

# 基于线性动态系统稀疏编码的异常事件检测

刘 洋<sup>1</sup> 李一波<sup>2</sup>

(南京航空航天大学自动化学院 南京 210016)<sup>1</sup> (沈阳航空航天大学自动化学院 沈阳 110136)<sup>2</sup>

**摘 要** 线性动态系统模型结合稀疏编码实现异常事件检测。线性动态系统可有效地捕捉动态纹理在时间和空间的转移信息,描述视频的时空小块。然而,线性动态系统属于非欧氏空间,无法直接用传统的稀疏编码进行异常检测。基于约束凸优化公式,将相似性变换与稀疏编码结合,可实现线性动态系统稀疏编码的优化求解。实验表明,所提出的方法具有更好的性能。

**关键词** 稀疏编码,异常事件检测,线性动态系统,相似性变换

**中图分类号** TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.10.063

## Abnormal Event Detection Using Linear Dynamical System Combined with Sparse Coding

LIU Yang<sup>1</sup> LI Yi-bo<sup>2</sup>

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)<sup>1</sup>

(College of Automation, Shenyang Aerospace University, Shenyang 110136, China)<sup>2</sup>

**Abstract** Linear dynamical system model combined with sparse coding is used to achieve abnormal event detection. Linear Dynamical System(LDS) as a description for dynamic texture can capture the transition of appearance and motion effectively. LDS is applied to describe spatio-temporal cuboids. Since LDS does not belong to Euclidean space, traditional sparse coding techniques can not be applied. Similarity transformation combined with sparse coding based on a principled convex optimization formulation can deal with the optimization of sparse coding. The results show that the proposed algorithm has better performance and outperforms the earlier approaches.

**Keywords** Sparse coding, Abnormal event detection, Linear dynamical system, Similarity transformation

## 1 引言

视频中异常事件检测是计算机视觉领域一个热点话题,至今已有很多方法被提出,包括基于跟踪的方法、基于运动的方法、基于时空的方法。基于跟踪的方法也称为基于轨迹的方法,主要是通过观察目标的空间位置判断其异常性,由目标的外观变化或运动变化带来的异常是不能检测的,因此这样的方法适用于单一或少量目标前景的研究。在实际的监控应用环境中,背景是复杂的,并且有多个目标,目标之间有交叉、遮挡、噪声等因素,这使得跟踪场景中每一个目标存在很大的困难。相对于基于跟踪的方法,很多学者应用基于运动的方法获得了较好的检测效果,例如光流的方法和相位相关的方法<sup>[1]</sup>。文献[1]研究在复杂背景中的实时异常事件检测,提出了时空相关混合高斯模型,将每帧图像分割为不重叠的小块,计算两帧相应位置小块的相位相关,建立正常事件和异常事件的时空混合高斯模型。基于运动的方法聚焦在运动信息上,没有考虑目标的空间信息,无法识别由目标的外观变化带来的异常情况,并且在拥挤的环境中是不适合的。基于时空的方法是应用较多,且能获得较好效果的方法,例如文献[2]提出的时空兴趣点属于对视频的局部描述,在分类识别和异常事件检测中应用广泛。在拥挤的环境中,无法建立视频

的整体特征,也很难准确获得所有运动目标,而时空兴趣点就是视频中动态的时空立体块,正是研究者感兴趣的部分。

时空方法在提取时空立体块之后,准确地描述时空立体块,这对于最后的检测非常重要。描述时空块的方法有很多,如光流直方图、梯度直方图或直接将时空块的像素串联为列向量。效果更好的方法分别捕捉和计算空间和时间轴的特性即观测特征和运动特征,然后将其混合串联为一个描述符作为描述时空块的特征,或者是直接扩展 2D 特征到 3D 特征,如 3D-SIFT (Scale-Invariant Feature Transform)、3D-HOG (Histograms of Oriented Gradients)。然而,每个视频提取的时空块成千上万,使得计算负担非常重,且虚警率高。特征必须能准确有效地描述视频数据在时间和空间中动态和外观的变化。线性动态系统用于描述动态纹理特征,同时捕捉目标的外观和运动特征,很好地描述视频块在时间和空间上的变化,可以同时识别由于时间轴或空间轴变化的异常。文献[3]中应用线性动态系统成功地捕捉了动态纹理的时空变化,参数矩阵  $A$  和  $C$  描述系统运动和外观变化,获得了很好的识别率。

本文应用线性动态系统描述视频时空块,对每一个训练样本和测试样本建模,得到模型参数  $\{A, C\}$ 。建立数据模型后,要计算样本之间的距离以进行分类识别。因为线性系统

到稿日期:2013-12-02 返修日期:2014-02-23 本文受国家自然科学基金青年基金项目(61103123)资助。

刘 洋(1977—),女,博士生,主要研究方向为视频异常事件处理、图像分类, E-mail: yang97\_net@163.com; 李一波(1952—),男,博士后,教授,主要研究方向为图像处理、模式识别。

