第 37 卷 第 4 期
Vol.37 No.4
计 算 机 工 程
Computer Engineering
2011 年 2 月
February 2011
·人工智能及识别技术· 文章编号：1000—3428(2011)04—0216—02 文献标识码：A 中图分类号：TP18

融合 LPC 与 MFCC 的特征参数

张学锋[1]，王 芳[1]，夏 萍[2]
(1. 安徽工业大学计算机学院，安徽 马鞍山 243002；
2. 中北大学机械工程与自动化学院，太原 030051)

摘 要： 在线性预测系数(LPC)的基础上，借鉴美尔倒谱系数(MFCC)计算方法，对 LPC 进行美尔倒谱计算，得到一种新的特征参数：线性预测美尔倒谱系数(LPMFCC)。在 Matlab7.0 平台上实现一个基于隐马尔可夫模型(HMM)的说话人识别系统，分别用 LPMFCC 及其一阶差分、MFCC 及其一阶差分和基于小波包分析的特征参数(WPDC)及其一阶差分作为识别参数进行对比实验。结果表明，以 LPMFCC 作为特征参数的系统具有较高的识别率。

关键词： 线性预测；美尔倒谱系数；说话人识别

Feature Factor Fused on LPC and MFCC

ZHANG Xue-feng[1], WANG Fang[1], XIA Ping[2]
(1. School of Computer Science, Anhui University of Technology, Maanshan 243002, China;
2. School of Mechanical Engineering and Automatic, North University of China, Taiyuan 030051, China)

【Abstract】 This paper obtains a new feature factor which is called linear prediction Mel Frequency Cepstral Coefficient(MFCC) based on the Linear Prediction Coefficient(LPC) and MFCC. In order to verify the validity of the LPMFCC, a speaker recognition system based on the HMM model using the Matlab7.0 is established. The speech recognition rate of LPMFCC and the other two feature factors are tested. The result suggests that the new feature factor is not only efficiency, but also better than other two feature factors. It can achieve high recognition rate.

【Key words】 linear prediction; Mel Frequency Cepstral Coefficient(MFCC); speaker recognition
DOI: 10.3969/j.issn.1000-3428.2011.04.078

1 概述

说话人识别是语音识别的一个分支，是利用表征说话人个性的特征参数来对说话人进行辨认的一种身份认证技术。它包括 2 个方面的内容：说话人辨认和说话人确认。说话人辨认是把要检测的语句判为 N 个训练说话人之一所说，是一个多选一的问题；说话人确认则是把待检测的语句与其参考说话人相比较，相符的即得到肯定(确认)，不相符的则得到否定(拒绝承认)，是二选一的问题。说话人辨认根据说话内容是否限制又分为文本无关和文本相关 2 类。特征提取是说话人识别中的一个重要步骤，特征参数的好坏直接影响到识别结果的正确率。

目前，在说话人识别中，频谱包络特征特别是倒谱特征用的比较多，主要有线性预测倒谱系数(LPCC)、美尔倒谱系数(MFCC)。在此基础上，人们提出了许多改进算法，如利用小波变换代替快速傅立叶变换和三角滤波器组提取的 WPDC 参数。LPCC 是通过对 LPC 进行复倒谱运算得到的，对噪声敏感，在存在外界干扰时识别率就会大大下降。MFCC 利用了人耳的听觉原理和倒谱解的相关特性，对噪声的敏感程度不如 LPCC。本文在 LPC 分析[1]的基础上，融入 MFCC 提取过程中模拟人耳听觉机理的特点，寻找到既具有 LPC 分析优点又同时具有了 MFCC 较好鲁棒性和利用人耳听觉原理的特点的新的特征参数。

本文在研究 MFCC、WPDC 和 LPC 算法原理的基础上，提出把 MFCC 计算过程应用于 LPC 上的新思路，进而得到一种新的特征参数：线性预测美尔倒谱系数(LPMFCC)。通过实验，证明了该方法的有效性且具有优于 MFCC 和 WPDC 的识别率。

