



(12) 发明专利

(10) 授权公告号 CN 102194105 B

(45) 授权公告日 2014. 03. 19

(21) 申请号 201110071738. 3

(22) 申请日 2011. 03. 18

(30) 优先权数据

12/727, 787 2010. 03. 19 US

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 A·费茨吉本 J·肖顿 M·科克

R·摩尔 M·芬诺基奥

(56) 对比文件

CN 101579238 A, 2009. 11. 18, 全文.

CN 101238981 A, 2008. 08. 13, 全文.

US 2006274947 A1, 2006. 12. 07, 全文.

US 6674877 B1, 2004. 01. 06, 全文.

US 2006098865 A1, 2006. 05. 11, 全文.

US 6999601 B2, 2006. 02. 14, 全文.

审查员 韩燕

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 钱静芳

(51) Int. Cl.

G06K 9/00 (2006. 01)

G06K 9/62 (2006. 01)

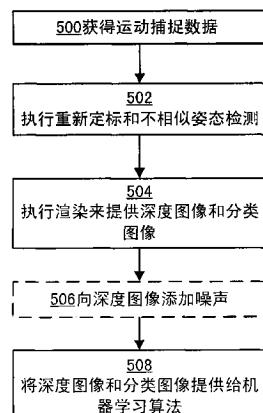
权利要求书2页 说明书18页 附图18页

(54) 发明名称

用于人体跟踪的代表训练数据

(57) 摘要

本发明公开了一种生成用于人体跟踪的代表训练数据的方法。为身体关节跟踪系统的机器学习算法生成合成身体图像。来自运动捕捉序列的帧被重新定标到若干不同的身体类型，来充分利用该运动捕捉序列。为了避免向机器学习算法提供冗余或相似的帧，并且为了提供紧凑但仍高度变化的图像集，可使用相似度度量来标识不相似的帧。该相似度度量用于定位根据阈值距离足够不同的帧。出于真实性起见，基于真实世界深度相机通常经历的噪声源来向深度图像添加噪声。也可引入其他随机变化。例如，可向重新定标添加一定程度的随机性。对于每一帧，提供深度图像和具有标记的身体部位的对应的分类图像。也可提供3-D场景元素。



1. 一种用于生成用于人体跟踪的代表训练数据的处理器实现的方法,包括以下处理器实现的步骤:

访问至少一个运动捕捉序列(500),所述至少一个运动捕捉序列标识其中行动者执行移动的时间段期间所述行动者的身体的姿态(1100);

基于所述至少一个运动捕捉序列执行到多个身体类型的重新定标以及执行不相似姿态选择,来提供多个不相似姿态(502),所述不相似姿态是彼此相隔至少一阈值距离的姿态;以及

根据用于所述多个身体类型中的一相应身体类型的3-D身体模型(1302、1360)来渲染每一所述不相似姿态,以提供所述不相似姿态的相应的深度图像,并提供标识所述不相似姿态的身体部位的相应的分类图像(1400),其中所述3-D身体模型是基于对应骨架的关节位置和关节角度中的至少一个来渲染的,使用多个3-D身体模型,对每一身体类型有一个3-D身体模型,并且相应的深度图像和相应的分类图像包括可由机器学习算法用于人体跟踪的像素数据。

2. 如权利要求1所述的处理器实现的方法,其特征在于:

所述重新定标将所述至少一个运动捕捉序列重新定标到所述多个身体类型,以提供多个重新定标的运动捕捉序列,每一重新定标的运动捕捉序列标识所述多个身体类型中的一相应身体类型的姿态;以及

所述不相似姿态选择从所述多个重新定标的运动捕捉序列中选择不相似姿态,使得所述不相似姿态是所述多个重新定标的运动捕捉序列中的姿态的子集。

3. 如权利要求1所述的处理器实现的方法,其特征在于:

所述不相似姿态选择从所述至少一个运动捕捉序列中选择不相似姿态,使得所述不相似姿态是所述至少一个运动捕捉序列中的姿态的子集;以及

所述重新定标将所述不相似姿态重新定标到所述多个身体类型,以提供多个重新定标的运动捕捉序列,每一重新定标的运动捕捉序列标识所述多个身体类型中的一相应身体类型的姿态。

4. 如权利要求1所述的处理器实现的方法,其特征在于,还包括以下处理器实现的步骤:

向至少一个所述不相似姿态引入变化,所述变化包括变化的骨长度和变化的关节角度中的至少一个。

5. 如权利要求1所述的处理器实现的方法,其特征在于:

所述渲染向至少一个所述不相似姿态添加3-D场景元素(1304、1306、1308、1310);以及

所述3-D场景元素中的至少一个被修改成适合所述不相似姿态的对应的身体姿态。

6. 如权利要求5所述的处理器实现的方法,其特征在于,还包括以下处理器实现的步骤:

通过组合来自所述至少一个运动捕捉序列的两个或更多运动捕捉帧的关节角度,来扩展所述行动者的身体的姿态以包括至少一个附加姿态。

7. 如权利要求1所述的处理器实现的方法,其特征在于:

所述渲染通过改变虚拟相机高度、虚拟相机角度和虚拟相机视野中的至少一个来改变

至少一个所述不相似姿态。

8. 如权利要求 1 所述的处理器实现的方法,其特征在于:

所述渲染通过改变至少一个所述不相似姿态的身体位置来改变至少一个所述不相似姿态。

9. 如权利要求 1 所述的处理器实现的方法,其特征在于:

所述渲染向至少一个所述不相似姿态添加 3-D 场景元素;以及

所述渲染组合不同姿态中的、但具有共同的虚拟相机位置和共同的 3-D 场景元素的两个或更多 3-D 身体模型。

10. 一种用于生成用于人体跟踪的代表训练数据的处理器实现的方法,包括以下处理器实现的步骤:

访问至少一个运动捕捉序列(500),所述至少一个运动捕捉序列标识其中行动者执行移动的时间段期间所述行动者的身体的姿态(1100);

基于所述至少一个运动捕捉序列执行不相似姿态选择以及执行到多个身体类型的重新定标,来提供多个不相似姿态(502),所述不相似姿态是彼此相隔至少一阈值距离的姿态;以及

根据用于所述多个身体类型中的一相应身体类型的 3-D 身体模型(1302、1360)来渲染每一所述不相似姿态,以提供所述不相似姿态的相应的深度图像,并提供标识所述不相似姿态的身体部位的相应的分类图像(1400),其中使用多个 3-D 身体模型,对每一身体类型有一个 3-D 身体模型,并且相应的深度图像和相应的分类图像包括可由机器学习算法用于人体跟踪的像素数据,并且所述渲染向至少一个所述不相似姿态的深度图像添加噪声(506)。

11. 如权利要求 10 所述的处理器实现的方法,其特征在于:

所述噪声是通过向至少一个所述不相似姿态的深度图像添加头部毛发(902)来添加的。

12. 如权利要求 10 所述的处理器实现的方法,其特征在于:

所述噪声是通过量化至少一个所述不相似姿态的深度图像的深度分辨率(904)来添加的,其中量化的精度取决于深度图像的深度。

13. 如权利要求 10 所述的处理器实现的方法,其特征在于:

所述噪声是通过标识至少一个所述不相似姿态的深度图像中的至少一个边缘(908),并将所述至少一个边缘处的至少一个像素的深度替换为所述至少一个边缘处的至少一个像素的更远离所述至少一个边缘的相邻像素的深度来添加的。

14. 如权利要求 10 所述的处理器实现的方法,其特征在于:

所述噪声是通过标识比至少一个所述不相似姿态的深度图像中的指定数量的像素更窄的至少一个薄结构(910),并用背景像素替换所述至少一个薄结构的像素来添加的。

15. 如权利要求 10 所述的处理器实现的方法,其特征在于:

所述噪声是通过标识至少一个所述不相似姿态的深度图像中具有大于阈值的深度不连续性的第一和第二边缘(912),并用背景像素替换所述第一和第二边缘之间的像素来添加的。

## 用于人体跟踪的代表训练数据

### 技术领域

[0001] 本发明涉及运动捕捉系统，尤其涉及运动捕捉系统中的人体跟踪。

### 背景技术

[0002] 跟踪系统获得关于人或其他对象在物理空间中的位置和移动的数据，并且可使用该数据作为计算系统中的某一应用的输入。某些系统确定身体的骨架模型，包括该骨架的关节，并且因此可被认为是身体关节跟踪系统。可能有许多应用，如出于军事、娱乐、体育和医疗目的。例如，人的运动可用于创建动画人物或化身。包括使用可见和不可见（例如，红外）光的系统在内的光学系统使用相机来检测视野中的人的存在。然而，需要通过提供合成图像形式的训练数据来方便身体关节跟踪系统的开发。

### 发明内容

[0003] 提供了用于生成用于身体关节跟踪系统中的人体跟踪的代表训练数据的处理器实现的方法、系统和有形计算机可读存储。

[0004] 在身体关节跟踪系统的开发中，使用深度相机来获得在该相机的视野中移动的人的深度图像。使用各种处理技术来检测该人的身体，并识别该人执行的移动或姿态。该过程可被认为是监督机器学习算法。该过程是监督的是因为人的位置和姿态都是已知的。目标是使得身体关节跟踪系统了解如何识别该人的位置和姿态。可对该学习算法作出各种调整，例如，过滤掉噪声、识别不同的身体类型、以及将该人的身体与可能存在于视野中的其他物体（如家具、墙等）进行区分。然而，使用真实世界环境中的活人来训练该学习算法是低效的，且无法准确地表示身体关节跟踪系统在其被部署为数千或甚至数百万用户家庭中的商用产品时将经历的各种场景。

[0005] 为了优化对学习算法的训练，可生成合成图像来作为用于真实的人的图像的代替或代表。该合成图像可用于扩充或替换真实的人的图像。此外，该合成图像可以用在计算上高效的方式来提供，同时是真实的且提供高度可变性来模拟身体关节跟踪系统在被部署时将经历的真实世界状况。

[0006] 在一个实施例中，提供了一种用于生成用于人体跟踪的代表训练数据的处理器实现的方法。该方法包括多个处理器实现的步骤。该方法包括访问至少一个运动捕捉序列，该至少一个运动捕捉序列标识了其中行动者执行移动的时间段期间该行动者的身体的姿态。例如，该序列可在运动捕捉工作室中通过在穿着带有标记的运动捕捉套装的行动者执行一系列规定移动时对该行动者进行成像来获得。该方法还包括基于至少一个运动捕捉序列执行对多个不同身体类型的重新定标以及不相似姿态选择，以便提供多个重新定标的不相似姿态。该方法还包括根据用于相应身体类型的 3-D 身体模型来渲染每一不相似姿态，以提供该不相似姿态的相应深度图像，并且提供标识该不相似姿态的身体部位的相应分类图像。使用多个不同的 3-D 身体模型，对每一身体类型使用一个。此外，相应深度图像和相应分类图像构成了可由机器学习算法用于人体跟踪的像素数据。

[0007] 在一种方法中,重新定标是在不相似姿态选择之前执行的,并且在另一方法中,重新定标是在不相似姿态选择之后执行的。任选地,向深度图像添加噪声来提供更真实的深度图像,该更真实的深度图像与真实世界环境中的深度相机将看见的深度图像相似。噪声可包括可因人有头发而引起的噪声、深度量化噪声、随机噪声、人的身体边缘所引起的噪声、因检测极薄结构而引起的噪声、以及因照相机几何结构引起的噪声。

[0008] 提供本概要以用简化形式介绍在下面的说明书中进一步描述的精选概念。本概述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,也不旨在用于限定所要求保护的主题的范围。

## 附图说明

- [0009] 在附图中,标号相同的元素彼此对应。
- [0010] 图 1 描绘了身体关节跟踪系统的示例实施例。
- [0011] 图 2 描绘了图 1 的身体关节跟踪系统的示例框图。
- [0012] 图 3 描绘了可以在图 1 的身体关节跟踪系统中使用的计算环境的示例框图。
- [0013] 图 4 描绘了可以在图 1 的身体关节跟踪系统中,或在用于重新定标、不相似姿态选择和渲染的计算机图形系统中使用的计算环境的另一示例框图。
- [0014] 图 5 描绘了用于提供用于人体跟踪的代表训练数据的方法。
- [0015] 图 6 提供了获得运动捕捉数据(图 5 的步骤 500)的进一步细节。
- [0016] 图 7a 提供了执行重新定标和不相似姿态检测(图 5 的步骤 502)的进一步细节,其中首先执行重新定标。
- [0017] 图 7b 描绘了用于不相似姿态选择的算法。
- [0018] 图 7c 提供了执行重新定标和不相似姿态检测(图 5 的步骤 502)的进一步细节,其中首先执行不相似姿态检测。
- [0019] 图 8 提供了执行渲染来提供深度图像和分类图像(图 5 的步骤 504)的进一步细节。
- [0020] 图 9 是向深度图像添加噪声(图 5 的步骤 506)的进一步细节。
- [0021] 图 10a 描绘了图 5 的过程的替换视图,其中重新定标在不相似姿态检测之前执行。
- [0022] 图 10b 描绘了图 5 的过程的替换视图,其中重新定标在不相似姿态检测之后执行。
- [0023] 图 11a 描绘了运动捕捉工作室中具有标记的行动者的第一姿态(图 5 的步骤 500)的示例视图。
- [0024] 图 11b 描绘了运动捕捉工作室中具有标记的行动者的第二姿态(图 5 的步骤 500)的示例视图。
- [0025] 图 12 描绘了从图 11a 的行动者的姿态中得到的骨架中的关节位置的示例。
- [0026] 图 13a 描绘了第一身体类型的 3-D 身体的深度图像的渲染,该 3-D 身体具有图 12 的对应骨架的覆盖物以及 3-D 场景元素。
- [0027] 图 13b 描绘了第二身体类型的 3-D 身体 1360 的深度图像的渲染,该 3-D 身体具有对应骨架的覆盖物。
- [0028] 图 14 描绘了对应于图 13a 的 3-D 身体的分类图像。
- [0029] 图 15a 描绘了一示例深度图像。

[0030] 图 15b 描绘了对应于图 15a 的深度图像的示例分类图像。

## 具体实施方式

[0031] 提供了用于生成供身体关节跟踪系统的机器学习算法使用的合成图像的技术。从运动捕捉工作室获得有限数量的运动捕捉序列。该运动捕捉序列包括行动者执行的姿态或移动。这些序列被充分利用来通过将各序列重新定标到多个不同身体类型来提供程度提高的可变性。通过选择不相似姿态，使得冗余姿态或接近冗余的姿态不被提供给机器学习算法，来达到效率。此外，通过添加预期在身体关节跟踪系统的真实世界部署中看见的各种类型的噪声来达到较大的真实性。也可引入其他随机变化。例如，可向重新定标添加一定程度的随机性。提供给该学习算法的数据包括深度和分类图像的已配准对形式的已标记训练数据，以及姿态数据。

[0032] 此处提供的技术避免向训练算法提供压倒性量的数据，同时仍覆盖了大范围的姿态和身体类型，包括例如较高和较低身体的独立移动。可提供能处理大范围姿态和身体类型的单个系统。

[0033] 各特征包括基于姿态之间的距离的样本选择、通过组合部分骨架来生成新样本、通过插入 3-D 模型来生成合成背景、以及通过扰动深度图来生成合成的含噪声图像。

