



研究与开发

基于语音静音段特征的手机来源识别方法

裴安山, 王让定, 严迪群

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 手机来源识别已成为多媒体取证领域重要的热点问题。提出了一种基于语音静音段特征的手机来源识别方法, 该方法先通过使用自适应端点检测算法得到语音的静音段; 然后将静音段的梅尔频谱系数(MFC)的均值作为分类特征; 最后结合 WEKA 平台的 CfsSubsetEval 评价函数按照最佳优先(BestFirst)搜索进行特征选择, 并采用支持向量机(SVM)对手机来源进行识别。实验部分对 23 款主流型号的手机进行了分类, 结果表明所提特征具有较好的分类性能, 在 TIMIT 数据库和自建的 CKC-SD 数据库上, 平均识别准确率分别为 99.23%和 99.00%。另外, 与语音段 MFC 特征和梅尔倒谱系数(MFCC)特征进行了对比, 实验结果证明所提特征具有更加优越的性能。

关键词: 多媒体取证; 手机来源识别; 静音段; 梅尔频谱特征

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017123

Source cell-phone identification from recorded speech using non-speech segments

PEI Anshan, WANG Rangding, YAN Diqun

College of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: Source cell-phone identification has become a hot topic in multimedia forensics. A novel cell-phone identification method was proposed based on the silent segments of recorded speech. Firstly, the silent segments were obtained using adaptive endpoint detection algorithm. Then, the mean of Mel frequency coefficients (MFC) was extracted as the characteristics for device identification. Finally, the CfsSubsetEval evaluation function of WEKA platform was selected according to the best priority (BestFirst) search, and support vector machine (SVM) was used for classification. Twenty-three popular models of the cell-phones were evaluated in the experiment. Experimental results show that the proposed method is feasible and the average recognition rates are 99.23% and 99.00% on the TIMIT database and the CKC-SD database. At the same time, the proposed feature performs was demonstrated better than the MFC features and the Mel frequency cepstrum coefficients (MFCC) features of the speech segments.

Key words: audio forensics, source cell-phone identification, silent segment, Mel frequency coefficient

收稿日期: 2017-01-17; 修回日期: 2017-03-30

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), The Scientific Research Foundation of Ningbo University (No.XKXL1405, No.XKXL1420, No.XKXL1509, No.XKXL1503), K.C. Wong Magna Fund in Ningbo University



1 引言

随着移动互联网和微芯片产业的发展^[1],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。而各种功能强大且易于操作的数字媒体编辑软件的出现,给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^[2]。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^[3]。多媒体来源取证技术是多媒体取证技术的重要研究内容,包括数字图像来源取证和数字音频来源取证等。数字图像来源取证的目的是对给定的图像来源的真实性和完整性进行研究,并且取得了很多成果^[4-6]。而数字音频来源取证是对音频来源的真实性、完整性等进行验证的。当一段录音样本被作为法庭证据时,对录音的设备痕迹和环境痕迹的检测往往是必须的。

近年来,数字音频来源取证得到了业界越来越多的关注,已取得了一些有影响的成果^[5-12]。Hanilci 等人^[13]通过提取录音文件的 MFCC (Mel frequency cepstrum coefficient, 梅尔倒谱系数) 特征,用于手机品牌和型号的识别,在 14 个不同型号的手机设备的闭集识别实验中,识别率可达到 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) 在手机来源识别中性能的优劣,得出 MFCC 特征在手机来源识别中分类效果最好^[14]。Kotopoulos^[15]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,然后通过映射到低维空间进行降维的方法,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%。Buchholz 等人^[16]采用 256 和 2 048 点窗长的静音段傅立叶系数特征对 7 种不同的麦克风进行鉴别,后者的识别准确率达

