

基于运动方向的异常行为检测

胡芝兰^{1,2} 江帆¹ 王贵锦¹ 林行刚¹ 严洪^{2,3}

摘要 提出了一种基于运动方向的异常行为检测方法. 根据不同行为的运动方向具有不同的规律性, 该方法采用块运动方向描述不同的动作, 并利用支持向量机 (Support vector machine, SVM) 对实时监控视频进行异常行为检测. 为了减少噪声运动的影响, 同时有效保留小幅度运动的前景目标, 在行为描述之前, 本文采用了背景边缘模型对每一视频帧进行前景帧 (有目标出现的视频帧) 判断. 在行为描述时, 先提取相应视频段的所有前景帧的块运动方向, 然后对这些运动方向进行归一化直方图统计得到该视频段的行为特征. 在走廊等公共场景中的实验结果表明, 该方法能够对单人以及多人的复杂行为进行有效检测, 对运动过程中目标大小的变化、光照的变化以及噪声等具有较好的鲁棒性, 而且计算复杂度小, 能够实现实时监控.

关键词 视频监控, 异常检测, 前景分割, 运动方向, 支持向量机
中图分类号 TP391

Anomaly Detection Based on Motion Direction

HU Zhi-Lan^{1,2} JIANG Fan¹ WANG Gui-Jin¹ LIN Xing-Gang¹ YAN Hong^{2,3}

Abstract A novel algorithm is proposed in this paper to detect anomalous human behaviors based on motion directions. According to different motion direction rules for different events, we introduce block-based motion directions to model those events, and use support vector machine (SVM) to detect the abnormalous actions from real-time monitoring video sequences. To increase the robustness against noise and to capture the slight movement of the object, we select the foreground frames (the frames having human object) with a background edge model before the action feature extraction. Then, action features are extracted using normalized histogram analysis from the motion directions of all the foreground frames. Our experiments on public areas such as hallway show that our algorithm is able to track complex actions of single and multiple people accurately and is robust against the variation of object size, lighting, and noise during their movements. Our algorithm is of low computation complexity thus it can be used for real time monitoring.

Key words Video surveillance, anomaly detection, foreground segmentation, motion direction, support vector machine (SVM)

近年来, 随着安全问题受到社会的日益关注, 视频监控^[1] 系统的需求正在逐渐扩大, 比如银行、商店、ATM 机等商业场所, 走廊、大厅、停车场、博物馆等公共场所, 以及交通管理系统等. 行为的自动理解和分析, 作为视频监控系统中的关键技术, 目前引起了国内外的广泛重视, 并在许多国际著名期刊和会议上进行了专门的讨论.

已有的行为分析方法大致可以分为以下三类:

1) 基于人体结构分析的方法. 该方法先采用人体模型提取每一帧视频的人体姿势, 然后利用姿势

随时间的变化来描述各种动作. 这类方法中, 有的采用 2D 人体模型^[2-4], 但是随着视角的变化, 人体的有些关节在 2D 图像中会表现为遮挡, 从而不易准确描述行为. 为了解决遮挡问题, 有人采用多摄像头得到人体 3D 姿势来描述各种行为^[5-6], 但是视频获取设备以及计算量都变得更加复杂. 通常, 这类方法需要手动初始化, 跟踪丢失后不易重新自动初始化, 而且计算复杂度高, 难以实时处理.

2) 基于时空特征的模式分析方法. 这类方法先提取每帧图像中的人体区域, 然后利用该动作的人体轮廓序列的各种形状特征对人体行为进行描述和识别. 例如, Bobick 等^[7] 通过提取运动历史图像 (Motion history image, MHI) 的 Hu 不变矩来描述人体行为. 文献 [8-9] 则提取运动序列中所有人体轮廓线所组成的时空边缘线 (Spatiotemporal contour of the human body) 的各种几何特征来分析人体动作. 这类方法依赖于每一帧图像中人体区域的准确提取, 局限于若干种简单的单人行为, 而不易扩展到多人间的复杂行为.

3) 基于图像统计处理的方法. 这类方法直接对

收稿日期 2007-09-27 收修改稿日期 2008-05-06
Received September 27, 2007; in revised form May 6, 2008
国家自然科学基金 (60472028), 高等学校博士学科点专项科研基金 (20040003015), 香港城市大学基金 (9610034) 资助
Supported by National Natural Science Foundation of China (60472028), Specialized Research Fund for the Doctoral Program of Higher Education of China (20040003015), and Grant of City University of Hong Kong (9610034)
1. 清华大学电子工程系 北京 100084 2. 香港城市大学电子工程系 香港 3. 悉尼大学电气信息学院 悉尼 00026A, 澳大利亚
1. Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China 2. Department of Electronic Engineering, City University of Hong Kong, Hong Kong, P. R. China 3. School of Electrical and Information Engineering, University of Sydney, Sydney 00026A, Australia
DOI: 10.3724/SP.J.1004.2008.01348

视频帧的低层信息进行统计分析, 从而对视频段进行理解. 比如, Ma^[10] 对视频帧进行主成分分析 (Principle component analysis, PCA), 然后对各种动作序列进行分类和识别, 但是图像的大数据量导致计算量非常大. Masoud^[11] 先对整幅图像进行 IIR 滤波得到特征图像, 然后利用这些特征图像对几个简单的单人动作进行训练和识别, 这种方法对于多人情况以及复杂行为不易处理. Au^[12] 采用基于压缩的匹配方法在原始视频流中直接检测新目标, 但是没有涉及到具体的行为理解和识别. 文献 [13–14] 与本文所提出的方法比较接近, 文献 [13] 针对家庭中病人护理的应用, 首先根据运动信息并结合图像分割以及形状模型得到运动区域, 然后利用光流信息得到每一个运动区域相对于人脸的运动方向, 从而判断出该动作的运动属性. 文献 [14] 则针对 ATM 机的监控, 先直接对一视频段的所有帧的前景分割图进行主成分分析得到该视频段的运动特征, 然后利用 1-SVM (One-class support vector machine) 对每一视频段进行异常程度检测. 由于涉及到图像分割处理^[13] 或者对大数据量的图像序列进行 PCA^[14], 因此这两种方法的计算量都较大; 同时它们所采用的行为描述特征局限于单人情况, 而不易描述多人间的行为, 所以难以扩展到走廊、广场等其他公共应用场景.

