

文章编号:1672-3961(2011)06-0024-07

面向监控视频的行为模式挖掘

梁浩哲,徐树奎,李国辉,张军

(国防科技大学信息系统与管理学院,湖南长沙410073)

摘要:行为模式挖掘技术是监控视频语义分析的重要组成部分,由于先验知识的缺乏与特征维数约束,难以准确定义参数化挖掘模型的结构复杂度,通过非参数化的无限高斯混合聚类运动特征得到原子行为模式,并估计其持续时间分布,使用局部特征维测试验证了挖掘模型的运动相似性假设。结果表明所得到的行为模式集准确刻画了场景的潜在运动语义,而通过行为中存在的时间多形态分布进一步发现了隐藏运动知识。

关键词:视觉监控;场景分析;无限高斯混合模型;行为模式挖掘

中图分类号:TP391 **文献标志码:**A

An activity mining model for surveillance video

LIANG Hao-zhe, XU Shu-kui, LI Guo-hui, ZHANG Jun

(School of Information System and Management, National University of Defense Technology, Changsha 410073, China)

Abstract: The activity pattern mining technique is the key component of semantic analysis for surveillance video. Because of the lack of prior and high-dimensional feature constraints, the complexity of the model structure of the parametric mining model is difficult to be precisely defined. Non-parametric clustering of motion feature by infinite Gaussian mixture was used to get the elementary activity patterns, based on which duration distribution was estimated. The partial-dimension test for feature validated the motion similarity hypothesis existing in the mining model. The results showed that the obtained activity patterns precisely reflected motion semantics of the scene, and that the multi-modality temporal distribution existing in activity can be further used to discover the hidden knowledge of motion.

Key words: visual surveillance; scene analysis; infinite Gaussian mixture model; activity pattern mining

0 引言

视频监控场景中运动对象的行为模式是一种重要的高层语义信息,在安全、智能监控等方面有广阔的应用前景^[1-4]。XIANG Tao 等人利用 HMM 分析场景中的行为结构^[5];WANG Xiaogang 等人基于主题模型挖掘行为的潜在并发模式^[6-7];R NATAVIA 与 G LAVEE 等人从知识建模的角度出发推理高层事件^[8-11]。统计挖掘描述底层视觉信息中的不确定性,具有强鲁棒性,但目前的方法大多需要手动设置模型结构。本研究使用无限高斯混合模型(infinite Gaussian mixture model, I-GMM)估计运动轮廓集

的潜在分布,实现非监督式的行为模式聚类,通过行为的时间建模进一步发现场景中存在的隐藏知识。

1 监控场景行为挖掘模型

行为模式挖掘针对监控帧序列 $F = \{f_1, f_2, \dots, f_n\}$,研究监控场景对象的运动特性,使监控设备具备场景高层语义的自动理解能力。模型分为3个步骤:

(1) 通过分析监控场景中的视觉运动特点,形成特征矢量集 M ,包括运动轮廓的提取及运动属性统计。

(2) 通过 I-GMM 估计特征集 M 的潜在分布,聚类得到场景中的原子行为模式(elementary activity

收稿日期:2011-06-26

基金项目:国家自然科学基金资助项目(60902093)

作者简介:梁浩哲(1983-),男,山东济南人,博士研究生,主要研究方向为多媒体信息处理. E-mail: Lianghaozhe@hotmail.com

pattern, EAP)。

(3) 建模 EAP 的时间关系。定义行为模式序列并通过统计分析进一步推理隐藏知识。

1.1 运动特征空间

在多数视频监控场景中,由于抖动、光强变化等因素的影响,运动对象的细节难以得到。利用时空运动模板可以有效地表示对象行为,例如运动历史图^[12]、像素历史图^[13]。

本模型使用运动轮廓矢量^[14]作为 EAP 的构成基元。针对每帧图像 f_i , 有映射 $S: (x, y) \rightarrow o$, 其中, (x, y) 为像素, o 为运动轮廓变量, 代表帧差图像的连通性区域, S 为帧序列的分割算子。定义映射 $F: o \rightarrow m$, 其中 m 为运动矢量, 见图 1; m 用 o 的 6 维特征描述: $(x, y, dx, dy, w, h)^T$, 其中, (x, y) : 空间像素, 代表位置坐标; (dx, dy) : 运动速度, 表示速度分量, 反映运动方向; (w, h) : 尺寸信息, 通过跟踪矩形框的大小描述(见图 1)。

