

DOI: 10.19650/j.cnki.cjsi.J2209132

基于级联深度神经网络的车辆位姿融合估计*

常 彬¹, 李 旭¹, 徐启敏¹, 李 娜², 胡玮明¹

(1. 东南大学仪器科学与工程学院 南京 210046; 2. 交通运输部公路科学研究院 北京 100088)

摘 要:在复杂的城市环境下,全球导航卫星系统信号失效时,现有利用单目相机或惯性导航系统估计车辆位姿的方法存在严重的累积误差。针对以上问题,提出了一种基于级联深度神经网络(CDNN)的车辆位姿融合估计算法。首先,设计了一种级联深度神经网络,减小单目相机因尺度模糊和尺度漂移造成的累积误差;其次,为了减小引入的器件噪声,利用简化惯性传感器系统(RISS)获取车辆横向、纵向加速度和横摆角速度。为了减小系统中不确定噪声的影响,利用 H_{∞} 滤波融合 CDNN 和 RISS 的输出,在准确估计车辆位姿的同时保证高频输出。在 KITTI 数据集上的实验结果表明,与基于卡尔曼滤波的方法相比,本文算法估计得到的东向位置均方根误差(RMSE)减小了 41.3%,北向位置 RMSE 减小了 70.6%,航向角 RMSE 减小了 66.6%。

关键词: 车辆位姿;级联深度神经网络;三维运动学参数;简化惯性传感器系统; H_{∞} 滤波

中图分类号: TH701 文献标识码: A 国家标准学科分类代码: 460.4020

Fusion estimation of vehicle pose based on the cascaded deep neural network

Chang Bin¹, Li Xu¹, Xu Qimin¹, Li Na², Hu Weiming¹

(1. School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China;

2. Research Institute of Highway Ministry of Transport, Beijing 100088, China)

Abstract: In a complex urban environment, when the GNSS signal fails, the existing methods for estimating the vehicle pose using a monocular camera or an inertial navigation system suffer serious cumulative errors. To address these issues, a vehicle pose fusion estimation algorithm based on the cascaded deep neural network (CDNN) is proposed. First, CDNN is designed to reduce the cumulative error caused by scale blur and scale drift in monocular cameras. Secondly, to reduce the introduced device noise, a simplified inertial sensor system (RISS) is used to obtain the vehicle lateral and longitudinal acceleration and yaw rate. To reduce the influence of uncertain noise in the system, H_{∞} filtering is used to fuse the outputs of CDNN and RISS to accurately estimate the vehicle pose while keeping high-frequency output. Compared with the method based on the Kalman filter, experimental results on the KITTI dataset show that, the root mean square error (RMSE) of the easting position estimated by the proposed algorithm is reduced by 41.3%, and the RMSE of the estimated northing position is reduced by 70.6%, and the RMSE of the estimated heading angle is reduced by 66.6%.

Keywords: vehicle pose; cascaded deep neural network; three-dimensional kinematics parameters; reduced inertial sensor system; H_{∞} filtering

0 引 言

近年来,随着智能驾驶服务的快速发展,对车辆位置和姿态的可靠估计技术成为其推广应用的关键。传统上,全球导航卫星系统(global navigation satellite system,

GNSS)和基于微机械(micro electro mechanical systems, MEMS)的惯性导航系统(inertial navigation system, INS)的组合导航技术是车辆位姿估计通用的技术手段。其中,在开阔环境中,GNSS 可以实时提供车辆的绝对位置和速度等信息^[1]。此外,MEMS-INS 是一种自主式导航系统,不易受到外部环境的干扰,可以在 GNSS 信号短时

收稿日期:2022-01-04 Received Date: 2022-01-04

* 基金项目:国家自然科学基金青年科学基金项目(4190402)、国家自然科学基金(61973079)、未来网络科研基金项目(FNSRFP-2021-YB-29)资助

间中断情况下跟踪车辆的位置姿态。然而,在复杂的城市环境中,由于 GNSS 信号受遮挡,会导致车辆定位输出精度下降甚至中断。仅依靠 MEMS-INS 可以在短时间内保持车辆位姿估计的精度,但在累积误差的影响下,特别是在 GNSS 信号长时间频繁丢失的情况下, MEMS-INS 无法满足车辆位姿可靠估计的需求^[2]。在复杂的城市环境下,当 GNSS 信号丢失时如何对车辆位姿进行准确可靠的估计,是智能驾驶服务应用亟需解决的问题。

与 MEMS-INS 类似,单目相机是一种体积小、成本低的车载感知传感器^[3]。已有方法利用单目相机捕获的连续图像对车辆位姿进行估计,并在低速运动环境下取得了较高的估计精度。在此基础上,结合单目相机和 MEMS-INS 两者优势的车辆位姿估计方法逐渐成为了研究热点。目前,单目相机或 MEMS-INS 进行车辆位姿估计方法主要分为两类,一类是基于视觉惯性里程计 (visual-inertial odometry, VIO) 的方法,另一类是基于深度学习的方法。

1) 基于 VIO 方法

VIO 通过融合图像和 INS 信息实现对刚性物体的位姿估计,视觉里程计 (visual odometry, VO) 是其实现的关键。文献[4]提出了包括特征提取、匹配和帧间位姿估计的经典 VO 框架并且为后续 VIO 相关研究奠定了关键基础。此外,基于卡尔曼滤波 (Kalman filter, KF) 技术融合单目相机/MEMS-INS 的方法在 VO 中得到了广泛应用。文献[5-6]提出了基于松耦合的 KF 方法融合双目相机和 INS 的输出,但是其没有充分使用 INS 的输出来辅助图像特征点匹配、跟踪和离群点移除等过程,在车辆高速运动过程中很快出现累积误差。文献[7]提出了基于多维随机特征算法的松耦合方案来实现视觉辅助的低精度惯性导航,但是该方案严格要求跟踪的特征点个数保持不变,大大限制了其鲁棒性和实用性。Bloesch 等^[8]提出了鲁棒视觉惯性里程计算法来融合图像和 INS 的信息,该算法基于扩展卡尔曼滤波的紧耦合框架,能够兼顾精度和计算量的要求。但是,当提取到的特征数量少于 20 个时,该算法估计车辆位姿时其精度将明显降低。在文献[9]中,作者提出了基于无迹卡尔曼滤波的 VIO 框架,并引入了对极几何约束来估计微型飞行器的姿态。该框架能够在较长时间内更准确地跟踪两个连续图像之间的特征。但是,当微型飞行器快速移动时,由于单目相机存在尺度模糊和尺度漂移问题,该算法框架的准确度会大大降低。

当车辆在光照变化剧烈和复杂动态的城市环境中快速运动时,单目相机存在尺度模糊和尺度漂移等问题。目前,融合单目相机/MEMS-INS 的 VIO 方法无法充分有效地捕捉图像的有效信息,在估计车辆位姿时会产生较