[0034] 图 1 描绘了其中个人 8 与应用交互的身体关节跟踪系统 10 的示例实施例。这示出了身体关节跟踪系统诸如在用户家中的真实世界部署。身体关节跟踪系统 10 包括显示器 196、深度相机系统 20、以及计算环境或装置 12。深度相机系统 20 可包括图像相机组件 22，其具有红外 (IR) 光发射器 24、红外相机 26 和红 - 绿 - 蓝 (RGB) 相机 28。也称为个人或玩家的用户 8 站在深度相机的视野 6 中。线 2 和 4 表示视野 6 的边界。在该示例中，深度相机系统 20 和计算环境 12 提供了其中显示器 196 上的化身 197 跟踪用户 8 的移动的应用。例如，当用户举起手臂时，化身可举起手臂。化身 197 在 3-D 虚拟世界中站在路 198 上。可定义笛卡尔世界坐标系，其包括沿着深度相机系统 20 的焦距（例如水平）延伸的 z 轴、垂直延伸的 y 轴、以及横向且水平延伸的 x 轴。注意，附图的透视被修改成简化表示，显示器 196 在 y 轴方向上垂直延伸，z 轴垂直于 y 轴和 x 轴且与用户 8 所站立的地平面平行地从深度相机系统延伸出来。

[0035] 一般而言，身体关节跟踪系统 10 用于识别、分析和 / 或跟踪人类目标。计算环境 12 可包括计算机、游戏系统或控制台等，以及执行应用的硬件组件和 / 或软件组件。

[0036] 深度相机系统 20 可以包括相机，相机用于在视觉上监视诸如用户 8 等的一个或多个人，从而可以捕捉、分析并跟踪用户所执行的姿态和 / 或移动，来执行应用中的一个或多个控制或动作，如使化身或屏幕上人物活动起来或选择用户界面 (UI) 中的一菜单项。

[0037] 身体关节跟踪系统 10 可以连接到诸如显示器 196 等可向用户提供视觉和音频输出的视听设备，如电视机、监视器、高清电视机 (HDTV) 等，甚至是墙或其他表面上的投影。音频输出也可经由单独的设备来提供。为驱动显示器，计算环境 12 可包括提供与应用相关联的视听信号的诸如图形卡等视频适配器，和 / 或诸如声卡等音频适配器。显示器 196 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。

[0038] 用户 8 可使用深度相机系统 20 来跟踪，使得用户的姿势和 / 或移动被捕获并用于使化身或屏幕上人物活动起来，和 / 或被解释为对计算机环境 12 所执行的应用的输入控

制。

[0039] 用户 8 的某些移动可被解释为可对应于除控制化身之外的动作的控制。例如,在一实施例中,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。玩家可使用移动来从主用户界面选择游戏或其他应用,或以其他方式导航选项菜单。由此,用户 8 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。

[0040] 个人可在与应用交互时抓握诸如道具等物体。在此类实施例中,个人和物体的移动可用于控制应用。例如,可以跟踪并利用玩家手持球拍的运动来控制模拟网球游戏的应用中的屏幕上球拍。在另一示例实施例中,可以跟踪并利用玩家手持诸如塑料剑等玩具武器的运动来控制提供海盗船的应用的虚拟世界中对应的武器。

[0041] 身体关节跟踪系统 10 还可用于将目标移动解释为游戏以及出于娱乐和休闲目的的其他应用范围之外的操作系统和 / 或应用控制。例如,操作系统和 / 或应用的基本上任何可控方面可由用户 8 的移动来控制。

[0042] 图 2 描绘了图 1a 的身体关节跟踪系统 10 的示例框图。深度相机系统 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,捕捉带有包括深度图像的深度信息的视频,该深度图像可包括深度值。深度相机系统 20 可将深度信息组织为“Z 层”,即可与从深度相机沿其视线延伸的 Z 轴垂直的层。

[0043] 深度相机系统 20 可包括图像相机组件 22,如捕捉物理空间中的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,其中该 2-D 像素区域中的每一像素具有代表距离图像相机组件 22 的线性距离的相关联的深度值。

[0044] 图像相机组件 22 可包括可用于捕捉场景的深度图像的红外 (IR) 光发射器 24、红外相机 26、以及红 - 绿 - 蓝 (RGB) 相机 28。3-D 相机由红外发射器 24 和红外相机 26 的组合来形成。例如,在飞行时间分析中,IR 光发射器 24 将红外光发射到物理表面上,并且红外相机 26 检测来自物理空间中的一个或多个目标和物体的表面的反向散射光。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与相应的入射光脉冲之间的时间并将其用于确定从深度相机系统 20 到物理空间中的目标或对象上的特定位置的物理距离。可将传出光波的相位与传入光波的相位进行比较来确定相移。然后可以使用相移来确定从深度相机系统到对象或目标上的特定位置的物理距离。

[0045] 飞行时间分析也可用于通过经由包括例如快门式光脉冲成像等各种技术来分析反射光束随时间的强度,来间接地确定从深度相机系统 20 到目标或对象上的特定位置的物理距离。

[0046] 在另一示例实施例中,深度相机系统 20 可使用结构化光来捕捉深度信息。在这一分析中,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 光发射器 24 被投影到场景上。在撞击到场景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红外相机 26 和 / 或 RGB 相机 28 来捕捉,然后可被分析以确定从深度相机系统到目标或对象上的特定位置的物理距离。

[0047] 深度相机系统 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。

[0048] 深度相机系统 20 还可包括话筒 30,话筒 30 包括例如接收声波并将其转换成电信号的换能器或传感器。另外,话筒 30 可用于接收由个人提供的诸如声音等音频信号,来控

制由计算环境 12 运行的应用。音频信号可包括诸如说出的单词、口哨、叫声和其他话语等个人的口声，以及诸如拍手或跺脚等非口声。

[0049] 深度相机系统 20 可包括与图像相机组件 22 进行通信的处理器 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三维像素网格的指令；用于移除包括在三维像素网格中的背景以便隔离与人类目标相关联的一个或多个三维像素的指令；用于确定隔离的人类目标的一个或多个骨端的位置或定位的指令；用于基于一个或多个骨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，这些将在下文中更详细描述。

[0050] 深度相机系统 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、以及存储 3-D 相机或 RGB 相机所捕捉的图像或图像帧、或任何其他合适的信息、图像等等。根据一示例实施例，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、或任何其他合适的有形计算机可读存储组件。存储器组件 34 可以是经由总线 21 与图像捕捉组件 22 和处理器 32 进行通信的单独组件。根据另一实施方式，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。

[0051] 深度相机系统 20 可以经由通信链路 36 与计算环境 12 进行通信。通信链路 36 可以是有线和 / 或无线连接。根据一实施例，计算环境 12 可经由通信链路 36 向深度相机系统 20 提供时钟信号，该信号指示何时从位于深度相机系统 20 的视野中的物理空间捕捉图像数据。

[0052] 另外，深度相机系统 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 / 或可由深度相机系统 20 生成的骨架模型。计算环境 12 然后可使用该模型、深度信息和捕捉的图像来控制应用。例如，如图 2 所示，计算环境 12 可包括诸如姿势过滤器集合等姿势库 190，每一姿势过滤器具有关于可由骨架模型（在用户移动时）执行的姿势的信息。例如，可为各种手姿势提供姿势过滤器，如手的挥击或投掷。通过将检测到的运动与每一过滤器进行比较，可标识个人执行的指定姿势或移动。也可确定执行移动的范围。

[0053] 可将由深度相机系统 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特定移动。那些移动可与应用的各种控制命令相关联。

[0054] 计算环境还可包括用于执行存储在存储器 194 中的指令以向显示设备 196 提供音频 - 视频输出信号并实现如此处所描述的其他功能的处理器 192。

[0055] 图 3 描绘了可以在图 1 的身体关节跟踪系统中使用的计算环境的示例框图。计算环境可用于解释一个或多个姿势或其他移动并作为响应来更新显示器上的视觉空间。上面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。多媒体控制台 100 包括具有一级高速缓存 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 临时存储数据并因此减少存储器访问周期数，由此改进处理速度和吞吐量。CPU 101 可以设置成具有一个以上的核，以及由此的附加的一级和二级高速缓存 102 和 104。诸如闪存 ROM 等存储器 106 可存储在多媒体控制台 100 通电时在引导过程的初始阶段期间加载的可执行代码。

[0056] 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、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。视频处理流水线将数据输出到 A/V( 音频 / 视频 ) 端口 140 以传输到电视机或其他显示器。存储器控制器 110 连接到 GPU 108 以便于处理器访问各种类型的存储器 112, 诸如 RAM( 随机存取存储器 )。

[0057] 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 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 (NW IF) 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、电缆调制解调器等的各种不同的有线或无线适配器组件中任何一种。

[0058] 提供系统存储器 143 来存储在引导进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、或其它可移动媒体驱动器。媒体驱动器 144 对于多媒体控制台 100 可以内置或外置的。应用数据可经由媒体驱动器 144 访问, 以由多媒体控制台 100 执行、回放等。媒体驱动器 144 经由诸如串行 ATA 总线或其它高速连接等总线连接到 I/O 控制器 120。

[0059] 系统管理控制器 122 提供涉及确保多媒体控制台 100 的可用性的各种服务功能。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对应的音频处理流水线。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。

[0060] 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。

[0061] CPU 101、GPU 108、存储器控制器 110、和多媒体控制台 100 内的各个其它组件经由一条或多条总线互连, 包括串行和并行总线、存储器总线、外围总线、和使用各种总线架构中任一种的处理器或局部总线。

[0062] 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 中并在 CPU 101 上执行。应用可呈现在导航到多媒体控制台 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, 媒体驱动器 144 中包含的应用和 / 或其它媒体可从媒体驱动器 144 启动或播放, 以向多媒体控制台 100 提供附加功能。

[0063] 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它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, 随着通过网络接口 124 或无线适配器 148 可用的宽带连接的集成, 多媒体控制台 100 还可作为较大网络社区中的参与者来操作。

[0064] 当多媒体控制台 100 通电时, 保留指定量的硬件资源以供多媒体控制台操作系统作系统使用。这些资源可以包括存储器保留 ( 例如, 16MB ) 、CPU 和 GPU 周期 ( 例如, 5% ) 、网络带宽 ( 例如, 8kbs ) 等。因为这些资源是在系统引导时保留的, 所以所保留的资源对应用而言是不存在的。

[0065] 具体地,存储器保留较佳地足够大,以包含启动内核、并发系统应用和驱动程序。CPU 保留较佳地为恒定,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。

[0066] 对于 GPU 保留,通过使用 GPU 中断来显示由系统应用生成的轻量消息(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,并且覆盖图较佳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,也就不会引起 TV 重新同步。

[0067] 在多媒体控制台 100 引导且系统资源被保留之后,就执行并发系统应用来提供系统功能。系统功能被封装在一组在上述所保留的系统资源中执行的系统应用中。操作系统内核标识是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。进行调度是为了把由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。

[0068] 当并发系统应用需要音频时,则由于时间敏感性而异步调度音频处理给游戏应用。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)。

[0069] 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。输入设备不是所保留的资源,但却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,并且驱动程序维持有关焦点切换的状态信息。控制台 100 可从包括相机 26 和 28 的图 2 的深度相机系统 20 接收附加输入。

[0070] 图 4 描绘了可以在图 1 的身体关节跟踪系统中,或在用于重新定标、不相似姿态选择和渲染的计算机图形系统中使用的计算环境的另一示例框图。

[0071] 在身体关节跟踪系统中,该计算环境可用于解释一个或多个姿势或其他移动并作为响应更新显示器上的视觉空间。计算环境 220 包括计算机 241,计算机 241 通常包括各种有形计算机可读存储介质。这可以是能由计算机 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。基本输入 / 输出系统 224(BIOS) 包括如在启动时帮助在计算机 241 内的元件之间传输信息的基本例程,它通常储存在 ROM 223 中。RAM 260 通常包含处理单元 259 可以立即访问和 / 或目前正在操作的数据和 / 或程序模块。图形接口 231 与 GPU 229 进行通信。作为示例而非局限,图 4 描绘了操作系统 225、应用程序 226、其它程序模块 227 和程序数据 228。

[0072] 计算机 241 还可包括其他可移动 / 不可移动、易失性 / 非易失性计算机存储介质,如从不可移动、非易失性磁介质读取或对其写入的硬盘驱动器 238,从可移动、非易失性磁盘 254 读取或对其写入的磁盘驱动器 239,以及从诸如 CDROM 或其他光介质等可移动、非易失性光盘 253 读取或对其写入的光盘驱动器 240。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 / 非易失性有形计算机可读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 238 通常由诸如接口 234 等不可移动存储器接口连接至系统总线 221,磁盘驱动器 239 和光盘驱动器 240 通常由

诸如接口 235 等可移动存储器接口连接至系统总线 221。

[0073] 以上讨论并在图 4 中描绘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。例如，硬盘驱动器 238 被描绘为存储操作系统 258、应用程序 257、其它程序模块 256 和程序数据 255。注意，这些组件可以与操作系统 225、应用程序 226、其他程序模块 227 和程序数据 228 相同，也可以与它们不同。操作系统 258、应用程序 257、其它程序模块 256 和程序数据 255 在这里被标注了不同的标号是为了说明至少它们是不同的副本。用户可以通过输入设备，诸如键盘 251 和定点设备 252（通常被称为鼠标、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 236 连接至处理单元 259，但也可以由其他接口和总线结构，诸如并行端口、游戏端口或通用串行总线（USB），来连接。包括相机 26 和 28 的图 2 的深度相机系统 20 可为控制台 100 定义附加输入设备。监视器 242 或其他类型的显示器也经由接口，诸如视频接口 232 连接至系统总线 221。除监视器以外，计算机也可以包括其它外围输出设备，诸如扬声器 244 和打印机 243，它们可以通过输出外围接口 233 连接。

[0074] 计算机 241 可使用至一个或多个远程计算机，诸如远程计算机 246 的逻辑连接在网络化环境中操作。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的网络节点，且通常包括许多或所有以上相对于计算机 241 描述的元件，但是在图 4 中仅示出了存储器存储设备 247。逻辑连接包括局域网（LAN）245 和广域网（WAN）249，但也可以包括其它网络。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。

[0075] 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 连接至 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 建立通信的其他装置。调制解调器 250 可以是内置或外置的，它可以经由用户输入接口 236 或其他适当的机制连接至系统总线 221。在网络化环境中，相对于计算机 241 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 示出了远程应用程序 248 驻留在存储器设备 247 上。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。

[0076] 该计算环境可包括其上包含有计算机可读软件的有形计算机可读存储，该计算机可读软件用于对至少一个处理器进行编程来执行此处描述的用于生成用于人体跟踪的代表训练数据的方法。该有形计算机可读存储可包括例如，组件 222、234、235、230、253、254 中的一个或多个。此外，该计算环境的一个或多个处理器可以提供用于生成用于人体跟踪的代表训练数据的处理器实现的方法，包括此处所描述的处理器实现的步骤。处理器可包括例如组件 229 和 259 中的一个或多个。

[0077] 图 5 描绘了用于提供用于人体跟踪的代表训练数据的方法。步骤 500 包括获得运动捕捉数据。例如，这可从运动捕捉工作室中的行动者获得。在运动捕捉工作室中，对一个或多个行动者的移动进行每秒多次，例如每秒 30 次的采样。每一样本对应于一数据帧。通常，行动者穿着具有将光反射到相机的标记的特殊套装。使用多个高分辨率相机来从工作室中的不同位置对行动者成像，可通过三角测量来检测每一标记的确切位置。参见图 11a

和 11b 来获得更多细节。步骤 500 的输出是一组运动捕捉序列。每一这样的序列描述了执行移动的行动者的身体上的一组 3-D 点的运动。所表示的各具体点对应于该行动者身体上的标记的定位，并且可以在序列之间有所不同。运动捕捉序列包括一组运动捕捉帧，其中每一帧对应于某一时刻的点的配置。步骤 500 在以下结合图 6 来进一步讨论。

[0078] 步骤 502 包括执行重新定标和不相似姿态检测。在一种方法中，重新定标是在不相似姿态选择之前执行的，而在另一种方法中，重新定标是在不相似姿态选择之后执行的。重新定标将从运动捕捉工作室中的行动者获得的标记位置转换成不同身体类型的骨架模型。给定身体类型的骨架模型可以通过基于标记的位置确定骨架模型中的关节的位置来获得。例如，当一个或多个标记被定位在行动者上相对于肩的已知位置时，可从该标记位置确定表示肩的关节的位置。

