

基于 DBSCAN 聚类算法的异常轨迹检测

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

(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. 航天恒星科技有限公司, 北京 100086; 4. 中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016)

摘要: 现有的异常轨迹检测算法往往侧重于检测轨迹的空域异常, 忽略了对轨迹时域异常的检测, 并且检测精确度不高, 针对此类问题, 提出了基于增强聚类的异常轨迹检测算法。首先, 采用基于速度的最小描述长度(VMDL)准则把轨迹简化成有序线段; 然后, 使用改进的线段间的距离定义, 基于 DBSCAN 算法把线段分为不同的类, 以建模局部正常运动模式; 最后, 采用先检测空间异常性再检测时间异常性的二级检测算法, 检测时空异常轨迹点。在多个测试集上的实验结果表明: 该算法可以检测位置、角度、速度等三种时空异常轨迹点, 相对于其他算法, 明显提高了异常轨迹检测的精确度。

关键词: 时空异常轨迹检测; VMDL 分割准则; DBSCAN 聚类算法; 二级检测算法

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA201746.0528001

Trajectory outlier detection based on DBSCAN clustering algorithm

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

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. Space Star Technology Co., Ltd., Beijing 100086, China;

4. Key Laboratory of Opt-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China)

Abstract: Existing traditional trajectory outlier detection algorithms always focus on spatial outliers and ignore temporal outliers, and the accuracy is relatively low. To solve these problems, a simple and effective approach based on enhanced clustering algorithm was proposed to detect spatio-temporal trajectory outliers. Firstly, each original trajectory was simplified into a set of sequential line segments with the velocity-based minimum description length (VMDL) partition principle. Secondly, the distance formula between line segments was improved to enhance the clustering performance. Using DBSCAN algorithm, the line segments were classified into different groups which could represent local normal behaviors. Thirdly, outliers were detected using two-level detection algorithm which first detected spatial outliers and then detected temporal outliers. Experimental results on multiple trajectory data sets demonstrate that the proposed algorithm could successfully detect three kinds of spatio-temporal outliers, position, angle and

收稿日期: 2016-09-11; 修订日期: 2016-10-20

作者简介: 周培培(1991-), 女, 博士生, 主要从事图像处理方面的研究。Email: zhoupeipei@sia.cn

导师简介: 丁庆海(1969-), 男, 研究员, 博士生导师, 博士, 主要从事光电成像、光电跟踪、目标识别等方面的研究。

Email: 13693689880@139.com

罗海波(1967-), 男, 研究员, 博士生导师, 博士, 主要从事图像处理、目标识别等方面的研究。Email: luohb@sia.cn

velocity. Compared with other methods, the precision and accuracy make great improvement.

Key words: spatio-temporal trajectory outlier detection; VMDL partition principle; DBSCAN clustering algorithm; two-level detection algorithm

0 引言

目前,使用先进的卫星设备及跟踪技术可以方便地获得大量移动目标的轨迹数据,包括飓风跟踪、动物迁徙、交通工具移动以及人类活动等各类轨迹。庞大的轨迹集隐含着很多有用的信息,特别是对于社会公共环境中的财产和人身安全,异常轨迹检测具有重要的意义。但是,异常事件的发生率往往较低,常用的方法是对环境中的正常行为进行建模,那么,新出现的行为就默认为异常。关于异常,笔者采用 V. Barnett 等人^[1]对异常的定义,该定义更加科学且易于理解:“An outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data”,文中的异常主要指那些与自身时空邻域特性不一致的轨迹点,特指位置、方向与速度异常的轨迹点。

对于由全球定位系统(GPS)或在视频图像中采用不同的跟踪算法^[2-4]得到的目标轨迹,研究人员提出了多种轨迹分析模型^[5-9]。总结起来,轨迹的相似性度量是建模过程中的难点之一,由于任意两条轨迹包含的轨迹点个数不同以及获取轨迹时的误差使得某些连续点的坐标变化较大,造成轨迹间的距离难以度量,因此,对轨迹的度量方式直接影响着异常轨迹检测的精度,一般有整体度量和分段度量两种方式。

整体度量方式是将轨迹看成一个整体进行分析,直接检测每条轨迹的整体异常性,比如,Hu 等人^[9]使用改进的 Hausdorff 距离度量轨迹距离,构建了一个轨迹数据集的空间相似性矩阵,该矩阵的元素值代表对应的两条轨迹之间的距离,然后使用空间聚类获得轨迹模式,最后通过模式匹配判断轨迹的空间异常性。Zhang 等人^[10]除了使用轨迹点的位置外,还综合考虑角度、速度、加速度几种因子的影响,把待测点与自身轨迹中的其他点进行比较,异常轨迹点超过一定比例时认为该轨迹异常。然而,直接分析轨迹整体异常性的方法计算效率低下,而且容易

忽略掉一些异常的轨迹片段。

分段度量方式是根据轨迹特性将其分割成一些线段的有序连接^[5,11],然后以分割后的轨迹片段为单位进行分析。此方式只需保留轨迹中的关键点,大大简化了轨迹。然而, Lee 等人^[5,11]在轨迹分割时并未考虑轨迹点的速度信息,导致在轨迹的子集中不能对轨迹的速度异常进行检测。

