

基于非单调共轭梯度算法的声纹识别 机器人控制系统设计

吴俊杰

(泉州信息工程学院, 福建 泉州 362000)

摘要: 传统声纹识别机器人控制系统识别准确率低, 存在语音识别噪声鲁棒性问题; 针对上述问题, 基于非单调共轭梯度算法设计了一种新的声纹识别机器人控制系统, 采用 BioVoice 2.0 标准声纹采集器采集数据, 提取声纹特征, 根据提取的声纹特征建立模型库, 同时引用了两个声纹数据采集终端, 型号分别是 TMC104-B 和 TMC104, 选用型号为 AS-MrobotR 的机器人配合采集器和采集终端实现工作; 在 Windows 平台下使用 C/C++ 语言研究了一种专用的程序, 在程序内部添加 mde-api 数据库, 完成训练程序和识别程序; 实验结果表明, 基于非单调共轭梯度算法的声纹识别机器人控制系统能够很好地解决语音识别噪声鲁棒性问题, 在有噪声环境下识别准确率提高 15.24%, 在无噪声环境下识别准确率提高 21.55%。

关键词: 非单调共轭梯度算法; 声纹识别; 识别机器人; 识别控制系统

Design of Voiceprint Recognition Robot Control System Based on Non-monotone Conjugate Gradient Algorithm

Wu Junjie

(Quanzhou Institute of Information Engineering, Quanzhou 362000, China)

Abstract: The recognition accuracy of traditional voiceprint recognition system is low, and there is a problem of speech recognition noise robustness. Aiming at the above problems, a new voiceprint recognition robot control system is designed based on non-monotone conjugate gradient algorithm. The BioVoice 2.0 standard voiceprint collector is used to collect data, extract voiceprint features, and build a model library based on the extracted voiceprint features. At the same time, two voiceprint data acquisition terminals are cited, the models are TMC104-B and TMC104 respectively, and the robot with model AS-MrobotR is selected to cooperate with the collector and the collection terminal to achieve work. In the Windows platform, a special program is studied using C/C++ language. The mde-api database is added inside the program to complete the training program and the recognition program. The experimental results show that the voiceprint recognition robot control system based on non-monotone conjugate gradient algorithm can solve the problem of speech recognition noise robustness. The recognition accuracy is improved by 15.24% in noisy environment. The accuracy rate increased by 21.55%.

Keywords: non-monotone conjugate gradient algorithm; voiceprint recognition; recognition robot; recognition control system

0 引言

随着现代科技技术的飞速进步, 各类识别技术也在不断发展, 识别技术作为新兴技术, 在各个领域都发挥着自已的作用。声纹识别与人脸识别、虹膜识别和指纹识别的性质相同, 都是利用生物特征对生物的身份进行认证。相比于其它识别技术, 声纹识别技术的成本更低, 更容易被用户接受, 且在使用声纹技术时, 用户不需要在额外使用隐私信息, 整个过程安全可靠, 通过动态码来防止各种信息被篡改, 使录音能够更加完好地被拼接到一起^[1]。在移动支付、智能硬件等领域, 声纹识别系统都有所应用^[2]。

声纹识别系统内部拥有声纹识别引擎, 能够对数据进行管理和存储, 由于声纹识别系统愈来愈受到人们关注, 所以更多的学者投入到声纹识别系统的研究中^[3]。目前主要应用的研究方法有基于线性预测倒谱系数的声纹识别预测方法和基于梅尔频率倒谱系数的声纹识别预测方法^[4]。

在众多优化方法中, 共轭梯度算法是最常见的优化算法之一, 该算法计算简单, 对存储量的要求较少, 在大规模优化问题中, 共轭梯度算法的效果最为显著。共轭梯度算法的收敛速度要远远快于其它算法, 共轭梯度算法来源于共轭性质, 从 1936 年至今, 该算法已经有近百年的历史, 在很多方面都已经相当成熟^[5]。

综上所述, 本文基于非单调共轭梯度算法设计了一种声纹识别机器人控制系统, 通过加入多种传感器实现远距离控制, 利用四自由度的 AS-MRobotE 机器人平台到的软件编程接口设计实时控制系统, 设计新的控制界面, 从而更好地操控客户端和服务端^[6]。

