



(12) 发明专利

(10) 授权公告号 CN 112328999 B

(45) 授权公告日 2021.04.06

(21) 申请号 202110005259.5

(22) 申请日 2021.01.05

(65) 同一申请的已公布的文献号
申请公布号 CN 112328999 A

(43) 申请公布日 2021.02.05

(73) 专利权人 北京远鉴信息技术有限公司
地址 100000 北京市海淀区西四环北路158
号1幢7层80001-2

(72) 发明人 白世杰 吴富章

(74) 专利代理机构 北京超凡宏宇专利代理事务
所(特殊普通合伙) 11463
代理人 李莎

(51) Int. Cl.
G06F 21/32 (2013.01)
G06K 9/62 (2006.01)

(56) 对比文件

- CN 105426723 A, 2016.03.23
- CN 111709004 A, 2020.09.25
- CN 111523103 A, 2020.08.11
- CN 107294730 A, 2017.10.24
- CN 107808145 A, 2018.03.16
- CN 104376250 A, 2015.02.25
- US 2005238207 A1, 2005.10.27
- US 2010039223 A1, 2010.02.18

审查员 刘梦瑶

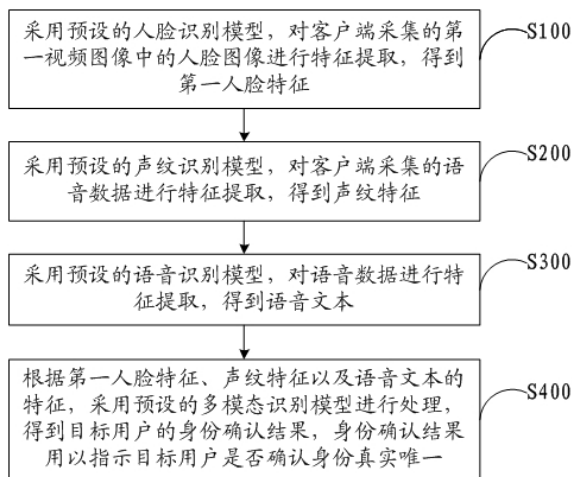
权利要求书3页 说明书17页 附图5页

(54) 发明名称

双录质检方法、装置、服务器及存储介质

(57) 摘要

本申请提供一种双录质检方法、装置、服务器及存储介质,涉及身份识别技术领域。该方法包括:采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人臉特征;采用预设的声纹识别模型,对客户端采集的语音数据进行特征提取,得到声纹特征;采用预设的语音识别模型,对语音数据进行特征提取,得到语音文本;根据第一人臉特征、声纹特征以及语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,身份确认结果用以指示目标用户是否确认身份真实唯一。本申请可有效保证用户在远程双录质检过程中身份认证的可靠性。



1. 一种双录质检方法,其特征在于,包括:

采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人脸特征;

采用预设的声纹识别模型,对所述客户端采集的语音数据进行特征提取,得到声纹特征;

采用预设的语音识别模型,对所述语音数据进行特征提取,得到语音文本;

根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,所述身份确认结果用以指示所述目标用户是否确认身份真实唯一;

其中,所述多模态识别模型为预先采用样本人脸特征、样本声纹特征,以及样本语音文本的特征进行训练得到的模型,所述样本人脸特征具有是否为真实人脸的标注信息,所述样本声纹特征具有是否为真实人声的标注信息,所述样本语音文本的特征具有是否为确认语音的标注信息;

所述方法还包括:

采用预设的面部动作模型,对所述第一视频图像进行动作检测,得到所述第一视频图像的动作检测结果;

若所述动作检测结果包括:识别到的预设面部动作的次数大于或等于预设次数阈值,且,所述身份确认结果通过,则确定所述目标用户为活体唯一身份。

2. 如权利要求1所述的方法,其特征在于,所述多模态识别模型包括:多模态融合层、自注意力层和全连接层,所述根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,包括:

采用所述多模态融合层,对所述第一人脸特征、所述声纹特征和所述语音文本的特征进行多模态特征融合,得到多模态融合特征;

采用所述自注意力层,对所述多模态融合特征进行特征学习,得到活体身份识别向量;

采用所述全连接层,对所述活体身份识别向量进行分类,得到所述目标用户的身份确认结果。

3. 如权利要求1所述的方法,其特征在于,所述根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果之前,所述方法还包括:

根据所述语音文本中各个文本段的起止时间段,从所述第一视频图像中截取所述起止时间段对应的唇语图像帧序列;

采用预设的唇语识别模型,判断所述唇语图像帧序列的动作是否与所述各个文本段对应的预设唇语动作匹配,得到所述目标用户的唇语匹配结果;

所述根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,包括:

若所述唇语匹配结果通过,则根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用所述多模态识别模型进行处理,得到所述身份确认结果。

4. 如权利要求3所述的方法,其特征在于,所述根据所述语音文本中各个文本段的起止时间段,从所述第一视频图像中截取所述起止时间段对应的唇语图像帧序列之前,所述方

法还包括：

从预设的声纹数据库中获取所述目标用户的备案声纹特征；

对所述声纹特征和所述备案声纹特征进行比对，得到所述目标用户的备案声纹比对结果；

所述根据所述语音文本中各个文本段的起止时间段，从所述第一视频图像中截取所述起止时间段对应的唇语图像帧序列，包括：

若所述备案声纹比对结果通过，则根据所述起止时间段，从所述第一视频图像中截取所述唇语图像帧序列。

5. 如权利要求4所述的方法，其特征在于，所述从预设的声纹数据库中获取所述目标用户的备案声纹特征之前，所述方法还包括：

从预设的人脸数据库中获取所述目标用户的备案人脸图像；

采用所述人脸识别模型，对所述备案人脸图像进行人脸特征提取，得到第二人脸特征；

对所述第一人脸特征和所述第二人脸特征进行比对，得到所述目标用户的备案人脸比对结果；

所述从预设的声纹数据库中获取所述目标用户的备案声纹特征，包括：

若所述备案人脸比对结果通过，则从所述声纹数据库中获取所述目标用户的备案声纹特征。

6. 如权利要求5所述的方法，其特征在于，所述从预设的人脸数据库中获取所述目标用户的备案人脸图像之前，所述方法还包括：

采用预设的身份证检测模型，对所述客户端采集的第二视频图像进行区域检测，得到身份证文本区域；

采用预设的身份字符识别模型，对所述身份证文本区域进行文字识别，得到第一身份文字信息；

对所述第一身份文字信息和预设的身份信息数据库中所述目标用户的第二身份文字信息进行比对，得到身份证文字比对结果；

所述从预设的人脸数据库中获取所述目标用户的备案人脸图像，包括：

若所述身份证文字比对结果通过，则从所述人脸数据库中获取所述备案人脸图像。

7. 如权利要求6所述的方法，其特征在于，所述采用预设的身份证检测模型，对所述客户端采集的第二视频图像进行区域检测，得到身份证文本区域，包括：

采用所述身份证检测模型，对所述第二视频图像进行区域检测，得到身份证人脸头像区域和所述身份证文本区域；

所述若所述身份证文字比对结果通过，则从所述人脸数据库中获取所述备案人脸图像，包括：

采用所述人脸识别模型，对所述身份证人脸头像区域进行人脸特征提取，得到第三人脸特征；

对所述第一人脸特征和所述第三人脸特征进行比对，得到所述目标用户的身份证人脸比对结果；

若所述身份证人脸比对结果和所述身份证文字比对结果均通过，则从所述人脸数据库中获取所述备案人脸图像。

8. 一种双录质检装置,其特征在于,所述装置包括:

人脸识别模块,用于采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人脸特征;

声纹识别模块,用于采用预设的声纹识别模型,对所述客户端采集的语音数据进行特征提取,得到声纹特征;

语音识别模块,用于采用预设的语音识别模型,对所述语音数据进行特征提取,得到语音文本;

身份确认模块,用于根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,所述身份确认结果用以指示所述目标用户是否确认身份真实唯一;

其中,所述多模态识别模型为预先采用样本人脸特征、样本声纹特征,以及样本语音文本的特征进行训练得到的模型,所述样本人脸特征具有是否为真实人脸的标注信息,所述样本声纹特征具有是否为真实人声的标注信息,所述样本语音文本的特征具有是否为确认语音的标注信息;

所述装置还包括:

面部动作检测模块,用于采用预设的面部动作模型,对所述第一视频图像进行动作检测,得到所述第一视频图像的动作检测结果;

活体确认模块,用于若所述动作检测结果包括:识别到的预设面部动作的次数大于或等于预设次数阈值,且,所述身份确认结果通过,则确定所述目标用户为活体唯一身份。

9. 一种服务器,其特征在于,包括:处理器、存储器,所述存储器中存储有所述处理器可执行的程序指令,当服务器运行时,所述处理器执行所述存储器中存储的所述程序指令,以执行如权利要求1至7任一所述的双录质检方法的步骤。

