

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

UDC <sup>注1</sup> \_\_\_\_\_

# 学 位 论 文

复杂场景运动目标跟踪算法研究

(题名和副题名)

卢耀坤

(作者姓名)

指导教师

彭真明

教 授

电子科技大学

成 都

(姓名、职称、单位名称)

申请学位级别 硕士 学科专业 信号与信息处理

提交论文日期 2019.04 论文答辩日期 2019.05

学位授予单位和日期 电子科技大学 2019 年 06 月

答辩委员会主席 \_\_\_\_\_

评阅人 \_\_\_\_\_

注 1：注明《国际十进分类法 UDC》的类号。

# **Research on Moving Object Tracking of Visual Target Under Complex Scenario**

A Master Thesis Submitted to

University of Electronic Science and Technology of China

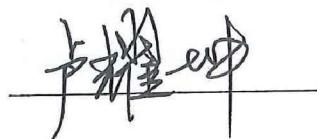
Discipline: **Signal and Information Processing**  
Author: **Yaokun Lu**  
Supervisor: **Prof. Zhenming Peng**  
School: **School of Information and Communication Engineering**

---

## 独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

作者签名：



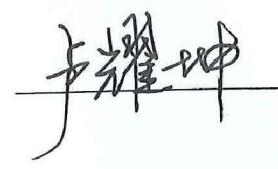
日期：2019 年 6 月 24 日

## 论文使用授权

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权电子科技大学可以将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、汇编学位论文。

(保密的学位论文在解密后应遵守此规定)

作者签名：



导师签名：



日期：2019 年 6 月 24 日

## 摘要

运动目标跟踪技术是近年来计算机视觉技术的热点研究方向之一，在智能监控、医疗影像、自动驾驶、无人机视觉导航等领域都有着广泛的应用。然而由于跟踪场景的复杂多变性，当面临快速运动、运动模糊、形状变化等目标外观改变，以及噪声干扰、光照变化、低分辨率等环境变化时，使用单种特征进行跟踪容易造成跟踪漂移<sup>[1]</sup>，而目标被遮挡是引起跟踪失败的主要原因。

本文针对复杂场景下目标跟踪的各种难点问题进行了分析，在此基础上介绍了传统的相关滤波算法，基于相关滤波算法的优缺点，提出了一种基于多特征融合以及重检测的目标跟踪算法框架。本文主要研究内容包括如下：

(1)研究了目标跟踪相关理论，对跟踪中常用的几种特征进行了研究，并对它们的性能和特点进行了分析。提出了多特征联合使用来增强跟踪场景适应性的特征的使用策略，使得跟踪算法更有利于在各种场景下进行跟踪。

(2)研究了相关滤波方法的相关理论，介绍了算法的基本理论和基本步骤。针对相关滤波跟踪器无法评价当前跟踪效果的不足，增加了跟踪效果评价模块，使得跟踪器可以判断目标是否已经丢失。

(3)针对跟踪中目标发生遮挡后无法继续跟踪的问题，引入了重检测模块，根据跟踪效果评价模块的输出，当判断目标丢失时，在图像中重新检测目标，待找到目标后重新启动跟踪。

(4)最后，本文在大规模公开数据集对目前主流的几种跟踪器和本文提出的跟踪器进行评测和比较。实验表明，相较于传统算法，本文算法在复杂场景下的跟踪效果取得了满意的效果。

**关键词：**目标跟踪，复杂场景，目标遮挡，相关滤波，预测跟踪

## ABSTRACT

Moving target tracking is a popular research directions of computer vision technology in recent years. It has been widely used in intelligent surveillance, medical imaging, automatic driving, visual navigation and other fields. However, due to the complexity and variability of tracking scene, while facing fast motion, motion blurring, shape change and other changes in the appearance of the target, as well as noise interference, illumination change, low resolution and other environmental changes, tracking drift is easily caused by using a single feature, and the main reason for tracking failure is that the target is occluded.

In this thesis, the difficulties of target tracking in complex scenes are analyzed. On this basis, the traditional correlation filtering algorithm is introduced. Based on the advantages and disadvantages of the correlation filtering algorithm, a framework of target tracking algorithm based on multi-feature fusion and re-detection is proposed. The main contents of this thesis are as follows:

(1) The related theory of target tracking is studied, several common features in target tracking are studied, and their performances and characteristics are analyzed. A strategy of using multi-feature to enhance the adaptability of tracking scene is proposed, which makes the tracking algorithm more conducive to tracking in various scenarios.

(2) The related theory of correlation filtering method is studied, and the basic theory and steps of the algorithm are introduced. In order to overcome the shortcomings of correlation filter tracker which can no evaluate current tracking effect, a tracking effect evaluation module is added to enable the tracker to judge whether the target has been lost.

(3) Aiming at the problem that the target cannot continue tracking after occlusion occurs in tracking, the re-detection module is introduced. According to the output of the tracking effect evaluation module, when the target is lost, the target is re-detected in the image, and the tracking is restarted after the target is found.

(4) Finally, this thesis evaluates and compares several mainstream trackers and the trackers proposed in this thesis on large-scale open datasets. Experiments show that compared with the traditional algorithm, the tracking effect of this algorithm in complex scenes has achieved satisfactory results.

**Keywords:** target tracking, complex scene, object occlusion, kernelized correlation filter

## 目 录

<b>第一章 绪论 .....</b>	<b>1</b>
1.1 研究背景及意义 .....	1
1.2 国内外研究现状以及发展动态 .....	2
1.3 本文主要研究内容以及技术路线 .....	6
1.4 本文主要结构安排 .....	7
<b>第二章 相关跟踪理论基础 .....</b>	<b>9</b>
2.1 相关滤波跟踪的基本理论 .....	10
2.1.1 岭回归模型 .....	11
2.1.2 循环矩阵 .....	11
2.1.3 核函数 .....	13
2.2 相关滤波跟踪的基本步骤 .....	15
2.2.1 特征提取 .....	16
2.2.2 余弦窗函数 .....	17
2.2.3 训练分类器 .....	20
2.2.4 快速目标检测 .....	23
2.2.5 模型更新 .....	24
2.3 本章小结 .....	25
<b>第三章 目标跟踪中的特征融合方法 .....</b>	<b>26</b>
3.1 运动目标跟踪中的特征 .....	26
3.1.1 HOG 特征 .....	26
3.1.2 LBP 特征 .....	30
3.1.3 HOG 特征与 LBP 特性能分析 .....	33
3.1.4 颜色特征 .....	34
3.2 特征融合方法 .....	36
3.3 本章小结 .....	39
<b>第四章 基于重检测的相关跟踪算法设计及实现 .....</b>	<b>40</b>
4.1 基于峰值旁瓣比的跟踪置信度 .....	41
4.2 被遮挡后的目标重检测方法 .....	44
4.2.1 级联分类器检测模块 .....	45
4.3 实验结果 .....	48

4.3.1 实验运行环境 .....	50
4.3.2 实验评价指标 .....	50
4.3.3 定性分析 .....	52
4.3.4 定量分析 .....	54
4.4 算法的工程实现 .....	58
4.5 本章小结 .....	61
<b>第五章 总结与展望 .....</b>	<b>62</b>
5.1 工作总结 .....	62
5.1.1 开展工作 .....	62
5.1.2 创新点及主要贡献 .....	63
5.2 后续研究展望 .....	63
<b>致谢 .....</b>	<b>64</b>
<b>参考文献 .....</b>	<b>65</b>
<b>攻读硕士学位期间取得的成果 .....</b>	<b>71</b>

# 第一章 绪论

## 1.1 研究背景及意义

在人类的各种感官中，负责与外界进行联系的主要有三种：听觉，触觉和视觉，其中人类的视觉系统是人类生产实践中最主要的信息来源接收系统，接收了 80% 的外界信息。计算机视觉(Computer Vision, CV)是一门研究如何让计算机学会去和我们人类一样“看”真实世界的科学，我们可以将计算机视觉看成是一种模拟人类视觉的方法，它延伸了人类有限的视觉官能，实现了对影像资料的自动化分析和理解。随着目前摄像机芯片生产工艺的提升，摄像机的价格变得越来越便宜，它在人们日常生产生活中得到了广泛的应用，人类产生了史无前例的大量的视频数据，所以人们产生了对这些海量影像资料进行智能分析的需求，这也从侧面成为了计算机视觉技术发展的重要驱动力。随着计算机硬件的飞速发展、分布式架构的海量运用，极大地促进了研究人员对目标跟踪邻域的深入探究。

视觉目标跟踪算法在人们生活的各方面都应用广泛，尤其是智能监控<sup>[2-4]</sup>和人机交互<sup>[5, 6]</sup>等方面。例如：在公安领域，公安相关人员在银行通过摄像头的跟踪技术对罪犯进行追踪并分析出罪犯的特征和意图，运用城市道路上广泛布设的安防摄像头对犯罪人员、车辆进行时空联合定位，使得犯罪分子无处可逃；在军事领域，战场上的环境是瞬息万变的，所有的军事武装都需要对敌军目标进行定位和跟踪，使用目标跟踪技术实现对敌军目标的精准打击；在无人机领域，无人机通过自身摄像头获得周围环境的视觉信息后，对视场中的固定目标进行检测、定位，随后反过来对自己的导航参数进行更新，帮助无人机实现更精确地导航与定位；在医疗领域，由于人体内部血管复杂，脏器众多，环境十分复杂，医生凭借自己的经验通过人工听诊的方式对病人病情进行判断时经常会发生误判。而凭借目标跟踪技术，医生们在电脑上对需要检查的部位进行精准的锁定，从而提升诊断的精确性，也使得诊断更为容易、详实、客观；在汽车无人驾驶领域，随着计算机硬件和激光雷达技术的发展，无人驾驶技术已经不再是空中楼阁，各大汽车厂商和互联网公司都纷纷斥巨资对相关技术进行研究。目标跟踪技术可以帮助无人驾驶汽车对实际行驶道路上的各种目标保持跟踪，例如正在行走的行人，道路两侧的人行道阶梯，交通标志牌，路面标志，信号灯等目标，借此帮助无人驾驶汽车决策，控制汽车的行驶状态，提升无人驾驶的安全性。目标跟踪技术还在虚拟现实(Virtual Reality, VR)、增强现实(Augmented Reality, AR)技术中有着广泛的应用，目标跟踪技术可以推动各方面相关技术更加广泛的运用到复杂环境之中<sup>[7]</sup>。

复杂场景下的目标跟踪技术在学术上具有重要研究价值，在实际应用中又有巨大的应用价值。目标在运动过程中，多种多样的因素会影响跟踪算法的性能和效率，算法不仅要解决目标自身受到的光照、形变、快速运动带来的干扰，还要解决目标被障碍物甚至被其他的目标部分遮挡或完全遮挡，或者在某些图像序列中目标移出视场的情况。这些目标自身的影响因素在实际的应用场景中是普遍存在的，而且在实际应用中，目标所在的场景本身也是在变化的，算法需要在动态场景下对目标进行跟踪<sup>[8]</sup>。而且由于目标跟踪应用场景的特殊性，通常都要求跟踪算法具备实时性，所以算法的复杂度不能太高，这也是目标跟踪算法的关键之一。在实际应用场景中有可能受到复杂气候条件变化、外界噪声过大等因素的干扰，尤其是在红外图像跟踪中，红外图像有着低信噪比、低对比度和目标信息少的特点，在这种条件下进行跟踪十分困难<sup>[9,10]</sup>。尽管各种去噪方法可以改善图像质量<sup>[11]</sup>，但是由此带来的额外计算量也不容小视。此外，有限的计算资源也变成算法面临的主要挑战之一。

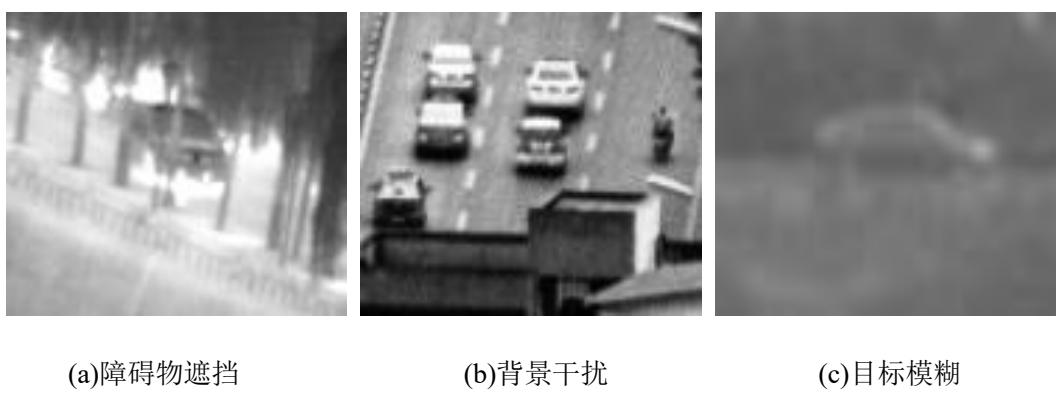


图 1-1 复杂场景的图像。 (a) 目标被障碍物部分遮挡； (b) 目标被背景干扰； (c) 目标由于运动或分辨率过低变得模糊

本课题来源于四川长虹科技股份有限公司与实验室的合作课题：《图像增强与电子稳像与扩展目标稳定跟踪技术算法研究》。课题旨在探究复杂场景下运动目标跟踪在实际使用环境中应用时的问题，实现一套鲁棒、准确、有效的算法并最终在 DSP 上部署。

## 1.2 国内外研究现状以及发展动态

西方发达国家在很早以前就发现了目标检测和跟踪技术的重要性并开始了相应地研究，并取得了一定的成果，技术上也比较成熟。但是在复杂环境下的对目标的监测和追踪技术还没有十分完善，因此广大科研工作者在该领域投入了大量精

力。Wax 在 1955 年首次提出了目标跟踪技术这个概念，到目前为止该领域经过大量专家学者的补充和完善，已经可以称为一个较成熟的领域。但由于现实世界中的不可控因素太多，目前还没有哪种单独的算法可以完全轻松的适应各种情况。这也大大激发了该领域科研工作者的研究热情，希望可以解决这个充满挑战性的难关，致使该领域成为计算机视觉方面的一个热门研究方向。GCA 公司早在上个世纪 50 年代就为海军部队自主研发了自动地形识别系统。1997 年以卡内基梅隆大学和麻省理工学院牵头的诸多高等学院参与了美国国防部设立的 DARPA 项目，该项目的主要内容是视觉跟踪。Maryland 大学通过自主研发，发明了可以利用将人体进行分割成不同部位来实现多人扫描跟踪的要求的实时监控系统—W4，该系统不仅能实现多人跟踪还可以对检测目标是否携带物品进行监控。U.K 雷丁大学基于道路上的行人和运动车辆的监控要求研发了智能视频跟踪系统，该新系统在一定程度上取得了不错的成果。文献<sup>[12]</sup>大致的对近年来提出的跟踪算法进行了总结和归纳：

目前国际上在影像处理和内容识别领域具有主要影响力的期刊有 IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)、IEEE Transactions on Image Processing (TIP)、International Journal of Computer Vision (IJCV) 等。在计算机视觉领域具有最高学术水平的会议代表是 :International Conference on Computer Vision (ICCV)、IEEE Conference on Computer Vision and Pattern Recognition (CVPR)、European Conference on Computer Vision (ECCV) 统称 ICE。上所涉及到的科学期刊和学术会议中都会大致的介绍以下当前科学界主流认为比较有价值和影响力关于目标追踪方面的高影响因子论文和受到广泛认可的观点。

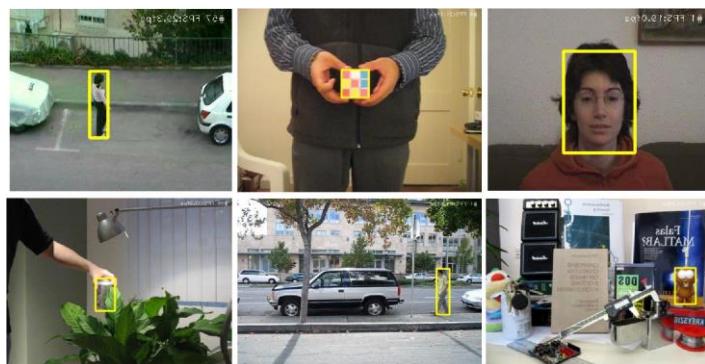


图 1-2 常见的几种跟踪场景

基于现有的不同原理可以将目标跟踪算法分为以下两种：以生成式为基础的跟踪方法和以判别式为基础的跟踪方法：

以生成式为基础的跟踪方法的基本原理是先将目标通过数据化处理使其重新建立为一个模型，然后在所监测的环境范围内对所有存在物进行搜索，把搜索结果

数据通过最小化重构误差，最后在数据库里选出最佳与最先建立的模型相匹配的目标。这种方法以搜索目标的外观为区分点，在搜索范围内进行目标的跟踪定位。

以判别式为基础的跟踪方法是将目标存在区域和目标不存在区域分为两个不同的样本，再设计一个的判别函数，以上两个数据库为基础分别建立一个正样本和负样本，也就是建立一个目标存在样本和目标不存在样本。在整合所监测区域的样品数据，和上面建立的两个标准样本对比，把其中和正样本符合情况最高的样本输出，以此来达到追踪目标的要求。可以把这种方法归结成一种把跟踪问题转换成二分类问题的方法。

### (1)传统跟踪方法

下面介绍几种优秀的传统跟踪方法，它们当中涌现出了许多优秀的方法和思路，后来的跟踪方法中大多能看到它们的思想：

归一化互相关(Normalized Cross Correlation, NCC)<sup>[13]</sup>: 这种跟踪方法选择的模板是灰度特征，从目标框中得到这个灰度特征以后，下一步就需要对目标进行直接匹配，这一步将会使用归一化互相关的方法。具体的过程如下：首先在每一帧对候选图块进行采样，这些图块处于目标附近，下一步就是对这些图块以及其他的目标模板使用归一化互相关的方法进行处理，最后选出峰值最大的候选图块，这就是需要的新目标位置。需要注意的是在 NCC 中的目标模板有些不同，因为它不存在更新过程。

前景后景跟踪(Foreground-Background Tracking, FBT)<sup>[14]</sup>: 这种方法中利用的判别分类器属于渐进式的，其中对于判别分类器的训练主要使用的信息是两个，分别是背景区域中的 Gabor 小波的纹理特征以及目标区域中的 Gabor 小波的纹理特征。这种方法会在选择上一帧目标旁边的位置处进行寻找，然后利用分类器确定新目标的具体位置。这种方法是通过新旧采样的样本来完成模型更新的。

均值移位法(Mean-Shift Tracking, MST)<sup>[15]</sup>: 这种方法将 RGB 颜色直方图作为基础信息，然后根据这些信息来进行匹配。因为不是利用像素灰度实现，所以在一些目标变形的情况下也可以使用。找到一个数据函数能够让巴氏距离达到最大，然后利用这个函数目标的最佳位置，这种方法没有模板更新。

增量视觉跟踪(Incremental Visual Tracking, IVT)<sup>[16]</sup>: 这种方法可以抓取目标外表的所有变化，因为这种方法可以对目标的外表模型进行扩展。首先获得目标的图像特征，这是主成分分析(Principal Component Analysis, PCA)得到的，在运动模型中通过粒子进行采样，最终得到候选窗口。下一步就需要考虑每个样本的置信度，具体方式是在候选窗口到目标这个范围内进行特征空间的测量，其中得分最小的一种情况就是候选窗口。

下面介绍的是基于判别分类的跟踪，这种方法和上文提到的有许多不同之处，这种方法的基础是目标的前景以及目标的背景，主要依据是这两个信息得到的区分模型，也就是检测跟踪，其中分类器的更新利用的新的样本数据。

多实例学习跟踪(Multiple Instance Learning, MIL)<sup>[17]</sup>: 这种方法中正样本选择的是离目标近的一部分区域。反之，离的较远的就是正样本，所以，对于 MIL 分类器来说，正样本指的是近距离的其他矩形框，负样本指的是远距离的矩形窗口，这里的特征选择的是 Haar。从上一帧中的圆形区域中可以得到候选窗口。

光流法(Kanade-Lucas-Tomasi Tracking, KLT)<sup>[18]</sup>: 所谓的光流场指的是在图像灰度模式下的一种表面运动，这个信息可以对两个临近帧之间的运动情况进行反馈，利用这种原理可以完成运动目标监测和跟踪的工作。这种方法第一步是找到仿射变换模板，然后确定目标运动之后的位置，这是通过观察目标旁边的像素的相关性以及灰度变化来完成的<sup>[19]</sup>。这种方式只要得到目标速度场的数据就可以得到目标的结构关系和运动关系。

跟踪-学习-检测 (Tracking-Learning-Detection, TLD)<sup>[20-22]</sup>算法主要依靠对于未知目标的跟踪来实现，这种方法非常的稳定，在很多非常复杂的情况下也能够运用。同时使用前后还能够对于目标的具体偏移量进行反映。在检测器得到的目标和实际不一致的情况下会对跟踪器进行重置，重置是依据正样本库信息实现的。同时其中的学习模块可以了解检测器以及使用的跟踪器性能。这种方法的不足点在于计算量巨大，让整体框架的效率下降。值得注意的是，通过对图像整体进行分割，进一步缩小检测器的检测范围，提高算法的效率和精度<sup>[23]</sup>。

## (2) 基于相关滤波的跟踪方法

这里介绍的利用相关滤波实现的跟踪方法在很多地方都有使用，因为其具备鲁棒性以及高效性的特点<sup>[24]</sup>。最开始在跟踪领域中使用相关滤波方法的是误差最小平方和滤波跟踪器(Minimum Output Sum of Squared Error, MOSSE)<sup>[25]</sup>，Bolme 等研究人员对于跟踪目标模型的解码提出了新的方案，即通过学习一个具备有自适应特点的相关滤波器来完成，这里的模型使用的是灰度特征，这个模型的速度有 669FPS；

Heriques 等人通过核化的循环结构跟踪器(Circulant-Structure of tracking-by-detect with Kernels, CSK)<sup>[24, 26, 27]</sup>对原本在信号处理领域中常用的相关滤波算法应用到跟踪中，为了提高模型的表达能力，他们利用了梯度直方图特征以及来自于目标相邻图像块的循环结构信息进行训练。对于核相关滤波来说，其中涉及到的核函数有三个，分别是高斯核，多项式核以及线性核。利用线性核的模型速度最快，但是会损失一部分性能。这些研究为以后的滤波器研究提供依据。

长程相关跟踪器(Long-Term Correlation tracking, LCT)<sup>[28]</sup>中为了解决在长期的对目标的跟踪监查过程中存在的目标可能被一些其他物体所遮挡的问题，特地为其增添了一个增加置信度滤波器。其原理是根据峰值旁瓣比(Peak-to-Sidelobe Ratio, PSR)来判断被跟踪的目标是不是存在为其他物体遮蔽的情况，若目标跟踪母体存在被其他物体遮蔽的情况，系统将通过随机分类器进行新的定位，对目标进行二次跟踪；

空间正则化相关滤波(Spatially Regularized Discriminative Correlation Filters, SRDCF)[29]用于解决跟踪过程中存在边界效应的问题，同时解决样本进行分类的过程中存在不完善信息的负样本等问题，以免对循环矩阵的构造造成不良影响，特地将目标函数的设定中添加了一个限定期程序，对整个样本数据库加以约束，其中心思想可以概括为对边界附近的滤波器系数接近于0，也可以理解为其忽略循环移位边界部分像素值。但由于采用了空间限制的方法，这也会使得初始设立的所要求解的函数方程并不能解除答案，为了解决这类问题我们可以选择采用高斯-赛德尔迭代法对设定函数进行求解，然而由于采用了高斯迭代法进行求解，导致 SRDCF 的速度只有 5 帧每秒，这又使运算速率大大下降。

由于跟踪算法具有巨大的应用价值，如何将算法与运算平台适当地结合起来，也是一个重要的问题。王珏<sup>[29]</sup>和唐龙<sup>[30]</sup>分别完成了基于安卓和 DSP 这两类嵌入式平台上的跟踪算法移植工作。

