



(12) 发明专利

(10) 授权公告号 CN 108628454 B

(45) 授权公告日 2022.03.22

(21) 申请号 201810442311.1

(22) 申请日 2018.05.10

(65) 同一申请的已公布的文献号

申请公布号 CN 108628454 A

(43) 申请公布日 2018.10.09

(73) 专利权人 北京光年无限科技有限公司

地址 100000 北京市石景山区石景山路3号
玉泉大厦四层常青藤青年创业工作室
193号

(72) 发明人 尚小维 李晓丹 俞志晨

(74) 专利代理机构 北京聿华联合知识产权代理
有限公司 11611

代理人 朱绘 张文娟

(51) Int. Cl.

G06F 3/01 (2006.01)

(56) 对比文件

CN 107632706 A, 2018.01.26

CN 107992191 A, 2018.05.04

CN 107765856 A, 2018.03.06

杨明浩等. 面向自然交互的多通道人机对话
系统.《计算机科学》.2014,第41卷(第10期),第
14页.

审查员 宫玉龙

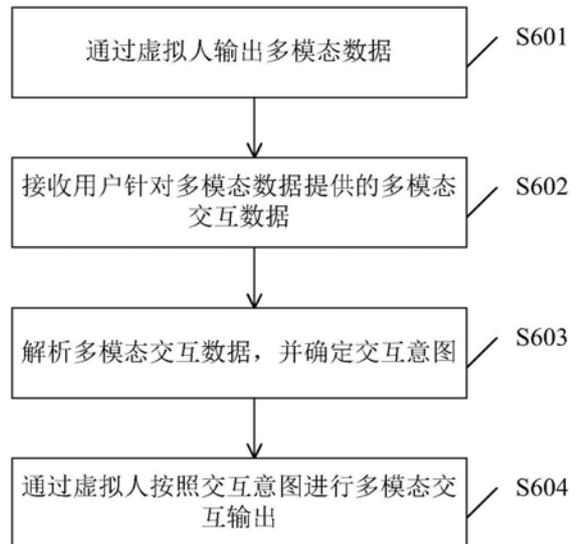
权利要求书2页 说明书11页 附图6页

(54) 发明名称

基于虚拟人的视觉交互方法及系统

(57) 摘要

本发明提供一种基于虚拟人的视觉交互方法,虚拟人通过智能设备展示,在处于交互状态时启动语音、情感、视觉以及感知能力,方法包含:通过虚拟人输出多模态数据;接收用户针对多模态数据提供的多模态交互数据;解析多模态交互数据,其中:通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图;通过虚拟人按照交互意图进行多模态交互输出。本发明提供的基于虚拟人的视觉交互方法及系统提供了一种虚拟人,虚拟人具备预设形象和预设属性,能够与用户进行多模态的交互。并且,本发明提供的虚拟人的视觉交互方法及系统还能够通过手部伸缩动作或手臂收放动作判断用户的意图,与用户展开交互,使得用户享受拟人的交互体验。



1. 一种基于虚拟人的视觉交互方法,其特征在於,所述虚拟人通过智能设备展示,在处于交互状态时启动语音、情感、视觉以及感知能力,所述方法包含以下步骤:

通过所述虚拟人输出多模态数据;

接收用户针对所述多模态数据提供的多模态交互数据;

解析所述多模态交互数据,其中:通过视觉能力检测并提取所述多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图;

通过所述虚拟人按照所述交互意图进行多模态交互输出;

当所述多模态交互数据中包含语音数据或表情数据时,依据所述手部伸缩动作或所述手臂收放动作作为交互意图,以上步骤还包含:

检测并提取所述多模态交互数据中的语音数据或表情数据;

解析所述语音数据或所述表情数据,判断所述语音数据或所述表情数据与所述手部伸缩动作或所述手臂收放动作的意图是否符合;

若符合,则根据解析的结果结合所述手部伸缩动作或所述手臂收放动作作为交互意图;

若不符合,则所述手部伸缩动作或所述手臂收放动作作为交互意图;

在判断周期内,如果所提取的用户的手臂动作与预先保存的手臂收放信息相似度大于第一阈值,则识别所述手臂动作作为手臂收放动作,手臂收放动作为两个手臂的闭合以及展开动作;

若所述手部伸缩动作为第二手势或所述手臂动作为手臂收放动作,则将所述交互意图识别为第二意图,所述第二意图表明用户的意愿为以所述第二手势或所述手臂收放动作确定的定点为中心放大或缩小目标操作对象,第二手势确定的定点为拇指与食指指腹闭合时对应智能设备显示屏上的定点,手臂收放动作确定的定点为左手与右手闭合时对应智能设备显示屏上的定点;

在通过视觉能力检测并提取所述多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图的步骤中,还包括:捕捉用户手部对应的智能设备展示画面中的操作对象,所述操作对象包括:虚拟人、系统功能画面、多媒体画面。

2. 如权利要求1所述的基于虚拟人的视觉交互方法,其特征在於,

在判断周期内,如果所提取的用户的手部动作与预先保存的手部伸缩信息相似度大于第一阈值,则识别所述手部动作作为手部伸缩动作。

3. 如权利要求2所述的基于虚拟人的视觉交互方法,其特征在於,依据所述手部伸缩动作或所述手臂收放动作将所述交互意图识别为第一意图或第二意图,其中,

若所述手部伸缩动作为第一手势,则将所述交互意图识别为所述第一意图,所述第一意图表明用户的意愿为以预设定点为中心放大或缩小目标操作对象。

4. 如权利要求1-3中任一项所述的基于虚拟人的视觉交互方法,其特征在於,所述虚拟人接收来自多个用户的针对所述多模态数据提供的多模态交互数据,识别出所述多个用户中的主要用户,并对所述主要用户的手部伸缩动作以及手臂收放动作进行检测;

或,

采集当前全部或部分用户的手部伸缩动作以及手臂收放动作,按照预设的用户采集比例确定被采集用户的交互意图。

5. 如权利要求4所述的基于虚拟人的视觉交互方法,其特征在于,通过所述虚拟人按照所述交互意图进行多模态交互输出,包括:通过所述虚拟人按照所述手部伸缩动作或所述手臂收放动作对应的交互意图输出并展示多模态交互输出,所述多模态交互输出包括:第一意图或第二意图对应的结果数据。

6. 一种存储介质,其包含用于执行如权利要求1-5中任一项所述的方法步骤的一系列指令。

7. 一种虚拟人,其特征在于,所述虚拟人具备特定的虚拟形象和预设属性,采用如权利要求1-5中任一项所述的方法进行多模态的交互。

8. 一种基于虚拟人的视觉交互系统,其特征在于,所述系统包含:

智能设备,其上装载有如权利要求7所述的虚拟人,用于获取多模态交互数据,并具备语言、情感、表情和动作输出的能力;

云端大脑,其用于对所述多模态交互数据进行语义理解、视觉识别、认知计算以及情感计算,以决策所述虚拟人输出多模态交互数据。

基于虚拟人的视觉交互方法及系统

技术领域

[0001] 本发明涉及人工智能领域,具体地说,涉及一种基于虚拟人的视觉交互方法及系统。

背景技术

[0002] 机器人多模态交互系统的开发致力于模仿人类对话,以试图在上下文之间模仿人类之间的交互。但是,目前来说,对于虚拟人相关的机器人多模态交互系统的开发还不太完善,尚未出现进行多模态交互的虚拟人,更为重要的是,尚无针对肢体,尤其针对手势交互,且对肢体、尤其手势交互有响应的基于虚拟人的视觉交互产品。

[0003] 因此,本发明提供了一种基于虚拟人的视觉交互方法及系统。

发明内容

[0004] 为解决上述问题,本发明提供了一种基于虚拟人的视觉交互方法,所述虚拟人通过智能设备展示,在处于交互状态时启动语音、情感、视觉以及感知能力,所述方法包含以下步骤:

[0005] 通过所述虚拟人输出多模态数据;

[0006] 接收用户针对所述多模态数据提供的多模态交互数据;

[0007] 解析所述多模态交互数据,其中:通过视觉能力检测并提取所述多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图;

[0008] 通过所述虚拟人按照所述交互意图进行多模态交互输出。

[0009] 根据本发明的一个实施例,在判断周期内,如果所提取的用户的手部动作与预先保存的手部伸缩信息相似度大于第一阈值,则识别所述手部动作作为手部伸缩动作;

