

DOI:10.19651/j.cnki.emt.1802047

一种基于声纹识别的智能门锁系统设计与实现

王涛¹ 王国中² 朱林林²

(1.上海大学上海电影学院上海电影特效工程技术研究中心 上海 200072; 2.上海大学通信与信息工程学院 上海 200444)

摘要: 针对目前市面上已有的指纹锁、人脸锁等,尚且缺乏一款成熟的语音锁,设计并实现了一种利用不同说话人声音之间的差异,而识别开锁人身份的智能语音门锁系统。系统以智能手机为客户端平台,采集用户的语音,上传至云服务器,利用声纹识别技术识别用户身份,当用户通过身份认证,系统再通过无线网络通信的方式,将开锁信号发送到以 NodeMCU 为硬件平台的门锁端,从而实现开锁操作。此外,系统采用 Web 网页的形式展示服务端后台,可以实现管理员对系统的远程管理和监控。实验结果表明,该门锁系统的具有 90% 左右的识别准确率,同时错误接受率较低,充分证明了系统的安全性和稳定性。

关键词: 智能门锁;声纹识别;智能手机;NodeMCU;网络通信

中图分类号: TN912.3 **文献标识码:** A **国家标准学科分类代码:** 510.99

Design and implementation of a smart door lock system based on voiceprint recognition

Wang Tao¹ Wang Guozhong² Zhu Linlin²

(1.Shanghai Film Special Effects Engineering Technology Research Center, Shanghai Film Academy, Shanghai University, Shanghai 200072,China;

2. School of Communication & Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: As the existing fingerprint locks and face locks on the market, there is still a lack of a mature voice lock, and a smart voice lock system that uses the difference between different speaker voices to identify the unlocked person is designed and implemented. The system uses the smart phone as the client platform, collects the user's voice, uploads it to the cloud server, and uses the voiceprint recognition technology to identify the user identity. When the user passes the identity authentication, the system sends the unlock signal to the NodeMCU which is the door lock end of the hardware platform through wireless network communication, thereby realizing the unlocking operation. In addition, the system uses the web page to display the server background, which enables the administrator to remotely manage and monitor the system. The experimental results show that the door lock system has a recognition accuracy of about 90%, and the error acceptance rate is very low, which fully proves the safety and stability of the system.

Keywords: smart door lock; voiceprint recognition; smart phone; NodeMCU; network communication

0 引言

锁的发展一直伴随着人类社会的发展,随着制作工艺与科学技术的不断提高,锁的形式也越发多元化。从最早的机械锁,到后来的射频卡锁、密码锁,进而发展到现在比较常见的指纹锁、人脸锁等,从中可以看出,锁的发展趋势是在保证其基本属性—安全的前提下,逐渐趋向于便捷化、智能化^[1-2]。如指纹锁、人脸锁,这类基于人体生物特征的智能门锁的投入使用,为用户带来极大的便利,不仅解决了用户忘带钥匙的烦恼,同时也避免了因钥匙遗失而带来的

安全隐患。

声纹如同指纹、人脸一样,是人体特有的一种生物特征^[3]。声纹识别技术又被称为说话人识别技术,是通过对话人语音信号特征分析处理,识别说话人身份的过程^[4]。由于每个人都有其独一无二的声纹特征,很难被模仿或者伪造,所以声纹识别技术一直被应用于身份认证领域^[5]。本文中声纹识别技术与门锁结合,设计并制作出一款基于声纹识别的智能门锁,相比于指纹识别、人脸识别类门锁,其以非接触式和不侵犯隐私等特点更为突出,更容易被用户所接受。

1 系统需求分析与总体设计

在基于声纹识别的智能门锁系统的构建中,用户需要从智能手机端进行声纹注册,发起开门请求,进行声纹验证。管理员可以在 PC 端进行操作,审核用户注册申请,管理用户组,以及查看门锁的操作记录。在用户通过声纹识别后,通过无线通信,实现打开门锁开关^[6]。为了满足这些需求,对声纹门锁系统设计如图 1 所示。

