**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**

# 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。今年来，视频监控中的许多关键领域已取得了很大的进步，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。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。

异常事件监测涉及图像处理、计算机视觉、模式识别等多个领域，本文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，主要研究内容包括：

1. 研究一种MPPCA特征处理方法并在时空MRF推断异常。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，而本文采用后者进行计算。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。
2. 基于稀疏表示和字典学习算法实现了一个快速的近似稀疏表示算法。
3. 探讨和研究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，而本文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，在不影响检测结果的情况下，很大程度上可以提高检测速度。

**关键词：**异常行为检测，智能视频监控，计算机视觉，拥挤场景，全局异常， 局部异常，MPPCA，MRF，稀疏表示，字典学习。

# Abstract

目录

[摘要 I](#_Toc364499813)

[Abstract II](#_Toc364499814)

[第1章 绪论 1](#_Toc364499815)

[1.1 课题背景 1](#_Toc364499816)

[1.2 课题研究的目的及意义 2](#_Toc364499817)

[1.3 国内外研究现状 3](#_Toc364499818)

[1.3.1 国内研究现状 3](#_Toc364499819)

[1.3.2 国外研究现状 3](#_Toc364499820)

[1.4 主要研究内容及论文结构 4](#_Toc364499821)

[第2章 学习运动模式并在时空MRF中推断视频中的异常 6](#_Toc364499822)

[2.1 引言 6](#_Toc364499824)

[2.2 马尔可夫随机场相关理论 7](#_Toc364499825)

[2.2.1 马尔可夫随机场理论 7](#_Toc364499826)

[2.2.2 MRF图模型的表示 8](#_Toc364499827)

[2.2.3 MRF能量最小化方法 9](#_Toc364499828)

[2.3 MPPCA学习局部状态的运动模式 12](#_Toc364499829)

[2.3.1 光流特征提取 12](#_Toc364499830)

[2.3.2 混合概率主成分分析 13](#_Toc364499831)

[2.3.3 学习局部状态的运动模式 14](#_Toc364499832)

[2.4 构建运动模式的时空MRF模型 15](#_Toc364499833)

[2.4.1 时空MRF模型的能量函数 15](#_Toc364499834)

[2.4.2 时空MRF模型的自更新 16](#_Toc364499835)

[2.5 本章小结 17](#_Toc364499836)

[第3章 字典学习与快速稀疏表示 17](#_Toc364499837)

[3.1 引言 17](#_Toc364499839)

[3.2 稀疏表示的相关理论 18](#_Toc364499840)

[3.3 字典学习算法 18](#_Toc364499841)

[3.3.1 最优方向法 19](#_Toc364499842)

[3.3.2 K-SVD算法 19](#_Toc364499843)

[3.4 一种快速稀疏表示算法 22](#_Toc364499844)

[3.4.1 有原型的稀疏表示 22](#_Toc364499845)

[3.4.2 有原型的快速稀疏近似 24](#_Toc364499846)

[3.5 本章小结 25](#_Toc364499847)

[第4章 稀疏表示方法检测视频中的异常 26](#_Toc364499848)

[4.1 引言 26](#_Toc364499851)

[4.2 稀疏表示实现视频异常检测 27](#_Toc364499852)

[4.3.1 多尺度光流直方图（MHOF）特征提取 28](#_Toc364499857)

[4.3.2 字典选取 29](#_Toc364499858)

[4.3.3 异常的衡量标准与优化 29](#_Toc364499859)

[4.3 使用快速稀疏近似算法提高检测速度 31](#_Toc364499860)

[4.4 本章小结 31](#_Toc364499861)

[第5章 系统设计与实验结果 31](#_Toc364499862)

[5.1 引言 31](#_Toc364499864)

[5.2 系统设计 31](#_Toc364499865)

[5.3 实验结果 32](#_Toc364499866)

[5.4 本章小结 32](#_Toc364499867)

[结论 33](#_Toc364499868)

[参考文献 34](#_Toc364499869)

[哈尔滨工业大学硕士学位论文原创性声明 37](#_Toc364499870)

[致谢 38](#_Toc364499871)

# 绪论

## 课题背景

近年来，异常行为检测已成为智能监控系统中的一个重要的研究领域，这主要得益于视频监控系统的快速发展和社会关注的公共安全问题。视频监控系统广泛应用在政府部门和金融、公安、交通等行业中，视频监控市场受到“平安城市”项目、奥运安防项目以及各地、各行业安防项目需求加速增长等因素的强劲刺激和拉动，取得了快速的发展，除了前面的传统行业，企业和个人市场也正在兴起，中国经济最活跃的中小企业和个人用户，对视频监控的应用前景都已经明显呈现出来。