到 93.5%。王志锋等人^[17]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%。

现有大多数手机来源识别的研究是基于录音语音部分相关特征的分类研究,本文从静音段中提取出表征录音设备“指纹”的特征^[16,17],并比较了从语音部分和静音部分提取相同的声学特征的手机来源识别的准确率,特征选择方面将离散余弦变换 (discrete cosine transform, DCT) 之前的参数作为分类特征 (MFC 特征)。DCT 的目的是获取频率谱的低频信息,对数据进行降维压缩。但是频率谱的高频信息和 DCT 去除的能量值之间的相关性可成为手机来源识别的分类特征,而且本文采用的特征优化降维算法,能对 MFC 特征进行选择降维,选出区分性最大的特征,同时提高计算效率,因此将 MFC 特征作为分类特征,并与 MFCC 特征进行了比较。实验结果表明,语音静音段 MFC 特征可以作为手机来源识别的特征,在 TIMIT 数据库和自建的 CKC-SD 数据库上,平均识别准确率可以达到 99.23% 和 99.00%,较语音段 MFCC 特征有更加优越的性能。

2 基于静音段特征的手机来源识别算法

基于语音的手机来源识别包括两个步骤:模型训练和语音样本识别。在训练过程中,需要提取数据库中每个手机训练语音样本的特征,并创建一个训练模型。在识别过程中,需从手机测试语音样本中提取特征,并计算与每个手机训练模型的相似度,检测出的相似度最大的手机模型即该测试语音样本的手机来源。

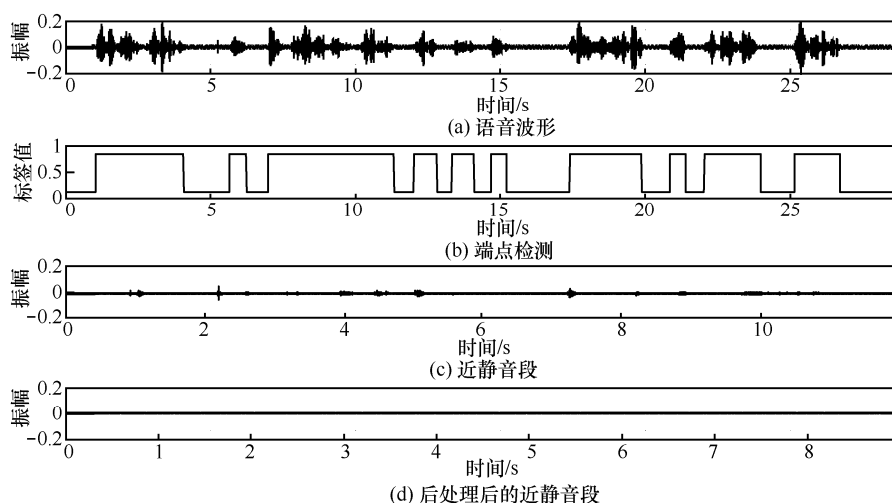


图 1 基于自适应端点检测和后处理操作的静音段估计

2.1 静音段估计

本文采用端点检测技术对静音段进行估计, 选择自适应端点检测算法^[18]进行静音段的提取。其过程为先求原始信号的 MFCC, 再采用谱减法进行去噪, 增大语音段和静音段的能量对比。在此基础上对所有帧能量由低到高进行排序, 寻找一个固定的百分比, 对语音段帧和静音段帧加标签, 然后利用这些帧的 MFCC 训练语音段模型和静音段模型, 最后用训练模型对所有帧加标签。由图 1 (a) 和图 1 (b) 可知, 该方法^[16]能较好地地区分语音段和静音段, 但从合并后的静音段来看, 如图 1 (c) 所示, 在原始信号语音段和静音段的交界处, 还残留少量语音信息, 为了进一步消除不必要的语音信息, 本文使用函数 $ns(\cdot)$ 对静音段进行后处理, 即:

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

其中, Thr 是 $ns(i)$ 升序排列后前 40% 的样本的均值, $\text{find}(\cdot)$ 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, 最后通过拼接得到处理后的静音段信号, 如图 1 (d) 所示, 可以看出本文所修改的端点检测算法能够更好地提取语音的静音段部分。

2.2 特征提取

在语音识别研究中大部分的特征集都是从信

号频谱中提取的, 其中 MFCC 特征是应用最广泛的特征之一, 在参考文献[13]中将以 MFCC 特征作为设备的分类特征, 用于实现语音信号的录音设备来源识别。MFCC 特征提取具体流程如图 2 所示。

