



(12) 发明专利申请

(10) 申请公布号 CN 104170318 A

(43) 申请公布日 2014. 11. 26

(21) 申请号 201280071927. 3

H04L 12/853(2006. 01)

(22) 申请日 2012. 04. 09

(85) PCT国际申请进入国家阶段日  
2014. 09. 26

(86) PCT国际申请的申请数据  
PCT/CN2012/000458 2012. 04. 09

(87) PCT国际申请的公布数据  
W02013/152453 EN 2013. 10. 17

(71) 申请人 英特尔公司  
地址 美国加利福尼亚州

(72) 发明人 X. 童 李文龙 杜杨洲 W. 胡  
Y. 张

(74) 专利代理机构 中国专利代理(香港)有限公司  
72001  
代理人 张金金 姜甜

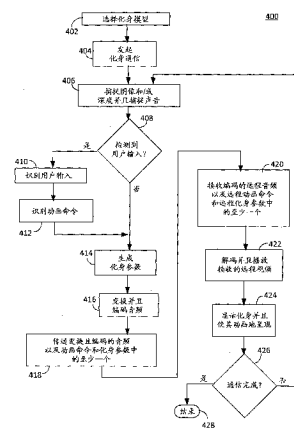
(51) Int. Cl.  
H04L 12/24(2006. 01)  
H04L 12/70(2006. 01)

权利要求书3页 说明书13页 附图5页

(54) 发明名称  
使用交互化身的通信

(57) 摘要

大体上本公开描述视频通信系统,其用动画呈现的化身替换参与用户的实际实况图像。方法可包括:选择化身;发起通信;检测用户输入;识别用户输入;基于用户输入识别动画命令;生成化身参数;以及传送动画命令和化身参数中的至少一个。



1. 一种系统,其包括:
  - 用户输入设备,其配置成捕捉用户输入;
  - 通信模块,其配置成传送以及接收信息;以及
  - 一个或多个存储介质,其上独立地或组合地存储有指令,所述指令在由一个或多个处理器执行时促成下列操作,其包括:
    - 选择化身;
    - 发起通信;
    - 检测用户输入;
    - 识别所述用户输入;
    - 基于所述用户输入识别动画命令;
    - 生成化身参数;以及
    - 传送所述动画命令和所述化身参数至少其中之一。
2. 如权利要求 1 所述的系统,其进一步包括:
  - 麦克风,其配置成捕捉声音并且将捕捉的声音转换成对应的音频信号,其中所述指令在由一个或多个处理器执行时促成下列额外操作:
    - 捕捉用户语音并且将所述用户语音转换成对应的用户语音信号;
    - 将所述用户语音信号变换成化身语音信号;以及
    - 传送所述化身语音信号。
3. 如权利要求 1 所述的系统,其包括拍摄装置,所述拍摄装置配置成捕捉图像,其中所述指令在由一个或多个处理器执行时促成下列额外操作:
  - 捕捉图像;
  - 检测所述图像中的面部;
  - 从面部提取特征;以及
  - 将所述特征转换成化身参数。
4. 如权利要求 1 所述的系统,其进一步包括显示器,其中所述指令在由一个或多个处理器执行时促成下列额外操作:
  - 显示至少一个化身;
  - 接收远程动画命令和远程化身参数中的至少一个;以及
  - 基于所述远程动画命令和所述远程化身参数中的至少一个使一个显示的化身动画地呈现。
5. 如权利要求 1 所述的系统,其进一步包括扬声器,所述扬声器配置成将音频信号转换成声音,其中所述指令在由一个或多个处理器执行时促成下列额外操作:
  - 接收远程化身语音信号;以及
  - 将所述远程化身语音信号转换成化身语音。
6. 如权利要求 1 至 5 中任一项所述的系统,其中所述用户输入设备是拍摄装置,其配置成捕捉距离,并且所述用户输入是手势。
7. 如权利要求 1 至 5 中任一项所述的系统,其中所述用户输入设备是触敏显示器并且所述用户输入是触摸事件。
8. 如权利要求 2 至 5 中任一项所述的系统,其中所述变换包括移调和/或时间延展中的至

少一个。

9. 一种方法,其包括:

选择化身;

发起通信;

检测用户输入;

识别所述用户输入;

基于所述用户输入识别动画命令;

生成化身参数;以及

传送所述动画命令和所述化身参数至少其中之一。

10. 权利要求 9 所述的方法,其进一步包括:

捕捉用户语音并且将所述用户语音转换成对应的用户语音信号;

将所述用户语音信号变换成化身语音信号;以及

传送所述化身语音信号。

11. 权利要求 9 所述的方法,其进一步包括:

捕捉图像;

检测所述图像中的面部;

从面部提取特征;以及

将所述特征转换成化身参数。

12. 如权利要求 9 所述的方法,其进一步包括:

显示至少一个化身;

接收远程动画命令和远程化身参数中的至少一个;以及

基于所述远程动画命令和远程化身参数中的至少一个使一个显示的化身动画地呈现。

13. 如权利要求 9 所述的方法,其进一步包括:

接收远程化身语音信号;以及

将所述远程化身语音信号转换成化身语音。

14. 如权利要求 9 至 13 中任一项所述的方法,其中所述用户输入是手势。

15. 如权利要求 9 至 13 中任一项所述的方法,其中所述用户输入是触摸事件。

16. 如权利要求 10 至 13 中任一项所述的方法,其中所述变换包括移调和/或时间延展中的至少一个。

17. 一种系统,其包括一个或多个存储介质,所述一个或多个存储介质在其上独立地或组合地存储有指令,所述指令在由一个或多个处理器执行时促成下列操作,其包括:

选择化身;

发起通信;

检测用户输入;

识别所述用户输入;

基于所述用户输入识别动画命令;

生成化身参数;以及

传送所述动画命令和所述化身参数至少其中之一。

18. 如权利要求 17 所述的系统,其中所述指令在由一个或多个处理器执行时促成下列

额外操作：

捕捉用户语音并且将所述用户语音转换成对应的用户语音信号；  
将所述用户语音信号变换成化身语音信号；以及  
传送所述化身语音信号。

19. 如权利要求 17 所述的系统，其中所述指令在由一个或多个处理器执行时促成下列额外操作：

捕捉图像；  
检测所述图像中的面部；  
从面部提取特征；以及  
将所述特征转换成化身参数。

20. 如权利要求 17 所述的系统，其中所述指令在由一个或多个处理器执行时促成下列额外操作：

显示至少一个化身；  
接收远程动画命令和远程化身参数中的至少一个；以及  
基于所述远程动画命令和远程化身参数中的至少一个使一个显示的化身动画地呈现。

21. 如权利要求 17 所述的系统，其中所述指令在由一个或多个处理器执行时促成下列额外操作：

接收远程化身语音信号；以及  
将所述远程化身语音信号转换成化身语音。

22. 如权利要求 17 至 21 中任一项所述的系统，其中所述用户输入是手势。

23. 如权利要求 17 至 21 中任一项所述的系统，其中所述用户输入是触摸事件。

24. 如权利要求 18 至 21 中任一项所述的系统，其中所述变换包括移调和/或时间延展中的至少一个。

## 使用交互化身的通信

### 技术领域

[0001] 下列公开涉及视频通信,并且更特定地涉及使用交互化身的视频通信。

### 背景技术

[0002] 在移动设备中可用的种类日益增加的功能性已经对用户造成除简单电话外还经由视频来通信的期望。例如,用户可发起“视频通话”、“视频会议”等,其中设备中的拍摄装置和麦克风捕捉用户的音频和视频,其实时被传送到一个或多个其他接收者,例如其他移动设备、台式计算机、视频会议系统等。视频的通信可牵涉传送大量数据(例如,根据拍摄装置的技术、处理捕捉的图像数据所采用的特定视频编解码器,等)。考虑到现有 2G/3G 无线技术的带宽限制,和新兴 4G 无线技术的仍然有限的带宽,实施并发视频通话的许多设备用户可超出现有无线通信基础设施中的可用带宽,这可能负面影响视频通话的质量。

### 附图说明

[0003] 要求保护的主旨的各种实施例的特征和优势将随着下列详细说明进行以及在参考图(其中类似的数字指示类似的部件)时变得明显,并且其中:

图 1A 图示根据本公开的各种实施例的示例的设备到设备系统;

图 1B 图示根据本公开的各种实施例的示例虚拟空间系统;

图 2 图示根据本公开的各种实施例的示例设备;

图 3 图示根据本公开的至少一个实施例的示例系统实现;以及

图 4 是根据本公开的至少一个实施例的示例操作的流程图。

[0004] 尽管下列详细说明将参考说明性实施例而进行,其许多备选、修改和变化对于本领域内技术人员将是明显的。

### 具体实施方式

[0005] 一般,该公开描述用于使用交互化身的视频通信的系统和方法。使用化身(如与实况图像不同)使要传送的数据的量大大减少,并且从而,化身通信需要较少的带宽。交互化身配置成通过基于用户输入修改选择的化身的显示来增强用户体验。此外,可捕捉用户语音并且将其变换来产生化身语音。化身语音然后可与用户语音相关,但可掩饰用户的身份。音频变换可包括,例如移调(pitch shifting)和/或时间延展。

