

视频监控中的人群异常行为检测与定位

周培培^{1,2,3,4*}, 丁庆海^{1,5**}, 罗海波^{1,3,4}, 侯幸林^{1,2,3,4}

¹中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;

²中国科学院大学, 北京 100049;

³中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016;

⁴辽宁省图像理解与视觉计算重点实验室, 辽宁 沈阳 110016;

⁵航天恒星科技有限公司, 北京 100086

摘要 人群中异常行为是一大潜在威胁, 自动检测监控中的异常行为成为近年的研究热点之一。然而, 由于异常的未知性与复杂性, 已有的检测方法仍然存在检测率低、定位精度差的问题。为此, 提出了对视频监控中的人群异常行为自动检测与定位的方法。结合灰度值与光流场的分布提取运动区域; 对运动区域分割得到有效的运动块, 从中提取表示外观和动态的两种特征, 即局部 H 梯度方向直方图 G 和局部 H 光流方向直方图 F 特征; 使用 k -means 方法对运动块进行聚类, 对每类样本使用一类分类器进行建模。最后, 加入运动连续性约束, 以抑制干扰噪声。在两个复杂的异常行为数据集上的实验结果表明, 本文方法明显优于已有的检测方法, 且可以满足正确率高、抗干扰能力强等实际工程需求。

关键词 机器视觉; 模式识别; 人群异常检测; 运动区域分割; 特征提取; 一类分类器; 运动连续性滤波

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/AOS201838.0815007

Anomaly Detection and Location in Crowded Surveillance Videos

Zhou Peipei^{1,2,3,4*}, Ding Qinghai^{1,5**}, Luo Haibo^{1,3,4}, Hou Xinglin^{1,2,3,4}

¹Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning 110016;

²University of Chinese Academy of Sciences, Beijing 100049;

³Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang, Liaoning 110016;

⁴Key Laboratory of Image Understanding and Computer Vision, Shenyang, Liaoning 110016;

⁵Space Star Technology Co., Ltd., Beijing 100086

Abstract The anomaly in the crowd is a great potential threat, and the automatic detection of abnormal behavior for surveillance has become a hot topic in recent years. However, because the anomaly is unknown and complex, the previous detection methods still suffer from a low detection rate and poor location accuracy. To this end, a method is proposed for anomaly detection and location in the crowded surveillance videos. First, the motion regions are extracted according to the distributions of the gray-scale value and the optical flow field. Second, the effective motion blocks are obtained by segmenting the motion regions. Two features, namely the local H histogram of gradient G and the local H histograms of flow F, are extracted from the motion blocks, representing the appearance and dynamics. Third, the motion blocks are clustered with the k -means method, and each cluster is modeled using a one-class classifiers. Finally, the motion continuity constraint is added to suppress the noisy noises. Experimental results on two complex abnormal behavior datasets show that the proposed method is obviously better than previous detection methods. It could meet the practical engineering needs such as high accuracy and strong anti-interference ability.

Key words machine vision; pattern recognition; crowd anomaly detection; motion region segmentation; feature extraction; one class classifier; motion continuity filtering

OCIS codes 150.0155; 100.5010; 100.2000; 100.2960

收稿日期: 2018-01-22; 修回日期: 2018-02-06; 录用日期: 2018-02-26

基金项目: 中国科学院国防科技创新基金(Y6A4160401)

* E-mail: zhoupeipei@sia.cn; ** E-mail: 13693689880@139.com

1 引 言

人群中的异常行为严重影响着人们的生命财产安全,随着监控设备的普及,海量的监控数据使得安防人员的压力空前剧增,自动检测与定位异常行为成为近年来的研究重点之一。在复杂的人群视频场景中,异常行为发生的次数较少且模式多变,因此,难以对异常行为进行准确描述。根据 Cong 等^[1]的定义,“异常事件”是指场景中发生概率较低的事件。本文针对仅有正常行为模式的训练集视频进行学习,对混有正常和异常行为的测试视频进行预测,标记出异常帧,并在异常帧中标记异常行为发生的位置。

不同监控场景中可能发生的异常行为差别较大,根据不同场景中不同的检测任务,研究人员提出了多种不同的检测方法^[2],大致分为以下三个方面。

1) 行为关系分析法。使用概率模型描述不同行为之间的关系,通过对这些关系进行统计分析来判断异常的发生概率。研究人员设计了多种统计模型如统计显著关联规则模型(SARM)^[3]、依赖狄利克雷过程的隐马尔科夫模型(DDP-HMM)^[4]高斯模型^[5]以及混合高斯模型^[6],对行为之间的关系进行建模。Kwon 等^[7-8]把异常检测问题转换成图模型编辑问题,一个视频表示成一个图模型,时空分割得到的视频块表示成节点,用节点之间的连接权值描述事件之间的关系,学习得到可靠的关联关系。该类方法效果较好,然而,需要大量的训练样本才能获得稳定可靠的关系模型。

