基于主题隐马尔科夫模型的人体异常行为识别

朱旭东 刘志镜
(西安电子科技大学计算机学院 西安 710071)

摘 要 针对基于监控视频的人体异常行为识别问题,提出了基于主题隐马尔科夫模型的人体异常行为识别方法,即通过无任何人工标注的视频训练集自动学习人体行为模型,并能够应用学到的人体行为模型实时检测异常行为和识别正常行为。这一方法主要围绕"低层视频表示-中层语义行为建模-高层语义分类"3个方面进行:1)基于时-空间兴趣点构建了一种紧凑的和有效的视频表示方法。2)提出一种新颖的语义主题模型(Topic Model,TM)——主题隐马尔科夫模型(Topic Hidden Markov Model,THMM),它能够自然地分组视频中检测到的人体行为。主题隐马尔科夫模型基于已有的马尔科夫模型和主题模型构造,不但聚类运动词汇成简单动作,而且聚类简单动作成全局行为,同时建模了行为时间上的相关性。THMM是一个4层贝叶斯主题模型,它将视频序列建模为行为的马尔科夫链,同时行为是视频序列中某些视频剪辑(Clip)的概率分布;将视频剪辑建模为动作的随机组合,同时动作是视频剪辑中运动词汇的概率分布。克服了传统隐马尔科夫模型和主题模型在人体复杂行为建模过程中精度、鲁棒性和计算效率上的不足。3)提出运行时累积的异常性测度及其在线异常行为检测方法和基于在线似然比检验(Likelihood Ratio Test,LRT)的实时正常行为分类方法,从而克服了实时行为识别过程中由于缺乏充分的视觉证据而引发的行为类型歧义,能完较好地完成监控场景中实时异常行为检测和在线正常行为识别的任务。取自实际监控场景的实验数据集上的实验结果证明了本方法的有效性。

关键词 计算机视觉,语义主题模型,异常检测,运动词包,行为聚类
中图法分类号 TP391 **文献标识码** A

Human Abnormal Behavior Recognition Based on Topic Hidden Markov Model

ZHU Xu-dong LIU Zhi-jing
(School of Computer Science and Technology, Xidian University, Xi'an 710071, China)

Abstract This paper aimed to address the problem of modeling human behavior patterns captured in surveillance videos for the application of online normal behavior recognition and anomaly detection. From the perspective of cognitive psychology, a novel method was developed for automatic behavior modeling and online anomaly detection without the need for manual labeling of the training data set. The work has been done with the hierarchical structure, following the routine of "Video Representation-Semantic Behavior (Topic) Model-Behavior Classification": 1) A compact and effective behavior representation method is developed based on spatial-temporal interest point detection. 2) The natural grouping of behavior patterns is determined through a novel clustering algorithm, topic hidden Markov model (THMM) built upon the existing hidden Markov model (HMM) and latent Dirichlet allocation (LDA), which overcomes the current limitations in accuracy, robustness, and computational efficiency. The new model is a four-level hierarchical Bayesian model, in which each video is modeled as a Markov chain of behavior patterns where each behavior pattern is a distribution over some segments of the video. Each of these segments in the video can be modeled as a mixture of actions where each action is a distribution over spatial-temporal words. 3) An online anomaly measure is introduced to detect abnormal behavior, whereas normal behavior is recognized by runtime accumulative visual evidence using likelihood ratio test (LRT) method. Experimental results demonstrate the effectiveness and robustness of our approach using noisy and sparse data sets collected from a real surveillance scenario.

Keywords Computer vision, Topic model, Anomaly detection, Bag of motion word, Behavior clustering

1 引言

人的行为分析在安全监控、高级人机交互、视频会议、基于行为的视频检索以及医疗诊断等方面有着广泛的应用前景和潜在的经济价值,是当前计算机视觉领域的一个研究热点[1,2]。视频智能监控系统(Intelligent Video Surveillance

到稿日期:2011-04-16 返修日期:2011-07-02 本文受国家自然科学基金(60573139)资助。
朱旭东(1973—),男,博士生,主要研究方向为视觉计算、数据挖掘,E-mail:zhudongxu@vip.sina.com;刘志镜(1956—),男,教授,主要研究方向为数据挖掘、视觉计算。

System,IVSS)是其最重要的应用之一。随着公共安全级别的不断提升和摄像机价格的持续下跌,IVSS已经广泛应用于火车站、飞机场、地铁站等公共场所。人的异常行为识别是部署IVSS的主要目的之一[3]。在动态场景中,IVSS基于视频序列侦测、跟踪和识别对象,甚至描述和理解人的行为。部署IVSS的最终目标是替代传统的被动监控系统,因为传统的被动监控系统通过人工观测和分析海量的监控视频[2],导致了高昂的人工成本、较低的识别率和较高的漏检率。总之,IVSS的目标不仅是通过摄像机代替人眼,而且尽可能自动化完成整个监控任务。人的异常行为的检测将有助于犯罪、事故、恐怖袭击等的侦测。但目前大部分的技术,尤其是实时系统还处在对运动目标的检测跟踪阶段,运动目标的行为分析仅是简单的走跑运动。本文主要讨论监控场景下人的异常行为识别问题,并将场景中少量发生的行为定义为异常行为,而将大量反复出现的一般行为定义为正常行为。

