



(12)发明专利

(10)授权公告号 CN 105745920 B

(45)授权公告日 2019.06.25

(21)申请号 201480063195.2

S·E·克伦卡克

(22)申请日 2014.11.17

(74)专利代理机构 永新专利商标代理有限公司
72002

(65)同一申请的已公布的文献号
申请公布号 CN 105745920 A

代理人 刘瑜 王英

(43)申请公布日 2016.07.06

(51)Int.Cl.

H04N 7/14(2006.01)

(30)优先权数据

14/084,090 2013.11.19 US

(56)对比文件

CN 101141608 A,2008.03.12,说明书第5页
倒数第4段-第7页末段.

(85)PCT国际申请进入国家阶段日
2016.05.19

CN 101141608 A,2008.03.12,说明书第5页
倒数第4段-第7页末段.

(86)PCT国际申请的申请数据

PCT/US2014/065855 2014.11.17

CN 102411783 A,2012.04.11,说明书第

(87)PCT国际申请的公布数据

W02015/077159 EN 2015.05.28

[0002]-[0003]段,第[0022]- [0081]段,附图1-
10.

(73)专利权人 微软技术许可有限责任公司
地址 美国华盛顿州

US 2008/0250458 A1,2008.10.09,全文.
US 2013/0100228 A1,2013.04.25,全文.

审查员 汤茂飞

(72)发明人 D·D·贝雷特 L·德拉涅

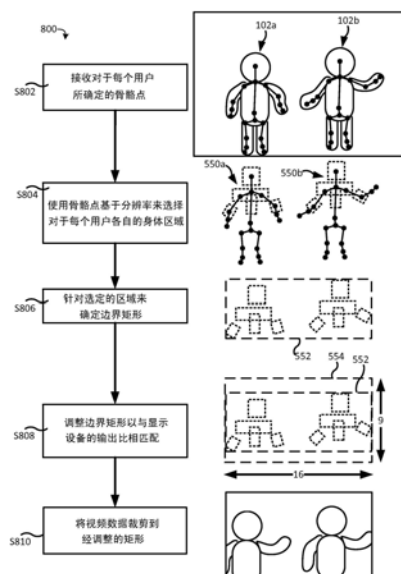
权利要求书2页 说明书14页 附图10页

(54)发明名称

视频传输

(57)摘要

公开了一种经由网络传输视频的方法,以及被配置为实现所述方法的用户设备和计算机程序产品。所述方法包括:将从图像捕捉设备接收的一个或多个用户的视频经由所述网络向至少另一个用户设备传输;接收关于在所述用户设备与另一个用户设备之间的通信信道的和/或关于所述用户设备和/或所述另一个用户设备的一个或多个资源的信息;基于接收到的信息从多个可视用户特征中选择特征;以及基于对选定的特征的检测来控制所述视频,以跟踪选定的特征。



1. 一种用户设备,其包括:

网络接口,其被配置为:将从图像捕捉设备接收的一个或多个用户的视频经由网络传输到至少另一个用户设备;

视频控制器,其被配置为:从多个可视用户特征中选择一个或多个特征,并基于对选定的特征的检测来控制所述视频以跟踪所述选定的特征,对所述选定的特征的所述检测包括:针对所述用户中的至少一个用户,接收该用户的多个检测到的骨骼点,所述视频是基于所述检测到的骨骼点来控制的;以及

资源管理器,其被配置为:接收关于在所述用户设备与所述另一个用户设备之间的通信信道的信息,并且基于接收到的信息来确定待用于所述视频的视频分辨率,其中,所述资源管理器被配置为在接收到的信息指示较好的信道质量的情况下确定较高的视频分辨率,并且在接收到的信息指示较差的信道质量的情况下确定较低的视频分辨率,并且其中,所述选择是基于所确定的视频分辨率的。

2. 根据权利要求1所述的用户设备,其中所述资源管理器被配置为接收关于所述用户设备和/或所述另一个用户设备的一个或多个资源的信息。

3. 根据前述权利要求中任一项所述的用户设备,其中,所述多个可视特征是用户的多个身体部分。

4. 根据权利要求3所述的用户设备,其中,所述多个身体部分包括:头和/或肩和/或中脊柱和/或下脊柱和/或全脊柱和/或肘和/或臀和/或腿和/或脚。

5. 根据权利要求4所述的用户设备,其中,对所述特征进行的所述选择是从四个身体区域中选择一个身体区域,所述四个身体区域中的第一身体区域包括:头和肩;所述四个身体区域中的第二身体区域包括:头、肩、中脊柱和肘;所述四个身体区域中的第三身体区域包括:头、中脊柱、肘、下脊柱和臀;所述四个身体区域中的第四身体区域包括:头、肩、全脊柱、肘、臀、腿和脚。

6. 根据权利要求1或权利要求2所述的用户设备,其中,所述视频是多个用户的视频,所述视频控制器还被配置为:基于对所述用户中的每个用户的所述选定的特征的检测来控制所述视频,以跟踪所述用户中的每个用户的所述选定的特征。

7. 根据权利要求1或权利要求2所述的用户设备,其中,所述网络接口还被配置为:将从所述图像捕捉设备接收的所述一个或多个用户的另外的视频经由设置网络传输到另外的用户设备;

其中,所述视频控制器还被配置为:执行从所述多个可视用户特征中对一个或多个特征的另外的选择,并基于对选定的另外的特征的检测来控制所述另外的视频以跟踪所述选定的另外的特征;以及

其中,资源管理器还被配置为:接收关于在所述用户设备与所述另外的用户设备之间的通信信道的另外的信息和/或关于所述另外的用户设备的一个或多个资源的另外的信息,并且基于接收到的另外的信息来控制由所述视频控制器进行的所述选择。

8. 根据权利要求1或权利要求2所述的用户设备,其中,所述控制包括:基于对所述选定的特征的所述检测来生成边界数据,以及下列各项中的至少一项:基于所生成的边界数据对从所述图像捕捉设备接收的视频数据进行处理,以及基于所生成的边界数据来操纵所述图像捕捉设备的运作方式。

9. 至少一种存储代码的计算机可读介质,所述代码当在一个或多个处理器上执行时被配置为实现经由网络传输视频的方法,所述方法包括:

将从图像捕捉设备接收的一个或多个用户的视频经由所述网络传输到至少另一个用户设备;

接收关于在所述用户设备与所述另一个用户设备之间的通信信道的信息;

基于接收到的信息来确定待用于所述视频的视频分辨率,其中,在接收到的信息指示较好的信道质量的情况下确定较高的视频分辨率,并且在接收到的信息指示较差的信道质量的情况下确定较低的视频分辨率;

基于所确定的视频分辨率从多个可视用户特征中选择特征;以及

基于对选定的特征的检测来控制所述视频,以跟踪所述选定的特征,对所述选定的特征的所述检测包括:针对所述用户中的至少一个用户,接收该用户的多个检测到的骨骼点,所述视频是基于所述检测到的骨骼点来控制的。

10. 根据权利要求9所述的至少一种存储代码的计算机可读介质,其中所述方法还包括:接收关于所述用户设备和/或所述另一个用户设备的一个或多个资源的信息。

11. 一种用户设备,其包括:

网络接口,其用于连接到网络;

一个或多个处理器,其被配置为:

将从图像捕捉设备接收的一个或多个用户的视频经由所述网络传输到另一个用户设备;

从多个可视用户特征中选择特征;

基于分别在第一时刻和第二时刻对选定的特征的检测来生成第一组边界数据和第二组边界数据;

使用动态模型基于所述第一组边界数据和所述第二组边界数据来生成过渡数据;以及基于所述过渡数据来控制所述视频,以跟踪所述选定的特征。

12. 根据权利要求11所述的用户设备,其中,所述动态模型包括弹性弹簧模型,其被定义为 $m * \frac{d^2x}{dt^2} = -k * x - D * \frac{dx}{dt}$, 其中,m是质量、k是刚度、D是阻尼,而x是位移、t是时间。

视频传输

背景技术

[0001] 传统通信系统允许设备(如个人计算机或移动设备)的用户在基于分组的计算机网络(例如互联网)上进行语音或视频呼叫。这样的通信系统包括互联网协议语音或视频电话(VoIP)系统。这些系统对用户是有益的,因为它们的成本与传统的固定线路网络或移动蜂窝网络相比通常低得多。对于长距离通信来说情况尤其如此。为了使用VoIP系统,用户在其设备上安装并执行客户端软件。客户端软件建立VoIP连接以及提供其它功能诸如注册和用户认证等。除了语音通信以外,客户端还可以建立连接用于其它通信介质,诸如即时消息收发(“IM”)、SMS消息收发、文件传输和语音邮件。

[0002] 最近,已经向用户设备中集成了互联网能力和功能,例如被布置为连接到其它(例如大屏幕)显示装置的电视机的游戏控制台、电视机本身(通常被称为“智能电视”)、布置为连接到电视机的机顶盒等。这包括将客户端软件集成到游戏控制台、电视机、机顶盒(或类似物)中,以实现在基于分组的计算机网络(例如互联网)上的通信。这种对客户端软件的集成,通过将从远端用户接收的视频信号输出到近端用户,允许将大的、高分辨率的屏幕利用于视频呼叫。另外,通过诸如机顶盒、电视等用户设备可提供显著的处理功率(processing power),尤其是由于大型的、市电供电的消费类电子设备的功率要求与例如移动设备相比较不严格。这可使得在嵌入式通信客户端中能够包括全方位的特征,例如对从连接到用户设备的摄像头或者从其它类似图像输入装置接收的视频数据进行的高质量的语音编码和视频编码。

发明内容

[0003] 提供本发明内容以便以简化的形式引入下面在具体实施方式中进一步描述的概念的选集。本发明内容并不旨在确定要求保护的的主题的关键特征或重要特征,也不旨在用于限制要求保护的的主题的范围。

[0004] 公开了一种用户设备,其包括网络接口、视频控制器和资源管理器。所述网络接口被配置为:将从图像捕捉设备接收的一个或多个用户的视频经由网络传输到至少另一个用户设备。所述视频控制器被配置为:从多个可视用户特征中选择一个或多个特征,并基于对选定的特征的检测来控制所述视频以跟踪所述选定的特征。所述资源管理器被配置为:接收关于在所述用户设备与另一个用户设备之间的通信信道的信息和/或关于所述用户设备和/或另一个用户设备的一个或多个资源的信息,并且基于接收到的信息来控制由所述视频控制器进行的选择。