2 LPC 分析

实际语音信号处理中最常用的模型是全极点模型，全极点模型可用线性预测分析的方法估计模型参数。线性预测分析的基本思想是：用过去 p 个样点值来预测现在或未来的样点值。设 $P\{x(n)\mid n=0,1,\cdots,N-1\}$ 为一帧语音采样序列，则第 n 个语音样点值 $s(n)$ 的 p 阶线性预测值为：

$$\widehat{s(n)}=\sum_{i=1}^{p}a_i s(n-i) \tag{1}$$

其中，p 为预测阶数；$a_i(i=1,2,\cdots,p)$ 是线性预测系数。

预测误差 $\varepsilon(n)$ 为：

$$\varepsilon(n)=s(n)-\widehat{s(n)}=s(n)-\sum_{i=1}^{p}a_i s(n-i) \tag{2}$$

这样就可以通过在某个准则下使预测误差 $\varepsilon(n)$ 达到最小值的方法来决定惟一的一组线性预测系数 $a_i(i=1,2,\cdots,p)$，这个准则通常采用均方误差准则。

某一帧内的短时平均预测误差定义为：

$$E\{\varepsilon^2(n)\}=E\{[s(n)-\sum_{i=1}^{p}a_i s(n-i)]^2\} \tag{3}$$

为使 $E\{\varepsilon^2(n)\}$ 最小，对 a_j 求偏导，并令其为 0，有

$$E\{[s(n)-\sum a_i s(n-i)]s(n-j)\}=0, j=1,2,\cdots,p \tag{4}$$

基金项目： 安徽省教育厅青年教师基金资助项目(2008jq1032)；安徽省教育厅基金资助一般项目(KJ2009B143Z)
作者简介： 张学锋(1978－)，男，副教授、博士，主研方向：模式识别，计算机仿真；王 芳、夏 萍，硕士研究生
收稿日期： 2010-06-24 **E-mail：** zxf_5218@163.com

上式表明采用最佳预测系数时，预测误差 $\varepsilon(n)$ 与过去的语音样点正交。对于一帧从 n 时刻开窗选取的 N 个样点的语音段 $s(n)$，记为：

$$\Phi_n(j,i)=E\{s_n(m-j)s_n(m-i)\} \tag{5}$$

对于语音段 $s(n)$，它的自相关函数为：

$$R_n(j)=\sum_{n=j}^{N-1}s_n(n)s_n(n-j),\ p=1,2,\cdots,p \tag{6}$$

自相关函数是偶函数且满足 $R_n(j-i)$ 只与 j 和 i 的相对大小有关，比较式(5)和式(6)可以定义 $\Phi_n(i,j)$ 为：

$$\Phi_n(i,j)=\sum_{m=0}^{N-1-|i-j|}s_n(m)s_n(m+|i-j|) \tag{7}$$

因此，有：

$$\sum_{i=1}^{p}a_iR_n(|i-j|)=R_n(j),\ j=1,2,\cdots,p \tag{8}$$

由式(8)可得 p 个方程，写成矩阵形式为：

$$\begin{pmatrix} R_n(0) & R_{n(1)} & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & \cdots & R_n(p-2) \\ \vdots & \vdots & & \vdots \\ R_n(p-1) & R_n(p-2) & \cdots & R_n(0) \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}=\begin{pmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{pmatrix} \tag{9}$$

由这 p 个方程，可以求出 p 个预测系数 a_i。上式方程左边的矩阵称为托普利兹矩阵(ToepLitz)，可用莱文逊-杜宾(Levinson-Durbin)递推算法[2]求解 p 个预测系数。

通过 LPC 分析，由若干帧语音可以得到若干组 LPC 参数。每组参数形成一个描绘该帧语音特征的矢量，即 LPC 特征参量。由 LPC 特征参量可以进一步得到很多派生参数，例如线性预测倒谱系数、线谱对特征、部分相关系数等。这里把 MFCC 计算过程应用于 LPC 参数，得到线性预测美尔倒谱系数。

