



(12)发明专利申请

(10)申请公布号 CN 107680229 A

(43)申请公布日 2018.02.09

(21)申请号 201710994250.5

(22)申请日 2017.10.23

(71)申请人 西安科技大学

地址 710054 陕西省西安市雁塔路中段58号

(72)发明人 秦学斌

(74)专利代理机构 上海精晟知识产权代理有限公司 31253

代理人 冯子玲

(51)Int.Cl.

G07C 9/00(2006.01)

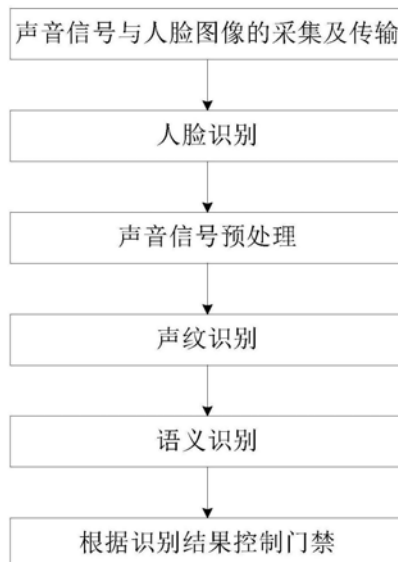
权利要求书3页 说明书9页 附图2页

(54)发明名称

基于语音特征和人脸识别的门禁系统及其控制方法

(57)摘要

本发明公开了一种基于语音特征和人脸识别的门禁系统及其控制方法,其系统包括用于拾取人的声音信号的麦克风和用于拍摄人脸图像的摄像头,以及计算机和微控制器模块,麦克风和摄像头均与计算机连接,计算上接有主机蓝牙模块,微控制器模块上接有用于与主机蓝牙模块无线通信的从机蓝牙模块,微控制器模块的输出端接有用于驱动门禁中电子锁开关的继电器电路模块。其方法包括步骤:一、声音信号与人脸图像的采集及传输;二、人脸识别;三、声音信号预处理;四、声纹识别;五、语义识别;六、根据识别结果控制门禁。本发明融合了人脸识别、声纹识别和语义识别为一体,设计合理,实现方便,鲁棒性很强,识别率较高,实用性强,推广应用价值高。



1. 一种基于语音特征和人脸识别的门禁系统,其特征在于:包括用于拾取人的声音信号的麦克风(2)和用于拍摄人脸图像的摄像头(3),以及用于处理采集的信号得到判别结果的计算机(4)和用于控制门禁中电子锁开关的微控制器模块(1),所述麦克风(2)和摄像头(3)均与计算机(4)连接,所述计算上接有主机蓝牙模块(5),所述微控制器模块(1)上接有用于与主机蓝牙模块(5)无线通信的从机蓝牙模块(6),所述微控制器模块(1)的输出端接有用于驱动门禁中电子锁开关的继电器电路模块(7)。

2. 按照权利要求1所述的一种基于语音特征和人脸识别的门禁系统,其特征在于:所述微控制器模块(1)包括单片机STC89C52。

3. 按照权利要求2所述的一种基于语音特征和人脸识别的门禁系统,其特征在于:所述主机蓝牙模块(5)为HC-06主机蓝牙模块,所述HC-06主机蓝牙模块与计算机(4)的USB接口连接;所述从机蓝牙模块(6)为HC-06从机蓝牙模块,所述HC-06从机蓝牙模块的信号接收端引脚RXD与单片机STC89C52的串口发送端引脚TXD连接,所述HC-06从机蓝牙模块的信号发送端引脚TXD与单片机STC89C52的串口接收端引脚RXD连接。

4. 按照权利要求2所述的一种基于语音特征和人脸识别的门禁系统,其特征在于:所述继电器电路模块(7)包括三极管Q1、继电器JDQ、电阻R1和二极管D1,所述三极管Q1的基级通过电阻R1与单片机STC89C52的第21引脚连接,所述三极管Q1的集电极接地,所述三极管Q1的发射极与继电器JDQ的线圈的一端和二极管D1的阳极连接,所述继电器JDQ的线圈的另一端和二极管D1的阴极均与外部电源模块的输出端VCC连接,门禁中电子锁接在所述继电器JDQ的常闭端。

5. 一种对如权利要求1所述基于语音特征和人脸识别的门禁系统进行控制的方法,其特征在于,该方法包括以下步骤:

步骤一、声音信号与人脸图像的采集及传输:人对着麦克风(2)说话时,所述麦克风(2)将其拾取到的人的声音信号传输给计算机(4);同时,所述摄像头(3)采集人脸图像并将其拍摄到的人脸图像传输给计算机(4);所述计算机(4)采集声音信号和人脸图像;

步骤二、人脸识别,具体过程为:

步骤201、所述计算机(4)调用OpenCV库函数对人脸图像进行分析处理,初步识别人脸图像,当判断为是人脸图像时,执行步骤202;当判断为非人脸图像时,返回步骤一重新采集人脸图像;

步骤202、所述计算机(4)调用支持向量机分类模块对人脸图像进行二分类,分为是人脸图像和非人脸图像两类,当判断为是人脸图像时,执行步骤203;当判断为非人脸图像时,返回步骤一重新采集人脸图像;

步骤203、所述计算机(4)将人脸图像正规化为 164×164 个像素大小;

步骤204、所述计算机(4)将人脸图像输入预先构建的Tensorflow深度学习人脸识别网络中,得到人脸识别的结果 $R_{\text{face}} = j$;其中, j 表示人脸识别结果为第 j 个人, j 的取值为非零自然数;

步骤三、声音信号预处理:所述计算机(4)调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;

步骤四、声纹识别:所述计算机(4)将步骤三得到的语谱图输入预先构建的Tensorflow深度学习声纹识别网络中,得到声纹识别的结果 $R_{\text{vp}} = i$;其中, i 表示声纹识别结果为第 i 个

人, i 的取值为非零自然数;