大误差,无法满足智能驾驶服务应用对于车辆位姿估计的精度要求。

2) 基于深度学习的方法

随着深度学习技术的不断发展,基于神经网络架构的端到端视觉/INS 融合方法得到了越来越多的关注。对比传统提取图像特征点的方法,基于深度学习的方法可以更加充分地利用图像中的信息并进一步从输入图像中提取充足的有效特征,在一定程度上消除复杂动态环境的不利影响,但是现有方法主要是利用端到端的方式从单目相机图像中直接估计车辆的位姿。DeepVO^[10]是基于深度学习的视觉里程计方法中的代表性工作。该方法首先利用卷积神经网络直接从原始 RGB 图像中提取有效信息,其次利用循环神经网络 (recurrent neural network, RNN) 来对提取到的信息之间的依赖关系进行自动建模,然后得到车辆位姿的相对变化量,最终由李代数合成层解算出车辆的位姿,实现了端到端的车辆相对位姿估计。在公开数据集上的实验验证了端到端的深度学习技术可以成为传统 VIO 框架的一种可行的补充。但是 DeepVO 采用了两个连续的 LSTM,且每个 LSTM 均有 1 000 个隐藏单元,参数量过大,容易出现过拟合现象,且位姿参数输出频率低。Clark 等^[11]使用双向光流提取网络 FlowNetCorr^[12]和多节点长短期记忆网络 (long short-term memory, LSTM) 设计了一种端到端的视觉/INS 组合里程计网络 VINet。VINet 利用构建的视觉特征提取器解析并融合两张单目相机图像的高维特征,实验表现性能突出。同时,文献[13]公开了 FlowNetCorr 网络的预训练参数,有利于发展基于深度学习的视觉-惯性位姿估计技术。然而,FlowNetCorr 网络的卷积神经网络 (convolution neural network, CNN) 层数较多,参数量大,训练过程中网络容易发散。针对该问题,文献[14]设计了仅由 6 层 CNN 构成的视觉特征提取器,并采用均值池化操作将高维视觉特征直接压缩为 6 维度的相邻图像帧位姿。但是该方法对于图像特征的解析能力比较弱,在位姿估计任务中神经网络的收敛速度较慢。

综上,基于深度学习的方法,大多采用由图像直接估计位姿的端到端的方式,在一定区域范围内可以取得较好的效果,但是在差异较大的新场景中泛化能力有限,存在位姿估计结果误差较大的问题。

针对以上问题,本文提出了一种基于级联深度神经网络 (cascaded deep neural network, CDNN) 的车辆位姿融合估计方法。首先,在 CDNN 架构中,以 5 层 CNN 结构作为特征提取模块自动提取图像中的有效几何特征,以 1 层 LSTM 作为时序学习模块来对图像特征的运动学关系进行隐式建模。其次,本文采用简化惯性传感器系统 (reduced inertial sensor system, RISS) 代替传统的全维 INS,建立车辆位姿参数动态偏差模型,并利用 H_{∞} 滤波

方法对 RISS 和 CDNN 的输出进行融合,其中 RISS 中包含纵向加速度计、横向加速度计与垂向陀螺仪。本文的创新点主要为:

1) 本文提出一种用于估计车辆运动状态参数的 CDNN 网络架构,自动提取连续图像中充足的有效特征,并学习有效特征之间的关系,准确估计位姿估计所需的间接量,包括车辆的横向速度、纵向速度和横摆角速度参数,弥补传统 VIO 方法中因无法提取充足图像特征而产生误差的缺陷,适用于复杂动态场景。

2) 本文提出一种基于单目相机和 MEMS 惯性传感器的车辆位姿融合估计方法,采用 RISS 代替全维 INS,减小了系统中引入的 MEMS 器件噪声,采用 H_∞ 滤波方法融合单目相机图像通过 CDNN 估计的速度和 RISS 的输出,减小了 RISS 不确定噪声的影响,同时保证了位姿输出频率,提升了车辆位姿估计的准确性与鲁棒性。

1 级联深度神经网络的车辆位姿融合估计总体框架

在城市环境下,当 GNSS 信号失效时,为了有效利用单目相机和 MEMS-INS 准确可靠地估计车辆位姿,本文提出了级联深度神经网络的车辆位姿融合估计算法,算法整体框架如图 1 所示。该框架包括用于学习单目相机图像特征及其关系的 CDNN 部分和用于融合 CDNN 与 RISS 输出信息的 H_∞ 滤波部分。

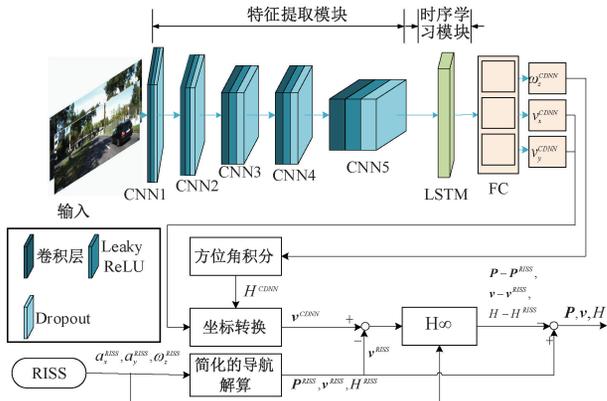


图 1 基于级联深度神经网络的车辆位姿融合估计算法框架
Fig. 1 Algorithm framework of vehicle poses fusion estimation based on CDNN

首先,本文设计了 CDNN 架构,主要包括基于 CNN 的特征提取模块和基于 LSTM 的时序学习模块。特征提取模块用于自动从图像中充分提取有效的几何特征,时序学习模块能够学习图像几何特征的运动学关系并对其运动学关系进行隐式建模。该网络能够以端到端的方式从原始的单目相机视频流中直接对估计车辆位姿所需的

三维运动参数进行估计,即车辆的横向速度、纵向速度以及横摆角速度。接着,本文引入 RISS 并将 RISS 和 CDNN 的输出解算至导航坐标系。最后,在建立车辆位姿参数动态偏差模型基础上,采用 H_∞ 滤波的方法融合 RISS 和 CDNN 的信息,获得准确可靠的车辆位姿。

2 基于级联深度神经网络的车辆三维运动参数估计

2.1 基于 CNN 的特征提取模块

考虑到 CNN 能够自动学习适用于 VO 问题的有效图像特征,本文采用 CNN 设计特征提取模块。本文提出的神经网络架构需要泛化并部署在未知环境中,因此,网络中提取的应是几何特征表示^[15]。为了估计图像中与运动相关的几何信息,常用做法是把堆叠的两张连续图像打包为一个张量作为 CNN 的输入^[13]。

