



(12)发明专利申请

(10)申请公布号 CN 113473238 A

(43)申请公布日 2021.10.01

(21)申请号 202010359143.7

H04N 21/437(2011.01)

(22)申请日 2020.04.29

H04N 21/43(2011.01)

(71)申请人 海信集团有限公司

地址 266071 山东省青岛市市南区东海西路17号

(72)发明人 矫佩佩 张玉 孙菁 陈维强

(74)专利代理机构 北京同达信恒知识产权代理有限公司 11291

代理人 朱佳

(51)Int.Cl.

H04N 21/4788(2011.01)

H04N 21/431(2011.01)

H04N 21/442(2011.01)

H04N 21/435(2011.01)

H04N 21/44(2011.01)

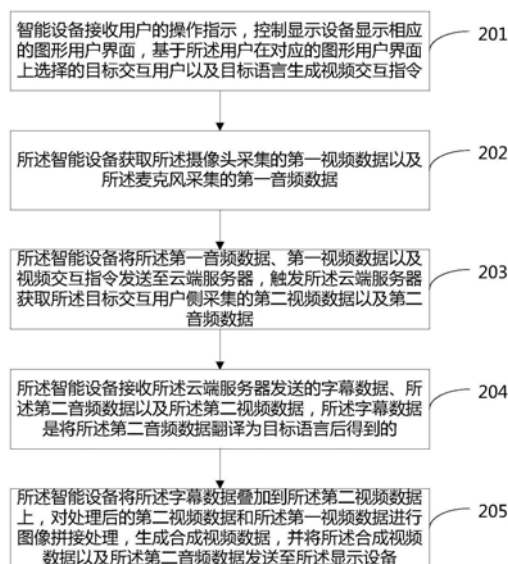
权利要求书2页 说明书12页 附图5页

(54)发明名称

一种智能设备及视频通话时的同声翻译方法

(57)摘要

本发明涉及智能家居领域,尤其涉及一种智能设备及视频通话时的同声翻译方法。方法包括:接收用户的操作指示,基于所述用户选择的目标交互用户以及目标语言生成视频交互指令,获取第一视频数据和第一音频数据,接收云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,并将生成的合成视频数据以及所述第二音频数据发送至显示设备。这样,无需为实现音频数据转换为文本数据而架设大规模处理设备,且通过将合成视频数据与第二音频数据的同步配置,避免了音频数据与字幕数据之间的延时,实现将同声翻译应用于日常的使用中。



1. 一种视频通话时的同声翻译方法,其特征在于,包括:

接收用户的操作指示,控制显示设备显示相应的图形用户界面,基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令,获取第一视频数据和第一音频数据;

将所述第一音频数据、所述第一视频数据以及所述视频交互指令发送至云服务器,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据;

接收所述云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,所述字幕数据是将所述第二音频数据翻译为目标语言后得到的;

将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二音频数据发送至所述显示设备。

2. 如权利要求1所述的方法,其特征在于,所述接收用户的操作指示,控制显示设备显示相应的图形用户界面之后,所述基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令之前,进一步包括:

确定所述用户基于所述显示设备上呈现的图形用户界面选择的目标语言,以及确定所述目标语言存在于预设的有效语言列表中,其中,所述有效语言列表中包含云服务器能够识别翻译的所有语言信息。

3. 如权利要求2所述的方法,其特征在于,进一步包括:

确定所述用户基于所述显示设备上呈现的图形用户界面选择的目标语言未包含在预设的有效语言列表中时,生成选择目标语言的提示信息;

将所述提示信息发送至所述显示设备进行显示,等待所述用户重新选择目标语言。

4. 如权利要求1-3任一项所述的方法,其特征在于,所述基于所述用户选择的目标交互用户以及目标语言生成视频交互指令时,包括:

确定所述用户选择开启同声翻译功能并选择目标交互用户时,至少基于所述用户选择的目标交互用户的ID信息以及开启同声翻译时选择的目标语言信息生成视频交互指令。

5. 如权利要求4所述的方法,其特征在于,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据之后,所述将所述合成视频数据以及所述第二音频数据发送至显示设备之前,进一步包括:

确定仅接收到所述云服务器发送的所述第二视频数据以及所述第二音频数据时,采用图像拼接技术对所述第二视频数据以及所述第一视频数据进行处理,生成合成视频数据。

6. 一种电子设备,其特征在于,包括:

存储器,用于存储可执行指令;

处理器,用于读取并执行存储器中存储的可执行指令,以实现如权利要求1至5中任一项所述的视频通话时的同声翻译方法。

7. 一种视频通话时的同声翻译方法,其特征在于,包括:

接收基于用户的操作指示确定呈现的图形用户界面,进行显示;

接收合成视频数据以及音频数据,并基于所述合成视频数据进行显示,以及基于所述音频数据进行播放,其中,所述合成视频数据是将目标交互用户侧采集的第二音频数据翻译为目标语言,生成字幕数据后,将所述字幕数据与所述目标交互用户侧采集的第二视频

数据进行叠加处理,并将叠加处理后的第二视频数据与本地采集的第一视频数据进行图像拼接处理后得到的,所述音频数据为所述目标用户侧采集的第二音频数据。

8.如权利要求7所述的方法,其特征在于,进一步包括:

确定接收选择目标语言的提示信息时,显示所述提示信息,并向所述用户呈现选择目标语言的图形用户界面。

9.一种显示设备,其特征在于,包括:

显示器,用于显示进行视频交互的图形用户界面,以及基于需要播放的视频交互数据进行显示;

扬声器,用于基于视频交互过程中目标交互用户的音频数据进行播放;

处理器,所述处理器用于执行:

接收智能设备基于用户的操作指示确定呈现的图形用户界面,调用所述显示器进行显示;

接收所述智能设备发送的合成视频数据以及音频数据,并基于所述合成视频数据在所述显示器上进行显示,以及基于所述音频数据在所述扬声器上进行播放,其中,所述合成视频数据是将目标交互用户侧采集的第二音频数据翻译为目标语言,生成字幕数据后,将所述字幕数据与所述目标交互用户侧采集的第二视频数据进行叠加处理,并将叠加处理后的第二视频数据与本地采集的第一视频数据进行图像拼接处理后得到的。

10.如权利要求9所述的显示设备,其特征在于,所述处理器进一步用于:

