

说明书摘要

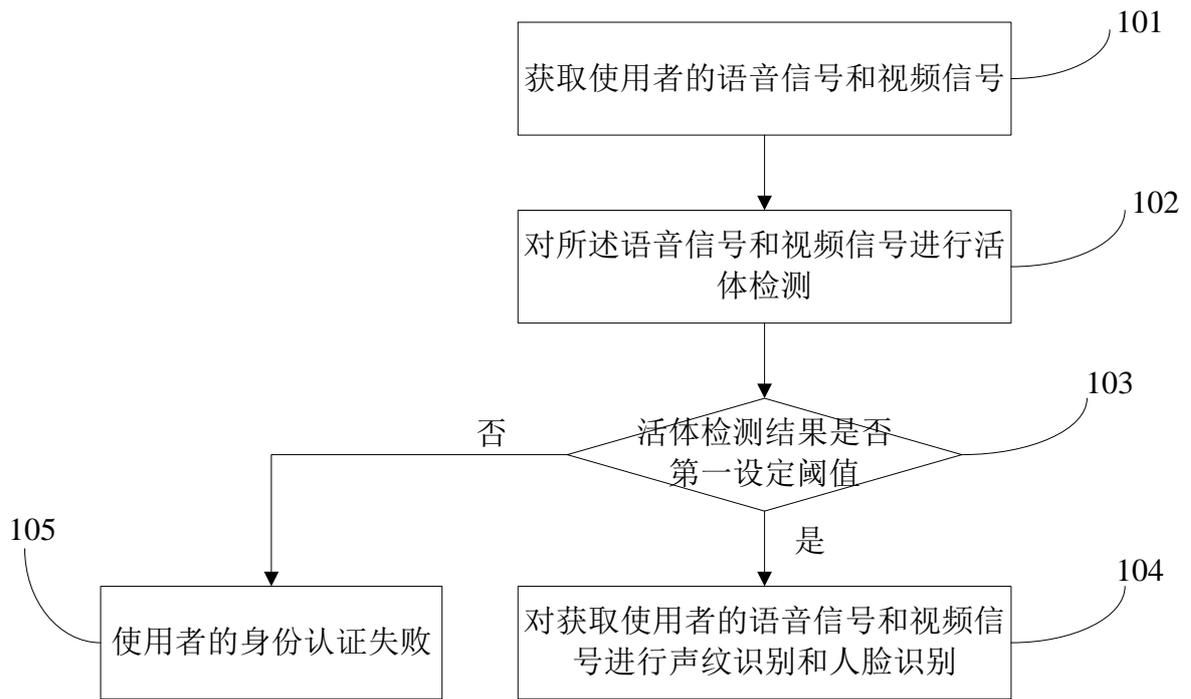
本申请提供了一种基于声纹识别、人脸识别以及活体检测的身份认证方法和系统，包括：获取使用者的语音信号和视频信号；对所述语音信号和视频信号进行活体检测，获得活体检测结果；若所述活体检测结果大于第一设定阈值，则对获取使用者的语音信号和视频信号进行声纹识别和人脸识别，若声纹识别和人脸识别全部通过，则身份认证通过，否则身份认证失败，通过对获取的语音信号和视频信号进行活体检测、声纹识别和人脸识别，从而极大的阻止了非法用户的闯入，通过对使用者的语音信号和视频信号进行多重检测，从而解决了现有技术单一生物特征识别技术的不足，进而提高了身份认证的准确性、安全性和可靠性。

10

15



摘要附图



权 利 要 求 书

1、一种基于声纹识别、人脸识别以及活体检测的身份认证方法，其特征在于，包括：

获取使用者的语音信号和视频信号；

对所述语音信号和视频信号进行活体检测，获得活体检测结果；

5 若所述活体检测结果大于第一设定阈值，则对获取使用者的语音信号和视频信号进行声纹识别和人脸识别，若声纹识别和人脸识别全部通过，则身份认证通过，否则身份认证失败。

2、根据权利要求1所述的方法，其特征在于，所述活体检测包括语音识别和唇语识别。

10 3、根据权利要求2所述的方法，其特征在于，所述对所述语音信号和视频信号进行活体检测，获得活体检测结果的步骤包括：

对所述语音信号进行语音识别，获得语音信号对应的语言信息；

对所述视频信号进行分帧处理，获得分帧后的每帧图像中的嘴唇位置；

15 对所述每帧图像中的嘴唇位置进行唇语识别，获得每帧图像的唇语对应的语言信息；

使用时间对准算法计算语音信号对应的语言信息与唇语识别对应的语言信息的相似度值，将所述相似度值作为活体检测结果。

20 4、根据权利要求1所述的方法，其特征在于，所述对获取使用者的语音信号和视频信号进行声纹识别和人脸识别的步骤包括：

对使用者的语音信号进行声纹识别，得到语音信号的声纹识别打分；

对使用者的人脸视频图像，进行分帧，获取人脸连续帧图像，对所述人脸连续帧图像进行人脸识别，对获得的人脸识别进行分数融合，得到人脸识别打分；

25 根据所述声纹识别打分和人脸识别打分，获得使用者身份认证的判决得分；

若所述判决得分大于第二设定阈值，则使用者身份认证成功；

若所述判决得分小于第二设定阈值，则使用者身份认证失败。



5、根据权利要求 1 所述的方法，其特征在于，还包括：使用视频图像采集器获取使用者视频信号，使用语音采集器获取使用者的语音信号。

6、一种基于声纹识别、人脸识别以及活体检测的身份认证系统，其特征在于，包括：

5 获取模块，用于获取使用者的语音信号和视频信号；

活体检测模块，用于对所述语音信号和视频信号进行活体检测，获得活体检测结果；

第一判断模块，用于若所述活体检测结果大于第一设定阈值，则对获取使用者的语音信号和视频信号进行声纹识别和人脸识别；若声纹识别和
10 人脸识别全部通过，则身份认证通过，否则身份认证失败。