[0006] 在一个实施例中,在耦合于拍摄装置、麦克风和扬声器的设备中激活应用。该应用可配置成允许用户选择化身用于在远程设备上、虚拟空间中等显示。设备然后可配置成发起与至少一个其他设备、虚拟空间等的通信。例如,通信可通过 2G、3G、4G 蜂窝连接而建立。备选地或另外,通信可经由 WiFi 连接通过因特网建立。在建立通信后,拍摄装置可配置成开始捕捉图像和/或到物体的距离并且麦克风可配置成开始捕捉声音(例如,用户语音),并且将用户语音转换成用户语音信号。

[0007] 然后可确定是否检测到用户输入。该用户输入可被用户输入设备捕捉。用户输入

包括由触敏显示器捕捉的触摸事件和由拍摄装置(例如,配置成捕捉到物体的距离的深度拍摄装置和/或web拍摄装置)捕捉的手势。从而,用户输入设备包括触敏显示器和/或拍摄装置。如果检测到用户输入,可识别该用户输入。对于触摸事件,用户输入标识符可与触摸类型以及一个或多个触摸位点有关。对于手势(例如,张开的手),用户输入标识符可与手势标识符有关。然后可基于用户输入识别动画命令。动画命令对应于与用户输入关联的期望响应,例如响应于显示的化身的面部上的单击而改变显示的化身的面部的颜色。

[0008] 然后可生成化身参数。化身参数可基于面部检测、头部移动和/或动画命令而生成。化身参数从而可包括基于例如面部检测和头部移动的被动组成,和基于动画命令的交互组成。化身参数可能用于使化身在至少一个其他设备上、在虚拟空间内等动画地呈现。在一个实施例中,化身参数可基于面部检测、头部移动和动画命令而生成。在该实施例中,所得的动画包括基于面部检测的被动动画和基于动画命令由交互动画修改的头部移动。从而,化身动画可包括基于例如面部检测和头部移动的被动动画,和基于用户输入的交互动画。

[0009] 然后可传送动画命令和化身参数中的至少一个。在一个实施例中,接收远程动画命令和远程化身参数中的至少一个。该远程动画命令可促使设备基于远程动画命令确定化身参数以便使显示的化身动画地呈现。远程化身参数可促使设备基于接收的远程化身参数使显示的化身动画地呈现。

[0010] 音频通信可伴随着化身动画。在建立通信后,麦克风可配置成捕捉音频输入(声音)(例如用户语音),并且将捕捉的声音转换成对应的音频信号(例如,用户语音信号)。在实施例中,用户语音信号可变换成化身语音信号,其然后可被编码和传送。接收的化身语音信号然后可由扬声器转换回声音(例如,化身语音)。化身语音从而可基于用户语音并且可保存内容但可更改与捕捉的语音关联的光谱数据。例如,变换包括但不限于,移调、时间延展和/或转换重放速率。

[0011] 用户输入设备(例如,触敏显示器和/或拍摄装置)可配置成捕捉用户输入,其配置成基于至少一个其他设备上的用户输入使化身动画地呈现。用户驱动的动画(基于动画命令)可以是基于面部表情和/或头部移动的动画的补充。动画命令可包括但不限于,化身的显示取向中的改变、面部特征失真、改变特征来表达情感,等。与基于面部检测/跟踪的动画相似或作为对其的补充,动画命令从而可修改化身动画。动画命令可导致有时间限制的动画并且可基于来自远程用户的输入,其中所得的动画在本地用户的显示化身上图示。

[0012] 从而,有限带宽视频通信系统可使用化身来实现。音频可被变换并且视频可基于检测的用户输入和识别的动画命令而动画地呈现来增强用户对化身通信的体验。此外,匿名可使用化身而得到保持,其包括如本文描述的音频变换。

[0013] 图1A图示与本公开的各种实施例一致的设备到设备系统100。系统100一般可包括经由网络122而通信的设备102和112。设备102包括至少拍摄装置104、麦克风106、扬声器107和触敏显示器108。设备112包括至少拍摄装置114、麦克风116、扬声器117和触摸显示器118。网络122包括至少服务器124。

[0014] 设备102和112可包括能够有线和/或无线通信的各种硬件平台。例如,设备102和112可包括但不限于,视频会议系统、台式计算机、便携式计算机、平板计算机、智能电话(例如,iPhones®、基于Android®的电话、Blackberries®、基于Symbian®的电话、基于Palm®

的电话,等)、蜂窝手持机等。拍摄装置 104 和 114 包括用于捕捉代表环境(其包括一个或多个个人)的数字图像的任何设备,并且可具有足够的分辨率用于面部分析和 / 或手势识别,如本文描述的。例如,拍摄装置 104 和 114 可包括静态拍摄装置(例如,配置成捕捉静态照片的拍摄装置)或视频拍摄装置(例如,配置成捕捉移动图像(其由多个帧组成)的拍摄装置)。拍摄装置 104 和 114 可配置成使用可见光谱中的光或利用电磁波谱(不限于红外光谱、紫外光谱)的其他部分来操作。在一个实施例中,拍摄装置 104 和 114 可配置成检测从拍摄装置的深度,即到物体和 / 或该物体上的点的距离。拍摄装置 104 和 114 可分别并入设备 102 和 112 内,或可以是配置成经由有线或无线通信与设备 102 和 112 通信的单独设备。拍摄装置 104 和 114 的特定示例可包括如可与计算机、视频监视器等关联的有线(例如,通用串行总线(USB)、以太网、火线等)或无线(例如,WiFi、Bluetooth,等)web 拍摄装置、深度拍摄装置、移动设备拍摄装置(例如在例如之前论述的示例设备中集成的手机或智能电话拍摄装置)、集成便携式计算机拍摄装置、集成平板计算机拍摄装置(例如,iPad®、Galaxy Tab® 及类似物),等。

[0015] 设备 102 和 112 可进一步包括麦克风 106 和 116 以及扬声器 107 和 117。麦克风 106 和 116 包括配置成感测(即,捕捉)声音并且将感测的声音转换成对应音频信号的任何设备。麦克风 106 和 116 可分别在设备 102 和 112 内集成,或可经由有线或无线通信(例如在上文关于拍摄装置 104 和 114 的示例中描述的)而与设备交互。扬声器 107 和 117 包括配置成将音频信号转换成对应声音的任何设备。扬声器 107 和 117 可分别在设备 102 和 112 内集成,或可经由有线或无线通信(例如在上文关于拍摄装置 104 和 114 的示例中描述的)而与设备交互。触敏显示器 108 和 118 包括配置成显示文本、静态图像、移动图像(例如,视频)、用户界面、图形等并且配置成感测例如点击、滑动等触摸事件的任何设备。触摸事件可包括触摸类型和触摸位点。触敏显示器 108 和 118 可分别在设备 102 和 112 内集成,或可经由有线或无线通信(例如在上文关于拍摄装置 104 和 114 的示例中描述的)而与设备交互。在一个实施例中,显示器 108 和 118 配置成分别显示化身 110 和 120。如本文引用的,化身限定为采用二维(2D)或三维(3D)的用户的图形表示。化身不必与用户的相貌相像,并且从而,尽管化身可以是逼真表示,它们还可以采取图、卡通、草图等的形式。在系统 100 中,设备 102 可显示代表设备 112 的用户(例如,远程用户)的化身 110,并且同样,设备 112 可显示代表设备 102 的用户的化身 120。这样,用户可看到其他用户的表示而不必交换对于采用实时图像的设备到设备通信所牵涉的大量信息。此外,化身可基于用户输入而动画地呈现。采用该方式,用户可与本地和 / 或远程化身的显示交互,由此增强用户体验。所得的动画可比可能仅使用面部检测和跟踪提供更广泛的动画。此外,用户可主动选择动画。

[0016] 如本文提及的,化身音频(即,声音)限定为变换的用户音频(声音)。例如,声音输入可包括用户的话音(即,用户语音),并且对应的化身音频可包括变换的用户语音。化身音频可与用户音频有关。例如,化身语音可对应于用户语音的移调、时间延展和 / 或其他变换。化身语音可与人语音相像或可对应于卡通人物,等。在系统 100 中,设备 102 可发出代表设备 112 的远程用户的化身音频并且相似地,设备 112 可发出代表由设备 102 捕捉的音频(例如,设备 102 的本地用户的语音)的化身音频。这样,用户可听见可被变换的其他用户语音的表示。

[0017] 网络 122 可包括各种第二代(2G)、第三代(3G)、第四代(4G)基于蜂窝的数据通信

技术、Wi-Fi 无线数据通信技术,等。网络 122 包括至少一个服务器 124,其配置成在使用这些技术时建立并且维持通信连接。例如,服务器 124 可配置成支持因特网相关通信协议,像用于创建、修改和终止双方(单播)和多方(多播)会话的会话发起协议(SIP)、用于呈现允许在字节流连接顶部建立协议的框架的交互连接性建立协议(ICE)、用于允许应用通过网络接入翻译器(NAT)操作来发现其他 NAT、IP 地址和端口(被分配给连接到远程主机的应用用户数据报协议(UDP)连接)的存在的网络接入翻译器或 NAT 的会话穿越实用性协议(STUN)、用于允许 NAT 或防火墙后面的元件通过传输控制协议(TCP)或 UDP 连接来接收数据的在 NAT 周围使用中继器的穿越(TURN),等。

