



(12)发明专利申请

(10)申请公布号 CN 107004279 A

(43)申请公布日 2017.08.01

(21)申请号 201580067125.9

(22)申请日 2015.11.30

(30)优先权数据

14/566,410 2014.12.10 US

(85)PCT国际申请进入国家阶段日

2017.06.09

(86)PCT国际申请的申请数据

PCT/US2015/062925 2015.11.30

(87)PCT国际申请的公布数据

W02016/094109 EN 2016.06.16

(71)申请人 微软技术许可有限责任公司

地址 美国华盛顿州

(72)发明人 D·巴斯蒂安 K·A·肯尼迪

(74)专利代理机构 上海专利商标事务所有限公司 31100

代理人 罗婷婷 胡利鸣

(51)Int.Cl.

G06T 7/80(2017.01)

G06T 7/73(2017.01)

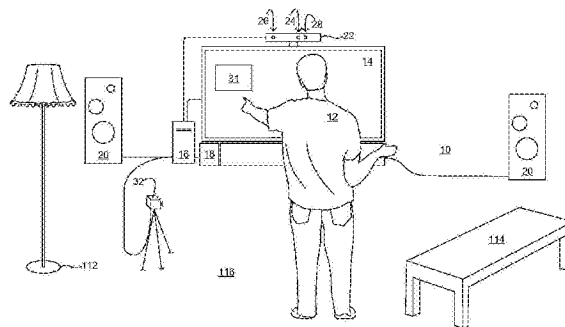
权利要求书2页 说明书12页 附图7页

(54)发明名称

自然用户界面相机校准

(57)摘要

提供了一种校准自然用户界面(NUI)相机相对于显示器的六自由度(6DoF)姿态的方法。从校准角度对环境进行成像的校准视频(该校准视频设置显示器和一个或多个特征)是从NUI相机或校准相机接收的。该环境的三维图(其定义显示器的6DoF姿态以及一个或多个特征中的每一者的三维位置)是根据该校准视频来建模的。从操作角度对该环境进行成像的主视频(该主视频设置该一个或多个特征)是从该NUI相机接收到的。NUI相机的6DoF姿态是基于该一个或多个特征的操作透视图在该环境的三维图内被寻找到的。



1. 一种校准自然用户界面 (NUI) 相机相对于显示器的六自由度 (6DoF) 姿态的方法:

从校准相机接收从校准角度对环境进行成像的校准视频, 所述校准视频设置所述显示器和一个或多个特征;

根据所述校准视频三维地建模所述环境的三维图, 所述环境的所述三维图定义所述显示器在所述环境内的6DoF姿态以及所述一个或多个特征中的每一者在所述环境内的三维位置;

从所述NUI相机接收从操作角度对所述环境进行成像的主视频, 所述主视频设置所述一个或多个特征; 以及

基于所述一个或多个特征的操作透视图, 在所述环境的所述三维图内寻找所述NUI相机的6DoF姿态。

2. 如权利要求1所述的方法, 其特征在于, 进一步包括:

从所述NUI相机接收对所述环境中的用户进行成像的主视频;

标识所述用户的第一身体部位和所述用户的第二身体部位相对于所述显示器的三维位置; 以及

在所述显示器上可视地指示所述显示器和从所述第一身体部位起延伸通过所述第二身体部位的线的交点。

3. 如权利要求1所述的方法, 其特征在于, 进一步包括:

将用户的第二身体部位朝向所述显示器的移动解释为针对显示在所述显示器和从所述用户的第一身体部位起延伸通过所述第二身体部位的线的交点处的对象的输入命令。

4. 如权利要求3所述的方法, 其特征在于, 所述第二身体部位朝向所述显示器的移动是基于所述第二身体部位和所述用户的身体之间的距离来确定的。

5. 如权利要求3所述的方法, 其特征在于, 所述第二身体部位朝向所述显示器的移动是基于所述第二身体部位和所述第一身体部位之间的距离来确定的。

6. 如权利要求3所述的方法, 其特征在于, 进一步包括:

在所述显示器上呈现显示对象; 以及

当所述交点与所述显示对象重合时, 响应于所述第二身体部位朝向所述显示器的移动而选择所述显示对象。

7. 如权利要求2所述的方法, 其特征在于, 所述第一身体部位是所述用户的眼睛, 且所述第二身体部位是所述用户的手指。

8. 如权利要求2所述的方法, 其特征在于, 所述第一身体部位是所述用户的肘部, 且所述第二身体部位是所述用户的手指。

9. 如权利要求1所述的方法, 其特征在于, 所述主视频设置所述校准相机。

10. 如权利要求9所述的方法, 其特征在于, 所述校准相机包括所述NUI相机可见的两个或更多个参考标记。

11. 如权利要求9所述的方法, 其特征在于, 所述校准相机包括显示屏, 所述显示屏被配置成呈现所述NUI相机可见的参考图像。

12. 如权利要求1所述的方法, 其特征在于, 所述校准相机被包括在用户可穿戴的头戴式增强现实设备中。

13. 一种计算系统, 包括:

深度相机;

逻辑机;以及

存储机,所述存储机保持能由所述逻辑机执行以执行进行以下的指令的指令:

接收相对于所述深度相机在环境内的6DoF姿态来定义显示器的6DoF姿态的三维模型;

接收所述环境内的用户的机器可读身体模型,所述机器可读身体模型包括所述用户的第一身体部位的三维位置以及所述用户的第二身体部位的三维位置;

在第一条件期间,将所述用户的所述第一身体部位的位置和所述用户的所述第二身体部位的位置解释成控制所述计算系统的物理交互区姿势;以及

在第二条件期间,将所述用户的所述第一身体部位相对于所述显示器的位置和所述用户的所述第二身体部位相对于所述显示器的位置解释成控制所述计算系统的指向姿势。

14. 如权利要求13所述的计算系统,其特征在于,所述存储机保持进行以下的指令:

响应于接收来自所述用户的命令而从所述第一条件转变到所述第二条件。

15. 如权利要求13所述的计算系统,其特征在于,所述存储机保持进行以下的指令:

响应于识别到所述第二身体部位相对于所述显示器的距离已减小到低于阈值,从所述第一条件转变到所述第二条件。

自然用户界面相机校准

[0001] 背景

[0002] 自然用户界面允许用户向计算设备提供姿势输入。

[0003] 概述

[0004] 提供本概述以便以简化的形式介绍以下在详细描述中进一步描述的一些概念。本概述并不旨在标识所要求保护主题的关键特征或必要特征,也不旨在用于限制所要求保护主题的范围。此外,所要求保护的主体不限于解决在本公开的任一部分中所提及的任何或所有缺点的实现。

[0005] 提供了一种校准自然用户界面(NUI)相机相对于显示器的六自由度(6DoF)姿态的方法。从校准角度对环境进行成像的校准视频(该校准视频设置(site)显示器和一个或多个特征)是从NUI相机或校准相机接收到的。该环境的三维图(该三维图定义显示器的6DoF姿态以及一个或多个特征中的每一者的三维位置)是根据该校准视频来建模的。从操作角度对该环境进行成像的主视频(该主视频设置一个或多个特征)是从该NUI相机接收到的。NUI相机的6DoF姿态是基于该一个或多个特征的操作透视图在该环境的三维图内找出的。

[0006] 附图简述

[0007] 图1示出包括自然用户界面相机的计算系统的示例环境。

[0008] 图2示出示例头戴式显示系统。

[0009] 图3示出用于校准自然用户界面相机相对于显示器的六自由度姿态的示例方法。

[0010] 图4示出示例计算系统的各方面。

[0011] 图5显示示例虚拟骨架的各方面。

[0012] 图6示出示例虚拟骨架的手部分的各方面。

[0013] 图7A示出用户经由视线指向来定向(target)显示器。

[0014] 图7B示出用户经由手臂瞄准指向来定向显示器。

[0015] 图8示出用户经由物理交互区来向计算系统提供输入。

[0016] 详细描述

[0017] 本公开涉及用于校准计算系统的自然用户界面(NUI)相机的方法。具体地,包括显示器和NUI相机,环境被成像,并且该环境的三维图被构造。在该三维图内,显示器和NUI相机相对于该环境及相对于彼此的位置和6自由度(6DoF)姿态被建立。

[0018] 图1示出在其中用户12站立在大尺寸、平面显示屏14前面的环境10。在一些实施例中,显示器14可以是立体3D显示器,或者在其它实施例中,可以是传统的2D显示器。计算系统16操作上耦合至该显示器以及其它组件,例如音频-视频(A/V)接收机18和扩音器20。在一个实施例中,计算系统可以是游戏控制台。在另一个实施例中,计算系统可以是多用途个人计算机或工作站。以下描述计算系统的附加方面。

[0019] 在一些实施例中,用户12可以通过传统的输入设备(例如键盘、触摸屏、鼠标、游戏系统控制器和/或无线遥控器)与计算系统16和相关联的组件进行交互。替换地或附加地,计算系统可对“自然用户输入”(NUI)作出响应,其中用户的语音、手势和/或眼睛移动被感测、解释并被用于控制计算系统。