7、根据权利要求 6 所述的系统，其特征在于，所述活体检测包括语音识别和唇语识别。

8、根据权利要求 7 所述的系统，其特征在于，活体检测模块包括：

15 语音识别模块，用于对所述语音信号进行语音识别，获得语音信号对应的语言信息；

分帧模块，用于对所述视频信号进行分帧处理，获得分帧后的每帧图像中的嘴唇位置；

唇语识别模块，用于对所述每帧图像中的嘴唇位置进行唇语识别，获得每帧图像的唇语识别对应的语言信息；

20 计算模块，用于使用时间对准算法计算语音信号对应的语言信息与唇语识别对应的语言信息的相似度值，将所述相似度值作为活体检测结果。

9、根据权利要求 6 所述的系统，其特征在于，判断模块包括：

声纹识别模块，用于对使用者的语音信号进行声纹识别，得到语音信号的声纹识别打分；

25 人脸识别模块，用于对使用者的人脸视频图像，进行分帧，获取人脸连续帧图像，对所述人脸连续帧图像进行人脸识别，对获得的人脸识别进行分数融合，得到人脸识别打分；

判决模块，用于根据所述声纹识别打分和人脸识别打分，获得使用者



身份认证的判决得分；

第二判断模块，用于若所述判决得分大于第二设定阈值，则使用者身份认证成功；

若所述判决得分小于第二设定阈值，则使用者身份认证失败。

- 5 10、根据权利要求6所述的系统，其特征在于，还包括：使用视频图像采集器获取使用者视频信号，使用语音采集器获取使用者的语音信号。

10

15

20



说明书

基于声纹识别、人脸识别以及同步活体检测的身份认证方法及系统

技术领域

本申请涉及计算机信息服务技术领域，特别是涉及一种基于声纹识别、人脸识别以及同步活体检测的身份认证方法及系统。

5 背景技术

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。

- 10 近几年，随着公众账户安全性意识的提高，现有技术常采用单一生物特征技术实现对账户的身份验证，但是单一生物特征技术对重复录音、模仿及合成的语音闯入识别性能不是很好，并且在安全性要求较高的场合单独使用单一生物特征认证存在一定的风险。例如，采用人脸技术实现对身份的验证，但是由于人脸的外形很不稳定，人可以通过脸部变化产生各种表情；此外，不同的观测角度和光照条件，人脸的视觉图像也相差较大。因此，当闯入者利用高像素图像闯入人脸识别系统，此时人脸识别系统通过了验证，但是实际上该闯入者是非法用户，从而给用户带来了安全隐患。
- 15

发明内容

- 20 本申请提供一种基于声纹识别、人脸识别以及同步活体检测的身份认证方法及系统，以解决身份识别不准确和安全性差的问题。

为了解决上述问题，本申请公开了一种基于声纹识别、人脸识别以及活体检测的身份认证方法，包括：获取使用者的语音信号和视频信号；



对所述语音信号和视频信号进行活体检测，获得活体检测结果；

若所述活体检测结果大于第一设定阈值，则对获取使用者的语音信号和视频信号进行声纹识别和人脸识别：若声纹识别和人脸识别全部通过，则身份认证通过，否则身份认证失败。

5 优选地，所述活体检测包括语音识别和唇语识别。

优选地，所述对所述语音信号和视频信号进行活体检测，获得活体检测结果的步骤包括：

对所述语音信号进行语音识别，获得语音信号对应的语言信息；

10 对所述视频信号进行分帧处理，获得分帧后的每帧图像中的嘴唇位置；

对所述每帧图像中的嘴唇位置进行唇语识别，获得每帧图像的唇语对应的语言信息；

使用时间对准算法计算语音信号对应的语言信息与唇语识别对应的语言信息的相似度值，将所述相似度值作为活体检测结果。

15 优选地，所述对获取使用者的语音信号和视频信号进行声纹识别和人脸识别的步骤包括：

对使用者的语音信号进行声纹识别，得到语音信号的声纹识别打分；

20 对使用者的人脸视频图像，进行分帧，获取人脸连续帧图像，对所述人脸连续帧图像进行人脸识别，对获得的人脸识别进行分数融合，得到人脸识别打分；

根据所述声纹识别打分和人脸识别打分，获得使用者身份认证的判决得分；

若所述判决得分大于第二设定阈值，则使用者身份认证成功；

若所述判决得分小于第二设定阈值，则使用者身份认证失败。

25 优选地，还包括：使用视频图像采集器获取使用者视频信号，使用语音采集器获取使用者的语音信号。

为了解决上述问题，本申请还公开了一种基于声纹识别、人脸识别以及活体检测的身份认证系统，包括：获取模块，用于获取使用者的语音信



号和视频信号；

活体检测模块，用于对所述语音信号和视频信号进行活体检测，获得活体检测结果；

5 第一判断模块，用于若所述活体检测结果大于第一设定阈值，则对获取使用者的语音信号和视频信号进行声纹识别和人脸识别；若声纹识别和人脸识别全部通过，则身份认证通过，否则身份认证失败。

优选地，所述活体检测包括语音识别和唇语识别。

优选地，活体检测模块包括：

10 语音识别模块，用于对所述语音信号进行语音识别，获得语音信号对应的语言信息；

分帧模块，用于对所述视频信号进行分帧处理，获得分帧后的每帧图像中的嘴唇位置；

唇语识别模块，用于对所述每帧图像中的嘴唇位置进行唇语识别，获得每帧图像的唇语识别对应的语言信息；

15 计算模块，用于使用时间对准算法计算语音信号对应的语言信息与唇语识别对应的语言信息的相似度值，将所述相似度值作为活体检测结果。

优选地，判断模块包括：

声纹识别模块，用于对使用者的语音信号进行声纹识别，得到语音信号的声纹识别打分；

20 人脸识别模块，用于对使用者的人脸视频图像，进行分帧，获取人脸连续帧图像，对所述人脸连续帧图像进行人脸识别，对获得的人脸识别进行分数融合，得到人脸识别打分；

判决模块，用于根据所述声纹识别打分和人脸识别打分，获得使用者身份认证的判决得分；

25 第二判断模块，用于若所述判决得分大于第二设定阈值，则使用者身份认证成功；

若所述判决得分小于第二设定阈值，则使用者身份认证失败。

优选地，还包括：使用视频图像采集器获取使用者视频信号，使用语



音采集器获取使用者的语音信号。

与现有技术相比，本申请包括以下优点：

本申请将获取的语音信号和视频信号先进行活体检测，当活体检测结果小于等于第一设定阈值时，则身份认证失败，当活体检测结果大于第一设定阈值，则继续对获取的语音信号和视频信号进行声纹识别和人脸识别，若声纹识别和人脸识别全部通过，则该使用者的身份认证通过，否则认证失败，通过对获取的语音信号和视频信号进行活体检测、声纹识别和人脸识别，从而极大的阻止了非法用户的闯入，通过对使用者的语音信号和视频信号进行多重检测，从而解决了现有技术单一生物特征识别技术的不足，进而提高了身份认证的准确性、安全性和可靠性。