10. 一种存储介质,其特征在于,所述存储介质上存储有计算机程序,所述计算机程序被处理器运行时执行如权利要求1至7任一所述的双录质检方法的步骤。

双录质检方法、装置、服务器及存储介质

技术领域

[0001] 本发明涉及身份识别技术领域,具体而言,涉及一种双录质检方法、装置、服务器及存储介质。

背景技术

[0002] 在金融或证券行业,需要在交易时对用户确认交易过程录音录像,以确认用户的真实意愿。

[0003] 现有的针对远程线上双录技术,大多采用将语音数据和视频数据分别传输给服务器,服务器对语音数据和视频数据进行单独识别,但在识别过程中无法分别远程客户端的用户是否为活体,即存在可以利用照片和录音假冒用户身份的漏洞。

[0004] 现有的判断客户端的用户是否为活体,通常采用指示客户点头、摇头、张嘴及眨眼等动作,确认用户为活体,但该方案无法对语音数据的真实性进行判断,使得难以确认用户身份的真实性。

发明内容

[0005] 本发明的目的在于,针对上述现有技术中的不足,提供一种双录质检方法、装置、服务器及存储介质,以便对远程用户身份的真实性进行准确判断。

[0006] 为实现上述目的,本申请实施例采用的技术方案如下:

[0007] 第一方面,本申请实施例提供了一种双录质检方法,包括:

[0008] 采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人脸特征;

[0009] 采用预设的声纹识别模型,对所述客户端采集的语音数据进行特征提取,得到声纹特征;

[0010] 采用预设的语音识别模型,对所述语音数据进行特征提取,得到语音文本;

[0011] 根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,所述身份确认结果用以指示所述目标用户是否确认身份真实唯一;

[0012] 其中,所述多模态识别模型为预先采用样本人脸特征、样本声纹特征,以及样本语音文本的特征进行训练得到的模型,所述样本人脸特征具有是否为真实人脸的标注信息,所述样本声纹特征具有是否为真实人声的标注信息,所述样本语音文本的特征具有是否为确认语音的标注信息。

[0013] 可选的,所述根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果之前,所述方法还包括:

[0014] 根据所述语音文本中各个文本段的起止时间段,从所述第一视频图像中截取所述起止时间段对应的唇语图像帧序列;

[0015] 采用预设的唇语识别模型,判断所述唇语图像帧序列的动作是否与所述各个文本

段对应的预设唇语动作匹配,得到所述目标用户的唇语匹配结果;

[0016] 所述根据所述第一人臉特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,包括:

[0017] 若所述唇语匹配结果通过,则根据所述第一人臉特征、所述声纹特征以及所述语音文本的特征,采用所述多模态识别模型进行处理,得到所述身份确认结果。

[0018] 可选的,所述根据所述语音文本中各个文本段的起止时间段,从所述第一视频图像中截取所述起止时间段对应的唇语图像帧序列之前,所述方法还包括:

[0019] 从预设的声纹数据库中获取所述目标用户的备案声纹特征;

[0020] 对所述声纹特征和所述备案声纹特征进行比对,得到所述目标用户的备案声纹比对结果;

[0021] 所述根据所述语音文本中各个文本段的起止时间段,从所述第一视频图像中截取所述起止时间段对应的唇语图像帧序列,包括:

[0022] 若所述备案声纹比对结果通过,则根据所述起止时间段,从所述第一视频图像中截取所述唇语图像帧序列。

[0023] 可选的,所述从预设的声纹数据库中获取所述目标用户的备案声纹特征之前,所述方法还包括:

[0024] 从预设的人脸数据库中获取所述目标用户的备案人脸图像;

[0025] 采用所述人脸识别模型,对所述备案人脸图像进行人脸特征提取,得到第二人脸特征;

[0026] 对所述第一人臉特征和所述第二人脸特征进行比对,得到所述目标用户的备案人脸比对结果;

[0027] 所述从预设的声纹数据库中获取所述目标用户的备案声纹特征,包括:

[0028] 若所述备案人脸比对结果通过,则从所述声纹数据库中获取所述目标用户的备案声纹特征。

[0029] 可选的,所述从预设的人脸数据库中获取所述目标用户的备案人脸图像之前,所述方法还包括:

[0030] 采用预设的身份证检测模型,对所述客户端采集的第二视频图像进行区域检测,得到身份证文本区域;

[0031] 采用预设的身份字符识别模型,对所述身份证文本区域进行文字识别,得到第一身份文字信息;

[0032] 对所述第一身份文字信息和预设的身份信息数据库中所述目标用户的第二身份文字信息进行比对,得到身份证文字比对结果;

[0033] 所述从预设的人脸数据库中获取所述目标用户的备案人脸图像,包括:

[0034] 若所述身份证文字比对结果通过,则从所述人脸数据库中获取所述备案人脸图像。

[0035] 可选的,所述采用预设的身份证检测模型,对所述客户端采集的第二视频图像进行区域检测,得到身份证文本区域,包括:

[0036] 采用所述身份证检测模型,对所述第二视频图像进行区域检测,得到身份证人脸头像区域和所述身份证文本区域;

[0037] 所述若所述身份证文字比对结果通过,则从所述人脸数据库中获取所述备案人脸图像,包括:

[0038] 采用所述人脸识别模型,对所述身份证人脸头像区域进行人脸特征提取,得到第三人脸特征;

[0039] 对所述第一人脸特征和所述第三人脸特征进行比对,得到所述目标用户的身份证人脸比对结果;

[0040] 若所述身份证人脸比对结果和所述身份证文字比对结果均通过,则从所述人脸数据库中获取所述备案人脸图像。

[0041] 可选的,所述方法还包括:

[0042] 采用预设的面部动作模型,对所述第一视频图像进行动作检测,得到所述第一视频图像的动作检测结果;

[0043] 若所述动作检测结果包括:识别到的预设面部动作的次数大于或等于预设次数阈值,且,所述身份确认结果通过,则确定所述目标用户为活体唯一身份。

[0044] 第二方面,本申请实施例还提供一种双录质检装置,所述装置包括:

[0045] 第一人脸识别模块,用于采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人脸特征;

[0046] 声纹识别模块,用于采用预设的声纹识别模型,对所述客户端采集的语音数据进行特征提取,得到声纹特征;

[0047] 语音识别模块,用于采用预设的语音识别模型,对所述语音数据进行特征提取,得到语音文本;

[0048] 身份确认模块,用于根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,所述身份确认结果用以指示所述目标用户是否确认身份真实唯一;

[0049] 其中,所述多模态识别模型为预先采用样本人脸特征、样本声纹特征,以及样本语音文本的特征进行训练得到的模型,所述样本人脸特征具有是否为真实人脸的标注信息,所述样本声纹特征具有是否为真实人声的标注信息,所述样本语音文本的特征具有是否为确认语音的标注信息。

[0050] 可选的,在所述身份确认模块之前,所述装置还包括:

[0051] 唇语图像帧获取模块,用于根据所述语音文本中各个文本段的起止时间段,从所述第一视频图像中截取所述起止时间段对应的唇语图像帧序列;

[0052] 唇语识别模块,用于采用预设的唇语识别模型,判断所述唇语图像帧序列的动作是否与所述各个文本段对应的预设唇语动作匹配,得到所述目标用户的唇语匹配结果;

[0053] 所述身份确认模块,用于若所述唇语匹配结果通过,则根据所述第一人脸特征、所述声纹特征以及所述语音文本的特征,采用所述多模态识别模型进行处理,得到所述身份确认结果。

[0054] 可选的,在所述唇语图像帧获取模块之前,所述装置还包括:

[0055] 备案声纹特征获取模块,用于从预设的声纹数据库中获取所述目标用户的备案声纹特征;

[0056] 声纹比对模块,用于对所述声纹特征和所述备案声纹特征进行比对,得到所述目

标用户的备案声纹比对结果；

[0057] 所述唇语图像帧获取模块,用于若所述备案声纹比对结果通过,则根据所述起止时间段,从所述第一视频图像中截取所述唇语图像帧序列。

[0058] 可选的,在所述备案声纹特征获取模块之前,所述装置还包括:

[0059] 备案人脸图像获取模块,用于从预设的人脸数据库中获取所述目标用户的备案人脸图像;

[0060] 第二人脸识别模块,用于采用所述人脸识别模型,对所述备案人脸图像进行人脸特征提取,得到第二人脸特征;

[0061] 人脸比对模块,用于对所述第一人脸特征和所述第二人脸特征进行比对,得到所述目标用户的备案人脸比对结果;

[0062] 所述备案声纹特征获取模块,用于若所述备案人脸比对结果通过,则从所述声纹数据库中获取所述目标用户的备案声纹特征。

[0063] 可选的,在所述备案人脸图像获取模块之前,所述装置还包括:

[0064] 身份证检测模块,用于采用预设的身份证检测模型,对所述客户端采集的第二视频图像进行区域检测,得到身份证文本区域;

[0065] 字符识别模块,用于采用预设的身份字符识别模型,对所述身份证文本区域进行文字识别,得到第一身份文字信息;

[0066] 文字比对模块,用于对所述第一身份文字信息和预设的身份信息数据库中所述目标用户的第二身份文字信息进行比对,得到身份证文字比对结果;

[0067] 所述备案人脸图像获取模块,用于若所述身份证文字比对结果通过,则从所述人脸数据库中获取所述备案人脸图像。

[0068] 可选的,所述身份证检测模块包括:

[0069] 头像及文本检测单元,用于采用所述身份证检测模型,对所述第二视频图像进行区域检测,得到身份证人脸头像区域和所述身份证文本区域;

[0070] 所述备案人脸图像获取模块包括:

[0071] 人脸特征识别单元,用于采用所述人脸识别模型,对所述身份证人脸头像区域进行人脸特征提取,得到第三人脸特征;

[0072] 人脸特征比对单元,用于对所述第一人脸特征和所述第三人脸特征进行比对,得到所述目标用户的身份证人脸比对结果;

[0073] 备案人脸图像获取单元,用于若所述身份证人脸比对结果和所述身份证文字比对结果均通过,则从所述人脸数据库中获取所述备案人脸图像。

[0074] 可选的,所述装置还包括:

[0075] 面板动作检测模块,用于采用预设的面部动作模型,对所述第一视频图像进行动作检测,得到所述第一视频图像的动作检测结果;

[0076] 活体确认模块,用于若所述动作检测结果包括:识别到的预设面部动作的次数大于或等于预设次数阈值,且,所述身份确认结果通过,则确定所述目标用户为活体唯一身份。

[0077] 第三方面,本申请实施例还提供一种服务器,包括:处理器、存储器,所述存储器中存储有所述处理器可执行的程序指令,当服务器运行时,所述处理器执行所述存储器中存

储的所述程序指令,以执行上述任一所述的双录质检方法的步骤。

[0078] 第四方面,本申请实施例还提供一种存储介质,所述存储介质上存储有计算机程序,所述计算机程序被处理器运行时执行上述任一所述的双录质检方法的步骤。

[0079] 本申请的有益效果是:

[0080] 本申请实施例提供的双录质检方法、装置、服务器及存储介质,采用人脸识别模型提取第一视频图像中的第一人脸特征,采用声纹识别模型提取语音数据中的声纹特征,采用语音识别模型提取语音数据中的语音文本,根据第一人脸特征、声纹特征以及语音文本的特征,采用多模态识别模型得到目标用户的身份确认结果,指示目标用户是否确认身份真实唯一。通过本申请实施例提供的方法,可采用多模态识别模型,对第一人脸特征、声纹特征及语音文本的特征进行处理后,得到指示目标用户是否确认身份真实唯一的概率值,可实现跨媒体数据,即视频图像、语音数据和文本数据的有效结合,无需指示用户执行指定动作,在用户无感知的情况下全程监控用户的身份真实唯一,提高用户体验,提升业务办理流程流畅性;采用多种深度学习模型,对用户的视频图像、语音数据和文本数据进行多方位识别,充分利用多模态融合模型识别用户身份,有效保证用户在远程双录质检过程中身份认证的可靠性,防止他人伪造身份信息非法操作,为远程双录质检提供安全保障。

[0081] 另外,通过服务器实现自动质检,代替了人工质检,降低了质检成本,提高了质检效率和质检准确率,避免人工质检导致的误检、漏检。

附图说明

[0082] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本发明的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。

[0083] 图1为本申请实施例提供的第一种双录质检方法的流程示意图;

[0084] 图2为本申请实施例提供的第二种双录质检方法的流程示意图;

[0085] 图3为本申请实施例提供的第三种双录质检方法的流程示意图;

[0086] 图4为本申请实施例提供的第四种双录质检方法的流程示意图;

[0087] 图5为本申请实施例提供的第五种双录质检方法的流程示意图;

[0088] 图6为本申请实施例提供的第六种双录质检方法的流程示意图;

[0089] 图7为本申请实施例提供的第七种双录质检方法的流程示意图;

[0090] 图8为本申请实施例提供的第八种双录质检方法的流程示意图;

[0091] 图9为本申请实施例提供的一种双录质检装置的结构示意图;

[0092] 图10为本申请实施例提供的服务器的示意图。

具体实施方式

[0093] 为使本发明实施例的目的、技术方案和优点更加清楚,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。通常在此处附图中描述和示出的本发明实施例的组件可以以各种不同的配置来布置和设计。

[0094] 因此,以下对在附图中提供的本发明的实施例的详细描述并非旨在限制要求保护的本发明的范围,而是仅仅表示本发明的选定实施例。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。

[0095] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步定义和解释。

[0096] 在本发明的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0097] 本申请实施例提供的双录质检方法,可由具有双录质检功能的服务器上进行。该服务器可以为该客户端对应的应用服务器,该客户端可以为安装在手机,或,计算机设备等电子设备上的客户端应用程序。该应用服务器可以为客户端应用程序对应的服务端应用程序所在的设备,其可以为本地服务器,也可以为云服务器。

[0098] 例如,该双录质检方法由客户端设备执行,则该服务器上预先部署有多种多种深度学习模型,如人脸识别模型、声纹识别模型、语义识别模型、多模态识别模型、唇语识别模型、身份证检测模型、身份字符识别模型、面部动作模型、人脸检测模型。其中,人脸识别模型可用于提取人脸图像中的人脸特征,声纹识别模型可用于提取语音数据中的声纹特征,语音识别模型可用于提取语音数据中的语音文本,多模态识别模型可用于确认用户身份,唇语识别模型可用于识别唇语动作是否与文本对应的预设唇语动作匹配,身份证检测模块可用于提取身份证文本区域和身份证人脸头像区域,身份字符识别模型可用于提取身份证文本区域的文字信息、面部动作模型可用于识别面部动作,人脸检测模型可用于检测视频图像中的人脸图像。

[0099] 图1为本申请实施例提供的第一种双录质检方法的流程示意图;如图1所示,该方法包括:

[0100] S100:采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人脸特征。

[0101] 具体地,预设的人脸识别模型为预先采用样本人脸图像进行训练的模型,其可用于提取人脸图像中的人脸特征。在实际的应用中,可由服务器向客户端发送双录请求指令,客户端根据双录请求指令控制客户端所在设备的摄像头采集视频图像,控制客户端所在设备的语音输入设备如麦克风采集语音数据。其中,摄像头采集的视频图像包括:摄像头针对用户面部采集的视频图像,如该第一视频图像。摄像头在采集到该第一视频图像之后,可由客户端将采集的第一视频图像传输给服务器,由服务器通过人脸识别模型提取第一视频图像中的人脸图像的特征,得到第一人脸特征。

[0102] 在一种可能的示例中,在S100之前,该方法还包括:

[0103] 采用预设的人脸检测模块,对第一视频图像进行人脸检测,并从第一视频图像中扣取人脸图像。

[0104] 具体的,预设的人脸检测模块为预先采用第一样本视频图像进行训练的模型,其可用于提取视频图像中的人脸图像。由于第一视频图像中,除了用户人脸之外还可能存在着其它图像,如背景图像等,为避免背景图像对人脸特征提取造成干扰,可在人脸特征提取之前,采用人脸检测模块对第一视频图像进行人脸检测,并从第一视频图像中扣取具有检测

到的人脸的人脸图像,将人脸图像传输给人脸识别模型,使得人脸识别模型基于该人脸图像进行第一人臉特征的提取。

[0105] S200:采用预设的声纹识别模型,对客户端采集的语音数据进行特征提取,得到声纹特征。

[0106] 具体地,预设的声纹识别模型为预先采用样本语音数据进行训练得到的模型,其可用于提取语音数据中的声纹特征。

[0107] 语音输入设备采集到的语音数据可包括:语音输入设备针对用户语音采集的语音数据。语音输入设备在采集到该语音数据之后,便可将其传输至服务器,由服务器采用声纹识别模型对语音数据进行声纹特征提取,得到声纹特征。

[0108] S300:采用预设的语音识别模型,对语音数据进行特征提取,得到语音文本。

[0109] 具体地,预设的语音识别模型为预先采用语音数据进行训练得到的模型,可用于提取语音数据中的语音文本,服务器通过语音识别模型提取语音数据中的语义特征,得到语音文本。该语音文本可以为表征该语音数据中语义的文本。

[0110] 示例的,语音数据中包含指示该客户端对应用户的身份确认语音如:本人XXX,确认身份信息真实有效。

[0111] S400:根据第一人臉特征、声纹特征以及语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,身份确认结果用以指示目标用户是否确认身份真实唯一。

[0112] 其中,多模态识别模型为预先采用样本人脸特征、样本声纹特征,以及样本语音文本的特征进行训练得到的模型,样本人脸特征具有是否为真实人脸的标注信息,样本声纹特征具有是否为真实人声的标注信息,样本语音文本的特征具有是否为确认语音的标注信息。

[0113] 具体的,多模态识别模型包括多模态融合层、自注意力层和全连接层。其中,多模态融合层可采用预设的融合方法将提取到的第一人臉特征 F 、声纹特征 V 和语音文本的特征 T 进行多模态特征融合,得到多模态融合特征。示例的,预设的融合方法可以为张量积融合方法,融合后的多模态融合特征 M 为: $M=F \otimes V \otimes T$ 。符号 \otimes 表示张量积。

[0114] 其中,自注意力层可采用预设的学习方法,学习多模态融合特征 M 中的第一人臉特征 F 、声纹特征 V 和语音文本的特征 T ,并以活体身份识别向量 S 表示学习结果。

