

一种用于异常行为检测的运动特征描述子



王昆仑 刘文璨 何小海 卿粼波 吴晓红

四川大学电子信息学院 成都 610065

(2017222050123@stu.scu.edu.cn)

摘要 目前,用于描述视频中人群的运动信息大多是基于光流的速度描述子。事实上,加速度蕴含丰富的运动信息,能够提供速度描述子在描述复杂运动模式时缺失的信息,以更好地表征复杂的运动模式。文中研究了一种运动特征描述子,使用受限玻尔兹曼机模型进行异常行为检测。首先,提取视频中的光流场信息,计算帧间加速度光流;然后,对一个时空块中的加速度信息进行直方图统计,将若干帧的所有时空块直方图特征进行拼接,从而获得加速度描述子;最后,在仅包含正常行为的训练集上建立受限玻尔兹曼机模型,在测试阶段根据测试视频重建特征与原始特征的误差大小进行异常检测。实验表明,所提出的加速度描述子结合速度描述子,在UMN数据集与UCF-Web数据集上,ROC曲线下的面积分别达到了0.984与0.958,相较于其他算法,所提方法取得了更高的异常行为检测准确率。

关键词: 异常行为;加速度光流;受限玻尔兹曼机;运动信息特征;特征提取

中图法分类号 TP391.41

Motion Feature Descriptor for Abnormal Behavior Detection

WANG Kun-lun, LIU Wen-can, HE Xiao-hai, QING Lin-bo and WU Xiao-hong

School of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

Abstract Modern motion description techniques for crowd motion in videos are mostly velocity descriptors based on optical flow. However, acceleration contains a wealth of motion information, which can provide information that the velocity descriptors are missing when describing complex motion patterns, and can better characterize complex motion patterns. This paper studies a motion descriptor, which uses an energy-based restricted Boltzmann machine model to perform anomalous behavior detection. Firstly, the optical flow information in the video is extracted, and the acceleration information is calculated through the optical flow information of two consecutive frames. Then, acceleration histogram feature is computed over spatial-temporal blocks, and all the spatial-temporal block histogram features of adjacent frames are spliced to obtain an acceleration descriptor. The Restricted Boltzmann Machine learns the normal motion patterns from the normal video training set, which is used for abnormal detection in terms of the errors of reconstructed data in detecting phase. The results show that the average area under the curve (AUC) of the UMN dataset reaches 0.984, and the area under the average curve (AUC) of UCF-Web reaches 0.958. Compared with other state-of-the-art algorithms, the proposed descriptor has superior performance on anomaly detection.

Keywords Abnormal behavior, Acceleration optical flow, Restricted boltzmann machine, Motion feature, Feature extraction

1 引言

随着新型智慧城市的建设,当今社会愈加重视公共场所的安全问题,对监控视频内容的自动化分析成为一项迫切的需求。

过去,视频监控系统主要依赖于数名工作人员同时观察多台监控设备,实时判断画面中是否出现紧急情况。随着平安城市的建设,大量的摄像头将耗费更多的人力资源,因此智能化分析监控摄像头的视频内容变得非常重要。其中,判断

监控视频中人群行为是否出现异常,是视频内容智能化分析的一项重要内容。异常事件是指行人个体或群体做出与当前场景下一般行为不同的或者与所处环境不相容的动作行为。异常行为发生次数较少、发生频率较低^[1],在公共场所中具体包括非法入侵、车辆逆行、人群聚集逃离、打架斗殴、人群骚乱等。

分析视频中人群的运动模式是全局异常行为检测的一类重要方法。对于运动模式的研究,大量方法都是受人脑处理运动信息的方式的启发而得到的。计算两帧图像间的光流来表征物体的运动信息,是其中一种典型的运动表征方法,与人

到稿日期:2019-03-01 返修日期:2019-07-06 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:四川省科技计划项目(2018HH0143);国家自然科学基金(61871278);成都市产业集群协同创新项目(2016-XT00-00015-GX)

This work was supported by the Sichuan Science and Technology Program (2018HH0143), National Natural Science Foundation of China (61871278) and Chengdu Science and Technology Project (2016-XT00-00015-GX).

