



(12) 发明专利申请

(10) 申请公布号 CN 102375541 A

(43) 申请公布日 2012. 03. 14

(21) 申请号 201110280552. 9

(22) 申请日 2011. 08. 19

(30) 优先权数据

12/859, 995 2010. 08. 20 US

(71) 申请人 微软公司

地址 美国华盛顿州

(72) 发明人 O·O·G·桑托斯 M·黑格

C·武切蒂奇 B·欣德尔

D·A·贝内特

(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

(51) Int. Cl.

G06F 3/01 (2006. 01)

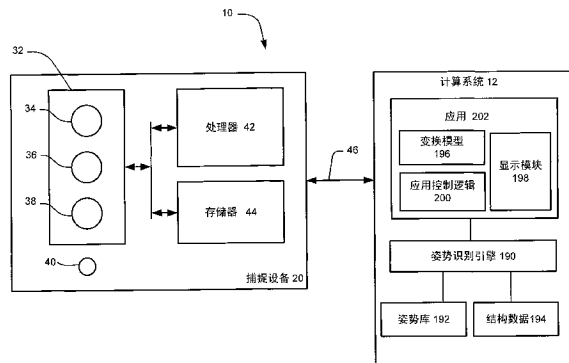
权利要求书 2 页 说明书 16 页 附图 14 页

(54) 发明名称

将用户运动转换成多个对象响应

(57) 摘要

本发明涉及将用户运动转换成多个对象响应的方法和系统。提供了一种基于计算设备上执行的用户的用户交互, 来将用户运动转换成屏幕上对象的多个对象响应的系统。从捕捉设备处接收来自一个或多个用户的用户运动数据。所述用户运动数据对应于与应用中呈现的屏幕上对象所进行的用户交互。该屏幕上对象对应于除由计算设备显示的用户的屏幕上表示以外的对象。将用户运动数据自动转换成屏幕上对象的多个对象响应。将屏幕上对象的多个对象响应同时显示给用户。



1. 一种基于与计算设备 (12) 上执行的应用 (202) 所进行的用户交互, 将用户运动转换成多个对象响应的方法, 包括:

从传感器 (34、36、38) 处接收一个或多个用户的用户运动数据;

确定所述用户运动数据是否匹配一个或多个预定义的姿势;

在所述用户运动数据匹配一个或多个预定义的姿势时, 将所述用户运动数据自动转换成屏幕上对象 (52) 的多个对象响应, 所述屏幕上对象 (52) 对应于除由执行所述应用 (202) 的计算设备 (12) 显示的用户的屏幕上表示以外的对象; 以及

基于所述转换来同时显示所述屏幕上对象 (52) 的所述多个对象响应。

2. 如权利要求 1 所述的方法, 其特征在于:

所述多个对象响应中的至少一个包括所述屏幕上对象的运动响应。

3. 如权利要求 1 所述的方法, 其特征在于:

所述多个对象响应中的至少一个是所述屏幕上对象的音频响应。

4. 如权利要求 1 所述的方法, 其特征在于:

所述多个对象响应中的至少一个是所述屏幕上对象的视觉响应。

5. 如权利要求 1 所述的方法, 其特征在于, 将所述用户运动数据转换成多个对象响应还包括:

基于确定所述用户运动数据是否匹配所述一个或多个预定义的姿势, 来访问将姿势与对象响应进行相关的数据结构;

访问所述屏幕上对象的相应响应; 以及

实现所述对象响应, 所述实现包括将所述用户的骨架模型表示映射成所述屏幕上对象的对象模型表示。

6. 如权利要求 1 所述的方法, 其特征在于, 从一个或多个用户处接收所述用户运动数据进一步包括:

从第一用户处接收第一用户运动数据, 并从第二用户处接收第二用户运动数据。

7. 如权利要求 6 所述的方法, 其特征在于, 还包括:

将所述第一用户运动数据转换成所述屏幕上对象的第一对象响应;

将所述第二用户运动数据转换成所述屏幕上对象的第二对象响应;

将所述屏幕上对象的所述第一对象响应和所述第二对象响应同时显示给所述第一用户和所述第二用户。

8. 如权利要求 7 所述的方法, 其特征在于:

当所述第二用户运动数据与所述第一用户运动数据不相同, 所述第二对象响应与所述第一对象响应不相同。

9. 如权利要求 7 所述的方法, 其特征在于:

当所述第二用户运动数据与所述第一用户运动数据相同时, 所述第二对象响应是所述第一用户响应的放大。

10. 如权利要求 1 所述的方法, 其特征在于:

接收所述用户运动数据包括从第一用户处接收第一用户运动数据以及从第二用户处接收第二用户运动数据;

将所述用户运动数据转换成多个对象响应包括将所述第一用户运动数据转换成所述

屏幕上对象的第一运动响应,将所述第二用户运动数据转换成所述屏幕上对象的第二运动响应;

将所述用户运动数据转换成多个对象响应包括将所述用户运动数据的骨架模型表示映射成所述屏幕上对象的对象模型表示;

基于所述转换来显示所述屏幕上对象的多个对象响应包括显示所述屏幕上对象的运动响应、音频响应或视觉响应中的至少一个。

11. 如权利要求 1 所述的方法,其特征在于:

所述一个或多个屏幕上对象包括在所述应用中呈现的无生命的对象;并且其中:

所述多个对象响应中的至少一个包括所述无生命的对象的运动。

12. 一个或多个处理器可读存储设备,在所述一个或多个处理器可读存储设备上包括有处理器可读代码,所述处理器可读代码用于对一个或多个处理器编程以执行一种方法,包括:

从传感器(34、36、38)处接收第一用户(18)的第一用户运动数据(54),所述第一用户运动数据(54)对应于与在计算设备(12)上执行的应用(202)中呈现的屏幕上对象(52)所进行的第一用户交互;

从所述传感器(34、36、38)处接收第二用户(19)的第二用户运动数据(53),所述第二用户运动数据(53)对应于与在计算设备(12)上执行的应用(202)中呈现的所述屏幕上对象(52)所进行的第二用户交互;

将所述第一用户运动数据(54)转换成所述屏幕上对象(52)的第一运动响应(56、57);

将所述第二用户运动数据(54)转换成所述屏幕上对象(52)的第二运动响应(59);以及

将所述第一运动响应(56、57)和所述第二运动响应(59)同时显示给所述第一用户(18)和所述第二用户(19)。

13. 如权利要求 12 所述的一个或多个处理器可读存储设备,其特征在于:

当所述第二用户运动数据与所述第一用户运动数据不相同,所述第二运动响应与所述第一对象响应不相同。

14. 如权利要求 12 所述的一个或多个处理器可读存储设备,其特征在于:

当所述第二用户运动数据与所述第一用户运动数据相同时,所述第二运动响应是所述第一运动响应的放大。

15. 一种基于与应用(202)的用户交互来触发对象响应的装置,包括:

捕捉用户运动数据的传感器(34、36、38);以及

连接至所述传感器(34、36、38)的计算设备(12),它将所述用户运动数据转换成无生命的对象(62)的运动响应、音频响应或视觉响应中的至少一个,所述计算设备(12)基于所述转换来显示所述无生命的对象(62)的所述运动响应、所述音频响应或所述视觉响应中的至少一个。

将用户运动转换成多个对象响应

技术领域

[0001] 本发明涉及将用户运动转换成多个对象响应的方法和系统。

背景技术

[0002] 过去,诸如计算机游戏和多媒体应用等计算应用使用控制器、遥控器、键盘、鼠标等来允许用户操纵游戏角色。近来,计算机游戏和多媒体应用已开始使用照相机和软件姿势识别来提供人机接口(“HCI”)。使用 HCI,用户姿势形式的用户交互被检测、解释和使用来控制游戏角色。

发明内容

[0003] 公开了一种技术,它通过允许用户与传统上为非交互式的和静态的各种屏幕上对象进行交互,来增强与应用进行的用户交互。例如,可使用用户运动来改变、控制或移动除用户的屏幕上表示以外的对象。捕捉与要和应用进行交互的用户意图有关的用户运动,并将该用户运动转换成屏幕上对象的多个对象响应。屏幕上对象的多个对象响应为与应用进行交互的用户提供了增强的和改进的用户体验,而不会改变用户对应用进行的典型的或以其他方式的被动交互的结果。

[0004] 在一个实施例中,提供了一种基于与在计算设备上执行的应用所进行的用户交互来将用户运动转换成屏幕上对象的多个对象响应的方法。从捕捉设备接收来自一个或多个用户的用户运动数据。该用户运动数据对应于与应用中呈现的屏幕上对象所进行的用户交互。该屏幕上对象对应于除由计算设备显示的用户的屏幕上表示以外的对象。将用户运动数据自动转换成屏幕上对象的多个对象响应。将屏幕上对象的多个对象响应同时显示给用户。

[0005] 提供本概述以便以简化的形式介绍将在以下详细描述中进一步描述的一些概念。本发明内容并不旨在标识出所要求保护的主题的关键特征或必要特征,也不旨在用于帮助确定所要求保护的主题的范围。此外,所要求保护的主题不限于解决在本公开的任何部分中提及的任何或所有缺点的各实现。

附图说明

[0006] 图 1 示出了用户正在玩游戏的跟踪系统的一个实施例。

[0007] 图 2 示出了可用作跟踪系统的一部分的捕捉设备的一个实施例。

[0008] 图 3 示出了可用来跟踪运动并基于所跟踪的运动来更新应用的计算设备的一个实施例。

[0009] 图 4 示出了可用来跟踪运动并基于所跟踪的运动来更新应用的计算设备的一个实施例。

[0010] 图 5 是描述了用于执行本公开技术中的各操作的过程的一个实施例的流程图。

[0011] 图 6 是描述了根据本公开技术的用于捕捉用户运动数据的过程的一个实施例的

流程图。

[0012] 图 7 示出了表示所扫描的人类目标的骨架模型或映射的示例。

[0013] 图 8 提供了图 2 中示出的姿势识别引擎的示例性实施例的进一步细节。

[0014] 图 9 是描述了根据本公开技术的用于确定用户运动数据是否匹配一姿势的过程的一个实施例的流程图。

[0015] 图 10-14 示出了根据本公开技术的描绘了与计算设备上执行的应用所进行的用户交互的各种用户界面屏幕。

具体实施方式

[0016] 公开了一种技术,通过该技术与在计算设备上执行的应用所进行的用户交互得到了增强,从而使得用户能够与不是用户的屏幕上表示的屏幕上对象进行交互。捕捉设备捕捉涉及与应用进行交互的用户意图的用户运动数据。计算设备连接至所述捕捉设备,并从所述捕捉设备接收与用户运动数据有关的信息,以控制应用的各个方面。应用可以包括例如在所述计算设备上执行的视频游戏应用或电影应用。一个或多个用户经由计算设备中的用户界面来与由该应用描绘的屏幕上对象进行交互。在一个实施例中,屏幕上对象对应于例如非传统的玩游戏元素,诸如在非交互式或静态体验(诸如例如由应用呈现的静态电影序列、故事序列、影片序列或动画)期间显示的有生命的对象或者无生命的对象。

[0017] 在一个实施例中,计算设备将用户运动数据转换成屏幕上对象的多个对象响应。用户运动数据可从与应用交互的一个或多个用户处接收。屏幕上对象的多个对象响应可包括屏幕上对象的运动响应、音频响应或视觉响应。可以经由计算设备中的用户界面来将对象响应同时显示给用户。

[0018] 图 1 示出了用于执行本公开技术的各操作的目标识别、分析和跟踪系统 10 的一个实施例(下文中通常称为跟踪系统)。目标识别、分析和跟踪系统 10 可用来识别、分析和/或跟踪诸如用户 18 等的人类目标。如图 1 所示,跟踪系统 10 可包括计算设备 12。计算设备 12 可以是计算机、游戏系统或控制台等等。根据一个实施例,计算设备 12 可以包括硬件组件和/或软件组件,从而使得计算设备 12 可用来执行诸如游戏应用、非游戏应用之类的应用。在一个实施例中,计算设备 12 可包括可执行在处理器可读存储设备上存储的、用于执行在此描述的过程的指令的处理器,诸如标准化处理器、专用处理器、微处理器等。

[0019] 如图 1 所示,跟踪系统 10 还可包括捕捉设备 20。捕捉设备 20 可以是例如照相机,该照相机可用于在视觉上监视诸如用户 18 等的一个或多个用户,从而使得可以捕捉、分析并跟踪由所述一个或多个用户作出的姿势,以便控制在计算设备 12 上执行的应用的各个方面。

[0020] 根据一个实施例,跟踪系统 10 可连接至可向诸如用户 18 等的用户提供游戏或应用视觉和/或音频的视听设备 16,诸如电视机、监视器、高清电视机(HDTV)等。例如,计算设备 12 可包括诸如图形卡等视频适配器和/或诸如声卡等音频适配器,这些适配器可提供与游戏应用、非游戏应用等相关联的视听信号。视听设备 16 可从计算设备 12 接收视听信号,然后可向用户 18 输出与视听信号相关联的游戏或应用视觉和/或音频。根据一个实施例,视听设备 16 可经由例如,S-视频电缆、同轴电缆、HDMI 电缆、DVI 电缆、VGA 电缆等连接到计算设备 12。

[0021] 该目标识别、分析和跟踪系统 10 可用来识别、分析和 / 或跟踪诸如用户 18 等的一个或多个人类目标。例如,可使用捕捉设备 20 来跟踪用户 18,使得可将用户 18 的移动解释为可用于影响由计算设备 12 执行的应用或操作系统的控制。

[0022] 图 2 示出了捕捉设备 20 和计算设备 12 的一个实施例,它们可以在目标识别、分析和跟踪系统 10 中使用,以便识别捕捉区域中的人类或非人类目标,并在三维空间中唯一地标识他们及跟踪他们。根据一个实施例,捕捉设备 20 可被配置成经由任何合适的技术,包括例如飞行时间、结构化光、立体图像等来捕捉具有深度信息的视频,该深度信息包括可以包括深度值的深度图像。根据一实施例,捕捉设备 20 可将所计算的深度信息组织为“Z 层”或可垂直于从深度相机沿其视线延伸的 Z 轴的层。