附图说明

图1是本申请实施例一所述一种基于声纹识别、人脸识别以及活体检测的身份认证方法的流程图；

图2是本申请实施例二一种基于声纹识别和人脸识别的身份认证方法的流程图；

图3是本申请实施例三所述一种基于声纹识别、人脸识别以及活体检测的身份认证方法的流程图；

图4是本申请采集器的结构示意图；

图5是本申请语音信号对应的语言信息与唇语识别对应的语言信息同步检测的示意图；

图6是本申请实施例四所述一种基于声纹识别、人脸识别以及活体检测的身份认证方法的流程图；

图7是本申请实施例五所述一种基于声纹识别、人脸识别以及活体检测的身份认证系统的结构框图。

具体实施方式

为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附



图和具体实施方式对本申请作进一步详细的说明。

实施例一

本申请基于声纹识别、人脸识别以及活体检测的身份认证方法可以在线使用或者离线状态使用。

- 5 在线使用该身份认证的方法是基于客户端与服务器的架构，使用者通过客户端向服务端发送身份验证请求，服务器对身份验证请求进行验证，当身份请求验证通过，服务端向客户端反馈信息采集指令，客户端接收服务端的信息采集指令，客户端根据该信息采集指令开启摄像头和麦克风。

服务端生成动态提示码，并将该动态提示码发送给客户端。

- 10 客户端根据提示码采集语音信号和视频信号。

客户端将采集得到的语音和视频信号发送至服务器，根据服务端的反馈信号，做出相应的响应。

离线状态可以再客户端单独使用，也可以在服务器端单独使用，对本申请不做具体限制。

- 15 参照图 1，示出了本申请一种基于声纹识别、人脸识别以及活体检测的身份认证方法，该方法包括：

步骤 101：获取使用者的语音信号和视频信号。

服务端接收语音信号和视频信号，调用语音识别模块对语音信号进行识别，获取语音信号对应的语言信息。

- 20 步骤 102：对所述语音信号和视频信号进行活体检测，获得活体检测结果。

步骤 103：若所述活体检测结果大于第一设定阈值，则执行步骤 104，若所述活体检测结果小于等于第一设定阈值，则执行步骤 105。

- 25 若活体检测结果大于第一设定阈值，则认为使用者的活体存在；反之，则认证失败。

服务端将检测结果反馈到用户端；若检测结果未通过，则服务端结束认证，并向客户端反馈认证失败信号，若检测结果通过，服务端继续进行后续的声纹识别和人脸识别。



其中，第一设定阈值和第二设定阈值的设定可以由本领域技术人员采用任意适当方式进行设定，如可以采用人工经验设定阈值，或者针对历史数据的差异值设定阈值，本申请对此不作限制。

5 步骤 104：对获取使用者的语音信号和视频信号进行声纹识别和人脸识别，若声纹识别和人脸识别全部通过，则身份认证通过，否则身份认证失败。

步骤 105：使用者的身份认证失败。

本实施例，通过将获取的语音信号和视频信号先进行活体检测，当活体检测结果小于等于第一设定阈值时，则身份认证失败，当活体检测结果大于第一设定阈值，则继续对获取的语音信号和视频信号进行声纹识别和人脸认识，若声纹识别和人脸识别全部通过，则该使用者的身份认证通过，否则认证失败，通过对获取的语音信号和视频信号进行活体检测、声纹识别和人脸识别，从而极大的阻止了非法用户的闯入，通过对使用者的语音信号和视频信号进行多重检测，提高了身份认证的准确性、安全性和可靠性，从而解决了现有技术单一生物特征识别技术不足的问题。

10
15

实施例二

参照图 2，示出了本申请一种基于声纹识别和人脸识别的身份认证方法，该方法包括：

步骤 201：使用采集器采集使用者的语音信号和视频信号。

20 步骤 202：对所述语音信号和视频信号进行声纹识别和人脸识别。

将采集到的语音信号和视频信号分别传入声纹识别认证模块与人脸识别认证模块，对语音信号进行声纹识别，对视频信号进行人脸识别。

需要说明的是可以对语音信号先进行声纹识别，然后对视频信号进行人脸识别，也可以先对视频信号进行人脸识别，然后对语音信号先进行声纹识别，还可以同时对语音信号进行声纹识别，对视频信号进行人脸识别，对此本申请不做具体限制。

25

步骤 203：若声纹识别和人脸识别全部通过，则执行步骤 204，若未通过则执行步骤 205。



步骤 204：身份认证通过。

步骤 205：身份认证失败。

本申请将获取的语音信号和视频信号进行声纹识别和人脸识别，若声纹识别和人脸识别全部通过，则该使用者的身份认证通过，否则认证失败，通过对获取的语音信号和视频信号进行声纹识别和人脸识别，从而极大的阻止了非法用户的闯入，通过对使用者的语音信号和视频信号进行多重检测，从而解决了现有技术单一生物特征识别技术的不足，进而提高了身份认证的准确性、安全性和可靠性。

实施例三

10 参照图 3，示出了本申请一种基于声纹识别、人脸识别以及活体检测的身份认证方法，该方法包括：

步骤 301：使用采集器采集使用者的语音信号和视频信号。

采集器包括视频图像采集器和语音采集器，使用视频图像采集器获取使用者视频信号，使用语音采集器获取使用者的语音信号。

15 参见图 4 示出了采集器的结构示意图，

使用者通过客户端向服务端发送身份验证请求，服务器对身份验证请求进行验证，当身份请求验证通过，服务端向客户端反馈信息采集指令，客户端开始视频图像采集器和语音采集器，实现视频信号和语音信号的同步采集，其中，语音采集器包括麦克风或手机或电话，使用麦克风或手机或电话采集语音信号，视频图像采集器采集使用者的人脸和嘴唇。

采集器主要包括：提示显示屏和人脸显示屏，使用提示显示屏显示服务器生成的文本串以及服务器反馈的识别结果。

人脸显示屏，提供用户交互功能，实时呈现用户脸部位置，方便用户实时调整使用者脸部位置和动作，根据具体情况采集器还可包括指示灯、启动钮和文本提示显示屏，其中，文本提示显示屏作为辅助设备，不局限于显示屏，还可以包括语音提示等方式。

