(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112534475 B
(45)授权公告日 2023.01.10

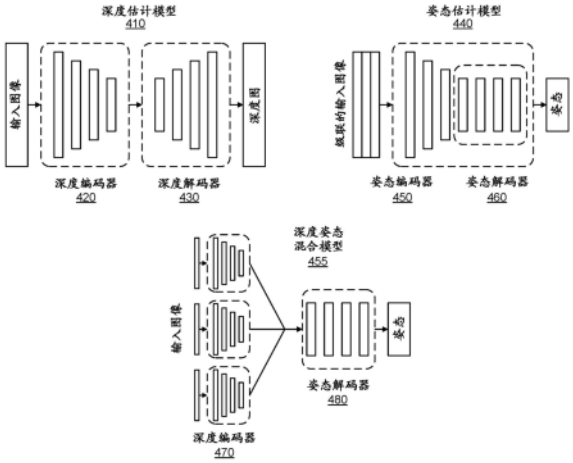
(21)申请号 201980047649.X
(22)申请日 2019.05.16
(65)同一申请的已公布的文献号
申请公布号 CN 112534475 A
(43)申请公布日 2021.03.19
(30)优先权数据
62/673,045 2018.05.17 US
(85)PCT国际申请进入国家阶段日
2021.01.15
(86)PCT国际申请的申请数据
PCT/US2019/032616 2019.05.16
(87)PCT国际申请的公布数据
WO2019/222467 EN 2019.11.21
(73)专利权人 奈安蒂克公司
地址 美国加利福尼亚州
(72)发明人 C•戈达德 O•麦克•奥达
G•J•布罗斯托 M•菲尔曼
(74)专利代理机构 北京市金杜律师事务所 11256
专利代理师 黄倩 杨飞
(51)Int.Cl.
*G06T 7/593*(2017.01)
*G06T 7/00*(2017.01)
*G06T 7/30*(2017.01)
(56)对比文件
CN 107438866 A,2017.12.05
JP 2015087851 A,2015.05.07
Tinghui Zhou 等.Unsupervised Learning of Depth and Ego-Motion from Video".《2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)》.2017,摘要,第2-3,5节.
审查员 巩瑜
权利要求书4页 说明书16页 附图7页

(54)发明名称

深度估计系统的自我监督训练

(57)摘要

描述了一种用于训练深度估计模型的方法及其使用方法。图像被获取并且被输入到深度模型中,以基于深度模型的参数提取针对多个图像中的每个的深度图。该方法包括将图像输入到姿态解码器中,以提取针对每个图像的姿态。该方法包括基于针对每个图像的深度图和姿态,生成多个合成帧。该方法包括基于合成帧和图像的比较利用输入比例遮挡和运动感知损失函数计算损失值。该方法包括基于损失值调节深度模型的多个参数。经训练模型可以接收场景的图像并且根据图像生成场景的深度图。




CN 112534475 B

1.一种非暂态计算机可读存储介质,存储:

经训练深度模型,所述经训练深度模型通过如下的过程而被获得,所述过程包括:

获取图像的集合,所述图像的集合包括在第一时间戳处的第一图像、在第二时间戳处的第二图像和在第三时间戳处的第三图像;

应用所述深度模型,以基于所述深度模型的参数生成针对所述第一图像和所述第三图像的深度图;

将所述第一图像和所述第三图像应用于姿态解码器,以生成针对所述第一图像和所述第三图像的姿态;

基于针对所述第一图像和所述第三图像的所述深度图和所述姿态,生成合成帧,其中第一合成帧是基于针对所述第一图像的所述深度图和所述姿态而针对所述第二时间戳生成的,并且第二合成帧是基于针对所述第三图像的所述深度图和所述姿态而针对所述第二时间戳生成的;

基于所述合成帧和所述第二图像的比较,利用输入比例遮挡感知和运动感知损失函数计算损失值,其中计算所述损失值包括:

计算所述第一合成帧与所述第二图像之间的第一差异以及所述第二合成帧与所述第二图像之间的第二差异;以及

标识所述第一差异与所述第二差异之间的最小差异,其中所述损失值基于所述最小差异;以及

基于所述损失值,调节所述深度模型的所述参数;以及

指令,所述指令在由计算设备执行时,使得所述计算设备将所述经训练深度模型应用于场景的图像,以生成所述场景的深度图。

2.根据权利要求1所述的非暂态计算机可读存储介质,其中所述图像的集合包括单眼视频,所述单眼视频的每个图像是在对应时间戳处捕获的,其中生成所述合成帧包括:

对于所述单眼视频的图像,在相邻时间戳处生成合成帧。

3.根据权利要求2所述的非暂态计算机可读存储介质,其中计算所述第一差异以及所述第二差异是以下过程的一部分,所述过程包括:

对于所述单眼视频的图像,计算所生成的合成帧与具有匹配时间戳的图像之间的差异。

4.根据权利要求1所述的非暂态计算机可读存储介质,其中:

计算所述第一差异以及所述第二差异是以下过程的一部分,所述过程包括:计算所述第一合成帧与所述第二图像之间的第一差异集合;并且所述第二差异是所述第二合成帧与所述第二图像之间的第二差异集合的一部分;以及

标识所述最小差异是以下过程的一部分,所述过程包括:对于所述第二图像的每个像素,标识所述第一差异集合与所述第二差异集合之间的最小差异,其中所述损失值基于所述最小差异。

5.根据权利要求3所述的非暂态计算机可读存储介质,其中利用所述输入比例遮挡感知和运动感知损失函数计算所述损失值还包括:

标识所述单眼视频中的一个或多个静态特征,

其中所述损失值基于滤除所述一个或多个静态特征的所述差异。

6.根据权利要求5所述的非暂态计算机可读存储介质，其中标识所述单眼视频中的一个或多个静态特征包括：

标识在所述单眼视频的第一时间戳处的第一图像中、并且在所述单眼视频的第二时间戳处的第二图像中的区域；

确定对象是否在所述第一图像与所述第二图像之间处于相似位置；以及

响应于确定所述对象在所述第一图像与所述第二图像之间处于相似位置，将所述区域定义为静态特征。

7.根据权利要求1所述的非暂态计算机可读存储介质，其中所述图像的集合处于第一分辨率，并且所生成的深度图处于低于所述第一分辨率的第二分辨率，并且其中所述获得所述深度模型的过程还包括：

将所生成的深度图从所述第二分辨率上采样成所述第一分辨率。

8.根据权利要求1所述的非暂态计算机可读存储介质，其中所述图像的集合包括立体图像对，其中每个立体图像对是由立体相机对捕获的，其中生成所述合成帧包括：

对于每个立体图像对，基于所述立体图像对的第一图像生成合成帧。

9.根据权利要求8所述的非暂态计算机可读存储介质，其中基于所述损失值调节所述深度模型的所述参数包括：

对于每个立体图像对，计算所生成的合成帧与所述立体图像对的第二图像之间的差异；以及

调节所述参数以使所述差异最小化。

10.一种计算机实现的方法，包括：

接收场景的图像；

将所述图像输入到经训练深度模型中，所述深度模型是利用过程而训练的，所述过程包括：

获取图像的集合，所述图像的集合包括在第一时间戳处的第一图像、在第二时间戳处的第二图像和在第三时间戳处的第三图像；

将所述第一图像和所述第三图像输入到所述深度模型中，以基于所述深度模型的参数提取针对所述第一图像和所述第三图像的深度图；

将所述第一图像和所述第三图像输入到姿态解码器中，以提取针对所述第一图像和所述第三图像的姿态；

基于针对所述第一图像和所述第三图像的所述深度图和所述姿态，生成合成帧，其中第一合成帧是基于针对所述第一图像的所述深度图和所述姿态而针对所述第二时间戳生成的，并且第二合成帧是基于针对所述第三图像的所述深度图和所述姿态而针对所述第二时间戳生成的；

基于所述合成帧和所述第二图像的比较，利用输入比例遮挡感知和运动感知损失函数计算损失值，其中计算所述损失值包括：

计算所述第一合成帧与所述第二图像之间的第一差异以及所述第二合成帧与所述第二图像之间的第二差异；以及

标识所述第一差异与所述第二差异之间的最小差异，其中所述损失值基于所述最小差异；以及

基于所述合成帧和所述图像的所述比较，调节所述深度模型的所述参数；以及

由经训练模型生成与所述场景的所述图像相对应的所述场景的深度图。

11.根据权利要求10所述的方法，其中所述图像的集合包括单眼视频，其中所述单眼视频的每个图像是在对应时间戳处捕获的，其中生成所述合成帧包括：

对于所述单眼视频的图像，在相邻时间戳处生成合成帧。

12.根据权利要求11所述的方法，其中计算所述第一差异以及所述第二差异是以下过程的一部分，所述过程包括：

对于所述单眼视频的图像，计算所生成的合成帧与具有匹配时间戳的图像之间的差异。

13.根据权利要求10所述的方法，其中：

计算所述第一差异以及所述第二差异是以下过程的一部分，所述过程包括：计算所述第一合成帧与所述第二图像之间的第一差异集合；并且所述第二差异是所述第二合成帧与所述第二图像之间的第二差异集合的一部分；以及

标识所述最小差异是以下过程的一部分，所述过程包括：对于所述第二图像的每个像素，标识所述第一差异集合与所述第二差异集合之间的最小差异，其中所述损失值基于所述最小差异。

14.根据权利要求12所述的方法，其中利用所述输入比例遮挡感知和运动感知损失函数计算所述损失值还包括：

标识所述单眼视频中的一个或多个静态特征，其中所述损失值基于滤除所述一个或多个静态特征的所述差异。

15.根据权利要求14所述的方法，其中标识所述单眼视频中的一个或多个静态特征包括：

标识在所述单眼视频的第一时间戳处的第一图像中、并且在所述单眼视频的第二时间戳处的第二图像中的区域；

确定对象在所述第一图像与所述第二图像之间是否处于相似位置；以及

响应于确定所述对象在所述第一图像与所述第二图像之间处于相似位置，将所述区域定义为静态特征。

16.根据权利要求10所述的方法，其中所述图像的集合处于第一分辨率，并且所提取的深度图处于低于所述第一分辨率的第二分辨率，用于训练所述深度模型的所述过程还包括：

将所提取的深度图从所述第二分辨率上采样成所述第一分辨率。

17.根据权利要求10所述的方法，其中所述图像的集合包括立体图像对，其中每个立体图像对是由立体相机对捕获的，其中生成所述合成帧包括：

对于每个立体图像对，基于所述立体图像对的第一图像生成合成帧。

18.根据权利要求17所述的方法，其中基于所述损失值调节所述深度模型的所述参数包括：

对于每个立体图像对，计算所生成的合成帧与所述立体图像对的第二图像之间的差异；以及

调节所述参数以使所述差异最小化。

19.根据权利要求10所述的方法,还包括:

基于所述场景的所述深度图,显示利用虚拟内容增强的所述场景的所述图像。

20.一种存储有指令的非暂态计算机可读存储介质,所述指令在由计算设备执行时使得所述计算设备执行操作,所述操作包括:

接收场景的图像;

将所述图像输入到经训练深度模型中,所述深度模型是利用过程而训练的,所述过程包括:

获取图像的集合,所述图像的集合包括在第一时间戳处的第一图像、在第二时间戳处的第二图像和在第三时间戳处的第三图像;

将所述第一图像和所述第三图像输入到深度编码器中,以基于所述深度编码器的参数提取针对所述第一图像和所述第三图像的抽象深度特征;

对针对所述第一图像和所述第三图像的所述抽象深度特征进行级联;

将所级联的所述抽象深度特征输入到姿态解码器中,以提取针对所述第一图像和所述第三图像的姿态;