模型不属于欧氏空间,应用马丁距离(Martin distance)或核计算样本之间的相似性,应用最近邻法(NN)或支持向量机(SVM)进行分类。这些方法的缺点是,随着训练样本数量的增加,计算量增大而致分类识别无法实现;另一方面,其要求纹理固定视角。为了解决视角问题,文献[4]将动态纹理模型和词袋方法结合。应用词袋方法,测试数据不用与训练数据中每个样本计算距离,即使训练数据不断增加,同样可以有效地实现分类识别。而其中重要的问题是如何建立词袋,因为线性系统不属于欧氏空间,因此不能用常规的方法获得词袋,用MDS(Multidimensional Scaling)或Isomap方法将模型降维,用低维空间的向量进行聚类,得到聚类中心<sup>[4]</sup>。然而,没有从低维空间映射回线性系统的方法,因此选择相应的低维向量与聚类中心最近的线性系统作为动态系统的词袋,属于一种近似的方法,无法从理论上加以证明。文献[5]和文献[6]提出新的建立系统词袋的方法。线性系统模型结合词袋的方法,适用于局部信息描述整体特征,即一个视频中提取很多个时空块,计算每个时空块的系统模型与词袋中的词之间的距离,判断该时空块归属于哪个词,从而计算出词在每个视频中发生的频率,将这个频率的直方图作为描述整体视频的特征进行分类识别。

在异常事件检测中,提取的时空块本身尺寸较小,不适用于在其中提取更小的时空块,这样,应用词袋的方法是无法进行分类和检测的;另一方面,异常事件检测面临的一个重要困难是数据量的巨大。直接将模型参数作为描述感兴趣时空块的特征进行分类或检测,利用少量具有重要信息的数据描述测试数据,即应用稀疏编码(Sparse Coding)解决上述问题。稀疏编码是最近被广泛采用的分类识别方法,与传统的奈奎斯特采样定理不同,它是人为定义或训练一个超完备字典,通过优化迭代算法将高维数据映射到低维空间中,得到一个编码系数向量,而这个编码系数是稀疏的,且能表征数据的主要特征。因此,稀疏编码用少量的字典基向量重构测试数据。2009年,Wright和Yang<sup>[7]</sup>等人验证了稀疏编码在人脸识别领域具有很好的性能,提高识别率的同时在大量噪声和遮挡的情况下仍具有很好的识别率,并将稀疏编码引入模式识别和语义理解领域。此后,很多学者进行深入研究和拓展,其中文献[8]应用稀疏编码实现基于线性动态系统的动态纹理识别,在概率框架下获得了较好的识别率。将稀疏编码应用到线性动态系统的动态纹理识别中,能提高识别率,减小计算负担且对噪声具有鲁棒性。

综上所述,线性动态系统具有较强的描述能力,稀疏编码用少量数据重构测试样本,将二者结合有利于实现异常事件检测。本文提出用于异常事件检测的线性动态系统稀疏编码方法,其核心内容分为稀疏编码的优化求解和字典的训练学习两部分。第一部分线性动态系统稀疏编码的优化求解,不能直接应用传统的稀疏编码,本文提出将控制理论中的相似性变换引入到稀疏编码学习中,稀疏编码重构的样本模型等价于原模型线性变换空间中的一个模型,这样,优化求解即为优化编码系数和线性变换矩阵的过程,应用约束凸优化公式实现稀疏编码的优化求解。第二部分是稀疏编码字典的训练学习,时空块的数目比较大,一次性投入到算法中从计算角度很难实现,需要压缩字典的尺寸。本文提出将样本分为若干子集,在每个子集中,提出一个样本,剩下样本作为字典,学习编码系数向量,此过程在整个集合中循环,统计每个样本被用

于重建的次数,选择被应用次数高的若干样本作为子集的字典。所有子集的字典组合作为最后整体算法的字典。线性动态系统作为时空块的数学模型,提高了描述的准确性,压缩了字典尺寸使得计算负担小,应用控制理论中的线性变换解决了线性动态系统非欧氏空间的计算问题。

## 2 时空块的线性动态系统描述

线性时不变动态系统,简称线性动态系统,它作为线性高斯状态空间模型,由式(1)中两个方程表示:

$$\begin{cases} x_{t+1} = Ax_t + Bv(t) \\ y_t = Cx_t + w(t) \end{cases} \quad (1)$$

两个方程为状态方程和观测方程,这里 $A, B, C$ 是不依赖于输入的常数矩阵, $y(t) \in R^m, x(t) \in R^n, A \in R^{n \times n}, C \in R^{m \times n}, B \in R^n$ 。状态转移矩阵 $A$ 描述隐含状态 $x(t)$ 在时间域内的行为,对于动态纹理来说它捕捉视频的运动信息。观测矩阵 $C$ 描述了隐含状态到输出变量 $y_t$ 的转移过程,捕捉视频的外观变化信息。 $v(t)$ 是过程噪声, $w(t)$ 是观测噪声,分别是具有协方差矩阵 $Q$ 和 $R$ 的零均值正态分布的随机变量,即 $v(t) \sim N(0, Q)$ 和 $w(t) \sim N(0, R)$ 。因此,线性动态系统的模型参数可以表示为 $\{A, B, C, Q, R\}$ 。可以说,线性动态系统模型是外观和运动的联合模型。当线性动态系统模型应用于动态纹理识别时,给定一个纹理视频,模型中的 $y(t)$ 即是一帧图像的像素灰度值串连而成的向量,建立一个动态纹理的线性动态系统模型意味着计算模型参数 $\{A, B, C, Q, R\}$ 。本文采用线性动态系统描述时空兴趣点,并采用文献[2]提出的方法提取时空兴趣点,如图1所示。