步骤五、语义识别: 所述计算机 (4) 将步骤三得到的语谱图输入预先构建的Tensorflow深度学习语义识别网络中, 得到语义识别的结果 $R_{se} = r$; 其中, r 表示语义识别结果为第 r 种语义, r 的取值为非零自然数;

步骤六、根据识别结果控制门禁: 所述计算机 (4) 先比较人脸识别与声纹识别的结果, 当 $i \neq j$ 时, 说明识别的结果表示不是同一个人, 计算机 (4) 不输出门禁控制信号; 当 $i = j$ 时, 说明识别的结果表示是同一个人, 计算机 (4) 再判断语义识别的结果 r 为 0 还是 1, 当语义识别的结果为 0 时, 说明人需要进行开门动作, 所述计算机 (4) 输出开门控制信号并通过主机蓝牙模块 (5) 发送出去; 当语义识别的结果为 1 时, 说明人需要进行关门动作, 所述计算机 (4) 输出关门控制信号并通过主机蓝牙模块 (5) 发送出去; 所述微控制器模块 (1) 通过从机蓝牙模块 (6) 接收到计算机 (4) 发送的开门控制信号或关门控制信号, 并输出相应的控制信号给继电器电路模块 (7), 控制电子门锁打开或锁上。

6. 按照权利要求 5 所述的方法, 其特征在于: 步骤 204 中所述Tensorflow深度学习人脸识别网络的构建方法为:

步骤A1、人脸图像的采集及传输: 人对着麦克风 (2) 说话时, 所述摄像头 (3) 采集人脸图像并将其拍摄到的人脸图像传输给计算机 (4); 所述计算机 (4) 采集人脸图像;

步骤A2、所述计算机 (4) 将其采集的人脸图像正规化为 164×164 个像素大小;

步骤A3、所述计算机 (4) 调用OpenCV库函数对人脸图像进行分析处理, 初步识别人脸图像, 当判断为是人脸图像时, 执行步骤A4; 当判断为非人脸图像时, 返回步骤A1重新采集人脸图像;

步骤A4、所述计算机 (4) 调用支持向量机分类模块对人脸图像进行二分类, 分为是人脸图像和非人脸图像两类, 当判断为是人脸图像时, 所述计算机 (4) 存储人脸图像并执行步骤A5; 当判断为非人脸图像时, 返回步骤A1重新采集人脸图像;

步骤A5、重复执行步骤A1至步骤A4, 直至所述计算机 (4) 存储了 10 个人、每个人不同状态下的 500 幅人脸图像;

步骤A6、所述计算机 (4) 构建一个卷积网络核的层数为五层、输入层节点为 164×164 个像素的人脸图像、输出层节点为 10 类的Tensorflow深度学习网络, 并将其存储的 10 个人、每个人不同状态下的 500 幅人脸图像作为训练样本, 对Tensorflow深度学习网络进行训练, 得到Tensorflow深度学习人脸识别网络; 所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$; 所述Tensorflow深度学习人脸识别网络的输出为 $R_{face} = j$, j 的取值为 1~10 的自然数。

7. 按照权利要求 5 所述的方法, 其特征在于: 步骤四中所述Tensorflow深度学习声纹识别网络的构建方法为:

步骤B1、声音信号的采集及传输: 人对着麦克风 (2) 说话时, 所述麦克风 (2) 将其拾取到的人的声音信号传输给计算机 (4); 所述计算机 (4) 采集声音信号;

步骤B2、所述计算机 (4) 调用语谱图绘制模块, 将接收到的声音信号绘制为语谱图;

步骤B3、重复执行步骤B1和步骤B2, 直至所述计算机 (4) 存储了 10 个人、每个人 200 组时长为 2s 的声音信号的语谱图;

步骤B4、所述计算机 (4) 构建一个卷积网络核的层数为五层、输入层节点为 256×256 个

像素的语谱图、输出层节点为10类的Tensorflow深度学习网络,并将其存储的10个人、每个人200组时长为2s的声音信号的语谱图作为训练样本,对Tensorflow深度学习网络进行训练,得到Tensorflow深度学习声纹识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$;所述Tensorflow深度学习声纹识别网络的输出为 $R_{vp} = i$, i 的取值为1~10的自然数。

8.按照权利要求5所述的方法,其特征在于:步骤五中所述Tensorflow深度学习语义识别网络的构建方法为:

步骤C1、声音信号的采集及传输:人对着麦克风(2)说话时,所述麦克风(2)将其拾取到的人的声音信号传输给计算机(4);所述计算机(4)采集声音信号;

步骤C2、所述计算机(4)调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;

步骤C3、重复执行步骤C1和步骤C2,直至所述计算机(4)存储了“开门”、“关门”和“你好”三种语义各500组时长为2s的声音信号的语谱图;

步骤C4、所述计算机(4)构建一个卷积网络核的层数为五层、输入层节点为 256×256 个像素的语谱图、输出层节点为三类的Tensorflow深度学习网络,并将其存储的“开门”、“关门”和“你好”三种语义各500组时长为2s的声音信号的语谱图作为训练样本,对Tensorflow深度学习网络进行训练,得到Tensorflow深度学习语义识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$;所述Tensorflow深度学习语义识别网络的输出为 $R_{se} = r$, r 的取值为0,1或2; $r=0$ 表示语义识别结果为“开门”, $r=1$ 表示语义识别结果为“关门”, $r=2$ 表示语义识别结果为“你好”。

