



(12)发明专利

(10)授权公告号 CN 106255978 B

(45)授权公告日 2019.09.03

(21)申请号 201580023604.0

(74)专利代理机构 北京市金杜律师事务所

(22)申请日 2015.04.23

11256

(65)同一申请的已公布的文献号

代理人 王茂华 辛鸣

申请公布号 CN 106255978 A

(51)Int.Cl.

G06K 9/00(2006.01)

(43)申请公布日 2016.12.21

G02B 27/00(2006.01)

(30)优先权数据

G02B 27/01(2006.01)

14/265,201 2014.04.29 US

A63F 13/213(2014.01)

(85)PCT国际申请进入国家阶段日

A63F 13/212(2014.01)

2016.10.28

A63F 13/56(2014.01)

(86)PCT国际申请的申请数据

A63F 13/327(2014.01)

PCT/US2015/027188 2015.04.23

(56)对比文件

(87)PCT国际申请的公布数据

US 2013117377 A1,2013.05.09,

W02015/167909 EN 2015.11.05

US 8068169 B2,2011.11.29,

(73)专利权人 微软技术许可有限责任公司

WO 2013152453 A1,2013.10.17,

地址 美国华盛顿州

CN 101283379 A,2008.10.08,

(72)发明人 张察 张正友

审查员 郑婷

B·罗梅拉帕雷德斯

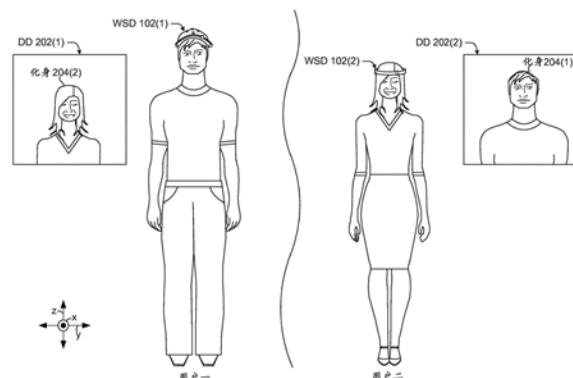
权利要求书2页 说明书10页 附图11页

(54)发明名称

面部表情跟踪

(57)摘要

本说明书涉及面部跟踪。一个示例可以包括定向结构，其被配置为相对于用户的面部来放置可穿戴设备。该示例还可以包括摄像机，其被定向结构固定为与用户的面部平行或成低角度以捕捉跨用户的面部的图像。该示例还可以包括处理器，其被配置为接收图像并且以及将图像映射到与化身模型相关联的参数。



1.一种系统,包括:

智能眼镜,其包括:

框架,其固定为戴着所述智能眼镜的用户看透而被放置的一对镜片;以及

至少一个传感器,其相对于所述框架被固定并且被配置为捕捉包括戴着所述智能眼镜的所述用户的面颊或下颚的图像的低角度传感器数据,

所述至少一个传感器相对于与戴着所述智能眼镜的所述用户的面部平行的基准平面、以小于45度的角度被定向;以及

处理器,其被配置为:

获得所述低角度传感器数据;

使用模型以使来自所述低角度传感器数据的所述面颊或下颚的所述图像与包括针对代表戴着所述智能眼镜的所述用户的化身的嘴部运动的动画参数相关;以及

输出所述动画参数。

2.根据权利要求1所述的系统,所述智能眼镜还包括:至少一个其他传感器,其在所述框架上朝着戴着所述智能眼镜的所述用户的眼睛被定向并且被配置为捕捉所述用户的眼睛的其他图像,

其中所述处理器被配置为通过使用所述模型以处理所述图像和所述其他图像两者来获得所述动画参数。

3.根据权利要求2所述的系统,其中所述至少一个传感器和所述至少一个其他传感器包括可见光摄像机或红外摄像机。

4.根据权利要求1所述的系统,其中所述至少一个传感器包括鱼眼镜片,所述鱼眼镜片能够在单独图像中捕捉所述用户的眼睛中的至少一只眼睛以及戴着所述智能眼镜的所述用户的嘴部。

5.根据权利要求1所述的系统,其中所述框架包括一对镜脚并且其中所述至少一个传感器大体上与所述一对镜脚中的单独镜脚的长轴垂直地被定向。

6.根据权利要求1所述的系统,还包括:

与所述智能眼镜分开的计算设备,所述处理器被体现在所述计算设备中,

其中所述智能眼镜还包括电路,所述电路被配置为向所述计算设备传输所述低角度传感器数据。

7.根据权利要求1所述的系统,其中所述处理器被提供在所述智能眼镜上。

8.根据权利要求1所述的系统,其中用于确定所述嘴部运动的所述面颊或下颚的所述图像不包括戴着所述智能眼镜的所述用户的嘴部的图像。

9.一种智能设备,包括:

定向结构,其被配置为相对于戴着所述智能设备的用户的面部来放置所述智能设备;

平行或低角度摄像机,其被所述定向结构固定为与所述用户的面部平行或成低角度以捕捉包括戴着所述智能设备的所述用户的面颊或下颚的图像的平行或低角度图像数据;

处理器,其被配置为:

从所述平行或低角度摄像机获得所述平行或低角度图像数据;

将来自所述平行或低角度图像数据的所述面颊或下颚的所述图像与代表戴着所述智能设备的所述用户的嘴部运动的嘴部动画参数相关;以及

输出所述嘴部动画参数。

10. 根据权利要求9所述的智能设备,表现为智能眼镜并且其中所述定向结构包括框架。

11. 根据权利要求10所述的智能设备,还包括由所述框架支撑的一对镜片,其中所述平行或低角度摄像机相对于与戴着所述智能设备的所述用户的面部平行的基准平面、以小于45度的角度被定向。

12. 根据权利要求11所述的智能设备,其中所述智能眼镜被配置为在所述一对镜片中的一个镜片或两个镜片上呈现图像。

13. 根据权利要求11所述的智能设备,其中所述一对镜片是视力矫正镜片。

14. 根据权利要求9所述的智能设备,表现为智能帽子并且其中所述定向结构包括所述智能帽子的帽舌。

15. 根据权利要求9所述的智能设备,其中所述定向结构包括头带。

16. 根据权利要求9所述的智能设备,其中所述定向结构包括框架并且还包括用于将所述框架固定到戴着所述智能设备的所述用户的衣服的夹子。

17. 根据权利要求9所述的智能设备,还包括由所述定向结构固定的另一摄像机并且其中所述另一摄像机指向戴着所述智能设备的所述用户的眼睛。

18. 根据权利要求17所述的智能设备,其中所述处理器还被配置为:

至少部分基于所述嘴部动画参数来控制代表戴着所述智能设备的所述用户的化身。

19. 一种计算机实现的方法,包括:

接收沿着正戴着可穿戴设备的用户的面部、以低角度被捕捉的低角度图像数据,所述低角度图像数据包括戴着所述可穿戴设备的所述用户的面颊或下颚的图像;

用被配置为将所述面颊或下颚的所述图像映射到嘴部动画参数的模型来处理所述用户的所述面颊或下颚的所述图像;以及

使得所述用户的动画表示根据所述嘴部动画参数来执行嘴部运动。

20. 根据权利要求19所述的方法,其中接收所述低角度图像数据包括通过网络从所述可穿戴设备接收所述低角度图像数据,所述可穿戴设备具有捕捉所述低角度图像数据的传感器。

21. 根据权利要求19所述的方法,还包括:

从在所述用户的前方被放置的基准摄像机获得前方训练图像;

从相对于所述用户的面部、以低角度被放置的低角度摄像机获得低角度训练图像,所述低角度训练图像捕捉所述用户的所述面颊或下颚;以及

使用所述前方训练图像和所述低角度训练图像来训练所述模型。

面部表情跟踪

背景技术

[0001] 各种可穿戴设备由于特定原因而上市。例如，智能眼镜和头戴式显示器可以确定用户正在看什么。然而，这些设备未提供用户的有用视图，因为设备被靠着用户放置并且具有向外指向的摄像机和/或为了眼睛跟踪而指向用户的眼睛的摄像机。本概念可以通过即使当摄像机被极接近用户放置时也提供用户的有用视图来提供关于可穿戴设备的附加功能。

发明内容

[0002] 本说明书涉及面部跟踪。一个示例可以包括定向结构，其被配置为相对于用户的面部来放置可穿戴设备。该示例还可以包括摄像机，其被定向结构固定为与用户的面部平行或成低角度以捕捉跨用户的面部的图像。该示例还可以包括处理器，其被配置为接收图像以及将图像映射到与化身(avatar)模型相关联的参数。

[0003] 以上列出的示例旨在提供快速参考以帮助读者，并且不旨在限定在此描述的概念的范围。

附图说明

[0004] 附图图示了在本文档中传达的概念的实现方式。所示实现的特征可通过结合附图参考以下描述而更容易地理解。只要可行，各附图中相似的标号就用来指代相似的元素。另外，每个标号的最左边的数字传达其中首次引入该标号的附图以及相关联的讨论。

[0005] 图1和图4至图5示出了根据本概念的一些实现方式的可穿戴智能设备的示例。

[0006] 图2至图3示出了根据本概念的一些实现方式的可视化情形的示例。

[0007] 图6是根据本概念的一些实现方式的示例可视化过程。

[0008] 图7至图11B根据本概念的一些实现方式更详细地示出了示例可视化系统的各方面。

[0009] 图12示出了根据本概念的一些实现方式的示例可视化系统。

概述

[0011] 本说明书涉及捕捉关于用户的有用视觉信息，诸如来自被极接近用户放置和/或以与用户成极低角度放置的摄像机的面部信息。这些概念可以被实现在运用低角度摄像机的各种可穿戴智能设备上。

[0012] 这些概念的一个方面包括从用户的低角度图像得出有价值信息的能力。例如，一个或多个摄像机可被靠近用户放置并且捕捉与用户的面部几乎平行的图像。正常情况下这样的图像几乎不具有价值。本实现方式可以从这些图像中得出有用信息。有用信息可以用于各种目的，诸如以逼真方式控制用户的化身。

