

基于道路结构特征的智能车单目视觉定位

俞毓锋^{1,2} 赵卉菁^{1,2} 崔锦实^{1,2} 查红彬^{1,2}

摘要 高精度定位是实现自动驾驶的关键. 在城市密集区域, 全球定位系统 (Global positioning system, GPS) 等卫星定位系统受到遮挡、干扰、多路径反射等影响, 无法保障自动驾驶所需的定位精度. 视觉定位技术通过图像特征匹配进行位置估计, 被广泛研究. 然而传统基于特征点的方法容易受到移动目标的干扰, 在高动态交通场景中的应用面临挑战. 在结构化道路场景中, 车道等线特征普遍存在, 为人类驾驶员的视觉理解与决策提供重要线索. 受该思路的启发, 本文利用场景中的三垂线和点特征构建道路结构特征 (Road structural feature, RSF), 并在此基础上提出一个基于道路结构特征的单目视觉定位算法. 本文利用在北京市区的典型路口、路段、街道等场所采集的车载视频数据进行实验验证, 以同步采集的高精度 GPS 惯性导航组合定位系统数据为参照, 与传统视觉定位算法进行比较. 结果表明, 本文算法在朝向估计上明显优于传统算法, 对环境中的动态干扰有更高的鲁棒性. 在卫星信号易受干扰的区域, 可以有效地弥补 GPS 等定位系统的不足, 为满足自动驾驶所需的车道级定位要求提供重要的技术手段.

关键词 单目相机, 视觉定位, 道路结构特征, 三垂线

引用格式 俞毓锋, 赵卉菁, 崔锦实, 查红彬. 基于道路结构特征的智能车单目视觉定位. 自动化学报, 2017, 43(5): 725–734

DOI 10.16383/j.aas.2017.c160413

Road Structural Feature Based Monocular Visual Localization for Intelligent Vehicle

YU Yu-Feng^{1,2} ZHAO Hui-Jing^{1,2} CUI Jin-Shi^{1,2} ZHA Hong-Bin^{1,2}

Abstract Precise localization is an essential issue for autonomous driving applications, while global positioning system (GPS)-based systems are challenged to meet requirements such as lane-level accuracy, especially in crowded urban environment. This paper introduces a new visual-based localization approach in dynamic traffic environments, focusing on structured roads like straight roads or intersections. Such environments show several line segments on lane markings, curbs, poles, building edges, etc., which demonstrate the road's longitude, latitude and vertical directions. Based on this observation, we define a road structural feature (RSF) as sets of segments along three perpendicular axes together with feature points, and propose an RSF based monocular visual localization method. Extensive experiments are conducted on three typical scenarios, including highway, intersection and downtown streets. Results show better accuracy compared with a state-of-the-art visual localization method using feature points. We demonstrate that the proposed method can help improving localization accuracy in GPS restricted area, and discuss the remained challenges leading to future studies.

Key words Monocular camera, visual localization, road structural feature (RSF), perpendicular axes

Citation Yu Yu-Feng, Zhao Hui-Jing, Cui Jin-Shi, Zha Hong-Bin. Road structural feature based monocular visual localization for intelligent vehicle. *Acta Automatica Sinica*, 2017, 43(5): 725–734

智能车或自动驾驶技术受到广泛的关注, 精确的定位技术是其重要环节. 面向城市道路环境, 车道级定位是实现智能车轨迹规划或决策的关键. 全球定位系统 (Global positioning system, GPS) 被

广泛运用于汽车导航系统. 单点 GPS 定位误差为数米, 且受到信号遮挡、干扰、多路径反射等影响, 在城市密集区域, 无法保障自动驾驶所需的定位精度. 而以视觉里程计为代表, 通过环境数据匹配进行运动估计的视觉定位方法可以有效提高相对定位精度. 但城市环境中来自车辆、行人等动态目标的干扰给传统视觉定位算法带来了巨大挑战. 如图 1 (a) 所示, 在一帧图像中, 共检测了 785 个特征点, 其中有 277 个在周围车辆上. 移动干扰特征点的比率已经超出了视觉定位算法中外点滤除的极限, 导致定位结果产生偏差.

在人类驾驶员视觉理解过程中, 相比于特征点, 车道等结构化道路信息起到更重要作用. 以图 1 (b) 为例, 图像中包括沿道路方向、水平垂直道路方向和竖直方向的三种线特征, 这些信息为驾驶员有效定

收稿日期 2016-05-23 录用日期 2016-12-27
Manuscript received May 23, 2016; accepted December 27, 2016
国家高技术研究发展计划 (863 计划) (2012AA011801), 国家自然科学基金 (61573027) 资助

Supported by National High Technology Research and Development Program of China (863 Program) (2012AA011801), and National Natural Science Foundation of China (61573027)

本文责任编辑 王飞跃
Recommended by Associate Editor WANG Fei-Yue

1. 北京大学信息科学技术学院 北京 100871 2. 北京大学机器感知与智能教育部重点实验室 北京 100871

1. School of Electronic Engineering and Computer Science, Peking University, Beijing 100871 2. Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871

位自身车辆与道路的关系提供重要线索. 受该思路启发, 本文提出一个基于道路结构特征 (Road structural feature, RSF) 的智能车单目视觉定位算法, 用于解决高动态结构化道路环境下的定位问题. 本文结合场景中的三垂线和点特征, 定义道路结构特征, 并在此基础上建立定位算法, 包括道路结构预测、特征提取和车辆运动估计. 本文利用安装在智能车平台上的摄像头在北京市区的路口、拥堵路段、密集街道等典型复杂场景进行数据采集, 对本文提出的算法进行实验分析, 以同步采集的高精度 GPS 惯性导航组合定位系统 (GPS/惯导) 数据为参照, 与传统视觉定位算法进行比较, 验证本文算法的有效性.

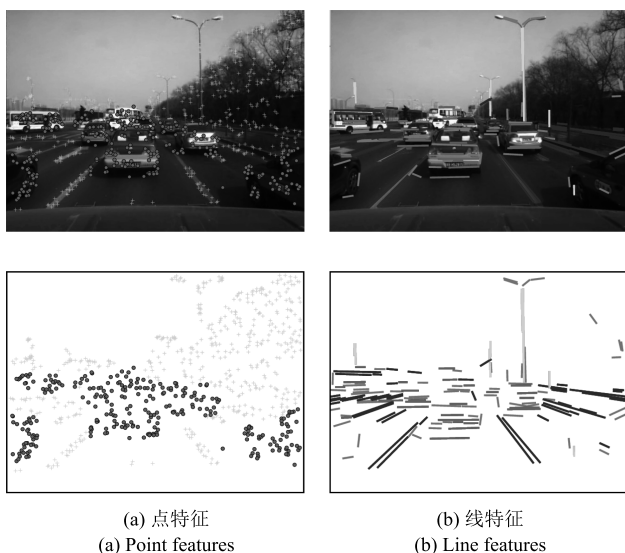


图 1 单帧图像的点特征和线特征
Fig. 1 Point feature and Line feature of one on-road image

本文结构如下: 第 1 节围绕近年来国内外视觉定位的前沿研究, 分析其在大范围高动态交通场景应用中所面临的关键难点; 第 2 节给出道路结构特征的定义, 并概述基于道路结构特征的定位算法; 第 3 节详细描述定位算法中的细节; 第 4 节给出实验结果, 并与高精度 GPS/惯导和传统视觉定位方法作比较; 第 5 节总结本文算法并给出今后的研究展望.

