

Video Abnormal Event Detection Based on Improved Dynamic Clustering

Liqiming Liu^{1,2}, Xianghua Xv^{1,2*}, Lingjun Zhang^{1,2}

¹School of Computer Science and Technology, HangZhou DianZi University, Hangzhou Zhejiang

²Zhejiang Provincial Key Laboratory of Data Storage and Transmission Technology, Hangzhou Zhejiang

Email: 152050072@hdu.edu.cn, *xhxu@hdu.edu.cn

Received: May 2nd, 2018; accepted: May 21st, 2018; published: May 28th, 2018

Abstract

Abnormal event detection is an important part of intelligent surveillance systems, especially for complex surveillance video scenes. In recent years, many algorithms have been proposed to detect abnormal events. However, most of them need to set a series of parameters in the model during the modeling process, which is not only troublesome in arranging the parameters, but also the parameters need to be reset when changing the detect scene. This paper proposes an abnormal detection algorithm based on non-parametric models, constructs and maintains a vector set based on the motion trend vector merging method, and uses clustering to generate different event clusters, and proposes a pre-detection step to improve the detection effect of the algorithm in sparse scenes. Finally, some existing detection algorithms are selected for comparison experiments. The results show that the model proposed in this paper has certain advantages in detection rate and time performance.

Keywords

Abnormal Event Detection, Monitoring Video, Vector Merge, Dynamic Clustering

基于改进的动态聚类的 视频异常事件检测

刘李启明^{1,2}, 徐向华^{1,2*}, 张灵均^{1,2}

¹杭州电子科技大学, 计算机学院, 浙江 杭州

²浙江省数据存储传输技术重点实验室, 浙江 杭州

Email: 152050072@hdu.edu.cn, *xhxu@hdu.edu.cn

收稿日期: 2018年5月2日; 录用日期: 2018年5月21日; 发布日期: 2018年5月28日

*通讯作者。

摘要

异常事件检测是智能监控系统中的重要一环，尤其是对复杂的监控视频场景而言。近年来研究者提出了很多的算法来检测视频中的异常事件，然而它们中的大多数都需要在建模过程中给模型设定一系列的参数，这样不仅调参麻烦，而且在更换检测场景时需要重新设定参数。本文提出了一个基于非参数模型的异常检测算法，通过基于运动趋势的向量合并方法来构造并维护一个向量集合，并运用聚类生成出不同的事件簇，同时提出了一个预检测步骤以此来提高算法在稀疏场景下的检测效果。本文选取了一些已有的检测算法进行了对比实验，最后的实验结果表明，本文提出的检测模型在检测率以及时间性能上均有一定的优势。

关键词

异常事件检测，监控视频，向量合并，动态聚类

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着社会的不断发展，人们对公共安全的需求变得越来越大，为了方便管理，如今越来越多的公共场合部署了视频监控设备，而监控系统也逐渐成为维护公共安全不可缺少的一环。然而一般意义下的异常事件拥有突发性等特点，为了能够在第一时间了解到异常事件的情况，通常需要花费较多的人力来长时间地盯着监控屏幕来进行观察[1]，这样不仅会使人产生疲劳，而且容易疏忽掉一些不显眼的事件。所以，复杂场景下的监控视频异常事件自动检测识别作为智能视频监控系统的核心课题，正在受到越来越多研究者的关注，然而由于光照变化、相互遮挡、重叠、无规律运动[2]以及异常事件定义不清晰[3]等的影响，使得复杂场景下的异常事件检测成为一个极具挑战性的问题。

对于复杂场景下的异常事件自动检测，如今已经有一些效果较好的方法被提出，比如 Feng [4]等人将监控视频中那些目前为止没有出现过的、或者是出现频率很低的运动事件定义为异常事件，并提出了一个两层的聚类模型，通过对运动特征向量的聚类分析来对视频中的运动事件进行分析判断，以此来进行异常事件检测。

然而本文在对该方法进行实验的过程中发现，当监控视频场景中存在正常事件和异常事件的运动方向一致、而运动速度存在差异时，算法的检测效果不理想，异常事件经常会被漏检，如图 1 第一行所示，该场景当前为止未出现过自行车，而算法也始终未能将其检测出来；同时该算法在稀疏监控场景下随着检测的进行，出现误检的概率会逐步地变高。本文针对上述的两个问题，分析了问题出现的原因，并提出了针对性的改进。最后的实验结果表明，本文提出的改进使得算法在检测率以及时间效率上均有了较大的提升。本文的主要工作如下：

