。

“自顶向下”显著性检测算法主要跟特定任务相关，如在对特定目标进行检测的过程中，可将该目标的特征模板当作先验信息，融入显著性检测算法当中，提高生成显著图的质量。

2.3.2 Itti模型

Itti等提出了一种基于生物视觉系统和心理学认知理论的视觉注意模型，该模型为典型的自下而上注意机制，能够快速选择感兴趣的目标位置区域。基于特征集成理论模拟人类视觉搜索机制，通过计算逐个像素点在不同特征和尺度下的显著度，最终组合得到图像的视觉显著图。随后许多科研学者在此经典模型的基础上结合图像的多维特征进行了改进和完善。

输入一帧图像，通过线性滤波操作提取图像初级颜色、强度和方向特征图，根据中央周围区分度和归一化原则，获取多级尺度不同特征的注意图，线性组合多级尺度不同特征的关注图，得到最终的视觉显著图，而后根据生物学中的竞争机制得到视觉注意焦点位置，最后根据返回抑制机制完成注意焦点转移。Itti 视觉注意模型的具体实施过程如图 2.3 所示。



对于一幅图像，分别为其三个单通道像素值，由于色调在低亮度下的变化不明显，因此将三通道图像扩展为四幅单通道图像，以提高颜色特征对目标的表达能力，计算过程如下：









视觉显著性体现在前景区域与其周围背景的差异上，而图像空间尺度的不同往往能体现出这种差异，因此利用不同空间尺度图像相减来表示图像中心与周围的差异。

令尺度参数，，，这样可以得到6幅亮度特征在不同尺度空间下的视觉激励图，计算公式为：



颜色特征在不同尺度空间下共有12幅视觉激励图，计算公式为：





其中，运算为不同图像之间相同坐标像素值的相减动作，体现了不同空间尺度下图像中心与四周边缘的差异性。

采用Gabor滤波方式提取方向特征，得到24幅方向刺激图，计算公式如下：



其中，表示Gabor滤波，表示方向，。

通过上述处理，共得到48幅视觉激励图，采用线性插值的方式将不同空间尺度的三种激励图分别归一化到相同尺度，并进行相加处理，计算过程如下：







其中，是一种对图像的局部非线性迭代操作，具体计算过程见[35]。

经过以上计算，分别得到了亮度、颜色，方向三种特征的视觉显著图，通过求平均值并相加，得到图像的视觉显著图:



2.3.2 GBVS模型

乔纳森哈雷尔等人在Itti模型的基础上提出的一种基于图论的显著性分析算法——GBVS算法,利用拓扑结构、图论的平行性、Markov链的平稳分布性来达到自然有效的显著输出。通过特异性、显著性来定义边的权重，计算出该点的活动显著值。GBVS计算原理概括如下：

获取多尺度的亮度信息：对输入的灰度图像，首先对其使用高斯金字塔低通滤波器进行滤波，高斯金字塔的每一阶都是一个二维高斯低通滤波器：



式中，为图像中心点的位置，为尺度因子，其值越小，图像平滑的范围越小。金字塔指的是对原始图像进行不断的降釆样和高斯低通滤波，随图像的变小而不断减小，每一阶的输入都是上一阶滤波后再降采样后的结果。最终得到一组不同尺度下的滤波结果，将这一组滤波结果记为，表示亮度通道。

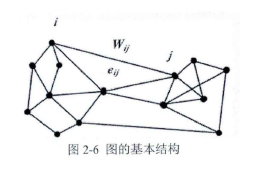
获取多尺度的方位信息：用Gabor金字塔滤波器组对原始灰度图像进行滤波，以得到方向上的信息。二维Gabor滤波器表示如下：



其中，为尺度因子，为频率，为方向，通常取，即在四个方向上滤波。这样，可以得到四组不同尺度下的滤波结果，记为，表示方向通道。

求不同尺度和不同特征图的马尔可夫平衡：对上述五组结果的每一组中的每一个尺度的滤波结果（记为），求其马尔可夫平衡分布。首先我们定义两个像素点和的差异性如下：





根据上式，我们可以用图像中任意两点计算差异性，构建成如图（）的全连接的图，图的每个顶点代表—个像素，每条边代表两个点之间的权重，点到点的权重为：



也就是说权重是由点间灰度差异和点间距离共同决定的。不难看出，任意两个相异点都有两条边：出边和入边，且权重相同。接着求图的平衡分布，代表着熵在图中不断流动，最终达到一种平衡态，与周围差异性大的点会得到更多的熵。求平衡态的过程就是求图所对应权矩阵的最大特征值对应的特征向量的过程。

归一化：目的是对不同特征图的平衡分布结果再一次进行马尔科夫计算，以进一步突出极值响应点。这时权矩阵由下式定义：



其中为平衡分布的像素值。

计算显著图：将所有组所有尺度的滤波结果依次计算出平衡分布后，将结果按照通道叠加起来并归一化。在同一通道内，小尺度的图像扩大后叠加到大尺度图像上，依次类推，亮度通道可以得到一张特征图，方向通道可以得到四张子方向特征图，合成另一张特征图。最后，再将这两张图线性相加并归一化得到最终的显著图，其大小与原始图像一致。

GBVS模型在复杂场景下有着很好的显著性分析效果，其马尔可夫链的建立与求解也类似于模型Itti中的中心周边差异步骤。但该方法由于马尔可夫过程的

计算消耗随图像矩阵的增大而大幅上升，所以该方法在计算速度上存在明显不足。

2.3.3 谱残差模型

谱残差(Spectral Residual, SR)模型从信息理论角度出发，认为信息可分为冗余部分与变化部分。人们的视觉对变化部分更敏感。视觉系统的一个基本原则就是抑制对频繁出现的特征的响应，同时对非常规的特征保持敏感。于是将图像分为如下两部分：



该研究者对图像的log频谱发下了如下规律（此处的g是自然对数）：



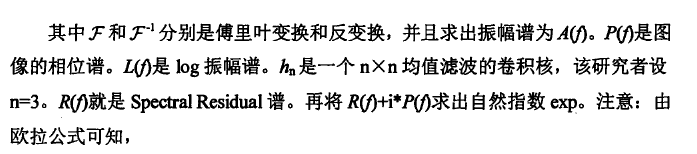
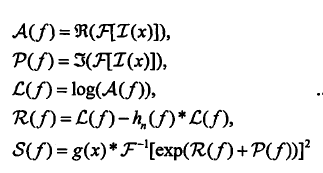
上式表示大量图像的log频谱的平均值是与频率倒数呈现正比关系的。该研究者根据大量试验提出大量图像的振幅谱都趋近一条直线，因此一副图像的log振幅谱减去平均振幅谱就是显著性部分。该研究者定义谱残差为：



式中L(f)是图像f的log振幅谱，A(f)是平均log振幅谱，作者就是将L(f)进

行3×3的均值滤波得到的A(f)。R(f)就是图像f的谱残差。总得来说,一幅图像

I(x)，我们有以下计算步骤：





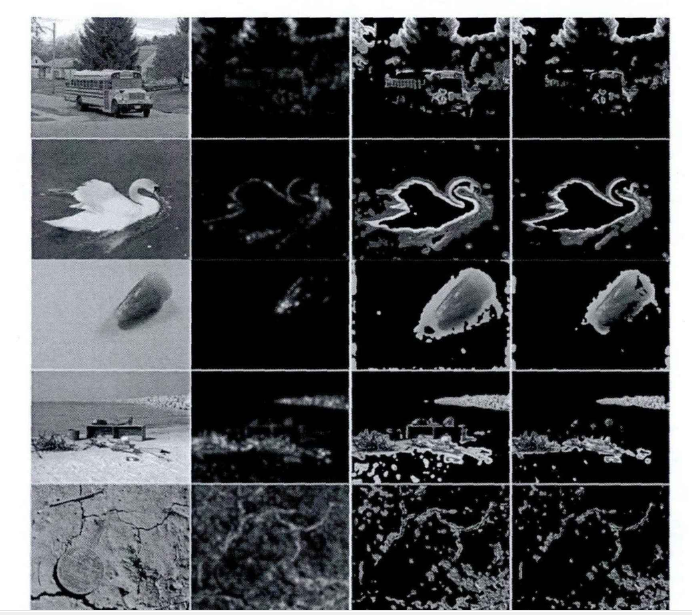
其中e是相位谱，e的正余弦也可由傅里叶的实部和虚部求出：



然后对其进行傅里叶反变换，再进行一个高斯模糊滤波就得到了显著性区域。根据以上原理用rmtlab2014a编程试验，分析600张图片（图片来源MSRASalientObjectDatabase)选出8张比较有代表性的图片效果如图所示。