中国视频监控市场经历了模拟监控、数字监控及网络监控三个主要阶段。第一代视频监控采用模拟监控技术，在图像质量、监控范围上具有较大的局限性。数字监控作为继模拟监控之后的第二代监控技术，无论在图像质量、保存时间以及可靠性上均有大幅度的提升，为用户提供了高性价比的监控解决方案。当前，数字监控已牢牢地成为主导整体视频监控市场的视频监控解决方案，并成熟部署于金融、商业等行业之中。随着近年来“平安城市”、“平安校园”等安防项目在全国范围的开展和深入，机场、地铁以及景区等用户对于视频监控覆盖范围、监控点数以及网络传输I/O 等要求的不断提升，网络监控正成为中国视频监控市场重要的拉动因素。网络监控设备厂商的视频监控整体解决方案，正得到越来越多用户的了解和认可。在国内大型的视频监控项目中，更是出现了视频监控系统中，除后端显示设备之外，全部设备IP 化的发展趋势。

对于传统的视频监控系统与设备，仍然存在效率低且人员易疲劳等缺点，这就需要一个更加智能的监控系统来辅助人工的检测。智能监控系统是采用图像处理、模式识别和计算机视觉技术，通过在监控系统中增加[智能视频分析](http://baike.baidu.com/view/3495170.htm)模块，借助计算机强大的[数据处理](http://baike.baidu.com/view/51860.htm)能力过滤掉视频画面无用的或干扰信息、自动识别不同物体，分析抽取视频源中关键有用信息，快速准确的定位事故现场，判断监控画面中的异常情况，并以最快和最佳的方式发出警报或触发其它动作，从而有效进行事前预警，事中处理，事后及时取证的全自动、全天候、实时监控的智能系统，其主要研究方向包括检测与追踪、人体动作分析、活动分析与事件监测等。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。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作出智能响应。从中不难看出，智能监控系统将成为智慧城市建设的重点之一。

拥挤场景中的异常检测是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，如何实时有效地监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

智能视频监控技术正在向着数字化、网络化、智能化的方向发展，其中智能化是“三化”的最高境界。传统的视频监控系统与设备虽然在功能和性能上得到了极大的提高，但实际的监控任务仍需要较多的人工工作来完成。在很多情况下，目前的视频监控系统所提供的信息是没有经过任何分析的视频裸数据，这就不能充分发挥监控系统应有的实时主动的监督作用，导致整个系统在安全性和实用性方面仍然没有达到人们期望。具体的制约因素如下:

1、人类自身的弱点:在很多情况下，人类并非是一个可以完全信赖的观察者，无论是观看实时的视频流还是观看录像回放，由于自身生理上的弱点，人类经常无法察觉安全威胁，从而导致漏报现象的发生。

2、监控时间:除了一些规模较小的监控应用之外，很少有视频监控系统会按照1: 1的比例为监控摄像机配置监视器。这意味着对于那些机场、港湾等大型的视频监控系统来讲，各个监控点并非每时每刻都处于监控当中。

3、误报和漏报:误报和漏报是目前视频监视系统中最常见的两大问题。漏报是指在监控点发生了某种安全威胁，但该威胁并没有被监控系统或安全人员发现;误报是指位于监控点的安全活动被误认为是安全威胁，从而产生错误的报警。漏报可能会导致非常危险的后果发生，而误报会浪费人力物力。并且这两种问题都会大大降低人们对监控系统的信任，从而降低监控系统的应用价值。

4、数据分析困难:报警发生后对录像数据进行分析通常是安全人员必须要做的工作之一，而误报和漏报现象则进一步加重了进行数据分析的工作负担。另外，安全人员经常被要求找出与报警事件相关的录像资料，找到肇事者、确定事故责任或评估该事件的安全威胁程度。但由于传统视频监控系统缺乏智能因素，录像数据无法被有效的分类存储更不用说其他智能分析，最多只能打上时间标签，因此数据分析工作变得极其耗时，并且很难获得全部的相关信息，而经常发生的误报漏报现象使得无用数据进一步增加有用数据经常缺失，从而给数据分析工作带来了更大的困难。

