



基于约束高斯混合模型的车道信息获取

唐炉亮¹ 杨 雪¹ 靳 晨¹ 刘 章¹ 李清泉^{1,2}

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

2 深圳大学土木工程学院空间信息智能感知与服务深圳市重点实验室,广东 深圳,518060

摘要:针对现有车道级道路信息获取方法大多存在数据采集成本高、更新周期长、数据处理难度大等缺点,提出了一种基于浮动车数据(floating car data,FCD)的城市车道数量信息快速获取方法。首先根据浮动车数据的空间分布特征,利用 Delaunay 三角网方法对数据进行优选,通过探测优选后浮动车数据覆盖的宽度间接得到道路宽度;然后将一部分已知车道数量及浮动车数据覆盖宽度的路段作为训练样本,分析其车道数量和浮动车数据覆盖宽度之间的关系构建基本分类器;最后按照待测路段的浮动车数据分布宽度查找基本分类器,获取待测路段可能存在的若干个车道数量类型候选值,并利用约束高斯混合模型对最终车道数量类型进行确认。实验结果表明,该方法实现了从低精度浮动车数据中快速获取车道数量信息,提取精度达到了 82.3%。

关键词:浮动车数据;基本分类器;约束高斯混合模型;车道数量

中图法分类号:P208

文献标志码:A

道路信息(道路级别信息和车道级别信息)是智能辅助驾驶系统应用的基础,其获取方法和精细程度随数据源和需求不断发展演变^[1]。现有的道路信息获取方法按数据源不同,主要分为从图像数据中提取道路信息^[2]、从激光点云数据中获取道路信息^[3-4]以及从时空 GPS 轨迹数据中挖掘道路信息^[5-6]三类。利用时空 GPS 轨迹大数据提取道路信息是目前研究的热点话题。Rogers^[5]利用时空 DGPS(difference global positioning system)轨迹数据提取道路中心线以及车道边界线。随后在 Rogers 等研究的基础上,利用时空 GPS 轨迹数据获取道路信息逐渐发展成为一种端对端的模式^[5-6],细化到城市道路的车道信息、复杂交叉口几何结构、转向信息等。文献[6]提出一种脱离原始地图的道路信息获取模式,该模式首先采用轨迹分类和融合方法从大量 DGPS 轨迹数据中提取道路级别信息,然后利用高斯混合模型从归属于每一条路段的大量轨迹数据中提取车道信息。利用高质量 DGPS 轨迹数据提取道路信息虽然具有精细程度高且精度可靠的特点,但是也存在数据采集成本高、周期长且无法反映因为交

通管制、城市建设和社会功能改变等引起的道路信息变化的缺陷。

安装 GPS 接收机的出租车通常被称为浮动车,其海量轨迹数据布满城市任何道路,是一种采集速度快、成本低、蕴含丰富车道信息的数据来源。利用低精度浮动车数据获取高精度道路信息是一个难点,而目前大量的研究仍然停留在道路级别信息的获取^[7-18]。Uduwaragod 等^[12]利用核密度聚类方法从车载 GPS 轨迹数据中探测车道数量和车道线位置。本文针对浮动车数据的特点,提出了一种城市车道数量信息快速获取方法。该方法首先根据 FCD(floating car data)的空间分布特征,利用 Delaunay 三角网方法优选数据,通过探测优选后 FCD 覆盖的宽度间接得到道路宽度;然后将一部分已知车道数量及 FCD 覆盖宽度的路段作为训练样本,分析其车道数量和 FCD 覆盖宽度之间的关系构建基本分类器;最后根据待测路段的 FCD 分布宽度查找基本分类器,获取待测路段归属的若干个车道数量类型候选值,并利用约束高斯混合模型确认最终车道数量类型。实验结果表明,该方法实现了车道数量信息的快

速获取,可以及时反映交通管制、城市建设车道功能改变等引起的车道数变化。

1 浮动车数据预处理

1.1 浮动车数据分布特征

数万辆出租车每天行驶在城市大街小巷,以60 s左右的采样频率24 h不间断采集,数据涵盖的城市交通信息内容丰富、实时性强,并且海量轨迹布满城市任何道路和车道。但是,由于采集环境和装备等限制,使得原始FCD中夹杂了大量噪音信息。根据FCD轨迹方向信息,将道路上行驶方向相反的FCD进行区分,然后对每一条路段上单向行驶的FCD进行特征分析。由于GPS误差服从正态分布,所以每一个轨迹点的位置与实际值的接近程度也服从正态分布。这种由单条轨迹分布蔓延至整条道路横截面上全部轨迹的分布状态叠加,使得分布在道路上轨迹点的密度要明显高于道路面以外漂移点的密度,并且位于道路中心线处的密度值最高,然后密度值依次向两边衰减。按照FCD在道路横截面分布的特点,本