[0005] 还公开了一种用户设备,其包括用于连接到网络的网络接口以及一个或多个处理器。所述一个或多个处理器被配置为:将从图像捕捉设备接收的一个或多个用户的视频经由所述网络传输到另一个用户设备。所述一个或多个处理器还被配置为:从多个可视用户特征中选择特征;以及基于分别在第一时间和第二时刻对选定的特征的检测来生成第一组边界数据和第二组边界数据。所述一个或多个处理器还被配置为:使用动态模型基于所述第一组边界数据和所述第二组边界数据来生成过渡数据;以及基于所述过渡数据来控制所

述视频,以跟踪所述选定的特征。

[0006] 还公开了相应的方法以及包括被配置用于这些方法中的每种方法的可执行代码的相应的计算机程序产品。

附图说明

[0007] 为了理解本发明主题以及示出本发明主题可以怎样来实施,将通过示例的方式来参考下列附图,在附图中:

[0008] 图1是一种通信系统的示意性说明;

[0009] 图2是一种用户设备的示意性框图;

[0010] 图3是一种视频控制技术的功能图;

[0011] 图4A和4B是用户与用户设备进行交互的示意性说明;

[0012] 图5是辐射图案的示意性说明;

[0013] 图6A、6B和6C分别是用户的示意性正视图、俯视图和侧视图;

[0014] 图7A是用户的检测到的骨骼点的示意性说明;

[0015] 图7B是外推出的用户身体部分的示意性说明;

[0016] 图8是一种控制视频的方法的流程图;

[0017] 图9是一种随着时间推移来控制视频的方法的流程图。

具体实施方式

[0018] 本文中公开了一种技术,由此在视频呼叫(一对一,或多方)期间,使用深度检测装置(深度检测器或类似物)来检测用户在房间中的位置,并且所述呼叫的视频在向(一个或多个)其它用户传输时被控制,以基于该检测来跟踪用户。更具体地说,该用户的各个可视特征(例如该用户的身体部分)被这样检测,并且该视频被控制以跟踪那些特征(例如,身体部分)中被选择的(但不一定是全部的)特征。特定可视特征是否被跟踪取决于通信信道状况(例如,信道带宽)和/或近端和/或远端设备资源—例如,针对较低的信道带宽和/或远端设备的较小屏幕尺寸可跟踪(例如)较少的特征(例如,身体部分),而针对较高的信道带宽和/或远端设备的较大屏幕尺寸可跟踪较多的特征(例如,身体部分)。

[0019] 图1示出了包括下列各项的通信系统100:与第一用户设备104相关联的第一用户102a(“用户A”)、与第二用户设备110相关联的第二用户108(“用户B”)、与第三用户设备114相关联的第三用户112(“用户C”)、以及在第一用户附近的第四用户102b(“用户D”)。在其它实施例中,通信系统100可包括任意数量的用户和相关联的用户设备。用户设备104、110和114可在通信系统100的网络106上通信,从而允许用户102a、102b、108和112在网络106上与彼此通信。图1中示出的通信系统100是基于分组的通信系统,但其它类型的通信系统也可被使用。网络106可以是例如互联网。用户设备104和110中的每个用户设备可以是例如移动电话、平板电脑、膝上电脑、个人计算机(“PC”) (包括例如Windows™、Mac OS™和Linux™个人计算机)、游戏控制台或其它游戏设备、电视机、个人数字助理(“PDA”)或其它能够连接到网络106的嵌入式设备。用户设备104被配置为:从用户102a、102b接收信息以及向用户102a、102b输出信息。用户设备104包括诸如显示器和扬声器的输出模块。用户设备104还包括输入模块,诸如小键盘、触摸屏、用于接收音频信号的麦克风和/或用于对由视频帧(图像)序

列组成的视频信号的图像进行捕捉的摄像头。用户设备104连接到网络106。

[0020] 用户设备104执行由与通信系统100相关联的软件提供商提供的通信客户端的实例。该通信客户端是在用户设备104的本地处理器上执行的软件程序。该客户端实现在用户设备104处要求的处理,以使用户设备104通过通信系统100传输和接收数据。

[0021] 用户设备110对应于用户设备104,并将与在用户设备104处执行的通信客户端相对应的通信客户端在本地处理器上执行。用户设备110处的客户端以与用户设备104处的客户端实现为允许用户102a和102b通过网络106进行通信所需的处理相同的方式来实现为允许用户108通过网络106进行通信所需的处理。用户设备114对应于用户设备104,并将与在用户设备104处执行的通信客户端相对应的通信客户端在本地处理器上执行。用户设备110处的客户端以与用户设备104处的客户端实现为允许用户102a和102b通过网络106进行通信所需的处理相同的方式来实现为允许用户108通过网络106进行通信所需的处理。用户设备104、110和114是通信系统100中的端点。为了清晰起见,图1仅示出了四个用户(102a、102b、108和112)和三个用户设备(104和110),但更多的用户和用户设备可包括在通信系统100中,并且可使用在各个用户设备上执行的各个通信客户端通过通信系统100进行通信。

[0022] 图2示出了用户设备104的详细视图,在用户设备104上执行了用于通过通信系统100进行通信的通信客户端实例206。用户设备104包括处理器或“处理模块”202,处理器包括一个或多个中央处理单元(“CPU”),其连接有:输出设备(如可实现为触摸屏的显示器208,用于输出音频信号的扬声器(或“扩音器”)210,以及用于输出不可见辐射的辐射投影仪224(如红外线(IR)投影仪));输入设备(如用于接收音频信号的麦克风212、用于接收图像数据的摄像头216、小键盘218以及用于接收不可见辐射信号的不可见图像的辐射传感器阵列226,例如,用于接收红外信号的图像的IR摄像头);用于存储数据的存储器214;以及网络接口220(如用于与网络106进行通信的调制解调器)。扬声器210经由音频信号处理模块209连接到处理器202。用户设备104可包括不同于图2中示出的元件的其它元件。显示器208、扬声器210、麦克风212、存储器214、摄像头216、小键盘218和网络接口220可集成到用户设备104中。或者,显示器208、扬声器210、麦克风212、存储器214、摄像头216、小键盘218和网络接口220中的一个或多个可以不被集成到用户设备104中,而是可经由各个接口连接到处理器202。这样的接口的一个示例是USB接口。例如,对于游戏控制台形式的用户设备来说,摄像头216、投影仪224、传感器226、显示器208、麦克风212、小键盘218(例如,游戏控制器)可以不被集成到游戏控制台中,而是可以经由各自的接口连接到处理器202,同时网络接口220、存储器214和处理器202可集成到游戏控制台中。如果用户设备104经由网络接口220到网络106的连接是无线连接,那么网络接口220可包括天线,天线用于以无线的方式向网络106传输信号以及以无线的方式从网络106接收信号。

[0023] 投影仪224和传感器226构成用于以三维(“3D”)捕捉不可见辐射数据的深度检测器222。在该实施例中,投影仪224在传感器226的前方投射辐射图案(其是通过传感器226可检测的);来自传感器226的传感器数据被用来基于检测到的辐射图案(radiation pattern)中的扭曲(distortion)来建立3D图像(如参考图5和图6A-6C更详细解释的)。深度检测器222和摄像头216可一起容纳于用户设备104外部的单个单元中,可能具有独立于用户设备104的电源,经由合适的接口(例如,USB或基于USB的)连接到处理器202。这样的单元的一个示例是微软Kinect传感器™。

[0024] 图2还示出了在处理器202上执行的操作系统(“OS”)204。除了OS 204之外,还运行通信系统100的客户端实例206的软件。操作系统204管理用户设备104的硬件资源,并处理正在经由网络接口220被传输去向和来自网络106的数据。客户端206与操作系统204通信,并管理通过通信系统的连接。客户端206具有客户端用户界面,其用于向用户102呈现信息以及接收来自用户102的信息。以这种方式,客户端206实现为允许用户102通过通信系统100进行通信所必需的处理。

[0025] 参考图3,现在将描述通过网络传输视频的方法。图3是用户设备104的部分的功能图。

[0026] 如图3中所示,用户设备104包括编码器308(例如,H264编码器)。这可以由处理器202向其供应视频数据以便在通过网络106传输之前进行编码的硬件编码器,由(例如)在处理器202上执行的软件实现的软件编码器,或者这二者的组合。用户设备还包括:控制器302(例如,实现为客户端106的软件的一部分)、摄像头216、深度检测器222和深度处理器310。控制器302包括资源管理器306和视频信号处理器304。摄像头216在可见光光谱(即,人眼可见的)中捕捉用户102a、102b的视频(图3中标记为“近端视频”)的图像,并将这些捕捉到的图像供应给视频信号处理器的第一输入。传感器226在不可见光光谱(即,人眼不可见的)中捕捉图像,并将这些捕捉到的图像供应给深度处理器310的输入。深度处理器310具有被耦合到视频信号处理器的第二输入的输出。投影仪224朝向用户102a、102b在传感器226的前方投射不可见辐射,该不可见辐射是通过传感器226可检测的。控制器300、编码器308和传感器310构成视频处理系统300。

[0027] 资源管理器306具有第一、第二和第三输入,以及输出。资源管理器306的第一输入被配置为:接收关于在用户设备104与网络106的一个或多个其它用户设备(例如,108、112)之间的一个或多个通信信道的信息。资源管理器306的第二输入被配置为:接收关于用户设备104和/或一个或多个其它用户设备(例如,110、114)的资源的信息。资源管理器的第三输入耦合到视频信号处理器304的输出。

[0028] 资源管理器302的输出耦合到视频信号处理器304的第三输入。视频信号处理器304具有耦合到编码器308的输入的输出。编码器被配置为:供应经编码的视频数据,这些视频数据在供应给网络接口220以便通过网络106传输到第二用户设备110和第三用户设备114中的至少一个用户设备之前,可能在设备104处经受进一步处理(例如,分组)。

[0029] 传感器处理器310可操作用于:对由传感器226捕捉的以及从传感器226接收的传感器数据进行处理,以检测在传感器226的视场中的对于每个用户的各自的多个可视用户特征。

[0030] 视频信号处理器302被配置为:基于从资源管理器306输出的信息而从多个可视用户特征中选择特征,并且控制被供应给编码器308的视频以便基于由传感器处理器310对选定的特征的检测来进行编码,其中视频被这样地控制以跟踪选定的特征(即,以使得视频随时间推移而跟踪选定的特征,最终使检测到的特征在视频中可见)。

[0031] 这将参考图4A和4B来进行描述。图4A和4B示出了用户设备104(其在该实施例中示为游戏控制台)附近的用户102(例如,102a、102b)。用户设备耦合到显示器208、摄像头216、投影仪224和传感器226,在该实施例中上述各项都在用户设备104外部。在图4A中,如由摄像头216捕捉的第一可视区域402a的视频(限于包括用户102的上部(例如,头、肩等)而不是

