

## 基于车牌识别系统车辆轨迹的行为异常检测

孙玉砚<sup>1,2</sup> 孙利民<sup>1</sup> 朱红松<sup>1</sup> 周新运<sup>1</sup>

<sup>1</sup>(物联网信息安全技术北京市重点实验室(中国科学院信息工程研究所) 北京 100093)

<sup>2</sup>(中国科学院大学 北京 100049)

(sunyuyan@iie.ac.cn)

## Activity Anomaly Detection Based on Vehicle Trajectory of Automatic Number Plate Recognition System

Sun Yuyan<sup>1,2</sup>, Sun Limin<sup>1</sup>, Zhu Hongsong<sup>1</sup>, and Zhou Xinyun<sup>1</sup>

<sup>1</sup>(Beijing Key Laboratory of IOT Information Security Technology (Institute of Information Engineering, Chinese Academy of Sciences), Beijing 100093)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100049)

**Abstract** Anomaly detection acts as the major direction of intelligent traffic management, but current studies may not yield the best results in the field of public safety. This paper proposes a machine-learning based technique to detect vehicle anomalies from vehicle trajectory data captured by automatic number plate recognition (ANPR) system. Our scheme is capable of detecting vehicles with the behavior of wandering round and unusual activity at specific time. Firstly the spatial and temporal quantitative indicators of vehicle activity features are extracted from historical vehicle trajectory data. The vehicles with unusual spatial feature are found and their cumulative rotation angles around the centroid of the route are calculated to detect spatial wandering round behavior. The distance from the center of clusters created by *K*-means classification algorithm based on the temporal features vectors are computed to find outliers. We collect the records from ANPR system with 315 cameras deployed in real-world for more than two months, and over 5.4 million vehicles are captured. The evaluation results based on the data set show the efficiency of the anomaly detection. More importantly, our scheme can significantly improve the detection robustness especially when the data collected by the ANPR system are noisy due to poor weather condition.

**Key words** anomaly detects; data mining; feature vectors; classification; trajectory analysis

**摘要** 目前已有很多面向智能交通管理的车辆异常行为检测方法,但是在公共安全领域的异常行为检测研究不足.为此提出了一种基于车牌识别系统车辆轨迹的行为异常检测机制,通过车牌识别系统获取抓拍记录,分析各个车辆在系统卡口的历史通行记录,提取车辆轨迹的时间空间特征,通过空间特征发现异常路线并计算路线的围绕质心累积转动角度值检测徘徊行为,用聚类算法获得时间特征的聚类中心并计算离群点检测特殊时间活跃行为.利用实际部署的车牌识别系统收集的数据测试了所提出的异常检测方法,实验结果表明该检测方法能够很好地检测面向公共安全领域的车辆异常行为,在卡口设备识别率不理想的情况下有效地提高了异常检测率.

收稿日期:2014-07-23;修回日期:2014-11-20

基金项目:国家“八六三”高技术研究发展计划基金项目(2012AA050804);国家科技支撑计划基金项目(2012BAH20B03);中国科学院战略性先导科技专项(XDA06040100)

关键词 异常检测;数据挖掘;特征向量;聚类算法;轨迹分析

中图法分类号 TP391.4

近年来中国道路上车辆数量增加异常迅猛,近10年年均增加1100多万辆,截至2013年底总量突破2.5亿辆,共有31个城市的保有量超过100万辆。交通管理压力和安全风险不断加大,监控以及处置突发危机的难度也随之加大,因此交通监控保障安全变得异常紧迫和重要。异常车辆行为检测是交通监控管理不可或缺的任务,及时发现行为异常的车辆并采取相应措施有效的处理不仅会使道路行驶更加安全,还可以中止犯罪行为或恐怖袭击活动。

## 1 相关工作

随着信息化进程的加快和设备成本的降低,海量监控摄像头被部署在公路、桥梁、路口和重点区域,各种定位装置如全球定位系统(global positioning system, GPS)和智能手机被部署在车内,车辆的行为轨迹越来越多地被记录下来。通过这些智能交通数据,人们可以对车辆轨迹的时空信息进行研究,对车辆的行为规律进行分析,监控整个交通的运行状态,一旦检测到有车辆行为异常时,就可以提前为之准备,采取相应的处置措施。异常检测一直是智能交通领域的一个研究热点,根据检测技术不同可以分为2类:基于视频的图像处理技术和基于定位的轨迹数据挖掘技术。

目前大部分车辆行为异常检测研究都是利用视频图像处理技术直接识别车辆的异常驾驶行为,例如检测监控区域发生的交通事故造成拥堵<sup>[1]</sup>,以及车辆违章逆行、变道、转向、蜿蜒、超速行驶、闯红灯等危险行为<sup>[2-6]</sup>或者检测高速公路上的行人<sup>[7]</sup>、自行车<sup>[8]</sup>等潜在的事故目标。只有少量研究针对公共安全领域,例如针对车辆在检查站或重要机构附近突然加速或者骤停等行为的检测<sup>[9]</sup>。基于视频图像处理的异常检测是分布式的,只能针对重点关注的几路视频独立的分析,目前还无法做到海量视频协同检测;另外只能监控视频可见区域内的异常行为,很难实现道路全程监控。

基于定位的轨迹数据挖掘是新兴的研究热点,车辆轨迹的定位信息获取来源包括车载GPS设备、基站定位等定位技术。基于定位装置的轨迹分析技术大部分都是将海量车辆的轨迹作时间序列切片,研究各个时间点的交通运行快照,检测其中的异常