针对监控场景下人体异常行为识别问题,提出了基于语义主题模型的人体异常行为识别方法。该方法建立低层人体运动特征描述与中高层人体行为认知之间的关系,确定人体行为的中间语义描述,实现人体异常行为的语义判别。整个研究路线围绕人体运动特征提取和运动词包表示、人体行为语义建模(人体行为聚类)以及实时语义分类(异常行为检测和正常行为识别)这3个主要部分展开,它们之间的组成结构和功能如图1所示。

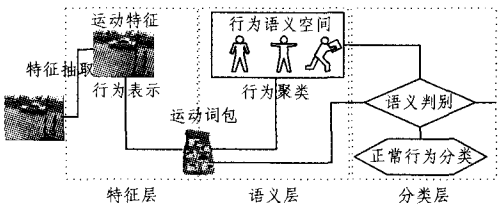


图1 人体异常行为识别系统流程图

1)基于运动词袋的视频表示。作为对低层运动特征分类性能的提高,语义行为类型可解决视频低级运动特征与高层语义类别之间的鸿沟问题。本文采用Dollar等人[4]提出的时-空间兴趣点作为低层运动特征。同类人体行为具有相似的概念分布,建立语义概念的运动分布就可以将人体行为划分为特定的语义类别。这些相似概念通常是与运动特征相关联的,如何在语义层表示中保留特征的不变性,体现运动信息,实现局部特征与全局特征的综合互补,是跨越低层特征与高层人体行为语义间鸿沟的关键,也是实现人体行为语义建模的关键。

2)人体行为语义建模。同类人体行为中,虽存在着相似概念,但因视频本身内容的变化而存在着多样性。应捕获这些共同出现的相似概念,并抽象为主题,以为视频描述提供进一步的中间语义描述。由于视频人体行为语义内容以不可知的方式进行组合,因此采用概率生成模型进行推导更为合理。本文针对具有较长时间跨度和包含多种基本动作的人体复杂行为建模问题,通过协同聚类算法同步建模运动词汇间和人体动作间的相关性,同时采用隐马尔科夫模型建模行为时间上的相关性,以提高人体复杂行为的识别精度。

3)在线人体异常行为检测和正常人体行为识别。当前常用于人体异常行为识别的语义主题模型是PLSA和LDA模型,它们只能通过离线和批量的推理完成人体异常行为识别任务,难以满足智能视频监控对人体异常行为识别在线和实时数据处理的要求。本文基于运行时累积思想提出一种异常性测度及其在线异常行为检测方法,并且基于在线似然比检验(Likelihood Ratio Test,LRT)方法提出一种实时正常行为识别方法。

2 相关工作

已有的异常行为检测方法大致可以分为两类。

基于异常行为建模的方法[5,6]。这一方法大致分为两步:首先,从视频序列中提取图像特征,特征通常通过侦测和跟踪运动对象[7],并计算其轨迹、速度以及形状描述符而获得[8]。然后,基于提取的特征通过人工或应用监督学习技术构建"正常"行为模型[9]。行为建模通常选择隐马尔科夫模型(HMM)[10,11]或其他图模型[12],这些模型将图像特征量化为一系列离散的状态并建模状态随时间的变化方式。为了检测异常行为,将视频与一系列正常模型相匹配,其中不适合模型的段落认为是异常。基于模型的方法在"正常"行为可以明确定义和约束的场景中相当有效。然而,在典型的现实生活视频中,如在我们实验中使用的视频,各种各样的"正常"行为很容易淹没"异常"行为。因此,在无约束的环境下界定和建模"正常"行为比确定"异常"行为更加困难。如果目标是在长视频中检测异常行为,基于模型的方法往往过于严格。

基于"异常"与"正常"行为之间相异度的方法[13-15]。这一方法又根据是否构建行为模型进一步细分为两种不同的子方法:不需要构建行为模型的方法,首先聚类观察到的行为模式,然后将其中小的聚类标注为异常[13,14];或首先构建"正常"行为集的时-空间碎片数据库,然后将不能通过数据库中数据表达的行为标注为异常行为[15]。Hua等人[13]提出的方法不能应用于样本集未出现过的行为,因此只适合于事后分析而不适合在线异常侦测。Boiman等[14]和Hamid等[15]分别针对这一问题提出了两种不同的解决方法。然而,在这些方法中,为了检测未出现过的异常行为,必须将所有先前观察到的正常行为以离散事件直方图[14]或时-空间碎片[15]的形式存储在数据库中。因此,这一方法的扩展性受到较大制约,很难适用于各种各样的应用场景。

基于行为模型的方法。首先基于语义主题模型聚类场景中的正常行为,然后通过已获得的聚类进行进一步的有监督分类实现聚类精化。Fleischman等[16]首先发现视频流中的频繁主题,然后构建主题的SVM分类器,标识家庭行为(例如"煮咖啡"、"洗碗")。Xing等[17]使用无向图模型发现新闻视频中交差模式的隐含结构,然后使用发现结果改进概念侦测。