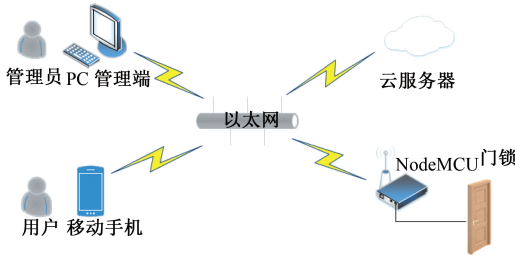


图 1 基于声纹识别的智能门锁系统框架

从系统组成上划分,该声纹门锁系统主要分为 4 个部分:1)客户端,采用现在持有率较高的移动智能手机平台,配合开发的特有手机应用,方便用户对门锁系统的操作和使用;2)电脑管理端,采用 Web 网页形式呈现,管理员可以通过管理端完成对用户申请的审核,管理用户组,以及查看开门记录等;3)云服务器端,服务器运行声纹识别算法,提供声纹识别接口,供客户端调用,此外服务器还运行系统后台服务程序,提供数据库操作接口,供网页端调用;4)门锁端,这部分主要由 NodeMCU,变压器模块和电磁门锁组成,其中 NodeMCU 内嵌 WiFi 模块,通过 TCP/IP 技术可与云服务器端进行数据通信,实现对门锁的控制。

2 系统硬件设计

系统硬件内容主要集中在门锁端,门锁端硬件部分有无线通信模块,电源降压模块及其外围电路组成的电磁门锁控制电路。硬件实物如图 2 所示。

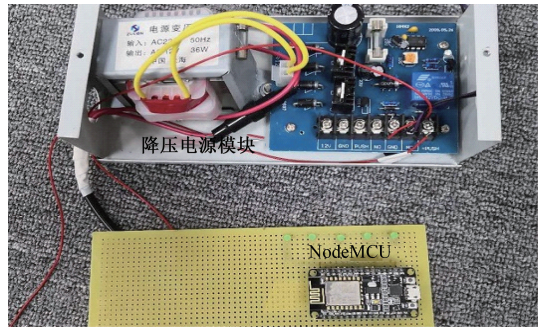


图 2 系统硬件实物

门锁端的控制模块采用的是 NodeMCU 可编程固件。NodeMCU 是一块开源的物联网开发板,它内置了一块基于 ESP8266 的 WiFi 模块,开发通常使用快捷的 Lua 语言和 Arduino IDE。在结构上具有一个集成完善得,USB 串口,通过 USB 口的接入省去了使用 FTDI 的 USB 转串口来开发编程的不便^[7]。在该系统的使用中,将 ESP8266 WiFi 芯片设置成透传模式,便可实现接入网络并和云服务器、智能手机进行数据通信。接收到无线控制信号,经过分析处理,转化成 NodeMCU 的 GPIO 口输出的高低电平。

电磁门锁采用的是 ZUCON 公司生产的祖程 303 W 电磁门锁电源,其内置变压器模块、继电器模块、延时调节模块等。变压器模块能够对 220 V 民用电压降压,达到电磁门锁的额定输入电压。NodeMCU 输出的高低电平信号输入到继电器模块,电平的高低控制了门锁电源的通断。电磁门锁在通电状态下,工作,产生巨大的吸引力,门关;反之,门关。延时调节模块能够控制电源每次断电状态的时长,即开门状态的时长,时间到了后,会自动切换回通电状态,门关。