对于整体度量和分段度量,研究人员采用了多种建模算法来挖掘轨迹信息,如聚类算法^[9,11-13]、马尔科夫模型^[14]、隐马尔科夫模型^[15]等。除此之外,参考文献[5]中首先采用二级分割策略简化轨迹,其次结合距离和密度检测异常点,不足之处是没有充分利用时间维度的信息。与参考文献[5]相似,Liu 等人^[8]使用基于密度的检测模型对轨迹点进行时空两个维度建模,这种方法相比于参考文献[5]可以检测出更多的有效异常点,但是同样在分割阶段忽略了速度信息。针对大数据集的时空异常轨迹点,Birant 等人^[12]提出一种简单的三步检测法:轨迹点聚类,空域异常检测,时域异常检测。显然,相比于仅仅使用空间信息,时空轨迹模型可以更合理地检测到潜在的异常点,但是很多算法往往不能适应实际应用中的多样性和复杂性。还有学者在使用原始轨迹点信息的基础上加入了高层语义信息来分析轨迹异常,比如, Lee 等人^[16]首先提取基于区域的高层特征和基于轨迹的低层特征,然后利用聚类方法分析轨迹集的信息;参考文献[7,17]重点分析了兴趣区域特征来分析轨迹的异常。尽管使用语义信息的模型在某些场景中分析效果较好,但也有其局限性,如这种模型不适用于长度差异太大的轨迹集数据。

总结以上检测算法可知,整体度量方法容易忽略轨迹中的异常片段且计算较为复杂,分段度量方法能有效地检测局部异常,检测精度较高。所以,文中采用分段度量的方式分析轨迹的异常。首先,通过分割来简化轨迹。针对参考文献[11]中分割后某些轨迹片段内部速度变化显著的问题,文中提出了基于速度的最小描述长度准则(VMDL)分割算法,把速

度变化对分割效果的影响考虑在内,以利于对轨迹点速度异常的检测。然后,采用聚类算法对线段建模,文中通过改进线段间的距离定义改善了聚类效果,使其对角度更加敏感,以利于对轨迹中角度异常的检测。最后,采用二级检测算法,先检测空域异常再检测时域异常。实验结果表明:文中算法成功地检测出了绝大部分的异常轨迹点,以较低的计算复杂度保证了高检测率和低虚警率。

1 异常轨迹检测

每条轨迹都是特定目标在不同时刻所处位置点的集合,表示为 $T_i=p_1 \cdots p_j \cdots p_{len_i} (1 \leq i \leq N)$,其中 $p_j (1 \leq j \leq len_i)$ 是三维空间中的一个点 (i, x_j, y_j, t_j) ,表示轨迹 T_i 所对应的目标在 t_j 时刻的位置坐标 (x_j, y_j) 。由于各目标的运动速度存在差异,导致每条轨迹上的点数 len_i 各不相同。对于同一轨迹 T 上两点之间的轨迹片段定义为子轨迹 $Ts(p_s, p_t)=p_s p_{s+1} \cdots p_t$ 。文中提出的异常轨迹检测算法的输入为 N 条轨迹数据集 $T=\{T_1, \cdots, T_i, \cdots, T_N\}$,输出是轨迹中所有的异常子轨迹的集合 $O=\{Ts_1, Ts_2, \cdots, Ts_n\}$ 。

参考文献[5]可以成功检测出空间异常轨迹,然而,异常轨迹还常常表现为速度异常。速度异常有两层含义:其一,轨迹的全局速度明显有别于其他轨迹,如横穿广场的摩托车;其二,轨迹的内部速度有显著变化,如行走的小偷偷东西之后快速跑开。事实上,如果相机视场较小,第二种情况就可能退化为第一种情况。文中工作的目标就是能够同时检测出角度、位置以及速度异常。

Lee 等人^[5]定义了轨迹点的位置异常和角度异常,文中在此基础上对速度异常进行了定义,并将其用于异常轨迹检测中。

(1) 位置异常:指位置不同于其邻域的轨迹^[5]。

(2) 角度异常:指角度不同于其邻域的轨迹^[5]。

(3) 速度异常:指速度不同于其邻域的轨迹。

这三种异常点涵盖了时域和空域的信息,对智能监控的应用来说更加合理。

以 ICAI 院内轨迹数据集为例,图 1 直观展示了文中算法流程,其中图 1(a)~(c)为训练阶段,图 1(d)~(f)为检测阶段。图 1(e)为分割线段的预测结果;图 1(f)为轨迹点预测结果,点表示预测为正常的正常轨迹点(TN),圈表示预测为异常的正常轨迹点(FP)。

