



(12)发明专利

(10)授权公告号 CN 105159111 B

(45)授权公告日 2019.01.25

(21)申请号 201510523179.3

(22)申请日 2015.08.24

(65)同一申请的已公布的文献号  
申请公布号 CN 105159111 A

(43)申请公布日 2015.12.16

(73)专利权人 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 葛行飞 李峥 林汉权

(74)专利代理机构 北京清亦华知识产权代理事务  
所(普通合伙) 11201

代理人 宋合成

(51)Int.Cl.

G05B 15/02(2006.01)

(56)对比文件

CN 1494711 A,2004.05.05,  
CN 104777910 A,2015.07.15,

审查员 尹婷婷

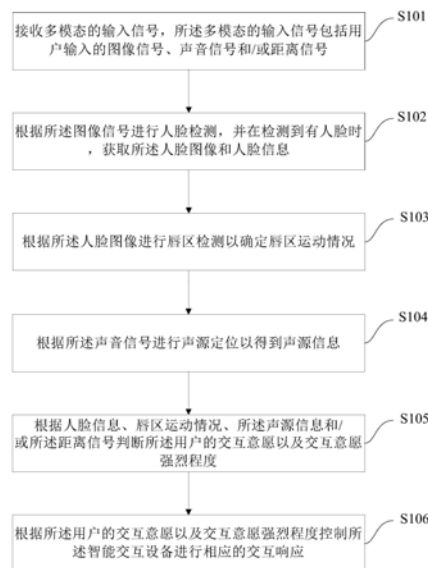
权利要求书4页 说明书10页 附图2页

(54)发明名称

基于人工智能的智能交互设备控制方法及系统

(57)摘要

本发明提出了一种基于人工智能的智能交互设备控制方法、系统及智能交互设备,该方法包括:接收多模态的输入信号,多模态的输入信号包括用户输入的图像信号、声音信号和/或距离信号;根据图像信号进行人脸检测,并在检测到有人脸时,获取人脸图像和人脸信息;根据人脸图像进行唇区检测以确定唇区运动情况;根据声音信号进行声源定位以得到声源信息;根据人脸信息、唇区运动情况、声源信息和/或距离信号判断用户的交互意愿以及交互意愿强烈程度;根据用户的交互意愿以及交互意愿强烈程度控制所述智能交互设备进行相应的交互响应。该方法能够提升用户与智能交互设备的交互体验,提升智能交互设备的智能性。



1. 一种基于人工智能的智能交互设备控制方法,其特征在于,包括以下步骤:

接收多模态的输入信号,所述多模态的输入信号包括用户输入的图像信号、声音信号和/或距离信号;

根据所述图像信号进行人脸检测,并在检测到有人脸时,获取人脸图像和人脸信息;

根据所述人脸图像进行唇区检测以确定唇区运动情况;

根据所述声音信号进行声源定位以得到声源信息;

根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,其中,所述交互意愿强烈程度包括:弱交互意愿、疑似交互意愿、强交互意愿、伴随交互意愿、强疑似交互意愿和弱疑似交互意愿;以及

根据所述用户的交互意愿以及交互意愿强烈程度控制所述智能交互设备进行相应的交互响应,其中,当所述用户具有弱交互意愿时,控制所述智能交互设备进行静默响应;当所述用户具有疑似交互意愿时,控制所述智能交互设备进行提高音量提示响应;当所述用户具有强交互意愿时,控制所述智能交互设备进行正式交互响应;当所述用户具有伴随交互意愿时,控制所述智能交互设备进行语音/聊天交互响应;当所述用户具有强疑似交互意愿时,控制所述智能交互设备转向声源方向并进行提示响应;当所述用户具有弱疑似交互意愿时,控制所述智能交互设备进行转向所述声源的响应。

2. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,在根据所述声音信号进行声源定位以得到声源信息之前,还包括:

判断所述声音信号中是否包含所述用户说话时的语音;

如果是,则保留所述声音信号中所述用户说话时的语音,并从所述声音信号中滤除其它的干扰噪音。

3. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,在根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度之前,还包括:

判断所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号是否满足预定条件;

如果满足所述预定条件,则执行用户的交互意愿以及交互意愿强烈程度的判断。

4. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,在根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度之前,还包括:对所述人脸信息和所述唇区运动情况进行量化处理。

5. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,还包括:

调整所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号的权重,其中,所述权重用于影响所述用户的交互意愿以及交互意愿强烈程度的判断结果;

所述判断所述用户的交互意愿以及交互意愿强烈程度,进一步包括:

根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号的权重判断所述用户的交互意愿以及交互意愿强烈程度。

6. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,所述人

脸信息包括人脸面积信息和人脸正面面对程度,所述声源信息包括声源方位信息和声音强度信息。

7. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,所述根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,包括:当判断所述用户正对所述智能交互设备、所述用户的嘴唇未运动、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有弱交互意愿。

8. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,所述根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,包括:当判断所述用户正对所述智能交互设备、所述用户的嘴唇产生运动、所述用户发声且声音强度小于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有疑似交互意愿。

9. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,所述根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,包括:当判断所述用户正对所述智能交互设备、所述用户的嘴唇产生运动、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有强交互意愿。

10. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,所述根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,包括:当判断所述用户侧面面对所述智能交互设备、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有伴随交互意愿。

11. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,所述根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,包括:当未检测到人脸图像、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有强疑似交互意愿。

12. 根据权利要求1所述的基于人工智能的智能交互设备控制方法,其特征在于,所述根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,包括:当未检测到人脸图像、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离大于预设距离时,判断所述用户具有弱疑似交互意愿。