[0020] 为此,在图1中的计算系统16操作上被耦合到机器视觉系统22。所示的机器视觉系统包括红外(IR)或近红外(NIR)照明源24、深度相机26以及平面图像相机28。机器视觉系统还可以包括用于从用户12接收语音或其它可听输入的话筒30(在图1中未示出)。

[0021] 深度相机26可被配置来获取用户12的时间分辨深度图序列和环境10的其它方面。如本文中所使用的,术语“深度图”指代与成像场景的对应区域(X_i, Y_i)配准(register)的像素阵列,其中对于每一像素,深度值 Z_i 指示对应区域的深度。“深度”被定义为与深度相机的光轴平行的坐标,该坐标随着距深度相机的距离的增加而增加。操作上,深度相机可被配置成获取2D图像数据,深度图是经由下游处理根据2D图像数据获得的。

[0022] 一般来说,深度相机26的性质在本公开的各实施例中可以是不同的。在一实施例中,来自深度相机中的两个立体地定向的成像阵列的亮度或颜色数据可被共同配准并用于构造深度图。在其它实施例中,照明源24可被配置成将包括众多离散特征(例如,线或点)的结构化照明图案投影到该主体上。该深度相机中的成像阵列可被配置成对从该主体反射回的结构化照明进行成像。基于所成像的主体的各个区域中的毗邻特征之间的间隔,可构造该主体的深度图。在又一其它实施例中,照明源可以将脉冲IR或NIR照明投向主体。深度相机中的一对成像阵列可被配置成检测从主体反射回的脉冲照明。这两个阵列均可包括与脉冲照明同步的电子快门,但这两个阵列的集成时间可不同,使得脉冲照明的从照明源到主体再接着到这两个阵列的像素解析的飞行时间可基于在两个阵列的相应元素中接收到的相对光的量来区别开来。

[0023] 通过合适的物镜系统,平面图像相机28在某一视野角范围内检测光,将这样的角度映射到矩形像素阵列中的对应像素。在一个实施例中,平面图像相机可检测多个波长通道—例如,红、绿、蓝等中的光,每个波长通道与该阵列中的像素的子集相关联。替换地,单色平面图像相机可被使用,其以灰度级对可见的、IR、NIR和/或紫外(UV)光进行成像。平面图像相机中曝光的所有像素的颜色或亮度值共同构成数字图像。在一些实施例中,深度相机26和平面图像相机28被布置成具有平行的、以相同方向定向的光轴。

[0024] 可使用深度相机26来在环境10内跟踪用户12,使得用户12的移动可被计算系统16解释为控制,而无需用户12持有任何特定的运动检测设备或佩戴任何特定的标记。在一个示例中,用户12可仅通过指向显示器14上出现的显示对象31来定向到该显示对象。为了允许用户通过直接指向显示对象来提供输入,显示器和机器视觉系统相对于彼此、相对于环境10以及相对于用户12的六自由度(6DoF= X, Y, Z 、俯仰、滚转、偏航)姿态必须是已知的。

[0025] 为了映射环境10内的显示器的6DoF姿态,可从不同的角度拍摄环境10的两个或更多个图像,只要这些图像中的至少一者包括显示器14即可。在一些示例中,这可通过以下来完成:在深度相机26以校准角度放置时对环境10进行成像,随后将深度相机26移动到操作角度(例如,图1中示出的角度)并对环境10重新进行成像。在其他示例中,单独的校准相机可被用于从校准角度对环境10进行成像。例如,辅助相机32可被通信地耦合到计算系统16。在另一示例中,用户12可使用设置在智能电话或平板计算设备内的相机对环境10进行成像。在又一示例中,环境10可由设置在增强现实设备(诸如由用户12穿戴的头戴式计算系统)内的相机来成像。

[0026] 图2描绘了示例头戴式显示系统40。头戴式显示系统40采用一对可佩戴的眼镜或护目镜的形式。头戴式显示系统40包括透视显示器42,透视显示器42可被配置成向透过透

视显示器观看物理环境的用户可视地增强物理环境的景象。头戴式显示系统40可被配置成增强真实世界环境的现实。

[0027] 例如,真实世界背景的景象可通过可用透视显示器42呈现的图形内容来增强,使得图形内容表现为与真实世界背景混合。具体地,图像产生系统43可被配置成用透视显示器42显示虚拟物体的图像。

[0028] 在一些配置中,头戴式显示系统可以是加性显示系统,其中全息光可被显示在透视显示器上。例如,一个或多个部分透明的像素可向用户的眼睛发射光。用户可穿过和/或绕过像素来观看真实世界物体,但是所发射的光可看起来照亮了穿过和/或绕过像素观看到的背景物体。

[0029] 头戴式显示系统40包括可包括一个或多个光学传感器的光学传感器系统44。在一个示例中,光学传感器系统44包括面向内的光学传感器46和面向外的光学传感器48。面向内的光学传感器可被配置成对用户的眼睛执行注视检测或其他分析。面向外的光学传感器可被配置成从与用户穿过透视显示器42所观察到的类似的有利视点(vantage point)(例如,视线)检测真实世界背景。在一些配置中,传感器子系统可包括两个或更多个不同的面向内的和/或面向外的光学传感器(例如,面向外的平面图像相机和面向外的深度相机)。

[0030] 头戴式显示系统40可进一步包括位置传感器系统50,该位置传感器系统50可包括输出位置传感器信息的一个或多个位置传感器(例如,(诸)加速度计、(诸)陀螺仪、(诸)磁力计、(诸)全球定位系统、(诸)多点定位跟踪器等),该位置传感器信息可用于评估相关传感器的位置、定向和/或运动。

[0031] 从光学传感器系统44接收的光学传感器信息和/或从位置传感器系统50接收的位置传感器信息可被用于评估透视显示器相对于其他环境物体的有利视点的位置和定向。在一些配置中,有利视点的位置和定向可用六个自由度(例如,世界空间X、Y、Z、俯仰、翻滚、偏航)来表征。可全局地或独立于真实世界背景地表征有利视点。可用机载计算系统(例如,机载计算系统52)和/或非机载计算系统(例如,图1的计算机系统16)来确定位置和/或定向。

[0032] 此外,光学传感器信息和位置传感器信息可被计算系统用来执行对真实世界背景的分析,诸如深度分析、表面重建、环境色彩和照明分析、或其他合适的操作。光学和位置传感器信息可被用于创建真实世界背景的虚拟模型。在一些配置中,可相对于此虚拟空间来表征有利视点的位置和定向。此外,虚拟模型可被用于确定虚拟物体在虚拟空间中的位置。

[0033] 图3示出用于校准自然用户界面(NUI)相机相对于显示器的六自由度(6DoF)姿态的示例方法100。在110,方法100包括从校准相机接收从校准角度对环境进行成像的校准视频,该校准视频设置显示器和一个或多个特征。如上所述,校准相机可以是辅助相机(诸如辅助相机32)、设置在增强现实设备内的相机(诸如面向外的光学传感器48)、设置在智能电话或平板计算机内的相机或其他合适的相机。替换地,NUI相机可被移动到校准角度并被用于完成该校准。在任一情况下,用于完成该校准的相机可被配置成获得2D图像数据和/或3D图像数据。

[0034] 一个或多个特征可包括环境10的可由环境10的经组装的深度图上的一个或多个点来表示的一个或多个对象、方面或特性。该一个或多个特征可以是静态特征,使得该特征的位置在校准过程期间不被改变。例如,环境10内的各对象(诸如,灯112或长凳114的特定方面)可被用作特征。环境10的一方面(诸如地板平面116、墙壁、天花板等)的全部或部分可

被用作特征。各种图像处理策略可被用于标识适合于6DoF姿态校准(例如,处理以推导出边缘图、强度曲线和/或点云)的特征。

[0035] 显示器14可被配置成呈现校准相机可见的校准图像。校准图像可被呈现为使得显示器的校准视频可被用于确定该显示器的特性。例如,显示器的大小、形状和/或分辨率可至少部分地基于校准视频中的校准图像的外观来确定。此外,显示器相对于校准相机的位置和倾斜可至少部分地基于校准视频中的校准图像的重点(keystone)来确定。校准图像可以是静态的,或者可被渐进地改变。通过改变校准图像,从显示器发出的光可相应地改变。环境特征的阴影或反光方面的改变可由此被记录在校准视频中。

[0036] 在120,方法100包括根据校准视频三维地建模该环境的三维图,该环境的三维图定义显示器在该环境内的6DoF姿态以及一个或多个特征中的每一者在该环境内的三维位置。

[0037] 在130,方法100包括从NUI相机接收从操作角度对该环境进行成像的主视频,该主视频设置该一个或多个特征。操作角度可根据NUI相机的位置和姿态,在计算系统的操作期间,将根据该位置和姿态来接收对于计算系统的输入。主视频可与校准视频同时被记录,或者可被分开地记录。

