



(12) 发明专利申请

(10) 申请公布号 CN 102385237 A

(43) 申请公布日 2012. 03. 21

(21) 申请号 201110285455. 9

(22) 申请日 2011. 09. 07

(30) 优先权数据

12/877, 595 2010. 09. 08 US

(71) 申请人 微软公司

地址 美国华盛顿州

(72) 发明人 S·卡茨 A·阿德勒

(74) 专利代理机构 上海专利商标事务所有限公司
31100

代理人 蔡悦

(51) Int. Cl.

G03B 35/08 (2006. 01)

G03B 15/05 (2006. 01)

H04N 5/232 (2006. 01)

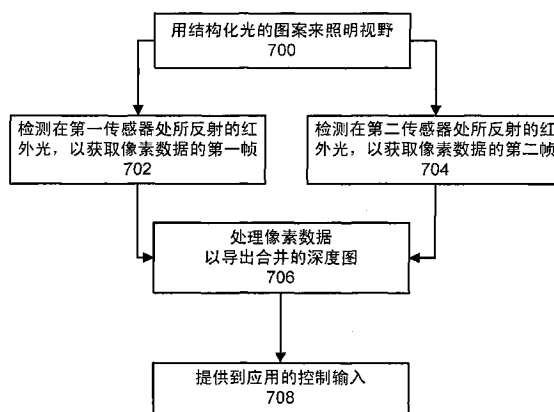
权利要求书 2 页 说明书 19 页 附图 13 页

(54) 发明名称

基于结构化光和立体视觉的深度相机

(57) 摘要

本发明公开了基于结构化光和立体视觉的深度相机。深度相机系统诸如在跟踪视野中的用户的运动的系统中使用结构化光照明器和多个传感器, 诸如红外光检测器。一个传感器可针对较短程检测来优化, 而另一传感器针对较长程检测来优化。这些传感器可具有距该照明器不同的基线距离、以及不同的空间分辨率、曝光时间和敏感性。在一个方法中, 通过匹配到结构化光图案从每一传感器获取各深度值, 并且各深度值被合并, 以获取作为到应用程序的输入来提供的最终深度图。该合并可涉及不加权平均、加权平均、准确性测量和/或置信度测量。在另一方法中, 使用各传感器的像素数据之间的立体匹配来获取包括在该合并中的各附加深度值。



1. 一种深度相机系统,包括:
照明器 (26),所述照明器 (26) 用结构化光的图案来照明视野 (6) 中的对象 (8);
第一传感器 (24),所述第一传感器 (24) 感测从所述对象反射的光,以获取像素数据的第一帧,所述第一传感器针对较短程成像来优化;
第二传感器 (29),所述第二传感器 (29) 感测从所述对象反射的光,以获取像素数据的第二帧,所述第二传感器针对较长程成像来优化;以及
至少一个控制电路 (32、192),所述至少一个控制电路 (32、192) 通过将所述像素数据的第一帧与所述结构化光的图案相比较来导出所述对象的第一结构化光深度图,通过将所述像素数据的第二帧与所述结构化光的图案相比较来导出所述对象的第二结构化光深度图,并且导出基于所述第一和第二结构化光深度图的合并的深度图。
2. 如权利要求 1 所述的深度相机系统,其特征在于:
所述第一传感器与所述照明器之间的基线距离 (BL1) 小于所述第二传感器与所述照明器之间的基线距离 (BL2)。
3. 如权利要求 2 所述的深度相机系统,其特征在于:
所述第一传感器的曝光时间短于所述第二传感器的曝光时间。
4. 如权利要求 2 所述的深度相机系统,其特征在于:
所述第一传感器的敏感性低于所述第二传感器的敏感性。
5. 如权利要求 2 所述的深度相机系统,其特征在于:
所述第一传感器的空间分辨率小于所述第二传感器的分辨率。
6. 如权利要求 1 所述的深度相机系统,其特征在于:
所述第二结构化光深度图包括深度值;以及
在导出所述合并的深度图时,对所述第二结构化光深度图中的深度值给予比所述第一结构化光深度图中的深度值更重的权重。
7. 如权利要求 1 所述的深度相机系统,其特征在于:
所述至少一个控制电路基于所述对象的至少一个立体深度图来导出所述合并的深度图,其中所述至少一个控制电路通过至少以下各项中的至少一个来导出所述至少一个立体深度图:(i) 所述像素数据的第一帧到所述像素数据的第二帧的立体匹配,以及(ii) 所述像素数据的第二帧到所述像素数据的第一帧的立体匹配。
8. 如权利要求 7 所述的深度相机系统,其特征在于:
所述第一和第二结构化光深度图以及所述至少一个立体深度图包括深度值;以及
所述至少一个控制电路将第一组权重分配给所述对象的第一结构化光深度图中的深度值,将第二组权重分配给所述对象的第二结构化光深度图中的深度值,将第三组权重分配给所述对象的第一立体深度图中的深度值,并且导出基于所述第一、第二和第三组权重的所述合并的深度图。
9. 如权利要求 8 所述的深度相机系统,其特征在于:
所述第一组权重基于所述第一传感器与所述照明器之间的基线距离来分配;
所述第二组权重基于所述第二传感器与所述照明器之间的基线距离来分配;以及
所述第三组权重基于所述第一传感器与第二传感器之间的基线距离来分配。
10. 一种用于处理深度相机系统中的图像数据的方法,包括:

用结构化光的图案来照明视野 (6) 中的对象 (8)；

在第一传感器 (24) 处,感测从所述对象反射的光,以获取像素数据的第一帧 (510)；

在第二传感器 (29) 处,感测从所述对象反射的光,以获取像素数据的第二帧；

通过将所述像素数据的第一帧与所述结构化光的图案相比较来导出所述对象的第一结构化光深度图,所述第一结构化光深度图包括所述像素数据的第一帧的像素的深度值；

通过将所述像素数据的第二帧与所述结构化光的图案相比较来导出所述对象的第二结构化光深度图,所述第二结构化光深度图包括所述像素数据的第二帧的像素的深度值；

确定是否期望对像素数据图的第一帧的一个或多个像素的深度值进行细化；以及

如果期望细化,则执行所述像素数据的第一帧的一个或多个像素到所述像素数据的第二帧的一个或多个像素的立体匹配。

11. 如权利要求 10 所述的方法,其特征在于：

在将所述像素数据的第一帧与所述结构化光的图案相比较中,当所述像素数据的第一帧的一个或多个像素没有被成功地匹配到所述结构化光的图案时,期望所述细化。

12. 如权利要求 10 所述的方法,其特征在于：

当所述深度值超出阈值距离时,期望所述细化。

13. 如权利要求 10 所述的方法,其特征在于：

所述第一传感器与所述第二传感器之间的基线距离 (BL1+BL2) 大于所述第一传感器与所述照明器之间的基线距离 (BL1),并且大于所述第二传感器与所述照明器之间的基线距离 (BL2)。

14. 如权利要求 10 所述的方法,其特征在于：

所述立体匹配是针对期望细化的所述像素数据的第一帧的一个或多个像素来执行的,而不是针对不期望细化的所述像素数据的第一帧的一个或多个像素。

15. 如权利要求 10 所述的方法,其特征在于：

如果期望所述细化,则基于所述立体匹配以及所述第一和第二结构化光深度图来提供合并的深度图。

基于结构化光和立体视觉的深度相机

技术领域

[0001] 本发明涉及图像处理技术,具体涉及深度数据的获取。

背景技术

[0002] 实时深度相机能够确定与该相机的视野内的人类或其他对象相距的距离,并且基于该相机的帧速率来基本实时地更新该距离。这样的深度相机可以在运动捕捉系统中例如用于获取关于物理空间中的人体或其他主体的位置和移动的数据,并且可以将该数据用作到计算系统中的应用的输入。可以有许多应用,诸如用于军事、娱乐、体育和医学目的。通常,深度相机包括照明视野的照明器、以及感测来自视野的光以形成图像的图像传感器。然而,各种挑战由于诸如光照条件、表面纹理和色彩、以及遮挡可能性等变量而存在。

发明内容

[0003] 提供了深度相机系统。该深度相机系统使用至少两个图像传感器以及结构化光图像处理和立体图像处理的结合来基本实时地获取场景的深度图。可为由传感器获得的像素数据的每一新的帧更新该深度图。此外,图像传感器可被安装在距照明器不同距离处,并可具有不同特征,以便允许获取更准确的深度图同时降低遮挡的可能性。

[0004] 在一个实施例中,深度相机系统包括用结构化光图案来照明视野中的对象的照明器、至少第一和第二传感器、以及至少一个控制电路。该第一传感器感测从该对象反射的光以获取像素数据的第一帧,并且针对较短程成像来优化。这一优化可按照例如第一传感器与照明器之间的相对较短距离,或第一传感器的相对少的曝光时间、空间分辨率和 / 或对光的敏感性来实现。该深度相机系统进一步包括感测从该对象反射的光以获取像素数据的第二帧的第二传感器,其中,该第二传感器针对较长程成像来优化。这一优化可按照例如第二传感器与照明器之间的相对较长距离,或第二传感器的相对长的曝光时间、空间分辨率和 / 或对光的敏感性来实现。

[0005] 该深度相机系统进一步包括至少一个控制电路,该控制电路可与传感器和照明器在公共外壳中、和 / 或在诸如计算环境等分开的组件中。该至少一个控制电路通过将像素数据的第一帧与结构化光的图案相比较来导出对象的第一结构化光深度图,通过将像素数据的第二帧与结构化光的图案相比较来导出对象的第二结构化光深度图,并且导出基于第一和第二结构化光深度图的合并的深度图。每一深度图可包括诸如在像素网格中的每一像素位置的深度值。

[0006] 在另一方面,立体图像处理还被用于细化深度值。例如,当像素数据的第一和 / 或第二帧中的一个或多个像素没有被成功地匹配到结构化光的图案时,或当深度值指示需要较大基线以实现良好的准确性的大距离时,立体图像处理的使用可被触发。以此方式,仅按需向深度值提供进一步细化,以便避免不必要的处理步骤。

[0007] 在某些情况下,由传感器获取的深度数据可基于该传感器的特征来分配权重,和 / 或基于深度值中的置信度来分配准确性度量。

[0008] 最终深度图可被用作对例如运动捕捉系统中的应用的输入,其中,该对象是由该运动捕捉系统跟踪的人类,并且该应用响应于诸如通过对化身动画化、导航屏幕上菜单、或执行某一其他动作之类的人类的姿势或移动来改变该运动捕捉系统的显示。

[0009] 提供本发明内容以便以简化形式介绍将在以下的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

[0010] 在附图中,相同编号的元件彼此对应。

[0011] 图 1 描绘了运动捕捉系统的示例实施例。

[0012] 图 2 描绘了图 1 的运动捕捉系统的示例框图。

[0013] 图 3 描绘了可以在图 1 的运动捕捉系统中使用的计算环境的示例框图。

[0014] 图 4 描绘了可以在图 1 的运动捕捉系统中使用的计算环境的另一个示例框图。

[0015] 图 5A 描绘了在结构化光系统中的照明帧和所捕捉的帧。

[0016] 图 5B 描绘了在立体光系统中的两个所捕捉的帧。

[0017] 图 6A 描绘了带有在照明器的公共侧的两个传感器的成像组件。

[0018] 图 6B 描绘了带有在照明器的一侧的两个传感器、以及在该照明器的相对侧的一个传感器的成像组件。

[0019] 图 6C 描绘了带有在照明器的公共侧的三个传感器的成像组件。

[0020] 图 6D 描绘了带有在照明器的相对侧的两个传感器的成像组件,其示出这两个传感器如何感测对象的不同部分。

[0021] 图 7A 描绘了用于获取视野的深度图的过程。

[0022] 图 7B 描绘了图 7A 的步骤 706 的进一步细节,其中两个结构化光深度图被合并。

[0023] 图 7C 描绘了图 7A 的步骤 706 的进一步细节,其中两个结构化光深度图和两个立体深度图被合并。

[0024] 图 7D 描绘了图 7A 的步骤 706 的进一步细节,其中,使用立体匹配按需对各深度值进行细化。

[0025] 图 7E 描绘了图 7A 的步骤 706 的另一方法的进一步细节,其中,使用立体匹配按需对合并的深度图的各深度值进行细化。

[0026] 图 8 描绘了如在图 7A 的步骤 708 中阐述的用于使用控制输入来跟踪人类目标的示例方法。

[0027] 图 9 描绘了如在图 8 的步骤 808 中阐述的人类目标的示例模型。

具体实施方式

[0028] 提供了深度相机以供在跟踪视场中的一个或多个对象中使用。在示例实现中,该深度相机被用在运动跟踪系统中以跟踪人类用户。该深度相机包括两个或更多个传感器,该两个或更多个传感器被优化以解决诸如光照条件、表面纹理和色彩、以及遮挡可能性之类的变量。该优化可包括优化各传感器相对于彼此之间的放置和相对于照明器的放置,以及优化各传感器的空间分辨率、敏感性和曝光时间。该优化还可包括诸如通过将像素数据

的帧匹配到结构化光的图案和 / 或通过将像素数据的一帧匹配到另一帧来优化如何获取深度图数据。

[0029] 如此处描述的多个传感器的使用提供了胜于其他方法的优点。例如,除立体相机以外的实时深度相机往往提供可嵌入在 2-D 矩阵上的深度图。这样的相机有时被称为 2.5D 相机,因为它们通常使用单个成像设备来提取深度图,使得没有给出被遮挡对象的信息。立体深度相机往往获取对两个或更多个相机可见的相当稀疏的位置测量。同样,当对诸如白墙之类的平滑的无纹理 (textureless) 表面进行成像时,它们运作得不好。某些深度相机使用结构化光来测量 / 标识由作为成像系统设备的传感器与远离传感器的作为光投影设备的照明器之间的视差造成的失真。这一方法固有地产生具有由于对传感器可见但对照明器不可见的阴影位置而引起的丢失信息的深度图。另外,外部光有时可使结构化图案对相机不可见。

[0030] 上述缺点可通过使用两个或更多个传感器和单个照明设备的星座图来克服,以便如同使用了三个深度相机那样有效地提取 3D 样本。该两个传感器可通过匹配到结构化光图案来提供深度图像,而第三相机通过应用立体技术对来自两个传感器的两个图像进行匹配来实现。通过应用数据融合,增强包括对相机间中断的稳健性等 3D 测量的稳健性是可能的。我们提供了使用两个传感器与单个投影仪来实现两个深度图,使用结构化光技术、将结构化光技术与立体技术结合、并且在融合过程中使用以上来实现具有减小的遮挡和增强的稳健性的 3D 图像。