本文提出一种新颖的语义主题模型——主题隐马尔科夫模型(Topic Hidden Markov Model,THMM),该模型克服了已有的马尔科夫模型和贝叶斯主题模型在精度、鲁棒性和计算效率上的不足。其主要创新点包括:(1)层次建模:不但聚类运动词汇成简单动作,而且聚类简单动作成全局性复杂行为;(2)时序建模:建模行为时序上的相关性。

3 视频表示

总体而言,存在3种常用的描述姿态和运动的特征:基于边缘和肢体形状的静态特征、基于光流的动态特征和基于局

部视频块的时-空间特征。特别地，局部时-空间特征为行为分类提供了丰富的说明和强大的表示手段。

局部时-空间特征的检测方法[4]类似于局部空间特征的检测方法，只是将图像 $I(x,y)$ 换成了图像序列 $I(x,y,t)$。为了从图像序列 $I(x,y,t)$ 中检测局部特征，针对单一固定摄像机视频应用可分线性过滤器构造下述响应函数 R：

$$R=(I\times g\times h_{ev})^2+(I\times g\times h_{od})^2 \tag{1}$$

其中，

$$g(x,y;\sigma)=\frac{e^{(-(x^2+y^2)/2\sigma_l^2-t^2/2\tau_l^2)}}{\sqrt{(2\pi)^3\sigma_l^4\tau_l^2}} \tag{2}$$

是二维 (x,y) 高斯平滑核；

$$h_{ev}(t;\tau,\omega)=-\cos(2\pi t\omega)e^{t^2/\tau^2} \tag{3}$$

和

$$h_{od}(t;\tau,\omega)=-\sin(2\pi t\omega)e^{t^2/\tau^2} \tag{4}$$

是一对正交的时间维 Gabor 滤波器；参数 τ 和 σ 分别对应于空间侦测器和时间侦测器的尺度；在各种情况下，ω 都取 $4/\tau$。

响应函数 R 的每个极大值点的邻接区域中包含了该图像序列中局部事件的运动和空间外观的信息。为了捕捉这些信息，从该区域抽取称为时-空间兴趣点的像素立方体，以构造局部特征的表示。兴趣点大小通过空间和时间尺度参数 (σ,τ) 确定。选取的兴趣点应包含对响应函数 R 取极大值有贡献的大多数点，因为这些点反映了底层图像结构和局部模式的变化率。

时-空间特征点的描述符 l 通过亮度梯度 G 构造，构造过程如下：首先，计算兴趣点中每一像素 (x,y,t) 亮度沿 x,y 和 t 方向的梯度 (G_x,G_y,G_t)；然后，将这些梯度值串联构成一个矢量，其中矢量大小＝立方体像素数×平滑尺度数×梯度方向数；最后，利用主成分分析(PCA)降维技术，将矢量投射到低维空间(不妨设维数为 n)，从而获得描述符

$$l=(\vec{G}_1,\vec{G}_2,\cdots,\vec{G}_n) \tag{5}$$

由实践观察知，同一行为的两个实例在整体外观和运动上存在较大差异，但其产生的时-空间兴趣点是相似的。因此，即使存在无限数量的时-空间兴趣点，但其类型应该是相对较少的。通过 K-means 聚类训练数据集检测到的所有兴趣点 l，构建一个时-空间词汇表 (W)，其中每个聚类中心定义为一个时-空间词 (w_i)。

4 人体行为聚类

4.1 主题隐马尔科夫模型

给定具有 M 条视频序列的样本集 D，记为

$$D=\{v_1,v_2,\cdots,v_M\} \tag{6}$$

若将 D 中每一样本 v 都均匀分割为 T 段时长一秒钟的视频剪辑 c_t，则 v 包含的行为模式 $\boldsymbol{W}$ 是视频剪辑 c_t 包含的子行为模式 w_t 的序列，记为

$$\boldsymbol{W}=[w_1,w_2,\cdots,w_t,\cdots,w_T] \tag{7}$$

式中，w_t 是 c_t 中出现的时-空间词 $w_{t,n}$ 的集合，记为

$$w_t=[w_{t,1},w_{t,2},\cdots,w_{t,n},\cdots,w_{t,N}] \tag{8}$$

同时，$w_{t,n}$ 是时-空间词汇表的 V 个基本词汇之一。

如图 2 所示，建模 $\boldsymbol{W}$ 成为具有 3 层隐含结构(即运动词汇、动作和行为)的层次图模型。模型基本思想：假设视频 v 是 T 段视频剪辑的序列，其中每段视频剪辑 c_t 展现出特定行为类型 y_t。同时建模序列 $\{y_t\}_{t=1}^{T}$ 为马尔科夫链，即行为类型 y_t 依据多项式分布(Multinomial)沿时间方向逐段演变，即