收稿日期: 2019-11-06; 修回日期: 2019-12-10。

基金项目: 福建省教育厅中青年教育科研项目(科技类)(JT180714)。

作者简介: 吴俊杰(1980-), 男, 湖北黄梅人, 讲师, 主要从事物联网、软件工程方向的研究。

1 基于非单调共轭梯度算法的声纹识别机器人控制系统硬件设计

目前研究的声纹识别系统技术越来越先进, 因此在设计声纹识别机器人控制系统时, 必须要充分使用计算机技术, 利用互联网控制系统的整体运行, 从而提高控制系统的稳定性、可靠性和实时性^[7]。对于声纹识别机器人控制系统来说, 稳定性和可靠性是确保整个系统正常运行的关键。Linux 具有性能稳定的优点, 因此本文将其应用到控制系统中, 用来确保控制系统能够稳定运行。由于系统内部加入了大量由 Windows 平台开发的软件程序, 因此系统也加入了 Windows, 将 Windows 和 Linux 一同使用, 可以很好地提升配置资源的合理性, 提高系统的工作性能^[8]。基于非单调共轭梯度算法的声纹识别机器人控制系统硬件结构图如图 1 所示。

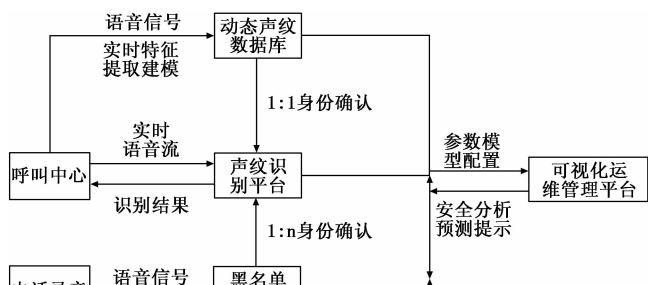


图 1 声纹识别机器人控制系统硬件结构图

1.1 BioVoice 2.0 标准声纹采集器

在声纹识别机器人控制系统中加入采集器, 用来进行声纹识别, 理解能够反映说话人生理和行为特征的语音参数, 从而确定说话人的身份^[9]。由于每个人的发声器官在尺寸和形态上有所差异, 所以声纹也会有所差异, 因此利用声纹可以很好地鉴定出说话人的身份^[10]。

使用的声纹识别采集器为 BioVoice 2.0 标准声纹采集器, 该声纹识别采集器不仅能够采集语音, 提取声纹特征, 同时也会根据提取的声纹特征建立模型库, 比对待识别的语音与所建立的模型库中的语音, 从而更好地完成识别^[11]。声纹识别采集器能够执行两个任务: 声纹采集、声纹建模。

声纹识别采集器与数据库紧密连接, 在确保采集的参数数据是准确的后, 与数据库对比, 在不同设备和不同通道上得到的声纹数据质量不同, 这会直接影响算法模型的准确性, 降低识别准确率^[12]。因此选择的声纹采集器要能够有效克服由于信道差异而带来的识别结果误差, 利用多个通道同时进行录音, 从而提高采集精度, 在采集时, 要充分分析采集内容是否完整, 标准的声纹采集器可以为后续的声纹识别、声纹鉴定和声纹比对打下坚实的基础^[13]。

BioVoice 2.0 标准声纹采集器内部拥有智能化麦克风集群, 既能够单向拾音, 也能够双向拾音, 内部加入了多种文本采集方式, 以配套的方式获得声纹信息, 在填入多标

签后加入数据库, 分类存储, 从而实现实时检索^[14]。采集器支持连续性创建采集, 可批量入库, 从而缩小被采集的数据入库的时间, 确保采集的内容真实完整, 提高声纹采集的工作效率^[15]。