5、响应时间长:由于安全威胁的响应速度关系到一个安全系统的整体性能，传统的视频监控系统通常都由安全工作人员对安全威胁做出响应和处理，这对于处理对实时响应要求较低的安全威胁来说可能足够。但是很多情况下，在威胁发生时，需要安全系统的多个功能部分，甚至多个安全相关的部门在最短的时间内协调配合，共同处理危机。这时候，监控系统的响应速度将直接关系到用户的人身或财产的损失情况。为了解决上述导致视频监控效率低下的问题和减少对繁琐人工的依赖，自动完成对复杂环境中的目标进行实时观测以及对感兴趣的对象的行为进行分析和描述，就需要智能视频监控技术作为支持。

异常检测与报警在某些场合，监控系统有必要自动地检测场景中人或车辆是否存在着异常的行为或事件。比如说，停车场或超市需要监控是否有盗窃行为，一般来说，可以有两种警告方式，一种是当此种情况发生的时候自动的做出公共的警告，或者是直接自动地联系警方。

对于场景可按是否拥挤分为拥挤场景与非拥挤场景，本文主要是针对拥挤场景中的异常事件进行检测。

## 国内外研究现状

### 国内研究现状

基于机器视觉的运动目标跟踪与识别在国内也得到了广泛的推广，863开展了一系列重大项目研究，公安部在全国开展城市报警与监控系统建设“3111”试点工程。中国科学院自动化研究所在相关研究领域取得了重大进展，自动化研究所生物识别与安全技术研究中心研究开发的CBSR（Center for Biometrics and Security Research）智能视频监控系统要功能包括人和车辆的多目标检测、跟踪，并对目标异常行为的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。

在国内，己有许多高等院校和研究机构在智能视频监控领域投入了相当程度的研究精力，例如:清华大学、中国科学院自动化研究所，上海交通大学、华中科技大学等。其中，中国科学院自动化研究所模式识别实验室在该领域作了大量工作，他们在人体运动分析、交通行为事件分析、交通场景视频监控和智能轮椅视觉导航等领域取得了许多科研成果。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

美国国防高级研究项目署DARPA（Defense Advanced Research Projects Agency）在2000年先后资助了视频监控项目（Video Surveillance and Monitoring, VSAM[1]）和远程人类识别项目（Human Identification at a Distance, HID[2]）。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。HID用于研究开发多模式的监控技术以实现远距离情况下人的检测、分类和识别。

CAVIAR[3][4]（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，日本也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[5]）。

MIT媒体实验室开发了实时人体跟踪与行为理解系统Pfinder[6]，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[7]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。

UMN UCSD等视频库。

当前，国际上的许多信息处理类权威杂志如: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)和重要年度学术会议: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。

## 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。

本文主要研究拥挤场景中的异常事件监测。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，如图1-1（a）中所示，红色目标的运动模式与相邻区域的运动模式不同。GAE是指整个场景的群体异常行为，如图1-1（b）中所示，由于异常状况行人突然散开。

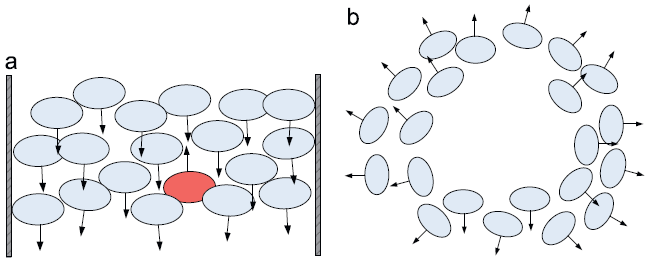


图 1-1 局部和全局异常事件

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下图1-2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。

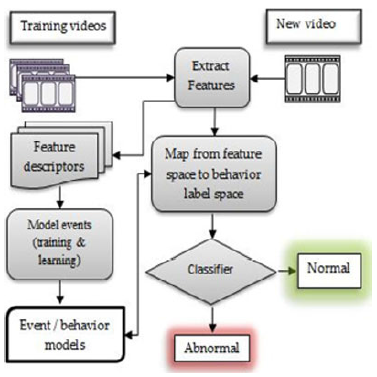


图 1-2基于特征的建模过程和视频序列中的异常检测方法

本文的主要研究内容包括两部分：基于MPPCA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[12]和图像去噪[2]。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特征建立一个概率模型，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如何通过对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表示方法对检测算法进行改进，可以有效地提高检测的速度。