基于所述模型的参数和针对所述第一图像和所述第三图像的所述姿态,生成合成帧,其中第一合成帧是基于针对所述第一图像的深度图和姿态而针对所述第二时间戳生成的,并且第二合成帧是基于针对所述第三图像的深度图和姿态而针对所述第二时间戳生成的;

基于所述合成帧和所述第二图像的比较,利用输入比例遮挡感知和运动感知损失函数计算损失值,其中计算所述损失值包括:

计算所述第一合成帧与所述第二图像之间的第一差异以及所述第二合成帧与所述第二图像之间的第二差异;

标识所述第一差异与所述第二差异之间的最小差异,其中所述损失值基于所述最小差异;以及

基于所述合成帧和所述图像的比较,调节所述模型的所述参数;以及

由经训练模型生成与所述场景的所述图像相对应的所述场景的深度图。

# 深度估计系统的自我监督训练

## 技术领域

[0001] 所描述的主题总体上涉及从单色输入图像估计深度图，并且具体地涉及使用视频数据和/或立体图像数据而训练的用于估计深度图的机器学习模型。

## 背景技术

[0002] 深度感测在导航和场景理解方面都有应用。很多方法使用经训练模型或网络来从单色输入图像确定深度图。有几种方法使用不同种类的训练数据来训练深度估计系统。可以使用检测和测距系统来训练深度估计系统，以为环境中的对象建立地面真实深度(即，无线电检测和测距(RADAR)、光检测和测距(LIDAR)等)，该环境与由相机获取的同一场景的图像配对。尽管检测和测距系统可以提供对象深度的地面真相，但是不断利用检测和测距系统来感测很多不同环境的深度可能是一种在时间和资源方面昂贵的尝试。此外，检测和测距系统不能确定一些对象(例如，反射对象)的深度，这些对象的材料性质可能使得检测和测距系统无法检测到它们。

[0003] 训练深度估计系统的另一种方法是利用同一场景的立体图像对。在单个时刻捕获立体图像对取决于使用两个相机，这两个相机聚焦在同一场景上但相距一定距离。深度估计系统通过从立体图像对中的一个立体图像投影到另一立体图像来进行操作。为了彼此投影，除了用于获取立体图像对的两个相机的物理位置之间的视差(深度的缩小倒数scaled inverse of depth)和相对变换，深度估计系统还要考虑当前立体图像。在与所捕获的立体图像相比使投影的光度重构误差(photometric reconstruction error)最小化时，深度估计系统可以确定场景的深度。

[0004] 一些更新颖的方法利用不断变化的场景的单眼视频数据来训练深度估计系统。深度估计系统通过以下方式进行训练：将单眼视频数据中的一个时间图像(temporal image)投影到后续时间图像，同时使光度重构误差最小化。但是，从一个时间图像到另一时间图像，这样的系统可能会错误地解释进入或离开视野的对象，这会导致深度图和深度图边界附近出现伪影。此外，在对深度图进行上采样之前，传统系统当前输入用于确定深度图的低分辨率图像，这易于产生深度上采样伪影。

## 发明内容

[0005] 本公开描述了一种用于训练和使用深度估计模型的方法。为了训练模型，系统获取图像。系统将图像输入到深度模型中，以基于深度模型的参数提取针对每个图像的深度图。系统将图像输入到姿态解码器中，以提取针对每个图像的姿态。系统基于针对每个图像的深度图和姿态，生成合成帧。系统基于合成帧和图像的比较，利用输入比例遮挡和运动感知损失函数计算损失值。输入比例遮挡和运动感知损失函数计算损失值以用于优化深度模型的参数。损失函数包括计算合成帧与输入图像之间每个像素的光度重构误差。损失函数还可以考虑从来自单眼视频的多个时间图像投影的两个合成帧之间的最小光度重构误差，这些时间图像在时间上相邻。在合成帧的生成期间也可以使用上采样深度特征，这将影响

外观匹配损失的计算。损失函数还可以实现所生成的掩蔽，该掩蔽在计算损失值时减少图像中的静态特征。系统基于损失值调节深度模型的参数。利用经训练模型，设备可以接收场景的图像并且根据图像生成场景的深度图。

[0006] 在一些实施例中，深度估计模型利用包括单眼视频的图像数据而进行训练。单眼视频的每个图像在一个不同的时间被捕获，并且与对应时间戳相关联。在使用具有第一时间戳的第一图像和具有第二时间戳的第二图像的示例讨论中，模型使用来自单眼视频的具有第一时间戳的第一图像来在第二时间戳处生成合成帧。该模型计算合成帧与具有第二时间戳的第二图像之间的光度重构误差。该模型遵循上述过程，其中来自单眼视频的其他图像对具有相邻时间戳。该模型调节参数以使误差最小化。在其他实施例中，该模型利用包括立体图像对的图像数据进行训练，其中每个立体图像对由立体相机对捕获。当生成合成帧时，该模型从立体图像对获取图像之一(例如，左图像)，并且在另一图像(例如，右图像)处生成合成帧。该模型计算合成帧与其他图像之间的光度重构误差。该模型调节参数以使误差最小化。在其他实施例中，该模型利用包括单眼视频和立体图像对的图像数据而进行训练。

## 附图说明

[0007] 图1示出了根据一个或多个实施例的联网计算环境。

[0008] 图2描绘了根据一个或多个实施例的具有与真实世界平行的地理环境的虚拟世界的表示。

[0009] 图3描绘了根据一个或多个实施例的并行现实游戏的示例性游戏界面。

[0010] 图4示出了根据一个或多个实施例的使用单独的姿态估计模型与深度-姿态混合模型的概念比较。

[0011] 图5是描述根据一个或多个实施例的训练深度估计模型的一般过程的流程图。

[0012] 图6是描述根据一个或多个实施例的使用深度估计模型的一般过程的流程图。

[0013] 图7是根据一个或多个实施例的计算设备的示例架构。

[0014] 附图和以下描述仅通过说明的方式描述了某些实施例。本领域技术人员将从下面的描述中容易地认识到，在不脱离所描述的原理的情况下，可以采用结构和方法的替代实施例。现在将参考几个实施例，其示例在附图中示出。

## 具体实施方式

[0015] 示例性的基于位置的并行现实游戏系统(parallel reality gaming system)

[0016] 并行现实游戏是具有虚拟世界地理环境的基于位置的游戏，该虚拟现实地理环境与真实世界地理环境的至少一部分平行，使得玩家在真实世界中的移动和动作影响在虚拟世界中的动作，反之亦然。使用本文中提供的公开内容的本领域普通技术人员将理解：所描述的主题适用于其他情况，在这些情况中，期望根据图像数据确定深度信息。另外，基于计算机的系统的固有灵活性允许在系统的组件之间进行任务和功能的多种可能的配置、组合和划分。例如，根据本公开的各方面的系统和方法可以使用单个计算设备或跨多个计算设备(例如，在计算机网络中连接)来实现。

[0017] 图1示出了根据一个或多个实施例的联网计算环境100。联网计算环境100提供了

虚拟世界中的玩家的交互，该虚拟世界具有与真实世界平行的地理环境。特别地，真实世界中的地理区域可以被直接链接或映射到虚拟世界中的对应区域。通过移动到真实世界中的各个地理位置，玩家可以在虚拟世界中移动。例如，玩家在真实世界中的位置可以被跟踪并且用于更新玩家在虚拟世界中的位置。通常，玩家在真实世界中的位置是通过如下的方式而被确定：找到客户端设备110(玩家正在通过该客户端设备110与虚拟世界交互的)的位置，并且假定玩家处于与该客户端设备110相同(或近似相同)的位置。例如，在各种实施例中，如果玩家在真实世界中的位置在真实世界位置(该真实世界位置与虚拟元素在虚拟世界中的虚拟位置相对应)的阈值距离(例如，十米、二十米等)之内，则玩家可以与虚拟元素交互。为了方便起见，参考“玩家的位置”描述各种实施例，但是本领域技术人员将理解，这样的引用可以是指玩家的客户端设备110的位置。

[0018] 现在参考图2，图2描绘了根据一个实施例的与真实世界200(该真实世界200可以充当并行现实游戏的玩家的游戏板)平行的虚拟世界210的概念图。如图所示，虚拟世界210可以包括与真实世界200的地理环境平行的地理环境。具体地，定义真实世界200中的地理区域或空间的坐标范围被映射到用于定义虚拟世界210中的虚拟空间的对应坐标范围。真实世界200中的坐标范围可以与城镇、街区、城市、校园、场所、国家、大陆、整个地球或其他地理区域相关联。地理坐标范围内的每个地理坐标被映射到虚拟世界中的虚拟空间中的对应坐标。

[0019] 玩家在虚拟世界210中的位置与在玩家真实世界200中的位置相对应。例如，位于真实世界200中的位置212的玩家A在虚拟世界210中具有对应位置222。类似地，位于真实世界中的位置214的玩家B在虚拟世界中具有对应位置224。当玩家在真实世界中的地理坐标范围内移动时，玩家也在定义虚拟世界210中的虚拟空间的坐标范围内移动。特别地，当玩家在真实世界中的地理坐标范围内导航时，与由玩家携带的移动计算设备相关联的定位系统(例如，GPS系统)可以用来跟踪玩家的位置。与玩家在真实世界200中的位置相关联的数据用于更新玩家在定义虚拟世界210中的虚拟空间的对应坐标范围内的位置。以这种方式，通过简单地在真实世界200中的对应地理坐标范围内行进，玩家就可以沿着定义虚拟世界210中的虚拟空间的坐标范围内的连续轨迹进行导航，而无需在真实世界200中的特定离散位置处报到或周期性地更新位置信息。

[0020] 基于位置的游戏可以包括多个游戏目标，该游戏目标要求玩家行进到分散在虚拟世界中的各个虚拟位置处的各个虚拟元素和/或虚拟对象，和/或与之交互。玩家可以通过行进到虚拟元素或对象在真实世界中的对应位置来行进到这些虚拟位置。例如，定位系统可以连续地跟踪玩家的位置，使得随着玩家在真实世界中连续导航，玩家也在并行虚拟世界中连续导航。然后，玩家可以在特定位置与各种虚拟元素和/或对象交互，以实现或执行一个或多个游戏目标。

[0021] 例如，游戏目标具有如下的玩家，这些玩家与位于虚拟世界210中的各个虚拟位置处的虚拟元素230交互。这些虚拟元素230可以链接到真实世界200中的地标、地理位置或对象240。真实世界地标或对象240可以是艺术品、纪念碑、建筑物、企业、图书馆、博物馆或其他合适的真实世界地标或对象。交互包括捕获、主张所有权、使用某个虚拟物品、花费一些虚拟货币等。为了捕获这些虚拟元素230，玩家必须行进到真实世界中的被链接到虚拟元素230的地标或地理位置240，并且必须执行与虚拟世界210中的虚拟元素230的任何必要的交

互。例如,图2中的玩家A可能必须行进到真实世界200中的地标240,以便与虚拟元素230(其与该特定地标240链接)交互、或捕获该虚拟元素230。与虚拟元素230的交互可能需要真实世界中的动作,诸如获取照片和/或验证、获取或捕获关于与虚拟元素230相关联的地标或对象240的其他信息。

[0022] 游戏目标可能要求玩家使用由玩家在基于位置的游戏中收集的一个或多个虚拟物品。例如,玩家可以在虚拟世界210中行进,以寻找对完成游戏目标有用的虚拟物品(例如,武器、生物、道具或其他物品)。通过行进到真实世界200中的不同位置、或通过在虚拟世界210或真实世界200中完成各种动作,可以找到或收集这些虚拟物品。在图2所示的示例中,玩家使用虚拟物品232来捕获一个或多个虚拟元素230。特别地,玩家可以将虚拟物品232部署在虚拟世界210中的靠近虚拟元素230的位置处、或在虚拟元素230中的位置处。以这种方式部署一个或多个虚拟物品232可以导致针对特定玩家或针对特定玩家的团队/阵营的虚拟元素230的捕获。