确定接收所述智能设备发送的选择目标语言的提示信息时,将所述提示信息显示在所述显示器上,并向所述用户呈现选择目标语言的图形用户界面。

一种智能设备及视频通话时的同声翻译方法

技术领域

[0001] 本发明涉及智能家居技术领域,尤其涉及一种智能设备及视频通话时的同声翻译方法。

背景技术

[0002] 随着智能电视的发展,智能电视不再只能满足人们的娱乐消遣需要,人们可以借助于智能电视,实现视频通话,又由于全球化趋势的推进,视频通话的对象可能使用的是不同的语言,消除不同语言之间的沟通障碍变得尤为重要,同声翻译技术也就应运而生。

[0003] 对于目前存在的同声翻译技术来说,一种情况下,通过在大规模国际会议或者直播等公共场景的现场,部署用于同声翻译的大型通信设备,实现会议或者直播过程中的同声传译,另一种情况下,同声翻译以应用软件的形式存在于具有较高计算能力的终端设备上,进而实现将语音内容转换为翻译的文字内容。

[0004] 然而,现有技术中,在第一种情况下,由于部署大规模通信设备的成本较高且部署困难,因而难以在日常的应用中普及,无法应用于智能电视上的视频通话场景中,而对于第二种情况,终端设备在得到语音数据后,在转译为文本数据的过程中需要一定的处理时间,这样会导致听到的音频数据和看到的文字数据之间不同步,存在一定的延时,极大影响了用户的使用体验,无法应用于智能电视上的视频通话过程中对语音数据进行处理。

发明内容

[0005] 本发明实施例提供一种智能设备及视频通话时的同声翻译方法,用以解决现有技术中存在无法在保证翻译得到的字幕数据与音频数据同步的基础上,将同声翻译应用于智能电视上的视频交互中的问题。

[0006] 本发明实施例提供的具体技术方案如下:

[0007] 一种视频通话时的同声翻译方法,包括:

[0008] 接收用户的操作指示,控制显示设备显示相应的图形用户界面,基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令,获取第一视频数据和第一音频数据;

[0009] 将所述第一音频数据、所述第一视频数据以及所述视频交互指令发送至云服务器,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据;

[0010] 接收所述云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,所述字幕数据是将所述第二音频数据翻译为目标语言后得到的;

[0011] 将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二音频数据发送至所述显示设备。

[0012] 可选的,所述接收用户的操作指示,控制显示设备显示相应的图形用户界面之后,所述基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频

交互指令之前,进一步包括:

[0013] 确定所述用户基于所述显示设备上呈现的图形用户界面选择的目标语言,以及确定所述目标语言存在于预设的有效语言列表中,其中,所述有效语言列表中包含云服务器能够识别翻译的所有语言信息。

[0014] 可选的,进一步包括:

[0015] 确定所述用户基于所述显示设备上呈现的图形用户界面选择的目标语言未包含在预设的有效语言列表中时,生成选择目标语言的提示信息;

[0016] 将所述提示信息发送至所述显示设备进行显示,等待所述用户重新选择目标语言。

[0017] 可选的,所述基于所述用户选择的目标交互用户以及目标语言生成视频交互指令时,包括:

[0018] 确定所述用户选择开启同声翻译功能并选择目标交互用户时,至少基于所述用户选择的目标交互用户的ID信息以及开启同声翻译时选择的目标语言信息生成视频交互指令。

[0019] 可选的,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据之后,所述将所述合成视频数据以及所述第二音频数据发送至显示设备之前,进一步包括:

[0020] 确定仅接收到所述云服务器发送的所述第二视频数据以及所述第二音频数据时,采用图像拼接技术对所述第二视频数据以及所述第一视频数据进行处理,生成合成视频数据。

[0021] 一种电子设备,包括:

[0022] 存储器,用于存储可执行指令;

[0023] 处理器,用于读取并执行存储器中存储的可执行指令,以实现上述任一项所述的视频通话时的同声翻译方法。

[0024] 一种视频通话时的同声翻译方法,包括:

[0025] 接收基于用户的操作指示确定呈现的图形用户界面,进行显示;

[0026] 接收合成视频数据以及音频数据,并基于所述合成视频数据进行显示,以及基于所述音频数据进行播放,其中,所述合成视频数据是将目标交互用户侧采集的第二音频数据翻译为目标语言,生成字幕数据后,将所述字幕数据与所述目标交互用户侧采集的第二视频数据进行叠加处理,并将叠加处理后的第二视频数据与本地采集的第一视频数据进行图像拼接处理后得到的,所述音频数据为所述目标用户侧采集的第二音频数据。

[0027] 一种显示设备,包括:

[0028] 显示器,用于显示进行视频交互的图形用户界面,以及基于需要播放的视频交互数据进行显示;

[0029] 扬声器,用于基于视频交互过程中目标交互用户的音频数据进行播放;

[0030] 处理器,所述处理器用于执行:

[0031] 接收智能设备基于用户的操作指示确定呈现的图形用户界面,调用所述显示器进行显示;

[0032] 接收所述智能设备发送的合成视频数据以及音频数据,并基于所述合成视频数据

在所述显示器上进行显示,以及基于所述音频数据在所述扬声器上进行播放,其中,所述合成视频数据是将目标交互用户侧采集的第二音频数据翻译为目标语言,生成字幕数据后,将所述字幕数据与所述目标交互用户侧采集的第二视频数据进行叠加处理,并将叠加处理后的第二视频数据与本地采集的第一视频数据进行图像拼接处理后得到的。

[0033] 本发明有益效果如下:

[0034] 本公开中,接收用户的操作指示,控制显示设备显示相应的图形用户界面,基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令,进而,获取第一视频数据和第一音频数据,并将所述第一声音频数据、所述第一视频数据以及所述视频交互指令发送至云服务器,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据,然后,接收所述云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,所述字幕数据时将所述第二音频数据翻译为目标语言后得到的,再将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二音频数据发送至所述显示设备。

[0035] 这样,通过从云服务器获取翻译后的字幕数据,有效降低了本地设备处理的难度,无需为实现音频数据转换为文本数据而架设大规模处理设备,且通过将合成视频数据与第二音频数据的同步配置,避免了音频数据与字幕数据之间的延时,一方面借由显示设备实现了视频交互,另一方面能够实现将同声翻译应用于日常的使用中。

附图说明

[0036] 图1A为本公开实施例中显示设备、智能设备以及云服务器之间的操作场景示意图;

[0037] 图1B为本公开实施例中智能设备的功能模块示意图;

[0038] 图2为本公开实施例中智能设备实现视频通话时同声翻译的流程示意图;