特征提取模块包含 5 个 CNN 结构,以单目相机捕获的视频流中连续两帧图片堆叠成的张量作为输入。每个 CNN 结构包含一个卷积层,各卷积层的参数配置如表 1 所示。随着网络的加深,卷积层的卷积核大小由 7×7 逐渐减小为 3×3 ,以提取到更精细、更丰富的有效信息。Leaky ReLU 激活函数是 ReLU 函数的变体,保持导数始终为非零,可以避免因出现神经元坏死现象而导致的梯度消失问题^[16]。因此,每一个卷积层都选择 Leaky ReLU 作为激活函数。为了使卷积层能够适应卷积核的卷积操作,并且能够在卷积操作结束以后保持张量的空间维度,在每一个 CNN 中引入零填充操作。同时,为了避免网络出现过拟合问题,提升模型的泛化能力,在每一个 CNN 结构中引入 Dropout 层^[17]。

表 1 CNN 参数配置

Table 1 Parameters of CNN

网络层	输入尺寸	卷积核尺寸	零填充大小	步长	Dropout 比率	输出尺寸
CNN1	1 280×384×6	7×7	3	4	0.2	320×96×128
CNN2	320×96×128	5×5	2	2	0.2	160×48×256
CNN3	160×48×256	5×5	2	2	0.2	80×24×512
CNN4	80×24×512	3×3	1	2	0.2	40×12×512
CNN5	40×12×512	3×3	1	2	0.2	20×6×1024

特征提取模块在经过训练以后从高维的原始图片中学习到低维的有效几何特征表示。这些特征不仅能够将原始高维图像压缩成紧凑的描述形式,有效减小单目相机因尺度模糊和尺度漂移产生的不利影响,而且能够促进有效特征的时序学习过程。

2.2 基于 LSTM 的时序学习模块

RNN 不适合直接从高维的原始图像中学习时序关系,并且标准的 RNN 结构在学习长序列时会出现梯度消失的问题,而 LSTM 通过引入记忆单元能够学习序列的长期依赖关系。LSTM 通过丢弃或保留先前的隐藏状态以更新当前状态,能够对特征序列运动学关系自动隐式建模。每个 LSTM 单元都与特定的时间步长相关联,这使得 LSTM 能够由任意长度的输入特征序列对车辆的运动参数进行估计。因此,本文选用 LSTM 设计时序学习模块,以 CNN 提取的图像低维几何表示特征序列作为时序学习模块的输入。给定 k 时刻 CNN 提取的特征 \mathbf{D}_k 和 $k-1$ 时刻 LSTM 的隐藏状态 \mathbf{h}_{k-1} 与单元状态 \mathbf{c}_{k-1} , k 时刻 LSTM 的更新过程为:

$$\begin{aligned} \mathbf{i}_k &= \sigma(\mathbf{W}_{Di}\mathbf{D}_k + \mathbf{W}_{hi}\mathbf{h}_{k-1} + \mathbf{b}_i) \\ \mathbf{f}_k &= \sigma(\mathbf{W}_{Df}\mathbf{D}_k + \mathbf{W}_{hf}\mathbf{h}_{k-1} + \mathbf{b}_f) \\ \mathbf{g}_k &= \tanh(\mathbf{W}_{Dg}\mathbf{D}_k + \mathbf{W}_{hg}\mathbf{h}_{k-1} + \mathbf{b}_g) \\ \mathbf{c}_k &= \mathbf{f}_k \odot \mathbf{c}_{k-1} + \mathbf{i}_k \odot \mathbf{g}_k \\ \mathbf{o}_k &= \tanh(\mathbf{W}_{Do}\mathbf{D}_k + \mathbf{W}_{ho}\mathbf{h}_{k-1} + \mathbf{b}_o) \\ \mathbf{h}_k &= \mathbf{o}_k \odot \tanh(\mathbf{c}_k) \end{aligned} \quad (1)$$

其中, \odot 表示两个向量的元素积, σ 为 Sigmoid 非线性函数, \tanh 为双曲正切非线性函数, \mathbf{W} 为相应的权重矩阵, \mathbf{b} 为相应的偏置向量, \mathbf{i}_k 、 \mathbf{f}_k 、 \mathbf{g}_k 、 \mathbf{c}_k 和 \mathbf{o}_k 分别为 k 时刻的输入门、遗忘门、输入调制门、记忆单元和输出门。

LSTM 需要一定的深度来学习高级特征表示和复杂的车辆运动状态^[18],当 LSTM 的隐藏单元数量过大时,网络容易过拟合,综合考虑,本文以一个包含 32 个隐藏单元的 LSTM 结构作为时序学习模块。该模块输入的特征维数为 1 024,输出状态维数为 8。在时序学习模块之后引入一个全连接(full connection, FC)层,根据时序学习模块隐式建立的特征序列的运动学关系输出没有累积误差车辆的三维运动学参数,具体包括车辆的纵向速度 v_x^{CDNN} 、横向速度 v_y^{CDNN} 、横摆角速度 ω_z^{CDNN} 。

2.3 网络损失函数及优化

给定单目相机图像序列的条件下,CDNN 计算的车辆三维运动学状态参数的条件概率为:

$$p(\mathbf{Y}_k | \mathbf{X}_k) = p(\mathbf{y}_1, \dots, \mathbf{y}_k | \mathbf{x}_1, \dots, \mathbf{x}_k) \quad (2)$$

其中, $\mathbf{Y}_k = (\mathbf{y}_1, \dots, \mathbf{y}_k)$ 为给定时刻 k 之前的所有车辆状态参数, $\mathbf{X}_k = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ 为给定时刻 k 之前的张量序列,其中 $\mathbf{y}_k = [v_x^{CDNN}(k), v_y^{CDNN}(k), \omega_z^{CDNN}(k)]^T$ 为 k 时刻车辆运动参数的真实值,由高精度 MEMS-INS 系统采集得到; \mathbf{x}_k 为连续两帧图像打包形成的一个张量。

通过使式 (2) 中的概率最大化,可以得到网络的最优权重与偏置超参数 θ , 即:

$$\theta^* = \arg \max_{\theta} p(\mathbf{Y}_k | \mathbf{X}_k; \theta) \quad (3)$$

式中: θ^* 为式 (2) 中的概率最大化形式。由网络得到的估计值为 $\hat{\mathbf{y}}_k = [\hat{v}_x^{CDNN}(k), \hat{v}_y^{CDNN}(k), \hat{\omega}_z^{CDNN}(k)]^T$ 。使用权重平衡后的所有时刻车辆状态平均绝对误差来更新超参数 θ , 进而使输出的估计值达到最优,即最接近真实值,该过程为:

$$\theta^* = \arg \min_{\theta} \frac{1}{k} \sum_{i=1}^k \left(\begin{aligned} &\rho_1 |v_x(i) - \hat{v}_x(i)| + \\ &\rho_2 |v_y(i) - \hat{v}_y(i)| + \\ &\rho_3 |\omega_z(i) - \hat{\omega}_z(i)| \end{aligned} \right) \quad (4)$$