1 相关研究

基于视觉的定位算法被广泛用于智能车定位系统, 视觉里程计 (Visual odometry)^[1] 是最重要的领域之一. 视觉里程计利用单个或者多个相机的输入信息, 通过前后帧特征匹配, 估计本体的运动. 有关视觉里程计的研究最早可以追溯到 1980 年. 当时, Moravec 等^[2] 利用一个可滑动相机获取视觉信息, 并以此作为输入, 完成了机器人的室内导航. 1987 年, Matthies 等^[3] 提出了利用视觉做定位的

概念, 设计了包含特征提取、特征匹配与跟踪、运动估计的理论框架. 2004 年, Nistér 等^[4] 首次明确了视觉里程计这一概念, 并在 2006 年详细描述了视觉里程计系统^[5]. 他们设计了一种实时的视觉里程计系统, 真正意义上实现了机器人室外运动导航, 同时提出了两类视觉里程计的实现途径和流程, 即单目视觉和立体视觉的方法, 之后的研究大多在此基础上进行改进.

通用的视觉里程计以特征点匹配为基础, 定位结果受到累计误差和移动物体干扰两方面因素的影响. 针对累计误差, 光束平差法 (Bundle adjustment)^[6] 是一个常用的处理方法, 通过最小化三维特征点在不同图像上的重投影误差, 同时优化三维特征点和相机位置, 提高了定位精度, 但计算代价较高. 为降低时间复杂度, Mouragnon 等^[7] 使用局部光束平差法, 在处理累计误差和计算代价上做了一个平衡. 针对移动物体的干扰, 传统的方法将移动物体上的特征点当作噪声, 利用随机采样一致性方法 (Random sample consensus, RANSAC)^[8] 来滤除噪声, 但这只能处理少量干扰问题. Ozden 等^[9] 建立了一个多模型的定位与场景重建方法, 将视频中的特征点进行聚类, 提取移动物体, 去除干扰. Kundu 等^[10] 将光束平差法与多模型方法结合, 有效提高了定位精度, 但在移动物体较多的环境下, 有效性受到制约. 近几年的视觉里程计的研究将定位和建图结合起来, 建立视觉 SLAM 方法, 例如 lsd-SLAM^[11]、Orb-SLAM^[12] 等. 此类方法有效减少了累计误差, 但由于需要在定位的同时重建三维场景, 导致高动态环境下移动物体的干扰增加.

另一类视觉定位算法则采用与地图匹配的方式进行定位. Geiger 等^[13] 利用视觉里程计生成的轨迹和二维地图上的道路进行形状比对, 确定车辆实际位置. 算法可定位到所处道路, 但无法到达车道级精度. Ramalingam 等^[14-15] 直接利用城市的三维模型, 将图像中的特征点和线投影到三维模型中, 直接获取对应的三维位置, 避免了三维重建过程中的误差. 算法提高了定位精度, 但目前三维模型的可获取性不够广泛, 算法的局限性高. 另一些研究者则采用智能车预先建立地图进行定位. Achar 等^[16] 利用词袋 (Bag-of-words) 的方法, 在需要定位的街道上预先行驶若干圈采集数据, 生成特征地图, 在线定位时, 将图像特征点和数据库中特征点进行匹配并定位. Ziegler 等^[17] 同时建立了道路的几何地图与特征点云地图, 提高了在线匹配的精度. 该智能车完成了 103 千米的自动驾驶, 但地图建立存在大量手工操作, 人力和时间成本高.

传统的视觉定位算法使用特征点进行定位, 在高动态交通场景中, 特征点易受移动车辆的干扰, 导致定位结果出现偏差. Geiger 等^[18] 将图像特征投影到路面上, 利用标志线的角点对定位进行优化, 适

用于低动态场景, 但无法处理拥堵环境. 考虑到道路场景中线特征的朝向特殊性, Zhang 等^[19] 利用室外的垂直线特征进行定位, 将原三维定位问题转换为二维定位问题, 降低了复杂度, 但垂直线特征不通用, 局限性大. Barinova 等^[20-21] 则将线段进行聚类, 并检测消失点, 利用消失点之间的几何约束确定三垂线. 该算法只在部分图片中进行测试, 缺乏有效性验证, 且计算代价高. 本文对三垂线的检测进行优化, 并将其用于结构化道路场景定位算法, 极大地提高了定位估计的朝向稳定性.

2 系统概述

本文提出的定位算法框架如图 2 所示. 以下首先给出道路结构特征的定义, 在此基础上简述算法的主要步骤, 包含道路结构的预测、道路结构特征提取和基于道路结构特征的位置估计.

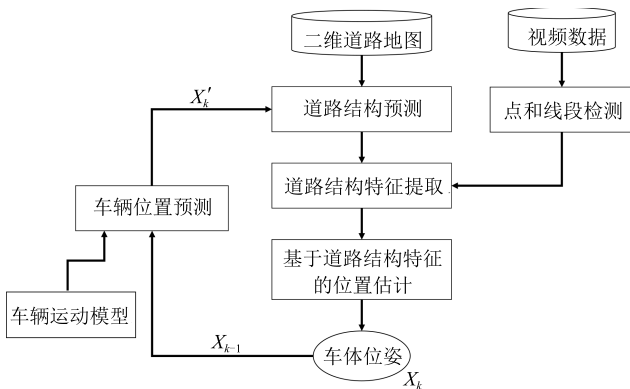


图 2 算法框架
Fig. 2 System outline

2.1 道路结构特征定义

道路结构特征是一组带有语义信息, 描述典型结构化道路场景特征的基本元素集合. 在本研究中, 道路结构特征由图像中对应于道路行进方向、水平垂直道路方向和竖直方向的三组线特征 L_1, L_2, L_3 , 以及点特征 P 构成, 即 $Z_{\text{RSF}}: \{L_1, L_2, L_3, P\}$. 其中线特征 L_1, L_2, L_3 在道路坐标系下相互垂直, 在本文中简称三垂线.

2.2 道路结构预测

随着定位服务 (Location based services, LBS) 技术的发展, 地图作为数据基础被广泛建立, 部分已开放使用 (例如 OpenStreetMap). 这些地图中包含道路中心线、路口节点等信息, 为预测局部道路结构提供依据. 如图 3 所示, 对于直道场景, 根据智能车当前的定位信息和相机标定参数, 可在地图中确定相机的可视区域, 结合地图中道路中心线信息, 建立局部道路坐标系. 利用车辆位置以及相机参数等信息, 将局部道路坐标系映射到图像中, 预测对应于道

路行进方向、水平垂直道路方向和竖直方向的三组线特征 L_1, L_2, L_3 的方向. 对于路口场景, 根据地图中道路中心线, 将相机可视区域分割为一系列小区域, 每个小区域都根据道路中心线定义自身的道路局部坐标系, 再映射到图像中, 将图像分为相应的区域, 并在每个区域中预测对应的三组线特征方向.

