(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104834849 B
(45)授权公告日 2018.09.18

(21)申请号 201510175828.5

(22)申请日 2015.04.14

(65)同一申请的已公布的文献号
申请公布号 CN 104834849 A

(43)申请公布日 2015.08.12

(73)专利权人 北京远鉴科技有限公司
地址 100142 北京市海淀区西四环北路158号慧科大厦东区9A

(72)发明人 张策 龚星 吴鉴 张齐 王黎明

(74)专利代理机构 北京中海智圣知识产权代理有限公司 11282
代理人 白凤武

(51)Int.Cl.
G06F 21/32(2013.01)
G06K 9/00(2006.01)
G10L 17/00(2013.01)

(56)对比文件
CN 103440686 A,2013.12.11,全文.
CN 103841108 A,2014.06.04,全文.
US 2014016835 A1,2014.01.16,全文.
CN 101075868 A,2007.11.21,全文.

审查员 赵洋

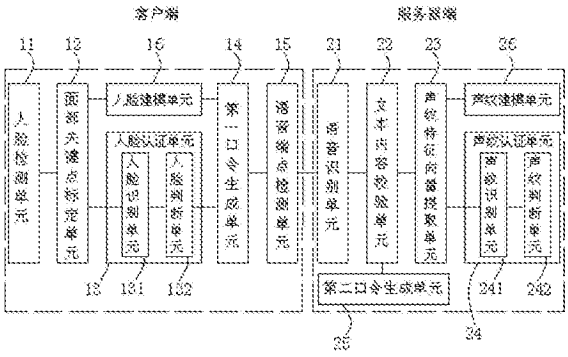
权利要求书2页 说明书6页 附图3页

(54)发明名称

基于声纹识别和人脸识别的双因素身份认证方法及系统

(57)摘要

本发明涉及基于声纹识别和人脸识别的双因素身份认证方法及系统，认证方法包括：检测用户人脸图像；标关键点；人脸认证单元校验人脸相似度；生成口令；采集语音数据及用户ID；对语音数据处理；文本内容校验单元校验口令；声纹认证单元校验声纹相似度。认证系统包括设于客户端的人脸检测单元、面部关键点标定单元、人脸认证单元、第一口令生成单元、语音端点检测单元，及设于服务器端的语音识别单元、文本内容校验单元、声纹特征向量提取单元、声纹认证单元。本发明的优越效果在于：在声纹和人脸识别基础上加入语音识别，提高了认证的安全性和可靠性；口令在客户端和服务器端独立完成，且不以文本或加密文本在服务器端和客户端传输，安全性高。



CN 104834849 B

1.基于声纹识别和人脸识别的双因素身份认证系统的认证方法，所述双因素身份认证系统包括设置于客户端且依次连接的人脸检测单元、面部关键点标定单元、人脸认证单元、第一口令生成单元、语音端点检测单元，所述面部关键点标定单元与第一口令生成单元之间设有人脸建模单元；还包括设置于服务器端且依次连接的语音识别单元、文本内容校验单元、声纹特征向量提取单元、声纹认证单元，所述文本内容校验单元与第二口令生成单元连接，以及与声纹特征向量提取单元连接的声纹建模单元；所述人脸检测单元设有视频采集装置；所述语音端点检测单元设有语音采集单元；所述声纹认证单元包括依次连接的声纹识别单元和声纹判断单元；所述认证系统包括人脸模型训练单元和声纹模型训练单元；

其特征在于，至少包括如下步骤：

S01：人脸检测单元检测请求认证用户的人脸区域图像；

S02：通过面部关键点标定单元在检测到人脸区域内标定面部关键点；

S03：人脸识别单元计算该用户的人脸与客户端存储的注册的人脸模型的相似度，人脸判断单元用于判断人脸相似度是否大于设定的阈值，若人脸相似度大于阈值则通过进入S04，若人脸相似度小于阈值则认证失败；

S04：第一口令生成单元通过随机算法以当前时间和用户ID作为种子生成动态口令文本，同时触发第二口令生成单元生成相同的动态口令文本；

S05：通过语音采集单元采集用户读取动态口令的语音数据；以及通过语音端点检测单元检测用户语音的起始端点和结束端点，并将检测的语音数据及用户ID发送至服务器端；

S06：所述服务器端接收语音数据及用户ID，语音识别单元对接收到的语音数据进行语音识别处理，并将语音数据转换为口令文本；

S07：文本内容校验单元对转换的口令文本与第二口令生成单元生成的动态口令文本进行比对，若口令文本相同则通过进入S08，若口令文本不同则认证失败；

S08：声纹特征向量提取单元从用户的语音数据中提取声纹特征向量，声纹识别单元通过内积计算得到用户声纹特征向量与服务器端存储的注册的声纹模型之间的相似度，声纹判断单元用于判断声纹相似度是否大于设定的阈值，若声纹相似度大于阈值则身份认证成功，若声纹相似度小于阈值则认证失败。