其中, ρ_1 、 ρ_2 、 ρ_3 分别为比例因子,用以平衡车辆的速度和角速度状态量所占的权重,在网络训练过程中根据训练效果进行调整。

3 基于 H_{∞} 滤波的车辆位姿估计

由训练完成 CDNN 网络通过原始视觉图像估计得到的车辆三维运动参数不受累积误差的影响,但是其输出频率低。而低精度 MEMS-RISS 的输出存在不确定噪声,在 GNSS 失效时仅依赖 RISS 估计车辆位姿存在累积误差。考虑到 H_{∞} 滤波方法受外界干扰信号的影响较小,能够减小 RISS 不确定噪声的影响,提高位姿估计结果的准确性,因此,在建立车辆位姿参数动态偏差模型后,本文利用 H_{∞} 滤波融合视觉和 RISS 的信息,对车辆的位姿参数进行高频准确地估计。

3.1 以导航坐标系为基准的参数解算

CDNN 和 RISS 的输出都是相对于车体坐标系的,而在实际应用中车辆的位姿通常以导航坐标系为基准,因此需要根据 CDNN 和 RISS 的输出分别解算出车辆在导航坐标系中的参数。当车辆行驶在平坦或者曲率、坡度很小的道路上时,车轮始终不会离开路面,车辆的俯仰角的侧倾角及其对应的角速度和垂向速度可以合理地假设为 0^[19]。本文只估计车辆的二维平面位置,因此把高度假设为 0。

根据上述假设,通过对车辆三维运动参数估计网络输出的车辆横摆角速度 ω_z^{CDNN} 积分得到车辆航向角 H^{CDNN} :

$$H^{CDNN}(k) = H^{CDNN}(k-1) + \omega_z^{CDNN}(k) T^{CDNN} \quad (5)$$

式中: T^{CDNN} 为 CDNN 的参数输出周期,由计算机的性能、图像尺寸、相机采集频率和 CDNN 参数量等因素所决定,并且大于与 RISS 的输出周期 T 。

导航坐标系和车体坐标系的原点重合,则将 CDNN 输出的相对于车体坐标系的纵向速度 v_x^{CDNN} 、横向速度 v_y^{CDNN} 转换到东北天导航坐标系中:

$$\mathbf{v}^{CDNN} = \begin{bmatrix} v_x^{CDNN} \\ v_y^{CDNN} \end{bmatrix} = \begin{bmatrix} \cos H^{CDNN} & -\sin H^{CDNN} \\ \sin H^{CDNN} & \cos H^{CDNN} \end{bmatrix} \begin{bmatrix} v_x^{CDNN} \\ v_y^{CDNN} \end{bmatrix} \quad (6)$$

由 RISS 输出的车辆横摆角速度 ω_z^{RISS} 解算车辆航向角 H^{RISS} :

$$H^{RISS}(k) = H^{RISS}(k-1) + \omega_z^{RISS}(k)T \quad (7)$$

将 RISS 输出的车辆纵向加速度 a_x^{RISS} 、横向加速度 a_y^{RISS} 转换到东北天坐标系中:

$$\begin{cases} a_e^{RISS} = a_x^{RISS} \cos H^{RISS} - a_y^{RISS} \sin H^{RISS} \\ a_n^{RISS} = a_x^{RISS} \sin H^{RISS} + a_y^{RISS} \cos H^{RISS} \end{cases} \quad (8)$$

由东北向加速度速度解算 k 时刻 RISS 的东向速度 $v_e^{RISS}(k)$ 和北向速度 $v_n^{RISS}(k)$:

$$\mathbf{v}^{RISS}(k) = \mathbf{v}^{RISS}(k-1) + \mathbf{a}_{en}^{RISS}(k-1)T \quad (9)$$

其中, $\mathbf{v}^{RISS} = [v_e^{RISS}, v_n^{RISS}]^T$, $\mathbf{a}_{en}^{RISS} = [a_e^{RISS}, a_n^{RISS}]^T$ 。

由经度 λ^{RISS} 、纬度 L^{RISS} 表示的 k 时刻车辆位置 \mathbf{P}^{RISS} 的解算方法为^[20]:

$$\mathbf{P}^{RISS}(k) = \mathbf{P}^{RISS}(k-1) + \begin{bmatrix} v_e(k-1) \sec L^{RISS}(k-1) \\ R_e(k-1) \\ v_n(k-1)/R_n(k-1) \end{bmatrix} T \quad (10)$$

其中, $\mathbf{P}^{RISS} = [\lambda^{RISS}, L^{RISS}]^T$, R_n 为所在地参考椭球子午线曲率半径, R_e 为所在地与参考椭球子午线垂直的法线平面的曲率半径,计算方法在文献[20]中给出。

3.2 基于 H^∞ 滤波的 RISS/视觉信息融合

与传统的 MEMS-INS 相比,RISS 既能够减小误差累积的程度,又能够减少引入到融合系统中的噪声^[21],因此本文引入 RISS 代替 MEMS-INS。基于卡尔曼滤波及其改进形式的信息融合方法将器件噪声视为高斯噪声, H^∞ 滤波将信号中可能存在的干扰信号视为能量有限的随机信号,使干扰信号到输出误差的闭环传递函数的 H^∞ 范数最小^[22]。RISS 的输入噪声为不确定噪声,因此基于 H^∞ 滤波方法更符合实际,更有利于提高融合估计结果的准确性。

根据 3.1 节的解算过程,本文的车辆位姿参数动态偏差模型在文献[20]的基础上简化为式(11),其中 δL 、 $\delta \lambda$ 为用经纬度表示的真实位置与 RISS 解算位置的误差, v_e 、 v_n 为车辆的东向、北向真实速度, δv_e 、 δv_n 为真实东向、北向速度与 RISS 解算的东向、北向速度的误差, δH 为车辆真实航向角与 RISS 解算航向角的误差, ∇_e^{RISS} 、 ∇_n^{RISS} 分别为东向、北向的等效加速度计偏置, ε_u^{RISS} 为天向等效陀螺漂移,上标 \cdot 表示变量对时间 t 的导数。

$$\begin{cases} \delta \dot{L} = \frac{1}{R_n} \delta v_n \\ \delta \dot{\lambda} = \frac{v_e}{R_e} \sec L \tan L \delta L + \frac{\sec L}{R_e} \delta v_e \\ \delta \dot{v}_e = \left(2v_n \omega_{ie} \cos L + \frac{v_e v_n \sec^2 L}{R_e} \right) \delta L + a_n^{RISS} \delta H + \left(2\omega_{ie} \sin L + \frac{v_e \tan L}{R_e} \right) \delta v_n + \frac{v_n \tan L}{R_e} \delta v_e + \nabla_e^{RISS} \\ \delta \dot{v}_n = - \left(2v_e \omega_{ie} \cos L + \frac{v_e^2 \sec^2 L}{R_e} \right) \delta L - a_e^{RISS} \delta H - \left(2\omega_{ie} \sin L + \frac{2v_e \tan L}{R_e} \right) \delta v_e + \nabla_n^{RISS} \\ \delta \dot{H} = \left(\omega_{ie} \cos L + \frac{v_e \sec^2 L}{R_e} \right) \delta L + \frac{\tan L}{R_e} \delta v_e + \varepsilon_u^{RISS} \end{cases} \quad (11)$$