### 1.3 本文主要研究内容以及技术路线

本文将研究一类能在复杂场景下完成跟踪任务的算法，利用计算机视觉中的特征提取，目标跟踪及检测等方法，对目前已经存在的跟踪器进行改进，以期其能在多种复杂的场景下完成跟踪任务。因此本文主要研究复杂场景下的运动目标跟踪算法，包含以下内容：

- (1) 研究相关滤波类算法的基本原理与具体步骤，结合其实现原理分析其当前不能完成多种复杂条件下的跟踪任务的原因，为后文将其改进奠下基础；
- (2) 研究了在跟踪中对跟踪结果至关重要的特征。研究了几种优秀的特征描述子，并分析它们的优劣。最后给出一种特征融合的方法使得在跟踪中将特征的优势互补，提高跟踪性能；
- (3) 研究了目标重检测以及跟踪置信度的计算方法，并用其改进了目前现有的跟踪器的不足；
- (4) 研究了改进后的跟踪器在大规模公开跟踪数据集上的表现效果，分析了其优点和不足。

本文的技术路线图如图 1-3 所示。

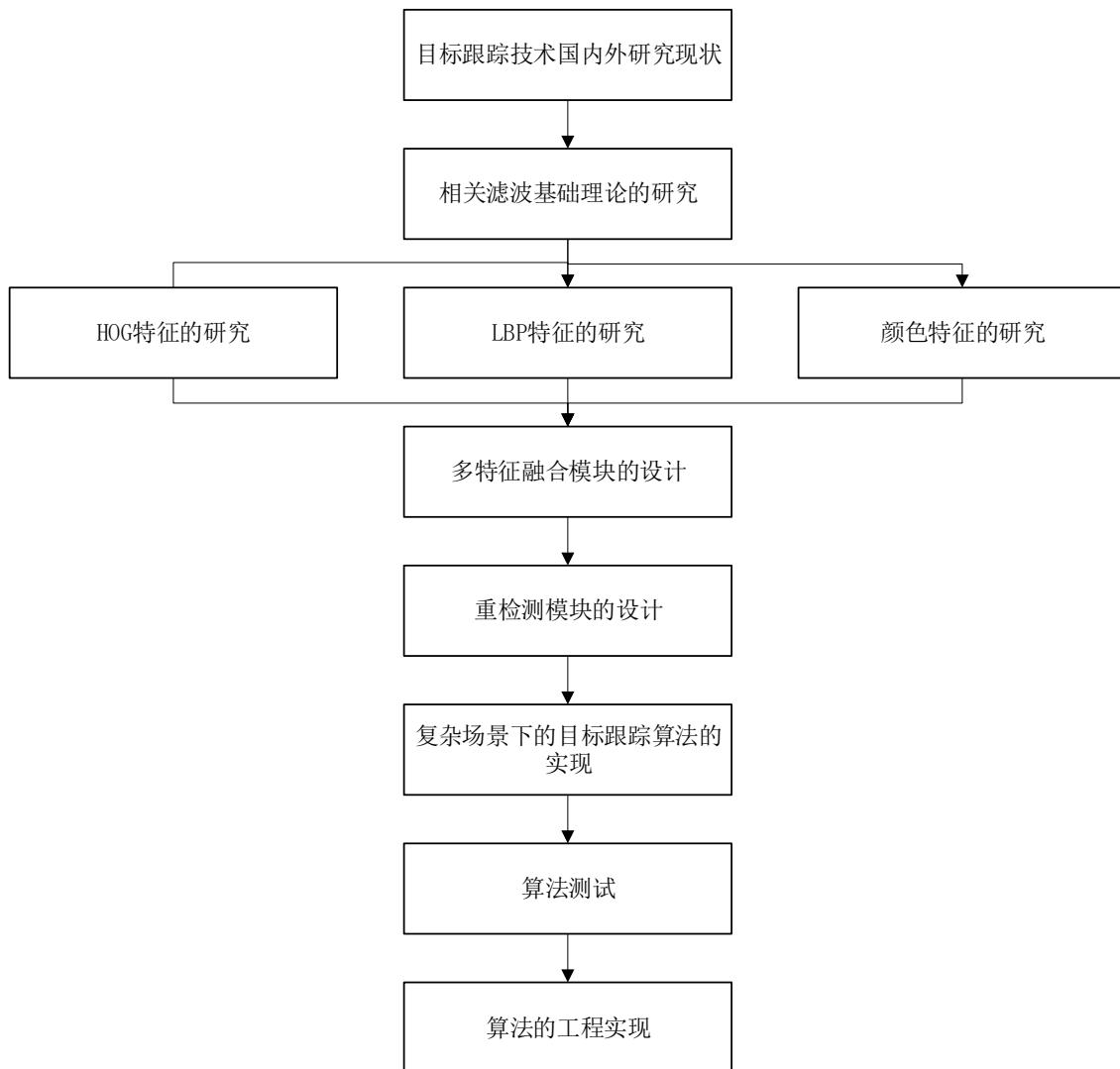


图 1-3 本文算法技术路线图

#### 1.4 本文主要结构安排

本文针对复杂场景下的运动目标跟踪算法进行研究，各章节内容安排如下：

**第一章：**先大概的对目标跟踪算法的发展和当前研究内容进行总结，再特别提出由于现实情况环境的复杂性导致了对运动目标跟踪存在一些亟待解决的问题，并对其中的主要研究方向和研究难点进行了分析。

**第二章：**对本文所采用的相关滤波类跟踪方法进行了简要描述，然后介绍了算法的基础理论并对其中重要的理论和步骤进行了介绍，为后文改进相关滤波类算法打下了理论基础。

**第三章：**对目标跟踪算法中的特征进行了简要的介绍。然后具体介绍了几种重

要的特征：方向梯度直方图(Histogram of Oriented Gradient, HOG)特征、局部二值模式(Local Binary Pattern, LBP)特征、颜色名字(Color Name, CN)特征，并对他们的性能进行了分析，最后提出了一种融合多特征进行跟踪的特征融合方法。

第四章：针对复杂场景下运动目标跟踪中最常见的目标遮挡问题，提出了基于运动目标轨迹预测和目标重检测的运动目标跟踪算法。结合第二章第三章的内容，最终形成了一个改进算法并对本文提出的算法进行了充分的实验。实验表明，这种算法在复杂场景下具有良好的跟踪效果，但也有其不足之处。

第五章：全文总结，总结本课题的主要研究成果及复杂场景下目标跟踪算法的主要创新点，并提出了算法的应用可能和缺点改进。

## 第二章 相关跟踪理论基础

在广大科研工作者的努力下，学术界已经发表了许多卓有成效的算法，但是由于跟踪中场景的复杂多样性，在目标跟踪领域中还存在着许多急需解决的问题，下面将简单的介绍几个主要的问题：

在目标跟踪中，目标的跟踪会受到环境因素的影响。例如光照明暗变化、外界物体遮挡、非刚性变形和刚性变形等干扰因素的影响，跟踪目标的外观会在不同的跟踪场景下发生改变。为了解决这类干扰因素，一般采用的方法是使用更加复杂的目标外观模型，但是在采用更加复杂的模型时往往会对计算机的计算能力提出更高的要求，这就导致跟踪算法的实时性降低。由于当前技术的限制，计算机的运行能力有限，采用越发复杂的模型将导致需要更大的计算量，故为了保证算法的运算正常，用来表征目标的外观模型并不能无限制地复杂化。综上所述为了尽可能的得到最优解，在实际的目标跟踪算法中，工作人员必须在算法的工作效率和跟踪精度中间做出平衡，不能盲目的提出更高的要求。

目标跟踪的目的是估计出目标在下一帧的位置，但是在基于检测的目标跟踪框架下，分类器的输出往往是样例的标签，这种“0-1”式的非黑即白的逻辑判断式决策机制显得过于生硬，所以通过最大化分类器响应的方法定位目标会带来一定的误差，且这种误差会随着运行时间的推移而越来越大<sup>[31]</sup>。

由于目标跟踪多是服务于在线视频分析，所以优秀的目标跟踪算法必须具有很高的处理速度。这也对跟踪算法提出了运算速度的要求：算法的运算速度必须高于视频的播放速度。目前的视频制式中常见的视频的播放速度为 25 帧每秒(FPS, Frames Per Second)，只要高于这个速度，我们就认为满足了实时性要求。

在目标跟踪这个研究领域里，美国的斯坦福大学和英国的哈佛大学等世界知名高校均在目标跟踪算法领域投入了大量的时间和精力，在该领域的发展上可以认为是行业的最前列。国内计算机领域的发展时间有限，在目标跟踪算法领域的研究时间没有国外高校这样长久，故很多方面仍有欠缺。但在不懈的努力下，如清华大学、华东科技大学、中国科学院等科研机构均取得了一定进展。

相关滤波(correlation filter) 理论最早在信号处理领域被提出并应用<sup>[32]</sup>。相关运算定义如式(2-1)所示。

$$(f \odot g)(\tau) = \int_{-\infty}^{+\infty} f(t)g(t + \tau)dt \quad (2-1)$$

其中， $\odot$  表示相关运算符， $f$  和  $g$  是两个待相关的信号的函数表达式。

相关滤波操作的物理意义可以看成：在某个同一时间 $t$ 时两者的相似程度。假若两个函数相关的输出值越大，则表示两个函数越相似。把这个相关的概念转换成频域里的公式可以表示为式(2-2)。

$$(f \odot g)(\tau) = F^{-1}(F(t) \otimes G(t + \tau)) \quad (2-2)$$

其中， $F(t)$  和  $G(t + \tau)$  分别表示  $f(t)$  和  $g(t + \tau)$  的傅里叶变换<sup>[33]</sup>，操作符  $\otimes$  即是点乘运算。

通过这种方法，频域上的计算就转化为傅里叶域内的计算，在实际的运算中，对算法的计算速度有很大提升。另外，如果将傅里叶变换的计算放在 GPU 上运算，则效率还可以进一步提高<sup>[34]</sup>。

我们将相关滤波<sup>[32]</sup>中求两个信号相似程度的思想和相关操作的这种具有特殊性质的操作运用到目标跟踪的领域里，在对图片进行比对时两张图片相似度越高，那么他们进行相关操作得到的值就越大。这样我们就可以通过前一帧图的结果来预测后一帧图中对象的坐标。然而这种跟踪方法对跟踪场景的要求很高，在复杂场景下很难得到理想的结果<sup>[35]</sup>，为使这种方法应用到实际中来，我们就必须考虑到各种复杂的环境因素，有针对性地设计出对在各种环境下都能更稳健地跟踪目标的目标跟踪算法。

## 2.1 相关滤波跟踪的基本理论

在基于检测思想的视觉跟踪算法中，如何选择合适的判别式和分类器，是非常关键的一步。分类器可以区分出跟踪目标和背景信息，同时为了应对图像可能出现的变化，分类器需要对特征样本进行训练操作，训练的过程中会出现冗余样本或者重叠像素，相关滤波算法在详尽的观察了训练样本的结构特征后，进行了分析总结，因此算法的解析模型效率很高<sup>[36]</sup>。

近些年来，这种相关滤波器面世之后，由于它具有高效率和高速的处理速度，吸引了大批学者的注意，该方向也逐渐成为该领域的热门方向。国外的 Bolme DS 和他的团队首次把相关滤波器应用到了目标跟踪领域：用相关滤波器在灰度图像上学习 MOSSE 的算法<sup>[37]</sup>；后来 Henriques 及其团队提出了核化的循环结构跟踪算法(Circulant-Structure of tracking-by-detect with Kernels, CSK)，这种算法在样本周围进行密集的取样，并利用 RLS 分类器进行归纳、分析和样本的分类。搜集样本的特征、分类器优化、查找目标坐标、更新分类器的是该算法的四个重要步骤。

我们的思路是针对该算法在目标跟踪领域的缺陷进行算法的更新而且会进一步的阐述我们对该算法的改进思路，本章详细介绍核相关滤波算法的相关知识要

点：2.1 节先阐述了核相关算法理论中的一些基本概念并就如何根据这些概念把核相关算法运用到目标跟踪算法领域中 2.2 节详细介绍和解释了核相关跟踪算法的几个基本流程；2.3 分析和总结了本章的主要内容。

### 2.1.1 岭回归模型

一维岭回归方法与传统的支持向量机（Support Vector Machine, SVM）分类器方法的基本思想相同，都是采取了对每个样本赋予相应的权值，通过样本的训练拟合出分类器参数。但不同的是，岭回归使用了 L2 正则化的线性回归模型<sup>[38]</sup>，岭回归方法可以得到一个简单的闭式解，却可以得到接近支持向量机分类器这种更为复杂方法的性能。岭回归训练的目标是拟合一个函数  $f(z) = w^t z$  使得样本  $x^i$  和它们的回归目标  $y^i$  之间的平方误差最小化，即：

$$\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2 \quad (2-3)$$

其中， $x_i$  代表第  $i$  个样本， $f(x_i)$  表示分类器预测  $x_i$  的输出， $y_i$  表示样本  $i$  的标签， $\lambda \|w\|^2$  为正则项， $\lambda$  为正则化参数， $w$  为权重系数。

类似在支持向量机分类算法中加入正则项的方法，(2-3)中  $\lambda$  为控制过拟合的一个正则化参数。最小化误差存在一个闭式解，可以由下式给出：

$$w = (X^T X + \lambda I)^{-1} X^T y \quad (2-4)$$

其中，在矩阵  $X$  中，每一行的元素  $x^i$  代表一个样本的向量， $y$  中的每一个元素相当于一个回归得分  $y_i$ ，式中的  $I$  为单位向量矩阵。

将(2-5)用复数形式表示，得到：

$$w = (X^H X + \lambda I) X^T y \quad (2-6)$$

其中，矩阵  $X^H$  是原矩阵的厄米特转置， $X^*$  是原矩阵的复共轭， $X^H = (X^*)^T$ 。

上述问题的求解需要求解一个相对复杂的高维线性方程组，求解这个方程所需的计算量在实时系统中是很难完成的，下面将介绍利用循环矩阵冗余采样时的特殊性质减少计算复杂度，绕开这个限制，使得这个方法在实时性敏感的跟踪场景中也得以使用。

### 2.1.2 循环矩阵

将目标可能出现的区域成为感兴趣的区域(Region Of Interest, ROI)可以看做  $n \times 1$  维的矩阵，用  $x$  表示，作为后续训练的样本，利用上文提到的岭回归学习方法，

训练一个分类器。岭回归分类器的训练中需要正样本和负样本，我们把目标可能出现的区域块 $x$ 作为正样本，负样本则是目标周围的除了目标的背景信息，通过循环移位算子(置换矩阵) $P$ 对区域块 $x$ 的一维进行移位建模。为了简洁起见，本节主要讨论单通道，一维信号下的情况，当然讨论的结果也可以拓展为多通道和多维图像。

$$P = \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad (2-7)$$

其中， $P$ 是代表区域块 $x$ 的一维循环移位得到的置换矩阵。 $(2-7)$ 表示将第一行样本循环移位得到的矩阵。

将一维向量 $x$ 通过置换矩阵进行循环采样，则乘积 $Px = [x_n, x_1, x_2, \dots, x_{n-1}]^T$ 将原来的样本进行了一个单位的移位转换。通过若干次同样的操作即可使用矩阵的幂运算 $P^n x$ 表示为对样本进行多个像素的移位。当 $n$ 是正数的时候，表示图像向右移位，当 $n$ 是负数的时候，表示图像向左移位。当循环移位应用到二维图像的时候。可以看到将基础图像分别左移右移数个单位后，效果类似于在图像周围使用滑动窗口采样，如图 2-1 所展示。

$$C(\text{---}) = \left[ \begin{array}{c} \text{基础样本} \\ \diagdown \text{移动一个元素} \\ \diagdown \text{移动两个元素} \\ \vdots \\ \diagup \text{移动} n-1 \text{个元素} \end{array} \right]$$

图 2-1 一维循环矩阵采样



图 2-2 二维循环采样



图 2-3 二维循环矩阵采样

循环移位构建出的训练样本集是跟踪目标在矩形框不同位置的，假设基础图像中的目标在矩形框中间，则通过循环移位后得到的新样本距离矩形框的位置可以由该样本的移位方向和次数获得。如图 2-2、图 2-3 展示了循环矩阵采样的两个实际例子。根据训练样本集移位的距离，我们可以生成一个线性回归的模型，这个模型由原本的图像经过循环位移得到。而且循环矩阵可以由傅里叶矩阵经过对角化获得，故岭回归分类器的训练速度得以大大提高。

### 2.1.3 核函数

核函数是一类用来表示空间中任意一点  $X$  到某个中心之间距离的函数。机器可以结合该方法对特征计算的问题进行降维处理。从而化繁为简，从一个比较简单的角度简化并解决了比较复杂的高维特征空间里的分类问题和回归问题。

核函数是一类将一个函数从原来空间映射到另外一个空间的一类函数，常用的核函数如表 2-1 所示。

如前所述，在有的情况下，我们的训练样本是线性可分的，但是在很多现实的任务中，比如在我们的分类任务中，原始样本空间里也许很难存在一个让我们可以简单而正确地将样本分成两类的平面或者超平面。对于这种问题，我们可以将函数从原来的样本空间映射到一个维度更加高的空间中去。是这个样本在这个维度更加高的空间中实现线性可分。如果原始样本空间的维度是有限的，即它的属性的数量是有限的，则一定可以找到一个高维特征空间将其线性可分。

表 2-2 常用的几种核函数

名称	表达式	参数
线性核	$\kappa(x_i, x_j) = x_i^T x_j$	
多项式核	$\kappa(x_i, x_j) = (x_i^T x_j)^d$	$d > 1$ 为多项式的次数
高斯核	$\kappa(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$	$\sigma > 0$ 为高斯核的带宽
拉普拉斯核	$\kappa(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ }{\sigma^2}\right)$	$\sigma > 0$
Sigmoid 核	$\kappa(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$	Tanh 为双曲正切函数， $\beta > 0, \theta < 0$

一般的分类器参数求解是需要样本内部有线性关系作为前提条件的。但在实际情况里，样本之间本身没有太多线性关系，这就使得分类器参数的求解过程与实际情况存在较大的出入。由于岭回归和目前流行的深度学习或者神经网络不同，它的网络是单层结构，在处理线性不可分问题的时候，不能通过多层网络映射的方法来解决，所以需要从空间映射的角度来解决。引入核函数后，可以减少样本线性不可分问题的影响，原本线性不可分的样本，通过核函数映射到高维空间后，原本在原空间线性不可分的数据很可能就变得线性可分了。常见的许多核函数都是可以满足这个性质的，径向基函数核，例如高斯核；累加核，例如交叉核，点乘核，例如多项式核，线性核等都可以减少线性不可分问题。本文中采用的是高斯核函数，将空间映射到了无穷维数的空间。

假设存在一个非线性映射函数  $\phi(x)$ ，样本经过此函数处理之后是线性可分的，因此，能够利用处理之后的数据去训练分类器  $f(x_i) = w^T \phi(x_i)$ ，再利用最小二乘法最小化平方误差(2-8)来计算分类器的权值  $w$ 。

$$w = \min_w \| \phi(X)w - y \|^2 + \lambda \| w \|^2 \quad (2-8)$$

其中， $\phi(X)$  是  $\phi(x)$  的矩阵形式， $w$  为  $\phi(X) = [\phi(x_1), \phi(x_2), \dots, \phi(x_{n-1})]^T$  行向量张成的空间中的一个向量， $y_i$  表示样本  $i$  的标签， $\lambda \| w \|^2$  为正则项， $\lambda$  为正则化参数， $w$  为权重系数。

令  $w = \sum \alpha_i \phi(x_i)$ ，上式就化简为：

$$\alpha = \min_w \| \phi(X)\phi(X)^T \alpha - y \|^2 + \lambda \| \phi(X)^T \alpha \|^2 \quad (2-9)$$

其中， $\alpha$  为分类器参数。

求导，使上式导数为零：

$$\alpha = (\phi(X)\phi(Y)^T + \lambda I)^{-1} y \quad (2-10)$$

对于核化的岭回归，令  $K = \phi(X)\phi(X)^T$  它的解(2-11)将会变成如下形式：

$$\alpha = (K + \lambda I)^{-1} y \quad (2-11)$$

其中， $K = C(k^{xx})$  是循环矩阵， $k^{xx}$  是矩阵  $K$  的第一行，它的每一个元素可以表示为  $K_{ij} = \kappa(x_i, x_j)$  且矩阵  $K$  是矩阵  $\phi(X)$  和它的转置的积。

故  $K^{-1}$  是一定存在的，但由于经过映射后的循环矩阵转变成非循环矩阵。因此，我们要给计算过程加一些约束，使得它的计算结果还是循环矩阵。

而  $K$  是循环矩阵的充要条件是  $C(x)$  满足： $\kappa(x, x') = K(Mx, Mx')$ ，对于任意的排列矩阵  $M$  成立。则式(2-11)可化简为：

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \quad (2-12)$$

其中  $k^{xx}$  为核矩阵  $K$  的基样本，即  $K = C(k^{xx})$ 。

至此，我们已经可以求得最优分类器参数  $\hat{\alpha}$ 。本小节简要介绍了岭回归分类器的构造以及其训练过程。接下来简要描述一下相关滤波跟踪算法的基本步骤

## 2.2 相关滤波跟踪的基本步骤

上一节讲述了相关滤波跟踪的基础理论，对相关滤波跟踪的大体思想有了一个整体的把握，下面再来对相关滤波跟踪方法的具体使用流程简要说明一下：

相关滤波类的跟踪算法一般有一个如图 2-3 的比较普通的流程<sup>[39]</sup>。这个流程可以总结概括为以下步骤：

- (1) 初始化：算法根据输入，找到到目标的区域范围，并自动提取与之对应的特征，然后结合算法训练之后得到相应的滤波模板。
- (2) 区域特征提取：从后续帧中和预测的区域内提取足够的特征
- (3) 目标定位：在傅里叶域进行图像数据的相关转换处理。通过加余弦窗处理来解决边缘效应的问题。把特征通过傅里叶变换和训练得到相应的滤波模板，并对该滤波模板进行乘积及相关，最后通过傅里叶反变换，把图中最大的响应的点的坐标位置作为目标的新的坐标位置。
- (4) 模型更新：对新的位置采集新的目标进行更新训练并获取最新的相关滤波

模板，并转入下一帧的跟踪。

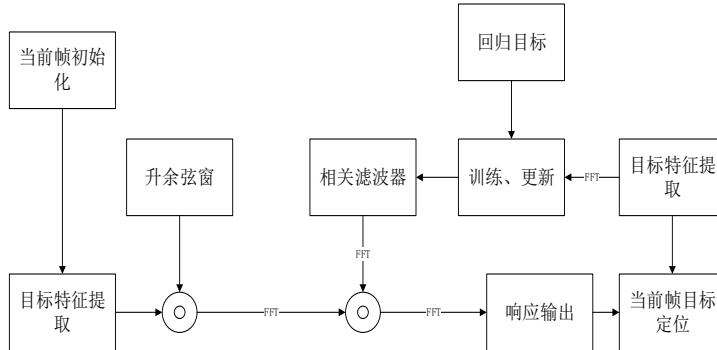


图 2-3 相关滤波跟踪流程图

### 2.2.1 特征提取

我们的特征提取分两步进行：采集样本以及提取视觉特征。根据图 2-3 所示结果显示，特征提取可以发生在目标检测阶段或是模型更新阶段。接下来我们对不同阶段的特征提取的过程做个简单的介绍：