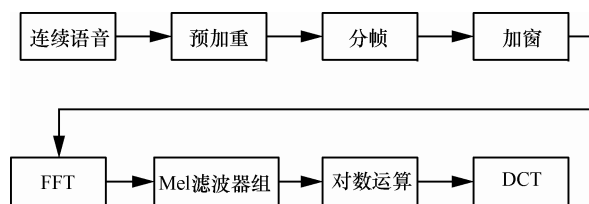


图 2 MFCC 特征提取流程

如第 1 节所述 MFC 特征含有 MFCC 特征所欠缺的高频谱信息和能量值之间的相关性信息, 而上述信息可成为手机来源识别的分类特征, 因此本文将 MFCC 特征提取流程中 DCT 之前所得的语音特征信息 (MFC 特征) 作为手机来源识别的分类特征。另外由于语音信号的内容和说话人信息等与录音设备来源识别无关, 因此本文通过获取静音段部分的 MFC 特征进行录音设备来源识别, 以减少语音内容、说话人情感等不相关因素对识别系统的干扰。MFC 特征提取流程如下。

首先是将第 2.1 节所得的静音段信号通过一个高通滤波器进行预加重, 即:

$$H(Z) = 1 - \mu z^{-1} \quad (2)$$



其中, $\mu=0.97$, 然后加汉明窗进行分帧, 帧长为 30 ms, 帧移为 15 ms, 对加窗后的信号 $x(n)$ 求傅里叶变换, 即:

$$X_a(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}, 0 \leq k \leq N \quad (3)$$

N 表示傅里叶变换的点数, 再按式 (4) 计算每个滤波器组输出的对数能量, 即:

$$s(m) = \ln\left(\sum_{k=0}^{N-1} |X_a(k)|^2 H_m(k)\right), 0 \leq m \leq M \quad (4)$$

其中, $H_m(k)$ 为三角滤波器的频率响应, 定义为:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m) \leq k \leq f(m+1) \\ 0, & k \geq f(m+1) \end{cases} \quad (5)$$

其中, $f(m)$ 为中心频率, $m=1,2,\dots,M$, 这里 M 取 27, 而 $H_m(k)$ 的约束条件是 $\sum_{m=0}^{M-1} H_m(k) = 1$ 。

将 27 维的梅尔频谱参数 $s(m)$ 作为语音的静态参数特征, 为了得到语音的动态参数特征, 求其一阶差分参数, 即:

$$\Delta s_i = s_{(i+j)} - s_{(i-j)}, i=1,2,\dots,T \quad (6)$$

其中, s_i 为式 (4) 得到的第 i 帧的 M 维梅尔频谱系数, j 在本文取 1、2 和 3, 用于计算第 $(i+j)$ 帧和第 $(i-j)$ 帧梅尔频谱系数的一阶差分参数 Δs_i , 得到 108 维的 MFC 特征 (27 维梅尔频谱参数 $s(m)$, 27×3 维一阶差分参数 Δs_i) 作为本文所提出的原始分类特征。

由于 MFC 特征维数较高, 可能导致特征不能完全独立不相关, 因此采用 WEKA 进行特征选择。首先采用最佳优先搜索^[19], 具体步骤如下。

步骤 1 选择 N 个得分最高的特征作为特征

子集, 将其加入一个优先队列。

步骤 2 每次从队列中取出得分最高的子集。

步骤 3 然后穷举向该子集加入 1 个特征后产生的所有特征集, 将这些特征集加入队列。

步骤 4 重复步骤 2、步骤 3 直到产生最优特征子集。

得分是根据 CfsSubsetEval 评价函数^[20] 通过考虑每个特征的预测能力以及它们之间的关联性来评价的, 预测能力为每个特征或特征子集的分类能力, 而关联性指的是特征之间的冗余 (重复) 程度, 首选的特征子集是: 与类具有高相关性, 同时特征之间相关度低的特征子集。将产生的特征子集通过 5 折交叉验证, 具体步骤如下。