[0115] 例如可对多模态融合特征 M 分别乘以 A 、 B 、 C 三个预设矩阵,得到向量 Q 、 K 、 V ,对 Q 和 K 进行向量内积的方式计算相似度,以得到权重向量 P ,对 P 和 V 进行哈达玛(Hadamard)乘积计算,得到活体身份识别向量 S 。其中,矩阵 A 、 B 、 C 是预先经过训练得到的参数矩阵。

[0116] 其中,全连接层可对自注意力层输出的活体身份识别向量 S 进行分类,输出分类识别结果,该分类识别结果可用于表示该客户端对应用户即目标用户的身份确认结果。示例的,该分类识别结果例如可以为一维向量,如一维概率值。若该概率值大于或等于预设概率值,则确定身份确认结果为:目标用户的身份真实唯一;若该概率值小于预设概率值,则可确定身份确认结果为目标用户的身份不真实唯一。

[0117] 采用多模态识别模型,对第一人臉特征、声纹特征和语音文本的特征进行多模态融合、自注意力层学习和全连接层分类,得到分类识别结果,表示用户的身份确认结果。相对于分别利用人脸特征识别模型识别人脸真实性、利用声纹特征识别模型识别声纹真实

性、利用语义识别模型识别语音文本的语义,并将人脸真实性、声纹真实性和语义结合判断目标用户的身份确认结果,多模态识别模型对特征的多模态融合、学习及识别,以得到用户的身份确认结果,通过各个模态之间的相互辅助和补充,使得身份确认结果的准确率更高,保证双录质检的安全性和可靠性。

[0118] 本申请实施例提供的双录质检方法,采用人脸识别模型提取第一视频图像中的第一人臉特征,采用声纹识别模型提取语音数据中的声纹特征,采用语音识别模型提取语音数据中的语音文本,根据第一人臉特征、声纹特征以及语音文本的特征,采用多模态识别模型得到目标用户的身份确认结果,指示目标用户是否确认身份真实唯一。通过本申请实施例提供的方法,可采用多模态识别模型,对第一人臉特征、声纹特征及语音文本的特征进行处理后,得到指示目标用户是否确认身份真实唯一的概率值,可实现跨媒体数据,即视频图像、语音数据和文本数据的有效结合,无需指示用户执行指定动作,在用户无感知的情况下全程监控用户的身份真实唯一,提高用户体验,提升业务办理流程流畅性;采用多种深度学习模型,对用户的视频图像、语音数据和文本数据进行多方位识别,充分利用多模态融合模型识别用户身份,有效保证用户在远程双录质检过程中身份认证的可靠性,防止他人伪造身份信息进行非法操作,为远程双录质检提供安全保障。

[0119] 另外,通过服务器实现自动质检,代替了人工质检,降低了质检成本,提高了质检效率和质检准确率,避免人工质检导致的误检、漏检。

[0120] 在上述实施例的基础上,本申请实施例还提供一种双录质检方法,图2为本申请实施例提供的第二种双录质检方法的流程示意图,如图2所示,在上述S400之前,该方法还包括:

[0121] S51:根据语音文本中各个文本段的起止时间段,从第一视频图像中截取起止时间段对应的唇语图像帧序列。

[0122] 具体的,为避免用户输入的语音数据为预先录音的数据,需要将第一视频图像中的唇语动作和语音文本对应的预设唇语动作进行比对。将上述S300得到的语音文本按照时间段划分为多个文本段,根据文本段的起始时间和终止时间,从第一视频图像中截取相同时间段对应的唇语图像帧序列。

[0123] S52:采用预设的唇语识别模型,判断唇语图像帧序列的动作是否与各个文本段对应的预设唇语动作匹配,得到目标用户的唇语匹配结果。

[0124] 具体的,预设的唇语识别模型为预先采用样本唇语动作和样本文本对应的预设唇语动作进行训练得到的模型,其可用于对唇语动作和预设唇语动作进行比对,判断唇语动作和预设唇语动作是否匹配。在实际的应用中,服务器中存有语音文本对应的预设唇语动作,将每个文本段和对应起止时间段的唇语图像帧序列输入至唇语识别模型,判断每个文本段的预设唇语动作与唇语图像帧序列中的唇语动作是否匹配,得到的唇语匹配结果用于指示目标用户是否为活体。若每个文本段的预设唇语动作与唇语图像帧序列中的唇语动作不匹配,则唇语匹配结果指示目标用户不是活体,双录质检失败,服务器在客户端指示目标用户唇语质检失败,用户认证无效,需重新认证。

[0125] 上述S400包括:

[0126] S400a:若唇语匹配结果通过,则根据第一人臉特征、声纹特征以及语音文本的特征,采用多模态识别模型进行处理,得到身份确认结果。

[0127] 具体的,若每个文本段的预设唇语动作与唇语图像帧序列中的唇语动作匹配,则唇语匹配结果指示目标用户是活体,则可将第一人脸特征、声纹特征以及语音文本的特征输入至多模态识别模型,对目标用户身份进行二次确认,得到身份确认结果。

[0128] 本申请实施例提供的双录质检方法,在采用多模态识别模型得到目标用户的身份确认结果之前,根据语音文本中各个文本段的起止时间段,从第一视频图像中截取起止时间段对应的唇语图像帧序列,采用唇语识别模型,判断唇语图像帧序列的动作是否与各个文本段对应的预设唇语动作匹配,得到目标用户的唇语匹配结果。通过本方法进行唇语匹配检测,判断目标用户读取指定文本的唇语动作是否与文本的预设唇语动作是否相同,在目标用户的唇语动作与文本的预设唇语动作相同时,确认目标用户为活体,提高双录质检的效果,避免他人以照片和录音冒充真实用户,为远程双录质检提供安全保障,确保双录质检的可靠性。

[0129] 在上述任一实施例的基础上,本申请实施例还提供一种双录质检方法,图3为本申请实施例提供的第三种双录质检方法的流程示意图,如图3所示,在上述S51之前,该方法还包括:

[0130] S41:从预设的声纹数据库中获取目标用户的备案声纹特征。

[0131] 具体的,在进行唇语识别之前,还可通过判断目标用户的声纹特征确定用户身份是否真实。预设的声纹数据库中预先存有多个用户真实的备案声纹特征,每个用户具有唯一的用户标识和备案声纹特征,可根据每个用户唯一的用户标识从预设的声纹数据库中调用该用户的备案声纹特征。示例的,预设的声纹数据库可以为用户在公安部门备案的声纹特征底库,用户标识为身份证号码,根据目标用户的身份证号码,从公安部门的声纹特征底库中调用该目标用户的备案声纹特征。

[0132] S42:对声纹特征和备案声纹特征进行比对,得到目标用户的备案声纹比对结果。

[0133] 具体的,将通过声纹识别模型提取的目标用户的声纹特征和目标用户的备案声纹特征进行比对,得到声纹特征和备案声纹特征的声纹相似度值,若声纹相似度值小于声纹相似度阈值,则目标用户的备案声纹比对结果表示目标用户的声纹识别错误,目标用户与备案声纹特征对应的用户不是同一人,双录质检失败,服务器在客户端指示目标用户声纹质检失败,用户认证无效,需重新认证。

[0134] 上述S51包括:

[0135] S51a:若备案声纹比对结果通过,则根据起止时间段,从第一视频图像中截取唇语图像帧序列。

[0136] 具体的,若声纹相似度值大于或等于声纹相似度阈值,则目标用户的备案声纹比对结果表示目标用户的声纹识别正确,目标用户与备案声纹特征对应的用户为同一人,则可进行唇语匹配。

[0137] 本申请实施例提供的双录质检方法,在进行唇语识别和多模态识别之前,通过从预设的声纹数据库中获取目标用户的备案声纹特征,对声纹特征和备案声纹特征进行比对,以得到目标用户的备案声纹比对结果。通过本方法,可通过将从语音数据中提取的目标用户的声纹特征和目标用户的备案声纹特征进行比对,以确认目标用户的身份是否真实,避免他人利用变声装置等冒充目标用户,有效保证用户在远程双录质检过程中身份认证的可靠性,防止他人伪造身份信息进行非法操作,为远程双录质检提供安全保障。

[0138] 在上述任一实施例的基础上,本申请实施例还提供一种双录质检方法,图4为本申请实施例提供的第四种双录质检方法的流程示意图,如图4所示,在上述S41之前,该方法还包括:

[0139] S31:从预设的人脸数据库中获取目标用户的备案人脸图像。

[0140] 具体的,在进行声纹比对之前,还可通过判断目标用户的人脸特征确定用户身份是否真实。预设的人脸数据库中预先存有多个用户真实的备案人脸图像,每个用户具有唯一的用户标识和备案人脸图像,可根据每个用户唯一的用户标识从预设的人脸数据库中调用该用户的备案人脸图像。示例的,预设的人脸数据库可以为用户在公安部门备案的人脸图像底库,根据目标用户的身份证号码,从公安部门的人脸图像底库中调用该目标用户的备案人脸图像。

[0141] S32:采用人脸识别模型,对备案人脸图像进行人脸特征提取,得到第二人脸特征。

[0142] 具体的,为保证比对结果的准确性,备案人脸图像和第一视频图像的人脸图像可采用相同的人脸识别模型提取人脸特征,服务器通过人脸识别模型提取备案人脸图像中的人脸特征,得到第二人脸特征。