首先在目标检测阶段，机器现根据要求选出候选样本，再从中提取视觉特征。系统对于样本的选取方式是根据传统的跟踪算法在前一帧中目标位置坐标的周围随机取样，从而得到一些候选样本，但传统方法由于受到跟踪实时性的约束，机器选出的候选样本一般不是特别多。相对于传统的算法，新的核相关滤波跟踪算法结合循环矩阵的方式在前一帧的目标附近提取作为基准的样本构成循环矩阵从而达到候选样本的密集取样，这样就得到了详细的目标附近的候选样本，从而解决了如何精准定位的问题。而且，由 2.1.2 节中提到的循环矩阵的性质我们可以得知，把基准样本  $x$  转化成循环矩阵  $C(x)$  之后的计算转换到了频率域计算，因此，整个计算由始至终只结合了基准样本  $x$ ，而且循环矩阵可以不利用显示的计算方式和储存，这样我们及时大大的增加了候选样本的同时也不会消耗太多的计算时间。这给整个计算过程节约了很多的时间成本和储存空间。

然后在模型更新阶段，机器根据要求选出训练样本，再从中提取视觉特征。然后对于如何去选取训练样本，传统的跟踪算法采取的是直接从当前帧中的目标坐标附近提取训练样本并做一套仿射变换转换成训练后的正样本，于训练样本的选取方法，最后在距离目标较远的地方提取目标样本也采取类似的仿射变换转换成训练后的负样本。传统的方法受到跟踪实时性的束缚，机器采集的训练样本一般不

会太多，得到的模型对目标不能具备足够的判断力并在一定程度上影响了算法的追踪的性能。相对于传统的算法，新的核相关滤波跟踪算法结合循环矩阵的方式在前一帧的目标附近提取作为基准的样本构成循环矩阵从而达到候选样本的密集取样，这样就得到了详细的目标附近的候选样本，从而解决了如何精准定位的问题。与候选样本的选取过程类似，我们及时增加了候选样本的同时也不会消耗太多的计算时间。这给整个计算过程节约了很多的时间成本和储存空间。

目前，有学者<sup>[40]</sup>结合实验论证提出跟踪算法中最重要的一步就是提取视觉特征(颜色、边缘、纹理、光流等)，而且选择适合的鲁棒特征可以大幅度的提高算法跟踪的准确性。我们采用的核相关滤波跟踪算法结合了以方向为梯度的直方图 HOG 的特征<sup>[41]</sup>来对目标进行一系列的跟踪。但在比较复杂的跟踪场景及环境里，单一的特征无法涵盖整体的信息，所以需要整合多个特征来对目标进行稳定而又持续的跟踪。所以怎么去选择视觉特征，怎么去整合这些复杂的视觉特征都是很有前景和意义的问题。

## 2.2.2 余弦窗函数

为了能够更好的解决其中存在的训练样本不足以及训练的问题，这里会利用到核相关滤波跟踪算法，因为这个算法可以进行循环矩阵近似密集采样。其中对目标样本的提取位置就是当前帧的目标位置，将这里提取到的目标样本作为基准样本。最终实现其中的循环矩阵模拟样本平移，可以对训练样本实现密集采样，这样对各种采样的具体数量有一个较好的提升作用。在这种情况下，其中的分类器将会具备更强的实力来对目标以及其中的背景进行区分。最终保证对目标位置进行精准定位。

需要注意的是，这里使用的循环矩阵本身也有一定的不足存在，具体情况可以参考图 2-4 所示，在这个图中展现的是一个目标在垂直的方向上进行循环唯一的运动，其中可以很明显的看见有一条分界线，产生这种问题的原因在于边缘像素不连续。如果样本的运动方向从原来的垂直改为垂直加水平就会产生更多这样的分界线，这些分界线对于样本特征的提取来说是一种不小的干扰因素，由于这些干扰的存在，我们在跟踪的过程中不能保证得到可靠的样本，在进行频域转换的过程中使用的是傅里叶变换，存在这些干扰因素会出现频谱泄漏的情况，没有办法保证最终的运输结果可靠。出现这种情况，一般会添加窗函数来进行处理，这样便可尽可能地解决边界的不连续的问题。



图 2-4 垂直循环采样矩阵示例

常用的窗函数有余弦窗、汉宁窗、高斯窗等。KCF 目标跟踪算法中对提取的样本采用了乘上与余弦窗的处理方式。

$$\cos\_window = h(m) * h(n)' \quad (2-13)$$

其中， $m$  和  $n$  分别对应样本的宽和高， $h$  函数如(2-14)所定义， $N$  为余弦窗的尺寸。

$$h(N) = \frac{1}{2}(1 - \cos(2\pi \frac{i}{N})), 0 \leq i \leq N \quad (2-14)$$

跟踪样本在加窗后的可视化结果如下图 2-5 所示。

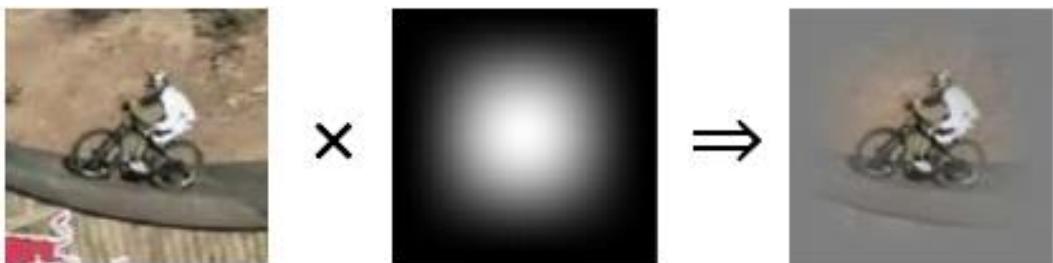


图 2-5 目标样本加窗效果对比

对于图 2-5 进行分析，这个图是目标样本在加了窗函数以后的可视化结果，从图中可以发现，窗函数对于边缘信息具有很好的抑制作用。同时还可以将中间的信息进行了突出，文献<sup>[42]</sup>对于窗函数的具体作用有一个比较详细的介绍。一句话概括窗函数的作用就是，将使用者感兴趣的信号进行提取，同时可以减少频谱泄漏的一种协助观察的方法。

图 2-5 加窗后的样本做垂直循环移位，可得下图 2-6：



图 2-6 加窗后的纯循环移位图片

其中的图 2-6 就是经过窗函数处理以后的图像，从这个结果来看，其中的循环移位样本可以体现出目标正确的运动情况。但是进一步分析就会发现，如果目标移动的速度过快的话，那么这个目标就会出现被分割的情况，表现出来的样本不能如实地表达目标的信息，就没有办法得到正确样本。如果目标处在搜索的区域以外，那么这个样本就不会将目标包括在内。针对这两种情况，一般会扩大样本的搜索范围。需要注意的是样本的描述范围情况是通过候选基准样本决定的，候选基准样本从模型的大小来看和目标的情况是一样的。在得到了目标具体大小情况以后，一般会使用余弦窗函数来进行处理，利用这个函数处理以后可以对于其中作用的具体区域以及形状固定情况进行明确，换句话说就是即目标大小的 $(1 + padding)$ 倍，同时其中得到的  $padding$  数值是比较小的。这个数据不能够太大，因为太大就会无法精确地对其中的模型情况进行确定，对后面的连续跟踪会有影响。一般来说，确定了目标的大小信息以后，其中使用的函数以及其中样本的搜索范围这两个信息也会被确定下来，这两个数据信息不能随意的进行改变，如果目标的移动速度比较快，搜索范围又没有办法进行覆盖，很容易出现跟踪失败的情况。在一些训练目标模型中也经常会出现这样的情况，比如目标在尺度上发生了改变，那么其中窗函数的具体区域就没有办法自动适应这些信息的变化情况，最终调整失败，影响后来的跟踪准确性。

图 2-7 是采用固定尺寸采样时，随着目标尺度改变，余弦窗对样本的加窗效果。可以看出来不同尺度的窗口对目标的有效信息的加强和周围背景信息的减弱是不一样的，过大的窗口可能会包含过多背景信息；而过小的窗口可能会遗漏部分目标信息。所以，选择合理的窗函数对基于相关滤波的跟踪方法也是不可忽略的一个环节。



图 2-7 不同尺度下的余弦窗口使用效果

### 2.2.3 训练分类器

使用相关滤波的第一步是训练分类器。一般对于样本的标记会使用连续标记的方式进行，也就是在利用一个 1 区间内的数值大小来判断物体和选框之间的距离信息，数值上越接近于 1，代表着目标距离越近，如果数值趋向于零，那么代表目标距离在变远<sup>[43]</sup>，当然也可以通过其他的函数对于其进行赋值，比如可以使用正弦函数或者是高斯函数对物体和选框之间的距离进行映射。

下面介绍一下损失函数(2-15)，损失函数可以看成是一种最简单的滤波算法，损失函数的性能需要大量的训练数据才能够不断的进行迭代优化。而这种迭代学习的模式使得损失函数有可能在学习的过程中产生过拟合的现象。过拟合指的是在对样本进行数学模型拟合的时候，为了得到与训练样本一致的假设而使模型变得过度严格而导致变得过拟合。这种情况可能会由两个原因导致，第一个原因可能是由于得到的模型维度过高，表现出过于复杂的特点。第二个原因可能是由于模型本身的泛化性能不好，就不能够很容易的对新数据进行适应，为了避免出现这种情况，会在权值向量中加入一个约束条件。一般的做法就是缩减(shrinkage)，因为这样就可以使得分类器得到一个较小的权值，我们常在这种情况下在等式的后面加上  $\lambda \|h\|^2$ ，这种处理方式被称为正则化，这项式子  $\lambda \|h\|^2$  就成为正则化项。正则化项中有一个重要参数  $\lambda$ ，它代表的正则化强度的大小，是一个标量，用来控制分类器得到的权值的大小。正则项约束权值的特性还可用于图像的去噪应用中<sup>[11]</sup>。

在(2-15)中的这个函数代表的就是岭回归的函数。其中可以设置一些数据来进行检验，一般会设置好训练样本以及期望输出值  $\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\}$ ，训练的最终目标是找到一个函数  $f(z)=w^T z$  使得代价函数(2-15)最小：

$$\min \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2 \quad (2-15)$$

其中,  $x_i$  代表第  $i$  个样本,  $f(x_i)$  表示分类器预测  $x_i$  的输出,  $y_i$  表示样本  $i$  的标签,  $\lambda \|w\|^2$  为正则项,  $\lambda$  为正则化参数,  $w$  为权重系数。

写成矩阵形式:

$$\min \|Xw - y\|^2 + \lambda \|w\|^2 \quad (2-16)$$

对  $X = [x_1, x_2, \dots, x_n]^T$  求导, 令其等于 0:

$$w = (X^T X + \lambda I)^{-1} X^T y \quad (2-17)$$

其中,  $w$  为权重系数,  $X$  是  $x_i$  的矩阵形式,  $X^T$  是它的转置,  $y$  是  $y_i$  的矩阵形式。

在样本采样的步骤中, 我们使用的是循环移位矩阵采样作为训练器的训练样本, 用单个采样构造整个训练集。

设:

$$X = [x_1, x_2, \dots, x_n]^T \quad (2-18)$$

$$X = [x_1, x_2, \dots, x_n]^T \quad (2-19)$$

$$P = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \ddots & & & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad (2-20)$$

则有:

$$PX = [x_n, x_1, \dots, x_{n-1}]^T \quad (2-21)$$

所以, 我们可以由单个采样样本构建出循环采样矩阵从而构成了整个数据集, 循环矩阵每一行右移后得到一个新的样本, 放入循环矩阵的另一行中。上文中提到, 循环矩阵拥有一个可以简化训练阶段计算量的十分有利于我们减少计算量的性质: 循环矩阵能被傅里叶矩阵对角化:

假设有离散傅里叶矩阵  $F$ , 用复数表示, 如(2-22)所示。其中  $K$  是矩阵  $F$  的尺寸。例如假设它的尺寸为 4, 则有(2-23)。

$$\omega = e^{-2\pi i / K} \quad (2-22)$$

$$F = \frac{1}{\sqrt{K}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega & \omega^2 & \omega^3 \\ 1 & \omega^2 & \omega^4 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega^9 \end{bmatrix} \quad (2-23)$$

更一般的，当  $K=n$  时：

$$F = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 1 & \omega & \dots & \omega^{n-2} & \omega^{n-1} \\ 1 & \omega^2 & \dots & \omega^{2(n-2)} & \omega^{2(n-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \omega^{n-1} & \dots & \omega^{(n-1)(n-2)} & \omega^{(n-1)^2} \end{bmatrix} \quad (2-24)$$

矩阵  $F$  中的每一行可以看成是一个旋转向量，这个向量从上到下的角速度越来越大。由于傅里叶矩阵拥有性质(2-25)，所以可以称其为一个酉矩阵(unitary)。

$$F^H F = FF^H = I \quad (2-25)$$

在原矩阵中两边乘以逆矩阵：

$$F^{-1} \cdot X \cdot (F^H)^{-1} = \text{diag}(\hat{x}) \quad (2-26)$$

再利用前述(2-25)中的酉矩阵性质：

$$F^{-1} \cdot X \cdot (F^H)^{-1} = F^H X F = \text{diag}(\hat{x}) \quad (2-27)$$

整理可得(2-28)：

$$X = F \text{diag}(\hat{x}) F^H \quad (2-28)$$

其中， $F^H$  是  $F$  的复共轭， $\text{diag}(\hat{x})$  表示它是一个对角矩阵， $\hat{x}$  如(2-29)所示。

$$\hat{x} = F(x) = \sqrt{n} F x \quad (2-29)$$

将(2-29)-(2-22)代入(2-17)中可以解得：

$$\begin{aligned} w &= (F \text{diag}(\hat{x}^*) \cdot F^H F \text{diag}(\hat{x}) F^H + \lambda I)^{-1} F \text{diag}(\hat{x}^*) F^H y \\ &= (F \text{diag}(\hat{x}^* \odot \hat{x} + \lambda) F^H)^{-1} F \text{diag}(\hat{x}^*) F^H y \\ &= F \text{diag}\left(\frac{\hat{x}^*}{\hat{x}^* \odot \hat{x} + \lambda}\right) F^H y \\ &= C(F^{-1}\left(\frac{\hat{x}^*}{\hat{x}^* \odot \hat{x} + \lambda}\right)) y \end{aligned} \quad (2-30)$$

其中， $C(x)=X$ ， $\hat{x}$  表示  $x$  的傅里叶变换， $\hat{x}^*$  与  $\hat{x}$  是共轭关系。 $\odot$  表示元素对应相乘， $\lambda$  代表正则化系数， $y$  是  $y_i$  的矩阵形式。

由于循环矩阵的卷积特性，即：

$$F(C(x)y) = x^* \odot \hat{y} \quad (2-31)$$

将等式两边同时做二维离散傅里叶变换，可以解得目标训练器参数  $\hat{w}$ ：

$$\hat{w} = \frac{\hat{x} \odot \hat{y}}{\hat{x}^* \odot \hat{x} + \lambda} \quad (2-32)$$

其中， $\hat{w}$  表示  $w$  的离散傅里叶变换。

再通过核函数的非线性映射，将  $w$  表示成关于  $x_{m,n}$  的高特特征矩阵的线性加权。这里直接放出结论：

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \quad (2-33)$$

其中， $\hat{\alpha}$  表示在频域中的系数矩阵， $\hat{y}$  代表频域中的样本， $\hat{k}^{xx}$  表示  $x$  在傅里叶域中进行自相关。

我们可以从化简结果中看出，化简过程中由于利用了循环矩阵对角化的性质，最终化简得到的式(2-33)中不包含矩阵  $K$ ，所以无需计算它，从而大大减少了计算复杂度。

## 2.2.4 快速目标检测

在进行目标位置检测的时候，首先需要进行的就是选择候选样本，这个候选样本一般会从搜索区域中进行选择。需要注意的是其中的分类器对于不一样的样本会有不一样的响应情况。一般选择一个输出最大的位置作为这个帧的新位置。接下来就是收集待检测样本集，一般的收集区域由两部分组成，一个是预测区域，另外一个是其循环移位。进行快速目标检测的时候，除了开始的一帧外，当前帧的中心点一般会从上一帧中检测得分最大的地方得到，其中具体得到检测得分  $\hat{f}(z)$  的公式如下。

$$\hat{f}(z) = \hat{k}(x_{m,n}, z_{m,n}) \odot \hat{\alpha} \quad (2-34)$$

其中， $f(z)$  表示图像块  $z$  经过分类器后的输出，而  $\hat{f}(z)$  则表示在傅里叶域中进行计算的结果。 $z_{m,n}$  表示候选目标块， $x_{m,n}$  表示上一帧中找到的目标， $\hat{k}(x_{m,n}, z_{m,n})$  表示候选目标块与上一帧中找到的目标进行相关运算， $\hat{\alpha}$  表示在频域中的系数矩阵， $\odot$  表示两个矩阵对应位置元素相乘。

通过(2-34)的运算，得到的是与上一帧相关且经过岭回归器分类后的最大响应位置处，我们这个位置认为是当前帧目标所在的具体位置。

KCF 跟踪算法如果从深层的逻辑来进行分析，可以发现其还是将对于目标的检测作为基础的信息进行处理，需要注意的是在这种算法中有两个步骤是一起进行的，一个是采样，还有一个是检测。其中采样指的是对于目标附近一定范围内进行循环采样，采样的主要的目的是为了获得一定数量的正负样本。在实际的使用中，首先需要做的是对于训练样本矩阵进行构建，为了能够适应目标外观的变化，跟踪器在对于每一帧进行跟踪的时候都会更新检测模型，更新模型的算法如(2-35)所示。

$$\hat{\alpha}' = (1 - \gamma)\hat{\alpha}^{-1} + \gamma\hat{\alpha} \quad (2-35)$$

其中， $\hat{\alpha}'$  为更新后的分类器系数的频域表示， $\hat{\alpha}^{-1}$  为频域中的跟踪器系数的逆， $\gamma$  表示更新系数。

更新系数主要的含义是，当前帧的信息以及之前学习到的信息在整体的模型中所占的比例。这种算法可以让背景得到简化，如果目标没有出现遮挡的情况或者是目标姿态变化很快的情况，都可以适用，同时结果具有良好的鲁棒，过程非常的高效，然而其中也具有一定的缺点，比如这个策略没有办法对于目标是否有遮挡的情况进行判断。同时不具备漂移矫正的方法，所以可能因为干扰因素的累计导致跟踪不成功。

但在实际目标的跟踪过程中可能会出现各种原因导致跟踪，比如遮挡问题时跟踪器学习遮挡处的纹理，造成模型被“污染”，或者如光照变化下，目标的外观造成巨大变化，跟踪器来不及学习，造成跟踪失败等情况。

## 2.2.5 模型更新

在 KCF 目标跟踪算法中，模型的更新是按照固定的学习率进行学习的，所以当面对遮挡、光照变化等复杂的目标跟踪场景时，算法学习到的模型很可能已经不是我们想要跟踪的目标，或者目标已经和模型的差异较大，就会造成跟踪精度下降甚至跟踪失败。一个良好的目标样本模型可以让跟踪器更好地跟踪目标，而选择什么时候更新模型就显得至关重要了，如果能判断当前跟踪结果是否可靠，然后再去决定是否要更新模型，将会大大提高跟踪算法的精确度和鲁棒性。在第四章中我们将会提出一种根据峰值旁瓣比计算跟踪置信度，进而决定是否要更新模型的目标跟踪算法，使得跟踪更为精确，鲁棒，更能适应复杂场景下的跟踪任务。

### 2.3 本章小结

对相关滤波类跟踪方法进行了简要描述，接着介绍了相关滤波跟踪算法的基础理论并对其中重要的理论和步骤进行了介绍。通过对相关滤波类跟踪器的介绍，阐述了它的结构和运作机制，对后文针对它的机制与不足进行跟踪算法的改进奠定了理论基础。

## 第三章 目标跟踪中的特征融合方法

### 3.1 运动目标跟踪中的特征

目标跟踪是计算机视觉领域中一个十分热门的研究方向。在计算机软件科学以及计算机硬件技术蓬勃发展的今天，除了我们一般认为的视频监控外，智能机器人和虚拟现实等这类研究方向也离不开目标跟踪技术的支持。但是受到复杂现实环境的影响即在实际应用中仍面临光照条件变化、被跟踪物体本身会发生非刚性变形、目标的尺度会发生改变等多因素造成的跟踪不稳定的困难<sup>[44]</sup>，目标跟踪算法在投入到实际的生产生活中时还存在这一定的局限。

为了解决上面描述的跟踪目标在受到光照条件变化、被跟踪物体本身会发生非刚性变形、目标的尺度会发生改变等这些跟踪场景的不确定因素导致跟踪失败的问题，本章将 HOG 和 CN 这两个具有不同特性的目标跟踪算法结合在一起。第一步是分别利用 HOG 特征和 CN 特征这两个不同的算法来采集两种不同的特征；第二步对上面两种算法得到的各自的跟踪结果，根据它们各自的跟踪效果分配不同的权重，以这种动态权重自适应地融合多个特征；第三步通过前面得到的结果利用依靠建立的不同的模型和它们各自在算法中所占的比重加以结合，通过算法预测跟踪目标下一帧出现的位置。本论文提出的目标跟踪算法的优点有以下两点：

- (1) 在复杂不固定的环境中，利用多特征更加具体和完善的对跟踪目标进行描述。
- (2) 在目标跟踪的过程中，作为跟踪依据的不同特征算法会根据自身，对跟踪目标的要求自适应地对目标的细节进行不断的细微的改变，用于下一步预测目标位置。

#### 3.1.1 HOG 特征

方向梯度直方图(Histogram of Oriented Gradient, HOG)特征描述算子在 2005 年的 CVPR 上被 Dalal 提出<sup>[45]</sup>，对视觉跟踪领域产生了巨大的影响。HOG 特征已经被用于各种各样的计算机视觉和计算机图像的各种应用场景中，例如在跟踪和检测中 HOG 特征描述算子已经被广泛使用，它的优点是可以对几何和光学的形变保持很好的不变性，或者说，对环境变化具有很强的鲁棒性。但是它对一个目标旋转前后或者尺度变化后得到的特征相差较大，也就是说，没有旋转不变性和尺度不变性。通过计算归一化后的 HOG 特征，提取跟踪目标的特征信息。它通过计算图像

一个个局部区域中的梯度和边缘信息，从一个个局部区域中统计成直方图，构成梯度直方图特征。HOG 的基本思想是待检测的目标的纹理和形状可以由它在图像中的边缘方向密度或梯度信息来描述出来<sup>[26]</sup>。

要获取 HOG 特征，第一步是将图像进行归一化。由于 HOG 特征提取的是纹理特征，颜色信息对它来说无法使用，所以将彩色图像转换为灰度。其次，为了提高特征对几何和光学的形变等干扰因素的鲁棒性，需要对图像进行归一化操作。也就是对图像进行 Gamma 校正，以完成对整个图像的归一化，来调节图像的对比度，来降低图像中局部光照和阴影造成的干扰，此外，也能够减少环境噪声造成的影响 [46]。

第二，需要计算图像像素的梯度。我们需要计算的梯度分为水平方向和竖直方向这两个方向上的梯度，并且计算出每个像素位置上的梯度方向和大小。假设图像中某个坐标为  $(x, y)$  的像素  $I(x, y)$  处的水平梯度  $G_x$  为：

$$G_x(x, y) = I(x+1, y) - I(x-1, y) \quad (3-1)$$

竖直方向上的梯度为  $G_y$  为：

$$G_y(x, y) = I(x, y+1) - I(x, y-1) \quad (3-2)$$

其中， $I(x, y)$  为图像中某个坐标为  $(x, y)$  的像素。

在实际中，要计算水平和竖直方向上的梯度，其实只需要使用水平模板  $[-1, 0, 1]$  和竖直模板  $[-1, 0, 1]^T$  对图像进行卷积，就可以得到。

在点  $I(x, y)$  处的梯度幅值  $G(x, y)$  和梯度方向  $\alpha(x, y)$  为分别为：

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (3-3)$$

$$\alpha(x, y) = \tan^{-1} \frac{G_y(x, y)}{G_x(x, y)} \quad (3-4)$$

其中， $\tan^{-1}$  为反正切函数， $G_x(x, y)$ 、 $G_y(x, y)$  分别为点  $I(x, y)$  处水平和竖直方向上的梯度， $\alpha(x, y)$  为梯度的方向。

