



(12) 发明专利

(10) 授权公告号 CN 102549619 B

(45) 授权公告日 2013. 11. 20

(21) 申请号 201080044919. 0

(51) Int. Cl.

(22) 申请日 2010. 10. 06

G06T 7/20 (2006. 01)

H04N 5/262 (2006. 01)

(30) 优先权数据

12/575, 388 2009. 10. 07 US

审查员 吴黄飞

(85) PCT申请进入国家阶段日

2012. 04. 06

(86) PCT申请的申请数据

PCT/US2010/051662 2010. 10. 06

(87) PCT申请的公布数据

W02011/044257 EN 2011. 04. 14

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 T·莱瓦德 J·李 S·斯塔奇亚克

C·佩普尔 S·刘

(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

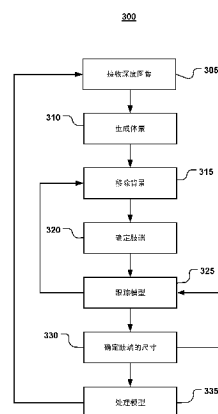
权利要求书2页 说明书23页 附图19页

(54) 发明名称

人类跟踪系统

(57) 摘要

可由设备接收、观察或捕捉诸如场景的深度图像等图像。接着可基于深度图像生成体素的网格,使得可对该深度图像进行降采样。可移除体素的网格中所包括的背景,以隔离与前景对象(诸如人类目标)相关联的一个或多个体素。可以确定被隔离的人类目标的一个或多个肢端的位置或方位,并且可以基于一个或多个肢端的位置或方位来调整模型。



1. 一种用于跟踪用户的方法,所述方法包括:
接收深度图像;
基于所述深度图像生成体素的网格;
移除所述体素的网格中所包括的背景以隔离与人类目标相关联的一个或多个体素;
确定被隔离的人类目标的一个或多个肢端的位置或方位,包括为所述人类目标定义边界框以确定所述人类目标的核体积,定义边界框包括沿 X 方向水平地搜索以确定人类目标的宽度以及沿 Y 方向垂直地搜索以确定所述人类目标的高度;以及
基于所述一个或多个肢端的位置或方位来调整模型。
2. 如权利要求 1 所述的方法,其特征在于,所述模型包括具有关节和骨骼的骨架模型。
3. 如权利要求 2 所述的方法,其特征在于,基于一个或多个所确定的肢端来调整所述模型包括将所述一个或多个肢端的位置或方位指派到所述模型的对应关节。
4. 如权利要求 1 所述的方法,其特征在于,移除所述体素的网格的背景以隔离与人类目标相关联的一个或多个体素还包括:
确定所述网格中的物体是否在运动中;以及
当基于所述确定所述物体不在运动中时,丢弃与所述物体相关联的体素。
5. 如权利要求 1 所述的方法,其特征在于,还包括确定所述一个或多个肢端的尺寸,以及基于所确定的尺寸来调整所述模型的关节和骨骼。
6. 如权利要求 1 所述的方法,其特征在于,所述一个或多个肢端包括如下的至少一个:头、质心、肩、髋、腿、臂、手、肘、膝和脚。
7. 如权利要求 1 所述的方法,其特征在于,还包括处理经调整的模型。
8. 如权利要求 1 所述的方法,其特征在于,所述网格中的每个体素是基于所接收的深度图像中所包括的像素部分或块来生成的。
9. 一种计算机执行的方法,所述方法包括:
接收包括像素的深度图像;
对所接收的深度图像中的所述像素进行降采样以生成一个或多个体素;
隔离与人类目标相关联的一个或多个体素;
确定被隔离的人类目标的一个或多个肢端的位置或方位,包括为所述人类目标定义边界框以确定所述人类目标的核体积,定义边界框包括沿 X 方向水平地搜索以确定人类目标的宽度以及沿 Y 方向垂直地搜索以确定所述人类目标的高度;以及
基于所述一个或多个肢端的位置或方位来调整模型。
10. 如权利要求 9 所述的方法,其特征在于,所述模型包括具有关节和骨骼的骨架模型。
11. 一种用于跟踪与用户相关联的模型的系统,所述系统包括:
捕捉设备,其中所述捕捉设备包括接收场景的深度图像的相机组件;以及
与所述捕捉设备可操作地通信的计算设备,其中所述计算设备包括处理器,所述处理器基于从所述捕捉设备接收的所述深度图像中的一个或多个像素来生成经降采样的深度图像,移除所述经降采样的深度图像的背景以隔离人类目标,确定被隔离的人类目标的一个或多个肢端的位置或方位,并且基于所述一个或多个肢端的位置或方位来调整模型;
其中,确定被隔离的人类目标的一个或多个肢端的位置或方位包括为所述人类目标定

义边界框以确定所述人类目标的核体积,定义边界框包括沿 X 方向水平地搜索以确定人类目标的宽度以及沿 Y 方向垂直地搜索以确定所述人类目标的高度。

12. 如权利要求 11 所述的系统,其特征在于,所述处理器通过为所述一个或多个像素部分或块生成体素,来基于从所述捕捉设备接收的所述深度图像中的一个或多个像素生成所述经降采样的深度图像。

13. 如权利要求 12 所述的系统,其特征在于,所述模型包括具有关节和骨骼的骨架模型。

14. 如权利要求 13 所述的系统,其特征在于,所述处理器通过将所述一个或多个肢端的位置或方位指派到所述模型的对应关节,来基于一个或多个所确定的肢端调整所述模型。

15. 如权利要求 11 所述的系统,其特征在于,其中所述一个或多个肢端包括如下的至少一个:头、质心、肩、髋、腿、臂、手、肘、膝和脚。

人类跟踪系统

背景技术

[0001] 诸如计算机游戏、多媒体应用等的许多计算应用使用控制来允许用户操纵游戏角色或应用的其他方面。通常使用例如控制器、遥控器、键盘、鼠标等等来输入这样的控制。不幸的是,这些控制可能是难以学习的,由此造成了用户和这些游戏及应用之间的障碍。此外,这些控制可能与这些控制所用于的实际游戏动作或其他应用动作不同。例如,使得游戏角色挥动棒球球棒的游戏控制可能不与挥动棒球球棒的实际运动相对应。

发明内容

[0002] 此处所公开的是用于跟踪场景中的用户的系统和方法。例如,可以接收或观察诸如场景的深度图像等图像。可基于深度图像生成体素 (voxel) 的网格 (grid) 以便降采样该深度图像。例如,深度图像可包括多个像素,这些像素可被划分成部分或块。然后可为每个部分或块生成体素,使得所接收的深度图像被降采样到体素的网格中。

[0003] 根据一个实施例,然后可移除体素的网格中所包括的背景以隔离开与前景对象 (诸如人类目标) 相关联的一个或多个体素。可以确定所隔离开的人类目标的一个或多个肢端 (诸如质心或中心、头、肩、髋、臂、手、肘、腿、脚、膝等) 的位置或方位。此外,可以确定这些肢端的尺寸,诸如包括宽度、长度等的测量值。

[0004] 随后可基于由此所确定的一个或多个肢端的位置或方位和 / 或尺寸来调整模型。例如,该模型可以是可包括诸如关节和 / 或骨骼的骨架模型。可以调整模型的一个或多个关节,使得该一个或多个关节可被指派到与其对应的一个或多个肢端的位置或方位,和 / 或在其间定义的骨骼可被调整至与其对应的一个或多个肢端的尺寸。

[0005] 可处理经调整的模型。例如,在一个实施例中,可以将经调整的模型映射到化身或游戏角色,使得该化身或游戏角色可以被动画化 (animated) 以模仿该用户,和 / 或经调整的模型可被提供到计算环境中的姿势库,该姿势库可用于基于该模型中的各身体部分的方位来确定在应用内执行的控制。

[0006] 提供本发明内容以便以简化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。本发明内容并不旨在标识出所要求保护的主题的关键或必要特征,也不旨在用于限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一部分中提及的任何或所有缺点的实现。

附图说明

[0007] 图 1A 和 1B 示出了目标识别、分析和跟踪系统的示例实施例,其中用户正在玩游戏。

[0008] 图 2 示出了可在目标识别、分析和跟踪系统中使用的捕捉设备的示例实施例。

[0009] 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和 / 或动画化目标识别、分析和跟踪系统所显示的化身或屏幕上角色的计算环境的示例实施例。

[0010] 图 4 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和 / 或动画

化目标识别、分析和跟踪系统所显示的化身或屏幕上角色的计算环境的另一示例实施例。

[0011] 图 5 描绘了用于跟踪场景中的用户的示例方法的流程图。

[0012] 图 6 示出可被捕捉或观察的深度图像的示例实施例。

[0013] 图 7A-7B 示出被降采样的深度图像的一部分的示例实施例。

[0014] 图 8 示出了为人类目标估计的质心或中心的示例实施例。

[0015] 图 9 示出了可被定义为确定核体积的边界框的示例实施例。

[0016] 图 10 示出了可被创建为对头候选项进行打分的头柱体和躯干柱体的示例实施例。

[0017] 图 11 示出了基于人类目标的头以及质心或中心的头 - 到 - 中心向量的示例实施例。

[0018] 图 12 示出了基于头 - 到 - 中心向量所确定的肩体块和髋体块的示例实施例。

[0019] 图 13 示出可基于肩体块和髋体块来计算的肩和髋的示例实施例。

[0020] 图 14 示出了可表示核体积的柱体的示例实施例。

[0021] 图 15A-15C 示出了基于锚点来确定的手的示例实施例。

[0022] 图 16 示出可基于臂和腿的平均方位和 / 或锚点来计算的手和脚的示例实施例。

[0023] 图 17 示出了可被生成的模型的示例实施例。

[0024] 说明性实施例的具体实施方式

[0025] 图 1A 和 1B 示出目标识别、分析和跟踪系统 10 的配置的示例实施例,其中用户 18 正在玩拳击游戏。在一示例实施例中,目标识别、分析和跟踪系统 10 可用于识别、分析和 / 或跟踪诸如用户 18 等的人类目标。

[0026] 如图 1A 所示,目标识别、分析和跟踪系统 10 可以包括计算环境 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。在一个实施例中,计算环境 12 可包括可执行指令的标准化处理器、专用处理器、微处理器等,这些指令可包括例如用于接收深度图像的指令;用于基于深度图像来生成体素网格的指令;用于移除体素网格中所包括的背景以隔离与人类目标相关联的一个或多个体素的指令;用于确定被隔离的人类目标的一个或多个肢端的位置或方位的指令;用于基于一个或多个肢端的位置或地点来调整模型的指令;或任何其他合适的指令,下面将对它们进行更详细的描述。

[0027] 如图 1A 所示,目标识别、分析和跟踪系统 10 还可包括捕捉设备 20。捕捉设备 20 可以是例如照相机,所述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可以捕捉、分析并跟踪一个或多个用户所执行的姿势和 / 或移动,来执行应用程序中的一个或多个控制命令或动作和 / 或动画化化身或屏幕上人物,如将在下面更详细地描述的。

[0028] 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。

[0029] 如图 1A 和 1B 所示,目标识别、分析和跟踪系统 10 可用于识别、分析和 / 或跟踪诸如用户 18 等的人类目标。例如,可使用捕捉设备 20 来跟踪用户 18,从而可以捕捉用户 18 的姿势和 / 或移动来动画化化身或屏幕上人物,和 / 或可将用户 18 的姿势和 / 或移动解释为可用于影响计算环境 12 所执行的的应用的控制命令。因此,根据一个实施例,用户 18 可移动他或她的身体来控制应用和 / 或动画化化身或屏幕上人物。

[0030] 如图 1A 和 1B 所示,在一示例性实施例中,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。例如,计算环境 12 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。计算环境 12 还可使用视听设备 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。例如,如图 1B 所示,用户 18 可以在物理空间中挥拳以致使玩家化身 40 在游戏空间中挥拳。因此,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 可用于识别和分析物理空间中用户 18 的重拳从而使得该重拳可被解释为对游戏空间中的玩家化身 40 的游戏控制和 / 或该重拳的运动可用于动画化游戏空间中的玩家化身 40。

[0031] 用户 18 的其他移动也可被解释为其他控制命令或动作,和 / 或用于动画化玩家化身,如上下快速摆动、闪避、滑步、封堵、用拳猛击或挥动各种不同的有力重拳的控制命令。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,在一个实施例中,玩家可使用移动来结束、暂停或保存游戏、选择级别、查看高分、与朋友交流等。根据另一实施例,玩家可使用移动来从主用户界面选择游戏或其他应用程序。因此,在示例实施例中,用户 18 的全范围运动可以用任何合适的方式来获得、使用并分析与应用程序进行交互。

[0032] 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。

[0033] 根据其他示例实施例,目标识别、分析和跟踪系统 10 还可用于将目标移动解释为游戏领域之外的操作系统和 / 或应用控制。例如,事实上操作系统和 / 或应用程序的任何可控方面都可由诸如用户 18 等目标的移动来控制。

[0034] 图 2 示出可在目标识别、分析和跟踪系统 10 中使用的捕捉设备 20 的示例实施例。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。根据一实施例,捕捉设备 20 可将深度信息组织为“Z 层”或者可与从深度相机沿其视线延伸的 Z 轴垂直的层。

[0035] 如图 2 所示,捕捉设备 20 可包括图像相机组件 22。根据一个示例性实施例,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。深度图像可以包括所捕捉场景的二维 (2-D) 像素区域,其中 2-D 像素区域中的每个像素都可表示深度值,比如例如所捕捉场景中的对象距相机的以厘米、毫米等为单位的长度或距离。

[0036] 如图 2 所示,根据一示例性实施例,图像相机组件 22 可以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,可以使用脉冲红外光,使得可以测量出射光脉冲与相应入射光脉冲之间的时间,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。附加地,在其他示例性实施例中,可以将出射光波的相位与入射光波的相位相比较来确定相移。然后可以使用该相移来确定从捕捉设备到目标或物体上的特定位置的物理距离。

[0037] 根据另一示例性实施例,可以使用飞行时间分析来通过经由包括例如快门式光脉冲成像在内的各种技术分析反射光束随时间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。

[0038] 在另一示例性实施例中,捕捉设备 20 可以使用结构化光来捕捉深度信息。在这样的分析中,图案化光(即被显示为诸如网格图案或条纹图案之类的已知图案的光)可以经由例如 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上以后,作为响应,图案可以变为变形的。图案的这种变形可由例如 3-D 相机 26 和 / 或 RGB 相机 28 来捕捉,然后可被分析来确定从捕捉设备到目标或物体上的特定位置的物理距离。

[0039] 根据另一实施例,捕捉设备 20 可包括两个或更多物理上分开的相机,这些相机可从不同角度查看场景以获得视觉立体数据,该视觉立体数据可被解析以生成深度信息。

[0040] 捕捉设备 20 还可包括话筒 30。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 与计算环境 12 之间的反馈。附加地,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、非游戏应用等之类的应用。

[0041] 在一示例性实施例中,捕捉设备 20 还可包括可与图像相机组件 22 可操作地通信的处理器 32。处理器 32 可包括可执行指令的标准化处理器、专用处理器、微处理器等,这些指令可包括例如用于接收深度图像的指令;用于基于深度图像来生成体素网格的指令;用于移除体元网格中所包括的背景以隔离与人类目标相关联的一个或多个体素的指令;用于确定被隔离的人类目标的一个或多个肢端的位置或方位的指令;用于基于一个或多个肢端的位置或定位来调整模型的指令;或任何其他合适的指令,下面将对它们进行更详细的描述。

[0042] 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 相机或 RGB 相机捕捉到的图像或图像的帧、或任何其他合适的信息、图像等。根据一示例性实施例,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。

[0043] 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,该时钟可用于确定何时捕捉例如场景。

[0044] 另外,捕捉设备 20 可经由通信链路 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 捕捉的图像,和 / 或可由捕捉设备 20 生成的骨架模型。计算环境 12 然后可使用该模型、深度信息、以及所捕捉的图像来例如控制诸如游戏或字处理程序等的应用和 / 或动画化化身或屏幕上人物。例如,如图 2 所示,计算环境 12 可包括姿势库 190。姿势库 190 可包括姿势过滤器集合,每一姿势过滤器包括关于骨架模型可执行(在用户移动时)的姿势的信息。可将由相机 26、28 和捕捉设备 20 捕捉的骨架模型形式的数据以及与其相关联的移动与姿势库 190 中的姿势过滤器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。那些姿势可与应用的各种控制相关联。因此,计算环境 12 可使用姿势库 190 来解释骨架模型的移动并基于该移动来控制应用。

[0045] 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和 / 或动画化目标识别、分析和跟踪系统所显示的化身或屏幕上的人物的计算环境的示例实施例。上面参考图 1A-2 所描述的诸如计算环境 12 等计算环境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。CPU101 可被提供为具有一个以上的核,并且由此具有附加的一级高速缓存 102 和二级高速缓存 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。

[0046] 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。数据经由总线从图形处理单元 108 输送到视频编码器 / 视频编解码器 114。视频处理流水线向 A/V(音频 / 视频)端口 140 输出数据,以便传输到电视机或其他显示器。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)。

[0047] 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、以及外置存储器设备 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的任何一种。

[0048] 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,并且其可包括 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 可内置或外置于多媒体控制台 100。应用数据可经由媒体驱动器 144 访问,以供多媒体控制台 100 执行、回放等。媒体驱动器 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。