13. 根据权利要求1-12任一项所述的基于人工智能的智能交互设备控制方法,其特征在于,所述根据所述人脸图像进行唇区检测以确定唇区运动情况,具体包括:根据多帧人脸图像之间的唇区形状差异确定所述唇区运动情况。

14. 一种基于人工智能的智能交互设备控制系统,其特征在于,包括:

接收模块,用于接收多模态的输入信号,所述多模态的输入信号包括用户输入的图像信号、声音信号和/或距离信号;

人脸检测模块,用于根据所述图像信号进行人脸检测,并在检测到有人脸时,获取人脸

图像和人脸信息;

唇区检测模块,用于根据所述人脸图像进行唇区检测以确定唇区运动情况;

声源定位模块,用于根据所述声音信号进行声源定位以得到声源信息;

决策模块,所述决策模块用于根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度,其中,所述交互意愿强烈程度包括:弱交互意愿、疑似交互意愿、强交互意愿、伴随交互意愿、强疑似交互意愿和弱疑似交互意愿;以及

复合输出控制模块,用于根据所述用户的交互意愿以及交互意愿强烈程度控制所述智能交互设备进行相应的交互响应,其中,当所述用户具有弱交互意愿时,控制所述智能交互设备进行静默响应;当所述用户具有疑似交互意愿时,控制所述智能交互设备进行提高音量提示响应;当所述用户具有强交互意愿时,控制所述智能交互设备进行正式交互响应;当所述用户具有伴随交互意愿时,控制所述智能交互设备进行语音/聊天交互响应;当所述用户具有强疑似交互意愿时,控制所述智能交互设备转向声源方向并进行提示响应;当所述用户具有弱疑似交互意愿时,控制所述智能交互设备进行转向所述声源的响应。

15. 根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,还包括:

语音活动检测模块,用于在所述声源定位模块根据所述声音信号进行声源定位以得到声源信息之前,判断所述声音信号中是否包含所述用户说话时的语音,如果是,则保留所述声音信号中所述用户说话时的语音,并从所述声音信号中滤除其它的干扰噪音。

16. 根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块还用于在根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度之前,判断所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号是否满足预定条件;如果满足所述预定条件,则执行用户的交互意愿以及交互意愿强烈程度的判断。

17. 根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块还用于在根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度之前,对所述人脸信息和所述唇区运动情况进行量化处理。

18. 根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块还用于:

调整所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号的权重,其中,所述权重用于影响所述用户的交互意愿以及交互意愿强烈程度的判断结果;

所述判断所述用户的交互意愿以及交互意愿强烈程度,包括:

根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号的权重判断所述用户的交互意愿以及交互意愿强烈程度。

19. 根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述人脸信息包括人脸面积信息和人脸正面面对程度,所述声源信息包括声源方位信息和声音强度信息。

20. 根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述

决策模块用于:当判断所述用户正对所述智能交互设备、所述用户的嘴唇未运动、所述用户发声且声音强度大于预定强度以及用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有弱交互意愿。

21.根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块用于:当判断所述用户正对所述智能交互设备、所述用户的嘴唇产生运动、所述用户发声且声音强度小于预定强度以及用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有疑似交互意愿。

22.根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块用于:当判断所述用户正对所述智能交互设备、所述用户的嘴唇产生运动、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断用户具有强交互意愿。

23.根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块用于:当判断所述用户侧面面对所述智能交互设备、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有伴随交互意愿。

24.根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块用于:当未检测到人脸图像、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有强疑似交互意愿。

25.根据权利要求14所述的基于人工智能的智能交互设备控制系统,其特征在于,所述决策模块用于:当未检测到人脸图像、所述用户发声且声音强度大于预定强度以及所述用户与所述智能交互设备之间的距离大于预设距离时,判断所述用户具有弱疑似交互意愿。

26.根据权利要求14-25任一项所述的基于人工智能的智能交互设备控制系统,其特征在于,所述唇区检测模块用于:根据多帧人脸图像之间的唇区形状差异确定所述唇区运动情况。

27.一种智能交互设备,其特征在于,包括:根据权利要求14-26任一项所述的基于人工智能的智能交互设备控制系统。

## 基于人工智能的智能交互设备控制方法及系统

### 技术领域

[0001] 本发明涉及智能终端技术领域,特别涉及一种基于人工智能(Artificial Intelligence,简称:AI)的智能交互设备控制方法、控制系统及智能交互设备。

### 背景技术

[0002] 现在的智能交互设备,如电视机、生活电器等,通常是采用遥控或者提前设定好的程序来执行相关的动作。这种通过遥控或者提前设定好的程序来执行相关动作的智能交互设备存在以下缺点:

[0003] 与人类之间的交互方式单一且互动性差,这是由于遥控操作功能有限,智能交互设备无法完成遥控操作功能以外的动作,同样,智能交互设备按照提前设定好的程序进行动作,也是存在无法完成设定程序以外的其它动作,不能针对不同用户需求进行不同的运动。另外,这些交互方式均是在用户遥控或者触发某个功能按键后进行的,因此,完全是被动的交互方式。

[0004] 虽然有一些如视频会议跟踪系统可以根据说话人的声音将摄像头等转向说话人,但是并不能够准确判断说话人是否存在交互意愿,也不能够根据交互意愿做成适当的反应。