第三，是计算每个细胞单元中的方向梯度直方图：将图像划分为一个个更小的单元，或者叫做细胞单元(Cell)，将上一步计算得到的梯度方向投影到  $0\sim180^\circ$  的范围内。将计算得到的梯度幅值作为投影的权值。用梯度的方向决定这个细胞应该向哪一维度进行投影。细胞单元中的梯度方向以多少角度等分，即把所有梯度方向分为多少个 bin，则可得到多少该细胞方向梯度向量的维度。例如，将梯度方向以 20 度等分，则细胞单元的 HOG 特征是一个 9 维的向量。如下图 3-1 是一幅图像的一个细胞内的方向梯度直方图。

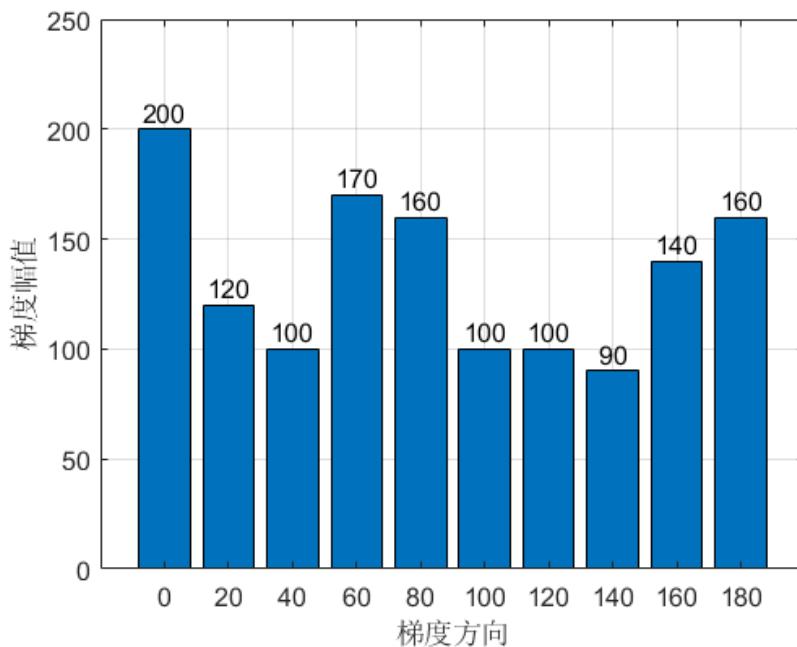


图 3-1 一个细胞单元内的方向梯度直方图

第四，由于局部光照可能会变化，以及前景背景的对比度也有可能会产生变化。所以每个细胞单元的梯度强度变化范围就会十分大，所以采用梯度强度归一化的方法进一步对光照、阴影和边缘进行压缩。由于 L2 范数(L2-Norm)简单且在检测中效果相对较好，故使用 L2 范数针对每个块进行对比度归一化先对每个细胞单元内的方向梯度直方图后进行统计，再将他们联合起来，得到更宏观的梯度方向直方图：将每个细胞单元的梯度直方图作为每个细胞单元的描述子，由胞组成的更大的描述子，称为块(Block)。一个块由四个相邻的细胞中的特征向量串联起来，作为这个块的梯度方向直方图。假设细胞单元的梯度方向向量有 9 个维度，则块的维度有  $4 \times 9 = 36$  个维度。

第五，最后一步就是将图像内所有的块的 HOG 特征描述子串联起来，就可以得到整幅图像的 HOG 特征描述子了，这也就是最终分类时使用的特征向量。

在 HOG 特征的提取中，细胞层面的梯度计算相当于局部特征提取，块是比细胞层面更大的局部特征，而块与块之间是有重叠的，所以每个细胞的直方图都会被重复多次被用于最终的整幅图像的直方图的计算中。

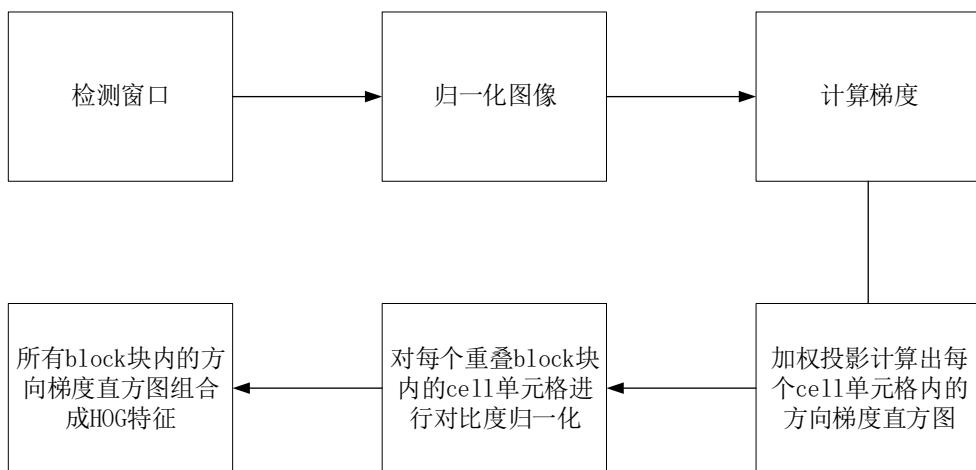


图 3-2 方向梯度直方图提取过程流程图

相互重叠可以区分出更细的局部特征，最后由于图像中还有局部光照和背景的变化，再对块内的特征进行归一化。这样的冗余操作显著提升了 HOG 特征的性能。小尺度梯度、精细的方向采样、相对粗糙的空间采样，和重叠描述子块中的高质量局部对比度归一化都对 HOG 特征描述子的性能至关重要。

HOG 特征通过运用梯度或边缘的方向密度分布统计信息，可以采集图像的纹理信息，如图 3-3 中展示的就是人脸信息经过 HOG 特征采集后可视化的图像，从可视化信息可以看出来 HOG 特征保留了纹理信息，而颜色等信息被省略了。

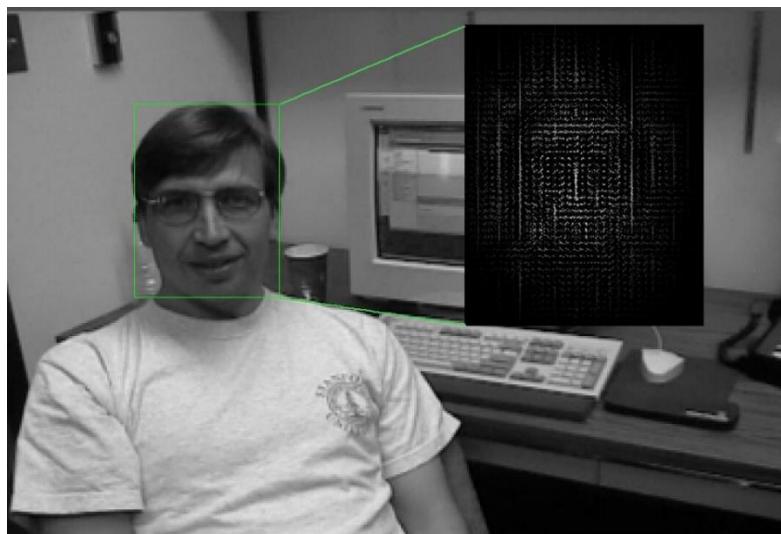


图 3-3 HOG 特征采集的人脸纹理

### 3.1.2 LBP 特征

局部二值模式(Local Binary Pattern, LBP)特征是一种可以有效的描述纹理的特征，它是一种有着灰度和旋转不变性的优质特征。近年来，随着 LBP 算法的不断完善和发展，其应用已经被大大开发，特别是在纹理分割、纹理分类等方面起到了至关重要的作用。LBP 算法与其它算法最大的不同点在于它的侧重点是提取局部特征。LBP 算法从检测目标的局部特征<sup>[47]</sup>上提取到不同于其他物体的局部不变量，因此在整个跟踪目标的过程中可以对因由于目标外观的变化(如光照、遮挡)，而造成的干扰情况有着比较好的排除效果。LBP 算法以监测目标的纹理作为重点，再通过其纹理从上面提取处目标的特征信息。上面多次提到的纹理，是以图像的颜色和灰度这类存在于跟踪目标外观表面的特征作为判断的基本点，来对影像文件进行分析的一种定义。纹理特征重点针对的是跟踪目标图像在灰度上的特征不同点，因此纹理特征不光可以反映图像像素点自身的情况，还可以对所测目标的像素点的周围环境像素点进行分析。纹理特征不仅反映图像的微观特性，还可以对目标整个环境的整体信息进行一定的描绘。正是由于纹理特征有着诸如此类的优点，越来越多的研究者选择纹理特征作为判断依据来对跟踪算法加以研究。

#### (1) 标准 LBP 算法

LBP<sup>[48]</sup>算法的原理可以看成如图 1-2 所示的一个九宫格，我们将九宫格中的中心位置所标示的像素大小作为基准点，和周围 8 个位置的像素灰度值进行对比，如果该位置的像素灰度值大于中心的基准点，我们就将其标记为 1；反之若该位置的像素灰度值比中心基准点小，则标记为 0。

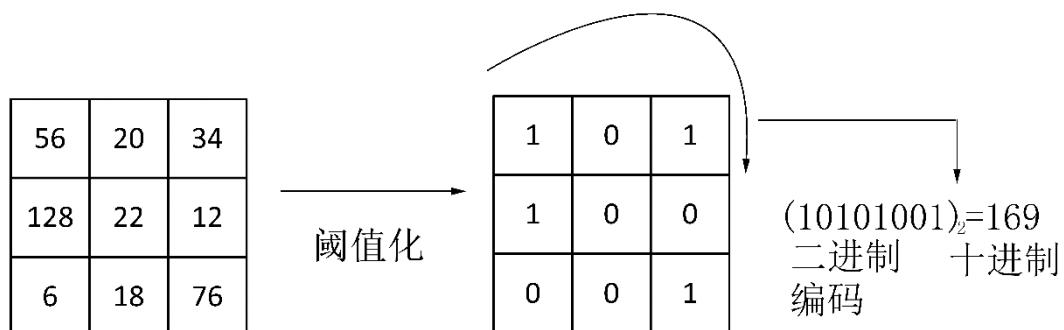


图 3-2 基本 LBP 算子示意图

以此我们就可以在原先九宫格的基础上建立一个新的九宫格，在以除中心位置以外的 8 个位置从最左端顺时针取值，设立一个新的二进制数值，即为 10101001，再通过十进制转化成具体数值数，这个数值就是中心像素的 LBP 值，即为 169。九宫格的 LBP 值可以通过公式(3-5)进行计算：

$$LBP(x_c, y_c) = \sum_{k=0}^7 Z(I_k, I_c) 2^k \quad (3-5)$$

其中,  $I_c$  是九宫格中心位置像素点的灰度值,  $I_k$  除去中心位置以外周围 8 个 是像素点的灰度值,  $Z(I_k, I_c)$  可以由公式(3-6)计算。

$$Z(I_c, I_k) = \begin{cases} 1, & I_k - I_c > 0 \\ 0, & I_k - I_c \leq 0 \end{cases}, k = 0, 1, 2, \dots, 7 \quad (3-6)$$

## (2) 改进的 LBP 算法:

在上面提到的初始 LBP 算法中, 我们不难发现该算法中是只具备了  $3 \times 3$  方形的 9 个像素灰度点的九宫格, 但这也恰恰反映了其具备包含观察范围小的缺点, 因此采用这种初始方式所得到的信息是有限的。为了改善这个情况, 得到更加丰富的目标信息, Ojala 等人提出了一种改进的方法, 即放弃以正方形作为取值的标准范围, 采用任意半径的圆这种涵盖的数据量更大的方式来采集数据, 且采集的空间更大<sup>[49]</sup>。

改良的 LBP 算法可以采取  $(P, R)$  来表示,  $P$  是与中心相邻的像素灰度点的数量,  $R$  是以一个像素灰度点作为 1 的圆的半径。如图 3-4:

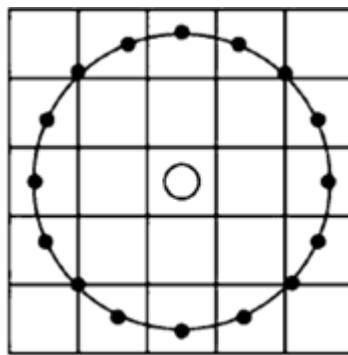


图 3-4 圆形(8,2)邻域的 LBP 算法

图 3-4 为一个半径为 2, 相邻像素点数量为 8 的圆形, 每个方格都是表示为一个像素灰度值; 对于部分不再方格中央的黑点(倾斜 45 度方向), 可以通过双线性插值方法依据周围的数值对其进行求解。把采取这种方法求出的 LBP 特征记为  $LBP_{P,R}$ 。

在图 3-5 所示中, 将不同位置的灰度值用  $I_{i,j}$  来表示, 其中  $i$  和  $j$  分别表示第  $i$  行, 第  $j$  列的位置, 来指明该灰度值所对应的位置。其中除圆心外的圆点表示较重要的 8 个领域,, 除去可以直接得到的点外, 还有部分点不在方格中央, 这个时候就可以利用双线性插值计算对其周围 4 个像素点(即 4 个空心点)进行计算。

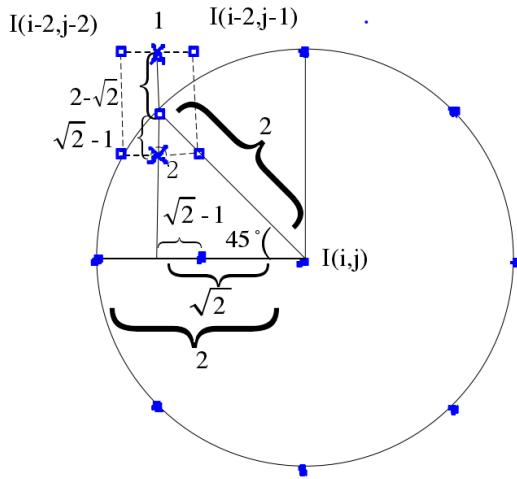


图 3-5 双线性差值求取在不同像素中心位置领域点的取值

根据双线性插值规则，需要先算出图中 1 和 2 位置的水平插值，其中点 1 处的值是通过和它在同一行的  $I(i-2, j-2)$  和  $I(i-2, j-1)$  的线性插值得到：

$$Value(1) = I(i-2, j-2) + (2 - \sqrt{2}) \times (I(i-2, j-1) - I(i-2, j-2)) \quad (3-7)$$

其中， $I_{i,j}$  表示第  $i$  行，第  $j$  列的位置的灰度值

类似地，可以求出点 2 位置处的值：

$$Value(2) = I(i-1, j-2) + (2 - \sqrt{2}) \times (I(i-1, j-1) - I(i-1, j-2)) \quad (3-8)$$

计算出点 1 和点 2 位置处竖直线性插值计算得到的结果：

$$Value = Value(1) + (2 - \sqrt{2}) \times (Value(2) - Value(1)) \quad (3-9)$$

结合上面的计算公式，就可以计算出灰度值(不处于中心像素点位置)，在和中心点的灰度值作为基准加以对比，若该像素点的灰度值大于中心像素值，用 0 表示；反之，用 1 表示。得到新的二进制结果后将每个邻域点给予一个权值之后，将所有结果加在一起，得到的即  $LBP_{8,2}$ ，就是中心像素点的局部纹理特征值的大小。

我们可以发现，标准 LBP 算法与扩展 LBP 算法基本上都是采用了一样的原理，只是取值的方法略有不同，如图 3-6 所示通过改变取值半径，可以得到不同尺度的 LBP 算子。扩展的 LBP 算法的优点有：表达方式灵活性强、准确性高、监测的范围大和传达的信息量更大。

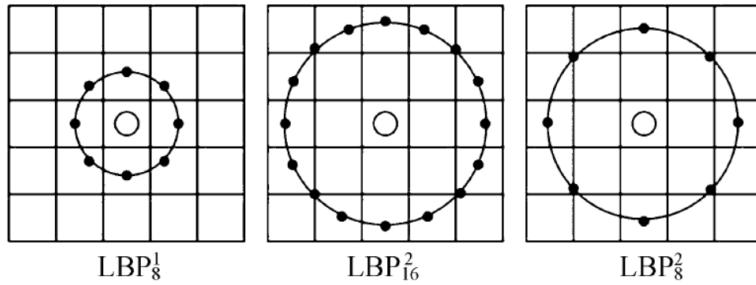


图 3-6 几种不同尺度的 LBP 算子

### 3.1.3 HOG 特征与 LBP 特征性能分析

上面介绍了两种特征, LBP 和 HOG 特征较多的被应用于目标检测和跟踪领域行业, 都获得了不错的效果。这两种特征都是手工设计出来的且它在所有特征里是性能较好的, 下面我们具体对两种特征自身存在的优缺点进行分析和总结, 并讨论如何更好地使用这两种特征。

HOG 特征是一种获取图像方向梯度直方图的特征描述子。在目前的研究领域里, 我们可以结合目标边缘的方向密度分布来描述目标的具体形状。提取 HOG 特征步骤为: 先把图像平均划分为不同的基本单元, 然后得到基本单元中各个像素的边缘方向直方图, 再统计和分析所有直方图已获得具体特征。各个 cell 的方向直方图足够能具体反映各个小块区域的特征信息, 同时不足以对大块区域进行表征。

根据以上提取 HOG 特征的流程, 光照的变化对这个描述子的抑制性比较强, 然而过程中对遮挡的光照变化问题处理不到位, 而且没有选取主动的方向去提取梯度方向的相应的信息。所以该 HOG 特征的提取方式没有旋转不变性。

LBP 特征是记录图像小部分信息的特征描述子。原始的 LBP 特征提取把模板信息转化为 8 个二进制的数。所以把这八个二进制数组合在一起并换算成一个十进制的数, 这个十进制的值就是模版对应的 LBP 值。一般 LBP 值大小为 0 到 255, 可以通过统计和分析所有模版的 LBP 值来构造对应直方图的信息。

相对的该特征的收集有这些不足: 首先由于模板是固定的所以会导致解析出来的局部纹理的信号与实际需求有所出入; 然后由于 LBP 特征描述子长度固定且其维数随着采样点增加而增加, 这会大幅度的增加计算量。对于第一个缺陷的一种解决方法是把  $3 \times 3$  的固定模版延伸到任意形状任意大小的模版同时选取模版内的部分进行计算。LBP 描述子通过这种方式就获得了一定的灰度不变形; Maenpaa 等人<sup>[50]</sup>提出对得到的  $a$ (对应得到的任意二进制数)个二进制数进行不同的排序并构成不同的  $a$  位二进制数, 把其中最小的  $a$  位二进制数取小值作为 LBP 值。LBP 描述子通过这种方式就获得了良好的对目标旋转的鲁棒性。Ojala 及其团队利用一种等

价模式对特征进行降维。

总之，LBP 特征灰度不变性和旋转不变性方面占据优势。HOG 特征在复杂的背景、光照强度变化、目标不清晰的环境因素影响条件下表现更佳。所以分别结合它们的优缺点，联合利用 HOG 特征和 LBP 特征进行目标的跟踪定位。

### 3.1.4 颜色特征

颜色名字(Color Name, CN)特征的出现使得近几年来目标分类和目标跟踪领域得到了不小的突破和成就。Berlin 和 Kay<sup>[51]</sup>认为颜色特征是对日常生活里出现的不同颜色赋予不同的标签。比如英语中用 11 个不同的单词来定义不同的颜色，这是由人类定义的用来表达在客观世界中人类通过肉眼直接观察得到的事物的外观颜色，这些最基本的颜色可以用来作为颜色语义特征的 11 维特征通道，又叫做语义颜色标签。

由于图像中的颜色分配是在像素层面进行的，所以需要一种方式把可以用到的颜色特征和计算机里的 RGB 值一一对应起来。用一个矩阵来将标准的 RGB 三通道映射到颜色名字特征空间内。也就是所谓的概率潜在语义分析(Probabilistic Latent Semantic Analysis, PLSA)方法<sup>[52]</sup>。从 Google 得到含有这 11 种颜色的 250 张图像并对这些图像进行训练操作得到每个基本 Cell 属于 11 个对应颜色标签的概率值，经过统计整合就得到了这些颜色名字概率字典。

PLSA 是一项来自自然语言处理领域迁移到计算机视觉领域的语义分析方法，自然语言处理领域中采用概率的形式表示计算的过程从而获得目标文档里的潜在内容。如果存在一组图片集  $D = \{d_1, d_2, d_3, \dots, d_n\}$  的文档，一个确定的字典单元集合  $W = \{w_1, w_2, w_3, w_4, \dots, w_n\}$  为词汇的组合，颜色名字  $Z = \{z_1, z_2, z_3, z_4, \dots, z_n\}$  是潜在主题集合。联合概率分布  $p(d, w)$  被定义为：

$$p(d, w) = p(d)p(w|d) \quad (3-10)$$

$$p(w|d) = \sum_{z \in Z} p(w|z)p(z|d) \quad (3-11)$$

其中， $p(z|d)$  指的是文档  $d$  出现时， $z$  出现的可能性，也就是条件概率， $p(w|d)$  指的单位时间内单词  $w$  在文档里  $d$  出现的条件概率， $p(w|z)$  中， $z$  是潜在内容在给定词典中出现的概率， $p(z|d)$  表示文档  $d$  里潜在主题  $z$  的分布概率密度， $w$  是字典里的词汇。

通过最大期望(Expectation-Maximization algorithm, EM)算法，使用极大似然原则来估计潜在变量模型参数，得到的后验概率如下：

$$L(\theta) = \sum_{d \in D} \sum_{w \in W} n(d, w) \log p(d, w) \quad (3-12)$$

其中,  $L(\theta)$  是似然函数,  $n(d, w)$  表示某篇文档  $d$  中的词  $w$  出现的次数,  $p(d, w)$  就是观测到的变量  $d$  与  $w$  的联合概率分布。

PLSA 训练方法中, 数据文档  $d$  代表图片的信息, 在 LAB 颜色空间创建直方图,  $w$  对应 LBA 的数据库直方图对应的列向量里的每一个值,  $p(w|d)$  单位时间内该值在数据文档里出现的次数。图像的颜色属性则与潜在主题  $z$  相关,  $p(z|d)$  表示每幅图像颜色属性的混合系数。机器在图像进行处理时, 在进行训练时, 通过最大期望算法给  $p(w|z)$  和  $p(z|d)$  赋值, 最大化期望如下:

$$p(z|w, d) \propto p(z|d)p(w|z) \quad (3-13)$$

其中,  $p(z|w, d)$  是已知在观测  $d$ ,  $w$  下  $z$  的概率分布。

在最大化过程中, 已知  $n(d|w)$  后, 应用拉格朗日乘数发去更新  $p(w|z)$  和  $p(z|d)$  的最大后验概率估计得到如下:

$$p(w|z) \propto \sum_d n(d, w) p(z|w, d) \quad (3-14)$$

$$p(z|d) \propto \sum_w n(d, w) p(z|w, d) \quad (3-15)$$

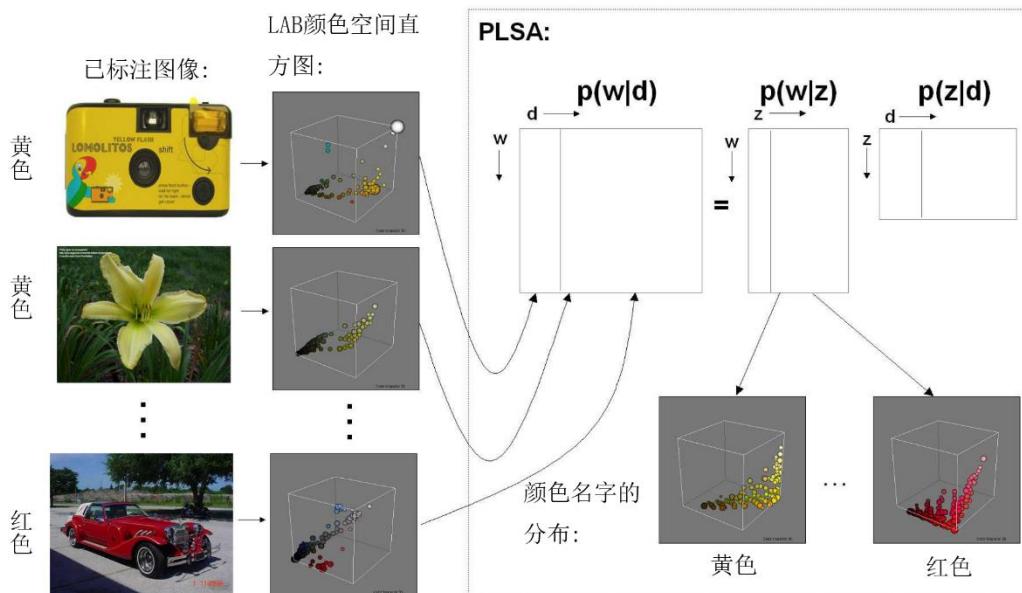


图 2-7 CN 特征建模过程

根据 Google 的包含 11 种颜色的与图片中的颜色属性进行对比，做出的分布概率图，采集到的像素颜色重合度最高的颜色  $p(z|d)$  概率值最大，即被分类到该种颜色中去，完成颜色特征的采集。颜色特征与之前描述的 HOG 特征、LBP 特征相比，颜色特征完全不描述目标的形状和纹理，只描述目标的颜色信息，在目标发生形变、旋转、或者发生较大尺度变化，或者是运动模糊<sup>[53]</sup>，分辨率较低的情况下，颜色名字特征对这些变化都有很好的适应性，可以很好地和这两种特征形成互补关系，得到更好的跟踪效果。