2) 运动轨迹分析法。基于已有的目标跟踪方法^[9],提取兴趣目标的运动轨迹,分析判断是否为异常轨迹点。常用的方法使用跟踪轨迹的速度与方向等特征^[10-11]来建模正常轨迹。此外,另有时空轨迹检测方法,如 Tran 等^[12]使用时空路径搜索方法检测与定位复杂运动目标的异常行为;Nguyen 等^[13]构建了共享结构的分层隐马尔科夫模型(HHMM),以从正常模式中区分异常目标轨迹。该类方法主要对轨迹数据建模,然而,由于人群视频中的目标之间存在重叠、遮挡等问题,难以跟踪到精确的目标轨迹。

3) 底层特征分析法。通过提取运动特征、外观特征或者基于深度学习的视觉特征等底层特征,建模分析判断是否发生了异常行为。Mehran 等^[14]使用视频的光流特征建立了社会力流图像,对力流图像采用隐狄利克雷分布(LDA)模型区分异常帧。基于光流信息,研究人员还建立了多种模型对正常

行为进行建模,如运动影响系数矩阵^[15]和运动能量模型^[16],从而检测异常。然而,光流特征无法表达运动目标的外观信息。兼顾外观与运动特征,研究人员提出了多种检测模型分析人群中的异常行为,如基于梯度的时空运动模式(STG-LMP)描述子^[17],联合使用运动目标的速度、尺寸和纹理信息的综合模型^[18],以及对时间和空间异常行为分别设计的互补检测器^[19-20]等。然而,已有的时空模型仍然存在准确率低、虚警率高的问题。

近年来,随着卷积神经网络在行为识别领域研究的逐步深入,基于深度学习的视觉特征逐渐应用到异常行为检测模型中。Sabokrou 等^[21]设计了一个非监督的自编码神经网络,使用隐含层的输出作为时空视频块的全局特征描述子;Sabokrou 等^[22]还设计了三维(3D)深度自编码网络来表征正常视频块;Zhou 等^[23]使用时空卷积神经网络(CNN)提取运动和外观特征;有学者采用基于主成分分析网络(PCANet)的深度特征^[24-25]表征视频块,然后设计分类器检测出异常视频块。由于异常数据集难以搜集,深度网络不能直接用于异常预测,主要用来提取视频特征。然而,基于深度网络提取视频特征的计算量较大,难以满足实时性要求。

本文提出了一种稳健性好、准确率高、可解释性强的人群异常行为检测与定位方法,基于视频的时空特征,兼顾外观与动态信息,提出了两种时空特征描述子,使用一类分类器对特征建模,并引入运动连续性约束对一类分类器预测的异常目标进行滤波。在加州大学圣地亚哥分校(UCSD)^[26]和明尼苏达大学(UMN)^[27]搜集的两个常用的基准数据集上,本文算法取得了较好的检测效果。

2 异常行为检测与定位

如图 1 所示,算法流程大致分为 4 个步骤,即运动区域分割、特征提取、分类器训练/预测和运动连续性滤波。

2.1 运动区域分割

异常行为的发生应满足以下两个条件:1)是运动目标;2)目标的行为与场景中其他行为明显不同。因此,首先提取运动区域,对运动目标的正常行为建模,然后区分出异常行为。

目前,快速有效的背景建模方法是视频背景估计(ViBE)算法^[28],该算法为背景建立像素级的样本估计模型,若当前像素判断为背景像素,则更新背景模型,随机选择模型中某个像素样本进行替换,随

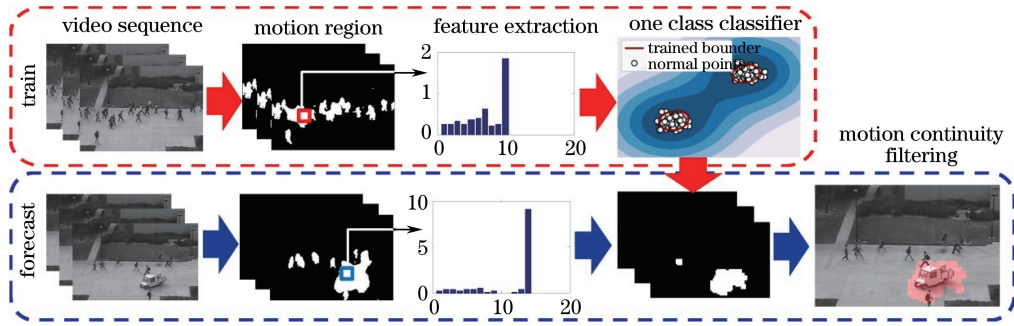


图 1 异常行为检测与定位算法流程图

Fig. 1 Flow chart of abnormal behavior detection and location

机选择邻域像素进行替换;否则背景模型不做更新。背景模型的两种随机更新策略使得该算法对噪声的稳健性较好。然而,对于图 2(a)中与背景颜色相近的运动目标,该算法无法区分目标区域,检测结果如图 2(b)所示。对此,提出了融合运动信息的 ViBE 算法,在 ViBE 算法的基础上加入光流强度信息,以增强运动目标与背景模型的区分度,如图 2(c)所示,运动目标得以增强。

