

文章编号:1006-2475(2009)11-0163-03

基于声纹识别的嵌入式防盗系统

茅剑,林奇

(集美大学计算机工程学院,福建 厦门 361021)

摘要:为了克服传统防盗锁的缺点,嵌入式防盗系统是在对声纹识别算法进行相关的研究和改进的基础上基于 ARM9 嵌入式开发的,本文结合 Linux 操作系统,选用 MINI2440 芯片,并配合外围驱动电路和电子锁开发实现。研究结果表明,该系统具有较高的语音识别率,性能稳定,有一定的应用价值。

关键词:嵌入式系统; Linux; 声纹识别; MFCC; DTW; 防盗系统

中图分类号:TP311 **文献标识码:**A **doi:** 10.3969/j.issn.1006-2475.2009.11.047

Embedded Anti-theft System Based on Voiceprint Recognition

MAO Jian, LIN Qi

(Computer Engineering College of Jimei University, Xiamen 361021, China)

Abstract: The anti-theft system is based on voice print identification algorithm. This paper uses a chip design MINI2440 to make voiceprint lock hardware system, combines with the development ARM9 embedded operating system, Linux hardware driver programming design method, and peripheral drive circuit and electronic lock module. After system testing, it can be verified to be useful in certain area.

Key words: embedded system; Linux; voiceprint identification; MFCC; DTW; anti-theft system

0 引言

传统的机械锁及 IC、TM 等电子锁,容易出现钥匙丢失、仿造、密码盗用等现象,安全性能不佳。而生物特征识别则是利用人体生物特征有异、终身不变、随身携带等特点开发的,安全性能高。人声属于内藏型身体特征,只要当事人不开口说话,则被盗用的可能性很小。同时声纹识别相比于指纹识别、掌形识别、虹膜等相关识别等其他生物识别技术,具有简便、准确、经济、扩展性良好等优势,可广泛应用于安全验证、控制等领域,如电话银行、电话炒股、电子购物等方面。

智能声纹识别保密安防锁是基于声纹识别技术,结合 ARM9 嵌入式硬件系统,开发出来的一个小型实用型嵌入式系统。作为成品它包括了声纹识别模块组件、小键盘组件、驱动电路、电子锁。

1 声纹识别系统的硬件设计

系统硬件结构如图 1 所示。

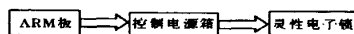


图1 系统硬件结构框图

系统开发使用 MINI2440 开发板。该 ARM 板采用 Samsung S3C2440A CPU,主频 400MHz,板上 SDRAM 内存 64M,使用 Linux 作为嵌入式操作系统。

ARM 板处理录音、训练、匹配等信息。控制电源箱接收来自 ARM 板的识别信号,向柔性电子锁提供开锁瞬间电压,并为终端电子锁提供 12V 工作电压。硬件模块之间通过电信号进行通信。

操作流程如图 2 所示。

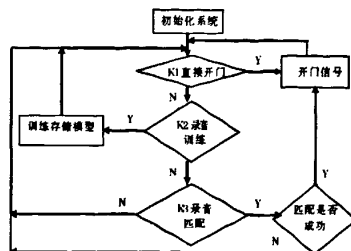


图2 系统操作流程框图

收稿日期:2009-07-13

作者简介:茅剑(1980-),男,福建沙县人,集美大学计算机工程学院助教,硕士,研究方向:嵌入式系统;林奇(1989-),男,福建福州人,本科,研究方向:嵌入式系统。

按键 K1:作为测试按键,可屏蔽,按下 K1 直接打开门锁。

按键 K2:按下 K2,LED1 号亮,开始约为 3 秒的录音,录音结束后回放录音内容,并自动进行模型训练与存储,完成后 LED1 号灭。

按键 K3:按下 K3,LED2 号亮,开始约为 3 秒的录音,录音结束后回放录音内容,并自动进行模型匹配。当匹配正确时,门锁打开;反之则无任何动作。完成后 LED2 号灭。

2 声纹识别算法设计

声纹识别算法的原理如图 3 示。

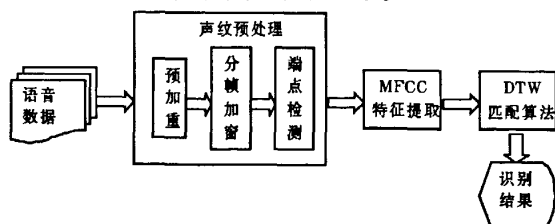


图 3 声纹识别算法的原理框图

2.1 语音数据

系统输入的声音数据格式为 44.1k 的取样频率, 16 位量化数字的 WAV 文件。按下系统采集按键后, 使用 ARM 板载麦克风采集环境语音, 以 WAV 文件格式保存在 SDRAM 中。