[0049] 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。

[0050] 前面板 I/O 子部件 130 支持暴露在多媒体控制台 100 的外表面上的电源按钮 150 和弹出按钮 152、以及任何 LED（发光二极管）或其他指示器的功能。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。

[0051] 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。

[0052] 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。应用可在导航到多媒体控制台 100 上可用的不同媒体类型时呈现提供一致的用户体验的图形用户界面。在操作中，媒体驱动器 144 中所包含的应用和/或其他媒体可从媒体驱动器 144 启动或播放，以将附加功能提供给多媒体控制台 100。

[0053] 多媒体控制台 100 可通过简单地将该系统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。

[0054] 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，所保留的资源从应用的视角而言是不存在的。

[0055] 具体而言，存储器保留量优选地足够大，以包含启动内核、并发系统应用程序和驱动程序。CPU 保留量优选地为恒定，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。

[0056] 对于 GPU 保留量，通过使用 GPU 中断来调度代码以将弹出窗口呈现为覆盖图，从而显示由系统应用程序生成的轻量消息（例如，弹出窗口）。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。

[0057] 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。操作系统内核标识出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，以便提供对应用而言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的高速缓存分裂最小化。

[0058] 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（如下所描述的）在系统应用活动时控制游戏应用的音频级别（例如，静音、衰减）。

[0059] 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。

[0060] 图 4 示出计算环境 220 的另一示例实施例,该计算环境可以是用于解释目标识别、分析和跟踪系统中的一个或多个姿势和 / 或动画化目标识别、分析和跟踪系统所显示的化身或屏幕上人物的图 1A-2 中示出的计算环境 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转换成等价的硬件结构,而硬件结构本身可被转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之一并留给实现者。

[0061] 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被计算机 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。

[0062] 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。

[0063] 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。

[0064] 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、企业范围的计算机网络、内联网和因特网中是常见的。

[0065] 当在 LAN 联网环境中使用时，计算机 241 通过网络接口或适配器 237 连接到 LAN 245。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。

[0066] 图 5 描绘了用于跟踪场景中的用户的示例方法 300 的流程图。示例方法 300 可使用例如，参考图 1A-4 所描述的目标识别、分析和跟踪系统 10 的捕捉设备 20 和 / 或计算环境 12 来实现。在一示例实施例中，示例方法 300 可采取可由例如参考图 1A-4 所描述的目标识别、分析和跟踪系统 10 的捕捉设备 20 和 / 或计算环境 12 执行的程序代码（即指令）的形式。

[0067] 根据一个实施例，在 305，可接收深度图像。例如，目标识别、分析和跟踪系统可包括诸如参考图 1A-2 描述的捕捉设备 20 等捕捉设备。捕捉设备可捕捉或观察可包括一个或多个目标的场景。在一示例实施例中，捕捉设备可以是深度相机，该深度相机被配置成使用诸如飞行时间分析、结构化光分析、立体视觉分析等任何合适的技术来获得场景的深度图像。

[0068] 深度图像可以是多个观测到的像素，其中每个观测到的像素具有观测到的深度值。例如，深度图像可包括所捕捉的场景的二维（2D）像素区域，其中 2D 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捕捉设备的长度或距离。

[0069] 图 6 示出了可在 305 处被接收的深度图像 400 的示例实施例。根据一示例实施例，深度图像 400 可以是由例如以上参考图 2 所描述的捕捉设备 20 的 3D 相机 26 和 / 或 RGB 相机 28 所捕捉的场景的图像或帧。如图 6 所示，深度图像 400 可包括与例如诸如参考图 1A 和 1B 所描述的用户 18 等的用户对应的人类目标 402a 以及诸如所捕捉的场景中的墙、桌子、监视器等的一个或多个非人类目标 404。如上所述，深度图像 400 可包括多个观测到的

像素,其中每一观测到的像素具有与其相关联的观测到的深度值。例如,深度图像 400 可包括所捕捉的场景的二维 (2D) 像素区域,其中该 2D 像素区域中的特定 X 值和 Y 值处的每一像素可具有诸如以例如厘米、毫米等为单位的所捕捉的场景中的目标或对象离捕捉设备的长度或距离之类的深度值。

[0070] 在一个实施例中,深度图像 400 可被着色 (colorized) 从而使得深度图像的像素的不同颜色对应于和 / 或从视觉上描绘了人类目标 402a 和非人类目标 404 离捕捉设备的不同距离。例如,与深度图像中最接近捕捉设备的目标相关联的像素可用红色和 / 或橙色阴影来着色,而与深度图像中较远的目标相关联的像素可用绿色和 / 或蓝色阴影来着色。

[0071] 回头参考图 5,在一个实施例中,一接收到该图像,在 305,可从深度图像中移除和 / 或平滑掉一个或多个高变度和 / 或含噪声的深度值;可填入和 / 或重构缺少的和 / 或被移除的深度信息的各部分;和 / 或可对接收到的深度图像执行任何其他合适的处理,以使与该深度图像相关联的深度信息可用于生成诸如骨架模型之类的模型,这将在下文更详细地描述。

[0072] 根据一示例实施例,在 310,可基于所接收到的深度图像生成一个或多个体素的网格。例如,目标识别、分析和跟踪系统可通过使用所接收到的深度图像中所包括的信息来生成一个或多个体素来对所接收的深度图像进行降采样,以便可以生成经降采样的深度图像。在一个实施例中,该一个或多个体素可以是可在经子采样 (sub-sampled) 的网格上表示所接收的深度图像中所包括的信息的数据或值的体元素。

[0073] 例如,如上所述,深度图像可包括所捕捉的场景的 2D 像素区域,其中每一像素具有与其相关联的 X 值、Y 值和深度值 (或 Z 值)。在一实施例中,深度图像可通过将 2D 像素区域中的像素减少为一个或多个体素的网格来降采样。例如,深度图像可被分成各个像素部分或像素块,诸如 4x4 的像素块、5x5 的像素块、8x8 的像素块、10x10 的像素块等等。每个部分或块可被处理以生成深度图像的体素,该体素可表示与该 2D 深度图像的像素相关联的部分或块在真实世界空间中的方位。根据一示例实施例,每个体素的方位可以基于例如该体素可能表示的块或部分中的像素的有效或非零深度值的平均深度值、或者该体素可能表示的块或部分中的像素的最小、最大和 / 或中间深度值、该体素可能表示的部分或块中的具有有效深度值的像素的 X 值和 Y 值的平均、或该深度图像所提供的任何适当的信息。因此,根据一示例实施例,每个体素可表示该深度图像中具有诸如以下值的子体 (sub-volume) 部分或块:该体素可能表示的块或部分中的像素的有效或非零深度值的平均深度值、该体素可能表示的部分或块中的像素的最小、最大和 / 或中间深度值、该体素可能表示的部分或块中的具有有效深度值的像素的 X 值和 Y 值的平均、或基于在 305 所接收的深度图像的像素的相应部分或块的 X 值、Y 值和深度值由该深度图像所提供的任何其他适当信息。

[0074] 在一个实施例中,经降采样的深度图像中的一个或多个体素的网格可被分层。例如,目标识别、分析和跟踪系统可生成如上所述的体素。然后该目标识别、分析和跟踪系统可将所生成的体素堆叠在该网格中所生成的一个或多个其他体素上。

[0075] 根据一示例实施例,该目标识别、分析和跟踪系统可以将体素堆叠在该网格中,例如该深度图像中所捕捉的场景中的物体的边缘周围。例如,在 305 接收的深度图像可包括人类目标和非人类目标 (诸如墙)。例如在人类目标的边缘处,人类目标可以与非人类目

标（诸如墙）重叠。在一个实施例中，重叠边缘可包括诸如与该深度图像中可捕捉的该人类目标和非人类目标相关联的深度值、X 值、Y 值等。该目标识别、分析和跟踪系统可以生成该重叠边缘处的与该人类目标相关联的体素以及与该人类目标相关联的体素，以便该体素可被堆叠而该重叠边缘的信息（诸如深度值、X 值、Y 值等）可被保持在该网格中。

[0076] 根据另一实施例，在 310，通过例如将深度图像中像素的信息（诸如深度值、X 值、Y 值等）投影到三维（3D）空间中，可以生成一个或多个体素的网格。例如，目标识别、分析和跟踪系统可以使用诸如相机、图像或透视变换等变换将深度图像中的像素的信息（诸如 X 值、Y 值等）映射到 3D 空间中的 3D 点，使得该信息在 3D 空间中可被变换为梯形或锥形。在一个实施例中，具有梯形或锥形的 3D 空间可被划分成诸如立方体的块，所述块可创建体素的网格，使得这些块或立方体中的每一个可表示该网格中的体素。例如，该目标识别、分析和跟踪系统可以将 3-D 网格叠加在与该深度图像中的物体相对应的 3-D 点上。随后该目标识别、分析和跟踪系统可以将该网格划分或切割为表示体素的块以将该深度图像降采样为更低的分辨率。根据一示例实施例，网格中的每个体素可包括该体素可表示的与网格中的 3D 空间相关联的像素的有效或非零深度值的平均深度值、该体素可表示的与网格中的 3D 空间相关联的像素的最小和 / 或最大深度值、该体素可表示的与网格中的 3D 空间相关联的具有有效深度值的像素的 X 值和 Y 值的平均、或该深度图像所提供的任何其他适合的信息。

[0077] 图 7A-7B 示出被降采样的深度图像的一部分的示例实施例。例如，如图 7A 所示，以上参考图 6 所述的深度图像 400 的一部分 410 可包括多个像素 420，其中每一像素 420 可具有与其相关联的 X 值、Y 值和深度值（或 Z 值）。根据一个实施例，如上所述，深度图像（诸如深度图像 400）可通过将 2D 像素区域中的像素减少为一个或多个体素的网格来降采样。例如，如图 7A 所示，深度图像 400 的部分 410 可以被划分为像素 420 的部分或块 430，诸如像素 420 的 8x8 块。目标识别、分析和跟踪系统可以处理该部分或块 430 以生成体素 440，体素 440 可表示与像素 420 相关联的部分或块 430 在真实世界中的方位，如图 7A-7B 所示。

[0078] 回头参考图 5，在 315，可将背景从经降采样的深度图像中移除。例如，可以移除经降采样的深度图像中的背景（诸如非人类目标或物体）以隔离前景物体（诸如与用户相关联的人类目标）。在一个实施例中，如上所述，目标识别、分析和跟踪系统可以通过为所捕捉或观察的深度图像生成一个或多个体素的网格来对所捕捉或观察的深度图像进行降采样。目标识别、分析和跟踪系统可以分析经降采样的深度图像中的每个体素以确定体素是否可与背景对象（诸如该深度图像的一个或多个非人类目标）相关联。如果体素可与背景物体相关联，则该体素可从经降采样的深度图像被移除或丢弃，使得前景物体（诸如人类目标）和该网格中与该前景物体相关联的一个或多个体素可以被隔离。

[0079] 根据一个实施例，目标识别、分析和跟踪系统可以分析每个体素以确定与其相关联的物体。例如，如上所述，可在 305 处作为深度图像（诸如上面参考图 6 所描述的深度图像 400）来观察或捕捉的场景可包括多个物体。物体可包括一个或多个人类目标和 / 或一个或多个非人类目标，诸如墙、桌子、沙发椅、灯等。在一个实施例中，目标识别、分析和跟踪系统可以分析网格中的每个体素以确定该体素可能与场景中的哪个物体相关联，使得在 315 处，目标识别、分析和跟踪系统可以标识与场景中的每个物体相关联的体素。因此，根据一

示例实施例,如果人类目标或一个人站在场景中的墙的前面,则目标识别、分析和跟踪系统可以分析每个体素以确定该体素可能是与人类目标还是与墙相关联。

[0080] 为了确定体素可能与场景中的哪个物体相关联,目标识别、分析和跟踪系统可以比较以下值,诸如该体素可表示的块或部分中的像素的有效或非零深度值的平均深度值、该体素可表示的块或部分中的像素的最小和 / 或最大深度值、该体素可表示的具有有效深度值的像素的 X 值和 Y 值的平均、或者邻近或附近体素的任何其他适合的信息。例如,在一个实施例中,可以将与网格中正被分析的特定体素相关联的平均深度值与同网格中正被分析的特定体素相邻的每个体素的平均深度值进行比较。如果正被分析的特定体素的平均深度值与相邻体素的平均深度值的差可以小于阈值,则特定体素与相邻体素可被标识为属于相同的物体。如果正被分析的特定体素的平均深度值与相邻体素的平均深度值的差可以大于阈值,则特定体素与相邻体素可被标识为属于不同的物体。根据一个示例实施例,该阈值可以是例如由目标识别、分析和跟踪系统所生成的预定值,该预定值可以基于该体素可能是相同的物体的一部分的可能性或概率。因此,根据一示例实施例,如果人类目标或一个人站在由深度图像所捕捉或观察的场景中的墙的前面,则目标识别、分析和跟踪系统可以分析为该深度图像生成的每个体素以确定该体素可能是与人类目标还是与墙相关联。

[0081] 在标识了接收到的深度图像的场景中的物体以及与之相关联的体素之后,目标识别、分析和跟踪系统随后可以计算与每个所标识的物体相关联的信息。例如,目标识别、分析和跟踪系统可以为每个所标识的物体计算最大世界空间、最小世界空间方位和平均世界空间方位等。

[0082] 在一个实施例中,在 315 处,目标识别、分析和跟踪系统还可确定场景中所标识的一个或多个物体是否应与场景中的其他物体合并。例如,可将在 305 处接收到的深度图像中的物体的一部分与该物体的另一部分相分离。根据一个实施例,可通过由例如场景中的一个物体或另一个物体等所投射的红外阴影来将该物体的一部分与该物体的另一部分相分离。在另一个示例实施例中,可通过例如与一个物体相关联的颜色、纹理、图案等来将该物体的一部分与该物体的另一部分相分离。例如,可通过例如面部毛发、多件服装等沿 Y 方向中的 Y 平面将人类目标的头与该人类目标的躯干相分离。

[0083] 为了确定场景中所标识的一个物体是否可能实际上是场景中所标识的另一个物体的一部分,目标识别、分析和跟踪系统可以对同该物体相关联的体素的 X 值、深度值与同附近物体相关联的体素的 X 值、深度值进行比较。例如,目标识别、分析和跟踪系统可以对同例如场景中所标识的第一物体相关联的一个或多个体素的 X 值、深度值与同该第一物体附近或邻近的第二物体相关联的一个或多个体素的 X 值、深度值进行比较。因此,根据示例实施例,目标识别、分析和跟踪系统可以分析场景中的体素以确定第一和第二物体是否沿着 X 方向中定义的 X 平面和 / 或 Z 方向中定义的 Z 平面重叠,使得可以合并第一和第二物体并将其标识为同一物体的各部分。

[0084] 根据一个实施例,如果同第一物体相关联的一个或多个体素的 X 值和深度值可能与同第二物体相关联的一个或多个体素的 X 值和深度值相重叠,则目标识别、分析和跟踪系统可以合并第一和第二物体,使得目标识别、分析和跟踪系统可将第一和第二物体标识为共同物体的各部分。例如,如果同第一物体相关联的第一体素可具有沿 X 方向的 X 值 5 以及位于第一物体右侧外缘的深度值 10mm,并且同第二物体相关联的第二体素可具有沿 X

方向的X值3以及位于第二物体左侧外缘的深度值10mm,则目标识别、分析和跟踪系统可以确定第一和第二物体可能重叠。目标识别、分析和跟踪系统随后可合并第一和第二物体,使得目标识别、分析和跟踪系统可将第一和第二物体标识为同一物体的各部分。

[0085] 另外,为了确定场景中所标识的一个物体是否可能实际上是场景中所标识的另一个物体的一部分,目标识别、分析和跟踪系统可以确定在场景中为一个物体所定义的边界框是否与另一个物体的边界框相重叠。例如,目标识别、分析和跟踪系统可为每个所标识的物体定义边界框。目标识别、分析和跟踪系统随后可以基于例如如上所述的包括在其中的一个或多个体素的X值、Y值和/或深度值来确定一个或多个物体的边界框是否重叠。

[0086] 根据另一个示例实施例,目标识别、分析和跟踪系统可以通过例如对物体中所包括的体素的X值、Y值和深度值取平均来确定每个物体的中心或质心。目标识别、分析和跟踪系统随后可以确定场景中各物体的质心或中心之间的距离,以便确定场景中所标识的一个物体是否可能实际上是场景中所标识的另一个物体的一部分。基于各物体之间的距离,目标识别、分析和跟踪系统可以合并一个或多个物体。例如,目标识别、分析和跟踪系统可以确定第一物体的质心或中心与第二物体的中心或质心之间的距离。如果第一物体与第二物体的质心或中心之间的距离可能位于指示应当合并第一和第二物体的预定范围以内,则目标识别、分析和跟踪系统可以合并这两个物体,使得目标识别、分析和跟踪系统可将第一和第二物体标识为同一物体的各部分。