图3.2中由输入图像(a)通过SR获得显著区域强度图(b)，根据图(b)所示显著区域强度图提取原图75分位显著区域图像(c)，根据图(b)所示显著区域强度图提取原图85分位显著区域图像(d)。

此算法优点明显，算法实时性优秀，400\*300像素的单幅图像处理的平均耗时76毫秒。但是其缺点也明显（可能由于本文使用的区域分割算法与该显著图不匹配而造成），根据(C)(D)容易看出此方法的几个明显缺点：第一，显著区域成散点状，无法用这样的区域为目标跟踪提供目标坐标；第二，显著区域图没有包括完整的目标園像特征（目标图像内部有空洞），不能为后续图像处理提供特征信息数据；第三，显著目标提取错误，例如样本图一一土地上一枚硬币的样本图中显著区域提取的是泥土的沟壑而非目标——硬币



2.3.4 RC 模型

Cheng等人提出了基于区域颜色对比度的显著性检测模型，该模型从人类视觉系统对颜色对比度的敏感度出发，认为在显著图中显著度较高的像素点或区域应与其周边区域有较大的颜色差异。并加入了特征的空间关系，认为相比于较远区域的高对比度，近邻区域的高对比度比能为显著性带来更大的贡献。

首先，为了提高计算效率，减少计算量，将RGB三个颜色分量全部重新量化为12个级别，这样就将颜色数量缩减为种颜色。由于一副图像中的颜色范围一般有限，只选择出现频率占95%的颜色，剩余的5%颜色的像素合并到直方图中距离最近的颜色中。

RC模型采用基于图的分割方法将图像分成小区域，保证这些区域在空间和颜色上具有一定的同质性和完整。对于区域，进行区域级的颜色对比度计算，区域和的颜色对比度可表示如下式：





其中，为颜色在区域的所有种颜色中出现的概率，。为区域中的颜色和区域中的颜色在颜色空间的距离。



对于区域，通过计算它与其它区域的空间距离加权颜色对比度得到该区域的显著性值。定义为下式：



（8）



其中，表示区域的权重，可以用区域包含的像素数计算；为区域和几何中心的距离；为权值控制系数，值越大，空间距离对显著性的影响越小，距离较近的区域对显著性的贡献变大。



RC模型是目前效果最好的自底向上的显著性检测模型之一。将图像按照局部的特征分割成区域，不仅可以较好地克服噪声的影响，而且区域级的显著性计算使得计算复杂度与时间成本极大的降低。

2.4本章小结

本章主要介绍了当下主流视觉跟踪算法的原理和优缺点，通过分析给出了本文的研究对象和研究目标。为后续跟踪算法的实现和性能优化奠定了基础。然后阐述了不同视觉模型的显著区域检测算法，模拟人类的神经视觉认知系统来处理庞大的图像信息。视觉显著区域对比度较高，其对应的视觉特征突出，如能将其运用于目标跟踪过程中，将能够大大改善跟踪的准确性和鲁棒性。

第三章 时空信息融合的显著性检测模型

3.1 显著性的相关概念

显著性检测：模仿视觉系统自动搜索感兴趣区域的行为，使用图像处理的方法将图像中最易引起注意的区域提取出来，显著性区域检测可以计算和衡量图像中各个位置吸引注意的可能性，对于图像的分析处理以及计算资源的分配具有重要的意义。

显著图：显著性检测的结果以显著图进行表示，显著图是根据图像的视觉特征显著度生成的二维灰度图像，每个像素点的灰度值表征该点显著度的大小，其中，显著区域的像素值较大，而背景区域的像素值较小。

显著性特征：所谓视觉显著性特征，是一种能直观表现出目标区域与背景区域差异的特征，具体在显著图中体现为目标区域显著度较高，而背景区域显著度较低。显著性特征不需要建立将征模板，从而避免了跟踪过程中经常出现的模板漂移现象。

视频显著性：动态视频序列比静态图像增加了一维时间信息，也就引入了时间维的运动特征，并且运动特征对视觉注意力分配的影响较大。由于视频序列内部具有较强的时间关联性，将每个视频帧单独考虑计算显著性显然是不合理的。因此，视频显著性包含图像固有底层特征所造成的空间显著性，以及时域上的运动信息所造成的时间显著性。其中，空间显著性指的是视频帧的颜色、亮度、纹理等底层信息对视觉注意力的吸引程度，它是视频帧在某一个时刻所表现出显著特性；而时间显著性由视频中前景目标的运动造成，关注的是目标与背景之间的运动差异度，通常我们也称之为运动显著性。相应的，视频显著性特征包括空间特征和时间特征，即静态特征和运动特征，同时，也包括一些容易被人认知、对人类视觉注意产生影响的其它特征。

3.2 （时空信息融合显著性检测）模型框架

当前出现了很多显著性检测算法，归纳起来主要有两种，分别是基于底层特征驱动的“自底向上”(bottom-up)显著性检测算法［６］和基于顶层任务驱动的“自顶向下”(top-down)显著性检测算法［７］，它们各有优劣，互为补充。其中，“自底向上”显著性检测算法，主要是利用目标与周围背景的颜色、尺度等方面不同，提取图像的各种特征，分别生成不同的特征显著图，然后采用相应的图像融合方法，生成最终显著图。“自顶向下”显著性检测算法主要跟特定任务相关，如在对特定目标进行检测的过程中，可以将该目标的特征模板当作先验信息，融入显著性检测算法当中，提高生成显著图的质量。

人类视觉系统具有从海量场景信息中迅速识别目标的特点，因此，视觉显著性对于目标跟踪和识别等任务具有重大的指导意义，近年来，显著性检测在目标跟踪方面的研究引起广泛重视，并且取得了实质性的进展和突破[16]。

目前在目标跟踪方面应用的大多是基于自底向上的显著性检测模型，在简单场景且前景目标较为突出的情况下，自底向上的模型能够取得较好的检测效果。然而实际跟踪任务中，目标所处的环境往往很复杂，当目标的尺度较小且位于图片边缘位置时，可能导致误检和漏检。因此将自顶向下的视觉注意引入显著性检测模型，以增强目标的显著度十分必要。

基于以上理论，本文提出了一种适用于目标跟踪任务的显著性检测模型，该模型基于时空信息自适应融合的架构。空域上，引入基于任务的“自顶向下”的思想，采用上一帧目标模板信息对“自底向上”的空域显著性检测模型进行改进，利用目标位置先验对视觉注意焦点进行引导，以提取空域显著特征；时域上，对于视频监控下的目标跟踪任务，动态语义特征的显著性远高于与静态特征显著性[34]，运动信息对于人类视觉注意的引导也具有十分重要的作用。因此添加运动特征通道作为视觉注意引导的方式，利用运动信息构造时域显著模型；时空显著信息融合阶段，采用自适应加权的融合方式，以适应不同背景情形的目标跟踪任务。

在上一章中论文学习了目标跟踪和视觉显著性检测的相关理论，并给出了视觉显著性的基本概念，分析了目标跟踪任务中的显著性检测模型所需要具备的性质，并介绍了目前主流的显著性检测方法。

本章将进一步对显著性检测技术进行研究，研究内容主要分为三个方面：首先对目前公认性能较优的基于颜色对比度的空域显著性检测模型进行合理的改进；二是以人类视觉对运动信息的感知为出发点，提出一种时域显著性计算方法；最后采用自适应加权的融合方式，构造时空显著性检测模型，以适应不同背景情形的目标跟踪任务。

将运动特征引入本文空域静态显著性模型，构造适用于目标跟踪的时空信息融合显著性检测模型，总体流程如图A所示。

对于视频监控下的目标检测，动态语义特征的显著性远高于与静态特征显著性[34]。运动信息对于人类视觉注意的引导也具有十分重要的作用。因此，除了基于空域静态特征的视觉注意模型，基于时域动态的视觉注意模型也得到了研究与发展。

针对当前一些“自底向上”显著性检测算法得到的显著图中，目标区域显著度不高的问题，提出一种利用目标先验信息的显著性检测算法。引入基于任务的“自顶向下”的思想，利用第一帧目标模板信息对“自底向上”的显著性检测模型进行改进，将目标信息模板的颜色特征作为高级语义特征对视觉注意焦点进行引导，以提取空间显著特征；

利用本章提出的显著性特征对跟踪算法中的反向投影图进行加权，使得跟踪的结果不仅仅在颜色信息上与目标是相似的，而且还具有较高的显著性，以克服跟踪过程中采用单一颜色特征导致跟踪鲁棒性不强的问题。



