**哈尔滨工业大学**

**硕士学位论文中期报告**

**题 目：拥挤场景中视频异常事件检测方法研究**

**院 （系） 计算机科学与技术**

**学 科 计算机科学与技术**

**导 师 王轩**

**研 究 生 刘凤桐**

**学 号 11S051058**

**中期报告日期 2013.3.26**

**研究生院制**

**二〇一三年三月**

**硕士学位论文**

**拥挤场景中视频异常事件检测方法研究**

**RESEARCH ON VIDEO ABNORMAL EVENT DETECTION IN CROWDED SCENE**

**刘凤桐**

哈尔滨工业大学

2013年12月

国内图书分类号： 学校代码：10213

国际图书分类号： 密级：公开

工学硕士学位论文

拥挤场景中视频异常事件检测方法研究

|  |  |  |
| --- | --- | --- |
| 硕士研究生 | ： | 刘凤桐 |
| 导 师 | ： | 王轩教授 |
| 申请学位 | ： | 工学硕士 |
| 学科 | ： | 计算机科学与技术 |
| 所 在 单 位 | ： | 深圳研究生院 |
| 答 辩 日 期 | ： | 2013年12月 |
| 授予学位单位 | ： | 哈尔滨工业大学 |

Classified Index: TP

U.D.C:

Dissertation for the Master Degree of Engineering

RESEARCH ON VIDEO ABNORMAL EVENT DETECTION IN CROWDED SCENE

|  |  |
| --- | --- |
| **Candidate:** | Liu Fengtong |
| **Supervisor:** | Prof. Wang Xuan |
| **Academic Degree Applied for:** | Master of Engineering |
| **Specialty:** | Computer Science |
| **Affiliation:** | Shenzhen Graduate School |
| **Date of Defence:** | December, 2013 |
| **Degree-Conferring-Institution:** | Harbin Institute of Technology |

# 摘 要

视频中的异常事件检测是智能视频监控中一个重要研究方向，近年来逐渐成为深受关注的应用领域。这主要是因为传统的视频监控系统与设备存在效率低、人员易疲劳等缺点，因此需要一个更加智能的监控系统来辅助检测。而针对于人口流动性大、稠密度高的公共场所，我们称之为拥挤场景，相应的公共安全问题自然受到更多的关注。基于视频的异常检测又可划分为全局异常和局部异常检测，如何兼顾二者进行有效的检测同样是研究的重点。在非拥挤场景中通常采用跟踪轨迹方法进行异常判断；而在拥挤场景中，很难分离每个目标或事件，因此更多有效的方法选择使用局部特征来表示异常事件，包括时空信息、抽取运动特征以及建立光流直方图等。异常事件检测涉及图像处理、计算机视觉、模式识别等多个领域，本课题重点研究拥挤场景中的异常事件检测，并对涉及的相关问题进行了深入的研究。

本课题首先研究一种基于时空MRF框架的视频异常检测方法。虽然构建时空MRF后的能量最小化问题是NP难，但可以转化为置信传播进行估算。在特征提取阶段，为了更好的表示区域的运动特征，提取视频中图像的光流后，采用MPPCA方法对低层的光流进行学习，因为MPPCA关联了概率模型和PCA，通过组合多个PCA模型可以更好的对复杂数据建模。而相应区域就构成了时空MRF结构中节点，从而可以选取一段时域内的视频帧，构成一个完整的时空MRF图模型。这样，通过在MRF图模型中应用置信传播最小化能量函数后，可得出图模型中的各节点的标号，即由此推断节点是否异常来实现视频异常检测。

稀疏表示在信号处理和压缩感知领域已有广泛的应用，而本课题通过构建样本的稀疏表示代价函数对异常进行判定，从而达到检测视频中异常事件的目标。由于正常样本和异常样本经过稀疏表示后得到的稀疏系数幅值相差较大，所以可以先对正常样本进行字典学习，再用学习的字典对测试样本进行稀疏表示，根据计算得出的稀疏表示代价对样本判断是否为异常。

在稀疏表示过程中，虽然样本的特征向量经过稀疏编码后变得很稀疏，但应用的字典比较稠密，所以可以通过消除字典中的冗余项来实现一种近似稀疏表示方法，从而提高稀疏表示过程的速度。在此基础上，将其融合到视频异常检测框架中，可以大幅度的提高帧的处理速度，从而保证了基于稀疏表示的视频异常检测框架的实时性。

**关键词：**异常检测；拥挤场景；MPPCA；MRF；稀疏表示；字典学习

# Abstract

Abnormal event detection in video is an important research in the intelligent monitoring system, and it is becoming a focused application field gradually. Since there are some defects in the traditional intelligent monitoring system such as low efficiency and so on, a more intelligent monitoring system is required to aid the anomaly detection. Regarding the crowded scene where the crowd has a big liquidity and density, the corresponding public security problem is paid more attention. The abnormal detection based on the video contains two types: global anomaly detection and local anomaly detection, so it is also an important point how to consider both of two types. In recent years, many key areas has made a progress, including background modeling, target tracking, pedestrian detection, motion recognition, crowd metering and traffic monitoring. In the uncrowded scene, the way of tracking the target is usually used to infer the anomaly. However, in the crowded scene, it is hard to separate every target or event, so more efficient methods are using local features to present abnormal event, which contains space-time information, motion features and flow histogram.

Abnormal event detection referred to several fields, for example, Image Processing, Computer Vision and Pattern Recognition. In this paper, the main research content is the abnormal event detection in crowded scene.

Belief propagation theory has been used widely in early vision. Although the energy minimization problem of space-time MRF (Markov Random Field) is NP-Hard, it can be solved by transforming it to belief propagation. At first, the paper research a method based on space-time MRF framework to detect the anomaly in the video. In the stage of extracting features, it is better to present the motion feature of region by using MPPCA (Mixture of Probabilistic Principal Component Analysis) to learn the optical flow. Because the method of MPPCA combine the probabilistic model and PCA (Principal Component Analysis), and the combination of several PCA model can model the complex data better. And the corresponding region becomes the node of space-time MRF; therefore, a part of frames in video can construct a complete space-time MRF model. Then the labels of the graph model can be inferred by using belief propagation, which means the achievement of video anomaly detection.

Spares representation is often involved in signal processing and compression sense. And the paper focuses on the anomaly inference by constructing sparse representation cost function, which can achieve the target to detect abnormal event in the video. The sparse efficient between the normal and abnormal sample exists difference, which is computed by using sparse representation. Therefore, the dictionary learned from normal samples can be used to represent the test samples, and then the state of frame can be inferred by computing the sparse representation cost.

In the process of sparse representation, the features vector of samples is sparse after sparse representation, while the applied dictionary is still dense. So if the reduce item in dictionary can be reduced, the method can approximate the process of sparse representation, which can improve the speed of the process. Then it can be added to the video anomaly detection framework, and improve the processing speed of each frame. As a result, this method can promise the just-in-time of the process of detection.

**Keywords:** crowded scene, mppca, mrf, sparse representation, dictionary learning

# 目 录

[摘 要 I](#_Toc371065348)

[Abstract II](#_Toc371065349)

[目 录 IV](#_Toc371065350)

[第1章 绪论 1](#_Toc371065351)

[1.1 课题背景 1](#_Toc371065352)

[1.2 课题研究的目的及意义 1](#_Toc371065353)

[1.3 国内外研究现状 2](#_Toc371065354)

[1.3.1 国内研究现状 2](#_Toc371065355)

[1.3.2 国外研究现状 3](#_Toc371065356)

[1.4 主要研究内容及论文结构 4](#_Toc371065357)

[第2章 基于时空MRF的视频异常检测 6](#_Toc371065358)

[2.1 引言 6](#_Toc371065360)

[2.2 早期视觉问题与MRF相关理论 6](#_Toc371065361)

[2.2.1 早期视觉问题 6](#_Toc371065362)

[2.2.2 MRF与图模型表示 7](#_Toc371065363)

[2.2.3 MRF能量最小化方法 8](#_Toc371065364)

[2.3 MPPCA学习局部状态的运动模式 12](#_Toc371065365)

[2.3.1 光流特征提取 12](#_Toc371065366)

[2.3.2 混合概率主成分分析 12](#_Toc371065367)

[2.3.3 学习局部状态的运动模式 13](#_Toc371065368)

[2.4 构建运动模式的时空MRF模型 14](#_Toc371065369)

[2.4.1 时空MRF模型的能量函数 14](#_Toc371065370)

[2.4.2 时空MRF模型的自更新 15](#_Toc371065371)

[2.5 本章小结 16](#_Toc371065372)

[第3章 字典学习与稀疏表示 17](#_Toc371065373)

[3.1 引言 17](#_Toc371065375)

[3.2 稀疏表示 17](#_Toc371065376)

[3.3 字典学习 19](#_Toc371065377)

[3.3.1 最优方向法 19](#_Toc371065378)

[3.3.2 K-SVD算法 19](#_Toc371065379)

[3.4 本章小结 22](#_Toc371065380)

[第4章 稀疏表示方法检测视频中的异常 23](#_Toc371065381)

[4.1 引言 23](#_Toc371065384)

[4.2 基于稀疏表示方法的视频异常检测 24](#_Toc371065385)

[4.2.1 光流直方图特征提取 24](#_Toc371065389)

[4.2.2 视频异常检测框架 26](#_Toc371065390)

[4.3 视频异常检测方法的改进 28](#_Toc371065391)

[4.3.1 快速稀疏表示近似算法 28](#_Toc371065393)

[4.3.2 时间与空间复杂度分析 29](#_Toc371065394)

[4.3.3 融合快速稀疏近似算法 30](#_Toc371065395)

[4.4 本章小结 31](#_Toc371065396)

[第5章 系统实现与实验结果 32](#_Toc371065397)

[5.1 引言 32](#_Toc371065399)

[5.2 系统实现 32](#_Toc371065400)

[5.2.1 实验开发环境 32](#_Toc371065404)

[5.2.2 系统实现框架 32](#_Toc371065405)

[5.3 实验结果分析 34](#_Toc371065406)

[5.3.1 时空MRF的实验结果 34](#_Toc371065408)

[5.3.2 稀疏表示的实验结果 34](#_Toc371065409)

[5.4 本章小结 37](#_Toc371065410)

[结论 39](#_Toc371065411)

[参考文献 40](#_Toc371065412)

[哈尔滨工业大学硕士学位论文原创性声明 43](#_Toc371065413)

[致谢 44](#_Toc371065414)

# 绪论

## 课题背景

近年来，公共安全备受关注并逐渐成为重要的社会问题。政府为了更好的保障人民的人身财产安全在公共安全上加大了投入，越来越多的网络摄像头等监控设备被安置在公路、公园、机场及地铁等公共场所。同时，对于智能视频监控系统也提出了全新的需求。视频监控系统已经在政府部门、金融行业、公安系统和交通管理系统中广泛地部署与应用，但传统的视频监控系统仍然存在效率上的诸多问题。

智能视频监控已经成为一个重要的研究领域。目前的技术发展处于这样一个阶段，监控设备成本低而操作和观察监控系统的人力成本高。监控设备的普及度越来越高，几乎每一个小的社区超市都安装了监控设备。但当发生偷盗等犯罪事件时，通常只能查看视频记录进行确认，发现的延迟性很高。因此，对于具有实时性的智能监控系统的需求在不断增加。

对于传统的视频监控系统与设备，存在着效率低、人员易疲劳等缺点，这就需要一个更加智能的监控系统来辅助人工的检测。在新的需求下，通过融合图像处理、计算机视觉以及模式识别等关键技术，智能视频监控系统应能提取视频中的有效信息，分析和判断视频中的异常情况，并能最早的发出警报，从而实现有效进行事前预警，事中处理，事后及时取证的全自动、全天候及实时监控。其中异常事件检测作为视频监控中的重要应用，已经引起了极大的兴趣。根据场景的差异，可以把异常检测分为拥挤场景和非拥挤场景两种情况。在非拥挤的场景中，由于只是针对单个或少数目标的分析，前景目标易于从背景中分离，可以对目标进行识别以及基于跟踪的分析方法进行相应的异常分析。而在拥挤场景中，这种方法则并没有很好的效果，因为很难取得很好的跟踪效果。