[0013] 介绍性的图1示出了若干可穿戴智能设备102的示例。在这种情况下可穿戴智能设备102(1)表现为智能帽子，可穿戴智能设备102(2)表现为智能头带，可穿戴智能设备102(3)表现为智能眼镜，并且可穿戴智能设备102(4)表现为智能麦克风。当然，其他示例被预

期到。可穿戴智能设备102可以包括一个或多个传感器。在这一示例中，传感器表现为低角度摄像机104、眼睛跟踪摄像机106和/或麦克风108。(注意，可穿戴智能设备102(1)和102(3)包括低角度摄像机和眼睛跟踪摄像机两者。相比之下，可穿戴智能设备102(2)和102(4)不包括眼睛跟踪摄像机)。摄像机可以是可见光摄像机和/或红外摄像机，和/或深度摄像机等。这些元件的功能在下面被说明。可穿戴智能设备还可以包括定向结构110。定向结构可以表现为框架、带、帽舌，或者帮助保持传感器相对于用户的位置和/或朝向的其他结构。如在上面提到的，所例示的实现方式中的一些实现方式利用低角度摄像机104和眼睛跟踪摄像机106两者。其他实现方式可以利用单个摄像机，该单个摄像机可以捕捉否则将由两个摄像机提供的数据。例如，具有鱼眼镜片的低角度摄像机104可以用来捕捉用户的面部的各部分，诸如眼睛、嘴部、面颊和/或下颌。

[0014] 总而言之，可穿戴智能设备正变得更加常见。可穿戴智能设备经常包括摄像机。然而，因为摄像机如此接近用户和/或被以不寻常的朝向放置，所以摄像机无法提供用户(特别是用户的面部)的有用图像。本实现方式可以按照产生关于用户的面部表情、嘴部运动、眨眼、眼睛注视方向等的有用信息的方式放置这些摄像机并且处理它们的图像。

[0015] 图2示出了涉及穿着可穿戴智能设备102(1)的“用户一”和穿着可穿戴智能设备102(2)的“用户二”以及两个显示设备202(1)和202(2)的用例情形。假定用户一位于靠近显示设备202(1)的第一位置并且用户二位于与第一位置分开某一距离的第二位置并且用户二靠近显示设备202(2)。在这一示例中，(如在图1中示出的)，可穿戴智能设备102(1)包括低角度摄像机104(1)和指向用户一的眼睛的另一摄像机(例如，眼睛跟踪摄像机)106(1)。可穿戴智能设备102(2)包括低角度摄像机104(2)。

[0016] 在这种情形下，可穿戴智能设备102(1)可以捕捉用户一的图像。这些图像可以用来控制用户一的表示。在这种情况下，用户一的表示表现为在靠近用户二的显示设备202(2)上呈现的用户一的化身204(1)。类似地，可穿戴智能设备102(2)可以捕捉用户二的图像。这些图像可以用来控制用户二的表示。在这种情况下，用户二的表示表现为在靠近用户一的显示设备202(1)上呈现的用户二的化身204(2)。在这种情形下，用户可以相互通信，并且它们的相应的化身可以向另一用户反映它们的眼睛运动、面部表情、下颌运动和/或嘴部运动(诸如由发音引起的)。

[0017] 化身204可以表现为任何形式，诸如捕捉眼睛移动、面部表情和/或嘴部运动的卡通角色或者计算机生成的角色，但是不一定像用户。在其他实现方式中，化身可以是用户的真实表示(例如，增强的逼真图像)。用于在讨论期间看见相应的化身的能力可以促进比只有语音的通信更加鲁棒和令人愉快的讨论。例如，用于在讨论期间看见其他用户的反应的能力可以提供模仿面对面交谈的反馈。

[0018] 还注意到，虽然在图2的情形中仅涉及两个用户，但是其他实现可以涉及附加用户。另外，因为低角度摄像机104和/或眼睛跟踪摄像机106可以跟踪用户的眼睛，所以实现方式可以辨别单独用户正在看着哪个用户(例如，用户一和用户二正在看着用户三的化身)。这种信息可以被反映在用户的化身中并且正如在面对面群组情形中一样可以是通信过程的有用部分。

[0019] 图3示出了与图2的用例情形类似的另一用例情形。这一特定用例情形涉及两组可穿戴智能设备102(3)(A)和102(3)(B)。虽然未被明确示出，但是假定可穿戴智能设备被图2

的用户一和用户二穿着。类似于上面的相对于图2的讨论,可穿戴智能设备102(3)(A)可以捕捉用户一的低角度图像以及使用图像数据来控制可以为用户二显示的用户一的表示(例如,化身204(1))。类似地,可穿戴智能设备102(3)(B)可以执行相对于用户二的相同功能以控制可以为用户一显示的用户二的表示(例如,化身204(2))。

[0020] 因此,可穿戴智能设备102(3)(A)示出了用户一在看透可穿戴智能设备102(3)(A)时看见用户二的化身204(2)并且可穿戴智能设备102(3)(B)示出了用户二在看透可穿戴智能设备102(3)(B)时看见用户一的化身204(1)。总之,可穿戴智能设备102(3)(A)和102(3)(B)从用户一和用户二的角度被图示,尽管用户自身未被示出。可穿戴智能设备102(3)(A)和102(3)(B)可以捕捉用来控制穿着者的(例如,用户的)化身的关于穿着者的图像信息。化身被另一用户的可穿戴智能设备示出,该可穿戴智能设备同时捕捉用于控制该用户的化身的图像数据。换言之,用户二的化身204(2)被可穿戴智能设备102(3)(A)为用户一显示。类似地,用户一的化身204(1)被可穿戴智能设备102(3)(B)为用户二显示。因此,这一实现方式可以提供促进视频通话和/或其他用途的独立可穿戴智能设备。

[0021] 图4和图5分别示出了用户三的正视图和侧视图。在这种情况下,“用户三”正穿着可穿戴智能设备102(3)(例如,智能,眼镜)和可穿戴智能设备102(4)(例如,智能耳挂式麦克风)。在这种情况下,用户三为了说明目的而穿着两个可穿戴智能设备。备选地,用户可以使用这些可穿戴智能设备中的任一个而没有另一个。图4至图5还示出了基准平面402,其可以帮助图示当前讨论中的术语“低角度摄像机”意味的含义的示例。在这一实现方式中,基准平面402穿过或者包括用户的眼睛和用户的嘴部的前面。

[0022] 在这一示例中,低角度摄像机104可以被按照相对于基准平面402的少于大约+/-45度的角度被定向(例如,低角度摄像机可以捕捉大体上与用户的面部的表面平行的图像)。相比之下,诸如在智能电话、平板或视频会议系统上运用的那些的传统摄像机被配置为大体上与基准平面402垂直地被定向并且无法从低角度捕捉有意义的图像信息。

[0023] 在这一示例中,可穿戴智能设备102(3)的低角度摄像机104(3)一般被放置在基准平面402内或靠近基准平面402被放置并且被配置为在沿着基准平面的方向或与基准平面平行的方向上捕捉由箭头404表示的图像。类似地,可穿戴智能设备102(4)的低角度摄像机104(4)一般被放置在基准平面402内或靠近基准平面402被放置并且被配置为在沿着基准平面402的方向或与基准平面402平行的方向上捕捉由箭头406表示的图像。

[0024] 图6示出了用于完成本概念的示例处理流程600。为了说明的目的,处理流程600可以被分成训练阶段602和部署阶段604。训练阶段可以在606处为用户构建化身模型。如在上面提到的,化身可以是逼真的或是某一其他形式的。用户的化身可以经由化身模型608被定义。化身模型可以包括诸如对应于眼睛注视、嘴部运动等的多个面部参数或者动画参数。另外,处理可以在610处针对传感输入与化身动画参数之间的相关进行训练以创建映射612。处理流程的这一部分可以通过利用可穿戴智能设备(例如,利用可穿戴智能设备上的传感器)和诸如红色、蓝色、绿色加深度(RGBD)摄像机之类的另一成像设备同时记录用户而被完成。RGBD摄像机可以处于正常(一般与图4的平面402垂直)朝向。用户动作可以被诸如低角度摄像机、麦克风和/或其他传感器(见图1)之类的可穿戴智能设备和RGBD摄像机捕捉。映射612可以使如由RGBD摄像机和可穿戴智能设备捕捉的用户的特定用户动作与特定的化身动画参数相关联。

[0025] 部署阶段604可以单独利用可穿戴智能设备被实现。在这种情况下,各种类型的传感器数据614可以被可穿戴智能设备收集。在这一示例中,根据可穿戴智能设备的传感器配置,传感器数据614可以包括视频数据、音频数据和/或接触数据等。(接触数据可以被感测皮肤运动、肌肉收缩等的与用户的皮相接触的传感器收集)传感器数据614可以在616处被处理以得出传感器输入618。传感器输入618在620处可以利用来自训练阶段602的化身模型608和映射612而被映射到化身动画参数。这一方面输出用于部署阶段604的化身动画参数622。化身动画参数620(和化身模型608)可以通过网络624被发送以用于在626处的远程化身渲染和动画制作。下面关于图7至图12更详细地描述处理流程600的特定方面。

[0026] 总之,本概念可以使得需要一种用于基于身体安装的、向内看(例如,眼睛跟踪)和/或低角度摄像机进行面部表情跟踪的系统。来自这些摄像机的图像可以被利用以使得用户可以用远程方处的动画的化身被表示。

[0027] 图7示出了可以从可穿戴智能设备102(3)提供完整面部表情跟踪的示例系统700。(图7对于图6是有点冗余的,但是具有帮助说明相对于图6引入的概念的不同重点)。为了说明的目的,系统700被分成输入列701(1)和输出列701(2)。

[0028] 在这种情况下,系统700包括基准摄像机,诸如可见光摄像机(例如,黑白或者红绿蓝(RGB))或者红色、绿色、蓝色+深度(RGBD)摄像机702。这样的深度摄像机的一个示例是由微软®公司提供的Kinect®牌深度摄像机。RGBD摄像机702可以按照正常方式面向用户(例如,大体上与在上面相对于图4至图5描述的基准平面垂直地指向用户)。(注意,在图7的示例中,用户面向前方并且RGBD摄像机702在用户一边。RGBD摄像机将实际上在用户前方(例如,被定位为捕捉用户的完整面部图像)。