1 方法概述

针对走廊等公共场合上的视频监控系统, 本文的方法分为两个阶段: 1) 行为描述阶段, 即在原始视频中提取块运动方向对各种行为进行描述; 2) 异常行为检测阶段, 这里通过支持向量机 (Support vector machine, SVM) 训练得到异常行为检测器. 系统框图如图 1 所示.

在公共场所, 场景中的人数变化不定, 人与人之间可能存在一定的行为关系, 同时环境光照等也会随时改变, 这些都显著增加了人物分割的难度 (例如如图 2, 要准确分割出每个人非常困难). 所以, 针对单个人的监控技术^[7–11, 13] 很难推广到这些公共场所的视频监控中. 与已有的行为描述方法不同, 本文根据运动方向在同类行为中的规律性以及不同行为之间的差异性, 采用块运动方向描述并区分不同的行为, 避开了对人物位置的确定以及对人物的准确分割, 从而大大降低了算法复杂度, 同时能够有效地分析多人之间的行为. 例如图 2 (a) 为走廊上普通的多人行走, 运动方向集中在两个相向的方向上, 而图 2 (b) 则为多人站立聊天, 身体会沿各个方向做无规律的摆动, 所以利用运动方向能够有效区分这两种不同的行为.

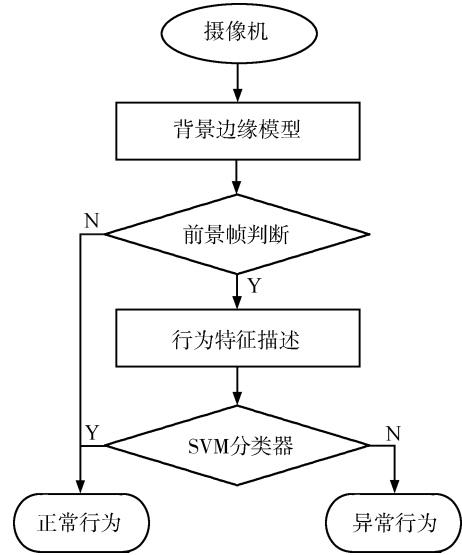


图 1 异常行为检测算法流程图
Fig. 1 Flowchart of the proposed anomaly detection algorithm



(a) 多人交叉行走
(a) Multiple people walking with face to face



(b) 四人聊天
(b) Four persons chatting together

图 2 多人在走廊上的监控视频示例

Fig. 2 Examples of multiple people in the corridor

采用运动方向描述行为时, 由于光照变化以及其他噪声等影响, 可能存在一些杂乱无章的运动, 即引入噪声运动. 为了减少噪声运动对人体行为描述的影响, 我们进行了如下处理: 首先, 与文献 [13] 不同, 我们采用基于块的运动方向来代替基于像素的

光流信息,这不仅可以滤除大部分噪声点运动,同时还大大减少了计算量;其次,背景帧(即没有目标出现的视频帧)由于光照等影响也会出现一些噪声运动块,因此在行为描述之前,我们提出了一种比较稳健的基于背景边缘模型的前景检测方法对每帧进行前景帧(即存在目标的视频帧)判断,然后用检测得到的前景帧的运动方向对该视频段的行为进行有效描述;最后,对这些块运动方向进行归一化直方图统计,进一步去除噪声运动的影响,并突出该动作的主要运动模式。

每一个阶段的具体方法如下:

在行为描述阶段,先对某一视频段的每一帧进行前景帧判断,然后提取该视频段中所有前景帧的块运动方向,最后对这些块运动方向进行归一化直方图统计,得到该视频段的行为描述特征.利用该方法得到的行为描述特征具有位置、运动速度以及目标大小不变性,对单人和多人都能进行有效描述.

在异常检测阶段,根据走廊的普遍用途,我们把正常行为定义为走廊上来回走动或者停留,除此之外视为可疑行为.本文利用 SVM 对所标定的训练样本进行训练得到异常行为检测器.对于每一监控视频,则通过该检测器进行异常行为判断.

2 前景检测

与一般的前景检测目的(即获得干净的前景区域)不同,本文的前景检测是为了判断当前帧是否为前景帧.目前,前景检测算法主要有两大类,一类是基于像素的前景检测算法,另一类是基于轮廓邻域信息的前景检测算法.基于轮廓邻域信息的算法通常用到比较复杂的理论,虽然检测结果较好,但是计算复杂度高,难以达到实时.基于像素的前景检测算法则相对简单,主要包括帧差法(Temporal differencing)^[15]和背景法(Background subtraction)^[16].帧差法对噪声比较敏感,不易检测运动缓慢或者静止的前景目标;而背景法对噪声相对鲁棒,但受光照影响大,且背景的更新与小幅度运动或者静止的前景目标检测之间的矛盾不易控制.因此,本文提出了一种新的前景检测方法,即通过背景边缘模型来检测前景边缘,从而判断是否存在前景目标.在静止背景的场景中,背景的边缘位置相对固定,对光照变化较鲁棒,所以通过建立背景边缘模型,可以有效地检测运动以及相对静止的前景目标边缘,从而实现前景帧的判断.

2.1 背景边缘模型

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得到.设当前帧为 t 帧,

每个图像位置 (i, j) 是背景边缘的概率 $P_b(i, j, t)$ 为

$$P_b(i, j, t) = \sum_{k=t-T}^{t-1} \frac{E(i, j, k)}{T}$$

$$i = 1, 2, \dots, W, j = 1, 2, \dots, H \quad (1)$$

其中, $E(\cdot, \cdot, k)$ 为第 k 帧所对应的边缘图像,该边缘图像通过 Canny 算子检测得到. H 和 W 分别为每帧图像的高和宽. T 为背景边缘模型的更新时间.对于前景目标来说,即使相对静止的行为(比如站立),同一姿势的滞留时间也不会很长,即前景目标的轮廓边缘仍处于小幅度运动中,所以 T 的取值不需很大,以便及时适应背景边缘的更新.根据实验分析,这里更新时间 T 取 1500 帧,即为 60 秒时间(25 fps).