### 3.2 特征融合方法

本节之前的部分介绍了 HOG 特征、LBP 特征、CN 特征三种重要而又能相互补充不足的特征，并对他们的原理，采集方式进行了描述，得到了他们各自的优点和缺点，本文将用一种特征融合的方式将他们三种特征联合使用，相互协力完成更好的跟踪效果。

目标跟踪中的目标特征融合在追求实时性的算法中应用不广泛，因为在实际的生产生活中对跟踪算法最为看重的性能是目标跟踪算法的实时性。在目标跟踪算法中采取特征融合的方法可以加强算法的鲁棒性，提高目标的跟踪精度，但是会一定程度上增加算法的复杂度，降低实时性<sup>[54]</sup>，所以也有将特征进行选择后使用的方法<sup>[55]</sup>。

常规图像匹配模式主要利用了像素的灰度信息和形状信息<sup>[56]</sup>，而当目标处于比较复杂的环境中时，因为单个特征涵盖的目标信息太少<sup>[57]</sup>，这常常导致对目标跟踪的不够准确甚至丢失。所以很有必要开发一种能够整合多个特征信息的方法。对多个特征信息进行融合是指搜集和分析多个特征的信息，并让这些信息相互之间补足来完整的描述目标。针对在复杂的场景下的目标跟踪开发出了一种新的高效的描述方式：多特征描述。多特征描述主要采用两种方式：一、结合利用不同的特征描述目标外观构成多特征；二、对相同特征采用固定的分块个数构成多特征。大部分情况下，多特征描述方法可以提升一定的跟踪算法的性能，但这种方法也存在如下缺陷：

- (1) 目标特征信息会因为图中目标的颜色过于单一而发生冗余。
- (2) 通过单一特征的信息来对目标进行全面的描述会因为画面中的目标处于过于复杂的背景时出现较多的误差。

目前，融合特征的方式方法很多，比如加权特征融合、最小最大规则融合还有乘性特征融合的方式方法。也有利用多聚焦图像清洗区域和模糊区域进行图像融合的方式<sup>[58]</sup>。在实际情况中，我们应用不同的特征融合方法能够得到不同的效果。

例如常见的加权融合系数的方式，根据不同的场景采用不同的特征，从整体来看，虽然它可以提高目标跟踪的准确度，但会使计算变得更加复杂，它的权值计算如式(3-16)所示，使用加权融合的模式；再比如另一种特征融合方式，其式子如式(3-17)所示，它采用的是直接相乘融合的方式，这种方法虽然比前者简单的多，但也会给画面增加噪声。

$$w_i = a_1 \cdot w_i^{e1} + a_2 \cdot w_i^{e2} + \cdots + a_n \cdot w_i^{en} \quad (3-16)$$

$$w_i = \prod_{j=1}^m w_i^{c_j} \quad (3-17)$$

其中， $w_i^{e1}, w_i^{e2}, \dots, w_i^{en}$  分别代表  $n$  种特征。 $a_1, a_2, \dots, a_n$  表示对应的特征加权权重系数，由于它们是权重系数，所以它们必然满足  $\sum a_i = 1$ 。

最小最大规则特征融合和乘性特征融合方式可以表示为公式(3-18)和(3-19)所示。

$$p = \sum_{i=1}^m \alpha_i \cdot p_i \quad (3-18)$$

$$\alpha_i = M \setminus \sum_{i=1}^m M_i \quad (3-19)$$

其中， $p_i, i=1, 2, \dots, m$  指的是特征  $i$  的概率分布模型， $M, i=1, 2, \dots, m$  指的是特征  $i$  的总频数， $\alpha_i, i=1, 2, \dots, m$  表示特征  $i$  的权值。

对于特征融合中各特征的权值难以确定的问题，本文将不采用固定各特征权值的方式。由于本文的算法试图解决的是在复杂场景下的跟踪问题，由于复杂背景噪声、干扰较多，会增加目标跟踪的难度<sup>[59]</sup>，通常的结合许多先验(3-20)信息得到的固定特征权值模式的特征融合方法将很难适应多变的跟踪任务。

本文采用一种基于特征对跟踪效果贡献的动态权值特征融合方法，根据特征跟踪目标后计算出来的跟踪置信度 PSR 来决定该特征在最终跟踪结果中所占的权值。跟踪置信度 PSR 的计算见第三章中的式(3-23)所示。本文算法将用三种特征独立对目标进行跟踪后计算出跟踪置信度，再根据跟踪置信度动态调整特征权值。最后再利用各特征跟踪所输出的位置加权得到最后的融合过后的目标位置。其中最后输出的目标位置  $P$  可以用公式(3-21)表示：

$$P = \eta_c P_c + \eta_h P_h + \eta_l P_l \quad (3-21)$$

其中， $P_i$  为各个特征输出的目标位置， $\eta_i$  为各特征的权值，可以由(3-22)得到。

$$\eta_i = \frac{psr_i}{psr_c + psr_h + psr_l} \quad (3-22)$$

其中  $psr_c$ 、 $psr_h$ 、 $psr_l$  分别为颜色名字、HOG 特征、LBP 特征单独跟踪得到的跟踪置信度。多特征融合模块的流程图如图 3-8 所示：

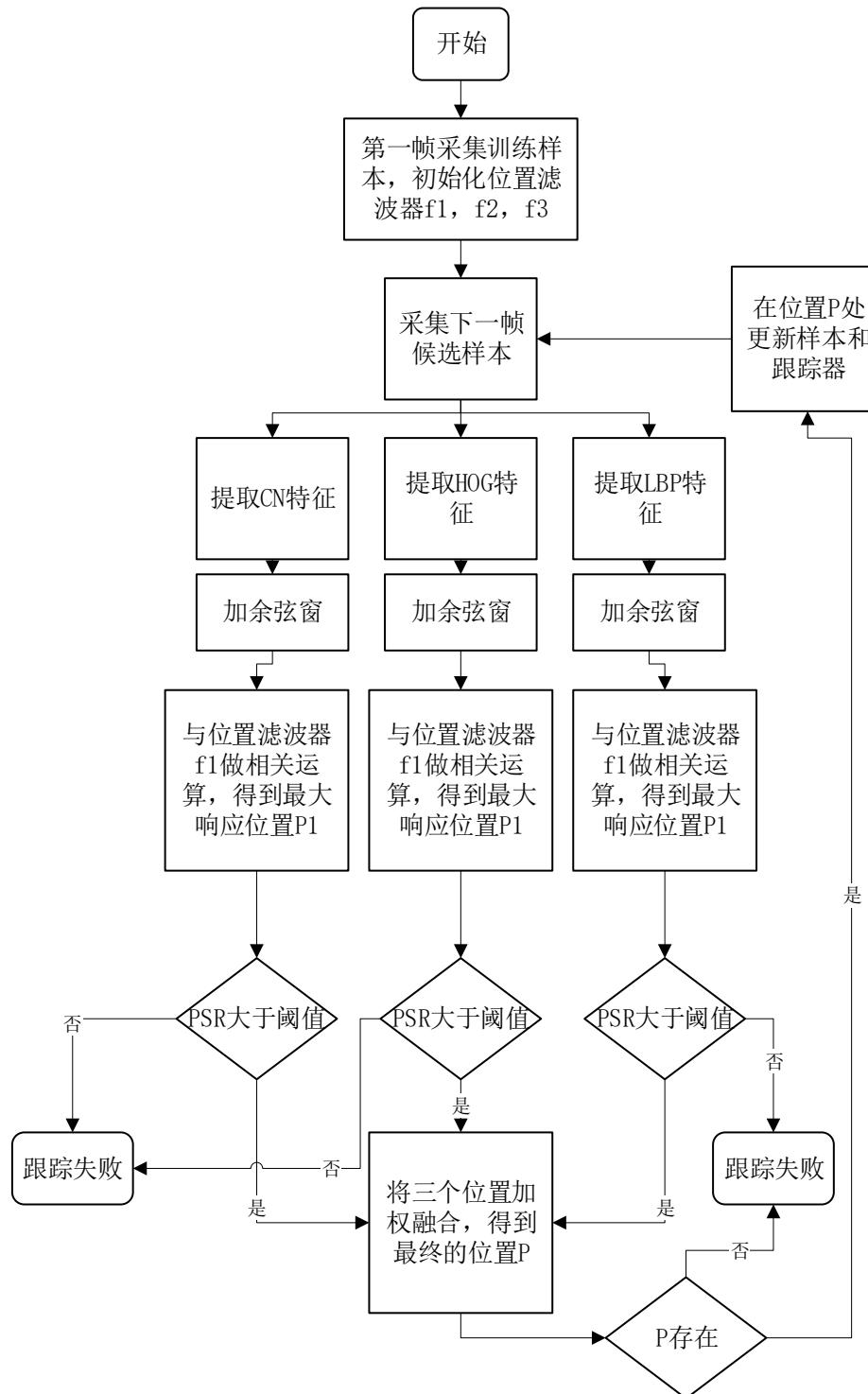


图 3-8 多特征融合流程图

本论文中的 HOG 特征的一个细胞单元的尺寸为  $4 \times 4$  像素，维数为 31；LBP 特征采用的算子是  $LBP_{8,1}^{u^2}$ 。LBP 特征的采样点有 8 个且  $R_{\text{像素}}=1$ 。和前者相同，cell 结构单元的像素为  $4 \times 4$  的结构，用来采集各小块直方图的信息。由于多次计算了各种特征，所以计算量较大，故采用了尺度适中的特征来平衡跟踪精度和计算量。

### 3.3 本章小结

本章对目标跟踪的算法的主要方法进行了概述，然后对目标跟踪算法中的特征进行了简要的介绍。然后具体介绍了几种重要的特征：HOG 特征、LBP 特征、CN 特征，并对他们的性能进行了分析，最后提出了一种融合多特征进行跟踪的特征融合方法。为后续整体算法所使用的特征进行了铺垫。

## 第四章 基于重检测的相关跟踪算法设计及实现

对目标跟踪技术不管是在军事还是民事上都有举重若轻的作用。近年来，对目标跟踪问题有广泛的研究，并取得了很大进展，有很多目标跟踪算法涌现。但是现实生活中的目标跟踪受遮挡、光照变换、目标与背景相似、目标背景复杂等因素的影响，目标跟踪算法的准确性和实时性受到了一定的限制。然而遮挡问题却是研究的重点和难点，尤其是目标受到严重遮挡或者完全遮挡时目标遮挡问题按目标被遮挡程度可分为三类，部分遮挡、严重遮挡、完全遮挡。遮挡处理一般分为三部分：进入遮挡、遮挡保持、离开遮挡。

目前，对于遮挡问题的处理有两种比较有效的解决方案，方案一，在目标出现了部分遮挡的情况时，利用局部特征匹配的方法进行处理，实现对目标的再次定位，如果遇到严重遮挡或者是完全遮挡的情况，将会利用重新检测的方式来对被跟踪目标的具体位置进行二次判别。方案二，使用的方案是对于轨迹预测，首先是在目标没有被遮挡的情况下就进行计算得到相应的轨道预测，然后在重新进入遮挡以后利用跟踪算法来进行预测，这种预测一般会使用到峰值旁瓣比，具体的预测方法是比较跟踪成功阈值和峰值旁瓣比的数值大小，如果前者小于后者，那么就认为跟踪成功。经过借鉴红外弱小目标中的检测方法<sup>[60]</sup>，并观察了 KCF 跟踪算法的许多实验数据后，我们发现目标重新进入遮挡的情况下会导致漂移，这也是这个算法的不足之处。这是因为这个算法不具备遮挡主动检测机制，没有这项机制的帮助，在遮挡以后跟踪器还是会进行更新，由于导入了错误的信息，导致跟踪失败。

针对上述问题，本章算法根据跟踪-学习-检测 (Tracking-Learning-Detection, TLD)跟踪器的检测部分进行改进，整个算法包括三个部分：相关滤波跟踪器模块、多特征融合策略模块、级联分类器重检测模块。多特征融合策略模块利用第三章算法来采集和融合，采用第二章描述的相关滤波跟踪器进行跟踪后计算跟踪置信度。最后采用级联分类器在跟踪失败时进行重检测。

由于 KCF 核相关跟踪滤波器本身不能得到自身跟踪效果的质量评价，在遇到目标遮挡时会一直学习背景，导致跟踪失败，故本文在这种方法中增加了遮挡主动检测机制，基于上述考虑，本文设计了如下的算法流程。

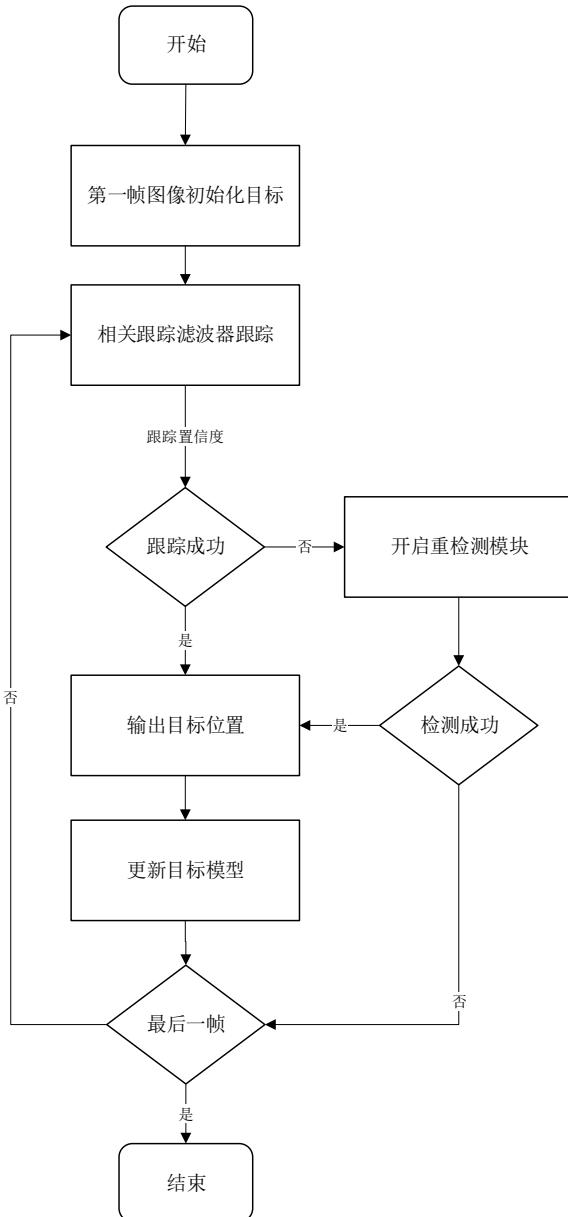


图 4-1 本文算法流程图

#### 4.1 基于峰值旁瓣比的跟踪置信度

上面提到的遮挡包含了有三个遮挡模式，一种是进入遮挡，一种是遮挡保持，还有一种是离开遮挡，具体情况可以参考图 4-1 所示，其中的 505 帧位置展示的情况就属于进入遮挡的典型情况，其中的 517 帧展示的就是典型的完全遮挡情况，其中的 571 帧位置展示的就是目标出遮挡的典型情况。其中需要注意的是完全遮挡，这意味着目标信息将完全丢失，也正因为得不到目标的相关信息，很容易导致跟踪失败，同时得到的跟踪响应图表现出来的也是最小值。所以要解决跟踪遮挡这

个问题的关键主要在于对不同类型的遮挡进行判断<sup>[61]</sup>。



图 4-2 目标被遮挡的情况

本文文中对于判断目标是否被遮挡这个问题采用的测量方法是峰值旁瓣这个指标来衡量跟踪置信度。其公式如下：

$$PSR = \frac{p - \mu}{\sigma} \quad (3-23)$$

其中的  $p$  代表的是最大峰值，通常情况下会将这个数据周围的一部分区域称作为旁瓣，用  $\mu, \sigma$  来表示旁瓣的方差以及旁瓣的均值。

在实际的算法使用过程中我们会设置一个阈值  $T$ ，将大于这个阈值  $T$  的跟踪判断为成功的跟踪帧，小于这个阈值的看做失败的跟踪帧。

我们可以看到采用 PSR 来做跟踪置信度的指标相对于采用相关跟踪器利用自身模板与图像相关后得到的响应图的最大输出值相比，机械地根据响应图最大输出的大小作为跟踪置信度的判据无法告知当前跟踪的状态，因为响应图的输出不是一个相对的值，它的值的大小只能表示当前模板和图像的相似程度，在模板被污染后，甚至时常出现被遮挡时反而响应图的峰值输出比没有遮挡时大。而 PSR 可以有效地告知跟踪器当前的跟踪质量，因为它的本质是峰值与附近的比值，是一个可以被归一化的量，所以可以借此判断跟踪质量，作为跟踪置信度的判据。当跟踪置信度低于阈值时停止模板的更新，防止了模板学习背景，从而避免了跟踪器的模板被“污染”情况的发生。

如下图 4-5 图 4-6 所示，PSR 作为跟踪置信度可以有效地判断目标是否被遮挡，并且是一个比值量，是一个相对的值，对参数较为不敏感，而峰值响应是相关操作后的直接输出，其范围不能确定且变化较大，对参数十分敏感。

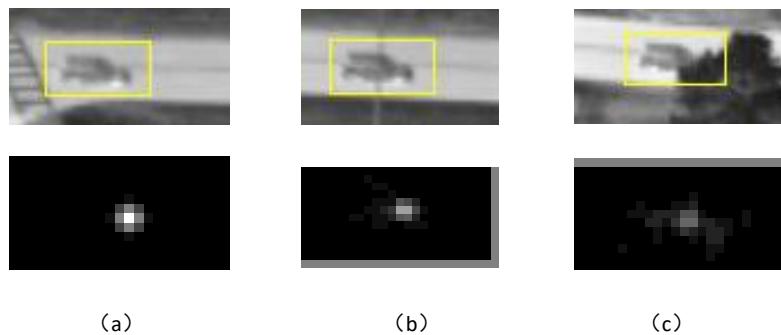


图 4-3 目标被遮挡时响应图的变化情况 (a)为目标没有遮挡时的响应图 (b)为部分遮挡时响应图 (c)为遮挡较为严重时响应图

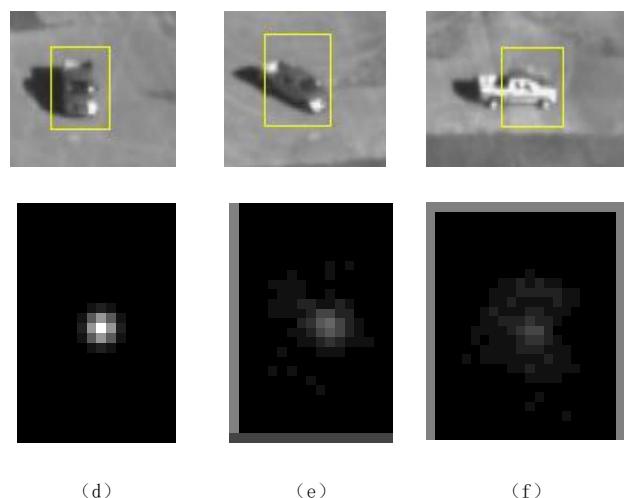


图 4-4 目标被遮挡时响应图的变化情况(d)为目标一开始时响应图 (e)为目标发生轻微形变时响应图 (f)为目标发生较大形变时响应图

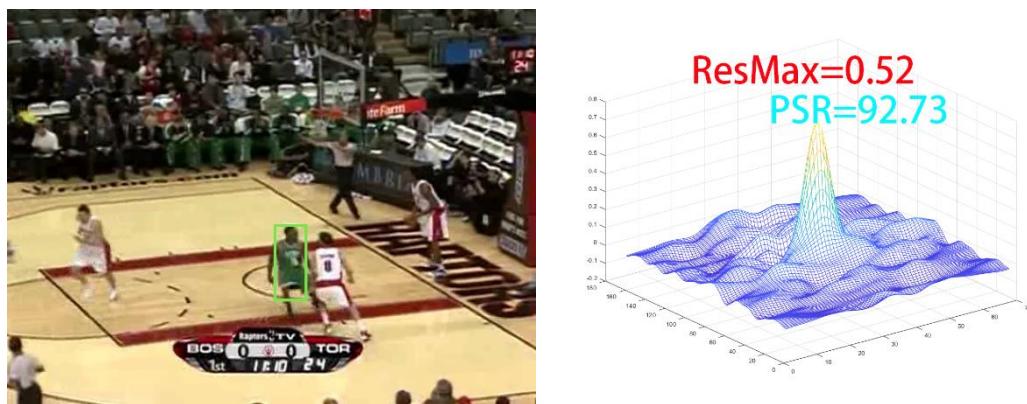


图 4-5 跟踪良好情况下的跟踪置信度和峰值响应

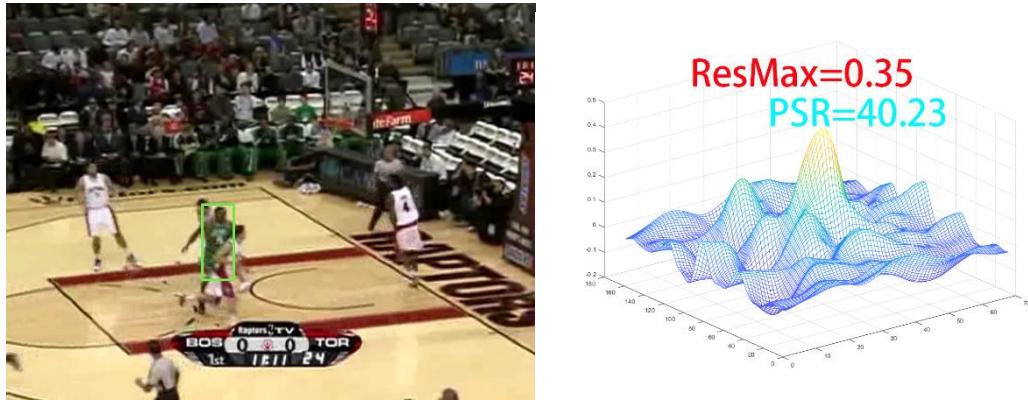


图 4-6 目标被遮挡时的跟踪置信度和峰值响应

## 4.2 被遮挡后的目标重检测方法

从第三章相关滤波类跟踪的理论可以知道，相关滤波跟踪里的每一帧其实都对目标进行了检测，通过寻找最大响应所在的位置，才得到的每一帧的目标位置。所以相关滤波跟踪器在进行跟踪的过程中也包含了检测的过程。但由于遮挡或其他原因，跟踪器失效后，就需要一个更“强”的检测器来重新检测目标，在成功重新检测到目标后，重启跟踪器。值得注意的是，虽然 CSK、KCF 等相关滤波类算法都对目标进行了检测，但是他们都不对跟踪结果进行可靠性的验证，每一帧的跟踪结果都会被用来更新模型，所以当目标进入遮挡时，跟踪器学习遮挡物的信息，造成模型被“污染”，最终导致跟踪失效。

