



基于设备本底噪声频谱特征的手机来源识别

裴安山,王让定,严迪群

(宁波大学,浙江 宁波 315211)

摘要:随着手机录音设备的普及以及各种功能强大且易于操作的数字媒体编辑软件的出现,手机来源识别已成为多媒体取证领域的热点问题。将本底噪声作为手机的“指纹”,提出了一种基于本底噪声的手机来源识别方法。该方法先通过静音段录音的估计得到本底噪声;然后计算本底噪声的频谱特征在时间轴方向上的均值,将其作为手机来源识别的分类特征;最后采用主成分分析(PCA)法对特征进行降维,并采用支持向量机(SVM)进行分类。实验部分对24款主流型号的手机进行了分类,结果表明本文方法的平均识别准确率(accuracy)和平均召回率(recall)达到了99.24%,同时也验证了相比MFCC,本底噪声有更加优越的性能。

关键词:多媒体取证;手机来源识别;本底噪声;频谱特征

中图分类号:TP391

文献标识码:A

doi: 10.11959/j.issn.1000-0801.2017019

Cell-phone origin identification based on spectral features of device self-noise

PEI Anshan, WANG Rangding, YAN Diquan

Ningbo University, Ningbo 315211, China

Abstract: With the widespread availability of cell-phone recording devices and the availability of various powerful and easy-to-use digital media editing software, source cell-phone identification has become a hot topic in multimedia forensics. A novel cell-phone identification method was proposed based on the recorded speech. Firstly, device self-noise (DSN) was considered as the fingerprint of the cell-phone and estimated from the silent segments of the speech. Then, the mean of the noise's spectrum was extracted as the identification. Principal components analysis (PCA) was applied to reduce the feature dimension. Support vector machine (SVM) was adopted as the classifier to determine the source of the detecting speech. Twenty-four popular models of the cell-phones were evaluated in the experiment. The experimental results show that the average identification accuracy and recall of the method can reach up to 99.24% and demonstrate that the self-noise feature has more superior performance than the MFCC feature.

Key words: multimedia forensics, cell-phone origin identification, self-noise, spectral feature

收稿日期:2016-12-11; **修回日期:**2017-01-10

通信作者:王让定, wangrangding@nbu.edu.cn

基金项目:国家自然科学基金资助项目(No.61672302, No.61300055);浙江省自然科学基金资助项目(No.LZ15F020002, No.LY17F020010);宁波大学科研基金资助项目(No.XKXL1405, No.XKXL1420, No.XKXL1509, No.XKXL1503);宁波大学科研创新基金资助项目(No.G16079);宁波大学王宽诚幸福基金资助项目

Foundation Items: The National Natural Science Foundation of China (No.61672302, No.61300055), Natural Science Foundation of Zhejiang Province of China (No.LZ15F020002, No.LY17F020010), Scientific Research Foundation of Ningbo University(No.XKXL1405, No.XKXL1420, No.XKXL1509, No.XKXL1503), Scientific Research Foundation of Graduate School of Ningbo University(No.G16079), K.C. Wong Magna Fund in Ningbo University



1 引言

如今,随着移动互联网和微芯片产业的发展,移动终端成为人们生活中不可或缺的部分。越来越多的人开始用智能手机、Pad 等便携设备捕捉和采集他们看到或听到的情景,而不是用相机、录音笔、DV 等专业设备。大量数字采集设备和采集数据的可用性带来了新的问题和挑战——多媒体的安全问题。作为一种检测多媒体数据独创性、真实性、完整性的技术,多媒体取证技术是信息安全领域的热点研究问题^[1]。

基于语音的手机来源识别是与多媒体取证最相关的应用之一,用来确定录制语音文件的手机型号。这一研究方向引起了大量取证研究者的关注,并在近年获得重大进展^[2-5]。Hanilci 等人^[6]通过提取录音文件的 MFCC(梅尔倒谱系数)特征进行手机品牌和型号的识别。在 14 个不同型号手机的闭集识别实验中,识别率达到 96.42%。此后,又比较了 MFCC、LFCC(线性倒谱系数)、BFCC(Bark 频率倒谱系数)和 LPCC(线性预测倒谱系数)^[7]在手机来源识别中效果的优劣,实验得出 MFCC 特征在手机来源识别中效果最好。在手机的来源识别领域还有很多重要的成果^[8-10], Kotropoulos^[11]通过对不同手机采集的语音信号频谱取对数,然后沿时间轴取平均,或者通过对每一帧的特征参数利用高斯混合模型建模得到高维的特征向量,然后映射到低维空间进行降维。在 7 个品牌 21 个型号的手机来源识别实验中,识别率达到 94%。王志锋等人^[12]考虑到语音段中的静音包含了与正常语音一样的设备信息,并且不受说话人、文本、情感等因素的影响,因此从静音段中提出表征录音设备的特征。然后利用设备通用背景模型构建设备模型。最后通过归一化似然度得分,对输入的录音设备的语音样本特征进行分类判决。实验结果表明,他们的算法对于 9 种不同设备的平均识别率为 87.42%。邹领等人^[13]提出了一种基于噪声估计的录音设备源识别方法,

通过录音设备源识别实验以及与其他几种常用的录音设备源识别特征,包括 MFCC 均值特征、基于 MFCC 的 GLDS kernel 超向量特征以及 SSF 特征做比较,证明了该方法的有效性。