通信作者:何小海(nic5602@scu.edu.cn)

眼观察运动的方式非常相似。到目前为止,仍有非常多的研究人员致力于获得能更准确地表述运动的光流计算方式^[2-3]。

在获取到原始的运动信息的基础上,需要合适的描述子来进一步提炼其中的信息。已有许多特征描述子用于提取视频中的运动信息,并成功应用于各类计算机视觉任务中。得益于方向梯度直方图(Histogram of Oriented Gradients, HOG)被成功应用于静态图像中的目标检测,Klaser等^[4]将其扩展到时间域,得到3维时空方向梯度直方图(Histogram of Oriented 3D Spatio-temporal Gradients, HOG3D)。SURF与SIFT算子也被进一步扩展到时域,分别得到了拓展的SURF与3维SIFT。Kratz等^[5]对局部区域的富运动模式建模,将时空梯度算子成功应用于人群高度密集场景下的异常行为检测。然而,这类从空域扩展到时域的特征描述子,将空间与时间等同对待,使得空间和时间域上信息的差异性不能得到充分体现。Laptev等^[6]将HOF(Histogram of Optical Flow)与HOG算子级联,以能够同时对场景信息和运动信息进行表述。Cong等^[7]将两种不同尺度规模的8箱HOF描述子级联,构成多尺度光流直方图MHOF(Multi-scale Histogram of Optical Flow),该描述子结合了稀疏重建模型,被成功应用于异常行为检测。在此基础上,Li等^[8]进一步探索了HMOFP(Histogram of Maximal Optical Flow Projections),这是HOF的另一种变体,用于拥挤场景下的异常检测。以上方法都是基于速度信息来描述视频中的运动。同时一些学者也对视频中运动的加速度信息进行了研究,例如Nallai-varothayan等^[9]首次将运动中的加速度信息融入到异常检测模型中,由于行人肢体存在摆动,其身体周围的光流方向不断变化,存在明显的光流加速度,而非行人个体不具有此特征,因此他们据此现象来检测道路中的非行人实体。Direkolu等^[10]计算两帧光流矢量的角度差来修正当前帧光流的强度,由光流角度差和强度信息组成的特征能够表征更为复杂的行为模式,在一类SVM分类器上取得了不错的检测效果。

通常,将视频中的异常行为检测作为一类典型分类问题。训练集中包含部分正常视频及部分异常视频,首先需要人工对异常视频中的异常行为进行时间维度和空间维度上的标注,读取标注良好的训练集数据来训练分类器。随后,分类器对测试集中的视频内容做出正常、异常的判断。这类方法需要大量的标注清晰的正常、异常视频,人工成本高,获取困难。另外一种研究思路是使用离群点检测来实现异常行为检测。根据正常行为视频数据构建一个正常行为特征所在的子空间,若测试集中的行为特征落入之前构建的子空间中,则认为其是正常行为,否则认为其是离群点,将其判定为异常行为。这类方法仅仅需要正常行为视频作为训练集,数据量大,标注容易,易于实现。

在真实环境中,人群在正常走动的情况下不会受到过多的外界因素干扰,速度不会突然剧烈变化。然而,在发生某种异常事件后,人群出于本能而远离危险,在这个过程中由于慌乱,运动速度可能发生显著的变化。据此,本文研究了一种加速度信息特征描述子,基于夹角的加速度直方图HAVA(Histogram of Angle Between Velocity and Acceleration)将该描述子与速度描述子级联,利用受限玻尔兹曼机模型建立正常行

为模式,通过判断测试数据的重建误差完成异常行为的检测。

2 运动信息及特征提取

本节介绍了在每帧光流信息的基础上计算帧间的光流加速度的方法,并构建了一个矢量特征描述子来描述视频的加速度信息。

2.1 加速度计算

当视频中的运动目标速度发生变化时,光流也产生相应的变化。与光流计算相似,可以根据亮度平滑性约束条件计算光流加速度,根据光流的变化可得到像素级的加速度光流^[11]。在描述复杂运动时,加速度光流能够补充速度光流缺失的运动信息。速度信息结合加速度信息与现实中描述物体机械运动的方式相符合。

加速度光流信息是通过计算光流对时间的偏导得到的。在视频中的某一帧图像中,像素点 (x, y) 的亮度值为 $E(x, y)$,光流约束方程为:

$$uE_x + vE_y + E_t = 0 \quad (1)$$

式(1)对时间求偏导可得到光流加速度约束方程:

$$aE_x + gE_y + [u \frac{\partial E_x}{\partial t} + v \frac{\partial E_y}{\partial t} + \frac{\partial E_t}{\partial t}] = 0 \quad (2)$$

