

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2019 年 3 月 14 日 (14.03.2019)



(10) 国际公布号
WO 2019/047646 A1

(51) 国际专利分类号:
G05D 1/02 (2006.01)

(21) 国际申请号: PCT/CN2018/098637

(22) 国际申请日: 2018 年 8 月 3 日 (03.08.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710790602.5 2017 年 9 月 5 日 (05.09.2017) CN

(71) 申请人: 百度在线网络技术(北京)有限公司 (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。

(72) 发明人: 郑超 (ZHENG, Chao); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。 郁浩 (YU, Hao); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。 闫泳杉 (YAN, Yongshan); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。 唐坤 (TANG, Kun); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。 张云飞 (ZHANG, Yunfei); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。 姜雨 (JIANG, Yu); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。

(74) 代理人: 北京英赛嘉华知识产权代理有限公司 (INSIGHT INTELLECTUAL PROPERTY

(54) Title: OBSTACLE AVOIDANCE METHOD AND DEVICE FOR VEHICLE

(54) 发明名称: 车辆避障方法和装置

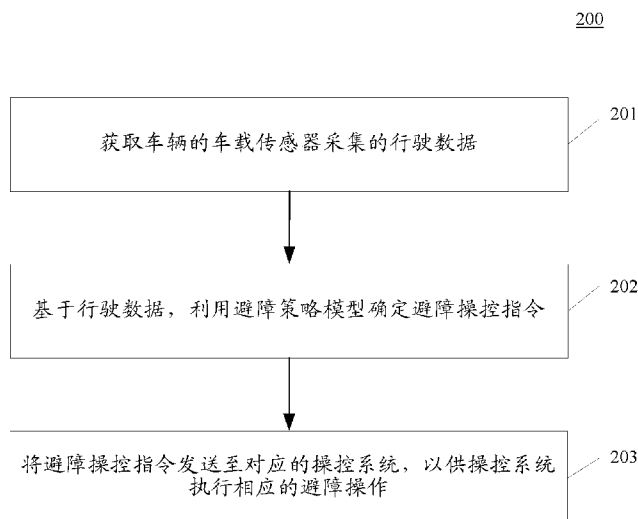


图 2

- 201 Acquire driving data collected by an onboard sensor of a vehicle
202 Determine an obstacle avoidance control instruction on the basis of the driving data and by means of an obstacle avoidance strategy model
203 Send the obstacle avoidance control instruction to a corresponding control system, such that the control system executes a corresponding obstacle avoidance operation

(57) Abstract: An obstacle avoidance method and a device for a vehicle, comprising: acquiring driving data collected by a onboard sensor (102) of a vehicle (101) (201, 302), wherein the driving data comprises obstacle information of a driving path and sensor data of the vehicle (101); determining an obstacle avoidance control instruction on the basis of the driving data by means of an obstacle avoidance strategy model (202, 303), wherein the obstacle avoidance strategy model is obtained in an end-to-end manner on the basis of a historical obstacle avoidance record by means of a deep reinforcement learning algorithm (301); and sending the obstacle avoid-

LIMITED); 中国北京市海淀区知春路甲48号盈都大厦A座19A, Beijing 100098 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

ance control instruction to a corresponding control system, such that the control system executes a corresponding obstacle avoidance operation (203, 304). Thus, the invention improves the success rate of obstacle avoidance by the vehicle (101).

(57) 摘要: 一种车辆避障方法和装置, 包括: 获取车辆(101)的车载传感器(102)采集的行驶数据(201,302), 其中行驶数据包括行驶路径的障碍物信息和车辆(101)的传感器数据; 基于行驶数据, 利用避障策略模型确定避障操控指令(202,303), 其中避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的(301); 将避障操控指令发送至对应的操控系统, 以供操控系统执行相应的避障操作(203,304)。由此可以提升车辆(101)避障的成功率。

车辆避障方法和装置

本专利申请要求于 2017 年 9 月 5 日提交的、申请号为 201710790602.5、申请人为百度在线网络技术（北京）有限公司、发明名称为“车辆避障方法和装置”的中国专利申请的优先权，该申请的全文以引用的方式并入本申请中。

技术领域

本申请涉及车辆控制技术领域，具体涉及车辆的行驶安全控制技术领域，尤其涉及车辆避障方法和装置。

背景技术

随着深度学习技术的迅速发展、人工智能领域的深入研究，给汽车工业带来革命性的变化。

自动驾驶，是在没有任何人类主动的操作下，利用电脑自动安全地操作机动车辆的技术。在车辆行驶过程中碰到障碍物时，现有的自动驾驶技术需要首先需要利用目标识别技术对障碍物进行识别，并基于预先定义的规则或基于模仿学习得出的深度学习系统确定避障策略。然而现有的自动驾驶技术的避障操作只能覆盖有限情况下的道路状况，并不具有推理能力，很难在实际道路中应对各种复杂的场景，避障成功率有待提升。

发明内容

为了解决上述背景技术部分提到的一个或多个技术问题，本申请实施例提供了车辆避障方法和装置。

第一方面，本申请实施例提供了一种车辆避障方法，包括：获取车辆的车载传感器采集的行驶数据，行驶数据包括行驶路径的障碍物信息和车辆的传感器数据；基于行驶数据，利用避障策略模型确定避

障操控指令，避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的；将避障操控指令发送至对应的操控系统，以供操控系统执行相应的避障操作。

在一些实施例中，上述方法还包括：基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出避障策略模型的步骤，其中，历史避障记录包括历史避障结果、与历史避障结果对应的历史行驶数据和历史操控数据；上述基于历史避障记录、采用深度增强学习算法训练得出避障策略模型的步骤包括：获取各历史避障结果的历史评估指数；基于历史行驶数据、历史评估指数和历史操控数据进行深度增强学习，以得出使避障结果最优的避障策略模型。

在一些实施例中，上述基于历史行驶数据、历史评估指数和历史操控数据进行深度增强学习，以得出使避障结果最优的避障策略模型的策略参数，包括：对每条历史避障记录，将历史行驶数据作为状态数据，历史操控数据作为操作数据，历史评估指数作为价值数据；利用历史避障记录构建学习轨迹，学习轨迹包括与各条历史避障记录一一对应的多个轨迹点，每个轨迹点包括对应的状态数据、操作数据以及价值数据；基于学习轨迹以及总价值数据计算总价值数据的期望关于策略参数的梯度，其中，策略参数用于将状态数据映射至操作数据；基于总价值数据的期望关于策略参数的梯度对策略参数进行调整，生成避障策略模型。

在一些实施例中，上述基于行驶数据，利用避障策略模型确定避障操控指令，包括：将车载传感器采集的行驶数据作为当前的状态数据输入避障策略模型；基于调整后的策略参数将当前的状态数据映射至当前的操作数据，作为避障操控指令。

在一些实施例中，在将避障操控指令发送至对应的操控系统之后，上述方法还包括：获取车辆的当前避障结果的评估指数，将当前避障结果的评价指数作为当前价值数据；基于当前的状态数据、当前的操作数据以及当前价值数据更新学习轨迹、总价值数据以及总价值数据的期望；基于更新后的学习轨迹以及各价值数据之和计算更新后的总价值数据的期望关于策略参数的梯度；基于更新后的总价值数据的期