- 对原算法在某些特定场景下检测出现偏差的原因进行了实验分析；
- 基于对问题的实验分析，分别采用基于运动趋势合并的方法来维护大小固定的字典集合以及预检测的方法来排除背景特征对检测的干扰。



Figure 1. Detection result in corresponding scenes
图 1. 对应场景检测结果

本文剩余部分的内容组织如下：第二章为相关工作，主要展示了一些已有的检测方法并将其分为两类来进行了介绍；第三章则对出现的问题进行了详细的分析；第四章对于检测效果不理想的场景，根据实验分析提出了针对性的检测模型；第五章展示实验分析，对提出的算法进行了实验评估；总结则位于最后的第六章。

2. 相关工作

对于视频当中异常事件的自动检测识别，针对不同的视频场景以及异常事件定义，已经有大量检测算法被提出并应用到实际的场景当中。对于已有的方法，按照算法的思想大致上可以分为：基于追踪的和基于模型的两大类。

其中，基于追踪的方法属于一般性的方法。在该模型下，算法对视频当中的运动物体分别进行追踪，形成连续不间断地运动轨迹[5] [6] [7] [8]，并用一条或多条轨迹来代表其中的某个运动物体。Shu *et al.* [9] 对于无人机航拍得到的视频从中提取出大量轨迹特征，用与或图来对事件类型进行分类，并运用马尔柯夫链及动态规划来识别轨迹特征段所代表的具体事件类型；Cui *et al.* [10]则是对时空兴趣点进行追踪来得到相应的运动轨迹，通过计算不同轨迹之间的相互作用能来描述与目标对应行为之间的关系，并运用支持向量机(Support Vector Machines, SVM)来对轨迹特征进行分类来检测异常事件；而在[11]中，Wu 等则是从视频人群中计算出光流质点轨迹并用概率模型进行建模，最后运用极大似然估计来判断异常。

基于追踪的方法一般情况下都是通过追踪视频当中每一个运动物体的轨迹来实现的，在该模型下，异常事件的检测是通过对提取出来的轨迹特征进行比较计算来识别，由于轨迹只考虑运动物体的前景信息，从而该方法能够尽可能地降低背景噪声对检测造成的影响。但是另一方面，对于拥挤的视频场景而言，场景当中同一时间通常包含有较多的运动物体，这种情况下同时追踪视频当中所有运动物体的运动轨迹是十分困难的[12]：不仅计算量十分巨大[13]，会降低算法对于异常事件的检测速度；而且由于场景中有较多的运动物体，那么运动物体之间出现重叠、交错的情况[14] [15]也就在所难免，而这种情况下的轨迹可能无法被正确追踪，导致轨迹前后所描述的并非是一个运动物体。这就使得基于追踪的方法对于拥挤场景而言存在诸多挑战，不太适用。

综上所述，轨迹特征在复杂场景中的描述能力较差，为了解决基于追踪的方法在拥挤场景当中所面临的问题，基于模型的方法通常情况下用其他更加一般的特征来对事件进行描述[16]-[21]，通过对特征向

量的计算处理来判断异常事件的出现。Zhao *et al.* [22]将方向梯度直方图或光流直方图(Histogram of Optical Flow, HoF)特征描述子运用稀疏编码的方式计算出一组相互正交的基向量,运用正交基向量的稀疏重构来进行运动特征的异常判断;而 Lu *et al.* [23]则将稀疏编码检测进行了改进,将正交基向量组拆分成一系列小的向量组,并运用这些向量组分别计算各自的重构误差,最后通过所有误差联合判别的方式来进行场景异常事件的判断;Li *et al.* [24]则是针对 HEVC (High Efficiency Video Coding)视频压缩域的特征提出了一种特定的运动特征描述方法,并通过对该特征进行线性预测的方式来判断异常。同时,机器学习算法也被大量运用在模型的异常检测当中, Wang [25]等人则将 one-class SVM 以及核化主成分分析应用其中,通过正常的训练数据集来训练分类器,然后通过训练好的分类器来对检测视频进行异常事件检测;而在[26]中作者将视频图像当中的显著性信息与 HoF 特征相结合,通过简化的卷积神经网络来提取出高层运动特征,并运用 SVM 来判断高层特征所描述的运动是否属于异常。

3. 问题分析

在[4]中,作者对于视频当中异常事件的定义,是在视频当中从来没有出现过的事件、或者是那些出现频次很低的运动特征。为此通过构造一个两层聚类模型来对 HOF 运动特征进行归类判断:对于模型输入的运动特征,首先通过向量合并的方法来维护一个大小固定的向量集合;然后对集合当中的向量进行聚类,通过计算特征向量与各个聚类中心之间的误差来判断输入的运动特征是否为异常事件特征。然而根据图 1 的检测结果显示,原算法在前述的一些视频场景中的检测结果不理想,为此接下来本文针对出现的检测结果进行分析。

