



(12) 发明专利

(10) 授权公告号 CN 114972425 B

(45) 授权公告日 2025. 01. 10

(21) 申请号 202210546841.7

(22) 申请日 2022.05.18

(65) 同一申请的已公布的文献号

申请公布号 CN 114972425 A

(43) 申请公布日 2022.08.30

(73) 专利权人 北京地平线机器人技术研发有限公司

地址 100086 北京市海淀区丰豪东路9号院
2号楼3层1单元302

(72) 发明人 朱红梅 张骞 任伟强 孟文明

(74) 专利代理机构 北京弘权知识产权代理有限公司 11363

专利代理师 李少丹 许伟群

(51) Int. Cl.

G06T 7/246 (2017.01)

G06T 7/292 (2017.01)

G06V 10/80 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

G06V 10/62 (2022.01)

G06V 20/58 (2022.01)

(56) 对比文件

CN 111626219 A, 2020.09.04

CN 113033523 A, 2021.06.25

审查员 严凯丽

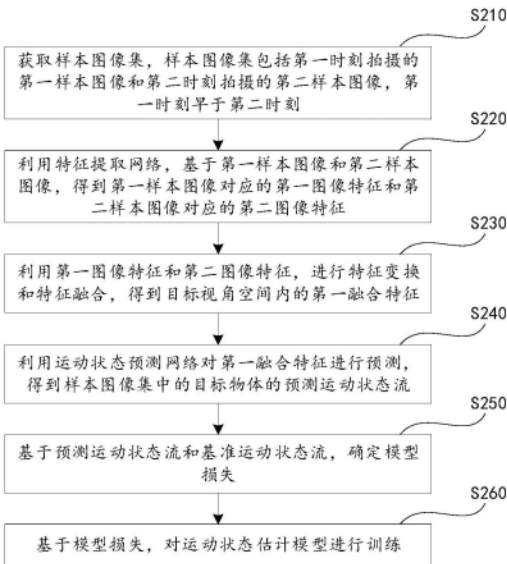
权利要求书3页 说明书16页 附图7页

(54) 发明名称

运动状态估计模型的训练方法、运动状态估计方法及装置

(57) 摘要

本公开实施例公开了一种运动状态估计模型的训练方法、运动状态估计方法及装置。该方法包括：利用特征提取网络，基于第一时刻拍摄的第一样本图像和第二时刻拍摄的第二样本图像，得到第一样本图像对应的第一图像特征和第二样本图像对应的第二图像特征；利用第一图像特征和第二图像特征，进行特征变换和特征融合，得到目标视角空间内的第一融合特征；利用运动状态预测网络对第一融合特征进行预测，得到样本图像集中的目标物体的预测运动状态流；基于预测运动状态流和基准运动状态流，确定模型损失；基于模型损失，对运动状态估计模型进行训练。这样，训练好的运动状态估计模型可以直接利用车载相机拍摄的多张图像估计目标物体的运动状态，提高了对目标物体运动状态的估计效率。



1. 一种运动状态估计模型的训练方法,所述运动状态估计模型包括特征提取网络和运动状态预测网络,所述方法包括:

获取样本图像集,所述样本图像集包括第一时刻拍摄的第一样本图像和第二时刻拍摄的第二样本图像,所述第一时刻早于所述第二时刻;

利用所述特征提取网络,基于所述第一样本图像和所述第二样本图像,得到所述第一样本图像对应的第一图像特征和所述第二样本图像对应的第二图像特征;

利用所述第一图像特征和所述第二图像特征,进行特征变换和特征融合,得到目标视角空间内的第一融合特征;

利用所述运动状态预测网络对所述第一融合特征进行预测,得到所述样本图像集中的目标物体的预测运动状态流;

基于所述预测运动状态流和基准运动状态流,确定模型损失;所述基准运动状态流为所述目标视角空间内用于表征所述目标物体的运动状态真值的张量;

基于所述模型损失,对所述运动状态估计模型进行训练;

其中,利用所述第一图像特征和所述第二图像特征,进行特征变换和特征融合,得到目标视角空间内的第一融合特征,包括:利用第一转移矩阵将所述第一图像特征和第二图像特征投影至所述目标视角空间内,得到所述第一图像特征在所述目标视角空间内对应的第一投影特征和所述第二图像特征在所述目标视角空间内对应的第二投影特征;利用第二转移矩阵,对所述第一投影特征进行变换,得到所述第一投影特征在所述第二时刻对应的第三投影特征;将所述第三投影特征与所述第二投影特征进行特征融合,得到所述第一融合特征;所述第一转移矩阵为目标视角坐标系到图像坐标系的单应性矩阵;所述第二转移矩阵为第一时刻到第二时刻的车辆姿态变化矩阵。

2. 根据权利要求1所述的方法,其中,所述特征提取网络包括第一特征提取网络和第二特征提取网络,所述利用所述特征提取网络,基于所述第一样本图像和所述第二样本图像,得到所述第一样本图像对应的第一图像特征和所述第二样本图像对应的第二图像特征,包括:

利用所述第一特征提取网络,基于所述第一样本图像,得到所述第一图像特征;

利用所述第二特征提取网络,基于所述第二样本图像,得到所述第二图像特征。

3. 根据权利要求1所述的方法,其中,所述基准运动状态流,通过以下步骤确定:

获取点云数据集,所述点云数据集包括所述第一时刻采集的第一点云数据和所述第二时刻采集的第二点云数据;

利用目标检测算法,确定所述目标物体在所述第一点云数据中的第一方位信息和所述目标物体在所述第二点云数据中的第二方位信息;

利用所述第一方位信息和所述第二方位信息,确定所述目标物体从所述第一时刻到所述第二时刻的方位变化信息;

利用所述方位变化信息更新初始的运动状态流,得到所述基准运动状态流。

4. 根据权利要求3所述的方法,其中,所述利用所述第一方位信息和所述第二方位信息,确定所述目标物体从所述第一时刻到所述第二时刻的方位变化信息,包括:

利用第三转移矩阵,将所述第一方位信息和所述第二方位信息投影至目标视角空间内,得到所述第一方位信息在所述目标视角空间内对应的第一投影方位信息和所述第二方

位信息在所述目标视角空间内对应的第二投影方位信息,所述第三转移矩阵为车载雷达坐标系到所述目标视角坐标系的外参矩阵;

利用所述第二转移矩阵,对所第一投影方位信息进行变换,得到所述第一投影方位信息在所述第二时刻对应的第三投影方位信息;

利用所述第二投影方位信息和所述第三投影方位信息,确定所述方位变化信息。

5.根据权利要求3所述的方法,其中,所述样本图像集中包括多个所述目标物体,所述利用所述第一方位信息和所述第二方位信息,确定所述目标物体从所述第一时刻到所述第二时刻的方位变化信息,包括:

利用目标追踪算法,确定多个所述目标物体中的每个所述目标物体的所述第一方位信息和所述第二方位信息的对应关系;

基于所述对应关系,分别利用每个所述目标物体的所述第一方位信息和所述第二方位信息,确定每个所述目标物体从所述第一时刻到所述第二时刻的所述方位变化信息。

6.一种运动状态估计方法,包括:

获取待处理图像集,所述待处理图像集包括第三时刻拍摄的第一待处理图像和当前时刻拍摄的第二待处理图像,所述第三时刻早于当前时刻;

利用权利要求1-5任一项所述的运动状态估计模型的训练方法训练得到的运动状态估计模型,基于所述待处理图像集,得到目标物体在当前时刻的运动状态流。

7.一种运动状态估计模型的训练装置,所述运动状态估计模型包括特征提取网络和运动状态预测网络,所述装置包括:

样本获取模块,用于获取样本图像集,所述样本图像集包括第一时刻拍摄的第一样本图像和第二时刻拍摄的第二样本图像,所述第一时刻早于所述第二时刻;

特征提取模块,用于利用所述特征提取网络,基于所述第一样本图像和所述第二样本图像,得到所述第一样本图像对应的第一图像特征和所述第二样本图像对应的第二图像特征;

特征融合模块,用于利用所述第一图像特征和所述第二图像特征,进行特征变换和特征融合,得到目标视角空间内的第一融合特征;

预测模块,用于利用所述运动状态预测网络对所述第一融合特征进行预测,得到所述样本图像集中的目标物体的预测运动状态流;

损失确定模块,用于基于所述预测运动状态流和基准运动状态流,确定模型损失;所述基准运动状态流为所述目标视角空间内用于表征所述目标物体的运动状态真值的张量;

训练模块,用于基于所述模型损失,对所述运动状态估计模型进行训练;

其中,所述特征融合模块,具体用于利用第一转移矩阵将所述第一图像特征和第二图像特征投影至所述目标视角空间内,得到所述第一图像特征在所述目标视角空间内对应的第一投影特征和所述第二图像特征在所述目标视角空间内对应的第二投影特征;利用第二转移矩阵,对所述第一投影特征进行变换,得到所述第一投影特征在所述第二时刻对应的第三投影特征;将所述第三投影特征与所述第二投影特征进行特征融合,得到所述第一融合特征;所述第一转移矩阵为目标视角坐标系到图像坐标系的单应性矩阵;所述第二转移矩阵为第一时刻到第二时刻的车辆姿态变化矩阵。

8.一种运动状态估计装置,包括:

图像获取模块,用于获取待处理图像集,所述待处理图像集包括第三时刻拍摄的第一待处理图像和当前时刻拍摄的第二待处理图像,所述第三时刻早于当前时刻;

运动状态估计模块,用于利用权利要求1-5任一项所述的运动状态估计模型的训练方法训练得到的运动状态估计模型,基于所述待处理图像集,得到目标物体在当前时刻的运动状态流。

9.一种计算机可读存储介质,所述存储介质存储有计算机程序,所述计算机程序用于执行上述权利要求1-5任一所述的运动状态估计模型的训练方法或权利要求6所述的运动状态估计方法。

10.一种电子设备,所述电子设备包括:

处理器;

用于存储所述处理器可执行指令的存储器;

所述处理器,用于从所述存储器中读取所述可执行指令,并执行所述指令以实现上述权利要求1-5任一所述的运动状态估计模型的训练方法或权利要求6所述的运动状态估计方法。

运动状态估计模型的训练方法、运动状态估计方法及装置

技术领域

[0001] 本申请涉及自动驾驶技术领域,尤其涉及一种运动状态估计模型的训练方法、运动状态估计方法、装置、存储介质及电子设备。

背景技术

[0002] 在自动驾驶场景中,自动驾驶系统可以利用相邻时刻拍摄的车辆周围的多张图像,来估计车辆周围的目标物体的运动状态。

[0003] 目前,自动驾驶系统需要首先利用目标追踪算法,对多张图像中的目标物体进行关联匹配,然后才能进一步计算目标物体的运动状态,无法直接从多张图像中获取目标物体的运动状态。