2.根据权利要求1所述的基于声纹识别和人脸识别的双因素身份认证系统的认证方法，其特征在于，所述S01中，所述人脸检测单元检测注册用户的人脸区域图像，且从中截取多张人脸区域图像作为人脸样本，并存储于人脸建模单元内。

3.根据权利要求1所述的基于声纹识别和人脸识别的双因素身份认证系统的认证方法，其特征在于，所述S06中，所述服务器端接收注册用户的语音数据及用户ID，通过声纹特征向量提取单元将语音数据转换成固定长度的声纹特征向量，并以用户ID为索引存储于声纹建模单元内。

4.根据权利要求3所述的基于声纹识别和人脸识别的双因素身份认证系统的认证方法，其特征在于，所述注册用户在注册时录入一次或多次语音数据。

5.根据权利要求1所述的基于声纹识别和人脸识别的双因素身份认证系统的认证方法，其特征在于，所述S08中，所述声纹特征向量提取单元提取声纹特征向量时，将用户语音转化为短时频谱特征序列，计算每一帧频谱特征在全局背景模型各高斯分量上的后验概率，利用最大后验概率准则自适应训练得出用户的高斯混合模型，将高斯混合模型中高斯

分量的均值拼接形成高维向量，所述高维向量为声纹特征向量。

基于声纹识别和人脸识别的双因素身份认证方法及系统

技术领域

[0001] 本发明属于模式识别技术领域，涉及远程身份认证技术，具体涉及一种基于声纹识别和人脸识别的双因素身份认证方法及系统。

背景技术

[0002] 随着移动互联网的高速发展以及手持终端设备如智能手机、平板电脑的普及，互联网安全问题日益突出。目前，无论是银行的硬件数字证书还是动态口令牌，都只做到了对可信终端的管理，无法对用户身份进行验证。

[0003] 生物特征识别技术是利用人的生理特征或行为特征，来进行个人身份的鉴定。已被用于生物识别的生物特征有声音、指纹、人脸、虹膜、视网膜等，而麦克风和摄像头普遍存在于现有的移动终端，因此通过声音和人脸来进行身份认证是最方便、最经济的解决方案。

[0004] 人脸识别是基于人的脸部特征信息进行身份识别的一种生物特征识别技术，主要包括人脸注册和人脸认证两大模块。人脸识别利用摄像头采集含有人脸的图像或视频流，并自动在图像中检测和跟踪人脸，进而对检测到的人脸进行脸部的一系列相关技术处理。

[0005] 人的声音涵盖了多个维度的信息，如说话内容、说话语气、声音特质等。声纹识别是一种通过人的声音特质来辨别不同说话人的技术，不同的声道结构决定了声纹的唯一性。声纹识别主要包括两大模块：声纹注册模块和声纹认证模块。声纹注册是指采用预先选定的模型对用户的语音样本进行建模，生成该用户的声纹模型；在用户请求身份验证时，利用对应的声纹模型对请求语音进行认证。只有经过声纹注册的用户才能使用声纹认证功能。声纹识别结合说话内容，能够有效避免重放攻击。

发明内容

[0006] 本发明的目的在于克服现有技术中的不足，提供一种基于声纹识别和人脸识别的双因素身份认证方法及系统。

[0007] 本发明是通过以下技术方案实现的：

[0008] 基于声纹识别和人脸识别的双因素身份认证方法，至少包括如下步骤：

[0009] S01：所述人脸检测单元检测请求认证用户的人脸区域图像；

[0010] S02：通过面部关键点标定单元在检测到人脸区域内标定面部关键点；

[0011] S03：所述人脸识别单元计算该用户的人脸与客户端存储的注册的人脸模型的相似度，所述人脸判断单元用于判断人脸相似度是否大于设定的阈值，若人脸相似度大于阈值则通过进入S04，若人脸相似度小于阈值则认证失败；

