



(12) 发明专利

(10) 授权公告号 CN 102470274 B

(45) 授权公告日 2014. 10. 29

(21) 申请号 201080033128. 8

A63F 13/00 (2014. 01)

(22) 申请日 2010. 07. 27

A63F 13/20 (2014. 01)

G06T 19/00 (2011. 01)

(30) 优先权数据

12/511, 850 2009. 07. 29 US

(85) PCT国际申请进入国家阶段日

2012. 01. 20

(86) PCT国际申请的申请数据

PCT/US2010/043291 2010. 07. 27

(87) PCT国际申请的公布数据

W02011/014467 EN 2011. 02. 03

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 K·S·佩雷兹 A·基普曼

N·D·伯顿 A·威尔逊

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 杨洁

(51) Int. Cl.

G06F 19/00 (2011. 01)

(56) 对比文件

US 2004/0175039 A1, 2004. 09. 09,

KR 10-2008-0059865 A, 2008. 07. 01,

CN 101305401 A, 2008. 11. 12,

US 2007/0260984 A1, 2007. 12. 08,

审查员 王俊

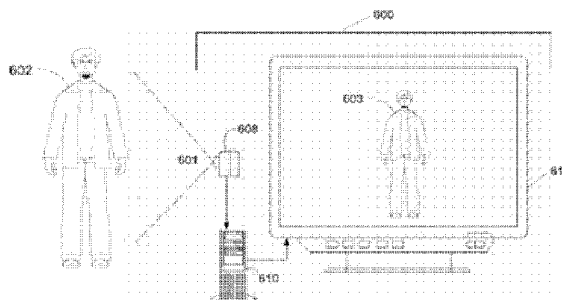
权利要求书2页 说明书25页 附图10页

(54) 发明名称

自动生成视觉表示

(57) 摘要

自动生成目标的视觉表示的技术可以减少或消除生成目标的视觉表示所需要的手动输入。例如,具有捕捉设备的系统可在物理空间中检测用户的各个特征,并且基于所检测到的特征从视觉表示特征选项库中作出特征选择。系统可以基于所检测到的特征自动地将选择应用于用户的视觉表示。或者,系统可以作出选择,该选择缩减了用户从中选择特征的选项的数量。系统可以实时地将选择应用于用户,以及实时地对所选择的和被应用到目标的视觉表示的特征进行更新。



1. 一种用于生成目标的视觉表示的方法,所述方法包括:
接收场景的数据,其中所述数据包括表示物理空间中的目标的数据(802);
从所述数据检测至少一个目标特征(806);
将至少一个所检测到的目标特征与视觉表示特征选项进行比较(806),其中所述视觉表示特征选项包括被配置为应用到所述目标的视觉表示的可选择的选项;
从所述视觉表示特征选项选择视觉表示特征(810);
将所述视觉表示特征应用于所述目标的所述视觉表示(816);
呈现所述视觉表示;
监视所述目标,并且检测所述至少一个所检测到的目标特征中的变化(818);
基于所述至少一个所检测到的目标特征中的变化,通过更新被应用于所述视觉表示的所述视觉表示特征,实时地更新所述目标的所述视觉表示(816)。
2. 如权利要求1所述的方法,其特征在于,从所述至少一个所检测到的特征与所述视觉表示特征选项的比较中自动生成所述视觉表示,使得对所述视觉表示特征的选择是在没有用户(18、602)的手动选择的情况下执行的。
3. 如权利要求1所述的方法,其特征在于,选择所述视觉表示特征包括选择与所述至少一个所检测到的目标特征相似的视觉表示特征(810)。
4. 如权利要求1所述的方法,其特征在于,所述视觉表示特征是面部特征、身体部位、颜色、大小、高度、宽度、形状、饰品、或服装项中的至少一个。
5. 如权利要求1所述的方法,其特征在于,还包括:
从所述视觉表示特征选项为所述视觉表示特征(810)生成视觉表示特征选项的子集(702);以及
提供所生成的特征选项的子集以供用户(18、602)选择要应用于所述视觉表示(810)的视觉表示特征。
6. 如权利要求5所述的方法,其特征在于,所生成的视觉表示特征选项的子集包括选择与所述至少一个所检测到的目标特征相似的多个视觉表示特征选项。
7. 如权利要求5所述的方法,其特征在于,还包括接收用户(18、602)从所生成的特征选项的子集中对所述视觉表示特征的选择(812),其中从所述视觉表示特征选项中选择所述视觉表示特征包括选择与所述用户(18、602)选择相对应的视觉表示特征。
8. 如权利要求1所述的方法,其特征在于,实时地呈现具有所述视觉表示特征的所述视觉表示。
9. 如权利要求1所述的方法,其特征在于,还包括在所述目标是人类目标的情况下,检测用户(18、602)的眼睛、嘴、鼻子或眉毛中至少一个的位置,并且使用所述位置来将对应的视觉表示特征与所述视觉表示对齐(816)。
10. 如权利要求1所述的方法,其特征在于,还包括基于提供所需修改的设置,修改所选择的视觉表示特征(816)。
11. 如权利要求10所述的方法,其特征在于,所述修改是基于计算尺的,所述计算尺可为所述视觉表示特征提供不同级别的修改。
12. 一种设备,所述设备包括:
捕捉设备(20),所述捕捉设备(20)用于接收场景的数据,其中所述数据包括表示物理

空间中的目标的数据；以及

处理器 (32)，所述处理器 (32) 用于执行计算机可执行指令，所述计算机可执行指令包括用于以下操作的指令：

从所述数据检测至少一个目标特征 (806)；

将至少一个所检测到的目标特征与视觉表示特征选项进行比较 (806)，其中所述视觉表示特征选项包括被配置为应用到视觉表示的可选择的选项；

从所述视觉表示特征选项中选择视觉表示特征 (810)；

将所述视觉表示特征应用于所述目标的所述视觉表示 (816)；

监视所述目标，并且检测所述至少一个所检测到的目标特征中的变化 (818)；

基于所述至少一个所检测到的目标特征中的变化，通过更新被应用于所述视觉表示的所述视觉表示特征，实时地更新所述目标的所述视觉表示 (816)。

13. 如权利要求 12 所述的设备，其特征在于，还包括用于实时地呈现所述视觉表示的显示设备 (193) (816)，其中所述处理器 (32) 从所述至少一个所检测到的特征与所述视觉表示特征选项的比较中自动生成所述视觉表示，使得对所述视觉表示特征的选择是在没有用户 (18、602) 的手动选择的情况下执行的。

14. 如权利要求 12 所述的设备，其特征在于，所述计算机可执行指令还包括用于以下操作的指令：

从所述视觉表示特征选项为所述视觉表示特征 (810) 生成视觉表示特征选项的子集 (702)；以及

在显示设备上提供所生成的特征选项的子集，以供用户 (18、602) 选择要应用于所述视觉表示 (810) 的视觉表示特征。

自动生成视觉表示

背景技术

[0001] 应用通常显示用户通过特定动作来控制的与该用户相对应的视觉表示,诸如选择遥控器上的按钮或以特定方式来移动控制器。视觉表示可以是化身、想象的人物、卡通图像或动物、光标、手等形式。视觉表示是在诸如计算机游戏、视频游戏、聊天、论坛、社区、即时消息收发服务等各种应用中通常采取二维(2D)或三维(3D)模型的形式计算机表示。诸如计算机游戏、多媒体应用、办公室应用等很多计算应用提供对预定义动画人物的选择,该预定义动画人物可被选择以在应用中用作用户的化身。

[0002] 允许创建化身的大多数系统也允许通过提供可被应用于该化身的可选择特征的数据库来定制该人物的外貌。例如,用户可访问应用中可用的衣服和配件的储存库并对化身的外貌作出修改。通常,用户将选择与该用户的自身特征最相似的特征。例如,用户可以选择具有与该用户相似身体结构的化身,并且随后该用户可以从特征的目录选择相似的眼睛、鼻子、嘴、头发等。然而,用于那些特征中每一个的特征的数量和选项的数量可能导致从中选择的过多的选项,而用户的视觉表示的手动生成可能变得繁重。系统可以限制可选择特征的数量以减少用户所需要的工作,但是这不合需要地限制了用户可用来生成唯一化身的特征。

发明内容

[0003] 可能期望的是应用或系统代表用户来作出用于用户的视觉表示的特征选择。使用所选择的特征,系统可以自动生成用户的视觉表示。例如,系统可以检测用户的各个特征,并基于所检测到的特征来作出特征选择。系统可以基于所检测到的特征自动地将选择应用于用户的视觉表示。或者,系统可以作出选择,该选择缩减了用户从中选择特征的选项的数量。如果系统可以代表用户作出决定,则用户可以无需作出很多决定或不从很多选项中进行选择。由此,所公开的技术可以移除用户大量的工作,并且可以代表用户作出选择并将其应用于用户的视觉表示。

[0004] 在示例实施例中,系统可以执行身体扫描并使用面部识别技术和/或身体识别技术来标识用户的特征。系统可以对与用户的所标识的特征最接近地类似的用户视觉表示作出选择。在另一个示例实施例中,系统可在将选择应用于视觉表示之前修改该选择。用户可在将选择应用于该用户的视觉表示之前指示系统作出修改。例如,如果用户是超重的,则该用户可以指示系统选择用于该用户的视觉表示的较苗条的身材。

[0005] 系统可以实时地将选择应用于用户。还期望系统从物理空间捕捉数据,标识用户的特性,并且实时地更新用户的视觉表示的特征。

[0006] 提供本发明内容以便以简化的形式介绍将在以下具体实施方式中进一步描述的一些概念。本发明内容既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,所请求保护的主体不限于解决在本公开的任一部分中提及的任何或所有缺点的实现。

附图说明

[0007] 参考附图来进一步描述根据本说明书的用于作出特征选择并自动生成视觉表示的系统、方法和计算机可读介质,在附图中:

[0008] 图 1 示出目标识别、分析、及跟踪系统的示例实施例,用户正在玩游戏。

[0009] 图 2 示出了可在目标识别、分析和跟踪系统中使用且结合了链接和动画混合技术的捕捉设备的示例实施例。

[0010] 图 3 示出了其中可实现本文描述的动画技术的计算环境的示例实施例。

[0011] 图 4 示出了其中可实现本文描述的动画技术的计算环境的另一示例实施例。

[0012] 图 5 示出了从深度图像生成的用户的骨架映射。

[0013] 图 6A-6B 每个描绘了示例目标识别、分析、及跟踪系统以及自动生成的视觉表示的示例实施例。

[0014] 图 7 描绘了为应用提供针对目标的视觉表示的特征选项的子集的示例目标识别、分析、及跟踪系统。

[0015] 图 8 描绘了自动生成视觉表示或为应用自动生成针对视觉表示的特征选项的子集的方法的示例流程图。

[0016] 图 9 描绘了使用目标数字化技术在物理空间中标识目标的示例目标识别、分析、及跟踪系统。

[0017] 说明性实施例的具体实施方式

[0018] 此处所公开的是用于提供目标的视觉表示的技术,诸如物理空间中的用户或非人类物体。用户的视觉表示例如可以是化身、屏幕上的光标、手、或物理空间中与用户相对应的任何其他虚拟物体的形式。人的骨架模型或网格模型的各方面可以基于捕捉设备所捕捉的图形数据来生成,并且可被求值以检测用户的特性。捕捉设备可以检测用户的特征并通过从特征的目录中选择特征来自动生成用户的视觉表示,该特征类似于那些所检测到的特征,诸如面部表情、头发颜色和类型、肤色和类型、服装、身体类型、身高、体重等。例如,系统可以使用面部识别和姿势/身体姿态识别技术从与已识别的特征相对应的特征选项的目录或数据库中自动地选择特征。系统可以实时地将所选择的特征、以及对那些特征的更新应用于用户的视觉表示。类似地,系统可在物理空间中检测非人类目标的特征,并从特征选项的目录中选择用于虚拟物体的特征。系统可显示与所检测到的特征相对应的虚拟物体。

[0019] 该计算环境可基于例如已被识别出并被映射到系统自动生成的视觉表示的用户的姿势来确定在计算机环境上执行的应用中要执行哪些控件。由此,可以显示虚拟用户,并且用户可通过在物理空间中作出姿势来控制虚拟用户的运动。所捕捉的运动可以是物理空间中由诸如相机等捕捉设备捕捉的任何运动。所捕捉的运动可包括物理空间中诸如用户或物体等目标的运动。所捕捉的运动可包括转换成操作系统或应用中的控制的姿势。运动可以是动态的,如奔跑运动,或者运动可以是静态的,例如以很少的移动摆姿态的用户。

[0020] 基于可检测的用户特性作出用于视觉表示的选择的面部和身体识别的系统、方法、技术和组件可在诸如游戏控制台等多媒体控制台中被具体化,或可在期望显示目标的视觉表示的任何其他计算设备中被具体化,作为示例但非有意的限制,这些其他计算设备包括卫星接收机、机顶盒、电子游戏机、个人计算机(PC)、便携式电话、个人数字助理(PDA)、以及其他手持式设备。

[0021] 图 1 示出了目标识别、分析、及跟踪系统 10 的配置的示例实施例,该目标识别、分析、及跟踪系统 10 可以采用将用户的特性应用于化身的技术。在示例实施例中,用户 18 正在玩拳击游戏。在示例实施例中,系统 10 可识别、分析和 / 或跟踪诸如用户 18 的人类目标。系统 10 可以收集与物理空间中用户的运动、面部表情、身体语言、情绪等相关的信息。例如,系统可以标识并扫描人类目标 18。系统 10 可使用身体姿态识别技术来标识人类目标 18 的身体类型。系统 10 可以标识用户 18 的身体各部位以及它们如何移动。系统 10 可以将所检测到的用户特征与可选择的视觉表示特征的目录进行比较。

[0022] 如图 1 所示,目标识别、分析及跟踪系统 10 可包括计算环境 12。计算环境 12 可以是计算机、游戏系统或控制台等等。根据一示例实施例,计算环境 12 可包括硬件组件和 / 或软件组件,使得计算环境 12 可用于执行诸如游戏应用、非游戏应用等应用。

[0023] 如图 1 所示,目标识别、分析及跟踪系统 10 还可包括捕捉设备 20。捕捉设备 20 可以是例如相机,该相机可用于在视觉上监视诸如用户 18 等一个或多个用户,以使得可以捕捉、分析并跟踪一个或多个用户所执行的姿势以执行应用中的一个或多个控制或动作,如将在下面更详细地描述的。

[0024] 根据一个实施例,目标识别、分析和跟踪系统 10 可连接到可向诸如用户 18 等用户提供游戏或应用视觉和 / 或音频的视听设备 16,如电视机、监视器、高清电视机 (HDTV) 等。例如,计算环境 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 可从计算环境 12 接收视听信号,然后可向用户 18 输出与该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 电缆等连接到计算环境 12。

[0025] 如图 1 所示,目标识别、分析及跟踪系统 10 可用于识别、分析和 / 或跟踪诸如用户 18 等人类目标。例如,可使用捕捉设备 20 来跟踪用户 18,以使得可将用户 18 的移动解释为可用于影响由计算机环境 12 执行的的应用的控制。由此,根据一个实施例,用户 18 可移动他或她的身体来控制应用。系统 10 可以跟踪用户的身体以及用户的身体所作的运动,包括控制诸如应用、操作系统等系统的各方面的姿势。

[0026] 系统 10 可将捕捉设备 20 的输入转换成动画,该输入表示用户的运动,使得动画由该输入驱动。由此,用户的运动可映射到化身 40,使得物理空间中该用户的运动由该化身 40 执行。用户的运动可以是可适用于应用中的控件的姿势。如图 1 所示,在示例实施例中,在计算环境 12 上执行的应用可以是用户 18 可能正在玩的拳击游戏。