发明内容

[0004] 目前,自动驾驶系统无法直接从多张图像中获取目标物体的运动状态,导致对目标物体运动状态的估计效率较低。

[0005] 为了解决上述技术问题,提出了本公开。本公开的实施例提供了一种运动状态估计模型的训练方法、运动状态估计方法及装置。

[0006] 根据本公开的第一方面,提供了一种运动状态估计模型的训练方法,该运动状态估计模型包括特征提取网络和运动状态预测网络,该训练方法包括:获取样本图像集,样本图像集包括第一时刻拍摄的第一样本图像和第二时刻拍摄的第二样本图像,第一时刻早于第二时刻;利用特征提取网络,基于第一样本图像和第二样本图像,得到第一样本图像对应的第一图像特征和第二样本图像对应的第二图像特征;利用第一图像特征和第二图像特征,进行特征融合,进行特征变换和特征融合,得到目标视角空间内的第一融合特征;利用运动状态预测网络对第一融合特征进行预测,得到样本图像集中的目标物体的预测运动状态流;基于预测运动状态流和基准运动状态流,确定模型损失;基于模型损失,对运动状态估计模型进行训练。

[0007] 根据本公开的第二方面,提供了一种运动状态估计方法,包括:获取待处理图像集,待处理图像集包括第三时刻拍摄的第一待处理图像和当前时刻拍摄的第二待处理图像,第三时刻早于当前时刻;利用本公开第一方面提供的运动状态估计模型的训练方法训练得到的运动状态估计模型,基于待处理图像集,得到目标物体在当前时刻的运动状态流。

[0008] 根据本公开的第三方面,提供了一种运动状态估计模型的训练装置,该运动状态估计模型包括特征提取网络和运动状态预测网络,该训练装置包括:样本获取模块,用于获取样本图像集,样本图像集包括第一时刻拍摄的第一样本图像和第二时刻拍摄的第二样本图像,第一时刻早于第二时刻;特征提取模块,用于利用特征提取网络,基于第一样本图像和第二样本图像,得到第一样本图像对应的第一图像特征和第二样本图像对应的第二图像特征;特征融合模块,用于利用第一图像特征和第二图像特征,进行特征融合,进行特征变换和特征融合,得到目标视角空间内的第一融合特征;预测模块,用于利用运动状态预测网

络对第一融合特征进行预测,得到样本图像集中的目标物体的预测运动状态流;损失确定模块,用于基于预测运动状态流和基准运动状态流,确定模型损失;训练模块,用于基于模型损失,对运动状态估计模型进行训练。

[0009] 根据本公开的第四方面,提供了一种运动状态估计装置,包括:图像获取模块,用于获取待处理图像集,待处理图像集包括第三时刻拍摄的第一待处理图像和当前时刻拍摄的第二待处理图像,第三时刻早于当前时刻;运动状态估计模块,用于利用上述第一方面所述的运动状态估计模型的训练方法训练得到的运动状态估计模型,基于待处理图像集,得到目标物体在当前时刻的运动状态流。

[0010] 根据本公开的第五方面,提供了一种电子设备,电子设备包括:处理器;用于存储处理器可执行指令的存储器;处理器,用于从存储器中读取可执行指令,并执行指令以实现上述的运动状态估计模型的训练方法或运动状态估计方法。

[0011] 本公开实施例的技术方案,首先利用特征提取网络,从先后拍摄的第一样本图像和第二样本图像中提取第一图像特征和第二图像特征;然后对第一图像特征和第二图像特征进行特征融合,得到第一融合特征;接下来,利用运动状态预测网络对第一融合特征进行预测,得到目标物体的预测运动状态流;基于预测运动状态流和基准运动状态流,确定模型损失;基于模型损失,对运动状态估计模型进行训练,得到了训练好的运动状态估计模型。这样,训练好的运动状态估计模型可以直接利用车载相机拍摄的多张图像估计目标物体的运动状态,提高了对目标物体运动状态的估计效率。

附图说明

[0012] 通过结合附图对本申请实施例进行更详细的描述,本申请的上述以及其他目的、特征和优势将变得更加明显。附图用来提供对本申请实施例的进一步理解,并且构成说明书的一部分,与本申请实施例一起用于解释本申请,并不构成对本申请的限制。在附图中,相同的参考标号通常代表相同部件或步骤。

[0013] 图1是适用于本公开实施例的各方法和装置的示例性系统架构示意图。

[0014] 图2是本公开一示例性实施例提供的运动状态估计模型的训练方法的流程图。

[0015] 图3是本公开一示例性实施例提供的运动状态估计模型的算法结构的示意图。

[0016] 图4是本公开一示例性实施例提供的运动状态估计模型的训练方法步骤S230的流程图。

[0017] 图5是本公开一示例性实施例提供的确定基准运动状态流的流程图。

[0018] 图6是本公开一示例性实施例提供的确定基准运动状态流的步骤S330的流程图。

[0019] 图7是本公开一示例性实施例提供的确定基准运动状态流的步骤S320的流程图。

[0020] 图8是本公开一示例性实施例提供的从点云数据生成基准运动状态流的数据处理框图。

[0021] 图9是本公开一示例性实施例提供的运动状态估计方法的流程图。

[0022] 图10是本公开一示例性实施例提供的运动状态估计模型的训练装置的结构图。

[0023] 图11是本公开一示例性实施例提供的运动状态估计模型的训练装置的另一结构图。

[0024] 图12是本公开一示例性实施例提供的运动状态估计装置的结构图。

[0025] 图13是本公开一示例性实施例提供的电子设备的结构框图。

具体实施方式

[0026] 下面,将参考附图详细地描述根据本申请的示例实施例。显然,所描述的实施例仅仅是本申请的一部分实施例,而不是本申请的全部实施例,应理解,本申请不受这里描述的示例实施例的限制。

[0027] 应注意到:除非另外具体说明,否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本公开的范围。

[0028] 本领域技术人员可以理解,本公开实施例中的“第一”、“第二”等术语仅用于区别不同步骤、设备或模块等,既不代表任何特定技术含义,也不表示它们之间的必然逻辑顺序。

[0029] 还应理解,在本公开实施例中,“多个”可以指两个或两个以上,“至少一个”可以指一个、两个或两个以上。

[0030] 还应理解,对于本公开实施例中提及的任一部件、数据或结构,在没有明确限定或者在前后文给出相反启示的情况下,一般可以理解为一个或多个。

[0031] 另外,本公开中术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。另外,本公开中字符“/”,一般表示前后关联对象是一种“或”的关系。

[0032] 还应理解,本公开对各个实施例的描述着重强调各个实施例之间的不同之处,其相同或相似之处可以相互参考,为了简洁,不再一一赘述。

[0033] 同时,应当明白,为了便于描述,附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

[0034] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的,决不作为对本公开及其应用或使用的任何限制。

[0035] 对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论,但在适当情况下,技术、方法和设备应当被视为说明书的一部分。

[0036] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步讨论。

[0037] 本公开实施例可以应用于终端设备、计算机系统、服务器等电子设备,其可与众多其它通用或专用计算系统环境或配置一起操作。适于与终端设备、计算机系统、服务器等电子设备一起使用的众所周知的终端设备、计算系统、环境和/或配置的例子包括但不限于:个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境,等等。

[0038] 终端设备、计算机系统、服务器等电子设备可以在由计算机系统执行的计算机系统可执行指令(诸如程序模块)的一般语境下描述。通常,程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等,它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施,分布式云计算环境中,任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中,程序模块可以位于包括存储

设备的本地或远程计算系统存储介质上。

[0039] 申请概述

[0040] 在自动驾驶场景中,自动驾驶系统可以利用车载相机对其周围环境进行拍摄,并基于拍摄到的图像估计其周围环境中的各个目标物体的运动状态。其中,目标物体例如可以是行人、其他车辆、动物等。

[0041] 目前,对目标物体的运动状态估计方法一般会包括以下步骤:首先,从车载相机在相邻两个时刻拍摄的图像中估计出各个目标物体的边界,例如3D box;然后,利用目标追踪算法对两张图像中的各个3D box进行匹配,以确定各个目标物体在两张图像中的匹配关系;最后,根据车辆自身的运动状态和目标物体的匹配关系,计算出目标物体的运动状态,例如:运动方向、速度、角速度等。

[0042] 可见,目前的自动驾驶系统无法直接从车载相机拍摄的多张图像中获取目标物体的运动状态,导致对目标物体运动状态的估计效率较低。

[0043] 示例性系统

[0044] 图1是适用于本公开实施例的各方法和装置的示例性系统架构示意图。

[0045] 如图1所示,系统架构100可以包括终端设备101、网络102、服务器103和采集设备104。

[0046] 终端设备101可以是各种电子设备,包括但不限于移动电话、笔记本电脑、平板电脑PAD、车载终端(例如车载导航终端)、自动驾驶计算平台、车载计算模块,移动数据中心(mobiledata center,MDC)等。终端设备101上可以安装有各种通讯客户端应用,例如导航类应用,电子地图应用、搜索类应用、网页浏览器应用、即时通信工具等。在一些实现方式中,终端设备101可以设置在各类移动物体上,例如车辆(汽车)、轮船、火车、地铁、飞机等,以及各种机器人,例如:服务机器人、运输机器人、自主导引机器人(automated guided vehicle,AGV)、无人地面车(unmanned ground vehicle,UGV)等。

[0047] 网络102用于在终端设备101和服务器103之间提供通信链路的介质。网络102可以包括2G、3G、4G和5G等多种模式的网络,用于使终端设备101与服务器103通过网络102实现通信,以接收或发送消息等,例如实现车辆的车联万物(vehicle-to-everything,V2X)等。

[0048] 服务器103可以是提供各种服务的服务器,例如可以是利用采集设备104采集的样本图像集进行模型训练,或者利用采集设备104采集的待处理图像集预测目标物体的运动状态的人工智能(artificial intelligence,AI)服务器等。示例性的,AI服务器可以利用样本图像集进行模型训练,得到运动状态估计模型,或者利用运动状态估计模型对接收到的待处理图像集进行预测,得到目标物体的运动状态。需要补充说明的是,本公开中的服务器103可以是终端设备的厂商或者自动驾驶系统的服务商提供的后台服务器,也可以是互联网服务供应商(internet service provider,ISP)或者云服务商提供的云服务器,还可以是在不同的网络节点部署的分布式服务器集群,本公开实施例对服务器103的实现方式不做具体限定。

