



## (12) 发明专利申请

(10) 申请公布号 CN 106650555 A

(43) 申请公布日 2017. 05. 10

(21) 申请号 201510731925. 8

(22) 申请日 2015. 11. 02

(71) 申请人 苏宁云商集团股份有限公司

地址 210042 江苏省南京市玄武区苏宁大道  
1 号苏宁总部

(72) 发明人 蒋明 莫全益 徐瑾 王匡宇  
何蔼 乐正宜

(74) 专利代理机构 江苏圣典律师事务所 32237  
代理人 许峰

(51) Int. Cl.  
G06K 9/00(2006. 01)

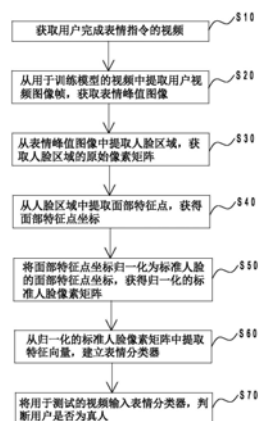
权利要求书2页 说明书9页 附图4页

### (54) 发明名称

一种基于机器学习的真人验证方法及系统

### (57) 摘要

本发明公开了一种基于机器学习的真人验证方法,包括:S10 获取用户完成表情指令的视频;S20 从视频中提取用户视频图像帧,获取表情峰值图像;S30 从表情峰值图像中提取人脸区域,获取人脸区域的原始像素矩阵;S40 从人脸区域中提取面部特征点,获得面部特征点坐标;S50 将面部特征点坐标归一化为标准人脸的面部特征点坐标,获得归一化的标准人脸像素矩阵;S60 从归一化的标准人脸像素矩阵中提取特征向量,建立表情分类器;S70 将用于测试的视频输入表情分类器,判断用户是否为真人。还公开了基于机器学习的真人验证系统。该真人验证方法及系统利用视频技术,识别用户面部表情动作,提高真人识别率。



1. 一种基于机器学习的真人验证方法,其特征在于,该验证方法包括以下步骤:

S10 获取用户完成表情指令的视频,视频包括用于测试的视频和用于训练模型的视频;

S20 从用于训练模型的视频中提取用户视频图像帧,获取表情峰值图像;

S30 从 S20 获取的表情峰值图像中提取人脸区域,获取人脸区域的原始像素矩阵;

S40 从人脸区域中提取面部特征点,获得面部特征点坐标;

S50 将 S40 得到的面部特征点坐标归一化为标准人脸的面部特征点坐标,获得归一化的标准人脸像素矩阵;

S60 从归一化的标准人脸像素矩阵中提取特征向量,建立表情分类器;

S70 将 S10 中用于测试的视频输入表情分类器,判断用户是否为真人。

2. 按照权利要求 1 所述的基于机器学习的真人验证方法,其特征在于,所述的步骤 S20 具体包括:从用于训练模型的视频中逐帧提取图像,选取与中立表情图像区别最大的图像作为表情峰值图像。

3. 按照权利要求 1 所述的基于机器学习的真人验证方法,其特征在于,所述的步骤 S30 具体包括:对 S20 获取的表情峰值图像,利用人脸探测器检测出人脸区域,并输出人脸区域的原始像素矩阵。

4. 按照权利要求 1 所述的基于机器学习的真人验证方法,其特征在于,所述的 S50 具体包括:

S501 将 S40 提取的面部特征点映射到标准人脸的特征点上,并计算线性仿射变换;

S502 将步骤 501 得到的线性仿射变换应用到 S30 获取的人脸区域的原始像素矩阵中,获得归一化的标准人脸像素矩阵。

5. 按照权利要求 1 所述的基于机器学习的真人验证方法,其特征在于,所述的 S60 具体包括:

S601 提取 S50 获得的标准人脸的面部特征点,构造特征区域,并从特征区域中选择最有区分度的特征区域;

S602 将最有区分度的特征区域中的每个像素点作为一个特征向量,对最有区分度的特征区域中所有像素点构成全部的特征向量进行机器学习,建立表情分类器。

6. 按照权利要求 1 所述的基于机器学习的真人验证方法,其特征在于,所述的 S70 具体包括:将 S10 中用于测试的视频输入表情分类器,依据式 (1) 判断用户是否为真人。

$$s = \frac{w_1}{w} \quad \text{式 (1)}$$

其中,  $s$  表示用于测试的视频中,用户成功完成表情帧数百分比,  $w_1$  为单位时间内用户成功完成表情帧数,  $w$  表示单位时间内视频总帧数;当  $s \geq s_0$  时,用户为真人;当  $s < s_0$  时,用户不是真人;  $s_0$  表示判断阈值。

7. 一种基于机器学习的真人验证系统,其特征在于,该验证系统包括:

第一获取模块:用于获取用户完成表情指令的视频;

第二获取模块:用于从用于训练模型的视频中提取用户视频图像帧,获取表情峰值图像;