拥挤场景中的异常检测作为智能监控系统中的一个重要研究课题，它主要针对人口流动性大、稠密度高等公共场所，因此相应的公共安全问题受到很大的关注，如2013年波士顿马拉松赛事发生的爆炸事件、2013年北京首都机场T3爆炸事件及频发的聚集和斗殴事件等，如何实时有效地检测场景中的异常行为仍然是很大的挑战。

## 课题研究的目的及意义

计算机视觉领域中的一个重要挑战就是拥挤场景中人的活动分析。目前，基于跟踪等技术个人的行为已经能被很好的理解，而拥挤场景下的分析仍然面临着巨大的挑战，如突发事件和自发性的活动。群体行为分析在研究群体活动上逐渐成为一个新的领域，并有可能在将来发展成为一个新的应用方向。近年来群体事件不断发生，公共安全问题依然严峻，为了避免群体事件对社会带来的危害，有效的群体行为分析可以及时地对事件做出响应，例如自动检测拥挤场景中的骚乱、暴动等行为和发生异常的区域。

本课题的研究内容主要是针对拥挤场景中的异常事件进行检测，在实际的公共安全和社会生活中具有重要的意义和研究价值。传统监控系统中存在着诸多问题，由于人类自身生理条件的限制，无法准确地对监控平台中的多个窗口进行实时观察，易发生漏看，易产生疲劳，并不能真正实现24小时全天候监控。对于传统监控系统中的不足，一个智能的异常监控系统应具备以下几个条件：

有效性：监控系统应能够准确有效的检测异常事件并发出报警，这就需要监控系统对场景进行合理的建模，对于采集的视频数据进行一定处理，尽量排除噪声数据并获取有价值的信息，在检测到异常事件时能对相应数据进一步提取、存储，供后期的调查取证所用。

实时性：监控系统还能够实时地检测并在最早的时刻发出报警，这进一步提高了对监控系统的要求，对于突发事件，尽早的发出警报可以避免事态的恶化并减少事件带来的不必要损失，因此实时性也是一个很重要的标准。对于视频数据的分析，需要更强大的数据处理能力的支撑，同时一个快速、好的分析算法也可以保证监控系统能做出最快的分析和判断。

容错性：监控系统应该具有一定的容错能力，一定的错报、误报是不可避免的，但是过多的错报、误报现象就会会加重工作人员的负担，影响工作效率。这就需要监控系统具有对数据中的噪点做更多的处理，既要避免漏报，也要控制错报、误报的比率。

## 国内外研究现状

### 国内研究现状

基于计算机视觉的智能视频监控系统在国内得到了广泛的推广，863开展了一系列重大项目研究，“平安城市”的建设在一些试点城市开始陆续展开，2005年，公安部在全国开展城市报警与监控系统建设“3111”试点工程，此举有力地推动了“平安城市”的建设步伐。[近年来，上海、深圳、南京、武汉、成都、杭州、宁波、佛山、昆山等城市相继推出了“智慧城市”（Smart City）的发展战略，努力争夺先发优势。](file:///C:\Users\tong\Desktop\paperpass%E6%A3%80%E6%B5%8B%E7%BB%93%E6%9E%9C\htmls\sentence_detail\14.htm)[IBM给出“智慧城市”的定义为：运用信息和通信技术手段感测、分析、整合城市运行核心系统的各项关键信息，从而对包括民生、环保、公共安全、城市服务、工商业活动在内的各种需求做出智能响应。](file:///C:\Users\tong\Desktop\paperpass%E6%A3%80%E6%B5%8B%E7%BB%93%E6%9E%9C\htmls\sentence_detail\15.htm)

国内的许多高等院校和科研机构在智能视频监控领域投入了相当程度的研究，例如:清华大学、上海交通大学、中国科学院自动化研究所等。其中，自动化研究所生物识别与安全技术研究中心研究开发除了CBSR（Center for Biometrics and Security Research）智能视频监控系统，主要功能包括人和车辆的多目标检测与跟踪，并进一步对目标异常行为进行识别与报警，以及人群和交通流量评估、车辆计数和拥堵报警等。为了促进国内智能视频监控的发展，中国科学院自动化研究所在2002年和2003年分别举办了第一届和第二届全国智能视觉监控会议。《自动化学报》在2003年5月出版了一期视觉监控专刊。《计算机学报》、《软件学报》和《控制与决策》等一些杂志也对智能监控系统中一些关键技术进行刊载。

### 国外研究现状

国外对智能监控系统的研究很早就开始了，无论是在国家层面、重要的研究机构还是科技公司都对该领域投入了大量的资金开发相应的智能视频监控系统，而系统的应用场景涉及了从未来城市、公共区域甚至于战场等。

美国国防高级研究项目署DARPA（Defense Advanced Research Projects Agency）在2000年资助了视频监控项目（Video Surveillance and Monitoring, VSAM）的主要目标是利用视频理解、网络通信、多种传感器融合等技术实现对未来城市、战场等你的自动监控。CAVIAR（Context Aware Vision using Image-based Active Recognition）是在2002到2005年间欧盟的资助下开展的项目，主要用于研究城市闹市区的监控问题和商业顾客行为分析。此外，日本开展了用于公共区域及智能小区的视觉监控的计划（The Cooperative Distributed Vision Project，CDVP）。

MIT媒体实验室开发了实时人体跟踪与行为理解系统Pfinder，它基于颜色和形状特征建立多类别统计模型，能够在自由试点条件下构造人体头和手等部位的二维表示形式。IBM研究中心开发的IBM S3智能视频系统用于满足机场的安全监控需求，S3系统是一个开放框架的智能视频监控，可以使各种处理和分析模块更方便的嵌入到S3系统中。

当前，国际上的许多图像处理类权威杂志和重要的计算机视觉学术会议都将智能视频监控作为主题内容之一，为该领域研究人员提供了广泛的交流机会，而拥挤场景中的异常检测也作为智能视频监控中的重要一部分得到了很大的关注。权威杂志包括模式分析和机器智能((PAMI: IEEE Transactions on Pattern Analysis and Machine Intelligence，图像和视觉计算(IVC: Image and Vision Computing)等，重要的学术会议如计算机视觉和模式识别会议 (CVPR:IEEE Computer Society Conference on Computer Vision and Pattern Recognition，国际计算机视觉会议(ICCV: International Conference on Computer Vision，欧洲计算机视觉会议(ECCV: European Conference on Computer Vision)视觉监控会议(IWVS: IEEE International Workshop on Visual Surveillance)等。

## 主要研究内容及论文结构

异常检测也称离群点检测，通常指基于给定的样本空间来检测不属于已建立的正常行为的模式，它有着广泛的应用，如入侵检测、错误检测、欺诈检测及传感器网络中的事件检测等。本课题主要研究基于视频的拥挤场景中的异常事件检测。异常事件可以定义为区别于正常行为的反常行为，是一种不规律事件，具有偶发性、反常性等特点。通常，异常事件可以分为两类：局部异常事件（Local Abnormal Event, LAE）和全局异常事件（Global Abnormal Event, GAE）。LAE是指个别行为区别与它的相邻行为，而GAE是指整个场景的群体异常行为，由于异常状况行人突然散开的情形等。

一般的，基于特征的建模过程和视频序列中的异常检测方法如图 1‑1所示。首先，对于给定的训练视频序列提取特征并建立特征描述符，然后就是学习得到事件模型，用同样的方法对测试视频序列提取特征，并根据学习得到的事件模型进行分类，判断是否为异常事件。

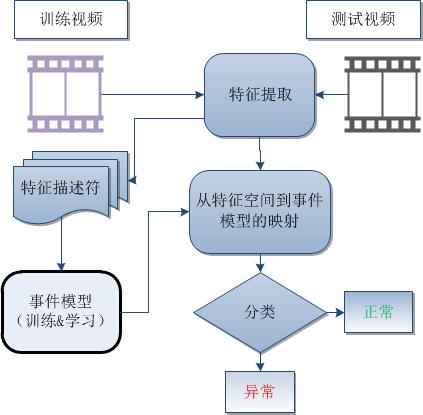


图 1‑1 基于视频的异常检测框架

本课题的主要研究内容包括两部分：基于时空MRF的视频异常检测方法和基于稀疏表示的视频异常检测方法。

第一部分主要研究通过MPPCA学习运动模式，并建立视频流的时空马尔可夫随机场（Space-Time MRF）模型来检测视频中出现的异常状态。视频中的每帧被分割成网格，每个格子对应于MRF图模型中的节点并提取相应的光流特征，同时根据时空关系建立相邻节点的链接。这种模型已经在一些低层的视频处理中得到应用，如立体匹配[14]和图像去噪[4]。光流是一种低层的运动特征表述，这里采用混合概率主成分分析（MPPCA）方法对节点处的光流建立一个概率模型，用于学习节点处活动的状态（正常或异常）模式。基于这种学习模式和建立的MRF图，计算局部节点正常度的最大后验概率，作为节点状态判断的依据，并且随着新观察值的进入，整个模型可以进行自更新。

第二部分将重点介绍稀疏表示和字典学习算法，以及如何通过运动特征的稀疏表示建立相应的稀疏表示代价函数，以此推断视频中的异常事件。同时，根据一种快速的稀疏近似算法对检测算法进行改进，可以有效地提高检测的速度。

本课题结构如下：第2章主要研究如何使用MPPCA学习运动模式，并在时空MRF中推断视频中的异常；第3章将介绍稀疏表示和字典学习的方法；第4章则主要研究基于稀疏表示代价算法实现的视频异常检测框架，并融合快速稀疏近似算法提高检测速度；最后，第5章将介绍系统的整体设计和实验结果的对比情况。

# 基于时空MRF的视频异常检测



## 引言

视频中的异常检测研究由于在实际的视频监控系统中有很好的实际意义，因此引起了相当大的兴趣和关注。但异常检测仍面临着很多技术性的挑战，毕竟很难智能的去定义视频中的异常事件，同时也存在很多不可预见性的情况。

本章主要研究一种构建时空MRF模型检测视频中异常的方法。对于低层的光流特征，采用混合概率主成分分析（Mixture of Probabilistic Principal Component Analyzers, MPPCA）的方法进行学习，从而得到原子的运动模式，它可以更好的表达运动状态。然后使用学习得到的运动模式，构建时空MRF能量函数并计算MRF中的相应参数，接下来就是MRF能量最小化的问题，即通过在MRF图模型中应用置信传播推理得出图模型中的各节点的二标号问题，最后通过标号可推断节点是否异常。

## 早期视觉问题与MRF相关理论

### 早期视觉问题

这里首先介绍一个计算机视觉的应用方法，在文献[18]中主要讨论有效置信传播应用在早期视觉的情况，其中早期视觉包括立体视觉（Stereo），光流（Optical Flow）以及图形修复（Image Restoration）等问题。应用的方法是将问题构建成MRF的模型，即将问题转化为在MRF 模型中计算能量最小化（Energy Minimization）问题。由于在MRF中，能量最小化问题是NP难（NP-hard），因此应用置信传播的演算法做估算。

许多早期视觉问题需要建立一些空间变量（如强度或差异），这些变量倾向于分段光滑。对于在某个有限集合L中每个像素点p需要被指定标号的问题，在运动和立体视觉领域标号代表差异，而在图像修复中则表示强度。因此，研究目标是为每个像素指定一个标号fp，这里f既是分段光滑又与观察数据保持一致。这些视觉问题自然地变成能量最小化问题。在这个框架下，最小化能力函数如下：

(2-1)