文采用Delaunay三角网方法对轨迹密度值低的数据进行剔除,把留下的点作为路面点。

1.2 基于Delaunay三角网粗差处理

通过分析Delaunay三角网的空间剖分特性及FCD空间分布特征,本文提出采用Delaunay三角网方法对原始FCD进行粗差处理。

采用Delaunay三角网进行粗差处理,主要以三角形面积和边长为约束条件进行处理。如图1所示,图1(a)为区域内若干条路段上分布的浮动车数据;图1(b)为分布在该区域内的所有FCD的Delaunay三角网。其中对于FCD高密度聚集的区域,三角网的边长和面积比较小,而对于FCD低密度聚集的区域,其三角网出现狭长三角形或者面积非常大三角形。针对三角网在这两种聚集情况下出现的反差,本文提出以三角形面积和边长作为约束条件剔除原始FCD中夹杂的粗差(图1(c))。三角形面积和边长阈值为对应粗差剔除比率下的结果。同时,为了避免一些路段上分布的GPS轨迹数据量稀少而被误作为粗差删除,本文选择以路段分割段为基本单元进行处理,并且根据GPS轨迹方向将道路单双向路段进行区分。

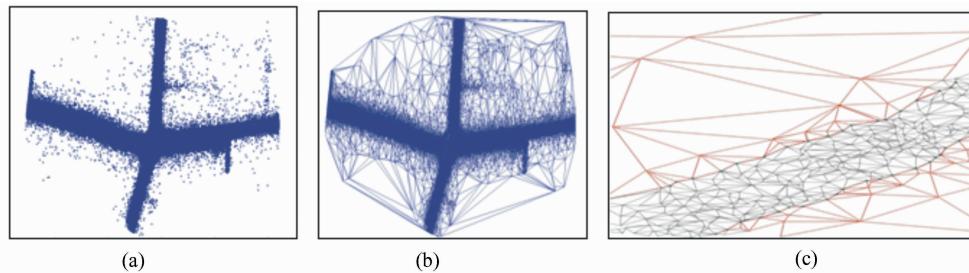


图1 浮动车数据的Delaunay三角网

Fig. 1 D-TIN of Floating Car Data

2 车道信息获取方法

对比高精度、高采样频率且专门采集的GPS轨迹数据,FCD具有海量性与实时性两个非常明显的特点。本文从这两个特点出发,对一定周期内分布在目标路段上的FCD进行覆盖宽度和轨迹特征探测,提出了一种从FCD中提取车道数量信息的新方法。

2.1 基本分类器构建

2.1.1 浮动车数据分布宽度探测

为了获取FCD在路面分布的宽度,本文采用自适应宽度探测算法对目标路段上覆盖的所有FCD进行宽度探测(图2)。

假设目标路段Road总长为 L ,按照分割尺度 h 依次可以得到 $n = (L/h)$ 个分割段,记为 $\{x_1,$

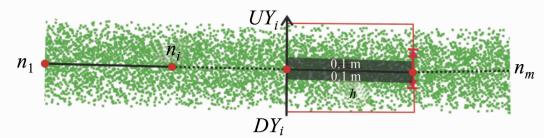


图2 浮动车数据覆盖宽度探测

Fig. 2 Detection of Floating Car Data Covered Width

$x_2, \dots, x_n\}$,其分割点记为 $\{n_1, n_2, \dots, n_m\}$ 。根据文献[13]提出的轨迹聚类方法获取各分割段上覆盖的FCD中心线,以中心线作为相应分割段上坐标系的横轴,各分割点即为每段分割段上对应坐标系的坐标原点, UY_i 和 DY_i 分别为每一个分割段内移动窗口沿坐标系上下移动后的纵坐标值,其测宽结果为 $\{w_1, w_2, w_3, \dots, w_m\}$ 。其算法具体步骤如下:

```
/* Initialization */
```