用户102的下部(例如,脚、腿等)被供应给编码器308用于编码,并最终通过网络106传输到另一个用户(例如,112、108)。如图4中所示,区域408a的视频的传输到另一个用户的一个版本也可选地被显示在显示器208上,覆盖在从该另一个用户接收的视频上。在图4B中,由摄像头216捕捉的第二可视区域402b(其既包括用户102的上部又包括用户102的下部)的视频被供应给编码器308用于编码,并最终通过网络106传输到另一个用户。如图4中所示,第二区域408b的视频的传输到另一个用户的一个版本也可选地显示在显示器208上,覆盖在从该另一个用户接收的视频上。

[0032] 第一和第二可视区域402a、402b具有如下文解释的由视频控制模块基于从资源管理器306接收的数据以及从传感器处理器310接收的数据来确定的各自的尺寸和位置。

[0033] 图5示出了由投影仪224发出的辐射图案。如图5中所示,辐射图案是系统性不均匀的,其包括交替亮度的多个系统性布置的区域。具体来说,图5的辐射图案包括辐射点的基本均匀的阵列。在该实施例中,辐射图案是红外线(IR)辐射图案,并且还是通过传感器226可检测的。图5的辐射图案是示例性的,并且可以设想其它辐射图案的使用。

[0034] 辐射图案500由投影仪226在传感器224的前方投射。传感器224捕捉投射在其视场中的不可见辐射图案的图像。这些图像由传感器处理器310处理,以便计算传感器226的视场中的用户的深度(有效地构建该用户的三维表示),从而允许对不同用户及其各自不同身体部分的识别。

[0035] 图6B示出了用户设备104附近的用户102的俯视图,图6A是相应的正视图(如摄像头216和检测器222的传感器226所见)并且图6C是相应的侧视图。如图所示,用户102正摆出其左手朝检测器222伸出的姿势(尽管用户当然可能摆出其它姿势)。用户的头伸出超出其躯干,躯干在其右臂前方。辐射图案500由投影仪224投射到用户身上。

[0036] 如图6A中所示,因此,用户102具有的形体起到对传感器224检测到的所投射的辐射图案进行扭曲的作用,与该辐射的被投射到用户的更靠近投影仪的部分上(即,在这种情况下,以使辐射图案的点间隔较小)的部分相比,该辐射图案的被投射到用户的更加远离投影仪的部分上(即,在这种情况下,以使辐射图案的点间隔较大)的部分有效地伸展,其中伸展的量随着距投影仪的间隔来缩放,并且所述辐射的被投射到显著地在用户后方的对象上的部分对于传感器226来说是有效地不可见的。因为辐射图案500是系统性不均匀的,因此辐射图案的由用户的形体造成的扭曲可用于辨别该形体(例如,识别用户的头、左手、右臂、躯干等),这是通过传感器处理器310处理由传感器226捕捉的扭曲的辐射图案的图像。例如,用户的区域与传感器的间隔可通过测量用户的该区域中检测到的辐射图案的点的间隔来确定。

[0037] 尽管在图5和图6A中,辐射图案被示为人眼可见的,但这仅是为了帮助理解,并且在现实中投射到用户102上的辐射图案将是人眼不可见的。

[0038] 如现在将参考图7A和图7B所描述的,由传感器226捕捉的不可见辐射图案500的图像(传感器数据)由传感器处理器310处理。该处理包括骨骼检测处理。骨骼检测处理是本领域中已知的,并且例如,当前实现在微软Xbox 360™中(有时结合微软Kinect传感器™一起使用),其结果通过应用编程接口(API)的方式对于软件开发者为可用的。

[0039] 传感器检测器310从传感器226接收传感器数据,并对数据进行处理,以确定传感器226的视场中的多个用户(例如,102a、102b),并使用本领域中已知的骨骼检测技术来识

别对于每个用户的各自的多个骨骼点。每个骨骼点表示视频中相应的人体关节的近似位置。

[0040] 具体而言,在该实施例中,传感器检测器310检测到传感器226的视场中的对于每个用户的各自二十个骨骼点。每个骨骼点与二十个识别出的人体关节中的一个关节相对应,当(一个或多个)用户在传感器226的视场内移动时,每个关节在空间和时间中变化。基于由传感器226检测到的用户的(或多个用户各自的)三维形体,来计算这些关节在任意时刻的位置。骨骼点还具有跟踪状态:其可以针对清晰可见的关节而被“跟踪”,当关节不清晰可见但传感器处理器310正在推测其位置时可以对其进行“推测”,或者是“非跟踪”的(例如,针对坐姿模式跟踪中较低的关节(即,当检测到:用户坐着并且较低关节不被传感器处理器310跟踪时))。

[0041] 每个骨骼点可设有指示相应关节已被正确检测的概率的各自的置信值。具有低于某个阈值的置信值的点可从视频模块304的处理中排除。

[0042] 图7A中示出了这二十个骨骼点,表1中示出了相应的人体关节。

[0043] 表1-骨骼点

骨骼点的名称:	相应的人体关节:	标记为:
AnkleLeft	左脚踝	722b
AnkleRight	右脚踝	722a
ElbowLeft	左肘	706b
ElbowRight	右肘	706a
FootLeft	左脚	724b
FootRight	右脚	724a
HandLeft	左手	702b
HandRight	右手	702a
Head	头	710
HipCenter	臀之间的中心	716
HipLeft	左臀	718b
HipRight	右臀	718a
KneeLeft	左膝盖	720b
KneeRight	右膝盖	720a
ShoulderCenter	肩之间的中心	712
ShoulderLeft	左肩	708b
ShoulderRight	右肩	708a
Spine	脊柱	714
WristLeft	左腕	704b
WristRight	右腕	704a

[0046] 骨骼点和来自摄像头216的视频相互关联,以使得由传感器处理器在特定时刻报告的骨骼点的位置与该时刻视频的帧(图像)内的相应人体关节的位置相对应。传感器检测器310将这些检测到的骨骼点作为骨骼点数据供应给视频控制器304以便由其使用。对于视频数据中的每个帧,由传感器处理器310供应的骨骼点数据包括骨骼点在该帧内的位置(例如,表示为关于视频帧尺寸而定边界的坐标系的笛卡尔坐标(x,y))。

[0047] 视频控制器304接收一个或多个用户(102a、102b)的检测到的骨骼点,并且被配置为:由此确定该用户的多个可视特征(或这些用户的各自的多个可视特征)。在该实施例中,可视用户特征采取人体部分的形式。身体部分由视频控制器检测,分别通过根据由视频处理器310提供的一个或多个骨骼点进行外推(extrapolation)的方式来检测,并且与来自摄像头216的视频的相应视频帧内的区域(即,定义为在前述坐标系内的区域)相对应。图7B示

出了检测到的身体部分,其已经基于图7A的骨骼点而被检测到。

[0048] 表2-身体部分(可视特征)

[0049]

身体部分的名称(可视特征)	图7B中检测到的特征:
头	750
肩	752
中脊柱(脊柱的上部)	756
下脊柱(脊柱的下部)	758
全脊柱(脊柱的全部)	760
臀	762
肘	754a, 754b
腿	764a, 764b
脚	766a, 766b

[0050] 应该指出的是:这些可视特征在它们表示用户身体的特征的意义上是可视的,所述特征在现实中是可见并且是可分辨的;然而在该实施例中,视频控制器没有“看到”它们(例如,没有在由摄像头216捕捉的视频数据中检测到它们);而是视频控制器根据由传感器处理器310提供的二十个骨骼点的布置,在来自摄像头216的视频的帧内对这些特征的(近似)相对位置、形状和尺寸进行外推(extrapolate)(而不是基于例如对该帧的处理)——例如,通过将每个身体部分近似为具有根据与该身体部分有密切关系的骨骼点的检测到的布置来计算出的位置和尺寸(以及可选地,方向)的矩形(或类似物)。

[0051] 现在将参考图8来描述基于检测到的用户特征(在该实施例中,身体部分)来控制将在网络上传输的视频的方法800。图8在左边示出了方法800的流程图,并且在右边示出了每个方法步骤的相应可视表示。

[0052] 在该实施例中,该方法在算法上实现为客户端106的软件的一部分——具体而言由控制器302实现。

[0053] 方法800将在使用第一用户(近端)设备104和第二用户(远端)设备110在设备的用户之间进行的实时视频呼叫的上下文中描述。

[0054] 该方法考虑以下身体区域:

[0055] -区域1:头、肩

[0056] -区域2:头、肩、中脊柱、肘

[0057] -区域3:头、肩、中脊柱、肘、下脊柱、臀

[0058] -区域4:头、肩、全脊柱、肘、臀、腿、脚

[0059] 在视频呼叫的开始,近端设备104的客户端106的资源管理器306基于由此接收的信息来确定视频分辨率(要被用于传输到远端设备110的视频)。例如,该信息可包括关于下列各项中的一项或多项的信息:

[0060] -网络资源,如在近端设备和远端设备之间的通信信道(视频通过该信道从近端设备传输到远端设备)的信道带宽;

[0061] -当从远端设备接收时的远端设备110的资源(例如,屏幕尺寸),从而与具有较少像素的屏幕相比,具有较多像素的屏幕拥有较多的屏幕资源;远端设备的屏幕的宽高比(例

如,4:3,16:9等),其为屏幕的水平长度与屏幕的竖直长度之比;远端设备的一个或多个处理器的处理资源(例如,以兆赫兹、千兆赫兹等单位的可用处理器周期);远端设备的存储器资源(例如,以兆赫兹、千兆赫兹等单位的可用存储器);视频解码能力

[0062] -近端设备的资源,例如,近端设备的一个或多个处理器的处理资源(例如,以兆赫兹、千兆赫兹等单位的可用处理器周期);远端设备的存储器资源(例如,以兆赫兹、千兆赫兹等单位的可用存储器);视频编码能力

[0063] 可以通过与远端设备110协商来至少部分确定视频分辨率(例如,从而近端资源管理器306从其请求关于远端设备的资源的信息)。

[0064] 在呼叫期间,近端设备104的资源管理器306监测可用资源(例如,可用带宽),做出对正在向近端设备108发送的视频的分辨率进行增加或减少的决定,并向视频控制器304传达那些决定。尽管由于经由不可靠无线(例如,WiFi)连接而连接到网络106的近端设备和远端设备中的一者或二者而出现的例如波动信道带宽的呼叫,但视频分辨率可因此而动态变化。