3 系统软件设计与声纹识别算法

3.1 系统软件设计

系统的软件设计分为两部分,一部分是用户注册,另一部分是开门操作。软件框图如图 3 所示。

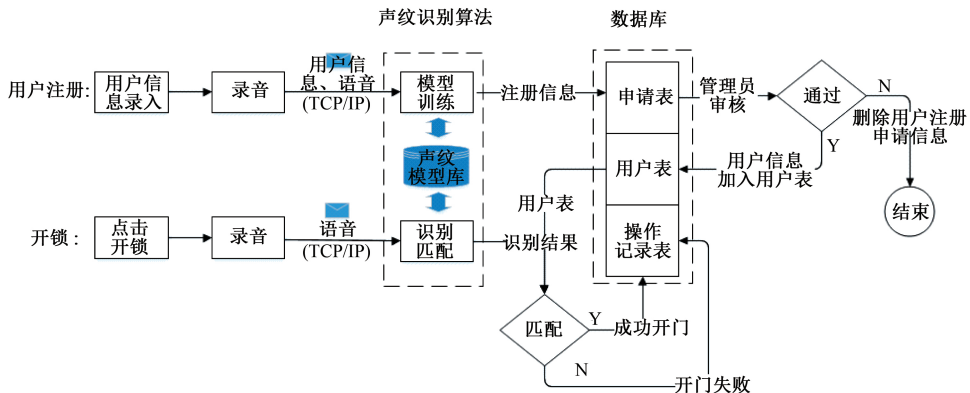


图 3 系统软件框图

用户注册流程如下:

- 1) 输入注册用户相关信息;
- 2) 录音,对同一开门指令,如“芝麻开门”,让注册者录 5 段语音口令,完成后通过 TCP 协议上传到云服务器;
- 3) 对注册语音进行训练,得到注册用户的声纹模型,存入声纹模型库,同时将用户申请信息加入服务端数据库中的申请表中;
- 4) 管理员后台通过 PC 端网页可以对用户的注册申请进行审核,若审核通过,注册用户将被添加进用户表,注册成功,否则,申请信息将被删除,注册失败。

开锁操作流程如下:

- 1) 点击开锁;
- 2) 录音,输入语音开门指令;
- 3) 对开门指令进行声纹识别,与库中的声纹模型进行匹配,将识别结果与数据库中的用户列表进行比对;
- 4) 若声纹识别结果存在于用户表中,则操作用户为注册用户,开锁操作通过,服务端向门锁端发送开门指令,并将成功开锁记录存入操作记录表,反之,用户为非法用户,将开锁失败记录存入。

3.2 声纹识别算法

本文中采用的是基于 Mel 频率倒谱系数 (Mel-frequency cepstral coefficients, MFCC) 特征和高斯混合模型 (Gaussian mixture model, GMM) 的声纹识别算法。

MFCC 是声音的短期功率谱的表示,基于非线性频谱上的对数功率谱的线性余弦变换。在自动语音识别领域, MFCC 是使用最广泛的特征之一,同时,它也广泛应用于声纹识别领域^[8]。MFCC 特征提取过程^[9]如图 4 所示。

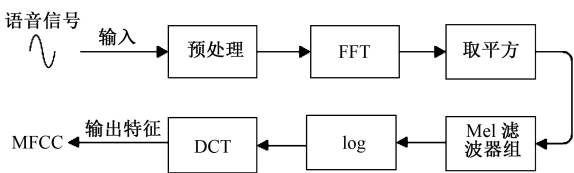


图 4 MFCC 提取流程

1) 对原始语音进行预加重、分帧和加窗等预处理操作,得到短时信号 $x(n)$;

2) 对每一个短时信号 $x(n)$ 进行快速傅里叶变换 (FFT) 得到对应的线性频谱 $Xa(k)$;

3) 对 $Xa(k)$ 取模的平方,得到离散功率谱 $X(k)$;

4) 将得到的频谱 $X(k)$ 通过 Mel 滤波器^[10]组进行滤波,再对滤波器组的输出求对数能量 m_i ;

5) 对 m_i 进行离散余弦变换 (discrete cosine transform, DCT) 得到 MFCC,此变换式可简化为:

$$C_n = \sum_{i=1}^M m_i \cos \left[\frac{\pi n (i - 0.5)}{M} \right], n = 1, 2, \dots, L \quad (1)$$

式中: C_n 表示的是 MFCC 的系数; L 表示 MFCC 的阶数。

实验表明,当阶数升高到一定程度,系统识别性能的改善将变得很小,系统的复杂度却大大增加。因此实际应用中,只需取 12~16 阶倒谱系数就可以达到很高的识别效率^[11]。

高斯混合模型用于说话人识别是由麻省理工学院林肯实验室的 Reynolds 等于 1995 年提出的,它是一个状态的连续隐马尔科夫模型,该模型用多个高斯分布的密度函数的组合来描述特征矢量在概率空间的分布情况^[12]。高斯混合模型是目前声纹识别中比较常用的一个模型。基于高斯混合模型的声纹识别可以分为训练阶段和识别阶段。