3.1. 同向异速漏检

图 1 上面一行的检测结果显示,视频当中的汽车虽然被正常检出,但是对于自行车这个异常事件算法全程都没有检测出来,为此本文在相关场景下做了验证性实验,如图 2 所示,算法同样无法检测出运动方向与行人一致的这辆自行车。对于这种在正常事件和异常事件的运动方向一致、而运动速度存在差异时算法造成异常事件漏检这一现象,本文经过分析之后发现这种漏检的出现是由于算法自身的模型设计缺陷所造成的:对于每一个大小为 $n \times m \times t$ 像素的视频块(grid),算法提取特征时计算 grid 中每一个像素点的光流信息 $(v_{x_j}, v_{y_j}), j \in \text{grid}$, 根据像素点的光流值计算出一个均值权重 HoF 特征描述子 $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$

$$v_i = \bar{M} \times \sum_{j \in \text{grid}} g(D_j), i = 1, 2, \dots, n \quad (1)$$

$$\begin{cases} \bar{M} = \frac{1}{nmt} \sum_{j \in \text{grid}} \sqrt{v_{x_j}^2 + v_{y_j}^2} \\ D_j = \arctan\left(\frac{v_{y_j}}{v_{x_j}}\right) \\ g(D_j) = \begin{cases} 1, & D_j \in \left[\frac{2\pi i}{n}, \frac{2\pi i}{n} + 1\right) \\ 0, & \text{else} \end{cases} \end{cases} \quad (2)$$

由上述(1)、(2)式可知,物体的运动方向影响的是 D 的取值,而 D 又只与直方图当中的投影区间有关,所以在物体运动方向不变的前提下,AMHOF 特征的不同维度的取值是相差不大的;而物体的运动速度不同,则统计之后得到的光流平均大小 \bar{M} 不同,在 D 相同的前提下,得到的两个特征向量在不同维度之间大致上是呈一种等比例关系的。

算法模型在输入特征向量 \mathbf{v} 之后，由于需要维护向量集合的大小为固定值 N ，所以选取欧氏距离最近的两个向量进行合并，用合并之后的新向量来表示原来两个向量所代表的运动信息

$$\begin{cases} \mathbf{v} = (1 - \alpha) \times \mathbf{v}_a + \alpha \times \mathbf{v}_b \\ r(\mathbf{v}) = r(\mathbf{v}_a) + r(\mathbf{v}_b) \end{cases} \quad (3)$$

在论文描述的模型当中，向量的更新由(3)式所示，只是简单地将两个向量通过学习率 α 进行加法运算。在算法检测过程中，刚开始一系列正常的行人特征会被添加进向量集合当中，而在接下来的检测过程中为了维持集合大小 N 不变，会不停地进行向量合并，在经过一段时间的向量合并之后，向量各个维度的值都要比原先的特征向量的值大，导致向量所代表的运动信息发生偏移；而异常事件自行车又恰好与行人的运动方向大致相同，只是在运动速度上存在差异，根据前面的分析，两者的特征向量之间大致上呈现的也是等比关系。所以可以发现当视频中出现自行车这个异常事件时，此时集合中的向量经过一系列合并之后，每一个维度的值相比于行人都要高，集合中的向量与描述自行车运动的特征向量之间的相似程度更高，无法识别出两者是不同的运动，从而导致了漏检。

3.2. 稀疏场景误检

原方法在稀疏场景下的检测也不怎么好，如图 1 第二行所示，随着检测的进行算法会出现越来越多的误检，即算法将那些原本属于正常事件范畴的运动识别为了出现频次很低的运动，然后根据文章对异常的定义将其划归为异常事件。原方法通过对集合向量进行计数的方式来表示向量所描述的运动信息在视频中出现的次数，用 $r(\mathbf{v})$ 表示；而进行向量合并时，由于合并之后的新向量同时包含有原来两个向量的运动信息，所以在(3)式中合并之后新向量的计数值为原来两个向量之和；接着在对集合中的向量进行聚类的时候，论文中对于每一个类 C_i ，定义该类的频数 $r(C_i)$ 及占比 $w(C_i)$

$$\begin{cases} r(C_i) = \sum_{j \in C_i} r(\mathbf{v}_j) \\ w(C_i) = \frac{r(C_i)}{\sum_{i=1}^K \sum_{j \in C_i} r(\mathbf{v}_j)} \end{cases} \quad (4)$$