交通状况,例如文献[10]利用出租车GPS轨迹提取出租车的移动速度,最终获得城市异常拥挤区域的分布信息;文献[11]利用车辆密度检测拥塞等异常情况;文献[12]利用GPS定位数据检测车辆的超速异常行为。基于GPS定位轨迹的异常检测虽然精度很高但成本更高,如果要面向公共安全领域监控所有车辆,目前很难给每辆车都额外安装定位装置,同时还要考虑轨迹数据的定期传输收集成本。

车牌识别技术(automatic number plate recognition, ANPR)是计算机视频图像识别技术在车辆牌照识别中的一种成熟应用,已经开始在各个城市推广部署,目前的检测率和识别率在理想情况下可达到90%以上。目前对于车牌识别系统记录的数据挖掘处理还处于起步阶段,文献[13]利用ANPR数据检测车辆尾随(也称车辆伴随)异常行为,文献[14]利用ANPR数据预测车辆行驶轨迹。

使用车牌识别系统记录的车辆轨迹分析检测异常行为的车辆具有以下优点:

- 1) 监测记录道路上行驶的所有车辆,不需要在每个车辆上安装任何额外的专业设备或硬件;
- 2) 在不同的条件下保持较高的检测率,包括恶劣天气、光线和交通拥塞状况;
- 3) 较低的系统安装和维护成本,数据收集和處理相对比较容易。

本文提出了一种基于车牌识别系统车辆轨迹的行为异常检测机制,检测面向公共安全领域的特殊时间活跃及徘徊异常行为。主要贡献包括:提出了一种车辆轨迹时空特征提取方法;提出了一种基于围绕质心累计转动角度(cumulative rotation angles around the centroid, CRAC)的异常徘徊行为检测方法,以及基于K-means聚类中心点距离(distances from centroids of K-means clustering, DCKC)的离群点检测,在卡口设备识别率不理想的情况下有效地提高了异常检测率。

## 2 轨迹特征值提取

通过处理车牌识别系统的ANPR记录数据,我们可以得到每辆车在所有系统卡口通行的历史记录,每一条记录包含抓拍的照、抓拍的时间、网关的

编号、抓拍照片中车辆的车牌号码等信息。

我们分析的主要对象是普通车辆,因此需要根据国内车辆的车牌号码编排规则<sup>[15]</sup>过滤特定的车辆。特殊车牌包括警车、军车、外籍车牌等,车牌号码中有特定的标示字符如“WJ”等。特定车辆还包括政

府公车、出租车、公交车等,其车牌号码都有预留号段。出于隐私保护的考虑,我们对过滤后的普通车辆车牌号码使用单向 Hash 函数进行变换加密。

接下来我们讨论如何从普通车辆的 ANPR 记录中提取车辆轨迹的空间时间特性,如图 1 所示:

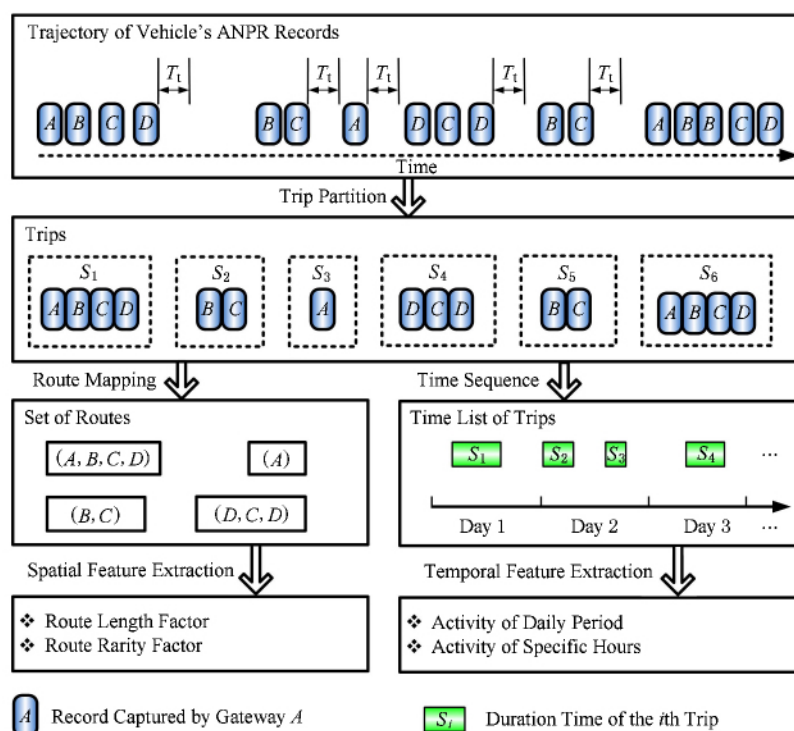


Fig. 1 Trajectory feature extraction example.

图 1 轨迹特征提取流程示例

## 2.1 行程切分

车辆轨迹用一个多维点序列表示为  $S_{\text{trajectory}} = \{p_1, \dots, p_i, \dots, p_n\}$ , 其中  $p_i$  代表第  $i$  条 ANPR 记录,  $n$  是该车辆 ANPR 记录的总数, 按照时间先后顺序排列。第  $i$  条 ANPR 记录表示为  $p_i = (t_i, g_i, l_i)$ , 其中  $t_i$  是抓拍时间,  $g_i$  是抓拍卡口编号,  $l_i$  是抓拍后识别出的车牌号码。

将轨迹作为一个整体很难提取车辆的行为特征, 我们将整个轨迹 trajectory 分割成多个子轨迹  $\{S_1, \dots, S_i, \dots, S_m\}$ , 每个子轨迹代表一个独立的短期行程 trip, 切分完之后总共  $m$  个行程。轨迹划分独立行程的标准是时间间隔, 如果 2 个连续的 ANPR 记录的时间间隔超过一个阈值  $T_t$ , 则将这 2 个记录划分到 2 个子轨迹中。第  $i$  个行程用一个多维点序列表示为  $S_i = \{p_a, \dots, p_{a+i}, \dots, p_{a+k}\}$ , 其中  $p_a$  是行程的起点记录,  $p_{a+k}$  是行程的终点记录, 数量  $k$  就是行程中记录的数量。如图 1 所示, 该车辆的轨迹被划分为 6 个行程。