### 发明内容

[0005] 本发明的目的旨在至少解决所述技术缺陷之一。

[0006] 为此,本发明的一个目的在于提出一种基于人工智能的智能交互设备控制方法。该方法能够提升用户与智能交互设备的交互体验,提升智能交互设备的智能性。

[0007] 本发明的另一个目的在于提出一种基于人工智能的智能交互设备控制方法。

[0008] 本发明的再一个目的在于提出一种智能交互设备。

[0009] 为达到上述目的,本发明的第一方面的实施例公开了一种基于人工智能的智能交互设备控制方法,包括以下步骤:接收多模态的输入信号,所述多模态的输入信号包括用户输入的图像信号、声音信号和/或距离信号;根据所述图像信号进行人脸检测,并在检测到有人脸时,获取所述人脸图像和人脸信息;根据所述人脸图像进行唇区检测以确定唇区运动情况;根据所述声音信号进行声源定位以得到声源信息;根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度;以及根据所述用户的交互意愿以及交互意愿强烈程度控制所述智能交互设备进行相应的交互响应。

[0010] 根据本发明实施例的基于人工智能的智能交互设备控制方法,能够实时的采集用户的语音信号、图像信号和/或距离信号,经过人工智能的分析后确定出用户是否存在交互意愿,并可以确定出交互意愿的强烈程度,然后自主地控制智能交互设备进行相应的动作,主动地与用户进行交互且交互手段丰富,进而提升用户的使用体验。

[0011] 本发明第二方面的实施例公开了一种基于人工智能的智能交互设备控制系统,包

括:接收模块,用于接收多模态的输入信号,所述多模态的输入信号包括用户输入的图像信号、声音信号和/或距离信号;人脸检测模块,用于根据所述图像信号进行人脸检测,并在检测到有人脸时,获取所述人脸图像和人脸信息;唇区检测模块,用于根据所述人脸图像进行唇区检测以确定唇区运动情况;声源定位模块,用于根据所述声音信号进行声源定位以得到声源信息;决策模块,所述决策模块用于根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度;以及复合输出控制模块,用于根据所述用户的交互意愿以及交互意愿强烈程度控制所述智能交互设备进行相应的交互响应。

[0012] 根据本发明实施例的基于人工智能的智能交互设备控制系统,能够实时的采集用户的语音信号、图像信号和/或距离信号,经过人工智能的分析后确定出用户是否存在交互意愿,并可以确定出交互意愿的强烈程度,然后自主地控制智能交互设备进行相应的动作,主动地与用户进行交互且交互手段丰富,进而提升用户的使用体验。

[0013] 本发明第三方面的实施例公开了一种智能交互设备,包括:根据上述的第二方面实施例所述的基于人工智能的智能交互设备控制系统。该智能的智能交互设备能够实时的采集用户的语音信号、图像信号和/或距离信号,经过人工智能的分析后确定出用户是否存在交互意愿,并可以确定出交互意愿的强烈程度,然后自主地控制智能交互设备进行相应的动作,主动地与用户进行交互且交互手段丰富,进而提升用户的使用体验。

[0014] 本发明附加的方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本发明的实践了解到。

## 附图说明

[0015] 本发明所述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0016] 图1是根据本发明一个实施例的基于人工智能的智能交互设备控制方法的流程图;以及

[0017] 图2是根据本发明一个实施例的基于人工智能的智能交互设备控制系统的结构框图;以及

[0018] 图3是根据本发明一个实施例的基于人工智能的智能交互设备控制系统的原理图。

## 具体实施方式

[0019] 下面详细描述本发明的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,仅用于解释本发明,而不能解释为对本发明的限制。

[0020] 在本发明的描述中,需要说明的是,除非另有规定和限定,术语“安装”、“相连”、“连接”应做广义理解,例如,可以是机械连接或电连接,也可以是两个元件内部的连通,可以是直接相连,也可以通过中间媒介间接相连,对于本领域的普通技术人员而言,可以根据具体情况理解所述术语的具体含义。

[0021] 为了解决相关技术中存在的智能交互设备智能性差且不能很好与人类进行交互

的问题,本发明基于人工智能实现了智能性高且与人类交互体验好的智能交互设备控制方法、控制系统及智能交互设备,其中,人工智能(Artificial Intelligence,简称:AI),是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支,企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器,该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。

[0022] 人工智能是对人的意识、思维的信息过程的模拟。人工智能不是人的智能,但能像人那样思考,也可能超过人的智能。人工智能是包括十分广泛的科学,由不同的领域组成,如机器学习,计算机视觉等等,总的说来,人工智能研究的一个主要目标是使机器能够胜任一些通常需要人类智能才能完成的复杂工作。

[0023] 以下结合附图描述根据本发明实施例的基于人工智能的智能交互设备控制方法、控制系统及智能交互设备。

[0024] 图1是根据本发明一个实施例的基于人工智能的智能交互设备控制方法的流程图。如图1所示,该方法包括如下步骤:

[0025] S101:接收多模态的输入信号,所述多模态的输入信号包括用户输入的图像信号、声音信号和/或距离信号。

[0026] 具体地,用户输入的声音信号可以是用户通过麦克风输入的;上述的图像信号可以通过摄像头采集得到的;上述的距离信号可以通过红外距离传感器采集得到。

[0027] S102:根据图像信号进行人脸检测,并在检测到有人脸时,获取人脸图像和人脸信息。其中,人脸信息包括但不限于人脸面积信息和人脸正面面对程度。

[0028] 具体地,对于如摄像头采集到的图像,可以通过人脸检测手段检测图像中是否存在人脸、人脸在图像中占据的面积、人脸是否正对智能交互设备等。

[0029] 当检测到图像中存在人脸之后,可以从图像中截取人脸图像,并保存人脸信息。

[0030] S103:根据人脸图像进行唇区检测以确定唇区运动情况。

[0031] 具体地,当步骤S102中检测到图像中存在人脸后,可以通过唇区检测手段从截取的人脸图像中进行唇区运动情况的检测。例如:检测结果为唇区发送动作或者唇区未发生动作。

