



(12) 发明专利

(10) 授权公告号 CN 107972028 B

(45) 授权公告日 2020.10.23

(21) 申请号 201710632059.6

B25J 13/00 (2006.01)

(22) 申请日 2017.07.28

B25J 13/08 (2006.01)

(65) 同一申请的已公布的文献号
申请公布号 CN 107972028 A

审查员 潘玉芬

(43) 申请公布日 2018.05.01

(73) 专利权人 北京物灵智能科技有限公司
地址 100102 北京市朝阳区望京东园四区
13号楼101内A座36层3602室

(72) 发明人 张锦华

(74) 专利代理机构 北京风雅颂专利代理有限公司 11403

代理人 李莎 李弘

(51) Int. Cl.

B25J 9/16 (2006.01)

B25J 11/00 (2006.01)

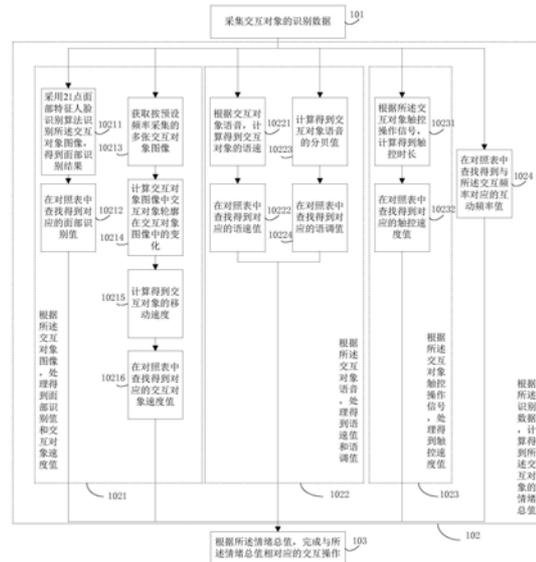
权利要求书3页 说明书20页 附图3页

(54) 发明名称

人机交互方法、装置及电子设备

(57) 摘要

本发明公开了一种人机交互方法,包括:采集交互对象的识别数据;所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率中的至少两种;根据所述识别数据,计算得到所述交互对象的情绪总值;根据所述情绪总值,完成与所述情绪总值相对应的交互操作;所述交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种。本发明还提供了一种人机交互装置和电子设备。



1. 一种人机交互方法,其特征在于,包括:

采集交互对象的识别数据;所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率;

根据所述识别数据,计算得到所述交互对象的情绪总值;

根据所述情绪总值,完成与所述情绪总值相对应的交互操作;所述交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种;

其中,根据所述识别数据,计算得到所述交互对象的情绪总值的步骤,包括:

根据所述交互对象图像,处理得到面部识别值和交互对象速度值;

根据所述交互对象语音,处理得到语速值和语调值;

根据所述交互对象触控操作信号,处理得到触控速度值;

根据所述交互频率,处理得到互动频率值;

所述情绪总值为所述面部识别值、交互对象速度值、语速值、语调值、触控速度值和互动频率值之和。

2. 根据权利要求1所述的方法,其特征在于,根据所述交互对象图像,处理得到面部识别值的步骤,包括:

采用21点面部特征人脸识别算法识别所述交互对象图像,得到面部识别结果;

在预先设定的面部识别结果与面部识别值对照表中,查找得到与所述面部识别结果对应的面部识别值。

3. 根据权利要求1所述的方法,其特征在于,根据所述交互对象图像,处理得到交互对象速度值的步骤,包括:

获取按预设频率采集的多张交互对象图像;

计算交互对象图像中交互对象轮廓在交互对象图像中的变化;

根据所述预设频率和变化,计算得到交互对象的移动速度;

在预先设定的移动速度与交互对象速度值对照表中,查找得到与所述移动速度对应的交互对象速度值。

4. 根据权利要求1所述的方法,其特征在于,根据所述交互对象语音,处理得到语速值的步骤,包括:

根据交互对象语音,计算得到交互对象的语速;

在预先设定的语速与语速值对照表中,查找得到与所述语速对应的语速值。

5. 根据权利要求1所述的方法,其特征在于,根据所述交互对象语音,处理得到语调值的步骤,包括:

计算得到交互对象语音的分贝值;

在预先设定的分贝值与语调值对照表中,查找得到与所述分贝值对应的语调值。

6. 根据权利要求1所述的方法,其特征在于,根据所述交互对象触控操作信号,处理得到触控速度值的步骤,包括:

根据所述交互对象触控操作信号,计算得到触控时长;

在预先设定的触控时长与触控速度值对照表中,查找得到与所述触控时长对应的触控速度值。

7. 根据权利要求1所述的方法,其特征在于,根据所述交互频率,处理得到互动频率值

的步骤,包括:

在预先设定的交互频率与互动频率值对照表中,查找得到与所述交互频率对应的互动频率值。

8. 一种人机交互装置,其特征在于,包括:

采集模块,用于采集交互对象的识别数据;所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率;

计算模块,用于根据所述识别数据,计算得到所述交互对象的情绪总值;

操作模块,用于根据所述情绪总值,完成与所述情绪总值相对应的交互操作;所述交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种;

所述计算模块,具体用于:

根据所述交互对象图像,处理得到面部识别值和交互对象速度值;

根据所述交互对象语音,处理得到语速值和语调值;

根据所述交互对象触控操作信号,处理得到触控速度值;

根据所述交互频率,处理得到互动频率值;

所述情绪总值为所述面部识别值、交互对象速度值、语速值、语调值、触控速度值和互动频率值之和。

9. 根据权利要求8所述的装置,其特征在于,所述计算模块,具体用于:

采用21点面部特征人脸识别算法识别所述交互对象图像,得到面部识别结果;

在预先设定的面部识别结果与面部识别值对照表中,查找得到与所述面部识别结果对应的面部识别值。

10. 根据权利要求8所述的装置,其特征在于,所述计算模块,具体用于:

获取按预设频率采集的多张交互对象图像;

计算交互对象图像中交互对象轮廓在交互对象图像中的变化;

根据所述预设频率和变化,计算得到交互对象的移动速度;

在预先设定的移动速度与交互对象速度值对照表中,查找得到与所述移动速度对应的交互对象速度值。

11. 根据权利要求8所述的装置,其特征在于,所述计算模块,具体用于:

根据交互对象语音,计算得到交互对象的语速;

在预先设定的语速与语速值对照表中,查找得到与所述语速对应的语速值。

12. 根据权利要求8所述的装置,其特征在于,所述计算模块,具体用于:

计算得到交互对象语音的分贝值;

在预先设定的分贝值与语调值对照表中,查找得到与所述分贝值对应的语调值。

13. 根据权利要求8所述的装置,其特征在于,所述计算模块,具体用于:

根据所述交互对象触控操作信号,计算得到触控时长;

在预先设定的触控时长与触控速度值对照表中,查找得到与所述触控时长对应的触控速度值。

14. 根据权利要求8所述的装置,其特征在于,所述计算模块,具体用于:

在预先设定的交互频率与互动频率值对照表中,查找得到与所述交互频率对应的互动频率值。

15. 一种电子设备,包括:
用于采集图像的摄像头、用于采集语音的声音传感器和用于接收触摸信号的触摸装置;
至少一个处理器;以及,
与所述至少一个处理器通信连接的存储器;其中,
所述存储器存储有可被所述一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1-7任一项所述的方法。

人机交互方法、装置及电子设备

技术领域

[0001] 本发明涉及数据处理技术领域,特别是指一种人机交互方法、装置及电子设备。

背景技术

[0002] 随着中国人口红利消失,机器人不仅在制造业上正在逐渐替代工人,未来还将在军事、服务、娱乐等领域代替人类完成许多工作。当今世界,服务机器人的市场化程度仍处于起步阶段,但在受简单劳动力不足及老龄化等刚性驱动和科技发展促进的影响下,其增长也较为迅速。服务机器人是除工业机器人之外的、用于非制造业并服务于人类的各种先进机器人,主要包括个人/家用服务机器人和专业服务机器人。其中,个人/家用机器人主要包括:家庭作业机器人、娱乐休闲机器人、残障辅助机器人、住宅安全和监视机器人等;专业服务机器人主要包括:场地机器人、专业清洁机器人、医用机器人、物流用途机器人、检查和维护保养机器人、建筑机器人、水下机器人,以及国防、营救和安全应用机器人等等。现在机器人行业大多数都以你问我答方式来为人们服务,即下发指令方式来做出相应的功能表现。比如:用户询问天气“今天天气怎么样?”,机器人则回答相应的天气情况,或者,用户通过语音命令机器人扫地等等。

[0003] 服务机器人行业目前基本上都是以目的性服务为主。没有通过机器人本体去观察人为的需求,在人们需要服务时不能及时主动的为人们服务,需要用户主动地下发相应的指令。这样一来,使得服务性能上显得有些滞后,即,机器人反应速度比较慢。

[0004] 目前日常见到的机器人大都是以人脸识别、语音技术其中的单一技术做为产品生产的基础技术。例如,常见的故事机(应用到了语音技术)、儿童教育机器人(应用到了人脸识别技术来绑定设备、语音技术做为正常交互)等。这些机器人大都是以用户主动下发指令并等待机器人做出响应为实现方式,也就不能及时为人们做出主动的、恰当的、合理的服务。