卡口的车牌识别相机在检测到车辆的运动后会抓拍大量的图片。而根据实际经验, 在交通拥挤条件下一个车辆有可能被抓拍 2 次或更多次, 因此需要将行程  $S_i$  中的相同卡口的连续记录丢弃掉。例如如图 1 中, 第 6 个行程  $S_6$  序列内有 2 个卡口 B 的连续记录, 只有第 1 个记录被保留, 其他记录被丢弃。需要注意的是相同的卡口可以不连续地出现多次, 例如公交车之类的车辆行驶在往返线路中, 行程中会多次出现相同的卡口。

## 2.2 空间特征提取

车辆轨迹划分为多个行程后, 我们可以从行程的路线中提取空间特征。第  $i$  个行程  $S_i$  的路径用一个元组表示为  $P_i = (g_1, \dots, g_j, \dots, g_k)$ , 其中  $g_j$  是行程第  $j$  个记录的卡口编号。元组中卡口的数量  $k$  就是路径  $P_i$  的长度。

所有行程路径的并集就是该车辆的路线集合  $S_R$ , 表示如下:

$$S_R = \bigcup_{i=1}^m \{P_i\}. \quad (1)$$

路线集合  $S_R$  中每一个元素都是该车辆曾经行驶过的一条路径,可能只行驶过一次,也可能行驶过多次.一些有固定行驶线路的车辆,例如城市公共汽车和班车,可能会有很多个行程,但是路线集合  $S_R$  中只有少量的路径.

定义 1. 路线长度系数 (route length factor,  $RLF$ ). 路线集合  $S_p$  中最大路径长度值表示为  $L_{\max} = \max(|P_i|), P_i \in S_R$ . 最大路径长度的归一化值称为路线长度系数 ( $RLF$ ), 定义如式 (2) 所示, 其中  $T_{\text{Lenth}}$  是阈值. 车辆行驶线路越长,  $RLF$  值越接近 1.

$$RLF = \sqrt{\frac{L_{\max} - 1}{T_{\text{Lenth}} - 1}}. \quad (2)$$

定义 2. 路线稀有系数 (route rarity factor,  $RRF$ ). 对于出现的每种路径  $P$ , 将数据集中行驶过该路径的所有车辆的集合称为  $V_P$ , 路径  $P$  的稀有度  $R_P$  就是集合  $V_P$  中元素的数量, 表示为  $R_P = |V_P|$ . 路径的稀有度代表曾经有多少辆车行驶过该路径. 计算车辆路线集合  $S_R$  中所有路径的稀有度后, 我们可以获得其中的最小值  $R_{\min}$ , 表示为  $R_{\min} = \min(|R_P|), P \in S_R$ . 每个路径的路线稀有因素  $RRF$  计算方法如式 (3) 所示. 其中  $T_{\text{Rarity}}$  是稀有度上限阈值, 凡是超过该阈值的路线稀有度, 归一化之后取值为 0. 行驶过该线路的车辆越少,  $RRF$  值越接近 1.

$$RRF = \begin{cases} \frac{T_{\text{Rarity}} - R_{\min}}{T_{\text{Rarity}} - 1}, & R_{\min} \leq T_{\text{Rarity}}; \\ 0, & R_{\min} > T_{\text{Rarity}}. \end{cases} \quad (3)$$

### 2.3 时间特征提取

车辆轨迹划分为多个行程后, 我们根据每个行程的时间提取时间特征. 车辆轨迹的时间序列是一个有序的元组, 表示为  $T = (t_1, \dots, t_i, \dots, t_m)$ , 其中  $t_i$  是第  $i$  个行程第 1 条 ANPR 记录的时间. 从长期来看, 车辆常常在某些日期出现或者在某些时刻活动.

定义 3. 日常活跃度 (activity of daily period,  $ADP$ ). 假定收集的 ANPR 数据集有  $N$  天, 在  $N$  天的时间里车辆活跃度列表表示为  $f = [d_1, \dots, d_i, \dots, d_N]$ ,  $d_i \in \{0, 1\}$ . 其中  $d_i$  是车辆在第  $i$  天的活跃情况, 如果车辆轨迹的时间序列中有第  $i$  天的行程记录, 即车辆在第  $i$  个日期出现过, 则  $d_i$  取值为 1; 否则  $d_i$  取值为 0. 在此我们使用快速傅里叶变换处理车辆活跃度列表  $f$ , 计算公式如式 (4) 所示. 傅里叶系数  $F(0)$  的平方定义为每个车辆的日常活跃度  $ADP$ , 表示为  $ADP = |F(0)|^2$ .

$$F(m) = \sum_{n=0}^{N-1} f(n) \exp\left(-\frac{i2\pi mn}{N}\right). \quad (4)$$

定义 4. 特殊时间活跃度 (activity of specific hours,  $ASH$ ). 异常行为通常发生在特殊的时间段, 例如夜间活跃度异常的车辆很可能是潜在的安全威胁. 统计某车辆在特殊时间出现的行程次数  $shCount$ , 归一化为特殊时间活跃度之后表示为  $ASH = shCount/m$ , 其中  $m$  是该车辆行程的总数. 车辆在特殊时期越活跃,  $ASH$  越接近 1.