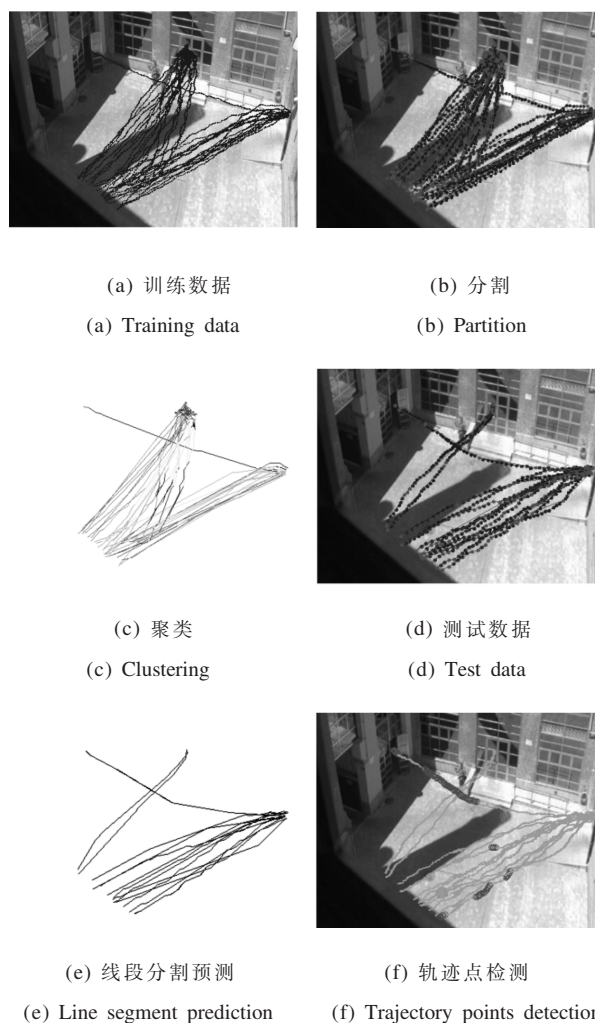


图 1 文中算法流程图

Fig.1 Flow chart of the proposed method

1.1 轨迹分割

轨迹分割是提取其显著特征点,依次连接显著特征点得到的新轨迹。同一轨迹上相邻特征点的连线定义为线段。轨迹分割有两个优化目标,一是分割后新轨迹与原轨迹差异尽量小,二是新轨迹中的线段尽量少。

如图 2 所示,图 2(a)是分辨率为 200×200 的图像中 4 个目标的轨迹点,目标均从左下角往右上角移动,采样频率相同。参考文献[11]使用最小描述长度(MDL)准则分割得到图 2(b)所示的结果,显然, T_4 上的子轨迹 $Ts_4(p_2, p_6)$ 包含两段速度明显不同的轨迹片段,而后续对轨迹的处理以线段为最小单位,所以这种分割方法会影响后续对异常速度的检测。

文中提出了基于速度的最小描述长度(VMDL)准则来简化轨迹描述,对于任意一条轨迹 $T=p_1 \cdots p_i$,

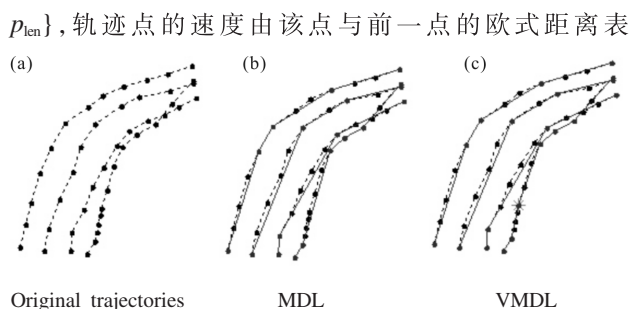


图 2 图像中 4 个目标的轨迹分割效果,其中虚线是原始轨迹,实线是分割后的新轨迹

Fig.2 Results of partition algorithms about four objects on an image, where the dotted lines is the original trajectories and the solid ones are segmented trajectories

示,输入速度阈值 V_T ,输出为轨迹 T 的特征点集 C ,依次连接 C 即可得到分割后的轨迹,采用 VMDL 准则进行轨迹分割的算法流程如下所示:

/* 速度分割 */

- (1) 把 p_1 放入特征点集 C ;
初始化变量值: $startId=1, length=1$;
- (2) 令 $current=startId+length$;
- (3) 计算点 $p_{startId} \sim p_{current-1}$ 的线段的速度均值,表示为 $v_m = velocity_mean(p_{startId}, p_{current-1})$;
- (4) 若 $\left| \frac{v_m - v(p_{current})}{\min(v_m, v(p_{current}))} \right| < V_T$, 则令 $length=length+1$,转到 06;
- (5) 若 $\left| \frac{v_m - v(p_{current})}{\min(v_m, v(p_{current}))} \right| \geq V_T$,把当前点的前一点 $p_{current-1}$ 放入 C ;更新变量值:
 $startId=curreIndex-1; length = 1$;
- (6) 重复步骤(2)~(5)直至 $startId+length>len$;
- (7) 把轨迹最后一个点加入特征点集 C .
/*MDL 准则分割^[11]*/
- (8) 重新初始化变量: $startId=1; length=1$;
- (9) 令 $current=startId+length$;
- (10) 基于 MDL 准则计算当前点作为特征点与否的代价: $cost_{par}=MDL_{par}(p_{startId}; p_{current})$;
 $cost_{nopar}=MDL_{nopar}(p_{startId}; p_{current})$;
- (11) 若当前点 $p_{current}$ 已经是 C 内特征点,则 $startId=current; length=1$;跳转到(14)
- (12) 若点 $p_{current}$ 不是 C 内特征点,判断若 $cost_{par} < cost_{no_par}$,则 $length=length+1$,转到(14);
- (13) 若 $cost_{par} \geq cost_{no_par}$,把当前点的前一点 $p_{current-1}$ 放