坐标原点: n_1 ; 横坐标轴: x_1 的中心线; 纵坐标轴: UY_i

$=0$; $DY_i=0$; 移动窗口: 移动窗口长为 h ; 移动窗口宽为 0.1 m ; 窗口内 FCD 占总体 FCD 比例为 0 。

```
/* Assignment */
for each  $x_j$  ( $j=1, 2, \dots, n$ ), do
repeat
```

沿纵轴上下平移移动窗口, 每次计算移动窗口内 FCD 占分割段上分布的总体 FCD 的比例(Proportion), 并进行累加。until $\text{Proportion} == 100\%$

```
set  $Dw_i = \max |UY_i| + |\max DY_i|$ ;
set 坐标原点更新为  $n_{t+1}$ ;  $UY_i = 0$ ;  $DY_i = 0$ ; ( $t=1, 2 \dots m, i=0$ )
end for
```

2.1.2 训练数据的选择

城市路网及浮流动车数据采集环境非常复杂。本文为了提高待测路段的分类度, 将训练数据的选择定位在有遮挡类和无遮挡类两个类别, 同时训练数据与待测数据在总体样本占的比例为 $2:1$ 。

2.1.3 基本分类器构建

FCD 在道路面的分布宽度一定程度上表征

表 1 基本分类器
Tab. 1 Primary Classifier

有遮挡类/无遮挡类	车道数量	浮流动车数据分布宽度/m			样本个数
		最小值	平均值	最大值	
单向两车道	2	$Dw_2 \min$	$Dw_2 \text{average}$	$Dw_2 \max$	n_1
单向三车道	3	$Dw_3 \min$	$Dw_3 \text{average}$	$Dw_3 \max$	n_2
单向四车道	4	$Dw_4 \min$	$Dw_4 \text{average}$	$Dw_4 \max$	n_3
单向五车道	5	$Dw_5 \min$	$Dw_5 \text{average}$	$Dw_5 \max$	n_4

由于 FCD 自身精度, 即使经过 Delaunay 三角网粗差处理后, 噪音信息仍然会存在。所以相邻类型的车道之间, 其 FCD 宽度存在重叠区域。为了获得待测路段分割段的最终车道数量, 本文提出了采用限制性高斯混合模型。

2.2 基于约束高斯混合模型的车道数获取

本文根据 GPS 轨迹点误差分布特征, 提出利用约束高斯混合模型确定最终的车道数量。其原理就是通过模拟道路上浮流动车数据的分布状态, 比较不同高斯成分个数组合后的高斯混合模型优劣, 选取模型最优时所对应的高斯成分个数作为该路面分布的车道数量。对于待测路段, 当获悉其分割段的浮流动车数据分布宽度后, 参照基本分类器即可确定该分割段可能归属的若干个车道类型, 然后利用约束高斯混合模型确定最终的车道数量。约束高斯混合模型可以定义为:

$$p(x) = \sum_{j=1}^k \omega_j \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_j)^2}{2\sigma^2}\right) \quad (1)$$

式中, k 是高斯混合模型中高斯成分的个数, 即分割段根据基本分类器初步确定的若干个车道数量候选值; ω_j 是每一个高斯成分的权重, 其中 $i=1,$

即实际道路面的宽度信息, 而实际路面宽度则与车道数量息息相关。根据现有国内的车道建设相关规范和标准, 道路车道数量主要有单向双车道、单向三车道、单向四车道、单向五车道等 4 种。不同等级的道路车道建设宽度也会不同, 例如根据我国道路建设标准, 3 级以上道路的车道宽度一般为 $3.5 \sim 3.75\text{ m}$ 。假设实验区域内有 m 条已知道路车道数量的目标路段, 其 FCD 覆盖宽度探测值分别为 W_{i_1}, \dots, W_{i_m} , 将这些目标路段的覆盖宽度和对应车道数量进行统计分析, 构建车道数量与浮流动车数据分布宽度对应关系的基本分类器, 如表 1 所示。表 1 中, \min 表示浮流动车数据分布宽度的最小值, average 表示浮流动车数据分布宽度的平均值; \max 表示浮流动车数据分布的最大值。 $Dw_{i \min}, Dw_{i \text{ average}}, Dw_{i \max}$ 分别表示每一种车道数量其浮流动车数据分布宽度所对应的具体值, 其中 $i=2, 3, 4, 5$ 。