发明内容

[0005] 有鉴于此,本发明的目的在于提出一种人机交互方法、装置及电子设备,能够主动地提供及时的服务。

[0006] 基于上述目的本发明提供的人机交互方法,包括:

[0007] 采集交互对象的识别数据;所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率中的至少两种;

[0008] 根据所述识别数据,计算得到所述交互对象的情绪总值;

[0009] 根据所述情绪总值,完成与所述情绪总值相对应的交互操作;所述交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种。

[0010] 可选的,所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率;

[0011] 根据所述识别数据,计算得到所述交互对象的情绪总值的步骤,包括:

- [0012] 根据所述交互对象图像,处理得到面部识别值和交互对象速度值;
- [0013] 根据所述交互对象语音,处理得到语速值和语调值;
- [0014] 根据所述交互对象触控操作信号,处理得到触控速度值;
- [0015] 根据所述交互频率,处理得到互动频率值;
- [0016] 所述情绪总值为所述面部识别值、交互对象速度值、语速值、语调值、触控速度值和互动频率值中至少两项之和。
- [0017] 可选的,据所述交互对象图像,处理得到面部识别值的步骤,包括:
- [0018] 采用21点面部特征人脸识别算法识别所述交互对象图像,得到面部识别结果;
- [0019] 在预先设定的面部识别结果与面部识别值对照表中,查找得到与所述面部识别结果对应的面部识别值。
- [0020] 可选的,根据所述交互对象图像,处理得到交互对象速度值的步骤,包括:
- [0021] 获取按预设频率采集的多张交互对象图像;
- [0022] 计算交互对象图像中交互对象轮廓在交互对象图像中的变化;
- [0023] 根据所述预设频率和变化,计算得到交互对象的移动速度;
- [0024] 在预先设定的移动速度与交互对象速度值对照表中,查找得到与所述移动速度对应的交互对象速度值。
- [0025] 可选的,根据所述交互对象语音,处理得到语速值的步骤,包括:
- [0026] 根据交互对象语音,计算得到交互对象的语速;
- [0027] 在预先设定的语速与语速值对照表中,查找得到与所述语速对应的语速值。
- [0028] 可选的,根据所述交互对象语音,处理得到语调值的步骤,包括:
- [0029] 计算得到交互对象语音的分贝值;
- [0030] 在预先设定的分贝值与语调值对照表中,查找得到与所述分贝值对应的语调值。
- [0031] 可选的,根据所述交互对象触控操作信号,处理得到触控速度值的步骤,包括:
- [0032] 根据所述交互对象触控操作信号,计算得到触控时长;
- [0033] 在预先设定的触控时长与触控速度值对照表中,查找得到与所述触控时长对应的触控速度值。
- [0034] 可选的,根据所述交互频率,处理得到互动频率值的步骤,包括:
- [0035] 在预先设定的交互频率与互动频率值对照表中,查找得到与所述交互频率对应的互动频率值。
- [0036] 可选的,所述情绪总值为所述面部识别值、交互对象速度值、语速值、语调值、触控速度值和互动频率值之和。
- [0037] 本发明实施例的第二个方面,提供了一种人机交互装置,包括:
- [0038] 采集模块,用于采集交互对象的识别数据;所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率中的至少两种;
- [0039] 计算模块,用于根据所述识别数据,计算得到所述交互对象的情绪总值;
- [0040] 操作模块,用于根据所述情绪总值,完成与所述情绪总值相对应的交互操作;所述交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种。
- [0041] 可选的,所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率;

- [0042] 所述计算模块,具体用于:
- [0043] 根据所述交互对象图像,处理得到面部识别值和交互对象速度值;
- [0044] 根据所述交互对象语音,处理得到语速值和语调值;
- [0045] 根据所述交互对象触控操作信号,处理得到触控速度值;
- [0046] 根据所述交互频率,处理得到互动频率值;
- [0047] 所述情绪总值为所述面部识别值、交互对象速度值、语速值、语调值、触控速度值和互动频率值中至少两项之和。
- [0048] 可选的,所述计算模块,具体用于:
- [0049] 采用21点面部特征人脸识别算法识别所述交互对象图像,得到面部识别结果;
- [0050] 在预先设定的面部识别结果与面部识别值对照表中,查找得到与所述面部识别结果对应的面部识别值。
- [0051] 可选的,所述计算模块,具体用于:
- [0052] 获取按预设频率采集的多张交互对象图像;
- [0053] 计算交互对象图像中交互对象轮廓在交互对象图像中的变化;
- [0054] 根据所述预设频率和变化,计算得到交互对象的移动速度;
- [0055] 在预先设定的移动速度与交互对象速度值对照表中,查找得到与所述移动速度对应的交互对象速度值。
- [0056] 可选的,所述计算模块,具体用于:
- [0057] 根据交互对象语音,计算得到交互对象的语速;
- [0058] 在预先设定的语速与语速值对照表中,查找得到与所述语速对应的语速值。
- [0059] 可选的,所述计算模块,具体用于:
- [0060] 计算得到交互对象语音的分贝值;
- [0061] 在预先设定的分贝值与语调值对照表中,查找得到与所述分贝值对应的语调值。
- [0062] 可选的,所述计算模块,具体用于:
- [0063] 根据所述交互对象触控操作信号,计算得到触控时长;
- [0064] 在预先设定的触控时长与触控速度值对照表中,查找得到与所述触控时长对应的触控速度值。
- [0065] 可选的,所述计算模块,具体用于:
- [0066] 在预先设定的交互频率与互动频率值对照表中,查找得到与所述交互频率对应的互动频率值。
- [0067] 可选的,所述情绪总值为所述面部识别值、交互对象速度值、语速值、语调值、触控速度值和互动频率值之和。
- [0068] 本发明实施例的第三个方面,提供了一种电子设备,包括:
- [0069] 用于采集图像的摄像头、用于采集语音的声音传感器和/或用于接收触摸信号的触摸装置;
- [0070] 至少一个处理器;以及,
- [0071] 与所述至少一个处理器通信连接的存储器;其中,
- [0072] 所述存储器存储有可被所述一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如上任一项所述的人机交互方法。

[0073] 从上面所述可以看出,本发明实施例提供的人机交互方法、装置及电子设备,通过主动采集更为丰富的交互对象的识别数据,并基于所述识别数据得到更接近交互对象的当前状态的情绪总值,再根据情绪总值完成相应的较为丰富的交互操作,从而主动为交互对象提供更合理的服务。

附图说明

[0074] 图1为本发明提供的人机交互方法的一个实施例的流程示意图;

[0075] 图2为本发明提供的人机交互方法的一个实施例中21点面部特征人脸识别算法在人脸上的关键点分布示意图;

[0076] 图3为本发明提供的人机交互装置的一个实施例的结构示意图;

[0077] 图4为本发明提供的电子设备的一个实施例的结构示意图。

具体实施方式

[0078] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明进一步详细说明。

[0079] 需要说明的是,本发明实施例中所有使用“第一”和“第二”的表述均是为了区分两个相同名称非相同的实体或者非相同的参量,可见“第一”“第二”仅为了表述的方便,不应理解为对本发明实施例的限定,后续实施例对此不再一一说明。

[0080] 基于前述目的,本发明实施例的第一个方面,提供了一种能够主动地提供及时的服务的人机交互方法的一个实施例。如图1所示,为本发明提供的人机交互方法的一个实施例的流程示意图。

[0081] 所述人机交互方法,包括:

[0082] 步骤101:采集交互对象的识别数据;所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率中的至少两种;

[0083] 所述交互对象,通常是指所述人机交互方法中的用户一方;所述交互对象图像可以是指通过摄像头采集得到的包含交互对象的图像数据。用于采集所述交互对象图像的摄像头,可以是设置在用于完成所述人机交互方法的专用设备上(例如集成在可以完成人机交互的智能机器人上),也可以是设置在具有较好的拍摄角度的室内的某个位置,例如能够拍摄到大门处的较为完整的画面的玄关的某个位置,或者是正对着大门的客厅的某个位置等等。所述交互对象语音,可以是指通过声音传感器、拾音器或麦克风等能够采集声音的装置采集得到的包含交互对象发出的声音的语音数据。用于采集所述交互对象语音的装置,可以是设置在用于完成所述人机交互方法的专用设备上(例如集成在可以完成人机交互的智能机器人上),也可以是设置在具有较好的声音采集效果的室内的某个位置,例如能够采集到用户进门以后所发出的声音的位置(例如玄关的某个位置),或者是用户经常活动的区域,例如客厅的某个位置,等等。所述交互对象触控操作信号,可以是通过能够接收触控信号的触控装置采集得到的,所述触控信号则可以是交互对象触摸该触控装置而发出的触控信号;用于采集所述交互对象触控操作信号的触控装置,可以是设置在室内如玄关、客厅、卧室等位置的触控板等触控装置,也可以是设置在用于完成所述人机交互方法的专用设备上的(例如集成在可以完成人机交互的智能机器人上)。当然,可以想到,用于采集所述交互