采用 Zach 等^[29]提出的光流提取方法,对第 t 帧图像计算 x 和 y 两个方向的光流图像: $f_x(t)$ 和 $f_y(t)$,光流强度 $M(t)$ 的定义为

$$M(t, i, j) = \sqrt{f_x(t, i, j)^2 + f_y(t, i, j)^2}, \quad (1)$$

式中 (t, i, j) 表示第 t 帧的像素位置。对于当前帧图像 $f(t)$,增强后的图像 $F(t)$ 表示为

$$F(t, i, j) = f(t, i, j) + \lambda M(t, i, j), \quad (2)$$

式中 λ 是正实数,本文实验中取 $\lambda = 10$ 。

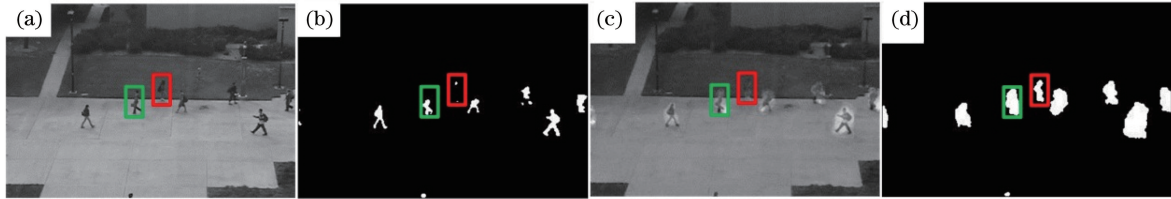


图 2 运动区域分割效果图。(a)原始图像;(b) ViBE 算法得到的运动区域;
(c)加入运动信息的图像;(d)本文算法得到的运动区域

Fig. 2 Segmented motion-region images. (a) Original image; (b) motion regions achieved by the ViBE method;
(c) image with motion magnitude; (d) motion regions achieved by the proposed algorithm

采用融合运动信息的 ViBE 算法分割得到的运动区域如图 2(d)所示。对比图 2 方框中的两个行人可见,本文算法可以区分与背景相似的运动目标,得到更完整的运动区域,但是会将某些运动区域边界放大,这是由于在光流图像中这些区域也存在明显的运动信息,但并不影响后续异常行为检测。提出的运动区域分割算法可以准确检测运动目标,为后续异常检测降低了背景干扰。

2.2 特征提取

运动区域分割之后,从运动区域中提取特征。传统的梯度方向直方图(HOG)特征^[30]是针对行人整体检测设计的,在行人无遮挡情况下效果良好。然而,由于视频中人与人之间常常相互遮挡,导致身体某些部分无法检测到。因此 HOG 特征在这类视频中表现不佳。据此,针对灰度图像设计了局部

HOG 特征描述子(H_α),对运动区域提取外观特征。此外,为充分利用视频的时域信息,针对光流图像设计了局部光流方向直方图(HOF)特征描述子(H_β),以获取运动特征。联合使用两种特征,对运动区域的时空信息进行描述。

运动区域被划分成有重叠的运动块,特征提取算法描述如下:

输入:第 t 帧灰度图像 $f(t)$,第 t 帧的光流图像 $f_x(t)$ 和 $f_y(t)$,运动区域分割图像 $F_{\text{mask}}(t)$,运动块的宽高尺寸 b_w 和 b_h ,运动块在图像横纵方向上的步进长度 s_w 和 s_h ,直方图的主方向个数 b_{in} 。

输出:图像 $f(t)$ 的特征描述子 $g(t)$ 。

1) 获取有效运动块。以大小为 $b_w \times b_h$ 、横纵向步长分别为 s_w 和 s_h 的矩形遍历图像 $f(t)$,矩形框所覆盖的图像区域视为一个图像块,比较图像

$F_{\text{mask}}(t)$ 在该位置处的运动像素的个数,若运动像素数超过 50%,则该图像块视为一个有效运动块,记录其左上角坐标。

2) 计算梯度。①使用文献[30]的方法计算图像 $f(t)$ 的每个像素点灰度值的梯度,得到外观梯度 G_α 的幅值 A_α 和方向 Θ_α ,其中 $\Theta_\alpha \in [0, \pi]$;②计算图像 $f(t)$ 运动量的梯度 G_β ,其中 G_β 的幅值 A_β 根据(1)式计算得到, G_β 的方向值 Θ_β 由(3)式计算得到,即

$$\Theta_\beta(t, i, j) = \begin{cases} \arctan \frac{q}{p + \epsilon}, p \geq 0, q \geq 0 \\ \arctan \frac{q}{p + \epsilon} + \pi, p < 0 \\ \arctan \frac{q}{p + \epsilon} + 2\pi, p \geq 0, q < 0 \end{cases}, \quad (3)$$