其中 $r(C_i)$ 为属于该类的运动特征出现的总次数，由所有属于该聚类的向量的计数值之和得到；而 $w(C_i)$ 则表示为第 C_i 类所描述的运动出现的概率。视频中的主流运动特征指的是在视频中经常出现的运动信息 [27]，那么当主流运动出现次数增多之后，在集合中表示这一类运动的向量的计数值也会随之变大，聚类之后相应的聚类中心的频数 $r(C_i)$ 也会增加，对应的 $w(C_i)$ 也会变大。

本文在分析出现低频误检的原因时发现，当视频中主流运动为正常的行人运动时，算法对于低频率事件能够正常检测；而当场景变为稀疏场景时，视频中的主流特征则转变为了背景特征，此时再经过长时间的异常检测之后，集合中向量表示的大部分信息就变为了背景信息，即在聚类之后表示背景的那一类的 $w(C_i)$ 值变得过大而其他类所占比重则较低，当 $w(C_i) < w_{th}$ 时，属于 C_i 类的特征向量就会被视为低频异常从而造成误检。

4. 模型改进

根据前述的分析，[4]当中所提出的检测方法在异常事件与正常事件的运动方向基本一致、只在运动速度上存在差异该种情形下导致检测效果不理想，是由于原方法在合并时采用的是加法运算，使得向量的运动信息发生了偏移；而对于低频误检的出现，则是由于在稀疏场景当中背景特征变成了主流特征，导致聚类时代表背景特征的那一类所占的信息比重偏高。对于这两种检测率较低的场景，本文根据第三节分析并结合原算法提出一个新的两层检测模型，使得在上述两种视频场景中的检测率得到提高。

4.1. 预处理及特征提取

低频误检出现的主要原因就是从视频流提取出来的特征向量中表示背景信息的向量占多数，导致在检测过程中信息分布地过于集中，导致其中一个类 C_i 的 $w(C_i)$ 值过高。为此，本文对于输入的视频流，首先判断当前帧是否存在活动区域，活动区域分布在视频帧的哪个部分，在确定活动区域之后再对视频帧进行异常检测，在检测过程中对无用的背景信息进行剔除，只留下运动特征来进行检测判断。

同时，为了能够定位异常事件在视频当中的位置，对于输入的视频流，算法将其分割成为非重叠的 grid，从每个 grid 中提取一个特征向量来表示该 grid 中所包含的运动信息，这里本文沿用原方法中的 AMHOF 特征，如(1)、(2)式所示。另一方面在光流的定义中，若像素点的光流向量取值大则说明该像素点在前后帧之间的运动幅度大，所以对于待检测的视频流，本文通过在提取 AMHOF 特征时计算出的像素点的光流向量信息来进行运动物体的判断。在稀疏场景当中，运动物体在视频中不占主流，且在运动过程中很大程度上也不会占满整个画面，为了尽可能地去掉背景信息对检测的影响，本文对划分好的每一个 grid 分别进行判断来确定该 grid 所示区域中是否存在显著的运动特征，对于当前 grid，如果其满足

$$\sum_{j \in \text{grid}} \text{count} \left(\sqrt{v_{x_j}^2 + v_{y_j}^2} \right) > c_{th} \quad (5)$$

$$\text{count} \left(\sqrt{v_{x_j}^2 + v_{y_j}^2} \right) = \begin{cases} 1, & \sqrt{v_{x_j}^2 + v_{y_j}^2} > M_{th} \\ 0, & \sqrt{v_{x_j}^2 + v_{y_j}^2} < M_{th} \end{cases} \quad (6)$$

就认为该 grid 中存在显著的运动特征，不再认为是背景信息，从而对该 grad 提取相应的 AMHOF 特征描述子，进而进行异常判断；否则，就直接跳过该 grad 继续对下一个进行判断，不再对其进行异常检测。

4.2. 集合维护及聚类

对于提取出来的特征向量，算法在第一层维护一个大小固定为 N 的向量集合，该步骤可以视为一个简单的聚类过程：集合中新加入一个特征向量然后将最相近的两个向量进行合并，实质上就是将 $N + 1$ 个特征向量聚成 N 个类，而采用向量合并则能够在保证准确性的情况下提高检测速度。在模型训练的初始阶段，直接将提取出的特征向量 \mathbf{v} 添加进行即可；若集合中的向量数已经等于 N ，则在加入新向量时需要进行合并操作来使得集合中的向量数保持不变：对于集合当中待合并的向量 $\mathbf{v}_a = \{v_{a_1}, v_{a_2}, \dots, v_{a_n}\}$ 和 $\mathbf{v}_b = \{v_{b_1}, v_{b_2}, \dots, v_{b_n}\}$ ，(3)式表明，原先论文当中所使用的方法是直接将两者通过学习率 α 进行加权求和，而根据 3.1 节的分析，在向量合并的时候将两个向量的值直接加权求和，在经过训练数据集的模型训练之后，会使得集合当中的向量所表示的运动信息发生偏移。