[0023] 在一个特定实现中,玩家可能必须收集作为并行现实游戏的一部分的虚拟能量。如图2所示,虚拟能量250可以散布在虚拟世界210中的不同位置处。玩家可以通过进行到虚拟能量250的在真实世界200中的对应位置来收集虚拟能量250。虚拟能量250可以用于为虚拟物品供能和/或在游戏中执行各种游戏目标。失去所有虚拟能量250的玩家可能与游戏断开连接。

[0024] 根据本公开的各方面,并行现实游戏可以是大型多玩家的基于位置的游戏,其中游戏中的每个参与者共享同一虚拟世界。玩家可以分为不同的团队或阵营,并且可以共同努力实现一个或多个游戏目标,诸如捕获或主张虚拟元素的所有权。以这种方式,并行现实游戏本质上可以是鼓励游戏内玩家之间合作的社交游戏。在并行现实游戏期间,来自对方团队的玩家可以互相对抗(或有时合作以达到共同的目标)。玩家可以使用虚拟物品攻击或阻碍对方团队的玩家的前进。在某些情况下,鼓励玩家聚集在真实世界的多个位置处,以进行并行现实游戏中的合作或互动事件。在这些情况下,游戏服务器试图确保玩家确实在场并且没有欺骗。

[0025] 并行现实游戏可以具有各种特征以增强和鼓励并行现实游戏中的游戏玩法。例如,玩家可以累积在整个游戏中可以使用(例如,以购买游戏中物品,兑换其他物品,制作物品,等等)的虚拟货币或其他虚拟奖励(例如,虚拟代币、虚拟积分、虚拟材料资源等)。随着玩家完成一个或多个游戏目标并且在游戏中获取经验,玩家可以前进到各个级别。在一些实施例中,玩家可以通过在游戏中提供的一个或多个通信接口彼此通信。玩家还可以获取增强的“力量”或虚拟物品,这些力量”或虚拟物品可以用于完成游戏中的游戏目标。使用本文中提供的公开内容的本领域普通技术人员应当理解,在不脱离本公开的范围的情况下,并行现实游戏可以包括各种其他游戏特征。

[0026] 再次参考图1,联网计算环境100使用客户端服务器架构,其中游戏服务器120通过网络105与客户端设备110通信,以向客户端设备110处的玩家提供并行现实游戏。联网计算环境100还可以包括其他外部系统,诸如赞助商/广告商系统或业务系统。尽管在图1中仅示出了一个客户端设备110,但是,任何数目的客户端110或其他外部系统可以通过网络105连接到游戏服务器120。此外,联网计算环境100可以包含不同或附加的元素,并且功能可以以与下文所述不同的方式分布在客户端设备110与服务器120之间。

[0027] 客户端设备110可以是玩家可以用来与游戏服务器120接口的任何便携式计算设备。例如，客户端设备110可以是无线设备、个人数字助理(PDA)、便携式设备。游戏设备、蜂窝电话、智能电话、平板电脑、导航系统、手持GPS系统、可穿戴计算设备、具有一个或多个处理器的显示器、或其他这样的设备。在另一实例中，客户端设备110包括常规计算机系统，诸如台式计算机或膝上型计算机。仍然，客户端设备110可以是具有计算设备的车辆。简而言之，客户端设备110可以是能够使得玩家能够与游戏服务器120交互的任何计算机设备或系统。作为计算设备，客户端设备110可以包括一个或多个处理器和一个或多个计算机可读存储介质。计算机可读存储介质可以存储引起处理器执行操作的指令。客户端设备110优选地是便携式计算设备，其可以与诸如智能电话或平板电脑等玩家容易地携带或以其他方式运输。

[0028] 客户端设备110与游戏服务器120通信，以向游戏服务器120提供物理环境的感测数据。客户端设备110包括相机组件125，该相机组件125在客户端设备110所在的物理环境中捕获场景的二维图像数据。客户端设备110还包括深度估计模型130，该深度估计模型130是例如由游戏服务器120训练的机器学习模型。在图1所示的实施例中，每个客户端设备110包括诸如游戏模块135和定位模块140等软件组件。客户端设备110可以包括各种其他输入/输出设备，以用于从玩家接收信息和/或向玩家提供信息。示例输入/输出设备包括适合于语音识别的显示屏、触摸屏、触摸板、数据输入键、扬声器和麦克风。客户端设备110还可以包括用于记录来自客户端设备110的数据的其他各种传感器，包括但不限于运动传感器、加速度计、陀螺仪、其他惯性测量单元(IMU)、气压计、定位系统、温度计、光传感器等。客户端设备110还可以包括用于通过网络105提供通信的网络接口。网络接口可以包括用于与一个或多个网络进行接口连接的任何合适的组件，包括例如发射器、接收器、端口、控制器、天线或其他合适的组件。

[0029] 相机组件125捕获客户端设备110所在的环境的场景的图像数据。相机组件125可以利用各种不同的光电传感器，这些光电传感器具有以不同的捕获率变化的颜色捕获范围。相机组件125可以包含广角镜头或远距镜头。相机组件125可以被配置为捕获单个图像或视频作为图像数据。另外，相机组件125的取向可以平行于地面，其中相机组件125对准地平线。相机组件125捕获图像数据并且与客户端设备110上的计算设备共享图像数据。图像数据可以附加有描述图像数据的其他细节的元数据，元数据包括感测数据(例如，温度、环境的亮度)或捕获数据(例如，曝光、暖度、快门速度、焦距、捕获时间等)。相机组件125可以包括可以捕获图像数据的一个或多个相机。在一个实例中，相机组件125包括一个相机，并且被配置为捕获单眼图像数据。在另一实例中，相机组件125包括两个相机，并且被配置为捕获立体图像数据。在各种其他实现中，相机组件125包括多个相机，每个相机被配置为捕获图像数据。

[0030] 深度估计模型130接收场景的输入图像，并且基于输入图像输出场景的深度。深度估计模型130由深度估计训练系统训练，并且可以由深度估计训练系统更新或调节，这将在下面更详细地讨论。所接收的输入图像可以由相机组件125的相机或另一客户端设备110的另一相机捕获。在一些实施例中，所接收的输入图像具有附加到图像的元数据，该元数据指定输入图像的固有特征。图像的固有特征是指在捕获图像时相机的一个或多个几何性质，例如，在捕获图像时相机的焦距、相机的主点偏移、相机的偏斜等。利用该固有特征，深度估

计模型130可以生成考虑该固有特征的本证矩阵(intrinsic matrix)。在一些实施例中,深度估计模型130确定输入图像是否高于阈值分辨率。如果否,则深度估计模型130可以在确定场景的深度图之前将输入图像上采样成期望分辨率。深度估计模型130输入图像(所接收的或在上采样之后)并且确定场景的深度图。机器学习算法可以在深度估计模型130中实现,以用于训练和/或推理。

[0031] 游戏模块135为玩家提供用于参与并行现实游戏的接口。游戏服务器120通过网络105向客户端设备110传输游戏数据,以供客户端设备110处的游戏模块135使用,以向远离游戏服务器120的位置处的玩家提供游戏的本地版本。游戏服务器120可以包括用于通过网络105提供通信的网络接口。网络接口可以包括用于与一个或多个网络进行接口连接的任何合适的组件,包括例如发射器、接收器、端口、控制器、天线或其他合适的组件。

[0032] 由客户端设备110执行的游戏模块135在玩家与并行现实游戏之间提供接口。游戏模块135可以在与客户端设备110相关联的显示设备上呈现用户界面,该用户界面显示与游戏相关联的虚拟世界(例如,渲染虚拟世界的图像)、并且允许用户在虚拟世界中进行交互以执行各种游戏目标。在一些其他实施例中,游戏模块135呈现来自真实世界的图像数据(例如,由相机组件125捕获的),该图像数据利用来自并行现实游戏的虚拟元素被增强。在这些实施例中,游戏模块135可以根据从客户端设备的其他组件接收的其他信息来生成虚拟内容、和/或调节虚拟内容。例如,根据在图像数据中捕获的场景的深度图(例如,由深度估计模型130确定的),游戏模块135可以调节虚拟对象以在用户界面上显示。

[0033] 游戏模块135还可以控制各种其他输出,以允许玩家与游戏交互,而无需玩家观看显示屏。例如,游戏模块135可以控制各种音频、振动或其他通知,这些通知允许玩家在不看显示屏的情况下玩游戏。游戏模块135可以访问从游戏服务器120接收的游戏数据,以向用户提供游戏的准确表示。游戏模块135可以接收和处理玩家输入、并且通过网络105向游戏服务器120提供更新。游戏模块135还可以生成和/或调节要由客户端设备110显示的游戏内容。例如,游戏模块135可以基于深度信息(例如,由深度估计模型130确定的)生成虚拟元素。

[0034] 定位模块140可以是用于监测客户端设备110位置的任何设备或电路系统。例如,定位模块140可以通过使用卫星导航定位系统(例如,GPS系统、伽利略定位系统、全球导航卫星系统(GLONASS)、北斗卫星导航和定位系统、惯性导航系统、航位推算系统、基于IP地址、通过使用三角测量和/或到蜂窝塔或Wi-Fi热点的接近度、和/或用于确定位置的其他合适的技术来确定实际或相对位置。定位模块140还可以包括各种其他传感器,其可以帮助准确地定位客户端设备110的位置。

[0035] 当玩家在真实世界中与客户端设备110一起移动时,定位模块140跟踪玩家的位置并且将玩家位置信息提供给游戏模块135。游戏模块135基于玩家在真实世界中的实际位置来更新与游戏相关联的虚拟世界中的玩家位置。因此,玩家可以简单地通过在真实世界中携带或运输客户端设备110来与虚拟世界交互。特别地,玩家在虚拟世界中的位置可以对应于玩家在真实世界中的位置。游戏模块135可以通过网络105向游戏服务器120提供玩家位置信息。作为响应,游戏服务器120可以制定各种技术来验证客户端设备110的位置,以防止作弊者欺骗客户端设备110的位置。应当理解,只有在已经向玩家通知要访问玩家的位置信息以及在游戏的上下文中将如何利用该位置信息(例如,以更新虚拟世界中的玩家位置)之

后，在给予许可的情况下，与玩家相关联的位置信息才被利用。另外，与玩家相关联的任何位置信息将以能够保护玩家隐私的方式来存储和维护。

[0036] 游戏服务器120可以是任何计算设备，并且可以包括一个或多个处理器和一个或多个计算机可读存储介质。计算机可读存储介质可以存储引起处理器执行操作的指令。游戏服务器120可以包括游戏数据库115或可以与之通信。游戏数据库115存储游戏数据，该游戏数据在并行现实游戏中被使用，以通过网络105而被供应或被提供给(多个)客户端120。

[0037] 存储在游戏数据库115中的游戏数据可以包括：(1)与并行现实游戏中的虚拟世界相关联的数据(例如，用于在显示设备上渲染虚拟世界的图像数据、虚拟世界中的位置的地理坐标等)；(2)与并行现实游戏的玩家相关联的数据(例如，玩家个人资料，包括但不限于玩家信息、玩家经验等级、玩家币种、虚拟世界/真实世界中的当前玩家位置、玩家能量等级、玩家偏好、团队信息、阵营信息等)；(3)与游戏目标相关联的数据(例如，与当前游戏目标、游戏目标的状态、过去游戏目标、未来游戏目标、期望游戏目标等相关联的数据)；(4)与虚拟世界中的虚拟元素相关联的数据(例如，虚拟元素的位置、虚拟元素的类型、与虚拟元素相关联的游戏目标；虚拟元素的对应实际世界位置信息；虚拟元素的行为、虚拟元素的相关性等)；(5)与真实世界对象、地标、链接到虚拟世界元素的位置相关联的数据(例如，真实世界对象/地标的位置、真实世界对象/地标的描述、链接到真实世界对象的虚拟元素的相关性等)；(6)游戏状态(例如，当前玩家人数、游戏目标的当前状态、玩家排行榜等)；(7)与玩家动作/输入相关联的数据(例如，当前玩家位置、过去玩家位置、玩家移动、玩家输入、玩家查询、玩家通信等)；以及(8)在并行现实游戏的实现期间所使用、所涉及或所获取的任何其他数据。存储在游戏数据库115中的游戏数据可以由系统管理员离线或实时填充、和/或通过从系统100的用户/玩家(诸如通过网络105从客户端设备110)接收的数据而被离线或实时填充。

