



(12) 发明专利申请

(10) 申请公布号 CN 112513944 A

(43) 申请公布日 2021.03.16

(21) 申请号 201980049016.2

(74) 专利代理机构 北京市中咨律师事务所

(22) 申请日 2019.07.22

11247

代理人 杨晓光 于静

(30) 优先权数据

62/702,309 2018.07.23 US

(51) Int.Cl.

G06T 19/00 (2006.01)

(85) PCT国际申请进入国家阶段日

A63F 13/211 (2006.01)

2021.01.22

A63F 13/428 (2006.01)

(86) PCT国际申请的申请数据

G02B 27/01 (2006.01)

PCT/US2019/042846 2019.07.22

G06F 1/16 (2006.01)

(87) PCT国际申请的公布数据

W02020/023399 EN 2020.01.30

(71) 申请人 奇跃公司

地址 美国佛罗里达州

(72) 发明人 L·巴拉克 G·罗森塔尔

A·佩里

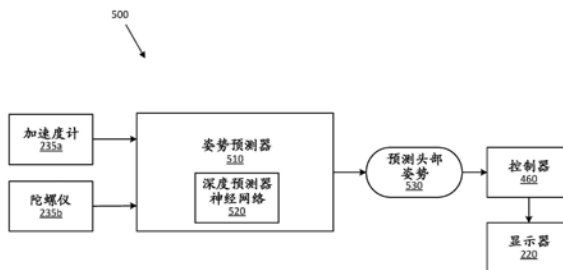
权利要求书2页 说明书17页 附图9页

(54) 发明名称

用于头部姿势预测的深度预测器循环神经网络

(57) 摘要

用于预测用于增强或虚拟现实设备的渲染引擎的头部姿势的系统和方法可以包括接受时间序列的头部姿势数据并输出预测头部姿势的循环神经网络(RNN)。该循环神经网络可以包括一个或多个长短期记忆(LSTM)单元或门控循环单元(GRU)。全连接(FC)层可以接受来自RNN的输入,并输出3自由度(DOF)头部姿势(例如,角度取向或空间位置)或6DOF头部姿势(例如,角度取向和空间位置两者)。渲染引擎可以使用预测头部姿势生成虚拟内容并且在用户看向虚拟内容的位置时向用户显示该虚拟内容,这减少了系统延迟并改善了用户体验。



1. 一种可穿戴系统,包括:
 - 显示器,其被配置为被设置在所述可穿戴系统的佩戴者的眼睛前方,所述显示器被配置为向所述可穿戴系统的所述佩戴者显示虚拟内容;
 - 头部姿势传感器,其被配置为提供头部姿势数据;
 - 非暂时性存储器,其被配置为存储头部姿势预测循环神经网络;
 - 与所述头部姿势传感器、所述显示器、以及所述非暂时性存储器通信的硬件处理器,所述硬件处理器被编程为:
 - 接收所述头部姿势数据;
 - 将所述头部姿势数据输入到所述头部姿势预测循环神经网络;
 - 执行所述头部姿势预测循环神经网络以输出在时间范围内的预测头部姿势;以及
 - 使所述显示器至少部分地基于在所述时间范围内的所述预测头部姿势来渲染所述虚拟内容。
2. 根据权利要求1所述的可穿戴系统,其中,所述显示器被配置为在多个深度平面呈现所述虚拟内容。
3. 根据权利要求1所述的可穿戴系统,其中,所述头部姿势传感器包括惯性测量单元。
4. 根据权利要求3所述的可穿戴系统,其中,所述惯性测量单元包括加速度计和陀螺仪。
5. 根据权利要求1所述的可穿戴系统,其中,所述头部姿势数据包括速度或加速度。
6. 根据权利要求5所述的可穿戴系统,其中,来自所述头部姿势传感器的输出由扩展卡尔曼滤波器来滤波以提供所述头部姿势数据。
7. 根据权利要求1所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括循环神经网络(RNN)和全连接(FC)层。
8. 根据权利要求1所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括多个重复模块。
9. 根据权利要求8所述的可穿戴系统,其中,所述多个重复模块中的至少一个重复模块包括:长短期记忆单元(LSTM)或门控循环单元(GRU)。
10. 根据权利要求1所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括:包括一个或多个长短期记忆(LSTM)单元的堆叠或包括一个或多个门控循环单元(GRU)的堆叠。
11. 根据权利要求10所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括三个LSTM单元或三个GRU。
12. 根据权利要求1所述的可穿戴系统,其中,所述时间范围在从10ms至50ms的范围内。
13. 根据权利要求1所述的可穿戴系统,其中,所述时间范围在从20ms至40ms的范围内。
14. 根据权利要求1所述的可穿戴系统,其中,所述预测头部姿势包括3自由度DOF姿势。
15. 根据权利要求14所述的可穿戴系统,其中,所述3DOF姿势包括头部取向。
16. 根据权利要求14所述的可穿戴系统,其中,所述3DOF姿势包括头部位置。
17. 根据权利要求1所述的可穿戴系统,其中,所述预测头部姿势包括6自由度DOF姿势,所述6DOF姿势包括取向和位置。
18. 根据权利要求1所述的可穿戴系统,其中,所述预测头部姿势包括头部取向,并且所

述硬件处理器还被编程为：使用预测性非神经网络模型来预测在所述时间范围内的头部位置。

19. 根据权利要求18所述的可穿戴系统，其中，所述预测性非神经网络模型包括恒定加速度模型。

20. 一种用于渲染虚拟内容的方法，所述方法包括：

在包括计算机硬件的渲染管线的控制下：

接收针对用户的头部的头部姿势数据；

至少使用循环神经网络 (RNN) 和所述头部姿势数据，预测在预测范围内的预测头部姿势；以及

至少部分地基于所述预测头部姿势，渲染在所述预测范围内的所述虚拟内容。

21. 根据权利要求20所述的方法，其中，所述头部姿势数据包括以下一项或多项的时间序列：(1) 来自惯性测量单元 (IMU) 的数据；(2) 针对所述用户的所述头部的速度数据或加速度数据；或(1) 和(2) 两者。

22. 根据权利要求21所述的方法，还包括：对数据进行滤波以提供所述速度数据或所述加速度数据。

23. 根据权利要求22所述的方法，其中，利用扩展卡尔曼滤波器 (EKF) 来执行所述滤波。

24. 根据权利要求20所述的方法，其中，所述RNN包括：包括一个或多个长短期记忆 (LSTM) 单元的堆叠或包括一个或多个门控循环单元 (GRU) 的堆叠。

25. 根据权利要求24所述的方法，其中，所述RNN包括具有三个LSTM单元的堆叠或具有三个GRU的堆叠。

26. 根据权利要求20所述的方法，其中，所述RNN被连接到被配置为输出所述预测头部姿势的全连接 (FC) 层。

27. 根据权利要求20所述的方法，其中，所述预测头部姿势包括所述用户的所述头部的取向，并且所述方法还包括：利用恒定加速度模型来预测在所述预测范围内的所述用户的所述头部的位置。

28. 根据权利要求20所述的方法，其中，所述预测范围在从10ms至50ms的范围内。

用于头部姿势预测的深度预测器循环神经网络

[0001] 相关申请的交叉引用

[0002] 本申请要求于2018年7月23日提交的题为“DEEP PREDICTOR RECURRENT NEURAL NETWORK FOR HEAD POSE PREDICTION (用于头部姿势预测的深度预测器循环神经网络)”美国专利申请号62/702,309的优先权的权益,该申请的全部内容通过引用并入在此。

技术领域

[0003] 本公开总体上涉及使用循环神经网络来预测头部姿势(例如,取向、位置、旋转或平移)的系统和方法,更具体地涉及利用这样的循环神经网络用于头部姿势预测的增强或虚拟现实系统。

背景技术

[0004] 现代计算和显示技术已经促进了用于所谓的“虚拟现实”、“增强现实”或“混合现实”体验的系统的开发,其中,数字再现的图像或其部分以它们似乎是真实的或可能被感知为真实的方式呈现给用户。虚拟现实或“VR”场景通常涉及数字或虚拟图像信息的呈现,而对其它实际真实世界视觉输入没有透明度;增强现实或“AR”场景通常涉及数字或虚拟图像信息的呈现,作为对用户周围的实际世界可视化的增强;与合并真实世界和虚拟世界有关的混合现实或“MR”产生了新的环境,在该环境中物理和虚拟对象共存并实时交互。事实证明,人类的视觉感知系统非常复杂,并且产生促进虚拟图像元素在其它虚拟或真实世界图像元素之间的舒适自然感觉丰富的呈现的VR、AR或MR技术具有挑战性。在此公开的系统和解决方法解决了与VR、AR和MR技术有关的各种挑战。

发明内容

[0005] 用于预测用于增强或虚拟现实设备的渲染引擎的头部姿势的系统和方法可以包括循环神经网络(RNN),该循环神经网络接收时间序列的头部姿势数据并输出预测头部姿势(例如,取向、位置、旋转或平移)。循环神经网络可以包括一个或多个长短期记忆(LSTM)单元或门控循环单元(GRU)。全连接(FC)层可以接受来自RNN的输入,并输出3自由度(DOF)头部姿势(例如,角度取向或空间位置)或6DOF头部姿势(例如,角度取向和空间位置两者)。渲染引擎可以使用预测头部姿势来生成虚拟内容,并且在用户看向虚拟内容的位置时向用户显示该虚拟内容,这样减少了系统延迟并改善了用户体验。

[0006] 在各个方面,可穿戴显示系统包括:显示器,被配置为设置在可穿戴系统的佩戴者的眼睛前方,该显示器被配置为向可穿戴系统的佩戴者显示虚拟内容;头部姿势传感器,被配置为提供头部姿势数据;非暂时性存储器,被配置为存储头部姿势预测循环神经网络;与头部姿势传感器、显示器和非暂时性存储器通信的硬件处理器,该硬件处理器被编程为:接收头部姿势数据;将头部姿势数据输入到头部姿势预测循环神经网络;执行头部姿势预测循环神经网络以在时间范围内输出预测头部姿势;以及使显示器至少部分地基于时间范围内的预测头部姿势来渲染虚拟内容。

[0007] 本文描述的循环神经网络头部姿势预测技术不限于AR、MR或VR应用,并且可以用于需要人的头部姿势的其他应用中,例如,游戏、医学(例如,神经病学、眼科、康复)等。

[0008] 本说明书中描述的主题的一种或多种实现方式的细节在附图和以下描述中阐述。根据说明书、附图和权利要求书,其它特征、方面和优点将变得显而易见。该发明内容或以下详细描述均不旨在限定或限制本发明主题的范围。

附图说明

[0009] 图1描绘了具有某些虚拟现实对象以及人观看的某些物理对象的增强现实场景的图示。

[0010] 图2示意性地示出可穿戴系统的示例。

[0011] 图3示意性地示出用于使用多个深度平面来模拟三维图像的方法的方面。

[0012] 图4示意性地示出用于向用户输出图像信息的波导堆叠的示例。

[0013] 图5是示意性示出用于向可穿戴系统的用户显示虚拟内容的渲染管线的一部分的示例的框图。

[0014] 图6是示意性示出用于头部姿势预测的神经网络的示例的框图。神经网络的此示例包括循环神经网络(RNN)和全连接(FC)层。

[0015] 图7是示意性示出用于头部姿势预测的神经网络的示例的框图。在该示例中,神经网络包括RNN,该RNN包括连接到FC层的三个长短期记忆(LSTM)模块。

[0016] 图8是示意性示出在时间预测范围 dt 内的头部姿势预测的示例的曲线图。

[0017] 图9是示出了用于利用循环神经网络预测头部姿势的方法的示例的流程图。

[0018] 贯穿整个附图,参考标记可以被重复使用以指示参考元件之间的对应关系。提供附图是为了说明在此描述的示例实施例,并不旨在限制本公开的范围。

具体实施方式

[0019] 概述

[0020] AR、MR或VR可穿戴系统可以向用户的眼睛渲染虚拟内容(例如,参见图1)。由于生成可渲染的内容并将其递送给AR、MR或VR显示器需要花费时间,因此,虚拟内容渲染管线中通常会有一定程度的渲染延迟。在各种AR、MR、VR实施方式中,延迟时间可以在从约10ms至约50ms的范围内,例如,对于一些可穿戴系统约30ms。

