(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112099628 A
(43)申请公布日 2020.12.18

(21)申请号 202010936390.9

(22)申请日 2020.09.08

(71)申请人 平安科技(深圳)有限公司
地址 518000 广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼

(72)发明人 邹芳 龙文甜

(74)专利代理机构 深圳市世联合知识产权代理有限公司 44385
代理人 汪琳琳

(51)Int.Cl.
*G06F 3/01*(2006.01)
*G06T 19/00*(2011.01)
*G06F 16/332*(2019.01)
*G10L 15/16*(2006.01)
*G10L 15/22*(2006.01)
*G06F 40/30*(2020.01)
*G06F 16/33*(2019.01)
*G06T 13/40*(2011.01)

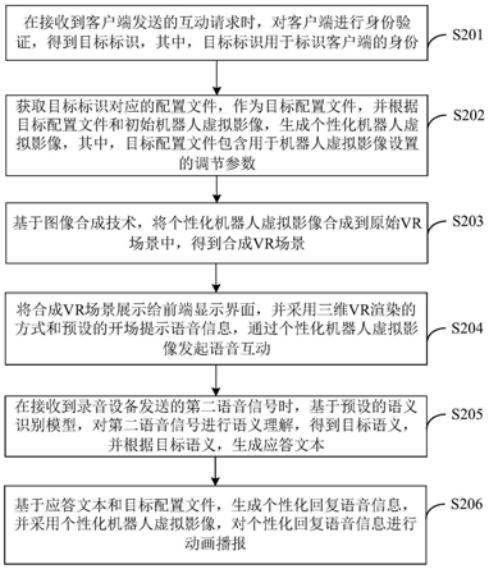
权利要求书2页 说明书13页 附图3页

(54)发明名称

基于人工智能的VR互动方法、装置、计算机设备及介质

(57)摘要

本发明涉及通信领域，公开了一种基于人工智能的VR互动方法、装置、计算机设备及介质，所述方法包括：在接收到互动请求时，对客户端进行身份验证，得到目标标识，获取目标标识对应的配置文件，作为目标配置文件，并根据目标配置文件生成个性化机器人虚拟影像，并合成到原始VR场景中，得到合成VR场景，将合成VR场景展示给前端显示界面，同时，采用个性化机器人虚拟影像发起语音互动，在接收到录音设备发送的第二语音信号时，对第二语音信号进行语义理解，并根据得到的目标语义，生成应答文本，基于应答文本和目标配置文件，生成个性化回复语音信息，并采用个性化机器人虚拟影像，对个性化回复语音信息进行动画播报，本发明可增强VR交互性和沉浸感。




CN 112099628 A

1.一种基于人工智能的VR互动方法,其特征在于,包括:

在接收到客户端发送的互动请求时,对客户端进行身份验证,得到目标标识,其中,所述目标标识用于标识客户端的身份;

获取所述目标标识对应的配置文件,作为目标配置文件,并根据所述目标配置文件和初始机器人虚拟影像,生成个性化机器人虚拟影像,其中,所述目标配置文件包含用于机器人虚拟影像设置的调节参数;

基于图像合成技术,将所述个性化机器人虚拟影像合成到原始VR场景中,得到合成VR场景;

将所述合成VR场景展示给前端显示界面,并采用三维VR渲染的方式和预设的开场提示语音信息,通过所述个性化机器人虚拟影像发起语音互动;

在接收到录音设备发送的第二语音信号时,基于预设的语义识别模型,对所述第二语音信号进行语义理解,得到目标语义,并根据所述目标语义,生成应答文本;

基于所述应答文本和所述目标配置文件,生成个性化回复语音信息,并采用所述个性化机器人虚拟影像,对所述个性化回复语音信息进行动画播报。

2.如权利要求1所述的基于人工智能的VR互动方法,其特征在于,所述互动请求为语音信号,所述在接收到客户端发送的互动请求时,对客户端进行身份验证,得到目标标识包括:

获取所述互动请求中包含的第一语音信息;

从所述第一语音信息中提取目标声纹特征;

采用动态声纹识别模型,从预设的声纹数据库中,获取目标声纹对应的用户标识,作为所述目标标识。

3.如权利要求2所述的基于人工智能的VR互动方法,其特征在于,所述从所述第一语音信息中提取目标声纹特征包括:对所述第一语音信息进行声纹解析,得到初始声纹样本;

对所述初始声纹样本进行预加重处理,生成具有平坦频谱的加重处理声纹样本;

采用分帧和加窗的方式,对所述加重处理声纹样本进行分帧处理,得到初始语音帧;

对所述初始语音帧信号进行静默音分离,得到目标语音帧;

基于所述目标语音帧,提取所述目标声纹特征。

4.如权利要求1所述的基于人工智能的VR互动方法,其特征在于,根据所述目标配置文件和初始机器人虚拟影像,生成个性化机器人虚拟影像包括:

获取所述初始机器人虚拟影像对应的初始配置文件;

将所述初始配置文件与所述目标配置文件进行对比分析,得到差异配置参数;

使用所述差异配置参数对所述初始机器人虚拟影像进行更新,得到所述个性化机器人虚拟影像。

5.如权利要求1所述的基于人工智能的VR互动方法,其特征在于,所述根据所述目标语义,生成应答文本包括:

将所述目标语义输入到训练好的相似问模型中,通过所述训练好的相似问模型,确定所述目标语义对应的目标相似问,其中,所述训练好的相似问模型为采用VR问答语料对Transform模型训练得到;

从问答语料数据库中,获取所述目标相似问对应的标准答,作为所述应答文本。

6.如权利要求1至5任一项所述的基于人工智能的VR互动方法，其特征在于，所述基于所述应答文本和所述目标配置文件，生成个性化回复语音信息包括：

根据所述目标配置文件与预设的偏好判断条件，确定所述目标标识对应的个性偏好；

选取与所述个性偏好对应的语音合成方式，对所述应答文本进行语音合成，得到个性化回复语音信息。

7.一种基于人工智能的VR互动装置，其特征在于，包括：

身份识别模块，用于在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，其中，所述目标标识用于标识客户端的身份；

影像生成模块，用于获取所述目标标识对应的配置文件，作为目标配置文件，并根据所述目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，其中，所述目标配置文件包含用于机器人虚拟影像设置的调节参数；

场景合成模块，用于基于图像合成技术，将所述个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景；

语音交互模块，用于将所述合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过所述个性化机器人虚拟影像发起语音互动；