[0031] 图 1 描绘了运动捕捉系统 10 的示例实施例,其中,人 8 如在用户的家中与应用程序进行交互。运动捕捉系统 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 可包括成像组件 22,其具有诸如红外 (IR) 光发射器之类的照明器 26、诸如红外照相机之类的图像传感器 26、以及色彩 (诸如红 - 绿 - 蓝 (RGB)) 照相机 28。一个或多个对象,诸如也被称为用户、个人或玩家之类的人 8,站在深度相机的视野 6 内。线 2 和 4 表示视野 6 的边界。在此示例中,深度相机系统 20 以及计算环境 12 提供应用程序,其中,显示器 196 上的化身 197 追踪人 8 的移动。例如,当人抬起手臂时,化身可以抬起手臂。化身 197 站在 3-D 虚拟世界的公路 198 上。可以定义笛卡儿世界坐标系,其包括沿着深度相机系统 20 的焦距延伸 (例如,水平地) 的 z 轴、垂直地延伸的 y 轴、以及横向地并且水平地延伸的 x 轴。注意,因为显示器 196 在 y 轴方向垂直地延伸,且 z 轴从深度相机系统垂直于 y 轴和 x 轴并平行于用户 8 所站立的地面地伸出,附图的透视被修改为简化。

[0032] 一般而言,运动捕捉系统 10 被用来识别、分析、和 / 或跟踪一个或多个个人目标。计算环境 12 可包括计算机、游戏系统或控制台等等,以及执行应用程序的硬件组件和 / 或软件组件。

[0033] 深度相机系统 20 可被用来在视觉上监视一个或多个诸如人 8 之类的人,以便可以捕捉、分析、以及跟踪由人执行的姿势和 / 或移动,以在应用程序内执行一个或多个控制或动作,如动画化化身或屏幕上人物或选择用户界面 (UI) 中的一个菜单项。深度相机系统 20 在下文中更详细讨论。

[0034] 运动捕捉系统 10 可以连接到诸如显示器 196 之类的视听设备,例如,电视机、监视器、高清晰度电视机 (HDTV) 等等,或者甚至向用户提供视觉和音频输出的墙或其他表面上的投影。还可以经由单独的设备来提供音频输出。为驱动显示器,计算环境 12 可包括提供

与应用程序相关联的视听信号的诸如图形卡之类的视频适配器和 / 或诸如声卡之类的音频适配器。显示器 196 可连接到计算环境 12。

[0035] 可以使用深度相机系统 20 来跟踪人 8, 以使得用户的姿势和 / 或移动被捕捉并用于动画化身或屏幕上人物, 和 / 或被解释为对正在由计算机环境 12 执行的应用程序的输入控制。

[0036] 人 8 的一些移动可被解释为可对应于除了控制化身之外的动作的控制。例如, 在一个实施方式中, 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, 与朋友进行交流等等。玩家可以使用移动来从主用户界面中选择游戏或其他应用程序, 或以别的方式导航选项菜单。如此, 可以有人 8 的全套运动可用, 可使用它们, 并以任何合适的方式分析它们, 以与应用程序进行交互。

[0037] 运动捕捉系统 10 还可以用于将目标移动解释成处于游戏或意味着娱乐和休闲的其他应用程序的领域之外的操作系统和 / 或应用程序控制。例如, 操作系统和 / 或应用程序的实质上任何可控方面都可以由人 8 的移动来控制。

[0038] 图 2 描绘了图 1a 的运动捕捉系统 10 的示例框图。深度相机系统 20 可以被配置为通过任何合适的技术捕捉带有深度信息 (包括可以包括深度值的深度图像) 的视频, 这些技术包括例如飞行时间、结构化光、立体图像等等。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。

[0039] 深度相机系统 20 可包括成像组件 22, 该成像组件 22 捕捉物理空间中的场景的深度图像。深度图像或深度图可包括所捕捉的场景的二维 (2-D) 像素区域, 其中 2-D 像素区域中的每一像素都可具有表示与成像组件 22 的线性距离的相关联的深度值, 从而提供 3-D 深度图像。

[0040] 成像组件 22 的各种配置是可能的。在一个方法中, 成像组件 22 包括照明器 26、第一图像传感器 (S1) 24、第二图像传感器 (S2) 29、以及可见色彩相机 28。传感器 S1 和 S2 可被用于捕捉场景的深度图像。在一个方法中, 照明器 26 是红外 (IR) 光发射器, 并且第一和第二传感器是红外光传感器。通过照明器 26 和一个或多个传感器 26 的组合来形成 3-D 深度相机。

[0041] 使用各种技术可由每一传感器获取深度图。例如, 深度相机系统 20 可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, 被显示成诸如网格图案或条纹图案等已知图案的光) 由照明器 26 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传感器 24 或 29 和 / 或色彩相机 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。

[0042] 在一个可能的方法中, 传感器 24 和 29 位于照明器 26 的相对侧, 并且处于与该照明器的不同基线距离。例如, 传感器 24 位于与照明器 26 的距离 BL1 处, 并且传感器 29 位于与照明器 26 的距离 BL2 处。传感器与照明器之间的距离可按照传感器与照明器的中心点 (诸如光轴) 之间的距离来表示。使各传感器在照明器的相对侧的一个优点是视野中对象的遮挡区域可被减小或消除, 因为各传感器从不同角度看到对象。同样, 可通过将传感器放置在相对接近于照明器来优化该传感器以供查看视野内较近的对象, 而可通过将另一传感器放置在相对远离照明器来优化该另一传感器以供查看视野内较远的对象。例如, 在 $BL2 > BL1$ 的情况下, 传感器 24 可被认为是针对较短程成像来优化, 而传感器 29 可被认为是针

对较长程成像来优化的。在一个方法中,传感器 24 和 29 可以共线,使得它们被沿着穿过照明器的公共线放置。然而,关于传感器 24 和 29 的定位的其他配置是可能的。

[0043] 例如,传感器可围绕要扫描的对象圆周地安排、或围绕全息图要被投影的位置圆周地安排。安排多个深度相机系统,各自具有围绕对象的一照明器和各传感器也是可能的。这可允许查看对象的不同侧,从而提供围绕该对象的旋转视图。通过使用更多深度相机,我们添加了该对象的更多可见区域。可以有两个深度相机,一个在对象的前方而一个在对象的后方,只要它们在其照明上彼此不盲就彼此对准。每一深度相机可感测从对象反射的其自身的结构化光图案。在另一示例中,两个深度相机被安排成彼此成 90 度。

[0044] 深度相机系统 20 可包括与 3-D 深度照相机 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元网格的指令;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元的指令;用于确定被隔离的人类目标的一个或多个肢端的位置或定位的指令;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。

[0045] 处理器 32 可访问存储器 31 以使用导出结构化光深度图的软件 33、导出立体视觉深度图的软件 34、以及执行深度图合并计算的软件 35。处理器 32 可被认为是至少一个控制电路,该至少一个控制电路通过将像素数据的帧与由照明平面中的照明器发射的结构化光的图案相比较来导出对象的结构化光深度图。例如,使用软件 33,该至少一个控制电路可通过将由传感器 24 获取的像素数据的第一帧与由照明器 26 发射的结构化光的图案相比较来导出对象的第一结构化光深度图,并且通过将由传感器 29 获取的像素数据的第二帧与该结构化光的图案相比较来导出对象的第二结构化光深度图。该至少一个控制电路可使用软件 35 来导出基于第一和第二结构化光深度图的合并的深度图。结构化光深度图在下文进一步讨论,例如参考图 5A。

[0046] 同样,该至少一个控制电路可使用软件 34 通过由传感器 24 获取的像素数据的第一帧到由传感器 29 获取的像素数据的第二帧的立体匹配来导出对象的至少第一立体深度图,并且通过像素数据的第二帧到像素数据的第一帧的立体匹配来导出对象的至少第二立体深度图。软件 25 可合并一个或多个结构化光深度图和 / 或立体深度图。立体深度图在下文进一步讨论,例如参考图 5B。

[0047] 该至少一个控制电路可由同样在深度相机系统外部的处理器来提供,诸如处理器 192 或任何其他处理器。该至少一个处理电路可访问来自存储器 31 的软件,该存储器 31 例如可以是有形计算机可读存储,该有形计算机可读存储其上包含有用于对至少一个处理器或控制器 32 进行编程以执行如此处描述的用于处理深度相机系统中的深度数据的方法的计算机可读软件。

[0048] 存储器 31 可存储由处理器 32 执行的指令,以及存储诸如由传感器或色彩相机捕捉的像素数据 36 的帧之类的图像。例如,存储器 31 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 31 可以是经由总线 21 与图像捕捉组件 22 和处理器 32 通信的单独的组件。根据另一实施例,存储器组件 31 可被集成到处理器 32 和 / 或图像捕捉组件 22 中。

[0049] 深度相机系统 20 可以经由诸如有线和 / 或无线连接等通信链路 37 与计算环境 12

进行通信。计算环境 12 可经由通信链路 37 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。

[0050] 另外,深度相机系统 20 可以通过通信链路 37 向计算环境 12 提供深度信息和由例如图像传感器 24 和 29 和 / 或色彩相机 28 捕捉到的图像,和 / 或可以由深度相机系统 20 所生成的骨架模型。然后,计算环境 12 可以使用模型、深度信息、以及捕捉到的图像来控制应用程序。例如,如图 2 所示,计算环境 12 可包括诸如姿势过滤器的集合之类的姿势库 190,每一个过滤器都具有关于可以由骨架模型(随着用户移动)执行的姿势的信息。例如,可以为各种手势(如手的猛击或投掷)提供姿势过滤器。通过将检测到的运动与每一个过滤器进行比较,可以标识由人执行的指定的姿势或移动。还可以确定执行移动的程度。

[0051] 可以将由深度相机系统 20 以骨架模型的形式捕捉的数据以及与它相关联的移动与姿势库 190 中的姿势过滤器进行比较,以标识(如由骨架模型所表示的)用户何时执行了一个或多个特定移动。那些移动可以与应用程序的各种控制相关联。

[0052] 计算环境也可以包括处理器 192,其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。

[0053] 图 3 描绘了可以在图 1 的运动捕捉系统中使用的计算环境的示例框图。可以使用计算环境来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。上文所描述的诸如计算环境 12 等的计算环境可包括诸如游戏控制台等的多媒体控制台 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内核,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 之类的存储器 106 可存储当多媒体控制台 100 通电时在引导过程的初始阶段期间加载的可执行代码。

[0054] 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,以便于处理器对各种类型的存储器 112,比如 RAM(随机存取存储器)的访问。

[0055] 多媒体控制台 100 包括优选地在模块 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 以及前面板 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、电缆调制解调器等各种不同的有线或无线适配器组件中任何一种。

[0056] 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。提供了媒体驱动器 144,其可以包括 DVD/CD 驱动器、硬盘驱动器、或其他可移动媒体驱动器。媒体驱动器 144 可以在多媒体控制台 100 的内部或外部。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 总线或其他高速连接等总线连接到 I/O 控制器 120。

[0057] 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。

[0058] 前面板 I/O 子部件 130 支持暴露在中媒体控制台 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。

[0059] CPU 101、GPU 108、存储器控制器 110、和多媒体控制台 100 内的各个其他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。

[0060] 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/或其他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。

[0061] 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,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。

[0062] 当多媒体控制台 100 通电时,可以保留指定量的硬件资源以供多媒体控制台操作系统使用。这些资源可包括预留存储器(例如,16MB)、CPU 和 GPU 周期(例如,5%)、网络带宽(例如,8kbs)等等。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。

[0063] 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。

[0064] 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,从而无需改变频率并引起 TV 重新同步。

[0065] 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。系统功能被封装在上述所保留的系统资源中执行的一组系统应用中。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。

[0066] 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)。

[0067] 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，并且驱动程序维护有关焦点切换的状态信息。控制台 100 可以从图 2 的包括传感器 24 和 29 的深度相机系统 20 接收附加的输入。

[0068] 图 4 描绘了可以在图 1 的运动捕捉系统中使用的计算环境的另一个示例框图。在运动捕捉系统中，可以使用计算环境来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 包括通常包括各种有形计算机可读存储介质的计算机 241。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。作为示例而非限制，图 4 描绘了操作系统 225、应用程序 226、其他程序模块 227，以及程序数据 228。

[0069] 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，例如，读写不可移动、非易失性磁性介质的硬盘驱动器 238，读写可移动、非易失性磁盘 254 的磁盘驱动器 239，以及读写诸如 CD ROM 或其他光学介质之类的可移动、非易失性光盘 253 的光盘驱动器 240。可以在该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。

[0070] 上文所讨论的并且在图 4 中所描绘的驱动器以及其相关联的计算机存储介质为计算机 241 提供了计算机可读指令、数据结构、程序模块及其他数据的存储。例如，硬盘驱动器 238 被描绘为存储了操作系统 258、应用程序 257、其他程序模块 256、以及程序数据 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。在此给操作系统 258、应用程序 257、其他程序模块 256、以及程序数据 255 提供了不同的编号，以说明至少它们是不同的副本。用户可以通过诸如键盘 251 和定点设备 252（通常被称为鼠标、跟踪球或触摸垫）之类的输入设备向计算机 241 输入命令和信息。其他输入设备（未示出）可包括话筒、游戏杆、游戏手柄、圆盘式卫星天线、扫描仪等。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，但也可以由诸如并行端口、游戏端口或通用串行总线 (USB) 等其他接口和总线结构来进行连接。图 2 的深度相机系统 20，包括传感器 24 和 29，可以为控制台 100 定义附加的输入设备。监视器 242 或其他类型的显示器也通过接口，诸如视频接口 232，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。

[0071] 计算机 241 可使用至诸如远程计算机 246 之类的一个或多个远程计算机的逻辑连接在网络化环境中操作。远程计算机 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,通常包括上文相对于计算机 241 所描述的许多或全部元件,但是图 4 中只描绘了存储器存储设备 247。逻辑连接包括局域网 (LAN) 245 和广域网 (WAN) 249,但也可以包括其他网络。这样的联网环境在办公室、企业范围计算机网络、内联网和因特网中是常见的。

[0072] 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 建立通信的其他手段。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 描绘了驻留在存储器设备 247 上的远程应用程序 248。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。

[0073] 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,计算机可读软件用于对至少一个处理器进行编程以执行如此处描述的用于处理深度相机系统中的图像数据的方法。有形计算机可读存储可包括,例如,组件 31、194、222、234、235、230、253 和 254 中的一个或多个。处理器可包括,例如,组件 32、192、229 和 259 中的一个或多个。

