文章编号:1672 - 058X(2012)02 - 0052 - 06

基于改进 MFCC 的异常声音识别算法*

贺玲玲,周 元
(重庆工商大学 计算机科学与信息工程学院,重庆 400067)

摘 要:在声音识别系统中,特征参数的获取对声音识别和训练有着重要的影响;MFCC 算法作为典型的声音特征参数提取方法,性能稳定,识别率高;针对 MFCC 算法存在较大计算量的情况,提出一种改进的特征参数提取算法 MFCC_E;相比于标准的 MFCC 算法,MFCC_E 算法减少了约 50% 的运算量,并且易于硬件实现;实验结果表明,MFCC_E 算法与 MFCC 算法的识别率大致相同,而计算复杂度却小很多。

关键词:声音识别;特征提取;MFCC;MFCC_E;GMM
中图分类号:TP393 **文献标志码**: A

在目标跟踪领域,视频和音频数据是最关键的两类信息,而在过去的 20 a 中,视频跟踪一直处于主导地位,但是当被观测目标离开观测范围时,视频跟踪的性能将会大幅度降低。与视频监控相比,声学传感器具有成本低、体积小、效率高等优点,而且声音信号随时间变化比较慢,采集到的信号比较稳定和可靠。因此,声音识别、声目标定位成为近年来的研究热点[1]。

在音频跟踪领域,声音特征参数和分类器的选择直接影响跟踪系统的复杂度和识别性能。其中对特征参数选取经典的算法主要有线性预测系数(Linear Prediction Coefficient,LPC)、倒谱参数(Linear Prediction Cepstrum Coefficient,LPCC)及基于人耳听觉特性的梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient,MFCC)等[2-4]。但这些算法应用都需要经过大量的计算,不仅提高了成本,限制了其应用范围,更为重要的是降低了其硬件实现的可行性。

通过对标准 MFCC 算法的分析研究,提出一种改进的算法提取声音的特征参数,得到了较好的识别效果。与单独采用 MFCC 系数做特征参数相比,识别率相当,计算量却有明显降低,且该算法更适合于硬件实现。

1 声音识别原理

语音识别本质上是一种模式识别过程,其基本结构原理如图 1 所示,主要包括语音信号预处理、特征参数提取、特征建模(建立参考模式库)、模式匹配等几个功能模块。

一个声音识别系统主要包括训练和识别两个阶段。无论是训练还是识别,都需要对输入的原始声音进行预处理,并进行特征提取。

收稿日期:2011 - 07 - 08;**修回日期**:2011 - 09 - 06.
* **基金项目**:重庆市科委重大攻关项目(CSTC,2010AA2036);重庆市教委项目(KJ100709).
作者简介:贺玲玲(1975-),女,四川自贡人,讲师,从事软件开发研究.

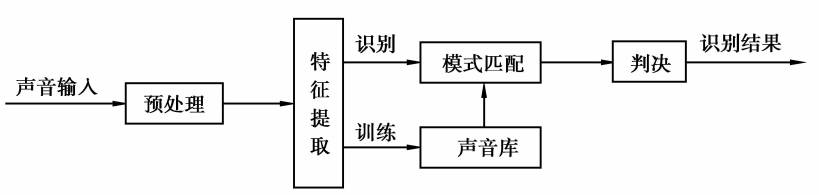

图 1　声音识别系统框图

1.1　预处理

声音信号的预处理主要包括滤波、A/D 转换、预加重、分帧、端点检测等,假设经 A/D 转换后的数字音频信号为 $x(n)$,预处理过程如下[5]:

1) 归一化处理。归一化处理的目的是消除不同样本声音之间的大小差异,将样本幅度值限定在[-1, +1]范围内。

2) 预加重。预加重通常使用具有 6 dB/倍频程的一阶数字滤波器来实现,如式(1)所示:

$$H(z) = 1 - \mu z^{-1} \tag{1}$$

其中 μ 为常数,通常取 0.97。

3) 对音频信号进行加窗分帧。虽然声音信号是非线性时变信号,但它具有短时平稳的特点,对其进行分帧可以提取其短时特性。通常取帧长为 10 ~ 30 ms,为了避免帧与帧之间的特性变化过大,帧移通常取帧长的 1/2,即相邻帧之间有 1/2 的重叠数据。为了进行短时分析,必须通过加窗来选取窗口内的声音信号,窗口外的声音信号为 0,最常用的窗口函数是汉明窗。一般取 256 点为一帧,帧间重叠为 128 点。

1.2　特征参数提取

声音特征的选择取决于具体的系统,比较有代表性的特征包括幅度(或功率)、过零率、线性预测系数特征矢量(LPC)、LPC 倒谱特征矢量(LPCC)、梅尔倒谱系数(MFCC)等。特征提取完成对声音信号进行分析处理,去掉与声音识别无关的冗余信息,获得影响声音识别的重要信息。由于倒频谱(cepstrum)有着能将频谱上的高低频分开的优点,因此被广泛地应用在声音识别领域,如 LPCC 和 MFCC。