2.2 语音预处理

预加重目的是去噪, 语音信号的频谱上效果是高频提升, 使语音信号的变得平坦, 便于进行频谱分析或声道参数分析。预加重可以压缩信号的动态范围, 有效地提高信噪比。预加重数字滤波器的传递函数为:

$$H(z) = 1 - m \cdot z^{-1} \quad (m \text{ 的值接近 } 1) \quad (1)$$

分帧的目的是对语音信号采用短时平稳性分析, 一般每秒的帧数约为 33 ~ 100, 利用可移动的有限长度窗口进行加权的方法来实现。加窗是为了让波形不易丢失以及利于对能量的处理, 为了保证波形细节选择海明窗:

$$w(k) = 0.54 - 0.46 \cos\left(2\pi \frac{k}{N-1}\right) \quad k = 1, 2, \dots, N \quad (2)$$

端点检测是指从背景噪声中找出语音的开始和终止点, 可以减少系统的计算量, 提高系统的运行效率。端点检测采用短时能量 E_n 与过零率 Z_n 两个阈值指标:

$$E_n = \sum_{m=0}^{N-1} [s(m)w(n-m)]^2 \quad (3)$$

$$Z_n = \sum_{m=0}^{N-1} |\text{sgn}[s(m)] - \text{sgn}[s(m-1)]| * w(n-m) \\ = |\text{sgn}[s(n)] - \text{sgn}[s(n-1)]| * w(n) \quad (4)$$

2.3 MFCC 倒谱系数

MFCC 倒谱系数可以精确地表征语音的频域特性, 为 DTW 识别算法提供良好的前端输入。对于预处理之后的语音数据, 每帧提取的 12 阶的 MFCC 倒谱系数, N 帧就构成 $N \times 12$ 的二维数组, 作为下一步语音识别 DTW 算法的输入。实验证明, MFCC 比 LPCC 具有更良好的鲁棒性。MFCC 特征参数提取过程如下:

首先对一帧采样语音 $\{x(t)\} (t = 1, 2, 3, \dots, N)$ 预加重处理, 加 Hamming 窗后作 N 点 FFT, 取得信号幅度谱 $|x(k)|$ 。

其次将实际频率尺度转换为 Mel 频率尺度:

$$\text{Mel}(f) = 2595 \lg(1 + f/700) \quad (5)$$

接着在整个 Mel 轴配置 L 个三角形滤波器。第 L 个三角形滤波器的下限, 中心和上限频率 $o(L)$ 、 $c(L)$ 、 $h(L)$ 之间的关系:

$$c(L) = h(L-1) = o(L+1) \quad (6)$$

根据语音信号幅度 $|x(k)|$ 求每一个三角形滤波器的输出:

$$m(l) = \sum_{k=o(l)}^{h(l)} W_l(k) |X_n(k)|, l = 1, 2, \dots, L \quad (7)$$

$$W_l(k) = \begin{cases} \frac{k - o(l)}{c(l) - o(l)} & o(l) \leq k \leq c(l) \\ \frac{h(l) - k}{h(l) - c(l)} & c(l) \leq k \leq h(l) \end{cases} \quad (8)$$

最后对所有滤波器输出作对数运算, 再进行离散余弦变换, 可以得到 MFCC:

$$c_{mfcc}(i) = \sqrt{2/N} \sum_{l=1}^L \log m(l) \cos\left[(l-1/2) \frac{i\pi}{L}\right], i = 1, 2, \dots, Q \quad (9)$$

其中, Q 为 MFCC 参数的阶数, 本系统取 $Q = 12$, $c(i)$ 为所求的 MFCC 的参数。

2.4 动态时间规整 (DTW) 算法

要建立一个性能好的语音识别系统仅有好的语音特征还不够, 还要有好的语音识别的模型和算法。语音识别过程是根据模式匹配原则, 计算未知语音模式与语音模板库中的每一个模板的距离测度, 从而得到最佳的匹配模式。DTW 是较早的一种模式匹配和模型训练技术, 它应用动态规划方法成功解决了语音信号特征参数序列比较时时长不等的难题, 在孤立词语音识别中获得了良好的性能。基于动态时间规整匹配的 DTW 算法从目前来看, 可能是一个最为小巧的语音识别的算法, 系统开销小, 识别速度快, 在小词汇量的语音命令控制系统中是一个非常有效的算法。

DTW 思想就是把未知量均匀地伸长或缩短, 直到它与参考模式的长度一致时为止。在时间规整过程中, 未知单词的时间轴要不均匀地扭曲或弯折, 以便使其特征与模型特征对正。算法过程如下:

参考模板特征矢量序列为 $A = \{a_1, a_2, \dots, a_I\}$, 输入语音特征矢量序列为 $B = \{b_1, b_2, \dots, b_J\}$, $I \neq J$ 。为了让被测语音模板的时间轴 i , 非线性地映射到参考模板的时间轴 j , 并使总的累计失真量最小, 设时间规正函数为: $C = \{c(1), c(2), \dots, c(N)\}$, 其中 N 为路径长度, $c(n) = (i(n), j(n))$ 表示第 n 个匹配点对, 它是由参考模板的第 $i(n)$ 个特征矢量与被测模板的第 $j(n)$ 个特征矢量构成的匹配点对。两者之间的距离 (或失真值) $d(a_{i(n)}, b_{j(n)})$ 称为局部匹配距离。DTW 算法就是通过局部优化的方法实现加权距离总和最小, 即:

$$D = \min_c \frac{\sum_{n=1}^N [d(a_{i(n)}, b_{j(n)}) \cdot W_n]}{\sum_{n=1}^N W_n} \tag{10}$$