采集器通过端口与对应主控电路相连，在实际中，使用者点击启动钮，主控电路接收信号并启动语音视频采集器；待指示灯和文本提示显示



屏给出信息采集提示后，使用者人脸显示屏提示的文本内容；在信息采集过程中，服务器会要求使用者整个人脸图像能够实时呈现在人脸显示屏中；若人脸显示屏显示的人脸不全时，则服务器会向使用者发出提示信号，进行调整，当使用者点击按钮，完成视频信号采集；

- 5 采集器将采集得到的语音信号和视频信号通过认证接口进行发送到服务器进行身份认证，该认证接口可以在本地完成身份认证。

步骤 302：对所述语音信号和视频信号进行活体检测，获得活体检测结果。

所述活体检测包括语音识别和唇语识别。

- 10 需要说明的是可以对语音信号先进行语音识别，然后对视频信号进行唇语脸识别，也可以先对视频信号进行唇语脸识别，然后对语音信号进行语音识别，还可以同时对语音信号进行语音识别，对视频信号进行唇语识别，对此本申请不做具体限制。

步骤 302 包括以下子步骤：

- 15 步骤 3021：对所述语音信号进行语音识别，获得语音信号对应的语言信息。

使用预先训练的语音语言模块对语音信号进行识别获得语音信号对应的语言信息，其中语言模型可以隐马尔科夫 HMM-GMM 或者深度神经网络模型。

- 20 步骤 3022：对所述视频信号进行分帧处理，获得分帧后的每帧图像中的嘴唇位置；

步骤 3023：服务端对视频信号进行分帧，捕获分帧后的每帧图像中嘴唇位置，调用唇语识别模块，获得分帧后的每帧图像中的嘴唇位置。

- 25 对捕获分帧后的每帧图像中嘴唇位置和语音信号进行采样，采样精度可以为每 20ms 取一帧图像，并以 10ms 的窗长移动，保证语音信号和视频信号的同步采样；捕获分帧后的每帧图像中嘴唇位置方法有很多，例如跟踪每帧嘴唇位置之间设定关键点的位置变化情况，通过计算设定关键点之间的差值变化，得到每帧嘴唇图像中的嘴唇位置，本申请对此不做限制。



步骤 3024：对所述每帧图像中的嘴唇位置进行唇语识别，获得每帧图像的唇语对应的语言信息。

使用预先训练的唇语语言模块对唇语进行识别获得每帧图像的唇语对应的语言信息。

5 步骤 3025：使用时间对准算法计算语音信号对应的语言信息与唇语识别对应的语言信息的相似度值，将所述相似度值作为活体检测结果。

参见图 5，示出了本申请语音信号对应的语言信息与唇语识别对应的语言信息的同步检测的示意图。

10 采用时间对准算法语音信号对应的语言信息与唇语识别对应的语言信息的同步检测，常采用的时间对准算法包括：动态时间规整（Dynamic Time Warping, DTW）、Viterbi 算法等。

15 现以当语音信号为数字串为例进行说明，当语音信号发出的数字串为 8579，将语音信号中对应的语音信息 8579 与唇语识别对应的语言信息进行比较，计算两组识别结果的相似度，需要说明的是，语音信号还可以为其他文本内容，对此本申请不做具体限制。

20 以动态时间规整算法为例，计算出两组识别结果序列的相似度。例如，对语音信号进行语音识别，获得的语音识别序列为： $S = s_1, s_2, \dots, s_K$ ； K 为语音信号的帧数，对视频信号进行唇语识别后，获得的唇语识别序列为： $V = v_1, v_2, \dots, v_N$ ； N 为唇语图像的采样帧数；在实际使用中， K 和 N 可相同也可不同，二者数值相差较小。

由于语言特征向量 s_k 和唇语识别序列 v_N 维度相同，因此，也可以采用欧式距离、余弦距离等来衡量二者的相似度，还可以采用动态规划算法完成序列 S 和 v 的相似度计算，动态规划算法的核心公式为：

$$r(i, j) = d(s_i, v_j) + \min\{r(i-1, j-1), r(i-1, j), r(i, j-1)\}$$

25 其中， $d(s_i, v_j)$ 为 s_i 和 v_j 当前距离；通过距离的不断累积， $r(i, j)$ 即为序列相似度值，其中 i, j 分别表示语音信号的帧位置和视频信号的帧位置。

步骤 303：若所述活体检测结果大于第一设定阈值，则执行步骤 304。

步骤 304：对获取使用者的语音信号和视频信号进行声纹识别和人脸识



别。

步骤 305：若声纹识别和人脸识别全部通过，则执行步骤 306，否则执行步骤 307。

步骤 304 包括以下子步骤：

5 步骤 3041：对使用者的语音信号进行声纹识别，得到语音信号的声纹识别打分；

步骤 3042：对使用者的人脸视频图像，进行分帧，获取人脸连续帧图像，对所述人脸连续帧图像进行人脸识别，对获得的人脸识别进行分数融合，得到人脸识别打分。

10 人脸识别包括人脸表情、人脸姿态、人脸五官变化等等；例如，人脸面部表情包括高兴、痛苦、悲伤、愤怒和面无表情等；人脸姿态包括正面、左侧面、右侧面、上侧面和下侧面；人脸五官变化包括张嘴、闭嘴、皱眉、眨眼、耸鼻等。

可以对采集到的人脸视频图像进行随机抽样，对抽样后的人脸视频图
15 样进行分数融合，得到人脸识别打分，其中，分数融合包括加权平均。

步骤 3043：根据所述声纹识别打分和人脸识别打分，获得使用者身份认证的判决得分。

$S'_v = 1/(1 + e^{S_v - R_v})$ (1) 其中， S'_v 代表声纹识别得分， S_v 代表声纹识别
打分， R_v 代表声纹识别错误概率所对应的数值。

20 $S'_f = 1/(1 + e^{S_f - R_f})$ (2) 其中， S'_f 代表人脸识别得分， S_f 代表声人脸识别打
分， R_f 代表人脸识别错误概率所对应的数值。

根据公式 (1) 和 (2) 可以得到身份认证的判决得分 S ，

$S = \alpha \times S'_v + (1 - \alpha) \times S'_f$ 其中， S'_v 代表声纹识别得分， S'_f 代表人脸识别得
分， S 代表身份认证的判决得分， α 代表插值加权系数，一般取 (0, 1) 之间
25 的一个数值。