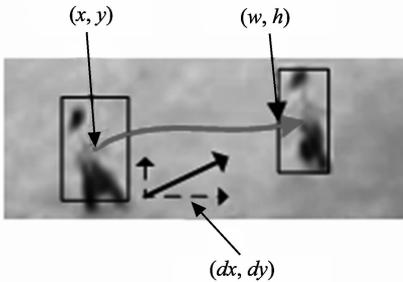


图 1 轨迹点 6 维矢量

Fig. 1 6-dimensional contour feature of trajectory

监控场景上下文约束对象行为,例如人通过斑马线过马路,运动对象需要右行进等交通规则,因此运动矢量具有潜在规律。如图 2 所示,对象 A、B、C 在移动的过程中,基本保持匀速运动,并具备相似的方向,矩形框尺寸发生微小变化。3 对象的运动反映了场景中的基本运动规则:右行,沿道路移动。模型假设运动行为的相似性关系构成原子行为模式 EAP,基于运动矢量样本集 $M = \{m_1, m_2, \dots, m_n\}$ 挖掘其潜在知识。

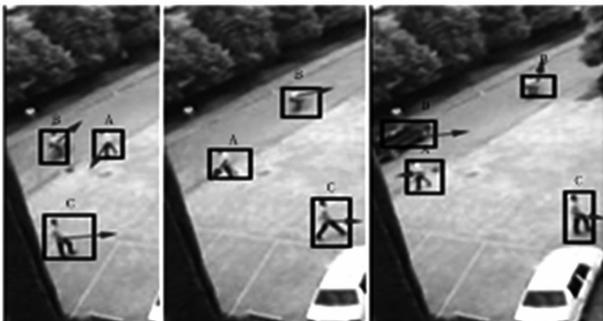


图 2 视频监控中的运动相似性

Fig. 2 Visual motion similarity of surveillance

高斯混合模型可有效近似拟合矢量集 M 的概率分布。对随机矢量 $m_i = (x_i, y_i, dx_i, dy_i, w_i, h_i) \in M$, 其概率密度函数 p 为混合形式^[5]:

$$p(m_i | \mu, \sigma, \pi) = \sum_{k=1}^n \pi_k N(\mu_k, \sigma_k), \quad (1)$$

其中, (μ, σ, π) 分别表示均值矢量阵, 协方差阵, 混合系数, $N(\mu_k, \sigma_k)$ 高斯分布函数。

在式(1)中, p 函数参数估计需要预定义“正确”的成分参数 n , 以确定模型结构^[5]。在文献[5]中, 使用 BIC 进行模型选择, 随着样本量的增大, 可以渐进收敛到“真实”模型结构, 但其在小样本量较少时会存在数据欠拟合的问题。因此本研究采用数据驱动式的 I-GMM 模型对 p 进行估计, 利用随机采样方法推理场景中的潜在 EAP 集。在 1.2 节中, 介绍了 I-GMM 的数值解法, 并通过二维仿真数据验证了其有效性。

1.2 基于 I-GMM 的运动矢量聚类

I-GMM 基于 Gibbs 采样, 迭代采样数据的类别索引, 蒙特卡洛马尔科夫模型可验证随迭代次数增加, 索引样本服从其真实分布, 是一种优良的非参数贝叶斯估计模型。I-GMM 扩展了传统的 GMM 模型, 假设先验关系 $\mu \sim \text{Gaussian}$, $\sigma \sim \text{Wishart}$, $\pi \sim \text{Dirichlet}$, 使用“中国餐馆过程(china restaurant process)”进行采样^[15-16]。

首先假设 GMM 中成分数量 n 有限时, 有样本类别的先验分布表示为

$$p(c_i = j | c_{-i}, \alpha) = \frac{k_{-i,j} + \alpha/n}{k - 1 + \alpha}, \quad (2)$$

其中, i 为样本号; j 为类别索引值; α 为 Dirichlet 分布参数; ‘-’ 指示除当前样本外的其他样本点, 例如 $k_{-i,j}$ 表示为除样本 i 之外的属于 j 类的样本量。当 $n \rightarrow \infty$ 时, 则有 Gibbs 先验迭代采样分布:

$$\begin{aligned} \text{if: } k_{-i,j} > 0: & p(c_i = j | c_{-i}, \alpha) \propto \frac{k_{-i,j}}{k - 1 + \alpha}, \\ \text{else: } & p(c_i = j | c_{-i} \neq j) \propto \frac{\alpha}{n - 1 + \alpha}. \end{aligned} \quad (3)$$