望关于策略参数的梯度对策略参数进行调整。

第二方面，本申请实施例提供了一种车辆避障装置，包括：获取单元，配置用于获取车辆的车载传感器采集的行驶数据，行驶数据包括行驶路径的障碍物信息和车辆的传感器数据；确定单元，配置用于基于行驶数据，利用避障策略模型确定避障操控指令，避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的；发送单元，配置用于将避障操控指令发送至对应的操控系统，以供操控系统执行相应的避障操作。

在一些实施例中，上述装置还包括：训练单元，配置用于基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出避障策略模型，其中，历史避障记录包括历史避障结果、与历史避障结果对应的历史行驶数据和历史操控数据；训练单元配置用于按照如下方式训练得出避障策略模型：获取各历史避障结果的历史评估指数；基于历史行驶数据、历史评估指数和历史操控数据进行深度增强学习，以得出使避障结果最优的避障策略模型。

在一些实施例中，上述训练单元进一步配置用于按照如下方式进行深度增强学习，以得出使避障结果最优的避障策略模型的策略参数：对每条历史避障记录，将历史行驶数据作为状态数据，历史操控数据作为操作数据，历史评估指数作为价值数据；利用历史避障记录构建学习轨迹，学习轨迹包括与各条历史避障记录一一对应的多个轨迹点，每个轨迹点包括对应的状态数据、操作数据以及价值数据；基于学习轨迹以及总价值数据计算总价值数据的期望关于策略参数的梯度，其中，策略参数用于将状态数据映射至操作数据；基于总价值数据的期望关于策略参数的梯度对策略参数进行调整，生成避障策略模型。

在一些实施例中，上述确定单元进一步配置用于按照如下方式确定避障操控指令：将车载传感器采集的行驶数据作为当前的状态数据输入避障策略模型；基于调整后的策略参数将当前的状态数据映射至当前的操作数据，作为避障操控指令。

在一些实施例中，上述装置还包括更新单元，配置用于：在将避障操控指令发送至对应的操控系统之后，获取车辆的当前避障结果的

评估指数，将当前避障结果的评价指数作为当前价值数据；基于当前的状态数据、当前的操作数据以及当前价值数据更新学习轨迹、总价值数据以及总价值数据的期望；基于更新后的学习轨迹以及各价值数据之和计算更新后的总价值数据的期望关于策略参数的梯度；基于更新后的总价值数据的期望关于策略参数的梯度对策略参数进行调整。

第三方面，本申请实施例提供了一种设备，包括：一个或多个处理器；存储装置，用于存储一个或多个程序，当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现上述的车辆避障方法。

本申请实施例提供的车辆避障方法和装置，通过获取车辆的车载传感器采集的行驶数据，行驶数据包括行驶路径的障碍物信息和车辆的传感器数据；而后基于行驶数据，利用避障策略模型确定避障操控指令，避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的；之后将避障操控指令发送至对应的操控系统，以供操控系统执行相应的避障操作，能够实现车辆的自动避障，保证车辆在无人驾驶场景中的安全行驶；并且利用通过基于深度增强学习方法训练的避障模型，可以应对复杂的道路场景，能够提升避障成功率。

附图说明

通过阅读参照以下附图所作的对非限制性实施例详细描述，本申请的其它特征、目的和优点将会变得更明显：

图 1 是本申请可以应用于其中的示例性系统架构图；

图 2 是根据本申请的车辆避障方法的一个实施例的流程图；

图 3 是根据本申请的车辆避障方法的另一个实施例的流程图；

图 4 是根据本申请的采用深度增强学习算法训练得出所述避障策略模型的方法的一种具体实现方式的示意性流程图；

图 5 是本申请的车辆避障装置的一个实施例的结构示意图；

图 6 是适于用来实现本申请实施例的终端设备或服务器的计算机系统的结构示意图。

具体实施方式

下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

图 1 示出了可以应用本申请的车辆避障方法或车辆避障装置的实施例的示例性系统架构 100。

如图 1 所示，系统架构 100 可以包括车辆 101，安装在车辆 101 上的车载传感器 102 和车载控制单元 103、以及服务器 104。其中车辆 101 可以为无人驾驶车辆，车载控制单元 103 可以通过网络与服务器 104 连接，该网络可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

车载传感器 102 可以采集车辆行驶过程中的道路数据和车辆状态数据。车载传感器 102 可以包括车载摄像头、激光雷达传感器、毫米波雷达传感器、碰撞传感器、速度传感器、空气压力传感器等。车载控制单元 103 可以为 ECU (Electronic Control Unit, 电子控制单元)，或者可以为车载电脑，用于对车辆的各部件的工作状态进行分析和控制。车载控制单元 103 可以获取车载传感器 102 采集的数据，并对数据进行处理和响应，还可以控制车载传感器采集的数据的存储和传输。

服务器 104 可以通过网络与车载控制单元 103 建立连接，车载控制单元 103 可以将传感器数据发送至服务器 104。服务器 104 可以对传感器数据进行分析，将分析结果反馈至车载控制单元 103。车载控制单元 103 可以根据接收到的分析结果进行响应。

通常车辆 101 在行驶过程中会遇到障碍物，车载控制单元 103 可以通过车载传感器 102 获取障碍物的信息，例如利用车载摄像头采集障碍物的图像、利用红外探测器或超声波探测器探测障碍物的距离、

形状等信息，然后车载控制单元 103 可以根据障碍物信息进行分析处理，最终做出操控决策，或者车载控制单元 103 可以将障碍物信息传输至服务器 104 进行分析处理并执行操控决策，由服务器 104 将操控决策传回车载控制单元 103。车载控制单元 103 根据操控决策对车辆 101 的对应部件进行控制。

需要说明的是，本申请实施例所提供的车辆避障方法可以由车载控制单元 103 或服务器 104 执行，相应地，车辆避障装置可以设置于车载控制单元 103 或服务器 104 中。

应该理解，图 1 中的车辆、车载传感器、车载控制单元和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的车辆、车载传感器、车载控制单元和服务器。

继续参考图 2，示出了根据本申请的车辆避障方法的一个实施例的流程 200。该车辆避障方法，包括以下步骤：

步骤 201，获取车辆的车载传感器采集的行驶数据。

在本实施例中，上述车辆避障方法运行于其上的电子设备可以通过与车载传感器的有线或无线连接获取车载传感器采集的行驶数据，也可以通过与用于存储车载传感器数据的存储设备的连接、发出数据获取请求来获取车载传感器采集的行驶数据。其中，行驶数据包括行驶路径的障碍物信息和车辆的传感器数据。障碍物信息可以为道路上的行人、车辆、路障标志物等的相关信息，车辆传感器数据可以包括当前车辆的各个传感器的数据，可以包括但不限于压力传感器数据、图像传感器数据、激光雷达传感器数据、ABS（antilock brake system，制动防抱死系统）传感器数据、速度传感器数据、位置和转速传感器数据等。这里的车辆传感器数据可以包括传感器采集的车辆的状态数据以及车辆所处道路上的可被车辆传感器感知的数据。

本实施例中的车辆可以为无人驾驶车辆，或者说是具有自动驾驶功能的车辆。通常无人驾驶车辆在行驶过程中，车辆传感器可以持续地采集相应的数据，上述电子设备可以直接与各车辆传感器连接，实时地获取车辆传感器采集的数据，也可以通过与车载控制单元获取车辆传感器传输至车载控制单元的传感器数据。