然而,现有大多数手机来源识别的研究是基于 MFCC 特征或 MFCC 的统计分类特征的。虽然相关特征在手机来源识别中取得了令人满意的效果,但是梅尔频带近似人类听觉系统的响应,分类效果可能受到许多不确定条件的干扰,如说话人的性别、情感、语音内容等。为了有效捕获手机的内在“指纹”,将本底噪声作为相同型号手机特有的身份特征。本底噪声是设备自身产生的信号,它的特性主要受电路设计和电子元器件的影响。因为不同型号的手机电路设计和电子元器件的使用不一致性,所以不同型号的手机没有完全相同的本底噪声。本文提出了一种基于本底噪声频谱分布特征的手机来源识别方法。实验结果表明,本底噪声的频谱分布特征能够对数据库中的手机进行识别分类,平均识别准确率和平均召回率可以达到 99.24%。

2 本底噪声

2.1 本底噪声的定义

手机的本底噪声是手机本身产生的噪声,是一种在没有输入信号的情况下也能产生的噪声。在电子学中,所有组件的温度在绝对零度以上时,都将产生自己的噪声。而录音在采集过程中可能会受到录音软件和系统操作版本的影响,另外,部分手机可能包含语音压缩等信号处理过程,因此本文研究的本底噪声是广义上的本底噪声。在研究中默认使用系统自带的录音软件,对手机电子元器件和电路设计对本底噪声的影响进行整体研究,针对系统版本对本底噪声的影响挑选几款典型的手机进行对比研究^[13-15]。图 1 为数字语音录音系统的简化示意,可以看出,背景噪声和本底噪声是独立的,本底噪声主要由与麦克风相关的电子元件及编码传输失真产生的噪声组成。

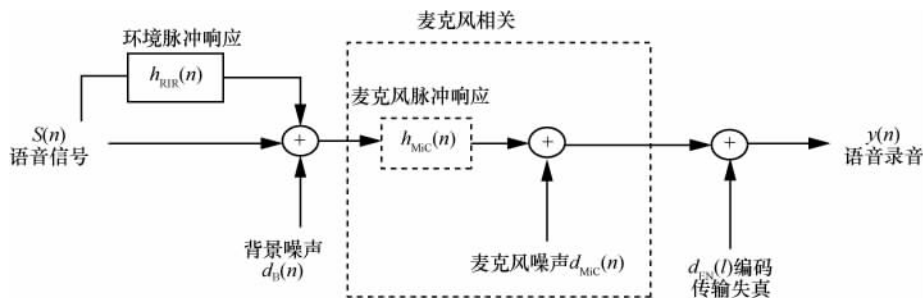


图 1 数字语音录音系统简化示意

本底噪声的定义表明,本底噪声的特性与手机的电路设计和电子元器件的使用密切相关。由于不同型号的手机在电路设计和电子元器件的使用上存在部分差异。所以,本底噪声可以作为手机的“指纹”进行手机的来源识别。

2.2 本底噪声的性质

为了研究本底噪声的特性,对现下流行的 7 个品牌 24 个型号的 25 台设备(其中有两台 iPhone 5)进行了研究,具体手机型号见表 1。实际本底噪声样本是在无回声的录音棚的静音环境下录制的。为了避免其他电气设备噪声对实验的干扰,当手机在录音时,关闭了整个录音棚的电源开关。

语谱是语音信号的频谱随时间变化的直观展示。为了研究本底噪声的频谱特性,图 2 给出了其中 8 个手机的本底噪声语谱。通过观察图 2 可以看出,不同品牌手机的本底噪声语谱存在很大差异。例如,魅族 MX4 本底噪声语谱的振幅曲线是随频率呈波动变化趋势的,HTC D610t 的本底噪声语谱在频率为 4 000 Hz 附近有大幅下降,三星 Galaxy S5 本底噪声的能量在所有的频点间隔(0~16 kHz)是最弱的,而相同型号的手机(如 iPhone 5)的语谱比较相似。

图 3 以 iPhone 手机为例,给出了相同品牌不同型号手机的语谱。通过图 3 可以看出,同一品牌手机的本底噪声语谱在一定程度上是相似的,不同型号手机间的语谱区别较小,但是在低频部分的噪声强度上仍有一定区别,需要构建特征并选用合适的分类算法进行区分。基于图 2 和图 3

表 1 实验中手机的品牌和型号以及类名

类名	品牌	型号
H1	HTC	D610t
H2		D820t
H3		One M7
W1	华为	Honor6
W2		Honor7
W3		Mate7
O1	OPPO	Find7
O2		Oneplus1
O3		R831S
S1	三星	Galaxy Note2
S2		Galaxy S5
S3		Galaxy GT-I8558
A1	苹果	iPhone 4s
A2		iPhone 5
A3		iPhone 5s
A4		iPhone 6
A5		iPhone 6s
Z1	魅族	Meilan Note
Z2		MX2
Z3		MX4
M1	小米	Mi 3
M2		Mi 4
M3		红米 Note1
M4		红米 Note2

