

视频监控中人体行为识别的研究*

温向兵¹, 满君丰^{1,2}

(1.湖南工业大学 计算机与通信学院, 湖南 株洲 412008;

2.中南大学 信息科学与工程学院, 湖南 长沙 410083)

摘要: 提出一种人体行为识别模型和前景提取方法。针对人体运动过程中产生新的行为问题, 该模型用分层 Dirichlet 过程聚类人体特征数据来判断人体运动过程中是否有未知的人体行为模式; 用无限隐 Markov 模型对含有未知行为模式的特征向量进行行为模式的有监督的学习, 由管理者将其添加到规则与知识库中。当知识库的行为模式达到一定规模时, 系统便可以无监督地对人体行为进行分析, 其分析采用 Markov 模型中高效的 Viterbi 解码算法来完成。对于前景的提取, 提出了基于背景边缘模型与背景模型相结合的前景检测方法, 此方法能够有效避免光照、阴影等外部因素的影响。仿真实验证明, 本文提出的方法在实时视频监控中的人体行为识别方面有独特的优势。

关键词: 行为模式; 嵌套的狄利克雷过程; 无限隐 Markov 模型; 行为识别

中图分类号: TP391.4

文献标识码: A

文章编号: 1674-7720(2010)23-0026-05

Study of human behavior recognition based on video surveillance

WEN Xiang Bing¹, MAN Jun Feng^{1,2}

(1.College of Computer and Communication, Hunan University of Technology, Zhuzhou 412008, China;

2.College of Information Science and Technology, Central South University, Changsha 410083, China)

Abstract: A behavior recognition model and a model-based approach to extract foreground are presented. Because the process of human motion are likely to result in new events and behaviors, in this model, hierarchical Dirichlet process is used to cluster feature data monitored of the human body to determine whether there are unknown behaviors. Infinite hidden Markov model is used to learn unknown behavior patterns with supervise, and then update the knowledge base and rule base. When the rules and knowledge repository reach a certain scale, system can analyze behaviors without supervise. The Viterbi decoding algorithm in Markov model is adopted to analyze the current behavior of the human body. Foreground detection method is proposed based on the background edge model combined with background model, which can effectively avoid the light, shadows and other external factors. The simulation experiments show that this method in real-time video surveillance in the detection of human behavior has an unique advantage.

Key words: behavior pattern; nested Dirichlet process; infinite hidden Markov model; behavior recognition

人的行为理解与描述是近年来被广泛关注的研究热点, 它是指对人的运动模式进行分析和识别, 并用自然语言等加以描述。行为理解可以简单地被认为是时变数据的分类问题, 即将测试序列与预先标定的代表典型行为的参考序列进行匹配。对于人的行为识别, 参考文献[1]概括为以下两种方法:

(1)模板匹配方法。参考文献[2-5]都采用模板匹配

* 基金项目: 湖南工业大学研究生创新基金 (CX1003); 国家自然科学基金 (60773110); 湖南省自然科学基金 (09JJ6087)

技术的行为识别方法。首先将图像序列转换为一组静态形状模式, 然后在识别过程中和预先存储的行为标本来解释图像序列中人的运动。

(2)空间方法。基于状态空间模型的方法定义每个静态姿势作为一个状态, 这些状态之间通过某种概率联系起来。目前, 状态空间模型已经被广泛地应用于时间序列的预测、估计和检测, 最有代表性的是 HMMs。每个状态中可用于识别的特征包括点、线或二维小区域。

本文从两个方面来阐述视频监控系统中的人体行

为识别:(1)行为描述,即在视频帧中提取人体特征,并对人体行为进行描述;(2)行为识别,通过基于数理统计的 Markov 模型训练得到的行为检测器来实现。针对行为描述,本文采用背景边缘法来提取视频前景,通过背景边缘法来获取人体的边界轮廓,背景法可获取前景人体区域。由于背景法受光照影响较大,通过这种方法提取的人体区域不够完整,但通过人体边界和人体区域相加,再进行形态学的闭运算,就能得到较完整的前景目标。对于行为识别,首先利用 HDP-iHMM 进行人体行为状态的确定,即确定是否有新的人体行为模式产生,如果有新的行为状态,则进行 iHMM 的行为模式的学习;如果没有新的行为状态,则用已训练的 HMM 进行行为检测。