第三获取模块:用于从表情峰值图像中提取人脸区域,获取人脸区域的原始像素矩

阵；

第四获取模块：用于从人脸区域中提取面部特征点，获得面部特征点坐标；

第五获取模块：用于将面部特征点坐标归一化为标准人脸的面部特征点坐标，获得归一化的标准人脸像素矩阵；

建立模块：用于从归一化的标准人脸像素矩阵中提取特征向量，建立表情分类器；

判断模块：用于将用于测试的视频输入表情分类器，判断用户是否为真人。

8. 按照权利要求 7 所述的基于机器学习的真人验证系统，其特征在于，所述的第五获取模块包括：

计算单元：用于将面部特征点映射到标准人脸的特征点上，并计算线性仿射变换；

应用单元：用于将线性仿射变换应用到 S30 获取的人脸区域的原始像素矩阵中，获得归一化的标准人脸像素矩阵。

9. 按照权利要求 7 所述的基于机器学习的真人验证系统，其特征在于，所述的建立模块包括：

构造单元：用于提取标准人脸的面部特征点，构造特征区域，并从特征区域中选择最有区分度的特征区域；

建立单元：用于将最有区分度的特征区域中的每个像素点作为一个特征向量，对最有区分度的特征区域中所有像素点构成全部的特征向量进行机器学习，建立表情分类器。

10. 按照权利要求 7 所述的基于机器学习的真人验证方法，其特征在于，所述的判断模块依据式 (1) 判断用户是否为真人。

$$s = \frac{w_1}{w} \quad \text{式 (1)}$$

其中， $s$  表示用于测试的视频中，用户成功完成表情帧数百分比， $w_1$  为单位时间内用户成功完成表情帧数， $w$  表示单位时间内视频总帧数；当  $s \geq s_0$  时，用户为真人；当  $s < s_0$  时，用户不是真人； $s_0$  表示判断阈值。

## 一种基于机器学习的真人验证方法及系统

### 技术领域

[0001] 本发明涉及一种真人验证方法及系统,具体来说,涉及一种基于机器学习的真人验证方法及系统。

### 背景技术

[0002] 传统金融机构提供业务一般使用营业厅开户验证以防止身份欺诈,费时费力,且受地域限制不易拓展业务。

[0003] 目前开户审批等大多数业务场景仍采用人工比对,用户上传的手持身份证照片和公安证件照片,业务量庞大,亟需自动化系统。

[0004] 电商,基金公司、农信社、银行远程真人识别开户时用户开启设备摄像头,并需要通过上传证件照验证客户身份。

[0005] 在如上场景中,上传的身份证等证件照容易伪造,证件和人照片影像方式不能确保是否有录像,照片 PS,假人偶操作等。

### 发明内容

[0006] 技术问题:本发明所要解决的技术问题是:提供一种基于机器学习的真人验证方法及系统,利用视频技术,识别用户面部表情动作,提高真人识别率。

[0007] 技术方案:为解决上述技术问题,本发明实施例采用如下技术方案:

[0008] 第一方面,本实施例提供一种基于机器学习的真人验证方法,该验证方法包括以下步骤:

[0009] S10 获取用户完成表情指令的视频,视频包括用于测试的视频和用于训练模型的视频;

[0010] S20 从用于训练模型的视频中提取用户视频图像帧,获取表情峰值图像;

[0011] S30 从 S20 获取的表情峰值图像中提取人脸区域,获取人脸区域的原始像素矩阵;

[0012] S40 从人脸区域中提取面部特征点,获得面部特征点坐标;

[0013] S50 将 S40 得到的面部特征点坐标归一化为标准人脸的面部特征点坐标,获得归一化的标准人脸像素矩阵;

[0014] S60 从归一化的标准人脸像素矩阵中提取特征向量,建立表情分类器;

[0015] S70 将 S10 中用于测试的视频输入表情分类器,判断用户是否为真人。

[0016] 对于第一方面,作为第一种可能实施的方式,所述的步骤 S20 具体包括:从用于训练模型的视频中逐帧提取图像,选取与中立表情图像区别最大的图像作为表情峰值图像。

[0017] 对于第一方面,作为第二种可能实施的方式,所述的步骤 S30 具体包括:对 S20 获取的表情峰值图像,利用人脸探测器检测出人脸区域,并输出人脸区域的原始像素矩阵。

[0018] 对于第一方面,作为第三种可能实施的方式,所述的 S50 具体包括:

[0019] S501 将 S40 提取的面部特征点映射到标准人脸的特征点上,并计算线性仿射变换;

[0020] S502 将步骤 501 得到的线性仿射变换应用到 S30 获取的人脸区域的原始像素矩阵中,获得归一化的标准人脸像素矩阵。

[0021] 对于第一方面,作为第四种可能实施的方式,所述的 S60 具体包括:

[0022] S601 提取 S50 获得的标准人脸的面部特征点,构造特征区域,并从特征区域中选择最有区分度的特征区域;

[0023] S602 将最有区分度的特征区域中的每个像素点作为一个特征向量,对最有区分度的特征区域中所有像素点构成全部的特征向量进行机器学习,建立表情分类器。

[0024] 对于第一方面,作为第五种可能实施的方式,所述的 S70 具体包括:将 S10 中用于测试的视频输入表情分类器,依据式 (1) 判断用户是否为真人。

[0025]  $s = \frac{w_1}{w}$  式 (1)

[0026] 其中, s 表示用于测试的视频中,用户成功完成表情帧数百分比,  $w_1$  为单位时间内用户成功完成表情帧数, w 表示单位时间内视频总帧数;当  $s \geq s_0$  时,用户为真人;当  $s < s_0$  时,用户不是真人;  $s_0$  表示判断阈值。

[0027] 第二方面,本实施例提供一种基于机器学习的真人验证系统,该验证系统包括:

[0028] 第一获取模块:用于获取用户完成表情指令的视频;

[0029] 第二获取模块:用于从用于训练模型的视频中提取用户视频图像帧,获取表情峰值图像;

[0030] 第三获取模块:用于从表情峰值图像中提取人脸区域,获取人脸区域的原始像素矩阵;

[0031] 第四获取模块:用于从人脸区域中提取面部特征点,获得面部特征点坐标;

[0032] 第五获取模块:用于将面部特征点坐标归一化为标准人脸的面部特征点坐标,获得归一化的标准人脸像素矩阵;

[0033] 建立模块:用于从归一化的标准人脸像素矩阵中提取特征向量,建立表情分类器;

[0034] 判断模块:用于将用于测试的视频输入表情分类器,判断用户是否为真人。

[0035] 对于第二方面,作为第一种可能实施的方式,所述的第五获取模块包括:

[0036] 计算单元:用于将面部特征点映射到标准人脸的特征点上,并计算线性仿射变换;