基于语音特征和人脸识别的门禁系统及其控制方法

技术领域

[0001] 本发明属于智能控制技术领域,具体涉及一种基于语音特征和人脸识别的门禁系统及其控制方法。

背景技术

[0002] 随着人们生活水平的不断提高,人们对公共场所和家居的安全智能提出了更高的要求,安防问题已成为人们逐渐增加关注的焦点,安防设备越来越受到人们的重视,各种各样门禁锁层出不穷,发明一种先进技术的门禁系统对智能化场所是非常有必要的。与此同时,生物识别技术也取得了巨大的发展,例如语音识别、指纹识别、虹膜识别和人脸识别都已经广泛应用于人们的生活中。其中人脸识别因为其使用方便、不易仿冒、识别率高等特点被广泛应用于智能门禁系统。但是,单一生物识别技术由于复杂的环境可能会出现识别错误或识别不准确,具体表现为:(1) 门禁声纹识别技术是一种生物认证的方法,从说话人能反映自己生理和行为个性特征的语音参数中提取出说话人是谁的信息的过程;但是由于以下几个声纹识别的难点问题可能会出现误判:A、可靠的语音特征参数;B、语音信号的不稳定性;C、语音易被模仿;D、实际环境复杂;(2) 语音语义识别由于感冒,环境噪声等影响语义的识别精度;(3) 人脸识别技术已比较成熟,取得较大的进步,但当人们在化妆或环境光照变换较大的情况下,则人脸识别的误判率较高。

[0003] 为了解决以上问题,更好地将生物识别技术应用于智能门禁系统中,提高门禁识别的可靠性,专家学者们进行了大量研究。例如,(1) 申请号为200310118507.9的中国专利公开了一种结合语义和声纹信息的说话人身份确认系统,采用说话人的声纹特征通过GMM模型建立声纹模型,通过电话或录音设备对声音进行预处理,采用特征提取、声学模型建模、基于语义的说话人确认、文本有关和文本无关声纹确认系统等,文本有关的声纹确认系统是基于HMM声学建模的;其存在的缺点是:GMM,HMM模型,由于该模型对语音大数据的建模能力不是很好,对噪声的鲁棒性也比较差。(2) 申请号为201410198929.X的中国发明专利公开了一种基于语音及人脸识别的门禁道闸管理方法及装置,通过智能终端采集当前用户的人脸生物特征数据,并使所述文字密码与当前人脸生物特征数据构成一一对应的关系,判断所述文字信息与预设的文字密码是否一致,若一致则开门,若不一致则识别错误,判断所述发出解密语音信号的用户的人脸生物特征数据与预设的人脸生物特征数据是否一致,如是,所述门禁道闸管理系统发出驱动信号控制门禁道闸开启或关闭;否则,识别错误;其存在的缺点是:从人脸识别角度来说,由于周围环境,即光照和日光灯等因素的影响,与预设的人脸生物特征变化可能较大,识别率不高;另一方面,文字密码很容易被其他说话人用相同的问题密码。(3) 申请号为201510199387.2的中国发明专利公开了一种基于语音识别、面部识别和指纹识别的门禁控制方法,语音提取模块对即时语音信息进行特征提取,人脸提取模块对即时人脸信息进行特征提取,指纹提取模块对即时指纹信息进行特征提取;其存在的缺点是:需要对语音和人脸特征提取。(4) 申请号为201510903466.7的中国专利公开了一种基于FPGA的人脸和语音及重量复合识别门禁系统,人脸采集器、语音采集器、重量采集

器将各自采集的人脸图像信息、语音信息和重量范围信息传输至FPGA数据处理模块,然后通过FPGA数据处理模块传输至数据存取模块保存或通过FPGA数据处理模块将采集到的信息与数据存取模块内保存的信息进行对比;当三种身份识别都为“一致”时,最终识别结果才为“通过”,识别方式为:采集本人的人脸图像信息、语音信息、重量范围信息等存入数据存取模块,作为以后身份识别时的样本匹配数据;其存在的缺点是:选取库里的样本数据,以及周围的环境因素等对识别精度都有很大的影响。

发明内容

[0004] 本发明所要解决的技术问题在于针对上述现有技术中的不足,提供一种结构简单、设计合理、实现方便、实用性强、推广应用价值高的基于语音特征和人脸识别的门禁系统。

[0005] 为解决上述技术问题,本发明采用的技术方案是:一种基于语音特征和人脸识别的门禁系统,其特征在于:包括用于拾取人的声音信号的麦克风和用于拍摄人脸图像的摄像头,以及用于处理采集的信号得到判别结果的计算机和用于控制门禁中电子锁开关的微控制器模块,所述麦克风和摄像头均与计算机连接,所述计算机上接有主机蓝牙模块,所述微控制器模块上接有用于与主机蓝牙模块无线通信的从机蓝牙模块,所述微控制器模块的输出端接有用于驱动门禁中电子锁开关的继电器电路模块。

[0006] 上述的一种基于语音特征和人脸识别的门禁系统,其特征在于:所述微控制器模块包括单片机STC89C52。

[0007] 上述的一种基于语音特征和人脸识别的门禁系统,其特征在于:所述主机蓝牙模块为HC-06主机蓝牙模块,所述HC-06主机蓝牙模块与计算机的USB接口连接;所述从机蓝牙模块为HC-06从机蓝牙模块,所述HC-06从机蓝牙模块的信号接收端引脚RXD与单片机STC89C52的串口发送端引脚TXD连接,所述HC-06从机蓝牙模块的信号发送端引脚TXD与单片机STC89C52的串口接收端引脚RXD连接。

[0008] 上述的一种基于语音特征和人脸识别的门禁系统,其特征在于:所述继电器电路模块包括三极管Q1、继电器JDQ、电阻R1和二极管D1,所述三极管Q1的基级通过电阻R1与单片机STC89C52的第21引脚连接,所述三极管Q1的集电极接地,所述三极管Q1的发射极与继电器JDQ的线圈的一端和二极管D1的阳极连接,所述继电器JDQ的线圈的另一端和二极管D1的阴极均与外部电源模块的输出端VCC连接,门禁中电子锁接在所述继电器JDQ的常闭端。

[0009] 本发明还公开了一种方法步骤简单,融合人脸识别、声纹识别和语义识别为一体,鲁棒性很强,识别率较高的基于语音特征和人脸识别的门禁控制方法,其特征在于,该方法包括以下步骤:

[0010] 步骤一、声音信号与人脸图像的采集及传输:人对着麦克风说话时,所述麦克风将其拾取到的人的声音信号传输给计算机;同时,所述摄像头采集人脸图像并将其拍摄到的人脸图像传输给计算机;所述计算机采集声音信号和人脸图像;

[0011] 步骤二、人脸识别,具体过程为:

[0012] 步骤201、所述计算机调用OpenCV库函数对人脸图像进行分析处理,初步识别人脸图像,当判断为是人脸图像时,执行步骤202;当判断为非人脸图像时,返回步骤一重新采集人脸图像;

[0013] 步骤202、所述计算机调用支持向量机分类模块对人脸图像进行二分类,分为是人脸图像和非人脸图像两类,当判断为是人脸图像时,执行步骤203;当判断为非人脸图像时,返回步骤一重新采集人脸图像;

[0014] 步骤203、所述计算机将人脸图像正规化为 164×164 个像素大小;

[0015] 步骤204、所述计算机将人脸图像输入预先构建的Tensorflow深度学习人脸识别网络中,得到人脸识别的结果 $R_{face} = j$;其中, j 表示人脸识别结果为第 j 个人, j 的取值为非零自然数;

[0016] 步骤三、声音信号预处理:所述计算机调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;

[0017] 步骤四、声纹识别:所述计算机将步骤三得到的语谱图输入预先构建的Tensorflow深度学习声纹识别网络中,得到声纹识别的结果 $R_{vp} = i$;其中, i 表示声纹识别结果为第 i 个人, i 的取值为非零自然数;

[0018] 步骤五、语义识别:所述计算机将步骤三得到的语谱图输入预先构建的Tensorflow深度学习语义识别网络中,得到语义识别的结果 $R_{se} = r$;其中, r 表示语义识别结果为第 r 种语义, r 的取值为非零自然数;

[0019] 步骤六、根据识别结果控制门禁:所述计算机先比较人脸识别与声纹识别的结果,当 $i \neq j$ 时,说明识别的结果表示不是同一个人,计算机不输出门禁控制信号;当 $i = j$ 时,说明识别的结果表示是同一个人,计算机再判断语义识别的结果 r 为0还是1,当语义识别的结果为0时,说明人需要进行开门动作,所述计算机输出开门控制信号并通过主机蓝牙模块发送出去;当语义识别的结果为1时,说明人需要进行关门动作,所述计算机输出关门控制信号并通过主机蓝牙模块发送出去;所述微控制器模块通过从机蓝牙模块接收到计算机发送的开门控制信号或关门控制信号,并输出相应的控制信号给继电器电路模块,控制电子门锁打开或锁上。

[0020] 上述的方法,其特征在于:步骤204中所述Tensorflow深度学习人脸识别网络的构建方法为:

[0021] 步骤A1、人脸图像的采集及传输:人对着麦克风说话时,所述摄像头采集人脸图像并将其拍摄到的人脸图像传输给计算机;所述计算机采集人脸图像;

[0022] 步骤A2、所述计算机将其采集的人脸图像正规化为 164×164 个像素大小;

[0023] 步骤A3、所述计算机调用OpenCV库函数对人脸图像进行分析处理,初步识别人脸图像,当判断为是人脸图像时,执行步骤A4;当判断为非人脸图像时,返回步骤A1重新采集人脸图像;

[0024] 步骤A4、所述计算机调用支持向量机分类模块对人脸图像进行二分类,分为是人脸图像和非人脸图像两类,当判断为是人脸图像时,所述计算机存储人脸图像并执行步骤A5;当判断为非人脸图像时,返回步骤A1重新采集人脸图像;

[0025] 步骤A5、重复执行步骤A1至步骤A4,直至所述计算机存储了10个人、每个人不同状态下的500幅人脸图像;

[0026] 步骤A6、所述计算机构建一个卷积网络核的层数为五层、输入层节点为 164×164 个像素的人脸图像、输出层节点为10类的Tensorflow深度学习网络,并将其存储的10个人、每个人不同状态下的500幅人脸图像作为训练样本,对Tensorflow深度学习网络进行训练,

得到Tensorflow深度学习人脸识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$;所述Tensorflow深度学习人脸识别网络的输出为 $R_{face} = j$, j 的取值为1~10的自然数。

[0027] 上述的方法,其特征在于:步骤四中所述Tensorflow深度学习声纹识别网络的构建方法为:

[0028] 步骤B1、声音信号的采集及传输:人对着麦克风说话时,所述麦克风将其拾取到的人的声音信号传输给计算机;所述计算机采集声音信号;

[0029] 步骤B2、所述计算机调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;

[0030] 步骤B3、重复执行步骤B1和步骤B2,直至所述计算机存储了10个人、每个人200组时长为2s的声音信号的语谱图;

[0031] 步骤B4、所述计算机构建一个卷积网络核的层数为五层、输入层节点为 256×256 个像素的语谱图、输出层节点为10类的Tensorflow深度学习网络,并将其存储的10个人、每个人200组时长为2s的声音信号的语谱图作为训练样本,对Tensorflow深度学习网络进行训练,得到Tensorflow深度学习声纹识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$;所述Tensorflow深度学习声纹识别网络的输出为 $R_{vp} = i$, i 的取值为1~10的自然数。

[0032] 上述的方法,其特征在于:步骤五中所述Tensorflow深度学习语义识别网络的构建方法为:

[0033] 步骤C1、声音信号的采集及传输:人对着麦克风说话时,所述麦克风将其拾取到的人的声音信号传输给计算机;所述计算机采集声音信号;

[0034] 步骤C2、所述计算机调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;

[0035] 步骤C3、重复执行步骤C1和步骤C2,直至所述计算机存储了“开门”、“关门”和“你好”三种语义各500组时长为2s的声音信号的语谱图;