文本确定模块，用于在接收到录音设备发送的第二语音信号时，基于预设的语义识别模型，对所述第二语音信号进行语义理解，得到目标语义，并根据所述目标语义，生成应答文本；

动画应答模块，用于基于所述应答文本和所述目标配置文件，生成个性化回复语音信息，并采用所述个性化机器人虚拟影像，对所述个性化回复语音信息进行动画播报。

8.如权利要求7所述的基于人工智能的VR互动装置，其特征在于，所述身份识别模块包括：

语音信息获取单元，用于获取所述互动请求中包含的第一语音信息；

声纹特征提取单元，用于从所述第一语音信息中提取目标声纹特征；

目标标识识别单元，用于采用动态声纹识别模型，从预设的声纹数据库中，获取目标声纹对应的用户标识，作为所述目标标识。

9.一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如权利要求1至6任一项所述的基于人工智能的VR互动方法。

10.一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1至6任一项所述的基于人工智能的VR互动方法。

# 基于人工智能的VR互动方法、装置、计算机设备及介质

## 技术领域

[0001] 本发明涉及电通信技术领域，尤其涉及一种基于人工智能的VR互动方法、装置、计算机设备及介质。

## 背景技术

[0002] 随着互联网技术的进步，人类的交流方式逐渐向虚拟现实时代迈进。虚拟现实(Virtual Reality，VR)技术能够为用户提供更加逼真的三维环境，使用户完全沉浸于VR之中。

[0003] 目前，我们现在体验到的很多VR在游戏、地产、影视、教育、医疗等行业的应用，市面上的VR设备主要分为两种，一种是作为游戏的控制游览设备的专业VR设备，这种VR设备需要通过线缆接入高配置的电脑并使用手柄进行相应操作。另一种是VR眼镜，需要在智能终端中安装应用程序来播放相应的VR片源进行观看，但内容的介绍都采用画外音的方式进行介绍，用户操作需要通过人体移动来操作按钮和VR场景进行交互，有内容表现效果较差、不够智能的问题，使得VR的互动性较差。

## 发明内容

[0004] 本发明实施例提供一种基于人工智能的VR互动方法、装置、计算机设备和存储介质，以提高VR交互的互动性。

[0005] 为了解决上述技术问题，本申请实施例提供一种基于人工智能的VR互动方法，包括：

[0006] 在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，其中，所述目标标识用于标识客户端的身份；

[0007] 获取所述目标标识对应的配置文件，作为目标配置文件，并根据所述目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，其中，所述目标配置文件包含用于机器人虚拟影像设置的调节参数；

[0008] 基于图像合成技术，将所述个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景；

[0009] 将所述合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过所述个性化机器人虚拟影像发起语音互动；

[0010] 在接收到录音设备发送的第二语音信号时，基于预设的语义识别模型，对所述第二语音信号进行语义理解，得到目标语义，并根据所述目标语义，生成应答文本；

[0011] 基于所述应答文本和所述目标配置文件，生成个性化回复语音信息，并采用所述个性化机器人虚拟影像，对所述个性化回复语音信息进行动画播报。

[0012] 可选地，所述互动请求为语音信号，所述在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识包括：

[0013] 获取所述互动请求中包含的第一语音信息；

[0014] 从所述第一语音信息中提取目标声纹特征；

[0015] 采用动态声纹识别模型，从预设的声纹数据库中，获取目标声纹对应的用户标识，作为所述目标标识。

[0016] 可选地，所述从所述第一语音信息中提取目标声纹特征包括：

[0017] 对所述第一语音信息进行声纹解析，得到初始声纹样本；

[0018] 对所述初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

[0019] 采用分帧和加窗的方式，对所述加重处理声纹样本进行分帧处理，得到初始语音帧；

[0020] 对所述初始语音帧信号进行静默音分离，得到目标语音帧；

[0021] 基于所述目标语音帧，提取所述目标声纹特征。

[0022] 可选地，根据所述目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像包括：

[0023] 获取所述初始机器人虚拟影像对应的初始配置文件；

[0024] 将所述初始配置文件与所述目标配置文件进行对比分析，得到差异配置参数；

[0025] 使用所述差异配置参数对所述初始机器人虚拟影像进行更新，得到所述个性化机器人虚拟影像。

[0026] 可选地，所述根据所述目标语义，生成应答文本包括：

[0027] 将所述目标语义输入到训练好的相似问模型中，通过所述训练好的相似问模型，确定所述目标语义对应的目标相似问，其中，所述训练好的相似问模型为采用VR问答语料对Transform模型训练得到；

[0028] 从问答语料数据库中，获取所述目标相似问对应的标准答，作为所述应答文本。

[0029] 可选地，所述基于所述应答文本和所述目标配置文件，生成个性化回复语音信息包括：

[0030] 根据所述目标配置文件与预设的偏好判断条件，确定所述目标标识对应的个性偏好；

[0031] 选取与所述个性偏好对应的语音合成方式，对所述应答文本进行语音合成，得到个性化回复语音信息。

[0032] 为了解决上述技术问题，本申请实施例还提供一种基于人工智能的VR互动装置，包括：

[0033] 身份识别模块，用于在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，其中，所述目标标识用于标识客户端的身份；

[0034] 影像生成模块，用于获取所述目标标识对应的配置文件，作为目标配置文件，并根据所述目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，其中，所述目标配置文件包含用于机器人虚拟影像设置的调节参数；

[0035] 场景合成模块，用于基于图像合成技术，将所述个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景；

[0036] 语音交互模块，用于将所述合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过所述个性化机器人虚拟影像发起语音互动；

[0037] 文本确定模块，用于在接收到录音设备发送的第二语音信号时，基于预设的语义

识别模型，对所述第二语音信号进行语义理解，得到目标语义，并根据所述目标语义，生成应答文本；

[0038] 动画应答模块，用于基于所述应答文本和所述目标配置文件，生成个性化回复语音信息，并采用所述个性化机器人虚拟影像，对所述个性化回复语音信息进行动画播报。

