**哈尔滨工业大学**

**硕士学位论文中期报告**

**题 目：MPPCA与稀疏表示在视频异常检测中的研究**

**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

# Abstract

目录

[摘要 1](#_Toc361641031)

[Abstract 2](#_Toc361641032)

[第1章 绪论 1](#_Toc361641033)

[1.1 课题背景（5k字） 1](#_Toc361641034)

[1.2 课题研究的目的及意义 1](#_Toc361641035)

[1.3 国内外研究现状 2](#_Toc361641036)

[1.4 主要研究内容及论文结构 2](#_Toc361641037)

[第2章 学习运动模式并在时空MRF中推断视频中的异常 4](#_Toc361641038)

[2.1 引言 4](#_Toc361641040)

[2.2 马尔科夫随机场相关理论 5](#_Toc361641041)

[2.2.1 马尔科夫随机场理论 5](#_Toc361641042)

[2.2.2 MRF能量最小化方法 6](#_Toc361641043)

[2.2.3 MRF图模型的表示 8](#_Toc361641044)

[2.3 MPPCA学习局部状态的运动模式 12](#_Toc361641045)

[2.3.1 光流特征提取 12](#_Toc361641046)

[2.3.2 混合概率主成分分析 12](#_Toc361641047)

[2.3.3 学习局部状态的运动模式 13](#_Toc361641048)

[2.4 构建运动模式的时空MRF模型 14](#_Toc361641049)

[2.4.1 时空MRF模型的能量函数 14](#_Toc361641050)

[2.4.2 时空MRF模型的自更新 16](#_Toc361641051)

[2.5 本章小结 16](#_Toc361641052)

[第3章 快速稀疏表示与字典学习的方法 16](#_Toc361641053)

[3.1 引言（5k字） 16](#_Toc361641055)

[3.2 稀疏表示的相关理论 17](#_Toc361641056)

[3.3 字典学习算法 18](#_Toc361641057)

[3.3.1 最优方向法（Method of Optimal Directions, MOD） 18](#_Toc361641058)

[3.3.2 K-SVD算法 19](#_Toc361641059)

[3.4 一种快速稀疏表示算法 20](#_Toc361641060)

[3.4.1 有原型的稀疏表示 20](#_Toc361641061)

[3.4.2 有原型的快速稀疏近似 22](#_Toc361641062)

[3.5 本章小结 23](#_Toc361641063)

[第4章 稀疏表示方法检测视频中的异常 23](#_Toc361641064)

[4.1 引言（5k字） 23](#_Toc361641067)

[4.2 多尺度HOF的特征提取 23](#_Toc361641068)

[4.3 基于稀疏表示的视频异常检测 24](#_Toc361641069)

[4.3.1 初始训练样本的字典学习 24](#_Toc361641074)

[4.3.2 测试样本的稀释表示代价函数计算 24](#_Toc361641075)

[4.4 通过快速稀疏近似算法提高检测速度 25](#_Toc361641076)

[4.5 本章小结 25](#_Toc361641077)

[第5章 系统设计与实验结果 25](#_Toc361641078)

[5.1 引言（5k字） 25](#_Toc361641080)

[5.2 系统设计 25](#_Toc361641081)

[5.3 实验结果 26](#_Toc361641082)

[5.4 本章小结 26](#_Toc361641083)

[结论 26](#_Toc361641084)

[参考文献 26](#_Toc361641085)

[哈尔滨工业大学硕士学位论文原创性声明 28](#_Toc361641086)

[致谢 29](#_Toc361641087)

# 绪论

## 课题背景（5k字）

课题主要研究建立视频流的时空马尔可夫随机场（Space-Time MRF）模型，用于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[11]和图像去噪[1]。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。