[0039] 图3A-图3C为本公开实施例中智能设备控制显示设备呈现图形用户界面的示意图;

[0040] 图3D为本公开实施例中显示设备上呈现的视频交互界面示意图;

[0041] 图4为本公开实施例中,显示设备、智能设备以及云服务器实现视频时同声翻译的交互示意图;

[0042] 图5为本公开实施例中,智能设备的逻辑结构示意图;

[0043] 图6为本公开实施例中,显示设备的逻辑结构示意图。

具体实施方式

[0044] 为了使本技术领域的人员更好地理解本公开中的技术方案,下面将结合本公开实施例中的附图,对本公开实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本公开一部分实施例,而不是全部实施例。基于本公开中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都应当属于本公开保护的范围。

[0045] 为了解决现有技术中存在的无法在保证音视频同步的基础上,实现将同声翻译应

用于在智能电视上进行视频通话的场景中,同时为了保证在显示设备上进行视频时的交互质量以及对于目标语言的翻译效果,本申请提出了一种智能设备及视频通话时的同声翻译方法。

[0046] 图1A中示例性示出了智能设备、显示设备以及云服务器之间操作场景的示意图。如图1A所示,控制装置100和显示设备200之间可以采用有线或无线方式进行通信,控制装置100与智能设备300之间也可以通过有线或无线方式进行通信。

[0047] 智能设备300与显示设备200采用有线连接,并采用USB视频类(USB Video Class, UVC)协议格式,或者,实时流传输协议(Real Time Streaming Protocol, RTSP)格式传输视频流和音频流。所述智能设备300与云服务器400之间采用无线通信连接,传输的视频流和音频流采用RTSP格式,所述智能设备300内部的组件间指令的传输方式可以采用socket套接字格式。

[0048] 需要说明的是,目标交互用户侧的显示设备201、智能设备301,以及云服务器400之间的连接关系和处理方式与所述发起视频交互的用户侧的各个设备之间的连接相同,在此不再赘述。

[0049] 控制装置100一方面被配置为控制显示设备200,如:用户通过操作控制装置100上频道加减键,显示设备200响应频道加减的操作,所述控制装置100另一方面被配置为控制智能设备300,智能设备300控制显示设备200显示图形用户界面,并接收用户基于所述图形用户界面做出的选择或输入操作。

[0050] 控制装置100可以是遥控器100A,包括红外协议通信或蓝牙协议通信,及其他短距离通信方式等,通过无线或其他有线方式来控制智能设备300。用户基于显示设备200上呈现的图形用户界面,可以通过遥控器上按键、语音输入、控制面板输入等输入用户指令,来控制智能设备300。如:用户可以通过遥控器上音量加减键、频道控制键、上/下/左/右的移动按键、语音输入按键、菜单键、开关机按键等输入相应控制指令,来实现控制智能设备300的功能。

[0051] 控制装置100也可以是诸如移动终端100B、平板电脑、计算机、笔记本电脑等终端设备。例如,使用在终端设备上运行的应用程序控制智能设备300。该应用程序通过配置在终端设备关联的屏幕上,通过直观的用户界面(UI)为用户提供各种控制。

[0052] 示例性的,移动终端100B可与智能设备300安装软件应用,通过网络通信协议实现连接通信,实现一对一控制操作的和数据通信的目的。如:可以使移动终端100B与智能设备300建立控制指令协议,通过操作移动终端100B上提供的用户界面的各种功能键或虚拟按钮,来实现如遥控器100A布置的实体按键的功能。

[0053] 显示设备200,可以是液晶显示器、有机发光显示器、投影设备。具体显示设备类型、尺寸大小和分辨率等不作限定。

[0054] 智能设备300与显示设备200通过多种通信方式进行数据通信。这里可允许通过有线连接、局域网(Local Area Network, LAN)、无线局域网(Wireless Local Area Networks, WLAN)等方式进行通信连接。智能设备300基于用户通过控制装置100发起的操作指示,获取本地采集的第一视频数据和第一音频数据,并借由云服务器400获取进行视频交互的目标交互用户侧采集的第二视频数据和第二音频数据,以及字幕数据。进而将字幕数据与第二视频数据进行叠加处理,并对处理后的第二视频数据和所述第一视频数据进行拼

接处理。

[0055] 云服务器400,与智能设备采用局域网(LAN)或者无线局域网(WLAN)进行通信连接。所述云服务器400负责接收智能设备发送的音频数据和视频数据,并将其中的音频数据翻译为目标语言后,生成字幕数据下发相关用户。

[0056] 在本公开的一些实施例中,一台智能设备300控制管理一台显示设备200,智能设备300用于接收用户基于显示设备200显示的图形用户界面选择发起的视频交互指示,生成视频交互指令,以及接收本地采集的第一音频数据和第一视频数据,并将所述视频交互指令以及所述第一音频数据和所述第一视频数据,发送至云服务器400,由云服务器400获取目标交互用户侧采集的第二视频数据和第二音频数据,并在云服务器400将所述第二音频数据翻译为目标语言得到字幕数据后,所述智能设备300接受所述云服务器400发送的字幕数据以及所述第二音频数据和第二视频数据后,进而将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二音频数据发送至所述显示设备200,这样,智能设备300能够使显示设备200得到的视频数据和音频数据是同步的,避免出现音视频数据之间的延时,而且能够有效保护用户隐私。

[0057] 在本公开的另一些实施例中,一台智能设备300能够管理在一定空间内能够相互信任的不同显示设备200,如,同一家庭或同一公司的不同显示设备200可以共用一台智能设备300。用户可以选择在不同显示设备200上显示视频交互画面,智能设备300在接收到用户基于显示设备上显示的图形用户界面进行的操作指示后,生成相应的视频交互指令,并接收本地采集的第一音频数据和第一视频数据,进而将所述视频交互指令、所述第一音频数据以及所述第一视频数据,发送至云服务器400,由云服务器400获取目标交互用户侧采集的第二视频数据和第二音频数据,并在所述云服务器400将所述第二音频数据翻译为目标语言得到字幕数据后,所述智能设备300接收所述云服务器400发送的字幕数据以及所述第二视频数据和第二音频数据后,进而将所述第二视频数据与所述字幕数据进行叠加处理,并将处理后的第二视频数据与第一视频数据进行图像拼接处理得到的合成视频数据,将所述合成视频数据与第二音频数据同步发送至显示设备200,这样,智能设备300能够使显示设备200得到的视频数据和音频数据是同步的,避免出现音视频数据之间的延时,并通过为能够相互信任的显示设备配置共用同一智能设备,能够有效降低配置成本。