[0027] 计算环境 12 可使用视听设备 16 来提供用户 18 可通过他的或她的移动来控制的玩家化身 40 的视觉表示。系统可将运动和 / 或姿势应用于用户的视觉表示,该视觉表示可以是系统基于用户的所检测到的特征而自动生成的自动生成的视觉表示。例如,用户 18 可在物理空间中挥拳猛击,这使得玩家化身 40 在游戏空间中挥拳猛击。玩家化身 40 可以具有捕捉设备 20 所标识的用户的特性,或者系统可以为映射到用户的运动的视觉表示使用知名拳击运动员的特征或描绘专业拳击运动员的体格。系统 10 可以跟踪用户,并且可基于物理空间中用户可检测的特征来修改用户化身的特性。计算环境 12 还可使用视听设备 16 来向用户 18 提供拳击对手 38 的视觉表示。根据一示例实施例,目标识别、分析和跟踪系统 10 的计算环境 12 和捕捉设备 20 可用于识别和分析用户 18 在物理空间中的出拳,使得该出

拳可被解释为对游戏空间中的玩家化身 40 的游戏控制。多个用户可从远程位置彼此交互。例如,拳击对手 38 的视觉表示可以表示另一个用户,诸如物理空间中的第二用户与用户 18 或第二物理空间中的联网用户。

[0028] 用户 18 的其他移动也可被解释为其他控制或动作,诸如上下快速摆动、闪避、滑步、格挡、直拳或挥动各种不同力度的拳等控制。此外,某些移动可被解释为可对应于除控制玩家化身 40 之外的动作的控制。例如,玩家可以使用移动来结束、暂停或保存游戏、选择级别、查看高分、与朋友交流等。另外,用户 18 的全范围运动可以用任何合适的方式来获得、使用并分析以与应用进行交互。

[0029] 在各示例实施例中,诸如用户 18 等人类目标可持有一物体。在这些实施例中,电子游戏的用户可手持物体,使得可使用玩家和物体的运动来调整和 / 或控制游戏的参数。例如,可以跟踪并利用玩家手持球拍的运动来控制电子运动游戏中的屏幕上球拍。在另一示例实施例中,可以跟踪并利用玩家手持物体的运动来控制电子格斗游戏中的屏幕上武器。

[0030] 用户的姿势或运动可被解释为可对应于除控制玩家化身 40 之外的动作的控制。例如,玩家可以使用移动来结束、暂停或保存游戏、选择级别、查看高分、与朋友交流等。玩家可以使用移动来将修改应用于化身。例如,用户可在物理空间中摇晃他的或她的手臂,而这可以是被系统 10 标识为请求使化身的手臂更长的姿势。事实上操作系统和 / 或应用的任何可控方面可由诸如用户 18 等目标的移动来控制。根据其他示例实施例,目标识别、分析和跟踪系统 10 可解释目标移动用于控制游戏领域之外的操作系统和 / 或应用的各方面。

[0031] 用户的姿势可以是可适用于操作系统、游戏的非游戏方面、或非游戏应用的控制。用户的姿势可被解释为物体操纵,例如控制用户界面。例如,考虑具有从左向右垂直排列的叶片 (blades) 或选项卡式界面的用户界面,其中对每个叶片或选项卡的选择打开应用或系统内的各种控制的选项。系统可标识移动选项卡的用户的手姿势,其中物理空间中的用户的手虚拟地与应用空间中的选项卡对齐。包括暂停、抓取运动以及然后的手向左扫动在内的姿势可被解释为对选项卡的选择,然后将其移开以打开下一选项卡。

[0032] 图 2 示出了可用于目标识别、分析和跟踪的捕捉设备 20 的示例实施例,其中目标可以是用户或物体。根据一示例实施例,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、立体图像等来捕捉包括深度图像的带有深度信息的视频,该深度信息可包括深度值。根据一个实施例,捕捉设备 20 可将所计算的深度信息组织成“Z 层”,或与从深度相机沿其视线延伸的 Z 轴垂直的层。

[0033] 如图 2 所示,捕捉设备 20 可包括图像相机组件 22。根据一个示例实施例,图像相机组件 22 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、毫米等单位的、所捕捉场景中的物体距相机的长度或距离。

[0034] 如图 2 所示,根据一示例实施例,图像相机组件 22 可包括可用于捕捉场景的深度图像的 IR 光组件 24、三维 (3-D) 相机 26、和 RGB 相机 28。例如,在飞行时间分析中,捕捉设备 20 的 IR 光组件 24 可将红外光发射到场景上,并且随后可使用传感器 (未示出)、用例如 3-D 相机 26 和 / 或 RGB 相机 28 来检测从场景中的一个或多个目标和物体的表面反向散射的光。在一些实施例中,可使用脉冲红外光,从而可测量出射光脉冲和相应入射光脉冲之间

的时间差,并且将其用于确定从捕捉设备 20 到场景中的目标或物体上的特定位置的物理距离。另外,在其他示例实施例中,可将出射光波的相位与入射光波的相位作比较来确定相移。该相移随后可用于确定从捕捉设备 20 到目标或物体上特定位置的物理距离。

[0035] 根据另一示例实施例,飞行时间分析可用于通过经由包括例如快门式光脉冲成像在内的各种技术来随时间分析反射光束的强度来间接地确定从捕捉设备 20 到目标或物体上的特定位置的物理距离。

[0036] 在另一示例实施例中,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,被显示为例如网格图案或条纹图案的已知图案的光)可经由例如 IR 光组件 24 被投影到场景上。在落到场景中的一个或多个目标或物体的表面时,作为响应,图案可变形。图案的此类变形可被例如 3-D 相机 26 和 / 或 RGB 相机 28 捕捉,并且随后可被分析以确定从捕捉设备 20 到目标或物体上的特定位置的物理距离。

[0037] 根据另一实施例,捕捉设备 20 可包括可以从不同的角度观察场景的两个或更多个在物理上分开的相机,以获取可以被解析以生成深度信息的视觉立体数据。

[0038] 在另一示例实施例中,捕捉设备 20 可使用点云数据(point cloud data)和目标数字化技术来检测用户的特征。这些技术在以下关于图 2B 更详细地描述。

[0039] 捕捉设备 20 还可包括话筒 30 或话筒阵列。话筒 30 可包括可接收声音并将其转换成电信号的变换器或传感器。根据一个实施例,话筒 30 可用于减少目标识别、分析和跟踪系统 10 中的捕捉设备 20 和计算环境 12 之间的反馈。另外,话筒 30 可用于接收也可由用户提供的音频信号,以控制可由计算环境 12 执行的诸如游戏应用、非游戏应用程序等应用。

[0040] 在一个示例实施例中,捕捉设备 20 还可包括可与图像相机组件 22 有效通信的处理器 32。处理器 32 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接收深度图像的指令、用于确定合适的目标是否可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。

[0041] 例如,计算机可读介质可包括用于接收场景的数据的计算机可执行指令,其中该数据包括表示物理空间中的目标的数据。该指令包括用于从该数据中检测至少一个目标特征并将至少一个所检测到的目标特征与来自特征库 197 的视觉表示特征选项进行比较的指令。视觉表示特征选项可包括被配置为应用到视觉表示的可选择选项。其他指令允许从视觉表示特征选项中选择视觉表示特征,将视觉表示特征应用于目标的视觉表示,以及呈现该视觉表示。视觉表示可从至少一个检测到的特征与视觉表示特征选项的比较中自动生成,使得对视觉表示特征的选择是在没有用户的手动选择的情况下执行的。

[0042] 对视觉表示特征的选择可包括选择类似于所检测到的目标特征的视觉表示特征。视觉表示特征可以是脸部特征、身体部位、颜色、大小、高度、宽度、形状、饰品、或服装项中的至少一个。该指令可允许从用于视觉表示特征的视觉表示特征选项中生成视觉表示特征选项子集,并提供所生成的特征选项子集以供用户选择要应用于视觉表示的视觉表示特征。所生成的视觉表示特征选项子集可包括类似于所检测到的目标特征的多个视觉表示特征选项。该指令可允许接收用户从所生成的特征选项子集中对视觉表示特征的选择,其中从视觉表示特征选项中选择视觉表示特征包括选择对应于该用户选择的视觉表示特征。具有视觉表示特征的视觉表示可被实时呈现。此外,该指令可允许监视目标并检测所检测到

的目标特征中的变化,并通过基于所检测到的目标特征中的变化实时地更新应用于视觉表示的视觉表示特征来更新目标的视觉表示。

[0043] 捕捉设备 20 还可包括存储器组件 34,存储器组件 34 可存储可由处理器 32 执行的指令、3-D 相机 26 或 RGB 相机 28 所捕捉的图像或图像的帧、或任何其他合适的信息、图像等等。根据一个示例实施例,存储器组件 34 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,在一个实施例中,存储器组件 34 可以是与图像捕捉组件 22 和处理器 32 进行通信的单独的组件。根据另一实施例,存储器组件 34 可被集成到处理器 32 和 / 或图像捕捉组件 22 中。

[0044] 如图 2 所示,捕捉设备 20 可经由通信链路 36 与计算环境 12 进行通信。通信链路 36 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 连接等的无线连接。根据一个实施例,计算环境 12 可以经由通信链路 36 向捕捉设备 20 提供时钟,可以使用该时钟来确定何时捕捉例如场景。

[0045] 另外,捕捉设备 20 可以通过通信链路 36 向计算环境 12 提供深度信息和由例如 3-D 相机 26 和 / 或 RGB 相机 28 捕捉到的图像,以及可以由捕捉设备 20 生成的骨架模型。计算环境 12 然后可使用该骨架模型、深度信息和捕捉的图像来例如控制例如游戏或文字处理程序之类的应用。例如,如图 2 所示,计算环境 12 可包括姿势库 192。

[0046] 如图所示,在图 2 中,计算环境 12 可包括姿势库 192 和姿势识别引擎 190。姿势识别引擎 190 可包括姿势过滤器 191 的集合。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。每一过滤器 191 可包括定义姿势以及该姿势的参数或元数据的信息。例如,包括一只手从身体背后到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到身体前方的移动的信息的姿势过滤器 191,该移动将由深度相机来捕捉。随后可为该姿势设定参数。在姿势是投掷的情况下,参数可以是该手必须达到的阈值速度、该手必须行进的距离(绝对的,或相对于用户的整体大小)、以及识别器引擎对发生了该姿势的置信评级。用于姿势的这些参数可随时间在各应用之间、在单个应用的各个上下文之间、或在一个应用的一个上下文内变化。

[0047] 尽管构想了姿势识别引擎 190 可包括姿势过滤器的集合,其中过滤器可包括表示用于处理深度、RGB 或骨架数据的代码或以其他方式表示用于进行这样的处理的组件,但对过滤器的使用不旨在将分析限于过滤器。过滤器是分析系统所接收的场景的数据,并将该数据与表示姿势的基础信息进行比较的示例组件或代码部分的表示。作为分析的结果,系统可产生对应于输入数据是否对应于姿势的输出。表示姿势的基础信息可被调整为对应于代表用户的捕捉运动的数据历史中重复出现的特征。例如,基础信息可以是如上所述的姿势过滤器的一部分。但是,构想了任何合适的用于分析输入数据和姿势数据的方式。

[0048] 在示例实施例中,姿势可被识别为进入修改模式的触发,在修改模式中,用户可修改系统自动生成的视觉表示。例如,姿势过滤器 191 可包括用于识别修改触发姿势的信息。如果识别到修改触发姿势,则应用可进入修改模式。修改触发姿势可在应用之间、在系统之间、在用户之间等变化。例如,网球游戏应用中相同的姿势可能不是保龄球游戏应用中相同的修改触发姿势。考虑包括用户运动在该用户身体前方的用户的右手,同时食指向上指并以圆圈运动移动的示例修改触发姿势。为该修改触发姿势设置的参数可用于标识用户的手在用户身体的前方,用户的食指以向上运动指向,并标识食指正在以圆圈运动移动。

[0049] 某些姿势可被标识为进入修改模式的请求,其中如果应用当前正在执行,则修改模式中断应用的当前状态并进入修改模式。修改模式可使得应用暂停,其中应用可在用户离开修改模式时在暂停点继续。或者,修改模式可能不导致对应用的暂停,且应用可在用户作出修改的同时继续执行。

[0050] 可将由相机 26、28 和设备 20 捕捉的骨架模型以及与其相关联的移动的形式的数据与姿势库 192 中的姿势过滤器 191 进行比较来标识用户(如骨架模型所表示的)何时执行了一个或多个姿势。由此,对诸如过滤器 191 等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,像在关节处相交的骨所形成的角度、来自场景的 RGB 色彩数据、以及用户的某一方面的变化速率等内容。如所提到的,可为姿势设置参数。来自过滤器 191 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、以及姿势发生的时间等内容。

[0051] 计算环境 12 可包括处理器 195,处理器 196 可处理深度图像来确定场景中有什么目标,如房间中的用户 18 或物体。这可例如通过将深度图像中共享相似距离值的像素分组在一起来实现。该图像也可被解析以产生用户的骨架表示,其中标识例如关节和各关节之间的组织等特征。存在骨架映射技术,其使用深度相机来捕捉个人,并从中确定该用户骨架上的多个点,手、腕、肘、膝、鼻、踝、肩的关节,以及骨盆与脊椎相交之处。其他技术包括将图像转换为人的身体模型表示以及将图像转换为人的网格模型表示。

[0052] 在实施例中,处理是在捕捉设备 20 本身上执行的,且深度和色彩(其中捕捉设备 20 包括 3D 相机 26)值的原始图像数据经由链路 36 被发送到计算环境 12。在另一实施例中,处理由耦合到相机 402 的处理器 32 来执行,然后经解析的图像数据被发送到计算环境 12。在又一实施例中,原始图像数据和经解析的图像数据两者被发送到计算环境 12。计算环境 12 可接收经解析的图像,但是它仍可接收原始数据来执行当前过程或应用。例如,如果场景的图像通过计算机网络发送到另一用户,则计算环境 12 可发送供另一计算环境处理的原始数据。

[0053] 处理器可具有特征比较模块 196。特征比较模块 196 可将目标的所检测到的特征与特征库 197 中的选项进行比较。特征库 197 可提供视觉表示特征选项,如颜色选项、面部特征选项、身体类型选项、大小选项等,并且选项可以对人类和非人类目标变化。库可以是存储视觉表示的特征的目录、数据库、存储器等。库可以是特征选项的有组织的或无组织的集合。系统或用户可以将特征添加到目录。例如,应用可具有一组预封装的特征选项,或者系统可具有默认数量的可用特征。附加特征选项可被添加到特征库 197 或在其中更新。例如,用户可在虚拟市场购买附加特征选项,用户可将特征选项赠予另一用户,或者系统可通过拍摄用户的所检测到的特征的快照来生成特征选项。

[0054] FCM 196 可诸如从特征选项目录中作出最接近地类似于目标的所检测到的特征的特征选择。系统可自动生成具有所检测到的特征的虚拟物体。例如,考虑对物理空间中的红色的、两个座位的长椅的检测。系统可从特征库 197 中标识单独地或组合地类似于该长椅的所检测到的目标特征的特征。在示例实施例中,从特征库 197 中的选择可以就像选择具有物理目标的至少一个特征的虚拟目标那样简单。例如,特征库 197 可具有用于家具的众多特征选项,并且可包括对红色的、两个座位的长椅的虚拟图像或描绘。这些特征可以是预封装的,并随应用或随系统一起提供。在另一示例中,系统可拍摄该物理长椅的快照,并创建具有该物理长椅的形状的卡通或虚拟图像。由此,所选特征可以是来自系统先前拍摄

并添加到特征库 197 的该物理长椅的快照。

[0055] 系统可以基于所检测到的目标特征来调整所选特征的颜色、定位或比例。例如，系统可从特征库 197 中选择类似于所检测到的目标的特征的特征或组合若干特征。系统可将特征添加到所选特征或虚拟图像来更完全地类似于所检测到的目标。在所检测到的长椅的示例中，系统可执行在特征库 197 中的特征查找，并标识具有类似于物理长椅的特征的至少一个特征的长椅的虚拟框架。例如，系统可最初选择形状类似于所检测到的物理长椅的虚拟长椅。如果虚拟的两个座位的长椅是可用特征选项，则系统可选择该虚拟的两个座位的长椅。颜色可以是可由系统选择的特征选项。在该示例中，如果红色长椅明确不是特征库 197 中的选项，则系统可从特征库 197 中选择颜色并将其应用于所选的虚拟框架。系统可在特征库 197 中选择类似于物理长椅的所检测到的红色的现有颜色，或者系统可拍摄物理长椅的颜色的快照并将其作为特征选项添加到特征库。系统可将所选红色特征应用于虚拟长椅图像。

