(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 111272174 B
(45)授权公告日 2021.11.23

(21)申请号 202010125661.2

(22)申请日 2020.02.27

(65)同一申请的已公布的文献号
申请公布号 CN 111272174 A

(43)申请公布日 2020.06.12

(73)专利权人 中国科学院计算技术研究所
地址 100190 北京市海淀区中关村科学院南路6号

(72)发明人 罗海勇 高喜乐 包林封 宁勃坤 龚依林 肖逸敏

(74)专利代理机构 北京泛华伟业知识产权代理有限公司 11280
代理人 王勇

(51)Int.Cl.
G01C 21/16(2006.01)
G01C 21/28(2006.01)
G01S 19/47(2010.01)

(56)对比文件
CN 101846734 A,2010.09.29
CN 105652306 A,2016.06.08
CN 106500695 A,2017.03.15
CN 107014376 A,2017.08.04
CN 110322017 A,2019.10.11
CN 110196443 A,2019.09.03
CN 110727968 A,2020.01.24
US 2009132164 A1,2009.05.21

审查员 沈新华

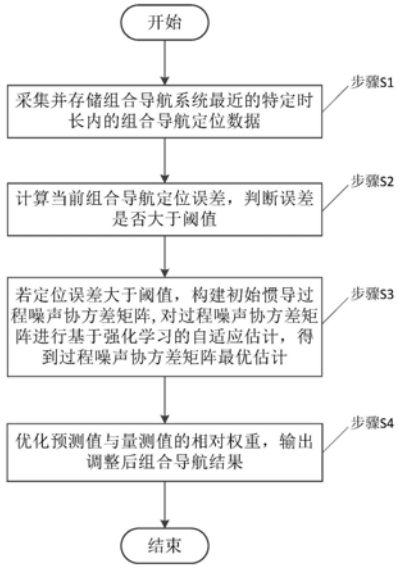
权利要求书2页 说明书11页 附图8页

(54)发明名称

一种组合导航方法和系统

(57)摘要

本发明公开了一种组合导航方法和系统，所述方法包括下述步骤：采集并存储组合导航系统最近的特定时长内的组合导航定位数据；计算当前组合导航定位误差，判断误差是否大于阈值；若定位误差大于阈值，构建初始惯导过程噪声协方差矩阵，对过程噪声协方差矩阵进行基于强化学习的自适应估计，得到过程噪声协方差矩阵最优估计；优化预测值与量测值的相对权重，输出调整后组合导航结果。本发明能够为单个具体惯导设备提供个性化的组合导航方案，避免对先验知识的依赖；所采用的模型学习能力强，调优速度快，解决了梯度回传困难的挑战；提高组合导航系统的定位准确度，在不同环境下具有强鲁棒性。



CN 111272174 B

1.一种组合导航方法,其特征在于,包括下述步骤:

步骤S1、采集并存储组合导航系统最近的特定时长内的组合导航定位数据;

步骤S2、计算当前组合导航定位误差,判断误差是否大于阈值;

步骤S3、若定位误差大于阈值,构建初始惯导过程噪声协方差矩阵,对过程噪声协方差矩阵进行基于强化学习的自适应估计,得到过程噪声协方差矩阵最优估计;

步骤S4:优化预测值与量测值的相对权重,输出调整后组合导航结果。

2.根据权利要求1所述的一种组合导航方法,其特征在于,所述步骤S1的组合导航系统为基于卡尔曼滤波构建,作为强化学习模型的环境。

3.根据权利要求1或2所述的一种组合导航方法,其特征在于,所述步骤S1的最近组合导航定位数据作为强化学习模型的训练数据,训练数据的时间长度由定位精度、训练所需时间和数据存储性能确定。

4.根据权利要求1所述的一种组合导航方法,其特征在于,所述步骤S3的过程噪声协方差矩阵由过程噪声及其变化量确定,作为强化学习模型的状态,其定义为:

$$Q = E[e^{W_k+\alpha_k}(e^{W_k+\alpha_k})^T],$$

其中,过程噪声$W_k = [w_{\delta r}, w_{\delta v}, w_{\varphi}, w_q, w_{\nabla}, w_{\delta kit}]^T$,$w_{\delta r}$为位置游走系数,$w_{\delta v}$、$w_{\varphi}$、$w_q$、$w_{\nabla}$、$w_{\delta kii}$分别速度、角度、角速率、加速度和比例误差随机游走;α_k为过程噪声W_k变化量,作为强化学习模型的动作。

5.根据权利要求1所述的一种组合导航方法,其特征在于,对所述步骤S3基于强化学习的自适应估计方法是在深度确定性策略梯度DDPG基础上,以所述步骤S1的组合导航定位数据为输入,以过程噪声协方差矩阵最优估计为输出,包括如下步骤:

步骤S31、构造动作评估网络和值评估网络;

步骤S32、初始化两个网络的权重;

步骤S33、初始化经验池大小和每轮训练数据个数;

步骤S34、对于每一轮的每组训练数据分别计算值评价网络和动作评价网络的梯度,并更新网络权值;

步骤S35、得到过程噪声协方差矩阵最优估计。

6.根据权利要求5所述的一种组合导航方法,其特征在于,所述步骤S31的动作评估网络输入维度与状态的维度相等,输出维度与动作的维度相等;所述值评估网络输入由状态和动作共同构成,其维度为状态和动作维度之和,输出维度为1。

7.根据权利要求6所述的一种组合导航方法,其特征在于,所述步骤S34包括如下步骤:

步骤S341、动作评估网络根据当前状态计算得到一个动作估计;

步骤S342、组合导航模型计算得到下一个状态;

步骤S343、计算过程噪声协方差矩阵,对定位误差取反得到奖励;

步骤S344、保存转换关系到经验池中;

步骤S345、当经验池样本数大于等于经验池容量时,通过随机采样确定训练集,分别计算值评估网络和动作评估网络梯度,并对动作目标网络和值目标网络参数进行更新。

8.一种组合导航系统,其特征在于,至少包括组合导航模块和自适应估计模块,其中,

所述组合导航模块,被配置为基于卡尔曼滤波构建,能够提供实时导航结果,采集并存

储组合导航系统最近的固定时长IMU和GNSS数据，计算定位误差，若定位误差大于阈值，构建初始惯导过程噪声协方差矩阵，并将数据发送至自适应估计模块，根据自适应估计输出调整后组合导航结果；

所述自适应估计模块，被配置为接收组合导航模块的数据，对过程噪声协方差矩阵进行基于强化学习的自适应估计，获得过程噪声协方差矩阵最优估计。

9.一种电子设备，包括中央处理器以及存储计算机可执行指令的存储器，其特征在于，所述计算机可执行指令在被执行时使所述处理器执行根据权利要求1-7中任一项所述的方法。

10.一种非易失性存储介质，其中存储有计算机程序，其特征在于，所述计算机程序在执行时实现权利要求1-7中任一项的方法。

一种组合导航方法和系统

技术领域

[0001] 本发明属于导航控制技术领域,具体涉及一种组合导航方法和系统。

背景技术

[0002] 近年来,高可靠性的自主导航系统越来越受到人们的关注。在行业需求方面,自主导航系统可用于移动测量,提供精密的定位定姿技术,为测绘行业带来革命性的变化;在大众需求方面,以智能手机,无人机,自动驾驶汽车,移动机器人为代表的智能化载体,在进行自主运动时,高度依赖精密位置信息,自主导航系统是其环境感知与决策控制的基础和核心。就自动驾驶而言,自主导航系统的市场潜力巨大。与巨大的市场潜力相对的是目前自动驾驶技术的不成熟,与其相关的精准定位技术仍有欠缺。目前定位-导航-授时(Position,Navigation,Timing,PNT)技术的重要发展方向,即为多传感器集成,多源异质信息融合。基于实时动态差分技术(Real-time Kinematic,RTK)和惯性导航系统(Inertial Navigation System,INS)的组合导航是目前主流的室外高精度算法。为了提高在复杂场景下,RTK不可用时的定位精度,近年来学者们提出了各种自适应算法,主要分为伪量测自适应估计,量测噪声协方差矩阵自适应估计和过程噪声协方差矩阵估计三个方面。伪量测自适应估计是指在卡尔曼滤波的过程中,通过训练好的模型对当前的量测值进行估计,用伪量测值进行卡尔曼滤波的量测更新,以减小组合导航过程中量测不可用时的定位误差。量测噪声协方差矩阵(R阵)是衡量当前量测的可靠度的方式。通过对其进行自适应学习,可以有效的衡量当前量测值和预测值在最终导航结果中的正确占比,从而得到较好的定位结果。本专利主要针对对过程噪声协方差矩阵(Q阵)进行自适应学习的情况,目的在于通过传感器数据自主自适应地得到最能表征当前惯性传感器状态的过程噪声协方差矩阵,从而摆脱过程噪声协方差矩阵对先验知识的依赖,减小温度,气压的变化对定位结果的影响。