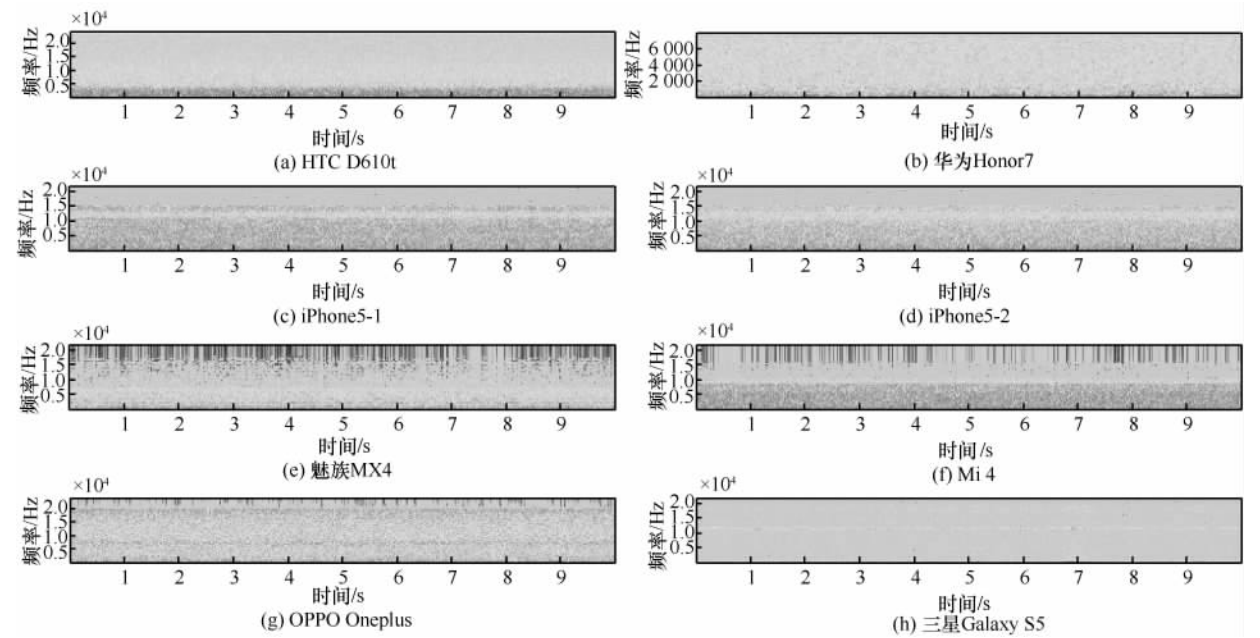


图 2 不同手机品牌的本底噪声语谱

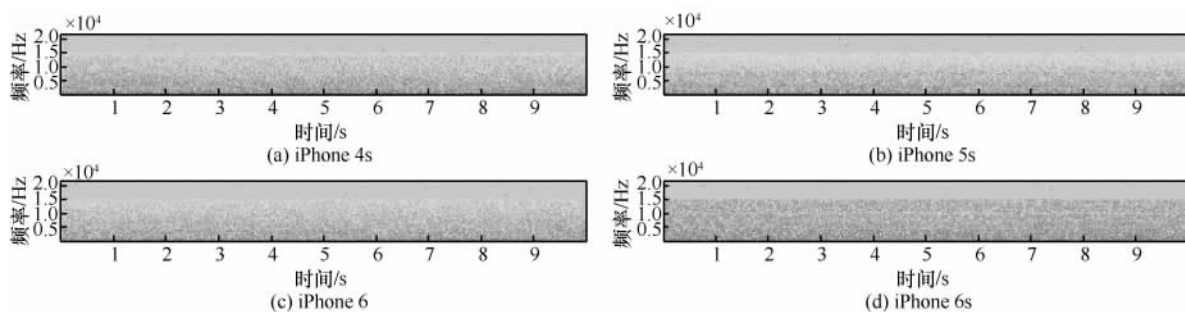


图3 苹果品牌手机的语谱

的分析可得,显示语谱内在变化的频谱分布特征可以作为手机设备来源识别的特征。

以 HTC D610t、iPhone 5 和 Mi 4 为例,对 3 台手机的操作系统版本进行了升级,以验证系统版本对本底噪声的影响。图 4 是升级后 3 台手机本底噪声的语谱,由图 4 和图 2 中 3 台手机的语谱对比可得,HTC D610t 和 iPhone 5 系统升级前后本底噪声语谱没有明显变化,Mi 4 手机操作系统升级前后本底噪声语谱变化较两者略大,主要是高频部分本底噪声的能量有些许差别。

3 基于本底噪声的特征提取和来源识别

3.1 本底噪声的估计

本文将本底噪声作为手机来源识别的基本特征。因为本底噪声不等同于语音的噪声,所以本底噪声的估计是一项具有挑战性的工作。因此,本文提出两步实现本底噪声估计的方法。

(1) 静音段提取

静音段提取原因是:语音的静音段主要由本底噪声和背景噪声构成,不会被语音部分的综合噪声中占主导地位的声电响应不一致噪声所污染^[6]。本文采用自适应端点检测算法进行静音段估计^[17]。如图 5(a)和图 5(b)所示,该方法可以很好地识别静音段。从图 5(c)可以看出,经过自适应端点检测算法得到的静音段还含有少量语音信息。为了进一步消除不必要的语音信息,使用函数 $ns(\cdot)$ 对静音段进

行了后处理,如式(1)所示:

$$nss(j)=ns(find|ns(i)|<5\times Thr) \quad (1)$$

式(1)中 Thr 是 $ns(i)$ 升序排列后前 40% 样本的均值, $find(\cdot)$ 的功能是当 $ns(i)$ 的绝对值小于 5 倍的 Thr 时,定位和输出 $ns(i)$ 的位置。

经步骤 1 处理后,得到一个语音样本的几段静音段,将它们拼接起来得到一个静音段信号。

(2) 背景噪声抑制

为了从静音段得到实际的本底噪声,要尽可能地抑制背景噪声。首先,采用算法 1(改进的频谱减法)的伪代码抑制背景噪声。然后根据式(2)计算 5 760 个语音样本(24 个手机,每个手机 240 个样本)的通用背景噪声模型。

$$|BN_{mean}(k)|=\frac{1}{24}\sum_{i=1}^{24}\frac{1}{N}\sum_{n=1}^N|BN_i(k,n)| \quad (2)$$