[0079] 骨架模型或骨架是通过肢或骨连接的身体的 3-D 关节或其它点的虚拟构造，使得骨架的构造可以通过列出 3-D 点的位置，或者另选地通过枚举将各条骨相关到骨架中的另一条骨的关节角度来表示。该相对定位可将每一条骨相关到骨架的树结构分解中的其父骨。另外，可以用关节角度来指定形状参数，例如指定骨长度。

[0080] 在对身体关节跟踪系统的真实世界使用中，所成像的用户将具有许多不同的身体类型，包括基于高度、宽度、体重、姿态、年龄、性别、发型和发量、服饰等的变化。由此，仅使用行动者的身体类型或某一其它标准身体类型来提供用于运动捕捉的学习算法的训练数据将不能提供足够的可变性。重新定标到不同的身体类型提供了增加的可变性而不需要从运动捕捉工作室中的许多不同身体类型的不同行动者获得运动捕捉数据，从而节省了成本和时间。

[0081] 运动捕捉数据的重新定标可涉及将来自运动捕捉序列的 3-D 数据表示为身体类型的预定义骨架的参数，特别是在转换运动捕捉序列的每一帧的 3-D 标记位置的意义上，将该 3-D 数据表示为一系列关节角度（每一帧有一组关节角度）和形状参数（每一序列有一组形状参数）。从同一行动者捕捉的序列一般将共享形状参数。运动捕捉序列的每一帧的 3-D 标记位置可作为诸如 (x, y, z) 坐标等一系列坐标来提供。类似地，骨架的关节位置可作为另一系列 (x, y, z) 坐标来提供。可执行到不同骨架和身体类型的重新定标。作为一示例，可使用 10-15 个不同的身体类型。重新定标还可引入身体类型的进一步变化，如略微变化的骨或肢长度，以便增加可变性程度。一般而言，目标是在基于一定范围的真实人类可变性的边界内提供身体姿态之间的最高量的可变性。

[0082] 不相似姿态选择分析从每一重新定标的运动捕捉序列获得的所有姿态的集合。采用每秒 30 帧的帧速率、长度为例如 1-2 分钟的运动捕捉序列、以及对每一帧到 15 个不同身体类型的重新定标，可以看到帧 / 姿态的数量可变得巨大。为了提高效率并避免向学习算法提供具有高度冗余的过量数据，可使用每一帧来运行不相似姿态选择过程，以获得指定的、减少数量的不相似姿态。该不相似姿态选择过程标识根据一距离度量相隔指定距离的帧。步骤 502 在以下结合图 7a-7c 来进一步讨论。不相似姿态选择提供了具有最大程度不同的骨架的运动捕捉样本子集。这些样本中的每一个用于渲染 3-D 身体模型，从而创建深度图以及分类图或 Harlequin（哈利昆）图。多个任选阶段可被插入到该过程中。

[0083] 步骤 504 包括执行渲染来提供深度图和分类图像。渲染指的是在像素空间中生成合成图像。深度图像可以从相对于深度图像所表示的身体位于指定位置的虚拟相机的角度

来渲染。在渲染深度图像时也可指定诸如虚拟相机的视野等其它因素。本质上，深度图像通过模拟视野中的 3-D 身体并且任选地模拟诸如地板、墙、天花板、家具和其它家用物品等场景元素，来模拟真实的深度相机将在真实环境中看见的东西。深度图像可具有与真实深度相机相似的像素分辨率。此外，在深度图像中，每一像素可以标识从虚拟相机到 3-D 身体的距离、从虚拟相机到 3-D 场景物体的距离、或背景空间，背景空间是既不表示 3-D 身体也不表示 3-D 场景物体的像素。

[0084] 分类图像或分类图标识并标记 3-D 身体的不同身体部位或不同 3-D 场景元素。例如，每一像素可以标识最接近虚拟相机的身体部位的编号，或 3-D 场景物体的唯一索引。在用于身体关节跟踪系统的学习算法中，深度图像是使用诸如过滤器设置等设置来处理的，并且生成对应的分类图，在该分类图中学习算法试图标识身体部位和场景元素。由学习算法生成的分类图可与随深度图像一起提供的分类图进行比较来确定该学习算法有多准确。随深度图像一起提供的分类图本质上是提供了正确答案的基准图，并且该学习算法可以在处理深度图像时重复地调整其设置，例如训练其自身，直到它能够尽可能准确地复制正确的结果。此外，深度图像的处理以及与相关联的分类图的比较可对该学习算法作为输入数据集接收的众多不相似帧重复。

[0085] 一旦优化了学习算法，就记录对应的设置并可随身体关节跟踪系统一起发送该学习算法以供最终用户使用。

[0086] 渲染也可随每一深度图像和分类图一起提供描述姿态的文本文件，如按照用于提供 3-D 身体姿态的骨架模型的关节坐标来描述。也可提供诸如在运动捕捉工作室中使用的设置等其它数据。步骤 504 在以下结合图 8 来进一步讨论。

[0087] 步骤 506 包括向部分或全部深度图像添加噪声。在一种方法中，噪声的量和类型以及要添加噪声的深度帧的选择可被随机化。步骤 506 在以下结合图 9 来进一步讨论。步骤 508 包括将深度图像和分类图像提供给机器学习算法。

[0088] 图 6 提供了获得运动捕捉数据（图 5 的步骤 500）的进一步细节。步骤 600 包括在运动捕捉工作室中，当行动者随之间执行移动时捕捉帧序列。在一种方法中，执行大量种类的移动，这些移动预期适用于描述预期在用户参与身体关节跟踪系统的不同应用时会遇到的用户移动。在另一种方法中，移动是特定于诸如游戏等具体应用的。例如，交互式网球游戏可具有诸如挥动球拍等规定移动。任选地，行动者可在移动期间在他或她的手中持有道具。

[0089] 每一序列由连续的帧组成，且每一帧标识了行动者的身体上的标记的位置。每一序列可以基于一脚本来执行。例如，一个脚本可指定特定的手臂和腿移动。可获得一个或多个序列。

[0090] 在步骤 602，对于序列中的每一帧，提供标记的 3-D 坐标的数据集。如上所述，每一标记的确切位置可以使用运动捕捉工作室中的不同相机通过三角测量来确定。步骤 604 输出运动捕捉序列的一个或多个数据集。

[0091] 在一个示例实现中，捕捉 N 个运动捕捉序列，表示为  $\{S_i | i = 1 \dots N\}$ 。每一序列包括帧  $S_i = \{F_{it} | t = 1 \dots N_i\}$ 。由此， $S_i$  表示一系列或一组帧，且其所包含的对象是  $F_{it}$ 。每一  $F_{it}$  是 3-D 点位置的向量或关节角度的向量，并表示在时间 t 序列  $S_i$  的帧 (F)。每一帧由一组 M 个标记位置来表示，因此  $F_{it}$  由  $M \times 3$  的矩阵来表示，每一行编码了一个标记的 3-D 位

置。注意,  $N$  和每一个  $N_i$  表示不同的变量。 $N_i$  是序列  $S_i$  中的帧数。

[0092] 图 7a 提供了执行重新定标和不相似姿态检测(图 5 的步骤 502)的进一步细节, 其中首先执行重新定标。在一种方法中, 将每一运动捕捉帧重新定标到多个不同的身体类型。该重新定标在步骤 700 处开始。步骤 702 从来自图 6 的步骤 606 的运动捕捉序列的数据集中选择一序列。步骤 704 从当前序列中选择一个帧。步骤 706 选择一身体类型。

[0093] 步骤 708 基于所选身体类型将 3-D 标记位置转换成关节位置。作为指定类型的身体的骨架模型的一部分的关节位置可基于标记的位置来获得。例如, 当一个或多个标记被定位在行动者上相对于肩的已知位置时, 可从该标记位置确定表示肩的关节的位置。此外, 指定类型的身体中的关节的位置可以基于该身体的模型以及适合该身体模型的对应的骨架来确定。也可在步骤 708 期间添加诸如基于骨或肢长度等随机变化。

[0094] 在一示例实现中, 输入骨架序列被重新定标到一个或多个标号为 1...B 的身体形状, 从而产生经重新定标的帧  $\{F'_{itk} | k = 1 \dots B\}$ 。 $F'_{itk}$  是对于身体类型  $k$  在时间  $t$  运动捕捉序列  $S_i$  中的一个帧。可以选择身体形状的范围来覆盖该系统的较大比例的潜在用户, 并且可包括以下各项的变化: 性别(男性、女性)、年龄(成人、孩子)、身体类型(诸如 100 磅、150 磅和 200 磅等指定体重; 或胖、瘦或普通体形)、身高(例如, 5 英尺、5.5 英尺、6 英尺)、头部毛发类型(小胡子、胡须、长发/短发)、服饰(宽松、绷紧、短裙)、等等。例如, 身体类型 1 可以是男性、成人、体重 150 磅、5.5 英尺高、具有短发和宽松的服饰, 身体类型 2 可以是男性、成人、体重 200 磅、6 英尺高、具有长发和绷紧的服饰, 等等。

[0095] 该阶段可任选地包括对诸如手指角度等未跟踪关节的调整。用于渲染的身体模型具有比输入骨架多得多的参数(关节角度), 因此对于大多数未跟踪的关节, 不具有关于它们在哪里的信息。为了填入该信息, 可以设置一默认值(例如, 将手指角度设为对应于张开的手)。或者, 每一渲染的图像可以随机地设置这些值, 从而生成渲染中的更多变化。人们知道这些手指在哪里是因为人们知道手在哪里(因为手是跟踪的关节), 且给定手指定向, 人们可以使用人类骨架的运动学模型以及手指如何与手相关。

[0096] 在一个具体实例中, 使用了 15 个基础骨架, 其体重和/或骨或肢长度随机变化。由此,  $B$  实际上非常大, 但可考虑所有可能的  $F'_{itk}$  的随机子集。参见图 12 可获得身体模型的骨架的进一步细节。步骤 710 将重新定标的帧添加到重新定标的帧的数据集。

[0097] 在第一关节位置标识方案中, 骨架模型中的关节位置可以由(x, y, z)坐标来标识。一示例骨架可具有大约 40 个关节。该数据集可包括对应于每一帧的矩阵, 其中每一行表示骨架模型中的一个具体关节, 并且有表示坐标系中的关节的(x, y, z)位置的三列: 对应于 x 的第一列、对应于 y 的第二列、以及对应于 z 的第三列。例如, 可标识关节, 使得左肩是关节 #1, 右肩是关节 #2, 等等。

[0098] 在第二关节位置标识方案中, 骨架模型可以通过指定起始点以及一系列关节角度和诸如骨或肢长度等形状参数来定义。例如, 关节可以与一给定点相距沿着一向量的指定距离。在这一情况下, 该数据集可包括对应于每一帧的矩阵, 其中第一行表示起始关节, 且每一另外的行表示骨架模型中的相邻关节。在这一情况下, 可以有四列。前三个列表示例如分别在每一 x-y、y-z、x-z 平面中从先前关节到当前关节的向量的角度。第四个矩阵列可提供形状参数, 如从先前关节到当前关节的距离, 例如骨长度。也可使用其它关节标识方案。

[0099] 可做出关节位置标识方案之间的转换。例如,从第一到第二关节位置标识方案的转换可以涉及将 3-D 空间中的两个连续关节位置相减来获得它们之间的向量,该向量按照每一 x-y、y-z、x-z 平面中的角度以及作为形状参数的该向量的大小。从第二到第一关节位置标识方案的转换可涉及添加由两个连续关节位置之间的角度和大小定义的向量。

[0100] 在判定步骤 712,如果存在下一身体类型,则对当前帧重复步骤 706-710 的过程。如果当前帧已被重新定标到所有身体类型,则判定步骤 712 为假,并且判定步骤 714 确定当前序列中是否还有另一帧要处理。如果判定步骤 714 为真,则在步骤 704 选择一个新的帧,并且对该新的帧重复步骤 706-710 的过程。如果当前序列中的所有帧都已被重新定标,则判定步骤 714 为假,并且判定步骤 716 确定是否有另一序列要处理。如果有另一序列要处理,则在步骤 702 选择新序列并且执行步骤 704-710 的过程。当已经处理了最后一个序列时,判定步骤 716 为假,并且重新定标在步骤 718 结束。步骤 720 输出重新定标的帧的数据集。

[0101] 不相似姿态选择在步骤 722 开始。在所有可能姿态的空间中,不相似姿态选择提供了稀疏采样。结果,向学习算法提供了较少的帧,从而减少了计算花费,但不损失大量质量。例如,帧数可以按照大小的数量级来减少。

[0102] 在所描绘的方法中,不相似姿态检测使用用于重新定标的骨架模型中的关节位置,而非来自运动捕捉序列的标记位置来执行。步骤 724 从步骤 720 提供的重新定标的帧的数据集中选择并移除姿态。步骤 726 将该姿态添加到所选不相似姿态的新数据集。步骤 728 确定每一所选不相似姿态和重新定标的帧的数据集中的所有剩余姿态之间的距离,这些剩余姿态是候选姿态。步骤 730 任选地排除距离任何所选不相似姿态不足至少阈值距离 T 的候选姿态。这些候选姿态被认为与一个或多个所选不相似姿态太相似。步骤 732 确定哪一候选姿态具有与任何所选不相似姿态的最大的最小距离。由此,对于每一候选姿态,可以确定其距离每一所选不相似姿态的距离。然后可对每一候选姿态取这些距离的最小值。然后,确定在所有候选姿态中哪一个最小值是最大的。在步骤 734,将所选候选姿态添加到所选不相似姿态的数据集,并从重新定标的帧的数据集中移除。如果在判定步骤 736 在重新定标的帧的数据集中有下一帧要处理,则处理在步骤 728-734 继续。在一种方法中,当在重新定标的、不相似的帧的数据集中已经提供了指定数量 D 的帧时,判定步骤 736 为假。不相似姿态检测在步骤 738 结束。步骤 740 输出所选不相似帧的数据集。不相似姿态是重新定标的运动步骤序列中的姿态的子集。

[0103] 图 7b 描绘了用于不相似姿态选择的算法。将每一元组  $(i, t, k)$  与一整数  $r$  相关联,所有重新定标的运动捕捉帧被表示为集合  $S' = \{F'_{r_i} | r = 1 \dots R\}$ ,其中  $R = B \times \prod_i N_i = B \times N_1 \times \dots \times N_N$ 。不相似姿态选择的目标是选择这些帧的(大小为 D 的)子集,由整数集  $P_I = \{r_1, \dots, r_D\}$  来表示,使得该子集不包括由相似度函数  $\Phi$  定义的相似帧对,该相似度函数也可被认为是距离度量。该函数(其示例在下文中给出)将来自  $S'$  的一对帧映射到一正实数,该正实数对于相似帧较小,在其它情况下则较大。为清楚起见,如果帧由  $M \times 3$  的矩阵 G 和 H 来表示,则  $\Phi(G, H)$  返回 G 和 H 的相似度,其中对于相同姿态  $\Phi = 0$ ,且越大的  $\Phi$  表示越不相似的姿态。用于该算法的伪代码在图 7b 中示出。输出索引  $P_I$  定义了作为对后续阶段的输入的一组运动捕捉帧 P。如果未执行不相似姿态检测,则  $P' := S'$ 。或者,不相似姿态检测可以在重新定标之前在运动捕捉帧上运行(见图 7c)。

[0104] 不相似姿态检测算法使用姿态相似度函数  $\Phi(G, H)$ 。给定矩阵 G,其 M 行由  $\{g_j\}$

$j = 1:M$  表示,且类似地  $H$  的行由  $\{h_j\}_{j=1:M}$  表示。第一个可能的相似性定义是最大关节距离  $\phi(G, H) := \max_j \left\| \mathbf{g}_j - \mathbf{h}_j \right\|_2^2$ 。