由于 $w_1+w_2+\dots+w_k=1$; σ 是属于每个高斯成分的标准差, 由于每一个车道宽度一致, 所以每一个高斯成分的 σ 都相同; μ_1, \dots, μ_k 是每一个高斯成分的均值。构建高斯混合模型前, 需要分解各类参数值。

采用最大期望算法(expectation maximization, EM)来解算未知高斯参数 $\theta_{j(m)}(\omega_{j(m)}, \mu_{j(m)}, \sigma^{(m)})$, 其中 m 则为迭代计算的次数。高斯参数的初始值 $\omega_{j(0)}, \mu_{j(0)}, \sigma^{(0)}$ ($j=1, \dots, k$) 可以被设定为 $\omega_{1(0)}=\omega_{2(0)}=\dots=\omega_{k(0)}=1/k$, 而 $\mu_{1(0)}, \dots, \mu_{k(0)}$ 则为分割段有 k 个车道时, 每个车道的中心线, $\sigma^{(0)}$ 根据道路建设标准设定为 1.75 m 。

E 步:

$$\gamma_{i,j}^{(m)} = \frac{\omega_j^{(m)} \Phi(x_i \mid \mu_j^{(m)}, \sigma_j^{(m)})}{\sum_{j=1}^k \omega_j \Phi(x_i \mid \mu_j^{(m)}, \sigma_j^{(m)})} \quad (2)$$

$$\varphi(x \mid \mu, \sigma) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3)$$

$$n_j^m = \sum_{i=1}^k \gamma_{i,j}^{(m)} \quad (4)$$

M 步

$$\omega_j^{(m+1)} = \frac{n_j^{(m)}}{n} \quad (5)$$

$$\mu_j^{(m+1)} = \frac{1}{n_j^{(m)}} \sum_{j=1}^k \gamma_{i,j}^{(m)} \quad (6)$$

$$\sigma_j^{(m+1)} = \sqrt{\frac{1}{n_j^{(m)}} \sum_{j=1}^k \gamma_{i,j}^{(m)} (x_i - \mu_j^{(m+1)})^2} \quad (7)$$

根据 EM 算法,可以得到在不同车道数量(k)条件下,未知高斯参数 $\theta_{j(m)}$ ($\omega_{j(m)}, \mu_{j(m)}, \sigma_{j(m)}$)的值。为了获取分割段最终的车道数量,本文提出采用模型熵对 k 属于不同值时所对应的约束高斯混合模型进行评价,如式(8)所示。当熵值最小时表示该模型最优,其对应的 k 值即为该分割段的车道数量,如式(9)所示,其中 Num 为待测分割段最终的车道数量, n 为分布在待测路段分割段上所有浮动车数据点的个数。

$$F(k) = \left(-\frac{1}{n} \sum_{j=1}^k \lg p(x_i | \theta_k) + R(D, \theta_k) \right) \quad (8)$$

$$\text{Num} = \operatorname{argmin} F(k) \quad (9)$$

其中, $R(D, \theta_k)$ 为赤池信息量准则(AIC), 定义为:

$$R_{\text{AIC}}(D, \theta_k) = \frac{d}{n}, d = k + 2 \quad (10)$$

3 实验

本文以武汉市出租车采集的浮动车数据作为研究对象,选取 1 000 条路段作为样本,其中 665 条路段为训练样本,其余 335 条路段为待测样本,如图 3 所示。

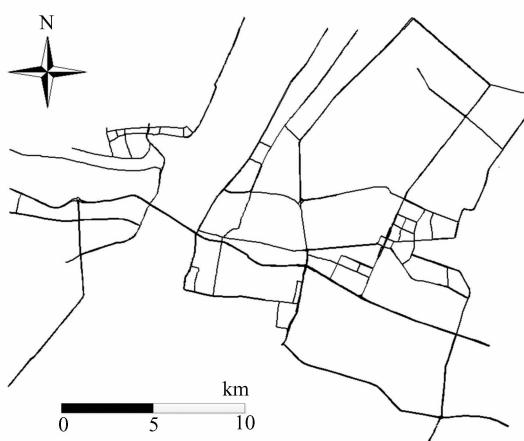


图 3 实验路网数据

Fig. 3 Experimental Road Network

3.1 浮动车数据预处理

1) 选取非高峰期时段浮动车数据,并剔除载客状态为 0 以及行驶速度为 0 的信息记录,减少

因交通特征不同和驾驶员驾驶行为导致的 FCD 在道路横截面上分布异常;