[0049] 采集设备104可以是各种用于采集图像或点云数据的设备,例如车载相机和/或车载雷达等。

[0050] 车载相机可以用于对车辆周围进行拍摄,得到样本图像集或者待处理图像集。采集设备104可以包括一个或者多个车载相机,多个车载相机可以设置在车辆的不同位置,朝

向不同的方向,以对车辆周围的不同方向进行拍摄。当采集设备104包括多个车载相机时,采集设备104在同一时刻可以得到不同车载相机拍摄的多张图像。车载相机可以是鱼眼相机、窄角相机等,本公开实施例对车载相机的类型不作限定。

[0051] 车载雷达可以用于对车辆周围环境进行扫描,得到车辆周围环境对应的点云数据集。采集设备104可以包括一个或者多个车载雷达,多个车载雷达可以设置在车辆的不同位置,朝向不同的方向,以对车辆周围的不同方向进行扫描。车载雷达可以是毫米波雷达、激光雷达、超声波雷达、红外雷达等,本公开实施例对车载雷达的类型不作限定。

[0052] 需要说明的是,本公开的实施例所提供的运动状态估计模型的训练方法或运动状态估计方法可以由终端设备101执行,也可以由服务器103执行,相应地,运动状态估计模型的训练装置或运动状态估计装置可以设置于终端设备101中,也可以设置于服务器103中。

[0053] 应该理解,图1中的终端设备、网络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终端设备、网络和服务器的。在用于模型训练的样本图像集和用于运动状态估计的待处理图像集不需要从远程获取的情况下,上述系统架构可以不包括网络,只包括终端设备101和采集设备104。

[0054] 示例性方法

[0055] 图2是本公开一示例性实施例提供的运动状态估计模型的训练方法的流程图。

[0056] 本实施例可应用在电子设备上(例如图1所示的终端设备101或服务器103)。其中,电子设备中可以部署有待训练的运动状态估计模型,该待训练的运动状态估计模型可以包括一个或者多个神经网络,例如,包括一个或者多个特征提取网络,以及一个或者多个运动状态预测网络等。

[0057] 如图2所示,该运动状态估计模型的训练方法可以包括如下步骤:

[0058] 步骤S210,获取样本图像集,样本图像集包括第一时刻拍摄的第一样本图像和第二时刻拍摄的第二样本图像,第一时刻早于第二时刻。

[0059] 样本图像集可以是车载相机对车辆周围进行拍摄得到的。样本图像集中可以包括拍摄于不同时刻的多张样本图像。本公开实施例中,第一时刻和第二时刻指的是先后到来的任意两个不同时刻。第一样本图像和第二样本图像是指在先后到来的任意两个不同时刻拍摄的图像。其中,当车载相机的数量为多个时,第一样本图像可以包括多个车载相机在第一时刻拍摄的多张图像,第二样本图像可以包括多个车载相机在第二时刻拍摄的多张图像。

[0060] 示例性的,车载相机可以以特定的时间间隔 Δt 拍摄样本图像,例如:在 $t-1$ 时刻拍摄一张样本图像,然后间隔 Δt ,在 t 时刻再拍摄一张样本图像。那么,对于在 $t-1$ 时刻和 t 时刻拍摄的图像来说,在 $t-1$ 时刻拍摄的样本图像为第一样本图像 I_{t-1} ,在 t 时刻拍摄的样本图像为第二样本图像 I_t 。

[0061] 示例性的,车载相机可以以一定的帧率拍摄视频流。那么,第一样本图像和第二样本图像可以是视频流中的任意相邻两帧样本图像。例如:第一样本图像 I_{t-1} 可以是车载相机在 $t-1$ 时刻拍摄的相邻两帧样本图像中的前一帧样本图像,第二样本图像 I_t 可以是车载相机在 t 时刻拍摄的相邻两帧样本图像中的后一帧样本图像。

[0062] 示例性的,电子设备还可以对车载相机拍摄的视频流进行帧采样,从视频流中每隔 n 帧图像选出一帧图像作为样本图像。例如:将车载相机在 $t-1$ 时刻拍摄的第 m 帧样本图

像作为第一样本图像,将车载相机在t时刻拍摄的第m+n帧样本图像作为第二样本图像。

[0063] 步骤S220,利用特征提取网络,基于第一样本图像和第二样本图像,得到第一样本图像对应的第一图像特征和第二样本图像对应的第二图像特征。

[0064] 其中,第一图像特征是第一样本图像在相机视角空间内的图像特征,第二图像特征是第二样本图像在相机视角空间内的图像特征。第一图像特征和第二图像特征包括但不限于样本图像中的各目标物体的纹理特征、边缘轮廓特征、类别特征、灰度特征、直方图特征等,这里不做具体限定。

[0065] 其中,特征提取网络可以由一个或者多个神经网络构成,例如:卷积神经网络CNN、深度残差网络resnet、循环神经网络RNN、Transformer网络、densenet、mobilenet、BERT网络等,各神经网络可以包括一个或者多个神经网络层,例如:卷积层、池化层、全连接层等,这里不做具体限定。

[0066] 具体实现中,电子设备可以将第一样本图像和第二样本图像输入到特征提取网络中,特征提取网络可以利用其内部的神经网络层对第一样本图像和第二样本图像进行卷积、池化、全连接操作,得到第一图像特征和第二图像特征。

[0067] 步骤S230,利用第一图像特征和第二图像特征,进行特征变换和特征融合,得到目标视角空间内的第一融合特征。

[0068] 其中,目标视角空间可以是鸟瞰视角(birds eye views,BEV)空间,或者是其他视角空间,本公开实施例对此不作限定。

[0069] 可以理解的是,由于第一样本图像和第二样本图像是相机拍摄的图像,因此,第一图像特征和第二图像特征是相机视角空间内的图像特征,其采用的坐标系是相机视角空间内的图像坐标系,当车载相机数量为多个时,不同相机所对应的图像坐标系也是不同的,不利于进行特征融合。另外,自动驾驶场景也通常需要在BEV空间内估计目标物体的运动状态。因此,为了便于特征融合以及后续在BEV空间内估计目标物体的运动状态,电子设备可以利用特征变换,将第一图像特征和第二图像特征转移同一个的目标视角空间内,例如BEV空间内,然后,对转换后的第一图像特征和第二图像特征进行融合,得到第一融合特征。

[0070] 本公开实施例中,特征融合可以采用concat融合、add融合、或者其他的融合方式实现,本公开实施例对此不作限定。

[0071] 步骤S240,利用运动状态预测网络对第一融合特征进行预测,得到样本图像集中的目标物体的预测运动状态流。

[0072] 其中,运动状态预测网络可以由一个或者多个神经网络构成,例如:循环神经网络、长短期记忆网络LSTM、门控循环单元网络GRU等,这里不做具体限定。

[0073] 电子设备可以利用机器学习方法,将第一融合特征输入到待训练的运动状态预测网络中,通过运动状态预测网络对样本图像集中的目标物体进行预测,得到目标物体的预测运动状态流。

[0074] 在一种实现方式中,预测运动状态流可以是目标视角空间内用于表征目标物体的运动状态预测值的张量。该张量的尺寸例如可以是 $H*W*3$ 。其中:H和W表示目标视角空间的尺寸;3表示目标物体的运动状态的预测值的3个参数,分别为目标物体从第一时刻到第二时刻沿x方向位移的预测值,目标物体从第一时刻到第二时刻沿y方向位移的预测值,目标物体从第一时刻到第二时刻的朝向角变化量的预测值。其中,x方向为车辆前后方向,y方向

为车辆左右方向。

[0075] 示例性的,假设目标视角空间内的位置 (i, j) 在某个目标物体内,那么,预测运动状态流中的 $(i, j, 0)$ 位置的参数值即表示该目标物体从第一时刻到第二时刻沿x方向位移的预测值, $(i, j, 1)$ 位置的参数值即表示该目标物体从第一时刻到第二时刻沿y方向位移的预测值, $(i, j, 2)$ 位置的参数值即表示该目标物体从第一时刻到第二时刻的朝向角变化量的预测值。

[0076] 可以理解的是,如果预测运动状态流中的 (i, j) 位置不属于任何一个目标物体,那么 $(i, j, 0)$ 、 $(i, j, 1)$ 、 $(i, j, 2)$ 的值均为0。

[0077] 步骤S250,基于预测运动状态流和基准运动状态流,确定模型损失。

[0078] 在一种实现方式中,基准运动状态流可以是目标视角空间内用于表征目标物体的运动状态真值的张量。基准运动状态流优选与预测运动状态流的张量尺寸相同,例如可以是 $H \times W \times 3$,以便于计算。其中:H和W表示目标视角空间的尺寸;3表示目标物体的运动状态的真值的3个参数,分别为目标物体从第一时刻到第二时刻沿x方向位移的真值,目标物体从第一时刻到第二时刻沿y方向位移的真值,目标物体从第一时刻到第二时刻的朝向角变化量的真值。

[0079] 示例性的,假设目标视角空间内的位置 (i, j) 在某个目标物体内,那么,基准运动状态流中的 $(i, j, 0)$ 位置的参数值即表示该目标物体从第一时刻到第二时刻沿x方向位移的真值, $(i, j, 1)$ 位置的参数值即表示该目标物体从第一时刻到第二时刻沿y方向位移的真值, $(i, j, 2)$ 位置的参数值即表示该目标物体从第一时刻到第二时刻的朝向角变化量的真值。

[0080] 可以理解的是,如果基准运动状态流中的 (i, j) 位置不属于任何一个目标物体,那么 $(i, j, 0)$ 、 $(i, j, 1)$ 、 $(i, j, 2)$ 的值均为0。

[0081] 本公开实施例中,电子设备可以基于L1损失函数、L2损失函数、Smooth L1损失函数、交叉熵损失函数、log损失函数、最大似然损失函数等任意损失函数确定模型损失。以基于L1损失函数为例,模型损失Loss可以为:

[0082]
$$\text{Loss} = |\text{motion_flow}_{\text{pred}} - \text{motion_flow}_{\text{gt}}|$$

[0083] 其中, $\text{motion_flow}_{\text{pred}}$ 表示预测运动状态流, $\text{motion_flow}_{\text{gt}}$ 表示基准运动状态流。 $|\text{motion_flow}_{\text{pred}} - \text{motion_flow}_{\text{gt}}|$ 表示预测运动状态流与基准运动状态流的误差的绝对值。

[0084] 步骤S260,基于模型损失,对运动状态估计模型进行训练。

[0085] 基于网络损失,运动状态估计模型可以以基准运动状态流作为其预测输出的目标,进行多次迭代,从而不断更新其内部的模型参数,使模型损失不断收敛。其中,当模型损失收敛至预设条件,例如模型损失Loss的值小于预设阈值时,结束训练,得到训练好的运动状态估计模型,或者,当运动状态预测网络的训练迭代达到预设次数时,结束训练,得到训练好的运动状态估计模型。