[0003] 过程噪声协方差矩阵由惯性测量单元(Inertial Measurement Unit,IMU)的性能决定。一般情况下,传感器器件在出厂时,商家都会给各个噪声参数一个初始值的范围,但该人工标定的值往往不能衡量当前器件的真实水平。由于自身特性,过程噪声协方差矩阵在一次导航过程中不会有太大的变化。但是由于会受到温度,气压等因素的影响,过程噪声协方差矩阵的估计也无法做到一劳永逸,需要对其进行自适应估计。但现有的组合导航方法中对过程噪声协方差矩阵的估计算法存在灵活性低,鲁棒性差,梯度反向传播困难的问题。

发明内容

[0004] 针对现有技术中命名数据网络泛滥式缓存数据的问题,以及上述现有方案存在的不足,本发明提出了一种组合导航方法和系统。

[0005] 为达到以上目的,一方面,本发明提出了一种组合导航方法,包括以下步骤:

[0006] 步骤S1、采集并存储组合导航系统最近的特定时长内的组合导航定位数据;

[0007] 步骤S2、计算当前组合导航定位误差,判断误差是否大于阈值;

[0008] 步骤S3、若定位误差大于阈值，构建初始惯导过程噪声协方差矩阵，对过程噪声协方差矩阵进行基于强化学习的自适应估计，得到过程噪声协方差矩阵最优估计；

[0009] 步骤S4：优化预测值与量测值的相对权重，输出调整后组合导航结果。

[0010] 优选地，所述步骤S1的组合导航系统为基于卡尔曼滤波构建，作为强化学习模型的环境。

[0011] 优选地，所述步骤S1的最近组合导航定位数据作为强化学习模型的训练数据，训练数据的时间长度由定位精度、训练所需时间和数据存储性能确定。

[0012] 优选地，所述步骤S3的过程噪声协方差矩阵由过程噪声及其变化量确定，作为强化学习模型的状态，其定义为：$Q = E[e^{W_k+\alpha_k}(e^{W_k+\alpha_k})^T]$，其中，过程噪声为：$W_k = [w_{\delta r}, w_{\delta v}, w_{\varphi}, w_q, w_{\nabla}, w_{\delta kii}]^T$，$w_{\delta r}$为位置游走系数，$w_{\delta v}$、$w_{\varphi}$、$w_q$、$w_{\nabla}$、$w_{\delta kii}$分别速度、角度、角速率、加速度和比例误差随机游走；α_k为过程噪声W_k变化量，作为强化学习模型的动作。

[0013] 优选地，对所述步骤S3基于强化学习的自适应估计方法是在深度确定性策略梯度DDPG基础上，以所述步骤S1的组合导航数据为输入，以过程噪声协方差矩阵最优估计为输出，包括如下步骤：

[0014] 步骤S31、构造动作评估网络和值评估网络；

[0015] 步骤S32、初始化两个网络的权重；

[0016] 步骤S33、初始化经验池大小和每轮训练数据个数。

[0017] 步骤S34、对于每一轮的每组训练数据分别计算值评价网络和动作评价网络的梯度，并更新网络权值；

[0018] 步骤S35、得到过程噪声协方差矩阵最优估计。

[0019] 优选地，所述步骤S31的动作网络输入维度与所述状态的维度相等，输出维度与所述动作的维度相等；所述值评估网络输入由状态和动作共同构成，其维度为状态和动作维度之和，输出维度为1；

[0020] 优选地，所述步骤S34包括如下步骤：

[0021] 步骤S341、动作评估网络根据当前状态计算得到一个动作估计；

[0022] 步骤S342、组合导航模型计算得到下一个状态；

[0023] 步骤S343、计算过程噪声协方差矩阵，对定位误差取反得到奖励；

[0024] 步骤S344、保存转换关系到经验池中；

[0025] 步骤S345、当经验池样本数大于等于经验池容量时，通过随机采样确定训练集，分别计算值评估网络和动作评估网络梯度，并对动作目标网络和值目标网络参数进行更新。

[0026] 另一方面，本发明还提供一种组合导航系统，至少包括组合导航模块和自适应估计模块，其中，

[0027] 所述组合导航模块，被配置为基于卡尔曼滤波构建，能够提供实时导航结果，采集并存储组合导航系统最近的固定时长IMU和GNSS数据，计算定位误差并将数据发送至自适应估计模块，根据自适应估计输出调整后组合导航结果；

[0028] 所述自适应估计模块，被配置为接收组合导航模块的数据，对过程噪声协方差矩阵进行基于强化学习的自适应估计，获得过程噪声协方差矩阵最优估计。

[0029] 再一方面,本发明提供一种电子设备,包括中央处理器以及存储计算机可执行指令的存储器,所述计算机可执行指令在被执行时使所述处理器执行所述方法。

[0030] 第四方面,本发明提供一种非易失性存储介质,其中存储有计算机程序,所述计算机程序在执行时实现所述方法。

[0031] 本发明相对于现有技术取得了如下的技术效果:

[0032] 本发明的方法和系统能够为单个具体惯导设备提供个性化的组合导航方案,避免对先验知识的依赖;所采用的模型学习能力强,调优速度快,解决了梯度回传困难的挑战;提高组合导航系统的定位准确度,在不同环境下具有强鲁棒性。

附图说明

[0033] 以下,结合附图来详细说明本发明的实施例,其中:

[0034] 图1示出本发明实施例的一种组合导航方法流程图;

[0035] 图2示出本发明实施例的一种基于强化学习的组合导航系统示意图;

[0036] 图3示出本发明实施例的DDPG算法应用在过程噪声协方差矩阵自适应估计的示意图;

[0037] 图4示出本发明实施例的DDPG内部动作和值网络架构图;

[0038] 图5示出本发明实施例的2019年7月17号数据轨迹图;

[0039] 图6示出本发明实施例的学习到的Q阵在2019年7月17号数据上的导航误差累积概率图;

[0040] 图7示出本发明实施例的2019年7月17号真值与组合导航结果对比图;

[0041] 图8示出本发明实施例的不同算法在2019年7月17号GNSS/INS系统下GNSS缺失10s时的导航准确度对比图;

[0042] 图9示出本发明实施例的不同算法在2019年7月10号GNSS/INS/ODO系统下GNSS缺失300s时的导航准确度对比图;

[0043] 图10示出本发明实施例的不同算法在2019年7月10号GNSS/INS/NHC系统下GNSS缺失300s时的导航准确度对比图。

具体实施方式

[0044] 下面结合附图和具体实施方式对本发明做进一步说明。

[0045] 为了使本发明的目的、技术方案、设计方法及优点更加清楚明了,以下结合附图通过具体实施例对本发明进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。下面结合附图和具体实施方式对本发明作进一步描述。

[0046] 首先给出本发明用到的缩略语和解释:

[0047] PNT:Position,Navigation,Timing,定位、导航、授时;

[0048] INS:Inertial Navigation System,惯性导航系统;

[0049] RTK:Real-time kinematic,实时动态载波相位差分技术;

[0050] IMU:Inertial Measurement Unit,惯性测量单元;

[0051] GPS:Global Positioning System,全球定位系统;

[0052] GNSS,Global Satellite Navigation System,全球卫星定位导航系统;

[0053] DDPG,Deep Deterministic Policy Gradient,深度确定性策略梯度,是针对连续行为的策略学习方法,该算法是Actor-Critic和DQN(Deep QNetwork,深度Q网络)的结合体,克服了Actor-Critic收敛慢的问题;

[0054] RL,Reinforcement Learning,强化学习,是利用环境反馈的奖励信息,通过试错法获得最优估计的算法框架,具有较强的自主探索能力。

[0055] 实施例

[0056] 本发明的技术方案涉及到如下几个方面:1)过程噪声协方差矩阵的取值对先验知识的依赖。在传统卡尔曼滤波组合导航算法中,过程噪声协方差矩阵Q和量测噪声协方差矩阵R的取值对卡尔曼滤波的性能有很大的影响。若使用有较大误差的Q阵和R阵,则必然会降低组合导航的准确度,甚至有时会导致滤波发散。在实际应用中,R矩阵的参数取值往往取决于当前收到GNSS量测的质量,而Q矩阵则依赖于特定IMU器件性能的先验知识。2)Q矩阵没有真值可做校准,且参与的计算极为复杂。过程噪声协方差矩阵依赖于传感器的固有性能,无法通过实验装置测得其真实值,只能通过当前时刻的定位误差来间接反映出当前所使用的Q矩阵的优劣。然而,Q矩阵在组合导航系统中参与的计算过于复杂。若要通过定位误差对Q矩阵的取值进行调整,需要克服梯度回传困难的挑战。3)Q矩阵的半正定性要求。Q矩阵实际上是传感器的噪声向量的协方差矩阵,组合导航系统要求该矩阵要时刻保持半正定性。因此,在对Q阵进行自适应估计的过程中,必须保证对过程噪声协方差矩阵预测的每一个状态都是半正定的。4)Q矩阵会随着环境的变化而变化。通常情况下,传感器生产厂商会提供某一类型号的IMU期间在标准工作环境(温度,湿度,气压等)下的默认Q参数。然而,同一类型号的不同器件本身的噪声性能并不能一概而论,而且器件的Q阵参数会随着工作环境而缓慢变化。为了提高组合导航系统的性能,并减小Q阵估计对先验知识的依赖,发明人对单个传感器的过程噪声协方差矩阵自适应估计进行了长期深入的研究。

