

基于倒谱分量的融合参数应用于声纹识别

郑凯鹏,周 萍,张上鑫,柯晶晶

(桂林电子科技大学 电子工程与自动化学院,广西 桂林 541004)

摘 要: 声纹识别技术作为一种新型的生物特征认证技术,互联网的快速发展给声纹识别带来很多商业上的应用,对于声纹识别技术的研究越来越受到科学和市场的重视. 优化声纹识别算法速度和正确率的重要做法是提高语音信号特征参数的鲁棒稳健性. 因此,本文以实验室录制的语音作为信息库,利用 Mel 频率倒谱系数、差分以及加权倒谱系数三个信息进行融合,采用增减阶数法分析获取高重要度的倒谱分量,组成新的融合参数,建立基于 VQ 的声纹识别模型,采用 LBG 算法设计不同码本容量的实验. 通过仿真验证,本实验的倒谱分量重要度分析得到的融合特征参数在声纹识别上正确率得到提升.

关键词: 声纹识别;MFCC;倒谱重要度分析;VQ

中图分类号: TN912. 34

文献标识码: A

文章编号: 1000—7180(2017)08—0029—04

Base on the Fusion Parameters of Cepstrum Components Used in Speaker Recognition

ZHENG Kai-peng, ZHOU Ping, ZHANG Shang-xin, KE Jing-jing

(School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: Voiceprint recognition technology is a new kind of biological certification technology. With the rapid development of Internet research bring a lot of business applications for voiceprint recognition, the technology has get more and more attention of science and market. Extracting more robust speech feature parameters is the major aspect to develop the voiceprint recognition algorithm speed and correct rate. With the recording voice library in the laboratory, the paper extracted Mel frequency cepstrum coefficient and its difference, weighted cepstrum coefficient. Then get new hybrid parameters composed of parameter vector with high Importance based on analysis of increase and decrease component. Finally, the paper build a speaker recognition model based on the VQ and designed different codebook capacity experiments with the LBG algorithm. Experimental results show that hybrid parameters through this paper's analysis got a higher recognition rate in the speaker recognition.

Key words: voiceprint recognition; MFCC; cepstrum importance analysis; VQ;

1 引言

随着日常生活信息交流的多元化发展,身份鉴定是人们参与社会活动的重要保障^[1-2]. 对比新科技的多种人体识别技术,如指纹匹配、虹膜辨别、掌纹识别等,声纹识别有更多应用上的优点,但同时声纹识别应用上的技术难度也比其他生物识别技术大得多,主要体现在大多数生物认证技术归根结底都是已经发

展得非常成熟的图像识别技术,且同一人物对象的生物图像可变性不大. 语音信号虽然是一维信号数据,但是每次录音即使是同一个对象其提取的特征参数都存在极大的差异,每个特征参数都为最后声纹识别提供不同的重要度,而本文的目的就是比较这些特征参数重要度的大小,以提高声纹识别率.

声纹识别的研究虽然能够做到中英文词汇 95% 以上的识别率,但是他们都是建立在大数据和

大服务器上运行的,在现实离线电脑的算法测试上存在大量速度慢和识别率低的问题,实际生活中使用的问题也很多,在提取特征参数和识别匹配算法两方面仍存在未能解决的问题^[3]. 为提高声纹识别系统的整体性能,本文采取增减阶数法融合 Mel 频率倒谱系数(MFCC)、差分系数(ΔMFCC)以及加权倒谱系数(WMFCC)三个主要参数对声纹识别的重要度和稳定性分析,提取重要度和稳定性较高的融合数据进行声纹识别^[4].

2 基于倒谱分量重要度分析的融合参数

作为仿真人耳听觉机理特殊结构的特征参数, Mel 频率倒谱参数(MFCC)在很多声纹识别算法中都发挥很大的作用,大大提高了识别率. 然而 MFCC 仅仅表现的是声纹信号的静态特点,语音帧与帧之间的增加和减少趋势变化曲线的特点没能体现出来,需要将 MFCC 的进一步数据融合才能获得. 先将语音数据进行傅里叶变化后得到频谱,再进一步提取 MFCC,然后进行一阶差分变换,就可以得到ΔMFCC,ΔMFCC 可以更好的描述声纹特征参数的变化情况^[5]. 在声纹识别的研究中,对各个特征参数的融合(如 MFCC 和 LPCC)时要用到加权系数^[6],而在为了提高特征参数的有效性和稳定性,也会在参数的求解分量之前对特征进行系数加权. 提取出来的特征参数中会包含词汇信息和声纹信息,而声纹识别要注重于保留其每个人特殊的声纹信息,滤除词汇信息部分,声纹信息的特征参数经过加权以后就更能体现声纹的个性特征.