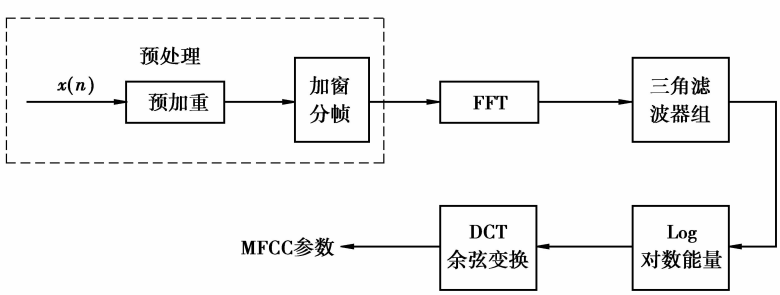

图 2　MFCC 参数提取过程

由于声音信号在时域上的变化快速且不稳定,通常将它转换到频域上来分析其特征参数。梅尔倒频谱特征参数提取过程如图 2 所示[6]。预处理后的声音数据经过快速傅里叶变换(FFT),计算出每帧数据的频谱参数,再将每帧数据的频谱参数通过一组 N(N 的值通常为 20 ~ 40)个三角形带通滤波器构成的梅尔频率滤波器做卷积运算,之后对每个频带的输出取对数,求出每个输出的对数能量(Log Energy)$S(m)$,$m = 1,2,3,\cdots,N$。最后对此 N 个参数进行离散余弦变换,求出梅尔倒谱系数作为声音特征参数,如公式(2)。

$$C_i(n) = \sum_{m=1}^{M} S(m)\cos\left[\frac{\pi n(m-0.5)}{M}\right], 0 \leqslant m < M \tag{2}$$

其中,n 为所取 MFCC 个数;$C_i(n)$ 为第 i 帧的第 n 个 MFCC 系数;$S(m)$ 为音频信号的对数功率谱;M 为三角滤波器个数[5]。

1.3 训练与识别

声音识别系统需要建立声音训练库,对样本声音进行预处理和特征参数提取,并通过分类器训练声音库。普遍采用的分类器有支持向量机(Support Vector Machine,SVM)[7]、隐马尔科夫模型(Hidden Markov Model,HMM)[8]和高斯混合型(Gaussian Mixture Model,GMM)[9]等。

GMM 本质上是一种基于参数估计的多维概率统计模型,它认为每一种声音的特征在特征空间中都形成特定的分布,并且可以用多个高斯分布组合对它的特征分布进行拟合。不同参数的高斯分布组合可以用来表征不同的声音,即每种声音的特征参数对应一个 GMM。

GMM 已经广泛应用于说话人识别和语音识别中,其训练过程是按照文献[10]描述的方法,采用从训练样本中提取特征参数矢量来训练 GMM,对于有多种异常声音的声音识别系统,每种声音用一个 GMM 来代替,得到每种异常声音的模型参数,最终得到描述每种声音的整个 GMM 的三元式,如公式(3)。

$$\lambda = \{P_i, \mu_i, \Sigma_i\}, i = 1, 2, \cdots, N \tag{3}$$

其中,P_i 为混合分量的权值;μ_i 为均值矢量;Σ_i 为协方差矩阵;N 为混合阶数。

识别过程是采用从测试样本中提取的特征矢量,结合 GMM 分类器,通过求取后验概率的最大值得到每类测试样本的识别结果,最后将每一类所有测试样本的识别结果相加,求出每类声音的总体识别率。

2 改进算法

在特征参数提取过程中,预加重、加窗、分帧、FFT、滤波器组、求对数能量、余弦变换等环节都包含大量的乘法操作,而 FFT 过程所处理的乘法运算量占整个处理过程的绝大部分。大量的乘法运算导致系统处理能力要求高、能耗大、稳定性降低、适用范围窄。

2.1 改进特征参数提取过程

将预处理过程的分帧环节调整到三角形滤波器组之后,改进后的 MFCC 参数提取过程如图 3 所示,改进后的算法称为 MFCC_E。

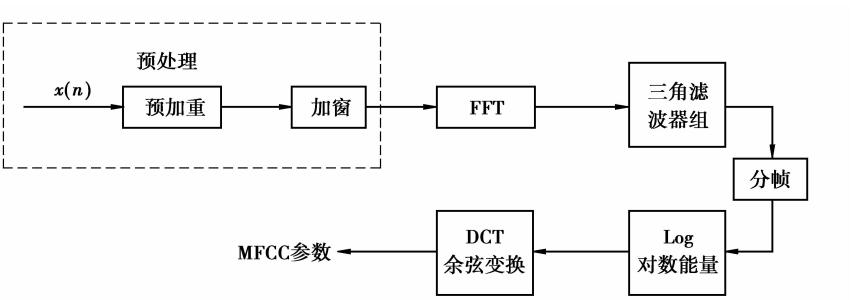

图 3 MFCC_E 参数提取过程

本质上看,加窗环节本身就带有分帧功能,一个窗口提取一帧声音数据。设置窗口长度,使得每帧声音数据包含 128 个点,连续帧间没有重叠数据,因此,尽管帧长变短,总的声音帧数与原算法保持一致。