[0087] 在一个实施例中,在315处,目标识别、分析和跟踪系统还可确定场景中所标识的一个或多个物体是否应被分离。例如,在315处场景中所标识的物体可能实际上是两个分离的物体。为了确定场景中的一个物体是否应被分离,目标识别、分析和跟踪系统可以标识为之前接收到的帧所确定的每个物体的中心的位置。根据一个实施例,目标识别、分析和跟踪系统随后可以同时从由之前接收到的帧的物体所确定的中心的位置开始,泛色填充场景中为在305处接收到的帧的深度图像所生成的体素。目标识别、分析和跟踪系统随后可以使用物体的之前位置来确定经泛色填充的体素可能更接近于之前接收到的帧中的哪个物体。如果经泛色填充的体素可能更接近于之前接收到的帧中所标识的另一个物体,则目标识别、分析和跟踪系统可在315处拆分一个物体。

[0088] 在315处,目标识别、分析和跟踪系统随后可以确定所标识的物体是否可能是诸如非人类目标的背景物体、还是诸如人类目标的前景物体。根据一示例实施例,目标识别、分析和跟踪系统可以基于所标识的物体是否可能在运动中或正在移动来确定所标识的物体可能是背景物体还是前景物体。例如,目标识别、分析和跟踪系统可包括诸如场景的基准图像等基准平面,该基准平面可包括例如每个体素的非运动深度信息。根据示例实施例,基准平面可包括体素的最小世界空间方位(诸如通过一系列帧确定的网格中体素的最小X值、Y值和深度值),体素的平均世界方位(通过一系列帧确定的网格中体素的平均X值、Y值和深度值),或者任何其他适合的基准平面。根据另一个实施例,基准平面可包括与场景中的每个体素相关联的移动平均。移动平均可包括例如通过一系列之前接收到的帧所确定的体素的平均深度值。

[0089] 根据一个实施例,目标识别、分析和跟踪系统可将例如在305处接收的深度图像的场景中所标识的物体相关联的每个体素的诸如最大深度值、平均深度值、最小深度值等

的深度信息与基准平面中所包括的每个对应体素的非运动深度信息进行比较。基于基准平面中对应体素的深度信息和非运动深度信息的比较,目标识别、分析和跟踪系统可以标识一个体素在移动。例如,在一个实施例中,如果诸如体素的最小深度值、最大深度值、和 / 或平均深度值的深度值可小于基准平面中对应体素的移动平均从而使得体素可在移动平均前方,则该体素可被标识为移动。根据另一个示例实施例,目标识别、分析和跟踪系统可以计算与体素和基准平面中对应体素相关联的值之间的差。如果例如诸如体素的平均深度值、最大深度值、和 / 或最小深度值的深度值与基准平面中对应体素的非运动信息所包括的深度值之间的差可大于运动阈值,则该体素可由目标识别、分析和跟踪系统标识为移动。

[0090] 根据另一个示例实施例,目标识别、分析和跟踪系统可将诸如体素和与其邻近的体素的最大深度值、平均深度值、最小深度值等的深度信息与基准平面中所包括的每个对应体素的非运动深度信息进行比较。例如,为了处理边缘噪声,目标识别、分析和跟踪系统可将特定体素和可与其邻近的体素的最小深度值相对于基准平面中对应体素的最小深度值进行比较,以确定体素和 / 或与其相关联的物体是否可能正在移动。如果例如特定体素和可与其邻近的体素的最小深度值与基准平面中对应体素的非运动信息所包括的最小深度值之间的差可能大于运动阈值,则该特定体素可由目标识别、分析和跟踪系统标识为移动。

[0091] 目标识别、分析和跟踪系统随后基于移动体素的百分比为每个所标识的物体计算前景分数。在一个实施例中,目标识别、分析和跟踪系统可以用可被标识为移动的岛中的体素的数量除以岛中所包括的体素总数来计算前景分数。

[0092] 目标识别、分析和跟踪系统随后可以隔离具有超过分数阈值的前景分数的物体。分数阈值可以是由目标识别、分析和跟踪系统所定义的可指示一个物体可处于运动中的值或百分比。例如,目标识别、分析和跟踪系统可基于前景分数将可能未在移动的背景物体从经降采样的深度图像中移除或丢弃,使得可在经降采样的深度图像中对的前景物体(诸如具有可能超过分数阈值的前景分数的人类目标)进行隔离。根据一示例实施例,为了移除或丢弃可能未正在移动的物体,目标识别、分析和跟踪系统可以通过用 0 值、或另一个适合的指示符、或可指示体素可能无效的标记代替 X 值、Y 值和 / 或深度值来移除或丢弃与未移动物体相关联的体素。

[0093] 在 320,可以为被隔离的前景物体(诸如人类目标)确定一个或多个肢端(诸如一个或多个身体部位)。例如,在一个实施例中,该目标识别、分析和跟踪系统可将一个或多个试探(heuristic)或规则应用到经隔离的人类目标以确定例如与该经隔离的人类目标相关联的例如质心或中心、头、肩、躯干、臂、腿等。根据一个实施例,基于对肢端的确定,目标识别、分析和跟踪系统可以生成和 / 或调整被隔离的人类目标的模型。例如,如果在 305 所接收的深度图像可以被包括在由捕捉设备(诸如上面参考图 1A-2 描述的捕捉设备 20)所观察或捕捉的初始帧中,则可基于在 320 确定的肢端(诸如质心、头、肩、臂、手、腿等等)的位置,通过例如将骨架模型的关节分配到肢端的所确定的位置来生成模型,这将在下面更详细地描述。或者,如果深度图像可以被包括在由捕捉设备所观察或捕捉的后续或非初始帧中,则可以基于在 320 确定的肢端(诸如质心、头、肩、臂、手、腿等)的位置来调整可在之前已经生成的模型。

[0094] 根据一示例实施例,在 315 隔离前景物体(诸如人类目标)之后,该目标识别、

分析和跟踪系统可在 320 计算该人类目标中的体素的平均以例如估算该人类目标的质心或中心。例如,该目标识别、分析和跟踪系统可计算该人类目标中包括的体素的平均方位,其可提供对该人类目标的质心或中心的估算。在一个实施例中,该目标识别、分析和跟踪系统可基于与体素相关联的 X 值、Y 值和深度值来计算与该人类目标相关联的体素的平均方位。例如,如上所述,该目标识别、分析和跟踪系统可通过对与体素相关联的像素的 X 值求平均来计算该体素的 X 值,通过对与该体素相关联的像素的 Y 值求平均来计算该体素的 Y 值,以及通过对与该体素相关联的像素的深度值求平均来计算该体素的深度值。在 320,目标识别、分析和跟踪系统可对该人类目标中所包括的体素的 X 值、Y 值和深度值求平均以计算平均方位,该平均方位可提供该人类目标的质心或中心的估算。

[0095] 图 8 示出了为人类目标 402b 估计的质心或中心的示例实施例。根据示例实施例,质心或中心的位置或方位 802 可以基于与如上所述的被隔离的人类目标 402b 相关联的体素的平均方位或位置。

[0096] 回头参考图 5,目标识别、分析和跟踪系统随后可在 320 为人类目标定义边界框以确定例如可包括人类目标的头和 / 或躯干的人类目标的核体积。例如,在确定对人类目标的质心或中心的估计时,目标识别、分析和跟踪系统可以沿 X 方向水平地搜索以确定可用于定义与核体积相关联的边界框的人类目标的宽度。根据一个实施例,为了沿 X 方向水平地搜索以确定人类目标的宽度,目标识别、分析和跟踪系统可以从质心或中心沿 X 轴在向左方向和向右方向上搜索,直到目标识别、分析和跟踪系统可到达无效的体素,诸如可能不包括与其相关联的深度值的体素、或者可能与场景中所标识的另一个物体相关联的体素。例如,如上所述,可以在 315 处移除与背景相关联的体素以隔离人类目标和与其相关联的体素。如上所述,根据示例实施例,为了在 315 处移除体素,目标识别、分析和跟踪系统可以使用 0 值、或另一个适合的指示符、或可指示体素无效的标记来代替与背景物体的体素相关联的 X 值、Y 值、和 / 或深度值。在 320 处,目标识别、分析和跟踪系统可从人类目标的质心在向左方向上搜索直至到达人类目标左侧的第一无效体素,并且可从人类目标的质心在向右方向上搜索直至到达人类目标右侧的第二无效体素。目标识别、分析和跟踪系统随后可基于例如向左方向上到达的第一无效体素邻近的第一有效体素的 X 值与向右方向上第二无效体素邻近的第二有效体素的 X 值之间的差来计算或测量长度。

[0097] 目标识别、分析和跟踪系统随后可沿 Y 方向垂直地搜索以确定人类目标的高度,例如可用于定义与核体积相关联的边界框的从头到髋的高度。根据一个实施例,为了沿 Y 方向垂直地搜索以测量人类目标的宽度,目标识别、分析和跟踪系统可以从质心或中心沿 Y 轴在向上方向和向下方向上搜索,直到目标识别、分析和跟踪系统可到达无效的体素、或者可能不包括与其相关联的 X 值、Y 值、或深度值的体素。例如,在 320 处,目标识别、分析和跟踪系统可从人类目标的质心在向左方向上搜索直至到达人类目标上部的第三无效体素,并且可从人类目标的质心在向下方向上搜索直至到达人类目标下部的第四无效体素。目标识别、分析和跟踪系统随后可基于例如向上方向上到达的第三无效体素邻近的第三有效体素的 Y 值与向下方向上第四无效体素邻近的第四有效体素的 Y 值之间的差来计算或测量高度。

[0098] 根据示例实施例,目标识别、分析和跟踪系统还可以在 X 和 Y 轴上沿 X 和 Y 方向以各种角度(诸如 30 度、45 度角、60 度角等)正交地搜索,以便确定可用于定义与核体积相

关联的边界框的其他距离和值。

[0099] 另外,目标识别、分析和跟踪系统可以基于距离或值的比率来定义与核体积相关联的边界框。例如,在一个实施例中,目标识别、分析和跟踪系统可以基于如上所述确定的高度乘以常量(诸如 0.2、0.25、0.3、或任何其他适合的值),来定义边界框的宽度。

[0100] 目标识别、分析和跟踪系统随后可以基于由沿着 X 轴的水平搜索所确定的第一和第二有效体素、由沿着 Y 轴的垂直搜索所确定的第三和第四有效体素、或者由例如正交搜索所确定的其他距离或值来定义可表示核体积的边界框。例如,在一个实施例中,目标识别、分析和跟踪系统可在第一有效体素的 X 值处沿 Y 轴生成边界框的第一垂直线,以及在第二有效体素的 X 值处沿 Y 轴生成边界框的第二垂直线。另外,目标识别、分析和跟踪系统可在第三有效体素的 Y 值处沿 X 轴生成边界框的第一水平线,以及在第四有效体素的 Y 值处沿 X 轴生成边界框的第二水平线。根据示例实施例,第一和第二水平线可以与第一和第二垂直线交叉以形成矩形或正方形,该矩形或正方形可以表示与人类目标的核体积相关联的边界框。

[0101] 图 9 示出了可被定义为确定核体积的边界框 804 的示例实施例。如图 9 所示,边界框 804 可以基于如上所述所确定的第一垂直线 VL1、第二垂直线 VL2 与第一水平线 HL1、第二水平线 HL2 的交叉来形成矩形。

[0102] 回头参考图 5,在 320,目标识别、分析和跟踪系统随后可确定人类目标的头。例如,在一个实施例中,在确定了核体积和定义了与其相关联的边界框之后,目标识别、分析和跟踪系统可以确定人类目标的头的位置或方位。

[0103] 为了确定头的方位或位置,目标识别、分析和跟踪系统可在适合于头的方位或位置搜索各个候选项,可以对各个候选项打分,并且随后可以基于分数从各个候选项中选择头的方位。例如,根据一个实施例,目标识别、分析和跟踪系统可以搜索该人类目标的绝对最高体素和 / 或邻近或接近该绝对最高体素的体素、基于为之前的帧确定的头的位置的一个或多个增量体素、可(从例如质心或中心和 / 或邻近或接近为之前的帧确定的之前向上向量上的最高体素的体素)竖直延伸的向上向量上的最高体素、在中心和为之前帧确定的最高体素之间的之前的向上向量上的最高体素、或可以是头的候选项的任何其他适合的体素。

[0104] 该目标识别、分析和跟踪系统然后可对这些候选项打分。根据一个实施例,可基于 3-D 模式匹配对这些候选项打分。例如,该目标识别、分析和跟踪系统可创建头柱体和肩柱体。该目标识别、分析和跟踪系统随后可基于头柱体中可包括的与候选项相关联的体素的数量来计算候选项的得分,这将在下面更为详细地描述。

[0105] 图 10 示出了可被创建为对与头相关联的候选项进行打分的头柱体 806 和肩 808 的示例实施例。根据示例实施例,目标识别、分析和跟踪系统可基于与头柱体 806 和肩柱体 808 中可包括的头候选项相关联的体素的数量来计算候选项的得分。例如,目标识别、分析和跟踪系统可基于与头候选项相关联的体素的位置来确定头柱体 806 和 / 或肩柱体 808 内部的头候选项的总数,以及基于与头候选项相关联的体素来确定头柱体 806(例如,区域 807 以内)和 / 或肩柱体 808 外部的头候选项的总数。目标识别、分析和跟踪系统还可基于肩柱体 808 的左半部 LH 中头候选项的数量与肩柱体 808 的右半部 RH 中头候选项的数量之间的差的绝对值的函数来计算对称的度量。在示例实施例中,目标识别、分析和跟踪系统随

后可通过从头柱体 806 和 / 或肩柱体 808 内部的候选项的总数中减去头柱体 806 和 / 或肩柱体 808 外部的候选项的总数,并进一步从头柱体 806 和 / 或肩柱体 808 内部和外部的候选项的总数之差中减去对称度量来计算候选项的分数。根据一个实施例,目标识别、分析和跟踪系统可将头柱体 806 和 / 或肩柱体 808 内部和外部的候选项的总数乘以由目标识别、分析和跟踪系统在减法之前所确定的常数。

[0106] 回头参考图 5,根据一个实施例,如果与候选项之一相关联的分数超过头阈值分数,则目标识别、分析和跟踪系统可在 320 基于与该候选项相关联的体素来确定头的方位或位置。例如,在一个实施例中,目标识别、分析和跟踪系统可以基于最高点、可(从例如质心或中心和 / 或邻近或接近为之前帧所确定的之前向上向量上的最高体素的体素)垂直延伸的向上向量上的最高体素、在中心和为之前帧确定的最高体素之间的之前向上向量上的最高体素、之前帧中头的方位或位置周围的区域(诸如框、立方体等)内全部体素的平均方位等,来选择头的方位或位置。根据其他示例实施例,目标识别、分析和跟踪系统可以计算与可能超过头阈值分数的候选项相关联的体素的值(诸如 X 值、Y 值、和深度值)的平均,以确定头的方位或位置,或者目标识别、分析和跟踪系统可以基于可超过头阈值分数的候选项中所包括的体素的拟合线或最佳拟合线来选择头的方位或位置。

[0107] 此外,在一个实施例中,如果一个以上候选项超过头阈值分数,则目标识别、分析和跟踪系统可选择可具有最高分数的候选项,并且随后可基于与可具有最高分数的候选项相关联的体素来确定头的方位或位置。如上所述,目标识别、分析和跟踪系统可以基于例如与可具有最高分数的候选项相关联的体素的平均值(诸如 X 值、Y 值和深度值)来选择头的方位或位置。

[0108] 根据一个实施例,如果与候选项相关联的分数都未超过头阈值分数,则目标识别、分析和跟踪系统可使用其中头部分数可已超过头阈值分数的之前帧的深度图像相关联的人类目标中所包括的体素确定的头的之前方位或位置,或者如果在 305 接收的深度图像可能是在由该捕捉设备捕捉或观察的初始帧,则目标识别、分析和跟踪系统可使用人类目标的默认姿态(诸如,T 型姿态、自然站立姿态等)中的头的默认方位或位置。

[0109] 根据另一实施例,该目标识别、分析和跟踪系统可包括与例如头的形状相关联的一个或多个二维(2D)样式。目标识别、分析和跟踪系统随后可基于与候选项相关联的体素可能是一个或多个 2D 样式的头的形状的可能性来对与头相关联的候选项打分。例如,目标识别、分析和跟踪系统可对可以指示定义头的形状的邻近或附近的体素的深度值进行采样。如果可以指示定义头的形状的体素之一的经采样的深度值可能偏离头的形状的体素的一个或多个预期或预定义的深度值,则目标识别、分析和跟踪系统可以减少默认分数或初始分数,以指示该体素可能不是头。在一个实施例中,目标识别、分析和跟踪系统随后可以选择具有最高值的分数,并且可以基于与具有最高分数的候选项相关联的体素的位置或方位来指派头的位置或方位。

[0110] 根据另一个实施例,默认分数或初始分数可以是与使用如上所述的头和 / 或肩柱体来计算的头相关联的候选项的分数。如果候选项可能不在与一个或多个 2D 样式相关联的头的形状中,则目标识别、分析和跟踪系统可以减少这个分数。如上所述,目标识别、分析和跟踪系统随后可以选择超过头阈值分数的候选项的分数,并且可以基于候选项的位置或方位来指派头的位置或方位。

[0111] 该目标识别、分析和跟踪系统可在 320 进一步确定人类目标的肩和髌。例如，在一个实施例中，在确定人类目标的头的位置或方位之后，目标识别、分析和跟踪系统可确定人类目标的肩和髌的位置或方位。目标识别、分析和跟踪系统还可确定肩和髌的定向，诸如肩和髌的旋转或角度。