步骤 3044：若所述判决得分大于第二设定阈值，则使用者身份认证成功。

使用者的身份认证成功，说明使用者的声纹和人脸均来自于使用者本



人反之，认证不通过，则说明使用者的声纹和人脸均不是来自于使用者本人。

步骤 3045：若所述判决得分小于第二设定阈值，则使用者身份认证失败。

5 步骤 306：身份认证通过。

步骤 307：身份认证失败。

本实施例，通过将获取的语音信号和视频信号先进行活体检测，当活体检测结果小于等于第一设定阈值时，则身份认证失败，当活体检测结果大于第一设定阈值，则继续对获取的语音信号和视频信号进行声纹识别和人脸认识，若声纹识别和人脸识别全部通过，则该使用者的身份认证通过，否则认证失败，通过对获取的语音信号和视频信号进行活体检测、声纹识别和人脸识别，从而极大的阻止了非法用户的闯入，通过对使用者的语音信号和视频信号进行多重检测，提高了身份认证的准确性、安全性和可靠性，从而解决了现有技术单一生物特征识别技术不足的问题。

15 实施例四

参照图 6，示出了本申请一种基于声纹识别、人脸识别以及活体检测的身份认证方法，该方法包括：

步骤 601：使用采集器采集使用者的语音信号和视频信号。

20 步骤 602：对所述语音信号和视频信号进行活体检测和人脸识别，获得综合检测结果。

步骤 603：若所述综合检测结果大于第三设定阈值，则执行步骤 604，若所述活体检测结果小于等于第三设定阈值，则执行步骤 607。

其中，第三设定阈值的设定可以由本领域技术人员采用任意适当方式进行设定，如可以采用人工经验设定阈值，或者针对历史数据的差异值设定阈值，本申请对此不作限制。

25 步骤 604：对获取使用者的语音信号和视频信号进行声纹识别和人脸识别。

步骤 605：若声纹识别和人脸识别是否全部通过，若是，则执行步骤 606，



若否，则执行步骤 607。

步骤 606：身份认证通过。

步骤 607：使用者的身份认证失败。

本申请将获取的语音信号和视频信号先进行活体检测和人脸检测，当
5 活体检测结果大于第一设定阈值，则继续对获取的语音信号和视频信号进行
声纹识别和人脸识别，若声纹识别和人脸识别全部通过，则该使用者的
身份认证通过，否则认证失败，通过对获取的语音信号和视频信号进行活
体检测、声纹识别和人脸识别，从而极大的阻止了非法用户的闯入，通过
10 对使用者的语音信号和视频信号进行多重检测，从而解决了现有技术单一
生物特征识别技术的不足，进而提高了身份认证的准确性、安全性和可靠
性。

基于上述方法实施例的说明，本申请还提供了相应的一种基于声纹识
别、人脸识别以及活体检测的身份认证系统的实施例，来实现上述方法实
施例所述的内容。

15 实施例五

参见图 7，示出了本申请实施例五中的一种基于声纹识别、人脸识别以
及活体检测的身份认证系统的结构框图，具体可以包括：获取模块 701，用
于获取使用者的语音信号和视频信号。

活体检测模块 702，用于对所述语音信号和视频信号进行活体检测，获
20 得活体检测结果。

第一判断模块 703，用于若所述活体检测结果大于第一设定阈值，则对
获取使用者的语音信号和视频信号进行声纹识别和人脸识别；若声纹识别
和人脸识别全部通过，则身份认证通过，否则身份认证失败。

若所述活体检测结果小于等于第一设定阈值，则使用者身份认证失
25 败。

优选地，所述活体检测包括语音识别和唇语识别。

优选地，活体检测模块包括：语音识别模块，用于对所述语音信号进
行语音识别，获得语音信号对应的语言信息；



分帧模块，用于对所述视频信号进行分帧处理，获得分帧后的每帧图像中的嘴唇位置；

唇语识别模块，用于对所述每帧图像中的嘴唇位置进行唇语识别，获得每帧图像的唇语识别对应的语言信息；

5 计算模块，用于使用时间对准算法计算语音信号对应的语言信息与唇语识别对应的语言信息的相似度值，将所述相似度值作为活体检测结果。

优选地，判断模块包括：声纹识别模块，用于对使用者的语音信号进行声纹识别，得到语音信号的声纹识别打分；

人脸识别模块，用于对使用者的人脸视频图像，进行分帧，获取人脸连续帧图像，对所述人脸连续帧图像进行人脸识别，对获得的人脸识别进行分数融合，得到人脸识别打分；

判决模块，用于根据所述声纹识别打分和人脸识别打分，获得使用者身份认证的判决得分；

15 第二判断模块，用于若所述判决得分大于第二设定阈值，则使用者身份认证成功；

若所述判决得分小于第二设定阈值，则使用者身份认证失败。

优选地，还包括：使用视频图像采集器获取使用者视频信号，使用语音采集器获取使用者的语音信号。

本实施例，通过将获取的语音信号和视频信号先进行活体检测，当活体检测结果小于等于第一设定阈值时，则身份认证失败，当活体检测结果大于第一设定阈值，则继续对获取的语音信号和视频信号进行声纹识别和人脸识别，若声纹识别和人脸识别全部通过，则该使用者的身份认证通过，否则认证失败，通过对获取的语音信号和视频信号进行活体检测、声纹识别和人脸识别，从而极大的阻止了非法用户的闯入，通过对使用者的语音信号和视频信号进行多重检测，从而解决了现有技术单一生物特征识别技术的不足，进而提高了身份认证的准确性、安全性和可靠性。



对于系统实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相
5 参见即可。

以上对本申请所提供的一种基于声纹识别、人脸识别以及活体检测的身份认证方法和系统，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。
10



