doi:10.3969/j.issn.1001-893x.2013.11.014

基于时域统计特征的音频内容取证新算法*

谢 玲[1,**],范明泉[2]

(1.中国西南电子技术研究所,成都 610036;2.西南交通大学 信息科学与技术学院,成都 610031)

摘 要:针对现有音频内容取证算法采用二值图像作为辨识水印所带来的安全隐患,以及基于音频内容或特征生成的辨识水印稳定性不高,易被常规信号处理操作淹没的问题,提出了一种新的基于时域统计特征的音频内容取证算法。通过对音频信号时域统计平均值进行非均匀量化生成辨识水印。理论和实验结果表明通过该方法生成的辨识水印能够抵抗常规信号处理操作,稳定性高。生成的辨识水印存储于认证中心,组建辨识水印库。对音频内容进行取证时,将由该音频生成的辨识水印与从水印库中提取的对应辨识水印进行比对,即可对待取证音频的真实性、完整性进行鉴定。该取证方法操作简便,对不同类型音频均能实现篡改定位,对常规音频信号处理操作的鲁棒性高,有效扩大了基于内容音频取证算法的应用范围。

关键词:音频内容取证; 辨识水印;篡改定位;非均匀量化;时域统计特征;混沌系统

中图分类号:TN912.3;TN919 **文献标志码**:A **文章编号**:1001-893X(2013)11-1476-06

A Novel Audio Content Forensics Scheme Based on Time Domain Statistical Characteristic

XIE Ling[1], FAN Ming-quan[2]

(1. Southwest China Institute of Electronic Technology, Chengdu 610036, China;
2. School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: Many previous audio content forensics schemes adopt binary image as identifying watermark, which introduces security holes to forensics systems. On the other hand, partial content-based or feature-based identifying watermarks have feeblish stability and may be damaged under various signal processing operations. To overcome these problems, a novel audio content forensics scheme based on time domain statistical characteristic is proposed in this paper. The statistical average value of continuous audio samples is used to generate identifying watermark by non-uniform quantization. Theoretical analysis and experimental results show that the generated identifying watermark is robust against various signal processing operations. Various identifying watermarks generated from different audio signals are stored at CA (Center of Authentication). When authenticating the veracity and integrity of audio content, firstly identifying watermark is generated from the to be detected audio, then corresponding identifying watermark is extracted from database of CA, finally the two identifying watermarks for audio content forensics are compared. The proposed forensics scheme has lower computation complexity, and the ability of tamper localization and tolerance against common signal processing operations are excellent. It greatly expands the applicability of content-based audio forensics scheme.

Key words: audio content forensics; identifying watermark; tamper localization; non-uniform quantization; time domain statistical characteristic; chaotic system

1 引 言

近年来,伴随着互联网技术的迅猛发展以及音频压缩技术的日益成熟,以MP3为代表的音乐在互联网上广泛传播,极大地便利和丰富了人们的生活。

* 收稿日期:2013-10-18;修回日期:2013-11-04 Received date:2013-10-18;Revised date:2013-11-04
** 通讯作者:wangyangxie@126.com **Corresponding author**: wangyangxie@126.com

然而,由于网络信息的全透明性和易操作性,以及各种音频信号处理工具的涌现,使得恶意攻击者可以从感知上不留痕迹地对音频数据进行篡改、伪造。在一些重要的实际应用(如新闻媒体、法律证据、电子商务)中,人们需要确切地知道所接收或要使用的音频数据是否真实、是否完整、是否还具有使用价值。因此,如何有效地对音频数据进行真实性、完整性鉴别取证,已成为学术界当前迫切需要解决的难题之一[1-3]。