2.2 前景帧判断

对于 t 帧,先用 Canny 算子得到该帧的边缘图像 $E(\cdot, \cdot, t)$,然后对每一边缘点 (i, j) 根据背景边缘模型判断是否为前景点,见式(2).若背景边缘概率低于阈值 TB ,该边缘点为前景边缘点,否则为背景边缘点.

$$F(i, j, t) = \begin{cases} 1, & E(i, j, t) = 1 \text{ 且 } P_b(i, j, t) < TB \\ 0, & \text{其他} \end{cases} \quad (2)$$

通常,一个前景边缘点在 T 时间内出现在同一个位置的时间比较短,而背景边缘点出现的位置则相对比较稳定,同时考虑到背景边缘会因为相机的微小抖动以及噪声的影响而出现小范围偏移,因此阈值 TB 既不能太大也不能太小.根据实验,当 TB 在 $(0.1 \sim 0.3)$ 之间时前景检测性能相当,本文 TB 取 0.2.

由于光照以及灰尘噪声等影响,前景边缘图(Foreground edge image, FEI)中会存在一定的噪声边缘点,如图 3(b)~3(d)(见下页)所示.一般地,背景帧的 FEI 只有少量噪声边缘点,且分布散,如图 3(b)所示;而在前景帧的 FEI 中,边缘点数量则明显增多,并且分布相对集中,如图 3(d)所示.根据前景边缘点的数量和分布,我们通过如下的方法对前景帧和背景帧进行有效区分:首先根据邻域信息分析每一个前景边缘点是否为噪声点,若与该前景边缘点连在一起的边缘点小于 2 个,则此点以及与其相连的点均为噪声点而去除;然后统计剩余前景边缘点的个数 (Nt) ,当 Nt 小于阈值 TN 时,判断该帧为背景帧,否则为前景帧.根据实验统计,本文 TN 取 30.

这种前景帧判断方法相对于其他常见的前景判断方法来说简单且鲁棒.图 4 给出了几种方法的比

较实例, 实验中背景法采用文献 [16] 的方法, 帧差法则简单地通过两帧灰度值相减取阈值^[15] 得到前景区域. 图 4 中, 图 4(a) 为人朝着门行走, 背景法和背景边缘法都得到较好的检测结果, 而帧差法则由于均一的衣服颜色, 只得到部分轮廓, 并且因为灯光影响存在大量的噪声; 图 4(b) 为目标走到门口停住, 此时背景法和帧差法基本检测不到静止目标, 而背景边缘法则有相对稳定的检测结果; 图 4(c) 为灯光突然熄灭, 背景法和帧差法因为光照的突变使得检测结果为整个画面, 而背景边缘法则受光线的变化影响较小. 由此可见, 对于前景检测, 背景边缘法对光线以及运动速度的变化较鲁棒, 能够较好检测

相对静止的目标.

3 行为描述

行为描述的目的是利用某些特征有效地表示某一行为, 并区别于其他不同的行为. 在走廊等公共场合, 同一场景中的人数以及个体大小千变万化, 而且同一动作发生的位置以及速度也会因人因时而变, 所以要求行为描述特征具有位置不变性以及尺寸大小不变性, 并且对人数以及运动速度的变化比较鲁棒. 本文所采用的块运动方向恰好满足上述要求. 下面给出该特征的详细提取方法和相关的讨论.