[0058] 本公开的一些实施例中,智能设备300包括处理器。

[0059] 所述处理器用于,

[0060] 接收用户的操作指示,控制显示设备显示相应的图形用户界面,基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令,获取第一视频数据和第一音频数据;

[0061] 将所述第一音频数据、所述第一视频数据以及所述视频交互指令发送至云服务器,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据;

[0062] 接收所述云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,所述字幕数据是将所述第二音频数据翻译为目标语言后得到的;

[0063] 将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二

音频数据发送至所述显示设备。

[0064] 需要说明的是所述智能设备200可以通过调用可交互的其他摄像头102和麦克风103,获取所述摄像头102和麦克风103采集的第一音频数据和第一音频数据。

[0065] 在本公开的另一些实施例中,智能设备200本身可以配置有可移动的摄像头102和麦克风103,其中,所述摄像头102可以灵活安放于显示设备200顶部的中央位置,也可以安放于显示设备200的其他位置,或者,安放在显示设备200之外能够采集用户的视频数据的位置,用于拍摄发起视频交互指示的用户产生的第一视频数据;所述麦克风103可以灵活安放于便于采集用户的音频数据的位置,用于采集发起视频交互指示的用户产生的第一音频数据。

[0066] 本公开的另一些实施例中,参阅图1B所示,智能设备300各组件按照实现的功能不同,划分为模块,分别包括视频交互承载模块、业务指令收发模块、业务控制模块、音视频数据收发模块、音视频数据采集模块、数据处理模块、接口模块以及通信模块。

[0067] 其中,所述业务指令收发模块用于接收用户基于显示设备200上显示的图形用户界面,通过选择和输入操作进而发起的视频交互指示,并向所述显示设备200传递相应的控制指令;

[0068] 所述业务控制模块,用于控制音视频数据收发模块获取音视频数据采集模块采集的第一视频数据以及第一音频数据,并控制将处理完成的合成视频数据以及第二音频数据经由接口模块发送至显示设备200,并控制音视频数据收发模块发送第一音频数据和第一视频数据,以及接收第二视频数据、第二音频数据,以及字幕数据。

[0069] 所述数据处理模块,用于将字幕数据与第二视频数据进行叠加,并将叠加处理后的第二视频数据与第一视频数据进行图像拼接;

[0070] 所述音视频数据采集模块,用于采集本地发起视频交互的用户产生的第一音频数据以及第一视频数据。

[0071] 下面结合附图2,对本公开的一些实施例中,显示设备200与智能设备300之间的交互过程进行说明:

[0072] 步骤201:智能设备接收用户的操作指示,控制显示设备显示相应的图形用户界面,基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令。

[0073] 智能设备上预先安装有支持用户与其他用户进行视频交互的相关应用,并控制显示设备呈现所述相关应用的图形用户界面,具体的,所述智能设备控制所述显示设备显示提示所述用户选择进行视频交互的目标交互用户的图形用户界面,以及控制所述显示设备显示提示用户选择开启同声翻译功能以及选择目标语言的图形用户界面。

[0074] 进一步的,所述智能设备确定所述用户在对应的图形用户界面上完成目标交互用户和目标语言的选择并确认发起视频交互的操作时,确定所述用户发起视频交互指示。

[0075] 本公开的一些实施例中,所述智能设备响应于所述用户基于所述显示设备上显示的图形用户界面进行的选择操作,确定所述用户开启同声翻译功能、确定所述用户选择的目标语言以及确定所述用户选择的目标交互用户后,进一步的,在确定所述目标语言存在于预设的有效语言列表中时,至少基于所述用户选择的目标交互用户的ID信息以及开启同声翻译时选择的目标语言信息生成视频交互指令,其中,所述有效语言列表中包含云服务

器能够识别翻译的所有语言信息。

[0076] 本公开的另一实施例中,所述智能设备确定所述用户基于显示设备上的图形用户界面选择的目标语言未包含在预设的有效语言列表中时,生成选择目标语言的提示信息,并将所述提示信息发送至所述显示设备进行显示,等待所述用户重新选择目标语言。并在确定所述用户选择开启同声翻译功能并选择有效的目标语言,以及所述用户选择目标交互用户后,至少基于所述用户选择的目标交互用户的ID信息以及重新选择的目标语言信息生成视频交互指令。

[0077] 需要说明的是,所述智能设备控制将提示所述用户选择目标交互用户的图形用户界面显示在显示设备上后,支持所述用户在图形用户界面上输入用户标识信息,如身份标识(Identity,ID)信息,以确定目标交互用户,或者,支持所述用户在所述显示设备上呈现的关联用户列表中选择目标交互用户,其中,所述关联用户列表中包括与所述用户在进行视频交互的相关应用上预先建立好友关系的其他用户以及曾经与所述用户进行视频交互的其他用户。

[0078] 例如,参阅图3A-图3C所示,智能设备控制显示设备呈现的初始图形用户界面中,呈现了能够进行视频交互的相关应用的应用标识,确定用户通过遥控器操作点击所述应用标识后,令显示设备呈现图3B示意性展示的图形用户界面,并支持所述用户通过键入目标交互用户的ID信息,或者直接选择关联用户列表中呈现的其他用户的方式确定目标交互用户,进一步的,控制显示设备呈现图3C示意性展示的图形用户界面,提示用户选择开启同声翻译功能,并选择目标语言,确定接收所述用户指示开启同声翻译功能并选择目标语言后的确定操作后,确定所述用户发起了发起视频交互指示。

[0079] 步骤202:所述智能设备获取第一视频数据和第一音频数据。

[0080] 智能设备确定用户发起了视频交互指示后,通过视频采集装置和音频采集装置,控制采集所述用户的第一视频数据,以及所述用户的第一音频数据,其中,所述视频采集装置包括但不限于摄像头,所述音频采集装置包括但不限于麦克风。

[0081] 步骤203:所述智能设备将所述第一音频数据、第一视频数据以及视频交互指令发送至云服务器,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据。

[0082] 智能设备基于用户的选择操作生成视频交互指令后,将所述视频交互指令协同获取的第一音频数据以及第一视频数据发送至云服务器,触发所述云服务器基于所述视频交互指令中携带的目标交互用户ID信息以及目标语言信息,获取目标交互用户侧采集的第二音频数据和第二视频数据,令所述云服务器将所述第二音频数据翻译为目标语言后,得到字幕数据。