[0056] 在另一示例中，系统可以组合来自特征库的特征以生成类似于所检测到的目标的可视物体。例如，系统可以通过从来自特征库 197 的长椅特征选项（诸如臂、腿、座位、坐垫、靠背、脊等）中进行选择来生成两个座位的长椅并使用所选择的特征来拼凑长椅。

[0057] 在另一示例中，目标是人类用户，且该系统检测用户的特征，诸如眼睛颜色、大小和形状，头发颜色、类型和长度等。系统可将所检测到的特征与特征选项目录进行比较，并将所选特征应用于视觉表示。如上所述，系统可组合特征并更改这些特征。例如，特征可通过向目标应用颜色、定位或比例来更改。特征可通过从特征库 197 中选择诸如颜色等附加特征或通过使用来自目标的快照的图像数据来更改。例如，应用可在特征库 197 中提供一组通用的纯色裤子、T 恤和鞋子类型。系统可从该通用服装特征中进行选择，并通过向服装应用颜色来反映系统检测到的目标的服装的颜色来更改所选的服装特征。

[0058] 在另一示例中，系统可标识特征库 197 中类似于用户的特征的特征子集，并提供用户可从中进行选择的该子集。由此，提供给用户的用于某一特征的选项数量可以被智能地过滤以便使用户更容易定制视觉表示。

[0059] 特征库可适用于应用或者可以是系统范围的。例如，游戏应用可以定义指示了适用于该游戏的各种性情的特征。特征选项可包括具体特征和通用特征。还注意到，对查找表或数据库的引用是示例性的，并且构想了对与此处公开的技术相关的特征选项的供应可以通过任何适合的方式来访问、存储、封装、提供、生成等。

[0060] 计算环境 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。计算环境 12 可对用户的表示进行建模和显示，例如采用诸如显示设备 193 之类的显示器上的化身或指针的形式。显示设备 193 可包括计算机监视器、电视机屏幕、或任何合适的显示设备。例如，相机控制的计算机系统可捕捉用户图像数据，并在电视机屏幕上显示映射到用户的姿势的用户反馈。用户反馈可被显示为屏幕上的化身，如图 1A 和 1B 中所示。化身的运动可直接通过将化身的移动映射到用户的移动来控制。用户的姿势可以被解释来控制应用的某些方面。

[0061] 根据一示例实施例，目标可以是诸如站着或坐着等处于任何位置的人类目标、具有物体的人类目标、两个或更多人类目标、一个或多个人类目标的一个或多个附件等，目标可被扫描、跟踪、建模和 / 或评估来生成虚拟屏幕，将用户与一个或多个所存储的简档进行

比较和 / 或在诸如计算环境 12 等计算环境中存储关于目标的简档信息 198。简档信息 198 可以采用用户简档、个人简档、应用简档、系统简档的形式,或用于存储数据以供稍后访问的任何其他合适的方法。简档信息 198 可以经由应用来访问,或者可以是例如系统范围可用的。简档信息 198 可包括用于加载特定用户简档信息的查找表。虚拟屏幕可与可由以上关于图 1A-1B 描述的计算环境 12 执行的应用进行交互。

[0062] 系统可通过基于存储在用户简档中的信息自动生成视觉表示来呈现诸如用户等目标的视觉表示。根据各示例实施例,查找表可包括用户专用简档信息。在一个实施例中,诸如计算环境 12 等计算环境可在查找表中包括关于一个或多个用户的所存储的简档数据 198。所存储的简档数据 198 可特别地包括目标的所扫描的或所估计的身体大小、骨架模型、身体模型、语音样本或口令、目标的性别、目标年龄、先前的姿势、目标限制、以及目标对系统的标准使用,例如坐、惯用左手或惯用右手的趋势,或站在非常靠近捕捉设备之处的趋势。该信息可用于确定在捕捉场景中的目标与一个或多个用户简档 198 之间是否有匹配,在一个实施例中,该匹配可允许系统将虚拟屏幕自适应到用户,或根据简档 198 来自适应计算或游戏体验的其他元素。

[0063] 先前为目标的视觉表示所选择的特征可被存储在简档中。例如,用户专用简档可存储所选并被应用来自动生成用户的视觉表示的特征。位置专用简档可存储所选并被应用来自动生成并显示类似于物理空间的虚拟场景的特征。例如,对应于诸如房间中的家具等物理空间中的物体的虚拟物体可通过从特征库 197 中选择选项来生成。可检测颜色,并且可从特征库 197 中选择可用颜色。在系统识别或初始化时,可加载位置专用简档,从而显示对应于该位置的家具和颜色。

[0064] 一个或多个个人简档 198 可被存储在计算机环境 12 中,并在多个用户会话中使用,或者一个或多个个人简档可仅为单个会话而创建。用户可以有建立简档的选项,其中他们可向系统提供信息,诸如语音或身体扫描、年龄、个人偏好、惯用右手或左手、化身、姓名等。还可以为除步入捕捉空间以外不向系统提供任何信息的“访客”提供个人简档。可以为一个或多个访客建立临时的个人简档。在访客会话的结尾,可以存储或删除访客个人简档。

[0065] 姿势库 192、姿势识别引擎 190、特征库 197、特征比较器 196 和简档 198 可以用硬件、软件或两者的组合来实现。例如,姿势库 192 和姿势识别引擎 190 可被实现为在计算环境 12 的诸如处理器 195 等处理器上执行(或在图 3 的处理单元 101 或图 4 的处理单元 259 上)执行的软件。

[0066] 要强调的是,以下描述的图 3-4 中描绘的框图是示例性的,且不在暗示一具体实现。由此,图 1 的处理器 195 或 32、图 3 的处理单元 101、和图 4 的处理单元 259 可被实现为单个处理器或多个处理器。多个处理器可以分布式或集中式地定位。例如,姿势库 192 可被实现为在捕捉设备的处理器 32 上执行的软件,或者它可被实现为在计算环境 12 中的处理器 195 上执行的软件。构想了适用于执行此处公开的技术的处理器器的任意组合。多个处理器可无线地、经由硬连线、或以其组合来通信。

[0067] 此外,如此处所使用的,计算环境 12 可以指的是单个计算设备或计算系统。计算环境可包括非计算组件。计算环境可包括显示设备,如图 2 所示的显示设备 193。显示设备可以是与计算环境分开但与其耦合的实体,或者显示设备可以是例如进行处理和显示的计

算设备。由此,计算系统、计算设备、计算环境、计算机、处理器或其他计算组件可被互换地使用。

[0068] 姿势库和过滤器参数可由姿势工具为应用或应用的上下文来调节。上下文可以是文化上下文,并且可以是环境上下文。文化上下文指的是使用系统的用户的文化。不同的文化可使用相似的姿势来赋予显著不同的含义。例如,希望叫另一个用户“看”或“使用他的眼睛”的美国用户可将他的食指放在他头上靠近他的眼睛的远端处。然而,对意大利用户而言,该姿势可被解释为对黑手党的引用。

[0069] 类似地,在单个应用的不同环境之中可能有不同的上下文。以涉及操作摩托车的第一用户射击游戏为例。当用户在走路时,将手指朝向地面握拳并向前且从身体向外伸出拳头可表示出拳姿势。当用户在驾驶上下文中时,相同的运动可表示“换挡”姿势。关于对视觉表示的修改,不同的姿势可依赖于环境而触发不同的修改。不同的修改触发姿势可用于进入相对于系统范围修改模式的应用专用修改模式。每个修改模式可以与对应于该修改模式的独立的姿势集包装在一起,作为修改触发姿势的结果而进入。例如,在保龄球游戏中,挥动手臂运动可以是被标识为挥动保龄球以便向下释放到虚拟保龄球道的姿势。然而,在另一个应用中,挥动手臂运动可以是被标识为请求延长屏幕上所显示的用户化身的手臂的姿势。可能还有一个或多个菜单环境,其中用户可保存他的游戏、在他的人物装备之间选择或执行类似的不包括直接玩游戏的动作。在该环境中,该游戏姿势可具有第三个含义,如选择某样东西或前进到另一屏幕。

[0070] 姿势可被一起分组到可能由该风格中的应用使用的互补姿势风格包中。互补姿势 – 或者如通常一起使用的那些姿势中那样是互补的,或者如一个姿势的参数改变将改变另一姿势的参数中那样是互补的 – 被一起分组到风格包中。这些包可被提供给应用,应用可选择至少其中一个。应用可调节或修改姿势或姿势过滤器 191 的参数来最佳地适合应用的独特方面。当调节该参数时,也调节该姿势或第二姿势的第二互补参数(在相互依赖的意义上)使得这些参数保持互补。用于视频游戏的风格包可包括诸如第一用户射击、动作、驾驶和体育等风格。

[0071] 图 3 示出可用于解释目标识别、分析和跟踪系统中的一个或多个姿势的计算环境的示例实施例。以下参考附图 1-2 所描述的诸如计算环境 12 的计算环境可以是多媒体控制台 100,诸如游戏控制台。图 3 的计算环境 100 可以是诸如游戏控制台等多媒体控制台 -2。如图 3 所示,多媒体控制台 100 具有含有一级高速缓存 102、二级高速缓存 104 和闪存 ROM(只读存储器)106 的中央处理单元(CPU)101。一级高速缓存 102 和二级高速缓存 104 临时存储数据,并且因此减少存储器访问周期的数量,由此改进处理速度和吞吐量。CPU 101 可被设置成具有一个以上的内核,并且由此附加一级和二级高速缓存 102 和 104。闪存 ROM 106 可存储在多媒体控制台 100 通电时引导过程的初始阶段期间加载的可执行代码。

[0072] 图形处理单元(GPU)108 和视频编码器/视频编解码器(编码器/解码器)114 形成用于高速和高分辨率图形处理的视频处理流水线。经由总线从图形处理单元 108 向视频编码器/视频编解码器 114 运送数据。视频处理流水线向 A/V(音频/视频)端口 140 输出数据,用于传输至电视或其他显示器。存储器控制器 110 连接到 GPU 108 以方便处理器访问各种类型的存储器 112,诸如但不限于 RAM(随机存取存储器)。

[0073] 多媒体控制台 100 包括较佳地在模块 118 上实现的 I/O 控制器 120、系统管理控

制器 122、音频处理单元 123、网络接口控制器 124、第一 USB 主控制器 126、第二 USB 控制器 128 和前面板 I/O 子部件 130。USB 控制器 126 和 128 用作外围控制器 142(1)-142(2)、无线适配器 148、和外置存储器设备 146(例如闪存、外置 CD/DVD ROM 驱动器、可移动介质等)的主机。网络接口 124 和 / 或无线适配器 148 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、蓝牙模块、电缆调制解调器等各种不同的有线或无线适配器组件中任何一种。

[0074] 提供系统存储器 143 来存储在引导过程期间加载的应用数据。提供媒体驱动器 144,且其可包括 DVD/CD 驱动器、硬盘驱动器、或其他可移动媒体驱动器等。媒体驱动器 144 可以是对多媒体控制器 100 内置的或外置的。应用数据可经由介质驱动器 144 访问,以供多媒体控制台 100 执行、回放等。介质驱动器 144 经由诸如串行 ATA 总线或其他高速连接(例如 IEEE 1394)等总线连接到 I/O 控制器 120。

[0075] 系统管理控制器 122 提供与确保多媒体控制台 100 的可用性相关的各种服务功能。音频处理单元 123 和音频编解码器 132 形成具有高保真度和立体声处理的相应音频处理流水线。音频数据经由通信链路在音频处理单元 123 与音频编解码器 132 之间传输。音频处理流水线将数据输出到 A/V 端口 140,以供外置音频播放器或具有音频能力的设备再现。

[0076] 前面板 I/O 子部件 130 支持电源按钮 150 和弹出按钮 152 以及暴露在多媒体控制台 100 外表面上的任何 LED(发光二极管)或其它指示器的功能。系统供电模块 136 向多媒体控制台 100 的组件供电。风扇 138 冷却多媒体控制台 100 内的电路。

[0077] CPU 101、GPU 108、存储器控制器 110、以及多媒体控制台 100 内的各种其他组件经由一条或多条总线互连,该总线包括串行和并行总线、存储器总线、外围总线、以及使用各种总线体系结构中的任一种的处理器或局部总线。作为示例,这些架构可以包括外围部件互连(PCI)总线、PCI-Express 总线等。

[0078] 当多媒体控制台 100 通电时,应用数据可从系统存储器 143 加载到存储器 112 和 / 或高速缓存 102、104 中,并且可在 CPU 101 上执行。应用可呈现在导航到在多媒体控制台 100 上可用的不同媒体类型时提供一致的用户体验的图形用户界面。在操作中,介质驱动器 144 中所包含的应用和 / 或其他媒体可从介质驱动器 144 启动或播放,以将附加功能提供给多媒体控制台 100。

[0079] 多媒体控制台 100 可通过将该系统简单地连接到电视机或其他显示器而作为独立系统来操作。在该独立模式中,多媒体控制台 100 允许一个或多个用户与该系统交互、看电影、或听音乐。然而,在通过网络接口 124 或无线适配器 148 可用的宽带连接集成的情况下,多媒体控制台 100 还可作为更大网络社区中的参与者来操作。

[0080] 当多媒体控制台 100 通电时,可以保留设定量的硬件资源以供多媒体控制台操作系统作系统使用。这些资源可包括存储器的保留量(诸如,16MB)、CPU 和 GPU 周期的保留量(诸如,5%)、网络带宽的保留量(诸如,8kbs),等等。因为这些资源是在系统引导时间保留的,所保留的资源对于应用视角而言是不存在的。

[0081] 具体而言,存储器保留优选地足够大,以包含启动内核、并发系统应用程序和驱动程序。CPU 保留优选地为恒定,使得若所保留的 CPU 用量不被系统应用使用,则空闲线程将消耗任何未使用的周期。

[0082] 对于 GPU 保留,通过使用 GPU 中断来显示由系统应用程序生成的轻量消息(例如,弹出窗口),以调度代码来将弹出窗口呈现为覆盖图。覆盖图所需的存储器量取决于覆盖区域大小,并且覆盖图优选地与屏幕分辨率成比例缩放。在并发系统应用使用完整用户界面的情况下,优选使用独立于应用分辨率的分辨率。定标器可用于设置该分辨率,从而无需改变频率和引起 TV 重新同步。

[0083] 在多媒体控制台 100 引导且系统资源被保留之后,执行并发系统应用来提供系统功能。系统功能被封装在上述所保留的系统资源内执行的一组系统应用中。操作系统内核标识作为系统应用线程而非游戏应用线程的线程。系统应用优选地被调度为在预定时间并以预定时间间隔在 CPU 101 上运行,以便为应用提供一致的系统资源视图。调度是为了使在控制台上运行的游戏应用所引起的高速缓存分裂最小化。

[0084] 当并发系统应用需要音频时,由于时间敏感性而将音频处理异步地调度给游戏应用。多媒体控制台应用管理器(如下所述)在系统应用活动时控制游戏应用的音频水平(例如,静音、衰减)。

[0085] 输入设备(例如,控制器 142(1)和 142(2))由游戏应用和系统应用共享。输入设备不是保留资源,而是在系统应用和游戏应用之间切换以使其各自具有设备的焦点。应用管理器较佳地控制输入流的切换,而无需知晓游戏应用的知识,并且驱动程序维护有关焦点切换的状态信息。相机 26、28 和捕捉设备 20 可为控制台 100 定义附加输入设备。

[0086] 图 4 示出了可用于在目标识别、分析和跟踪系统中解释一个或多个姿势的计算环境 220 的另一示例实施例,该计算环境可以是图 1A-2 所示的计算环境 12。计算系统环境 220 只是合适的计算环境的一个示例,并且不旨在对当前公开的主题的使用范围或功能提出任何限制。也不应该将计算环境 220 解释为对示例性操作环境 220 中示出的任一组件或其组合有任何依赖性 or 要求。在某些实施例中,所描绘的各种计算元素可包括被配置成实例化本发明的各具体方面的电路。例如,本公开中使用的术语电路可包括被配置成通过固件或开关来执行功能的专用硬件组件。其他示例中,术语电路可包括由实施可用于执行功能的逻辑的软件指令配置的通用处理单元、存储器等。在电路包括硬件和软件的组合的示例实施例中,实施者可以编写体现逻辑的源代码,且源代码可以被编译为可以由通用处理单元处理的机器可读代码。因为本领域技术人员可以明白现有技术已经进化到硬件、软件或硬件/软件组合之间几乎没有差别的地步,因而选择硬件或是软件来实现具体功能是留给实现者的设计选择。更具体地,本领域技术人员可以明白软件进程可被变换成等价的硬件结构,而硬件结构本身可被变换成等价的软件进程。由此,对于硬件实现还是软件实现的选择是设计选择并留给实现者。