[0036] 步骤C4、所述计算机构建一个卷积网络核的层数为五层、输入层节点为 256×256 个像素的语谱图、输出层节点为三类的Tensorflow深度学习网络,并将其存储的“开门”、“关门”和“你好”三种语义各500组时长为2s的声音信号的语谱图作为训练样本,对Tensorflow深度学习网络进行训练,得到Tensorflow深度学习语义识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$;所述Tensorflow深度学习语义识别网络的输出为 $R_{se} = r$, r 的取值为0,1或2; $r=0$ 表示语义识别结果为“开门”, $r=1$ 表示语义识别结果为“关门”, $r=2$ 表示语义识别结果为“你好”。

[0037] 本发明与现有技术相比具有以下优点:

[0038] 1、本发明基于语音特征和人脸识别的门禁系统,硬件上仅需要麦克风、摄像头、微控制器模块和计算机,结构简单,设计合理,实现方便。

[0039] 2、本发明微控制器模块与计算机之间采用蓝牙进行通信,无需接线,使用方便整洁。

[0040] 3、本发明的基于语音特征和人脸识别的门禁系统的控制方法,融合了人脸识别、声纹识别和语义识别为一体,使用时,人只需要再门口面朝摄像头,对着麦克风说一声“开门”,就能够识别出主人并开门,智能化程度高,使用方便,且提高了识别的准确率,提高了

门禁系统的安全性。

[0041] 4、本发明的基于语音特征和人脸识别的门禁系统的控制方法,通过提取固定长度、符合语序的语音片段,形成时间线上的有序语谱图,通过Tensorflow深度学习网络提取特征,自适应的提取说话人的声纹和语义;即使说话人说同样的话采用声纹技术也能识别出说话人是否为预定数据库内的人;而且通过Tensorflow深度学习网络进行识别,对光照、环境等因素影响的鲁棒性很强。

[0042] 5、本发明的基于语音特征和人脸识别的门禁系统的控制方法,无需对语音和人脸特征进行提取,直接采用Tensorflow深度学习网络进行自动提取特征,方法步骤简单,识别效率高。

[0043] 6、本发明的基于语音特征和人脸识别的门禁系统的控制方法,只需要准备不同状态下的训练样本,对于人脸识别来说,在不同光照、不同姿态下的人脸模型;对于声纹识别来说,对同一个人样本可以说不同的话语,然后预处理后作为样本数据进行训练;对于语义识别来说,可以对多人采集语音信息,采用对不同的人说的统一词语的语义进行识别;对人脸识别、声纹识别和语义识别分别搭建三个不同的深度学习Tensorflow网络,能够精确的识别出人脸、声纹和语义特征。

[0044] 7、本发明的基于语音特征和人脸识别的门禁系统的控制方法,融合了声纹识别和人脸识别,判断说话人的声纹和人脸是否为同一个人,这样通过多生物特征信息提高了对说话人识别的精度。

[0045] 8、本发明采用了的训练样本不同,训练中采用两个字词语,每个词语形成一个2秒的语谱图,该训练样本容易判断出训练样本的开始和结束位置,既能识别出说话人,又能识别出相应的语义,识别率较高。

[0046] 9、本发明的人脸识别,无需提取人脸图像,首先在输入的图片上采用opencv函数找出可能是人脸的区域;由于opencv找出的人脸区域块可能出现误判,则再次通过支持向量机(SVM)的方法找出正负样本,正样本为人脸,负样本为非人脸;最后将识别的人脸区域作为Tensorflow深度学习网络的输入,其输出为某某人。

[0047] 10、本发明采用人工智能的方法对进门人员进行智能识别,融合了声纹、语义和人脸三方面特征进行识别,采用了三个独立的Tensorflow深度学习网络结构进行声纹、语义和人脸的识别,不仅通过声纹和人脸特征识别出进门人员,而且基于语义特征识别出说话人的语义,该方法的采用大大提高了对门禁系统中人员的识别精度,而且方便可靠,实用性强,推广应用价值高。

[0048] 综上所述,本发明设计合理,实现方便,融合了人脸识别、声纹识别和语义识别为一体,鲁棒性很强,识别率较高,实用性强,推广应用价值高。

[0049] 下面通过附图和实施例,对本发明的技术方案做进一步的详细描述。

附图说明

[0050] 图1为本发明基于语音特征和人脸识别的门禁系统的电路原理框图。

[0051] 图2为本发明微控制器模块与从机蓝牙模块的电路连接图。

[0052] 图3为本发明继电器电路模块的电路原理图。

[0053] 图4为本发明基于语音特征和人脸识别的门禁的控制方法的方法流程框图。

[0054] 附图标记说明:

[0055] 1—微控制器模块; 2—麦克风; 3—摄像头;

[0056] 4—计算机; 5—主机蓝牙模块; 6—从机蓝牙模块;

[0057] 7—继电器电路模块; 8—晶振电路; 9—复位电路。

具体实施方式

[0058] 如图1所示,本发明的基于语音特征和人脸识别的门禁系统,包括用于拾取人的声音信号的麦克风2和用于拍摄人脸图像的摄像头3,以及用于处理采集的信号得到判别结果的计算机4和用于控制门禁中电子锁开关的微控制器模块1,所述麦克风2和摄像头3均与计算机4连接,所述计算机4上接有主机蓝牙模块5,所述微控制器模块1上接有用于与主机蓝牙模块5无线通信的从机蓝牙模块6,所述微控制器模块1的输出端接有用于驱动门禁中电子锁开关的继电器电路模块7。

[0059] 本实施例中,如图2所示,所述微控制器模块1包括单片机STC89C52。具体实施时,所述单片机STC89C52上接有晶振电路8和复位电路9,所述晶振电路包括晶振X、电容C1和电容C2,所述晶振X的一端和电容C1的一端均与所述单片机STC89C52的第19引脚连接,所述晶振X的另一端和电容C2的一端均与所述单片机STC89C52的第18引脚连接,所述电容C1的另一端和电容C2的另一端均接地;所述复位电路9包括复位开关S、电容C3和电阻R1,所述复位开关S的一端、电容C3的一端和电阻R1的一端均与所述单片机STC89C52的第9引脚连接,所述复位开关S的另一端与电容C3的另一端连接,所述电阻R1的另一端接地。