对象图像、交互对象语音和交互对象触控操作信号的设备,并不局限于仅在某个地方设置一个,为了达到更全面的效果,可以在需要的位置进行设置,并且用户可以根据自己的实际需求进行增删,等等。所述交互频率则是根据用户在预定时间范围内进行人机交互的次数而定的,具体的参数可以根据用户需要进行设置,也可以采用系统默认设定值。

[0084] 步骤102:根据所述识别数据,计算得到所述交互对象的情绪总值;所述情绪总值基于当前采集的所述识别数据而得出,可以用于表征所述交互对象的情绪状态。

[0085] 步骤103:根据所述情绪总值,完成与所述情绪总值相对应的交互操作;在得到情绪总值后,能够根据预先设定的情绪总值(也可以是情绪总值的范围区间)与所述交互操作的对应关系,选择相应的交互操作并执行;根据情绪总值的不同,所述交互操作也可以不同,或者,在情绪总值对应的交互操作为多种时,不同情绪总值(或者情绪总值的范围区间)对应的交互操作可以有部分重叠,所述交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种。

[0086] 所述播放多媒体文件,可以是播放预先存储于某设备中的语音或视频,并且不同语音或视频与相应的情绪总值相关联,在不同情绪总值时,可根据设定的方式播放所述语音或视频;除此之外,所述语音或视频还可以是通过网络连接到的互联网中的语音或视频,这种方式可以通过提取用户的所述交互对象语音中的关键词来实现;可以想到,除了语音和视频外,还可以包括歌曲、电视剧、电影等各种多媒体文件,在此不再赘述。

[0087] 所述互动问答,可以是基于预先存储于某设备中的对话数据,根据交互对象发出的语音中的关键字而进行的交互对象与设备之间的一问一答;此外,当交互对象的语音中的关键字不能在本地找到匹配数据时,还可以通过网络从互联网中查找。

[0088] 所述控制家庭环境中的任意设备运行,可以是指,控制处于用户的家庭室内环境中的任何能够建立连接的电子设备或电器设备等,实现开启、关闭、调节等操作,例如开启音响并播放音乐,调节灯的颜色、亮度等,打开热水器并对浴缸进行注水,打开电视并播放用户喜欢的节目,等等。所述家庭环境中的任意设备,可以是指家庭环境中安装的具有常用功能的电子设备或电器设备,例如电视、冰箱、音响设备等等,但也可以是将一些能够集成在一起的功能集成于一体的专用设备(例如可以完成人机交互的智能机器人)。

[0089] 上述的交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种,在特殊情况下,可以同时包括这三种,从而使用户获得更丰富的反馈,有利于调节用户的情绪状态。

[0090] 从上述实施例可以看出,本发明实施例提供的人机交互方法,通过主动采集更为丰富的交互对象的识别数据,并基于所述识别数据得到更接近交互对象的当前状态的情绪总值,再根据情绪总值完成相应的较为丰富的交互操作,从而主动为交互对象提供更合理的服务。

[0091] 在一些可选实施方式中,所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率;

[0092] 继续参照附图1,根据所述识别数据,计算得到所述交互对象的情绪总值的步骤102,可具体包括以下步骤:

[0093] 步骤1021:根据所述交互对象图像,处理得到面部识别值a和交互对象速度值b。

[0094] 可选的,所述面部识别值a可通过人脸识别技术来处理得到。

[0095] 所述人脸识别技术可包含以下三个部分：

[0096] (1) 人脸检测

[0097] 面貌检测是指在动态的场景与复杂的背景中判断是否存在面像，并分离出这种面像。一般有下列几种方法：

[0098] ①参考模板法

[0099] 首先设计一个或数个标准人脸的模板，然后计算测试采集的样品与标准模板之间的匹配程度，并通过阈值来判断是否存在人脸；

[0100] ②人脸规则法

[0101] 由于人脸具有一定的结构分布特征，所谓人脸规则的方法即提取这些特征生成相应的规则以判断测试样品是否包含人脸；

[0102] ③样品学习法

[0103] 这种方法即采用模式识别中人工神经网络的方法，即通过对面像样品集和非面像样品集的学习产生分类器；

[0104] ④肤色模型法

[0105] 这种方法是依据面貌肤色在色彩空间中分布相对集中的规律来进行检测；

[0106] ⑤特征子脸法

[0107] 这种方法是将所有面像集合视为一个面像子空间，并基于检测样品与其在子空间的投影之间的距离判断是否存在面像。

[0108] 值得提出的是，上述5种方法在实际的人脸识别过程中也可综合采用。

[0109] (2) 人脸跟踪

[0110] 面貌跟踪是指对被检测到的面貌进行动态目标跟踪。具体采用基于模型的方法或基于运动与模型相结合的方法。此外，利用肤色模型跟踪也不失为一种简单而有效的手段。

[0111] (3) 人脸比对

[0112] 面貌比对是对被检测到的面貌像进行身份确认或在面像库中进行目标搜索。这实际上就是说，将采样到的面像与库存的面像依次进行比对，并找出最佳的匹配对象。所以，面像的描述决定了面像识别的具体方法与性能。主要采用特征向量与面纹模板两种描述方法：

[0113] ①特征向量法

[0114] 该方法是先确定眼虹膜、鼻翼、嘴角等面像五官轮廓的大小、位置、距离等属性，然后再计算出它们的几何特征量，而这些特征量形成一描述该面像的特征向量。

[0115] ②面纹模板法

[0116] 该方法是在库中存贮若干标准面像模板或面像器官模板，在进行比对时，将采样面像所有象素与库中所有模板采用归一化相关量度量进行匹配。此外，还有采用模式识别的自相关网络或特征与模板相结合的方法。

[0117] 人脸识别技术的核心实际为“局部人体特征分析”和“图形/神经识别算法。”这种算法是利用人体面部各器官及特征部位的方法。如对应几何关系多数数据形成识别参数与数据库中所有的原始参数进行比较、判断与确认。一般要求判断时间低于1秒。

[0118] 在一些可选实施方式中，根据所述交互对象图像，处理得到面部识别值a的步骤，具体可包括以下步骤：

[0119] 步骤10211:采用21点面部特征人脸识别算法识别所述交互对象图像,得到面部识别结果;

[0120] 所述21点面部特征人脸识别算法具体的实现方式包括:

[0121] 在人脸上的关键位置进行标记,如图2和表1所示。

[0122] 表1 21点面部特征人脸识别算法中21点在人脸上的分布

[0123]	左眉	A、B、C	右眉	D、E、F
	左眼	G、Q、H	右眼	I、R、J
	鼻子	K、L、M、S		
	嘴唇	N、O、P	嘴角	T、U

[0124] 根据21点中各个点的变动偏离值给出一个数值,赋予成员变量中;然后根据预设算法得到相应的情绪识别结果,例如悲痛、高兴等等。

[0125] 可选的,21点面部特征人脸识别算法可直接调用商汤提供的接口,接口方法中需要传递相应的参数,参数1为bitmap类型的图片,参数2为该图片的方向;根据该参数,商汤根据21点识别技术,返回一个情绪的识别结果(例如悲痛、高兴等等,参见表2),具体函数为CvAttributeResult attribute(Bitmap image,int rotation),商汤会对这个图片进行处理,返回一个自定义的类CvAttributeResult,该类包含两个数组,一个为人脸特征(戴太阳镜、性别、种族等),另一个数组为人脸情绪(悲痛、高兴等),具体细节请参见商汤提供的FaceSDK_WhitePaper v6.2.pdf技术文档。

[0126] 对于能够返回的所有人脸情绪(面部识别结果),赋予其相应面部识别值,从而得到预先设定的面部识别结果与面部识别值对照表。

[0127] 步骤10212:在预先设定的面部识别结果与面部识别值对照表(可选的,所述面部识别结果与面部识别值对照表如表2所示)中,查找得到与所述面部识别结果对应的面部识别值a;在所述面部识别结果与面部识别值对照表中,判定交互对象为平静状态时,面部识别值a赋值为0~1。

[0128] 表2面部识别结果与面部识别值对照表

[0129]	识别结果	面部识别值
	悲痛	-8
	恐惧	-8
	哭泣	-7
	愤慨	-7
	悲伤	-6
	沮丧	-5
	忐忑	-5
	郁郁寡欢	-5
	悲观	-4
	尴尬	-3
	颓然	-3
	忧愁	-2
	忧郁	-2

郁闷	-2
纠结	-1
无奈	-1
正常	0
期待	1
微笑	1
感兴趣的	2
欣喜	2
激动	3
喜出望外	3
兴奋	4
高兴	4
愉悦	5
快乐	5
欢快	6
大笑	7
开怀大笑	8

[0130] 在一些可选实施方式中,根据所述交互对象图像,处理得到交互对象速度值的步骤,还可具体包括以下步骤:

[0131] 步骤10213:获取按预设频率采集的多张交互对象图像;所述预设频率可以根据不同的需求进行设定,可选的,具体的设定值以能够通过连续采集得到的图像判断交互对象的移动速度和移动方式为准;