[0038] 可选地,校准相机可包括NUI相机可见的两个或更多个参考标记。参考标记可指示校准相机相对于NUI相机以及相对于环境的位置和方向。在其中校准相机包括显示屏的示例中,诸如当校准相机被包括在智能相机或平板计算机中时,显示屏可被配置成呈现NUI相机可见的参考图像。基于显示屏的尺寸和分辨率,校准相机和NUI相机之间的距离可被确定。

[0039] 在140,方法100包括基于一个或多个特征的操作透视图在该环境的三维图内找出NUI相机的6DoF姿态。通过这种方式,NUI相机和显示器相对于彼此以及相对于环境的6DoF姿态是已知的。作为一个示例,NUI相机的6DoF姿态可从NUI相机到所映射的特征中的一者或多者的距离和角度中推导出。

[0040] 如以上所介绍的,NUI相机可被用于捕捉校准视频以及主视频两者。例如,该方法可包括从NUI相机接收从校准角度对环境进行成像的校准视频,该校准视频设置显示器和一个或多个特征。该方法可进一步包括从NUI相机接收从操作角度对环境进行成像的主视频,该主视频设置该一个或多个特征。主视频可与校准视频分开地被记录。NUI相机可被耦合到一个或多个运动传感器,使得该相机从校准角度到操作角度的移动被跟踪。来自相机跟踪的空间信息可被用来确定该相机相对于环境和显示器的位置和姿态。

[0041] 还可使用校准NUI相机相对于显示器的6DoF姿态的其他方法。在一些示例中,NUI相机可被配置成从固定于显示器的6DoF姿态的操作角度利用。例如,NUI相机可位于显示器的边框。在这样的示例中,环境的一个或多个特征的成像对于确定NUI相机的6DoF姿态而言可能是足够的。例如,通过使用地板平面作为特征,该传感器的高度和角度可被确定。在一些场景中,显示器的特性(诸如显示器尺寸和分辨率)可被计算系统知道。对于一些显示器,显示屏可被假设为与重力平行。

[0042] 用户输入也可被用来校准NUI相机相对于显示器的6DoF姿态。例如,NUI相机可记录用户指向显示器上的诸特征(诸如显示器拐角或作为校准过程的一部分被显示的对象)的视频。该用户可被指令通过指向和拖动对象来使显示器上出现的显示对象相对于彼此移

动。用户可用手指或其他身体部位指向屏幕,或者可用诸如激光指示器或可伸缩天线之类的工具指向屏幕。在一些示例中,用户可使相机或可伸缩透镜瞄准显示器上的特征,或者可举起包括NUI相机可见的边框的反光镜。当这样的工具被使用时,可选地,它们可包括标记,NUI相机可对这些标记进行成像以解析该工具相对于NUI相机的6DoF姿态。如果NUI相机的操作角度被更改了,则可从该新的操作角度重复该校准过程,或者可基于该相机的关于环境的一个或多个特征的新视图来推导出新的NUI相机位置。

[0043] 如在图4中所示,来自机器视觉系统22的传感数据由计算系统16的一个或多个NUI引擎处理以揭示可行的用户输入。这样的输入可以由操作系统(OS) 160或由应用级代码来解析,由此控制计算系统的操作。在图4的实施例中,计算系统16包括语音识别引擎162、姿势识别引擎164以及眼睛跟踪引擎166。

[0044] 语音识别引擎162被配置成处理来自话筒30的音频数据、识别用户的语音中的单词或短语以及生成对应的到OS 160或应用170的可操作输入。姿势识别引擎164被配置成至少处理来自机器-机器视觉系统22的深度视频(即时间分辨深度图序列),标识深度视频中的一个或多个人类主体、计算所标识的主体的各种几何(例如,骨架)特征,以及从几何特征收集要被用作NUI的各姿势或手势信息。

[0045] 在一个非限制性实施例中,姿势识别引擎164标识深度视频中的一个或多个人类主体的至少一部分。通过合适的深度图像处理,深度图的给定位点可被识别为属于人类主体。在更特定实施例中,属于人类主体的像素可(例如通过分割出深度图的在合适的时间标度内展示出高于阈值的运动的部分)被标识出,并且人类的通用几何模型可从这些像素中推导出。

[0046] 在一个实施例中,深度图的每一像素可被分配将该像素标识为属于特定人类主体或非人类元素的人物索引。作为示例,与第一人类主体相对应的像素可被分配等于一的人物索引,与第二人类主体相对应的像素可被分配等于二的人物索引,而不与人类主体相对应的像素可被分配等于零的人物索引。人物索引可用任何适当的方式来确定、分配和保存。

[0047] 姿势识别引擎164还可以作出关于哪个(或哪些)人类主体将向计算系统16提供用户输入(即哪个人类主体将被标识为用户)的确定。在一些实施例中,人类主体可基于到显示器14或深度相机26的邻近度和/或在深度相机的视野中的位置被选为用户。更具体地,所选的用户可以是最接近深度相机或最接近深度相机的视野的中心的人类主体。

[0048] 姿势识别引擎164还可以处理来自用户的姿势信息。该姿势信息是通过计算从深度视频推导出的。在一些实施例中,附加的传感输入—例如,来自平面图像相机28的图像数据或来自话筒30的音频数据—可与姿势信息一起被处理。

[0049] 在一些实施例中,姿势识别引擎164可被配置成分析深度图的对应用于用户的像素,以便确定每个像素对应于用户身体的什么部位。至此,各种不同的身体部位分配技术可被使用。在一个示例中,深度图中具有适当的人物索引(见上文)的每一像素可被分配一身体部位索引。身体部位索引可包括离散标识符、置信度值和/或指示像素可能对应的一个或多个身体部位的身体部位概率分布。

[0050] 在一些实施例中,机器学习可被用于向每一像素分配身体部位索引和/或身体部位概率分布。机器学习方法参考从先前训练的已知姿态集合中学习到的信息来分析用户。在受监督训练阶段期间,例如,各种人类主体被观察到处于各种姿态;训练器在观察到的数

据中提供标记各机器学习分类器的真实注释。所观察到的数据和注释接着被用于生成将输入(例如,深度视频)映射到所需输出(例如,相关像素的身体部位索引)的一个或多个机器学习算法。

[0051] 此后,虚拟骨架被适配到深度视频的对应于用户的像素。图5示出示例虚拟骨架172。该虚拟骨架包括枢转地耦合在多个关节176处的多个骨骼部分174。在一些实施例中,身体部位指定可被分配给每个骨骼部分和/或每个关节。在图5中,每个骨骼部分174的身体部位指定由所附字母表示:A针对头部、B针对锁骨、C针对上臂、D针对前臂、E针对手、F针对躯干、G针对骨盆、H针对大腿、J针对小腿、而K针对脚。类似地,每个关节176的身体部位指定由所附字母表示:A针对颈部、B针对肩膀、C针对肘、D针对手腕、E针对下背、F针对髋关节、G针对膝盖以及H针对踝。自然地,图5中示出的骨骼部分和关节的安排不是为了以任何方式进行限制。符合本公开的虚拟骨架可包括实质上任何类型和数量的骨骼部分和关节。

[0052] 在更具体的实施例中,对应于用户的手的点云(深度图的各部分)可以被进一步处理以揭示手的骨骼子结构。图6示出用户的虚拟骨架172的示例手部分178。手部分包括腕关节180、指关节182以及毗邻的手指部分184。

[0053] 通过任何合适的最小化方式,骨骼部分的长度和关节的位置和旋转角可被调整以符合深度图的各个轮廓。以此方式,每个关节被分配各种参数—例如,指定关节位置的笛卡尔坐标、指定关节旋转的角、以及指定对应的身体部位的形态(手张开、手闭合等)的附加参数。该虚拟骨架可采取数据结构的形式,该数据结构包括针对每个关节的这些参数中的任一个、一些或全部。该过程可以限定所成像的人类主体的位置和姿态。一些骨架拟合算法可以使用深度数据结合其他信息,诸如彩色图像数据和/或指示一个像素轨迹如何相对于另一个来移动的运动数据。在上述方式中,虚拟骨架可被适配到深度视频的帧序列中的每一帧中。通过分析各个骨架关节和/或部分的位置变化,被成像的用户的对应移动—例如姿势或动作—可被确定。

[0054] 上述描述不应当被理解为限制可用于构造虚拟骨架172的方法的范围,因为可以用任何合适的方式从深度图中导出虚拟骨架而不背离本公开的范围。此外,尽管使用虚拟骨架来建模用户12有优点,但这个方面并不意味着是必需的。代替虚拟骨架,原始点云数据或其他机器可读模型可被用来提供合适的姿态信息。