式(3)分别表示已有成分与出现新成分的混合采样分布。在 I-GMM 中, 使用共轭先验, 积分掉成分参数, 避免了成分 n 的预定义。结合数据似然 $p(x_i | c^{-i}, x^{-i})$, 使用 Gibbs 采样器可以对数据索引集 c 进行重复采样。由于假设 $n \rightarrow \infty$ 进行采样, 模型具有足够的复杂度, 可以避免数据欠拟合; 采样参数仅为索引集合 $c = \{c_1, c_2, \dots, c_i\}$, 不会随分量 n 而增加, 可避免数据过拟合问题。可使用 I-GMM 中“活性”成分的均值与方差刻画数据分布, 例如聚类质心 μ_j 的“截断”估计式为

$$\mu_j = \frac{\sum_{i=1}^N \mathbf{x}_i \delta(c_i = j)}{N_j}, \quad (4)$$

其中, N 为样本总量; N_j 为 j 类样本点的数目; δ 为 Delta 函数; \mathbf{x}_i 为样本值。

首先使用仿真二维数据演示 I-GMM 的聚类效果, 验证其有效性并与传统的模型选择方法贝叶斯准则 (Bayesian information criteria, BIC)、Aike 信息准则 (Aike information criteria, AIC) [17-18] 的效果进行了比较。数据为 5 个高斯成分产生的 300 个空间点; 同时, 在两相邻高斯均值间产生 20 个均匀分布

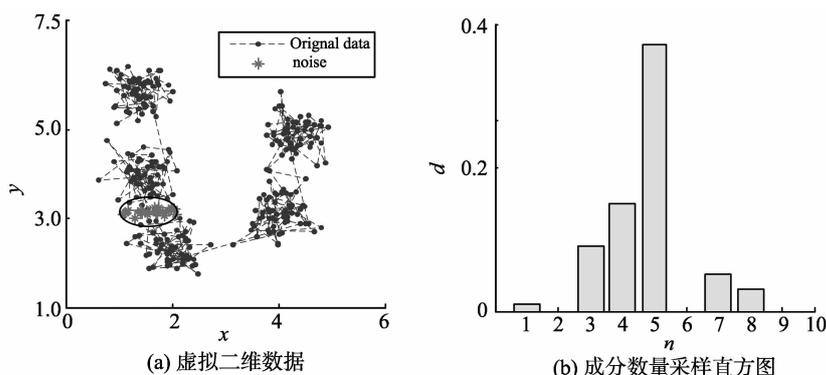


图3 I-GMM 聚类测试
Fig. 3 Test of I-GMM clustering

1.3 EAP 时态序列分析

如 1.2 所述, 通过 I-GMM 聚类可形成原子行为模式集 $EAPS = \{EAP_1, EAP_2, \dots, EAP_n\}$, n 为“活性”高斯分布数目 (存在运动样本支持)。

对轨迹集 $T = \{T_1, T_2, \dots, T_i\}$, 存在轨迹到 EAPS 集的映射:

$$\text{MAP: } T_i = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_l, y_l, t_l)\} \rightarrow A_i = \{A_1, A_2, \dots, A_l\}.$$

其中, l 为轨迹帧长, $A_i \in EAPS, T_i \in T$ 。MAP 为最大后验概率分类。可降维 T_i , 将其映射到语义行为空间 EAPS 中处理 [19-20]。

定义 1.3 时序行为模式序列 (temporal elementary activity pattern sequence, TEAPS)。对于 $\forall (x, y, t) \in T_i$, 有最大后验行为模式分类 MAP: $(x, y, t) \rightarrow A_i$, 则 TEAPS 实例表示为

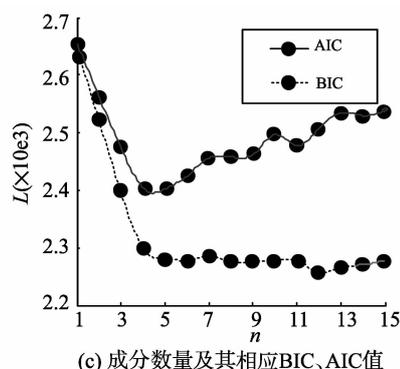
$$\text{TEAPS}_i = A_1 \xrightarrow{t_{A_1}} A_2 \xrightarrow{t_{A_2}} \dots \xrightarrow{t_{A_{n-1}}} A_n$$