[0032] 在本发明的一个实施例中,可以根据多帧人脸图像之间的唇区形状差异确定唇区运动情况。例如:前一帧人脸图像中唇区部分显示上下嘴唇闭合,而后一帧人脸图像中唇区部分显示上下嘴唇张开,此时,可以判定出用户唇区产生动作,可能是用户在开口说话等。

[0033] 需要说明的是,正常情况下即使用户没有说话等,上下嘴唇也可能在某个时刻产生动作,例如打哈欠。这种情况下,不应该认为用户的唇区产生了与说话等相关的动作,因此,为了避免误判的发生,可以通过连续的多帧图像间唇区部分的比较来确定上下嘴唇是否产生了动作,即用户是否存在说话等行为。此外,还可以通过对声音信号进行语音活动检测的方式判断用户是否存在说话等行,例如:判断声音信号中是否包含了用户说话时的语音(即:说话声),具体可以通过人工智能中语音识别的功能实现。当识别出声音信号中包含了说话人的语音(即:说话声),则可以判断出用户存在说话行为。这样,也可以避免上述中误判的发生。

[0034] S104:根据声音信号进行声源定位以得到声源信息。其中,声源信息包括但不限于



声源方位信息和声音强度信息。

[0035] 具体地,例如对于通过麦克风阵列接收到的多方位的声音信号,可以据此通过声源定位手段进行声源定位,从而确定出声源方位信息(即:声源角度信息)和声音强度信息。

[0036] 需要说明的是,通常声音信号中包括了多种声音,如:说话声和其它的噪声,因此,为了能够准确地对说话人的说话声进行声源定位,因此,在根据声音信号进行声源定位以得到声源信息之前,可以对声音信号进行去噪以便滤除其它的噪声干扰,提升对说话人的说话声进行声源定位的定位精度。具体而言:判断声音信号中是否包含用户说话时的语音;如果是,则保留声音信号中用户说话时的语音,并从声音信号中滤除其它的干扰噪音,在上述示例中,可以通过人工智能中语音识别的功能实现,即通过语音识别功能识别出声音信号中包含的说话人的说话声,进而滤除其它的噪声,由此,提升对说话人的说话声进行声源定位的定位精度。

[0037] S105:根据人脸信息、唇区运动情况、声源信息和/或距离信号判断用户的交互意愿以及交互意愿强烈程度。

[0038] 可以理解的是,在上述描述中,可以根据人脸信息、唇区运动情况、声源信息和距离信号中的任意一个来判断用户的交互意愿以及交互意愿强烈程度,也可以根据人脸信息、唇区运动情况、声源信息和距离信号中的多个或者全部一起来判断用户的交互意愿以及交互意愿强烈程度。相对于通过一个或者少数几个信息来判断用户的交互意愿以及交互意愿强烈程度而言,通过多个或者全部的上述信息来判断用户的交互意愿以及交互意愿强烈程度的准确性和可靠性相对较高。

[0039] 如下所述:

[0040] 1、当判断用户正对智能交互设备、用户的嘴唇未运动、用户发声且声音强度大于预定强度以及用户与智能交互设备之间的距离小于预设距离时,判断用户具有弱交互意愿。其中,预定强度可以根据经验确定,其目的是区别高强度的声音和相对低强度的声音,例如:预定强度可以以分贝的形式存在,预定强度例如为50分贝,当声音强度小于50分贝,则认为是低强度声源,反之则认为其为高强度声源,当然,在本发明的其它示例中,声音强度也可以用语音活动性指数来代替;预设距离也可以根据经验确定,例如:预设距离为1米。也就是说,如果判断出用户正脸面对本智能交互设备,距离近(如1米内),嘴唇未运动,无高强度声源,则判定为用户对智能交互设备感兴趣,存在弱交互意愿。

[0041] 2、当判断用户正对智能交互设备、用户的嘴唇产生运动、用户发声且声音强度小于预定强度以及用户与智能交互设备之间的距离小于预设距离时,判断用户具有疑似交互意愿。其中,预定强度可以根据经验确定,其目的是区别高强度的声音和相对低强度的声音,例如:预定强度可以以分贝的形式存在,预定强度例如为50分贝,当声音强度小于50分贝,则认为是低强度声源,反之则认为其为高强度声源;预设距离也可以根据经验确定,例如:预设距离为1米。也就是说,如果用户正脸面对智能交互设备,距离近(如1米内),嘴唇产生动作,不存在高强度声源,此时判定为疑似交互意愿。

[0042] 3、当判断用户正对智能交互设备、用户的嘴唇产生运动、用户发声且声音强度大于预定强度以及用户与智能交互设备之间的距离小于预设距离时,判断用户具有强交互意愿。其中,预定强度可以根据经验确定,其目的是区别高强度的声音和相对低强度的声音,例如:预定强度可以以分贝的形式存在,预定强度例如为50分贝,当声音强度小于50分贝,

则认为低强度声源,反之则认为其为高强度声源,当然,在本发明的其它示例中,声音强度也可以用语音活动性指数来代替;预设距离也可以根据经验确定,例如:预设距离为1米。也就是说,如果用户正脸面对智能交互设备,距离近(如1米内),嘴唇产生动作,存在高强度声源,则判定用户存在强交互意愿。