[0074] 图 5A 描绘了在结构化光系统中的照明帧和所捕捉的帧。照明帧 500 表示照明器的图像平面,该照明器将结构化光发射到照明器的视野中的对象 520 上。照明帧 500 包括具有 x_2 、 y_2 和 z_2 正交轴的轴系统。 F_2 是照明器的焦点,以及 O_2 是轴系统的原点,诸如位于照明帧 500 的中心处。所发射的结构化光可包括条纹、点或其他已知的照明图案。类似地,所捕捉的帧 510 表示诸如参考图 2 讨论的传感器 24 或 29 之类的传感器的图像平面。所捕捉的帧 510 包括具有 x_1 、 y_1 和 z_1 正交轴的轴系统。 F_1 是传感器的焦点,以及 O_1 是轴系统的原点,诸如位于所捕捉的帧 510 的中心处。尽管不是必需的,但出于简明起见,在这一示例中, y_1 和 y_2 是共线对齐的,并且 z_1 和 z_2 是平行的。同样,可使用两个或更多传感器,但出于简明起见,这里只描绘了一个传感器。

[0075] 所投影的结构化光的光线是从照明器平面中的不同 x_2 、 y_2 位置发射的,诸如从照明帧 500 上的点 P_2 所发射的示例光线 502。光线 502 撞击对象 520,例如,点 P_0 处的人,并且在许多方向上反射。光线 512 是从 P_0 行进至所捕捉的帧 510 上的点 P_1 的示例反射光线。 P_1 由传感器中的像素表示,使得它的 x_1 、 y_1 位置是已知的。根据几何原理, P_2 位于包括 P_1 、 F_1 和 F_2 的平面上。这一平面与照明帧 500 相交的一部分是外极线 (epi-polar) 505。通过标识结构化光的哪一部分由 P_2 投影, P_2 沿外极线 505 的位置可被标识。 P_2 是 P_1 的对应点。对象的深度约接近,外极线的长度就越长。

[0076] 随后, P_0 沿 z_1 轴的深度可通过三角测量来确定。这是深度图中分配给像素 P_1 的深度值。对于照明帧 500 中的某些点,在所捕捉的帧 510 中可能没有对应像素,诸如由于遮挡或由于传感器的有限视野。对于所捕捉的帧 510 中对应点在照明帧 500 中被标识的每一像素,深度值可被获取。所捕捉的帧 510 的深度值集提供所捕捉的帧 510 的深度图。对于附加传感器及其相应的所捕捉的帧可实现相似的过程。此外,当视频数据的连续帧被获取后,可对每一帧实现该过程。

[0077] 图 5B 描绘了在立体光系统中的两个所捕捉的帧。立体处理类似于图 5A 中描述的处理,因为两个帧中的对应点被标识。然而,在这一情况下,两个所捕捉的帧中的对应像素

被标识,并且照明器被分开地提供。照明器 550 提供投影在照明器视野中的对象 520 上的光。这一光可例如由对象反射并由两个传感器感测。第一传感器获取像素数据的帧 530,而第二传感器获取像素数据的帧 540。示例光线 532 从对象上的点 P_0 延伸到帧 530 中的像素 P_2 ,穿过相关联的传感器的焦点 F_2 。类似地,示例光线 542 从对象上的点 P_0 延伸到帧 540 中的像素 P_1 ,穿过相关联的传感器的焦点 F_1 。从帧 540 的角度来看,立体匹配可涉及标识外极线 545 上对应于 P_1 的点 P_2 。类似地,从帧 530 的角度来看,立体匹配可涉及标识外极线 548 上对应于 P_2 的点 P_1 。由此,立体匹配可分开地执行,每次针对一对帧中的一个帧。在某些情况下,可执行从第一帧到第二帧在一个方向上的立体匹配,而不执行从第二帧到第一帧在其他方向上的立体匹配。

[0078] P_0 沿 z_1 轴的深度可通过三角测量来确定。这是深度图中分配给像素 P_1 的深度值。对于帧 540 中的某些点,在帧 530 中可能没有对应像素,诸如由于遮挡或由于传感器的有限视野。对于帧 540 中对应像素在帧 530 中被标识的每一像素,深度值可被获取。帧 540 的深度值集提供帧 540 的深度图。

[0079] 类似地, P_2 沿 z_2 轴的深度可通过三角测量来确定。这是深度图中分配给像素 P_2 的深度值。对于帧 530 中的某些点,在帧 540 中可能没有对应像素,诸如由于遮挡或由于传感器的有限视野。对于帧 530 中对应像素在帧 540 中被标识的每一像素,深度值可被获取。帧 530 的深度值集提供帧 530 的深度图。

[0080] 对于附加传感器及其相应的所捕捉的帧可实现相似的过程。此外,当视频数据的连续帧被获取后,可对每一帧实现该过程。

[0081] 图 6A 描绘了带有在照明器的公共侧的两个传感器的成像组件 600。照明器 26 是用结构化光图案来照明视野内的人类目标或其他对象的投影仪。光源可以是例如具有 700nm-3,000nm 波长的红外激光,包括具有 0.75 μm -1.4 μm 波长的近红外光、具有 3 μm -8 μm 波长的中等波长的红外光、以及具有 8 μm -15 μm 波长的长波长的红外光,该光源是最接近于由人类所发射的红外辐射的热成像区域。照明器可包括接收激光以及输出多个衍射的光束的衍射光学元件 (DOE)。一般而言,DOE 用于从单个准直光束提供多个较小的光束,如数千个较小的光束。每一个较小的光束都具有单个准直光束的功率的一小部分,并且较小的衍射光束可具有名义上相等的强度。

[0082] 较小光束在所希望的预定图案中定义照明器的视野。DOE 是光束复制器,因此所有输出光束将具有与输入光束相同的几何形状。例如,在运动跟踪系统中,可能希望以允许跟踪站在或坐在房间里的人类目标的方式来照亮房间。为跟踪整个人类目标,视野应该在足够宽的角度,在高度和宽度方面延伸,以便照亮人类的整个高度和宽度以及当与运动跟踪系统的应用程序进行交互时人类可以来回移动的区域。可以基于诸如人类的预期高度和宽度之类的因素来设置适当的视野,这些因素包括当手臂被抬高到头上或被抬到侧面时手臂的跨度,当与应用程序进行交互时人类可以移动的区域的大小,人类与照相机的预期的距离,以及照相机的焦距。

[0083] 如先前讨论的,还可提供 RGB 相机 28。还可在图 6B 和 6C 中提供 RGB 相机,但出于简明起见而未描绘。

[0084] 在这一示例中,传感器 24 和 29 位于照明器 26 的公共侧。传感器 24 位于与照明器 26 的基线距离 BL1 处,并且传感器 29 位于与照明器 26 的基线距离 BL2 处。由于传感

器 29 的较短基线,针对较短程成像来优化传感器 29,而由于传感器 24 的较长基线,针对较长程成像来优化传感器 24。此外,对于通常包括大小有限的外壳的成像系统 600 的固定大小,通过将两个传感器放置在照明器的一侧,可为距离照明器最远的传感器实现较长基线。另一方面,较短的基线改进了较短程成像,因为假设一给定焦距,传感器可聚焦于较近的对象,由此允许对较短距离的更准确的深度测量。较短基线导致较小差别以及最小遮挡。

[0085] 较长基线改进了较长程成像的准确性,因为对应点的各光射线之间存在较大角度,这意味着图像像素可检测距离中的更小差异。例如,在图 5A 中,可以看到,如果帧 500 与 510 更远离,则光线 502 与 512 之间的角将更大。并且,在图 5B 中,可以看到,如果帧 530 与 540 更远离,则光线 532 与 542 之间的角将更大。当传感器更远离使得光射线之间的角更大时,用于确定深度的三角测量的过程更准确。

[0086] 除了在成像组件 600 的外壳的约束下根据是较短程成像还是较长程成像被优化来为传感器设置最优基线以外,传感器的其他特征可被设置成优化较短程成像或较长程成像。例如,相机的空间分辨率可被优化。诸如电荷耦合器件 (CCD) 之类的传感器的空间分辨率是像素数量及其相对于所投影的图像的大小的函数,并且是对该传感器可检测出的细节有多精细的测量。与针对较长程成像来优化的传感器相比,对于针对较短程成像来优化的传感器,较低的空间分辨率是可接受的。较低的空间分辨率可通过使用帧中的相对较少的像素和 / 或相对较大的像素来实现,因为由于视野中所检测的对象的较短深度,像素大小相对于投影图像是相对较大的。这可导致成本节省以及降低的能耗。另一方面,与针对较短程成像来优化的传感器相比,对于针对较长程成像来优化的传感器,应使用较高的空间分辨率。较高的空间分辨率可通过使用帧中的相对较多的像素和 / 或相对较小的像素来实现,因为由于视野中所检测的对象的较长深度,像素大小相对于投影图像是相对较小的。较高的分辨率在深度测量中产生较高的准确性。

[0087] 可被设置成优化较短程或较长程成像的传感器的另一特征是敏感性。敏感性指的是传感器对入射光作出反应的程度。敏感性的一个测量是量子效率,该量子效率是入射在传感器的光反应表面(诸如像素)上的光子的百分比,这将产生电子-空穴对。对于针对短程成像来优化的传感器,较低的敏感性是可接受的,因为由于将光子反射回传感器的对象的较近距离,相对更多光子将入射在每一像素上。较低的敏感性可通过例如导致成本节省的较低质量的传感器来实现。另一方面,与针对较短程成像来优化的传感器相比,对于针对较长程成像来优化的传感器,应使用较高的敏感性。较高的敏感性可通过较高质量的传感器来实现,以允许其中由于把光子反射回传感器的对象的更远距离而导致相对较少的光子将入射在每一像素上的检测。

[0088] 可被设置成优化较短程或较长程成像的传感器的另一特征是曝光时间。曝光时间是在获取图像数据的帧的过程期间光被允许落在传感器的像素上的时间量,例如,相机快门打开的时间。在曝光时间期间,传感器的像素累积或集成电荷。曝光时间与敏感性有关,因为较长曝光时间可补偿较低的敏感性。然而,为准确地捕捉较短范围处的运动序列,较短的曝光是合乎需要的,因为成像的对象的给定移动在对象较接近时转换成较大的像素偏移。较短曝光时间可被用于针对较短程成像来优化的传感器,而较长曝光时间可被用于针对较长程成像来优化的传感器。通过使用适当的曝光时间,可避免较近对象的过度曝光 / 图像饱和以及较远对象的曝光不足。

[0089] 图 6B 描绘了带有在照明器的一侧的两个传感器、以及在该照明器的相对侧的一个传感器的成像组件 610。以此方式添加第三传感器可导致对象的成像有较少的遮挡、以及由于所获取的附加深度测量而导致的更准确的成像。诸如传感器 612 之类的一个传感器可被定位在接近于照明器,而其他两个传感器位于照明器的相对侧。在这一示例中,传感器 24 位于距照明器 26 的基线距离 BL1 处,传感器 29 位于距照明器 26 的基线距离 BL2 处,以及第三传感器 612 位于距照明器 26 的基线距离 BL3 处。

[0090] 图 6C 描绘了带有在照明器的公共侧的三个传感器的成像组件 620。以此方式添加第三传感器可导致由于所获取的附加深度测量而导致的更准确的成像。此外,每一传感器可针对不同深度范围来优化。例如,位于距照明器的较大基线距离 BL3 处的传感器 24 可针对较长程成像来优化。位于距照明器的中等基线距离 BL2 的传感器 29 可针对中程成像来优化。并且,位于距照明器的较短基线距离 BL1 的传感器 612 可针对较短程成像来优化。类似地,可为传感器 24、传感器 29 和传感器 612 将空间分辨率、敏感性和 / 或曝光时间优化至较长程级别、中程级别,以及较短程级别。

[0091] 图 6D 描绘了带有在照明器的相对侧的两个传感器的成像组件 630,其示出这两个传感器如何感测对象的不同部分。传感器 S1 24 位于距照明器 26 的基线距离 BL1 处,并且针对较短程成像来优化。传感器 S2 29 位于距照明器 26 的基线距离 $BL2 > BL1$ 处,并且针对较长程成像来优化。还描绘了 RGB 相机 28。对象 660 出现在视野中。注意,因为从正视图示出成像组件 630 并且从俯视图示出对象 660,附图的透视被修改为简化。光线 640 和 642 是由照明器 26 投影的光的示例光线。光线 632、634 和 636 是由传感器 S1 24 感测的反射光的示例光线,而光线 650 和 652 是由传感器 S2 29 感测的反射光的示例光线。

[0092] 对象包括由感测器 S1 24 和 S2 29 感测的五个表面。然而,由于遮挡,不是所有表面都被两个传感器感测到。例如,表面 661 仅由传感器 S1 24 感测,并且从传感器 S2 29 的角度来看是被遮挡的。表面 662 同样仅由传感器 S1 24 感测,并且从传感器 S2 29 的角度来看是被遮挡的。表面 663 由传感器 S1 和 S2 两者感测。表面 664 仅由传感器 S2 感测,并且从传感器 S1 的角度来看是被遮挡的。表面 665 仅由传感器 S2 感测,并且从传感器 S1 的角度来看是被遮挡的。表面 666 由传感器 S1 和 S2 两者感测。这指示第二传感器或其他附加传感器的添加如何可用于对原本会被遮挡的对象的部分进行成像。此外,对于使遮挡最小化,将传感器放置得实际尽可能远离照明器常常是期望的。

[0093] 图 7A 描绘了用于获取视野的深度图的过程。步骤 700 包括用结构化光的图案来照明视野。可使用任何类型的结构化光,包括编码的结构化光。步骤 702 和 704 可至少部分地并发执行。步骤 702 包括检测第一传感器处所反射的红外光,以获取像素数据的第一帧。这一像素数据可指示例如在曝光时间期间由每一像素所累积的电荷量,作为从视野中入射在像素上的光量的指示。类似地,步骤 704 包括检测第二传感器处所反射的红外光,以获取像素数据的第二帧。步骤 706 包括处理来自两个帧的像素数据以导出合并的深度图。这可涉及不同的技术,诸如参考图 7B-7E 进一步讨论的。步骤 708 包括基于合并的深度图来提供对应用程序的控制输入。这一控制输入可用于各种目的,诸如更新化身在显示器上的位置、在用户界面 (UI) 中选择菜单项、或许多其他可能的动作。

[0094] 图 7B 描绘了图 7A 的步骤 706 的进一步细节,其中两个结构化光深度图被合并。在该方法中,第一和第二结构化光深度图分别从第一和第二帧中被获取,并且两个深度图被

合并。该过程可延伸至合并两个或更多中的任意数量个深度图。具体地,在步骤 720,对于像素数据的第一帧中的每一像素(在图 7A 的步骤 702 中获取),通过匹配结构化光的图案来尝试确定照明帧中的对应点。在某些情况下,由于遮挡或其他因素,可能无法为第一帧中的一个或多个像素成功地确定照明帧中的对应点。在步骤 722,提供了第一结构化光深度图。这一深度图可标识第一帧中的每一像素以及对应的深度值。类似地,在步骤 724,对于像素数据的第二帧中的每一像素(在图 7A 的步骤 704 中获取),尝试确定照明帧中的对应点。在某些情况下,由于遮挡或其他因素,可能无法为第二帧中的一个或多个像素成功地确定照明帧中的对应点。在步骤 726,提供了第二结构化光深度图。这一深度图可标识第二帧中的每一像素以及对应的深度值。步骤 720 和 722 可至少部分地与步骤 724 和 726 并发执行。在步骤 728,结构化光深度图被合并,以导出图 7A 的步骤 706 的合并的深度图。