[0038] 游戏服务器120可以被配置为从客户端设备110接收对游戏数据的请求(例如，经由远程过程调用(RPC))并且经由网络105响应于这些请求。例如，游戏服务器120可以将游戏数据编码在一个或多个数据文件中，并且将数据文件提供给客户端设备110。此外，游戏服务器120可以被配置为经由网络105从客户端设备110接收游戏数据(例如，玩家位置、玩家动作、玩家输入等)。例如，客户端设备110可以被配置为周期性地向游戏服务器120发送玩家输入和其他更新，游戏服务器120使用该玩家输入和其他更新来更新游戏数据库115中的游戏数据，以反映游戏的任何和所有已改变的条件。

[0039] 在所示的实施例中，服务器120包括通用游戏模块145、商业游戏模块150、数据收集模块155、事件模块160和深度估计训练系统170。如上所述，游戏服务器120与游戏数据库115交互，该游戏数据库115可以是游戏服务器120的一部分或可以被远程地访问(例如，游戏数据库115可以是经由网络105访问的分布式数据库)。在其他实施例中，游戏服务器120包含不同和/或附加的元素。另外，功能可以以与所描述的不同的方式在元素之间分配。例如，游戏数据库115可以被集成到游戏服务器120中。

[0040] 通用游戏模块145为所有玩家托管并行现实游戏，并且用作用于所有玩家的并行现实游戏的当前状态的权威来源。作为主机，通用游戏模块145生成游戏内容以例如经由其相应客户端设备110呈现给玩家。通用游戏模块145可以在托管并行现实游戏时访问游戏数据库115，以取回和/或存储游戏数据。针对并行现实游戏的所有玩家，通用游戏模块145还

从客户端设备110接收游戏数据(例如,深度信息、玩家输入、玩家位置、玩家动作、地标信息等)、并且将所接收的游戏数据并入整体并行现实游戏中。通用游戏模块145还可以管理游戏数据通过网络105向客户端设备110的传递。通用游戏模块145还可以管理客户端设备110的安全方面,包括但不限于保护客户端设备110与游戏服务器120之间的连接,在各种客户端设备110之间建立连接,并且验证各种客户端设备110的位置。

[0041] 在包括商业游戏模块150的实施例中,商业游戏模块150可以与通用游戏模块145分离或者作为其一部分。商业游戏模块150可以管理在并行现实游戏中是否包括与真实世界中的商业活动链接的各种游戏特征。例如,商业游戏模块150可以通过网络105(经由网络接口)从诸如赞助商/广告商、企业或其他实体等外部系统接收如下的请求,该请求用于将与商业活动链接的游戏特征包括在并行现实游戏中。然后,商业游戏模块150可以布置是否将这些游戏特征包括在并行现实游戏中。

[0042] 游戏服务器120还可以包括数据收集模块155。在包括数据收集模块155的实施例中,数据收集模块155可以与通用游戏模块145分离或者作为其一部分。数据收集模块155可以管理在并行现实游戏中是否包括与真实世界中的数据收集活动链接的各种游戏特征。例如,数据收集模块155可以修改存储在游戏数据库115中的游戏数据,以在并行现实游戏中包括与数据收集活动链接的游戏特征。数据收集模块155还可以分析由玩家根据数据收集活动而收集的数据、并且提供数据以供各种平台访问。

[0043] 事件模块160管理玩家对并行现实游戏中的事件的访问。尽管为了方便起见而使用术语“事件”,但是应当理解,该术语不必指代在特定位置或时间的特定事件。相反,它可以指代提供任何访问控制的游戏内容,其中使用一个或多个访问标准来确定玩家是否可以访问该内容。这样的内容可以是较大的并行现实游戏的一部分,该较大的并行现实游戏包括如下的游戏内容,该游戏内容具有较少访问控制或没有访问控制,或者可以是独立的访问控制的并行现实游戏。

[0044] 深度估计训练系统170训练深度估计模型,例如提供给客户端设备110的深度估计模型130。深度估计训练系统170接收用于训练深度估计模型的图像数据。一般而言,深度估计训练系统170处理图像数据,将图像数据输入到深度估计模型和姿态估计模型中,将图像投影到其他图像上作为合成帧,并且迭代地调节深度估计模型的参数。深度估计训练系统170还可以基于合成帧和图像的比较,利用输入比例遮挡(scale occlusion)和运动感知损失函数(motion aware loss function),从而定义损失值(loss value),然后在细化参数时,该损失值被最小化。损失值还可以指示深度估计模型是否被充分训练、和/或在估计深度信息时是否足够精确。损失函数还可以考虑从来自单眼视频的多个时间图像投影的两个合成帧之间的最小光度重构误差,多个时间图像在时间上相邻。在合成帧的生成期间,也可以使用上采样深度特征,这将影响外观匹配损失的计算。损失函数还可以实现所生成的掩蔽(mask),该掩蔽在计算损失值时减少图像中的静态特征(static feature)。一旦深度估计模型被训练,深度估计模型就接收图像数据、并且在图像数据中输出环境的深度信息。深度估计训练系统170将经训练模型提供给客户端设备110。深度估计训练系统170的训练将在下面进一步描述。

[0045] 网络105可以是任何类型的通信网络,诸如局域网(例如,内联网)、广域网(例如,互联网)或其某种组合。网络还可以包括客户端设备110与游戏服务器120之间的直接连接。

通常，游戏服务器120与客户端设备110之间的通信可以使用各种通信协议(例如，TCP/IP、HTTP、SMTP、FTP)、编码或格式(例如，HTML、XML、JSON)和/或保护方案(例如，VPN、安全HTTP、SSL)、使用任何类型的有线和/或无线连接经由网络接口来承载。

[0046] 本文中讨论的技术参考服务器、数据库、软件应用和其他基于计算机的系统、以及与这样的系统获取的动作和发送的信息。本领域普通技术人员将认识到，基于计算机的系统的固有灵活性允许组件之间的任务和功能的多种可能的配置、组合和划分。例如，本文中讨论的服务器进程可以使用单个服务器或组合工作的多个服务器来实现。数据库和应用可以在单个系统上实现，也可以分布在多个系统上。分布式组件可以顺序地或并行操作。

[0047] 另外，在本文中讨论的系统和方法访问和分析关于用户的个人信息或者使用诸如位置信息等个人信息的情况下，可以为用户提供机会，以控制程序或特征是否收集信息并且控制是否从系统或其他应用接收内容和/或如何从系统或其他应用接收内容。直到已经向用户提供了有关要收集哪些信息以及如何使用该信息的有意义的通知，才收集或使用这样的信息或数据。除非用户提供同意(用户可以随时撤消或修改该同意)，否则不会收集或使用该信息。因此，用户可以控制有关用户的信息如何收集以及如何由应用或系统使用。另外，某些信息或数据可以在存储或使用之前以一种或多种方式处理，因此可以删除个人身份信息。例如，可以处理用户的身份，使得无法确定该用户的个人身份信息。

[0048] 示例性游戏界面

[0049] 图3描绘了可以在客户端120的显示器上呈现的、作为玩家与虚拟世界210之间的界面的一部分的游戏界面300的一个实施例。游戏界面300包括显示窗口310，该显示窗口310可以用于显示虚拟世界210和游戏的其他各个方面，诸如玩家位置222以及虚拟世界210中的虚拟元素230、虚拟物品232和虚拟能源250的位置。用户界面300还可以显示其他信息，诸如游戏数据信息、游戏通信、玩家信息、客户位置验证指令和与游戏相关联的其他信息。例如，用户界面可以显示玩家信息315，诸如玩家名称、经验等级和其他信息。用户界面300可以包括用于访问各种游戏设置和与游戏相关联的其他信息的菜单320。用户界面300还可以包括通信界面330，该通信界面330启用游戏系统与玩家之间的通信、以及并行现实游戏的一个或多个玩家之间的通信。

[0050] 根据本公开的各方面，玩家可以通过简单地在真实世界中随身携带客户端设备110来与并行现实游戏交互。例如，玩家可以通过简单地访问与智能电话上的并行现实游戏相关联的应用、并且与智能电话一起在真实世界中移动来玩游戏。在这点上，玩家不必为了玩基于位置的游戏而在显示屏上连续观看虚拟世界的视觉表示。结果，用户界面300可以包括允许用户与游戏交互的多个非视觉元素。例如，当玩家接近游戏中的虚拟元素或对象时，或者当在并行现实游戏中发生重要事件时，游戏界面可以向玩家提供可听通知。玩家可以使用音频控件340控制这些可听通知。可以根据虚拟元素或事件的类型，向用户提供不同类型的可听通知。可听通知的频率或音量可以根据玩家到虚拟元素或对象的接近度而增大或减小。可以向用户提供其他非视觉通知和信号，诸如振动通知或其他合适的通知或信号。

[0051] 使用本文中提供的公开内容，本领域普通技术人员将理解，根据该公开内容，很多游戏界面配置和底层功能将是很清楚的。本公开内容不旨在限于任何一种特定配置。

[0052] 深度估计训练

[0053] 深度估计训练系统170训练深度估计模型130以供客户端设备110使用。在图1所示

的实施例中，深度估计训练系统170包括深度和姿态模型175、图像合成模块180、误差计算模块185、外观匹配损失模块190、缩放模块195和掩蔽模块(mask module)197。在其他实施例中，深度估计训练系统170可以包括不同和/或其他组件，例如数据存储库、反馈模块、平滑模块等。例如，数据存储库可以在训练深度和姿态模型175时存储训练数据或训练参数。在另一示例中，平滑模块可以处理深度图，诸如对深度图中的深度值进行平滑化。深度和姿态模型175包括接收图像并且可以确定图像的深度特征和/或姿态的一个或多个模型。如将在下面讨论的，深度和姿态模型175可以配置有用于深度模型的参数，该参数不同于用于姿态模型的参数。替代地，深度和姿态模型175可以被配置为使得来自姿态模型的一个或多个参数与深度模型共享。

[0054] 现在参考图4，深度估计训练系统170可以彼此分开地训练深度估计模型410和姿态估计模型440，以使其能够从输入图像确定场景的深度图和输入图像的姿态。在该实施例中，深度估计模型410和姿态估计模型440分别操作，每个使用计算时间和资源来操作。

[0055] 深度估计模型410接收输入图像以确定与该输入图像相对应的深度图。在一个实施例中，深度估计模型410将输入图像馈送通过深度编码器420，以提取抽象深度特征。深度编码器420可以使用不同的机器学习算法和技术来提取这些特征。在该图示中，深度编码器420是包括多个层的卷积神经网络，其中每个后续层减小所提取的特征的维数。例如，在第一层之后，将数量级为10^6个像素或数据点的输入图像缩小为数量级为10^5的一组特征。通过深度编码器420中的最后一层，抽象深度特征可以在10^4或更小的数量级。这些数字纯粹是出于说明目的。实际上，深度编码器可以具有不同数目的层，并且像素和深度特征的数目可以变化。