[0039] 可选地，所述身份识别模块包括：

[0040] 语音信息获取单元，用于获取所述互动请求中包含的第一语音信息；

[0041] 声纹特征提取单元，用于从所述第一语音信息中提取目标声纹特征；

[0042] 目标标识识别单元，用于采用动态声纹识别模型，从预设的声纹数据库中，获取目标声纹对应的用户标识，作为所述目标标识。

[0043] 可选地，语音信息获取单元包括：

[0044] 声纹解析子单元，用于对所述第一语音信息进行声纹解析，得到初始声纹样本；

[0045] 预加重子单元，用于对所述初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

[0046] 分帧加窗子单元，用于采用分帧和加窗的方式，对所述加重处理声纹样本进行分帧处理，得到初始语音帧；

[0047] 静默音分离子单元，用于对所述初始语音帧信号进行静默音分离，得到目标语音帧；

[0048] 目标声纹特征提取子单元，用于基于所述目标语音帧，提取所述目标声纹特征。

[0049] 可选地，所述影像生成模块包括：

[0050] 初始配置文件获取单元，用于获取所述初始机器人虚拟影像对应的初始配置文件；

[0051] 分析对比单元，用于将所述初始配置文件与所述目标配置文件进行对比分析，得到差异配置参数；

[0052] 影像更新单元，用于使用所述差异配置参数对所述初始机器人虚拟影像进行更新，得到所述个性化机器人虚拟影像。

[0053] 可选地，所述文本确定模块包括：

[0054] 目标相似问确定单元，用于将所述目标语义输入到训练好的相似问模型中，通过所述训练好的相似问模型，确定所述目标语义对应的目标相似问，其中，所述训练好的相似问模型为采用VR问答语料对Transform模型训练得到；

[0055] 应答文本确定单元，用于从问答语料数据库中，获取所述目标相似问对应的标准答，作为所述应答文本。

[0056] 可选地，所述动画应答模块包括：

[0057] 偏好判断单元，用于根据所述目标配置文件与预设的偏好判断条件，确定所述目标标识对应的个性偏好；

[0058] 个性回复合成单元，用于选取与所述个性偏好对应的语音合成方式，对所述应答文本进行语音合成，得到个性化回复语音信息。

[0059] 为了解决上述技术问题，本申请实施例还提供一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述基于人工智能的VR互动方法的步骤。

[0060] 为了解决上述技术问题，本申请实施例还提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现上述基于人工智能的VR互动方法的步骤。

[0061] 本发明实施例提供的基于人工智能的VR互动方法、装置、计算机设备及存储介质，一方面，通过在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，获取目标标识对应的配置文件，作为目标配置文件，并根据目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，基于图像合成技术，将个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景，将合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过个性化机器人虚拟影像发起语音互动，实现通过无感认证并生产与认证用户对应的个性化影像，有利于提高用户的沉浸感，另一方面，在接收到录音设备发送的第二语音信号时，基于预设的语义识别模型，对第二语音信号进行语义理解，得到目标语义，并根据目标语义，生成应答文本，基于应答文本和目标配置文件，生成个性化回复语音信息，并采用个性化机器人虚拟影像，对个性化回复语音信息进行动画播报，实现通过语音交互的方式与用户进行互动，增强了交互性，同时，采用个性化的动画播报方式，也有利于提高沉浸感。

## 附图说明

[0062] 为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例的描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0063] 图1是本申请可以应用于其中的示例性系统架构图；

[0064] 图2是本申请的基于人工智能的VR互动方法的一个实施例的流程图；

[0065] 图3是根据本申请的基于人工智能的VR互动装置的一个实施例的结构示意图；

[0066] 图4是根据本申请的计算机设备的一个实施例的结构示意图。

## 具体实施方式

[0067] 除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同；本文中在申请的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在于限制本申请；本申请的说明书和权利要求书及上述附图说明中的术语“包括”和“具有”以及它们的任何变形，意图在于覆盖不排他的包含。本申请的说明书和权利要求书或上述附图中的术语“第一”、“第二”等是用于区别不同对象，而不是用于描述特定顺序。

[0068] 在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

[0069] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发

明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0070] 请参阅图1，如图1所示，系统架构100可以包括终端设备101、102、103，网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

[0071] 用户可以使用终端设备101、102、103通过网络104与服务器105交互，以接收或发送消息等。

[0072] 终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、MP3播放器(Moving Picture E界面显示perts Group Audio Layer III，动态影像专家压缩标准音频层面3)、MP4(Moving Picture E界面显示perts Group Audio Layer IV，动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机和台式计算机等等。

[0073] 服务器105可以是提供各种服务的服务器，例如对终端设备101、102、103上显示的页面提供支持的后台服务器。

[0074] 需要说明的是，本申请实施例所提供的基于人工智能的VR互动方法由服务器执行，相应地，基于人工智能的VR互动装置设置于服务器中。

[0075] 应该理解，图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器，本申请实施例中的终端设备101、102、103具体可以对应的是实际生产中的应用系统。

[0076] 请参阅图2，图2示出本发明实施例提供的一种基于人工智能的VR互动方法，以该方法应用在图1中的服务端为例进行说明，详述如下：

[0077] S201：在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，其中，目标标识用于标识客户端的身份。

[0078] 具体地，在用户需要开启VR的某项体验时，通过客户端向服务端发送互动请求，服务端在接收到该互动请求后，通过互动请求中包含的信息，对客户端进行身份验证，得到标识客户端用户身份的目标标识。

[0079] 在本实施例中，客户端具体可以是VR设备端。

[0080] 其中，目标标识为用于唯一标识客户端的用户身份的标识符号，具体可以是数字、字母、汉字等中的一种或多种组合。

[0081] 优选的，为提高VR交互性，本实施例的互动请求以语音信号的形式进行发送，例如，用户通过说出预设的唤醒词“开启虚拟过山车”，来发送互动请求。服务端从语音信息中进行声纹提取，根据声纹来对客户端用户进行身份认证，得到目标标识，具体过程可参考后续实施例，为避免重复，此处不再赘述。

[0082] S202：获取目标标识对应的配置文件，作为目标配置文件，并根据目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，其中，目标配置文件包含用于机器人虚拟影像设置的调节参数。

[0083] 具体地，在用户第一次使用该VR设备，并录入声纹信息后，服务端会提供机器人虚拟影像的参数，来供用户自主配置，得到用户标识对应的配置文件，在本实施例中，用户自主配置的参数不同，使得机器人的虚拟影像会发生改变，基于此，用户可根据自己的偏好设

置各项用于机器人虚拟影像设置的调节参数，得到个性化的配置文件，在获取到目标标识，也即，确定当前使用VR设备的用户身份后，根据用以指示用户身份的目标标识，获取对应的配置文件，作为目标配置文件，进而根据该目标目标配置文件和VR设备提供的初始机器人虚拟影像，来生成符合用户偏好的个性化机器人虚拟影像。