[0018] 图 1B 图示与本公开的各种实施例一致的虚拟空间系统 126。系统 126 可采用设备 102、设备 112 和服务器 124。设备 102、设备 112 和服务器 124 可继续采用与在图 1A 中图示的相似的方式通信,但用户交互可在虚拟空间 128 中而不是采用设备到设备格式发生。如本文提及的,虚拟空间可限定为物理位点的数字模拟。例如,虚拟空间 128 可与户外位点(像城市、道路、人行道、田地、森林、岛等)或内部位点(像办公室、住宅、学校、购物中心、商店,等)相像。由化身表示的用户可表现为如现实世界中那样在虚拟空间 128 中交互。虚拟空间 128 可在耦合于因特网的一个或多个服务器上存在,并且可由第三方维持。虚拟空间的示例包括虚拟办公室、虚拟会议室、虚拟世界(像 Second Life®)、大型多人在线角色扮演游戏(MMORPG)(像 World of Warcraft®)、大型多人在线真实游戏(MMORLG)(像 Sims Online®)等。在系统 126 中,虚拟空间 128 可包含对应于不同用户的多个化身。显示器 108 和 118 可显示虚拟空间(VS) 128 的封装(例如,较小)版本,而不是显示化身。例如,显示器 108 可显示对应于设备 102 的用户的化身在虚拟空间 128 中所“看到”的事物的透视图。相似地,显示器 118 可显示对应于设备 112 的用户的化身在虚拟空间 128 中所“看到”的事物的透视图。化身可能在虚拟空间 128 中看到的事物的示例包括但不限于,虚拟结构(例如,建筑)、虚拟车辆、虚拟物体、虚拟动物、其他化身等。

[0019] 图 2 图示根据本公开的各种实施例的示例设备 102。尽管仅描述一个设备 102,设备 112(例如,远程设备)可包括配置成提供相同或相似功能的资源。如之前论述的,示出设备 102 包括拍摄装置 104、麦克风 106、扬声器 107 和触敏显示器 108。拍摄装置 104、麦克风 106 和触敏显示器 108 可对拍摄装置、音频和触屏框架模块 200 提供输入,并且拍摄装置、音频和触屏框架模块 200 可对扬声器 107 提供输出(例如,音频信号)。拍摄装置、音频和触屏框架模块 200 可包括定制、专用、已知和 / 或之后开发的音频和视频处理代码(或指令集),其一般被明确定义并且能操作成控制至少拍摄装置 104、麦克风 106、扬声器 107 和触敏显示器 108。例如,拍摄装置、音频和触屏框架模块 200 可促使拍摄装置 104、麦克风 106、扬声器 107 和触敏显示器 108 记录图像、到物体的距离、声音和 / 或触摸,可处理图像、声音、音频信号和 / 或触摸,可促使图像和 / 或声音再现,可向扬声器 107 提供音频信号,等。拍摄装置、音频和触屏框架模块 200 可根据设备 102 并且更特定地根据在设备 102 中运行的操作系统(OS)而改变。示例操作系统包括 iOS®、Android®、Blackberry® OS、Symbian®、Palm® OS,等。扬声器 107 可从拍摄装置、音频和触屏框架模块 200 接收音频信息并且可配置成再现本地声音(例如,以便提供用户话音的音频反馈(变换或未变换的))和远程声音(例如,参加电话、视频通话或在虚拟空间中交互的其他方的声音(变换或未变换的))。

[0020] 面部检测和跟踪模块 202 可配置成识别并且跟踪由拍摄装置 104 提供的图像内的

头部、面部和 / 或面部区域。例如,面部检测模块 204 可包括定制、专用、已知和 / 或之后开发的面部检测代码(或指令集)、硬件和 / 或固件,其一般被明确定义并且能操作成接收标准格式图像(例如,但不限于,RGB 彩色图像)并且至少在一定程度上识别图像中的面部。面部检测和跟踪模块 202 还可配置成通过一系列图像(例如,以每秒 24 个帧的视频帧)来跟踪检测的面部并且基于检测的面部确定头部位置。可由面部检测 / 跟踪模块 202 采用的已知跟踪系统可包括粒子滤波、均值偏移、卡尔曼滤波等,其中的每个可利用边缘分析、平方和差值分析、特征点分析、直方图分析、肤色分析等。

[0021] 特征提取模块 204 可配置成识别由面部检测模块 202 检测的面部中的特征(例如,例如眼睛、眉毛、鼻子、嘴等面部标志的位点和 / 或形状)。在一个实施例中,化身动画可直接基于感测的面部动作(例如,面部特征中的改变)而没有面部表情识别。化身面部上的对应特征点可遵循或模仿真实人的面部的移动,其称为“表达克隆”或“表现驱动的面部动画”。特征提取模块 204 可包括定制、专用、已知和 / 或之后开发的面部特性识别代码(或指令集),其一般被明确定义并且能操作成从拍摄装置 104 接收标准格式图像(例如,但不限于,RGB 彩色图像)并且至少在一定程度上提取图像中的一个或多个面部特性。这样的已知面部特性系统可包括但不限于,科罗拉多州立大学的 CSU 面部识别评估系统。

[0022] 特征提取模块 204 还可配置成识别与检测的特征关联的表情(例如,识别之前检测的面部是否愉快、悲伤、微笑、皱眉、惊讶、激动等)。从而,特征提取模块 204 可进一步包括定制、专用、已知和 / 或之后开发的面部表情检测和 / 或识别代码(或指令集),其一般被明确定义并且能操作成检测和 / 或识别面部中的表情。例如,特征提取模块 204 可确定面部特征(例如,眼睛、嘴、脸颊、牙齿等)的大小和 / 或位置并且可将这些面部特征与面部特征数据库比较,该面部特征数据库包括具有对应面部特征分类(例如,微笑、皱眉、激动、悲伤等)的多个样本面部特征。

[0023] 音频变换模块 206 配置成将用户的话音变换成化身语音,即变换的用户语音。变换包括调整节奏(例如,时间延展)、音调(例如,移调)和重放速率。例如,音频变换模块 206 可包括定制、专用、已知和 / 或之后开发的音频变换代码(或指令集),其一般被明确定义并且能操作成接收代表用户语音的话音数据并且将该话音数据转换成变换的话音数据。话音数据可与基于由麦克风 106 捕捉并且被拍摄装置、音频和触屏框架模块 200 处理的声音的音频信号有关。这样的已知话音变换系统包括但不限于 SoundTouch 开源音频处理库,配置成调整音频流或音频文件的节奏、音调(pitch)和重放速率。

[0024] 音频变换模块 206 可包括多个预定义话音风格,其对应于与变换用户的话音关联的变换参数。例如,变换参数可配置成维持具有不同音调和 / 或节奏的人类发声变换话音输出。音调可移到女人或像孩子的话音的较高频率,音调可移到男人话音的较低频率,节奏可向上或向下调整来增加或减小语音的速度,等。在另一个示例中,变换参数可配置成产生变换的话音输出,其对应于像动物的话音(例如,猫)和 / 或卡通人物型话音。这可通过调整用户语音的音调、其他频率分量和 / 或采样参数而实现。

[0025] 用户可在发起通信之前选择期望的音频变换输出和 / 或可在通信期间选择期望的音频变换。音频变换模块 206 可配置成响应于来自用户的请求提供样本音频变换输出。在实施例中,音频变换模块 206 可包括这样的实用程序(utility),其允许用户选择音频变换参数来产生定制音频变换输出。该实用程序可配置成基于用户的话音输入来提供样本变

换音频输出。用户然后可调整音频变换参数(例如,通过试错法)直到实现合适的变换输出。与对于用户合适的输出关联的音频变换参数然后可被存储和 / 或利用以用于化身通信,如本文描述的。

[0026] 触摸检测模块 208 配置成从拍摄装置、音频和触屏框架模块 200 接收触摸数据并且基于接收的触摸数据识别触摸事件。触摸事件标识符可包括触摸类型和 / 或触摸位点。触摸类型可包括单击、双击、点击并按住、点击并移动、收缩和延展、滑动等。触摸位点可包括触摸起始位点、触摸结束位点和 / 或中间移动触摸位点等。触摸位点可对应于触敏显示器 108 的坐标。触摸检测模块 208 可包括定制、专用、已知和 / 或之后开发的触摸检测代码(或指令集),其一般被明确定义并且能操作成接收触摸数据并且识别触摸事件。

[0027] 手势检测模块 210 配置成从拍摄装置、音频和触屏框架模块 200 接收深度和 / 或图像数据、基于接收的深度和 / 或图像数据识别对应的手势以及基于识别的手势确定手势标识符。深度对应于从拍摄装置到物体的距离。手势标识符与识别的手势有关。手势检测模块 210 可包括定制、专用、已知和 / 或之后开发的手势检测代码(或指令集),其一般被明确定义并且能操作成基于接收的深度和 / 或图像数据识别手势。