步骤 202, 基于行驶数据, 利用避障策略模型确定避障操控指令。

在本实施例中, 避障策略模型用于将输入的行驶数据映射至避障操控指令。避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的。其中, 历史避障记录可以为车辆在行驶过程中已发生的多次避障过程的相关记录。

在本实施例中, 可以将上述行驶数据输入已训练的避障策略模型, 利用避障策略模型来确定相应的避障操控指令。避障策略模型可以根据障碍物信息和车辆的传感器数据确定需要调用的操控部件及待发送至待调用的操控部件的操控指令。

上述避障策略模型可以基于 DRL (Deep Reinforcement Learning, 深度增强学习) 技术以端到端的方式训练得出。

端到端的方式是指由直接获取的图像、音频等输入数据映射至期望的输出结果的方式, 在本申请实施例的场景中, 端到端是指由输入的行驶数据映射至操控指令的方式。区别于传统的操控决策方法, 端到端的方式将数据清洗、特征提取、编解码等过程融合在一起, 由深度神经网络模型建立输入数据到最终的目标输出数据之间的映射关系。以端到端的方式训练得出避障策略模型, 就是将历史避障记录中的历史行驶数据作为模型的输入, 对应的避障操控指令作为模型的输出而不断学习避障操控策略的过程。

在本实施例中, 上述避障策略模型是以端到端的方式训练得出的, 在训练过程中, 只需提供原始的车载传感器采集的行驶数据和避障操控得到的避障结果的标记值即可, 避障策略模型可以通过深度增强学习来优化模型参数。

DRL 是将深度学习与强化学习的算法相结合, 使训练得出的模型同时具有感知能力和决策能力。具体地, 深度增强学习算法可以包括基于价值函数的深度增强学习、基于策略梯度的深度增强学习以及基于搜索与监督的深度增强学习。

举例来说, 基于价值函数的深度学习将卷积神经网络与强化学习中的 Q 学习算法结合, 利用 DQN (Deep Q Network, 深度 Q 网络) 模型, 输入的数据经过多层神经网络和连接层的非线性变换之后生成

每个动作的 Q 值。在训练 DQN 模型时, 首先可以构建一个深度神经网络作为 Q 值的网络, 其中网络参数为 w , 然后计算 Q 值的误差函数, 之后计算参数 w 关于误差函数的梯度, 之后采用随机梯度下降方法来更新网络参数 w 。

步骤 203, 将避障操控指令发送至对应的操控系统, 以供操控系统执行相应的避障操作。

在利用避障策略模型制定避障策略、得到避障操控指令之后, 可以将避障操控指令发送至对应的操控系统。在这里, 操控系统可以包括制动系统、EPS (Electric Power Steering, 电子助力转向) 系统、ESP (Electronic Stability Program, 电子稳定系统) 等。操控系统接收到避障操控指令之后执行相应减速、转向等操作, 使车辆避开障碍物。

具体地, 避障操控指令可以包括制动指令、转向指令、报警指令等, 可以将制动指令发送至制动系统, 转向指令发送至 EPS 系统、报警指令发送至报警灯和报警喇叭。可选地, 车辆的各操控系统与 CAN (Controller Area Network, 控制器局域网) 总线相连, 这些避障操作指令可以通过 CAN 总线传输至对应的操控系统。

本申请上述实施例的车辆避障方法, 通过获取车辆的车载传感器采集的行驶数据, 行驶数据包括行驶路径的障碍物信息和车辆的传感器数据; 而后基于行驶数据, 利用避障策略模型确定避障操控指令, 避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的; 之后将避障操控指令发送至对应的操控系统, 以供操控系统执行相应的避障操作, 能够实现车辆的自动避障, 保证车辆在无人驾驶场景中的安全行驶; 并且利用通过基于深度增强学习方法训练的避障模型, 可以应对复杂的道路场景, 能够提升避障成功率。

请参考图 3, 其示出了根据本申请的车辆避障方法的另一个实施例的流程图。如图 3 所示, 本实施例的车辆避障方法的流程 300, 可以包括以下步骤:

步骤 301, 基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出避障策略模型。

在本实施例中, 历史避障记录可以包括历史避障结果、与历史避

障结果对应的历史行驶数据和历史操控数据。其中，历史避障结果可以为车辆的历史行程中避障操作的结果，可以表征车辆在避障操作之后的状态，例如可以包括“完美避开”、“勉强避开”、“避障失败”等。历史行驶数据可以包括历史行驶路径的障碍物信息和车辆的历史传感器数据。历史操控数据可以包括在车辆避障时上述电子设备（如车载控制单元）发出的历史避障操控指令。

在本实施例中，避障策略模型用于将输入的行驶数据映射至避障操控指令。上述电子设备可以具体按照如下方式来训练得出避障策略模型：获取历史避障结果的历史评估指数，基于历史行驶数据、历史评估指数和历史操控数据进行深度增强学习，以得出使避障结果最优的避障策略模型。

具体地，可以对每一条历史避障记录的历史避障结果进行评估，得出评估指数，例如可以由专业的驾驶人员为历史避障结果打分，避障结果越好，分数越高，将打分作为该历史避障结果对应的历史评估指数。然后将历史行驶数据、历史操控数据和历史评估指数输入模型进行深度增强学习。上述端到端的方式是指在学习过程中，将历史行驶数据作为模型的输入、历史操控数据作为模型的输出，由车辆直接采集的数据到操控指令的“端到端”学习方式，在学习过程中将历史评估指数作为学习后的模型的效果评价指标，则可以在学习过程中不断调整模型参数使避障结果逼近评估指数较高的结果。这样，通过将避障结果量化为评估指数来进行学习，可以使模型更容易地学习出避障结果较好的避障策略的逻辑，加快学习速度，提升避障策略模型的效果。

在本实施例的一些可选的实现方式中，步骤 301 可以按照如下方式执行：从历史避障记录中抽样以构建样本数据集，样本数据集包括多条样本数据，每条样本数据与一条历史避障记录对应，样本数据包括对应的历史行驶数据、历史评估指数和历史操控数据；基于样本数据集，对已构建的深度增强学习模型进行训练，得到避障策略模型。

具体地，可以例如采用从历史避障记录中随机抽样的方式构建出样本数据集，然后构建 DQN 模型，并利用样本数据集训练 DQN 模型，

根据每个样本数据对应的评估指数调整 DQN 模型中每个卷积层和连接层的参数，得出使评估指数最高的参数，即可生成避障策略模型。

步骤 302，获取车辆的车载传感器采集的行驶数据。

在本实施例中，上述车辆避障方法运行于其上的电子设备可以通过与车载传感器的有线或无线连接获取车载传感器采集的行驶数据，也可以通过与用于存储车载传感器数据的存储设备的连接、发出数据获取请求来获取车载传感器采集的行驶数据。其中，行驶数据包括行驶路径的障碍物信息和车辆的传感器数据。障碍物信息可以为道路上的行人、车辆、路障标志物等的相关信息，车辆传感器数据可以包括当前车辆的各个传感器的数据，包括压力传感器数据、图像传感器数据、激光雷达传感器数据、ABS (antilock brake system, 制动防抱死系统) 传感器数据、速度传感器数据、位置和转速传感器数据等。这里的车辆传感器数据可以包括传感器采集的车辆的状态数据以及车辆所处道路上的可被车辆传感器感知的数据。

步骤 303，基于行驶数据，利用避障策略模型确定避障操控指令。

在本实施例中，可以采用上述步骤 301 训练得出的避障策略模型来处理获取的行驶数据，得出避障操控指令。避障操控指令可以为用于控制车辆的操控单元进行避障操作的指令。