## 3 异常检测

本文提出的车辆异常行为检测主要针对的是特殊时间异常活跃和区域徘徊 2 种异常行为. 由第 2 节的定义我们可以根据轨迹信息得到车辆的一组特征值, 包括路线长度系数  $RLF$ 、路线稀有系数  $RRF$ 、日常活跃度  $ADP$ 、特殊时间活跃度  $ASH$ . 在此基础上我们提出了 2 个异常检测方法: 基于围绕质心累计转动角度的空间异常检测算法、基于  $K$ -means 聚类的时间离群点检测算法.

### 3.1 空间特征异常检测

基于围绕质心累计转动角度  $CRAC$  的空间特征异常检测算法, 首先根据车辆轨迹的空间特征值路线长度系数  $RLF$  和路线稀有系数  $RRF$  挑选候选的异常车辆; 然后计算这些车辆路线的围绕质心累计转动角度, 判断是否有徘徊行为.

当车辆轨迹的路线稀有系数  $RRF$  超过阈值  $T_{RRF}$  且路线长度系数  $RLF$  超过阈值  $T_{RLF}$  时, 该车辆将成为候选异常车辆. 意味着该车辆行驶的路线很少有其他车辆行驶过, 并且该车辆行驶的路线路径比较长, 有可能检测出徘徊行为.

候选车辆的第  $i$  个路径  $P_i = (g_1, \dots, g_j, \dots, g_k)$ , 其中  $k$  是路径长度, 路径中第  $j$  个卡口  $g_j$  的地理坐标为  $(x_j, y_j)$ , 则路径  $P_i$  质心  $O$  的地理坐标为  $(x_o, y_o)$ , 表示为  $x_o = (\sum_{i=1}^k x_i)/k, y_o = (\sum_{i=1}^k y_i)/k$ , 路径  $P_i$  围绕质心累计转动角度  $\varphi$  计算方法如式 (5) 所示:

$$\varphi = \sum_{j=1}^{k-1} \arccos\left(\frac{(x_i - x_o)(x_{j+1} - x_o) + (y_i - y_o)(y_{j+1} - y_o)}{(\sqrt{(x_j - x_o)^2 + (y_j - y_o)^2} \times \sqrt{(x_{j+1} - x_o)^2 + (y_{j+1} - y_o)^2})}\right). \quad (5)$$

在图 2 中, 车辆行驶路线为  $A \Rightarrow B \Rightarrow C \Rightarrow D \Rightarrow E \Rightarrow F$ , 路线的质心为  $O$ , 则围绕质心  $O$  的累计转动角度  $\varphi = \varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 + \varphi_5 + \varphi_6$ . 如果计算结果  $\varphi$

超过一定阈值  $T_\varphi$ , 则可以判定该车辆在该区域内徘徊行为异常。

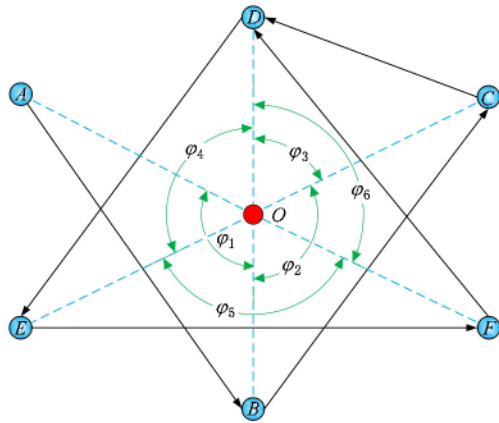


Fig. 2 CRAC for the trip with  $A \Rightarrow B \Rightarrow C \Rightarrow D \Rightarrow E \Rightarrow F$  route.

图2 CRAC 计算示例(行程路线为  $A \Rightarrow B \Rightarrow C \Rightarrow D \Rightarrow E \Rightarrow F$ )

基于围绕质心累计转动角度的空间异常检测算法 CRAC 过程描述如下:

#### 算法 1. CRAC.

输入: 普通车辆号码列表  $V$ 、车辆轨迹特征向量映射表  $F$ 、车辆路线集合映射表  $R$ 、路线长度系数阈值  $T_{RLF}$ 、路线稀有系数阈值  $T_{RRF}$ 、累计转动角度阈值  $T_\varphi$ ;

输出:  $L_V$  是具有徘徊行为的车辆列表.

- ① for each  $i$  in  $V$
- ② 从  $F$  获取  $RLF_i$ 、路线稀有系数  $RRF_i$ ;
- ③ if  $RRF_i \geq T_{RRF}$  且  $RLF_i \geq T_{RLF}$
- ④ 从  $R$  获取当前车牌号码  $i$  的路线集合  $S_R$ ;
- ⑤ for each  $P$  in  $S_R$
- ⑥ 按式(5)计算角度  $\varphi_P$ ;
- ⑦ if  $\varphi_P \geq T_\varphi$
- ⑧ 将车牌号  $i$  加入到列表  $L_V$  中;
- ⑨ endif
- ⑩ endfor
- ⑪ endif
- ⑫ endfor
- ⑬ 返回列表  $L_V$ .

### 3.2 时间特征异常检测

基于  $K$ -means 聚类中心点距离 DCKC 的时间特征异常检测算法分为 2 个阶段.

1) 选择一部分典型车辆及其轨迹时间特征值作为训练集, 车辆轨迹的时间特征向量表示为  $v = (ADP, ASH)$ ; 然后使用  $K$ -means<sup>[16]</sup> 聚类算法方法

计算训练集特征空间各个聚类的质心, 表示为  $S_c = [c_1, c_2, \dots, c_k]$ , 其中  $k$  是聚类的个数.

