

## 基于半监督学习的行为建模与异常检测\*

李和平, 胡占义<sup>+</sup>, 吴毅红, 吴福朝

(中国科学院 自动化研究所 模式识别国家重点实验室, 北京 100080)

### Behavior Modeling and Abnormality Detection Based on Semi-Supervised Learning Method

LI He-Ping, HU Zhan-Yi<sup>+</sup>, WU Yi-Hong, WU Fu-Chao

(National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-10-62616540, E-mail: huzy@nlpr.ia.ac.cn

**Li HP, Hu ZY, Wu YH, Wu FC. Behavior modeling and abnormality detection based on semi-supervised learning method. *Journal of Software*, 2007,18(3):527-537. <http://www.jos.org.cn/1000-9825/18/527.htm>**

**Abstract:** A simple and efficient method based on semi-supervised learning technique is proposed for behavior modeling and abnormality detection. The method is composed of the following steps: (1) Dynamic time warping (DTW) based spectral clustering method is used to obtain a small set of samples to initialize the hidden Markov models (HMMs) of normal behaviors; (2) The HMMs' parameters are further trained by the method of iterative learning from a large data set; (3) Maximum a posteriori (MAP) adaptation technique is used to estimate the HMMs' parameters of abnormal behaviors from those of normal behaviors; (4) The topological structure of HMM is finally constructed to detect abnormal behaviors. The main characteristic of the proposed method is that it can automatically select the number of normal behavior patterns and samples from the training dataset to build normal behavior models and can effectively avoid the running risk of over-fitting when the HMMs of abnormal behaviors are learned from sparse data. Experimental results demonstrate the effectiveness of the proposed method in comparison with other related works in the literature.

**Key words:** behavior modeling; abnormality detection; semi-supervised learning; HMM (hidden Markov models); computer vision

**摘要:** 提出了一种基于半监督学习的行为建模与异常检测方法. 该算法包括以下几个主要步骤: (1) 通过基于动态时间归整(DTW)的谱聚类方法获取适量的正常行为样本, 对正常行为的隐马尔可夫模型(HMM)进行初始化; (2) 通过迭代学习的方法在大样本下进一步训练这些隐马尔可夫模型参数; (3) 以监督的方式, 利用最大后验(MAP)自适应方法估计异常行为的隐马尔可夫模型参数; (4) 建立行为的隐马尔可夫拓扑结构模型, 用于异常检测. 该方法的主要特点是: 能够自动地选择正常行为模式的种类和样本以建立正常行为模型; 能够在较少样本的情况下避免隐马尔可夫模型欠学习的问题, 建立有效的异常行为模型. 实验结果表明, 该算法与其他方法相比具有更高的可靠性.

**关键词:** 行为建模; 异常检测; 半监督学习; 隐马尔可夫模型; 计算机视觉

中图法分类号: TP18 文献标识码: A

\* Supported by the National Natural Science Foundation of China under Grant No.60303021 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2005AA118020 (国家高技术研究发展计划(863))

Received 2006-04-03; Accepted 2006-06-09

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<sup>[1,2]</sup>.行为分析的最终目标是要通过对行为特征数据的分析来获取行为的语义描述与理解,而行为建模和检测是达到上述目标的一个关键步骤.在很多实际应用(如基于行为的视频检索、视觉监控)中,人们往往更为关注异常行为的检测.因此,本文主要讨论行为建模与异常检测,其中,我们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.

传统的基于监督学习的建模方法<sup>[3-7]</sup>虽然能够在固定的场景下建立准确的行为模型,但是需要手工标记大量的行为序列以获取足够的训练样本,这会造成大量人力资源的浪费.因此,人们又提出了基于无(半)监督的建模方法<sup>[8-14]</sup>.这类方法能够自动地(半自动地)建立行为模型,可以减轻人的负担,同时增强算法的适用性.Zelnik-Manor 与 Irani<sup>[10]</sup>首次通过多时间尺度特征和谱聚类的方法自动地建立了行为事件模型.这种方法简单,适用于大量数据中的事件检测、检索以及时间序列分割等.但是他们在文献[10]中同时也指出,这种利用多时间尺度特征建立的统计距离判别模型没有较好的行为识别精度.Zhong 等人<sup>[11]</sup>提出了一种无监督的异常检测方法,这种方法没有直接建立行为模型,而是通过聚类分析和简单的相似性度量来判别异常行为,因此只适用于离线状态下大量数据中的异常检测问题.Xiang 与 Gong 在文献[12]中提出了一种新的行为建模与异常检测方法:首先,通过谱聚类的方法自动地建立正常行为的隐马尔可夫模型(hidden Markov model,简称 HMM);然后,通过给定的阈值来判断异常行为.这种方法不仅解决了行为的时变问题,而且能够进行实时的异常检测.该方法的不足之处在于:在建立相似性矩阵的过程中,对每个行为的时间序列片段直接学习得到了一个 HMM,这样会由于欠学习而带来 HMM 参数估计的不可靠性问题<sup>[15]</sup>;此外,在异常检测中利用固定的阈值会漏检一些与正常行为较相似的异常行为.解决漏检的一个有效途径是,在建立正常行为模型的同时也建立各种异常行为的模型.异常行为的样本通常是少量的,这给直接建模带来了一定的困难.Zhang 等人<sup>[13]</sup>提出了一种半监督建模方法,即首先通过基于监督学习的方法建立一种正常行为模型,然后通过迭代自适应的方法在此基础上自动地建立异常行为的模型.这种方法考虑了存在一种正常行为模型下的异常行为建模问题,没有讨论存在多种正常行为模型的情况;此外,也没有讨论最佳迭代次数以确定异常行为的模型个数这一问题.