根据容忍音频数据被篡改的程度来划分,数字音频信号主动性取证技术主要可以分为两类。第一类不允许有任何修改,被称为精确取证,这类取证可用脆弱水印来实现。文献[4]通过修改音频信号混合变换域低、中频系数嵌入二值图像辨识水印实现对音频内容的取证;文献[5]通过修改音频信号小波变换域细节分量嵌入二值图像辨识水印,而在基于音频特征生成的秘密密钥上嵌入二值图像标识水印,实现对音频内容取证和版权保护的双重功能。这类方案大多采用二值图像作为辨识水印,其劣势在于[6]:一是二值图像的使用,增加了信息的传输量,浪费了网络传输带宽;二是若二值图像在传输过程中被篡改,将会增加取证的虚警概率;三是若二值图像在传输过程中被替换,同时对传输的音频处理后嵌入了用来替换的二值图像,则取证时即使音频内容发生了篡改,取证方也觉察不到。第二类允许不改变音频内容的修改,如音频转码、重采样、D/A-A/D 转换、有损压缩、音量调节、去除噪声等,被称为模糊取证,这类取证可利用半脆弱水印、音频感知哈希(Perceptual Hashing)来实现。这类方法大多基于音频内容或特征点生成辨识水印,然而大部分算法生成的辨识水印稳定性较差,容易被常规的音频信号处理操作所淹没。文献[7]将各音频段的重要比特位的能量和作为特征,对该特征进行二值编码生成辨识水印;文献[8]基于语音信号的重要频率带上的能量变化来编码,生成基于内容的特征矢量,用作取证的辨识水印。它们共同面临的问题是音频特征点不稳定,部分特征点易被常规信号处理操作淹没,影响取证准确率[9]。

鉴于此,本文利用音频信号连续采样时域统计平均值这一特征,通过基于混沌系统的非均匀量化手段生成安全的二值辨识水印,解决了传统使用二值图像作为辨识水印带来的安全隐患问题。理论分析和实验结果表明,通过本文方法生成的辨识水印不仅能有效地抵抗常规音频信号处理操作,而且能准确地锁定音频内容被篡改的区域,适合于音频信号的实际取证应用。

2　基于时域统计特征的辨识水印生成

假设原始音频信号表示为 $A=\{a(i)\mid i=0,\cdots,L-1\}$,辨识水印的生成过程框图如图 1 所示。

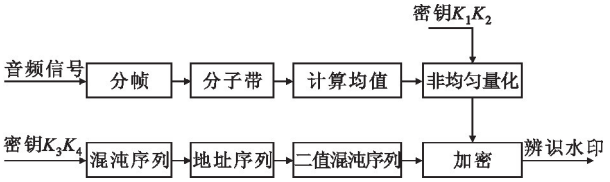

图 1　辨识水印生成过程框图

Fig. 1 Diagram of identifying watermark generation

下面介绍具体步骤。

步骤 1:划分音频帧

将原始音频信号 A 均匀地划分成 M 个互不重叠的音频帧,记为 $A_1(p)$,$p=1,2,\cdots M$,音频帧的长度记为 N,$N=L/M$。

步骤 2:划分音频子带

将每个音频帧均匀地划分成 M_1 个互不重叠的音频子带,记为 $A_2(p,q)$,$q=1,2,\cdots M_1$,音频子带的长度为 N/M_1。

步骤 3:计算时域统计平均值

计算每个音频子带的时域统计平均值,记为 $D(p,q)$,如公式(1)所示:

$$D(p,q)=\frac{\sum_{t=1}^{N/M_1}A_2(p,q,t)}{N/M_1} \tag{1}$$

步骤 4:非均匀量化

首先,将归一化的音频幅值区间[-1,1]均匀地划分为子区间的组合,记为

$$[-1,h(1)),[h(1),h(2)),\cdots,$$
$$[h(i),h(i+1)),\cdots,[h(2/S-1),1]$$

这里 S 是均匀量化的间隔,并且 $h(i)=-1+i\times S$,$i=1,\cdots,2/S-1$。

其次,基于密钥 K_1 和 K_2,通过混沌系统生成伪随机序列 $Q=\{Q(i)\mid i=1,2,\cdots,2/S-1\}$,这里密钥 K_1 是混沌系统的初值,密钥 K_2 是混沌系统的参数。

接着,通过伪随机序列 Q 来扰乱均匀的子区间 $[-1,h(1)),[h(1),h(2)),\cdots,[h(i),h(i+1)),\cdots,[h(2/S-1),1]$,记扰乱后的子区间为

$$[-1,h'(1)),[h'(1),h'(2)),\cdots,$$

$$[h'(i),h'(i+1)),\cdots,[h'(2/S-1),1]$$

其中,$h'(i)=-1+i\times S+\Delta\times Q(i)$,$\Delta$ 是调制参数,$\Delta<S$,$i=1,\cdots,2/S-1$。