[0029] 在训练阶段602期间,RGBD摄像机702可以捕捉完整面部图像704。类似地,可穿戴智能设备的低角度摄像机104(3)和眼睛跟踪摄像机106(3)与RGBD摄像机702同时地分别捕捉用户的图像706和708。注意,由于附图的比例,这些元件在图7中不是容易辨别的,但是在图1、图4和图5中是容易显见的。由于低角度摄像机104(3)和眼睛跟踪摄像机106(3)的有限视场,在每个摄像机104(3)和106(3)中仅见到用户的面部的一部分。具体而言,低角度摄像机104(3)如在图像706中指示的沿着面颊指向嘴部和下巴,而眼睛跟踪摄像机106(3)如由图像708所指示的指向用户的眼睛。(注意,图示出的图像为了说明目的而是代表性的,并且不是旨在是限制性的。另外,在低角度摄像机上使用的镜片的类型可以影响图像以及对用户的定向角。例如,鱼眼镜片可以按照给定定向比标准镜片捕捉更多的用户的面部。在这样的实现方式中,可以利用单个低角度摄像机104(3),其捕捉图像706中的眼睛和嘴部并且不利用两组图像(例如,不利用图像708或者捕捉它们的摄像机)。

[0030] 在又一实现方式中,其他传感器数据可以扩充来自低角度摄像机104(3)的图像数据。例如,回想图6指示超出图像(例如,视频)的各种其他形式的传感器数据614(诸如音频数据和/或用户接触数据)可以被利用。在这样的情况下,低角度摄像机104(3)可能是智能可穿戴设备上的唯一摄像机。低角度摄像机104(3)例如可捕捉用户的眼睛和面颊,而非嘴部。在这样的情况下,来自麦克风(见图1的麦克风108)的音频数据可以捕捉可以在模型建立606中使用的音频数据。例如,由用户发出的单独声音在训练阶段602期间可以被映射到与嘴部运动和/或面部表情有关的单独化身参数值。因此,音频数据可以用于某些化身参数并且来自低角度摄像机104(3)的图像数据可以用于其他化身参数。

[0031] 在例示出的配置中,在训练阶段602期间,来自可穿戴智能设备102(3)的两个摄像机104(3)和106(3)的部分面部图像流当鉴于完整面部图像704被评估时可以用来估计用户的完整面部表情。该估计可以包括关于多个面部表情参数的信息。多个面部表情参数可以用来驱动用户的动画化身710。例如,一些实现方式可以运用基于深度的面部表情跟踪算法来自动地采集用于用户的训练数据。多个机器学习算法然后可以用来构建模型以便从训练数据估计面部表情参数。在许多可能的技术之中,岭回归、多任务特征学习和/或卷积神经网络等可以被运用。

[0032] 换言之,在训练阶段602期间,由可穿戴智能设备102(3)捕捉的一组训练数据(例如,视频)以及来自RGBD摄像机702的对应的地面实况面部表情(例如,图像704)可以用来构建描述它们的关系的化身模型(图6的608)。注意,尽管在这一实现方式中利用RGBD摄像机来获得地面实况图像,但是其他实现方式可以使用其他类型的传感器或摄像机。例如,地面实况图像可以利用标准的RGB摄像机、红外摄像机和/或其他类型的传感器被获得。在部署阶段604中,模块可以用来将来自可穿戴智能设备的输入图像706(1)和708(1)(例如,视频对流)映射到可以用来驱动远程方处的化身710(1)的一组面部表情参数。在下面更详细地描述在上面相对于图6至图7介绍的特定方面。

[0033] 数据获取

[0034] 从图4至图5回想,在这一实现方式中,可穿戴智能设备102(3)的摄像机被不对称地放置(例如,低角度摄像机104(3)与面部平行地指向嘴部并且眼睛跟踪摄像机106(3)指向用户的眼睛)。不对称性的一个动机是只有一个摄像机用来捕捉冗余的面部部分(例如,注视可以主要由一个眼的位置来描述),但是诸如面颊之类的可以是不对称的面部的部分由两个摄像机来捕捉。当然,诸如对称和/或不对称配置中的三个或者更多个摄像机之类的其他配置被预期到。

[0035] 回到图7,各种方式可以用来跟踪来自RGBD摄像机输入(例如,图像704)的面部表情。例如,诸如微软的 Kinect® 化身跟踪算法之类的RGBD跟踪算法等可以被利用。一些RGBD跟踪算法实现方式可以对涉及一组个性化的网格模型 $\mathcal{B} = \{B_0, \dots, B_n\}$ 的假定进行操作,其中 B_0 是静止姿势并且 B_i ($i > 0$) 是表示静止姿势与某些面部表情之间的差异的附加位移。任意面部表情因而可以被生成为:

[0036]

$$M_j = B_0 + \sum_{i=1}^n y_{ij} B_i, \quad (1)$$

[0037] 其中 y_{ij} 是用于姿势 M_j 的混合权重。

[0038] RGBD跟踪算法可以使用多种混合形状,诸如48种混合形状,因此对于被捕捉的每种框架,其将生成48个混合权重。因此,估计面部表情的任务可以被重组为学习48个回归函数的问题,它们中的每一个将从可穿戴智能设备图像接收到的信息映射到对应的混合形状权重。

[0039] 后续讨论涉及以图像预处理开始的面部表情跟踪。

[0040] 一些实现方式可以执行用于输入图像数据的预处理以增加算法对穿着不同衣服的用户、当穿上和脱下可穿戴智能设备时的小运动以及用户之间的差异的鲁棒性。

[0041] 第一类型的预处理可以涉及皮肤过滤。这一处理可以基于三个步骤。首先,简单的

分类器可以用来估计每个像素是皮肤的概率。仅使用像素分类可能不是足够鲁棒的,因为一些假阳性像素可能被发现,当极值在背景中出现时尤其是如此。为了使该处理更加鲁棒,一些实现方式可以检测图像的边缘并且在边缘之中寻找使皮肤与背景区域分开的(潜在的)最佳的一个。

[0042] 为此,这些实现方式可以首先对整个图像应用Canny边缘(或者其他)检测器。检测器的灵敏度可以被设置为高以使得一组许多边缘被检测到,从而使得任何有用边缘被包括在该组中。最后,这些实现方式可以通过使用利用在前两个处理中采集的信息的动态编程算法来寻找(潜在的)最佳边缘。这一方法可以例如从左向右对图像的列进行迭代。相对于每个列,处理可以针对属于边缘的每个像素来计算其是(潜在的)最佳边缘的一部分的概率。这可以基于在其上方和下方的像素的皮肤概率以及在之前迭代中计算出的左邻像素的似然性被计算。

[0043] 图8示出了一个这样的示例。实例一示出了在处理之前的来自向内看的眼睛跟踪摄像机106(3)(图4至图5)的图像800(1)。这一说明必须处理难以在行图中图示的颜色。在实例一中,以交叉影线示出的区域802是具有与邻近皮肤804类似的颜色的背景并且因此可能是难以辨别的。边界806可以被检测到并且在实例二中背景区域的颜色被改变为对比色(在这种情况下为由对角线表示的白色),如在经处理的图像800(2)中的808处指示的。

[0044] 图9示出了由一些实现方式使用的第二类型的预处理的示例。这一第二类型的预处理可以处理小摄像机运动或者不同用户之间的差异。在这样的情况下,基于模版匹配使所有图像(在这种情况下为左图像900(1)和右图像900(2))对齐的对准(registration)方法可以被使用。分别在左摄像机图像900(1)和右摄像机图像900(2)中搜索鼻一面颊皱纹902和外右眼角904,并且相应地调节输入图像以减少图像之间的对准误差。如在实验结果中示出的,这一对准步骤对于训练新用户而言可以是极有帮助的。

[0045] 在预处理之后,两种形式的输入图像可以被利用:原始像素描述和局部二值模式(LBP)描述。

[0046] 为了说明的目的,通过矩阵 $X \in \mathbb{R}^{d \times N}$ 来表示训练输入数据,该矩阵包含N个训练实例,这些训练实例具有维度d以及 $Y \in \mathbb{R}^{N \times 48}$ 个训练标签(从RGBD跟踪算法获得的混合形状权重)。 y_t 可以用来表示Y的第t列,即,用于第t个混合形状的地貌实况权重。鉴于机器学习问题可以存在于多输出回归中,因此三种机器学习算法(岭回归、多任务特征学习和卷积神经网络)已被运用。当然,在其他实现方式中可以使用其他机器学习算法。

[0047] 岭回归

[0048] 岭回归(Ridge Regression,RR)可以通过求解以下优化问题来学习用于每个输出的线性模型:

$$[0049] \underset{w_t}{\operatorname{argmin}} \|X^T w_t - y_t\|_2^2 + \lambda \|w_t\|_2^2, \quad (2)$$

[0050] 其中 w_t 是用于混合形状t的d维权重向量。注意,以上问题针对每个混合形状被独立地求解。

[0051] 多任务特征学习

[0052] 考虑到混合形状权重一起用来描述面部表情,并且这些混合形状是非正交的,因

此它们可以被预计是线性相关的。作为结果,可以针对当前任务运用多任务特征学习(MTFL)。MTFL可以通过同时学习所有任务来设法利用在混合形状权重之间的共性。MTFL可以尝试求解以下优化问题:

$$\begin{aligned} [0053] \quad & \underset{\mathcal{C}, D}{\operatorname{argmin}} \quad \| X^T D C - Y \|_{Fro}^2 \\ & + \frac{\lambda}{2} (\| C \|_{Fro}^2 + \| D \|_{Fro}^2) \end{aligned} \quad , \quad (3)$$

[0054] 其中 $\|\cdot\|_{Fro}^2$ 表示矩阵的Frobenius范数的平方,即,它的元素的平方的总和。如在图10中示出的,对于通过将输入乘以矩阵D形成的所有任务可以存在共有层,其表示共享的特征1002。然后通过将共享的特征乘以矩阵C来形成输出层1004。两个层都是线性的。