[0055] 环境的三维图和用户的三维骨架图可被用于允许用户仅通过指向显示器来与计算系统交互。在一个示例方法中,计算系统16可从对环境10中的用户12进行成像的经6DoF校准的机器视觉系统22接收主视频,并可基于该主视频来相对于环境的三维图确定用户的三维骨架图。计算系统16可标识用户的一个或多个身体部位相对于显示器的三维位置。

[0056] 图7A示出用户12指向显示器14的一个示例。在该示例中,用户的眼睛被指定为第一身体部位200。用户的眼睛的位置可由眼睛跟踪引擎166和/或姿势识别引擎164来确定。例如,眼睛跟踪引擎166可基于用户的眼睛的中心点、用户的瞳孔的中心点来确定用户的眼睛的位置,和/或姿势识别引擎164可基于虚拟骨架的头部关节的位置来估计眼睛的位置。眼睛跟踪引擎166可被校准以确定每一用户的主视眼,以便选择一眼睛作为第一身体部位200。在该示例中,用户的手指被指定为第二身体部位202,但在其他示例中,第二身体部位可以是用户的手。它也可以是棒或指示笔、手指的组合、手套的一部分、脚趾、脚、鞋或任何其它合适的对象。第二身体部位202的位置可基于虚拟骨架的单个手关节位置以及来自姿

势识别引擎164的关于手是张开还是闭合的指示来确定。此外,用户12的身体可被分配例如指定用户的躯干、或重心的中心点203。

[0057] 参考图4和7A,指示器投影引擎204从姿势识别引擎164接收包括第二身体部位202的3D坐标在内的姿势信息。指示器投影引擎还从眼睛跟踪引擎166接收第一身体部位200(用户的眼睛)的3D坐标。指示器投影引擎实际上被配置成延伸线205,该线205有一端点在第一身体部位的位置处并通过第二身体部位的位置。线205可被延伸到显示器14。可选地,交点206可在显示器上以光标、图标或其他合适的显示对象的形式来指示。该视线可被用于标识显示器14上的定向到的对象。作为示例,该交点被用于相对于显示器的拐角计算指示器投影坐标(X_{pp}, Y_{pp}),其以配准数据207的形式被提供给OS 160。

[0058] 各种图形显示对象31可被呈现在显示器14上。这样的对象可例如包括窗口、照片、文本框、图标和UI控件元素。每一对象在显示器上的坐标被存储在位置数据208中(参考图2),该位置数据是OS 160可访问的。位置数据可包括每一对象与显示器的可观看区域的给定拐角相对的拐角的X,Y偏移209。在其中显示器14是立体或3D显示器的实施例中,深度偏移Z也可以被指定。位置数据208还包括针对显示器14的配准数据207。配准数据可包括定义显示器的可观看区域的物理大小、位置和定向的坐标—例如有效显示面积的每个拐角的X、Y和Z。

[0059] 显示对象31的偏移209可随后被与交点206的配准数据进行比较。当交点206与显示对象31的全部或部分重合时,显示对象31可被定向到。显示对象31的定向可视觉地(诸如通过更改显示对象的外观)、听觉地或通过其他合适的手段来指示。一旦一对象被定向到,用户12可以发出要在该对象上采取进一步行动的信号。计算系统16的一个或多个NUI引擎可以被配置为检测用户要在被定向的对象上行动的意图。例如,在对象被定向到后,用户可通过朝向显示屏移动第二身体部位202(在202a处被指示)来发信号表示选择该对象的意图。例如,当交点206与显示对象31重合时,显示对象31可响应于第二身体部位202朝向显示器14的移动而被选择。第二身体部位朝向显示器的移动可基于第二身体部位202相对于第一身体部位200之间的距离来确定。随着该距离增加,朝向显示器的移动可被推断出。随着该距离减小,远离显示器的移动可被推断出。第二身体部位202朝向显示器的移动可基于第二身体部位和显示器之间的距离。在一些示例中,用户将第二身体部位维持在同一X-Y位置达阈值时间段也可构成输入命令。

[0060] 在一些实施例和场景中,可选地,在用户的选择之后可激活定向到的对象。激活可包括执行与定向到的对象相关联的OS或应用级代码。例如,如果定向到的对象是文档,则激活可包括启动文字处理器,该文字处理器自动打开文档以供编辑。如果定向到的对象是幻灯片,则激活可包括在显示器上呈现该幻灯片。姿势识别引擎34可被配置成检测这种类型的激活姿势。在其它示例中,用户要选择或激活目标对象的意图可由语音识别引擎162所检测到的口述的单词或短语来发信号表示。

[0061] 图7B示出用户12指向显示器14的另一示例。在该示例中,用户的肘部被指定为第一身体部位210,而用户的手指被指定为第二身体部位212。用户12的身体被分配中心点213。线215从用户的肘部起延伸通过用户的手指,并延伸到显示器14,如交点216所指示的。通过该方式,用户可按类似于激光指示笔的方式来定向显示器。当显示对象31与交点216重合时,用户12可通过相对于显示器14移动第二身体部位212(在212a处示出)来选择或引起

对显示对象31的动作。第二身体部位朝向显示器的移动可基于第一身体部位210的位置和/或第二身体部位212的位置相对于用户身体中心点213的位置之间的距离来确定。

[0062] 可在计算系统16上建立并存储针对视线指向(如图7A所示)和/或手臂瞄准指向(如图7B所示)的用户偏好,使得用户可基于应用使用一种或两种类型的指向。姿势识别引擎164和眼睛跟踪引擎166可通过机器学习或另一技术来训练以基于用户的习惯来调整显示器上的交点的位置。例如,如果用户一般将偏离期望显示对象的线瞄准显示器,则该线的方向和/或交点的位置可相应地被补偿。

[0063] 尽管该指示器输入模式在许多场景下对于用户可能都是优选的,但可存在其中不同的输入模式被偏好的其他应用。由此,计算系统16可支持多个输入模式(在本文中为“物理交互区或PhIZ”)并进一步支持多个PhIZ之间的转变。

[0064] 图8示出用户12经由人体工程学PhIZ来向计算系统16提供输入的示例。可在世界空间中概念化交互区220,并可在模型空间中实例化交互区28。交互区可被配置成当用户在世界空间中移动时与该用户一起移动。结果,环境中的经实例化的交互区的可移动的界面空间坐标系225可相对于固定的环境坐标系230而改变。交互区220的位置和定向可基于用户12的位置和定向(和/或用户12的三维骨架模型)来建立。在交互区跟踪用户时,交互区可如图8所示保持在用户的正前方或所建立的任何空间定向。用户可在环境内到处移动和/或改变定向,并且交互区将自动地跟踪用户。由此,在该操作模式期间,用户的姿势命令被确定为与交互区和用户的位置的坐标有关,而非与显示器14有关。

[0065] 由此,计算系统16可被配置成在不同的条件下将用户的身体部位的位置解释为交互区姿势或指向区姿势。例如,计算系统16可接收相对于深度相机在环境内的6DoF姿态来定义显示器的6DoF姿态的三维模型,并且还可接收环境内的用户的机器可读身体模型,该身体模型包括用户的第一身体部位的三维位置和用户的第二身体部位的三维位置。在第一条件期间,计算系统可将用户的第一身体部位的位置和用户的第二身体部位的位置解释成交互区姿势。在第二条件期间,计算系统可将用户的第一身体部位的位置和用户的第二身体部位的位置解释成定向姿势。

[0066] 计算系统16可响应于接收到来自用户的命令而从第一条件转变到第二条件。例如,用户可发出口头命令或姿势命令(诸如,闭合非主视眼)。用户可朝向显示器迈步并指向显示器,以指示到第二条件的转变。用户可通过其他输入或通信系统提供输入。

[0067] 在一些示例,计算系统16可响应于接收来自在计算系统上执行的应用的指令而从第一条件转变到第二条件。应用可基于上下文和/或显示界面来指示第一或第二条件。例如,游戏应用可指示某些任务将在人工工程学PhIZ中执行,而其他任务将在指示器PhIZ中执行。应用可呈现要在指示器PhIZ中与其交互的菜单,而一旦选择了菜单项,就可在人工工程学PhIZ中与这些菜单项进行交互。

[0068] 在一些示例中,响应于识别到第二身体部位相对于显示器的距离已减小到低于阈值,计算系统可从第一条件转变到第二条件。例如,用户可朝向显示器14伸出手或手指。当手或手指在显示器的阈值距离内时,指示器PhIZ可被调用。

[0069] 尽管本文中的应用已相关于单个显示器来描述,但多个显示器可在环境内被校准。在一些示例中,投影屏可被用作显示器。在其中用户正佩戴着增强现实设备(诸如头戴式显示系统40)的示例中,虚拟显示器可在环境内的三维位置处被呈现给用户。用户可使用