式中 $p = f_x(t, i, j), q = f_y(t, i, j), \epsilon$ 是一个很小的正数。

3) 获取图像 $f(t)$ 的第 n 个运动块的特征向量 $g_n(t)$ 。对于外观梯度 G_α ,把 $[0, \pi]$ 均匀分成 b_{in} 个主方向,把该运动块的像素按 Θ_α 逐个投影到 b_{in} 个直方图中,根据幅值 A_α 对 $g_n(t)$ 的直方图进行加权投影,得到局部 HOG 描述子 H_{α_n} 。同理,根据运动量梯度 G_β 得到 $1 \times b_{in}$ 的向量 H_{β_n} ,从而得到第 n 个运动块的特征向量 $g_n = [H_{\alpha_n}, H_{\beta_n}]$ 。

4) 图像 $f(t)$ 的特征描述子表示为 $g(t) = [g_1; g_2; \dots; g_n; \dots; g_N]$,其中, N 为图像 $f(t)$ 中提取的有效运动块的总个数。

2.3 分类器设计

常见的分类问题是二分类或多分类问题,训练数据集中有多种标记样本。文中涉及的问题为一类分类器学习问题,训练数据中绝大部分是正常行为样本甚至只有正常行为的样本,这里记为正样本,标签数据是 1,测试过程需要把异常行为预测出来,即区分混杂在正样本中的负样本。这种情况下,需要训练一个对正样本紧凑的分类边界,超出这个边界的视为负样本,即异常行为数据。

直观上来看,尽管训练数据中的行为均属于正常行为,有一定的相似性,如行走的人、相似的移动速度等,但是这些行为依然存在类内差异,如人身体的不同部分、不同的速度方向等。因此,直观上来看,从不同运动块中提取的特征应该被归到不同的类。若能充分利用先验信息,则能增加分类的准确率。由于数据样本无标注,因此采用了聚类方法对

训练数据分组,不同组中的数据再使用一类分类器进行建模,得到正常行为的边界。

设计下述分类算法:1)训练阶段,首先,对训练集视频提取的运动块特征使用 k -means 方法进行聚类,记录每个类的聚类中心;然后,对每类数据学习样本的紧凑边界,得到一个一类分类器。2)测试阶段,对于一个运动块的特征向量样本,首先计算待测样本到不同聚类中心的距离,将该样本归属于距离最小的类中,然后使用该类的一类分类器对待测样本进行预测。

对于聚类后的训练样本,一类分类器的设计采用文献[31]中的分类思想。对于训练数据集中的 K 个运动块,提取 K 个向量的训练数据集 $X = \{x_1, x_2, \dots, x_K\}$ 。由于高维特征向量往往线性不可分,因此采用核函数把特征向量映射到内积空间 $\Phi: X \rightarrow F$,核函数为

$$k(x, y) = [\Phi(x) \cdot \Phi(y)], \quad (4)$$

式中 $x, y \in X$ 。常用的核函数如线性核函数、多项式核函数、径向基核函数等。训练集正常样本的边界学习转化成二次规划问题,即

$$\min_{\omega \in F, \xi \in R, \rho \in R} \frac{1}{2} \|\omega\|^2 + \frac{1}{vK} \sum_i \xi_i - \rho, \quad (5)$$

$$\text{s.t. } [\omega \cdot \Phi(x_i)] \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad (6)$$

式中 $i \in \{1, 2, \dots, K\}$,每个样本被赋予一个松弛变量,以增加分类器的容错性。惩罚因子 $v \in (0, 1]$,是一个标量, v 代表了训练样本中离群点的比例。通过对训练集数据的学习,得到(5)式和(6)式的参数 ω 和 ρ 。最后,分类器的决策函数为

$$f(x) = \text{sgn}[\omega \cdot \Phi(x_i) - \rho]. \quad (7)$$

高维数据的分类一直是一个难点问题,本文充分利用了训练数据中的正常行为可分为多个类这一先验知识,并在一类分类器设计时采用核函数和引入松弛变量,把数据转化成高维空间中线性可分的数据。实际数据集上的测试结果表明了本文分类器的有效性。

2.4 运动连续性滤波

由于正常行为的运动块存在某些离群点和噪声,分类器有时会检测到一些离散的异常运动块。在连续帧中,这些异常运动块呈现无规律分布的特点,图 3(a)为使用分类器得到的当前帧(红色框)及其前后 3 帧的预测图像,白色为异常区域,图中的一些零散分布的小块可能是虚警异常。