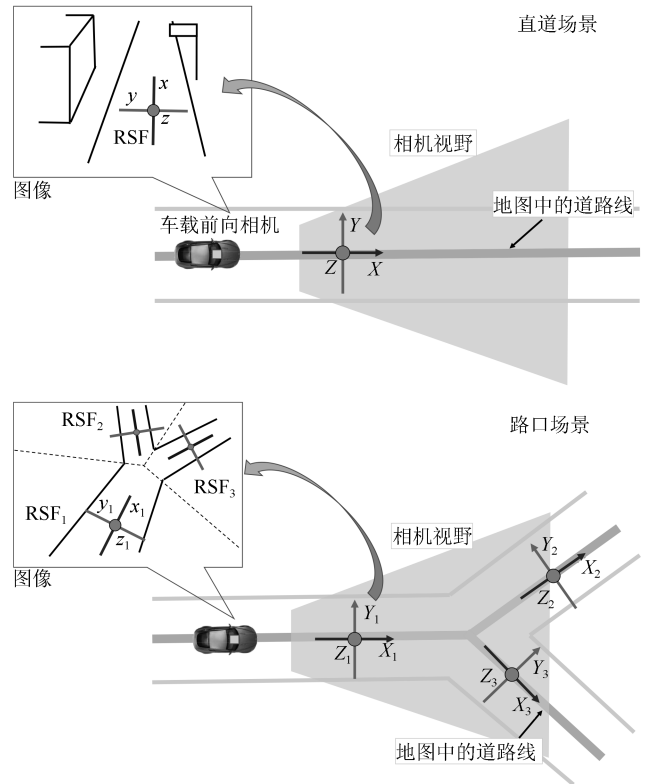


图 3 道路局部三垂线结构预测, 包含直道和路口两种情况
Fig. 3 Map-based RSF prediction, including straightway and intersection

2.3 道路结构特征提取

道路结构特征的提取以线段检测算法和特征点检测算法为基础, 实时检测图像中的线段和特征点, 构建道路结构特征. 对于线特征, 定义线和道路坐标系三方向的距离函数, 以此将图像中的线段进行分类. 通常情况下, 路面上的车道线、道路边界、车辆侧面的线段等会被分成第一类 L_1 ; 路面上的停止线、车辆后侧的横向线段、路牌上的横向线段会被分成第二类 L_2 ; 路灯、路边建筑物、大型车辆等的竖直方向线段会被分为第三类 L_3 . 对于点特征, 直接根据通用特征点检测结果, 生成特征点集合 P .

2.4 基于道路结构特征的位置估计

车辆的位置估计采用随机采样一致性方法 (RANSAC). 在检测到所有道路结构特征 Z_{RSF} 后, 基于随机采样构建样本集合 $\{C_{\text{RSF}}^i\}$. 一个有效的采样 C_{RSF} 包含三条线段和两个点. 其中三条线段中

两条相互平行, 并且垂直于第三条.

$$C_{\text{RSF}} = \{l_1, l_2 \in L_u, l_3 \in L_v, u \neq v, p_1, p_2 \in P\} \quad (1)$$

图4所示为当 $u = 1, v = 3$ 的一个样例. l_1 和 l_2 为沿道路方向上的线段, l_3 为竖直方向上的线段, p_1 和 p_2 为图像中任意两个特征点. 在利用采样 C_{RSF} 进行位置姿态 x 求解的过程中, 使用其中的线特征求解旋转, 再使用点特征估计平移, 平移尺度的估计则使用车速信息累计求解. 位置姿态求解结果使用后验概率 $p(Z_{\text{RSF}}|x_k^i)p(x_k^i|x_{k-1})$ 进行评价. 其中 $p(Z_{\text{RSF}}|x_k^i)$ 为观测模型, $p(x_k^i|x_{k-1})$ 为车辆运动模型. 通过一定次数的随机采样 C_{RSF} (本文实验中采样次数为 200), 最优位置估计结果 \hat{x}_k 通过最大化后验概率求得.

$$\hat{x}_k = \arg \max_{x_k^i \sim C_{\text{RSF}}^i} \{p(Z_{\text{RSF}}|x_k^i)p(x_k^i|x_{k-1})\} \quad (2)$$

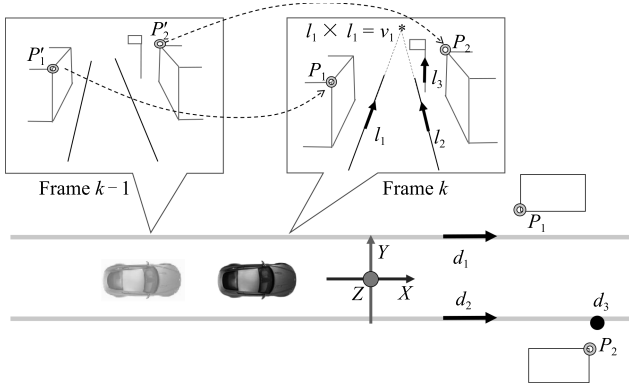


图4 道路结构特征采样样例, 包含三条线段和两个特征点
Fig.4 Sample of RSF candidate, including three line segments and two points

3 算法细节

3.1 坐标系定义

车体与相机的坐标系定义如图5所示. 定义 $x_v = \{R_v, s_v\}$ 为车体在世界坐标系下的位置姿态. 其中, R_v 为旋转矩阵, s_v 为平移向量. 车体坐标系定义为右手系准则, 其中原点位于车辆后轴中心, x, y, z 的方向分别为车辆右侧, 前方和上方. 相机固定在车顶前侧上方, 定义 $x_c = \{R_c, s_c\}$ 为相机的在世界坐标系下的位置姿态. 相机坐标系定义为右手系准则, 其中原点为相机中心, x, y, z 的方向分别为相机右侧、下方和前方. 相机和车体的刚性固定, 车体坐标系到图像映射 $P_{vc} = K[R_{vc}|s_{vc}]$ 可通过标定得到. 其中 K 为相机内参数, R_{vc} 和 s_{vc} 分别为外参数中的旋转矩阵和平移向量, 即车体在相机坐标系下的旋转矩阵和平移向量.

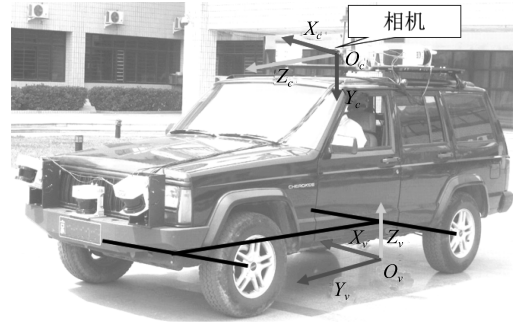


图5 车体与相机的坐标系定义

Fig.5 Coordinate frames of the vehicle and the camera

3.2 道路结构预测

定义 $e_1 = (1, 0, 0)^T$, $e_2 = (0, 1, 0)^T$, $e_3 = (0, 0, 1)^T$ 为 x, y, z 方向的单位向量, 车体在世界坐标系下的三个方向可表示为

$$\begin{bmatrix} d_x & d_y & d_z \end{bmatrix} = R_v \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} \quad (3)$$

其中, $d_y = R_v e_2$ 为车体朝向. 定义 $\{R_{v,k}, s_{v,k}\}$ 为车体在 k 时刻的位置姿态, $\{\hat{R}_{v,k}, \hat{s}_{v,k}\}$ 为对应时刻的预测结果. 假设车体的运动在朝向上变化均匀, 则车辆在 k 时刻的旋转矩阵可以通过前一帧的旋转矩阵和前后两帧之间的相对旋转求解

$$\hat{R}_{v,k} = R_{rel} R_{v,k-1} = (R_{v,k-1} R_{v,k-2}^T) R_{v,k-1} \quad (4)$$