其中, E_x, E_y 和 E_t 分别表示该点的亮度 $E(x, y)$ 对 x, y, t 的偏导, (u, v) 为光流矢量, (a, g) 为加速度矢量。

相邻像素点具有一致的速度变化率,由此得到平滑约束方程为:

$$\min a_x^2 + a_y^2 \text{ and } \min g_x^2 + g_y^2 \quad (3)$$

其中, a_x, a_y, g_x, g_y 分别是 a, g 对 x, y 的偏导。因此,计算光流加速度的问题被转换为根据光流加速度约束条件 (ξ_a) 和平滑约束条件 (ξ_g) 来求解光流加速度约束方程中最小误差 ξ 的问题。最小误差求解公式为:

$$\min \xi^2 = \iint (\xi_a^2 + \lambda^2 \xi_g^2) dx dy \quad (4)$$

其中,

$$\xi_a^2 = [uE_x + vE_y + E_t]^2 \quad (5)$$

$$\xi_g^2 = a_x^2 + a_y^2 + g_x^2 + g_y^2 \quad (6)$$

根据变分法求解该最小误差,计算方式如下:

$$\nabla^2 a = \lambda [aE_x^2 + gE_xE_y + uE_xE_{xt} + vE_xE_{yt} + E_xE_{tt}] \quad (7)$$

$$\nabla^2 g = \lambda [aE_xE_y + gE_y^2 + uE_yE_{xt} + vE_yE_{yt} + E_yE_{tt}] \quad (8)$$

其中, E_{xt}, E_{yt}, E_{tt} 分别是 E_x, E_y, E_t 对 t 的偏导。 a, g 的离散拉普拉斯近似解可以使用式(9)、式(10)进行迭代求解,计算式如下:

$$a_{i,j}^{(n+1)} = \bar{a}_{i,j}^{(n)} - \frac{\lambda E^{(n)} E_x}{1 + \lambda (E_x^2 + E_y^2)} \quad (9)$$

$$g_{i,j}^{(n+1)} = \bar{g}_{i,j}^{(n)} - \frac{\lambda E^{(n)} E_y}{1 + \lambda (E_x^2 + E_y^2)} \quad (10)$$

其中,

$$E^{(n)} = [a_{i,j}^{(n)} E_x + g_{i,j}^{(n)} E_y + u_{i,j} E_{xt} + v_{i,j} E_{yt} + E_{tt}] \quad (11)$$

$$\bar{a}_{i,j}^{(n)} = \frac{1}{4} [a_{i-1,j}^{(n)} + a_{i+1,j}^{(n)} + a_{i,j-1}^{(n)} + a_{i,j+1}^{(n)}] \quad (12)$$

由此得到该帧图像的加速度近似解。

2.2 加速度描述子

Nallai-varothayan等^[9]首次提出了基于光流加速度的特

征描述子,其对于检测非行人个体有较好的效果。该加速度描述子仅考虑了速度的大小,忽略了其方向上的变化,在复杂运动表述上有所欠缺。Direkoglu等^[10]通过计算光流矢量间夹角的方式,考虑了多帧光流间的联系,但除了光流角度差信息,光流强度的变化也蕴含了丰富的运动信息。计算加速度光流能够更加充分地利用光流间角度差及强度变化的信息。本文在计算速度光流图和加速度光流图的基础之上研究了一个加速度描述子——基于夹角的加速度直方图(Histogram of Angle Between Acceleration and Velocity, HAVA)。描述子的特点在于以某一区域所有像素点的速度与加速度夹角作为统计量,同时以加速度强度作为权重来对直方图分箱投票加权。

首先,将视频按照 $m \times n$ 的网格在空域上划分为多个大小相同的矩形网状空间块,同时将连续 t 帧视频中子区域对应位置相同的空间块拼接为一个完整的时空块。然后,对各时空块中所有像素点按照 2.1 节所述方法计算稠密光流图谱和加速度光流。其中,第 t 帧中点 (x, y) 的光流加速度水平方向和垂直方向上的分量分别为 $a_{(x,y)}^{(t)}$ 和 $g_{(x,y)}^{(t)}$,则该点的光流加速度矢量为 $(a_{(x,y)}^{(t)}, g_{(x,y)}^{(t)})$,光流矢量为 $(u_{(x,y)}^{(t)}, v_{(x,y)}^{(t)})$ 。加速度光流的强度及其与光流矢量的夹角分别为:

$$|OA| = \sqrt{|a_{(x,y)}^{(t)}|^2 + |g_{(x,y)}^{(t)}|^2} \quad (13)$$

$$\theta_{AV} = \langle (a_{(x,y)}^{(t)}, g_{(x,y)}^{(t)}), (u_{(x,y)}^{(t)}, v_{(x,y)}^{(t)}) \rangle \quad (14)$$