[0028] 例如,手势检测模块 210 可包括预定义手势的数据库。这些预定义手势可包括至少一些相对普通、相对简单的手势,其包括张开的手、紧握的手(即,拳头)、挥手、用手做圆周运动、使手从右到左地移动、使手从左到右地移动等。从而,手势可包括静止、非移动的手部手势、主动移动手部手势和 / 或其组合。在实施例中,手势检测模块 210 可包括训练实用程序,其配置成允许用户定制预定义手势和 / 或训练新的手势。定制手势和 / 或新的手势然后可与手势标识符关联并且该手势标识符可与动画命令关联,如本文描述的。例如,用户可从动画命令的预定义列表中选择动画命令以与手势关联。

[0029] 从而,动画命令与对用户输入的期望响应有关。动画命令可与识别的用户输入(例如,触摸事件标识符和 / 或手势标识符)关联。采用该方式,用户可与显示的化身交互和 / 或可做手势以便修改显示的化身的动画。

[0030] 化身选择模块 212 配置成允许设备 102 的用户选择化身用于在远程设备上显示。化身选择模块 212 可包括定制、专用、已知和 / 或之后开发的用户界面构造代码(或指令集),其一般被明确定义并且能操作成向用户呈现不同的化身使得用户可选择这些化身中的一个。在一个实施例中,一个或多个化身可在设备 102 中预定义。预定义化身允许所有设备具有相同的化身,并且在交互期间仅化身的选择(例如,预定义化身的识别)需要传达到远程设备或虚拟空间,这使需要交换的信息的量减少。在建立通信之前选择化身,但也可在有效通信过程中改变化身。从而,在通信期间的任何点发送或接收化身选择,并且接收设备根据接收的化身选择改变显示的化身,这可以是可能的。

[0031] 化身控制模块 214 配置成基于到设备 102 的用户输入来接收用户输入标识符。该用户输入标识符可包括由触摸检测模块 208 基于触摸事件数据确定的触摸事件标识符或由手势检测模块 210 确定的手势标识符。触摸事件数据包括触摸类型和触摸位点。触摸位点可对应于与触敏显示器 108 关联的坐标。触摸位点可映射到显示的化身上的一个或多个点,例如映射到特征(例如,鼻端、嘴、嘴唇、耳朵、眼睛,等)。显示的化身上的点可与化身动画的期望响应(即,化身命令)有关。

[0032] 化身控制模块 214 配置成基于用户输入标识符(即,识别的用户输入)确定动画命

令。动画命令配置成识别期望的化身动画。例如,期望的动画包括改变显示的化身面部的颜色、改变显示的化身的特征的大小(例如,使鼻子更大)、递眼色(winking)、眨眼、微笑、去除特征(例如,耳朵)等。从而,化身控制模块 214 配置成接收用户输入标识符并且基于该用户输入标识符确定动画命令。

[0033] 化身控制模块 214 配置成基于动画命令实现化身动画。在实施例中,对于在远程设备(例如,设备 112)上显示的交互动画,动画命令可被传送并且远程化身控制模块然后可实现动画。在另一个实施例中,可传送化身参数,其配置成用于化身动画的立即实现。

[0034] 实现的基于动画命令的交互动画可具有有限持续时间,在其之后化身动画可回到如本文描述的基于例如面部检测和跟踪的被动动画。实现的影响特征大小的交互动画可配置成逐渐改变大小并且逐渐回到初始大小。另外或备选地,影响特征大小的动画可配置成具有效应梯度。也就是说,大小改变的相对幅度可取决于相对于例如关键顶点的位点。显示的化身上更接近关键顶点的点可比显示的化身上相对更远的点经历更大的改变。

[0035] 从而,化身控制模块 214 可基于用户输入来接收用户输入标识符,并且可基于该用户输入标识符确定动画命令并且可基于该动画命令实现动画。基于动画命令的交互动画在时间上可局限于一段时间(持续时间)和/或可包括效应梯度。动画在该时段后可回到基于面部检测和跟踪的被动化身动画。

[0036] 化身控制模块 214 配置成生成用于使化身动画地呈现的参数。本文所称的动画可限定为更改图像/模型的外观。动画包括基于例如面部表情和/或头部移动的被动动画和基于用户输入的交互动画。单个动画(其可包括被动和交互动画)可更改 2D 静态图像的外观,或多个动画可依次出现来模拟图像中的运动(例如,头部转动、点头、眨眼、讲话、皱眉、微笑、发笑、递眼色、眨眼,等)。对于 3D 模型的动画示例包括使 3D 线框模型变形、应用纹理映射以及重新计算正常用于渲染的模型顶点。检测的面部和/或提取的面部特征的位置中的改变可转换成这样的参数,其促使化身的特征与用户面部的特征相像。在一个实施例中,检测的面部的一般表情可转换成促使化身展现相同表情的一个或多个参数。化身的表情还可夸大来突出表情。对选择的化身的知悉在化身参数一般可应用于预定义化身中的全部时可不是必需的。然而,在一个实施例中,化身参数对于选择的化身可以是特定的,并且从而,如果选择另一个化身则可更改化身参数。例如,人像化身可能需要与动物化身、卡通化身等不同的参数设置(例如,可更改不同的化身特征)来证明像快乐、悲伤、生气、惊讶等情感。

[0037] 化身控制模块 214 可包括定制、专用、已知和/或之后开发的图形处理代码(或指令集),其一般被明确定义并且能操作成生成用于使化身动画地呈现的参数,该化身由化身选择模块 212 基于由面部检测和跟踪模块 202 检测的面部/头部位置、特征提取模块 204 检测的面部特征和/或由触摸检测模块 208 和/或手势检测模块 210 确定的用户输入标识符来选择。对于基于面部特征的动画方法,2D 化身动画可利用例如图像扭曲或图像变形来进行,而 3D 化身动画可利用自由变形(FFD)或通过利用在头部的 3D 模型中限定的动画结构来进行。Oddcast 是可用于 2D 化身动画的软件资源的示例,而 FaceGen 是可用于 3D 化身动画的软件资源的示例。

[0038] 例如,对于包括延长 3D 化身显示的鼻子的交互动画,关键顶点  $v_k$  可与鼻子的尖端有关地限定(例如,选择)。对于关键顶点  $v_k$  可限定关联的 3D 运动矢量  $d_k$  ( $dx, dy, dz$ )和作用半径  $R$ 。作用半径  $R$  内的其他顶点可在交互动画中改变(即,移动),而作用半径  $R$  外部的

顶点可根据交互动画而保持不变。交互动画可具有关联的持续时间,即动画时间  $T$ ,其可延续多个帧。时间效应参数  $\eta_t$  可基于时间  $t$  和动画时间  $T$  而限定为:

$$\eta_t = \begin{cases} t/T, & 0 \leq t < T \\ 2 - t/T, & T \leq t < 2T \\ 0, & \text{其他} \end{cases}$$

相对更接近  $v_k$  的作用半径  $R$  内的顶点可比相对更远离关键顶点  $v_k$  的顶点改变得相对更大。对于顶点  $v_i$  的空间效应参数  $\eta_i$  可限定为:

$$\eta_i = \exp\left(-\frac{\|v_k - v_i\|^2}{R^2}\right)$$

并且顶点  $v_i$  在时间  $t$  的运动矢量  $d_i^t$  可限定为  $d_i^t = \eta_t \cdot \eta_i \cdot d_k$ 。交互动画呈现的化身的新的坐标则是  $v_i^t = v_i^0 + d_i^t$ , 其中  $v_i^0$  对应于基于面部检测和跟踪(即被动动画)的顶点  $v_i$  的坐标。

[0039] 从而,可对于显示的化身(其包括修改被动动画的交互动画)实现动画。交互动画可在总持续时间方面受限制并且动画效应的幅度可在该持续时间内改变。交互动画可配置成仅影响化身的一部分并且效应对于更靠近关键顶点的点更大。在已经完成交互动画后,动画可基于如本文描述的面部检测和跟踪而持续。

[0040] 另外,在系统 100 中,动画控制模块 214 可接收远程化身选择和远程化身参数,其可用于在远程设备处显示对应于用户的化身并且使其动画地呈现。动画可包括被动动画以及交互动画。化身控制模块可促使显示模块 216 在显示器 108 上显示化身 110。显示模块 216 可包括定制、专用、已知和 / 或之后开发的图形处理代码(或指令集),其一般被明确定义并且能操作成根据示例设备到设备实施例在显示器 108 上显示化身并且使其动画地呈现。例如,化身控制模块 214 可接收远程化身选择并且可解释该远程化身选择以对应于预定化身。显示模块 216 然后可在显示器 108 上显示化身 110。此外,可解释在化身控制模块 214 中接收的远程化身参数,并且命令可提供给显示模块 216 来使化身 110 动画地呈现。在一个实施例中,超过两个用户可参加视频通话。当超过两个用户在视频通话中交互时,显示器 108 可划分或分割以允许对应于远程用户的超过一个化身同时显示。备选地,在系统 126 中,化身控制模块 214 可接收信息,该信息促使显示模块 216 显示对应于设备 102 的用户的化身在虚拟空间 128 中所“看到”的事物(例如,从化身的视角)。例如,显示器 108 可显示在虚拟空间 128 中表示的建筑、对象、动物、其他化身等。

