

分类号 TP391

学号 13050039

UDC 681

密级 公开

工学硕士学位论文

监控视频异常事件检测方法研究

硕士生姓名 谭瀚霖

学科专业 控制科学与工程

研究方向 多媒体信息系统与虚拟现实技术

指导教师 张茂军 教授

国防科学技术大学研究生院

二〇一五年十二月

Researches on Abnormal Events Detection of Video Surveillance

Candidate: Hanlin Tan

Advisor: Professor Maojun Zhang

A dissertation

Submitted in partial fulfillment of the requirements

for the degree of Master of Engineering

in Control Science and Engineering

Graduate School of National University of Defense Technology

Changsha, Hunan, P. R. China

December 2, 2015

独创性声明

本人声明所呈交的学位论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除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和撰写过的研究成果，也不包含为获得国防科学技术大学或其他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。

学位论文题目： 监控视频异常事件检测方法研究

学位论文作者签名： 谭伦霖 日期： 2015 年 11 月 27 日

学位论文授权使用授权书

本人完全了解国防科学技术大学有关保留、使用学位论文的规定。本人授权国防科学技术大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档，允许论文被查阅和借阅；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。

(保密学位论文在解密后适用本授权书。)

学位论文题目： 监控视频异常事件检测方法研究

学位论文作者签名： 谭伦霖 日期： 2015 年 11 月 27 日

作者指导教师签名： 张 毅 日期： 2015 年 11 月 27 日

目 录

摘 要	i
ABSTRACT	ii
第一章 绪论	1
1.1 课题背景及意义	1
1.1.1 课题背景	1
1.1.2 研究意义	2
1.2 国内外相关研究现状	3
1.2.1 对异常视频序列建模	4
1.2.2 对正常视频序列建模	4
1.2.3 其他方法	5
1.3 论文研究的主要问题	5
1.4 论文的主要工作与结构	7
1.4.1 本文的主要工作	7
1.4.2 本文的组织结构	8
第二章 监控视频异常检测的基础知识	10
2.1 描述视频事件的特征提取	10
2.1.1 描述视频事件的特征概述	10
2.1.2 光流直方图	10
2.1.3 3D 梯度	13
2.1.4 前景计数器	13
2.1.5 复合描述子	14
2.2 异常检测方法	15
2.2.1 异常检测的基本原理	15
2.2.2 基于聚类的异常检测	16
2.2.3 基于统计模型的异常检测	16
2.2.4 基于分类的异常检测	16
2.2.5 基于稀疏表示的异常检测	19
2.3 本章小结	20
第三章 监控视频中的异常事件检测方法	22
3.1 引言	22

3.2	基于运动先验信息的视频光流直方图特征提取	22
3.2.1	改进 ViBe 算法的运动检测	22
3.2.2	像素特征：快速鲁棒视频光流提取	23
3.2.3	聚合特征：区域光流直方图	26
3.3	异常检测模型	28
3.3.1	基于朴素最大值模型的异常检测	29
3.3.2	基于单类 SVM 的异常检测	29
3.3.3	基于稀疏表示的异常检测	30
3.4	监控视频异常检测系统的实现	30
3.5	本章小结	31
第四章	实验与分析	33
4.1	引言	33
4.2	数据集	33
4.3	评价准则	35
4.3.1	异常检测性能评价	36
4.3.2	时间复杂度评价	36
4.4	监控视频分块光流直方图的特征分析	37
4.4.1	参数配置	37
4.4.2	特征分析	37
4.5	参数分析	40
4.5.1	特征参数寻优	40
4.5.2	单类 SVM 参数寻优	41
4.6	异常结果与分析	41
4.6.1	UCSD Ped1 数据集	41
4.6.2	采集的实际监控视频	44
4.7	结论	50
第五章	总结与展望	52
5.1	总结	52
5.2	展望	52
	致谢	54
	参考文献	55
	作者在学期间取得的学术成果	59

表 目 录

表 1.1	几种对正常视频序列建模方法的比较	6
表 3.1	重要的特征聚合参数	28
表 4.1	特征提取的参数寻优	41
表 4.2	单类 SVM 的参数寻优	41
表 4.3	三种模型异常检测性能对比	42
表 4.4	不同方法的性能比较。前六列为近年来 CVPR、ICCV 和 TPAMI 上发表的异常检测论文方法的结果，最后三列为本文三种模型的结果。	42
表 4.5	运行时间比较	44

图 目 录

图 1.1	人群四散逃离的异常事件（全局异常事件）	1
图 1.2	摩托车违章逆行的异常事件（局部异常事件）	2
图 1.3	行人横穿干道的异常事件	3
图 1.4	特定类别异常事件——打架斗殴	4
图 1.5	基于轨迹的方法一般流程	6
图 2.1	光流示意图。图中红色箭头代表运动方向。	11
图 2.2	从光流到光流直方图	12
图 2.3	对 UMN 数据集广场片段某 3D 块 3D 梯度随帧数变化曲线	14
图 2.4	UMN 数据集广场片段第 9 秒（265 帧）	14
图 2.5	USCDPed1 数据集中车辆出现时的 3D 梯度	17
图 2.6	梯度特征 x, y, t 通道两两之间的直方图	17
图 2.7	第 5 分块特征拟合三维高斯分布概率密度变化曲线	18
图 2.8	单类 SVM 对二维高斯分布随机数据的分类结果。(a) 是单类训练数据（红色）：单位圆内服从二维均匀分布的随机变量 (x, y) ；(b) 是单类 SVM 对测试数据的分类结果（红色为同类，蓝色异类）。	20
图 3.1	异常检测方法的训练和检测流程	22
图 3.2	在 UCSD Ped1 数据集上的运动检测结果。图 (a) 是 UCSD Ped1 数据集训练视频第 5 帧；图 (b) 是 MTP 模型建立初始背景模型演化成的背景；图 (c) 是 ViBe 方法得到的前景图象。	24
图 3.3	鲁棒光流计算流程	24
图 3.4	光流计算结果示例	25
图 3.5	点像素光流聚合为光流直方图的过程	26
图 3.6	光流直方图特征。图 (a) 为某路口场景的前景图象；图 (b) 为光流计算结果；图 (c) 为光流直方图向量可视化的结果，不同的颜色表示不同的直方图通道（交叠的网格未画出）。可以看出，逆行的摩托车直方图柱值是黄色，与正常行驶的汽车直方图的红紫色明显不同，说明光流直方图特征有效表达出了此运动方向异常。	27
图 3.7	视频监控异常检测系统流程图	31
图 4.1	UCSD ped1 数据集部分异常示例。在此数据集中，所有非行人运动被视为异常。图 (a) 为快速运动的自行车；图 (b) 为滑板运动；图 (c) 为机动车运动；图 (d) 为轮椅运动。	33

图 4.2 自采集数据集部分异常示例。在此数据集中，明显违反交通规则的行为被视为异常。图 (a) 为强穿马路的行人；图 (b) 为雨夜逆行的卡车；图 (c) 为路口逆行的摩托车；图 (d) 为路口强穿马路的行人。	34
图 4.3 第 139 块特征与 t-SNE 降维投影	38
图 4.4 第 173 块特征与 t-SNE 降维投影	39
图 4.5 在 UCSD Ped1 数据集上朴素的最大值模型的结果。图 (a) 是帧水平 ROC 曲线；图 (b) 是像素水平 ROC 曲线。图中除了最大值模型，还有 Adam 等的方法、MDT、Social Force、Sparse 和 Sparse Combination 五种方法的结果。	43
图 4.6 过马路数据集异常检测结果截图。(a) 是急转弯的车辆；(b) 是在非机动车道上逆行的摩托车；(c) 和 (d) 均为横穿马路的行人。	45
图 4.7 路口一数据集异常检测结果截图	46
图 4.8 路口二数据集异常检测结果截图。(a) 是横穿马路的行人；(b) 是逆行的摩托车；(c) 也是横穿马路的行人；(d) 是越黄线行驶的摩托车；(e) 是头顶扛麻袋的行人；(f) 是双层巴士。	47
图 4.9 雨夜数据集异常检测结果截图。(a)、(b)、(c) 均为跨过道路中线逆行的两辆卡车；(d) 为横穿道路的汽车。	48
图 4.10 异常检测误检的例子	49

摘要

监控视频异常检测是智能监控技术的重要组成部分，其研究价值源于其潜在的广阔应用需求。异常检测从原理上看，可以对异常数据建模，也可以对正常数据建模。但异常数据难于收集，因而对正常数据建模是更通用的方法。在对正常数据建模的方法中，本文总结了基于分块处理的视频异常检测算法框架，具体进行了以下四项工作：

(1) 总结了基于分块处理的视频异常检测的原理、常用的特征和模型。视频异常检测常用的特征有光流直方图、3D 梯度、前景计数器和复合描述子。光流直方图能直观反映运动目标速度大小和方向，3D 梯度能综合纹理和运动信息，前景计数器能感知场景的运动密度，复合描述子可以综合两种以上的基本特征。异常检测是找出行为很不同于预期对象的过程，从数据挖掘的角度看，有基于聚类、基于统计模型、基于分类等多种方法。异常检测是高度依赖于应用的一类问题，只能依据具体应用去选择方法。本文综合近年来优秀论文中的异常检测模型，最终选取了朴素的最大值模型、单类 SVM 和稀疏模型作为候选方法。

(2) 实现了一个视频异常检测系统。它以光流直方图特征为主的特征描述子作为特征，以朴素的最大值模型、单类 SVM 和稀疏模型作为模型，可以用这些特征模型组合进行视频异常检测。对有真值的数据集，可以对方法性能进行测评。在准备好训练视频和测试视频后，该系统可以一键完成模型训练、异常检测和结果评估。

(3) 在标准数据集上对比了本文实现的方法和近年来的六种优秀视频异常检测方法的性能和速度，并测试了本文方法在实际监控视频上的效果。在标准数据集上的对比结果说明了本文光流直方图特征和最大值模型组合在视频异常检测，特别是对异常事件的定位上的有效性。本文方法在标准数据集上达到实时 (30.67 FPS) 运行的速度也超越了除了 Sparse Combination 外的所有方法。

(4) 在实际监控视频中，不同视角、距离、光照、气象条件下，本文方法能够检测出车流人流中的有意义的异常事件。这也说明了本文光流直方图特征和最大值模型组合方法的有效性。另外，从实验结果还可以分析出误检的漏检的主要原因分别是遮挡、光照差和目标尺度小、光照差。

关键词: 智能监控; 异常检测; 光流直方图; 单类 SVM; 稀疏表示

ABSTRACT

Abnormal event detection of video surveillance is a cross research which requires knowledge and skills on both video processing and data mining. The significance of this research lies in the potential and vast demand on video analysis. Modeling normal data and abnormal data are two possible solutions for anomaly detection. However, it is difficult to collect adequate abnormal data. Therefore, modeling normal data is the more commonly used method. The thesis summarizes the framework for patch based visual anomaly detection methods. The major work consists of the following four parts:

(1) We summarize the principles, features and models the state-of-art methods adopt. The commonly used features for visual anomaly detection includes HOF, 3D-gradient, foreground counter and a descriptor of combined features. HOF is able to catch the speed and direction of motion; 3D-gradient is able to combine texture information with motion; foreground counter is a map of motion density; and the descriptor contains information from multiple features. Anomaly detection is a process digging out objects with unexpected behaviors. In data mining, the methods can be classified into cluster-based, stochastic, classification based, etc. Anomaly detection heavily relies on the application. We extract features and models from the recent published works seeking for the best combination. We select the maximum model, one-class SVM and sparse model as candidate models for evaluation.

(2) We implement an abnormal event detection system of video surveillance. The system uses HOF as the major part of the descriptor and implement the maximum model, one-class SVM and sparse model for anomaly detection. It is able to evaluate methods if the dataset contains ground-truth videos. All the training, detection and evaluation work can be done with one-click.

(3) We compare our methods with six state-of-art methods on a benchmark dataset. The results suggest that the combination of HOF feature with the maximum model is effective and rank third among all the methods on the dataset. The running speed of our method reaches real-time (30.67 FPS) and exceeds all other methods except for the Sparse Combination method, which claims to reach a detection speed of 150 FPS.

(4) Detection results on real surveillance videos suggest our method is effective in both day and night. Our method is able to detect abnormal behaviors of vehicles, mo-

torcycles and pedestrians in complicated and crowded scenes. By analyzing the results, we also find the main causes for false alarm and missed target lies in occlusion, bad light conditions and small size of targets, bad light conditions, respectively.

Key Words: Intelligent Surveillance; Anomaly Detection; HOF; One-class SVM; Sparse Representation

符号使用说明

HOG	光流直方图 (Histogram of Optical Flow)
$V = (V_x, V_y)^T$	光流向量 (Optical Flow)
$I(x, y, t)$	图像序列 (x, y, t) 处的亮度
I_x, I_y, I_t	$I(x, y, t)$ 分别对 x, y, t 的偏导数
H	光流直方图向量
B	某个时空三维块
$B(x, t)$	分块 x 在 t 时刻的忙碌度 (Busyness)
b_H	分块高度
b_W	分块宽度
b_T	分块时域长度
δ_H	分块纵向间隔, 一般取为 $0.5b_H$
δ_W	分块横向间隔, 一般取为 $0.5b_W$
m	光流直方图柱数
AUC_f	帧水平 ROC 曲线下面积
AUC_p	帧水平 ROC 曲线下面积
ΔAUC	$AUC_f - AUC_p$
AUC_H	AUC_f 和 AUC_p 的调和平均数

第一章 绪论

1.1 课题背景及意义

1.1.1 课题背景

异常，新华词典的解释是“不同于平常”。从分类的角度看，异常与正常是两个大类别，异常内部又可以分成打架、撞车等小类别。从概率的角度看，“平常”是大多数，而“异常”就是少数，所以异常事件，可解释为“小概率事件”^[2]。

异常事件的分类有很多角度。根据场景运动目标的多少，可以分为拥挤场景的异常事件和不拥挤场景的异常事件。这种分类主要根据基于跟踪和轨迹分析的异常检测方法能否适用。拥挤场景现有的跟踪方法都会失效，而不拥挤场景基于跟踪和轨迹分析的方法是可能奏效的。根据异常事件的规模，可以分为全局异常事件（如图1.1）和局部异常事件（如图1.2）。这种分类可以用于决定异常警报的级别。



图 1.1 人群四散逃离的异常事件（全局异常事件）



图 1.2 摩托车违章逆行的异常事件（局部异常事件）

1.1.2 研究意义

随着视频监控在商场、银行、小区、道路等公共场所的广泛部署，监控视频数据大量产生。目前监控视频主要还是用于威慑犯罪和事后调取，但视频智能分析的需要一直存在。近期发生了一些引起公众关注的事件再次体现了监控视频异常检测需求的迫切性。IBM 深圳公司的一名女经理在地铁口突发心脏病跌倒，虽然正对着监控，却因为监控无人查看而耽误了抢救时间，最终不幸去世。对于这种紧急情况，仅有八分钟的黄金抢救时间，不能及时发现险情和施救生命就会逝去。

监控视频的异常检测在安防领域、交通管理、城市管理方面有广阔的应用前景。从监控视频中自动发打架斗殴、交通违章、交通事故、人群聚集等事件具有及时发现事故险情，及时发现安全隐患的作用。例如图1.3中的行人违规横穿马路，说明此路段存在交通安全隐患，有必要派出交警或者增设警示标志。如果这种情况持续发生，可以考虑架设人行天桥来引导行人。而这种安全隐患靠人工是很难发现和统计的。



图 1.3 行人横穿干道的异常事件

1.2 国内外相关研究现状

对于需要检测异常事件种类的问题，多采用分类的方法。而只需要判断是否有异常事件的问题，多采用概率方法。本课题研究如何从异常视频中检测到是否有异常事件，但并不细分异常事件的种类，故以下所述的方法，都是基于概率的方法。

按照建模的角度，视频异常事件检测的方法可以分为两类：对异常视频序列建模，或者相反地，对正常视频序列建模^[3]。前者的主要思想是一种对异常事件建立分类器，对输入视频序列分类，故而又可以称为基于分类器的方法^[4]。后者的主要思想是通过统计学习对正常视频序列（或称为“参考序列”）建模，那么任何与正常视频序列模型不一致的序列就是异常序列。这种方法也被称为基于模型的方法^[4]。



图 1.4 特定类别异常事件——打架斗殴

1.2.1 对异常视频序列建模

对异常视频序列建模的方法需要按异常类别收集的大量视频数据作为训练集，提取某种特征并训练分类器，再用分类器判别测试集视频是否属于异常事件。这种方法的优势在于可以利用比较成熟的分类器技术来解决异常检测问题，但问题在于：1) 异常事件发生概率小，很难收集大量的异常视频数据；2) 训练和测试视频可能来自很多不同场景，实现跨场景的异常检测存在很大挑战。例如图 1.4 所示的四个场景中都存在打架斗殴的异常事件，他们发生的场景、视频角度、光照各异，想要跨场景检测确有难度。

1.2.2 对正常视频序列建模

对正常视频序列建模的方法针对要使用的场景提取正常视频作为训练集，提取某种特征并建立模型，用模型来判别测试集视频是否属于异常事件。这种方法不需要收集大量异常视频提供先验信息，也不存在跨场景识别的问题，因而具有

更广泛的适用性。本课题将对基于模型的方法进行研究。对正常视频序列建模的异常检测，本质上是一个概率问题，即判断一个事件是否是大概率事件。

表1.1梳理最近十年来领域会议和期刊对正常视频序列建模的异常检测方法，可以发现这些方法^[2, 3, 5-10]大多可以归纳为以下几个步骤（或其子集）：

1) 计算像素特征。常用的像素特征有光流直方图^[2, 7]和空间梯度。考虑时间维度时梯度可以加上时间维变成三维(3D)梯度^[6, 10]。Jodoin 等利用前景计数器^[3, 8]作为特征，即统计每个像素在单位时间内处于前景的次数。Kratz 等将以上特征组合起来作为局部特征描述子^[9]获得了更好的检测效果。

2) 在时空域上聚合特征。时空聚合的主要方法就是在时空 3D 块内对像素特征求和，削减噪声的影响，以增强特征的鲁棒性。也有文章直接基于 3D 块来提取特征，如 Manadevan 等人使用的动态纹理特征^[5]。

3) 对聚合特征建模。对于前景计数器特征，可以直接使用训练特征各个通道的最大值比较作为模型。对于光流、梯度等向量特征，有稀疏编码^[2, 6]、多维高斯模型^[10]、单类 SVM^[7]模型。对于动态纹理特征，有显著性模型^[5]。

4) 利用模型对比测试序列和参考序列。比较方法因模型而异，大多需要设置一个阈值来区分正常与异常。对于计数器模型，就直接对比计数值。对于稀疏编码，对比稀疏重构系数。对于多维高斯等概率模型，对比事件出现的概率。对于单类 SVM、显著性模型等，其相应的理论方法能够直接判别测试序列是否异常。

1.2.3 其他方法

关于视频事件检测、识别、分类还有众多方法^[11]。早期的一些研究提取稀疏（角点、固定点）特征^[12-16]，近期的方法更倾向于使用稠密特征，因为稠密特征具有更好的性能^[17]，如 2.2 节中提到的 3D 块中的光流直方图、3D 梯度等特征。光流、梯度特征具有互补性，融合它们可以取得更好的检测识别效果^[18]。也有将亮度、光流、梯度联合作为特征的，为了削减特征维度，可能还会使用 PCA 等降维方法^[19]处理特征。关于视频事件分类，还有启发式规则、马尔科夫模型、贝叶斯网和核方法（Kernel methods，如 SVM）^[11]。

另一类异常检测方法是基于轨迹分析的方法^[20-23]。基于轨迹方法的一般流程如图 5 所示。例如^[20, 24]基于跟踪获取运动目标轨迹再进行两阶段聚类获取场景模型。但是基于轨迹的方法的缺陷在于场景如果是拥挤的（存在大量交叠的运动目标），跟踪算法往往会失效，基于轨迹的方法也就失效了。