在本文中,我们提出了一种新的基于半监督<sup>\*</sup>学习的行为建模与异常检测方法.不同于上述方法<sup>[10-12]</sup>,本文的方法采用基于动态时间归整(dynamic time warping,简称 DTW)的归一化距离来建立相似性矩阵,并将它用于谱聚类.DTW 能够解决行为的时变问题,是一个动态规划过程.为了减少运算量,我们从大量样本中先随机抽取部分行为样本利用基于 DTW 的谱聚类进行样本初分类,然后通过迭代学习的方法在大样本下进一步学习正常行为的 HMM.这种方法不仅能够自动地选择正常行为模式的种类和样本来建立正常行为的 HMM,解决行为的时变问题,而且还能有效地避免 Xiang 与 Gong<sup>[12]</sup>方法中的 HMM 参数估计的不可靠性问题.同时,本文算法以监督的方式,从多种正常行为的 HMM 中利用最大后验(maximum a posteriori,简称 MAP)自适应的方法得到了所有异常行为的 HMM,从而克服了 Zhang 等人<sup>[13]</sup>的方法只适用于存在一种正常行为模式下的局限性.最后,通过正常行为和异常行为的模型,建立了异常行为的判别模型,避免了文献[12]中方法的漏检问题,从而降低了错误率.

在后面的几节中,将对算法的关键步骤进行详细介绍.

## 1 时间序列分割与表示

在时间序列的分割中,通常有 3 种方法:(1) 根据信号的间断点分割;(2) 根据信号的突变点分割;(3) 按照等长有重叠的时间窗口进行分割.根据不同的应用场景,可以选择不同的分割方法.

---

\* 本文在建立异常行为模型的过程中,充分利用了正常行为无监督学习的结果和异常行为的标记样本,虽然与机器学习中的“半监督”概念有着明显的不同,但从异常行为模型获取的整个过程来看,与上述概念仍有相似之处,所以仍然将本文的学习方法称作“半监督”学习方法.

一个连续的视频序列  $V$  被分割成  $N$  小段,即  $V=\{v_1, v_2, \dots, v_n, \dots, v_N\}$ .理想情况下,分割后的每段表示一个行为.对每帧图像提取一个特征向量  $D_t=[d_{t1}, d_{t2}, \dots, d_{ti}, \dots, d_{tj}]$ (其中: $D_t$  表示第  $t$  帧图像的特征向量; $d_{ti}(1 \leq i \leq J)$  表示第  $t$  帧图像的第  $i$  个特征分量, $J$  为特征分量的个数),则每段序列可以表示为  $v_n=\{D_{n1}, D_{n2}, \dots, D_{nm}, \dots, D_{nM}\}$ (其中: $D_{nm}$  表示第  $n$  段视频序列第  $m$  帧图像的特征向量; $M$  表示第  $n$  段视频序列的长度).

本文利用等长有重叠的分割方法对序列进行了分割,采用时空滤波的方式进行了特征提取,具体细节见实验部分.

## 2 正常行为模型的建立

DTW 和 HMM 是时间序列建模过程中经常使用的数学工具<sup>[1,16]</sup>.它们都能较好地解决行为的时变问题.但是,DTW 需要为每类行为建立一个模板,是一个动态规划过程,具有较大的计算量.HMM 在学习能力和处理未分割的数据流方面比 DTW 具有更大的优越性<sup>[1,2]</sup>,但在使用中需要考虑初始化问题,因为 HMM 在少量样本的情况下,直接学习 HMM 属于欠学习问题<sup>[15]</sup>,得到的参数一般是不可靠的.本文的方法充分利用了它们各自的特点,提出了以下建模方法:首先,利用基于 DTW 的谱聚类方法得到适量的样本,对各类正常行为的 HMM 进行初始化;然后,利用阈值模型<sup>[16]</sup>进一步选择各类正常行为的样本,通过迭代学习的方法进一步训练 HMM.

为了减少计算量,我们从  $N$  段视频序列中随机抽取  $H$  段(一般取  $H=N/10$ )用于谱聚类.所抽取数据的集合表示为  $V'=\{v'_1, v'_2, \dots, v'_H\}$ ,将  $V$  中剩余的样本序列集合称为  $A$ .

### 2.1 基于DTW的谱聚类与行为的HMM形式

给定两个时间序列信号  $Q$  与  $P$ (设  $Q$  长为  $m$ , $P$  长为  $n$ ),本文中采用归一化后的距离来评价两个时间序列信号间的相似性

$$s=\exp[-\gamma(m,n)/G\sigma] \quad (1)$$

其中: $\gamma(m,n)$ 表示通过 DTW 得到的两个时间序列信号  $Q$  与  $S$  的最小累加距离; $G$  表示通过 DTW 得到的最佳路径长度,其取值范围为  $\max(m,n) \leq G \leq m+n-1$ ; $\sigma$  为常数因子.