[0095] 合并可基于不同方法,包括涉及不加权平均、加权平均、准确性测量和/或置信度测量的方法。在一个方法中,对于每一像素,在两个或更多深度图之间对深度值求平均。第一帧中第 i 像素的深度值 d_1 与第二帧中第 i 像素的深度值 d_2 的示例不加权平均为 $(d_1+d_2)/2$ 。第一帧中第 i 像素的权重为 w_1 的深度值 d_1 与第二帧中第 i 像素的权重为 w_2 的深度值 d_2 的示例加权平均为 $(w_1*d_1+w_2*d_2)/(w_1+w_2)$ 。合并深度值的一个方法基于传感器与照明器之间的基线距离来将权重分配给帧的深度值,使得在基线距离较大时被分配指示较高置信度的较高权重,而在基线距离较小时被分配指示较低置信度的较低权重。这么做是由于较大的基线距离产生较准确的深度值。例如,在图 6D 中,我们可将 $w_1 = BL_1/(BL_1+BL_2)$ 的权重分配给来自传感器 S1 的深度值,将 $w_2 = BL_2/(BL_1+BL_2)$ 的权重分配给来自传感器 S2 的深度值。为了说明,如果我们假设 $BL = 1$ 且 $BL = 2$ 距离单位,则 $w_1 = 1/3$ 且 $w_2 = 2/3$ 。可在每一像素或每一深度值的基础上应用权重。

[0096] 以上示例可以使用基于图 6D 中的距离 BL_1+BL_2 从来自传感器 S1 的图像到来自传感器 S2 的图像的立体匹配所获取的深度值来扩充。在这一情况下,我们可将 $w_1 = BL_1/(BL_1+BL_2+BL_1+BL_2)$ 分配给来自传感器 S1 的深度值,将 $w_2 = BL_2/(BL_1+BL_2+BL_1+BL_2)$ 的权重分配给来自传感器 S2 的深度值,并且将 $w_3 = (BL_1+BL_2)/(BL_1+BL_2+BL_1+BL_2)$ 的权重分配给从自 S1 到 S2 的立体匹配所获取的深度值。为了说明,如果我们假设 $BL = 1$ 且 $BL = 2$ 距离单位,则 $w_1 = 1/6$, $w_2 = 2/6$ 且 $w_3 = 3/6$ 。在进一步扩充中,深度值是从来自图 6D 中的传感器 S2 的图像到来自传感器 S1 的图像的立体匹配中获取的。在这一情况下,我们可将 $w_1 = BL_1/(BL_1+BL_2+BL_1+BL_2+BL_1+BL_2)$ 分配给来自传感器 S1 的深度值,将 $w_2 = BL_2/(BL_1+BL_2+BL_1+BL_2+BL_1+BL_2)$ 的权重分配给来自传感器 S2 的深度值,将 $w_3 = (BL_1+BL_2)/(BL_1+BL_2+BL_1+BL_2+BL_1+BL_2)$ 的权重分配给从自 S1 到 S2 的立体匹配所获取的深度值,并且将 $w_4 = (BL_1+BL_2)/(BL_1+BL_2+BL_1+BL_2+BL_1+BL_2)$ 的权重分配给从自 S2 到 S1 的立体匹配所获取的深度值。为了说明,如果我们假设 $BL = 1$ 且 $BL = 2$ 距离单位,则 $w_1 = 1/9$ 、 $w_2 = 2/9$ 、 $w_3 = 3/9$ 且 $w_4 = 3/9$ 。这只是一种可能性。

[0097] 权重还可基于置信度测量来提供,使得具有较高置信度测量的深度值被分配较高权重。在一个方法中,将初始置信度测量分配给每一像素,并且基于对象的深度将不会逐帧快速地变化的假设,为其中深度值在容忍度内是相同或接近相同的每一新帧增加置信度测量。例如,在每秒 30 帧的帧速率的情况下,所跟踪的人类将不会在各帧之间显著地移动。参见题为“Visual navigation and obstacle avoidance structured light system(可

视导航和障碍避免结构化光系统)”,8-13-91 授权的美国专利 5,040,116,其通过引用结合于此,以供进一步详述。在另一方法中,置信度测量是对深度值中的噪声的测量。例如,假设实际上不太可能出现相邻像素之间的深度值的大变化,这样的深度值的大变换可以是较大噪声量的指示,导致较低置信度测量。参见题为“System and method of using range image data with machine vision tools(使用带有机器视觉工具的范围图像数据的系统和方法)”,6-15-04 授权的美国专利 6,751,338,其通过引用结合于此,以供进一步详述。用于分配置信度测量的其他方法也是可能的。

[0098] 在一个方法中,定义了“主”相机坐标系统,并且我们将其他深度图像变换并重新采样至“主”坐标系统。一旦我们有了匹配图像,则我们可选择考虑一个或多个样本,其中我们可对其置信度进行加权。平均是一种方案,但不一定是最好的方案,因为它不能解决遮挡的情况,其中每一相机可能成功地观察空间中的不同位置。置信度测量可与深度图中的每一深度值相关联。另一方法是合并 3D 空间中的数据,其中不存在图像像素。在 3-D 中,可利用体积方法。

[0099] 为确定像素是否已经与图案正确地匹配并由此具有正确的深度数据,我们通常在图像与已知的所投影的图案之间执行相关或归一化相关。这是沿着传感器与照明器之间的外极线来完成的。成功的匹配由可与高置信度测量相关联的、相关的相对强的局部最大值来指示。另一方面,相关的相对弱的局部最大值可与低置信度测量相关联。

[0100] 权重还可基于准确性测量来提供,使得具有较高准确性测量的深度值被分配较高权重。例如,基于空间分辨率、各传感器与照明器之间以及各传感器之间的基线距离,我们可为每一深度样本分配准确性测量。用于确定准确性测量的各种技术是已知的。例如,参见加拿大不列颠哥伦比亚的里士满,Point Grey Research(Point Grey 研究所)于 2004 年 4 月 19 日,“Stereo Accuracy and Error Modeling(立体准确性和误差建模)”,<http://www.ptgrey.com/support/kb/data/kbStereoAccuracyShort.pdf>。随后,我们可基于这些准确性来计算加权平均。例如,对于所测量的 3D 点,我们分配权重 $W_i = \exp(-\text{准确性}_i)$,其中准确性_i为准确性测量,并且平均的 3D 点为 $P_{\text{平均}} = \sum(W_i * P_i) / \sum(W_i)$ 。随后,使用这些权重,在 3-D 中接近的点样本可使用加权平均来合并。

[0101] 为合并 3D 中的深度值数据,我们可使用 $(X, Y, Z) = \text{深度} * \text{光线} + \text{原点}$ 来将所有深度图像投影到 3D 空间,其中,光线为从像素至传感器的焦点的 3D 向量,原点是传感器的焦点在 3D 空间中的位置。在 3D 空间中,我们为每一深度数据点的计算法线方向。此外,对于每一数据点,我们从其他源查找一附近的数据点。在以下情况中:其他数据点足够接近并且这些点的法向量之间的点积是正的,这意味着它们被相似地定向且不是对象的两侧,则我们将这些点合并成单个点。这一合并可例如通过计算这些点的 3D 位置的加权平均来执行。权重可由测量的置信度来定义,其中置信度测量基于相关分数。

[0102] 图 7C 描绘了图 7A 的步骤 706 的进一步细节,其中两个结构化光深度图和两个立体深度图被合并。在这一方法中,第一和第二结构化光深度图分别从第一和第二帧中获取。另外,一个或多个立体深度图被获取。第一和第二结构化深度图和一个或多个立体深度图被合并。该过程可延伸至合并两个或更多中的任何数量个深度图。步骤 740 和 742 可至少部分地与步骤 744 和 746、步骤 748 和 750、以及步骤 752 和 754 并发执行。在步骤 740,对于像素数据的第一帧中的每一像素,我们确定照明帧中的对应点,并且在步骤 742,我们提

供第一结构化光深度图。在步骤 744, 对于像素数据的第一帧中的每一像素, 我们确定像素数据的第二帧中的对应像素, 并且在步骤 746, 我们提供第一立体深度图。在步骤 748, 对于像素数据的第二帧中的每一像素, 我们确定照明帧中的对应点, 并且在步骤 750, 我们提供第二结构化光深度图。在步骤 752, 对于像素数据的第二帧中的每一像素, 我们确定像素数据的第一帧中的对应像素, 并且在步骤 754, 我们提供第二立体深度图。步骤 756 包括合并不同的深度图。

[0103] 合并可基于不同方法, 包括涉及不加权平均、加权平均、准确性测量和 / 或置信度测量的方法。

[0104] 在这一方法中, 两个立体深度图与两个结构化光深度图合并。在一个选项中, 该合并并在单个合并步骤中一起考虑所有的深度图。在另一可能的方法中, 该合并发生在多个步骤中。例如, 可合并结构化光深度图以获取第一个合并的深度图, 可合并立体深度图以获取第二个合并的深度图, 合并该第一个和第二个合并的深度图以获取最终合并的深度图。在其中合并发生在多个步骤中的另一选项中, 第一结构化光深度图与第一立体深度图合并以获取第一个合并的深度图, 第二结构化光深度图与第二立体深度图合并以获取第二个合并的深度图, 以及合并该第一个和第二个合并的深度图以获取最终合并的深度图。其他方法也是可能的。

[0105] 在另一方法中, 只有一个立体深度图与两个结构化光深度图合并。该合并可在一个或多个步骤中发生。在多步骤方法中, 第一结构化光深度图与立体深度图合并以获取第一个合并的深度图, 以及第二结构化光深度图与立体深度图合并以获取最终合并的深度图。或者, 合并两个结构化光深度图以获取第一个合并的深度图, 以及第一个合并的深度图与立体深度图合并以获取最终合并的深度图。其他方法是可能的。

[0106] 图 7D 描绘了图 7A 的步骤 706 的进一步细节, 其中, 使用立体匹配按需对各深度值进行细化。这一方法是自适应的, 因为立体匹配被用于响应于检测到指示期望细化的条件来细化一个或多个深度值。立体匹配可仅针对帧中的像素子集来执行。在一个方法中, 在像素不能匹配到结构化光图案时, 对该像素的深度值细化是期望的, 使得该深度值为空或默认值。像素可能由于遮挡、阴影、光照条件、表面纹理或其他原因而没有被匹配到结构光图案。在这一情况下, 立体匹配可提供深度值 (其中先前没有获取深度值), 或在某些情况下, 由于传感器由比传感器与照明器之间的基线间隔更大的基线所间隔开, 立体匹配可提供更准确的深度值。例如, 参见图 2、6B 和 6D。

[0107] 在另一方法中, 在像素的深度值超出阈值距离时, 这指示对象上的对应点距离传感器相对远, 则对该深度值的细化是期望的。在这一情况下, 立体匹配可在各传感器之间的基线大于各传感器中的每一个与照明器之间的基线的情况下提供更准确的深度值。

[0108] 细化可涉及提供深度值 (其中之前没有提供过), 或例如基于涉及不加权平均、加权平均、准确性测量和 / 或置信度测量的不同方法来合并深度值。此外, 在深度值被合并之前, 细化可针对每一传感器的各帧分开地执行。

[0109] 通过仅对于检测到指示期望细化的条件来针对像素执行立体匹配, 避免了不必要的处理。对于没有检测到指示期望细化的条件, 不对像素执行立体匹配。然而, 当检测到一条件, 该条件该指示细化对于帧的一个或多个像素是期望的时, 对整个帧执行立体匹配也是可能的。在一个方法中, 当对于帧中像素的最小数量部分的细化被指示时, 发起对整个帧

的立体匹配。

[0110] 在步骤 760, 对于像素数据的第一帧中的每一像素, 我们确定照明帧中的对应点, 并且在步骤 761, 我们提供对应的第一结构化光深度图。判定步骤 726 确定对深度值的细化是否被指示。可评估对于像素数据的第一帧中每一像素的准则, 并且在一个方法中, 准则可指示对与该像素相关联的深度值的细化是否是期望的。在一个方法中, 当相关联的深度值不可用或不可靠时, 细化是期望的。不可靠性可基于例如准确性测量和 / 或置信度测量。如果置信度测量超出阈值置信度测量, 则深度值可被认为是可靠的。或者, 如果准确性测量超出阈值准确性测量, 则深度值可被认为是可靠的。在另一方法中, 为使深度值被认为是可靠的, 置信度测量和准确性测量必须都超出相应的阈值水平。

[0111] 在另一方法中, 当相关联的深度值指示深度相对远时, 诸如当深度超出阈值深度时, 细化是期望的。如果期望细化, 则步骤 736 执行像素数据的第一帧的一个或多个像素到像素数据的第二帧的一个或多个像素的立体匹配。这得到像素数据的第一帧的一个或多个附加深度值。

[0112] 类似地, 对于像素数据的第二帧, 在步骤 764, 对于像素数据的第二帧中的每一像素, 我们确定照明帧中的对应点, 并且在步骤 765, 我们提供对应的第二结构化光深度图。判定步骤 766 确定对深度值的细化是否被指示。如果期望细化, 则步骤 767 执行像素数据的第二帧的一个或多个像素到像素数据的第一帧的一个或多个像素的立体匹配。这得到像素数据的第二帧的一个或多个附加深度值。

[0113] 步骤 768 合并像素数据的第一帧和第二帧的深度图, 其中该合并包括从步骤 763 和 / 或 767 的立体匹配中获取的深度值。合并可基于不同方法, 包括涉及不加权平均、加权平均、准确性测量和 / 或置信度测量的方法。

[0114] 注意, 对于细化被指示的给定像素, 该合并可将来自第一结构化光深度图的深度值、来自第二结构化光深度图的深度值、以及来自立体匹配的一个或多个深度值进行合并。与丢弃来自结构化光深度图的深度值并使用来自立体匹配的深度值来替换它的方法相比, 这一方法可提供更可靠的结果。

[0115] 图 7E 描绘了图 7A 的步骤 706 的另一方法的进一步细节, 其中, 使用立体匹配按需对合并的深度图的各深度值进行细化。在这一方法中, 对通过到结构化光图案的匹配所获取的深度图的合并发生在细化过程之前。步骤 760、761、764 和 765 与图 7D 中相同编号的步骤相同。步骤 770 合并结构化光深度图。该合并可基于不同方法, 包括涉及不加权平均、加权平均、准确性测量和 / 或置信度测量的方法。步骤 771 类似于图 7D 的步骤 762 和 766, 并涉及确定对深度值的细化是否被指示。