[0012] S04：所述第一口令生成单元通过随机算法以当前时间和用户ID作为种子生成动态口令文本，同时触发第二口令生成单元生成相同的动态口令文本；

[0013] 所述第一口令生成单元和第二口令生成单元利用精确到分钟的当前时间和用户ID生成动态口令。

[0014] S05：通过语音采集单元采集用户读取动态口令的语音数据；以及通过语音端点检

测单元检测用户语音的起始端点和结束端点，并将检测的语音数据及用户ID发送至服务器端；

[0015] S06：所述服务器端接收语音数据及用户ID，所述语音识别单元对接收到的语音数据进行语音识别处理，并将语音数据转换为口令文本；

[0016] S07：所述文本内容校验单元对转换的口令文本与第二口令生成单元生成的动态口令文本进行比对，若口令文本相同则通过进入S08，若口令文本不同则认证失败；

[0017] S08：所述声纹特征向量提取单元从用户的语音数据中提取声纹特征向量，所述声纹识别单元通过内积计算得到用户声纹特征向量与服务器端存储的注册的声纹模型之间的相似度，所述声纹判断单元用于判断声纹相似度是否大于设定的阈值，若声纹相似度大于阈值则身份认证成功，若声纹相似度小于阈值则认证失败。

[0018] 所述的技术方案优选为，所述S08中，所述声纹特征向量提取单元提取声纹特征向量时，将用户语音转化为短时频谱特征序列，计算每一帧频谱特征在全局背景模型各高斯分量上的后验概率，利用最大后验概率准则自适应训练得出用户的高斯混合模型，将高斯混合模型中高斯分量的均值拼接形成高维向量，所述高维向量为声纹特征向量。

[0019] 所述的技术方案优选为，所述短时频谱特征采用梅尔频率倒谱系数或感知线性预测系数。

[0020] 所述的技术方案优选为，所述S01中，所述人脸检测单元检测注册用户的人脸区域图像，且从中截取多张人脸区域图像作为人脸样本，并存储于人脸建模单元内。

[0021] 所述的技术方案优选为，所述人脸样本的采集要求：相邻的两张人脸区域图像的时间间隔至少为500毫秒、且相邻两张人脸区域图像在灰度值上的差异大于预先设定的阈值。

[0022] 所述的技术方案优选为，所述S06中，所述服务器端接收注册用户的语音数据及用户ID，通过声纹特征向量提取单元将语音数据转换成固定长度的声纹特征向量，并以用户ID为索引存储于声纹建模单元内。

[0023] 所述的技术方案优选为，所述注册用户在注册时录入一次或多次语音数据。

[0024] 本发明提供一种基于声纹识别和人脸识别的双因素身份认证系统，采用所述双因素身份认证方法，所述双因素身份认证系统包括设置于客户端且依次连接的人脸检测单元、面部关键点标定单元、人脸认证单元、第一口令生成单元、语音端点检测单元，所述面部关键点标定单元与第一口令生成单元之间设有人脸建模单元；还包括设置于服务器端且依次连接的语音识别单元、文本内容校验单元、声纹特征向量提取单元、声纹认证单元，所述文本内容校验单元与第二口令生成单元连接，以及与声纹特征向量提取单元连接的声纹建模单元。

[0025] 所述的技术方案优选为，所述人脸检测单元设有视频采集装置；所述视频采集装置用于采集用户人脸区域图像。

[0026] 所述人脸检测单元用于用户在发出身份认证请求时，通过视频采集装置采集人脸区域图像；所述面部关键点标定单元用于确定所述人脸区域内的五官位置及轮廓；所述第一口令生成单元用于生成动态口令文本，通过随机算法以当前时间和用户ID作为种子，同时触发设置于服务器端的第二口令生成单元生成相同的动态口令文本；所述语音端点检测单元用于检测用户语音的起始端点和结束端点，将检测的语音数据发送至服务器端。

[0027] 所述语音识别单元用于将客户端发来的语音数据转换成文本内容;所述文本内容校验单元用于将第二口令生成单元生成的动态口令文本与语音识别单元发来的文本内容进行比对,若比对结果一致则通过,若不一致则认证失败。所述声纹特征向量提取单元用于从语音数据中提取能够代表用户声纹的声纹特征向量。所述人脸建模单元用于从注册用户录入的视频数据中选取多张人脸图像样本,进而建立人脸模型,所述声纹建模单元用于从注册用户录入的语音数据中提取声纹特征向量,进而建立声纹模型。