[0010] 在判断周期内,如果所提取的用户的手臂动作与预先保存的手臂收放信息相似度大于第一阈值,则识别所述手臂动作作为手臂收放动作。

[0011] 根据本发明的一个实施例,依据所述手部伸缩动作或所述手臂收放动作将所述交互意图识别为第一意图或第二意图,其中,

[0012] 若所述手部伸缩动作作为第一手势,则将所述交互意图识别为所述第一意图,所述第一意图表明用户的意愿为以预设定点为中心放大或缩小目标操作对象;

[0013] 若所述手部伸缩动作作为第二手势或所述手臂动作作为手臂收放动作,则将所述交互意图识别为所述第二意图,所述第二意图表明用户的意愿为以所述第二手势或所述手臂收放动作确定的定点为中心放大或缩小目标操作对象。

[0014] 根据本发明的一个实施例,在通过视觉能力检测并提取所述多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图的步骤中,还包括:捕捉用户手部对应的智能设备展示画面中的操作对象,所述操作对象包括:虚拟人、系统功能画面、多媒体画面。

[0015] 根据本发明的一个实施例,所述虚拟人接收来自多个用户的针对所述多模态数据提供的多模态交互数据,识别出所述多个用户中的主要用户,并对所述主要用户的手部伸

缩动作以及手臂收放动作进行检测；

[0016] 或，

[0017] 采集当前全部或部分用户的手部伸缩动作以及手臂收放动作，按照预设的用户采集比例确定被采集用户的交互意图。

[0018] 根据本发明的一个实施例，当所述多模态交互数据中包含语音数据或表情数据时，依据所述手部伸缩动作或所述手臂收放动作作为交互意图，以上步骤还包含：

[0019] 检测并提取所述多模态交互数据中的语音数据或表情数据；

[0020] 解析所述语音数据或所述表情数据，判断所述语音数据或所述表情数据与所述手部伸缩动作或所述手臂收放动作的意图是否符合；

[0021] 若符合，则根据解析的结果结合所述手部伸缩动作或所述手臂收放动作作为交互意图；

[0022] 若不符合，则所述手部伸缩动作或所述手臂收放动作作为交互意图。

[0023] 根据本发明的一个实施例，通过所述虚拟人按照所述交互意图进行多模态交互输出，包括：通过所述虚拟人按照所述手部伸缩动作或所述手臂收放动作对应的交互意图输出并展示多模态交互输出，所述多模态交互输出包括：第一意图或第二意图对应的结果数据。

[0024] 根据本发明的另一个方面，还提供了一种程序产品，其包含用于执行如上任一项所述的方法步骤的一系列指令。

[0025] 根据本发明的另一个方面，还提供了一种虚拟人，所述虚拟人具备特定的虚拟形象和预设属性，采用如上任一项所述的方法进行多模态的交互。

[0026] 根据本发明的另一个方面，还提供了一种基于虚拟人的视觉交互系统，所述系统包含：

[0027] 智能设备，其上装载有如上所述的虚拟人，用于获取多模态交互数据，并具备具备语言、情感、表情和动作输出的能力；

[0028] 云端大脑，其用于对所述多模态交互数据进行语义理解、视觉识别、认知计算以及情感计算，以决策所述虚拟人输出多模态交互数据。

[0029] 本发明提供的基于虚拟人的视觉交互方法及系统提供了一种虚拟人，虚拟人具备预设形象和预设属性，能够与用户进行多模态的交互。并且，本发明提供的虚拟人的视觉交互方法及系统还能够通过手部伸缩动作或手臂收放动作判断用户的意图，与用户展开交互，使得用户与虚拟人之间能够进行流畅的交流，并使得用户享受拟人的交互体验。

[0030] 本发明的其它特征和优点将在随后的说明书中阐述，并且，部分地从说明书中变得显而易见，或者通过实施本发明而了解。本发明的目的和其他优点可通过在说明书、权利要求书以及附图中所特别指出的结构来实现和获得。

附图说明

[0031] 附图用来提供对本发明的进一步理解，并且构成说明书的一部分，与本发明的实施例共同用于解释本发明，并不构成对本发明的限制。在附图中：

[0032] 图1显示了根据本发明的一个实施例的基于虚拟人的视觉交互系统的结构框图；

[0033] 图2显示了根据本发明的一个实施例的基于虚拟人的视觉交互系统的结构框图；

[0034] 图3显示了根据本发明的另一个实施例的基于虚拟人的视觉交互系统的模块框图；

[0035] 图4显示了根据本发明的另一个实施例的基于虚拟人的视觉交互系统的结构框图；

[0036] 图5显示了根据本发明的一个实施例的基于虚拟人的视觉交互系统进行视觉交互的示意图；

[0037] 图6显示了根据本发明的一个实施例的基于虚拟人的视觉交互方法的流程图；

[0038] 图7显示了根据本发明的一个实施例的基于虚拟人的视觉交互方法的确定交互意图的流程图；

[0039] 图8显示了根据本发明的另一个实施例的基于虚拟人的视觉交互方法的确定交互意图的流程图；

[0040] 图9显示了根据本发明的一个实施例的基于虚拟人的视觉交互方法的另一流程图；以及

[0041] 图10显示了根据本发明的一个实施例的在用户、智能设备以及云端大脑三方之间进行通信的流程图。

具体实施方式

[0042] 为使本发明的目的、技术方案和优点更加清楚，以下结合附图对本发明实施例作进一步地详细说明。

[0043] 为表述清晰，需要在实施例前进行如下说明：

[0044] 本发明提到的虚拟人搭载于支持感知、控制等输入输出模块的智能设备；以高仿真3d虚拟人物形象为主要用户界面，具备显著人物特征的外观；支持多模态人机交互，具备自然语言理解、视觉感知、触摸感知、语言语音输出、情感表情动作输出等AI能力；可配置社会属性、人格属性、人物技能等，使用户享受智能化及个性化流畅体验的虚拟人物。

[0045] 虚拟人所搭载的智能设备为：具备非触摸、非鼠标键盘输入的屏幕（全息、电视屏、多媒体显示屏、LED屏等），并携带有摄像头的智能设备，同时，可以是全息设备、VR设备、PC机。但并不排除其他智能设备，如：手持平板、裸眼3D设备、甚至智能手机等。

[0046] 虚拟人在系统层面与用户进行交互，所述系统硬件中运行操作系统，如全息设备内置系统，如PC则为windows或MAC OS。

[0047] 虚拟人为系统应用程序，或者可执行文件。

[0048] 虚拟机器人基于所述智能设备的硬件获取用户多模态交互数据，在云端大脑的能力支持下，对多模态交互数据进行语义理解、视觉识别、认知计算、情感计算，以完成决策输出的过程。

[0049] 所提到的云端大脑为提供所述虚拟人对用户的交互需求进行语义理解（语言语义理解、动作语义理解、视觉识别、情感计算、认知计算）的处理能力的终端，实现与用户的交互，以决策所述虚拟人的输出多模态交互数据。

[0050] 下面结合附图对本发明的各个实施例进行详细描述。

[0051] 图1显示了根据本发明的一个实施例的基于虚拟人的视觉交互系统的结构框图。如图1所示，进行多模态交互需要用户101、智能设备102、虚拟人103以及云端大脑104。其

中,与虚拟人交互的用户101可以为真实人、另一个虚拟人以及实体的虚拟人,另一虚拟人以及实体虚拟人与虚拟人的交互过程与单个的人与虚拟人的交互过程类似。因此,在图1中仅展示的是用户(人)与虚拟人的多模态交互过程。

[0052] 另外,智能设备102包括显示区域1021以及硬件支持设备1022(实质为核心处理器)。显示区域1021用于显示虚拟人103的形象,硬件支持设备1022与云端大脑104配合使用,用于交互过程中的数据处理。虚拟人103需要屏显载体来呈现。因此,显示区域1021包括:全息屏、电视屏、多媒体显示屏以及LED屏等。

[0053] 图1中虚拟人与用户101之间交互的过程为:

[0054] 交互所需的前期准备或是条件有,虚拟人搭载并运行在智能设备102上,并且虚拟人具备特定的形象特征。虚拟人具备自然语言理解、视觉感知、触摸感知、语言输出、情感表情动作输出等AI能力。为了配合虚拟人的触摸感知功能,智能设备上也需要安装有具备触摸感知功能的部件。根据本发明的一个实施例,为了提升交互的体验,虚拟人在被启动后就显示在预设区域内。

[0055] 在此需要说明的是,虚拟人103的形象以及装扮不限于一种模式。虚拟人103可以具备不同的形象以及装扮。虚拟人103的形象一般为3D高模动画形象。虚拟人103可以具备不同的外貌以及装饰。每种虚拟人103的形象还会对应多种不同的装扮,装扮的分类可以依据季节分类,也可以依据场合分类。这些形象以及装扮可以存在于云端大脑104中,也可以存在于智能设备102中,在需要调用这些形象以及装扮时可以随时调用。