其中, i 是手机型号的索引, BN_i 是背景噪声的语谱, $BN(k,n)$ 是第 $k(1\leq k\leq K)$ 个频率点、第 $n(1\leq n\leq N)$ 帧在短时傅里叶变换(short-time Fourier transform, STFT)域的频谱特征, K 表示语音样本频率点的个数, N 表示语音样本的总帧数。注意,本文语音样本 STFT 的点数为 4 096 个,由于 STFT 输出的对称性,所以 K 的值为 2 049(4 096/2+1)个, N 随着语音长度的变化而变化。 k 、 N 、 K 和 N 的定义也适用于其他式子。

算法 1 背景噪声计算

输入 静音段语谱 $|NS(k,n)|$, 本底噪声的频谱 $|SN_{mean}(k)|$

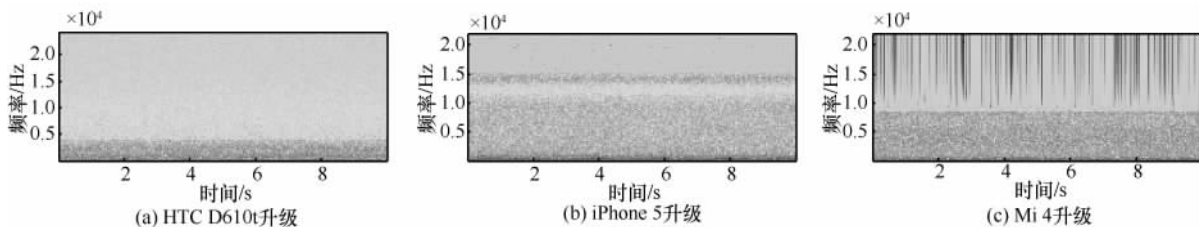


图4 操作系统升级后手机的语谱

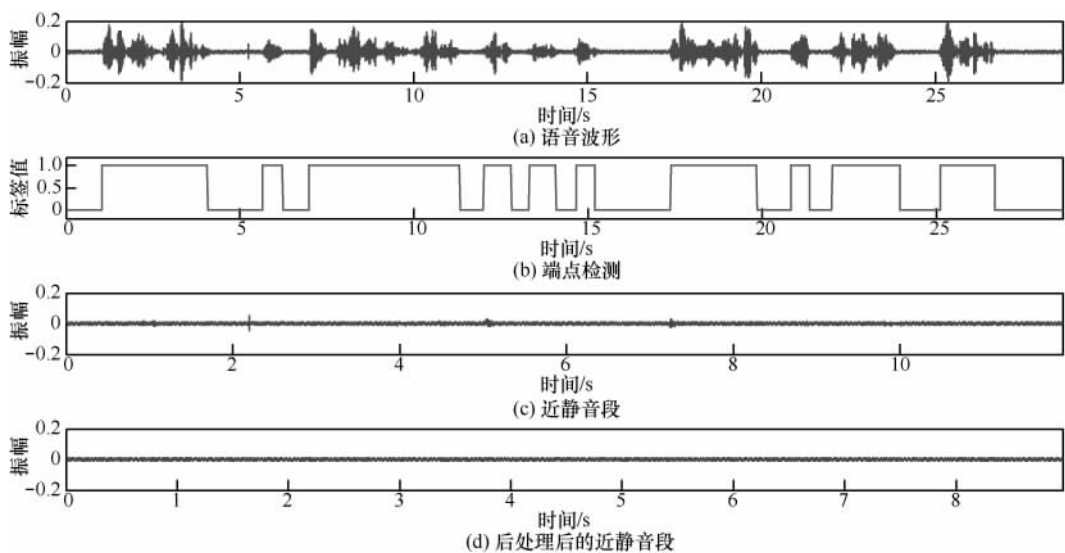


图5 基于端点检测的静音段估计

输出 背景噪声语谱 $|BN(k,n)|$

参数: 比例因子 $\alpha_1=3$, 增益补偿因子 $\beta_1=0.1$

if $|NS(k,n)| > \alpha_1 \times |SN_{mean}(k)|$

$|BN(k,n)| = |NS(k,n)| - \alpha_1 \times |SN_{mean}(k)|;$

else $|BN(k,n)| = \beta_1 \times |NS(k,n)|;$

end

因此,未知来源手机的本底噪声可以通过静音段信号减去通用背景噪声得到。为了使估计的本底噪声不含残余的背景噪声,设计了一个滤波方案。本底噪声估计的完整过程的伪代码如算法2所示。

算法2 本底噪声估计

输入 静音段语谱图 $|NS(k,n)|$, 背景噪声的频谱 $|BN_{mean}(k)|$

第一步的输出: 估计本底噪声的语谱 $|SN_{estimated}(k)|$

第二步输出: 估计的本底噪声滤波输出 $|SN_{filter}(k)|$

参数: 比例因子 $\alpha_2=3, \varphi=3, \omega=1$, 增益补偿因子 $\beta_2=0.9$

功能: $median(\cdot)$ 计算每一个频率区间沿时间轴的中值

$ifft(\cdot)$ 逆快速傅里叶变换重构时域信号

第一步: if $|NS(k,n)| > \alpha_2 \times |BN_{mean}(k)|$

$|SN_{estimated}(k,n)| = |NS(k,n)| - \alpha_2 \times |BN_{mean}(k)|;$

else $|SN_{estimated}(k,n)| = \beta_2 \times |NS(k,n)|;$

end

第二步: $SN_{median}(k) = median(|SN_{estimated}(k,n)|);$

if $|NS(k,n)| \geq \varphi \times |SN_{median}(k)|$

$|SN_{filter}(k,n)| = \omega \times |SN_{median}(k)|;$

else $|SN_{filter}(k,n)| = |SN_{estimated}(k,n)|;$

end

$|SN_{filter}(n)| = ifft(|SN_{estimated}(k,n)|)$

图6展示了iPhone 6的实际本底噪声语谱和它的估计本底噪声的语谱。对比图3(c)和图6(b)可以看出,iPhone 6估计的本底噪声语谱与实际的本底噪声语谱是最相似的,图6(c)也表明了估计本底噪声与实际本底噪声的频谱是很相似的。因此,本文所提出的两步实现对未知来源手机语音的本底噪声估计的方法是可行的。