本文结构如下：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，并提出一种快速稀疏表示方法；第4章则根据第3章中提出的内容应用在视频异常检测中，主要是提出稀疏表示代价算法实现有效地检测视频中的异常，并采用快速稀疏表示算法提高检测速度；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。

# 学习运动模式并在时空MRF中推断视频中的异常



## 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。

这里首先介绍一个计算机视觉的应用方法，在[16]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：

(2.1.1)

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：

(2.1.2)

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。

正是基于上面的理论框架，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。

## 马尔可夫随机场相关理论

### 马尔可夫随机场理论

马尔可夫过程是指下一个时刻的值仅和当前时刻的值有关，而与以前无关，即未来决定于现在而不是过去。这种已知“现在”的条件下，“未来”与“过去”彼此独立的特性被称为马尔可夫性，具有这种性质的随机过程叫做马尔可夫过程，其原始的模型就是马尔可夫链。有这样的一个例子，一只被切除了大脑的白鼠在若干个洞穴间的蹿动可以称之为马尔可夫链，因为这只白鼠没有了记忆，瞬间而生的念头决定了它从一个洞穴窜到另一个洞穴，当其所在位置确定时，它下一步蹿往何处与它以往经过的路径无关。

马尔可夫随机场（Markov Random Field，MRF）包含两层意思：马尔可夫性质和随机场。上面已经说明了马尔可夫性质，下面将介绍随即场的概念。当给每一个位置中按照某种分布随机赋予相空间的一个值之后，其全体就叫做随机场。我们不妨拿种地来打个比方，其中有两个概念：位置（site），[相空间](http://www.hudong.com/wiki/%E7%9B%B8%E7%A9%BA%E9%97%B4)（phase space）。“位置”好比是一亩亩农田；“相空间”好比是种的各种庄稼。我们可以给不同的地种上不同的庄稼，这就好比给随机场的每个“位置”，赋予相空间里不同的值。所以，随机场就是在哪块地里种什么庄稼的事情。现在拿种地比作马尔可夫随机场：如果任何一块地里种的庄稼的种类仅仅与它邻近的地里种的庄稼的种类有关，与其它地方的庄稼的种类无关，那么这些地里种的庄稼的集合，就是一个马尔可夫随机场。

MRF应用在视觉中，相当于一个标记问题，更具体点，是通过MAP推理来确定图中每个节点的标号。MRF相比其他方法的优势是：1）提供了一种原则性方法来对先验知识建模，2）MRF可以很容易用定量的方法描述上下文信息。因此，相比其它pixel-based, 或local-based 方法，它可以考虑到环境知识的影响，如果能建立适合的图模型，进而可能获得全局最优解，这更接近于人类视觉。

要想实现MRF的推理，首先必须有图结构和参数学习。图的创建一般是对问题本身的建模，比如在图像修复（image restoration）和图像分割（image segmentation）中，常用到四邻接（4-neighborhood）或八邻接（8-neighborhood）的成对模型，这样，四或八相邻的像素中间便用边连接，这样的模型就是典型的成对马尔可夫模型,如果要加入高阶（>=3)的势能，相当于我们引入了更多的约束，比如：连通性约束、非基督分类结果的约束，说到非监督分类结果的约束，要注意的是：一定是其他分类方法，而非MRF本身的分类结果构成新的约束。

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。下面介绍图模型的表示方法以及两种能量最小化方法：图割法和循环置信传播。

### MRF图模型的表示

图模型(graphical model)是一类用图来表示概率分布的一类技术的总称。它的主要优点是把概率分布中的条件独立用图的形式表达出来，从而可以把一个概率分布（特定的，和应用相关的）表示为很多因子的乘积，从而简化在边缘化一个概率分布的计算，这里的边缘化指的是给定n个变量的概率分布，求取其中m个变量的概率分布的计算（m<n)。

图模型主要有两大类，一类是贝叶斯网络（又称有向图模型）；另外一类是马尔可夫网络（又称无向图模型）。图模型主要有三个主要的关注点：

1）图模型的表示（representation）：指的是一个图模型应该是什么样子的；

2）图模型的推断（inference）：指的是已知图模型的情况下，怎么去计算一个查询的概率，例如已经一些观察节点，去求其它未知节点的概率；

3）图模型的学习（learning）：这里又分为两类，一类是图的结构学习；一类是图的参数学习。

这里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。无向图模型为了表示一个概率分布，需要把变量之间的条件独立编码在图表示中，从而使得概率分布的表示可以被表示为因子乘积的形式。与有向图模型表示不同的是：无向图模型是建立在无向图基础上，而有向图模型是建立在有向图基础之上。我们先看下图中的例子：