[0043] 4、当判断用户侧面面对智能交互设备、用户发声且声音强度大于预定强度以及用户与智能交互设备之间的距离小于预设距离时,判断用户具有伴随交互意愿。其中,预定强度可以根据经验确定,其目的是区别高强度的声音和相对低强度的声音,例如:预定强度可以以分贝的形式存在,预定强度例如为50分贝,当声音强度小于50分贝,则认为低强度声源,反之则认为其为高强度声源,当然,在本发明的其它示例中,声音强度也可以用语音活动性指数来代替;预设距离也可以根据经验确定,例如:预设距离为1米。也就是说,如果用户侧脸面对本设备,距离近(如1米内),有高强度声源,则判定用户存在伴随交互意愿。

[0044] 5、当未检测到人脸图像、用户发声且声音强度大于预定强度以及用户与智能交互设备之间的距离小于预设距离时,判断用户具有强疑似交互意愿。其中,预定强度可以根据经验确定,其目的是区别高强度的声音和相对低强度的声音,例如:预定强度可以以分贝的形式存在,预定强度例如为50分贝,当声音强度小于50分贝,则认为低强度声源,反之则认为其为高强度声源,当然,在本发明的其它示例中,声音强度也可以用语音活动性指数来代替;预设距离也可以根据经验确定,例如:预设距离为1米。也就是说,如果有高强度声源,摄像头检测不到人脸,距离近(如1米内):判断为用户存在强疑似交互意愿(即需确认强交互意愿)。

[0045] 6、当未检测到人脸图像、用户发声且声音强度大于预定强度以及用户与智能交互设备之间的距离大于预设距离时,判断用户具有弱疑似交互意愿。其中,预定强度可以根据经验确定,其目的是区别高强度的声音和相对低强度的声音,例如:预定强度可以以分贝的形式存在,预定强度例如为50分贝,当声音强度小于50分贝,则认为低强度声源,反之则认为其为高强度声源,当然,在本发明的其它示例中,声音强度也可以用语音活动性指数来代替;预设距离也可以根据经验确定,例如:预设距离为1米。也就是说,如果有高强度声源,检测不到人脸,距离远(如大于1米),则判定为弱疑似交互意愿(即弱疑似交互意愿)。

[0046] 7、以上为各种示例情况,综合来讲,是根据输入的多个独立特征构造针对多种交互意愿的多分类器,并依据多模态输入信号的值进行综合判断,来准确判定交互意愿并做出相应的反应。

[0047] S106:根据用户的交互意愿以及交互意愿强烈程度控制智能交互设备进行相应的交互响应。

[0048] 例如:当上述步骤中判断出存在弱交互意愿时,可以智能地控制智能交互设备进行静默响应,如:显示不同表情,简单的机械动作等,而无需发声。

[0049] 当上述步骤中判断出存在疑似交互意愿时,可以控制智能交互设备进行提高音量提示响应,如发出提高音量的提示。

[0050] 当上述步骤中判断存在强交互意愿时,可以控制智能交互设备进行正式交互响应,即:正式与用户进行交互。

[0051] 当上述步骤中判断存在伴随交互意愿时,可以控制智能交互设备进行语音/聊天交互响应,即:以语音/聊天交互方式为主。

[0052] 当上述步骤中判断存在强疑似交互意愿时,可以控制智能交互设备转向声源方向并进行提示响应,例如:将麦克风转向声源方向,并对用户进行提示。

[0053] 当上述步骤中判断存在弱疑似交互意愿时,可以仅仅控制智能交互设备转向声源方向。例如:仅将麦克风转向声源方向而不进行提示。

[0054] 另外,为了更加准确地判断出用户的交互意愿以及交互意愿强烈程度而避免误判的发生,在本发明的一个实施例中,可以在根据人脸信息、所述唇区运动情况、声源信息和/或距离信号判断用户的交互意愿以及交互意愿强烈程度之前,判断人脸信息、唇区运动情况、声源信息和/或距离信号是否满足预定条件;如果满足预定条件,则执行用户的交互意愿以及交互意愿强烈程度的判断。

[0055] 具体地说,可以通过一个计时器来实现上述条件的判断,例如:当检测到有正脸面对智能交互设备之后,启动计时器,并在正脸面对智能交互设备的时间超过一个特定的时间(如3秒)后,判定用户确实是面对智能交互设备。这样可以避免误判的发生,设想一下,用户如果仅是活动一下头部,则也可能在某个时刻正脸面对智能交互设备,而通过上述的计时判断,可以对用户活动头部而在某个时刻正脸面对智能交互设备进行忽略,因此,可以降低误判的概率甚至消除误判。

[0056] 此外,为了进一步提升用户的交互意愿以及交互意愿强烈程度的判断的准确性,在根据人脸信息、唇区运动情况、声源信息和/或距离信号判断用户的交互意愿以及交互意愿强烈程度之前,可以对人脸信息和唇部运动情况进行量化处理。如:30%正脸面对智能交互设备、50%正脸面对智能交互设备等。在进行量化之后,可以为用户的交互意愿以及交互意愿强烈程度的判断提供统一的标准,从而提升判断精度。

[0057] 在本发明的一个实施例中,该方法还包括:调整人脸信息、唇区运动情况、声源信息和/或距离信号的权重,其中,权重用于影响用户的交互意愿以及交互意愿强烈程度的判断结果;判断用户的交互意愿以及交互意愿强烈程度,进一步包括:根据人脸信息、唇区运动情况、声源信息和/或距离信号的权重判断用户的交互意愿以及交互意愿强烈程度。具体地,通过调整各个输入信号的敏感程度(即权重),如:调高正脸面对信号和唇部运动的权重,降低声源输入强度的权重,则在用户仅动嘴唇,并不实际发声的情况下,也判断为有交互意愿,这样,可以针对不同的场景,进行不同的交互行为的响应,提升智能交互设备的交互体验。