[0086] 本公开实施例的技术方案,首先利用特征提取网络,从先后拍摄的第一样本图像和第二样本图像中提取第一图像特征和第二图像特征;然后对第一图像特征和第二图像特征进行特征融合,得到第一融合特征;接下来,利用运动状态预测网络对第一融合特征进行预测,得到目标物体的预测运动状态流;基于预测运动状态流和基准运动状态流,确定模型

损失;基于模型损失,对运动状态估计模型进行训练,得到了训练好的运动状态估计模型。这样,训练好的运动状态估计模型可以直接利用车载相机拍摄的多张图像估计目标物体的运动状态,提高了对目标物体运动状态的估计效率。

[0087] 图3是本公开一示例性实施例提供的运动状态估计模型的算法结构的示意图。

[0088] 如图3所示,在一种实现方式中,运动状态估计模型的特征提取网络可以包括第一特征提取网络和第二特征提取网络,其中,第一特征提取网络和第二特征提取网络可以是相同类型的神经网络,也可以是不同类型的神经网络。

[0089] 基于图3示出的运动状态估计模型,在上述图2所示的实施例的基础上,步骤S220可以包括以下步骤:

[0090] 步骤S221,利用第一特征提取网络,基于第一样本图像,得到第一图像特征。

[0091] 具体实现中,电子设备可以将第一样本图像输入到第一特征提取网络中。第一样本图像的尺寸可以为 $(v*3*h*w)$,其中, v 表示第一样本图像的数量, 3 表示样本图像的通道数为3(例如RGB通道), h 表示第一样本图像的高度, w 表示第一样本图像的宽度。第一特征提取网络可以利用其内部的神经网络层对第一样本图像进行卷积、池化、全连接操作,得到第一图像特征。

[0092] 第一图像特征可以是第一特征提取网络输出的预设尺寸的特征张量,该特征张量的尺寸可以为 $(v*c*h1*w1)$,其中, v 表示样本图像的数量, c 表示特征张量的通道数, $h1$ 表示特征张量的高度, $w1$ 表示特征张量的宽度。

[0093] 步骤S222,利用第二特征提取网络,基于第二样本图像,得到第二图像特征。

[0094] 具体实现中,电子设备可以将第二样本图像输入到第二特征提取网络中。第二样本图像的尺寸可以为 $(v*3*h*w)$,其中, v 表示第二样本图像的数量, 3 表示样本图像的通道数为3(例如RGB通道), h 表示第二样本图像的高度, w 表示第二样本图像的宽度。第二特征提取网络可以利用其内部的神经网络层对第二样本图像进行卷积、池化、全连接操作,得到第二图像特征。

[0095] 第二图像特征可以是第二特征提取网络输出的预设尺寸的特征张量,该特征张量的尺寸可以为 $(v*c*h1*w1)$,其中, v 表示样本图像的数量, c 表示特征张量的通道数, $h1$ 表示特征张量的高度, $w1$ 表示特征张量的宽度。

[0096] 这里需要补充说明的是,本公开实施例中,第一样本图像和第二样本的图像的尺寸可以相同,也可以不同,第一图像特征的尺寸与第二图像特征的尺寸可以相同,也可以不同,本公开实施例对此均不作具体限定。其中,第一样本图像和第二样本的图像的尺寸相同,第一图像特征的尺寸与第二图像特征的尺寸相同,为优选的实现方式,以利于后续对第一图像特征的尺寸与第二图像特征进行特征融合。

[0097] 本公开实施例的技术方案,利用两个特征提取网络分别对第一样本图像和第二样本图像进行特征提取,得到第一图像特征和第二图像特征,以便于后续进行特征融合。

[0098] 图4是本公开一示例性实施例提供的运动状态估计模型的训练方法步骤S230的流程图。

[0099] 如图4所示,在上述图2和图3所示实施例的基础上,步骤S230可以包括如下步骤:

[0100] 步骤S231,利用第一转移矩阵,将第一图像特征和第二图像特征投影至目标视角空间内,得到第一图像特征在目标视角空间内对应的第一投影特征和第二图像特征在目标

视角空间内对应的第二投影特征。

[0101] 其中,第一转移矩阵为目标视角坐标系到图像坐标系的单应性矩阵。

[0102] 具体实现中,电子设备首先利用第一转移矩阵,对第一图像特征和第二图像特征进行特征投影,得到第一图像特征在BEV空间内对应的第一投影特征和第二图像特征在BEV空间内对应的第二投影特征。

[0103] 第一投影特征和第二投影特征均可以是预设尺寸的特征张量。其中,第一投影特征的尺寸和第二投影特征的尺寸可以相同,也可以不同,本公开实施例对此不作限定。为了便于特征融合,第一投影特征和第二投影特征的尺寸优选相同。

[0104] 示例性的,第一投影特征和第二投影特征的尺寸均可以为 $(1*c*h2*w2)$,其中,c表示特征张量的通道数,h2表示特征张量的高度,w2表示特征张量的宽度。

[0105] 步骤S232,利用第二转移矩阵,对第一投影特征进行变换,得到第一投影特征在第二时刻对应的第三投影特征。

[0106] 其中,第二转移矩阵为第一时刻到第二时刻的车辆姿态变化矩阵。

[0107] 可以理解的是,在车辆移动时,随着车辆位姿的变化,车辆所处的BEV空间也在时刻发生变化。因此,第一投影特征和第二投影特征虽然均是BEV空间内的投影特征,但是并不是同一时刻的BEV空间内的投影特征,不利于直接进行特征融合。对此,电子设备利用第二转移矩阵,对第一投影特征进行变换,将第一投影特征变换至第二时刻的BEV空间内,得到第三投影特征。这样,第一投影特征和第三投影特征位于同一时刻的BEV空间内,有利于后续进行特征融合。

[0108] 示例性的,第三投影特征和第二投影特征的尺寸优选相同,均可以为 $(1*c*h2*w2)$,其中,c表示特征张量的通道数,h2表示特征张量的高度,w2表示特征张量的宽度。

[0109] 步骤S233,将第三投影特征与第二投影特征进行特征融合,得到第一融合特征。

[0110] 具体实现中,对第二投影特征和第三投影特征进行特征融合,可以采用concat融合、add融合、或者其他的融合方式实现,本公开实施例对此不作限定。第一融合特征的尺寸与第二投影特征或第三投影特征的尺寸可以相同,也可以不同,具体可以与融合方式有关。例如:当采用concat融合时,第一融合特征的通道数相比于第二投影特征或第三投影特征的通道数可以增加一倍,尺寸可以达到 $(1*2c*h2*w2)$ 。

[0111] 本公开实施例的方案,将不同时刻的相机视角空间内的第一图像特征和第二图像特征转换至同一时刻的BEV空间内,并进行特征融合,使第一融合特征能够在BEV空间表达目标物体的图像特征随时间的变化。

[0112] 图5是本公开一示例性实施例提供的确定基准运动状态流的流程图。

[0113] 如图5所示,在上述图2所示的实施例的基础上,确定基准运动状态流可以包括以下步骤:

[0114] 步骤S310,获取点云数据集,点云数据集包括第一时刻采集的第一点云数据和第二时刻采集的第二点云数据。

[0115] 点云数据集可以是车载雷达对车辆周围进行扫描得到的。点云数据集中可以包括车载雷达在不同时刻扫描得到的多个点云数据。本公开实施例中,第一点云数据和第二点云数据是指点云数据集中的在先后任意两个不同时刻扫描得到的点云数据。

[0116] 示例性的,车载雷达可以以特定的时间间隔 Δt 对车辆周围环境进行扫描,例如:

在 $t-1$ 时刻进行一次扫描,获取一个点云数据,然后间隔 Δt ,在 t 时刻再进行一次扫描,获取另一个点云数据。那么,对于在 $t-1$ 时刻和 t 时刻扫描得到的点云数据来说,在 $t-1$ 时刻扫描得到的点云数据为第一点云数据 P_{t-1} ,在 t 时刻扫描得到的点云数据为第二点云数据 P_t 。

[0117] 步骤S320,利用目标检测算法,确定目标物体在第一点云数据中的第一方位信息和目标物体在第二点云数据中的第二方位信息。

[0118] 第一方位信息可以包括目标物体在第一时刻的雷达坐标系中的位置信息和朝向信息,第二方位信息可以包括目标物体在第二时刻的雷达坐标系中的位置信息和朝向信息。例如,位置信息可以包括目标物体在雷达坐标系中的三维坐标 (x, y, z) ,其中,雷达坐标系的 x 方向为车辆前后方向, y 方向为车辆左右方向, z 方向为车辆高度方向。位置信息还可以包括目标物体的三维尺寸 (w, h, l) ,其中, w, h, l 分别表示目标物体的宽、高、长。朝向信息可以包括目标物体的朝向与 x 方向的夹角 yaw 。

[0119] 在一种实现方式中,目标物体的方位信息可以通过边界框(bounding box, BBox)来描述。那么,对于任意目标物体 k ,其边界框可以表示为:

$$[0120] \quad BBox_t^k = (x, y, z, w, h, l, yaw)$$

[0121] 为便于区分,本公开实施例以下将通过边界框描述的第一方位信息记作 $BBox_{t-1}$,将通过边界框描述的第二方位信息记作 $BBox_t$ 。

[0122] 步骤S330,利用第一方位信息和第二方位信息,确定目标物体从第一时刻到第二时刻的方位变化信息。

[0123] 具体实现中,电子设备可以将第一方位信息和第二方位信息转移至同一个目标视角空间内,例如同一个时刻的BEV空间内,以便于对目标物体在第一时刻和第二时刻的位置变化和朝向变化进行比较。然后,在BEV空间内,确定出目标物体从第一时刻到第二时刻的方位变化信息,其中,方位变化信息可以包括目标物体从第一时刻到第二时刻的位置变化量和朝向变化量。位置变化量可以包括目标物体的沿 x 方向的位移 dx ,沿 y 方向的位移 dy ,朝向变化量可以包括目标物体的朝向与 x 方向的夹角变化量 d_{yaw} 。

[0124] 步骤S340,利用方位变化信息更新初始的运动状态流,得到基准运动状态流。

[0125] 具体实现中,电子设备可以初始化一个运动状态流,该初始的运动状态流的尺寸优选与预测运动状态流的尺寸相同,例如均为 $H*W*3$ 。其中,初始的运动状态流中的各参数值均为0。接下来,对位于任意一个目标物体,电子设备可以将其方位变化信息(例如: dx, dy, d_{yaw})更新到初始的运动状态流中的目标物体对应的位置中,得到目标物体的基准运动状态流。