[0132] 步骤10214:计算交互对象图像中交互对象轮廓在交互对象图像中的变化;所述变化可以是交互对象在交互对象图像中的位置变化、轮廓大小的变化等等;

[0133] 步骤10215:根据所述预设频率和变化,计算得到交互对象的移动速度;根据两张相邻图像中交互对象的轮廓的变化,可以计算得到交互对象相对于采集装置的移动距离,而基于所述预设频率,则可计算得到交互对象移动该移动距离所耗费的时间,基于计算得到的移动距离和时间,就可以计算得到交互对象的移动速度;较佳的,为了保证计算结果的准确性,可以将多张连续的图片分别计算移动速度后取平均值;

[0134] 步骤10216:在预先设定的移动速度与交互对象速度值对照表(可选的,所述移动速度与交互对象速度值对照表如表3所示)中,查找得到与所述移动速度对应的交互对象速度值b。

[0135] 表3移动速度与交互对象速度值对照表

移动速度	慢 (小于 1m/s ²)	缓 慢 (1~2m/s ²)	正 常 (2~4m/s ²)	较 快 (4~5m/s ²)	快 (大于 5m/s ²)
速度值	-2	-1	0	1	2

[0137] 除了采用上述方式计算移动速度外,还可以采用其他方式来计算所述移动速度。根据运动学原理,人体重心位于第二骶骨前缘,两髋关节中央。直线运动时该中心是身体摆

动幅度最小的部位。行走时人体重心不仅在水平方向,而且在垂直方向上不断改变着位置和速度。其中身体重心在垂直方向的速度变化与各关节及其活动肌肉的力学状况有密切关系。以人体重心加速度为衡量依据。因此,可以通过计算重心加速度来判断交互对象的移动速度,例如,利用交互对象随身携带的手机中的陀螺仪采集交互对象的重心加速度变化,并通过网络接收所述计算结果;或者,若所述人机交互方法对应有专用于该方法的设备(例如智能机器人)时,可通过交互对象携带所述设备运动时,经该设备中的陀螺仪等装置来计算重心加速度的变化。

[0138] 步骤1022:根据所述交互对象语音,处理得到语速值c和语调值d;

[0139] 可选的,所述语速值c可通过语音识别技术来处理得到。

[0140] 一般来说,语音识别的方法有三种:基于声道模型和语音知识的方法、模板匹配的方法以及利用人工神经网络的方法。

[0141] (1) 基于语音学和声学的方法

[0142] 通常认为常用语言中有有限个不同的语音基元,而且可以通过其语音信号的频域或时域特性来区分。这样该方法分为两步实现:

[0143] 第一步,分段和标号

[0144] 把语音信号按时间分成离散的段,每段对应一个或几个语音基元的声学特性。然后根据相应声学特性对每个分段给出相近的语音标号。

[0145] 第二步,得到词序列

[0146] 根据第一步所得语音标号序列得到一个语音基元网格,从词典得到有效的词序列,也可结合句子的文法和语义同时进行。

[0147] (2) 模板匹配的方法

[0148] 模板匹配的方法发展比较成熟,目前已达到了实用阶段。在模板匹配方法中,要经过四个步骤:特征提取、模板训练、模板分类、判决。常用的技术有三种:动态时间规整(DTW)、隐马尔可夫(HMM)理论、矢量量化(VQ)技术。

[0149] 动态时间规整(DTW)

[0150] 语音信号的端点检测是进行语音识别中的一个基本步骤,它是特征训练和识别的基础。所谓端点检测就是在语音信号中的各种段落(如音素、音节、词素)的始点和终点的位置,从语音信号中排除无声段。在早期,进行端点检测的主要依据是能量、振幅和过零率。但效果往往不明显。60年代日本学者Itakura提出了动态时间规整算法(DTW: DynamicTimeWarping)。算法的思想就是把未知量均匀的升长或缩短,直到与参考模式的长度一致。在这一过程中,未知单词的时间轴要不均匀地扭曲或弯折,以使其特征与模型特征对正。

[0151] 隐马尔可夫法(HMM)

[0152] 隐马尔可夫法(HMM)是70年代引入语音识别理论的,它的出现使得自然语音识别系统取得了实质性的突破。HMM方法现已成为语音识别的主流技术,目前大多数大词汇量、连续语音的非特定人语音识别系统都是基于HMM模型的。HMM是对语音信号的时间序列结构建立统计模型,将之看作一个数学上的双重随机过程:一个是用具有有限状态数的Markov链来模拟语音信号统计特性变化的隐含的随机过程,另一个是与Markov链的每一个状态相关联的观测序列的随机过程。前者通过后者表现出来,但前者的具体参数是不可测的。人的

言语过程实际上就是一个双重随机过程,语音信号本身是一个可观测的时变序列,是由大脑根据语法知识和言语需要(不可观测的状态)发出的音素的参数流。可见HMM合理地模仿了这一过程,很好地描述了语音信号的整体非平稳性和局部平稳性,是较为理想的一种语音模型。

[0153] 矢量量化(VQ)

[0154] 矢量量化(VectorQuantization)是一种重要的信号压缩方法。与HMM相比,矢量量化主要适用于小词汇量、孤立词的语音识别中。其过程是:将语音信号波形的k个样点的每一帧,或有k个参数的每一参数帧,构成k维空间中的一个矢量,然后对矢量进行量化。量化时,将k维无限空间划分为M个区域边界,然后将输入矢量与这些边界进行比较,并被量化为“距离”最小的区域边界的中心矢量值。矢量量化器的设计就是从大量信号样本中训练出好的码书,从实际效果出发寻找到好的失真测度定义公式,设计出最佳的矢量量化系统,用最少的搜索和计算失真的运算量,实现最大可能的平均信噪比。

[0155] 核心思想可以这样理解:如果一个码书是为某一特定的信源而优化设计的,那么由这一信息源产生的信号与该码书的平均量化失真就应小于其他信息的信号与该码书的平均量化失真,也就是说编码器本身存在区分能力。

[0156] 在实际的应用过程中,人们还研究了多种降低复杂度的方法,这些方法大致可以分为两类:无记忆的矢量量化和有记忆的矢量量化。无记忆的矢量量化包括树形搜索的矢量量化和多级矢量量化。

[0157] (3) 神经网络的方法

[0158] 利用人工神经网络的方法是80年代末期提出的一种新的语音识别方法。人工神经网络(ANN)本质上是一个自适应非线性动力学系统,模拟了人类神经活动的原理,具有自适应性、并行性、鲁棒性、容错性和学习特性,其强的分类能力和输入-输出映射能力在语音识别中都很有吸引力。但由于存在训练、识别时间太长的缺点,目前仍处于实验探索阶段。

[0159] 由于ANN不能很好的描述语音信号的时间动态特性,所以常把ANN与传统识别方法结合,分别利用各自优点来进行语音识别。

[0160] 一个完整的基于统计的语音识别系统可大致分为三部分:

[0161] (1) 语音信号预处理与特征提取;

[0162] 选择识别单元是语音识别研究的第一步。语音识别单元有单词(句)、音节和音素三种,具体选择哪一种,由具体的研究任务决定。

[0163] 单词(句)单元广泛应用于中小词汇语音识别系统,但不适合大词汇系统,原因在于模型库太庞大,训练模型任务繁重,模型匹配算法复杂,难以满足实时性要求。

[0164] 音节单元多见于汉语语音识别,主要因为汉语是单音节结构的语言,而英语是多音节,并且汉语虽然有大约1300个音节,若不考虑声调,约有408个无调音节,数量相对较少。因此,对于中、大词汇量汉语语音识别系统来说,以音节为识别单元基本是可行的。

[0165] 音素单元以前多见于英语语音识别的研究中,但目前中、大词汇量汉语语音识别系统也在越来越多地采用。原因在于汉语音节仅由声母(包括零声母有22个)和韵母(共有28个)构成,且声韵母声学特性相差很大。实际应用中常把声母依后续韵母的不同而构成细化声母,这样虽然增加了模型数目,但提高了易混淆音节的区分能力。由于协同发音的影响,音素单元不稳定,所以如何获得稳定的音素单元,还有待研究。

[0166] 语音识别一个根本的问题是合理的选用特征。特征参数提取的目的是对语音信号进行分析处理,去掉与语音识别无关的冗余信息,获得影响语音识别的重要信息,同时对语音信号进行压缩。在实际应用中,语音信号的压缩率介于10-100之间。语音信号包含了大量各种不同的信息,提取哪些信息,用哪种方式提取,需要综合考虑各方面的因素,如成本,性能,响应时间,计算量等。非特定人语音识别系统一般侧重提取反映语义的特征参数,尽量去除说话人的个人信息;而特定人语音识别系统则希望在提取反映语义的特征参数的同时,尽量也包含说话人的个人信息。

[0167] 线性预测(LP)分析技术是目前应用广泛的特征参数提取技术,许多成功的应用系统都采用基于LP技术提取的倒谱参数。但线性预测模型是纯数学模型,没有考虑人类听觉系统对语音的处理特点。