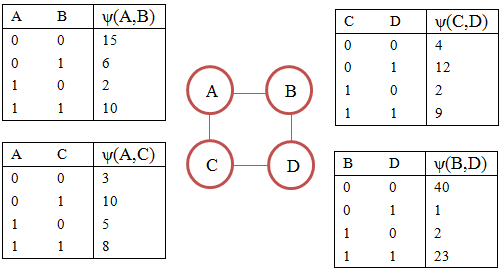


图 2.1

上图是一个无向图模型的完整的表示，左侧是它的拓扑结构，右侧是它的参数。无向图模型是以最大团和定义在团上的势能函数为核心，具体来说，在这个例子中，它有四个团：AC , AB, BD ,CD。那么我们需要在四个团上定义相应的势能函数，如图中表格所示，必须注意势能函数必须为正。那么这个无向图模型表示的概率分布是：P(A,B,C,D)= (1/Z)\* ψ(A,C)\* ψ(A,B)\* ψ(C,D)\* ψ(B,D)。其中，Z是归一化因子，因为势能函数并没有归一化，而要概率是[0，1]，所以需要归一化；ψ是表示相应的势能函数。所以一个无向图模型表示的概率分布形式化地可以表示为：，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：

(2.2.1)

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：

(2.2.2)

(2.2.3)

(2.2.4)

则在标号L空间中，V称为公制；如果V只满足（2.2.2）和（2.2.3），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：

(2.2.5)

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。

|  |
| --- |
| 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 |
| 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 |

如上表所示，算法swap-move 和 expansion-move[17]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。

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下：

(2.2.6)

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

(2.2.7)

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)。[16]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，的关系图如下：

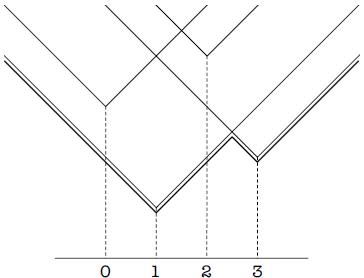


图 2.2

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:

(forward)，

For from k-2 to 0:

(backward)。

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forward 的过程后成为(3,1,2,2)，再经过backward后变成(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)。

在这里的讨论是在目前我们考虑的早期视觉问题中，可以用线性的方法在处理每个消息时将计算时间由O(k2)减少到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。

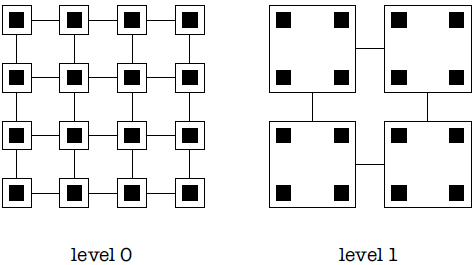


图 2.3

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[3]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[6]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（本文中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），如下图所示。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。

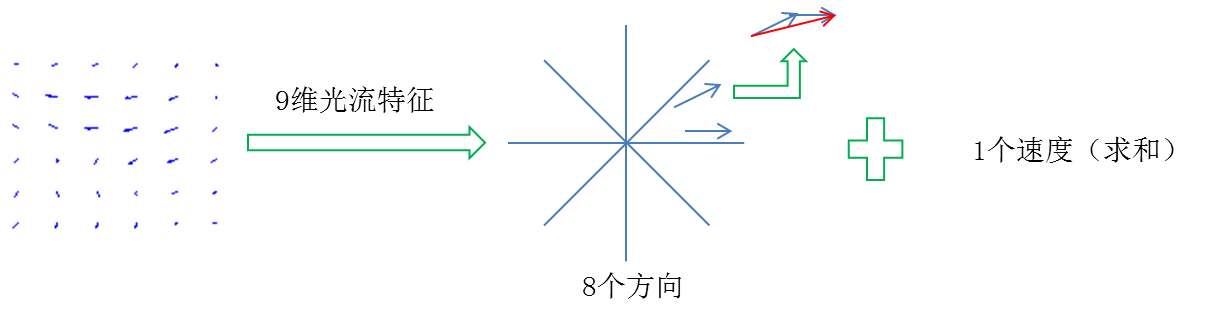


图 2.4

### 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2-1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 2.5

PPCA[10]（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[13]）。

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：

(2.3.1)