这里衡量f不是分段光滑的程度；衡量f与观察数据的差异。目前，针对不同应用已提出了许多不同的能量函数。典型的形式有：

(2-2)

其中衡量在已知的观察数据下标号符合像素p的程度。例如，在图像修复中，通常为，是像素p处观察到强度值。

### MRF与图模型表示

马尔可夫随机场（Markov Random Field, MRF）在视觉中的应用可以等价于一个标号问题，也就是根据最大后验概率（Maximum A Posteriori, MAP）推理得出MRF图模型中每个节点的标号。与其他方法比较，MRF主要有两点优势，其一是它提供了一种原则性方法来对先验知识建模，其二是它可以很容易用定量的方法描述上下文信息。

在实现MRF的推理之前，首先必须解决基于图结构建立相应的MRF图模型和参数学习问题。图的创建一般是对问题本身的建模，在图像修复（Image Restoration）和图像分割（Image Segmentation）中经常会处理图的创建问题，而图的创建其实本身就是对问题的一种建模。在图模型中，四邻接（4-neighborhood）或八邻接（8-neighborhood）的成对模型经常会被使用，四或八相邻的像素中间通过边进行连接，从而构成典型的成对马尔可夫模型。当要加入高阶（>=3)的势能时，等同于引入了更多的约束条件，例如：连通性约束和非基督分类结果的约束等。

MRF推理问题也就是求解能量函数最小化的问题。对于经典的只有一元和二元势能的MRF模型，图割法已经能够在线性时间内进行求解，如果加入更高阶的势能,虽然问题本身可能变成了NP-hard，仍然有很多近似算法，比如循环置信传播(LBP)、权重树消息传递(TRW)等，当然其实问题的实质都是能量最小化。在下一节将详细介绍两种能量最小化方法：图割法和循环置信传播。

关于MRF图模型的表示，首先应该知道图模型（Graphical Model）是用图来表示概率分布的一类技术的总称。通过引入图模型，可以建立起概率分布中的条件独立的表达式，这样做的优点是，它简化了边缘化概率分布的计算。这里所说的边缘化，就是在给定m个变量的概率分布中，计算其中n个变量的概率分布，且满足n小于m。

图模型根据图结构可以分为两类，即有向图模型和无向图模型。典型的有向图模型如贝叶斯网络，而马尔可夫网络是一种无向图模型。在讨论图模型时主要解决三个问题：图模型的表示、图模型的推理和图模型的学习。图模型的表示就是解决一个图模型应该是什么的数据结构和逻辑结构；在已知图模型如何表示的情况下，图模型的推理要解决如何计算一个查询的概率，例如已知一些观察节点，去求其它未知节点的概率；最后图模型还要处理对图的结构和图的参数的学习，以及采取什么样的学习策略。

下面主要讨论无向图模型的表示，即马尔可夫随机场图模型的表示。无向图模型为在表示一个概率分布时，需要在图表示中对变量之间的条件独立进行编码，进而以因子乘积的形式对概率分布进行表示。无向图模型与有向图模型由于基于的图结构不同，所以它们的表示也不相同，无向图模型的建立是基于无向图，而有向图模型的建立是基于有向图。

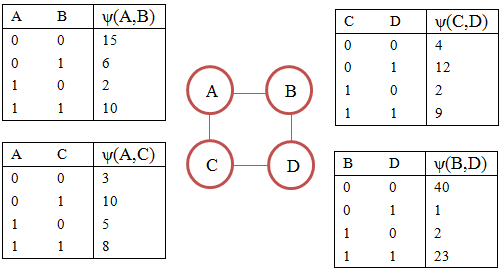


图 2‑1 无向图模型

如图 2‑1所示，是一个无向图模型的完整表示，中间的图形是它的拓扑结构，其他表格数据是它的参数。无向图模型是以最大团和定义在团上的势能函数为核心进行表示的，如上图所示，它有四个团：AC 、AB、 BD和CD。那么就需要在四个团上定义相应的势能函数，如图中表格所示，需要注意的是势能函数必须为正。那么这个无向图模型表示的概率分布是：P(A,B,C,D)= (1/N)\* ψ(A,C)\* ψ(A,B)\* ψ(C,D)\* ψ(B,D)。其中，N是一个归一化因子，由于要求概率在[0,1]这样的区间内，所以需要对势能函数做归一化操作；而ψ是表示相应的势能函数。这样一个无向图模型表示的概率分布可以形式化地表示为：，其中*Ci*表示的第*i*个团。

### MRF能量最小化方法

1. 图割法（Graph Cuts）

能量最小化主要的难点在于巨大的计算代价，这些能量函数通常有许多局部最小值（非凸的），而且标记空间维度可能会很大。由于计算全局最小无效，学者们更倾向于计算局部最小，可是通常一个局部最小值可能会偏离最优值很多。典型的使用标准移动的局部方法是条件迭代方法（Iterated Conditional Modes, ICM），它是一个求局部最小的贪婪算法。而对于一些多维度能量函数，图割法可以用于寻找全局最小值,它在图像分割中已经得到广泛的应用。

问题如公式（2-1）所示：

(2-3)

其中P是图像中的像素集合，L是标号集合，N是四链接网格图中的边，代表指定的labeling，对应于像素p∈P，表示对某个像素P我们指定的标号。对应于图模型中非连续消耗，代表对两个相邻的像素p和q来说，指定不同标号所需的消耗。对应于图模型中的数据消耗，是指定给像素的p所需的消耗。上式代表由标号产生的能量函数，代表quality的结果。因此我们的问题是对应这样定义的MRF，估计某种可以产生最小的能量的标号。而在具体的实际问题中，集合P可以是对图像进行划分后的所有分块的集合。

如果V满足如下条件：

则在标号L空间中，V称为公制；如果V只满足a和b，则称为半公制。

一般地，标号*f*是一个能量E的局部最小值，如果满足如下条件：

(2-4)

在离散标号情况下，*f*相邻的标号位于*f*的单个移动内。在某一时刻只有一个像素改变它的标号，这通常称为标准移动。下面将介绍两种允许更大数量的像素同时改变它们的标号的移动方法：和。

任何标号f可以表示为像素的一个划分，是分配了标号l像素的子集。因为f与P是一对一的关系，所以可以进行概念上的互换。给定一对标号，从划分P到新划分的移动称为。这就说明了P与仅有的差别，一些在P中标记为的像素在中被标记为，相反，一些在P中标记为的像素在中被标记为。给定一个标号，从划分P到新划分的移动就称为，且满足对于任意的标号有和。换句话说，移动允许任意子集将标号改为。

表 2‑1 Swap算法

|  |
| --- |
| 1. 初始标记任意的标号f 2. 赋值success为 0 3. 循环每对标号   3.1通过f的 swap在f’中寻找  3.2如果, 令f := 且success := 1   1. 如果success值为1 转向步骤2 2. 返回 f |

如上表 2‑1 和表 2‑2所示，算法swap-move 和 expansion-move[19]在结构上相近。给定一个f，算法共需要通过指数级的swap和expansion移动。在算法中，步骤3.1-3.1为迭代过程，步骤2-4位循环过程。在每个循环里，算法为每个标号或成对标号执行一次迭代。如果在任意迭代中找到了一个严格更好的标号，则循环成功。算法在第一个失败循环后停止，因为此时标号没有进一步提高的可能。在swap算法中，一个循环要花费|L|2次迭代，而在expansion算法中为|L|，这两个算法都能保证在有限的循环内终止。

表 2‑2 Expansion算法

|  |
| --- |
| 1. 初始标记任意的标号f 2. 赋值success为 0 3. 循环每个标号   3.1通过f的-expansion在f’中寻找  3.2如果, 令f := 且success := 1   1. 如果success值为1 转向步骤2 2. 返回 f |

1. 最大乘循环置信传播（Max-Product Loopy Belief Propagation, LBP）

在马尔可夫随机场中，BP是一种有效的推理方法。特别的，最大乘算法可以计算能量函数近似最小消耗的标号。最大乘BP算法建立了在四链接图中传递消息的有效机制。每个消息都是由标号数确定的一个向量。令代表在t时刻节点p传递给相邻节点q的消息。初始时刻设置为0，那么新消息的迭代计算公式如公式（2-5）：

(2-5)

N( p) \ q 代表p 附近除了q 的节点。在经过T次迭代后，每个节点的置信向量为

(2-6)

最后每个节点都可算出最小的来决定所要的标号。

这个算法的时间消耗共需要O(nk2T)，包括对T次每次需要O(n)时间的迭代，以及每次计算消息要花O(k2)。文献[18]中介绍一种改进算法，经过加速，可以降为O(nk)的时间。改善的方法有三个部分：

(1).消息计算的改进:

首先讨论的是在相邻像素间的消息传递代价，对一般的低层视觉问题代价函数基本上是两个标号和的差值。因此，如果我们把这个消息传递代价设定成，则消息传递代价函数可改写成，其中，如此对k层迭代计算消息的时间仅需要O(k)。

接着考虑一个截尾线性模型，即，这里的d即代表线性增加的代价上限。这里考虑两种情形，如果相差超过定的代价上限，情況就像上面考虑的情形，花的时间因此只需O(k)。另外要是，计算消息的公式就变成，的关系图如图 2‑2 消息值的关系图所示：

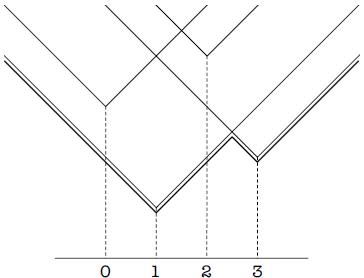


图 2‑2 消息值的关系图

其中各条不同的Ｖ形线是因为所选的值不同而产生，所以需要计算的是在选不同的时对应到消息的最小值，可以采用2路通过算法进行计算，如下所示：

For from 1 to k-1:

(向前)，

For from k-2 to 0:

(向后)。

举例来说，一开始最小的消息值为V形顶点的值，在上图中值从0-3 可看成(3,1,4,2)，在向前的过程后成为(3,1,2,2)，再经过向后过程变成(2,2,2,2)，即上图粗线处。这样进行2路通过后可以找出对应到最小消息和q的层值。这里用2路的方法是由于代价定为线性的，因此只要参考左右的最小值就可决定自己本身的最小值，而演算法所需的时间为O(k)。

(2). 第二部分是利用双向图置信传播来加速，这里将计算时的格子分为两组(A,B)，而传递的消息就只有两种: A -> B 和B -> A，因此在每一次消息的传递(更新)时，就只需要更新某一种的消息就可以。在设计时，就分別记录两种不同的消息，再随迭代的过程更新。如此，大约可以减少一半计算的时间。

(3). 多尺度置信传播: 这部分是利用减少消息传递的迭代技术来加速。采用的做法是由粗到细（Coarse-to-fine），将计算的范围由大至小切细成不同大小的子块，计算子块中更新的消息，因此就可以解决每个像素都要传递消息的计算量。

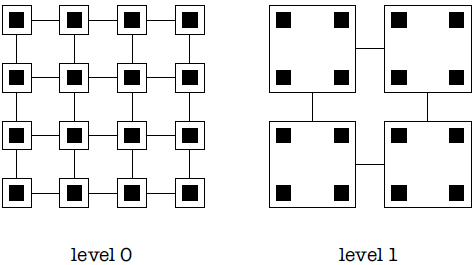


图 2‑3 多尺度置信传播

在计算由大到小的消息时，则将大块内每个小块的消息指定为大的消息。另外在计算时需要的数据消耗，如此的话则每个切出的子块可以有多个造成低代价的层值，而不会因为合并计算而和单一像素计算的结果随迭代有不趋近的走向。另外在计算的过程中，会比单一像素的计算更新的结果收敛至相近的结果。