对于随机抽取的  $H$  段,根据相似性距离计算公式(1),我们可以得到一个  $H \times H$  的相似性对称矩阵  $S=[s_{mn}]_{H \times H}$ .设  $S$  的特征值为  $\lambda_i(i \in [1, H]), \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_H$ ,其方差贡献率  $\mu$  定义为

$$\mu = \frac{\sum_{i=1}^L \lambda_i}{\sum_{j=1}^H \lambda_j} \quad (L \leq H) \quad (2)$$

选取方差贡献率大于  $T_\mu$ (本文取  $T_\mu=0.8$ )对应的  $L$  作为初始聚类数,利用谱聚类(关于谱聚类的详细介绍,请参考文献[17])的方法,将  $H$  段聚成  $L$  类,取样本数大于  $T$ (本文取  $T=H/10$ )的  $C$  类( $C \leq L$ )作为时间序列中大量出现的行为事件(正常行为)类别,分别记为  $W_i(i=1, 2, 3, \dots, C)$ .

利用  $C$  类中的每一类样本,建立一个 HMM.每个 HMM 包含  $e$  个隐藏节点,每个隐藏节点的输出概率密度函数为高斯混合函数:

$$p(y|\alpha, \theta) = \sum_{k=1}^K \alpha_k p_k(y|\mu_k, \Sigma_k) \quad (3)$$

其中: $y$  是各分量相互独立的  $D$  维随机变量; $p_k(y|\mu_k, \Sigma_k)$  是第  $k$  个成分的高斯密度函数,其均值为  $\mu_k$ ,协方差为  $\Sigma_k$ ; $\theta$  是成分密度函数的参数所构成的向量,包括各个高斯密度函数的均值  $\mu$ ,协方差  $\Sigma$ ; $\alpha=(\alpha_1, \alpha_2, \dots, \alpha_k)$ ( $\sum_{k=1}^K \alpha_k = 1$ )是混合系数向量, $\alpha_k$  表示第  $k$  个成分被选择的概率.

我们采用 BIC(Bayesian information criterion)<sup>[18]</sup>来自动确定混合高斯的成分个数:

$$\hat{K} = \arg \min_K \left\{ -\sum_{i=1}^T \log p(y_i|\alpha, \hat{\theta}(K)) + \frac{\zeta(K)}{2} \log T \right\} \quad (4)$$

其中: $T$  表示整个时间序列的长度; $\zeta(K)$  表示混合模型的参数个数.

## 2.2 阈值模型

利用上节  $C$  个已经初始化的 HMM 构造一个如图 1 所示的阈值模型<sup>[14,16]</sup>.它是一个各态遍历的 HMM.具体构造过程是:将  $C$  个 HMM 的所有隐藏节点作为阈值模型的隐藏节点,则隐藏节点个数  $N=e*C$ ,各状态的初始概率均设为  $1/N$ ,并且在新模型中保持各隐藏节点的自转移概率和输出概率不变,各隐藏状态之间的转移概率由下式给出:

$$a_{ij}=(1-a_{ii})/(N-1) \quad (i \neq j) \quad (5)$$

其中, $a_{ii}$  是第  $i$  个隐藏节点的自转移概率.图 1 中两个黑色的圆点分别表示虚拟的开始节点和结束节点,它们均没有输出,其间的转移概率为 1.

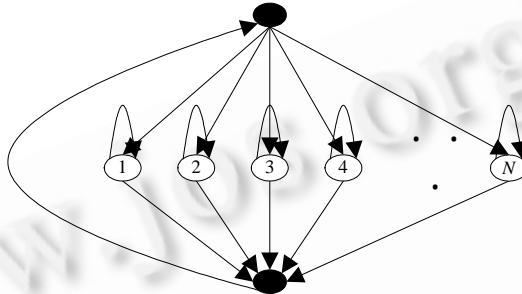


Fig.1 The structure of the threshold model

图 1 阈值模型结构

设当前观察序列为  $Y$ ,第  $i$  个 HMM 的参数集合为  $\lambda_i(i=1,2,3,\dots,C)$ ( $\lambda_i$  包括第  $i$  个 HMM 的初始状态分布向量、状态转移概率矩阵和输出概率分布矩阵),阈值模型的参数为  $\lambda_T$ ,如果  $Y$  属于第  $i$  类,则

$$i = \max_j \{P(Y|\lambda_j) | P(Y|\lambda_j) > P(Y|\lambda_T)\} \quad (j=1,2,3,\dots,C) \quad (6)$$

其中: $P(Y|\lambda_j)$  是第  $j$  个 HMM 产生观察序列  $Y$  的概率; $P(Y|\lambda_T)$  是阈值模型产生观察序列  $Y$  的概率.

## 2.3 迭代学习

通过基于 DTW 的聚类方法,我们得到了反映长时间序列  $V$  主要信息的  $C$  类行为的一些样本序列.利用这些样本,分别对每类正常行为的 HMM 进行初始化.由于这些样本的个数远少于整体样本数,为了增强各 HMM 的鲁棒性,使  $C$  个 HMM 模型更准确地反映时间序列的主要信息,我们使用了迭代学习的方法进一步训练各个 HMM,主要过程如下:

- (1) 通过阈值模型和已建立的各类正常行为的 HMM,判断  $A$  中每个样本序列的类别,并通过相应的 HMM 计算产生该样本序列的概率.设该步获取的每类正常行为的样本个数为  $n_i(i=1,2,\dots,C)$ ;
- (2) 从  $A$  中无放回地取各类中概率最大的  $n_i/5$  个(若  $n_i < T'$ ,则停止提取该类样本.其中, $T'$  为给定阈值)样本序列分别放入到相应的  $w_i$  中;
- (3) 利用  $w_i$  中的所有样本序列,以各类 HMM 的已有参数为初始值,通过 EM 算法,进一步训练 HMM 参数;
- (4) 根据(3)获取的参数更新阈值模型的参数;
- (5) 利用更新后的阈值模型和各类 HMM,判断  $A$  中是否存在满足各类正常行为 HMM 的序列样本:若存在,转到(1);否则,停止.