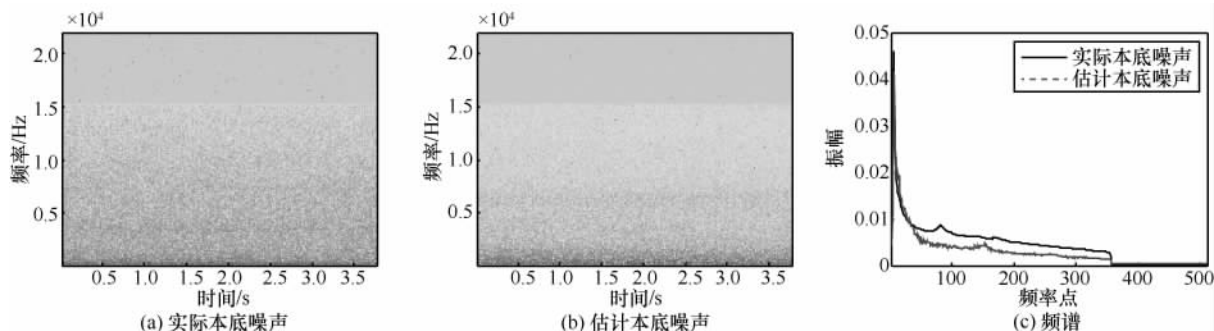


图6 估计本底噪声和实际本底噪声的对比



3.2 特征提取和识别过程

为了实现手机的来源识别,需要选择不同型号的手机之间具有差异性的本底噪声特征。本文选取本底噪声的频谱分布特征作为分类特征,如式(3)所示:

$$f_{sd}(k)=\frac{1}{N}\sum_{n=1}^N\lg|SN(k,n)| \quad (3)$$

特征的构建过程表述如下:首先求出本底噪声的频谱特征 $SN(k,n)$, 然后对频谱特征 $SN(k,n)$ 取对数,最后沿时间轴按帧取平均,作为一个样本的特征。

正如第 3.1 节提到的每帧的特征维度是 2 049, 由于特征维数太大,可能导致特征不能完全独立不相关。冗余的特征不仅不能提升识别的准确率,甚至会影响特征向量的识别效果。而 PCA(principal component analysis, 主成分分析)是一种常用的数据分析方法^[18,19],通过线性变换将原始数据变换为一组各维度线性无关的数据,可用于提取数据的主要特征分量,常用于高维数据的降维。因此利用 PCA 进行降维,形成最优的频谱分布特征空间。识别采用 PCA 降维和 MATLAB 自带的 SVM(support vector machine, 支持向量机)函数训练分类,SVM 是一种二类分类模型,其基本模型定义为特征空间上的间隔最大的线性分类器,其学习策略是间隔最大化,最终可转化为凸二次规划问题的求解。本文所采用的 SVM 函数使用的是 RBF(radial basis function, 径向基函数核函数)。实验中,对两个关键参数惩罚系数(C)和 Gamma 函数(γ)进行寻优处理。识别过程的步骤为:先将第 4.1 节所述的测试语音库中一半样本用于模型训练,另一半用于测试。训练集和测试集样本按第 3.1 节所述进行本底噪声估计,然后进行频谱分布特征提取,将各自所得的频谱分布特征分别构造特征空间。再利用 PCA 对训练特征空间进行降维,将降维后的特征空间归一化处理,测试集的特征空间根据训练集特征空间降维所采用的映射矩阵进行降维,对降维后的测试特征空间进行归一化处理。最后对降维后训练特征空间进行模型训练,对测试特征空间逐行进行 SVM 分类判别。

4 实验分析

4.1 实验设置

在实验中,建立了一个语音数据库来有效评估提出方法的有效性。该数据库由表 1 所示的 24 个型号的手机所采集的语音组成。邀请了 12 人(6 男 6 女)参与语音采集。

每个人需要录制两段语音,一段为用正常的语速朗读固定的内容,另一段为问答和主题演讲,时长均保证 5 min 以上。录音环境是一间相对安静的办公室,25 个手机由工作人员同时打开和关闭录音机,保证录制较好的同步性。每个手机采集了 24 段语音,对每段语音取中间固定的长度进行分割,将每段语音分割成 5 s 的语音片段,每个手机得到 1 200 个语音样本,作为原始语音库。对原始语音进行静音段估计提取,得到静音段数据库。由于静音段语音长度不一致。为了保证在构建特征空间时特征矩阵长度保持一致,选取每个型号手机 240 个语音帧数大于 40 帧的静音段语音,组成求取本底噪声频谱分布特征的测试语音库。对测试语音库中所有语音进行通用背景噪声模型训练,然后采用改进的谱减法得到估计本底噪声的语音库。构造特征空间时,取每段语音的前 40 帧本底噪声频谱分布特征的平均值。此处帧长为 64 ms,帧移为 32 ms。按照第 3.2 节所述,进行基于本底噪声的频谱分布特征的手机设备来源识别实验。