其中, t_{A_i} 为原子行为 A_i 的持续帧数, 通过行为持续帧数表示, A_i 为 t_{A_i} 的截断 Gibbs 索引估计。

时序行为模型表示了行为模式的时间结构, 通过 t_{A_i} 的统计分布可以建模相似行为中不同的持续时间偏差。图 4 为一条简单的 TEAPS 表述。

使用 A_i 的平均持续时间可以估计场景中运动对象的停留时间, 模型使用 GMM 建模 EAPS 的时

噪点 (outlier), 见图 3。图 3(a) 中标注处, 两成分存在小范围上的重叠。图 3(b) 为 Gibbs 的递归采样直方图, 最大采样数值为 5, 其密度值 d 为 0.37。聚类结果如图 3(c) 所示, 在 AIC 与 BIC 值曲线中: 由于其对参数增长的惩罚性不足与成分重叠的影响, AIC 出现明显的过拟合问题, 即偏好于复杂模型; BIC 值 L 在 4 处取得最小值, 由于小样本量的影响存在数据欠拟合。可以看出 I-GMM 可以有效的避免模型选择问题, 正确估计类别索引, 形成“活性”成分聚类。



序分布。

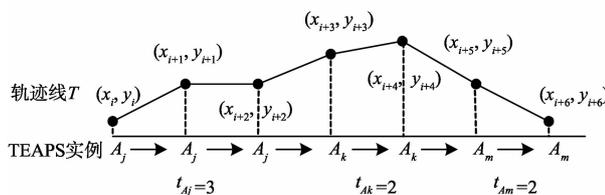


图4 TEAPS 模型
Fig. 4 TEAPS model

2 实验结果及分析

在满足一般监控条件下, 实验数据使用近远景视频序列, 运动对象的大小受视角的影响较小, 可忽略此因素。

特征提取使用快速收敛 FGMM [21] 对场景背景进行建模, 以此为基础, 使用 Kalman 滤波器跟踪运动对象, 帧速为 10 帧/s。如图 5 所示, 通过跟踪检测, 随运动对象演化, 提取运动轮廓, 图 5 中的空间区线形成运动对象的轨迹集 T 。

首先观察运动分量的空间分布, 见图 6。图 6(a), 选择 31 条场景轨迹作为训练集。图 6(b) 描述 M 的空间分布, 由于场景中的规则约束, 场景中的“规则轨道”上的方向具备相似性, 从直观上看存在 3 个行人运动方向。图 6(c) 为采样过程中分量 n

的直方图,可以观察到,随迭代次数 i 增加,Gibbs 采样分布与分量 n 的初值相关性迅速降低,且密度 d 在 $n=3$ 处达到最大值 0.69。图 6(d) 显示了聚类数目随递归采样次数的剧烈衰减,并收敛到聚类正确目标数量 3。图 6(e) 为方向样本 (dx, dy) 3 个主方向,图 6(f) 为其 BIC 曲线,聚类数量与 Gibbs 抽样相同。

使用相同的方法,讨论运动对象的 (w, h) 特征分量,见图 7。图 7(a) 为“规则轨道”中的运动对象尺寸样本;图 7(b) 为左图直观的表现一个高斯分布,即此“轨道”上仅仅有单类运动对象“行人”。类似的,图 7(a) 左图与图 7(a) 右图分别代表了“轨道”中的车辆对象与人对象的 (w, h) 分量;图 7(b) 右图表现为不同的高斯分布,运动特征在不同运动对象之间具有较强的分离性。通过局部特征维测试,