为此本文在向量合并过程中，将两个待合并向量之间的相关性也纳入了考虑范围，不再进行简单的向量相加，而是根据两个向量之间的关系来决定合并时的操作。进一步地，在合并时根据两个向量每一维的值，用线性运算来代替加权求和，假设将向量 \mathbf{v}_b 合并到向量 \mathbf{v}_a 当中，则有新向量 \mathbf{v}

$$\mathbf{v} = f(\mathbf{v}_a, \mathbf{v}_b)$$

$$f(\mathbf{v}_a, \mathbf{v}_b) = \sum_{i=1}^n (1 - \alpha) v_{a_i} + \text{sign}(i) \alpha v_{b_i} \quad (7)$$

其中 $\text{sign}(i)$ 为符号函数，其取值受向量 \mathbf{v}_a 和 \mathbf{v}_b 取值的影响

$$\text{sign}(i) = \begin{cases} 1, & v_{a_i} < v_{b_i} \\ -1, & v_{a_i} > v_{b_i} \end{cases} \quad (8)$$

在模型当中，两个向量进行合并的本质是通过一个新的向量来表示原来两个向量所描述的运动信息，由于合并肯定会带来误差，所以在合并时选取集合中最相近的两个向量进行合并，合并之后新的向量应当表现出由一个向量向另外一个向量靠拢的趋势：由于是将向量 \mathbf{v}_b 合并到向量 \mathbf{v}_a 中，所以当 $v_{a_i} < v_{b_i}$ 时，说明向量 \mathbf{v}_b 所描述的运动特征在该维度方向上的运动分量值要高于 \mathbf{v}_a ，合并时在第 i 维上采用加法，使得新的向量 \mathbf{v} 在这一维上的运动分量在 \mathbf{v}_a 的基础上偏向于 \mathbf{v}_b ；而当 $v_{a_i} > v_{b_i}$ 时，则表明 \mathbf{v}_a 在第 i 维的运动分量大于 \mathbf{v}_b ，此时若还是相加，则会使得新向量 \mathbf{v} 在该维上的分量出现了对于 \mathbf{v}_b 的偏离，所以本文此时采用减法运算，让 \mathbf{v} 在这一维上的运动分量信息在 \mathbf{v}_a 的基础上偏向于 \mathbf{v}_b 。

算法模型不仅考虑了那种从未出现过的运动，还将视频中出现频率较低的运动也纳入了异常范畴，所以对每个特征向量定义一个计数值 $r(\mathbf{v})$ ，代表该向量所示特征在视频中出现的次数，对于每个新提取出来的特征向量，其值初始为 1；而对于集合向量合并时生成的新向量，其 $r(\mathbf{v})$ 值为两个向量的计数值之和，如(3)式所示。对于更新之后的向量集合，模型在第二层对 N 个向量进行 K-Means 聚类，将其聚成 K 个类，得到对应的码本，并对每个聚类 C_i ，通过(4)式计算该类的频数 $r(C_i)$ 及占比 $w(C_i)$ ；同时根据每个类中所属向量与该聚类中心 \mathbf{c}_i 之间的关系计算每个类的最大类内距离 d_i

$$d_i = \max_{j \in C_i} \text{dis}(\mathbf{c}_i, \mathbf{v}_j) \quad (9)$$

4.3. 异常检测

视频当中的异常事件定义为：从来没有出现过、或是出现频率较低的运动。对于这两种类型的异常，模型在判断时也对其分别进行考虑。

在异常检测过程中，对于新输入模型的特征向量 \mathbf{v} ，首先判断该向量所描述的运动特征在之前的视频流中有没有出现过：若出现过，则向量 \mathbf{v} 与某一个码本 C_i 之间的误差应该较小，因为该码本中存有该运动的特征信息；若没有出现过，则在所有的码本中均找不到这类信息，那么 \mathbf{v} 与所有码本之间的误差均较大。所以对于新输入的特征向量 \mathbf{v} ，若有

$$\forall_{i \in K} \text{dis}(\mathbf{v}, \mathbf{c}_i) > d_i \quad (10)$$

即认为向量 \mathbf{v} 所表示的运动信息在之前的视频流中从未出现过，应被视为异常。

对于那些低频异常而言，这些运动信息在之前的视频流中已经出现过，且被记录在相应的码本当中，所以其不满足(10)式的条件；而另一方面，低频异常在视频中出现的频率很低，使得表示该类运动信息的向量的 $r(\mathbf{v})$ 值较小，从而包含该向量的码本 C_i 的频数 $r(C_i)$ 及占比 $w(C_i)$ 的取值均较小。那么对于低频异常，如果

$$\exists_{i \in K} \text{s.t.} \begin{cases} \text{dis}(\mathbf{v}, \mathbf{c}_i) < d_i \\ w(C_i) < w_{th} \end{cases} \quad (11)$$