车辆在坡度较小的道路上行驶时,俯仰角、侧倾角、纵向角速度、横向角速度和垂向速度可以合理地假设为 0,则:

$$\begin{cases} \nabla_e^{RISS} = \nabla_x^{RISS} \cos H - \nabla_y^{RISS} \sin H \\ \nabla_n^{RISS} = \nabla_x^{RISS} \sin H + \nabla_y^{RISS} \cos H \\ \varepsilon_u^{RISS} = 0 \end{cases} \quad (12)$$

其中, H 为估计的车辆真实的航向角。

由式(11)和(12)得到系统的状态方程为:

$$\dot{\mathbf{X}}(t) = \mathbf{F}(t)\mathbf{X}(t) + \mathbf{G}(t)\mathbf{W}_s(t) \quad (13)$$

其中,状态变量 $\mathbf{X}(t) = [\delta L, \delta \lambda, \delta v_e, \delta v_n, \delta H]^T$, $\mathbf{F}(t)$ 为状态转移矩阵, $\mathbf{G}(t)$ 为系统噪声输入矩阵, $\mathbf{W}_s = [w_{ax}, w_{ay}, w_{gz}]^T$ 为系统白噪声, w_{ax} 为 RISS 横向加速度计噪声, w_{ay} 为 RISS 纵向加速度计噪声, w_{gz} 为 RISS 垂向陀螺仪噪声。 $\mathbf{G}(t)$ 和 $\mathbf{F}(t)$ 分别如式(16)、(17)所示。

\mathbf{P} 为车辆的真实位置信息, \mathbf{v} 为车辆的真实速度信息,令 $\delta \mathbf{P} = [\delta L, \delta \lambda]^T$ 、 $\delta \mathbf{v} = [\delta v_e, \delta v_n]^T$ 则有:

$$\begin{cases} \delta \mathbf{P} = \mathbf{P} - \mathbf{P}^{RISS} \\ \delta \mathbf{v} = \mathbf{v} - \mathbf{v}^{RISS} \\ \delta H = H - H^{RISS} \end{cases} \quad (14)$$

由 CDNN 的输出根据式(6)解算得到的速度 \mathbf{v}^{CDNN} 和 RISS 的输出根据式(9)解算得到的速度 \mathbf{v}^{RISS} 的偏差定义观测量为:

$$\mathbf{Y}(t) = \mathbf{v}^{CDNN}(t) - \mathbf{v}^{RISS}(t) + \mathbf{N}_m(t) = \mathbf{H}_{ob}(t)\mathbf{X}(t) + \mathbf{N}_m(t) \quad (15)$$

其中,观测矩阵 $\mathbf{H}_{ob} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$, 观测噪声向量为 $\mathbf{N}_m(t) = [n_{ve}, n_{vn}]^T$, n_{ve} 为东向速度观测噪声, n_{vn} 为北向速度观测噪声,噪声统计特性未知。

$$\mathbf{G}(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \cos H & -\sin H & 0 \\ \sin H & \cos H & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (16)$$

$$\mathbf{F}(t) = \begin{bmatrix} 0 & 0 & 0 & 1/R_n & 0 \\ \frac{v_e}{R_e} \sec L \tan L & 0 & \frac{\sec L}{R_e} & 0 & 0 \\ 2v_n \omega_{ie} \cos L + \frac{v_e v_n \sec^2 L}{R_e} & 0 & \frac{v_n \tan L}{R_e} & 2\omega_{ie} \sin L + \frac{v_e \tan L}{R_e} & 0 \\ -2v_e \omega_{ie} \cos L - \frac{v_e^2 \sec^2 L}{R_e} & 0 & -2\omega_{ie} \sin L - \frac{2v_e \tan L}{R_e} & 0 & a_e^{RISS} \\ \omega_{ie} \cos L + \frac{v_e \sec^2 L}{R_e} & 0 & \frac{\tan L}{R_e} & 0 & 0 \end{bmatrix} \quad (17)$$

由于 H_∞ 最优估计问题的封闭解很难得到,因此求解 H_∞ 滤波问题通常采用次优 H_∞ 滤波策略^[23],即给定性能边界 γ ,使干扰信号到输出误差的闭环传递函数的 H_∞ 范数小于 γ 。根据式(13)建立的系统状态方程和式(15)建立的测量方程,得到状态线性组合的估计方程为:

$$\mathbf{Z}(k) = \mathbf{L}(k)\mathbf{X}(k) \quad (18)$$

其中, \mathbf{Z} 为系统的待估计向量, \mathbf{L} 为给定的状态线性组合矩阵,其中 $\mathbf{Z} = \mathbf{X}$,则 $\mathbf{L} = \mathbf{I}_{5 \times 5}$,是一个单位矩阵。

以次优 H_∞ 滤波策略估计车辆位姿的递推过程参考文献[21],递推结果输出周期与 RISS 的输出周期一致。结合状态变量 \mathbf{X} 的递推结果与式(14)能够得到准确的车辆位姿参数,包括车辆的经度、纬度与航向角。

4 实验验证

为了评估本文算法的性能,本文从 KITTI 数据集中选取 11 个数据序列作为原始数据集。该数据集在复杂动态城市环境由多种传感器采集形成,不同的车载传感器进行了严格的时空同步。数据集中包含不同光照强度的图像和直行、转弯以及不同速度的车辆行驶状态数据,其中车辆运动学参数由高性能组合导航系统 OXTS RT3003 采集得到,采集频率为 100 Hz;图像由彩色相机 FL2-14S3M 采集得到,采集频率为 10 Hz^[24],能够充分验证本文算法的有效性和准确性。

本文将数据集分为包含 5 718 个样本的训练集、439 个样本的验证集和 140 个样本的测试集,每个样本包含 5 张连续的单目视觉图像序列,其中测试集中包含 700 张连续的单目视觉图像序列。数据集中的车辆位姿