[0023] 如图 2 所示,捕捉设备 20 可以包括图像相机组件 32。根据一个实施例,图像相机组件 32 可以是可捕捉场景的深度图像的深度相机。深度图像可包括所捕捉的场景的二维 (2-D) 像素区域,其中 2-D 像素区域中的每一像素可表示深度值,诸如例如以厘米、毫米等计的所捕捉的场景中的对象距相机的距离。

[0024] 如图 2 所示,图像相机组件 32 可包括可被用来捕捉捕捉区域的深度图像的 IR 光组件 34、三维 (3-D) 相机 36、以及 RGB 相机 38。例如,在飞行时间分析中,捕捉设备 20 的 IR 光组件 34 可以将红外光发射到捕捉区域上,然后可以使用传感器,用例如 3-D 相机 36 和 / 或 RGB 相机 38 来检测从捕捉区域中的一个或多个目标和对象的表面反向散射的光。在某些实施例中,可以使用脉冲式红外光从而可以测量出射光脉冲和相应的入射光脉冲之间的时间差并将其用于确定从捕捉设备 20 到捕捉区域中的目标或对象上的特定位置的物理距离。此外,可将出射光波的相位与入射光波的相位进行比较来确定相移。然后可以使用该相移来确定从捕捉设备到目标或对象上的特定位置的物理距离。

[0025] 根据另一个实施例,可使用飞行时间分析,通过经由包括例如快门式光脉冲成像的各种技术来分析反射光束随时间的强度变化以间接地确定从捕捉设备 20 到目标或对象上的特定位置的物理距离。

[0026] 在另一示例中,捕捉设备 20 可使用结构化光来捕捉深度信息。在该分析中,图案化光 (即,被显示为诸如网格图案或条纹图案等已知图案的光) 可经由例如 IR 光组件 34 被投影到捕捉区域上。在撞击到捕捉区域中的一个或多个目标或对象的表面时,作为响应,图案可变形。图案的这种变形可由例如 3-D 相机 36 和 / 或 RGB 相机 38 来捕捉,然后可被分析来确定从捕捉设备到目标或对象上的特定位置的物理距离。

[0027] 根据一个实施例,捕捉设备 20 可包括可以从不同的角度观察捕捉区域的两个或更多在物理上分离的相机,以获取可以被解析以生成深度信息的视觉立体数据。也可使用其他类型的深度图像传感器来创建深度图像。

[0028] 捕捉设备 20 还可包括话筒 40。话筒 40 可包括可接收声音并将其转换成电信号的换能器或传感器。根据一个实施例,话筒 40 可用来减少在目标识别、分析和跟踪系统 10 中的捕捉设备 20 与计算设备 12 之间的反馈。另外,话筒 40 可用来接收也可由用户提供的音频信号,以控制可由计算设备 12 执行的诸如游戏应用、非游戏应用等应用。

[0029] 在一个实施例中,捕捉设备 20 还可以包括可以与图像相机组件 32 进行可操作的通信的处理器 42。处理器 42 可包括可执行指令的标准处理器、专用处理器、微处理器等,这些指令可包括用于存储简档的指令、用于接收深度图像的指令、用于确定合适的目标是否

被包括在深度图像中的指令、用于将合适的目标转换成该目标的骨架表示或模型的指令、或任何其他合适的指令。

[0030] 捕捉设备 20 还可包括存储器组件 44, 存储器组件 44 可存储可由处理器 42 执行的指令、由 3-D 相机或 RGB 相机所捕捉的图像或图像的帧、用户简档、或任何其他合适的信息、图像等等。根据一个示例, 存储器组件 44 可包括随机存取存储器 (RAM)、只读存储器 (ROM)、高速缓存、闪存、硬盘或任何其他合适的存储组件。如图 2 所示, 存储器组件 44 可以是与图像捕捉组件 32 和处理器 42 进行通信的单独的组件。在另一实施例中, 存储器组件 44 可被集成到处理器 42 和 / 或图像捕捉组件 32 中。在一个实施例中, 图 2 中示出的捕捉设备 20 的组件 32、34、36、38、40、42 和 44 中的一些或全部被容纳在单个外壳中。

[0031] 捕捉设备 20 可以经由通信链路 46 来与计算设备 12 进行通信。通信链路 46 可以是包括例如 USB 连接、火线连接、以太网电缆连接等的有线连接和 / 或诸如无线 802. 11b、802. 11g、802. 11a 或 802. 11n 连接等无线连接。计算设备 12 可以向捕捉设备 20 提供时钟, 该时钟可以用来确定何时经由通信链路 46 来捕捉例如场景。

[0032] 捕捉设备 20 可经由通信链路 46 来向计算设备 12 提供由例如 3-D 相机 36 和 / 或 RGB 相机 38 捕捉的深度信息和图像, 包括可由捕捉设备 20 生成的骨架模型。计算设备 12 然后可使用该骨架模型、深度信息和所捕捉的图像来例如创建视觉屏幕, 并控制诸如游戏或文字处理程序之类的应用。

[0033] 计算设备 12 包括姿势库 192、结构数据 194 和姿势识别引擎 190。姿势库 192 可包括姿势过滤器的集合, 每一姿势过滤器包括描述或定义可由骨架模型 (在用户移动时) 作出的运动或姿势的信息。结构数据 192 包括关于可被跟踪的对象的结构信息。例如, 可以存储人类的骨架模型以帮助理解用户的移动并识别身体部位。还可以存储关于非生命对象的结构信息以帮助识别这些对象并帮助理解移动。

[0034] 姿势识别引擎 190 可将由相机 36、38 和设备 20 捕捉的、骨架模型以及与其相关联的移动等形式的数据与姿势库 192 中的姿势过滤器进行比较, 以标识 (如由骨架模型表示的) 用户何时作出了一个或多个姿势。计算设备 12 可使用姿势库 192 来解释骨架模型的移动并基于该移动来控制应用。关于姿势识别引擎 190 的更多信息参见 2009 年 4 月 13 日提交的美国专利申请 12/422, 661 “Gesture Recognition System Architecture (姿势识别系统架构)”, 该申请通过整体引用合并于此。关于识别姿势的更多信息参见 2009 年 2 月 23 日提交的美国专利申请 12/391, 150 “Standard Gestures (标准姿势)”; 以及 2009 年 5 月 29 日提交的美国专利申请 12/474, 655 “Gesture Tool (姿势工具)”, 这两个申请都通过整体引用结合于此。关于运动检测和跟踪的更多信息可在 2009 年 12 月 18 日提交的美国专利申请 12/641, 788 “Motion Detection Using Depth Images (使用深度图像的运动检测)”, 以及美国专利申请 12/475, 308 “Device for Identifying and Tracking Multiple Humans over Time (用于随时间标识和跟踪多个人类的设备)” 中找到, 这两个申请通过整体引用合并于此。

[0035] 在一个实施例中, 计算设备 12 可包括应用 202。应用 202 可包括在计算设备 12 上执行的视频游戏应用、电影应用、购物应用、浏览应用或其他应用。应用 202 允许与由该应用 202 呈现的一个或更多屏幕上对象所进行的用户交互。所述屏幕上对象可对应于由应用 202 呈现的例如非传统的玩游戏元素, 诸如在非交互式或静态体验 (诸如例如静态的电影

序列、故事序列、影片序列或动画)期间显示的有生命的对象或者无生命的对象。应用 202 可包括一个或多个场景。应用 202 中的场景可包括例如用于在玩游戏期间发展情节的背景信息。例如,应用 202 中的场景可包括由应用 202 呈现的静态电影序列、故事序列、影片序列或动画,它们描绘了一个或多个有生命的或无生命的屏幕上对象。在一个示例中,屏幕上对象可对应于除用户的屏幕上角色表示以外的对象。在一个实施例中,用户可经由用户的计算设备 12 中的用户界面来与应用 102 进行交互。

[0036] 应用 202 可包括变换模型 196、显示模块 198 和应用控制逻辑 200。变换模型 196、显示模块 198 和应用控制逻辑 200 可以作为软件模块来实现,以执行本公开技术的一个或多个操作。应用控制逻辑 200 可包括与应用 202 的执行有关的预编程逻辑和 / 或规则的集合。在一个实施例中,由应用控制逻辑 200 指定的规则可基于与应用 202 所进行的用户交互来定义可用于控制由程序 202 呈现的屏幕上对象的方式。在一个示例中,由应用控制逻辑 200 指定的规则可基于与应用 202 所进行的用户交互来定义在屏幕上对象上执行的动作的类型。如将在下文中详细讨论的,用户可以作出一定范围的运动,以便与应用中描绘的一个或多个屏幕上对象进行交互,其中不同的用户运动可以导致相似的或不同的屏幕上对象的相应对象动作。

[0037] 在一个实施例中,姿势识别引擎 190 可向应用控制逻辑 200 提供人类目标的骨架模型和关于人类目标的移动的信息。在一个实施例中,关于人类目标的移动的信息可包括例如与人类目标相关联的各个身体部位的位置、方向、加速度和曲率。应用控制逻辑 200 可利用该信息来定义要在显示在应用 202 中的屏幕上对象上执行的动作的类型。在一个实施例中,由应用控制逻辑 200 指定的规则可被实现为将由姿势识别引擎 190 标识的一组姿势与屏幕上对象的一组对象响应进行相关的数据结构。

[0038] 在另一个实施例中,应用控制逻辑 200 也可以定义其中可以描绘应用 202 的屏幕上对象中的一个或多个的环境上下文。因此,要在屏幕上对象上执行的动作的类型可以基于与用户运动数据(人类目标的移动)有关的信息以及其中描绘有屏幕上对象的环境上下文。例如,如果在应用中描绘了诸如鲨鱼等有生命的屏幕上对象,则环境上下文可以对应于第一上下文。在这种上下文中,用户的髋部的从左到右的移动例如可造成使得鲨鱼的鳍部来回移动的动作,然而在描绘了诸如卡车等无生命的对象的第二上下文中,同样的动作却可造成在卡车上执行拐弯的动作。

[0039] 变换模型 196 可包括基于应用控制逻辑 200 所定义的规则而在特定屏幕上对象上执行特定动作的可执行指令。在一个实施例中,变换模型 196 可通过将用户运动数据自动转换成屏幕上对象的多个对象响应来在屏幕上对象上执行特定动作。在一个实施例中,变换模型 196 可以通过从由程序控制逻辑 200 定义的数据结构中访问屏幕上对象的相应响应来执行该转换。在一个实施例中,屏幕上对象的相应响应可包括该屏幕上对象的运动响应、视觉响应或音频响应。变换模型 196 然后可访问实现该对象响应的代码。在一个实施例中,可通过将用户的骨架模型表示映射成屏幕上对象的对象模型表示来实现该对象响应。例如,变换模型 196 可通过将从人类目标的骨架模型表示中获取的身体部位的移动映射成基于屏幕上对象的对象模型表示的屏幕上对象中的某部分的相应移动,来将用户运动数据转换成屏幕上对象的相应运动响应。类似地,变换模块 196 可通过利用屏幕上对象的对象模型表示来将用户运动数据转换成该屏幕上对象的相应的音频响应或视觉响应。

[0040] 在一个实施例中,变换模型 196 在执行具体移动以执行屏幕上对象的相应的所转换运动时,可利用与人类目标的移动有关的元数据,诸如身体部位的速度、身体部位的加速度或身体部位行进的距离。例如,人类目标的具体身体部位的更高速度的移动可造成将用户运动数据转换成屏幕上对象的运动响应,其中响应于更高速度的移动,屏幕上对象移动得更快。类似地,变换模型 196 可以利用与人类目标的移动有关的元数据(诸如身体部位的速度、身体部位的加速度或身体部位行进得距离等)来触发屏幕上对象的音频响应或视觉响应。

[0041] 显示模块 198 控制要向用户显示什么,并可包括用于基于从变换模型 196 处接收到的信息来在计算设备 12 的用户界面上显示屏幕上对象的多个对象响应的可执行指令。

[0042] 图 3 示出了可用来实现图 1-2 的计算设备 12 的计算设备 100 的示例。图 3 的计算设备 100 可以是诸如游戏控制台等多媒体控制台 100。如图 3 所示,多媒体控制台 100 具有中央处理单元(CPU)200 以及便于处理器访问各种类型存储器的存储器控制器 202,各种类型存储器包括闪速只读存储器(ROM)204、随机存取存储器(RAM)206、硬盘驱动器 208、以及便携式媒体驱动器 106。在一种实现中,CPU 200 包括 1 级高速缓存 210 和 2 级高速缓存 212,用于临时存储数据,并且因此减少对硬盘驱动器 208 进行的存储器访问周期的数量,从而提高处理速度和吞吐量。

[0043] CPU 200、存储器控制器 202、以及各种存储器设备经由一个或多个总线(未示出)互连在一起。在此实现中所使用的总线的细节对理解此处所讨论的主题不是特别相关。然而,应该理解,这样的总线可以包括串行和并行总线、存储器总线、外围总线、使用各种总线体系结构中的任何一种的处理器或局部总线中的一个或多个。作为示例,这样的体系结构可以包括工业标准体系结构(ISA)总线、微通道体系结构(MCA)总线、增强型 ISA(EISA)总线、视频电子标准协会(VESA)局部总线、以及也称为夹层总线的外围部件互连(PCI)总线。

[0044] 在一种实现中,CPU 200、存储器控制器 202、ROM 204、以及 RAM 206 被集成到公用模块 214 上。在此实现中,ROM 204 被配置为通过 PCI 总线和 ROM 总线(两者都没有示出)连接到存储器控制器 202 的闪速 ROM。RAM 206 被配置为多个双倍数据速率同步动态 RAM(DDR SDRAM)模块,它们被存储器控制器 202 通过分开的总线(未示出)独立地进行控制。硬盘驱动器 208 和便携式媒体驱动器 106 被示为通过 PCI 总线和 AT Attachment(ATA)总线 216 连接到存储器控制器 202。然而,在其他实现中,也可以备选地应用不同类型的专用数据总线结构。