平移的预测则将车辆在两帧间的位移进行求和. 在较短的时间内, 位移方向近似为前后帧车辆朝向的角平分线的方向, 即

$$\hat{s}_{v,k} \approx s_{v,k-1} + \frac{d_{y,k-1} + \hat{d}_{y,k}}{2} \int_{t_{k-1}}^{t_k} v_t dt \quad (5)$$

其中, $d_{y,k-1}$ 为车体 $k-1$ 时刻的朝向, $\hat{d}_{y,k}$ 为车体 k 时刻朝向预测结果, v 为车速. 相机在 k 时刻的位置姿态 $\{R_c, s_c\}$ 通过车体的位置姿态预测结果和标定参数进行求解.

$$\hat{R}_{c,k} = \hat{R}_{v,k} R_{cv} \quad (6)$$

$$\hat{s}_{c,k} = \hat{s}_{v,k} + \hat{R}_{v,k} s_{cv} \quad (7)$$

世界坐标系下到相机的投影矩阵的预测为

$$\hat{P}_c = K[\hat{R}_{c,k}^T | -\hat{R}_{c,k}^T \hat{s}_{c,k}] \quad (8)$$

车体和相机的位置预测完成后, 道路结构通过相机的投影矩阵变换到图像坐标系下, 如图3所示. 对于直道环境, 只有一条道路线被投影到图像中, 生成一个局部道路坐标系; 对于路口环境, 多条道路线被投影到图像中, 将图像分割成不同区域, 每个区域包含一个局部道路坐标系; 对于弯道环境, 用多段线段拟合弯道, 再将其投影到图像中, 与路口环境类似. 对于每一个区域, 道路坐标系对应三个朝

向 d_j ($j = 1, 2, 3$), 分别为道路行进方向、水平垂直道路方向和竖直方向. 虽然对于二维地图, 道路朝向只有一个自由度, 但为了系统的统一, 以及今后有三维道路地图后算法的可扩展性, 使用旋转矩阵 $R_r = [d_1 \ d_2 \ d_3]$ 来描述道路朝向.

3.3 道路结构特征提取

道路结构特征 RSF 定义为 $\{L_1, L_2, L_3, P\}$, 即三垂线和特征点的集合. 道路结构特征的提取中最重要的部分为从图像基础特征 (线段和点) 中建立道路结构特征. 对于特征点, 直接使用通用特征点检测跟踪结果; 对于三垂线特征, 利用预测的道路结构, 将通用线段特征依照朝向进行分类, 关键点在于定义图像线段和道路坐标系三方向的距离函数.

图像上的任意线段 l_i 转换到相机坐标系下, 构成一个过相机中心的平面 $\pi_i^l = [l_i^T K \ 0]^T$, 其法线方向为

$$n_i = \frac{K^T l_i}{\|K^T l_i\|} \quad (9)$$

定义线段 l_i 到道路方向 $R_r e_j$ 的距离函数为线段对应平面法线 n_i 和道路方向 $R_r e_j$ 夹角余弦值的平方, 即

$$d^2(l_i, R_r e_j) = (n_i^T R_c^T R_r e_j)^2 = \left(\frac{l_i^T K R_c^T R_r e_j}{\|K^T l_i\|} \right)^2 \quad (10)$$

对于任意线段 l_i , 计算它和局部坐标系下三方向的距离, 将其归类为距离最近的方向, 并将其朝向记录为 d_i . 当最小距离大于给定阈值 ϵ 时, 该线段则被当作噪声滤除.

3.4 车体位置姿态估计

给定一个道路结构特征的采样 C_{RSF} , 如式 (1) 所示. 其中 $l_1, l_2 \in L_u$ 为图像上的两条线段, 对应两条在道路坐标系下的平行线; $l_3 \in L_v$ 为第三条线段, 在道路坐标系下垂直于前两条线段. $p_1, p_2 \in P$ 为两个特征点. 令 $\hat{d}_1, \hat{d}_2, \hat{d}_3$ 为三垂线特征 l_1, l_2, l_3 的朝向估计结果, p_1 和 p_2 为当前帧检测到的特征点, p'_1 和 p'_2 为前一帧的跟踪结果.

3.4.1 旋转矩阵求解

根据线特征和对应朝向的距离定义, 有

$$l_1^T K R \hat{d}_1 = 0 \quad (11)$$

$$l_2^T K R \hat{d}_2 = 0 \quad (12)$$

$$l_3^T K R \hat{d}_3 = 0 \quad (13)$$

从特征采样 C_{RSF} 中已知 l_1 和 l_2 在道路坐标系下的朝向一致, 即 $\hat{d}_1 = \hat{d}_2$, 则 l_1 和 l_2 在图像上的交点对应于该朝向的消失点, 即

$$v_1 = l_1 \times l_2 = \lambda K R \hat{d}_1 \quad (14)$$

根据罗德里格斯矩阵, 旋转矩阵可改写为

$$R = (I - S)^{-1} (I + S) \quad (15)$$

$$S = \begin{bmatrix} 0 & -c & -b \\ c & 0 & -a \\ b & a & 0 \end{bmatrix} \quad (16)$$

将式 (15) 代入式 (14) 中, 可得

$$(I - S)K^{-1}(l_1 \times l_2) = \lambda(I + S)\hat{d}_1 \quad (17)$$

结合式 (16) 和式 (17), 共有 4 个参数 a, b, c, λ 需要求解. 式 (17) 和式 (13) 共给出 4 个约束方程, 通过联立方程组可求解这 4 个参数. 在求解过程中, 最多产生两组可行解, 这两组可行解对应 l_3 的方向相反. 通过和预测的旋转矩阵 \hat{R}_c 做比较, 可滤除错误解, 留下唯一可行解.

3.4.2 平移向量求解

在世界坐标系下, 令当前帧相机的投影矩阵为 $P = K[R|t]$, 前一帧相机的投影矩阵为 $P' = K[R'|t']$, 则相机的相对旋转和平移为

$$R_{\text{rel}} = R R'^T \quad (18)$$

$$t_{\text{rel}} = t - R_{\text{rel}}^T t' \quad (19)$$

根据对极几何约束可得

$$p^T F p' = 0 \quad (20)$$

其中,

$$F = K^{-T} [t_{\text{rel}}]_{\times} R_{\text{rel}} K^{-1} \quad (21)$$

由于旋转矩阵已知, 式 (20) 中待求解的只有平移向量 t_{rel} , 共三个自由度. 使用两个特征点, 即可通过式 (20) 求解出归一化的平移向量, 尺度因子 $\|t_{\text{rel}}\|$ 则需要通过其他方式求解. 在本文实验中, 尺度因子通过车轮编码器所获取的车速信息求解. 相对位移求解完成后, 通过式 (19) 求解当前相机的绝对位置, 再利用标定参数, 得到车体的绝对位置.