1.3 论文研究的主要问题

本课题基于正常视频序列建模的方法来研究异常检测问题，需要解决的主要问题有：特征的选取、模型的选取和在大量视频数据上对方法进行评估。

表 1.1 几种对正常视频序列建模方法的比较

方法	特征	模型	检测	优缺点
Jodoin 2008 ^[3]	前景计数器 (像素点处于前景的次数)	固定时间长度内累积特征的最大值	固定时间长度内累积特征大于模型最大值	对方向异常无效; 快速很快
Kratz and Nishino 2009 ^[10]	3D 梯度	三维混合高斯模型	用 KL 散度度量特征分布之间的距离	场景越拥挤效果越好
Mahadevan 2010 ^[5]	3D 块混合动态纹理	图像显著性模型	图像显著性检测	精度较好; 速度很慢
Cong 2011 ^[2]	3D 块多尺度光流直方图	光流直方图建立稀疏字典	稀疏重构代价阈值分割	精度高; 速度很慢; 使用的多尺度策略在异常目标本身像素不多时会导致漏检
WANG 2012 ^[7]	3D 块光流直方图	单类 SVM	由单类 SVM 输出分类结果	用于群体异常事件监测, 精度无定量评估
Saligrama and Chen 2012 ^[9]	融合前景计数器和光流的描述子	由自定义的正负检测器、GLRT 和熵组成的异常打分器	根据异常分数值的阈值分割	精度很好
Lu 2013 ^[6]	多尺度 3D 块的 3D 梯度 (Kratz and Nishino 2009)	稀疏基表示 (本质上还是稀疏表示)	稀疏重构代价阈值分割	精度很好; 速度很快; 使用的多尺度策略在异常目标本身像素不多时会导致漏检

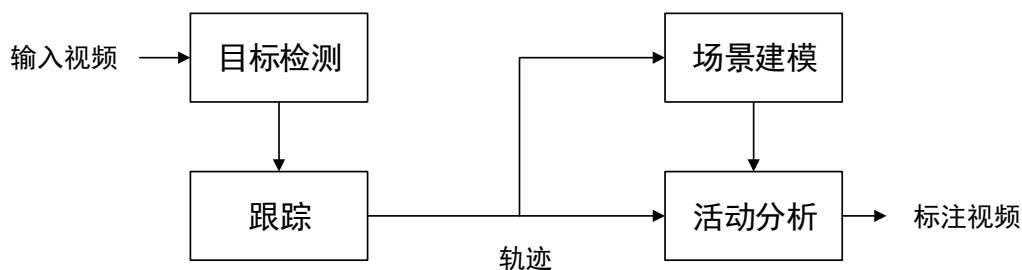


图 1.5 基于轨迹的方法一般流程

具体的, 将在以下两个方面开展研究:

一是研究适用于异常检测的视频活动的特征提取。现有的视频活动特征存在运动信息、纹理信息、前景背景信息不能兼顾的情况。因此有必要提出一种融合多种信息的视频活动特征描述子。

二是研究基于融合特征的视频异常事件检测的数学模型。针对提出的特征描述子，提出一种异常检测模型，对不同角度的特征信息归一化并学习。特征的选取思路是选用一种融合多种基本特征的描述子。首先需要解决的问题是基本特征的计算。前景计数器、梯度特征的计算流程比较简单，但光流的计算仍然是一个有待研究的问题。首先是精度，现有方法对相邻视频帧之间稠密光流计算都无法避免出错。对于颜色、纹理特征不明显的场景，计算出错的比例很大。其次是速度，对于 720p 的图像帧，计算其稠密光流的速度慢到数十秒一帧。这样的处理速度不仅无法满足异常检测的实际需求，就连实验效率也会大打折扣。在计算好基本特征的基础上，如何设计特征描述子也是一个问题。这里涉及到如何将几种不同性质、不同量纲的特征融合为一个特征描述子，如何采取适当的归一化手段使得模型的检测效果能够最优化。

关于模型，文献综述中已经提到，基本模型包括稀疏编码、多维高斯模型、单类 SVM、显著性模型等等。对于多种特征融合而成的描述子，是选用其中一种、还是多种结合，或者设计全新的模型也是一个问题。

为了测试方法的有效性，需要在各个已有的视频数据集和大量实际监控视频上进行实验评估。第一个问题就是去收集这些视频数据，另外对于没有标定真值的实际数据，需要人工判断来评估检测结果的正确性。

1.4 论文的主要工作与结构

1.4.1 本文的主要工作

针对监控视频异常检测问题，本文采用的是视频帧分块处理而非跟踪的方法，将检测问题转化为判定分块是否异常的问题，而分块与分块之间独立。因此问题可以抽象为通过学习一个块在一段时间的运动特征并建模，通过模型预测这个块在接下来一段时间内每一帧是否是异常。所以问题的核心分为两个部分：一是找出合适的描述分块运动模式的特征；二是对特征建立合适的一类学习模型。针对问题核心，本文的主要工作有：

(1) 基于运动先验信息的光流直方图特征提取的实现

在充分对比现有特征的基础上，本文选择了光流直方图作为基本特征。光流是反映运动速度大小和方向的最直接特征，光流直方图可以表征一定图象区域内的运动规律。针对图象的光流算法运算速度慢，计算结果中包含的错误多。本文利用视频的运动信息缩小光流计算的像素面积，大大加速了视频的光流提取。进

一步通过选择特征点和前后项矫正的方法过滤计算错误的光流，使得视频的光流提取错误的情况大大减少。

(2) 朴素的最大值模型、单类 SVM 模型和稀疏模型的实现

本文根据对特征分析的结果，结合近年来视频异常检测方法提出的模型，选取了朴素的最大值模型、单类 SVM 和稀疏模型作为考察模型。将这些模型分别与光流直方图组合，对比产生的检测结果，选取最优模型。

(3) 对所提特征与模型组合性能的评估分析

为了评估本文方法，选取近年来视频异常检测论文最常用的 UCSD Ped1 数据集作为标准数据集进行性能测试，并与近年来 CVPR、ICCV 和 TPAMI 上优秀论文的结果进行对比。此外，本文还采集了四段城市道路监控作为数据集，测试本文方法在实际监控中的效果，并分析方法的优点以及误检、漏检的原因。

1.4.2 本文的组织结构

第一章: 绪论。本章主要介绍了监控视频异常事件检测的背景及意义。从对异常视频建模和对正常视频建模两个方面，总结了国内外研究现状，概括了视频异常检测的最新研究成果。并明确了本文研究的主要问题。在此基础上，介绍了本文的主要工作和了论文结构。

第二章: 监控视频异常检测的基础知识。监控视频的异常检测分为特征提取和异常检测两大步骤，本章介绍了描述视频事件的特征提取和异常检测方法。特征提取方面，本章介绍了光流直方图、3D 梯度、前景计数器和复合描述子四种特征的含义、数学描述和提取方法。异常检测，也称为离群点检测 (Outlier Detection)，并非视频分析中独有的问题。在医疗处理、公共安全、网络入侵检测中也有类似的问题。本章将把异常检测抽象为一般问题，介绍其基本原理和方法，并具体介绍了基于聚类的异常检测方法、基于统计模型的异常检测方法、基于分类的异常检测方法和基于稀疏表示的异常检测方法。其中，基于分类的异常检测方法又包括朴素的最大值模型和单类 SVM 模型。

第三章: 监控视频中的异常事件检测方法。本章详细叙述了基于运动先验信息的视频光流直方图特征提取方法，包括改进 ViBe 的运动检测方法，快速鲁棒视频光流提取和区域光流直方图的计算；以数学模型形式描述了最大值模型、单类 SVM 模型和稀疏表示模型用于异常检测的公式。本章明确了特征提取和模型异常检测的主要参数符号，为实验评估打下基础。

第四章: 实验与分析。本章首先介绍了用于实验评估的数据集，包括标准数据集 UCSD Ped1 和自采集的数据集。其次，对视频异常检测的评价标准做了讨论，在原有的帧水平 AUC 和像素水平 AUC 两个评价指标的基础上，提出了综合性性能指标。接着将本文的方法模型在标准数据集上进行了测试对比，并与近年来

的优秀方法做了比较。本文还在自采集的实际监控视频上测试了最大值模型，证明了本文方法的有效性，也分析了误检、漏检的主要原因。

第五章：总结与展望。本章总结了本文的主要工作内容，并阐述了此领域的特点、存在的问题、本文方法的局限性和改进方向。本文总结了基于分块处理的视频异常检测的原理、常用的特征和模型；现了一个视频异常检测系统；并在标准数据集和实际监控上对本文方法进行了评估。相比一些比较成熟的研究领域，视频异常检测公开数据集和代码匮乏制约着领域发展。异常检测这个问题本身也并不是在所有场景下都有明确答案。更好的检测效果可能需要更加工程化的处理，包括对目标场景更细致的分类，以及讲目标分类与运动信息结合等处理。

第二章 监控视频异常检测的基础知识

上一章提到本文采用的是视频帧分块处理的方法，那么接下来的任务是如何提取分块的特征并建立模型。本章将介绍分块特征提取和建模的基础知识：常用特征和异常检测方法。通过对特征和方法的分析，将在下一章引出本文的方法。

2.1 描述视频事件的特征提取

2.1.1 描述视频事件的特征概述

为了进行异常检测，首先需要将视频帧数据转化为特征。数学上能够接受的特征形式，往往是向量形式。表1.1梳理的特征中，光流直方图、3D 梯度、前景计数器都是向量形式的特征。而动态纹理可以看成一种矩阵形式的特征。这些特征中，光流直方图使用最广泛，也是目前公认有效的特征；其次是 3D 梯度和前景计数器。动态纹理需要从整帧提取，不适用于分块方法，因此本文不予考虑。

特征按关注的方面划分，可以分为运动特征和纹理特征。具体采用哪一种特征也需要根据应用场景来确定。例如，在道路上，运动特征就比纹理特征好。因为人们不在意运动目标是什么问题，只在目标的运动模式是否“与众不同”。而如果监控场景是人行道，那么纹理特征可能更好。因为其他类别的运动目标都是异常。采用纹理特征需要运动场景的正常纹理能够被训练视频有效覆盖，否则误检会很多。

将视频分块信息转化为特征往往需要经过两步：计算像素特征和在时空域上聚合特征。特征提取的原子操作一般在像素级别上；为了增强鲁棒性平滑噪声，需要对这些像素级别的特征通过在一定的时间、空间范围内求和。所以所得的特征又称为“3D 块特征”。

2.1.2 光流直方图

19 世纪 40 年代，美国心理学家 Gibson 为了描述对动物的视觉刺激首次提出了光流的概念^[25]。光流 (Optical flow 或 Optic flow) 是物体、表面或者边缘在视觉场景中由于观察者和场景相对运动形成的可见运动模式。对于图像序列，光流可以看成同一个空间点在前后图像之间的像素位置对应关系，如图2.1。下面给出光流的数学描述：

设 (x, y, t) 表示图像序列在时间 t 处，图像位置 (x, y) 处； $I(x, y, t)$ 为该处图像亮度。如果下一帧图像移动为 $(\Delta x, \Delta y, \Delta t)$ ，根据亮度一致性假设，有

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (2.1)$$



图 2.1 光流示意图。图中红色箭头代表运动方向。

当图像移动很小，对 $I(x + \Delta x, y + \Delta y, t + \Delta t)$ 作泰勒展开可得

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + o(\Delta x, \Delta y, \Delta t) \quad (2.2)$$

略去高阶项，结合式 (2.1) 和式 (2.2)，可得

$$\frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0 \quad (2.3)$$

两端同时除以 Δt 得到

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial t} = 0 \quad (2.4)$$

记 $\frac{\partial I}{\partial x}$ 为 I_x ， $\frac{\partial I}{\partial y}$ 为 I_y ， $\frac{\partial I}{\partial t}$ 为 I_t ，则

$$I_x V_x + I_y V_y = -I_t \quad (2.5)$$

式 (2.5) 称为光流约束方程 (Optical Flow Constraint Equation)。其中 $(V_x, V_y)^T$ 为 $I(x, y, t)$ 的速度在空域上的分量，也就是光流。从光流的推导中，可以看出光流计算必须满足两个条件：一是亮度一致性假设；二是像素位移不能过大。另外，光流约束方程式一个欠定方程，需要增加额外的约束条件才能求解。

求解光流方程的经典方法是 1981 年卡耐基梅隆大学 Lucas 和 Kanade 提出的 LK 法^[26]。LK 法增加了一个假设：像素点 p 的邻域内的像素与像素点 p 具有相同的位移。则可以得到以下方程组：

$$\begin{aligned} I_x(q_1)V_x + I_y(q_1)V_y &= -I_t(q_1) \\ &\vdots \\ I_x(q_n)V_x + I_y(q_n)V_y &= -I_t(q_n) \end{aligned} \quad (2.6)$$

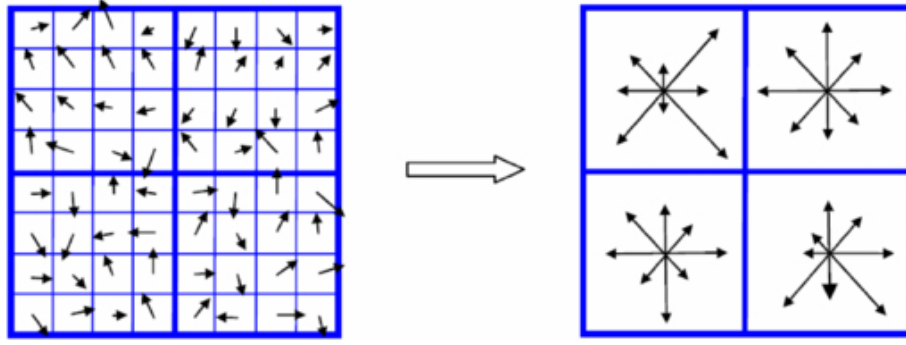


图 2.2 从光流到光流直方图

其中 q_1, \dots, q_n 为像素点 p 的邻域像素点。它可以写成矩阵形式 $Av = b$ ，其中

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ \vdots & \vdots \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, \quad b = \begin{bmatrix} -I_t(q_1) \\ \vdots \\ -I_t(q_n) \end{bmatrix}, \quad v = \begin{bmatrix} V_x \\ V_y \end{bmatrix} \quad (2.7)$$

式 (2.7) 是一个超定方程组，用最小二乘法可以写出最优解的形式为

$$v = (A^T A)^{-1} A^T b \quad (2.8)$$

以上方法中每一个邻域像素点的贡献是一样的。为了体现“越邻近的点有更大作用”的原则，可以考虑给距离像素点 p 近的邻域像素点更高的权重。设邻域点的权重向量为 $w \in R^n$, W 为对角矩阵且 $W_{ii} = w_i$ 。则最优解 v 的形式为

$$v = (A^T W A)^{-1} A^T W b \quad (2.9)$$

这就是 LK 法求解光流的基本原理。

光流是像素特征，伴随着大量噪声。由于实际图像序列可能不满足光流的两个条件，计算出错误的光流也在所难免。此时光流直方图 (Histogram of Optical Flow)^[27] 就应运而生了。图 2.2 直观展示了由光流得到光流直方图的过程：首先将图像分块；然后将每个分块中的所有光流投影到邻近的两个直方图角度上并累加。

设某图像分块的光流直方图 (向量) H 有 m 个分量，每个分量对应的角度分别为 h_1, \dots, h_m ，分块中有 n 个光流 v_1, \dots, v_n 。记光流集合

$$S_i = \{v_k \mid |\arccos(\frac{v_k^T h_i}{\|v_k\| \cdot \|h_i\|})| < \frac{2\pi}{m}\} \quad (2.10)$$

则光流直方图分量

$$H_i = \sum_{k \in S_i} \frac{v_k^T h_i}{\|h_i\|} \quad (2.11)$$

其中 $i = 1, 2, \dots, m$.

光流直方图累加了分块内光流的信息，平滑了噪声，也突出了块的主要运动方向和速度大小，是表示分块运动模式的一种有效特征。

2.1.3 3D 梯度

3D 梯度，又称为时空运动模式^[10]，是在视频时空三维块上定义的一种梯度特征。梯度特征对于光度变化和几何变换都具有局部不变性，因此是一种鲁棒的特征。对于三维块 B 中的像素 I ，其三维梯度定义为

$$\nabla B_I = [B_{I,x} \ B_{I,y} \ B_{I,t}]^T = \left[\frac{\partial I}{\partial x} \ \frac{\partial I}{\partial y} \ \frac{\partial I}{\partial t} \right]^T \quad (2.12)$$

其中 x, y, t 分别为水平、竖直和时间维。

3D 梯度是一种像素特征，每个像素的 3D 梯度都代表着分块内的一种运动模式和纹理信息。如何聚合它为模型所用也存在多种方法。此特征的提出者 Kratz 选择的是计算块内像素 3D 梯度的均值和协方差：

$$\mu = \frac{1}{N} \sum_{I \in B} \nabla B_I, \quad \Sigma = \frac{1}{N} \sum_{I \in B} (\nabla B_I - \mu)^T (\nabla B_I - \mu) \quad (2.13)$$

均值和协方差可以作为统计模型的参数对分块的运动模式建模。

图 2.3 展示了对 UMN 数据集^[28] 广场场景某分块 3D 梯度随帧数变化情况。图 2.4 是第 265 帧场景，可以看到人群在四散逃窜，同时 3D 梯度曲线各个通道出现剧烈抖动，说明 3D 梯度对于场景异常事件具有一定的敏感性。

2.1.4 前景计数器

前景计数器，顾名思义，是单位时间内固定位置像素点出现在前景中的次数。由 Jodoin 等 [3] 首先用于视频的异常检测。它也是一种像素特征。设 $\chi(x, y, t)$ 表示 t 时刻像素位置 (x, y) 是否处于前景，则

$$\chi(x, y, t) = \begin{cases} 1, & (x, y, t) \in \text{Foreground} \\ 0, & \text{otherwise} \end{cases} \quad (2.14)$$

设 $L(x, y, t)$ 表示像素位置 (x, y) 处自 t 时刻起 τ 时间段内该像素位置的前景计数器，则

$$L(x, y, t) = \sum_{i=t}^{t+\tau} \chi(x, y, i) \quad (2.15)$$

前景计数器一个像素位只有一个值，可以看成只有一个维度的向量。它已经在时间维上进行了聚合，也可以像之前的特征一样在空间维再进行聚合，得到分块的前景计数器。

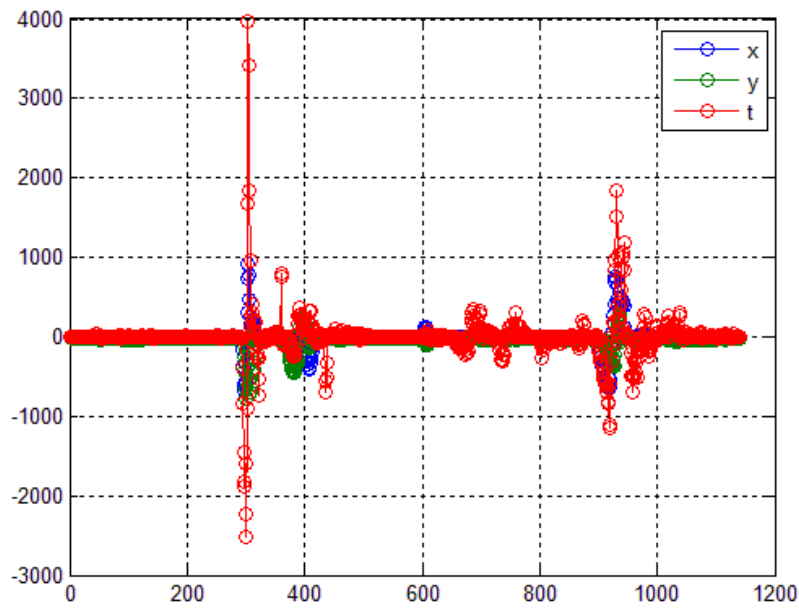


图 2.3 对 UMN 数据集广场片段某 3D 块 3D 梯度随帧数变化曲线



图 2.4 UMN 数据集广场片段第 9 秒 (265 帧)