3 Mel 倒谱系数

Mel 倒谱系数最重要的特点就是利用了人耳的听觉原理和倒谱的解相关的特性。另外，Mel 倒谱也具有对卷积性信道失真进行补偿的能力。由于这些原因，Mel 倒谱被证明是在语音相关的识别任务中应用最成功的特征描述之一[3]。其提取计算过程[4-5]如下：

(1)原始语音信号经过预加重、分帧、加窗等处理，得到每个语音帧的时域信号 $x(n)$，然后经过离散傅里立叶变换后得到离散频谱 $x(k)$。设语音信号的 DFT 为：

$$x(k)=\sum_{n=0}^{N-1}x(n)e^{-j2\pi nk/N},\ 0\leqslant k\leqslant N \tag{10}$$

然后取频谱模的平方 $|x(k)|^2$ 得到离散能量谱。其中，$x(n)$为输入的语音信号；N 表示傅立叶变换的点数。

(2)将上述离散能量谱通过 Mel 频率滤波器组，计算每个滤波器的输出对数能量：

$$s(m)=\ln[\sum_{k=0}^{N-1}|x(k)|^2H_m(k)],\ 0<m<M \tag{11}$$

(3)经离散余弦变换(DCT)得到 MFCC 系数：

$$c(n)=\sum_{m=0}^{N-1}s(m)\cos(\pi n(m-0.5)/M),\ 0<n<M \tag{12}$$

4 线性预测美尔倒谱系数

线性预测系数反映了语音信号的特性，可以作为语音信号特征参数用于语音识别，但它对噪声特别敏感。人的耳朵能够从嘈杂的背景噪声中听到语音信号并能加以辨认说话人的身份，这是因为人的耳朵对不同的频率，相应的临界带宽内的信号引起的基础膜振动的位置不同。Mel 倒谱系数正是利用了人耳听觉的频率的非线性特性，在噪声的环境会有比 LPC 系数更好的鲁棒性。把这 2 种方法融合起来得到的线性预测美尔倒谱系数即可以描述说话人的语音特征，同时又具有同 MFCC 系数一样的鲁棒性，应用于识别系统会有更好的识别效果。

LPMFCC 计算过程如下：

(1)语音信号 $x(n)$经过预加重、分帧、加窗，计算每一帧的 LPC 系数 α，α 的长度与一帧语音信号的长度相等。

(2)每帧的 LPC 系数经过离散傅立叶变换(DFT)得到离散频谱 $x_a(k)$。然后取频谱模的平方 $|x_a(k)|^2$ 得到离散能量谱。

(3)将上述能量谱经过三角滤波器组滤波，计算每个三角滤波器的输出对数能量 E_i。

(4)将 E_i 经过离散余弦变换得到 LPMFCC 系数。

5 WPDC 系数及各参数的差分计算

WPDC 系数[6]是根据小波包的多分辨分析，利用小波包变换代替 MFCC 提取过程中 FFT 与 Mel 滤波器组所得到的一种系数。因为人耳对语音信号的感知是非线性的，所以 Mel 滤波器组采用一种非线性的频率单位(Mel 频率)，以模拟人的听觉系统。在 WPDC 系数提取过程中就是利用 MEL 频率划分的指导思想来选取小波包分析后的结点频带，使选出的小波结点的频带与 MEL 滤波器组频带范围相一致。

WPDC 系数、MFCC 系数和 LPMFCC 系数都用下式计算差分系数：

$$\Delta C(i)=-2C(i-2)-C(i-1)+C(i+1)+2C(i+2) \tag{13}$$
$$3\leqslant i\leqslant N-2$$