[0065] 对特征的选择包括:如果接收到的信息指示较好的信道质量和/或较多的设备资源,则选择较多的特征;而如果接收到的信息指示较差的信道质量和/或较少的设备资源,则选择较少的特征。例如,对于较低带宽和/或较小屏幕尺寸来说,资源管理器306确定较低的分辨率(导致对具有较少身体部分的区域的选择);对于较高带宽和/或较大屏幕尺寸来说,与针对所述较低带宽和/或较小屏幕尺寸所确定的分辨率相比,资源管理器306确定较高的分辨率(导致对具有较多身体部分的区域的选择)。

[0066] 在该实施例中,视频是以1920x1080像素的固定分辨率从摄像头216捕捉的。然而,如下所述,可在传输之前对捕捉到的视频进行裁剪(裁剪是删除视频的图像的外围部分)。也就是说,仅将捕捉到的视频数据的选择的部分(如同从选定的可视区域捕捉的)供应给编码器308用于编码和随后向远端设备的传输。因此,供应经裁剪的视频数据意味着向编码器308供应较少的视频数据,因为所定义的区域外部的视频数据不被供应给编码器。

[0067] 在步骤S802,骨骼点数据由视频控制器从传感器处理器310接收。在图8的右手侧的示例性图示中,示出了并置(collocated)的用户102a和102b(二者都在传感器226的视场中)的骨骼数据。然而,将显而易见的是:方法800可应用于接收到的任意数量的(一个或多个)用户的骨骼数据。

[0068] 在步骤S804,视频控制器304基于从资源管理器306接收的确定的分辨率来选择在传感器216的视场中的对于每个用户的各自的身体区域。根据该视频分辨率,视频控制器从表2的可能的多个身体区域中选择身体区域如下:

[0069] -分辨率1920×1080: 区域4

[0070] -分辨率等于或高于1280×1024: 区域3

[0071] -分辨率等于或高于640×480: 区域2

[0072] -其它分辨率: 区域1

[0073] 或者可替换地如下:

[0074] -分辨率宽度1920: 区域4

[0075] -分辨率宽度等于或高于1280: 区域3

[0076] -分辨率宽度等于或高于640: 区域2

[0077] -其它分辨率: 区域1

[0078] 特征的选择包括:为较高的分辨率(较好的信道状况和/或较多的设备资源)选择较多的特征,以及为较低分辨率(较差的信道状况和/或较少的设备资源)选择较少的特征。

[0079] 对于传感器226的视场中的每个用户102a、102b来说,选定的区域是基于接收到的骨骼数据检测的(也就是说,检测到选定的区域的各个身体部分)。视频在编码和传输之前基于检测到的身体部分进行裁剪(如下所述)(例如,对于区域1来说,基于用户102a的头、用户102b的头、用户102a的肩以及用户102b的肩,而不基于那些用户的任何其它身体部分来裁剪视频,因为在区域1中只包括头和肩)。因此,不是所有的骨骼点都一定对裁剪有所贡献(因为这些骨骼点中一些可能只用来对不属于选定的区域的的部分的身体部分进行定义)。也就是说,对视频的控制可以基于多个检测到的骨骼点中的骨骼点的一个选集,而不是基于所述多个检测到的骨骼点中的其它骨骼点。

[0080] 例如,在图8的右手侧的示例性描绘中,已经选择了区域2(头、肩、中脊柱、肘),并且因此视频控制基于对于用户102a(图8中的550a)和用户102b(图8中的550b)二者的检测到的区域2的特征。

[0081] 作为视频控制的一部分(其在该实施例中包括根据所确定的分辨率来选择性地裁剪视频),在步骤S806,视频控制器基于对选定的区域的身体部分的检测来生成边界数据。边界数据定义矩形边界552(裁剪矩形)。裁剪矩形是根据所有被跟踪的用户的各自的检测到的区域的并集而形成的。

[0082] 在步骤S808,视频控制器304基于预先确定的宽高比(例如,远端设备110的显示器的宽高比,如4:3、16:9等)来修改所生成的边界数据,由此,裁剪矩形被调整到预先确定的宽高比。经调整的裁剪矩形(边界矩形)在图8的右手侧示为554(并且在该实施例中具有16:9的宽高比)。

[0083] 在步骤S810,如下文更详细解释的,基于经调整的矩形来对视频进行裁剪。然后在由编码器308进行编码之前,将经裁剪的视频缩放到输出分辨率(例如,将远端设备的屏幕的分辨率与正在传输的相匹配)。

[0084] 方法800针对视频数据的每个帧来执行。因为边界矩形在帧之间跟踪用户特征(当用户到处移动时,用户特征可能移动,或者例如在两个用户移动分开时,用户特征可能分离),裁剪矩形到处移动。

[0085] 然而,过渡数据是基于针对一个或多个先前的帧计算出的各自的经调整的矩形并且基于针对当前帧计算出的经调整的矩形生成的,而不是简单地将视频数据的每个帧裁剪到针对该帧确定的经调整的矩形(例如,通过仅供应来自自由经调整的矩形定义的该帧的部分的视频数据)。过渡数据是基于弹性弹簧模型生成的。

[0086] 在实施例中,弹性弹簧模型可定义如下:

$$[0087] \quad m * \frac{d^2x}{dt^2} = -k * x - D * \frac{dx}{dt}$$

[0088] 其中,m(“质量”)、k(“刚度”)和D(“阻尼”)是可配置的常数,而x(位移)和t(时间)是变量。这是“过渡的加速度”与“该过渡的位移和速度的加权和”成比例的模式。

[0089] 裁剪矩形根据弹性弹簧模型到处移动。这平滑了帧之间的过渡并且防止了抖动视

频。出于以下原因,其还增加了编码效率。因为弹性模型有效地‘阻滞’了裁剪矩形的移动,其减少了相邻帧之间的差异,这将显而易见导致更高效的差分编码。

[0090] 现在将参考图9来进行描述。图9示出了随时间推移来控制将要向远端用户传输的视频的方法900的流程图,并且在右侧示出了方法900的示例性胸部表示。

[0091] 在步骤S902,针对在捕捉到的视频的视频帧序列中当前的视频帧956,视频控制器基于分别在第一时间以及在一个或多个第二时刻对选定的可视用户特征的检测来生成第一组边界数据以及一或多第二组边界数据(例如,如上所述通过计算具有预先确定的宽高比的边界矩形),第一时间是当前帧956的时刻,而(一个或多个)第二时刻是先前帧的(一个或多个)时刻。较早的帧的边界数据可以在较晚的帧的边界数据之前生成(例如,基于逐个帧)。因此,S902可在跨度几个视频帧的时间段上进行。

[0092] 在图9的右侧示出了分别在第一时间 t_1 和第二时刻 t_2 的两个示例边界矩形950、952。针对 t_2 的矩形被示为比针对 t_1 的矩形更大,这可以是,例如,由于两个用户(102a、102b)正在被跟踪以及那些用户进一步移动分开,或者由于可用资源的变化(例如,信道带宽增加)导致资源管理器306增加所确定的宽高比,从而导致具有较多身体部分的区域被选择(例如,从区域1向区域2的切换),并且边界矩形由视频控制器304相应地自动调整。

[0093] 在步骤S904,视频模块304使用上述动态模型基于第一和第二组边界数据来生成过渡数据。所生成的过渡数据有效地定义过渡边界矩形954(图9中所示的),其表示从先前帧的边界矩形向当前帧956的边界矩形的过渡。

[0094] 例如,根据上文的弹性弹簧模型,边界矩形可由不同位置处的一个或多个点(边界数据的一种形式)来参数化。一个点在时刻 t_2 可以具有第二位置(“desiredPosition(所期望的位置)”),该第二位置是矩形952的参数,并且在时刻 t_1 具有第一位置“currentPosition(当前位置)”,该第一位置是矩形950的参数。在这种情况下,过渡数据可通过按如下方式更新“当前位置”来生成,其中,经更新的“当前位置”是过渡边界矩形954的参数:

```
[0095] velocity=0
[0096] previousTime=0
[0097] currentPosition=<some_constant_initial_value>
[0098] UpdatePosition(desiredPosition,time)
[0099] {
[0100] x=currentPosition-desiredPosition;
[0101] force=-stiffness*x-damping*m_velocity;
[0102] acceleration=force/mass;
[0103] dt=time-previousTime;
[0104] velocity+=acceleration*dt;
[0105] currentPosition+=velocity*dt;
[0106] previousTime=time;
[0107] }
```

[0108] 在步骤S908处,视频控制器基于过渡数据来控制视频,以跟踪选定的特征。在这里,这涉及裁剪视频,以使得仅仅把帧956的在边界矩形954内的图像数据(而不是帧956的

在边界矩形954之外的图像数据)供应给编码器308,用于编码以及向远端用户108传输。

[0109] 检测器222和传感器处理器310被配置以使得不对与检测器222相隔大于约1米(也就是说在传感器的视场之外)的用户进行跟踪,以使得他们不影响裁剪矩形。因此,例如,如果两个用户最初被跟踪,并且一个用户移动到了视场之外,那么根据方法800和900,此后计算出的裁剪矩形将仅基于剩下的用户的特征,并且将因此在剩下的用户上自动“放大”(弹性模型确保该缩放的平滑过渡)。

[0110] 在一些实施例中,视频控制器304可以反馈回到资源管理器306,以使得如果资源管理器认为存在用于1920x1080视频的足够带宽,但至少一个身体要被跟踪,那么视频发送分辨率切换到1280x720,以便能够进行动态跟踪(对于所传输的视频来说,分辨率等于其被捕捉的分辨率,没有“回旋余地”,因为对于每个帧来说,裁剪矩形有效地包覆该整个帧-例如,在不使用按比例放大的情况下,情况就是如此)。

[0111] 如上所述,传感器处理器310不仅供应识别每个检测到的骨骼点的信息,还识别那些骨骼点与传感器226的视场中的一个或多个用户中的哪个用户相对应。因此,所描述的方法800可关于任意数量的用户实现(只要传感器226仍然能够在他们之间进行区分),并且针对选定的区域中的每个特征,将跟踪每个区域的那些特征。因此,当多个用户走入和走出帧时,方法800自动适应,边界矩形以平滑的方式(由于弹性模型)自动过渡以便在新的用户走入帧时容纳他们(通过有效地缩小以针对视频中的所选区域包括该用户的任何身体部分),并且当用户走出该帧时进行调整(通过有效地放大以针对选定的区域排除之前由该用户的身体部分占据的任何区域,以使得仅针对剩下的用户的选定区域的身体部分保留在视频中)。