[0126] 本公开实施例的技术方案,首先利用目标检测算法,从先后扫描的第一点云数据和第二点云数据中提取目标物体的第一方位信息和第二方位信息;然后,利用第一方位信息和第二方位信息,确定目标物体从第一时刻到第二时刻的方位变化信息;最后,利用方位变化信息更新初始的运动状态流,得到基准运动状态流。从而为运动状态估计模型提供了训练目标,有利于提高运动状态估计模型的训练效果。

[0127] 图6是本公开一示例性实施例提供的确定基准运动状态流的步骤S330的流程图。

[0128] 如图6所示,在上述图5所示的实施例的基础上,步骤S330可以包括以下步骤:

[0129] 步骤S331,利用第三转移矩阵,将第一方位信息和第二方位信息投影至目标视角空间内,得到第一方位信息在目标视角空间内对应的第一投影方位信息和第二方位信息在

目标视角空间内对应的第二投影方位信息。

[0130] 其中,第三转移矩阵为车载雷达坐标系到目标视角坐标系的外参矩阵。目标视角坐标系可以是BEV坐标系,或者是其他坐标系,本公开实施例对此不作限定。

[0131] 可以理解的是,由于第一点云数据和第二点云数据是车载雷达获取的点云数据,因此,第一方位信息和第二方位信息是车载雷达坐标系下的方位信息。然而,自动驾驶场景通常需要在BEV空间内估计目标物体的运动状态,因此需要得到BEV空间内的基准运动状态流。对此,电子设备首先利用第三转移矩阵,对第一方位信息和第二方位信息进行投影,得到第一方位信息在BEV空间内对应的第一投影方位信息和第二方位信息在BEV空间内对应的第二投影方位信息。

[0132] 步骤S332,利用第二转移矩阵,对所第一投影方位信息进行变换,得到第一投影方位信息在第二时刻对应的第三投影方位信息。

[0133] 可以理解的是,在车辆移动时,随着车辆位姿的变化,车辆所处的BEV坐标系也在时刻发生变化。因此,第一投影方位信息和第二投影方位信息虽然均是BEV空间内的方位信息,但是并不是同一时刻的BEV空间内的方位信息,不利于直接计算目标物体的方位变化信息。对此,电子设备利用第二转移矩阵,对第一投影方位信息进行变换,将第一投影方位信息变换至第二时刻的BEV空间内,得到第三投影方位信息。这样,第一投影方位信息和第三投影方位信息位于同一时刻的BEV空间内,有利于后续目标物体的方位变化信息。

[0134] 步骤S333,利用第二投影方位信息和第三投影方位信息,确定方位变化信息。

[0135] 具体实现中,电子设备可以对目标物体在第二投影方位信息和第三投影方位信息中的方位进行计算,以确定出目标物体从第一时刻到第二时刻的方位变化信息。

[0136] 本公开实施例提供的方案,将不同时刻的雷达坐标系内的第一方位信息和第二方位信息转换至同一时刻的BEV空间内,从而更准确地确定目标物体的方位变化信息。

[0137] 图7是本公开一示例性实施例提供的确定基准运动状态流的步骤S320的流程图。

[0138] 可以理解的是,车辆在道路上行驶时,其周围可能会存在多个目标物体,例如多个其他车辆、多个行人等。在这种情况下,车辆需要分别确定每一个目标物体的方位变化信息。

[0139] 如图7所示,为了分别确定每一个目标物体的方位变化信息,在上述图5所示的实施例的基础上,步骤S330可以包括以下步骤:

[0140] 步骤S321,利用目标追踪算法,确定多个目标物体中的每个目标物体的第一方位信息和第二方位信息的对应关系。

[0141] 具体实现中,对于多个目标物体,电子设备可以建立第一时刻对应的第一方位信息集合以及第二时刻对应的第二方位信息集合。

[0142] 第一方位信息集合可以包括各个目标物体在第一时刻对应的第一方位信息,例如:

$$[0143] \quad A_{t-1} = [BBox_{t-1}^1, BBox_{t-1}^2, \dots, BBox_{t-1}^n]$$

[0144] 其中, A_{t-1} 表示第一方位信息集合, n 为目标物体的数量, $BBox_{t-1}^1 \sim BBox_{t-1}^n$ 为 n 个目标物体的第一方位信息。

[0145] 第二方位信息集合可以包括各个目标物体在第二时刻对应的第二方位信息,例如:

[0146] $A_t = [BBox_t^1, BBox_t^2, \dots, BBox_t^n]$

[0147] 其中, A_t 表示第二方位信息集合, n 为目标物体的数量, $BBox_t^1 \sim BBox_t^n$ 为 n 个目标物体的第二方位信息。

[0148] 接下来, 电子设备可以对第一方位信息集合和第二方位信息集合执行目标追踪算法, 以确定第一方位信息集合中的每一个第一方位信息与第二方位集合中的每一个第二方位信息的对应关系。这里的匹配关系指的是第一方位信息集合中的哪一个第一方位信息与第二方位信息中的哪一个第二方位信息属于同一个目标物体。

[0149] 其中, 目标追踪算法包括但不限于基于特征点的目标追踪算法、基于灰度的目标追踪算法或者其他目标追踪算法等, 例如: SIFT算法、SURF算法、PCA-SIFT算法、Moravec算法、Susan算法等, 本公开实施例对此不作限定。

[0150] 步骤S322, 基于对应关系, 分别利用每个目标物体的第一方位信息和第二方位信息, 确定每个目标物体从第一时刻到第二时刻的方位变化信息。

[0151] 具体实现中, 电子设备可以针对每个目标物体分别执行步骤S331-步骤S333, 以确定该目标物体的方位变化信息, 这里不再赘述。

[0152] 图8是本公开一示例性实施例提供的从点云数据生成基准运动状态流的数据处理框图。

[0153] 如图8所示, 基于上述图5-图7所示的实施例, 为了生成基准运动状态流, 电子设备可以首先利用目标检测算法, 对第一点云数据和第二点云数据进行3D目标检测, 以得到多个目标物体中的每一个目标物体的第一方位信息 (例如: 边界框 $BBox_{t-1}$) 和第二方位信息 (例如: 边界框 $BBox_t$)。接下来, 电子设备可以利用目标追踪算法, 对各个目标物体进行目标追踪, 以确定每一个目标物体在第一点云数据和第二点云数据中的对应关系。另外, 电子设备可以利用第三转移矩阵, 将第一方位信息和第二方位信息投影至BEV空间内, 得到第一投影方位信息和第二投影方位信息。接下来, 电子设备可以利用第二转移矩阵, 将所第一投影方位信息转换至第二时刻, 得到第三投影方位信息。最后, 电子设备可以基于对各个目标物体的目标追踪结果, 对各个目标物体的第二投影方位信息和第三投影方位信息进行比较, 以得到基准运动状态流。

[0154] 本公开实施例的方案, 针对车辆周围存在多个目标物体的场景, 利用目标追踪算法对多个目标物体的第一方位信息和第二方位信息的对应关系进行匹配, 并利用对应关系分别确定每一个目标物体的方位变化信息, 使运动状态估计模型能够具备同时分辨和追踪多个目标物体的运动状态的能力。

[0155] 图9是本公开一示例性实施例提供的运动状态估计方法的流程图。该方法可以应用于电子设备, 能够利用上述各实施例的方案训练的运动状态估计模型, 对目标物体的运动状态进行预测。如图9所示, 该方法包括:

[0156] 步骤S410, 获取待处理图像集, 待处理图像集包括第三时刻拍摄的第一待处理图像和当前时刻拍摄的第二待处理图像, 第三时刻早于当前时刻。

[0157] 待处理图像集可以车辆在行驶过程中, 车载相机对车辆周围进行实时拍摄的到的。待处理图像集可以包括车载相机在当前时刻之前的第三时刻拍摄的第一待处理图像, 以及在当前时刻拍摄的第二待处理图像。

[0158] 示例性的,车载相机可以以特定的时间间隔 Δt 拍摄样本图像,那么,第二待处理图像可以是车载相机在当前时刻 t_0 拍摄的图像,第一待处理图像可以是车载相机在当前时刻 t_0 之前的 $t_0 - \Delta t$ 时刻拍摄的图像。

[0159] 示例性的,车载相机可以以一定的帧率拍摄视频流。那么,第二待处理图像可以是车载相机在当前时刻拍摄的当前帧图像,第一待处理图像可以是车载相机拍摄的当前帧图像的前一帧图像。

[0160] 步骤S420,利用运动状态估计模型,基于待处理图像集,得到目标物体在当前时刻的运动状态流。

[0161] 具体实现中,电子设备可以将待处理图像输入到训练好的运动状态估计模型中,利用运动状态估计模型对待处理图像集进行预测,并输出目标物体在当前时刻的运动状态流。

[0162] 本公开实施例提供的技术方案,电子设备可以利用运动状态估计模型,直接从车载相机拍摄的多张图像估计目标物体的运动状态,提高了对目标物体运动状态的估计效率。

[0163] 示例性装置

[0164] 图10是本公开一示例性实施例提供的运动状态估计模型的训练装置的结构图。示例性的,该运动状态估计模型的训练置可以设置于电子设备中,用于执行本公开上述任一实施例的运动状态估计模型的训练方法。

[0165] 如图10所示,该图像处理装置包括:样本获取模块510,特征提取模块520,特征融合模块530,预测模块540,损失确定模块550,训练模块560。

[0166] 其中:

[0167] 样本获取模块510,用于获取样本图像集,样本图像集包括第一时刻拍摄的第一样本图像和第二时刻拍摄的第二样本图像,第一时刻早于第二时刻。

[0168] 特征提取模块520,用于利用特征提取网络,基于第一样本图像和第二样本图像,得到第一样本图像对应的第一图像特征和第二样本图像对应的第二图像特征。

[0169] 特征融合模块530,用于利用第一图像特征和第二图像特征,进行特征融合,得到第一融合特征。

[0170] 预测模块540,用于利用运动状态预测网络对第一融合特征进行预测,得到样本图像集中的目标物体的预测运动状态流。

[0171] 损失确定模块550,用于基于预测运动状态流和基准运动状态流,确定模型损失。

[0172] 训练模块560,用于基于模型损失,对运动状态估计模型进行训练。

[0173] 图11是本公开一示例性实施例提供的运动状态估计模型的训练装置的另一结构图。

[0174] 在一个实施例中,如图11所示,特征提取模块520具体包括:

[0175] 第一特征提取单元521,用于利用第一特征提取网络,基于第一样本图像,得到第一图像特征。

[0176] 第二特征提取单元522,用于利用第二特征提取网络,基于第二样本图像,得到第二图像特征。

[0177] 在一个实施例中,如图11所示,特征融合模块530具体包括:

[0178] 第一投影单元531,用于利用第一转移矩阵,将第一图像特征和第二图像特征投影至目标视角空间内,得到第一图像特征在目标视角空间内对应的第一投影特征和第二图像特征在目标视角空间内对应的第二投影特征,第一转移矩阵为目标视角坐标系到图像坐标系的单应性矩阵。

[0179] 第二投影单元532,用于利用第二转移矩阵,对第一投影特征进行变换,得到第一投影特征在第二时刻对应的第三投影特征,第二转移矩阵为第一时刻到第二时刻的车辆姿态变化矩阵。

[0180] 融合单元533,用于将第三投影特征与第二投影特征进行特征融合,进行特征变换和特征融合,得到目标视角空间内的第一融合特征。

[0181] 在一个实施例中,如图11所示,该运动状态估计模型的训练置还包括:点云获取模块610,目标检测模块620,确定模块630,生成模块640。其中:

[0182] 点云获取模块610,用于获取点云数据集,点云数据集包括第一时刻采集的第一点云数据和第二时刻采集的第二点云数据。

[0183] 目标检测模块620,用于利用目标检测算法,确定目标物体在第一点云数据中的第一方位信息和目标物体在第二点云数据中的第二方位信息。

[0184] 确定模块630,用于利用第一方位信息和第二方位信息,确定目标物体从第一时刻到第二时刻的方位变化信息。

[0185] 生成模块640,用于利用方位变化信息更新初始的运动状态流,得到基准运动状态流。

[0186] 在一个实施例中,如图11所示,确定模块630具体包括:

[0187] 第三投影单元631,用于利用第三转移矩阵,将第一方位信息和第二方位信息投影至目标视角空间内,得到第一方位信息在目标视角空间内对应的第一投影方位信息和第二方位信息在目标视角空间内对应的第二投影方位信息,第三转移矩阵为车载雷达坐标系到目标视角坐标系的外参矩阵。

[0188] 第四投影单元632,用于利用第二转移矩阵,对所第一投影方位信息进行变换,得到第一投影方位信息在第二时刻对应的第三投影方位信息。

[0189] 确定单元633,用于利用第二投影方位信息和第三投影方位信息,确定方位变化信息。

[0190] 在一个实施例中,如图11所示,确定模块630还包括:

[0191] 目标追踪单元634,用于利用目标追踪算法,确定多个目标物体中的每个目标物体的第一方位信息和第二方位信息的对应关系。

[0192] 确定单元633,还用于基于对应关系,分别利用每个目标物体的第一方位信息和第二方位信息,确定每个目标物体从第一时刻到第二时刻的方位变化信息。

[0193] 图12是本公开一示例性实施例提供的运动状态估计装置的结构图。示例性的,该运动状态估计装置可以设置于电子设备中,用于执行本公开上述任一实施例的运动状态估计方法。

[0194] 如图12所示,该运动状态估计装置包括:图像获取模块710,运动状态估计模块720。其中:

[0195] 图像获取模块710,用于获取待处理图像集,待处理图像集包括第三时刻拍摄的第

一待处理图像和当前时刻拍摄的第二待处理图像,第三时刻早于当前时刻。

[0196] 运动状态估计模块720,用于利用运动状态估计模型的训练方法训练得到的运动状态估计模型,基于待处理图像集,得到目标物体在当前时刻的运动状态流。

[0197] 示例性电子设备

[0198] 图13是本公开一示例性实施例提供的电子设备的结构框图。

[0199] 如图13所示,电子设备10包括一个或多个处理器111和存储器112。

[0200] 处理器111可以是中央处理单元(CPU)或者具有数据处理能力和/或指令执行能力的其他形式的处理单元,并且可以控制电子设备10中的其他组件以执行期望的功能。

[0201] 存储器112可以包括一个或多个计算机程序产品,计算机程序产品可以包括各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。易失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。非易失性存储器例如可以包括只读存储器(ROM)、硬盘、闪存等。在计算机可读存储介质上可以存储一个或多个计算机程序指令,处理器111可以运行程序指令,以实现上文的本公开的各个实施例的运动状态估计模型的训练方法、运动状态估计方法以及/或者其他期望的功能。在计算机可读存储介质中还可以存储诸如输入信号、信号分量、噪声分量等各种内容。

[0202] 在一个示例中,电子设备10还可以包括:输入装置113和输出装置114,这些组件通过总线系统和/或其他形式的连接机构(未示出)互连。

[0203] 此外,该输入设备113还可以包括例如键盘、鼠标等等。

[0204] 该输出装置114可以向外部输出各种信息,包括确定出的距离信息、方向信息等。该输出装置114可以包括例如显示器、扬声器、打印机、以及通信网络及其所连接的远程输出设备等等。

[0205] 当然,为了简化,图10中仅示出了该电子设备10中与本公开有关的组件中的一些,省略了诸如总线、输入/输出接口等等的组件。除此之外,根据具体应用情况,电子设备10还可以包括任何其他适当的组件。

[0206] 示例性计算机程序产品和计算机可读存储介质

[0207] 除了上述方法和设备以外,本申请的实施例还可以是计算机程序产品,其包括计算机程序指令,计算机程序指令在被处理器运行时使得处理器执行本说明书上述“示例性方法”部分中描述的根据本申请各种实施例的运动状态估计模型的训练方法、运动状态估计方法中的步骤。

[0208] 计算机程序产品可以以一种或多种程序设计语言的任意组合来编写用于执行本申请实施例操作的程序代码,程序设计语言包括面向对象的程序设计语言,诸如Java、C++等,还包括常规的过程式程序设计语言,诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。

[0209] 此外,本申请的实施例还可以是计算机可读存储介质,其上存储有计算机程序指令,计算机程序指令在被处理器运行时使得处理器执行本说明书上述“示例性方法”部分中描述的根据本申请各种实施例的运动状态估计模型的训练方法、运动状态估计方法中的步骤。

[0210] 计算机可读存储介质可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以包括但不限于电、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。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(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。

[0211] 以上结合具体实施例描述了本申请的基本原理,但是,需要指出的是,在本申请中提及的优点、优势、效果等仅是示例而非限制,不能认为这些优点、优势、效果等是本申请的各个实施例必须具备的。另外,上述公开的具体细节仅是为了示例的作用和便于理解的作用,而非限制,上述细节并不限制本申请为必须采用上述具体的细节来实现。

[0212] 本申请中涉及的器件、装置、设备、系统的方框图仅作为例示性的例子并且不意图要求或暗示必须按照方框图示出的方式进行连接、布置、配置。如本领域技术人员将认识到的,可以按任意方式连接、布置、配置这些器件、装置、设备、系统。诸如“包括”、“包含”、“具有”等等的词语是开放性词汇,指“包括但不限于”,且可与其互换使用。这里所使用的词汇“或”和“和”指词汇“和/或”,且可与其互换使用,除非上下文明确指示不是如此。这里所使用的词汇“诸如”指词组“诸如但不限于”,且可与其互换使用。

[0213] 还需要指出的是,在本申请的装置、设备和方法中,各部件或各步骤是可以分解和/或重新组合的。这些分解和/或重新组合应视为本申请的等效方案。

[0214] 提供所公开的方面的以上描述以使本领域的任何技术人员能够做出或者使用本申请。对这些方面的各种修改对于本领域技术人员而言是非常显而易见的,并且在此定义的一般原理可以应用于其他方面而不脱离本申请的范围。因此,本申请不意图被限制到在此示出的方面,而是按照与在此公开的原理和新颖的特征一致的最宽范围。

[0215] 为了例示和描述的目的已经给出了以上描述。此外,此描述不意图将本申请的实施例限制到在此公开的形式。尽管以上已经讨论了多个示例方面和实施例,但是本领域技术人员将认识到其某些变型、修改、改变、添加和子组合。