2) 计算数据库中每个车辆的特征值向量  $v$  与特征空间质心集合  $S_c$  中各个聚类质心  $c$  的距离, 距离计算采用欧氏距离公式如式(6)所示:

$$d(v, c) = \sqrt{(ADP_v - ADP_c)^2 + (ASH_v - ASH_c)^2}. \quad (6)$$

我们挑选那些与各个质心的距离都超过一定阈值  $T_d$  的离群点, 这些离群点就是异常检测结果. 基于  $K$ -means 聚类中心点距离的时间离群点检测算法 DCKC 描述如下:

#### 算法 2. DCKC.

输入: 全部车辆号码列表  $V$ 、训练集车辆号码列表  $S$ 、车辆轨迹特征向量映射表  $F$ 、离群点距离阈值  $T_d$ 、聚类数目  $k$ 、 $K$ -means 算法迭代次数  $r$ ;

输出: 检测的离群点列表  $L_{outlier}$ .

- ① 根据  $S$  构造训练集的特征空间数组  $A$ ;
- ② 使用  $K$ -means 算法对  $A$  进行聚类, 聚类数目为  $k$ , 迭代次数为  $r$ ;
- ③ 获取聚类中心点列表  $S_c$ ;
- ④ for each  $i$  in  $V$
- ⑤ 从  $F$  获取当前车牌号码  $i$  的特征向量  $v_i$ ;
- ⑥ 将车牌号码  $i$  加入离群点列表  $L_{outlier}$ ;
- ⑦ for each  $c$  in  $S_c$
- ⑧ 按式(6)计算  $v_i$  与  $c$  的距离  $dis$ ;
- ⑨ if  $dis \leq T_d$
- ⑩ 将  $i$  从离群点列表  $L_{outlier}$  中删除;
- ⑪ endif
- ⑫ endfor
- ⑬ endfor
- ⑭ 返回离群点列表  $L_{outlier}$ .

## 4 实验结果及分析

### 4.1 数据集

经过某市相关部门许可, 我们从该市已经实际运行的 ANPR 系统收集了从 2013-08-17—2013-10-24 的 ANPR 系统后台数据. 该 ANPR 系统包括了 314 个前端抓拍摄像头, 这些摄像头部署在 112 个卡口, 具体卡口位置如图 3 中地理信息系统 (geographic information system, GIS) 地图圆点所示. 实验使用的数据集包括了 ANPR 文本记录和相应的车辆正面抓拍照片文件.



由于车牌识别系统的识别率达不到 100%, 原始数据还需要通过数据清洗. 在这 69 d 内 ANPR 记录的总数超过 1.14 亿, 大约有 540 万个车牌号码被抓拍并成功识别. 在原始数据中未识别记录的占比为 15.1%, 重复记录的占比为 5.79%. 此外, 根据车牌号码规则过滤了超过 49 万个特定车牌号码, 保留了将近 490 万辆车的 ANPR 轨迹数据.

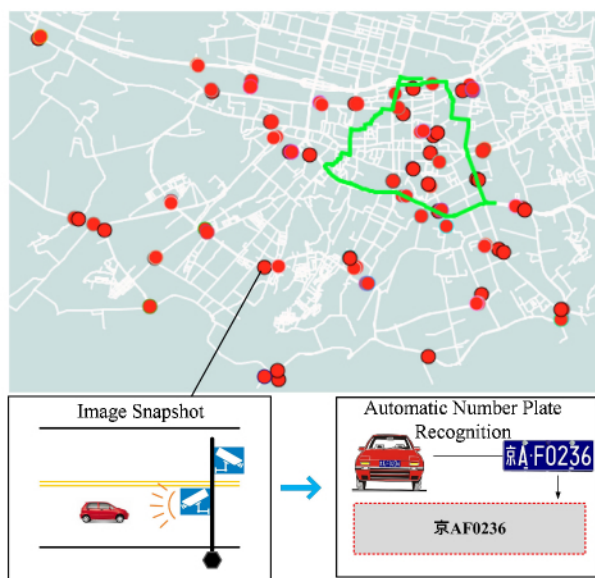


Fig. 3 ANPR system deployment.

图3 车牌识别系统实际部署

为了评估 CRAC 空间异常检测算法的性能, 我们使用模拟仿真数据集测试. 在该市 GIS 地图数据的基础上生成 5 个环形的路径, 其中一条路径如图 3 中环形路径所表示, 其测试路径中途经过 6 个卡口. 模拟测试车辆在每个环形路径上行驶 3 圈. 为了评估 DCKC 空间异常检测算法的性能, 我们使用真实数据测试, 选择  $T_d=0.5$  时检测出的离群点的轨迹数据.

时间特征异常检测 DCKC 算法训练数据的具体选择过程直接影响其聚类中心, 进而影响异常检测结果. 道路上出现的出租车、公交车及公务车已经被过滤, 本文使用人工手动标注图像中车辆的车型, 筛选出了比较常见的 5 类车辆类型, 包括本市上下班通勤家用的小客车、外地出差旅游的小客车、本市货运的卡车、外地长途货运的卡车、载客的大中型客车. 通过查看车辆的证明抓拍图片及历史轨迹, 最终挑选 1 万辆正常车辆, 包括 0.3 万辆市内小型客车、0.2 万辆外地小型客车、0.2 万辆本市卡车、0.2 万辆外地卡车、0.1 万辆大中型客车.

## 4.2 参数确定

在该区域内的实际驾驶测试中, 在高峰时刻最远的 2 个卡口之间的行驶时间约 30 min, 这意味着在一个短途行程的车牌识别记录间隔不超过 30 min. 此外我们计算并统计了所有车辆 ANPR 轨迹的连续记录的时间间隔, 其分布情况如图 4 所示, 约 67.9% 的时间间隔值小于 30 min. 因此, 分割行程的时间间隔阈值  $T_i=30$  min.