[0045] 图形处理单元 220 和视频编码器 222 构成了用于进行高速度和高分辨率(例如,高清晰度)的图形处理的视频处理流水线。数据通过数字视频总线(未示出)从图形处理单元 220 传输到视频编码器 222。音频处理单元 224 和音频编码解码器(编码器/解码器)226 构成了对应的音频处理流水线,用于对各种数字音频格式进行多通道音频处理。通过通信链路(未示出)在音频处理单元 224 和音频编码解码器 226 之间传输音频数据。视频和音频处理流水线向 A/V(音频/视频)端口 228 输出数据,以便传输到电视机或其他显示器。在所示出的实现中,视频和音频处理组件 220-228 安装在模块 214 上。

[0046] 图 3 示出了包括 USB 主控制器 230 和网络接口 232 的模块 214。USB 主控制器 230 被示为通过总线(例如,PCI 总线)与 CPU 200 和存储器控制器 202 进行通信,并作为外围控制器 104(1)-104(4)的主机。网络接口 232 提供对网络(例如因特网、家庭网络等)的

访问,并且可以是包括以太网卡、调制解调器、无线接入卡、蓝牙模块、电缆调制解调器等各种有线或无线接口组件中的任一种。

[0047] 在图 3 中所描绘的实现中,控制台 102 包括用于支持四个控制器 104(1)-104(4)的控制器支持子组件 240。控制器支持子组件 240 包括支持与诸如,例如,媒体和游戏控制器之类的外部控制设备的有线和无线操作所需的任何硬件和软件组件。前面板 I/O 子部件 242 支持电源按钮 112、弹出按钮 114,以及任何 LED(发光二极管)或暴露在控制台 102 的外表面上的其他指示灯等多个功能。子部件 240 和 242 通过一个或多个电缆部件 244 与模块 214 进行通信。在其他实现中,控制台 102 可以包括另外的控制器子部件。所示出的实现还示出了被配置成发送和接收可以传递到模块 214 的信号的光学 I/O 接口 235。

[0048] MU 140(1) 和 140(2) 被示为可以分别连接到 MU 端口“A”130(1) 和“B”130(2)。附加 MU(例如, MU 140(3)-140(6)) 被示为可连接到控制器 104(1) 和 104(3),即每一个控制器两个 MU。控制器 104(2) 和 104(4) 也可以被配置成接纳 MU(未示出)。每一个 MU 140 都提供附加存储器,在其上面可以存储游戏、游戏参数、及其他数据。在一些实现中,其他数据可以包括数字游戏组件、可执行的游戏应用,用于扩展游戏应用的指令集、以及媒体文件中的任何一种。当被插入到控制台 102 或控制器中时, MU 140 可以被存储器控制器 202 访问。系统供电模块 250 向游戏系统 100 的组件供电。风扇 252 可冷却控制台 102 内的电路。

[0049] 包括机器指令的应用程序 260 被存储在硬盘驱动器 208 上。当控制台 102 被接通电源时,应用 260 的各个部分被加载到 RAM 206,和 / 或高速缓存 210 以及 212 中以在 CPU 200 上执行,其中应用 260 是一个这样的示例。各种应用可以存储在硬盘驱动器 208 上以用于在 CPU 200 上执行。

[0050] 通过简单地将游戏和媒体系统 100 连接到监视器 150(图 1)、电视机、视频投影仪或其他显示设备,该系统 100 就可以作为独立系统来操作。在此独立模式下,游戏和媒体系统 100 允许一个或多个玩家玩游戏或欣赏数字媒体,例如观看电影或欣赏音乐。然而,随着宽带连接的集成通过网络接口 232 而成为可能,游戏和媒体系统 100 还可以作为较大的网络游戏社区的参与者来操作。

[0051] 图 4 例示了用于实现本公开技术的操作的通用计算设备。参考图 4,用于实现本公开技术的示例性系统包括以计算机 310 的形式呈现的通用计算设备。计算机 310 的组件可以包括,但不限于,处理单元 320、系统存储器 330,以及将包括系统存储器的各种系统组件耦合到处理单元 320 的系统总线 321。系统总线 321 可以是若干类型的总线结构中的任一种,包括使用各种总线体系结构中的任一种的存储器总线或存储器控制器、外围总线,以及局部总线。作为示例而非限制,这样的体系结构包括工业标准体系结构 (ISA) 总线、微通道体系结构 (MCA) 总线、增强型 ISA (EISA) 总线、视频电子标准协会 (VESA) 局部总线,以及也称为夹层总线的外围部件互连 (PCI) 总线。

[0052] 计算机 310 通常包括各种计算机可读介质。计算机可读介质可以是能由计算机 310 访问的任何可用介质,而且包含易失性和非易失性介质、可移动和不可移动介质。作为示例而非限制,计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其他数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括,但不限于, RAM、ROM、EEPROM、闪存或其他存储器技术, CD-ROM、数字多功能盘 (DVD) 或其他光盘存储设

备,磁带盒、磁带、磁盘存储设备或其他磁存储设备,或者能用于存储所需信息且可以由计算机 310 访问的任何其他介质。通信介质通常以诸如载波或其他传输机制等已调制数据信号来体现计算机可读指令、数据结构、程序模块或其他数据,并包括任意信息传送介质。术语“已调制数据信号”指的是一个或多个特征以在信号中编码信息的方式被设定或更改的信号。作为示例而非限制,通信介质包括如有线网络或直接线连接之类的有线介质,以及如声学、RF、红外及其他无线介质之类的无线介质。上述中任一组合也应包括在计算机可读介质的范围之内。

[0053] 系统存储器 330 包括易失性和 / 或非易失性存储器形式的计算机存储介质,如只读存储器 (ROM) 331 和随机存取存储器 (RAM) 332。基本输入 / 输出系统 333 (BIOS) 包括如在启动时帮助在计算机 310 内的元件之间传输信息的基本例程,它通常储存在 ROM 331 中。RAM 332 通常包含处理单元 320 可以立即访问和 / 或目前正在操作的数据和 / 或程序模块。作为示例而非限制,图 4 示出了操作系统 334、应用程序 335、其他程序模块 336,以及程序数据 337。

[0054] 计算机 310 也可以包括其他可移动 / 不可移动、易失性 / 非易失性计算机存储介质。仅作为示例,图 4 示出了对不可移动、非易失性磁介质进行读写的硬盘驱动器 340,对可移动、非易失性磁盘 352 进行读写的磁盘驱动器 351,以及对例如 CD ROM 或其它光学介质等可移动、非易失性光盘 356 进行读写的光盘驱动器 355。可以在示例性操作环境中使用的其他可移动 / 不可移动、易失性 / 非易失性计算机存储介质包括但不限于,磁带盒、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等等。硬盘驱动器 341 通常由诸如接口 340 的不可移动存储器接口连接至系统总线 321,并且磁盘驱动器 351 和光盘驱动器 355 通常由诸如接口 350 的可移动存储器接口连接至系统总线 321。

[0055] 上文所讨论的并且在图 4 中示出的驱动器及其相关联的计算机存储介质为计算机 310 提供了对计算机可读的指令、数据结构、程序模块及其他数据的存储。例如,图 4 中,硬盘驱动器 341 被示为存储操作系统 344、应用程序 345、其它程序模块 346 和程序数据 347。注意,这些组件可以与操作系统 334、应用程序 335、其他程序模块 336 和程序数据 337 相同,也可以与它们不同。在此操作系统 344、应用程序 345、其他程序模块 346 以及程序数据 347 被给予了不同的编号,以说明至少它们是不同的副本。用户可以通过诸如键盘 362 和定点设备 361 (通常被称为鼠标、跟踪球或触摸垫) 之类的输入设备向计算机 20 输入命令和信息。其他输入设备 (未示出) 可以包括话筒、操纵杆、游戏手柄、圆盘式卫星天线、扫描仪等等。这些和其他输入设备通常由耦合至系统总线的用户输入接口 360 连接至处理单元 320,但也可以由诸如并行端口、游戏端口或通用串行总线 (USB) 等其他接口和总线结构来进行连接。监视器 391 或其他类型的显示设备也通过诸如视频接口 390 的接口连接至系统总线 321。除监视器之外,计算机也可以包括诸如扬声器 397 和打印机 396 之类的其他外围输出设备,它们可以通过输出外围接口 390 来连接。

[0056] 计算机 310 可使用至诸如远程计算机 380 之类的一个或多个远程计算机的逻辑连接在网络化环境中操作。远程计算机 380 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其他公共网络节点,并且通常包括上文参考计算机 310 所描述的许多或全部元件,虽然图 4 中只示出了存储器设备 381。图 4 中所描述的逻辑连接包括局域网 (LAN) 371 和广域网 (WAN) 373,但是,也可以包括其他网络。这样的联网环境在办公室、企业范围计算

机网络、内联网和因特网中是常见的。

[0057] 当在 LAN 联网环境中使用时,计算机 310 通过网络接口或适配器 370 连接至 LAN 371。当在 WAN 联网环境中使用时,计算机 310 通常包括调制解调器 372 或用于通过诸如因特网等 WAN 373 建立通信的其他手段。调制解调器 372 可以是内置或外置的,它可以经由用户输入接口 360 或其他适当的机制连接至系统总线 321。在网络化环境中,相对于计算机 310 所描述的程序模块或其部分可被存储在远程存储器存储设备中。作为示例而非限制,图 4 示出了驻留在存储器设备 381 上的远程应用程序 385。可以理解,所示的网络连接是示例性的,且可以使用在计算机之间建立通信链路的其他手段。

[0058] 可以使用图 1-4 中的硬件设备来实现一系统,该系统允许用户与除用户的屏幕上表示以外的对象进行交互。图 5 是描述了用于执行本公开技术的各操作的过程的一个实施例的流程图。在一个实施例中,图 5 中的步骤可以由在系统 10 的计算设备 12 中的变换模型 196、显示模块 198 和 / 或应用控制逻辑 200 中的软件模块来执行。在步骤 500 中,先接收计算设备 12 上的用户的身份。在一个示例中,步骤 500 可使用面部识别来将来自接收到的视觉图像的用户的脸与参考视觉图像进行相关,以确定用户的身份。在另一示例中,确定用户身份可包括从用户处接收标识他们身份的输入。例如,可由计算机设备 12 来存储用户简档,并且用户可在屏幕上作出选择以将他们自己标识为对应于该用户简档。在成功进行了用户标识以后,在步骤 501 接收用户对诸如应用 202 等应用的选择。在步骤 501 中,可以通过用户的处理设备 12 中的用户界面来提示用户选择应用 202。如在上文中讨论的,应用 202 可以包括例如在计算设备 12 上执行的视频游戏应用、电影应用等。在步骤 502 中,向用户显示应用 202 的场景。

[0059] 在步骤 506 中,作出确定是否检测到了用户运动的检查。如果确定检测到了用户运动,则在步骤 512 中接收用户运动数据。在一个实施例中,捕捉设备 20 中的处理器 42 可以检测并接收与用户的运动有关的信息。在图 6 中讨论了由捕捉设备 20 来检测并捕捉用户运动数据的过程。如果没有检测到用户运动数据,则在步骤 508 中作出确定用户是否希望与应用 202 的附加场景进行交互的检查。如果确定用户希望与一个或多个附加场景进行交互,则在步骤 504 中获取应用 202 的下一个场景,并在步骤 502 中向用户显示该场景。如果确定用户不希望与应用 202 的任何附加场景进行交互,则在步骤 510 应用 202 退出。

[0060] 当在步骤 512 捕捉了用户运动以后,在步骤 514 作出确定用户运动数据是否匹配一姿势的检查。用户可以作出一定范围的运动,以便与应用中描绘的一个或多个屏幕上对象进行交互。在一个实施例中,步骤 514 可以包括将用户运动数据与由姿势识别引擎 190 指定的一个或多个预定义姿势进行比较。确定用户运动数据是否匹配一姿势的过程可以由姿势识别引擎 190 来执行,并且会在图 7 中进行讨论。如果用户运动数据匹配一姿势,则在步骤 518 中将该用户运动数据自动转换成屏幕上对象的多个对象响应。在一个实施例中,转换用户运动数据的步骤 518 可以如下执行。在步骤 519 中,利用在步骤 514 中获取的被匹配姿势来访问如图 2 所述的用于将姿势与对象响应进行相关的数据结构。在步骤 520 中,从数据结构处访问屏幕上对象的相应响应,其包括访问指向实现该响应的代码的指针。例如,屏幕上对象的相应响应可包括该屏幕上对象的运动响应、视觉响应或音频响应。在步骤 521 中,访问实现对象响应的代码。在一个实施例中,可如图 2 所述的通过将用户的骨架模型表示映射成屏幕上对象的对象模型表示来实现对象响应。

[0061] 如果用户运动数据不匹配姿势,则在步骤 516 中经由计算设备 12 的用户界面来向用户提供与可被用户用来与屏幕上对象进行交互的不同类型的运动有关的反馈。例如,可向用户提供具有文本的向导,该向导与用户可用来与各种屏幕上对象进行交互的不同类型的运动有关。如果在步骤 506 检测到了用户运动数据,则在步骤 512 中随后基于向用户提供的反馈来捕捉用户运动数据。

[0062] 在步骤 522,经由用户的计算设备 12 中的用户界面来将屏幕上对象的多个对象响应显示给用户。

[0063] 图 6-9 是提供了图 5B 中的各个步骤的更多细节的流程图。图 6 是描述了用于捕捉用户运动数据的过程的一个实施例(图 5B 中的步骤 512)的流程图。在步骤 526,捕捉设备 20 的处理器 42 从图像捕捉组件 32 处接收视觉图像和深度图像。在其他示例中,在步骤 526 仅接收深度图像。深度图像和视觉图像可由图像捕捉组件 32 中的任何传感器或本领域中已知的其他适合的传感器来捕捉。在一个实施例中,深度图像与视觉图像是分开被捕捉的。在某些实现中,深度图像和视觉图像是同时被捕捉的,而在其他实现中,它们是顺序地或在不同时刻被捕捉的。在其他实施例中,深度图像是与视觉图像一起被捕捉的,或者与视觉图像组合成一个图像文件,使得每个像素具有 R 值、G 值、B 值和 Z 值(表示距离)。