[0143] S33:对第一人脸特征和第二人脸特征进行比对,得到目标用户的备案人脸比对结果。

[0144] 具体的,将第一人脸特征和第二人脸特征进行比对,得到第一人脸特征和第二人脸特征的第一人脸相似度值,若第一人脸相似度值小于第一人脸相似度阈值,则目标用户的备案人脸比对结果表示目标用户的人脸识别错误,目标用户和备案人脸图像对应的用户不是同一人,双录质检失败,服务器在客户端指示目标用户人脸质检失败,用户认证无效,需重新认证。

[0145] 上述S41包括:

[0146] S41a:若备案人脸比对结果通过,则从声纹数据库中获取目标用户的备案声纹特征。

[0147] 具体的,若第一人脸相似度值大于或等于第一人脸相似度阈值,则目标用户的备案人脸比对结果表示目标用户的人脸识别正确,目标用户与备案人脸图像对应的用户为同一人,则可进行声纹比对。

[0148] 本申请实施例提供的双录质检方法,在进行声纹比对、唇语识别和多模态识别之前,通过从预设的人脸数据库中获取目标用户的备案人脸图像,采用人脸识别模型提取备案人脸图像的第二人脸特征,并对第一人脸特征和第二人脸特征进行比对,以得到目标用户的备案人脸比对结果。通过本方法,可通过将从第一视频图像中的提取的第一人脸特征和目标用户的备案人脸图像的第二人脸特征进行比对,以确认目标用户的身份是否真实,有效保证用户在远程双录质检过程中身份认证的可靠性,防止他人伪造身份信息进行非法操作,为远程双录质检提供安全保障。

[0149] 在上述任一实施例的基础上,本申请实施例还提供一种双录质检方法,图5为本申请实施例提供的第五种双录质检方法的流程示意图,如图5所示,在上述S31之前,该方法还包括:

[0150] S21:采用预设的身份证检测模型,对客户端采集的第二视频图像进行区域检测,得到身份证文本区域。

[0151] 具体的,在进行人脸比对之前,还可通过判断目标用户的身份文字信息确定用户身份是否真实。预设的身份证检测模型为预先采用第二样本视频图像进行训练得到的模型,其可用于检测视频图像中的身份证文本区域。在实际应用中,可由服务器向客户端发送身份证出示指令,客户端根据该身份证出示指令控制客户端所在设备输出提醒目标用户出示身份证的信息,该信息可以为在客户端所在设备的显示屏上显示的文字信息,也可以是客户端所在设备的扬声器播放的语音信息,摄像头采集的视频图像包括:摄像头针对用户出示的身份证采集的视频图像,如该第二视频图像。摄像头在采集到该第二视频图像之后,可由客户端将采集的第二视频图像传输给服务器,由服务器通过身份证检测模型提取对第二视频图像进行区域检测,得到身份证文本区域。

[0152] S22:采用预设的身份字符识别模型,对身份证文本区域进行文字识别,得到第一身份文字信息。

[0153] 具体的,预设的身份字符识别模型为预先采用样本身份证文字区域进行训练得到的模型,其可用于提取身份证文本区域的身份文字信息,示例的,该预设的身份字符识别模型可以为OCR(Optical Character Recognition,光学字符识别)识别模型。服务器在采用身份证检测模型得到身份证文本区域之后,将身份证文本区域输入至身份字符识别模型,对身份证文本区域进行文字识别,得到第一身份文字信息。示例的,该第一身份文字信息可以包括:姓名、性别、民族、出生年月日、身份证号码、住址。

[0154] S23:对第一身份文字信息和预设的身份信息数据库中目标用户的第二身份文字信息进行比对,得到身份证文字比对结果。

[0155] 具体的,预设的身份信息数据库中预先存有多个用户的真实身份文字信息,每个用户具有唯一的用户表示和真实身份文字信息,可根据目标用户唯一的用户标识从预设的身份信息数据库中调用该目标用户的真实身份文字信息,作为目标用户的第二身份文字信息。示例的,预设的身份信息数据库可以为用户在公安部门备案的身份文字信息底库,根据目标用户的身份证号码,从公安部门的身份文字信息底库中调用该目标用户的第二身份文字信息。

[0156] 将第一身份文字信息和第二身份文字信息进行比对,得到第二身份文字信息和第二身份文字信息的身份相似度值,若身份相似度值小于身份相似度阈值,则目标用户的身份文字比对结果表示目标用户的身份文字信息识别错误,目标用户与第二身份文字信息对应的用户不是同一人,双录质检失败,服务器在客户端指示目标用户身份文字信息质检失败,用户认证无效,需重新认证。

[0157] 上述S31包括:

[0158] S31a:若身份证文字比对结果通过,则从人脸数据库中获取备案人脸图像。

[0159] 具体的,若身份相似度值大于或等于身份相似度阈值,则目标用户的身份文字比对结果表示目标用户的身份文字信息识别正确,目标用户与第二身份文字信息对应的用户为同一人,则可进行人脸比对。

[0160] 本申请实施例提供的双录质检方法,在进行人脸比对、声纹比对、唇语识别和多模态识别之前,通过身份证检测模型提取第二视频图像的身份证文本区域,并通过身份字符识别模型提取第一身份文字信息,将第一身份文字信息和从身份信息数据库中调用的第二身份文字信息进行比对,以得到目标用户的身份证文字比对结果。通过本方法,可以将目标

用户提供的身份证的第一身份文字信息和身份信息数据库中的第二身份文字信息进行比对,以确认目标用户提供的身份证是否真实,避免使用假身份证进行认证,有效保证用户在远程双录质检过程中身份认证的可靠性,防止他人伪造身份信息进行非法操作,为远程双录质检提供安全保障。

[0161] 在上述任一实施例的基础上,本申请实施例还提供一种双录质检方法,图6为本申请实施例提供的第六种双录质检方法的流程示意图,如图6所示,上述S21包括:

[0162] 采用身份证检测模型,对第二视频图像进行区域检测,得到身份证人脸头像区域和身份证文本区域。

[0163] 具体的,身份证检测模型除了可以提取身份证文本区域外,还可以提取身份证人脸头像区域,服务器通过身份证检测模型对第二视频图像中的身份证区域进行检测,提取身份证人脸头像区域和身份证文本区域。

[0164] 上述S31a包括:

[0165] S31a1:采用人脸识别模型,对身份证人脸头像区域进行人脸特征提取,得到第三人脸特征。

[0166] 具体的,为保证比对结果的准确性,身份证人脸头像区域和第一视频图像的人脸图像可采用相同的人脸识别模型提取人脸特征,服务器通过人脸识别模型提取身份证人脸头像区域中的人脸特征,得到第三人脸特征。

[0167] S31a2:对第一人脸特征和第三人脸特征进行比对,得到目标用户的身份证人脸比对结果。

[0168] 具体的,将第一人脸特征和第三人脸特征进行比对,得到第一人脸特征和第三人脸特征的第二人脸相似度值,若第二人脸相似度值小于第二人脸相似度阈值,则目标用户的身份证人脸比对结果表示目标用户的身份证人脸识别错误,目标用户和身份证人脸图像对应的用户不是同一人,双录质检失败,服务器在客户端指示目标用户身份证人脸质检失败,用户认证无效,需重新认证。

[0169] S31a3:若身份证人脸比对结果和身份证文字比对结果均通过,则从人脸数据库中获取备案人脸图像。

[0170] 具体的,若第二人脸相似度值大于或等于第二人脸相似度阈值,则目标用户的身份证人脸比对结果表示目标用户的身份证人脸识别正确,目标用户和身份证人脸图像对应的用户为同一人,则身份证人脸比对结果通过,同时,若上述S23得到的身份证文字比对结果通过,则可进行人脸比对。

[0171] 本申请实施例提供的双录质检方法,在进行人脸比对、声纹比对、唇语识别和多模态识别之前,采用人脸识别模型提取身份证人脸头像区域的第三人脸特征,并对第一人脸特征和第三人脸特征进行比对,以得到目标用户的身份证人脸比对结果。通过本方法,可将第一视频图像中的第一人脸特征和目标用户提供的身份证中的第三人脸特征进行比对,确保人证统一,有效保证用户在远程双录质检过程中身份认证的可靠性,防止他人伪造身份信息进行非法操作,为远程双录质检提供安全保障。

[0172] 在上述任一实施例的基础上,本申请实施例还提供一种双录质检方法,图7为本申请实施例提供的第七种双录质检方法的流程示意图,如图7所示,该方法还包括:

[0173] S500:采用预设的面部动作模型,对第一视频图像进行动作检测,得到第一视频图

像的动作检测结果。

[0174] 具体的,预设的面部动作模型为预先采用第一样本视频图像经过训练得到的模型,其可用于检测视频图像中的面部动作。服务器在完成上述S400对目标用户的多模态融合特征的身份确认之后,还可以采用面部动作模型,对第一视频图像中连续多帧的面部动作进行检测,得到第一视频图像的动作检测结果,动作检测结果用于指示目标用户是否为活体。