说明书附图

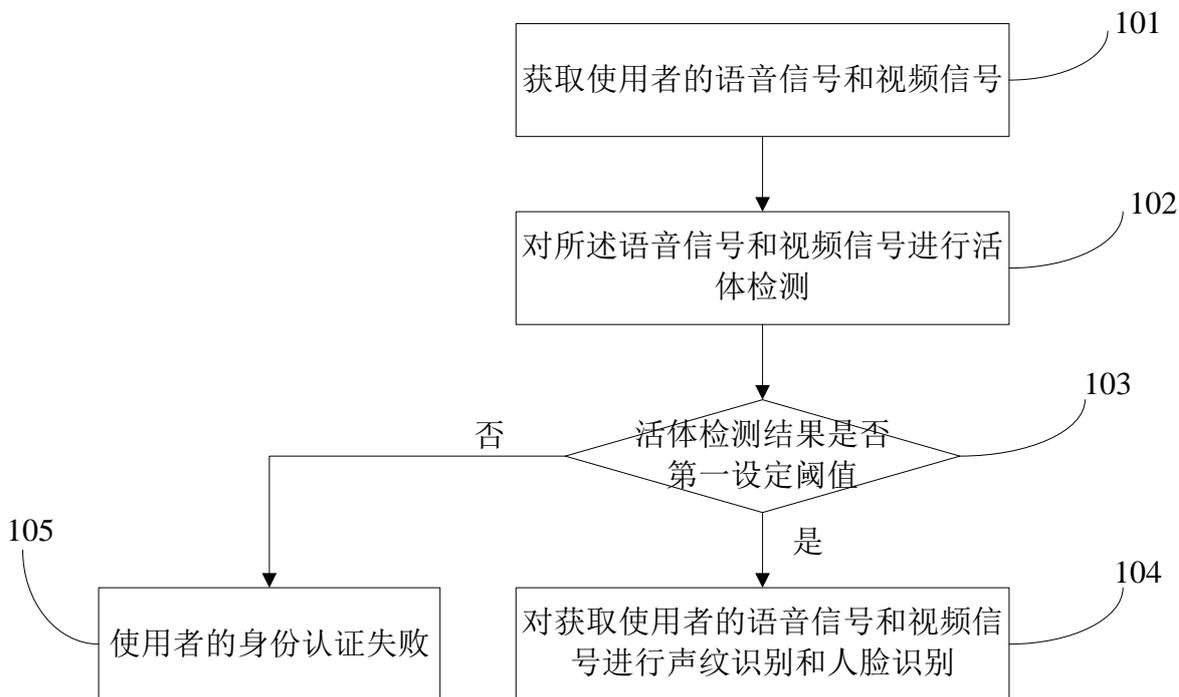


图 1

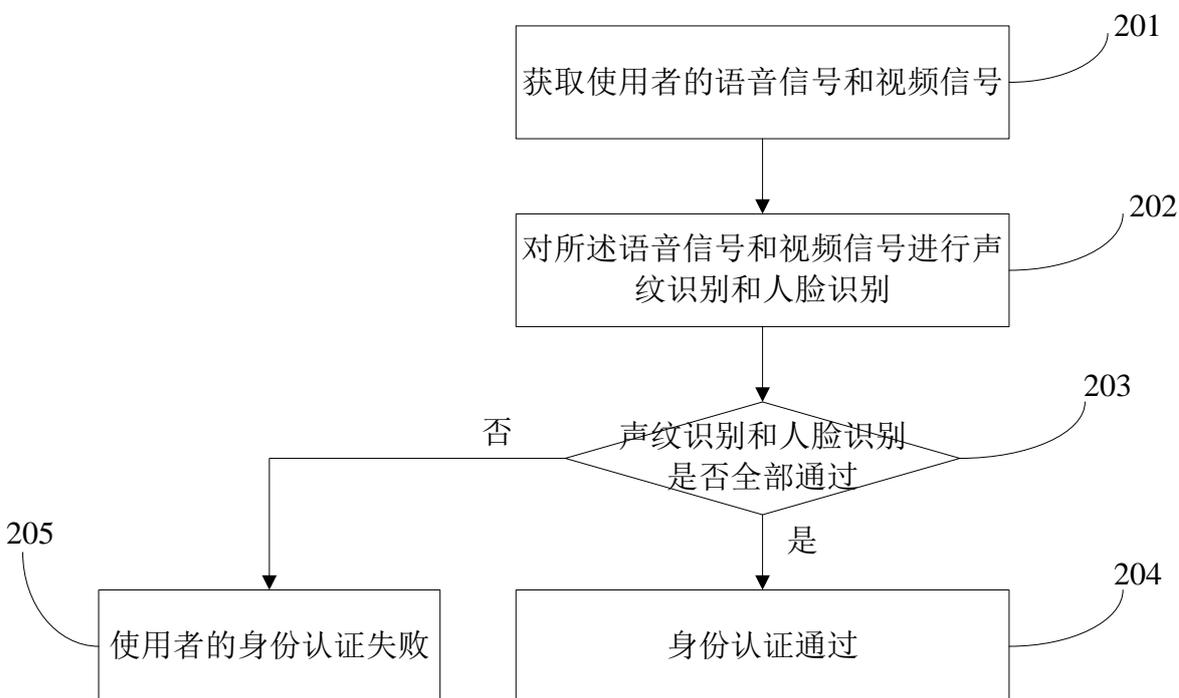


图 2