目前课题已经完成了光流特征提取和基于高斯混合模型（GMM）的运动区域检测。接下来将重点研究根据光流特征建立MPPCA模型，并进一步建立时空马尔科夫随机场模型。

## 课题研究的目的及意义

## 国内外研究现状

## 主要研究内容及论文结构

# 学习运动模式并在时空MRF中推断视频中的异常



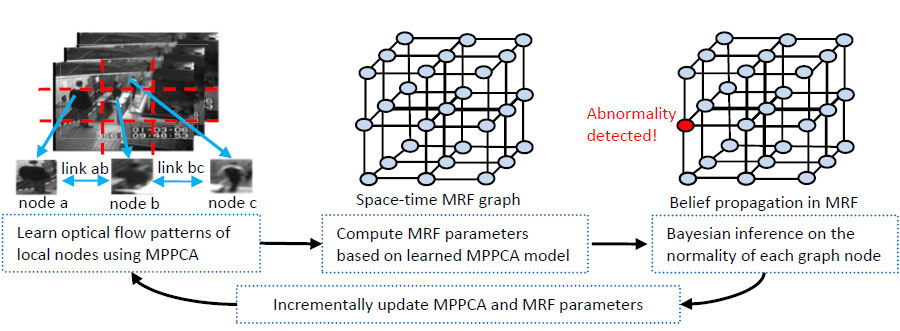
## 引言

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。

首先介绍一个计算机视觉的应用方法，在[]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stereo, optical flow 以及image restoration 等问题。应用的方法如下：首先将问题构建成MRF 的模型，即将问题转化为在MRF 模型下计算Energy Minimization。由于在MRF 中Minimization 问题是NP-hard，因此应用置信传播的演算法做估算。问题如下：

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 labeling产生的能力函数，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label可以产生最小的energy。

正是基于上面的理论基础，本章介绍一种构建时空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，这种模式可以更好的表达运动状态。基于这种运动模式，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。



## 马尔科夫随机场相关理论

### 马尔科夫随机场理论

马尔科夫过程是指下一个时间点的值只与当前值有关，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科夫性，具有这种性质的随机过程叫做马尔科夫过程，其原始的模型就是马尔科夫链。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科夫链，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。

马尔可夫性质：它指的是一个随机变量序列按时间先后关系依次排开的时候，第N+1时刻的分布特性，与N时刻以前的随机变量的取值无关。拿天气来打个比方。如果我们假定天气是马尔可夫的，其意思就是我们假设今天的天气仅仅与昨天的天气存在概率上的关联，而与前天及前天以前的天气没有关系。其它如传染病和谣言的传播规律，就是马尔可夫的。

随机场：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，俗气点说，随机场就是在哪块地里种什么庄稼的事情。

更形象一些，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Labeling问题，更具体点，是通过MAP inference来确定图中每个节点的label。MRF相比其他方法的优势是：1）提供了一种principled method来对Prior knowledge建模，2）MRF可以很容易用定量的方法描述contextual information。 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，如果建立的图模型得当，进而可能获得全局最优解释，这样正是向人类视觉更靠近了一步。

说到MRF的Inference，首先必须有 graph construction, parameter learning,最后才是Inference，图的创建一般是对问题本身的建模，比如在image restoration 和image segmentation中，常用到4-neighborhood或8-neighborhood的pairwise模型，这样，4-或8-相邻的像素中间便用边连接，这样的模型就是paradigmatic pairwise Markov model,如果要加入高阶（>=3)的potential，相当于我们引入了更多的约束，比如：connectivity 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本身的分类结果累构成新的约束。

下面重点说MRF inference问题，即求解能量函数最小化的问题。对于经典的只有unary 和 binary potential的MRF模型，graph cut已经能够在Linear time内进行求解，如果加入更高的potential,虽然问题本身可能变成了NP-hard，仍然有很多近似算法，比如Loopy belief propagation(LBP)、tree-reweighted message passing(TRW)等。撇开这些方法，其实问题的实质是能量最小化。

### MRF能量最小化方法

1. 图割法（Graph Cuts）

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：

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。典型的形式有：

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。

首先介绍什么是割与最小割。割：在一个图G（V，E）中V是点集，E是边集。在E中去掉一个边集C使得G（V，E-C）不连通，C就是图G（V，E）的一个割；最小割：在G（V，E）的所有割中，边权总和最小的割就是最小割。

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。有两种流行的图割算法：swap-move 和 expansion-move。它们都是重复的计算二标号问题的全局最小值，具体算法如下所示：

|  |
| --- |
| Swap algorithm   1. Start with an arbitrary labeling f 2. Set success := 0 3. For each pair of labels   3.1 Find among f’ within one swap of f  3.2 If , set f := and success := 1   1. If success = 1 goto 2 2. Return f |
| Expansion algorithm   1. Start with an arbitrary labeling f 2. Set success := 0 3. For each pair of label   3.1 Find among f’ within one -expansion swap of f  3.2 If , set f := and success := 1   1. If success = 1 goto 2 2. Return f |