这就要求重检测时使用的检测器必须满足：

(1) 重检测所使用的算法必须检测精度比原来的相关滤波器要高。否则如果重检测输出的结果相比相关滤波跟踪器要差，那么重检测也就失去了它的价值。

(2) 检测器不能使用帧间信息，需要在独立的每一帧内检测出目标。由于目标可能会消失一段时间后重新出现，那么对不可能找到目标的图像帧做过多的运算显然是效费比较低的，所以每隔一定帧数进行检测可以大幅提高整体的运算速度而不至于影响检测精度。

(3) 检测器使用的算法不能太过于复杂。由于目标跟踪有一定的实时性要求，如果重检测使用的算法破坏了算法整体的实时性，造成跟踪器时效不稳定，则限制了整个跟踪算法的应用范围。

(4) 检测器必须可以验证结果的可靠性。当检测的结果可靠性低于阈值时，判断为未检测到目标。由于检测成功后会重新启动跟踪器一直跟踪下去，如果在错误的目标上一直跟踪，会影响跟踪结果的正确性。

被跟踪目标出现了短时间的完全遮挡，那么想要对于其进行跟踪信息就会很

少，跟踪的过程也就非常不准确。为了应对这种情况，如果目标产生了被遮挡的问题，我们将采取重检测的方式来进行应对，其中可以利用的算法有很多，经过多年的发展，现在可以利用机器学习类方法来进行重检测模块的设计。最近较为流行的两类算法，一种是像 SVM、朴素贝叶斯等这样的经典学习算法，另一种方法是深度学习类检测方法如 YOLO<sup>[62]</sup>、SSD<sup>[63]</sup>等。这两种算法各有优缺点，机器学习类<sup>[64]</sup>方法不需要依赖图形处理单元(Graphics Processing Unit, GPU)，所需样本数量少，也无需对被检测物体有太多的先验知识。而深度学习类检测方法检测正确率很高，不过运算速度较慢，并且需要在 GPU 条件下运行。也有利用局部对比度，在图像局部的显著性区域中进行对比度计算，而非在整幅图像中检测目标<sup>[65]</sup>，从而加速运算的方法。

#### 4.2.1 级联分类器检测模块

本文采用的检测器是基于机器学习类的多个简单的分类器级联得到的检测器。相较于 SVM 分类器，这种级联分类器由于由多个简单的“弱分类器”级联而成，由于绝大多数错误的样本被前级分类器抛弃了，从而减少了很多计算量，提高了算法的实时性。如果采用 SVM 分类器这种“强分类器”的话，则面对海量的样本，计算速度无法达到实时。

在检测开始时，首先通过一个扫描窗口对图像逐一扫描，得到大量图像块，这些图像块都是有可能的目标。然后对每个图像块进行二分类，将其分成目标类和背景类，被分类为目标的样本被送到下一级分类器中，最后被级联分类器一致认为是目标的样板就是最后检测器的输出<sup>[66]</sup>。

扫描窗口使用了 21 个尺度，在原来选定的目标框大小的基础上，向上放大 10 个尺度，向下缩小 10 个尺度，加上原来的目标框，一共是 21 个尺度大小。每个梯度下都对图像进行扫描，每隔 10% 的窗口大小创造一个矩形框。按照这种设定进行扫描可以得到整幅图像的所有可能的目标候选框。

如下图所示，本文所使用的级联分类器检测模块可以由以下分类器级联而成：

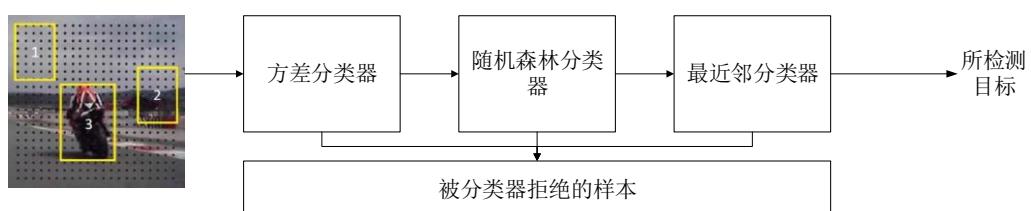


图 4-7 级联分类器检测模块示意图

首先介绍方差分类器，它作为检测器的最前端，要首先筛选出不可能作为目标的样本作为背景信息而不是有可能的目标信息传递给下一层分类器。它取灰度值来计算样本图像块和之前训练得到的模型之间的方差，如果方差小于阈值 50%，则认为该图像块与目标相差太大，被分类为背景，无法进入下一级分类器。

对于图像块  $p$ ，方差计算的公式为：

$$\text{var} = E(p^2) - E^2(p) \quad (3-24)$$

其中， $E(p)$  为图像块基于灰度值的期望。

这样一来，通过方差这样计算简单而又较为有效的分类方法，就可以将超过一半的不太可能是目标的样本，例如大量的天空、地面、水面等灰度变化平缓的背景区域就被抛弃了，有效减轻了后级分类器的工作负担。

而随机蕨分类器(Random Ferns Classifier)类似于随机森林(Random Forest)。但是它的每层“蕨”中只有一种判断准则，而随机森林则是每层都不一样。它是一个线性的结构，根据样本的特征值进行分类。在随机蕨中，从图像中随机取两点像素  $u_1$  和  $u_2$ ，并对这两个像素点进行二进制比较就可以得到特征  $f$ ，如果样本的点的亮度值高于样本的，则记为 1。如果不是则记为 0。

$$f(x; u_1, u_2) = I(x(u_1) > x(u_2)) \quad (3-25)$$

其中， $I(a)$  是个二值函数， $a$  为真时， $I(a)$  为 1。 $x(u_1)$ ， $x(u_2)$  分别是两点各自的像素值。

当选取了  $s$  个像素后，将会得到  $s$  对二进制的编码，将其转化到十进制，这就是这个图像块的特征，且特征的取值范围在  $[0, 2^s - 1]$  之间，如图 4-8 所示。

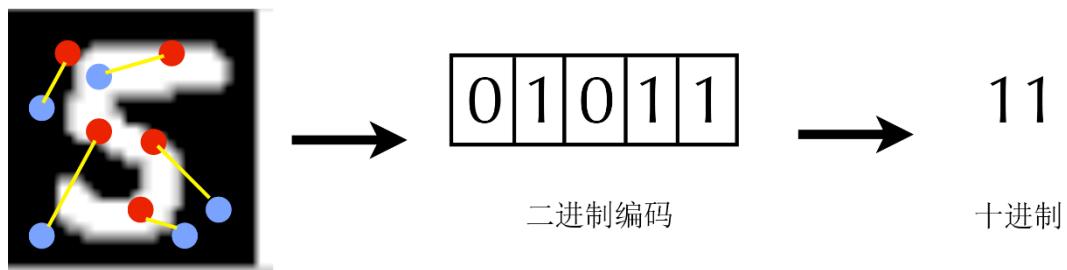


图 4-8 随机蕨特征点编码

正负样本经过随机蕨分类器后，可以得到正负样本各自的概率分布，当一个新的样本来的时候，只要这个样本经过这层蕨后在正负样本中寻找后验概率最大的一个，就是新样本的类别。经过多层蕨后，就可以提高随机蕨分类器的分类准确度。

设  $f_j (j=1,2,\dots,N)$  是待分类的图像块的特征的集合。 $c_i (i=0,1)$  为类别的集合  $C$ ，则有：

$$P = P(c_i | f_1, f_2 \dots f_N) \quad (3-26)$$

其中， $P$  表示该图像块为目标或者背景的概率。

根据贝叶斯定律：

$$P = \frac{P(f_1, f_2 \dots f_N | c_i) P(c_i)}{P(f_1, f_2 \dots f_N)} \quad (3-27)$$

其中， $P(c_i)$  符合均匀分布， $P(f_1, f_2 \dots f_N)$  是和分类无关的概率。

于是：

$$P = P(f_1, f_2 \dots f_N | c_i) \quad (3-28)$$

所以在某个类别下，每个特征的概率可以由下得到：

$$P = \prod_{1 \leq n \leq N} P(f_n | c_i) \quad (3-29)$$

其中， $\prod$  表示连乘算符。

由于  $f_j$  的形式十分简单，所以需要较多的特征才能正确分类，为了保证  $f_j$  之间有较大的相关性，假设如果将类别分组，他们之间的二值特征将会是相互独立的，将  $N_f$  个类划分为一个小组，可以得到。

$$P(C_k | f_1, f_2, \dots, f_{N_f}) = \prod_{m=1}^M P(F_m | C = C_i) \quad (3-30)$$

其中， $F_m = \{f_{\sigma(m,1)}, f_{\sigma(m,2)}, \dots, f_{\sigma(m,S)}\}$  是第  $m$  个蕨的特征的集合。 $\{\sigma(m,j) | m=1,2,\dots,M, j=1,2,\dots,S\}$  表示取值为 1 到  $N$  的随机函数。

所以该图块的类别可以由(3-31)得到：

$$k = \prod_{m=1}^M P(F_m | C = C_i) \quad (3-31)$$

其中， $k$  是该图像块的类别。

将分类为目标的图像块送入下一级分类器中，而被分类为背景的图像块则在用来更新簇模型后丢弃。

最后，在简单的分类以后用最近邻分类器对于那些没有被分类的目标进行区分，主要的原理就是寻找其中和目标比较相似度最高的目标块。如果其中的目标块计算得到的相关相似度值为  $S_z^{nn}$  ( $nn$  为阈值)，那么我们就将这种情况判断为目标

检测成功，反之就是失败，具体的定义如下所示：

$$S_z^+ = \max_{Z_i^+ \in N} S_{z,+} \quad (3-32)$$

$$S_z^- = \max_{Z_i^- \in N} S_{z,-} \quad (3-33)$$

$$S_{z,-} = S(p, p_i^-) \quad (3-34)$$

$$S_{z,+} = S(p, p_i^+) \quad (3-35)$$

$$S_z = \frac{S_z^+}{S_z^+ + S_z^-} \quad (3-36)$$

其中， $S_z^+$ 和 $S_z^-$ 分别表示正样本和负样本的最近邻相似度， $p_i^+$  $p_i^-$ 分别为正负样本目标模型， $p$ 是待分类的样本。

两个图像块之间的相似度 $s$ ，是由归一化互相关匹配算法计算得到的：

$$S(p, p_k) = 0.5(NCC(p, p_k) + 1) \quad (3-37)$$

其中 $p_k$ 是第 $k$ 个图像块样本。

分类器利用(3-36)计算相关相似度，相关相似度越大，则样本越有可能与模型相似，越有可能是我们要寻找的目标，寻找一个阈值 $S_z^{mn}$ ，当 $S_z > S_z^{mn}$ 时，我们认为图像块包含目标，分类器输出结果，认为找到了目标。数值越小表示越不可能是真正的目标，则将其丢弃。

### 4.3 实验结果

为了验证本文算法在各种复杂场景下的有效性，我们需要一种标准的测试基准(benchmark)来对比不同的跟踪算法在同一个测试序列中的结果，然后与人工标注的真值(ground-truth)进行比较，以此来评估跟踪器的性能。

目前在跟踪领域，一个名为 Object Tracking Benchmark(OTB)的，是一个来自韩国的团队的工作，它是一个包含了许多跟踪挑战的公开测试集，所有测试集对外公开。由于这个数据集不仅覆盖了常见的跟踪场景，还囊括了跟踪中存在困难的许多复杂场景，具备很强的挑战性。本文采用的是 OTB2013<sup>[67]</sup>目标跟踪数据集，它包含十一种跟踪的难点（光照变化、目标遮挡、尺度变化以及快速运动等）总长度超过五万帧，目前最新的主流跟踪算法几乎都给出了在这个数据集上的跟踪结测

试结果，从而方便研究人员将自己的算法与同领域其他算法进行对比，具有一定的权威性和可靠性。本文算法也将会在这个数据集中进行评估和分析。

OTB 基准测试集是由许多的视频序列以及每个序列对应的人工标注组成的。其中包括很多的种类视频序列，比如其中含有电影动画，同时还包括一些自然场景、基于人脸的场景、基于人体的场景、基于车辆的场景等。测试集的序列也长短不一，最短的只有几十帧，最长的超过 3000 帧。其中全部的视频序列都已经被标注号了目标的位置以及其目标框大小。其中主要有两种类型的视频，一种是 RGB 三通道视频，还有灰度视频，对于 RGB 三通道视频来说，其中一共含有五十一个，对于灰度视频来说一种含有一百个。其中视频挑战的程度非常的丰富，一共含有十一种类型的挑战，在这十一种不一样的挑战中可以对于算法在多种场景下的具体情况进行全面的分析，最终评估算法在各种场景下的效果。此外，还能对于算法的鲁棒性进行分析。这个数据集还包括了一个一致的代码接口，通过这个接口可以实现输入和输出的标准化，通过这个接口可以很方便地添加其他算法并对各种算法进行横向比较。

如下图是 Object Tracking Benchmark(OTB)数据集，每个序列都标明了这个序列所包含的不同的挑战任务。

OTB 数据集中对于算法会利用两种标准来进行评价，一个是算法的准确性，另一个是算法成功率。对于算法准确性的判断主要依据算法对于目标位置判断的准确性，这方面的考察主要判断的是这个算法跟踪目标过程中，目标和中心位置的差异大小。之后又主要通过准确率曲线来对结果进行判定，通过对这个曲线的分析可以了解到其跟踪结果和真值框中心点的距离比标准的阈值小的帧数的百分比情况。在这个数据集的测试中，使用的阈值标准为 20 个像素点。另外一个判断的标准是成功率，成功率是一个非常重要的判断指标，通过这个指标可以了解通过算法进行的跟踪得到的矩形框和真实的目标框之间的重叠情况。一般会利用  $R_T$  来表示算法得到的目标矩形框，会利用  $R_A$  来表示人工标注出来的准确的矩形框，最终计算重叠率的公式就是  $S = \frac{|R_T \cap R_A|}{|R_T \cup R_A|}$ ，对于成功率的判断除了需要对于其中目标中心位置进行考察以外，还需要对于其中的目标尺度信息进行判断。最后的成功率曲线就是利用各个计算得到的重叠率阈值画出来的，其中这个算法的成功率就是曲线以下的面积部分。

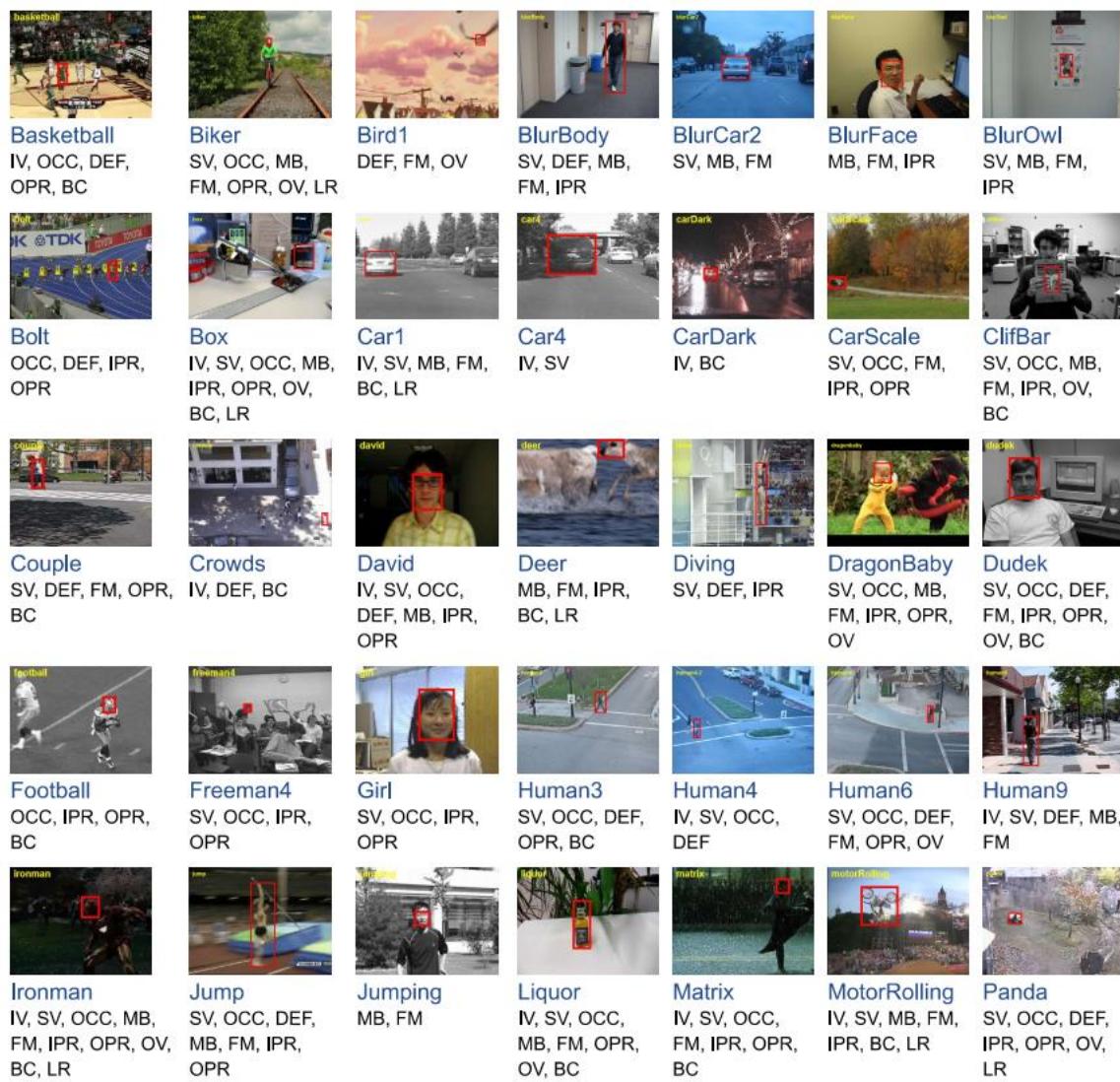


图 4-9 OTB 数据集

### 4.3.1 实验运行环境

本文的复杂场景下目标跟踪算法的测试实验平台是在 Matlab R2018b 64 位软件, Intel(R) Core(TM) i5-6500 @3.4Ghz 的 PC 机上进行测试的, 电脑运行内存 8G, 操作系统为 64 位的 Windows10 系统。

### 4.3.2 实验评价指标

在 OTB 数据集中, 有“成功率”、“距离精度”两种量化评价指标。

#### (1) 成功率 Success Rate

成功率的计算需要两个区域信息: 一个是人工标注的区域, 第二是跟踪器每帧

输出的目标框范围。根据每帧预测区域( $ROI_t$ )和人工标注区域( $ROI_{gt}$ )的交集与并集的比值如图 4- 10 所示。重叠率(Pascal VOC Overlap Ratio, VOR)<sup>[68]</sup>的计算公式如下：

$$VOR = \frac{area(ROI_t \cap ROI_{gt})}{area(ROI_t \cup ROI_{gt})} \quad (3-38)$$

其中， $ROI_t$  和  $ROI_{gt}$  分别是跟踪器输出的预测区域和人工标注的真值区域。

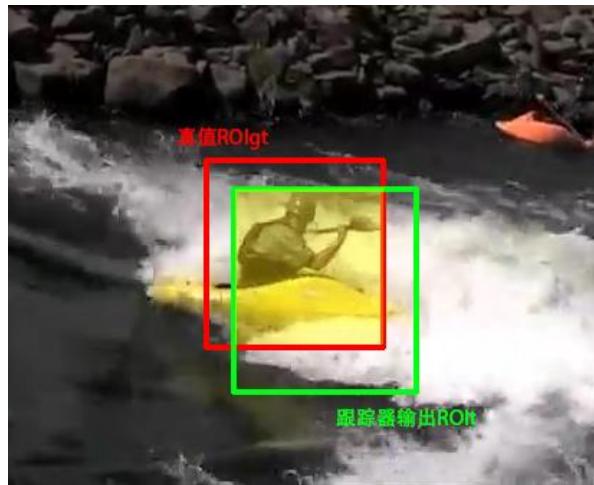


图 4- 10 重叠区域示意图

在成功率图中，每个跟踪器的曲线下面积(Area Under Curve, AUC)越大，说明跟踪器总体性能越好。

## (2) 距离精度 Precision

距离精度指标 Precision 使用中心定位误差(Center Location Error, CLE)来定义人工标注的目标框的中心位置和跟踪器输出的目标框的中心位置的欧氏距离。然后统计不同距离精度下的跟踪帧占测试序列中所有的图像数量的比值可以计算画出一条曲线，即精度曲线图(Precision plot)。来评价跟踪器在该序列上的总体性能。中心定位误差的计算可以由人工标注的目标框中心位置  $P_{gt}$  和跟踪器的输出目标框中心位置  $P_t$  求差再求绝对值得到，如公式(3-39)：

$$CLE = |P_{gt} - P_t| \quad (3-39)$$

其中， $P_{gt}$  和  $P_t$  分别是人工标注的目标框的真值的中心位置和踪器的输出目标框中心位置

和在成功率图中的结论类似，在距离精度图中，AUC 也代表了跟踪器性能的好坏。

### 4.3.3 定性分析

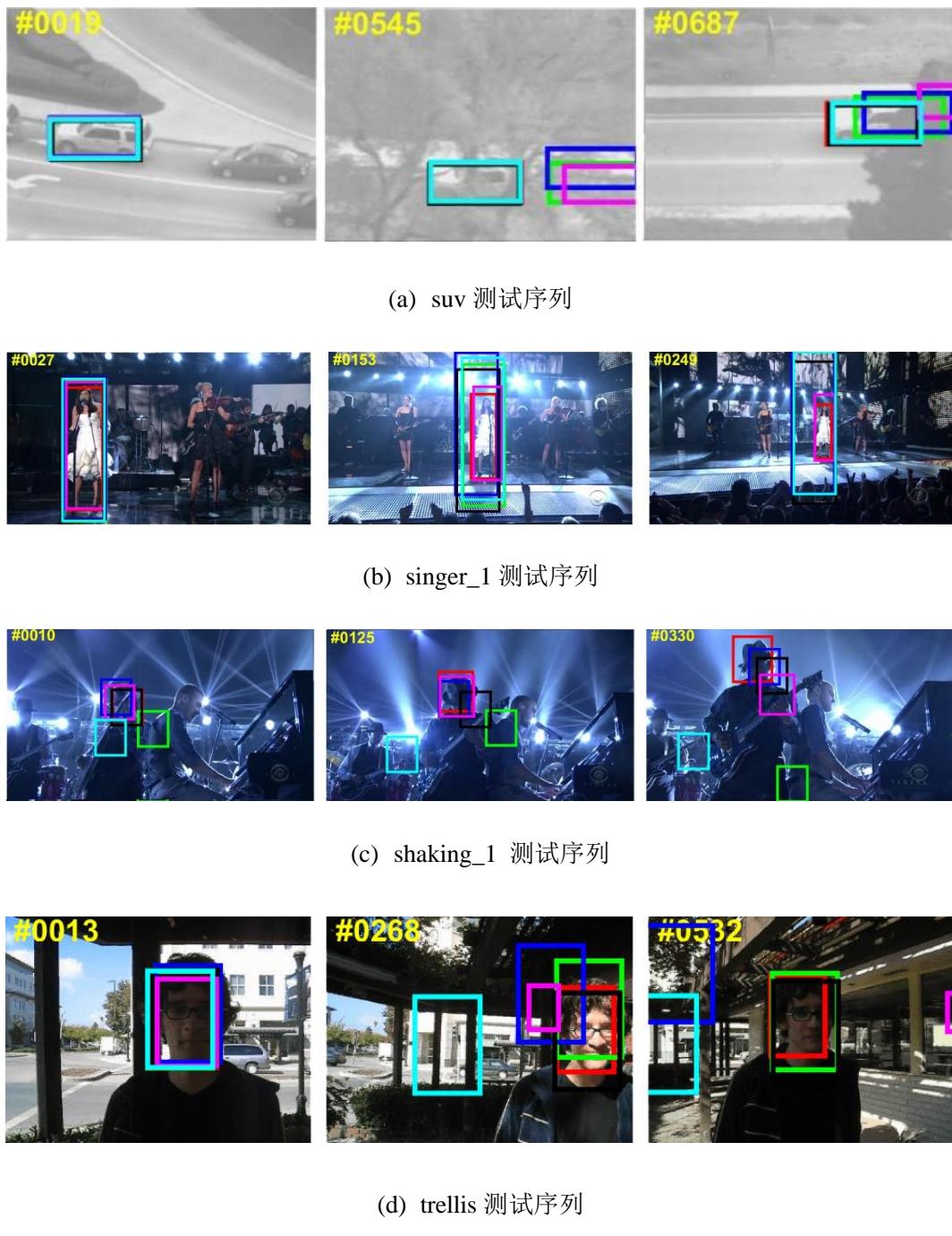


图 4-11 几种典型测试序列的跟踪结果比较

如图 4-11，截取了六个跟踪器在几种典型复杂场景下的跟踪效果比较结果。每个图中的框表示各个跟踪器输出的跟踪框，跟踪框的颜色代表了它所属的跟踪

器，跟踪器的颜色和名称的对应在图的最下方可以找到。每张图的左上角代表了当前帧数，在每一组测试序列中，分别截取了跟踪刚开始时，跟踪进行到一半，跟踪接近尾声的图片，来尽量表示出跟踪器在整个测试序列中的表现。下面对每种场景做具体分析：