[0041] 在一个实施例中,化身控制模块 214 可配置成促使显示模块 216 显示“反馈”化身 218。反馈化身 218 代表选择的化身如何在远程设备上、在虚拟空间中等出现。特别地,反馈化身 218 表现为由用户选择的化身并且可使用由化身控制模块 214 生成的相同参数而动画地呈现。这样,用户可确认远程用户在他们的交互期间所看到的。反馈化身 218 还可用于显示由到设备 112 的远程用户输入引起的交互动画。从而,本地用户可与他或她的反馈化身(例如,化身 218 和设备 102 的用户)交互来促使向设备 112 上的远程用户所显示的他或她关联的化身的交互动画。本地用户可相似地与远程用户的显示化身(例如,化身 110)交互来促使在设备 112 上显示的远程用户的反馈化身的交互动画。

[0042] 通信模块 220 配置成传送并且接收用于选择化身、显示化身、使化身动画地呈现、显示虚拟地方角度等的信息。通信模块 220 可包括定制、专用、已知和 / 或之后开发的通信处理代码(或指令集),其一般被明确定义并且能操作成传送化身选择、化身参数、动画命令、交互化身参数并且接收远程化身选择、远程化身参数、远程动画命令和远程交互化身参数。通信模块 220 还可传送并且接收对应于基于化身的交互的音频信息。通信模块 220 可经由如之前描述的网络 122 传送并且接收上文的信息。

[0043] 处理器 222 配置成执行与设备 102 以及其中包括的模块中的一个或多个关联的操作。

[0044] 图 3 图示根据至少一个实施例的示例系统实现。设备 102' 配置成经由 WiFi 连接 300 而无线通信(例如,在工作中),服务器 124' 配置成经由因特网 302 协商设备 102' 与 112' 之间的连接,并且装置 112' 配置成经由另一个 WiFi 连接 304 而无线通信(例如,在家)。在一个实施例中,基于设备到设备化身的视频通话应用在装置 102' 中激活。在化身选择之后,应用可允许选择至少一个远程设备(例如,设备 112')。应用然后可促使设备 102' 发起与设备 112' 的通信。通信可随着设备 102' 经由企业接入点(AP) 306 向设备 112' 传送连接建立请求而发起。企业 AP 306 可以是可在商业环境中使用的 AP,并且从而可支持比家庭 AP 314 更高的数据吞吐量和更多的并发无线客户端。企业 AP 306 可从设备 102' 接收无线信号并且可经由网关 308 通过各种商业网络进行传送连接建立请求。连接建立请求然后可穿过防火墙 310,其可配置成控制流入和流出 WiFi 网络 300 的信息。

[0045] 设备 102' 的连接建立请求然后可由服务器 124' 处理。服务器 124' 可配置成用于 IP 地址的注册、目的地地址的验证和 NAT 穿越使得连接建立请求可被指引到因特网 302 上的正确目的地。例如,服务器 124' 可从设备 102' 接收的连接建立请求中的信息解析计划的目的地(例如,远程设备 112'),并且可因此将信号路由通过正确的 NAT、端口并且到目的地 IP 地址。根据网络配置,这些操作可仅必须在连接建立期间执行。在一些实例中,操作可在视频通话期间重复以便向 NAT 提供使连接保持活动的通知。在已经建立连接后,媒体和信号路径 312 可直接将视频(例如,化身选择和 / 或化身参数)和音频信息运送到家庭 AP 314。设备 112' 然后可接收连接建立请求并且可配置成确定是否接受该请求。确定是否接受请求可包括例如向设备 112' 的用户呈现视觉叙述,询问是否接受来自设备 102' 的连接请求。如果设备 112' 的用户接受连接(例如,接受视频通话),可建立连接。拍摄装置 104' 和 114' 可分别配置成然后开始捕捉设备 102' 和 112' 的相应用户的图像,以供在使由每个用户选择的化身动画呈现中使用。麦克风 106' 和 116' 可配置成然后开始捕捉来自每个用户的音频。在设备 102' 与 112' 之间开始信息交换时,显示器 108' 和 118' 可显示对应于设备 102' 和 112' 的用户的化身并且使其动画地呈现。

[0046] 图 4 图示与本公开的实施例一致的示范性操作的流程图 400。这些操作可例如由设备 102 和 / 或 112 执行。特别地,流程图 400 描绘配置成实现化身动画(其包括被动动画和 / 或交互动画)和 / 或音频变换以用于通过网络的设备之间的通信的示范性操作。假设面部检测和跟踪、特征提取和被动化身动画如本文描述的那样实现和操作。

[0047] 可在操作 402 选择化身模型。该化身模型可包括视频化身选择和音频变换选择。可显示多个视频化身模型,用户可从其中选择期望的化身。在实施例中,选择视频化身模型可包括关联的音频变换。例如,像猫的化身可与像猫的音频变换关联。在另一个实施例中,

音频变换可独立于视频化身选择而选择。

[0048] 化身模型(其包括音频变换)可在激活通信之前选择,但也可在有效通信的过程期间改变。从而,在通信期间的任何点发送或接收化身选择和/或改变音频变换选择,并且接收设备根据接收的化身选择改变显示的化身,这可以是可能的。

[0049] 可在操作 404 激活化身通信。例如,用户可启动这样的应用,其配置成使用如本文描述的化身来传达音频和视频。操作 404 可包括配置通信并且建立连接。通信配置包括识别参与视频通话的至少一个远程设备或虚拟空间。例如,用户可从存储在应用内、与设备中的另一个系统关联地存储(例如,智能电话、手机等中的联系人列表)、远程存储(例如在因特网上(例如,在像 Facebook、LinkedIn、Yahoo、Google+、MSN 等的社交媒体网站中))的远程用户/设备列表选择。备选地,用户可选择在像 Second Life 的虚拟空间中上线。

[0050] 在操作 406,设备中的拍摄装置然后可开始捕捉图像和/或深度并且设备中的麦克风可开始捕捉声音。图像可以是静态图像或实况视频(例如,依次捕捉的多个图像)。深度可与图像一起捕捉或可独立捕捉。深度对应于在拍摄装置的视场中从拍摄装置到物体(和物体上的点)的距离。是否检测到用户输入可在操作 408 确定。用户输入包括由图像和/或深度拍摄装置捕捉的手势和在触敏显示器上检测的触摸输入。如果检测到用户输入,可在操作 410 识别该用户输入。用户输入标识符包括触摸标识符或手势标识符。该触摸标识符可基于触敏显示器上的触摸而确定并且可包括触摸类型和触摸位点。手势标识符可基于捕捉的图像和/或深度数据而确定并且可包括识别手势。

[0051] 动画命令可在操作 412 识别。该动画命令可配置成使在远程设备上显示的用户选择的化身动画地呈现或使也在远程用户的设备上显示的用户反馈化身动画呈现。动画命令对应于与用户输入关联的期望响应。例如,触摸显示的化身的面部(用户输入)可导致显示的化身的面部的颜色改变(由动画命令识别的期望响应)。动画命令可基于识别的用户输入而识别。例如,每个用户输入可与用户输入标识符和动画命令的数据库中的动画命令有关(例如,关联)。

[0052] 操作 414 包括生成化身参数。这些化身参数包括被动组成并且可包括交互组成。如果未检测到用户输入,化身参数可包括被动组成。如果检测到用户输入,化身参数是否可包括交互组成取决于动画命令并且因此取决于用户输入。对于对应于动画命令(其配置成使用户的选择化身动画地呈现)的用户输入,动画命令可与仅包括被动组成的化身参数一起传送或可在传送之间应用于化身参数使得传送的化身参数包括被动和交互组成两者。对于对应于动画命令(其配置成使在远程用户的设备上显示的远程用户的反馈化身动画地呈现)的输入,可仅传送动画命令。

[0053] 操作 416 包括变换并且编码捕捉的音频。捕捉的音频可转换成音频信号(例如,用户语音信号)。用户语音信号可根据操作 402 的化身选择的音频变换部分来变换。变换的用户语音信号对应于化身语音信号。化身语音信号可使用已知技术来编码以用于通过网络传送到远程设备和/或虚拟空间。可在操作 418 传送变换且编码的音频。操作 418 可进一步包括传送动画命令和化身参数中的至少一个。传送动画命令配置成允许远程设备通过根据动画命令修改化身参数来使本地显示的化身动画地呈现。在传送之前已经根据动画命令修改的传送的化身参数可直接用于使在远程设备上显示的化身动画地呈现。也就是说,对由动画命令表示的化身参数的修改可本地或远程执行。

[0054] 操作 420 包括接收远程编码音频,其可以是变换的音频。操作 420 进一步包括接收远程动画命令和远程化身参数中的至少一个。该远程动画命令可用于修改化身参数,其对应于远程用户的显示化身或本地用户的显示的反馈化身。动画命令和化身参数配置成导致化身动画,其基于用户输入而被修改。接收的音频可在操作 422 解码和播放并且化身可在操作 424 显示和动画地呈现。