根据常识推理,运动目标不可能凭空出现,也不可能凭空消失,若帧频为 $50 \text{ frame} \cdot \text{s}^{-1}$,那么持续

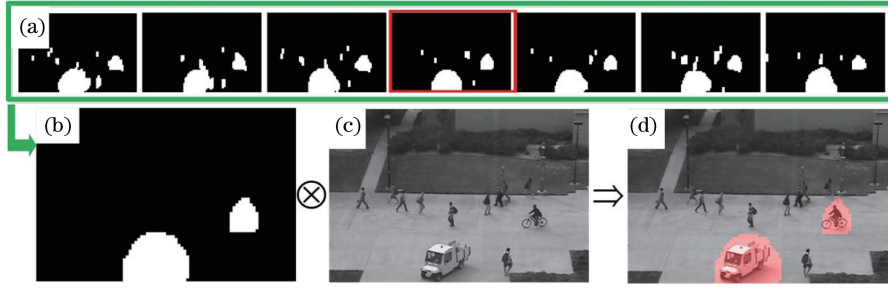


图 3 对分类器预测的异常区域使用运动连续性滤波后的结果

Fig. 3 Filtering results according to motion continuity for the abnormal regions which are predicted by the classifiers

0.2 s的异常行为至少被 10 帧图像检测到,0.2 s的时间内目标的移动距离是有限的。据此,本文引入了运动连续性约束:对于某一异常目标,连续 T 帧中其所在区域的交集为非空。

对使用分类器得到的第 t 帧的异常区域二值图像 $B(t)$ 进行分析:1)对图像 $B(t)$ 标记连通域 $\{C_1^t, C_2^t, \dots, C_l^t, \dots, C_L^t\}$;2)对于每个连通域 $C_l^t (1 \leq l \leq L)$,分别计算 C_l^t 与其前后 T 帧的异常区域的交集 I_{ot}^l 和 I_{pt}^l ,再对 I_{ot}^l 和 I_{pt}^l 取并集,得到 U_l^t ;3)若 U_l^t 的面积与 C_l^t 的比值超过阈值,则认为该 C_l^t 是异常区域,否则视为噪声。图 3(b)为滤波后的图像,由图可知,虚警噪声干扰得到了有效抑制;图 3(d)中阴影部分为滤波后的异常区域,效果较好。

3 实验与讨论

为了测试本文算法的有效性,在两个常用的基准数据集,即 UCSD^[26] 和 UMN^[27] 上进行了实验。这两个数据集涵盖了室内和室外、局部异常和全局异常等不同的情况。实验中采用帧级和像素级两种度量标准。帧级度量^[32]是指若某帧中检测到至少有一个像素是异常的,那么该帧被视为异常帧。像素级度量^[32]是指若某帧中有异常像素,实际标注的异常像素中 40%以上被异常检测算法检测到,那么该帧被视为正确检测的异常帧(真阳性帧),否则视为假阳性帧。实验中,从定性与定量两个方面对本文算法与已有算法进行了对比。定性评价标准采用同一帧的异常检测与定位情况进行显示,定量评价采用受试者工作特征(ROC)曲线、等错误率(EER)和检测率(DR)等指标。其中,EER定义为 ROC 曲线中错分正负样本概率相等的点所对应的错分概率值^[32]。

3.1 数据集

UCSD 数据集包括 ped1 和 ped2 两个数据子集,像素分辨率分别为 158 pixel \times 238 pixel 和 240 pixel \times 360 pixel,均为室外场景,移动目标主要

是正常行走的行人,其余行为均视为异常,包括了汽车、滑板车、自行车、轮椅等。实验选择在 ped2 数据集上评估算法性能。ped2 数据集包括 16 个训练集视频和 12 个测试集视频,其中 12 个测试集视频均有像素级异常标注,本文使用了所有的测试视频来评估算法。

UMN 数据集共包括 11 段视频,像素分辨率均为 240 pixel \times 320 pixel,包括了三个场景,如图 4 所示。一群人停住、行走或游荡,视为正常行为;人群奔跑(逃散),视为异常行为。



图 4 UMN 数据集中的正常与异常人群行为举例。
(a)正常行为;(b)异常行为

Fig. 4 Examples of normal and abnormal crowded activities in UMN dataset.

(a) Normal behaviors; (b) abnormal behaviors

3.2 实验结果

运动区域分割之后,对运动区域提取有效运动块,实验中的运动块尺寸固定为 8×16 ,运动块的描述子为局部 HOG 特征和局部 HOF 特征的组合,实验中两种特征的直方图主方向数均选为病句。运动连续性滤波中,通过选用不同数量的连续图像,对预测图像进行不同强度的干扰抑制,从而得到 ROC 曲线的不同点。

以 UCSD ped2 数据集为例,为了说明提出算法的有效性,本文选用的对比算法不仅包括一些经典算法,如社会力模型^[33]、光流混合(MPPCA)模型^[34]、空域动态纹理混合(S-MDT)模型^[32]和时域

动态纹理混合 (T-MDT) 模型^[32], 也包括时空卷积神经网络 (ST-CNN) 模型^[23] 和运动能量模型^[16] 等。

图 5 所示为不同检测算法的帧级和像素级异常检测的 ROC 曲线, 纵坐标为真正率 (TPR), 横坐标为假正率 (FPR)。从图 5 可以看出, 当虚警率要求较高 (如 0.05) 时, 算法表现不太理想, 这与场景本身有关; 当虚警率大于 0.1 时, 本文算法在两种度量方式下均明显优于已有的算法。表 1 列举了两种度量