可验证模型的运动相似性假设成立。



图 5 轨迹提取
Fig. 5 Trajectory training set

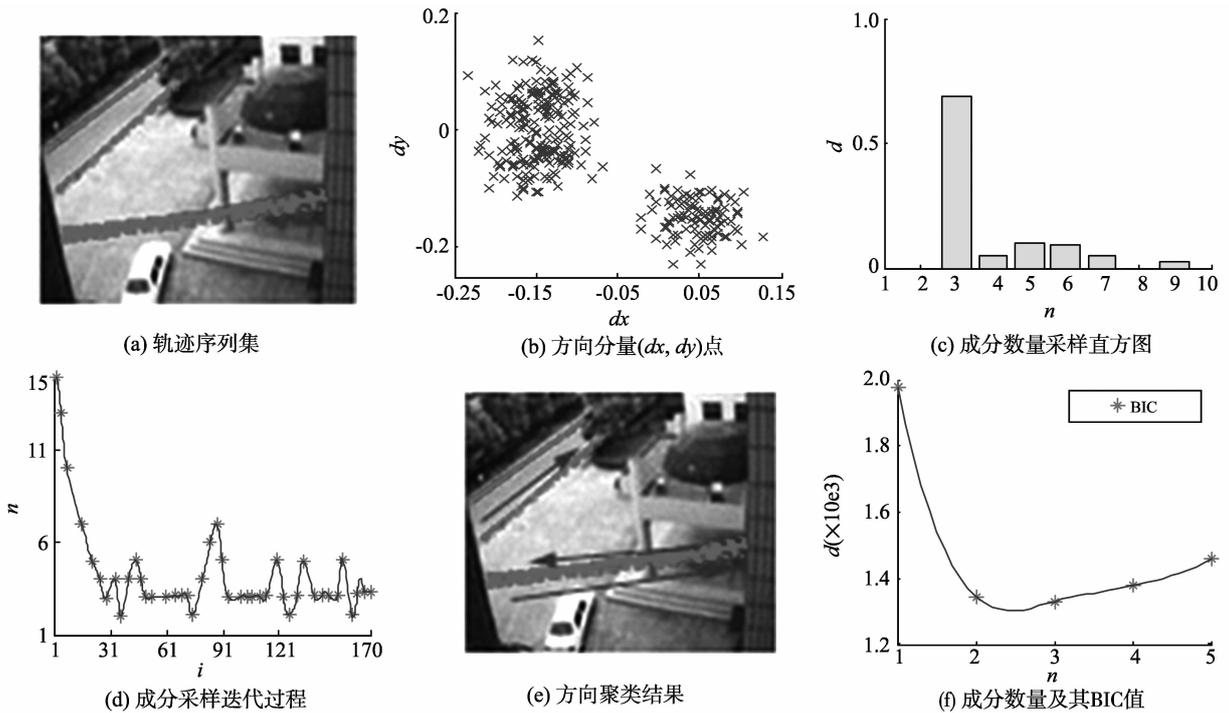


图 6 方向分量聚类

Fig. 6 Clustering result for direction component

针对轨迹样本 T ,在其基础上筛选了大量的平稳型轨迹,数量为 2 000,见图 8。图 8(a) 所示为部分轨迹,在演化过程中不会发生过度衰减并允许存在跟踪失效产生的间断轨迹。在监控场景中,此类轨迹训练集可以反应正常的场景运动规则。

在平稳轨迹集上使用 I-GMM&Gibbs 采样器,可观察到 14 种 EAP,反映了监控场景中的基本运动规则。图 8(b)、图 8(c) 分别代表了采样过程中的聚类直方图与迭代抽样过程,可以看出,GMM 聚类拟合为类球体形状,对运动对象的空间聚类结果近似为圆形。在图 8(d) 中,将各行行为模式映射到二维