本文的创新点是在人体前景获取的过程中利用了两种背景模型的结合。在行为检测方面,应用 HDP-iHMM 确定是否有未知人体行为,利用 HMM 来进行行为的检测,这样能使检测系统不断地学习,当知识库的行为模式达到一定规模时,系统便可以无监督地对人体行为进行检测。

1 人体行为描述

参考文献[2]、[6]为了理解人体行为,采用最常用的背景减法来提取运动的人体,利用当前图像与背景图像的差分来检测出前景运动区域的一种技术,但这种方法对光照和外来无关事件的干扰等特别敏感。为了解决这个问题,本文采用背景边界模型和背景模型的结合来检测前景,通过这两种模型的结合,再应用形态学运算,就能获得一个相对较完整的人体前景。

1.1 人体前景提取

背景边缘模型通过统计视频图像中每个位置在连续时间内出现边缘的概率计算得到:

(1)利用背景边界模型来获取前景边缘。设当前时刻帧为 $F(x, y)$, 每帧图像中的位 Location $(x, y) (x=1, 2, \dots, W, y=1, 2, \dots, H)$, 其中 H 和 W 分别为图像帧的宽和高)是背景边缘的概率 $P(x, y)$:

$$P(x, y) = \sum_{k=t-T}^{t-1} \frac{\text{edge}(x, y, k)}{T} \quad (1)$$

式中, $\text{edge}(\dots, \dots, k)$ 为第 k 帧对应的边缘图像, 通过 Sobel 算子检测得到; T 为背景边缘模型的更新时间, 由于同一姿势的滞留时间不会很长, 所以 T 的取值不需很大。若背景边缘概率低于设定阈值 α , 则该边缘点为前景边缘点, 否则背景边缘点为:

$$F_{\text{edge}}(x, y, k) = \begin{cases} 1, & \text{edge}(x, y, k) = 1 \& P(x, y) < \alpha \\ 0, & \text{其他} \end{cases} \quad (2)$$

式中, F_{edge} 为前景边缘。

(2)利用背景模型来获取前景。设背景为 $F_{\text{bg}}(x, y)$, 当前时刻帧为 $F(x, y)$, 则前景 $F_{\text{fg}}(x, y)$ 为:

$$F_{\text{fg}}(x, y) = \begin{cases} 1, & F(x, y) - F_{\text{bg}}(x, y) > 0 \\ 0, & F(x, y) - F_{\text{bg}}(x, y) < 0 \end{cases} \quad (3)$$

(3)通过一些数学运算结合两种模型获取 $f(x, y)$, 然后对

$f(x, y)$ 进行形态学运算, 来填充前景孔洞, 为特征计算奠定基础。

1.2 特征计算

在提取了前景后, 为了分析人的活动和行为模式, 进一步提取和计算一些人体特征数, 本文的研究着重于以下图像特征值:

(1)长宽比(A): $A=L/W$, A 值包含了行为模式识别的重要信息。这一特征可以识别人体是站立或是别的姿势。

(2)矩形度(R): $R=A_0/A_R$, 其中 A_0 是人体的面积, A_R 是最小封闭矩形的面积。矩形拟合因子的值限定在 0 和 1 之间。

(3)协方差矩阵(C):

$$C = \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T \quad (4)$$

坐标值 x_i 是人体范围内的像素值, 这个值包含着重要的信息(如人体的姿势)。

(4)不变矩(M): 具有两个变元的有界函数的矩集被定义为:

$$M_{jk} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k f(x, y) dx dy \quad (5)$$