[0064] 在步骤 528,确定对应于视觉图像和深度图像的深度信息。可分析在步骤 526 接收到的视觉图像和深度图像,以确定图像内的一个或多个目标的深度值。捕捉设备 20 可捕捉或观察可包括一个或多个目标的捕捉区域。

[0065] 在步骤 530,捕捉设备确定深度图像是否包括人类目标。在一个实施例中,可对深度图像中的每一目标进行泛色填充并将其与一图案进行比较来确定该深度图像是否包括人类目标。在一个实施例中,可确定深度图像的被捕捉场景中的每个目标的边缘。深度图像可包括被捕捉场景的二维像素区域。在 2D 像素区域中的每个像素可表示诸如例如可从相机测量到的长度或距离等深度值。可以通过将与例如深度图像的相邻或邻近的像素相关联的各种深度值进行比较来确定边缘。如果正被比较的各种深度值大于预定边容差,则这些像素可定义一条边缘。捕捉设备可将包括深度图像的所计算的深度信息组织成“Z 层”或可垂直于沿着其视线从相机延伸到观察者的 Z 轴的各层。可以基于所确定的边缘,对 Z 层的可能的 Z 值进行泛色填充。例如,可将与所确定的边缘相关联的像素和在所确定的边缘内的区域中的像素互相关联,以定义捕捉区域中的目标或对象。

[0066] 在步骤 532,捕捉设备扫描人类目标以寻找一个或多个身体部位。可以扫描人类目标来提供与用户的一个或多个身体部位相关联的诸如长度、宽度等的度量,使得可基于这些度量来生成该用户的准确模型。在一个示例中,隔离人类目标,并创建位掩模来扫描一个或多个身体部位。可例如通过对人类目标进行泛色填充来创建位掩模,使得该人类目标与捕捉区域元素中的其他目标或对象相分离。

[0067] 在步骤 534,基于在步骤 532 执行的扫描来生成人类目标的模型。可分析该位掩模来寻找一个或多个身体部位,以生成诸如人类目标的骨架模型、网格人类模型等模型。例如,可使用由所扫描的位掩模确定的度量值来定义骨架模型中的一个或多个关节。位掩模可包括人类目标沿 X、Y 和 Z 轴的值。该一个或多个关节可用于定义可对应于人类的身体部位的一根或多根骨骼。

[0068] 根据一个实施例,为了确定人类目标的颈部、肩膀等的位置,可以将例如在被扫描

的位置处的位掩模的宽度与关联于例如颈部、肩膀等的典型宽度的阈值进行比较。在一替换实施例中,可以使用与扫描的并与位掩模中的身体部位相关联的先前位置的距离来确定颈部、肩膀等的位置。

[0069] 在一个实施例中,为了确定肩膀的位置,可将肩膀位置处的位掩模的宽度与阈值肩膀值进行比较。例如,可以将肩膀位置处的位掩模的 X 值处的两个最外部 Y 值之间的距离与例如人类肩膀之间的典型距离的阈值肩膀值进行比较。由此,根据一示例实施例,该阈值肩膀值可以是与人类的身体模型中的肩膀相关联的典型宽度或宽度范围。

[0070] 在一个实施例中,诸如腿、脚等的某些身体部位可基于例如其他身体部位的位置来计算。例如,如上所述,可以扫描与人类目标相关联的诸如位、像素等信息,以确定人类目标的各个身体部位的位置。基于这些位置,随后可以为人类目标计算诸如腿、脚等的后续身体部位。

[0071] 根据一个实施例,在确定了例如某身体部位的值后,可以创建一数据结构,该数据结构可包括与人类目标的位掩模的扫描相关联的身体部位的诸如长度、宽度等的度量值。在一个实施例中,该数据结构可包括对多个深度图像求平均所得的扫描结果。例如,捕捉设备可捕捉各帧中的捕捉区域,每个帧包括深度图像。可按如上所述来分析每个帧的深度图像,以确定是否包括人类目标。如果帧的深度图像包括人类目标,则可扫描与该帧相关联的深度图像的人类目标的位掩模来寻找一个或多个身体部位。然后可以对为每一帧所确定的身体部位的值求平均,从而使得该数据结构可包括与每一帧的扫描相关联的身体部位的诸如长度、宽度等的平均度量值。根据一个实施例,可调整所确定的身体部位的度量值,如放大、缩小等,使得数据结构中的度量值更接近地对应于人类身体的典型模型。在步骤 534 中,可使用由所扫描的位掩模确定的度量值来定义骨架模型中的一个或多个关节。

[0072] 在步骤 536,使用骨架映射来跟踪在步骤 534 中创建的模型。例如,可在用户在视野内在相机前的物理空间中移动时调整和更新用户 18 的骨架模型。来自捕捉设备的信息可用于调整模型,使得骨架模型准确地表示用户。在一个示例中,这是通过向该骨架模型的一个或多个受力方面施加一个或多个力,以将该骨架模型调整成更接近地对应于人类目标和物理空间的姿态的姿态来实现的。

[0073] 在步骤 538 中,基于骨架映射来捕捉运动,以生成运动捕捉文件。在一个实施例中,捕捉运动的步骤 538 可包括计算由所述扫描标识的一个或多个身体部位的位置、方向、加速度和曲率。在 X、Y、Z 空间中计算身体部位的位置,以创建该身体部位在相机的视野内的三维位置表示。根据该位置来计算身体部位移动的方向。方向性移动可具有在 X、Y 和 Z 方向的任何一个或其组合上的分量。确定该身体部位在 X、Y、Z 空间中的移动的曲率,例如以表示身体部位在捕捉区域内的非线性移动。速度、加速度和曲率计算并不取决于方向。可以理解,对 X、Y、Z 笛卡尔映射的使用仅仅是作为示例来提供的。在其他实施例中,可使用不同的坐标映射系统来计算移动、速度和加速度。当检查身体部位自然地绕关节旋转的移动时,例如球面坐标映射可能是有用的。

[0074] 一旦分析了在扫描中的所有身体部位以后,就可以为目标更新在步骤 538 生成的运动捕捉文件。在一个示例中,运动捕捉文件是基于与所跟踪的模型相关联的信息来实时生成的。例如,在一个实施例中,运动捕捉文件可包括含 X、Y 和 Z 值的矢量,这些矢量在模型于各个时间点被跟踪时用于定义该模型的关节和骨骼。如上所述,被跟踪的模型可基于

各个时间点的用户运动来进行调整,并且可生成并存储运动的模型的运动捕捉文件。在与目标识别、分析和跟踪系统交互的用户进行自然移动期间,该运动捕捉文件可捕捉所跟踪的模型。例如,可生成运动捕捉文件,使得该运动捕捉文件可自然地捕捉用户在与目标识别、分析和跟踪系统交互期间进行的任何移动或运动。该运动捕捉文件可包括与例如用户在不同时间点的运动的快照相对应的帧。在捕捉了所跟踪的模型之后,可在运动捕捉文件的一帧中呈现与模型相关联的信息,该信息包括在一特定时间点应用于该模型的任何移动或调整。该帧中的信息可包括例如含 X、Y 和 Z 值的矢量以及一时间戳,这些矢量定义所跟踪的模型的关节和骨骼,该时间戳可指示例如用户执行了对于所跟踪的模型的姿态的移动的时间点。

[0075] 在一个实施例中,步骤 526-538 是由计算设备 12 执行的。此外,尽管步骤 526-538 被描述为是由捕捉设备 20 来执行的,但这些步骤中的各个步骤均可由诸如计算环境 12 之类的其他组件来执行。例如,捕捉设备 20 可向计算设备 12 提供视觉和 / 或深度图像,计算设备 12 接着将确定深度信息、检测人类目标、扫描目标、生成并跟踪模型、以及捕捉人类目标的运动。

[0076] 图 7 示出了可在图 6 的步骤 534 中生成的、表示所扫描的人类目标的骨架模型或映射 540 的示例。根据一个实施例,骨架模型 540 可包括可将人类目标表示为三维模型的一个或多个数据结构。每个身体部位可被表征为定义骨架模型 540 的关节和骨骼的数学矢量。

[0077] 骨架模型 540 包括关节 n1-n18。关节 n1-n18 中的每一个可使得在这些关节之间定义的一个或多个身体部位能相对于一个或多个其他身体部位移动。表示人类目标的模型可包括多个刚性和 / 或可变形身体部位,这些身体部位可由诸如“骨骼”等的一个或多个结构件来定义,而关节 n1-n18 位于相邻骨骼的交叉点处。关节 n1-n18 可使得与骨骼和关节 n1-n18 相关联的各个身体部位能够彼此独立地或彼此相对地移动。例如,在关节 n7 与 n11 之间定义的骨骼对应于前臂,该前臂可独立于例如在关节 n15 与 n17 之间定义的对应用于小腿的骨骼而移动。可以理解,某些骨骼可对应于人类目标中的解剖学骨骼,和 / 或某些骨骼在人类目标中可能不具有对应的解剖学骨骼。

[0078] 骨骼和关节可共同构成骨架模型,它们可以是该模型的构成元素。轴向滚动角可用于定义肢相对于其父肢和 / 或躯干的旋转定向。例如,如果骨架模型正示出手臂的轴向旋转,则滚动关节可用来指示相关联的腕所指的方向(例如,手掌向上)。通过检查肢相对于其父肢和 / 或躯干的定向,可确定轴向滚动角。例如,如果正在检查小腿,则可检查小腿相对于相关联的大腿和髌部的定向以便确定轴向滚动角。

[0079] 图 8 提供了如图 2 所示的姿势识别引擎 190 的示例性实施例的进一步细节。如所示出的,姿势识别引擎 190 可包括确定一个或多个姿势的至少一个过滤器 450。过滤器 450 包括定义姿势 452(以下称为“姿势”)以及该姿势的元数据 454 的参数。过滤器可包括可识别姿势或以其他方式处理深度、RGB 或骨架数据的代码和相关联的数据。例如,包括一只手从身体背后越过身体前方的运动的投掷可被实现为包括表示用户的一只手从身体背后越过身体前方的移动的信息的姿势 452,该移动将由深度相机来捕捉。然后可为该姿势 452 设定参数 454。在姿势 452 是投掷的情况下,参数 454 可以是该手必须达到的阈值速度、该手必须行进的距离(或者是绝对的,或者是作为整体相对于用户的大小)、以及由识别引擎

来作出评级的姿势发生的置信度。姿势 452 的这些参数 454 可以在各应用之间、在单个应用的各上下文之间、或在一个应用的一个上下文内随着时间而变化。姿势参数可包括阈值角度（例如髋部 - 大腿角度、前臂 - 二头肌角度等）、运动发生或不发生的周期数、阈值周期、阈值位置（开始、结束）、移动方向、速度、加速度、移动的坐标等。

[0080] 过滤器可包括可识别姿势或以其他方式处理深度、RGB 或骨架数据的代码和相关联的数据。过滤器可以是模块化的或是可互换的。在一个实施例中，过滤器具有多个输入和多个输出，这些输入中的每一个具有一类型，这些输出中的每一个具有一类型。在这一情形中，第一过滤器可用与该第一过滤器具有相同数量和类型的输入和输出的第二过滤器来替换，而无需更改姿势识别引擎体系结构的其他方面。例如，可能具有要驱动的第一过滤器，该第一过滤器将骨架数据作为输入，并输出与该过滤器相关联的姿势正在发生的置信度和转向角。在希望用第二驱动过滤器来替换该第一驱动过滤器的情况下（这可能是因为在第二驱动过滤器更高效且需要更少的处理资源），可以通过简单地用第二过滤器替换第一过滤器来这样做，只要第二过滤器具有同样的输入和输出——骨架数据类型的一个输入、以及置信度类型和角度类型的两个输出。

[0081] 过滤器不需要具有参数。例如，返回用户的高度的“用户高度”过滤器可能不允许可被调节的任何参数。替换的“用户高度”过滤器可具有可调节参数，如在确定用户的高度时是否考虑用户的鞋、发型、头饰以及体态。

[0082] 过滤器的输入可包括诸如关于用户的关节位置的关节数据、如在关节处相交的骨骼所形成的角度、来自捕捉区域的 RGB 色彩数据、以及用户的一方面的变化率等内容。来自过滤器的输出可包括诸如正作出给定姿势的置信度、作出姿势运动的速度、以及作出姿势运动的时间等内容。

[0083] 姿势识别引擎 190 可具有向姿势过滤器 450 提供功能的基础识别引擎 456。在一个实施例中，基础识别引擎 456 实现的功能包括跟踪所识别的姿势和其他输入的随时间输入 (input-over-time) 存档、隐马尔可夫模型实现（其中模型化系统被假定为具有未知参数的马尔可夫过程（当前状态封装了确定将来状态所需的任何过去状态信息，因此不必为此目的而维护任何其它过去状态信息的过程），并且隐藏参数从可观察数据中确定）、以及解决姿势识别的特定实例所需的其他功能。

[0084] 过滤器 450 在基础识别引擎 456 之上加载并实现，并且可利用引擎 456 提供给所有过滤器 450 的服务。在一实施例中，基础识别引擎 456 处理所接收到的数据，以便确定它是否满足任何过滤器 450 的要求。由于这些诸如对输入进行解析等所提供的服务是由基础识别引擎 456 一次性提供而不是由每一过滤器 450 来提供的，因此这种服务在一段时间内只需被处理一次而不是在该时间段期间由每一过滤器 450 处理一次，由此减少了确定姿势所需的处理。

[0085] 应用程序可使用由识别引擎 190 提供的过滤器 450，或者该应用程序可提供它自己的过滤器 450，该过滤器被插入到基础识别引擎 456 中。在一实施例中，所有过滤器 450 具有启用该插入特性的公共接口。此外，所有过滤器 450 可利用参数 454，因此可使用如下所述的单个姿势工具来诊断并调节整个过滤器系统。这些参数 454 可由姿势工具为应用或应用的上下文来调节。