指示器PhIZ基于用户的第一和第二身体部位相对于显示器的位置来与每一显示器交互。用户不需要与最近的显示器交互；相反，环境内的任何显示器都可被定向到。

[0070] 在一些示例中，指示器投影引擎82可被用于标识环境10中的任何定向到的真实世界对象。定向到的对象可包括可远程控制的设备，诸如有线电视盒、A/V接收机、打印机或环境控件。通过指示器投影引擎82定向和选择这样的对象的动作可具有在显示器14上呈现那个设备的控制界面使得用户能够通过NUI控制该设备的效果。例如，参考图1，用户12可指向灯112以打开或关闭该灯，或者改变调光设置。对任何现实世界对象的定向要求按与本文中描述的针对显示器的校准类似的方式来相关于该环境校准那个对象的位置和姿态。在其它环境中，现实世界对象或显示器可以是自校准的。自校准对象可包括具有内置的位置和定向跟踪（通过GPS和/或惯性传感器）的设备以及物理地耦合到无线位置查寻器的对象。在环境10内这些对象的位置和定向可被本地地计算并被传递给计算系统16以实现校准。这个特征的另一个优点在于自校准对象即使在该环境中被移动时也保持校准。

[0071] 在一些实施例中，本文中描述的方法和过程可与一个或多个计算设备的计算系统绑定。具体而言，这样的方法和过程可被实现为计算机应用程序或服务、应用编程接口（API）、库和/或其他计算机程序产品。

[0072] 参考图4，该图以简化形式示出用于执行本文中描述的方法和过程的计算系统16的一个非限制性示例。该计算系统包括逻辑机238和存储机240。该计算系统还包括显示器14、通信子系统242和附图中未示出的各种组件。

[0073] 逻辑机238包括被配置成执行指令的一个或多个物理设备。例如，逻辑机可被配置成执行作为以下各项的一部分的指令：一个或多个应用、服务、程序、例程、库、对象、组件、数据结构、或其他逻辑构造。这种指令可被实现以执行任务、实现数据类型、转换一个或多个组件的状态、实现技术效果、或以其他方式得到期望结果。

[0074] 逻辑机可包括被配置成执行软件指令的一个或多个处理器。作为补充或替换，逻辑机可包括被配置成执行硬件或固件指令的一个或多个硬件或固件逻辑机。逻辑机的处理器可以是单核或多核，且在其上执行的指令可被配置为串行、并行和/或分布式处理。逻辑机的各个组件可任选地分布在两个或更多单独设备上，这些设备可以位于远程和/或被配置成进行协同处理。逻辑机的各方面可由以云计算配置进行配置的可远程访问的联网计算设备来虚拟化和执行。

[0075] 存储机240包括被配置成保持可由逻辑机执行以实现此处所述的方法和过程的指令的一个或多个物理设备。在实现这些方法和过程时，可以变换存储机240的状态（例如，保存不同的数据）。

[0076] 存储机240可以包括可移动和/或内置设备。存储机240可包括光学存储器（例如，CD、DVD、HD-DVD、蓝光盘等）、半导体存储器（例如，RAM、EPROM、EEPROM等）和/或磁存储器（例如，硬盘驱动器、软盘驱动器、磁带驱动器、MRAM等）等等。存储机240可包括易失性、非易失性、动态、静态、读/写、只读、随机存取、顺序存取、位置可寻址、文件可寻址和/或内容可寻址设备。

[0077] 可以理解，存储机240包括一个或多个物理设备。然而，本文描述的指令的各方面可另选地通过不由物理设备在有限时长内持有的通信介质（例如，电磁信号、光信号等）来传播。

[0078] 逻辑机238和存储机240的各方面可被一起集成到一个或多个硬件逻辑组件中。这些硬件逻辑组件可包括例如现场可编程门阵列(FPGA)、程序和应用专用的集成电路(PASIC/ASIC)、程序和应用专用的标准产品(PSSP/ASSP)、片上系统(SOC)以及复杂可编程逻辑器件(CPLD)。

[0079] 术语“模块”、“程序”和“引擎”可用于描述被实现为执行一个特定功能的计算系统16的一方面。在某些情况下,可以通过执行由存储机240所保持的指令的逻辑机238来实例化模块、程序或引擎。将理解,可以从同一应用、服务、代码块、对象、库、例程、API、函数等实例化不同的模块、程序和/或引擎。同样,可以由不同的应用程序、服务、代码块、对象、例程、API、函数等来实例化同一模块、程序和/或引擎。术语模块、程序和引擎可涵盖单个或成组的可执行文件、数据文件、库、驱动程序、脚本、数据库记录等。

[0080] 应该理解,在此使用的“服务”是跨多个用户会话可执行的应用程序。服务可用于一个或多个系统组件、程序和/或其它服务。在某些实现中,服务可以在一个或多个服务器计算设备上运行。

[0081] 当包括显示器14时,显示器14可被用于呈现由存储机240保持的数据的视觉表示。此视觉表示可采取图形用户界面(GUI)的形式。由于本文所描述的方法和过程改变了由存储机保持的数据,并由此变换了存储机的状态,因此同样可以转变显示器14的状态以视觉地表示底层数据的改变。显示器14可包括使用实际上任何类型的技术的一个或多个显示设备。可将此类显示设备与逻辑机238和/或存储机240组合在共享封装中,或者此类显示设备可以是外围显示设备。

[0082] 当包括通信子系统242时,通信子系统242可被配置成将计算系统16与一个或多个其他计算设备通信地耦合。通信子系统242可包括与一个或多个不同通信协议兼容的有线和/或无线通信设备。作为非限制性示例,通信子系统可被配置成用于经由无线网络或者有线或无线局域网或广域网来进行通信。在一些实施例中,通信子系统可允许计算系统16经由诸如因特网这样的网络将消息发送至其他设备以及/或者从其他设备接收消息。

[0083] 将会理解,本文描述的配置和/或方式本质是示例性的,这些具体实施例或本文示例不应被视为限制性的,因为许多变体是可能的。本文描述的具体例程或方法可以表示任何数量的处理策略中的一个或多个。如此,所示和/或所述的各种动作可以以所示和/或所述顺序、以其他顺序、并行地执行,或者被省略。同样,上述过程的次序可以改变。

[0084] 本公开的主题包括本文公开的各种过程、系统和配置以及其他特征、功能、动作和/或性质的所有新颖和非显而易见的组合和子组合,以及其任何和所有等同物。

[0085] 一示例提供一种校准自然用户界面(NUI)相机相对于显示器的六自由度(6DoF)姿态的方法,所述方法包括:从校准相机接收从校准角度对环境进行成像的校准视频,该校准视频设置显示器和一个或多个特征,根据所述校准视频三维地建模所述环境的三维图,所述环境的所述三维图定义所述显示器在所述环境内的6DoF姿态以及所述一个或多个特征中的每一者在所述环境内的三维位置,从所述NUI相机接收从操作角度对所述环境进行成像的主视频,该主视频设置所述一个或多个特征,以及,基于所述一个或多个特征的操作透视图,在所述环境的三维图内寻找所述NUI相机的6DoF姿态。在这样的示例中,所述方法可附加地或替换地包括:从所述NUI相机接收对所述环境中的用户进行成像的主视频,标识所述用户的第一身体部位和所述用户的第二身体部位相对于所述显示器的三维位置,以及,

在所述显示器上可视地指示所述显示器和从所述第一身体部位起延伸通过所述第二身体部位的线的交点。在这样的示例中,所述方法可附加地或替换地包括将用户的第二身体部位朝向所述显示器的移动解释为针对显示在所述显示器与从所述用户的第一身体部位起延伸通过所述第二身体部位的线的交点处的对象的输入命令。在这样的示例中,所述第二身体部位朝向所述显示器的移动可附加地或替换地基于所述第二身体部位和所述用户的身体之间的距离来确定。在这样的示例中,所述第二身体部位朝向所述显示器的移动可附加地或替换地基于所述第二身体部位和所述第一身体部位之间的距离来确定。在这样的示例中,所述方法可附加地或替换地包括在所述显示器上呈现显示对象,并可附加地或替换地包括在所述交点与所述显示对象重合时,响应于所述第二身体部位朝向所述显示器的移动而选择所述显示对象。在这样的示例中,所述第一身体部位可附加地或替换地为所述用户的眼睛,并且所述第二身体部位可替换地或附加地为所述用户的手指。在这样的示例中,所述第一身体部位可附加地或替换地为所述用户的肘部,并且所述第二身体部位可替换地或附加地为所述用户的手指。在这样的示例中,所述主视频可附加地或替换地设置所述校准相机。在这样的示例中,所述校准相机可附加地或替换地包括所述NUI相机可见的两个或更多个参考标记。在这样的示例中,所述校准相机可附加地或替换地包括显示屏,所述显示屏被配置成呈现所述NUI相机可见的两个或更多个参考标记。在这样的示例中,所述NUI相机可附加地或替换地为深度相机。在这样的示例中,所述显示器可附加地或替换地被配置成呈现所述校准相机可见的校准图像。在这样的示例中,所述校准相机可附加地或替换地被包括在用户可穿戴的头戴式增强现实设备中。以上描述的示例中的任何一个或全部可按任何合适的方式被组合在各实现中。