式中, j, k 取所有的非负整数, 两者产生一个矩的无限集。

相对于主轴计算并用面积规范化的中心矩, 不变矩在物体放大、平移、旋转时保持不变。这些矩的幅值反映了物体的形状并能用于模式识别, 本文着重计算下面这些不变矩:

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \end{aligned} \quad (6)$$

以上所述为本系统所用到的所有特征值, 则人体行为可表示为如下特征向量: $F=[A, R, C, \phi_1, \phi_2, \phi_3, \phi_4]$: 通过特征向量 F 表示的一种人体行为, 将作为 iHMM 的训练样本。

2 行为的识别模型

对未知行为的学习过程如图 1 所示。当 HDP 聚类过

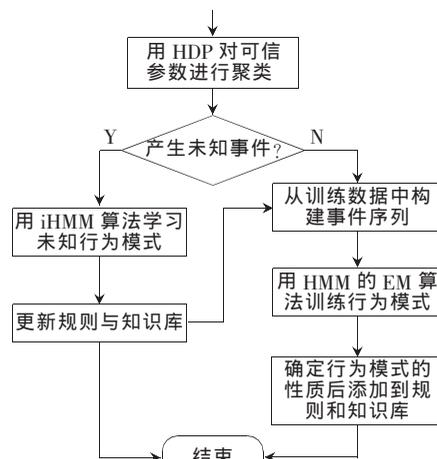


图 1 行为模式的训练过程

程中发现有新行为产生时,则用 iHMM 的 Beam 抽样算法学习未知行为模式,将定性的行为模式添加到规则和知识库。

被观察变量 X 是离散的,有 $r(r \geq 2)$ 个可能状态 x_1, \dots, x_r , 用多项式抽样所得的简单共轭先验就是 Dirichlet 分布:

$$p(\theta|\xi) = \text{dir}(\theta|\alpha_1, \dots, \alpha_r) \equiv \frac{\Gamma(\alpha)}{\prod_{k=1}^r \Gamma(\alpha_k)} \prod_{k=1}^r \theta_k^{\alpha_k-1} \quad (7)$$

式中, $\alpha_k > 0, k=1, \dots, r$; 数量 $\alpha_1, \dots, \alpha_r$ 指区别于 θ 的超参数。

Dirichlet 过程 DP(Dirichlet Process) 混合模型无需事先设定混合分量个数, DP 为混合分量个数提供一个非参数先验,其优点是能将选择模型分量个数和确定模型参数联合进行。多层 Dirichlet 过程 HDP(Hierarchical Dirichlet Process) 着重考虑多个相关数据组的训练问题,每个组用一个无限混合模型描述,混合成分可以跨组被共享。该模型具有混合成份数量无上界且能够从数据中自动推理出。

有限 HMM 包含一个隐含状态序列 $s=(s_1, s_2, \dots, s_r)$ 和相应的观察序列 $o=(o_1, o_2, \dots, o_r)$ 。每个状态变量 s_i 具有有限数量的状态,即 $1 \dots K$ 。由状态转移矩阵控制状态间的转换,本文 $\pi_{ij}=(s_i=j|s_{i-1}=i)$, 而最初的状态概率是 $\pi_{0j}=(s_1=j)$ 。对于每一个状态 $s_i \in \{1 \dots K\}$, 有一个参数 ϕ_{s_i} , 它描述状态的观察相似度: $o_i|s_i \sim F(\phi_{s_i})$ 。给定 HMM 的参数 $\{\pi_0, \pi, \phi, K\}$, 隐含状态 $s(s_0=0)$ 和观察 O 的联合分布可以写成:

$$p(s, o|\pi_0, \pi, \phi, K) = \prod_{i=1}^r p(s_i|s_{i-1})p(o_i|s_i) \quad (8)$$