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：

(2.3.2)

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，计算公式如下：

(2.3.3)

(2.3.4)

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，定义如下：

(2.3.5)

(2.3.6)

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。

（总结）

## 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：

(2.4.1)

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：

(2.4.2)

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：

(2.4.3)

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：

(2.4.4)

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：

(2.4.5)

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：

(2.4.6)

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：

(2.4.7)

根据公式（2.4.6）和（2.4.7）可得完整的成对势能函数为：

(2.4.8)

其中，是一个衡量光滑度的权值常量。

给定了公式（2.4.5）和（2.4.8）在节点和链接处的MRF参数，即可通过MAP最大化式（2.4.1）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：

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，而本文采用后者进行计算。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。

# 字典学习与快速稀疏表示



## 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[24,26]，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[25]和OMP[27]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）。本文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。假定我们得到一个学习来的字典，我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。

## 稀疏表示的相关理论

在压缩感知中经常会遇到这样的问题，信号x在经过测量矩阵A后得到测量值y，即y=Ax，其中测量矩阵（m远小于n），那么怎么从y中精确的恢复出信号x呢？

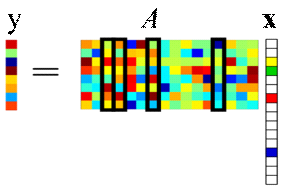


图 3.1

由于m远小于n，用m个方程求解n个未知数，因此y=Ax是个欠定方程，有无穷多个解。就像解优化问题一样，如果加上适当的限定条件，或者叫正则项，问题的解会变得清晰一些。这里我们加上的正则项是（零范数：值为0的元素个数），即使重构出的信号 x尽可能的稀疏。后来Donoho和Elad这对师徒证明了如果A满足某些条件，那么这个优化问题即有唯一解。唯一性确定了，仍然不能求解出该问题，后来就尝试使用l1和l2范数来替代l0范数，华裔科学家陶哲轩和candes合作证明了在A满足UUP原则这样一个条件下，l0范数可以使用l1范数替代，所以优化问题变成这样一个凸优化问题，可以通过线性优化的问题来解决，这就是稀疏表示要研究的问题。

对于测量值y，如果有一个已知的过完备字典D，就可以求出y在D上的稀疏表示x，这个过程被称作稀疏编码（Sparse Coding）。这里有两个未知量，过完备字典D和稀疏系数x，已知量是输入信号y，当然先验知识是输入信号在字典D上可以稀疏表示。问题模型如下：

同时获取字典D和稀疏系数x的方法是将该模型分解：第一步将D固定，求出x的值，这就是稀疏编码，也就是上面提到的字典D固定，求信号y在D上稀疏表示的问题；第二步是使用上一步得到的x来更新字典D，即字典更新（Dictionary Update）。如此反复迭代几次即可得到优化的D和x。

步骤1（稀疏编码）：

步骤2（字典更新）：

本文主要介绍两种方法：MOD和K-SVD。在MOD中，稀疏编码采用的方法是OMP贪婪算法，字典更新采用的是最小二乘法，下面将做进一步详细阐述。

## 字典学习算法

### 最优方向法

最优方向法（Method of Optimal Directions, MOD）是由Engan等[21,22]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：

(3.1)

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。对上面的公式求导就会得到，推出

(3.2)

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[23]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：

(3.3)

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：

(3.4)

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

(3.5)

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，详细步骤如下表所示。

|  |
| --- |
| Task: Find the best dictionary to represent the data samples by nearest neighbor, by solving  Initialization: Set the codebook matrix. Set J=1.  Repeat until convergence(use stop rule):   * Sparse Coding Stage: Partition the training samples Y into K sets   each holding the sample indices most similar to the column ,   * Codebook Update Stage: For each column k in , update it by * Set J = J+1. |

K-SVD对K-Means的扩展（泛化）

稀疏表示问题可视为对矢量化目标（16）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：

(3.6)

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：

(3.7)

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。

(3.8)

下面考虑字典的更新阶段，

(3.9)

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：

(3.10)

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，，。这样上面的公式可以变换为：

(3.11)