入 C ,并更新变量

$startId=current-1; length=1$;

(14) 若 $startId + length > len$,程序结束,输出 C ,否则重复步骤(9)~(13)。

使用提出的 VMDL 准则得到的分割结果如图 2(c)所示,可见,速度差别大于阈值的轨迹片段被分割到了不同的线段,同一线段内部轨迹点的速度基本保持一致。

1.2 训练集建模

假设训练集中已经包括了该场景中常见的正常轨迹模式,通过对轨迹集建模得到正常轨迹模式的定量表示。文中通过对训练集中的线段聚类得到不同的轨迹模式,鉴于 DBSCAN^[18]聚类算法可以得到任意形状的一类,采用 DBSCAN 作为基本的聚类算法。此处,聚类算法的输入量是连接相邻特征点所得的线段,输出是不同类的线段,每个类代表了一种运动模式,然后,筛选出满足一定条件的类,表示局部正常的运动模式。

聚类需要解决的首要问题就是如何度量线段之间的差异,参考文献[11,19]中定义的线段之间的距离由 3 部分构成,即垂直距离 d_{\perp} 、平行距离 d_{\parallel} 和角度距离 d_{θ} ,图 3 直观地展示了 3 种距离的定义及物理意义,线段 L_i 与线段 L_j 间的距离表示为:

$$\text{dist}(L_i, L_j) = \omega_{\perp} \cdot d_{\perp}(L_i, L_j) + \omega_{\parallel} \cdot d_{\parallel}(L_i, L_j) + \omega_{\theta} \cdot d_{\theta}(L_i, L_j) \quad (1)$$

式中: ω_{\perp} 、 ω_{\parallel} 和 ω_{θ} 分别对应垂直距离、平行距离和角度距离的权值。

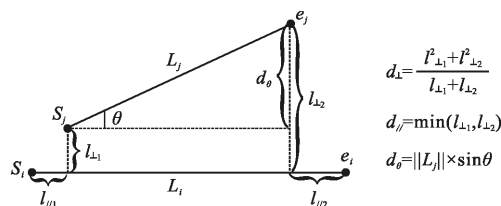


图 3 线段 $L_i(s_i, e_i)$ 和 $L_j(s_j, e_j)$ 距离函数的组成部分,其中 $L_j(s_j, e_j)$ 代表较短的线段, θ 是两线段较小的交叉角

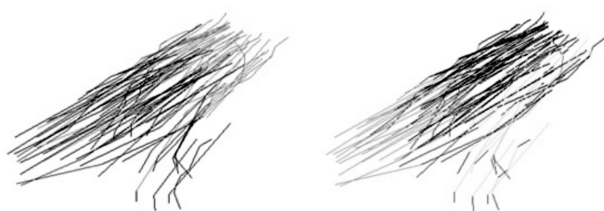
Fig.3 Components of distance function for two line segments, $L_i(s_i, e_i)$ and $L_j(s_j, e_j)$. Here, $L_j(s_j, e_j)$ is the longer one and θ is the smaller intersecting angle

由图 3 可以看出,角度距离与较短线段长度 L_i 、线段间的夹角 θ 成正比,当线段较短而角度 θ 很大时,大角度 θ 无法在角度距离上很好地体现出来,从而导致线段聚类后同一类中的角度差异较大,影响后

续对角度异常检测的准确度。对角度距离的定义进行了改进,使角度距离更好地体现角度的变化,而不受线段长度的影响。角度距离的定义需要满足以下两点要求:能直接反应角度的变化,量纲与垂直距离 d_{\perp} 、平行距离 d_{\parallel} 保持一致。文中对角度距离的定义如下:

$$d_{\theta}(L_i, L_j) = \|L_0\| \times \sin\theta \quad (2)$$

式中: θ 指线段 L_i 与 L_j 之间较小的夹角,即 $\theta \in [0, \pi/2]$; $\|L_0\|$ 表示线段间的距离常数,在文中,距离单位为像素,由于 $\sin\theta \in [0, 1]$,为了使 d_{θ} 与 d_{\perp} 、 d_{\parallel} 在数量级上保持一致,文中 $\|L_0\|$ 取经验值 20。综上,角度距离 $d_{\theta}(L_i, L_j) \in [0, \|L_0\|]$ 。由公式(2)所示的角度距离定义可见, d_{θ} 只与线段间的夹角相关,线段间的距离 $\text{dist}(L_i, L_j)$ 随线段的夹角增大而增大,因此,线段间的角度距离可以更好地区分线段间的夹角。如图 4 所示,在 UCSD 轨迹集上,把分割后得到的线段聚到不同的类,图 4(a)是传统算法的聚类结果^[11],有些角度变化较大的线段被归到了同一类,图 4(b)为使用改进角度距离后的聚类算法的分类结果,由图 4 可见,聚类效果有了很大的改善,更利于后续对角度异常的检测。文中实验中线段距离的 3 个组成部分垂直距离 d_{\perp} 、平行距离 d_{\parallel} 和角度距离 d_{θ} 的权值分别设置为 0.25、0.25、0.5。