[0028] 所述的技术方案优选为,所述人脸认证单元包括依次连接的人脸识别单元和人脸判断单元。所述人脸识别单元用于计算用户的人脸和客户端存储的注册人脸模型的相似度,所述人脸判断单元用于判断人脸相似度是否大于设定的阈值。

[0029] 所述的技术方案优选为,所述语音端检测单元设有语音采集单元;所述语音采集单元用于采集用户读取动态口令的语音数据。

[0030] 所述的技术方案优选为,所述声纹认证单元包括依次连接的声纹识别单元和声纹判断单元。所述声纹识别单元用于计算用户的声纹特征向量和服务器端存储的注册声纹模型之间的相似度,所述声纹判断单元用于判断声纹相似度是否大于设定的阈值。

[0031] 所述的技术方案优选为,所述语音端点检测单元设有实时数据发送单元;所述语音识别单元设有实时数据接收单元。所述实时数据发送单元用于将语音端点检测单元采集的语音数据发送至服务器端,所述实时数据接收单元用于接收客户端发来的语音数据。

[0032] 所述的技术方案优选为,所述语音端点检测单元与实时数据发送单元之间设有数据压缩单元,所述语音识别单元与实时数据接收单元之间设有数据解压单元。所述数据压缩单元用于将语音数据压缩,所述数据解压单元用于将语音数据解压。

[0033] 所述的技术方案优选为,所述认证系统包括重复性检查单元,用于检查请求用户在客户端及服务器端是否注册,以及是否注册对应的人脸模型和声纹模型。

[0034] 所述的技术方案优选为,所述认证系统包括人脸模型训练单元和声纹模型训练单元;所述人脸模型训练单元用于根据注册用户输入的人脸图像,训练所述注册用户的人脸模型,并存储于客户端,所述声纹模型训练单元用于根据注册用户输入的语音数据,训练所述注册用户的声纹模型,并存储于服务器端。

[0035] 与现有技术相比,本发明的优越效果在于:在声纹识别和人脸识别的基础上,加入语音识别对动态口令进行校验,提高了远程身份认证的安全性和可靠性;所述动态口令同时在客户端和服务器端独立完成,且动态口令不需要以文本或加密文本在服务器端和客户端之间传输,保障了所述动态口令不被网络传输中的第三人窃取和攻击;防范了录音重放的闯入攻击,增强了认证的安全性。

附图说明

[0036] 图1为基于声纹识别和人脸识别的双因素身份认证系统的结构示意图;

[0037] 图2为基于声纹识别和人脸识别的双因素身份认证方法实现流程图;

[0038] 图3为图2所述身份认证方法中注册过程实现流程图。

[0039] 附图标识如下:

[0040] 11-人脸检测单元、12-面部关键点标定单元、13-人脸认证单元、131-人脸识别单元、132-人脸判断单元、14-第一口令生成单元、15-语音端点检测单元、16-人脸建模单元、

21-语音识别单元、22-文本内容校验单元、23-声纹特征向量提取单元、24-声纹认证单元、241-声纹识别单元、242-声纹判断单元、25-第二口令生成单元、26-声纹建模单元。

具体实施方式

[0041] 下面结合附图对本发明具体实施方式作进一步详细说明。

[0042] 如附图1所示，本发明所述基于声纹识别和人脸识别的双因素身份认证系统，包括设置于客户端且依次连接的人脸检测单元11、面部关键点标定单元12、人脸认证单元13、第一口令生成单元14、语音端点检测单元15，所述面部关键点标定单元12与第一口令生成单元14之间设有人脸建模单元16；还包括设置于服务器端且依次连接的语音识别单元21、文本内容校验单元22、声纹特征向量提取单元23、声纹认证单元24，所述文本内容校验单元22与第二口令生成单元连接25，以及与声纹特征向量提取单元23连接的声纹建模单元26。

[0043] 所述人脸检测单元11设有视频采集装置(图中未示)；所述视频采集装置用于采集用户人脸区域图像。所述人脸检测单元11用于用户在发出身份认证请求时，通过视频采集装置采集人脸区域图像；所述面部关键点标定单元12用于确定所述人脸区域内的五官位置，眼睛位置、眉毛位置、鼻子位置及轮廓；所述第一口令生成单元14用于生成动态口令文本，通过随机算法以当前时间和用户ID作为种子，同时触发设置于服务器端的第二口令生成单元25生成相同的动态口令文本；所述语音端点检测单元15用于检测用户语音的起始端点和结束端点，将检测的语音数据发送至服务器端。所述人脸建模单元16用于从注册用户录入的视频数据中选取多张人脸图像样本，进而建立人脸模型。

