

文章编号:1671-6833(2016)06-0072-05

基于稀疏超完备的异常行为检测算法

逯 鹏, 张利亚, 黄石磊, 李奇航, 张 微

(郑州大学 电气工程学院, 河南 郑州 450001)

摘 要: 视频异常行为检测的核心问题是如何从海量数据中理解行为. 提出一种新的基于视觉稀疏超完备表示机制的特定场景中视频异常行为检测算法, 针对视频中感兴趣的点提取局部有效信息, 提高数据处理效率. 首先, 提取训练样本的时空兴趣点, 计算局部时空特征; 其次, 将该特征输入稀疏超完备模型, 训练得到一组稀疏基; 然后, 利用上述基对待测视频进行重构, 通过局部时空特征重构的差异实现异常行为检测; 最后, 提出对不同视频的稀疏基更新算法. 标准数据库的实验表明, 该算法能够有效解决异常行为检测问题, 检测准确率高, 错误警报率低.

关键词: 异常行为检测; 稀疏超完备; 时空兴趣点

中图分类号: TP2 **文献标志码:** A **doi:**10.13705/j.issn.1671-6833.2016.03.031

0 引言

视频中异常行为检测是目前计算机视觉的研究热点. 常用的方法是模型学习, 根据学习形成的数学统计模型或已知的先验模型, 找到视频行为的最佳匹配模型^[1]. 例如, 隐马尔可夫模型(HMMs)、动态贝叶斯网络(DBNs)、条件随机场模型(CRF)等经典模型, 以及一些在经典模型基础上进行改进的方法, 如 STR 匹配方法^[2]、主题隐马尔可夫模型方法^[3]和融合声音与视频信息的隐随机场模型^[4]等.

然而, 提取高度详细特征信息则会导致计算复杂度呈指数增长^[5]. 因此, 如何在保证提取足够的有效特征信息的基础上减少计算量是视频异常行为检测要解决的关键问题. 事实上, 根据检测任务仅提取少量关键特征^[6], 能有效提高算法的效率. 而人类视觉系统的稀疏响应能够从包含大量冗余的场景中快速提取有效信息, 解决了高维、海量数据的实时处理问题, 这为目标检测、异常行为检测等问题提供了减少数据存储量和计算量的新思路^[7]. 文献[8]利用该思路建立了基于稀疏超完备表示的目标检测算法, 实现了非结构化二维图像场景中的目标检测. 在此基础上, 解决运动目标的异常行为检测问题则需要增加时间维度.

基于该思路, 针对特定场景的视频异常行为检测任务, 笔者以包含视频高层语义的时空兴趣点作为切入点, 提取视频的局部特征信息, 建立稀疏超完备计算模型, 设计感受野学习和异常行为检测等关键算法, 通过多类型视频异常行为检测, 验证模型和算法的有效性.

1 时空兴趣点的提取与描述

如何表示视频序列及行为并建立可计算数学模型, 是异常行为检测要解决的首要问题. 目标跟踪和模型匹配是行为表示的常用方法, 但都需要提取高度详细的轮廓信息^[9], 这很难直接通过包含噪声信号的真实视频实现. 时空兴趣点(spatial temporal interesting point, STIP)只提取视频的局部特征而不是全局特征作为数据处理基础, 能有效解决该问题.

时空兴趣点即视觉关注的焦点, 针对视频三维场景特点, Laptev^[10]将二维图像中的 Harris 角点检测算子扩展到三维空间, 提出了时空兴趣点的概念, 不仅包含人体动作姿态的空间信息, 还包含运动的动态时间信息, 开始了时空局部特征点在视频领域的研究. 图 1 为针对皇家理工学院行为数据库^[11](KTH 数据库)视频段的时空兴趣点提取结果. 可以发现, 打拳动作的 STIP 集中在手

收稿日期:2015-10-28; 修订日期:2016-01-17

基金项目:国家自然科学基金资助项目(60841004, 60971110, 61172152); 河南省青年骨干教师资助计划(2012GGJS-005)

作者简介:逯鹏(1974—), 男, 河南滑县人, 郑州大学副教授, 博士, 主要从事复杂视觉图像信息处理、脑机接口、复杂网络等研究, E-mail: lupeng@zzu.edu.cn.