[0057] 根据本发明的一个实施例,提出一种组合导航方法(以下简称RL-AKF),如图1所示,该方法包括:首先,采集并存储组合导航系统最近的特定时长内的组合导航定位数据;其次,构建惯导过程噪声协方差矩阵差;计算当前组合导航定位误差,判断误差是否大于阈值;若定位误差大于阈值,对过程噪声协方差矩阵进行基于强化学习的自适应估计,得到过程噪声协方差矩阵最优估计,重新确定预测值与量测值权重,输出调整后组合导航结果;若定位误差小于等于阈值,直接输出组合导航结果。

[0058] 进一步地,RL-AKF主要由基于卡尔曼滤波的嵌入式平台组合导航和基于增强学习的自适应过程噪声协方差矩阵估计模块两部分组成,该系统的架构如图2所示。卡尔曼滤波组合导航模块运行在嵌入式平台上,通过IMU和GNSS数据组合导航获取初步的实时导航结果,同时,采集并存储最接近当前时刻的N秒的IMU和GNSS数据。在有RTK固定解信号时,对当前的定位误差进行评判。若定位误差大于阈值,则将最近N秒的传感器数据发送到基于强化学习的过程噪声协方差矩阵自适应估计模块。该模块的目标是学习得到一个最优的过程噪声协方差矩阵,能使采集到的数据的定位误差最小。

[0059] 强化学习的各元素定义和说明如下所示:

[0060] 状态s:定义为当前的过程噪声向量W的取值,本实施例中待估计的目标矩阵Q是过程噪声向量W的协方差;

[0061] 动作α:定义为噪声向量W的变化。由于共有6个不同的噪声来源构成了噪声向量W,

所以α的维度也是6。显然的是，该问题中的动作是定义在连续的实数域空间上的，无法进行穷举；

[0062] 奖励r：定义为定位误差的相反数，奖励值越高，说明当前的导航定位效果越好；

[0063] 环境E：定义为GNSS/INS组合导航模型，一旦环境系统受到一个新的动作，它会更新当前的状态，并计算此时的定位误差获得奖励值，然后将状态，动作，奖励等一系列信息送到代理A。

[0064] 代理A：该部分是强化学习中的自适应估计模块，本发明实施例使用DDPG算法来克服动作空间的连续性问题。

[0065] 该部分的具体算法流程为：

输入：连续 IMU 和 GNSS 传感器数据

输出：当前的位置、速度和姿态导航结果

1. 根据加速度和陀螺仪数据获取对下一时刻导航状态的预测；
2. 若当前 GPS 定位状态为差分定位，则转 3，否则转 4；

[0066]

3. GNSS 量测状态与 INS 预测状态经过卡尔曼滤波融合后输出当前的位置、速度和姿态导航结果，转 1；
4. 计算当前的 INS 预测结果的位置误差，若误差小于阈值，转 3，否则转 5；
5. 将距离当前时刻最近的 N 秒数据传输到后台，进行过程噪声协方差矩阵优化；
6. 优化后的过程噪声协方差矩阵替换之前的过程噪声协方差矩阵，继续组合导航。

[0067] 传感器制造商通常会为一种型号的传感器提供一组默认值，本发明实施例给出过程噪声协方差矩阵的各分量含义及其对应型号的初始值：

[0068] $$W_k = [w_{\delta r} \quad w_{\delta v} \quad w_{\varphi} \quad w_q \quad w_{\nabla} \quad w_{\delta kii}]^T$$

[0069] $$Q = E[W_k W_k^T]$$

[0070] 上式中W_k表示过程噪声向量，包括了6种噪声组成，共有21维，Q是过程噪声向量的协方差矩阵，其各个组成部分的含义及其初始值如下表所示。

[0071]

符号	含义	维度	M39	CPT
$w_{\delta r}$	位置游走系数	3	$0m/\sqrt{h}$	$0m/\sqrt{h}$
$w_{\delta v}$	速度随机游走	3	$0.09m/s/\sqrt{h}$	$\sim m/s/\sqrt{h}$
w_{ϕ}	角度随机游走	3	$0.12deg/\sqrt{h}$	$0.067deg/\sqrt{h}$
w_q	角速率随机游走	3	$180deg/\sqrt[2]{h^3}$	$\sim deg/\sqrt[2]{h^3}$
w_{∇}	加速度随机游走	3	$8\mu Gal/\sqrt{h}$	$\sim \mu Gal/\sqrt{h}$
$w_{\delta kii}$	比例误差随机游走	6	$1000ppm/\sqrt{h}$	$4000ppm/\sqrt{h}$

[0072] 为了保证Q阵的半正定性，本发明将动作定义在过程噪声向量上而非直接改变Q矩阵对角线上的值。同时，为了保证噪声值是有意义的，应对每一个状态下的噪声向量进行非负限制。为了满足以上条件，本发明将Q矩阵的计算方法重新定义为：

$Q = E[e^{W_k}(e^{W_k})^T]$。当环境在当前状态值下，接收到新的动作时，Q阵的计算方法：$Q = E[e^{W_k+\alpha_k}(e^{W_k+\alpha_k})^T]$。

[0073] 进一步地，过程噪声协方差矩阵的取值会极大的影响INS导航结果在组合导航系统中的性能，因此，组合导航系统会对当前的过程噪声协方差矩阵一个有效的反馈。本发明实施例将强化学习中的环境定义为基于卡尔曼滤波的组合导航系统。该系统会对当前的协方差矩阵的优劣进行评估，从而得到奖励值。当环境接收到状态更改的操作后，需要定义一个有效的奖励机制来评估当前的动作。对本问题而言，该系统将奖励建模为INS和GNSS组合导航系统在N秒内的定位误差的相反数。本实施例从N秒的数据中选出m个长度为l_m的数据段。为了模拟GNSS信号的缺失，在这些时间段内，不会进行GNSS的量测更新。计算每个数据段结束时的定位误差，统计其均方根，即可得到当前过程噪声协方差矩阵下的最终导航精度，其相反数也即为当前动作的奖励值。通过使奖励最大化，可以最终得到能够使定位误差最小的过程噪声协方差矩阵。

[0074] 进一步地，该问题的动作空间定义在实数域空间$[\mathcal{L}_p, \mho_p]$上。以下描述一种能够从连续空间中选择最优估计的网络架构DDPG。DDPG算法是一种基于确定性策略梯度的Model-Free算法，可以在连续的动作空间上执行。DDPG算法在自适应过程噪声协方差矩阵估计中的应用框架如图3所示。在每一轮中，DDPG都会执行一定的时间步长。在每个时间步，动作α_i由ACTOR模块的动作评估网络进行选择。GNSS/INS集成导航模型执行此操作，计算奖励并将转换数据存储在经验池中。当其中存储的样本条数满足要求时，DDPG算法将随机抽取批量大小的训练样本，分别计算值评价网络和动作评价网络的梯度，然后更新网络权值进行下一轮计算。

[0075] DDPG算法的网络框架主要分为Actor和Critic两个部分，每个部分各有一个评估网络和目标网络。各个网络的功能定位如下：

[0076] Actor评估网络：负责策略网络参数的迭代更新，根据当前的过程噪声协方差矩阵值，选择当前调整方案，即当前动作，用于和环境交互生成下一时刻的过程噪声协方差矩阵和奖励值；

[0077] Actor目标网络：负责根据从经验池中采样的下一过程噪声协方差矩阵选择下一个最优动作，网络参数从Actor评估网络参数中定期复制；

[0078] Cri t ic评估网络：负责价值计算网络参数的迭代更新，负责从当前的过程噪声协方差矩阵，动作中计算当前的奖励值，，目标Q值$y_i = R + \gamma Q'(S', A', w')$；

[0079] Critic目标网络：负责计算Q值中的$Q'(S', A', w')$部分，网络参数从Critic评估网络中定期复制。

[0080] 以上算法是图2中“过程噪声协方差矩阵自适应学习”模块的具体实施方案。经过DDPG算法的模拟和学习，可以在当前的定位环境下获得最能表征当前IMU性能的过程噪声协方差矩阵值，从而反馈到嵌入式组合导航平台，进行高精度组合导航。

输入：N秒的连续IMU和GNSS传感器数据

输出：学习得到的最优过程噪声协方差矩阵估计结果

1. 构造动作评估网络和价值评估网络。动作网络的输入维度与状态s的维度一致，为21，输出维度与动作一致，为6。值评估网络的输入由状态和动作共同构成，维度为27，输出维度为1。

[0081] 2. 初始化两个网络的权重θ^{μ}和θ^{v}。复制评估网络的参数到动作目标网络和价值目标网络：$\theta^{\mu} \rightarrow \theta^{\mu'}, \theta^{v} \rightarrow \theta^{v'}$。