平面进行可视化显示。表 1 为各 EAP 的语义信息短描述。

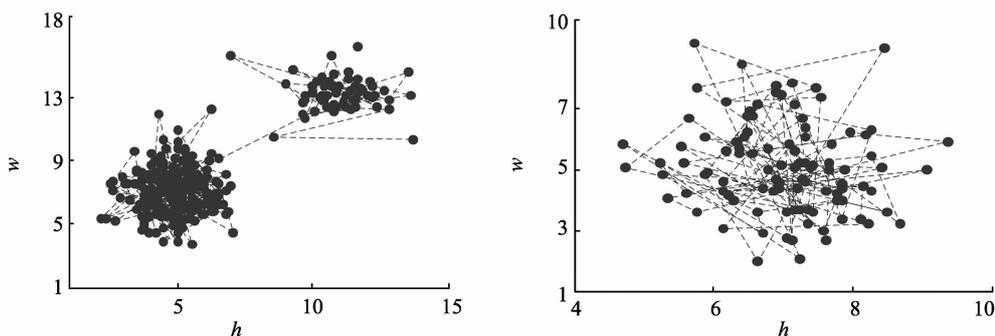
表 1 EAP 语义描述及时间统计

Table 1 Semantic description and temporal statistic of EAP

EAP: A_i	语义描述	时间均值/帧
(A_1, A_2)	行人沿左下路径进入大楼	(60,40)
(A_3, A_4)	行人沿左下路径离开大楼	(38,62)
(A_5, A_6)	行人由左上角进出大楼	(40,42)
(A_7, A_8)	行人在公路上行	(52,32)
(A_9, A_{10})	行人在公路下行	(35,55)
(A_{11}, A_{12})	车辆上行	(25,22)
(A_{13}, A_{14})	车辆下行	(23,26)

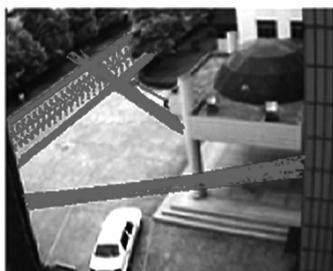


(a) 运动对象尺寸样本

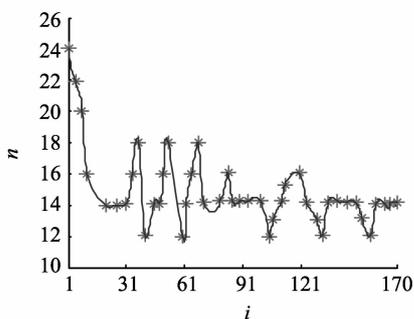


(b) 运动对象尺寸分布

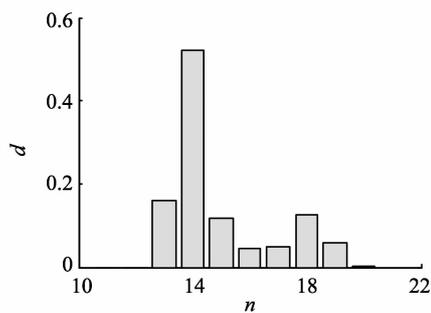
图7 尺寸分量空间分布
Fig.7 Clustering result of component size



(a) 轨迹样本



(b) 成分数量采样迭代过程



(c) 成分数量采样直方图



(d) 运动聚类结果

图8 运动轮廓聚类结果
Fig.8 Clustering result based on motion contour

如表1的语义描述,EAP的序列组合可以很方便的描述整个场景中规则的视觉运动。根据2.4,使用最大后验概率分类轨迹序列点,可以得到场景中的基本行为转移序列,例如 $A_1 \rightarrow A_2 \& A_2 \rightarrow A_1$ 表示进出主建筑物, $A_5 \rightarrow A_6 \& A_6 \rightarrow A_5$ 代表车辆在公路上的平稳行驶;但不存在 $A_5 \rightarrow A_{10}$ 的行为序列,因为行

人在离开大楼并向左下角行进只存在 $A_2 \rightarrow A_1$ 的行为序列,此序列不满足在此监控场景下行人的合理运动行为。基于I-GMM行为模式挖掘有效地发现了场景中的运动语义内容。

使用定义1.3完成轨迹样本向EAP空间转换,使用GMM对各个EAP的持续时间进行统计,图9

左图行为 A_1 的持续时间直方图与高斯拟合曲线,明显的呈现单峰形态。受篇幅所限, A_3 、 A_4 、 A_5 、 A_6 、 A_7 、 A_{10} 、 A_{11} 、 A_{12} 、 A_{13} 、 A_{14} 持续时间与之相似,其均值时间如表 1 所示。图 9 右图为 A_8 、 A_9 的时间统计直方图,对应图 8(d) (左 1,左 2) 中右上角行为模式,两种行为不呈现稳定的单峰性,近似于三峰,存在 3 种不同的持续时间,由于 I-GMM 的近似球形分布与多维度聚类下的特征冗余问题,此两类行为中存在不同的视觉运动形式。