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

在马尔科夫随机场中，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：

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

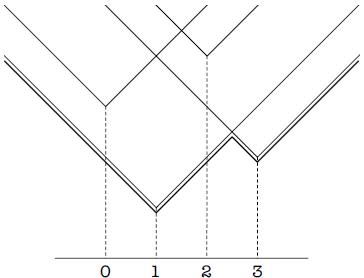
最后每个节点都可算出最小的来决定所要的标号。

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)。[]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cost ，对一般的低层视觉问题代价函数基本上是两个label和的差值。因此，如果我们把这个cost设定成，则cost function 可改写成，其中，如此对k层迭代计算 message 的时间仅需要O(k)。

接着考虑是一个 truncated linear model，即，这里的d即代表线性增加的cost上限。这里考虑两种情形，如果相差超过定的cost 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 message的公式就变成，的关系图如下：



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，2路通过算法如下所示：

For from 1 to k-1:

(forward)，

For from k-2 to 0:

(backward)。

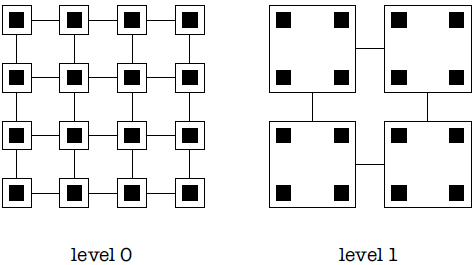
举例来说，一开始最小的消息值为V形顶点的值，在上图中值从0-3 可看成(3,1,4,2)，在forward 的过程后成为(3,1,2,2)，再经过backward后变成(2,2,2,2)，即上图粗线处。这样2-way pass后可以找出对应到最小消息，q的level值。这里2-way 的方法是由于cost定为linear 的公式，因此只要参考左右的最小值就可决定自己本身的min，而演算法所需的时间为O(k)。

在这边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 bipartite graph BP 来加速，在这里将计算时的grid 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BP : 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Coarse-to-fine，将计算的范围由大至小切细成不同大小的MB，计算MB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。



在计算由大到小的消息时，则将打block内每个小block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MB可以有多个造成低cost的level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。

### 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)。

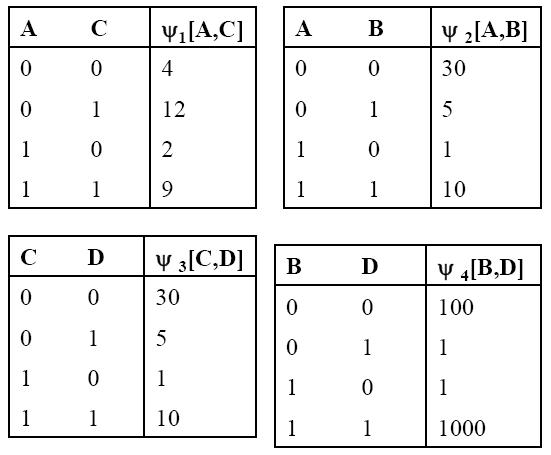
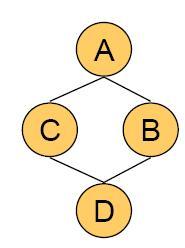
图模型主要有两大类，一类是贝叶斯网络（又称有向图模型）；另外一类是马尔可夫网络（又称无向图模型）。图模型主要有三个主要的关注点：

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。

      这里主要介绍无向图模型的表示，即马尔科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。我们先看一个例子：



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（potential function）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，如右侧所示，必须注意势能函数必须为正。

最终这个无向图模型表示的概率分布是：P(A,B,C,D)= (1/Z)\*/phi(A,C)\*/phi(A,B)\*/phi(C,D)\*/phi(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；/phi 是表示相应的势能函数（这里因为不能表示数学符号，故而用了latex中的符号表示）。

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：

P(**X**)=(1/Z)\*II\_{i=1}^{n}/phi(Ci(x))，其中Ci表示的第i个团（都是利用了latex中符号表示数学公式。

 因子图

// Index of factors and nodes (example)