[0087] 在图 4 中,计算环境 220 包括计算机 241,计算机 241 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 241 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。系统存储器 222 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)223 和随机存取存储器(RAM)260。包含诸如在启动期间帮助在计算机 241 内的元件之间传输信息的基本例程的基本输入/输出系统 224(BIOS)通常储存在 ROM 223 中。RAM 260 通常包含处理单元 259 可立即访问和/或当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 225、应用程序 226、其他程序模块 227 和程序数据 228。

[0088] 计算机 241 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 示出了从不可移动、非易失性磁介质中读取或向其写入的硬盘驱动器 238,从可移动、非易失性磁盘 254 中读取或向其写入的磁盘驱动器 239,以及从诸如 CD ROM 或其他光学介质等可移动、非易失性光盘 253 中读取或向其写入的光盘驱动器 240。可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。硬盘驱动器 238 通常由例如接口 234 等不可移动存储器接口连接至系统总线 221,而磁盘驱动器 239 和光盘驱动器 240 通常由例如接口 235 等可移动存储器接口连接至系统总线 221。

[0089] 以上讨论并在图 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 241 提供了对计算机可读指令、数据结构、程序模块和其他数据的存储。在图 4 中,例如,硬盘驱动器 238 被示为存储操作系统 258、应用程序 257、其他程序模块 256 和程序数据 255。注意,这些组件可与操作系统 225、应用程序 226、其他程序模块 227 和程序数据 228 相同,也可与它们不同。在此操作系统 258、应用程序 257、其他程序模块 256 以及程序数据 255 被给予了不同的编号,以说明至少它们是不同的副本。用户可以通过输入设备,例如键盘 251 和定点设备 252——通常是指鼠标、跟踪球或触摸垫——向计算机 241 输入命令和信息。其他输入设备(未示出)可包括话筒、操纵杆、游戏手柄、圆盘式卫星天线、扫描仪等。这些以及其他输入设备通常通过耦合到系统总线的用户输入接口 236 连接到处理单元 259,但也可通过诸如并行端口、游戏端口或通用串行总线(USB)之类的其他接口和总线结构来连接。相机 26、28 和捕捉设备 20 可为控制台 100 定义附加输入设备。监视器 242 或其他类型的显示设备也通过诸如视频接口 232 之类的接口连接至系统总线 221。除监视器之外,计算机还可以包括可以通过输出外围接口 233 连接的诸如扬声器 244 和打印机 243 之类的其他外围输出设备。

[0090] 计算机 241 可使用到一个或多个远程计算机(诸如,远程计算机 246)的逻辑连接而在联网环境中操作。远程计算机 246 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其他常见网络节点,并且通常包括许多或所有以上关于计算机 241 所描述的元件,但在图 4 中仅示出了存储器存储设备 247。图 2 中所描绘的逻辑连接包括局域网(LAN)245 和广域网(WAN)249,但还可包括其他网络。此类联网环境在办公室、企业级计算机网络、内联网和因特网中是常见的。

[0091] 当用于 LAN 网络环境中时,计算机 241 通过网络接口或适配器 245 连接到 LAN 237。当在 WAN 联网环境中使用时,计算机 241 通常包括调制解调器 250 或用于通过诸如因特网等 WAN 249 建立通信的其他手段。调制解调器 250 可以是内置的或外置的,可经由用户输入接口 236 或其他适当的机制连接到系统总线 221。在联网环境中,相对于计算机 241 所示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 示出了远程应用程序 248 驻留在存储器设备 247 上。应当理解,所示的网络连接是示例性的,并且可使用在计算机之间建立通信链路的其他手段。

[0092] 计算机可读存储介质可以包括用于修改视觉表示的计算机可读指令。该指令可包括用于呈现视觉表示、接收场景的数据、以及基于用户的修改姿势来修改视觉表示的指令,其中该数据包括表示物理空间中用户的修改姿势的数据,该修改姿势是映射到控件的姿势,该控件用于修改视觉表示的特性。

[0093] 图 5 描绘了可从捕捉设备 20 所捕捉的图像数据生成的用户的示例骨架映射。在该实施例中,标识出各个关节和骨骼:每一手 502、每一前臂 504、每一肘 506、每一二头肌 508、每一肩 510、每一髋 512、每一大腿 514、每一膝 516、每一小腿 518、每一足 520、头 522、躯干 524、脊椎的顶部 526 和底部 528,以及腰 530。在跟踪更多点的情况下,可标识出附加的特征,比如手指或脚趾的骨骼和关节,或面部的各个特征,如鼻和眼。

[0094] 用户可通过移动他的身体创建姿势。姿势包括用户的运动或姿态,其可被捕捉为图像数据并解析其意义。姿势可以是动态的,包括运动,如模仿投球。姿势可以是静态姿态,如在一个人的躯干 524 前面交叉握住他的前臂 504。姿势也可结合道具,如通过挥动仿制的剑。姿势可包括多于一个身体部位,如拍双手 502,或是较微小的运动,如撅起一个人的嘴唇。

[0095] 用户的姿势可用作一般计算上下文中的输入。例如,手 502 或其他身体部位的各种运动可对应于常见的系统级任务,如在分层列表中向上或向下导航、打开文件、关闭文件和保存文件。例如,用户能以手指向上指且掌心面向捕捉设备 20 来使他的手保持不动。他然后可以将手指朝向手掌收拢来形成拳头,并且这可以是指示基于窗口的用户界面计算环境中的焦点窗口应被关闭的姿势。姿势也可在视频游戏专用上下文中取决于游戏来使用。例如,对于驾驶游戏,手 502 和脚 520 的各种运动可对应于在一方向上操控车辆、换挡、加速和刹车。由此,姿势可指示映射到所显示的用户表示的、在诸如视频游戏、文本编辑器、文字处理、数据管理等各种各样应用中的各种各样的运动。

[0096] 用户可通过自己在原地行走或奔跑来生成对应于行走或奔跑的姿势。例如,用户可另选地提起并放下每一腿 512-520 来在不移动的情况下模拟行走。系统可通过分析每一髋 512 和每一大腿 514 来解析该姿势。当一个髋部-大腿角(如相对于垂直线测量的,其中站立的腿具有 0° 的髋部-大腿角,而向前水平伸展的腿具有 90° 的髋部-大腿角)超过相对于另一大腿的特定阈值时,可识别一步。行走或奔跑可在交替腿的某一数量的连续步之后被识别。两个最近的步之间的时间可被认为是一周期。在不满足阈值角度达某一数量的周期之后,系统可确定行走或奔跑姿势已停止。

[0097] 给定“行走或奔跑”姿势,应用可为与该姿势相关联的参数设定值。这些参数可包括上述阈值角度、发起行走或奔跑姿势所需的步数、结束姿势的没有发生步的周期数、以及确定姿势是行走还是奔跑的阈值周期。快周期可对应于奔跑,因为用户将快速地移动他的腿,而较慢的周期可对应于行走。

[0098] 姿势最初可以与一组默认参数相关联,应用可用其自己的参数来覆盖该组默认参数。在这一场景中,不迫使应用提供参数,而是应用可改为使用一组允许在没有应用定义的参数的情况下识别姿势的默认参数。与姿势有关的信息可被存储用于预录制的动画的目的。

[0099] 存在可以与姿势相关联的各种输出。可以有關於姿势是否正在发生的基线“是或否”。还可以有置信度水平,其对应于用户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应用中,它可仅使用那些具有高置信度水平,如至少 0.95 的已识别的姿势。在应用必须识别姿势的每一实例的情况下,即使以假肯定为代价,它可使用至少具有低得多的置信度水平的姿势,如仅仅大于 0.2 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。

[0100] 另一示例性姿势是“脚跟提起跳”。在该姿势中,用户可通过将他的脚跟提离地面,但保持他的脚趾着地来创建该姿势。另选地,用户可跳向空中,其中他的脚 520 完全离开地面。该系统可通过分析肩 510、髌 512 和膝 516 的角度关系来解析该姿势的骨架,以查看它们是否是等于直立的对齐位置。然后,可监视这些点和较高 526 和较低 528 脊椎点来发现任何向上加速。足够的加速组合可触发跳跃姿势。加速度与某一姿势的足够的组合可满足转变点的参数。

[0101] 给定该“脚跟提起跳”姿势,应用可为与该姿势相关联的参数设定值。参数可包括上述加速阈值,其确定用户的肩 510、髌 512 和膝 516 的某种组合必须向上移动多快来触发该姿势;以及肩 510、髌 512 和膝 516 之间仍可触发跳跃的最大对齐角。输出可包括置信度水平,以及用户在跳跃时的身体角度。

[0102] 基于将接收姿势的应用的细节来为姿势设定参数对于准确地标识姿势而言是重要的。正确地标识姿势以及用户的意图极大地有助于创建积极的用户体验。

[0103] 应用可以为与各种转变点相关联的参数设置值来标识使用预录制的动画的点。转变点可由各种参数来定义,如某一姿势的标识、速度、目标或物体的角度、或其任何组合。如果转变点至少部分地由某一姿势的标识来定义,则正确地标识姿势有助于提高转变点的参数已被满足的置信水平。

[0104] 对于姿势的另一参数可以是移动的距离。在用户的姿势控制虚拟环境中的化身的动作的情况下,该化身可以是手臂离球的长度。如果用户希望与该球交互并抓住它,则这可要求用户伸展他的手臂 502-510 到全长同时作出抓握姿势。在这一情形中,用户仅部分地伸展他的手臂 502-510 的类似的抓握姿势可能无法达到与球交互的结果。类似地,转变点的参数可以是对抓握姿势的标识,其中如果用户仅部分地伸展他的手臂 502-510,从而没有实现与球交互的结果,则用户的姿势将不满足转变点的参数。

[0105] 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 510a、且与投掷手臂 502a-310a 在头 522 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场景的边缘。

[0106] 图 6A 和 6B 描绘了可包括捕捉设备 608、计算设备 610 和显示设备 612 的系统 600。例如,捕捉设备 608、计算设备 610 和显示设备 612 的每个可包括执行所需功能的任何合适的设备,诸如参考图 1-5 所描述的设备。构想了单个设备可以执行系统 600 中的全部功能,或者合适的设备的任何组合可以执行所需的功能。例如,计算设备 610 可以提供参考图 2 所示的计算环境 12 或图 3 中的计算机所描述的功能。如图 2 所示,计算环境 12 可包括显示设备和处理器。计算设备 610 还可包括其自身的相机组件,或者可以耦合到具有相机组件的设备,诸如捕捉设备 608。

[0107] 在这些示例中,深度相机 608 捕捉用户 602 所存在于的物理空间 601 中的场景。深度相机 608 处理深度信息,和 / 或向诸如计算机 610 等计算机提供深度信息。深度信息可被解释以便显示用户 602 的视觉表示。例如,深度相机 608 或如所示的其耦合到的计算设

备 610 可向显示器 612 输出。所捕捉并显示的图像数据帧的速率可确定视觉表示的所显示的运动的连续性水平。虽然可以捕捉并显示附加图像数据帧,但图 6A 和 6B 中描绘的帧是出于示例性的目的来选择的。还注意到,视觉表示可以是物理空间 601 中另一个目标的视觉表示,诸如另一个用户或非人类物体,或者视觉表示可以是部分的或完整的虚拟物体。

[0108] 技术在此公开了系统自动生成目标的视觉表示的能力,该视觉表示具有与目标所检测到的特征相类似的特征。或者,系统可提供用户可从中进行选择的可选择特征子集。系统可基于目标所检测到的特征来选择特征,并将该选择应用于目标的视觉表示。或者,系统可以作出选择,该选择缩减了用户从中选择的选项的数量。如果系统可以代表用户作出决定,则用户可以无需作出很多决定或不从很多选项中进行选择。由此,所公开的技术可从用户移除大量的工作。例如,系统可代表用户作出选择,并将它们应用于用户的视觉表示。

[0109] 如图 6A 所示,系统呈现与物理空间 601 中的用户 602 相对应的视觉表示 603。在该示例中,系统通过检测用户 602 的特征,将所检测到的特征与特征选项库进行比较,选择类似于用户 602 所检测到的特征的特征选项,以及自动地将其应用于用户的视觉表示 603 来自动生成视觉表示 603。视觉表示的自动生成从用户 602 移除了工作,并为用户 602 创建如同他们不费力地被传送到游戏或应用体验中那样的不可思议的体验。

[0110] 还公开了用于实时地显示视觉表示并且实时地更新被应用到视觉表示的特征选择的技术。系统可以随着时间跟踪物理空间中的用户,并同样实时地应用修改或更新被应用于视觉表示的特征。例如,系统可以跟踪用户并标识用户已经脱下了运动衫。系统可以标识用户的身体移动,并识别用户的服装类型和颜色的变化。系统可以使用用户的所标识的特性中的任一个来帮助特征选择过程和 / 或更新从特征库中所选并应用于视觉表示的特征。由此,同样,系统可以不费力地将用户传送到应用体验中,并且更新视觉表示以便在用户的所检测到的特征改变时与其实时地对应。

[0111] 在示例实施例中,为了检测用户的特征并使用所检测到的特征来选择用于视觉表示的特征的选项,系统可以生成用户的模型。为了生成模型,捕捉设备可以捕捉场景的图像并扫描场景中的目标或物体。根据一个实施例,图像数据可包括来自深度相机 608 和 / 或 RGB 相机的深度图像或图像,或者任何其他检测器上的图像。系统 600 可从场景捕捉深度信息、图像信息、RGB 数据等。为了确定场景中的目标或物体是否对应于人类目标,可对每个目标进行泛色填充并将其与人体模型的模式作比较。可扫描与人类模式 (pattern) 相匹配的每个目标或物体以生成模型,诸如骨架模型、泛色模型、网格人类模型、或与其相关联的模型。随后可将该骨架模型提供给计算环境以便跟踪该骨架模型并呈现与该骨架模型相关联的化身。

[0112] 图像数据和 / 或深度信息可用于标识目标特征。关于人类目标的这些目标特征可包括,例如身高和 / 或臂长,并且可以基于例如身体扫描、骨架模型、用户 602 在像素区域上的延伸范围或任何其他合适的过程或数据来获得。使用例如与人类目标相关联的多个观察到的像素中的深度值和人类目标的一个或多个方面的延伸范围,如身高、头宽、或肩宽等,可确定人类目标的大小。相机 608 可以处理图像数据,并将其用于确定用户的各个部位的形状、颜色和大小,包括用户的头发、服装等。所检测到的特征可以与用于应用于视觉表示的特征选项的目录进行比较,如特征库 197 中的视觉表示特征选项。

[0113] 在另一示例实施例中,为了标识用户的特性,并使用所标识的特性来选择用于视觉表示的特征,系统可使用目标数字化技术,诸如参考图 2B 所描述的那些技术。该技术包括从得自诸如深度传感设备等捕捉设备的未经组织的点云中标识表面、纹理和物体尺寸。采用目标数字化可包括表面提取、标识点云中的点、标记表面法线、计算物体属性、随时间跟踪物体属性中的变化、以及随着捕捉额外的帧而增加物体边界和身份中的置信度。例如,可以接收或观察与物理空间中的物体相关的数据点的点云。然后可分析该点云来确定该点云是否包括物体。点云的集合可被标识为物体并融合在一起来表示单个物体。点云的表面可从所标识的物体中提取。

[0114] 可使用提供扫描已知 / 未知物体、扫描人类、以及扫描场景中的背景方面(例如,地板、墙)的能力的任何已知技术或此处公开的技术来检测物理空间中的目标的特征。可使用每一对象的所扫描的数据,包括深度和 RGB 数据的组合,来创建该物体的三维模型。该 RGB 数据被应用于该模型的对应区域。各帧之间的时间跟踪可以提高置信度并实时地自适应物体数据。由此,可使用物体属性以及随时间对物体属性中的变化的跟踪来可靠地跟踪其位置和定向实时地在各帧之间变化的物体。捕捉设备以交互式速率来捕捉数据,从而提高了数据的保真度,并允许所公开的技术处理原始深度数据,数字化场景中的物体,提取物体的表面和纹理,以及实时地执行这些技术中的任一种,以使得显示画面可以提供该场景的实时描绘。