臂上,而跑步动作的 STIP 则集中在腿、胳膊和头部这些运动的部位。

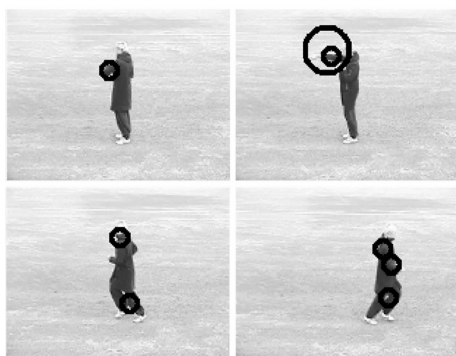


图1 时空特征点的提取结果

Fig.1 Results of extracting STIP

获得 STIP 后,结合周围时空信息构建 STIP 描述符,完成异常行为语义描述,形成特定行为特征.为同时表示行为的静态和动态特征,采用梯度信息(HOG)和光流信息(HOF)的组合特征作为描述符。

将 HOG/HOF 描述符推广到三维视频的动作识别中,具体方法如下:采用多尺度时空滑动窗,如图2,在每个 STIP 周围取一定大小的时空块,所取时空块为 $\Delta_t(\tau) = 8\tau, \Delta_x(\delta) = \Delta_y(\delta) = 18\delta$, 空间维度 $\delta = 1$, 时间维度 $\tau = 2.5$. 再将每个时空块划分为 $n_x \times n_y \times n_z$ 的网格,网格是更小的胞体积.对于每个胞体积,计算 4 维 HOG 和 5 维 HOF. 每个时空块中,胞体积个数是 18 ($3 * 3 \text{ pixels} * 2 \text{ frames}$),那么其 HOG 特征为 72 维,而 HOF 特征为 90 维,HOG/HOF 特征共 162 维。

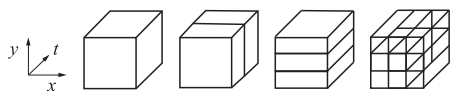


图2 时空滑动窗

Fig.2 Spatial temporal slipping window

据此,视频序列就可以用 HOG/HOF 特征表示为用 $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$, n 为检测出的 STIP 个数;而 $\mathbf{X}_i = \{\mathbf{X}_i^1, \dots, \mathbf{X}_i^{n_i}\}$ ($i = 1, 2, \dots, n$) 表示单个事件, $n_i = 18$ 为时空滑动窗内胞体积个数。

2 基于稀疏超完备的异常行为检测

特定场景下正常行为与异常行为的 HOG/HOF 特征存在差异,通过对特征数据的处理可以检测出视频中的异常行为.据此,视频异常行为检测可转化为超完备稀疏编码的问题。

视频异常行为检测整体基本思想是:首先,定

义特定环境下何种行为为正常行为(如机场的进口处定义行人进入为正常行为,行人出去则为异常行为),提取出无异常行为的视频序列 STIP,在每个 STIP 周围计算其特征描述符 \mathbf{X}_i ;其次,将 \mathbf{X}_i 输入稀疏超完备模型进行训练,生成包含预定义的正常行为的稀疏基 \mathbf{W} , \mathbf{W} 的列向量是用来重构信号的特征基;然后,提取待测视频的 STIP 和 HOG/HOF 局部特征 \mathbf{X}' ,计算 \mathbf{X}' 关于 \mathbf{W} 最优表示的重构误差,通过重构误差判断视频中的异常行为.需要指出的是,稀疏编码模型常见的输入是一个向量,而此处输入却是一个事件,包含时间和空间局部信息,是一簇向量。