[0058] 需要说明的是,智能交互设备可以为普通的生活电器、信息类电器(如:计算机、电视机等)、视频会议系统或者智能机器人等。

[0059] 根据本发明实施例的基于人工智能的智能交互设备控制方法,能够实时的采集用户的语音信号、图像信号和/或距离信号,经过人工智能的分析后确定出用户是否存在交互意愿,并可以确定出交互意愿的强烈程度,然后自主地控制智能交互设备进行相应的动作,主动地与用户进行交互且交互手段丰富,进而提升用户的使用体验。

[0060] 图2是根据本发明一个实施例的基于人工智能的智能交互设备控制系统的结构框图。

[0061] 如图2所示,并结合图3根据本发明一个实施例的基于人工智能的智能交互设备控制系统200,包括:接收模块210(如摄像头、红外距离传感器、麦克风阵列)、人脸检测模块220、唇区检测模块230、声源定位模块240、决策模块250(即决策中心)和复合输出控制模块

260。

[0062] 其中,接收模块210用于接收多模态的输入信号,所述多模态的输入信号包括用户输入的图像信号、声音信号和/或距离信号。人脸检测模块220用于根据所述图像信号进行人脸检测,并在检测到有人脸时,获取所述人脸图像和人脸信息。唇区检测模块230用于根据所述人脸图像进行唇区检测以确定唇区运动情况。声源定位模块240用于根据所述声音信号进行声源定位以得到声源信息。决策模块250用于根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度。复合输出控制模块260用于根据所述用户的交互意愿以及交互意愿强烈程度控制所述智能交互设备进行相应的交互响应。

[0063] 在本发明的一个实施例中,还包括:语音活动检测模块(图2中没有示出),用于在声源定位模块240根据声音信号进行声源定位以得到声源信息之前,判断声音信号中是否包含用户说话时的语音,如果是,则保留声音信号中用户说话时的语音,并从声音信号中滤除其它的干扰噪音。

[0064] 具体而言,通常声音信号中包括了多种声音,如:说话声和其它的噪声,因此,为了能够准确地对说话人的说话声进行声源定位,因此,在根据声音信号进行声源定位以得到声源信息之前,可以对声音信号进行去噪以便滤除其它的噪声干扰,后续可以提升对说话人的说话声进行声源定位的定位精度。具体而言:判断声音信号中是否包含用户说话时的语音;如果是,则保留声音信号中用户说话时的语音,并从声音信号中滤除其它的干扰噪音,在上述示例中,可以通过人工智能中语音识别的功能实现,即通过语音识别功能识别出声音信号中包含的说话人的说话声,进而滤除其它的噪声,由此,后续便可以提升对说话人的说话声进行声源定位的定位精度。

[0065] 在本发明的一个实施例中,决策模块250还用于在根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度之前,判断所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号是否满足预定条件;如果满足所述预定条件,则执行用户的交互意愿以及交互意愿强烈程度的判断。

[0066] 在本发明的一个实施例中,决策模块250还用于在根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号判断所述用户的交互意愿以及交互意愿强烈程度之前,对所述人脸信息和所述唇区运动情况进行量化处理。

[0067] 在本发明的一个实施例中,决策模块250还用于:调整所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号的权重,其中,所述权重用于影响所述用户的交互意愿以及交互意愿强烈程度的判断结果;所述判断所述用户的交互意愿以及交互意愿强烈程度,包括:根据所述人脸信息、所述唇区运动情况、所述声源信息和/或所述距离信号的权重判断所述用户的交互意愿以及交互意愿强烈程度。

[0068] 在本发明的一个实施例中,人脸信息包括人脸面积信息和人脸正面面对程度,所述声源信息包括声源方位信息和声音强度信息。

[0069] 在本发明的一个实施例中,决策模块250用于:当判断所述用户正对所述智能交互设备、所述用户的嘴唇未运动、所述用户发声且声音强度大于预定强度以及用户与所述智能交互设备之间的距离小于预设距离时,判断所述用户具有弱交互意愿,复合输出控制模

块260用于:控制所述智能交互设备进行静默响应。

[0070] 在本发明的一个实施例中,决策模块250用于:当判断所述用户正对所述智能交互设备、所述用户的嘴唇产生运动、所述用户发声且声音强度小于预定强度以及用户与所述智能交互设备之间的距离小于所述预设距离时,判断所述用户具有疑似交互意愿,复合输出控制模块260用于:控制所述智能交互设备进行提高音量提示响应。

[0071] 在本发明的一个实施例中,决策模块250用于:当判断所述用户正对所述智能交互设备、所述用户的嘴唇产生运动、所述用户发声且声音强度大于所述预定强度以及所述用户与所述智能交互设备之间的距离小于所述预设距离时,判断用户具有强交互意愿,复合输出控制模块260用于:控制所述智能交互设备进行正式交互响应。

[0072] 在本发明的一个实施例中,决策模块250用于:当判断所述用户侧面面对所述智能交互设备、所述用户发声且声音强度大于所述预定强度以及所述用户与所述智能交互设备之间的距离小于所述预设距离时,判断所述用户具有伴随交互意愿,复合输出控制模块260用于:控制所述智能交互设备进行语音/聊天交互响应。

[0073] 在本发明的一个实施例中,决策模块250用于:当未检测到人脸图像、所述用户发声且声音强度大于所述预定强度以及所述用户与所述智能交互设备之间的距离小于所述预设距离时,判断所述用户具有强疑似交互意愿,复合输出控制模块260用于:控制所述智能交互设备转向所述声源方向并进行提示响应。