2) 对 GPS 轨迹分方向处理,将位于同一条道路不同行驶方向的轨迹进行区分,并将路段上分布的 FCD 进行等距离分割(实验中分割距离为 10 m);

3) 采用基于 Delaunay 三角网的数据剔除方式,剔除多余漂移点(对一部分已知路面宽度的样本按照路段环境分为无遮挡类和有遮挡类,然后采取不同的粗差剔除比率进行粗差剔除。统计结果表明,对于无遮挡类数据,当剔除比率为 25% ~ 30% 时,浮动车数据分布宽度与实际路面宽度平均差异在 2 m 以内,也即路面宽度探测的精度为 2 m;对于有遮挡类数据,当剔除率为 35% ~ 40% 时,浮动车数据分布宽度与实际路面宽度平均差异在 2 m 以内,也即路面宽度探测精度为 2 m)。

按照 Delaunay 三角网粗差处理统计结果,对实验区内其他路段分割段逐一进行处理,如图 4 所示。图 4(a)为覆盖在基础路网上的原始浮动车数据;图 4(b)是经过 Delaunay 三角网处理后,红色轨迹点为保留下来的有效点,而黑色轨迹点则为被删除的粗差数据。将经过处理后的 FCD 与影像图叠加如图 4(c)和图 4(d),其中图 4(c)是剔除后的数据与真实影像叠加;图 4(d)为图 4(c)的放大图,从图 4(d)中可以发现 FCD 均匀地分布在道路边界线以内。

3.2 浮动车采集周期分析

训练数据的选取需要注意以下几个方面。

首先,路段类型要涵盖城市内部高等级道路、低等级道路,同时需要涉及高架桥、跨江大桥道路;

其次,浮动车数据采集周期选取需要满足一定条件。通常由于浮动车数量限制和城市不同区域客流量影响,不同路段每一天覆盖的 FCD 数量不同,为了使目标路段宽度探测更加贴近真实情况,本文通过随机抽取一部分路段,对不同周期内所覆盖的 FCD 进行宽度探测,如图 5 所示。

图 5(a)为实验路段中一小部分路段 1~8 d 不同周期浮动车覆盖宽度探测结果,当采集周期由 7 d 变为 8 d 时,路段上分布的 FCD 宽度不再发生变化。图 5(b)是同一条路段在相邻周期内探测宽度变化,7 d 内浮动车数据分布宽度与 8 d 内浮动车数据分布宽度变化基本为 0。因此,本文在后续车道数量信息探测过程中,选择一周以上的 FCD 作为实验样本。