2.1 稀疏超完备计算模型

视觉稀疏表示的图像可以逼近分解为:基函数集合分解系数的线性组合。

$$\mathbf{I} = \sum_{i=1}^m \mathbf{A}_i s_i + \mathbf{N}. \quad (1)$$

式中: \mathbf{I} 代表图像; \mathbf{A}_i 是视皮层 V1 区感受野特征基向量,为 n 维列向量; m 是基向量的个数; s_i 表示神经元对不同基图像的响应系数,对应主视皮层 V1 区简单细胞神经元的活动状态; \mathbf{N} 是高斯噪声; \mathbf{A}_i 与 s_i 均未知。

在视觉超完备机制下特征矩阵维数应大于图像数据的维数,即 $m > n$,此时 $\mathbf{A}_{m \times n}$ 为超完备集不可逆,使 \mathbf{I} 的稀疏分解是需要组合搜索的 NP-Hard 问题.为得到合理解,采用在能量模型基础上,最大化稀疏性来最优估计基函数 \mathbf{W} . 用非归一化对数似然函数来定义该能量模型:

$$\log L(\mathbf{x}; \mathbf{w}_1, \dots, \mathbf{w}_n) = -\log Z(\mathbf{W}) + \sum_{i=1}^n G_i(\mathbf{w}_i^T \mathbf{x}). \quad (2)$$

式中: \mathbf{x} 是单个样本数据; $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n]$,

$Z(\mathbf{W}) = \int \prod_{i=1}^n \exp(G_i(\mathbf{w}_i^T \mathbf{x})) d\mathbf{x}$ 是 \mathbf{w}_i 的归一化常数; $G_i(u) = -\alpha_i \log \cosh(u)$ 为 s_i 的对数概率密度,是 s_i 的稀疏性测度之和; \mathbf{w}_i 为待估计参数; α_i 也为未知数,随着 \mathbf{w}_i 同时进行估计。

对于 T 个样本数据 \mathbf{x}_t ($t = 1, 2, \dots, T$),则样本概率密度函数的对数似然函数:

$$\log L(\mathbf{w}_1, \dots, \mathbf{w}_n; \mathbf{x}_1, \dots, \mathbf{x}_T) =$$

$$T \log |\det(\mathbf{W})| + \sum_{i=1}^m \sum_{t=1}^T G_i(\mathbf{w}_i^T \mathbf{x}_t). \quad (3)$$

采用得分匹配直接估计 \mathbf{w}_k ,感受野 $\mathbf{w}_k = (w_{k1}, \dots, w_{kn})$ 约束为单位化形式.为简化 $Z(\mathbf{W})$ 的计算,取数据向量的对数密度函数 $\log p(\mathbf{x})$,由此定义得分函数为:

$$\psi(\mathbf{x}; \mathbf{W}, \alpha_1, \dots, \alpha_m) = \nabla_x \log p(\mathbf{x}; \mathbf{W}, \alpha_1, \dots, \alpha_m) = \sum_{k=1}^m \alpha_k \mathbf{w}_k g(\mathbf{w}_k^T \mathbf{x}). \quad (4)$$

g 是 G 的一阶导数, $g = G'$. 将数据得分函数和模型得分函数的平方距离作为目标函数:

$$J = \sum_{k=1}^m \alpha_k \frac{1}{T} \sum_{t=1}^T g'(\mathbf{w}_k^T \mathbf{x}(t)) + \frac{1}{2} \sum_{j,k=1}^m \alpha_j \alpha_k \mathbf{w}_j^T \mathbf{w}_k \frac{1}{T} \sum_{t=1}^T g(\mathbf{w}_k^T \mathbf{x}(t)) g(\mathbf{w}_j^T \mathbf{x}(t)). \quad (5)$$

通过最小化 $J(\mathbf{w})$, 求得最优参数 \mathbf{w} 估计值:

$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} J(\mathbf{w})$. 采用梯度下降学习算法极小化该目标函数, 就可以得到参数 \mathbf{w} 的一致性估计. 最速下降准则为:

$$\mathbf{w}(t) = \mathbf{w}(t-1) - \eta(t) \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} \Big|_{\mathbf{w}=\mathbf{w}(t-1)}. \quad (6)$$

式中: 学习速率 $\eta(t)$ 是负梯度方向步长, 由此得到感受野的超完备集表示. 正常行为的稀疏基 \mathbf{W} 是根据公式(5)最优化目标函数 $J(\mathbf{w})$ 的结果. 对于超完备稀疏编码模型 $\mathbf{S} = \mathbf{W}\mathbf{X}$, \mathbf{W} 学习算法如下所示.

输入: 样本视频特征数据集 \mathbf{X} ;

输出: 稀疏基 \mathbf{W} .

具体步骤:

(1) 去除 \mathbf{x}_i 的局部均值;

(2) 用 PCA 对 \mathbf{X} 白化、降维, 并投影到白化空间 $\mathbf{Z} = \mathbf{V}\mathbf{X}$;

(3) 初始化 \mathbf{W}_s 为单位阵; 开始迭代过程;

(4) 计算 $\mathbf{Y} = \mathbf{W}_s \times \mathbf{Z}$;

(5) 计算初始点梯度 $J(\mathbf{w}(0))$, 沿负梯度方向移动, 对 $t = 1, 2, \dots$, 按公式(6)最小化 \mathbf{W} , 并归一化为单位向量; 更新参数 α ;

(6) 如果 $\text{norm}(\Delta \mathbf{W}) < \varepsilon$ 则停止迭代; 否则, 返回步骤 4;

(7) 停止学习, 将 \mathbf{W}_s 反投影到原视频空间得 $\mathbf{W} = \mathbf{W}_s \mathbf{V}$;

(8) 算法结束.

2.2 异常行为检测算法

对于待测事件 \mathbf{X}' , 最小化目标函数 $J(\mathbf{w})$, 定义最优化结果 $J(\mathbf{X}', \mathbf{w})$ 为重构误差. 正常行为的特征能够通过 \mathbf{W} 进行表达, 重构误差比较小. 包含异常行为的视频特征通过 \mathbf{W} 进行表达的重构误差相对较大, 故可以通过重构误差判断视频中的异常行为. 如果重构误差大于阈值, 如式(7)所示, 则判断为异常行为; 否则为正常行为.

$$J(\mathbf{X}', \mathbf{w}, \alpha) > \xi. \quad (7)$$

阈值 ξ 的大小可根据检测异常行为的灵敏度需要设定. 正常行为视频通过稀疏基表达时也会产生较小的重构误差, 阈值 ξ 的设定越接近正常行为的重构误差, 检测的灵敏度越高; 阈值越偏离正常行为的重构误差, 检测的灵敏度越低. 算法流程如图 3 所示.