[0112] 由控制器302实现的方法800、900可在使用例如用户设备104、110和114(用户102a、102b、108和112之间的)的网络106上进行的多方呼叫期间实现,各个边界矩形由每个远端设备110(第二设备)和114(第三设备)的近端用户104(第一设备)确定。

[0113] 也就是说,除了上述之外,资源管理器还可接收关于下列各项中的至少一项的信息:在用户设备104与第三用户设备114之间的通信信道,以及另外的用户设备的资源(除了接收上述信息,如对于第二设备110的类似的信息)。资源管理器然后基于接收到的另外信息选择从多个可视用户特征(例如,身体部分)中选定的另外的特征,用来控制要向第三用户设备114传输的视频(除了选择用于控制要向第二用户设备110传输的视频的前述特征之外)。

[0114] 视频控制器304然后基于针对第三设备114选择的所述选定的另外的特征检测,来控制要向第三用户设备114传输的视频(以跟踪第三设备视频中的所述选定的另外的特征),同时基于对针对第二设备110选定的特征检测来控制要向第二用户设备110传输的视频。

[0115] 针对第三用户设备的另外的特征选择可以是独立的,并且不同于针对第二用户设备的特征选择。因此,向第二[第三]用户设备传输的另外的视频可以不同于向第二用户设备传输的视频,与向第二用户设备传输的视频相比,向第三用户设备传输的另外的视频跟踪更多或更少的用户特征。

[0116] 针对第二用户设备视频的特征选择独立于针对第三用户视频的特征选择。虽然某种情况(如第一用户设备104经由慢连接来连接到网络106)可导致针对二者的类似的裁剪,

其它情况(如第二和第三设备中的一个设备经由慢连接来连接到网络106,或者那些设备中的一个设备具有有限的资源)可导致不同的裁剪。

[0117] 例如,第三用户设备114可具有较小的屏幕(例如,智能电话屏幕)和/或经由慢连接来连接到网络106;与此相反,第二用户设备110可具有较大的屏幕(例如,连接到电视机屏幕)和/或经由快连接来连接到网络106。在这种情况下,向第二用户设备传输的视频可经受“区域4”裁剪(见表2),以便用户108接收从头到脚示出用户102a和102b的视频。与此相反,向用户设备114传输的视频可经受“区域1”裁剪(见表2),以便用户112接收仅示出用户102a、102b各自的头和肩的视频。

[0118] 尽管在上文中,要在网络上传输的视频是通过视频信号处理的方式控制的,但或者或另外,视频可由(控制器302的)视频控制器304通过基于对选定的特征的检测来操纵摄像头本身来控制,例如,操纵包括:通过操纵摄像头的运作方式(mechanics)来执行下列操作中的至少一项:平移操作、变焦操作和倾斜操作。例如,对于多方呼叫来说,光学变焦和数字变焦(裁剪)可结合例如用于捕捉示出最高选择数量的用户特征的视频的机械变焦(例如,在上述示例中,机械变焦可用于选择要向第二用户设备110展示的区域)来使用,其中该视频的裁剪用于针对将要向其发送具有较少用户特征的视频的用户(例如,上述示例中的第三设备114)来控制视频。

[0119] 另外,尽管在上文中使用了弹性模型,但作为替代方案,任何动态模型(例如,基于一个或多个时间上的微分方程)可用于生成过渡数据。

[0120] 另外,虽然在上文中,选定的可视特征(腿,手臂等)是基于由深度检测器(其在被配置用于检测非可视辐射图案的传感器的前方投射非可视辐射图案)供应的传感器数据来检测的,但可以设想替代的检测。例如,深度检测可以是基于飞行时间的检测,其中,辐射传播时间用于测量深度。或者,具有不同倾度的摄像头阵列(如与3D图像识别算法结合使用的全光摄像头)或类似物可用于构建3D图像(根据多个二维图像),其中,可视特征是从3D图像检测到的。

[0121] 通常,本文中描述的功能中的任意功能(例如,图3所示的功能模块以及图8和图9所示的功能步骤)可使用软件、固件、硬件(例如,固定逻辑电路系统)或这些实现的组合来实现。图3中单独示出的模块(视频处理系统300、控制器302、视频信号处理器304、资源管理器306、编码器308等)以及图8和图9中单独示出的步骤可以或可以不实现为单独的模块或步骤。本文中所使用的术语“模块”、“功能”、“组件”和“逻辑”通常表示软件、固件、硬件或者它们的组合。在软件实现的情况下,模块、功能或逻辑表示当在处理器(例如,CPU或一些CPU)上执行时执行特定任务的程序代码。程序代码可存储在一个或多个计算机可读存储器设备中。本文中描述的技术的特征是独立于平台的,意指这些技术可在具有多种处理器的多种商用计算平台上实现。例如,用户设备还可以包括使得用户设备的硬件来执行操作(例如,处理器功能块)的实体(例如,软件)等。例如,用户设备可包括计算机可读介质,计算机可读介质可被配置为:保持使用户设备(更具体地说,操作系统及用户设备的相关联硬件)执行操作。例如,图3的模块中的一些或所有模块可由在一个或多个处理器上执行的客户端应用的软件来实现。因此,这些指令用于配置操作系统和相关联的硬件执行操作并由此导致操作系统和相关联的硬件的变换来执行功能。这些指令可由计算机可读介质通过各种不同配置向用户设备提供。

[0122] 计算机可读介质的一种这样的配置是信号承载介质,并且因此被配置为向计算设备(如经由网络)传输指令(例如,作为载波波形)。计算机可读介质还可被配置为计算机可读存储介质,并且因此不是信号承载介质。计算机可读存储介质的示例包括:随机存取存储器(RAM)、只读存储器(ROM)、光盘、闪存存储器、硬盘存储器以及可使用磁、光及其它技术来存储指令和其它数据的其它存储器设备。

[0123] 虽然以特定于结构特征和/或方法动作的语言对发明主题进行了描述,但应当理解的是:所附权利要求书中定义的发明主题并不一定受限于上述具体特征或动作。更确切地说,上述具体特征或动作是作为实现权利要求书的示例形式而公开的。