## 2.4 正常行为的HMM拓扑结构模型

在获取了  $C$  个 HMM 模型之后,我们设计了一个如图 2 所示的 HMM 拓扑结构模型作为正常行为模型.

图 2 所示是一个具有  $C+2$  个节点的 HMM 拓扑结构.两个黑点分别表示虚拟的开始节点和结束节点,两个节点之间的转移概率为 1.每个正常行为的 HMM 作为该拓扑结构的一个状态,各状态之间具有相等的转移概率  $1/C$ .

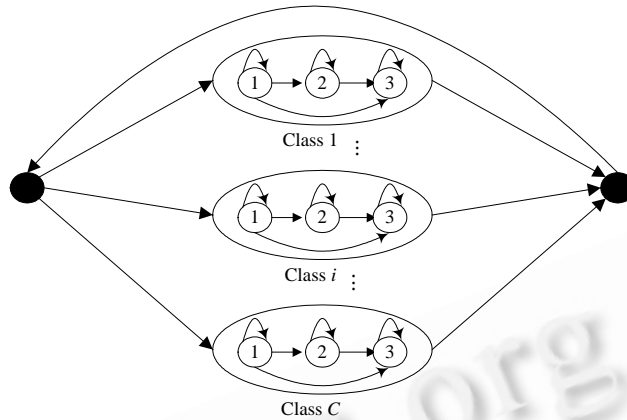


Fig.2 The structure of normal behavior model

图 2 正常行为模型结构

### 3 异常行为模型的建立

在 HMM 中,各状态之间的转移概率总是取固定的几个值,取值范围比较小;而各状态的输出概率的取值范围相对较大.所以在 Viterbi 解码算法中,状态的输出概率起主要作用<sup>[19]</sup>.因此,针对少样本行为,借助于语音识别中常用的方法<sup>[20]</sup>,我们可以在保持已有 HMM 的状态转移概率不变的条件下,通过 MAP 自适应的方法调整各状态的输出概率密度函数的参数,得到少样本异常行为的 HMM,从而避免 HMM 在少样本下的欠学习问题<sup>[15]</sup>.

#### 3.1 最大后验自适应(MAP adaptation)

通常情况下,最大似然估计在少量样本的情况下不能准确地估计参数.最大后验估计(MAP estimation)也称作贝叶斯学习,在训练过程中引入了先验知识,适用于少量样本的参数训练<sup>[20]</sup>.

设  $p(X|\theta)$  是相对于样本集  $X=\{x_1, x_2, \dots, x_N\}$  的似然函数,  $p(\theta)$  为先验分布,  $\theta$  为似然函数参数,根据最大后验原理,我们选择使后验概率密度最大的参数值作为当前参数的估计,即

$$\theta_{MAP} = \arg \max_{\theta} p(\theta | X) = \arg \max_{\theta} p(X | \theta) p(\theta) \tag{7}$$

根据式(7),可以将已知的 HMM 参数作为先验知识,利用新的样本序列(本文中是指手工标记的异常行为样本序列),自适应得到一个新的 HMM.在自适应过程中,HMM 的初始概率和状态转移概率保持不变.设当前样本序列为  $X=\{x_1, x_2, \dots, x_t, \dots, x_T\}$ ,类似于文献[13,20],采用两步法来估计未知的参数(包括各个混合高斯成分的混合系数  $\alpha$ 、均值  $\mu$  和方差  $\sigma$ ):

(1) 根据已知 HMM,计算当前样本的统计值

对于 HMM 状态  $s$  的输出概率密度(具有  $K$  个分量的混合高斯)的第  $i$  个分量,有

$$p_s(i | x_t) = \frac{\alpha_s p_{s_i}(x_t)}{\sum_{j=1}^K \alpha_s p_{s_j}(x_t)} \tag{8}$$

$$\alpha_{s_i}^{new} = \frac{1}{T} \sum_{t=1}^T p_s(i | x_t) \tag{9}$$

$$\mu_{s_i}^{new} = \frac{\sum_{t=1}^T x_t p_s(i | x_t)}{\sum_{t=1}^T p_s(i | x_t)} \tag{10}$$

$$\sigma_{s_i}^{new} = \frac{\sum_{t=1}^T p_s(i | x_t) (x_t - \mu_{s_i}^{new})(x_t - \mu_{s_i}^{new})^T}{\sum_{t=1}^T p_s(i | x_t)} \tag{11}$$

(2) 根据已知 HMM 参数和当前样本的统计值,自适应地得到新 HMM 的参数

新 HMM 的初始概率和状态转移概率与原 HMM 保持一样,其输出概率的参数由以下自适应公式得到:

$$\hat{\alpha}_{s_i} = \rho \cdot \alpha_{s_i}^{new} + (1 - \rho) \cdot \alpha_{s_i}^{old} \quad (12)$$

$$\hat{\mu}_{s_i} = \rho \cdot \mu_{s_i}^{new} + (1 - \rho) \cdot \mu_{s_i}^{old} \quad (13)$$

$$\hat{\sigma}_{s_i} = \rho \cdot \sigma_{s_i}^{new} + (1 - \rho) \cdot [\sigma_{s_i}^{old} + (\hat{\mu}_{s_i} - \mu_{s_i}^{old})(\hat{\mu}_{s_i} - \mu_{s_i}^{old})^T] \quad (14)$$

其中, $\rho$ 为比例因子,且  $0 \leq \rho \leq 1$ .