对位于同一时空块的所有像素点进行直方图统计。首先,将直方图设置为 k 箱,即将夹角 θ_{AV} 映射到 k 个区间:

$$\chi = \{\chi_1, \chi_2, \dots, \chi_k\} \quad (15)$$

其中,

$$\chi_i = \left[(i-1) \frac{360}{k}, i \frac{360}{k} \right), i=1, 2, \dots, k \quad (16)$$

然后,将加速度强度大小 $|OA|$ 作为直方图分箱投票权重,计算得到加速度光流直方图 HAVA,如式(17)所示:

$$h_i = \{ \sum |OA|, \theta_{AV} \in \chi_i, i=1, 2, \dots, k \} \quad (17)$$

最后,将这 t 帧中的所有 $m \times n$ 个子空间块的加速度信息直方图按固定顺序堆叠起来,从而获得一个更高维的特征矢量。这样即可将 t 帧视频图像中包含的运动信息表征为一个 $m \times n \times k$ 维的特征矢量。

3 基于能量的模型

在不同场景下,异常行为检测任务对异常行为的定义的差别非常大,即使在类似场景下,由于设置监控摄像头的目的各异,对异常行为也有不同的理解。目前,普遍将异常行为定义为在该场景下发生概率小、发生次数少、与周围环境不相容的一些事件。即便如此,异常行为的类型仍是多种多样的,事先穷举出所有的异常行为后再进行检测、识别,也几乎是不可能的。一种可行的方法是:基于给定的正常事件数据,通过建模来学习其中的规律,再根据待测数据与正常数据的差异来判断其是否与正常数据同分布,进而判断是否发生异常。

本文框架使用高斯伯努利受限玻尔兹曼机(Gaussian-Bernoulli Restricted Boltzmann Machine, GBRBM)^[12]作为核心模块来对复杂的数据分布进行建模,从而学习数据的规律及变化。该模型可以高效地重建数据中发生次数较多的正常

事件。异常检测阶段主要通过测试数据重建的误差大小来判断测试数据与训练模式的正常数据分布的差异^[13],进而判断视频中是否发生异常行为。

3.1 高斯伯努利受限玻尔兹曼机

受限玻尔兹曼机(RBM)是在玻尔兹曼机的基础上进一步发展而来的,由完全图变成了二分图,包括可见层与隐藏层^[14],各层神经元内无连接,层间全连接。其中,可见层神经元用来描述观察数据的一个方面或一个特征,隐藏层主要作为特征提取层。GBRBM是受限玻尔兹曼机的一种变体,将可见层输入变量由二值类型改为实值类型。

假设 GBRBM 有 m 个可见节点 $\mathbf{v} = \{v_1, v_2, \dots, v_m\} \in \mathbb{R}$ 和 n 个隐藏节点 $\mathbf{h} = \{h_1, h_2, \dots, h_n\} \in \{0, 1\}^n$ 。模型的参数包括可见层的偏置矩阵 $\mathbf{b} = \{b_1, b_2, \dots, b_m\} \in \mathbb{R}^m$ 、隐藏层偏置矩阵 $\mathbf{c} = \{c_1, c_2, \dots, c_n\} \in \mathbb{R}^n$ 以及权重矩阵 $\mathbf{W} \in \mathbb{R}^{m \times n}$, 权重矩阵 \mathbf{W} 中的元素 w_{ij} 表示可见单元 v_i 与隐藏单元 h_j 之间的连接权重。参数集合 $\phi = \{\mathbf{b}, \mathbf{c}, \mathbf{W}\}$ 可得到一个确定的受限玻尔兹曼机。

GBRBM 是一类基于能量的模型,其能量的定义如下:

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{i=1}^m \sum_{j=1}^n w_{ij} h_j v_i - \sum_{i=1}^m \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^n c_j h_j \quad (18)$$

根据 RBM 二分图的结构特点,可以计算其条件概率分布 $p(\mathbf{v}|\mathbf{h})$ 以及 $p(\mathbf{h}|\mathbf{v})$ 。RBM 层内单元无连接特点,即处于同一层的神经元相对于另一层神经元条件独立。针对两层神经元的 RBM,当给定可见层单元状态时,隐藏层单元是否激活是条件独立的;反之,当给定隐藏层单元状态时,可见层单元是否激活也是条件独立的。计算式如下:

$$p(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^m p(v_i|\mathbf{h}) \quad (19)$$

$$p(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^n p(h_j|\mathbf{v}) \quad (20)$$