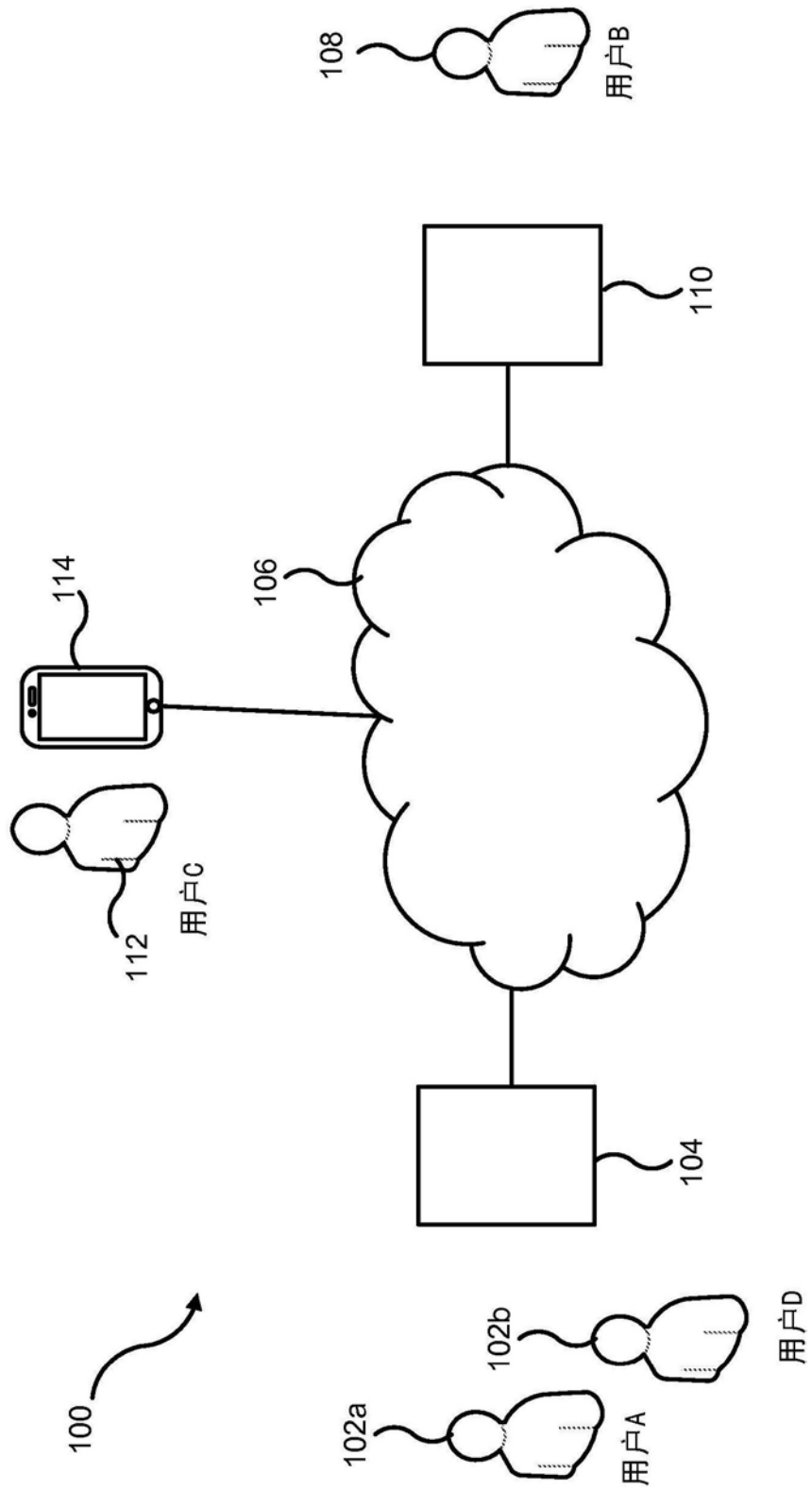


图1

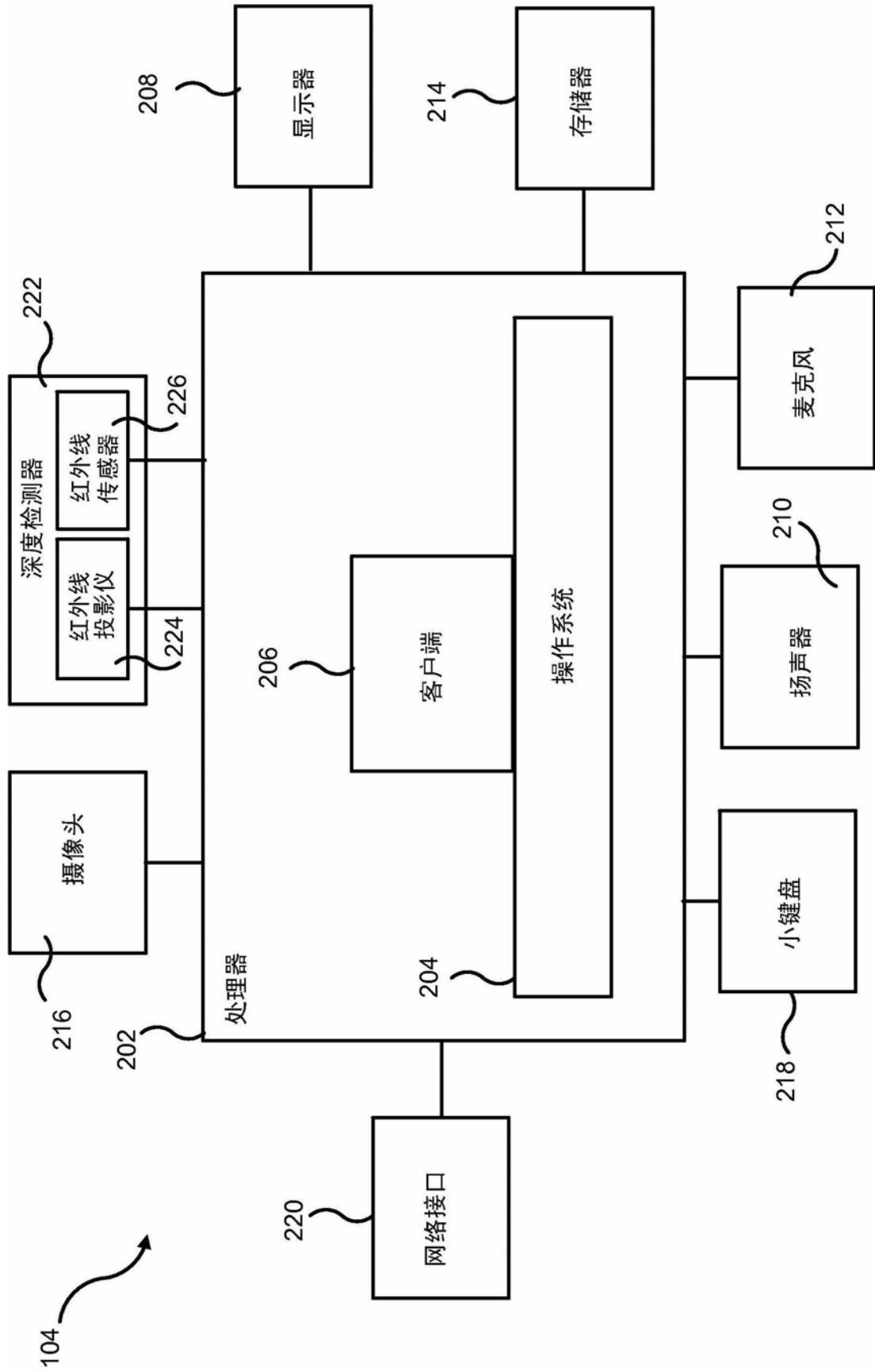


图2

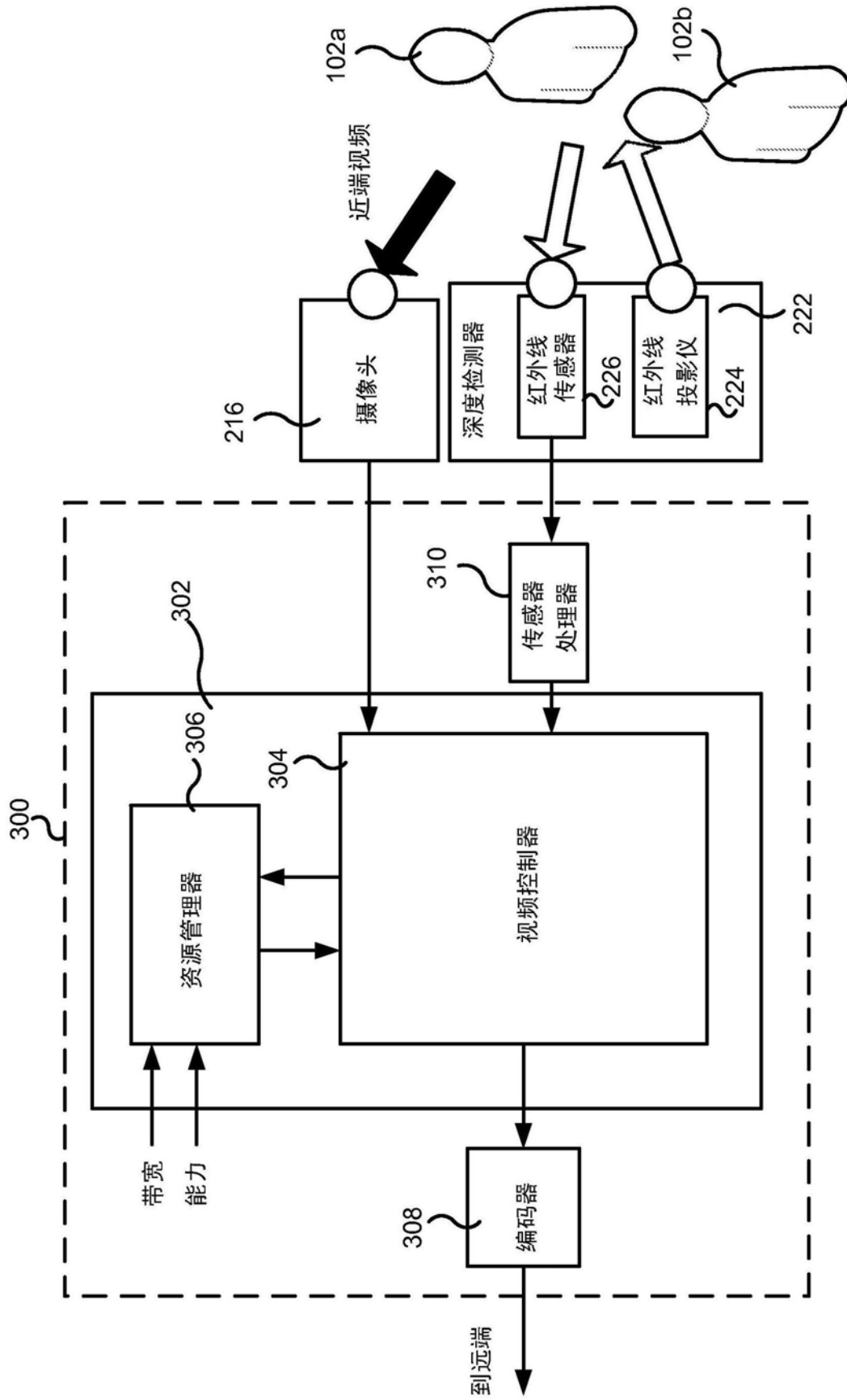


图3

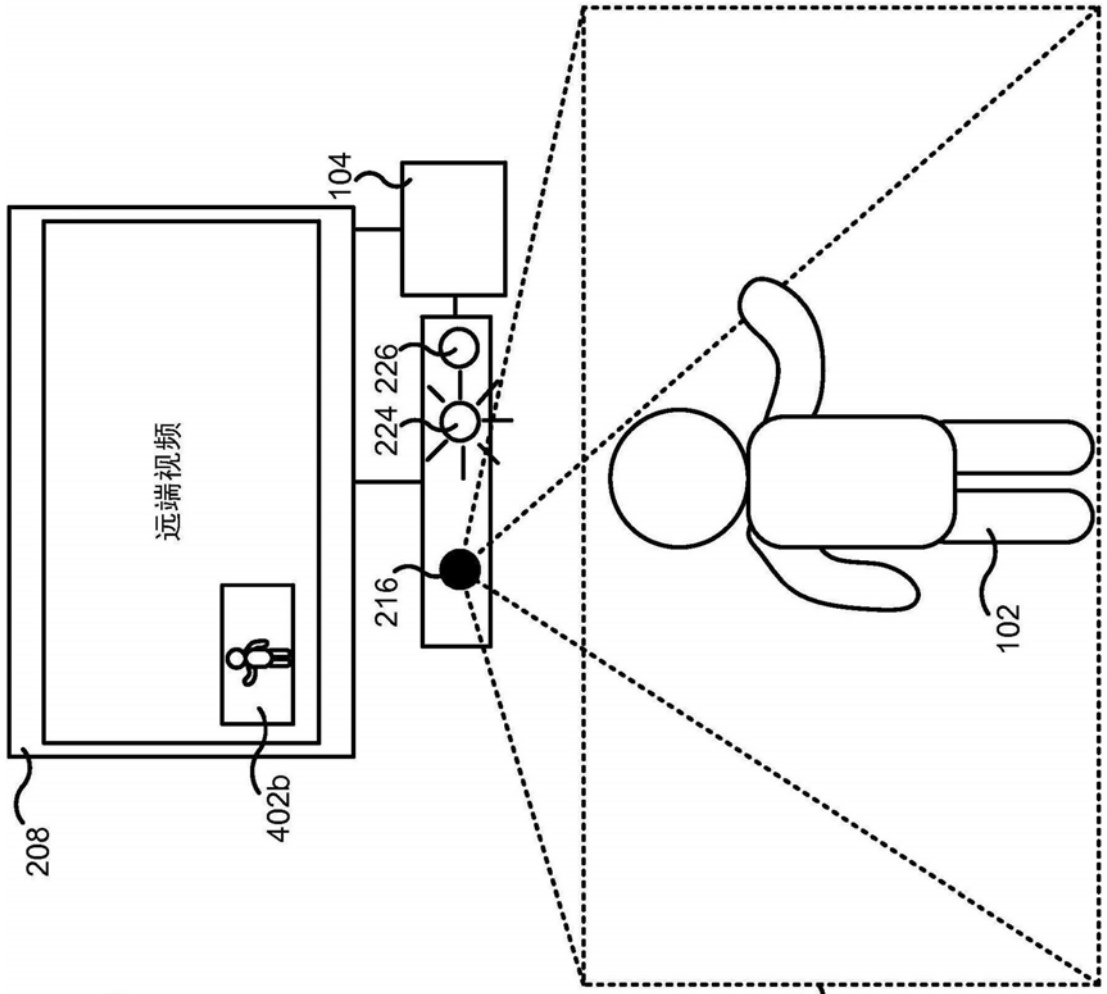


图4A

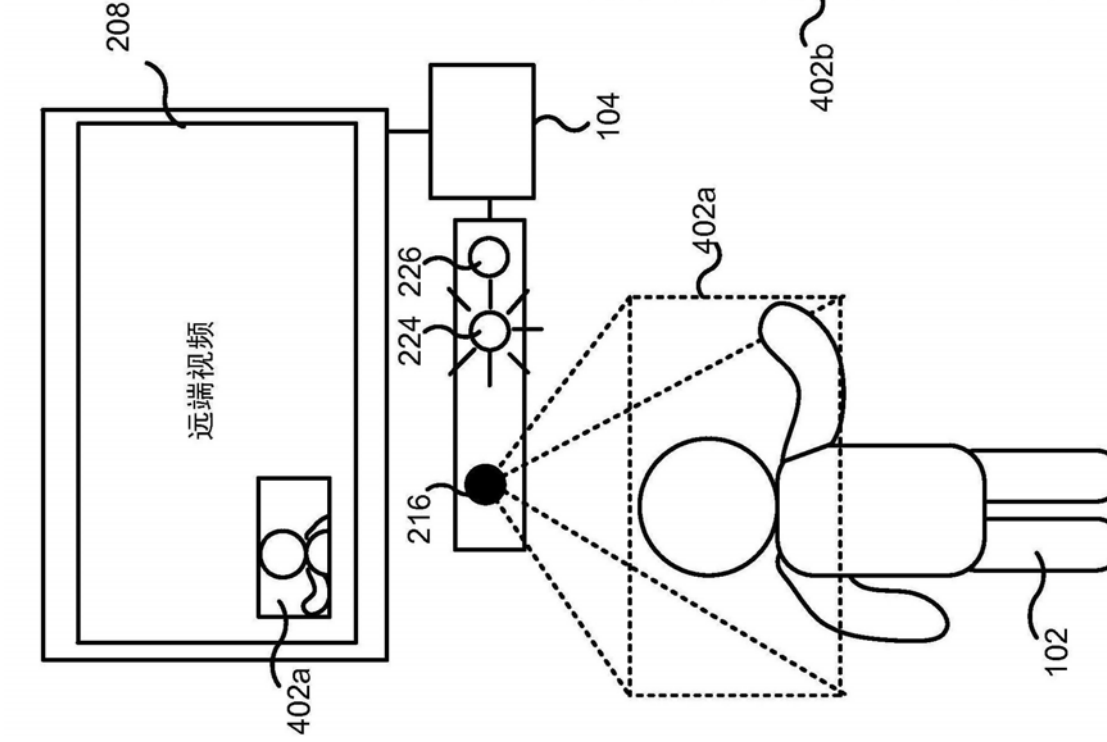