步骤 304，将避障操控指令发送至对应的操控系统，以供操控系统执行相应的避障操作。

在利用避障策略模型制定避障策略、得到避障操控指令之后，可以将避障操控指令发送至对应的操控系统。避障操控指令可以包括制动指令、转向指令、报警指令等，操控系统可以包括制动系统、EPS (Electric Power Steering, 电子助力转向) 系统、ESP (Electronic Stability Program, 电子稳定系统) 等。操控系统接收到避障操控指令之后执行相应减速、转向等操作，使车辆避开障碍物。

上述方法流程中的步骤 302、步骤 303、步骤 304 分别与前述实施例中的步骤 201、步骤 202、步骤 203 相同，上文针对步骤 201、步骤 202、步骤 203 的描述也适用于本实施例中的步骤 302、步骤 303、步骤 304，此处不再赘述。

在图 2 所示实施例的基础上，图 3 所示的实施例增加了基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出所述避障策略模型的步骤，本实施例的车辆避障方法可以将历史避障结果量化为评估指数，使得作为训练实体的上述电子设备可以将该评估指数为模型的效果指标来不断学习内在的避障逻辑，从而实现了避障效果的进一步优化。

在本实施例的一些可选的实现方式中，上述基于历史行驶数据、历史评估指数和历史操控数据进行深度增强学习，以得出使避障结果最优的避障策略模型的步骤可以包括：对每条历史避障记录，将历史行驶数据作为状态数据，历史操控数据作为操作数据，历史评估指数作为价值数据；利用历史避障记录构建学习轨迹，学习轨迹包括与各条历史避障记录一一对应的多个轨迹点，每个轨迹点包括对应的状态数据、操作数据以及价值数据；基于学习轨迹以及各价值数据之和计算总价值数据的期望关于策略参数的梯度，其中，策略参数用于将状态数据映射至操作数据；基于总价值数据的期望关于策略参数的梯度对策略参数进行调整，生成避障策略模型。

具体地，请参考图 4，其示出了根据本申请的采用深度增强学习算法训练得出所述避障策略模型的方法的一种具体实现方式的示意性流程图。

如图 4 所示，在步骤 401 中，基于历史避障记录构建深度增强学习算法的学习轨迹。

学习轨迹包括多个轨迹点，每个轨迹点对应一条历史避障记录。每个轨迹点包括对应的历史避障记录的状态数据、操作数据以及价值数据。在本实施例中，第 i 条历史避障记录的状态数据 s_i 为历史行驶数据，操作数据 a_i 为历史操控数据，价值数据 r_i 为历史评估指数， i 为整数且 $0 \leq i \leq T-1$ ， T 为历史避障记录的条数。则构建的学习轨迹 $\tau = (s_0, a_0, r_0, s_1, a_1, r_1, \dots, s_{T-1}, a_{T-1}, r_{T-1})$ 。

随后，在步骤 402 中，基于学习轨迹以及总价值数据计算总价值数据的期望关于策略参数的梯度。

具体地，假设策略参数为 θ ，可以根据各价值数据计算得出总价

值数据 R ，例如：

$$R = \sum_{t=0}^{T-1} r_t \quad (1)$$

总价值数的期望为 $E[R|\pi_\theta]$ ，其中 π_θ 表示由状态数据 s_i 映射至操作数据 a_i 的策略，总价值数据的期望关于策略参数的梯度可以表示为：

$$g = R \nabla_{\theta} \sum_{t=0}^{T-1} \log \pi(a_t | s_t; \theta) \quad (2)$$

其中， $\pi(a_t | s_t; \theta)$ 也可表示为 $\pi_\theta(a_t | s_t)$ ，表示由状态数据 s_t 映射至操作数据 a_t 的策略。

之后，在步骤 403 中，基于总价值数据的期望关于策略参数的梯度对策略参数进行调整。具体的调整策略为将 θ 调整为 $\theta + \alpha g$ ，其中 α 为控制策略参数更新速率的参数。

最终可以得出使得总价值数据的期望最大的策略参数，即为上述避障策略模型。

在进一步的实施例，上述基于行驶数据，利用避障策略模型确定避障操作指令可以包括：将车载传感器采集的行驶数据作为当前的状态数据输入避障策略模型；基于调整后的策略参数将当前的状态数据映射至当前的操作数据，作为避障操控指令。具体来说，在根据上述方法得出避障策略模型之后，可以将车载传感器采集的行驶数据作为当前时刻 T 的状态数据 s_T ，利用该模型中的策略参数 θ 将状态数据 s_T 映射至当前的操作数据 a_T ，该操作数据 a_T 即为避障操作指令。

进一步地，如图 4 所示，在将避障操控指令发送至对应的操控系统，以供操控系统执行相应的避障操作之后，上述车辆的避障方法还包括：

步骤 404，获取车辆的当前避障结果的评估指数，然后将当前避障结果的评估指数作为当前价值数据 r_T ，得到包括当前状态数据 s_T 、当前的操作数据 a_T 、以及当前价值数据 r_T 的轨迹点，在步骤 405 中利用该轨迹点更新上述学习轨迹 τ 、进而更新总价值数据 R 以及总价值数据的期望 $E[R|\pi_\theta]$ 。

然后可以根据上式 (2) 计算更新后的总价值数据的期望 $E[R|\pi_\theta]$ 对策略参数 θ 的梯度，并根据学习率 α 对策略参数 θ 进行调整，得到对当前避障操作学习后的避障策略模型。

从图 4 可以看出，本实施例的避障策略模型训练方法利用增强学习算法构建了基于状态数据、操作数据以及价值数据的学习轨迹，然后依据学习轨迹对每个轨迹点进行学习，得出最优的避障策略参数，从而生成避障策略模型。并且，在每一次避障操作后，可以依据避障结果来更新避障策略模型，实现了避障策略模型的进一步优化。

进一步参考图 5，作为对上述各图所示方法的实现，本申请提供了一种车辆避障装置的一个实施例，该装置实施例与图 2 所示的方法实施例相对应，该装置具体可以应用于各种电子设备中。

如图 5 所示，本实施例的车辆避障装置 500 包括：获取单元 501 确定单元 502 和发送单元 503。其中，获取单元 501 可以配置用于获取车辆的车载传感器采集的行驶数据，行驶数据包括行驶路径的障碍物信息和车辆的传感器数据；确定单元 502 可以配置用于基于行驶数据，利用避障策略模型确定避障操控指令，避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的；发送单元 503 可以配置用于将避障操控指令发送至对应的操控系统，以供操控系统执行相应的避障操作。

在本实施例中，获取单元 501 可以通过与车载传感器的有线或无线连接获取车载传感器采集的行驶数据，也可以通过与用于存储车载传感器数据的存储设备的连接来获取车载传感器采集的行驶数据。其中，行驶数据包括行驶路径的障碍物信息和车辆的传感器数据。

确定单元 502 可以利用基于 DRL 技术以端到端方式训练得出的避障策略模型确定避障操控指令。避障策略模型可以将行驶数据输入谁知对应的避障操控指令，可以利用历史行驶数据和历史操控数据训练得出。确定单元 502 可以将获取单元 501 获取的行驶数据输入避障策略模型，则可以得出相应的避障操控指令。