[0105] 描绘了平方欧几里得距离。所使用的第二个可能的定义是平方距离和:  $\phi(G, H) := \sum_{j=1}^M \left\| \mathbf{g}_j - \mathbf{h}_j \right\|_2^2$  可容易地定义一范围的替换距离函数。例如,在不相似姿态选择中,身体的不同部位可被不同地加权。例如,权重可以指示具有不相似的手比具有不相似的脚更重要,在这一情况下,涉及手的关节距离被加权更多。

[0106] 距离度量和 / 或阈值可被调整来达到所需数量的不相似姿态和对应的所需采样稀疏性。

[0107] 图 7c 提供了执行重新定标和不相似姿态检测(图 5 的步骤 502)的进一步细节,其中首先执行不相似姿态检测。在一种可能的方法中,不相似姿态检测使用来自运动捕捉序列的标记位置,而非重新定标的骨架模型中的关节位置来执行。此外,由于重新定标到多个身体类型尚未执行,因此在比较过程中只要使用较少的帧,因此其在计算上的花费较少。首先进行重新定标的优点是在不同人物中可能某些不相似姿态看上去非常相似,因此这可能是浪费的。

[0108] 不相似姿态选择在步骤 750 开始。步骤 752 从在图 6 的步骤 606 提供的运动捕捉序列的数据集中选择并移除一姿态。步骤 754 将该姿态添加到所选不相似姿态的新数据集。步骤 756 确定每一所选不相似姿态和运动捕捉序列的数据集中的所有剩余姿态之间的距离,这些剩余姿态是候选姿态。任选的步骤 758 排除距离所选不相似姿态不足至少阈值距离  $T$  的候选姿态。步骤 760 确定哪一候选姿态具有与所选不相似姿态的最大的最小距离。步骤 762 将该候选姿态添加到所选不相似姿态的数据集,并从运动捕捉序列的数据集中移除候选姿态。如果在判定步骤 764 在运动捕捉序列的数据集中有下一帧要处理,则处理在步骤 756-762 继续。在一种方法中,当在不相似的帧的数据集中已经提供了指定数量的帧时,判定步骤 764 为假。不相似姿态检测在步骤 766 结束。步骤 768 输出所选不相似姿态的数据集,这些姿态尚未被重新定标。不相似姿态是重新定标的运动步骤序列中的姿态的(少于全部的)子集。

[0109] 重新定标在步骤 770 处开始。步骤 772 从来自步骤 720 的所选不相似姿态的数据集中选择一帧。步骤 704 从当前序列中选择一个帧。步骤 774 选择一身体类型。步骤 776 如上所讨论地基于所选身体类型将 3-D 标记位置转换成关节位置。也可在步骤 776 期间添加诸如肢长度等随机变化。步骤 778 将重新定标的帧添加到重新定标的、不相似帧的数据集。

[0110] 在判定步骤 780,如果存在下一身体类型,则对当前帧重复步骤 774-778 的过程。如果当前帧已被重新定标到所有身体类型,则判定步骤 780 为假,并且判定步骤 782 确定不相似帧的数据集中是否还有另一帧要处理。如果判定步骤 782 为真,则在步骤 772 选择一个新的帧,并且对该新的帧重复步骤 774-778 的过程。如果已经重新定标了不相似帧的数据集中的所有帧,则判定步骤 782 为假,并且重新定标在步骤 784 结束。步骤 786 输出重新定标的、不相似帧的数据集。

[0111] 在图 7a 和 7c 中,有可能但不一定对所有帧的身体类型的所有组合进行重新定

标。相反,例如,对于每一帧,用于重新定标的身体类型可以是随机选择的。并且,并非所有帧都需要被重新定标。相反,例如,帧可被随机选择来进行重新定标。也可以对帧和身体类型两者做出随机选择。此外,如此处所描述的用于标识不相似姿态的技术可被修改来标识不同的相似姿态组,并提供来自每一组的代表性姿态同时丢弃该组中其它太相似的姿态。

[0112] 图 8 提供了执行渲染来提供深度图像和分类图像(图 5 的步骤 504)的进一步细节。一旦提供了重新定向的不相似帧,每一帧就被渲染为深度图像、分类图像和文本文件。即,给定一组 P 个运动捕捉帧或姿态,下一过程是渲染这些姿态。该步骤包括,对给定姿态,选择 3-D 场景几何结构的集合,将对应于该姿态的 3-D 身体模型放置在 3-D 场景中,选择该场景中的一个或多个相机位置,然后为姿态、相机位置和身体模型位置的许多组合中的每一种生成一对包括渲染的深度图像的图像(见图 15a),其中每一像素编码了从虚拟相机到 3-D 场景或身体模型的距离,或被编码为背景。该对图像还包括渲染的分类图像(见图 15b),其中每一像素包含编码了该像素处最接近相机的身体部位的编号,或 3-D 身体模型的唯一索引。身体部位使用应用于该对象的纹理图来定义,而诸如地平面等场景几何对象在该分类图像中被给予唯一索引。身体模型的某些项可以在分开的层中渲染,例如短裙或头发。代替用其自己的唯一颜色对它们进行着色,在分类图像中定义了一组新的索引来表示“头发加上头顶”、“头发加上头的左边”、“头发加上肩”等。可对被渲染到单独的层的对象容易地添加噪声。

[0113] 步骤 800 包括从重新定标的不相似帧的数据集中选择一帧。步骤 802 提供对应于所选帧,或者更具体地,对应于骨架模型及其在所选帧中的姿态的 3-D 身体模型。步骤 804 向 3-D 身体模型应用纹理图,其中该纹理图包括对应于不同身体部位的不同区域,并且每一区域被分配一索引或编号。关于进一步细节请参看图 14。

[0114] 步骤 806 选择场景中的虚拟相机位置。对应于每一对渲染的深度图像和分类图像的相机位置可由诸如相机在地面之上的高度、相机的观看方向(全景和倾斜)或相机角度、以及相机视野等参数来参数化,或者甚至可以瞄准身体位置来确保该人在视野中。

[0115] 步骤 808 选择一身体位置。一般而言,该身体模型位置可由虚拟地平面上的(x, z)位置来参数化,并围绕垂直于地平面的 y 轴旋转。z 轴表示距离虚拟相机的深度的轴,y 轴垂直地延伸,且 x 轴表示与地平面平行且垂直于 y 和 z 轴的从左到右方向。注意,并非相机位置、场景几何结构和身体模型位置的所有可能组合都需要被渲染。对于某些训练情形,一个或多个参数中的随机变化将足够。这些参数也被选择来确保人物至少部分地位于相机的视野中。

[0116] 任选地,可在步骤 810 提供也称为场景几何对象的 3-D 场景元素。步骤 812 为每一场景几何对象提供一唯一索引。关于场景几何结构的选择,所渲染的身体模型可被放置在 3-D 场景模型中。可能的场景模型包括:(1) 空场景,其中仅身体模型出现在所渲染的帧中,(2) 地面场景,其中一平面几何元素被放置在模型中,并且出现在渲染的帧中的模型下方,以及(3) 一般场景,其中诸如墙,如长椅、椅子或咖啡桌等家具,植物,宠物,其它家用物品,或可被用户在与身体关节跟踪系统交互时手持或穿戴的对象(如网球拍或棒球)等 3-D 元素可被获得并定位在合成场景中。使用碰撞检测来确保所渲染的身体被定位来避免与场景几何结构的非真实相交。对象可被放置在身体的前方和 / 或后方。

[0117] 作为一示例,运动捕捉序列中的行动者可在一位置中如同坐在长椅上那样摆出姿

态。在这一情况下,作为场景对象的长椅可被插入到具有对应于该姿态的身体模型的深度图像中,使得该身体模型表现为坐在长椅上。由此,可以将长椅或其它场景元素的位置修改成适合该身体姿态。这提供了可用于为此常见情形训练学习算法的真实深度图像。也可使用不同类型(样式/大小)的长椅。注意,某些场景元素可被修改成适合身体姿态,和/或可提供为所有姿态通用。

[0118] 基于这些步骤,渲染包括像素的深度图像。具体地,用从虚拟相机到3-D身体的距离来编码表示3-D身体部位的像素(步骤816)。表示3-D场景对象的像素用从虚拟相机到该3-D场景对象的距离来编码(步骤818)。表示背景区域的像素例如使用与3-D身体和场景对象的像素形成对比的指定颜色来编码为背景像素(步骤820)。可通过用最接近虚拟相机的身体部位的编号编码表示3-D身体部位的像素来渲染对应的分类图像(步骤822)。可为每一身体部位分配一单独的颜色。在某些情况下,身体的一个部位(例如,手臂)可以在另一部位(例如,躯干)的前方,在这一情况下标识较接近的部位。在步骤824,可用对象的唯一索引来编码表示3-D场景对象的像素。作为一示例,输出深度和分类图像的分辨率可以是 $640 \times 480$ 像素。该分辨率应与正被训练的真实世界深度相机的分辨率相似。在步骤814也可提供姿态数据的文本文件。

[0119] 其它选项包括通过组合来自两个或更多原始运动捕捉帧的关节角度或位置,如取来自一个帧的较低身体关节以及来自另一帧的较高身体关节,来扩展姿态集。由此,渲染可组合处于不同姿态、但具有共同的虚拟相机位置和共同的3-D场景元素的两个或更多3-D身体模型的各部分。

[0120] 图9提供了向深度图像添加噪声(图5的步骤506)的进一步细节。在模拟真实深度相机的成像过程时,为使所渲染的深度图像或深度图对训练最有用,可添加噪声来使得深度图像看上去与将从真实世界相机获得的图像相似。为此,可向所渲染的深度图像应用多个过程中的一个或多个来使其成为对将由真实世界相机获得的图像的更准确模拟。

[0121] 在图9中,用于深度图像(900)的噪声建模可基于头部毛发噪声(902)、深度量化噪声(904)、随机噪声(906)、边缘粗糙化(908)、洞切割(910)和投影(912)中的一个或多个。

[0122] 关于毛发噪声模拟(902),某些相机的一个具体特性是身体毛发(例如,头皮毛发、小胡子和胡须—统称为“头部毛发”)可能未被可靠地成像。为了模拟头部毛发的存在,被分类为头部毛发的某一比例的所渲染的像素可被随机地删除,并用深度图中的背景标记和分类图中的背景类来替换。被删除的像素可以是在某一小部分毛发像素的指定半径内的那些像素。例如,参考图14的分类图像,区域1402和1430可表示头皮。深度图像中的多个像素将表示每一区域。在每一区域中,可定义由一部分像素构成的子区域,例如10%、20%、30%等等。该子区域的指定半径,例如1、2、3或4个像素内的像素以及该子区域内的像素。作为另一示例,图14的区域1406和1434可表示其中可删除像素的胡须区域。

[0123] 关于深度量化噪声(904),所渲染的图像包含比真实世界相机的深度分辨率更高的高分辨率深度信息。为了模拟相机的效果,该分辨率可被量化为若干离散等级之一,其精度取决于深度(深度越大,精度越小)。作为一个示例,0到1米之间的深度被量化为5mm的增量;1到2米之间的深度被量化为21mm的增量;依此类推,深度3到3.5米被量化为65mm的增量。

[0124] 关于随机噪声 (906), 为了模拟相机中的随机噪声, 可向所渲染的深度图像中的每一非背景像素添加噪声。非背景像素可以是不表示 3-D 身体 (步骤 816) 或 3-D 场景元素 (步骤 818) 的像素。该噪声可从  $[-N, N]$  上的均匀分布中取得, 其中  $N$  是预设的噪声级。该噪声被添加到逆深度, 使得对于具有深度  $z$  的像素  $(i, j)$ , 添加了噪声的新深度由  $Z_{\text{含噪声}} = \frac{1}{z + N\eta}$  给出, 其中  $\eta$  是从范围  $(-1..1)$  中的均匀分布中的提取。

[0125] 关于边缘粗糙化 (908), 在接近深度图像中的大的不连续处, 深度信息较不可靠。应用称为“边缘粗糙化”的过程来模拟该效应。在一种方法中, 深度图像中其相邻像素离相机远得多 (例如, 10cm) 的点被标记为“边缘”。根据较远像素相对于中心的方向, 此类边缘也被标记为“东”/“西”等。由此, “西”边缘表示当从图像的左边行进到右边时深度的急剧减小。对于每一这样的边缘点, 该点处的值随机地由从较远的表面 (具有较大深度的表面) 中选择的相邻像素的深度来替换。例如, 如果边缘是边缘点 (对应于沿着边缘的每像素一个) 的序列, 则给定边缘将具有其像素被替换的随机子集。因此, 一个边缘可能替换了 2 个像素, 另一边缘可能没有替换像素, 等等。

[0126] 由此, 通过标识不相似姿态之一中的具有大于阈值的深度不连续性的第一和第二边缘, 并且用背景像素替换第一和第二边缘之间的像素来添加噪声。

[0127] 关于洞切割 (910), 可能需要仿真的真实世界深度相机中的传感器的另一特性是“薄”结构 (比如, 图像中比 4 个像素更窄) 可能完全检测不到的性质。这些结构可通过在这一距离内定位补充 (例如, 西, 则为东; 或者北, 则为南) 边缘并用背景替换中间值来检测。

[0128] 关于投影 (912), 可被建模的另一传感器特性是阴影的投影。阴影在场景的较近部分遮挡较远部分时发生, 从而阻止了深度相机报告准确深度。例如, 近的部分可以是另一条手臂前方的一条手臂 (例如, 用前臂形成了“X”)。阴影与其深度不连续性大于阈值斜率的“西”边缘相关联。该投影过程在深度图像中定位此类边缘, 找到满足斜率阈值的最东面的点, 并用背景来替换该边缘和最东面的点之间的像素。该示例适用于其中由于相机的几何结构, IR 发射器在 IR 相机的左侧, 因此阴影仅对垂直边缘形成的情形。可针对其它相机几何结构修改该示例。

[0129] 图 10a 描绘了图 5 的过程的替换视图, 其中重新定标在不相似姿态检测之前执行。此处, 运动捕捉数据 (1000) 经历重新定标 (1002) 来提供身体姿态数据 (1004)。身体姿态数据经历不相似姿态检测 (1006) 来提供唯一或足够不相似的姿态 (1008)。唯一姿态经历渲染 (1010) 来提供深度图 (1012)、分类图 (1014) 和位置信息文件 (1016)。噪声合成 (1018) 被任选地添加到深度图。每一步骤的结果被提供给表示监督学习算法的机器学习算法 (1020)。

[0130] 图 10b 描绘了图 5 的过程的替换视图, 其中重新定标在不相似姿态检测之后执行。此处, 运动捕捉数据 (1050) 经历不相似姿态检测 (1052) 来提供唯一或足够不相似的姿态 (1054)。唯一姿态经历重新定标 (1056) 来提供身体姿态数据 (1058)。身体姿态数据经历渲染 (1060) 来提供深度图 (1062)、分类图 (1064) 和位置信息文件 (1066)。噪声合成 (1068) 被任选地添加到深度图。每一步骤的结果被提供给机器学习算法 (1070)。

[0131] 图 11a 描绘了运动捕捉工作室中具有标记的行动者的第一姿态 (图 5 的步骤 500)

的示例视图。在由渲染 1101 描绘的运动捕捉工作室的简化视图中，多个相机 1140、1142 对正戴着标记的行动者成像。该行动者处于第一姿态，如渲染 1100 所描绘的。标记位于：头顶 (1102)、头的右侧 (1104)——指的是图像空间中的右侧——右肩 (1106)、右肘 (1108)、右腕 (1110 和 1112)、右髋 (1114)、右膝 (1116) 以及右脚 (1118)。其它标记位于：头的左侧 (1134)——指的是图像空间中的左侧——左肩 (1132)、左肘 (1126)、左腕 (1128 和 1130)、左髋 (1124)、左膝 (1122) 和左脚 (1120)。其它标记位于：下部的躯干 (1140)、中间的躯干 (1138) 以及上部的躯干 (1136)。描绘了 x 和 y 方向上的轴。这些可对应于图 1 中的 x 和 y 方向。