4.2 实验结果与分析

引入平均识别准确率 Accuracy 和召回率 Recall 作为客观指标评价分类特征的识别性能,如式(4)、式(5)所示:

$$Accuracy=(TP+TN)/(P+N) \quad (4)$$

$$Recall=TP/(TP+FN) \quad (5)$$

其中, TP 为正样本预测为正样本的个数, TN 为负样本预测为负样本的数目, FN 为正样本预测为负样本的数目, P 为正样本实际数, N 为负样本实际数, $P=TP+FN$, 式(5)可表示为 $Recall=TP/P$ 。在实验中,24 类手机来源识别的平均召回率为 $((TP_1+TP_2+\dots+TP_{24})/P)/24=(TP_1+TP_2+\dots+TP_{24})/24P$, 由于 $(TP_1+TP_2+\dots+TP_{24})=TP+TN$, $24P=P+N$, 则平均召回率即平均识别准确率。

表 2 给出了降维后特征为 28 维、参数 C 和 γ 分别为 30 和 0.001 时,24 个手机的识别结果,从中可算得平均识别准确率为 99.24%,算法可以较好地完成 24 个手机的分类识别,iPhone 6 识别率为 91.67%,其主要误判来自品牌类内区分,将之误判为 iPhone 4s 和 iPhone 5s。除 iPhone 6 以外,其他手机都有较高的识别准确率,其中有 18 个手机的识别率达到 100%,三星、OPPO、魅族等品牌可以做到无差错分类。

4.3 系统升级后实验结果与分析

对 3 台操作系统升级后的手机按第 4.1 节的实验设

表 2 本底噪声频谱分布特征的识别率

标签	预测标签																							
	H ₁	H ₂	H ₃	W ₁	W ₂	W ₃	A ₁	A ₂	A ₃	A ₄	A ₅	Z ₁	Z ₂	Z ₃	M ₁	M ₂	M ₃	M ₄	O ₁	O ₂	O ₃	S ₁	S ₂	S ₃
H ₁	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H ₂	2.5%	97.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H ₃	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W ₁	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W ₂	0	0	0	2.5%	97.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W ₃	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₁	0	0	0	0	0	0	97.5%	0	0	1.7%	0	0	0.8%	0	0	0	0	0	0	0	0	0	0	0
A ₂	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₃	0	0	0	0	0	0	0	0	98.3%	1.7%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₄	0	0	0	0	0	0	2.5%	0	5.8%	91.7%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₅	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0
Z ₁	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0
Z ₂	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
Z ₃	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0
M ₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0
M ₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
M ₃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
M ₄	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99.2%	0.8%	0	0	0	0	0
O ₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0
O ₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
O ₃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
S ₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
S ₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
S ₃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%

置进行语音录制和语音处理,每个手机得到 1 200 个语音样本,将其作为原始语音。对原始语音按照试验设置进行后续操作,每个设备取 100 个帧长大于 40 帧的估计本底噪声的语音片段构成测试语音库,对该测试语音库中的语音样本进行频谱分布特征提取,然后以升级后的测试样本代替升级前 3 台手机的测试样本,以每个设备 100 个测试样本在系统升级前的 24 台设备的训练模型上进行检测。

表 3 是 HTC D610t、iPhone 5 和 Mi 4 操作系统升级后的识别结果(实验中特征为 28 维、参数 C 和 γ 分别为 30 和 0.001),平均识别率为 98.04%。从表 3 中可以看出,HTC D610t、iPhone 5 和 Mi 4 这 3 台设备的识别准确率依次是 97%、97%和 99%。因此可以得出,手机操作系统版本的升级对手机本底噪声的影响较小,实验结果也验证了第 2.2 节中图 2 和图 4 对应 HTC D610t、iPhone 5 和 Mi 4 手机系统升级前后本底噪声语谱没有明显变化的结论。

4.4 Hanilci[®]实验结果与分析

为了综合评价所提出的方法,对参考文献[6]中提出的基于 MFCC 特征的手机来源识别算法在原始语音数据库上进行了测试。训练和测试样本数目与实验所用数目一致。在该工作中将 12 维的 MFCC 特征和基于帧间的 MFCC 一阶差分系数 36 维作为区分特征。对特征沿时间轴做平均,每个样本得到 48 维的特征。表 4 为实验结果,SVM 中的参数 C 和 γ 分别为 7 和 0.004。

从表 4 可知,与本文所提特征相比,部分手机的识别准确率严重下降,如 HTC D820t、iPhone 4s、魅族 Meilan Note、红米 Note 2、OPPO find 7、三星 Galaxy Note 2 等,尤其是 HTC D610t 识别率已经降到 50%。不过 24 台设备中有 15 台的识别正确率超过 90%,平均识别率为 88.23%。本文所提方法的平均识别率优于 Hanilci 所提算法,其原因是本文所提取的本底噪声是手机本身的特征,而 Hanilci 所提特征是基于语音信号的,所含的说话人信息可能对分类产生干扰。