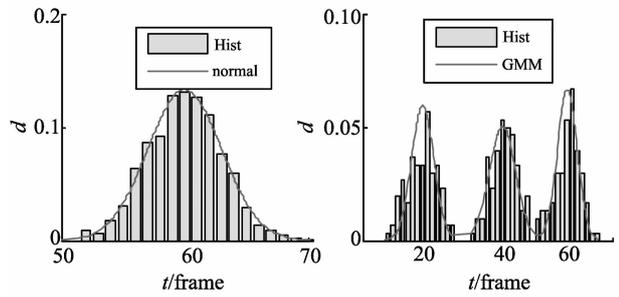


图 9 持续时间分布

Fig. 9 Distribution of EAP duration

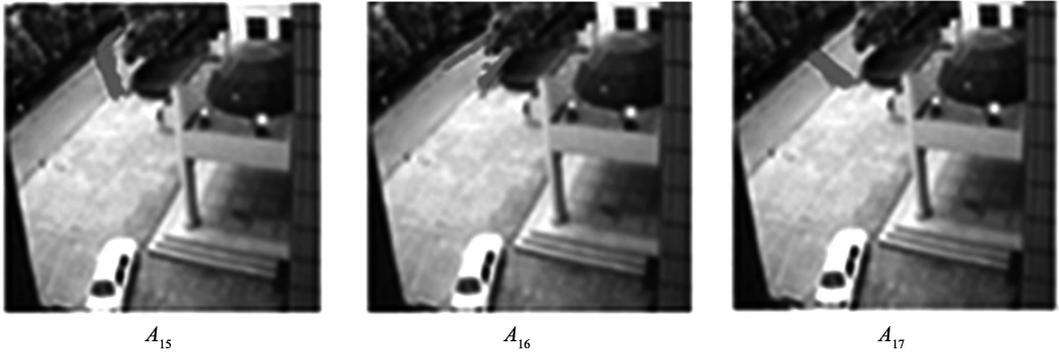


图 10 基于持续时间的异类运动模式

Fig. 10 Different class of EAP based on duration

模型使用 3 成分的 GMM 模型拟合 A_8 、 A_9 的持续时间分布,发现了 3 种不同的行为模式,如图 10: A_{16} , 右上运动或右下运动; A_{17} , 左上运动或左下运动; A_{15} , 左上转右上运动或右上转左下运动,同时,根据此行为序列的起点与终点,加深标注为场景中的隐藏出入口,而在直观的监控场景显示中很可能被监控用户忽略。

表 2 3 种行为的时间样本均值与方差
Table 2 Duration mean and variance of A_{15} , A_{16} and A_{17}

EAP: A_i	时间均值/帧	时间方差/帧
A_{15}	51	5.24
A_{16}	22	6.96
A_{17}	38	6.74

3 结论

本研究使用 I-GMM 拓展了文献 [13] 中的方法,实现了一种非监督式运动规则挖掘模型,通过运动特征聚类与行为时序建模解决了面向行为的监控场景语义知识挖掘问题,最后通过数值实验验证了模型的有效性并对实验结果进行了直观的展示。在实验过程中发现离线学习过程仍然存在较高的计算开销,因此面向监控轨迹的在线行为模式挖掘将是未来的研究重点。

参考文献:

[1] HU Weiming, TAN Tieniu, MAYBANK Steve. A survey on visual surveillance of object motion and behaviors[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2004, 34(3): 334-352.

[2] 焦礼成, 公茂果, 王爽, 等. 自然计算、机器学习与图像理解前沿[M]. 西安: 电子科技大学出版社, 2008: 143-147.

[3] TURAGA Pavan, CHELLAPPA Rama, SUBRAHMANNIAN V S, et al. Machine recognition of human activities: a survey[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(11): 1473-1488.

[4] MAKRIS Dimitrios, ELLIS Tim. Path detection in video surveillance[J]. Image and Vision Computing, 2002, 20(12): 895-903.

[5] XIANG Tao, GONG Shaogang. Beyond tracking: modeling activity and understanding behavior[J]. International Journal of Computer Vision, 2006, 67(1): 21-51.

[6] WANG Xiaogang, MA Keng Teck, NG Gee-Wah, et al. Trajectory analysis and semantic region modeling using a nonparametric bayesian model[C]//Proceedings of IEEE Conference On Computer Vision And Patter Recognition. Anchorage: IEEE Computer Society, 2008: 1-8.