改进后的工作原理如图 4 所示。f_n,f_{n+1},f_{n+2}为原算法的声音帧,帧长 256 点,帧间重叠 128 点。s_n,s_{n+1},s_{n+2}为 MFCC_E 的加窗模块输出帧,帧长 128 点,帧间没有重叠,经 FFT 运算和 Mel 滤波后的输出为 $S_{n,k}$,$S_{n+1,k}$,$S_{n+2,k}$,相邻输出叠加得到 Log 对数能量谱 $F_{n,k}$,$F_{n+1,k}$。FFT 运算和 Mel 滤波方法与原 MFCC 算法相同。然而,因为帧长只有原来算法的一半,由 256 个点变成 128 个点,FFT 模块只需要对这 128 个点的声音帧进行运算,运算量只有原算法的一半。同理,Mel 滤波的运算量也只有原算法的一半。因此,将分帧模块放到 Mel 滤波之后,可减少 50% 的计算量。

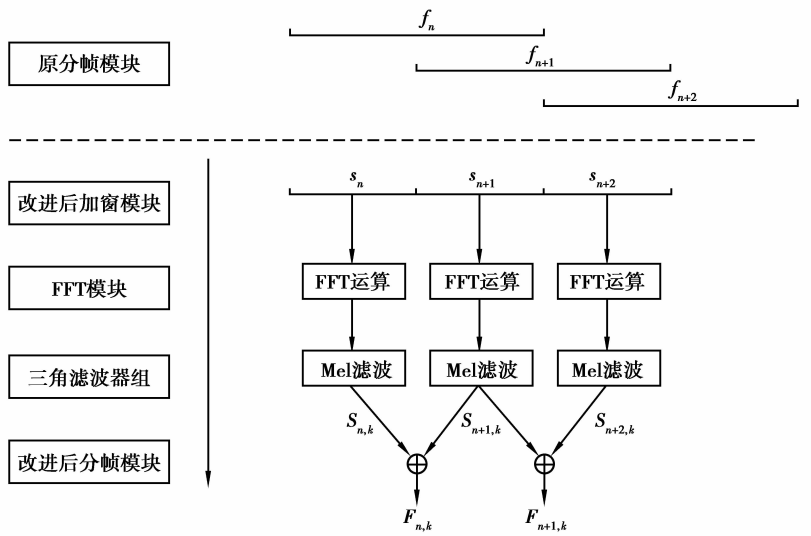

图 4　MFCC_E 算法分帧输出

2.2　改进预加重参数

预加重模块使用公式(1)来实现,原算法中参数 μ 的取值为 0.97。改用 31/32 代替 0.97,则公式(1)演变为公式(4):

$$H(z) = 1 - \frac{31}{32}z^{-1} = 1 - \left(1 - \frac{1}{32}\right)z^{-1} \tag{4}$$

尽管 31/32 与 0.97 在数值上没有多大差异,但带来的好处却非常明显。因为 1/32 可以通过数据移位操作(右移 5bit)来实现,这样,预加重模块可通过移位操作与加法运算的组合来代替原来的乘法运算,易于硬件实现。

3　实验结果

将改进算法应用于森林防盗砍系统中,选择森林为实验背景,搜集森林盗砍容易出现的 3 种异常声音,如砍树声、锯树声、树木倒塌声作为实验素材。

3.1　实验环境

实验运行在 PC 机的 Windows XP 操作平台上,PC 机的主频为 2.66 GHz,内存 2 GB,编程使用 Matlab 7.0。实验声音种类为砍树声、锯树声和树木倒塌声,每类声音有 40 个样本,采样率为 16 kHz,量化为 16 bit,原算法帧长 256 点,改进算法帧长为 128 点。训练样本随机选择总样本的 80%,识别样本为剩余的 20% 样本,每组实验做 5 次,列出每类声音的平均识别率,最后对相同 GMM 混合阶数下所有声音的识别率求平