[0060] 本实施例中,所述主机蓝牙模块5为HC-06主机蓝牙模块,所述HC-06主机蓝牙模块与计算机4的USB接口连接;如图2所示,所述从机蓝牙模块6为HC-06从机蓝牙模块,所述HC-06从机蓝牙模块的信号接收端引脚RXD与单片机STC89C52的串口发送端引脚TXD连接,所述HC-06从机蓝牙模块的信号发送端引脚TXD与单片机STC89C52的串口接收端引脚RXD连接。

[0061] 本实施例中,如图3所示,所述继电器电路模块7包括三极管Q1、继电器JDQ、电阻R1和二极管D1,所述三极管Q1的基级通过电阻R1与单片机STC89C52的第21引脚连接,所述三极管Q1的集电极接地,所述三极管Q1的发射极与继电器JDQ的线圈的一端和二极管D1的阳极连接,所述继电器JDQ的线圈的另一端和二极管D1的阴极均与外部电源模块的输出端VCC连接,门禁中电子锁接在所述继电器JDQ的常闭端。当所述单片机STC89C52的第21引脚输出高电平时,三极管Q1截止,所述继电器JDQ的线圈为开路,继电器JDQ不动作,门禁中电子锁闭合,处于关门状态;当所述单片机STC89C52的第21引脚输出低电平时,三极管Q1导通,所述继电器JDQ的线圈通电,所述继电器JDQ的常闭端断开,门禁中电子锁打开,实现了开门。

[0062] 如图4所示,本发明的基于语音特征和人脸识别的门禁控制方法,包括以下步骤:

[0063] 步骤一、声音信号与人脸图像的采集及传输:人对着麦克风2说话时,所述麦克风2将其拾取到的人的声音信号传输给计算机4;同时,所述摄像头3采集人脸图像并将其拍摄到的人脸图像传输给计算机4;所述计算机4采集声音信号和人脸图像;

[0064] 步骤二、人脸识别,具体过程为:

[0065] 步骤201、所述计算机4调用OpenCV库函数对人脸图像进行分析处理,初步识别人脸图像,当判断为是人脸图像时,执行步骤202;当判断为非人脸图像时,返回步骤一重新采集人脸图像;

[0066] 步骤202、所述计算机4调用支持向量机分类模块对人脸图像进行二分类,分为是人脸图像和非人脸图像两类,当判断为是人脸图像时,执行步骤203;当判断为非人脸图像时,返回步骤一重新采集人脸图像;

[0067] 步骤203、所述计算机4将人脸图像正规化为 164×164 个像素大小;

[0068] 步骤204、所述计算机4将人脸图像输入预先构建的Tensorflow深度学习人脸识别网络中,得到人脸识别的结果 $R_{\text{face}} = j$;其中, j 表示人脸识别结果为第 j 个人, j 的取值为非零自然数;

[0069] 本实施例中,步骤204中所述Tensorflow深度学习人脸识别网络的构建方法为:

[0070] 步骤A1、人脸图像的采集及传输:人对着麦克风2说话时,所述摄像头3采集人脸图像并将其拍摄到的人脸图像传输给计算机4;所述计算机4采集人脸图像;

[0071] 步骤A2、所述计算机4将其采集的人脸图像正规化为 164×164 个像素大小;

[0072] 步骤A3、所述计算机4调用OpenCV库函数对人脸图像进行分析处理,初步识别人脸图像,当判断为是人脸图像时,执行步骤A4;当判断为非人脸图像时,返回步骤A1重新采集人脸图像;

[0073] 步骤A4、所述计算机4调用支持向量机分类模块对人脸图像进行二分类,分为是人脸图像和非人脸图像两类,当判断为是人脸图像时,所述计算机4存储人脸图像并执行步骤A5;当判断为非人脸图像时,返回步骤A1重新采集人脸图像;

[0074] 步骤A5、重复执行步骤A1至步骤A4,直至所述计算机4存储了10个人、每个人不同状态下的500幅人脸图像;

[0075] 步骤A6、所述计算机4构建一个卷积网络核的层数为五层、输入层节点为 164×164 个像素的人脸图像、输出层节点为10类的Tensorflow深度学习网络,并将其存储的10个人、每个人不同状态下的500幅人脸图像作为训练样本,对Tensorflow深度学习网络进行训练,得到Tensorflow深度学习人脸识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$;所述Tensorflow深度学习人脸识别网络的输出为 $R_{\text{face}} = j$, j 的取值为1~10的自然数。

[0076] 本发明的人脸识别方法,首先基于OpenCV库函数初步设别人脸,然后采用支持向量机分类算法进行二分类,即分为是人脸图像和非人脸图像两类,克服了仅采用OpenCV库函数进行人脸识别会由于光照、背景等变化出现误检的缺陷;最后再采用Tensorflow深度学习人脸识别网络得到最终的人脸识别结果,人脸识别的精确度高。

[0077] 步骤三、声音信号预处理:所述计算机4调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;具体实施时,所述计算机4是在MATLAB软件中,运用spectrogram函数,将其接收到的声音信号绘制为语谱图的。

[0078] 步骤四、声纹识别:所述计算机4将步骤三得到的语谱图输入预先构建的Tensorflow深度学习声纹识别网络中,得到声纹识别的结果 $R_{\text{vp}} = i$;其中, i 表示声纹识别结果为第 i 个人, i 的取值为非零自然数;

[0079] 本实施例中,步骤四中所述Tensorflow深度学习声纹识别网络的构建方法为:

[0080] 步骤B1、声音信号的采集及传输:人对着麦克风2说话时,所述麦克风2将其拾取到的人的声音信号传输给计算机4;所述计算机4采集声音信号;

[0081] 步骤B2、所述计算机4调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;