(a) 传统算法 (a) Traditional algorithm
(b) 改进算法 (b) Improved algorithm

图 4 传统聚类算法与增强聚类算法的对比效果图
Fig.4 Comparison results between traditional and improved clustering algorithms

通过对线段进行聚类,角度、位置较为相似的被归为一类。对于训练数据集,既包括正常运动模式,也包括异常运动模式和噪声,其中可能存在误差。完成聚类操作后,把轨迹数超过阈值的类视为是正常类,代表局部运动的正常轨迹模式;低于该阈值的类视为异常类,代表训练集中的异常轨迹模式。由此,通过对训练集建模,得到了正常运动的数学模型。

1.3 异常轨迹检测

对于待检轨迹,先对其进行分割简化,表示为线段集,轨迹检测算法通过比较待测线段与其邻域中线段的轨迹模式来判断其是否属于异常轨迹,若某线段为异常,则与该线段对应的轨迹片段为异常,否则,把该轨迹片段视为正常。按照文中的定义,对于一个待检的轨迹片段,若检测到位置、角度和速度有一种异常,则该轨迹片段视为异常。对于线段的检测过程,文中采用二级检测算法,具体的流程图如图 5 所示。

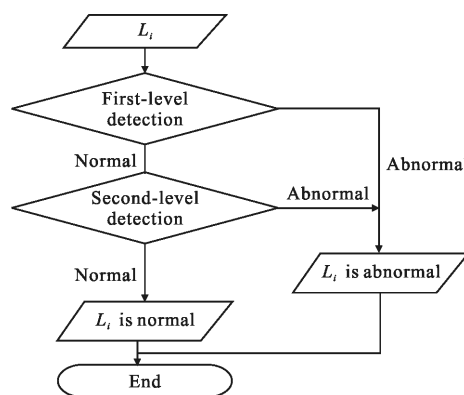


图 5 轨迹线段片段检测流程图

Fig.5 Flow chart of sub-trajectory detection

第一级检测中,检测线段 L_i 空间邻域内线段的异常性,判断 L_i 的位置和角度是否异常。对于测试集 T_e 中任意一个线段 $\forall L_i \in T_e$, 该线段在训练集中的 ε -邻域 $N_{\varepsilon}(L_i)$ 定义为:

$$N_{\varepsilon} = \{L_j | d_{\theta}(L_i, L_j) < \varepsilon, L_j \in T_r\} \quad (3)$$

式中: T_r 表示训练集中线段的集合。据此,计算得到待检线段 L_i 的空间邻域中正常线段的概率为:

$$P(L_i) = \frac{\eta_{\text{normal}}(L_i)}{\eta_{\text{total}}(L_i)} \quad \text{s.t. } L_j \in \{L_j | L_j \in N_{\varepsilon}(L_i), L_j \in T_r\} \quad (4)$$

式中:分母 $\eta_{\text{total}}(L_i)$ 代表训练集中处于 L_i 的邻域内的线段总数,分子 $\eta_{\text{normal}}(L_i)$ 代表这些线段中属于正常模式类的线段的数量。公式(4)为通过检测待测线段的空域内属于正常轨迹的概率来判断待检线段空域是否异常。

公式(4)中,若 $\eta_{\text{total}}(L_i) = 0$ 或 $P(L_i) < Th$,判定线段 L_i 为空域异常,对应的轨迹片段是异常的。若 $\eta_{\text{total}}(L_i) \neq 0$ 且 $P(L_i) \geq Th$,认为线段 L_i 在空域是正常的,此时把线段 L_i 归类为其邻域中个数最多的线段所在的类 G_k ,进一步检测线段的时域异常性。

第二级检测中,通过计算线段 L_i 时间邻域内异常线段的概率来判断该线段速度是否异常。线段的速度以线段对应的轨迹片段上点的平均速度来表示,通过比较线段 L_i 的速度 v_i 和所在类 G_k 的速度来判断线段速度是否异常。假设同一类内线段的速度服从正态分布,可以计算得到每个类 G_k 的所有线段的平均速度 \bar{V}_k 和方差 σ_k^2 。采用 3σ 准则判断线段 L_i 的异常性 $y(L_i)$,如公式(5)所示:

$$y(L_i) = \begin{cases} 0, & \bar{V}_k - 3\sigma_k \leq v_i \leq \bar{V}_k + 3\sigma_k \\ 1, & \text{else} \end{cases} \quad (5)$$

