



(12)发明专利申请

(10)申请公布号 CN 109166575 A

(43)申请公布日 2019.01.08

(21)申请号 201810848054.1

(22)申请日 2018.07.27

(71)申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号
百度大厦三层

(72)发明人 李士岩 李扬 吴淮

(74)专利代理机构 北京品源专利代理有限公司
11332

代理人 孟金喆

(51)Int.Cl.

G10L 15/22(2006.01)

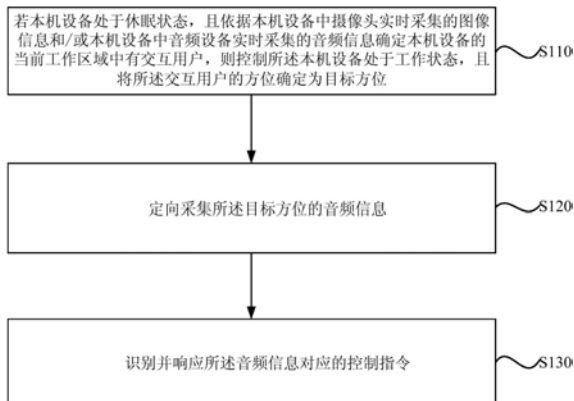
权利要求书2页 说明书11页 附图8页

(54)发明名称

智能设备的交互方法、装置、智能设备和存储介质

(57)摘要

本发明实施例公开了一种智能设备的交互方法、装置、智能设备和存储介质。该方法包括：若本机设备处于休眠状态，且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户，则控制所述本机设备处于工作状态，且将所述交互用户的方位确定为目标方位；定向采集所述目标方位的音频信息；识别并响应所述音频信息对应的控制指令。使得用户与智能设备之间的交互更加自然和拟人化，就像人与人之间的交流沟通一样，可以更自然的唤醒智能设备，提升用户体验和交互效率。同时，在目标方位上进行采集信息，得到的音频信息更准确，增加交互的准确和流畅性。



1. 一种智能设备的交互方法,其特征在于,包括:

若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位;

定向采集所述目标方位的音频信息;

识别并响应所述音频信息对应的控制指令。

2. 根据权利要求1所述的方法,其特征在于,依据本机设备中摄像头实时采集的图像信息确定本机设备的当前工作区域中有交互用户,包括:

若本机设备中摄像头实时采集的图像信息中包括用户正脸图像,且所述用户正脸图像的唇部状态发生变化,则将所述用户正脸图像所属的用户作为交互用户。

3. 根据权利要求1所述的方法,其特征在于,依据本机设备中摄像头实时采集的图像信息和本机设备中音频设备实时采集的音频信息,确定本机设备的当前工作区域中有交互用户,包括:

若依据本机设备中摄像头实时采集的图像信息确定本机设备的当前工作区域中不存在用户,则通过本机设备中音频设备实时采集环境音频信息;

若所述环境音频信息的分贝大于分贝阈值,且成功识别所述环境音频信息的语义,则将所述环境音频信息的发声人作为交互用户。

4. 根据权利要求1所述的方法,其特征在于,识别并响应所述音频信息对应的控制指令之前,还包括:

依据本机设备播放的音频信号的频率和/或振幅,对采集的音频信息做回声消除处理。

5. 根据权利要求1所述的方法,其特征在于,识别并响应所述音频信息对应的控制指令之后,还包括:

若依据摄像头实时采集的图像信息确定所述交互用户的位置移动,则控制摄像头旋转使所述摄像头朝向所述交互用户的正脸。

6. 根据权利要求1所述的方法,其特征在于,识别并响应所述音频信息对应的控制指令之后,还包括:

若依据实时采集的图像信息确定所述交互用户处于说话状态,则采集所述交互用户的新音频信息,识别且响应新音频信息对应的控制指令。

7. 根据权利要求1所述的方法,其特征在于,识别并响应所述音频信息对应的控制指令之后,还包括:

若在预设的时间长度内未接到交互用户的音频信息,且摄像头实时采集的图像信息中包括所述交互用户的侧脸图像;或者,摄像头实时采集的图像信息中没有所述交互用户的人脸图像,则控制本机设备切换到休眠状态。

8. 根据权利要求1所述的方法,其特征在于,识别并响应所述音频信息对应的控制指令,包括:

若在第一时间接收到所述目标方位的第一音频信息,则对在所述第一时间之前的预设时间长度内的第二音频信息以及所述第一音频信息进行识别。

9. 一种智能设备的交互装置,其特征在于,包括:

目标方位确定模块,用于若本机设备处于休眠状态,且依据本机设备中摄像头实时采

集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位;

音频信息采集模块,用于定向采集所述目标方位的音频信息;

控制指令响应模块,用于识别并响应所述音频信息对应的控制指令。

10.根据权利要求9所述的装置,其特征在于,所述目标方位确定模块包括图像识别单元,用于若本机设备中摄像头实时采集的图像信息中包括用户正脸图像,且所述用户正脸图像的唇部状态发生变化,则将所述用户正脸图像所属的用户作为交互用户。

11.一种智能设备,其特征在于,所述智能设备包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-8中任一所述的一种智能设备的交互方法。

12.一种包含计算机可执行指令的存储介质,所述计算机可执行指令在由计算机处理器执行时用于执行如权利要求1-8中任一所述的一种智能设备的交互方法。

智能设备的交互方法、装置、智能设备和存储介质

技术领域

[0001] 本发明实施例涉及计算机技术领域,尤其涉及一种智能设备的交互方法、装置、智能设备和存储介质。

背景技术

[0002] 随着互联网、人工智能、语音识别、语义判断的技术水平飞速发展,使得人机互动达到基本可用的状态,诸如智能机器人之类的智能设备广泛应用于各个领域。

[0003] 现有的智能设备唤醒方式主要有:(1)通过按压物理按键唤醒。由于每次的人机交互都按压物理键,以至于用户在语音的交互过程还需要手动操作,导致体验不佳,且交互不便。(2)通过说唤醒词唤醒。虽说比物理按键方式便捷,且在第一轮语音交互时没问题。但是在多轮的语音交互过程,每轮语音交互都需说唤醒词,这种交互方式是有违人们的自然语言沟通习惯,在用户感知上不够智能,在用户交互上繁琐重复。通过唤醒词进行智能硬件的唤醒或者通过物理按键的方式进行唤醒,这两种方式不智能且繁复,用户体验差。

发明内容

[0004] 本发明实施例提供一种智能设备的交互方法、装置、智能设备和存储介质,使得用户与智能设备的交互更加自然和拟人化,提升用户体验和交互效率。

[0005] 第一方面,本发明实施例提供了一种智能设备的交互方法,所述方法包括:

[0006] 若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位;

[0007] 定向采集所述目标方位的音频信息;

[0008] 识别并响应所述音频信息对应的控制指令。

[0009] 第二方面,本发明实施例还提供了一种智能设备的交互装置,所述装置包括:

[0010] 目标方位确定模块,用于若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位;

[0011] 音频信息采集模块,用于定向采集所述目标方位的音频信息;

[0012] 控制指令响应模块,用于识别并响应所述音频信息对应的控制指令。

[0013] 第三方面,本发明实施例还提供了一种智能设备,所述智能设备包括:

[0014] 一个或多个处理器;

[0015] 存储装置,用于存储一个或多个程序,

[0016] 当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如本发明任意实施例所述的一种智能设备的交互方法。

[0017] 第四方面,本发明实施例还提供了一种包含计算机可执行指令的存储介质,所述

计算机可执行指令在由计算机处理器执行时用于执行如本发明任意实施例所述的一种智能设备的交互方法。

[0018] 本发明实施例处于休眠状态的智能设备通过根据采集的图像和/或音频信息确定是否有交互用户,如果有交互用户,则进入工作状态并确定目标方位,在目标方位进行音频的采集和响应,使得用户与智能设备之间的交互更加自然和拟人化,就像人与人之间沟通交流一样,可以更自然的唤醒智能设备,提升用户体验和交互效率。同时,在目标方位上进行采集信息,得到的音频信息更准确,增加交互的准确和流畅性。