[0112] 根据示例实施例，为了确定肩和髌的位置或方位，目标识别、分析和跟踪系统可以基于头的位置或方位以及人类目标的质心或中心来定义头 - 到 - 中心的向量。例如，头 - 到 - 中心的向量可以是在头点的位置或方位的 X 值、Y 值和深度值（或 Z 值）与质点或中心点的位置或方位的 X 值、Y 值和深度值（或 Z 值）之间定义的向量或线。

[0113] 图 11 示出了基于人类目标的头以及质心或中心的头 - 到 - 中心向量的示例实施例。如上所述，可以确定诸如头的位置或方位 810 的位置或方位。如图 11 所示，目标识别、分析和跟踪系统随后可在头的位置或方位 810 与中心或质心的位置或方位 802 之间定义头 - 到 - 中心向量 812。

[0114] 回头参考图 5，在 320，目标识别、分析和跟踪系统随后可基于头 - 到 - 中心向量来定义肩体积框和髌体积框。例如，根据一个实施例，目标识别、分析和跟踪系统可以基于位移（诸如从诸如与头相关联的方位或位置的身体地标或质心或中心的长度）来定义或确定肩和髌的大致位置或方位。目标识别、分析和跟踪系统随后可以定义距离诸如与头相关联的方位或位置或质心或中心的身体地标的位移值周围的肩体块和髌体块。

[0115] 图 12 示出了基于头 - 到 - 中心向量 812 所确定的肩体块 SVB 和髌体块 HVB 的示例实施例。根据示例实施例，目标识别、分析和跟踪系统可以基于位移（诸如从诸如与头相关联的位置或方位 810、或与质心或中心相关联的位置或方位 802 的身体地标的长度）来定义或确定肩和髌的大致位置或方位。目标识别、分析和跟踪系统随后可以定义距离身体地标位移值周围的肩体块 SVB 和髌体块 HVB。

[0116] 回头参考图 5，目标识别、分析和跟踪系统还可在 320 基于诸如距离身体地标（诸如沿着头 - 到 - 中心向量的头）的长度的位移值来计算肩和髌的中心。例如，目标识别、分析和跟踪系统可以沿着头 - 到 - 中心向量向下或向上移动该位移值以计算肩和髌的中心。

[0117] 根据一个实施例，目标识别、分析和跟踪系统还可以确定诸如肩和髌的角度的定向。在一个实施例中，目标识别、分析和跟踪系统可以计算例如肩体块和髌体块内的深度值的拟合线，以便确定该定向，诸如肩和髌的角度。例如，目标识别、分析和跟踪系统可以基于与肩体块和髌体块相关联的体素的 X 值、Y 值和深度值来计算最佳拟合线，以便计算可通过肩的中心来定义肩骨的向量的肩斜度并可通过髌的中心来定义髌的关节之间的髌骨的向量的髌斜度。肩斜度和髌斜度可以定义相应的定向，诸如肩和髌的角度。

[0118] 根据另一个实施例，目标识别、分析和跟踪系统可以镜像人类目标的深度值，使得在 320 可以基于头 - 到 - 中心向量在人类目标的中心周围反映体素的深度值。因此，在一个实施例中，目标识别、分析和跟踪系统可通过围绕轴向量（诸如头 - 到 - 中心向量、从肩和 / 或髌边界框中计算的轴点等）来反映人类目标的体素的深度值来补偿身体的背面。

[0119] 目标识别、分析和跟踪系统随后可以计算包括例如肩体块和髌体块内所反映的深度值在内的深度值的拟合线，以便确定该定向，诸如肩和髌的角度。例如，目标识别、分析和跟踪系统可以基于（包括与肩体块和髌体块相关联的所反映的深度值在内的）体素的 X 值、Y 值和深度值来计算最佳拟合线，以便计算可通过肩的中心来定义肩骨的向量的肩斜度

以及可通过髋的中心来定义髋的关节之间的髋骨的向量的髋斜度。肩斜度和髋斜度可以定义相应的定向,诸如肩和髋的角度。

[0120] 图 13 示出可基于肩体块 SVB 和髋体块 HVB 来计算的肩和髋的示例实施例。如图 13 所示,可以如上所述地基于相应的肩体块 SVB 和髋体块 HVB 来确定肩的位置或方位 816a-b 和髋的位置或方位 818a-b。

[0121] 回头参考图 5,在 320,目标识别、分析和跟踪系统随后可确定人类目标的躯干。在一个实施例中,在确定肩和髋之后,目标识别、分析和跟踪系统可生成或创建躯干体,该躯干体可包括与头、肩、中心和髋相关联或围绕它们的体素。基于中心、头、肩和 / 或髋的位置或方位,躯干体可以是圆柱、药丸形(诸如两端为圆形的圆柱)等等。

[0122] 根据一个实施例,目标识别、分析和跟踪系统可以创建柱体,该柱体可表示具有基于肩、头、髋、中心等尺寸的核体积。例如,该目标识别、分析和跟踪系统可创建柱体,该柱体可具有基于肩的宽度的宽度或直径以及基于头和髋之间的距离的高度。然后该目标识别、分析和跟踪系统沿头 - 到 - 中心向量定向或倾斜(angle)可表示躯干体的柱体以便该躯干体可反映人类目标的躯干的定向(诸如角度)。

[0123] 图 14 示出了可表示核体积的柱体 820 的示例实施例。如图 14 所示,柱体 820 可以具有基于肩的宽度的宽度或直径以及基于头和髋之间的距离的高度。还可沿着头 - 到 - 中心向量 812 对柱体 820 定向或倾斜。

[0124] 回头参考图 5,在 320,目标识别、分析和跟踪系统随后可估计或确定人类目标的肢体。根据一个实施例,该目标识别、分析和跟踪系统可在生成或创建躯干体之后粗略地将躯干体外的体素标记为肢体。例如,目标识别、分析和跟踪系统可标识躯干体外的体素中的每一个,使得目标识别、分析和跟踪系统可将该体素标记为肢体的一部分。

[0125] 然后目标识别、分析和跟踪系统可确定与躯干体外的体素相关联的实际的肢体,诸如右臂和左臂、右手和左手、右腿和左腿、右脚和左脚等等。在一个实施例中,为了确定实际的肢体,目标识别、分析和跟踪系统可将所标识的肢体的之前方位或位置(诸如右臂、左臂、左腿、右腿等的之前方位或位置)与躯干体外的体素的方位或位置进行比较。根据示例实施例,之前标识的肢体的之前位置或方位可以是之前帧中所接收的深度图像中的肢体的位置或方位、基于之前移动的所投影的身体部位的位置或方位、或诸如人类目标的完全连接(fully articulated)骨架或体模型等人类目标的标识的任何其他适当之前的位置或方位。基于所述比较,目标识别、分析和跟踪系统随后可将躯干体外的体素与最接近的之前标识的肢体相关联。例如,目标识别、分析和跟踪系统可将包括躯干体外的体素中的每一个的 X 值、Y 值和深度值的方位或位置与包括之前标识的肢体(诸如之前标识的左臂、右臂、左腿、右腿等)的 X 值、Y 值和深度值的之前的方位或位置进行比较。目标识别、分析和跟踪系统随后可基于该比较将躯干体外的体素中的每一个与可能具有最接近的位置或方位的先前标识的肢体相关联。

[0126] 在一个实施例中,为了确定实际的肢体,目标识别、分析和跟踪系统可将人类目标的标识的默认姿态中的所标识的肢体(诸如右臂、左臂、左腿、右腿等)的默认方位或位置与躯干体外的体素的方位或位置进行比较。例如,在 305 接收的深度图像可被包括在捕捉设备所捕捉或观察的初始帧中。如果在 305 接收的深度图像可被包括在初始帧中,该目标识别、分析和跟踪系统可将肢体的默认方位或位置(诸如右臂、左臂、左腿、右腿等的默认

方位或位置)与躯干体外的体素的方位或位置进行比较。根据示例实施例,所标识的肢体的默认位置或方位可以是人类目标的表示(诸如默认姿态中的人类目标的完全连接的骨架或体模型)在默认姿态(诸如T型姿态、达芬奇姿态、自然姿态等)中的肢体的位置或方位。基于所述比较,该目标识别、分析和跟踪系统随后可将躯干体外的体素与与该默认姿态相关联的最接近的肢体相关联。例如,该目标识别、分析和跟踪系统可将包括躯干体外的体素中的每一个的X值、Y值和深度值的方位或位置与包括默认的肢体(诸如默认的左臂、右臂、左腿、右腿等)的X值、Y值和深度值的默认方位或位置进行比较。该目标识别、分析和跟踪系统随后可基于该比较将躯干体外的体素中的每一个与可能具有最接近的位置或方位的默认的肢体相关联。

[0127] 该目标识别、分析和跟踪系统还可基于所估算的肢体来重新标记躯干体内的体素。例如,在一个实施例中,臂(诸如左前臂)的至少一部分可被置于人类目标的躯干的前方。基于所标识的臂的默认方位或位置,如上所述,该目标识别、分析和跟踪系统可将该部分确定或估算为与所述臂相关联。例如,先前标识的肢的先前方位或位置可指示人类目标的肢体(诸如臂)的一个或多个体素可以在躯干体内。该目标识别、分析和跟踪系统随后可将包括先前标识的肢体(诸如先前标识的左臂、右臂、左腿、右腿等)的X值、Y值、和深度值的先前方位或位置与躯干体中所包括的体素的方位或位置进行比较。目标识别、分析和跟踪系统随后可基于该比较将躯干体内的体素中的每一个与可能具有最接近的位置或方位的先前标识的肢体相关联并将其重新标记。

[0128] 根据一个实施例,在标记与肢体相关联的体素后,目标识别、分析和跟踪系统可在320确定例如所标记的肢体的部分的位置或方位。例如,在标记与左臂、右臂、左腿和/或右腿相关联的体素之后,该目标识别、分析和跟踪系统可确定与右臂和左臂的手和/或肘、膝和/或脚、肘等的位置或方位。

[0129] 目标识别、分析和跟踪系统可基于所述肢体中的每一个的肢体平均来确定所述部分(诸如手、肘、脚、膝等)的位置或方位。例如,该目标识别、分析和跟踪系统可通过将与左臂相关联的体素的每一个的X值相加,将与左臂相关联的体素的每一个的Y值相加,以及将与左臂相关联的体素的每一个的深度值相加并将所加总的X值、Y值和深度值中的每一个的和除以与左臂相关联的体素的总数来计算左臂平均位置。根据一个实施例,该目标识别、分析和跟踪系统然后可定义左肩和左臂平均位置之间的向量或线以便该左肩和左臂平均位置之间的向量或线可定义左手的第一搜索方向。该目标识别、分析和跟踪系统然后可从肩至沿由该向量或线所定义的该第一搜索方向搜索最后有效体素或具有有效X值、Y值和/或深度值的最后体素并可该最后有效体素的位置或方位与左手相关联。

[0130] 根据另一实施例,该目标识别、分析和跟踪系统可计算锚点。目标识别、分析和跟踪系统然后可定义锚点与左臂平均位置之间的向量或线,使得锚点与左臂平均位置之间的向量或线可以定义左手的第二搜索方向。目标识别、分析和跟踪系统然后可从锚点沿由该向量或线所定义的第二搜索方向搜索最后有效体素或具有有效X值、Y值和/或深度值的最后体素并可该最后有效体素的位置或方位与左手相关联。

[0131] 在一示例实施例中,目标识别、分析和跟踪系统可基于从其他的确定的肢端(诸如头、髋、肩等)的一个或多个偏移来计算锚点的位置或方位。例如,目标识别、分析和跟踪系统可通过将肩的位置或方位在相应的X方向和Z方向上延伸与肩的位置或方位相关联的

X 值和深度值的一半来计算锚点的 X 值和深度值。该目标识别、分析和跟踪系统然后可在所延伸的位置或方位周围镜像锚点的 X 值和深度值的位置或方位。

[0132] 目标识别、分析和跟踪系统可基于左臂平均位置从头和 / 或髋的位移来计算锚点的 Y 值。例如,该目标识别、分析和跟踪系统可计算头的 Y 值和左臂平均的 Y 值之间的位移或差。该目标识别、分析和跟踪系统然后可将该位移或差加到例如髋的中心的 Y 值来计算锚点的 Y 值。

[0133] 图 15A-15C 示出了基于锚点 828a-828c 来确定的手的示例实施例。如图 15A-15C 所示,根据另一实施例,该目标识别、分析和跟踪系统可计算锚点 828a-828c。目标识别、分析和跟踪系统然后可定义锚点 828a-828c 与左臂平均位置 826a-826c 之间的向量或线,使得锚点与左臂平均位置之间的向量或线可以定义左手的第二搜索方向。目标识别、分析和跟踪系统然后可从锚点 828a-828c 沿由该向量或线所定义的第二搜索方向搜索最后有效体素或具有有效 X 值、Y 值和 / 或深度值的最后体素,并可将该最后有效体素的位置或方位与左手相关联。

[0134] 如上所述,在一示例实施例中,目标识别、分析和跟踪系统可基于从其他的确定的肢端(诸如如上所述的头、髋、肩等)的一个或多个偏移来计算锚点 828a-828c 的位置或方位。例如,目标识别、分析和跟踪系统可通过将肩的位置或方位在相应的 X 方向和 Z 方向上延伸与肩的位置或方位相关联的 X 值和深度值的一半来计算锚点 828a-828c 的 X 值和深度值。目标识别、分析和跟踪系统然后可在所延伸的位置或方位周围镜像锚点 828a-828c 的 X 值和深度值的位置或方位。

[0135] 目标识别、分析和跟踪系统可基于左臂平均位置从头和 / 或髋的位移来计算锚点 828a-828c 的 Y 值。例如,目标识别、分析和跟踪系统可计算头的 Y 值与左臂平均 826a-826c 的 Y 值之间的位移或差。目标识别、分析和跟踪系统然后可将该位移或差加到例如髋的中心的 Y 值来计算锚点 828a-828c 的 Y 值。

[0136] 回头参考图 5,根据示例实施例,目标识别、分析和跟踪系统可以在 320 计算右臂平均位置,该右臂平均位置可用于定义如上所述的诸如第一和第二搜索方向的搜索方向,该搜索方向可用于确定右手的位置或方位。目标识别、分析和跟踪系统还可以计算可用于定义如上所述的搜索方向的左腿平均位置和右腿平均位置,该搜索方向可用于确定左脚和右脚。

[0137] 图 16 示出可基于臂和腿的平均位置和 / 或锚点来计算的手和脚的示例实施例。如图 16 所示,可以基于由如上所述的相应的臂和腿平均方位和 / 或锚点确定的第一和第二搜索方向来确定手的位置或方位 822a-b 以及脚的位置或方位 824a-b。

[0138] 回头参考图 6,在 320,目标识别、分析和跟踪系统还可基于右臂和左臂平均位置、右腿和左腿平均位置、肩、髋、头等来确定肘和膝的位置或方位。在一个实施例中,目标识别、分析和跟踪系统可通过细化左臂平均位置的 X 值、Y 值和深度值来确定左肘的位置方位。例如,目标识别、分析和跟踪系统可确定最外面的体素,所述最外面的体素可能定义与左臂相关联的边缘。目标识别、分析和跟踪系统然后可将左臂平均位置的 X 值、Y 值和深度值调整为在中间或与各边缘等距。

[0139] 目标识别、分析和跟踪系统还可在 320 确定被隔离的人类目标的附加兴趣点。例如,目标识别、分析和跟踪系统可以基于诸如肩的角度的定向来确定离人体中心最远的体

素、最接近相机的体素、人类目标最前向的体素。

[0140] 目标识别、分析和跟踪系统随后可在 320 确定为肢端（诸如头、肩、髋、手、脚等）确定的位置或方位中的一个或多个是否可能不是人类目标的实际肢端的准确位置或方位。例如，在一个实施例中，右手的位置或方位可能是不准确的，使得右手的位置或方位可能停滞在或邻近于肩或髋的位置或方位。

[0141] 根据示例实施例，目标识别、分析和跟踪系统可包括或存储各肢端的体标记（marker）的列表，所述体标记可指示该肢端的不准确的位置或方位。例如，该列表可包括可能与手相关联的、在肩或髋周围的体标记。目标识别、分析和跟踪系统可基于该列表中与手相关联的体标记来确定手的位置或方位是否可能是准确的。例如，如果手的位置或方位可在列表中与手相关联的体标记之一内，该目标识别、分析和跟踪系统可确定手的位置或方位可能是不准确的。根据一个实施例，目标识别、分析和跟踪系统然后将手的位置或方位调整到之前帧中的手的之前准确位置到手的当前位置或方位。

[0142] 在 325，目标识别、分析和跟踪系统可跟踪基于在 320 确定的肢端而生成的模型。例如，目标识别、分析和跟踪系统可生成和 / 或可包括诸如可具有其间定义一个或多个关节和骨骼的骨架模型等模型。

[0143] 图 17 示出诸如可生成的骨架模型之类的模型 900 的示例实施例。根据一示例实施例，模型 900 可包括可表示例如人类的三维模型的一个或多个数据结构。每个身体部分可以被表征为一数学向量，该数学向量具有可定义模型 900 的关节和骨骼的 X、Y 和 Z 值。

