**哈尔滨工业大学**

**硕士学位论文中期报告**

**题 目：MPPCA与稀疏表示分别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**

目录

[第1章 绪论 1](#_Toc360824468)

[1.1 课题背景 1](#_Toc360824469)

[1.2 课题研究的目的及意义 1](#_Toc360824470)

[1.3 国内外研究现状 1](#_Toc360824471)

[1.4 主要研究内容及论文结构 1](#_Toc360824472)

[第2章 学习MPPCA模型并在时空MRF中推断视频中的异常 1](#_Toc360824473)

[2.1 引言 1](#_Toc360824475)

[2.2 局部状态的MPPCA模型 1](#_Toc360824476)

[2.2.1 光流特征提取 1](#_Toc360824477)

[2.2.2 混合概率主成分分析 2](#_Toc360824478)

[2.2.3 学习局部状态的MPPCA模型 2](#_Toc360824479)

[2.3 时空MRF的实现-图模型及信念传播（BP） 3](#_Toc360824480)

[2.3.1 马尔科夫随机场 3](#_Toc360824481)

[2.3.2 时空MRF模型的贝叶斯推理 5](#_Toc360824482)

[2.3.3 时空MRF模型自更新 7](#_Toc360824483)

[2.4 本章小结 7](#_Toc360824484)

[第3章 快速稀疏表示与字典学习的方法 7](#_Toc360824485)

[3.1 引言 7](#_Toc360824487)

[3.2 稀疏表示的相关理论 8](#_Toc360824488)

[3.3 字典学习算法 9](#_Toc360824489)

[3.3.1 最优方向法（Method of Optimal Directions, MOD） 9](#_Toc360824490)

[3.3.2 K-SVD算法 9](#_Toc360824491)

[3.4 一种快速稀疏表示算法 9](#_Toc360824492)

[3.4.1 有原型的稀疏表示 9](#_Toc360824493)

[3.4.2 有原型的快速稀疏近似 11](#_Toc360824494)

[3.5 本章小结 11](#_Toc360824495)

[第4章 稀疏表示方法检测视频中的异常 11](#_Toc360824496)

[4.1 引言 11](#_Toc360824499)

[4.2 多尺度HOF的特征提取 11](#_Toc360824500)

[4.3 稀疏表示的能量函数 12](#_Toc360824501)

[4.3.1 初始训练样本的字典学习 12](#_Toc360824506)

[4.3.2 测试样本的稀释表示能量函数计算 12](#_Toc360824507)

[4.4 本章小结 13](#_Toc360824508)

[第5章 系统设计与实验结果 13](#_Toc360824509)

[5.1 引言 13](#_Toc360824511)

[5.2 系统设计 13](#_Toc360824512)

[5.3 实验结果 14](#_Toc360824513)

[5.4 本章小结 14](#_Toc360824514)

[结论 14](#_Toc360824515)

[参考文献 14](#_Toc360824516)

# 绪论

## 课题背景

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MPPCA模型并在时空MRF中推断视频中的异常



## 引言

## 局部状态的MPPCA模型

### 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，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，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

PCA（Principal Component Analysis）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MPPCA[12]。

### 学习局部状态的MPPCA模型

在初始训练视频中，提取到所有局部区域的描述符（光流特征），并应用于下面定义的MPPCA模型：

(2.1)

其中，：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。各参数使用期望最大化方法计算。

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，也就是说一些大多数观察值在一些局部区域是自由运动状态。

根据上面的学习模型，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。

## 时空MRF的实现-图模型及信念传播（BP）

### 马尔科夫随机场

马尔科夫随机场(MRF)，又称为无向图模型或是马尔科夫网，是一种概率图模型。MRF是一维因果马尔科夫链扩展到二维空间的结果，在图像模型构造和处理过程中效果出众。

若某一随机过程，由一种状态转移至另一种状态的概率大小，仅由这两个状态之间的关系决定，与过去更早的状态或未来的状态等其他因素不相关，称该随机过程为马尔科夫随机过程。

若将tm看成是“现在”状态，则tm+1代表“未来”状态，t1,t2,..., tm-1就表示为“过去”的一些列状态。根据马尔科夫过程的描述，在tm时刻，随机过程ζ(t)取值xm的情况下，ζ(t)的“未来”状态，只有通过“现在”状态才能与“过去”状态产生联系，当“现在”状态被确定下来后，“未来”的状态，就会与“过去”的状态毫无关系。即在已知tm 时刻随机过程所处状态ζ(tm) = xm的条件下，时刻tm以后的过程将要到达的状态与时刻tm 以前的过程所处状态无关。这个性质就称作随机过程的马尔科夫性质。

MRF模型是由一组符合马尔科夫性质的随机变量构成的，可以使用无向图来表示这些随机变量的关系。MRF模型与贝叶斯网络有着相类似的变量关系依赖形式。但是，MRF模型与贝叶斯网络相比，能够表示出具有特定依赖关系（如有环的依赖关系）的变量集合，而贝叶斯网络却对此无能为力。除此之外，能够使用贝叶斯网络表示的特定依赖关系（如节点的父子关系），MRF模型却无法表示。

