

基于 MFCC 与改进 ACF 的汽车声音识别算法研究

蒋翠清, 邵宏波

(合肥工业大学 管理学院, 安徽 合肥 230009)

摘要:汽车声音识别是汽车声源定位等研究的基础,对交通事故鉴定、犯罪举证和犯罪现场还原等具有重要意义。现有汽车声音识别算法存在算法复杂度高和识别率相对较低等问题。针对现行问题,将以梅尔倒谱系数(MFCC)特征与自相关函数(ACF)方差作为混合特征的汽车声音识别算法应用到汽车声音识别系统中。该算法使用高斯混合模型(GMM)进行汽车声音建模和识别,获得比 MFCC 特征及其一阶差分特征组成的混合特征更好的识别效果。并通过仿真实验证明了该算法的有效性。

关键词:汽车声音识别;梅尔倒谱系数;自相关函数;高斯混合模型

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2015)02-0140-04

doi:10.3969/j.issn.1673-629X.2015.02.032

Research on Vehicle Audio Recognition Algorithm Based on MFCC and Improved ACF

JIANG Cui-qing, SHAO Hong-bo

(School of Management, Hefei University of Technology, Hefei 230009, China)

Abstract: Vehicle audio recognition is the foundation of vehicle sound source localization and other automotive research, it is very important for traffic accidents identification, crime scene evidence and crime reduction. The problem of high computational complexity and relatively low recognition rate has existed in current vehicle audio recognition. Concerning those problems above, the vehicle recognition algorithm taking Mel-Frequency Cepstrum Coefficients and improved Auto-Correlation Function as hybrid feature is applied in the vehicle audio recognition system. Modeling and classifying by the Gaussian Mixture Model, this feature vector outperforms MFCC and Differential MFCC features in recognition. The simulation results prove the effectiveness of the proposed algorithm.

Key words: vehicle audio recognition; MFCC; ACF; Gaussian mixture model

0 引言

在目标追踪领域,视频和音频材料是两类最重要的分析数据。目前以视频监控为核心技术的公共场所安全监控系统对减少犯罪率、提高案件侦破效率以及减低犯罪取证难度都起到了积极的作用。然而视频监控系统的的作用范围有限,一旦被观测目标离开视频监控范围,视频监控的性能就会大大降低^[1]。相比视频监控而言,音频监控有着复杂度小、成本低、作用范围大等优点,且声音信号变化较平稳,易于处理。因此,声音识别、音频监控、声目标定位等成为近年来研究人

员关注的热点^[2]。

Harma 等人提取了均方根、音调、频谱质心等特征参数,将办公室里的声音分成说话声、打印机声、键盘声等声音^[3];Atrey 等人给出了基于事件检测的多媒体监控系统,采用短时过零率以及对数频率倒谱系数等特征参数,运用高斯混合模型对叫声、哭声、锁门声、走路声等进行建模和识别^[4];管鲁阳等根据人的听觉系统里耳蜗的特性,建立了耳蜗模型,用以模仿人的听力处理机制,并由此提取了车辆声信号特征^[5];中南大学的周宴宇等利用 LPCC 和 MFCC 特征对滚动轴承声音

收稿日期:2014-02-20

修回日期:2014-05-26

网络出版时间:2014-12-27

基金项目:国家自然科学基金重点资助项目(71331002);高等学校博士学科点专项科研基金资助项目(20120111110027)

作者简介:蒋翠清(1965-),男,教授,博士,博士生导师,研究方向为数据挖掘、人工智能;邵宏波(1988-),男,硕士研究生,研究方向为数据挖掘、信息管理与信息系统。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20141227.1340.006.html>

信号进行描述,采用神经网络方法进行学习,对滚动轴承的故障诊断起到了很好的辅助作用^[6]。

文中采用梅尔倒谱系数(MFCC)和自相关函数(ACF)作为描述汽车声音特性的混合特征参数,通过GMM进行建模和识别,达到了较好的识别效果。与单纯采用MFCC特征或者MFCC特征及其一阶差分相结合作为特征的方式相比,文中算法运算复杂度较低,同时识别率也有所提升。

1 汽车声音特征分析与提取

音频信号特征参数提取是指从音频信号当中能够获得描述音频信号特性的特征矢量的过程。为了更为有效和准确地提取特征,在特征提取之前,需要对原始音频样本进行预处理。

1.1 预处理

假设待处理的音频信号为 $x(n)$,则预处理过程如下:

(1)低通滤波。通过对音频的频谱分析可知,列车鸣笛的频域主要集中在800~2 000 Hz之间,为消除音频中高频噪音的干扰,选取截止频率为2 000 Hz的低通滤波器对原始音频 $x(n)$ 进行低通滤波处理。