[0144] 如图 17 所示，模型 900 可包括一个或多个关节 j1-j16。根据一示例实施例，关节 j1-j16 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。例如，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，这些身体部位由一个或多个诸如“骨骼”等结构件来定义，而关节 j1-j16 位于毗邻骨骼的交叉点处。关节 j1-j16 可使得与骨骼和关节 j1-j16 相关联的各个身体部位能够彼此独立地移动。例如，如图 17 所示，关节 j10 与 j12 之间定义的骨对应于前臂，该前臂可独立于例如关节 j14 与 j16 之间的骨骼（其对应于小腿）移动。

[0145] 回头参考图 5，在 325，目标识别、分析和跟踪系统可基于在 320 为肢端确定的位置或方位来调整所生成的模型。例如，目标识别、分析和跟踪系统可将与头相关联的关节 j1 调整为与诸如在 320 确定的头的位置或方位 810 等的方位或位置相对应。因此，在一示例实施例中，关节 j1 可被赋予与如上所述的为头确定的位置或方位 810 相关联的 X 值、Y 值和深度值。如果一个或多个肢端基于例如如上所述的体标记的列表可能是不准确的，则目标识别、分析和跟踪系统可以基于之前的帧将不准确的关节保持在其之前的位置或方位。

[0146] 另外，如果目标识别、分析和跟踪系统可能不具有肢端的位置或方位，则该目标识别、分析和跟踪系统可以使用基于默认姿态（诸如 T 型姿态、达芬奇姿态等）的位置或方位。例如，目标识别、分析和跟踪系统可以吸附或调整模型的一个或多个关节 j1-j16，使其与默认姿态中最接近的体素的 X 值、Y 值、和 / 或深度值相关联。

[0147] 在 325，目标识别、分析和跟踪系统还可以基于在 330 确定的一个或多个身体测量（将在下面详细地描述）来调整对在模型的各关节之间定义的一个或多个骨骼的测量。例如，目标识别、分析和跟踪系统可在 330 确定人类目标的左前臂的长度。在 325，目标识别、分析和跟踪系统随后可以调整与左前臂相关联的骨骼的长度以模拟在 330 为人类目标确

定的左前臂的长度。例如，目标识别、分析和跟踪系统可以调整关节 j10 和 j12 的 X 值、Y 值和深度值（或 Z 值）中的一个或多个，使得在其间定义的骨骼可以等于在 330 为人类目标的左前臂确定的长度。

[0148] 在 325，目标识别、分析和跟踪系统还可以检查与所调整的模型相关联的关节的无效位置或方位。例如，在一个实施例中，目标识别、分析和跟踪系统可检查以确定诸如关节 j10 的关节是否可能凸出，使得该模型可能卡在小鸡舞的姿态。因此，在 325，目标识别、分析和跟踪系统可检查模型的已知位置或方位，其中该模型可能以不适当的方式折叠，诸如与肘相关联的关节。

[0149] 根据一个实施例，目标识别、分析和跟踪系统还可基于在 305 接收的非降采样的深度图像的 2D 像素区域中的 X 值、Y 值和深度值来细化关节的位置或方位。例如，在一个实施例中，目标识别、分析和跟踪系统可使用来自该非降采样的深度图像的数据来细化模型可能折叠之处的模型的关节的位置或方位。

[0150] 另外，目标识别、分析和跟踪系统可使用来自该非降采样的深度图像的数据来细化与频繁使用的姿势相关联的模型的关节的位置或方位。例如，根据一个实施例，该目标识别、分析和跟踪系统可对与手相关联的关节区分优先级。目标识别、分析和跟踪系统可本地化在 305 接收的非降采样的深度图像中的手周围的数据，使得目标识别、分析和跟踪系统可使用在 305 接收的非降采样的深度图像中的更高分辨率的数据来修改在 320 确定的手的位置或方位。

[0151] 如上所述，在 330，目标识别、分析和跟踪系统可扫描与被隔离的人类目标相关联的体素以确定与其相关联的肢端的尺寸。例如，可扫描被隔离的人类目标来确定例如与肢端（诸如臂、腿、头、肩、髋、躯干等）相关联的测量（诸如长度、宽度等）。

[0152] 为了确定尺寸，在 330，目标识别、分析和跟踪系统可生成为每个肢端建议的关节尺寸的估算器。目标识别、分析和跟踪系统可使用所述估算器来计算所建议的关节尺寸中的每一个的均值和标准差。目标识别、分析和跟踪系统可添加在定义的百分比偏差内的所建议的关节尺寸，而可拒绝离群点 (outlier) 以及在定义的百分比偏差外的所建议的关节尺寸。目标识别、分析和跟踪系统随后可基于在其标准差和所建议的关节尺寸的数量之间可具有最高比的估算器来确定所述肢端的尺寸。

[0153] 也可在 330 更新扫描所确定的与肢端相关联的尺寸。根据一个实施例，目标识别、分析和跟踪系统可包括一个或多个试探 (heuristic) 或规则以确定扫描所确定的尺寸是否可能是正确的。例如，目标识别、分析和跟踪系统包括可确定对称关节之间的欧几里德距离是否大致相等的试探或规则，可确定手和 / 或肘是否靠近身体的试探或规则，可确定头是否可锁定在一方位或位置的试探和 / 或规则、可确定手是否靠近头的试探和 / 或规则，或者可用于调整所述尺寸的其他试探或规则。如上所述，在 325，在 330 确定的尺寸可用于调整可被跟踪用于下一帧的模型。

[0154] 在 335，随后可以处理经调整的模型。例如，在一个实施例中，目标识别、分析和跟踪系统可以通过例如将应用于经调整的模型的一个或多个运动或移动映射到化身或游戏角色来处理该经调整的模型，使得该化身或游戏角色可被动画化以便模拟诸如以上参考图 1A 和 1B 描述的用户 18 等用户。例如，然后可响应于对经调整的模型的改变来改变屏幕上人物的视觉外观。

[0155] 在一个实施例中,经调整的模型可以通过将经调整的模型提供给诸如参照图 1A-4 描述的计算环境 12 等计算环境中的姿势库,来处理经调整的模型。姿势库可用于基于骨架模型中各身体部分的方位来确定在应用内执行的控制。

[0156] 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。此处所述的具体例程或方法可表示任何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,可以改变上述过程的次序。

[0157] 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、和此处所公开的其他特征、功能、动作、和 / 或特性、以及其任何和全部等效物。