[0115] 可使用相机识别技术来确定特征库 197 中的哪些元素最接近地类似于用户 602 的特性。系统可使用脸部识别和 / 或身体识别技术来检测用户 602 的特征。例如,系统可以基于从图像数据、点云数据、深度数据等中对模型的生成来检测用户的特征。可以进行脸部扫描,并且系统可以处理关于用户的脸部特征和 RGB 数据所捕捉的数据。在一示例实施例中,基于五个关键数据点(即,眼、嘴角点、和鼻)的位置,系统提出关于玩家的脸部建议。该脸部建议可包括至少一个所选的脸部特征、完整的脸部特征集,或者它可以是来自特征库 197 的关于脸部特征的经缩减的选项子集。系统可以执行身体识别技术,从而从身体扫描中标识各种身体部位 / 类型。例如,用户的身体扫描可提供关于用户的身高的建议。对于这些扫描中的任一个,可提示用户站在物理空间中提供最佳扫描结果的位置。

[0116] 可从所捕捉的数据中检测其他特征。例如,系统可以通过分析用户和 / 或用户的模型来检测颜色数据和服装数据。该系统可基于这些用户特性的标识来为用户推荐服装。服装推荐可基于用户的衣柜中的服装,或来自可供在虚拟世界市场中购买的服装。例如,用户可能具有个人衣柜,该衣柜具有特定虚拟表示所拥有且关联的商品的储存库。个人衣柜可包括允许用户查看并修改应用于用户的虚拟表示的服装和其他商品的界面。例如,可修改饰品、鞋子等。用户的性别可基于所捕捉的数据或作为访问与用户相关联的简档的结果来确定。

[0117] 系统可以检测用户的特征中的至少一个,并从特征库 197 中选择代表所检测到的特征的特征。系统可以将所选择的特征自动应用于用户的视觉表示 603。由此,用户的视觉表示 603 具有如系统所选择的用户的相像性。例如,特征提取技术可以映射用户的面部特征,并且从特征库选择的特征选项可用于创建用户的卡通表示。视觉表示 603 是使用从特征库选择的类似于用户的所检测到的特征的特征来自动生成的,但在该示例中,视觉表示是用户 602 的卡通版本。视觉表示具有用户 602 的头发、眼睛、鼻子、服装(例如,牛仔裤、

夹克、鞋)、身体位置和类型等的卡通版本。系统可以将视觉表示 603 呈现给用户 602, 该视觉表示 603 是通过应用特征并呈现自动生成的视觉表示 603 来创建的。用户 602 可以修改自动生成的视觉表示 603, 或者继续作出应用到视觉表示的选择。

[0118] 在物理空间 601 中检测到的用户的视觉表示还可以采取替换的形式诸如动画、人物、化身等。图 6B 中示出的示例视觉表示是猴子角色 605 的视觉表示。用户 602 可以从系统或应用所提供的多个库存模型中选择以供对用户的屏幕上的表示。例如, 在棒球游戏应用中, 用于可视地表示用户 602 的可用的库存模型可包括将知名棒球运动员表示为一块太妃糖, 或者将大象表示为想象的角色或符号, 诸如光标或手部符号。在图 6B 所示的示例中, 猴子角色 605 可以是系统或应用所提供的库存模型表示。库存模型可以专用于应用, 诸如与程序包装在一起, 或者库存模型可以是跨应用可用或系统范围内可用的。

[0119] 视觉表示可以是用户 602 的特征与动画或库存模型的组合。例如, 猴子表示 605 可以从猴子的库存模型来初始化, 但是猴子的各个特征可以根据系统 600 从特征选项目录所选择的(诸如特征库 197 中的那些)类似于该用户的特征来修改。系统可以用库存模型来初始化视觉表示, 但随后继续检测用户的特征, 将所检测到的特征与特征库 197 进行比较, 选择类似于用户的特征, 并且将所选择的特征应用于猴子角色 605。由此, 猴子 605 可以具有猴子的身体, 但具有用户的面部特征, 诸如眉毛、眼睛和鼻子。用户的面部表情、身体位置、所说的单词、或任何其他可检测的特性可被应用于虚拟猴子 605, 并且如果合适的话可被修改。例如, 用户正在物理空间中皱眉。系统检测该面部表情, 从特征库中选择最接近地类似于用户的皱眉的皱眉, 并且将所选择的皱眉应用于猴子, 使得虚拟猴子也在皱眉。此外, 猴子坐在与用户类似的位置, 除非被修改以对应于该位置中的猴子的身体类型和大小。系统 600 可以将所检测到的目标身体类型特征与特征库 197 进行比较, 该特征库 197 存储身体类型的可能的视觉表示特征的集合。系统可以从特征库的猴子特征子集中选择特征。例如, 应用可以提供特征库中的猴子专用的特征选项以便与和应用预封装在一起的库存模型猴子角色选项相对应。系统或用户可从最接近地类似于用户的所检测到的特征的猴子专用特征的选项中进行选择。

[0120] 期望的是系统提供来自特征库 197 的特征子集。例如, 特征库 197 中一个以上选项可以类似于用户的所检测到的特征。系统可提供用户从中进行选择的小的特征子集。代替用户手动地从数十、数百、甚至数千特征选项中选择, 系统可以提供经缩减的选项子集。例如, 图 7 描绘了图 6A 和 6B 中所示的系统 600。系统在显示器 612 上显示用于视觉表示的头发的示例特征选项集, 选项 1-10。在图 6A 中, 系统自动地选择被应用于用户的视觉表示的头发选项 #5。然而, 在图 7 所示的示例中, 系统已经选择了最接近地类似于用户的所检测到的头发特征的头发选项的子集 702。由此, 用户可从被应用于用户的视觉表示的选项子集 702 中选择。

[0121] 在该示例中, 头发的特征选项的子集 702 可包括最接近地类似于从身体和面部扫描所检测到的用户特征的选择, 包括用户的头发形状、颜色和类型。代替从中选择的大量的头发选项, 系统可以提供最接近地类似于用户的头发形状、颜色和类型的用于头发选项的短的选项列表。系统可以自动生成视觉表示, 但也可被设计为提供用户可从中选择一个以上选项, 使得用户可在最能取悦用户的特征选项之间作出最终的详细选择。选项子集减少了用户评估全部选项的需求。

[0122] 用户或应用可以具有用于在将对应于用户特性的特定特征应用于视觉表示之前对其进行修改的设置。例如,系统可以基于所捕捉的数据(例如,身体类型/大小)来检测用户的特定体重范围。然而,用户可以设置或者应用本身可以设置默认值,使得在特定体重范围而非实际的用户体重范围以内显示用户。由此,可为用户显示更令人愉悦的视觉表示,而不是例如可能超重的视觉表示。在另一个示例中,可以检测用户的面部特征,并且应用于用户的视觉表示的特征可以对应于所检测到的特征,使得视觉表示的面部特征在大小、比例、头上的空间安排等方面类似于用户的特征。用户可以通过改变特征来修改面部识别技术的真实效果。例如,用户可以通过改变计算尺来修改特征。用户可以改变计算尺以修改体重以便应用于视觉表示,或者改变将被应用于视觉表示的鼻子的大小。由此,可以应用系统所选择的某些特征,可以修改并随后应用其他特征。

[0123] 出于显示目的,可以修改系统所检测到的特定目标特性。例如,目标特性可被修改为对应于视觉表示的形式、应用、应用的状态等。例如,特定的特性可能不直接映射到用户的视觉表示,其中该视觉表示是想象的角色。诸如化身 603 或诸如猴子 605 的用户的角色表示等用户的任何视觉表示可被赋予例如类似于用户 602 的但为特定角色修改过的身体比例。例如,猴子表示 605 可被赋予类似于用户 602 的身高,但猴子的手臂可能比用户的手臂成比例地更长。猴子 605 的手臂的移动可以对应于用户手臂的移动,如系统所标识的,但是系统可以修改猴子手臂的动画以反应猴子手臂移动的方式。

[0124] 系统可以使用诸如扫描数据、图像数据或深度信息等所捕捉的数据来标识其他目标特性。目标特性可包括目标的任何其他特征,诸如:眼睛大小、类型和颜色,头发长度、类型和颜色,肤色,服装和服装颜色。例如,可以基于对应的 RGB 图像来标识颜色。系统还可将这些可检测的特征映射到视觉表示。例如,系统可以检测到用户佩戴眼镜以及穿一件红色衬衫,系统可以将眼镜和红色衬衫应用于在该示例中是用户的视觉表示的虚拟猴子 605。

[0125] 深度信息和目标特性还可与附加信息进行组合,该附加信息包括例如可与特定用户 602 相关联的信息,诸如具体的姿势、语音识别信息等。该模型随后可被提供给计算设备 610,使得计算设备 610 可跟踪该模型,呈现与该模型相关联的视觉表示,和/或基于例如该模型来确定在计算设备 610 上执行的应用中执行哪些控件。

[0126] 图 8 示出了向用户提供特征选择的示例方法。特征选择的提供可以通过显示视觉表示来提供,其中该视觉表示具有所应用的特征或用户可从中进行选择的具有经缩减的选项子集的特征库的子集。例如,在 802,系统从包括诸如用户或非人类物体等目标的物理空间接收数据。

[0127] 如上所述,捕捉设备可捕捉场景的数据,诸如场景的深度图像,并扫描场景中的目标。捕捉设备可以确定场景中的一个或多个目标是否对应于诸如用户等人类目标。例如,为了确定场景中的目标或物体是否对应于人类目标,可对每个目标进行泛色填充并将其与人体模型的图案作比较。然后可扫描匹配人体模型的每个目标或物体来生成与其相关联的骨架模型。例如,可扫描被标识为人类的目标来生成与其相关联的骨架模型。然后可将该骨架模型提供给计算环境来跟踪该骨架模型并呈现与该骨架模型相关联的视觉表示。在 804,系统可通过使用任何合适的技术,如身体扫描、点云模型、骨架模型、泛色填充技术等,来转换所捕捉的数据,以便标识物理空间中的目标的特征。

[0128] 在 806,系统可检测目标的特性并将其与诸如特征库中的特征选项等特征选项进

行比较。特征选项可以是用于目标的各种特征的选项的集合。例如,用于用户的特征选项可包括眉毛选项、头发选项、鼻子选项等。用于房间中的家具的特征选项可包括大小选项、形状选项、硬件选项等。

[0129] 在一示例实施例中,系统可以检测可供应用于类似于用户的所检测到的特征的视觉表示的若干特征。由此,在 806,系统可以检测用户的特征并将所检测到的特征与应用于用户的视觉表示的特征库 197 进行比较,并且在 810,系统可以基于所检测到的特征来选择特征选项的子集。系统可以通过比较特征库 197 中的特征与用户的所检测到的特性的相似度来选择该子集作为那些特征。有时,特征将非常相似,但是系统可能仍然在 810 向用户提供从中进行选择的选项子集。以此方式,用户可从该子集中选择至少与用户的对应特性相似的特征,但是可例如从该子集中选择更令人愉悦的特征。系统可在 812 接收用户从子集选项中进行的选择。由此,用户不必过滤具体特征的整个选项库来寻找与用户相似的特征。系统可以过滤选项库并向用户提供从中进行选择的特征子集。

[0130] 系统可以在 814 自动生成用户的视觉表示。由此,在将目标的所检测到的特征与特征库中的选项进行比较之后,系统可以通过自动选择要应用于视觉表示的特征来自动生成目标的视觉表示。当系统自动呈现对应于用户的视觉表示,自动从特征库中选择了类似于目标的所检测到的特征的特征时,目标被不费力地传送到系统或软件体验中。

[0131] 该视觉表示可具有自动选择的特征和用户基于系统所提供的选项子集选择的特征的组合。由此,该视觉表示可被部分地生成且部分地由用户来定制。

[0132] 在 816,可将系统和 / 或用户作出的选择应用于目标的视觉表示。系统可以向用户呈现该视觉表示。在 818,系统可以继续监视物理空间中的目标,从而随时间跟踪目标的可检测特征。对目标的视觉表示的修改可以实时地作出来反映对目标的所检测到的特征的任何改变。例如,如果目标是用户且该用户在物理空间中脱下了运动衫,则系统可以检测到新的衬衫样式和 / 或颜色,并从特征库中自动选择接近地类似于用户的衬衫的选项。

[0133] 所选选项可实时地应用于用户的视觉表示。由此,前述步骤中的处理可以实时地执行,使得显示画面实时地对应于物理空间。以此方式,物理空间中的物体、用户或运动可被转换以供实时显示,使得用户可以实时地与正在执行的应用进行交互。

[0134] 在 822,用户的所检测到的特征、系统所选择的特征、以及用户所选择的任何特征可成为简档的一部分。简档可以例如专用于某一物理空间或用户。包括用户的特征在内的化身数据可以成为用户的简档的一部分。可以在用户进入捕捉场景时访问简档。如果基于口令、用户的选择、身体大小、语音识别等,简档匹配用户,则该简档可用于确定用户的视觉表示。可监视用户的历史数据,从而将信息存储到用户的简档。例如,系统可以检测特定于用户的特征,如用户的面部特征、身体类型等。系统可以选择类似于所检测的特征的特征以便应用于目标的视觉表示并存储在目标简档中。

[0135] 图 9 描绘了来自图 6 的系统 600 的示例,该示例可以处理为物理空间 601 中的目标接收的信息,并且使用目标数字化技术来标识目标。所捕捉的目标可被映射到虚拟环境中那些目标的视觉表示。在该示例中,物理场景包括图 1A 所描绘的物理空间中示出的球 102、箱子 104、遮光帘 106、靠墙扶手 108、墙 #1110、墙 #2112、以及地板 115。场景中还示出了用户 602。在示例实施例中,系统 10 可以识别、分析和 / 或跟踪这些物体 102、104、106、108、110、112 和 115 中的任一个以及其他目标,如人类目标,诸如用户 602。系统 10 可在物理空

间中收集与物体 102、104、106、108、110、112 和 114 中每个和 / 或用户 602 的姿势相关的信息。诸如用户 602 的物理空间中的用户也可以进入物理空间。

[0136] 目标可以是物理空间 601 中的任何物体或用户。例如,捕捉设备 608 可以扫描物理空间 601 中的人类 602 或非人类物体,诸如球 607、纸板箱 609、或狗 605。在该示例中,系统 600 可以通过使用捕捉设备 608 来扫描物理空间 601 以便捕捉目标。例如,深度相机 608 可以接收原始深度数据。系统 600 可以处理原始深度数据,将深度数据解释为点云数据(point cloud data),将点云数据转换为表面法线。例如,深度缓冲区可被捕捉并转换为已排序的点云。

[0137] 深度缓冲区可以是记录被呈现的每个像素的深度的缓冲区。深度缓冲区可以在附加像素被呈现时保持对其的记录,并确定所呈现的不同像素的深度之间的关系。例如,深度缓冲区可以执行隐藏表面移除,并且将要被呈现的每个像素与已经位于帧缓冲区中该位置的像素进行比较。也被称为 z 缓冲区的深度缓冲区可包括帧缓冲区,该帧缓冲区存储对从捕捉设备到所捕捉的图像中每个可视点的距离的测量。

[0138] 基于所标识的点云和表面法线,系统 600 可以标记在场景中被解析的物体,清除噪声,并为每个物体计算定向。可在物体周围形成边界框。随后可逐帧地跟踪该物体以供纹理提取。

[0139] 根据一个实施例,图像数据可包括深度图像或来自深度相机和 / 或 RGB 相机的图像,或者任何其他检测器上的图像。例如,相机 608 可处理图像数据,并使用它来确定目标的形状、颜色和大小。在该示例中,物理空间 601 中的目标 602、102、104、106、108、110、112 和 114 被深度相机 608 捕捉,该深度相机 608 处理深度信息和 / 或将深度信息提供给计算机,诸如计算机 610。