### 3.2 自适应建立异常行为模型

由于异常行为的样本是少量的,所以花少量的人力就可以进行标记.而已经获取的正常行为模型是通过大量的样本训练得到的,它们反映了整个视频序列的主要信息,而且一般与异常行为存在着相关性,所以,可以通过以下过程得到各类异常行为的 HMM.首先,从  $C$  类正常行为的 HMM 中选取对当前异常行为样本估计概率最大的一个 HMM;然后,将该 HMM 的参数作为先验知识,通过 MAP 自适应方法得到相应异常行为的 HMM.

得到各类异常行为的 HMM 之后,同样可以建立如图 2 所示的异常行为的 HMM 拓扑结构模型.所不同的是,此时,节点的个数变为  $C'+2$ (其中, $C'$ 为异常行为的 HMM 个数),相应的状态由原来正常行为的 HMM 变为异常行为的 HMM,转移概率为  $1/C'$ .

## 4 异常检测

设正常行为的 HMM 拓扑结构模型的参数为  $\lambda_M$ ,异常行为的 HMM 拓扑结构模型的参数为  $\lambda_{AM}$ ,观察到的时间序列为  $O$ ,则当  $P(O|\lambda_M) > P(O|\lambda_{AM})$  时,观察到的是正常行为;否则是异常行为.当判断是正常行为时,可以进一步判断正常行为的类别:  $\hat{k} = \arg \max_k \{P(O|B_k)\}$ ,其中:  $1 \leq k \leq C; B_k$  是第  $k$  个正常行为的 HMM 参数.同样,当判断为异常行为时,可以由下式判断异常行为类别:  $\hat{v} = \arg \max_v \{P(O|\bar{B}_v)\}$ ,其中:  $1 \leq v \leq C'; \bar{B}_v$  表示第  $v$  个异常行为的 HMM 参数.

## 5 实验结果

本文采用平均错误率 HTER(half-total error rate)作为算法的评价标准<sup>[13]</sup>:

$$HTER = \frac{FAR + FRR}{2},$$

其中:错误报警率 FAR(false alarm rate)是将正常行为判为异常行为的错误率;错误拒绝率 FRR(false rejection rate)是对异常行为拒绝报警的错误率.计算公式分别为

$$FAR = \frac{\text{误报警数}}{\text{正常行为的样本数}} \times 100\%,$$

$$FRR = \frac{\text{拒报警数}}{\text{异常行为的样本数}} \times 100\%.$$

### 5.1 实验数据

已知文献中报道的行为分析算法通常在各自的小规模数据库上进行性能评估.我们从 Schuldt 等人<sup>[21]</sup>的一个数据库(<http://www.nada.kth.se/cvap/actions/>)中获取了如图 3 所示的 5 种行为的 21 327 帧图像(其中:“行走”7 584 帧;“挥手”8 043 帧;“拳击”、“鼓掌”、“慢跑”各 1 900 帧)进行实验.由于这个数据库中每个单独的视频序列仅包含一种行为,所以,在实验中使用的是由这 5 种行为的时间序列合成的数据.具体合成过程为:(1) 先对每种行为的时间序列进行无重叠的分割.对于正常行为的时间序列,每次取大于 100、小于 400 的随机整数作为每次分割的片段长度值,当该时间序列长度不大于 100 帧时,停止分割;对于异常行为的时间序列,每次取大于 30、小于 60 的随机整数作为每次分割的段长度值,当该时间序列长度不大于 30 帧时,停止分割;(2) 对所有的行为片

段进行随机排序,获取混合行为的时间序列.



Fig.3 The images of five behaviors' samples from the dataset of Schudlt, *et al.*

图3 Schudlt 等人数据库中 5 种行为的样本图像

通过上述合成过程,分别获取训练样本序列和测试样本序列.在实验中,采用了 8 种不同的训练样本序列:保持正常行为“行走”、“挥手”各 5 000 帧不变的情况下,分别混入异常行为“拳击”、“鼓掌”、“慢跑”各 60 帧、150 帧、350 帧、500 帧、700 帧、1 000 帧、1 300 帧和 1 500 帧;测试样本序列包括正常行为“走”2 584 帧,“挥手”3 043 帧;异常行为“拳击”、“鼓掌”、“慢跑”各 400 帧.在实验中,我们利用以上方法获取的时间序列数据在不同的参数(见第 5.3 节)和 4 种相关算法(见第 5.4 节)等多种情况下分别进行了一次实验.

5.2 时间序列分割与特征提取

我们利用等长有重叠的分割方式来分割所合成的时间序列,以获取相应的行为样本.在实验中,采用了宽为 35 帧,步长为 10 帧的滑动窗口.采用时空滤波的方式来提取行为的运动特征,具体操作为:

- (1) 对时间序列的每一帧图像进行时空滤波<sup>[11]</sup>:

$$I_t(x,y,t)=\|I(x,y,t)*G_t*G_{x,y}\|_2,$$

其中:  $G_t = te^{-(t/\sigma_t)^2}$ ,  $G_{x,y} = e^{-[(x/\sigma_x)^2+(y/\sigma_y)^2]}$ ;

- (2) 对获取的图像,利用阈值  $Th$  获取运动信息的二值图像(如图 4(b)所示);
- (3) 将前景运动区域矩形分成  $9 \times 5$  个小块(如图 4(c)所示),统计每个小块的前景像素点的个数,记为  $u_i, 1 \leq i \leq 45$ ,并将其归一化

$$d_i = u_i / \max(u_i), 1 \leq i \leq 45;$$

- (4) 将这些统计量按行排列,构成当前图像的特征向量

$$D_i = [d_{i1}, d_{i2}, \dots, d_{i9}, \dots, d_{i45}], J = 45.$$

在实际使用中,我们通过 PCA 将 45 维的特征向量降到了 15 维.