附图说明

- [0019] 图1是本发明实施例一中的一种智能设备的交互方法的流程图;
- [0020] 图2是本发明实施例二中的一种智能设备的交互方法的流程图;
- [0021] 图3是本发明实施例三中的一种智能设备的交互方法的流程图;
- [0022] 图4是本发明实施例四中的一种智能设备的交互方法的流程图;
- [0023] 图5是本发明实施例五中的一种智能设备的交互方法的流程图;
- [0024] 图6是本发明实施例六中的一种智能设备的交互方法的流程图;
- [0025] 图7是本发明实施例七中的一种智能设备的交互装置的结构示意图;
- [0026] 图8是本发明实施例七中的另一种智能设备的交互装置的结构示意图;
- [0027] 图9是本发明实施例八中的一种智能设备的结构示意图。

具体实施方式

[0028] 下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释本发明,而非对本发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与本发明相关的部分而非全部结构。

[0029] 实施例一

[0030] 图1为本发明实施例一提供的一种智能设备的交互方法的流程图,本实施例可适用于用户与机器人或智能设备与进行交互的情况,例如用户想要唤醒机器人或让机器人执行某种操作等,该方法可以由一种智能设备的交互装置来执行,该客装置可通过软件和/或硬件实现。如图1所示,该方法包括:

[0031] S110、若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位。

[0032] 其中,本机设备可以是具有通信能力、摄像功能、麦克风、图像识别和语音识别功能的智能设备。本机设备可以是用户交互的主体对象,例如智能机器人,用户可以通过与智能机器人进行交互,实现对智能机器人的控制,比如控制智能机器人执行信息查询或多媒体文件播放等。交互用户是要与本机设备进行交互对话的用户。

[0033] 在本实施例中,在处于休眠状态的过程中,本机智能设备的摄像头和/或音频设备保持实时采集信息,除了摄像头和音频设备之外的其他功能模块处于非工作状态。在处于工作状态过程中,本机设备上的硬件、应用程序或功能都可以运行或使用。当本机设备处于工作状态时,可以执行与用户进行交互,包括采集、识别以及响应用户的音频信息等操作。

[0034] 其中,工作区域是指本机设备能够与用户进行交互的区域范围,如果在对应的工作区域内有交互用户,则可以进行后续的信息采集和识别等操作。例如,可以将本机设备一定距离范围内的区域作为工作区域,不同的本机设备由于工作类别或本机设备中的硬件设备性能条件不同,工作区域范围可以不同。示例性的,可以预先设定本机设备采集信息的范围,例如可以预设采集5米内的图像和音频信息,机器人可以自主识别五米内有效人声并识别内容。当然,如果本机设备处于一个房间中,比如智能机器人,相应的,其工作区域便可以为该屋子的范围,如果识别到在该房间内有交互用户,则可以控制本机设备处于工作状态。另外,也可以将本机设备上摄像头的采集区域作为本机设备的工作区域。

[0035] 一般的,摄像头,可以是广角摄像头、可旋转摄像头或其他类型的摄像头,采集到的图像信息可以包括本机设备周围的环境图像;音频设备,比如麦克风采集到的音频信息可以包括本机设备周围的环境声音。

[0036] 示例性的,如果采集到的图像信息包括人脸图像和/或采集到的音频信息中包括人声,则可以确定当前工作区域内有交互用户,即可以识别为有用户想要跟本机设备进行交互,则控制本机设备处于工作状态,同时将所述交互用户的方位确定为目标方位。其中,交互用户的方位可以是用户的方向,具体的方位可以基于本机设备的位置为参考点确定。比如交互用户在本机设备的正北方位或东北方位等。目标方位即是交互用户所在的方位,以便于在目标方位与交互用户进行进一步的交互。

[0037] S120、定向采集所述目标方位的音频信息。

[0038] 其中,定向采集可以是在目标方位上对音频信息进行采集,即可以只对某一个方位上的音频信息采集。定向采集信息在一定程度上可以减少本机设备采集到的信息量,同时在与用户交互过程中采集到的无关的环境声音、噪音等也会减少,相应的,减小识别音频信息的计算量。

[0039] S130、识别并响应所述音频信息对应的控制指令。

[0040] 上述定向采集音频信息可以减少本机设备在对音频信息进行处理、识别过程中的计算量,从而提高交互速度和准确度。其中,控制指令可以是用于对本机设备进行相关操作对应的指令,比如关机、启动某项功能、进行某项操作等对应的指令。识别音频信息可以利用自动语音识别技术(Automatic Speech Recognition, ASR)对语音进行识别,将音频信息转化为文本信息。在识别出采集到音频信息时,可以对应该音频信息对应的控制指令进行响应或执行。示例性的,用户的音频信息识别结果为“打开音乐播放器1”,对应的控制指令可以是启动音乐播放器的相关操作,本机设备会响应于该指令,启动音乐播放器1。一般的,可以在数据库中预先设定相关的音频信息与控制指令之间的对应关系。

[0041] 可选的,在S130、识别并响应所述音频信息对应的控制指令之前,还包括:依据本机设备播放的音频信号的频率和/或振幅,对采集的音频信息做回声消除处理。

[0042] 一般情况下,采集到的环境音频信息中会包括发声者声音的回声,比如在一个比较空旷的空间时,如果声源发声,则会出现明显的回声,同时采集到的人声信息中便会包括回声,而回声严重时就会影响音频信息的识别。在本实施例中,将本机设备的扬声器播放的音频信息同步向音频采集设备传输,音频采集设备依据扬声器播放的音频信息对采集的音频信息进行回声消除。具体可以通过合理使用自适应算法和自适应滤波器来达到目的,以提高语音识别的效率和准确度。当然,对回声进行消除的同时,还可以对其与发声者声音无

关的环境噪音(或背景音)进行消除,进一步提高语音识别效率。

[0043] 本实施例的技术方案,在本机设备处于休眠状态时先根据采集的图像和/或音频信息确定是否有交互用户,如果有交互用户则进入工作状态进一步确定目标方位,在目标方位进行音频的采集和响应,使得用户与智能设备之间的交互更加自然和拟人化,就像人与人之间的交流沟通一样,可以更自然的唤醒智能设备,提升用户体验和交互效率。同时,在目标方位上进行采集信息,采集的音频更有定向性,得到的音频信息更准确,同时在对音频信息进行识别之前先对其进行回声消除,可以进一步增加识别以及交互的准确性和流畅性。

[0044] 实施例二

[0045] 图2是本发明实施例二中提供的一种智能设备的交互方法的流程图,在上述实施例的基础上,对如何确定在工作区域中无交互用户作了进一步的优化,如图2所示,该方法包括:

[0046] S210、若本机设备处于休眠状态,且本机设备中摄像头实时采集的图像信息中包括用户正脸图像,且所述用户正脸图像的唇部状态发生变化,则将所述用户正脸图像所属的用户作为交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位。

[0047] 在一种具体实施方式中,可以通过预先训练好的脸部分类模型或分类器根据输入的人脸图像信息来辨认摄像头采集到的图像信息中是否包括正脸图像。例如,在对采人脸图像进行识别时,可以通过对人脸68个特征点定位及识别,示例性的,在对识别到的人脸进行三次比对后可以确定交互用户的方位。识别唇部状态变化可以根据是否过度张嘴或长时无动作来判断,当然也可以利用预先训练好的唇部变化模型进行判断,根据输入的图像中的唇部部分确认唇部状态是否发生变化。如果唇部发生变化,则可以认为是用户在说话,此时便可以将该说话的正脸图像所属的用户确定为交互用户。

[0048] S220、定向采集所述目标方位的音频信息。

[0049] S230、识别并响应所述音频信息对应的控制指令。

[0050] 本实施例的技术方案,直接根据采集到的图像信息中包括用户正脸图像且唇部状态发生变化来确定有无交互用户,可以更加明确的确定交互用户,唤醒本机设备进入工作状态,以便在交互用户的方位进行音频信息采集和识别,实现交互用户与本机设备的准确交互。

[0051] 实施例三

[0052] 图3是本发明实施例三中提供的一种智能设备的交互方法的流程图,在上述实施例的基础上,对如何确定在工作区域中无交互用户作了进一步的优化,如图3所示,该方法包括:

[0053] S310、若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息确定本机设备的当前工作区域中不存在用户,则通过本机设备中音频设备实时采集环境音频信息。

[0054] 其中,依据图像信息确定本机设备的当前工作区域中不存在用户,可以是实时采集的图像信息中有没有人脸图像来判断,如果不包括人脸图像,则可以认为通过图像信息判断出工作区域内不存在交互用户,在此情况下,则通过进一步采集音频信息来判断是否

有交互用户。

[0055] S320、若所述环境音频信息的分贝大于分贝阈值，且成功识别所述环境音频信息的语义，则将所述环境音频信息的发声人作为交互用户，则控制所述本机设备处于工作状态，且将所述交互用户的方位确定为目标方位。

[0056] 其中，分贝阈值可以是预先设定的用于通过声音响度判断有无交互用户的数值。如果采集到的音频信息的分贝值大于分贝阈值，且成功识别到环境音频信息的语义，则可以将对应的发声人作为交互用户。具体的，该分贝阈值可以根据环境背景音的分贝值和/或可以接收到清晰人声的距离进行设定，示例性的，可以将环境背景音的分贝值设为阈值。其中，成功识别环境音频信息的语义是指识别出的音频信息具有一定的语义，比如“你好/您好/工作啦/……”等具有任何语义的信息。也就是说在音频信息高于分贝阈值且具有语义时，便可以将音频信息中对应的发声人作为交互用户，这样通过声音的响度以及语义情况对机器人进行唤醒，相比现有技术中，需要用户说出固定的唤醒词才可以对机器人进行唤醒，更加方便，并符合人与人之间的交流。示例性的，分贝值较高时可以表明该发声人距离本机设备较近或者发声人对着该本机设备发声，而这些情况可以说明用户与本机设备有较强的交互意图。所以通过分贝阈值和语义进行唤醒，一方面可以避免本机设备采集到一些分贝值较低的声音时实时的进行分析和响应，造成计算资源浪费的问题。另一方面，使得唤醒机器人的过程更加拟人化，提高了用户体验。

[0057] S330、定向采集所述目标方位的音频信息。

[0058] S340、识别并响应所述音频信息对应的控制指令。

[0059] 本实施例的技术方案，在依据图像信息判断不出交互用户的情况下，通过本机设备中音频设备实时采集环境音频信息来判断，若所述环境音频信息的分贝大于分贝阈值，且成功识别所述环境音频信息的语义，则将所述环境音频信息的发声人作为交互用户。使得确定交互用户唤醒本机设备的过程更加符合人与人之间的交流，提高了用户体验。

[0060] 实施例四

[0061] 图4是本发明实施例四中提供的一种智能设备的交互方法的流程图，在上述实施例的基础上，对该方法作了进一步的优化，如图4所示，该方法包括：

[0062] S410、若本机设备处于休眠状态，且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户，则控制所述本机设备处于工作状态，且将所述交互用户的方位确定为目标方位。

[0063] S420、定向采集所述目标方位的音频信息。

[0064] S430、识别并响应所述音频信息对应的控制指令。

[0065] S440、若依据摄像头实时采集的图像信息确定所述交互用户的位置移动，则控制摄像头旋转使所述摄像头朝向所述交互用户的正脸。

[0066] 其中，依据摄像头采集的图像信息确定所述交互用户的位置移动，具体可以是根据在预设时间内采集到的动态图像信息，比如视频信息，观察出交互用户的位置是否发生移动。如果发生移动，则控制摄像头旋转至摄像头朝向所述交互用户的正脸，以便于采集到交互用户的正脸图像信息或采集到更清楚的音频信息，便于对交互用户与本机设备的交互状态进一步做出准确识别和响应。

[0067] 本实施例的技术方案，通过在识别并响应所述音频信息对应的控制指令之后，若

依据摄像头实时采集的图像信息确定所述交互用户的位置移动,则控制摄像头旋转使所述摄像头朝向所述交互用户的正脸,可以采集交互用户的正脸图像。也就是说可以在交互方位的预定区域范围内进行动作追踪,实时更新发声用户的方位,并基于更新的交互用户的方位实时采集信息,提高交互用户的交互过程的识别效率。

[0068] 实施例五

[0069] 图5是本发明实施例四中提供的一种智能设备的交互方法的流程图,在上述实施例的基础上,对该方法作了进一步的优化,如图5所示,该方法包括:

[0070] S510、若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位。

[0071] S520、定向采集所述目标方位的音频信息。

[0072] S530、识别并响应所述音频信息对应的控制指令。

[0073] S540、若依据实时采集的图像信息确定所述交互用户处于说话状态,则采集所述交互用户的新音频信息,识别且响应新音频信息对应的控制指令。

[0074] 其中,在对一个指令进行识别并响应当前的音频信息对应的控制指令时或之后,可以实时的判断交互用户有没有新的指令发出。即用户可以实时打断与本机设备的交互过程,并发出新的信息指令。例如,如果根据实时采集到的图像信息中确定用户处于说话状态,说明用户有新的交互信息,则可以进一步对信息的音频信息进行采集、识别并响应。具体可以根据采集到图像是否包括正脸图像以及预设时间内的唇部状态是否在变化,来判断交互用户是否处于说话状态。如果包括正脸图像且在预设时间内唇部发生变化,则可以说明交互用户处于说话状态。

[0075] 本实施例的技术方案,可以通过在识别并响应所述音频信息对应的控制指令之后,若依据实时采集的图像信息确定所述交互用户处于说话状态,则采集所述交互用户的新音频信息,识别且响应新音频信息对应的控制指令。本机设备可以实时采集新的图像信息以确定用户是否在说话,采集交互用户新的音频信息,并作进一步的识别和响应。即用户可以实时打断与本机设备的交互过程,可以随时发出新的指令信息,本机设备可以直接继续对用户新发出的新的指令做出响应,而不需要再次进行唤醒。相比现有技术中,用户每发送一个指令之前都需要先说出固定的唤醒词进行唤醒,然后机器人才可以根据接收的新指令执行相应的操作的方案而言,本实施例的方案更加便捷、流畅和拟人化,进一步提高了用户体验。

[0076] 实施例六

[0077] 图6是本发明实施例四中提供的一种智能设备的交互方法的流程图,在上述实施例的基础上,对该方法作了进一步的优化,如图4所示,该方法包括:

[0078] S610、若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位。

[0079] S620、定向采集所述目标方位的音频信息。

[0080] S630、识别并响应所述音频信息对应的控制指令。

[0081] S640、若在预设的时间长度内未接到交互用户的音频信息,且摄像头实时采集的

图像信息中包括所述交互用户的侧脸图像;或者,摄像头实时采集的图像信息中没有所述交互用户的人脸图像,则控制本机设备切换到休眠状态。

[0082] 其中,侧脸图像可以通过预先训练好的脸部分类模型进行识别,比如可以识别出正脸和侧脸等姿态。如果在若在预设的时间长度内未接到交互用户的音频信息且检测图像为侧脸或无人脸图像,即该交互人5秒内侧脸无命令或检测无人脸无命令时,便可以认为该交互用户放弃了当前交互或不再进行下一步的交互操作,则可以控制本机设备切换到休眠状态。

[0083] 可选的,识别并响应所述音频信息对应的控制指令,包括:

[0084] 若在第一时间接收到所述目标方位的第一音频信息,则对在所述第一时间之前的预设时间长度内的第二音频信息以及所述第一音频信息进行识别。

[0085] 其中,第一时间可以是交本机设备在对采集到的交互用户的音频信息进行识别过程中的任意时刻。第一音频信息可以是用于表示用户有要打断当前与本机设备交互过程中意图的音频信息。示例性的,第一音频信息是如“停止/停一下/稍等/……”等类似表达停止当前交互的意图的信息。如果在交互用户与本机设备进行交互的过程中,接收到第一音频信息,则可以认为该交互用户暂时停止进一步发送后续的指令信息。第二音频信息可以是在第一时间之前的预设时间长度内采集到的音频信息。比如,如果出现打断后,识别前2秒内的有效语音内容。相比现有技术中,在遇到打断时,采集并识别打断时刻之前的音频信息的方案,本方案中根据发出打断指令时刻之前的预设时间长度内的音频信息进行识别,可以减少吃字情况的发生,增加语音识别的准确性。