[0055] 方程(3)中的问题是非凸的。相反,该过程可以优化由以下给出的它的等价凸问题:

$$[0056] \quad \underset{W}{\operatorname{argmin}} \quad \| X^T W - Y \|_{Fro}^2 + \lambda \| W \|_{Tr}, \quad (4)$$

[0057] 其中 $W = DC$; 并且 $\|\cdot\|_{Tr}$ 表示矩阵的迹范数,也即,其奇异值的总和。

[0058] 神经网络

[0059] 图11A和图11B示出了两个卷积神经网络1100 (1) 和1100 (2)。第一卷积神经网络1100 (1) 在卷积层1104 (1) 处接收输入图像1102 (1)。在这一示例中,卷积层1104 (1) 具有尺寸为3x3的16个内核。卷积层的输出是丢弃(dropout,简称DO)使能的并且被完全连接到具有50个节点的层1106 (1)。来自50节点层的输出然后被发送给具有48个节点的输出层1108 (1),其产生最终输出1110。

[0060] 第二卷积神经网络1100 (2) 包括第一层,其也是具有尺寸为3x3的16个内核的卷积层1104 (2)。第一卷积层1104 (2) 接收输入图像1102 (2)。第一卷积层的输出被密集地连接到第二隐藏层1106 (2)。在这一示例中,第二隐藏层具有200个隐藏单元。第二隐藏层的输出被引导至第三层1108 (2)。在这一示例中,第三层是全连接层,其也具有200个输出节点。在这一实例中,最终层1112接收第三层1108 (2) 的输出以产生最终输出1114。第三层1108 (2) 可以是具有48个输出的全连接层。在一些实现方式中,修正线性单元(ReLU)非线性被在除输出层之外的所有卷积和全连接层处被使用。此外,丢弃在一些全连接层处被使能。

[0061] 总之,以上描述说明了如何从以不寻常朝向(诸如与用户的面部几乎平行)被放置的可穿戴摄像机(例如,可穿戴智能设备的摄像机)执行面部表情跟踪。地面实况数据可以与可穿戴智能设备的摄像机同时地使用深度摄像机被获得。基于深度的面部表情跟踪算法可以用来使地面实况数据中的来自两个摄像机的图像信息相关。地面实况数据可以用来训练与面部参数相关联的回归模型。结果,来自可穿戴智能设备的图像可以被用回归模型处理以得出用于控制用户的化身的参数。

[0062] 图12示出了可视化系统1200。为了说明的目的,系统1200包括表现为智能眼镜的两个可穿戴智能设备102 (3) (A) 和102 (3) (B)。这些被图示的可穿戴智能设备表示两种设备配置,其中的任一种可被应用于除智能眼镜之外的其他可穿戴智能设备。简言之,可穿戴智能设备102 (3) (A) 表示操作系统中心配置并且可穿戴智能设备102 (3) (B) 表示片上系统配置。下面将更详细地描述这些方面。系统1200还包括不是可穿戴智能设备的设备1202。在这

种情况下,设备1202表现为智能电话,但是旨在表示可穿戴智能设备可与之通信的任何类型的设备。系统1200还包括可穿戴智能设备可经由一个或多个网络1206与之通信的诸如基于云的资源之类的远程资源1204。设备1202和远程资源1204中的任一个或两者可以与可穿戴智能设备协同工作以实现本概念。可穿戴智能设备102(3)(A)和102(3)(B)、设备1202和/或远程资源1204中的任一个可以支持或包括化身控制组件1208。下面将在对可穿戴智能设备102(3)(A)和102(3)(B)的其他元件的讨论之后更详细地描述化身控制组件。

[0063] 为了说明的目的,可穿戴智能设备102(3)(A)被组织成应用1210、操作系统1212和硬件1214。可穿戴智能设备102(3)(B)被组织成共享的资源1216、专用的资源1218以及其间的接口1220。

[0064] 可穿戴智能设备102(3)(A)和102(3)(B)还可以包括定向结构110,其可以包括框架1222并且与镜脚1224连接。镜脚可以沿着(与x基准轴平行的)长轴延伸并且终止于耳架1226。框架可以支撑(一对)镜片1228。可穿戴智能设备还可以包括存储装置1230、处理器1232、电池1234(或者其他电源)、传感器1236和/或通信组件1238。可穿戴智能设备备选地或者附加地可以包括其他元件,诸如输入/输出设备、总线、图形卡(例如,图形处理单元(GPU))等,其为了简明而未在这里被图示或讨论。在一种配置中,电池1234可以被放置在耳架1226中的一个中,而通信组件1238、存储装置1230和处理器1232被放置在另一个中。

[0065] 在这一示例中,传感器1236表现为低角度摄像机104、眼睛跟踪摄像机106和麦克风108。注意,如在这里使用的,摄像机可以被认为是可以捕捉关于用户特征的信息的任何类型的传感器。摄像机可以是可见光或者其他波长的。摄像机可以运用不同类型的镜片。例如,摄像机可以运用凸镜片、非球面镜片和/或鱼眼镜片等。鱼眼透镜可以允许单个摄像机捕捉否则可能需要使用两个摄像机来捕捉的用户的面部的大部分。上面在图1处开始描述了关于一个或多个摄像机的附加细节。当然,尽管摄像机被详细描述,但是其他实现方式备选地或者附加地可以使用其他类型的传感器1236。

[0066] 眼镜镜片1228可以是视力矫正或者非视力矫正的。在一些实现方式中,镜片可以具有显示图像以供用户查看的能力。显示可以通过投影或者通过将发光二极管或其他组件包括在镜片内被完成。

[0067] 通信组件1238可以包括接收器和发射器和/或用于与各种技术(诸如蜂窝、Wi-Fi(IEEE 802.xx)、蓝牙等)通信的其他射频电路。

[0068] 化身控制组件1208可以接收来自传感器1236的输入。化身控制组件1208可以利用化身模型来从传感器数据得出动画参数以控制用户的化身。单独可穿戴智能设备102(3)上的化身控制组件1208可以是相对鲁棒的,从而使得其处理传感器数据以确定动画参数的值。例如,可穿戴智能设备102(3)(A)可以从传感器1236(1)确定动画参数值。可穿戴智能设备102(3)(A)可以动画制作用户的化身并将动画发送给另一设备或者其可以将动画参数值发送给另一设备以用于执行。

[0069] 例如,相对于可穿戴智能设备102(3)(A),化身控制组件1208(1)可以将动画参数值发送给通信组件1238(1)以用于传送到诸如可穿戴智能设备102(3)(B)之类的远程设备。接收设备的化身控制组件1208(2)可以基于接收到的动画参数值来控制用户的化身。同时,远程设备的化身控制组件1208(2)可以正在接收来自传感器1236(2)的数据、处理传感器数据以确定动画参数值并且将动画制作参数值发送给可穿戴智能设备102(3)(A)。

[0070] 在其他配置中,可穿戴智能设备102(3)(A)可能是更不鲁棒的并且可能将传感器数据中的一些传感器数据发送给设备1202或远程资源1204以用于分别由化身控制组件1208(3)或1208(4)处理。这些化身控制组件1208(3)或1208(4)可以代表可穿戴智能设备102(3)(A)确定动画参数值并且代表可穿戴智能设备102(3)(A)将它们传送给可穿戴智能设备102(3)(B)。

[0071] 在又一种配置中,可穿戴智能设备102(3)(A)可以将动画参数值发送给设备1202。设备1202可以捕捉用户的‘正常’视频流(例如,在视频通话中)并且将该视频流发送给可穿戴智能设备102(3)(A)。因此,可穿戴智能设备102(3)(A)的用户将看见设备1202的用户的正常视频而设备1202的用户将看到可穿戴智能设备102(3)(A)的用户的动画化身。

[0072] 从一个角度,可穿戴智能设备102(3)(A)和102(3)(B)、设备1202和/或远程资源1204可以被看作计算机。

[0073] 如在此所使用的术语“设备”、“计算机”或“计算设备”可意味着具有某种量的处理能力和/或存储能力的任何类型的设备。处理能力可由一个或多个处理器提供,处理器可执行计算机可读指令形式的数据以提供功能。数据(诸如计算机可读指令和/或用户相关数据)可被存储在存储装置上,诸如对计算机而言可以是内部或外部的存储装置。存储装置可包括易失性或非易失性存储器、硬盘驱动器、闪存存储设备和/或光学存储设备(例如,CD、DVD等)、远程存储(例如,基于云的存储)等的任何一个或多个。如在此所使用的,术语“计算机可读介质”可包括信号。相比之下,术语“计算机可读存储介质”排除信号。计算机可读存储介质包括“计算机可读存储设备”。计算机可读存储设备的示例包括易失性存储介质(诸如RAM)和非易失性存储介质(诸如硬盘驱动器、光盘和闪存等等)。

[0074] 设备的示例可包括传统的计算设备,诸如个人计算机、台式计算机、笔记本计算机、蜂窝电话、智能电话、个人数字助理、平板型计算机、数字白板、摄像机、可穿戴设备(诸如智能眼镜)或任何不断演进或尚有待于开发的类型的计算设备的混合中的任一种。

[0075] 如在上面提到的,可穿戴智能设备102(3)(B)可被认为是片上系统(SOC)型设计。在这种情况下,由设备提供的功能可被集成在单个SOC或多个耦合的SOC上。一个或多个处理器可被配置成与共享的资源(诸如存储器、存储装置等)协调,和/或与一个或多个专用的资源(诸如被配置成执行某一特定功能的硬件块)协调。因此,如在此使用的术语“处理器”还可指代中央处理单元(CPU)、图形处理单元(GPU)、控制器、微控制器、处理器核、或其他类型的处理设备。

[0076] 一般而言,在此描述的任何功能都可以使用软件、固件、硬件(例如,固定逻辑电路)、人工处理或这些实现方式的组合被实现。在此使用的术语“组件”一般表示软件、固件、硬件、整个设备或网络或其组合。在软件实现方式的情况下,例如,这些可表示当在处理器(例如,一个或多个CPU)上执行时执行指定任务的程序代码。程序代码可被存储在诸如计算机可读存储介质之类的一个或多个计算机可读存储器设备中。组件的特征和技术是与平台无关的,从而意味着它们可以在具有各种处理配置的各种商用计算平台上被实现。