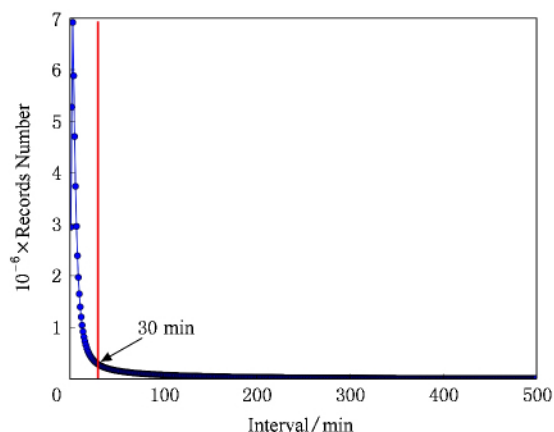


Fig. 4 Numbers of intervals between records when step is 1 min.

图4 步长为 1 min 时轨迹记录的时间间隔数值分布情况

不同时段交通流量统计如图 5 所示, 90% 的车辆如家用轿车等在日间行驶, 而夜间行驶的车辆较少, 特殊时间活跃度的特殊时间定义为夜间 0~6 点.

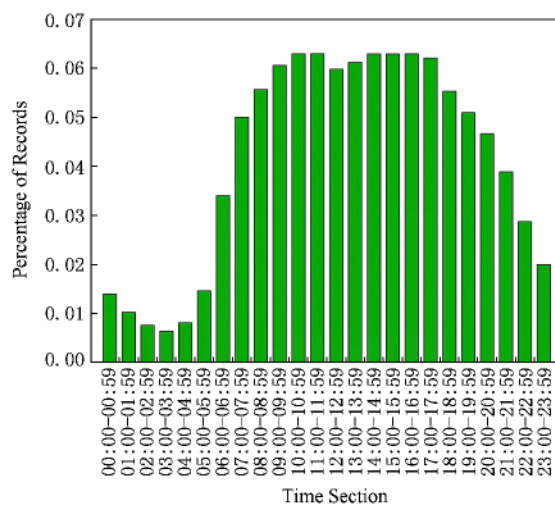


Fig. 5 Percentage of ANPR records numbers during different time.

图5 不同时段 ANPR 记录所占百分比

根据所有 180 多万种行程路径长度数值的统计结果, 90% 以上的路径长度为 1, 不足 1% 的路径长

度超过 20,因此式(2)中计算路线长度系数的阈值  $T_{Lenth}=20$ . 式(3)中计算路程稀有度的阈值  $T_{Rarity}=100$ ,即如果一个路线曾经有超过 100 辆车行驶过,则该路线的稀有度为 0. 空间特征异常检测算法中,挑选候选异常车辆的路线稀有系数阈值  $T_{RRF}=0.99$ ,路线长度系数阈值  $T_{RLF}=0.2$ ,累计转动角度阈值  $T_{\varphi}=4\pi$ ,即车辆围绕质心行驶超过 2 圈,则判定该路线是异常的徘徊路线. 时间特征异常检测算法中阈值  $T_d=0.5$ ,即如果车辆的特征向量距离其他分簇中心的最小距离超过 0.5,则判定该车辆在特定时刻活动异常.

### 4.3 异常检测性能分析

车辆 ANPR 轨迹提取特征值的分布情况如图 6 所示,70%左右的车辆其  $ADP$  数值接近于 0,意味着只出现过 1 d;90%的车辆  $ASH$  数值为 0,也就是从未在特定时刻行驶过;98%的车辆  $RLF$  数值小于 0.1,意味着这些车辆的行驶路线长度基本不超过 4 个卡口;88%的车辆  $RRF$  数值接近于 0,意味着这些车辆的行驶路线都是常见路线.

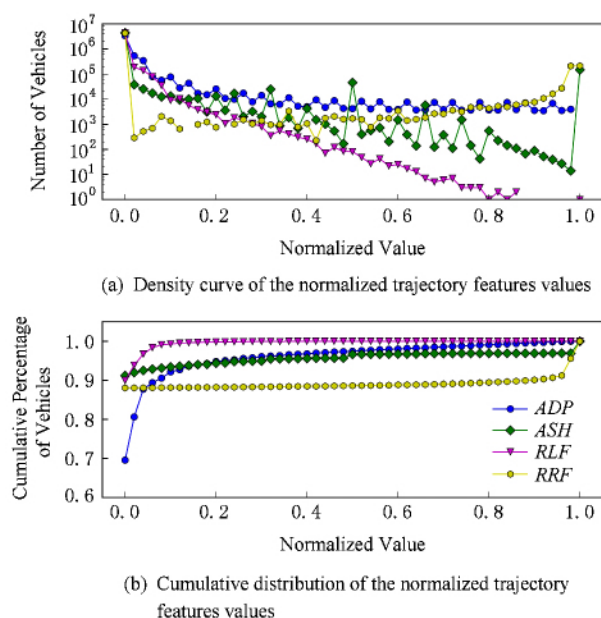


Fig. 6 Distribution of normalized trajectory features values when value step is 0.02.

图 6 车辆 ANPR 轨迹提取的特征值分布情况(步长为 0.02)

DCKC 时间离群点检测算法结果的示意图如图 7 所示,图 7 中使用  $K$ -means 算法对训练集内 1 万辆车的时间特征向量  $v=(ADP,ASH)$  进行分类,分类数  $k=5$ ,迭代次数  $r=10$ .