[0056] 以相反的方式，深度解码器430包括多个层以增加抽象特征的维数。按照上面的示例，深度解码器430可以采用数量级为10^4的抽象深度特征，并且在多个层上逐步导出输入图像的每个像素处的深度。然后，深度解码器430输出深度图，其中深度图上的每个像素对应于场景中的被投影到该像素的最近对象的距离。在替代实施例中，深度解码器430输出视差图，其中视差图上的每个像素对应于距离的倒数。在整个本公开中，参考深度图描述的原理容易地应用于具有视差图的实现中。例如，输入图像已经在给定像素处捕获了距相机某个未知距离的树。深度解码器430输出与从相机到该像素处的街区的距离相对应的深度值。在一些实施例中，输出深度值可以相对于另一深度值或被固有地定义。在其他实施例中，输出深度值按比例缩放，以提供对场景中的对象的真实测量，例如，一个街区在10英尺之外，或建筑物在25米之外。

[0057] 姿态估计模型440接收多个级联输入图像以确定每个输入图像的姿态。姿态通常是指两个图像的透视图之间的数学变换。在整个本公开中，姿态更一般地描述了图像的透视图，其中该透视图可以用于定义两个图像之间的变换。将多个级联输入图像放入姿态编码器450中，该姿态编码器450从多个级联输入图像中提取抽象姿态特征。然后抽象姿态特征被输入通过姿态解码器460，该姿态解码器460确定每个级联输入图像的姿态或每对输入图像之间的相对变换。姿态编码器450可以被配置为包括多个层的卷积神经网络，以用于提取抽象姿态特征并且然后推导每个级联输入图像的姿态。

[0058] 在替代配置中，深度姿态混合模型455与深度模型共享来自姿态估计模型的参数，这在给定较少的要训练的参数的情况下减少了总体计算时间，还有其他优点。在一个实施

例中，深度姿态混合模型455是接收场景的一个或多个图像、并且确定图像的一个或多个姿态的模型。深度姿态混合模型455结合了深度编码器470(其可以是深度估计模型410的深度编码器420)和姿态解码器480(其可以是姿态估计模型440的姿态解码器460)。在该实施例中，深度姿态混合模型455结合了在深度估计模型410和姿态估计模型440中使用的原理，因此能够减少总体计算时间和资源。此外，深度姿态混合模型455提供了在两个模型之间共享信息的途径，使得训练更容易。

[0059] 深度姿态混合模型455利用深度编码器470和姿态解码器480。在一个实施例中，深度姿态混合模型455获取多个输入图像，并且将每个输入图像馈送通过深度编码器470以提取抽象深度特征。然后，在将每个输入图像的抽象深度特征输入到姿态解码器480中之前，将它们级联在一起，从而得到每个输入图像的姿态、或两个后续输入图像之间的相对变换。在为每对输入图像提取姿态时，深度姿态混合模型455比姿态估计模型440在计算上更有效。深度姿态混合模型455将一些输入图像的抽象深度特征进行级联，而姿态估计模型440将输入图像进行级联。深度姿态混合模型455的姿态解码器480能够通过在深度编码器470与姿态解码器480之间共享训练参数来减少无关计算资源的使用。

[0060] 图像合成模块180将合成帧从一个训练图像投影到另一训练图像。在单眼视频数据的投影中，通过考虑第一时间图像的深度以及第一时间图像时间步长和第二时间图像时间步长之间的相对变换两者，图像合成模块180从第一时间步长的一个时间图像投影到第二时间步长的第二时间图像。深度是中间变量，而相对变换是从深度和姿态模型175获取的。

[0061] 在另一实施例中，图像合成模块180还考虑每个图像的固有特征。图像的固有特征是指用于捕获该图像的相机的几何特性，例如，包括相机的焦距、相机的主点偏移、相机的偏斜。在某些情况下，每个相机的固有特征在所获取的所有图像之间可以是恒定的，或者随着相机在获取各种图像时调节其参数，固有特征可以有所不同。在任何一种情况下，固有特征都可以表示为用于变换时间图像的本征矩阵。在另一实施例中，图像合成模块180还使用图像的姿态来利用单眼训练图像数据使图像扭曲。图像合成模块180将第一时间图像变换成第二时间图像的合成帧。

[0062] 在一个实施例中，图像合成模块180从单眼视频中获取三个连续时间图像的集合，并且从第一时间图像投影到第二时间图像时间步长上作为第一合成帧。图像合成模块180还从第三时间图像投影到第二时间图像时间步长上作为第二合成帧。在投影立体图像数据时，图像合成模块180从立体图像对中的一个(左图像)投影到立体图像对中的另一个(右图像)。在从一个图像投影到另一图像时，图像合成模块180(类似于单眼视频数据的投影)考虑立体图像对的深度和左图像与右图像之间的姿态两者。但是，与单眼视频数据不同，左图像与右图像之间的姿态由捕获立体图像对的两个相机的位置确定。图像合成模块180从左图像投影到右图像作为右合成帧，并且从右图像投影到左图像作为左合成帧。

[0063] 误差计算模块185计算合成帧与时间图像之间的差异。在利用单个输入图像计算光度重构误差的实施例中，误差计算模块185将从单个源图像投影的合成帧与另一图像之间的差异作为光度重构误差。

[0064] 外观匹配损失(appearance matching loss)模块190确定在利用多个输入图像进行计算时的光度重构误差(也称为外观匹配损失)。按照具有三个连续时间图像的集合的上

述实施例,误差计算模块185可以计算第一合成帧与第二时间图像之间的差异、以及第二合成帧与第二时间图像之间的差异。当一个时间图像中存在的特征在相邻时间图像中被遮挡(occlude)或被显现(disocclude)时,可能会出现问题。不幸的是,与这些特征相对应的像素会负面影响深度模型的训练。例如,如果针对这样的像素预测正确的深度,则被遮挡(或被显现)的源图像中的对应光度重构误差将可能非常大,从而尽管具有正确预测的深度,仍会导致较高的光度重构误差损失。这种有问题的像素来自两个主要类别:由于图像边界处的自我运动而导致的视线外像素(out-of-view pixel)、以及被遮挡(或被显现)的像素。在一个实施例中,外观匹配损失模块190标识与第一合成帧和第二合成帧的两个差异之间的最小值。在另一实施例中,外观匹配损失模块190对两个差异求平均。按照具有立体图像对的以上实施例,误差计算模块185可以计算左合成帧与左图像之间的左差异、以及右合成帧与右图像之间的右差异。外观匹配损失模块可以标识左差异与右差异之间的最小值、或者计算左差异与右差异之间的平均值。取两个差异之间的最小值有助于缓解由于一个视图中存在被遮挡的对象而另一视图中没有该对象而引起的问题,这可以避免产生伪像。事实证明,这在以下方面具有优势:显著减少图像边界处的伪影,改善遮挡边界的清晰度,以及使深度估计总体上具有更高的准确性。

[0065] 缩放模块195将深度图缩放为用于训练的输入图像的分辨率。常规地,外观匹配损失被计算为深度解码器中每一层的个体损失的组合。缩放模块195基于深度特征的分辨率和输入图像的分辨率,确定要被上采样的训练图像的深度特征的缩放因子。上采样可以使用多种图像上采样技术来实现,包括但不限于双线性采样或双三次采样(bicubic sampling)。上采样的深度特征用于生成合成帧和计算外观匹配损失。使用上采样的深度特征可以提供更好的训练结果,并且当在深度解码器中每一层的分辨率下计算图像的光度重构误差时,避免了可能会引入的纹理复制伪影(即,深度图中的细节从输入图像被错误地转移)。

[0066] 掩蔽模块(maskingmodule)197掩蔽训练图像数据中的一个或多个静态特征。静态特征可以被定义为例如在单眼视频中在两个或更多个图像之间的基本相似位置的一组像素。例如,以与捕获单眼视频的相机相同的速度移动的对象将在帧与帧之间显示为单眼视频中基本相似位置中的像素。换言之,对象在第一时间戳的第一图像和第二时间戳的第二图像之间可能出现在基本相同的位置中。当深度估计训练系统170计算外观匹配损失时,掩蔽模块197通过对静态特征施加掩蔽来考虑这些静态特征,该掩蔽将这些静态特征滤出。这样做防止了深度模型将单眼视频中的静态特征确定为处于非常不精确的深度,例如,趋向于无穷大,因为趋于无穷大的对象在帧与帧之间看起来是静态的。

[0067] 在一种实现中,掩蔽模块197基于所计算的损失来施加掩蔽。掩蔽模块197计算第一时间图像与第二时间图像之间的第一损失。掩蔽模块197分别计算第一时间图像与从第二时间图像投影的合成帧之间的第二损失。基于第一损失是否大于第二损失,掩蔽可以是克罗内克德尔塔函数(kronecker delta function)。然后可以在训练深度模型的参数期间,将掩蔽应用于合成帧与输入图像之间的损失计算。

[0068] 在利用训练图像训练其模型和模块之后,深度估计训练系统170可以为深度估计模型130提供参数,以接收颜色输入图像、并且基于由深度估计训练系统170训练的参数来生成深度图,该深度估计训练系统170包括深度和姿态模型175、图像合成模块180、误差计

算模块185、外观匹配损失模块190和缩放模块195。注意，尽管为了方便起见，深度估计训练系统170被示出为游戏服务器120的一部分，但是一些或全部模型可以由其他计算设备训练并且以各种方式提供给客户端设备110，包括作为操作系统的一部分，被包括在游戏应用中，或者按需在云中访问。

[0069] 图5是描述根据一个或多个实施例的训练深度估计模型的一般过程500的流程图。过程500产生多个参数，深度估计模型130可以使用该多个参数在给定输入图像的情况下生成深度图。

[0070] 深度估计训练系统170首先获取510训练图像数据，该训练图像数据包括多个单眼时间图像和/或多个立体图像对的组合。单眼视频数据可以从外部设备上的相机(例如，客户端设备110上的相机组件125)接收。立体图像对可以从外部设备上的一对双目相机(例如，客户端设备110上的相机组件125)接收。在一个实施例中，网络接口105接收训练图像数据。深度估计训练系统170可以将训练图像数据存储在各种数据存储库中，例如，将单眼视频数据存储在单眼视频数据存储库中、以及将立体图像对存储在立体图像数据存储库中。

[0071] 当使用单眼视频时，深度估计训练系统170将来自单眼视频数据的多个时间图像分组520为三个连续时间图像的多个集合。分组520为三个一组的这个步骤旨在利用投影到第三时间图像上的两个时间图像来计算光度重构误差。在其他实施例中，深度估计系统170可以将时间图像分组为四个一组或五个一组等。

[0072] 深度估计训练系统170将每个图像输入530到深度模型中以提取深度特征。在一个实施例中，将图像输入到深度估计模型(例如，深度估计模型410)中，该深度估计模型提取深度特征作为深度图，例如，该深度图可以是图像的分辨率。

[0073] 深度估计训练系统170将图像输入540到姿态解码器中以提取每个图像的姿态。在一个实施例中，将图像输入到提取图像的姿态的姿态估计模型(例如，姿态估计模型440)中。在具有深度姿态混合模型的实施例中，从深度编码器(例如，深度编码器470)确定的多个抽象深度特征被级联并且被输入到姿态解码器(例如，姿态解码器480)中，以提取针对每个时间图像的姿态。对于立体图像对，姿态定义或帮助定义立体图像对的两个透视图之间的变换。在一些实施例中，立体图像对的两个透视图之间的姿态是固定的和/或已知的。通过将单眼视频数据分组为三个连续时间图像(例如，第一、第二和第三时间图像)的多个集合，深度估计训练系统170提取从第一到第二的相对变换、以及从第二到第三的另一相对变换。