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。

|  |
| --- |
| Task: Find the best possible codebook to represent the data samples as sparse compositions, by solving  Initialization: Set the codebook matrix with 标准化的列. Set J=1.  Repeat until convergence(stopping rule):   * Sparse Coding Stage: Use any pursuit algorithm to compute the representation vectors xi for each example yi, by approximation the solution of * Codebook Update Stage: For each column k=1,2,…,K in , update it by   Define the group of examples that use this atom,  Compute the overall representation error matrix , by  Restrict by choosing only the columns corresponding to, and obtain .  Apply SVD decomposition . Choose the updated dictionary column dk to be the first column of U. Update the coefficient vector to be the first column of V multiplied by .   * Set J = J+1. |

## 一种快速稀疏表示算法

### 有原型的稀疏表示

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：

(3.12)

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：

(3.13)

其中，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：

(3.14)

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：

(3.15)

这里，是一个稀疏系数向量，通常是一个稠密矩阵,如下图所示。

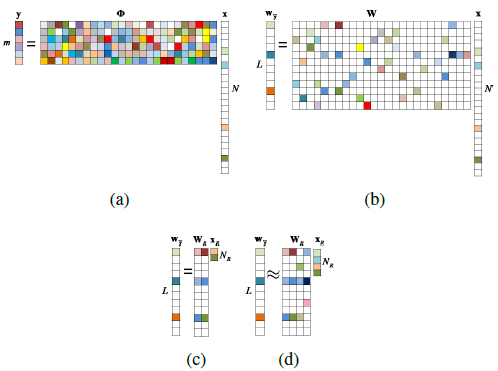


图 3.2

解决逆线性系统

对于上面的公式（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[31]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：

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[26,31,32]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：

这里，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：

其中是容错度。

通过线性转化的特征提取

矩阵, d<<m, 在公式（4）两边分别乘上T得到：

(9)

其中，

### 有原型的快速稀疏近似

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。

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：

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9中的矩阵F和公式11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：

(12)

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（12）可以重写为：

(13)

假定x稀疏，则称s-稀疏（只有s个非零值），那么有

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。

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

因此有

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：

如果字典D可以准确的重建信号，那么公式10中，问题简化为

如图b中所示。

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19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b中所示）。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：

其中，只包含相关列向量，如图c中所示。

## 本章小结

本章主要介绍了稀疏表示和字典学习算法。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。本章还介绍了一种有原型的快速稀疏近似算法，可以大幅提升稀疏表示的速度。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在视频的异常检测中应用稀疏表示。

# 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。异常事件监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[37]和3D时空前景掩码特征[34]。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也有广泛的研究[1,38,39,40]，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。[35]中使用直方图衡量局部光流特征的概率，[36]中通过抽取时空梯度来建立高斯模型，并HMM检测异常，[30]通过混合动态纹理对正常拥挤场景建模，[28]中提出了社会力模型，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。

本章提出的算法描述如下图所示，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。

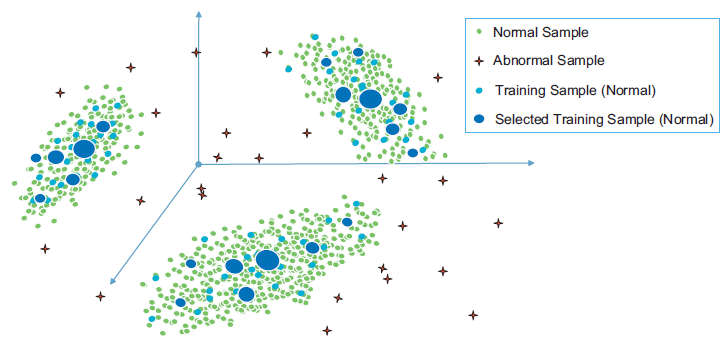


图 4.1

## 稀疏表示实现视频异常检测

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[34,37]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[35]，时空梯度[36]，社会力[28]，混合动态纹理[30]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[5,28,35,36]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）[5]、隐马尔可夫模型（HMM）[36]、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)

其中是一个阀值。

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，本文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。本文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。为了适应GAE和LAE这两种情况，样本基向量的选择可以根据不同情况进行灵活的选取。实验阶段将分别基于UMN和UCSD的视频数据库进行评价。



### MHOF特征提取

对于视频的运动特征计算，主要提取其多尺度光流直方图（MHOF），把视频帧分成若干子区域，每个子区域称之为一个特征单元，如图A中所示。将运动特征的方向分为八个方向，同时对于运动量级根据设定的阀值T分为大尺度（大于T）与小尺度（小于T）两类，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，大尺度使用后面的8个箱。