发送单元 503 可以将确定单元 502 确定出的避障操控指令发送至对应的操控系统，操控系统可以包括制动系统、EPS（Electric Power Steering，电子助力转向）系统、ESP（Electronic Stability Program，电子稳定系统）等。操控系统接收到避障操控指令之后执行相应减速、转向等操作，使车辆避开障碍物。

在一些实施例中，上述装置 500 还可以包括：训练单元，配置用于基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出避障策略模型，其中，历史避障记录包括历史避障结果、与历史避障结果对应的历史行驶数据和历史操控数据。训练单元配置用于按照如下方式训练得出避障策略模型：获取各历史避障结果的历史评估指数；基于历史行驶数据、历史评估指数和历史操控数据进行深度增强学习，以得出使避障结果最优的避障策略模型。

在进一步的实施例中，上述训练单元进一步配置用于按照如下方式进行深度增强学习，以得出使避障结果最优的避障策略模型的策略参数：对每条历史避障记录，将历史行驶数据作为状态数据，历史操控数据作为操作数据，历史评估指数作为价值数据；利用历史避障记录构建学习轨迹，学习轨迹包括与各条历史避障记录一一对应的多个轨迹点，每个轨迹点包括对应的状态数据、操作数据以及价值数据；基于学习轨迹以及总价值数据计算总价值数据的期望关于策略参数的梯度，其中，策略参数用于将状态数据映射至操作数据；基于总价值数据的期望关于策略参数的梯度对策略参数进行调整，生成避障策略模型。

在进一步的实施例中，上述确定单元进一步配置用于按照如下方式确定避障操控指令：将车载传感器采集的行驶数据作为当前的状态数据输入避障策略模型；基于调整后的策略参数将当前的状态数据映射至当前的操作数据，作为避障操控指令。

在进一步的实施例中，上述装置还包括更新单元，配置用于：在将避障操控指令发送至对应的操控系统之后，获取车辆的当前避障结果的评估指数，将当前避障结果的评价指数作为当前价值数据；基于当前的状态数据、当前的操作数据以及当前价值数据更新学习轨迹、总价值数据以及总价值数据的期望；基于更新后的学习轨迹以及各价值数据之和计算更新后的总价值数据的期望关于策略参数的梯度；基于更新后的总价值数据的期望关于策略参数的梯度对策略参数进行调整。

本申请上述实施例的车辆避障装置 500，通过获取单元获取车辆

的车载传感器采集的行驶数据，随后利用确定单元基于行驶数据，利用基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的避障策略模型确定避障操控指令，之后利用发送单元将避障操控指令发送至对应的操控系统，以供操控系统执行相应的避障操作，实现了车辆的自动避障，并且可以应对复杂的道路场景，能够提升避障成功率，提升车辆的安全性。

应当理解，装置 500 中记载的诸单元可以与参考图 2、图 3 和图 4 描述的方法中的各个步骤相对应。由此，上文针对方法描述的操作和特征同样适用于装置 500 及其中包含的单元，在此不再赘述。

下面参考图 6，其示出了适于用来实现本申请实施例的终端设备或服务器的计算机系统 600 的结构示意图。图 6 示出的终端设备或服务器仅仅是一个示例，不对本申请实施例的功能和使用范围带来任何限制。

如图 6 所示，计算机系统 600 包括中央处理单元 (CPU) 601，其可以根据存储在只读存储器 (ROM) 602 中的程序或者从存储部分 608 加载到随机访问存储器 (RAM) 603 中的程序而执行各种适当的动作和处理。在 RAM 603 中，还存储有系统 600 操作所需的各种程序和数据。CPU 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出 (I/O) 接口 605 也连接至总线 604。

以下部件连接至 I/O 接口 605：包括键盘、鼠标等的输入部分 606；包括诸如阴极射线管 (CRT)、液晶显示器 (LCD) 等以及扬声器等的输出部分 607；包括硬盘等的存储部分 608；以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 609。通信部分 609 经由诸如因特网的网络执行通信处理。驱动器 610 也根据需要连接至 I/O 接口 605。可拆卸介质 611，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器 610 上，以便于从其上读出的计算机程序根据需要被安装入存储部分 608。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程

序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信部分 609 从网络上被下载和安装，和/或从可拆卸介质 611 被安装。在该计算机程序被中央处理单元 (CPU) 601 执行时，执行本申请的方法中限定的上述功能。需要说明的是，本申请所述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器 (RAM)、只读存储器 (ROM)、可擦式可编程只读存储器 (EPROM 或闪存)、光纤、便携式紧凑磁盘只读存储器 (CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中，计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：无线、电线、光缆、RF 等等，或者上述的任意合适的组合。

附图中的流程图和框图，图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时

也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本申请实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括获取单元、确定单元和发送单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，获取单元还可以被描述为“获取车辆的车载传感器采集的行驶数据的单元”。

作为另一方面，本申请还提供了一种计算机可读介质，该计算机可读介质可以是上述实施例中描述的装置中所包含的；也可以是单独存在，而未装配入该装置中。上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该装置执行时，使得该装置：获取车辆的车载传感器采集的行驶数据，所述行驶数据包括行驶路径的障碍物信息和所述车辆的传感器数据；基于所述行驶数据，利用避障策略模型确定避障操控指令，所述避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的；将所述避障操控指令发送至对应的操控系统，以供所述操控系统执行相应的避障操作。

以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。

权 利 要 求 书

1、一种车辆避障方法，其特征在于，所述方法包括：

获取车辆的车载传感器采集的行驶数据，所述行驶数据包括行驶路径的障碍物信息和所述车辆的传感器数据；

基于所述行驶数据，利用避障策略模型确定避障操控指令，所述避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的；

将所述避障操控指令发送至对应的操控系统，以供所述操控系统执行相应的避障操作。

2、根据权利要求1所述的方法，其特征在于，所述方法还包括：

基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出所述避障策略模型的步骤，其中，所述历史避障记录包括历史避障结果、与所述历史避障结果对应的历史行驶数据和历史操控数据；

所述基于历史避障记录、采用深度增强学习算法训练得出所述避障策略模型的步骤包括：

获取各所述历史避障结果的历史评估指数；

基于所述历史行驶数据、所述历史评估指数和所述历史操控数据进行深度增强学习，以得出使避障结果最优的避障策略模型。

3、根据权利要求2所述的方法，其特征在于，所述基于所述历史行驶数据、所述历史评估指数和所述历史操控数据进行深度增强学习，以得出使避障结果最优的所述避障策略模型的策略参数，包括：

对每条所述历史避障记录，将所述历史行驶数据作为状态数据，所述历史操控数据作为操作数据，所述历史评估指数作为价值数据；

利用所述历史避障记录构建学习轨迹，所述学习轨迹包括与各条所述历史避障记录一一对应的多个轨迹点，每个轨迹点包括对应的状态数据、操作数据以及价值数据；

基于所述学习轨迹以及总价值数据计算总价值数据的期望关于策

略参数的梯度，其中，所述策略参数用于将所述状态数据映射至所述操作数据；

基于所述总价值数据的期望关于策略参数的梯度对所述策略参数进行调整，生成所述避障策略模型。

4、根据权利要求3所述的方法，其特征在于，所述基于所述行驶数据，利用避障策略模型确定避障操控指令，包括：

将所述车载传感器采集的行驶数据作为当前的状态数据输入所述避障策略模型；

基于所述调整后的策略参数将所述当前的状态数据映射至当前的操作数据，作为所述避障操控指令。

5、根据权利要求4所述的方法，其特征在于，在将所述避障操控指令发送至对应的操控系统之后，所述方法还包括：

获取所述车辆的当前避障结果的评估指数，将所述当前避障结果的评价指数作为当前价值数据；

基于所述当前的状态数据、所述当前的操作数据以及所述当前价值数据更新所述学习轨迹、所述总价值数据以及所述总价值数据的期望；

基于更新后的学习轨迹以及各价值数据之和计算更新后的总价值数据的期望关于策略参数的梯度；

基于所述更新后的总价值数据的期望关于策略参数的梯度对所述策略参数进行调整。

6、一种车辆避障装置，其特征在于，所述装置包括：

获取单元，配置用于获取车辆的车载传感器采集的行驶数据，所述行驶数据包括行驶路径的障碍物信息和所述车辆的传感器数据；

确定单元，配置用于基于所述行驶数据，利用避障策略模型确定避障操控指令，所述避障策略模型是基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出的；