[0021] 在AR、MR或VR体验期间,用户将会移动其头部以观看用户周围环境的不同部分,并且由于该头部移动,虚拟内容将进入和离开用户的视场(FOV)。可穿戴系统将渲染FOV中的虚拟内容以显示给用户。但是,随着用户移动其头部,当前渲染的虚拟内容的位置可能会发生变化(例如,从FOV的中心附近到FOV的更外围区域,反之亦然),某些虚拟内容将离开FOV并且不需要被渲染,而其他以前未被渲染的虚拟内容将进入用户的FOV并且将需要被渲染。因此,在使用时,虚拟内容的渲染位置将由于用户的头部移动而随时间变化。

[0022] 考虑到可穿戴系统渲染虚拟内容所花费的渲染延迟时间段(例如,30ms),对于系统而言,预测用户在(与渲染延迟时段相当的)未来时间将会看向的位置,以了解用户在该未来的时间将会观看哪些虚拟内容是有利的。系统可以在该未来时间之前(例如,在与渲染延迟时间相当的时间段之前)开始渲染虚拟内容,以使得当用户开始观看虚拟内容时,实际

上已经向用户渲染虚拟内容。从某种意义上讲,可穿戴系统的实施例可以开始渲染虚拟内容,以便当用户观看虚拟内容时及时显示该虚拟内容。这样的渲染管线可以例如减少延迟,在虚拟内容在FOV中移动时提供平滑的过渡,或减少所显示的虚拟内容的跳跃性,从而可以有益地改善用户的AR、MR或VR体验。

[0023] 头部姿势预测可以使用来自头部姿势传感器的信息,该头部姿势传感器可以包括加速度计或陀螺仪或其他位置或取向传感器。头部姿势可以指代(例如,以诸如偏航、俯仰和滚转的三个角度进行测量的)取向、(例如,以诸如笛卡尔x、y和z的三个空间坐标进行测量的)位置、或取向和位置两者。头部运动(例如,头部姿势的改变)可以通过取向的改变(例如,角度的改变)、空间位置的改变(例如,平移)或两者来测量。由于用于预测未来头部姿势的时间范围是与渲染延迟(例如,约30ms)相当的,因此一些实施方式利用恒定加速度模型来预测未来头部姿势。恒定加速度模型假设用户的头部的加速度在预测时间范围内大致恒定。恒定加速度模型可以使用来自头部姿势传感器(例如,加速度计或陀螺仪)的输入来测量加速度。可以对头部姿势传感器输入进行滤波(例如,使用扩展卡尔曼滤波器(EKF))以提供更平滑的速度或加速度估计。作为恒定加速度预测的一维非限制性示例,如果 ω 是用户的头部的角旋转速率,以及 α 是恒定角加速度,则在预测时间段 dt 内用户的头部的角位置的变化 $\Delta\theta$,为 $\Delta\theta = \omega dt + \alpha dt^2/2$ 。可以针对用户的头部位置的改变来开发类似的恒定加速度运动学模型。

[0024] 预测的未来头部姿势的误差可能由预测模型中使用的速度或加速度的误差以及预测模型本身的误差引起。例如,即使在相对较短的预测时间范围内,真实用户的头部也不会以恒定的加速度运动。例如,从用户测量中发现,均方根(rms)位置误差为约0.6mm,而95%分位数(q95)约为1.5mm。在这些示例用户测量中,大多数位置误差来自估计的速度和加速度误差(这些误差是从加速度计和陀螺仪数据的EKF滤波获得的)。对于一些可穿戴系统,用于头部平移的恒定加速度模型可以足够准确。

[0025] 在这些示例用户测量中,均方根角度误差为约1.5弧分,而q95为约4弧分。但是,与位置误差形成对比,大约30%的角度误差是由于使用了恒定加速度模型引起的。因此,可以通过利用比恒定加速度模型更具鲁棒性和更准确的角度预测模型来改善头部姿势预测。

[0026] 本公开提供用于头部姿势预测的循环神经网络(RNN)方法的示例,该方法可以比恒定加速度模型精确得多。在各种实施例中,循环神经网络方法可以应用于头部移动的角度预测、平移预测、或角度和平移预测两者。例如,在一些可穿戴系统中,与恒定加速度模型相比,已经发现用于角度头部姿势预测的循环神经网络模型将均方根角度误差减小25%,将角度q95改善约27%,并且总体上提供了角度预测方面约38%的提高(尤其是以较低的头速度)。

[0027] 可以在实际用户头部姿势数据上训练RNN方法的实施例,使得RNN可以从该数据中学习并提供准确的头部姿势预测。与恒定加速度模型(和其他一些预测算法)相比,RNN方法没有假设用于头部姿势预测的固定数学模型或对头部运动学进行假设(例如,恒定加速度)。RNN可以从实际用户的头部移动中学习,而无需假设头部移动,对头部移动进行假设会限制这样的固定模型的适用性和准确性。而且,RNN方法的实施例可以继续从在可穿戴系统的用户进行移动并且与真实和虚拟内容交互时获取的头部姿势数据中进行学习。因此,有利地,这样的实施例可以进行个性化以反映用户的头部动作的实际特征。这样的实施例在

用户患有导致头部移动(例如,震颤或抽搐)的残疾或疾病的情况下可以是特别有利的,而这些头部移动(例如,震颤或抽搐)可能不存在于标准化的默认用户群中。

[0028] 另外,头部姿势预测会受到短加速度(称为轻敲(tap))的影响,这些短加速度可能不代表真实的用户头部移动,而是由于对可穿戴系统的外部影响(例如,实际的物理轻敲或对可穿戴系统的头戴式框架的撞击)。这种“轻敲”可以导致恒定加速度模型中的预测误差,因为该模型假设这些加速度表示实际的头部移动。与之相比,循环神经网络方法的实施例表现出对轻敲的显著复原能力(resilience)以及较小的均方根误差(约37%)和q96分位数(约36%)。

[0029] 因此,利用本文描述的循环神经网络技术的可穿戴系统的实施例可以有利地提供更准确和更具鲁棒性的头部姿势预测,这导致虚拟内容的显示得以改善,并且可穿戴系统的用户的AR、MR或VR体验得到改善。

[0030] 本文描述的循环神经网络头部姿势预测技术不限于AR、MR或VR应用,并且可以用于需要人的头部姿势的其他应用,例如,游戏、医学(例如,神经病学、眼科、康复)等。

[0031] AR和VR的概述

[0032] 可穿戴系统(在本文中也称为增强现实(AR)系统)可以被配置为向用户呈现2D或3D虚拟图像。图像可以是静止图像、视频帧或视频及其组合等。可穿戴系统可以包括可穿戴设备,该可穿戴设备可以单独或组合地呈现VR、AR或MR环境以用于用户交互。可穿戴设备可以包括可穿戴显示设备,例如头戴式显示器(HMD)。可穿戴设备还可以包括腰包,该腰包可以包括中央处理单元和电池等,其中的中央处理单元用于处理可穿戴设备的一些数据处理。在一些情况下,可穿戴设备可以与增强现实设备(ARD)互换使用。

[0033] 图1描绘了具有某些虚拟现实对象以及由人观看的某些物理对象的混合现实场景的图示。在图1中,描绘了MR场景100,其中MR技术的用户看到了真实世界的公园状设置110,其以人、树、背景中的建筑物以及混凝土平台120为特征。除了这些项目之外,MR技术的用户还感知到他或她“看到”了站在真实世界平台120上的机器人雕像130,以及看起来是大黄蜂的化身的飞行着的卡通的化身角色140,即使这些元素在现实世界中不存在。

[0034] 为了使3D显示器产生真实的深度感,且更具体地,产生表面深度的模拟感,希望显示器视野中的每个点都生成与其虚拟深度相对应的适应性响应。如果对显示点的适应性响应不符合该点的虚拟深度(如由会聚和立体视的双目深度提示确定),则人眼可能会遇到适应冲突,从而导致成像不稳定,有害的眼睛疲劳,头痛,以及在缺乏适应信息的情况下,几乎完全没有表面深度。

[0035] VR、AR和MR体验可以由具有显示器的显示系统提供,在该显示器中,与多个深度平面相对应的图像被提供给观看者。图像对于每个深度平面可能是不同的(例如,提供场景或对象的稍有不同的表示),并且可能会被观看者的眼睛分别聚焦,从而有助于基于眼睛所需的适应向用户提供深度提示,为位于不同深度平面上的场景和/或基于观察不同深度平面上的不同图像特征失焦而聚焦不同图像特征。如本文其他地方所讨论的,这样的深度提示提供了对深度的可信感知。

[0036] 示例AR和VR系统和部件

[0037] 图2示出了可穿戴系统200的示例。可穿戴系统200可以包括显示器220以及支持显示器220的功能的各种机械和电子模块和系统。显示器220可以耦合至由用户、佩戴者或观

看者210可佩戴的框架230。显示器220可以位于用户210的眼睛前面。可穿戴系统的一部分(例如,显示器220)可以佩戴在用户的头部上。

[0038] 在图2中,扬声器240耦合到框架230并且位于用户的耳道附近(在一些实施例中,未示出的另一个扬声器位于用户的另一耳道附近以提供立体声/可成形的声音控制)。可穿戴系统200还可以包括面向外成像系统464(在图4中示出),其观察用户周围环境中的世界。可穿戴系统100还可以包括面向内成像系统462(图4所示),其可以跟踪用户的眼睛运动。面向内成像系统可以跟踪一只眼睛的运动或两只眼睛的运动。面向内成像系统可以附接到框架230,并且可以与处理模块260和/或270电连通,处理模块260和/或270可以处理由面向内成像系统获取的图像信息以确定例如用户210的瞳孔直径和/或眼睛的取向或眼睛姿势。

[0039] 作为示例,可穿戴系统200可以使用面向外成像系统464和/或面向内成像系统462来获取用户的姿势的图像。姿势可用于确定用户的移动或合成用户的图像。由面向外成像系统464和/或面向内成像系统462所获取的图像可以被传达给远程会话中的第二用户,以在第二用户环境中产生对用户的存在的真实感。

[0040] 可穿戴系统200可以包括头部姿势传感器235,该头部姿势传感器235可以确定用户210的头部姿势。头部姿势传感器235可以包括惯性测量单元(IMU),该惯性测量单元可以包括例如加速度计、陀螺仪或磁力计(例如,罗盘)。头部姿势传感器235可以被附接到框架230。头部姿势传感器235可以包括处理器,该处理器处理IMU数据或将IMU数据发送到本地处理和数据模块260以估计头部姿势。头部姿势可以被表示为头部取向(例如,以诸如偏航、俯仰和滚转的三个角度进行的测量)和/或头部位置(例如,通过诸如笛卡尔坐标系的三个空间坐标进行的测量)。例如,头部姿势的三个自由度(DOF)估计可以包括3DOF取向(例如,偏航、俯仰、滚转),3DOF位置(例如,笛卡尔 x, y, z)、或包括位置和取向两者(例如, x, y, z 、偏航、俯仰、滚转)的6DOF姿势。头部姿势传感器235可以检测用户的头部的运动,并且该运动被表示为取向角的(多个)旋转或空间坐标的(多个)平移。在其他实施方式中,头部姿势传感器235可以附加地或替代地包括其他传感器模式,例如电磁跟踪系统的部件。在一些实施例中,通过滤波(例如,卡尔曼滤波器或扩展卡尔曼滤波器(EKF))来处理IMU数据(例如,时间序列)以减少数据中的噪声或抖动的影响,从而提供头部姿势(3DOF或6DOF)更准确的估计。可选地,滤波可以附加地或替代地并入其他传感器数据(例如,相机图像、电磁跟踪数据等)以估计头部姿势。

[0041] 显示器220可以诸如通过有线引线或无线连接可操作地耦合250到本地数据处理模块260,该本地数据处理模块260可以以各种配置安装,诸如固定地附接到框架230,固定地附接到用户所戴的头盔或帽子,嵌入耳机中,或以其它方式可移除地附接到用户210(例如,以背包式配置,以皮带耦合式配置)。