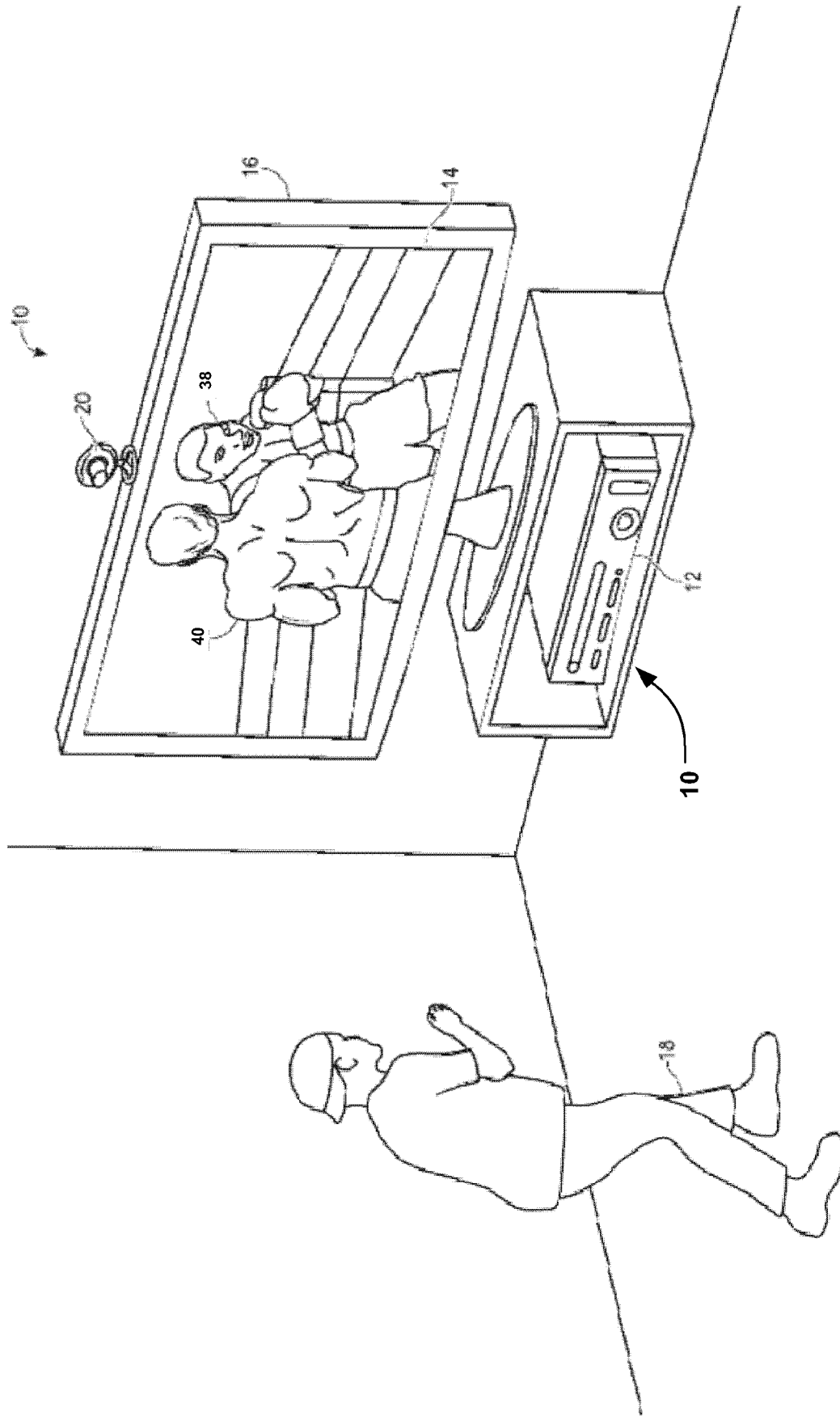


图 1A

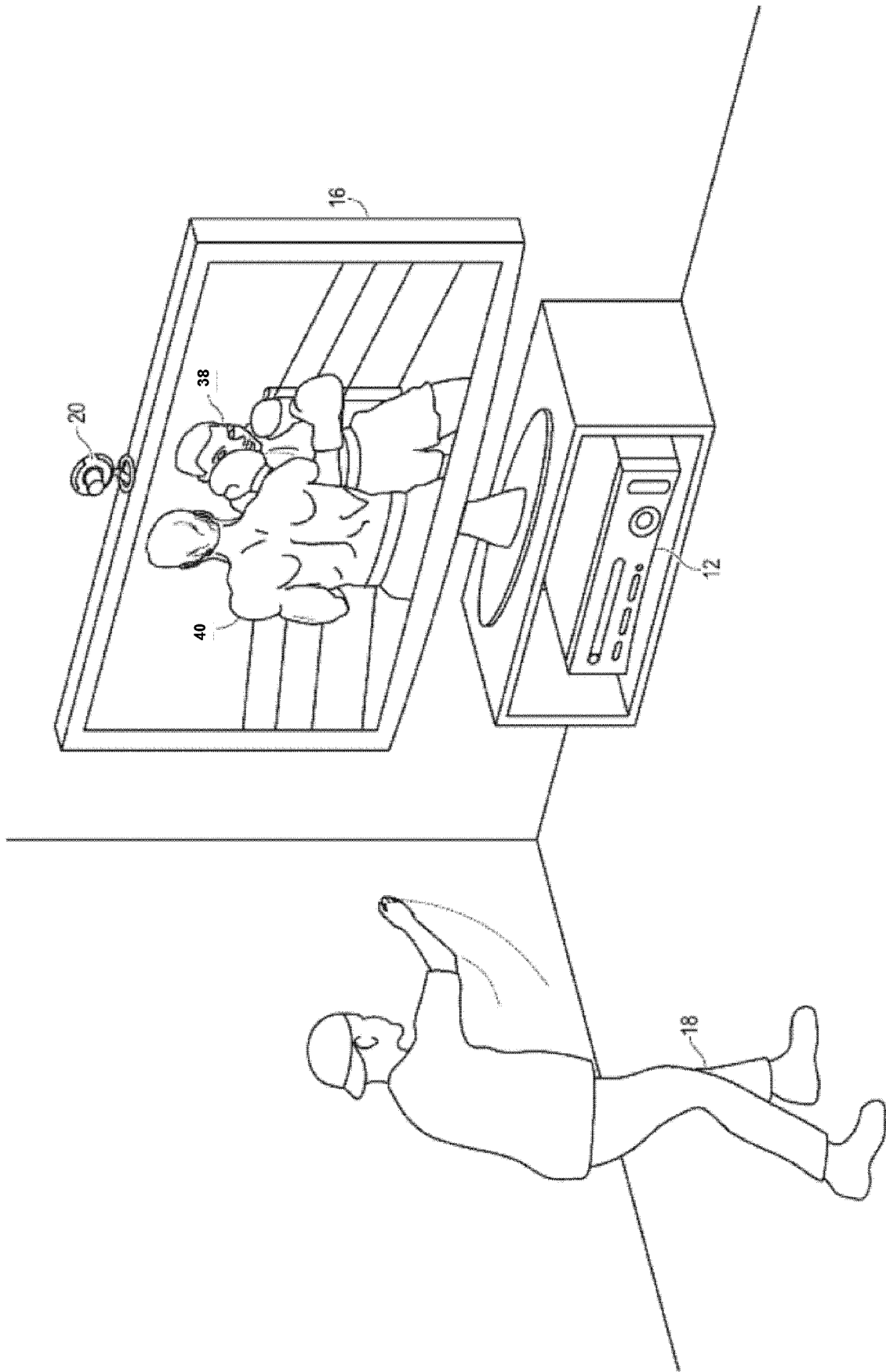


图 1B

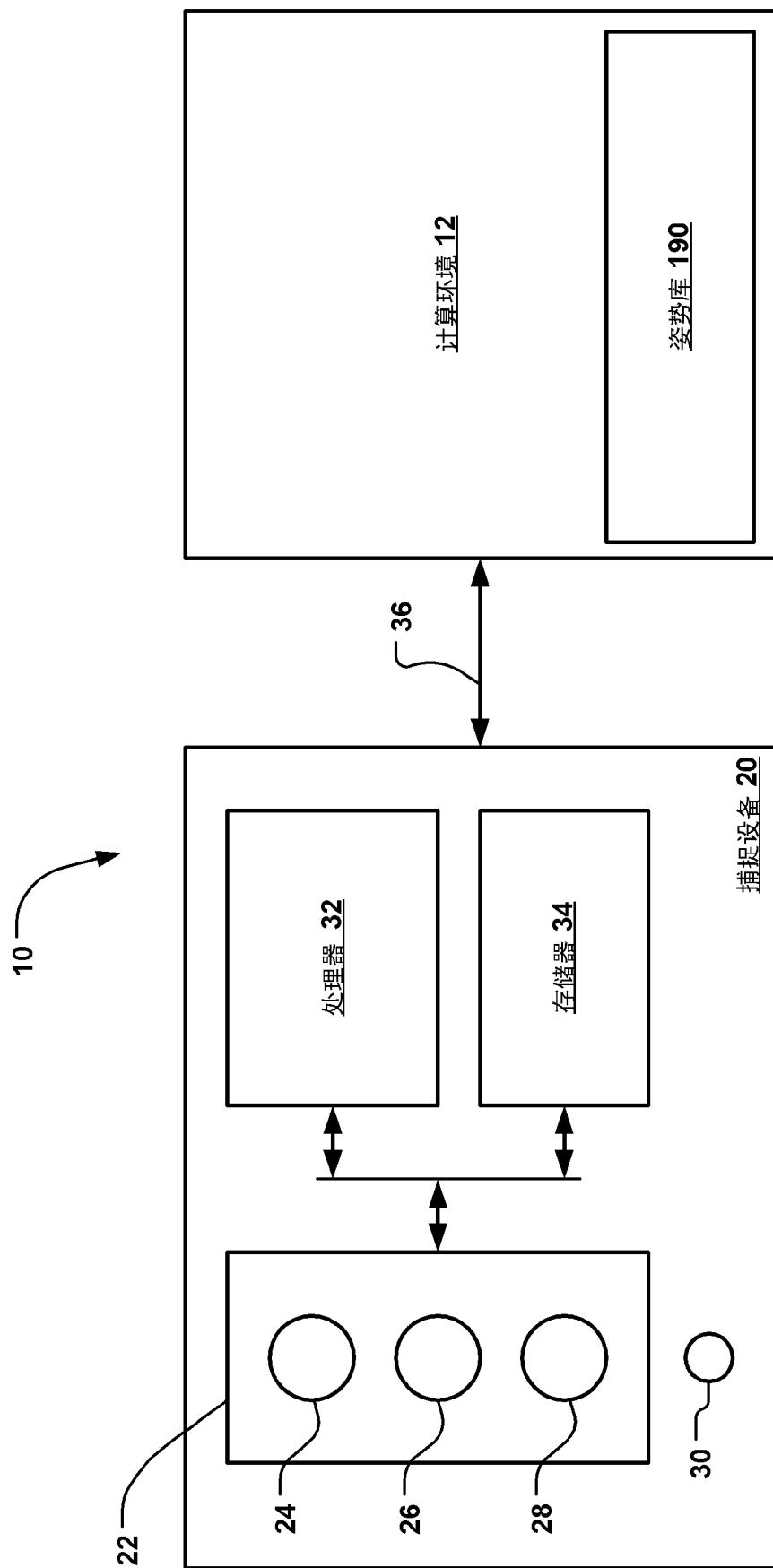


图 2

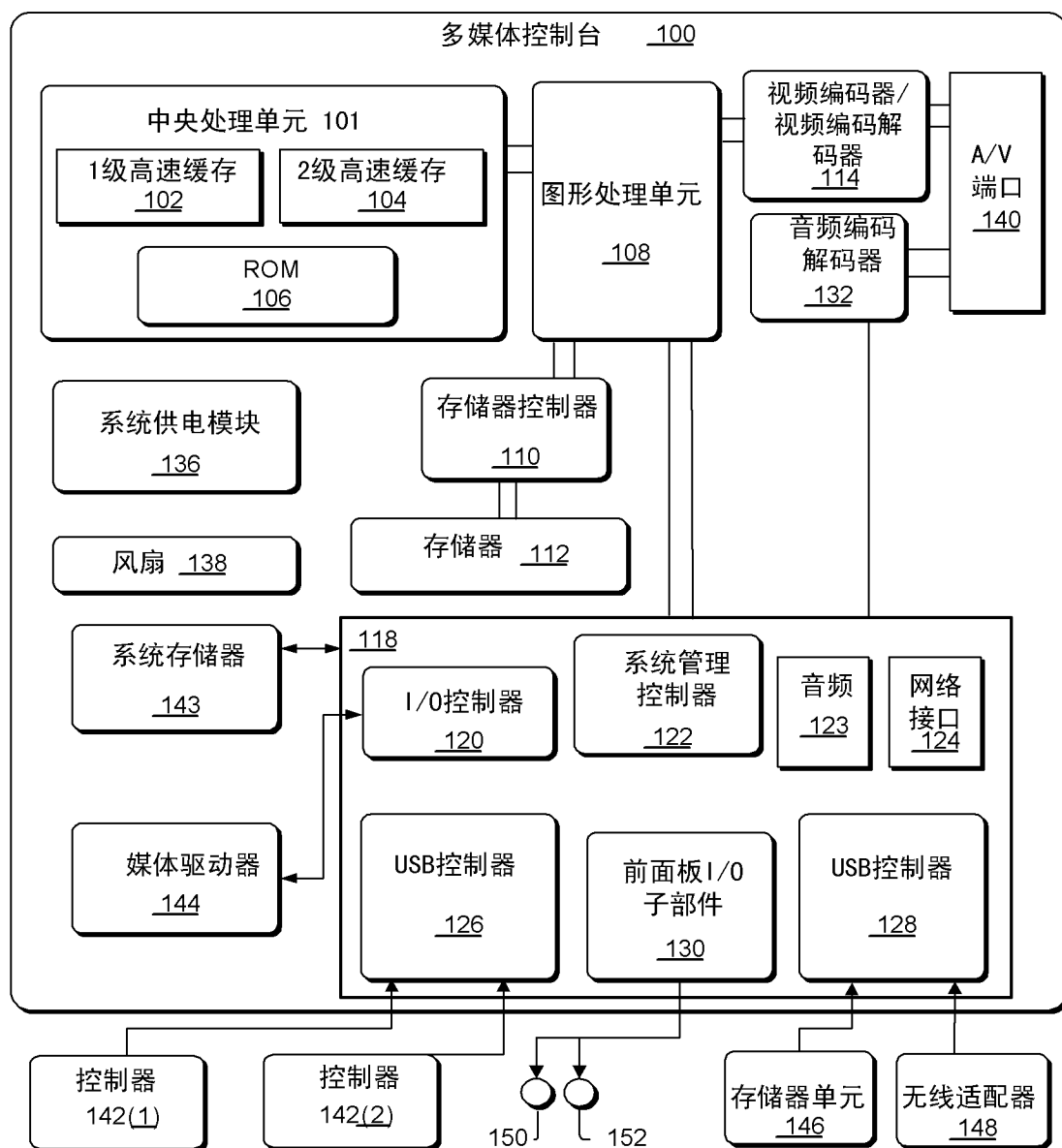


图 3

计算环境 220

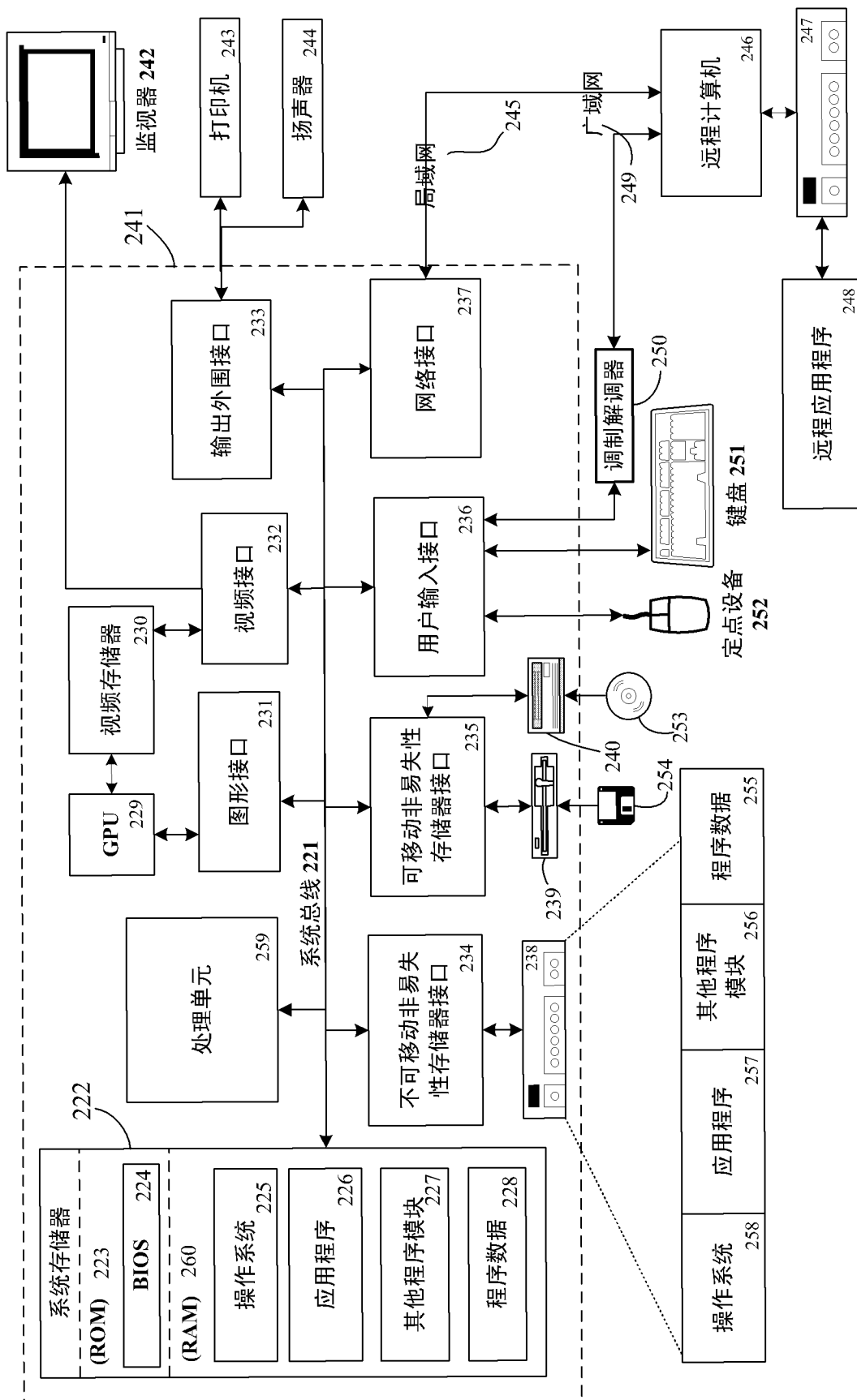


图 4

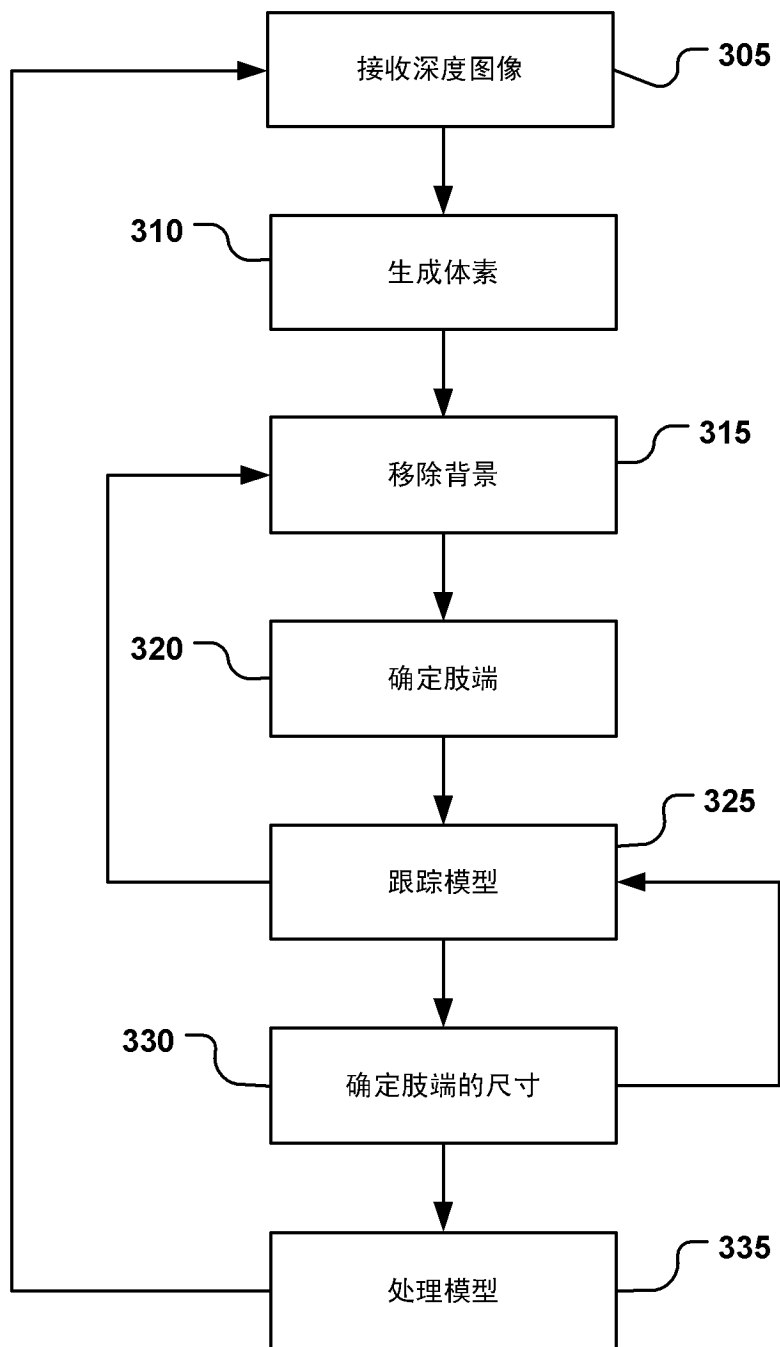
300

图 5

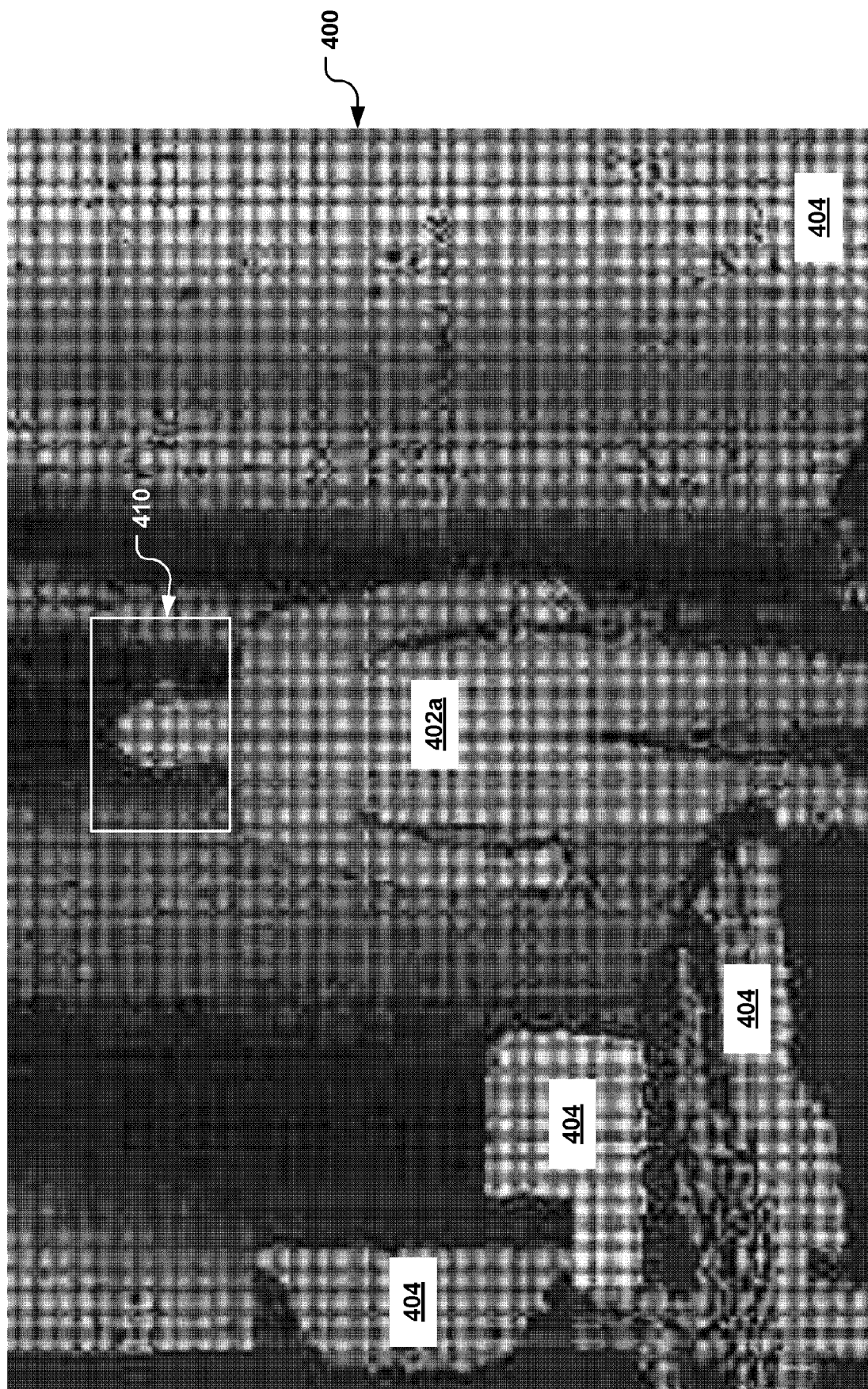


图 6

430

420

410

30	15	15	15	16	16	16	18	18	18	18	18	18	20	20	20	20	20	20	20	20	30	30		
30	15	15	15	16	16	16	18	18	18	18	18	18	20	20	20	20	20	20	20	20	30	30		
30	15	15	15	16	16	16	18	18	18	18	18	18	20	20	20	20	20	20	20	20	30	30		
30	15	15	15	16	16	16	18	18	18	18	18	18	9	20	20	19	19	20	20	20	30	30		
30	15	15	15	16	16	16	18	8	8	9	9	9	9	9	9	9	9	19	19	20	30	30		
30	15	15	15	16	16	16	8	8	7	8	8	8	8	8	8	8	8	9	9	9	19	20	30	30
30	15	15	16	16	16	8	7	7	6	7	7	7	7	7	7	7	8	8	9	9	9	20	30	30
30	15	15	16	16	8	8	6	6	5	5	5	5	5	6	7	7	7	8	8	9	19	20	30	30
30	15	15	16	7	7	7	6	5	5	5	5	5	5	5	7	6	7	8	8	9	19	20	30	30
30	15	15	16	7	7	7	5	5	5	5	5	5	5	5	5	6	7	8	8	9	9	19	30	30
30	15	15	16	7	7	6	5	5	5	5	5	5	5	5	5	6	7	8	8	9	9	19	30	30
30	15	15	7	7	6	6	5	5	5	5	5	5	5	5	5	6	7	8	8	9	19	19	30	30
30	15	15	6	6	5	6	5	5	5	4	4	5	4	4	6	7	8	8	9	19	20	30	30	
30	15	15	5	6	5	3	5	5	5	5	5	5	5	5	5	6	7	8	9	19	20	30	30	
30	15	15	4	5	4	4	5	5	5	5	5	5	5	5	5	6	7	8	9	19	19	30	30	
30	15	15	5	5	5	5	5	5	5	5	5	5	5	5	5	6	7	8	18	18	19	30	30	
30	15	15	15	5	4	5	5	5	5	5	5	5	5	5	5	6	7	8	18	18	18	30	30	
30	15	15	15	5	5	5	5	5	5	5	5	5	5	5	5	6	7	8	18	18	18	30	30	
30	15	15	15	15	5	5	5	5	5	5	5	5	5	5	5	6	7	8	18	18	18	30	30	
30	15	15	15	15	15	5	5	5	5	5	5	5	5	5	5	6	6	7	8	18	18	18	30	30
30	15	15	15	15	15	5	5	5	5	5	5	5	5	5	6	7	8	18	18	18	18	30	30	
30	15	15	15	15	15	5	5	5	5	5	5	5	5	6	7	7	18	18	18	18	18	30	30	
30	15	15	15	15	15	5	5	5	5	5	5	5	6	6	7	7	18	18	18	18	18	30	30	
30	15	15	15	15	15	5	5	5	5	5	5	5	6	6	7	7	18	18	18	18	18	30	30	