[0056] 虚拟人103的社会属性、人格属性以及人物技能也不限于一种或是一类。虚拟人103可以具备多种社会属性、多种人格属性以及多种人物技能。这些社会属性、人格属性以及人物技能可以分别搭配,并不固定于一种搭配方式,用户可以根据需要进行选择与搭配。

[0057] 具体来说,社会属性可以包括:外貌、姓名、服饰、装饰、性别、籍贯、年龄、家庭关系、职业、职位、宗教信仰、感情状态、学历等属性;人格属性可以包括:性格、气质等属性;人物技能可以包括:唱歌、跳舞、讲故事、培训等专业技能,并且人物技能展示不限于肢体、表情、头部和/或嘴部的技能展示。

[0058] 在本申请中,虚拟人的社会属性、人格属性和人物技能等可以使得多模态交互的解析和决策结果更倾向或更为适合该虚拟人。

[0059] 以下为多模态交互过程,首先,通过虚拟人输出多模态数据。在虚拟人103与用户101交流时,虚拟人103首先输出多模态数据,以等待用户101对于多模态数据的回应。在实际运用当中,虚拟人103可能输出一段话、一段音乐或一段视频。

[0060] 接着,接收用户针对多模态数据提供的多模态交互数据。多模态交互数据可以包含文本、语音、视觉以及感知信息等多种模态的信息。获取多模态交互数据的接收装置均安装或是配置于智能设备102上,这些接收装置包含接收文本的文本接收装置,接收语音的语音接收装置,接收视觉的摄像头以及接收感知信息的红外线设备等。

[0061] 然后,解析多模态交互数据,其中:通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图。在判断周期内,如果所提取的用户的手部动作与预先保存的手部伸缩信息相似度大于第一阈值,则识别手部动作作为手部伸缩动作。在判断周期内,如果所提取的用户的手臂动作与预先保存的手臂收放信息相似度大于第一阈值,则识别手臂动作作为手臂收放动作。

[0062] 所述手部伸缩信息包括:手部在判断周期内,手部关节关键点位置和手部关节关键点关系、手部关节关键点动态变化过程对应的意图,所述意图指示对操作对象的放大及缩小。

[0063] 手臂收放信息包括:手臂在判断短周期内,手掌关键点位置与手臂关键点关系、手肘关键点位置与肩膀关键点关系、两只手臂关键点之间的关系表示,所述关系表示可以是手臂动态变化过程的对应的意图,所述意图指示对操作对象的放大及缩小。

[0064] 最后,通过虚拟人按照交互意图进行多模态交互输出。

[0065] 另外,虚拟人103还可以接收来自多个用户的针对多模态数据提供的多模态交互数据,识别出多个用户中的主要用户,并对主要用户的手部伸缩动作以及手臂收放动作进行检测。或者,虚拟人103采集当前全部或部分用户的手部伸缩动作以及手臂收放动作,按照预设的用户采集比例确定被采集用户的交互意图。

[0066] 根据本发明的另一个实施例,一种虚拟人,虚拟人具备特定的虚拟形象和预设属性,采用基于虚拟人的视觉交互方法进行多模态的交互。

[0067] 图2显示了根据本发明的一个实施例的基于虚拟人的视觉交互系统的结构框图。如图2所示,通过系统完成多模态交互需要:用户101、智能设备102以及云端大脑104。其中,智能设备102包含接收装置102A、处理装置102B、输出装置102C以及连接装置102D。云端大脑104包含通信装置104A。

[0068] 在本发明提供的基于虚拟人的视觉交互系统需要在用户101、智能设备102以及云端大脑104之间建立通畅的通信通道,以便能够完成用户101与虚拟人的交互。为了完成交互的任务,智能设备102以及云端大脑104会设置有支持完成交互的装置以及部件。与虚拟人交互的对象可以为一方,也可以为多方。

[0069] 智能设备102包含接收装置102A、处理装置102B、输出装置102C以及连接装置102D。其中,接收装置102A用于接收多模态交互数据。接收装置102A的例子包括用于语音操作的麦克风、扫描仪、摄像头(采用可见或不可见波长检测不涉及触摸的动作)等等。智能设备102可以通过以上提到的输入设备来获取多模态交互数据。输出装置102C用于输出虚拟人与用户101交互的多模态输出数据,与接收装置102A的配置基本相当,在此不再赘述。

[0070] 处理装置102B用于处理交互过程中由云端大脑104传送的交互数据。连接装置102D用于与云端大脑104之间的联系,处理装置102B处理接收装置102A预处理的多模态交互数据或由云端大脑104传送的数据。连接装置102D发送调用指令来调用云端大脑104上的机器人能力。

[0071] 云端大脑104包含的通信装置104A用于完成与智能设备102之间的通信联系。通信装置104A与智能设备102上的连接装置102D之间保持通讯联系,接收智能设备102发来的请求,并发送云端大脑104发出的处理结果,是智能设备102以及云端大脑104之间沟通的介质。

[0072] 图3显示了根据本发明的另一个实施例的基于虚拟人的视觉交互系统的模块框图。如图3所示,系统包含交互模块301、接收模块302、解析模块303以及决策模块304。其中,接收模块302包含文本采集单元3021、音频采集单元3022、视觉采集单元3023以及感知采集单元3024。

[0073] 交互模块301用于通过虚拟人输出多模态数据。虚拟人103通过智能设备102展示,

在处于交互状态时启动语音、情感、视觉以及感知能力。在一轮交互中，虚拟人103首先输出多模态数据，以等待用户101对于多模态数据的回应。根据本发明的一个实施例，交互模块301包含输出单元3011。输出单元3011能够输出多模态数据。

[0074] 接收模块302用于接收多模态交互数据。其中，文本采集单元3021用来采集文本信息。音频采集单元3022用来采集音频信息。视觉采集单元3023用来采集视觉信息。感知采集单元3024用来采集感知信息。接收模块302的例子包括用于语音操作的麦克风、扫描仪、摄像头、感控设备，如采用可见或不可见波长射线、信号、环境数据等等。可以通过以上提到的输入设备来获取多模态交互数据。多模态交互可以包含文本、音频、视觉以及感知数据中的一种，也可以包含多种，本发明不对此作出限制。

[0075] 解析模块303用于解析多模态交互数据，其中：通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图。其中，解析模块303包含检测单元3031以及提取单元3032。检测单元3031用于通过视觉能力检测多模态交互数据中的手部伸缩动作或手臂收放动作。

[0076] 如果检测单元3031检测到多模态交互数据中存在手部伸缩动作或手臂收放动作，则提取单元3032提取手部伸缩动作或手臂收放动作，并将手部伸缩动作或手臂收放动作作为交互意图。根据本发明的一个实施例，交互意图分为两类，分别为第一意图以及第二意图。判断交互意图的类别的过程可以是：若手部伸缩动作为第一手势，则将交互意图识别为第一意图，第一意图表明用户的意愿为以预设定点为中心放大或缩小目标操作对象。根据本发明的一个实施例，第一手势为用户任一只手或两只手的所有手指从蜷缩到向外舒展或从舒展到向内蜷缩的动作，所述目标操作对象包括：虚拟人、系统功能画面、多媒体画面。

[0077] 另外，若手部伸缩动作为第二手势或手臂动作为手臂收放动作，则将交互意图识别为第二意图，第二意图表明用户的意愿为以第二手势或手臂收放动作确定的定点为中心放大或缩小目标操作对象。根据本发明的一个实施例，第二手势为用户任一只手大拇指与食指从指腹闭合到向外打开或从打开到指腹闭合的动作，所述目标操作对象包括：虚拟人、系统功能画面、多媒体画面。

[0078] 输出模块304用于通过虚拟人按照交互意图进行多模态交互输出。通过解析模块303确定交互意图后，输出模块304会输出符合交互意图的多模态交互输出。输出模块304包含输出数据单元3041，其能够根据交互意图确定需要输出的多模态交互输出，并通过虚拟人将多模态交互输出展示给用户101。

[0079] 图4显示了根据本发明的另一个实施例的基于虚拟人的视觉交互系统的结构框图。如图4所示，完成交互需要用户101、智能设备102以及云端大脑104。其中，智能设备102包含人机界面401、数据处理单元402、输入输出装置403以及接口单元404。云端大脑104包含语义理解接口1041、视觉识别接口1042、认知计算接口1043以及情感计算接口1044。