1) 训练阶段:基于前期得到的 MFCC 特征参数,然后依照该参数训练语音信号的高斯混合模型,组成高斯混合模型库。

2) 识别阶段:本文研究设计的声纹识别流程如图 5 所示。对待识别的语音提取特征参数,通过与高斯混合模型库做匹配,通过计算得到同模型库里匹配的概率最大值,再依据相似性准则决策出最优结果^[13-14]。本文的门锁系统应用于开集环境,所以需要设置一个阈值来判断待识别人是否是模型库中的某一个人,这里可以采用大量实验的方法,确定该阈值的合理取值。

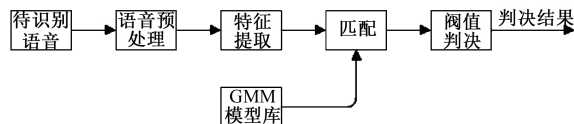


图 5 声纹识别流程

4 系统测试与分析

准确识别用户身份对于声纹门锁系统尤为重要,越低的错误率意味着系统越安全可靠,而对于系统的错误率,又分为错误接受率 (FAR) 和错误拒绝率 (FRR), FAR 与 FRR 的计算方法^[15]如下:

$$FAR = \frac{NFA}{NIRA} \times 100\% \quad (2)$$

$$FRR = \frac{NFR}{NGRA} \times 100\% \quad (3)$$

式中: $NIRA$ 和 $NGRA$ 分别是假冒者和合法用户分别尝试的总次数; NFA 和 NFR 分别是错误接受和错误拒绝的次数。

FAR 和 FRR 两个错误率反映了声纹门锁系统两个方面的特性: FAR 越低,假冒者被接受的可能性越低,系统的安全性也就越高; FRR 越低,合法用户被错误拒绝的可能性越低,系统的易用性和用户体验就越好。理想情况下这两个错误率都为 0,系统最优。但实际上这两个错误率是相互矛盾的,一个错误率下降会导致另一个错误率上升。可以通过调节阈值让两项指标的相对大小均达到一个合

理的水平,满足系统需求^[16]。系统测试内容主要包括 FAR 和 FRR 这两项指标。

首先组装好门锁系统,打开手机 APP,运行后台服务程序和声纹识别算法服务,如图 6 所示。然后选取 10 名实验者参与测试一共进行了 4 轮实验,每轮每人测试一次。对 FRR 与 FAR 两项指标实验测试结果如表 1 所示。

实验前,10 个用户,每个用户注册有 5 段语音,所以声纹库内有 50 个声纹模型,测试过程中,每个人提供 1 个待识别的语音,所以实验中 $NIRA = 5 \times (10 - 1) \times 10 = 450$, $NGRA = 50$ 。

从表 1 的数据可以看出,声纹门锁系统错误接收率较低,错误拒绝率相对稍高,但是处在可接受范围内。错误接受率影响系统的安全性,而错误拒绝率影响用户体验^[16]。通过实验者的反馈和实验数据分析可知,该声纹门锁系统,用户体验良好,系统的安全性较高。

针对实验过程中识别错误的语音,通过与正常识别的语音内容对比发现,该类语音内容中普遍出现了不同程度的噪音,由此分析可知,错误识别可能是由环境噪音引起,后续研究中应在语音预处理阶段继续改进,降低噪音的干扰。如在声纹识别过程中,对语音活动检测(voice activity detection, VAD)当前采用的是基于能量检测算法,这种方法对于噪声比较敏感,可以替换成抗噪声性能更好的长期频谱差异(long-term spectral divergence, LTSD)^[17]语音检测算法。