参数是由高性能组合导航系统采集得到的,累积误差小,作为实验验证的参考值。为了验证本文算法的性能,从与测试集图像序列对应的组合导航数据中取出横向加速度、纵向加速度、横摆角速度并加入噪声作为 RISS 的数据,其中加速度中加入 10 mg 的常值噪声,横摆角速度中加入 300 deg/s 的常值噪声^[25]。然后,对比使用不同的方法计算得到车辆的位姿估计结果。本文采用最大误差与均方根误差 (root mean square error, RMSE) 来衡量模型的性能。

4.1 CDNN 性能实验

CDNN 的具体结构由 PyTorch 实现,训练过程中的输入为由两种连续的单目视觉图像打包形成的张量,输出为与图像对应的车辆的横向速度、纵向速度和横摆角速度,真实值为数据集中图像对应的高精度 MEMS-INS 的输出。实验中计算机配置如表 2 所示,CDNN 训练的配置参数如表 3 所示。

表 2 实验计算机配置

Table 2 Computer configuration of the experiment

硬件配置	参数
GPU 型号	NVIDIA GeForce RTX2080 Ti
GPU 内存	8 GB
CPU 型号	AMD R3-3100
CPU 主频	3.6 GHz
计算机内存	16 GB
操作系统	64 bits Ubuntu18.04

表 3 CDNN 训练的配置参数

Table 3 Configuration parameters in CDNN training

参数项	参数值
迭代轮次	300
批大小	4
优化器	Adagrad
学习率	0.001

由于车辆的纵向速度、横向速度与横摆角速度存在数量级差别,为了确定式(4)中合适的比例因子以准确地更新网络中的超参数,在 CDNN 中对每一批次的状态估计值与真实值进行归一化处理。经过 300 轮迭代后,训练集与验证集的损失曲线如图 2 所示。由损失曲线可知,CDNN 在第 200 轮迭代之后训练集与验证集曲线开始收敛,且最终训练集与验证集曲线的泛化误差 (generalization gap, GG) 趋近于 0,证明 CDNN 处于良好拟合 (good fit, GF) 状态,具备泛化能力。

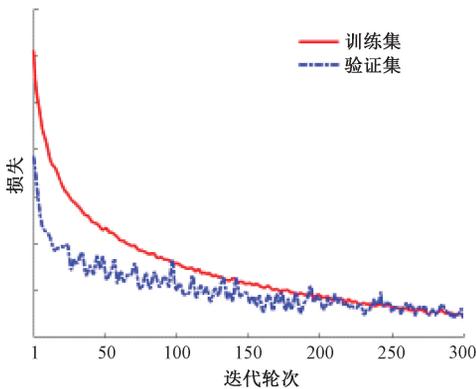


图 2 损失曲线

Fig. 2 Loss curves

对 CDNN 训练完成后,CDNN 从测试集图片序列中得到的车辆横向速度、纵向速度和横摆角速度估计结果如图 3(a)~(c) 所示,表 4 给出了估计值和真实值的误差与 RMSE。

由表 4 和图 3 可知,在测试集中图像的光照强度和车辆的行驶状态不断变化的情况下,CDNN 无需相机参数即可从原始图像序列中准确地估计车辆的三维运动参数,测试集中的样本与训练集、验证集中的样本没有重复。在新的环境中,纵向速度的 RMSE 为 0.287 9 m/s,横向速度的 RMSE 为 0.006 2 m/s,横摆角速度的 RMSE 为 0.004 7 rad/s。由实验结果可以看出,本文设计的 CDNN 在估计车辆的运动参数过程中保证了准确性。综上,本文设计的 CDNN 架构能够充分提取图像中的几何特征并根据这些特征准确估计车辆运动学参数,且具备泛化能力,能够部署在新的环境中。CDNN 对车辆三维运动参数的准确估计,能够提升位姿融合估计的精度。

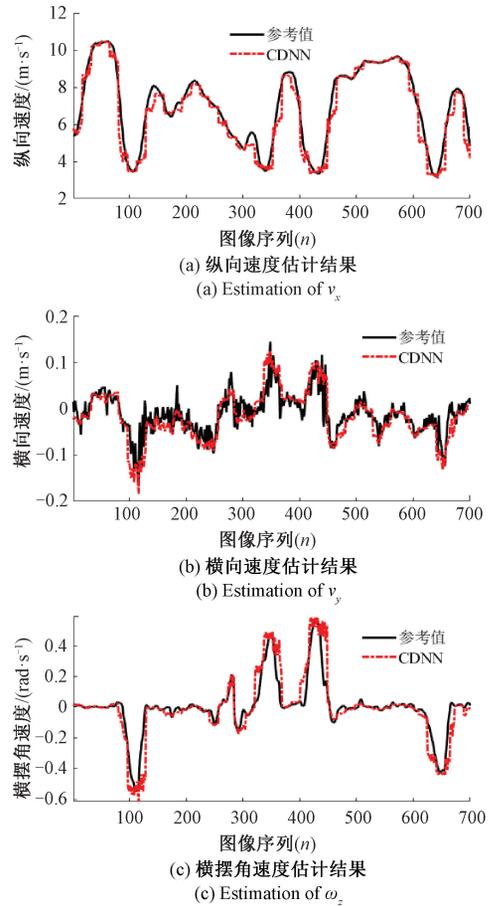


图 3 车辆三维运动参数估计结果

Fig. 3 Estimation results of vehicle three-dimensional motion parameters

表 4 车辆三维运动参数的最大误差和 RMSE

Table 4 Maximum error and RMSE of vehicle three-dimensional motion parameters

参数	最大误差	RMSE
纵向速度/($m \cdot s^{-1}$)	2.263 5	0.287 9
横向速度/($m \cdot s^{-1}$)	0.119 0	0.006 2
横摆角速度/($rad \cdot s^{-1}$)	0.447 0	0.004 7

4.2 基于 H^∞ 滤波的车辆位姿估计实验

在 CDNN 正确估计车辆运动学参数的基础上,分别使用不同的方法计算得到车辆的经度、纬度和航向角,具体包括利用 RISS 数据直接递推、基于 KF 算法融合视觉与 RISS 的数据进行递推、基于 H^∞ 滤波算法融合视觉与 RISS 数据进行递推。为了更直观地反映车辆位置及其误差,本文将车辆的经度、纬度由地心大地坐标系转换到平面导航直角坐标系中。图 4(a)、(b) 给出了车辆位置和姿态的估计结果,图 5(a)~(c) 给出了不同计算方法