[0042] 本地处理和数据模块260可以包括硬件处理器以及数字存储器(例如,非易失性存储器(例如,闪存)),两者均可以用于协助处理、缓存和存储数据。数据可以包括:a)从传感器(其可以例如可操作地耦合到框架230或以其他方式附接到用户210)捕获的数据,例如图像捕获设备(例如,面向内成像系统和/或面向外成像系统中的相机)、麦克风、惯性测量单元(IMU,235)(其可以包括例如加速度计,指南针,陀螺仪等)、全球定位系统(GPS)单元、无线电设备;和/或b)使用远程处理模块270或远程数据存储库280获取和/或处理的数据,可能在这样的处理或检索之后传递给显示器220。本地处理和数据模块260可以通过通信链路

262和/或264(例如,经由有线或无线通信链路)可操作地耦合到远程处理模块270和/或远程数据存储库280,使得这些远程模块可用作本地处理和数据模块260的资源。此外,远程处理模块280和远程数据存储库280可以可操作地彼此耦合。本地处理和数据模块260、远程处理模块270和远程数据存储库280可各自包括网络接口,以通过通信链路262、264提供通信。

[0043] 在一些实施例中,远程处理模块270可以包括被配置为分析和处理数据和/或图像信息的一个或多个硬件处理器。在一些实施例中,远程数据存储库280可以是数字数据存储设施,其可以通过互联网或“云”资源配置中的其他网络配置来使用。在一些实施例中,所有数据被存储并且所有计算在本地处理和数据模块中执行,从而允许来自远程模块的完全自主的使用。

[0044] 人的视觉系统是复杂的,并且提供对深度的真实感是具有挑战性的。不受理论的限制,据信,由于组合了聚散(vergence)和适应,对象的观看者可以将对象感知为三维。两只眼睛相对于彼此的聚散移动(例如,瞳孔相向或远离彼此的转动运动以收敛眼睛的视线以固定在对象上)与眼睛晶状体的聚焦(或“适应”)紧密相关联。在正常情况下,在称为“适应-聚散反射”(accommodation-vergence reflex)的关系下,改变眼睛晶状体的焦点或适应眼睛以将焦点从一个对象变化为不同距离的另一对象,将自动导致在相同距离上的聚散的匹配变化。同样,在正常情况下,聚散的改变将触发适应的匹配变化。在适应和聚散之间提供更好匹配的显示系统可以形成更逼真且舒适的三维图像模拟。

[0045] 图3示出了用于使用多个深度平面来模拟三维图像的方法的各方面。参考图3,眼睛302和304适应在z轴上距眼睛302和304各种距离处的对象,使得这些对象聚焦。眼睛302和304采取特定的适应状态,以沿z轴以不同距离进入聚焦对象。因此,可以说特定的适应状态与深度平面306中的特定一个深度平面相关联,并具有相关的焦距,使得当眼睛对于那个深度平面处于适应状态时,特定深度平面中的对象或对象的一部分处于聚焦状态。在一些实施例中,可以通过为眼睛302和304中的每一个眼睛提供图像的不同表示,并且还通过提供与每个深度平面相对应的图像的不同表示,来模拟三维图像。尽管为了图示清楚而被示出为分离的,但是将理解,例如,随着沿z轴的距离增加,眼睛302和304的视野可以重叠。另外,虽然为了便于说明而显示为平面的,但是应当理解,深度平面的轮廓可以在物理空间中弯曲,使得深度平面中的所有特征在特定的适应状态下都被眼睛聚焦。不受理论的限制,据信人眼通常可以解释有限数量的深度平面以提供深度感知。因此,通过向眼睛提供与这些有限数量的深度平面中的每一个深度平面相对应的图像的不同表示,可以实现感知深度的高度可信的模拟。

[0046] 波导堆叠组件

[0047] 图4示出了用于向用户输出图像信息的波导堆叠的示例。可穿戴系统400包括波导的堆叠或堆叠的波导组件480,其可以被用于使用多个波导432b、434b、436b、438b、400b向眼睛/大脑提供三维感知。在一些实施例中,可穿戴系统400可以对应于图2的可穿戴系统200,图4示意性地更详细地示出了可穿戴系统200的一些部分。例如,在一些实施例中,波导组件480可以集成到图2的显示器220中。

[0048] 继续参考图4,波导组件480还可包括波导之间的多个特征458、456、454、452。在一些实施例中,特征458、456、454、452可以是透镜。在其他实施例中,特征458、456、454、452可以不是透镜。而是,它们可以简单地是间隔物(例如,覆盖层和/或用于形成气隙的结构)。

[0049] 波导432b、434b、436b、438b、400b和/或多个透镜458、456、454、452可被配置为以各种级别的波前曲率或光线发散将图像信息发送到眼睛。每个波导级别可以与特定的深度平面相关联,并且可被配置为输出与该深度平面相对应的图像信息。图像注入设备420、422、424、426、428可被用于将图像信息注入波导440b、438b、436b、434b、432b,每个波导可被配置为将入射光分布在每个相应的波导上,以向眼睛410输出。光离开图像注入设备420、422、424、426、428的输出表面,并且被注入到波导440b、438b、436b、434b、432b的对应输入边缘中。在一些实施例中,可以将单个光束(例如,准直光束)注入每个波导中,以输出克隆的准直光束的整个视野,该准直光束以对应于与特定波导相关联的深度平面的特定角度(和发散量)指向眼睛410。

[0050] 在一些实施例中,图像注入设备420、422、424、426、428是分立的显示器,其各自产生图像信息以分别注入到对应的波导440b、438b、436b、434b、432b中。在一些其他实施例中,图像注入设备420、422、424、426、428是单个多路复用显示器的输出端,其可以例如经由一个或多个光导管(例如,光纤电缆)将图像信息输送到图像注入设备420、422、424、426、428中的每一个。

[0051] 控制器460控制堆叠的波导组件480和图像注入设备420、422、424、426、428的操作。控制器460包括编程(例如,存储在非暂时性计算机可读介质中的可执行指令),该编程调制定时和向波导440b、438b、436b、434b、432b提供图像信息。在一些实施例中,控制器460可以是单个集成设备,或者是通过有线或无线通信信道连接的分布式系统。在一些实施例中,控制器460可以是处理模块260和/或270(在图2中示出)的一部分。控制器460可以包括硬件处理器、专用集成电路(ASIC)、或其他可编程计算机组件。

[0052] 控制器460可以控制要显示给用户210的虚拟内容的渲染。控制器460可以从头部姿势传感器235接收输入,以帮助确定用户210看向的方向,从而部分地确定显示什么虚拟内容以及何时显示。例如,如本文中进一步描述的,控制器460和显示器220向用户渲染虚拟内容需要花费一段时间(例如,在一些实施例中,大约20ms至30ms)。控制器460可以使用对用户的头部未来(例如,在时间上向前约20ms至30ms)在哪的预测位置的估计,以便它可以确定用户将观看的方向以及相应地用户在AR/VR/MR环境中(在该方向上或附近)应该看到的虚拟内容。控制器460可以从头部姿势传感器235接受头部姿势传感器数据并计算预测头部姿势,或者在其他实施例中,头部姿势传感器235(在一些实施例中可以包括加速度计235a和陀螺仪235b)可以执行头部姿势预测并将预测头部姿势提供给控制器460。因此,在各个实施例中,控制器460、头部姿势传感器235、或本地处理和数据模块260可以获取头部姿势数据(例如,来自加速度计235a或陀螺仪235b),并使用本文描述的循环神经网络技术来计算预测头部姿势(例如,角度的或平移的)。在一些实施例中,头部姿势传感器235可以附加地或替代地包括磁力计。

[0053] 波导440b、438b、436b、434b、432b可被配置为通过全内反射(TIR)在每个相应波导内传播光。波导440b、438b、436b、434b、432b每个可以是平面的或具有另一种形状(例如,弯曲的),并且可以具有主要的顶部和底部表面以及在那些主要的顶部和底部表面之间延伸的边缘。在所示的配置中,波导440b、438b、436b、434b、432b可各自包括光提取光学元件440a、438a、436a、434a、432a,光提取光学元件440a、438a、436a、434a、432a被配置为通过使光重定向、在各个波导内传播从而从波导中提取光以将图像信息输出到眼睛410。提取的光

也可以称为出耦合光,并且光提取光学元件也可以称为出耦合光学元件。提取的光束由波导在波导中传播的光撞击光重定向元件的位置处输出。光提取光学元件(440a、438a、436a、434a、432a)可以例如是反射和/或衍射光学特征。虽然为了便于描述和绘图清楚起见,示出了设置在波导440b、438b、436b、434b、432b的底部主表面上,但是在一些实施例中,光提取光学元件440a、438a、436a、434a、432a可以设置在顶部主表面和/或底部主表面上,和/或可以直接设置在波导440b、438b、436b、434b、432b的体积中。在一些实施例中,光提取光学元件440a、438a、436a、434a、432a可以形成在材料层中,该材料层附接到透明基板上以形成波导440b、438b、436b、434b、432b。在一些其他实施例中,波导440b、438b、436b、434b、432b可以是单片材料,并且光提取光学元件440a、438a、436a、434a、432a可以形成在该片材料的表面上和/或内部中。

[0054] 继续参考图4,如本文所述,每个波导440b、438b、436b、434b、432b被配置为输出光以形成与特定深度平面相对应的图像。例如,最靠近眼睛的波导432b可被配置为将注入到这种波导432b中的准直光传送到眼睛410。准直光可以表示光学无限远焦平面。下一个上行(up)波导434b可被配置为在准直光到达眼睛410之前发出通过第一透镜452(例如,负透镜)的准直光。第一透镜452可被配置为产生轻微的凸波前曲率,从而眼睛/大脑将来自该下一上行波导434b的光解释为来自第一焦平面,该第一焦平面从光学无限远处向内更靠近朝向眼睛410。类似地,第三上行波导436b在到达眼睛410之前将其输出光通过第一透镜452和第二透镜454。第一透镜452和第二透镜454的组合光功率可被配置为产生另一增量的波前曲率,使得眼睛/大脑将来自第三波导436b的光解释为来自第二焦平面,该第二焦平面比来自下一个上行波导434b的光从光学无穷远向内更靠近人。

[0055] 其他波导层(例如,波导438b、440b)和透镜(例如,透镜456、458)被类似地配置,堆叠中的最高波导440b通过其与眼睛之间的所有透镜发送其输出,用于表示最靠近人的焦平面的聚合(aggregate)焦度。在观看/解释来自堆叠波导组件480的另一侧上的世界470的光时,为了补偿透镜堆叠458、456、454、452,可以在堆叠的顶部设置补偿透镜层430,以补偿下面的透镜堆叠458、456、454、452的聚合焦度。这样的配置提供与可用的波导/透镜对一样多的感知焦平面。波导的光提取光学元件和透镜的聚焦方面都可以是静态的(例如,不是动态的或电激励的)。在一些替代实施例中,其中一个或两者在使用电激励特征时可以是动态的。

[0056] 继续参考图4,光提取光学元件440a、438a、436a、434a、432a可以被配置为既将光重定向到其各自的波导之外,又对于与波导相关联的特定深度平面以适当的发散或准直量输出该光。结果,具有不同相关联的深度平面的波导可以具有光提取光学元件的不同配置,该光提取光学元件根据相关联的深度平面以不同的发散量输出光。在一些实施例中,如本文所讨论的,光提取光学元件440a、438a、436a、434a、432a可以是体积或表面特征,其可被配置为以特定角度输出光。例如,光提取光学元件440a、438a、436a、434a、432a可以是体积全息图、表面全息图和/或衍射光栅。在2015年6月25日公开的美国专利公开No.2015/0178939中描述了诸如衍射光栅的光提取光学元件,其通过引用全部并入在此。

[0057] 在一些实施例中,光提取光学元件440a、438a、436a、434a、432a是形成衍射图案或衍射光学元件(在本文中也称为“DOE”)的衍射特征。优选地,DOE具有相对较低的衍射效率,使得通过与DOE的每次交互,光束的仅一部分光朝着眼睛410偏转,而其余部分则通过全内

反射继续移动通过波导。携带图像信息的光因此被分成多个相关的出射光束,该出射光束在多个位置处离开波导,并且该结果对于在波导内反弹的该特定准直光束是朝向眼睛304的相当均匀图案的出射发射。