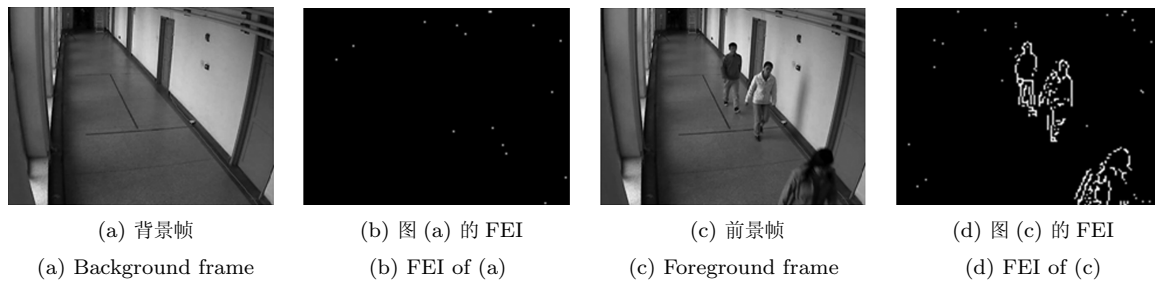


图 3 前景边缘点检测结果

Fig. 3 Foreground edge detection results

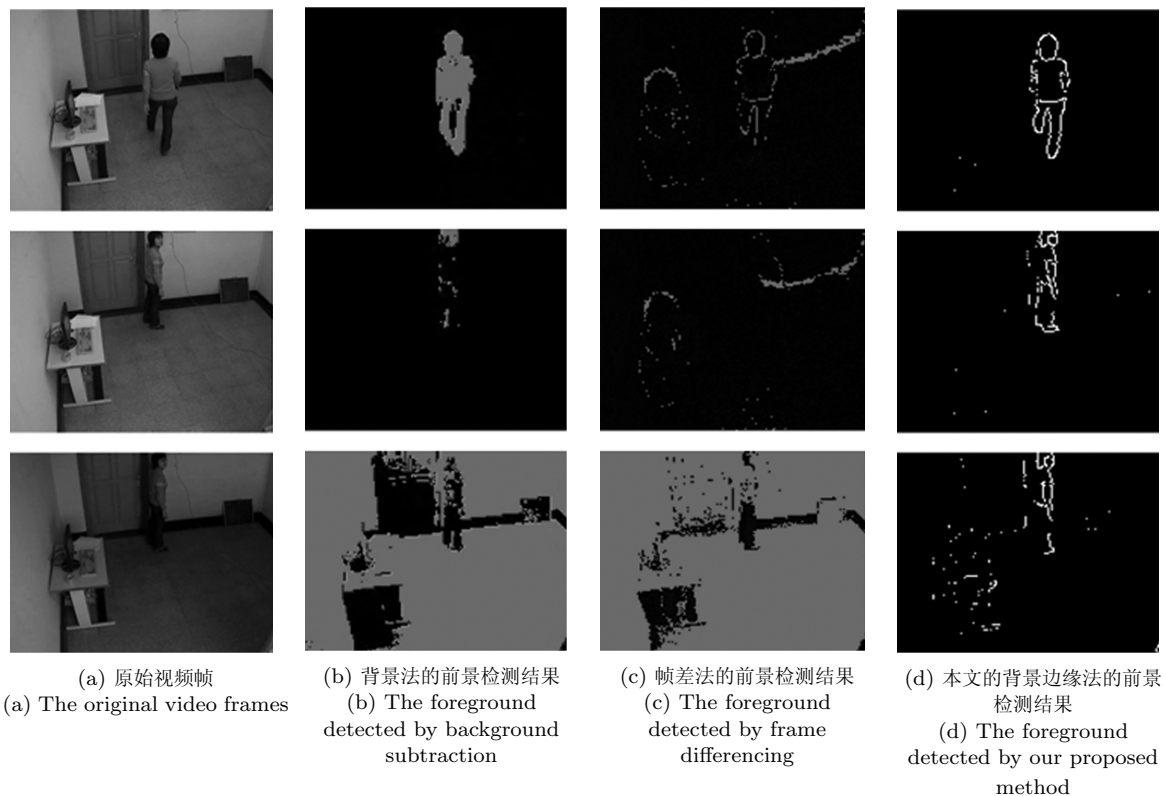


图 4 三种方法的前景检测结果

Fig. 4 Comparisons of three different foreground detection methods

3.1 视频分割

对于输入视频 V , 先将其分成 N 小段. 在理想的情况下, 每一视频段包含一个单独动作, 但复杂的行为和多变的目标导致自动视频分割非常困难. 这里我们简单地采用等间隔的交叠窗口把视频分成固定长度的视频段, 窗口长度为 4 秒 (100 帧), 相邻视频段的起始帧间隔为 10 帧. 虽然这种分割方法不能保证每一视频段对应一个完整的动作, 但是 4 秒的时间通常可以提供足够的信息来确定该视频段的主要动作属性. 比如在走廊上, 4 秒时间可以完成几步前行或一次捡东西的动作.

3.2 行为特征提取和描述

根据前面的方法得到的每一视频段的行为特征提取步骤如下.

步骤 1. 根据第 2 节介绍的前景帧判断方法得到该视频段的所有前景帧. 这里我们只对存在前景目标的视频段进行检测, 考虑到少数前景帧可能被漏检, 所以当该视频段中前景帧的比例小于 80% 时, 该段视频被认为是背景视频段, 不进行行为判断, 否则进入步骤 2, 提取该视频段的行为描述特征.

步骤 2. 提取该视频段中每一前景帧的块运动方向, 这里只考虑运动幅度不为零的块 (即运动块) 所对应的运动方向. 然后, 对该视频段的所有块运动方向做归一化直方图统计, 得到该视频段的行为描述特征. 对于每一前景帧, 运动块的选取不能太大也不能太小, 太大会使得块运动方向对目标的运动描述比较粗糙, 从而丢失一些有用的运动信息; 而如果太小则较易受到噪声的影响. 通过实验, 针对 640×480 的视频帧, 取块的大小为 8×8 . 运动矢量是在整帧图像中利用相邻两帧的信息计算得到的, 并采用菱形搜索法^[17] 有效地提高运算速度. 直方图统计过程中采用 N 个直方块 (bin), 每个 bin 的中心值在 $[0, 2\pi]$ 之间均匀求出. N 的选取将在下面讨论.

块运动方向归一化直方图可以有效地描述不同运动之间的差异. 如图 5 所示, 单人行走和多人行走的归一化直方图分布十分类似, 而打架以及聊天等异常行为的直方图分布则和正常行走有着很大的差异. 可见, 运动方向直方图可以有效地区分正常和异常行为, 同时对于个体的大小差异和人数变化比较

鲁棒.

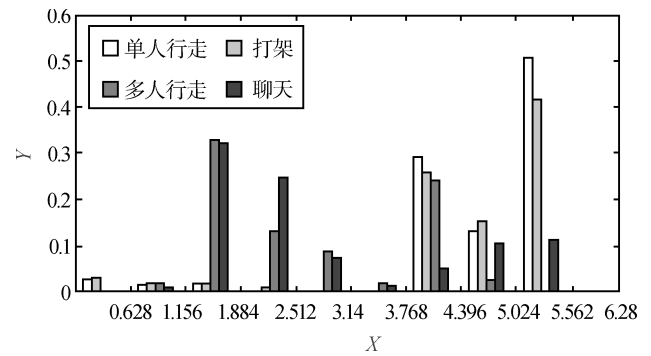


图 5 不同行为的运动方向归一化直方图 (X 轴表示每个 bin 的中心值, Y 轴表示该视频段属于对应 bin 的运动方向的比例)

Fig. 5 Motion direction normalized histograms of different activities (X is the center of each bin, Y is the motion direction proportion lying in the corresponding bin for a certain video section.)