[0168] Mel参数和基于感知线性预测(PLP)分析提取的感知线性预测倒谱,在一定程度上模拟了人耳对语音的处理特点,应用了人耳听觉感知方面的一些研究成果。实验证明,采用这种技术,语音识别系统的性能有一定提高。从目前使用的情况来看,梅尔刻度式倒频谱参数已逐渐取代原本常用的线性预测编码导出的倒频谱参数,原因是它考虑了人类发声与接收声音的特性,具有更好的鲁棒性(Robustness)。

[0169] 也有研究者尝试把小波分析技术应用于特征提取,但目前性能难以与上述技术相比,有待进一步研究。

[0170] (2) 声学模型与模式匹配;

[0171] 声学模型通常是将获取的语音特征使用训练算法进行训练后产生。在识别时将输入的语音特征同声学模型(模式)进行匹配与比较,得到最佳的识别结果。

[0172] 声学模型是识别系统的底层模型,并且是语音识别系统中最关键的一部分。声学模型的目的是提供一种有效的方法计算语音的特征矢量序列和每个发音模板之间的距离。声学模型的设计和语言发音特点密切相关。声学模型单元大小(字发音模型、半音节模型或音素模型)对语音训练数据量大小、系统识别率,以及灵活性有较大的影响。必须根据不同语言的特点、识别系统词汇量的大小决定识别单元的大小。

[0173] 以汉语为例:

[0174] 汉语按音素的发音特征分类分为辅音、单元音、复元音、复鼻尾音四种,按音节结构分类为声母和韵母。并且由音素构成声母或韵母。有时,将含有声调的韵母称为调母。由单个调母或由声母与调母拼音成为音节。汉语的一个音节就是汉语一个字的音,即音节字。由音节字构成词,最后再由词构成句子。

[0175] 汉语声母共有22个,其中包括零声母,韵母共有38个。按音素分类,汉语辅音共有22个,单元音13个,复元音13个,复鼻尾音16个。

[0176] 目前常用的声学模型基元为声韵母、音节或词,根据实现目的不同来选取不同的基元。汉语加上语气词共有412个音节,包括轻音字,共有1282个有调音节字,所以当在小词汇表孤立词语音识别时常选用词作为基元,在大词汇表语音识别时常采用音节或声韵母建模,而在连续语音识别时,由于协同发音的影响,常采用声韵母建模。

[0177] 基于统计的语音识别模型常用的就是HMM模型 $\lambda(N, M, \pi, A, B)$,涉及到HMM模型的相关理论包括模型的结构选取、模型的初始化、模型参数的重估以及相应的识别算法等。

[0178] (3) 语言模型与语言处理;

[0179] 语言模型包括由识别语音命令构成的语法网络或由统计方法构成的语言模型,语言处理可以进行语法、语义分析。

[0180] 语言模型对中、大词汇量的语音识别系统特别重要。当分类发生错误时可以根据语言学模型、语法结构、语义学进行判断纠正,特别是一些同音字则必须通过上下文结构才能确定词义。语言学理论包括语义结构、语法规则、语言的数学描述模型等有关方面。目前比较成功的语言模型通常是采用统计语法的语言模型与基于规则语法结构命令语言模型。语法结构可以限定不同词之间的相互连接关系,减少了识别系统的搜索空间,这有利于提高系统的识别。

[0181] 通过上述语音识别方法,能够得到采集得到的语音的文字表达,结合采集得到的语音的长度,即可得出交互对象的语速。

[0182] 在一些可选实施方式中,根据所述交互对象语音,处理得到语速值c的步骤,还可具体包括以下步骤:

[0183] 步骤10221:根据交互对象语音,计算得到交互对象的语速;经大量数据统计及分析,性别不同正常语速略有区别,因为是判断人的情绪,所以这里采用自发言语速度;

[0184] 步骤10222:在预先设定的语速与语速值对照表(可选的,所述语速与语速值对照表如表4所示)中,查找得到与所述语速对应的语速值c。

[0185] 表4语速与语速值对照表

[0186]	语速	慢 (SPM (男) =0~100.16 个、 SPM (女) =0~98.63 个)	较慢 (SPM (男) =150.16±50 个、 SPM (女) =128.63±30 个)	正常 (SPM (男) =226.13 ± 25.97 个、SPM (女) =198.73 ± 40.10 个)	较快 (SPM (男) =302.1±50 个、 SPM (女) =268.83±30 个)	快 (SPM (男) =352.1~∞ 个、 SPM (女) =298.83~∞ 个)
	语速 值	-2	-1	0	1	2

[0187] 其中,SPM指代每分钟的音节数,英文全称为:Syllables Per Minute。男性和女性的区分可以根据男性和女性的发音特点,通过设定分贝值的范围来区分;也可以通过初始设定时输入性别来确定。

[0188] 在一些可选实施方式中,根据所述交互对象语音,处理得到语调值d的步骤,还可具体包括以下步骤:

[0189] 步骤10223:计算得到交互对象语音的分贝值;通过采集得到的语音计算分贝值属于现有技术,在此不再赘述;

[0190] 步骤10224:在预先设定的分贝值与语调值对照表(可选的,所述分贝值与语调值对照表如下表5所示)中,查找得到与所述分贝值对应的语调值。

[0191] 表5分贝值与语调值对照表

[0192]	分贝值	低(30分贝以下)	较低(30~50分贝)	正常(50~60分贝)	较高(60~80分贝)	高(80分贝以上)
	语调值	-2	-1	0	1	2

[0193] 步骤1023:根据所述交互对象触控操作信号,处理得到触控速度值e;

[0194] 在一些可选实施方式中,根据所述交互对象触控操作信号,处理得到触控速度值e

的步骤,还可具体包括以下步骤:

[0195] 步骤10231:根据所述交互对象触控操作信号,计算得到触控时长;

[0196] 步骤10232:在预先设定的触控时长与触控速度值对照表(可选的,所述触控时长与触控速度值对照表如下表6所示)中,查找得到与所述触控时长对应的触控速度值;所述触控时长以交互对象与触控设备相触碰的时长为准,例如:交互对象有意识的触碰触控设备后,1.5s内快速收回,则认为触摸触控板的速度为快,对应的触控速度值e即为1。

[0197] 表6触控时长与触控速度值对照表

[0198]	触控时长	慢(4s以上)	正常(停留时长 1.5~4s)	快(1.5s以内)
	触控速度值	-1	0	1

[0199] 步骤1024:根据所述交互频率,处理得到互动频率值f;

[0200] 在一些可选实施方式中,根据所述交互频率,处理得到互动频率值f的步骤,还可具体包括以下步骤:

[0201] 在预先设定的交互频率与互动频率值对照表(可选的,所述交互频率与互动频率值对照表如下表7所示)中,查找得到与所述交互频率对应的互动频率值;所述交互频率,是指在预定时间间隔内,交互对象与人机交互装置之间的互动次数,参见下表7。

[0202] 表7交互频率与互动频率值对照表

[0203]	交互频率	0次/小时	2次/小时	4次/小时	6次/小时	8次/小时	10次/小时	12次/小时	14+∞次/小时
	互动频率值	-6	-4	-2	0	2	4	6	8

[0204] 在一些可选实施方式中,所述情绪总值为所述面部识别值a、交互对象速度值b、语速值c、语调值d、触控速度值e和互动频率值f中至少两项之和;这种任意的选择,可选的,以情绪总值足够表征交互对象的当前情绪状况为准。

[0205] 较佳的,在一些可选实施方式中,所述情绪总值T为所述面部识别值a、交互对象速度值b、语速值c、语调值d、触控速度值e和互动频率值f之和;即:情绪总值 $T=a+b+c+d+e+f$ 。在预先设定的情绪总值、情绪表现与交互操作(可选的,所述情绪总值、情绪表现与交互操作如下表8所示)中,查找得到与所述情绪总值对应的情绪表现。根据情绪表现找到对应的交互操作(参见下表8),并执行相应的交互操作。

[0206] 表8情绪总值、情绪表现与交互操作对照表

[0207]

情绪值	小于-16	-16-(-8)	-8-0	0	0-8	8-16	大于 16
情绪表现	低落 3	低落 2	低落 1	中性	亢进 1	亢进 2	亢进 3
交互操作	机器人开启心理医生技能、智能家居技能。例如：人在情绪极度低落时，机器人把家居调整到最适宜心理医疗的环境，然后开启心理医生技能对用户进行心理安抚。	机器人开启智能家居技能、音乐技能、chitchat（闲聊）技能。例如：人在疲劳，情绪低落时，机器人播放舒缓情绪的歌曲，家居环境调整到暖色，让人感觉到家的温馨，然后开启chitchat（闲聊）技能与用户进行聊天、讲笑话、热点话题等交互。	机器人开启音乐技能、chitchat（闲聊）技能、greeting（问候）技能。例如：在不怎么高兴时，机器人使用greeting（问候）技能可以对用户适当的问候、chitchat（闲聊）技能进行闲聊、播放一些调动情绪的歌曲。	机器人开启Greeting（问候）技能、天气技能、chitchat（闲聊）技能、音乐技能、故事技能、网上购物、预订业务、收发消息等技能。例如：一般情况下，机器人开启greeting（问候）技能和chitchat（闲聊）技能与用户进行正常的闲聊交互，在交互过程中可穿插的播放一些故事、音乐，可以网上购物、订餐、收发消息等业务。	机器人开启音乐技能、舞蹈技能、故事技能、chitchat（闲聊）、预订、网上购物、收发消息、视频播放等技能。例如：人在略微高兴时喜欢听一些故事、看一些视频、听听音乐、网上购物等活动，机器人在这种环境下都可以帮助用户完成相应的事件。	机器人开启游戏技能、故事技能、音乐技能、智能家居等技能。例如：人在高兴时，机器人与用户互动完成机器人本身设置的捉迷藏、木头人、你画我猜等游戏功能、播放一些欢快的音乐、家居环境也适当的调整到冷色调，从而达到一个舒适的娱乐环境。	机器人开启智能家居技能、游戏、故事、音乐、chitchat（闲聊）、教育等技能。例如：机器人调整环境至舒适的娱乐环境，可以给成人用户播放音乐、视频、故事，进行游戏互动、闲聊，对儿童进行英语教学、古诗词教学等等寓教于乐的技能。