2.1.5 复合描述子

复合描述子是将前面所述的多种特征融合为一个特征向量。特征向量的不同通道可能有不同的物理意义和量纲。这样的特征包含视频序列更丰富的信息，理论上利用复合描述子可能达到更好的异常检测精度。

但是利用复合描述子对于模型要求较高。不同量纲的特征直接放在一起反而可能引起模型精度下降。没有做适当的归一化和加权、特征之间有较高的相关性等等都会影响模型的精度。因此复合描述子可能涉及更多权重参数。对于模型建立是一个考验。

另一种思路是不同性质的特征通道使用不同的模型，最后来合并结果。这样做其实等价于多个特征和模型并用。这样做简单易行，也可以减少不同性质特征之间的干扰。缺点在于还是没有充分挖掘特征之间的关联性并以此提升检测精度。

2.2 异常检测方法

2.2.1 异常检测的基本原理

异常检测 (Anomaly Detection), 又称为离群点检测 (Outlier Detection), 是找出行为很不同于预期对象的过程。这种对象称为异常或者离群点^[29]。

从定义可以看出, 异常检测是有一定先决条件的, 即要存在一个正常对象的“群”, 这样才有可能找到群的“预期”, 从而判断新对象是否离群。基于光流特征的方法只关注物体的运动, 当某个区域没有物体运动时, 就没有光流, 即不存在一个正常的“群”, 或者说正常的“群”特征都是零。当由于训练特征集合退化为零, 一个有非零特征的物体出现在这个区域时, 模型一般会判定为异常。但实际上, 这个物体是否异常逻辑上是无法判别的。由此可以得到一个逻辑上的结论: 视频运动稀疏区域的运动模式是否异常无法判别。对于警戒级别高的应用, 可以不特殊处理; 对于一般应用, 可以假设运动稀疏区域无异常。

异常检测与传统的分类问题有着显著区别。传统的分类问题至少要有两类样本作为输入, 但因为异常样本往往难于获取, 在异常检测中, 往往只有正常样本作为输入。故而异常检测问题一类学习问题, 即给定一定数量的正常样本学习, 异常检测器需要预测接下来的样本是正常的还是异常的。

异常检测问题面临着巨大的挑战。因为异常样本往往难于获取, 从而很难对异常模式建模; 而正常样本虽然易于获取, 但是正常样本产生的模式也难以穷举。异常检测的实际效果高度依赖具体应用, 因而并没有可以保证效果的通用方法。如何去划定正常模式的边界, 必须具体问题具体分析。

一般而言, 异常检测问题可以考虑以下几种方法: 基于聚类的方法、基于统计模型的方法、基于分类的方法和其他方法。在选取方法前, 需要先考察样本的分布, 需要对样本做规范化、数据降维等处理, 通过人的观察与分析确定选择哪种方法。

2.2.2 基于聚类的异常检测

基于聚类的方法有一个默认假设：正常模式的样本是聚类的。也就是说正常模式产生的样本在某种距离度量下应该有聚簇结构。因此采用此方法前，需要考察样本是否具有聚簇结构。根据样本的聚簇结构选择相应的聚类方法。

能够椭圆形聚簇的聚类方法有 K-均值聚类、K-中心聚类。如果聚簇呈现不规则形状，可以考虑基于密度的聚类，如 DBSCAN 方法。

当样本分布呈现杂乱形态或者均匀分布时，聚类方法无法对其建模。此时应当考虑采用不同的距离度量或者采用其他异常检测方法。

2.2.3 基于统计模型的异常检测

当样本分布呈现某种统计分布形态的时候，基于统计模型的异常检测可能有效。当样本收到多方面因素影响时，其分布很有能呈现高斯分布。

多维高斯模型基于多维高斯分布^[30]。设 $x = (x_1, \dots, x_n)^T, i = 1, 2, \dots, n$ 是满足多维高斯分布的随机向量， μ 为其均值向量， K_{xx} 为其协方差矩阵，则其概率密度函数为

$$f_x(x) = \frac{1}{(2\pi)^{n/2} [\det(K_{xx})]^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)^T K_{xx}^{-1}(x - \mu)\right] \quad (2.16)$$

为了评估可行性，在 UCSDped1 数据集^[31] 上对 3D 特征和三维高斯分布进行了简单实验。图 2.5 是数据集某测试视频中的 42 帧，其中第 5 分块有一辆面包车经过，红绿蓝三色柱值表示 3D 梯度的三个维度值，自上而下的柱表示值为负。因为训练视频中只有行人，车辆在 UCSDped1 中被认为是异常，可以看到 3D 梯度缺失红色柱，与其他分块明显不同。图 2.6 是此分块在测试视频 200 帧内的三维梯度特征的二维投影，从图形直观上可以看出具有高斯分布的特征。图 2.7 是此分块 3D 梯度特征三维高斯概率密度随帧数变化曲线。可以看到在 40 帧前后明显出现谷值，表明出现了低概率事件。这与视频中车辆出现的情况对应，也说明了多维高斯分布具有一定的有效性。

2.2.4 基于分类的异常检测

传统的分类问题至少应具有两类训练数据。异常检测应用有大量的正常类别的数据，但是往往少有甚至没有异常类的数据。这意味着我们需要根据单类数据划定其边界。本节介绍朴素的最大值模型和单类 SVM 两种方法。

2.2.4.1 朴素的最大值模型

如何寻找高维训练数据的边界呢？一种最朴素的思想寻找 n 维训练数据每一个维度的最大值，构成一个忙碌度向量 (Busyness)，那么所有的训练数据一定包

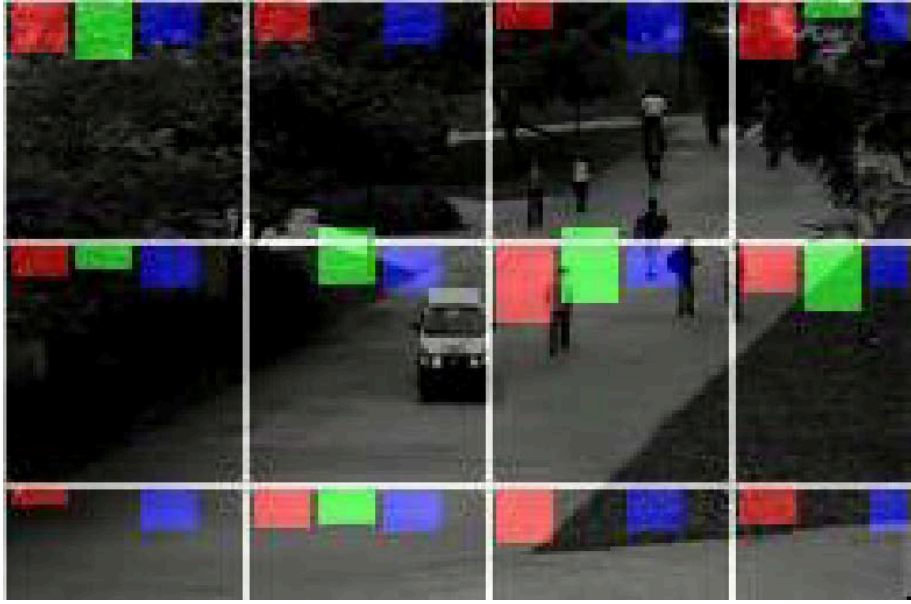
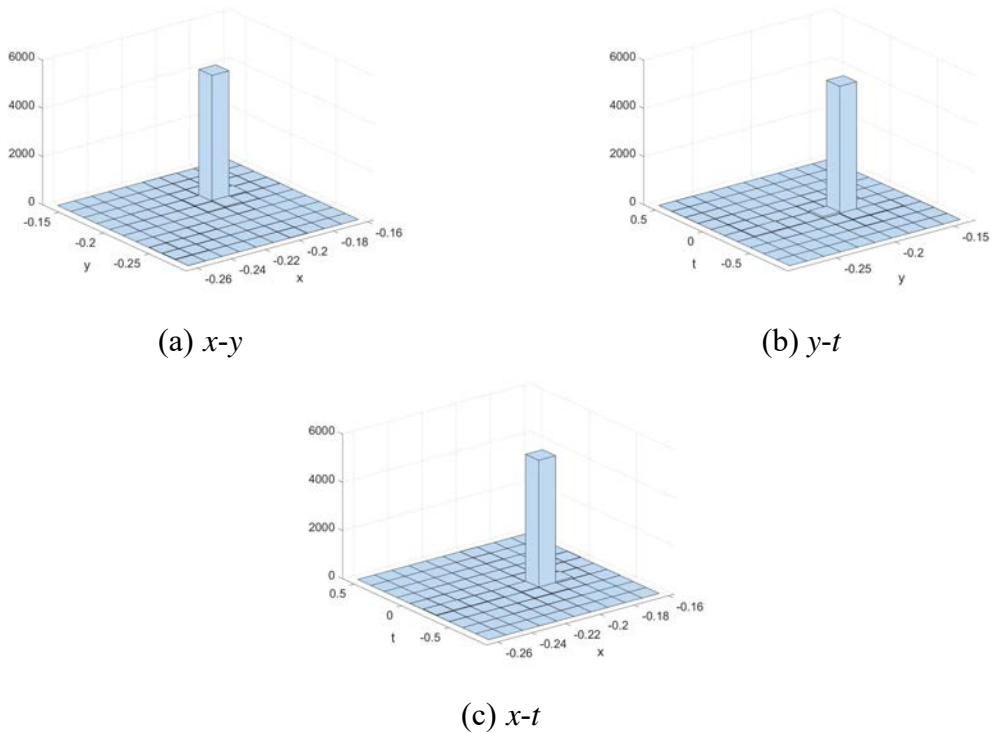


图 2.5 USCDPed1 数据集中车辆出现时的 3D 梯度

图 2.6 梯度特征 x, y, t 通道两两之间的直方图

含在零向量到 **Busyness** 边界划定的 n 维空间中。如果一个测试数据的某一个维度超过 **Busyness** 对应维度一定阈值，那么可以认为这个测试数据是异常数据。这个思想由 Jodoin^[3] 提出并用于行为分析。

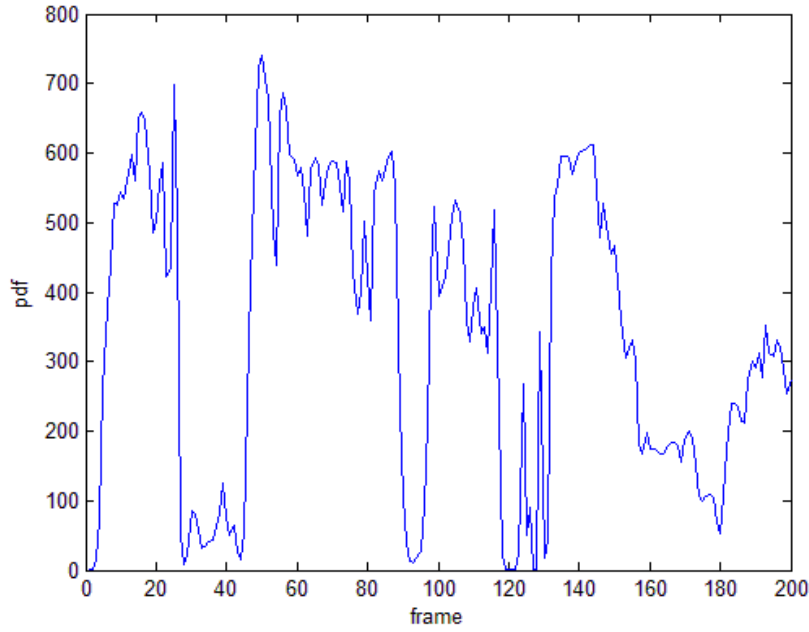


图 2.7 第 5 分块特征拟合三维高斯分布概率密度变化曲线

设训练数据为 n 维非负向量 $F(x, y, t)$ 的集合，其中 (x, y) 是空间像素位置， t 是时刻（或帧数），则训练数据的 Busyness 为

$$B_i(x, y) = \max_t F_i(x, y, t), i = 1, 2, \dots, n. \quad (2.17)$$

其中 F_i 、 B_i 分别表示 F 、 B 的第 i 维，所得的最大值 $B(x, y)$ 也是一个向量。这种朴素的寻找边界的思想具有简单快速的优点，也能保证一定的准确率和召回率。

2.2.4.2 基于单类 SVM 的异常检测

单类 SVM (one-class SVM)，也称为“一类 SVM”，是专门用于数据异常检测的一种支持向量机^[1, 32]。单类 SVM 的思想是先将训练样本向量映射到高维，寻找一个高维超球，尽可能包含大多数训练样本并且超球的体积较小。设向量集合 $x_1, \dots, x_n \in R^p$ 为训练数据， $\Phi: R^p \rightarrow R^q$ 为特征映射函数，单类 SVM 可以表示为如下优化问题：

$$\begin{aligned} \min_{r \in R, \zeta \in R^p, c \in R^q} \quad & r^2 + \frac{1}{vn} \sum_i \zeta_i \\ \text{s.t.} \quad & \|\Phi(x_i) - c\|^2 \leq r^2 + \zeta_i, \zeta_i \geq 0, i = 1, \dots, n \end{aligned} \quad (2.18)$$

其中 v 是参数，其值越大表示容纳越少的训练样本；反之，其值越小容纳越多的训练样本。为了求解此问题，写出式2.18的 Lagrange 函数

$$L(r, \zeta, c, \alpha, \beta) = r^2 + \frac{1}{vn} \sum_{i=1}^n \zeta_i + \sum_{i=1}^n \alpha_i [\|\Phi(x_i) - c\|^2 - r^2 - \zeta_i] - \sum_{i=1}^n \beta_i \zeta_i \quad (2.19)$$

其中 α 、 β 分别为两种约束条件的拉格朗日乘子。分别对 r 、 ζ 和 c 求偏导有

$$\begin{cases} \frac{\partial L}{\partial r} = 2r(1 - \sum_{i=1}^n \alpha_i) = 0 & \Rightarrow \sum_{i=1}^n \alpha_i = 1 \\ \frac{\partial L}{\partial \zeta_i} = \frac{1}{vn} - \alpha_i - \beta_i = 0 & \Rightarrow 0 \leq \alpha_i \leq \frac{1}{vn} \\ \frac{\partial L}{\partial c} = -2 \sum_{i=1}^n \alpha_i (\Phi(x_i) - c) & \Rightarrow c = \sum_{i=1}^n \alpha_i \Phi(x_i) \end{cases} \quad (2.20)$$

将式2.20中推导出的前两式作为约束条件，并将第一式、第三式回代入式2.19，则问题等价于一个二次优化问题^[33, 34]：

$$\begin{aligned} \min_{\alpha} & \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i K(x_i, x_i) \\ \text{s.t.} & \sum_{i=1}^n \alpha_i = 1, 0 \leq \alpha_i \leq \frac{1}{vn} \end{aligned} \quad (2.21)$$

其中训练数据仅以映射函数的乘 $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ 的形式出现，称 $K(x_i, x_j)$ 为核函数 (Kernel Function)。常用的核函数形式有多项式核函数、高斯径向基核函数、Sigmoid 核函数等。其中高斯径向基核函数的形式为

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (2.22)$$

高斯径向基核函数的参数 σ 可用于调节训练向量之间的距离。

图2.8展示了单类 SVM 对二维高斯分布随机数据的分类结果。(a) 是训练数据，用红色点表示；(b) 是单类 SVM 对测试数据的分类结果，其中红色点表示同类，蓝色点表示异类。训练数据为单位圆内随机分布的点，而单类 SVM 准确识别出了这一边界。可以看出单类 SVM 反映了数据对本类别的真实归属情况。

2.2.5 基于稀疏表示的异常检测

稀疏编码方法来源于神经科学，最早由康奈尔大学 Bruno 等人提出并用于自然图像的表达^[35]，后来稀疏编码被广泛应用于图像去噪^[36]，复原^[37]，运动目标跟

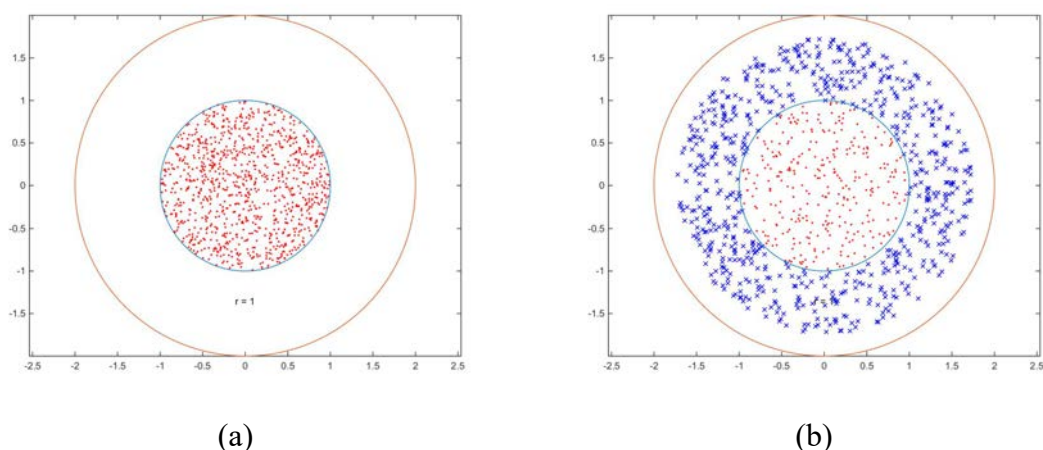


图 2.8 单类 SVM 对二维高斯分布随机数据的分类结果。(a) 是单类训练数据 (红色): 单位圆内服从二维均匀分布的随机变量 (x, y) ; (b) 是单类 SVM 对测试数据的分类结果 (红色为同类, 蓝色异类)。

踪^[38] 等各个方面。设信号向量 $x \in R^m$, 过完备字典矩阵 $D \in R^{m \times n} (m \ll n)$, 求解稀疏编码向量 $\alpha \in R^n$ 的问题可以表示为如下一个优化问题:

$$\min_{\alpha} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (2.23)$$

其中 λ 为常数参数。式 2.23 表示的问题也称为 LASSO (Least Absolute Shrinkage and Selection Operator)。

在异常检测中, 如果使用大量正常特征向量训练出字典 D , 那么正常特征通过此字典的稀疏编码的 1-范数应当较小。如果某个输入特征 x 在字典 D 的表示下编码 1-范数较大, 那么可以认为这是一个异常特征。通过设定阈值, 通过输入视频特征 x 在正常稀疏字典 D 下的稀疏编码的 1-范数 $\|\alpha\|_1$ 是否超过阈值来判断输入视频段是否有异常, 这就是稀疏编码模型用于异常检测的基本原理。

2.3 本章小结

本章针对视频帧分块异常检测的方法, 介绍了特征提取和异常检测模型的基础知识。对于每一个视频分块, 其特征往往是向量形式。常用的特征有光流直方图、3D 梯度、前景计数器和复合特征描述子。光流直方图直观反映运动速度的大小和方向分布; 3D 梯度结合运动信息和纹理信息; 前景计数器只关注动还是不动, 反映场景的运动密度, 可以用于检测异常滞留的前景目标; 复合特征描述子可以组合各个特征, 理论上具有最多的信息量, 但是直接放入模型效果不一定好, 因为各个维度的量纲和物理意义有较大差别, 因而需要对各个类别的属性做归一化或分别用不同的模型。

异常检测并非视频处理独有的问题，而是一个抽象的数学问题。异常检测与传统的分类问题有着显著区别，因为输入训练样本往往只有一类。异常检测的经典方法可以分为基于聚类的方法、基于统计的方法和基于分类的方法。具体地，本章介绍了基于多维高斯分布的异常检测方法、朴素的最大值模型和单类 SVM 模型。最后介绍了近年来提出的基于稀疏表示的异常检测方法，它是一种数学编码方法，利用编码值的范数大小来衡量特征的异常程度。