[0086] 在另一示例中,提供了一种使用深度相机来控制计算系统的方法,所述方法包括在所述计算系统处接收相对于所述深度相机在环境内的6DoF姿态来定义显示器的6DoF姿态的三维模型,在所述计算系统处接收所述环境内的用户的机器可读身体模型,所述机器可读身体模型包括所述用户的第一身体部位的三维位置以及所述用户的第二身体部位的三维位置,在第一条件期间,将所述用户的所述第一身体部位的位置和所述用户的所述第二身体部位的位置解释成控制所述计算系统的物理交互区姿势,以及,在第二条件期间,将所述用户的所述第一身体部位相对于所述显示器的位置和所述用户的所述第二身体部位相对于所述显示器的位置解释成控制所述计算系统的指向姿势。在这样的示例中,所述方法可附加地或替换地包括响应于接收来自所述用户的命令而从所述第一条件转变到所述第二条件。在这样的示例中,所述方法可附加地或替换地包括响应于接收来自在所述计算系统上执行的应用的指令而从所述第一条件转变到所述第二条件。在这样的示例中,所述方法可附加地或替换地包括响应于识别到所述第二身体部位相对于所述显示器的距离已减小到低于阈值而从所述第一条件转变到所述第二条件。在这样的示例中,所述方法可附加地或替换地包括在所述第二条件期间,在所述显示器上可视地指示所述显示器和从所述第一身体部位起延伸通过所述第二身体部位的线的交点。以上描述的示例中的任何一个或全部可按任何合适的方式被组合在各实现中。

[0087] 在又一示例中,提供了一种校准自然用户界面(NUI)相机相对于显示器的六自由度(6DoF)姿态的方法,所述方法包括从所述NUI相机接收从校准角度对环境进行成像的校准视频,所述校准视频设置所述显示器和一个或多个特征。在这样的示例中,所述方法可附

加地或替换地包括根据所述校准视频三维地建模所述环境的三维图,所述环境的所述三维图定义所述显示器在所述环境内的6DoF姿态以及所述一个或多个特征中的每一者在所述环境内的三维位置。在这样的示例中,所述方法可附加地或替换地包括从所述NUI相机接收从操作角度对所述环境进行成像的主视频,所述主视频设置所述一个或多个特征,并可附加地或替换地包括基于所述一个或多个特征的操作透视图,在所述环境的三维图内寻找所述NUI相机的6DoF姿态。以上描述的示例中的任何一个或全部可按任何合适的方式被组合在各实现中。