图 7 中  $ASH$  数值最大的一个分簇基本是外地

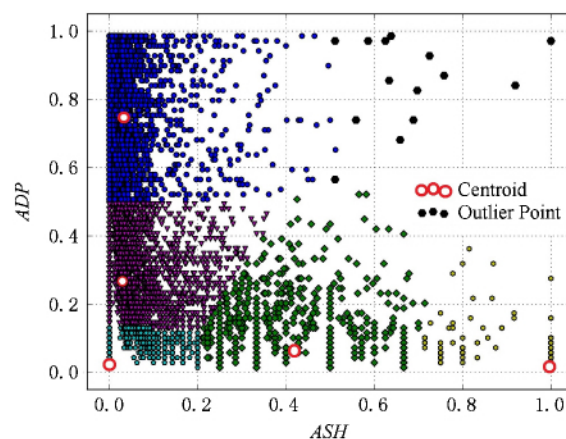


Fig. 7 DCKC temporal anomaly detection results.

图 7 DCKC 时间离群点检测结果示例

长途卡车,通常在夜间行驶; $ASH$  数值接近 0.5 的分簇基本是本市卡车;左下角  $ADP$  和  $ASH$  数值都较低的分簇基本是外地小客车,出现频率较低且通常在日间行驶;左上角的分簇基本是本市小客车, $ADP$  数值较高而  $ASH$  数值较低;左侧中间的分簇基本是本市大中型客车. 图 7 中白色空心圆点是各个分类的中心点;黑色正六边形点是检测出的离群点,其与各个中心点的距离超过 0.5. 使用该算法对其余 490 万辆车的轨迹数据进行检测,共检测出 340 个离群点.

CRAC 空间异常检测算法结果的示意图如图 8 所示. 图 8 中黑色的正六边形就是线路稀有系数与路线长度系数满足阈值同时围绕质心累计转动角度大于  $T_{\varphi}$  的异常车辆. 使用该算法共检测出异常车辆 12 万个,异常的路径 150 万多条. 由于在实验中检测到的区域徘徊车辆数量较多,因此我们选择了其中一部分活跃度比较高的车辆,分析这些车辆的

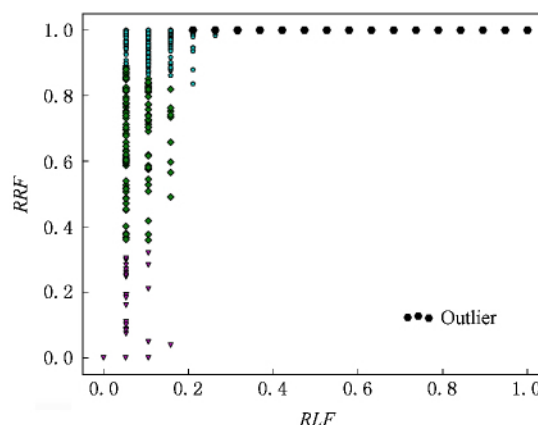


Fig. 8 CRAC spatial anomaly detection results.

图 8 CRAC 空间异常检测结果示例

抓拍图片后发现其中比较活跃的异常徘徊行为是无营业执照的黑出租车载客造成的。

基于车牌识别系统的异常检测的检测率与系统的识别率密切相关. 如图 9 所示, 我们测试了不同识别概率下, CRAC 空间异常检测算法以及 DCKC 时间离群点检测算法的检测率. 为了对比我们还实现了基于最长重复子串 (longest repeated subsequence) 的徘徊异常检测算法, 如果车辆的线路中最长重复子序列的长度大于等于 3 则判定车辆行为异常.

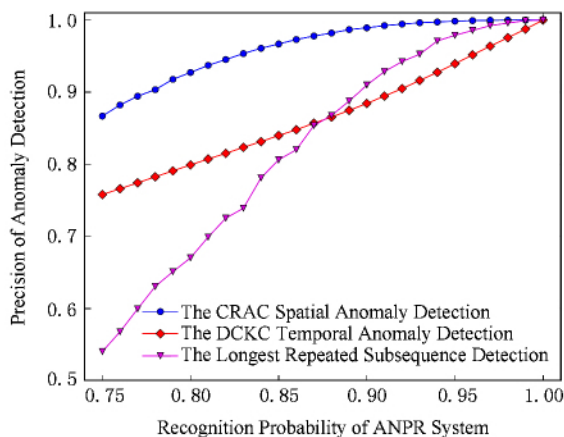


Fig. 9 Performance of proposed anomaly detection algorithm.

图 9 异常检测算法的检测率

性能测试中假设车牌识别系统的准确率为  $p$ , 也就是每次车辆行驶经过卡口时被抓拍和识别的概率为  $p$ , 因此以概率  $p$  保留每一条 ANPR 记录, 最终生成一个新的轨迹序列作异常检测. 图 9 中系统识别概率  $p$  从 0.75 到 1.0, 步长等于 0.01, 在每个识别概率  $p$  下生成 10 万个新测试集, 计算其中依然被检测出异常的轨迹数量就是相应的检测率.

由图 9 可以看出, CRAC 空间异常检测算法受车牌识别系统识别率的影响较小, 因此在恶劣天气或者夜间等对车牌识别系统不利的条件下也有较高的检测率, 能够检测出徘徊异常行为. DCKC 时间离群点检测算法与识别率基本成线性关系, 较低的识别率会导致离群点检测结果发生较大差异. 而作为对比的基于最长重复子串 LRS 的徘徊异常检测算法受车牌识别系统识别率的影响很大, 在车牌识别率为 75% 的情况下检测率只有 54.02%

## 5 结束语

本文提出了一种基于车牌识别系统车辆轨迹的

行为异常检测机制, 检测面向公共安全领域的特殊时间活跃及徘徊异常行为. 主要贡献包括: 提出了一种车辆轨迹时空特征提取方法, 提出了一种基于围绕质心累计转动角度 CRAC 的异常徘徊行为检测方法以及基于  $K$ -means 聚类中心点距离 DCKC 的离群点检测, 在卡口设备识别率不理想的情况下有效地提高了异常检测率.