为了更精确的反映不同声纹的特征差异,加权幅度大小应该与 MFCC 的各维分量大小分布位置相符,本文采用如下加权函数:

$$r_1 = 0.5 + 0.5 * \left| \sin(2\pi * \frac{i-1}{n}) \right|$$
$$i = 1, 2, \dots, n \quad n = 24 \tag{1}$$

鉴于不同的声学特征是基于不同的模型提取得到的^[4],如基音轮廓是基于频谱模型,人耳的听觉系统可以用 MFCC 描述,因此,可以用不同类别的特征参数用多个方法描述各个使用者的声纹特性. 若只是把 MFCC、ΔMFCC 或 WMFCC 这三个数值矩阵简简单单的依序组合成一个新的矩阵,在一定程度上都会比原单一的参数矩阵声纹识别率提高,但却无可避免计算量的指数增长,这在现实应用中都是会得不偿失的,没有真正解决算法问题. 为了克服这个缺点,本文会在特征矢量中对一些特征分量进

行增加和删除,然后进行识别率的测试,通过测试的结果判断改分量对识别率的重要性^[7].

本文将利用增加阶数法来考察每一阶倒谱系数的对识别结果的影响,其详细计算过程如公式(2)所示:

$$R(i) = \frac{1}{n} \left[\sum_{j>i} (p(i,j) - p(i+1,j)) + \sum_{j<i} (p(j,i) - p(j,i-1)) \right] \tag{2}$$

公式里 $R(i)$ 代表第 i 阶倒谱系数的相对重要度, n 为倒谱阶数, $p(i,j)$ 为从第 i 阶到第 j 阶倒谱系数特征的声纹识别正确率^[8].

若相对重要度 $R(i) > 0$,则说明该特征向量会增加识别率;若相对重要度 $R(i) < 0$,则说明该特征向量会降低识别率. 当然, $R(i)$ 只能表示当前参数的相对重要度,不可以完全根据这个参数判断各个向量彼此之间的关联性. 本文利用增减阶数法解析 MFCC、ΔMFCC 和 WMFCC 每个倒谱参数矩阵与声纹识别率的重要度之间的关联,分别将这多个特征参数间对声纹识别率重要度较高的 n 维倒谱矩阵融合成高效的矩阵向量.

3 矢量化

矢量化 VQ(Vector Quantization)是在香农的采样率失真的理论基础的建立的,它是一种高效的数据压缩和编码技术^[9].

本文将 VQ 算法运用于声纹识别的基本流程是:首先将同一说话人的声纹数据转换成多个特征参数并进行训练,训练的结果是每一个说话人的声纹特征会对应一个相应的向量码本,而该码本可以用聚类算法计算出来^[9].

识别时先按照此码本对待测参数编码,与之前训练的语音做匹配,选择失真测度最小的作为识别结果,系统原理框图如图 1 所示.

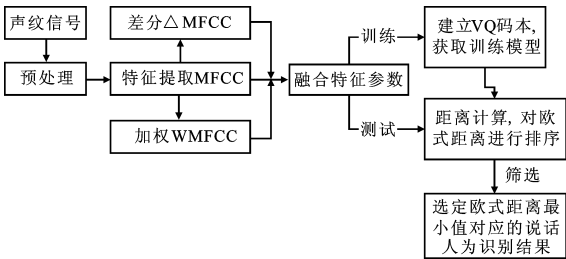


图 1 基于 VQ 融合特征矩阵的声纹识别算法流程图

4 声纹识别实验

(1)硬件环境:PC 个人计算机(AMD FX-7500 Radeon R7, 10 Computer Cores 4C + 6G 2. 10

GHz);麦克风+声卡+普通扬声器。

(2)软件环境:操作系统(Windows 7);开发平台(MATLAB R2016);录音软件(Adobe Audition CS6)。

(3)实验环境:普通实验研究室。

(4)语音库的建立:本使用使用笔记本内置麦克风录音,为了不使声纹变形,必须关闭笔记本内置硬件增强音频信号选项,在实验室相对安静的环境下录制,采用 8 KHz 的采样率,位数量化精度为 16bit,录音者(包括 15 个男性、16 个女性,共 31 人)进行约 3 分钟长度的随机朗读(与文本无关),截取训练语音与测试语音均为 3 s。