[0058] 在一些实施例中,一个或多个DOE可以在它们主动衍射的“开启”状态和它们不明显衍射的“关闭”状态之间切换。例如,可切换的DOE可以包括聚合物分散液晶层,其中,微滴(microdrop)在主体介质(host medium)中包含衍射图案,并且微滴的折射率可以切换为与主体材料的折射率基本匹配(在这种情况下,该图案不会明显衍射入射光),或者可以将微滴切换到与主体介质的折射率不匹配的折射率(在这种情况下,该图案会主动衍射入射光)。

[0059] 在一些实施例中,深度平面和/或景深的数量和分布可以基于观看者的眼睛的瞳孔大小和/或取向而动态地变化。景深可能与观看者的瞳孔大小成反比变化。结果,随着观看者眼睛的瞳孔大小减小,景深增加,使得由于该平面的位置超出眼睛的焦深而无法辨别的平面可能变得可辨别,并且随着瞳孔大小的减小和景深的相应增加表现为更聚焦。同样,随着瞳孔大小减小,可以减少用于向观看者呈现不同图像的间隔开的深度平面的数量。例如,在处于一个瞳孔大小下,不将眼睛的适应性从一个深度平面调整到另一个深度平面,则观看者可能无法清楚地感知第一深度平面和第二深度平面两者的细节。但是,这两个深度平面可以在不改变适应性的情况下,对于处于另一瞳孔大小的用户同时充分地聚焦。

[0060] 在一些实施例中,显示系统可以基于对瞳孔大小和/或取向的确定,或者基于接收到指示特定瞳孔大小和/或取向的电信号,来改变接收图像信息的波导的数量。例如,如果用户的眼睛不能区分与两个波导相关联的两个深度平面,则控制器460可以被配置或编程为停止向这些波导之一提供图像信息。有利地,这可以减少系统上的处理负担,从而增加系统的响应性。在波导的DOE可在开启和关闭状态之间切换的实施例中,当波导确实接收到图像信息时,DOE可被切换为关闭状态。

[0061] 在一些实施例中,可能希望使出射光束满足直径小于观看者眼睛的直径的条件。但是,鉴于观看者瞳孔大小的可变性,满足此条件可能是具有挑战性的。在一些实施例中,通过响应于观看者的瞳孔大小的确定而改变出射光束的大小,在宽范围的瞳孔大小上满足该条件。例如,随着瞳孔大小的减小,出射光束的大小也可以减小。在一些实施例中,可以使用可变孔径来改变出射光束的大小。

[0062] 可穿戴系统400可以包括对世界470的一部分进行成像的面向外成像系统464(例如,数码相机)。世界470的该部分可以被称为视场(FOV),成像系统464有时被称为FOV相机。观看者可用于观看或成像的整个区域可以被称为能视域(FOR)。FOR可以包括围绕可穿戴系统400的立体角的 4π 立体弧度。在可穿戴系统400的一些实施方式中,FOR可以包括围绕显示系统400的用户的基本上所有立体角,因为用户可以移动其头部和眼睛来观看用户周围的对象(用户的前方、后方、上方、下方或侧面)。从面向外成像系统464获得的图像可以用于跟踪用户做出的手势(例如,手部或手指手势),检测用户前方的世界470中的对象等。头部姿势传感器235可以用于确定FOR内的头部姿势的估计,该头部姿势的估计可用于估计用户的FOV。

[0063] 可穿戴系统400还可以包括面向内成像系统466(例如,数码相机),其观察用户的运动,例如眼睛运动和面部运动。面向内成像系统466可以用于捕获眼睛410的图像,以确定

眼睛304的瞳孔的大小和/或取向。面向内成像系统466可以用于获得确定用户正在观看的方向(例如,眼睛姿势)所使用的图像,或用于用户的生物特征识别(例如,通过虹膜识别)所使用的图像。在一些实施例中,可以针对每只眼睛使用至少一个相机,以独立地分别确定每只眼睛的瞳孔大小和/或眼睛姿势,从而允许向每只眼睛呈现的图像信息可以动态地适应该只眼睛。在一些其他实施例中,仅确定单只眼睛410的瞳孔直径和/或取向(例如,每双眼睛仅使用单个相机),并且假设该瞳孔直径和/或取向对于用户的两只眼睛是相似的。可以分析由面向内成像系统466获得的图像以确定用户的眼睛姿势和/或心情,可穿戴系统400可以使用该眼睛姿势和/或心情来决定应该向用户呈现哪些音频或视觉内容。如上所述,可穿戴系统400还可以使用例如头部姿势传感器235来确定头部姿势(例如,头部位置或头部取向)。

[0064] 可穿戴系统400可以包括用户输入设备466,用户可以通过该用户输入设备466向控制器460输入命令以与可穿戴系统400交互。例如,用户输入设备466可以包括触控板、触摸屏、游戏杆、多自由度(DOF)控制器、电容感测设备、游戏控制器、键盘、鼠标、方向盘(D-pad)、手杖、触觉设备、图腾、智能手机、智能手表、平板设备等,其组合等。多DOF控制器可以在控制器的一些或所有可能的平移(例如,向左/向右、向前/向后、或向上/向下)或旋转(例如,偏航、俯仰或滚转)上感测用户输入。用户可以通过以下来与用户输入设备466或他或她的环境中的对象(例如,虚拟或物理对象)进行交互:例如点击鼠标,敲击触摸板,在触摸屏上滑动,在电容式按钮上盘旋或触摸电容式按钮,按下键盘或游戏控制器(例如,5向d-pad)上的按键,将游戏杆、手杖或图腾指向对象,按下遥控器上的按钮,或与用户输入设备进行其他交互。用户输入设备466的致动可以使可穿戴系统执行用户界面操作,诸如例如,显示与对象相关联的虚拟用户界面菜单,在游戏中使用户的化身形成动画等。用户输入设备466可以被配置为发光。发光图案可以表示与用户环境中的对象相关联的信息,用户与用户输入设备466或可穿戴设备等的交互。

[0065] 在一些情况下,用户可以使用手指(例如,拇指)在触敏输入设备上按下或滑动以向可穿戴系统400提供输入(例如,向由可穿戴系统400提供的用户界面提供用户输入)。在可穿戴系统400的使用期间,用户输入设备466可以由用户的手部握持。用户输入设备466可以与可穿戴系统400进行有线或无线通信。用户输入设备466可以包括本文所述的图腾的实施例。图腾可以包括触摸表面,该触摸表面可以允许用户通过沿着轨迹滑动或轻敲等来致动图腾。

[0066] 利用深度预测器神经网络头部姿势预测器的示例渲染管线

[0067] 图5是示意性示出用于向可穿戴系统的用户显示虚拟内容的渲染管线500的一部分的示例的框图。渲染管线可以分别由参考图2和图4描述的可穿戴系统200或400来实现。

[0068] 在该示例中,提供头部姿势数据的头部姿势传感器235包括加速度计235a和陀螺仪235b。可选地,可以对头部姿势数据进行滤波,例如,使用扩展的卡尔曼滤波器(EKF)进行滤波,以提供对于用户的头部的速度和加速度更平滑的估计。头部姿势数据被输入到姿势预测器510,该姿势预测器510输出未来时间范围内的预测头部姿势530。未来时间范围可以与渲染管线生成虚拟内容并向系统200、400的用户显示虚拟内容所花费的时间相当,该时间例如可以是未来的从约10ms至约50ms。因此,渲染管线500可以在用户看向虚拟内容或靠近虚拟内容时向用户显示虚拟内容,这减少了渲染延迟并改善了用户体验。

[0069] 姿势预测器510可以由包括非暂时性存储器的硬件处理器来实现,例如本地处理和数据模块260、远程处理模块270和远程数据存储库280、控制器460或其他电子计算机电路。

[0070] 在各种实施例中,预测头部姿势530可以是包括头部取向(例如,偏航角、俯仰角和滚转角)的3DOF头部姿势,包括头部位置(例如,笛卡尔x、y和z坐标)的3DOF头部姿势,或包括头部取向和头部位置二者的6DOF姿势。

[0071] 预测头部姿势530被输入到控制器460(参考图4描述),该控制器460调节虚拟内容向可穿戴系统200、400的显示器220的定时和提供。

[0072] 姿势预测器510可以包括本文描述的深度预测器神经网络架构520的实施例(例如,参见图6和图7)。深度预测器神经网络架构520可以包括循环神经网络(RNN),该循环神经网络比恒定加速度预测模型更准确或更具有鲁棒性地执行姿势预测。在一些实施方式中,深度预测器神经网络架构520用于预测用户的头部的角度取向,并且恒定加速度模型用于预测用户的头部的空间平移。

[0073] 示例深度预测器神经网络架构

[0074] 图6是示意性示出用于头部姿势预测的深度预测器神经网络520的示例的框图。神经网络520的此示例包括循环神经网络(RNN)610和全连接(FC)层620。

[0075] 如上所述,可以将头部姿势数据(例如,来自加速度计235a和陀螺仪235b的头部姿势数据)输入到RNN 610。在一些实施例中,输入数据包括来自头部姿势传感器235的原始读数,作为补充或替代,利用头部速度和加速度的估计。例如,可以通过利用例如卡尔曼或扩展卡尔曼滤波器进行滤波来处理头部姿势传感器235的数据,以提供用户头部的速度和加速度。输入的头部姿势数据可以是一时间段(例如10ms至1s)上的时间序列。

[0076] 头部姿势数据被输入到RNN 610。RNN是具有“记忆”的神经网络,因为神经网络的各层利用了来自RNN先前层的状态信息。实际上,RNN具有对时间序列的数据进行操作的重复模块的环路,这允许时间序列中的信息得以保留(persist)(例如,被记住)并由RNN的未来层起作用。环路612示意性地示出了RNN 610的循环性质,其表示来自前一环路的输出(与其他输入一起)可以在RNN中的未来环路被处理。RNN 610因此使用其内部状态(例如,存储器)来处理时间序列的头部姿势数据以预测未来头部姿势530。

[0077] RNN 610的输出可以被馈送到全连接(FC)层620,全连接(FC)层620具有到RNN 610的最后层的所有神经元的连接。FC层620生成预测未来头部姿势530,例如3DOF(角度),3DOF(平移)或6DOF(旋转和平移)。

[0078] 在网络架构520的一些实施方式中,RNN 610的隐藏层包括32个神经元,并且FC层620提供3个(用于3DOF)或6个(用于6DOF)输出。在一些实施例中,FC层620包括修正线性单元(ReLU)或双曲正切(tanh)激活函数。可以使用其他数量的隐藏层神经元或其他类型的激活函数。例如,在一些实施方式中,RNN 610的隐藏层可以包括64个神经元。

[0079] 图7是示意性示出用于头部姿势预测的深度预测器神经网络520的另一示例的框图。在该示例中,神经网络架构包括RNN,该RNN包括三个堆叠的循环神经网络单元。这些循环神经网络单元是长短期记忆(LSTM)网络:LSTM 0 700,LSTM 1 701和LSTM 2 702。三个LSTM单元700、701、702堆叠在一起,以及来自LSTM 0 700的输出被馈送到LSTM 1 701,来自LSTM 1 701的输出被馈送到LSTM 2 702。LSTM 700、701、702的循环特性用箭头示意性地说明,这些

箭头循环回到每个LSTM单元,分别用 h_0 、 h_1 、 h_2 表示,示意性地指示每个LSTM单元的循环和记忆。

[0080] 三个LSTM单元702中的最后一个被连接到提供预测头部姿势530的FC层620。

[0081] LSTM单元特别适合于学习时间序列的输入数据(例如,来自头部姿势传感器235a、235b的时间序列的头部姿势数据)中的长期依赖性。在示例LSTM架构中,LSTM单元(cell)包括记忆(memory)单元、输入门、输出门和遗忘门。记忆单元存储长时间段或短时间段内的值(或状态)。这可以通过使用记忆单元的恒等映射(或非(no))激活函数来实现。LSTM门例如使用逻辑函数来计算激活。输入门可以控制新值流入单元的程度,遗忘门可以控制值保留在单元中的程度,以及输出门可以控制单元中的值被用于计算LSTM单元的输出激活的程度。这些门之间可以具有进出的连接。一些连接可以是循环的。这些连接的权重(能在训练期间学习)可以用于指导门的运算。每个门可以具有可能来自LSTM单元外部的其他单元的自己的参数,例如权重和偏差(bias)。