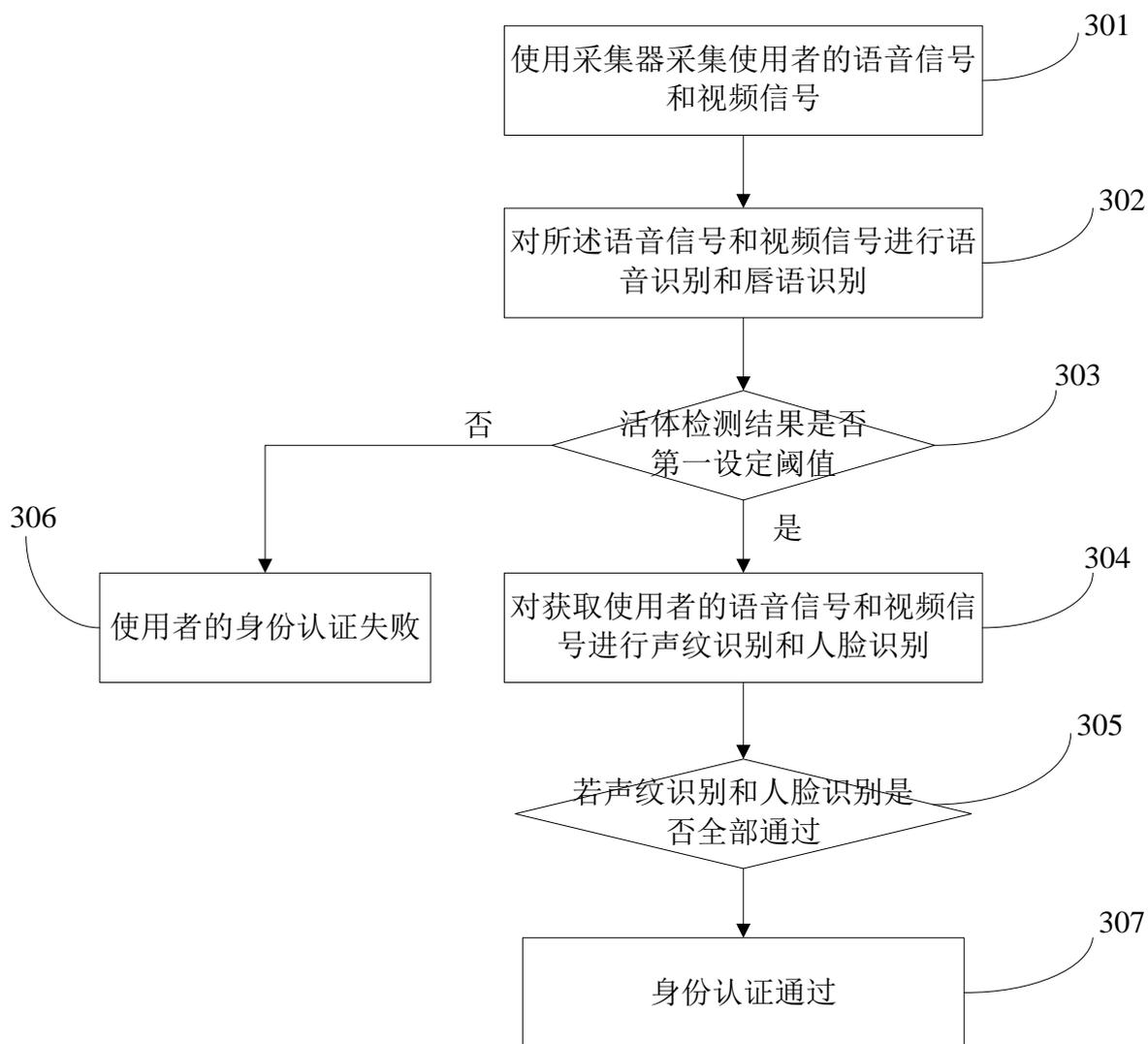


图 3



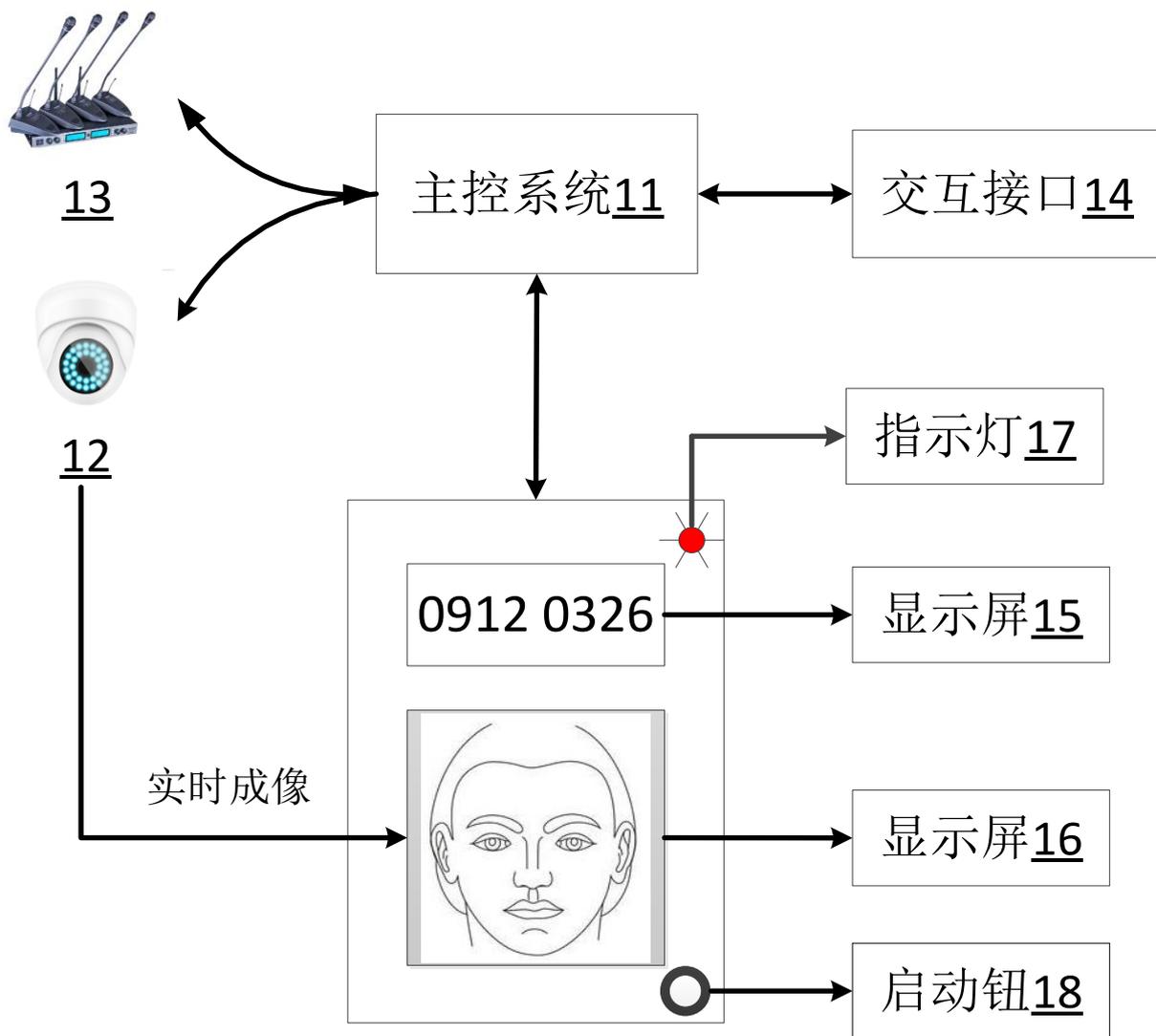


图 4

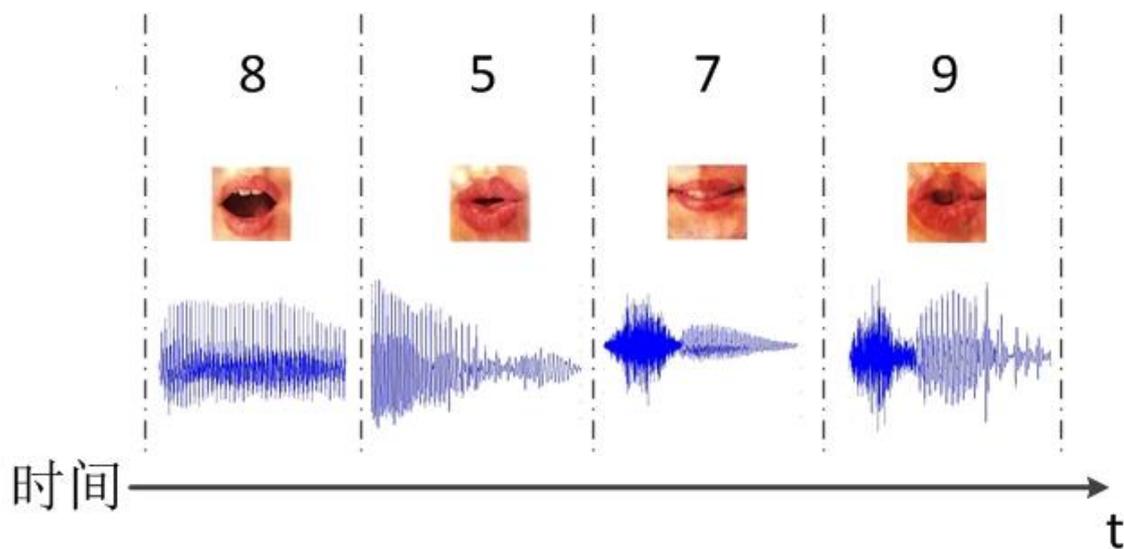