步骤 1 将不同机型相同数量语音样本的静音段 MFC 特征构成的数据集按照机型平均分割成 5 等份。

步骤 2 使用其中 1 份数据 (由每个机型 1/5 的语音样本的静音段 MFC 特征构成的数据) 作为测试数据, 其余作为训练数据。

步骤 3 计算测试准确率。

步骤 4 使用不同的测试集, 重复步骤 2、步骤 3。

步骤 5 对测试准确率取平均, 作为对未知数据预测准确率的估计。

验证其有效性, 选出最优的特征子集作为实验的分类特征。

2.3 语音手机来源检测算法

先通过使用本文改进的自适应端点检测算法得到语音的静音段; 提取静音段的 MFC 特征作为分类特征; 采用 WEKA 中的最佳优先搜索和 CfsSubsetEval 评价函数产生特征子集, 再通过 5 折交叉验证特征子集的有效性, 选出最优的特征子集; 然后对所得的特征构建的特征空间采用 SVM^[21] 进行 5 折交叉验证的模型训练和测试。另外由于语音信号幅度和频率特征是动态变化的, 在低维空间难于区分, 所以将它们映射到高维空

间。但这个办法会引起计算复杂度的增加,而核函数正好解决这个问题。因此本文选择径向基核函数(radial basis function, RBF),实验中需对惩罚系数(C)和 $\text{Gamma}(\gamma)$ 进行参数寻优。

3 手机来源识别语音库的建立

3.1 手机设备选择

由于手机和录音笔优越的便捷性,大多数人在录音取证时都会优先选择录音笔或具有录音功能的手机。随着智能手机的普及、其录音功能的完善以及比录音笔等专业录音设备更好的隐蔽性(即使在录音时将手机放在他人可见的位置,一般也不会引起关注和怀疑),越来越多的非专业人员将智能手机作为录音的首选。因此,本文主要以智能手机作为录制设备,选择了目前市面上7个主流品牌的23款流行型号的手机。手机的品牌和型号见表1。

表1 实验中手机的品牌和型号以及类名

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

3.2 语音库的建立

在本文实验中,使用两个不同的数据库,研究在不同的数据库上,手机来源识别系统的性能。第1个数据库是TIMIT库,TIMIT库是语音识别和说话人识别中应用最广泛的数据库。从TIMIT库中选取了100人(男女各一半)的1000个语音样本,通过使用高保真音箱(PhilipsDTM3500)进行回放,并用表1所示的设备同时录制。建立了23000个语音样本的数据库,其中每个手机录

制了1000个语音样本。

第2个数据库是笔者实验室建立的,称其为CKC-SD库。该数据库由表1所示的23台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,要求尽量使用正常的语速、语调、发声响度进行朗读或独白,时长均保证5 min以上。录音环境是一间相对安静的办公室,以参录人员为圆心,所有录音设备按圆弧形平放在办公桌上。每次录制时,均由多名工作人员控制所有设备录音功能的开启和关闭,以保证每段录音在不同设备之间有较好的同步性。因为,每个手机采集了12个人的24段语音,取每段语音中间固定的长度进行分割,将每段语音分割成3 s的语音片段,每个手机得到1200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1000个语音帧数大于8帧的静音段语音,组成求取MFC特征的语音库。

4 实验结果与分析

本文所构建的MFC特征为108维,但是不是所有特征都能优化分类效果,有些特征反而会干扰分类结果。因此,如第2.3节所述,先对特征进行寻优降维,然后采用基于RBF核函数的SVM进行模型训练和分类,同时对惩罚系数(C)和 $\text{Gamma}(\gamma)$ 进行交叉验证寻优。

4.1 MFC特征实验结果与分析

表2给出了经特征寻优和参数寻优后,特征为11维、参数 C 和 γ 分别为30和1.5时,23个手机基于静音段MFC特征在TIMIT库上的识别结果,从表2可算得平均识别准确率为99.23%。本算法可以在TIMIT库上较好地完成23个手机的识别,其中HTC D820t识别效果较差,识别率