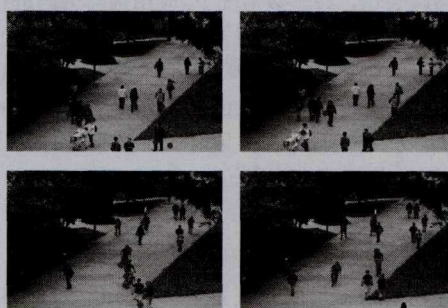


图1 采用文献[2]方法所提取的时空兴趣点

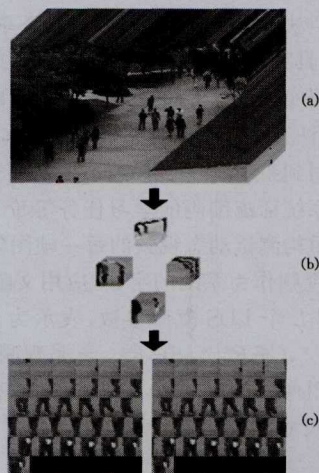


图2 (a)输入视频;(b)提取的时空小块举例;(c)左图为时空小块原始图像,右图为线性动态系统建立的时空小块图像



提取兴趣点之后,围绕兴趣点一定尺寸提取时空小块,用线性动态系统建模这些时空立体块,图2表示在视频中提取感兴趣的时空块,然后用线性动态系统描述。从图中可以看出,线性动态系统准确地捕捉了视频的动态和外观变化信息,具有很好的描述能力。

### 3 线性动态系统的稀疏编码

异常事件检测是计算机视觉领域的一个研究热点,至今尚未被很好地解决,存在的主要挑战是数据量巨大,而稀疏编码用少量包含重要信息的字典基向量重构测试数据可以解决海量数据的问题。线性动态系统模型能够很好地描述视频数据不确定的空间和时间变化转移特性,有利于分析数据在空间和时间轴的异常。因此,本文提出一种基于线性动态系统稀疏编码的异常事件检测,并获得了较好的性能。

#### 3.1 稀疏编码

稀疏编码作为分类器,一个测试样本可以由字典中具有代表性的样本线性组合而重构,字典可以人为定义或通过训练学习获得。稀疏编码用公式表达为式(2):

$$\alpha^* = \arg \min_{\alpha} \|y - D\alpha\|_2 + \lambda \|\alpha\|_1 \quad (2)$$

其中,  $y$  是测试样本,  $D$  是字典,  $\alpha$  是重构系数。第一部分  $\|y - D\alpha\|_2$  是重构误差项,基于字典能表达训练数据知识的假设,这一项值应该很小。第二部分  $\lambda \|\alpha\|_1$  是稀疏项,基于假设信号  $y$  是由字典中与测试样本来源于同一类的项线性组合,  $\alpha$  应该是稀疏的。稀疏编码方法进行识别需要采用最优重构系数  $\alpha^*$ , 它使得目标函数  $J = \|y - D\alpha\|_2 + \lambda \|\alpha\|_1$  达到最小。一些文献中利用获得的最优系数  $\alpha^*$  作为特征描述符,称为稀疏表达。在多数情况下,应用稀疏编码进行分类识别,重构误差  $e_i = \|y - D_i \alpha_i\|_2$  决定测试样本的类别,  $i=1, 2, \dots, N$ ,  $N$  是类别数,  $D_i \in R^{m \times k}$  是字典中第  $i$  类样本数据集,  $k$  表示第  $i$  类样本的数目,  $D_i$  的每一列代表第  $i$  类的一个样本,  $D = [D_1, D_2, \dots, D_N]$ 。  $\alpha_i$  是对应于类别  $i$  的编码系数向量。从而测试样本可以利用公式  $identity(y) = \arg \min_i \{e_i\}$  进行分类。

#### 3.2 线性动态系统的稀疏编码

文献[9]采用多尺度光流直方图(Multi-scale Histogram of Optical Flow, MHOF)描述时空块,文献[10]采用梯度直方图(HOG)和光流直方图(Histogram of Optical Flow, HOF)混合特征描述时空块。而本文采用的线性动态系统中的状态方程和观测方程,具有更强的时空信息描述能力。线性动态系统结合稀疏编码实现异常事件检测,然而,线性动态系统不属于欧氏空间,无法直接应用稀疏编码求解。本文提出稀疏编码模型可以映射到系统线性变换空间中,从而获得求解方法。