3.2.1 关于直方图的 bin 的个数的讨论

bin 个数 N 的选取会影响特征的描述效果, bin 个数越少, 则行为描述越粗糙, 比较容易提取出不同行为之间的共同特征; bin 个数越多, 则行为描述越细致, 较易区分不同行为之间的细节. 因此需要对 bin 的个数进行合理选择, 使得描述特征能够较好地地区分正常行为和异常行为. 表 1 给出了在我们的训练集 (将在第 5.1 节介绍) 中不同 bin 的情况下所得到的检测结果. 由表 1 可见, 当 bin 的个数过少时, 正常行为和异常行为的检测率相对较低, 尤其是异常行为. 随着 bin 个数的增加, 该特征的行为表征能力增强, 当 N 取 13 时, 异常行为和正常行为的检测率都达到最高值. 但是当 bin 个数过多时, 正常行为和异常行为的检测率开始下降, 这是由于 bin 个数的增加突出了每个行为的细节, 从而限制了对正常行为和异常行为的有效描述.

3.2.2 关于 PCA 的讨论

这里没有对运动方向直方图进行 PCA 处理, 从直观上分析, 通过 PCA 处理可以进一步对行为特征向量进行降维并去除一些非主要的特征分量的影响,

表 1 不同 bin 的个数所得到的检测性能比较

Table 1 Comparison of the detection results with different bin numbers

bin 的个数 (N)	5	7	9	10	11	12	13	14	15	16	18
正常行为检测率	0.953	0.979	0.979	0.975	0.974	0.976	0.982	0.982	0.982	0.981	0.980
异常行为检测率	0.667	0.719	0.722	0.732	0.723	0.717	0.751	0.747	0.726	0.738	0.738

但在我们的方法中却没有进行这一步处理, 这是因为: 通过前景帧判断、基于块的运动方向提取以及归一化直方图处理, 运动噪声的影响已经很小, 同时, 我们采用的行为特征向量只要 13 维, 数据量低. 表 2 给出了经过 PCA 处理后采用不同数目的主分量描述行为特征以及直接利用原始行为特征向量在我们的训练集中所得到的检测率. 通过该实验可以发现, 经过 PCA 处理后的检测性能要略低于原始的行为特征向量. 这是因为在噪声影响较小的情况下, 原始数据比较丰富地保留了各类行为的特征. 其次, PCA 处理会增加算法的复杂度, 所以本文采用原始行为特征描述各种行为.

4 SVM 分类器训练

在诸多公共场合中, 虽然人的行为可以丰富多变, 但是常见行为相对比较一致, 比如在走廊上通常是行人来回走动, 运动方向集中, 而其他行为 (如打架等) 的运动方向则不同于正常的行走, 所以可以对运动方向较集中的正常行为和较分散的异常行为进行训练, 从而得到异常行为检测器. SVM 是一种通过求解一个线性约束的二次规划, 构造出一个超平面作为决策平面, 使得特征空间中两类模式之间的距离最大, 并保证得到的解为全局最优解的方法. 它在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势, 已经在模式识别、函数逼近和概率密度估计等方面取得了良好的效果^[18]. 因此, 本文采用 SVM 进行异常行为检测器训练, 其中核函数为径向基核函数 (Radial basis function), 即

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (3)$$

其中参数 γ 为形状参数, Vapnik^[19] 的研究表明该参数影响样本数据在高维特征空间中分布的复杂程度. 在监控视频中, 正常行为比较容易得到, 而异常行为则较少且情况多变, 所以在确定参数 γ 时, 我们在保证异常行为检测率较高的情况下, 尽量降低正常行为的误判率. 按照这一原则, γ 可通过实验确定, 并将第 5.1 节中详细介绍.

5 实验结果

本文的实验数据为走廊上的监控视频. 该视频中行为、光照、人数和个体大小都有比较大的变化, 每一帧图像大小为 640×480 . 本算法在 Intel PD 2.8 GHz CPU 上可以达到实时处理 (25 fps).

5.1 异常行为检测器训练

在训练集中, 正面样本为走廊上的普通行走, 包括单人和多人, 沿着同方向或反方向行走等, 如图 6 所示, 共 1081 个视频段; 负面样本包括打架、聊天、病倒、在门附近徘徊等异常动作, 如图 7 (见下页) 所示, 共有 321 个视频段. 测试集为 624 个正面样本, 53 个负面样本.

在进行训练之前, 先按照第 3 节的方法提取每一训练样本的行为特征, 然后利用这些样本选择合适参数 γ . 根据第 4 节介绍的原则 (即在保证异常行为检测率较高的情况下, 尽量降低正常行为的误判率), 图 8 给出了不同 γ 下训练集和测试集中正面

表 2 经过 PCA 处理和采用原始行为特征所得到的检测性能比较

Table 2 Comparison of detection results by PCA-based action features and the original action features

主成分个数	基于 PCA										原始特征
	4	5	6	7	8	9	10	11	12	13	
正常行为检测率	0.960	0.974	0.974	0.981	0.979	0.980	0.980	0.982	0.982	0.982	0.982
异常行为检测率	0.732	0.723	0.735	0.735	0.741	0.750	0.750	0.751	0.751	0.751	0.751