$$P(y_t\mid y_{t-1},\psi)=\text{Multinomial}(\psi_{y_{t-1}}) \tag{9}$$

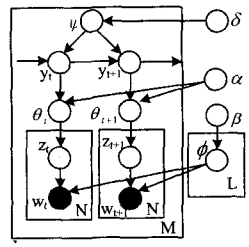


图 2 主题隐马尔科夫模型

而每一段视频剪辑 c_t 都是由各种动作随机组合而成的，则各种动作混合比例 θ_t 基于行为类型 y_t 确定，即

$$P(\theta_t\mid y_t,\alpha)=\text{Dirichlet}(\alpha_{y_t}) \tag{10}$$

同时在视频剪辑 c_t 中，N_t 个简单动作类型 $\{z_{t,n}\}_{n=1}^{N_t}$ 基于 θ_t 确定，即

$$P(z_{t,n}\mid\theta)=\text{Multinomial}(\theta_{z_{t,n}}) \tag{11}$$

最终，基于相关动作类型 $z_{t,n}$ 确定每个观察到的时-空间词 $w_{t,n}$，即

$$P(w_{t,n}\mid z_{t,n},\phi)=\text{Multinomial}(\phi_{z_{t,n}}) \tag{12}$$

形式化讲，视频样本集 D 中，任一行为模式 $\boldsymbol{W}$ 的生成过程如下：

1)确定行为模式 $\boldsymbol{W}$ 所含视频剪辑的数目 T：$T\sim\text{Poisson}(\mu)$；

2)确定各种行为类型的混合比例 ψ：$\psi\sim\text{Dirichlet}(\delta)$；

3)确定各种时-空间词汇的混合比例 ϕ：$\phi\sim\text{Dirichlet}(\beta)$；

4)T 段视频剪辑中任一段 c_t 生成过程如下：

(1)确定 c_t 对应的行为类型 y_t：$y_t\sim\text{Multinomial}(\psi_{y_{t-1}})$；

(2)确定 c_t 所含时-空间词汇的数目 N_t：$N_t\sim\text{Poisson}(\varepsilon)$；

(3)确定各种动作类型的混合比例 θ_t：$\theta_t\sim\text{Dirichlet}(\alpha_{y_t})$；

(4)N_t 中任一时-空间词汇 $w_{t,n}$ 生成过程如下：

i. 确定 $w_{t,n}$ 对应的动作类型 $z_{t,n}$：$z_{t,n}\sim\text{Multinomial}(\theta_{z_{t,n}})$；

ii. 确定时-空间词汇 $w_{t,n}$：$w_{t,n}\sim\text{Multinomial}(\phi_{z_{t,n}})$。

设时-空间词汇表的基本词汇数目 V 和行为分类表的基本类型数目 K 是预先知道且固定不变的，则变量 y 和 z 的维度就是已知和确定的。条件概率 $P(w_{t,n}\mid z_{t,n},\phi)$ 通过 $L\times V$ 维矩阵 ϕ 计算，其中元素 $\phi_{i,j}$ 表示已知动作类型 z^i 条件下时-空间词汇 w^j 出现的概率，即

$$\phi_{i,j}=P(w^j=1\mid z^i=1) \tag{13}$$

假设 $\phi_{i,j}$ 固定不变且可通过学习过程评估。φ 表示行为模式 $\boldsymbol{W}$ 中各种行为类型的混合比例，并且确定了一个 K 维 Multinomial 分布的参数，行为类型 y 是这一 Multinomial 分布的采样。θ_t 是 Dirichelt 分布的采样，并且 θ_t 确定了视频剪辑 c_t 中各种动作类型的混合比例，同时 θ_t 依赖于视频剪辑 c_t 的行为类型 y_t。每一种行为类型 y_t 都是若干动作类型的组合，这一事实是通过 θ_t 矩阵的“超参数” α 建模。

给定“超参数” α,β,δ，变量 $\{y_t,z_t,w_t\}_{t=1}^{T}$ 和参数 ψ,θ,φ 的联合概率分布为

$$\begin{aligned}&P(\{y_t,z_t,w_t\}_{t=1}^{T},\psi,\theta,\phi\mid\alpha,\beta,\delta)\\&=P(\psi\mid\delta)P(\phi\mid\beta)\prod_{t=1}^{T}p(y_t\mid y_{t-1},\psi)\,p(\theta_t\mid\alpha,y_t)\prod_{n=1}^{N_t}p(z_{tn}\mid\\&\quad\theta_t)\,p(w_{tn}\mid z_{tn},\phi)\end{aligned} \tag{14}$$

4.2 人体行为模型的无监督学习

如 LDA 模型、THMM 模型的精确推理计算复杂度过高，但基于马尔科夫-蒙特卡洛(Markov chain Monte Carlo, MCMC)近似学习和推理方法的 Gibbs 抽样是可行的。由于 THMM 模型 Dirichelt 分布与 Multinomial 分布的共轭关系，使得参数$\{\psi,\theta,\phi\}$在 Gibbs 抽样过程中自动消去。给定其他变量，积分条件分布 $P(y_t|\boldsymbol{y}_{-t},\boldsymbol{z},\boldsymbol{w})$中消去 ψ 和 θ，并且考虑沿行为马尔科夫链、行为 y_{t-1}和 y_{t+1}的各种可能转换概率，导出行为类型 y_t 的 Gibbs 抽样更新，即

$$p(y_t|\boldsymbol{y}_{-t},\boldsymbol{z},\boldsymbol{w})\propto \frac{\prod_z \Gamma(\alpha+n_{y_t}^{(z)})\Gamma(L\alpha+\sum_z n_{\neg y_t}^{(z)})}{\prod_z \Gamma(\alpha+n_{\neg y_t}^{(z)})\Gamma(L\alpha+\sum_z n_{y_t}^{(z)})} \frac{(n_{y_t}^{(y_{t+1})}+\delta)(n_{y_t}^{(y_{t+1})}+I(y_{t-1}=y_t)I(y_t=y_{t+1})+\delta)}{\sum_{y_{t+1}} n_{y_t}^{(y_{t+1})}+I(y_{t-1}=y_t)+K\delta} \tag{15}$$