式中: $y(L_i)$ 值为 0 表示正常,值为 1 表示异常。若线段 L_i 的速度 v_i 不满足 3σ 准则,则线段 L_i 的速度被判断为异常,即该线段在时域是异常的,那么线段 L_i 对应的轨迹片段就是异常轨迹;否则,线段 L_i 速度正常,也就是说,该线段在空域和时域均正常,其对应的轨迹片段属于正常运动模式。

2 实验结果与参数分析

2.1 实验结果

为了验证文中算法的有效性,对 3 个实际的视频序列进行测试:(1) UCSD 异常检测数据集^[20]:是室外人行道的视频,实际标注的异常轨迹是速度较快的自行车和滑板的轨迹;(2) AVSS PV Easy 数据集^[21]:是繁忙的街道与一条幽静小路的视频,其训练集与测试集以一个白色货车在禁停位置停车为分界,货车停车之前作为训练集,之后作为测试集;(3) ICAI:俯角拍摄的院内视频,门与门之间为正常通道,另有一些人选择其他异常的路径,异常轨迹是异常行为的所有轨迹,均作为测试集,剩余正常轨迹被分成训练集和测试集两部分。图 6 分别是 3 个训练集的轨迹数据。



图 6 3 个实际的轨迹集的训练数据

Fig. 6 Training data of three real trajectory set

文中算法与 2014 年 Alvar 等人^[13]提出的基于 DS (Dominant Sets) 的异常检测算法有很多相似之处,如

均考虑了时空信息、均把正常轨迹表示为不同的类、与正常轨迹差别较大的视为异常等。因此,将文中的实验结果主要与参考文献[13]的结果做对比,其中所使用的轨迹数据集均是由 Alvar 等人^[13]采集的^[22],把每条轨迹 T_i 上的轨迹点 p_i 表示成 4 维格式 (i, x_j, y_j, t_j) 。

由于实验数据是人工获取并且帧频较高,所以难免会有局部波动情况,例如,连接轨迹点发现目标的运动轨迹上有个别点后移或者与整体运动趋势明显不同的现象,这属于数据集中的噪声干扰。为减轻噪声干扰,首先对轨迹进行均值滤波,然后每隔 t 点取一个点作为实验轨迹,这里 t 取值较小,仅为了去除局部干扰,以免丢失有效数据。

文中采用参考文献[13]中对异常点的标记方法,把异常点标记为 P ,把正常点标记为 N ,那么,漏报(FN)是指把异常点标记成了正常点,虚警(FP)是指把正常点标记成了异常点,通过比较相同数据集中的虚警数和漏报数,将文中检测算法与 2014 年 Alvar 等人^[13]提出的 DS 算法做比较。图 7 所示为两种算法在 4 个实际测试集上检测的定性对比结果,同一数据集上第一列为 DS 算法的检测结果,第二列为文中方法的检测结果;“.”和“x”分别表示正常点预测为正常点(TN)和异常点预测为正常点(FN),“o”和“*”分别表示异常点预测成为异常点(TP)和正常点预测成为异常点(FP),即“x”(FN)和“*” (FP)代表的均是检测错误的样本,可以看出,本文算法检测错误的轨迹点只聚集于数量较少的几条轨迹上,对轨迹的误判率较低;而 DS 算法在 AVSS 数据集中检测错误的轨迹点几乎遍及每条轨迹。表 1 是在 4 个数据集上两种算法的定量评估对比结果,表格中的数字均代表对应项中轨迹点的个数。

综合分析图 7 和表 1,文中方法在 3 个数据集中表现较为平稳,准确度明显优于 DS 算法,在 UCSD 数据集中, FN 和 FP 均明显下降,即准确度得到提升;在 AVSS 数据集中,被误判的轨迹数明显降低,即精度得到改善;使用与 ICAI 相同的参数,对只有正样本的 ICAI2 测试集,文中算法表现出了明显改善的效果。对于测试数据集中异常检测错误的现象,主要因为这些轨迹点与所定义的正常轨迹在时空特性上表现较为相似,即空间位置相近、速度相差较小,所以文中算法不可避免地会在检测这类异常点时能力变弱。对于 ICAI 测试集,效果略差,原因是文中采用了分段局部比较方法,若训练集中待测线段附近有类似线段,则判断该线段是正常轨迹