3.4.3 后验概率估计

每一个采样 C_{RSF} 都可求解出相应的旋转矩阵和平移向量. 由于采样的随机性, 需要多次采样来得到最优解. 而最优解的估计可以通过最大化后验概率 $p(Z_{\text{RSF}} | x_k^i) p(x_k^i | x_{k-1})$ 得到. 其中, $p(x_k^i | x_{k-1})$ 为运动模型, 通过车辆的基础运动模型加上高斯噪声求解; $p(Z_{\text{RSF}} | x_k^i)$ 为观测模型, 需要定义观测值与真值的误差函数来求解.

本文误差函数定义为线特征与点特征的误差加权和, 即

$$E = E_l + \lambda E_p \quad (22)$$

对于三垂线特征, 误差定义为线段长度的平方乘以它到对应道路朝向的距离, 即

$$E_l = \sum_i \text{len}^2(l_i) \times d^2(l_i, \hat{d}_i) \quad (23)$$

其中, $\text{len}(l_i)$ 是线段 l_i 的长度, $d^2(l_i, \hat{d}_i)$ 是线段和对应道路朝向的距离, 即

$$d^2(l_i, \hat{d}_i) = \left(\frac{l_i^T K R_c^T \hat{d}_i}{\|K^T l_i\|} \right)^2 \quad (24)$$

对于点特征, 误差定义为点到对应极线距离的平方, 即

$$E_p = \sum_i [d^2(p, Fp') + d^2(p', F^T p)] \quad (25)$$

其中, 距离 $d(p, Fp')$ 是点 p 到对极线 Fp' 的距离.

3.5 鲁棒性优化

道路结构特征的采样 C_{RSF} 需要包含三条线段和两个特征点, 其中三条线段 l_1, l_2, l_3 需要满足 $d_1 = d_2, d_1 \times d_3 = 0$. 因此, 算法需要检测至少两个主方向上的线段. 对于空旷环境, 即移动车辆、路灯、建筑物较少的情况, 水平垂直道路方向和垂直方向上的线特征均较少, 只有道路行进方向的线特征可被稳定检测. 在这种情况下, 假设路面平整, 车辆颠簸不明显, 车辆运动只包含三个自由度 (x, y, yaw), 对应旋转矩阵 R 退化为一个自由度. 利用公式 $l_1^T K R d_1 = 0$, 通过一条线段求解旋转矩阵. 同理, 当仅能检测到其他两个主方向上的线段时, 也将车辆运动退化为二维运动求解. 此外, 当无法检测到有效数目的主方向线段时, 则根据通用的特征点定位算法求解. 算法根据当前帧图像所能检测到的三垂线数目动态切换.

4 实验与分析

4.1 概述

本文利用如图 5 所示的智能车平台, 在北京市区的路口、拥堵路段、密集街道等典型复杂场景进行数据采集和算法验证¹. 实验中利用 PointGray 公司的 Flea2 相机采集图像序列, 图像分辨率为 640 像素 \times 480 像素, 关键帧采样帧率为 3 Hz; 利用车轮编码器采集车速信息, 用于平移尺度的求解. 为了对算法的精度进行比较验证, 实验中使用一台拥有光纤陀螺的 GPS 惯性导航组合定位系统对车辆位置方向进行高精度测量.

实验中使用线段检测工具包 (Line segment detector, LSD)^[22] 进行图像线段检测, 参考文献 [23] 中的算法进行图像特征点检测跟踪. 本文算法的实

验结果与目前被广泛使用的单目视觉里程计算法 libviso2^[13]、最新发表的 Orb-SLAM 算法^[12] 以及高精度 GPS/惯导进行比较. 以下分别对路口、拥堵路段和密集街道等三种典型场景的实验结果进行展示, 最后给出分析讨论. 在路口和拥堵路段场景, Orb-SLAM 在初始化时特征点集中于前方移动车辆上, 导致定位失败. 在密集街道场景中, 移动车辆少, Orb-SLAM 结果较为稳定. 因此本文仅在密集街道场景中加入与 Orb-SLAM 定位结果的比较.

4.2 案例分析

4.2.1 路口实验

路口实验场景起始点坐标为 (39.9754760, 116.3911050), 行驶距离约为 600 米, 时间约为 2 分钟, 采集了共 350 帧关键帧, 车辆行驶过程中包括两次右换道、一次右转和一次靠右并入辅路等驾驶行为, 如图 6 所示.

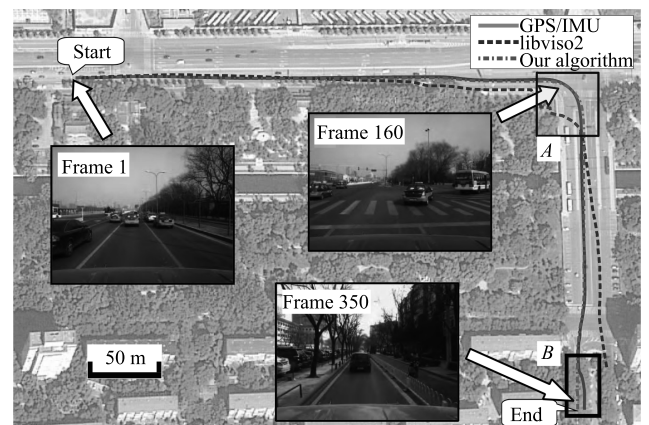


图 6 路口实验场景与定位结果

Fig. 6 The scenario and localization results of intersection environment

图 6 中点虚线、虚线和实线分别为本文算法、libviso2 和 GPS/IMU 的定位结果. 如图 6 中 B 区域所示, 在本次实验终点处, 虽然实验车沿直线行驶进入辅路, 然而由于周边树木建筑物密集, 卫星信号受到干扰, GPS 定位结果有明显偏移. 在 B 区域以外的其他区域中, 卫星信号条件良好, GPS 定位有较高的精度, 因此本文利用 B 区域以外的 GPS 数据作为参照, 定量化验证本算法的定位精度, 并与 libviso2 的结果进行比较. 实验中, libviso2 的结果在车辆右转过程中产生明显偏差, 如图 6 中 A 区域所示, 而本文算法表现良好, 轨迹和 GPS 接近. 图 7 展示了算法对车辆朝向的估计结果. 本文算法与 GPS/IMU 的结果非常接近, 而 libviso2 的结果却存在明显偏差. 虽然偏差只有 5 度, 但由此引起的位置偏差非常大, 在进入 B 区域前与 GPS 相比有

¹实验相关视频已上传至优酷: http://v.youku.com/v_show/id_XMTU4MDIzOTM5Mg==.html?from=y1.7-1.2

约 15 米的误差. 说明在定位中, 朝向的估计精度有着重要的影响, 而实验结果表明本文算法在朝向估计上精度明显高于 libviso2. 需要说明的是, 本文算法在朝向结算上只使用了当前帧的信息, 导致结果存在局部噪声, 呈现轻微锯齿状. 这个现象可以通过今后引入前后帧跟踪予以缓解.