式中，$\boldsymbol{y}_{-t}$表示除 y_t 以外的所有 $\boldsymbol{y}$ 变量；$n_{y_t}^{(z)}$表示指定为行为类型 y_t 的动作类型 z 数目；$n_{\neg y_t}^{(z)}$表示没有指定为行为类型 y_t 的动作类型 z 数目；$\sum_z n_{y_t}^{(z)}$表示指定行为类型 y_t 的动作总数；$\sum_z n_{\neg y_t}^{(z)}$表示没有指定行为类型 y_t 的动作总数；$n_{y_t}^{(y_{t+1})}$表示行为类型 y_{t+1}跟随 y_t 出现的次数；$\sum_{y_{t+1}} n_{y_t}^{(y_{t+1})}$表示跟随行为类型 y_t 的行为类型总数。L 表示动作类型 z 的数目，K 表示行为类型 y 的数目。

给定其他变量，积分条件分布 $P(z_{t,n}|\boldsymbol{z}_{-t,n},\boldsymbol{y},\boldsymbol{w})$，消去 θ 和 ϕ，导出行为类型 $z_{t,n}$的 Gibbs 抽样更新，即

$$p(z_{t,n}|\boldsymbol{z}_{-t,n},\boldsymbol{y},\boldsymbol{w})\propto \frac{n_{z_{t,n}}^{(w)}+\beta}{\sum_w n_{z_{t,n}}^{(w)}+V\beta}\frac{n_y^{(z)}+\alpha}{\sum_z n_y^{(z)}+L\alpha} \tag{16}$$

式中，$\boldsymbol{z}_{-t,n}$表示除 $z_{t,n}$以外的所有 $\boldsymbol{z}$ 变量；$n_{z_t}^{(w)}$表示指定为动作类型 z_t 的时-空间词汇 w 数目；$\sum_w n_{z_t}^{(w)}$表示指定动作类型 z_t 的时-空间词汇总数；$n_y^{(z)}$ 表示指定为行为类型 y 的动作类型 z 数目；$\sum_z n_y^{(z)}$ 表示指定行为类型 y 的动作总数。L 表示动作类型 z 的数目，V 表示时-空间词汇表基本词汇 w 数目。

通过后验分布 $P(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{w})$的抽样，能够评估隐含参数 ψ,θ 和 ϕ，即

$$\hat{\phi}=\frac{n_z^{(w)}+\beta}{\sum_w n_z^{(w)}+V\beta} \tag{17}$$

$$\hat{\theta}=\frac{n_y^{(z)}+\alpha}{\sum_z n_y^{(z)}+L\alpha} \tag{18}$$

$$\hat{\psi}=\frac{(n_{y_t}^{(y_{t+1})}+\delta)(n_{y_t}^{(y_{t+1})}+I(y_{t-1}=y_t)I(y_t=y_{t+1})+\delta)}{\sum_{y_{t+1}} n_{y_t}^{(y_{t+1})}+I(y_{t-1}=y_t)+K\delta} \tag{19}$$

5 在线异常行为检测和正常行为识别

基于 Gibbs 采样算法的模型近似推理方法是一种离线的和批量处理的方法。为了实现监控视频的在线异常行为检测和正常行为识别，本节在离线的和批量的模型学习过程之后，构造了一种新颖的和实时性的 THMM 模型推理方法。

给定含 T_{tr}段视频的训练集，并且假设这个训练集是有代表性的，即

$$P(\alpha,\beta,\delta|\boldsymbol{w}_{t'>T_{tr}})=P(\alpha,\beta,\delta|\boldsymbol{w}_{1,T_{tr}}) \tag{20}$$

设一个新的、未分类的行为模式 $\boldsymbol{W}$ 为：

$\boldsymbol{W}=[\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_T]$

本节采用截至当前(第 t)段剪辑累积信息 $\boldsymbol{w}_{1,t}$在线检测异常。首先，定义 $\boldsymbol{w}_{1,t}$的模型相似度为

$$l_t\triangleq\frac{1}{t}\log P(\boldsymbol{w}_{1,t}|\alpha,\beta,\delta) \tag{21}$$

然后，基于 l_t 定义异常测量函数 Q_t 为

$$Q_t=\begin{cases} l_1, & t=1 \\ (1-\alpha)Q_{t-1}+\alpha(l_t-l_{t-1}), & t\neq 1 \end{cases} \tag{22}$$

式中，α 是确定从当前段剪辑抽取的运动信息对于异常检测重要性的系数，取值范围为

$$0<\alpha\leqslant 1 \tag{23}$$

与直接采用 l_t 判断异常/正常相比较，Q_t 增加了更多的权重到新的观测量。最后，第 t 段剪辑是否异常的判别式为

$$Q_t<Th_A \tag{24}$$

式中，Th_A 是异常检测的门限，其值依据不同监控应用要求的检测率和误警率而确定。

若当前段(第 t)视频剪辑的 $Q_t>Th_A$，则 $\boldsymbol{w}_{1,t}$为 K 类正常行为之一。本节采用在线似然比检验(Likelihood Ratio Test, LRT)方法分类 $\boldsymbol{w}_{1,t}$：