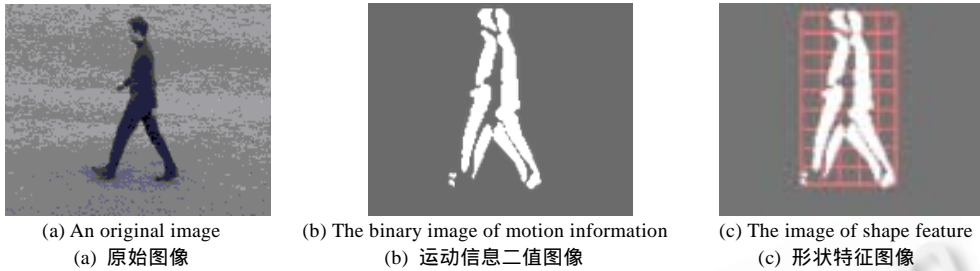
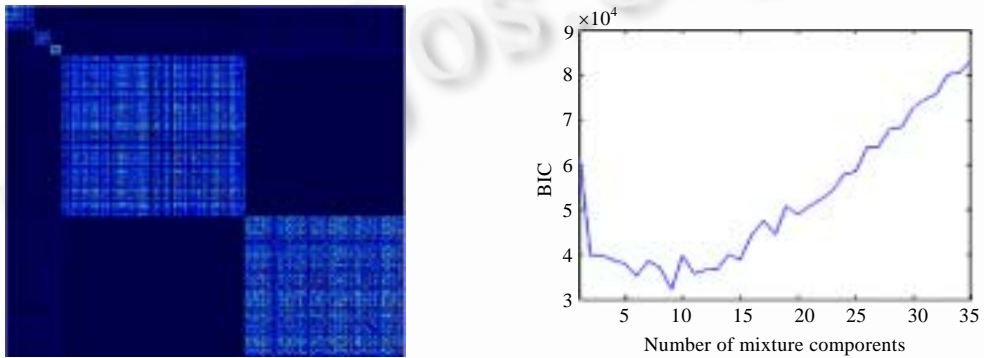


Fig.4 Examples of feature extraction  
图 4 特征提取示例

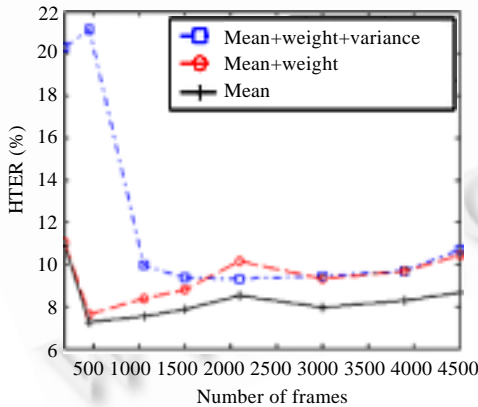
5.3 聚类与参数选择

图 5 给出了聚类与参数选择结果示意图.

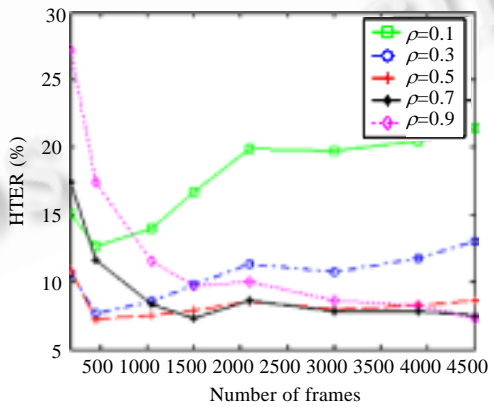


(a) The affinity matrix ordered according to the result of clustering  
(a) 谱聚类重排序后的相似性矩阵

(b) BIC curve  
(b) BIC 曲线图



(c) Results of abnormality detection for adapting different parameter combinations  
(c) 选择不同自适应参数的异常检测结果



(d) Results of abnormality detection for different  $\rho$  values  
(d) 选择不同比例因子的异常检测结果

Fig.5 Results of clustering and parameter selection

图 5 聚类与参数选择结果

图 5(a)表示通过第 5.1 节的方法合成的训练样本序列(其中包括异常行为样本‘拳击’、‘鼓掌’、‘慢跑’各 350 帧)经分割、采样、谱聚类、重排序后的相似性矩阵.此步骤的主要目标是获取适量的正常行为样本.可以看出,训练集被聚成了 5 类,其中:样本最多的两类就分别表示两类正常行为‘行走’与‘挥手’的样本;其余 3 类主要包含了异常行为样本和一些误分样本.图 5(b)是利用式(6)得到的 BIC 曲线图,当  $k=9$  时 BIC 值达到了最小值,所以在