[0116] 可评估对于合并的深度图中每一像素的准则, 并且在一个方法中, 准则可指示对与像素相关联的深度值的细化是否是期望的。在一个方法中, 当相关联的深度值不可用或不可靠时, 细化是期望的。不可靠性可基于例如准确性测量和 / 或置信度测量。如果置信度测量超出阈值置信度测量, 则深度值可被认为是可靠的。或者, 如果准确性测量超出阈值准确性测量, 则深度值可被认为是可靠的。在另一方法中, 为使深度值被认为是可靠的, 置信度测量和准确性测量必须都超出相应的阈值水平。在另一方法中, 当相关联的深度值指示深度相对远时, 诸如当深度超出阈值深度时, 细化是期望的。如果期望细化, 则可执行步骤 772 和 / 或步骤 773。在某些情况下, 通过将一个帧中的像素匹配到另一帧中的像素来在

一个方向上执行立体匹配是足够的。在其他情况下,可在两个方向上执行立体匹配。步骤 772 执行像素数据的第一帧的一个或多个像素到像素数据的第二帧的一个或多个像素的立体匹配。这得到像素数据的第一帧的一个或多个附加深度值。步骤 773 执行像素数据的第二帧的一个或多个像素到像素数据的第一帧的一个或多个像素的立体匹配。这得到像素数据的第二帧的一个或多个附加深度值。

[0117] 步骤 774 为已执行立体匹配的一个或多个所选择的像素对步骤 770 的合并的深度图进行细化。该细化可涉及基于不同方法来合并深度值,该不同方法包括涉及不加权平均、加权平均、准确性测量和 / 或置信度测量的方法。

[0118] 如果在判定步骤 771 处不期望细化,则该过程在步骤 775 处结束。

[0119] 图 8 描绘了如在图 7A 的步骤 708 中阐述的用于使用控制输入来跟踪人类目标的示例方法。如上所述,深度相机系统可被用于跟踪用户的移动,诸如姿势。该移动可在应用程序处作为控制输入来处理。例如,这可包括更新显示器上的化身的位置(其中如在图 1 中所描绘的,该化身表示用户),选择用户界面(UI)中的一个菜单项,或者许多其他可能的动作。

[0120] 示例方法可以使用例如深度相机系统 20 和 / 或计算环境 12、100 或 420 来实现,如参考图 2-4 所讨论的。可以扫描一个或多个人类目标,以生成诸如骨架模型、网格人模型之类的模型,或人的任何其他合适的表示。在骨架模型中,每一个身体部位都可以被表征为定义骨架模型的关节和骨头的数学矢量。身体部位在关节处可以相对于彼此移动。

[0121] 然后,可以使用该模型来与由计算环境执行的应用程序进行交互。扫描以生成模型可以在应用程序被启动或运行时发生,或在如被扫描的人通过应用程序控制的其他时候发生。

[0122] 人可以被扫描以生成骨架模型,该骨架模型可以被跟踪,以使用户的物理移动或运动可以充当调整和 / 或控制应用程序的参数的实时用户界面。例如,所跟踪的个人的移动可用于在电子角色扮演游戏中移动化身或其他屏幕上人物,在电子赛车游戏中控制屏幕上车辆,在虚拟环境中控制对象的构成或组织,或执行应用的任何其他合适的控制。

[0123] 根据一个实施方式,在步骤 800 中,例如,从深度相机系统接收深度信息。深度相机系统可以捕捉或观察可包括一个或多个目标的视野。深度信息可包括具有多个观察到的像素的深度图像或图,其中,每一个观察到的像素都具有观察到的深度值,如所讨论的。

[0124] 可以将深度图像降采样到较低的处理分辨率,以便可以更加轻松地利用较少的计算开销来使用和处理它。另外,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 / 或移除的深度信息的部分;和 / 或可对所接收的深度信息执行任何其他合适的处理,使得该深度信息可用于生成诸如骨架模型等模型(见图 9)。

[0125] 步骤 802 判断深度图像是否包括人类目标。这可包括对深度图像中的每一个目标或对象进行泛色填充,将每一个目标或对象与图案进行比较以判断深度图像是否包括人类目标。例如,可以将深度图像的选定区域或点中的像素的各种深度值进行比较,以确定可以定义目标或对象的边缘,如上文所描述的。可以基于所确定的边缘,对 Z 层的可能的 Z 值进行泛色填充。例如,与确定的边缘相关联的像素和该边缘内的区域的像素可以彼此相关联,以定义可以与图案相比较的捕捉区域中的目标或对象,下面将更详细地描述。

[0126] 如果深度图像包括人类目标,则在判断步骤 804,执行步骤 806。如果判断步骤 804 为假,则在步骤 800 中接收附加的深度信息。

[0127] 每一个目标或对象与其比较的图案可包括一个或多个数据结构,这些数据结构具有共同地定义人的典型身体的变量集合。可以将与例如视野内的人类目标和非人类目标的像素相关联的信息与变量进行比较,以标识人类目标。在一个实施例中,可以基于身体部位,对该集合中的每一个变量赋予权重。例如,图案中的诸如头和 / 或肩之类的各种身体部位可以具有与其相关联的权重值,这些权重值可以大于诸如腿之类的其他身体部位的权重值。根据一个实施方式,当将目标与变量进行比较以判断目标是否可能是人类以及哪些目标可能是人类时,可以使用权重值。例如,变量和目标之间具有较大的权重值的匹配与具有较小权重值的匹配相比可产生目标是人类的更大似然性。

[0128] 步骤 806 包括扫描人类目标以寻找身体部位。可以扫描人类目标,以提供与人的一个或多个身体部位相关联的诸如长度、宽度等等之类的测量值,以提供该人的准确模型。在一示例实施例中,人类目标可以被隔离,并且可以创建人类目标的位掩模来扫描一个或多个身体部位。可以通过例如对人目标进行泛色填充来创建位掩模,以便人类目标可以与捕捉区域元素中的其他目标或对象分离。然后,可以对于一个或多个身体部位分析位掩模,以生成人类目标的模型,如骨架模型、网格人类模型等等。例如,根据一个实施例,可以使用由扫描的位掩模确定的测量值来定义骨架模型中的一个或多个关节。可以使用一个或多个关节来定义可以对应于人类的身体部位的一个或多个骨骼。

[0129] 例如,人类目标的位掩模的顶部可以与头的顶部的位置相关联。在确定头的顶部之后,可以向下扫描位掩模,以便随后确定颈的位置、肩的位置等等。可以将例如位于被扫描的位置的位掩模的宽度与和例如颈、肩等等相关联的典型宽度的阈值进行比较。在一替换实施例中,可以使用与扫描的并与位掩模中的身体部位相关联的先前位置的距离来确定颈部、肩膀等的位置。可以基于例如其他身体部位的位置来计算诸如腿、脚等等之类的某些身体部位。在确定一个身体部位的值后,创建包括该身体部位的测量值的数据结构。数据结构可包括来自自由深度相机系统在不同的时间点提供的多个深度图像的平均扫描结果。

[0130] 步骤 808 包括生成人类目标的模型。在一个实施方式中,可以使用由扫描的位掩模确定的测量值来定义骨架模型中的一个或多个关节。一个或多个关节被用来定义对应于人类的身体部位的一个或多个骨头。

[0131] 可以调整一个或多个关节,直到各关节在一个关节和人类的一个身体部位之间的典型距离范围之内,以生成更加准确的骨架模型。还可以基于例如与人类目标相关联的身高来进一步调整模型。

[0132] 在步骤 810 中,通过每秒钟多次更新人的位置,来跟踪模型。随着用户在物理空间中移动,使用来自深度相机系统的信息来调整骨架模型,以使该骨架模型表示人。具体而言,可以向骨架模型的一个或多个受力面施加一个或多个力,以将骨架模型调整为更加紧密地对应于物理空间中的人类目标的姿态的姿态。

[0133] 一般而言,可以使用用于跟踪人的移动的任何已知技术。

[0134] 图 9 描绘了如在图 8 的步骤 808 中阐述的人类目标的示例模型。模型 900 在图 1 的 $-z$ 方向上面向深度相机,以致于所示出的截面位于 $x-y$ 平面。该模型包括许多参考点,如头的顶部 902、头的底部或下巴 913、右肩 904、右肘 906、右手腕 908 和右手 910,如例如

由指尖区域所表示的。右侧和左侧是从面向相机的用户的角度定义的。该模型还包括左肩 914、左肘 916、左手腕 918 和左手 920。还描绘了腰部区域 922, 以及右臀 924、右膝 926、右脚 928、左臀 930、左膝 932 和左脚 934。肩线 912 是肩 904 和 914 之间的线, 通常是水平的。例如, 还描绘了在点 922 和 913 之间延伸的上躯干中心线 925。

[0135] 因此, 可以看到, 提供了具有多个优点的深度相机系统。一个优点是减小的遮挡。由于较宽的基线被使用, 一个传感器可以看到对于另一个传感器是遮挡的信息。与由单个传感器所产生的图相比, 两个深度图的融合产生具有更可观察的对象的 3D 图像。另一优点是减小的阴影效果。结构化光方法在对传感器可见但对光源不“可见”的位置固有地产生阴影效果。通过在这些区中应用立体匹配, 这一效果可被减小。另一优点是对外部光的稳健性。存在许多情形, 其中外部光照可能中断结构化光相机, 使得结构化光相机不能产生有效结果。在那些情况下, 立体数据被获取作为附加的测量, 因为外部光照可实际地帮助它测量距离。注意到外部光可来自正在看相同场景的相同相机。换言之, 操作所建议的相机中正在看相同场景的两个或更多相机变成可能。这是由于以下事实: 即使由一个相机所产生的光图案可中断其他相机正确地匹配图案, 立体匹配仍然有可能成功。另一优点是, 由于两个传感器具有较宽的基线的事实, 使用所建议的配置来实现远距离处的更大的准确性是可能的。结构化光和立体测量准确性两者都严重地取决于各传感器 / 投影仪之间的距离。

[0136] 前面的对本技术的详细描述只是为了说明和描述。它不是为了详尽的解释或将本技术限制在所公开的准确的形式。鉴于上述教导, 许多修改和变型都是可能的。所描述的实施例只是为了最好地说明本技术的原理以及其实际应用, 从而使精通本技术的其他人在各种实施例中最佳地利用本技术, 适合于特定用途的各种修改也是可以的。本技术的范围由所附的权利要求进行定义。