[0175] 在一种可能的示例中,面部动作模型为眨眼模型,采用眨眼模型,对第一视频图像中连续多帧的眨眼动作进行检测,得到眨眼检测结果。

[0176] 在另一种可能的示例中,面部动作模型为张闭嘴模型,采用张闭嘴模型,对第一视频图像中连续多帧的张闭嘴动作进行检测,得到张闭嘴检测结果。

[0177] S600:若动作检测结果包括:识别到的预设面部动作的次数大于或等于预设次数阈值,且,身份确认结果通过,则确定目标用户为活体唯一身份。

[0178] 具体的,根据面部动作模型识别的在连续多帧内的预设面部动作的次数,若识别到预设面部动作的次数大于或等于预设次数阈值,则表明目标用户为活体,同时经过上述S400的身份确认结果通过,则确定目标用户为活体唯一身份。

[0179] 在一种可能的示例中,面部动作模型为眨眼模型,根据眨眼模型识别的在连续多帧内的眨眼的次数,则动作检测结果包括:识别到的眨眼动作的次数大于或等于预设次数阈值。

[0180] 在另一中可能的示例中,面部动作模型为张闭嘴模型,根据张闭嘴模型识别的在连续多帧内的张闭嘴的次数,则动作检测结果包括:识别到的张闭嘴动作的次数大于或等于预设次数阈值。

[0181] 本申请实施例提供的双录质检方法,采用面部动作模型检测第一视频图像中的动作,得到第一视频图像的动作检测结果,若该动作检测结果中识别到的预设面部动作的次数大于或等于预设次数阈值,且身份确认结果通过,则确定目标用户为活体唯一身份。通过本方法,检测第一视频图像的面部动作,无需指示用户执行指定动作,在用户无感知的情况下判断目标用户是否为活体,提高用户体验,并在身份确认结果通过的情况下,确定目标用户为活体唯一身份,有效保证用户在远程双录质检过程中身份认证的可靠性。

[0182] 在上述任一实施例的基础上,本申请实施例还提供一种双录质检方法,图8为本申请实施例提供的第八种双录质检方法的流程示意图,如图8所示,该方法包括:

[0183] S100:采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人臉特征。

[0184] S200:采用预设的声纹识别模型,对客户端采集的语音数据进行特征提取,得到声纹特征。

[0185] S300:采用预设的语音识别模型,对语音数据进行特征提取,得到语音文本。

[0186] S21:采用预设的身份证检测模型,对客户端采集的第二视频图像进行区域检测,得到身份证文本区域。

[0187] S22:采用预设的身份字符识别模型,对身份证文本区域进行文字识别,得到第一身份文字信息;

[0188] S23:对第一身份文字信息和预设的身份信息数据库中目标用户的第二身份文字

信息进行比对,得到身份证文字比对结果。

[0189] S31a:若身份证文字比对结果通过,则从人脸数据库中获取备案人脸图像。

[0190] 具体的,S31a可采用上述S31及S31a1-S31a3的方法,在此不做赘述。

[0191] S32:采用人脸识别模型,对备案人脸图像进行人脸特征提取,得到第二人脸特征。

[0192] S33:对第一人脸特征和第二人脸特征进行比对,得到目标用户的备案人脸比对结果。

[0193] S41a:若备案人脸比对结果通过,则从声纹数据库中获取目标用户的备案声纹特征。

[0194] 具体的,S41a采用上述S41的方法,在此不做赘述。

[0195] S42:对声纹特征和备案声纹特征进行比对,得到目标用户的备案声纹比对结果。

[0196] S51a:若备案声纹比对结果通过,则根据起止时间段,从第一视频图像中截取唇语图像帧序列。

[0197] 具体的,S51a采用上述S51的方法,在此不做赘述。

[0198] S52:采用预设的唇语识别模型,判断唇语图像帧序列的动作是否与各个文本段对应的预设唇语动作匹配,得到目标用户的唇语匹配结果。

[0199] S400a:若唇语匹配结果通过,则根据第一人脸特征、声纹特征以及语音文本的特征,采用多模态识别模型进行处理,得到身份确认结果。

[0200] 具体的,S400a采用上述S400的方法,在此不做赘述。

[0201] S500:采用预设的面部动作模型,对第一视频图像进行动作检测,得到第一视频图像的动作检测结果。

[0202] S600:若动作检测结果包括:识别到的预设面部动作的次数大于或等于预设次数阈值,且,身份确认结果通过,则确定目标用户为活体唯一身份。

[0203] 在上述任一实施例的基础上,本申请实施例还提供一种双录质检装置,图9为本申请实施例提供的一种双录质检装置的结构示意图,如图9所示,该装置包括:

[0204] 第一人脸识别模块100,用于采用预设的人脸识别模型,对客户端采集的第一视频图像中的人脸图像进行特征提取,得到第一人脸特征;

[0205] 声纹识别模块200,用于采用预设的声纹识别模型,对客户端采集的语音数据进行特征提取,得到声纹特征;

[0206] 语音识别模块300,用于采用预设的语音识别模型,对语音数据进行特征提取,得到语音文本;

[0207] 身份确认模块400,用于根据第一人脸特征、声纹特征以及语音文本的特征,采用预设的多模态识别模型进行处理,得到目标用户的身份确认结果,身份确认结果用以指示目标用户是否确认身份真实唯一;

[0208] 其中,多模态识别模型为预先采用样本人脸特征、样本声纹特征,以及样本语音文本的特征进行训练得到的模型,样本人脸特征具有是否为真实人脸的标注信息,样本声纹特征具有是否为真实人声的标注信息,样本语音文本的特征具有是否为确认语音的标注信息。

[0209] 可选的,在身份确认模块400之前,该装置还包括:

[0210] 唇语图像帧获取模块,用于根据语音文本中各个文本段的起止时间段,从第一视

频图像中截取起止时间段对应的唇语图像帧序列；

[0211] 唇语识别模块,用于采用预设的唇语识别模型,判断唇语图像帧序列的动作是否与各个文本段对应的预设唇语动作匹配,得到目标用户的唇语匹配结果；

[0212] 身份确认模块,用于若唇语匹配结果通过,则根据第一人臉特征、声纹特征以及语音文本的特征,采用多模态识别模型进行处理,得到身份确认结果。

[0213] 可选的,在唇语图像帧获取模块之前,该装置还包括：

[0214] 备案声纹特征获取模块,用于从预设的声纹数据库中获取目标用户的备案声纹特征；

[0215] 声纹比对模块,用于对声纹特征和备案声纹特征进行比对,得到目标用户的备案声纹比对结果；

[0216] 唇语图像帧获取模块,用于若备案声纹比对结果通过,则根据起止时间段,从第一视频图像中截取唇语图像帧序列。

[0217] 可选的,在备案声纹特征获取模块之前,该装置还包括：

[0218] 备案人脸图像获取模块,用于从预设的人脸数据库中获取目标用户的备案人脸图像；

[0219] 第二人脸识别模块,用于采用人脸识别模型,对备案人脸图像进行人脸特征提取,得到第二人脸特征；

[0220] 人脸比对模块,用于对第一人臉特征和第二人脸特征进行比对,得到目标用户的备案人脸比对结果；

[0221] 备案声纹特征获取模块,用于若备案人脸不低结果通过,则从声纹数据库中获取目标用户的备案声纹特征。

[0222] 可选的,在备案人脸图像获取模块之前,该装置还包括：

[0223] 身份证检测模块,用于采用预设的身份证检测模型,对客户端采集的第二视频图像进行区域检测,得到身份证文本区域；

[0224] 字符识别模块,用于采用预设的身份字符识别模型,对身份证文本区域进行文字识别,得到第一身份文字信息；

[0225] 文字比对模块,用于对第一身份文字信息和预设的身份信息数据库中目标用户的第二身份文字信息进行比对,得到身份证文字比对结果；

[0226] 备案人脸图像获取模块,用于若身份证文字比对结果通过,则从人脸数据库中获取备案人脸图像。

[0227] 可选的,身份证检测模块包括：

[0228] 头像及文本检测单元,用于采用身份证检测模型,对第二视频图像进行区域检测,得到身份证人脸头像区域和身份证文本区域；

[0229] 备案人脸图像获取模块包括：

[0230] 人脸特征识别单元,用于采用人脸识别模型,对身份证人脸头像区域进行人脸特征提取,得到第三人臉特征；

[0231] 人脸特征比对单元,用于对第一人臉特征和第三人臉特征进行比对,得到目标用户的身份证人脸比对结果；

[0232] 备案人脸图像获取单元,用于若身份证人脸比对结果和身份证文字比对结果均通

过,则从人脸数据库中获取备案人脸图像。

[0233] 可选的,该装置还包括:

[0234] 面板动作检测模块,用于采用预设的面部动作模型,对第一视频图像进行动作检测,得到第一视频图像的动作检测结果;