## MPPCA学习局部状态的运动模式

### 光流特征提取

光流是空间运动物体在观测成像平面上的像素运动的“瞬时速度”，利用图像序列中的像素强度数据的时域变化和相关性来确定各自像素位置的运动，进而从图片序列中近似得到不能直接得到的运动场。计算光流的两种主要方法是稠密光流和稀疏光流：Horm-Schunck[5]方法计算的就是稠密光流的速度场，即将图像中的每个像素都与速度关联；金字塔Lucas-Kanade[8]光流是一种流行的稀疏光流计算方法，它可以计算基于块的多尺度光流值。

这里采用金字塔LK光流方法计算视频中每帧的光流特征。对于每一帧分割成m乘n个局部区域（本课题中m=6，n=8），每个局部区域（对应于MRF图中的一个节点）的原子活动构建一个特征描述符，又分成u乘v个子区域（u=2，v=2），统计子区域内所有像素点的光流生成一个9维向量（8个方向，1个速度）。这样局部区域将得到9uv维的活动描述符，其中子区域的划分（u和v的值）取决于获取运动信息的细度。

### 混合概率主成分分析

主成分分析（Principal Component Analysis，PCA）是一种统计学上常用来降低问题维度的方法。如图图 2‑4 PCA示意图所示，对于空间中的点，寻找直线L使得所有空间中的点到直线的垂直距离和最小，这就是PCA所处理的问题。

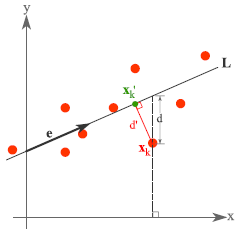


图 2‑4 PCA示意图

PPCA[12]（Probabilistic PCA）是将PCA用概率的方式表示。简单来说，就是给定一个观察序列（高维度），PPCA模型建立潜在的可变序列（低维度），同时根据最大似然估计准则找到最优的参数。它的优点是避免了带有丢失数据的PCA处理情况，也可以混合多组PPCA来表示更复杂的情况，即混合概率主成分分析（Mixture of Probabilistic Principal Component Analyzers, MPPCA[15]）。

MPPCA关联了概率模型和PCA，这样通过组合多个PCA模型可以更好的对复杂数据建模，而所有模型参数可通过最大似然法确定。对于这样一个混合模型，观察数据的log似然如下：

(2-7)

其中，代表一个单独的PPCA模型而是相应的混合系数（），并且每个组件关联一个独立的均值向量及参数和。

### 学习局部状态的运动模式

在初始训练视频中，首先提取每一帧所有局部区域的描述符（光流特征）。用于学习局部状态的MPPCA模型定义如下：

(2-8)

其中，t：局部区域的描述符，：第i个PPCA组件的系数； ：第i个PPCA组件的概率密度函数；：第i个PPCA组件的协方差矩阵；：第i个PPCA组件的均值向量。这样，就可以把局部区域的描述符（特征）建立起一个混合的概率模型，而这个模型类似于高斯混合概率模型GMM，各参数可以通过期望最大化方法计算得到。

这里为所有的局部区域定义一个更通用的MPPCA模型而不是单一模型，是因为在初始视频中，一些局部区域并没有足够的样本来保持EM算法中合并的稳定性，也就是说大多数观察值在一些局部区域是自由运动状态。因此，这种混合概率模型可以更好的表示局部区域的运动模式。

根据上面的MPPCA学习模型，主要计算节点处两类直方图的分布情况，分别是节点的频率直方图和链接的同现直方图。频率直方图代表节点处MPPCA中每个PPCA组件被观察到的频率，而同现直方图代表相邻节点处两个PPCA组件同时被观察到的频率。定义节点i处的频率直方图为H­i，节点i和j处的同现直方图为Hi,j，计算公式如（2-9）和（2-10）：

(2-9)

(2-10)

其中， ：频率直方图 的第*l*个箱； ：同现直方图 的第（*l*, *m*）个箱；而和分别是基于描述符ti,k和tj,k在第k帧节点i和j处的后验概率，定义如公式（2-11）和（2-12）：

(2-11)

(2-12)

因此，累计在节点i处观察到的所有活动描述符为PPCA组件l的后验概率，表达了发生在视频中相应区域低层运动类型的可能性。相似的，表示在节点i和j处PPCA组件为l和m的可能行，进一步说明相邻区域的相关性。

## 构建运动模式的时空MRF模型

### 时空MRF模型的能量函数

根据视频中最新的一帧和历史最近的帧序列中选取一个固定长度构建时空MRF模型。在时空MRF中定义了两个函数：节点置信函数与成对势能函数。他们的计算依据于上面定义的MPPCA，然后通过在MRF图上的置信推理产生最大后验概率（MAP）类别，判断节点的状态为正常或异常。这里时空MRF的能量函数定义如下：

(2-13)

其中，是节点置信函数的权值，是节点置信度函数，是成对势能函数，表示活动状态的类别（意味着节点i是异常，反之则正常）。

节点置信函数本身包含两个术语：频率术语nf和适合度术语ns。频率术语用于衡量一个状态模式与给定节点处的当前状态描述符相近的频率，适合度术语用于评价存在的MPPCA模型生成当前状态描述符的可能性。

频率术语强调每个节点PPCA组件对的约束关系，简单的说，如果节点i处的状态描述符被检测属于已被观察到的高频PPCA组件中的一个，那么的值变得更高（反之，变低），即。频率术语计算公式如下：

(2-14)

其中，是节点i的归一化频率直方图，是给定描述符的组件c的后验概率。Tk是一个控制对异常敏感度的转移函数，定义如下：

(2-15)

可以看出，降低参数k的值会导致较少的异常节点被检测。总之，描述了频率直方图Hi和MPPCA组件的概率分布关于节点i当前状态描述符的归一化关系。

适合度术语反应当前MPPCA模型解释新状态描述符的好坏，即是否适合。定义如下：

(2-16)

其中，是状态描述符t和MPPCA组件c的马氏距离（Mahalanobis Distance），并归一化。Fc是所有之前观察值在PPCA组件c处累积距离分布。

最后，完整的节点证据函数如下：

(2-17)

其中，为权值，它的值总是大于0.5。

成对势能函数由同现频率术语和光滑度术语组成。同现频率术语定义如下：

(2-18)

它衡量相邻节点同时发生两个动作的正常性。

平滑度术语强调相邻节点运动相似的平滑程度，其计算方法为：

(2-19)

因此，根据公式（2-18）和（2-19）可得完整的成对势能函数为：

(2-20)

其中，是一个衡量光滑度的权值常量。

给定了公式（2-17）和（2-20）在节点和链接处的MRF参数，即可通过MAP最大化式（2-13）中定义的函数，并使用循环的置信传播（传递最大的和信息）来推断节点的状态。

### 时空MRF模型的自更新

在建立了时空MRF模型，可以使用新视频帧的活动描述符对参数进行不断地更新。根据更新的MPPCA参数，所有的直方图参数和MRF参数也会相应的调整。

为了使用新视频帧的活动描述符更新MPPCA的参数，需要先为描述符选择最有可能的PPCA组件，然后更新组建Cmax的协方差矩阵C和均值向量µ。

混和系数的更新如下所示：

其中，和分别为在t和t+1的时间里被观察到的活动描述符的总数，和分布是在t和t+1的时间里属于组件i的活动描述符的总数，为组件i在t+1时刻更新的混合系数。

这个更新算法实现比较简单。但需注意一个必要的假定：一旦新的观察值（描述符）加入到模型中时，每个组建的后验概率将不变。

## 本章小结

本章介绍了通过MPPCA学习运动模式并建立视频流的时空MRF模型来检测视频中出现的异常状态的方法。采用MPPCA方法对低层的光流特征进行学习，从而得到原子的运动模式，这种模式可以更好的表达运动状态。正是基于这种运动模式，构建时空MRF，并通过解决MRF能量函数最小化问题进行异常的推断。建立时空MRF后，首先要计算MRF中的相应参数，然后就是MRF能量最小化的问题，本章介绍了两种最小化问题的解决方法：图割法和最大乘循环置信传播算法，而本课题采用后者进行计算。这样，通过在MRF图模型中应用置信传播最小化能量函数后，可得出图模型中的各节点的标号，即由此推断节点是否异常。

# 字典学习与稀疏表示



## 引言

稀疏表示在最近的字典学习和压缩感知的研究中扮演了重要的角色。对于给定的一个基向量集合（如一个字典），找一个信号的稀疏表示经常等同于带有L0范式和L1范式的最优化问题[26][28]，并最终转化为解决一个欠定的线性系统。然后，每个样本表示成为基向量的一个稀疏线性组合。我们都知道，解决L0范式最小化问题是一个NP难问题并且数值上不稳定。对于提出的一些近似L0范式解决方法如匹配追踪[27]和OMP[29]等，尽管这些方法相对简单有效，但它们都不是最优的。最近的关于稀疏编码的研究表明，在某种假设的情况下，解决L0范式最小化问题等价于L1范式最小化问题，而后者可由顶点最优计算。

在计算机视觉、机器学习、信号处理及计算机图形学等领域中，会有许多这样的问题，一个简单且有效的方法是假设同一个类中的样本可以依据原型建模。原型可以是样本本身，或者从样本中学习得到的结果（如：特征向量和均值）。本课题中假定某一个类的样本可以根据同一类的原型子集建模。

在原型学习算法中，最优方向法（Method of Optimal Directions ,MOD）和K-SVD算法引起了极大的兴趣，因为它们可以通过一个字典或者原型的稀疏组合表示每个样本。假定已经得到一个学习来的字典，就可以使用字典中的原型通过稀疏表示近似得到基向量，其中，原始的稀疏表示问题可以简化为一个L1范式的约束问题。下面将分别介绍稀疏表示算法和MOD与K-SVD两种字典学习算法。

## 稀疏表示

在许多问题里面，经常会给定一个已标记类别的样本集，并通过学习这个样本集来准确地推断未知样本的类别。假设我们从K个类别中的第i个类别得到ni个样本，并为每一个类i定义一个矩阵：

(3-1)

其中，表示类i中第j个样本。然后，合并所有K个类别的全部样本为一个矩阵：

(3-2)

这里，N是所有类别的样本总数。给定了类别i的充足样本数，如果待观察样本y属于类别i，那么样本y可以通过样本的线性组合近似得到：

(3-3)

其中，表示样本j在重建样本y时的加权贡献。

然而在大多数情况下，我们并不知道样本y属于哪个类。因此，我们可以重写y的线性表示：

(3-4)

这里，是一个稀疏系数向量，通常是一个稠密矩阵，如图 3‑1中（a）所示。

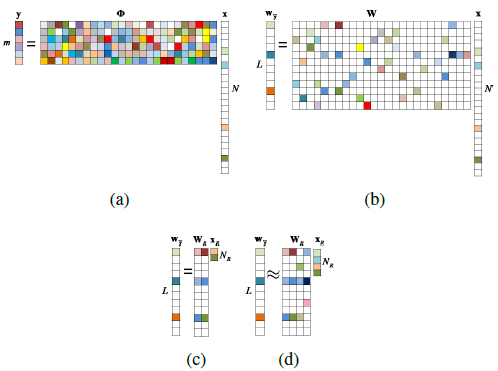


图 3‑1 稀疏表示算法示意图

对于上面的公式（3-4），每个观察样本y可以根据相应系数向量x表示，当然这需要先解决线性形态。如果出现观察数据y的维度大于所有样本数，那么唯一的解决办法是求解超定系统。然而，在大多数应用中，这个线性系统通常是欠定的，最终会导致无限的解决这个逆问题。因此，规则约束对得到有效的解决办法至关重要。

因为假定一个观察样本属于一个特定类别，它可以被同一类中的其它样本很好的表示。这个属性以及被广泛的探究，如局部线性嵌入、图像聚类及人脸识别[33]等。在每个类都有充足样本的情况下，我们期望系数向量x尽可能稀疏，最好只有一小部分为非零值。通过稀疏约束，我们寻求观察样本y的表示为：