[0083] 需要说明的是,所述智能设备将视频交互指令发送至所述云服务器后,触发所述云服务器基于所述目标交互用户ID信息确定目标交互用户,进而使所述云服务器将视频交互请求下发至所述目标交互用户侧的其他智能设备,并使所述云端服务器在所述目标交互用户同意参与视频交互后,获取所述其他智能设备发送的所述目标交互用户侧采集的第二音频数据和第二视频数据。

[0084] 步骤204:所述智能设备接收所述云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,所述字幕数据是将所述第二音频数据翻译为目标语言后得到的。

[0085] 智能设备将第一音频数据和第一视频数据,以及视频交互指令发送至云服务器后,触发云服务器获取所述目标交互用户侧的其他智能设备上传的所述目标交互用户侧采集的第二视频数据和第二音频数据,进而在所述云服务器完成将第二音频数据翻译为目标语言并生成字幕数据后,接收所述云服务器发送的第二音频数据、第二视频数据以及所述字幕数据。

[0086] 这样,通过在云服务器上进行音频数据到字幕数据的翻译,解决了进行语义分析和翻译时需要强大算力支撑的问题,无需在本地架设大规模的处理设备即可实现视频交互。

[0087] 步骤205:所述智能设备将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二音频数据发送至所述显示设备。

[0088] 智能设备接收云服务器基于第二音频数据以及目标语言生成的字幕数据、第二视频数据以及第二音频数据后,将所述字幕数据叠加所述第二视频数据上,由于所述字幕数据是基于所述第二音频数据生成的,且第二视频数据与第二音频数据是同步采集的,故第二视频数据、第二音频数据,以及字幕数据之间时间是同步的,进而采用数据叠加技术,实现字幕数据与第二视频数据的叠加,其中,字幕数据与视频数据的叠加为本领域的成熟技术,在此不再赘述。

[0089] 进一步的,分别将叠加有字幕数据的第二视频数据,与第一视频数据中时间同步的两个图像帧进行图像拼接处理,合并为一个图像帧,最终基于处理后得到的各个图像帧得到合成视频数据。

[0090] 进一步的,将得到的合成视频数据以及所述第二音频数据同步通过预设的音视频接口发送至显示设备,控制所述显示设备基于所述合成数据进行显示,以及基于所述第二音频数据进行播放。

[0091] 例如,参阅图3D所示,智能设备将合成视频数据与第二音频数据发送至显示设备后,呈现如图3D所示的播放画面,在显示设备上实时显示用户与目标交互用户之间的带有字幕数据的视频通话画面。

[0092] 这样,智能设备可以通过预设的视频接口将合成视频数据发送至显示设备,同时,通过预设的音频接口将第二音频数据同步发送至显示设备,能够保证显示设备播放的视频数据和音频数据是同步的,避免出现音频数据与视频数据之间存在延时,影响用户体验的问题。

[0093] 参阅图4所示,下面结合附图对本公开在视频通话时的同声翻译过程进行说明:

[0094] 步骤401:智能设备接收用户的操作指示,控制显示设备显示相应的图形用户界面。

[0095] 智能设备300接收用户的操作指示,并确定需要呈现给所述用户的图形用户界面,并针对性的基于所述用户在对应的图像用户界面上进行的选择操作,确定需要切换的图形用户界面,并控制显示设备200显示相应的图形用户界面。

[0096] 步骤402:智能设备基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令。

[0097] 智能设备300基于用户的操作,控制显示设备200显示相应的图形用户界面,并确

定所述用户在对应的图形用户界面上选择的目标交互用户,以及确定所述用户在选择开启同声翻译功能时选择的目标语言。

[0098] 进一步的,在本公开的一些实施例中,所述智能设备300在确定所述目标语言存在于预设的有效语言列表中,即,确定云服务器400能够将音频数据翻译为所述目标语言时,至少基于所述目标交互用户的ID信息以及所述目标语言信息生成视频交互指令。

[0099] 在本公开的另一些实施例中,所述智能设备300确定得到的目标语言不存在于预设的有效语言列表中时,则生成选择目标语言的提示信息,以及将所述提示信息发送至显示设备200进行显示,等待所述用户重新选择目标语言,并基于重新选择的有效目标语言信息以及目标交互用户的ID信息生成视频交互指令。

[0100] 步骤403:智能设备获取本地采集的第一音频数据以及第一视频数据。

[0101] 智能设备300基于用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令后,调用摄像头102采集第一视频数据,以及调用麦克风103采集第一音频数据,同时接收所述第一视频数据以及所述第一音频数据。

[0102] 步骤404:智能设备发送第一音频数据、第一视频数据以及所述视频交互指令至云服务器。

[0103] 智能设备300获取本地采集的第一视频数据以及第一音频数据后,将所述第一视频数据、第一音频数据,携同生成的视频交互指令,通过局域网发送至云服务器400。

[0104] 步骤405:云服务器确定成功接收数据信息以及指令信息,以及确定目标交互用户同意进行视频交互。

[0105] 云服务器400确定成功接收智能设备300发送的视频交互指令以及第一视频数据和第二视频数据后,基于所述视频交互指令中携带的目标交互用户信息,向对应的目标交互用户发送视频交互请求。

[0106] 本公开的一些实施例中,所述云服务器400将视频交互请求经由目标交互用户侧的智能设备301,显示在目标交互用户侧的显示设备201上,并确定得到所述目标交互用户同意加入视频交互的指示后,触发智能设备301控制所述目标交互用户侧的麦克风采集第二音频数据,以及控制所述目标交互用户侧的摄像头采集第二视频数据,并使所述智能设备301接收所述第二音频数据和第二视频数据。

[0107] 本公开的另一些实施例中,所述云服务器400将视频交互请求经由目标交互用户侧的智能设备301发送至显示设备201后。确定所述目标显示设备在所述显示设备201上发起拒绝加入视频交互的指示后,直接将所述拒绝进行视频交互的信息反馈至智能设备300,经由所述智能设备300转发至显示设备200上显示,并触发所述智能设备300结束当前的视频交互。

[0108] 步骤406:云服务器接收所述目标交互用户侧采集的第二视频数据以及第二音频数据。

[0109] 云服务器400确定成功接收视频交互指令,获得目标语言信息以及目标交互用户信息后,以及确定目标交互用户同意进行视频交互时,获取所述目标交互用户侧的智能设备301上报的所述目标交互用户侧采集的第二音频数据和第二视频数据。