[0132] 图 11b 描绘了运动捕捉工作室中具有标记的行动者的第二姿态（图 5 的步骤 500）的示例视图。在运动捕捉工作室 1103 的这一渲染 1103 中，同一行动者处于第二姿态，由渲染 1160 来描绘。

[0133] 图 12 描绘了从图 11a 的行动者的姿态中得到的骨架中的关节位置的示例。骨架 1200 包括以下关节 / 点：头 (1202)、颈 (1204)、右肩 (1206)、右肘 (1208)、右腕 (1210)、右手 (1212)、下部的躯干 (1220)、右大腿 (1218)、右膝 (1216) 和右脚 (1214)。骨架 1200 还包括：左肩 (1234)、左肘 (1232)、左腕 (1230)、左手 (1228)、左大腿 (1226)、左膝 (1224) 和左脚 (1222)。如上所述，每一关节或点可沿着其位置来标识。位置可由绝对 (x, y, z) 位置来标识，或者由起始点以及一系列关节角度和距离来标识。例如，左肩 1234 可以是起始点，其中右肩被定义为离左肩的距离 L1 (骨 1236)。假定肩线为关节 1234 和 1206 之间直接在 x 方向上延伸的向量，且在 y 或 z 方向上没有分量，则关节 1206 相对于关节 1234 的位置的位置可由 (x-y, x-z, y-z) 平面中的角度 ( $\alpha, 0, 0$ ) 并由长度 L1 来描述。作为另一示例，右肘关节 1208 相对于右肩关节 1206 的位置可由 (x-y, x-z, y-z) 平面中的角度 ( $\alpha, 0, 0$ ) 并由长度 L2 来描述，其中 L2 是右肘关节 1208 和肩关节 1206 之间的距离 (骨 1238)。在该简化示例中，关节 1206 和 1208 两者都被假定为在 x-y 平面上。

[0134] 图 13a 描绘了第一身体类型的 3-D 身体的深度图像的渲染，该深度图像具有图 12 的对应骨架的覆盖图以及 3-D 场景元素。深度图像 1300 包括渲染的 3-D 身体模型 1302。骨架未在该深度图像中渲染，但此处被示出指示身体如何对应于骨架。如上所述，不同身体类型将对应于不同骨架。3-D 场景元素或对象包括地平面或地板 1308、侧墙 1306、后墙 1304 和椅子 1310。

[0135] 图 13b 描绘了第二身体类型的 3-D 身体 1360 的深度图像的渲染，该深度图像具有对应骨架的覆盖图。身体 1360 是比图 13a 中的身体 1302 更瘦、更小的身体类型。

[0136] 图 14 描绘了对应于图 13a 的 3-D 身体的分类图像。如图 8 的步骤 822 中所提到的，分类图像中的每一身体部位可被分配一标识号。也可向每一身体部位分配唯一颜色。身体 1400 中的区域通过向身体模型应用纹理图来获得。示例身体区域包括，在图像空间的右侧：头皮 (1402)、上部的脸 (1404)、下部的脸 (1406)、肩 (1408)、上部的手臂 (1410)、肘 (1412)、下部的手臂 (1414)、手 (1416)、上部的躯干 (1418)、下部的躯干 (1420)、大腿 (1422)、膝 (1424)、小腿 (1426) 和脚 (1428)。示例身体区域包括，在图像空间的左侧：头皮 (1430)、上部的脸 (1432)、下部的脸 (1434)、肩 (1436)、上部的手臂 (1444)、肘 (1442)、下部的手臂 (1440)、手 (1438)、上部的躯干 (1446)、下部的躯干 (1448)、大腿 (1450)、膝 (1452)、小腿 (1454) 和脚 (1456)。

[0137] 使用若干身体部位的图 14 仅是一个示例。任何其他模式或数量 ( $>= 2$ ) 的身体部位也是可能的。

[0138] 图 15a 描绘了一示例深度图像。该深度图像可包括灰色阴影来表示深度，且较浅的颜色表示距离虚拟相机的较大深度。这些灰色阴影可被校准到真实的物理单位。例如，灰值 100 可意味着  $100\text{mm} = 0.1\text{m}$ ，且灰值 2500（在每像素 16 位的深度图像中）可意味着  $2500\text{mm} = 2.5\text{m}$ 。背景可被给予灰值 0。

[0139] 图 15b 描绘了对应于图 15a 的深度图像的示例分类图像。该分类图像包括对应于图 14 中的区域的若干区域。每一区域可用不同的颜色来渲染。

[0140] 在此所述技术的上述详细描述是为了说明和描述而提供的。并非旨在穷举本技术或将其限于所公开的精确形式。鉴于上述教导，许多修改和变型都是可能的。选择上述实施例来最好地解释本技术的原理及其实践应用，从而使本领域其他人能够在各种实施例中并用各种适于所构想的特定用途的修改一起最好地利用本技术。本技术的范围旨在由所附权利要求书来定义。