[0074] 利用深度特征和姿态，深度估计训练系统170将时间图像投影550到后续时间图像上、和/或将每个立体图像投影到立体图像对中的另一立体图像上。对于每组中的三个时间图像，深度估计训练系统170将第一时间图像投影到第二时间步长上作为第一合成帧，并且将第三时间图像投影到第二时间步长上作为第二合成帧。在以第一时间图像的深度为中间变量的情况下，基于第一时间图像的姿态或从第一时间图像到第二时间图像的相对变换，深度估计训练系统170将第一时间图像投影到第二时间步长上。同样，在以第三时间图像的深度为中间变量的情况下，利用从第二时间图像到第三时间图像的逆相对变换，深度估计训练系统170也将第三时间图像投影到第二时间步长上。在一实施例中，图像合成模块180执行从一个时间图像到合成帧的投影。对于立体图像对，深度估计训练系统170将立体图像对的左图像投影到立体图像对的右图像上作为右合成帧，并且类似地从右图像投影到左图

像作为左合成帧。在一个实施例中，图像合成模块180执行从左图像到右图像的投影，反之亦然。

[0075] 深度估计训练系统170基于合成帧和图像的比较、利用输入比例遮挡(input scale occlusion)和运动感知损失函数(motion aware loss function)来计算560损失值。输入比例遮挡和运动感知损失函数计算损失值，以用于训练深度模型。损失函数包括计算合成帧与输入图像之间每个像素的光度重构误差。损失函数还可以考虑从单眼视频的多个时间图像投影的两个合成帧之间的最小光度重构误差，多个时间图像在时间上相邻，如上面在外观匹配损失模块190中所述。上采样的深度特征(由缩放模块195)也可以在合成帧的生成期间被使用，这将影响外观匹配损失的计算。损失函数还可以实现由掩蔽模块197生成的掩蔽，该掩蔽在计算损失值时减少静态特征。

[0076] 深度估计训练系统170通过使每个像素的光度重构误差最小化来训练570深度模型。对于三个时间图像的集合，深度估计训练系统170基于第一合成帧和第二合成帧与第二时间图像的差异来标识每个像素的最小光度重构误差。在另一实施例中，深度估计训练系统170可以基于合成帧和图像来定义深度估计模型上的整体误差。整体误差可以定义为例如一对图像上的光度重构误差的平均值、多个或所有输入图像上的光度重构误差的平均值等。在使光度重构误差(或整体误差)最小化时，深度估计训练系统170细化深度模型的参数。姿态模型的参数也可以被细化作为使光度重构误差最小化的一部分。在一个实施例中，深度估计训练系统170将光度重构误差计算为两个差异之间的绝对最小值。在一个实施例中，外观匹配损失模块190与图像合成模块180串联地使光度重构误差最小化。在另一实施例中，缩放模块195以变化的分辨率对图像的深度图进行缩放，以调节深度模型中每一层的参数。在另一实施例中，掩蔽模块197标识具有静态特征的一个或多个区域，并且在计算光度重构误差时掩蔽这些区域。

[0077] 深度估计模型

[0078] 图6是描述根据一个或多个实施例的使用深度估计模型的一般过程600的流程图。过程600产生给定输入图像的深度图。过程600可以由具有经训练深度估计模型的客户端设备来完成。客户端设备可以是通用计算设备，也可以具有相机。在一些实施例中，客户端设备在以上图1-3中所述的并行现实游戏中实现。尽管以下描述在客户端设备的上下文内，但过程600可以在其他计算设备上执行。

[0079] 该方法包括接收610场景的图像。场景的图像可以由作为客户端设备的组件或在客户端设备外部的相机捕获。在并行现实游戏的上下文中，场景可以是可以映射到虚拟世界中的虚拟位置的真实世界位置。场景的图像还可以具有与捕获图像的相机的几何性质相对应的固有特征。图像可以是由相机捕获的单个图像。替代地，图像可以是由相机捕获的视频中的一帧。

[0080] 该方法包括将场景的图像输入620到经训练深度估计模型中。深度估计系统170可以例如经由图5的过程500来训练深度估计模型。深度估计模型接收场景的图像，并且可选地也接收图像的固有特征。

[0081] 该方法包括由经训练深度估计模型生成630与场景的图像相对应的场景的深度图。深度图的每个像素具有深度值，该深度值描述场景的图像中的对应像素处的表面的相对距离。深度估计接收场景的图像、并且基于根据图5而训练的参数来输出深度图。

[0082] 该方法包括基于场景的深度图生成640虚拟内容。虚拟内容可以源自于例如存储在游戏数据库115中的用于并行现实游戏的内容。所生成的虚拟内容可以是可以被增强到场景的图像上的增强现实内容。例如，生成虚拟角色，该虚拟角色可以在了解场景的深度的情况下在场景中移动。在一个实例中，当虚拟角色在街道上朝着用户行走时，虚拟角色的尺寸可以增大。在另一实例中，虚拟角色可以躲在树后面，在那里，虚拟角色的一部分随后被树遮挡。

[0083] 该方法包括显示650利用虚拟内容而被增强的场景的图像。客户端设备包括电子显示器。电子显示器可以提供由相机捕获的具有增强的虚拟内容的恒定视频。

[0084] 按照上面的示例，并行现实游戏可以提供与作为目标的虚拟角色的交互。为了与虚拟角色交互，移动设备的用户可能需要四处移动其移动设备，同时将虚拟角色保持在相机的视场中。当用户四处移动移动设备时，随着场景在用户移动移动设备的情况下变化，移动设备可以连续捕获视频或图像数据，这些视频或图像数据可以用于迭代地生成场景的深度信息。移动设备可以更新显示器上的视频馈送，同时还可以基于所生成的深度信息更新虚拟角色，以便用户将虚拟角色感知为始终在场景中适当地交互，例如，没有走过对象，不具有被切掉的部分并且没有任何对象遮挡这些部分，等等。

[0085] 示例计算系统

[0086] 图7是根据一个实施例的计算设备的示例架构。虽然图7描绘了高层框图，该高层框图示出了用作本文中描述的一个或多个实体的一部分或全部的计算机的物理组件，但是根据一个实施例，与图7中提供的相比，计算机可以具有更多组件、更少组件或组件变型。虽然图7描绘了计算机700，但是该图旨在作为计算机系统中可能存在的各种特征的功能描述，而不是作为本文中描述的实现的结构示意图。在实践中，并且如本领域普通技术人员所认识的，单独示出的项目可以组合，并且一些项目可以分离。

[0087] 图7中示出了耦合到芯片组704的至少一个处理器702。存储器706、存储设备708、键盘710、图形适配器712、定点设备714和网络适配器716也耦合到芯片组704。显示器718耦合到图形适配器712。在一个实施例中，芯片组704的功能由存储器控制器集线器720和I/O集线器722提供。在另一实施例中，存储器706直接耦合到处理器702而不是芯片组704。在一些实施例中，计算机700包括用于将这些组件互连的一个或多个通信总线。一个或多个通信总线可选地包括互连并且控制系统组件之间的通信的电路系统(有时称为芯片组)。

[0088] 存储设备708是任何非暂态计算机可读存储介质，诸如硬盘驱动器、光盘只读存储器(CD-ROM)、DVD或固态存储器设备或其他光学存储、盒式磁带、磁带、磁盘存储或其他磁性存储设备、磁盘存储设备、光盘存储设备、闪存设备、或者其他非易失性固态存储设备。这样的存储设备708也可以被称为持久性存储器。定点设备714可以是鼠标、跟踪球或其他类型的定点设备，并且与键盘710结合使用以将数据输入到计算机700中。图形适配器712在显示器718上显示图像和其他信息。网络适配器716将计算机700耦合到局域网或广域网。

[0089] 存储器706保存由处理器702使用的指令和数据。存储器706可以是非持久性存储器，其示例包括高速随机存取存储器，诸如DRAM、SRAM、DDR RAM、ROM、EEPROM、闪存。

[0090] 如本领域中已知的，计算机700可以具有与图7所示的组件不同的组件和/或其他组件。另外，计算机700可以缺少某些示出的组件。在一个实施例中，充当服务器的计算机700可以缺少键盘710、定点设备714、图形适配器712和/或显示器718。此外，存储设备708可

以在计算机700本地中和/或远离该计算机700(例如,体现在存储区域网络(SAN)中)。

[0091] 如本领域中已知的,计算机700适于执行计算机程序模块以提供本文中描述的功能。如本文中使用的,术语“模块”是指用于提供指定功能的计算机程序逻辑。因此,模块可以用硬件、固件和/或软件来实现。在一个实施例中,程序模块存储在存储设备708上,被加载到存储器706中,并且由处理器302执行。

[0092] 其他注意事项

[0093] 在题为“Digging Into Self-Supervised Monocular Depth Estimation”的说明书附录中可以找到实施例的附加讨论,该申请的全部内容通过引用合并于此。

[0094] 以上描述的某些部分在算法过程或操作方面描述了实施例。这些算法的描述和表示通常由数据处理领域的技术人员用来将其工作的实质有效地传达给本领域其他技术人员。虽然在功能上、计算上或逻辑上进行描述,但是这些操作应当被理解为由计算机程序实现,该计算机程序包括用于由处理器或等效电路、微代码等执行的指令。此外,在不失一般性的情况下,功能操作的这些布置有时称为模块也是方便的。

[0095] 如本文中使用的,对“一个实施例”或“实施例”的任何引用表示结合该实施例描述的特定元件、特征、结构或特性被包括在至少一个实施例中。短语“在一个实施例中”在说明书中各个地方的出现不一定全都是指同一实施例。

[0096] 一些实施例可以使用表达“耦合”和“连接”及其派生词来描述。应当理解,这些术语并不旨在彼此等同。例如,一些实施例可以使用术语“连接”来描述,以表示两个或更多个元件彼此直接物理或电接触。在另一示例中,一些实施例可以使用术语“耦合”来描述,以表示两个或更多个元件直接物理接触或电接触。然而,术语“耦合”也可以表示两个或更多个元件不彼此直接接触,但是仍然彼此协作或相互作用。实施例不限于该上下文。

[0097] 如本文中使用的,术语“包括(comprises)”、“包括(comprising)”、“包括(includes)”、“包括(including)”、“具有(has)”、“具有(hasing)”或其任何其他变型旨在覆盖非排他性包含。例如,包括一系列元素的过程、方法、物品或装置不一定仅限于这些元素,而是可以包括未明确列出或这样的过程、方法、物品或装置所固有的其他元素。此外,除非明确相反地指出,否则“或”是指包含性的“或”而不是排他性的“或”。例如,以下任一项满足条件A或B:A为真(或存在)并且B为假(或不存在),A为假(或不存在)并且B为真(或存在),以及A和B都为真(或存在)。

[0098] 此外,“一个(a)”或“一个(an)”的使用用于描述实施例的元件和组件。这样做仅仅是为了方便并且给出本公开的一般意义。该描述应当被理解为包括一个或至少一个,并且单数也包括复数,除非很清楚的是另有说明。

[0099] 在阅读本公开之后,本领域技术人员将理解用于验证在线服务提供商的账户对应于真实业务的系统和过程的另外的替代结构和功能设计。因此,尽管已经图示和描述了特定的实施例和应用,但是应当理解,所描述的主题不限于本文中公开的精确构造和组件,并且可以对所公开的方法和装置的布置、操作和细节做出对于本领域技术人员而言很清楚的各种修改、改变和变化。保护范围应当仅由所附权利要求书限制。