[0037] 应用单元:用于将线性仿射变换应用到 S30 获取的人脸区域的原始像素矩阵中,获得归一化的标准人脸像素矩阵。

[0038] 对于第二方面,作为第二种可能实施的方式,所述的建立模块包括:

[0039] 构造单元:用于提取标准人脸的面部特征点,构造特征区域,并从特征区域中选择最有区分度的特征区域;

[0040] 建立单元:用于将最有区分度的特征区域中的每个像素点作为一个特征向量,对最有区分度的特征区域中所有像素点构成全部的特征向量进行机器学习,建立表情分类器。

[0041] 对于第二方面,作为第三种可能实施的方式,所述的判断模块依据式 (1) 判断用户是否为真人。

[0042]  $s = \frac{w_1}{w}$  式 (1)

[0043] 其中,  $s$  表示用于测试的视频中, 用户成功完成表情帧数百分比,  $w_1$  为单位时间内用户成功完成表情帧数,  $w$  表示单位时间内视频总帧数; 当  $s \geq s_0$  时, 用户为真人; 当  $s < s_0$  时, 用户不是真人;  $s_0$  表示判断阈值。

[0044] 有益效果: 与现有技术相比, 本发明实施例具有以下有益效果:

[0045] 本发明实施例首先建立各种表情的分类器模型, 然后向用户发出表情指令, 根据用户完成的表情视频, 验证用户是否为真人。本发明实施例采用动态视频, 而非静态图像验证用户是否为真人。静态图像易于伪造, 动态视频需用户在指定时间内成功完成随机指定的任务, 即时互动增加伪造难度。本发明实施例能够提高真人识别率。

## 附图说明

[0046] 图 1 为本发明实施例真人验证方法的流程框图;

[0047] 图 2 为本发明实施例真人验证方法中步骤 S50 的流程框图;

[0048] 图 3 为本发明实施例真人验证方法中步骤 S60 的流程框图;

[0049] 图 4 为本发明实施例中人脸特征点的布局示意图;

[0050] 图 5 为本发明实施例真人验证系统的结构框图;

[0051] 图 6 为本发明实施例真人验证系统中第五获取模块的结构框图;

[0052] 图 7 为本发明实施例真人验证系统中建立模块的结构框图。

## 具体实施方式

[0053] 下面结合附图, 对本发明实施例进行详细的说明。

[0054] 如图 1 所示, 本实施例提供一种基于机器学习的真人验证方法, 该验证方法包括以下步骤:

[0055] S10 获取用户完成表情指令的视频, 视频包括用于测试的视频和用于训练模型的视频;

[0056] S20 从用于训练模型的视频中提取用户视频图像帧, 获取表情峰值图像;

[0057] S30 从 S20 获取的表情峰值图像中提取人脸区域, 获取人脸区域的原始像素矩阵;

[0058] S40 从人脸区域中提取面部特征点, 获得面部特征点坐标;

[0059] S50 将 S40 得到的面部特征点坐标归一化为标准人脸的面部特征点坐标, 获得归一化的标准人脸像素矩阵;

[0060] S60 从归一化的标准人脸像素矩阵中提取特征向量, 建立表情分类器;

[0061] S70 将 S10 中用于测试的视频输入表情分类器, 判断用户是否为真人。

[0062] 上述实施例首先通过步骤 S10—步骤 S60 建立各种表情的分类器模型, 然后将用户完成的用于测试的视频输入分类器模型中, 从而验证用户是否为真人。分类器模型验证用户时, 是从用户完成表情指令的视频中提取各种表情图像帧。视频是动态连续的图像。本实施例从动态视频中提取用户的表情图像帧, 用于建立模型和进行真伪分辨, 提高了验证用户真伪的准确性。该实施例采用动态视频, 而非静态图像验证用户是否为真人。静态图像易于伪造, 动态视频需用户在指定时间内成功完成随机指定的任务, 即时互动增加伪造

难度。本发明实施例通过使用一系列随机生成的面部表情指令,提高伪造难度。本发明实施例能够提高真人识别率。

[0063] 在 S10 中,系统向用户发出各种表情指令,用户根据表情指令作出各种相应的表情。用户作出各种表情的过程,系统将其拍摄成视频。视频中含有中立表情图像。中立表情是指用户不做任何表情时的自然表情。中立表情图像是从视频中提取的。作为优选,系统发出的第一个表情指令可以是要求用户做出不带表情的指令。这样便于后续表情峰值图像的比较与提取。

[0064] 用户完成表情指令的视频,包括用于测试的视频和用于训练模型的视频。用于测试的视频是用于分辨用户是否为真人。用于训练模型的视频是用于对建立的表情分类器模型进行训练。视频中的所有图像帧分为不重合的两部分,一部分作为训练数据的图像帧,另一部分作为测试数据的图像帧。

[0065] 在 S20 中,从用于训练模型的视频中逐帧提取图像,选取与中立表情区别最大的图像作为表情峰值图像。表情峰值图像是与中立表情的五官位移最大的一帧图像,比如:笑的最灿烂的一帧,眨眼眨的最厉害的一帧。测试时,不取峰值帧,而是对每一帧做处理。训练模型时,优选表情峰值图片。但是训练模型时,也可以不取表情峰值图片,而选取任何一张响应的指令表情图片,比如选取响应时间范围 1.5-2.5 秒内中点第 2 秒的图片帧,但是这样与中立表情图片区别不太大的话,训练出来的模型效果就会差一些。

[0066] 作为优选方案,在 S30 中,对 S20 获取的表情峰值图像,利用人脸探测器检测出人脸区域,并输出人脸区域的原始像素矩阵。从图像帧中检测出人脸的区域采用开源软件包中的人脸探测器。例如,利用开源软件 openCV 开源包中的 Viola-Jones 的人脸探测器。人脸探测器输出的是一个像素矩阵。该像素矩阵显示人脸位置在整个原图像中的坐标。当然还可以采用其他开源软件中的人脸探测器。