[0208] 上表8是以举例的方式给出的本发明实施例中交互操作与情绪总值的对应关系，可以知道，除了上述的例举之外，还可以增删或变化上述的交互操作内容，这些增删或变化都应当属于本发明的保护范围。较佳的，这些增删或变化应当以能够较好地调节用户情绪的目的为准。

[0209] 以下以举例方式示意性地介绍所述人机交互方法如何实现：

[0210] 假设用户很晚下班，回家后身心疲惫。当用户开门回家后，采集所述交互对象图像、交互对象语音、交互对象触控操作信号和交互频率；根据所述交互图像，检测到交互对象（用户）的面部表情为不开心，其对应分值为0~-2；并且交互对象（用户）的行为速度较慢分值为-1；在进行互动时，频率不是很高，1小时内很可能也及2~3次分值为-4；说话的语调也许会低沉分值为-1~-2；语速也较慢分值为-1；人累的时候也基本不会去触摸机器人所以分值为-1；这时的情绪总值 $T=a+b+c+d+e+f=-8\sim-10$ ，则判断交互对象（用户）需要照顾及情感呵护，此时，可将智能家居的灯调到温馨暖色，给交互对象（用户）打开热水器，给交互对象（用户）播放一首缓解疲劳的歌曲，如果交互对象（用户）有聊天需求并主动进行说话

聊天,人机交互设备还可以说一些比较有意思的热点话题。

[0211] 从上述实施例可以看出,本发明实施例提供的人机交互方法,通过采集交互对象图像、交互对象语音、交互对象触控操作信号和交互频率,从而捕获到交互对象的情绪并且根据交互对象的情绪去影响交互功能的实现,对应不同的情绪主动推送不同的功能、满足用户的需求,无需用户主动下发指令,更智能化,从而达到主动为交互对象提供服务的效果;此外,在用户有需求时也能够及时响应满足用户需求。

[0212] 对比之下,单一的数据采集(例如只采集语音或只采集图像),其给出的情绪总值区间过于宽泛,对应的交互操作也就更不准确,难以达到较为精确地识别交互对象情绪的效果。

[0213] 基于前述目的,本发明实施例的第二个方面,提供了一种能够主动地提供及时的服务的人机交互装置的一个实施例。如图3所示,为本发明提供的人机交互装置的一个实施例的结构示意图。

[0214] 所述人机交互装置,可选的,为可以完成人机交互的智能机器人,具体包括:

[0215] 采集模块201,用于采集交互对象的识别数据;所述识别数据包括交互对象图像、交互对象语音、交互对象触控操作信号和交互频率中的至少两种;相应地,所述采集模块201可以包括摄像头、采集声音的装置和触控装置中的至少两种;所述交互对象,通常是指与所述人机交互装置进行交互的用户一方。

[0216] 所述交互对象图像可以是指通过摄像头采集得到的包含交互对象的图像数据。用于采集所述交互对象图像的摄像头,可以是设置在用于所述人机交互装置上(例如集成在可以完成人机交互的智能机器人上),也可以是设置在具有较好的拍摄角度的室内的某个位置,例如能够拍摄到大门处的较为完整的画面的玄关的某个位置,或者是正对着大门的客厅的某个位置等等;当摄像头不集成在所述人机交互装置上时,摄像头可以与所述人机交互装置建立连接(通常是通过网络连接),从而向所述人机交互装置发送采集得到的数据。

[0217] 所述交互对象语音,可以是指通过声音传感器、拾音器或麦克风等能够采集声音的装置采集得到的包含交互对象发出的声音的语音数据。用于采集所述交互对象语音的装置,可以是设置在所述人机交互装置上(例如集成在可以完成人机交互的智能机器人上),也可以是设置在具有较好的声音采集效果的室内的某个位置,例如能够采集到用户进门以后所发出的声音的位置(例如玄关的某个位置),或者是用户经常活动的区域,例如客厅的某个位置,等等;当采集声音的装置不集成在所述人机交互装置上时,该采集声音的装置可以与所述人机交互装置建立连接(通常是通过网络连接),从而向所述人机交互装置发送采集得到的数据。

[0218] 所述交互对象触控操作信号,可以是通过能够接收触控信号的触控装置采集得到的,所述触控信号则可以是交互对象触摸该触控装置而发出的触控信号;用于采集所述交互对象触控操作信号的触控装置,可以是设置在室内如玄关、客厅、卧室等位置的触控板等触控装置,也可以是设置在所述人机交互装置上的(例如集成在可以完成人机交互的智能机器人上);当触控装置不集成在所述人机交互装置上时,该触控装置可以与所述人机交互装置建立连接(通常是通过网络连接),从而向所述人机交互装置发送采集得到的数据;当该触控装置设置在所述人机交互装置上时,其设置的位置可以是所述人机交互装置任意部

位,例如顶部、侧面、背面等。

[0219] 当然,可以想到,用于采集所述交互对象图像、交互对象语音和交互对象触控操作信号的设备,并不局限于仅在某个地方设置一个,为了达到更全面的效果,可以在需要采集数据的室内任意位置进行设置,并且用户可以根据自己的实际需求进行增删,等等。所述交互频率则是根据用户在预定时间范围内进行人机交互的次数而定的,具体的参数可以根据用户需要进行设置,也可以采用系统默认设定值。

[0220] 计算模块202,用于根据所述识别数据,计算得到所述交互对象的情绪总值;所述情绪总值基于当前采集的所述识别数据而得出,可以用于表征所述交互对象的情绪状态。

[0221] 操作模块203,用于根据所述情绪总值,完成与所述情绪总值相对应的交互操作;在得到情绪总值后,能够根据预先设定的情绪总值(也可以是情绪总值的范围区间)与所述交互操作的对应关系,选择相应的交互操作并执行;根据情绪总值的不同,所述交互操作也可以不同,或者,在情绪总值对应的交互操作为多种时,不同情绪总值(或者情绪总值的范围区间)对应的交互操作可以有部分重叠,所述交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种。

[0222] 所述播放多媒体文件,可以是播放预先存储于某设备(该设备可以是所述人机交互装置,也可以是能够被所述人机交互装置控制的其他设备)中的语音或视频,并且不同语音或视频与相应的情绪总值相关联,在不同情绪总值时,可根据设定的方式播放所述语音或视频;除此之外,所述语音或视频还可以是通过网络连接到的互联网中的语音或视频,这种方式可以通过提取用户的所述交互对象语音中的关键词来实现;可以想到,除了语音和视频外,还可以包括歌曲、电视剧、电影等各种多媒体文件,在此不再赘述。

[0223] 所述互动问答,可以是基于预先存储于某设备(该设备可以是所述人机交互装置,也可以是能够被所述人机交互装置控制的其他设备)中的对话数据,根据交互对象发出的语音中的关键字而进行的交互对象与设备之间的一问一答;此外,当交互对象的语音中的关键字不能在本地找到匹配数据时,还可以通过网络从互联网中查找。

[0224] 所述控制家庭环境中的任意设备运行,可以是指,所述人机交互装置控制处于用户的家庭室内环境中的任何能够建立连接的电子设备或电器设备等,实现开启、关闭、调节等操作,例如开启音响并播放音乐,调节灯的颜色、亮度等,打开热水器并对浴缸进行注水,打开电视并播放用户喜欢的节目,等等。所述家庭环境中的任意设备,可以是指家庭环境中安装的具有常用功能的电子设备或电器设备,例如电视、冰箱、音响设备等等,但也可以是将一些能够集成在一起的功能集成于一体的专用设备(例如可以完成人机交互的智能机器人)。

[0225] 上述的交互操作包括播放多媒体文件、互动问答、控制家庭环境中的任意设备运行中的至少一种,在特殊情况下,可以同时包括这三种,从而使用户获得更丰富的反馈,有利于调节用户的情绪状态。

[0226] 从上述实施例可以看出,本发明实施例提供的人机交互装置,通过主动采集更为丰富的交互对象的识别数据,并基于所述识别数据得到更接近交互对象的当前状态的情绪总值,再根据情绪总值完成相应的较为丰富的交互操作,从而主动为交互对象提供更合理的服务。