方式下的不同指标的对比结果。由表 1 可知, 当采用帧级度量时, 本文算法的 EER 明显降低, 并显著提升了 ROC 曲线下的面积 (AUC) 指标; 当采用像素级度量时, 在 DR、AUC 以及每帧的平均耗时方面, 与运动能量模型^[16] 相比, 本文算法耗时相对较多, 但是 DR 和 AUC 均明显占优; 与 ST-CNN 模型相比, 本文算法在耗时方面稍微改善, 在检测率方面提升了近 8%, AUC 也提高了 0.02。

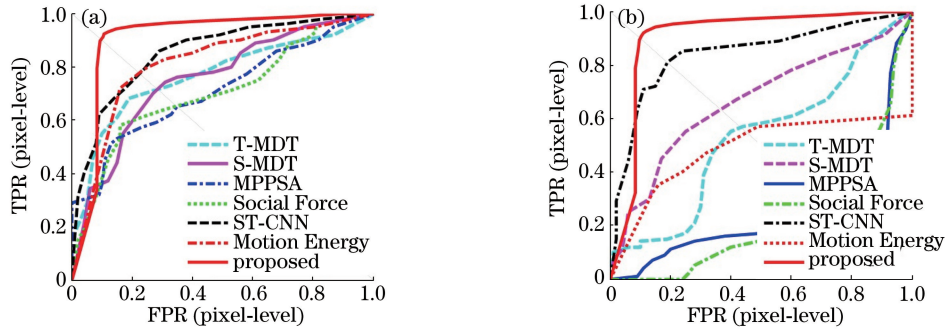


图 5 数据集 UCSD ped2 上, 不同算法在两种度量方式下的 ROC 曲线图。(a) 帧级度量; (b) 像素级度量

Fig. 5 ROC curves of different methods in two criteria on the UCSD ped2 dataset. (a) Frame-level; (b) pixel-level

表 1 UCSD ped2 数据集上, 不同算法在帧级和像素级不同标准的比较结果

Table 1 Comparison results of different methods on the UCSD ped2 dataset for frame-level and pixel-level detection

Method	Frame-level criterion		Pixel-level criterion		
	EER / %	AUC	DR / %	AUC	Time / s
Social Force ^[33]	42.0	0.702	27.6	0.217	-
MPPCA ^[34]	31.1	0.710	22.4	0.222	-
S-MDT ^[32]	28.7	0.750	63.4	0.665	0.69
T-MDT ^[32]	27.9	0.765	56.8	0.522	0.64
ST-CNN ^[23]	24.4	0.860	81.9	0.880	0.37
Motion Energy ^[16]	22.0	0.810	55.0	0.580	0.08
Proposed	10.3	0.905	89.7	0.902	0.26

此外, 在 UCSD ped2 数据集上对算法进行定性估计, 图 6 所示为本文算法与 4 种经典算法的像素级检测结果。由图 6 可知, 本文算法可以准确检测出与行人外貌差别较大的自行车、汽车等区域, 也可

以准确检出速度明显不同的滑板车等目标。此外, 对比已有的算法发现, 本文算法能较好地覆盖到异常区域, 检测率较高。

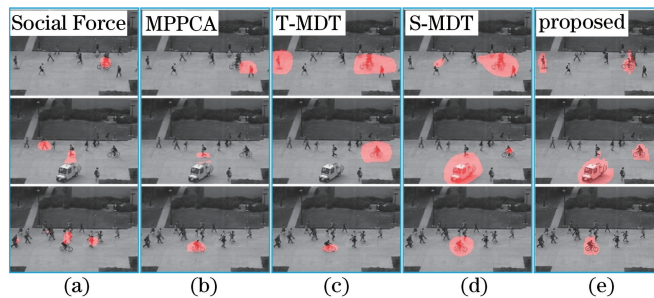


图 6 UCSD ped2 数据集上, 不同算法的像素级检测结果。(a) Social Force; (b) MPPCA; (c) T-MDT; (d) S-MDT; (e) 本文算法

Fig. 6 Pixel-level detection results with different methods on the UCSD ped2 dataset.

(a) Social Force; (b) MPPCA; (c) T-MDT; (d) S-MDT; (e) proposed method

另外,在 UMN 数据集上对算法作了实验验证。UMN 数据集视频中无训练和测试的分界标记,对于每个场景的视频,实验中采用每个视频的前一半作为训练集,后一半作为测试集。由于该数据集不含有像素级标注,仅含有帧级标注,因此,这里采用帧级度量下的 EER 和 AUC 两个指标评估不同的算法。

图 7 所示为在 UMN 数据集上不同算法的 ROC 曲线,可见,在虚警率小于 0.02 时,提出的算法的检测率较低,即虚警率要求严苛时,本文算法性能稍差,原因是该算法在群体有散开的动作时即视为异常,此时行人处于加速阶段,已明显异于正常行走速度,而数据集中的实际标注为群体完全散开时视为异常;当虚警率高于 0.02 时,提出的算法性能明显提高,即该算法对于明显的异常可以准确捕获。