[0080] 本发明提供的基于虚拟人的视觉交互系统包含智能设备102以及云端大脑104。虚拟人103在智能设备102中运行，且虚拟人103具备预设形象和预设属性，在处于交互状态时可以启动语音、情感、视觉和感知能力。

[0081] 在一个实施例中，智能设备102可以包括：人机界面401、数据处理单元402、输入输出装置403以及接口单元404。其中，人机界面401在智能设备102的预设区域内显示处于运行状态的虚拟人103。

[0082] 数据处理单元402用于处理用户101与虚拟人103进行多模态交互过程中产生的数据。所用的处理器可以为数据处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等,处理器是终端的控制中心,利用各种接口和线路连接整个终端的各个部分。

[0083] 智能设备102中包含存储器,存储器主要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、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据智能设备102的使用所创建的数据(比如音频数据、浏览记录等)等。此外,存储器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,例如硬盘、内存、插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0084] 输入输出装置403用于获取多模态交互数据以及输出交互过程中的输出数据。接口单元404用于与云端大脑104展开通信,通过与云端大脑104中的接口对接来调取云端大脑104中的虚拟人能力。

[0085] 云端大脑104包含语义理解接口1041、视觉识别接口1042、认知计算接口1043以及情感计算接口1044。以上这些接口与智能设备102中的接口单元404展开通信。并且,云端大脑104还包含与语义理解接口1041对应的语义理解逻辑、与视觉识别接口1042对应的视觉识别逻辑、与认知计算接口1043对应的认知计算逻辑以及与情感计算接口1044对应的情感计算逻辑。

[0086] 如图4所示,多模态数据解析过程中各个能力接口分别调用对应的逻辑处理。以下为各个接口的说明:

[0087] 语义理解接口1041,其接收从接口单元404转发的特定语音指令,对其进行语音识别以及基于大量语料的自然语言处理。

[0088] 视觉识别接口1042,可以针对人体、人脸、场景依据计算机视觉算法、深度学习算法等进行视频内容检测、识别、跟踪等。即根据预定的算法对图像进行识别,给出定量的检测结果。具备图像预处理功能、特征提取功能、决策功能和具体应用功能;

[0089] 其中,图像预处理功能可以是对获取的视觉采集数据进行基本处理,包括颜色空间转换、边缘提取、图像变换和图像阈值化;

[0090] 特征提取功能可以提取出图像中目标的肤色、颜色、纹理、运动和坐标等特征信息;

[0091] 决策功能可以是对特征信息,按照一定的决策策略分发给需要该特征信息的具体多模态输出设备或多模态输出应用,如实现人脸检测、人物肢体识别、运动检测等功能。

[0092] 认知计算接口1043,其接收从接口单元404转发的多模态数据,认知计算接口1043用以处理多模态数据进行数据采集、识别和学习,以获取用户画像、知识图谱等,以对多模态输出数据进行合理决策。

[0093] 情感计算接口1044,其接收从接口单元404转发的多模态数据,利用情感计算逻辑

(可以是情绪识别技术)来计算用户当前的情绪状态。情绪识别技术是情感计算的一个重要组成部分,情绪识别研究的内容包括面部表情、语音、行为、文本和生理信号识别等方面,通过以上内容可以判断用户的情绪状态。情绪识别技术可以仅通过视觉情绪识别技术来监控用户的情绪状态,也可以采用视觉情绪识别技术和声音情绪识别技术结合的方式来监控用户的情绪状态,且并不局限于此。在本实施例中,优选采用二者结合的方式来监控情绪。

[0094] 情感计算接口1044是在进行视觉情绪识别时,通过使用图像采集设备收集人类面部表情图像,而后转换成可分析数据,再利用图像处理等技术进行表情情绪分析。理解面部表情,通常需要对表情的微妙变化进行检测,比如脸颊肌肉、嘴部的变化以及挑眉等。

[0095] 图5显示了根据本发明的一个实施例的基于虚拟人的视觉交互系统进行视觉交互的示意图。如图5所示,解析多模态交互数据,其中:通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图。智能设备102上配置能够进行视觉能力检测的硬件设备,用来监测用户的手部伸缩动作以及手臂收放动作。

[0096] 在本发明的一个实施例中,在判断周期内,如果所提取的用户的手部动作与预先保存的手部伸缩信息相似度大于第一阈值,则识别手部动作作为手部伸缩动作。在判断周期内,如果所提取的用户的手臂动作与预先保存的手臂收放信息相似度大于第一阈值,则识别手臂动作作为手臂收放动作。

[0097] 并且,可以判定为手部伸缩动作的手部动作包含第一手势以及第二手势。第一手势为用户任一只手或两只手的所有手指从蜷缩到向外舒展或从舒展到向内蜷缩的动作,第二手势为用户任一只手大拇指与食指从指腹闭合到向外打开或从打开到指腹闭合的动作。

[0098] 当识别手部伸缩动作为第一手势,则将交互意图识别为第一意图。第一意图表明用户的意愿为以预设定点为中心放大或缩小目标操作对象。当识别手部伸缩动作为第二手势或手臂动作为手臂收放动作,则将交互意图识别为第二意图,第二意图表明用户的意愿为以第二手势或手臂收放动作确定的定点为中心放大或缩小目标操作对象。

[0099] 另外,在通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图的步骤中,还包括:捕捉用户手部对应的智能设备展示画面中的操作对象,操作对象包括:虚拟人、系统功能画面、多媒体画面。

[0100] 确定用户的交互意图后,可以通过虚拟人按照所述交互意图进行多模态交互输出,包括:通过虚拟人按照所述手部伸缩动作或手臂收放动作对应的交互意图输出并展示多模态交互输出,多模态交互输出包括:第一意图或第二意图对应的结果数据。

[0101] 如图5所示,第一手势501可以是用户任一只手或两只手的所有手指从蜷缩到向外舒展或从舒展到向内蜷缩的动作,第二手势502可以是用户任一只手大拇指与食指从指腹闭合到向外打开或从打开到指腹闭合的动作。手臂收放动作503可以是两个手臂的闭合以及展开动作。

[0102] 当手部伸缩动作被识别为第一手势,则按照用户的意图以预设定点为中心放大或缩小目标操作对象。预设定点可以设置为智能设备显示屏的中心,也可以是用户提前设置的其他位置,本发明不对此做出限制。

[0103] 当手部伸缩动作被识别为第二手势,则按照用户的意愿以第二手势确定的定点为中心放大或缩小目标操作对象。根据本发明的一个实施例,第二手势确定的定点可以为拇指与食指指腹闭合时对应智能设备显示屏上的定点。

[0104] 当手臂动作被识别为手臂收放动作时,则按照用户的意愿以手臂收放动作确定的定点为中心放大或缩小目标操作对象。根据本发明的一个实施例,手臂收放动作确定的定点可以为左手与右手闭合时对应智能设备显示屏上的定点。

[0105] 需要说明的是,第一手势以及第二手势一般应用在屏幕尺寸不大的智能设备上,例如手持终端桌面等设备。手臂收放动作一般应用于屏幕尺寸较大的智能设备上,例如,多媒体显示屏或LED屏等设备。

[0106] 另外,第一手势可以是左手,也可以是右手,也可以是两只手。第二手势可以是拇指与食指,也可以是其他能够实现缩放功能的任两只手指。手臂动作的展开与闭合方向可以是任意方向,本发明不对以上做出限制。

[0107] 图6显示了根据本发明的一个实施例的基于虚拟人的视觉交互方法的流程图。

[0108] 如图6所示,在步骤S601中,通过虚拟人输出多模态数据。在本步骤中,智能设备102中的虚拟人103向用户101输出多模态数据,以期在一轮交互中与用户101展开对话或其他交互。虚拟人103输出的多模态数据可以是一段话、一段音乐或一段视频。

[0109] 在步骤S602中,接收用户针对多模态数据提供的多模态交互数据。在本步骤中,智能设备102会获取多模态交互数据,智能设备102会配置有获取多模态交互数据的相应装置。多模态交互数据可以是文本输入、音频输入以及感知输入等形式的输入。

[0110] 在步骤S603中,解析多模态交互数据,其中:通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图。多模态交互数据中可能会包含手部伸缩动作或手臂收放动作,也可能不包含手部伸缩动作或手臂收放动作,为了确定交互意图,需要检测多模态交互数据中是否包含手部伸缩动作或手臂收放动作。在判断周期内,如果所提取的用户的手部动作与预先保存的手部伸缩信息相似度大于第一阈值,则识别手部动作作为手部伸缩动作。在判断周期内,如果所提取的用户的手臂动作与预先保存的手臂收放信息相似度大于第一阈值,则识别手臂动作作为手臂收放动作。