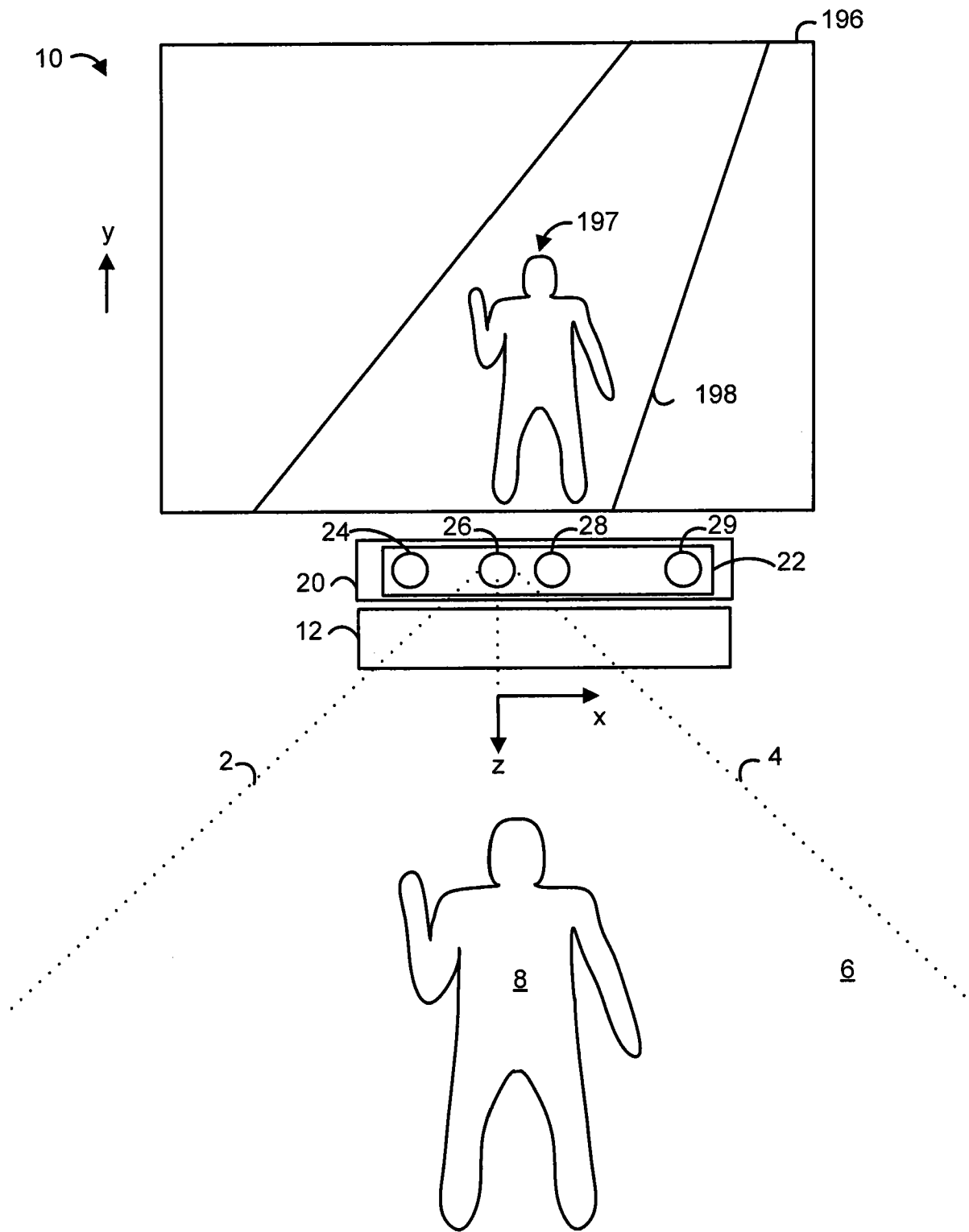


图 1

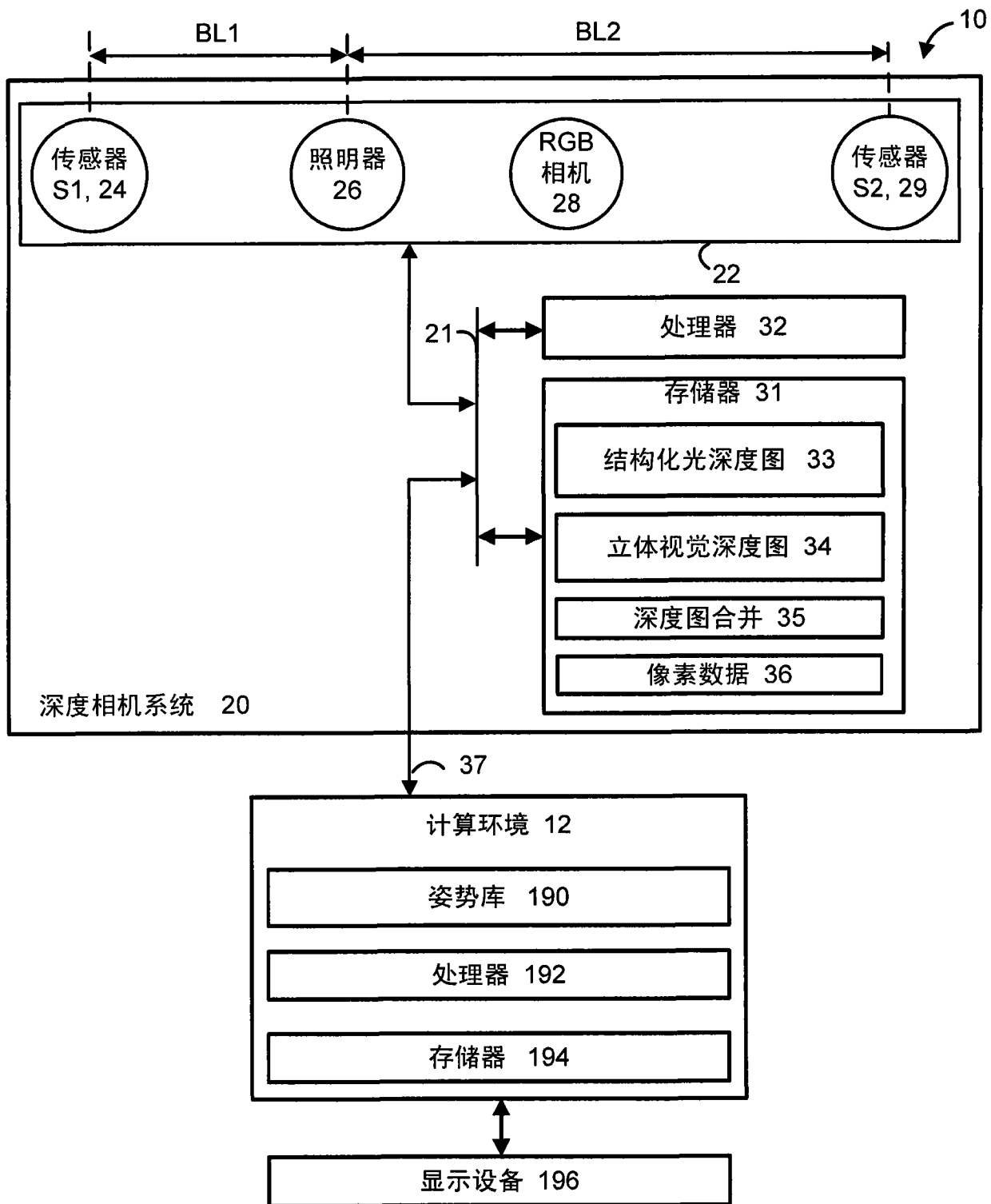


图 2

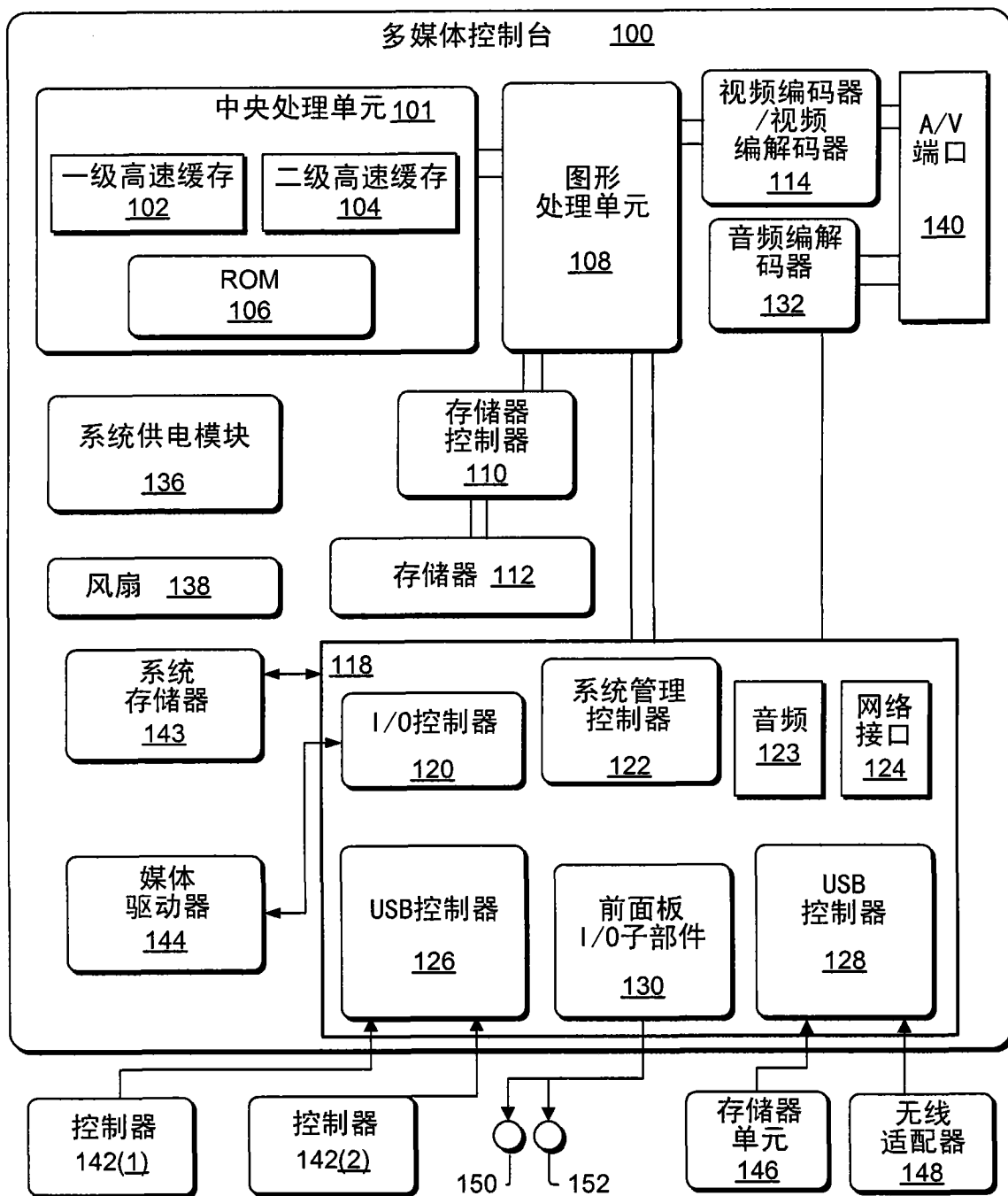


图 3

计算环境 220

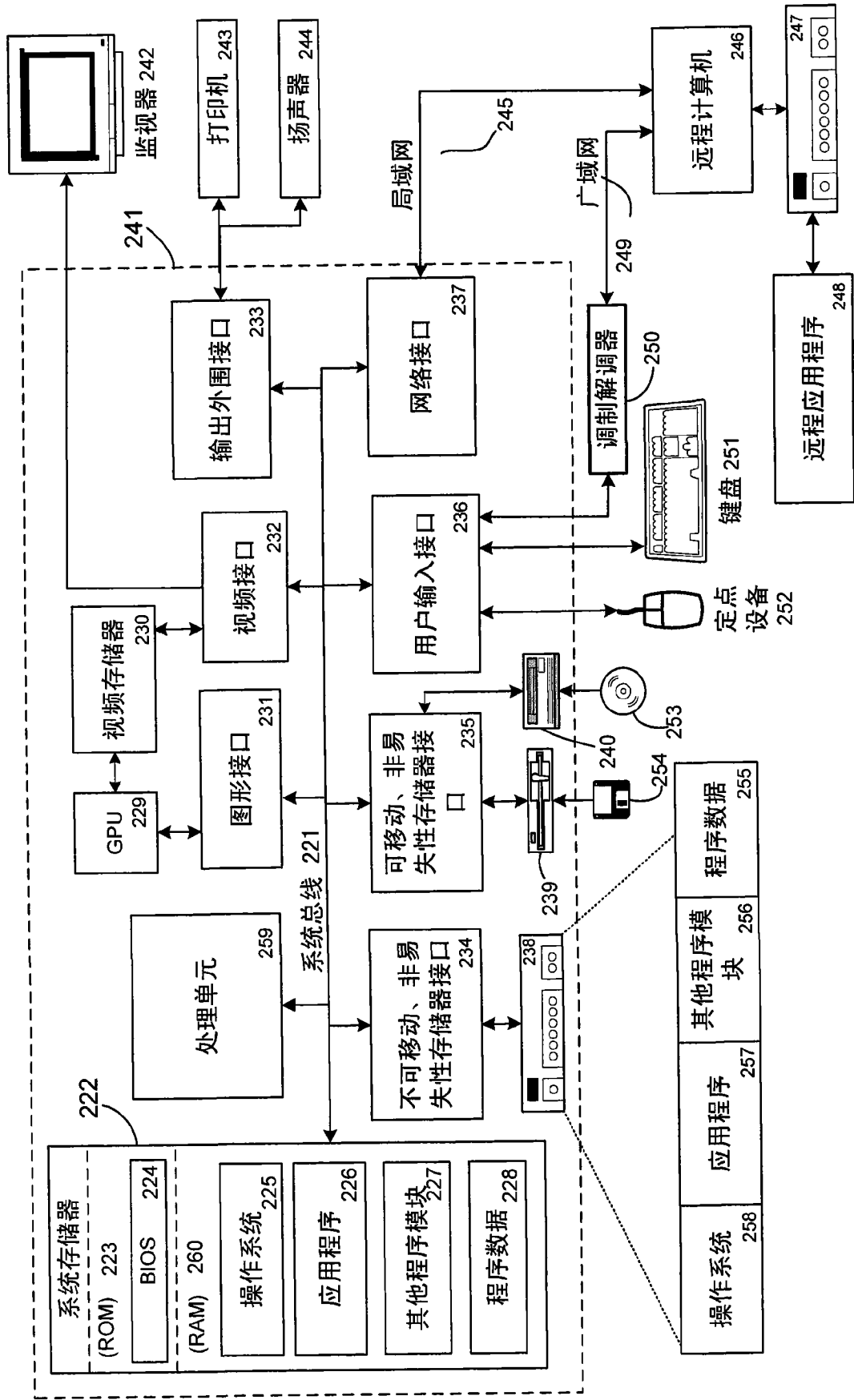


图 4

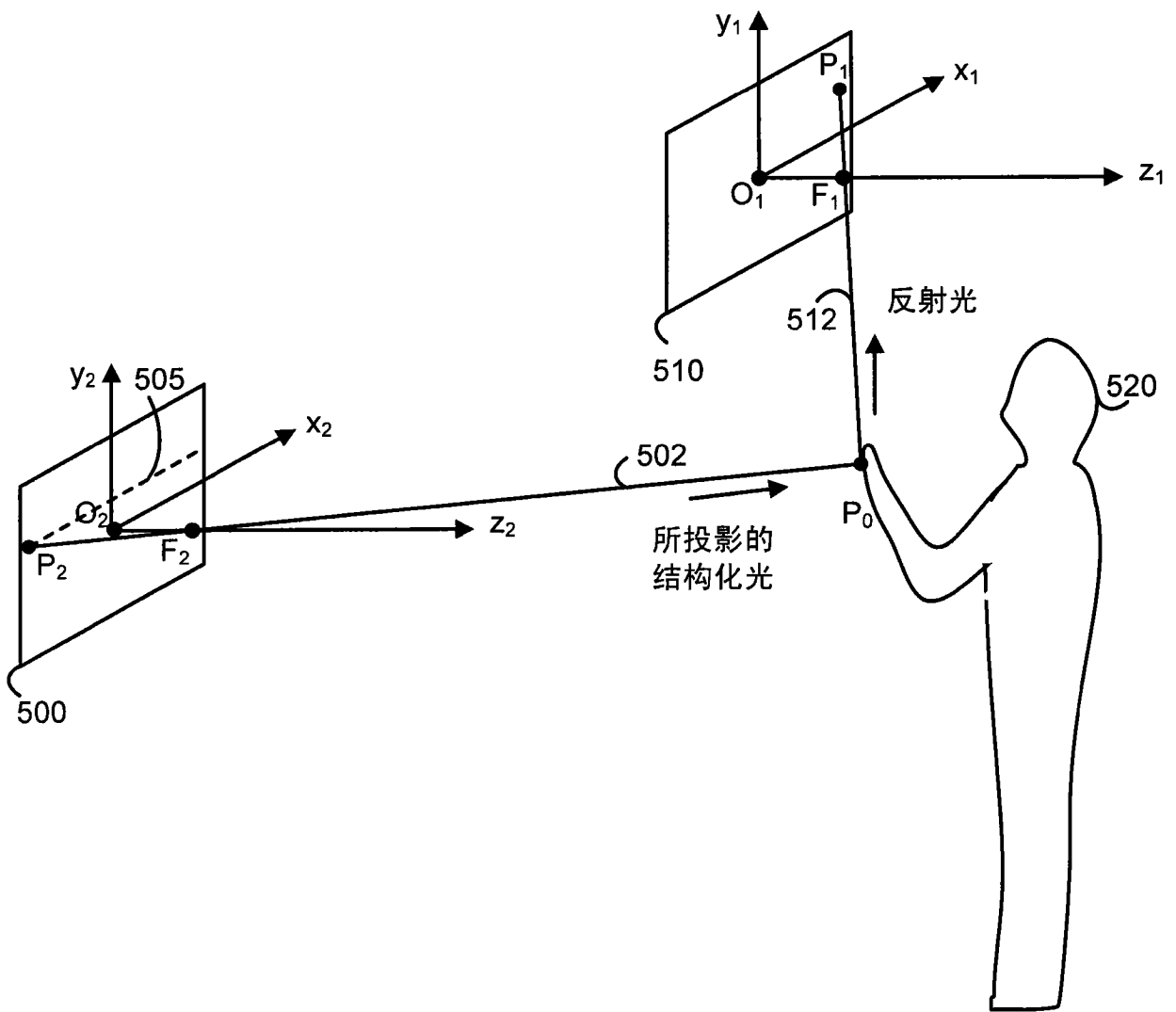


图 5A

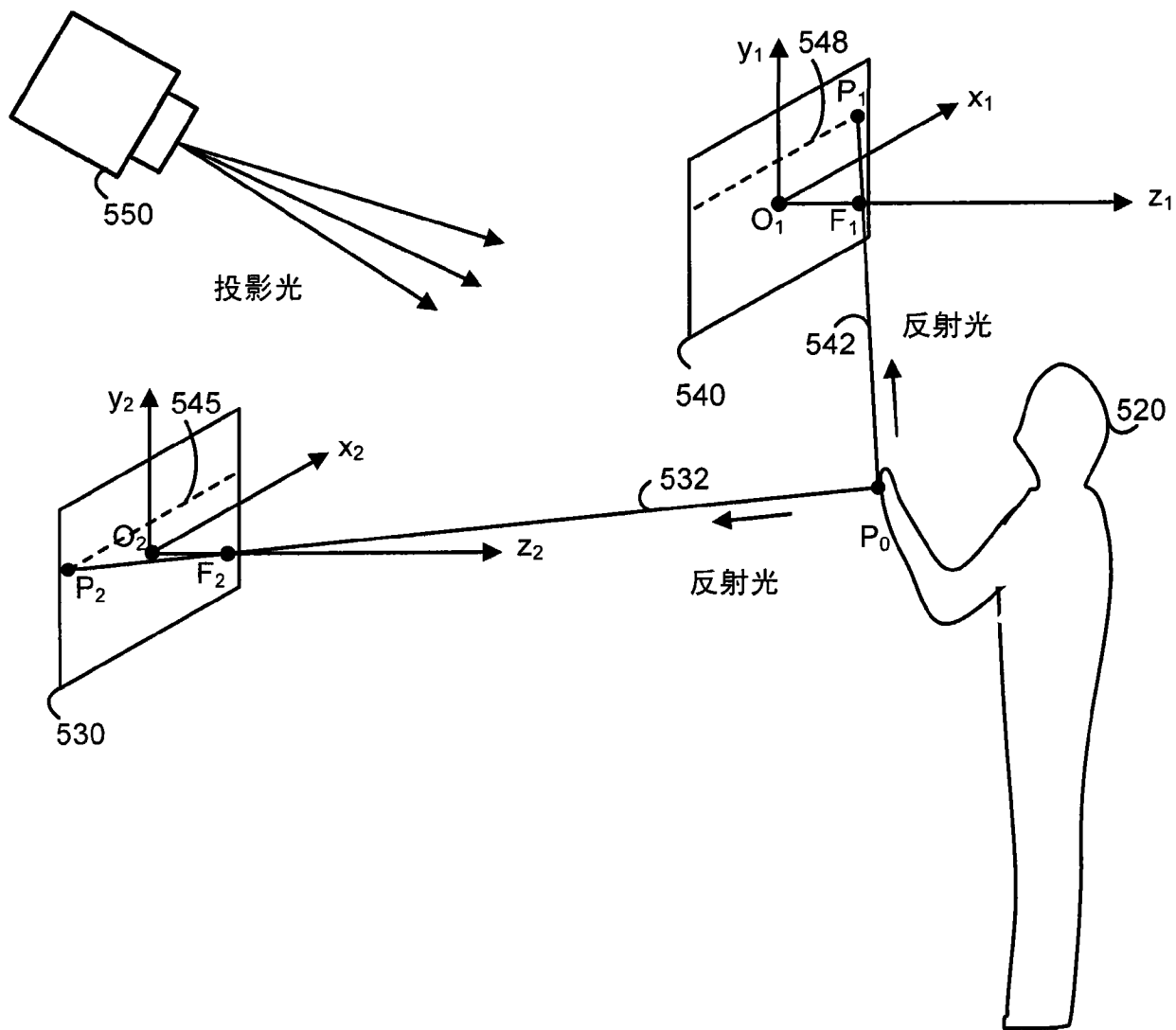