首先，考虑下述假设检验：

1) H_k：$\boldsymbol{w}_{1,t}$由模型 y_k 产生，对应于第 k 类正常行为；

2) H_0：$\boldsymbol{w}_{1,t}$不是由模型 y_k 产生，不属于第 k 类正常行为。H_0 称为替代假设。

然后，定义假设 H_k 和 H_0 的似然比为

$$r_k=\frac{P(\boldsymbol{w}_{1,t};H_0)}{P(\boldsymbol{w}_{1,t};H_k)} \tag{25}$$

式中，H_k 通过行为聚类 y_k 表示，因此计算 r_k 的关键是如何构造替代模型 y_0 来表示 H_0。一般情况下，由于可能的 y_0 是无穷多的，因此 $P(\boldsymbol{w}_{1,t};H_0)$只能通过近似方法计算。但在本节中，已经确定 $\boldsymbol{w}_{1,t}$属于正常行为，必然由 K 类正常行为之一产生。因此，设 y_0 是除 y_k 外 $K-1$ 类正常行为的组合模型是合理的。

H_k 和 H_0 的似然比按下式计算为

$$r_k=\frac{\sum_{i\neq k}P(y_i)P(\boldsymbol{w}_{1,t}|y_i)}{P(\boldsymbol{w}_{1,t}|y_k)}=\frac{\sum_{i\neq k}\frac{N_i}{N-N_k}P(\boldsymbol{w}_{1,t}|y_i)}{P(\boldsymbol{w}_{1,t}|y_k)} \tag{26}$$

式中，r_k 是时间 t 的函数，N 是训练集中行为模式的总数，N_k 是属于第 k 类行为类型的行为模式的数目。

最后，通过 r_k 的 χ^2 检验，分类 $\boldsymbol{w}_{1,t}$，即当 $\chi^2=-2\ln(r_k)>Th_r$ 时，$\boldsymbol{w}_{1,t}$识别为第 k 类行为类型。其中 Th_r 通过置信水平 α(一般取 0.05)确定，即

$$Th_r=\chi^2_{1-\alpha} \tag{27}$$

若存在多个 $r_k<Th_r$，则 $\boldsymbol{w}_{1,t}$识别为 r_k 值最小的第 k 类行为类型。

6 实验与结果分析

本节首先采用基于时-空间特征点的运动词袋表示，然后由 THMM 模型生成语义行为表示，最后采用上节描述的方法在线检测异常行为和实时分类正常行为。

6.1 实验设置

本节采用学校实验大楼门口的监控视频进行测试。该场景中摄像机位于实验大楼门口上方，用于监控进出实验大楼行人的行为。场景中常见典型行为包括进入、离开和路过，其中每一行为实例一般持续数秒，如图 3(a)所示。本节从场景

监控视频中选取 947 条视频数据进行实验。

使用第 3 节介绍的方法提取时-空间特征点，检测参数 $\sigma=2,\tau=2.5$，如图 3(b)所示。每个时-空间特征点由时间和空间梯度的级联矢量描述。然后，把描述符映射到低维空间。为了建立运动词典，对所训练视频集所有特征描述符进行聚类，构建一个含 1800 个基本运动词汇的词典。

(a) 监控测试场景的典型帧　(b) 由时-空间特征点表示的动作类型　(c) 场景中发现的行为类型及其动态性

图 3

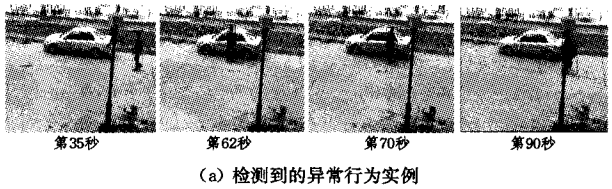
6.2 行为聚类

6.2.1 动作类型数目和行为类型数目估计

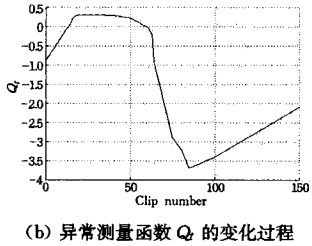
为了估计场景中动作类型数目 L 和行为类型数目 K，当 (L,K)取(9,3)，(13,5)，(18,7)和(22,8)并且 $\alpha=8,\beta=0.05$，$\delta=1$ 时，采用 Gibbs 抽样算法对隐含变量 y 和 z 的后验分布 $P(\{y_t,z_t\}_{t=1}^{T_{tr}}|\{w_t\}_{t=1}^{T_{tr}},\alpha,\beta,\delta)$采样，并计算概率 $P(\boldsymbol{w}|L,K)$。对每一对(L,K)的 Gibbs 采样总迭代 1500 次迭代，并抛弃前 1000 次，然后每隔 100 次取 5 个采样作为 $P(\boldsymbol{z},\boldsymbol{y}|\boldsymbol{w},\alpha,\beta,\delta)$独立采样。$P(\boldsymbol{w}|L,K)$作为$(L,K)$的函数，其值开始时是递增的，在(13,5)处达到顶点，随后递减。这一结果表明场景含 13 种常见动作类型和 5 种常见行为类型。