[0067] S40 中,使用开源工具从人脸区域中提取面部特征点。例如,CSIRO 软件开发工具包 CSIRO sdk 从人脸区域中提取面部特征点。预先可以在开源工具中设置提取的面部特征点的数量。例如,图 4 为一人脸区域,该人脸区域中设定了 66 个面部特征点。为更加准确的反应脸部表情,面部特征点主要涉及五官和脸的轮廓。开源工具输出面部特征点的坐标。坐标系的原点是原图像的左下角,x 轴是图像下边缘,方向向右为正方向;y 轴是图像下边缘,方向向上为正方向。

[0068] 作为优选方案,如图 2 所示,所述的 S50 具体包括:

[0069] S501 将 S40 提取的面部特征点映射到标准人脸的特征点上,并计算线性仿射变换;

[0070] S502 将步骤 501 得到的线性仿射变换应用到 S30 获取的人脸区域的原始像素矩阵中,获得归一化的标准人脸像素矩阵。

[0071] 所述的 S50 的几何校准,可以降低训练噪音,提高模型训练效果。在 S501 中,利用训练数据上训练模型,然后把模型应用到测试数据上看效果。训练模型采用表情峰值图像帧。测试模型效果时,是每一帧图片都获取设定数量的面部特征点。如图 4 所示,例如 66 个。将获取的一图像帧中的相关面部特征点映射到标准人脸的对应面部特征点,计算线性仿射变换。标准人脸适用于所有人。例如,相关面部特征点采用眼睛特征点中心(a, b)。选择两眼中心坐标(c, d)来确定标准人脸。两眼距离及两眼中心坐标(c, d)可以预先设定,

适用于每个人。当标准人脸以两个眼睛的坐标为标准时,则将 S30 得到的两个眼睛的中心坐标  $(a, b)$ ,映射到标准人脸的到的两个眼睛的中心  $(c, d)$ ,据此构造线性映射。其中, $a$  表示一只眼睛中含有的面部特征点的坐标均值,如图 4 中编号 36—41 的面部特征点的坐标均值, $b$  表示另一只眼睛中含有的面部特征点的坐标均值,如图 4 中编号 42—47 的面部特征点的坐标均值。 $c$  表示标准人脸中一只眼睛中心的坐标, $d$  表示标准人脸中另一只眼睛中心的坐标。

[0072] 计算线性仿射变换的过程为现有技术,例如在下述网址中公开:[http://baike.baidu.com/link?url=WW\\_iy9f2u1\\_FkGqEz82tnfQz2TB71ls\\_K702PoCplw0XxHP4axL\\_VaQ0dYpnmcvpbvRWBcpK6Zhsc19JWGxR3oa](http://baike.baidu.com/link?url=WW_iy9f2u1_FkGqEz82tnfQz2TB71ls_K702PoCplw0XxHP4axL_VaQ0dYpnmcvpbvRWBcpK6Zhsc19JWGxR3oa)。仿射变换可以通过一系列的原子变换的复合来实现,包括:平移(Translation)、缩放(Scale)、翻转(Flip)、旋转(Rotation)和错切(Shear)。此类变换可以用一个  $3 \times 3$  的矩阵来表示,其最后一行为  $(0, 0, 1)$ 。该变换矩阵将原坐标  $(a', b')$  变换为新坐标  $(c', d')$ 。原坐标和新坐标皆视为最末一行为  $(1)$  的三维列向量,原列向量左乘变换矩阵得到新的列向量。

[0073] S502 中,步骤 40 中每个面部特征点坐标会一一对应一个像素色彩  $(g, b, r)$ 。将线性仿射变换应用到步骤 40 中人脸区域中每个面部特征点的坐标上,得到人脸区域面部特征点的新坐标,但每个面部特征点对应的像素色彩不变,获得归一化的标准人脸像素矩阵。将线性仿射变换应用到所有图片的原始像素矩阵上就是归一化的过程。

[0074] 作为优选方案,如图 3 所示,所述的 S60 具体包括:

[0075] S601 提取 S50 获得的标准人脸的面部特征点,构造特征区域,并从特征区域中选择最有区分度的特征区域;

[0076] S602 将最有区分度的特征区域中的每个像素点作为一个特征向量,对最有区分度的特征区域中所有像素点构成全部的特征向量进行机器学习,建立表情分类器。

[0077] 在 S601 中,除了利用规则多边形区域(例如三角形)提取标准人脸的面部特征点,还可以利用不规则多边形(例如,长方形、四条边以上的其他多边形)提取标准人脸的面部特征点。利用不规则多边形提取标准人脸的特征点,可以增加备选特征,捕捉更多相关脸部特征细节以供机器学习。传统提取方法使用三角形,特征点组合得到的特征数量有限。而使用不规则多边形后,特征点组合方式有多种,供选特征数量成倍增加,弥补了传统方法中可能丢失的重要特征。提取的这些面部特征点供输入机器学习模型使用。如果只取眼部特征,那可能只能提取 100 个特征,但是如果全脸范围提取特征,那能提取百万特征,从而不丢失关键区分度高的特征。本实施例通过全脸范围提取特征,安全性高于单区域(眨眼)特征。

[0078] 任意 3 个及以上特征点组合,可以构成一个特征区域。每个特征区域包括不同的特征点组合,包括五官特征、德朗奈三角区域等等。

[0079] 采用统计模型中逐步向前选择的方法从特征区域中选择最有区分度的特征区域。最有区分度的特征区域是相对可以显著区分中立表情和指令表情的区域变量。例如,需要区分中立表情(如  $y_0$ )和指令表情(如  $y_1$ ),而有的特征区域的取值(如  $x_1$ )在中立表情和指令表情中表现类似,不易区分中立表情和指令表情,但是有的特征区域的取值(如  $x_3$ )在中立表情和指令表情中表现差异极大,容易区分中立表情和指令表情。这样,选择  $x_3$  作为最有区分度的特征区域的区域变量。