以上介绍的特征和模型是本文方法的重要来源和基础。下一章介绍的本文方法正是在他们的组合改进上得来。

第三章 监控视频中的异常事件检测方法

3.1 引言

图3.1展示了本文异常检测方法的主要流程。异常检测从步骤上看分为训练和检测两大步骤。训练可以分为特征提取和训练模型两步；检测可以分为特征提取和应用模型检测两步。其中特征提取是共同步骤，它分为像素特征提取和特征聚合两部分。而训练和检测则统一于所使用的模型。

本文选取光流直方图为主要特征，利用朴素的极大值模型、稀疏表示模型和单类 SVM 模型进行异常检测。针对视频中的光流计算的特点，提出了基于运动先验信息的视频光流计算方法。

3.2 基于运动先验信息的视频光流直方图特征提取

3.2.1 改进 ViBe 算法的运动检测

为了缩小光流计算的像素区域，同时不对计算资源有过多占用，需要一个快速鲁棒的运动检测算法。Barnich 等人提出的 ViBe 算法^[39] 在近年来的运动检测算法中具有公认的鲁棒效果，同时计算效率也较高。ViBe 算法存在的背景模型初始化仅采用第一帧的信息，这虽然简单易行，但是如果第一帧存在运动目标将会在后续检测中导致鬼影、拖尾和空洞现象。针对此问题，采用基于时域概率最大化

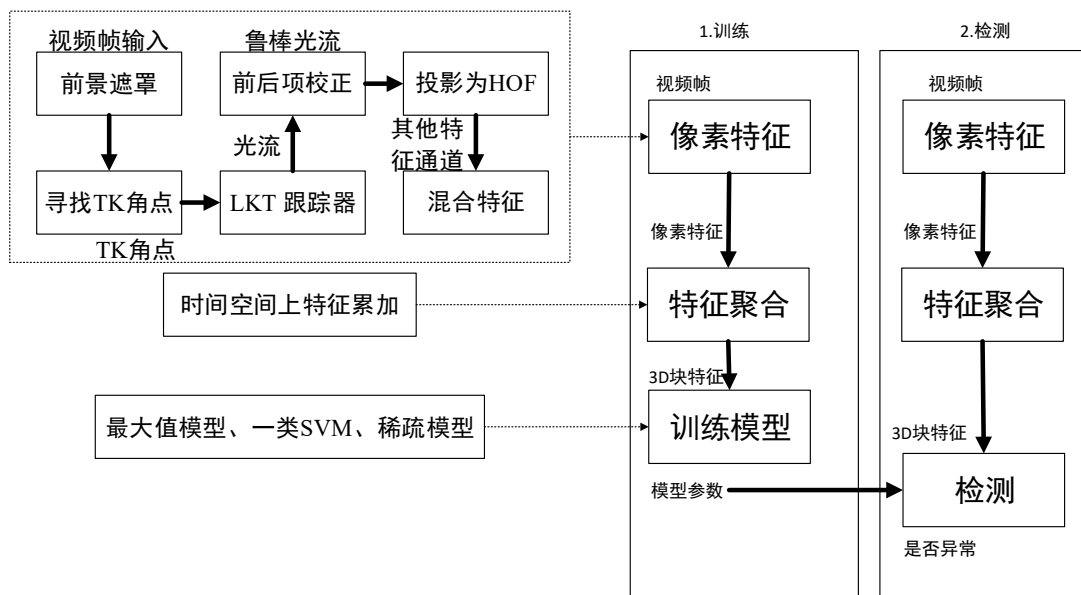


图 3.1 异常检测方法的训练和检测流程

(Maximum Temporal Probability, MTP) 的背景模型初始化方法来构造一个可靠的初始背景。

MTP 方法的主要思想是对图象中的每一个像素位置在时域上建立颜色直方图：将前 N 帧此像素位置邻域内的 RGB 像素通道分别建立直方图，取柱值最大的颜色作为此像素的颜色。当 N 取到 50-100 时，对多数视频可以得到干净的背景。

在 MTP 方法得到初始背景模型后，采用 ViBe 方法进行运动检测。设 $p_t(x)$ 为像素位置 x 的像素值， $S_R(p_t(x))$ 为以 x 为中心， R 为半径的球形区域内像素值的集合， $\{p_1, \dots, p_n\}$ 为 x 邻域内的背景样本点集合，若

$$S_R(p_t(x)) \cap \{p_1, \dots, p_n\} < \#_{min} \quad (3.1)$$

则认为 x 为前景点，否则认为 x 是背景点。其中 $\#_{min}$ 为阈值。

最后，对运动检测结果进行形态学滤波。本文采取的方法是删除面积小于一定阈值的块并且对前景图象做三次闭操作。图 3.2 展示了改进的 ViBe 方法在 UCSD Ped1 数据集上的运动检测结果。图 (a) 是 UCSD Ped1 数据集训练视频第 5 帧；图 (b) 是 MTP 模型建立初始背景模型演化成的背景；图 (c) 是 ViBe 方法得到的前景图象。可以看出虽然数据集一开始就有大量的运动目标，改进的 ViBe 方法基本能准确地区分出前景区域和背景区域，这为后续的光流计算打下了良好的基础。

3.2.2 像素特征：快速鲁棒视频光流提取

2.1.2 节已经介绍过光流计算的基本原理，然而还有一个很重要的问题是计算效率。传统的光流算法设计针对的是两幅图像，计算逐像素进行。一对 400×300 的图像在 2015 年的高配计算机上计算光流耗时不小于 10 秒。视频每秒有 25-30 帧图像，这样的时间开销不仅实际使用难以承受，即便是做实验也大大影响了实验周期和算法改进的效率。其实视频光流计算与图像光流计算有一个很大的不同，就是有运动先验信息。监控视频中运动的像素往往只占很小的比例，即前景所占比例是很小的。在前景区域计算光流结果才有意义，在背景区域计算光流一方面浪费了计算资源，另一方面计算结果不是接近零就是错误的。在视频前景区域计算光流，可以几十倍上百倍地提升光流的计算速度。在一些图像尺寸较小的数据集上计算速度更是可以达到实时。

图 3.3 展示了本文计算光流的流程。详细步骤如下：

1. 首先通过改进的 ViBe 算法计算出每一帧视频的前景区域。用前景区域作为遮罩进行后面的计算，一方面大大减少了需要计算光流的像素个数，而获取前景的运动检测算法本身耗时很少，故前景遮罩有加速作用；另一方面，背景中不计算光流也减少了光流出错的机会，提升了算法的鲁棒性。

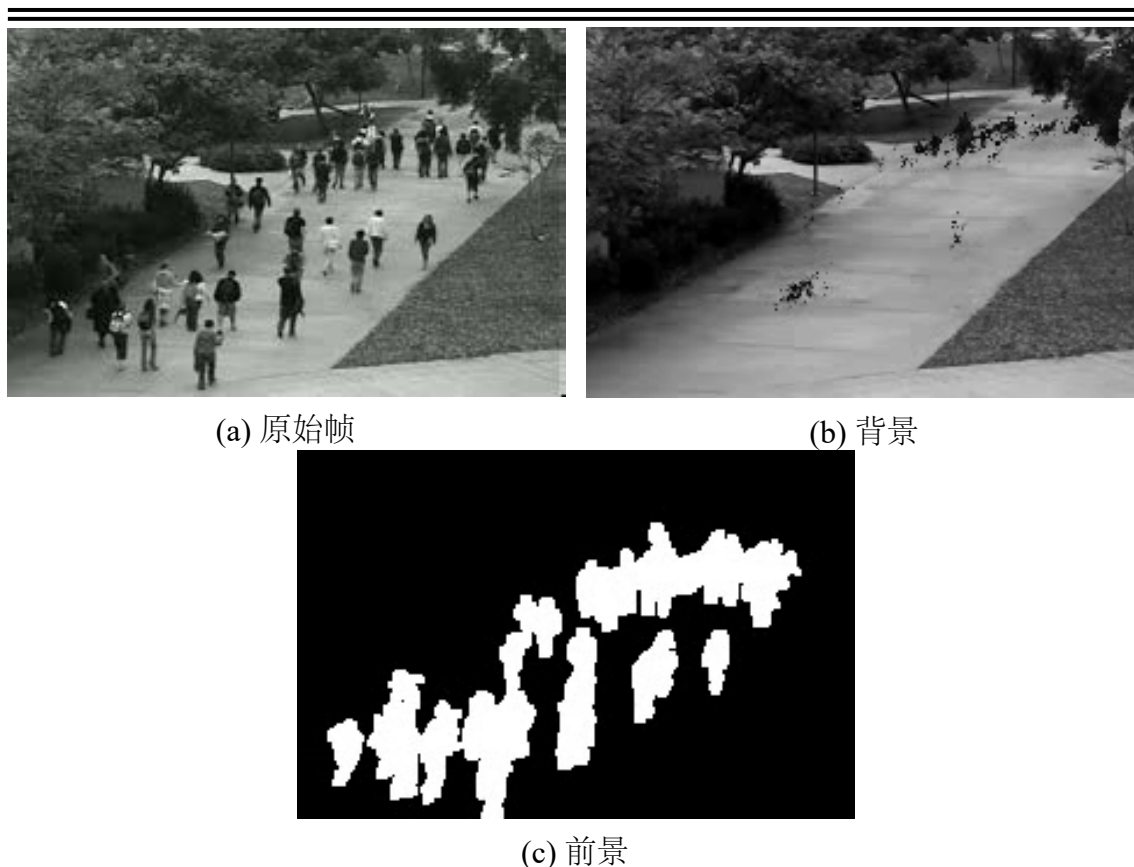


图 3.2 在 UCSD Ped1 数据集上的运动检测结果。图 (a) 是 UCSD Ped1 数据集训练视频第 5 帧；图 (b) 是 MTP 模型建立初始背景模型演化成的背景；图 (c) 是 ViBe 方法得到的前景图象。

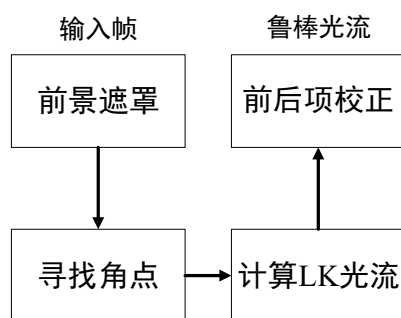


图 3.3 鲁棒光流计算流程

- 其次在前景区域的空域网格上计算 Shi-Tomasi 角点^[40]。2.1.2 节的讲过 LK 算法求解光流方程增加了一个约束假设：像素点 p 的邻域内的像素与像素点 p 具有相同的位移。为了确认像素点 p 的位移就需要进行图像块匹配。在纹理稀疏的图像区域块匹配的结果又多义性，因而由此求解出的光流也不可靠。而在纹理丰富的角点处计算块匹配可靠性高，求解出的光流可靠性也高。为了进一步加速，角点寻找也不必逐像素，二是可以设立空域上的网格，横竖方向每间隔 g 个像素取一个点用于计算。



(a) 城市公路的光流



(b) 雨夜场景的光流

图 3.4 光流计算结果示例

3. 接着角点上计算 LK 光流^[26]。在前景区域的空域网格上使用经典的 LK 算法求解光流方程，得到光流。
4. 最后通过 Kalal 博士论文中使用的前后项校正方法^[41] 过滤不可靠的光流，最终得到鲁棒稀疏光流。即使在前景的角点上计算光流，仍有一些情况可能导致光流计算错误。例如被遮挡的背景区域与前景运动物体具有相似纹理时，光流容易出现错误。此时可以将前后帧对调，重新计算光流，如果前后两个

光流的起点终点误差大于某一个阈值，则可认为此光流是不可靠的，故而应该过滤掉。

以上步骤相互配合，使得与单纯的 LK 方法相比，光流计算的速度和可靠性有了很大提升。图3.4展示了光流计算效果。可以看出，对于图 (a) 中所示的城市公路场景，本文的光流计算结果几乎没有错误；图 (b) 的雨夜公路场景，光流结果也基本准确，说明了本文方法的有效性。另一方面，图中也展示了光流计算的难点。一是公交车顶部有大片的白色区域没有计算出光流，这是典型的无纹理区域光流难于计算。二是左侧的林荫道中靠近图像边缘的两辆摩托车没有计算出光流。这是摩托车前一帧被树木遮挡或处于遮挡边缘，光流也难于计算。三是远处的车辆，由于运动距离过小，也难于计算光流，这是尺度问题。无纹理、遮挡和尺度问题正是光流计算的三大难点。如果场景光照条件不佳，如夜晚场景，光照不一致、反光和噪声等问题也会严重影响光流计算的精度。虽然光流算法的研究历史已经有五十余年，但算法测评多是针对某些数据集的，想获得在实际监控场景中广泛适用的光流并不容易。本文提出的步骤在白天、夜晚、下雨等多种实际监控场景中测试，均能获得稳定正确的光流。

3.2.3 聚合特征：区域光流直方图

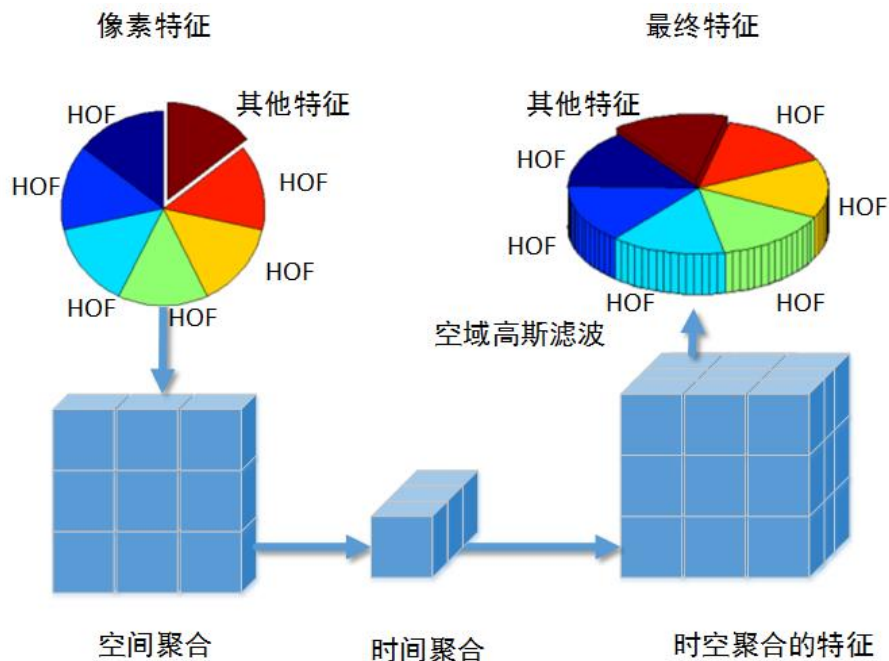


图 3.5 点像素光流聚合为光流直方图的过程

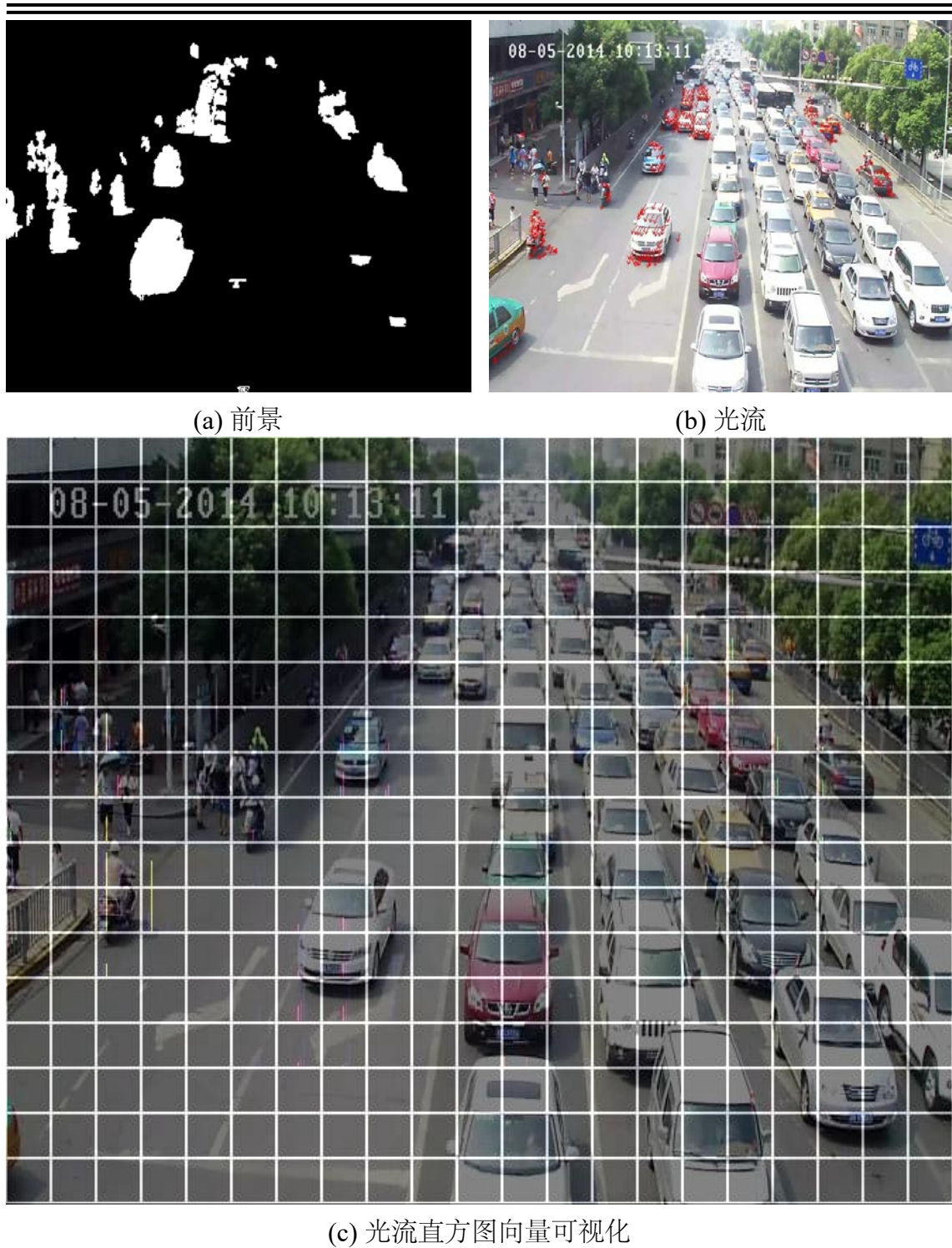


图 3.6 光流直方图特征。图 (a) 为某路口场景的前景图象；图 (b) 为光流计算结果；图 (c) 为光流直方图向量可视化的结果，不同的颜色表示不同的直方图通道（交叠的网格未画出）。可以看出，逆行的摩托车直方图柱值是黄色，与正常行驶的汽车直方图的红紫色明显不同，说明光流直方图特征有效表达出了此运动方向异常。

表 3.1 重要的特征聚合参数

聚合参数	含义
b_H	分块高度
b_W	分块宽度
b_T	分块时域长度
δ_H	分块纵向间隔, 一般取为 $0.5b_H$
δ_W	分块横向间隔, 一般取为 $0.5b_W$
m	光流直方图柱数

上一小节获得光流是像素特征, 在用于模型之前仍然需要聚合。绪论中提到本文采用的是视频帧分块处理的方法, 那么接下来的任务是在提取分块中聚合特征。图3.5展示了点像素的光流特征如何聚合成直方图特征, 详细步骤如下:

1. 首先根据上一节的方法计算鲁棒稀疏光流。
2. 其次在时空域上聚合特征。聚合步骤根据2.1.2节的式2.11。这里计算直方图的不同在于选择的分块是一个 3D 块, 即光流累加时包括时域。
3. 最后将在直方图在空域上进行一次高斯滤波。使得上下左右块的直方图平滑本块直方图, 实验表明此步骤可以进一步平滑噪声, 增强特征的鲁棒性。