图 7A

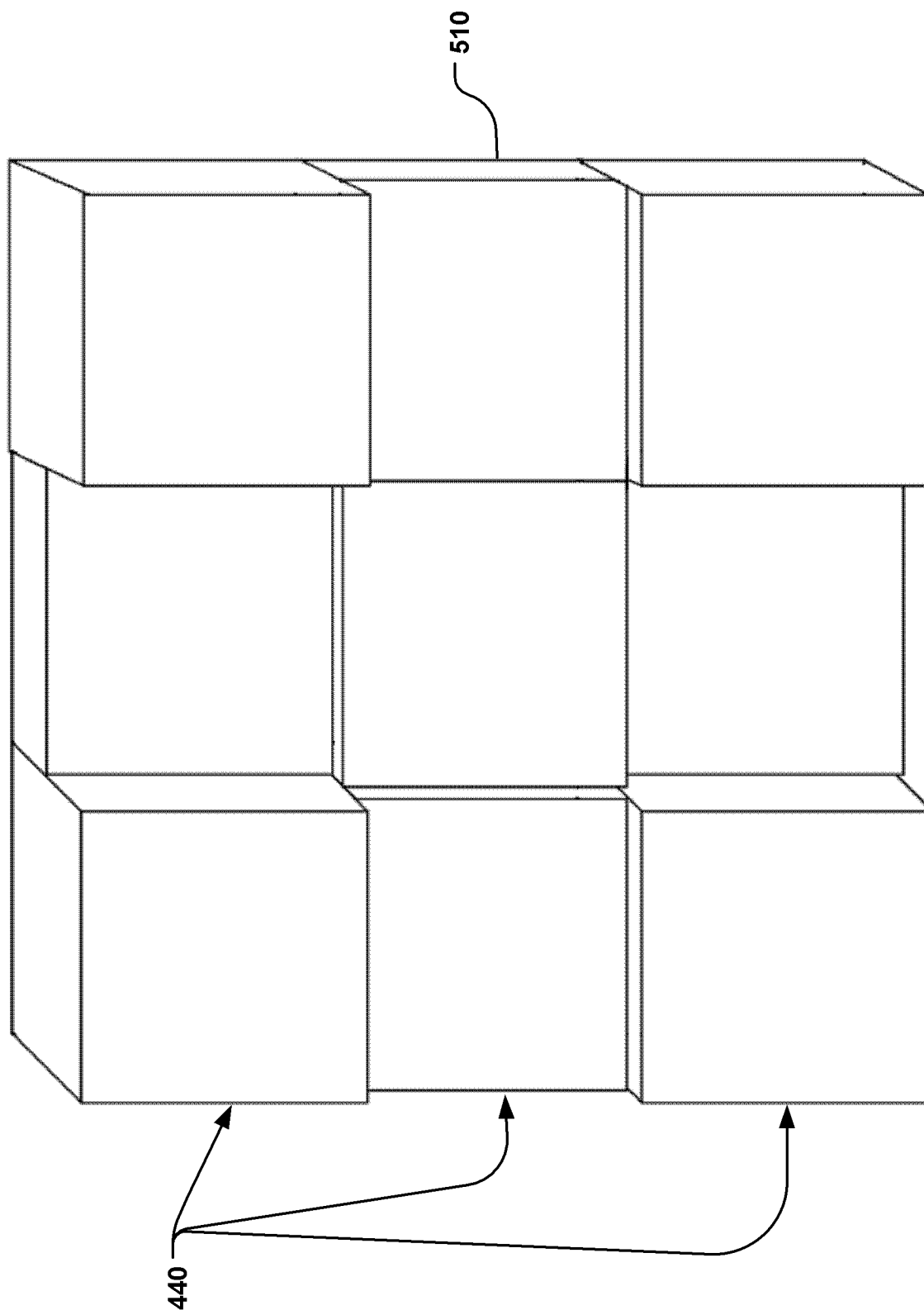


图 7B

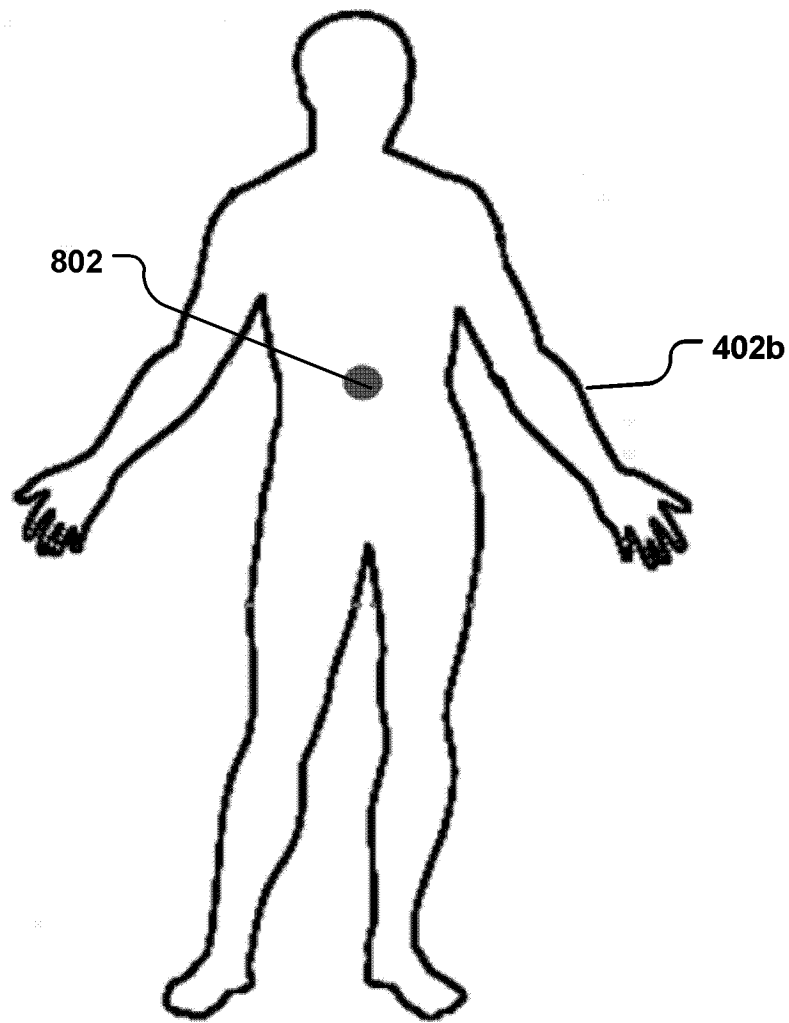


图 8

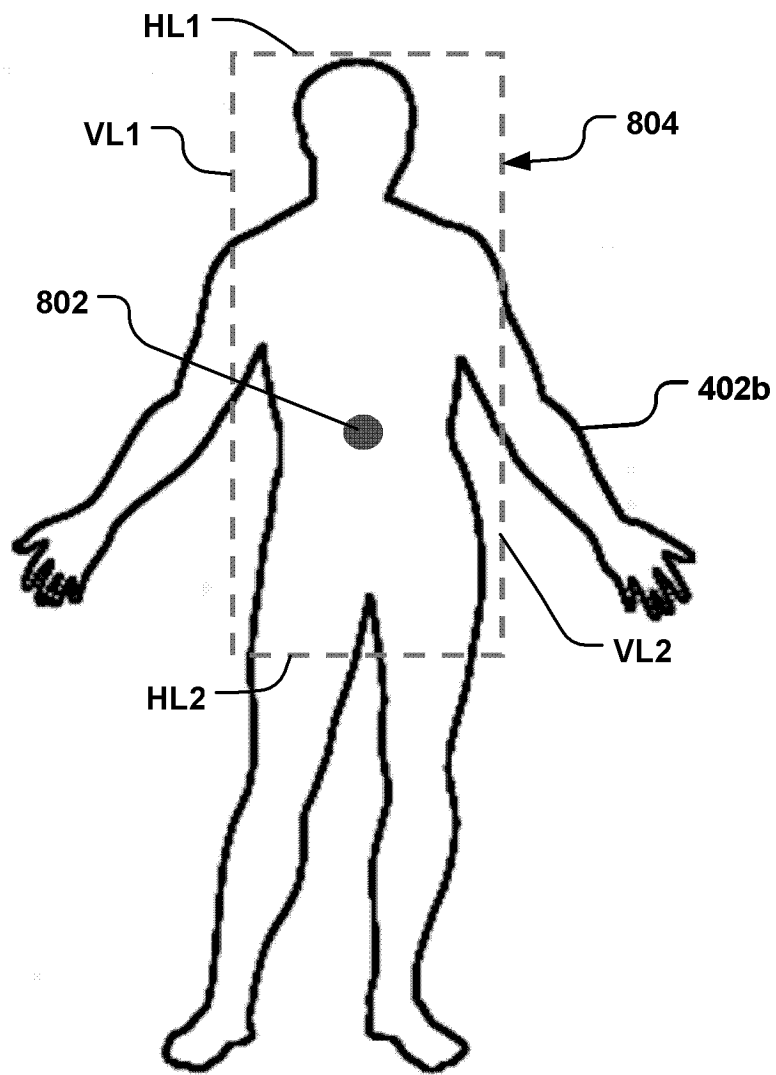


图 9

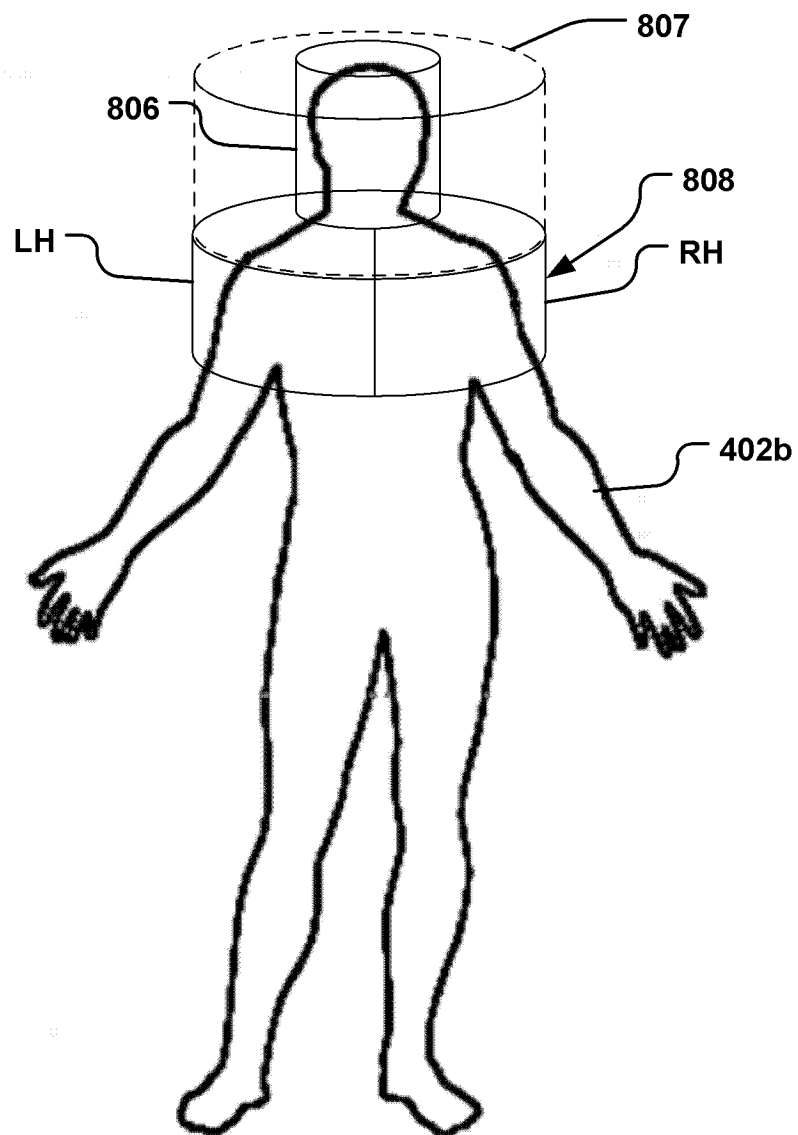


图 10

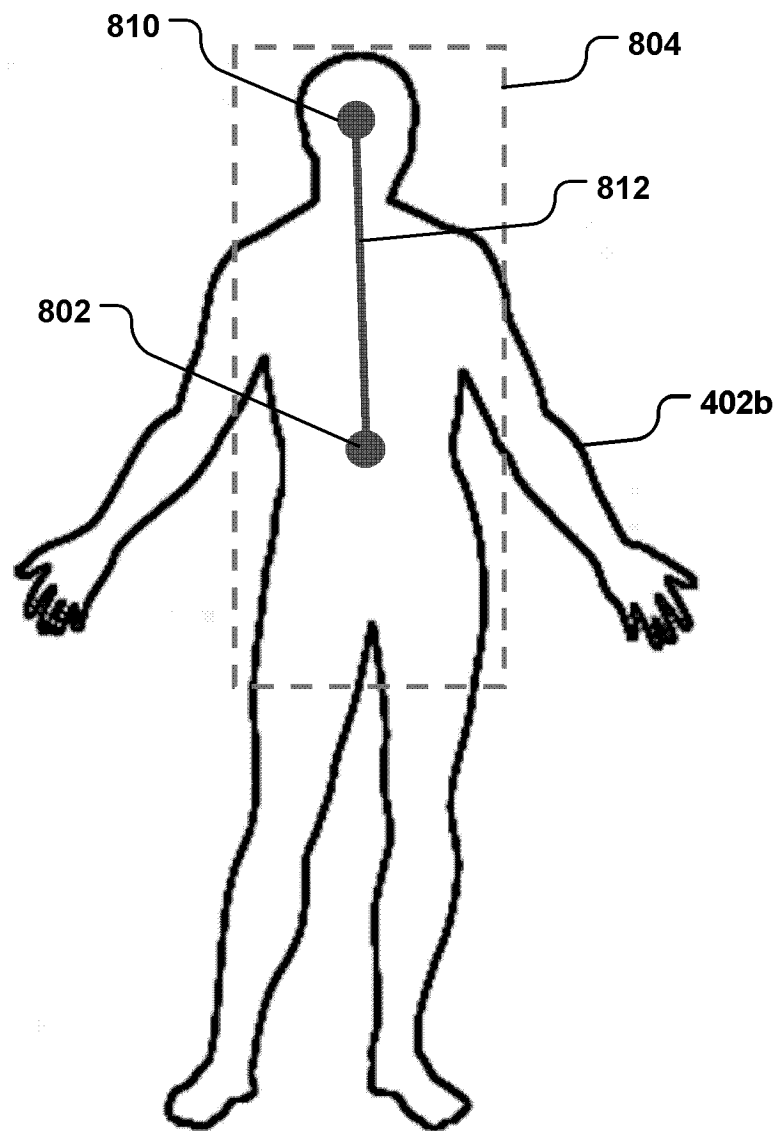


图 11

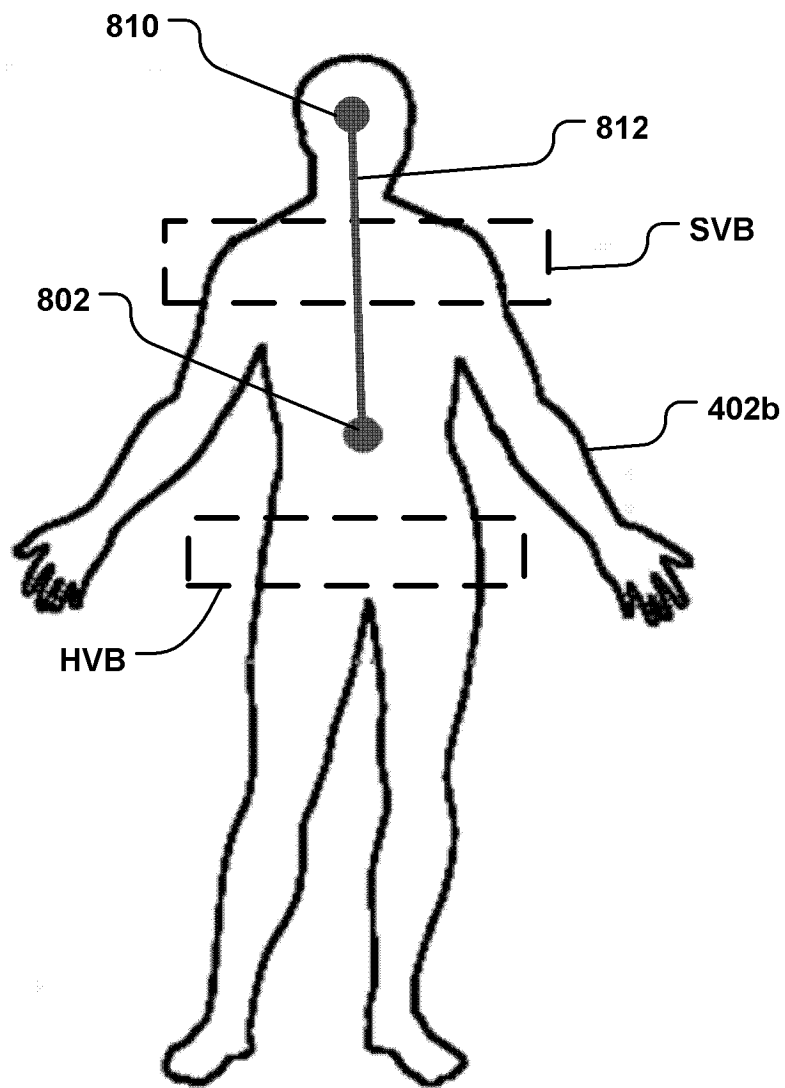


图 12

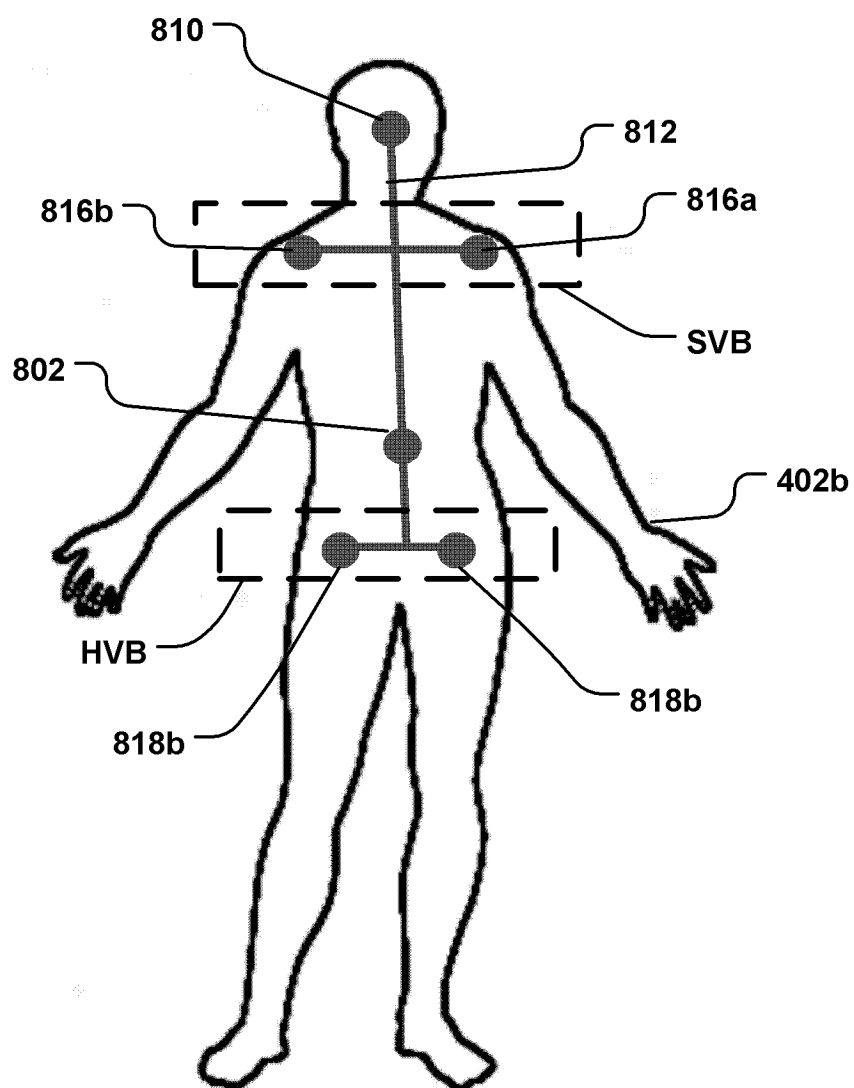