// n : node, f : factor

///////////////////////////////////////

// n0-f0-n1-f1-n2

// | | |

// f2 f3 f4

// | | |

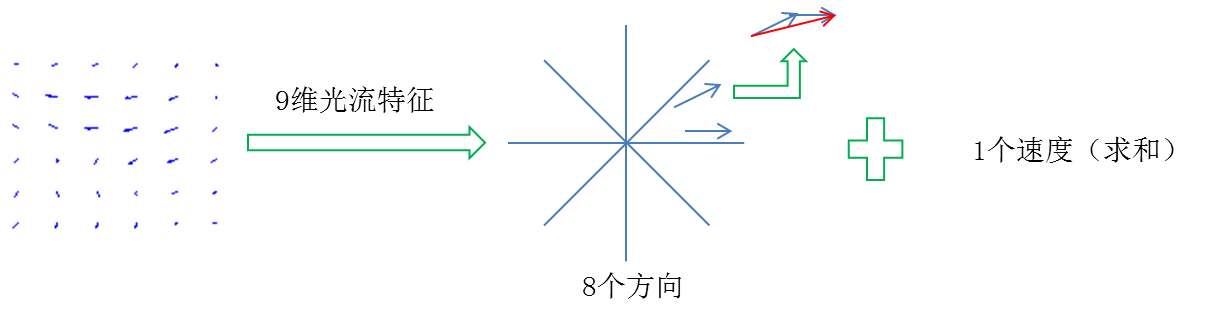
// n3-f5-n4-f6-n7

## 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[2]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[5]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。



### 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 2-1 PCA

PPCA[9]（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[12]）。

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：

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：

(2.1)

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：

(2.2)

(2.3)

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：

(2.4)

(2.5)

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：

(3.1)

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：

(3.2)

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：

(3.3)

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：

(3.4)

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：

(3.5)

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：

(3.6)

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：

(3.7)

根据公式（3.6）和（3.7）可得完整的成对势能函数为：

(3.8)

其中，是一个衡量光滑度的权值常量。

给定了公式（3.5）和（3.8）在节点和链接处的MRF参数，即可通过MAP最大化式（3.1）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的方法进行学习，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。基于这种运动模式，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。

# 快速稀疏表示与字典学习的方法



## 引言（5k字）

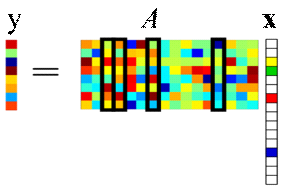
稀疏表示在最近的字典学习和压缩感知的研究中扮演了重要的角色。对于给定的一个基向量集合（如一个字典），找一个信号的稀疏表示经常等同于带有l0范式和l1范式的最优化问题，并最终转化为解决一个欠定的线性系统。然后，每个样本表示成为基向量的一个稀疏线性组合。我们都知道，解决l0范式最小化问题是一个NP难问题并且数值上不稳定。对于提出的一些近似l0范式解决方法如匹配追求和OMP等，尽管这些方法相对简单有效，但它们都不是最优的。最近的关于稀疏编码的研究表明，在某种假设的情况下，解决l0范式最小化问题等价于l1范式最小化问题，而后者可由顶点最优计算。

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，MOD（Method of Optimal Directions）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？



由于m远小于n，用m个方程求解n个未知数，因此y=A\*x是个欠定方程，有无穷多个解。就像我们解优化问题一样，如果我们加上适当的限定条件，或者叫正则项，问题的解会变得明朗一些！这里我们加上的正则项是norm(x,0)，即使重构出的信号 x尽可能的稀疏（零范数：值为0的元素个数），后来Donoho和Elad这对师徒证明了如果A满足某些条件，那么argmin norm(x,0) s.t.y=A\*x 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argmin norm(x,1) s.t.y=A\*x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。

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）。所以，我们想通过训练学习的方法获取字典D,这将是下一小节介绍的内容。

## 字典学习算法

继续上面的问题，在我们的未知量有两个，过完备字典D，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…我们再次列出sparse-land模型： [D,x]=argmin norm(y-D\*x,2)^2 s.t.norm(x,1)<=k。如何同时获取字典D和稀疏系数x呢？方法是将该模型分解：第一步将D固定，求出x的值，这就是你常听到的稀疏分解（Sparse Coding），也就是上一节提到的字典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。

Sparse Coding：x=argmin norm(y-D\*x,2) s.t.norm(x,1)<=k  
Dictinary Update:D=argmin norm(y-D\*x,2)^2