线性动态系统稀疏编码的学习任务等价于用  $L$  个 LDS 模型尽可能地重构测试动态视频的每一帧图像<sup>[8]</sup>, 见式(3)。给定  $L$  个动态视频作为系统的字典,应用文献[3]所提出的方法,学习得到  $L$  个 LDS 模型参数,表示为  $M = \{(A_i, B_i, C_i)\}_{i=1}^L$ ,  $A \in R^{n \times n}$ ,  $C \in R^{m \times n}$ ,  $B \in R^n$ 。  $m$  是观测特征向量的尺寸或一帧视频图像的像素个数,  $n$  是 LDS 模型的尺寸,等于模型参数  $A$  的尺寸,也等于隐含变量的个数。

$$\begin{cases} \dot{\vec{x}}_{t+1} = \sum_{i=1}^L \alpha_i A_i \vec{x}_t + \sum_{i=1}^L \alpha_i B_i \vec{v}(t) \\ y_t = \sum_{i=1}^L \alpha_i C_i \vec{x}_t + \vec{w}(t) \end{cases} \quad (3)$$

稀疏编码需要求解系数向量  $\alpha$ , 而式(3)属于非欧氏空间模型,无法直接用传统的稀疏编码求解方法,本文应用控制理论中的线性变换解决求解问题。

当给定一个线性动态系统时,模型表示为  $\Sigma(A, B, C)$ , 见式(1), 状态变量的选择并非唯一的, 当  $x(t)$  选取不同时, 状态变量之间存在一个坐标变换, 即  $\hat{x} = P^{-1}x$ ,  $P$  是线性非奇异变换矩阵, 则有相似性线性变换, 见式(4), 得到线性系统  $\hat{\Sigma}(\hat{A}, \hat{B}, \hat{C})$ 。

$$\begin{cases} \dot{\hat{x}}_{t+1} = P^{-1}AP \hat{x}_t + P^{-1}B \vec{v}(t) \\ y_t = CP \hat{x}_t + \vec{w}(t) \end{cases} \quad (4)$$

则线性系统  $\Sigma(A, B, C)$  在线性变换矩阵  $P$  下的各个矩阵具有式(5)的对应关系:

$$\hat{A} = P^{-1}AP, \hat{B} = P^{-1}B, \hat{C} = CP \quad (5)$$

$P \in GL(n)$ ,  $GL(n)$  是非奇异  $n \times n$  矩阵组成的空间。所有通过线性变换得到的系统组成一个大家庭, 大家庭中的每个成员之间是相似系统, 这属于控制理论范畴的知识。相似系统的特点是具有不同的转移矩阵和隐含状态, 但是拥有相同的输出变量。一个家庭中的相似系统具有相同的输出变量意味着家庭成员可以描述同一个视频, 尽管它们的模型参数不同。由非奇异矩阵  $P$  变换得到的所有系统所组成空间称为线性变换空间。线性变换就是选择不同的隐含状态来描述系统, 并未改变系统的本质。

仔细观察式(3)和式(4), 不难发现它们具有相同的形式, 换句话说, 用稀疏编码重构的系统等价于相似性变换空间大家庭中的一个成员。更详细地解释, 给定一个测试 LDS 模型  $\{A, B, C\}$ , 用训练集中模型的动态转移和外观变化过程直接重构测试模型的动态转移和外观变化过程几乎是不可能的, 即

$$B_{test} \neq \sum_{i=1}^L \alpha_i B_i, A_{test} \neq \sum_{i=1}^L \alpha_i A_i, C_{test} \neq \sum_{i=1}^L \alpha_i C_i$$

但是用式(3)可以重构测试动态视频的每一帧图像, 重构是有误差的, 对于包含大量运动的动态视频而言, 甚至误差很大。然而, 由于空间或时间的变化特性, 异常数据将产生更大的误差, 从而能够进行异常事件检测。这种由稀疏编码重构的过程等价于相似性变换的过程, 因此得到式(6)所描述的变换关系:

$$\sum_{i=1}^L \alpha_i A_i \cong P^{-1}AP, \sum_{i=1}^L \alpha_i B_i \cong P^{-1}B, \sum_{i=1}^L \alpha_i C_i \cong CP, \hat{x} = P^{-1}x \quad (6)$$

即在线性变换空间能够找到一个线性变换模型, 等于或无限接近于稀疏编码重构的系统, 则稀疏编码公式问题转化为式(7), 如下所示:

$$J = \|CP - \sum_{i=1}^L \alpha_i C_i\|_F^2 + \|P^{-1}AP - \sum_{i=1}^L \alpha_i A_i\|_F^2 + \|P^{-1}B - \sum_{i=1}^L \alpha_i B_i\|_F^2 \quad (7)$$

式(7)解释为在系统的线性变换空间中, 能够找到一个线性非奇异矩阵  $P$ , 经过线性变换, 等价于或无限接近于稀疏编码重构的系统, 它们是等价系统。因此, 稀疏编码的求解任务转化为优化求解矩阵  $P$  和编码系数  $\alpha$ 。求解中有一个难题, 即线性变换空间是无穷的, 从理论和计算的角度是无法实现的。解决的办法是抛弃空间  $GL(n)$ , 压缩线性变换空间, 本文