图4B

投射的图案

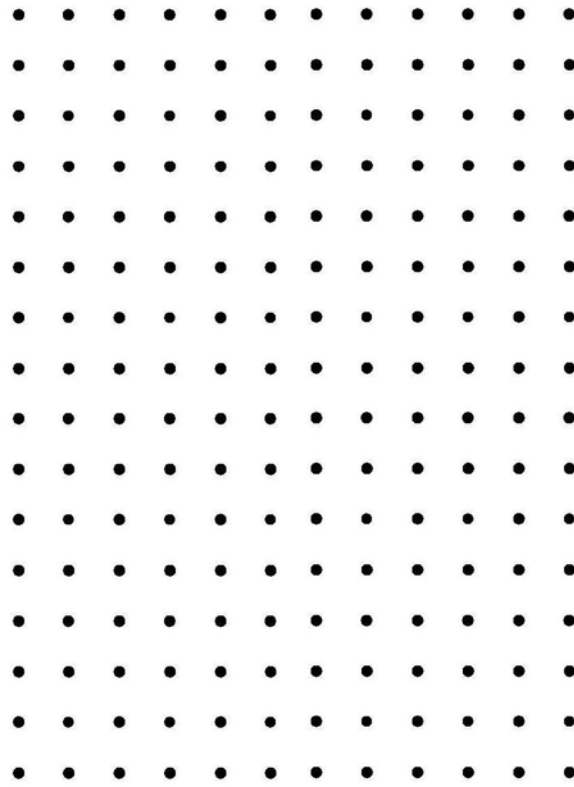


图5

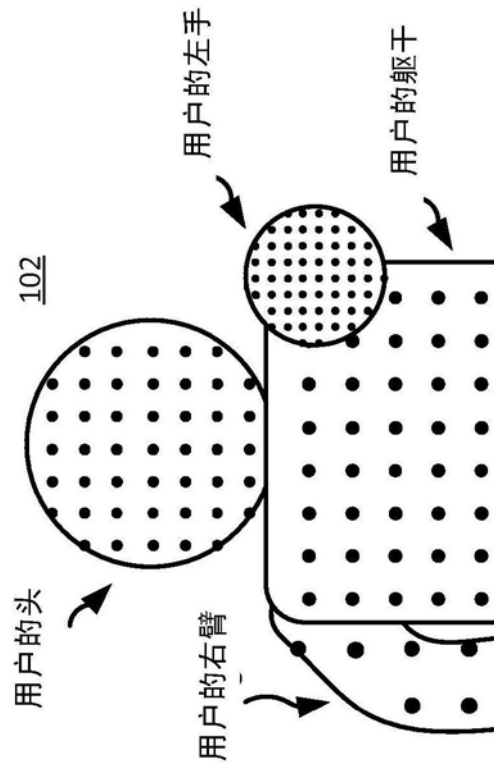


图6A正视图

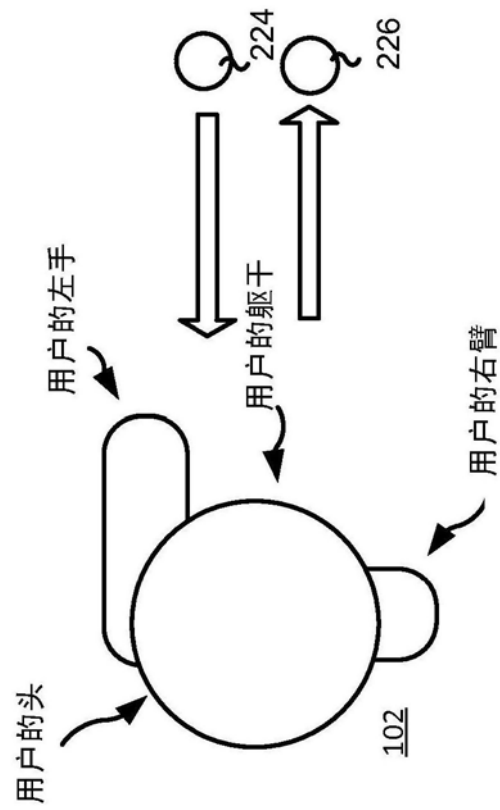


图6B俯视图

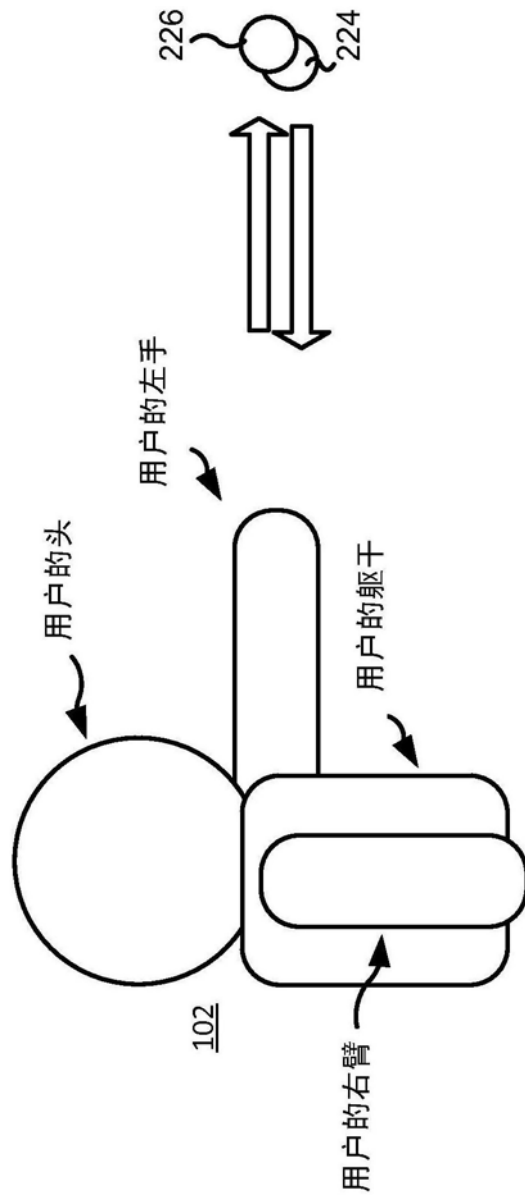