(2)归一化处理。归一化处理的目的是为了消除不同声音样本音强的差异,将音频样本的幅值限定在 $[-1,1]$ 之间。

(3)预加重。预加重选取如式(1)所示的一阶数字滤波器进行音频信号加重处理。

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

其中, μ 为常数,通常为0.97。

(4)加窗分帧。音频信号具有短时平稳的特性,经过分帧操作之后的音频信号可以有效体现该特性。选取30 ms为一帧音频的长度,每帧间重叠长度为15 ms。为了降低分帧造成的频谱能量泄露,需要对每一

帧音频加窗,文中窗函数选择的是汉明窗^[7]。

1.2 MFCC

大量研究表明,人耳对于不同频率的声音具有不同的感知能力,并且这种感知能力不是随着频率的增加而线性增加的。人们根据人耳在不同频率下的音调感应能力,提出了Mel频率的概念^[8]。MFCC是在Mel标度频率域提取出来的倒谱参数,在声音信号处理领域中得到了广泛的应用。

求Mel倒谱系数的方法是将预处理之后的时域信号通过FFT(快速傅里叶变换)求取其频域谱^[9]后,对每帧数据的频谱参数通过一组 M 个依照Mel刻度均等分布的三角滤波器组做卷积,之后对每个频带的输出取对数,求出每个输出的对数能量,再对滤波器组的输出向量做离散余弦变换(DCT),这样得到的前 M 维向量称为MFCC,如式(2)所示:

$$C_t(n) = \sum_{m=1}^M S_t(m) \cos(\frac{\pi n(m-0.5)}{M}), 0 \leq n \leq M$$
(2)

其中, $C_t(n)$ 为第 t 帧的第 n 个MFCC参数; $S_t(m)$ 为音频信号的对数功率谱; M 为三角滤波器的个数,文中取 M 为12。

由于MFCC反映了音频帧数据静态的特征,且音频是一个动态的时变过程,所以通常还需要提取音频数据的变化特征,这就需要计算MFCC的一阶差分 $D_t(n)$,如式(3)所示:

$$D_t = \frac{\sum_{i=1}^n i * (C_{t+i} - C_{t-i})}{2 \sum_{i=1}^n i^2}$$
(3)

其中, i 为取一阶差分的窗口宽度; D_t 为在时域第 t 帧的一阶差分。

图1为MFCC参数提取流程。

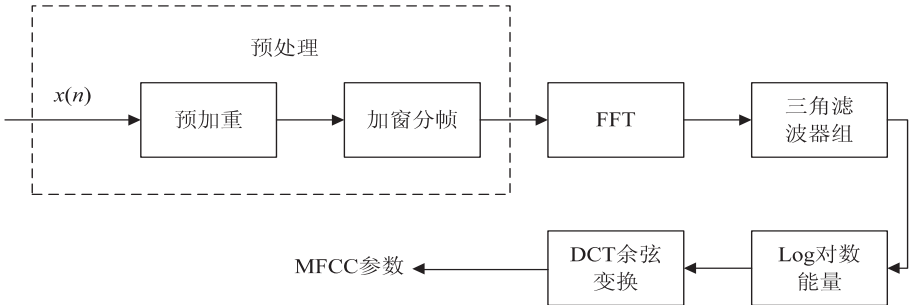


图1 MFCC参数提取流程

1.3 自相关函数

自相关函数(Auto-Correlation Function, ACF),常用于音频基音周期检测,这是音频处理的一种时域方法,是一帧音频信号 $s(n)$ 与其自身延迟后的信号 $s(n + \tau)$ 之间相似性的度量:

$$acf(\tau) = \sum_{i=0}^{n-1-\tau} s(i)s(i + \tau)$$
(4)

一般而言,对音频一帧信号通过自相关函数提取出一组自相关序列参数后,会提取该组序列中的次高值求该帧信号的基音频率^[10],以此作为信号的特征参

数,但是从 Matlab 仿真实验可以看出,鸣笛段与非鸣笛段的音频特征区分并不明显。然而笔者通过大量实验发现,鸣笛帧的 ACF 序列的波形振幅明显要比非鸣笛帧的 ACF 波形振幅要大,因此可以通过求 ACF 序列的方差作为区分鸣笛与其他汽车声音的特征,如图 2 所示。

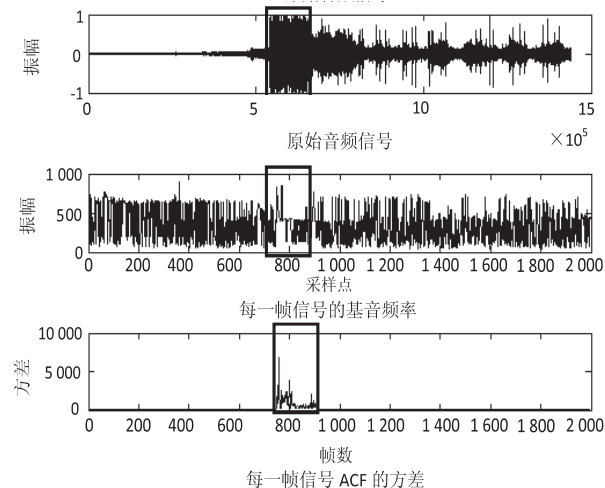


图 2 音频基音频率特征与 ACF 方差特征
其中方框内为鸣笛部分的音频。

1.4 混合特征

文中选取了 MFCC 与 ACF 相结合的特征参数组合,其理论依据如下:

(1)由一帧音频信号求出的自相关函数方差是一个标量值,因此与其他特征构成混合特征后也不会使原有特征矢量的维数显著增加。特征矢量维数越低,运算复杂度越小。同时,获取自相关函数方差的运算复杂度较低。

(2)MFCC 是人耳听觉感知特征,属于音频的频域特征。ACF 是音频的时域特征,标示的是一帧音频信号与其自身延迟后信号相似性的度量,与 MFCC 的相关性不大。并且 ACF 计算复杂度较低,可以有效提高算法的识别效率。

文中 MFCC 取 12 维, MFCC 的一阶差分作为 13 到 24 维,将每一帧音频的 ACF 方差作为第 25 维特征,因此混合特征矢量的维度为 25 维,文中将 ACF 方差特征记为 A,该混合特征矢量记为 MFCC_A。

2 特征参数分类

声音识别系统需要建立声音模型库,并利用分类器对提取的汽车声音特征矢量进行训练。目前广泛使用的分类器主要有支持向量机 (Support Vector Machine, SVM)、隐马尔可夫模型 (Hidden Markov Model, HMM) 和高斯混合模型 (Gaussian Mixture Model, GMM) 等。

GMM 是基于参数估计的概率统计模型,它的优点是可以平滑地逼近任意形状的概率密度函数,并且是易于处理、相当稳定的参考模型^[11]。一个 N 阶的 GMM 模型的概率密度函数是由 N 个高斯分布通过加权求和得到的,被赋予不同权重的高斯分布加以组合即可用来表征不同的汽车声音。

GMM 训练模型实质上就是提供一组训练样本,基于某种准则来确定模型的参数^[12]。文中训练样本即为所要训练的声音特征矢量。文中需对汽车声音进行识别,主要为汽车鸣笛声,非鸣笛声主要有车门关闭声、引擎发动声、汽车刹车声等声音。对每类汽车声音分别进行 GMM 模型训练,对于给出的初始模型参数通过 EM 算法进行参数估计,可以得到描述每一类汽车声音特征的 GMM 三元式^[13],如式(5)所示:

$$\lambda = \{P_i, \mu_i, \Sigma_i\}, i = 1, 2, \dots, N \quad (5)$$

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,文中采用的混合阶数 N 为 12 阶。

识别的过程用测试音频的模型与不同异常声音的模型进行对比,根据后验概率 $P(\lambda_i | X)$ 来选择最接近的异常声音模型作为识别结果,并以此求出样本的识别正确率^[14]。

3 实验

3.1 实验准备

文中实验所处环境均为 PC 机的 Windows 7 操作系统,PC 机主频为 2.3 GHz,内存为 4 GB,实验所用软件主要是 Matlab R2012a。

实验样本如下:声音为鸣笛和非鸣笛两种,鸣笛种类为电鸣笛和汽鸣笛,非鸣笛种类为车门关闭声、引擎发动声和汽车刹车声等。其中鸣笛类声音有 100 个样本,非鸣笛类声音有 100 个,所有声音样本均为道路路口录制。样本采样率为 48 kHz,精度为 16 bit。特征提取帧长为 20 ms,帧移为 10 ms。训练样本为样本总数的 80%,测试样本为剩余的 20% 的样本。每组实验做 10 次,以此求得每类声音的平均识别率,并对不同特征组合下声音的识别率及效率进行比较。

3.2 实验结果与分析

表 1 为不同特征组合对鸣笛和非鸣笛声音的识别率,表 2 为不同特征组合对鸣笛和非鸣笛声音的时间复杂度,其中特征组合包括 12 维的 MFCC 特征、24 维的 MFCC 特征、MFCC(12 维)_A 特征组合以及 MFCC(24 维)_A 特征组合。从实验数据可以看出, MFCC(12 维)_A 特征的鸣笛识别率明显较高, MFCC(12 维)特征的非鸣笛识别率较高。而就识别的时间复杂度而言, MFCC(12 维)_A 识别鸣笛和非鸣笛的样本的