3. 初始化经验池，容量为100；初始化batchsize，大小为32。
4. While 每一个轮
   a) While 每一个时间步
      i. 动作评估网络根据当前状态s_t计算得到一个动作估计a_t；
      ii. 导航模型计算得到下一个状态：$s_{t+1} = s_t \times e^{a_t}$；
      iii. 计算新的过程噪声协方差矩阵，得到定位误差e_t和奖励$r_t = -e_t$；
      iv. 保存转换关系(s_t, a_t, s_{t+1}, r_t)到经验池中；
      v. IF 经验池的样本数≥经验池的容量
      vi. 随机采样batchsize个样本数作为训练集；
      vii. 计算值评估网络的梯度并更新θ^{v}，计算公式为：

$$\theta_{t+1}^{v} = \theta_t^{v} - \eta \times \nabla_{\theta_t^{v}} L(\theta_t^{v}) = \theta_t^{v} - \eta \times \nabla_{\theta_t^{v}} \left(\frac{1}{N}\sum_{i=1}^{N}(y_i - v(s_i, a_i|\theta_t^{v}))^2\right)$$

[0082] viii. 计算动作评估网络的梯度并更新θ^{μ}，计算公式为：

$$\theta_{t+1}^{\mu} = \theta_t^{\mu} - \eta \times \nabla_{\theta_t^{\mu}} L\left(\theta_t^{\mu}\right)$$

$$= \theta_t^{\mu} - \eta \times \nabla_{\theta_t^{\mu}} \left(\frac{1}{N}\sum_{i=1}^{N}(\nabla_{\alpha} v(\mu(s_i), a_i|\theta_t^{v}) \times \nabla_{\theta_t^{\mu}} \mu(s_i|\theta_t^{\mu}))\right)$$

      ix. 对动作目标网络和值目标网络的参数进行软更新

$$\theta^{\mu\prime} = \beta\theta^{\mu} + (1-\beta)\theta^{\mu'}, \theta^{v\prime} = \beta\theta^{v} + (1-\beta)\theta^{v'}$$

      x. End if
   b) End 时间步
5. End 当前计算轮
6. return $\mu(s_0|\theta^{\mu})$

[0083] DDPG算法中所采用的网络架构为MLP(Multi-Layer Perception,多层感知机)，但ACTOR部分和CRITIC部分的具体实现方法略有不同。由于该问题中的状态和动作的维度固定，所以两个网络结构也不会有变化，具体的网络结构图如图4所示。动作评估网络和动作目标网络是一个三层全连通网络。输入层的维数为21，隐含层的维数为30，接以ReLU激活函数。输出层的大小等于与动作的维度相同。最终，将tanh激活函数应用于输出层神经元，即

可得到预测的动作，其网络结构如图4(a)所示。价值评估网络与价值目标网络的网络结构如图4(b)所示。输入维度是s_dim+a_dim＝27。状态输入和动作输入分别获得30维隐藏层h1和h2。对两个激活层的值相加后的结果应用ReLU激活函数，得到隐含层的输出结果。最后，通过全连通层获得当前的值估计。

[0084] 下面对本发明实施例的实验验证效果进行说明。

[0085] 本发明实施例的RL-AKF算法在不同的训练序列长度上得到了不同的训练结果和定位性能。不同训练序列下所得到的优化Q矩阵罗列如下表所示，其中Q[1:3]恒为0。

[0086]

序列长度	设备型号	Q[4:6]	Q[7:9]	Q[10:12]	Q[13:15]	Q[16:21]	训练时间
100s	M39	7.1668×10^{-7}	3.3580×10^{-9}	8.2571×10^{-13}	2.0095×10^{-9}	1.0617×10^{-10}	7889
	CPT	1.5860×10^{-6}	8.6049×10^{-10}	1.2858×10^{-12}	1.9014×10^{-9}	5.9887×10^{-10}	3176
200s	M39	6.8444×10^{-7}	3.2846×10^{-9}	8.2571×10^{-13}	2.4998×10^{-9}	1.2317×10^{-10}	7918
	CPT	1.4703×10^{-6}	1.0001×10^{-10}	1.4058×10^{-12}	1.8014×10^{-9}	6.5556×10^{-10}	4926
300s	M39	6.8444×10^{-7}	3.3161×10^{-9}	8.2571×10^{-13}	1.9000×10^{-9}	1.3424×10^{-10}	10240
	CPT	1.5320×10^{-6}	9.0008×10^{-10}	1.3958×10^{-12}	1.9014×10^{-9}	5.3925×10^{-10}	6558
400s	M39	6.8444×10^{-7}	3.4921×10^{-9}	9.3571×10^{-13}	1.9762×10^{-9}	1.5427×10^{-10}	12521
	CPT	1.5316×10^{-6}	6.7272×10^{-10}	1.2558×10^{-12}	2.2013×10^{-9}	5.3867×10^{-10}	7410

[0087] 不同的参数在测试集上的定位准确度可参见下表，其中，p、v、c分别表示位置，速度和航向角，*_67％和*_90％分别表示对误差序列进行排序后，67％处(即一倍中误差)和90％处的误差值。*_rms表示误差序列均方根，即root mean square。可以看出，定位误差会随着训练序列长度的变化而变化。一般来说，训练序列的长度越长，定位误差越小，同时，训练用时也会增长。以M39为例，当训练序列的长度从100s增加到200s时，定位误差降低了12.84％，而训练时长仅增加了0.36％。当训练序列长度增加到300s时，训练时长增加了29.32％，但定位误差仅减小了0.92％。SPAN-CPT设备上的实验结果也能得到类似的结论。训练序列的长度也会直接影响到嵌入式平台的存储容量需求。以SPAN-CPT设备为例，IMU的数据采集频率为100HZ，每条数据包含时间，三轴加速度和三轴陀螺仪。假设数据以32位的格式存储，则IMU数据每秒需要2.8KB的存储空间。GNSS信号采集频率为1Hz，每秒包含7个数据，分为时间，位置和速度。因此，GNSS数据每秒需要0.028KB的存储空间。若训练序列长度为N，则嵌入式平台需要2.828*N KB的额外存储空间。综合考虑定位准确度，训练时间和存储需求，本发明选定训练序列为200作为最终的测试方案。

[0088]

设备	评价指标	100s	200s	300s	400s
M39	p_67%(m)	0.6762	0.5782	0.5775	**0.5693**
	p_90%(m)	0.8564	0.7625	0.7600	**0.7568**
	p_rms(m)	0.7477	0.6517	0.6457	**0.6424**
	v_67%(m/s)	0.0578	0.0579	0.0578	**0.0574**
	v_90%(m/s)	0.1433	0.0987	**0.0931**	0.0958
	v_rms(m/s)	0.1023	0.0904	0.0896	**0.0826**
	c_67%(deg)	0.1717	0.1437	0.1423	**0.1256**
	c_90%(deg)	0.3024	0.2974	0.2972	**0.2961**
	c_rms(deg)	0.2415	0.1663	0.1659	**0.1543**
CPT	p_67%(m)	0.4504	0.4012	0.3928	**0.3498**
	p_90%(m)	0.6982	0.6787	0.6361	**0.5918**
	p_rms(m)	0.5629	0.4963	0.4937	**0.4846**
	v_67%(m/s)	0.0369	0.0351	0.0346	**0.0337**
	v_90%(m/s)	0.0979	0.0828	0.0812	**0.0710**
	v_rms(m/s)	0.0736	0.0586	0.0573	**0.0524**
	c_67%(deg)	0.3924	0.3809	0.3736	**0.3294**
	c_90%(deg)	0.8649	0.8184	0.7841	**0.6943**
	c_rms(deg)	0.4710	0.4651	0.4525	**0.4186**

[0089] 当导航环境不发生明显变化时，同一个设备的过程噪声协方差矩阵也不会有明显的改变。为了测试在7月10号训练数据上学到的结果是否可以在一段时间内保持鲁棒性，本发明在2019年7月17号使用M39设备收集了另一组数据。该段数据从GPS时间294514s开始，采集至303058s，路线图如图5所示。可以看出，该路段中包含较多的直线和高速行驶路段，与训练数据集有很大的不同。图6绘制了从7月10号训练路段学习到的过程噪声协方差矩阵在7月17号整个路段上的导航误差累积概率分布图。从图中可以看出，经过10s的GNSS缺失后，90％的测试点定位误差都保持在亚米级，速度误差仅为0.11m/s。图7对比了组合导航系统的结果和真值的导航轨迹。图中星号表示组合导航系统对当前位置的预测，圆圈表示真值所在的位置。在三角区域代表的时刻，本系统将不执行GNSS量测更新用以模拟当前时间段GNSS信号的终端。为了显示定位细节，本发明截取了测试数据中299770至302500的时间段，这期间包括三种特殊情况：1)拔下GNSS天线100s，真实模拟GNSS信号的缺失；2)车辆以1m/s的速度缓慢转弯；3)GNSS位置漂移。试验结果表明，该组合导航系统在各种环境下都能实现有效定位。图8对比了不同算法在7月17号测试数据上的导航误差，对比数据如下表。

[0090]

Item	RL-AKF	N-M	Cov-scale	NN-fb	Default
p_67%(m)	0.5959	0.9847	0.9174	1.2035	1.2309
p_90%(m)	0.7558	4.3011	1.3387	1.7144	1.6722
p_rms(m)	0.6075	2.8496	0.9978	1.4251	1.3425
v_67%(m/s)	0.0703	0.1354	0.1371	0.2145	0.2034
v_90%(m/s)	0.1105	0.4911	0.1709	0.3725	0.3459
v_rms(m/s)	0.0741	0.6116	0.1341	0.2457	0.2478
c_67%(deg)	1.4520	2.3270	1.2643	2.3326	2.3425
c_90%(deg)	1.8238	6.7497	1.5780	4.4513	4.4615
c_rms(deg)	1.6548	3.3338	1.2132	3.0215	3.0524