[0082] 在图7中所示的示例网络架构520中,使用了三个堆叠的LSTM单元。在一些实施方式中,可以使用单个LSTM单元。在其他实施方式中,可以使用2、4、5、6或更多个堆叠的LSTM单元。对于头部姿势预测,使用用户头部姿势数据的实验表明,三个LSTM单元提供的性能优于一个、两个或其他数量的单元。例如,增加LSTM单元的数量可以增加架构520学习头部姿势时间序列数据中的模式的能力,但是这种增加的预测能力是以训练和运行网络的增加的计算成本为代价的。发现三个LSTM单元在头部姿势预测精度和计算效率方面提供了出色的性能。应当理解,可以采用不同数量的LSTM单元以实现不同的目的。在一些示例中,可以使用更少的LSTM单元(例如,一个或两个LSTM单元),但是所采用的每个LSTM单元可以包括更多数量的神经元、连接或两者。类似的设计原则也可以反映在FC层的架构中。

[0083] 在深度预测器RNN架构的其他实施方式中,除了或代替图7中所示的LSTM单元,可以使用其他类型的循环单元。例如,部分或全部LSTM单元700、701、702可以用门控循环单元(GRU)或关注(attention)单元来代替。在其他实施方式中,RNN 610可以包括门控反馈RNN、时钟(CW)RNN或具有记忆的任何其他神经网络。

[0084] 训练深度预测器网络的示例

[0085] 使用从佩戴可穿戴显示器200、400的实施例的用户所收集的数据来训练深度预测器RNN架构520的实施方式。头部姿势传感器235包括以1000Hz运行的IMU,并进行扩展卡尔曼滤波以500Hz提供角速度和角加速度数据。头部姿势时间序列数据中的数据点的数量在长度上在从10,000至100,000范围内。训练数据包括126个用户头部移动的记录,其中102个记录用于训练,24个记录用于验证。使用多对多训练技术对深度预测器架构的实施方式进行了训练,并利用使用L2范数损失函数随时间的截断反向传播。合计了N个过去步骤的梯度(其中 $N=200$)并进行了优化,而沿着序列保持了隐藏状态。

[0086] 对于深度预测器网络的这种实施方式,训练的输出姿势是在30ms时间范围内所预测的3DOF角度头部姿势。对于其他时间范围,可以相对于30ms时间范围内的值对来自网络520的输出进行线性外推。

[0087] 深度预测器网络头部姿势预测的示例

[0088] 图8是示意性示出在时间预测范围 dt 上的头部姿势预测的示例的曲线图。如上所述,预测范围 dt 可以大致对应于可穿戴系统200、400的渲染管线500中的渲染延迟。例如,预

测范围 dt 可以在从大约10ms至大约50ms的范围内,例如,30ms。

[0089] 在图8中,用户的实际头部姿势随时间的变化情况由线804示意性地描绘(在这种情况下,出于说明目的为直线)。头部姿势传感器数据由线808示意性地示出,并且示出了由于噪声、抖动、EKF滤波中的不精确性等导致的围绕实际头部姿势线804的变化。在时间 t_1 ,头部姿势传感器数据指示头部姿势在点814。姿势预测器510可以利用深度预测器神经网络520和时间序列的头部姿势数据(例如,来自线808的时间 t_1 之前的时间序列),来预测在时间 t_1+dt 的头部姿势。该预测可以包括角度偏移、空间平移、或角度偏移和空间平移两者,如箭头812示意性示出。姿势预测器510将预测812应用于时间 t_1 处的头部姿势813,来计算在时间 t_1+dt 处的预测头部姿势816。在图8中还示出了在时间 t_2 处的类似计算。预测允许姿势预测器510跟踪实际的头部姿势804。

[0090] 用于使用深度预测器神经网络预测头部姿势的示例方法

[0091] 图9是示出用于利用深度预测器神经网络架构520的实施例来预测头部姿势的方法900的示例的流程图。方法900可以由可穿戴显示系统200、400来执行,例如由本地处理和数据模块260、姿势预测器510、控制器460、或由另一硬件处理器来执行。

[0092] 在框904处,方法900接收头部姿势数据。例如,头部姿势数据可以包括来自头部姿势传感器235(例如,加速度计235a和陀螺仪235b)的数据。另外或可替代地,头部姿势数据可以包括用户的头部的速度或加速度。例如,可以利用扩展卡尔曼滤波器(EKF)对来自头部姿势传感器235(或其他传感器,例如电磁跟踪传感器或来自相机的图像)的数据进行滤波,以提供对头部速度或加速度的估计。

[0093] 在框908处,方法900使用深度预测器神经网络架构520的实施例来预测在未来的预测范围(例如,未来从约10ms到约50ms)内的头部姿势(例如,参见图7和图8)。头部姿势可以是包括头部取向(例如,偏航角、俯仰角和滚转角)的3DOF头部姿势,包括头部位置(例如,笛卡尔 x 、 y 和 z 坐标)的3DOF头部姿势,或包括头部取向和头部位置两者的6DOF姿势。在一些实施例中,方法900使用深度预测器神经网络架构520来预测角度头部姿势,并且针对头部平移使用另一个预测模型(例如,恒定加速度)。深度预测器神经网络架构520可以包括RNN,该RNN包括一个或多个记忆单元的堆叠,记忆单元例如LSTM单元、GRU等。在一些实施方式中,利用了具有三个LSTM或GRU的堆叠。深度预测器神经网络架构520可以包括FC层,该FC层接受来自RNN的输出并计算预测头部姿势。

[0094] 在框912处,方法900利用预测头部姿势向用户渲染虚拟内容。例如,预测头部姿势提供关于用户在未来的时间段(例如10ms至50ms)将会看向哪里的信息。渲染管线500可以计算用户可能正在看向的虚拟内容,并且可以在用户实际转向到预测姿势之前开始渲染过程。因此,方法900的实施例可以减少渲染延迟,提供虚拟内容的及时渲染,并实现令人愉悦的用户体验。

[0095] 附加方面

[0096] 方面1.一种可穿戴系统,包括:

[0097] 显示器,其被配置为被设置在所述可穿戴系统的佩戴者的眼睛前方,所述显示器被配置为向所述可穿戴系统的所述佩戴者显示虚拟内容;

[0098] 头部姿势传感器,其被配置为提供头部姿势数据;

[0099] 非暂时性存储器,其被配置为存储头部姿势预测循环神经网络;

[0100] 与所述头部姿势传感器、所述显示器、以及所述非暂时性存储器通信的硬件处理器,所述硬件处理器被编程为:

[0101] 接收所述头部姿势数据;

[0102] 将所述头部姿势数据输入到所述头部姿势预测循环神经网络;

[0103] 执行所述头部姿势预测循环神经网络以输出在时间范围内的预测头部姿势;以及

[0104] 使所述显示器至少部分地基于在所述时间范围内的所述预测头部姿势来渲染所述虚拟内容。

[0105] 方面2.根据方面1所述的可穿戴系统,其中,所述显示器被配置为在多个深度平面呈现所述虚拟内容。

[0106] 方面3.根据方面1或方面2所述的可穿戴系统,其中,所述头部姿势传感器包括惯性测量单元。

[0107] 方面4.根据方面3所述的可穿戴系统,其中,所述惯性测量单元包括加速度计和陀螺仪。

[0108] 方面5.根据方面1至4中任一项所述的可穿戴系统,其中,所述头部姿势数据包括速度或加速度。

[0109] 方面6.根据方面5所述的可穿戴系统,其中,来自所述头部姿势传感器的输出由扩展卡尔曼滤波器来滤波以提供所述头部姿势数据。

[0110] 方面7.根据方面1至6中任一项所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括循环神经网络(RNN)和全连接(FC)层。

[0111] 方面8.根据方面1至7中任一项所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括多个重复模块。

[0112] 方面9.根据方面8所述的可穿戴系统,其中,所述多个重复模块中的至少一个重复模块包括:长短期记忆单元(LSTM)或门控循环单元(GRU)。

[0113] 方面10.根据方面1至9中任一项所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括:包括一个或多个长短期记忆(LSTM)单元的堆叠或包括一个或多个门控循环单元(GRU)的堆叠。

[0114] 方面11.根据方面10所述的可穿戴系统,其中,所述头部姿势预测循环神经网络包括三个LSTM单元或三个GRU。

[0115] 方面12.根据方面1至11中任一项所述的可穿戴系统,其中,所述时间范围在从10ms至50ms的范围内。

[0116] 方面13.根据方面1至11中任一项所述的可穿戴系统,其中,所述时间范围在从20ms至40ms的范围内。

[0117] 方面14.根据方面1至13中任一项所述的可穿戴系统,其中,所述预测头部姿势包括3自由度DOF姿势。

[0118] 方面15.根据方面14所述的可穿戴系统,其中,所述3DOF姿势包括头部取向。

[0119] 方面16.根据方面14或方面15所述的可穿戴系统,其中,所述3DOF姿势包括头部位置。

[0120] 方面17.根据方面1至16中任一项所述的可穿戴系统,其中,所述预测头部姿势包括6自由度DOF姿势,所述6DOF姿势包括取向和位置。

[0121] 方面18.根据方面1至17中任一项所述的可穿戴系统,其中,所述预测头部姿势包括头部取向,并且所述硬件处理器还被编程为:使用预测性非神经网络模型来预测在所述时间范围内的头部位置。

[0122] 方面19.根据方面18所述的可穿戴系统,其中,所述预测性非神经网络模型包括恒定加速度模型。

[0123] 方面20.一种用于渲染虚拟内容的方法,所述方法包括:

[0124] 在包括计算机硬件的渲染管线的控制下:

[0125] 接收针对用户的头部的头部姿势数据;

[0126] 至少使用循环神经网络(RNN)和所述头部姿势数据,预测在预测范围内的预测头部姿势;以及

[0127] 至少部分地基于所述预测头部姿势,渲染在所述预测范围内的所述虚拟内容。

[0128] 方面21.根据方面20所述的方法,其中,所述头部姿势数据包括以下一项或多项的时间序列:(1)来自惯性测量单元(IMU)的数据;(2)针对所述用户的所述头部的速度数据或加速度数据;或(1)和(2)两者。

[0129] 方面22.根据方面21所述的方法,还包括:对数据进行滤波以提供所述速度数据或所述加速度数据。

[0130] 方面23.根据方面22所述的方法,其中,利用扩展卡尔曼滤波器(EKF)来执行所述滤波。

[0131] 方面24.根据方面20至23中任一项所述的方法,其中,所述RNN包括:包括一个或多个长短期记忆(LSTM)单元的堆叠或包括一个或多个门控循环单元(GRU)的堆叠。

[0132] 方面25.根据方面24所述的方法,其中,所述RNN包括具有三个LSTM单元的堆叠或具有三个GRU的堆叠。

[0133] 方面26.根据方面20至25中任一项所述的方法,其中,所述RNN被连接到被配置为输出所述预测头部姿势的全连接(FC)层。

[0134] 方面27.根据方面20至26中任一项所述的方法,其中,所述预测头部姿势包括所述用户的所述头部的取向,并且所述方法还包括:利用恒定加速度模型来预测在所述预测范围内的所述用户的所述头部的的位置。

[0135] 方面28.根据方面20至27中任一项所述的方法,其中,所述预测范围在从10ms至50ms的范围内。

[0136] 方面29.一种可穿戴显示系统,包括:显示器,其被配置为设置在所述可穿戴系统的佩戴者的眼睛前方,所述显示器被配置为向所述可穿戴系统的所述佩戴者显示虚拟内容;头部姿势传感器,其被配置为提供头部姿势数据;非暂时性存储器,其被配置为存储头部姿势预测循环神经网络;与所述头部姿势传感器、所述显示器和所述非暂时性存储器通信的计算硬件,所述计算硬件被编程为包括渲染管线,所述渲染管线被编程为执行方面20至28中任一项的所述方法。

[0137] 其它考虑

[0138] 本文描述和/或附图中描绘的每个过程、方法和算法可以体现在由一个或多个被配置为执行专门的和特定的计算机指令的物理计算系统、硬件计算机处理器、专用电路和/或电子硬件中以及全部或部分地由其自动化。例如,计算系统可以包括用特定计算机指令