用无限状态数量获得非参数 HMM 的一个简单的方法是在转换概率的基础上使用对称 Dirichlet 分布先验,即利用参数 α/K , 且 $k \rightarrow \infty$ 。该方法已经成功应用于产生 DP 混合模型,但不适合应用于 HMM。因为转换概率是通过独立先验的形式给出的,没有耦合的跨越不同状态的转换。引入耦合的跨越转换,可以使用分层贝叶斯形式, Dirichlet 先验共享参数且拥有更高层次的先验,即:

$$\pi_k \sim \text{Dirichlet}(\alpha, \beta), \beta \sim \text{Dirichlet}(\gamma/K, \dots, \gamma/K) \quad (9)$$

式中, π_k 是超越状态 K 的转换概率, β 是共享先验参数。当 $k \rightarrow \infty$, 多层先验方法即为多层 Dirichlet 过程 HDP。当描述从 $k \sim k'$ 的转换概率 $\pi_{kk'}$ 和产生的用 ϕ_k 参数化的分布时,要鉴别每一个 G_k , 则正式定义 iHMM:

$$\beta \sim \text{GEM}(\gamma), \pi_k|\beta \sim \text{DP}(\alpha, \beta), \phi_k \sim H$$

$$s_i|s_{i-1} \sim \text{Multinomial}(\pi_{s_{i-1}}), o_i|s_i \sim F(\phi_{s_i}) \quad (10)$$

该多层模型的对应图形化表示如图 2 所示。在本文

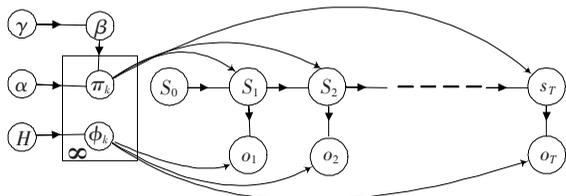


图 2 HDP-iHMM 的图形化表示

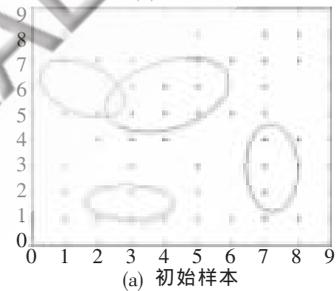
中, $\beta_{k'}$ 为转移到状态 k' 的转换概率的先验均值, α 为控制针对先验均值的可变性。如果固定 $\beta=(1/k, \dots, 1/k, 0, 0 \dots)$, 本文 K 个条目的值为 $1/k$, 而其余为 0; 当且仅当 $k' \in \{1, \dots, K\}$ 时, 达到状态 k' 的转换概率为非零。

3 系统仿真实验

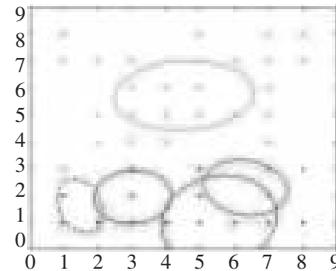
3.1 未知行为模式的定性

(1) 设初始行为状态为 4 个, 然后进行抽样获取训练 HDP-iHMM 模型的样本, 对模型进行训练, 同时对样本进行聚类, 可得到如图 3(a) 的聚类图, 模型状态转移矩阵如图 4(a) 所示, 模型观察值转移矩阵如图 4(d)。

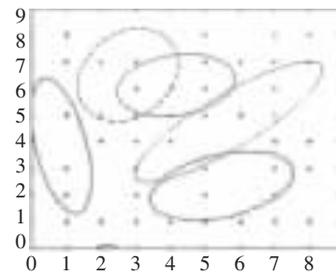
(2) 获取一个检测样本, 通过已经训练好的模型来验证模型的有效性。将包含 5 个状态的样本进行检测, 会发现有一种新的行为, 系统提示需要进行知识库更新。将这种新产生的行为进行定义后添加到知识库中, 并用这个样本作为训练样本来训练一个与此行为相对应的模型, 将训练结果保存到知识库中, 同时对样本进行聚类, 可得到如图 3(b) 所示的聚类图, 模型状态转移矩阵如图 4(b), 模型观察值转移矩阵如图 4(e)。由图可看出, 行为状态增加了 1 个, 由 4 个变成 5 个。最后, 用包含状态 5 和状态 6 的样本进行检测, 同样, 系统就会有新的行为的提示信息。重复上述过程, 可得到如图 3(c) 所示的聚类图, 模型状态转移矩阵如图 4(c) 所示, 模型观察值转移矩阵如图 4(f) 所示。由图 4 可看出, 状态又