为了更好的处理局部异常事件（LAE）和全局异常事件（GAE），可以分别采用下面几个不同的策略。对于GAE，主要采用整帧的特征（TYPE A）进行分析，而对于LAE可以选择时间域（TYPE B）或者时空域（TYPE C）进行检测，如下图B所示。

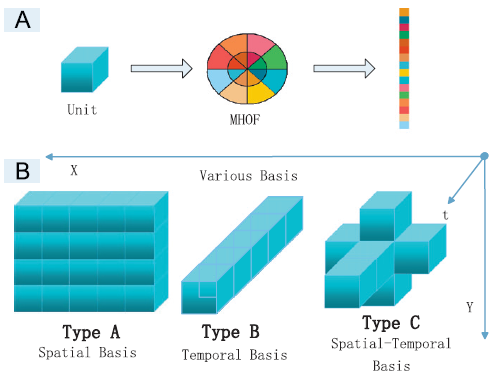


图 4.2

类型A因为是考虑全局的异常情况，所以提取整帧的信息，当然还是进行子区域划分；在考虑局部异常，所以对在每一帧进行划分后，单独分析每个区域的时域特征，如类型B，而类型C则把相邻区域的特征也融合进来，这样增强了关联性，更是考虑了空间上的关系。

### 字典选取

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：

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：

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，本文主要使用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。

### 异常的衡量标准与优化

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。

**关于稀疏系数在正常样本和异常样本的差异表现说明。**

使用K-SVD方法对训练集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。

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：

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。

视频异常检测框架如下所示：

|  |
| --- |
| 视频异常检测框架 |
| 输入：训练字典，测试样本  输出：稀疏表示代价  for i=1,…,N do  //计算重建的稀疏系数矩阵  //计算稀疏表示代价  end for |

WOMP

设计一个光滑处理方法

异常判断

## 使用快速稀疏近似算法提高检测速度

在第三章中提到一种快速稀疏近似算法，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。

对于给定的测试样本y，采用快速稀疏表示算法后，稀疏表示代价函数为：

## 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，而本文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，在不影响检测结果的情况下，很大程度上可以提高检测速度。

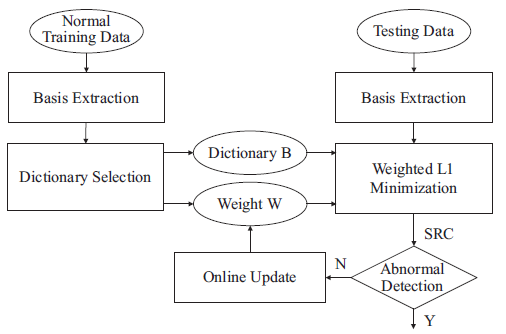
# 系统设计与实验结果



## 引言

## 系统设计

MATLAB



## 实验结果

## 本章小结

# 结论

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。本文主要研究了拥挤场景中的异常事件检测方法。研究得出的结论如下：

1. 研究一种MPPCA特征处理方法并在时空MRF推断异常。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，而本文采用后者进行计算。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。
2. 基于稀疏表示和字典学习算法实现了一个快速的近似稀疏表示算法。
3. 探讨和研究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，而本文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，在不影响检测结果的情况下，很大程度上可以提高检测速度。

# 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. 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.
3. Horn B K P, Schunck B G. Determining optical flow[J]. Artificial intelligence, 1981, 17(1): 185-203.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. Sun D, Roth S, Lewis J, et al. Learning optical flow[J]. Computer Vision–ECCV 2008, 2008: 83-97.
10. 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.
11. Tipping M E, Bishop C M. Mixtures of probabilistic principal component analyzers[J]. Neural computation, 1999, 11(2): 443-482.
12. 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.
13. 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.
14. Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision[J]. International journal of computer vision, 2000, 40(1): 25-47.
15. 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.
16. Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision[J]. International journal of computer vision, 2006, 70(1): 41-54.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. Engan K, Aase S O, Husøy J H. Multi-frame compression: Theory and design[J]. Signal Processing, 2000, 80(10): 2121-2140.
23. 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.
24. 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.
25. Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. Signal Processing, IEEE Transactions on, 1993, 41(12): 3397-3415.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]//Advances in neural information processing systems. 2006: 801-808.

# 哈尔滨工业大学硕士学位论文原创性声明

# 致谢