[0086] 有各种可以与姿势相关联的输出。在一个示例中，可以有于姿势是否正在发生

的基线“是或否”。在另一个示例中,还可以有置信水平,其对应于用户的被跟踪的移动对应于姿势的可能性。这可以是范围为 0 和 1 之间(包括端点)的浮点数的线性标度。在接收该姿势信息的应用不能接受假肯定作为输入的情况下,它可仅使用具有如至少 0.95 的高置信水平的那些所识别的姿势。在应用甚至以假肯定为代价而必须识别姿势的每一实例的情况下,它可使用至少具有低得多的置信水平的姿势,如仅仅大于 .2 的那些姿势。姿势可具有在两个最近步之间的时间的输出,并且在仅注册了第一步的情况下,这可被设为保留值,如 -1(因为任何两步之间的时间必须为正)。姿势也可具有关于在最近一步期间达到的最高大腿角的输出。

[0087] 姿势或其一部分可将它必须在其中发生的空间体作为参数。在姿势包括身体移动的情况下,该空间体通常可相对于身体来表达。例如,对于惯用右手的用户的美式足球投掷姿势可仅在不低于右肩 410a、且与投掷手臂 402a-410a 在头 422 的相同侧的空间体中识别。可能不必要定义空间体的所有边界,如对于该投掷姿势,其中从身体向外的边界留着不被定义,并且该空间体无限地向外延伸,或者延伸到正被监视的捕捉区域的边缘。

[0088] 另外,姿势可堆叠在彼此之上。即,用户一次可表达多于一个的姿势。例如,并非在作出投掷姿势时不允许除了投掷之外的任何输入,也不要求用户除了该姿势的分量之外保持不动(例如,在作出仅涉及一条手臂的投掷姿势时站着不动)。在姿势堆叠时,用户可同时作出跳跃姿势和投掷姿势,并且这两个姿势都将被姿势引擎识别出。

[0089] 图 9 是描述了根据本公开技术的用于执行图 5B 中的步骤 514 的确定用户运动数据是否匹配一姿势的过程的一个实施例的流程图。图 9 描述了一种基于规则的方法,该方法用于由姿势识别引擎 190 应用一个或多个姿势过滤器来确定用户的运动数据是否匹配一具体的姿势。可以理解,尽管在该具体示例中描述了对单个姿势的检测,但是图 9 的过程可被多次执行以检测在活动姿势集中的多个姿势。所描述的过程可针对多个活动姿势而并行或顺序地执行。

[0090] 在步骤 602,姿势识别引擎访问具体目标的骨架跟踪数据以开始确定该目标是否作出了所选的姿势。如图 6 所述,可从运动捕捉文件处访问骨架跟踪数据。在步骤 604,姿势识别引擎针对一个或多个预先确定的身体部位来过滤骨架跟踪数据,该一个或多个预先确定的身体部位与在所选的姿势过滤器中标识的所选的姿势有关。步骤 604 可包括仅访问与所选的姿势有关的数据,或访问目标的全部骨架跟踪数据并忽略或丢弃与所选的姿势不相关的信息。例如,手姿势过滤器可指示出仅人类目标的手与所选的姿势有关,从而使得与其他身体部位有关的数据可以被忽略。这种技术可通过将处理限于被预先确定为对所择的姿势是显著的信息来提高姿势识别引擎的性能。

[0091] 在步骤 606,姿势识别引擎为预先确定的轴向移动过滤骨架跟踪数据。例如,所选的姿势的过滤器可指定仅沿着轴的某一子集的移动是相关的。

[0092] 在步骤 608,姿势识别引擎访问在姿势过滤器中指定的规则 j。在图 9 的过程的第一次迭代中,j 等于 1。姿势可包括为了姿势被识别而需要被满足的多个参数。这些参数中的每个参数可在单独的规则中指定,虽然在单个规则中可包括多个分量。规则可指定目标的身体部位为了使姿势被满足而必须满足的阈值距离、位置、方向、曲率、速度和 / 或加速度以及其他参数。规则可应用于一个身体部位或多个身体部位。而且,规则可指定诸如位置之类的单个参数,或诸如位置、方向、距离、曲率、速度和加速度之类的多个参数。

[0093] 在步骤 610, 姿势识别引擎将在步骤 604 和 606 过滤的骨架跟踪数据与规则的指定参数进行比较, 以确定该规则是否被满足。例如, 姿势识别引擎可确定手的起始位置是否位于起始位置参数的阈值距离内。规则可进一步指定并且引擎确定手是否在指定方向上移动、在指定方向上从起始位置移动了阈值距离; 沿指定轴在阈值曲率内移动、以指定速度或超过指定速度移动; 达到或超过指定加速度。如果引擎确定骨架跟踪信息不满足过滤器规则中指定的参数, 则在步骤 612 中引擎返回失败或姿势过滤器未被满足的响应。向在计算设备 12 上执行的应用 202 返回该响应。

[0094] 在步骤 614, 姿势识别引擎确定姿势过滤器是否指定了对于要被完成的姿势必须被满足的附加规则。如果过滤器中包括附加规则, 则 j 递增 1 且过程返回到步骤 608, 在步骤 608 访问下一规则。如果没有附加规则, 则在步骤 618 姿势识别引擎返回姿势过滤器已被满足的指示。

[0095] 图 9 的步骤 612 和 618 对正被分析的姿势返回简单的通过 / 失败的响应。在其他示例中, 图 9 并不返回简单的通过 / 失败的响应, 而是将返回姿势过滤器被满足的置信水平。对于过滤器中的每个规则, 确定目标的移动满足或不满足指定参数的量。基于这些量的聚集, 识别引擎返回目标确实执行了姿势的置信水平。

[0096] 图 10-14 示出了描绘有与在计算设备上执行的应用所进行的各种用户界面屏幕。图 10-11 示出了经由计算设备 12 中的用户界面与应用进行的示例性用户交互, 以及与应用进行的用户交互的结果。图 10 示出了用户 18 正经由计算设备 12 中的用户界面与应用 202 进行交互。图 11 示出了与应用进行的如图 10 所示的用户交互的结果。在图 11 所示的示例性示图中, 用户 18 通过执行诸如髋部运动等示例性运动来与在应用 202 的场景 50 中描绘的鲨鱼对象 52 进行交互。示出了该用户交互的结果, 在该结果中髋部运动 54 被转换成了鲨鱼的鳍部的多个运动响应 56、57。如进一步示出的, 经由用户界面来将多个对象响应 56、57 同时显示给用户 18。

[0097] 在另一个实施例中, 一个以上的用户同时与在计算设备 12 中执行的应用 202 进行交互。因此, 可从第一用户处接收第一用户运动数据, 并可从与应用 202 中描绘的屏幕上对象进行交互的第二用户处接收第二用户运动数据。可将第一用户运动数据转换成第一运动响应, 并可从第二用户运动数据转换成第二运动响应。可经由计算设备 12 中的用户界面来将屏幕上对象的第一和第二响应同时显示给用户。在一个实施例中, 当第二用户运动数据与第一用户运动数据不相同, 第二对象响应可以与第一对象响应不相同。作为替换, 当第二用户运动数据与第一用户运动数据相同时, 第二对象响应可以是第一用户响应的放大。放大的响应可例如基于所传感的用户运动的速度或加速度来确定, 其中响应于更高速度的移动屏幕上对象可移动得更快。

[0098] 图 12 是第一用户 18 和第二用户 19 的示例性示图, 该第一用户 18 和第二用户 19 正在与在计算设备 12 上执行的应用 202 中描绘的屏幕上对象 52 进行交互。在该示例性示图中, 第一用户 18 执行髋部运动来与在应用 202 的场景 50 中描绘的鲨鱼对象 52 进行交互。第二用户 19 执行手部运动 53 来与相同的鲨鱼对象 52 进行交互。在图 12 中同时显示了与鲨鱼对象 52 进行的两个用户交互的结果, 其中髋部运动 54 被转换成了导致鲨鱼对象 52 的鳍部的多个运动 56、57 的第一运动响应, 而手部运动 53 被转换成了导致鲨鱼对象 52 的身体的运动 59 的第二运动响应。

[0099] 图 13 示出了用户 18 正经由计算设备 12 中的用户界面与应用 202 的另一个场景进行交互。在图 13 所示的示例性示图中,用户 18 与在应用 202 的具体场景 60 中描绘的诸如电灯泡对象 62 等无生命的对象进行交互。图 14 示出了与电灯泡对象 62 进行的用户交互的结果,其中用户的拍手运动 66 被转换成了电灯泡对象 62 的视觉响应。如所示出的,如由参考标记 64 所指示的该视觉响应将电灯泡 62 打开了。

[0100] 尽管用结构特征和 / 或方法动作专用的语言描述了本主题,但是可以理解,所附权利要求书中定义的主题不必限于上述具体特征或动作。更确切而言,上述具体特征和动作是作为实现权利要求的示例形式公开的。本发明的范围由所附的权利要求进行定义。