图3.5的特征通道除了增加了相邻块的高斯滤波平滑噪声, 还预留了“其他特征”, 这里的其他特征可以根据模型需要, 设置为前景计数器、3D 梯度等2.1节介绍的特征, 从而构成一个复合描述子。

特征聚合时有一些重要的参数会影响到后续异常检测效果, 表3.1总结了这 6 个参数, 前 5 个参数控制交叠网格的大小和交叠程度; 最后 1 个参数控制对光流角度分组的粒度。在后续实验中会对这些参数进行调优。

图3.6展示了对某路口进行特征提取的例子。图 (a) 为某路口场景的前景图象; 图 (b) 为光流计算结果; 图 (c) 为光流直方图向量可视化的结果, 不同的颜色表示不同的直方图通道 (交叠的网格未画出)。可以看出, 逆行的摩托车直方图柱值是黄色, 与正常行驶的汽车直方图的红紫色明显不同, 说明光流直方图特征有效表达出了此运动方向异常

3.3 异常检测模型

异常检测问题是非常依赖具体应用的问题, 并不存在通用的有效方法。故而对于监控视频的异常检测问题, 事先也无法确定哪些特征、哪些异常检测模型是有效的。根据现有文献的记载, 按照由简单到复杂的原则, 本文选择了朴素的最大值模型、单类 SVM 和稀疏表示三种模型作为候选方法, 分析它们在标准数据集和实际监控上的效果, 以求寻找到最有效的监控视频异常检测方法。

3.3.1 基于朴素最大值模型的异常检测

针对像素值特征的朴素的 \max 模型在2.2.4.1节已经做了介绍，对于分块特征的最大值模型类似。对于每一个分块，设 $H(x, t) \in R^m$ 为第 x 分块在第 t 时刻（帧）的光流直方图特征。对于输入的训练视频，分块的 B_{business} 为

$$B_i(x) = \max_t H_i(x, t), i = 1, 2, \dots, n. \quad (3.2)$$

其中 H_i 、 B_i 分别表示 H 、 B 的第 i 维。根据最大值模型的思想和本文的实验探索，对于输入的检测视频，分块 x 在 t 时刻（帧）是否异常由下式判别：

$$a(x, t) = \begin{cases} 1, & \exists i \in \{1, \dots, m\}, H_i(x, t) > B_i(x) + \delta \wedge (H_i(x, t) - B_i(x))/B_i(x) > \theta \\ 0, & \text{otherwise} \end{cases} \quad (3.3)$$

其中 δ 为加法阈值； θ 为比例阈值。在直方图柱值较小时，加法阈值起作用；在直方图柱值较大时，由乘法阈值起作用。实验表明，比例阈值 θ 取 0.2 可以取得较好的检测效果，不必改变。加法阈值 δ 的选取则与分块特征参数密切相关。本文设定

$$\delta = \eta(b_H \cdot b_W \cdot b_T) \quad (3.4)$$

其中 η 为非负常数，表示 δ 与分块像素总数的比值。通过设定不同的 η 值，最大值模型将表现出不同的检测性能。

最大值模型使用的光流直方图特征并不需要归一化，其优点在于此方法很容易用于在线检测。

3.3.2 基于单类 SVM 的异常检测

单类 SVM 在2.2.4.2节做了介绍，它是一种专门用于异常检测的支持向量机。SVM 往往要配合核函数使用。在给定适当的参数时，多项式、高斯和 Sigmoid 核函数具有相似的性能^[42]，本文选用高斯核函数。这样，单类 SVM 的重要参数由两个：式2.18中的优化尺度参数 ν 和式2.22中的高斯核函数标准差 σ 。在后续的实验中，这两个参数也需要调优。

在求解式2.21得到拉格朗日乘子 α 之后，输入样本 H 的单类 SVM 决策值通过下式确定^[33]：

$$f(H) = r^2 - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j) + 2 \sum_{i=1}^n \alpha_i K(x_i, H) - K(H, H) \quad (3.5)$$

设训练样本的监督值为 +1，则 $f(x)$ 越接近 +1 则 x 越正常， $f(x)$ 越接近 -1 则 x 越异常。因此分块 x 在 t 时刻（帧）是否异常由下式判别：

$$a(x, t) = \begin{cases} 1, & f(H) > \theta \\ 0, & otherwise \end{cases} \quad (3.6)$$

其中 θ 是阈值。

在数学上，训练单类 SVM 分类器就是寻找拉格朗日乘子 α 。实际运算中，保留所有训练样本来计算决策值是没有必要的。因而训练算法还需要挑选出决策边界上的一些样本作为支持向量保存在模型中。

3.3.3 基于稀疏表示的异常检测

基于稀疏表示的异常检测基本原理在 2.2.5 节中已经介绍。在异常检测中，用大量正常特征向量训练出字典 D ，按式 2.23 求解出带检测样本 H 的稀疏编码 α ，稀疏模型判断样本异常的标准是：

$$a(x, t) = \begin{cases} 1, & |\alpha|_1 > \theta \\ 0, & otherwise \end{cases} \quad (3.7)$$

其中 θ 是阈值。

过完备字典的训练本身是稀疏表示领域一个重要的研究点，经典的方法有 K-SVD 方法^[43]、基于非负矩阵因子化的 SPAMS 方法^[44, 45] 等。Mairal 等人开发了 spams-matlab^[46] 工具包，实现了基于非负矩阵因子化的过完备字典训练和稀疏编码求解 (LASSO) 算法。本文将直接采用 SPAMS 提供的算法实现稀疏模型。

3.4 监控视频异常检测系统的实现

根据所提方法，本文实现了一个监控视频异常检测系统，其流程图如图 3.7 所示。该系统有两个开始，分别是训练和检测，共用特征提取模块和模型评价模块。准备好训练视频和测试视频后，该系统能全自动运行，完成特征提取、模型训练和异常检测工作。其中光流直方图特征提取和最大值模型作为一个子模块用 C++ 实现，可以独立运行，并可以导出所提取的向量特征。稀疏模型的字典训练方法较为复杂，本文的实现借助了 Matlab 接口的 SPAMS 库^[46]；单类 SVM 模型的实现借助于 libSVM 库^[47]，故这两个模型的训练和检测均用 Matlab 实现。而它们使用的特征数据来自 C++ 程序的导出。模型的评价部分也用 Matlab 实现。

Matlab 模块还可以将模型检测的结果合成二值视频与真值 (Ground Truth) 视频比较，绘制 ROC 曲线和计算帧水平 AUC 和像素水平 AUC，用于评估模型的性能。评估模型性能需要设置不同的阈值多次运行异常检测模块，这一步在图 3.7 中没有显式画出，在此说明。

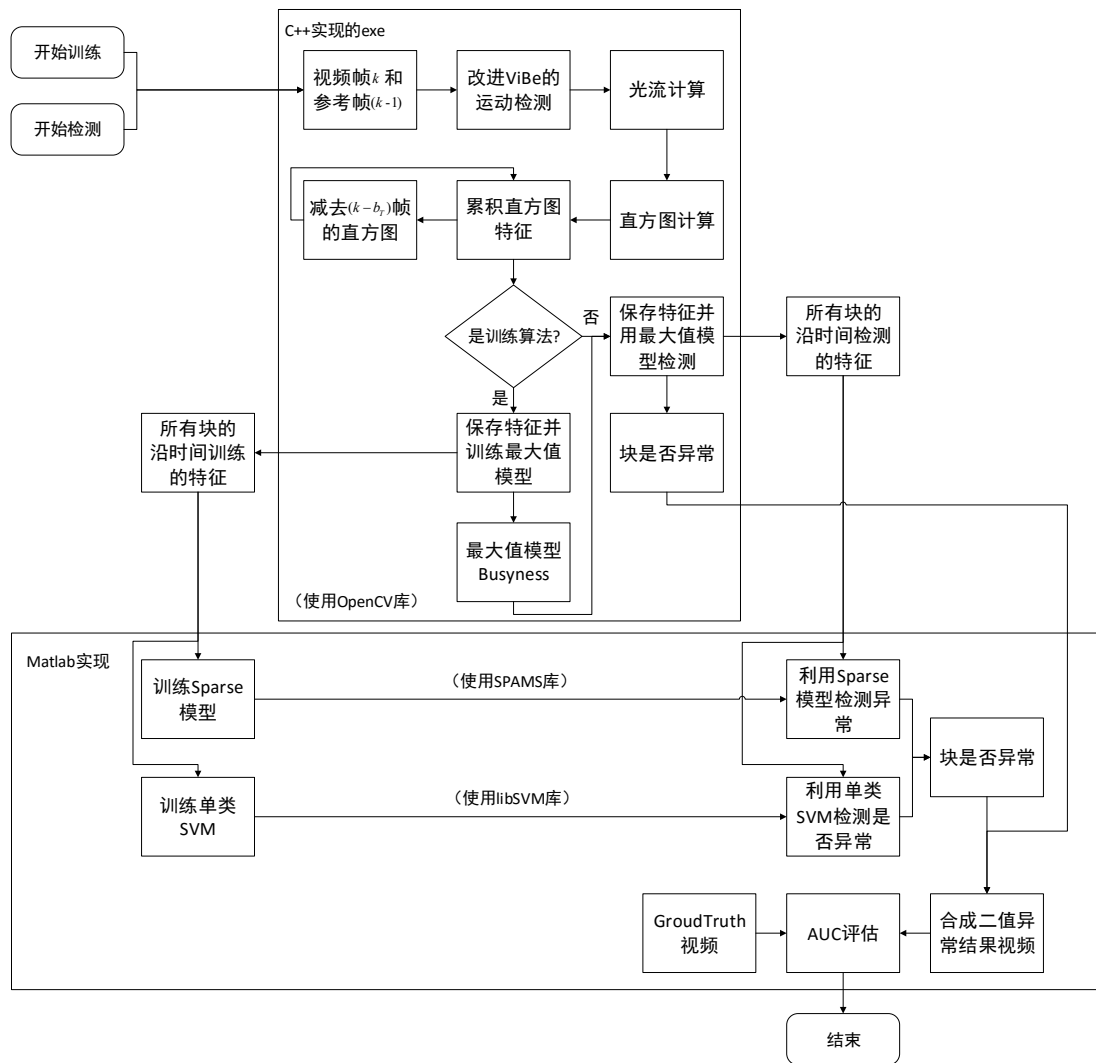


图 3.7 视频监控异常检测系统流程图

异常检测算法需要多个步骤串行执行，为了能调优每一个步骤，系统出了会输出检测结果，还可以保存下前景视频、光流效果视频、提取的向量特征、模型参数、训练和检测时间等中间结果和参数，便于定位问题和调试。

3.5 本章小结

本章说明了本文的异常检测方法。首先将视频划分成 3D 块，将检测问题转化为判定性问题。其次，对每一个 3D 块，用基于运动先验信息的视频光流直方图提取方法提取向量特征；从训练视频中提取到的特征集合将用于模型训练：这里的模型包括最大值模型、单类 SVM 和稀疏模型。从检测视频中提取到的特征送入模型检测，根据模型的打分和判断阈值判断此 3D 块是否为异常块。

本文方法对光流算法在视频上的速度和效果进行了优化，并大胆尝试使用朴素的最大值模型，同时采用了稀疏模型和单类 SVM 模型进行检测，力求在具体应用中找到最佳的特征模型组合。最后实现的监控视频异常检测系统能自动使用三种模型进行异常检测；对含有真值的数据集还可以进行性能评价。这为下一章的实验分析奠定了基础。

第四章 实验与分析

4.1 引言

本章通过实验检验光流直方图特征和最大值模型、稀疏表示模型以及单类SVM模型在监控视频异常检测中的效果。首先介绍实验所使用的数据集，为了便于与现有方法比较，采用一个标准数据集；为了检验算法在实际监控视频上的效果，采集了若干实际监控作为实际数据集。其次，明确评判异常检测结果的标准。主要从检测性能、时间开销和实用性三个方面进行评价。接着，分析实验中发现的光流直方图特征的特点。然后分析三种模型在两个数据集上的结果。最后得出结论。

4.2 数据集

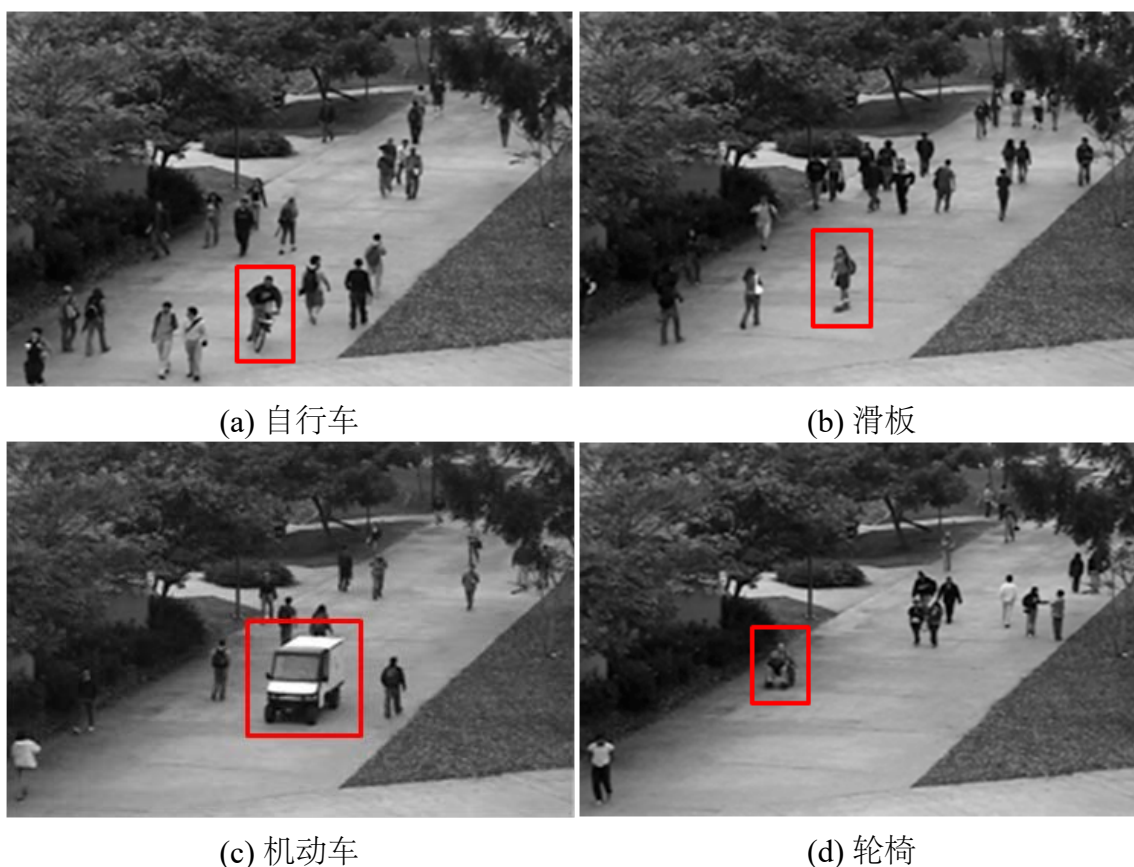


图 4.1 UCSD ped1 数据集部分异常示例。在此数据集中，所有非行人运动被视为异常。图 (a) 为快速运动的自行车；图 (b) 为滑板运动；图 (c) 为机动车运动；图 (d) 为轮椅运动。



图 4.2 自采集数据集部分异常示例。在此数据集中，明显违反交通规则的行为被视为异常。图 (a) 为强穿马路的行人；图 (b) 为雨夜逆行的卡车；图 (c) 为路口逆行的摩托车；图 (d) 为路口强穿马路的行人。

本文使用的数据集包括近年来异常检测常用的 UCSD ped1 数据集^[31]和调取的若干道路监控视频。UCSD ped1 数据集使用固定摄像机俯瞰行人拍摄的一组视频。其中训练视频 36 个，每个 200 帧，包含固定场景的行人行走情形；测试视频 34 个，均有帧水平的真实结果，其中 10 个包含像素水平真实结果。训练视频和测试视频每个 200 帧，帧大小为 238×158 ，包含自行车、滑板、机动车等非行人运动。数据集将非行人目标认为是异常目标。图 4.1 展示了四种异常场景。图 (a) 为快速运动的自行车；图 (b) 为滑板运动；图 (c) 为机动车运动；图 (d) 为轮椅运动。它们均不是行人运动，因此被视为异常。

虽然此数据集为近年来最常用的异常检测数据集，在其他研究者没有公布代码实现的情况下，使用此数据集可以比较各个算法的性能。但此数据集也存在一些问题：

1. UCSD Ped1 的异常判定标准其实与本文有出入。数据集的判断标准是非行人，是纹理异常；本文的判断标准是运动异常。比如，相对于行人，一辆自行车在快速的运动时可以算为异常。但是自行车在以近似行人的速度运动时运动模式与行人并没有什么不同，从运动角度看真值应当为不异常。此数据集的判定标准会使得基于纹理特征的算法在此数据集上效果更好。举个极端的例子，用行人检测器检测非行人可以在本数据集上得到很好的结果，然而这种方法并不能推广到其他场景。
2. 数据集的图像近乎灰度图像。缺少色彩信息会使得特征点定位的难度加大。
3. 数据集的帧大小较小。数据集视频分辨率仅有 238×158 ，这使得算法对细节的处理能力难以被考察。另外由于帧过小对算法的速度测试也不充分。

为了克服公用数据集存在的问题，并测试算法在更多实际场景监控视频中的效果，本文采集了四段彩色道路监控视频，帧分辨率为 960×540 或 720×576 ，每一段训练视频时长为 5 到 10 分钟，保证有足够长度的训练视频充分反映场景的运动模式图。4.2 为自采集数据集部分异常示例。在此数据集中，明显违反交通规则的行为被视为异常。图 (a) 为强穿马路的行人；图 (b) 为雨夜逆行的卡车；图 (c) 为路口逆行的摩托车；图 (d) 为路口强穿马路的行人。

UCSD Ped1 数据集上的结果可以与现有方法进行比较；而采集的数据集则用于测试算法在实际应用中的效果。

4.3 评价准则

对于异常检测结果的评价，可以从三个方面来考察：性能、算法复杂度和实用性。性能是最主要的评价因素，在标准数据集上容易评判。但是实际监控中没有标准答案，有些事件不同的人也会做出不同的判断，因此性能的评判结果其实有一定的波动区间。算法时间复杂度是另一个重要的评价因素。时间复杂度高，运行时间长的算法在处理视频这样的数据时，其价值会打折扣。编程复杂度是最后需要考虑的因素。编程复杂度高的算法不易于部署于产品中，将其实用需要更多的开发投入。

对于方法比较，还有一个现实的困难是近年来的异常检测论文均没有公布代码。由于异常检测的串行步骤多，实现难度大，其他人实现的结果未必达到作者论文中的效果，这使得比较的数据只能取自于作者的论文。这既限制了新数据集的使用，也使得比较的客观性存疑。因此在比较各个方法的结果时，应当关注指标是否具有显著差别。

4.3.1 异常检测性能评价

异常检测可以看为一个二分类问题，因此性能评价可以直接采用二分类问题的评价体系。接受者操作特性 (Receiver Operating Characteristic, ROC) 曲线是对分类器性能客观的评价方法，它采用分类器的假正例率 (False Positive Rate, FPR) 和真正例率 (True Positive Rate, TPR) 作为横纵坐标轴，将不同阈值下的分类器 (FPR, TPR) 连成曲线，曲线下面积 (Area Under Curve, AUC) 可以作为评价分类器的指标，其值越大则分类器性能越好。