[0084] 应理解，各项调节参数的类别和取值范围，可根据机器人虚拟影像的实际需求来进行设定，此处不做具体限定。

[0085] 例如，在一具体实时方式中，涉及的各项调节参数类型包括形象、人设、个性等，其中，形象包括头发、皮肤、眼睛、嘴巴、脸型、鼻子、耳朵、身材等多个参数，根据各种类型的参数，可以生成个性化的机器人影像，如宅男可以定制选择日系卡哇伊外形、人设、性格的智能机器人影像，宅女可以定制选择韩系小鲜肉外形、人设、性格的智能机器人影像。通过这些设计极大的提升了VR内容的表现力，还能满足不同用户千差万别的个性化诉求，很大程度的优化了用户体验。

[0086] 根据目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像的具体实现过程，可参考后续实施例的描述，为避免重复，此处不再赘述。

[0087] S203:基于图像合成技术，将个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景。

[0088] 具体地，通过图像合成技术，将个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景。

[0089] 其中，原始VR场景是指VR设备启用的初始场景画面或者预设场景画面。

[0090] 其中，图像合成技术具体可以是但不限于:基于深度学习的语义控制图像合成、基于视点合成的深度图编码和最小通信开销的Direct Send并行图像合成方法等。

[0091] 优选地，本实施例为提高合成速度，采用最小通信开销的Direct Send并行图像合成方法。

[0092] S204:将合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过个性化机器人虚拟影像发起语音互动。

[0093] 具体地，服务端在得到合成VR场景后，通过网络传输的方式将该合成VR场景显示到VR设备上，并启用个性化机器人虚拟影像，通过VR设备向用户发起语音互动。

[0094] 本实施例中，每个VR场景，均预先设置有开场提示语音信息，并存储该VR场景对应的非临时性可读计算机存储介质上，具体语音信息的内容，可根据VR场景的实际需要来设定，此处不做限定。

[0095] 其中，三维VR渲染的方式具体包括但不限于:基于图像的实时稠密三维重建、基于空间patch扩散的方法、基于网格的高精度实时三维重建(High-resolution Mesh-based Real-time 3D Reconstruction)和基于RGB-D实时三维重建等。

[0096] 优选的，本实施例采用基于空间patch扩散的方法，对个性化机器人虚拟影像在发起语音互过程中的三维动画进行渲染。优选的，本实施例选用5G网络进行数据传输。

[0097] S205:在接收到录音设备发送的第二语音信号时，基于预设的语义识别模型，对第二语音信号进行语义理解，得到目标语义，并根据目标语义，生成应答文本。

[0098] 具体地，在现有的交互方式中，用户通过手势进行VR交互，这对用户本身的操作有较高要求，同时，准确程度也不是很高，本实施例中，用户通过语音的方式，与VR设备进行交

互，用户在与VR交互中，通过录音设备，向服务端发送第二语音信号，服务端通过预设的语音识别模型，对第二语音信号中包含的语义进行理解，得到目标语义，再根据目标语义，生成应答文本。

[0099] 其中，预设的语义识别模型具体包括但不限于：Bert模型、NLP模型和Transform模型等，相应地，对第二语音信号进行识别和语义理解，可以通过NLP的方式，或者Bert模型的方式来实现，由于这些技术较为成熟，属于本领域的技术人员所知晓，此处不进行赘述。

[0100] 本实施例中，根据目标语义，生成应答文本，具体可以是采用规则库的方式，或者，智能问答机器人的方式，考虑到VR相关的问答与其他领域语料通用性不强，因而，本实施例采用预先设置VR交互相关的问答语料数据库，并采用训练好的相似问模型对目标语义进行识别，并根据识别结果，从问答语料数据库中，选取合适的应答文本，具体过程可参考后续实施例的描述，为避免重复，此处不再赘述。

[0101] S206：基于应答文本和目标配置文件，生成个性化回复语音信息，并采用个性化机器人虚拟影像，对个性化回复语音信息进行动画播报。

[0102] 具体地，服务端根据目标配置文件，预测目标标识个性偏好，根据该个性偏好和应答文本，采用TTS合成个性化回复语音信息。

[0103] 其中，TTS是Text To Speech的缩写，即“从文本到语音”，是人机对话的一部分，让机器能够说话。

[0104] 例如，预测到目标标识的个性偏好为动漫，则采用TTS根据应答文本合成偏向中二语调的个性化回复语音信息。

[0105] 基于应答文本和目标配置文件，生成个性化回复语音信息的具体实现方式，可参考后续实施例的描述，为避免重复，此处不再赘述。

[0106] 本实施例中，通过在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，获取目标标识对应的配置文件，作为目标配置文件，并根据目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，基于图像合成技术，将个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景，将合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过个性化机器人虚拟影像发起语音互动，实现通过无感认证并生产与认证用户对应的个性化影像，有利于提高用户的沉浸感，同时，在接收到录音设备发送的第二语音信号时，基于预设的语义识别模型，对第二语音信号进行语义理解，得到目标语义，并根据目标语义，生成应答文本，基于应答文本和目标配置文件，生成个性化回复语音信息，并采用个性化机器人虚拟影像，对个性化回复语音信息进行动画播报，实现通过语音交互的方式与用户进行互动，增强了交互性，同时，采用个性化的动画播报方式，也有利于提高沉浸感。

[0107] 在本实施例的一些可选的实现方式中，步骤S201中，互动请求为语音信号，在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识包括：

[0108] 获取互动请求中包含的第一语音信息；

[0109] 从第一语音信息中提取目标声纹特征；

[0110] 采用动态声纹识别模型，从预设的声纹数据库中，获取目标声纹对应的用户标识，作为目标标识。

[0111] 本实施例中，具体地，用户通过与VR设备语音交互的方式，向VR设备发送互动请

求，服务端接收该互动请求，并获取互动请求中包含的第一语音信息。

[0112] 其中，第一语音信息具体可以是用于开启或唤醒VR设备的语音指令，也可以是预设好的其他任意语音信号。

[0113] 需要说明的是，通过该第一语言信息激活VR设备，同时，在后续通过该第一语音信息进行解析，确定当前使用该VR的用户身份信息。

[0114] 进一步地，在服务端预设有声纹数据库，在用户使用VR设备时，通过第一语音信号判断用户是否为第一次使用该VR设备，若是，则从第一语音信号中提取声纹特征，并将该声纹特征和用户标识存储到声纹数据库中；若否，则从第一语音信号中提取目标声纹特征，并采用动态声纹识别模型，对该目标声纹特征与声纹数据库中存储的声纹特征进行识别，判断目标声纹对应的用户标识，得到目标标识，也即，确定当前使用该VR设备的用户。

[0115] 其中，声纹特征包括但不限于：声学特征、词法特征、韵律特征、语种方言口音信息和通道信息等。

[0116] 其中，动态声纹识别模型是通过对采集到预设的声纹数据库中每个声纹特征进行训练，进而得到每个用户标识对应的综合声纹特征，该综合声纹特征用于识别预设的声纹数据库中的唯一用户标识，并基于用户标识与其对应的综合声纹特征，构建动态声纹识别模型。