成立，则认为向量 \mathbf{v} 所对应的运动为低频运动，同样属于异常范畴。

5. 实验

本章将呈现对比实验结果，本文对于所提出的检测模型，通过 UCSD 以及 Subway 这两个数据集来进行效果测试。

5.1. UCSD Ped2 数据集

UCSD Ped2 dataset [28]是一个被广泛认可及应用的视频异常检测数据集。其包含的图像序列均为背景相同且固定的监控视频场景，定义行人的运动为一般性运动，均是平行于摄像机的水平运动；而数据

集定义的异常事件则包括自行车、汽车、滑板等与行人运动速度不在同一水平上的运动特征；数据集中包含有 16 个训练图像序列以及 12 个测试图像序列，图像大小均为 360×240 像素，每个 grid 中包含 $45 \times 30 \times 8$ 个像素点。

部分检测结果如图 2 所示，从检测结果中可以看出，原方法由于模型在向量合并过程中直接采用简单加法的原因从而导致的检测效果不理想。本文用向量的线性运算替换了原来的简单加法，使得算法在正常运动物体和异常运动物体的运动方向基本一致、而运动速度存在一定差异时的检测效果得到提高，在大部分场景下能够将异常事件检测出来；另一方面，本文针对稀疏场景的特点进行的剔除背景信息的改进，在拥挤场景中对检测率造成的影响较小，改进前后的检测效果基本一致，但检测效率方面得到了一定的提高，相比于已有的检测方法，在检测率相差不多的情况下检测速率得到了明显提升。

UCSD 数据集已有现有的评价体系，本文采用帧级别的异常检测来评价算法的检测效果，对于所有的视频帧将其视为 01 标签的单幅图像，若图像中存在异常事件且算法报警，则为检测正确，反之为误检。总的检测结果如表 1 所示。



Figure 2. Detection result of UCSD Ped2 dataset
图 2. UCSD Ped2 数据集检测结果

Table 1. Detection comparison on UCSD Ped2 dataset
表 1. UCSD Ped2 数据集量化检测结果

模型	异常事件			异常事件检测率	帧级别等错误率(%)	时间(FPS)
	自行车	滑板	汽车			
真实值	15	3	1	19		
方法[4]	3	0	1	4/19	43.1	10
方法[28]	15	3	1	19/19	25	0.04
方法[29]	15	3	1	19/19	18.5	0.72
Ours	15	3	1	19/19	18.6	18

5.2. Subway 数据集

同时, 本文还对改进前后的算法模型在 Adam's Subway dataset [30]做了检测效果测试, [4]中采用的是地铁出站口的监控视频, 所以在本文的实验当中, 采用的也是地铁出站口的监控视频。根据 Ground Truth, 视频当中行人错误的行走方向被定义为异常事件。地铁出站口的监控视频有 43 分钟共 64,900 帧图像, 视频的分辨率大小为 512×384 像素, 每一个 grid 的大小为 $64 \times 48 \times 8$ 像素, 模型的其余参数保持不变, 取视频前面的一部分作为训练数据, 训练时间约为 5 分钟; 剩余部分进行异常事件检测。部分检测结果如图 3 所示, 左边为原方法的低频误检结果, 右边一列为改进后的实验结果, 数据集当中所定义的异常事件均被检测出来。详细检测结果如表 2 所示。本文所有的实验均运行在 4 GB RAM 和 3.3 GHz CPU 的 Windows 10 系统上, 运行速度约为 25 FPS, 与一些现有方法相比, 在检测率相差无几的情况下, 本文的算法检测速率更高。



Figure 3. Detection result of subway dataset

图 3. Subway 数据集检测结果

Table 2. Detection comparison on subway dataset

表 2. Subway 数据集量化检测结果

模型	异常事件检测率	稀疏场景误检数	总误检数	时间(FPS)
真实值	9			
方法[4]	100% (9/9)	9	12	10
方法[22]	100% (9/9)		2	0.5
方法[31]	100% (9/9)		0	0.2
Ours	100% (9/9)	1	3	25