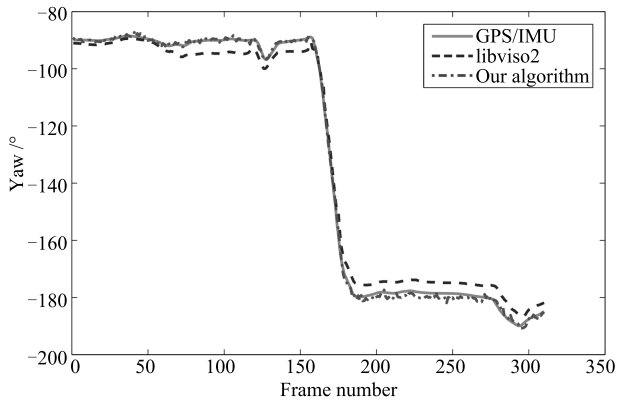


图 7 路口场景朝向计算结果

Fig. 7 Yaw results in intersection environment

为了使结果更加直观, 将图像中的沿道路线方向的线段结合定位结果投影到地平面上, 得到如图 8 所示结果. 通过投影结果, 可以较清晰地看到斑马线、道路标线、道路边界等信息. 结合车体运动轨迹, 可以检测出一些高层的运动信息, 例如车辆换道等.

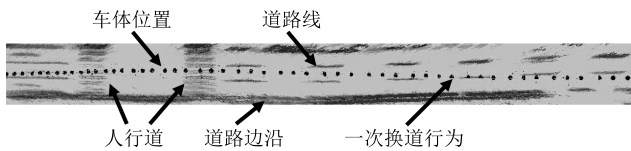


图 8 线特征鸟瞰投影结果

Fig. 8 Projected line segments on bird eye view

4.2.2 拥堵路段

拥堵路段实验场景在四环主路, 车流量大, 行驶速度缓慢, 道路笔直, 多处有过街天桥. 车辆由北向南行驶, 起始点坐标为 (39.9544800, 116.3031370), 行驶时间约为 2 分 30 秒, 全长约 1800 米, 共采集 450 帧关键帧, 期间做了若干次换道. 实验中未使用地图信息, 而利用前一帧的处理结果, 推测当前帧道路结构. 实验结果如图 9 所示, 其中点虚线、虚线和实线轨迹分别表示本文算法、libviso2 和 GPS/IMU 的结果. libviso2 的结果在第 22 帧和第 229 帧附近有两处较大偏差 (如图 9 中 A 区域和 B 区域所示), 原因是一辆环境车占据了图像中很大区域, 且进行换道, 即包含相对于车道的横向运动. 图像中该车上的特征点为实验车的运动估计带来干扰, 特别是在朝向求解上 (如图 10 所示). 另外, 实验路线中多处经过过街天桥. 这些天桥对卫星信号造成遮挡或干

扰, 使得经过天桥时, GPS 定位结果出现很大偏差. 如图 9 中 C 区域所示, GPS 轨迹呈现约 10 米的偏差. 相比之下, 本文算法的结果在朝向及位置估计上均稳定可靠, 结果优于 GPS 和 libviso2.

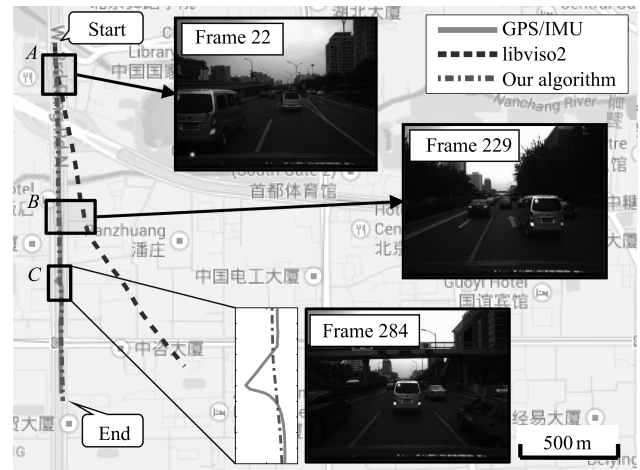


图 9 拥堵路段中的定位结果

Fig. 9 Localization results in high traffic environment

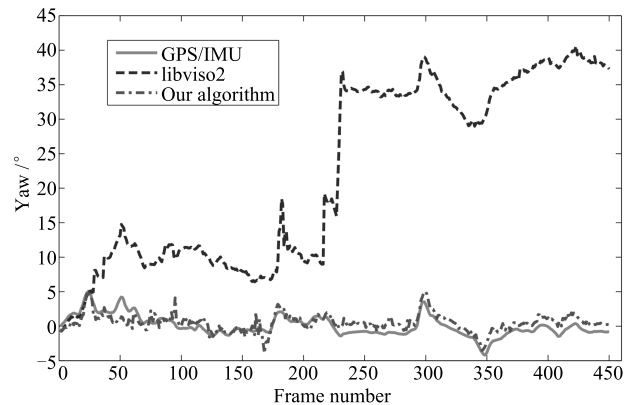


图 10 拥堵路段中的朝向计算结果

Fig. 10 Yaw results in high traffic environment

4.2.3 密集街道

密集街道场景在北京上地信息路附近区域. 车辆行驶了约 6 分 30 秒, 全长约 2000 米, 采集了 3800 帧视频数据, 期间做了若干次换道和左右转弯. 定位结果如图 11 所示, 其中点虚线、点线和虚线, 实线轨迹分别表示本文算法、Orb-SLAM、libviso2 和 GPS/IMU 的结果. 由于 Orb-SLAM 算法存在平移尺度漂移, 为了更好地比较结果, 利用车速信息将平移尺度修正. 图 11 所示结果中, libviso2 的结果在几次转弯后出现明显偏差, 而本文算法和 Orb-SLAM 误差相对较小. 以 GPS 值作为参考值计算误差, 本文算法精度高于 Orb-SLAM (数值对比结果见表 1). 在本场景中, 部分区域没有车道线信息, 且由于路边停放的车辆, 道路边沿被遮挡, 道路结

构信息不明显 (如图 11 中第 500 帧和第 2088 帧图像). 尽管如此, 但车辆大多沿道路方向停放, 其所提供的线特征与道路行驶方向以及垂直方向吻合. 在本文算法的三垂线检测过程中, 环境车辆上提供了大量符合场景约束的线段特征, 使得朝向估计稳定, 定位准确.

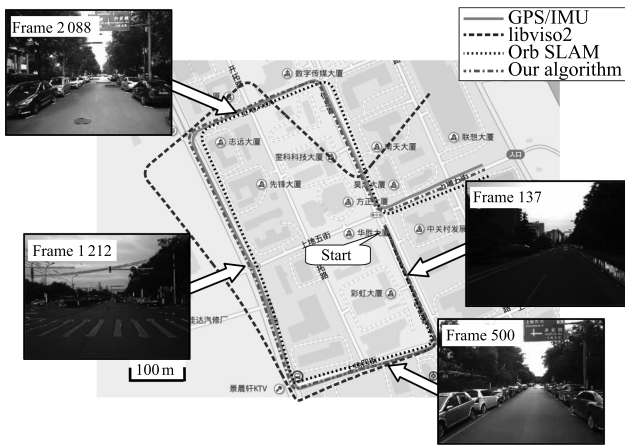


图 11 密集街道中的定位结果

Fig. 11 Localization results in downtown streets

4.3 分析讨论

在路口、拥堵路段、密集街道三组实验中, 对每组数据进行分段, 每 100 米为一段统计平移和朝向误差 (百分比), 得到多段测试的结果; 对每组实验求平均误差和 95% 置信度区间, 得表 1 (参考 Kitti 数据库中的关于视觉里程计的误差定义² 进行评价). 从表 1 中可知, 本文方法的平移误差均值在三组实验中分别为 1.07%、0.89% 和 0.69%, 算法精度较高. 误差独立统计中 95% 的误差值在 3% 以下 (分别为 2.94%、1.82% 和 1.23%), 说明算法的鲁棒性较高. 基于特征点的 libviso2 算法整体误差较大, 尤其在拥堵路段, 由于移动障碍物干扰较大, 导致 libviso2 的平均误差达到 18.49%. 在前两组实验中由于移动车辆干扰过大, Orb-SLAM 在初始化时特征点大多在前方车辆上, 导致定位完全失败. 在第三

组密集街道实验中, 移动车辆数目较少, Orb-SLAM 算法平移误差为 3.18%, 定位精度高于 libviso2, 但低于本文算法.