[0117] 其中，对每个用户标识对应的声纹特征进行训练，得到每个每个用户标识对应的综合声纹特征，所采用的训练方式包括但不限于：模板匹配方法、最近邻方法、神经网络方法、隐式马尔可夫模型(Hidden Markov Model，HMM)、矢量量化(Vector Quantization，VQ)方法、多项式分类器(Polynomial Classifiers)方法等。

[0118] 优选的，服务端与VR设备通过5G网络连接，结合5G网络高带宽特点，将目前在主机或者头显中进行的大规模复杂计算移至云端和网络边缘，从而提升VR内容渲染质量。

[0119] 在本实施例中，通过从互动请求中，提取出第一语音信息，进而通过对第一语音信息进行声纹识别，实现无感对客户端用户的识别认证，有利于提高身份认证的准确性和效率。

[0120] 在本实施例的一些可选的实现方式中，从第一语音信息中提取目标声纹特征包括：

[0121] 对第一语音信息进行声纹解析，得到初始声纹样本；

[0122] 对初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

[0123] 采用分帧和加窗的方式，对加重处理声纹样本进行分帧处理，得到初始语音帧；

[0124] 对初始语音帧信号进行静默音分离，得到目标语音帧；

[0125] 基于目标语音帧，提取目标声纹特征。

[0126] 具体地，在获取到第一语音信息后，对第一语音信息进行语音信号提取，得到初始声纹样本，在对初始声纹样本进行预加重处理、分帧加窗和静默音分离，得到目标语音帧，进而基于目标语音帧，提取目标声纹特征。

[0127] 进一步地，由于声门激励和口鼻辐射会对语音信号的平均功率谱产生影响，导致高频在超过800Hz时会按6dB/倍频跌落，所以在计算语音信号频谱时，频率越高相应的成分越小，为此要在预处理中进行预加重(Pre-emphasis)处理，预加重的目的是提高高频部分，使信号的频谱变得平坦，保持在低频到高频的整个频带中，能用同样的信噪比求频谱，以便

于频谱分析或者声道参数分析。预加重可在语音信号数字化时在反混叠滤波器之前进行，这样不仅可以进行预加重，而且可以压缩信号的动态范围，有效地提高信噪比。预加重可使用一阶的数字滤波器来实现，例如：有限脉冲响应(Finite Impulse Response，FIR)滤波器。

[0128] 值得说明的是，利用设备获取的语音信号都是模拟信号，在对这些模拟信号进行预加重处理之前，需要经过采样和量化将模拟信息转化为数字信号，根据语音的频谱范围200-3400Hz，采样率可设置为8KHz，量化精度为16bit。

[0129] 应理解，此处采样率和量化精度的数值范围，为本发明优选范围，但可以根据实际应用的需要进行设置，此处不做限制。

[0130] 语音信号在经过预加重后，频谱的高频部分得到提升，信号也变得平坦，生成具有平坦频谱的加重处理声纹样本，有利于后续的声纹特征提取。

[0131] 语音信号具有短时平稳的性质，语音信号在经过预加重处理后，需要对其进行分帧和加窗处理，来保持信号的短时平稳性，通常情况下，每秒钟包含的帧数在33～100帧之间。为了保持帧与帧之间的连续性，使得相邻两帧都能平滑过渡，采用交叠分帧的方式，如图3所示，图3示出了交叠分帧的样例，图3中第k帧和第k+1帧之间的交叠部分即为帧移。

[0132] 优选地，帧移与帧长的比值的取值范围为(0,0.5)。

[0133] 例如，在一具体实施方式中，预加重后的语音信号为s’(n)，帧长为N个采样点，帧移为M个采样点。当第1帧对应的采样点为第n个时，原始语音信号$x_1(n)$与各参数之间的对应关系为：

[0134] $x_1(n)=x[(1-1)M+n]$

[0135] 其中，n＝0,1,...,N-1,N＝256。

[0136] 进一步地，声纹样本经过分帧之后，使用相应的窗函数w(n)与预加重后的语音信号s’(n)相乘，即得到加窗后的语音信号$S_w$，将该语音信号作为初始语音帧信号。

[0137] 其中，窗函数包括但不限于：矩形窗(Rectangular)、汉明窗(Hamming)和汉宁窗(Hanning)等。

[0138] 矩形窗表达式为：

[0139] $$w(n)=\begin{cases}1 & (0\le n\le N-1)\\ 0 & (n<0,\ n>N)\end{cases}$$

[0140] 汉明窗表达式为：

[0141] $$w(n)=\begin{cases}0.54-0.46\cos(2*pi*n/(N-1)) & (0\le n\le N-1)\\ 0 & (n<0,\ n>N)\end{cases}$$

[0142] 汉宁窗表达式为：

[0143] $$w(n)=\begin{cases}0.5(1-\cos(2*pi*n/(N-1))) & (0\le n\le N-1)\\ 0 & (n<0,\ n>N)\end{cases}$$

[0144] 对经过预加重处理的声纹样本进行分帧和加窗处理，使得声纹样本保持帧与帧之间的连续性，并剔除掉一些异常的信号点，提高了声纹样本的鲁棒性。

[0145] 本实施例中，具体地，在采集到的语音信息中，语音信号可分为激活期和静默期两个状态，静默期不传送任何语音信号，上、下行链路的激活期和静默期相互独立。在进行声

纹特征提取的时候，需要检测出静默期状态，进而将静默期与激活期进行分离，以得到持续的激活期，将保留下来的持续的激活期的语音信号作为目标语音帧。

[0146] 其中，检测静默音状态的方式包括但不限于：语音端点检测、FFMPEG探测音频静音算法和语音活动检测(Voice Activity Detection，VAD)算法等。

[0147] 在经过预加重处理、分帧和加窗和静默音分离之后，获取了稳定性强的声纹样本，使用该样本进行声纹特征的提取。

[0148] 其中，声纹特征提取是提取并选择对说话人的声纹具有可分性强、稳定性高等特性的声学或语言特征。

[0149] 优选地，本发明选择提取的声纹特征为声学特征中的线性倒谱特征。

[0150] 在本实施例中，通过对语音信息进行预加重处理、分帧和加窗和静默音分离之后，再进行声纹提取，提高了声纹提取的准确性，有利于后续快速准确地确定目标标识。

[0151] 在本实施例的一些可选的实现方式中，步骤S202中，根据目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像包括：