编程的通用计算机(例如服务器)或专用计算机、专用电路等。代码模块可被编译并链接到可执行程序中、安装在动态链接库中,或者可以用解释性编程语言来编写。在一些实现中,特定的操作和方法可以由特定于给定功能的电路执行。

[0139] 此外,本公开的功能的某些实现在数学上、计算上或技术上都足够复杂,以致于可能需要专用硬件或一个或多个物理计算设备(利用适当的专用可执行指令)来执行功能,例如由于所涉及计算的量或复杂性,或为了实质上实时地提供结果。例如,视频可以包括许多帧,每个帧具有数百万个像素,并且需要专门编程的计算机硬件来处理视频数据,以在商业上合理的时间量内提供所需的图像处理任务或应用。此外,可以在20毫秒至50毫秒(例如20Hz至50Hz)的时间尺度上反复预测头部姿势,以便可穿戴系统的渲染引擎可以渲染虚拟内容,该虚拟内容提供足够低的延迟以提供令人满意的用户体验以及可能需要硬件处理电路来执行本文所述的循环神经网络头部姿势预测。

[0140] 代码模块或任何类型的数据可以存储在任何类型的非暂时性计算机可读介质上,例如物理计算机存储器,包括硬盘驱动器、固态存储器、随机存取存储器(RAM)、只读存储器(ROM)、光盘、易失性或非易失性存储设备、它们和/或类似存储器的组合等。方法和模块(或数据)也可以作为生成的数据信号(例如作为载波或其他模拟或数字传播信号的一部分)在包括基于无线的和基于有线/电缆的介质的各种计算机可读传输介质上进行发送,并且可以采用多种形式(例如作为单个或多路复用模拟信号的一部分,或作为多个离散数字分组或帧)。所公开的过程或过程步骤的结果可以永久地或以其他方式存储在任何类型的非暂时性有形计算机存储器中,或者可以经由计算机可读传输介质进行传递。

[0141] 本文描述的和/或在附图中描绘的流程图中的任何过程、框、状态、步骤或功能应被理解为潜在地表示代码模块、代码段或代码部分,其包括用于在流程中实现特定功能(例如逻辑或算术)或步骤的一个或多个可执行指令。各种过程、框、状态、步骤或功能可以组合、重新布置、添加到本文提供的说明性示例、从本文提供的说明性示例中删除、修改或以其他方式改变。在一些实施例中,附加的或不同的计算系统或代码模块可以执行本文描述的一些或全部功能。本文描述的方法和过程也不限于任何特定的顺序,并且与之相关的框、步骤或状态可以以适当的其他顺序(例如串行、并行或以某些其他方式)来执行。可以将任务或事件添加到所公开的示例实施例中或从中删除。此外,本文描述的实现中的各种系统组件的分离是出于说明的目的,并且不应被理解为在所有实现中都需要这种分离。应当理解,所描述的程序组件、方法和系统通常可以被集成在一起在单个计算机产品中或包装到多个计算机产品中。许多实现变型是可能的。

[0142] 可以在网络(或分布式)计算环境中实现过程、方法和系统。网络环境包括企业范围的计算机网络、企业内网、局域网(LAN)、广域网(WAN)、个人区域网(PAN)、云计算网络、众包计算网络、卫星网络、互联网和万维网。该网络可以是有线或无线网络或任何其他类型的通信网络。

[0143] 本公开包括可以使用主题设备执行的方法。该方法可以包括提供这种合适的设备的动作。这样的提供可以由终端用户执行。换句话说,“提供”动作仅要求终端用户获得、访问、接近、定位、设置、激活、加电或以其他方式提供以提供本主题方法中的必要设备。本文所列举的方法的元素、动作、步骤或框可以以逻辑上可能的所列举事件的任何顺序以及事件所列举的顺序来执行。

[0144] 本公开的系统和方法各自具有若干创新方面,其中,没有一个对本文公开的期望属性负有单独责任或要求。上述的各种特征和过程可以彼此独立地使用,或者可以以各种方式组合。所有可能的组合和子组合旨在落入本公开的范围。对本公开中描述的实现的修改对于本领域技术人员而言是显而易见的,并且在不脱离本公开的精神或范围的情况下,本文中定义的一般原理可以应用于其他实现。因此,权利要求书无意限于本文中所示的实现,而是应被赋予与本文中所揭示的本发明、原理及新颖特征一致的最广范围。

[0145] 在本说明书中在单独的实现的上下文中描述的某些特征也可以在单个实现中组合实现。相反,在单个实现的上下文中描述的各种特征也可以分别在多个实现中或以任何合适的子组合来实现。而且,尽管以上可以将特征描述为以某些组合起作用并且甚至最初如此宣称,但是在某些情况下可以从组合中切除所要求保护的组合中的一个或多个特征,并且所要求保护的组合可以针对子组合或子组合的变体。对于每个实施例,没有单个特征或一组特征是必要的或必不可少的。

[0146] 这里使用的条件语言,尤其例如“能够”、“会”、“可能”、“可以”、“例如”等,除非另有明确说明,否则在所使用的上下文中理解为通常意在传达某些实施例包括而其他实施例不包括某些特征、元素和/或步骤。因此,这样的条件语言通常不旨在暗示特征、元素和/或步骤以任何方式对于一个或多个实施例是必需的,或者一个或多个实施例必然包括用于在有或没有作者输入或提示的情况下决定这些特征、元素和/或步骤是否在任何特定实施例中被包括或将执行的逻辑。术语“包括”、“包含”、“具有”等是同义词,以开放式方式包含地使用,并且不排除附加要素、特征、动作、操作等。而且,术语“或”以其包含的含义使用(而不是以其排他的含义使用),因此例如在用于连接元素列表时,术语“或”表示列表中的一个、一些或全部元素。另外,在本申请和所附权利要求书中使用的“一”、“一个”和“该”应被解释为表示“一个或多个”或“至少一个”,除非另有说明。除本文中明确定义外,在保持权利要求有效性的同时,本文中使用的所有技术和科学术语应尽可能广泛地赋予通常理解的含义。

[0147] 如本文所使用的,指代项目列表中的“至少一个”的短语是指那些项目的任何组合,包括单个成员。例如,“A、B或C中的至少一个”旨在涵盖:A、B、C、A和B、A和C、B和C以及A、B和C。除非另外特别说明,否则诸如短语“X、Y和Z中的至少一个”之类的词语应与上下文一起理解,该上下文通常用于传达项目、术语等可以是X、Y或Z中的至少一个。因此,这种联合语言通常不旨在暗示某些实施例要求X中的至少一个、Y中的至少一个和Z中的至少一个存在。

[0148] 类似地,尽管可以以特定顺序在附图中描绘操作,但是要认识到,不需要以所示的特定顺序或相继顺序来执行这样的操作,或者不需要执行所有示出的操作来实现理想的结果。此外,附图可以以流程图的形式示意性地描绘一个或多个示例过程。然而,未示出的其他操作可以结合在示意性示出的示例方法和过程中。例如,可以在任何所示操作之前、之后、同时或之间执行一个或多个附加操作。另外,在其他实现中,操作可以重新布置或重新排序。在某些情况下,多任务和并行处理可能是有利的。此外,在上述实现中的各种系统组件的分离不应被理解为在所有实现中都需要这种分离,并且应当理解,所描述的程序组件和系统通常可以集成在单个软件产品中或打包到多个软件产品中。另外,其他实现在所附权利要求的范围内。在某些情况下,可以以不同的顺序执行权利要求中记载的动作,并且仍然实现期望的结果。