[0110] 步骤407:云服务器将所述第二音频数据翻译为目标语言后,生成字幕数据。

[0111] 云服务器400得到目标语言信息,以及目标交互用户侧采集的第二音频数据后,调

用语音翻译包,将所述第二音频数据翻译为目标语言,待翻译完成后生成与所述第二音频数据对应的字幕数据。

[0112] 步骤408:云服务器发送所述第二音频数据、第二视频数据以及所述字幕数据到智能设备。

[0113] 云服务器完成将第二音频数据翻译为目标语言的字幕数据后,将得到的字幕数据、目标交互用户侧采集的第二音频数据和第二视频数据,发送至智能设备300。

[0114] 本公开的一些实施例中,云服务器成功将第二音频数据翻译为目标语言,得到字幕数据后,则将所述字幕数据,携同所述第二音频数据和第二视频数据,发送至智能设备300。

[0115] 本公开的另一些实施例中,云服务器未能成功将第二音频数据翻译为目标语言,未能得到字幕数据时,则仅将所述第二音频数据以及所述第二视频数据发送至智能设备300。

[0116] 步骤409:智能设备将所述第二视频数据与所述字幕数据进行叠加处理。

[0117] 本公开的一些实施例中,智能设备300接收到云服务器400翻译得到的字幕数据后,按照所述字幕数据与第二音频数据的对应关系,以及第二音频数据与第二视频数据时间上的同步关系,将所述字幕数据叠加在第二视频数据上,其中,字幕数据的字体大小,字体颜色以及文字位置可灵活调整。

[0118] 本公开的另一些实施例中,智能设备300未接收到云服务器400发送的字幕数据时,直接执行步骤410限定的操作。

[0119] 步骤410:智能设备将叠加处理后的第二视频数据与所述第一视频数据进行图像拼接处理,生成合成视频数据。

[0120] 本公开的一些实施例中,智能设备300基于云服务器400发送的字幕数据,完成字幕数据与第二视频数据的叠加处理后,进一步的,将叠加有字幕数据的第二视频数据与本地采集的第一视频数据,以对应的图像帧为处理对象进行图像拼接,进而得到合成视频数据。

[0121] 本公开的另一些实施例中,智能设备300确定仅接收到所述云服务器400发送的所述第二视频数据以及第二音频数据时,直接采用图像拼接技术对所述第二视频数据以及所述第一视频数据进行图像拼接处理,生成合成视频数据。

[0122] 步骤411:智能设备将所述合成视频数据以及所述音频数据发送至显示设备。

[0123] 智能设备300处理得到合成视频数据后,将所述合成视频数据与第二音频数据同时发送至显示设备200,令所述显示设备200使用显示器基于所述合成视频数据进行显示,同步的,以及令所述显示设备200使用扬声器基于所述第二音频数据进行播放。

[0124] 基于同一发明构思,本公开实施例中,提供一种智能设备,参阅图5所示,至少包括:获取单元501、发送单元502、接收单元503和处理单元504,其中,

[0125] 获取单元501,接收用户的操作指示,控制显示设备显示相应的图形用户界面,基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令,获取第一视频数据和第一音频数据;

[0126] 发送单元502,将所述第一音频数据、所述第一视频数据以及所述视频交互指令发送至云服务器,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二

音频数据；

[0127] 接收单元503,接收所述云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,所述字幕数据是将所述第二音频数据翻译为目标语言后得到的；

[0128] 处理单元504,将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二音频数据发送至所述显示设备。

[0129] 基于同一发明构思,本公开实施例中,提供一种显示设备,参阅图6所示,至少包括:接收单元601和显示单元602,其中,

[0130] 接收单元601,接收基于用户的操作指示确定呈现的图形用户界面,进行显示；

[0131] 显示单元602,接收合成视频数据以及音频数据,并基于所述合成视频数据进行显示,以及基于所述音频数据进行播放,其中,所述合成视频数据是将目标交互用户侧采集的第二音频数据翻译为目标语言,生成字幕数据后,将所述字幕数据与所述目标交互用户侧采集的第二视频数据进行叠加处理,并将叠加处理后的第二视频数据与本地采集的第一视频数据进行图像拼接处理后得到的,所述音频数据为所述目标用户侧采集的第二音频数据。

[0132] 基于同一发明构思,本公开实施例提供一种存储介质,当所述存储介质中的指令由处理器执行时,使得所述处理器能够执行上述流程中智能设备实现的任一项视频通话时的同声翻译方法。

[0133] 基于同一发明构思,本公开实施例提供一种存储介质,当所述存储介质中的指令由处理器执行时,使得所述处理器能够执行上述流程中显示设备实现的任一项视频通话时的同声翻译方法。

[0134] 本公开中,接收用户的操作指示,控制显示设备显示相应的图形用户界面,基于所述用户在对应的图形用户界面上选择的目标交互用户以及目标语言生成视频交互指令,进而,获取第一视频数据和第一音频数据,并将所述第一声音频数据、所述第一视频数据以及所述视频交互指令发送至云服务器,触发所述云服务器获取所述目标交互用户侧采集的第二视频数据以及第二音频数据,然后,接收所述云服务器发送的字幕数据、所述第二音频数据以及所述第二视频数据,所述字幕数据时将所述第二音频数据翻译为目标语言后得到的,再将所述字幕数据叠加到所述第二视频数据上,对处理后的第二视频数据和所述第一视频数据进行图像拼接处理,生成合成视频数据,并将所述合成视频数据以及所述第二音频数据发送至所述显示设备。

[0135] 这样,通过从云服务器获取翻译后的字幕数据,有效降低了本地设备处理的难度,无需为实现音频数据转换为文本数据而架设大规模处理设备,且通过将合成视频数据与第二音频数据的同步配置,避免了音频数据与字幕数据之间的延时,一方面借由显示设备实现了视频交互,另一方面能够实现将同声翻译应用于日常的使用中。

[0136] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统、或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0137] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0138] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0139] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0140] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[0141] 显然,本领域的技术人员可以对本发明实施例进行各种改动和变型而不脱离本发明实施例的精神和范围。这样,倘若本发明实施例的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