[0152] 获取初始机器人虚拟影像对应的初始配置文件；

[0153] 将初始配置文件与目标配置文件进行对比分析，得到差异配置参数；

[0154] 使用差异配置参数对初始机器人虚拟影像进行更新，得到个性化机器人虚拟影像。

[0155] 具体地，获取服务端预先设置的初始机器人虚拟影像对应的初始配置文件，该初始配置文件中包含机器人虚拟影像的各个参数配置，并将该初始配置文件与目标配置文件进行对比，获取目标配置文件中与初始配置文件不同的参数，作为差异配置参数，并采用差异配置参数中的数值，对初始机器人虚拟影像进行调整更新，并进行图像边缘处理，得到个性化机器人虚拟影像。

[0156] 在本实施例中，通过用户预先设置的目标配置文件，生成个性化机器人虚拟影像，有利于提高用户的沉浸感。

[0157] 在本实施例的一些可选的实现方式中，步骤S205中，根据目标语义，生成应答文本包括：

[0158] 将目标语义输入到训练好的相似问模型中，通过训练好的相似问模型，确定目标语义对应的目标相似问，其中，训练好的相似问模型为采用VR问答语料对Transform模型训练得到；

[0159] 从问答语料数据库中，获取目标相似问对应的标准答，作为应答文本。

[0160] 其中，相似问是指在自然语言处理领域由于语言习惯等一些因素影响，使得提问问题与标准问虽然语义上相同，但是字面上有所区别的若干问题。在本实施例中，针对获取到的第二语音信号，进行识别得到目标语义，VR影像机器人在作答时，需要参考问答语料数据库中的回答，因而，需要先进行相似问的识别。

[0161] 在本实施例实施的过程中，还包括预先对问答语料数据库的构建，问答语料数据库的构建包括但不限于：相似问模型的训练和相似问的识别存储等。其中，相似问模型的训练采用VR问答语料对Transform模型训练来实现。

[0162] 在本实施例中，通过训练好的相似问模型对目标语义进行相似问的确定，进而得到对应的应答文本，提高应答的准确性和效率。

[0163] 在本实施例的一些可选的实现方式中，步骤S206中，基于应答文本和目标配置文件，生成个性化回复语音信息包括：

[0164] 根据目标配置文件与预设的偏好判断条件，确定目标标识对应的个性偏好；

[0165] 选取与个性偏好对应的语音合成方式，对应答文本进行语音合成，得到个性化回复语音信息。

[0166] 具体地，服务端预先设置有偏好判断条件，根据目标配置文件与预设的偏好判断条件，确定目标标识对应的个性偏好，再选择个性偏好对应的语音合成方式，对应答文本进行语音合成，得到个性化回复语音信息。

[0167] 例如，在一具体实施方式中，根据目标配置文件与预设的偏好判断条件，确定目标标识对应的个性偏好为中二爱好者，则采用中二爱好者对应的语调，对应答文本进行语音合成，得到偏向中二语调的回复语音信息。

[0168] 其中，预设的偏好判断条件，可以根据对不同参数的取值范围已经组合来进行个性化设置，此处不做具体设定。

[0169] 本实施例中，通过目标配置文件与预设的偏好判断条件，生成个性化回复语音信息，有利于提高用户体验和沉浸感。

[0170] 应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

[0171] 图3示出与上述实施例基于人工智能的VR互动方法一一对应的基于人工智能的VR互动装置的原理框图。如图3所示，该基于人工智能的VR互动装置包括身份识别模块31、影像生成模块32、场景合成模块33、语音交互模块34、文本确定模块35和动画应答模块36。各功能模块详细说明如下：

[0172] 身份识别模块31，用于在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，其中，目标标识用于标识客户端的身份；

[0173] 影像生成模块32，用于获取目标标识对应的配置文件，作为目标配置文件，并根据目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，其中，目标配置文件包含用于机器人虚拟影像设置的调节参数；

[0174] 场景合成模块33，用于基于图像合成技术，将个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景；

[0175] 语音交互模块34，用于将合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过个性化机器人虚拟影像发起语音互动；

[0176] 文本确定模块35，用于在接收到录音设备发送的第二语音信号时，基于预设的语义识别模型，对第二语音信号进行语义理解，得到目标语义，并根据目标语义，生成应答文本；

[0177] 动画应答模块36，用于基于应答文本和目标配置文件，生成个性化回复语音信息，并采用个性化机器人虚拟影像，对个性化回复语音信息进行动画播报。

[0178] 可选地，身份识别模块31包括：

[0179] 语音信息获取单元，用于获取互动请求中包含的第一语音信息；

[0180] 声纹特征提取单元，用于从第一语音信息中提取目标声纹特征；

[0181] 目标标识识别单元，用于采用动态声纹识别模型，从预设的声纹数据库中，获取目标声纹对应的用户标识，作为目标标识。

[0182] 可选地，语音信息获取单元包括：

[0183] 声纹解析子单元，用于对第一语音信息进行声纹解析，得到初始声纹样本；

[0184] 预加重子单元，用于对初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

[0185] 分帧加窗子单元，用于采用分帧和加窗的方式，对加重处理声纹样本进行分帧处理，得到初始语音帧；

[0186] 静默音分离子单元，用于对初始语音帧信号进行静默音分离，得到目标语音帧；

[0187] 目标声纹特征提取子单元，用于基于目标语音帧，提取目标声纹特征。

[0188] 可选地，影像生成模块32包括：

[0189] 初始配置文件获取单元，用于获取初始机器人虚拟影像对应的初始配置文件；

[0190] 分析对比单元，用于将初始配置文件与目标配置文件进行对比分析，得到差异配置参数；

[0191] 影像更新单元，用于使用差异配置参数对初始机器人虚拟影像进行更新，得到个性化机器人虚拟影像。

[0192] 可选地，文本确定模块35包括：

[0193] 目标相似问确定单元，用于将目标语义输入到训练好的相似问模型中，通过训练好的相似问模型，确定目标语义对应的目标相似问，其中，训练好的相似问模型为采用VR问答语料对Transform模型训练得到；

[0194] 应答文本确定单元，用于从问答语料数据库中，获取目标相似问对应的标准答，作为应答文本。

[0195] 可选地，动画应答模块36包括：

[0196] 偏好判断单元，用于根据目标配置文件与预设的偏好判断条件，确定目标标识对应的个性偏好；

[0197] 个性回复合成单元，用于选取与个性偏好对应的语音合成方式，对应答文本进行语音合成，得到个性化回复语音信息。

[0198] 关于基于人工智能的VR互动装置的具体限定可以参见上文中对于基于人工智能的VR互动方法的限定，在此不再赘述。上述基于人工智能的VR互动装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