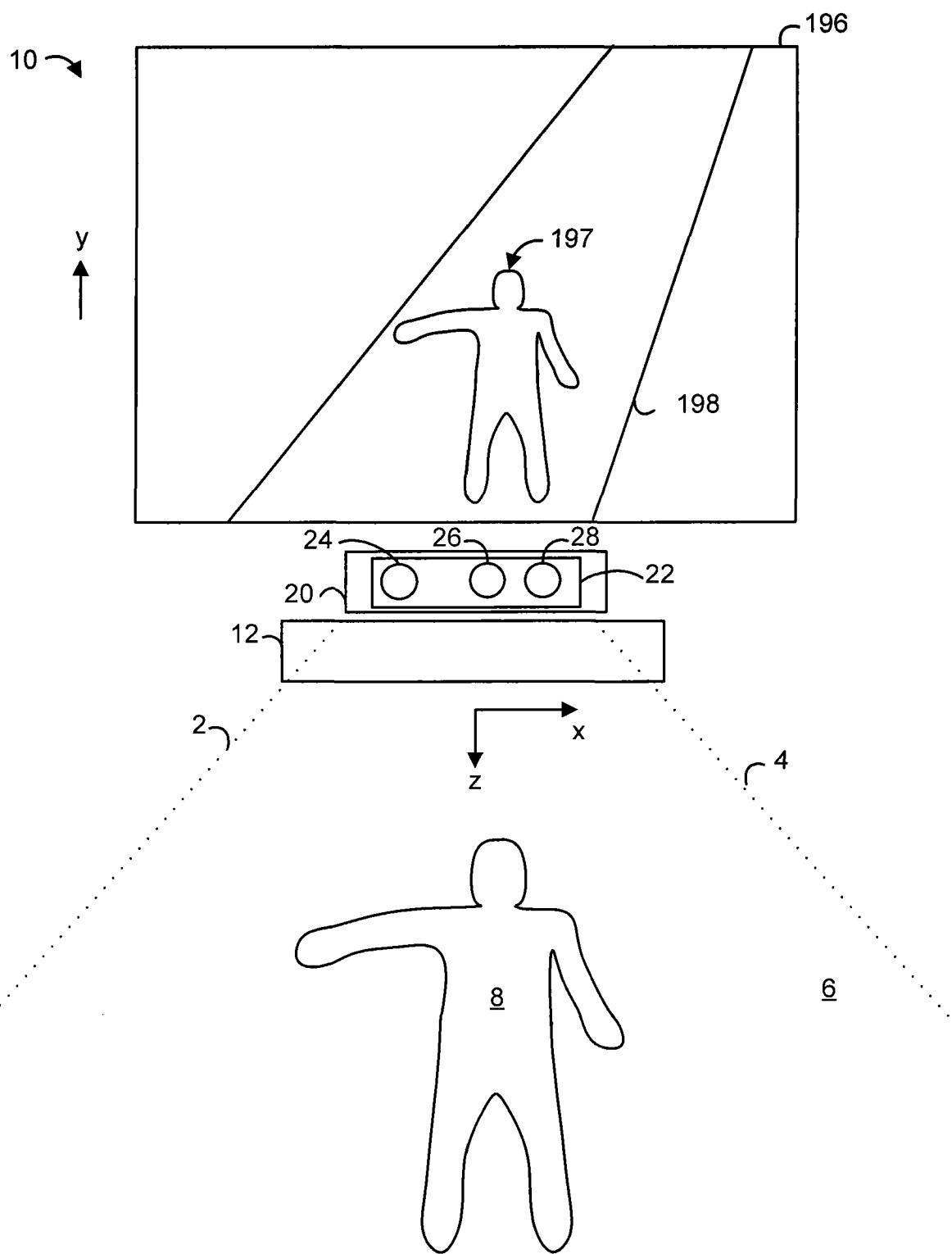


图 1

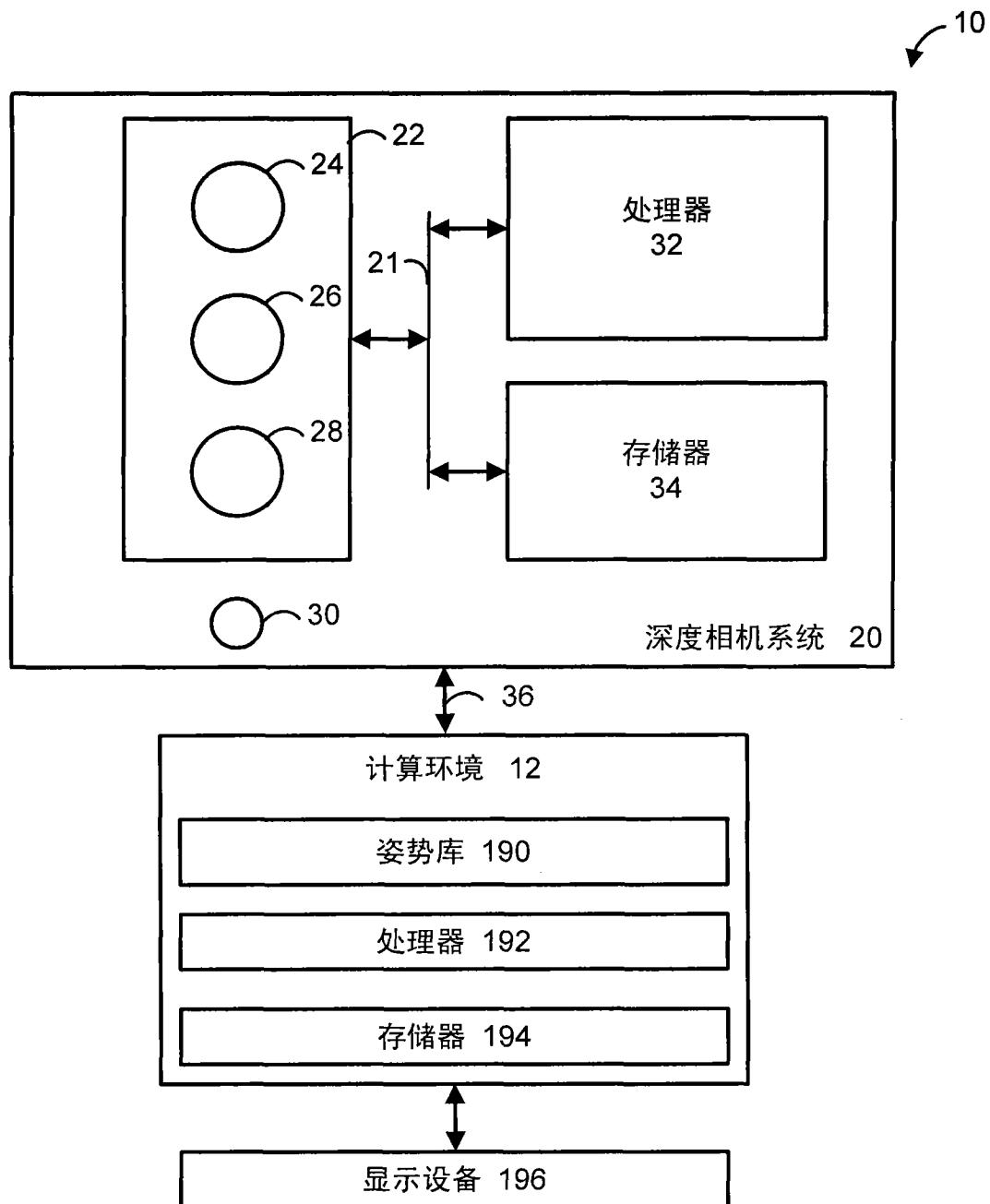


图 2

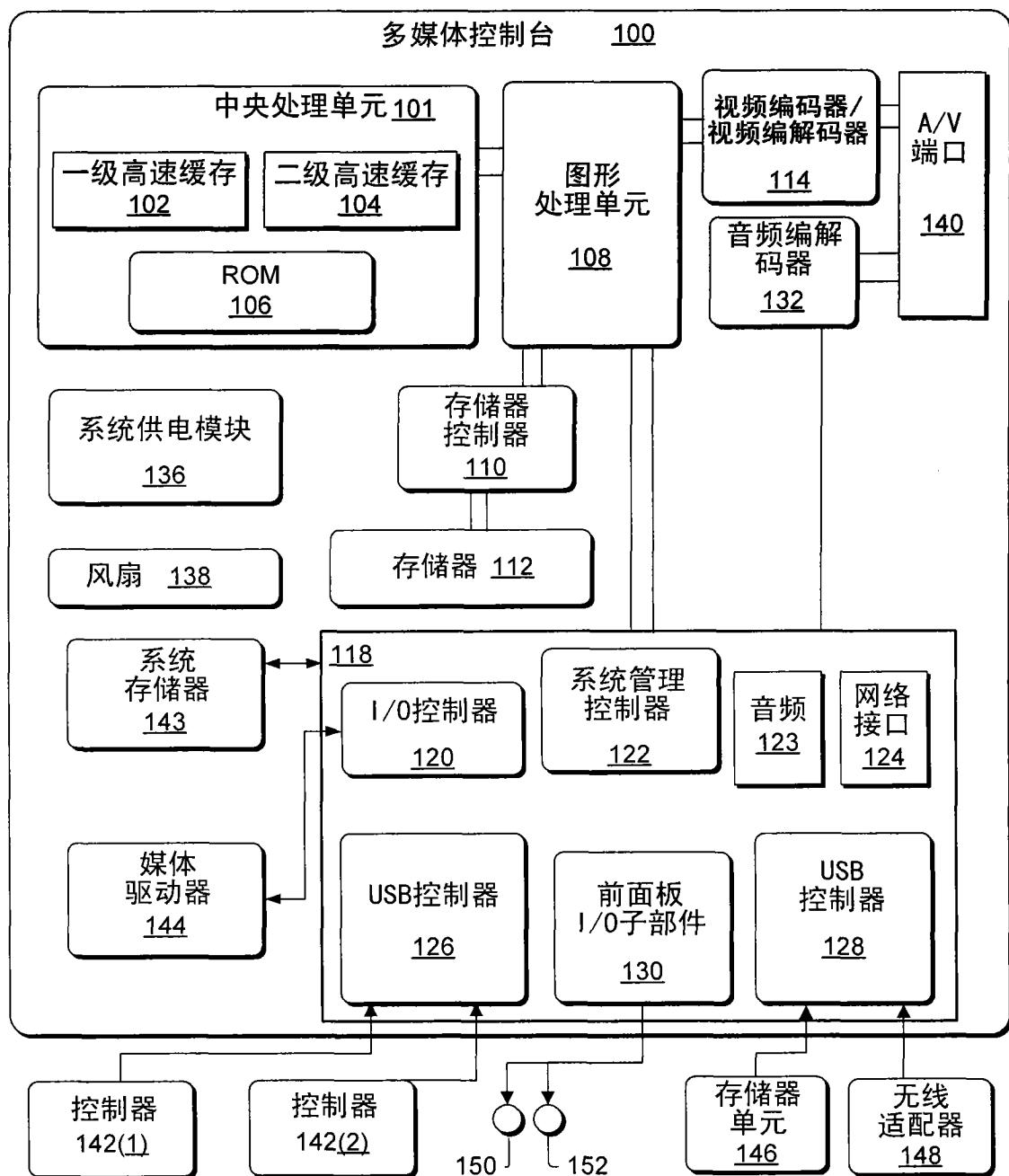


图 3

计算环境 220

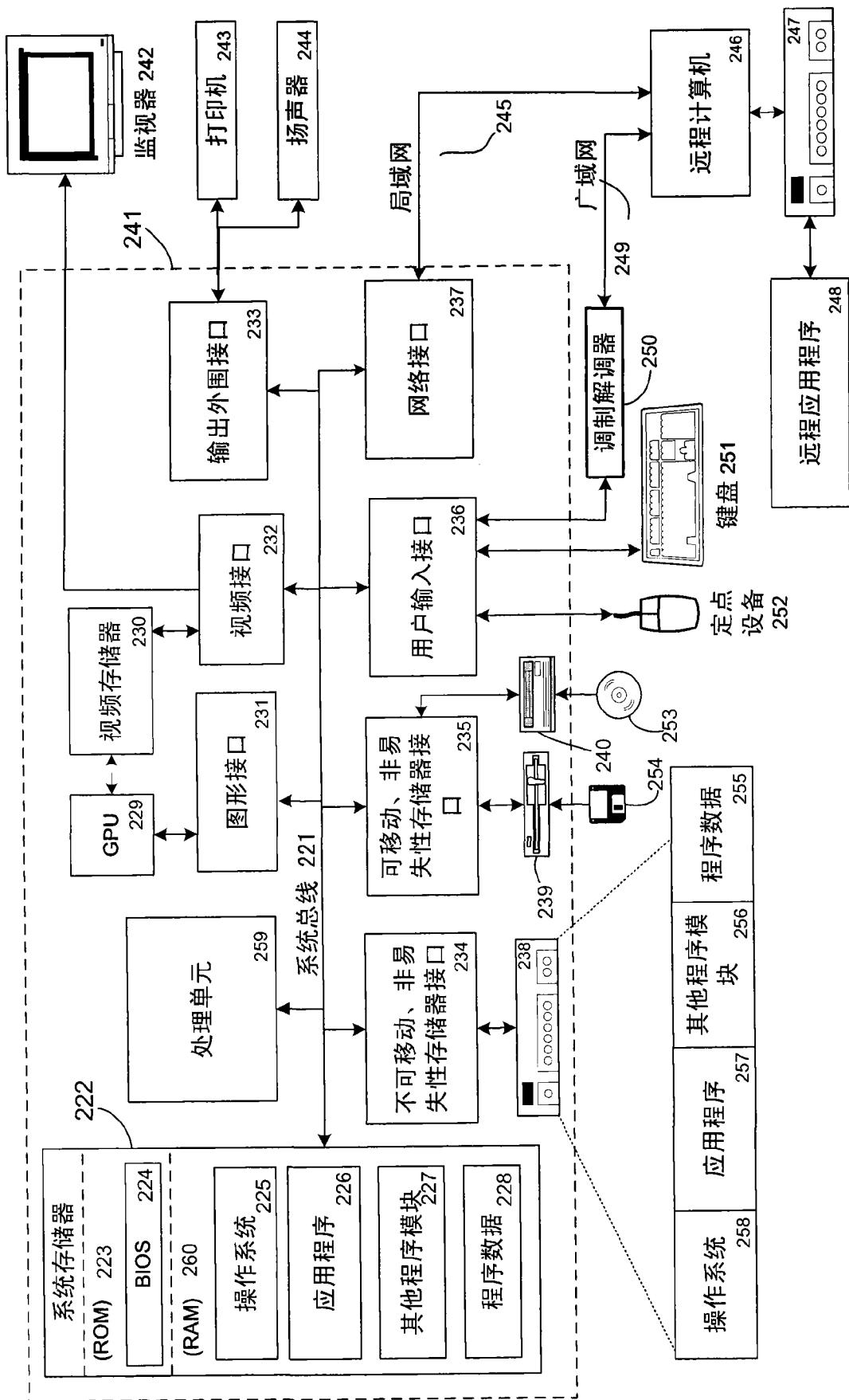


图 4

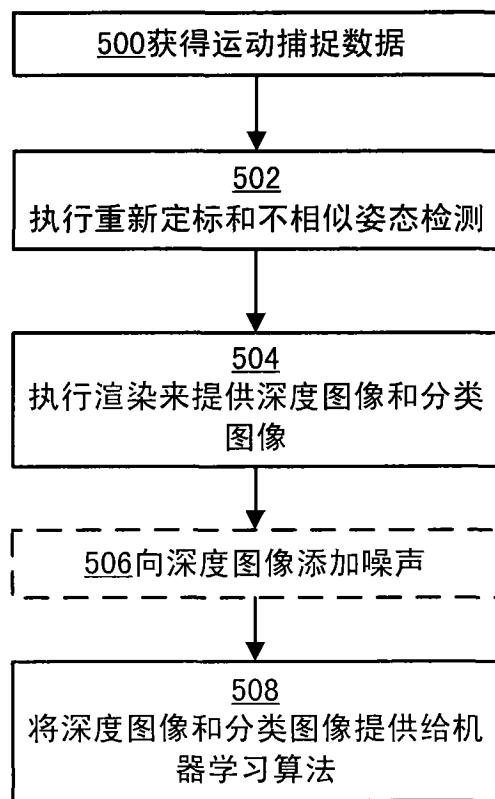


图 5

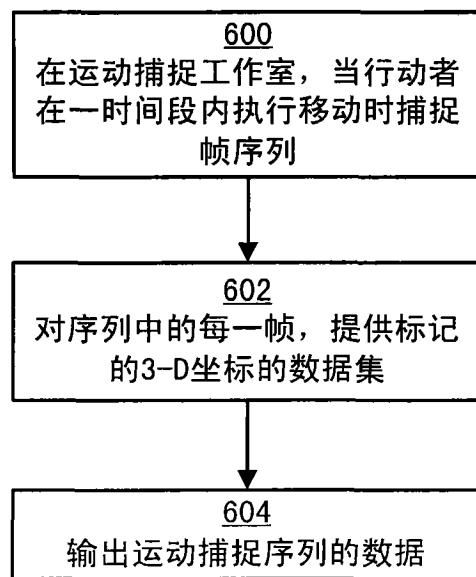


图 6

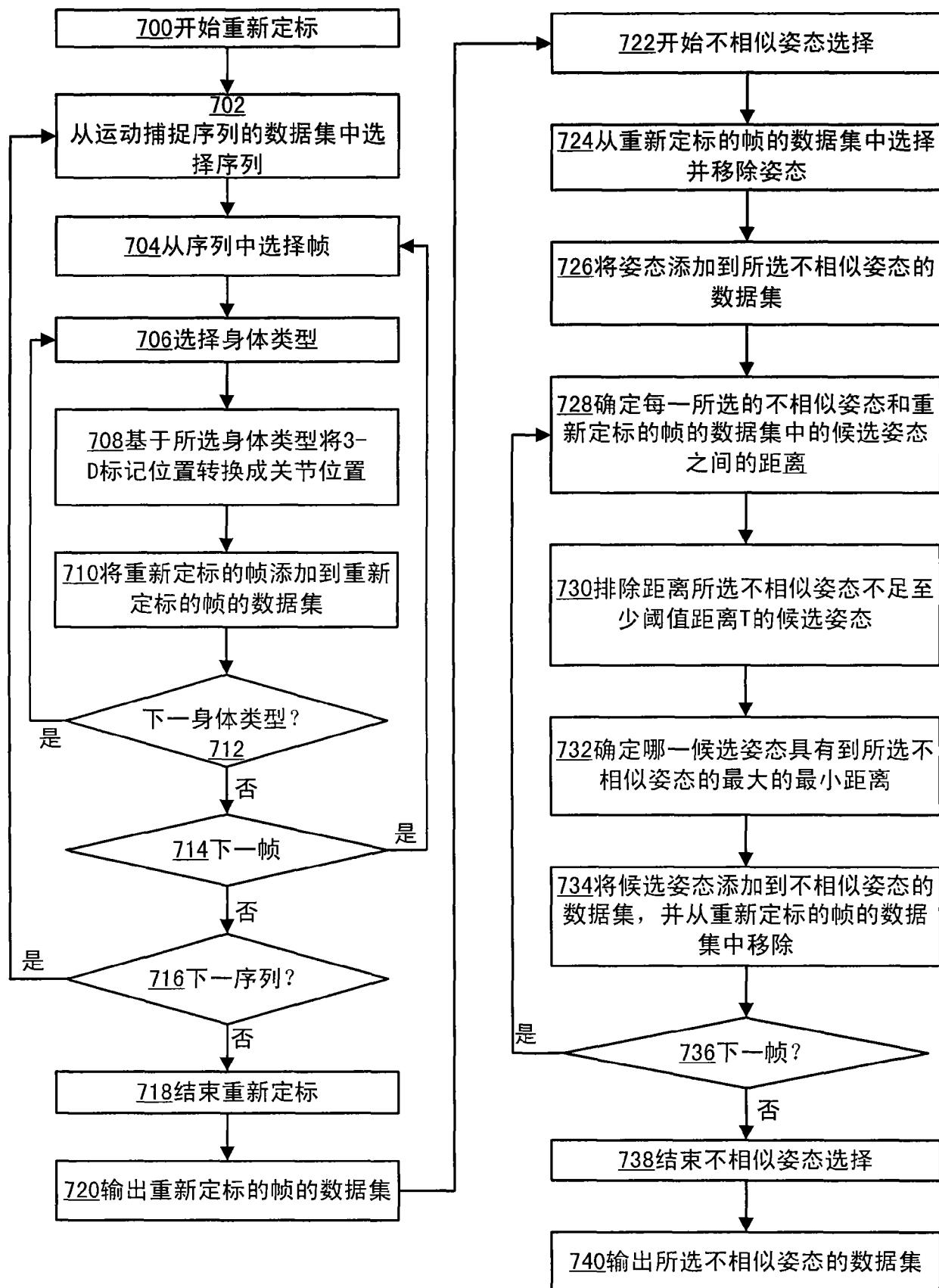


图 7a

用于不相似姿态检测的算法

输入：运动捕捉帧集合 $S'$ 。  
 输入：整数 $D \geq 0$ : 要保留的帧数。  
 输入：距离函数 $\phi$ : 测量帧对之间的距离。  
 输入：距离阈值 $T$ : 保留的帧对应不接近低于 $T$   
 输出：索引到 $S'$  的整数集 $PI$ 。

```

1: 初始化PI为包含 $S'$  的任何单个元素
2: 重复
3:           //1. 计算 $F_{max} = \arg\max_{f \in S'} \min_{g \in PI} \phi(f, S' [g])$ 
4:    $F_{max}$  = 空
5:    $\phi_{max}$  = 0
6:   for  $S'$  中的所有 $f$ , do
7:      $G_{min}$  = 空
8:     for PI中的所有 $g$ , do
9:       if  $G_{min}$  = 空或 $\phi(f, S' [g]) < \phi(f, G_{min})$ 
then
10:       $G_{min}$  =  $S' [g]$ 
11:    end if
12:  end for
13:  if  $\phi(f, G_{min}) > \phi_{max}$  then
14:     $F_{max}$  =  $f$ 
15:     $\phi_{max}$  =  $\phi(f, G_{min})$ 
16:  end if
17: end for
18: //2. 如果距离小于T则终止
19: return if  $\phi_{max} < T$ 
20: //3. 将 $F_{max}$ 添加到PI
21: 将 $F_{max}$ 添加到PI
22: 直到 $size(PI) = D$ 