只研究最紧凑子空间即  $n \times n$  矩阵  $O(n)$  空间, 在  $O(n)$  中  $P^T P = I$ , 在这个空间中最小化式(7)。

### 3.3 线性动态系统的稀疏编码的优化求解

优化求解的任务是学习参数  $\alpha^*$ 、 $P^*$ , 即优化式(7),  $(\alpha^*, P^*) = \arg \min_{\alpha, P} J(\alpha, P)$ 。求解的难点在于非线性  $P^{-1}AP$  项和矩阵相乘的项。本文优化求解的策略是固定一个参数, 求解另一个参数。从公式可以看出, 当变换矩阵  $P$  固定时, 式(7)优化编码系数  $\alpha$  是凸优化问题; 当  $\alpha$  固定时, 式(7)关于  $P$  是非线性的。优化过程即在这两个变量交替固定时, 最小化式(7)求解得到另一个参数。但是存在的问题是, 这样的求解得到的是局部最优, 且优化的结果与选择的初值关系密切。

优化求解编码系数  $\alpha$ ; 优化学习参数  $\alpha$ , 需要固定线性非奇异矩阵  $P$ , 优化问题则转化为传统的稀疏编码求解, 见式(8):

$$\min_{\alpha} \lambda_C \|CP - \sum_{i=1}^L \alpha_i C_i\|_F^2 + \lambda_A \|P^{-1}AP - \sum_{i=1}^L \alpha_i A_i\|_F^2 + \lambda_B \|P^{-1}B - \sum_{i=1}^L \alpha_i B_i\|_F^2 + \lambda_{\alpha} \|\alpha\|_1 \quad (8)$$

将式(8)变形为:

$$\min_{\alpha} \lambda_C \|C' - D_C \alpha\|_F^2 + \lambda_A \|A' - D_A \alpha\|_F^2 + \lambda_B \|B' - D_B \alpha\|_F^2 + \lambda_{\alpha} \|\alpha\|_1 \quad (9)$$

其中,  $D_C = [C_1, C_2, \dots, C_L]$ ,  $D_B = [B_1, B_2, \dots, B_L]$ ,  $D_A = [A_1, A_2, \dots, A_L]$ ,  $\alpha = [\alpha_1; \alpha_2; \dots; \alpha_L]$ ,  $C' = CP$ ,  $A' = P^{-1}AP$ ,  $B' = P^{-1}B$ 。已有一些方法求解上述优化问题, 如 generic QP solvers (如 CVX)、内点法、修正的最小二乘角回归方法 (LARS) 和移植方法。本文采用文献[11]提出的 feature-sign 搜索方法求解上述  $l_1$  优化问题。

优化求解学习相似性变化矩阵  $P$ ; 学习参数  $P$ , 需要固定非奇异矩阵  $\alpha$ , 此时优化问题等价于最小化式(10):

$$\min_P \lambda_C \|CP - \sum_{i=1}^L \alpha_i C_i\|_F^2 + \lambda_A \|P^{-1}AP - \sum_{i=1}^L \alpha_i A_i\|_F^2 + \lambda_B \|P^{-1}B - \sum_{i=1}^L \alpha_i B_i\|_F^2 \quad (10)$$

在最紧凑的子空间中寻找  $P$ , 即  $n \times n$  矩阵空间  $O(n)$ , 在  $O(n)$  中  $P^T P = I$ , 在这个空间中最小化式(10)。优化的方法应用文献[12]提出的快速 Jacobi-type 算法。

这种固定一个参数而优化另一个参数的方法将收敛于局部最优, 因此, 变量初值的选择非常重要。在实验中, 首先固定  $\alpha$ , 优化  $P$ 。  $\alpha$  初值的选择参考文献[8],  $\alpha^{(0)}$  中对应测试样本与  $M$  训练样本距离最近的若干个位置设为 1, 其它位置设为 0, 距离计算采用马丁距离, 本文中  $\alpha^{(0)}$  非零元素的个数设为 10。

### 4 异常事件检测

稀疏编码字典的学习: 对视频数据进行异常事件检测时, 将检测出大量的时空兴趣点, 提取的时空块样本数目巨大, 很多样本之间彼此是相似的, 一个时空块可以由其它若干时空块线性组合而重构。在这些数目巨大的样本中, 选择那些具有代表性的样本作为稀疏编码的字典。给定正常事件样本集合  $S$ , 将其分为  $N$  个子集  $S = \{S_1, S_2, \dots, S_N\}$ ,  $S_i = \{M_1^i, M_2^i, \dots, M_n^i\}$ ,  $M_j^i$  代表第  $i$  个子集中第  $j$  个样本的线性动态系统模型。在一个子集  $S_i$  中, 选择一个样本  $M_j^i$  作为测试样本, 剩下的样本作为字典, 应用稀疏编码进行重构:

$$M_j^i = \sum_{m=1, m \neq j}^n \alpha_m^i M_m^i \quad (11)$$

计算系数向量  $\alpha_j^i$ , 即优化式(7)。  $\alpha_j^i$  中非零元素的位置对应到字典中的样本是用于重构  $M_j^i$  的, 零元素表示字典中相应位置的样本与  $M_j^i$  不相似或相似很小。这样的重构工作在整个  $S_i$  子集的每个样本中循环, 将得到  $n$  个系数向量  $\{\alpha_1^i, \alpha_2^i, \dots, \alpha_n^i\}$ , 统计每个样本被用于重构其它样本的次数, 选择次数多的若干样本作为子集  $S_i$  中选出的子集字典  $D_i$ 。在所有子集中循环上述过程, 将得到的  $N$  个子集字典合并在一起形成最终字典  $D = \{D_1, D_2, \dots, D_N\}$ 。图 3 描述了在子集数据中学习字典的过程, 在划分的每个子集数据中都将学习得到这样的一个子集字典, 所有的子集字典组合在一起称为稀疏编码检测的字典。

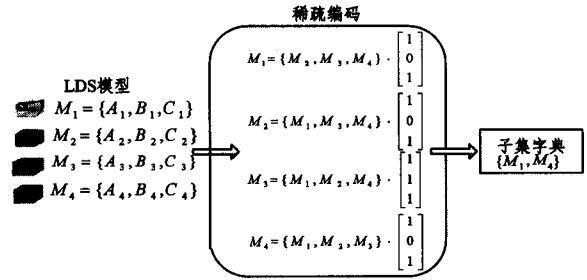


图 3 子集字典的学习过程

实验采用软件 MATLAB 7.0, 计算机配置为 2.4GHz CPU, 2GB RAM, 学习一个视频数据的编码系数  $\alpha$  平均需要 1.2 秒。建立字典所需时间根据数据库中提取的时空块数量和线性动态系统尺寸的不同而变化。本文中 UCSD 数据库字典的建立需要 1.9 个小时, UMN 数据库 3 个场景字典的建立平均需要 9.2 分钟。

异常事件检测: 综述前面内容, 给定一个新的观测事件  $y'$  和字典  $D$ , 本文提出的方法优化学习结束后, 得到最优的系数向量  $\alpha^*$ ,  $y'$  被检测为异常事件时满足以下条件:

$$J(y', \alpha^*, D) > \epsilon \quad (12)$$

这里  $\epsilon$  是人为定义的阈值,  $\epsilon$  的大小能控制算法对异常事件的敏感性。

### 5 实验及结果分析

为了验证方法的有效性, 分别在 UCSD Ped1 数据库<sup>[13]</sup> 和 UMN 数据库<sup>[14]</sup> 进行实验分析。UCSD 数据库属于局部异常事件检测, 即在视频图像中同时存在正常和异常事件; UMN 数据库属于整体异常事件检测, 即在视频图像中, 只出现正常事件或异常事件, 这两个数据库包含了典型的异常情况, 并用像素级和图像帧级 ROC 曲线定性和定量地验证方法的性能。

UCSD 数据库由 34 个正常事件视频和 36 个异常事件视频组成。每个视频有 200 帧, 图像的尺寸是  $238 \times 158$ 。正常事件是在人行道上的行人, 而自行车、滑板、汽车、行人在草坪上等属于异常事件。实验过程为: (1) 检测时空兴趣点, 提取时空块的尺寸是  $13 \times 13 \times 31$ ; (2) 用线性动态系统描述时空块, 得到模型参数  $\{A, B, C\}$ ; (3) 用正常事件建立字典, 字典尺寸为 1000; (4) 计算系数向量  $\alpha$  和重构残差  $J$ , 并与阈值  $\epsilon$  比较, 判断异常。提取兴趣点时所用的参数  $\sigma=1, \tau=2.5$ , 并为高斯和 Gabor 滤波器设定阈值 0.02。应用式(12)进行检测, 结果如图 4 所示, 阈值根据经验人为主观设定, 本文中阈值设



定为  $\epsilon=1.5 \times 10^3$ , 正常数据的重构误差多数在  $10^2$  数量级, 而异常数据的重构误差在  $10^4$  数量级及以上。算法能检测出自行车、滑板、汽车以及走在草坪上的人。评价算法性能用图像帧级 ROC 曲线和像素级 ROC 曲线, 并与 MDT 方法、Social force 和 MPPCA 等方法作比较。



图4 异常事件检测结果举例

如图5所示, 从 ROC 曲线可以看出, 本文提出的方法比其它方法具有更好的性能。检测结果的定量分析用图像帧级 EER(the Equal Error Rate) 和 AUC(Area Under Curve) 表示, 结果如表1所列, 像素级 EER 和 AUC 结果如表2所列。图像帧级检测结果中, EER 为 13.89%, 比文献[9]中 EER (19%) 低, ROC 曲线的面积 AUC 为 93.24%, 比文献[9]中 AUC(90.99%) 高, 表明本文方法具有较好的性能。