得到的车辆位姿估计结果与真实值的误差,位姿估计结果的最大误差和 RMSE 如表 5 所示。

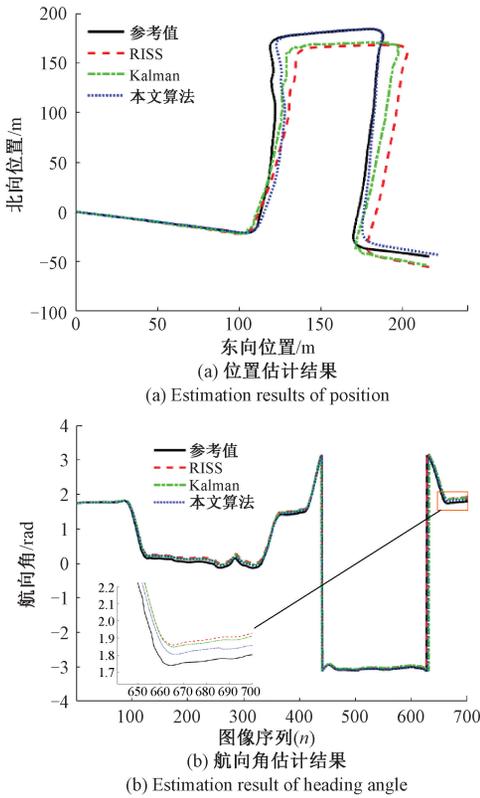


图 4 车辆位姿参数估计结果

Fig. 4 Estimation results of vehicle position and attitude estimation parameters

由图 4 与 5 可以看出,在直行、转弯等不同的行驶状态下,与使用 RISS 输出直接递推方法相比,采用基于 KF 方法和基于本文提出的 H^∞ 滤波方法对 CDNN 与 RISS 的输出进行融合均能够更准确地估计车辆位姿,且本文算法的准确度最高。通过对表 5 的定量分析可以得到,与直接递推方法相比,本文算法的东向位置的最大误差减小了 63.7%,东向位置的 RMSE 减小了 63.7%,北向位置的最大误差减小了 54.7%,北向位置的 RMSE 减小了 75.6%,航向角的最大误差减小了 15.9%,航向角的 RMSE 减小了 71.6%;与基于 KF 的融合方法相比,本文算法的东向位置的最大误差减小了 51.3%,东向位置的 RMSE 减小了 41.3%,北向位置的最大误差减小了 52.3%,北向位置的 RMSE 减小了 70.6%,航向角的最大误差减小了 13.4%,航向角的 RMSE 减小了 66.6%。

由以上实验分析可以看出,本文提出的基于 H^∞ 滤波融合 RISS 信息与视觉信息的方法误差最小,可靠性和准确性最高,综合性能表现最好,能够有效估计车辆位姿。

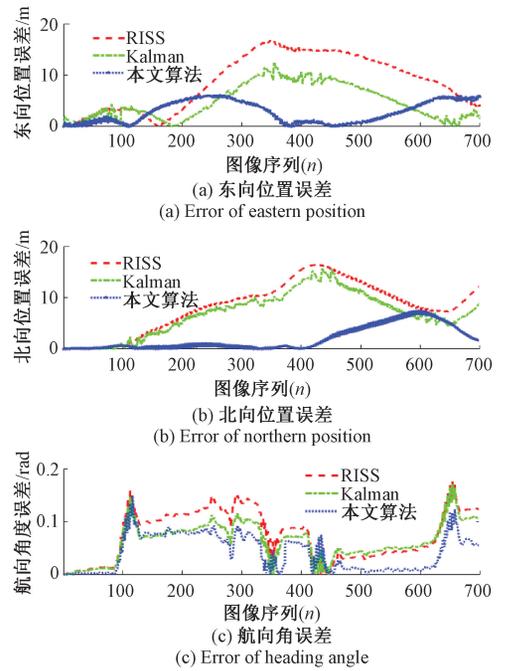


图 5 车辆位姿估计参数误差

Fig. 5 Error of vehicle position and attitude estimation parameters

表 5 车辆位姿估计参数的最大误差和 RMSE

Table 5 Maximum error and RMSE of vehicle position and attitude estimation parameters

算法	最大误差			RMSE		
	东向位置/m	北向位置/m	航向角/rad	东向位置/m	北向位置/m	航向角/rad
RISS	16.783	16.429	0.176	8.726	8.268	0.074 2
Kalman	12.503	15.615	0.171	4.821	6.855	0.063 2
本文算法	6.092	7.445	0.148	2.832	2.015	0.021 1

5 结 论

在复杂的城市环境下,当 GNSS 信号失效时,已有的融合单目相机和 INS 信息的方法无法为智能驾驶服务应用提供可靠的车辆位姿。针对此问题,本文提出了一种基于级联深度神经网络的车辆位姿融合估计方法。基于 KITTI 数据集的实验表明,本文算法能够从单目相机图像中直接准确地估计车辆的三维运动参数,与 RISS 的输出融合后能够准确地估计车辆的位姿参数,包括车辆的经度、纬度和航向角。从实验结果可以看出,本文设计的 CDNN 网络架构在不同光照环境和车辆的不同行驶状态

下,能够从单目相机中提取充足的有效特征,弥补了传统VIO方法因单目相机的尺度模糊和尺度漂移问题而无法提取充足有效特征的不足,能够在新场景中准确地对车辆的速度、角速度运动参数进行估计,具有一定的泛化能力。本文采用 H_{∞} 滤波算法融合CDNN与RISS的输出,减小了系统中引入的器件噪声,减小了单目相机和RISS不确定噪声的影响,提升了位姿估计精度,同时保证了输出频,具有较强的鲁棒性。下一步本文将继续优化神经网络,并考虑引入激光雷达等传感器以进一步提高位姿估计精度和环境适应性。