[0199] 为解决上述技术问题，本申请实施例还提供计算机设备。具体请参阅图4，图4为本实施例计算机设备基本结构框图。

[0200] 所述计算机设备4包括通过系统总线相互通信连接存储器41、处理器42、网络接口43。需要指出的是，图中仅示出了具有组件连接存储器41、处理器42、网络接口43的计算机设备4，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。其中，本技术领域技术人员可以理解，这里的计算机设备是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专

用集成电路(Application Specific Integrated Circuit,ASIC)、可编程门阵列(Field-Programmable Gate Array,FPGA)、数字处理器(Digital Signal Processor,DSP)、嵌入式设备等。

[0201] 所述计算机设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机设备可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

[0202] 所述存储器41至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如,SD或D界面显示存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中,所述存储器41可以是所述计算机设备4的内部存储单元,例如该计算机设备4的硬盘或内存。在另一些实施例中,所述存储器41也可以是所述计算机设备4的外部存储设备,例如该计算机设备4上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。当然,所述存储器41还可以既包括所述计算机设备4的内部存储单元也包括其外部存储设备。本实施例中,所述存储器41通常用于存储安装于所述计算机设备4的操作系统和各类应用软件,例如电子文件的控制的程序代码等。此外,所述存储器41还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0203] 所述处理器42在一些实施例中可以是中央处理器(Central Processing Unit,CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器42通常用于控制所述计算机设备4的总体操作。本实施例中,所述处理器42用于运行所述存储器41中存储的程序代码或者处理数据,例如运行电子文件的控制的程序代码。

[0204] 所述网络接口43可包括无线网络接口或有线网络接口,该网络接口43通常用于在所述计算机设备4与其他电子设备之间建立通信连接。

[0205] 本申请还提供了另一种实施方式,即提供一种计算机可读存储介质,所述计算机可读存储介质存储有界面显示程序,所述界面显示程序可被至少一个处理器执行,以使所述至少一个处理器执行如上述的基于人工智能的VR互动方法的步骤。

[0206] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本申请各个实施例所述的方法。

[0207] 显然,以上所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例,附图中给出了本申请的较佳实施例,但并不限制本申请的专利范围。本申请可以以许多不同的形式来实现,相反地,提供这些实施例的目的是使对本申请的公开内容的理解更加透彻全面。尽管参照前述实施例对本申请进行了详细的说明,对于本领域的技术人员来而言,其依然可以对前述各具体实施方式所记载的技术方案进行修改,或者对其中部分技术特征进行等效替换。凡是利用本申请说明书及附图内容所做的等效结构,直接或间接运用在其他相关的技术领域,均同理在本申请专利保护范围之内。