[0091] 实验结果表明,本发明提出的RL-AKF算法整体表现良好,在67%处的定位误差仅为0.60米,Nelder-Mead算法的学习能力最弱,NN-feedback的定位误差与默认的过程噪声协方差矩阵相当。大量的实验表明,本发明所提出的RL-AKF算法在一段数据集上学习到的过程噪声协方差矩阵能够很好的反映当前设备的固有噪声水平,从而可以在一段时间内依然保持鲁棒性,避免了频繁的前后端数据传输需求。

[0092] 在实际的定位场景中,GNSS中断时刻难以预测,中断的持续时间更是受到多方面因素的影响。为了评估本发明提出的RL-AKF算法在不同的GNSS终端时间周期下的定位精度,本发明分别比较了不同的算法在GNSS缺失10s,20s,…,60s时的定位误差。位置误差统计结果如下表所示,从表中可以看出,RL-AKF算法的定位误差与GNSS信号缺失时间之间存在二次项关系。采用Nelder-Mead算法学习到的过程噪声协方差矩阵在GNSS缺失较长时间时,会快速发散。Cov-Scale算法可以在一定程度上拟合传感器设备的固有噪声参数,降低定位误差。但从整体上来看,RL-AKF算法在全方位上优于其他算法。

[0093]

Periods	RL-AKF	N-M	Cov-scale	NN-fb	Default
10s	0.6517	0.7724	0.8773	1.1024	1.0578
20s	1.9730	5.7627	2.4160	3.9617	4.2157
30s	5.1794	28.8819	7.2712	9.2458	9.6138
40s	8.3244	46.1954	11.8819	17.3697	16.3254
50s	14.9237	-	18.5435	25.2157	25.1732
60s	20.4279	-	24.7773	31.2648	32.2154

[0094] 根据所携带设备和定位场景的不同,组合导航系统中所用到的量测也随时可能变化。因此,作者希望从GNSS/INS系统中学习得到的过程噪声协方差矩阵也能对各种组合导航量测更新方案保持鲁棒性。里程计通过提供可靠,低噪声的速度量测,在组合导航领域得到的越来越多的重视。在GNSS信号缺失期间,可以为卡尔曼滤波系统提供可靠度高的速度量测更新。我们对比了从7月10号训练数据得到的过程噪声协方差矩阵在测试数据上使用GNSS/INS/ODO组合导航系统时的定位性能,误差对比结果如图9所示,下表给出了对比数据。由于里程计可以将二次发散的定位误差减小到线性发散,作者将GNSS信号缺失时间扩大至300s。从实验结果可以看出,RL-AKF相较于其他算法,能够获得更为准确的位置和航向估计。

[0095]

Item	RL-AKF	N-M	Cov-scale	NN-fb	Default
p_67%(m)	9.3911	26.4577	14.5501	21.4873	20.4578
p_90%(m)	28.7662	34.2566	27.2024	30.2443	31.2548
p_rms(m)	14.9426	29.9687	18.0201	25.1547	24.2454
v_67%(m/s)	0.1272	0.1427	0.0998	0.1143	0.1025
v_90%(m/s)	0.1853	0.1987	0.1367	0.2156	0.2438
v_rms(m/s)	0.1249	0.1723	0.0991	0.1467	0.1587
c_67%(deg)	0.7566	1.2543	0.9635	1.0259	1.2578
c_90%(deg)	1.0196	1.7628	1.7159	1.5946	1.5687
c_rms(deg)	0.6920	1.5324	1.1488	1.2913	1.2744

[0096] NHC(Non-Holonomic Constraints,非完整性约束)是另一种被广泛应用于基于卡尔曼滤波的车辆导航系统中的量测更新方法。该方法假设车辆在载体坐标系中的右向速度和天向速度总保持为0。发明人在该部分比较了使用不同算法估计得到的过程噪声协方差矩阵在GNSS/INS/NHC组合导航系统下的定位误差,统计结果如图10所示,下表给出了对比数据。从实验结果可以看出,NHC提供的速度约束在一定程度上弱于ODO系统,但在GNSS缺失300s时,定位误差在67%处也仅有12.7米。整体来说,RL-AKF算法可以在GNSS/INS/NHC组合导航系统中保持鲁棒性。

[0097]

Item	RL-AKF	N-M	Cov-scale	NN-fb	Default
p_67%(m)	12.7463	34.2574	17.7241	27.8423	26.2548
p_90%(m)	22.4210	40.5821	39.1731	40.2518	41.2658
p_rms(m)	15.3380	36.2158	20.9373	33.4861	32.2581
v_67%(m/s)	0.1668	0.2156	0.1730	0.2186	0.2027
v_90%(m/s)	0.4637	0.2542	0.2198	0.2437	0.2256
v_rms(m/s)	0.2370	0.2467	0.1675	0.2259	0.2122
c_67%(deg)	0.6523	1.2957	1.1481	1.2999	1.3284
c_90%(deg)	0.9581	1.9524	1.7248	1.7654	1.8766
c_rms(deg)	0.7211	1.4755	1.2218	1.5284	1.4657

[0098] 最后所应说明的是,以上实施例仅用以说明本发明的技术方案而非限制。尽管参照实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,对本发明的技术方案进行修改或者等同替换,都不脱离本发明技术方案的精神和范围,其均应涵盖在本发明的权利要求范围当中。

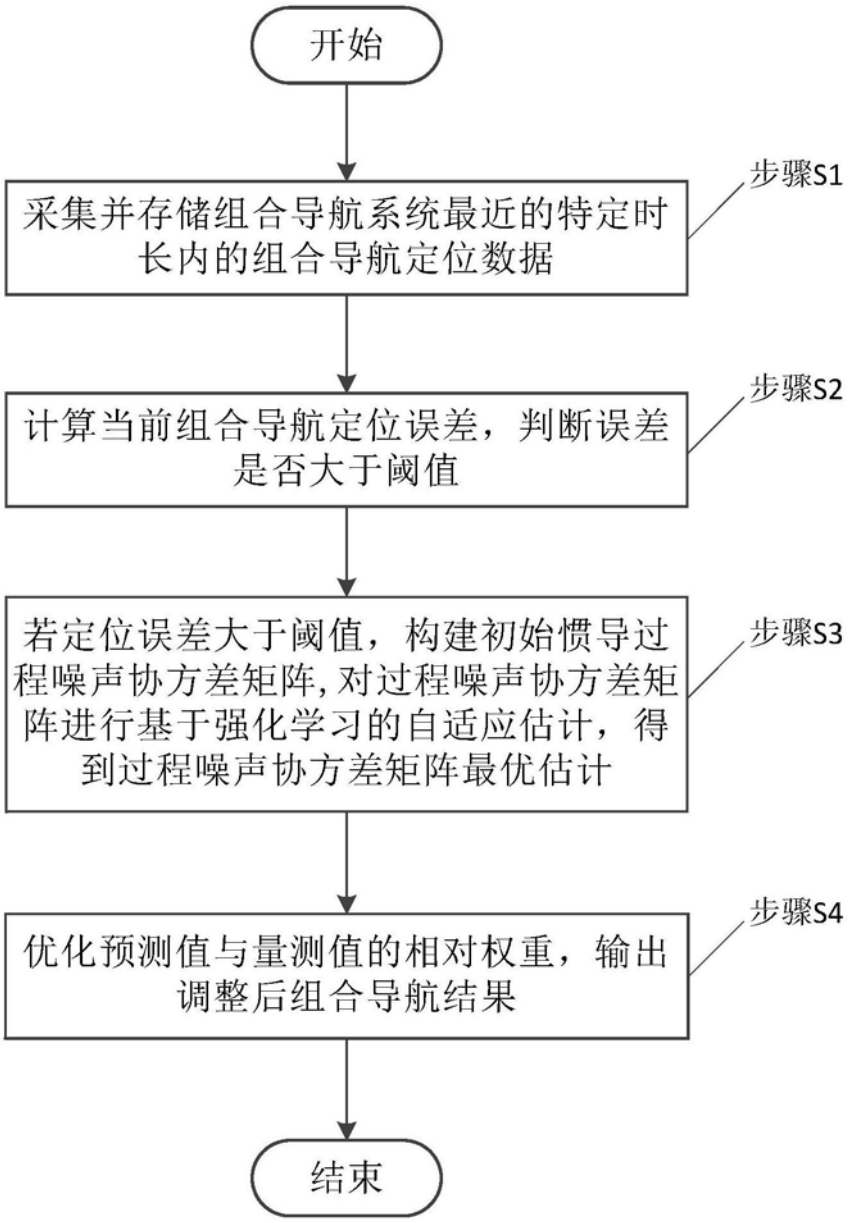

开始
采集并存储组合导航系统最近的特定时长内的组合导航定位数据
步骤S1
计算当前组合导航定位误差，判断误差是否大于阈值
步骤S2
若定位误差大于阈值，构建初始惯导过程噪声协方差矩阵,对过程噪声协方差矩阵进行基于强化学习的自适应估计，得到过程噪声协方差矩阵最优估计
步骤S3
优化预测值与量测值的相对权重，输出调整后组合导航结果
步骤S4
结束