BioVoice 2.0 标准声纹采集器在设计时严格遵照《声纹数据采集终端技术要求》, 通过声纹识别可以 1:1 确定身份, 实现声音辨认。

1.2 声纹数据采集终端

为了更好地实现声纹数据采集, 本文同时引用了两个声纹数据采集终端, 型号分别是 TMC104-B 和 TMC104。

TMC104-B 声纹数据采集终端的阵列单元为 4MES 麦克风, 能够频响的有效范围为 100~8 000 Hz, 使用的输出接口有两个, 分别是 USB 出口和模拟音频输出 (AUX) 出口, 采用的输出声道为双声道, 能够进行语音识别的有效范围为 5 m, 灵敏度超过 60 mV/Pa, 本机噪声级小于 30 dB, 信噪比大于 60 dB。在以 1 000 Hz 的正弦波信号为参考时, TMC104-B 声纹数据采集终端在低于 $\pm 45^\circ$ 正面入射角的范围内, 声压级会从 -7 dBFS 降低到 -9 dBFS; 当超过 $\pm 60^\circ$ 正面入射角的范围内, 声压级会从 -12 dBFS 降低到 -14 dBFS; 当超过 $\pm 90^\circ$ 反面入射角的范围内, 声压级会从 -12 dBFS 降低到 -14 dBFS, 被识别人所对应的音频没有输出信号, 因此可以从指定声道得到音频信号。

TMC104 声纹数据采集终端灵敏度超过 60 mV/pa, 本机噪声级小于 36 dB, 信噪比大于 30 dB, 能够频响的有效范围为 50~8 000 Hz, 波形失真度低于 7%, TMC104 声纹数据采集终端具有低中高多级噪声抑制功能, 以入网管理的方式管理各种数据, 采集的语音格式为 Windows PCM WAV, 以单声道的方式输入, 16 位量化精度, 采集终端的采样率 16 kHz, 能够实现静音段检测、信噪比检测、有效时长检测、截幅比例检测、平均能量检测、是否多人说话检测功能。

在以 1 000 Hz 的正弦波信号为参考时, TMC104B 声纹数据采集终端的性质与 TMC104-B 声纹数据采集终端完全一致。引入两个终端后, 系统在执行语音信号处理、声纹特征提取、声纹建模、声纹比对、判别决策等任务后, 执行能力更强, 在截取蕴含声纹特征的语音时, 更加方便自然, 使用的获取设备成本更低。

1.3 控制机器人

本文使用的机器人由上海广茂达公司生产, 型号为 AS-MrobotR, 该机器人为智能型机器人, 能够与 BioVoice 2.0 标准声纹采集器、TMC104-B 和 TMC104 声纹数据采集终端很好地配合。AS-MRobotR 模块化机器人内部拥有领先于世界的机器人技术, 具有可重构、全开放、安全可靠的特性, 相较于进口版的 AS-MRobot 系列机器人, AS-MRobotR 模块化机器人的性加密更高, 应用 AS-MRobotR 模块化机器人系统, 可以制造工业机器人、类人机器人、步行机器人、蛇行机器人。全开放的架构体系,

使机器人可以被重构。在软件中选用的操作系统为 Windows 2000/XP，使用的开发工具为 VC，机器人底层的控制工具系统设定为开放系统，用户可以随时操作机器人底层，引入新的控制算法。在底层的控制系统中加入精确运动控制模块，该控制模块拥有 2000 线高精度编码器和电机电流传感器，对于机器人的运动速度、加速度、夹持力和终点位置都有着很强的控制能力，确保机器人能够按照规定路径精确行走。

在 AS-MrobotR 机器人内部还加入了多个可扩展的配件，主要配件为全景摄像机、数字罗盘和力矩传感器，使用 CAN 总线进行通讯，设定 CAN 总线使用电压为直流电压，电压值为 24 V，在应用 CAN 总线进行通讯后，通讯产品可直接连接移动机器人 AS-R/RF/RE，与其进行配合。

2 基于非单调共轭梯度算法的声纹识别机器人控制系统应用程序设计

为了更好地控制本文使用的 AS-MrobotR 机器人，在 Windows 平台下使用 C/C++ 语言研究了一种专用的程序，在程序内部添加 mde-api 数据库，从而提高控制效果。AS-MrobotR 机器人在 VC++ 和 OpenGL 平台上进行开发，机器人内部设计的控制系统同时具有仿真能力和控制能力。