(3-5)

这里， 统计非零值的数量。然而，解决欠定系统的L0范式最小化问题是一个NP难问题并且数值上不稳定。

最近，稀疏表示和压缩感知[33][34]的研究表明，如果x的解决方案是足够稀疏的，那么最稀疏的解决方法可以通过L1范式最小化恢复：

(3-6)

其中，L1范式计算x中所有数值的绝对权值和。上式中，存在一定的噪声，更好的形式如下：

(3-7)

其中是容错度。上面两个公式（3-6）和（3-7）的解问题是一个凸优化问题，可以分别由线性规划和二次锥规划来求解。

通常直接在图像的原始空间上操作是复杂的，因为数据的维度过高带来了极大的计算开销。目前有多种特征提取方法可以将原始的数据映射到低维度的向量空间中，假定存在一个转换矩阵T使得原始数据映射到低维度的向量空间中，且矩阵, d<<m, 在公式（3-4）两边分别乘上T可得到：

(3-8)

其中，是观察样本y的特征向量，包含所有样本的特征向量。因为线性等式系统是欠定的，并且期望x的解足够稀疏，那么x的求解可以通过解决下面L1范式最小化问题近似得到：

(3-9)

使用Ll范式替换L0范式后，原优化问题也就变成了一个凸优化问题，可以通过线性优化的问题来解决。

一个凸优化（Convex Optimization）问题的标准形式如公式（3-10）所示：

(3-10)

这里的f0到fm是凸函数。在这个凸优化问题中，有下面几个约束条件：目标函数必须是凸的；不等约束函数必须是凸的；不等约束函数必须是可放射的。这里一个重要的特性就是：凸优化问题的可行集合必须是凸的。所以凸优化就是在凸集上寻找凸函数的全局最值的过程。

## 字典学习

### 最优方向法

最优方向法（Method of Optimal Directions, MOD）是由Engan等[23][24]提出的一个实用的字典训练算法，它采取了一种简单有效的字典更新方式。假定每个样本的稀疏编码已知，误差为。那么整个表示的均值平方误差如下：

(3-11)

其中，Y是所有样本的矩阵（），X为表示稀疏矩阵（）。表达式为Frobenius范数，定义为。

如果固定X的值，就可以找到一个D的更新，使得上面的误差值最小。对上面的公式（3-11）求导就会得到，推出

(3-12)

这就是字典D的更新公式，这个更新可以通过固定X很好的实现。

### K-SVD算法

K-SVD算法[25]是在MOD的基础上做了顺序更新列的优化。K-SVD算法很灵活，可以很好的和任何追踪算法联合使用。它简单有效，可以看作是K-Means的泛化。下面先简单介绍下K-Means算法。

一个包含K个密码词（codewords）的密码书（codebook）可以通过最近邻方式表示一个向量（信号）族。这样在选定的密码词周围，可以有效的实现信号压缩或描述。通过K-Means算法可以训练得到矢量化密码词的字典。定义密码书矩阵为，而密码词即列向量。在C给定的情况下，每个信号由离它最近的密码词表示（根据L2范式距）。记，其中是来自平凡基的一个向量，在平凡基中，除了第j个位置外所有向量都为0向量。索引j的选择满足如下：

(3-13)

这是稀疏编码的一个极端情况，即只有一个原子可以构建，并且系数强制为1。表示每个的均值平方误差（Mean Square Error, MSE）为，于是全部MSE的和为：

(3-14)

这样，矢量化训练问题变为找到一个密码书C使得误差E最小，且满足X的有限结构，X的列向量取自平凡基，即

(3-15)

K-Means算法采用迭代方式来求解最优的密码书，每次迭代有两个步骤：稀疏编码与密码书的更新，详细步骤如表 3‑1所示。

表 3‑1 K-Means算法求解

|  |
| --- |
| 任务：通过解决如下问题，并根据最近邻寻找最优的字典来表示数据样本  初始化：密码书矩阵，令J=1  循环直到合并结束(除了使用中断条件)：   * 稀疏编码阶段：划分训练样本Y为K个集合   每个集合拥有与列最相近的样本标记   * 密码书更新阶段：对每一列向量使用下面的公式进行更新 * 令J = J+1 |

稀疏表示问题可视为对矢量化目标（3-14）的泛化，只是在稀疏表示中，每个信号可以由多个密码词的线性组合来表示，密码词被称为字典元素。因此，因子系数将不只一个非0实例，从而相应的问题变成在样本集Y中寻找稀疏表示的最可能字典，满足下面的约束条件：

(3-16)

对于这样一个问题，首先，固定D去确定最优的稀疏矩阵X，可通过近似追踪算法如OMP等找到最优的X，这步也就是稀疏编码；然后就是去找一个更好的字典，在这个过程中，一次只更新一列，即先固定除了当前列外的其它列，再更新得到一个先例，并且想要系数可以更好的减小MSE。通过更新D同时改变相应的系数，从这个意义上说，K-SVD是对K-means算法的泛化。一次只更新一列的过程可以由奇异值分解（SVD）直接求解，而且更新字典列的同时改变系数可以加速合并过程，因为子序列的合并将基于更新后更相关的系数，下面将详细的介绍K-SVD算法。

先考虑稀疏编码阶段，我们假定D是确定的，然后考虑相应的优化问题，惩罚表达式可以重写为：

(3-17)

这样，相应的问题可以分解为如下N个不同问题，并有追踪算法解决，如果越小，这个解会越接近理想值。

(3-18)

下面考虑字典的更新阶段，

(3-19)

关于用SVD得到与应该注意下面这个问题。SVD可以找到最近的rank-1矩阵，这个矩阵近似于，它可以有效地减小上面公式（3-19）中的误差。然而这一步是错的，因为新的向量可能是稠密而非稀疏的，因为在更新时，没有强制进行稀疏约束。对于这个问题的解决比较简单，定义是指向使用了原子的样本的一组索引：

(3-20)

定义是的矩阵，在处有实值而其它都为0，进行相乘得到，通过丢弃0值可以缩减行向量，从而行向量的长度变为。相似的，，。这样上面的公式（3-19）可以变换为：

(3-21)

这样它就可以直接用SVD进行求解。SVD对矩阵进行分解为，定义的解为U的第1列而系数向量为V的第1列。要注意的是矩阵D的列向量需要进行标准化。完整的K-SVD算法如下表所示。

表 3‑2 完整的K-SVD算法

|  |
| --- |
| 任务：通过解决如下问题，寻找最优的字典来表示数据样本，如稀疏组合  初始化：密码书矩阵且带有标准化的列，令J=1  循环直到合并结束(除了使用中断条件)：   * 稀疏编码阶段：通过解决下面的问题，对于每个样本yi，用任何一种追踪算法计算稀疏向量xi * 密码书更新阶段：在中每一列（k=1,2,…,K）, 通过定义的一组使用这个原子的样本进行更新，，计算整体的表示误差矩阵, 根据   通过只选择与相关的列来约束，并得到。  应用奇异值分解（SVD），选择被更新的字典列dk作为U的第一列，更新系数向量作为V的第一列，并乘上。   * 令J = J+1. |

## 本章小结

本章主要介绍了稀疏表示和字典学习算法的相关概念和理论。在字典学习算法中，MOD是一种简单实用的字典学习算法，而K-SVD算法是在MOD的基础上做了顺序更新列的优化，更加灵活有效。目前，稀疏表示已广泛应用于信号处理和压缩感知领域，下一章会重点介绍如何基于稀疏表示实现视频的异常事件检测。

# 稀疏表示方法检测视频中的异常



## 引言

近年来，视频监控中的许多关键领域取得了很大的进步，如背景建模、目标跟踪、行人检测、动作识别、拥挤计量和交通监控等。异常事件检测作为视频监控中的重要应用，已经引起了极大的兴趣。根据场景的差异，可以把异常检测分为拥挤场景和非拥挤场景两种情况。

对于非拥挤场景的情况，如果前景目标可以很容易的从背景中抽取到，通常会采用基于背景模型的二值特征，如标准化割聚类[39]和3D时空前景掩码特征[36]。由于在非拥挤场景中目标模板可以初始化，基于目标跟踪的方法[1][40][41]也有广泛的研究，通过使用目标位置的不同帧生成目标追踪轨迹进行分析。这些方法在车辆监控中表现不错，但在拥挤场景中则失效，因为无法取得很好的目标轨迹。

在拥挤场景下，会有许多目标或事件同时出现，像地铁站里很难分离每个目标或事件，因此更多有效的方法选择使用局部特征来表示异常事件，包括时空信息、抽取运动特征以及建立光流直方图，而且通常用协方差矩阵来描述上下文信息。这其中包括使用直方图衡量局部光流特征的概率[37]；通过抽取时空梯度[38]来建立高斯模型，并HMM检测异常；通过混合动态纹理[32]对正常拥挤场景建模；社会力模型[30]的提出与应用，首先提取光流特征，然后计算社会力并结合LDA(Latent Dirichlet Allocation)模型进行异常检测。

本章的视频异常检测算法描述如下图 4‑1 样本空间所示，其中每个点代表一个高维特征，绿色和红色的点分别代表正常和异常的测试样本。如果大多数测试样本事件是正常的，那么绿色的点稠密而红色的点稀疏。图中浅蓝色点作为训练特征，通过字典学习后，得到深蓝色的点作为最优的特征来构建字典，点的大小代表其权重，越大那么表示越接近正常。异常检测就是衡量测试样本（绿色和红色的点）在正常样本学习到的字典（深蓝色的点）的稀疏表示代价。

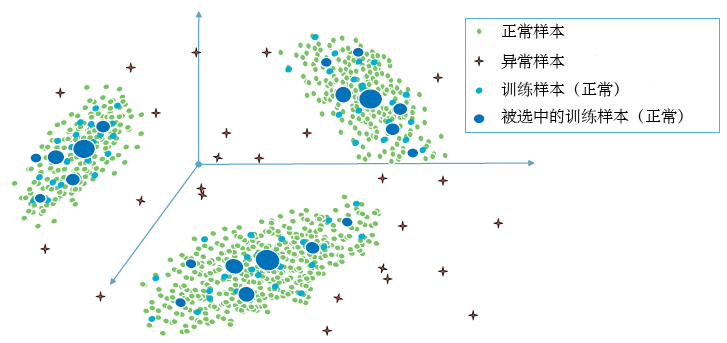


图 4‑1 样本空间

## 基于稀疏表示方法的视频异常检测

对于给定的训练集，N为训练样本总数，代表一个训练样本（d是特征的维度），它可以表示一般的对象，如一个像素、一块图像区域、混合动态纹理、运动描述符等。假设有这样一个测试样本，那么异常事件检测就是设计一个判断函数来决定z是否正常。即

(4-1)

要实现这样的检测，两个关键问题需要被确定：事件描述和异常判断。

对于异常事件描述，在文献[36][39]中采用了基于背景模型的二值特征。一些其他的方法考虑的是时空信息，如光流直方图[37]，时空梯度[38]，社会力[30]，混合动态纹理[32]等。而且，协方差矩阵经常用来描述空间上的关系。

异常判断实际是一分类学习问题，大多数传统的算法[7][30][37][38]倾向于这样一种观点，当测试样本有较低的概率符合训练样本的概率模型时，判定这个测试样本为异常。几种统计模型如：高斯模型、高斯混合模型或者混合概率主成分分析（MPPCA）[7]、隐马尔可夫模型（HMM）[38]、马尔可夫随机场模型（MRF）或者时空马尔可夫随机场模型等。通常的过程是，首先用训练样本Y建立上面提到的某种随机概率模型，然后根据给定的模型计算测试样本z的后验概率：

(4-2)