基于时空信息融合的显著性检测模型

**3.2** 空间（域）显著性检测模型

3.3.1 模型框架

本章空域显著性检测模型框架如图3.2所示。在目前公认性能较优的基于全局对比度的空域模型基础上，引入SLIC方法对图像进行超像素分割，计算超像素级别的显著特征，以更好地适应目标跟踪的要求。基于全局对比度的显著模型是由初级视觉特征驱动的自底向上模型，能够准确检测出与背景具有颜色差异的前景区域。但对于目标跟踪任务，缺少目标先验信息的引导。本章引入基于任务的“自顶向下”的思想，利用上一帧目标位置先验对视觉注意焦点进行引导，以提取空域显著特征；



3.2.1颜色空间理论

颜色是人类视觉感知的一种重要信息来源。相比于几何特征，颜色特征具有与生俱来的旋转不变性和尺度不变性。依据颜色差别，人类可以很容易地将场景中的主体与背景区分开来。因此，颜色特征已经成为显著性检测领域应用最为广泛的特征之一。根据计算方式的不同可以对颜色空间进行分类，下面对RGB、CIE L\*a\*b，HSV三种常用的颜色空间进行介绍。

（1）RGB颜色空间

RGB空间是最常见的颜色空间，RGB颜色空间以红(Red)、绿(Green)、蓝(Blue)三种基本色为基础，进行不同程度的叠加，产生丰富而广泛的颜色，俗称三基色模式。

RGB颜色空间采用一个长度为1的立方体对颜色进行表示，如图3.3。三个坐标轴分别表示红绿蓝三个颜色分量，各个基色的取值范围均为：[0,255]，并经过归一化到[0,1]之间。立方体中的每个点对应RGB空间中相应的颜色。

RGB颜色空间较为直观且易于理解，但是两种颜色之间的感知差异不等同于在RGB空间中两点间的距离，因此RGB颜色空间的均匀性较差。在计算机视觉领域中，通常将RGB图像转换为符合人类视觉感知的颜色空间，再进行后续的图像处理。

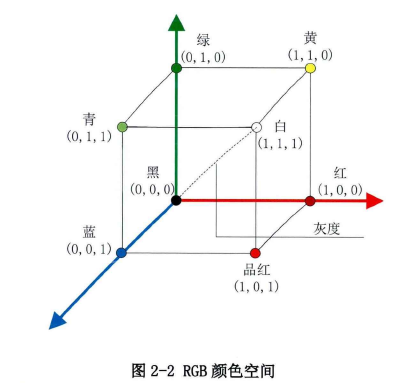


图3.3 RGB颜色空间

（2）CIE L\*a\*b颜色空间

CIE L\*a\*b，也称为L\*a\*b颜色空间，是由国际照明委员会(CIE)制定的一种接近人类视觉感知的颜色系统。L\*a\*b采用数字化的方法描述人的视觉感应，模型如图3.4所示。其中的L分量表示像素的亮度，取值范围为[0,100]，表示从纯黑到纯白；a表示从红色到绿色的范围，b表示从黄色到蓝色的范围，a、b的取值范围均是[127,-128]。

L\*a\*b颜色空间不依赖于设备，并且拥有较为宽阔的色域。它不仅包含了RGB的所有色域，还能表现出人眼所能感知到的所有色彩，弥补了RGB颜色空间色彩分布不均的缺点。在数字图像的处理过程中，为了保留尽量宽阔的色域和丰富的色彩，通常选用L\*a\*b颜色空间。

在图像处理过程中，经常会涉及到RGB与L\*a\*b颜色之间的转换，颜色空间的转换公式为：

:

 (3-1)

:

 (3-2)

式中，，，，

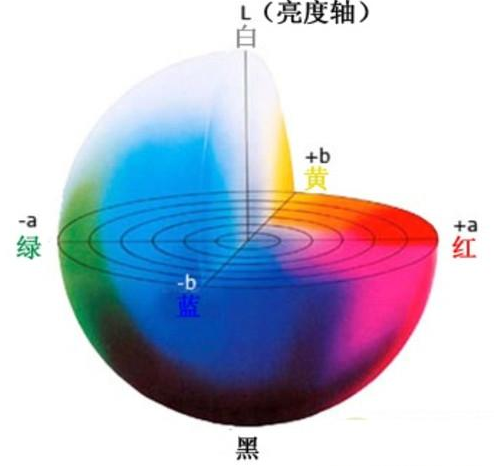


图3.4 L\*a\*b颜色空间

（3）HSV颜色空间

HSV由A.R.Smith于1978年创建，是一种能够反映颜色直观特性的颜色空间，也称为六角椎体模型。其中颜色参数H表示色调，S表示饱和度，V表示明度。HSV颜色空间模型如图3.5所示，色调H利用角度进行度量，取值范围为，红、绿、蓝分别相隔，互补色分别相差。饱和度S取值范围为0%~100%，值越大，颜色接近光谱色的程度越高，越饱和。明度V表示颜色明亮的程度，取值范围为0%（黑）~100%（白）。

在数字图像处理过程中，经常会涉及到RGB与HSV颜色空间之间的转换，转换公式如下：

:

 (3-3)

式中，；若，则。

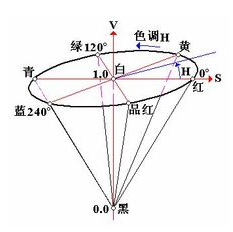


图3.5 HSV颜色空间

（4）颜色直方图

颜色直方图是描述颜色信息的一种有效方法，能够反映彩色图像的颜色信息在颜色空间的分布情况，研究表明，颜色直方图具有很好的特征表现能力，现已经被广泛应用于图像处理任务中。颜色直方图可以通过统计每种颜色在图像内的像素数而构造。首先，将颜色空间均匀地划分为若干个小的颜色区间，每个小区间作为直方图的一个，然后，计算颜色落在每个内的像素数即可得到颜色直方图。

在构建颜色直方图时，颜色区间数目的设定与应用的性能和效率要求有关，过多的颜色直方图会带来非常大的计算量。为了有效地减少数目，可以采用基于直方图的加速方法，通过保留数值较大的来构造图像特征，因为这些对应的颜色能够反映图像大部分像素的颜色信息。【实验证明】基于直方图的加速不会对图像的颜色表达带来太大的影响，并且由于忽略了数值较小的颜色区间，颜色直方图的抗噪能力也有所增强，对于某些应用来说效果更好。

3.2.2 SLIC超像素分割

超像素是由一系列位置相近且颜色、亮度、纹理等特征相似的像素点组成的小区域，这些小区域大多保留了进一步进行图像分割的有效信息，并且不会破坏图像中的边界轮廓。计算机视觉领域中，超像素分割指的是利用图像冗余信息将特征相似度较高的像素划分为一类，即构造超像素的过程。人类视觉倾向于捕捉场景中可定义的大块区域，且对单个像素并不敏感，因此利用超像素代替像素级进行图像处理，能够降低后续任务的复杂度，并在一定程度上排除噪声像素的影响。

Achanta等人于2010年提出了简单线性迭代聚类(simple linear iterative clustering, SLIC)超像素分割算法，对颜色、亮度相似且空间相干的像素进行局部聚类，进而生成紧凑、近似均匀的超像素。SLIC算法运算速度较快，在保持物体轮廓、超像素形状等方面具有较好的效果，目前已经被广泛应用于图像处理中。该算法的核心步骤如下：

（1）初始化聚类中心

在图像中设置均匀分布的聚类中心，为了保证超像素块具有紧凑均匀的分布，并且形状规则，若输入图像包含个像素点，聚类为个超像素时，每个超像素应包含个像素点，相邻聚类中心之间的距离大概为。

（2）校正聚类中心

对于每个聚类中心，计算其（通常取3）邻域范围内所有像素点的梯度值，将聚类中心移到该邻域内梯度最小的位置。以避免在聚类过程中，聚类中心落在梯度较大的轮廓边界上，从而影响后续的聚类效果。

（3）距离度量

在每个聚类中心的邻域内为每个像素点分配类标签。当期望的超像素尺寸为时，搜索范围为。对于每个搜索到的像素点，分别计算它与该聚类中心的距离，计算公式如下：

(3-4)

式中，表示L\*a\*b\*空间内的颜色距离；表示欧式距离；为最终距离度量；为平衡参数，用于控制颜色距离和欧式距离在聚类阈值中所占比重，取值范围为[1,40]，一般取10；为类内最大空间距离，取值为。