```

图 7b

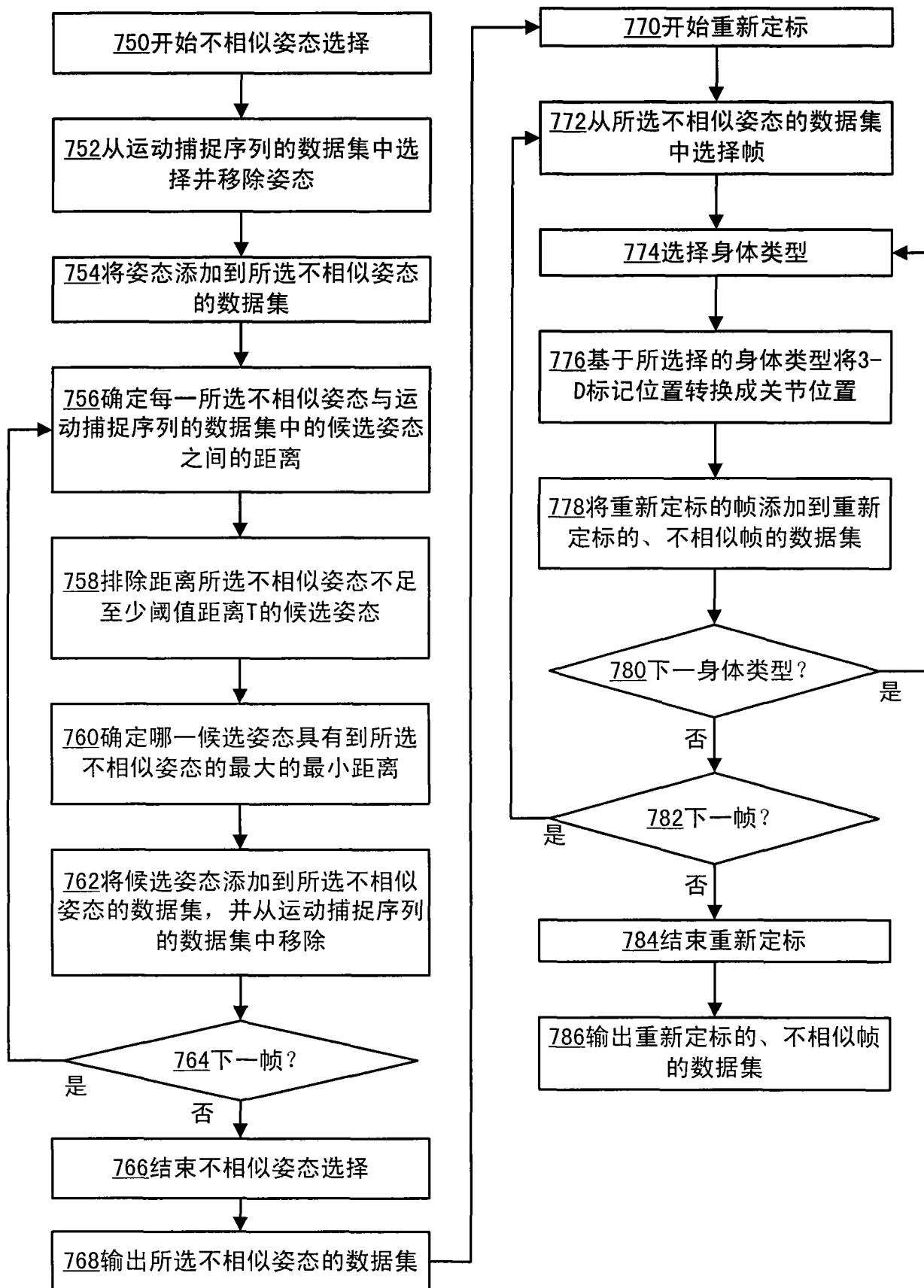


图 7c

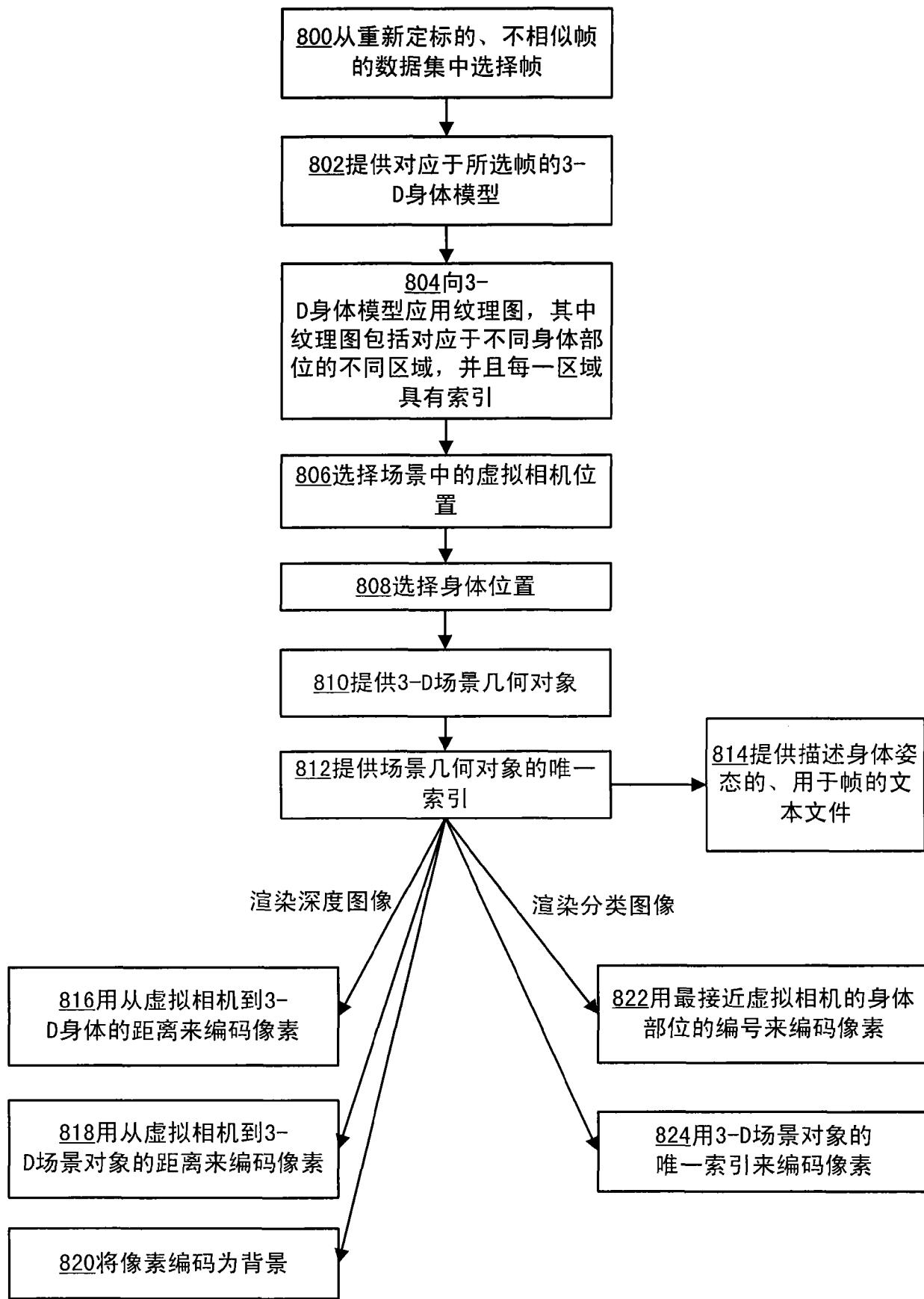


图 8

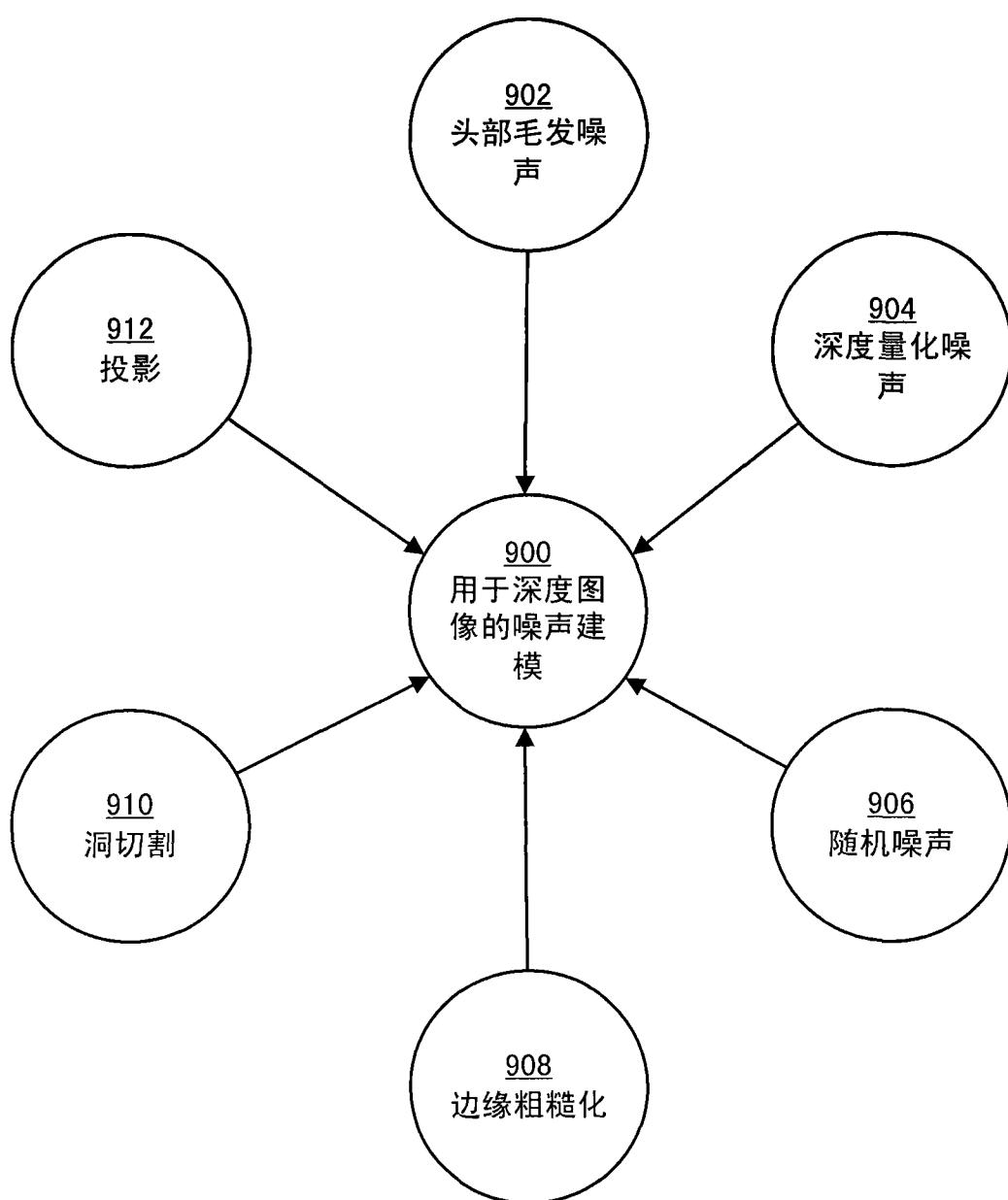


图 9

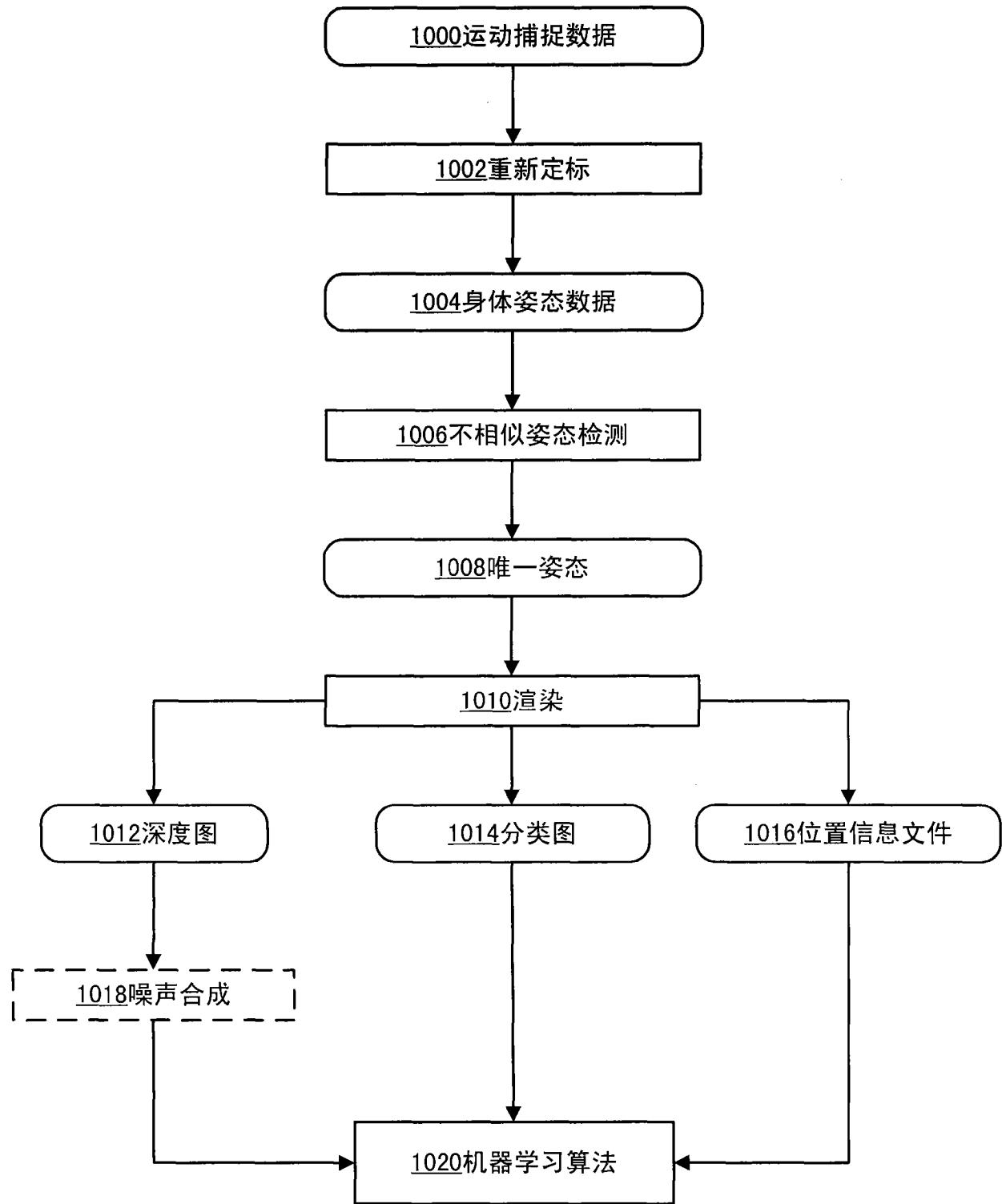


图 10a

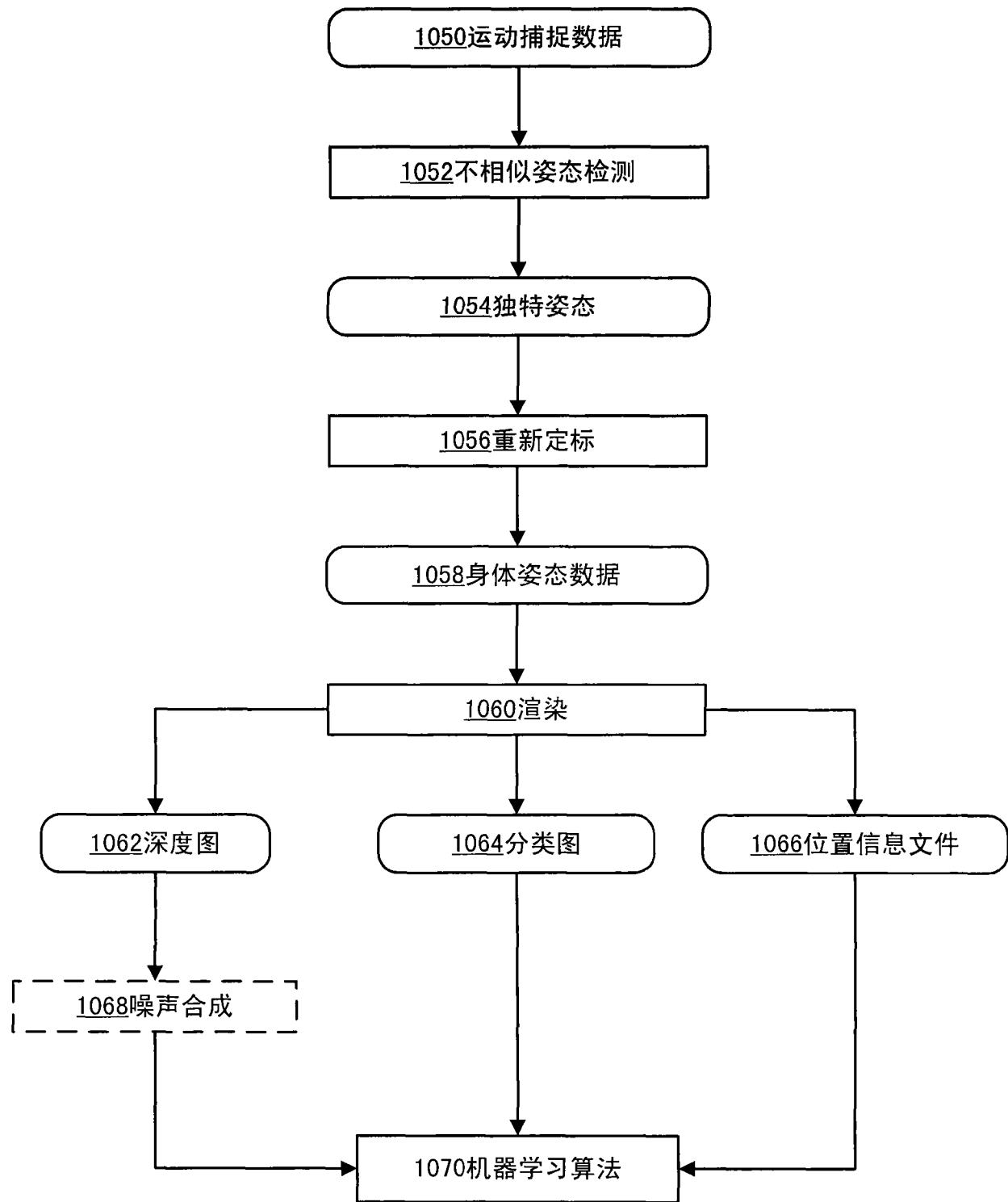


图 10b

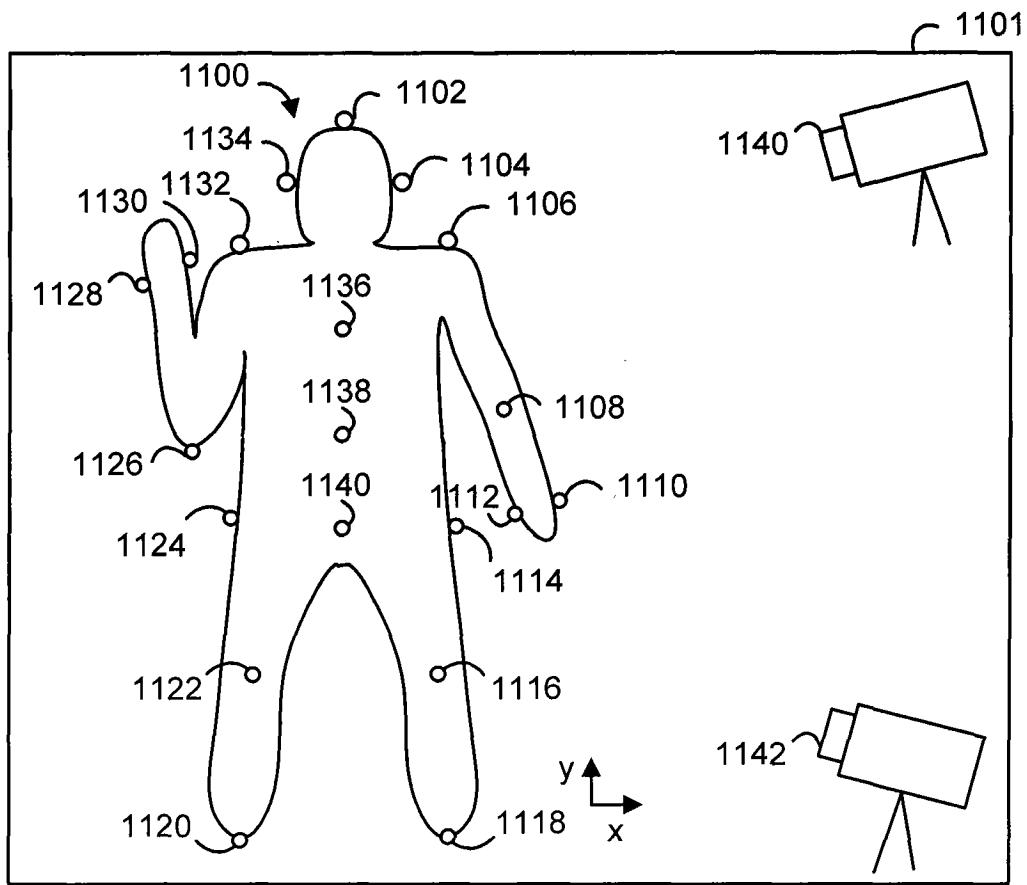


图 11a