可见层单元和隐层单元的条件概率分布如下:

$$p(v_i = \mathbf{v}|\mathbf{h}) = N(\mathbf{v}|b_i + \sum_j h_j w_{ij}, \sigma_i^2) \quad (21)$$

$$p(h_j = 1|\mathbf{v}) = f(c_j + \sum_i w_{ij} \frac{v_i}{\sigma_i^2}) \quad (22)$$

其中, $N(\cdot|\mu, \sigma^2)$ 代表均值为 μ 、标准差为 σ 的高斯概率密度函数。 $f(x)$ 为 sigmoid 激活函数。

RBM 是一种基于能量的模型,其学习过程主要为寻找一组参数 $\hat{\phi}$,使得网络能量最小化。对比散度算法(Contrastive Divergence, CD)是一种非常成功的 RBM 训练算法。对于训练样本集中的每一个样本 \mathbf{v} ,首先根据式(20)计算得到隐层神经元状态的概率分布,再对该概率分布通过 Gibbs 采样得到 \mathbf{h} ,类似地根据式(19)从 \mathbf{h} 中产生 \mathbf{v}' ,再根据 \mathbf{v}' 产生 \mathbf{h}' ,最终得到连接权重的更新公式,如式(23)所示:

$$\Delta w = \eta(\mathbf{v}\mathbf{h}^T - \mathbf{v}'\mathbf{h}'^T) \quad (23)$$

3.2 异常检测

图 1 是本文框架流程图,整个过程分为训练阶段和测试阶段。在训练阶段,通过获取训练集中所有视频的运动特征,构建一个正常行为的 GBRBM 模型。在测试阶段,根据模型重建测试数据误差的大小来判断是否发生异常。其中利用受限玻尔兹曼机进行异常行为检测是一种无监督方法,不需要任何先验知识来对异常行为标记标签。首先使用正常行为的视频数据构建训练集,GBRBM 模型学习其正常行为的模式,

得到一组最优化参数 $\hat{\psi}$;然后将待测视频的特征矢量送入GBRBM模型,可以在隐藏层获得该特征新的表达形式 $\tilde{\mathbf{h}}=(\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_k)$, \tilde{h}_k 的表达式如下:

$$\tilde{h}_k \sim p(h_k = 1 | \mathbf{v}) = \sigma(b_k + \sum_m w_{mk} v_m) \quad (24)$$

随后将隐藏层矢量 $\tilde{\mathbf{h}}$ 映射到可见层,得到输入的重建数据 $\tilde{\mathbf{v}}=[\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_m]^T$, \tilde{v}_m 的表达式如下:

$$\tilde{v}_m \sim p(v_m = 1 | \tilde{\mathbf{h}}; \psi) = \sigma(a_m + \sum_k w_{mk} \tilde{h}_k) \quad (25)$$

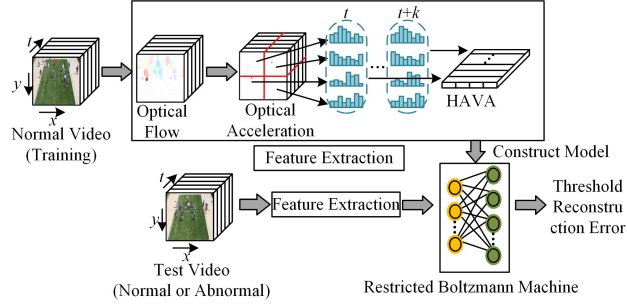


图1 本文方法的整体框架

Fig. 1 Framework of proposed abnormal detection method

由于RBM良好的因式分解特性,上述前向传播和后向传播过程的计算非常高效。在实际应用中,重建得到的数据 $\tilde{\mathbf{v}}$ 可以用来恢复在原始数据 \mathbf{v} 中被噪声干扰的元素;或者根据重建数据 $\tilde{\mathbf{v}}$ 与原始数据 \mathbf{v} 之间的重建误差来处理二分类问题。