参考文献

- [1] 曹相,王庆,高成发,等. 基于BDS-3,GPS和Galileo重叠频率观测值的紧组合RTK定位方法[J]. 仪器仪表学报, 2019, 40(10):138-144.
- CAO X, WANG Q, GAO CH F, et al. Tightly combined RTK positioning method based on the overlapping-frequency observations of BDS-3, GPS and Galileo[J]. Chinese Journal of Scientific Instrument, 2019, 40(10): 138-144.
- [2] IASECHKO M, SHELUKHIN O, MARANOV A, et al. Evaluation of the use of inertial navigation systems to improve the accuracy of object navigation [J]. International Journal of Computer Science & Network Security, 2021, 21(3): 71-75.
- [3] 燕必希,朱立夫,董明利,等. 卡尔曼滤波单目相机运动目标定位研究[J]. 仪器仪表学报, 2018, 39(8): 220-229.
- DONG B X, ZHU L F, DONG M L, et al. Research on Kalman filtering for moving target location on monocular camera[J]. Chinese Journal of Scientific Instrument, 2018, 39(8):220-229.
- [4] MATTHIES L, SHAFER S A. Error modeling in stereo navigation[J]. IEEE Journal on Robotics & Automation, 2003, 3(3):239-248.
- [5] LUIS R G C, ALEJANDRO E D L, LOZANO R, et al. Combining stereo vision and inertial navigation system for a quad-rotor UAV[J]. Journal of Intelligent & Robotic Systems, 2012, 65(1-4):373-387.
- [6] KELLY J, SARIPALLI S, SUKHATME G S. Combined visual and inertial navigation for an unmanned aerial vehicle [C]. Field and Service Robotics. Springer, Berlin, Heidelberg, 2008:255-264.
- [7] VETH M, RAQUET J. Fusing low-cost image and inertial sensors for passive navigation [J]. NAVIGATION, Journal of the Institute of Navigation, 2007, 54(1):11-20.
- [8] BLOESCH M, OMARI S, HUTTER M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2015:298-304.
- [9] OMOTUYI O, KUMAR M. UAV visual-inertial dynamics (VI-D) odometry using unscented Kalman filter [J]. IFAC-Papers Online, 2021, 54(20):814-819.
- [10] WANG S, CLARK R, WEN H, et al. Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks [C]. 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2017:2043-2050.
- [11] CLARK R, WANG S, WEN H, et al. Vinet: Visual-inertial odometry as a sequence-to-sequence learning problem[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1):1-10.
- [12] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: Learning optical flow with convolutional networks [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015:2758-2766.
- [13] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2462-2470.
- [14] ZHAN H, GARG R, WEERASEKERA C S, et al.

- Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 340-349.
- [15] SONG S Y, PARK S W, KIM H G, et al. Deep learning based visual-inertial drone odometry estimation [C]. Proceedings of the Korea Information Processing Society Conference, Korea Information Processing Society, 2020: 842-845.
- [16] LIU X, WANG Y, YANG Y. Fault diagnosis based on sparse semi-supervised GAN model [C]. 2020 Chinese Control and Decision Conference (CCDC), IEEE, 2020: 5620-5624.
- [17] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
- [18] YE F, ZHANG S, WANG P, et al. A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles[C]. 2021 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2021:1073-1080.
- [19] LI X, XU Q, LI B, et al. A highly reliable and cost-efficient multi-sensor system for land vehicle positioning[J]. Sensors, 2016, 16(6):755.
- [20] 田佳玉, 胡柏青, 李开龙, 等. 基于欧拉角的捷联惯性组合导航滤波算法比较[J]. 舰船电子工程, 2021, 41(9): 41-45.
- TIAN J Y, HU B Q, LI K L, et al. Comparison of strapdown inertial integrated navigation filtering based on euler angle [J]. Ship Electronic Engineering, 2021, 41(9): 41-45.
- [21] 徐启敏. 城市环境下车辆智能融合定位技术研究[D]. 南京:东南大学, 2018.
- XU Q M. Intelligent fusion positioning technology for vehicles in urban environment [D]. Nanjing: Southeast University, 2018.
- [22] ZHANG Q, STEPHENSON S, MENG X, et al. A new robust filtering for a GPS/SINS loosely coupled integration system[J]. Survey Review, 2016, 48(348): 181-187.
- [23] 宋丽君, 秦永元, 严恭敏. H_{∞} 次优滤波在速度姿态匹配传递对准中的应用[J]. 传感技术学报, 2012, 25(1):49-52.
- SONG L J, QIN Y Y, YAN M G. Velocity and attitude matching of transfer alignment by using H_{∞} filter[J]. Chinese Journal of Sensors and Actuators, 2012, 25(1): 49-52.
- [24] CVIŠIĆ I, MARKOVIĆ I, PETROVIĆ I. Recalibrating the KITTI dataset camera setup for improved odometry accuracy [C]. 2021 European Conference on Mobile Robots (ECMR), IEEE, 2021:1-6.
- [25] KIM K H, LEE J G, CHAN G P. Adaptive two-stage extended Kalman filter for a fault-tolerant ins-gps loosely coupled system[J]. IEEE Transactions on Aerospace and Electronic Systems, 2009, 45(1):125-137.

作者简介



常彬, 2018 年于南京工业大学获得电力系统及其自动化学士学位, 现为东南大学仪器科学与工程学院研究生, 主要研究方向为智能车路系统测评和智能车辆感知与定位。
E-mail: bin_chang@qq.com

Chang Bin received his B.Sc. degree in the major of electrical engineering and automation from Nanjing Tech University in 2018. He is currently a master student in the School of Instrument Science and Engineering at Southeast University. His current research interests include assessment of intelligent vehicle and infrastructure systems and intelligent vehicle perception and positioning.



李旭(通信作者), 2006 年于东南大学获得仪器科学与技术博士学位, 现为东南大学仪器科学与工程学院教授, 主要研究方向为智能车路系统的协同感知与控制、信息融合、自动驾驶和主动安全。

E-mail: lixu.mail@163.com

Li Xu (Corresponding author) received his Ph. D. degree in the major of instrument science and technology from Southeast University in 2006. He is currently a professor in the School of Instrument Science and Engineering at Southeast University. His main research interests are the collaborative perception and control of intelligent vehicle and infrastructure systems, information fusion, automated vehicles, and active safety.



徐启敏, 分别在 2011 年、2014 年和 2018 年于东南大学获得仪器科学与技术学士、硕士、博士学位, 现为东南大学仪器科学与工程学院讲师, 主要研究方向为车辆状态估计、车辆定位和自动驾驶。

E-mail: jimmy.xqm@gmail.com

Xu Qimin received his B. Sc., M. Sc. and Ph. D. degrees in the major of instrument science and technology from Southeast University in 2011, 2014, and 2018, respectively. He is currently a lecturer in the School of Instrument Science and Engineering at Southeast University. His main research interests are vehicle state estimation, vehicle positioning, and autonomous driving.



李娜, 2009 年于北京交通大学获得交通运输规划与管理硕士学位, 现为交通运输部公路科学研究院研究人员, 主要研究方向为智能车路系统和自动驾驶。

E-mail: na.li@rioh.cn

Li Na received her M. Sc. degree in the major of transportation planning and management from Beijing Jiaotong University in 2009. She is currently a research worker at the Research Institute of Highway Ministry of Transport. Her current research interests are intelligent vehicle and infrastructure systems and autonomous driving.



胡玮明, 2017 年于东北大学获得机械工程专业硕士学位, 现为东南大学仪器科学与工程学院博士研究生, 主要研究方向为自动驾驶决策和智能车路系统测试评估。

E-mail: hwm_seu@163.com

Hu Weiming received his M. Sc. degree in the major of mechanical engineering and automation from Northeast University in 2017. He is currently a Ph. D. candidate in the major of instrument science and technology at Southeast University. His current research interests include driving decision-making and assessment of the intelligent vehicle-infrastructure system.