首先,将由 MFCC、 Δ MFCC 和 WMFCC 分别组成的三个参数矩阵进行声纹识别测试^[9],采用的 VQ 码本容量 $M=8,16,32,64$,实验结果如表 1。

表 1 三个参数矩阵对应码本容量的识别率 %			
M	特征参数		
	MFCC	Δ MFCC	WMFCC
8	67.74	67.74	74.19
16	80.65	80.65	83.87
32	83.87	77.42	90.32
64	80.65	83.87	83.87

按照[MFCC; Δ MFCC]、[MFCC; WMFCC]、[Δ MFCC; WMFCC]的组合将特征矩阵进行融合,运用对照性实验,识别率如表 2。

表 2 两个码本融合的特征矩阵识别率 %			
M	特征参数		
	[MFCC; Δ MFCC]	[MFCC; WMFCC]	[Δ MFCC; WMFCC]
8	67.74	70.97	77.42
16	74.19	77.42	80.65
32	80.65	77.42	87.10
64	87.10	80.65	87.10

从表 1 和表 2 中可以看出,运算 VQ 码本的声纹识别过程中,码本维度的上升可以让算法的识别率相应地上升,

然而当码本维度达到一定数量后,声纹识别率曲线开始平缓不再随之上升,同时提高码本维度会导致计算时间延长、矩阵需求内存增多等缺陷;综合上述实验说明,将多个特征矩阵进行简单的相互组合并不能实现声纹识别率的提高,因此还要将多个特征矩阵之间的融合进一步改进。

为了进一步让声纹识别率提高,并且有效的将特征矩阵融合,利用增减阶数法分析出 MFCC、 Δ MFCC 和 WMFCC 三个倒谱分量间对声纹识别

的重要度,并将特征矩阵中对声纹识别率的有所的 n 阶倒谱矩阵融合到构成新的特征矩阵,并且为减少运算的复杂度,将 VQ 的码本维度设定为 $M=16$ 。图 2~图 4 为 MFCC、 Δ MFCC 和 WMFCC 的各维分量对应的重要度图。

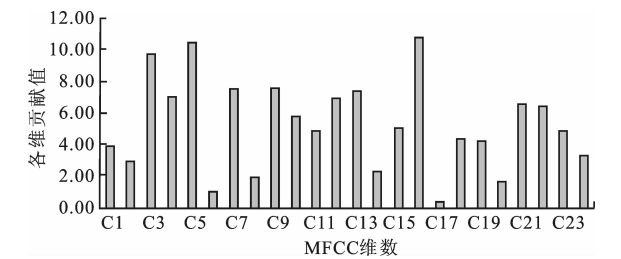


图 2 MFCC 各维分量重要度分析

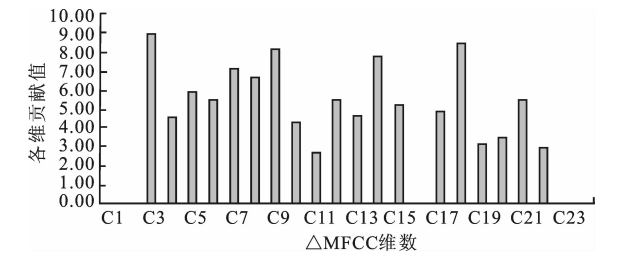


图 3 Δ MFCC 各维分量重要度分析

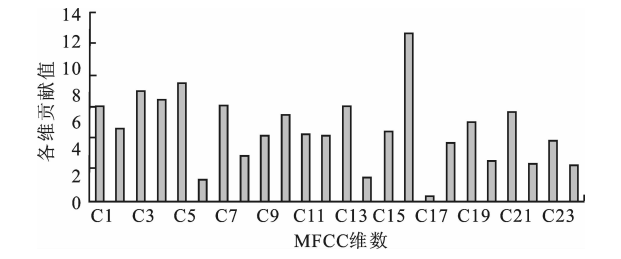


图 4 WMFCC 各维分量重要度分析

在获取到特征矩阵的每个倒谱分量对声纹识别的相对重要度后,算法把 MFCC 重要度较低的 6 维矩阵删减,使用剩下的 18 维 MFCC 作为识别算法运算的特征矩阵,依次对 Δ MFCC 和 WMFCC 做同样的处理,按照[MFCC; Δ MFCC]、[MFCC; WMFCC]和[Δ MFCC; WMFCC]三种特征融合方式这样程序就只有 36 维的特征矩阵来运用与声纹识别运算,表 3 为声纹识别结果。