从路口、拥堵路段、密集街道三组实验中可知, 在视觉定位算法中, 朝向 (旋转) 求解的准确性极大地影响了定位精度. 本文提出的基于道路结构特征的定位算法, 通过可以利用结构化道路场景所普遍拥有的三垂线信息, 提高了对车体朝向估计的准确性. 图 12 简要列举了在不同场景下的三垂线检测结果. 其中图 12(a) 为开阔场景, 道路上来往车辆很少; 图 12(b) 为拥堵场景, 周围车辆数目较多; 图 12(c) 为接近天桥场景, 且存在地面标识. 这三个结果说明在任何拥堵状态下, 三垂线检测都很稳定, 尤其在可视范围内有天桥和建筑物时, 三垂线的数目明显增加. 图 12(d) 和图 12(e) 为车道线不明显和相机过曝光下的检测结果, 三垂线的检测没有受到过多影响. 图 12(f)~12(h) 为主体车进行换道, 转弯等驾驶行为时的检测结果. 在车辆转弯的过程中, 斑马线特征明显, 因此转弯时朝向求解精度很高, 与前面的定位结果一致. 图 12(i) 和图 12(j) 为直道实验中, 基于点特征的视觉定位算法受到的两次环境车干扰. 环境车分别从左侧和右侧进入相机视野且进行换道行为, 导致点特征不稳定, 影响了基于点特征的定位结果. 而三垂线特征对于环境车的干扰具有较好的鲁棒性.

实验结果表明, 在结构化道路场景中, 本文算法定位准确, 特别是在朝向估计上, 明显优于传统基于特征点的方法. 在卫星信号易受干扰的城市密集区域, 本文算法可以有效弥补 GPS 定位的不足, 满足自动驾驶所需的车道级定位精度. 然而本文算法在实际应用中仍然面临以下问题: 1) 算法为开环算法, 存在累计误差, 特别是缺乏地图信息的区域. 如何构建与 GPS 和地图的融合定位算法是实际应用的关键. 2) 算法无法解决弯道场景下的定位问题. 虽然理论上可以将弯道分解为若干个短直道, 然而这样的近似为描述真实的道路结构带来一定的误差, 导致算法在实际应用中缺乏足够的精度和鲁棒性.

表 1 三组实验的平移和朝向误差
Table 1 Translation and rotation errors

	路口实验		拥堵路段		密集街道	
	平移 (%)	朝向 (°/m)	平移 (%)	朝向 (°/m)	平移 (%)	朝向 (°/m)
	均差, 95 %	均差, 95 %	均差, 95 %	均差, 95 %	均差, 95 %	均差, 95 %
libviso2	3.94, 7.85	0.0144, 0.0323	18.60, 34.09	0.0249, 0.0428	6.63, 11.70	0.0137, 0.0227
Orb-SLAM	—	—	—	—	3.18, 5.33	0.0101, 0.0218
Our	1.07, 2.94	0.0024, 0.0049	0.89, 1.82	0.0016, 0.0040	0.69, 1.23	0.0031, 0.0075

²http://www.cvlibs.net/datasets/kitti/eval_odometry.php

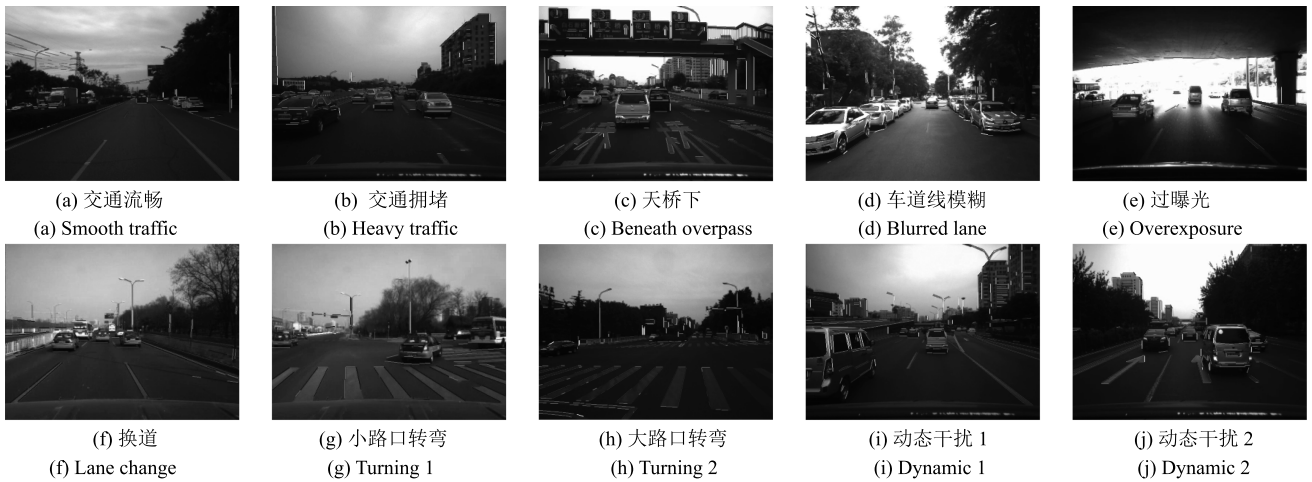


图 12 不同场景下三垂线特征检测结果

Fig. 12 Perpendicular line segment detection results in different situations

5 总结与展望

本文结合道路场景中的线特征与点特征, 定义道路结构特征, 提出基于该特征的单目视觉定位算法, 有效提高了高动态结构化道路环境下的定位精度. 在结构化道路场景中, 线特征方向性明显, 呈现三个主要方向(道路行进方向、水平垂直道路方向和竖直方向)且相互垂直. 本文算法利用路面、周围车辆和道路设施上的线段, 提取三垂线特征, 估计车辆与局部道路坐标系之间的朝向关系, 并结合图像点特征进行定位. 实验结果表明, 在结构化道路场景中, 本文算法优于传统基于特征点的视觉定位算法, 在高动态交通环境中对来自车辆的干扰有很好的鲁棒性. 在卫星信号易受干扰的区域, 可以有效地弥补 GPS 等定位系统的不足, 为满足自动驾驶所需的车道级定位要求提供重要的技术手段.