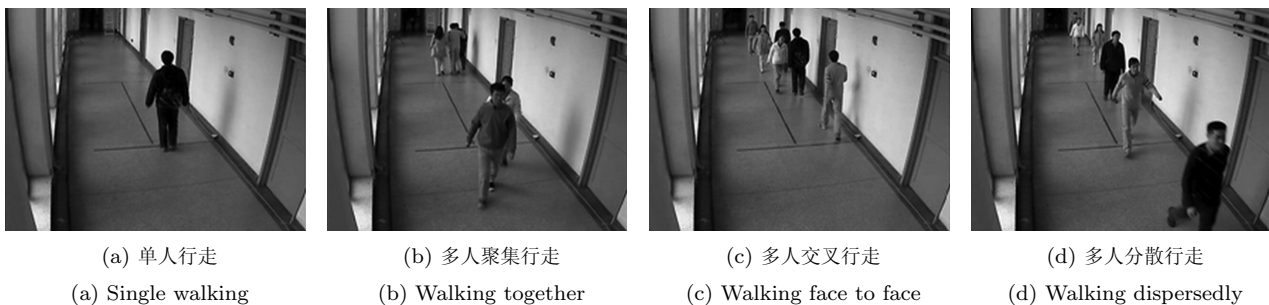


图 6 走廊上的正常行为示例

Fig. 6 Examples of normal activities in the corridor

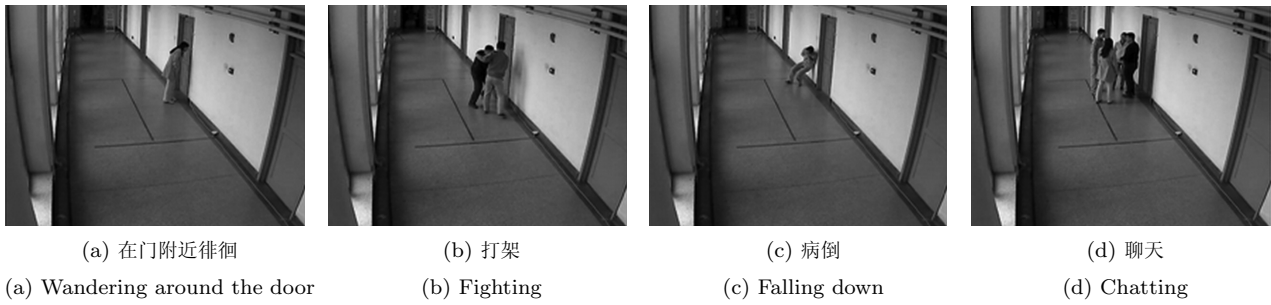


图 7 走廊上的异常行为示例

Fig. 7 Some anomalies in the corridor

样本的误检率以及负面样本的检测率. 由图 8 可见, 训练集中正常行为误判率比较稳定, 随着 γ 的增加略有下降, 但是测试集中的正常行为在 γ 较小时误判率高; 随着 γ 的增加, 训练集中的异常行为检测率有所提高, 但测试集中的负面样本检测率有所下降.

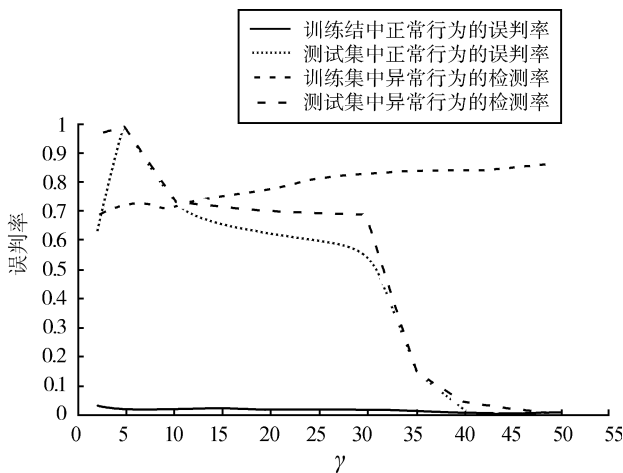


图 8 异常行为检测率和正常行为误判率随 γ 的变化曲线
Fig. 8 Anomaly detection rate curve and normal action false detection rate curve under different γ

折中考虑检测器在训练样本中的检测性能以及在测试集中的扩展性能, 本实验中 γ 取 30. 当 $\gamma = 30$ 时, 训练集中的异常行为检测率较高, 在测试样本中扩展性好, 训练集和测试集中的正面样本都具有较低的误判率. 虽然测试集的正面样本误判率比

较高, 但是在实际情况下, 这些误判的样本零散地分布在视频中, 很少连续出现, 所以根据本文后面提出的异常行为判别准则, 在实际中很少会引起正常行为的误检.

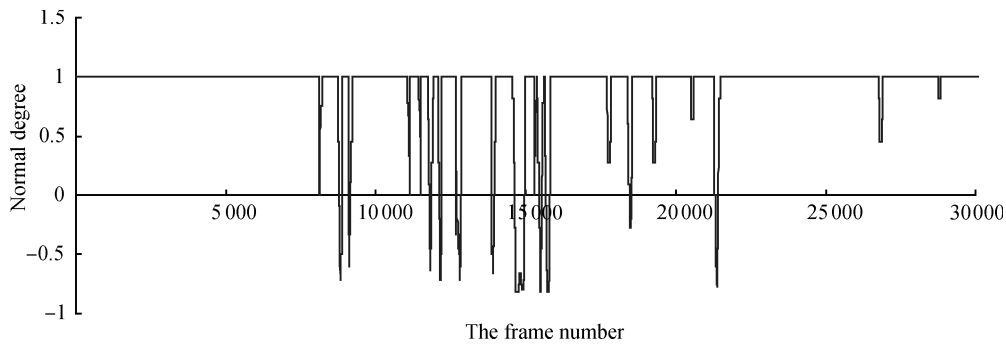
5.2 异常行为检测结果及分析

若视频段属正常行为, 本文的异常行为检测器输出值为 1, 属异常行为时则输出值为 -1. 用每一帧所在的所有视频段输出值的均值表征该帧的正常程度, 以走廊上的不同时间段的监控视频为测试数据, 图 9(a) (见下页) 给出了部分视频的检测结果. 当某一部分视频流中出现两个以上的连续异常视频段, 即某一帧的输出值小于 0.5 的时候, 我们认为该视频流中存在异常行为, 这样可以减少正常行为中由于光照等噪声所引起的误判, 同时有效保留异常行为. 图 6 和图 7 的检测结果见表 3, 其中图 6 中的正常行为的输出值都为 1, 而异常行为则都小于 0.5, 即本异常行为检测器可以正确区分正常行为和异常行为. 在我们的测试视频 (不同于第 5.1 节的测试集) 中, 除了大量的正常行走 (共 300 个独立的视频流), 共模拟了 78 个异常动作, 包括打架、聊天、病倒、放包、取包、在门附近徘徊以及进出门等, 每个动作的持续时间都不一样, 比如聊天持续时间可以长达半分钟, 而取包则只需 2、3 秒. 利用前面得到的异常行为检测器, 按照前面的判别准则, 异常行为的检测率为 84.6%, 正常行为的误判率为 1.3%. 在漏检的异常动作中, 大部分为取/放包动作, 这是由于取/