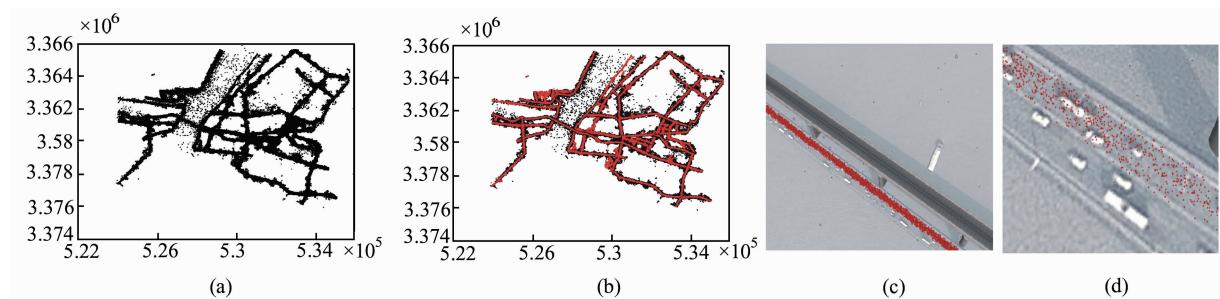


图 4 浮动车数据异常值处理

Fig. 4 Process of Floating Car Data

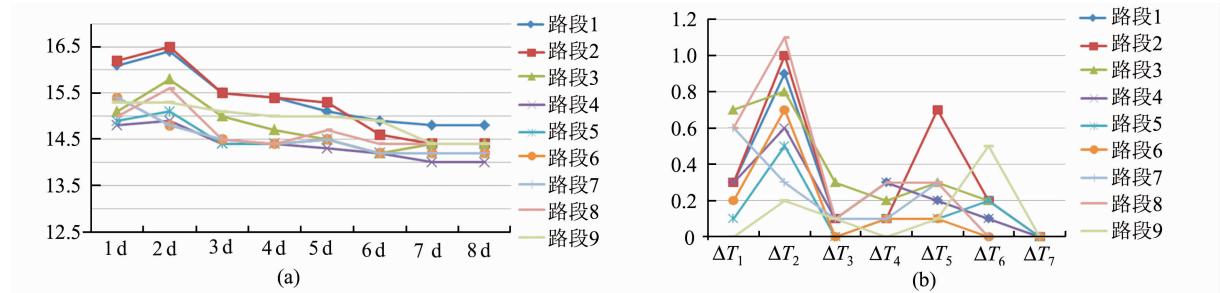


图 5 FCD 覆盖宽度周期研究

Fig. 5 Research on Floating Car Data Collecting Period

3.3 车道数量探测与讨论

对 665 条路段分别进行 FCD 覆盖宽度和其在道路横截面状态进行探测, 并将探测结果与路段对应实际车道数量进行统计构建基本分类器, 如表 2 所示。表 2 中前者宽度为有遮挡类路段分

布宽度, 后者宽度为无遮挡类路段分布宽度。例如, 表中的 7.9/7.5, 7.9 为有遮挡类时, 单向两车道上浮动车数据分布宽度的最小值; 7.5 为无遮挡类时, 单向两车道上浮动车数据分布宽度的最小值。

表 2 实验区车道数量探测基本分类器

Tab. 2 Basic Classifier of Lane Numbers Detection in Experimental Area

有遮挡类/无 遮挡类	车道数量	浮动车数据分布宽度/m			训练样本个数
		最小值	平均值	最大值	
单向两车道	2	7.90/7.5	10.02/8.9	12.20/11.3	分割段 3 841/2 108
单向三车道	3	9.96/10.3	12.15/12.1	19.8/15.4	分割段 9 660/3 213
单向四车道	4	13.45/13.4	17.82/14.2	25.92/18.9	分割段 13 230/986
单向五车道	5	17.8/16.7	22.78/20.2	27.79/22.8	分割段 1 650/523

假设待测路段分割段的浮动车数据分布宽度为 12.1 m, 其路段环境为由遮挡类, 参照实验区车道数量探测基本分类器, 可以大致推测该分割段可能属于单向两车道和单向三车道, 可分别计算当 $k=2$ 和当 $k=3$ 时的高斯参数。

当 $k=2$ 时, $w_1=0.5068$, $w_2=0.4931$; $\sigma=1.7996$; $\Delta\mu=4.2828$; $F(2)=0.0051$ 。

当 $k=3$ 时, $w_1=0.4277$, $w_2=0.5274$, $w_3=0.048$; $\sigma=1.8197$; $\Delta\mu=3.9012$; $F(3)=0.07$, 其模拟图如图 6 所示。图 6(a)表示当 $k=2$ 时的高斯混合模型分布; 图 6(b)表示当 $k=3$ 时的高斯混合模型分布。

根据计算结果, 当 $k=2$ 时, 模型熵 $F(2)$ 值最

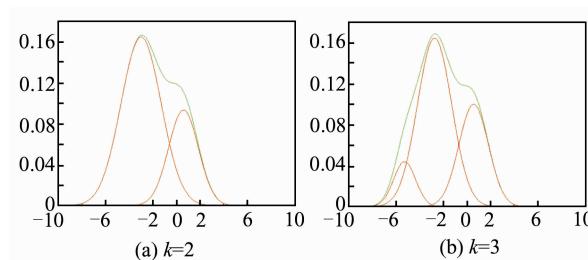


图 6 分割段对应的高斯混合模型

Fig. 6 Gaussian Mixture Model for Segment

小, 模型最优, 也即该分割段的车道数量为单向两车道。实验采用本文所提方法对其余目标路段进行车道数量探测。结果表明, 采用本文方法探测车道数量的总体精度为 82.3%, 同时存在 17.7%