其中是一个阈值。

高维的特征通常可以更好的表示事件。然而，为了符合一个好的概率模型，需要的训练样本数会随着特征维度按指数级增长，实际中，收集足够多的训练样本也是不现实的。因此，对于很多优秀的方法，仍未解决用高维特征表示事件和模型复杂度之间的问题。例如，本课题中的全局异常事件检测（GAE）方法中，仅有400个训练样本，而样本维度是320。对于这样有限的训练样本，甚至很难鲁棒的适合一个高斯模型。对于这种较少训练样本表示高维的样本的情况，稀疏表示方法可以很好的适用。本课题将通过稀疏表示方法实现异常事件的检测，首先根据训练样本学习得到字典D，这样测试样本z基于这个字典D进行稀疏表示后，可以得到相应的稀疏表示系数，最后建立一个稀疏表示代价函数用于判定样本z是否异常。



### 光流直方图特征提取

基于金字塔LK光流方法计算视频中每帧的光流特征后，把每帧划分成m×n个子区域，每个子区域将构成一个特征单元，并进一步计算它的光流直方图。

光流直方图（Histogram of Optical Flow, HOF）就是根据光流的方向统计相应的光流值，通常对一个方向的光流进行求和计算。选取光流方向时，最简单的就是根据坐标系中的象限划分为四个方向，为了取得更好的特征描述效果可以做更多的划分。如下图所示，对于一个区域中的光流特征，可以划分出八个方向，最后得到这八个方向的光流直方图特征。

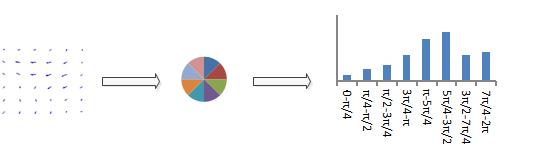


图 4‑2 光流直方图

多尺度光流直方图（Multi-Histogram of Optical Flow, MHOF）的特征表示如下图所示，为了更好的描述特征增加了尺度的划分。可以根据运动量级（光流大小）来设定不同的尺度，如下图所示，设定一个阈值T后，可以分为大尺度（大于T）与小尺度（小于T）两类特征，这样每个单元提取得到一个16箱的直方图特征，其中小尺度使用前面的8个箱，大尺度使用后面的8个箱，这样做可以细化特征值，更好的表示该区域的运动特征。

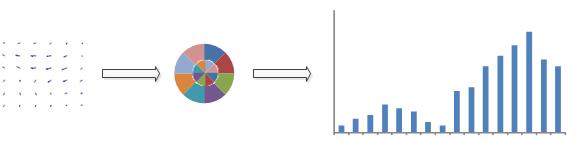


图 4‑3 多尺度光流直方图

为了更好的处理LAE和GAE，可以分别采用下面不同的策略。对于全局异常事件，主要采用图像的整帧特征进行分析；而对于局部异常事件可以选择时空域特征进行检测。考虑全局的异常情况，需要提取图像的整帧特征，即先对一帧的图像进行子区域划分，然后提取出每个子区域的特征向量；在考虑局部异常时，对于某个特定的局部区域可以进行单独的分析，即要分析这个区域的时域特征，也要考虑相邻区域的空间特征，这样融合时空域上的特征可以同时保证时间与空间上关联性。对于两种不同类型的特征域，可以根据异常检测实际需要进行有效的选取。对于得到的特征空间，还应该进一步做归一化操作，便于后面数据的进一步处理，并保证程序运行时收敛加快。

### 视频异常检测框架

在这一节，主要讨论如何根据给定的初始候选特征池学习字典的问题。特征池如，每个列向量代表一个正常特征，最后的目标就是找到一个最优的子集构建字典，其中，这样样本集B可以由很好的构建，并且尽可能的小。最简单的方式就是随机的选取候选列来构建字典，但这明显无法充分利用B中所有的候选，很有可能丢失重要的候选列或者包含了噪音列，这将极大的影响稀疏重建。为了选择最优的候选集构建字典，我们考虑这样的问题：

(4-3)

其中，Frobenius范数定义为，L1范式定义为。然而，这样得到的X的解近似于一个特定矩阵I，它会使上式中的第一个表达式为0且非常稀疏。这就需要保持解的稀疏一致性，对于解中为0的行向量，意味着相应的特征不会用于进行稀疏重建。

因此，改变上式中的L1范式约束，而用L2,1范式进行约束，于是采用下面优化问题选择字典：

(4-4)

其中，为X中的第i行。

因为K-SVD算法灵活有效，可以有效地与追踪算法结合。所以基于上面的理论，本课题主要使用K-SVD字典学习算法对训练样本进行字典学习，从而得到正常样本的字典。

本节会详细介绍如何衡量或决定一个测试样本y是否正常。基于正常样本的训练的得到的字典进行稀疏表示，得到测试样本空间中的正常样本和异常样本的稀疏系数会表现出较大的差异，在正常帧中，样本的稀疏表示系数值幅度较小，而在异常帧中，样本的稀疏系数值会表现出较大的幅度。正是从这一点出发，构建关于稀疏系数的能量函数，以此推断样本是否异常。

使用K-SVD方法对训练集Y进行字典学习得到字典Ф。通过字典Ф对测试样本进行稀疏重建，进而得到重建的稀疏系数矩阵。

(4-5)

对于给定的测试样本y，我们构建这样一个稀疏表示代价函数：

(4-6)

稀疏表示代价值越高意味着稀疏重建过程的消耗越大，那么异常的可能性也就越大，这也正是本法进行异常检测的依据。事实上，这个稀疏表示代价函数也可以映射到贝叶斯决策理论中。从贝叶斯的角度看，正常样本属于较高概率的点，而异常样本则有着较低概率。考虑正常样本最大化后验概率如下：

这里，第1个表达式和第2个表达式是先验分布。这样，就与稀疏表示代价函数保持了一致性，异常样本相应的越小，意味着得到稀疏表示代价越大。本课题的视频异常检测框架如下表所示。

表 4‑1

|  |
| --- |
| 视频异常检测框架 |
| 输入：训练字典，测试样本  输出：样本状态state（normal 或 abnormal）  for i=1,…,N do  //计算重建的稀疏系数矩阵  //计算稀疏表示代价  //基于测试样本的前k帧构建稀疏表示代价的阈值Threshold（已知这k帧正常）  if i == k  Threshold = 1/k \*  end if  //从第k+1帧开始进行检测  if  state = abnormal  else  state = normal  end if  end for |

整个异常检测的过程是，依据选取的前400帧正常样本训练得到的字典Ф和测试样本Y作为输入，样本的状态即是否异常作为输出。对于每个测试样本y，首先基于字典Ф进行稀疏表示，得到稀疏系数矩阵x\*，并由此计算相应的稀疏表示代价。对于如何判定样本的状态，这里依据前面几帧正常样本的稀疏表示代价值的均值作为阈值，并乘上一个敏感系数ω，当样本的稀疏表示代价大于这个乘积时，就判定为异常帧，否则为正常帧。同时关于具体的报警策略，由于样本中存在噪音，在检测过程中必然存在误检的情况，所以只有当连续出现4次以上异常帧时，系统才发出报警信号，以免由于误检导致报警过于频繁。

## 视频异常检测方法的改进



### 快速稀疏表示近似算法

这一节介绍如何使用稀疏近似实现一种更快的稀疏表示算法。通过使用预定义的字典，上式中的基础矩阵F可以根据少数的字典原型与他们相应的稀疏系数的线性组合近似得到。基础矩阵F上的近似表示使得问题成为寻找在稀疏字典上的稀疏解。因此，最初的带有稠密矩阵的L1范式最小化问题可以简化为一个小的稀疏矩阵求解问题，从而使求解更有效。

信号的稀疏和冗余建模已被证明在信号重建和分类中非常有效。使用一个过完备字典D（），它包含L个原型（或原子）作为列，那么信号f（）可以用这些原子的一个稀疏线性组合表示。特别地，存在一个稀疏系数向量w，使得f的表示正好满足f=DW或近似满足f≈DW。

给定一个样本集合，使用K-SVD算法可以找到最优的字典D，并通过最小化下面的L2范式重新表示样本：

(4-7)

其中，是类i的第j个样本的稀疏表示，为允许的最大非零值。

假定我们已经通过从样本集中学习得到一个字典D，那么就可以得到公式（3-8）中的矩阵F和公式（4-7）中的稀疏表示W。对于一个新的观察者，我们可以找到它在D上的原子分解。那么，在这个特征空间中的线性等式可以重写为：

(4-8)

如果学习的字典D能够很好的近似这个信号（），我们可以把表示为，其中是冗余项满足。相似的有，。这样引入残余信号后，公式（4-8）可以重写为：

(4-9)

假定x稀疏，则称s-稀疏（只有s个非零值），那么有

(4-10)

另，它也是一个稀疏向量，那么有。使用约束等距性（Restricted Isometry Property, RIP）可以决定这个稀疏编码是否稳定的获取。

(4-11)

假定z是sz-稀疏的且字典满足sz的RIP，通过公式（3-4）和（3-5）推出z的上界为

(4-12)

因此有

(4-13)

具体的RIP约束是未知的，然而，假设RIP不变，D近似sz-稀疏信号的欧氏距，有一个确定的上届，这样稀疏 解x可以通过解决更新的问题来恢复，即X可以通过解决下面的l1最小化问题得到：

(4-14)

如果字典D可以准确的重建信号，那么公式（3-9）中，问题简化为

(4-15)

见图 3‑1 中（b）所示。

下面阐述简化的l1最小化问题如何实现相同稀疏问题的加速。首先考虑19中的重建，W现在的维度是L×N（原来为d×N）。开始要先解决一个大的线性规划问题，然而因为W仅包含稀疏的列向量，在公式（4-15）中它可以通过定义相关列进行简化，这些列都有相同的支持如。这些列的定义和矩阵的简化过程如下：给定，先确定中的非0值，并记且（如图 3‑1 中（b）所示）。扫描W中所有列，找到不一致的支持列并删除，同时设置相应的系数值为0，这样得到的一个简化的WR将小于W。举例来说支持的只有三个实值：第1、第4和第9元素，然后扫描W中的所有列，仅获取使用相应支持的向量如图b所示，在确定用来表示的列后，就可以使用同样准确的解决方法最小化下面简化的L1范式：

(4-16)

其中，只包含相关列向量，如图 3‑1 中（c）所示。正如比N小得多，简化后的矩阵是一个精简矩阵。

### 时间与空间复杂度分析

在解决公式（4-16）中的简化问题时，主要有三个步骤。首先，矩阵可以用OMP算法高效地计算，时间复杂度为。其次，通过K－SVD算法计算W需要花费地时间复杂度，之后为了获得WR需要一次在列向量上进行一次扫描，时间复杂度为O(LN)。L1范式最小化问题的高效算法会将最初的问题转化为一个二阶锥规划问题，它可以通过对数障碍法解决。在对数障碍法的核心部分，解决的就是一个带二次限制形式的线性系统。这样，这个计算消耗分别为与。因此对于稠密矩阵相乘和稀疏矩阵相乘的时间复杂度分别是和。可以进一步得出，最初的稀疏表示与改进后的时间复杂度的比率为，这里比N小得多，并且L和m的尺度通常相等。因此改进后的算法可以在很大程度上提高速度。

在中间步骤中，存储学习的字典和矩阵W的空间复杂度是O( LK+KN)。然而，简化后的矩阵的空间复杂度只有。在解决L1范式最小化问题时，与的计算存在二次方的约束，因此原形式在空间复杂度比改进后的方法要高的多，这个比例是。

### 融合快速稀疏近似算法

基于上面提到的快速稀疏近似算法，这里将在视频异常检测算法中替换原先的稀疏表示算法，从而达到提供检测速度的目标，同时也保证了检测结果。

同样的方法，首先使用K-SVD方法对训练集Y进行字典学习得到字典Ф。基于前面的快速稀疏近似算法，对于上述问题的稀疏表示可以简化为公式（4-16）所要解决的问题。对于简化后的字典，原稀疏系数矩阵计算公式(4-5)经过简化后的计算公式如下：