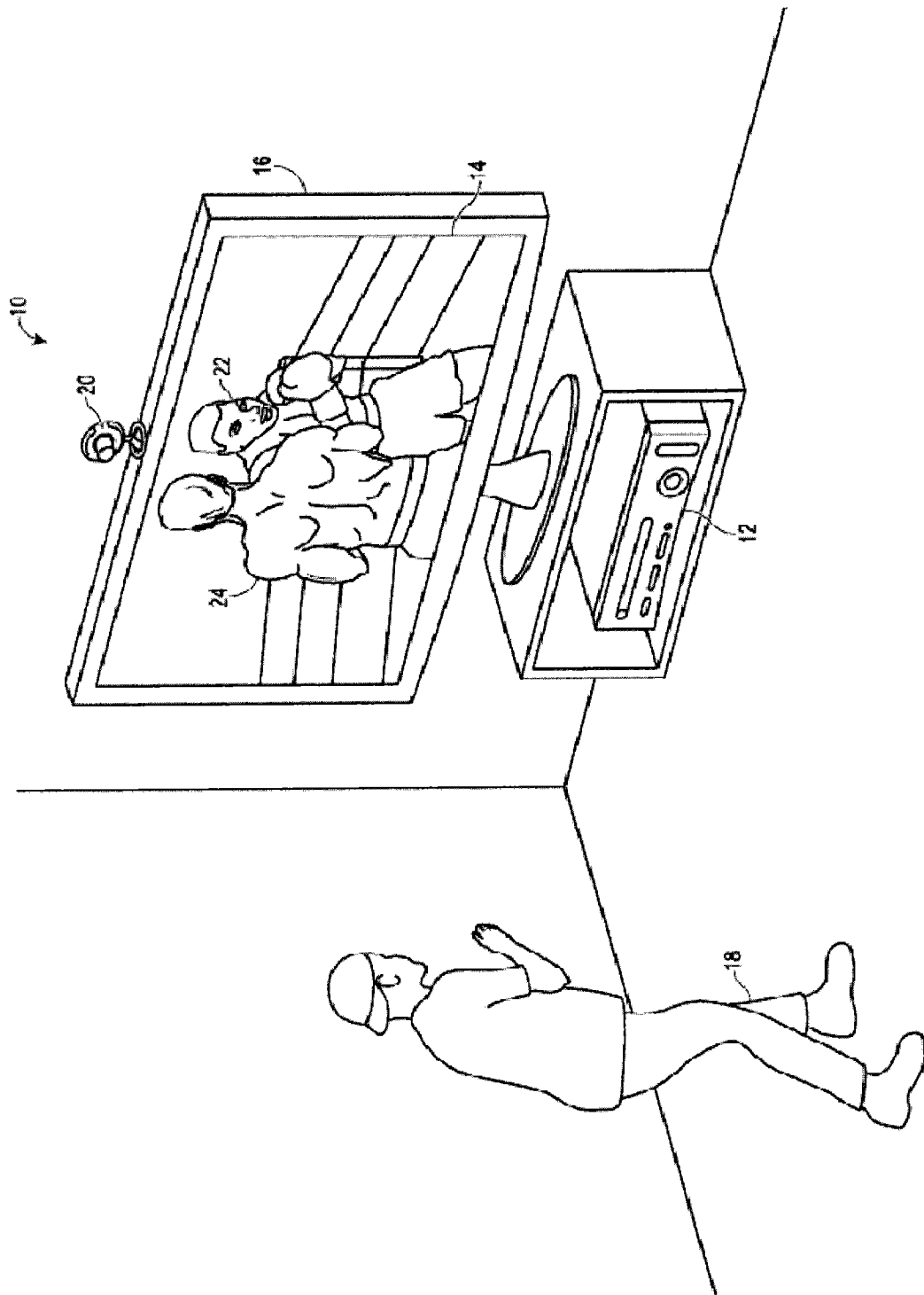


图 1

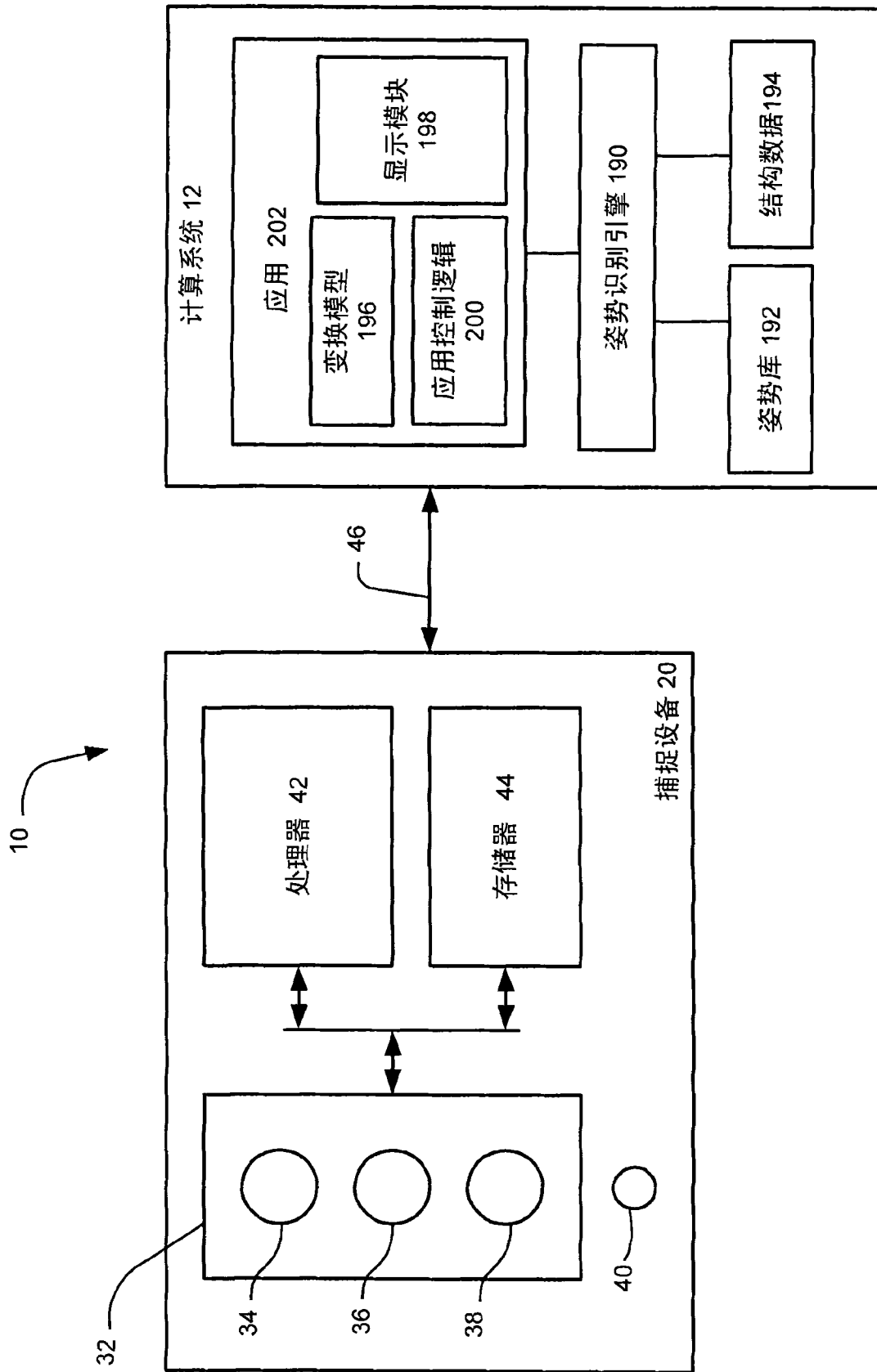


图 2

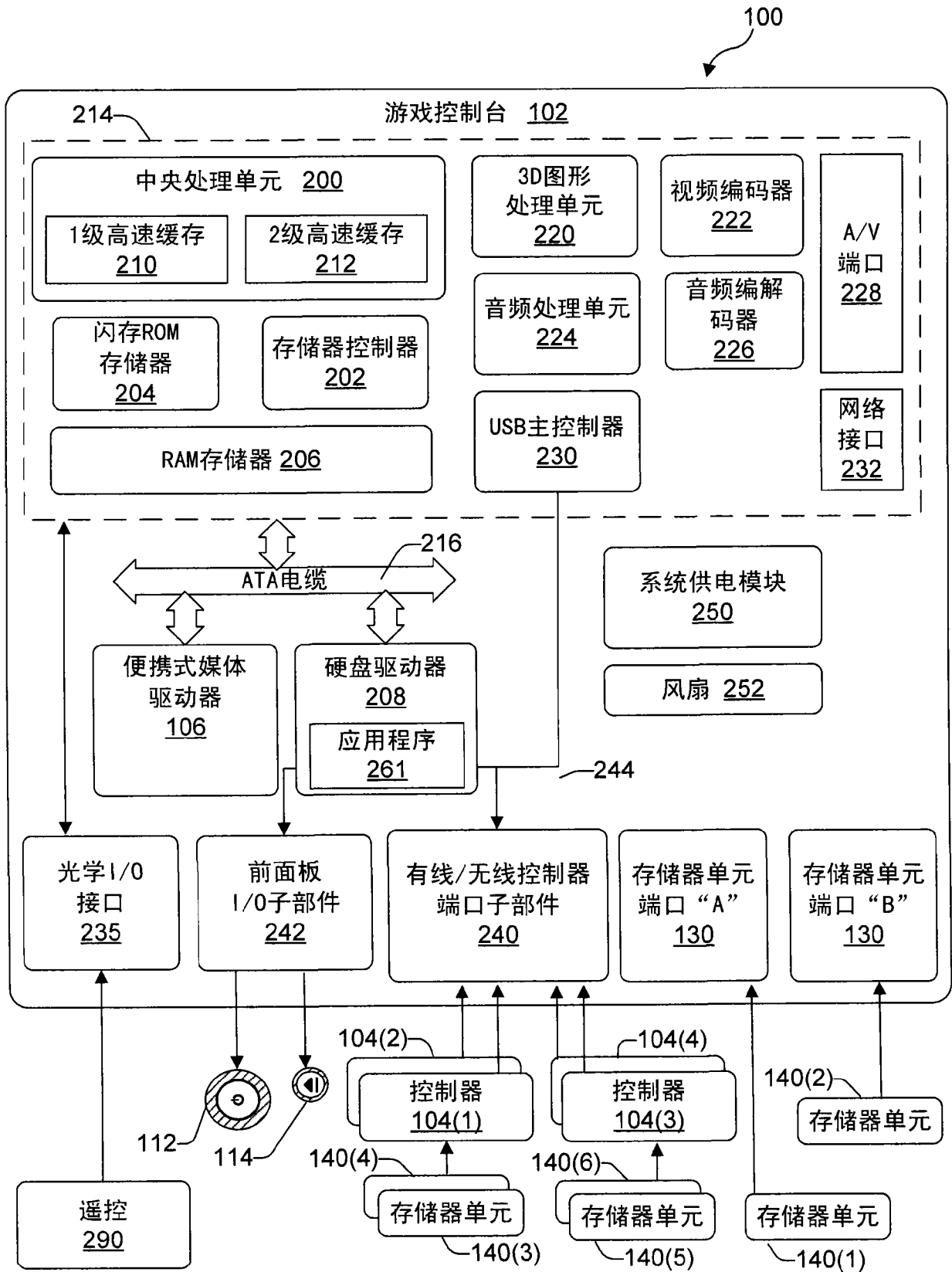


图 3

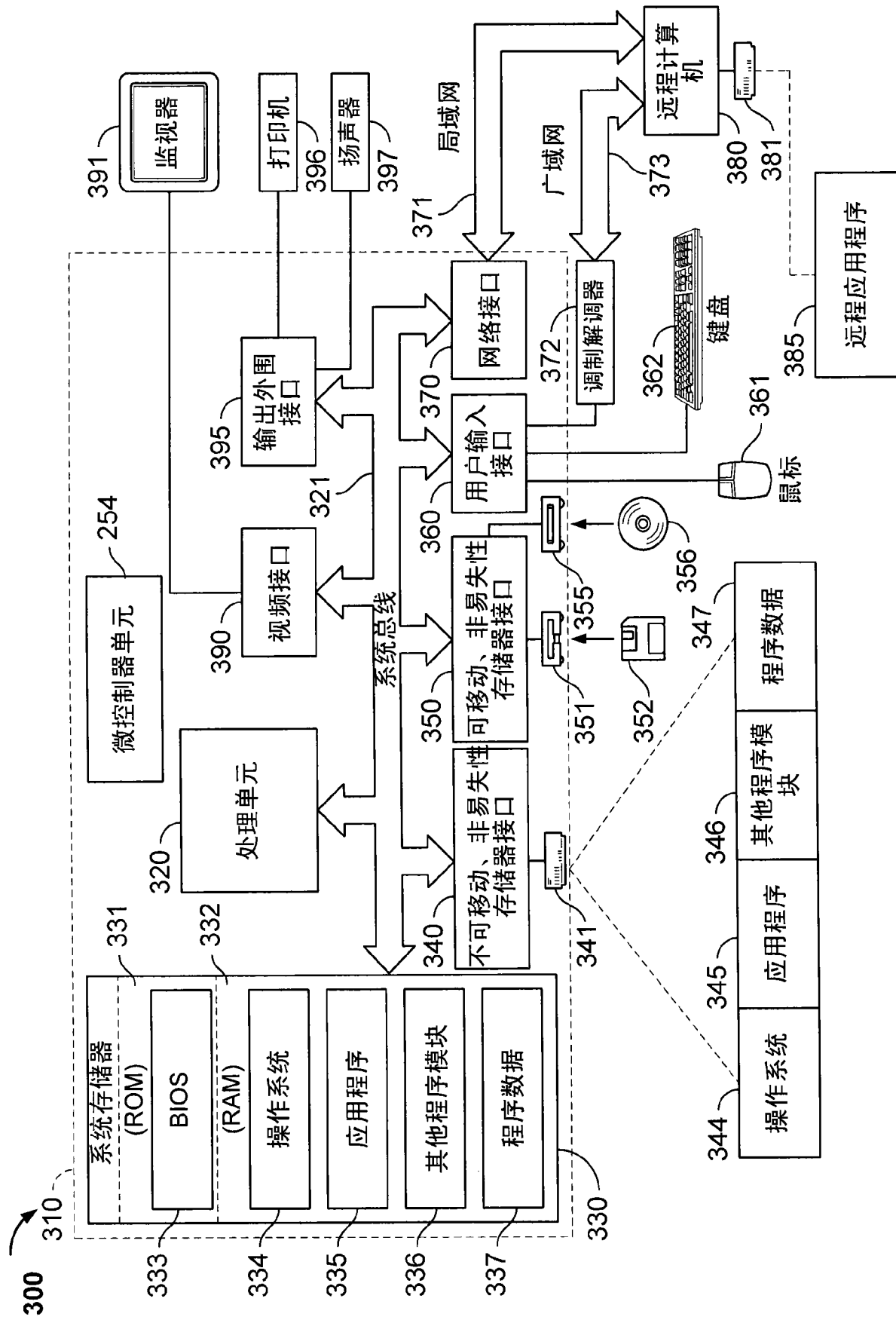


图 4

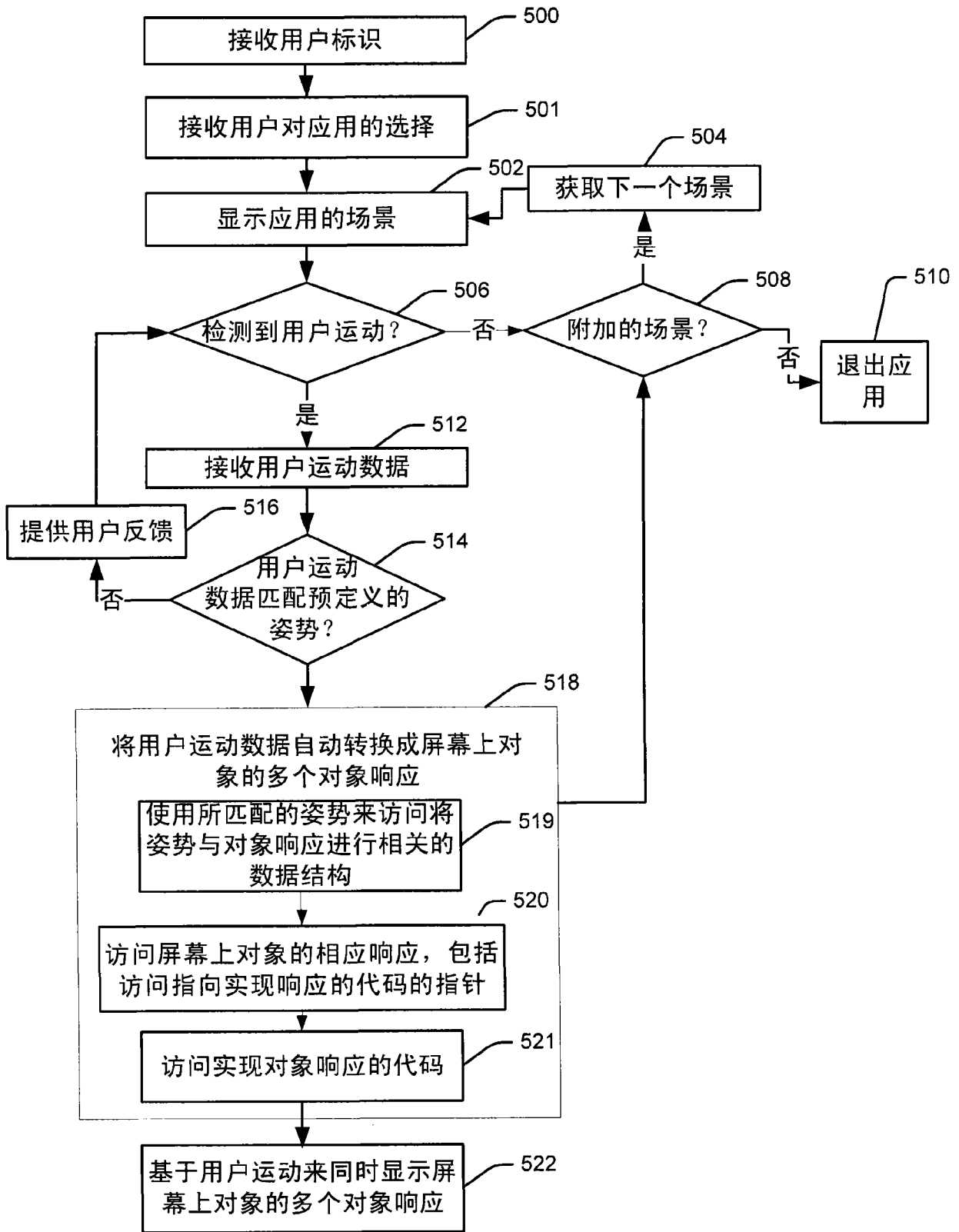