图 5B

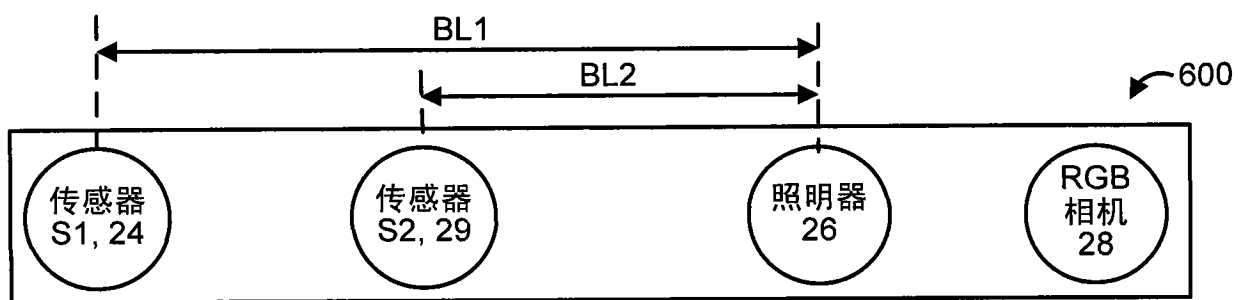


图 6A

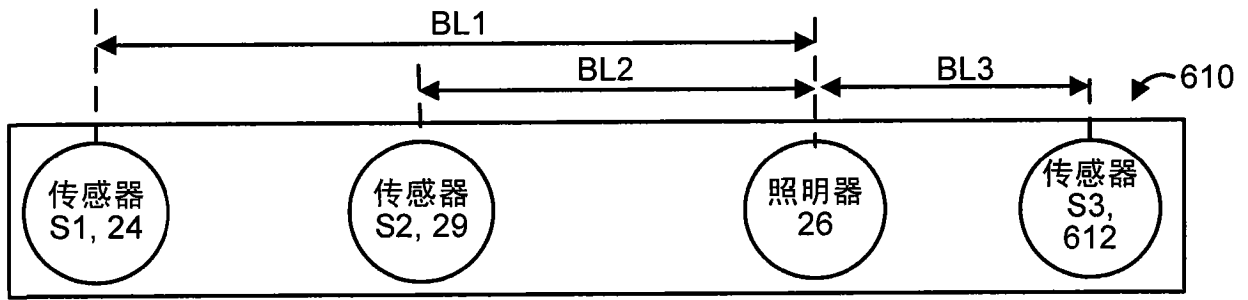


图 6B

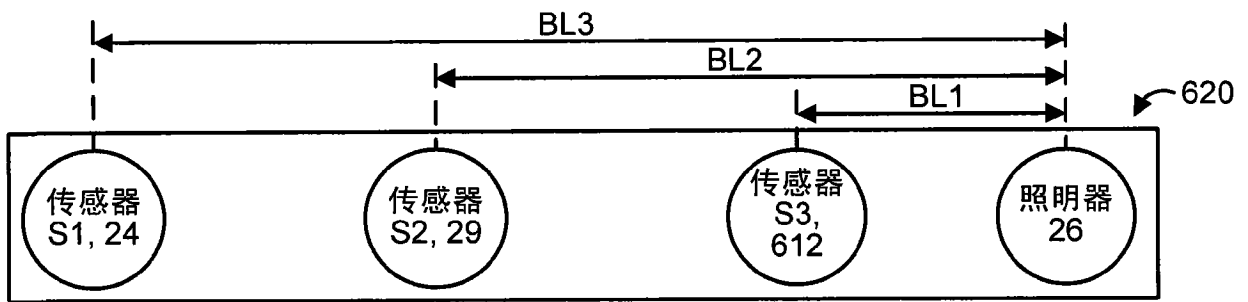


图 6C

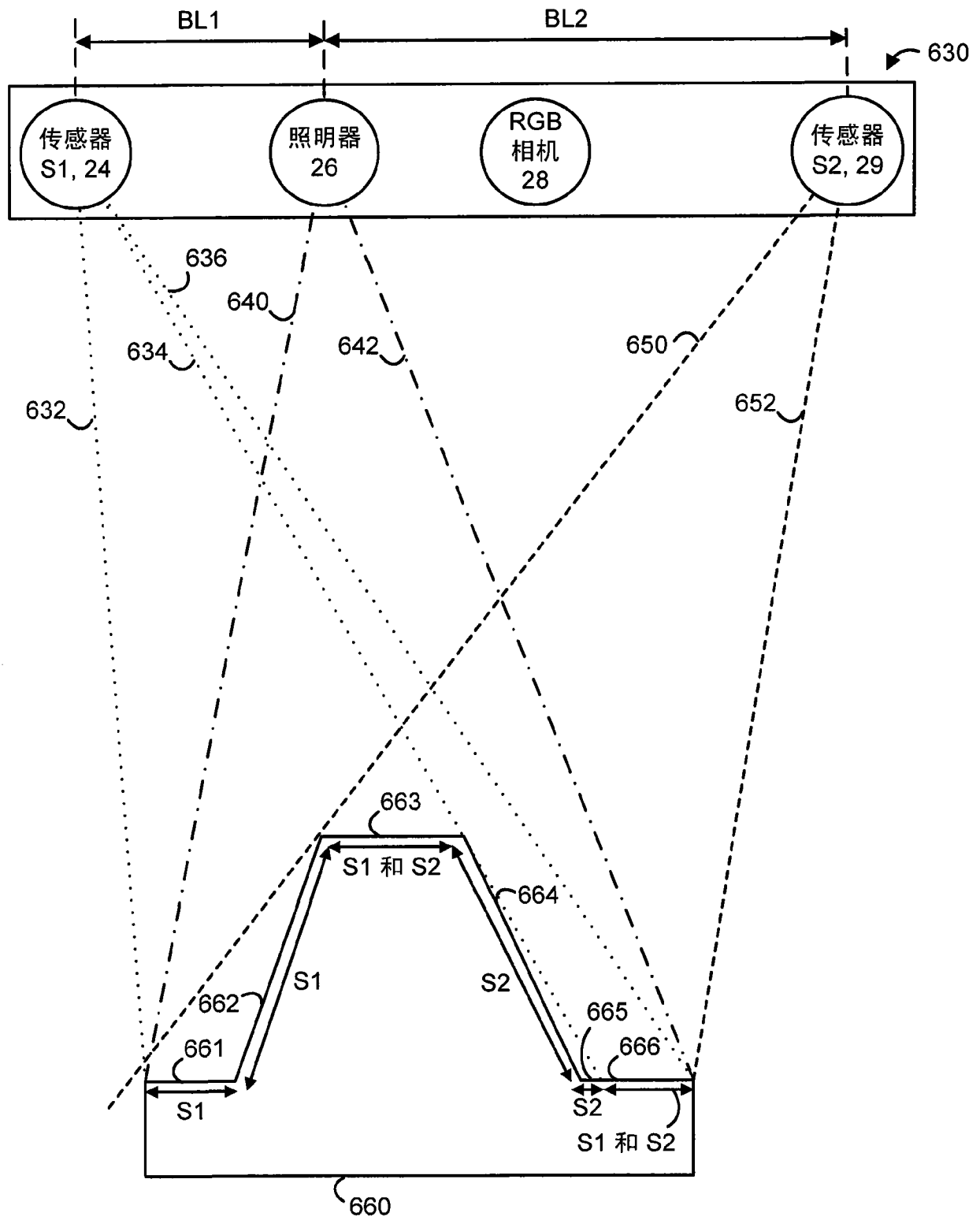


图 6D

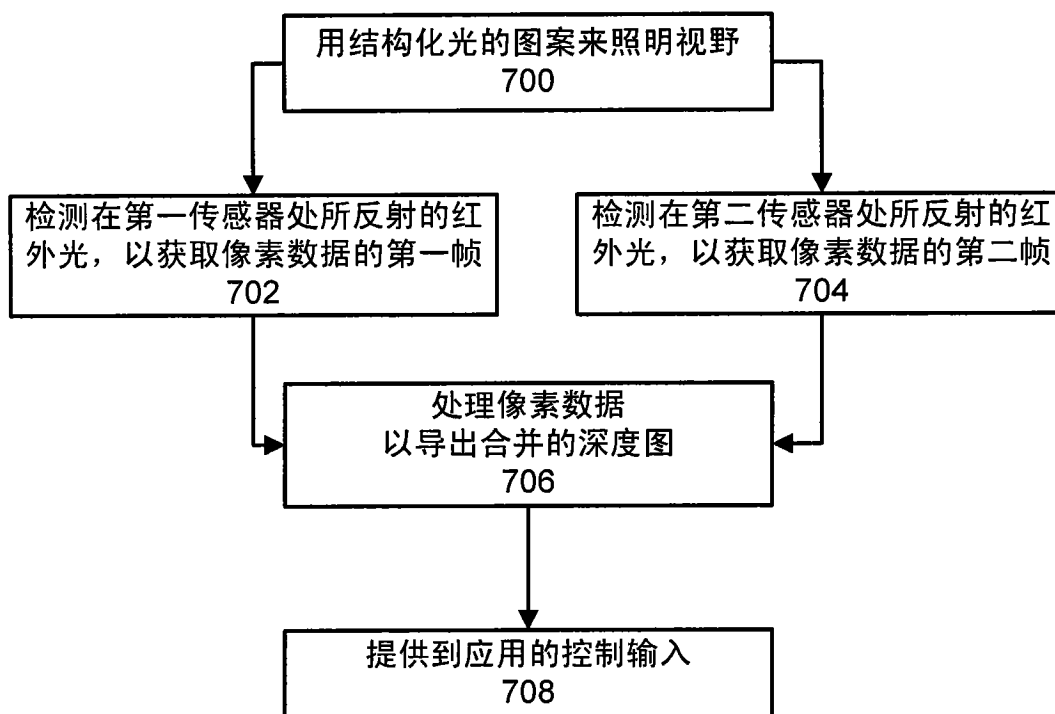


图 7A