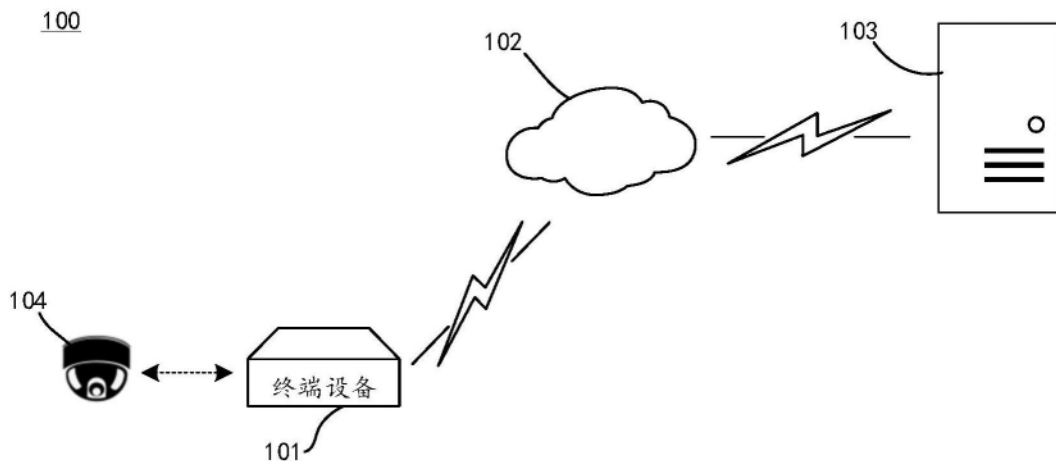


图1

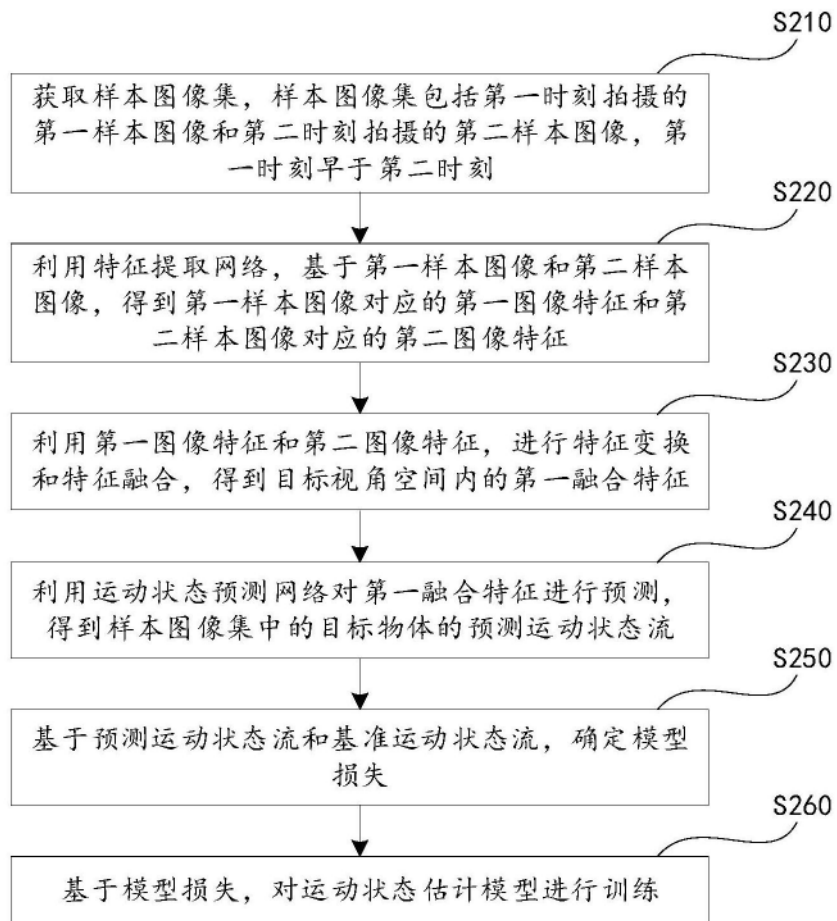


图2

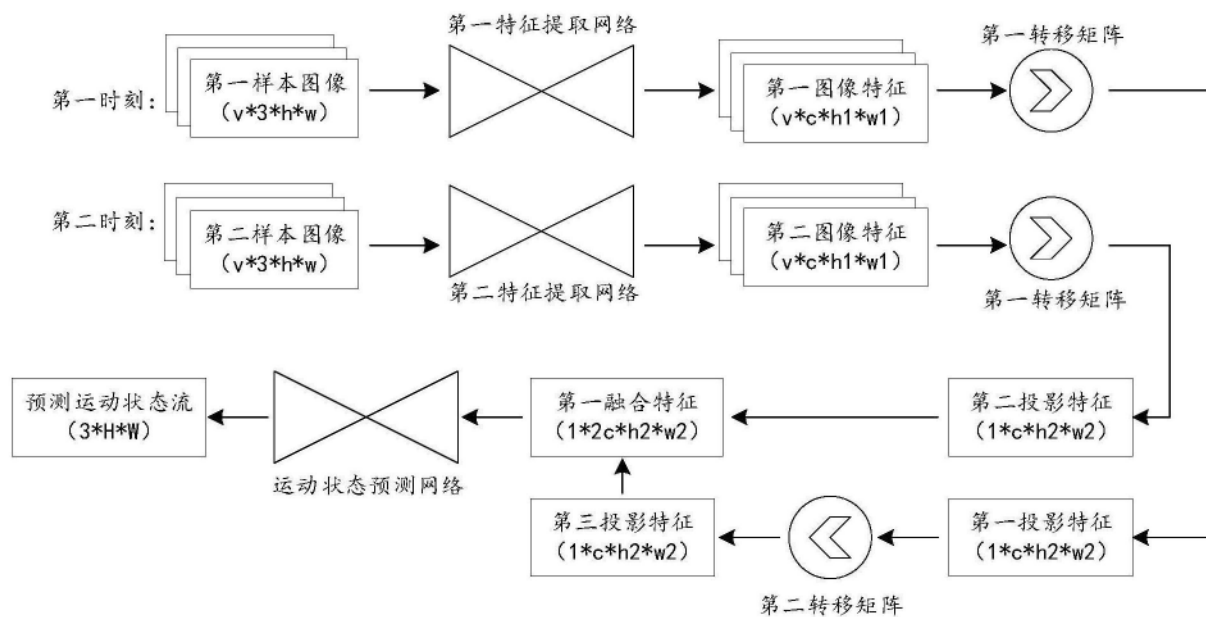


图3

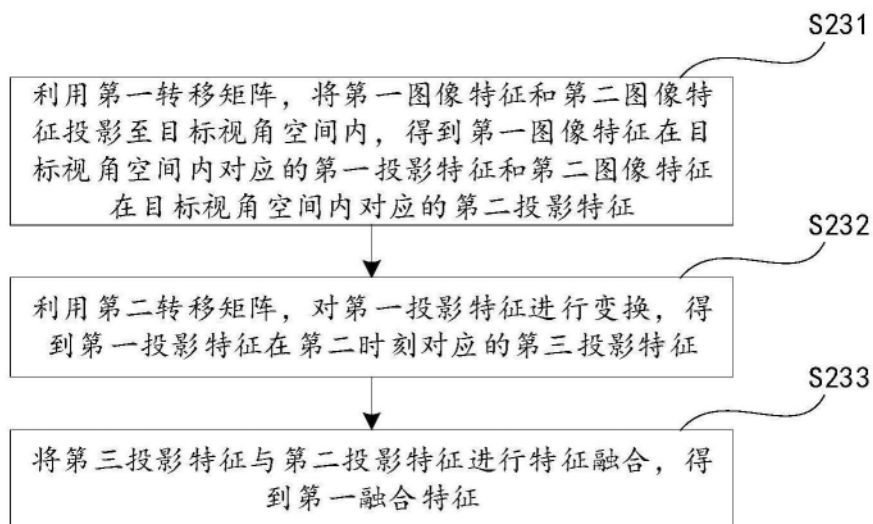


图4

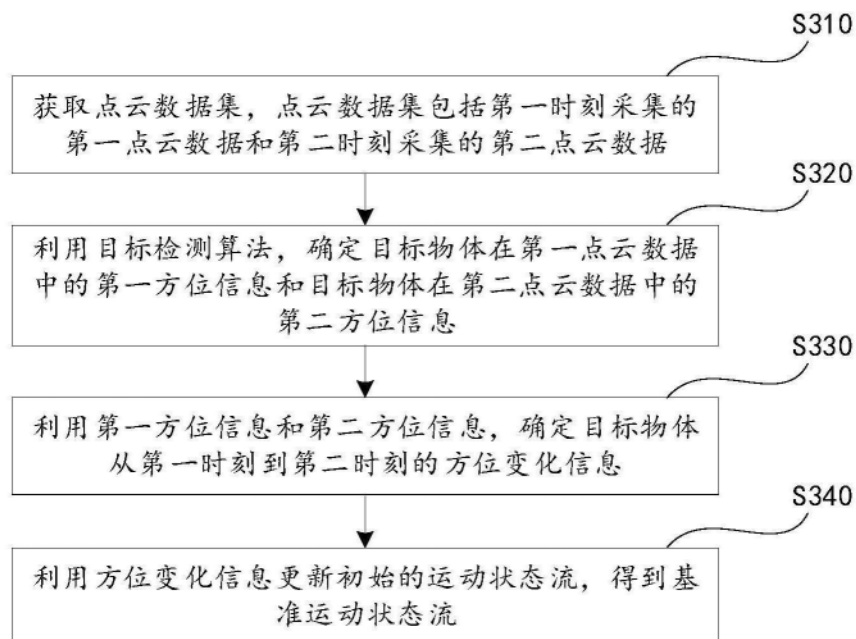


图5

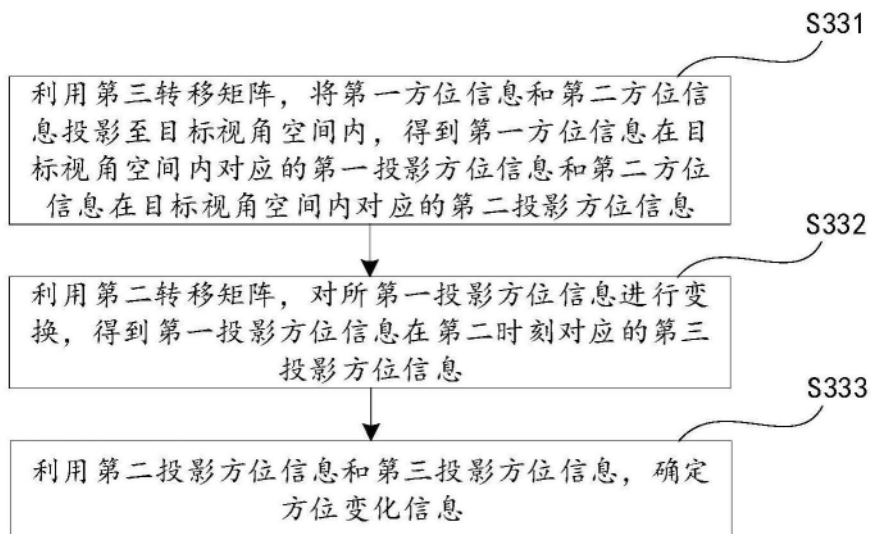


图6

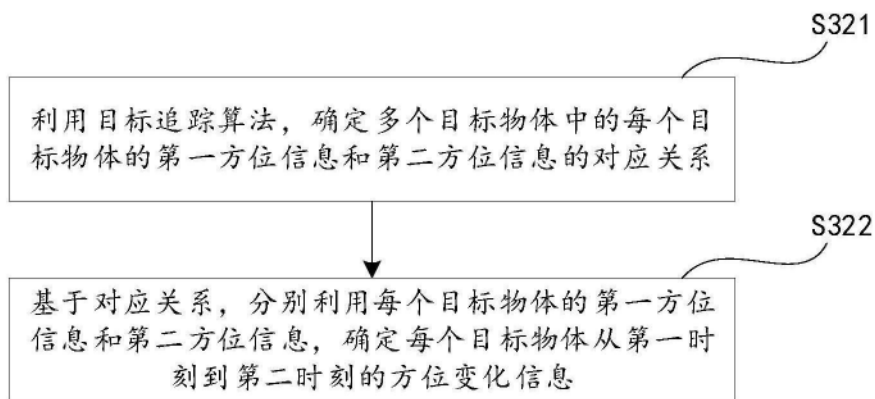