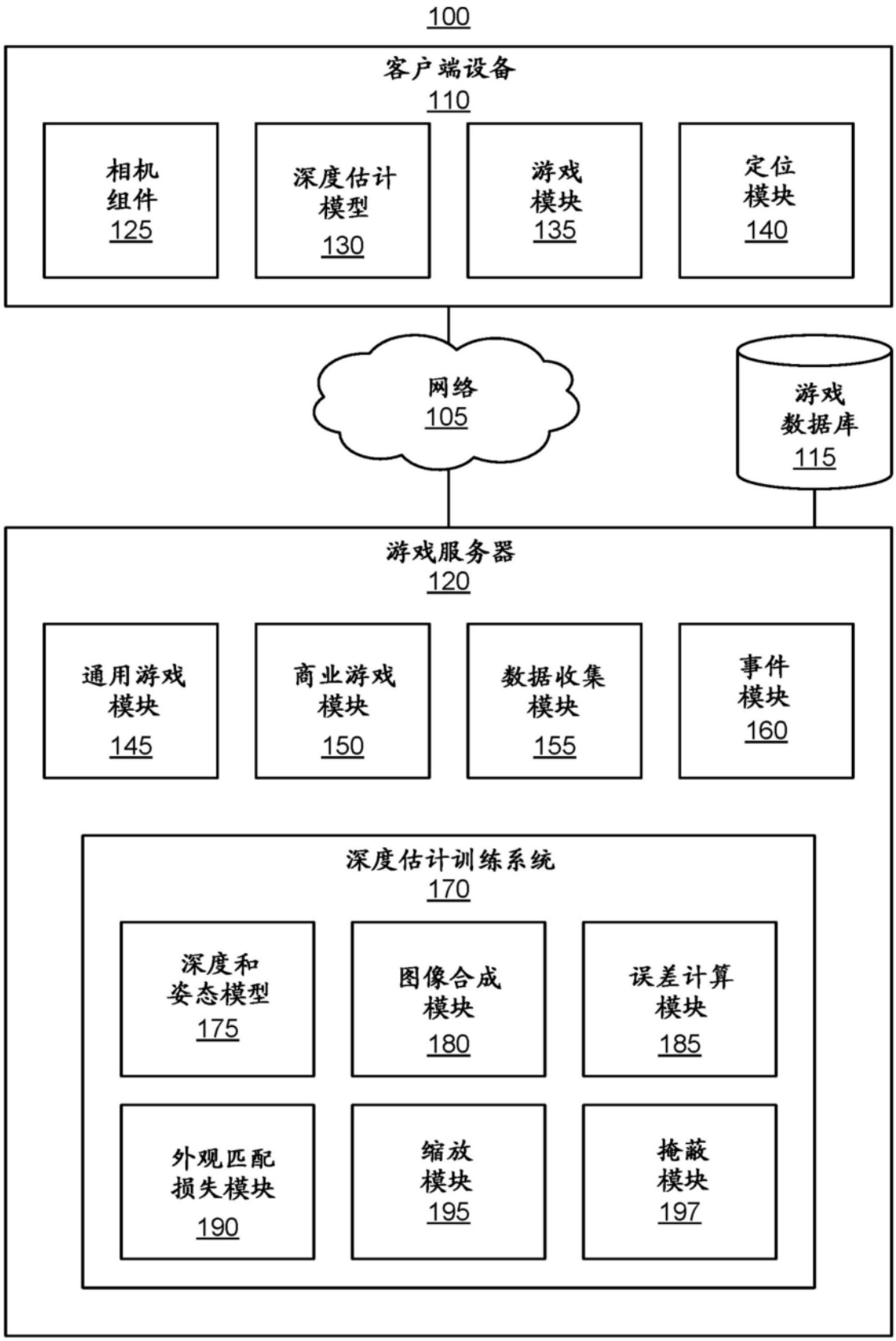

100
客户端设备
110
相机
组件
125
深度估计
模型
130
游戏
模块
135
定位
模块
140
网络
105
游戏
数据库
115
游戏服务器
120
通用游戏
模块
145
商业游戏
模块
150
数据收集
模块
155
事件
模块
160
深度估计训练系统
170
深度和
姿态模型
175
图像合成
模块
180
误差计算
模块
185
外观匹配
损失模块
190
缩放
模块
195
掩蔽
模块
197

图1

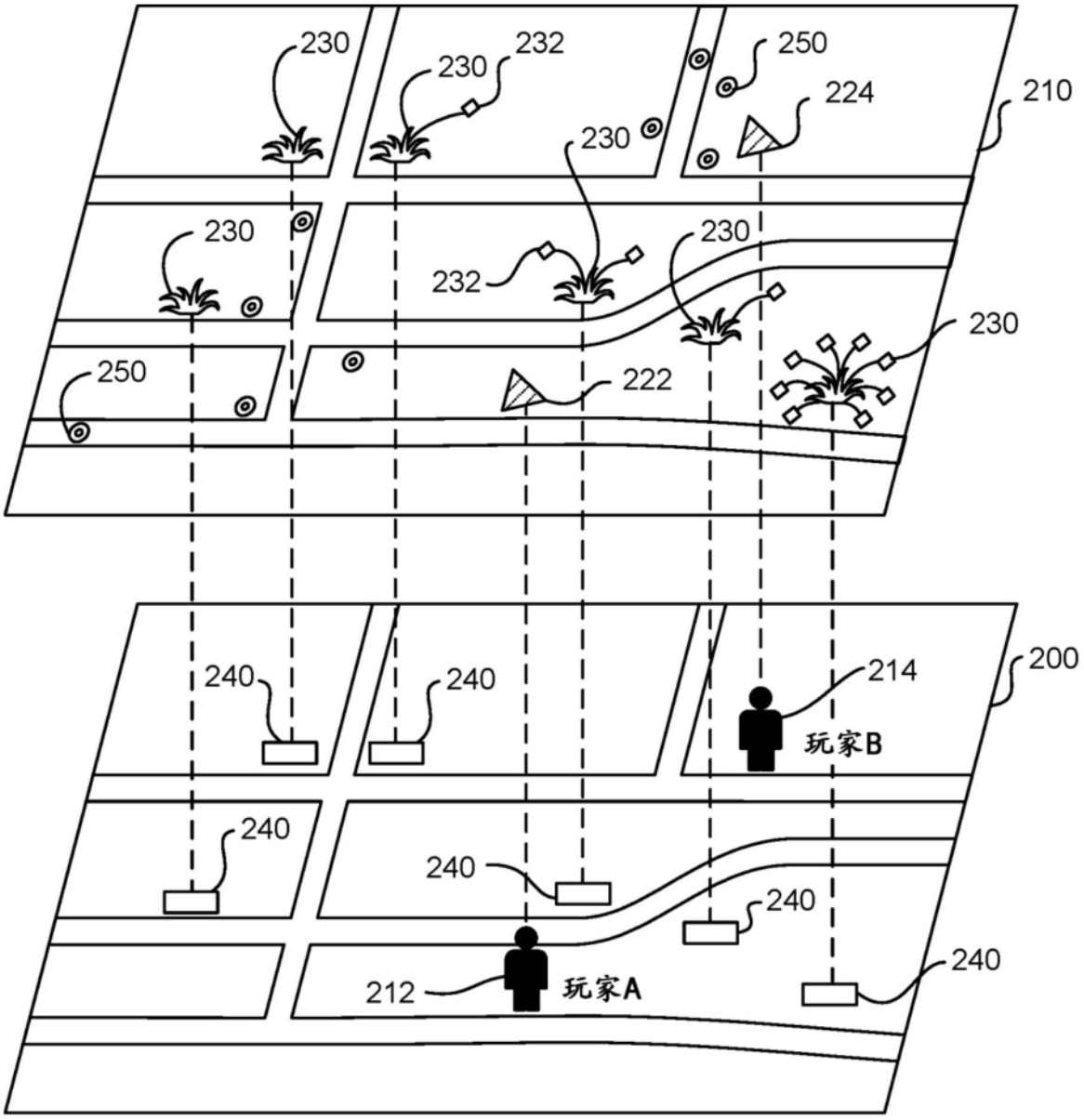

230
230
232
250
224
210
230
230
230
232
230
250
222
240
240
214
200
玩家B
240
240
240
240
212
玩家A

图2

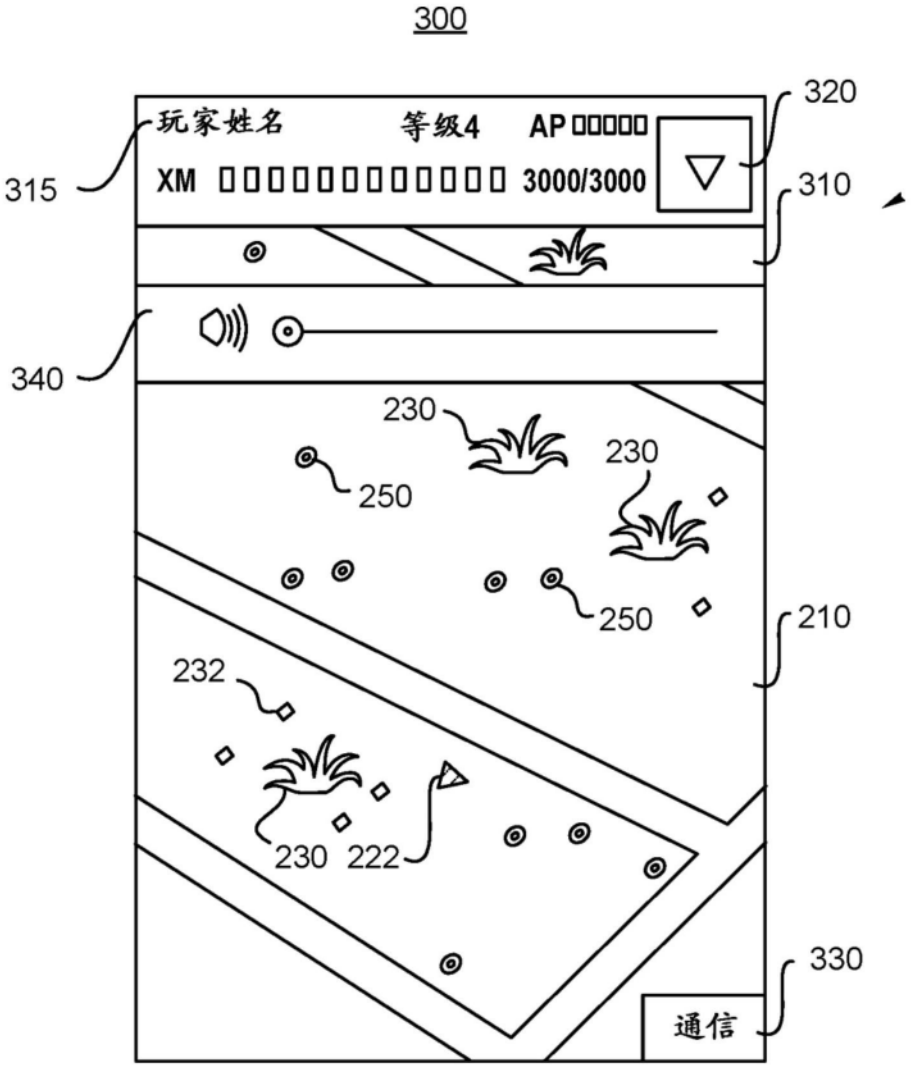

300
玩家姓名
等级4
AP
XM
3000/3000
315
320
310
340
230
230
250
250
210
232
230
222
330
通信