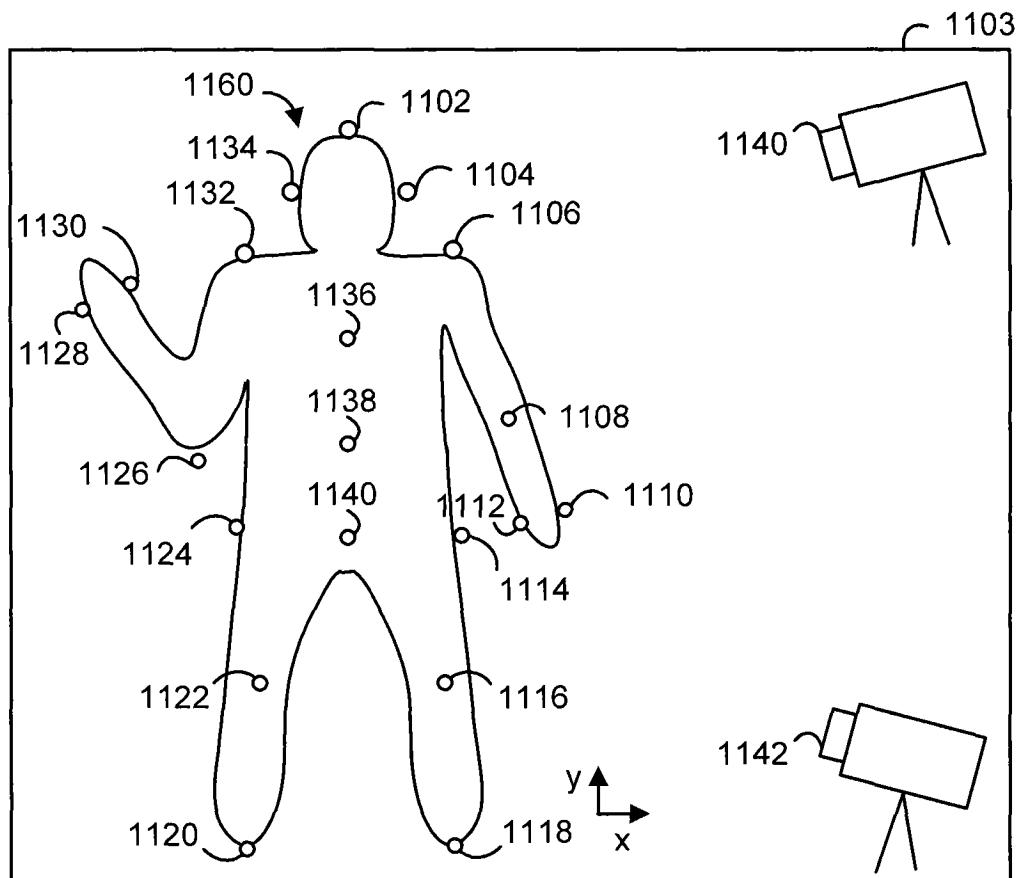


图 11b

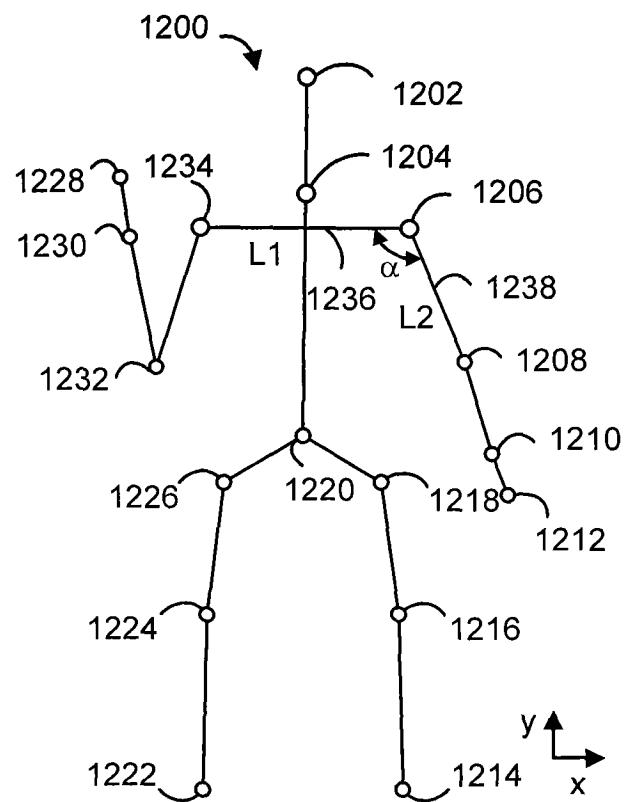


图 12

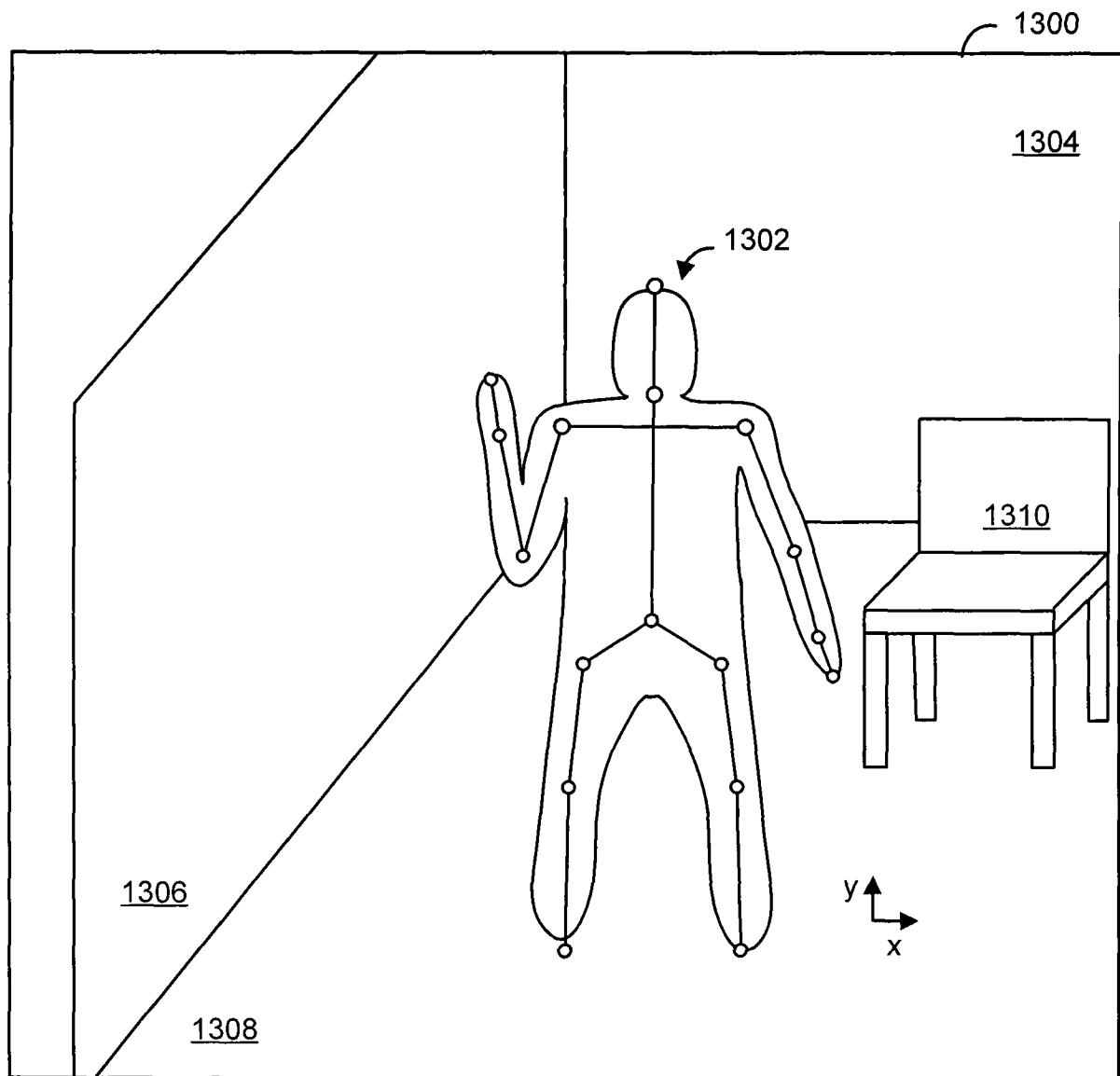


图 13a

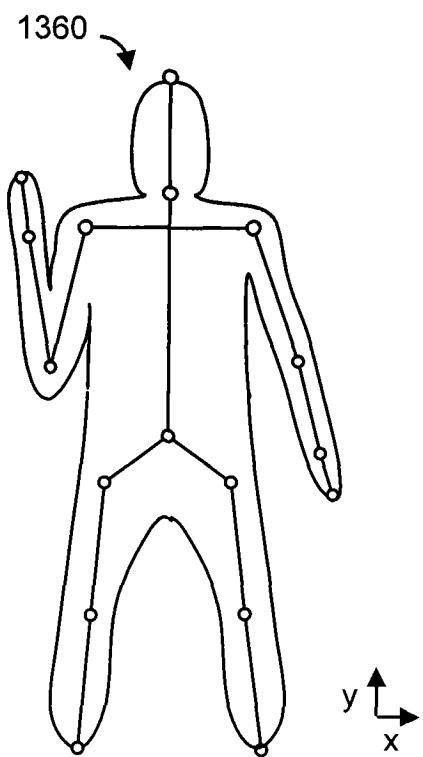


图 13b

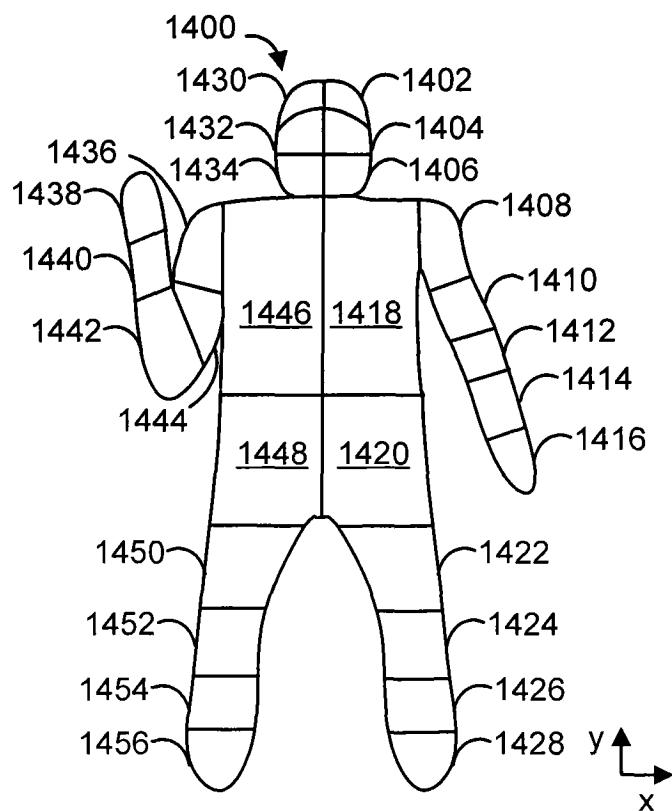


图 14



图 15a

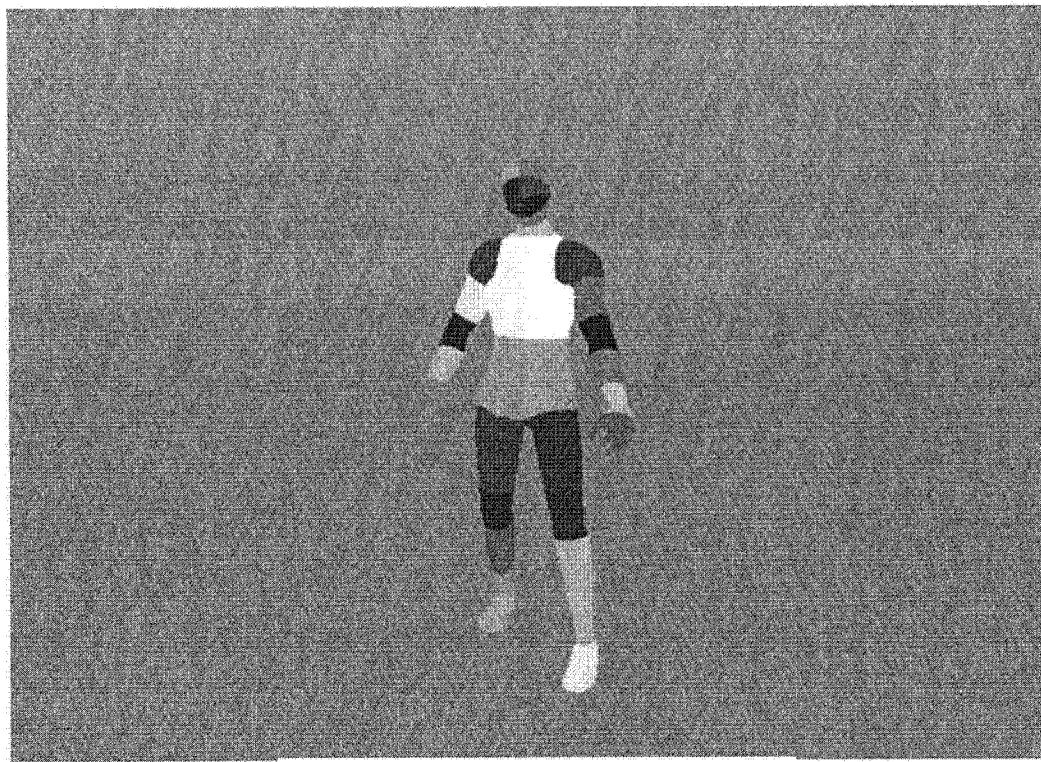


图 15b