[0055] 显示的化身的动画可基于如本文描述的检测且识别的用户输入。在设备到设备通信(例如,系统 100)实例中,远程化身选择或远程化身参数中的至少一个可从远程设备接收。对应于远程用户的化身然后可基于接收的远程化身选择而显示,并且可基于接收的远程化身参数而动画地呈现。在虚拟地方交互(例如,系统 126)的实例中,可接收这样的信息,其允许设备显示对应于设备用户的化身所看到的事物。

[0056] 通信是否完成可在操作 426 确定。如果完成通信,程序流可在操作 428 结束。如果通信未完成,程序流可进行到操作 406 :捕捉图像、深度和 / 或音频。

[0057] 尽管图 4 图示根据实施例的各种操作,要理解不是图 4 中描述的操作中的全部对于其他实施例都是必需的。确实,在本文充分预想在本公开的其他实施例中,在图 4 中描绘的操作和 / 或本文描述的其他操作可采用未专门在图中的任一个中示出但仍完全与本公开一致这一方式组合。从而,针对未确切在一个图中示出的特征和 / 或操作的权利要求认为在本公开的范围和内容内。

[0058] 如在本文的任何实施例中使用的,术语“应用”可在代码或指令中体现,这些代码或指令可在例如主机处理器等可编程电路或其他可编程电路上执行。

[0059] 如在本文的任何实施例中使用的,术语“模块”可指应用、软件、固件和 / 或电路,其配置成执行前面提到的操作中的任一个。软件可体现为在至少一个非暂时性计算机可读存储介质上记录的软件封装、代码、指令、指令集和 / 或数据。固件可体现为在存储器设备中硬编码(例如,非易失性)的代码、指令或指令集和 / 或数据。

[0060] 如在本文的任何实施例中使用的“电路”可单独或采用任何组合地包括例如硬接线电路、可编程电路(例如计算机处理器,其包括一个或多个个体指令处理核)、状态机电路和 / 或存储由可编程电路执行的指令的固件。模块可以共同或单独体现为电路,其形成较大系统(例如,集成电路(IC)、片上系统(SoC)、台式计算机、便携式计算机、平板计算机、服务器、智能电话等)的一部分。

[0061] 本文描述的操作中的任一个可在包括一个或多个存储介质的系统中实现,该一个或多个存储介质其上可以独立地或组合地存储有指令,这些指令在被一个或多个处理器执行时执行方法。在这里,处理器可包括,例如服务器 CPU、移动设备 CPU 和 / 或其他可编程电路。而且,规定本文描述的操作可跨多个物理设备(例如在超过一个不同的物理位点处的处理结构)分布。存储介质可包括任何类型的有形介质:例如任何类型的盘,其包括硬盘、软盘、光盘、压缩盘只读存储器(CD-ROM)、压缩盘可重写(CD-RW)和磁光盘;例如只读存储器(ROM)、随机存取存储器(RAM)(例如动态和静态 RAM)、可擦除可编程只读存储器(EPROM)、电可擦除可编程只读存储器(EEPROM)、闪速存储器、固态硬盘(SSD)等半导体器件;磁或光卡;或适合于存储电子指令的任何类型的介质。其他实施例可实现为由可编程控制设备执行的软件模块。存储介质可以是非暂时性的。

[0062] 从而,本公开提供用于交互地使得用于实施视频通信的化身动画地呈现而不是实

况图像的方法和系统。与发送实况图像相比,化身的使用使交换的信息的量减少。系统和方法进一步配置成通过例如移调和 / 或时间延展捕捉的音频信号来将用户语音变换成化身语音。化身的交互动画可基于检测的用户输入,其包括触摸和手势。交互动画配置成修改基于面部检测和跟踪而确定的动画。

[0063] 根据一个方面,提供一种系统。该系统可包括:用户输入设备,其配置成捕捉用户输入;通信模块,其配置成传送并且接收信息;和一个或多个存储介质。另外,该一个或多个存储介质其上独立地或组合地存储有指令,这些指令在由一个或多个处理器执行时促成下列操作,其包括:选择化身;发起通信;检测用户输入;识别用户输入;基于用户输入识别动画命令;生成化身参数;以及传送动画命令和化身参数中的至少一个。

[0064] 另一个示例系统包括前面的部件并且进一步包括麦克风,其配置成捕捉声音并且将捕捉的声音转换成对应的音频信号,并且指令在由一个或多个处理器执行时促成下列额外操作:捕捉用户语音并且将用户语音转换成对应的用户语音信号;将该用户语音信号变换成化身语音信号;以及传送该化身语音信号。

[0065] 另一个示例系统包括前面的部件并且进一步包括拍摄装置,其配置成捕捉图像,并且指令在由一个或多个处理器执行时促成下列额外操作:捕捉图像;检测该图像中的面部;从面部提取特征;以及将特征转换成化身参数。

[0066] 另一个示例系统包括前面的部件并且进一步包括显示器,并且指令在由一个或多个处理器执行时促成下列额外操作:显示至少一个化身;接收远程动画命令和远程化身参数中的至少一个;以及基于该远程动画命令和远程化身参数中的至少一个使一个显示的化身动画地呈现。

[0067] 另一个示例系统包括前面的部件并且进一步包括扬声器,其配置成将音频信号转换成声音,并且指令在由一个或多个处理器执行时促成下列额外操作:接收远程化身语音信号;以及将该远程化身语音信号转换成化身语音。

[0068] 另一个示例系统包括前面的部件并且用户输入设备是拍摄装置,其配置成捕捉距离,并且用户输入是手势。

[0069] 另一个示例系统包括前面的部件并且用户输入设备是触敏显示器并且用户输入是触摸事件。

[0070] 另一个示例系统包括前面的部件并且变换包括移调和/或时间延展中的至少一个。

[0071] 根据另一个方面,提供一种方法。该方法可包括:选择化身;发起通信;检测用户输入;识别用户输入;基于用户输入识别动画命令;基于动画命令生成化身参数;以及传送动画命令和化身参数中的至少一个。

[0072] 另一个示例方法包括前面的操作并且进一步包括:捕捉用户语音并且将该用户语音转换成对应的用户语音信号;将该用户语音信号变换成化身语音信号;以及传送该化身语音信号。

[0073] 另一个示例方法包括前面的操作并且进一步包括:捕捉图像;检测该图像中的面部;从面部提取特征;以及将特征转换成化身参数。

[0074] 另一个示例方法包括前面的操作并且进一步包括:显示至少一个化身;接收远程动画命令和远程化身参数中的至少一个;以及基于该远程动画命令和远程化身参数中的至少一个使一个显示的化身动画地呈现。

[0075] 另一个示例方法包括前面的操作并且进一步包括：接收远程化身语音信号；以及将该远程化身语音信号转换成化身语音。

[0076] 另一个示例方法包括前面的操作并且用户输入是手势。

[0077] 另一个示例方法包括前面的操作并且用户输入是触摸事件。

[0078] 另一个示例方法包括前面的操作并且变换包括移调和/或时间延展中的至少一个。

[0079] 根据另一个方面，提供一种系统。该系统可包括一个或多个存储介质，其上独立地或组合地存储有指令，这些指令在由一个或多个处理器执行时促成下列操作，其包括：选择化身；发起通信；检测用户输入；识别用户输入；基于用户输入识别动画命令；生成化身参数；以及传送动画命令和化身参数中的至少一个。

[0080] 另一个示例系统包括指令，其在由一个或多个处理器执行时促成前面的操作并且还包含：捕捉用户语音并且将该用户语音转换成对应的用户语音信号；将该用户语音信号转换成化身语音信号；以及传送该化身语音信号。

[0081] 另一个示例系统包括指令，其在由一个或多个处理器执行时促成前面的操作并且还包含：捕捉图像；检测该图像中的面部；从面部提取特征；以及将特征转换成化身参数。

[0082] 另一个示例系统包括指令，其在由一个或多个处理器执行时促成前面的操作并且还包含：显示至少一个化身；接收远程动画命令和远程化身参数中的至少一个；以及基于该远程动画命令和远程化身参数中的至少一个使一个显示的化身动画地呈现。

[0083] 另一个示例系统包括指令，其在由一个或多个处理器执行时促成前面的操作并且还包含：接收远程化身语音信号；以及将该远程化身语音信号转换成化身语音。

[0084] 另一个示例系统包括指令，其在由一个或多个处理器执行时促成前面的操作并且用户输入是手势。

[0085] 另一个示例系统包括指令，其在由一个或多个处理器执行时促成前面的操作并且用户输入是触摸事件。

[0086] 另一个示例系统包括指令，其在由一个或多个处理器执行时促成前面的操作并且变换包括移调和/或时间延展中的至少一个。

[0087] 本文已经采用的术语和表达用作说明的术语并且无限制，并且在这样的术语和表达的使用中没有意图要排除示出和描述的特征（或其部分）的任何等同，并且认识到各种修改在权利要求的范围内是可能的。因此，权利要求意在涵盖所有这样的等同。