(a) 初始样本



(b) 5 个状态样本



(c) 6 个状态样本

图 3 HDP 聚类图

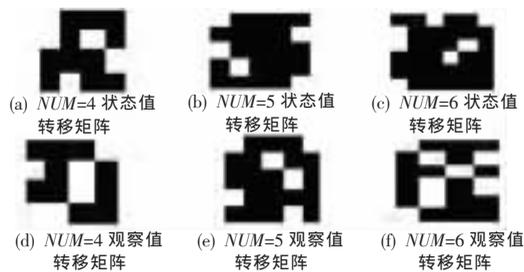


图4 训练过程中状态和观察值转移矩阵

增加了1个,由5个变成6个。

通过聚类图可看出,能将代表新的人体行为特征向量聚为一类,图3中的虚线椭圆表示一个新的聚类,从最初的4类到新增一类后的5类和新增两类后的6类。

以上描述了HDP-iHMM在识别未知行为方面的有效性,通过iHMM可以对未知行为进行确定和描述,为行为检测和预测做好准备。在本仿真统中,本文用HDP-iHMM确定未知行为,在事件数目确定后,用HMM实现行为的识别。通过iHMM和HMM的结合,增加了行为识别的主动性和智能性。

3.2 行为识别

3.2.1 前景获取

背景边缘模型是记录背景模型的边缘像素位置信息,通过背景边缘图与当前视频帧的边缘图像在相同位置像素的比较来判断该位置的像素点是否为前景目标像素点。通过与背景检测的比较,验证本文方法的优点,这种前景帧判断方法相对于其他常见的前景判断方法不但简单而且鲁棒性好,图5给出了两种方法的比较实例。从图中可以看出,由于受光照和阴影等因素的影响,背景法因为光照的突变使得检测结果为整个画面,不太理想。而本文的方法则受光线的变化影响较小。由此可见,对于前景检测,本方法能够很好地避免光照和人体阴影的影响,能够较好地检测前景目标。



图5 背景法检测和本文方法检测结果比较

3.2.2 模型的学习和行为识别

对系统行为识别能力的实验以视频监控中人体行为的识别为例。首先通过iHMM对未知行为模式进行定性和描述。表1为系统已经能识别的行为描述,让人体

特征数据通过本文的识别系统,识别系统将会返回一个行为标识,通过行为标识索取行为描述。

表1 行为描述

行为标识	行为描述	行为标识	行为描述
0	不能识别	4	侧身站立
1	向后走动	5	侧身走动
2	向前走动	6	弯腰站立或走动
3	站立	7	蹲下

对于本文仿真实验,采用的训练样本就是获取的人体特征向量。以“站立”、“侧身走动”和“蹲下”所对应的特征向量为训练样本,分别用S1、S2、S3表示来说明模型的训练过程。训练的收敛误差用联合相关性的稳定性来衡量,收敛误差根据收敛的精确度而定。训练过程中随着迭代次数的增加,最大似然估计值的对数值也在不断地增加,直到达到收敛误差为止。由于训练样本的差异,联合相关性稳定在不同的迭代次数之后。但从图6中可以看出,每个训练样本都达到了收敛。