基于非单调共轭梯度算法的声纹识别机器人控制系统应用程序分为两个程序，分别是训练程序、识别程序，在训练程序中，系统建立基本的声学模型，各基本单元都在训练程序中设立，在识别程序中，将采集到的声音特征参数按照一定的标准与系统模型进行对比，从而获得更加精准地识别模型。基于非单调共轭梯度算法的声纹识别机器人控制系统应用程序工作流程如图 2 所示。

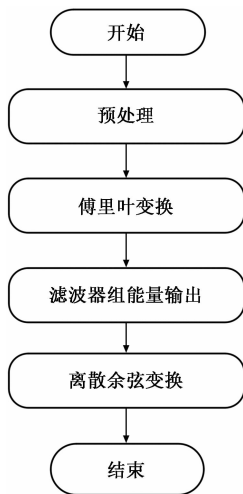


图 2 声纹识别机器人控制系统应用程序工作流程

利用麦克风采集声纹特征数据，将采集到的声纹特征数据上传给系统，使用计算机声卡对数据，使用 A/D 转换，将得到的语音原始数据存储起来，对端点进行检测，从而得到有效语音信号的起始点和终点。

在完成语音信号预处理后，进行特征提取，在短时语音帧中确定出有用的信号，使用的提取方法为非单调共轭梯度算法，该算法能够很好地求解出线性方程组和线性非方程组的数值，在给定初始量后，确定出代数特征值。由客户机和服务器端启动控制工作，客户机和服务器端的通讯过程如图 3 所示。

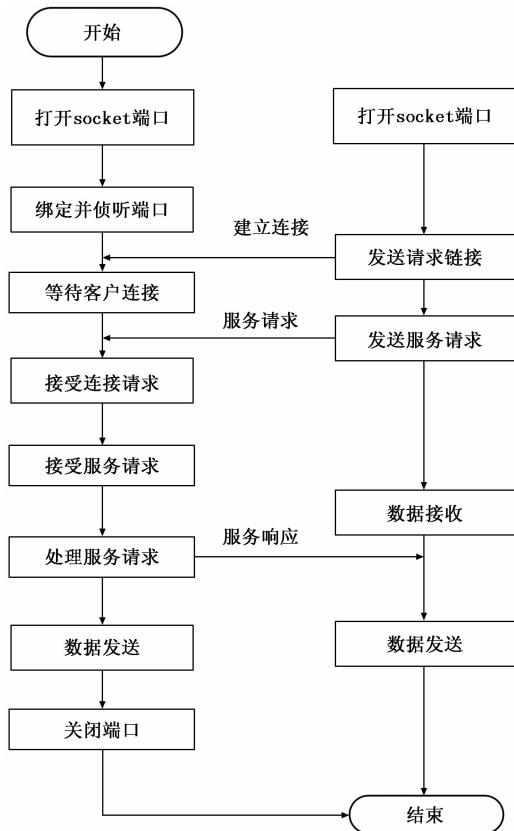


图 3 客户机和服务器端通讯过程

本文使用的 Windows 和 Linux 两个操作系统都支持 TCP 协议，因此采用标准的 socket 就可以调用，在对 socket 进行编程时，客户机/服务器会采用主动请求的方式了解信息，由 Linux 系统确定客户机的传送指令，由 Windows 系统的服务器来接收指令。

3 实验研究

3.1 实验目的

为检测本文基于非单调共轭梯度算法设计的声纹识别机器人控制系统，与传统声纹识别控制系统进行实验对比，验证系统的有效性。

3.2 实验参数

该实验的实验参数如表 1 所示。

3.3 实验结果与分析

根据上述参数，选择本文的系统 and 传统系统各自进行识别，识别次数为 10 次，识别环境分别为无噪声环境、白噪声环境、汽车噪声环境、工厂噪声环境以及 F16 机噪声环境，对识别结果进行统计。在无噪声情况下得到的实验

结果如图 4 所示。

表 1 测试实验应用参数

项目	参数
数字语音数据库	T1-46 数字语音数据库
采样频率	12.5kHz
量化	16b
训练集样本数	3000 个
测试集样本数	1000 个
特征提取模块帧长	256
特征提取模块帧移	80
窗函数	汉明窗
可见输入数据归一结果	(0,1)
正态分布随机数	(0,0.01)
计算机配置内存	4GB
双核	I5
处理器工作频率	2.67GHz
CPU 型号	NVIDIA GT540