图7

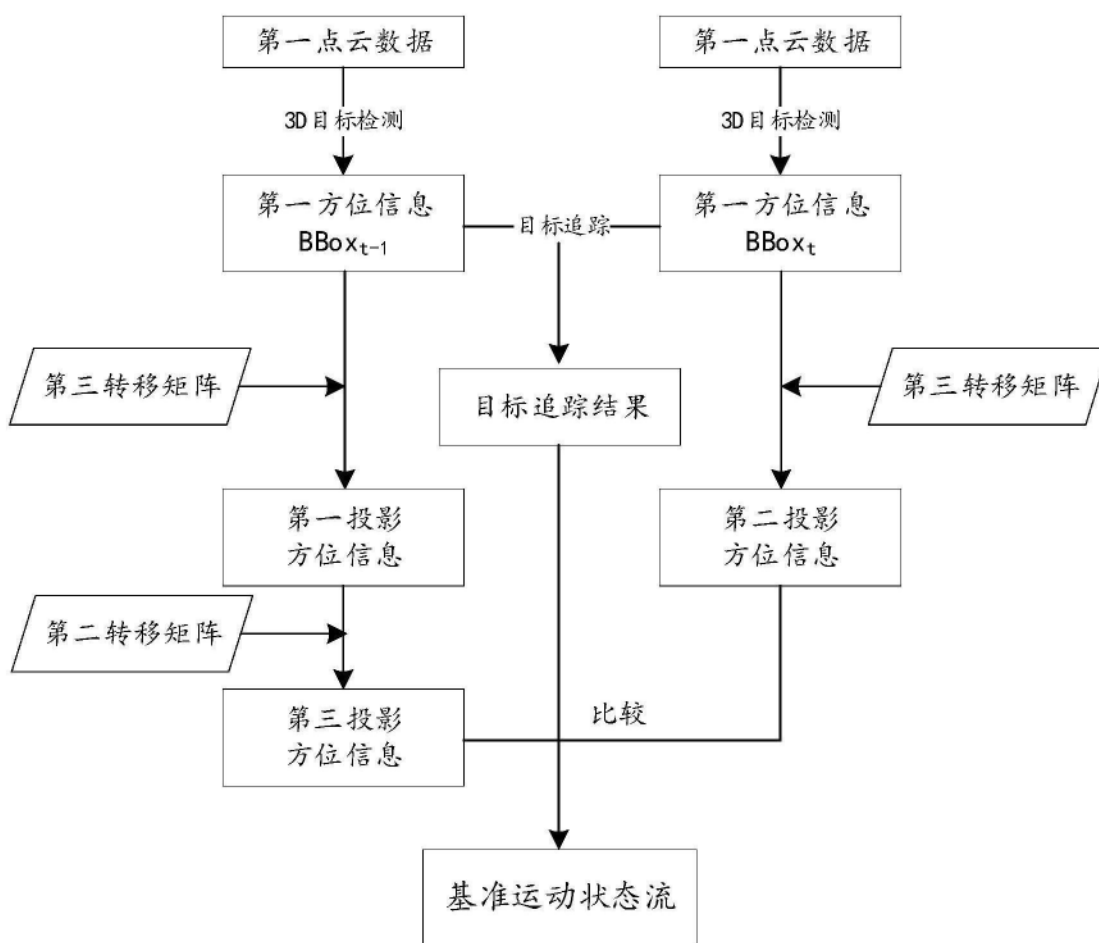


图8

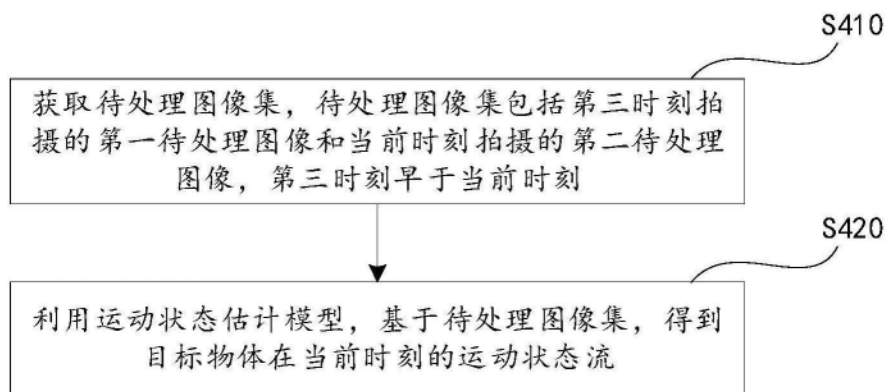


图9

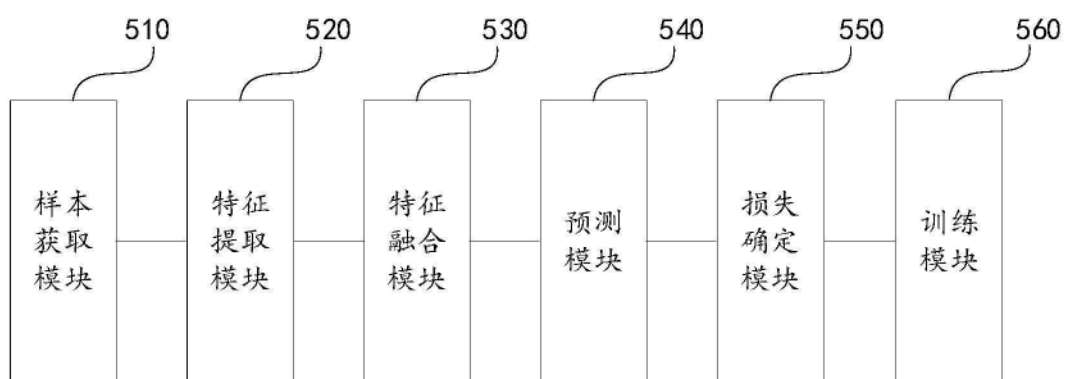


图10

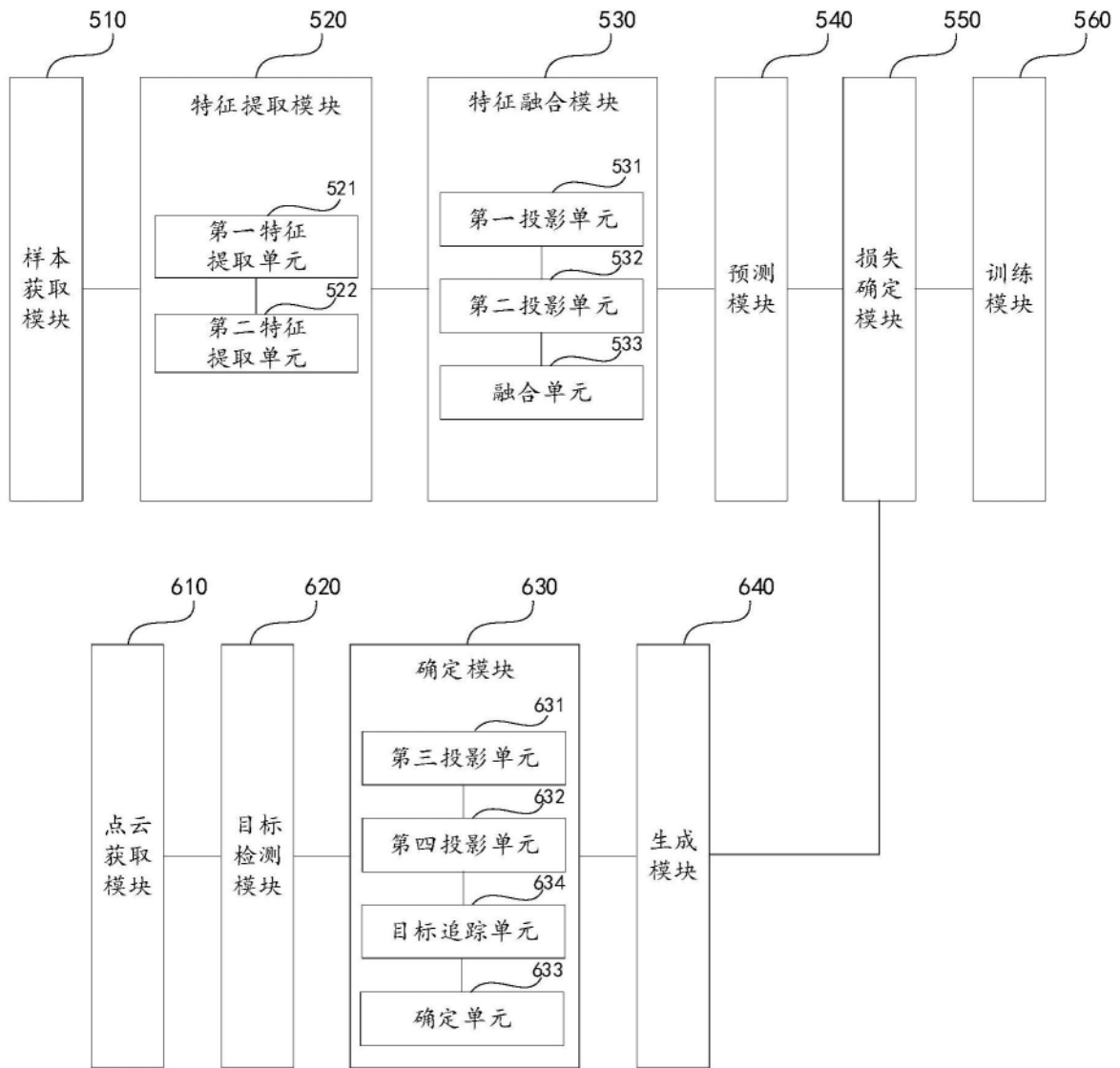


图11

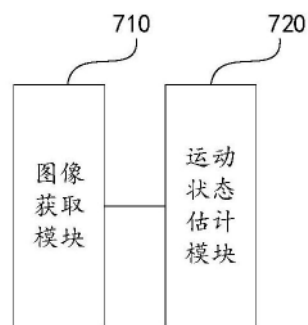


图12

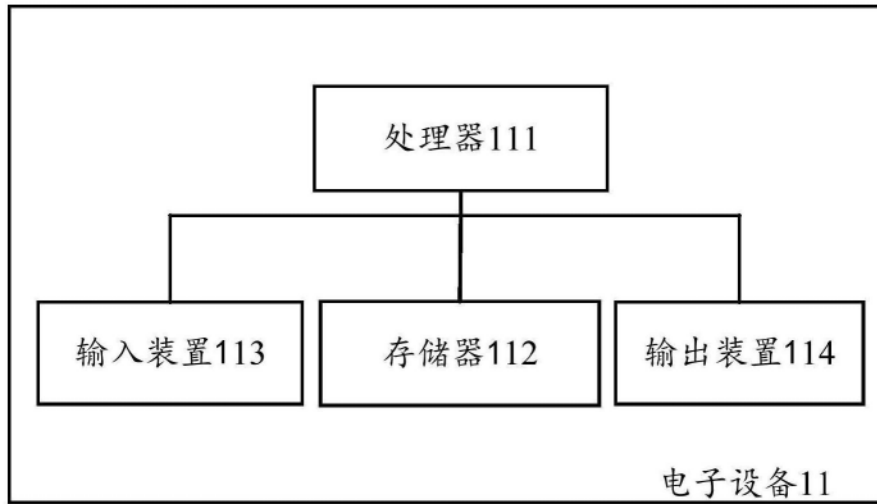


图13