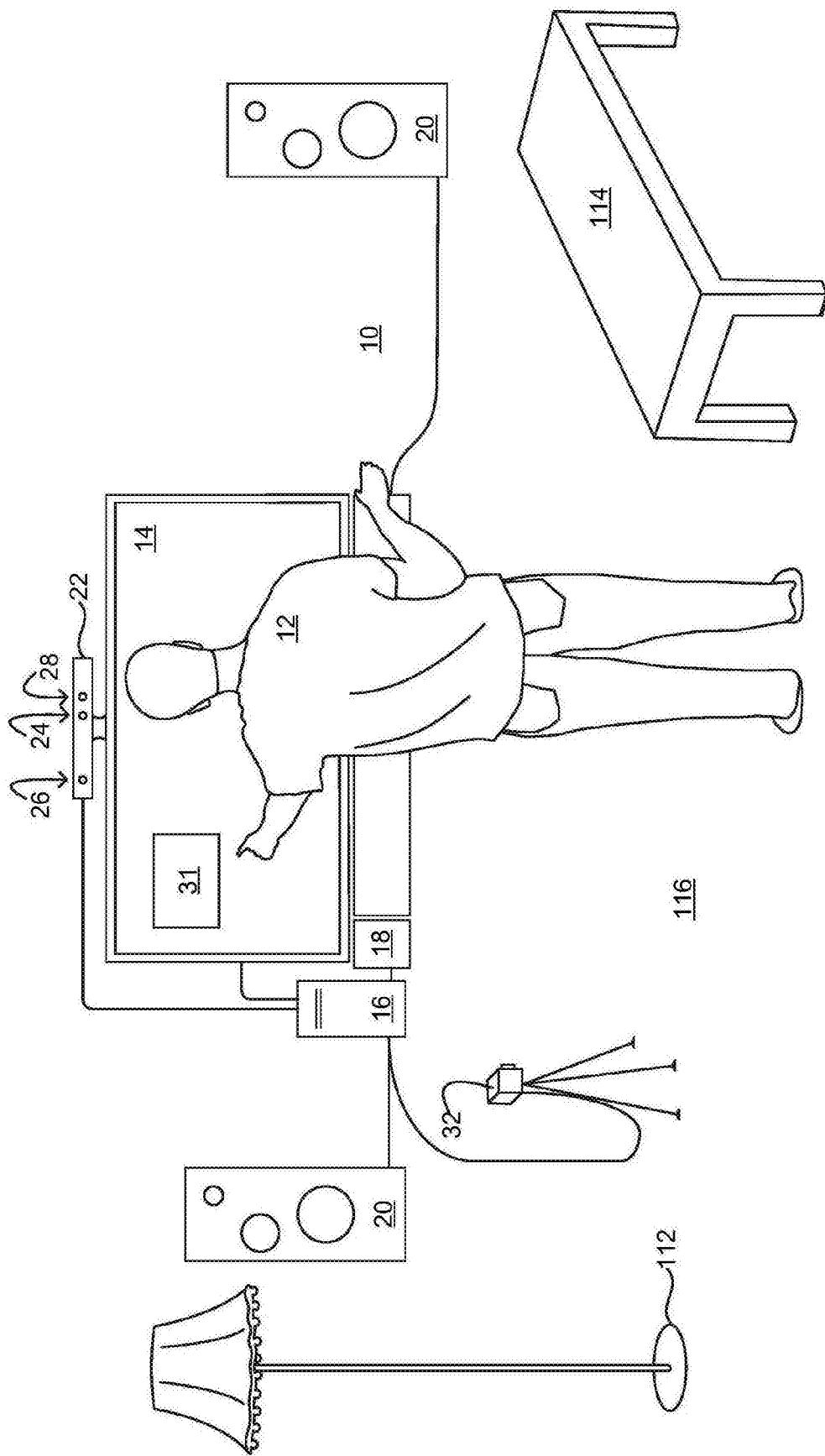


图1

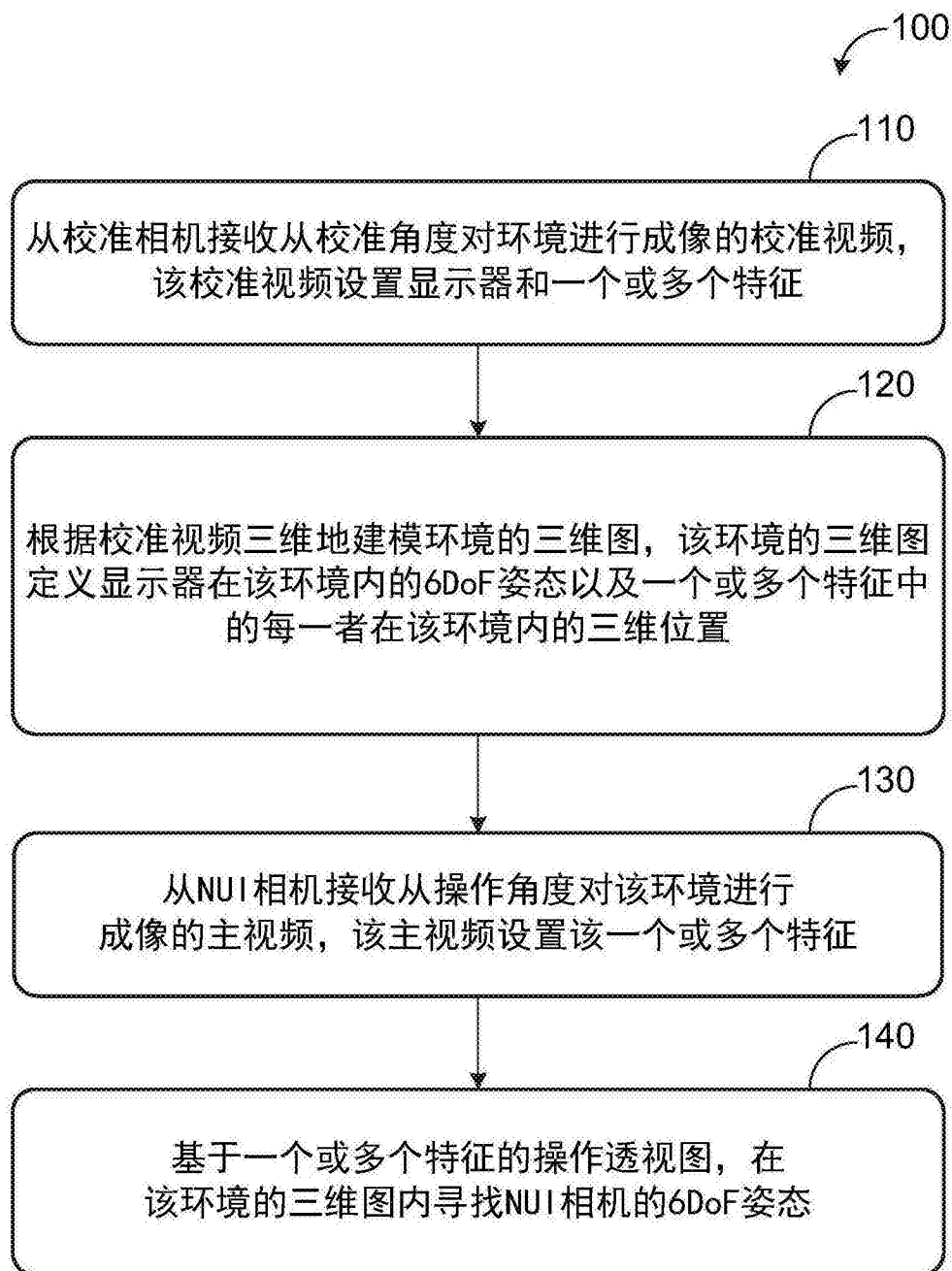


图3

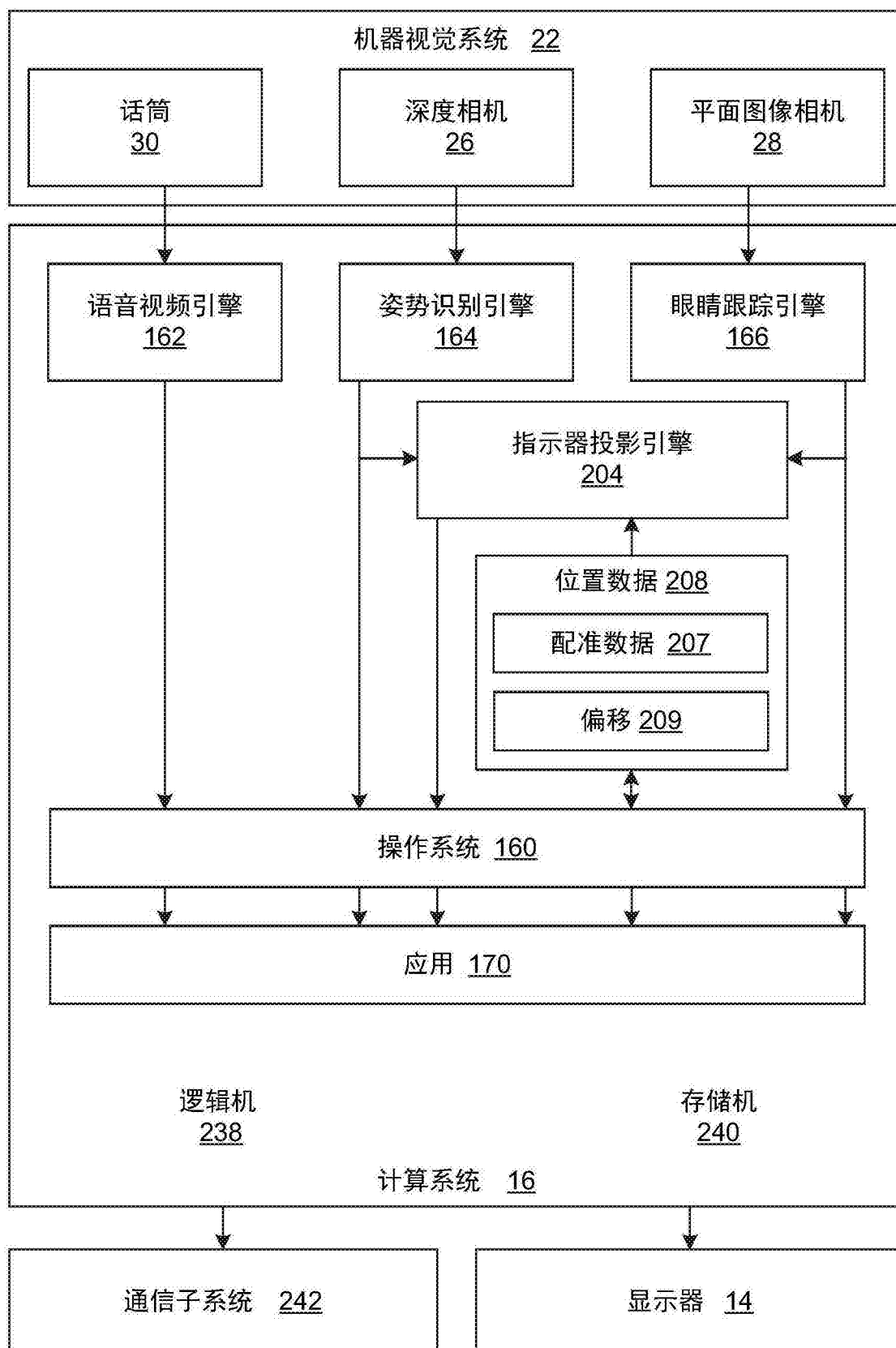


图4

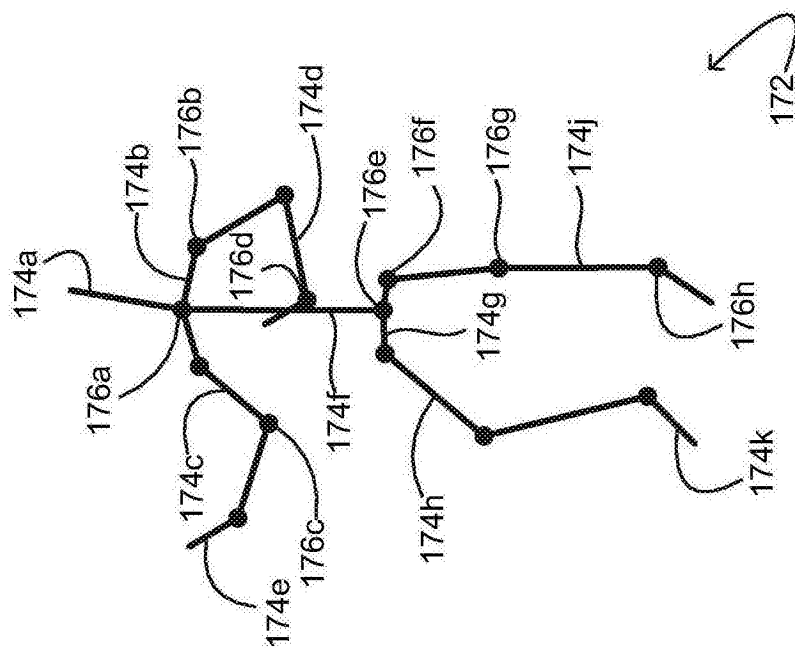


图5