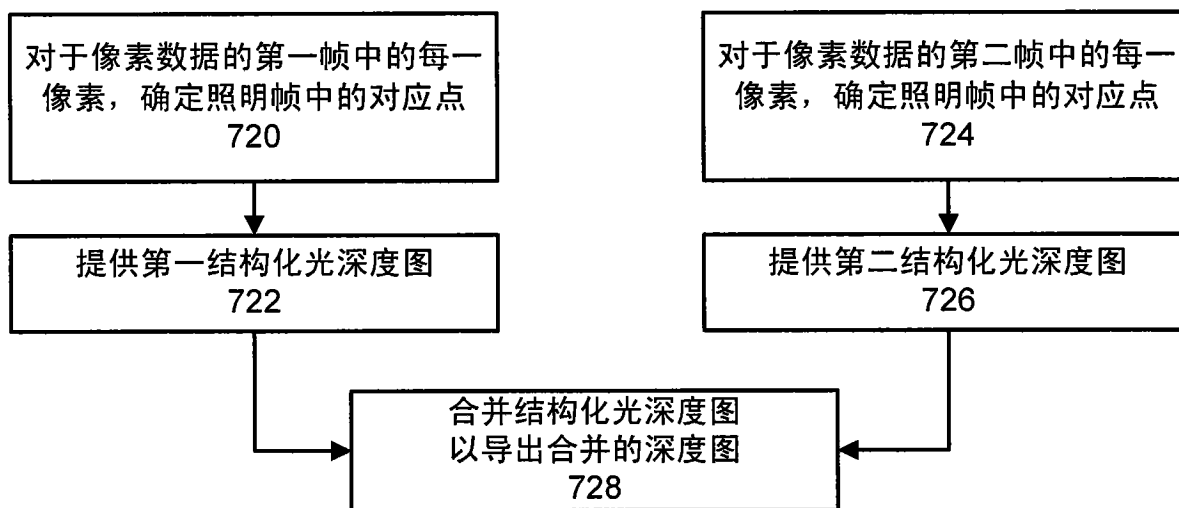


图 7B

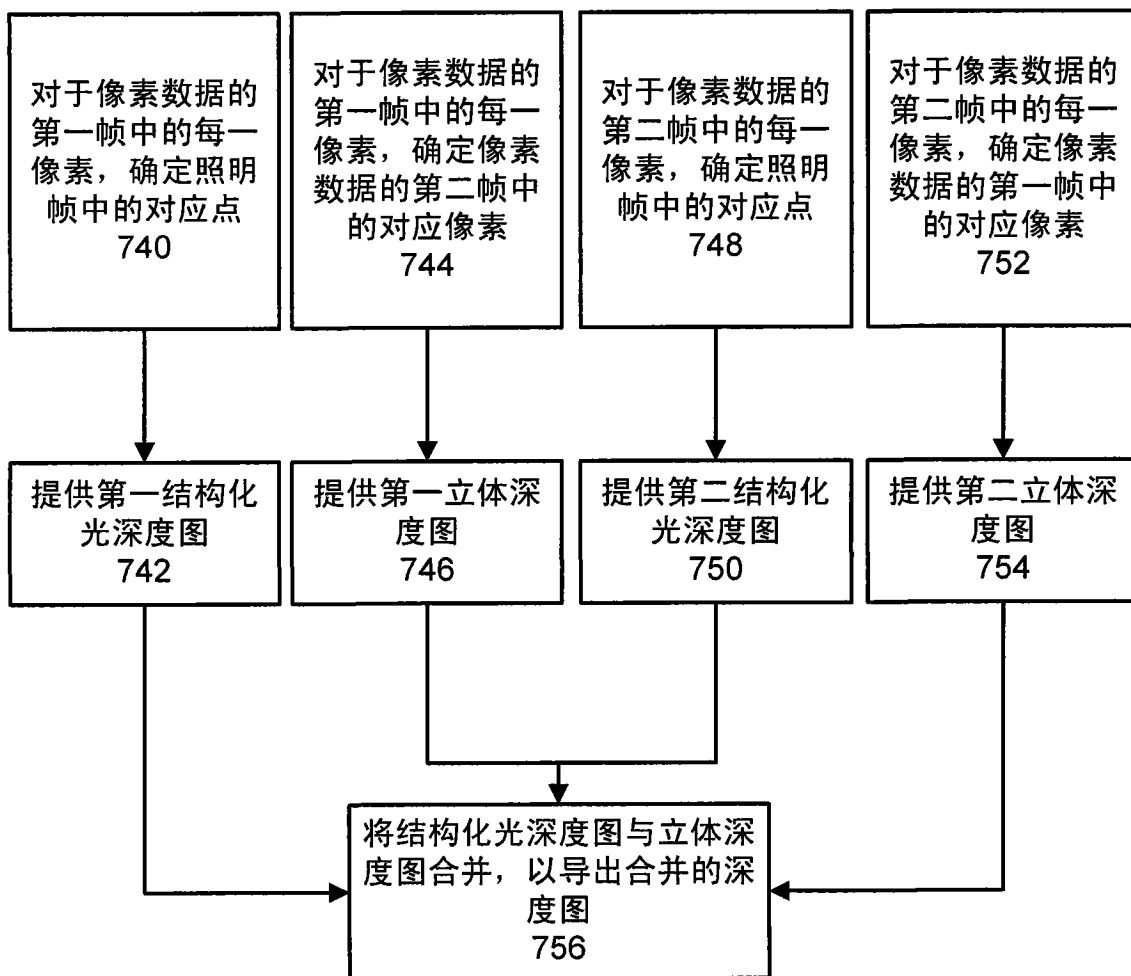


图 7C

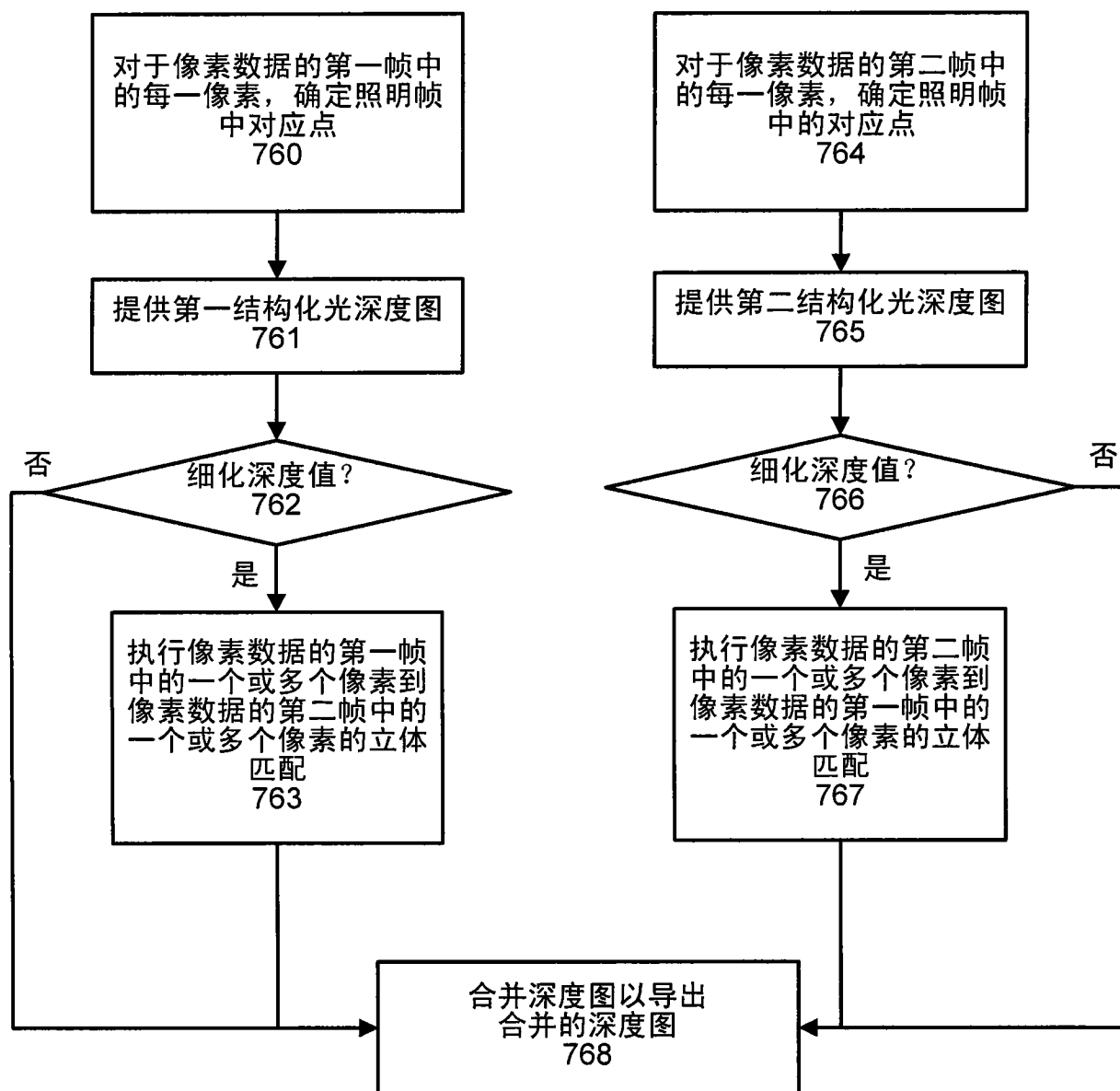


图 7D

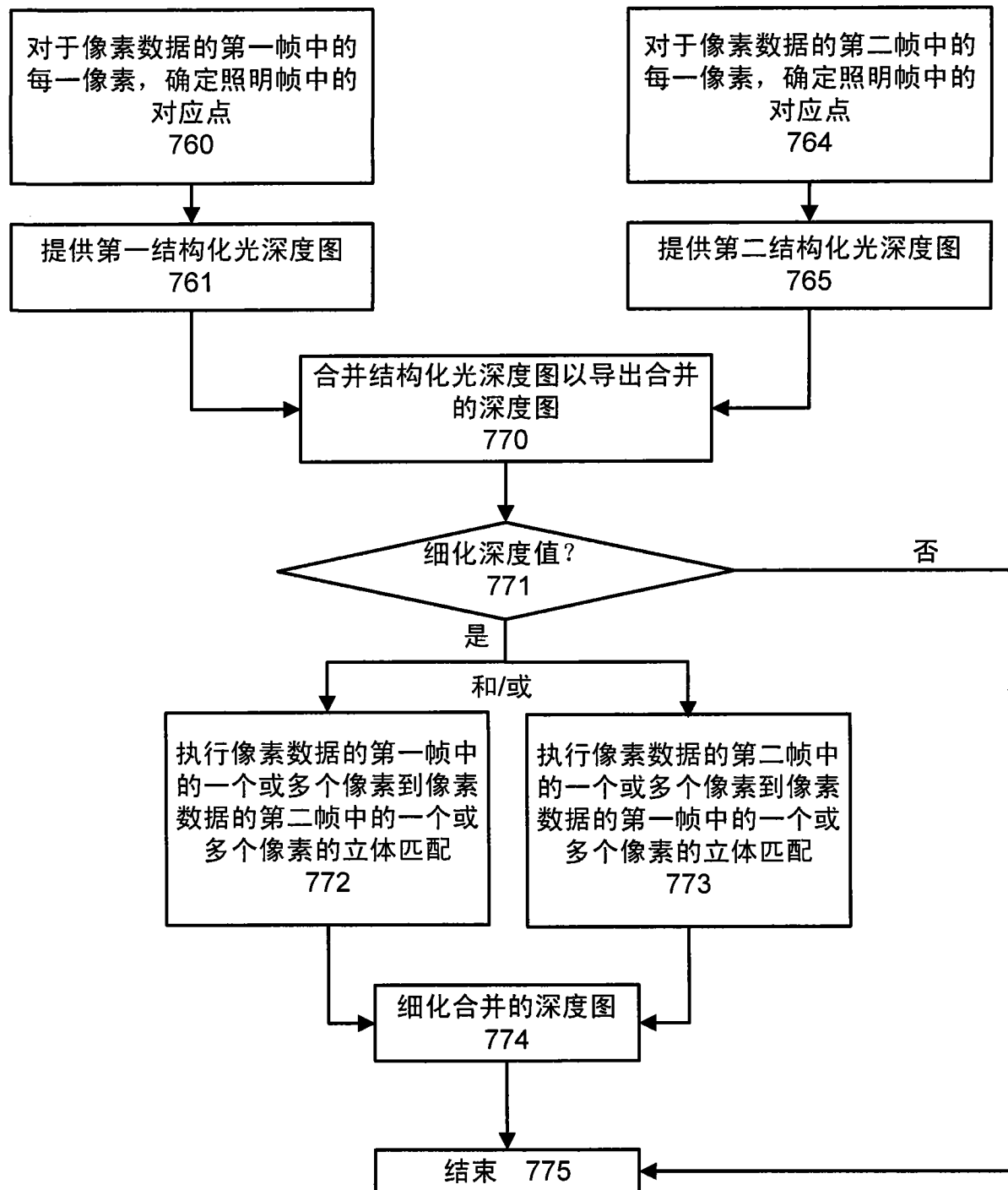


图 7E

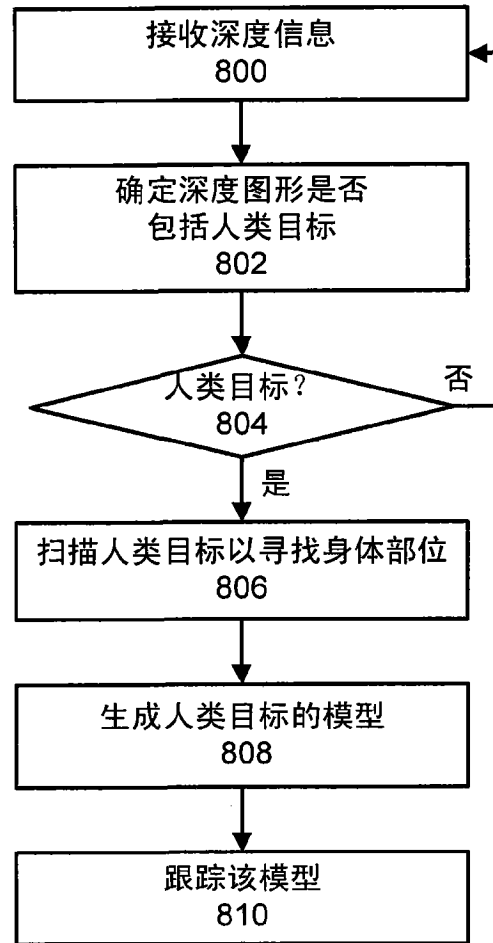


图 8

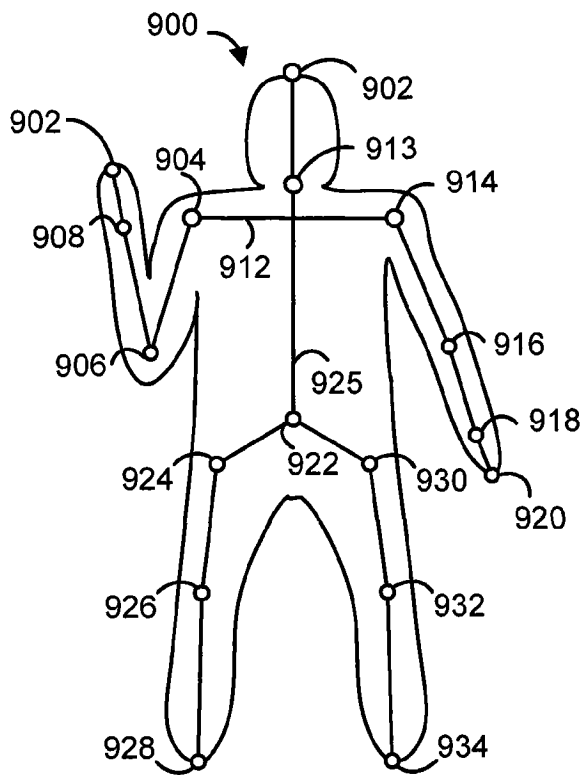


图 9