对于异常检测，性能评价可以分为两个层次。一是帧水平 (Frame-level)。如果某一帧有异常而算法检测结果也有异常，则认为检测在帧水平正确。二是像素水平 (Pixel-level)。这将进一步考察算法对异常的定位是否准确。按照近年来论文采用的 [5] 文的标准：如果该帧异常像素有 40% 被检测到，则认为算法检测结果在像素水平正确。可以看出，像素水平的检测标准是相当宽松的，而且也有漏洞。比如，直接将帧检测结果认为异常的帧全部像素置为异常像素。那么按这个标准像素水平的性能不会损失。之所以指定这样宽松的标准也是因为如果采用传统的 Jaccard 距离衡量定位准确度现有算法的 AUC 会很难看。例如 [5] 文采用的动态纹理特征加显著性检测的方法会将异常运动目标轨迹上相当大的一部分像素作为异常像素。按照 Jaccard 距离 (0.5 阈值) 这样的定位结果不能算正确。

值得注意的是对于帧水平的异常检测随机猜测 (Random Guess) 的 AUC 应为 0.5，而对于像素水平的异常检测随机猜测的 AUC 为 0。因此检测结果在这两种尺度下 AUC 的差值小于 0.5 才能说明定位比随机猜测有意义。因此本文还引入一个评价指标：帧水平像素水平 AUC 差值。这个值接近零，则说明帧水平检测结果与像素水平定位结果相当，说明定位效果好；如果此值大于 0.5 则说明异常检测的定位结果没有意义。

为了方便表述，记帧水平 AUC 为 AUC_f ，像素水平 AUC 为 AUC_p ，帧水平像素水平 AUC 差值为 ΔAUC 。

为了防止方法对帧水平或像素水平 AUC 有偏倚，并获得一个单一值的综合评价指标，引入 AUC_f 和 AUC_p 的调和平均数作为方法综合评价：

$$AUC_h = \frac{2AUC_f \times AUC_p}{AUC_f + AUC_p} \quad (4.1)$$

4.3.2 时间复杂度评价

异常检测算法往往是很多步骤的串行，而且输入为视频数据，分析算法复杂度的数学表达式意义不大。总体运行时间和各个步骤的运行时间可以作为时间复杂度的评价指标。评价时间复杂度时需要注意的问题是部分算法对原始视频帧进

行了降采样。例如 [6] 文声称其方法可以在 UCSD Ped1 数据集上达到 150FPS 的处理速度。但这是因为其方法在处理前将视频帧缩放到 20×20 、 30×40 、 120×160 三个尺度上做金字塔。这样做确实有加速作用，但是缺点也显而易见。720p 的实际监控缩小到 20×20 尺度上行人、车辆等重要目标都变成了几个像素甚至消失，其实是无法进行特征提取和异常检测的。所以所谓的 150FPS 有一定的水分。因此在时间复杂度上，本文将不与 [6] 文进行比较。

另外由于无法获取代码，算法实现的软硬件环境也不尽相同。对于不同平台的实现，应关注运行时间是否有显著不同。

4.4 监控视频分块光流直方图的特征分析

异常检测采用特征和模型是否合适非常依赖于具体应用。本节通过数据降维、可视化和统计方法分析监控视频分块光流直方图本身的特性，力求揭示这一特征本身的特点，为选取合适的模型提供依据。

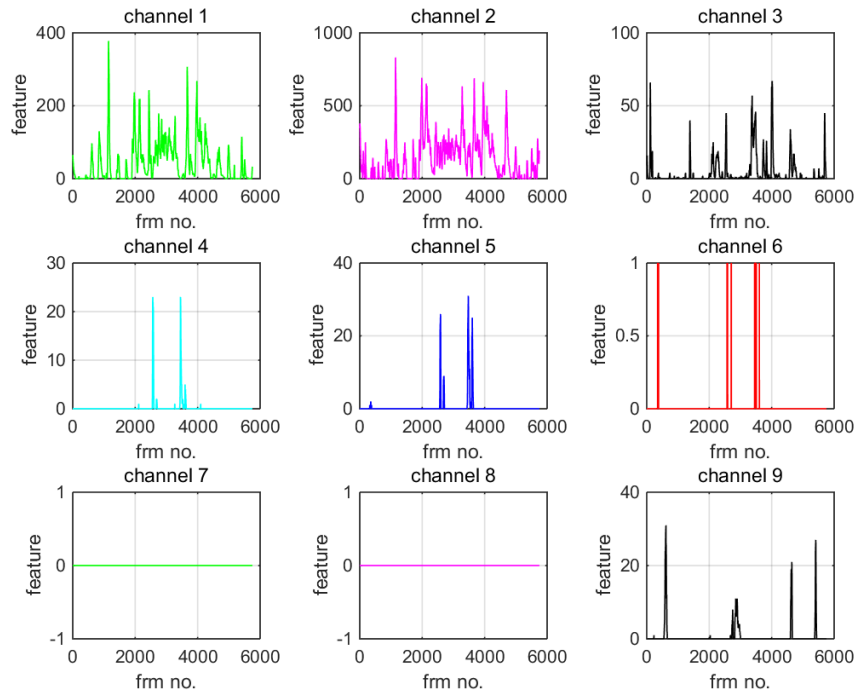
4.4.1 参数配置

特征提取涉及的参数主要有视频分块参数，光流参数和直方图参数。对于分块，在空间域上，文本将视频的每一帧划分为 32×32 的交叠块，相邻的块横竖间隔均为 16 个像素；在时间域上，每 5 帧构成一个块。对于光流，计算窗口大小为 15×15 ，金字塔层数为 8。对于直方图，方向柱为 9 个，即每个柱将吸收正负 40 度的光流在方向柱上的投影。

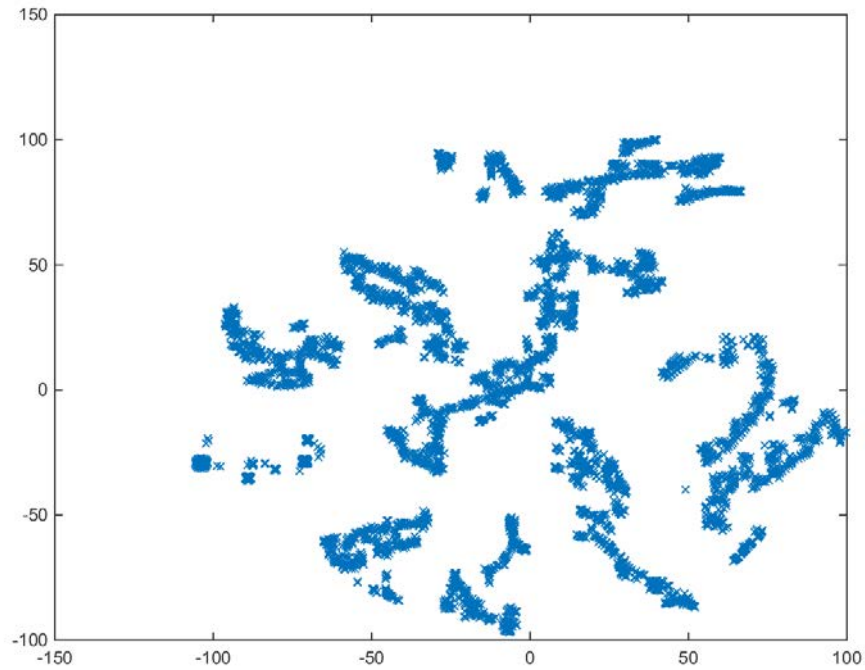
4.4.2 特征分析

为了把握光流直方图特征的直观特点，首先对自采集的过马路数据集的训练视频的每一个块的特征进行了可视化：分别绘制出光流直方图九个通道值随时间的变化曲线。为了进一步分析特征的特点，采用了近年来公认效果较好的 t-分布随机近邻嵌入法 (t-distributed Stochastic Neighbor Embedding, t-SNE)[48] 对特征做降维和可视化。对所有块特征分析后的结果大致可以分为两类。下面取具有代表性的第 139 块和第 173 块代表这两类做具体分析。

图 4.3 展示了第 139 块的特征。(a) 展示了光流直方图九个通道的值在 6000 帧内的变化情况。(b) 展示了用 t-SNE 法降维投影到二维平面特征的分布情况。可以看出，特征通道在时域上的分布呈散乱分布的尖峰形态。有效的特征分布在 1、2、3、4、5 和 9 通道上，并以 1、2 通道为主。这表明在时域上很难找到一个有效的回归模型来对光流直方图建模。但是朴素的最大值模型可能是有效的，因为尖峰的最大值只要不超过相应通道的历史最大值加一个阈值，就有理由认为它不是一

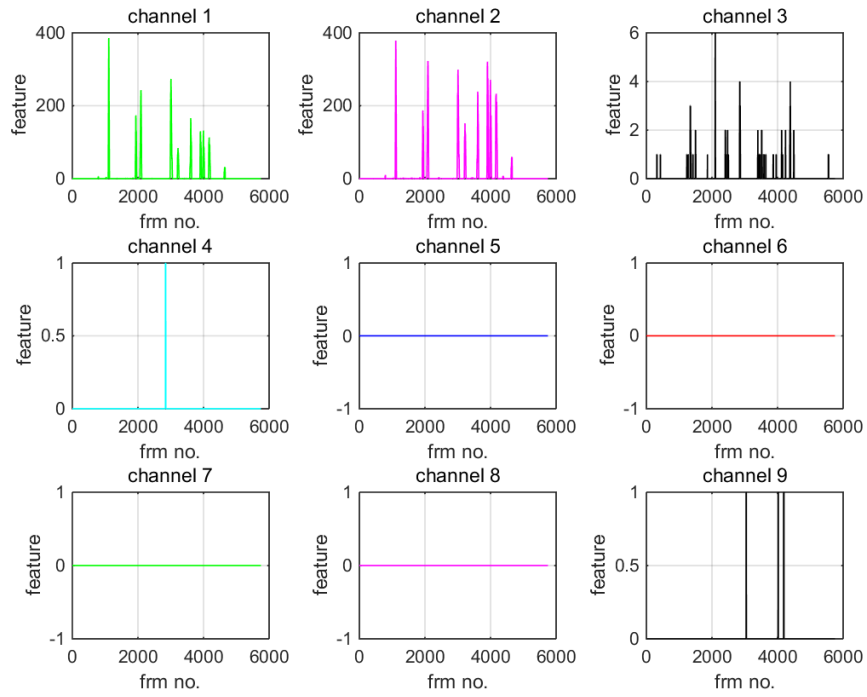


(a) 第 139 块特征

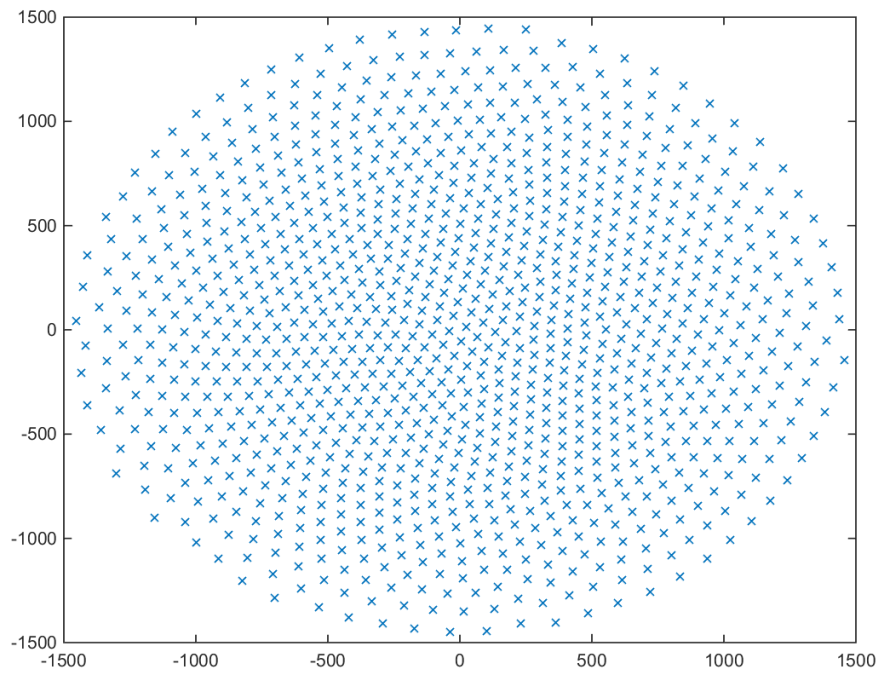


(b) 第 139 块特征 t-SNE 降维

图 4.3 第 139 块特征与 t-SNE 降维投影



(a) 第 173 块特征



(b) 第 173 块特征 t-SNE 降维

图 4.4 第 173 块特征与 t-SNE 降维投影

个异常事件。当然也可能存在一些异常事件，每个通道值并不超过历史最大值，但其通道分布是异常的。这样的事件最大值模型无法检测到。

采用 t-SNE 降维后的特征呈现不规则聚簇分布，伴随有个别离群点。t-SNE 法能保持特征之间的距离和聚簇结构。(b) 图展示的特征分布是多个松散的聚簇。这表明多中心的聚类方法可能有效。由于聚簇形状并不规则，聚类当优先选择基于密度的方法等能发现任意形状聚簇的方法。实际监控视频不能保证没有异常事件。个别离群点对应着这些小概率事件。这表明在聚类时，对象个数过少的聚簇应当被舍弃。

图4.4展示了第 173 块的特征。它的特征在时域上集中在通道 1、2。t-SNE 降维后的特征在平面上分布在一个松散的球状区域内。这表明此特征没有明显的聚簇结构，聚类方法是不适合的。但是它具有明显的边界。这说明朴素的最大值模型是有效的，因为各个通道的最大值向量就是高维空间中一个边界。比朴素最大值模型更复杂的边界寻找方法是单类 SVM 模型。单类 SVM 模型可用于系统地寻找特征边界。

图4.3和图4.4展示的共同特点是直方图特征中分别有 3 和 7 个通道值始终为 0 或者接近 0。这种零通道的存在有可能造成高等分类器的性能退化。一种合理的尝试是将零通道分离出来采用朴素的最大值模型，其他有意义的通道采用单类 SVM、稀疏表示等复杂模型建模。

4.5 参数分析

本文的方法由特征提取和模型穿行构成，含有许多参数。参数值不同会对方法在数据集上的性能有较大影响。为了选取到使异常检测方法达到较好性能的参数，采用帧水平和像素水平 AUC 作为评价指标，进行参数寻优。选取 UCSD Ped1 具有像素真值的 10 个测试视频作为数据集。

4.5.1 特征参数寻优

首先，对特征提取的参数寻优。特征提取特征提取的主要参数如表3.1所示。它们分别是分块大小 $b_H \cdot b_W \cdot b_T$ 、分块间隔 $\delta_H \cdot \delta_W$ 和光流直方图柱数 m 。为了减少变量，实验中取敏感性不强的分块的时间长度取为 $b_T = 5$ 帧，不变化。为了减小方法的影响，选择不需要额外参数的朴素最大值模型作为检测方法，所得的 AUC 结果如表4.1所示。

可以看出，空间块大小不宜过小，最佳的块大小为 32×32 (间隔 16×16)；直方图柱数 16 比 9 好。故特征提取参数选择第 5 组较好。

表 4.1 特征提取的参数寻优

组号	块大小 (b_H, b_W)	块间隔 (δ_H, δ_W)	直方图柱数 m	AUC_f	AUC_p	AUC_h
1	8,8	4,4	16	0.707	0.347	0.466
2			9	0.701	0.317	0.437
3	16,16	8,8	16	0.714	0.490	0.581
4			9	0.714	0.490	0.581
5	32,32	16,16	16	0.749	0.578	0.652
6			9	0.670	0.555	0.607

表 4.2 单类 SVM 的参数寻优

ν, σ	AUC_f	AUC_p	AUC_h
0.01,0.01	0.734	0.458	0.564
0.01,0.02	0.759	0.464	0.576
0.02,0.01	0.767	0.478	0.589
0.02,0.02	0.774	0.477	0.590
0.023,0.023	0.773	0.479	0.591
0.03,0.03	0.762	0.472	0.583

4.5.2 单类 SVM 参数寻优

在第 5 组特征提取参数下，还需考虑模型参数。对于单类 SVM 模型，需对优化尺度参数 ν 和高斯核函数标准差 σ 寻优，结果如表 4.2 所示。

表 4.2 表明 $\nu = 0.02, \sigma = 0.02$ 是单类 SVM 的较优参数。本文后续实验将采用这一组参数值。

4.6 异常结果与分析

4.6.1 UCSD Ped1 数据集

朴素的最大值模型仅有一个加法阈值参数可变，通过改变这个阈值参数，可以得到不同的 (FPR, TPR) 值，从而获得最大值模型的 ROC 曲线和 AUC 值。表 4.3 展示了 UCSD Ped1 数据集上三种异常检测模型在光流柱数 $m = 9, 16$ 下的性能。可以看出，16 光流柱的结果优于 9 光流柱的结果。而在 $m = 16$ 时单类 SVM 具有最好的帧水平性能，而最大值模型具有最好的像素水平性能。从 ΔAUC 来看，最大值模型具有最小的 ΔAUC ，这表明最大值模型的像素水平异常检测性能与帧水平异常检测性能最接近，意味着帧水平“乱猜”的成分最少。

图 4.5 展示了朴素的最大值模型与近年来 CVPR、ICCV 和 TPAMI 上发表的视频异常检测论文方法的结果在 UCSD Ped1 数据集上的性能对比。其中图 (a) 是帧

表 4.3 三种模型异常检测性能对比

分块大小 (b_H, b_W, b_T)	分块间隔 (δ_H, δ_W)	光流柱 m	方法	AUC_f	AUC_p	ΔAUC	AUC_h
32,32,5	16,16	9	最大值	0.700	0.555	0.145	0.619
			单类 SVM	0.750	0.465	0.285	0.574
			稀疏模型	0.696	0.413	0.283	0.518
		16	最大值	0.749	0.578	0.171	0.652
			单类 SVM	0.774	0.477	0.297	0.590
			稀疏模型	0.710	0.440	0.270	0.543

表 4.4 不同方法的性能比较。前六列为近年来 CVPR、ICCV 和 TPAMI 上发表的异常检测论文方法的结果，最后三列为本文三种模型的结果。

	SF	MDT	Sparse	Adam	Antic ^[50]	Sparse Combination	最大 值	单类 SVM	稀疏 模型
AUC_f	67.5%	81.8%	86%	65.0%	91.0%	91.8%	74.9%	77.4%	71.0%
AUC_p	20.5%	44.1%	13.3%	46.1%	76.0%	63.8%	57.8%	47.7%	44.0%
ΔAUC	47.0%	37.7%	72.7%	18.9%	15.0%	28.0%	17.1%	29.7%	27.0%
AUC_h	31.4%	57.3%	23.0%	53.9%	82.8%	75.3%	65.2%	59.0%	54.3%

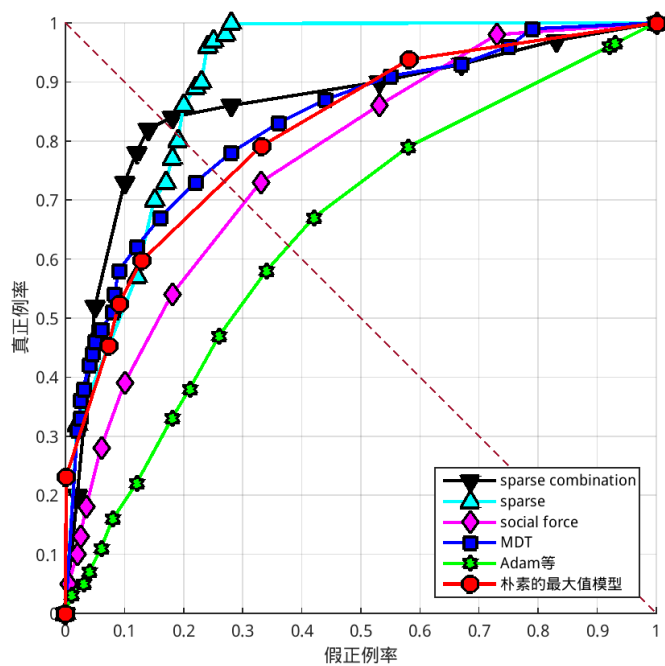
水平 ROC 曲线，图 (b) 是像素水平 ROC 曲线。图中除了最大值模型，还有 Adam 等的方法^[16]、MDT^[5]、Social Force^[49]、Sparse^[2] 和 Sparse Combination^[6] 五种方法的结果。由于没有获取代码，这些方法的结果数据摘自文献。