在本文的异常行为检测框架中,测试阶段视频中若存在异常行为,其运动特征与训练阶段中仅包含正常行为的视频特征存在明显的差异。而GBRBM模型是通过正常数据训练得到的,不能够很好地描述异常行为特征。因此,重建后的异常行为特征与原始特征存在较大的差异,重建误差数值较大。而测试集中正常行为特征仍然符合训练阶段中的正常行为模式,重建后的误差较小,因此可以将其与预设门限进行比较来实现异常行为检测。使用原始数据与重建数据之差的二范数来表示该特征的重建误差,计算公式如下:

$$E = \|\mathbf{v} - \tilde{\mathbf{v}}\|_2 \quad (26)$$

将重建误差与预先设置的门限值进行比较,判断是否发生异常,计算式如下:

$$Abnormal = \begin{cases} 0, & E < Th \\ 1, & E \geq Th \end{cases} \quad (27)$$

其中, Th 为预先设定的阈值。

4 实验结果分析

为了验证本文所提描述子的有效性,在多个公共数据集上进行实验验证,数据集包括明尼苏达大学(University of Minnesota, UMN)数据集与更具挑战性的UCF-Web数据集。本文采用ROC(Receiver operating characteristic curve)作为评测标准,AUC(Area under curve)表示ROC曲线下的面积。本文算法主要是计算重建数据与原始数据的误差,与预设阈值相比较来判断视频中是否发生异常,基于GBRBM模型的重建误差用来进行异常行为检测,即判断测试数据的重建误差是否超过预设门限,进而判定视频中是否发生异常情况。

ROC曲线及AUC能够较好地表达多阈值对模型表现性能的影响。

4.1 UMN数据集

UMN数据集是明尼苏达大学提出的一个用于异常行为检测的公共数据集,包含草坪、室内、广场3个场景下的共11段视频,视频中发生的异常行为主要是人群异常恐慌四散逃离,视频原始分辨率均为 320×240 像素。数据集中每段视频中的人群最初都按照各自的行进方向自由走动,假定中途发生某种突发事件,以人群迅速四散逃离作为结束。数据集的部分图像如图2所示。



图2 UMN正常行为与异常行为图例

Fig. 2 Sample frames of normal and abnormal behaviors in UMN dataset

在进行异常检测之前,使用数据集3个场景中的正常数据建立RBM模型,提取同一场景下的部分视频数据作为训练集,从而建立一个RBM模型。

计算光流加速度特征时,将视频图像分为 2×2 的矩形网状方格,即将一帧视频图像分为4个大小相同的空间块,然后将连续5帧视频图像中对应位置相同的空间块组成一个完整的时空块,5帧视频图像共被分为4个时空块。分别对每一个时空块统计其加速度特征,得到以强度加权的方向直方图,该直方图描述子即为HAVA特征。将4个方向直方图堆叠起来得到一个高维的特征,来表述整个时空块中的运动信息。

针对UMN数据集中3个不同的场景,分别选取各场景下第一段视频中的正常视频段作为训练样本,各场景下的训练视频帧分别为1-480帧,1453-1763帧,5596-6150帧。从检测结果中发现,在室内场景下检测结果出现了少量虚警,但虚警样本数量较少且分布较为分散,不能组成连续稳定的异常状态,而异常行为的发生具有持续性,这些孤立的虚警样本可以通过后处理进行修正。检测结果对异常发生的起始与结束定位不是十分准确,相比给定的真实标签有一定程度的提前或延后。这是由于异常行为发生前后,少量样本处于不确定的状态,这些样本既可以被定义为正常样本,也可以被定义为异常样本。HAVA特征描述子是对视频中的运动信息进行表征,这些状态模糊的样本处于异常未完全发生或即将结束的阶段,其运动信息没有异常完全发生时丰富,因此针对这部分样本出现了少量误判。

表1列出了本文基于HAVA描述子的方法与当前流行算法在UMN数据集上的检测效果的比较。本文算法的检测性能在草坪与室内两个场景下取得了与最优算法相接近的结果,分别为0.9928,0.9677,并且在Plaza场景下有一定提升,达到了0.9934。与其他运动描述子HOF0^[16]与HOA^[11]相比,本文HAVA描述子在各场景下的检测准确率相近,在部分场景下有一定提升。与Direkoglu等^[10]提出的方法相比,本文方法在光流间角度差的基础上利用了光流强

度变化信息,在各场景下检测准确率均有提升,这表明合理地结合更多的运动信息能够得到更好的检测效果。结合了速度信息与加速度信息的级联描述子 HAVA+HOF,相较于单个描述子的检测更加准确。这表明在表征目标复杂运动时,加速度描述子能够有效弥补速度类型描述子丢失的信息。