通过 ROC 曲线计算可得 EER 和 AUC 的值,如表 2 所示为对比结果,可见,提出算法的 AUC 指标与已有的算法相当,但是在 EER 指标上略优于其他算法。

图 8 所示为本文算法在三个场景中的异常预测情况,可见,本文算法可以较为准确地定位到异常行为发生的位置,且虚警较少。

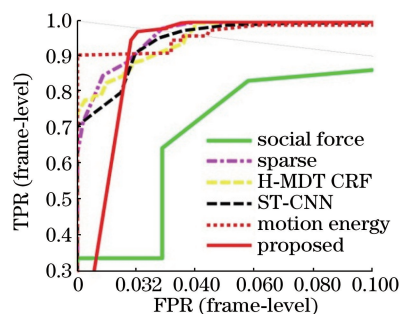


图 7 数据集 UMN 上,不同算法在帧级度量方式下的 ROC 曲线图

Fig. 7 ROC curves of different methods for frame-level detection on the UMN dataset

表 2 UMN 数据集上,不同的算法在帧级度量时 AUC 与 EER 的比较结果

Table 2 Comparison results of different methods on the UMN dataset in AUC and EER criterion for frame-level detection

Method	AUC	EER / %
Social Force ^[33]	0.949	12.6
Sparse ^[1]	0.996	2.8
H-MDT CRF ^[19]	0.995	3.7
ST-CNN ^[23]	0.996	3.3
Motion Energy ^[16]	0.989	4.1
Proposed	0.989	2.7



图 8 UMN 数据集三种不同场景的异常行为检测与定位效果图

Fig. 8 Detection and location results of three different scenes on UMN dataset

3.3 参数讨论

对于 k -means 聚类中的参数 k ,随着 k 值的增加,计算复杂度增加,系统性能呈现先增后减的分布趋势,以检测率为依据,可以得到一个最优的 k 值。在 UCSD ped2 和 UMN 数据集上,选用帧级检测率作为系统性能指标,得到参数 k 与系统性能的关系曲线如图 9 所示。本文实验中,UCSD ped2 数据集中 $k=5$,UMN 数据集中 $k=3$ 。

在(5)式中,一类分类器设计中的惩罚因子 $v \in (0, 1]$ 代表了训练集中噪声所占的比例,训练过程中,分类器把比例为 $1-v$ 的紧凑样本学习到分类边界以内。随着 v 增加,噪声点被分离出去,正样本更加紧凑,系统性能提升;当 v 继续增加时,越来

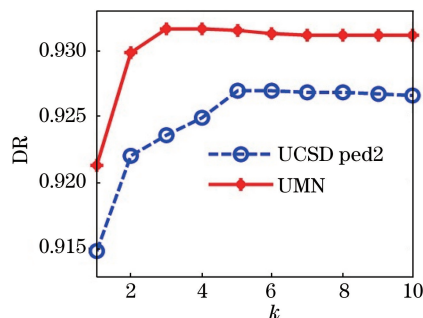


图 9 在两个数据集上,参数 k 与系统性能的关系曲线

Fig. 9 Relationship curves between the parameter k and the system performance on two datasets

越多的训练样本被视为噪声数据,当该比例超过了

实际的噪声点比例时,系统的性能开始下降。通过测试结果可以多次对惩罚因子进行调整,得到不同分类效果的分类器。实验中,测试得到两个数据集上的 $v=0.02$ 时,系统可以得到最高的检测率。

4 结 论

针对监控视频中的人群行为,提出了一种自动检测与定位异常行为的方法。提出了一种运动区域分割方法,对于与背景灰度值相近的目标仍然可以快速完整检出,降低了背景信息对后续计算的干扰,减少了后续计算量。对运动区域分割得到有效运动块,提出了局部 HOG 和局部 HOF 两种特征,联合二者得到运动块的时空描述子,较好地表征了运动块的外观与运动特性。采用先聚类再分类的思想,对运动块的特征向量建模,得到了较好的一类分类器。引入了运动连续性约束以抑制噪声干扰,大幅降低了虚警率。在 UCSD 和 UMN 数据集上的实验结果表明,相对于已有的算法,本文算法对于异常检测与定位的准确率更高。