为了更准确地比较各个方法，表 4.4 列出了各个方法的 AUC 等评价值。表中新列出的 Antic 等^[50] 提出的学习全局运动目标的方法是目前在 UCSD Ped1 数据集上有最好的性能指标的方法。这种方法的原理类似于用行人检测器去检测非行人，要求场景的运动目标类别较少，适用范围有限。

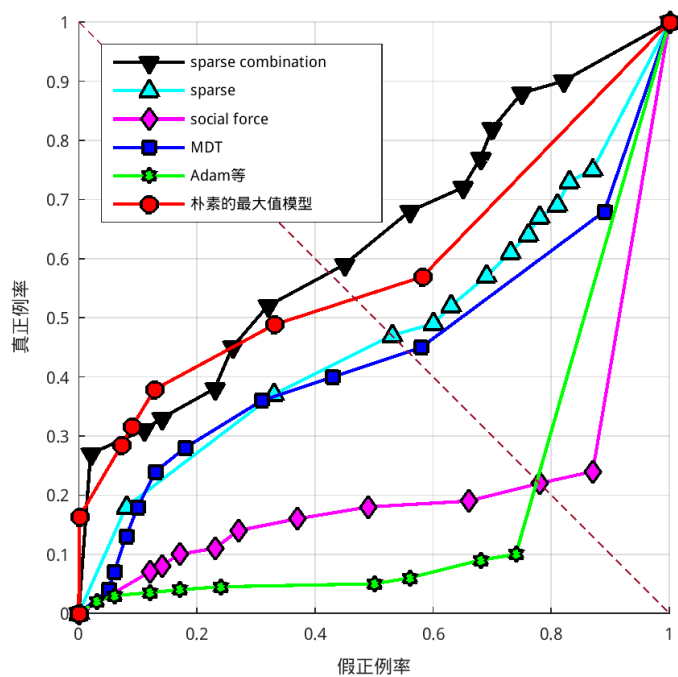
从像素水平 AUC 来看，最大值模型次于 Antic 的方法和 Sparse Combination 方法，排在第三位。从 ΔAUC 来看，最大值模型排在第二位，仅次于 Antic 的方法。从帧水平和像素水平 AUC 的调和平均数 AUC_h 来看，最大值模型排在第三位。

总的来看，在 UCSD Ped1 数据集上，最大值模型的性能超越了 MDT、Sparse、Adam 的方法，不及 Antic 和 Sparse Combination 两种方法。在异常定位能力上，尤其是在 FPR 较低时，最大值模型的性能与 Sparse Combination 相当。

以上方法中，提供了运行时间的方法并不多。表 4.5 比较了本文的光流直方图特征加最大值模型与 MDT、Sparse 两种方法的运行时间。可以看出光流直方图特



(a) 帧水平结果



(b) 像素水平结果

图 4.5 在 UCSD Ped1 数据集上朴素的最大值模型的结果。图 (a) 是帧水平 ROC 曲线；图 (b) 是像素水平 ROC 曲线。图中除了最大值模型，还有 Adam 等的方法、MDT、Social Force、Sparse 和 Sparse Combination 五种方法的结果。

表 4.5 运行时间比较

方法	秒 / 帧	FPS	平台	CPU (GHZ)	内存 (GB)
MDT ^[5]	25	0.04	-	3.0	2.0
Sparse ^[2]	3.8	0.26	-	2.6	2.0
最大值模型	0.0326	30.67	C++	2.5	8.0

征加最大值模的方法速度在 UCSD Ped1 数据集上运行速度达到 30.67FPS (实时), 远远超越 MDT 和 Sparse 方法的速度。

综合方法性能和运行时间, 可以看出光流直方图特征加最大值模型是一种有效的异常检测方法。它具有良好的异常事件定位能力, 且运行速度快。最大值模型的逻辑简单, 便于硬件实现和在线使用。

4.6.2 采集的实际监控视频

标准数据集便于与其他方法比较, 但也有场景单一的局限性。下面在自采集的数据集上测试光流直方图特征加最大值模型的效果。

实际视频的颜色、帧大小与 UCSD Ped1 数据集有较大差别, 特征采集的参数需要重新调整。经过试验, 在过马路、路口一、路口二这三个数据集上, 块大小参数选取为 $b_H = b_W = 16, d_T = 10, \delta_H = \delta_W = 4$, 光流直方图柱数为 $m = 9$; 最大值模型阈值为 $\delta = 0.2$ 。最后一个雨夜数据集, 块大小选取为 $b_H = b_W = 32, d_T = 9, \delta_H = \delta_W = 16$, 光流直方图柱数为 $m = 16$ 。另外, 特征除了光流直方图, 还增加了一维——前景计数器, 使特征变为了复合描述子。

实际采集的监控视屏没有真值, 只能根据检测结果能否找到人认为合理的解释判定检测的结果。下面的分析中将截取各个数据集正确和错误检测结果, 分析算法的有效性和存在的问题。

4.6.2.1 过马路数据集

图4.6展示了过马路数据集上异常检测结果的四幅截图, 图中红色覆盖的区域是被检测到的异常区域。其中 (a) 是急转弯的车辆; (b) 是在非机动车道上逆行的摩托车; (c) 和 (d) 均为横穿马路的行人。

过马路数据集的场景为城市环线, 机动车道双向六车道, 中间设置有围栏; 两侧有机动车道, 由绿化带隔开。画面中可见公交站, 车流量较大。

可以看出, 本文的以光流直方图特征为主的特征描述子加最大值模型的方法可以有效发现过马路数据集中行人横穿马路、摩托车逆行、车辆急转弯等事件。这些事件虽然没有造成严重的后果, 但是确实是事故隐患和风险所在。

4.6.2.2 路口一数据集

路口一数据集从城市十字路口上方的人行天桥上, 模拟交通信号灯的角度拍摄, 场景中央是道路, 路面上有大量行驶或等待的公交车、小汽车和摩托车。由

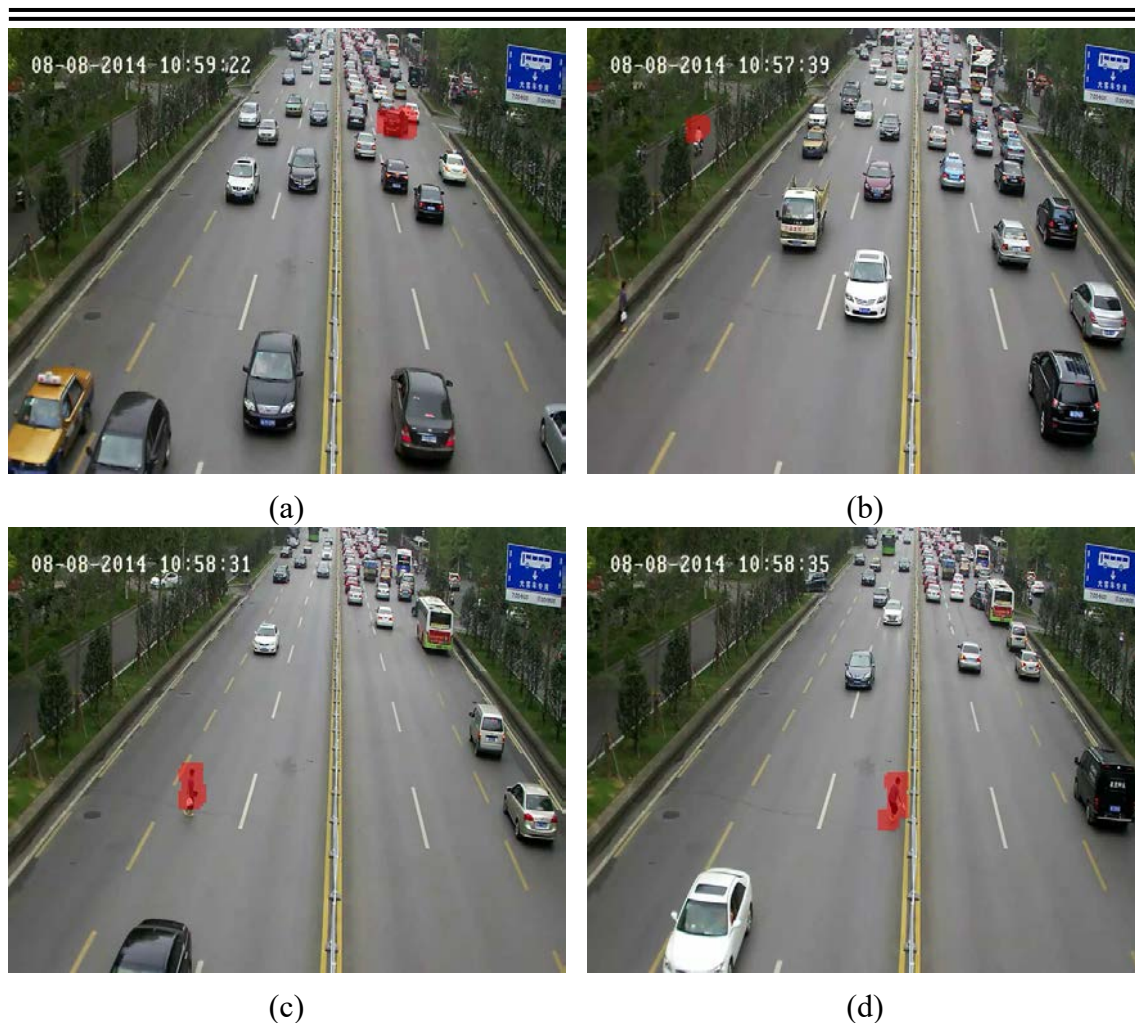


图 4.6 过马路数据集异常检测结果截图。(a) 是急转弯的车辆；(b) 是在非机动车道上逆行的摩托车；(c) 和 (d) 均为横穿马路的行人。

于人行天桥有轻微震动，路口一的视频画面有持续而轻微的上下抖动。这给特征提取增加了难度。

图4.7展示了路口一数据集上异常检测结果的四幅截图，异常内容均为逆行的摩托车。其中 (d) 中的两辆逆行的摩托车均被检测到。检测结果说明本文方法能有效地检测到在十字路口逆行的摩托车，说明此路口存在安全隐患，需要更好的交通疏导。

4.6.2.3 路口二数据集

路口二数据集与路口一数据集相似，也从城市十字路口上方的人行天桥上，模拟交通信号灯的角度拍摄。不同的是场景目标距离摄像机的位置更远一些，目标尺寸更小一些。

图4.8展示了路口二数据集中六福检测到的异常。(a) 是横穿马路的行人；(b) 是逆行的摩托车；(c) 也是横穿马路的行人；(d) 是越黄线行驶的摩托车；(e) 是头

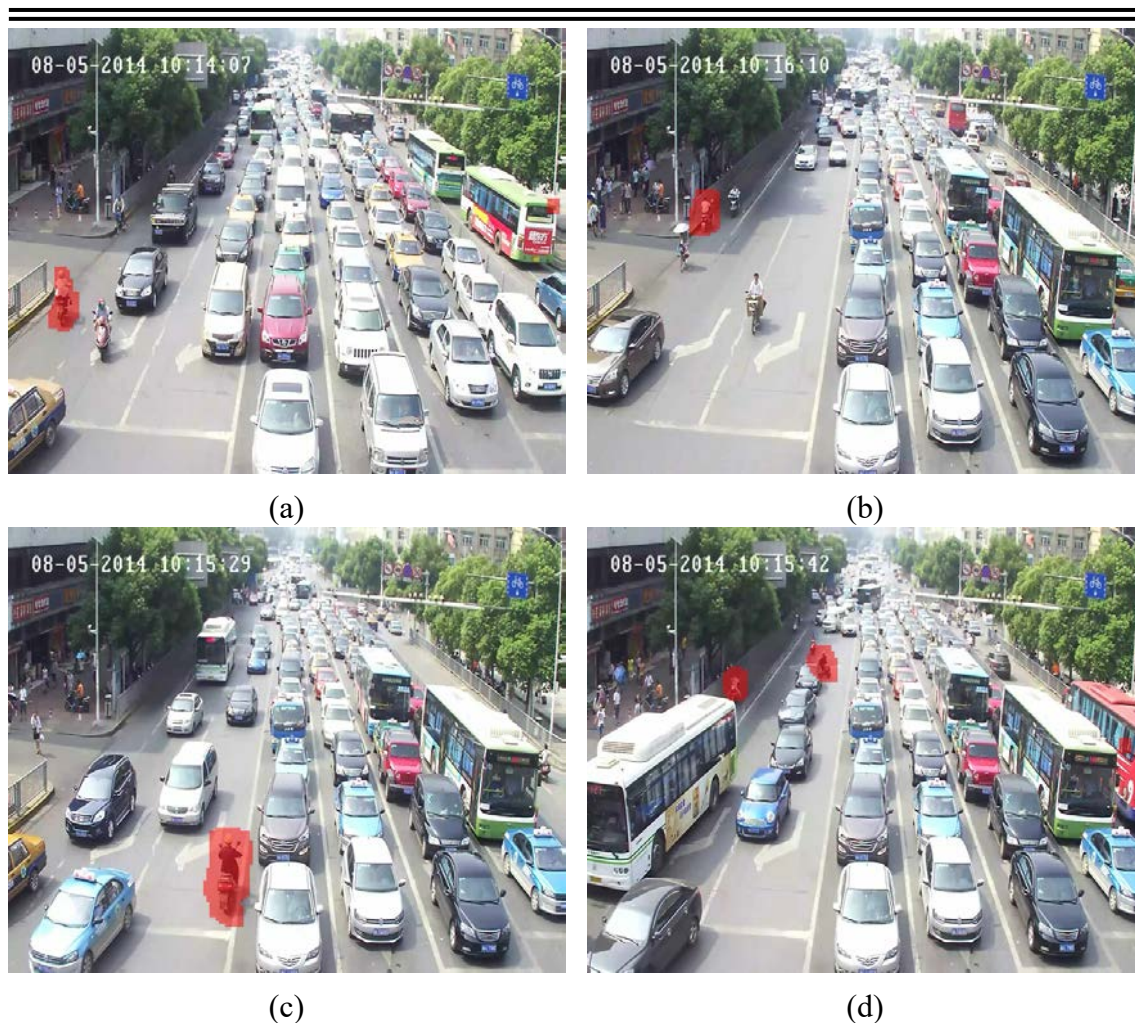


图 4.7 路口一数据集异常检测结果截图

顶扛麻袋的行人；(f) 是双层巴士。其中检测到的行人异常面积很小，这说明运动目标的尺度变小会导致被检测为异常的概率下降。检测到的摩托车与前两个数据集类似，有逆行或越界行驶的行为。有意思的结果是 (e)：一个头顶着一麻袋东西的行人，他所顶着的麻袋被检测为异常。这是因为这个行人顶着麻袋后整体高度显著高于其他行人，高出部分的运动由于在训练视频中少有而被检测为异常。(f) 的双层巴士被检测的原因与 (e) 相同，均是高度过高。它所遮挡的图像区域原本是静止的街道，与巴士的运动规律不同，因而被检测为异常。但是在人看来，一般认为双层巴士在城市道路上出现并不算异常事件。

在人的角度看，(e) 为异常似乎可以理解，(f) 为异常则显得不能接受。但是它们被检测到的原理却是一样的：均是由于遮挡了运动稀少的图像区域。被遮挡区域是否异常答案是不唯一的，在逻辑上不确定，这是监控视频异常检测无法避免的问题。只能说运动稀少的图像区域被运动目标遮挡是值得关注的事件，可能是异常事件。



图 4.8 路口二数据集异常检测结果截图。(a)是横穿马路的行人；(b)是逆行的摩托车；(c)也是横穿马路的行人；(d)是越黄线行驶的摩托车；(e)是头顶扛麻袋的行人；(f)是双层巴士。



图 4.9 雨夜数据集异常检测结果截图。(a)、(b)、(c) 均为跨过道路中线逆行的两辆卡车；(d) 为横穿道路的汽车。

4.6.2.4 雨夜数据集

为了检验算法在夜晚光照环境下和下雨天的效果，本文特意采集了雨夜数据集。它由安装在城市道路侧面的全景摄像机在一个下雨天的夜里拍摄。画面主要的运动目标集中在横着的干道上。

图4.9展示了雨夜数据集检测到的部分异常截图。(a)、(b)、(c) 均为跨过道路中线逆行的两辆卡车；(d) 为横穿道路的汽车。这说明本文方法在夜晚光照和下雨条件下也可以检测到车辆逆行这样的异常事件。

4.6.2.5 误检结果分析

图4.10为四个数据集中异常检测误检的例子。(a) 误检了正常行驶的出租车。原因是快速行驶的出租车是被树木遮挡，导致出租车的局部光流计算错误，从而引发误检。(b) 误检了正常行驶的公交车。原因是靠路边行驶的公交车比普通车辆高大，遮挡了人员穿梭的街道。公交车的运动规律与街道上行人的运动规律不同，

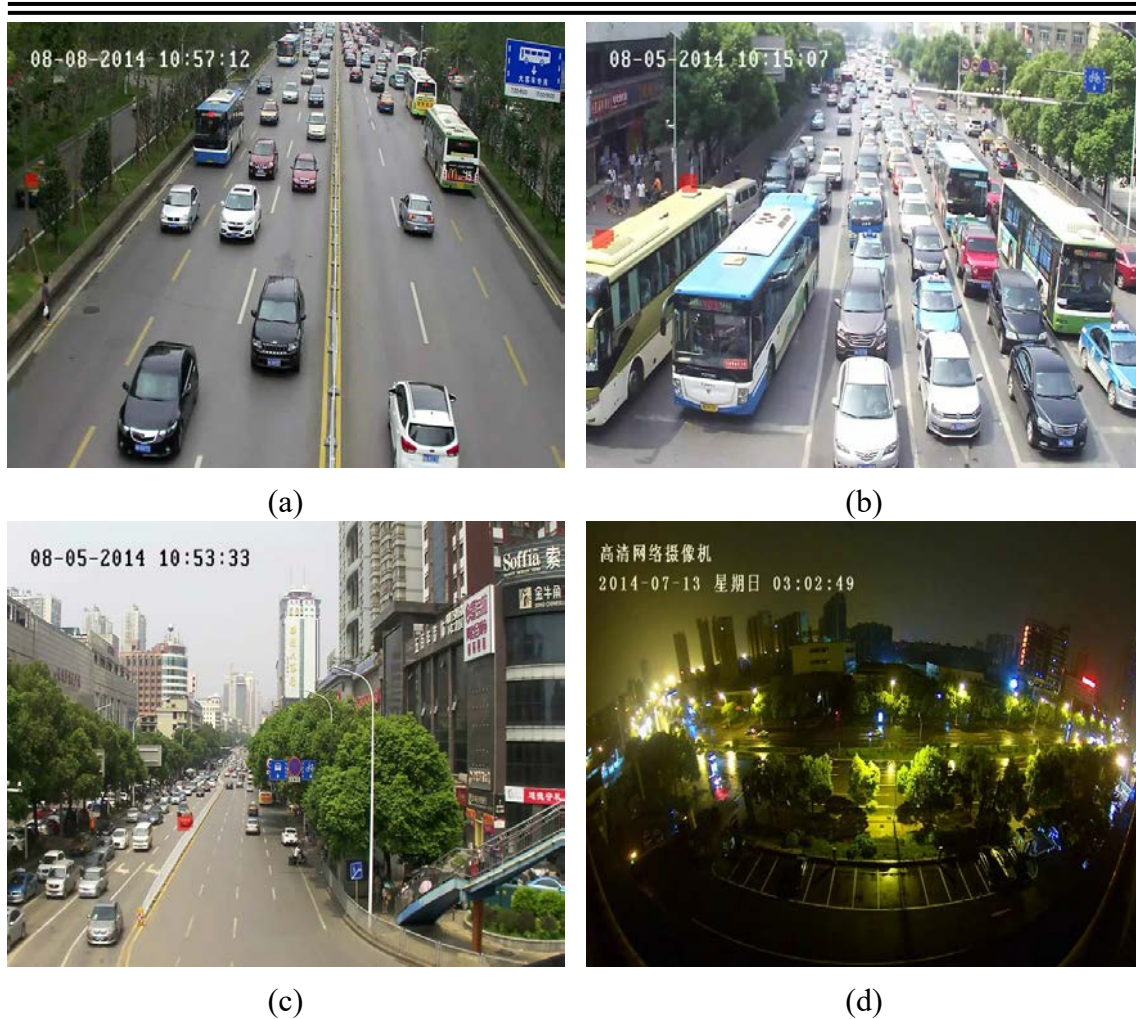


图 4.10 异常检测误检的例子

故而误检测为异常。(c) 误检了正常行驶的出租车。原因是出租车部分被护栏遮挡，导致局部光流计算错误，引发误检。(d) 误检了路面反光的积水。在下雨的夜里，车灯的光亮通过积水反射，被误检为运动目标；而在积水处本来没有运动目标，故而引发误检。