### (1) 遮挡及复杂场景下的跟踪结果分析

Suv 测试序列是 OTB 目标跟踪数据集中挑战较大的一个测试序列，因为其包含了部分遮挡、完全遮挡、其他相似物体干扰的场景。而 shaking 数据集则是其中目标自身形变较大，背景复杂，光照条件多变的复杂场景。(a)是 suv 数据集中车辆先后行驶经过道路标示线、道路指示牌、树林等环境事物，造成被跟踪目标被部分、完全遮挡的场景。可以看出跟踪场景是十分复杂的，车辆行驶过程中除了目标车辆还有别的车辆在一同行进。目标行驶到树林背后时，目标被完全遮挡。此时 KCF 跟踪器由于使用的是固定学习率的更新策略，无法判断遮挡情况，模型被污染，导致跟踪到了树和目标之间的位置，符合模型既包含车辆目标，又包含新学习的树木目标的理论分析。目标在被部分遮挡后，TLD 跟踪器由于其使用了模板匹配，会将跟踪框转移到非目标的另一辆车上，其移出视野后原目标已经脱离遮挡，所以 TLD 跟踪器又回到了原来的目标上。只有本文算法和 MIL 算法准确跟踪到了目标。而(c)shaking 数据集目标在剧烈抖动，属于快速运动情况，且伴随着人脸的快速变形，其帽子由于角度变化，视觉效果也在快速变化。可以看到在 120 帧时，目标头部朝下，相比 10 帧时面部正面对准屏幕的样子有了很大改变，此时 KCF、Struck、MIL 都已经跟踪失败。330 帧时，由于目标变化太大，TLD 跟踪器也发生了漂移，输出框维持在目标右下部也就是肩膀的位置持续跟踪。

### (2) 光照条件改变情况下的跟踪结果分析

由于在摄像机的原理其实是接收的是外界照明光入射到物体表面反射进入到 CCD 中得到的光信息，所以当外界照明条件变化时，物体的颜色、纹理都将发生改变。特别是如(d)trellis 数据集这样，目标在树荫底下前进，外界照明光不均匀，或者是走到暗处，目标被照射不充足时，颜色、纹理信息大量丢失。到光照强度较大处时又突然增加了很多边缘和纹理，和刚开始在较暗处时给定的目标发生了很大的变化。从 10 帧时，目标从较暗处开始出发，此时目标的五官较不清晰。到 268 帧时，目标走到了光照充足处，五官变得明显，和一开始给定的目标有了很大变化，所以 MIL、TLD、CSK 跟踪器都选择了跟踪背景较暗处而不是较亮的目标本身。在(b)singer1 数据集中，我们可以清楚地看到 27 帧和 153 帧时舞台灯光发生了很大的变化，同时目标的尺度也发生了缩小。此时相机焦距，目标形变也对跟踪造成了不小的挑战，此时只有本文算法和 TLD 成功跟踪到了目标。

综上，我们发现相对于其他几种算法，本文算法在复杂背景及光照变化下时能够实现较为准确的跟踪。

#### 4.3.4 定量分析

如图 4-12，在 OTB 目标跟踪数据集上测试了 50 个视频序列，其中包含了有照明变化、比例变化、遮挡、变形、运动模糊、快速运动、平面内旋转、超出视野、背景杂波、低分辨率等多种挑战<sup>[69]</sup>。本文的测试将采用一次通过评估(OPE, One Pass Evaluation)模式，这种评估模式在众多数据集综合测试之后，得到了这几个跟踪器的精度曲线和成功率曲线，在相同条件下，跟踪器的曲线面积越大，说明跟踪器效果越好。

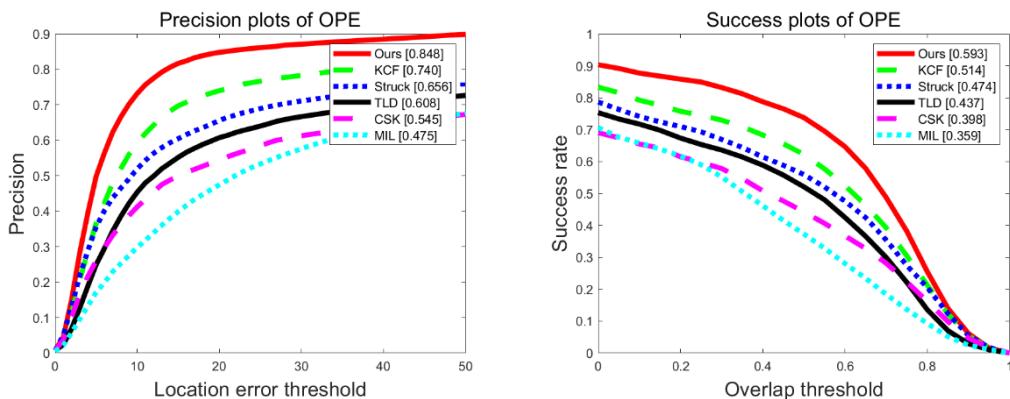
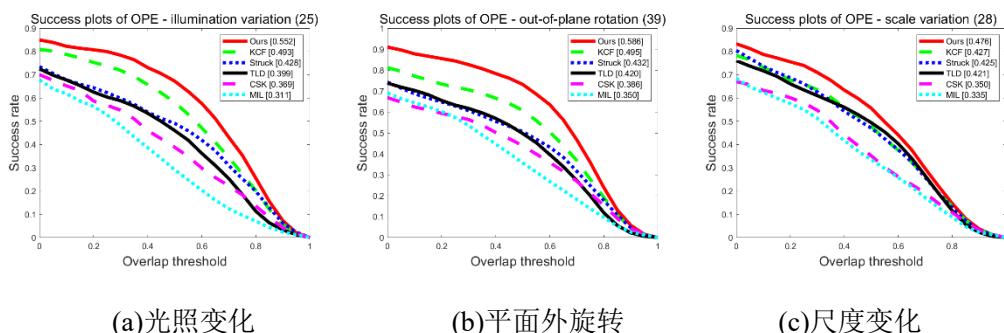


图 4-12 几种不同跟踪算法的精度和成功率对比曲线图

其中 Ours 为本文所提出算法，在使用了多特征融合以及目标重检测模块之后，从上图实验结果看出，本章提出的算法无论精度还是成功率都比其他算法有着更好的表现。具体而言，跟踪精度相对 KCF 跟踪算法提升了 0.108，跟踪成功率则提升了 0.099，整体而言有很大的提升。为了更好地观察算法在具体的跟踪中所要面临的各种挑战下跟踪器的表现，下面给出针对 OTB 数据集的 11 中挑战项目测试得到的精度和成功率如图 4-13。



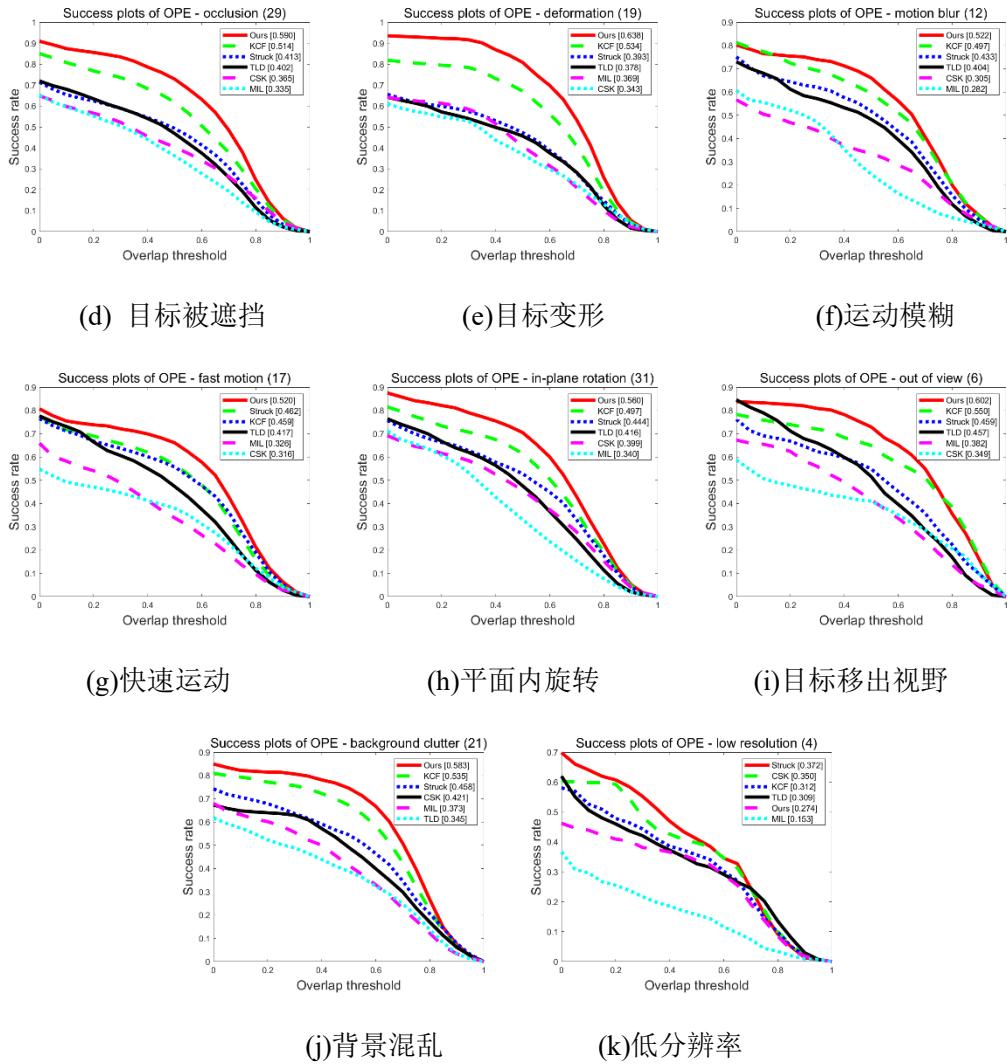
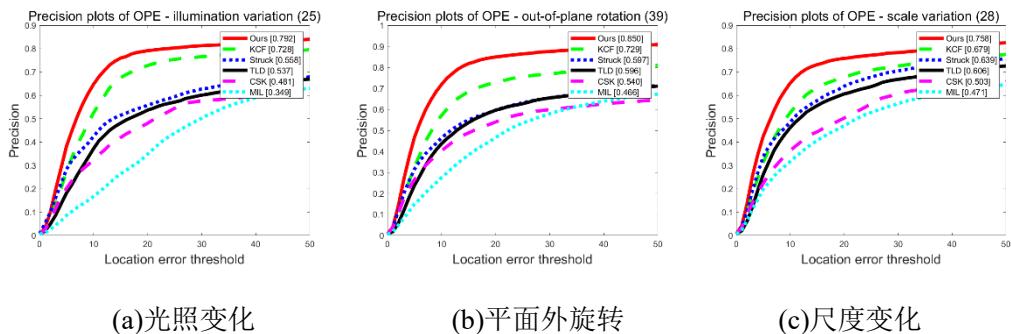


图 4-14 11 种跟踪挑战下的成功率图



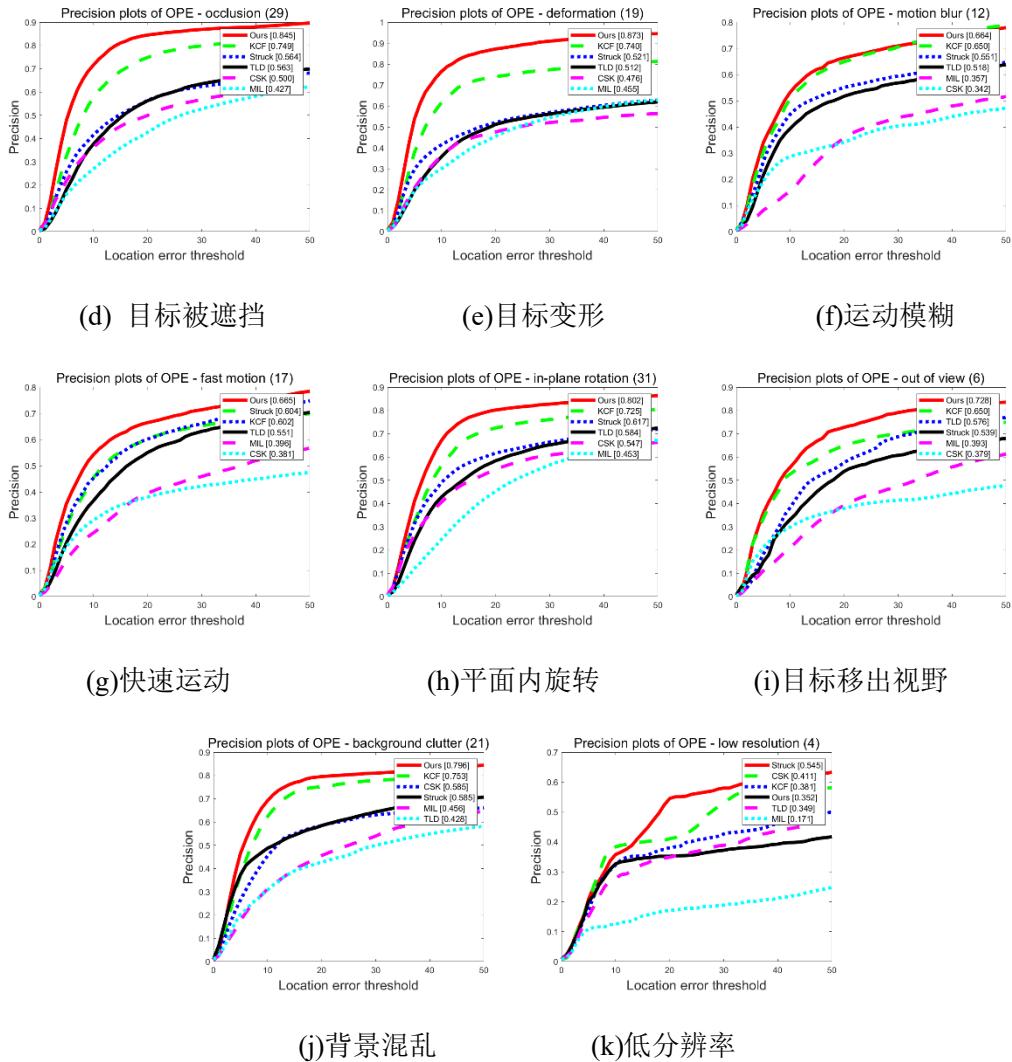


图 4-15 11 种跟踪挑战下的精度图

由于采用了多种特征融合的方法，在各种复杂的跟踪场景下，本文算法都有着较为稳定的表现。例如在(h)平面内旋转和 (b)平面外旋转 两种跟踪挑战下，由于采用了 LBP 特征，在目标旋转时，利用了 LBP 特征的旋转不变这种优秀的特质，使得跟踪得以更加稳定。再如(e) 目标变形时，由于目标虽然变形，其纹理和形状大幅变化，而它们的颜色直方图则是不变的，由于采用了 CN 特征，使得在这种场景下的跟踪更为稳定。由于使用了重检测的机制，所以在目标被遮挡如(d)目标被遮挡，甚至移除出视野的时候如(i)目标移出视野的时候，相对其他跟踪器，本文算法仍可以重新检测目标，然后重启跟踪器，得到较为稳定地跟踪效果。但是从(k)低分辨率 的跟踪挑战中也可以看出来，由于在低分辨率下的纹理形状较为不清晰，HOG 特征和 LBP 特征的表达能力减弱，重检测模块的检测能力也降低了。值得注意的是，与本文算法一样使用 HOG 特征的 KCF 跟踪算法的排名也有了一定下降。

从 11 种不同的跟踪挑战下的成功率图和精度图中我们可以看到在绝大多数的情况下，本文提出的算法的跟踪效果要优于其他几种算法。

下表给出几个测试序列的具体跟踪成功率和跟踪精度：

表 4-1 跟踪结果成功率比较

不同算法的成功率(%)						
视频序列 \ 算法	Ours	Struck	MIL	KCF	TLD	CSK
david	0.727	0.538	0.402	0.239	0.718	0.428
deer	0.786	0.624	0.746	0.740	0.602	0.124
singer1	0.835	0.355	0.359	0.358	0.725	0.356
shaking	0.765	0.040	0.575	0.349	0.390	0.427
soccer	0.384	0.422	0.145	0.185	0.127	0.171
car4	0.863	0.484	0.466	0.489	0.633	0.258
matrix	0.457	0.119	0.031	0.098	0.156	0.122
Ave	0.688	0.369	0.389	0.351	0.479	0.269

表 4-2 跟踪结果距离精度比较

不同算法的距离精度(px)						
视频序列 \ 算法	Ours	Struck	MIL	KCF	TLD	CSK
david	4.537	8.062	17.690	42.801	5.122	16.860
deer	4.472	21.155	4.963	5.265	30.931	100.729
singer1	3.564	12.846	14.012	14.510	7.991	16.365
shaking	7.118	112.500	17.163	30.695	37.106	24.027
soccer	19.434	15.373	70.509	71.358	77.101	77.853
car4	2.282	9.876	19.133	8.688	12.837	50.777
matrix	26.753	76.423	113.692	194.528	57.229	55.029
Ave	9.737	36.605	36.737	52.549	32.617	48.806

对于各个算法，为了验证他们各自算法的复杂度，测试其实时性，探究他们在不同应用场景中的使用条件。本文对上述的 6 种经典算法测试的时候还统计了他们的计算速度。本文选用跟踪帧率的平均值作为衡量实时性的标准，即用跟踪了的总的帧数除以总时间得到平均帧率(FPS, Frames Per Second)。

平均帧率计算公式如(3-40):

$$FPS = \frac{N_{frames}}{T_{time}} \quad (3-40)$$

其中， $N_{frames}$  代表测试了的帧数的总和，在这里是所有 50 个测试序列的总帧数。 $T_{time}$  是每个序列中算法的运行时间的总和。

表 4-3 跟踪算法运算平均帧率

算法名称	Ours	Struck	MIL	KCF	TLD	CSK
平均帧率 (FPS)	22.36	19.51	31.46	248.99	19.33	198.21

上表 4-3, 表 4-4 给出了不同测试序列下跟踪结果的跟踪成功率和跟踪结果的距离精度。其中, 当算法输出的目标框与人工标注的目标的“真实”位置中心距离相差小于 20 像素的时候就认为跟踪成功了。之前分析时提到本文算法对目标被遮挡, 光照条件变化, 目标发生形变时都能较好地跟踪目标, 有着较高的跟踪成功率, 故才会有较高的跟踪误差距离精度。但值得注意的是, 由于本文使用了 HOG 特征、LBP 特征和 CN 特征, 对 CPU 的要求比较高, 所以运算速度相比其他的算法并不是很高, 接近了每秒 25 帧的实时性要求, 但在实时性要求较高的应用上仍有一定限制, 如表 4-3 所示。

综上所述, 从成功率、精度两个评价指标上可以发现, 本章算法的整体性能最好。从平均中心误差和平均帧率评价指标上可以看出来, 本章算法的平均中心误差最小, 即跟踪精度最高; 平均帧率排在算法 KCF 和算法 CSK 两种相关滤波类算法和 MIL 之后, 但同样可以满足目标跟踪算法的实时性。综合所有评价指标可以得出本章算法的效果最好。

#### 4.4 算法的工程实现

目前大多数的目标跟踪算法都是在 Matlab 平台上开发实现的。Matlab 平台作为一个算法仿真、验证的平台是十分优秀的。但是在工程应用上, 要缅甸对实际应用场景下对跟踪算法部署的轻量化、可裁剪的需求, 对算法的优化效率的需求, 代码可维护性的需求, 代码可移植性的需求就远远不能满足了。

为了将算法应用落地, 在 Matlab 平台下对本文算法仿真验证, 借鉴了团队之前的跟踪方法移植方案<sup>[29, 70-72]</sup>之后, 我们开发了一套基于 C++程序开发语言实现的本文描述的目标跟踪算法。移植到 C++平台下的算法在算法的原理和正确性上与本文算法是一样的。但是由于 C++优秀的编译加速性能, 算法的实际运行效率得到了提高。程序的各种接口也基本实现了接口化, 代码解耦后程序的可维护性大大提高了。在此之上, 为了方便代码在各个平台上的测试、使用。还开发了一套易于使用、美观的基于 Qt-GUI 的用户界面程序。

软件的主体界面如图 4-16 所示。这套复杂场景下的目标跟踪算法仿真平台软件的界面是按照各个模块的功能以及需要的参数分成各个页面来实现的。界面上包括目标跟踪监视窗口模块，跟踪状态显示模块，跟踪参数设置模块，测试数据流控制模块和一些按钮。在软件主体界面中，可以看到界面中间是跟踪视频的图片显示，在图像显示层中用黄色的目标框表示跟踪器输出的跟踪结果。图像下方的状态信息栏中显示的是当前跟踪器的相关信息输出，从左到右依次是：当前跟踪器输出的目标中心坐标；当前帧序号与整段视频的长度；当前跟踪器的跟踪状态，跟踪状态分为跟踪正常，目标丢失两种。当目标发生丢失后，将会发出告警，提示用户当前目标已经丢失，并启用重检测模块。