表2 静音段 MFC 特征在 TIMIT 库上的识别准确率

AL (标签)	识别准确率																							
	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	Q ₁	Q ₂	Q ₃	S ₁	S ₂	S ₃	
H1	97.2%	2.8%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
H2	5.8%	94.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
H3	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
W1	0	0	0	99.8%	0.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
W2	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
W3	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A1	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A2	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A3	0	0	0	0	0	0	0	0	98.6%	0.4%	0	0	0	0	0	0	0	0	0	0	0	0	0	
A4	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	
A5	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	
Z1	0	0	0	0	0	0	0	0	0	0	0	95.6%	0	4.4%	0	0	0	0	0	0	0	0	0	
Z2	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	
Z3	0	0	0	0	0	0	0	0	0	0	0	3.2%	0	96.8%	0	0	0	0	0	0	0	0	0	
M1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	
M2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	
M3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	
Q1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	
Q2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	
Q3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	
S1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	
S2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	
S3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	

为 94.20%，其次是魅族 Meilan Note 和魅族 MX4，识别率分别为 95.60%和 96.80%，其主要误判为相同品牌内的误判，可能原因是相同品牌内手机语音特征的区别性较不同品牌间手机语音特征的区别性小，更容易被混淆。其他手机都有较高的识别准确率，其中有 18 个手机取得接近完美的识别效果，三星、OPPO、小米等品牌基本可以做到无差错识别。

表3 是经特征寻优和参数寻优后，特征为 20 维、参数 C 和 γ 分别为 35 和 1.7 时，23 个手机基于静音段 MFC 特征在 CKC-SD 库上的识别率，从表 3 可算得平均识别准确率为 99.00%。本算法在 CKC-SD 库上可以较好地完成 23 个手机的分类识别，HTC D610t 和 HTC D820t 识别效果相对较差，识别率分别为 95.80%和 97.40%，其次是 iPhone 6 的识别率为 97.80%，其主要误判均为同品牌内的误判，其他手机都有较高的识别准确率，三星、

OPPO、小米、魅族等品牌基本可以做到无差错识别。

为了研究静音段 MFC 特征较语音段 MFC 特征识别性能的优劣，对基于 TIMIT 库和 CKC-SD 库的两个特征的识别准确率进行对比。语音段 MFC 特征在 TIMIT 库上的特征的维数为 14 维， C 和 γ 的值为 28 和 1.1，手机来源的平均识别准确率为 94.22%；在 CKC-SD 库上的特征维数为 9 维， C 和 γ 的值为 34 和 1.2，手机来源的平均识别准确率为 96.66%。虽然 4 组实验最终采用了不同维度的分类特征和不同的参数进行识别，但是 4 组实验构建的初始分类特征都是 108 维的 MFC 特征，都采用了相同的特征选择方法对原始特征进行寻优降维，而且参数 C 和 γ 是在模型训练和测试时采用相同的交叉验证方式进行参数寻优所得，所以不同组实验之间分类特征识别准确率对比是合理的。不同特征各设备的识别准确率对比