[0080] 在 S602 中,最有区分度的特征区域中所有像素点构成全部的特征向量。例如,提取 S50 获得的标准人脸的 66 个特征点,通过构造多边形区域,比如任意 5 个点构造一个五边形,得到比如一百万个区域,通过逐步向前选择的方法从特征区域中选择最有区分度的特征区域,例如选出 1000 个区域。1000 个区域内每个像素点作为一个特征向量,1000 个区域内的特征向量合起来得到所有的特征向量。使用机器学习算法训练分类器。

[0081] 作为优选,采用下述三种特征向量作为备选的特征向量:(1) 将最有区分度的特征区域内所有的直方图均衡化后的区域像素灰度作为第一类高维特征向量;(2) 局部二值模式 (LBP) 提取特征向量;(3) 面部关键点的相对几何位置:将最有区分度的德朗奈三角区域的顶点坐标,边长作为第三类特征向量。通过以上 3 类特征作为备选的特征向量,进行机器学习。

[0082] 作为优选例,选择 6 个分类器。每个分类器对应一种表情,用以检测相对于中立表情的 6 类表情:微笑,撇嘴,吐舌头,张嘴,闭左眼,闭右眼。当然用本实施例的方法可扩展至更多表情。分类器是不同的机器学习模型,例如采用 svm 模型。

[0083] 作为优选方案,所述的 S70 具体包括:将 S10 中用于测试的视频输入表情分类器,依据式 (1) 判断用户是否为真人。

$$[0084] \quad s = \frac{w_1}{w} \quad \text{式 (1)}$$

[0085] 其中,  $s$  表示用于测试的视频中,用户成功完成表情帧数百分比,  $w_1$  为单位时间内用户成功完成表情帧数,  $w$  表示单位时间内视频总帧数;当  $s \geq s_0$  时,用户为真人;当  $s < s_0$  时,用户不是真人;  $s_0$  表示判断阈值。例如,1 分钟内 100 帧里有 70 帧被分类器模型判断为成功完成指令,30 帧被模型判断为指令完成失败,那么这个指令得分  $70/100 = 0.7$ 。如果阈值为 0.6,则判定用户为真人。

[0086] 计算所有指令的平均分数作为最终分数,用判断阈值  $s_0$  来界定真人与否。判断阈值  $s_0$  的值没有特殊的要求。判断阈值  $s_0$  是可调整的,判断阈值  $s_0$  设的高,用户就难以通过测试;判断阈值  $s_0$  设的低,用户就容易通过测试。判断阈值  $s_0$  可以用于用户体验与安全性的折衷。

[0087] 本发明实施例要求用户在电脑/手机前录制视频,跟随一系列随机生成的指令做出相应面部表情,并且获得一定分数以通过验证测试。由于有多种方式能够伪造用户的人脸信息(照片,录像,玩偶),所以判断登录对象是否是真人,对人脸识别系统的安全性很重要。本实施例采取在实时系统交互中,根据用户是否能做出系统指令的表情来判断人脸信息的真实性,防止用户隐藏真实身份。

[0088] 如图 5 所示,本实施例提供一种基于机器学习的真人验证系统,该验证系统包括:

[0089] 第一获取模块:用于获取用户完成表情指令的视频;

[0090] 第二获取模块:用于从用于训练模型的视频中提取用户视频图像帧,获取表情峰值图像;

[0091] 第三获取模块:用于从表情峰值图像中提取人脸区域,获取人脸区域的原始像素矩阵;

[0092] 第四获取模块:用于从人脸区域中提取面部特征点,获得面部特征点坐标;

[0093] 第五获取模块:用于将面部特征点坐标归一化为标准人脸的面部特征点坐标,获

得归一化的标准人脸像素矩阵；

[0094] 建立模块：用于从归一化的标准人脸像素矩阵中提取特征向量，建立表情分类器；

[0095] 判断模块：用于将用于测试的视频输入表情分类器，判断用户是否为真人。

[0096] 上述实施例通过设置第一获取模块、第二获取模块、第三获取模块、第四获取模块、第五获取模块和建立模块，建立各种表情的分类器模型。然后通过判断模块，将用户完成的用于测试的视频输入分类器模型中，从而验证用户是否为真人。分类器模型验证用户时，是从用户完成表情指令的视频中提取各种表情图像帧。本实施例从动态视频中提取用户的表情图像帧，用于建立模型和进行真伪分辨，提高了验证用户真伪的准确性。该实施例第一获取模块采集的是用户完成表情指令而拍摄的视频，而非静态图像。静态图像易于伪造，动态视频需用户在指定时间内成功完成随机指定的任务，即时互动增加伪造难度。本发明实施例通过使用一系列随机生成的面部表情指令，提高伪造难度，提高真人识别率。

[0097] 在第一获取模块中，系统向用户发出各种表情指令，用户根据表情指令作出各种相应的表情，系统将其拍摄成视频。视频中含有中立表情图像。中立表情是指用户不做任何表情时的自然表情。中立表情图像是从视频中提取的。

[0098] 第一获取模块获取的用户完成表情指令的视频，包括用于测试的视频和用于训练模型的视频。用于测试的视频是用于分辨用户是否为真人。用于训练模型的视频是用于对建立的表情分类器模型进行训练。视频中的所有图像帧分为不重合的两部分，一部分作为训练数据的图像帧，另一部分作为测试数据的图像帧。

[0099] 作为优选方案，所述的第二获取模块具体包括：从用于训练模型的视频中逐帧提取图像，选取与中立表情图像区别最大的图像作为表情峰值图像。表情峰值图像是与中立表情的五官位移最大的一帧图像，比如：笑的最灿烂的一帧。测试时，不取峰值帧，而是对每一帧做处理。训练模型时，优选表情峰值图片。但是训练模型时，也可以不取表情峰值图片，而选取任何一张响应的指令表情图片，但是这样与中立表情图片区别不太大的话，训练出来的模型效果就会差一些。