本文提出了一个利用道路结构化特征点视觉定位算法. 然而面向真实复杂场景的实际应用, 仍面临以下研究课题: 1) 构建闭环定位的应用系统, 从而减少累积误差; 2) 拓展算法, 解决弯道场景的定位问题; 3) 建立多传感器或多源信息融合的解决方案, 提高算法的鲁棒性. 针对累积误差, 算法可与路口检测或交通场景理解算法(如文献[24–25])相结合, 利用道路结构的拓扑性, 消除累积误差, 构建闭环定位的应用系统. 针对弯道场景, 算法可融合地图信息进行拓展, 建立基于曲线车道结构特征的视觉定位算法. 曲线车道结构特征的提取可参考文献[26], 建立基于车道平行平面曲线的提取与映射. 这些将在今后的研究中逐步开展.

References

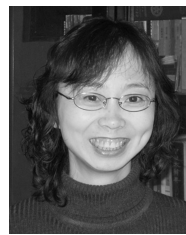
- Li Yu-Bo, Zhu Xiao-Zhou, Lu Hui-Min, Zhang Hui. Review on visual odometry technology. *Application Research of Computers*, 2012, **29**(8): 2801–2805, 2810 (李宇波, 朱效洲, 卢惠民, 张辉. 视觉里程计技术综述. 计算机应用研究, 2012, **29**(8): 2801–2805, 2810)
- Moravec H P. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover, Technical Report ADA092604, Department of Computer Science, Stanford University, USA, 1980.
- Matthies L, Shafer S. Error modeling in stereo navigation. *IEEE Journal on Robotics and Automation*, 1987, **3**(3): 239–248
- Nistér D, Naroditsky O, Bergen J. Visual odometry. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE, 2004. 652–659
- Nistér D, Naroditsky O, Bergen J. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 2006, **23**(1): 3–20
- Triggs B, McLauchlan P F, Hartley R I, Fitzgibbon A W. Bundle adjustment — a modern synthesis. *Vision Algorithms: Theory and Practice*. Berlin Heidelberg, Germany: Springer-Verlag, 2000. 298–372
- Mouragnon E, Lhuillier M, Dhome M, Dekeyser F, Sayd P. Real time localization and 3D reconstruction. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE, 2006. 363–370
- Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981, **24**(6): 381–395
- Ozden K E, Schindler K, Van Gool L. Multibody structure-from-motion in practice. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(6): 1134–1141
- Kundu A, Krishna K M, Jawahar C V. Realtime multibody visual SLAM with a smoothly moving monocular camera. In: Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 2080–2087

- 11 Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM. In: Proceedings of the 13th European Conference on Computer Vision. Berlin Heidelberg, Germany: Springer, 2014. 834–849
- 12 Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 2015, **31**(5): 1147–1163
- 13 Geiger A. Libviso2: library for visual odometry 2 [Online], available: <http://www.cvlibs.net/software/libviso/>, January 3, 2017
- 14 Ramalingam S, Bouaziz S, Sturm P, Brand M. Skyline2gps: localization in urban canyons using omni-skylines. In: Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China: IEEE, 2010. 3816–3823
- 15 Ramalingam S, Bouaziz S, Sturm P. Pose estimation using both points and lines for geo-localization. In: Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE, 2011. 4716–4723
- 16 Achar S, Jawahar C V, Krishna K M. Large scale visual localization in urban environments. In: Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE, 2011. 5642–5648
- 17 Ziegler J, Latégahn H, Schreiber M, Keller C G, Knöppel C, Hipp J, Hauéis M, Stiller C. Video based localization for bertha. In: Proceedings of the 2014 IEEE Intelligent Vehicles Symposium. Dearborn, MI, USA: IEEE, 2014. 1231–1238
- 18 Geiger A. Monocular road Mosaicing for urban environments. In: Proceedings of the 2009 IEEE Intelligent Vehicles Symposium. Xi'an, China: IEEE, 2009. 140–145
- 19 Zhang J, Song D Z. On the error analysis of vertical line pair-based monocular visual odometry in urban area. In: Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, MO, USA: IEEE, 2009. 3486–3491
- 20 Barinova O, Lempitsky V, Tretyak E, Kohli P. Geometric image parsing in man-made environments. In: Proceedings of the 11th European Conference on Computer Vision. Berlin Heidelberg, Germany: Springer, 2010. 57–70
- 21 Tretyak E, Barinova O, Kohli P, Lempitsky V. Geometric image parsing in man-made environments. *International Journal of Computer Vision*, 2012, **97**(3): 305–321
- 22 von Gioi R G, Jakubowicz J, Morel J M, Randall G. LSD: a line segment detector [Online], available: <http://www.ipol.im/pub/art/2012/gjmr-lsd/>, January 3, 2017
- 23 Geiger A, Ziegler J, Stiller C. Stereoscan: dense 3d reconstruction in real-time. In: Proceedings of the 2011 IEEE Intelligent Vehicles Symposium. Baden-Baden, Germany: IEEE, 2011. 963–968
- 24 Geiger A, Lauer M, Wojek C, Stiller C, Urtasun R. 3D traffic scene understanding from movable platforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(5): 1012–1025
- 25 Corral-Soto E R, Elder J H. Automatic single-view calibration and rectification from parallel planar curves. In: Proceedings of the 13th European Conference on Computer Vision. Berlin Heidelberg, Germany: Springer, 2014. 813–827

- 26 Zhuang Yan, Chen Dong, Wang Wei, Han Jian-Da, Wang Yue-Chao. Status and development of natural scene understanding for vision-based outdoor mobile robot. *Acta Automatica Sinica*, 2010, **36**(1): 1–11
(庄严, 陈东, 王伟, 韩建达, 王越超. 移动机器人基于视觉室外自然场景理解的研究与进展. *自动化学报*, 2010, **36**(1): 1–11)



俞毓锋 北京大学信息科学技术学院博士研究生. 2011 年获北京大学学士学位. 主要研究方向为智能车和计算机视觉.
E-mail: yuyufeng@pku.edu.cn
(**YU Yu-Feng** Ph.D. candidate at the School of Electronics Engineering and Computer Science, Peking University. He received his bachelor degree from Peking University in 2011. His research interest covers intelligent vehicle and computer vision.)



赵卉菁 北京大学机器感知与智能教育部重点实验室研究员. 1999 年获日本东京大学博士学位. 主要研究方向为智能车, 机器感知和移动机器人. 本文通信作者. E-mail: zhaohj@cis.pku.edu.cn
(**ZHAO Hui-Jing** Professor at the Key Laboratory of Machine Perception (Ministry of Education), Peking University. She received her Ph.D. degree from the University of Tokyo, Japan in 1999. Her research interest covers intelligent vehicle, machine perception, and mobile robot. Corresponding author of this paper.)



崔锦实 北京大学机器感知与智能教育部重点实验室副教授. 2004 年获清华大学博士学位. 主要研究方向为计算机视觉与智能系统.
E-mail: cjs@cis.pku.edu.cn
(**CUI Jin-Shi** Associate professor at the Key Laboratory of Machine Perception (Ministry of Education), Peking University. She received her Ph.D. degree from Tsinghua University in 2004. Her research interest covers computer vision and intelligent systems.)



查红彬 北京大学机器感知与智能教育部重点实验室教授. 1990 年获日本九州大学博士学位. 主要研究方向为计算机视觉, 智能人机交互.
E-mail: zha@cis.pku.edu.cn
(**ZHA Hong-Bin** Professor at the Key Laboratory of Machine Perception (Ministry of Education), Peking University. He received his Ph.D. degree from Kyushu University, Japan in 1990. His research interest covers computer vision and intelligent human-machine interaction.)