[0140] 深度信息可被解释以便在显示器 612 上显示视觉表示。系统可使用信息从特征库 197 选择选项以生成与物理空间中的目标相对应的虚拟物体。可扫描与人类样式(pattern)相匹配的每个目标或物体以生成与其相关联的模型,诸如骨架模型、网格人类模型等。可以扫描与已知物体库相匹配的每个目标或物体以生成可用于该特定物体的模型。还可以扫描未知物体以生成模型,该模型对应于点云数据、RGB 数据、表面法线、定向、边界框、以及对与该未知物体相对应的原始深度数据的任何其他处理。

[0141] 由于目标可以在物理空间中移动,因此所捕捉并显示的图像数据帧的速率确定对视觉表示的显示的连续性水平。此外,逐帧图像的数量可以随时间提高点云数据被解析为单独标记的物体的方式的置信度。物体的移动可以给出关于表面法线和定向的进一步的深度信息。系统 600 还可以能够将噪声与所需点数据区分开。系统 600 还可通过评估用户在单个捕捉数据帧中或在一系列帧中的位置来从用户 602 的运动中标识姿势。

[0142] 系统 600 可在物理空间 601 中跟踪目标 602、102、104、106、108、110、112 和 114 中的任一个,使得显示器 612 上的视觉表示映射到目标 602、102、104、106、108、110、112 和 114 以及物理空间 601 中所捕捉的那些目标中任一个的运动。物理空间中的物体可以具有捕捉设备可捕捉并扫描以与诸如图 2 所示的特征库 197 等特征库中的特征选项进行比较的特性。系统可从特征库中选择特征,该特征最接近地类似于目标的所检测到的特征。

[0143] 此处公开了用于与目标数字化的实现有关的计算机视觉的技术。这些技术可用于使系统能够将高保真地捕捉的特征与类似于目标特征的来自特征库的最佳选择特征进行

比较。计算机视觉是通过根据诸如原始深度或图像数据等所捕捉的数据在物理空间中创建物体模型来理解场景内容的概念。例如,技术可包括表面提取,基于邻近度解释点云中的点以恢复表面法线,计算物体属性,随时间跟踪该物体属性,随时间提高物体识别和形状中的置信度,并扫描人类或已知/未知物体。

[0144] 捕捉设备可以扫描物理空间,并且接收关于物理空间 601 中各个物体的范围数据。扫描可包括对物体表面的扫描、或对整个固体的扫描。通过对原始深度数据采用二维深度缓冲区的形式,任何适合的计算设备可以解释物体表面上的大量的点并输出点云。点云可以是在三维坐标系统中定义的数据点集,诸如由 x、y 和 z 坐标定义的数据点。点云数据可以表示物理空间中已被扫描过的物体的可视表面。由此,物体可以通过将场景中的物体表示为离散点集来数字化。点云数据可在数据文件中被保存为二维数据集。

[0145] 可以使用诸如深度相机或深度传感设备等捕捉设备来实时地捕捉范围数据。例如,可以使用深度缓冲区形式的深度传感相机以至少 20 赫兹的频率捕捉数据帧。数据可被解释为结构化的样本点云,其中每个点可包括相关联的目标的特性,诸如位置、定向、表面法线、颜色或纹理属性。点云数据可被存储在二维数据集中。由于捕捉设备的光学属性是已知的,因此范围数据可被投影到可由此被存储在正则化数据结构中的完全三维的点云。三维点云可以指示物体表面的拓扑结构。例如,可以从云中的邻居点确定表面的邻近点之间的关系。可将点云数据转换为表面,并且可以通过评估点云数据表面上的表面法线来提取点云数据所表示的物体表面。正则化数据结构可以类似于二维深度缓冲区。

[0146] 点云可包括与物理空间中各个物体相关的多个数据点。捕捉设备可以接收或观察点云数据,诸如此处所描述的。随后可分析该点云以确定该点云是否包括一个物体或一组物体。如果数据包括一个物体,则可以生成该物体的模型。物体识别中置信度的提高可在捕捉帧时发生。可以生成与特定物体相关联的模型的反馈并将其实时地提供给用户。此外,响应于物理空间中物体的任何移动,可以跟踪该物体的模型,使得该模型可被调整为模仿该物体的移动。

[0147] 所有这些可以以用于处理以及实时显示结果的速率来完成。实时显示指的是姿势的视觉表示的显示或视觉帮助的显示,其中该显示与该姿势在物理空间中的执行同时或几乎同时显示。例如,系统可以提供对用户和该用户环境进行回应的显示的显示更新速率可以是 20Hz 的速率或更高,其中无关紧要的处理延迟造成最小的显示延迟或对用户而言根本不可见。由此,实时包括与被自动数据处理所需的时间延迟了的数据的及时性有关的任何无关紧要的延迟。

[0148] 捕捉设备以交互式速率来捕捉数据,提高了数据的保真度,并允许所公开的技术处理原始深度数据,数字化场景中的物体,提取物体的表面和纹理,以及实时地执行这些技术中的任一种,以使得显示画面可以提供该场景的实时描绘。对于任何给定帧为了将云中的点组聚类成场景中的离散物体,可以使深度缓冲区在扫描线中从左到右并随后从上到下的走一遍。可在扫描时处理云中的每个对应的点或点集群。

[0149] 相机可以捕捉深度和颜色数据,并将颜色分配给与该颜色数据相对应的点云。由此,在捕捉设备从相机的观点查看深度数据时,相机可以解释该深度数据以用三维来表示物理空间。三维点云数据可被合并和结合,使得多个点变为点云,并且云中点的子集可被标记为特定的物体。可以从该被标记的点云为所创建的每个被标记的物体和对应的网格模型

恢复三维数据。由于颜色信息与深度信息相关,物体的纹理和表面也可被提取。这种目标数字化可能对游戏应用或非游戏应用是有用的,诸如操作系统或软件应用。在显示设备上提供关于捕捉和处理深度数据的实时的反馈提供了有价值的交互式体验,诸如玩游戏。

[0150] 在图 8 所描绘的示例中,墙、天花板和地板都位于物理空间中。系统可以从对点云数据的分析来标记墙和地板,该点云数据来自处理捕捉设备所接收的原始深度数据,诸如图 7B 中所表示的点云数据。随后,可以提取关于物理场景的附加信息,诸如房间的形状。系统可以使用关于物理空间的基本信息来从特征库中进行选择以生成与物理空间相对应的虚拟空间。例如,特征库可包括各个特征的卡通绘图,而因此自动生成的虚拟空间可以是物理空间的卡通版本。However, the cartoon version

[0151] 深度缓冲区中的信息可用于分离从原始深度数据标识的物体的表面。深度缓冲区的第一遍行走可用于基于从点云得出的表面法线来计算深度缓冲区的法线映射图。由此,系统可以得出表面指向的方向,而不是空间中独立的点。系统可以从深度缓冲区恢复表面法线,并将表面法线和与该表面法线相关联的云中的多个点一同存储。表面法线可用于标识物体的形状和轮廓。例如,球在整个表面上的各法线方向上具有逐渐恒定的变化。各个物体的表面法线可在各个物体过滤器中有所不同以便与场景中所检测到的表面法线进行比较。

[0152] 虽然对表面法线的计算和普通映射计算是此处所公开的用于从点云数据标识表面的常见技术,但是可以使用任何合适的表面分离或提取技术,诸如霍夫变换、普通映射、傅立叶变换、小曲线变换等。例如,用于从点云分离和 / 或提取表面的计算可以使用用于平坦表面的霍夫变换来实现。在这种实例中普通映射将是不必要的,相反可以生成点云的霍夫变换。由此,当云的多个点被合并成物体并被标记时,对每个点的霍夫空间的评估可以指示一个点是否与邻居点一起位于平面上,使系统能够分别地标记组成特定物体的具体的平坦表面。可以使用任何合适的分离 / 提取技术,并且可以依赖于场景将其调节至整体标记性能和特性。虽然使用各种表面分离 / 提取技术可以改变标记试探,但任何合适的技术都用于这种标识和标记,并仍然使系统能够实时地处理深度数据以便生成和刷新对用户的实时显示。

[0153] 噪声可以产生自所使用的深度传感器的类型。第一行走阶段可包括原始数据的噪声抑制轮。例如,可以执行平滑轮以从普通映射移除噪声。

[0154] 针对数据集在二维扫描轮中对云中的多个点进行标记,其中接近并已标识出相似表面的选项可被标记为属于相同的物体。例如,如果表面分离技术包括生成普通映射,则接近并具有相似表面法线的数据集可被标记为属于相同的物体。标记提供了平坦表面和微曲表面之间的区别,而空间上结合或不相交的表面(像地板和墙)可被分别地标记。与邻居点连接的多个点可以基于那些点与指向相似方向的对应的表面法线之间的距离来标记。调解距离阈值和法线相似度阈值可以导致物体不同的大小和曲率、以及被离散地标记的表面。已知物体的阈值和预期结果可被存储在物体过滤器中。

[0155] 如图 7C 所示,示出了球 102 和箱子 104 的点云。对邻近的点云数据和从点云集合标识的表面法线的评估可以将球与箱子相区分。由此,可以标记每个物体 102 和 104。标记可以简单地是唯一标识。云中多个点的位置与表面法线的组合对于在表面上的多个物体或构成一个物体的多个物体之间进行区分是有用的。例如,如果杯子位于箱子 104 顶部,则可

以用与赋予箱子相同的唯一 ID 来标识这个杯子,因为可能尚未从点云数据中确定这些物体是不相交的。然而,通过随后考虑表面法线,系统可以确定在各表面法线之间存在九十度差别,并且确定物体应当基于点的邻近度和点云来分别地标记。由此,点云中与结构表面元素一致的各数据点组可被相关联并被标记。

[0156] 系统可以重新投射各个点云的已确定的表面定向并且对纹理进行重新对齐,就好像它在平坦表面上。技术使系统能够更准确地对物体进行重新处理。例如,如果用户举起带有打印文本的杂志,则对于用户朝向捕捉设备举起杂志的定向没有限制。捕捉设备可以重新投射杂志表面的所捕捉的纹理,并且重新投影该纹理,包括颜色信息、文本以及任何纹理。

[0157] 对于被标记并具有其所涵盖的计算的参数组的物体,系统可以出于提高的保真度、机构和结构的目的,对虚拟场景执行或继续执行分析。例如,最适合边界框可以是区分特定物体的更准确的方式。最适合边界框可在特定帧中给出物体的定向。例如,其顶部带有咖啡杯的箱子最初可被赋予一个边界框,该边界框包括箱子的点云和表示该咖啡杯的点云二者。在每帧中,系统可以评估物体在空间上位于与最后一帧中相同的位置,并且确定定向是否相似。咖啡杯可以逐帧地移动,而系统可以标识出杯子与箱子是分开的,并因此为杯子生成新的边界框并为纸板箱重新定义边界框。

[0158] 有时由于房间中无关紧要的粒子或物体,或者基于所使用的传感器的类型,噪声被引入到系统中。例如,云中的一组点可以表示苍蝇的点云,或者所使用的传感器的类型可能导致过量的无关点。为了减少噪声,可以执行清洁阶段以对传感器数据进行清洁或者移除非常小的物体和仅具有少量组成点样本的物体。例如,可以捕捉场景中的灰尘粒子或苍蝇,但是表示苍蝇的少量组成点样本可能不够重要以触发与该点云相关联的表面法线的身份。由此,可以从分析中提取表示苍蝇的少量的组成点样本。点云数据的初始轮可一起使用物体中的在空间上与给出大型物体阵列有关的多个点。例如,点的大型集合可能是长椅并使用特定 ID 来标记;另一个物体可能是地板。特定阈值可被设置为标识应从分析移除的点集。例如,如果仅有 20 个点被标识用于一个物体,并且与物理空间或场景中的其他物体相比这 20 个点的空间安排位于相对小的区域,则系统可以消除这 20 个点。

[0159] 轴对齐的边界框可用作对物体所占据的全部体积/空间的快速测量。轴对齐是指诸如 X、Y 或 Z 等特殊轴,而不是空间中的物体的轴。例如,系统可以计算表面是复杂的还是简单的(例如,球或杂志具有简单的表面;玩偶或植物具有复杂的表面)。物体的旋转对于系统分析并确定物体的更精细的特性可能是有用的。捕捉设备可以执行物体的固体扫描以便体积估计。捕捉设备还可提供点云与场景中物体之间的引用,使得参考物理空间可以标识物体的特定位置。

[0160] 对物体属性的计算和随着时间对这些变化的跟踪建立了用于跟踪物体的可靠的技术,该物体的位置和定向实时地在各帧之间变化。在捕捉更多帧时,使用临时信息以捕捉变化可以对场景中的物体的解析、标识、和标记赋予进一步的置信度。由于普通数据集的大小,诸如 640x480 点,使用所公开的技术甚至可以实现复杂的处理。可在帧序列中以至少 20 赫兹的频率捕捉数据。

[0161] 可将物体参数与前一帧的参数进行比较,并且物体可被重新标记以允许移动物体以便被实时地跟踪,而还维持从静态物体的持续地标记。可以为每个物体计算置信度,并且

置信度因子可以随时间而增加。由此，在物体的置信度可能保持为高时，静态物体可由于遮挡而移入和移出视野。临时分析可包括对上一帧和当前帧的评估。如果物体在每帧中是相同的，则可以使用它在前一帧中的标签来重新标记该物体，以便逐帧地向标签和物体赋予一致性。物体以及表面定向和位置可用于估计深度相机的定向以及收集与相机周围相关的统计数据。例如，在很多情形中，主要平坦表面的位置将等于墙和地板。

[0162] 应该理解，此处所述的配置和 / 或方法在本质上是示例性的，且这些具体实施例或示例不被认为是限制性的。本文中所述的具体例程或方法可表示任意数量的处理策略中的一个或多个。由此，所示出的各个动作可以按所示顺序执行、按其他顺序执行、并行地执行等等。同样，可改变上述过程的次序。

[0163] 此外，尽管已经结合某些方面按各附图所示描述了本发明，但要理解，可使用其它相似方面或者可对所述方面进行修改或添加来执行本发明的相同功能而不脱离本发明。本公开的主题包括各种过程、系统和配置的所有新颖和非显而易见的组合和子组合、和此处所公开的其它特征、功能、动作、和 / 或属性、以及其任何和全部等效物。由此，所公开的各实施例的方法和装置或其某些方面或部分可采用包含在诸如软盘、CD-ROM、硬盘驱动器或任何其他机器可读存储介质等有形介质中的程序代码（即，指令）的形式。当程序代码被加载到诸如计算机等机器并由其执行时，该机器变为被配置成实施所公开的各实施例的装置。

[0164] 除了此处明确阐述的具体实现之外，考虑此处所公开的说明书，其它方面和实现将对本领域的技术人员是显而易见的。因此，本发明不应该仅限于任何单个方面，而是应该在根据所附权利要求书的广度和范围内解释。例如，本文描述的各种过程可用硬件或软件、或两者的组合来实现。