表1 UMN数据集上不同算法的对比结果

Table 1 Comparison results of different methods on UMN dataset

Method	Area under ROC		
	Lawn	Indoor	Plaza
Optical flow ^[15]	0.84	—	—
HOFO ^[16]	0.9779	0.9223	0.9849
SRC ^[7]	0.9950	0.9750	0.9640
STCOG ^[17]	0.9362	0.7759	0.9661
OC-SVM ^[18]	0.9845	0.9037	0.9815
Direkoglu et al. ^[10]	0.9910	0.9485	0.9776
HOA ^[11]	0.9770	0.9631	0.9881
HOA+HOF ^[11]	0.9928	0.9673	0.9931
HAVA (proposed)	0.9766	0.9623	0.9894
HAVA+HOF (proposed)	0.9928	0.9677	0.9934

4.2 UCF-Web 数据集

为了进一步评估该算法的有效性及其鲁棒性,在更具有挑战性的 UCF-Web 数据集^[19]上进行实验。该数据集包含 12 段正常的人群活动视频,包括人群正常走动、马拉松比赛等,另外包含了 8 段诸如斗殴、恐慌逃离等异常行为视频。该数据集中有视距远近各异的多个场景,而 HAVA 是一种运动描述子,对视频中个体运动幅度的大小敏感,这导致相同大小的光流矢量在不同场景下表示的人群实际运动幅度大小存在较大的差异,这对本文描述子表征运动信息带来了一定的干扰。因此,根据视距将数据集进一步划分为两个子数据集,分别由视距较远的视频段与视距较近的视频段组成。具体划分如图 3 所示,其给出每段视频中具有代表性的关键帧。图 3(a)为远视距场景,第一行为 3 段异常视频,第二行与第三行为 8 段正常视频;图 3(b)为近视距场景,第一行为 5 段异常视频,第二行为 3 段正常视频。



(a) UCF-Web 数据集中远视距场景划分



(b) UCF-Web 数据集中近视距场景划分

图3 UCF-Web 中正常行为与异常行为图例

Fig. 3 Sample frames of normal and abnormal behaviors in UCF-Web dataset

采用与 UMN 数据集相同的实验参数,将视频划分为 2×2 的网状区域,连续 5 帧构成一个时空块,在此时空块中提取 HAVA 特征。HAVA 特征设置为 8 箱的直方图,时空块的 4 个子区域分别计算 HAVA 特征,随后将其按照固定顺序堆叠,得到一个 $8 \times 4 = 32$ 维的融合特征矢量。在异常检测阶段,由于两个子数据集视距远近不同,运动幅度大小存在一定的差异,需要分别建立各自的 RBM 模型。与 UMN 数据集建立模型方式类似,首先提取出子数据集中所有视频段的特征,由于 UCF-Web 数据集中的视频片段与 UMN 数据集有所差异,没有同时包含正常行为和异常行为的视频段,因此其训练集与测试集的划分方式与 UMN 不同,具体为随机选取正常视频特征中的 $3/4$ 作为训练集,学习得到一个 RBM 模型,将剩余的 $1/4$ 与异常视频放入测试集中进行测试。将 10 次实验结果的平均值作为模型最终的表现。

表 2 列出了本文算法与当前流行算法的结果,以曲线下的面积 AUC 为评价指标。针对真实环境下复杂场景中的人群运动模式,本文结合了加速度信息与速度信息的级联描述子(HAVA+HOF),本文方法具有更好的表征能力,在该数据集下取得了更好的结果。表 3 进一步比较了本文算子 HAVA 与其他同类型描述子 HOA 在单独和级联情况下的表现。本文的 HAVA 算子在单独使用的情况下较同类型其他算子能够取得更好的效果。与单独的加速度算子相比,加速度特征描述子与速度特征描述子(HAVA+HOF)相结合能够得到更好的表现效果,这说明了加速度信息与速度信息相结合有更好的表现效果,且加速度信息与速度信息相互补充,对复杂的人群运动模式具有较好的表征效果。

表2 UCF-Web数据集上不同算法的对比结果

Table 2 Comparison results of different methods on UCF-Web dataset

Method	AUC
SRC ^[7]	0.84
SFM ^[15]	0.73
OF ^[15]	0.66
WIF ^[20]	0.85
NMF ^[19]	0.89
CSR ^[21]	0.900
HAVA+HOF (ours)	0.958