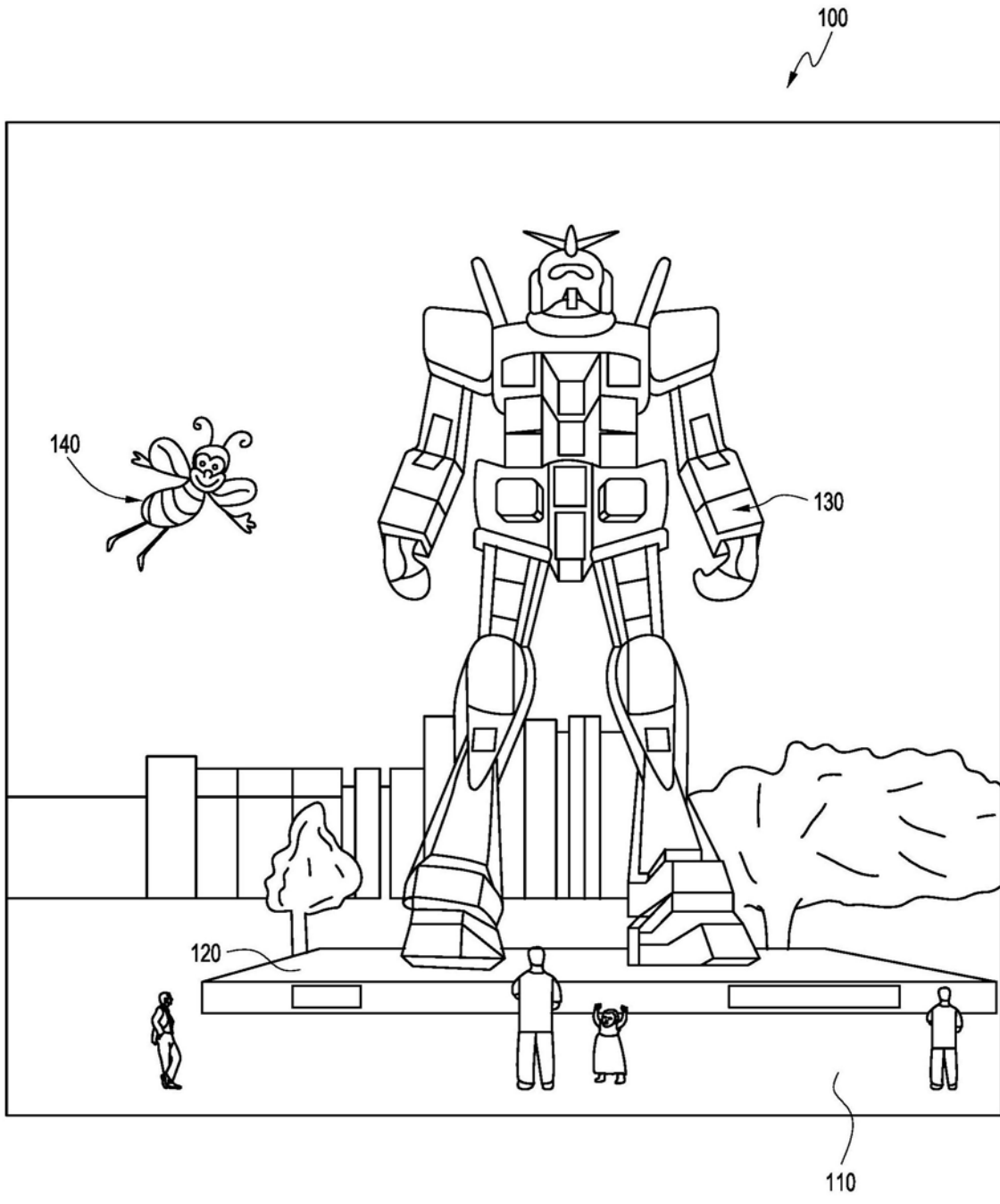


图1

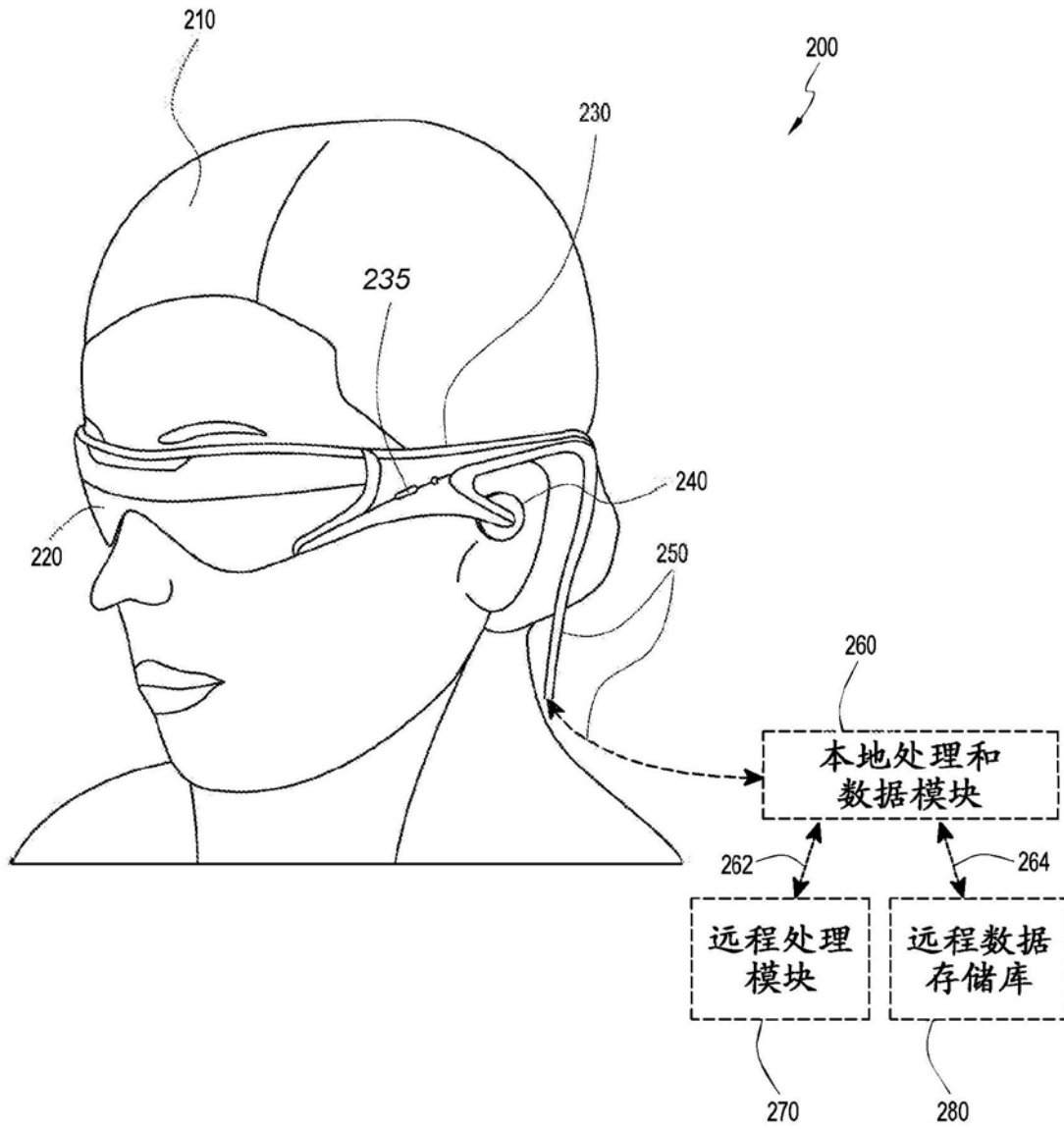


图2

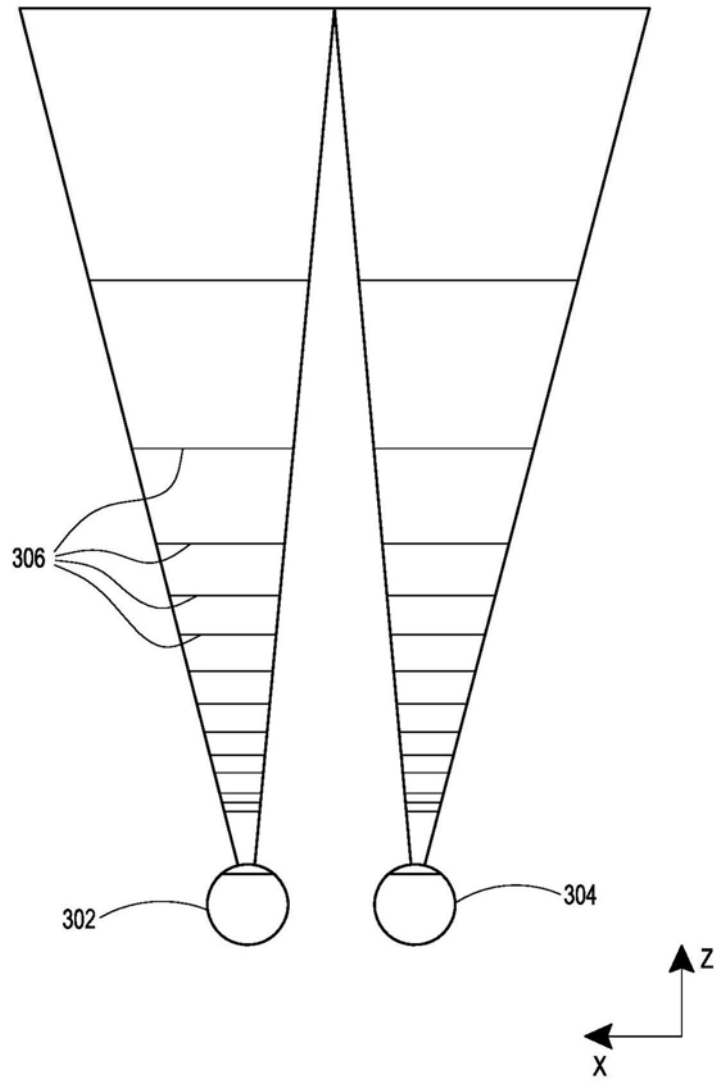


图3

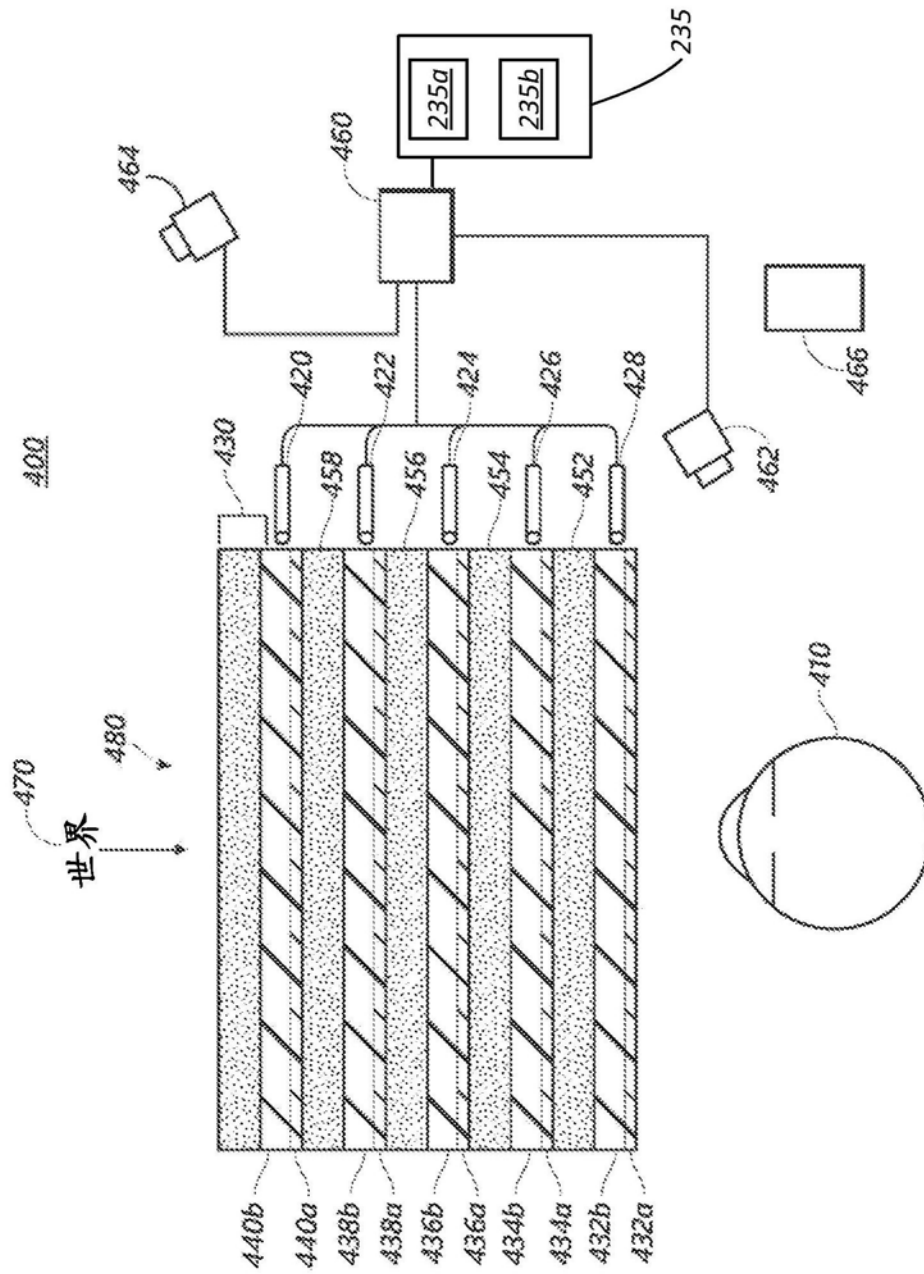


图4

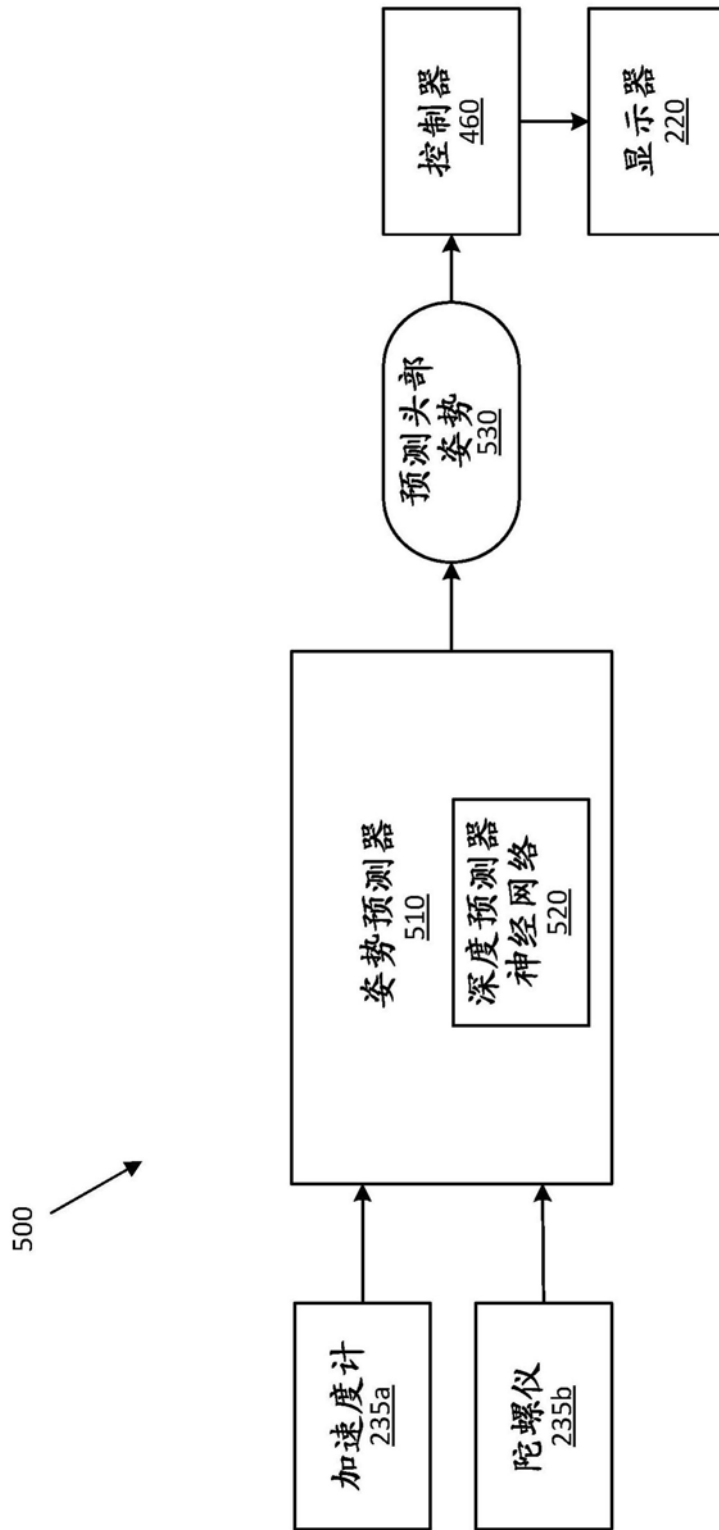