其中，$C(i)$ 表示第 i 帧的系数；$\Delta C(i)$ 表示第 i 帧的差分系数；N 表示原始语音的总的帧数。

6 实验验证及结果分析

本文实验采用 Matlab7.0 平台实现，以短时能量和平均过零率相结合的两级阈值判别法进行端点检测，HMM 为识别模型。利用语音工具箱里的 LPC 函数直接求取各帧的 LPC 系数。采用实验室录音环境，录音软件为 Windows XP 自带录音机，采样频率为 44.1 kHz，PCM 方式，量化精度为 16 bit。共有 10 位说话人，每人任意录 16 句话，每句话约为 4 s~ 6 s 长度，每句话保存为一个 WAV 格式的语音文件，每个人的 16 句录音 5 句用于训练，11 句用于识别。识别和训练的语句不重叠。分别用 LPMFCC 及其一阶差分、MFCC 及其一阶差分和 WPDC 及其一阶差分作为识别参数进行对比实验，实验结果如表 1 所示。

表 1 各特征参数下的系统识别率 (%)

说话人	LPMFCC	MFCC	WPDC
第 1 个人	90.9	45.5	54.5
第 2 个人	90.9	81.8	90.9
第 3 个人	100.0	81.8	72.7
第 4 个人	90.9	72.7	81.8
第 5 个人	81.8	81.8	81.8
第 6 个人	100.0	100.0	90.9
第 7 个人	81.8	72.7	81.8
第 8 个人	72.7	63.6	63.6
第 9 个人	81.8	81.8	72.7
第 10 个人	100.0	72.7	81.8
平均识别率	89.08	75.44	77.25

实验中 WPDC 系数的提取过程用 db6 小波进行 8 层分解。结点频带选取如表 2 所示。

(下转第 229 页)

算法取得了更好的效果。在 Welsh 算法中，对第 1 幅图像而言，Welsh 算法中灰度图像的树木颜色匹配到彩色图像中的草坪颜色，而本文算法则很好地将彩色图像中的树木颜色传递过来了。第 2 幅图像出现许多“麻点”，主要原因为 Welsh 算法是针对像素点进行匹配的，而像素点之间特征提取和相似性度量对此产生较大影响，出现了像素点的误匹配。本文算法克服了这一缺点，彩色化后的图像颜色过渡更加自然，彩色化的效果得到很大提高。

算法的具体运行时间如表 1 所示。由表 1 可知，基于 SOFM 的颜色传递算法，其颜色传递时间较 Welsh 算法有较大提高，但如果加入样本图像的训练时间，其时间消费将是很大一笔开销。但是，如果能够建立好样本图像库，存储好相应的网络结构及参数，则该算法实用性将会获得极大提高。特别是在以一幅图像为参考样本图像的情况下，批处理大量的灰度图像时，算法的速度具有绝对的优势，这也是本类算法的一个发展趋势。

表 1 颜色传递时间 s

图 3 中的图像	本文算法		Welsh 算法
	训练时间	颜色传递时间	
第 1 幅图像(149×159)	108.578	1.641	61
第 2 幅图像(258×168)	183.860	2.750	190

4 结束语

本文提出了一种基于自组织特征映射神经网络的灰度图像彩色化算法，算法采用多维的邻域特征向量作为像素点的特征，能够更好地反映像素的特征。利用 SOFM 神经网络，解决了多维特征向量搜索匹配速度的问题，同时，网络的自组织特性更好地反映了像素的内在特征，使彩色化的效果得到很大提高。本算法特别适合于在以一幅图像为参考样本图像时，批处理灰度图像的彩色化。进一步改进后，可以应用于视频图像的彩色化处理。

参考文献

[1] Ruderman D L, Cronin T W, Chiao C C. Statistics of Cone Responses to Natural Images: Implications for Visual Coding[J]. Journal of the Optical Society of American, 1998, 15(8): 2036-2045.

[2] Welsh T, Ashikhmin M, Mueller K. Transferring Color to Greyscale Images[J]. ACM Trans. on Graphics, 2002, 21(3): 277-280.

[3] Abadpour A, Kasaei S. An Efficient PCA-based Color Transfer Method[J]. Journal of Visual Communication and Image Representation, 2007, 18(1): 15-34.

[4] 李志永，滕升华，杜 坤，等. 基于不平度颜色混合的图像彩色化方法[J]. 电子与信息学报, 2008, 30(3): 514-517.