表 3 静音段 MFC 特征在 CKC-SD 库上的识别准确率

AL (标签)	识别准确率																									
	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	Q ₁	Q ₂	Q ₃	S ₁	S ₂	S ₃			
H1	95.8%	4.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
H2	2.2%	97.4%	0.2%	0	0	0	0	0	0	0	0	0	0	0.2%	0	0	0	0	0	0	0	0	0	0		
H3	0	0	99.4%	0.4%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2%	0	0	0		
W1	0	0	0.4%	98.2%	1.2%	0	0	0	0	0	0	0	0	0	0	0.2%	0	0	0	0	0	0	0	0		
W2	0	0	0	1.6%	98%	0	0	0	0.2%	0	0	0	0	0	0	0.2%	0	0	0	0	0	0	0	0		
W3	0	0	0	0.6%	0.4%	98.8%	0	0	0	0	0	0	0	0.2%	0	0	0	0	0	0	0	0	0	0		
A1	0	0	0	0	0	0	99.4%	0	0.4%	0	0	0	0	0	0	0	0	0	0	0	0.2%	0	0	0		
A2	0	0	0.2%	0	0	0	0	99.4%	0.2%	0	0	0	0	0	0	0	0	0	0	0	0.2%	0	0	0		
A3	0	0	0	0	0.2%	0	0.2%	0	98.2%	1.4%	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
A4	0	0	0	0	0	0	1%	0	1.4%	97.6%	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
A5	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0		
Z1	0	0	0	0.2%	0	0	0	0	0	0	0	99.4%	0.2%	0	0	0.2%	0	0	0	0	0	0	0	0		
Z2	0	0	0	0	0	0	0	0	0	0	0	0	99.8%	0	0	0	0	0	0.2	0	0	0	0	0		
Z3	0	0	0	0	0	0	0	0	0	0	0	0	0	99.6%	0	0	0	0	0	0	0	0	0	0		
M1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99.6%	0	0	0	0	0	0	0	0	0.4%		
M2	0	0	0	0.2%	0	0	0	0	0	0	0	0	0	0	0	99.8%	0	0	0	0	0	0	0	0		
M3	0	0	0.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	99.6%	0	0	0.2%	0	0	0	0		
Q1	0.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99.2%	0	0	0.6	0	0	0		
Q2	0	0	0	0	0	0	0	0	0	0	0	0.2%	0.2%	0	0	0	0	0	99.6%	0	0	0	0	0		
Q3	0	0	0	0	0	0	0	0	0.4%	0	0	0	0	0	0	0	0	0	0	99.6%	0	0	0	0		
S1	0	0	0.2%	0	0	0	0	0.2%	0	0	0	0	0	0	0.2%	0	0	0	0	0	99.2%	0.2%	0	0		
S2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4%	99.6%	0	0		
S3	0	0	0.2%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99.8%	0		

情况如图 3 所示,可以看出静音段 MFC 特征较语音段 MFC 特征在手机来源识别中有更加优越的性能。可能的原因是静音段能更好地提取设备特有信息,避免了语音段的内容信息、说话人情感和语调等的干扰。

为了排除相同型号不同个体对实验结果产生

影响的可能性,本文补充了 1 部 iPhone 5 和 1 部 iPhone 6,按照第 3.2 节所述对两部手机进行 TIMIT 库翻录,将得到的每段语音分割成 3 s 的语音片段,接着进行静音段估计提取;然后各取 500 个帧数大于 8 帧的静音样本替换原 TIMIT 静音段样本库中 iPhone 5 和 iPhone 6 的后 500 个静音样本;

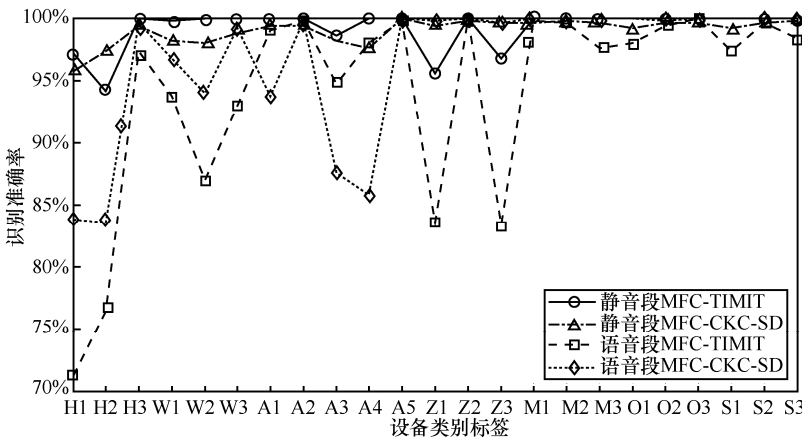


图 3 MFC 特征识别准确率对比



接下来对 23 个型号的手机 TIMIT 静音样本进行 MFC 特征提取,按表 2 所用特征进行特征选择得到 11 维特征;然后取每个型号前 500 个样本特征构成训练集,后 500 个样本构成测试集,参数 C 和 γ 仍为 30 和 1.5。补充实验与表 2 对应实验的识别准确率对比见表 4。