6.2.2 聚类运动词汇为动作类型

THMM 模型学习到的主题 z 对应于人体动作类型，其是运动词汇 w 的多项分布，即 $w_{t,n}\sim\text{Multinomial}(\phi_{z_{t,n}})$。从训练视频数据集中发现动作类型，其发现过程如下：首先，给定分配到动作类型 z 的运动词汇 w，通过式(12)很容易估计出“超参数”$\hat{\phi}_z$。然后，计算概率分布 $p(w|z,\hat{\phi}_z)$，取其中概率值较大的 w。最后，聚类 w 发现相应的动作类型 z。发现的动作类型 z 都具有清晰的语义，如图 4(a)所示。



(a) 检测到的异常行为实例



(b) 异常测量函数 Q_t 的变化过程

图 4　监控场景中的异常检测实例

6.2.3 发现行为类型及其动态性

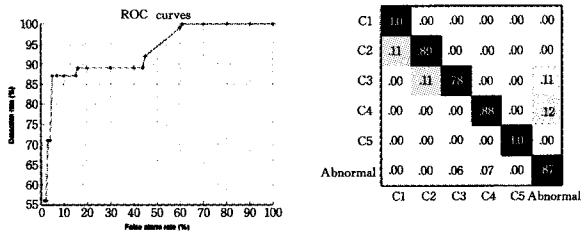
通过参数矩阵 θ_y 自动聚类共现动作类型(主题)成为行为类型 y。图 3(c)展示了场景中发现的 5 类行为类型：(B1)进入大楼、(B2)路过大楼、(B3)离开大楼、(B4)下车和(B5)上车。同时，通过学习获得的转换分布$\hat{\psi}$能够发现行为类型的持续时间和相互顺序，例如 B4 和 B1、B2 和 B5 的相继关系。

6.3 在线异常行为检测和实时正常行为识别

如上节所述，已经通过离线学习获得了监控场景的人体行为模型。下面基于这些模型，在线检测实时获得的监控视频流。对于未知的行为模式 $\boldsymbol{W}$，若其当前检测帧的异常测量函数 Q_t 满足式(24)，则实时判定其为异常行为，其中累积因子 α 设为 0.1。图 4 展示了监控场景中的一个异常检测实例，其中(a)显示了一行人刮划停放在该区域车辆的异常行为；(b)显示了 Q_t 的变化过程，其中直到第 62 段剪辑才判定行为异常。

为了测量算法识别率，手动标注测试集中的行为模式 $\boldsymbol{W}$ 成为不同的行为类型。若 $\boldsymbol{W}$ 通过学习获得的人体行为模型判定为正常且分类为某种人体行为，则认为其是正确识别。图 5(b)表明若一个正常行为未被正常识别，但其通常也被识别为另一个正常行为。

图 5(a)给出了检测率和误警率随 Th_A 变化的 ROC 曲线。ROC 曲线指受试者工作特征曲线(receiver operating characteristic curve)，是反映检测率和误警率连续变量的综合指标，是用构图法揭示检测率和误警率的相互关系，它通过将连续变量设定出多个不同的临界值(Th_A)，从而计算出一系列检测率和误警率，再以检测率为纵坐标、误警率为横坐标绘制成曲线，曲线下面积越大，诊断准确性越高。在 ROC 曲线上，最靠近坐标图左上方的点为检测率和误警率均较高的临界值。图 5(b)是通过异常检测算法生成的混淆矩阵(Confusion Matrices)。



(a)随 Th_A 变化的 ROC 曲线　(b)混淆矩阵($Th_A=0.1$)

图 5

6.4 实验结果分析

现将本节提出的方法与现有主流异常行为识别方法进行比较，比较结果如表 1 所列。算法性能比较的主要发现如下：

基于 THMM 模型的异常行为检测率优于其它方法，尤其对于复杂的异常行为。这里将异常行为定义为动作或行为在时间上非典型的共现，即异常是通过动作间的时-空间相关性定义的而非简单的个体动作。LSA[20]，PLSA 和 LDA[19]能推理动作，但不能同步推理其时序关系。HMM[18]和 n-gram[15]能够推理其时序关系，但由于其缺乏中间表示而陷入过匹配。THMM 结合了 LDA 和 HMM 模型的优点，从人体行为语义与语法的角度来捕获运动词语的共现，以实现运动词袋在中间语义行为上的映射。

表 1　算法性能比较

方法	检测率(%)
THMM	90.26
HMM (Xiang et al.[18])	82.76
n-grams (Hamid et al.[15])	82.38
LDA (Niebles et al.[19])	81.50
MAP-based (Boiman et al.[20])	79.32

(下转第 275 页)

[9] Atallah M J, Cole R, Goodrich M T. Cascading divide-and-conquer: A technique for designing parallel algorithms[J]. SIAM J. Comput, 1989

[10] Sun X H, Rover D T. Parallelism in computation problems[J]. SIAM J. Comput. IEEE Trans. on Parallel and Distributed Systems, 1994