[0077] 注意,在当前实现方式中的任意实现方式中,可以通过仅在用户给出其明确同意之后使能可视化特征来保护用户的隐私。所有的隐私和安全过程都可被实现以保护用户。例如,用户可提供授权(和/或限定授权条件)。可穿戴智能设备仅根据授权的条件继续进行面部跟踪用户。否则,用户信息不被收集。类似地,用户可被允许限定包括可视化数据的其

数据的用途。可视化数据的任何用途必须与限定的用户条件一致。

[0078] 结论

[0079] 当前实现方式可以从仅捕捉用户的面部的部分视图的低角度摄像机(和/或其他传感器)得出有用信息。当前实现方式可以使用机器学习算法基于在深度传感器(诸如由微软®公司提供的Kinect®牌深度传感器)的帮助下采集的训练数据来估计面部表情参数。当前实现方式即使从摄像机的极有限视角也可以可靠地跟踪人的面部表情。

[0080] 所描述的方法或处理可以由在上面描述的系统和/或设备执行,和/或由其他设备和/或系统执行。描述方法的次序并不旨在被解释为限制,并且任何数量的所描述动作都可以按任何次序被组合以实现本方法或备选方法。此外,该方法可以用任何合适的硬件、软件、固件或其组合被实现,从而使得设备可实现该方法。在一种情况下,该方法作为指令集被存储在计算机可读存储介质上,从而使得计算设备的处理器的执行使得该计算设备执行该方法。

[0081] 尽管已用特定于结构特征和/或方法动作的语言描述了涉及可视化信息的技术、方法、设备、系统等,但是,将会明白,所附权利要求书中定义的主题不一定限于所描述的具体特征或动作。相反,具体特征和动作作为实现所要求保护的方法、设备、系统等的示例性形式而被公开。