[0100] 作为优选方案，所述的第三获取模块具体包括：对从第二获取模块获取的表情峰值图像，利用人脸探测器检测出人脸区域，并输出人脸区域的原始像素矩阵。从图像帧中检测出人脸的区域采用开源软件包中的人脸探测器。例如，利用开源软件 openCV 开源包中的 Viola-Jones 的人脸探测器。人脸探测器输出的是一个像素矩阵。该像素矩阵显示人脸位置在整个原图像中的坐标。当然还可以采用其他开源软件中的人脸探测器。

[0101] 作为优选方案，所述的第四获取模块中，使用开源工具从人脸区域中提取面部特征点。例如，CSIRO 软件开发工具包 CSIRO sdk 从人脸区域中提取面部特征点。预先可以在开源工具中设置提取的面部特征点的数量。例如，图 4 为一人脸区域，该人脸区域中设定了 66 个面部特征点。为更加准确的反应脸部表情，面部特征点主要涉及五官和脸的轮廓。开源工具输出面部特征点的坐标。坐标系的原点是原图像的左下角，x 轴是图像下边缘，方向向右为正方向；y 轴是图像下边缘，方向向上为正方向。

[0102] 作为优选，如图 6 所示，上述的真人验证系统中，所述的第五获取模块包括：

[0103] 计算单元：用于将面部特征点映射到标准人脸的特征点上，并计算线性仿射变换；

[0104] 应用单元：用于将线性仿射变换应用到 S30 获取的人脸区域的原始像素矩阵中，获得归一化的标准人脸像素矩阵。

[0105] 利用训练数据训练模型，然后把模型应用到测试数据上看效果。训练模型采用表情峰值图像帧。测试模型效果时，是每一帧图片都获取设定数量的面部特征点。将获取的一图像帧中的相关面部特征点映射到标准人脸的对应面部特征点，计算线性仿射变换。标准人脸适用于所有人。例如，相关面部特征点采用眼睛特征点中心 (a, b)。选择两眼中心坐标 (c, d) 来确定标准人脸。两眼距离及两眼中心坐标 (c, d) 可以预先设定，适用于每个人。当标准人脸以两个眼睛的坐标为标准时，则将第三获取模块得到的两个眼睛的中心坐标 (a, b)，映射到标准人脸的到的两个眼睛的中心 (c, d)，据此构造线性映射。其中，a 表示一只眼睛中含有的面部特征点的坐标均值，如图 4 中编号 36—41 的面部特征点的坐标均值，b 表示另一只眼睛中含有的面部特征点的坐标均值，如图 4 中编号 42—47 的面部特征点的坐标均值。c 表示标准人脸中一只眼睛中心的坐标，d 表示标准人脸中另一只眼睛中心的坐标。计算线性仿射变换的过程为现有技术。

[0106] 第四获取模块获取的每个面部特征点坐标会一一对应一个像素色彩 (g, b, r)。应用单元将线性仿射变换应用到第四获取模块中人脸区域中每个面部特征点的坐标上，得到人脸区域面部特征点的新坐标，但每个面部特征点对应的像素色彩不变，获得归一化的标准人脸像素矩阵。将线性仿射变换应用到所有图片的原始像素矩阵上就是归一化的过程。

[0107] 作为优选，如图 7 所示，所述的真人验证系统中，建立模块包括：

[0108] 构造单元：用于提取标准人脸的面部特征点，构造特征区域，并从特征区域中选择最有区分度的特征区域；

[0109] 建立单元：用于将最有区分度的特征区域中的每个像素点作为一个特征向量，对最有区分度的特征区域中所有像素点构成全部的特征向量进行机器学习，建立表情分类器。

[0110] 在构造单元中，除了利用规则多边形区域（例如三角形）提取标准人脸的面部特征点，还可以利用不规则多边形（例如，长方形、四条边以上的其他多边形）提取标准人脸的面部特征点。利用不规则多边形提取标准人脸的特征点，可以增加备选特征，捕捉更多相关脸部特征细节以供机器学习。使用不规则多边形后，特征点组合方式有多种，供选特征数量成倍增加，弥补了传统方法中可能丢失的重要特征。提取的这些面部特征点供输入机器学习模型使用。如果只取眼部特征，那可能只能提取 100 个特征，但是如果全脸范围提取特征，那能提取百万特征，从而不丢失关键区分度高的特征。本实施例通过全脸范围提取特征，安全性高于单区域（眨眼）特征。

[0111] 任意 3 个及以上特征点组合，可以构成一个特征区域。每个特征区域包括不同的特征点组合，包括五官特征、德朗奈三角区域等等。

[0112] 采用统计模型中逐步向前选择的方法从特征区域中选择最有区分度的特征区域。最有区分度的特征区域是相对可以显著区分中立表情和指令表情的区域变量。例如，需要区分中立表情（如 y0）和指令表情（如 y1），而有的特征区域的取值（如 x1）在中立表情和指令表情中表现类似，不易区分中立表情和指令表情，但是有的特征区域的取值（如 x3）在中立表情和指令表情中表现差异极大，容易区分中立表情和指令表情。这样，选择 x3 作为最有区分度的特征区域的区域变量。

[0113] 在建立单元中,最有区分度的特征区域中所有像素点构成全部的特征向量。例如,提取第五获取单元获得的标准人脸的 66 个特征点,通过构造多边形区域,比如任意 5 个点构造一个五边形,得到比如一百万个区域,通过逐步向前选择的方法从特征区域中选择最有区分度的特征区域,例如选出 1000 个区域。1000 个区域内每个像素点作为一个特征向量,1000 个区域内的特征向量合起来得到所有的特征向量。使用机器学习算法训练分类器。