图 5

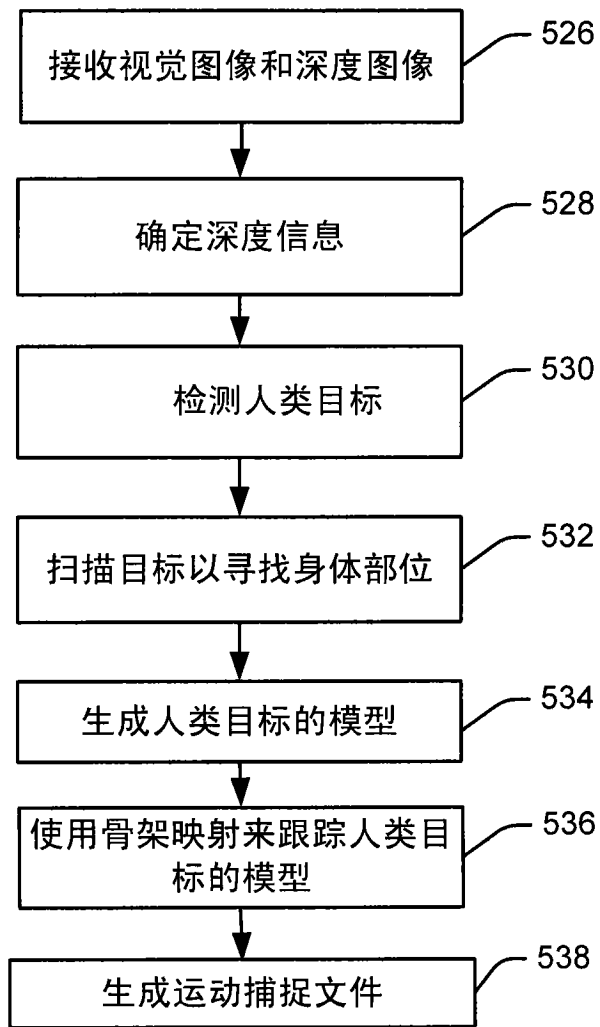


图 6

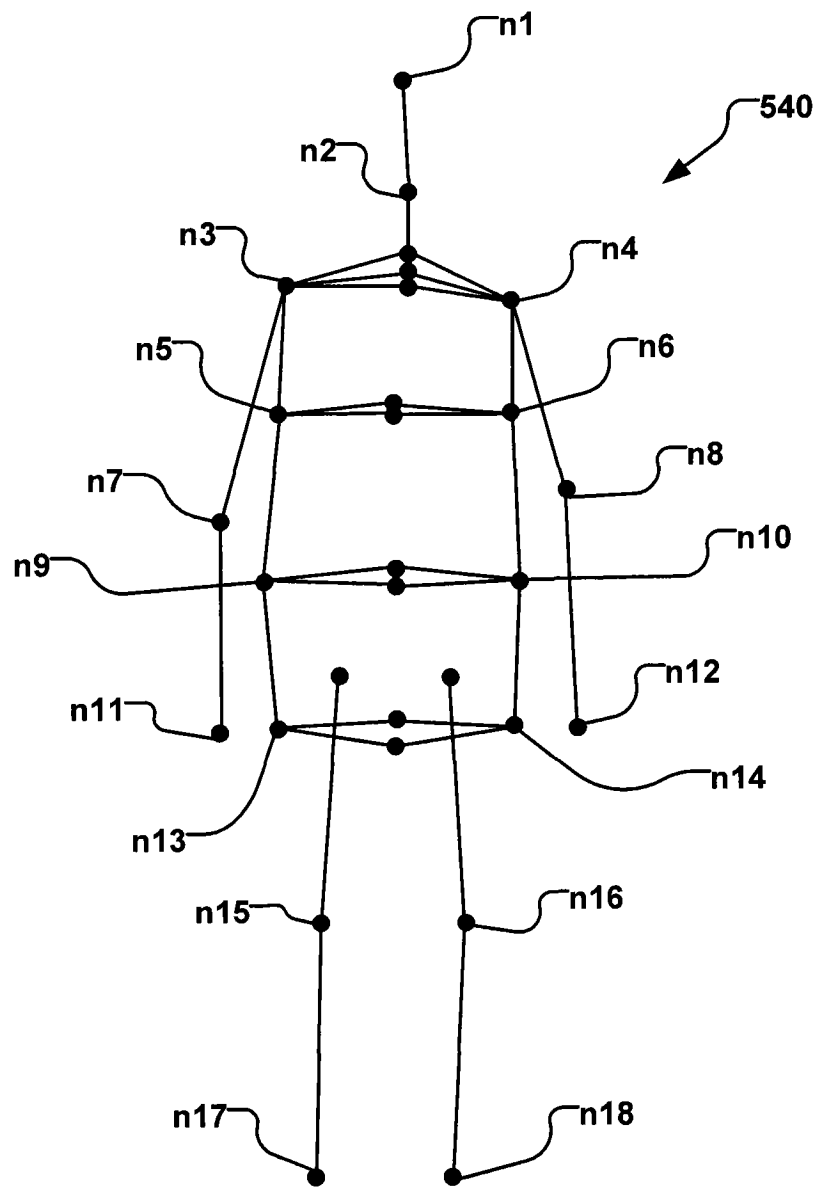


图 7

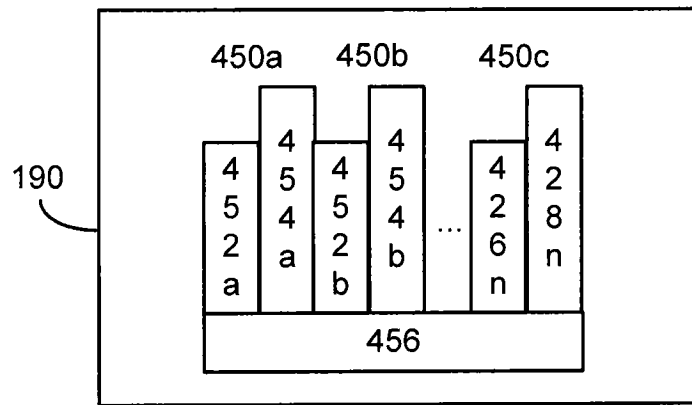


图 8

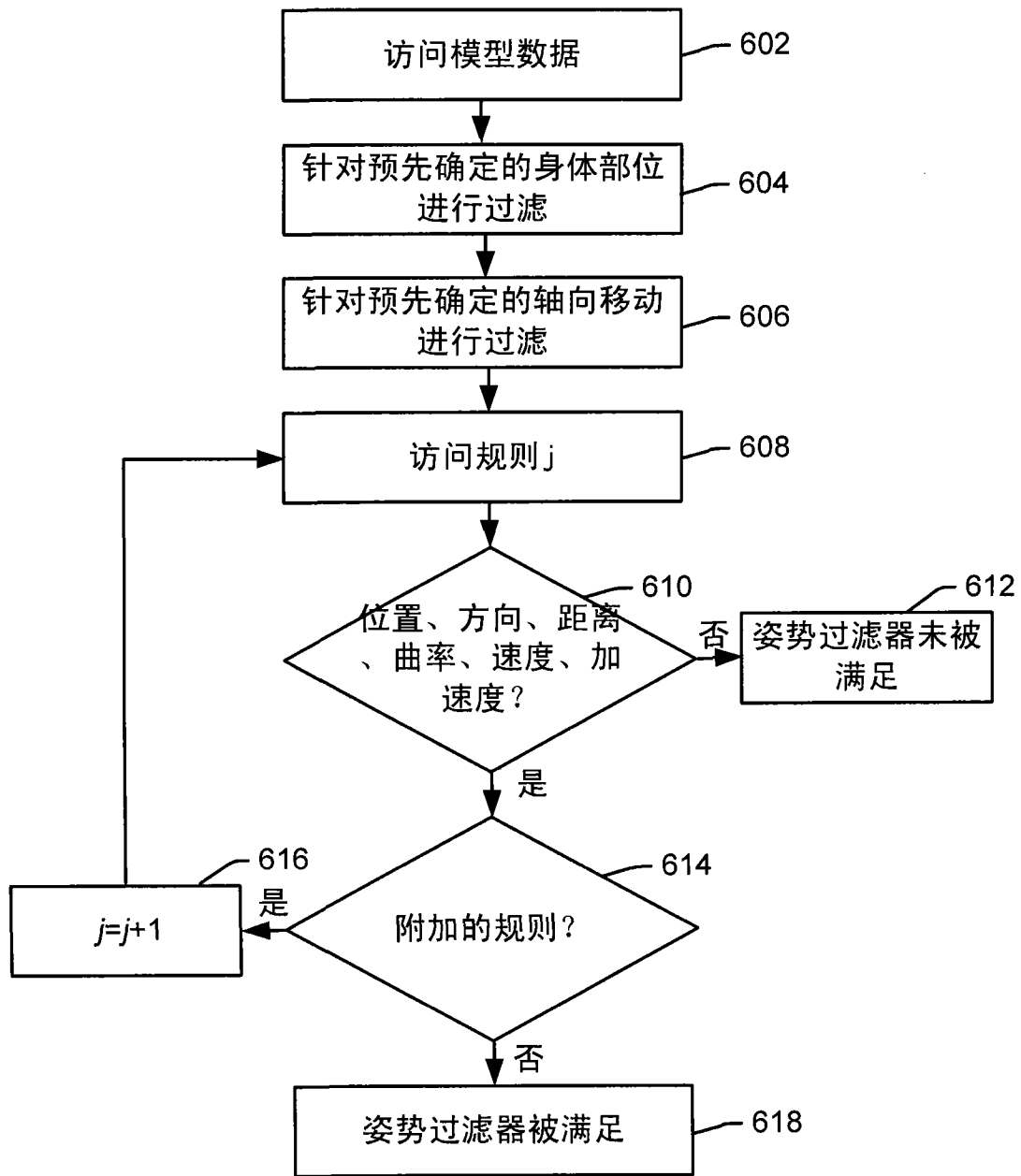


图 9