每个像素点会被多个聚类中心搜索到，统计该像素点与近邻聚类中心的最终距离，选取最小值对应的聚类中心为该像素点分配类标签。

（4）迭代优化

重复上述过程，直到每个像素点的聚类中心不在发生变化为止，实验表明大部分图片经过10次迭代便可以得到较为理想的分割效果，所以迭代次数通常取10。

3.3.2 超像素级对比度特征提取

生物学研究发现，在颜色、亮度、纹理，对比度等众多图像底层空间特征中，人类视觉系统对颜色对比度最为敏感，能够快速捕获图像中与周围相比对比度较大的区域。基于以上理论，2011年，Cheng等人提出了基于直方图对比度(Histogram Contrast, HC)的空间显著性检测模型，利用基于颜色直方图的方法对图像进行处理，依据与其他像素的色彩差异分配像素的显著性值，生成具有全分辨率的显著图。

HC为像素级的显著性检测模型，显著值定义为各像素点与图像中其它点颜色的对比度之和。像素级的对比度计算会造成较大的运算负担，无法满足目标跟踪等任务的实时要求，并且人类视觉倾向于捕捉场景中大块的可定以区域，更容易注意到和周围物体相比对比度较大的区域。因此，本文对HC模型进行改进，利用SLIC超像素分割将图像分割为若干区域，利用超像素代替像素进行显著性计算，在减少计算量的同时，使得检测结果更符合人类视觉感知。

首先，利用SLIC超像素分割方法对输入图像进行划分，得到若干个由颜色、亮度相似且空间相干的像素组成的超像素区域。对于每个超像素块，统计该块区域的颜色直方图，保留覆盖90%像素的颜色，剩下的颜色用直方图中距离最近的颜色代替。进行超像素级的颜色对比度计算，超像素块和的颜色对比度可表示如下：

 (3-5)

式中，为颜色在超像素的所有种颜色中出现的概率，可由颜色直方图统计得到，；为中的颜色和中的颜色在颜色空间的距离。

除了对比度，空间距离对于人类的注意分配也起到很重要的作用，近邻区域的高对比度比较远区域的高对比度会为显著性带来更大的贡献。因此将空间距离信息引入本文超像素级对比度的计算中，强调近邻区域的颜色差异影响，弱化较远区域的影响。对于超像素块，通过计算它与其它超像素的距离加权颜色对比度得到其对比度特征值，计算公式如下：

 (3-6)

式中，为超像素中的像素数；为和的欧氏距离；为权值控制系数，值越大，空间距离的影响越小，本文取值为0.4。

分别计算各个区域的对比度特征值，进而得到该帧图像基于颜色对比度的显著图。

1. 原图 (b) 基于超像素级对比度的显著图

图3.6 基于超像素级对比度的显著图

为了提高计算效率，减少参与计算的颜色数目，首先将RGB中每个通道的颜色量化到12个不同的值，颜色数量减小到，保留覆盖95%像素的颜色，剩下5%像素的颜色用颜色直方图中距离最近的颜色代替。

空间距离对于人类的注意分配也起到很重要的作用，近邻区域的高对比度比较远区域的高对比度会为显著性带来更大的贡献。本文空域显著模型同时考虑图像颜色对比度信息和空间相干性，利用距离加权超像素级的颜色对比度计算空间显著性值，保证每帧图像中显著区域的空间完整性。

3.3.3 结合目标先验信息的空间显著图/空域显著性检测模型

在“自底向上”显著性检测算法中，目标区域的显著程度依赖于目标区域与背景区域的颜色差异性大小、目标尺度大小、目标区域是否靠近图片中心等因素。但在实际用于跟踪的图片序列中，目标所处的环境往往很复杂，目标的尺度有时很小，目标有时也不在靠近图片中心的位置，这些都对现有的显著性检测算法造成了挑战。为了使得显著性检测模型能够适用于目标跟踪的任务，本文引入基于任务的“自顶向下”的思想，利用上一帧目标位置信息作为先验知识，对视觉注意焦点进行引导，通过引入空间掩膜抑制不相干区域的干扰，计算空域显著图。

在基于超像素级对比度特征的显著图3.6(b)中，与其他区域颜色差异较大的区域具有较高的显著度，例如前景目标以及背景中的建筑。在目标跟踪任务中，我们只需要关注其中的目标区域，并应削弱其它区域的显著度。因此，在计算区域级颜色对比度时，还应注意到目标在图像中通常是集中分布的特性，即以目标的先验位置为参考的空间关系。根据上一帧中目标区域的中心位置定义空间掩膜的质心坐标，并定义点与的位置关系度量和权重系数：

 (3-7)

 (3-8)

式中，为空间关系权重因子，用来衡量空间距离权重对显著度的影响程度，本文取0.3。

本文的空域显著性检测模型由结合目标先验信息的构成，定义如下：

 (3-9)

本文空域模型在考虑像素级颜色对比度对显著性贡献的同时，以目标先验位置信息作为显著性计算的权重参数，根据先验信息调控底层对比度特征的空间分布。对于与目标先验位置相关性较高的区域，提高该区域的对比度特征在空域显著图中的贡献度。由图3.7的实验结果可以看出，由于引入了空间掩膜，结合了位置先验信息的显著图更能突显真实目标区域，符合目标跟踪的实际需求。

(a) 基于超像素级对比度的显著图 (b) 结合目标先验信息的显著图

图3.7 引入空间掩膜前后的显著图对比

给定一幅图像，人的视觉系统往往根据图像对比度特征，对图像中颜色较为突出的区域关注度比较高。当带有基于任务的先验知识时，人的视觉系统则会自动地对这些特征进行调控，发现与顶层先验知识相关的视觉目标。基于以上理论，本文提出了底层对比度信息和先验位置信息协同感知的空间（域）显著性检测模型。

目标在第一帧的位置是己知的，如何利用目标先验信息来构造显著性检测算法，是解决以上问题的关键。

对于第一帧目标模板区域图像，采用3.3.2节的方法该模板图像的颜色直方图进行量化操作，并提取出归一化直方图，直方图中各分量的取值反应了在目标区域中，各个颜色出现的频率情况。对于当前帧A，利用SLIC将其分割为若干个区域，每个区域由具有相似纹理、颜色、亮度等特征的相邻像素构成，对于区域，统计该区域中出现频率最高的颜色在第一帧模板图像直方图中的概率，利用计算区域的颜色先验特征值：

， 

同理，遍历当前帧图像A，计算所有区域的颜色先验特征值，进而得到利用目标先验信息的显著图。【反向投影？？？】。

3.4 时间（域）显著性检测模型

3.4.1模型框架

视频由连续多帧图像组成，与单帧图像相比，帧之间的强相关性使得视频图像在时间维度上包含更丰富的运动信息。神经生物学研究发现，相比纹理、颜色等图像底层空间特征，人类视觉系统对场景中的运动信息更为敏感[A]，能够迅速捕捉到连续帧中的运动区域，因此运动信息对于视频显著性的贡献至关重要。图像序列在时域上的运动信息造成的显著性称为时域显著性，运动分为真实的显著运动以及背景干扰运动，例如树叶、草的晃动等[A]，好的时间显著性检测模型应关注视频中真正的运动信息，并能排除背景中干扰的影响。

本文提出一种鲁棒的时域显著性检测模型，结构如图3.8所示，首先采用帧差法提取出的运动区域，然后对于具有运动信息的像素点利用光流法计算运动矢量场。通常光流计算是全局均匀取点，计算量大并且缺少针对性。因此，对帧差法提取的运动区域进行光流特征的计算，能够有效减少计算量。其次，利用光流法提取出的运动矢量场计算“运动熵”和“方向一致性值”，从运动强度和空间一致性两方面生成时域显著图。



图3.8 时域显著性检测模型

3.4.2 运动特征的提取

常用的运动检测方法主要包括：背景减除法，帧差法和光流法。其中，背景减除法能够给出运动目标完整的区域表示，但对于场景中光照的变化和阴影的干扰较为敏感。帧差法通过像素的时间差分提取图像中的运动区域，具有计算量小、更新速度快的特点，但对于面积较大且颜色一致的目标，可能在内部产生空洞。光流法具有不需要场景先验信息的优点，并且在摄像机运动的情况下也能检测出独立的运动目标，但光流法的计算量较大，全局光流很难实现实时处理。

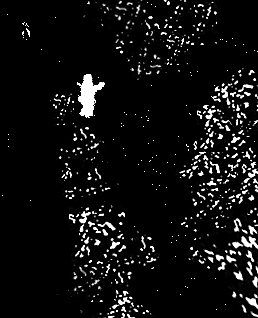
因此本文结合帧差法和光流法，利用帧差法大致定位场景中的运动区域，针对运动区域进行光流检测以提取运动特征，从而提高计算效率。本节运动特征的提取包括三个步骤：帧差法提取具有运动的区域、光流法计算运动矢量场、滤波去除非显著的运动矢量。