发送单元,配置用于将所述避障操控指令发送至对应的操控系统,以供所述操控系统执行相应的避障操作。

7、根据权利要求6所述的装置,其特征在于,所述装置还包括:训练单元,配置用于基于历史避障记录、采用深度增强学习算法以端到端的方式训练得出所述避障策略模型,其中,所述历史避障记录包括历史避障结果、与所述历史避障结果对应的历史行驶数据和历史操控数据;

所述训练单元配置用于按照如下方式训练得出所述避障策略模型:

获取各所述历史避障结果的历史评估指数;

基于所述历史行驶数据、所述历史评估指数和所述历史操控数据进行深度增强学习,以得出使避障结果最优的避障策略模型。

8、根据权利要求7所述的装置,其特征在于,所述训练单元进一步配置用于按照如下方式进行深度增强学习,以得出使避障结果最优的所述避障策略模型的策略参数:

对每条所述历史避障记录,将所述历史行驶数据作为状态数据,所述历史操控数据作为操作数据,所述历史评估指数作为价值数据;

利用所述历史避障记录构建学习轨迹,所述学习轨迹包括与各条所述历史避障记录一一对应的多个轨迹点,每个轨迹点包括对应的状态数据、操作数据以及价值数据;

基于所述学习轨迹以及总价值数据计算总价值数据的期望关于策略参数的梯度,其中,所述策略参数用于将所述状态数据映射至所述操作数据;

基于所述总价值数据的期望关于策略参数的梯度对所述策略参数进行调整,生成所述避障策略模型。

9、根据权利要求8所述的装置,其特征在于,所述确定单元进一步配置用于按照如下方式确定避障操控指令:

将所述车载传感器采集的行驶数据作为当前的状态数据输入所述避障策略模型；

基于所述调整后的策略参数将所述当前的状态数据映射至当前的操作数据，作为所述避障操控指令。

10、根据权利要求 9 所述的装置，其特征在于，所述装置还包括更新单元，配置用于：

在将所述避障操控指令发送至对应的操控系统之后，获取所述车辆的当前避障结果的评估指数，将所述当前避障结果的评价指数作为当前价值数据；

基于所述当前的状态数据、所述当前的操作数据以及所述当前价值数据更新所述学习轨迹、所述总价值数据以及所述总价值数据的期望；

基于更新后的学习轨迹以及各价值数据之和计算更新后的总价值数据的期望关于策略参数的梯度；

基于所述更新后的总价值数据的期望关于策略参数的梯度对所述策略参数进行调整。

11、一种设备，其特征在于，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-5 中任一所述的方法。

12、一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求 1-5 中任一所述的方法。

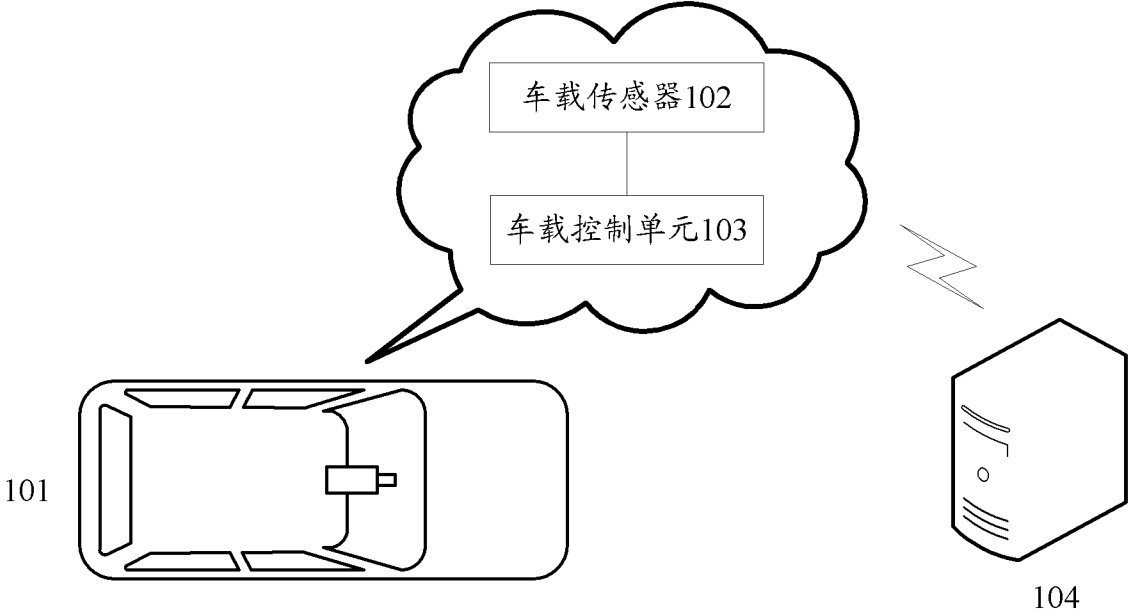


图 1

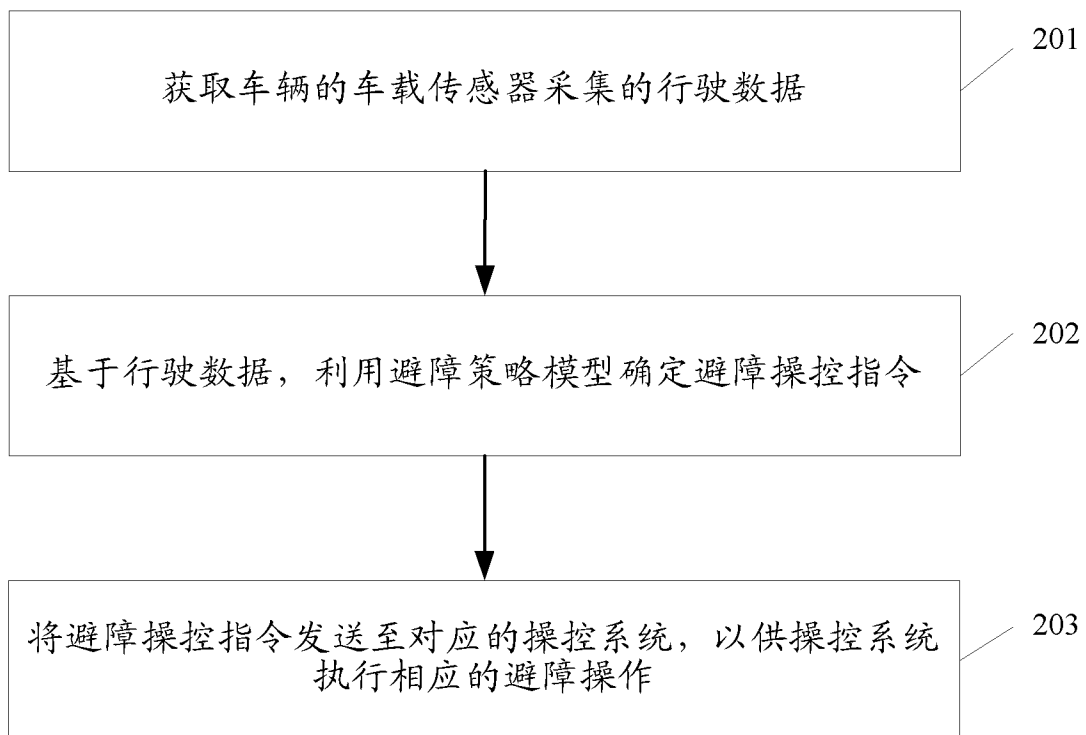
200

图 2

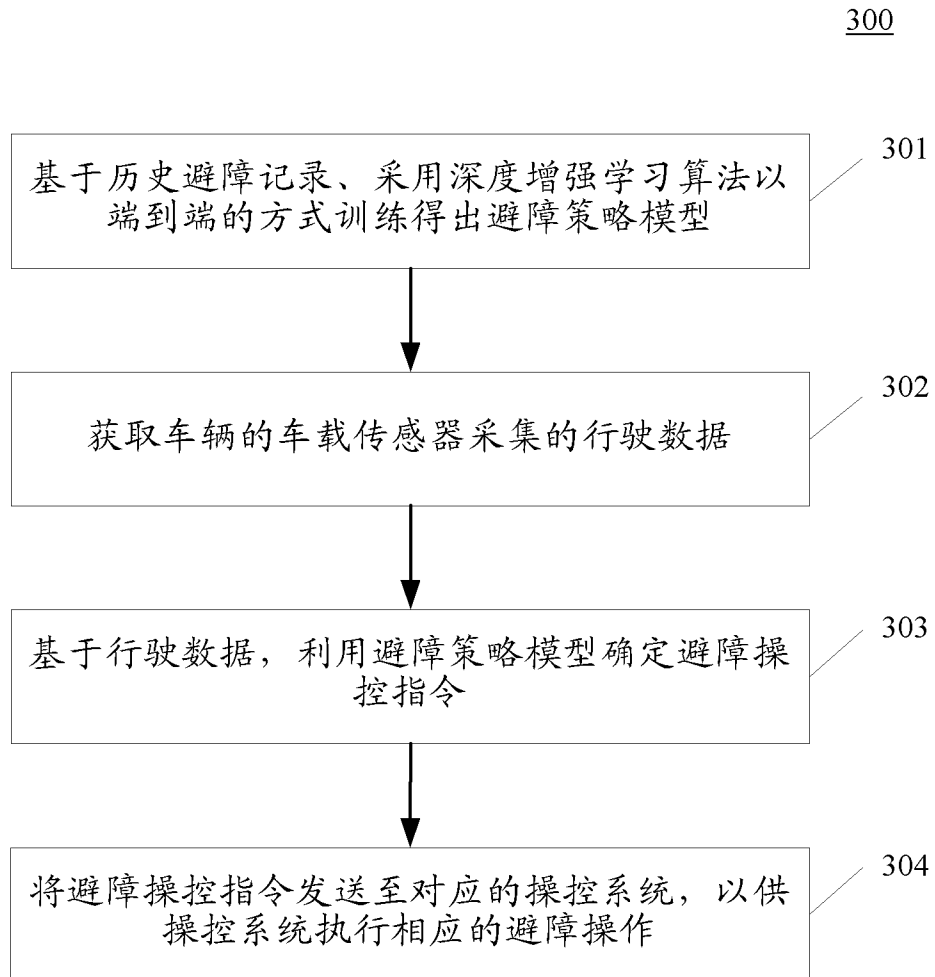


图 3

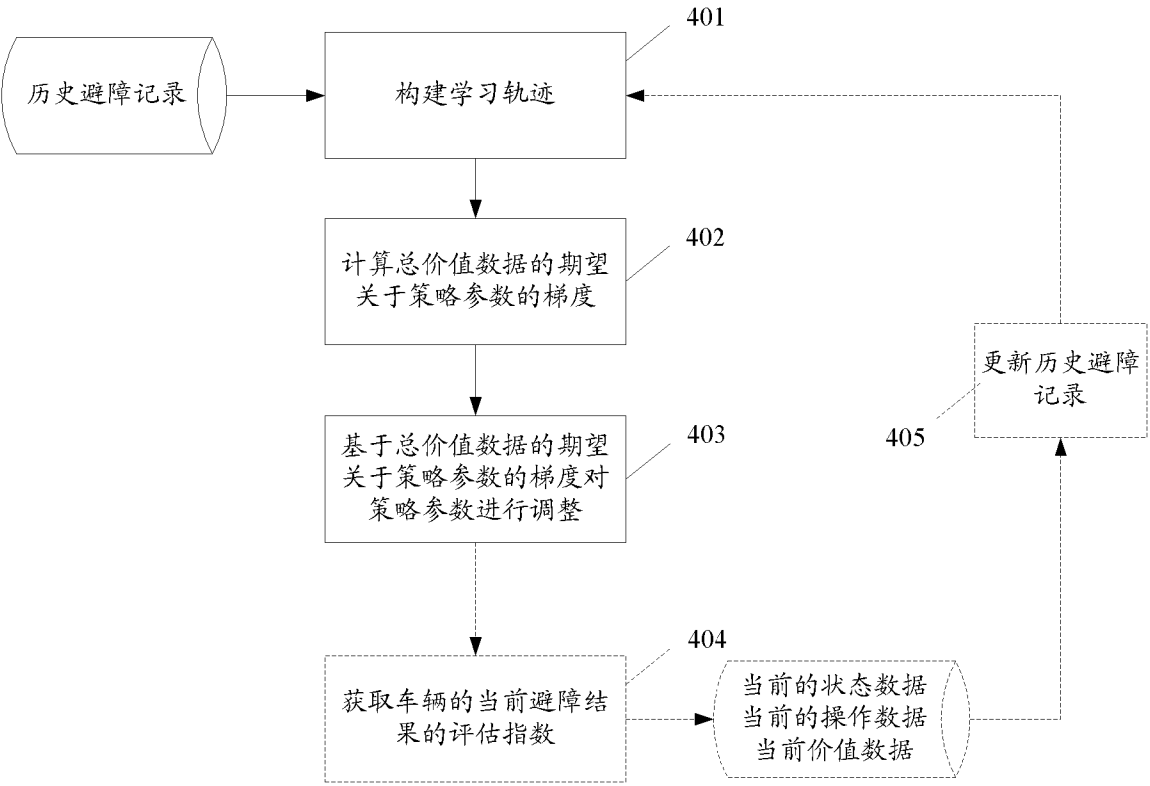


图 4

500

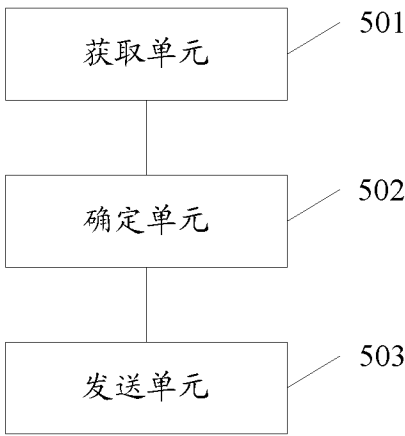


图 5

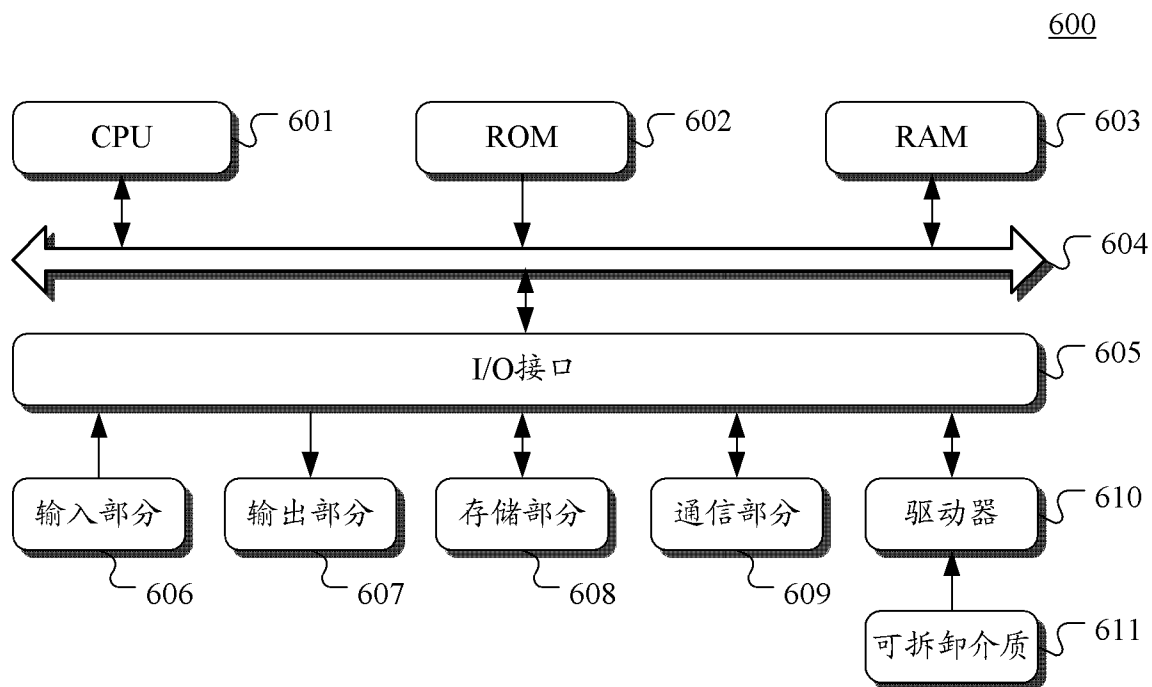


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/098637

A. CLASSIFICATION OF SUBJECT MATTER

G05D 1/02(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G05D; G06N; G06T; B25J

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: 自动驾驶 or 自动驾驶 or 无人驾驶 or 无人车, 加强学习 or 增强学习 or 强化学习, 避障 or 障碍 or 避碰, 训练 p 模型, 神经网络, Q网络 or Q学习, 映射 or 端对端 or 端到端, 梯度, 指令 or 命令 or 策略, 历史 or 记录, 奖励 or 奖赏, 强化 or 增强; 评估 or 打分 or 分数 or 估分 or 分值 or 评分, DRL or DQN or (reinforce+ w learn+), Q w (network or learn+), obstacle? or avoid+, scor+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107491072 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 19 December 2017 (2017-12-19) claims 1-12, description, paragraphs [0026]-[0098], and figures 1-6	1-12
X	CN 107065890 A (BEIHANG UNIVERSITY) 18 August 2017 (2017-08-18) description, paragraphs [0018]-[0046] and [0055]-[0062], and figures 1 and 3	1-12
X	CN 105139072 A (DONGHUA UNIVERSITY) 09 December 2015 (2015-12-09) description, paragraphs [0012]-[0016] and [0025]-[0052], and figures 1-7	1-12
A	CN 104317297 A (SHENYANG UNIVERSITY OF CHEMICAL TECHNOLOGY) 28 January 2015 (2015-01-28) entire document	1-12
A	CN 106558058 A (BEIJING TUSIMPLE FUTURE TECHNOLOGY CO., LTD.) 05 April 2017 (2017-04-05) entire document	1-12
A	WO 0178951 A1 (LIN, ZHIMIN) 25 October 2001 (2001-10-25) entire document	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