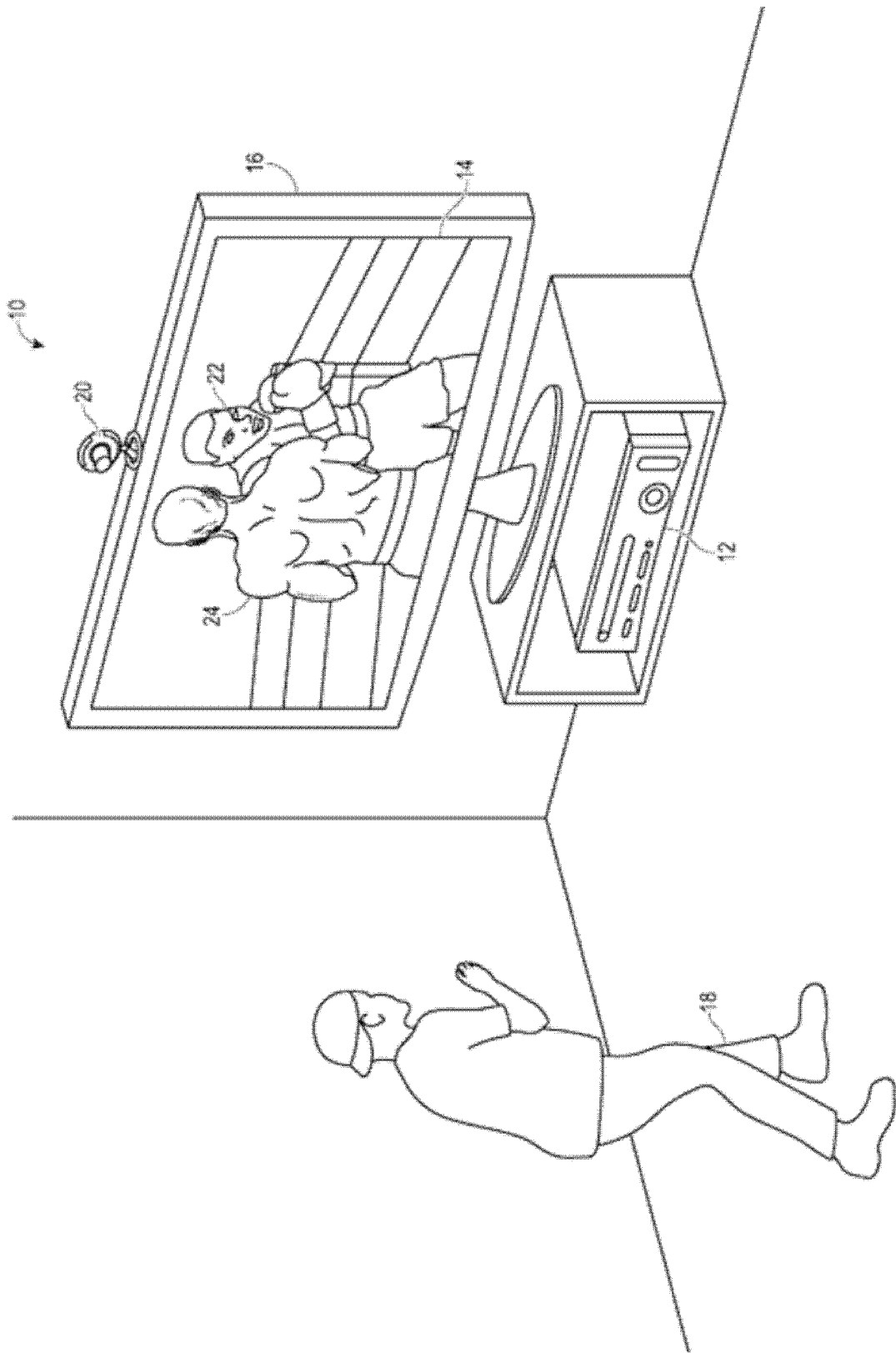


图 1

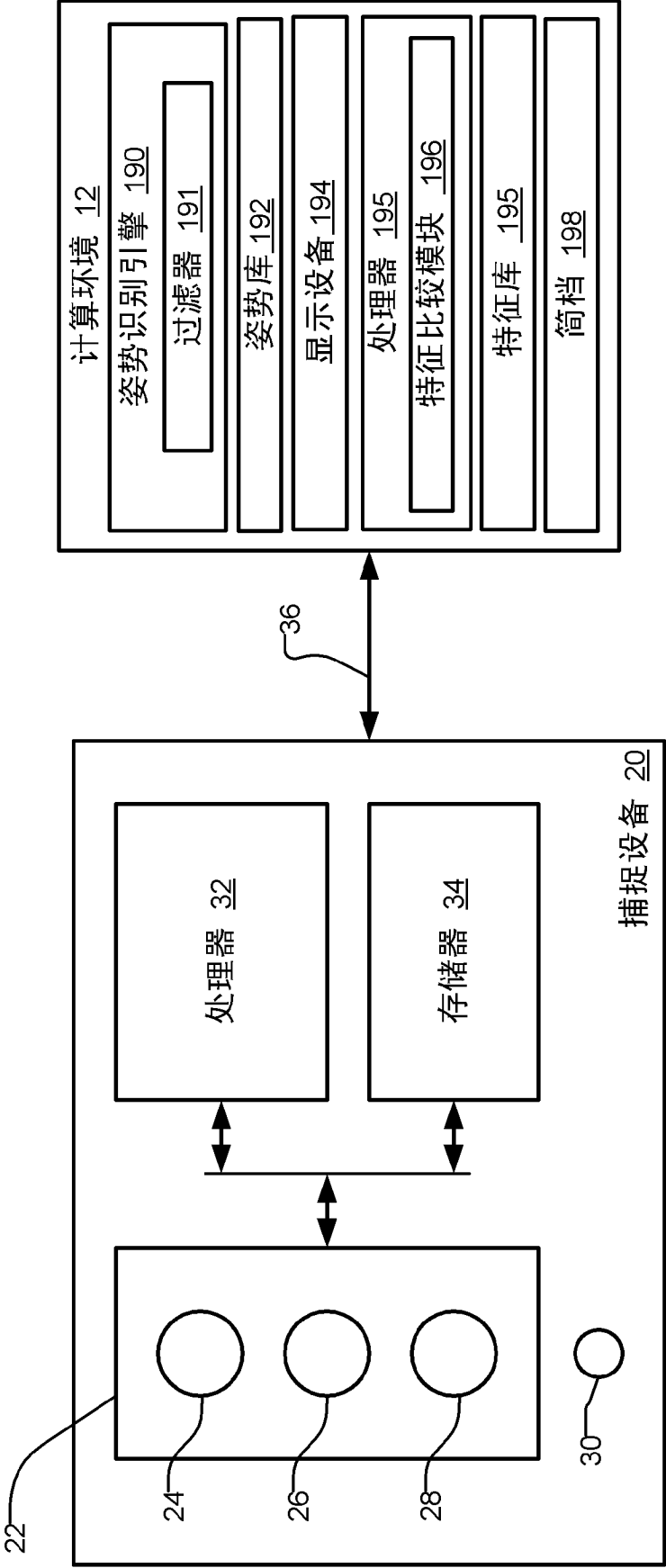


图 2

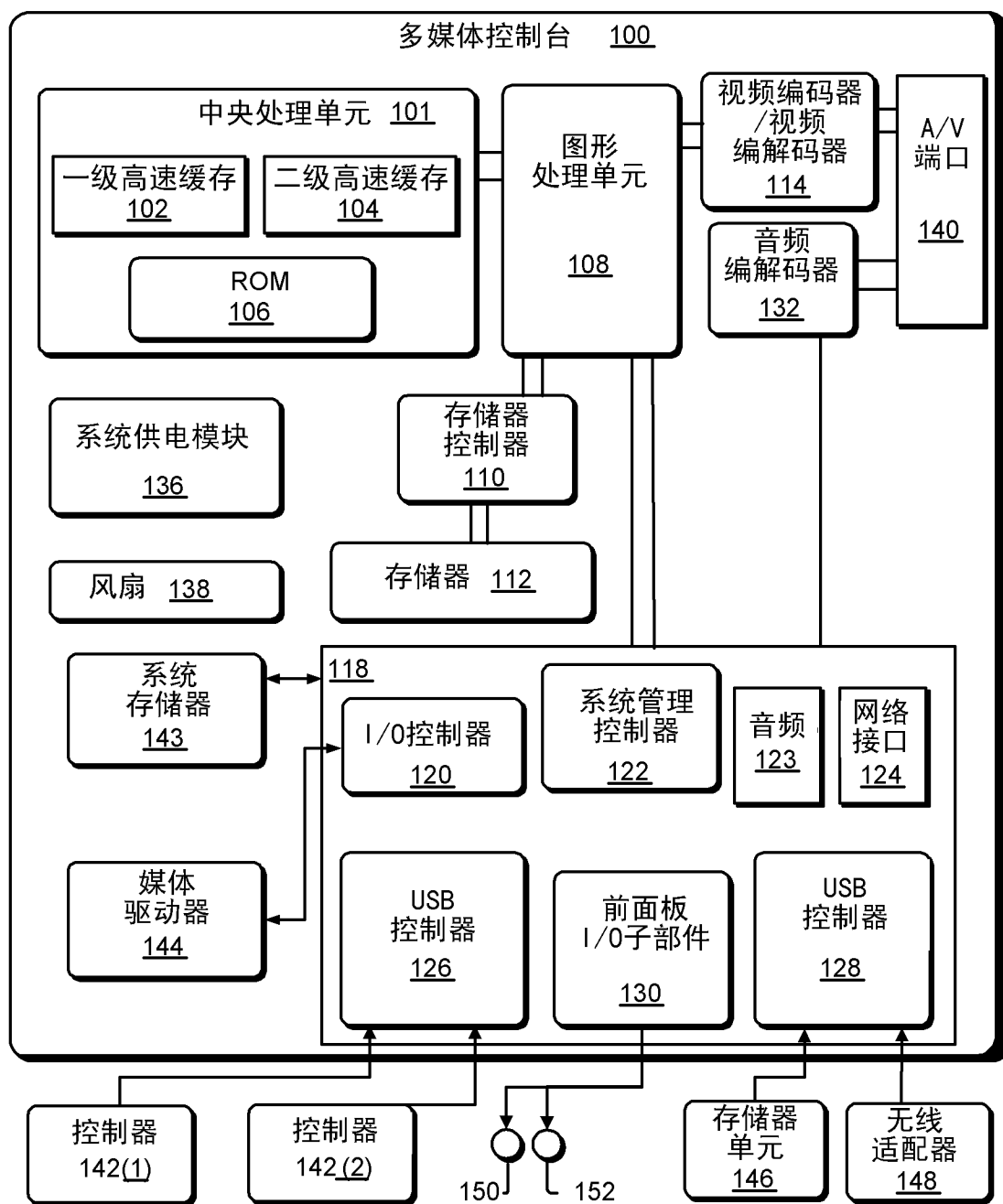


图 3

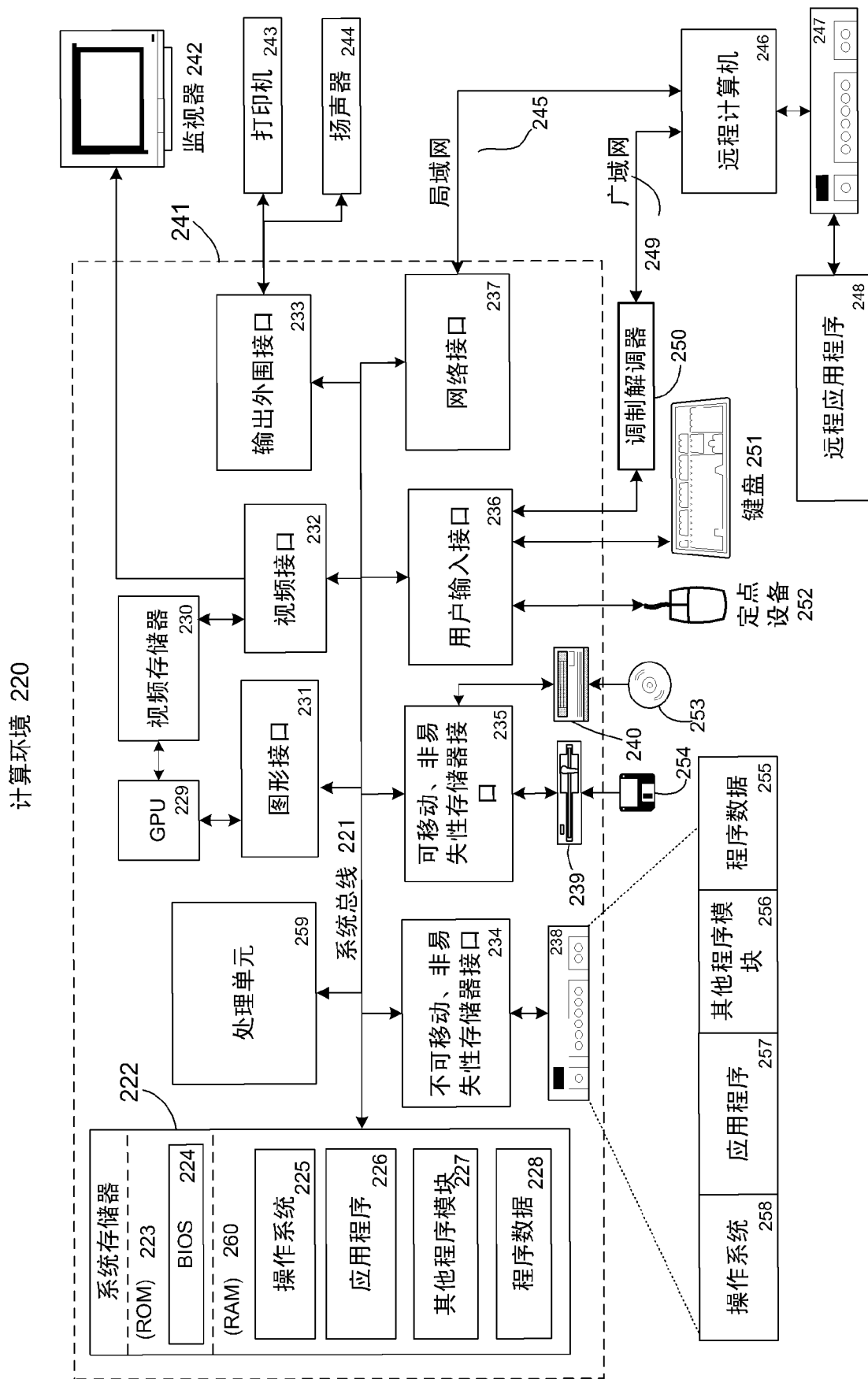


图 4

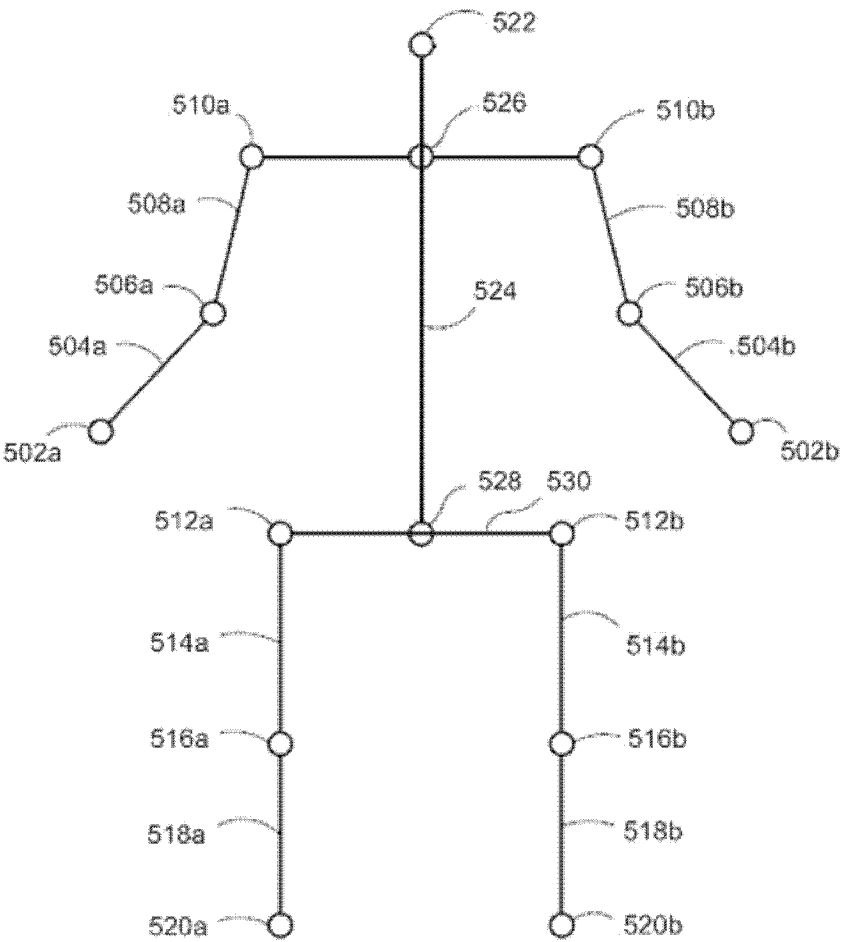


图 5