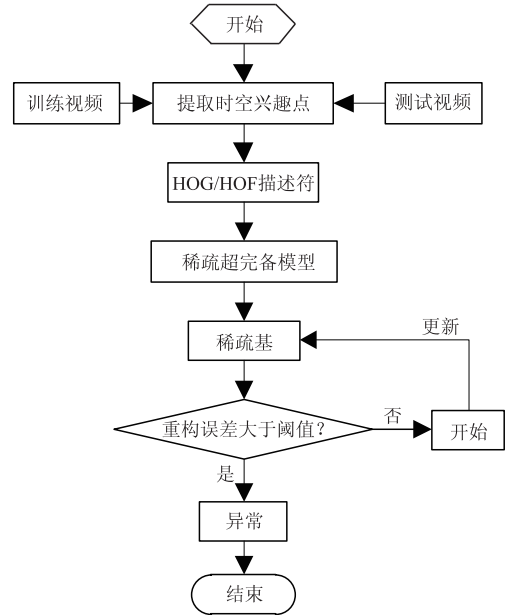


图 3 整体算法流程图

Fig. 3 The whole algorithm flow chart

2.3 稀疏基更新算法

由于视频环境因素的动态性, 最初训练获得的稀疏基, 在后续视频异常行为检测中就可能不是最优基. 因此, 采用稀疏基更新算法, 将已判断为正常的行为加入新的稀疏基. 对于已检测的视频中 τ 个时空块, 为节省数据的内存空间及优化计算速度, 稀疏基更新算法采用随机梯度下降法, 新稀疏基 \mathbf{W}_τ 由 $\mathbf{W}_{\tau-1}$ 和 \mathbf{X}_τ 计算得出, 如式(8):

$$\mathbf{W}_\tau = \prod_c [\mathbf{W}_{\tau-1} - \frac{\eta}{t} \nabla_{\mathbf{W}} l(\mathbf{X}_\tau, \mathbf{W}_{\tau-1})]. \quad (8)$$

式中: $l(\mathbf{X}_\tau, \mathbf{W}_{\tau-1}) = \frac{1}{2} \sum_{j=1, \dots, n_t} \|\mathbf{X}_\tau^j - \mathbf{W}_{\tau-1} \alpha_\tau^j\|_2^2$;

η 为学习率; \prod_c 为 C 正交投影.

3 实验结果及分析

目前对异常行为的检测, 多数定位于简单的肢体运动. 而在实际应用中通常需要结合特定场景判断行为异常. 为此本文结合一定场景设计实验. 表 1 为重构误差值和异常行为检测结果.

实验视频数据取自 P. M. Jodoin 和 Adam 数据库, 分别对 Canoe、Subway Exit 和 Subway Entrance 三段视频做出异常检测. 源代码开发工具

为 MatlabR2010b 和 STIP1.1. Matlab 工具用于形成稀疏基和计算重构误差,STIP1.1 用于提取时空兴趣点,并用高维特征表示.

表 1 重构误差值和异常行为检测结果
Tab.1 Reconstruction error and the result of abnormal behavior detection

训练视频段	待测视频段	重构误差	异常行为检测结果
Canoe 无船只经过	无船只	0.53	正常
	有船只	1.62	异常
Subway Entrance	行人进去	3.25	正常
	行人出去	5.51	异常
subway exit	行人出去	3.38	正常
	行人进去	5.32	异常

取 Adam 两段视频(地铁进口,96 min;地铁出口,43 min)的前 10 min 进行异常行为检测,本文算法与文献[12]及[7]检测结果如表 2 所示.



图 4 水面船只运动视频异常行为检测

Fig.4 The video abnormal behavior detection of moving boat on water



图 5 地铁进口视频异常行为检测

Fig.5 The video abnormal behavior detection of subway entrance



图 6 地铁出口视频异常行为检测

Fig.6 The video abnormal behavior detection of subway exit

4 结论

基于视觉稀疏超完备计算理论,提出一种新的异常行为检测算法.由实验结果可知,通过时空兴趣点、梯度和光流局部信息特征细致描述物体

可以看出,本文算法不但异常行为发现率高,而且产生错误警报率低.

表 2 地铁进口与出口处视频异常行为分析结果
Tab.2 The result of video abnormal behavior detection on subway exit and Subway entrance

算法	地铁进口		地铁出口	
	出站人数	错误警报	进站人数	错误警报
真实视频	21	—	9	—
文献[12]	17	4	9	2
文献[7]	21	4	9	0
本文算法	21	1	9	0

图 4~图 6 为异常行为检测结果.图 4 中,水面上无船只经过为正常情况,当出现行驶船只检测出异常.图 5 和图 6 分别为地铁进口和出口视频异常行为检测结果.图中进口处,出现行人出站时检测出异常;而出口处则相反.