1. 帧差法提取具有运动的区域

帧差法，又名时间差分(Temporal Difference)，基本原理就是在连续的图像序列中，在几个相邻帧间计算基于像素的时间差分，并通过阈值化来提取出图像中的运动区域。最简单形式的帧差法即在相邻两帧间进行差分，利用二值差分图像表示帧与帧之间的变化：

 (3-10)

若相邻两帧之间对应像素的变化小于阈值，判定此处为背景像素；反之，则认为这是由图像中的运动引起的，并将这些区域标记为前景像素。本文中，帧差阈值取值10。选取数据库中的相邻两帧图像进行实验，结果如图3.9所示，实验结果表明，帧差法能够检测出包含运动的所有区域，其中真实的显著运动为鸟的运动，树叶的抖动造成了大量的干扰运动区域。

(a) 时刻视频帧 (b) 时刻视频帧 (c) 帧差法结果

图3.9 帧差法实验结果

（2）光流法计算运动矢量场

光流场为空间物体在被观测面上的像素点运动而产生的瞬间速度场，是物体的三维速度矢量在成像平面上的投影，它表示了物体在图像位置的瞬时变化。光流法是针对图像在时间和空间的运动关系进行的一系列光流场分析，根据计算方法和数学理论基础，光流法可以分为基于梯度、区域匹配、能量、相位以及神经动力学等方法。

本文采用应用最为广泛的基于梯度的Lucas-Kanade方法提取光流场，以计算像素点在相邻两帧之间的运动矢量信息。当物体保持连续运动时，图像上相应位置像素的亮度随之变化，从而形成了光流场的连续变化。假设图像上一点在时刻的亮度为，在后该像素点的亮度变为，当时认为该点亮度不变，则可以得到：

 (3-11)

假设该点的亮度变化足够小，则对后的像素亮度进行泰勒展开得到：

 (3-12)

忽略上式中的高阶无穷小，当时，可得：

 (3-13)

记，，，并令，，上式可表示为：

 (3-14)

上式即为基本的光流约束方程，其中，分别表示速度场的水平，垂直分量，光流场的计算就是求解，的过程。由于方程中有两个待求解的未知数，若想得到唯一解则需要添加约束条件，此处利用Lucas-Kanade光流法中的局部平滑约束条件进行求解。

Lucas-Kanade光流法基于三个基本假设：a.亮度恒定，目标在被跟踪期间的外观亮度保持不变；b.相邻帧之间的运动较为缓慢，即在一个时间间隔内图像是趋于不变的；c.空间一致，同一场景同一表面上的近邻点运动情况相似，且这些点投影在图像上后位置也是近邻的。基于以上假设，在像素的邻域内根据加权平方和最小化计算光流矢量：

 (3-15)

式中，为像素的权重函数，为邻域大小，利用小二乘法进行求解可得：

 (3-16)

式中，，，，，表示邻域内的像素点，分别为像素点在方向上的梯度。

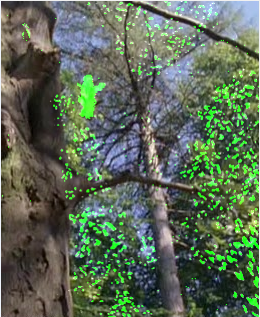
利用帧差法提取运动区域，对其中的像素点计算光流特征，各点的运动矢量可表示为：

 (3-17)

（3）滤波去除干扰运动矢量

帧差法能够检测出相邻帧之间存在运动的区域，但是时域显著性不完全等同于帧间的全部运动，好的时域检测模型应关注视频中真正的运动信息，并能排除背景中干扰的影响。因此，对于计算得到的运动矢量场进行滤波，滤除幅值过小的运动矢量，以排除帧差检测结果中微小的背景扰动对后续显著性计算的影响，滤波的计算如下式：

 (3-18)

经过帧差法检测运动像素点、光流法提取运动矢量场、滤波排除干扰运动矢量即得到用于时域显著性计算的运动矢量场。运动特征提取的结果如图3.10所示，由于帧差结果中存在大量的树叶干扰区域，使得图3.10(a)光流法的提取结果中有大量的干扰运动矢量。经过光流矢量幅值的滤波，从图3.10(b)中可以看出，大量的干扰区域被滤除。可以减少下节中显著图计算的运算量，减小时间消耗。 



3.4.3 时域显著图的计算

目标区域中像素点的运动具有连续性和方向一致性的特点，从而可以很好地与背景区分开来。本文引入“运动熵”和“方向一致性值”共同描述时域显著性，分别生成运动能量图和方向一致图，并采用相乘的方式融合成能够凸显真实目标区域的时域显著图。

视频中的显著性区域通常保持持续稳定的运动。依据运动的连续性，本文引入“运动熵”表征像素点在连续多帧内的运动强度，点在第帧的运动熵可通过计算该点在连续帧运动矢量幅值的平方和得到，为了进一步提升算法性能，在计算的过程中引入指数衰落因子以弱化较远帧对当前帧的影响，计算公式如下：

 (3-18)

式中，，为指数衰落因子，为参与计算的连续帧数，本文中取值为5。

运动熵能够反映目标在一段时间内的运动情况，目标区域由于保持持续稳定的运动，因此运动熵值较高，而背景中存在突发扰动的区域呈现不规律的运动，运动熵值较低。通常幅值越大的运动越能吸引人眼的注意，因此运动熵值与时域显著性成正相关。

目标区域内点的运动具有一致或近似一致的方向，若某区域各点的运动方向差异较大，则该区域可能位于存在微小扰动的背景中。本文引入“方向一致性值”表征像素点与邻域点运动矢量的方向相似程度。夹角余弦值可度量矢量之间角度的大小，其值越大，两个矢量的方向越相近，矢量与之间的夹角余弦值计算公式如下：

 (3-19)

对于第帧中的点，选定大小为的邻域（本文中取4），分别计算该点与其邻域内其它点的运动矢量的夹角余弦值，并进行累加求和，得到该点的“方向一致性”值：

 (3-20)

式中，为以点为中心，窗口长度为的邻域像素集合。

通常情况强度较大的运动能够吸引人眼的注意力，因此运动熵越大表示该运动的时域显著性越高。但是对于强度较低的运动，运动熵无法做到准确检测，因此要利用其他的通道去弥补运动熵的缺陷。空间一致性值表明，同一个物体的运动矢量的相位应趋于一致，如果区域内运动矢量的方向混乱，说明该位置的运动是由背景中的干扰造成，比如树叶、草的晃动，因此空间一致性值的计算时能够降低这种干扰运动的显著性。

基于以上分析将两个显著特征进行组合，得到能够表征场景运动情况的时域显著性检测模型，定义如下：

 (3-20)

本文时域模型综合考虑像素点在连续帧中的运动强度，以及与近邻点的方向相似程度，利用“运动熵”和“方向一致性值”共同描述时域显著性。时域显著性检测的结果如图3.11所示，由于具有连续稳定的运动，飞鸟所在区域的运动强度和方向一致性较高，因此具有较高的显著度。背景中抖动树叶等干扰区域的运动呈现间歇性，在连续多帧内的运动熵较小，其次干扰区域内运动矢量的方向各异，即方向一致程度较小，以上两种因素导致扰动区域的显著度较低。实验结果表明本文提出的时域模型能够实现鲁棒的时域显著性估计，并克服背景微小扰动带来的影响。

时间显著图的分析

图 4.5 显示了最终生成的运动显著图，可以看出树叶干扰区域的显著性很低，主要有几方面的原因，首先运动强度因子 I 较小，其次运动矢量的相位混乱，熵值大，即空间一致性因子sC较大，最后树叶的晃动在多帧图像序列中的运动不具有连续性，时间相位一致性因子tC 较小。而行人区域，特别是腿部的运动显著性很高，因为在图 4.3(a)和图 4.3(b)相邻两帧图像中，行人进行了抬腿的动作，运动强度较大，运动相位比较一致，并且在多帧中抬腿是一个连续的动作。



3.5 时空显著图的融合

在分别得到底层空域特征和时域运动特征的显著图之后，下面的工作就是根据场景空域和时域显著信息的关系，对两者进行融合从而生成整体显著图。心理学研究发现，当背景静止不动时，相对于颜色、纹理，对比度等图像底层空域特征，场景中的运动信息更能吸引人的注意**[16]**。在这种情况下，时空融合应该更加关注该物体的运动信息，即时域显著图比空域显著图占据更为主导的地位；当摄像头平动时，视频中的背景随前景目标一起移动，此时局部运动信息不再突出，图像对比度等空域特征更能引起人的注意，应提高空域显著图在整体显著图中所占的比重。由以上分析可知，不同背景条件下时域运动信息与底层空域信息对于显著性的贡献并不是均等的。