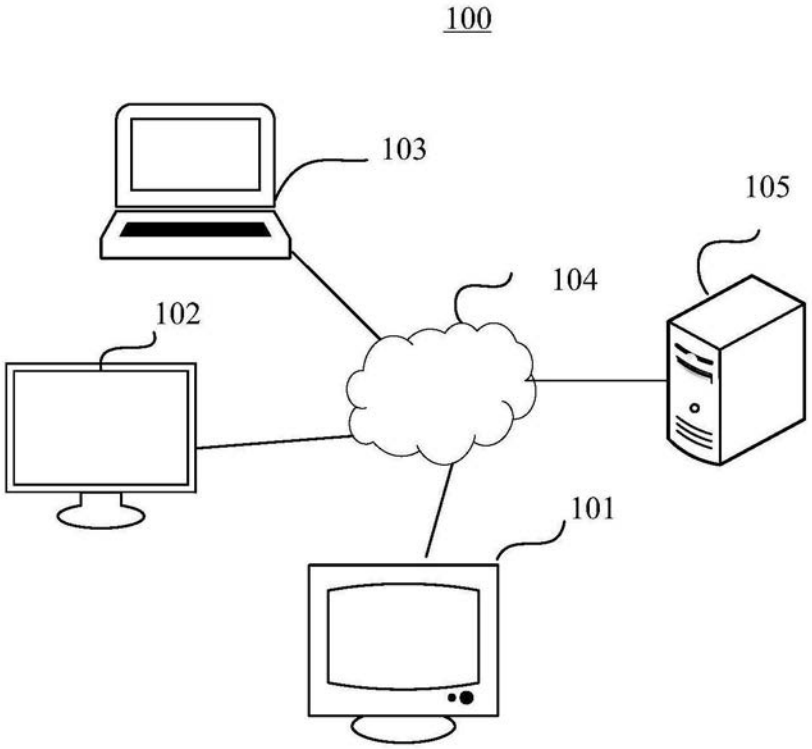



图1

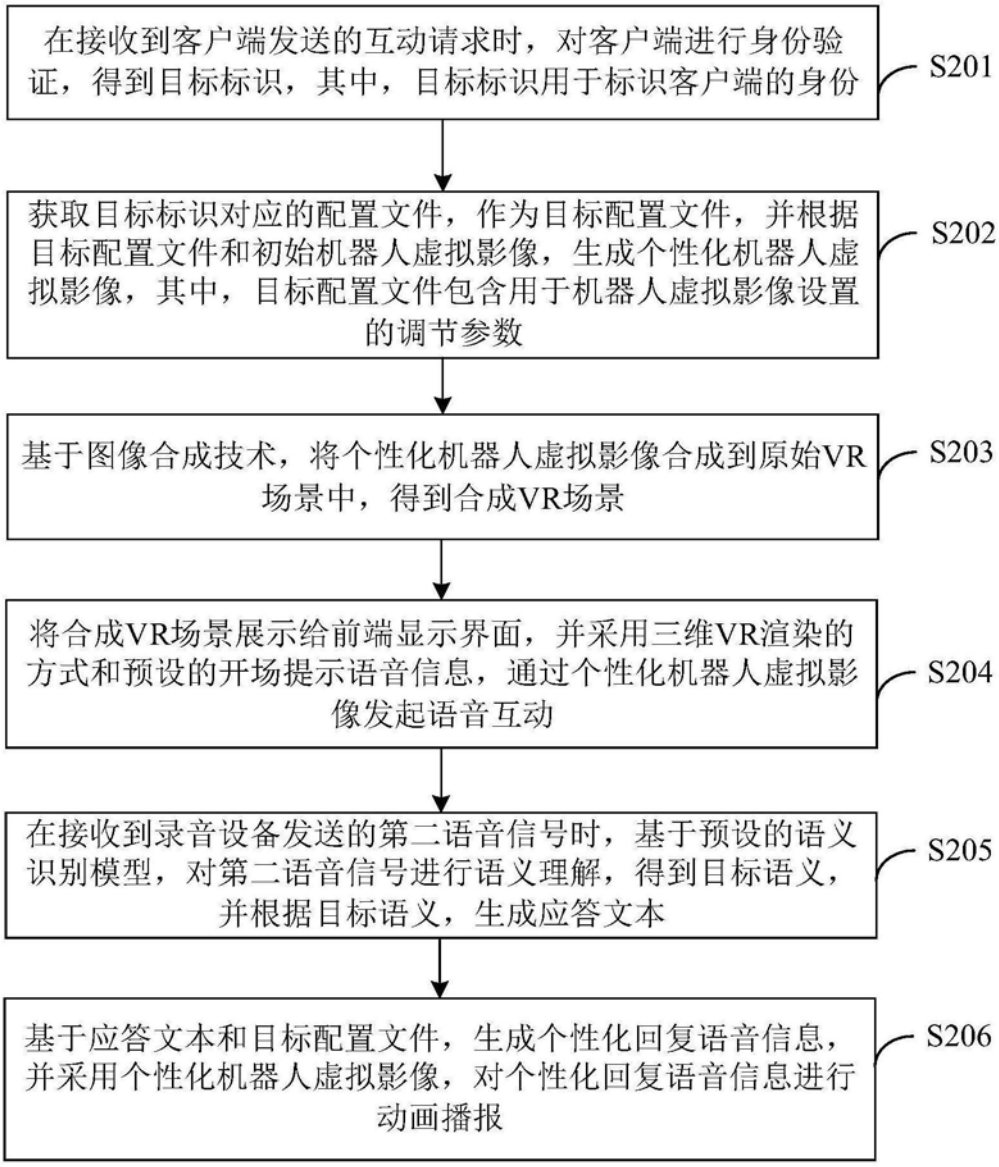

在接收到客户端发送的互动请求时，对客户端进行身份验证，得到目标标识，其中，目标标识用于标识客户端的身份
S201
获取目标标识对应的配置文件，作为目标配置文件，并根据目标配置文件和初始机器人虚拟影像，生成个性化机器人虚拟影像，其中，目标配置文件包含用于机器人虚拟影像设置的调节参数
S202
基于图像合成技术，将个性化机器人虚拟影像合成到原始VR场景中，得到合成VR场景
S203
将合成VR场景展示给前端显示界面，并采用三维VR渲染的方式和预设的开场提示语音信息，通过个性化机器人虚拟影像发起语音互动
S204
在接收到录音设备发送的第二语音信号时，基于预设的语义识别模型，对第二语音信号进行语义理解，得到目标语义，并根据目标语义，生成应答文本
S205
基于应答文本和目标配置文件，生成个性化回复语音信息，并采用个性化机器人虚拟影像，对个性化回复语音信息进行动画播报
S206

图2

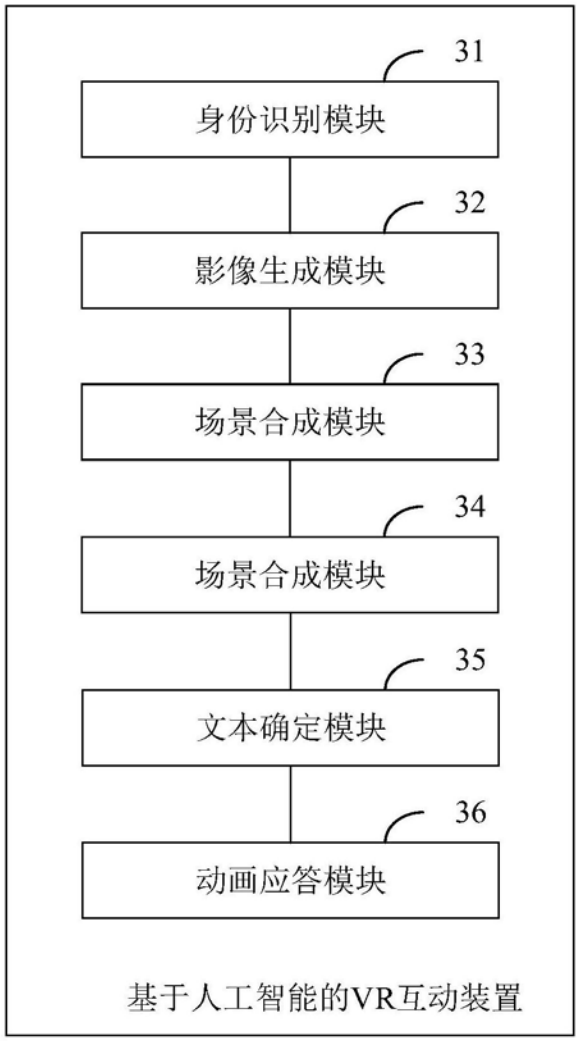

31
身份识别模块
32
影像生成模块
33
场景合成模块
34
场景合成模块
35
文本确定模块
36
动画应答模块
基于人工智能的VR互动装置

图3

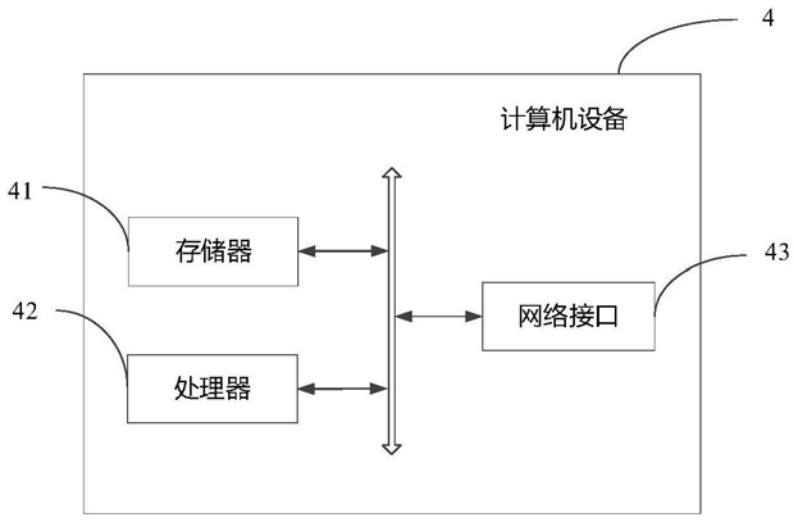

4
计算机设备
41
存储器
42
处理器
43
网络接口

图4