[0111] 在本步骤中,首先检测多模态交互数据中是否包含手部伸缩动作或手臂收放动作,如果多模态交互数据中包含手部伸缩动作或手臂收放动作,那么将手部伸缩动作或手臂收放动作作为本轮交互的交互意图。如果多模态交互数据中不包含手部伸缩动作或手臂收放动作,那么将根据多模态交互数据中的其他数据作为交互意图。

[0112] 在本发明的一个实施例中,交互意图分为第一意图以及第二意图。第一意图表示用户的意愿为以预设定点为中心放大或缩小目标操作对象。第二意图表示用户的意愿为以第二手势或手臂收放动作确定的定点为中心放大或缩小目标操作对象。

[0113] 最后,在步骤S604中,通过虚拟人按照交互意图进行多模态交互输出。确定了交互意图后,虚拟人103就可以根据确认的交互意图输出相应的多模态交互输出。

[0114] 此外,本发明提供的基于虚拟人的视觉交互系统还可以配合一种程序产品,其包含用于执行完成虚拟人的视觉交互方法步骤的一系列指令。程序产品能够运行计算机指令,计算机指令包括计算机程序代码,计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。

[0115] 程序产品可以包括:能够携带计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质等。

[0116] 需要说明的是,程序产品包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,程序产品不包括电载波信号和电信信号。

[0117] 图7显示了根据本发明的一个实施例的基于虚拟人的视觉交互方法的确定交互意图的流程图。

[0118] 在步骤S701中,解析多模态交互数据,其中:通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图。在本步骤中,需要对多模态交互数据进行解析,多模态交互数据包含多种形式的交互数据。为了获知交互意图,需要检测多模态交互数据中是否包含手部伸缩动作或手臂收放动作。当检测到多模态交互数据中包含手部伸缩动作或手臂收放动作后,需要提取出检测到的手部伸缩动作或手臂收放动作,并以手部伸缩动作或手臂收放动作作为交互意图。

[0119] 根据本发明的一个实施例,交互意图分为两类,分别是第一意图以及第二意图。在步骤S702中,当时别手部伸缩动作为第一手势,则将交互意图识别为第一意图,其中,第一意图表明用户的意愿为以预设定点为中心放大或缩小目标操作对象。

[0120] 同时,在步骤S703中,当识别手部伸缩动作为第二手势或手臂收放动作,则将交互意图识别为第二意图,其中,第二意图表明用户的意愿为以第二手势或手臂收放动作确定的定点为中心放大或缩小目标操作对象。最后,在步骤S704中,通过虚拟人按照交互意图进行多模态交互输出。

[0121] 图8显示了根据本发明的另一个实施例的基于虚拟人的视觉交互方法的确定交互意图的流程图。

[0122] 在步骤S801中,检测并提取多模态交互数据中的语音数据或表情数据。在多模态交互数据中包含多种形式的交互数据,这些数据都可能包含用户101当前的交互意愿。在本步骤中,检测多模态交互数据中是否包含语音数据或是表情数据,以为确定交互意图做出参考。

[0123] 接着,在步骤S802中,解析语音数据或表情数据。如果多模态交互数据中包含语音数据或是表情数据,在本步骤中,解析语音数据或表情数据,获知语音数据或表情数据中用户的交互意愿,得到解析结果。

[0124] 然后,在步骤S803中,判断语音数据或表情数据与手部伸缩动作或手臂收放动作的意图是否符合。如果语音数据或表情数据与手部伸缩动作或手臂收放动作的意图符合,则进入步骤S804,根据解析的结果结合手部伸缩动作或手臂收放动作作为交互意图。如果语音数据或表情数据与手部伸缩动作或手臂收放动作的意图不符合,则进入步骤S805,手部伸缩动作或手臂收放动作作为交互意图。

[0125] 图9显示了根据本发明的一个实施例的基于虚拟人的视觉交互方法的另一流程图。

[0126] 如图9所示,在步骤S901中,智能设备102向云端大脑104发出请求。之后,在步骤S902中,智能设备102一直处于等待云端大脑104回复的状态。在等待的过程中,智能设备102会对返回数据所花费的时间进行计时操作。

[0127] 在步骤S903中,如果长时间未得到返回的应答数据,比如,超过了预定的时间长度5S,则智能设备102会选择进行本地回复,生成本地常用应答数据。然后,在步骤S904中,输出与本地常用应答配合的动画,并调用语音播放设备进行语音播放。

[0128] 图10显示了根据本发明的一个实施例的在用户、智能设备以及云端大脑三方之间进行通信的流程图。

[0129] 为了实现智能设备102与用户101之间的多模态交互,需要用户101、智能设备102以及云端大脑104之间建立起通信连接。这种通信连接应该是实时的、通畅的,能够保证交互不受影响的。

[0130] 为了完成交互,需要具备一些条件或是前提。这些条件或是前提包含,智能设备102中装载并运行虚拟人,并且智能设备102具备感知以及控制功能的硬件设施。虚拟人在处于交互状态时启动语音、情感、视觉以及感知能力。

[0131] 完成前期准备后,智能设备102开始与用户101展开交互,首先,智能设备102通过虚拟人103输出多模态数据。多模态数据可以是在一轮交互中,虚拟人输出的一段话、一段音乐或一段视频。此时,展开通信的两方是智能设备102与用户101,数据传递的方向是从智能设备102传向用户101。

[0132] 然后,智能设备102接收多模态交互数据。多模态交互数据是用户针对多模态数据提供的回应。多模态交互数据中可以包含多种形式的的数据,例如,多模态交互数据中可以包含文本数据、语音数据、感知数据以及动作数据等。智能设备102中配置有接收多模态交互数据的相应设备,用来接收用户101发送的多模态交互数据。此时,展开数据传递的两方是用户101与智能设备102,数据传递的方向是从用户101传向智能设备102。

[0133] 接着,智能设备102向云端大脑104发送请求。请求云端大脑104对多模态交互数据进行语义理解、视觉识别、认知计算以及情感计算,以帮助用户进行决策。此时,通过视觉能力检测并提取多模态交互数据中的手部伸缩动作或手臂收放动作作为交互意图。然后,云端大脑104将回复数据传送至智能设备102。此时,展开通信的两方是智能设备102以及云端大脑104。

[0134] 最后,当智能设备102接收到云端大脑104传输的数据后,智能设备102会通过虚拟人按照交互意图进行多模态交互输出。此时,展开通信的两方为智能设备102与用户101。

[0135] 本发明提供的基于虚拟人的视觉交互方法及系统提供了一种虚拟人,虚拟人具备预设形象和预设属性,能够与用户进行多模态的交互。并且,本发明提供的虚拟人的视觉交互方法及系统还能够通过手部伸缩动作或手臂收放动作判断用户的意图,与用户展开交互,使得用户与虚拟人之间能够进行流畅的交流,并使得用户享受拟人的交互体验。

[0136] 应该理解的是,本发明所公开的实施例不限于这里所公开的特定结构、处理步骤或材料,而应当延伸到相关领域的普通技术人员所理解的这些特征的等同替代。还应当理解的是,在此使用的术语仅用于描述特定实施例的目的,而并不意味着限制。

[0137] 说明书中提到的“一个实施例”或“实施例”意指结合实施例描述的特定特征、结构或特性包括在本发明的至少一个实施例中。因此,说明书通篇各个地方出现的短语“一个实施例”或“实施例”并不一定均指同一个实施例。

[0138] 虽然本发明所公开的实施方式如上,但所述的内容只是为了便于理解本发明而采用的实施方式,并非用以限定本发明。任何本发明所属技术领域内的技术人员,在不脱离本发明所公开的精神和范围的前提下,可以在实施的形式上及细节上作任何的修改与变化,但本发明的专利保护范围,仍须以所附的权利要求书所界定的范围为准。