表 3 操作系统版本升级后本底噪声频谱分布特征识别率

标签	预测标签																							
	H ₁	H ₂	H ₃	W ₁	W ₂	W ₃	A ₁	A ₂	A ₃	A ₄	A ₅	Z ₁	Z ₂	Z ₃	M ₁	M ₂	M ₃	M ₄	O ₁	O ₂	O ₃	S ₁	S ₂	S ₃
H ₁	97%	1%	0	0	2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H ₂	2%	98%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H ₃	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W ₁	0	0	0	97%	3%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W ₂	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W ₃	0	0	0	0	1%	99%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₁	0	0	0	0	0	0	99%	0	0	1%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₂	0	0	0	0	2%	0	0	97%	1%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₃	0	0	0	0	0	0	1%	3%	94%	2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₄	0	0	0	0	0	0	5%	0	0	95%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A ₅	0	0	0	0	0	0	1%	0	0	0	97%	0	2%	0	0	0	0	0	0	0	0	0	0	0
Z ₁	0	0	0	0	1%	0	0	0	0	0	0	98%	0	0	0	0	0	0	0	1%	0	0	0	0
Z ₂	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
Z ₃	0	0	0	0	0	0	0	0	0	0	0	1%	2%	97%	0	0	0	0	0	0	0	0	0	0
M ₁	0	0	0	0	0	0	0	0	0	0	0	0	1%	0	99%	0	0	0	0	0	0	0	0	0
M ₂	0	0	0	0	0	0	0	0	1%	0	0	0	0	0	0	99%	0	0	0	0	0	0	0	0
M ₃	0	0	0	0	1%	0	0	0	0	0	0	0	0	0	0	0	99%	0	0	0	0	0	0	0
M ₄	0	0	0	0	1%	0	0	0	0	0	0	0	0	0	0	0	0	99%	0	0	0	0	0	0
O ₁	0	0	0	0	0	0	0	0	0	0	0	0	1%	0	0	0	0	0	99%	0	0	0	0	0
O ₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
O ₃	0	0	0	0	0	0	0	0	0	0	0	0	2%	0	0	0	0	0	0	0	98%	0	0	0
S ₁	0	0	0	0	0	0	0	0	0	0	0	0	2%	0	0	0	0	0	1%	0	0	97%	0	0
S ₂	0	0	0	0	3%	0	0	0	0	0	0	0	1%	0	0	0	0	0	0	0	0	0	96%	0
S ₃	0	0	0	0	0	0	0	0	0	0	0	0	1%	0	0	0	0	0	0	0	0	0	0	99%

从上述实验结果分析,手机的本底噪声可以作为手机的“指纹”很好地进行手机来源识别。在手机来源识别的语音取证领域,手机的本底噪声特征是一个具有区分性的特征。

5 结束语

本文提出了一种新的手机来源识别方法,将本底噪声作为手机的“指纹”进行手机来源识别。由于本底噪声主要取决于手机本身的电路设计和电子元器件的不同,因此以本底噪声作为手机来源识别的区分特征是可行的。实验结果验证了所提出的本底噪声可以作为手机来源识别的特征。而且,较 MFCC 特征,基于本底噪声频谱分布特征的分类识别效果更好。

然而,本文的方法是应用本底噪声做手机来源识别的初步研究,仍有许多局限性。如估计本底噪声的步骤是:先

通过录音自适应端点检测处理得到静音段,然后利用改进的谱减法减去背景噪声得到的。所以,能否准确地估计本底噪声,背景噪声很关键。在实际情况中,不同录音环境的背景噪声是不同的,会影响本底噪声估计的准确性。因此,本底噪声的估计是一项具有挑战性的工作。而综合基准数据库的建立是一项费力的工作,将之作为长远目标坚定地研究下去。

参考文献:

[1] 杨锐, 骆伟祺, 黄继武. 多媒体取证[J]. 中国科学:信息科学, 2013, 43(12): 1654-1672.
YANG R, LUO W Q, HUANG J W. Multimedia forensics[J]. Science China: Information Science, 2013, 43(12): 1654-1672.
[2] GUPTA S, CHO S, KUO C, et al. Current developments and future trends in audio authentication [J]. IEEE Multimedia, 2012, 19(1): 50-59.

表 4 Hanilci 特征的识别率

标签	预测标签																								
	H ₁	H ₂	H ₃	W ₁	W ₂	W ₃	A ₁	A ₂	A ₃	A ₄	A ₅	Z ₁	Z ₂	Z ₃	M ₁	M ₂	M ₃	M ₄	O ₁	O ₂	O ₃	S ₁	S ₂	S ₃	
H ₁	50%	49.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8%	0	0	0	0	0	0	
H ₂	20.8%	79.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
H ₃	2.5%	0	96.7%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8%	0	0	0	0	0	
W ₁	0	0	0	96.7%	1.7%	0.8%	0	0	0	0	0	0	0	0	0	0.8%	0	0	0	0	0	0	0	0	
W ₂	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
W ₃	0	0	0	4.2%	0	94.2%	0	0	0	0	0	0	0.8%	0	0	0.8%	0	0	0	0	0	0	0	0	
A ₁	0	0	0	0	0	0	62.5%	0.8%	5%	31.7%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A ₂	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A ₃	0	0	0	0	0	0	15%	1.7%	75.8%	7.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A ₄	0	0	0	0	0	0	21.7%	0.8%	10.8%	66.7%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A ₅	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	
Z ₁	0	0	0	0	0	0	0	0	0	0	0	85%	0	15%	0	0	0	0	0	0	0	0	0	0	
Z ₂	0	0	0.8%	1.7%	0	0	0	0	0	0	0	0	97.5%	0	0	0	0	0	0	0	0	0	0	0	
Z ₃	2.5%	0	0	0	0	0	0	0	0	0	0	0	0	96.7%	0	0	0	0	0	0.8%	0	0	0	0	
M ₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99.2%	0	0	0	0.8%	0	0	0	0	0	
M ₂	0	0	0	2.5%	0	0.8%	0	0	0	0	0	0	0	0	0	96.7%	0	0	0	0	0	0	0	0	
M ₃	0	0	1.7%	0	0	0	0	0	0	0	0	0	0	0	0	0	97.5%	0	0	0	0.8%	0	0	0	
M ₄	10%	1.7%	5.8%	0	0	0.8%	0	0	0	0	0	0	0.8%	4.2%	0	0	0	75%	0	0	0	1.7%	0	0	
O ₁	0	0	0	0	0	0	0	0	0	0	0	0	2.5%	0.8%	0.8%	0	0	0.8%	87.5%	5.8%	0	0	0	1.7%	
O ₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8%	0	0	0	0	0	93.3%	0	0	0	5.8%	
O ₃	0	0	2.5%	0	0	0	0	0	0	0	0.8%	0	0	0	0	0	4.2%	0	0	0	92.5%	0	0	0	
S ₁	5%	0	4.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5%	4.2%	0	0	80.8%	0	0.8%	
S ₂	0	0	0%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	
S ₃	0.8%	0	4.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8%	0	0	0	94.2%	