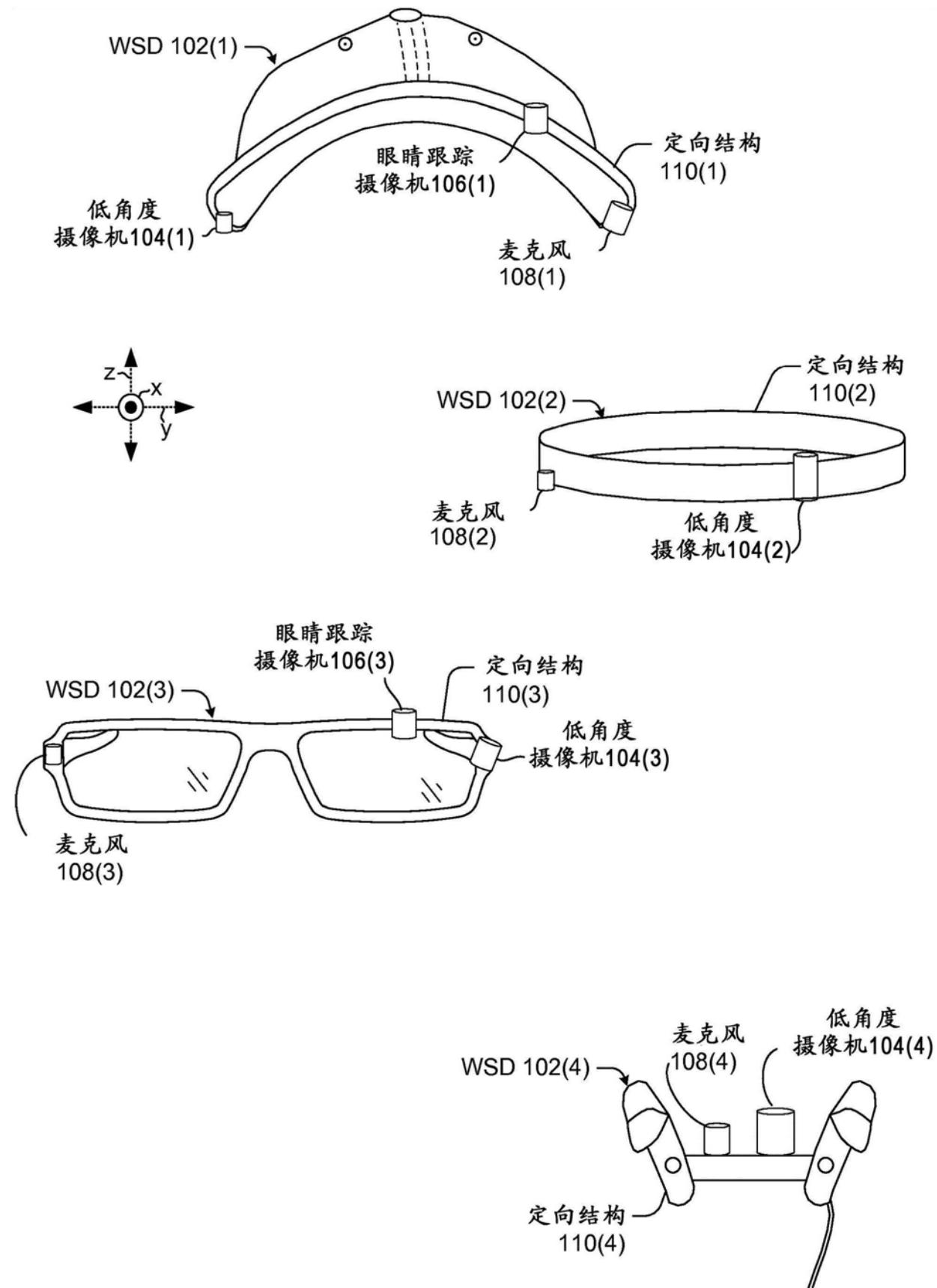


图1

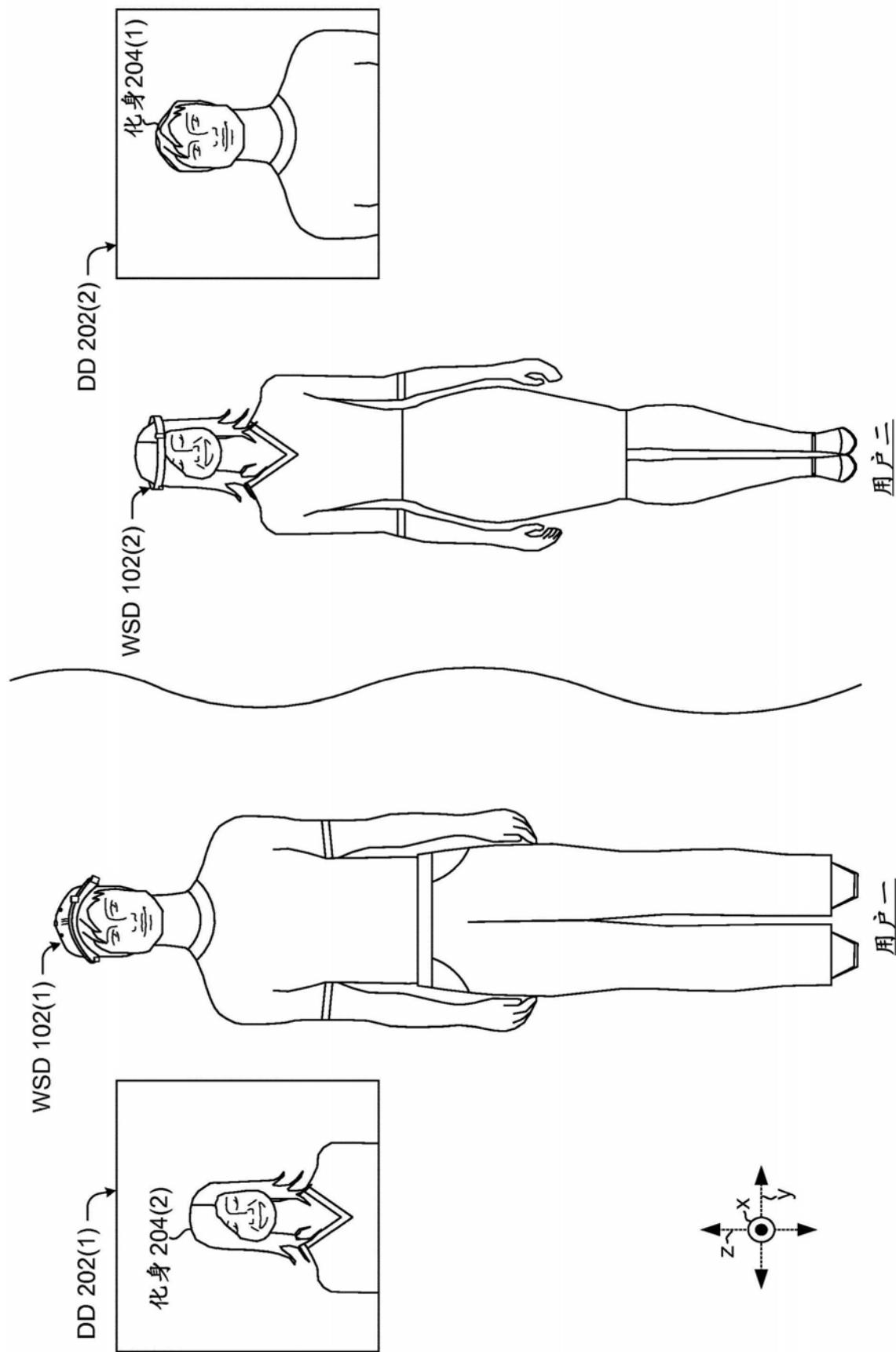


图2

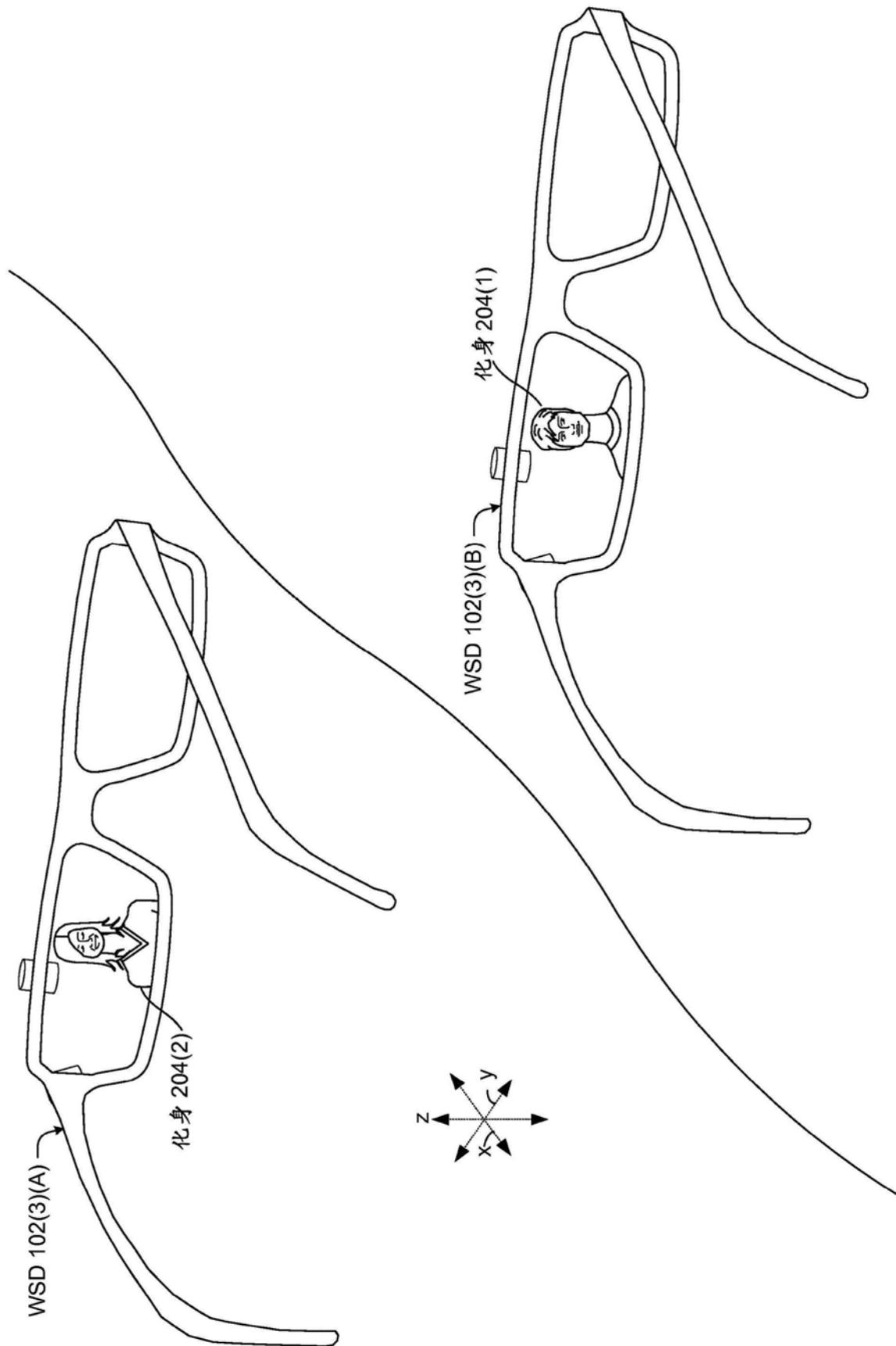


图3