[11] Horowitz E, Zorat A. Divide and conquer for parallel processing [J]. IEEE Trans. On Comput, 1983

[12] Shiloach Y, Vishkin U. Finding the maximum, merging and sorting in parallel computation model[J]. J. of Algorithms, 1981

[13] Valiant L G. Parallelism in comparison problems[J]. SIAM J. Comput, 1991

[14] Malyshkin V. Parallel Computing Technologies [C]// Proc of PACT'05. Krasnoyarsk, Russia: [s. n.], 2005

[15] Hill M, Marty M. Amdahl's Law in the Multicore Era[J]. IEEE Computer, 2008, 41(7): 33-38

(上接第 255 页)

文献[19]聚类行为模式成为中间语义行为，然后标注低内聚性聚类为异常行为，但这一方法难以实现在线异常检测。文献[15]可实现在线检测，但其仍需要观察到整个行为模式后才能判决。本文提出的在线异常检测和正常识别方法能够实现实时检测，其通过延时决策解决由于缺乏充分视觉证据而引发的行为类型歧义问题。

结束语 提出了一种具有层次结构的语义主题模型——主题隐马尔科夫模型（Topic Hidden Markov Model, THMM），它从中间语义描述的层次结构角度捕获运动词语的共现信息、动作的共现信息以及行为之间的时序信息；解决了 PLSA 和 LDA 等语义主题模型没有建模“运动词袋”之间的关联关系，从而难以确定“词袋”采集时间窗口大小的问题；其不但聚类运动词汇成简单动作，而且聚类简单动作成全局行为，同时建模了行为时间上的相关性。本模型构建两个具备层次结构的语义主题空间，使语义主题表示具备更强的判别力。在取自实际监控场景的实验数据集上的性能比较说明，本模型无论是在运动词包描述还是中间语义建模方面均能取得最好性能。

参 考 文 献

[1] Wang L, Hu W M, Tan T N. Recent developments in human motion analysis[J]. Pattern Recognition, 2003, 36(3): 585-601

[2] Johnson N, Hogg D. Learning the distribution of object trajectories for event recognition[J]. Image and Vision Computing, 1995, 14(8): 609-615

[3] Brand M, Oliver N, Pentland A. Coupled hidden markov models for complex action recognition [C]// Proceedings of IEEE International Conference on Computer Vision. San Juan, Puerto Rico: IEEE, 1997: 994-999

[4] Dollar P, Rabaud V, Cottrell G. Behavior recognition via sparse spatio-temporal features [C]// Proc. 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. China, 2005: 65-72

[5] Hongeng S, Nevatia R. Multi-agent event recognition[C]// Proceedings of Eighth International Conference on Computer Vision. Vancouver, BC, Canada: IEEE, 2001: 84-91

[6] Russo R, Shah M, Lobo N. A computer vision system for monitoring production of fast food [C]// Proceedings of The 5th Asian Conference on Computer Vision. Vancouver, Melbourne, Australia, 2002

[7] Wren C, Azarbayejani A, Darrell T. Pfinder: Real-time tracking of the human body[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780-785

[8] Haritaoglu I, Harwood D, Davis L S. W4: Who when where what a real time system for detecting and tracking people[C]// Proceedings of International Conference on Face and Gesture Recognition. Nara, Japan: IEEE, 1998

[9] Johnson N, Hogg D. Learning the distribution of object trajectories for event recognition[J]. Image and Vision Computing, 1995, 14(8): 609-615

[10] Brand M, Oliver N, Pentland A. Coupled hidden markov models for complex action recognition[C]// Proceedings of IEEE International Conference on Computer Vision. San Juan, Puerto Rico: IEEE, 1997: 994-999

[11] Medioni G, Cohen I, Bremond F. Event detection and analysis from video streams[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(8): 873-889

[12] Naphade M R, Huang T S. A probabilistic framework for semantic indexing and retrieval in video [C] // Proceedings of IEEE International Conference on Multimedia and Expo. New York, NY, USA: IEEE, 2000: 475-478

[13] Zhong H, Shi J B, Visontai M. Detecting Unusual Activity in Video [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, D. C., USA: IEEE, 2004: 819-826

[14] Hamid R, Johnson A, Batta S, et al. Detecting Unusual Activity in Video [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005: 1031-1038

[15] Boiman O, Irani M. Detecting Irregularities in Images and in Video [C] // Proceedings of Tenth IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005: 462-469

[16] Fleischman M, Decamp P, Roy D. Mining temporal patterns of movement for video content classification [C]// Proc. 8th ACM Int. Workshop Multimedia Information Retrieval. 2006: 183-192

[17] Xing E P, Yan R, Hauptmann A G. Mining associated text and images using dual-wing harmoniums [C] // Proc. Uncertainty Artifical Intelligence. 2005

[18] Xiang T, Gong S. Beyond tracking: modelling activity and understanding behaviour[J]. International Journal of Computer Vision, 2006, 67(1): 21-51

[19] Niebles J C, Wang H C, Li F F. Unsupervised learning of human action categories using spatial-temporal words[J]. International Journal of Computer Vision, 2008, 79(3): 299-318

[20] Zhong H, Shi J, Visontai M. Detecting unusual activity in video [C]// Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004: 819-826