表 3 采用不同的方法对图 6 和图 7 的检测结果 (输出值为该行为的正常程度)

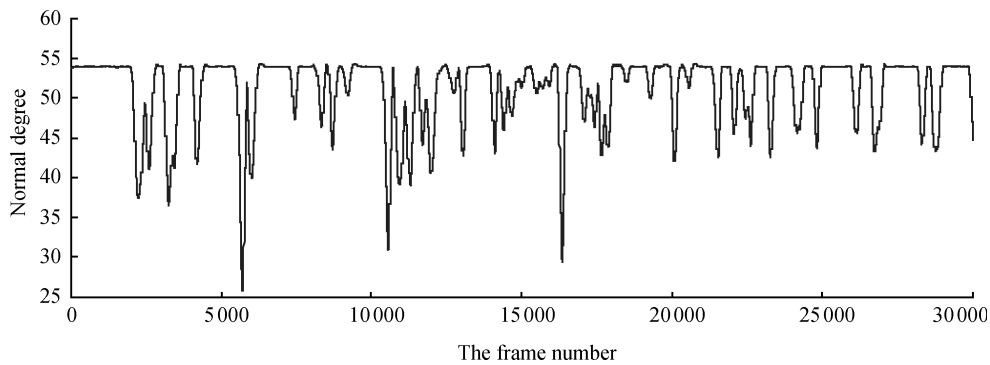
Table 3 Detection results of Fig. 6 and Fig. 7 with different methods (The output is the normal degree of the action.)

方法	图 6 (正常行为)				图 7 (异常行为)			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
本文的方法	1	1	1	1	-0.14	0	-0.33	-0.2
文献 [14] 的方法	50.41	43.47	48.58	29.82	51.48	49.48	53.03	48.58



(a) 采用本文方法得到的异常检测结果

(a) The anomaly detection results by our method



(b) 采用文献 [14] 方法得到的异常检测结果

(b) Anomaly detection result by [14]

图 9 两种不同方法的检测结果比较

Fig. 9 Detection result comparison of two different methods

放包的时候不需要人特别弯曲和停留, 因此速度很快 (一般 60 帧就能完成一个动作), 并且和普通的行走差别很小, 如图 10 (a) (见下页) 所示; 另一个则为快速病倒的动作, 该动作为先站立后蹲下, 速度很快 (30 帧左右完成该动作), 而且上下着装颜色一致, 所以不容易被检测, 如图 10 (b) 所示. 在所有正常行为中, 被误检的动作是因为行人在走路的过程中进行了比较夸张的肢体运动, 比如单个目标在行走的同时双手慢慢从身体两侧上举, 这和一般的行走有着明显的差别. 由检测结果可见, 本文方法能够有效检测复杂场景中的异常行为, 这是由于我们采用的块运动方向行为特征所描述的是一段视频中运动目标的运动方向的总体分布, 所以与目标的绝对位置没有关系. 同时, 经过归一化直方图处理, 对目标数目变化以及个体变化都比较鲁棒; 基于块的运动方向统计以及前景帧判断, 有效地去除了噪声运动的干扰.

此外, 我们还将本文的异常行为检测算法和与本文较类似的文献 [13–14] 的方法进行了比较. 文献 [14] 直接对一个视频段的所有帧的运动目标分割结果进行主成分分析, 从而提取出该视频段的运动特征, 然后利用 1-SVM 对每一视频段进行异常程度

检测. 图 9 (b) 给出部分走廊监控视频的检测结果, 图 6 和图 7 的图例所对应的检测结果见表 3. 根据实验结果可见, 正常度的最大输出值对应于背景帧, 输出值较小的对应于运动目标面积较大的前景帧, 即检测结果主要取决于前景目标的大小, 而无法正确区分走廊上异常行为和正常行为 (从表 3 的结果对应于图 6 和图 7 可以更加直观地说明这一点). 这主要是因为文献 [14] 中所采用的行为描述特征利用了每一帧的前景目标信息, 包括目标的位置、个数、运动轨迹等, 同时没有进行前景目标的归一化处理, 所以该行为描述特征对目标的位置、个数、运动轨迹以及在运动过程中目标大小的变化不鲁棒, 故只适用于单个目标、行为轨迹单一的情况 (如文献 [14] 所提到的 ATM 机场景, 因为在 ATM 机中, 正常行为一般为一个人进 ATM 机取款, 取好后离开的过程), 而不能对走廊等复杂的公共场合进行有效监控.

与本文类似, 文献 [13] 也采用了运动信息. 该方法首先根据运动信息并结合图像分割以及形状模型得到运动区域, 然后利用光流信息得到每一个运动区域相对于人脸的运动方向, 从而判断出该动作的运动属性. 该方法首先需要自动定位人脸的位置. 在诸多公共场合中 (如图 7), 由于人体之间的相互遮挡



(a) 放包动作全过程示例 (从左到右所对应的帧数分别为: 37 543, 37 575, 37 586)

(a) Examples of putting a bag (The corresponding frames in the video from left to right are: 37 543, 37 575, 37 586.)



(b) 病倒动作全过程示例 (从左到右所对应的帧数分别为: 35 460, 3 475, 35 490)

(b) Examples of falling down (The corresponding frames in the video from left to right are: 35 460, 3 475, 35 490.)

图 10 漏检的异常行为示例

Fig. 10 Examples of the missing anomalies

以及人脸朝向的不规则, 自动人头定位是十分困难的. 采用现有的比较成熟且普遍在科研中使用的技术^[20]对图 7 进行人脸定位, 无法准确检测到任何一张人脸. 其次, 文献 [13] 要求有效确定人体的各个部位 (手、胳膊、腿等), 以便通过各个部位相对人脸的运动情况来判别行为. 但是, 在二维图像中, 由于人体自身的内部遮挡 (如图 7(a)) 以及人体相互之间的遮挡 (如图 7(b)), 几乎不可能准确确定人体的各个部位. 所以, 在这种人多、情况复杂的公共场合, 文献 [13] 的方法并不适用.