[0074] 在本发明的一个实施例中,决策模块250用于:当未检测到人脸图像、所述用户发声且声音强度大于所述预定强度以及所述用户与所述智能交互设备之间的距离大于所述预设距离时,判断所述用户具有弱疑似交互意愿,复合输出控制模块260用于:控制所述智能交互设备转向所述声源的响应。

[0075] 在本发明的一个实施例中,唇区检测模块230用于:根据多帧人脸图像之间的唇区形状差异确定所述唇区运动情况。

[0076] 根据本发明实施例的基于人工智能的智能交互设备控制系统,能够实时的采集用户的语音信号、图像信号和/或距离信号,经过人工智能的分析后确定出用户是否存在交互意愿,并可以确定出交互意愿的强烈程度,然后自主地控制智能交互设备进行相应的动作,主动地与用户进行交互且交互手段丰富,进而提升用户的使用体验。

[0077] 需要说明的是,本发明实施例的基于人工智能的智能交互设备控制系统的具体实现方式与本发明实施例的基于人工智能的智能交互设备控制方法的具体实现方式类似,具体请参见方法部分的描述,为了减少冗余,此处不做赘述。

[0078] 进一步地,本发明的实施例公开了一种智能交互设备,包括:根据上述任意一个实施例所述的基于人工智能的智能交互设备控制系统。该智能交互设备能够实时的采集用户的语音信号、图像信号和/或距离信号,经过人工智能的分析后确定出用户是否存在交互意愿,并可以确定出交互意愿的强烈程度,然后自主地控制智能交互设备进行相应的动作,主动地与用户进行交互且交互手段丰富,进而提升用户的使用体验。

[0079] 在本发明的描述中,需要理解的是,术语“中心”、“纵向”、“横向”、“长度”、“宽度”、“厚度”、“上”、“下”、“前”、“后”、“左”、“右”、“竖直”、“水平”、“顶”、“底”、“内”、“外”、“顺时针”、“逆时针”、“轴向”、“径向”、“周向”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必

须具有特定的方位、以特定的方位构造和操作，因此不能理解为对本发明的限制。

[0080] 此外，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本发明的描述中，“多个”的含义是至少两个，例如两个，三个等，除非另有明确具体的限定。

[0081] 在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不必针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0082] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本发明的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本发明的实施例所属技术领域的技术人员所理解。

[0083] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤，例如，可以被认为是在于实现逻辑功能的可执行指令的定序列列表，可以具体实现在任何计算机可读介质中，以供指令执行系统、装置或设备（如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统）使用，或结合这些指令执行系统、装置或设备而使用。就本说明书而言，“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例（非穷尽性列表）包括以下：具有一个或多个布线的电连接部（电子装置），便携式计算机盘盒（磁装置），随机存取存储器（RAM），只读存储器（ROM），可擦除可编程只读存储器（EPROM或闪速存储器），光纤装置，以及便携式光盘只读存储器（CDROM）。另外，计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质，因为可以例如通过对纸或其他介质进行光学扫描，接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序，然后将其存储在计算机存储器中。

[0084] 应当理解，本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列（PGA），现场可编程门阵列（FPGA）等。

[0085] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

[0086] 此外，在本发明各个实施例中的各功能单元可以集成在一个处理模块中，也可以

是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机可读取存储介质中。

[0087] 上述提到的存储介质可以是只读存储器,磁盘或光盘等。尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