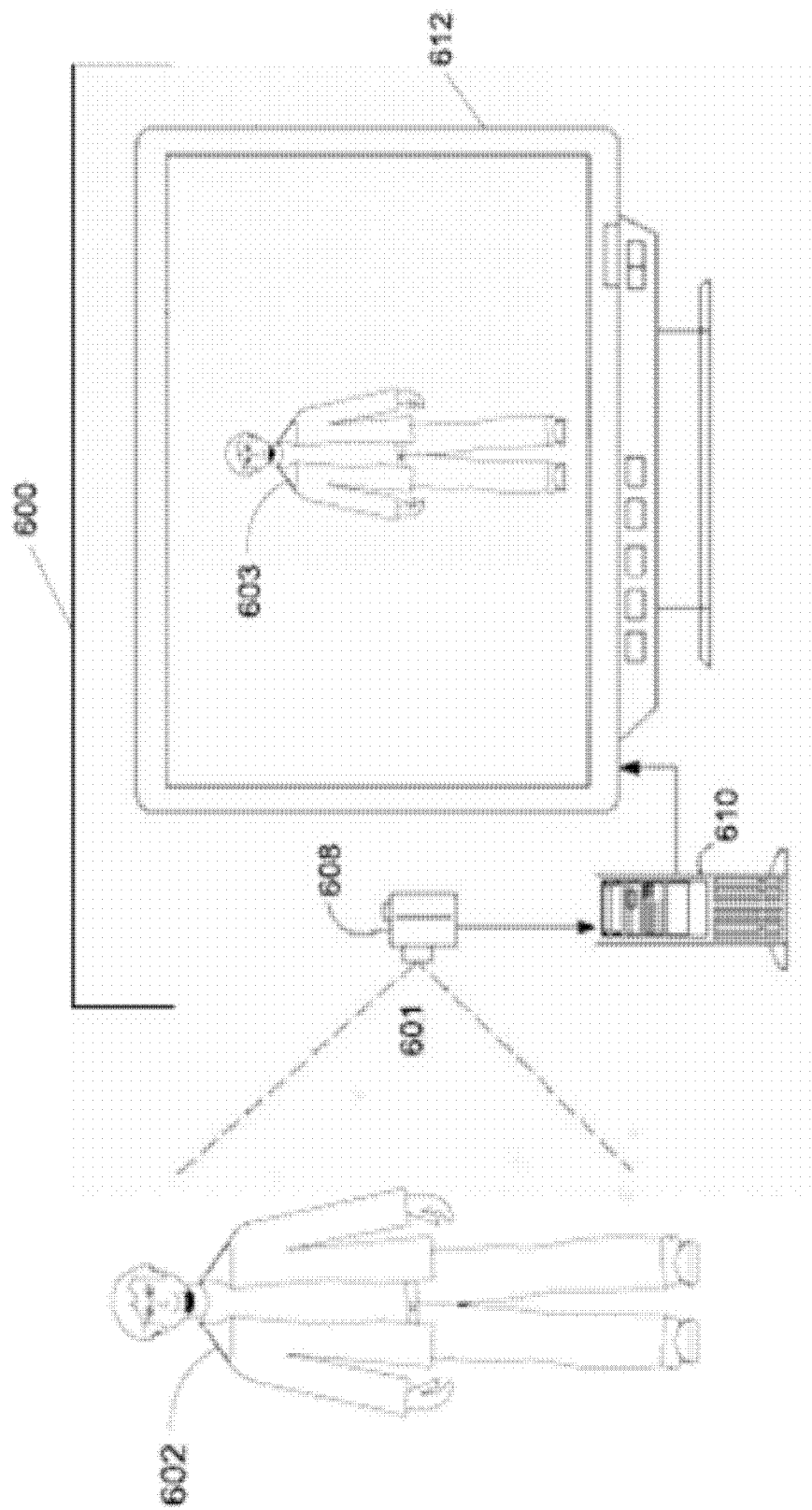


图 6A

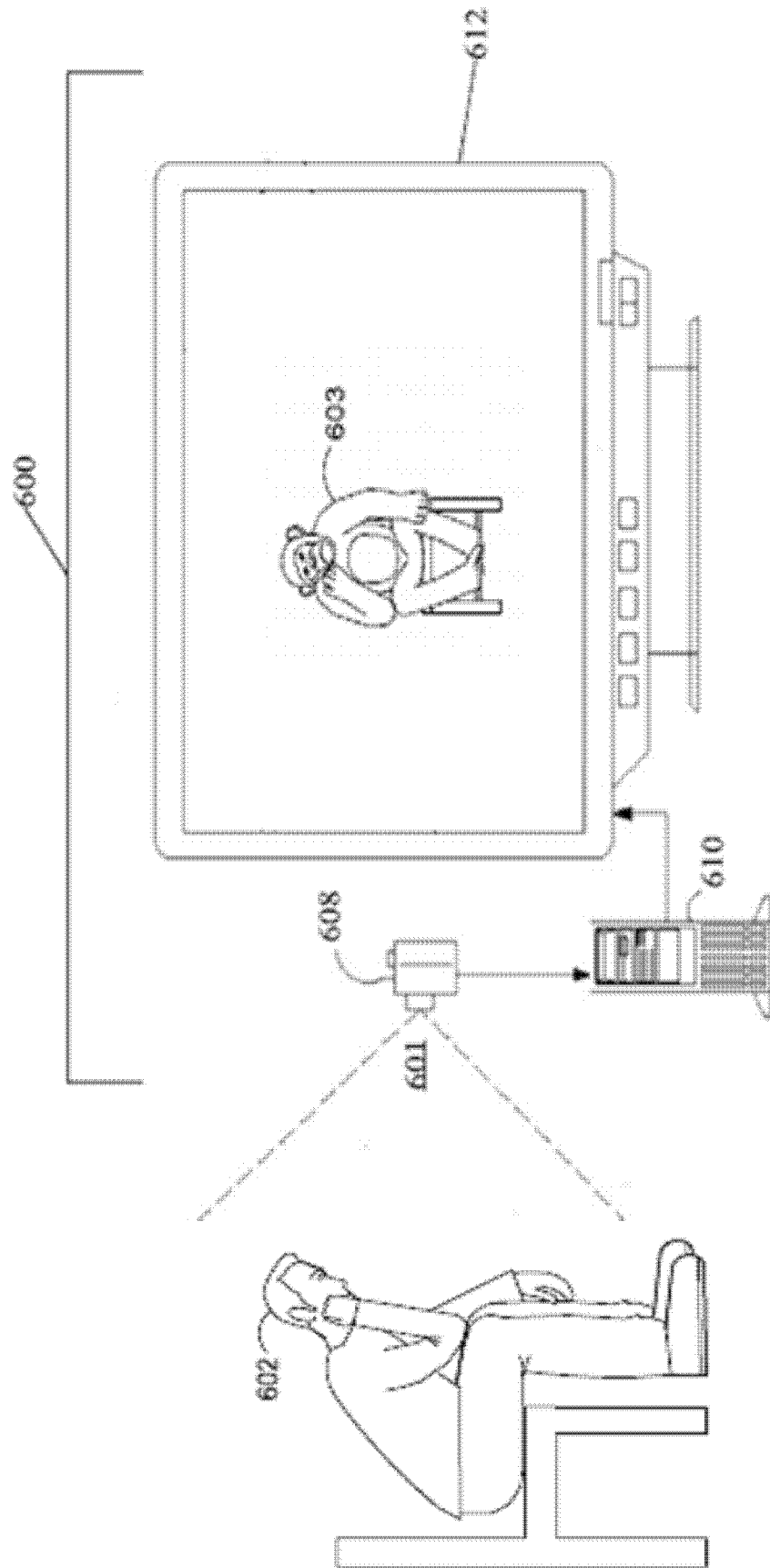


图 6B

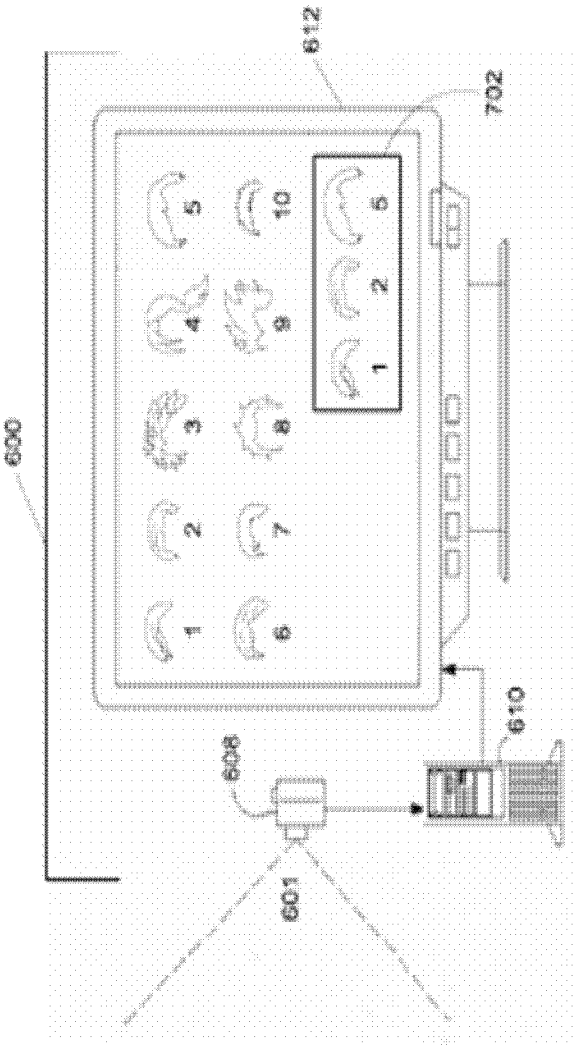


图 7

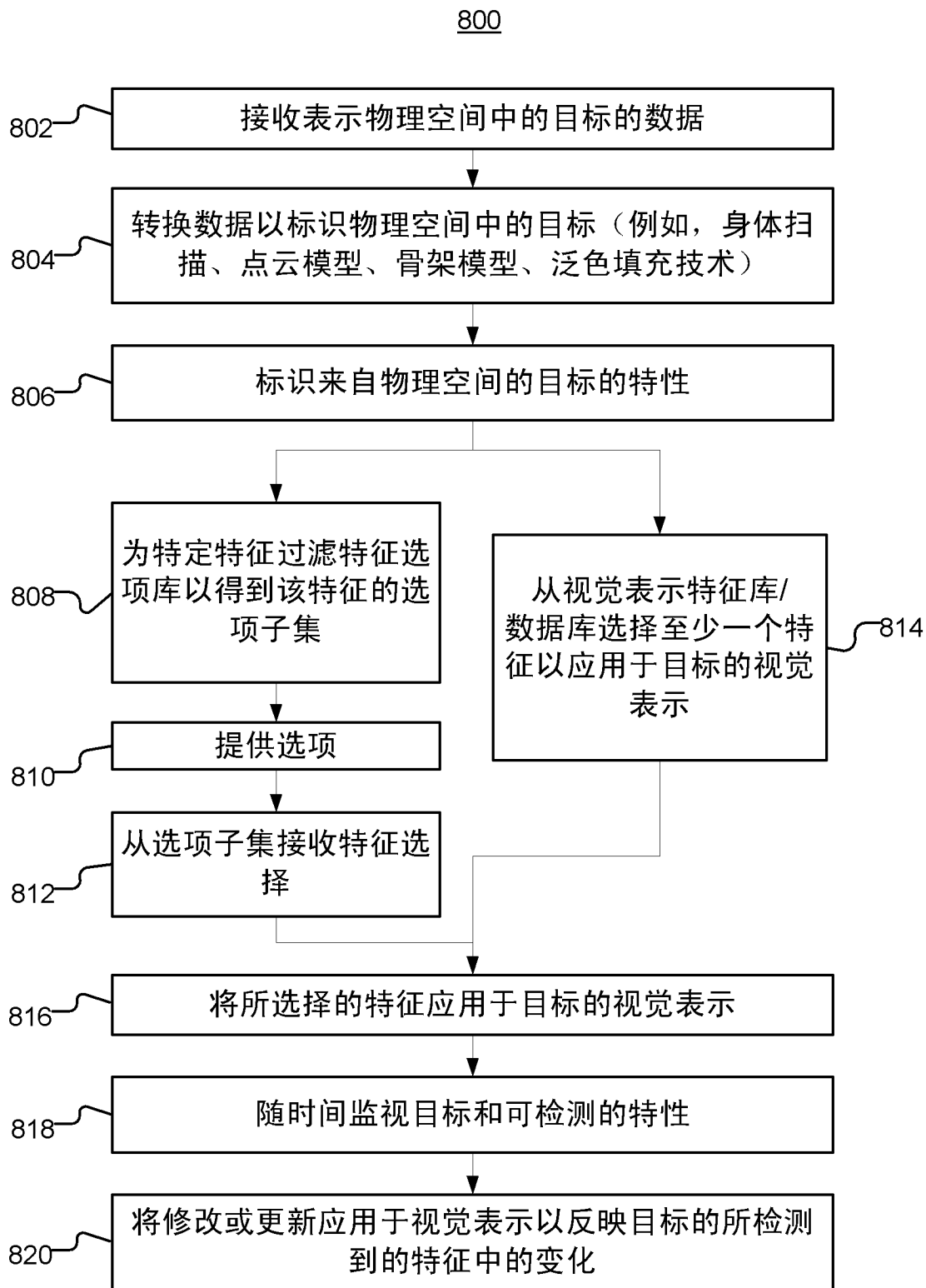


图 8

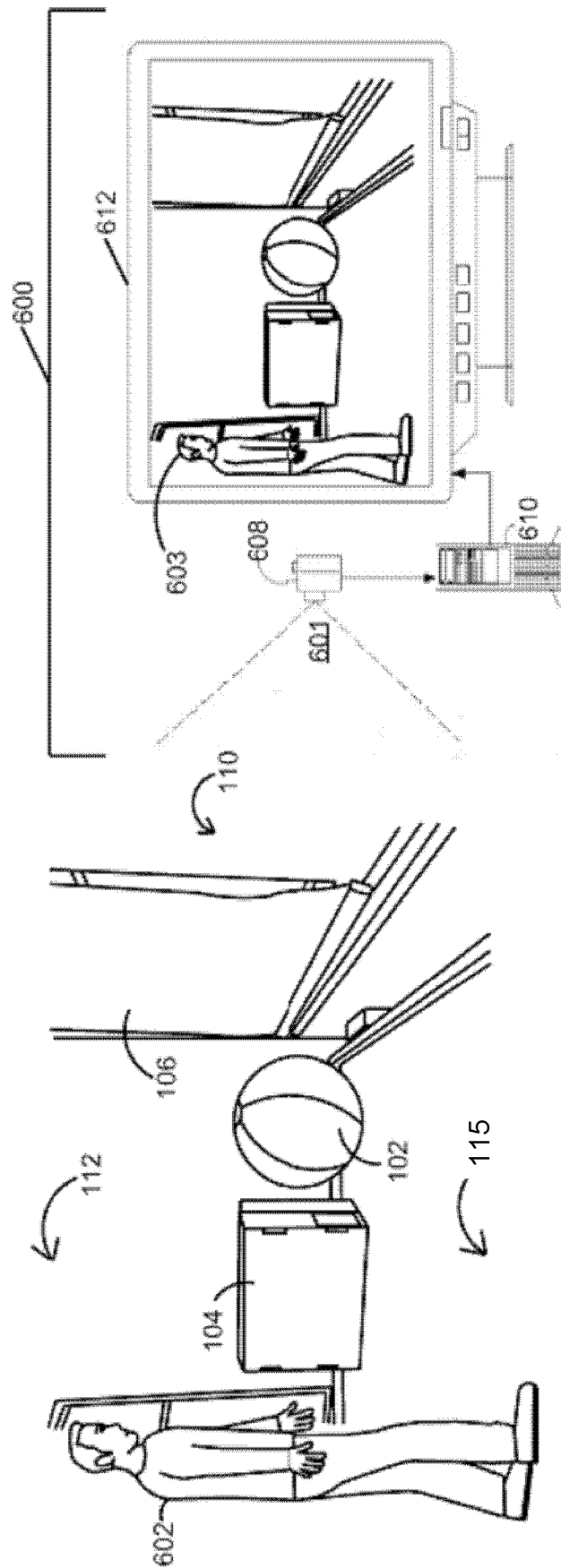


图 9