(4-17)

对于给定的测试样本y，采用快速稀疏表示算法后，则简化为，重新计算的稀疏表示代价函数为：

(4-18)

下面是加入了快速稀疏近似算法后的视频异常检测框架。

表 4‑2

|  |
| --- |
| 视频异常检测框架-采用快速稀疏近似算法 |
| 输入：训练字典，测试样本  输出：样本state（normal 或 abnormal）  //快速稀疏近似  Y->WY  Ф->WA  for i=1,…,N do  WA->WR  //计算重建的稀疏系数矩阵  //计算稀疏表示代价  //基于测试样本的前k帧构建稀疏表示代价的阈值Threshold（已知这k帧正常）  if i == k  Threshold = 1/k \*  end if  //从第k+1帧开始进行检测  if  state = abnormal  else  state = normal  end if  end for |

在这个检测框架中，加入了快速稀疏近似算法，这样字典Ф和测试样本Y都得到了简化，在检测速度上有了明显的提升。

## 本章小结

本章详述了一种基于稀疏表示对拥挤场景中的异常进行检测的方法。稀疏表示在信号处理和压缩感知领域已有广泛的应用，而本课题通过稀疏表示过程中代价对异常进行判定从而达到检测视频中异常事件的目标。对于训练样本（正常样本）进行字典学习得到的字典，在用它对测试样本进行稀疏表示后，异常样本会得到较大的稀疏表示代价，这个稀疏表示代价将作为异常检测的判断依据。同时，本章还在稀疏表示过程中应用了第3章提到的快速稀疏表示方法，在很大程度上可以提高检测速度。

# 系统实现与实验结果



## 引言

在基于时空MRF框架下的视频异常检测中，本课题选用了Adam[37]提供的地铁出口视频作为实验视频，在这个场景中，存在着人从错误方向通过出口引起的异常等。而为了测试基于稀疏表示代价的视频异常检测方法及其改进算法，本课题主要对UMN[43]视频库中的3个场景进行检测。UMN视频库中包含了三个场景：草地、大厅和广场，在三个场景中都出现了不同程度的异常群体行为。下面将首先介绍实验的开发环境和系统实现的架构设计，在实验结果分析中，将会对本课题提到的方法进行进一步的评价。

## 系统实现



### 实验开发环境

本课题的实验开发环境如下表5-1所示：

表 5‑1 实验开发环境

|  |  |
| --- | --- |
| 名称 | 配置 |
| 操作系统 | Windows7 Ultimate 64bit |
| 开发环境 | MATLAB R2012b / Visual Studio 2008 |
| CPU | Intel Core2 Quad CPU 2.33GHz |
| 内存 | DDR3 1333 4G |

在时空MRF框架下实现的异常检测中，因为主要是基于opencv进行开发，所以开发环境是Visual Studio 2008；而在基于稀疏表示的框架中主要在MATLAB R2012b环境下进行实验研究。

### 系统实现框架

基于时空MRF的视频异常检测框架如图 5‑2所示，第一步是要使用MPPCA学习局部节点的光流特征模式，在这里要将视频中的图像划分成若干个区域，每个区域就构成了MRF图中的一个节点，在选取一定的时域后，就可以构成为时空MRF；接下来就是对构建的时空MRF进行参数学习；最后通过置信传播理论来对节点进行推断，从而得出异常节点。可以看出整个框架通过MPPCA和MRF参数的更新实现系统的不断学习能力，并能适应新的情况。



图 5‑1 基于时空MRF的视频异常检测框架

系统的实现框架如图 5‑2所示，首先对训练样本和测试样本进行相应的特征提取，即得到相应的MHOF特征；然后基于训练样本的特征采用K-SVD算法学习得到字典D，这里的训练样本全部是正常样本；接下来就是根据稀疏表示算法对测试样本进行稀疏表示，从而可以得到测试样本的稀疏表示代价；最后，根据测试样本的稀疏表示代价并基于一定的判定条件来得出样本是否异常。



图 5‑2基于稀疏表示的视频异常检测框架

基于稀疏表示的系统框架，本课题搭建了一个可视化的操作平台，便于实验的操作和结果显示，系统实现的UI显示界面如图 5‑3所示：



图 5‑3 基于稀疏表示的视频异常检测UI

## 实验结果分析



### 时空MRF的实验结果

对Adam提供的地铁出口实验数据进行检测，检测效果图如图 5‑4所示，在图中的六个帧中，红框标记了发生方向错误的区域，作为一种异常进行处理。



图 5‑4 地铁出口视频的检测效果图

整个视频的检测结果在表 5‑2中作出了具体的表述和说明，基于时空MRF的方法对地铁出口方向错误的异常作出了很好的检测，但是对于徘徊这种相对复杂的运动模式无法检测，在整个过程中也有一定的错误报警情况。总的来说，该法可以对地铁出口逃票等异常行为进行有效的检测。

表 5‑2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 方向错误 | 徘徊 | 共计 | 错误报警 |
| 真实数据 | 9 | 3 | 11 | 0 |
| 检测结果 | 9 | 0 | 9 | 2 |

### 稀疏表示的实验结果

前面已经提到，正常样本和异常样本经过稀疏表示后得到的稀疏系数会表现出明显的差异，讲过实验证明，异常样本较之于正常样本的稀疏系数的幅值要大的多。如图 5‑5所示，上图的绿色曲线表示的是正常样本的稀疏系数杆状图，下图中的红色曲线表示的是异常样本的稀疏系数杆状图。从图中不难看出，正常样本的幅值在[-1,1]区间内，而异常样本的的幅值在[-30,30]区间内，异常样本稀疏系数的值更大。

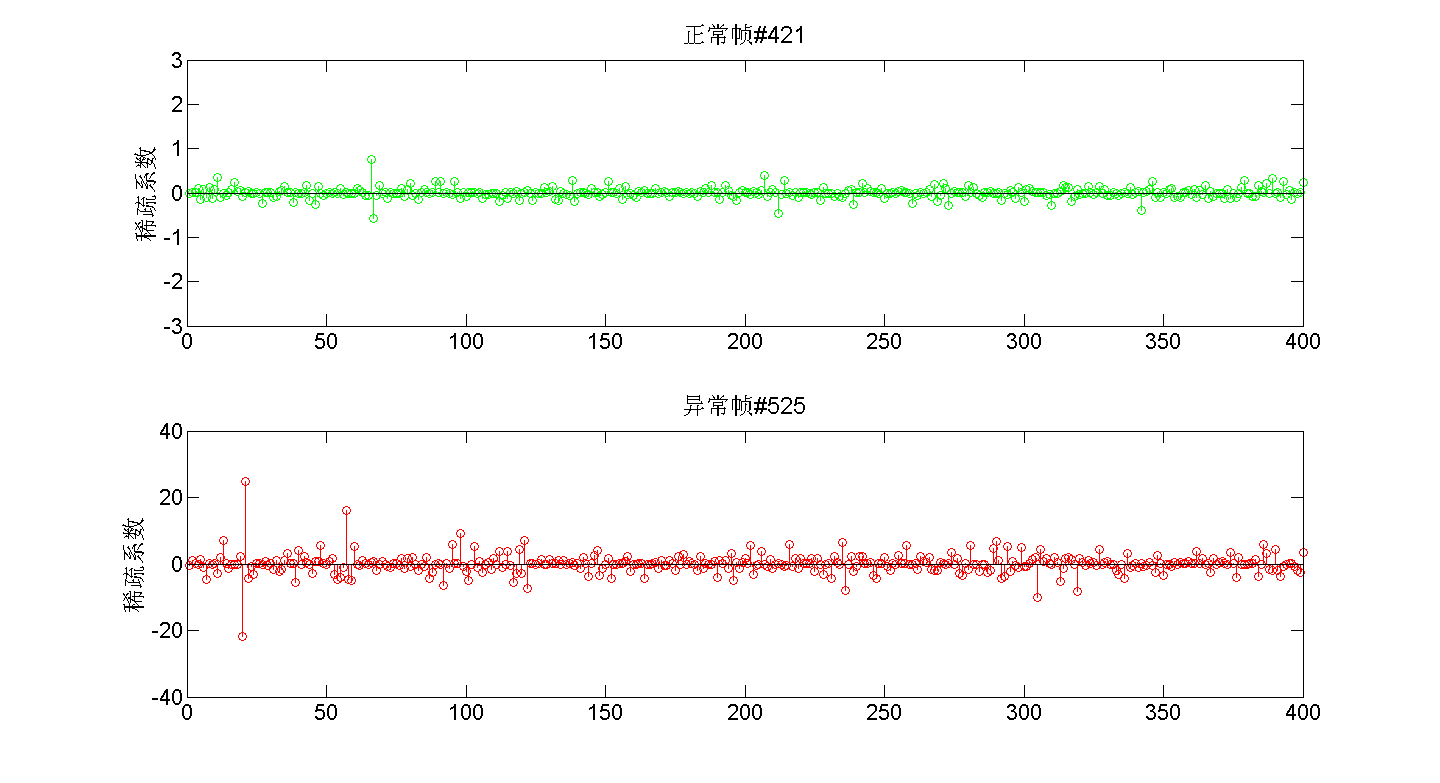


图 5‑5 正常样本和异常样本的稀疏系数

针对UMN视频库，本实验基于稀疏表示的方法对实验数据进行异常检测。实验中的检测效果如图 5‑6、图 5‑7、图 5‑8所示，分别给出了场景1、场景2、场景3的正常与异常的检测效果，在异常情况下，系统会做出报警。



图 5‑6 场景1（草地）



图 5‑7 场景2（大厅）



图 5‑8 场景3（广场）

在对整个实验数据进行检测后，可以通过稀疏表示代价曲线图更好的表现出现异常帧的位置，图 5‑9和图 5‑10分布给出了场景1的两种稀疏表示方法的代价的曲线，其中红色区域表示实验检测到的异常帧。

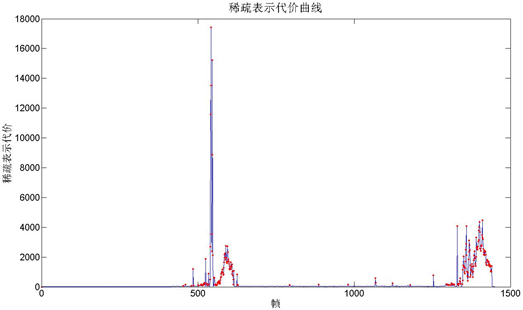


图 5‑9 稀疏表示方法

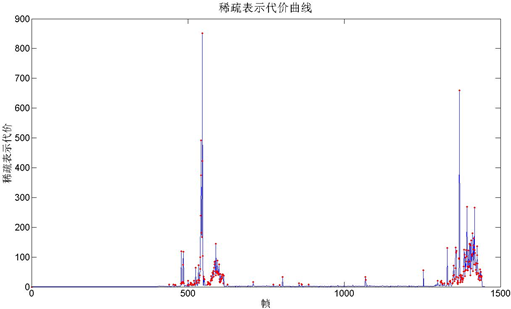


图 5‑10快速稀疏近似方法

对整个实验的检测结果进行统计后的示意图如图 5‑11 所示：



图 5‑11 UMN视频库异常检测结果示意图

三个场景的异常检测的实验结果分析如表所示：

表 5‑3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 总帧数（除去训练帧） | 异常帧 | 检测到的异常帧 | TPR | FPR | ACC |
| 场景一 | 1049 | 224 | 195 | 0.8705 | 0.0388 | 0.9419 |
| 场景二 | 3830 | 614 | 476 | 0.7752 | 0.2497 | 0.7543 |
| 场景三 | 1739 | 123 | 87 | 0.7073 | 0.1058 | 0.8810 |