图 5



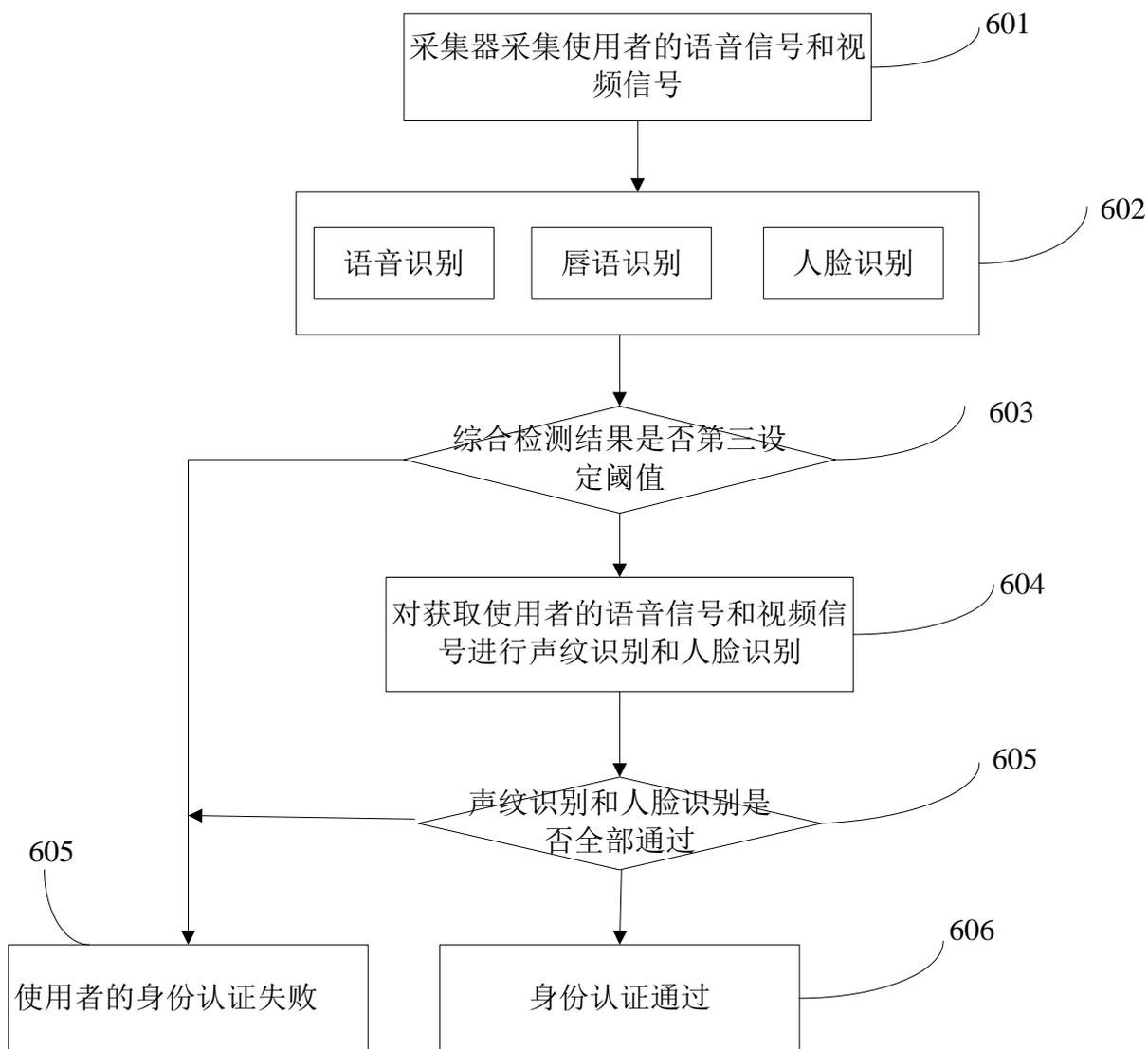


图 6

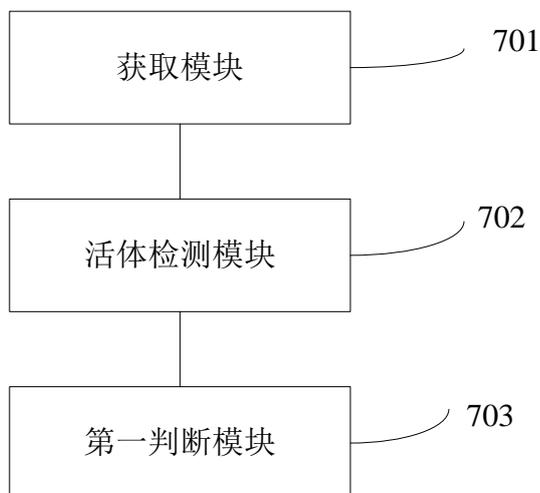


图 7