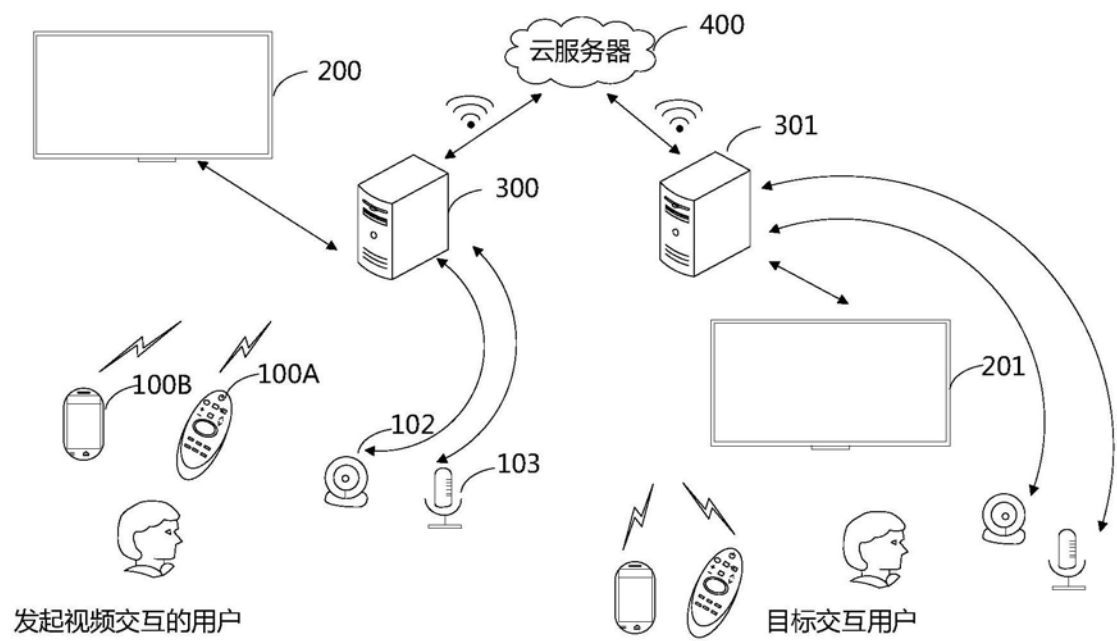


图1A

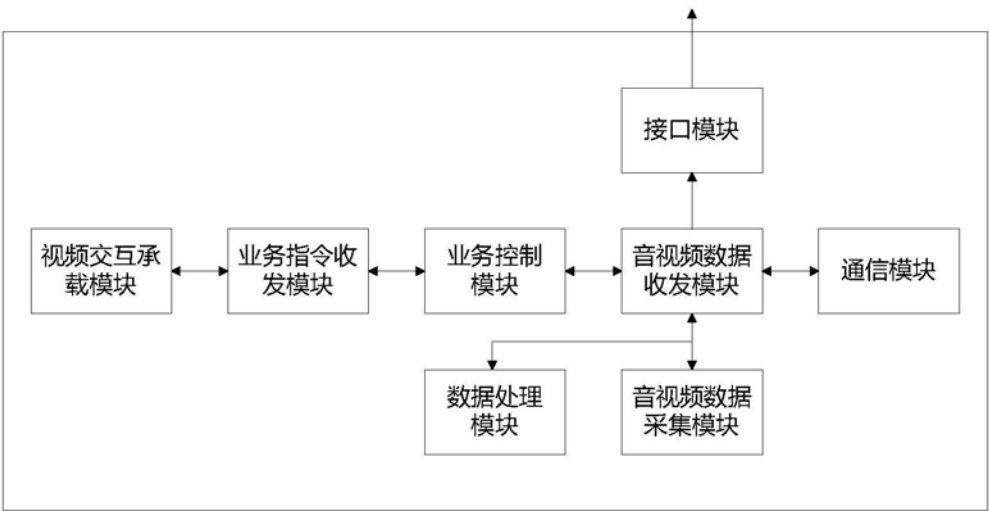


图1B

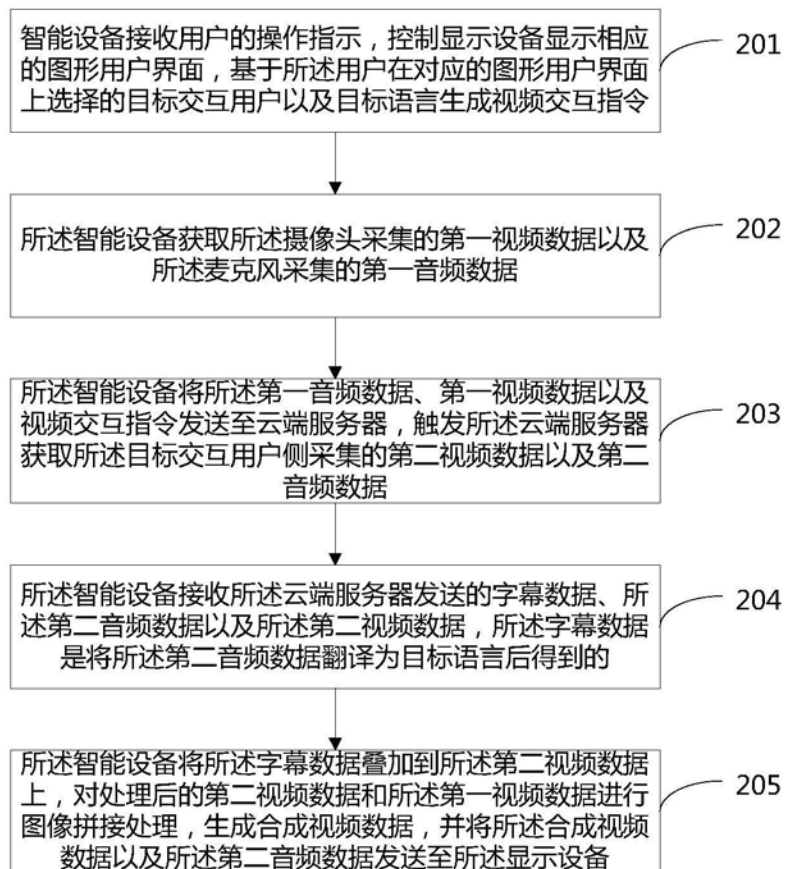


图2



图3A



图3B



图3C