表 4 相同型号的不同设备对静音段 MFC 特征在 TIMIT 库上识别准确率的影响

特征库	相同型号相同设备的语音手机来源识别准确率	相同型号不同设备的语音手机来源识别准确率
TIMIT 库	99.23%	98.13%

由表 4 可以看出,测试集中部分样本用相同型号不同设备的样本替换后识别准确率有一定下降,但是影响较小,对测试集语音样本仍然有较高的识别准确率。因此可得本文算法可以较为准确地对待测语音样本进行来源手机型号的判决,基本不受相同型号不同设备之间差异的影响。

4.2 对比实验结果与分析

为了综合评价所提方法,对比 Hanilci^[13]提出的基于语音段 MFCC 特征的手机来源识别算法。在本文的原始语音数据库上进行测试,训练和测试样本数目与本文实验所用数目一致。该工作中将 48 维的 MFCC 特征作为区分特征,对特征沿时间轴取平均,每个样本得到一个 48 维的特征向量。表 5 是参考文献[13]算法和本文所提算法手机来源识别结果平均识别准确率的对比,对比算法中 SVM 的参数 C 和 γ 分别为 30 和 1.5。

表 5 本文算法与参考文献[13]算法识别准确率对比

特征库	参考文献[13]算法	本文算法(静音段 MFC 特征)
TIMIT 库	93.38%	99.23%
CKC-SD 库	94.77%	99.00%

由表 5 可看出,本文算法识别准确率优于参考文献[13]算法,在 TIMIT 库和 CKC-SD 库上识别准确率分别优化了 5.85%和 4.23%。可能的原因是:参考文献[13]所提特征是从语音段提取的 48 维

的 MFCC 特征,48 维特征可能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,甚至会影响特征的识别效果,而且语音段包含的说话人信息和文本内容可能会对分类产生干扰,同时 MFCC 特征提取过程中的 DCT 损失了语音信号的高频部分信息,对特征的降维也无法达到效果最优。而本文算法将从 TIMIT 库和 CKC-SD 库中静音段提取的 108 维的 MFC 特征,按照最佳优先原则对特征进行交叉验证降维,得到的 11 维和 20 维的 MFC 特征作为分类特征,有效降低了特征维度,避免了冗余特征对识别准确率的影响,同时保留了语音信号的高频部分信息,并降低了计算复杂度,提高了计算效率。因此,相比参考文献[13]算法,本文算法具有特征维数更低和识别准确率更高等优点。

5 结束语

本文提出了一种新的手机来源识别的方法,将语音静音段的 MFC 特征作为手机的“指纹”进行手机来源识别。考虑到语音的静音段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,因此从静音段中提取表征录音设备“指纹”特征进行手机来源识别的方法是可行的^[16,17]。实验结果验证了本文所提出的静音段 MFC 特征是可以作为手机来源识别的特征的。而且,相较于语音段的 MFC 特征和语音段的 MFCC 特征,静音段的 MFC 特征的分类识别效果更好。该方法是应用语音静音段做手机来源识别的初步研究,也存在一定的局限性,如未考虑噪声等其他因素攻击情况下算法的顽健性,而且综合基准数据库的建立也是一项费力的工作,因此在今后的工作中将会对上述问题进行更深入的研究。

参考文献:

- [1] 卢卫, 陆希玉. 4G 时代移动互联网的发展趋势[J]. 电信科学, 2014, 30(5): 51-54.
LU W, LU X Y. Mobile internet trends in 4G era [J]. Telecom-