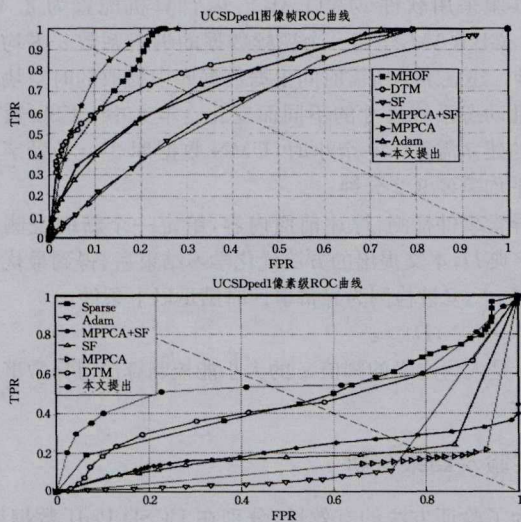


图5 UCSD 图像帧级和像素级 ROC 曲线与其它方法进行比较

表1 UCSD 图像帧级 EER 和 AUC 与其它方法进行比较

|                            | EER    | AUC    |
|----------------------------|--------|--------|
| Adam <sup>[15]</sup>       | 38%    | 77.05% |
| SF <sup>[16]</sup>         | 31%    | 58.24% |
| MPPCA <sup>[16]</sup>      | 40%    | 67.03% |
| MPPCA+SF <sup>[16]</sup>   | 32%    | 76.96% |
| DTM <sup>[16]</sup>        | 25%    | 83.77% |
| MHOF+Sparse <sup>[9]</sup> | 19%    | 90.99% |
| 本文方法                       | 13.89% | 93.24% |

表2 UCSD 像素级 EER 和 AUC 与其它方法进行比较

|                            | EER    | AUC    |
|----------------------------|--------|--------|
| Adam <sup>[15]</sup>       | 76.49% | 13.33% |
| SF <sup>[16]</sup>         | 78.53% | 17.99% |
| MPPCA <sup>[16]</sup>      | 82.58% | 20.57% |
| MPPCA+SF <sup>[16]</sup>   | 72.52% | 21.36% |
| DTM <sup>[16]</sup>        | 55.35% | 44.13% |
| MHOF+Sparse <sup>[9]</sup> | 54.25% | 47.05% |
| 本文方法                       | 46.60% | 55.48% |

从定性和定量分析结果可以看出, 对于局部异常事件检测, 与其它方法相比, 本文提出的方法具有较好的性能。

UMN 数据库属于整体异常数据, 在图像中只存在正常事件或异常事件。UMN 数据库由 3 个场景组成: 草坪(1—1453 帧)、大厅(1454—5596 帧)和广场(5597—7739 帧), 图像尺寸为  $320 \times 240$ 。用正常事件的 400 帧图像作为训练数据, 剩下的作为测试数据。字典的尺寸为 300, 提取时空块的高斯和 Gabor 滤波器参数为  $\sigma=1.5$  和  $\tau=2.5$ 。在图6中, 浅色条状表示正常事件帧数, 深色条状表示异常事件帧数, 从定性结果可以看出, 提出的方法可以检测到每个异常事件, 检测的结果比标注的真值帧数更多, 且提前于真值, 这是因为给定真值的标注要晚于异常事件发生的帧数, 从而也反映了本文提出的方法对异常事件具有更快的反应速度。



图6 UMN 3 个场景异常事件检测定性结果

定量分析结果在表3中给出, 像素帧级的 AUC 与其它方法的比较: 同文献[17]具有可比性, 而 Chaotic Invariants 方法只处理整体异常情况, 本文方法处理整体异常和局部异常都获得较好的结果; 与文献[9]的方法比较, 草坪场景和广场场景异常事件检测效果更好, AUC 为 99.69% 和 97.37%, 大厅场景异常事件检测的 AUC 稍低于文献[9]的 97.5%, 为 97.11%。大厅场景中出现剧烈的光照变化, 是导致检测率低的主要原因。UMN 数据库 ROC 曲线在图7中给出。

表3 UMN 3 个场景异常事件检测定性结果

|                                    | AUC    |
|------------------------------------|--------|
| Chaotic Invariants <sup>[17]</sup> | 99%    |
| Social Force <sup>[18]</sup>       | 96%    |
| Optical flow <sup>[18]</sup>       | 84%    |
| 草坪场景 <sup>[9]</sup>                | 99.5%  |
| 大厅场景 <sup>[9]</sup>                | 97.5%  |
| 广场场景 <sup>[9]</sup>                | 96.4%  |
| 草坪场景(本文方法)                         | 99.69% |
| 大厅场景(本文方法)                         | 97.11% |
| 广场场景(本文方法)                         | 97.37% |