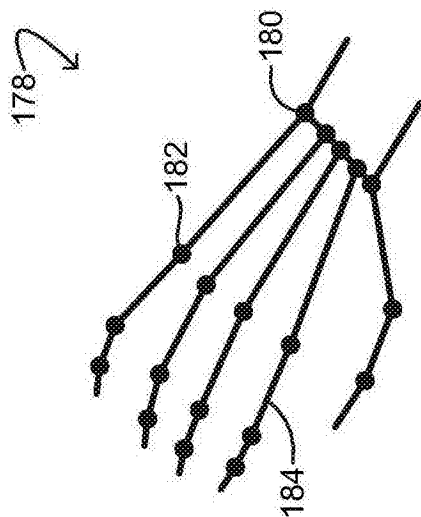


图6

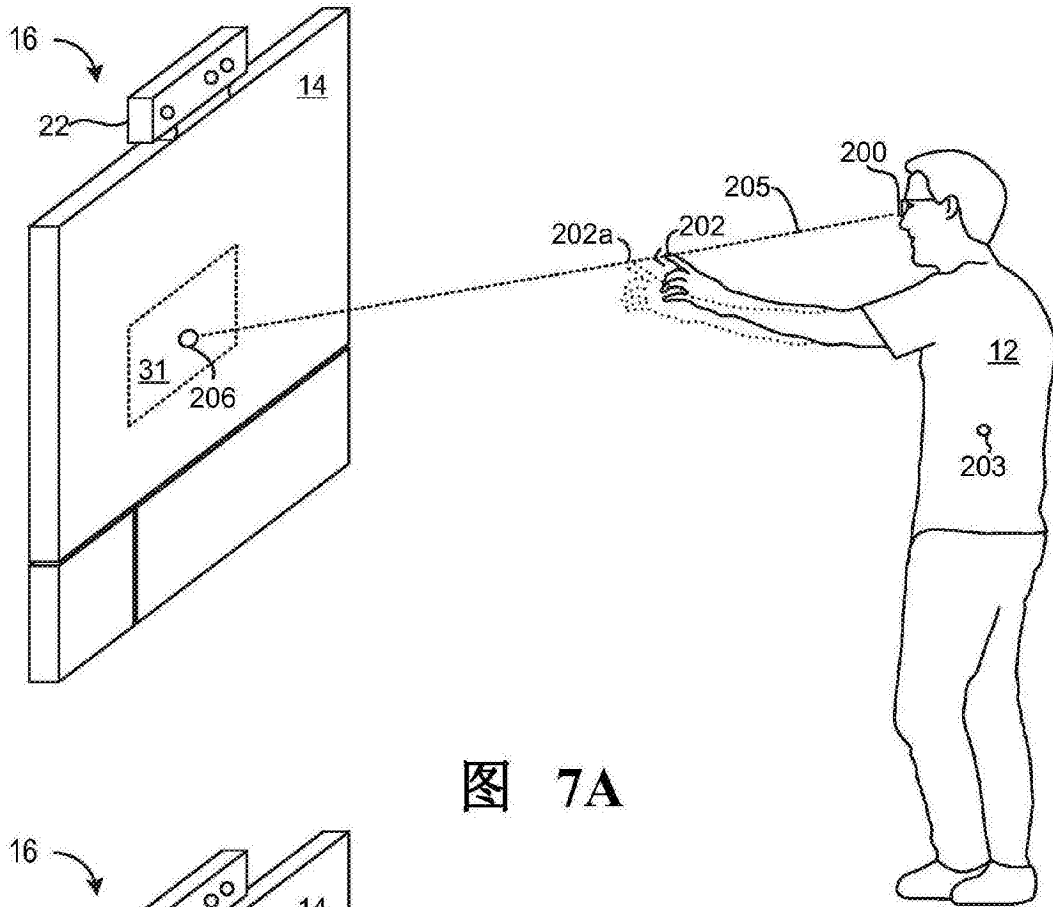


图 7A

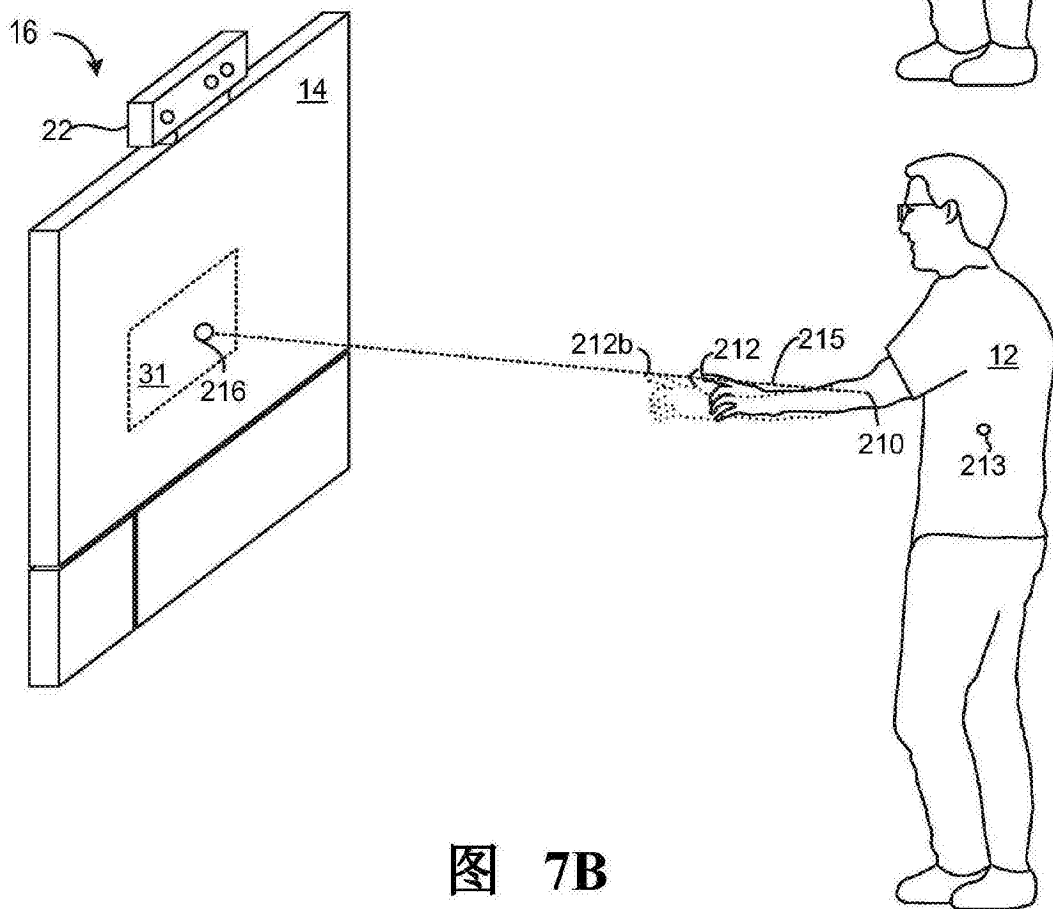


图 7B

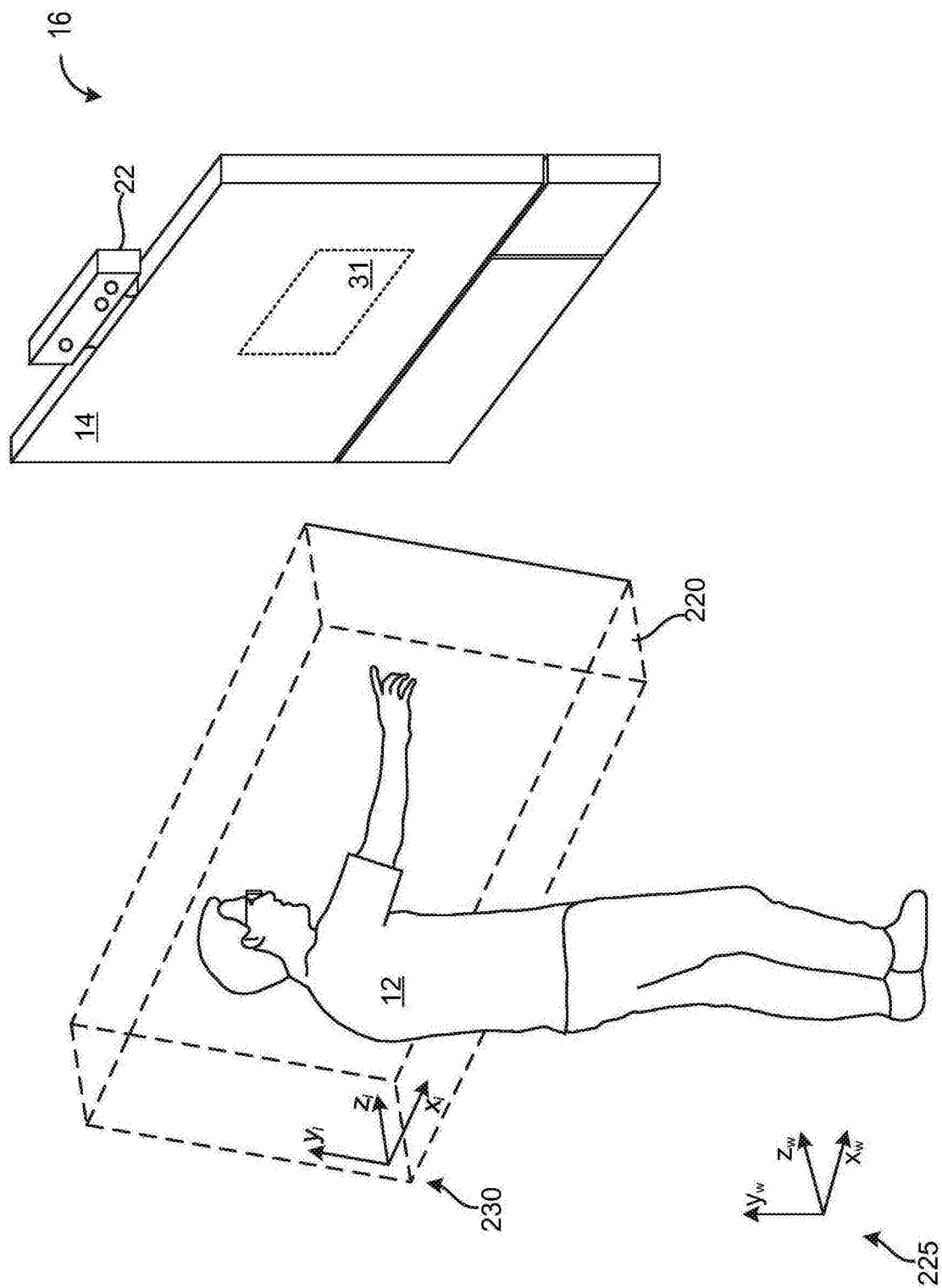


图8