5 结 论

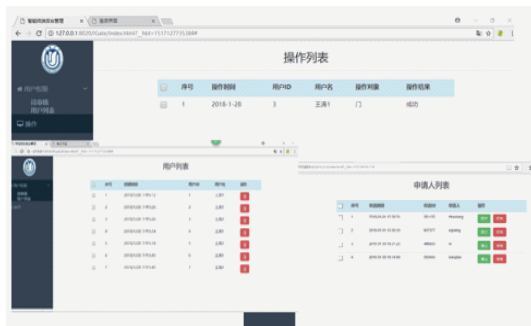
本文通过对智能门锁领域的现状和发展需求进行分析,设计了一套基于声纹识别的智能门锁系统。并从简单操作的智能手机客户端,到基于 Web 技术的 PC 管理端,再到结合声纹识别技术的云服务器端,最后到基于 NodeMCU 开发板的门锁端,通过无线网络通信技术,构建了一套完整的数据交互体系,实现了本文所设计的门锁系统。经过测试,该系统用户体验良好,且安全可靠。本文的方法可以广泛应用于其他安防领域。

参考文献

- [1] 张力平.智能门锁渐成趋势[J].电信快报,2015(2):10.
- [2] 薛泽宇,徐明星.一种基于声纹识别的门禁系统设计[C].中国中文信息学会语音信息专业委员会第十四届全国人机语音通讯学术会议(NCMMSC'2017),2017:4.
- [3] 裴鑫.声纹识别系统关键技术研究[D].哈尔滨:哈尔滨理工大学,2014.
- [4] 李湾湾.说话人声纹识别的算法研究[D].杭州:浙江大学,2017.
- [5] SUKHWAL A, KUMAR M. Comparative study between different classifiers based speaker recognition system using MFCC for noisy environment [C]. International Conference on Green Computing & Internet of Things, IEEE, 2016: 955-960.
- [6] 贾璇.基于用户体验的智能门锁设计研究[D].济南:山东大学,2017.
- [7] 寇世文.一种基于 NodeMCU 的智能锁控制系统[J].电子技术与软件工程,2018(10):100.



(a) 客户端截图



(b) PC 网页管理端



(c) 门锁端实物

图 6 声纹门锁系统

表 1 声纹门锁系统 4 轮实验测试结果

类别	第 1 轮	第 2 轮	第 3 轮	第 4 轮
正确识别次数	8	9	9	9
错误识别次数	2	1	1	1
NFA	1	0	1	0
NFR	1	1	0	1
FAR/%	0.22	0	0.22	0
FRR/%	2	2	0	2

- [8] 李虹,徐小力,吴国新,等.基于 MFCC 的语音情感特征提取研究[J].电子测量与仪器学报,2017,31(3):448-453.
- [9] 刘翔,孙静,赵洋,等.基于 MFCC 的心音信号特征提取及识别研究[J].电子测量技术,2018,41(2):1-5.
- [10] 胡政权,曾毓敏,宗原,等.说话人识别中 MFCC 参数提取的改进[J].计算机工程与应用,2014,50(7):217-220.
- [11] 周萍,李晓盼,李杰,等.混合 MFCC 特征参数应用于语音情感识别[J].计算机测量与控制,2013,21(7):1966-1968,1986.
- [12] 田昕,唐皓,余江,等.GMM-UBM 语种识别技术在无线电监管中的应用[J].电子测量技术,2015,38(8):82-84.
- [13] 蒋晔,唐振民.GMM 文本无关的说话人识别系统研究[J].计算机工程与应用,2010,46(11):179-182,195.
- [14] 曹洁,潘鹏.基于 GMM 的说话人识别技术研究[J].计算机工程与应用,2011,47(11):114-117.
- [15] 张建新.自动掌纹识别理论和算法研究[D].大连:大连理工大学,2009.
- [16] 周春晖.基于 MFCC 特征和 GMM 模型的说话人识别系统研究[D].兰州:兰州理工大学,2013.
- [17] RAMIREZ J, SEGURA J C, BENITEZ C, et al. Efficient voice activity detection algorithms using long-term speech information[J]. Speech Communication, 2004, 42(3):271-287.

作者简介

王涛, 硕士研究生, 主要研究方向为物联网技术、智能家居场景化应用。

E-mail: wangtao_sss@163.com

王国中, 博士、教授, 主要研究方向为中国音视频编解码标准 AVS、三网融合, 基于 AVS 和 DTMB 的双国标数字地面电视系统和 3DTV 等。

E-mail: Wanggz@shu.edu.cn

朱林林, 硕士研究生, 主要研究方向为图像处理、机器学习。

E-mail: 2528636507@qq.com