[0235] 活体确认模块,用于若动作检测结果包括:识别到的预设面部动作的次数大于或等于预设次数阈值,且,身份确认结果通过,则确定目标用户为活体唯一身份。

[0236] 上述装置用于执行前述实施例提供的方法,其实现原理和技术效果类似,在此不再赘述。

[0237] 以上这些模块可以是配置成实施以上方法的一个或多个集成电路,例如:一个或多个特定集成电路(Application Specific Integrated Circuit,简称ASIC),或,一个或多个微处理器(digital signal processor,简称DSP),或,一个或者多个现场可编程门阵列(Field Programmable Gate Array,简称FPGA)等。再如,当以上某个模块通过处理元件调度程序代码的形式实现时,该处理元件可以是通用处理器,例如中央处理器(Central Processing Unit,简称CPU)或其它可以调用程序代码的处理器。再如,这些模块可以集成在一起,以片上系统(system-on-a-chip,简称SOC)的形式实现。

[0238] 图10为本申请实施例提供的服务器的示意图,如图10所示,该服务器500包括:处理器501、存储器502,存储器502中存储有处理器501可执行的程序指令,当服务器500运行时,处理器501执行存储器502中存储的程序指令,以执行上述任一方法实施例。具体实现方式和技术效果类似,这里不再赘述。

[0239] 可选地,本发明还提供一种存储介质,存储介质上存储有计算机程序,计算机程序被处理器运行时执行上述任一方法实施例。

[0240] 在本发明所提供的几个实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0241] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0242] 另外,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单元中,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单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实现。

[0243] 上述以软件功能单元的形式实现的集成的单元,可以存储在一个计算机可读存储介质中。上述软件功能单元存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)或处理器(英文:processor)执行本发明各个实施例所述方法的部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(英文:Read-Only Memory,简称:ROM)、随机存取存储器(英文:Random Access Memory,简

称:RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0244] 上仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