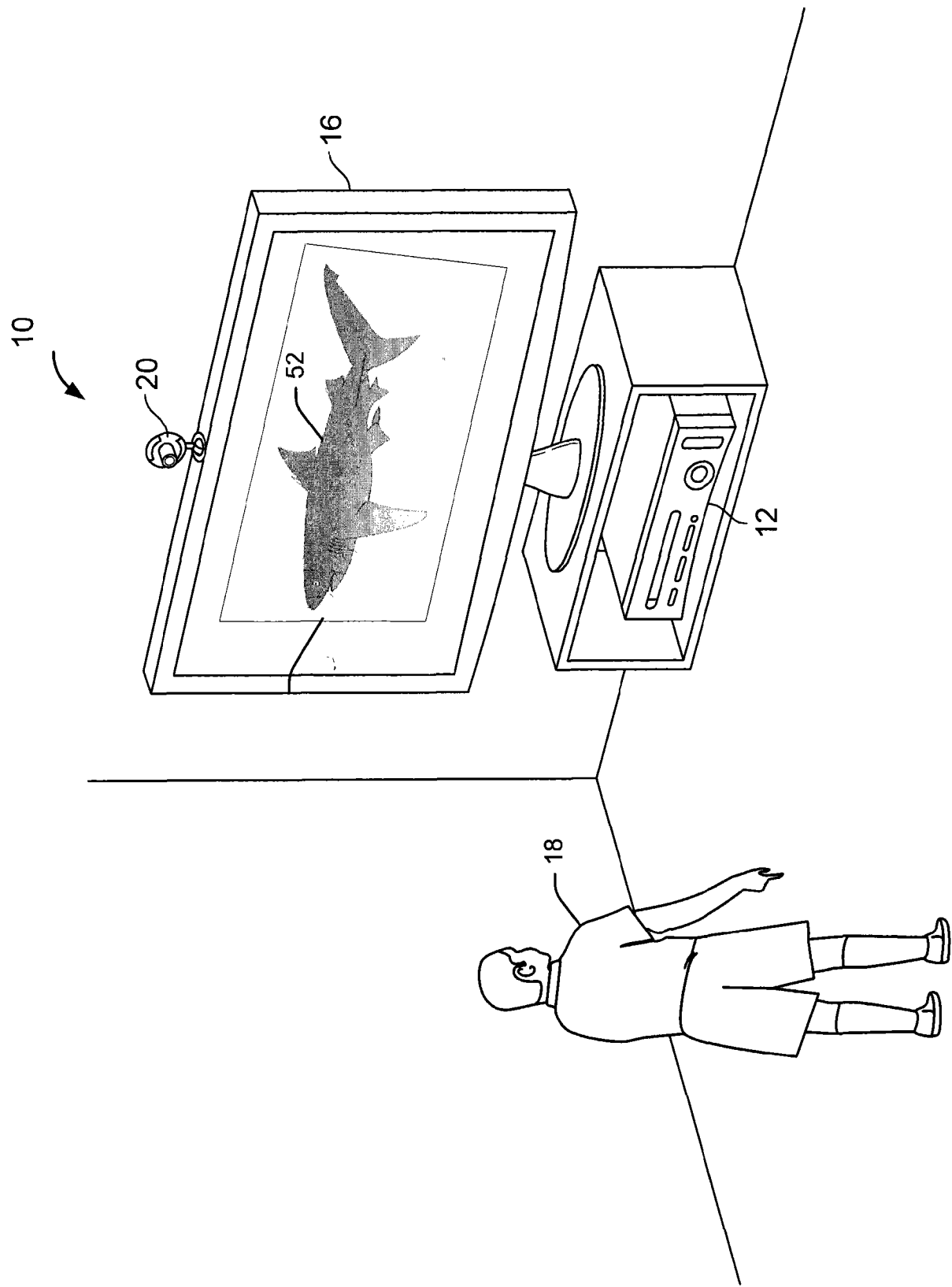


图 10

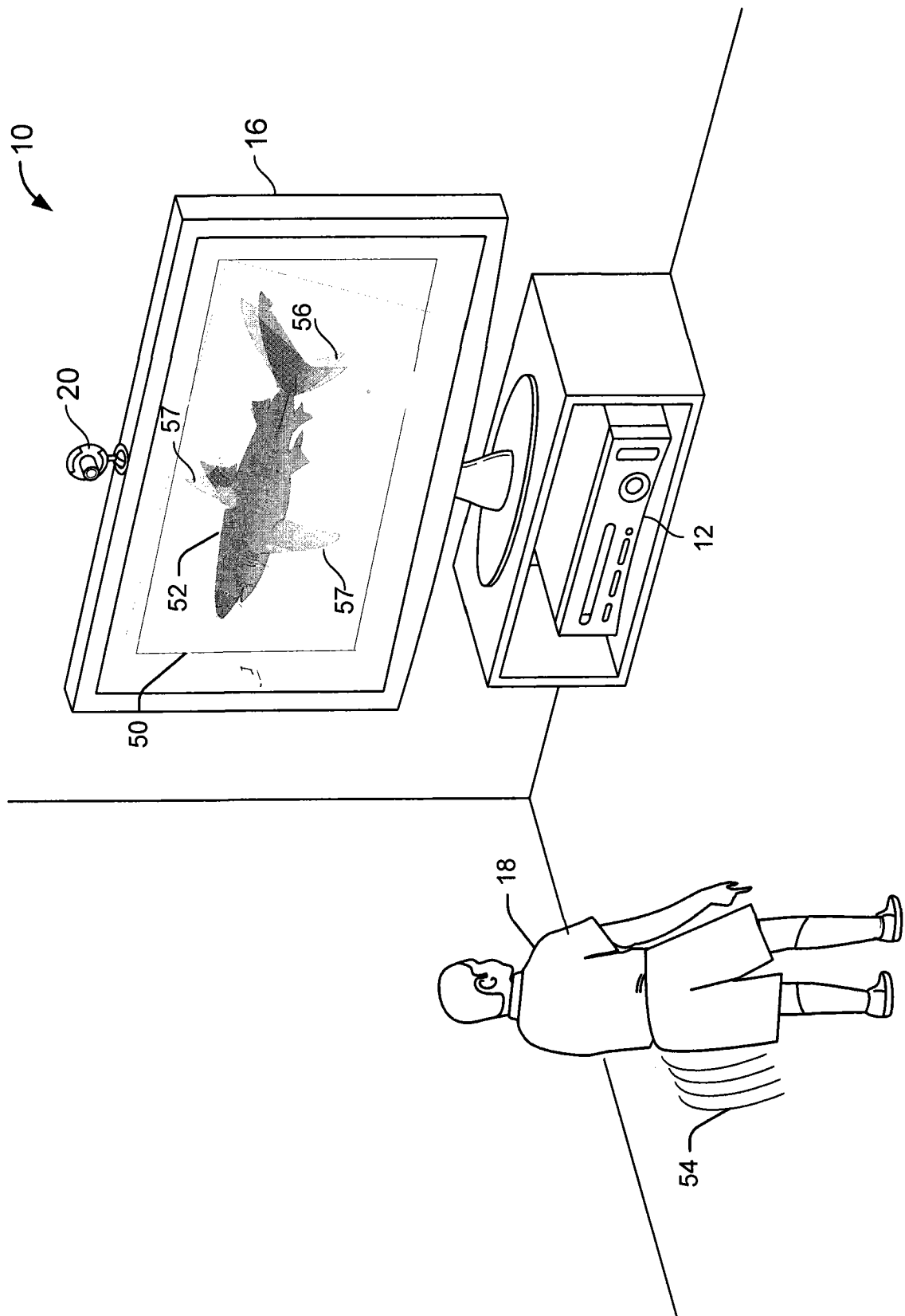


图 11

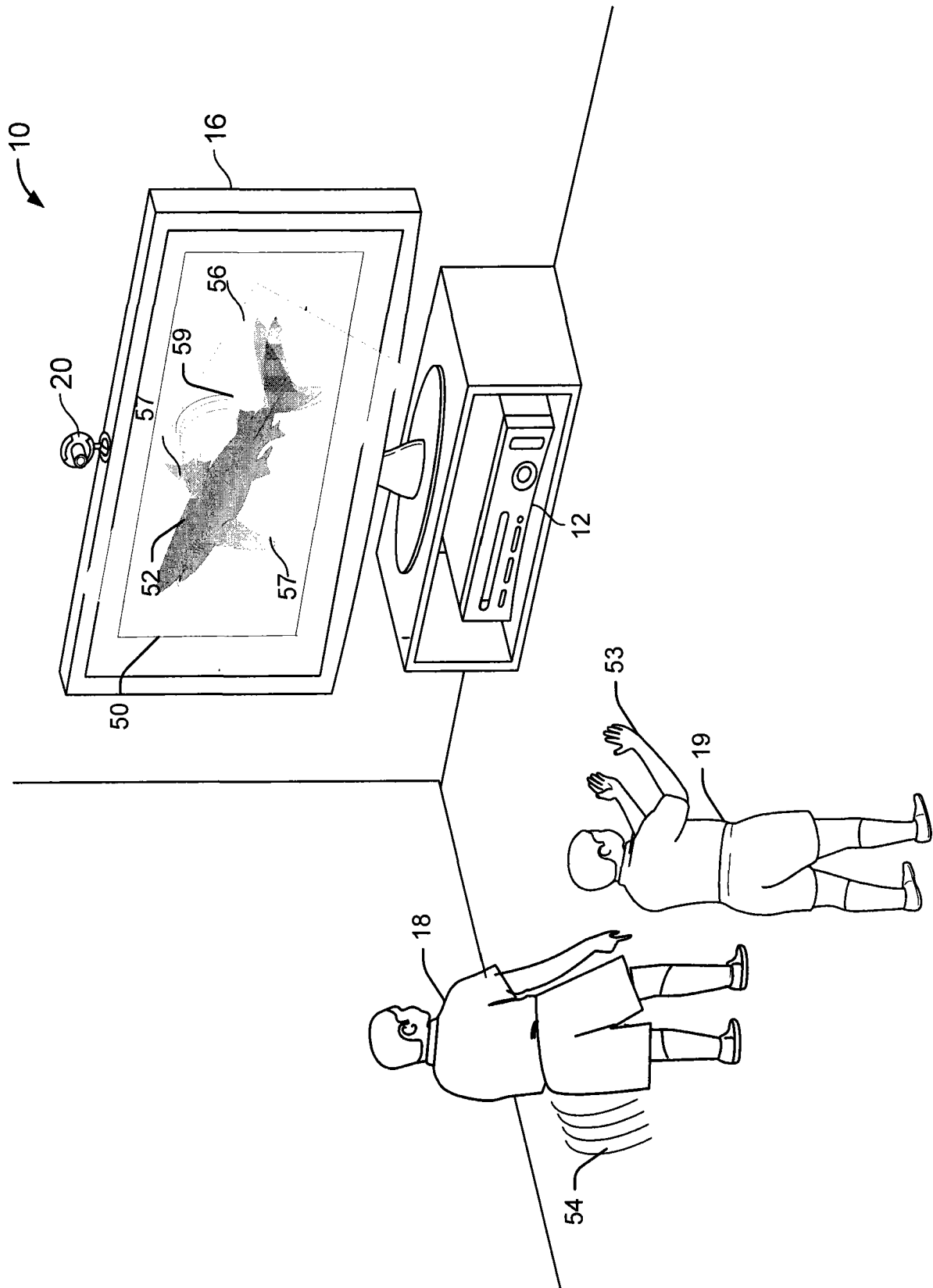


图 12

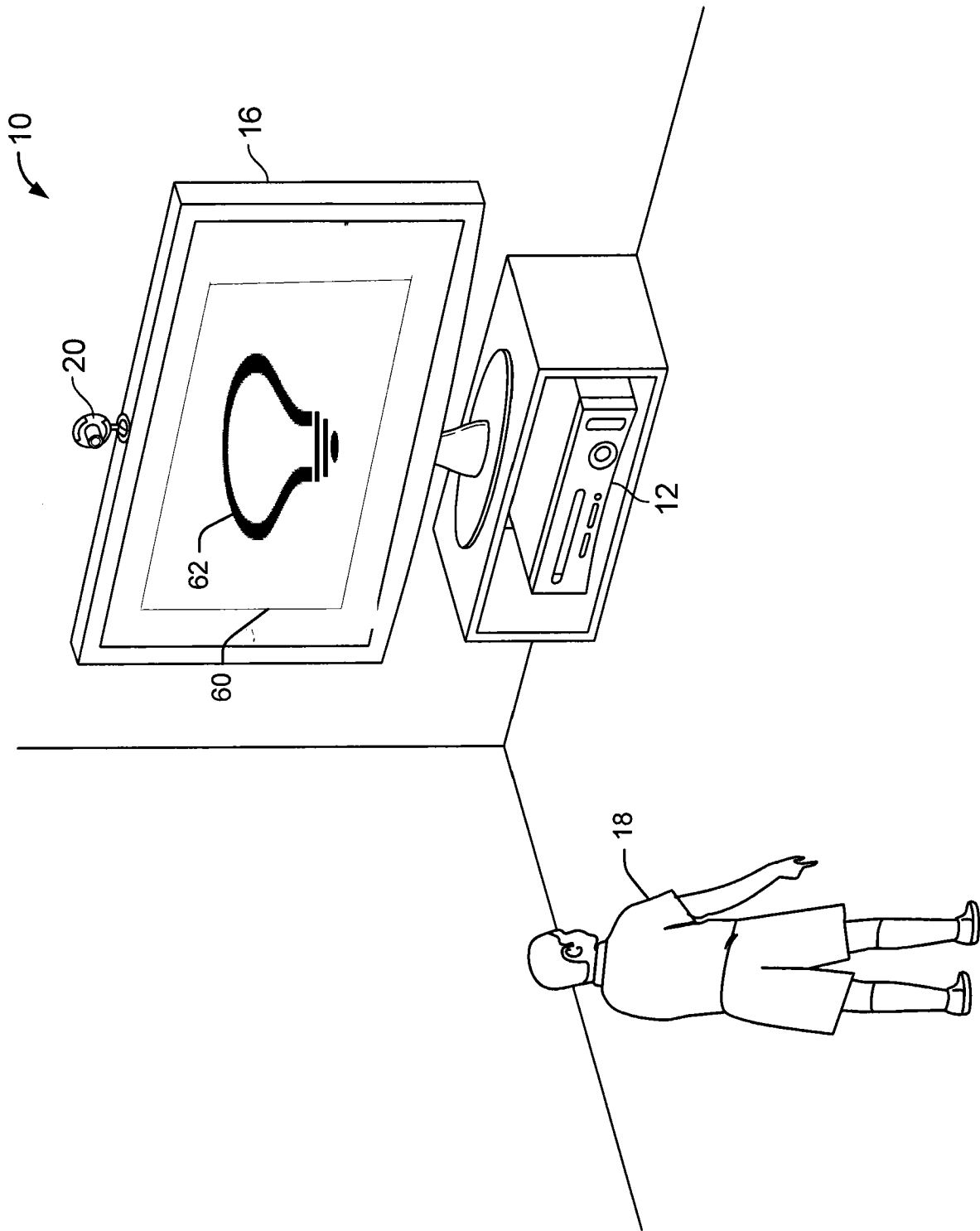


图 13

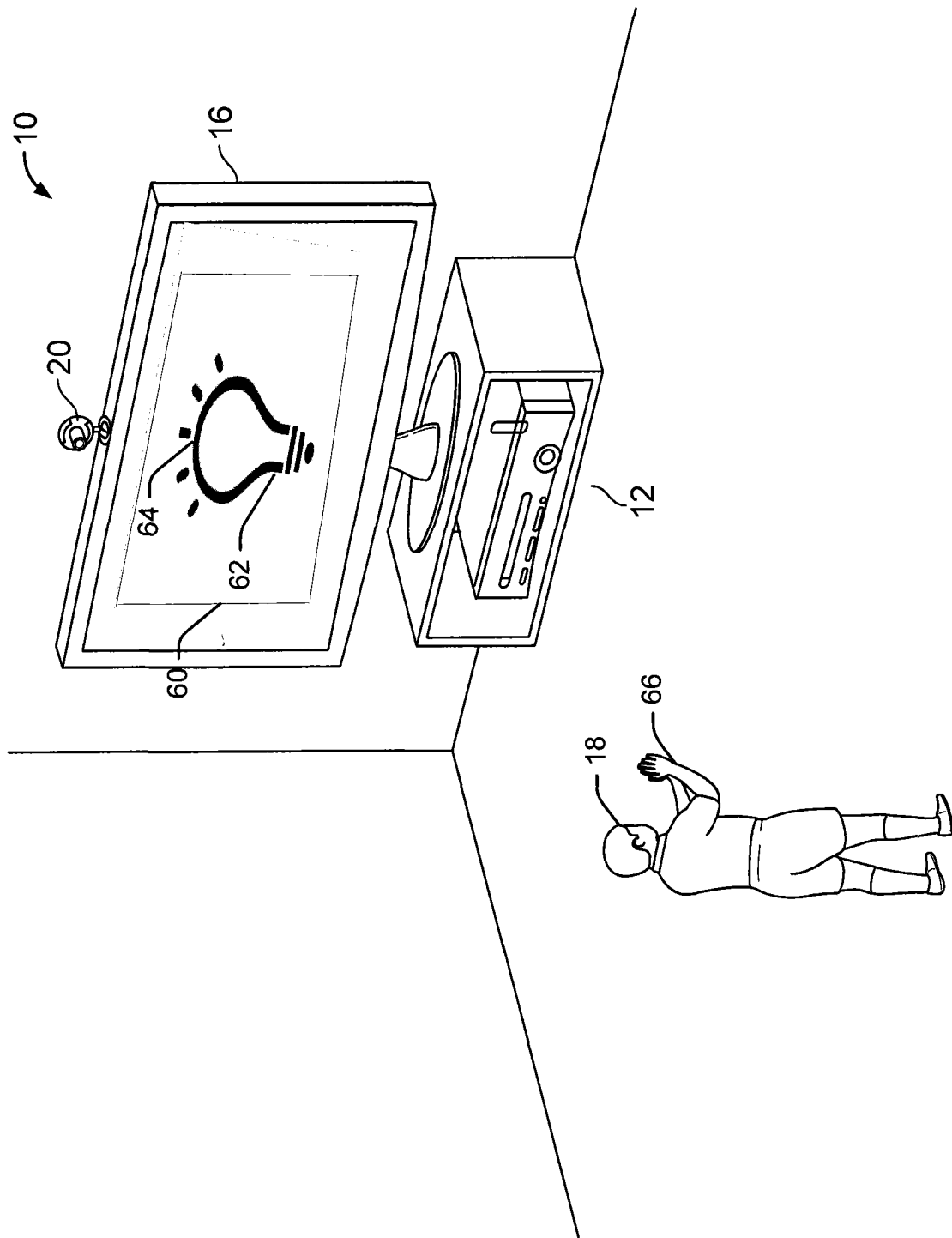


图 14