[5] 张 引，饶 娜，张三元，等. 自动采集样本的图像颜色传递算法[J]. 中国图象图形学报, 2005, 10(10): 1258-1263.

[6] 钱小燕，肖 亮，吴慧中. 模糊颜色聚类在颜色传输中的应用[J]. 计算机辅助设计与图形学学报, 2006, 18(9): 1332-1336.

[7] 马大伟，敬忠良，孙韶媛，等. 基于彩色传递的红外与可见光图像融合方法[J]. 计算机工程, 2006, 32(14): 172-173, 232.

[8] 滕秀花，陈昭炯，叶东毅. 基于多维特征向量及 ANN 技术的色彩传递算法[J]. 计算机应用, 2006, 26(12): 2866-2868.

[9] 赵国英，李 华. 人体脸部灰度图像上色的改进算法[J]. 计算机辅助设计与图形学报, 2004, 16(8): 1051-1056.

编辑 顾逸斐

(上接第 217 页)

表 2 小波包分析后的结点选取

序号	结点	频带范围/Hz
1	(8, 0)	0~86
2	(8, 1)	86~172
3	(8, 2)	172~258
4	(8, 3)	258~344
5	(8, 5)	430~516
6	(8, 6)	516~602
7	(8, 8)	688~774
8	(8, 9)	774~860
9	(8, 11)	946~1 032
10	(7, 6)	1 032~1 204
11	(7, 7)	1 204~1 376
12	(6, 4)	1 376~1 720
13	(6, 5)	1 720~2 064
14	(6, 6)	2 064~2 408
15	(6, 7)	2 408~2 752
16	(6, 8)	2 752~3 096
17	(6, 9)	3 096~3 440
18	(6, 11)	3 784~4 128
19	(5, 7)	4 816~5 504
20	(5, 8)	5 504~6 192
21	(5, 9)	6 192~6 880
22	(4, 6)	8 256~9 632
23	(3, 3)	9 632~11 008
24	(3, 4)	11 008~13 760

由以上实验结果可得出如下结论：LPMFCC 及其一阶差分作为特征参数的识别率比 MFCC 及其一阶差分和 WPDC 及其一阶差分作为识别参数的平均识别率有所提高。相比 MFCC 提高了 13.64%，比 WPDC 提高了 11.83%。而且 LPMFCC 系数的提取过程吸收了 MFCC 提取过程中的优点，具有同 MFCC 系数一样的鲁棒性。LPMFCC 系数比 MFCC 系数多了一步 LPC 系数的计算，在一定程度上增加了计算量。

7 结束语

本文在对 LPC 分析以及 MFCC 和 WPDC 特征参数的分析过程中，吸取 MFCC 系数的优点，将反映人耳听觉特性的计算 MFCC 的计算过程应用在 LPC 上，得到一种新的特征参数 LPMFCC，相应的新的特征参数也同时具有了与 MFCC 系数相同的鲁棒性。实验表明，该特征参数不仅能有效表征说话人的个性特征，而且同其他 2 种参数相比，还具有较高的识别率。

参考文献

[1] 赵 力. 语音信号处理[M]. 北京：机械工业出版社, 2003.

[2] 范新伟，申瑞民，杜彦蕊. 用 LPC 及 DTW 进行语音模式比较的设计与实现[J]. 计算机工程, 2004, 30(1): 126-128.

[3] Quatieri T F. 离散时间语音信号处理——原理与应用[M]. 赵胜辉，刘家康，谢 湘，等，译. 北京：电子工业出版社, 2004.

[4] 郭春霞，裘雪红. 基于 MFCC 的说话人识别系统[J]. 电子技术, 2005, (11): 53-56.

[5] 王让定，柴佩琪. 语音倒谱特征的研究[J]. 计算机工程, 2003, 29(13): 31-33.

[6] 胡文吉，王让定. 基于小波包分析的特征参数提取[J]. 宁波大学学报, 2007, 20(1): 51-54.

编辑 金胡考