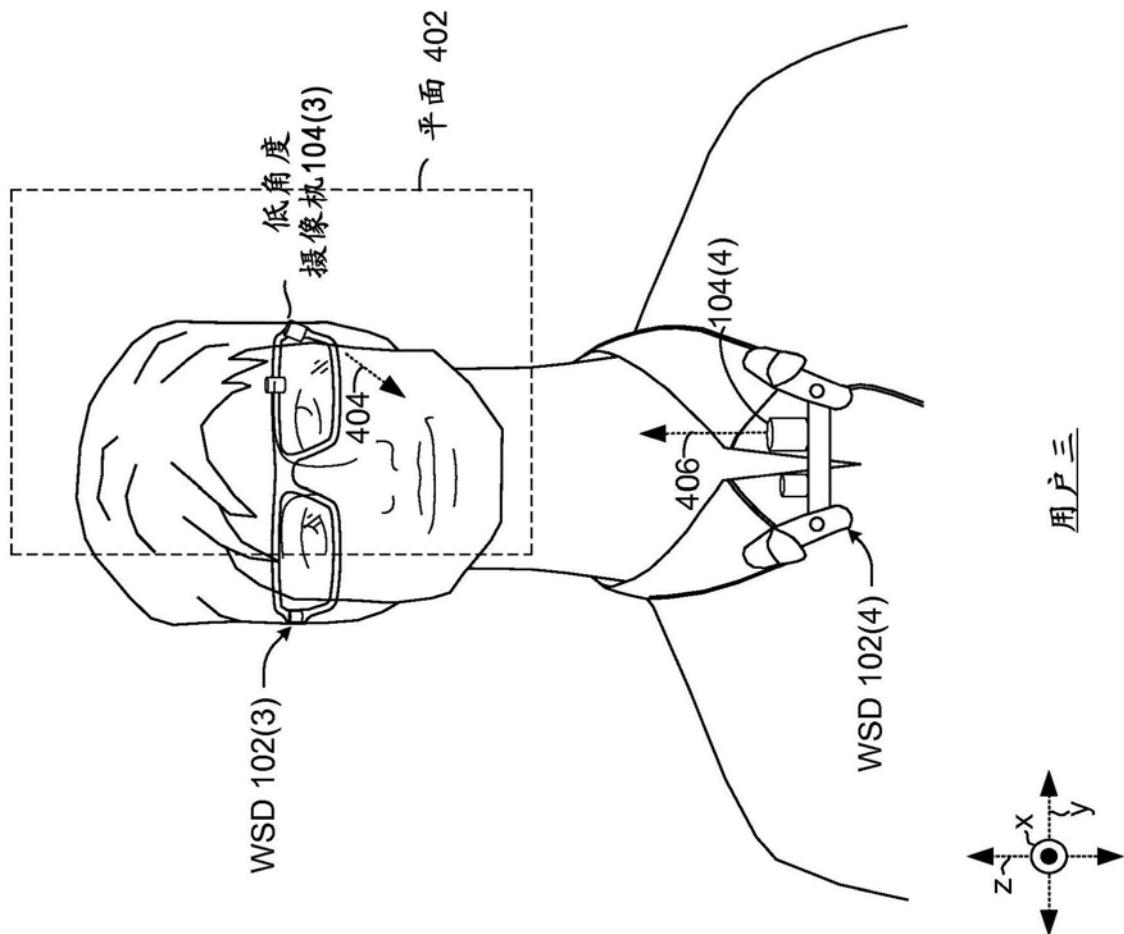


图4

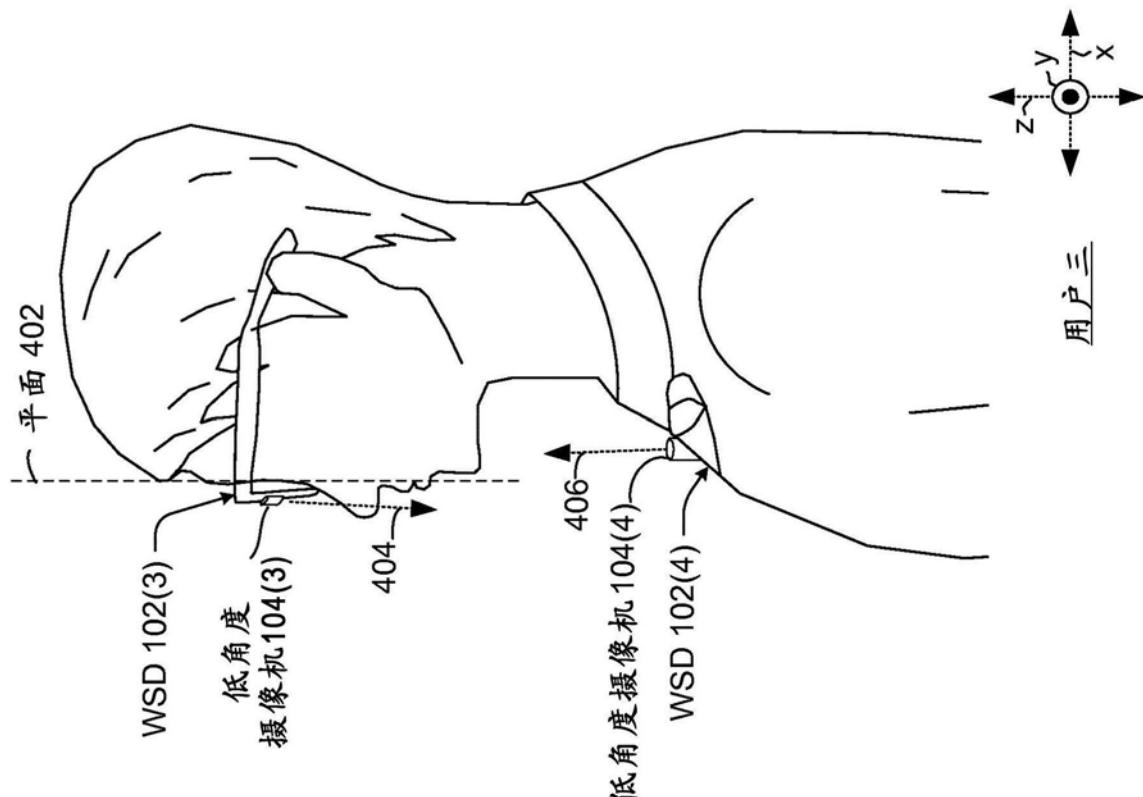


图5

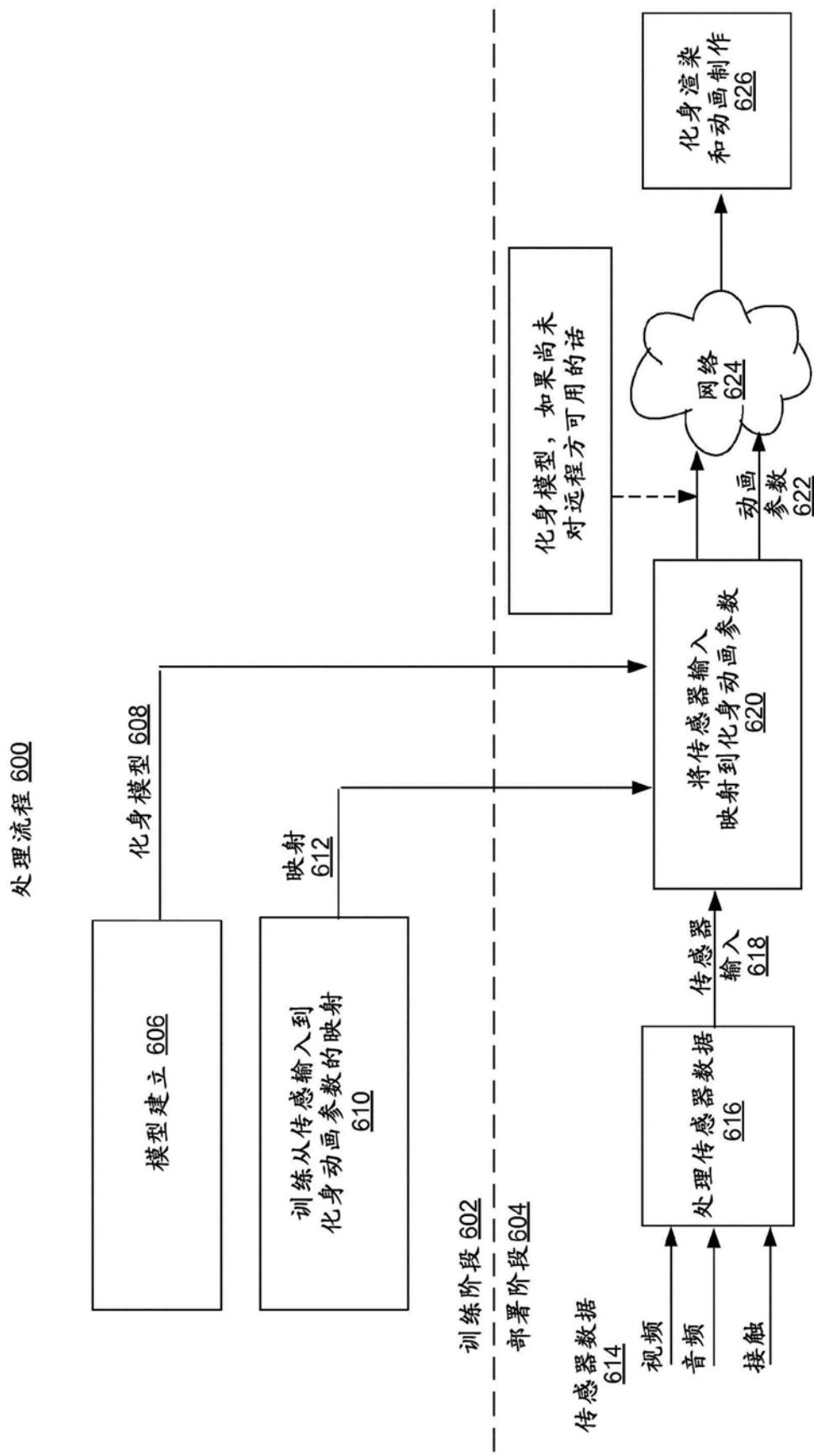


图6

系统 700

输入 701(1)

输出 701(2)