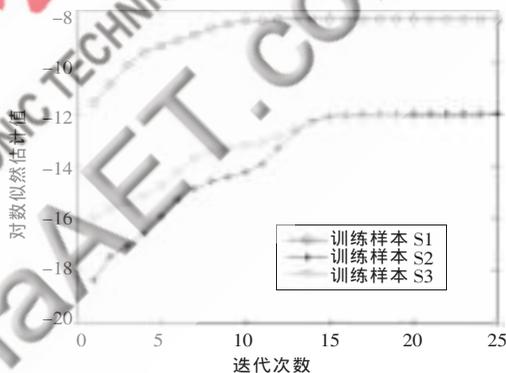


图6 模型训练曲线

模型对视频监控中人体行为的识别能力,HMM通过搜索最佳状态序列,以最大后验概率为准则来找到识别结果。在本系统中采用25帧/s的视频输入,来分析视频序列中人体的行为,同时验证本文识别系统的识别准确率。图7为对人体“站立”行为的识别,图8为对人体“蹲下”行为的识别。从图中可以看出,采用的基于统计学的行为识别模型能够很好地识别不同时刻人的同一行为。

在整个小男孩的行为识别过程中,由于存在许多相关因素的影响,会出现识别错误的情况。本文在整个跟踪过程中统计了跟识别错误率,其结果如图9所示。从图9可以看出,随着跟踪处理帧数的增加,跟踪错误率总围绕某一值上下波动,本文统计跟踪错误率大约是18%。

从时间复杂度方面考虑,整个系统包括两个部分:(1)离线的未知行为确定和行为模式学习系统;(2)在线的行为识别系统。对于在线系统,其行为识别算法采用应用比较广泛的Viterbi算法。因为利用全概率公式虽然可以计算系统的输出概率,但无法找到一条最佳的状态转移路径。而Viterbi算法,不仅可找到一条足够好的转移路径,而且可得到该路径对应的输出概率。同时,



图7 对不同时间下“站立”的识别



图8 对不同时间下“蹲下”的识别

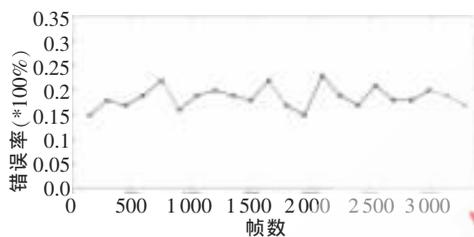


图9 系统识别错误率变化曲线

Viterbi 算法计算输出概率所需要的计算量要比全概率公式的计算量小很多。这些可以说明本文的行为识别系统实时性较好,识别算法时间复杂度小。对于离线系统,仿真试验已经验证了对未知行为的确定能力和行为模式的学习能力,而且离线系统对实时性要求较低。

本文的重点是对视频流中人体行为识别的研究,这是计算机视觉中一个重要的研究领域之一。仿真实验演示了视频监控中人体行为识别的全过程,提出了用背景边缘模型来提取前景图像,从仿真实验可看出此方法有

较好的提取效果,而且能够有效避免光照和阴影等外部因素的影响。此外,在行为识别方面,应用 NDP_iHMM 来确定行为状态数,在状态数确定以后将无限 iHMM 变成有限 HMM,这样提高了系统的普适性,通过 iHMM 与 HMM 结合,解决了在系统行为状态可变情况下的人体行为识别问题。

参考文献

- [1] 王亮,胡卫明,谭铁牛.人运动的视觉分析综述[J].计算机学报,2005,25(3).
- [2] CUI Y, WENG J. Hand segmentation using learning-based prediction and verification for hand sign recognition [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997: 88-93.
- [3] POLANA R, NELSON R. Low level recognition of human motion [C]. Proceedings of IEEE Workshop on Motion of Non-Rigid and Articulated Objects, Austin, TX, 1994: 77-82.
- [4] BOBICK A, DAVIS J. Real-time recognition of activity using temporal templates[C]. Proceedings of IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, 1996: 39-42.
- [5] DAVIS J, BOBICK A. The representation and recognition of action using temporal templates [R]. MIT Media Lab, Perceptual Computing Group, Technical report: 1997: 402.
- [6] XIANG Tao, GONG Shao Gang. Video behavior profiling for anomaly detection[J]. IEEE Transactions On Pattern Analysis and Machine Intelligence, 2008, 30(5):893-908.

(收稿日期:2010-07-01)

作者简介:

温向兵,男,1982 生,硕士,主要研究方向:智能信息处理。