图1

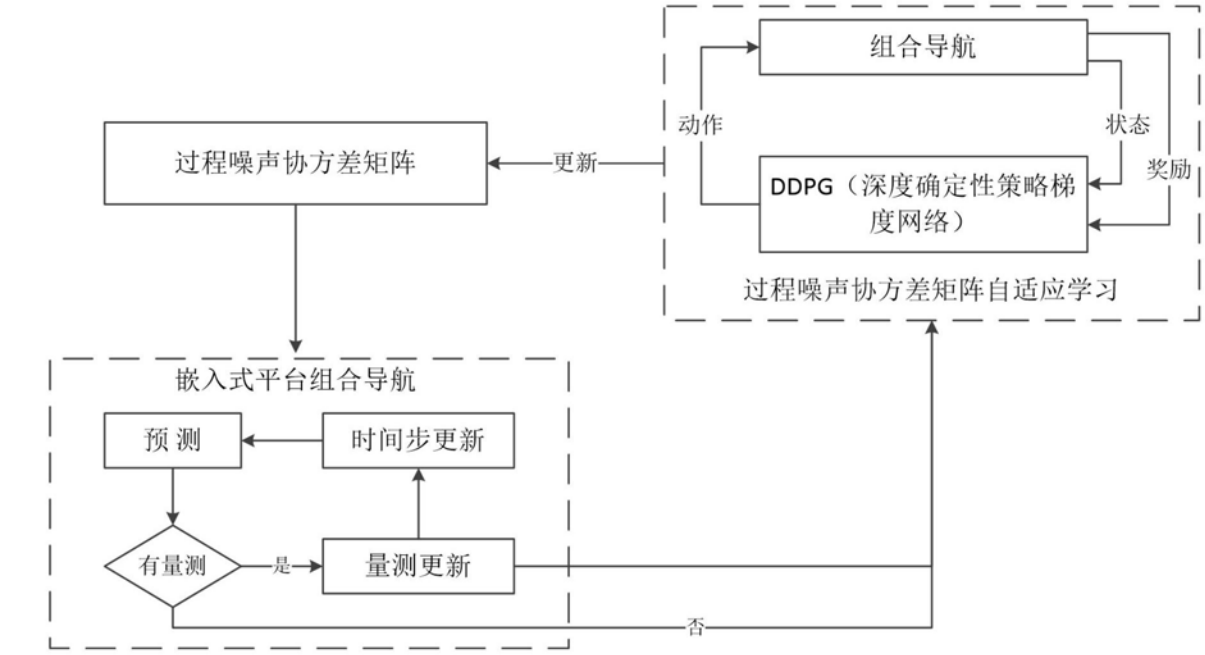

组合导航
动作
状态
奖励
过程噪声协方差矩阵
更新
DDPG（深度确定性策略梯度网络）
过程噪声协方差矩阵自适应学习
嵌入式平台组合导航
预 测
时间步更新
有量测
是
量测更新
否

图2

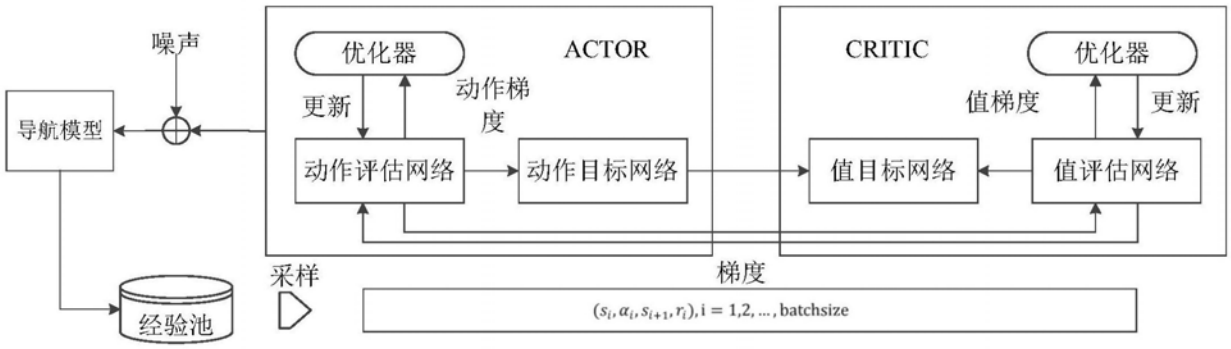

噪声
导航模型
优化器
ACTOR
更新
动作梯度
动作评估网络
动作目标网络
CRITIC
优化器
值梯度
更新
值目标网络
值评估网络
采样
梯度
经验池
(s_i, a_i, s_{i+1}, r_i), i = 1,2, … , batchsize

图3

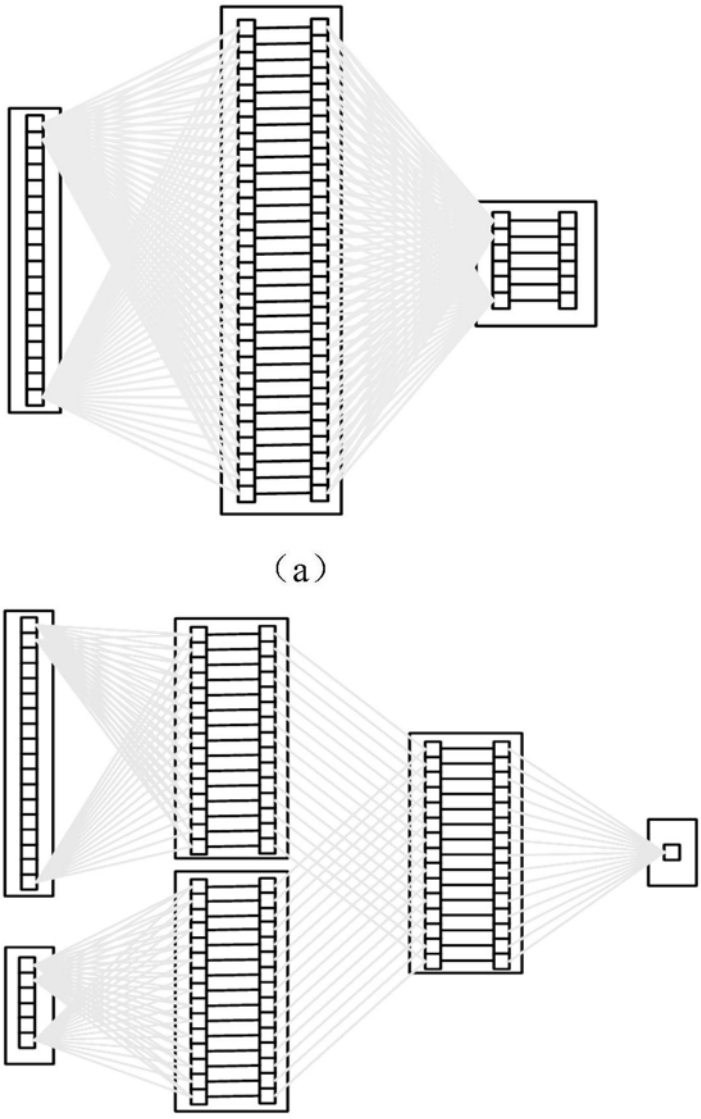


图4

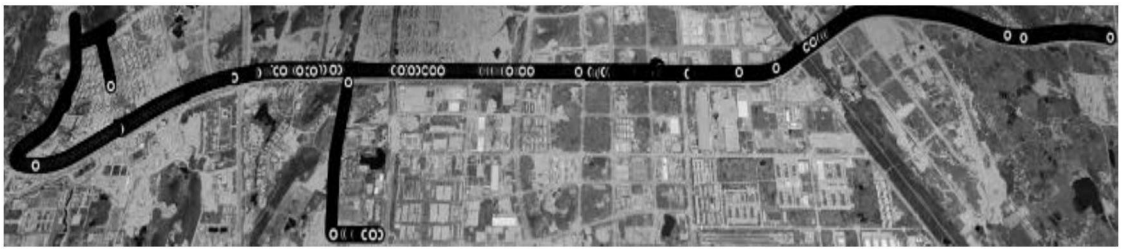
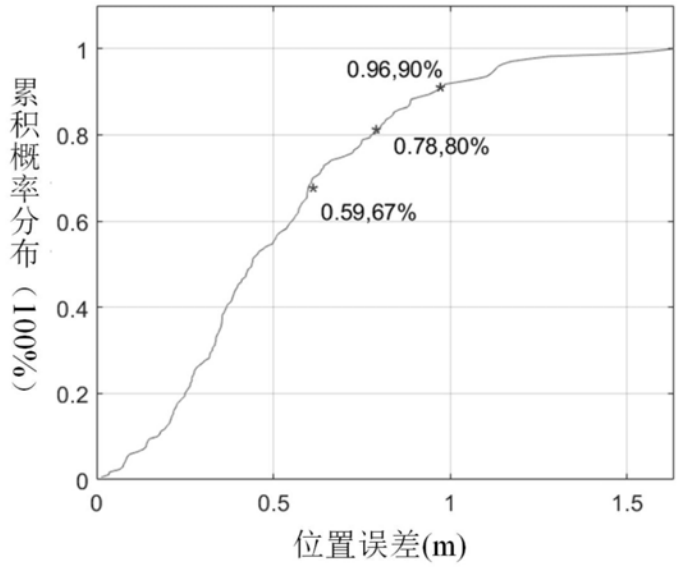
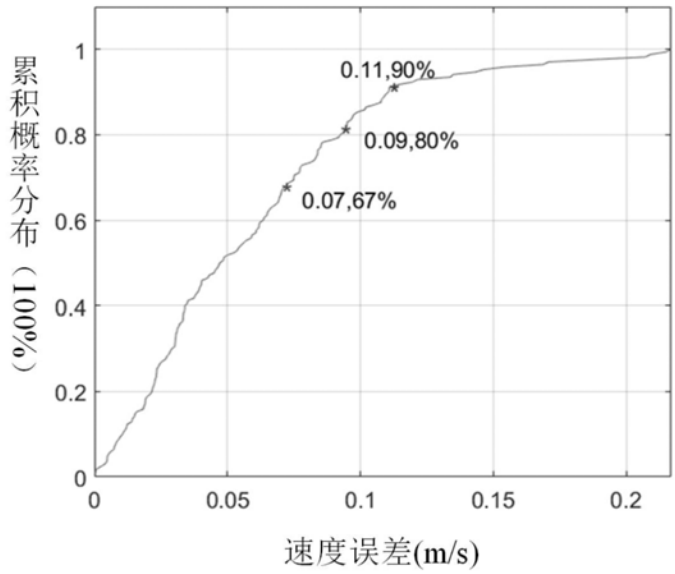