- [3] BAO Y, LIANG R. Research progress on key technologies of audio forensics [J]. Journal of Data Acquisition & Processing, 2016, 31(2): 252-259.
- [4] 王昌海, 张建忠, 徐敬东, 等. 基于 HMM 的动作识别结果可信度计算方法[J]. 通信学报, 2016, 37(5): 143-151.
WANG C H, ZHANG J Z, XU J D, et al. Identifying the confidence level of activity recognition via HMM [J]. Journal on Communications, 2016, 37(5): 143-151.
- [5] 贺前华, 王志锋, RUDNICKYA I, 等. 基于改进 PNCC 特征和两步区分性训练的录音设备识别方法 [J]. 电子学报, 2014, 42(1): 191-198.
HE Q H, WANG Z F, RUDNICKYA I, et al. A recording device identification algorithm based on improved PNCC feature and two-step discriminative training[J]. Acta Electronica Sinica, 2014, 42(1): 191-198.
- [6] HANILCI C, ERTAS F, ERTAS T, et al. Recognition of brand and models of cell-phones from recorded speech signals [J]. IEEE Transactions on Information Forensics & Security, 2012, 7(2): 625-634.
- [7] HANILCI C, CEMAL I, ERTAS F. Optimizing acoustic features for source cell-phone recognition using speech signals[C]//ACM Workshop on Information Hiding and Multimedia Security, June 20-22, 2013, Montpellier, France. New York: ACM Press, 2013: 141-148.
- [8] JAHANIRAD M, WAHAB A W A, ANUAR N B, et al. Blind source mobile device identification based on recorded call[J]. Engineering Applications of Artificial Intelligence, 2014(36): 320-331.
- [9] PANAGAKIS Y, KOTROPOULOS C. Automatic telephone handset identification by sparse representation of random spectral features [C]//MM&Sec'12, September 6-7, 2012, Coventry, UK. New York: ACM Press, 2012: 91-96.
- [10] PANAGAKIS Y, KOTROPOULOS C. Telephone handset identification by feature selection and sparserepresentations[C]//IEEE International Workshop on Information Forensics and Security, December 2-5, 2012, Costa, CA, USA. New Jersey: IEEE Press, 2012: 73-78.
- [11] KOTROPOULOS C L. Source phone identification using sketches of features[J]. Biometrics IET, 2014, 3(2): 75-83.



- [12] 王志锋, 贺前华, 李艳雄. 录音设备的建模和识别算法[J]. 信号处理, 2013(4): 419-428.
WANG Z F, H Q H, LI Y X. A modeling and identification algorithm of recording devices [J]. Journal of Signal Processing, 2013(4): 419-428.
- [13] 邹领, 贺前华, 邝细超, 等. 基于设备噪声估计的录音设备源识别[J]. 吉林大学学报(工学版), 2016: 1-8.
ZOU L, HE Q H, KUANG X C, et al. Source recording device recognition based on device noise estimation [J]. Journal of Jilin University(Engineering and Technology Edition), 2016: 1-8.
- [14] AGGARWAL R, SINGH S, ROUL A K, et al. Cellphone identification using noise estimates from recorded audio [C]// IEEE International Conference on Communications and Signal Processing (ICCSP), April 3-5, 2013, Melmaruvathur, India. New Jersey: IEEE Press, 2014: 1218-1222.
- [15] ZHAO H, MALIK H. Audio recording location identification using acoustic environment signature [J]. IEEE Transactions on Information Forensics & Security, 2013, 8(11): 1746-1759.
- [16] LUKAS J, FRIDRICH J, GOLJAN M. Digital camera identification from sensor pattern noise[J]. IEEE Transactions on Information Forensics & Security, 2006, 1(2): 205-214.
- [17] KINNUNEN T, RAJAN P. A practical, self-adaptive voice activity detector for speaker verification with noisy telephone and microphone data [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 26-30, 2013, Vancouver, BC, Canada. New Jersey: IEEE Press, 2013: 7229-7233.
- [18] MALHI A, GAO R X. PCA-based feature selection scheme for machine defect classification [J]. IEEE Transactions on

Instrumentation and Measurement, 2004, 53(6): 1517-1525.

- [19] LU Y, COHEN I, ZHOU X S, et al. Feature selection using principal feature analysis [C]//The 15th ACM International Conference on Multimedia, September 27-29, 2007, Baltimore, MD, USA. New York: ACM Press, 2007: 301-304.

[作者简介]



裴安山(1992-),男,宁波大学信息科学与工程学院硕士生,主要研究方向为多媒体通信、信息安全、移动终端来源检测等。



王让定(1962-),男,博士,宁波大学高等技术研究院教授,主要研究方向为多媒体通信与取证、信息隐藏与隐写分析、智能抄表及传感网络技术。



严迪群(1979-),男,博士,宁波大学信息科学与工程学院副教授,主要研究方向为多媒体通信、信息安全、基于深度学习的数字语音取证等。