[0044] 所述语音端检测单元15设有语音采集单元(图中未示)，如麦克风；所述语音采集单元用于采集用户读取动态口令的语音数据。所述语音端点检测单元15检测时，每隔20毫秒计算当前语音片段的能量，若当前能量高于预先设定的阈值，则将当前片段标记为有效语音片段，否则将当前片段标记为无效语音片段。记录有效语音片段和无效语音片段的数目，若连续出现10个无效语音片段且有效语音片段数大于50，则表明用户读取动态口令结束，所述语音采集单元停止采集语音数据。然后将第一个有效语音片段出现的位置标记为语音起始端点，将最后一个有效语音片段出现的位置标记为语音结束端点。所述声纹建模单元26用于从注册用户录入的语音数据中提取声纹特征向量，进而建立声纹模型。

[0045] 所述人脸认证单元13包括依次连接的人脸识别单元131和人脸判断单元132。所述人脸识别单元131用于计算用户的人脸和客户端存储的注册的人脸模型的相似度，所述人脸判断单元132用于判断人脸相似度是否大于设定的阈值。

[0046] 所述声纹认证单元24包括依次连接的声纹识别单元241和声纹判断单元242。所述声纹识别单元241用于计算用户的声纹特征向量和服务器端存储的注册的声纹模型之间的相似度，所述声纹判断单元242用于判断声纹相似度是否大于设定的阈值。

[0047] 所述语音端点检测单元15设有实时数据发送单元(图中未示)；所述语音识别单元21设有实时数据接收单元(图中未示)。所述实时数据发送单元用于将语音端点检测单元15采集的语音数据发送至服务器端，所述实时数据接收单元用于接收客户端发来的语音数据。进一步地，所述语音端点检测单元15与实时数据发送单元之间设有数据压缩单元(图中未示)，所述语音识别单元21与实时数据接收单元之间设有数据解压单元(图中未示)。所述数据压缩单元用于将语音数据压缩，所述数据解压单元用于将语音数据解压。通过对客户

端对输入语音压缩和实时网络传输技术，进一步降低了网络传输的带宽要求；同时在服务器端实时接收以及数据解压，极大地提高了用户注册及认证过程的响应速度。

[0048] 本发明一种基于声纹识别和人脸识别的双因素身份认证方法，如图2所示，采用上述双因素身份认证系统，包括如下步骤：

[0049] S01：所述人脸检测单元11检测请求认证用户的人脸区域图像；所述人脸检测单元11检测注册用户的人脸区域图像，且从中截取多张人脸区域图像作为人脸样本，并存储于人脸建模单元16内；所述人脸样本的采集要求：相邻的两张人脸区域图像的时间间隔至少为500毫秒、且相邻两张人脸区域图像在灰度值上的差异大于预先设定的阈值。

[0050] S02：通过面部关键点标定单元12在检测到人脸区域内标定面部关键点。

[0051] S03：所述人脸识别单元131计算该用户的人脸与客户端存储的注册的人脸模型的相似度，所述人脸判断单元132用于判断人脸相似度是否大于设定的阈值，若人脸相似度大于阈值则通过进入S04，若人脸相似度小于阈值则认证失败。

[0052] S04：所述第一口令生成单元14通过随机算法以当前时间和用户ID作为种子生成动态口令文本，同时触发第二口令生成单元25生成相同的动态口令文本。所述第一口令生成单元14和第二口令生成单元25利用精确到分钟的当前时间和用户ID生成动态口令文本。

[0053] S05：通过语音采集单元采集用户读取动态口令的语音数据；以及通过语音端点检测单元15检测用户语音的起始端点和结束端点，并将检测的语音数据及用户ID发送至服务器端。

[0054] S06：所述服务器端接收语音数据及用户ID，所述语音识别单元21对接收到的语音数据进行语音识别处理，并将语音数据转换为口令文本。所述服务器端接收注册用户的语音数据及用户ID，通过声纹特征向量提取单元23将语音数据转换成固定长度的声纹特征向量，并以用户ID为索引存储于声纹建模单元26内；所述注册用户在注册时录入一次或多次语音数据。

[0055] S07：所述文本内容校验单元22对转换的口令文本与第二口令生成单元25生成的动态口令文本进行比对，若口令文本相同则通过进入S08，若口令文本不同则认证失败。