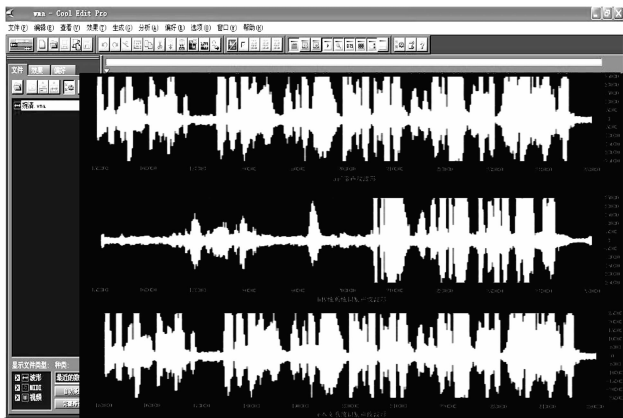


图 4 无噪声下实验结果对比图

分析图 4 可知，本文研究的数字语音信号共有 9 个，在针对 0~9 的数字语音信号进行识别时，本文研究的声纹识别机器人控制系统的控制能力始终高于传统系统，因此控制的声音识别准确率要始终高于传统系统，整个系统运行的稳定性也要更高。

在有噪声情况下进行测试，设定噪声的信噪比在 0~45 dB 之间，得到的实验结果如图 5 所示。

针对有噪声进行详细实验研究，分别在白噪声环境、汽车噪声环境、工厂噪声环境以及 F16 机噪声环境进行识别，得到的识别结果记录如下。

在白噪声情况下，对传统系统和本文系统进行测试，白噪声的信噪比在 0~45 dB 之间，得到的实验结果如图 6 所示。

分析图 6 可知，当信噪比为 10 dB 时，传统系统的识别准确率为 74.2%，本文系统的识别准确率为 70.3%；当信噪比为 20 dB 时，传统系统的识别准确率为 69.1%，本文系统的识别准确率为 94.8%；当信噪比为 30 dB 时，传统系统的识别准确率为 81.3%，本文系统的识别准确率为



图 5 有噪声下实验结果对比图

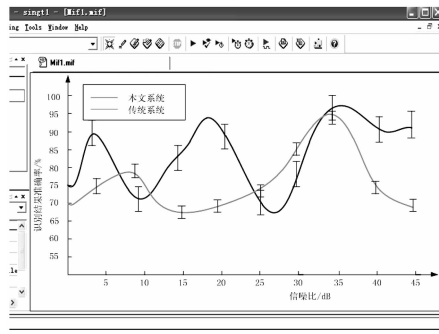


图 6 白噪声下实验结果对比图

75.1%；当信噪比为 40 dB 时，传统系统的识别准确率为 64.4%，本文系统的识别准确率为 91.7%。由上述数据可知，在不同信噪比下，本文设计的系统和传统系统的识别准确率是有所波动的，但是从整体来看，基于非单调共轭梯度算法设计的声纹识别机器人控制系统识别准确率要高于传统控制系统。

在汽车噪声情况下进行测试，汽车噪声的信噪比在 0~45 dB 之间，得到的实验结果如图 7 所示。

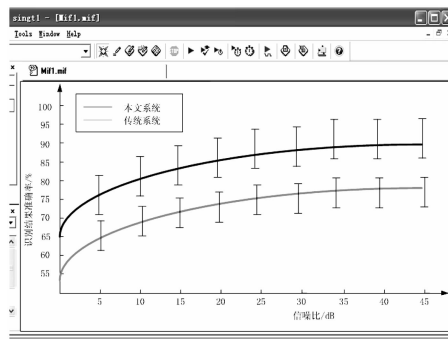


图 7 汽车噪声下实验结果对比图

分析图 7 可知，两个系统都属于比较平稳的状态。随着信噪比的增加，本文系统的识别准确率和传统系统的识别准确率逐渐增加，当信噪比为 45 dB 时，传统系统的识别准确率为 78%，本文控制系统的识别准确率为 90%。

在工厂噪声情况下进行测试，得到的实验结果如图 8 所示。