MRF又被称作是马尔科夫网，与贝叶斯网络不同，MRF是一种无向图，能够通过无向图G(V,E)表示出来，如图2-1所示。



图2-2 MRF模型的无向图表示

在MRF模型中，每个对应的模型节点i 都包含一个观察值Yi和一个隐状态Xi，其中，Xi既可以是连续的随机变量，又可以是离散的值。本文主要是构建一个状态空间为离散值的MRF模型。状态空间为离散值的MRF模型是指，在MRF模型中，每个节点的状态，均取自一组候选的离散值之中，因而，MRF模型的求解，也往往成为确定模型节点状态标记的问题。也就是说，使用一系列的候选值的状态，为每个节点确定一个标记状态，从而使得模型能够取得最大后验概率(MAP)。

定义3.3马尔科夫随机场： 给定无向图G(V,E)，指明变量间的连接关系，再给定随机变量集合，一组离散的状态空间，所有变量组成的状态解空间：，那么称X是一个马尔可夫随机场，当且仅当：

时，对所有

其中， 表示节点i的邻域节点的集合。

按照MRF模型的定义能够推导出下面的性质：

成对的马尔可夫性 当其他变量全部给定后，两个任意不相邻的变量条件独立。记为：当且仅当{u,v}E。

局部的马尔可夫性 当某一变量的邻域变量被给定后，该变量与其他所有变量条件独立。记为：， 表示节点v的所有邻域节点的集合。

整体的马尔可夫性 当变量的割集被确定，任意两个变量的子集条件独立。记为：。

通常情况，在马尔可夫随机场模型中，设节点i处的观测值是，隐状态该观测值间拥有相当程度上的统计相关性，可以定义一个相容性函数 来衡量它们之间的相关性，该相容函数又经常被称为隐状态的“证据”函数。按照对马尔可夫随机场所下的定义，一个隐状态的概率分布同它的邻域节点相关，即无向图的每条链接边都包涵着该边连接的两个变量和间的相关性，可以使用另一个相容性函数 来表示这层相邻节点之间的相关性。那么最后的联合概率分布如式(2-6)：

(2-6)

其中，i, j为模型中所有的相邻节点，而Z是归一化常量。遍历状态空间中所有的X的可能值，使得P(X, Y)得到最大后验概率时的，就是通过马尔可夫随机场模型求出的最优解。

因为马尔可夫随机场模型具有表述问题清晰、模拟结构以及描述数据规律的能力强，所以广泛应用于医学、多媒体应用等诸多领域，特别是计算机视觉领域，MRF有着更为广泛的应用前景。在计算机视觉分析领域，通常使用马尔可夫随机场对视频图像进行处理，在这种应用环境下，马尔可夫随机场模型的变量个数至少会是百万量级的，采用遍历所有解空间的方法，其算法复杂度会随着变量数量的增加而成指数级增长，这就存在着问题求解计算量过大过于复杂的问题。因此，如何高效的求解马尔可夫随机场模型逐渐成为了使用马尔科夫随机场模型解决问题时可避免的一个问题。

### 时空MRF模型的贝叶斯推理

本文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

# 快速稀疏表示与字典学习的方法



## 引言

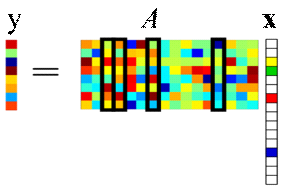
稀疏表示在最近的字典学习和压缩感知的研究中扮演了重要的角色。对于给定的一个基向量集合（如一个字典），找一个信号的稀疏表示经常等同于带有l0范式和l1范式的最优化问题，并最终转化为解决一个欠定的线性系统。然后，每个样本表示成为基向量的一个稀疏线性组合。我们都知道，解决l0范式最小化问题是一个NP难问题并且数值上不稳定。对于提出的一些近似l0范式解决方法如匹配追求和OMP等，尽管这些方法相对简单有效，但它们都不是最优的。最近的关于稀疏编码的研究表明，在某种假设的情况下，解决l0范式最小化问题等价于l1范式最小化问题，而后者可由顶点最优计算。

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）

目标方程

### 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。

## 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。

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：

其中，

### 有原型的快速稀疏近似

## 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

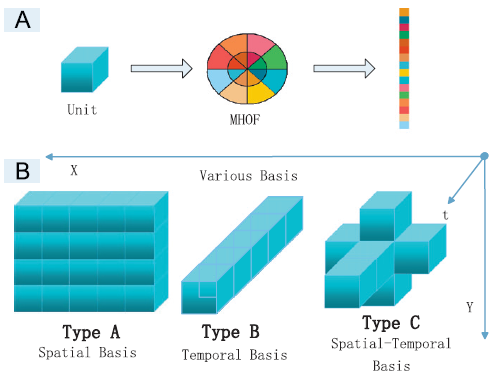


## 引言

本章介绍通过稀疏表示方法实现视频中的异常检测。

## 多尺度HOF的特征提取

多尺度光流直方图（MHOF）



## 稀疏表示的能量函数

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能量函数：

稀疏表示能量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。

下面阐述对应的贝叶斯决策过程。



### 初始训练样本的字典学习

### 测试样本的稀释表示能量函数计算

## 本章小结

# 系统设计与实验结果

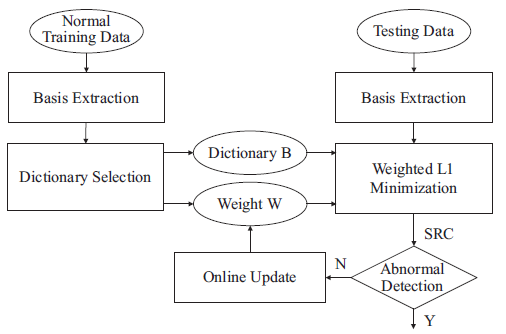


## 引言

## 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.