时空显著特征之间的关系表明，采用固定权重的线性组合来生成整体显著图是不符合实际情况的，应采用自适应动态加权的方式融合时域和空域显著图。若场景中的运动信息较为集中，则赋予时域显著图一个较大的权值；类似地，如果场景内运动信息较为平均，此时局部运动信息已经不再显著，则对空域显著图赋予一个较大的权值，强调图像底层特征对整体显著性的贡献。以上融合策略可以使得模型能够根据视频内容变化，自适应调整时域与空域显著图的比重，整体显著图的定义如下：

 (3-21)

式中，，分别为空间显著图和时间显著图的自适应权值，可由下式计算得到：

 (3-22)

式中，，分别为时域显著图的最大值和平均值。和随运动信息分布情况的变化曲线如下：

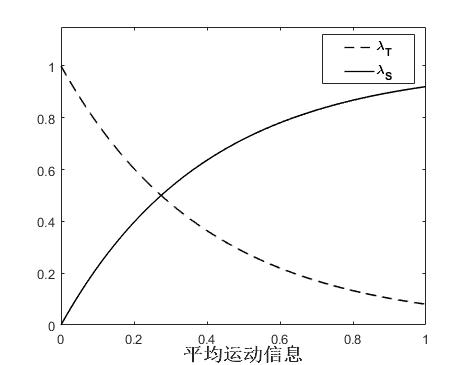


图3.12 动态权值变化示意图

当背景不动，视频中只有小范围区域存在运动时，场景中的运动信息较为集中，则较小，且与相差很大，此时增大，强调时域显著性的影响；当背景随显著区域一起运动时，视频中的运动信息较为平均，则增大，为空域显著图赋予一个较大的权值，强调对比度特征对显著性的影响。自适应动态加权的融合方式，能够根据不同背景情况，自适应调整时域和空域显著图的比重，既考虑了运动容易引起注意的特点，又合理地度量了空域显著性，符合人类感知过程的相关规律。

3.6实验结果及分析

为了验证本文提出的时空显著性检测模型在目标跟踪任务中的有效性和先进性，将本节模型与目前4种主流显著性检测模型进行定性和定量的对比实验。实验的软件环境为Visual Studio2013，硬件条件为Inter Celeron CPU G1840，4.00 GB内存。测试样本来自Segtrack数据集[17]，Dataset2014数据集[18]和VOT2016数据集[19]。所选数据集中的视频全面并具有代表性，涵盖了目标大小不一、背景信息不同的各种情况。数据集提供了手工标记的真实显著性区域Ground Truth，用于后续定量分析比较各算法的检测效果。

本节实验的对比模型包括图论注意模型 (GBVS)[A]、四元傅里叶变换相位谱模型(PQFT)[A]、基于频率调谐的模型(FT)[A]、基于感知测量的模型(SAG)[A]。其中，GBVS运用马尔可夫随机场的特点构建二维图像的马尔可夫链，通过求其平衡分布得到视频显著图。PQFT将像素点的灰度、颜色和运动特征组成的四元数组进行四元傅里叶变换，用于提取图像序列在时空域上的相位谱, 通过相位谱分析得到图像序列的显著区域。FT利用颜色特征的中央-周边算子实现单帧图像的显著性度量，虽然FT为针对静态图像的显著性检测模型，但因其计算速度较快，部分文献把它用于视频的检测。SAG利用时域上具有运动的轮廓边界预测前景目标的位置，通过计算超像素与边缘像素的测地距离生成时空显著图，同时构建前景和背景的全局外观模型和动态位置模型，最终实现时空一致的视频显著性检测。

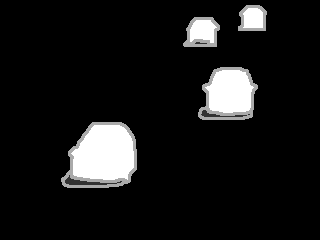
GBVS 算法的鲁棒性较好，但在显著区域边缘容易造成大面积的模糊，不能确定显著区域的明显边界。PQFT该方法仅采用连续两帧之间的差异估计运动信息，当背景的运动较为剧烈时，无法准确定位显著目标的位置。FT算法对于简单场景中的大目标检测能力较强，但对于复杂背景中噪声的干扰鲁棒性一般。SAG 算法主要依赖于图像的边缘特征和运动梯度特征，在复杂场景下容易产生噪声区域。

定性比较

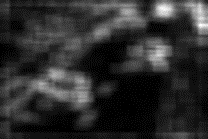
首先从定性角度对本文模型与对比模型的性能进行评价，选取不同情形的视频序列进行测试，用以分析不同背景条件下的显著性检测效果。下面对选取的测试数据集对应的场景进行介绍。

图3.13-图3.15显示了不同场景下的检测结果，分别采用本文提出的模型、GBVS模型、PQFT 模型以及FT模型进行实验对比。

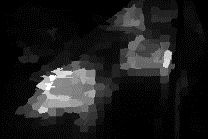
图3.13为背景静止且包含大块匀质区域视频的检测结果。highway视频中，车辆由远及近高速运动，真实的目标区域为车辆所在位置。由实验结果可知，对于此类视频，GBVS模型的检测结果中目标区域呈现大面积模糊，且无法检测出图像边缘位置的显著区域；PQFT模型仅采用连续两帧之间的差异估计运动信息，对于背景中树叶的扰动鲁棒性较差，易将干扰区域误判为显著区域，不能很好地区分目标区域与背景；FT模型为仅利用颜色特征的空域显著性计算方法，无法筛选出视频中的运动信息，检测效果较差，进一步强调了运动信息对视频显著性的重要性；本文模型均取得了较好的检测效果，本文提出了鲁棒的时间显著性计算方法，能够较为准确地提取出视频中的运动信息，并且该视频中运动信息较为集中，在时空信息融合阶段，本文提出的自适应融合方法能够自主调节时域显著性在整体显著性中所占比例，因此能够较好地检测出真实运动区域。

1. highway第189帧 (b) Ground Truth

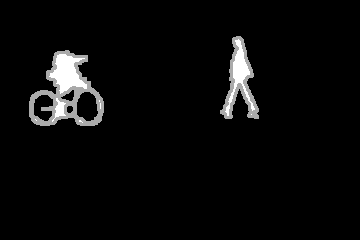
(c) GBVS模型 (d) PQFT模型

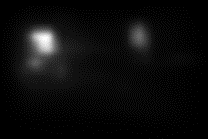
(e) FT模型 (f) 本文模型

图3.13 highway视频第189帧显著性检测结果

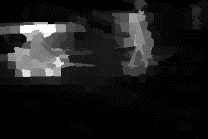
图3.14为目标区域较小且存在微小抖动视频的检测结果。pedestrians视频中，行人在街道上相向运动，真实的显著性区域为行人所在位置，摄像机在拍摄过程中发生抖动，因此该视频背景存在动态干扰。由实验结果可知，对于此类视频，GBVS模型能够实现对目标区域的大致定位，但检测结果较为模糊且不能确定显著区域的边界；PQFT模型将背景中的栏杆以及草地等位置误判为显著区域，无法区分背景和前景目标；FT模型只能检测出大致轮廓，不能均匀地突出目标区域；本文模型不仅能确定前景目标的位置，而且对两个前景目标的显著度存在一定的区分。自行车相较于行人的运动速度更快，能够更加吸引人类视觉的注意，相应地，图3.14(f)本文模型计算得到的显著图中，自行车区域的灰度值高于行人区域的灰度值，即显著图中自行车区域的显著度更高，因此本文模型的检测效果更加符合人类感知。利用“运动熵”和“方向一致性”共同描述时间显著性，使得本文模型能够克服摄像机的轻微抖动带来的干扰，并且较为准确地检测出连续帧中的运动目标区域。