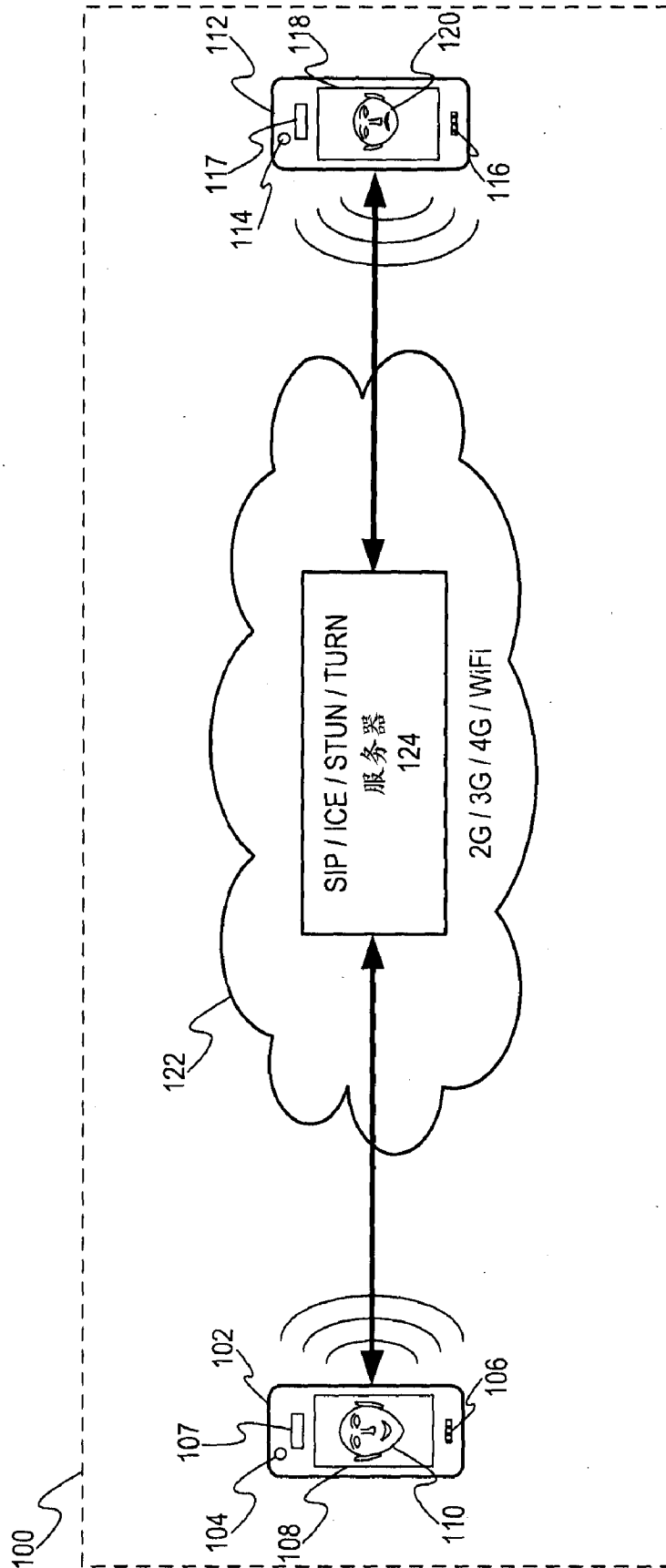


图 1A

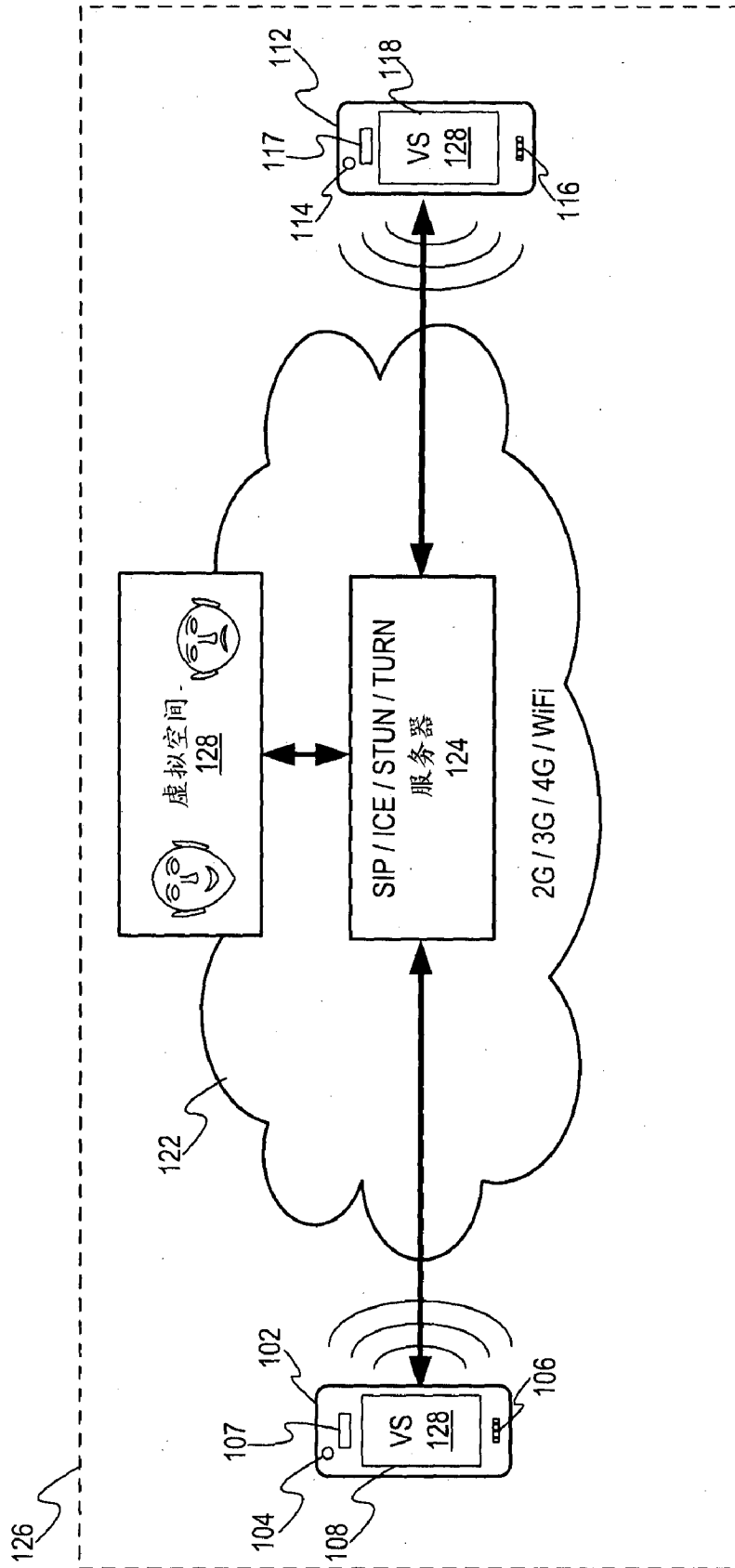


图 1B

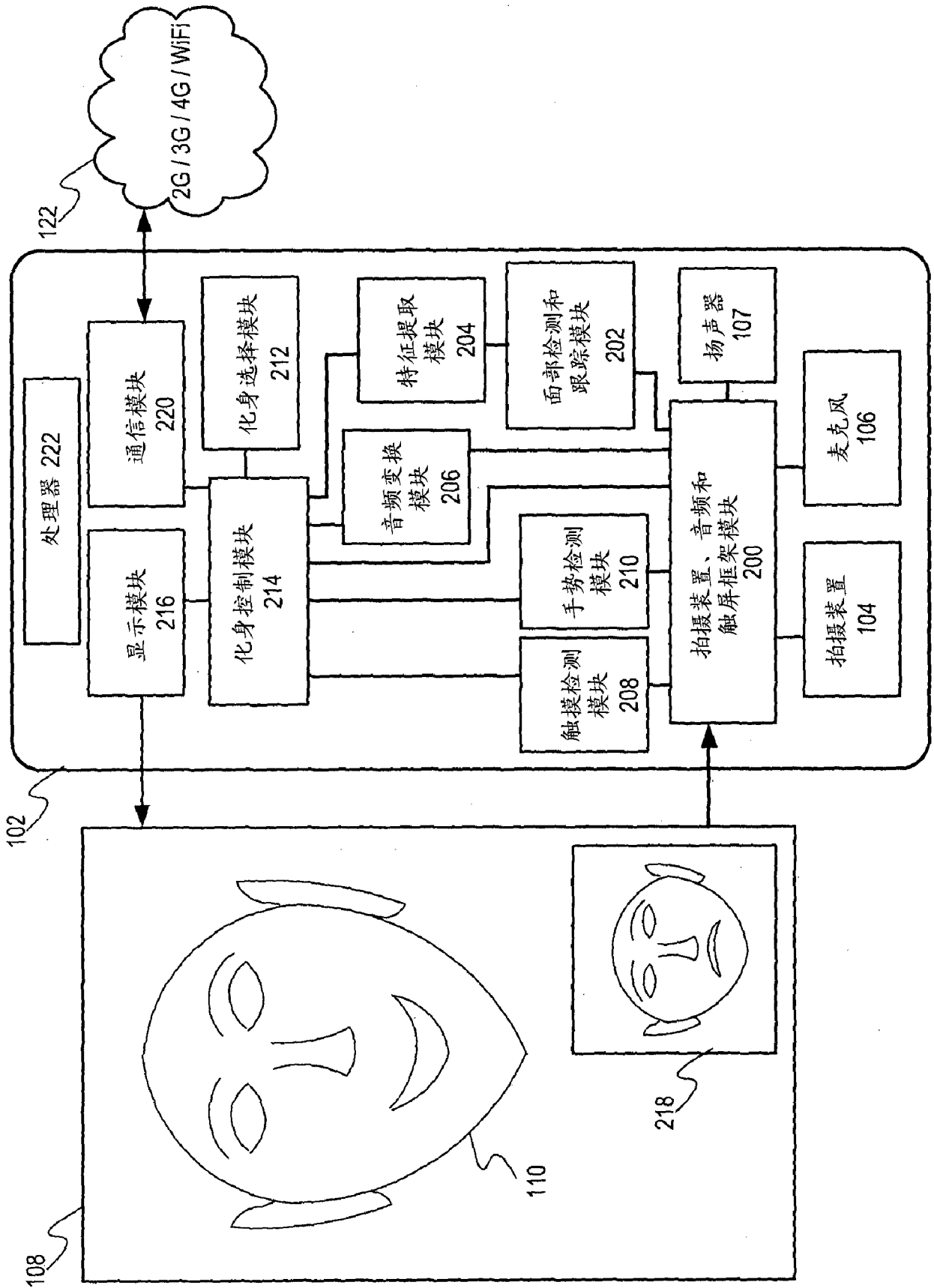