图 13

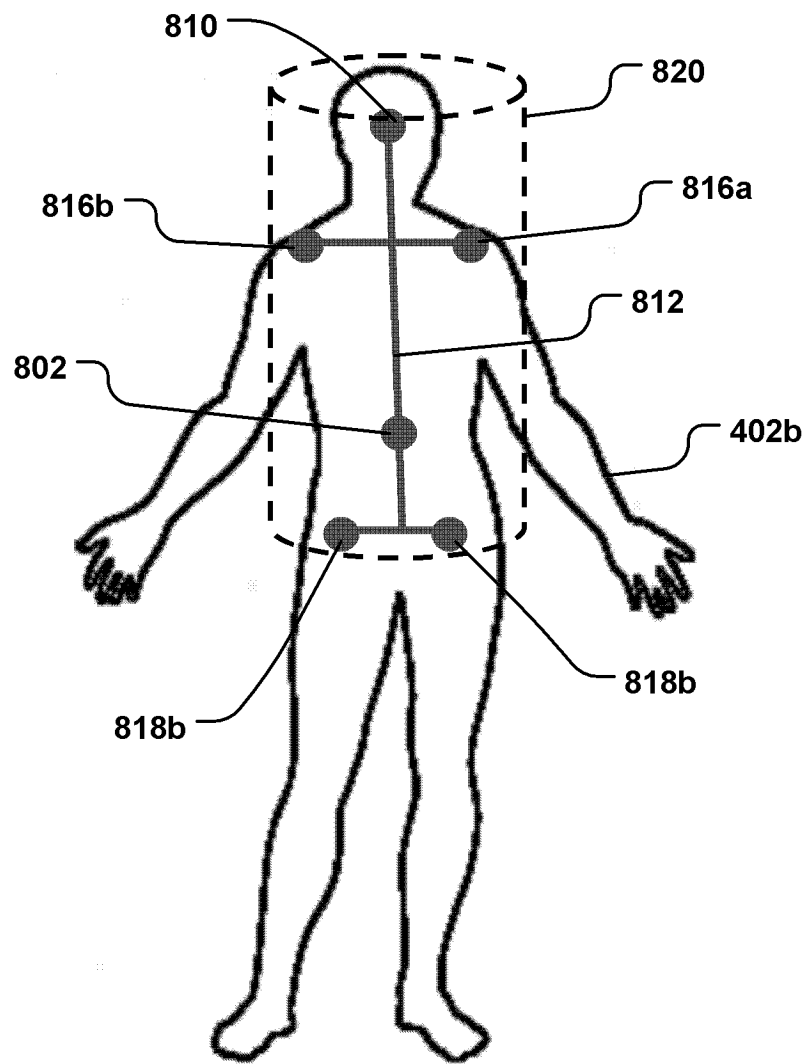


图 14

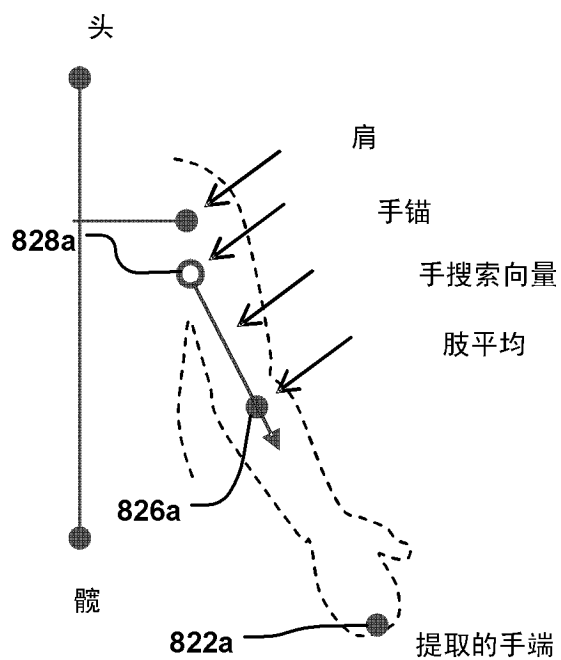


图 15A

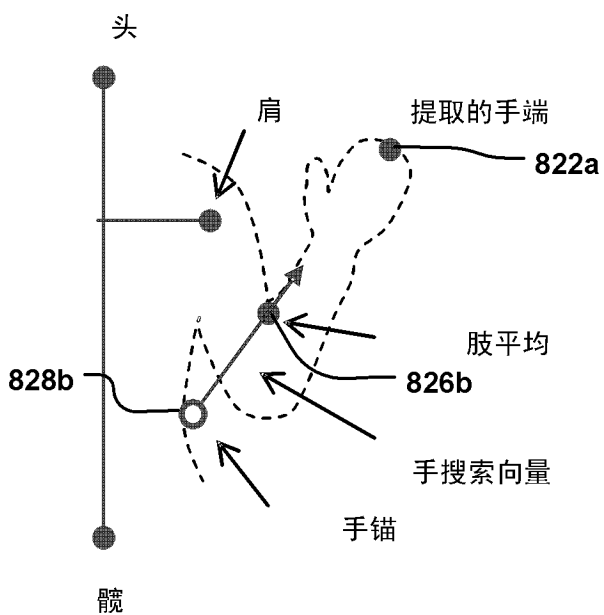


图 15B

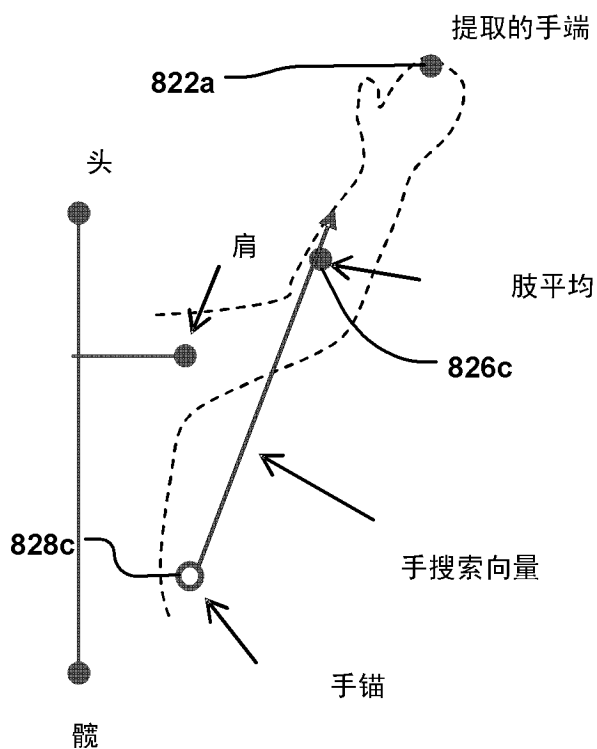


图 15C

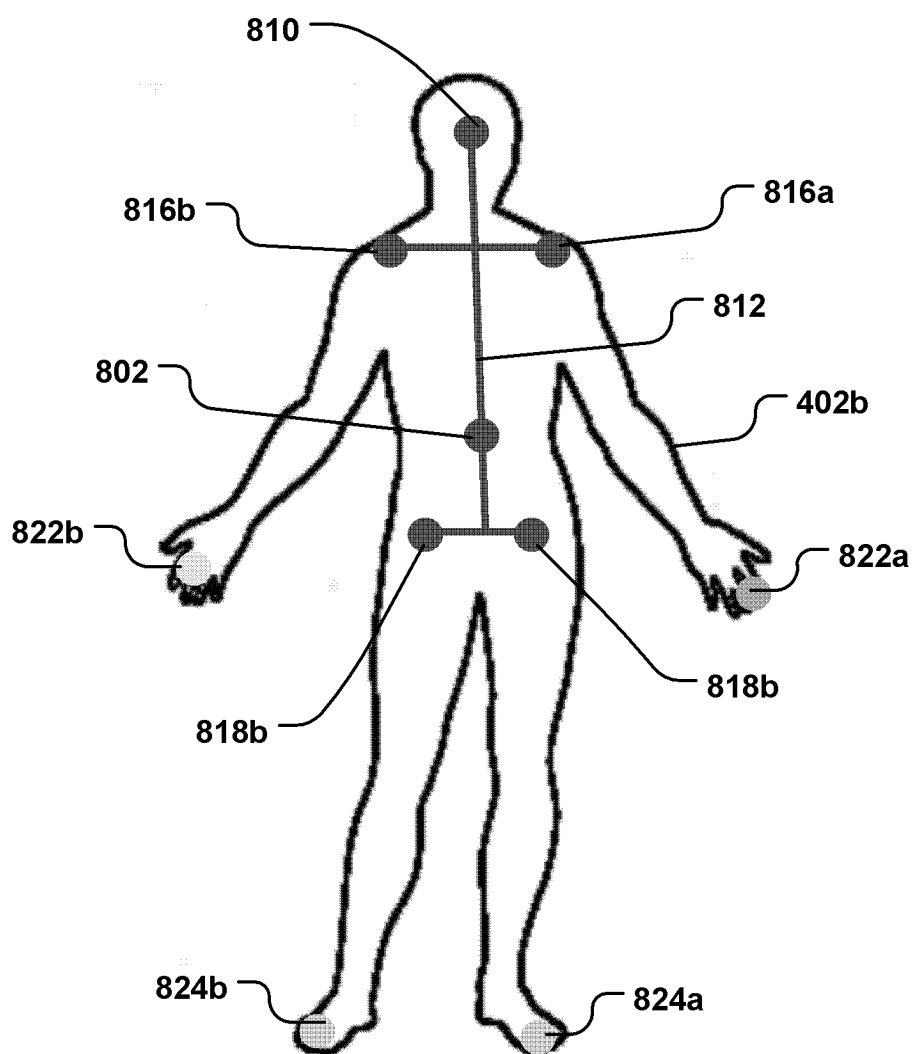


图 16

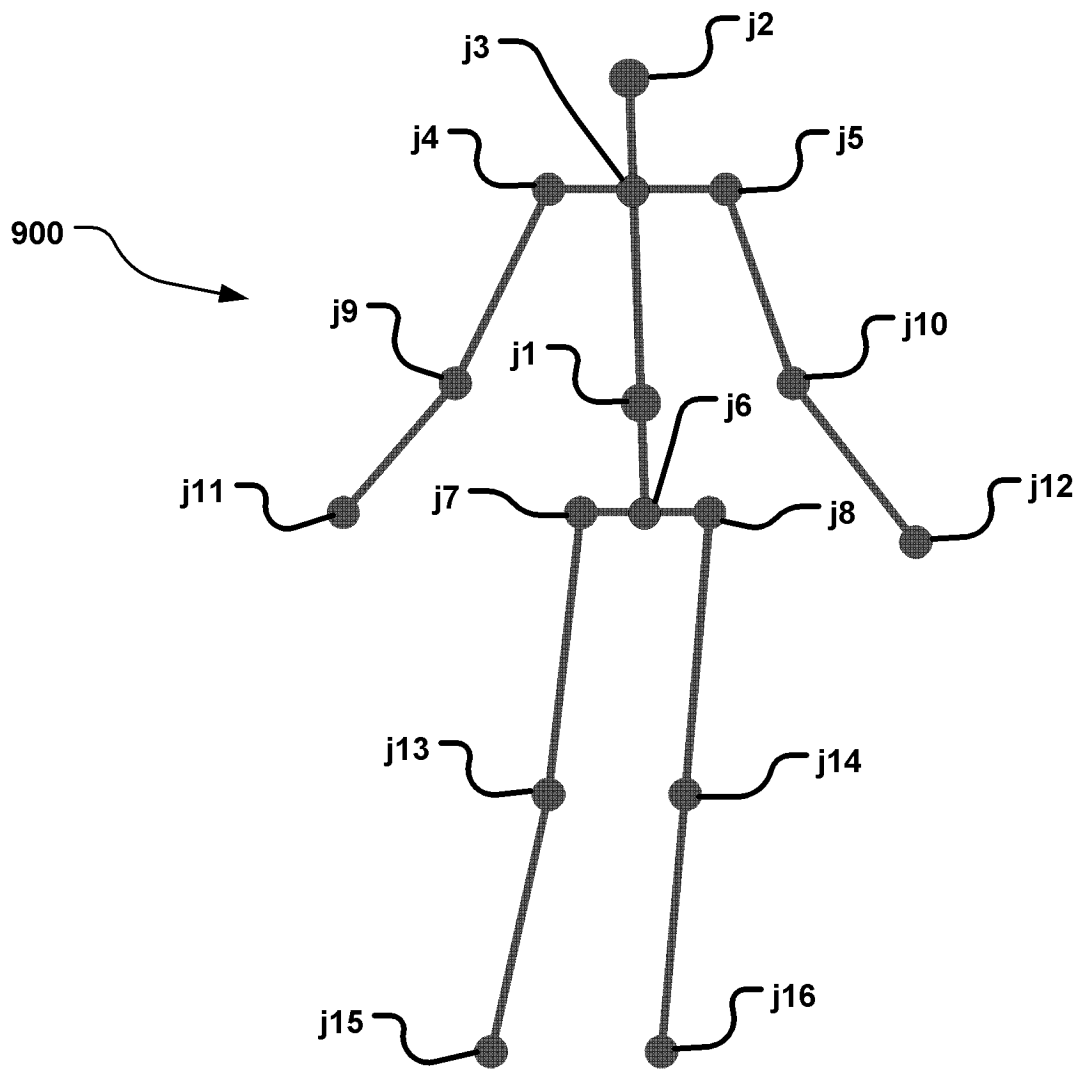


图 17