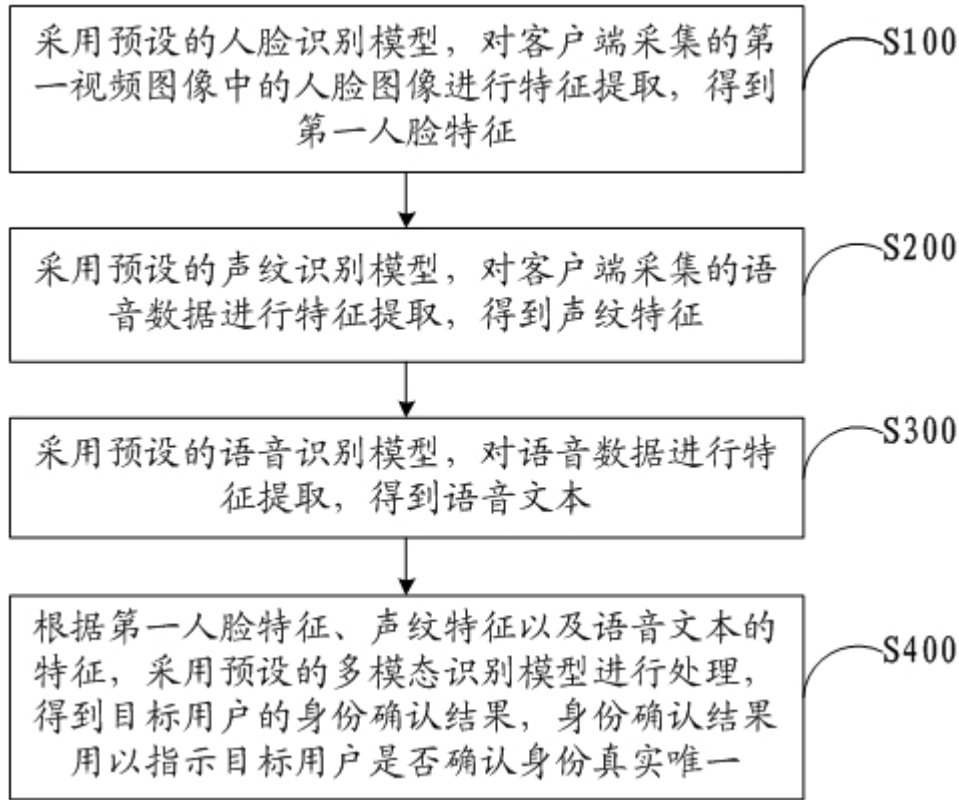


图1

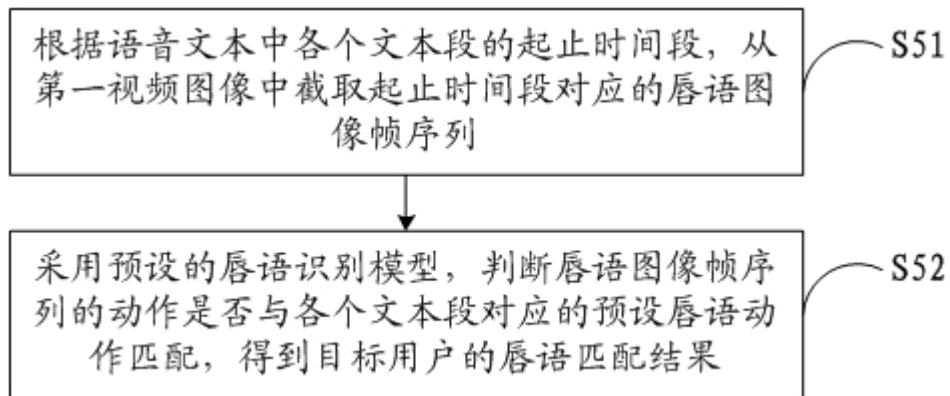


图2

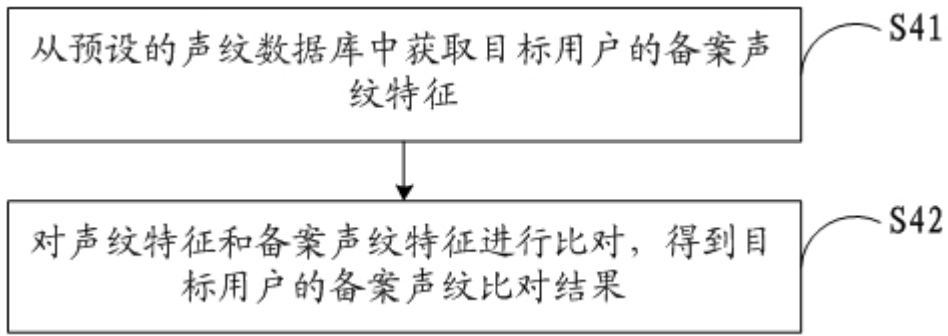


图3

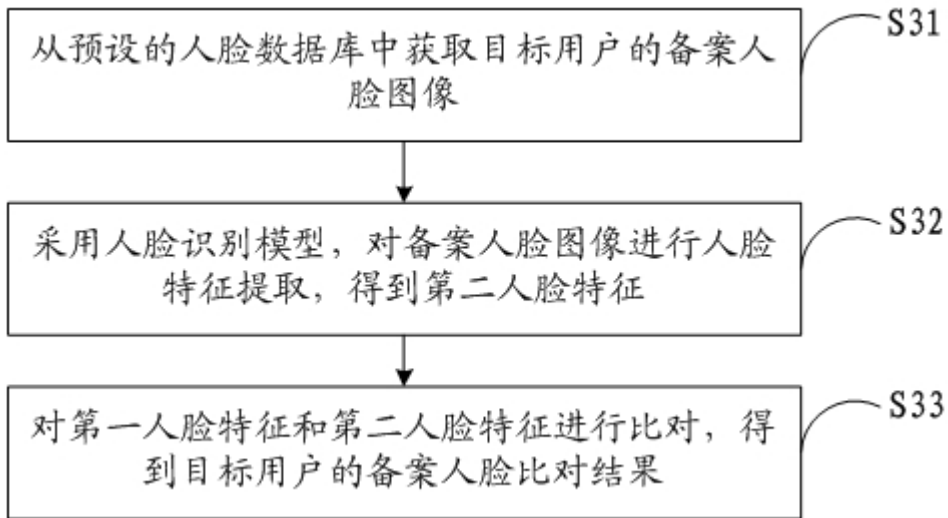


图4

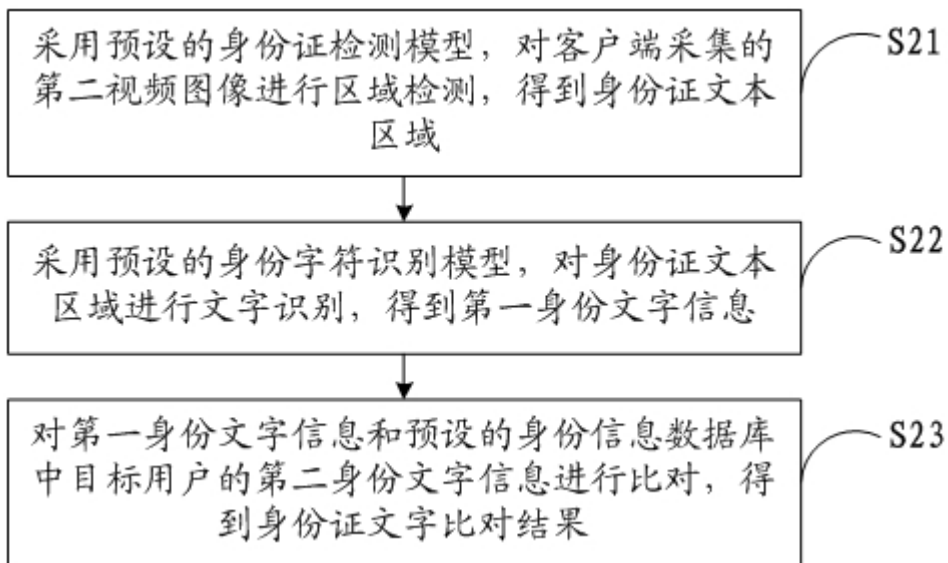


图5

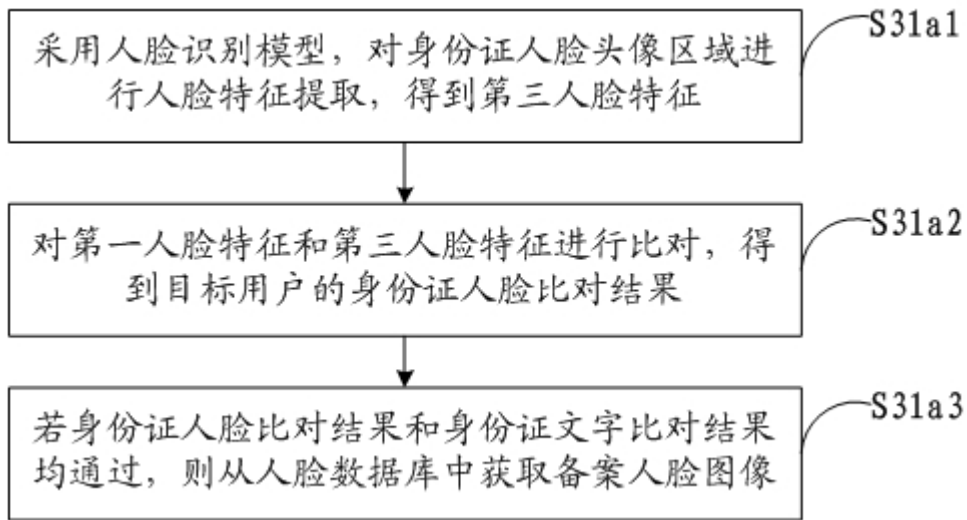


图6

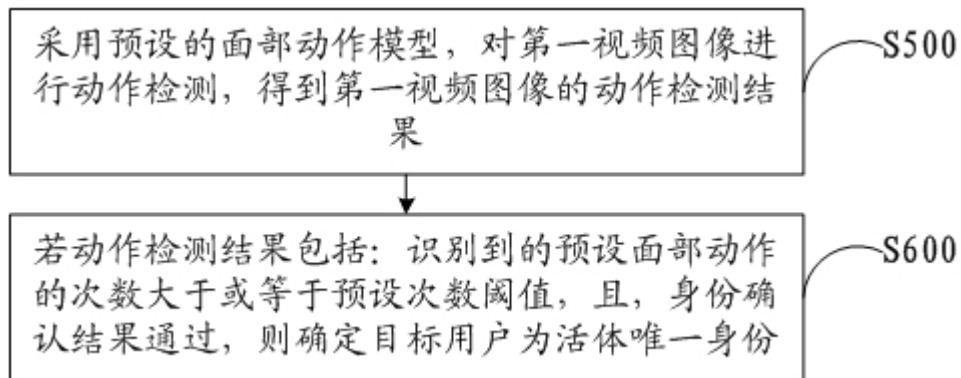


图7

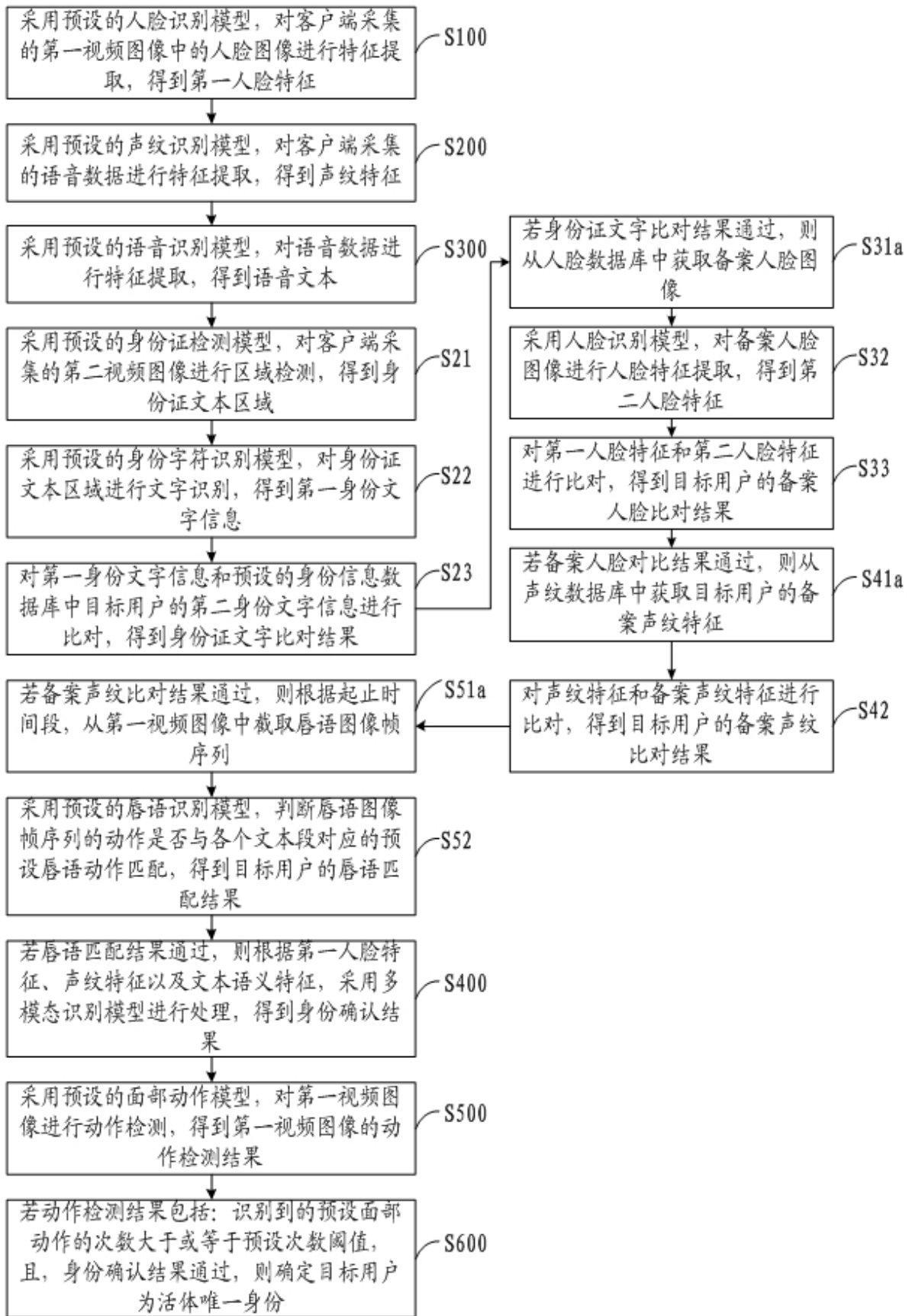


图8

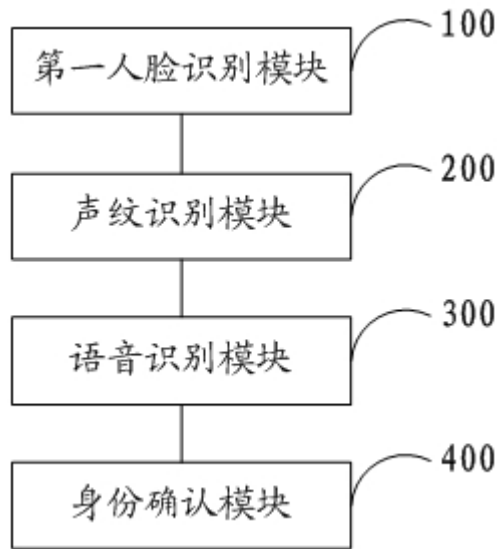


图9

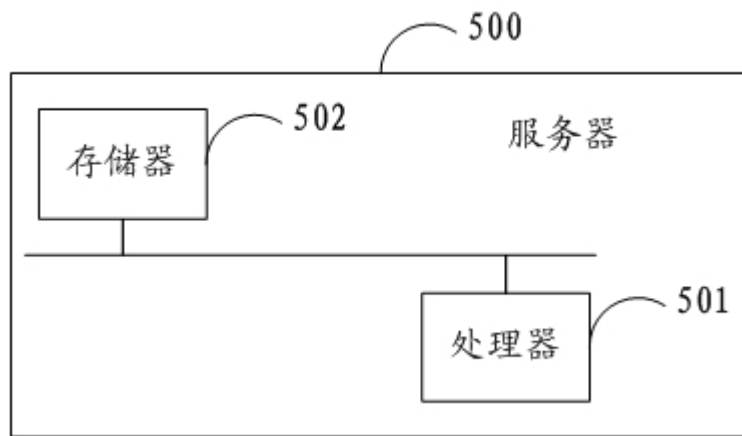


图10