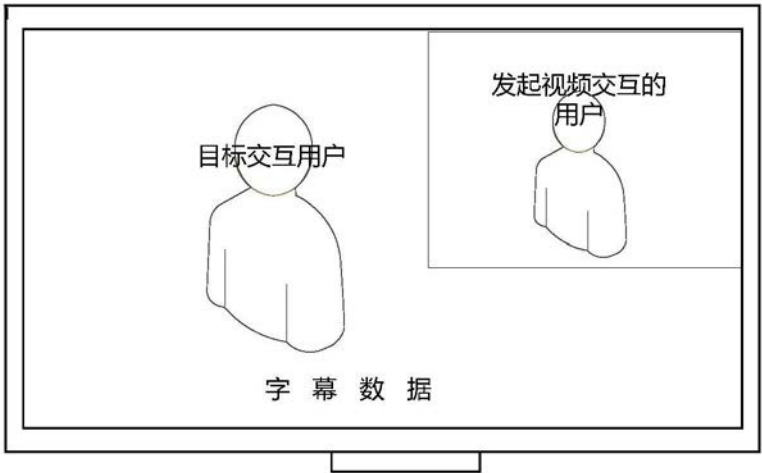


图3D

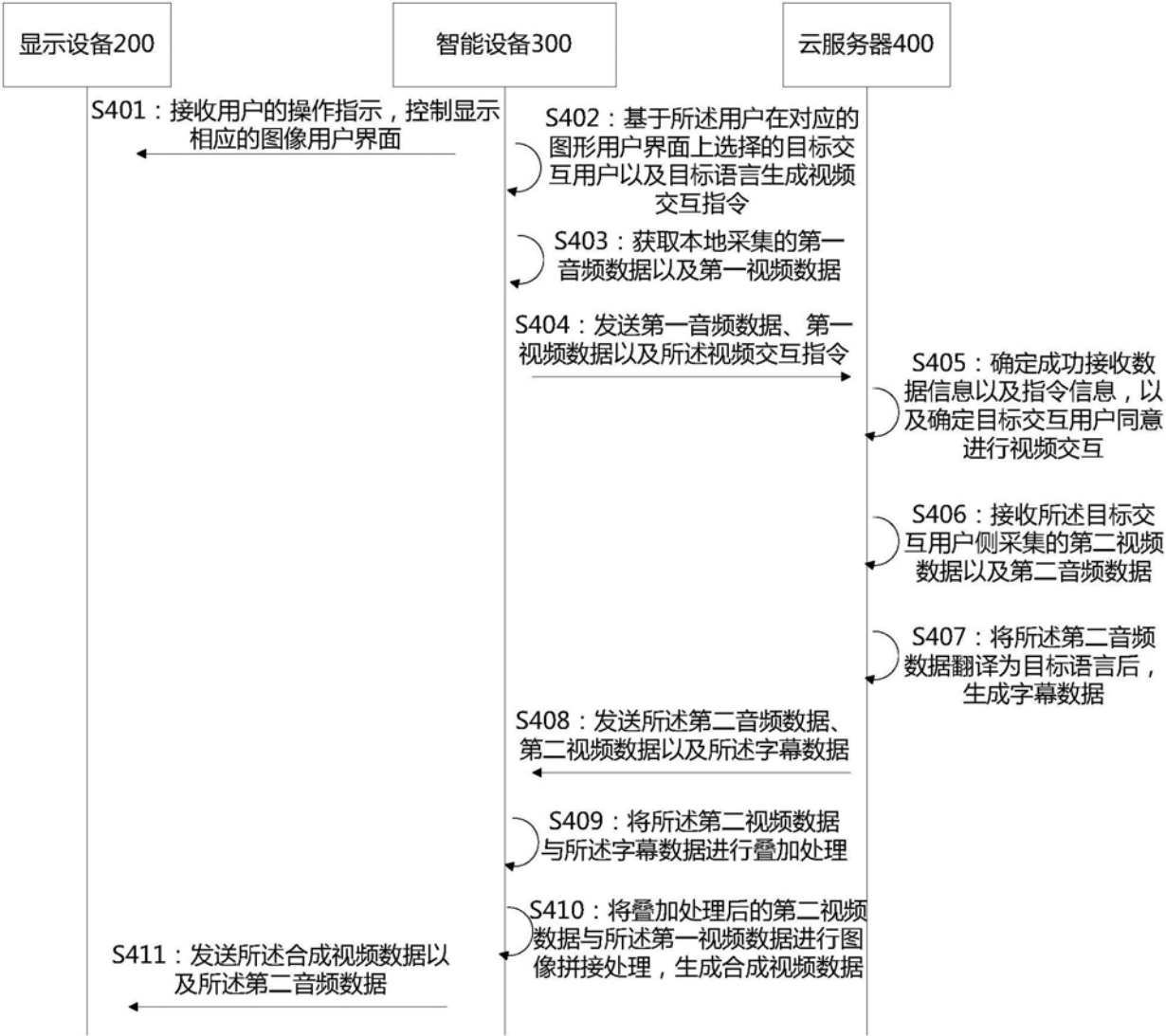


图4

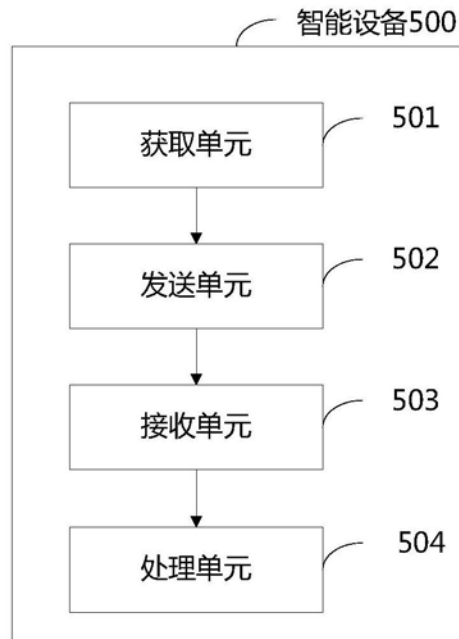


图5

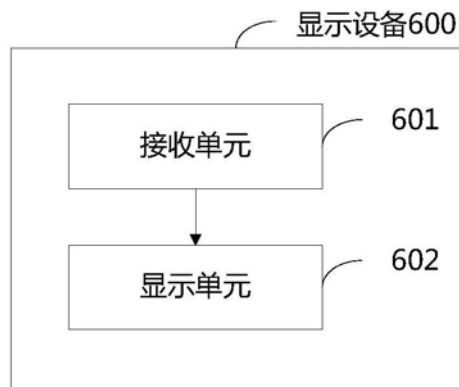


图6