图5

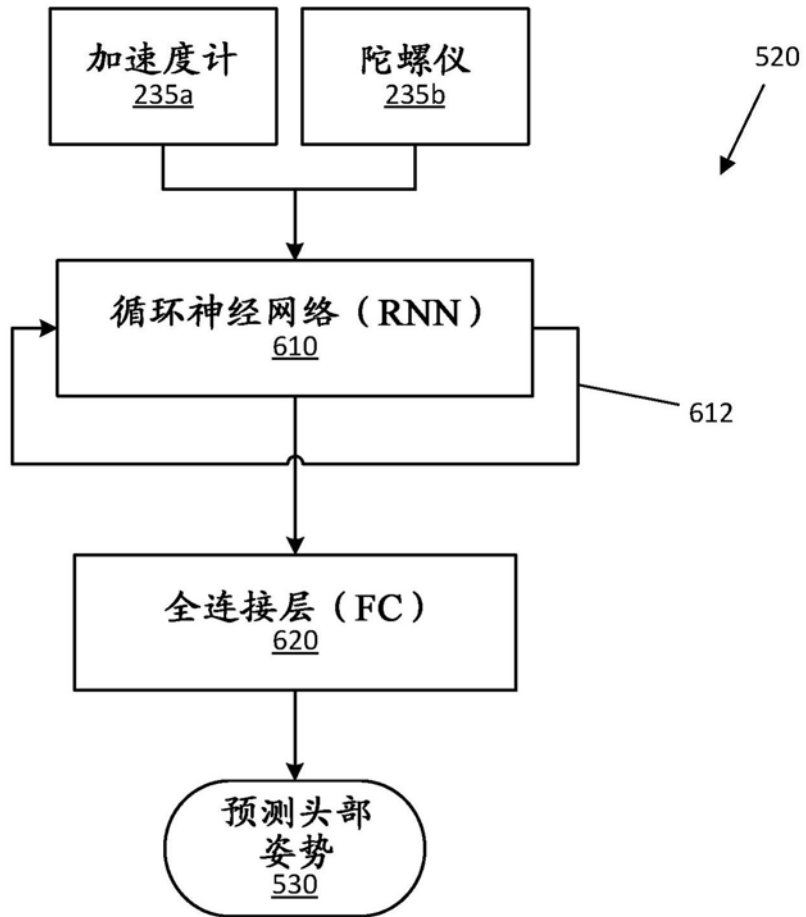


图6

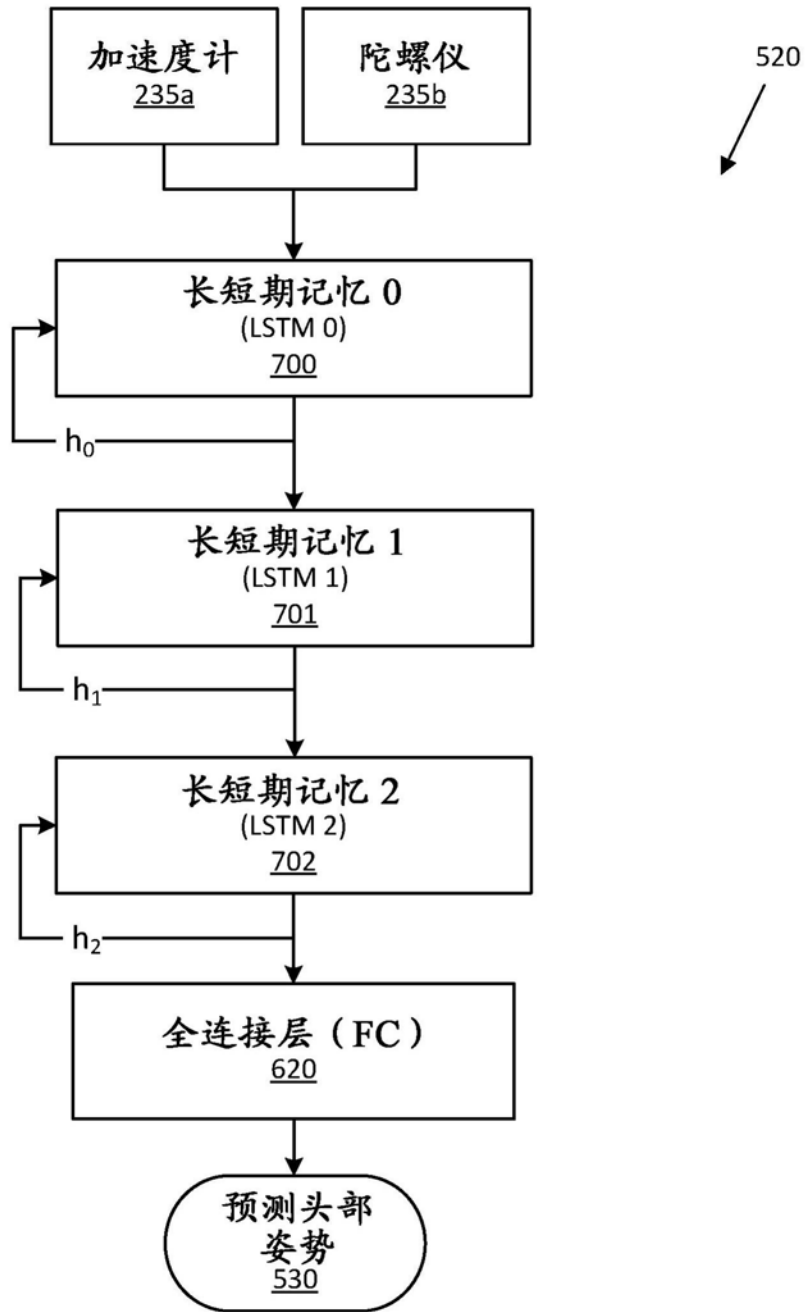


图7

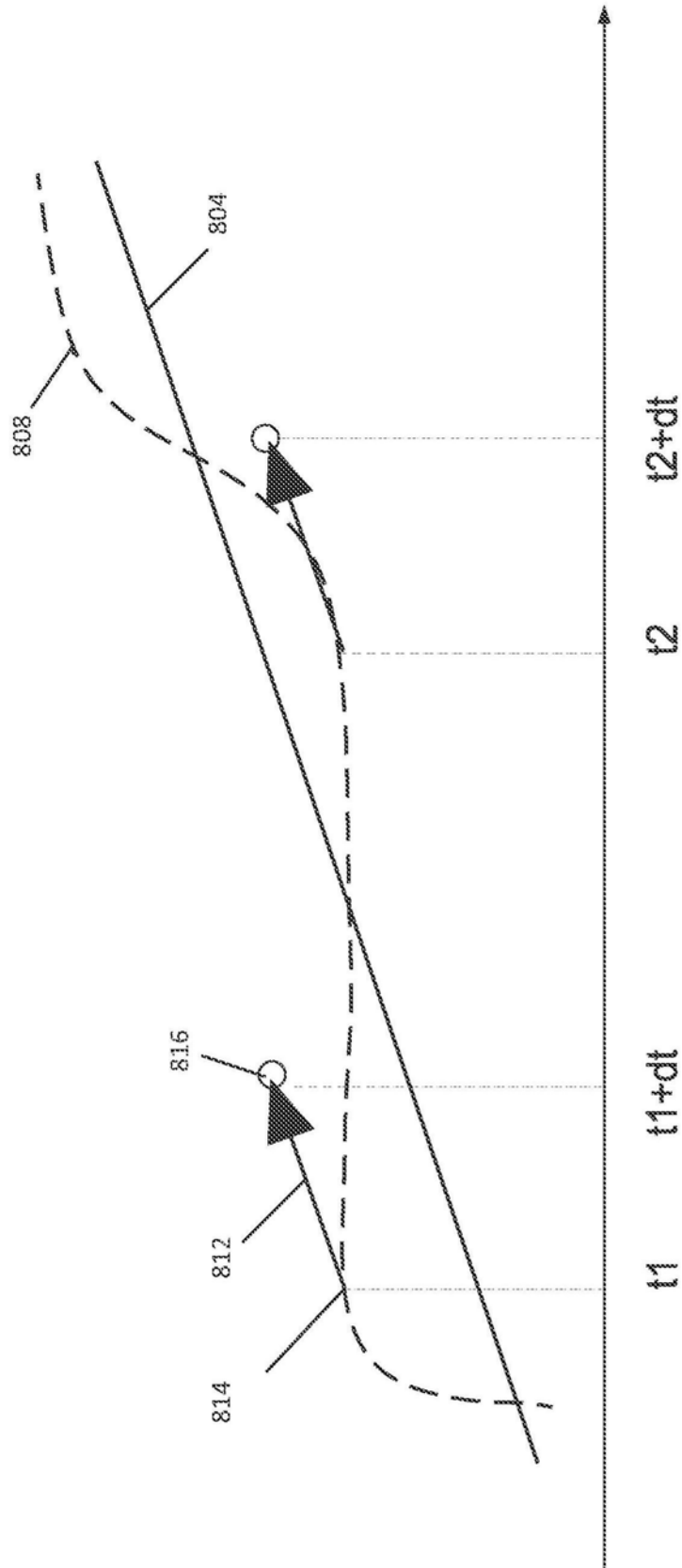


图8

900

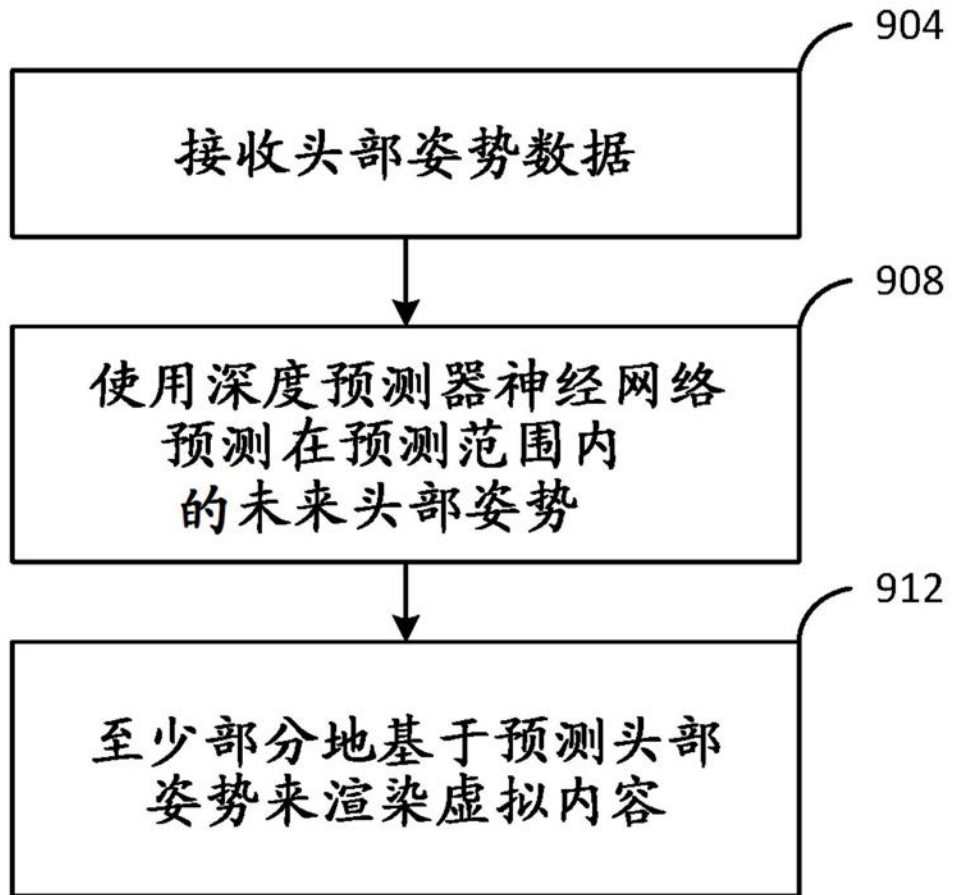


图9