时间复杂度均为最低。综合识别率和时间复杂度两方面的因素来看,MFCC(12 维)_A 较单独使用 MFCC 特征以及 24 维 MFCC 与 ACF 方差的混合特征而言识别效果更好且时间复杂度较低。因此,采用 MFCC(12 维)_A 混合特征进行汽车鸣笛与非鸣笛区分是可行并且有效的,在提高识别率的同时也降低了时间的消耗。

表 1 不同特征组合声音的识别率

样本	识别率			
	MFCC (12 维)	MFCC (24 维)	MFCC (12 维)_A	MFCC (24 维)_A
鸣笛	0.887	0.761	0.945	0.809
非鸣笛	0.947	0.826	0.936	0.831

表 2 不同特征组合声音识别的时间复杂度

样本	时间复杂度			
	MFCC (12 维)	MFCC (24 维)	MFCC (12 维)_A	MFCC (24 维)_A
鸣笛	2.369	2.466	2.338	2.394
非鸣笛	1.805	1.875	1.790	1.869

4 结束语

文中对 MFCC、一阶差分 MFCC 以及 ACF 方差三种特征所构成的混合特征进行 GMM 的建模与识别,用于列车鸣笛识别。该混合特征全面考虑了列车鸣笛声的时域特性和频域特性,对列车鸣笛信号特征的描述更加全面。经过多组实验验证,12 维 MFCC+短时能量+ACF 方差的混合特征比传统的 12 维 MFCC+12 维 MFCC 一阶差分和 12 维 MFCC 特征的识别率高,且维度较低,时间复杂度低。GMM 模型阶数的选择是一个重要的问题,混合阶数过少时训练得到的模型不足以完整地描述各类汽车声音的特征,混合阶数过多时模型较为复杂,需要的训练样本较多;而此次实验的训练样本共两类,每类 80 个样本,数据相对较少,训练出的模型对汽车声音的描述不够精确。文中采用的 GMM 混合阶数为 12 阶。合适的 GMM 混合阶数选择以及混合阶数对于识别性能的影响研究,将是下一阶段的研究重点。

参考文献:

[1] 朱慧芬,汤啸天,陈冬沪,等. 公共场所监控图像采集利用与隐私权保护研究报告[C]//2009 年政府法制研究. 上海:上海市行政法制研究所,2009:480-536.

[2] Zajdel W, Krijnders J, Andringa T. Audio-video sensor fusion for aggression detection[C]//Proceedings of the 2007 IEEE international conference on advanced video and signal based surveillance. London:IEEE Computer Society,2007.

[3] Harma M, McKinney F, Skowronek J. Automatic surveillance of the acoustic activity in our living environment[C]//Proc of IEEE international conference on multimedia and expo digital signal process. Netherlands:IEEE,2005:152-152.

[4] Atrey P K, Maddage N C, Kankanhalli M S. Audio based event detection for multimedia surveillance[C]//Proc of IEEE international conference on acoustics, speech and signal processing. Toulouse:IEEE,2006:3-5.

[5] 管鲁阳,鲍明,李晓东. 基于简化耳蜗模型的车辆声信号特征提取算法[J]. 声学技术,2007,26(5):227-228.

[6] 周宴宇,杨胜跃,何正明,等. 基于 MFCC 与 LPCC 的滚动轴承故障音频诊断法[J]. 微计算机信息,2009(31):123-124.

[7] 张斌,孔敏,吴从兵. 基于窗函数下频谱泄露的研究[J]. 信息化纵横,2009(11):10-12.

[8] 王华朋,杨洪臣. 声纹识别特征 MFCC 的提取方法研究[J]. 中国人民公安大学学报:自然科学版,2008(1):28-30.

[9] 冷建华. 傅里叶变换[M]. 北京:清华大学出版社,2006:27-28.

[10] 刘何来,张正炳,杨顺辽. 高噪声环境下的基音频率检测方法研究[J]. 长江大学学报:自然科学版,2013,10(1):46-48.

[11] 张文增,陈强,都东,等. 直线检测的灰度积分投影方法[J]. 清华大学学报:自然科学版,2005,45(11):1446-1449.

[12] 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.


[13] 吕霄云. 基于 MFCC 和 GMM 的异常声音识别算法研究[D]. 成都:西南交通大学,2010.

[14] 余清清,李应,李勇. 基于高斯混合模型的自然环境声音的识别[J]. 计算机工程与应用,2011,47(25):152-155.

基于MFCC与改进ACF的汽车声音识别算法研究

作者：[蒋翠清](#)，[邵宏波](#)，[JIANG Cui-qing](#)，[SHAO Hong-bo](#)

作者单位：[合肥工业大学 管理学院, 安徽 合肥, 230009](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2015 (2)

引用本文格式：[蒋翠清](#), [邵宏波](#), [JIANG Cui-qing](#), [SHAO Hong-bo](#) [基于MFCC与改进ACF的汽车声音识别算法研究](#)[期刊论文]-[计算机技术与发展](#) 2015 (2)