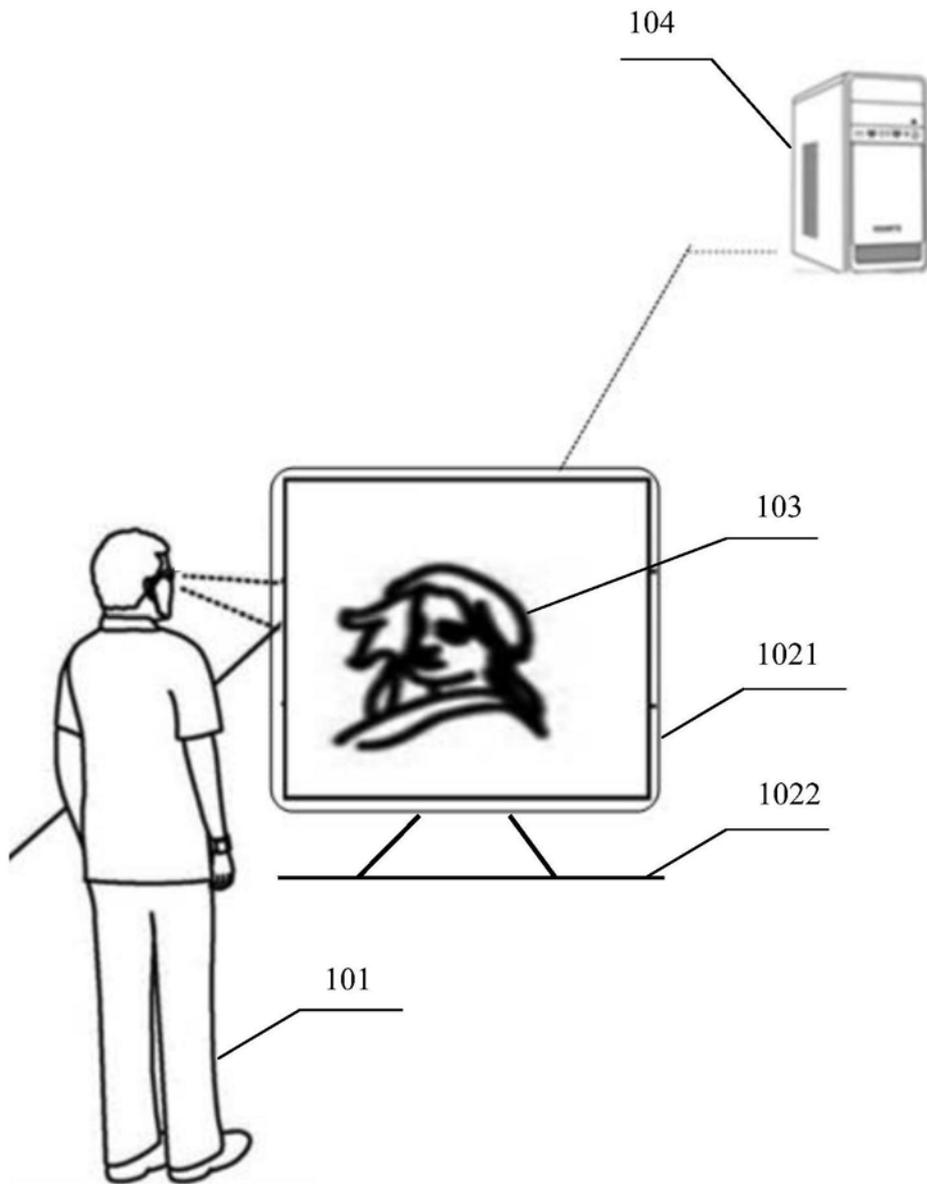


图1

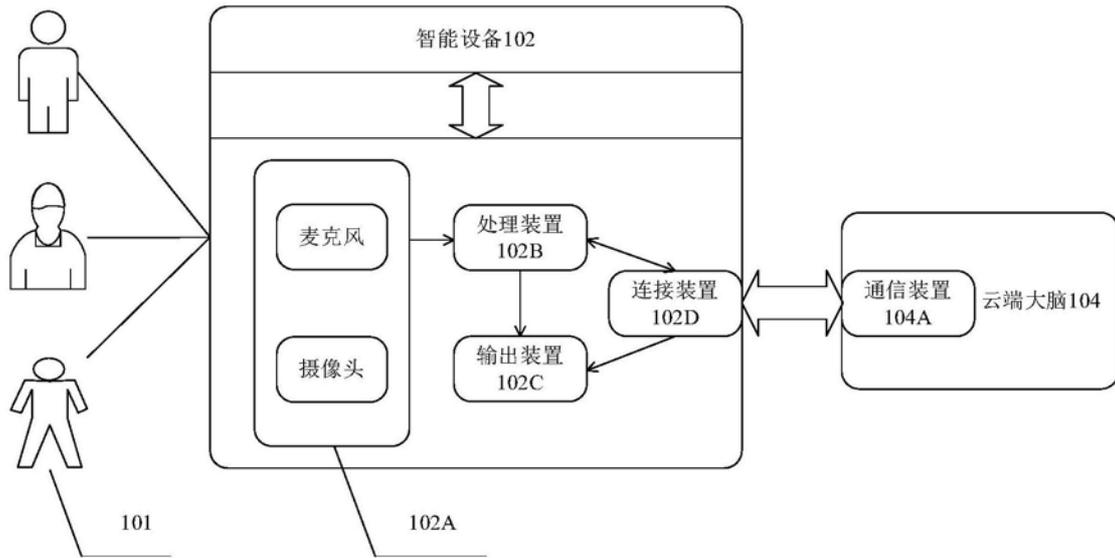


图2

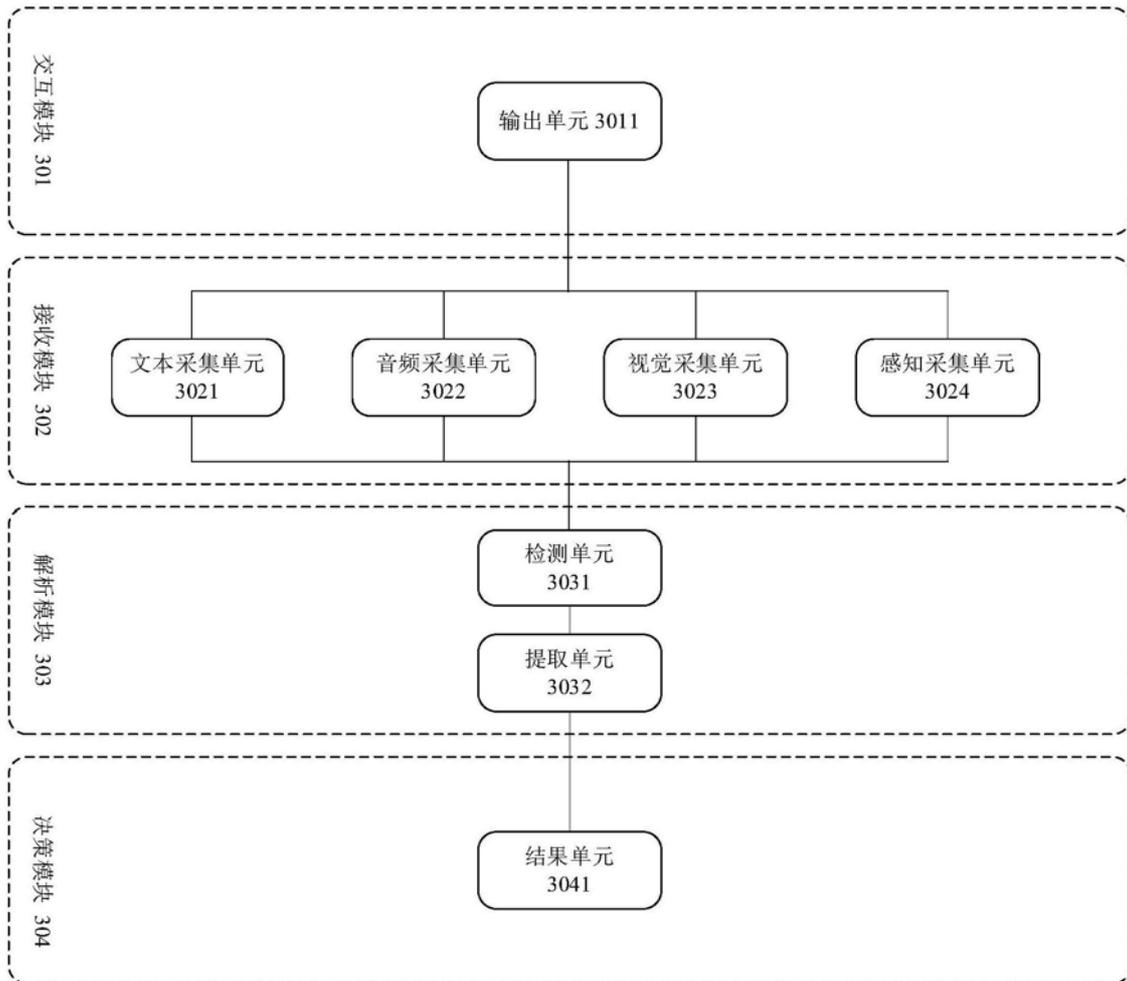


图3