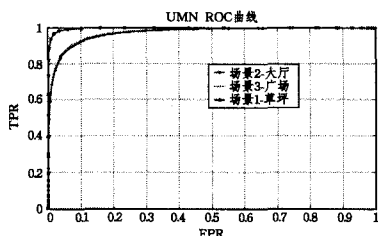


图7 UMN 3个场景异常事件检测 ROC 曲线

**结束语** 线性动态系统模型描述提取的视频时空块,稀疏编码进行异常事件检测。然而,线性动态系统属于非欧氏空间模型,无法直接应用传统的稀疏编码。输入视频时空小块的每帧图像可以由训练纹理的图像线性重构,重构系统与输入系统是等价的,它们之间可以通过非奇异线性变换矩阵进行转化。因此,算法的优化学习任务即是学习编码  $\alpha$  和相似性变换矩阵  $P$ 。由于线性变换空间是无限的,使得学习  $P$  无法实现。本文压缩线性变换空间,在正交子空间内寻找  $P$ ,将训练事件集合分为若干子集合,在每个子集中将用于重构次数最多的若干样本作为子集的字典,子集字典的集合构成算法的字典。实验结果表明,本文所提出的方法具有较好的检测结果。由于优化过程得到局部最优解,参数初值的选择至关重要。寻找到最佳的初值使识别率更高,或者寻找到获得全局最优解的优化方法,是本文以后的研究方向。

## 参考文献

- [1] Shi Ying-huan, Gao Yang, Wang Rui-li. Real-Time Abnormal Event Detection in Complicated Scenes[C]// International Conference on Pattern Recognition. 2010
- [2] Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features[C]// VS-PETS. October 2005
- [3] Soatto S, Doretto G, Wu Y N. Dynamic textures [J]. IJCV, 2003, 52(2): 91-109
- [4] Ravichandran A, Chaudhry R, Vidal R. View-invariant dynamic texture recognition using a bag of dynamical systems[C]// CVPR. 2009: 1651-1657
- [5] Ravichandran A, Chaudhry R, Vidal R. Categorizing Dynamic Textures using a Bag of Dynamical Systems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(2): 342-353
- [6] Coviello E, Mumtaz A, Chan A B, et al. Lanckriet. Growing a Bag of Systems Tree for Fast and Accurate Classification[C]// IEEE CVPR. 2012
- [7] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227
- [8] Ghanem B, Ahuja N. Sparse Coding of Linear Dynamical Systems with an Application to Dynamic Texture Recognition[C]// Proc. of the International Conference on Pattern Recognition (ICPR). 2010: 987-990
- [9] Cong Yang, Yuan Jun-song, Liu Ji. Sparse reconstruction cost for abnormal event detection[C]// CVPR. 2011: 3449-3456
- [10] Zhao Bin, Li Fei-fei, Xing E P. Online Detection of Unusual Events in Videos via Dynamic Sparse Coding[C]// CVPR. 2011: 3313-3320
- [11] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]// NIPS. 2006
- [12] Jimenez N D, Afsari B, Vidal R. Fast Jacobi-type algorithms for computing distances between linear dynamical systems[C]// IECC. 2013
- [13] Mahadevan V, Li W X, Vasconcelos N. UCSD dataset [OL]. [2012-01-11]. <http://www.svcl.ucsd.edu/projects/anomaly/>
- [14] UMN dataset [OL]. [2012-01-11]. <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>
- [15] Adam A, Rivlin E, Shimshoni I, et al. Robust real-time unusual event detection using multiple fixed-location monitors[J]. TPA-MI, 2008, 30(3): 555-560
- [16] Mahadevan V, Li W, Bhalodia V, et al. Anomaly detection in crowded scenes[C]// CVPR. 2010
- [17] Wu S, Moore B, Shah M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes[C]// CVPR. 2010
- [18] Mehran M S R, Oyama A. Abnormal crowd behavior detection using social force model[C]// CVPR. 2009
- [19] Jindal N, Liu B. Opinion spam and analysis[C]// Proceedings of the International Conference on Web Search and Web Data Mining. Palo Alto, California, USA: ACM, 2008: 219-230
- [20] Ott M, Choi Y, Cardie C, et al. Finding deceptive opinion spam by any stretch of imagination[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland, Oregon: 2011: 309-319
- [21] 李霄, 丁晨春. 垃圾商品评论信息的识别研究[J]. 现代图书情报技术, 2013, 229: 63-68
- [22] Lim E P, Nguyen V A, Jindal N, et al. Detecting product review spammers using rating behaviors [C]// Proceedings of the 19th ACM International Conference on Information and Knowledge Management. Toronto, ON, Canada: ACM, 2010: 930-948
- [23] Mukherjee A, Liu B, Wang J, et al. Spotting fake reviewer groups in consumer reviews[C]// Proceedings of 21st International Conference on World Wide Web. New York, NY, USA: ACM, 2012: 191-200
- [24] Wang G, Xie S, Liu B, et al. Review graph based online store review spammer detection[C]// Proceedings of the 11th International Conference on Data Mining. Vancouver, BC, Canada: IEEE, 2011: 1242-1247
- [25] Jindal N, Liu B, Lim E P. Finding unusual review patterns using unexpected rules [C]// Proceedings of the 19th International Conference on Information and Knowledge Management. Toronto, ON, Canada: ACM, 2010: 1549-1552
- [26] 《编程之美》小组. 编程之美[M]. 北京: 电子工业出版社, 2008: 141
- [27] 栾建安, 苏炳华. 多类别多评估人的 Kappa 分析[J]. 中国卫生统计, 1995, 12(6): 20-22
- [28] Landis J R, Koch G G. The measurement of observer agreement for categorical data[J]. biometrics, 1977, 33(1): 159-174

(上接第 299 页)