采用快速近似算法的实验结果分析如表所示：

表 5‑4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 总帧数（除去训练帧） | 异常帧 | 检测到的异常帧 | TPR | FPR | ACC |
| 场景一 | 1049 | 224 | 190 | 0.8482 | 0.0412 | 0.9352 |
| 场景二 | 3830 | 614 | 322 | 0.5244 | 0.1253 | 0.8185 |
| 场景三 | 1739 | 123 | 78 | 0.6342 | 0.0439 | 0.9333 |

每帧稀疏表示的平均时间消耗对比如表所示：

表 5‑5

|  |  |  |  |
| --- | --- | --- | --- |
|  | 每帧稀疏表示的平均时间消耗（s） | | |
|  | 场景1 | 场景2 | 场景3 |
| 一般方法 | 0.44562 | 0.24265 | 0.58433 |
| 快速稀疏近似算法 | 0.0046517 | 0.0049247 | 0.0046036 |

## 本章小结

本章对前面阐述的两个异常检测框架进行了实验研究。在基于时空MRF的异常检测框架中，采用了Adam提供的地铁出口视频作为实验数据，实验结果表明该方法可以有效检测出地铁口出现的局部异常情况。UMN视频库中三个场景检测的实验结果证明：在拥挤场景中基于稀疏表示代价的异常检测方法有较好的效果。同时，再采用快速稀疏近似算法对原有算法进行改进后，检测方法的执行速度有了大幅度的提高，这样的检测速度可以更好地适应实时性检测。而且，实验结果表明采用快速稀疏近似算法并未对检测效果产生影响。

# 结论

本课题的主要研究内容是拥挤场景中的视频异常事件检测，对于深受关注的公共安全问题，这样的群体事件研究具有重要的研究价值和实际应用意义。研究得出的结论如下：

1. 研究了一种应用MPPCA进行特征处理并建立时空MRF模型进行异常推断的方法。为了更好的表示时空MRF模型中节点的运动特征，采用MPPCA方法对低层的光流特征进行学习，通过混合概率模型的组合实现对特征的有效表示。基于这种运动特征表示方法来构建时空MRF，通过在MRF图模型中应用置信传播最小化能量函数后，可得出图模型中的各节点的标号，即由此推断节点是否异常。本课题使用Adam提供的地铁出口视频作为实验数据，对该框架进行验证。经实验结果论证，对于一段时域内的视频帧构成的时空MRF模型，该方法可以对其中的异常节点进行有效的检测。
2. 建立了一种基于稀疏表示对拥挤场景中的异常进行检测的框架。对于训练样本（正常样本）进行字典学习得到的字典，在用它对测试样本进行稀疏表示后，异常样本会得到的稀疏系数幅值较大。本课题基于稀疏系数进一步构建稀疏表示代价函数，并以此对异常进行判定从而达到检测视频中异常事件的目标。本课题同归对来自于UMN的视频库进行分析，发现异常样本的稀疏系数幅值比正常样本打10倍甚至更多。依据构建的异常检测框架，对UMN中的三个场景进行检测后，可以论证该方法在拥挤场景中对由于人群突然散开引起的异常具有很好的检测效果。
3. 在基于稀疏表示进行异常检测框架下，进一步融合一种快速近似稀疏算法。在实际的视频处理中，通常会处理海量的视频数据，对于时间和空间复杂度具有很大的要求，因此该方法具有一定的研究意义。在最初使用的字典比较稠密，所以可以通过消除字典中的冗余项来实现一种近似稀疏表示方法，从而提高稀疏表示过程的速度并节省存储空间。在此基础上，将其融合到视频异常检测框架中，可以大幅度的提高帧的处理速度，从而保证了基于稀疏表示的视频异常检测框架的实时性。经实验论证，采用快速近似稀疏算法后，进行异常检测时，每帧的处理速度可以提高几十倍甚至上百倍，大大提高了异常检测的性能。

本课题重点对拥挤场景情况下的群体异常进行了研究，其中仍存在一些问题有待于进一步研究去解决。本课题提出的异常检测方法虽然可以对群体异常进行检测，但并不能分析出异常的类型，没有进行分类处理。还有，在检测过程中如何消除不必要的噪音数据，尽可能避免漏报和错报等状况。

# 参考文献

1. 张一.  [智能视频监控中的目标识别与异常行为建模与分析](http://www.cnki.net/kcms/detail/detail.aspx?filename=2010113583.nh&dbcode=CDFD&dbname=CDFD2010&v=Mjg0OTl2RkNyZ1YxMjZIcks1SGRURXJPQTBZK2dPRFhROXVSUmw3RTBJUG5ibXd4QTFlTEdVUXJpY1p1Qg==)[D]. 上海交通大学 2010
2. 李科煌. 道路监控下的运动目标检测和异常行为检测有效方法研究[D].上海:上海交通大学.2011:1-75.
3. 陈颖鸣，陈树越，张显亭. 智能视频监控中异常行为识别研究[J]. 微电子学与计算机学，2010，27(11):102-105.
4. Chen J, Tang C K. Spatio-temporal markov random field for video denoising[C]//Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007: 1-8.
5. Horn B K P, Schunck B G. Determining optical flow[J]. Artificial intelligence, 1981, 17(1): 185-203.
6. KaewTraKulPong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection[C]//Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems. 2001, 25: 1-5.
7. Kim J, Grauman K. Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 2921-2928.
8. Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//Proceedings of the 7th international joint conference on Artificial intelligence. 1981.
9. Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[C]//Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. IEEE, 1999, 2.
10. Shum H Y, Ikeuchi K, Reddy R. Principal component analysis with missing data and its application to polyhedral object modeling[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1995, 17(9): 854-867.
11. Sun D, Roth S, Lewis J, et al. Learning optical flow[J]. Computer Vision–ECCV 2008, 2008: 83-97.
12. Tipping M E, Bishop C M. Probabilistic principal component analysis[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1999, 61(3): 611-622.
13. Tipping M E, Bishop C M. Mixtures of probabilistic principal component analyzers[J]. Neural computation, 1999, 11(2): 443-482.
14. Williams O, Isard M, MacCormick J. Estimating disparity and occlusions in stereo video sequences[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 2: 250-257.
15. Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction[C]//Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. IEEE, 2004, 2: 28-31.
16. Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision[J]. International journal of computer vision, 2000, 40(1): 25-47.
17. Cui X, Liu Q, Gao M, et al. Abnormal detection using interaction energy potentials[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 3161-3167.
18. Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision[J]. International journal of computer vision, 2006, 70(1): 41-54.
19. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001, 23(11): 1222-1239.
20. Huang J B, Yang M H. Fast sparse representation with prototypes[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 3618-3625.
21. Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 3449-3456.
22. Szeliski R, Zabih R, Scharstein D, et al. A comparative study of energy minimization methods for markov random fields with smoothness-based priors[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008, 30(6): 1068-1080.
23. Engan K, Aase S O, Hakon Husoy J. Method of optimal directions for frame design[C]//Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on. IEEE, 1999, 5: 2443-2446.
24. Engan K, Aase S O, Husøy J H. Multi-frame compression: Theory and design[J]. Signal Processing, 2000, 80(10): 2121-2140.
25. M. Aharon , M. Elad and A. M. Bruckstein  "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation",  IEEE Trans. Signal Process.
26. Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit[J]. SIAM journal on scientific computing, 1998, 20(1): 33-61.
27. Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. Signal Processing, IEEE Transactions on, 1993, 41(12): 3397-3415.
28. Donoho D L. For most large underdetermined systems of linear equations the minimal L1‐norm solution is also the sparsest solution[J]. Communications on pure and applied mathematics, 2006, 59(6): 797-829.
29. Chen S, Billings S A, Luo W. Orthogonal least squares methods and their application to non-linear system identification[J]. International Journal of control, 1989, 50(5): 1873-1896.
30. Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 935-942.
31. Cui X, Liu Q, Gao M, et al. Abnormal detection using interaction energy potentials[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 3161-3167.
32. Mahadevan V, Li W, Bhalodia V, et al. Anomaly detection in crowded scenes[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 1975-1981.
33. Bruckstein A M, Donoho D L, Elad M. From sparse solutions of systems of equations to sparse modeling of signals and images[J]. SIAM review, 2009, 51(1): 34-81.
34. Candes E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements[J]. Communications on pure and applied mathematics, 2006, 59(8): 1207-1223.
35. Candes E J, Tao T. Near-optimal signal recovery from random projections: Universal encoding strategies?[J]. Information Theory, IEEE Transactions on, 2006, 52(12): 5406-5425.
36. Benezeth Y, Jodoin P M, Saligrama V, et al. Abnormal events detection based on spatio-temporal co-occurences[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 2458-2465.
37. 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.
38. Kratz L, Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 1446-1453.
39. Zhong H, Shi J, Visontai M. Detecting unusual activity in video[C]//Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 2: II-819-II-826 Vol. 2.
40. Hu W, Xiao X, Fu Z, et al. A system for learning statistical motion patterns[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006, 28(9): 1450-1464.
41. Jiang F, Yuan J, Tsaftaris S A, et al. Anomalous video event detection using spatiotemporal context[J]. Computer Vision and Image Understanding, 2011, 115(3): 323-333.
42. Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]//Advances in neural information processing systems. 2006: 801-808.
43. Unusual crowd activity dataset of University of Minnesota, from <http://mha.cs.umn.edu/movies/crowdactivity-all.avi>.

# 哈尔滨工业大学硕士学位论文原创性声明

本人郑重声明：此处所提交的硕士学位论文《基于视频的行人检测及异常行为检测》，是本人在导师指导下，在哈尔滨工业大学攻读硕士学位期间独立进行研究工作所取得的成果。据本人所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。本声明的法律结果将完全由本人承担。

作者签字 日期： 年 月 日

哈尔滨工业大学硕士学位论文使用授权书

《基于视频的行人检测及异常行为检测》系本人在哈尔滨工业大学攻读硕士学位期间在导师指导下完成的硕士学位论文。本论文的研究成果归哈尔滨工业大学所有，本论文的研究内容不得以其它单位的名义发表。本人完全了解哈尔滨工业大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅。本人授权哈尔滨工业大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。

保 密□，在 年解密后适用本授权书。

本学位论文属于

不保密□。

（请在以上相应方框内打“√” ）

作者签名： 日期： 年 月 日

导师签名： 日期： 年 月 日

# 致谢

两年半的研究生时光即将结束，回忆这宝贵的学习生活感慨万分，在此向所有帮助以及关心我的老师、同学、亲人、朋友表示感谢。

首先，要感谢我的导师王轩教授。王轩老师渊博的学识以及严谨的科研态度深深的感染了我，能成为王老师的学生令我深感荣幸。从课题的选定到论文的完成，王轩老师都给我的极大的帮助，对于论文中每个细节，老师都给出了宝贵的意见，令我受益匪浅。跟随王轩老师学习的两年半中，我不仅学习到了丰富的知识，更加培养了我对科学研究的严谨态度，老师的每一句教诲都将成为我前进路上一笔宝贵财富。更重要的是，老师豁达的人生态度，对生活细微深刻的观察和思考，值得我们一生学习。

其次，感谢深研院曾给我授课老师们，他们的言传身教为我树立了一个个榜样，教会我赖以生存的技能和品质。感谢位老师认真地审查我的论文，为我的论文和答辩提出宝贵的意见。感谢在我参与的科研项目和毕业课题中给我耐心指导的于成龙师兄、李鑫鑫师兄、张自力师兄等，与他们的每一次交谈总能让我受益匪浅。

再要感谢我的11级同门以及实验室的全体同学们，在这两年多的时间里，我与他们共同探讨，共同学习，共同成长，正因为有了他们的帮助，我的论文才能顺利完成。还要感觉我的舍友段成刚和王珏，感谢他们在这段时间对我的关怀与包容。

最后要感谢我的家人，感谢我的父母赋予了我生命，感谢我的姐姐对我的关心，感谢他们这么多年对我的付出与支持。