18 September 2018

Date of mailing of the international search report

27 September 2018

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/098637

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107491072	A	19 December 2017	None			
CN	107065890	A	18 August 2017	None			
CN	105139072	A	09 December 2015	None			
CN	104317297	A	28 January 2015	None			
CN	106558058	A	05 April 2017	None			
WO	0178951	A1	25 October 2001	AU	3678500	A	30 October 2001

国际检索报告

国际申请号

PCT/CN2018/098637

A. 主题的分类

G05D 1/02 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G05D; G06N; G06T; B25J

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT, CNKI, WPI, EPDOC: 自动驾驶 or 自动驾驶 or 无人驾驶 or 无人车, 加强学习 or 增强学习 or 强化学习, 避障 or 障碍 or 避碰, 训练 p 模型, 神经网络, Q网络 or Q学习, 映射 or 端对端 or 端到端, 梯度, 指令 or 命令 or 策略, 历史 or 记录, 奖励 or 奖赏, 强化 or 增强; 评估 or 打分 or 分数 or 估分 or 分值 or 评分, DRL or DQN or (reinforc+ w learn+), Q w (network or learn+), obstacle? or avoid+, scor+

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 107491072 A (百度在线网络技术北京有限公司) 2017年 12月 19日 (2017 - 12 - 19) 权利要求1-12、说明书第[0026]-[0098]段、图1-6	1-12
X	CN 107065890 A (北京航空航天大学) 2017年 8月 18日 (2017 - 08 - 18) 说明书第[0018]-[0046], [0055]-[0062]段、图1, 3	1-12
X	CN 105139072 A (东华大学) 2015年 12月 9日 (2015 - 12 - 09) 说明书第[0012]-[0016], [0025]-[0052]段、图1-7	1-12
A	CN 104317297 A (沈阳化工大学) 2015年 1月 28日 (2015 - 01 - 28) 全文	1-12
A	CN 106558058 A (北京图森未来科技有限公司) 2017年 4月 5日 (2017 - 04 - 05) 全文	1-12
A	WO 0178951 A1 (LIN, ZHIMIN) 2001年 10月 25日 (2001 - 10 - 25) 全文	1-12

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2018年 9月 18日

国际检索报告邮寄日期

2018年 9月 27日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10) 62019451

受权官员

李思思

电话号码 86-(10)-53962608

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/098637

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	107491072	A	2017年 12月 19日	无			
CN	107065890	A	2017年 8月 18日	无			
CN	105139072	A	2015年 12月 9日	无			
CN	104317297	A	2015年 1月 28日	无			
CN	106558058	A	2017年 4月 5日	无			
WO	0178951	A1	2001年 10月 25日	AU	3678500	A	2001年 10月 30日

表 PCT/ISA/210 (同族专利附件) (2015年1月)