[0082] 步骤B3、重复执行步骤B1和步骤B2,直至所述计算机4存储了10个人、每个人200组时长为2s的声音信号的语谱图;

[0083] 步骤B4、所述计算机4构建一个卷积网络核的层数为五层、输入层节点为 256×256 个像素的语谱图、输出层节点为10类的Tensorflow深度学习网络,并将其存储的10个人、每个人200组时长为2s的声音信号的语谱图作为训练样本,对Tensorflow深度学习网络进行训练,得到Tensorflow深度学习声纹识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 $3 \times 3, 2 \times 2, 3 \times 3, 2 \times 2, 2 \times 2$;所述Tensorflow深度学习声纹识别网络的输出为 $R_{vp}=i$, i 的取值为1~10的自然数。

[0084] 具体实施时,构建Tensorflow深度学习网络的程序为:

[0085] $W_conv1=weight_variable([3,3,1,16])$

[0086] $b_conv1=bias_variable([16])$

[0087] $W_conv2=weight_variable([2,2,16,32])$

[0088] $b_conv2=bias_variable([32])$

[0089] $W_conv3=weight_variable([3,3,32,64])$

[0090] $b_conv3=bias_variable([64])$

[0091] $W_conv4=weight_variable([2,2,64,128])$

[0092] $b_conv4=bias_variable([128])$

[0093] $W_conv5=weight_variable([2,2,128,256])$

[0094] $b_conv5=bias_variable([256])$

[0095] $W_fc1=weight_variable([4*4*256,256])$

[0096] $b_fc1=bias_variable([256])$

[0097] $W_fc2=weight_variable([256,256])$

[0098] $b_fc2=bias_variable([256])$

[0099] $W_fc3=weight_variable([256,10])$

[0100] $b_fc3=bias_variable([10])$

[0101] 步骤五、语义识别:所述计算机4将步骤三得到的语谱图输入预先构建的Tensorflow深度学习语义识别网络中,得到语义识别的结果 $R_{se}=r$;其中, r 表示语义识别结果为第 r 种语义, r 的取值为非零自然数;

[0102] 本实施例中,步骤五中所述Tensorflow深度学习语义识别网络的构建方法为:

[0103] 步骤C1、声音信号的采集及传输:人对着麦克风2说话时,所述麦克风2将其拾取到的人的声音信号传输给计算机4;所述计算机4采集声音信号;

[0104] 步骤C2、所述计算机4调用语谱图绘制模块,将接收到的声音信号绘制为语谱图;

[0105] 步骤C3、重复执行步骤C1和步骤C2,直至所述计算机4存储了“开门”、“关门”和“你好”三种语义各500组时长为2s的声音信号的语谱图;

[0106] 步骤C4、所述计算机4构建一个卷积网络核的层数为五层、输入层节点为 256×256 个像素的语谱图、输出层节点为三类的Tensorflow深度学习网络,并将其存储的“开门”、“关门”和“你好”三种语义各500组时长为2s的声音信号的语谱图作为训练样本,对Tensorflow深度学习网络进行训练,得到Tensorflow深度学习语义识别网络;所述Tensorflow深度学习人脸识别网络五层卷积网络核的大小从一层到第五层分别为 3×3 ,

2x2, 3x3, 2x2, 2x2; 所述Tensorflow深度学习语义识别网络的输出为 $R_{se}=r$, r 的取值为0, 1或2; $r=0$ 表示语义识别结果为“开门”, $r=1$ 表示语义识别结果为“关门”, $r=2$ 表示语义识别结果为“你好”。

[0107] 步骤六、根据识别结果控制门禁: 所述计算机4先比较人脸识别与声纹识别的结果, 当 $i \neq j$ 时, 说明识别的结果表示不是同一个人, 计算机4不输出门禁控制信号; 当 $i = j$ 时, 说明识别的结果表示是同一个人, 计算机4再判断语义识别的结果 r 为0还是1, 当语义识别的结果为0时, 说明人需要进行开门动作, 所述计算机4输出开门控制信号并通过主机蓝牙模块5发送出去; 当语义识别的结果为1时, 说明人需要进行关门动作, 所述计算机4输出关门控制信号并通过主机蓝牙模块5发送出去; 所述微控制器模块1通过从机蓝牙模块6接收到计算机4发送的开门控制信号或关门控制信号, 并输出相应的控制信号给继电器电路模块7, 控制电子门锁打开或锁上。具体实施时, 当人需要进行关门动作时, 所述单片机STC89C52的第21引脚输出高电平, 当人需要进行开门动作时, 所述单片机STC89C52的第21引脚输出低电平。

[0108] 以上所述, 仅是本发明的较佳实施例, 并非对本发明作任何限制, 凡是根据本发明技术实质对以上实施例所作的任何简单修改、变更以及等效结构变化, 均仍属于本发明技术方案的保护范围内。