[0086] 本实施例的技术方案,通过在识别并响应所述音频信息对应的控制指令之后,若在预设的时间长度内未接到交互用户的音频信息,且摄像头实时采集的图像信息中包括所述交互用户的侧脸图像;或者,摄像头实时采集的图像信息中没有所述交互用户的人脸图像,则控制本机设备切换到休眠状态。实现在本机设备对当前的控制指令响应后,如果在预设时长内没有接收到新的音频信息,则可以表示该交互用户没有新的指令或新的交互,可以控制本机设备切换到休眠状态,以降低本机设备的功耗。

[0087] 实施例七

[0088] 图7是本发明实施例七提供了一种智能设备的交互装置的结构示意图,适用于用户与智能设备之间的交互的情况,比如用户与机器人之间交互的情景,如图7所示,该装置包括:

[0089] 目标方位确定模块710,用于若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位;

[0090] 音频信息采集模块720,用于定向采集所述目标方位的音频信息;

[0091] 控制指令响应模块730,用于识别并响应所述音频信息对应的控制指令。

[0092] 可选的,所述目标方位确定模块包括图像识别单元,用于若本机设备中摄像头实时采集的图像信息中包括用户正脸图像,且所述用户正脸图像的唇部状态发生变化,则将所述用户正脸图像所属的用户作为交互用户。

[0093] 可选的,所述目标方位确定模块还包括音频识别单元,用于

[0094] 若依据本机设备中摄像头实时采集的图像信息确定本机设备的当前工作区域中不存在用户,则通过本机设备中音频设备实时采集环境音频信息;

[0095] 若所述环境音频信息的分贝大于分贝阈值,且成功识别所述环境音频信息的语义,则将所述环境音频信息的发声人作为交互用户。

[0096] 可选的,所述装置还包括:音频信息处理模块,用于在识别并响应所述音频信息对应的控制指令之前,依据本机设备播放的音频信号的频率和/或振幅,对采集的音频信息做回声消除处理。

[0097] 可选的,所述装置还包括:摄像头旋转模块,用于在识别并响应所述音频信息对应的控制指令之后,若依据摄像头实时采集的图像信息确定所述交互用户的位置移动,则控制摄像头旋转使所述摄像头朝向所述交互用户的正脸。

[0098] 可选的,所述装置还包括:新控制指令响应模块,用于在识别并响应所述音频信息对应的控制指令之后,若依据实时采集的图像信息确定所述交互用户处于说话状态,则采集所述交互用户的新音频信息,识别且响应新音频信息对应的控制指令。

[0099] 可选的,所述装置还包括:休眠状态切换模块,用于在识别并响应所述音频信息对应的控制指令之后,若在预设的时间长度内未接到交互用户的音频信息,且摄像头实时采集的图像信息中包括所述交互用户的侧脸图像;或者,摄像头实时采集的图像信息中没有所述交互用户的人脸图像,则控制本机设备切换到休眠状态。

[0100] 可选的,所述控制指令响应模块,还用于若在第一时间接收到所述目标方位的第一音频信息,则对在所述第一时间之前的预设时间长度内的第二音频信息以及所述第一音频信息进行识别。

[0101] 示例性的,本发明实施例所述的一种智能设备的交互装置可以运行于IntelX86、安卓、LINUX等系统架构下。图8是本实施例提供的另一种智能设备的交互装置的结构示意图,在具体实施的过程中,该装置还可以包括如下硬件平台、软件平台和系统平台,如图8所示,以实施本发明任意实施例所述的智能设备的交互方法:

[0102] 其中,硬件平台可以包括:

[0103] a) 核心模块,包括CPU+GPU+内存;

[0104] b) 视频模块:包括摄像头,比如全彩低照和/或红外摄像头;

[0105] c) 音频模块:音频SPDIF out+功放模块,其中,SPDIF (Sony/Philips DigitalInterface Format,数字音频接口);

[0106] d) 微控模块:用于控制LED和电机以实现驱动功能

[0107] e) 电动结构:包括动力传动结构与电机结构

[0108] f) 全向麦阵:例如CX20921双麦克风,可以支持180度拾音,SER可达-15dB;CX20924四麦克风,可以支持360度拾音,SER可达-30dB。

[0109] 软件平台可以包括:

[0110] g) 集成摄像头、麦阵、声音以及屏显及相关接口驱动。

[0111] h) 引入视频及数据处理库。

[0112] i) 应用程序框架能支持整体系统运行以及完成显示和外部控制接口定义。

[0113] j) 应用层输出。

[0114] 系统平台可以包括:

[0115] 1) openCV:用于提供原始图像采集、检测及处理,并提供机器视觉底层算法。

[0116] 2) dlib:用于实现图像识别算法,例如人脸细节和识物能力。

[0117] 3) MicArray:用于实现基础拾音和背景音消除以及对声源方向的识别。

[0118] 4) 通用接口总线:包括UART、I2C、GPIO和USBHUB等,用于对连接内外部设备以及数据接口协议支持。

[0119] 本发明实施例所提供的一种智能设备的交互装置,可执行本发明任意实施例所提供的一种智能设备的交互方法,具备执行方法相应的功能模块和有益效果。未在本实施例中详尽描述的技术细节,可参见本发明任意实施例提供的一种智能设备的交互方法。

[0120] 实施例八

[0121] 参见图9,本实施例提供了一种智能设备900,其包括:一个或多个处理器920;存储装置910,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器920执行,使得所述一个或多个处理器920实现本发明实施例所提供的一种智能设备的交互方法,包括:

[0122] 若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位;

[0123] 定向采集所述目标方位的音频信息;

[0124] 识别并响应所述音频信息对应的控制指令。

[0125] 当然,本领域技术人员可以理解,处理器920还可以实现本发明任意实施例所提供的一种智能设备的交互方法的技术方案。

[0126] 图9显示的智能设备900仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0127] 如图9所示,智能设备900以通用计算设备的形式表现。智能设备900的组件可以包括但不限于:一个或者多个处理器920,存储装置910,连接不同系统组件(包括存储装置910和处理器920)的总线950。

[0128] 总线950表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构 (ISA) 总线,微通道体系结构 (MAC) 总线,增强型ISA总线、视频电子标准协会 (VESA) 局域总线以及外围组件互连 (PCI) 总线。

[0129] 智能设备900典型地包括多种计算机系统可读介质。这些介质可以是任何能够被智能设备900访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0130] 存储装置910可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器 (RAM) 911和/或高速缓存存储器912。智能设备900可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统913可以用于读写不可移动的、非易失性磁介质(图9未显示,通常称为“硬盘驱动器”)。尽管图9中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如CD-ROM, DVD-ROM或者其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线950相连。存储装置910可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本发

明各实施例的功能。

[0131] 具有一组(至少一个)程序模块915的程序/实用工具914,可以存储在例如存储装置910中,这样的程序模块915包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块915通常执行本发明所描述的任意实施例中的功能和/或方法。

[0132] 智能设备900也可以与一个或多个外部设备960(例如键盘、指向设备、显示器970等)通信,还可与一个或者多个使得用户能与该智能设备900交互的设备通信,和/或与使得该智能设备900能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口930进行。并且,智能设备900还可以通过网络适配器940与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图9所示,网络适配器940通过总线950与智能设备900的其它模块通信。应当明白,尽管图中未示出,可以结合智能设备900使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0133] 处理器920通过运行存储在存储装置910中的程序,从而执行各种功能应用以及数据处理,例如实现本发明实施例所提供的一种智能设备的交互方法。

[0134] 实施例九

[0135] 本实施例提供一种包含计算机可执行指令的存储介质,所述计算机可执行指令在由计算机处理器执行时用于执行一种智能设备的交互方法,该方法包括:

[0136] 若本机设备处于休眠状态,且依据本机设备中摄像头实时采集的图像信息和/或本机设备中音频设备实时采集的音频信息确定本机设备的当前工作区域中有交互用户,则控制所述本机设备处于工作状态,且将所述交互用户的方位确定为目标方位;