6. 总结

本文通过大量的实验,发现在[4]中作者所提出的方法在两种视频场景中的检测效果不理想:一是当物体的运动速度有差异而方向一致时;另一种是在稀疏场景当中。经过深入分析之后得出由于算法模型在向量更新过程中采用简单加法这种设计不合理,导致在向量合并时会使得向量所描述的运动信息发生偏移,从而造成漏检;同时由于稀疏场景中主流特征的特殊性,导致算法出现低频误检的概率较高。本文针对这两个特殊场景提出了新的检测模型,以期能够在这些特定场景下达到理想的检测效果,最后通过实验验证,经过改进之后的模型不仅在前面列举的场景中的检测效果明显提升,且运行效率也得到了明显提高。

基金项目

浙江省科技厅重点研发计划(2017C01065)。

参考文献

- [1] Yogameena, B. and Priya, K.S. (2015) Synoptic Video Based Human Crowd Behavior Analysis for Forensic Video Surveillance. *Eighth International Conference on Advances in Pattern Recognition*, Kolkata, 4-7 January 2015, 1-6. <https://doi.org/10.1109/ICAPR.2015.7050662>
- [2] Wang, L. and Dong, M. (2012) Real-Time Detection of Abnormal Crowd Behavior Using a Matrix Approximation-Based Approach. *IEEE International Conference on Image Processing*, Orlando, 30 September-3 October 2012, 2701-2704.
- [3] Yuan, Y., Fang, J. and Wang, Q. (2015) Online Anomaly Detection in Crowd Scenes via Structure Analysis. *IEEE Transactions on Cybernetics*, **45**, 562-575. <https://doi.org/10.1109/TCYB.2014.2330853>
- [4] Feng, J., Zhang, C. and Hao, P. (2012) Online Anomaly Detection in Videos by Clustering Dynamic Exemplars. *International Conference on Image Processing*, Orlando, 30 September-3 October 2012, 3097-3100. <https://doi.org/10.1109/ICIP.2012.6467555>
- [5] Wang, H., Kläser, A. and Schmid, C. (2013) Dense Trajectories and Motion Boundary Descriptors for Action Recognition. *International Journal of Computer Vision*, **103**, 60-79. <https://doi.org/10.1007/s11263-012-0594-8>
- [6] Wang, H. and Schmid, C. (2013) Action Recognition with Improved Trajectories. *IEEE International Conference on Computer Vision*, Sydney, December 2013, 3551-3558. <https://doi.org/10.1109/ICCV.2013.441>
- [7] Wang, L., Qiao, Y. and Tang, X. (2015) Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors. *Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 4305-4314. <https://doi.org/10.1109/CVPR.2015.7299059>
- [8] Murthy, O.V.R. and Goecke, R. (2013) Ordered Trajectories for Large Scale Human Action Recognition. *IEEE International Conference on Computer Vision Workshops*, Washington DC, 2-8 December 2013, 412-419. <https://doi.org/10.1109/ICCVW.2013.61>
- [9] Shu, T., Xie, D., Rothrock, B., et al. (2015) Joint Inference of groups, Events and Human Roles in Aerial Videos. *The IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 4576-4584.
- [10] Cui, X., Liu, Q., Gao, M., et al. (2011) Abnormal Detection Using Interaction Energy Potentials. *The IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 20-25 June 2011, 3161-3167. <https://doi.org/10.1109/CVPR.2011.5995558>
- [11] Wu, S., Moore, B.E. and Shah, M. (2010) Chaotic Invariants of Lagrangian Particle Trajectories for Anomaly Detection in Crowded Scenes. *IEEE Conference on Computer Vision & Pattern Recognition*, San Francisco, 13-18 June 2010, 2054-2060. <https://doi.org/10.1109/CVPR.2010.5539882>
- [12] Guo, Z., Li, N., Xu, D., et al. (2013) A Novel Statistical Learning-Based Framework for Automatic Anomaly Detection and Localization in Crowds. *IEEE International Conference on Robotics and Biomimetics*, Shenzhen, 12-14 December 2013, 1211-1215. <https://doi.org/10.1109/ROBIO.2013.6739629>
- [13] Sabokrou, M., Fathy, M., Hoseini, M., et al. (2015) Real-Time Anomaly Detection and Localization in Crowded Scenes. *Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 56-62.
- [14] Reddy, V., Sanderson, C. and Lovell, B.C. (2011) Improved Anomaly Detection in Crowded Scenes via Cell-Based Analysis of Foreground Speed, Size and Texture. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Colorado Springs, 20-25 June 2011, 55-61.