表3 单独描述子与级联描述子结果对比

Table 3 Comparison results of every single descriptor and combined descriptor

Method	近视距 AUC	远视距 AUC
HAVA(ours)	0.9452	0.9051
HOA	0.9363	0.8950
HAVA+HOF(ours)	0.9751	0.9403
HOA+HOF	0.9651	0.9348

结束语 本文研究了一种表征加速度运动信息的描述子,首先通过对每个像素点计算速度光流来得到加速度光流,应用直方图对运动信息进行编码得到特征描述子 HAVA。将 HAVA 与 HOF 算子级联,在基于能量的模型受限玻尔兹曼机中完成全局异常行为检测。多个数据集上的实验结果表明,加速度信息描述子能够表达部分速度描述子丢失的信息,在全局异常行为检测中更加准确。

参 考 文 献

- [1] SODEMANN A A, ROSS M P, BORGHETTI B J. A review of anomaly detection in automated surveillance[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012, 42(6):1257-1272.
- [2] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017:6.
- [3] YANG Y, SOATTO S. Conditional Prior Networks for Optical Flow[C]// *European Conference on Computer Vision*. Springer, 2018:282-298.
- [4] KLASER A, MARSZALEK M, SCHMID C. A spatio-temporal descriptor based on 3d-gradients[C]// *19th British Machine Vision Conference*. British Machine Vision Association, 2008:275:1-10.
- [5] NISHINO K, KRATZ L. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models [C]// *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, IEEE, 2009:1446-1453.
- [6] LAPTEV I, MARSZALEK M, SCHMID C, et al. Learning realistic human actions from movies[C]// *Computer Vision and Pattern Recognition*. IEEE, 2008:1-8.
- [7] CONG Y, YUAN J, LIU J. Sparse reconstruction cost for abnormal event detection[C]// *Computer Vision and Pattern Recognition*. IEEE, 2011:3449-3456.
- [8] LI A, MIAO Z, CEN Y, et al. Abnormal event detection based on sparse reconstruction in crowded scenes [C] // *Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016: 1786-1790.
- [9] NALLAIVAROTHAYAN H, FOOKES C, DENMAN S, et al. An MRF based abnormal event detection approach using motion and appearance features[C]// *Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2014:343-348.
- [10] DIREKOGLU C, SAH M, O'CONNOR N E. Abnormal crowd behavior detection using novel optical flow-based features[C]// *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2017:1-6.
- [11] EDISON A, JIJI C V. Optical Acceleration for Motion Description in Videos[C]// *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Honolulu, 2017:21-26.
- [12] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786):504-507.
- [13] VU H, NGUYEN T D, TRAVERS A, et al. Energy-based localized anomaly detection in video surveillance[C]// *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer. Cham, 2017:641-653.
- [14] ACKLEY D H, HINTON G E, SEJNOWSKI T J. A learning algorithm for Boltzmann machines[J]. *Cognitive Science*, 1985, 9(1):147-169.
- [15] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model[C]// *Computer Vision and Pattern Recognition*. IEEE, 2009:935-942.
- [16] WANG T, QIAO M, DENG Y, et al. Abnormal event detection based on analysis of movement information of video sequence [J]. *Optik-International Journal for Light and Electron Optics*, 2018, 152:50-60.
- [17] SHI Y, GAO Y, WANG R. Real-time abnormal event detection in complicated scenes[C]// *Pattern Recognition (ICPR)*. IEEE, 2010:3653-3656.
- [18] WANG T, SNOUSSI H. Detection of abnormal visual events via global optical flow orientation histogram[J]. *IEEE Transactions on Information Forensics and Security*, 2014, 9(6):988-998.
- [19] ZHU X, LIU J, WANG J, et al. Sparse representation for robust abnormality detection in crowded scenes[J]. *Pattern Recognition*, 2014, 47(5):1791-1799.
- [20] ZHU X, LIU J, WANG J, et al. Weighted interaction force estimation for abnormality detection in crowd scenes[C]// *Asian Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2012:507-518.
- [21] CHEN H, GAI J, ZHANG S, et al. Abnormal event detection based on cospase reconstruction[J]. *The Journal of Engineering*, 2018, 2018(5):254-256.



WANG Kun-lun, born in 1995, post-graduate. His main research interests include computer vision and pattern recognition.



HE Xiao-hai, born in 1964, Ph.D, professor, Ph.D supervisor. His main research interests include image processing, pattern recognition, computer vision, image communication, and software engineering.