(a) pedestrians第469帧 (b) Ground Truth

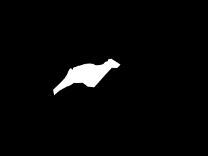
(c) GBVS模型 (d) PQFT模型

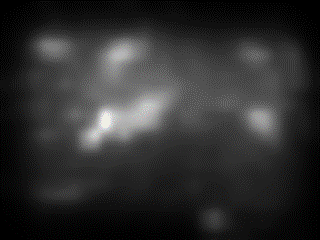
(e) FT模型 (f) 本文模型

图3.14 pedestrian第469帧显著性检测效果

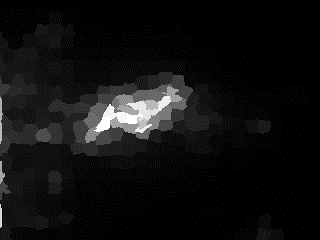
图3.15为背景纹理复杂且摄像头平动视频的检测结果。cheetah视频中，一只羚羊在草原上快速奔跑，目标区域为羚羊所在位置，该视频背景与目标区域颜色相似，场景纹理较为复杂，并且摄像机随羚羊的运动平动。由实验结果可知，对于此类视频，GBVS模型的检测结果存在大面积模糊区域，且将背景中的不相关位置误判为显著目标区域，检测效果较差；PQFT模型能够确定目标的大致位置，但检测结果较为模糊；FT模型能够检测出显著目标的部分边缘，无法均匀地突出显著位置；本文模型能够对目标进行大致的定位，但由于前景目标运动速度过快，而产生了一定的拖尾现象。实验结果表明本文算法同样适用于背景发生变化（例如摄像头平动）的情况，当场景中运动信息较为平均时，目标区域的运动信息已经不再突出，此时图像底层空域特征对于显著性的贡献更大，本文时空显著性模型根据场景中运动信息的分布情况自适应调节融合权重，提高空域显著图的权值，以适应摄像机平动的情况，从而克服了背景运动带来的动态干扰。

(a) cheetah第28帧 (b) Ground Truth

(c) GBVS模型 (d) PQFT模型

(e) FT模型 (f) 本文模型

图3.15 cheetah第28帧显著性检测效果

定量比较

为了更客观、精确地评价几种模型的性能，选用显著性检测领域常用的准确率—召回率曲线（P-R曲线）、F指标[19]对本文模型与对比模型进行定量评价。

（1）准确率—召回率曲线

准确率(precision)定义为正确检出的目标区域点数与检出的像素点总数之比，召回率(recall)定义为正确检出的目标区域点数与Ground Truth中目标区域像素点总数之比。用[0,255]区间内256个整数阈值对显著图进行二值化分割，计算每个阈值下的准确率和召回率，分别作为纵坐标和横坐标即得到P-R曲线。P-R曲线的评价准则为：在相同的召回率下，准确率越高则模型的性能越好。本文模型与对比模型的准确率-召回率曲线如图3.16所示，由实验结果可知，本文模型的显著性检测效果明显优于对比模型。

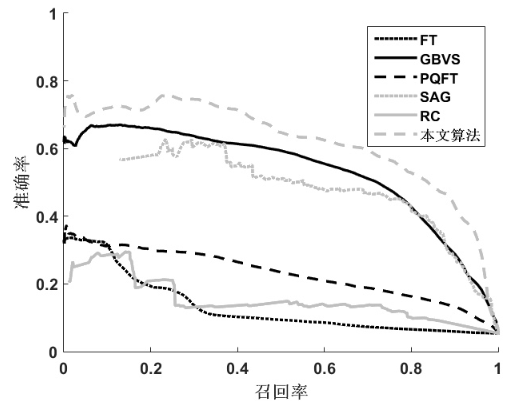


图3.16 P-R曲线

（2）F指标

F指标是融合准确率和召回率的综合评价指标，具体计算公式如下：



与文献[19]相同，式中的值设置为0.3。F指标表示模型提取真实目标区域的同时抑制背景区域的能力，其值越大，模型的检测性能越强。由于F指标受分割方法的影响较大，为了准确比较模型的性能，本文实验分别采用Saliency Cut[19]，Frequency-tuned[16]方法对显著图进行分割，并计算在不同分割方法下的F指标。图3.17为本文模型与对比模型在测试数据集上的F指标评价结果，实验结果表明，在不同分割方式下，本文模型均取得了最高的F值，表明本文模型在提取真实显著目标区域的同时，抑制背景区域的能力较强。

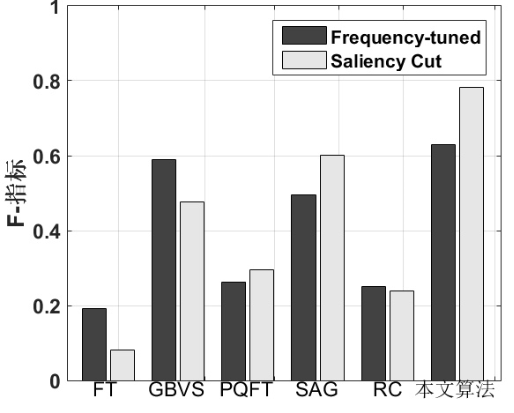


图3.17 F指标比较

综上，本文模型在定性分析和定量比较上均取得了较好的效果。不同测试视频的定性结果表明，本文模型能够较为精确地检测出视频中的显著目标区域，并且能够很好地克服背景中的微小扰动以及相机运动等干扰的影响，适用于不同背景条件下目标检测。准确率—召回率曲线、F指标的定量评价结果进一步验证了本文模型的有效性和鲁棒性。

3.7本章小结

本章在前文基于直方图对比度的空域显著性检测模型的基础上，引入目标先验信息和时域运动特征，提出了基于目标先验信息的时空显著性检测模型。该模型主要分为空域显著性计算、时域显著性计算以及时空信息融合三个部分。空域上，通过超像素级的对比度计算自底向上的空域显著特征，引入基于任务的自顶向下思想，利用目标先验信息对当前视频帧的空域特征进行调整，并构建空域显著图。时域上，在运动信息提取阶段，首先采用帧差法提取运动像素点，然后通过光流法计算运动矢量场，最后滤波排除幅值较小的运动矢量。利用运动熵、空间一致性值对运动矢量的显著性进行分析。运动熵表征像素点在连续帧中的运动强度，空间一致性值表示与近邻点的运动方向相似度，综合以上两个运动特征图谱，计算得到时域显著图。时空信息融合阶段，根据人眼对不同情形视频的主观感知，利用场景中运动信息的分布情况，设置动态权重自适应融合时空显著图，以得到最终视频显著图。在本章最后，将本文提出的模型与目前主流的显著性检测模型进行了对比实验。实验表明，本文提出的时空显著性检测模型在不同场景下均具有较好的鲁棒性，在准确率和召回率等方面优于对比模型，能够实现场景中目标的大致定位，从而进一步验证了显著性特征在目标跟踪中的有效性，以及通过模拟人类视觉注意进行目标跟踪的可行性。

第四章 基于视觉显著性特征的Camshift跟踪算法

前面的章节提出了模拟人类视觉感知的时空显著性检测模型，提出了模型的原理和建立过程，并通过实验证实了显著性检测应用于目标跟踪的可行性和有效性。本章主要研究运动目标的跟踪问题，在借鉴第三章视觉显著性检测研究的基础上，提出一种基于视觉显著性特征的Camshift目标跟踪算法。采用颜色特征和视觉显著性特征共同描述目标模板，利用视觉显著性特征在光照变化、姿态变化等干扰环境下的不变性来提高跟踪精度。

文献[A]通过显著性检测提取视觉显著性特征，并将其引入粒子滤波框架中，与颜色特征进行自适应融合，从而完成跟踪。将视觉显著性特征应用于目标跟踪任务中，以克服单一特征在跟踪中的不稳定性。基于以上思想，

由于Camshift算法具备较优的局部匹配和快速搜索能力，本文运动目标跟踪算法是在Camshift算法的框架下完成的。在目标跟踪过程中，u需要提取特征以准确地构建目标模板，颜色特征作为目标最直观的自然特征，对部分遮挡以及姿态变化具有较好的鲁棒性，因此经常被应用于目标跟踪。但是，颜色特征对于场景中的噪声和光照变化十分敏感，随着光照强度的改变，视频图像的像素值可能发生较大的变化，为跟踪的鲁棒性带来极大的挑战。为了克服利用单一颜色特征构造目标模板的局限性，本章在借鉴第三章视觉显著性检测研究的基础上，提出一种基于视觉显著性特征的Camshift目标跟踪算法。采用颜色特征和视觉显著性特征共同描述目标模板，利用视觉显著性特征在光照变化、姿态变化等干扰环境下的不变性来提高跟踪精度。

4.1算法流程

针对复杂环境下的运动目标跟踪问题，本文提出了一种基于视觉显著性特征的Camshift目标跟踪算法。首先提取目标区域的颜色直方图并计算其反向投影图，采用本文提出的视觉显著性检测模型对视频图像进行检测，生成时空显著图，然后在经过显著性加权的反向投影图基础上计算均值漂移向量，更新目标团块，通过Bhattacharrya距离确定收敛窗口，最后利用Camshift算法对收敛窗口进行调整，作为下一帧的初始搜索窗口，从而完成跟踪。本章算法流程如图A所示。



基于视觉显著性特征的Camshift跟踪流程图

4.2 Mean Shift算法

4.2.1模型描述

目标的颜色特征不易受外界因素的干扰，因此MeanShift采用颜色直方图对目标模型和候选模型进行描述[A]。在初始帧中确定目标区域，将区域中总共个像素点的值域划分为均匀的区间，得到由个相等的区间构成的颜色直方图， 利用颜色特征的概率密度函数表示目标模型，如公式：