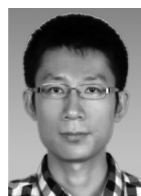
虽然本文提出异常行为检测取得了比较精确的检测结果, 但仍然存在一些问题和改进之处. 车牌识别系统的检测率和识别率高低会影响异常行为检测的准确性. 因此未来的工作将集中于处理分析大量未识别的 ANPR 记录, 保证车辆轨迹的完整性和准确性.

## 参 考 文 献

- [1] Barria J A, Thajchayapong S. Detection and classification of traffic anomalies using microscopic traffic variables [J]. IEEE Trans on Intelligent Transportation Systems, 2011, 12 (3): 695-704
- [2] Saruwatari K, Sakaue F, Sato J. Detection of abnormal driving using multiple view geometry in space-time [C] //Proc of the 4th IEEE Intelligent Vehicles Symp. Piscataway, NJ: IEEE, 2012: 1102-1107
- [3] Sang Haifeng, Wang Hui, Wu Danyang. Vehicle abnormal behavior detection system based on video [C] //Proc of the 5th IEEE Int Symp on Computational Intelligence and Design. Piscataway, NJ: IEEE, 2012: 132-135
- [4] Srivastava S, Ka K N, Delp E J. Co-ordinate mapping and analysis of vehicle trajectory for anomaly detection [C] //Proc of the 12th IEEE Int Conf on Multimedia and Expo. Piscataway, NJ: IEEE, 2011: 1-6
- [5] Hao Jiuyue, Hao Sheng, Li Chao, et al. Vehicle behavior understanding based on movement string [C] //Proc of the 12th IEEE Int Conf on Intelligent Transportation Systems. Piscataway, NJ: IEEE, 2009: 1-6
- [6] Qiu Lingyun, Han Jun, Gu Ming. Abnormal behavior detection for highway vehicle based on lane model [J]. Journal of Computer Applications, 2014, 34(5): 1378-1382 (in Chinese)  
(邱凌媛, 韩军, 顾明. 车道模型的高速公路车辆异常行为检测方法[J]. 计算机应用, 2014, 34(5): 1378-1382)
- [7] Bouttefroy P, Beghdadi A, Bouzerdoum A, et al. Markov random fields for abnormal behavior detection on highways [C] //Proc of the 2nd European Workshop on Visual Information Processing. Piscataway, NJ: IEEE, 2010: 149-154



- [8] Ryan D, Denman S, Fookes C, et al. Textures of optical flow for real-time anomaly detection in crowds [C] //Proc of the 8th IEEE Int Conf on Advanced Video and Signal-Based Surveillance. Piscataway, NJ: IEEE, 2011: 230-235
- [9] Srivastava S, Delp E. Standoff video analysis for the detection of security anomalies in vehicles [C] //Proc of the 39th IEEE Applied Imagery Pattern Recognition Workshop. Piscataway, NJ: IEEE, 2010: 1-8
- [10] Siyuan L, Yunhuai L, Ni L, et al. Detecting crowdedness spot in city transportation [J]. IEEE Trans on Vehicular Technology, 2013, 62(4): 1527-1539
- [11] Bacon J, Bejan A, Beresford A, et al. Using real-time road traffic data to evaluate congestion [G] //LNCS 6875: Proc of Dependable and Historic Computing. Berlin: Springer, 2011: 93-117
- [12] Mohamad I, Ali M, Ismail M. Abnormal driving detection using real time global positioning system data [C] //Proc of 2011 IEEE Int Conf on Space Science and Communication. Piscataway, NJ: IEEE, 2011: 1-6
- [13] Homayounfar A, Ho A, Zhu N, et al. Multi-vehicle convoy analysis based on ANPR data [C] //Proc of the 4th IEEE Int Conf on Imaging for Crime Detection and Prevention. Piscataway, NJ: IEEE, 2011: 1-5
- [14] Wu Zijun, Yu Chongchong, Sun Limin, et al. A method of prediction on driving track of specific vehicles in potential group [J]. Journal of Application Research of Computers, 2014, 31(7): 1-6 (in Chinese)  
(吴子珺, 于重重, 孙利民. 一种针对特定车辆潜在群体的行驶轨迹预测方法[J]. 计算机应用研究, 2014, 31(7): 1-6)
- [15] Wikimedia. Vehicle registration plates of China [EB/OL]. 2010 [2014-07-23]. [http://en.wikipedia.org/wiki/Vehicle\\_registration\\_plates\\_of\\_China](http://en.wikipedia.org/wiki/Vehicle_registration_plates_of_China)
- [16] Macqueen J. Some methods for classification and analysis of multivariate observations [C] //Proc of the 5th Berkeley Symp on Mathematical Statistics and Probability. Oakland, CA: University of California Press, 1967: 281-297



**Sun Yuyan**, born in 1982. PhD candidate and assistant professor in the Institute of Information Engineering, Chinese Academy of Sciences. Member of China Computer Federation. His main research interests include data-mining, ad-hoc network, RFID, wireless sensor network, Internet of things.



**Sun Limin**, born in 1966. PhD. Professor and PhD supervisor in the Institute of Information and Engineering, Chinese Academic of Sciences. Member of China Computer Federation. His main research interests include new generation of Internet technology, wireless mobile networks, mobile IP, etc.



**Zhu Hongsong**, born in 1973. PhD. Associate professor in the Institute of Information and Engineering, Chinese Academic of Sciences. Member of China Computer Federation. His main research interests include wireless system security, biometric security, and the security on Internet of things.



**Zhou Xinyun**, born in 1979. PhD. Senior engineer in the Institute of Information and Engineering, Chinese Academic of Sciences. Member of China Computer Federation. His main research interests include wireless mobile networks, mobile IP, multimedia integrated access network, etc.