均值。

3.2 GMM 训练库

样本训练过程如图5所示。对训练样本进行特征参数提取,并对各种异常声音的样本做模型训练,得出3种异常声音模型。

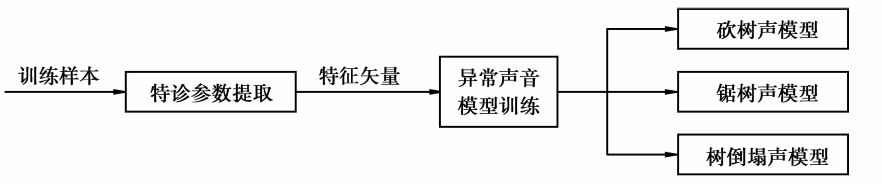

图5 GMM 训练模块

3.3 实验数据

表1为原MFCC算法对异常声音识别率,表2为MFCC_E算法对异常声音的识别率,N为GMM混合阶数。从测试数据可知,MFCC与NFCC_E算法的识别率相差不大,都具有较高的识别率,而且识别率随着GMM混合阶数的增加而提高。

表1 MFCC 算法对异常声音的识别统计

样本	识别率		
	$N=12$	$N=16$	$N=20$
砍树声	0.925	0.975	0.975
锯树声	0.850	0.900	0.950
树木倒塌声	0.750	0.775	0.875
平均识别率	0.841	0.883	0.933

表2 MFCC_E 算法对异常声音的识别统计

样本	识别率		
	$N=12$	$N=16$	$N=20$
砍树声	0.925	0.925	0.975
锯树声	0.825	0.900	0.925
树木倒塌声	0.750	0.800	0.900
平均识别率	0.833	0.875	0.933

4 结束语

介绍一种声音特征参数MFCC_E提取办法,将MFCC_E特征用于GMM的训练与识别中,实验证明使用MFCC_E算法同样能够达到MFCC的识别效果,当选择合适的GMM混合阶数,识别率能够达到90%以上。MFCC_E算法的关键优势在于降低了算法的复杂度,与MFCC算法相比,运算量减少了50%,而且其预加重阶段可直接使用硬件实现。

实验结果显示,GMM混合阶数越高,则对异常声音的识别效果越好,但阶数高使得模型的参数增多,更有可能在训练数据时得不到收敛的模型。如何选择合适的混合阶数将是下阶段的研究重点。

参考文献:

[1] ZAJDEL W,KRIJNDERS J D,ANDRNGA T. Audio-video sensor fusion for aggression detection [A]. Proceedings of the 2007 IEEE International Conference on Advanced Video and Signal based Surveillance[C]. London:IEEE Computer Society,2007

[2] 张玲华,郑宝玉,杨震. 基于LPC分析的语音特征参数研究及其在说话人识别中的应用[J]. 南京邮电学院学报,2005,25(6):1-6

[3] 荣薇,陶智,顾济华. 基于改进 LPC 和 MFCC 的汉语耳语音识别[J]. 计算机工程与应用,2007,43(30):213-216

[4] LEE C H,CHOU C H,HAN C C. Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis[J]. Pattern Recognition Letters,2006,27(2):93-101

[5] 吕霄云,王宏霞. 基于 MFCC 和短时能量混合的异常声音识别算法[J]. 计算机应用,2010,30(3):796-798

[6] WANG J C,WANG J F,WANG Y S. Chip design of MFCC extraction for speech recognition [J]. Integration,2002,32 (1/2) :111 -131

[7] RABAOU I A,DAVY M,ROSSIGNOL S. Using one-class SVMs and wavelets for audio surveillance [J]. IEEE Transactions on Information Forensics and Security,2008,3(4):763-775

[8] RABAOU I,LACHIR I Z,ELLOUZE N. Using HMM-based classifier adapted to background noises with improved sounds features for audio surveillance application [J]. International Journal of Signal Processing,2008,5(1):46-55

[9] RADHAKR I R,DIVAKARAN A,SMARAGDIS A. Audio analysis for surveillance applications [A]. Proceedings of the 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics[C]. Washington,DC:IEEE Computer Society,2005

[10] 胡益平. 基于 GMM 的说话人识别技术研究与实现[D]. 厦门:厦门大学,2007

Abnormal Audio Recognition Algorithm Based on Improved MFCC

HE Ling-ling, ZHOU Yuan

(School of Computer Science and Information Engineering,
Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: In audio recognition system, the acquisition of characteristic parameters has important influence on audio recognition and training. MFCC algorithm, as a typical audio characteristic parameter extraction method, has stable performance and high recognition rate. According to the situation that MFCC algorithm has larger amount of computation, a kind of improved characteristic parameter extraction algorithm, MFCC_E, was pointed out. Compared with standard MFCC algorithm, MFCC_E algorithm reduced about 50% computation amount and was easily realized on hardware. Experiment results show that the recognition rate of MFCC_E algorithm is approximately the same as MFCC algorithm but the computation difficulty of MFCC_E is largely smaller than that of MFCC algorithm.

Key words: audio recognition; characteristic extraction; MFCC; MFCC_E; GMM

责任编辑:代小红
校　　对:李翠薇