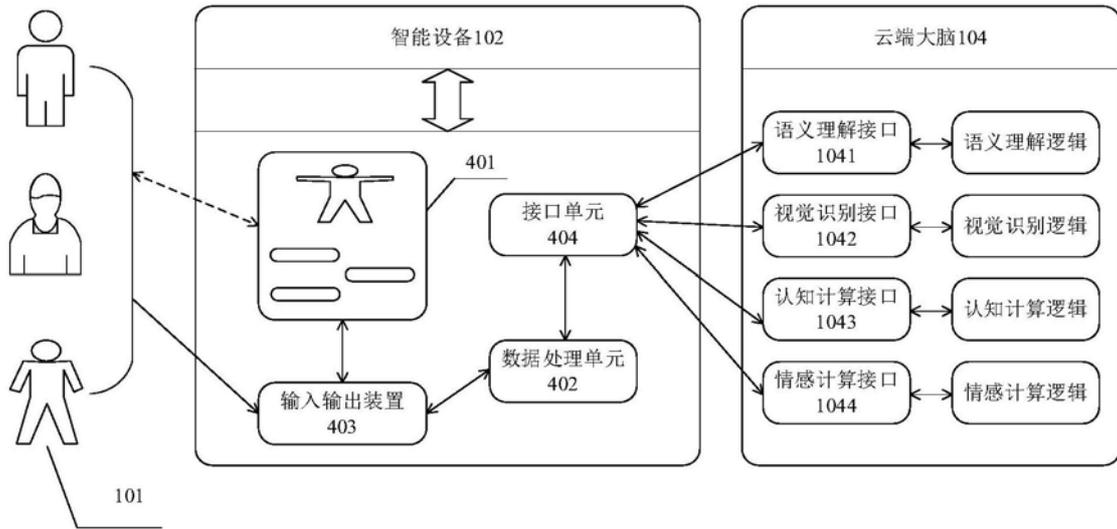


图4

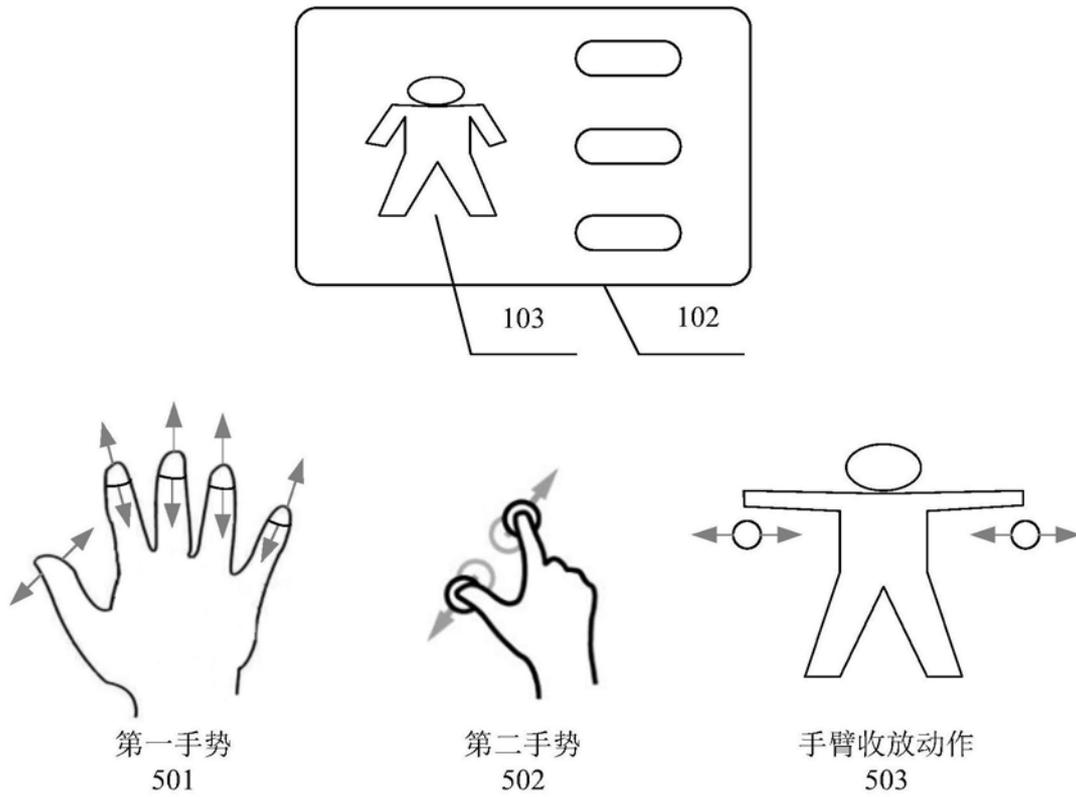


图5

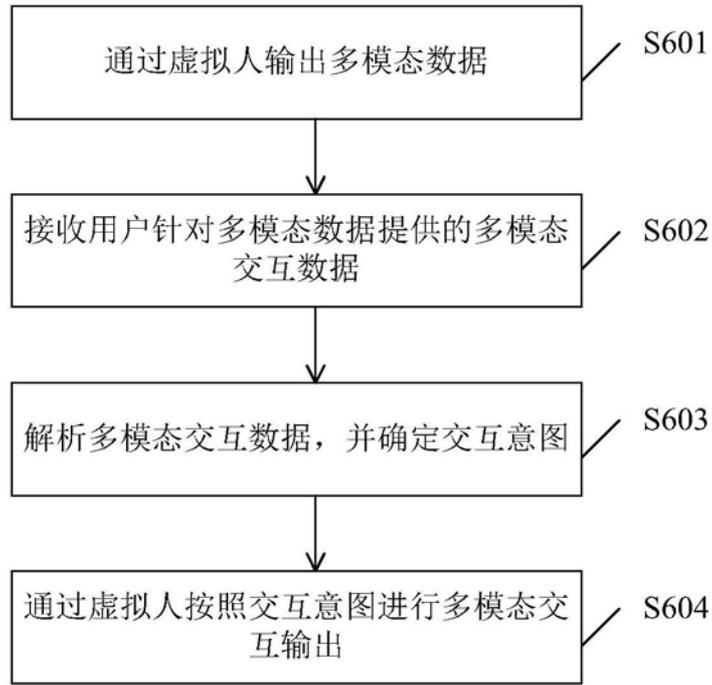


图6

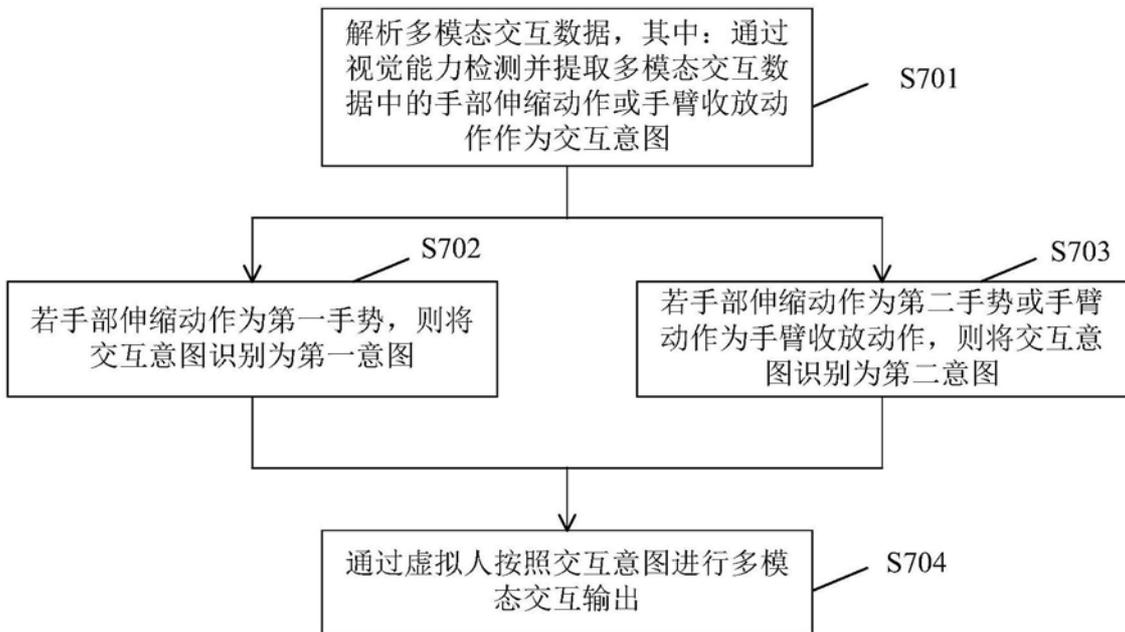