然而本文所采用的行为描述特征对目标的出现位置、大小以及个数比较鲁棒, 不需要对人体进行分割以及对人体各个部位进行确定, 所以可以有效地描述各种复杂情况下的不同行为, 从而可以检测出异常行为, 对应的实验结果见图 9(a) 和表 3.

6 结论

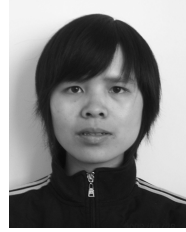
本文提出了一种基于块运动方向的行为描述特征, 该特征具有位置、大小不变性, 而且对噪声、人数、运动速度以及个体大小变化比较鲁棒. 在行为描述之前, 通过背景边缘法先对每一帧进行前景帧判断, 从而去除背景帧中噪声运动对行为描述产生的影响. 同时, 由于本文中的行为描述特征提取方法计算复杂度低, 利用 SVM 对监控视频可以实时有效地实现异常行为检测. 但是, 仅仅使用运动方向特征不易进行更具体的行为分析, 在以后的工作中, 我们

将进一步结合其他特征完善该算法, 并把该算法推广到更广泛的公共场合的监控视频中.

References

- Collins R T, Lipton A J, Kanade T. Introduction to the special section on video surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 745–746
- Hogg D. Model-based vision: a program to see a walking person. *Image and Vision Computing*, 1983, **1**(1): 5–20
- Cham T J, Rehg J M. A multiple hypothesis approach to figure tracking. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Ford Collins, USA: IEEE, 1999. 239–244
- Yilmaz A, Shah S. Recognizing human actions in videos acquired by uncalibrated moving cameras. In: *Proceedings of the 10th IEEE International Conference on Computer Vision*. Beijing, China: IEEE, 2005. 150–157
- Cheung G K M, Kanade T, Bouguet J Y, Holler M. A real time system for robust 3D voxel reconstruction of human motions. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Hilton Head Island, USA: IEEE, 2000. 714–720
- Lu C M, Ferrier N J. Repetitive motion analysis: segmentation and event classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(2): 258–263
- Bobick A F, Davis J W. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(3): 257–267

- 8 Weinland D, Ronfard R, Boyer E. Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, 2006, **104**(2): 249–257
- 9 Yilmaz A, Shah M. Actions sketch: a novel action representation. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 984–989
- 10 Ma G Y, Lin X Y. Typical sequences extraction and recognition. *Lecture Notes in Computer Vision*, 2004, **3058**: 60–71
- 11 Masoud O, Papanikolopoulos N. A method for human action recognition. *Image and Vision Computing*, 2003, **21**(8): 729–743
- 12 Au C E, Scaff S, Clark J J. Anomaly detection for video surveillance application. In: Proceedings of IEEE International Conference on Pattern Recognition. Hong Kong, China: IEEE, 2006. 888–891
- 13 Gao J, Collins R T, Hauptmann A G, Wactlar H D. Articulated motion modeling for activity analysis. In: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop. Washington D. C., USA: 2004. 20
- 14 Sudo K, Osawa T, Wakabayashi K, Koike H. Detecting the degree of anomaly in security video sequences. In: Proceedings of Signal and Image Processing. Tokyo, Japan: ACTA Press, 2006. 3–7
- 15 Anderson C, Burt P, van der W G. Change detection and tracking using pyramid transformation techniques. In: Proceedings of Intelligent Robots and Computer Vision. Cambridge, England: SPIE, 1985. 72–78
- 16 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Ford Collins, USA: IEEE, 1999. 246–252
- 17 Tham J Y, Ranganath S, Ranganth M, Kassim A A. A novel unrestricted center-biased diamond search algorithm for block motion estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 1998, **8**(4): 369–377
- 18 Bian Zhao-Qi, Zhang Xue-Gong. *Pattern Recognition*. Beijing: Tsinghua University Press, 2002
(边肇祺, 张学工. 模式识别. 北京: 清华大学出版社, 2002)
- 19 Vapnik V N. *Statistical Learning Theory*. New York: John Wiley and Sons, 1998
- 20 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Hawaii, USA: IEEE, 2001. 511–518



胡芝兰 博士研究生. 主要研究方向为计算机视觉和视频监控. 本文通信作者.
E-mail:
huzhilan00@mails.tsinghua.edu.cn
(**HU Zhi-Lan** Ph. D. candidate. Her research interest covers computer vision and video surveillance. Corresponding author of this paper.)



江帆 博士研究生. 主要研究方向为计算机视觉, 模式识别和视频监控.
E-mail:
jiang_f99@mails.tsinghua.edu.cn
(**JIANG Fan** Ph. D. candidate. His research interest covers computer vision, pattern recognition, and video surveillance.)



王贵锦 清华大学电子工程系副教授. 主要研究方向为视频处理与通信.
E-mail: wangguijin@tsinghua.edu.cn
(**WANG Gui-Jin** Associate professor in the Department of Electronic Engineering, Tsinghua University. His research interest covers video processing and communication.)



林行刚 清华大学电子工程系教授. 主要研究方向为图像/视频处理与分析, 视觉信息压缩编码及应用, 模式识别.
E-mail: xglin@tsinghua.edu.cn
(**LIN Xing-Gang** Professor in the Department of Electronic Engineering, Tsinghua University. His research interest covers image/video processing and analysis, video coding and communications, and pattern recognition.)



严洪 香港城市大学、悉尼大学教授. 主要研究方向为模式识别, 计算机视觉, 生物信息学.
E-mail: h.yan@cityu.edu.hk
(**YAN Hong** Professor at City University of Hong Kong and University of Sydney, Australia. His research interest covers pattern recognition, computer vision, and biology information.)