式中，表示区域中的像素点；为归一化常量。为Epanechnikov核函数；用于判断目标区域中像素出的灰度值是否属于直方图中第个单元，属于则为1，否则为0。

同理，当前帧中候选模型的可表示为；

 （3-17）

式中，表示候选区域中的像素点；为候选区域的中心点；为核窗口半径；为归一化常量。

4.2.2 相似性度量

相似性函数表征目标模型和候选模型之间的相似程度[43]，MeanShift算法中采用目标和候选位置颜色直方图的Bhattacharyya系数计算得到，定义如公式A：

 888

式中，，候选模型与目标模型越相似，越大，反之越小。

4.2.3 MeanShift迭代

为得到最大的相似函数，将在上一帧目标区域的中心位置处进行泰勒展开：

 (3-19)

由公式（3-17）、（3-19）求得下列形式：

 (3-20)

式中， 为权重，见式（3-21）：



公式（3-20）的第二项表示当前帧中点处的核密度估计，其值随变化，求取最大即为使该式中第二项最大的过程。当前帧中目标的位置可通过候选区域中心向真实区域中心的迭代完成，

4.2.4 算法流程

MeanShift 算法流程图如图 3-2 所示。该算法在视频序列中对特定运动目标进行跟踪一般包括以下几个步骤：

（1）在初始帧中选择目标区域，计算其颜色特征概率密度函数。

（2）用初始化当前帧的候选目标位置，计算其颜色特征概率密度函数。

（3）计算目标模型和候选模型之间的相似性函数。

（4）利用公式（3-21）计算当前窗口内各点的权重 。

（5）通过MeanShift迭代，得到目标的新位置。

 (3-23)

式中，。

（6）更新，并计算新的相似性函数。

（7）若，执行并返回步骤（6）；否则，执行步骤（8）。

（8）若，停止迭代；否则，并返回到步骤（3）。

4.3基于视觉显著性特征的Camshift跟踪算法

Camshift算法由Gray R.Bradski等人提出[45]，作为改进的MeanShift算法，Camshift解决了MeanShift在跟踪过程中不能自适应调整搜索窗口大小的缺点。基本思想就是对视频图像的每一帧进行MeanShift运算，将上一帧结果作为下一帧的初始值，如此迭代下去。该算法采用不变矩对目标的尺寸进行估算，能够连续自适应地调整跟踪窗口的大小和位置，实现了彩色图像序列中运动目标的快速跟踪。MeanShift是在单张图片中寻找最优的迭代结果，而Camshift则是利用MeanShift的方法对连续的图像序列进行分析。

4.2.1显著性加权的反向投影图

反向投影图是一种色彩的概率分布图，用输入图像的各个像素在直方图中对应上的值代替该像素值，即可得到输入图像对应的反向投影图。从统计学的角度，反向投影图中像素点的值为观测数组在目标区域的颜色直方图分布下的概率，因此像素值越高，表明这个点属于目标的概率越大。

RGB颜色空间对光照变化敏感，为了减小光照变化的影响，本文将输入的RGB图像转换为HSV格式，根据初始帧目标区域H分量的构建颜色直方图。对于输入图像，查找其中每一个像素的H分量在目标直方图中对应的，将这个对应的纵坐标代替该点的像素值，得到一副表示色彩概率分布的图像，即反向投影图。反向投影图为单通道灰度图像，每一个点的值反映了输入图像中对应点属于目标区域的概率，因此像素值越高，该点属于目标区域的可能性越大。对于输入的图像图A，其行人区域的反向投影图如图A所示。

Camshift跟踪方法本质上根据颜色特征进行目标匹配，通过在当前帧中查找与目标颜色信息相似的区域实现对目标的跟踪。该方法利用颜色直方图描述目标模板，但是颜色特征对噪声和遮挡不敏感，当背景与目标信息相似情况下，跟踪的准确性和鲁棒性受到较大的影响。本文对Camshift跟踪算法进行了改进，以第三章基于时空信息融合的显著性研究为基础，采用视觉显著性特征与颜色特征共同描述目标模型，以克服利用单一颜色特征进行跟踪所带来的的不稳定。本文利用时空显著图对反向投影图A进行加权，计算公式如A所示。

对显著性加权的反向投影图进行MeanShift迭代计算，使得跟踪结果不仅与目标具有较高的颜色相似度，并且具有较高的视觉显著度，能够反映场景中的真实情况AAA。



4.3.2 MeanShift

在反向投影图中通过使用MeanShift算法进行迭代搜索，直到达到最大迭代次数或者收敛，使得搜索窗口收敛到目标位置。在2D概率分布图中，MeanShift向量可由图像的零阶矩和一阶矩计算得到：



搜索窗口做MeanShift迭代的过程如下：

（1）根据初始帧中目标区域的大小和位置初始化搜索窗口。

（2）计算搜索窗口的质心。

，

式中，为搜索窗口的零阶矩，、为一阶矩。







式中，为反向投影图中点的概率值。

（3）将求得的质心作为新的搜索窗口的中心。

（4）当搜索窗口的移动距离小于一定的阈值或达到最大迭代次数时结束；否则返回（2）。

4.3.3 Camshift调整收敛窗口

MeanShift在跟踪中搜索框的大小一直不变，对目标的尺度变化不具有鲁棒性。针对此问题，Camshift跟踪算法利用MeanShift迭代所得收敛窗口的不变矩对目标的尺寸进行估算，以调整当前帧的收敛窗口，并确定下一帧搜索窗口的初始大小，实现了跟踪窗口大小的连续自适应调整。

收敛窗口的二阶矩，，可通过公式（1-3）计算得到。







定义，，为下列形式，并通过收敛窗口的不变矩和质心坐标计算求得。







计算目标的方向角、目标分布的长和宽，如公式A：







下一帧搜索窗口的初始宽度和高度见公式A。





4.3.4跟踪算法流程

Step1：读入第一帧视频

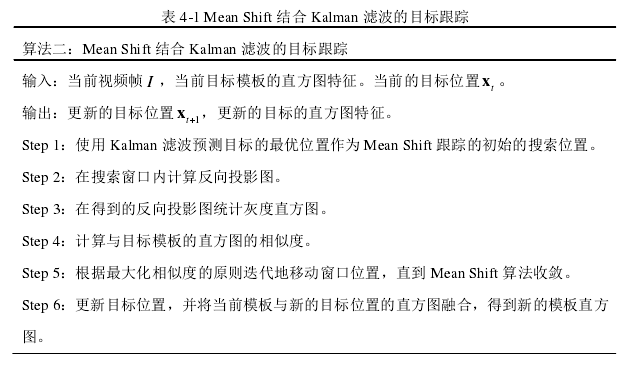
1. 初始化目标中心位置，模板半径。
2. 将目标区域从RGB格式转换为HSV格式，提取H分量的直方图作为颜色特征模板。

Step2：读入下一帧视频，进行MeanShift迭代优化。

1. 以，初始化搜索窗口。
2. 提取中各像素点的H分量，依据计算的反向投影图；
3. 对进行时空显著性检测，利用时空显著图对反向投影图进行加权，得到。
4. 在上调用MeanShift算法，计算均值漂移向量，搜索得到收敛窗口，。

Step3：利用Camshift算法更新收敛窗口的大小为。

Step4：令，，重复Step2。



（1）在初始帧中选择目标区域，对其进行颜色直方图统计。

（2）若当前帧图像不为空，计算它的颜色概率分布图（即反向投影图）。

（3）根据反向投影图和搜索窗口，调用MeanShift算法，得到收敛窗口。

（4）利用Camshift算法对目标尺寸进行估计，以调整收敛窗口的大小。

（5）在下一帧图像中，用步骤（4）得到的调整过的收敛窗口作为搜索窗口，并跳转到步骤（2）继续执行。Camshift法流程图如图3-3所示。

具体跟踪流程：

在初始帧中手动选取待跟踪的目标，将目标区域图像从RGB转换为HSV格式，并提取H分量的直方图作为颜色特征模板。从第二帧开始，首先以上一帧目标位置初始化搜索窗口，提取搜索窗口中各像素点的H分量，依据初始帧目标区域的直方图计算反向投影图；同时，对当前帧进行时空显著性检测，以提取目标区域的视觉显著性特征；然后，在经过时空显著图加权的反向投影图上调用MeanShift算法，计算均值漂移向量，搜索得到收敛窗口；最后利用Camshift算法对目标尺寸进行估计，以调整收敛窗口的大小。