我们主要通过实例介绍三种方法：MOD，K-SVD,Online… 首先是MOD(Method of Optimal Direction)。Sparse Coding其采用的方法是OMP贪婪算法，Dictionary Update采用的是最小二乘法，即D=argmin norm(y-D\*x,2)^2 解的形式是D=Y\*x'\*inv(x\*x’)。

### 最优方向法（Method of Optimal Directions, MOD）

最优方向法是由Engan等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：

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

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

K-SVD对K-Means的扩展

## 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：

(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：

(2)

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：

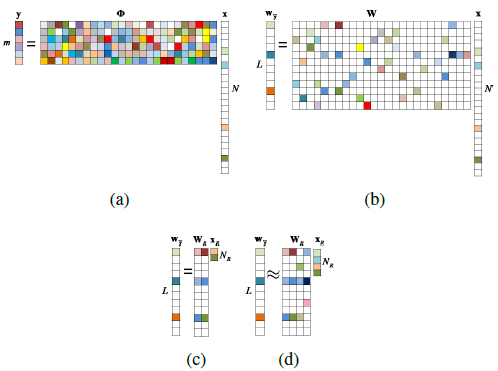
(3)

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：

(4)

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。



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。

假定我们已经通过从样本集中学习得到一个字典D，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

重写公式（12）

(13)

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。

因此

X可以通过解决下面的l1最小化问题得到：

简化为

如图b中所示。

…

如图c中所示

## 本章小结

本章在原有的稀疏表示算法和字典学习算法的基础上，提出了一种有原型的快速稀疏近似算法。

# 稀疏表示方法检测视频中的异常



## 引言（5k字）

本章介绍通过稀疏表示方法实现视频中的异常检测。

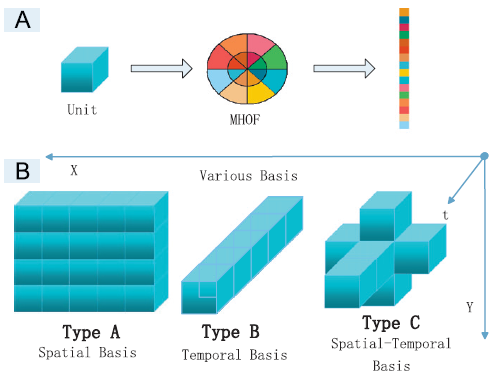
关于平滑处理的方法。

## 多尺度HOF的特征提取

多尺度光流直方图（MHOF）

对于视频中的帧计算它的运动特征，这里获取其光流特征，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。



## 基于稀疏表示的视频异常检测

本节会说明如何应用稀疏表示来检测视频中的异常事件。



### 初始训练样本的字典学习

对于给定的训练集D={x1,x2,…,xN}，N为训练样本总数，xi∈Rd代表一个训练样本，d是特征的维度。

对于训练集D，采用K-SVD算法构建一个字典。

### 测试样本的稀释表示代价函数计算

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对测试集进行稀疏重建，进而得到重建的稀疏系数矩阵X。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。

视频异常检测框架

## 通过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。

## 本章小结

# 系统设计与实验结果

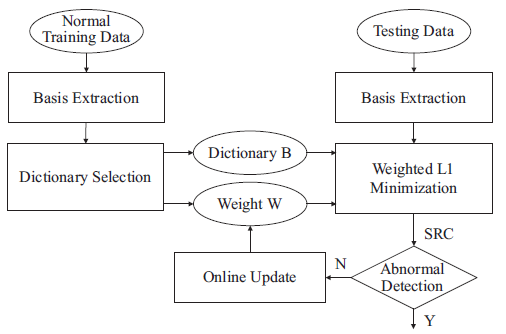


## 引言（5k字）

## 系统设计

MATLAB

SD



## 实验结果

## 本章小结

# 结论

# 参考文献

1. 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.
2. Horn B K P, Schunck B G. Determining optical flow[J]. Artificial intelligence, 1981, 17(1): 185-203.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. Sun D, Roth S, Lewis J, et al. Learning optical flow[J]. Computer Vision–ECCV 2008, 2008: 83-97.
9. 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.
10. Tipping M E, Bishop C M. Mixtures of probabilistic principal component analyzers[J]. Neural computation, 1999, 11(2): 443-482.
11. 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.
12. 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.

# 哈尔滨工业大学硕士学位论文原创性声明

# 致谢