[0056] S08：所述声纹特征向量提取单元23从用户的语音数据中提取声纹特征向量，所述声纹识别单元241通过内积计算得到用户声纹特征向量与服务器端存储的注册的声纹模型之间的相似度，所述声纹判断单元242用于判断声纹相似度是否大于设定的阈值，若声纹相似度大于阈值则身份认证成功，若声纹相似度小于阈值则认证失败。

[0057] 所述声纹特征向量提取单元23提取声纹特征向量时，将用户语音转化为短时频谱特征序列，计算每一帧频谱特征在全局背景模型各高斯分量上的后验概率，利用最大后验概率准则自适应训练得出用户的高斯混合模型，将高斯混合模型中高斯分量的均值拼接形成高维向量，所述高维向量为声纹特征向量。所述短时频谱特征采用梅尔频率倒谱系数、感知线性预测系数。

[0058] 本发明通过主成分分析对所述声纹特征向量降维，压缩声纹模型大小、提高声纹相似度计算模块速度；具体包括基于背景语音训练数据集，利用声纹特征向量提取单元23提取数据集中每一条语音的声纹特征向量，使用主成分分析计算出对应的特征值和特征向量，将特征值从大到小排序，选取前N个特征值所对应的特征向量构成降维后的特征子空间，利用所述子空间对注册声纹特征向量和认证声纹特征向量进行降维，得到N维声纹特征

向量。

[0059] 本发明并不限于上述实施方式，在不背离本发明的实质内容的情况下，本领域技术人员可以想到的任何变形、改进、替换均落入本发明的范围。

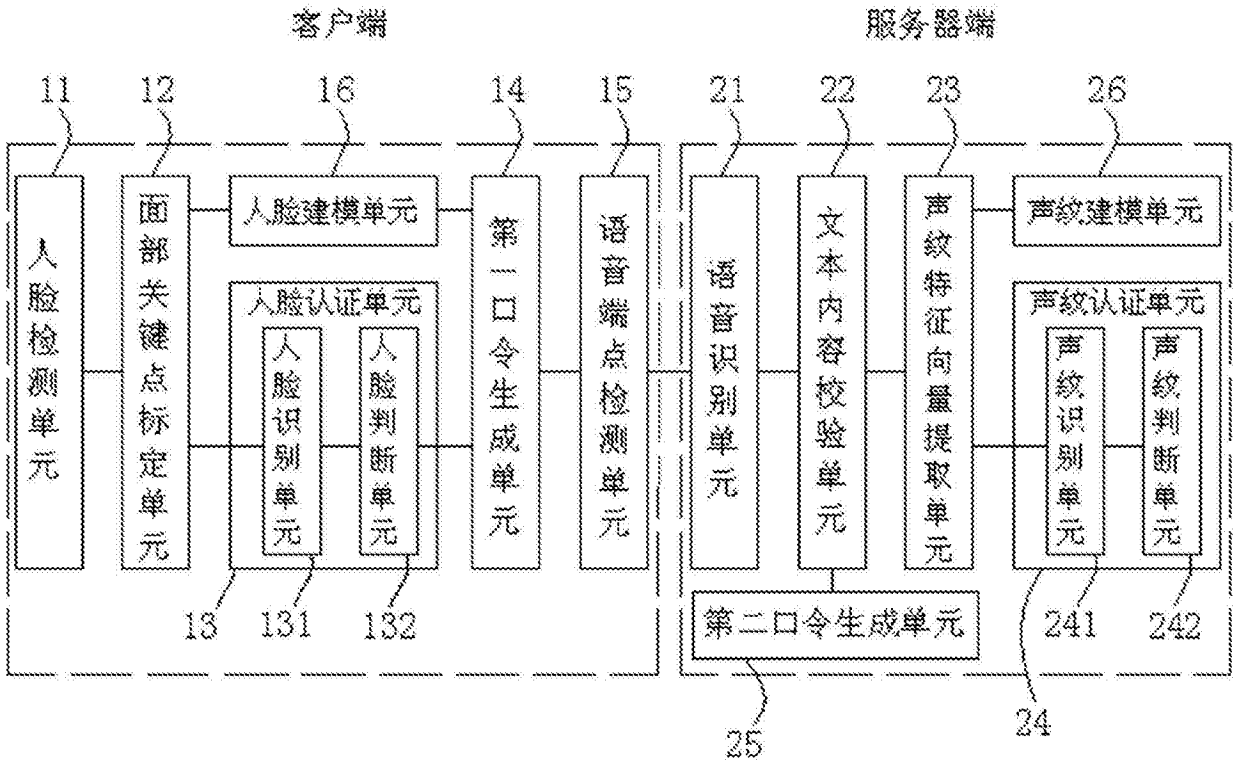

客户端
服务器端
11
12
16
14
15
21
22
23
26
人脸检测单元
面部关键点标定单元
人脸建模单元
人脸认证单元
人脸识别单元
人脸判断单元
第一口令生成单元
语音端点检测单元
语音识别单元
文本内容校验单元
声纹特征向量提取单元
声纹建模单元
声纹认证单元
声纹识别单元
声纹判断单元
第二口令生成单元
13
131
132
241
242
25
24