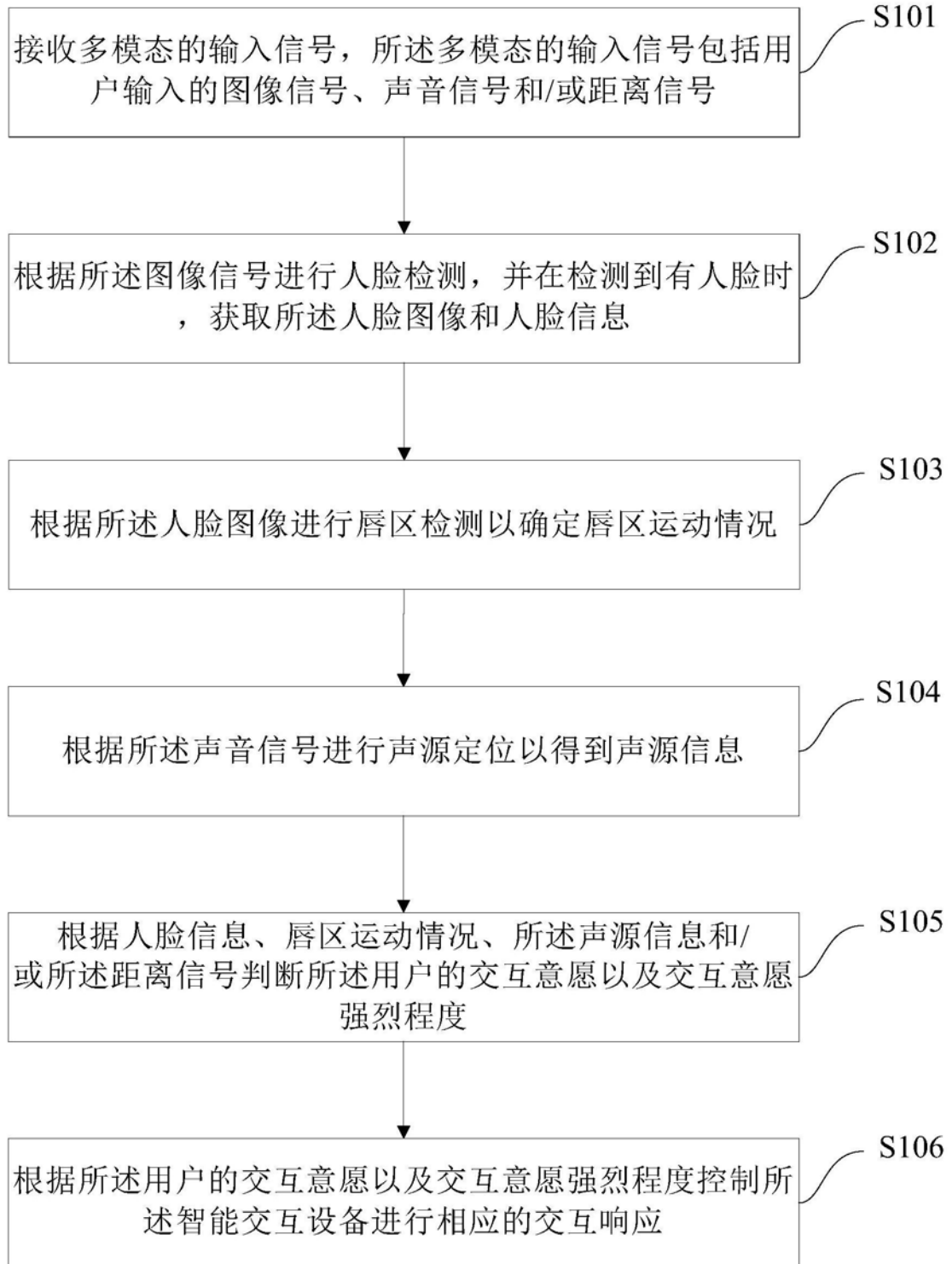


图1



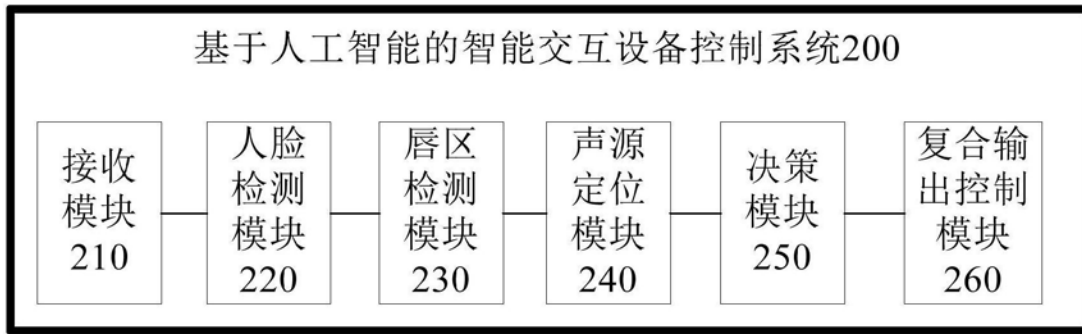


图2

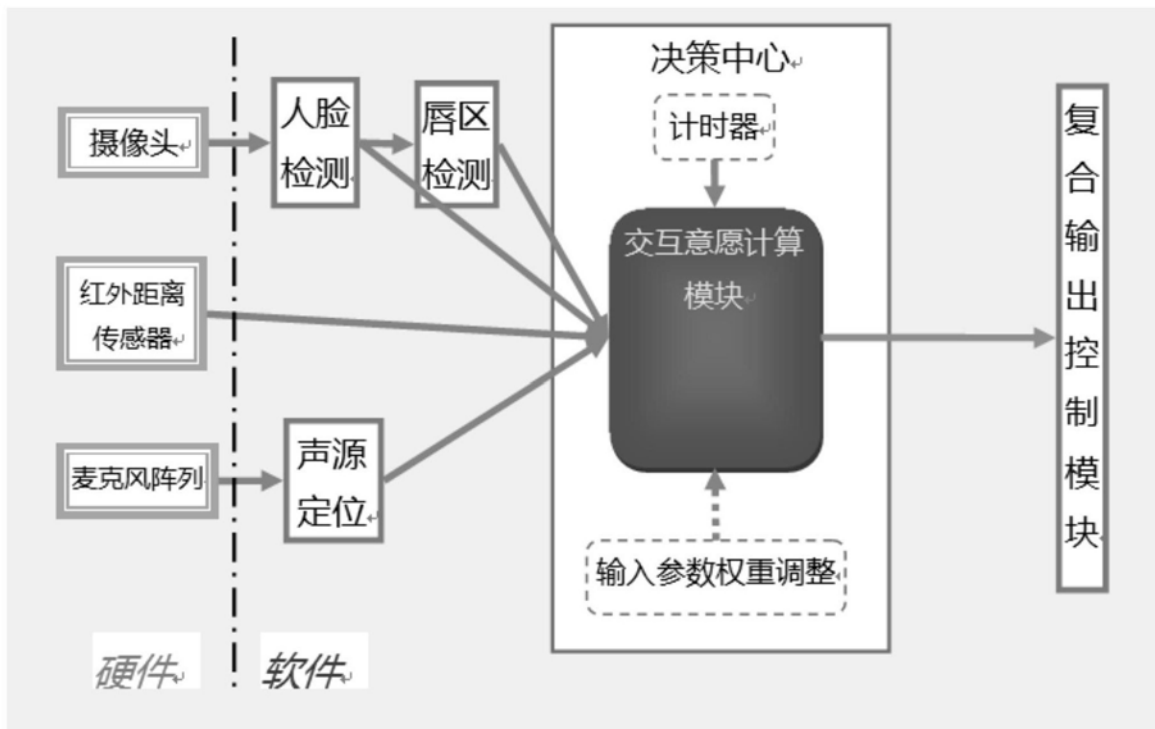


图3