[0137] 定向采集所述目标方位的音频信息;

[0138] 识别并响应所述音频信息对应的控制指令。

[0139] 当然,本发明实施例所提供的一种包含计算机可执行指令的存储介质,其计算机可执行指令不限于如上所述的方法操作,还可以执行本发明任意实施例所提供的一种智能设备的交互方法中的相关操作。

[0140] 本发明实施例的计算机存储介质,可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPR0M或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0141] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可

读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0142] 计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括——但不限于无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0143] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码,所述程序设计语言包括面向对象的程序设计语言——诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言——诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0144] 注意,上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解,本发明不限于这里所述的特定实施例,对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此,虽然通过以上实施例对本发明进行了较为详细的说明,但是本发明不仅仅限于以上实施例,在不脱离本发明构思的情况下,还可以包括更多其他等效实施例,而本发明的范围由所附的权利要求范围决定。

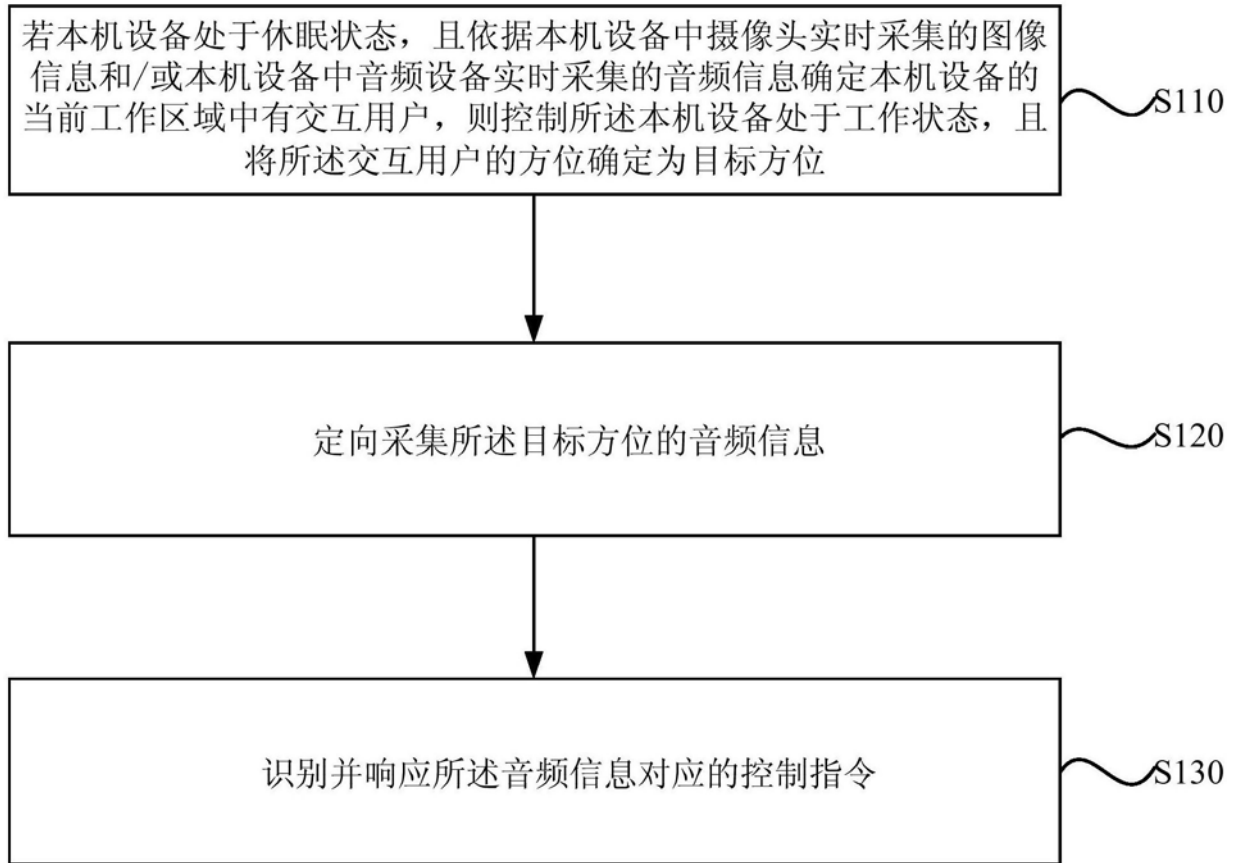


图1

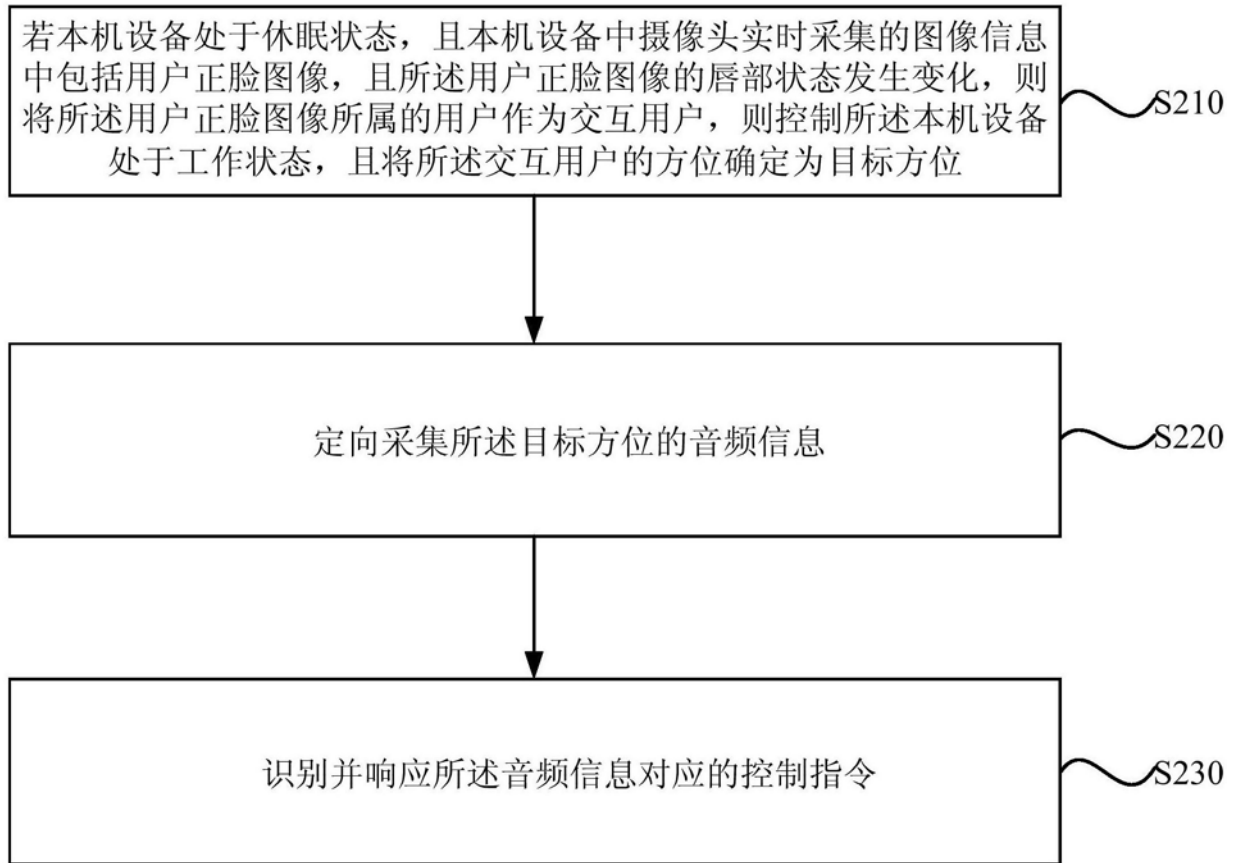