实验中选择的高斯混合成分个数为 9.图 5(c)和图 5(d)分别表示选择不同自适应参数、比例因子的异常检测结果曲线图,横坐标表示在训练集中加入异常行为样本的总帧数(各种异常行为的样本量相等),纵坐标表示平均错误率 HTER.图 5(c)中的 3 条曲线分别表示:当比例因子  $\rho=0.5$  时:(1) 对均值、混合系数和方差 (mean+weight+variance)都进行自适应的 HTER 曲线;(2) 对均值和混合系数(mean+weight)进行自适应的 HTER 曲线;(3) 仅对均值(mean)进行自适应的 HTER 曲线.当仅对均值进行自适应时,平均错误率 HTER 最小,这是因为在少量的样本下,对过多的参数进行自适应,会降低性能.图 5(d)中 5 条曲线分别表示:仅在对均值进行自适应的情况下,比例因子  $\rho$  分别取 0.1,0.3,0.5,0.7,0.9 时的平均错误率 HTER 曲线图.当样本比较少时, $\rho$  取较小的值,有较好的性能(一般取  $0.25 \leq \rho \leq 0.5$ );反之, $\rho$  取较大的值(一般取  $0.7 \leq \rho \leq 0.9$ ),有较好的性能.其原因可能为:当样本较少时,估计参数对先验有较大的依赖;否则,估计参数对当前样本信息有较大的依赖.

在后面的实验中,我们选择比例因子  $\rho=0.5$ ,仅对均值进行自适应算法.

### 5.4 算法比较与分析

我们将本文方法和下述 3 种方法进行了比较:(1) 基于监督的方法(supervised method).对所有样本均手工标记,然后通过 EM 算法学习各个 HMM;(2) 基于半监督的方法(semi-supervised method).首先通过本文自动建立正常行为模型的方法建立正常行为的 HMM,然后手工标记异常行为样本,通过 EM 算法学习得到异常行为的 HMM;(3) 另一种半监督方法(XiangTao+MAP).采用 Xiang 和 Gong<sup>[12]</sup>的方法,自动建立正常行为的 HMM,然后手工标记异常行为样本,通过本文 MAP 自适应方法得到异常行为的 HMM.

图 6 给出了各种算法的比较曲线图,横、纵坐标的意义均与图 5(c)相同.

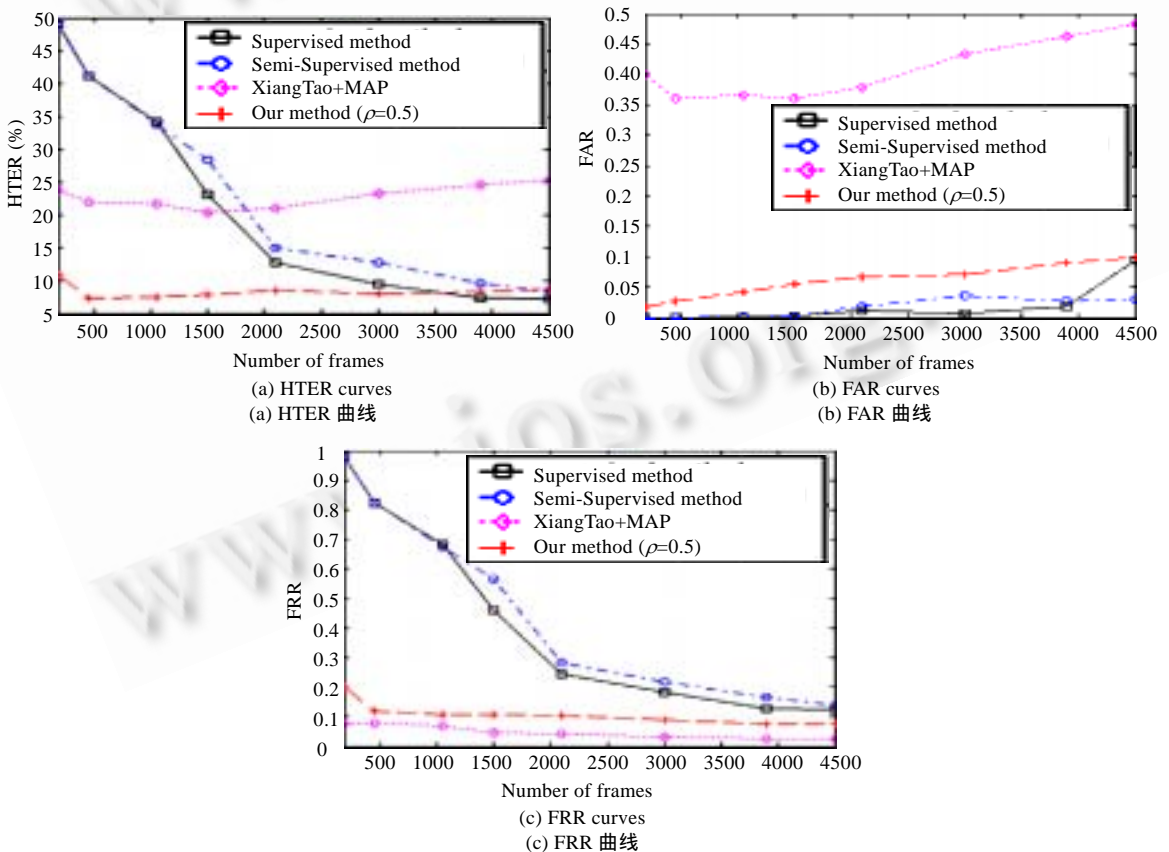


Fig.6 Comparison of the related algorithms

图 6 相关算法结果的比较

从图中可以看出:方法(1)、方法(2)在少量异常行为样本的情况下有较高的错误拒绝率和平均错误率,但随着样本的增加,错误拒绝率和平均错误率越来越低.这是因为在少量样本的情况下,直接利用 EM 算法学习 HMM 属于欠学习问题,得到的参数是非常不可靠的,当存在大量样本时,这样学习得到的参数才是可靠的;方法(3)具有较高的错误报警率和平均错误率,主要是因为直接采用了 HMM 距离<sup>[12]</sup>建立相似性矩阵,没有考虑 HMM 在少量样本下的初始化问题.而本文在建立 HMM 过程中,利用基于 DTW 的谱聚类方法选择适量的初始化样本,有效地避免了这个问题.同时,我们的方法在少量异常行为样本的情况下,采用 MAP 自适应的方法建立了异常行为的 HMM,所以,在少量样本的情况下也能得到很好的检测结果.