的错误率。通过分析车道数量判断错误的路段,发现其原因如下。

1)交叉口红绿灯、转向限制及其他交通特征引起不同车道车流量差异,导致接近交叉口附近路段车道数量判断错误率较高,从而影响整体车道数判断的精度;

2)由于现有FCD丢失了高程信息,所以暂时无法区分高架桥与其下遮挡的普通道路上的轨迹数据,尤其对于叠合部分比较多的双层路段,导致其车道数量判断结果与实际结果不一致;

3)包含隧道的路段,由于FCD丢失,造成无法获取其车道数量信息。

4 结语

本文在分析目前国内外车道信息探测方法的基础上,提出了一种基于约束高斯混合模型的车道信息获取新方法。实验结果表明,该方法从浮动车数据的空间分布特征出发实现了目标路段车道数量的提取,缩短了车道数量信息获取的周期,降低信息提取成本。此外,本文提出的以路段分割段为基本单元的车道数量探测模式还可以应用到城市路段变化快速检测方面。然而,对于交叉口附近路段及高架桥与普通路段叠合的部分,车道数量探测结果精度不高。未来研究中将会进一步对本文方法进行改进和完善,并将研究如何从浮动车数据中提取各车道中心线和转向信息。

参考文献

- [1] Hillel A B, Lerner R, Levi D, et al. Recent Progress in Road and Lane Detection: A Survey[J]. *Machine Vision and Applications*, 2014, 25(3): 727-745
- [2] Chen Long, Li Qingquan, Mao Qingzhou. Lane Detection and Following Algorithm Based on Imaging Model[J]. *China Journal of Highway and Transport*, 2011, 24(6):96-102(陈龙,李清泉,毛庆洲. 基于成像模型的车道线检测与跟踪方法[J]. 中国公路学报, 2011, 24(6):96-102)
- [3] Lin Zongjian, Liu Zhengrong. Methods and Prospects of Road and Linear Structure Extraction from Remote Sensing Image[J]. *Geomatics and Information Science of Wuhan University*, 2003, 28(1):90-93 (林宗坚,刘政荣. 从遥感影像提取道路信息的方法评述[J]. 武汉大学学报·信息科学版, 2003, 28(1):90-93)
- [4] Wu Xiaobo, Yang Liao, Shen Jinxiang, et al. Road Extraction from High-resolution Remote Sensing Images with Spatial Continuity[J]. *Acta Geodaetica et Cartographica Sinica*, 2011, 36(11): 1 298-1 301 (吴小波,杨辽,沈金祥,等. 利用空间连续性进行高分辨率遥感影像道路提取[J]. 武汉大学学报·信息科学版, 2011, 36(11):1 298-1 301)
- [5] Huang Xianfen, Li Na, Zhang Fan, et al. Automatic Power Lines Extraction Method from Airborne LiDAR Point Cloud[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(12): 1 563-1 569 (黄先锋,李娜,张帆,等. 利用LiDAR点云强度的十字剖分线法道路提取[J]. 武汉大学学报·信息科学版, 2015, 40(12):1 563-1 569)
- [6] Fang Lina, Yang Bisheng. Automated Extracting Structural Roads from Mobile Laser Scanning Point Clouds[J]. *Acta Geodaetica et Cartographica Sinica*, 2014, 42(1):260-267(方莉娜,杨必胜. 车载激光扫描数据的结构化道路自动提取方法[J]. 测绘学报, 2014, 42(1):260-267)
- [7] Wagstaff K, Cardie C, Rogers S, et al. Constrained k-means Clustering with Background Knowledge [C]. *The Eighteenth International Conference on Machine Learning*, San Francisco, 2001
- [8] Chen Yihua, Krumm J. Probabilistic Modeling of Traffic Lanes from GPS Traces [C]. *The 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, California, 2010
- [9] Tang Luliang, Yang Xue, Kan Zihan, et al. Lane-level Road Information Mining from Vehicle GPS Trajectories Based on Naïve Bayesian Classification [J]. *ISPRS International Journal of Geo-Information*, 2015, 4(4):2 660-2 680
- [10] Tang Luliang, Yang Xue, Dong Zhen, et al. CLRIC: Collecting Lane-based Road Information via Crowdsourcing[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 9(17):2 552-2 562
- [11] Tang Luliang, Yang Xue, Kan Zihan, et al. Traffic Lane Numbers Collecting Based on the Naïve Bayesian Classification[J]. *China Journal of Highway and Transport*, 2016, 29(3):116-123 (唐炉亮,杨雪,阚子涵,等. 一种基于朴素贝叶斯分类的车道数量探测[J]. 中国公路学报, 2016, 29(3):116-123)
- [12] Zhen Yu. Introduction to Urban Computing[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(1):1-13(郑宇. 城市计算概述[J]. 武汉大学学报·信息科学版, 2015, 40(1):1-13)
- [13] Li Qingquan, Huang Lian, Tan Wenxia. Mass GPS Tracking Data Compression Method Based on Road Features [J]. *Geomatics and Information Science*