[7] WANG Xiaogang, MA Xiaoxu, GRIMSON Eric. Unsupervised activity perception in crowded and complicated

- scenes using hierarchical bayesian models [J]. IEEE Transactions on PAMI, 2009, 31(3): 539-555.
- [8] LAVEE G, BORZIN A, RIVLIN E, et al. Building petri nets from video event ontology [C]//Proceedings of 3rd International Conference on Advances in Visual Computing. Nevada: Springer-Verlag, 2007:442-451.
- [9] BORZIN A, RIVLIN E, RUDZSKY M. Surveillance interpretation using generalized stochastic petrinets [C]//8th International Workshop on Image Analysis for Multimedia Interactive Services. Santorini: IEEE Computer Society, 2007:102-118.
- [10] LAVEE G, RUDZSKY M, RIVLIN E, et al. Video event modeling and recognition in generalized stochastic Petri nets [J]. IEEE Transaction on Circuits and Systems for Video Technology, 2010, 20(1): 102-118.
- [11] LAVEE G, RIVLIN E, RUDZSKY M. Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2009, 39(5):489-504.
- [12] DAVIS James W, BOBICK Aaron F. The representation and recognition of action using temporal templates [C]//IEEE Conference on Computer Vision and Pattern Recognition. San Juan: IEEE Computer Society, 1997:928-934.
- [13] XIANG Tao, GONG Shaogang. Incremental and adaptive abnormal behavior detection [J]. Computer Vision and Image Understanding, 2008, 111(1):59-73.
- [14] SIMON C. Using decision trees to recognize visual events [C]//Proceeding of the 1st ACM Workshop on Analysis and Retrieval of Events/Actions and Workflows in Video Streams. Vancouver: ACM New York, 2008: 41-48.
- [15] TEH Yee Whye, JORDAN Michael I, BEAL Matthew J, et al. Hierarchical dirichlet processes [J]. Journal of the American Statistical Association, 2006, 101(476): 1566-1581.
- [16] RASMUSSEN C. The infinite gaussian mixture model [J]. Advances in Neural Information Processing Systems, 2000, 12: 554-560.
- [17] GHEISSARI Niloofar, HADIASHAR Alireza Bab. Model selection criteria in computer vision: are they different? [C]//Proceedings of Digital Image Computing Techniques and Applications. Australia: CSIRO, 2003: 185-194.
- [18] ZUCCHINI Walter. An introduction to model selection [J]. Journal of Mathematical Psychology, 2000, 44:41-61.
- [19] KANG J. Mining trajectory patterns by incorporating temporal properties [C]//Proceedings of the 1st International Conference on Emerging Databases. Busan: KIISE, 2009:63-68.
- [20] KANG J, YONG H. Spatial-temporal discretization for sequential pattern mining [C]//Proceedings of International Conference Ubiquitous Information Management and Communication. Suwon: ACM New York, 2008: 218-224.
- [21] 焦波. 面向智能视频监控的运动目标检测与跟踪方法研究 [D]. 湖南:国防科学技术大学信息系统与管理学院, 2009.
- JIAO Bo. Research on moving object detection and tracking methods for intelligent video surveillance [D]. Hunan: School of Information System and Management, National University of Defense Technology, 2009.

(编辑:陈燕)

(上接第23页)

- [16] 中国社科院语言研究所词典编辑室. 现代汉语词典(增补本) [M]. 北京:商务印书馆, 2002.
- [17] 德范克. ABC 汉英大词典 [M]. 上海:汉语大词典出版社, 2003.
- [18] 英国培生教育出版有限公司. 朗文当代高级英语辞典(英英、英汉双解新版) [M]. 北京:外语教学与研究出版社, 2004.
- [19] 崔宝今, 林鸿飞, 张霄. 基于半监督学习的蛋白质关系抽取研究 [J]. 山东大学学报:工学版, 2009, 39(3): 16-21.
- CUI Baojin, LIN Hongfei, ZHANG Xiao. Research of protein-protein interaction extraction based on semi-supervised learning [J]. Journal of Shandong University: Engineering Science, 2009, 39(3): 16-21.
- [20] 郭剑毅, 雷春雅, 余正涛, 等. 基于信息熵的半监督领域实体关系抽取研究 [J]. 山东大学学报:工学版, 2011, 41(4): 7-12.
- GUO Jianyi, LEI Chunya, YU Zhengtao, et al. A semi-supervised learning method based on information entropy to extract the domain entity relation [J]. Journal of Shandong University: Engineering Science, 2011, 41(4): 7-12.

(编辑:胡春霞)