图5


累积概率分布（100%）
0.96,90%
0.78,80%
0.59,67%
位置误差(m)


累积概率分布（100%）
0.11,90%
0.09,80%
0.07,67%
速度误差(m/s)

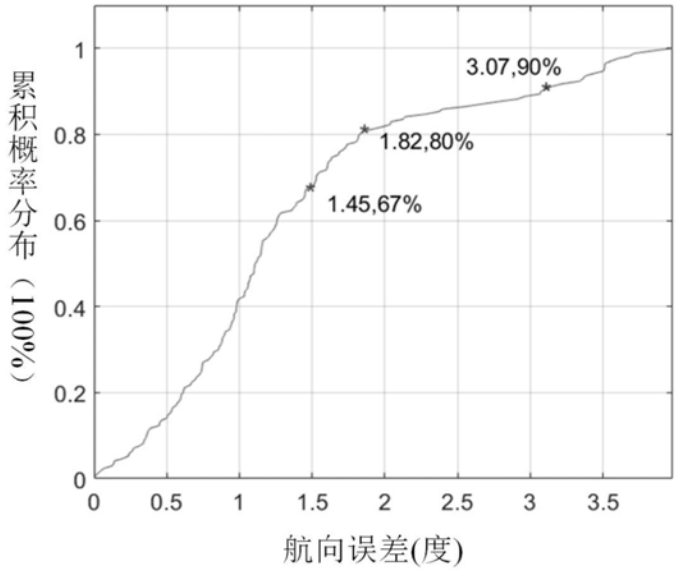

累积概率分布（100%）
3.07,90%
1.82,80%
1.45,67%
航向误差(度)

图6

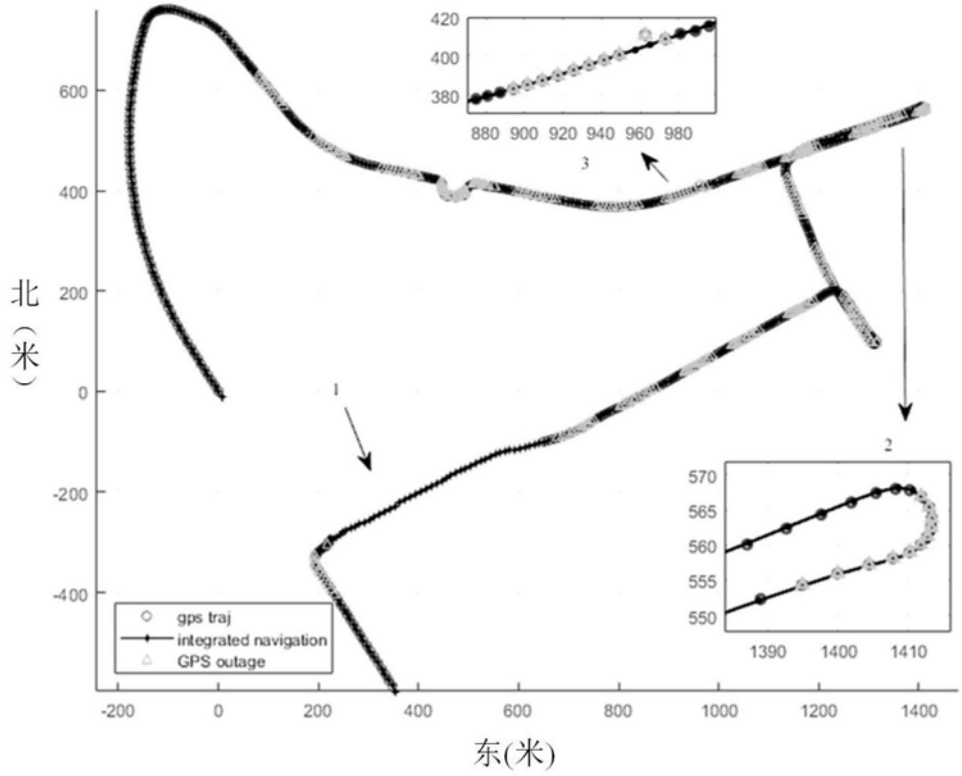
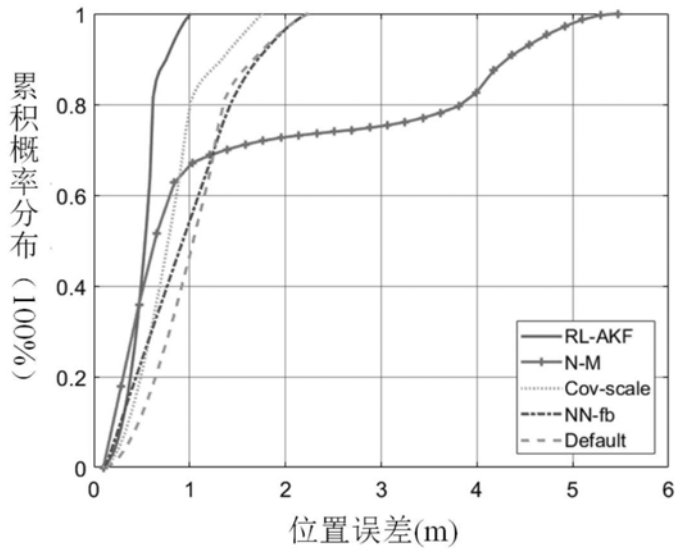
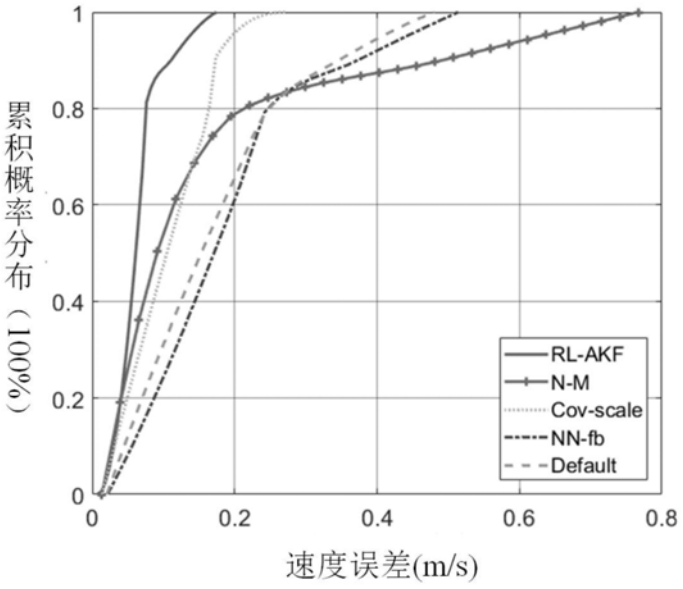