线段,所以出现了图 7(c)左侧“x”的错误检测结果。总之,采用实际数据集的实验结果表明:文中算法的精确度均有明显提升。

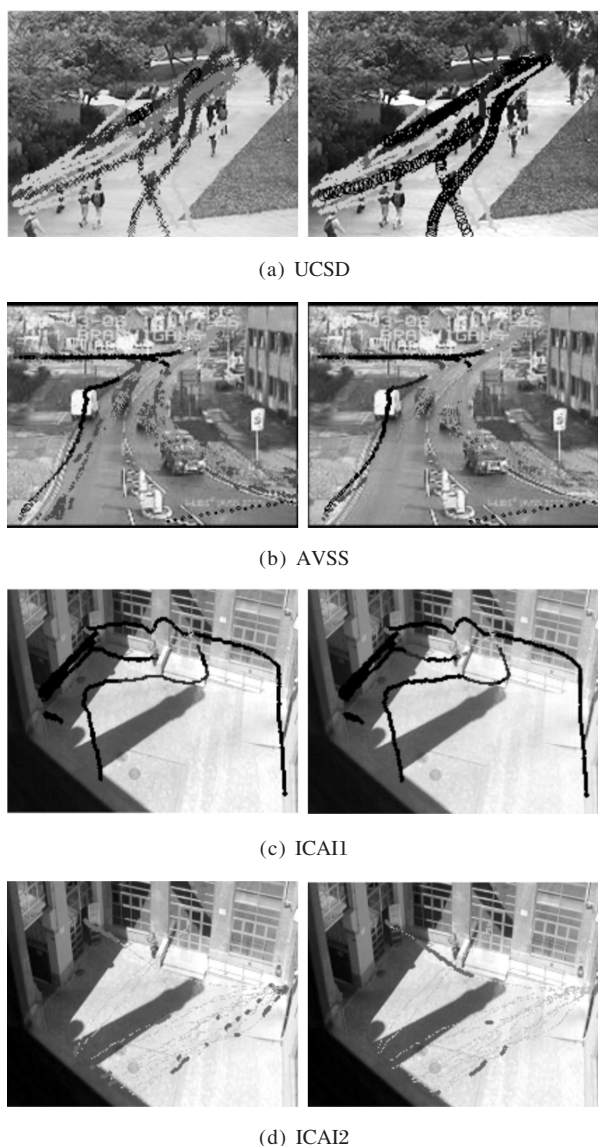


图 7 DS^[13]检测方法(左)和文中方法(右)定性对比结果
Fig.7 Comparison of qualitative results with DS^[13] method (left) and proposed method (right)

表 1 文中算法与传统算法的定量对比结果

Tab.1 Comparison of quantitative assessment between traditional and proposed methods

Video sequence	Training data	Test data	Abnormal samples	FN/FP of DS ^[13]	FN/FP of proposed method
UCSD	13 787	2 803	454	437/400	126/18
AVSS	10 821	15 203	4 445	65/728	126/315
ICAI1	10 032	3 125	2 866	175/41	150/55
ICAI2	10 032	3 191	0	0/211	0/135

为了比较算法的整体性能,对文中基于增强聚类的异常轨迹检测方法和基于 DS 的检测方法在同一数据集 AVSS 上绘制 ROC 曲线,如图 8 所示。可以发现,当虚警率 $FPR < 0.15$ 时,文中算法的正确率 TPR 明显高于参考文献[13]中的 DS 方法;当虚警率 $FPR > 0.2$ 时,两种算法均表现良好且稳定。由此可见,若要求虚警率较低时,文中算法性能明显优于传统算法,因此,实用性更好。

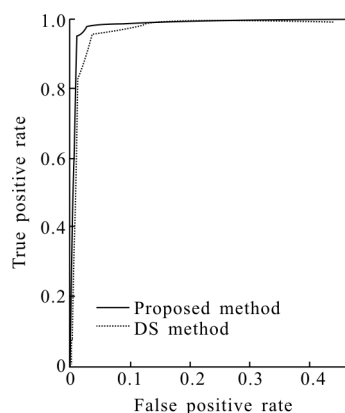


图 8 ROC 空间系统性能比较

Fig.8 Comparison of system performance in ROC space

2.2 参数影响分析

文中算法最重要的参数是 DBSCAN 聚类算法中邻域半径的大小 ϵ , ϵ 增大,聚类得到的类数减少,类内线段数目增多,线段间的差异变大,漏报率增加;反之,类内线段差异变小,虚警率增加。调整 ϵ 可优化 FN 和 FP,从而改进检测效果,对于每个场景而言,在含有正负样本的测试集上,采用寻优方法可以确定一个最优的 ϵ 值,使用该值对该场景进行异常检测可以达到较好的检测效果。例如,对于 ICAI1 测试集,调整 ϵ 的大小,当 FP 和 FN 均达到最优时,对应的 ϵ 值即为最优值。

3 结论

通过分析传统异常轨迹检测方法,发现存在时空异常轨迹点检测不全、检测精度不高等问题,文中提出了一种先将轨迹分割成线段,然后再对线段进行分类的异常轨迹检测方法。

相对于现有算法,文中算法在准确率和精度方面都有了较大的提高,且计算复杂度较低,可以满足实时性要求。然而,该算法中模型参数需要针对同一

场景的数据集监督学习, 这有待于在未来工作中研究改进方法, 使算法更具适用性。

参考文献:

- [1] Barnett V, Lewis T. Outliers in Statistical Data [M]. Chichester: John Wiley, 1995: 584.
- [2] Cui Xiongwen, Wu Qinzhang, Jiang Ping, et al. Affine-invariant target tracking based on subspace representation[J]. *Infrared and Laser Engineering*, 2015, 44(2): 769-774. (in Chinese)
崔雄文, 吴钦章, 蒋平, 等. 子空间模型下的仿射不变目标跟踪[J]. *红外与激光工程*, 2015, 44(2): 769-774.
- [3] Hi H, Ling H, Wu Y., et al. Real time robust l1 tracker using accelerated proximal gradient approach[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, 157(10): 1830-1837.
- [4] Milan A, Leal-Taixé L, Schindler K, et al. Joint tracking and segmentation of multiple targets [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 5397-5406.
- [5] Lee J G, Han J, Li X. Trajectory outlier detection: A partition-and-detect framework [C]//IEEE 24th International Conference on Data Engineering, 2008: 140-149.
- [6] Yuan G, Xia S, Zhang L, et al. Trajectory outlier detection algorithm based on structural features [J]. *Journal of Computational Information Systems*, 2011, 11(7): 4137-4144.
- [7] Aquino A R de, Alvares L O, Renso C, et al. Towards semantic trajectory outlier detection [C]//Proceedings of GeoInfo, 2013: 115-126.
- [8] Liu Z, Pi D, Jiang J. Density-based trajectory outlier detection algorithm[J]. *Journal of Systems Engineering and Electronics*, 2013, 24(2): 335-340.
- [9] Hu H Y, Li Z H. Multi-level trajectory learning for traffic behavior detection and analysis [J]. *Journal of the Chinese Institute of Engineers*, 2014, 37(8): 995-1006.
- [10] Zhang L, Hu Z, Yang G. Trajectory outlier detection based on multi-factors[J]. *IEICE Transactions on Information and Systems*, 2014, E97.D(8): 2170-2173.
- [11] Lee J G, Han J, Whang K Y. Trajectory clustering: a partition-and-group framework [C]//Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, 2007: 593-604.
- [12] Kut A, Birant D. Spatio-temporal outlier detection in large databases [J]. *CIT Journal of Computing and Information Technology*, 2006, 14(4): 291-297.
- [13] Alvar M, Torsello A, Sanchez-Miralles A, et al. Abnormal behavior detection using dominant sets [J]. *Machine Vision and Applications*, 2014, 25(5): 1351-1368.
- [14] Wang Y, Wang D, Chen F. Abnormal behavior detection using trajectory analysis in camera sensor networks [J]. *International Journal of Distributed Sensor Networks*, 2014, 2014(1): 1-9.
- [15] Jeong H, Chang H J, Choi J Y. Modeling of moving object trajectory by spatio-temporal learning for abnormal behavior detection [C]//2011 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2011: 119-123.
- [16] Lee J G, Han J, Li X, et al. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering [J]. *Proceedings of the VLDB Endowment*, 2008, 1(1): 1081-1094.
- [17] Fontes V C, Alencar L A de, Renso C, et al. Discovering trajectory outliers between regions of interest[C]//Proceedings of GeoInfo, 2013: 49-60.
- [18] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Proceedings of Kdd Location, 1996: 226-231.
- [19] Chen J, Leung M K, Gao Y. Noisy logo recognition using line segment Hausdorff distance [J]. *Pattern Recognition*, 2003, 36(4): 943-955.
- [20] Li Weixin. Anomaly detection and localization in crowded scenes [DB/OL]. (2016-8-24) [2016-8-30] <http://www.svcl.ucsd.edu/projects/anomaly/>.
- [21] 2007 IEEE International Conference on Advanced Video and Signal based Surveillance[DB/OL]. (2007-03-02) [2016-5-30] http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html.
- [22] Alvar M, Torsello A, Sánchez, et al. Abnormal behavior detection using dominant sets [EB/OL]. (2002-12-19) [2016-5-30] <http://iit.upcomillas.es/malvar/DominantSets.php>.



红外与激光工程



《红外与激光工程》系中国光学工程学会会刊，由中国航天科工集团公司主管，天津航技术物理研究所主办，创刊于1972年，月刊，是国家新闻出版署批准的国内外公开发行的国家级学术刊物。国内刊号CN12-1261/TN，国际刊号ISSN1007-2276。中国科技论文统计源期刊，中文核心期刊。被美国“工程索引(EI) compendex”，英国“科学文摘(SA)”，美国“剑桥科学文摘(CSA)”，俄罗斯“文摘杂志(AJ)”，日本“科学技术社数据库(JST)”，荷兰“斯高帕斯(Scopus)数据库”收录。

本刊是中国航天界光电子技术领域内学术性与工程应用性集于一体的综合性刊物，主要刊登国内红外与激光技术方面的学术论文和工程研究报告，集中反映了中国光电技术的工程应用水平。

多年来，期刊被连续评为天津市一级期刊，最新影响因子达到1.144，2012年、2014年被评为“中国国际影响力优秀学术期刊”。

主要栏目

- ▶ 红外技术及应用
- ▶ 激光与光电子技术应用
- ▶ 光学遥感与空间探测
- ▶ 光通信与光传感
- ▶ 光电测量
- ▶ 太赫兹
- ▶ 光谱成像



网址: www.irla.cn
 QQ群: 275664062 275667167
 地址: 天津空港经济区中环西路58号, 300308

联系方式: 022-58168883 付宁 翟远征
 022-58168884 王红琨 刘燕荣
 022-58158885 刘俊杰 许文颖