通过以上分析，可以发现在白天，视频中的遮挡是引发误检的最主要原因。正常的运动目标如果遮挡运动稀疏区域，虽然在物理空间上处于不同的位置，但由于遮挡，在视频中被认为出在同一个位置。由于两者运动规律本来就不同，这将难以避免地会引发误判。反过来，如果静止的物体遮挡运动目标，有可能引起局部光流计算错误，从而造成误检。

造成误检的另一个原因是光照条件不好。例如在夜晚，灯光和反射等现象容易引发误判。

4.6.2.6 漏检结果分析

四个数据集中展示了异常检测结果，图4.6和图4.7展示的异常对象被检测算法“跟踪”得较好。而图4.9中的逆行卡车只能断断续续地检测到。图4.8(a)(c)中的横穿马路的行人只在少数帧中被检测到。

造成漏检的最主要原因是运动目标所占的像素面积小。图4.8(a)(c)中的横穿马路的行人相对于其他异常目标所占的像素面积很小，这会导致提取到的光流数目不足，从而不容易检测到异常。

光照条件不好也容易造成异常事件的漏检。夜晚提取到的光流也明显少于白天，对运动规律建模并发现异常更加困难。

4.7 结论

本章对目前监控视频异常检测的评价准则进行探讨。对于异常检测的性能评价，除了现有的帧水平 AUC 和像素水平 AUC，还引入了它们的差值 ΔAUC 和它们的调和平均数 AUC_H 作为评价指标，防止通过参数调整突出某一个水平的检测性能而牺牲另一个水平检测性能的方法获得好评。

在测试模型前，本章对提取的光流直方图特征进行了数据降维和可视化。选取的两个块 9-柱光流直方图特征分别有两个或七个维度是接近零的。如此多的维度为零可能对训练算法产生影响。在选用训练算法时应当注意零维度的影响。值得注意的是，最大值模型不受零维度影响。

从降维和可视化的结果来看，特征分布呈小范围内的不规则形状聚簇或在一个圆形区域内分布。两类特征分布都可以找到分布的边界。最大值模型直接对应着训练特征在高维空间上的物理边界。单类 SVM 则可能寻找到一个拟合度更好的边界。稀疏表示不是从边界的角度考虑，而是试图选取一些基特征来最好的线性表示所有正常特征。

在实验中，首先选用了使用最频繁的 UCSD Ped1 数据集进行测试。本文的三种模型性能，最大值好于单类 SVM 好于稀疏模型；其中单类 SVM 在帧水平性能上小幅超越最大值模型。与目前在此数据集上效果最好的六种算法对比，本文光流直方图加最大值模型的方法性能上排在第三位；在速度可比的三种算法中，除了号称可以达到每秒 150 帧的 Sparse Combination^[6] 算法，本文方法以实时运行速度 (30.67FPS) 远远超越了 MDT^[5] 和 Sparse^[2] 方法。能达到这个运行速度，一方面是由于光溜特征提取时的前景遮罩和优化，另一方面是最大值模型本身逻辑简单，运算速度快。

在对自采集的四段城市道路监控的测试中，本文方法可以在少有误检的情况下，发现行人横穿马路、车辆和摩托车逆行、行人头顶麻袋等异常事件，说明了本文方法对实际监控视频的有效性。

异常检测存在的主要问题是误检和漏检。通过实验分析，视频中的遮挡是引发误检的最主要原因；夜晚的灯光和反射等现象也容易引起误检。造成漏检的最主要原因是运动目标所占的像素面积小；光照条件不好也容易造成漏检。

第五章 总结与展望

5.1 总结

监控视频异常检测涉及数字视频处理和数据挖掘两个学科，是智能监控技术的重要组成部分，也是一项非常具体的智能监控应用，具有广阔的应用前景。本文总结了基于分块处理的视频异常检测算法框架，并进行了以下几个方面的工作：

(1) 总结了基于分块处理的视频异常检测的原理、常用的特征和模型。视频异常检测常用的特征有光流直方图、3D 梯度、前景计数器和复合描述子。光流直方图能直观反映运动目标速度大小和方向，3D 梯度能综合纹理和运动信息，前景计数器能感知场景的运动密度，复合描述子可以综合两种以上的基本特征。异常检测是找出行为很不同于预期对象的过程，从数据挖掘的角度看，有基于聚类、基于统计模型、基于分类等多种方法。异常检测是高度依赖于应用的一类问题，只能依据具体应用去选择方法。本文综合近年来优秀论文中的异常检测模型，最终选取了朴素的 \max 模型、单类 SVM 和稀疏模型作为候选方法。

(2) 实现了一个视频异常检测系统。它以光流直方图特征为主的特征描述子作为特征，以朴素的 \max 模型、单类 SVM 和稀疏模型作为模型，可以用这些特征模型组合进行视频异常检测。对有真值的数据集，可以对方法性能进行测评。在准备好训练视频和测试视频后，该系统可以一键完成模型训练、异常检测和结果评估。

(3) 在标准数据集上对比了本文实现的方法和近年来的优秀视频异常检测方法的性能和速度，并测试了本文方法在实际监控视频上的效果。在标准数据集上的对比结果说明了本文光流直方图特征和 \max 模型组合在视频异常检测，特别是对异常事件的定位上的有效性。本文方法在标准数据集上达到实时运行的速度也超越了绝大多数现有方法。

(4) 在实际监控视频中，不同视角、距离、光照、气象条件下，本文方法能够检测出车流人流中的有意义的异常事件。这也说明了本文光流直方图特征和 \max 模型组合方法的有效性。另外，从实验结果还可以分析出误检的漏检的主要原因分别是遮挡、光照差和目标尺度小、光照差。

5.2 展望

在研究视频异常检测的过程中，也发现了此领域的特点、存在的问题、本文方法的局限性和改进方向：

(1) 监控视频异常检测是一项工程性较强的研究。就本文基于分块的方法而言，特征提取属于数字图像处理领域，异常检测则属于数据挖掘的范畴。一个完

整的视频异常检测方法往往是多个步骤视频处理和数据挖掘步骤的串行，其中涉及到很多的方法、模型和参数，任何一个步骤不佳则整体异常检测的效果都会大打折扣。这也使得实现方法、定位问题和改进方法的难度较大。

(2) 监控视频异常检测领域高质量的数据集和经典方法代码严重匮乏。视频异常检测的研究早在十年前就有，但至今高质量数据集仍然匮乏。UCSDPed 数据集是仅有的一个有像素水平真值的数据集，致使众多研究论文都采用此数据集测试分析方法。但此单一场景的局限性显而易见：由于场景中非行人为异常，采用类似行人检测器的方法会取得很好的效果。另一个问题是研究者们均不公布代码。前面提到视频异常检测的工程性强，步骤多，实现难度大，让他人来实现研究者本人的方法难度就更大了。可能也正是因为如此，研究者们小心保存自己的代码不予公布，以至于后来的研究者只能摘抄以前研究论文中的数据用于方法的对比。这不仅提高了进入此领域的门槛，也容易造成结果对比的误差，更限制了新数据集的使用。从长远看，对于监控视频异常检测的研究非常不利。

(3) 特定场景下光流计算易出错。本文使用的光流直方图特征虽然经过精心优化和过滤，仍然存在特定场景下准确度不高的问题。第四章总结过的遮挡、反光的情容易造成光流出错，容易引发误检。

(4) 运动稀疏区域是否异常无法判别。运动稀疏区域的光流样本过少，以至于不足以训练出可靠的模型来检测异常。这其实上升到了逻辑层次：从来没有运动目标的区域出现了运动目标，这是不是异常呢？这不是单靠运动模式信息能回答的问题，还需要目标的类别等其他逻辑信息才能回答。故而本文只能采用假设运动稀疏区域没有异常的方法来处理。

(5) 更好的异常检测效果需要更针对特定场景的方法。本文的方法和做比较的一些方法都是一般的异常检测方法，理论上可以用于所有的场景；实际上不同场景的视角、目标尺寸、光照、目标运动规律千差万别，需要更多的先验信息才能做出效果更好的异常检测应用。比如，运动模式信息与目标分类信息结合，将有望取得更好的结果。然而这项工作更加工程化，目标分类本身也是没有彻底解决的问题。所以针对某一类场景做好目标分类与运动模式信息结合可能更有希望获得成功。

致 谢

时光荏苒，岁月如梭，转眼两年半的硕士生活即将结束。硕士阶段是学术生涯的开端，这两年半，在实验室的学习生活令我的知识积累和科研能力上了一个台阶，为今后的学习工作打下了坚实基础。在此向所有关心和帮助过我的老师、同学、队干部、亲人表示最衷心的感谢。

首先衷心感谢我的导师张茂军教授。我的硕士生活是在张老师领导的计算摄影教研室度过的。张老师对我的课程学习，科研方向给予了重要的指导。更重要的是，张老师勇挑重担的决心和工作的激情对我的科研状态产生了巨大的鼓舞！张老师领导攻关，完成了一项项发明，更是我既引以为豪又想奋力赶超的榜样！要做好一件事就必须带着激情去做，用心去做，拼搏着去做，有了这样的闯劲，科研的山头也不是那么高不可攀。

同时，我要感谢实验室的刘煜老师、张政老师。两位老师经验丰富、知识全面，而且与我在同一个屋内，对于的学习生活提供了许多帮助，为我遇到的难题答疑解惑。我还要感谢王炜教授在直方图计算思想上给与我的指导。

感谢实验室左承林师兄、彭杨师兄、尹晓晴师兄、李卫丽师姐、李靖师兄、肖华欣师兄，还有已经毕业的陈立栋师兄、赖世明师兄、王斌师兄、高晨旭师兄，他们对于我平时的学习生活提供了诸多帮助。还要感谢于浩同学对我实验视频数据准备提供的帮助。

感谢计算摄影工程中心翟永平博士和向北海工程师在光流算法、运动检测算法和视频数据的采集上给与我的指导和帮助。在工程中心实习的两个多月，他们教会了我很多实用的技术。特别感谢翟永平博士在我研究异常检测算法的初期对于我方向和技术上的指导，他的帮助使我快速入门，找到了方向。

感谢五队干部、同学们两年半的陪伴！

最后，衷心感谢我的妈妈和姥姥，远在千里之外一直对我默默的关系、支持和付出。

参考文献

- [1] Chen Y, Zhou X S, Huang T S. One-class SVM for learning in image retrieval [C]. In Image Processing, 2001. Proceedings. 2001 International Conference on: 34–37.
- [2] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection [C]. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on: 3449–3456.
- [3] Jodoin P, Saligrama V, Konrad J. Behavior subtraction [C]. In PROCEEDINGS-SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING. 2008: 6822.
- [4] 王素玉, 沈兰荪. 智能视觉监控技术研究进展 [J]. 中国图象图形学报. 2007, 12: 1505–1514.
- [5] Mahadevan V, Li W, Bhalodia V, et al. Anomaly detection in crowded scenes [C]. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on: 1975–1981.
- [6] Lu C, Shi J, Jia J. Abnormal event detection at 150 fps in matlab [C]. In Computer Vision (ICCV), 2013 IEEE International Conference on: 2720–2727.
- [7] WANG T, SNOUSSI H, SMACH F. Detection of Visual Abnormal Events via One-class SVM [C]. In Proceedings of the 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV): 113–119.
- [8] Jodoin P, Saligrama V, Konrad J. Behavior subtraction [J]. Image Processing, IEEE Transactions on. 2012, 21 (9): 4244–4255.
- [9] Saligrama V, Chen Z. Video anomaly detection based on local statistical aggregates [C]. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on: 2112–2119.
- [10] Kratz L, Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models [C]. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on: 1446–1453.
- [11] Ballan L, Bertini M, Del Bimbo A, et al. Event detection and recognition for semantic annotation of video [J]. Multimedia Tools and Applications. 2011, 51 (1): 279–302.

-
-
- [12] Wong K-Y K, Cipolla R. Extracting spatiotemporal interest points using global information [C]. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*: 1–8.
 - [13] Laptev I. On space-time interest points [J]. *International Journal of Computer Vision*. 2005, 64 (2-3): 107–123.
 - [14] Oikonomopoulos A, Patras I, Pantic M. Spatiotemporal salient points for visual recognition of human actions [J]. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*. 2005, 36 (3): 710–719.
 - [15] Willems G, Tuytelaars T, Van Gool L. An efficient dense and scale-invariant spatio-temporal interest point detector [M]. In *Computer Vision–ECCV 2008*. Springer, 2008: 650–663.
 - [16] Adam A, Rivlin E, Shimshoni I, et al. Robust real-time unusual event detection using multiple fixed-location monitors [J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2008, 30 (3): 555–560.
 - [17] Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification [M]. In *Computer Vision–ECCV 2006*. Springer, 2006: 490–503.
 - [18] Ballan L, Bertini M, Del Bimbo A, et al. Recognizing human actions by fusing spatio-temporal appearance and motion descriptors [C]. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*: 3569–3572.
 - [19] Dollár P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features [C]. In *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*: 65–72.
 - [20] Fu Z, Hu W, Tan T. Similarity based vehicle trajectory clustering and anomaly detection [C]. In *Image Processing, 2005. ICIIP 2005. IEEE International Conference on*: II–602–5.
 - [21] Basharat A, Gritai A, Shah M. Learning object motion patterns for anomaly detection and improved object detection [C]. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*: 1–8.
 - [22] Morris B T, Trivedi M M. A survey of vision-based trajectory learning and analysis for surveillance [J]. *Circuits and Systems for Video Technology, IEEE Transactions on*. 2008, 18 (8): 1114–1127.

-
-
- [23] Piciarelli C, Micheloni C, Foresti G L. Trajectory-based anomalous event detection [J]. *Circuits and Systems for Video Technology*, IEEE Transactions on. 2008, 18 (11): 1544–1554.
- [24] Wang X, Tieu K, Grimson E. Learning semantic scene models by trajectory analysis [M] // Wang X, Tieu K, Grimson E. *Computer Vision—ECCV 2006*. Springer, 2006: 2006: 110–123.
- [25] Gibson J J. The perception of the visual world. [J]. 1950.
- [26] Lucas B D, Kanade T, et al. An iterative image registration technique with an application to stereo vision. [C]. In *IJCAI*. 1981: 674–679.
- [27] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005*. IEEE Computer Society Conference on: 886–893.
- [28] Unusual crowd activity dataset of University of Minnesota, <http://mha.cs.umn.edu/movies/crowdactivity-all.avi>. 2011. <http://mha.cs.umn.edu/movies/crowdactivity-all.avi>.
- [29] Han J, Kamber M. *Data mining: concepts and techniques*. 2001.
- [30] Kim H. *Probability, Statistics, and Random Processes for Engineers* [M]. 2004.
- [31] UCSDped1 dataset, <http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm>. 2013.
- [32] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution [J]. *Neural computation*. 2001, 13 (7): 1443–1471.
- [33] Chen Y, Zhou X S, Huang T S. One-class SVM for learning in image retrieval [C]. In *Image Processing, 2001. Proceedings. 2001 International Conference on*. 2001: 34–37.
- [34] Tax D M, Duin R P. Support vector data description [J]. *Machine learning*. 2004, 54 (1): 45–66.
- [35] Olshausen B A. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. *Nature*. 1996, 381 (6583): 607–609.
- [36] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries [J]. *Image Processing, IEEE Transactions on*. 2006, 15 (12): 3736–3745.
- [37] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration [J]. *Image Processing, IEEE Transactions on*. 2008, 17 (1): 53–69.

-
-
- [38] Zhang K, Zhang L, Yang M-H. Fast compressive tracking [J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2014, 36 (10): 2002–2015.
- [39] Barnich O, Van Droogenbroeck M. ViBe: a powerful random technique to estimate the background in video sequences [C]. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. 2009: 945–948.
- [40] Shi J, Tomasi C. Good features to track [C]. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. 1994: 593–600.
- [41] Kalal Z. TRACKING LEARNING DETECTION [D]. United Kingdom: Centre for Vision, Speech and Signal Processing Faculty of Engineering and Physical Sciences, University of Surrey, 2011.
- [42] Schölkopf B, Smola A J. *Learning with kernels: Support vector machines, regularization, optimization, and beyond* [M]. MIT press, 2002.
- [43] Aharon M, Elad M, Bruckstein A. SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation [J]. *Signal Processing, IEEE Transactions on*. 2006, 54 (11): 4311–4322.
- [44] Mairal J, Bach F, Ponce J, et al. Online dictionary learning for sparse coding [C]. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 2009: 689–696.
- [45] Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding [J]. *Statistics*. 2009, 1050: 1.
- [46] SPARSe Modeling Software, <http://spams-devel.gforge.inria.fr/>. 2013.
- [47] Chang C-C, Lin C-J. LIBSVM: A library for support vector machines [J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2011, 2 (3): 27.
- [48] Van der Maaten L, Hinton G. Visualizing data using t-SNE [J]. *Journal of Machine Learning Research*. 2008, 9 (2579-2605): 85.
- [49] Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model [C]. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 2009: 935–942.
- [50] Antić B, Ommer B. Video parsing for abnormality detection [C]. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2011: 2415–2422.

作者在学期间取得的学术成果

发表的学术论文

- [1] Tan, Hanlin, and Lidong Chen. "An approach for fast and parallel video processing on Apache Hadoop clusters." Multimedia and Expo (ICME), 2014 IEEE International Conference on. IEEE, 2014. (EI 收录, 检索号:20153001066496.)
- [2] Tan, Hanlin, Zhai, Yongping, Liu, Yu, and Zhang, Maojun. "Fast Anomaly Detection in Traffic Surveillance Video based on Robust Sparse Optical Flow." Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). 2016.(审稿中)
- [3] Zhao, Yawei, Honghui Chen, Jia Xu, Hanlin Tan, and Tao Chen. "Mirror Data to the Ceiling: A New Completely Wireless Data Center." In Frontiers in Internet Technologies, pp. 32-45. Springer Berlin Heidelberg, 2014.

研究成果

- [1] 徐玮, 谭瀚霖, 赖世铭, 王伟, 刘煜. 一种基于 Hadoop 的视频处理方法及系统: 中国, 公开号: CN 104853061 A. (中国专利公开号, 实质审查中.)

竞赛获奖

- [1] 谭瀚霖, 王志飞, 赵彦鹏. 全国研究生数学建模竞赛二等奖. 2013.
- [2] 谭瀚霖, 赵亚威, 刘瑶. 全国物联网大赛挑战赛二等奖. 2014.
- [3] 谭瀚霖, 王志飞, 赵亚威. 全国研究生数学建模竞赛二等奖. 2014.
- [4] 谭瀚霖, 王志飞, 贺云岳, 邓经升, 罗蕾. 基于相机阵列的多视图立体成像和超分辨率成像系统. 长城信息杯三等奖. 2015