参 考 文 献

- [1] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection [C]. 2011 IEEE Conference on Computer Vision and Pattern Recognition, 2011: 3449-3456.
- [2] 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, 2012, 42(6):1257-1272.
- [3] Hämäläinen W, Nykänen M. Efficient discovery of statistically significant association rules [C]. 8th IEEE International Conference on Data Mining, 2008:203-212.
- [4] Kuettel D, Breitenstein M D, Gool L V, *et al.* What's going on? Discovering spatio-temporal dependencies in dynamic scenes [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010:1951-1958.
- [5] Amraee S, Vafaei A, Jamshidi K, *et al.* Anomaly detection and localization in crowded scenes using connected component analysis [J]. Multimedia Tools and Applications, 2018, 77(12): 14767-14782.
- [6] Yuan Y, Feng Y, Lu X. Statistically hypothesis detector for abnormal event detection in crowded scenes [J]. IEEE Transactions on Cybernetics, 2017, 47(11): 3597-3608.
- [7] Kwon J, Lee K M. A unified framework for event summarization and rare event detection from multiple views [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1737-1750.
- [8] Lee K M. A unified framework for event summarization and rare event detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012:1266-1273.
- [9] Wang X, Hou Z Q, Yu W S, *et al.* Target scale adaptive robust tracking based on fusion of multilayer convolutional features [J]. Acta Optica Sinica, 2017, 37(11): 1115005.
王鑫, 侯志强, 余旺盛, 等. 基于多层卷积特征融合的目标尺度自适应稳健跟踪 [J]. 光学学报, 2017, 37(11): 1115005.
- [10] Zhou S, Shen W, Zeng D, *et al.* Unusual event detection in crowded scenes by trajectory analysis [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2015: 1300-1304.
- [11] Zhou P P, Ding Q H, Luo H B, *et al.* Trajectory outlier detection based on DBSCAN clustering algorithm [J]. Infrared and Laser Engineering, 2017, 46(5): 0528001.
周培培, 丁庆海, 罗海波, 等. 基于 DBSCAN 聚类算法的异常轨迹检测 [J]. 红外与激光工程, 2017, 46(5): 0528001.
- [12] Tran D, Yuan J, Forsyth D. Video event detection: from subvolume localization to spatiotemporal path search [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(2): 404-416.
- [13] Nguyen N T, Phung D Q, Venkatesh S, *et al.* Learning and detecting activities from movement trajectories using the hierarchical hidden Markov models [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2005:955-960.
- [14] Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2009: 935-942.
- [15] Lee D G, Suk H I, Lee S W. Crowd behavior representation using motion influence matrix for anomaly detection [C]. 2nd IAPR Asian Conference on Pattern Recognition, 2013: 110-114.
- [16] Chen T, Hou C, Wang Z, *et al.* Anomaly detection in crowded scenes using motion energy model [J]. Multimedia Tools and Applications, 2018, 77(11): 14137-14152.
- [17] Hu X, Hu S, Luo L, *et al.* Abnormal event detection in crowded scenes via bag-of-atomic-events-based topic model [J]. Turkish Journal of Electrical Engineering and Computer Sciences, 2016, 24(4): 2638-2653.
- [18] Reddy V, Sanderson C, Lovell B C. Improved

- anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture [C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013: 55-61.
- [19] Li W, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scenes[J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 2014, 36(1): 18-32.
- [20] Zhang Y, Lu H, Zhang L, *et al.* Combining motion and appearance cues for anomaly detection [J]. Pattern Recognition, 2016, 51: 443-452.
- [21] Sabokrou M, Fathy M, Moayed Z, *et al.* Fast and accurate detection and localization of abnormal behavior in crowded scenes[J]. Machine Vision and Applications, 2017, 28(8): 965-985.
- [22] Sabokrou M, Fayyaz M, Fathy M, *et al.* Deep-cascade: cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes [J]. IEEE Transactions on Image Processing, 2017, 26(4): 1992-2004.
- [23] Zhou S, Shen W, Zeng D, *et al.* Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes [J]. Signal Processing: Image Communication, 2016, 47: 358-368.
- [24] Bao T, Karmoshi S, Ding C, *et al.* Abnormal event detection and localization in crowded scenes based on PCANet [J]. Multimedia Tools and Applications, 2017, 76(22): 23213-23224.
- [25] Feng Y, Yuan Y, Lu X. Learning deep event models for crowd anomaly detection [J]. Neurocomputing, 2017, 219: 548-556.
- [26] UC San Diego. Statistical visual computing lab[EB/OL]. [2018-01-22]. <http://www.svcl.ucsd.edu/projects/anomaly>.
- [27] Umn anomaly dataset [EB/OL]. [2018-01-22]. <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>.
- [28] Barnich O, Droogenbroeck M V. ViBE: a powerful random technique to estimate the background in video sequences [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2009: 945-948.
- [29] Zach C, Pock T, Bischof H. A duality based approach for realtime TV- L^1 , optical flow[C]. 29th DAGM Conference on Pattern Recognition, Springer-Verlag, 2007: 214-223.
- [30] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2005:886-893.
- [31] Schölkopf B, Platt J C, Shawe-Taylor J, *et al.* Estimating the support of a high-dimensional distribution[J]. Neural Computation, 2001, 13(7): 1443.
- [32] Mahadevan V, Li W, Bhalodia V, *et al.* Anomaly detection in crowded scenes[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 1975-1981.
- [33] Helbing D, Molnar P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282.
- [34] Kim J, Grauman K. Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2009: 2921-2928.