- munications Science, 2014, 30(5): 51-54.
- [2] 卿斯汉. Android 安全的研究现状与展望[J]. 电信科学, 2016, 32(10): 2-6.
QING S H. Research status and outlook of android security[J]. Telecommunications Science, 2016, 32(10): 2-6.
- [3] 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.
- [4] 胡永健, 简超, 俞兵华. 利用彩色图像信息的源相机辨识策略[J]. 计算机应用, 2010, 30(10): 2694-2697.
HU Y J, JIAN C, YU B H. Source camera identification schemes with color image information[J]. Journal of Computer Applications, 2010, 30(10): 2694-2697.
- [5] 周翠娟, 周治平. 基于模式噪声大分量的手机相机来源检测[J]. 计算机工程, 2013, 39(3): 306-310.
ZHOU C J, ZHOU Z P. Mobile phone camera origin detection based on large components of pattern noise[J]. Computer Engineering, 2013, 39(3): 306-310.
- [6] 王波, 杨福龙. 数字图像来源取证现状与趋势[J]. 信息安全研究, 2016, 6(2): 501-511.
WANG B, YANG F L. An overview and trends on digital image source forensics[J]. Journal of Information Security Research, 2016, 6(2): 501-511.
- [7] GARCIA R D, ESPY W C. Automatic acquisition device identification from speech recordings[C]//2010 IEEE International Conference on Acoustics, Speech and Signal Processing, March 15-19, 2010, Dallas, TX, USA. New Jersey: IEEE Press, 2010: 1806-1809.
- [8] MALIK H, ZHAO H. Recording environment identification using acoustic reverberation[C]//2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 25-30, 2012, Kyoto, Japan. New Jersey: IEEE Press, 2012: 1833-1836.
- [9] GUPTA S, CHO S, KUO C, et al. Current developments and future trends in audio authentication[J]. IEEE Multimedia, 2012, 19(1): 50-59.
- [10] MALIK H, MILLER J. Microphone identification using higher-order statistics[C]//46th International Conference: Audio Forensics, June 1, 2012, Denver, CO, USA. [S.l.:s.n.], 2012: 2-5.
- [11] 贺前华, 王志锋, 朱铮宇, 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, 2013, 42(1): 191-198.
HE Q H, WANG Z F, ZHU Z Y, et al. A recording device identification algorithm based on improved PNCC feature and two-step discriminative training[J]. Acta Electronica Sinica, 2013, 42(1): 191-198.
- [12] JAHNIRAD 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.
- [13] 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.
- [14] HANILCI C, CEMAL I, ERTAS F. Optimizing acoustic features for source cell-phone recognition using speech signals[C]//2013 ACM Workshop on Information Hiding and Multimedia Security, June 17-19, 2013, Montpellier, France. New York: ACM Press, 2013: 141-148.
- [15] KOTROPOULOS C L. Source phone identification using sketches of features[J]. Biometrics IET, 2014, 3(2): 75-83.
- [16] BUCHHOLZ R, KRAETZER C, DITTMANN J. Microphone classification using Fourier coefficients[C]//2009 International Workshop on Information Hiding, June 8-10, 2009, Darmstadt, Germany. Heidelberg: Springer Berlin Heidelberg, 2009: 235-246.
- [17] 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, 29(4): 419-428.
WANG Z F, HE Q H, LI Y X. A modeling and identification algorithm of recording devices[J]. Journal of Signal Processing, 2013, 29(4): 419-428.
- [18] KINNUNEN T, RAJAN P. A practical, self-adaptive voice activity detector for speaker verification with noisy telephone and microphone data[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 26-31, 2013, Vancouver, Canada. New Jersey: IEEE Press, 2013: 7229-7233.
- [19] XU L, YAN P, CHANG T. Best first strategy for feature selection[C]//1988 9th International Conference on Pattern Recognition, May 14-November 17, 1988, Rome, Italy. New Jersey: IEEE Press, 1988: 706-708.
- [20] HALL M A. Correlation-based feature selection for machine learning[D]. Hamilton: The University of Waikato, 1999.
- [21] 何振红. 基于特征加权 FDCT 和模糊最小二乘支持向量机的虹膜识别算法[J]. 电信科学, 2016, 32(3): 92-98.
HE Z H. Iris recognition algorithm based on feature weighting fast discrete curvelet transform and fuzzy LS-SVM[J]. Telecommunications Science, 2016, 32(3): 92-98.

[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, 宁波大学信息科学与工程学院教授、博士生导师, 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、硕士生导师, 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。