图1

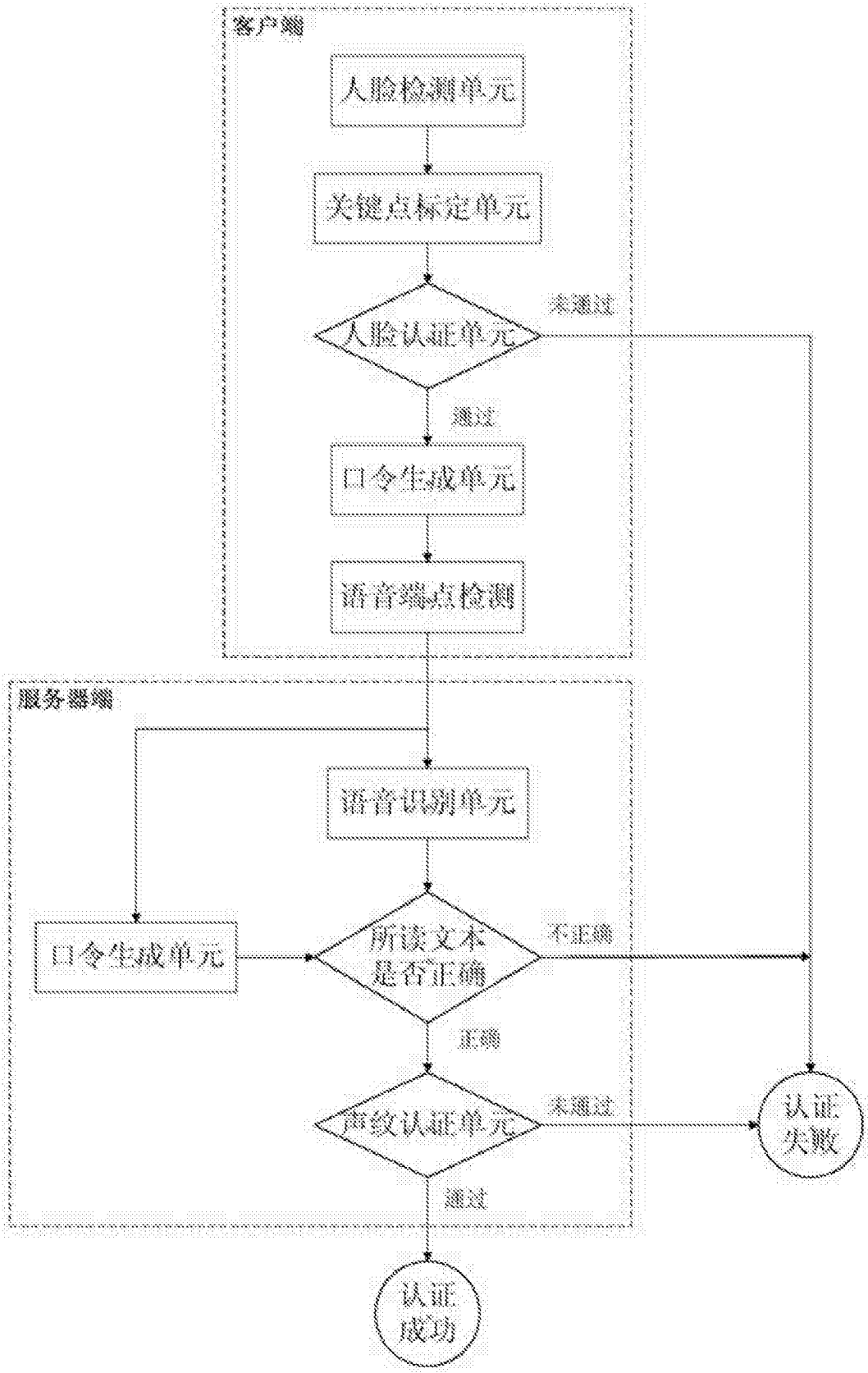

客户端
人脸检测单元
关键点标定单元
人脸认证单元
未通过
通过
口令生成单元
语音端点检测
服务器端
语音识别单元
口令生成单元
所读文本是否正确
不正确
正确
声纹认证单元
未通过
通过
认证失败
认证成功