[0114] 作为优选,上述的真人验证系统中,所述的判断模块依据式 (1) 判断用户是否为真人。

[0115] 
$$s = \frac{w_1}{w} \quad \text{式 (1)}$$

[0116] 其中,  $s$  表示用于测试的视频中,用户成功完成表情帧数百分比,  $w_1$  为单位时间内用户成功完成表情帧数,  $w$  表示单位时间内视频总帧数;当  $s \geq s_0$  时,用户为真人;当  $s < s_0$  时,用户不是真人;  $s_0$  表示判断阈值。计算所有指令的平均分数作为最终分数。判断阈值来界定真人与否。判断阈值  $s_0$  的值没有特殊的要求。判断阈值  $s_0$  是可调整的。判断阈值  $s_0$  设的高,用户就难以通过测试。判断阈值  $s_0$  设的低,用户就容易通过测试。判断阈值  $s_0$  可以用于用户体验与安全性的折衷。

[0117] 本领域技术人员应该知晓,实现上述实施例的方法或者系统,可以通过计算机程序指令来实现。该计算机程序指令装载到可编程数据处理设备上,例如计算机,从而在可编程数据处理设备上执行相应的指令,用于实现上述实施例的方法或者系统实现的功能。

[0118] 本领域技术人员依据上述实施例,可以对本申请进行非创造性的技术改进,而不脱离本发明的精神实质。这些改进仍应视为在本申请权利要求的保护范围之内。

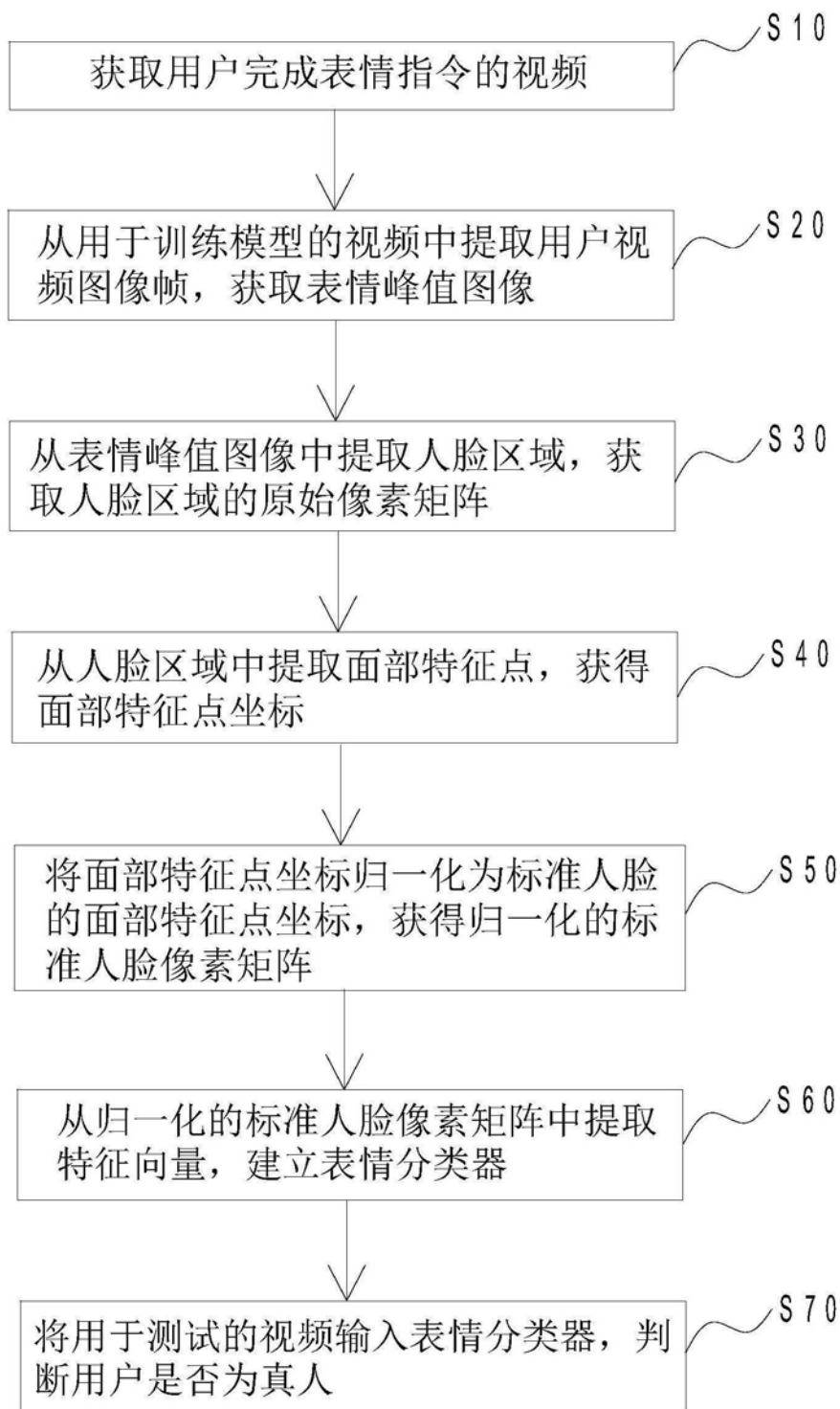


图 1

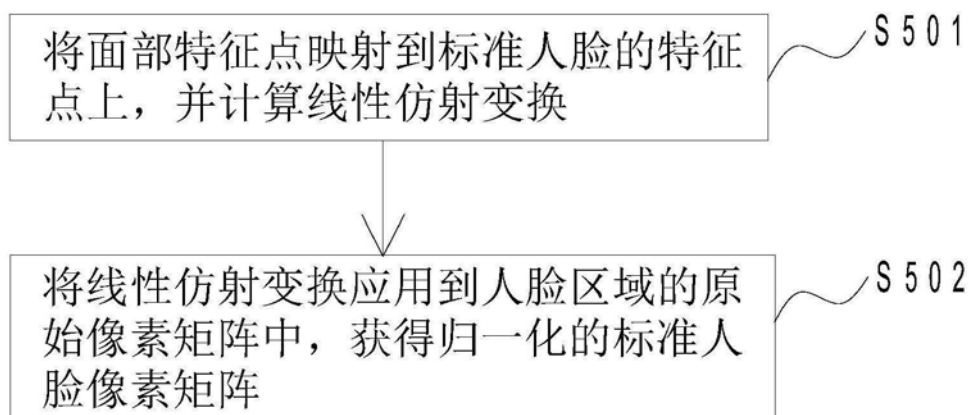


图 2

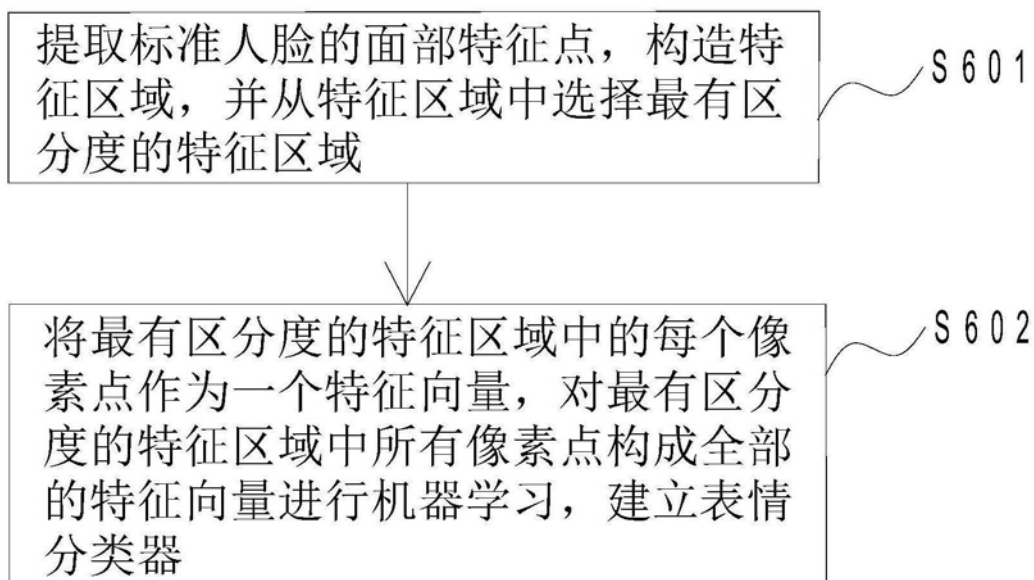


图 3

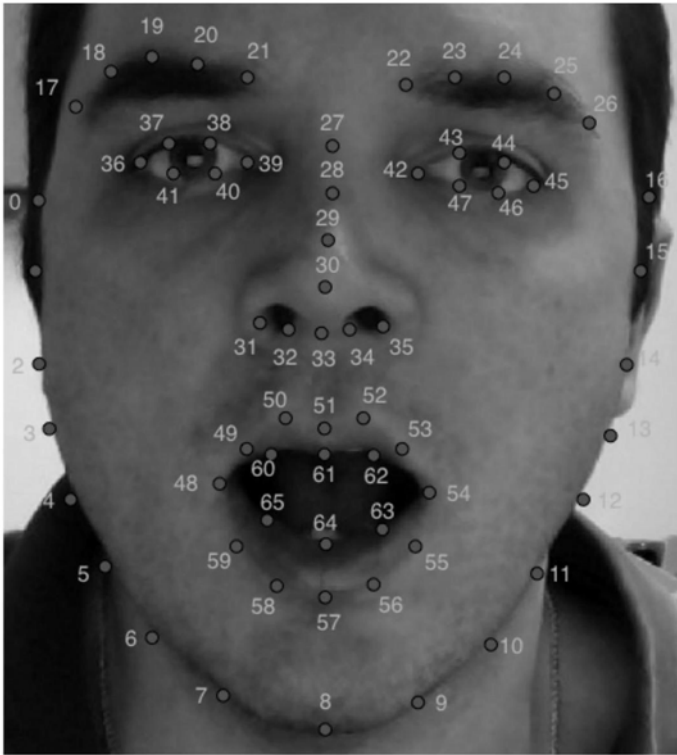


图 4



图 5

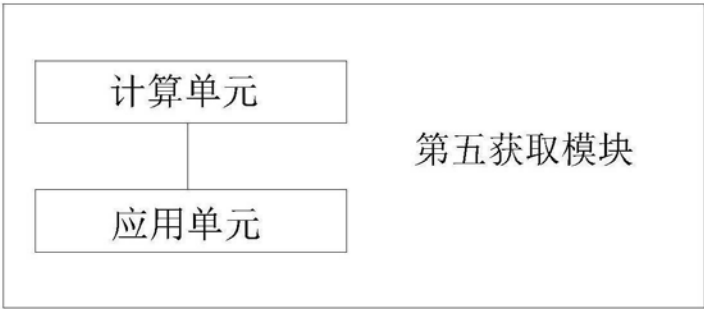


图 6

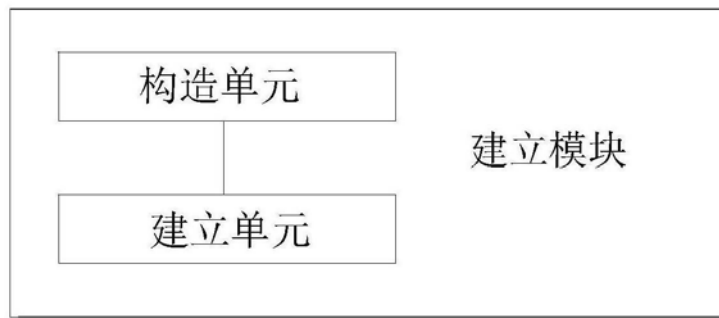


图 7