图6C侧视图

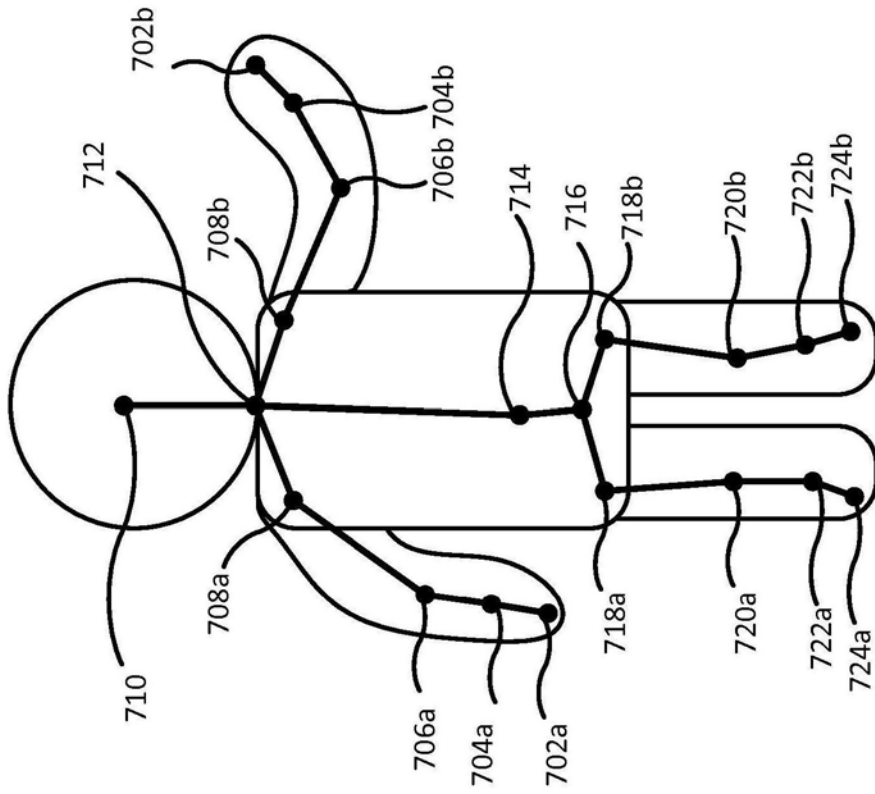


图7A

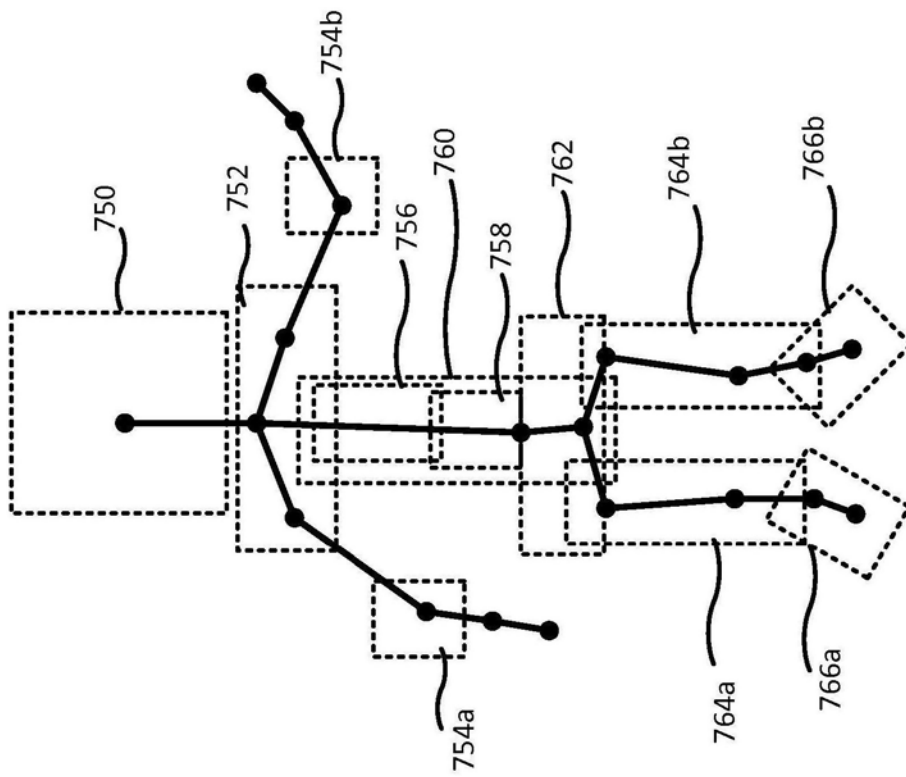


图7B

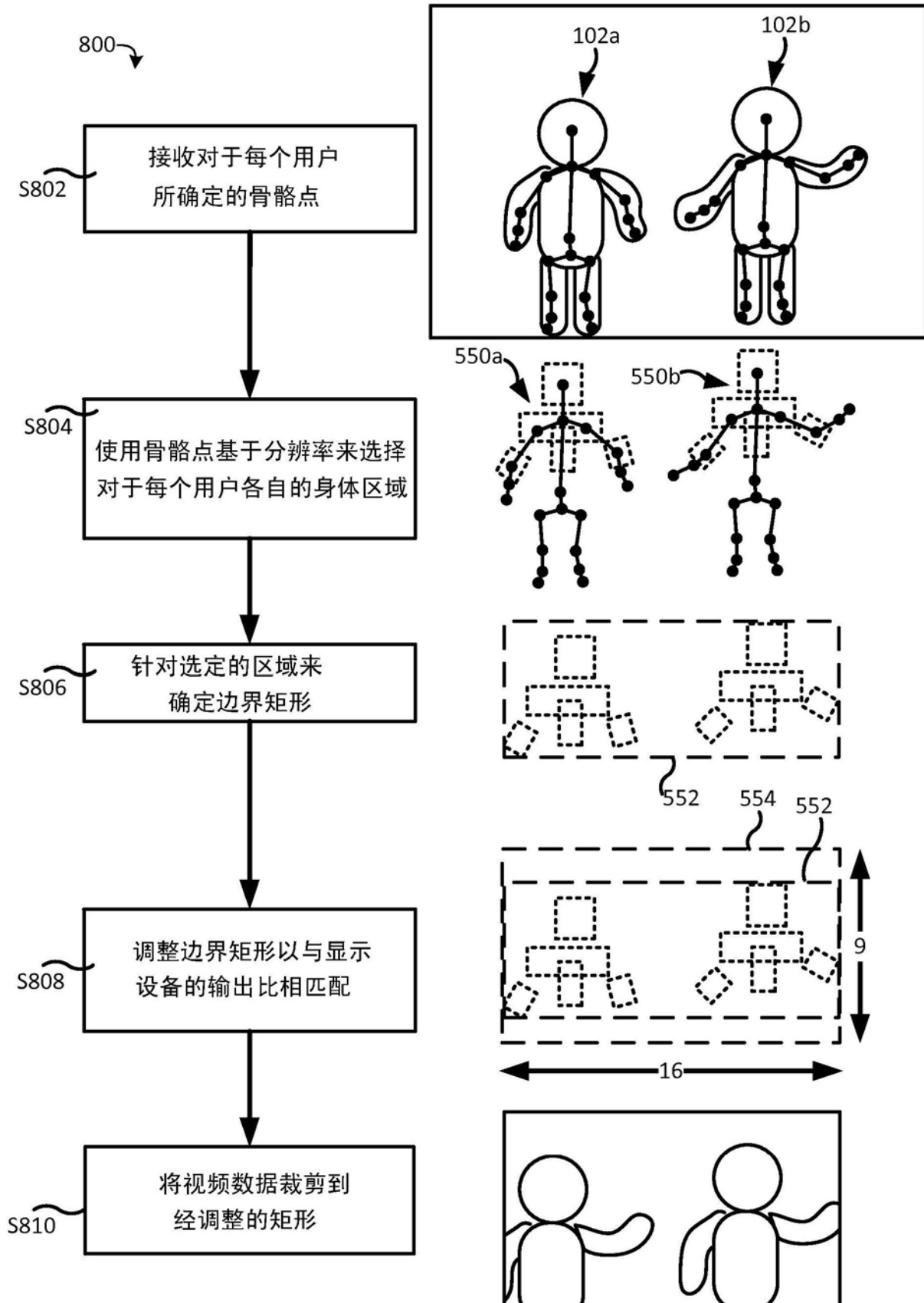


图8

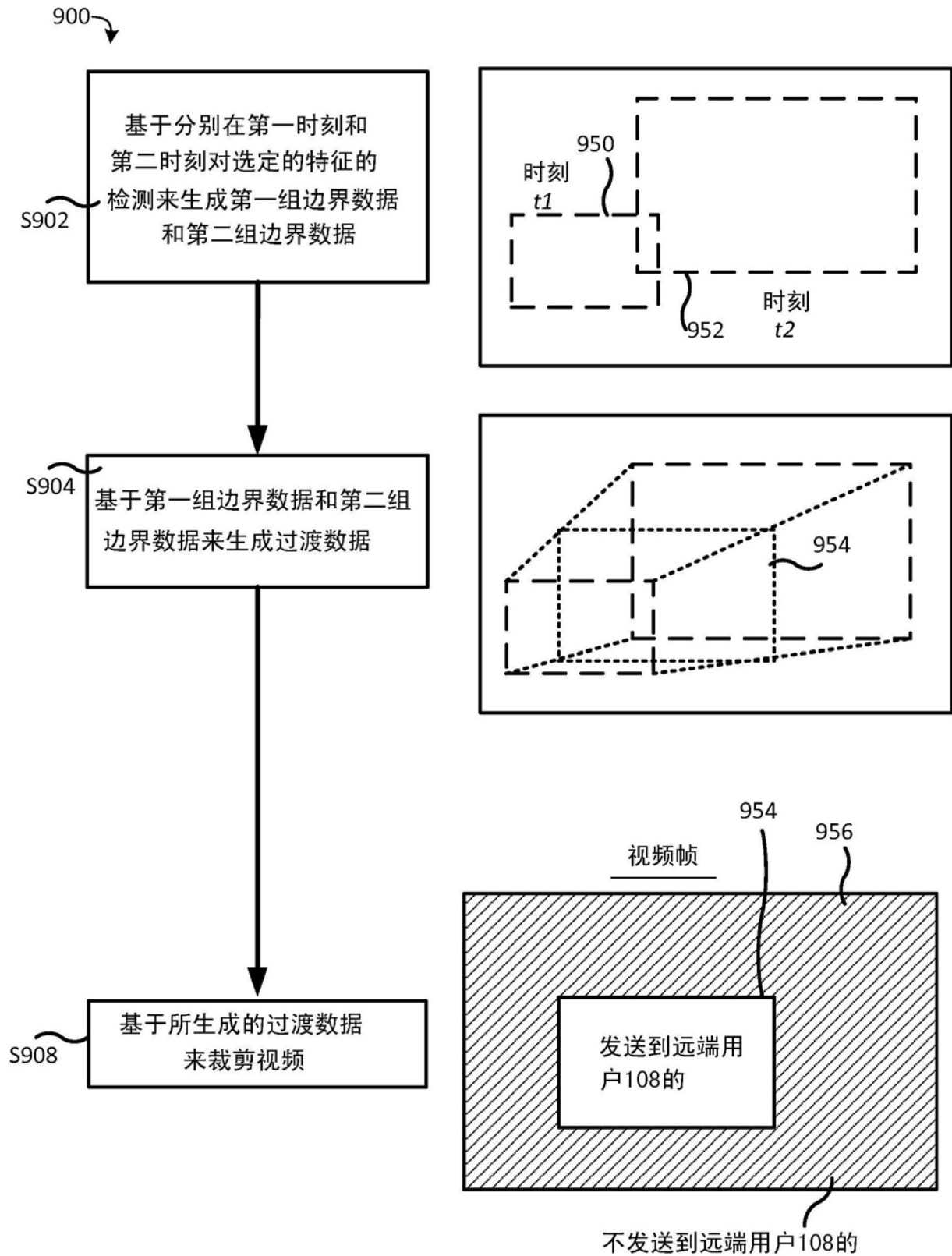


图9