图7


累积概率分布（100%）
位置误差(m)
RL-AKF
N-M
Cov-scale
NN-fb
Default


累积概率分布（100%）
速度误差(m/s)
RL-AKF
N-M
Cov-scale
NN-fb
Default

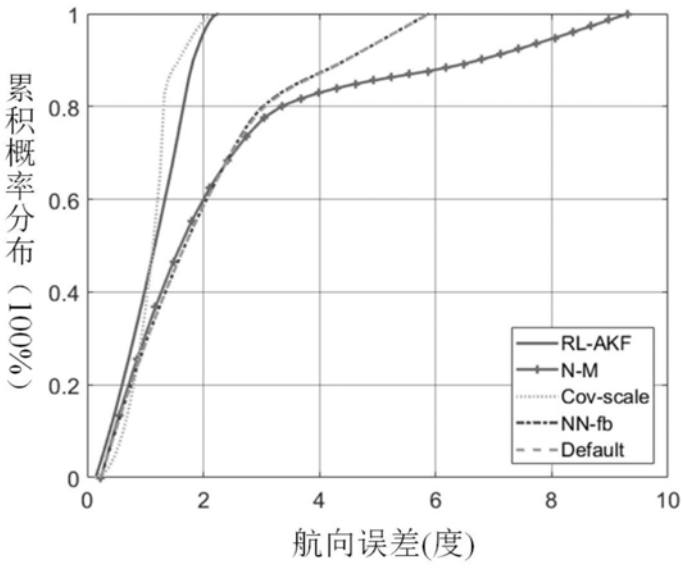
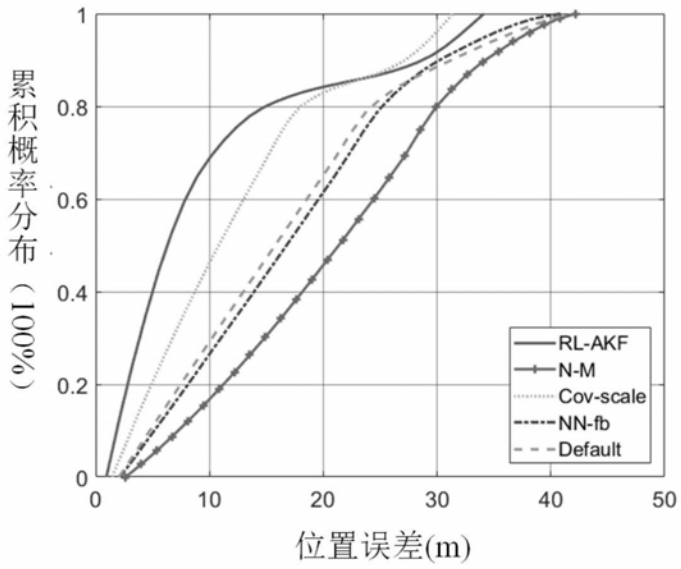
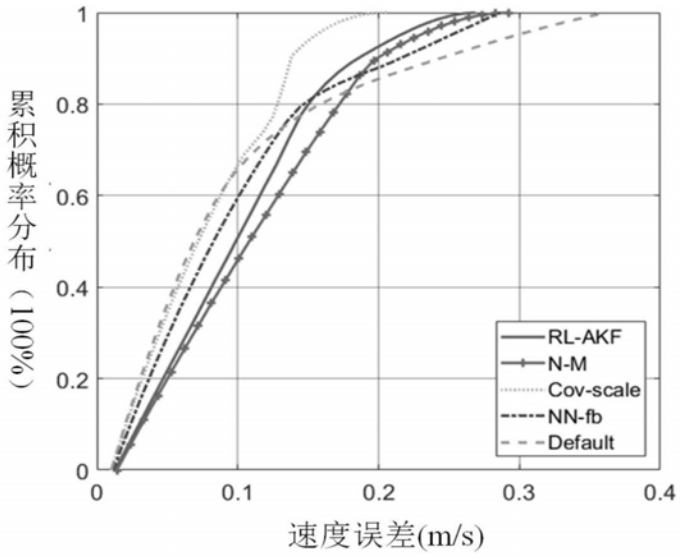

累积概率分布（100%）
航向误差(度)
RL-AKF
N-M
Cov-scale
NN-fb
Default

图8


累积概率分布（100%）
位置误差(m)
RL-AKF
N-M
Cov-scale
NN-fb
Default


累积概率分布（100%）
速度误差(m/s)
RL-AKF
N-M
Cov-scale
NN-fb
Default

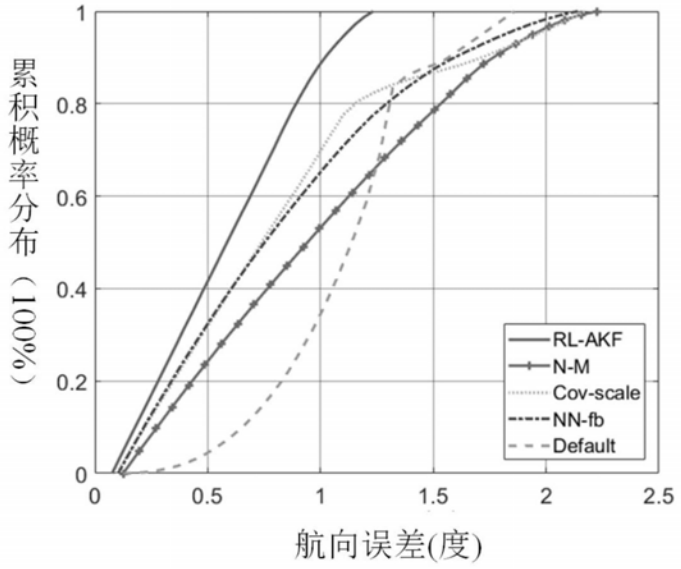
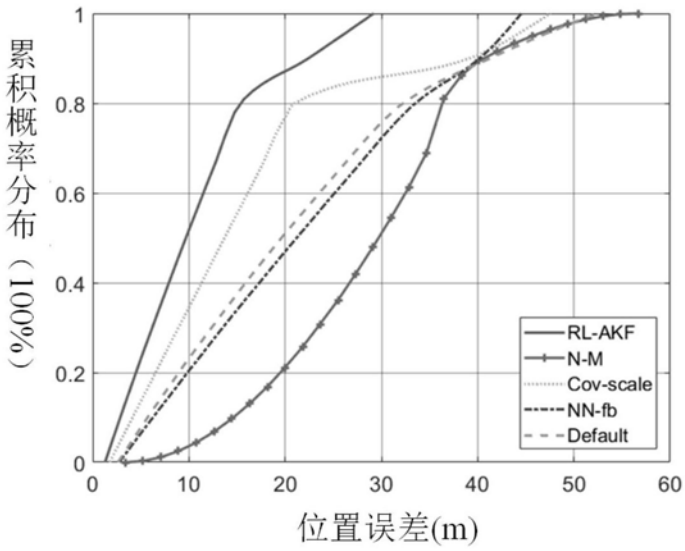
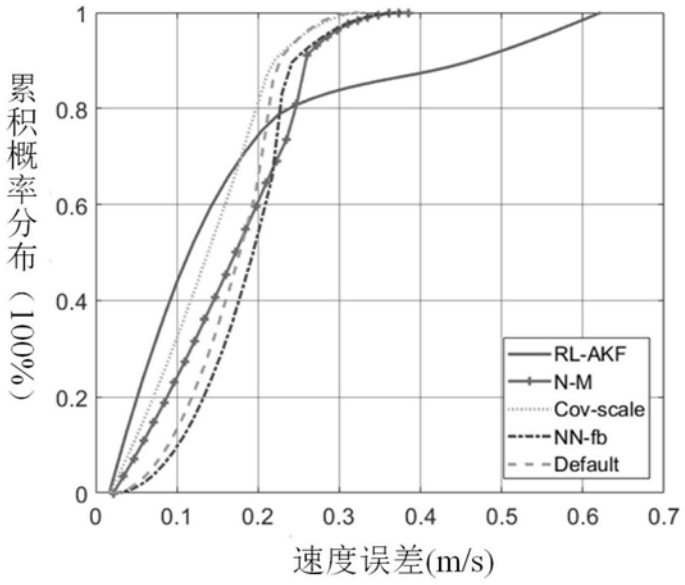

累积概率分布（100%）
航向误差(度)
RL-AKF
N-M
Cov-scale
NN-fb
Default

图9


累积概率分布（100%）
位置误差(m)
RL-AKF
N-M
Cov-scale
NN-fb
Default


累积概率分布（100%）
速度误差(m/s)
RL-AKF
N-M
Cov-scale
NN-fb
Default

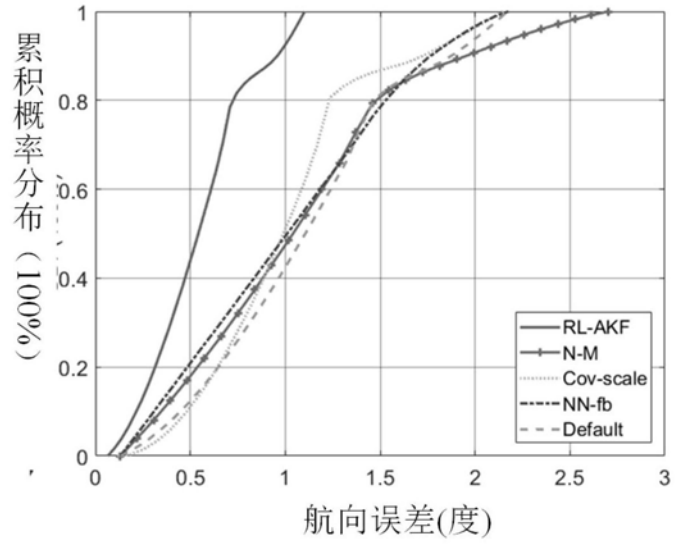

累积概率分布（100%）
航向误差(度)
RL-AKF
N-M
Cov-scale
NN-fb
Default

图10