图3

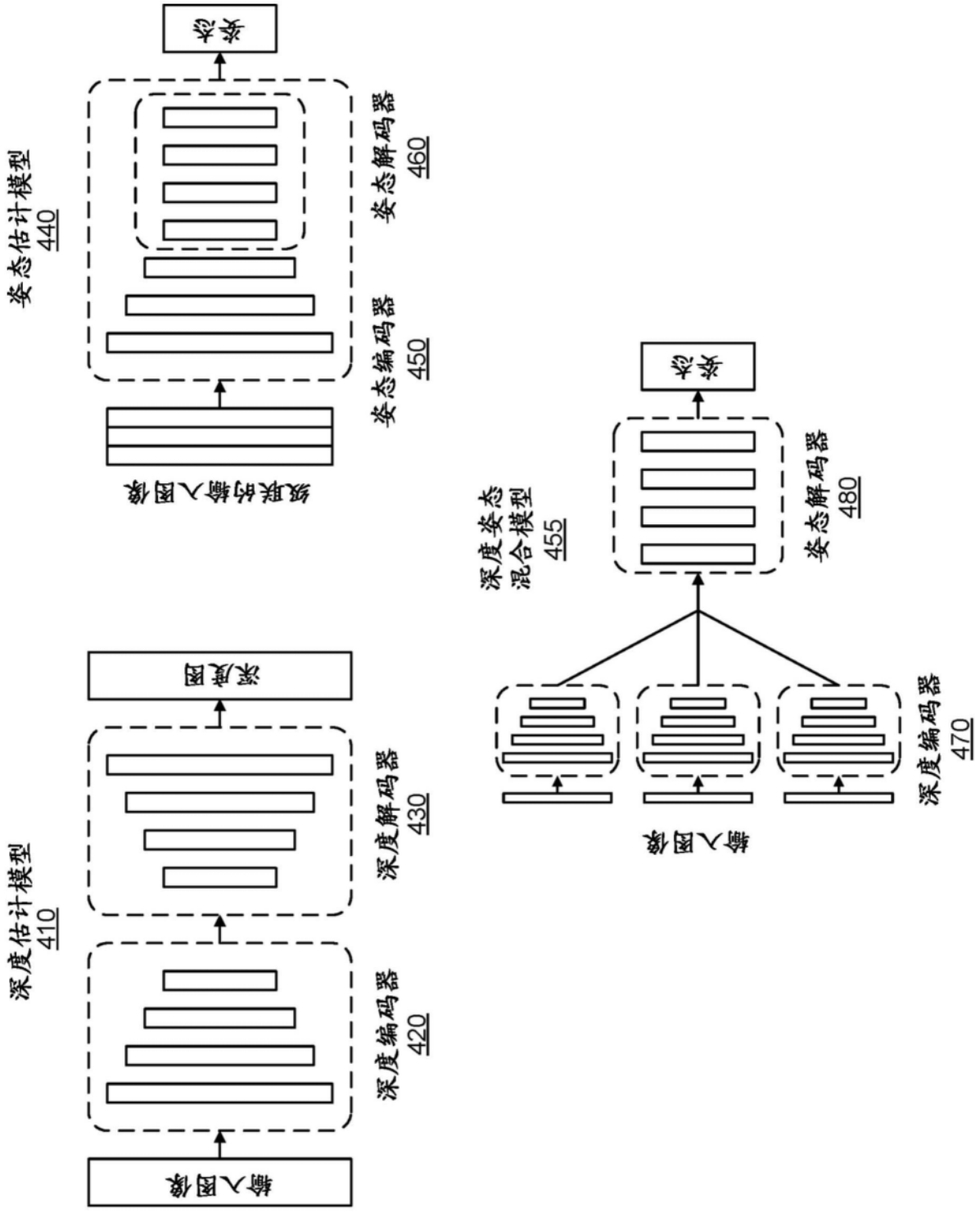

输入图像
深度估计模型
410
深度编码器
420
深度解码器
430
深图
姿态估计模型
440
姿态编码器
450
姿态解码器
460
级联的输入图像
深度姿态
混合模型
455
深度编码器
470
输入图像
姿态解码器
480
姿态

图4

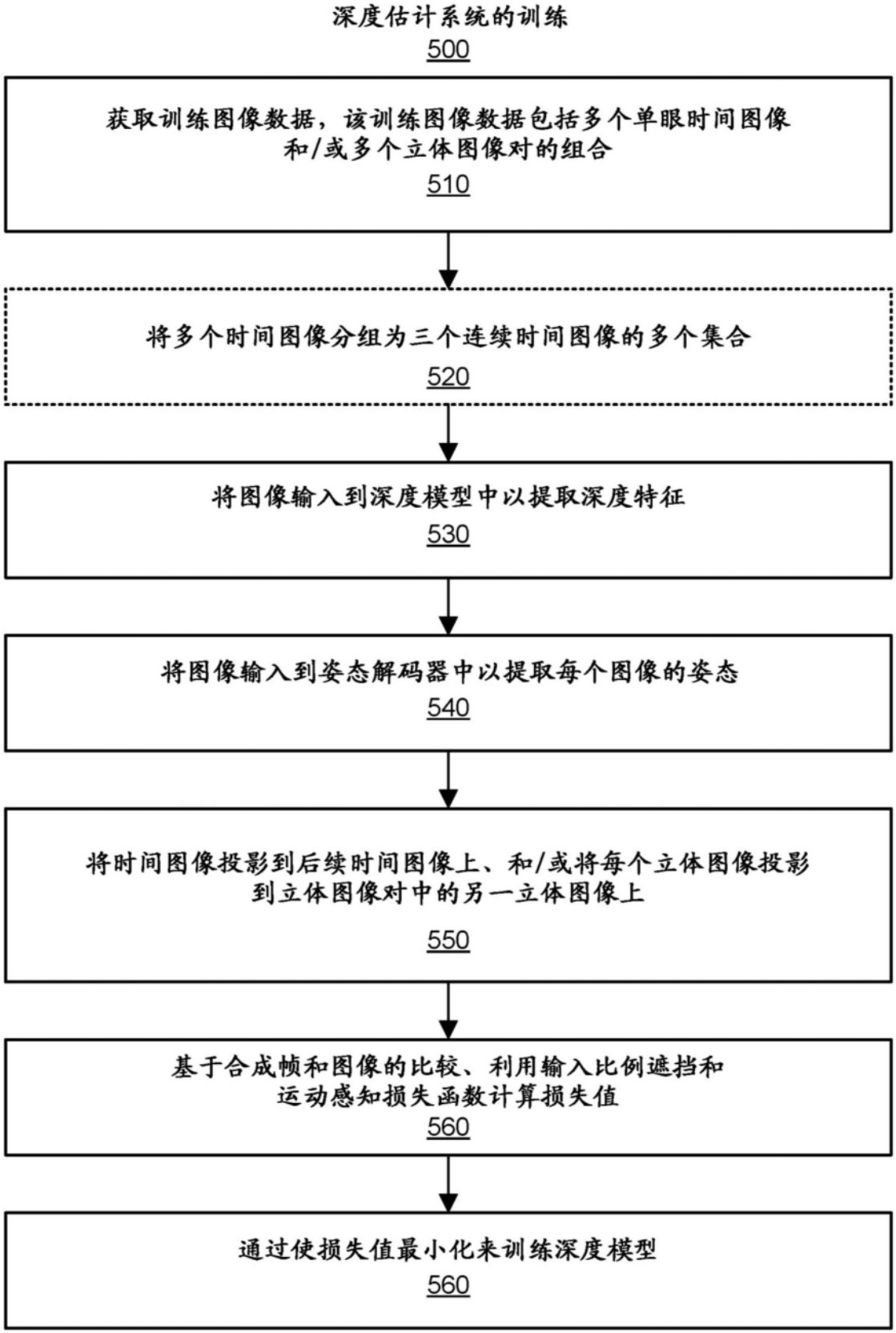

深度估计系统的训练
500
获取训练图像数据，该训练图像数据包括多个单眼时间图像和/或多个立体图像对的组合
510
将多个时间图像分组为三个连续时间图像的多个集合
520
将图像输入到深度模型中以提取深度特征
530
将图像输入到姿态解码器中以提取每个图像的姿态
540
将时间图像投影到后续时间图像上、和/或将每个立体图像投影到立体图像对中的另一立体图像上
550
基于合成帧和图像的比较、利用输入比例遮挡和运动感知损失函数计算损失值
560
通过使损失值最小化来训练深度模型
560

图5


深度估计
600
接收场景的图像
610
将场景的图像输入到经训练
深度估计模型中
620
由经训练深度估计模型生成与场景的
图像相对应的场景的深度图
630
基于场景的深度图生成虚拟内容
640
显示利用所生成的虚拟内容增强
的场景的图像
650

图6

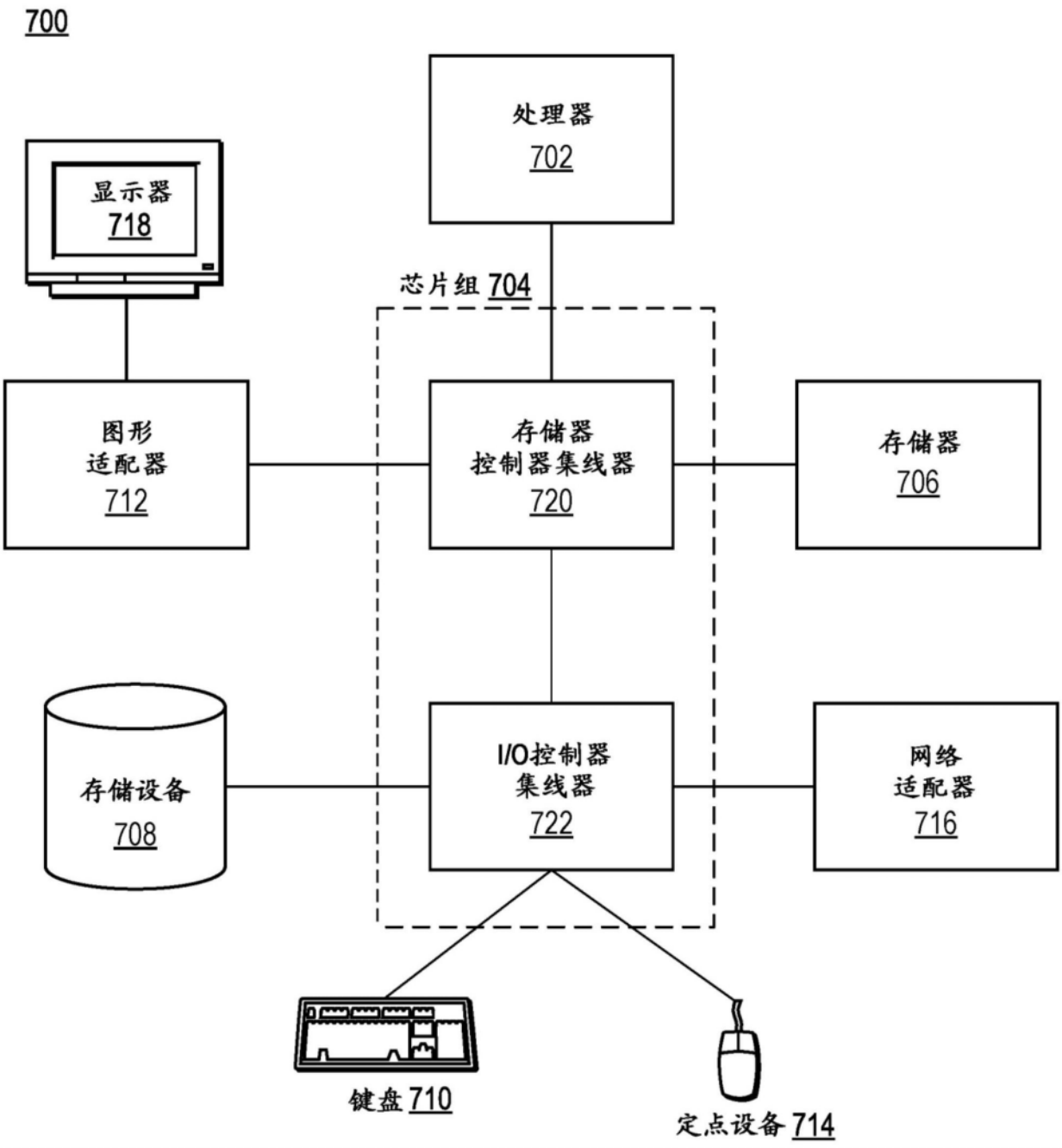

700
处理器
702
显示器
718
芯片组 704
图形
适配器
712
存储器
控制器集线器
720
存储器
706
存储设备
708
I/O控制器
集线器
722
网络
适配器
716
键盘 710
定点设备 714

图7