## 6 结论与展望

人的运动行为分析是近年来被广泛关注的研究热点.本文提出了一种基于半监督学习的行为建模与异常检测方法.该方法能够在大量样本的情况下自动地选择正常行为模式的种类和样本来建立正常行为模型,能够在少样本的情况下避免欠学习问题,建立可靠的异常行为模型.实验同时验证了该算法的有效性.

本文算法适用于训练集包括各类行为样本的情况.未来的工作将集中于获取更多的数据,对算法进行性能测试和研究无法预先获取每类异常行为样本情况下的行为建模与异常检测问题.此外,本文算法中,正常行为的种类在谱聚类过程中被选择之后,在学习过程中并没有被进一步调整改变.这在正常行为的训练样本集为大样本集的情况下是适用的;当有部分正常行为的训练样本集为小样本集时,基于谱聚类确定的正常行为的种类需要进一步调整取优.因此,在少样本训练集下选择正常行为的种类和异常检测也是我们下一步要研究的重点.

### References:

- [1] Wang L, Hu WM, Tan TN. Recent developments in human motion analysis. *Pattern Recognition*, 2003,36(3):585–601.
- [2] Hu WM, Tan TN, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. on Systems, Man, and Cybernetics—Part C*, 2004,34(3):334–352.
- [3] Haritaoglu I, Harwood D, Davis LS. W<sup>4</sup>: Real-Time surveillance of people and their activities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000,22(8):809–830.
- [4] Bobick AF, Davis JW. The recognition of human movement using temporal templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001,23(3):257–267.
- [5] Gong SG, Xiang T. Recognition of group activities using dynamic probabilistic networks. In: Werner B, ed. *Proc. of the Int'l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2003. 742–749.
- [6] Li H, Greenspan M. Multi-Scale gesture recognition from time-varying contours. In: Werner B, ed. *Proc. of the Int'l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2005. 236–243.
- [7] Laptev I, Linderberg T. Space-Time interest points. In: Werner B, ed. *Proc. of the Int'l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2003. 432–439.
- [8] Hu WM, Xie D, Tan TN, Shen J. A hierarchical self-organizing approach for learning the patterns of motion trajectories. *Chinese Journal of Computers*, 2003,26(4):417–426 (in Chinese with English abstract).
- [9] Shechtman E, Irani M. Space-Time behavior based correlation. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Press, 2005. 405–412.
- [10] Zelnik-Manor L, Irani M. Event-Based analysis of video. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2001. 123–130.
- [11] Zhong H, Shi JB, Visontai M. Detecting unusual activity in video. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2004. 819–826.
- [12] Xiang T, Gong SG. Video behaviour profiling and abnormality detection without manual labeling. In: Werner B, ed. *Proc. of the Int'l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2005. 1238–1245.
- [13] Zhang D, Gatica-Perez D, Bengio S, McCowan I. Semi-Supervised adapted HMMs for unusual event detection. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2005. 611–618.

- [14] Ma GY. Temporal sequence clustering and recognition [Ph.D. Thesis]. Beijing: Tsinghua University, 2004 (in Chinese with English abstract).
- [15] Oliver N, Garg A, Horvitz E. Layered representations for learning and inferring office activity from multiple sensory channels. *Computer Vision and Image Understanding*, 2004,96(2):163–180.
- [16] Lee HK, Kim JH. An HMM-based threshold model approach for gesture recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1999,21(10):961–973.
- [17] Weiss Y. Segmentation using eigenvectors: A unifying view. In: Werner B, ed. *Proc. of the Int'l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 1999. 975–982.
- [18] Xiang T, Gong SG. Visual learning given sparse data of unknown complexity. In: Werner B, ed. *Proc. of the Int'l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2005. 701–708.
- [19] Bengio Y. Markovian models for sequential data. *Neural Computing Surveys*, 1999,2:129–162.
- [20] Reynolds DA, Quatieri TF, Dunn RB. Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 2000,10(1-3):19–41.
- [21] Schudt C, Laptev I, Caputo B. Recognizing human actions: A local SVM approach. In: Werner B, ed. *Proc. of the Int'l Conf. on Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2004. 32–36.

#### 附中文参考文献:

- [8] 胡卫明,谢丹,谭铁牛,沈俊. 轨迹分布模式学习的层次自组织神经网络方法. *计算机学报*, 2003,26(4):417–426.
- [14] 马赓宇. 时间序列的聚类与识别[博士学位论文]. 北京:清华大学, 2004.



李和平(1978 - ),男,湖北荆门人,博士生,主要研究领域为模式识别,计算机视觉,行为分析.



吴毅红(1973 - ),女,博士,副研究员,博士生导师,主要研究领域为多项式消元理论及应用,几何不变量的计算及应用,几何定理机器证明,摄像机标定,三维重建,视觉几何.



胡占义(1961 - ),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为摄像机标定,三维重建,主动视觉,机器人导航,基于图像的建模和绘制.



吴福朝(1957 - ),男,研究员,博士生导师,主要研究领域为计算机视觉,人工智能.