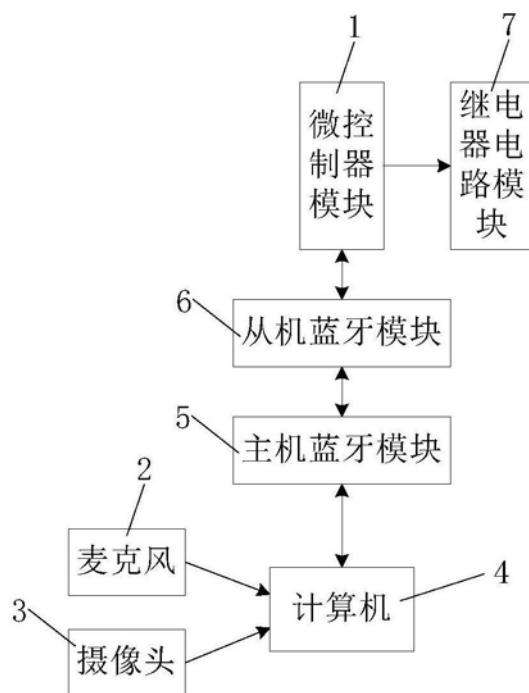


图1

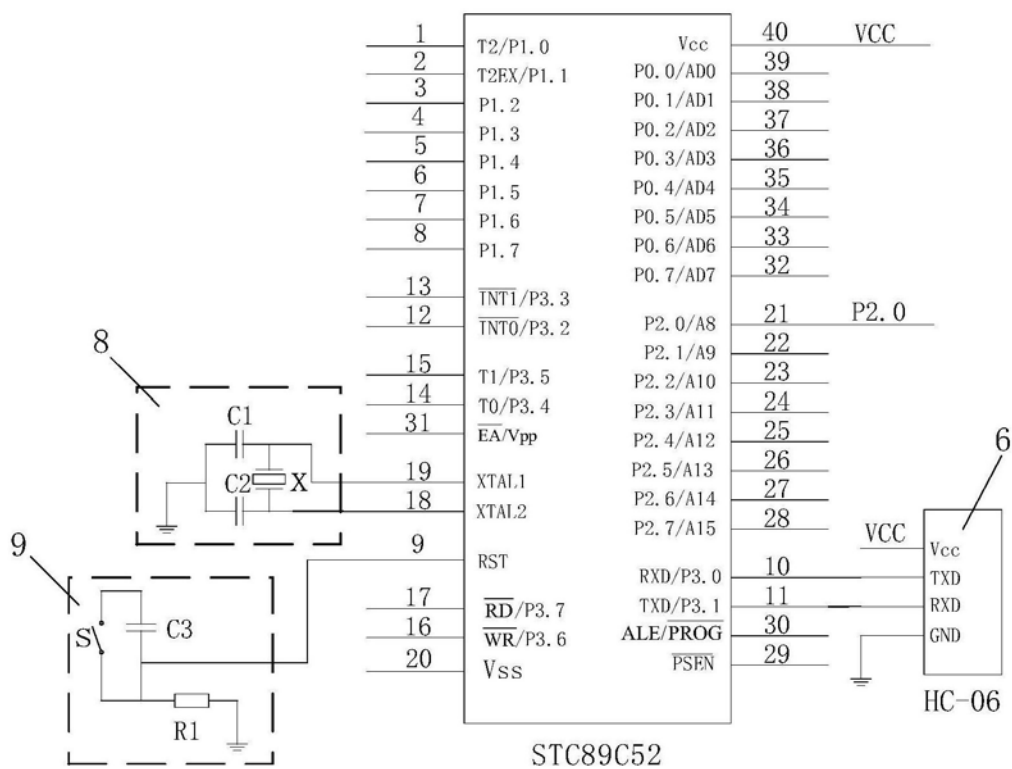


图2

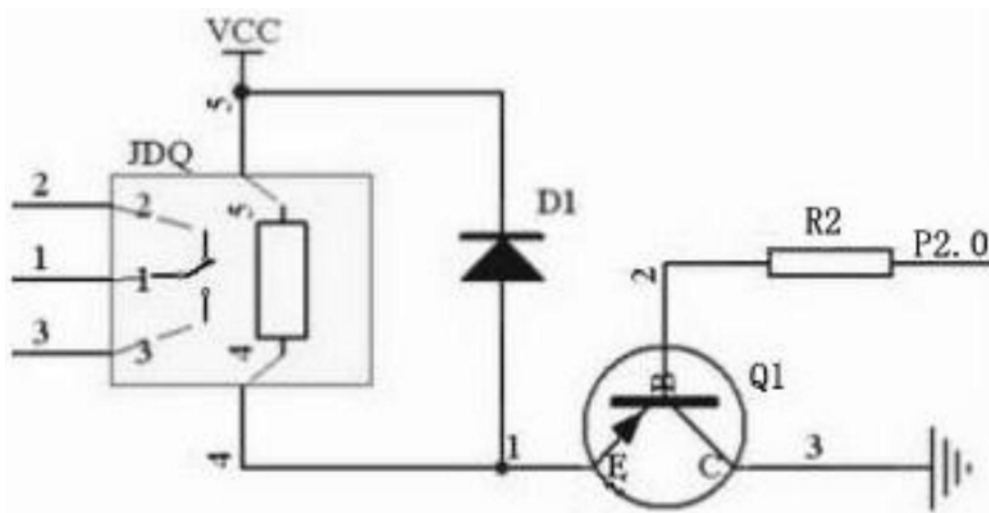


图3

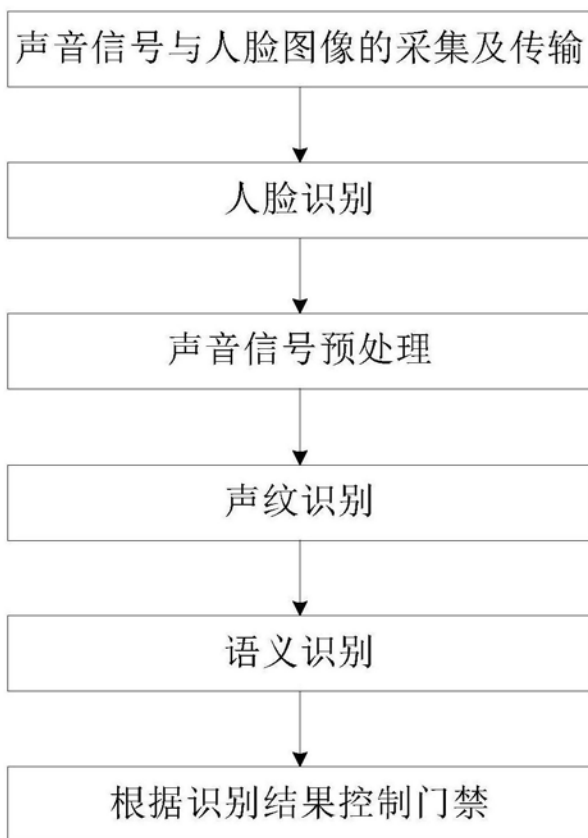


图4