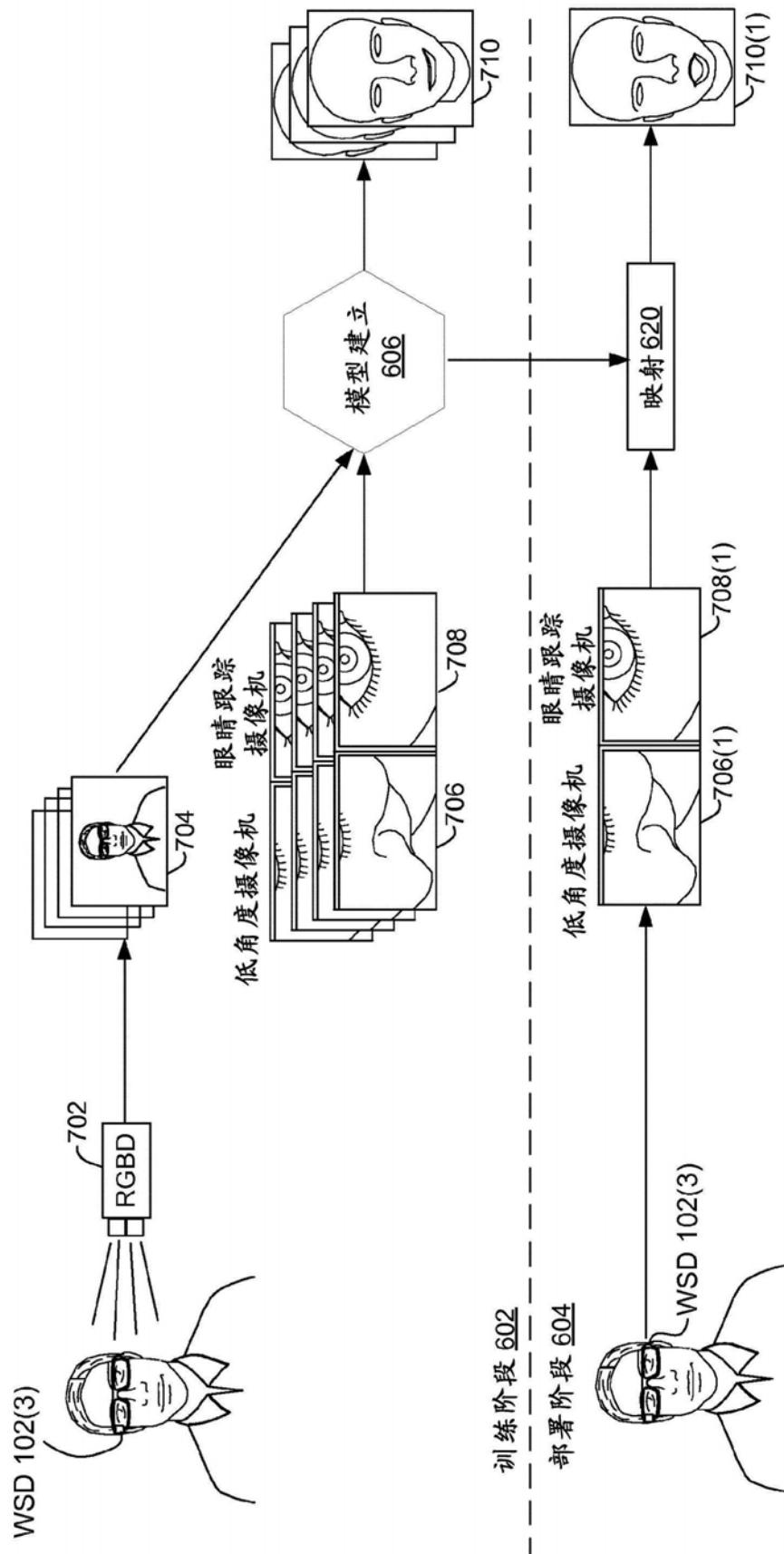


图7

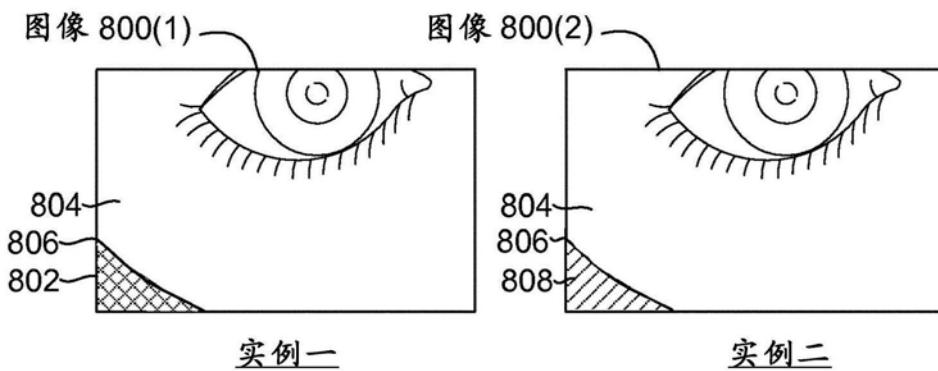


图8

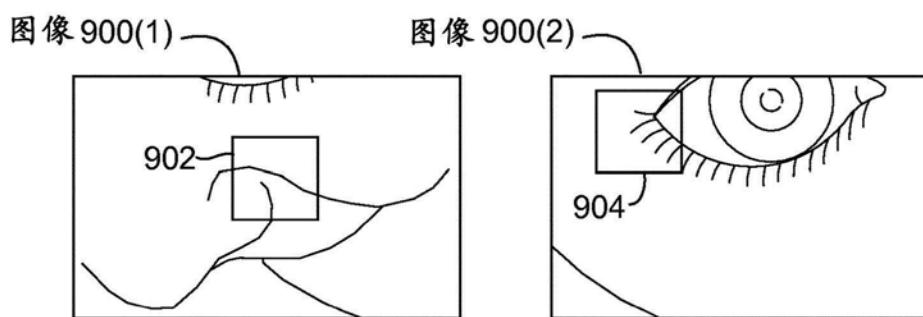


图9

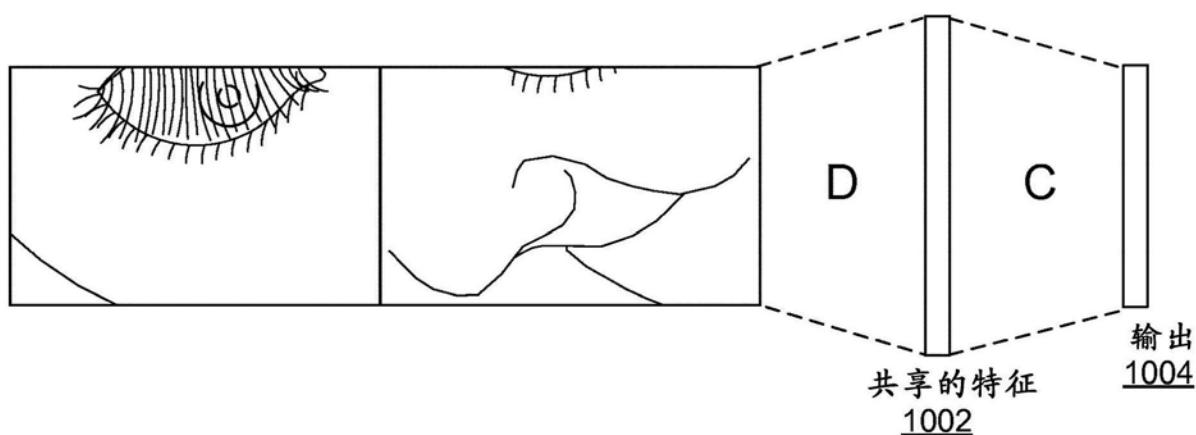


图10

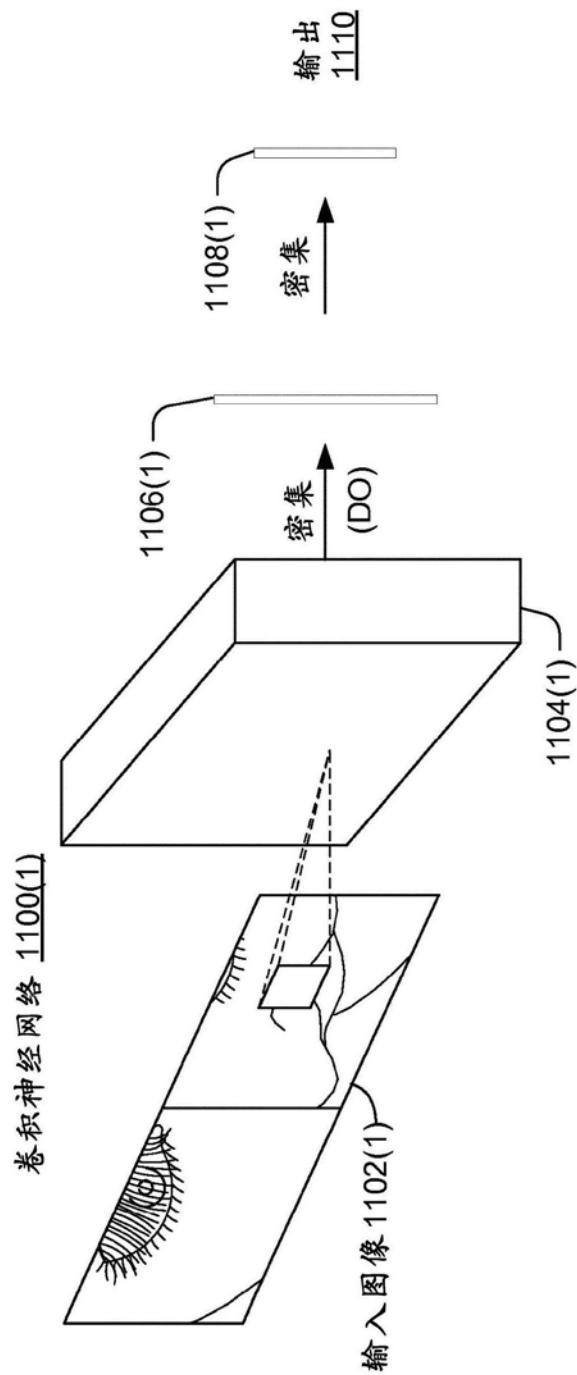


图11A

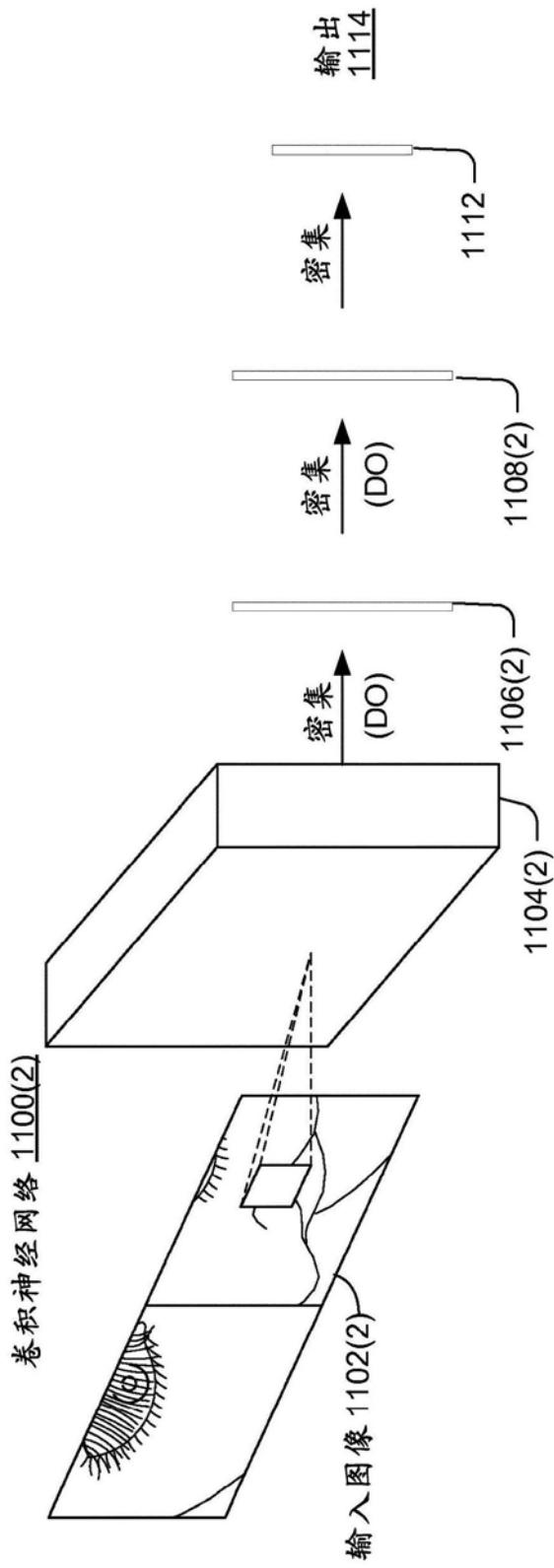


图11B

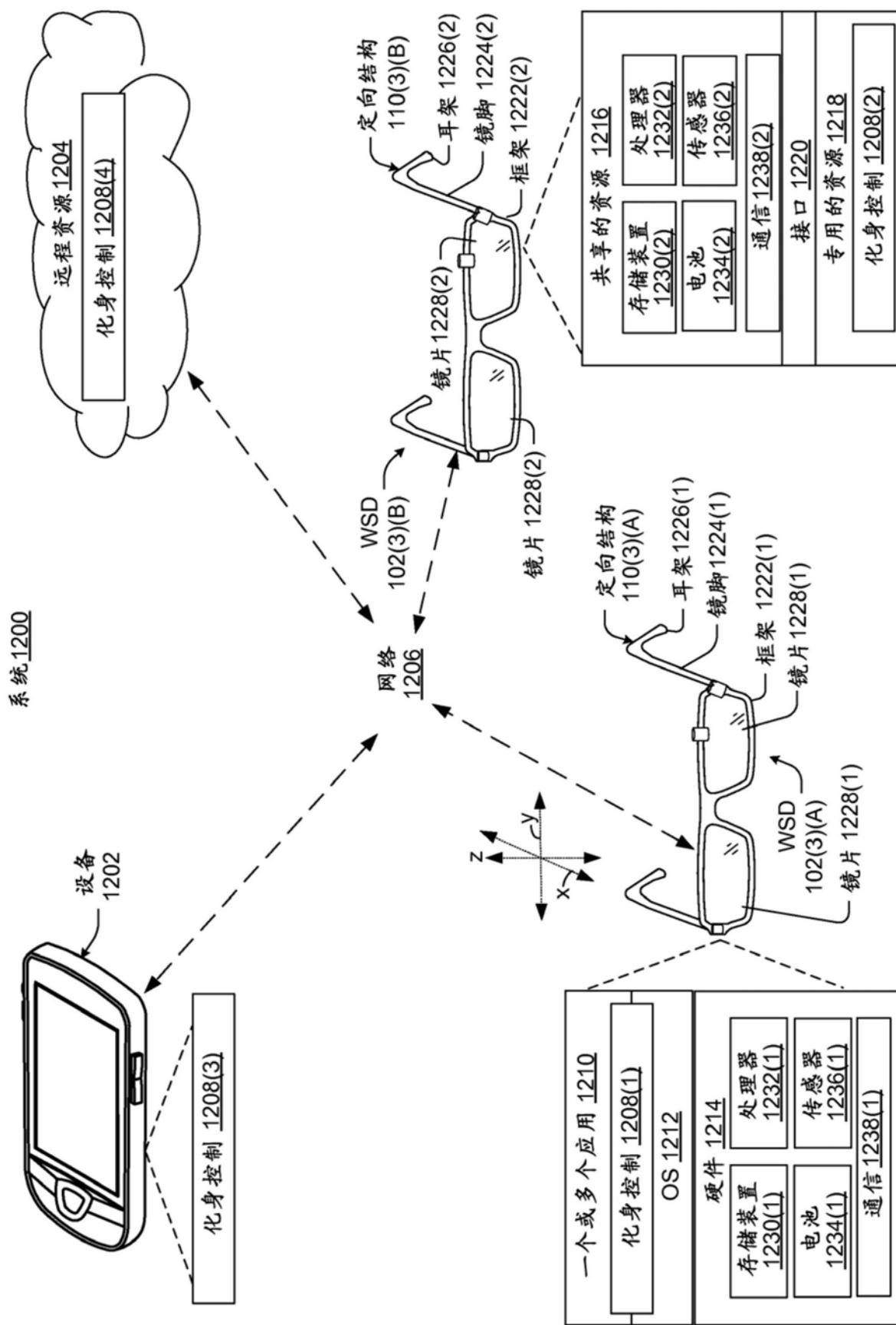


图12