- [15] Kratz, L. and Nishino, K. (2009) Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models. *IEEE Conference on Computer Vision & Pattern Recognition*, Miami, 20-25 June 2009, 1446-1453. <https://doi.org/10.1109/CVPR.2009.5206771>
- [16] Feng, J., Zhang, C. and Hao, P. (2010) Online Learning with Self-Organizing Maps for Anomaly Detection in Crowd Scenes. *International Conference on Pattern Recognition*, Istanbul, 23-26 August 2010, 3599-3602. <https://doi.org/10.1109/ICPR.2010.878>
- [17] Cong, Y., Yuan, J. and Liu, J. (2011) Sparse Reconstruction Cost for Abnormal Event Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 20-25 June 2011, 3449-3456. <https://doi.org/10.1109/CVPR.2011.5995434>
- [18] Ryan, D., Denman, S., Fookes, C., *et al.* (2011) Textures of Optical Flow for Real-Time Anomaly Detection in Crowds. *IEEE International Conference on Advanced Video & Signal-Based Surveillance*, Klagenfurt, 30 August-2 September 2011, 1-6. <https://doi.org/10.1109/AVSS.2011.6027327>
- [19] Roshtkhari, M.J. and Levine, M.D. (2013) An On-Line, Real-Time Learning Method for Detecting Anomalies in Videos Using Spatio-Temporal Compositions. *Computer Vision & Image Understanding*, **117**, 1436-1452. <https://doi.org/10.1016/j.cviu.2013.06.007>
- [20] Xu, D., Wu, X., Song, D., *et al.* (2013) Hierarchical Activity Discovery within Spatio-Temporal Context for Video Anomaly Detection. *IEEE International Conference on Image Processing*, Melbourne, 15-18 September 2013, 3597-3601. <https://doi.org/10.1109/ICIP.2013.6738742>
- [21] Mehran, R., Oyama, A. and Shah, M. (2009) Abnormal Crowd Behavior Detection Using Social Force Model. *Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 935-942.
- [22] Zhao, B., Li, F.F. and Xing, E.P. (2011) Online Detection of Unusual Events in Videos via Dynamic Sparse Coding. *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 20-25 June 2011, 3313-3320. <https://doi.org/10.1109/CVPR.2011.5995524>
- [23] Lu, C., Shi, J. and Jia, J. (2013) Abnormal Event Detection at 150 FPS in MATLAB. *IEEE International Conference on Computer Vision*, Sydney, 1-8 December 2013, 2720-2727. <https://doi.org/10.1109/ICCV.2013.338>
- [24] Li, H., Zhang, Y., Yang, M., *et al.* (2014) A Rapid Abnormal Event Detection Method for Surveillance Video Based on a Novel Feature in Compressed Domain of HEVC. *IEEE International Conference on Multimedia and Expo*, Chengdu, 14-18 July 2014, 1-6. <https://doi.org/10.1109/ICME.2014.6890212>
- [25] Wang, T. and Snoussi, H. (2015) Detection of Abnormal Events via Optical Flow Feature Analysis. *Sensors*, **15**, 7156-7171. <https://doi.org/10.3390/s150407156>
- [26] Fang, Z., Fei, F., Fang, Y., *et al.* (2016) Abnormal Event Detection in Crowded Scenes Based on Deep Learning. *Multimedia Tools & Applications*, **75**, 14617-14639. <https://doi.org/10.1007/s11042-016-3316-3>
- [27] Roshtkhari, M.J. and Levine, M.D. (2013) Online Dominant and Anomalous Behavior Detection in Videos. *IEEE Conference on Computer Vision & Pattern Recognition*, Portland, 23-28 June 2013, 2611-2618. <https://doi.org/10.1109/CVPR.2013.337>
- [28] Mahadevan, V., Li, W., Bhalodia, V., *et al.* (2010) Anomaly Detection in Crowded Scenes. *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 13-18 June 2010, 1975-1981. <https://doi.org/10.1109/CVPR.2010.5539872>
- [29] Li, W., Mahadevan, V. and Vasconcelos, N. (2014) Anomaly Detection and Localization in Crowded Scenes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **36**, 18-32. <https://doi.org/10.1109/TPAMI.2013.111>
- [30] Adam, A., Rivlin, E., Shimshoni, I., *et al.* (2007) Robust Real-Time Unusual Event Detection Using Multiple Fixed-Location Monitors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **30**, 555-560. <https://doi.org/10.1109/TPAMI.2007.70825>
- [31] Cong, Y., Yuan, J. and Liu, J. (2013) Abnormal Event Detection in Crowded Scenes Using Sparse Representation. *Pattern Recognition*, **46**, 1851-1864. <https://doi.org/10.1016/j.patcog.2012.11.021>

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2326-3415，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：airr@hanspub.org