图7

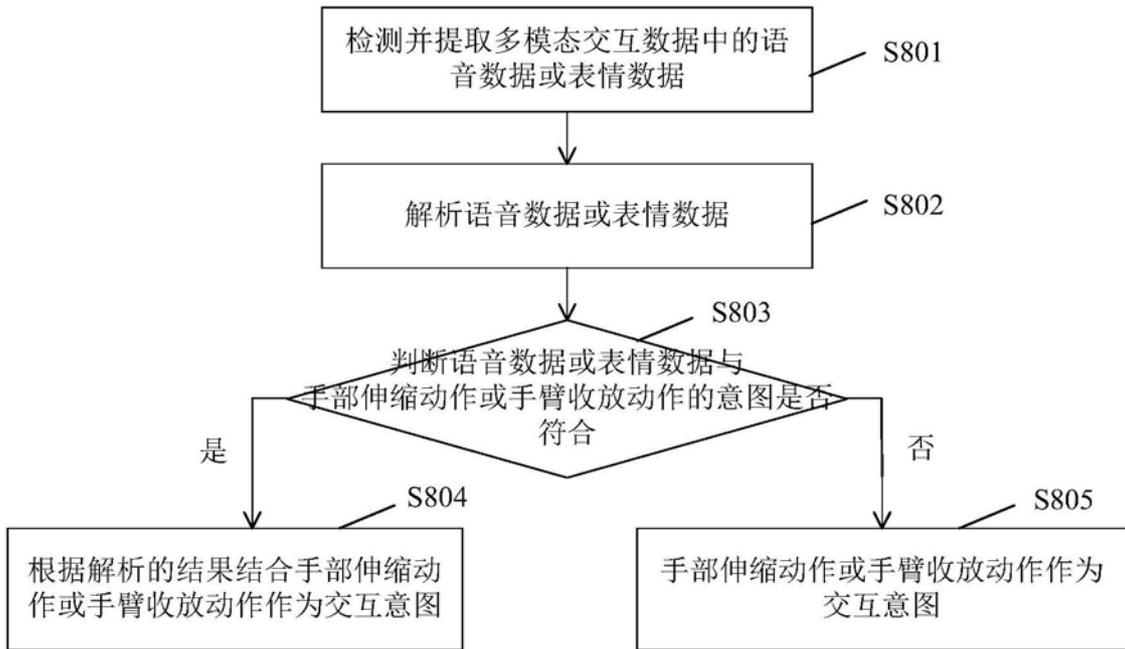


图8

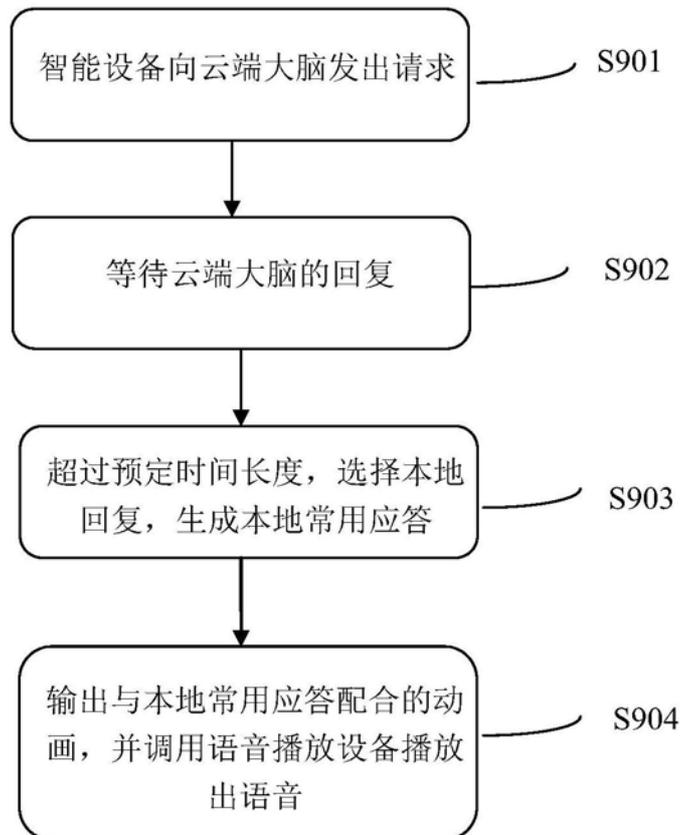


图9

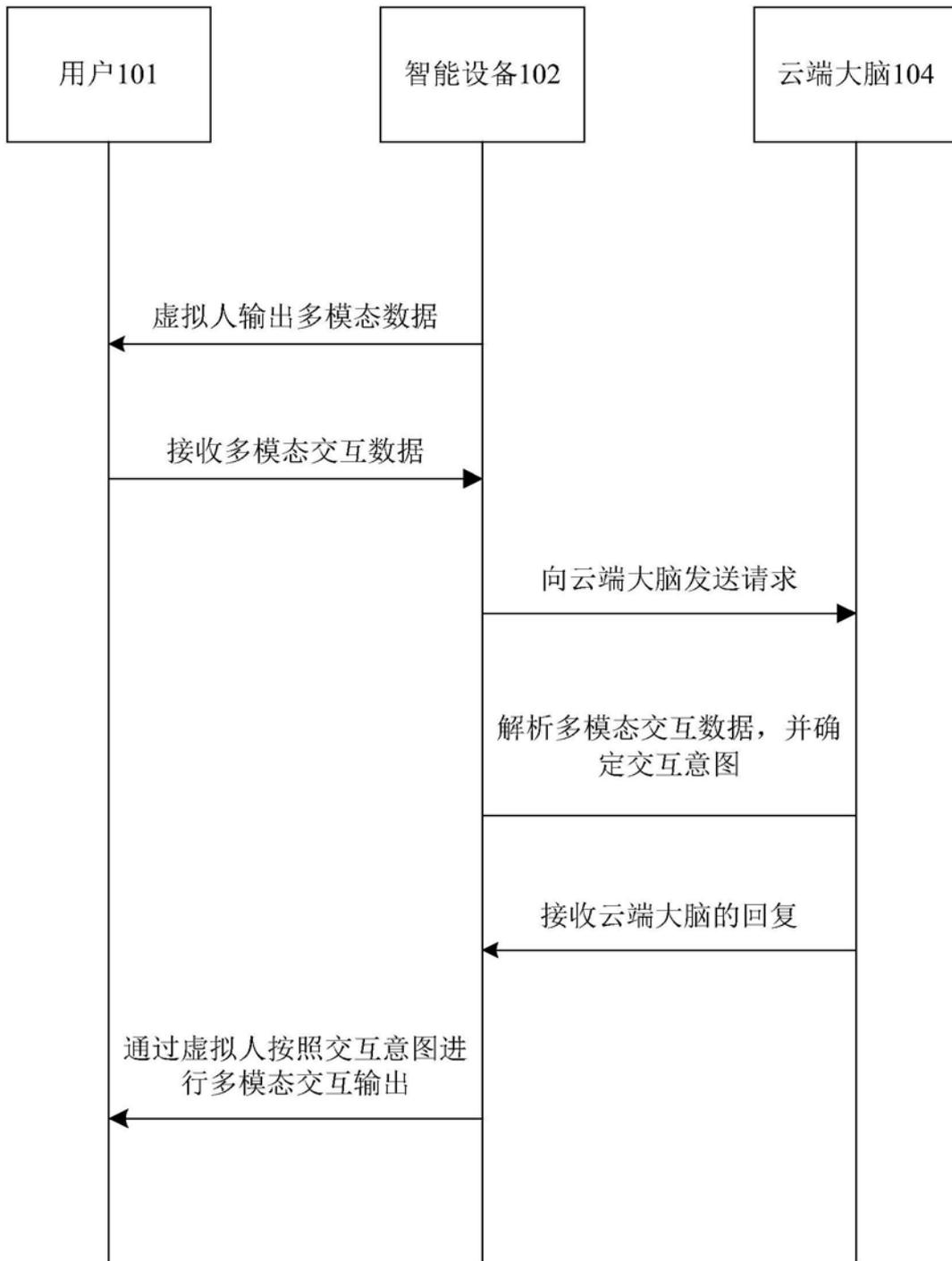


图10