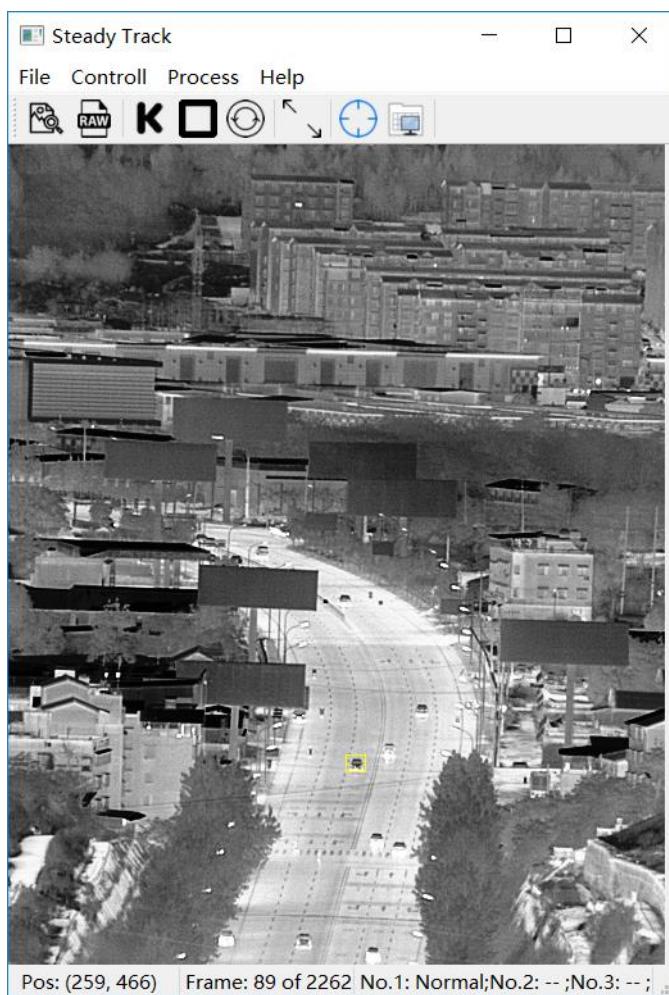


图 4-16 软件主体界面

待重新找到目标重启跟踪器后，则会重新返回跟踪正常状态。在上方的快捷按钮则依次是：“打开图像”按钮、“读取 raw 格式文件”按钮、“继续”按钮、“暂停”按钮、“重启数据流按钮”、“适配窗体”按钮、“跟踪参数选择”按钮，“增强

“算法选项”按钮。除了按钮，对应的功能还可以在上方的菜单栏中选择。界面设计风格简洁，参数清晰，易于修改，易于操作。各个模块按照其功能分布合理。用户可以很容易获得当前的图像信息和跟踪信息，十分方便。当前软件的功能满足了项目的需求。

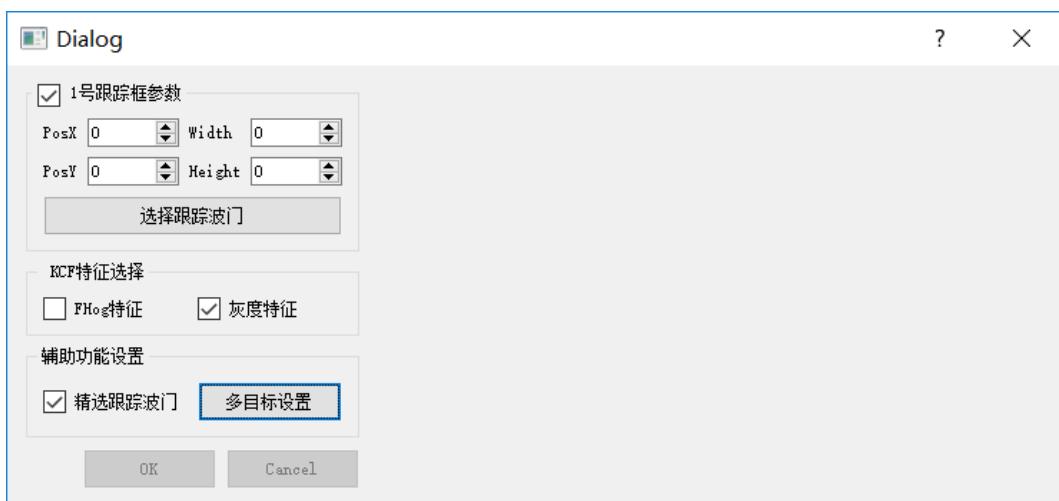


图 4-17 跟踪参数设置界面

为了方便设置跟踪算法中的各项参数，还设置了跟踪参数设置对话框。对话框分成三个功能模块，分别是跟踪框参数设置模块、特征选择模块、辅助功能设置模块。在跟踪框参数设置模块中可以使用“选择跟踪波门”功能在图像中框选待跟踪的初始目标框，也可以手动填写目标在图像中的精确坐标，方便重复试验；在特征选择功能模块中可以选择用于跟踪的特征、在辅助功能模块中可以选择是否使用精选跟踪波门功能。

经过不断地修改和改进，这套跟踪方案已经成为了功能完善，性能优秀，功能齐全，使用方便的一套跟踪算法系统方案。算法代码和界面代码由于采用 QT-GUI 和 C++ 的方案，可移植性强，可以移植到 Windows 操作系统、基于 Linux 内核的 ARM 开发板和 FPGA、DSP 开发板上，在各个平台上部署算法十分方便。算法内部的代码架构设计合理，各部分代码之间解耦性强，接口齐备，文档齐全。很好地满足了工程应用方的需求。使得我们对跟踪算法在实际落地应用的过程加深了体会，也锻炼了思维和代码能力。

本文提出的算法在移植后，很好地完成了各项设计指标，在项目需求方给出的各个数据集上大多表现良好，圆满完成了工作。

## 4.5 本章小结

本章中引入了基于相关滤波跟踪输出的响应图的峰值旁瓣比来评价跟踪质量，根据跟踪质量来选择是否更新模板，防止模板被污染，是本文提出的跟踪器中的重要模块。接着介绍了本文针对目标被遮挡这种跟踪中常见的复杂场景引入了目标重检测模块。最后介绍了一种大规模公开跟踪数据集，对本文提出的算法进行了充分的实验。实验表明，这种算法在复杂场景下具有良好的跟踪效果。在验证了本文算法的有效性之后，将本文算法在工程项目中的实现进行了介绍，展示了算法的工程价值。

## 第五章 总结与展望

### 5.1 工作总结

随着现代计算机算力的发展和人们发现的越来越多，越来越巧妙的算法，我们已经俨然进入了人工智能时代，越来越多的本来需要人们亲自动手去做的事情现在已经可以逐渐由机器，由电脑替代了。特别是工业生产自动化，无人自动驾驶领域的发展更是迅速。目标跟踪能让计算机更好地学会像人类一样去看的本领，甚至可以作为计算机去理解我们人类赖以生存的物质世界的一种重要途径，总之，对我们生活各方面都具有重要的意义。虽然有许多优秀的目标跟踪算法已经被提出来，但要让算法适应实际生产生活复杂条件，能真正提高人们劳动生产率，还存在不少问题没有解决。

#### 5.1.1 开展工作

本文针对目标跟踪中存在的一些问题，阅读了大量国内外关于目标跟踪的理论方法，学习并理解了数字图像处理中的很多基础知识和理论。研究并改进了原有的跟踪方法，并通过实验验证了本文提出的方法的有效性。具体而言，本文的贡献和工作如下：

(1)研究了目标跟踪相关理论，对跟踪中常用的几种特征进行了研究，并对它们的性能和特点进行了分析。提出了多特征联合使用来增强跟踪场景适应性的特征的使用策略，使得跟踪算法更有利于在光照条件明显改变、目标外观快速变化、背景复杂多变等多种复杂的场景下进行跟踪。

(2)研究了相关滤波方法的相关理论，介绍了算法的基本理论和基本步骤。针对相关滤波跟踪器无法评价当前跟踪效果的不足，增加了跟踪效果评价模块，使得跟踪器可以判断目标是否已经丢失，弥补了现有滤波器无法判断跟踪状态导致跟踪失败的缺点。

(3)针对跟踪中目标发生遮挡后无法继续跟踪的问题，引入了重检测模块，根据跟踪效果评价模块的输出，当判断目标丢失时，在图像中重新检测目标，待找到目标后重新启动跟踪，使得跟踪算法在应对部分或者全部遮挡后，可以有效地重新找回目标，来达到长时间跟踪的效果。

(4)本文在大规模公开数据集对目前主流的几种跟踪器和本文提出的跟踪器进行评测和比较。实验表明，相较于传统算法，本文算法在复杂场景下的跟踪效果取得了满意的效果。最后，在验证了本文算法的有效性之后，还将本文算法移植到了

适合生产中使用的易于移植落地的 C++ 程序语言平台并设计了功能完善的界面方便使用。

### 5.1.2 创新点及主要贡献

本文创新点及主要贡献为：

(1) 基于相关滤波类方法的特性，取长补短，利用了其采取计算相关运算来跟踪目标的方式，不限定于某种特征的特点，分析了它在目标被遮挡后模板被污染的不足，改进了它的模板更新模式和融合了多特征来使用。增加了跟踪器在其中一种特征失效或是目标被遮挡场景时跟踪的鲁棒性。

(2) 将算法移植到了高效率、灵活、易于维护的 C++ 程序语言平台并设计了软件使用界面，使其方便工程应用。

## 5.2 后续研究展望

在复杂场景下的运动目标跟踪是当今学术界和工业界都十分热门的话题，在无人驾驶，智能工厂中也有较大的应用前景。由于时间有限等原因的限制，本文所提出方法仍有一下不足待改进的地方：

(1) 目前多特征融合的方式需要多次计算相关来获取动态的特征权值，尝试改进计算权值的方法，使得权值可以不必做完整的相关计算流程后再计算。

(2) 目前使用的跟踪置信度判据仍依赖跟踪输出的结果，寻找一种更为独立、更为鲁棒的跟踪置信度判据会有利于算法在各种条件下发挥更稳定的效果。

(3) 深度神经网络在目标检测任务中取得了十分好的成绩，尝试将网络裁剪加速至能在计算资源有限的设备上仍能达到实时的效果的话，对提高跟踪器重检测目标的效果应该会大大提高，也能提高算法的应用前景。

## 致谢

时光荏苒，不知不觉在成电已经快要度过第七个年头了。回首这三年的生活，回忆里的只有欢笑和甜蜜。感谢过去三年中陪伴着我的家人，朋友，老师，同学们，是你们用一言一行陪伴着我，带给我欢笑，送给我祝福，陪着我打闹，教授我知识，指引我方向。

首先要感谢的是我的恩师，彭真明教授。他在我的本科阶段就开始教授我们《光电图像处理》一课，这也是我学习图像处理的启蒙课程。之后有幸被彭老师面试选中，正式成为了实验室的一份子。在我整个硕士阶段，彭老师都以他渊博的学识，严谨的治学态度，崇高的敬业精神，阳光的做人理念，成为了我人生和学术上的标杆，这份信念也将成为我今后在人生大海上航行的舵和帆，带给我方向也给我动力。同时也感谢和我们最亲近的何艳敏老师，每年中秋的月饼，生日礼物和每个在实验室日日夜夜的陪伴，让我们这些在外求学子们感到了家的温暖，同时也带给我们思考和宁静。每天坚持运动的好习惯更是让我感到自愧不如。还有陶冰洁老师组织、引导的汇报，让我懂得了做汇报要从平时积累素材，也要有面对观众展示自己的自信。感谢王卓然老师在科研项目上的管理，让我们有条不紊地按时完成任务。

感谢教研室的李梦雪、王璟璟、田明师兄师姐，让我在教研室开始了第一个项目，从而快速学习，快速成长。在项目组里一起工作的曹思颖、龙鸿峰、李健、李美惠几位同学，和你们一起头脑风暴，灵感碰撞的讨论使我受益匪浅，也明白了团队协作的意义。感谢张明英、杨佳晴、饶紫鹏、汪春宇、黄景雨、吴逢信、成祎珊、夏清几位同学，让我在平时学习生活、找工作期间充满了温馨和欢乐。同窗之谊，我将终生难忘！遇到你们，真幸运。

最后要感谢我的父母，在我犹豫困惑的时候，总是可以给我一个温暖的依靠，让我知道即使外面风雨再大，我还是有一个温暖的港湾可以停靠。感谢你们多年如一日的教导和付出，让我在没有太多压力的情况下完成了学业，也让我学会了如何做事，如何做人。他们对我无私的爱是我不断前进的动力和方向。

## 参考文献

- [1] 赖建宏. 复杂环境目标跟踪点漂移校正与抑制方法研究[D]. 成都: 电子科技大学, 2013, 20-37
- [2] Y. Hua, K. Alahari, C. Schmid. Occlusion and motion reasoning for long-term tracking[C]. European Conference on Computer Vision, 2014: 172-87
- [3] H. Grabner, M. Grabner, H. Bischof. Real-time tracking via on-line boosting[C]. Bmvc, 2006: 6
- [4] Z.-C. Mu. Ear detection based on skin-color and contour information[C]. 2007 International Conference on Machine Learning and Cybernetics, 2007: 2213-17
- [5] C. Bibby, I. Reid. Robust real-time visual tracking using pixel-wise posteriors[C]. European Conference on Computer Vision, 2008: 831-44
- [6] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt. A system for video surveillance and monitoring[J]. VSAM final report, 2000: 1-68
- [7] 王洪兵. 增强现实环境下基于视觉的高精度目标跟踪技术研究[D]. 成都: 电子科技大学, 2010, 1-8
- [8] 冯谦. 基于局部特征的视频目标跟踪方法研究[D]. 成都: 电子科技大学, 2013, 11-15
- [9] 吴大. 远距离前视红外多目标跟踪方法研究[D]. 成都: 电子科技大学, 2012, 9-16
- [10] 孟晔铭. 成像探测系统中的红外弱小目标跟踪点漂移抑制方法研究[D]. 成都: 电子科技大学, 2015, 52-60
- [11] 彭真明, 陈颖频, 蒲恬, 王雨青, 何艳敏. 基于稀疏表示及正则约束的图像去噪方法综述[J]. 数据采集与处理, 2018, 33, (1): 1-11
- [12] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, M. Shah. Visual tracking: An experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36, (7): 1442-68
- [13] K. Briechle, U. D. Hanebeck. Template matching using fast normalized cross correlation[C]. Proc. SPIE, 2001: 95-102
- [14] H. T. Nguyen, A. W. Smeulders. Robust tracking using foreground-background texture discrimination[J]. International Journal of Computer Vision, 2006, 69, (3): 277-93
- [15] D. Comaniciu, V. Ramesh, P. Meer. Real-time tracking of non-rigid objects using mean shift[C]. Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, 2000: 142-49

- [16] D. A. Ross, J. Lim, R.-S. Lin, M.-H. Yang. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77, (1): 125-41
- [17] B. Babenko, M.-H. Yang, S. Belongie. Visual tracking with online multiple instance learning[C]. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009: 983-90
- [18] S. Baker, I. Matthews. Lucas-kanade 20 years on: A unifying framework[J]. International journal of computer vision, 2004, 56, (3): 221-55
- [19] 刘洁. 基于特征点的目标检测与跟踪方法研究[D]. 成都: 电子科技大学, 2008, 44-55
- [20] Z. Kalal, J. Matas, K. Mikolajczyk. Pn learning: Bootstrapping binary classifiers by structural constraints[C]. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010: 49-56
- [21] Z. Kalal, K. Mikolajczyk, J. Matas. Tracking-learning-detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34, (7): 1409-22
- [22] Z. Kalal, K. Mikolajczyk, J. Matas. Face-tld: Tracking-learning-detection applied to faces[C]. Image Processing (ICIP), 2010 17th IEEE International Conference on, 2010: 3789-92
- [23] 徐东. 基于轮廓的图像分割及目标跟踪方法研究[D]. 成都: 电子科技大学, 2012
- [24] J. F. Henriques, R. Caseiro, P. Martins, J. Batista. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37, (3): 583-96
- [25] D. S. Bolme, J. R. Beveridge, B. A. Draper, Y. M. Lui. Visual object tracking using adaptive correlation filters[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 2544-50
- [26] J. F. Henriques, R. Caseiro, P. Martins, J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels[C]. European conference on computer vision, 2012: 702-15
- [27] J. F. Henriques, J. Carreira, R. Caseiro, J. Batista. Beyond hard negative mining: Efficient detector learning via block-circulant decomposition[C]. proceedings of the IEEE International Conference on Computer Vision, 2013: 2760-67
- [28] C. Ma, X. Yang, C. Zhang, M.-H. Yang. Long-term correlation tracking[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 5388-96
- [29] 王珏. 基于 Android 平台的多目标实时跟踪技术研究[D]. 成都: 电子科技大学, 2016, 22-28

- [30] 唐龙. 基于 DSP 的视频运动目标实时跟踪系统研究与实现[D]. 成都: 电子科技大学, 2014, 50-58
- [31] 黄斌. 基于粒子滤波的视频目标跟踪方法研究[D]. 成都: 电子科技大学, 2008, 74-94
- [32] B. Li, W. Xiong, W. Hu, B. Funt, J. Xing. Multi-cue illumination estimation via a tree-structured group joint sparse representation[J]. International Journal of Computer Vision, 2016, 117, (1): 21-47
- [33] Z. Chen, Z. Hong, D. Tao. An experimental survey on correlation filter-based tracking[J]. arXiv preprint arXiv:1509.05520, 2015:
- [34] 张全, 鲍华, 饶长辉, 彭真明. GPU 平台二维快速傅里叶变换算法实现及应用[J]. 光电工程, 2016, 43, (02): 69-75
- [35] N. Wang, J. Shi, D.-Y. Yeung, J. Jia. Understanding and diagnosing visual tracking systems[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 3101-09
- [36] V. Naresh Boddeti, T. Kanade, B. Vijaya Kumar. Correlation filters for object alignment[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2291-98
- [37] H. Kiani Galoogahi, T. Sim, S. Lucey. Multi-channel correlation filters[C]. Proceedings of the IEEE international conference on computer vision, 2013: 3072-79
- [38] M. Li, Z. Peng, P. Zhang. Robust object tracking based on structural local sparsity via a global L<sub>2</sub> norm constraint[C]. Infrared Technology and Applications, and Robot Sensing and Advanced Control, 2016: 1015719
- [39] A. Yilmaz, O. Javed, M. Shah. Object tracking: A survey[J]. ACM computing surveys (CSUR), 2006, 38, (4): 13
- [40] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan. Object detection with discriminatively trained part-based models[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 32, (9): 1627-45
- [41] T. Liu, G. Wang, Q. Yang. Real-time part-based visual tracking via adaptive correlation filters[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 4902-12
- [42] P. Ade. Planck 2013 results. IV. Low Frequency Instrument beams and window functions[J]. 2014:
- [43] M. Wang, Y. Liu, Z. Huang. Large margin object tracking with circulant feature maps[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4021-29

- [44] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection[C]. international Conference on computer vision & Pattern Recognition (CVPR'05), 2005: 886--93
- [45] T. Kanade, Y. Tian, J. F. Cohn. Comprehensive database for facial expression analysis[C]. fg, 2000: 46
- [46] 汪劲希, 燕肇一, 彭真明. 复杂背景下海天线的连通域检测法[J]. 舰船科学技术, 2016, 38, (15): 108-12
- [47] T. Ahonen, A. Hadid, M. Pietikainen. Face description with local binary patterns: Application to face recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, (12): 2037-41
- [48] T. Mäenpää, M. Pietikäinen. Classification with color and texture: jointly or separately?[J]. Pattern recognition, 2004, 37, (8): 1629-40
- [49] X. Feng, M. Pietikainen, A. Hadid. Facial expression recognition with local binary patterns and linear programming[J]. Pattern Recognition And Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii, 2005, 15, (2): 546
- [50] B. Berlin, P. Kay. Basic color terms: Their universality and evolution[M]. Univ of California Press, 1991
- [51] L. D. Griffin. Optimality of the basic colour categories for classification[J]. Journal of the Royal Society Interface, 2005, 3, (6): 71-85
- [52] K. Choi, K.-A. Toh, H. Byun. Realtime training on mobile devices for face recognition applications[J]. Pattern recognition, 2011, 44, (2): 386-400
- [53] 许俊平, 彭真明, 张启衡, 徐智勇. 强背景下模糊闪烁成像目标探测与跟踪[J]. 强激光与粒子束, 2008, (04): 537-41
- [54] 曾义, 彭真明. 一种基于视觉特征的多分辨率快速图像融合方法[J]. 成都信息工程学院学报, 2007, (04): 509-12
- [55] 刘世军, 彭真明, 赵书斌, 张启衡. 基于混沌粒子滤波的视频目标跟踪[J]. 光电工程, 2010, 37, (07): 16-23+35
- [56] 彭真明, 张启衡, 魏宇星, 张覃平. 基于多特征融合的图像匹配模式  $\Xi$  [J]. 强激光与粒子束, 2004, 16, (3):
- [57] E. Erdem, S. Dubuisson, I. Bloch. Fragments based tracking with adaptive cue integration[J]. Computer vision and image understanding, 2012, 116, (7): 827-41

- [58] 彭真明, 景亮, 何艳敏, 张萍. 基于多尺度稀疏字典的多聚焦图像超分辨融合[J]. 光学 精密工程, 2014, 22, (1): 169-76
- [59] 钟露. 复杂背景下红外弱小目标的正演模拟与仿真技术研究[D]. 成都: 电子科技大学, 2016, 10-14
- [60] 史漫丽, 彭真明, 张启衡, 李全忠, 林志勤. 基于自适应侧抑制网络的红外弱小目标检测[D]. 成都: , 2011, 20-22
- [61] 官安全. 复杂背景下扩展目标亚像素分级跟踪方法研究[D]. 成都: 电子科技大学, 2007, 30-56
- [62] J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 779-88
- [63] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg. Ssd: Single shot multibox detector[C]. European conference on computer vision, 2016: 21-37
- [64] H. Nam, M. Baek, B. Han. Modeling and propagating cnns in a tree structure for visual tracking[J]. arXiv preprint arXiv:1608.07242, 2016:
- [65] 王晓阳, 彭真明, 张萍, 孟晔铭. 局部对比度结合区域显著性红外弱小目标检测[J]. 强激光与粒子束, 2015, 27, (09): 32-38
- [66] 方浩宇. 序列图像中的目标检测与跟踪方法研究[D]. 成都: 电子科技大学, 2006, 32-43
- [67] Y. Wu, J. Lim, M.-H. Yang. Online object tracking: A benchmark[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2013: 2411-18
- [68] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88, (2): 303-38
- [69] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Čehovin Zajc, T. Vojir, G. Hager, A. Lukežić, A. Eldešokey. The visual object tracking vot2017 challenge results[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1949-72
- [70] 黄振星. 基于跟踪—学习—检测的稳健目标跟踪算法设计与实现[D]. 成都: 电子科技大学, 2015, 15-22
- [71] 郑欣. 返导航系统中目标检测与跟踪技术研究及其 DSP 实现[D]. 成都: 电子科技大学, 2007, 60-66

[72] 刘杰才. 成像跟踪系统中摄像机标定算法研究及 DSP 实现[D]. 成都: 电子科技大学, 2012, 45-

57

## 攻读硕士学位期间取得的成果

### 一、专利申请

- [1] 彭真明, 李梦雪, 田明, 王璟璟, **卢耀坤**, 李健, 张兰丹, 曹思颖, 张萍, 蒲恬, 一种流水作业平台烟条自动视觉检测与自动计数方法[P], 申请号: 201710044136.6, 申请日: 2017-01-19.
- [2] 彭真明, 吴逢信, 张明英, **卢耀坤**, 黄苏琦, 刘雨菡, 梁航, 贲庆妍, 张鹏飞, 张兰丹, 杨春平, 一种遥感图像高空卷云的检测方法[P], 申请号: 201810442692.3, 申请日: 2018-05-10.
- [3] 周子玉, 彭真明, 张明英, **卢耀坤**, 饶紫鹏, 曹思颖, 宋立, 马沪敏, 刘安, 王卓然, 杨春平, 一种针对恶劣天气条件下的低质图像增强方法[P], 申请号: 201810366499.6, 申请日: 2018-04-23.

### 二、参与科研项目

- [1] 北京优创新港科技股份有限公司合作项目, “基于视觉的传动平台烟条分类与识别算法研究”, 课题年限: 2016.05-2017.05. 主要负责烟条分割识别算法的算法仿真、算法结果性能测试。
- [2] 四川长虹科技股份有限公司合作项目, “图像增强与电子稳像技术算法研究”, 课题年限: 2017.10-2018.05. 作为主研人之一, 主要负责图像增强模块的算法设计、代码编写、性能测试以及软件平台移植。
- [3] 四川长虹科技股份有限公司合作项目, “扩展目标稳定跟踪技术算法研究”, 课题年限: 2017.10-2018.05. 作为主研人之一, 主要负责运动目标跟踪模块的算法设计、代码编写、性能测试以及软件平台移植。