3 实验及数据分析

经过对算法的应用, 测试实验结果如表 1 ~ 表 4 所示。

表 1 相同人相同语音 (✓为结果识别, ×为结果不识别)

测试编号	1	2	3	4	5	6	7	8	9	10
预计结果	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
测试结果	✓	✓	✓	✓	✓	✓	✓	×	✓	✓

表 2 相同人不同语音 (✓为结果识别, ×为结果不识别)

测试编号	1	2	3	4	5	6	7	8	9	10
预计结果	×	×	×	×	×	×	×	×	×	×
测试结果	×	×	×	×	×	✓	✓	×	×	×

表 3 不同人相同语音 (✓为结果识别, ×为结果不识别)

测试编号	1	2	3	4	5	6	7	8	9	10
预计结果	×	×	×	×	×	×	×	×	×	×
测试结果	×	×	×	×	×	✓	×	✓	×	×

表 4 不同人不同语音 (✓为结果识别, ×为结果不识别)

测试编号	1	2	3	4	5	6	7	8	9	10
预计结果	×	×	×	×	×	×	×	×	×	×
测试结果	×	✓	×	×	×	×	×	×	×	×

为了降低误识率, 可以尝试对一个语音进行多次训练并匹配, 再根据结果求出最终值; 也可以采用二级决策来判定语音。该设计对特定人语音的识别率达到了 85%, 有一定实用价值。

4 结束语

声纹识别嵌入式系统设计的重点和难点是在如何处理大量声音数据的问题, 它不仅要考虑到声音的特征问题, 对容易造成声音干扰的噪声去除, 还要对硬件

的各个接口进行匹配和编写驱动, 而且还要考虑 Linux 操作系统录音方面的缺陷及可能遇到的问题, 这需要多学科知识综合应用。从总体上看, 该嵌入式系统实现了基于声纹识别基本防盗功能。随着硬件性能和算法质量的提高, 该系统会有广阔的应用前景。

参考文献:

[1] 张跃辉. 语音识别的抗噪性研究[D]. 南京: 南京理工大学硕士学位论文, 2006.

[2] 郭军. 语音识别中的说话人自适应研究[D]. 北京: 北京邮电大学博士学位论文, 2007.

[3] Charlet D, Jouvett D, Collin O. An alternative normalization scheme in HMM-based text-dependent speaker verification [J]. Speech Communication, 2000, 31(2-3): 113-120.

[4] Li Zhipeng, Chen Shanguang, Xue Liang. Some mistakes in the algorithm for solving underflow problem in baum welch algorithm and their corrections [J]. Chinese Journal of Acoustics, 2003(2): 154-165.

[5] Deller J P, Proakis J G, Hansen H L. Discrete Time Processing of Speech Signals [M]. Piscataway N J: IEEE Press, 2000.

[6] 王社国, 赵建光. 基于 ARM 的嵌入式语音识别系统研究[J]. 微计算机信息, 2007, 23(5): 149-150.

[7] 朱民雄, 闻新, 黄健群, 等. 计算机语音技术[M]. 北京: 北京航空航天大学出版社, 2002.

[8] 陈显治. 语音处理技术的发展[J]. 通信工程学院学报, 1996, 10(26): 50-57.

[9] Rabiner L, Juang B H. 语音识别基本原理 (影印版) [M]. 北京: 清华大学出版社, 1999.

[10] Young S, Evermann G, Gales M, et al. The HTK Book (for HTK version 3. 4) [EB/OL]. <http://www.ee.ucla.edu/~weichu/htkbook/>, 2006-12-06.

[11] Zhang Wenjun, Xie Jianying. Adaptive endpoint detection based on subband speech [J]. Journal of Systems Engineering and Electronics, 2003, 14(1): 54-57.

[12] 何致远, 胡起秀, 徐光. 两级决策的开集说话人辨认方法[J]. 清华大学学报 (自然科学版), 2003, 43(4): 516-520.

[13] 杨大利, 徐明星, 吴文虎. 语音识别特征参数选择方法研究[J]. 计算机研究与发展, 2003, 40(7): 963-969.

[14] 于世功, 田岚, 李传林. 不依赖于文本的说话人识别研究及其应用[J]. 计算机工程与应用, 2000, 36(4): 71-73.

[15] 罗俊光, 汤荣江. 基于 DSP 的非特定人语音识别系统 [J]. 微计算机信息, 2007(4): 195-196, 178.

[16] 李锦, 何培宇. 一种改进的基于小波去噪 HMM 非特定人语音识别算法[J]. 四川大学学报 (自然科学版), 2007, 44(1): 69-72.