图2

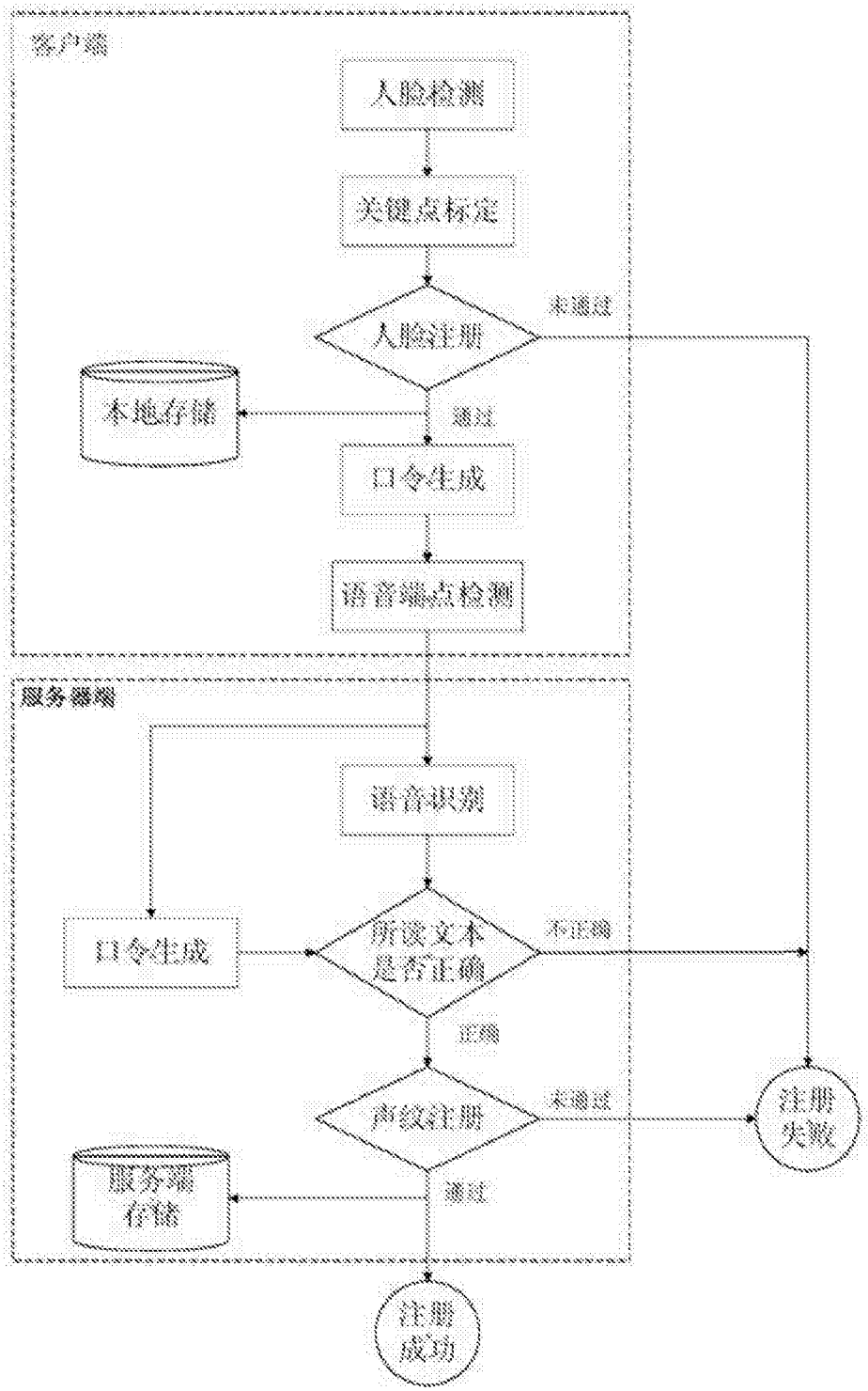

客户端
人脸检测
关键点标定
人脸注册
未通过
本地存储
通过
口令生成
语音端点检测
服务器端
语音识别
口令生成
所读文本是否正确
不正确
正确
声纹注册
未通过
注册失败
服务端存储
通过
注册成功

图3