图2

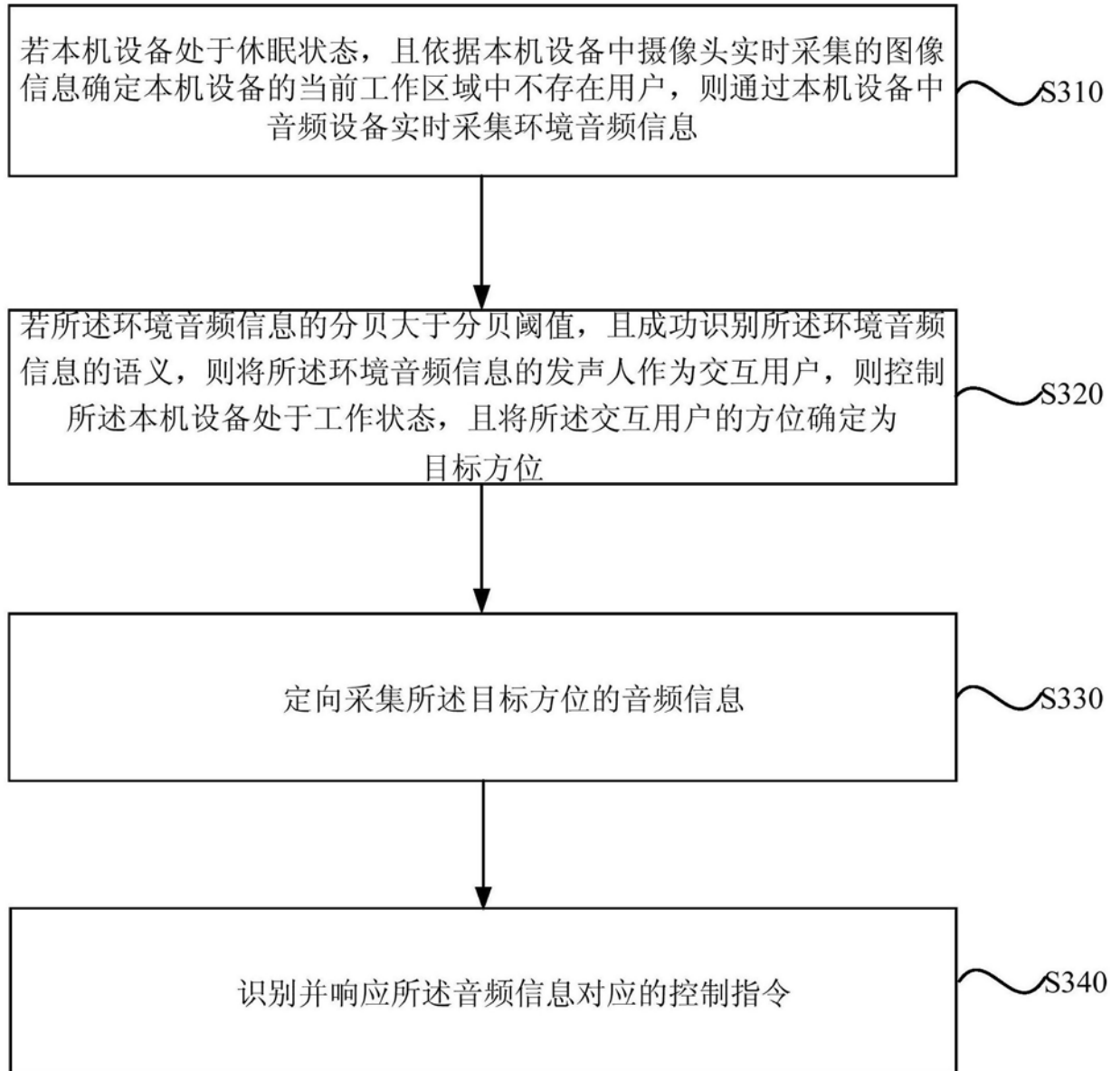


图3

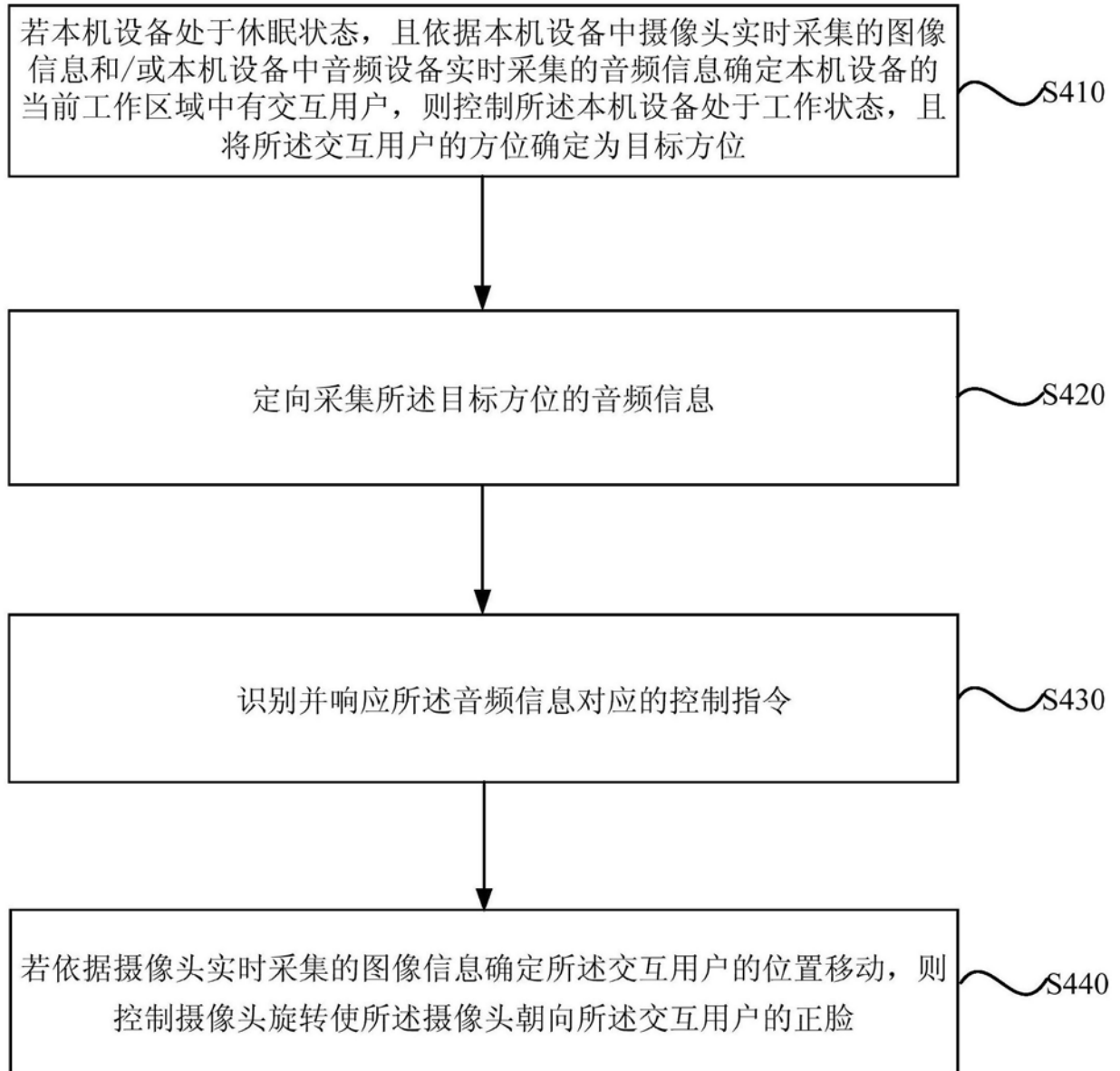


图4

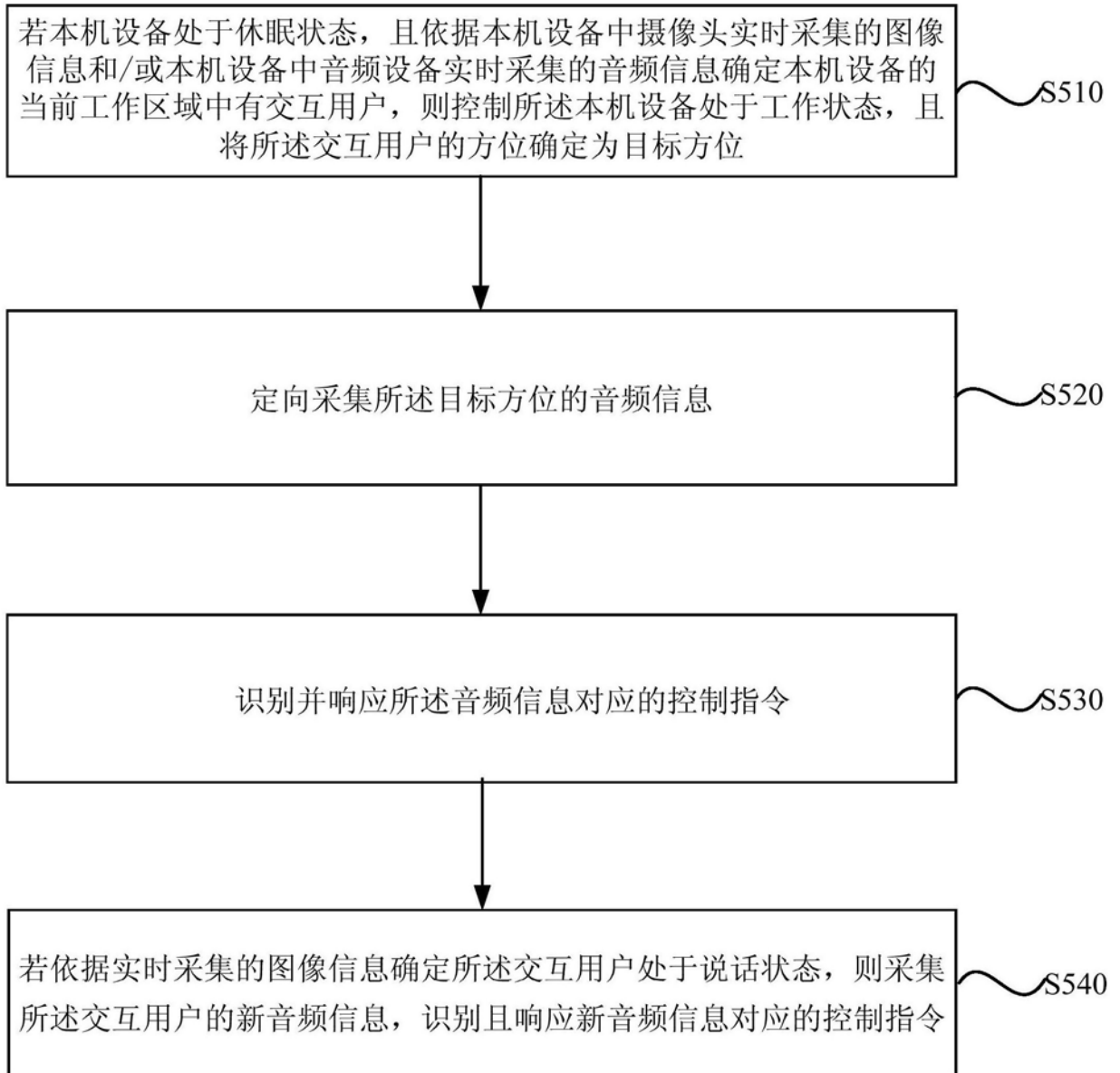


图5

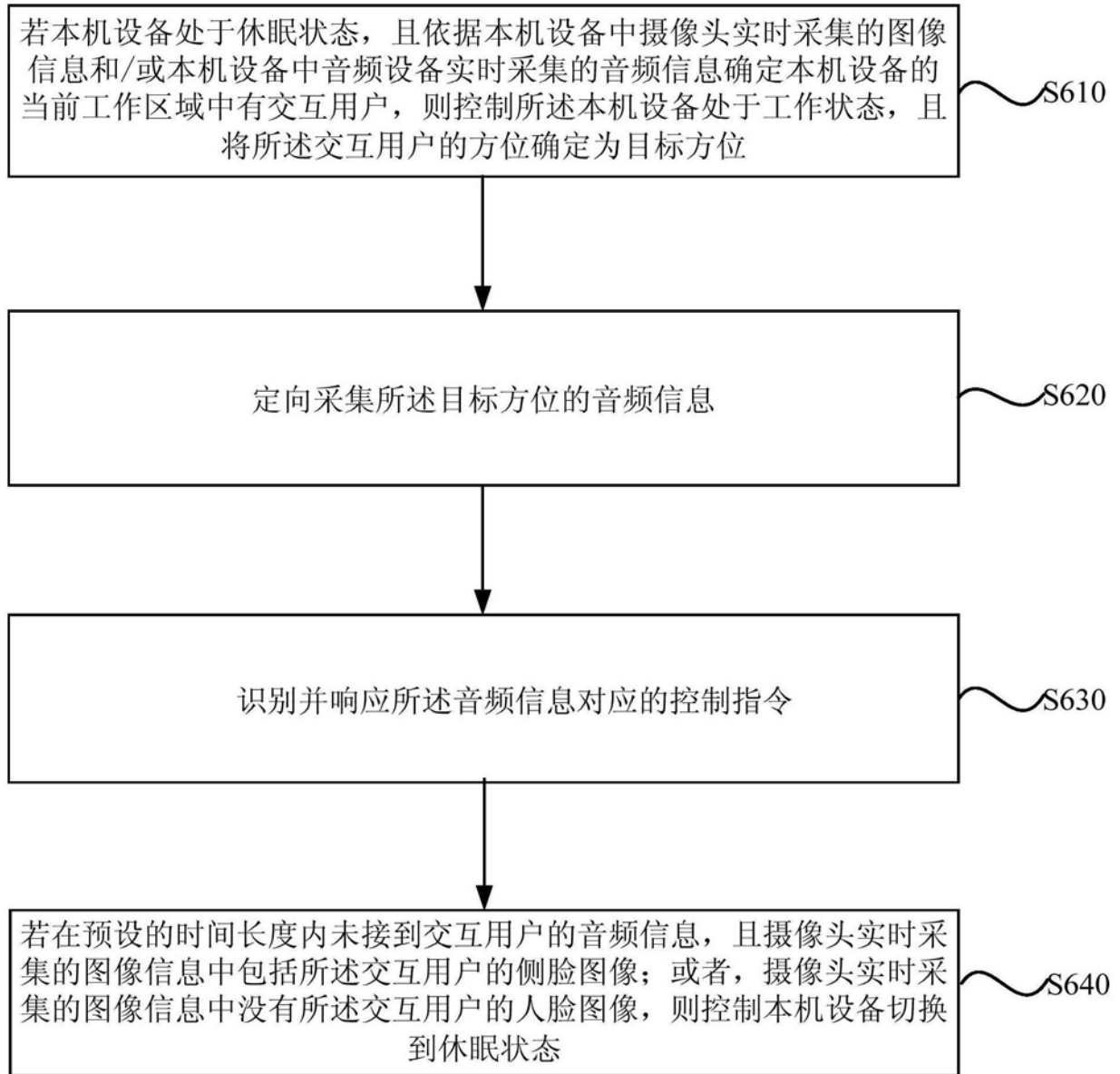


图6

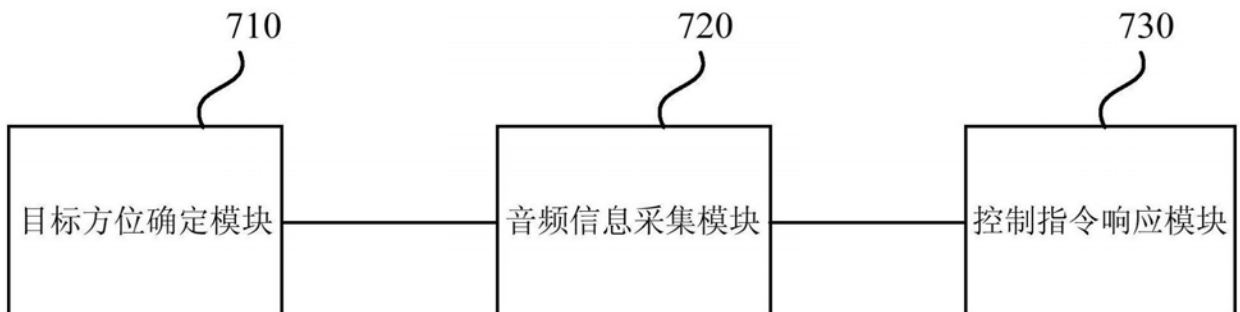


图7

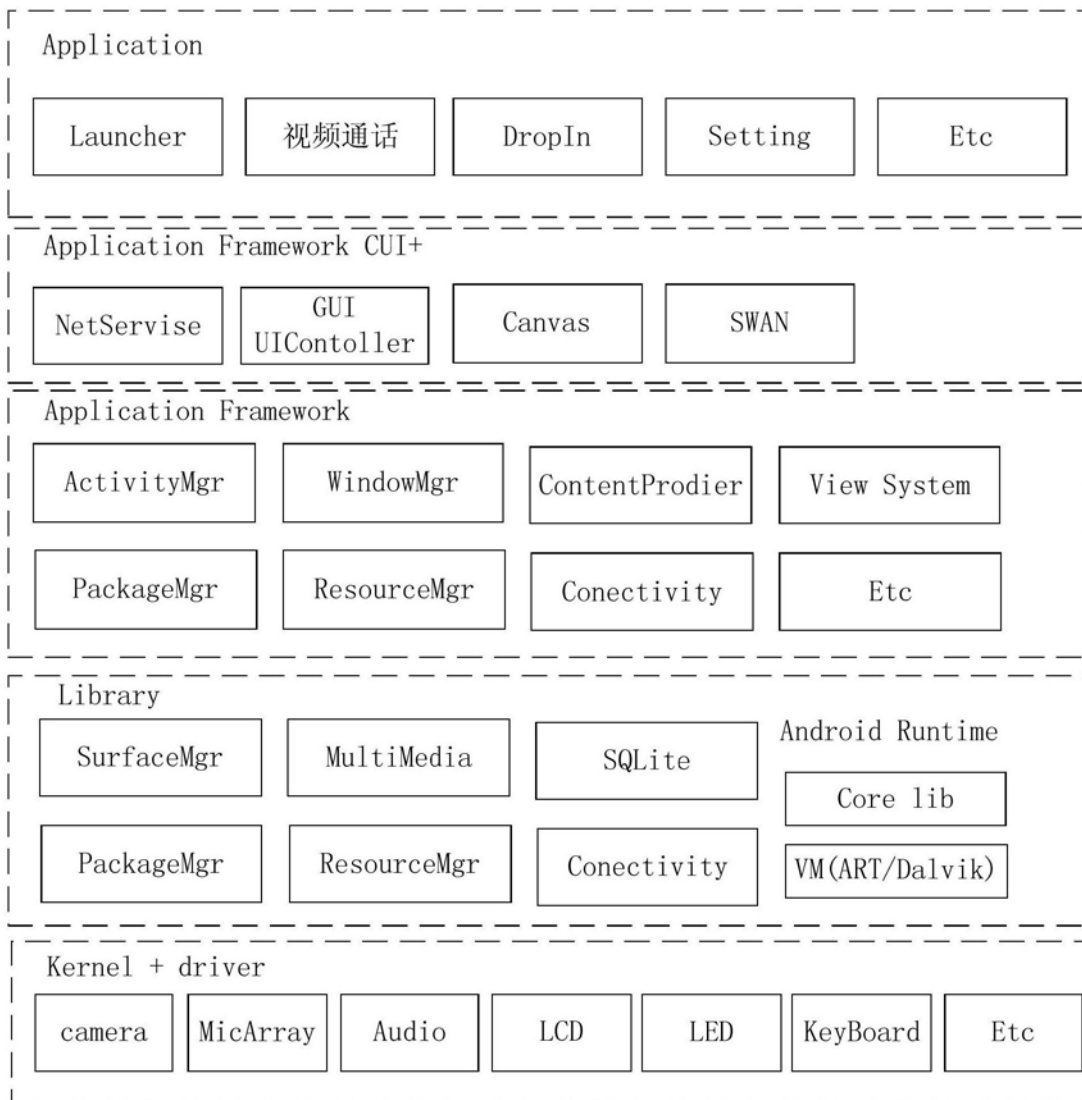


图8

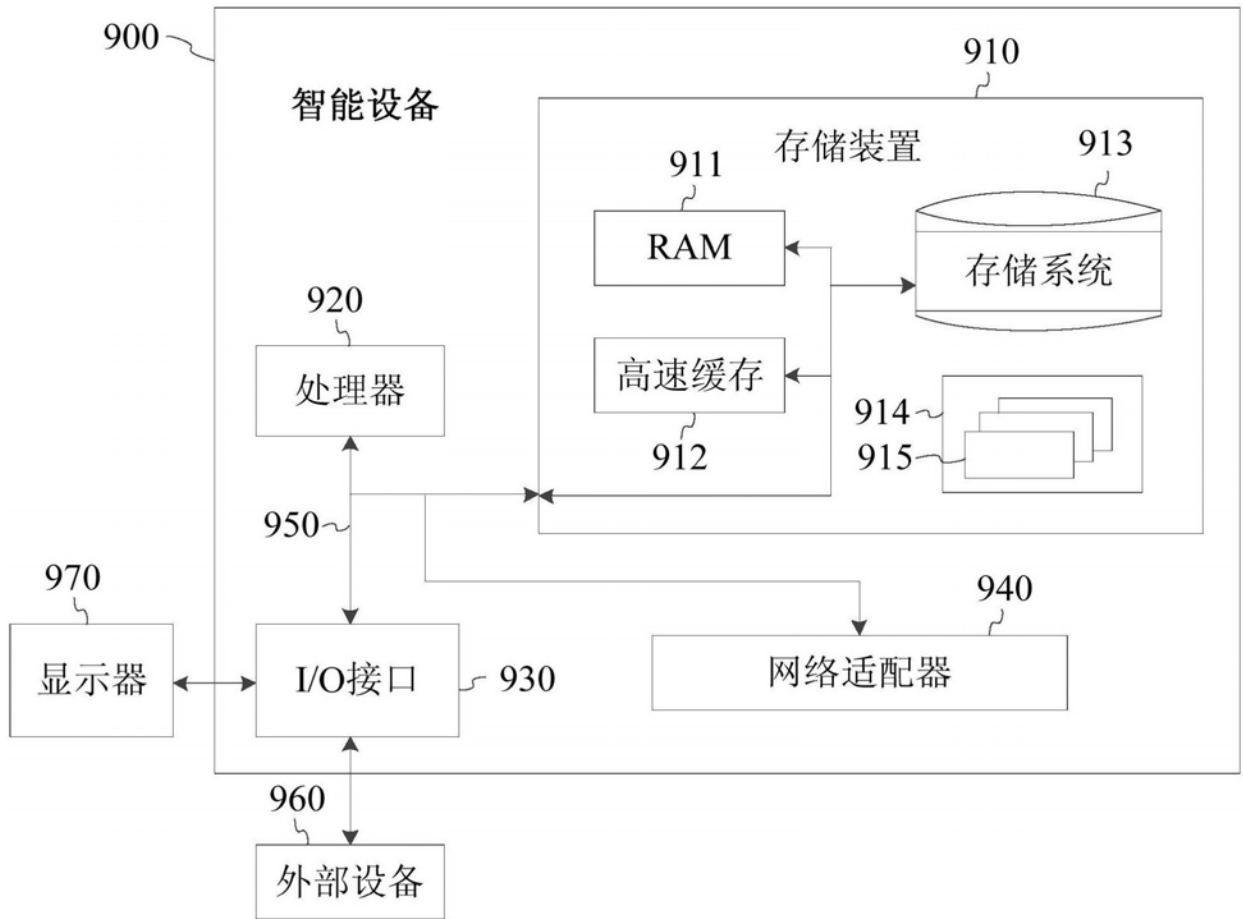


图9