[0227] 在一些可选实施方式中,所述识别数据包括交互对象图像、交互对象语音、交互对

象触控操作信号和交互频率；

[0228] 所述计算模块202,具体用于:

[0229] 根据所述交互对象图像,处理得到面部识别值a和交互对象速度值b;可选的,所述面部识别值a可通过人脸识别技术来处理得到;

[0230] 根据所述交互对象语音,处理得到语速值c和语调值d;可选的,所述语速值c可通过语音识别技术来处理得到;

[0231] 根据所述交互对象触控操作信号,处理得到触控速度值e;

[0232] 根据所述交互频率,处理得到互动频率值f;

[0233] 所述情绪总值为所述面部识别值a、交互对象速度值b、语速值c、语调值d、触控速度值e和互动频率值f中至少两项之和;这种任意的选择,可选的,以情绪总值足够表征交互对象的当前情绪状况为准。

[0234] 在一些可选实施方式中,所述计算模块202,具体用于:

[0235] 采用21点面部特征人脸识别算法识别所述交互对象图像,得到面部识别结果;

[0236] 在预先设定的面部识别结果与面部识别值对照表(可选的,所述面部识别结果与面部识别值对照表如表2所示)中,查找得到与所述面部识别结果对应的面部识别值a。

[0237] 在一些可选实施方式中,所述计算模块202,具体用于:

[0238] 获取按预设频率采集的多张交互对象图像;所述预设频率可以根据不同的需求进行设定,可选的,具体的设定值以能够通过连续采集得到的图像判断交互对象的移动速度和移动方式为准;

[0239] 计算交互对象图像中交互对象轮廓在交互对象图像中的变化;所述变化可以是交互对象在交互对象图像中的位置变化、轮廓大小的变化等等;

[0240] 根据所述预设频率和变化,计算得到交互对象的移动速度;根据两张相邻图像中交互对象的轮廓的变化,可以计算得到交互对象相对于采集装置的移动距离,而基于所述预设频率,则可计算得到交互对象移动该移动距离所耗费的时间,基于计算得到的移动距离和时间,就可以计算得到交互对象的移动速度;较佳的,为了保证计算结果的准确性,可以将多张连续的图片分别计算移动速度后取平均值;

[0241] 在预先设定的移动速度与交互对象速度值对照表(可选的,所述移动速度与交互对象速度值对照表如表3所示)中,查找得到与所述移动速度对应的交互对象速度值b。

[0242] 除了采用上述方式计算移动速度外,还可以采用其他方式来计算所述移动速度。根据运动学原理,人体重心位于第二骶骨前缘,两髋关节中央。直线运动时该中心是身体摆动幅度最小的部位。行走时人体重心不仅在水平方向,而且在垂直方向上不断改变着位置和速度。其中身体重心在垂直方向的速度变化与各关节及其活动肌肉的力学状况有密切关系。以人体重心加速度为衡量依据。因此,可以通过计算重心加速度来判断交互对象的移动速度,例如,利用交互对象随身携带的手机中的陀螺仪采集交互对象的重心加速度变化,并通过网络接收所述计算结果;或者,若所述人机交互方法对应有专用于该方法的设备(例如智能机器人)时,可通过交互对象携带所述设备运动时,经该设备中的陀螺仪等装置来计算重心加速度的变化。

[0243] 在一些可选实施方式中,所述计算模块202,具体用于:

[0244] 根据交互对象语音,计算得到交互对象的语速;通过语音识别方法,能够得到采集

得到的语音的文字表达,结合采集得到的语音的长度,即可得出交互对象的语速;经大量数据统计及分析,性别不同正常语速略有区别,因为是判断人的情绪,所以这里采用自发言语速度;

[0245] 在预先设定的语速与语速值对照表(可选的,所述语速与语速值对照表如表4所示)中,查找得到与所述语速对应的语速值c。

[0246] 在一些可选实施方式中,所述计算模块202,具体用于:

[0247] 计算得到交互对象语音的分贝值;通过采集得到的语音计算分贝值属于现有技术,在此不再赘述;

[0248] 在预先设定的分贝值与语调值对照表(可选的,所述分贝值与语调值对照表如表5所示)中,查找得到与所述分贝值对应的语调值。

[0249] 在一些可选实施方式中,所述计算模块202,具体用于:

[0250] 根据所述交互对象触控操作信号,计算得到触控时长;

[0251] 在预先设定的触控时长与触控速度值对照表(可选的,所述触控时长与触控速度值对照表如表6所示)中,查找得到与所述触控时长对应的触控速度值;所述触控时长以交互对象与触控设备相触碰的时长为准,例如:交互对象有意识的触碰触控设备后,1.5s内快速收回,则认为触摸触控板的速度为快,对应的触控速度值e即为1。

[0252] 在一些可选实施方式中,所述计算模块202,具体用于:

[0253] 在预先设定的交互频率与互动频率值对照表(可选的,所述交互频率与互动频率值对照表如表7所示)中,查找得到与所述交互频率对应的互动频率值;所述交互频率,是指在预定时间间隔内,交互对象与人机交互装置之间的互动次数,参见表7。

[0254] 较佳的,在一些可选实施方式中,所述情绪总值T为所述面部识别值a、交互对象速度值b、语速值c、语调值d、触控速度值e和互动频率值f之和;即:情绪总值 $T=a+b+c+d+e+f$ 。在预先设定的情绪总值与情绪表现对照表(可选的,所述情绪总值与情绪表现对照表如表8所示)中,查找得到与所述情绪总值对应的情绪表现。根据情绪表现找到对应的交互操作(参见表8),并执行相应的交互操作。

[0255] 以下以举例方式示意性地介绍所述人机交互方法如何实现:

[0256] 假设用户很晚下班,回家后身心疲惫。当用户开门回家后,采集所述交互对象图像、交互对象语音、交互对象触控操作信号和交互频率;根据所述交互图像,检测到交互对象(用户)的面部表情为不开心,其对应分值为0~-2;并且交互对象(用户)的行为速度较慢分值为-1;在进行互动时,频率不是很高,1小时内很可能也及2~3次分值为-4;说话的语调也许会低沉分值为-1~-2;语速也较慢分值为-1;人累的时候也基本不会去触摸机器人所以分值为-1;这时的情绪总值 $T=a+b+c+d+e+f=-8~-10$,则判断交互对象(用户)需要照顾及情感呵护,此时,可将智能家居的灯调到温馨暖色,给交互对象(用户)打开热水器,给交互对象(用户)播放一首缓解疲劳的歌曲,如果交互对象(用户)有聊天需求并主动进行说话聊天,人机交互设备还可以说一些比较有意思的热点话题。

[0257] 从上述实施例可以看出,本发明实施例提供的人机交互装置,通过采集交互对象图像、交互对象语音、交互对象触控操作信号和交互频率,从而捕获到交互对象的情绪并且根据交互对象的情绪去影响交互功能的实现,对应不同的情绪主动推送不同的功能、满足用户的需求,无需用户主动下发指令,更智能化,从而达到主动为交互对象提供服务的效

果;此外,在用户有需求时也能够及时响应满足用户需求。

[0258] 对比之下,单一的数据采集(例如只采集语音或只采集图像),其给出的情绪总值区间过于宽泛,对应的交互操作也就更不准确,难以达到较为精确地识别交互对象情绪的效果。

[0259] 基于上述目的,本发明实施例的第三个方面,提供了一种能够主动地提供及时的服务的电子设备。如图4所示,为本发明提供的电子设备的一个实施例的结构示意图。

[0260] 如图4所示,所述电子设备包括:

[0261] 用于采集图像的摄像头、用于采集语音的声音传感器和/或用于接收触摸信号的触摸装置;

[0262] 一个或多个处理器301以及存储器302,图4中以一个处理器301为例。

[0263] 所述执行所述人机交互方法的电子设备还可以包括:输入装置303和输出装置304。

[0264] 处理器301、存储器302、输入装置303和输出装置304可以通过总线或者其他方式连接,图4中以通过总线连接为例。

[0265] 存储器302作为一种非易失性计算机可读存储介质,可用于存储非易失性软件程序、非易失性计算机可执行程序以及模块,如本申请实施例中的所述人机交互方法对应的程序指令/模块(例如,附图3所示的采集模块201、计算模块202和操作模块203)。处理器301通过运行存储在存储器302中的非易失性软件程序、指令以及模块,从而执行服务器的各种功能应用以及数据处理,即实现上述方法实施例的人机交互方法。

[0266] 存储器302可以包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需要的应用程序;存储数据区可存储根据数据推荐装置的使用所创建的数据等。此外,存储器302可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实施例中,存储器302可选包括相对于处理器301远程设置的存储器,这些远程存储器可以通过网络连接至会员用户行为监控装置。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0267] 输入装置303可接收输入的数字或字符信息,以及产生与人机交互装置的用户设置以及功能控制有关的键信号输入。输出装置304可包括显示屏等显示设备。

[0268] 所述一个或者多个模块存储在所述存储器302中,当被所述一个或者多个处理器301执行时,执行上述任意方法实施例中的人机交互方法。所述执行所述人机交互方法的电子设备的实施例,其技术效果与前述任意方法实施例相同或者类似。