图 2

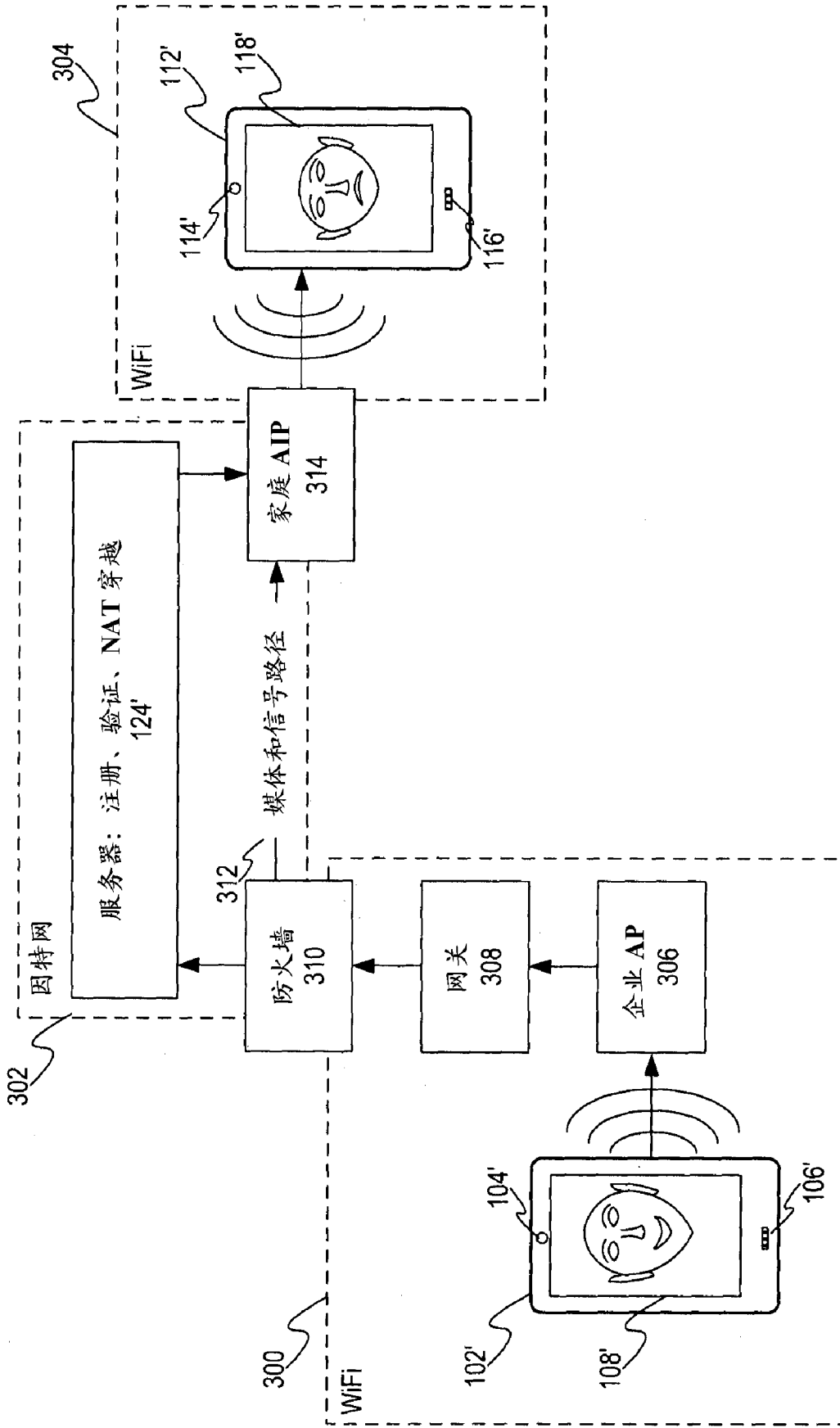


图 3

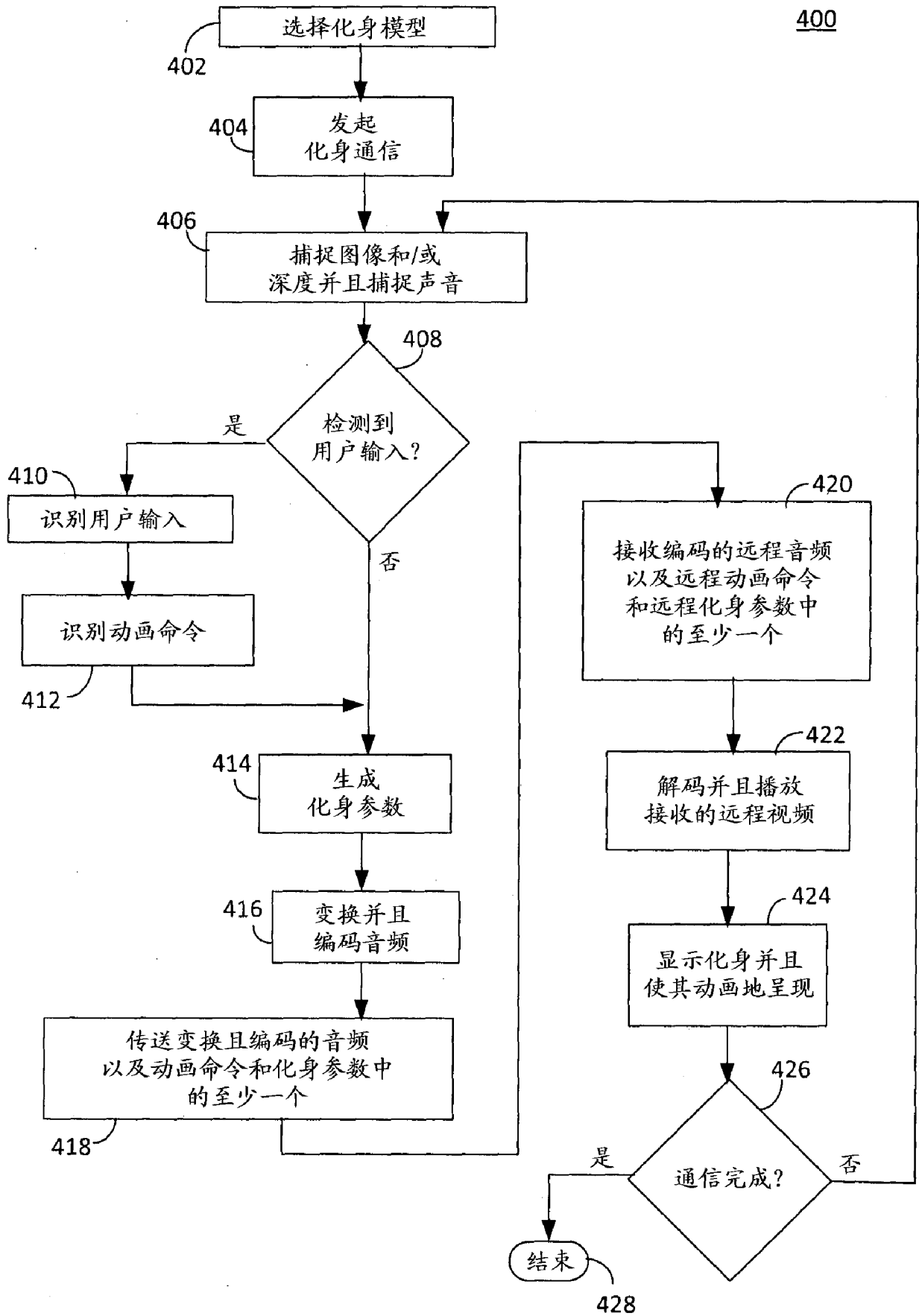


图 4