表 3 三个码本容量组成 36 维特征矩阵的识别率 %			
M	特征参数		
	[MFCC; Δ MFCC]	[MFCC; WMFCC]	[Δ MFCC; WMFCC]
8	74.19	77.42	77.42
16	77.42	83.87	87.10
32	83.87	90.32	93.55
64	90.32	87.10	90.32

从表 3 可以看出,使用加权与差分融合参数大大提高声纹识别率. 这说明由 $[\Delta\text{MFCC};\text{WMFCC}]$ 得到的融合矩阵可以分析声纹信号中帧与帧之间的短时动态特性,还能有效增强个人声纹特性上的分辨率,因此本文获取的融合参数具有更高的识别率和稳定性.

5 结束语

本文的实验数据和表格都是在实验室环境下基于 matlab 的平台实现的,利用多个说话人建立声纹样本,利用 VQ 的声纹识别模型,在不同类型的的码本和码本容量下,采用单一特征参数和融合特征参数进行多个码本容量的实验参数. 在特征参数的计算上,本文采用增减阶数法计算特征参数中每个倒谱特征参数对声纹识别系统正确率的重要度,选择倒谱特征矩阵中对声纹识别重要度高的组成新的融合特征矩阵. 最后,根据实验可得出,本文获取的融合特征参数可以使得识别率提升,在以后声纹识别中将具有更好的优化与应用.

参考文献:

[1] Garg A, Sharma P. Survey on acoustic modeling and feature extraction for speech recognition[C]//Computing for Sustainable Global Development (INDIA-Com), 2016 3rd International Conference on. New Delhi, India,IEEE, 2016; 2291-2295.

[2] 孙冬梅, 裘正定. 生物特征识别技术综述[J]. 电子学报, 2001, 29(12): 1744-1748.

[3] Shabani S, Norouzi Y. Speech recognition using prin-

cipal components analysis and neural networks[C]//Intelligent Systems (IS), 2016 IEEE 8th International Conference on. Subotica, Serbia,IEEE, 2016; 90-95.

[4] 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.

[5] 柯晶晶, 周萍, 景新幸, 等. 差分和加权 Mel 倒谱混合参数应用于说话人识别[J]. 微电子学与计算机, 2014 ,31(9): 88-91.

[6] 申小虎, 万荣春, 张新野. 一种改进动态特征参数的话者语音识别系统[J]. 计算机仿真, 2015, 32(4): 154-158.

[7] MarkoviC D, Popat J, Antonacci F, et al. An informed separation algorithm based on sound field mapping for speech recognition systems[C]//Acoustic Signal Enhancement (IWAENC), 2016 IEEE International Workshop on. Xi'an, China, IEEE, 2016; 1-5.

[8] 胡航. 现代语音信号处理[M]. 北京: 电子工业出版社, 2014.

[9] 张雪英. 数字语音处理及 MATLAB 仿真[M]. 北京: 电子工业出版社, 2012.

作者简介:

郑凯鹏 男, (1992-), 硕士研究生. 研究方向为语音识别. E-mail: 1049771392@qq. com.

周 萍 女, (1961-), 硕士, 教授. 研究方向为语音识别与智能控制.

张上鑫 男, (1992-), 硕士研究生. 研究方向为语音识别.

柯晶晶 女, (1988-), 硕士研究生. 研究方向为语音信号处理和说话人识别技术语音识别.

(上接第 28 页)

作者简介:

徐德智 男, (1963-), 博士后, 教授. 研究方向为 Web 计算、语义数据库.

易晓媛(通讯作者) 女, (1991-), 硕士研究生. 研究方向为语义网、本体映射. E-mail: 624906884@qq. com.

汤 哲 男, (1977-), 副教授. 研究方向为人工智能.