[0269] 所属领域的普通技术人员应当理解:以上任何实施例的讨论仅为示例性的,并非旨在暗示本公开的范围(包括权利要求)被限于这些例子;在本发明的思路下,以上实施例或者不同实施例中的技术特征之间也可以进行组合,步骤可以以任意顺序实现,并存在如上所述的本发明的不同方面的许多其它变化,为了简明它们没有在细节中提供。

[0270] 另外,为简化说明和讨论,并且为了不会使本发明难以理解,在所提供的附图中可以示出或不示出与集成电路(IC)芯片和其它部件的公知的电源/接地连接。此外,可以以框图的形式示出装置,以便避免使本发明难以理解,并且这也考虑了以下事实,即关于这些框图装置的实施方式的细节是高度取决于将要实施本发明的平台的(即,这些细节应当

完全处于本领域技术人员的理解范围内)。在阐述了具体细节(例如,电路)以描述本发明的示例性实施例的情况下,对本领域技术人员来说显而易见的是,可以在没有这些具体细节的情况下或者这些具体细节有变化的情况下实施本发明。因此,这些描述应被认为是说明性的而不是限制性的。

[0271] 尽管已经结合了本发明的具体实施例对本发明进行了描述,但是根据前面的描述,这些实施例的很多替换、修改和变型对本领域普通技术人员来说将是显而易见的。例如,其它存储器架构(例如,动态RAM(DRAM))可以使用所讨论的实施例。

[0272] 本发明的实施例旨在涵盖落入所附权利要求的宽泛范围之内的所有这样的替换、修改和变型。因此,凡在本发明的精神和原则之内,所做的任何省略、修改、等同替换、改进等,均应包含在本发明的保护范围之内。

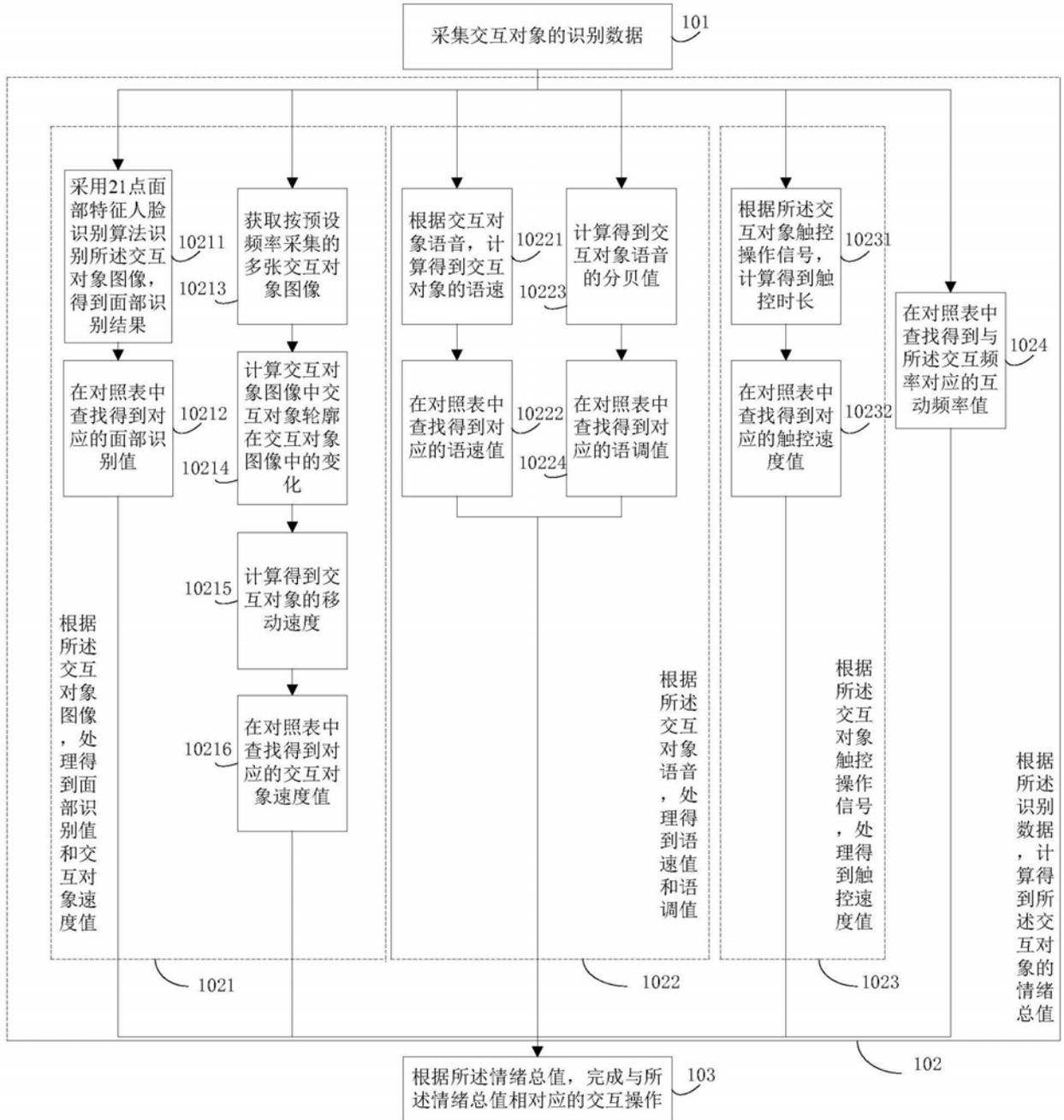


图1

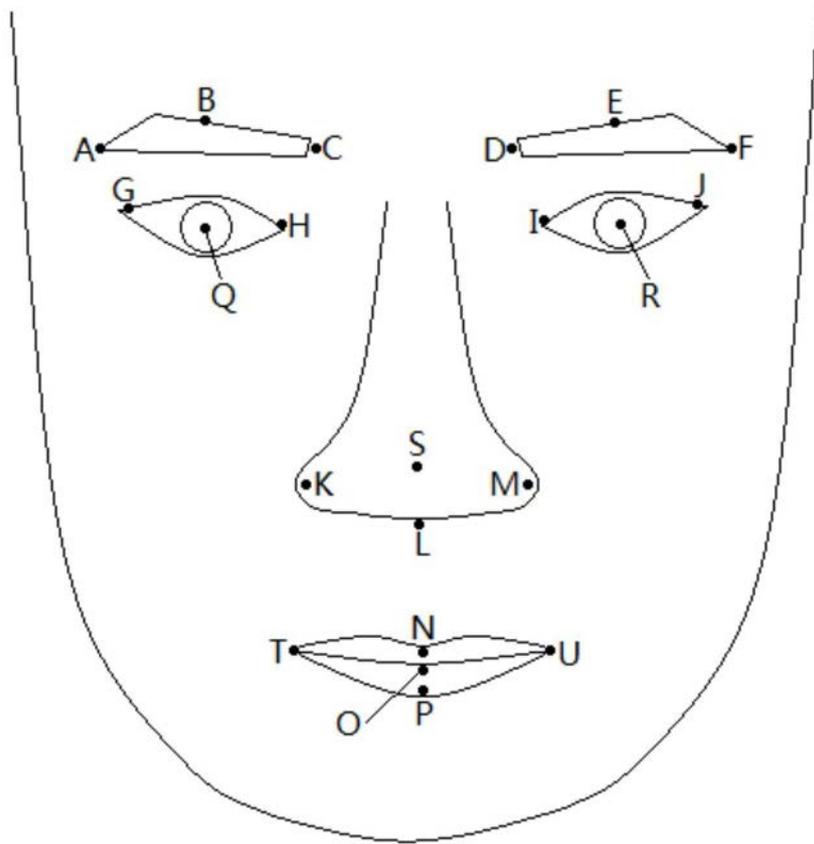


图2



图3

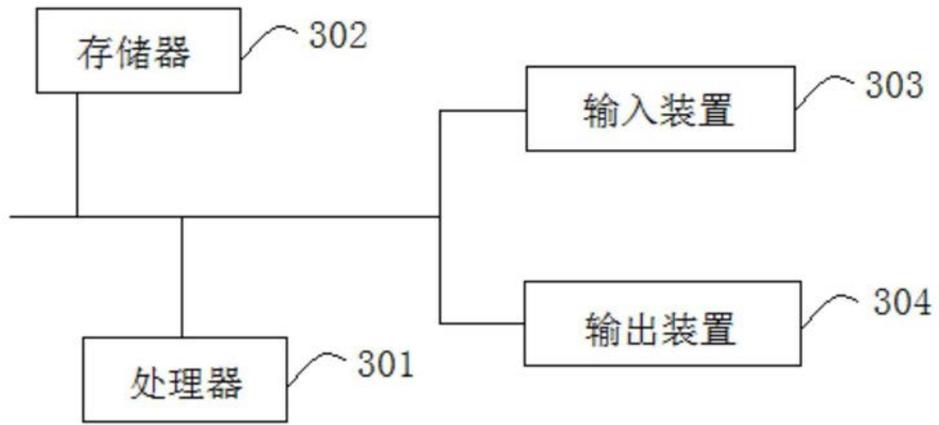


图4