的运动属性,进而利用稀疏超完备方法,能够准确的检测出视频中的异常行为.验证了该算法适用于特定场景的异常行为检测,表明算法的可行性和有效性.群体异常行为检测也是计算机视觉领域的重要前沿课题^[13],后续工作是研究这种方法

对于群体异常行为检测的效果和方式.

参考文献:

- [1] POPOOLA O P, WANG K J. Video-based abnormal human behavior recognition-a review [J]. IEEE Trans Syst Man Cybern, 2012, 42(6): 865-878.
- [2] RYOO M S, AGGARWAL J K. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities[C]//IEEE International Conference on Computer Vision. New York, USA: IEEE Press, 2009:1593-1600.
- [3] 朱旭东,刘志镜. 基于主题隐马尔科夫模型的人体异常行为识别[J]. 计算机科学, 2012, 39(3): 251-255.
- [4] LIN J, SUN Y, WANG W Q. Violence detection in movies with auditory and visual cues [C]//Computational Intelligence and Security. Los Alamites, USA: IEEE Press, 2010, 561-565.
- [5] UMAKANTHAN S, DENMAN S, FOOKES C, et al. Semi-binary based video features for activity representation [C]//IEEE International Conference on Digital Image Computing: Techniques and Applications. New York, USA: IEEE Press, 2013: 1-7.
- [6] ZHAO B, LI F F, XING E P. Online detection of unusual events in videos via dynamic sparse coding [C]//IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE Press, 2011:3313-3320
- [7] CONG Y, YUAN J S, LIU J. Sparse reconstruction cost for abnormal event detection [C]//IEEE International Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2011: 3449-3456.
- [8] 逯鹏,张姗姗,刘驰,等. 基于稀疏超完备表示的目标检测算法[J]. 仪器仪表学报, 2013, 34(6): 1273-1278.
- [9] BREGONZIO M, XIANG T, GONG S G. Fusing appearance and distribution information of interest points for action recognition [J]. Pattern Recognition. 2011, 45(3): 1220-1234.
- [10] LAPTEV I, LINDBERG T. Space-time interest points[C]//In Proceedings of the International Conference on Computer Vision. Los Alamites, USA: IEEE Press, 2003: 432-439.
- [11] SCHULDT C, LAPTEV I, CAPUTO B. Recognizing human actions: a local SVM approach [C]. ICPR, Los Alamites, USA: IEEE Press, 2004: 32-36.
- [12] ADAM A, RIVLIN E, SHIMSHONI I, et al. Robust real-time unusual event detection using multiple fixed-location monitors [J]. IEEE Trans on Pattern Anal Mach Intell, 2008, 30(3): 555-560.
- [13] 逯鹏,梁玉,陈树伟. 基于角点动能的视频群体异常行为检测[J]. 郑州大学学报(工学版), 2015, 36(3): 20-24.

Abnormal Behavior Detection Algorithm Based on Sparse Overcomplete Representation

LU Peng, ZHANG Liya, HUANG Shilei, LI Qihang, ZHANG Wei

(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: The focus question of video abnormal behavior detection was how to illustrate the behavior correctly through analysis of huge amounts of data. A new algorithm was proposed based on visual sparse overcomplete representation mechanism to extract local effective information about the interest points in the video of specific scenario, which could improve the efficiency of data processing. Firstly, the algorithm extracted the local spatial temporal interesting points (STIP) in training samples. At the same time it calculated the local spatial temporal characteristics. Then it put the characteristics into sparse overcomplete representation model to get a set of sparse matrix after training. Finally, it reconstructed the query video using the aforementioned matrix to detect abnormal behavior through the reconstruction error of local spatial temporal characteristics. In addition, the updated algorithm of sparse matrix function for different videos was proposed. Experiment results on standard database showed that our algorithm could detect abnormal behavior effectively and with higher accuracy and lower false alarm.

Key words: abnormal behavior detection; sparse overcomplete representation; spatial temporal interesting points