- of Wuhan University*, 2008, 33(4):337-340 (李清泉, 黄练, 谭文霞. 基于道路特征的海量 GPS 监控数据压缩方法[J]. 武汉大学学报·信息科学版, 2008, 33(4):337-340)
- [14] Li Qingquan, Yin Jianzhong, He Fenqin. A Coverage Rate Model of GPS Floating Car for Road Networks[J]. *Geomatics and Information Science of Wuhan University*, 2009, 34(6):715-718 (李清泉, 尹建忠, 贺奋琴. 面向道路网的 GPS 浮动车覆盖率模型研究[J]. 武汉大学学报·信息科学版, 2009, 34(6):715-718)
- [15] Tang Luliang, Chang Xiaomeng, Li Qingquan. Public Travel Route Optimization Based on Ant Colony Optimization Algorithm and Taxi GPS Data[J]. *China Journal of Highway and Transport*, 2011, 24(2):89-95 (唐炉亮, 常晓猛, 李清泉. 基于蚁群优化算法与出租车 GPS 数据的公众出行路径优化[J]. 中国公路学报, 2011, 24(2):89-95)
- [16] Lu Shiwei, Fang Zhixiang, Shaw Shihlung, et al. Quantitative Analysis of the Effects of Spatial Scales on Intra-urban Human Mobility[J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(9): 1 199-1 204 (鲁仕维, 方志祥, 萧世伦, 等. 城市群体移动模式研究中空间尺度影响的定量分析[J]. 武汉大学学报·信息科学版, 2016, 41(9): 1 199-1 204)
- [17] Uduwaragoda A, Perera A S, Dias S A D. Generating Lane Level Road Data from Vehicle Trajectories using Kernel Density Estimation[C]. The 16th International IEEE Annual Conference on Intelligent Transportation Systems, New York, 2013
- [18] Cao L, Krumm J. From GPS Traces to A Routable Road Map[C]. The 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Washington D C, 2009

Traffic Lane Number Extraction Based on the Constrained Gaussian Mixture Model

TANG Luliang¹ YANG Xue¹ JIN Chen¹ LIU Zhang¹ LI Qingquan^{1,2}

1 State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing,
Wuhan University, Wuhan 430079, China

2 Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

Abstract: The present methods are expensive and time-consuming that can't keep up with changes in the roads due to construction or disasters. Instead of this, an attractive alternative is proposed to acquire lane numbers from Floating Car Data collected by taxis. Firstly, the Delaunay triangulation network method is used to optimize the raw GPS trajectories by considering the spatial distribution characteristics of floating car data, and then the road width is set by detecting the trajectories covered width. The primary classifier is established from training data that shows the relation between lane numbers and road width, and some candidate lane numbers can be inferred by detecting the trajectory covered width with the primary classifier. The final lane numbers are determined by fitting GPS traces using the Constrained Gaussian mixture model. Experiments results show the method proposed in this paper can acquire lane numbers from low-quality floating car data and the overall accuracy is 82.3%.

Key words: floating car data; primary classifier; constrained Gaussian mixture model; lane number

First author: TANG Luliang, PhD, professor, specializes in T-GIS, GIS-transportation. E-mail: tll@whu.edu.cn

Corresponding author: YANG Xue, PhD candidate. E-mail: yangxue@whu.edu.cn

Foundation support: The National Natural Science Foundation of China, Nos. 41271442, 40801155, 41571430, 41671442; the Open Research Fund of the Academy of Satellite Application, No. 2014_CXJJ-DSJ_02; the Fund of the Shenzhen Science Technology Bureau.