最后,根据每个音频子带的时域统计均值,生成对应的二值比特。若统计均值 $D(p,q)$ 属于第 j 个子间隔,$j=1,2,\cdots,2/S$,那么对应的二值比特 $W(p,q)$ 为

$$W(p,q)=\begin{cases}0, & \text{if } \mathrm{mod}(j,2)=0\\ 1, & \text{if } \mathrm{mod}(j,2)=1\end{cases} \tag{2}$$

由此,可得最终整个音频信号对应的二值比特序列 $W_1=\{W_1(k)\mid k=1,2,\cdots,M\times M_1\}$。

步骤 5:地址序列的生成

基于密钥 K_3 和 K_4,通过混沌系统生成伪随机序列

$$Q_1=\{Q_1(i)\mid i=1,2,\cdots,M\}$$

这里密钥 K_3 是混沌系统的初值,密钥 K_4 是混沌系统的参数。将伪随机序列 Q_1 按降序排序,如公式(3)所示:

$$Q_{a(i)}=\mathrm{descend}(Q_1(i)),i=1,2,\cdots,M \tag{3}$$

其中,$a(i)$是混沌序列排序后的地址索引序列。

步骤 6:二值混沌序列的生成

将每个十进制数地址索引 $a(i)$转化为长度为 m 的二值序列,记为 $(a_1a_2\cdots a_d\cdots a_m)_2$,其中,$a_d\in\{0,1\}$,$m=n\times M_1$,$n$ 是整数。接着,将长度为 m 的二值序列均匀地分为 n 组,各组比特相互异或,如公式(4)所示:

$$W_2(i)=(a_1a_2\cdots a_{M_1})\oplus(a_{M_1+1}a_{M_1+2}\cdots a_{2\times M_1})\oplus\cdots\oplus(a_{m-M_1+1}a_{n-M_1+2}\cdots a_m) \tag{4}$$

这样,连接所有的异或值可得到流密码序列 $Q_c=\{Q_c(i)\mid i=1,2,\cdots,M\times M_1\}$。

步骤 7:加密

通过公式(5)获得该音频信号的二值辨识水印 W_c:

$$W_c=Q_c\oplus W_1 \tag{5}$$

最后,将密钥(K_1,K_2,K_3,K_4)及二值辨识水印 W_c 存储于可信认证中心(Authentication Center, CA),组建辨识水印库;当需要对某音频进行取证时,从认证中心 CA 提取对应的密钥及辨识水印用于音频内容的取证。

3 辨识水印的提取及音频内容取证

辨识水印的提取及音频内容取证过程框图如图 2 所示。

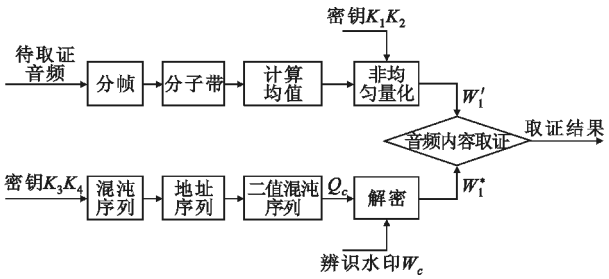

图 2 辨识水印的提取及音频内容取证框图

Fig. 2 Diagram of identifying watermark extraction and audio content forensics

下面介绍具体步骤。

步骤 1:类似于辨识水印的生成过程步骤 1 ~ 4,获得待取证音频信号 A^* 对应的二值比特序列 W'_1。

步骤 2:类似于辨识水印的生成过程步骤 5 ~ 6,获得流密码序列 Q_c,用流密码序列 Q_c 对二值辨识水印 W_c 进行解密,得二值比特序列 W_1^*。

步骤 3:音频内容取证。定义取证序列 $T=\{T(i)\in\{0,1\}\}$,$i=1,2,\cdots,M\times M_1$,T 由公式(6)计算获得:

$$T=W'_1\oplus W_1^* \tag{6}$$

将长度为 $M\times M_1$ 的取证序列 T 依次等分成 M 组,每组的 M_1 个比特对应相应的一个音频帧的内容取证,计算每组元素之和得

$$S(p)=\sum_{i=(p-1)\times M_1+1}^{p\times M_1}T(i),p=1,2,\cdots,M \tag{7}$$

定义

$$T_A(p)=\begin{cases}1, & S(p)\neq 0\\ 0, & S(p)=0\end{cases} \tag{8}$$

当 $T_A(p)=0$ 时,表示对应音频帧的内容没有发生变化;当 $T_A(p)=1$ 时,表示对应音频帧的内容被篡改。

4 性能分析

4.1 辨识水印规模分析

假设原始音频信号的采样率为 f_s(Hz),则通过本文算法生成的辨识水印 W_c 的规模 C_W(b/s)为

$$C_W=\frac{M_1\times f_s}{N} \tag{9}$$

其中,N 是音频帧的长度,M_1 是每个音频帧中的音频子带数。

4.2 辨识水印稳定性分析

本文算法基于音频子带时域统计均值生成辨识水印,辨识水印的稳定性主要取决于音频子带时域统计均值的稳定性。文献[10-11]给出了音频子带时域统计均值对时间尺度修改(Time-Scale Modifi-

cation,TSM)的近似不变性。实际上音频子带时域统计均值对常规音频信号处理操作也具有较强的鲁棒性。

设 $A(t)\mid t\in T_0$ 是音频子带的模拟表示,$n(t)\mid t\in T_0$是音频信号遭受常规信号处理操作后的变化量,这样受污染的音频信号可表示为 $A'(t)\mid t\in T_0$,$A'(t)=A(t)+n(t)$,那么有

$$\overline{A'}=\frac{1}{T_0}\int_{t=0}^{T_0}A'(t)\,\mathrm{d}t=\frac{1}{T_0}\int_{t=0}^{T_0}(A(t)+n(t))\,\mathrm{d}t=\frac{1}{T_0}\int_{t=0}^{T_0}A(t)\,\mathrm{d}t+\frac{1}{T_0}\int_{t=0}^{T_0}n(t)\,\mathrm{d}t \tag{10}$$

一般地,$n(t)\mid t\in T_0$ 服从均匀分布 $N(0,\sigma^2)$,这样公式(10)可演化为

$$\bar{A'}=\frac{1}{T_0}\int_{t=0}^{T_0}A'(t)\,\mathrm{d}t=\frac{1}{T_0}\int_{t=0}^{T_0}A(t)\,\mathrm{d}t+\frac{1}{T_0}\int_{t=0}^{T_0}n(t)\,\mathrm{d}t=\frac{1}{T_0}\int_{t=0}^{T_0}A(t)\,\mathrm{d}t=\bar{A} \tag{11}$$

由公式(11)可以看出,音频子带的时域统计均值在常规信号处理操作前后是不变的。进而可知,音频信号在遭受常规信号处理操作前后,由本文算法生成的辨识水印也是近似不变的。

4.3　辨识水印篡改检测性能分析

由辨识水印的生成及音频内容取证过程可以看出,如果第 p 个音频帧被恶意篡改,那么对应音频帧的二值比特序列将会发生变化,从而 $T_A(p)=1$。

取证序列 $T(i)$ 的元素可以假设为独立随机变量,那么 $T(i)$ 元素全为 0 的概率为 $1/(C_2^1)^{M_1}$。显然在这样的情况下,即使对应的音频帧内容发生变化,也无法取证得到,即漏警概率为 $1/2^{M_1}$。因此,当恶意篡改发生时,篡改检测的理论概率 P_r 为

$$P_r=(1-1/2^{M_1}) \tag{12}$$

由公式(12)可以看出,当音频帧划分的音频子带数越多时,篡改检测的理论概率越高。

5　实验结果

为了验证本文算法的检测可靠性、对恶意篡改的脆弱性及对常规音频信号处理操作的鲁棒性,选取了几类音频信号进行实验,它们均为 WAVE 格式、采样率为44.1 kHz、16 比特量化的音频信号。限于篇幅这里只选取其中具有代表性的音频信号来报道结果,如图 3 所示。算法重要的参数设置如下:测试音频信号样本数 $L=409\,600$,划分的音频帧数 $M=1\,024$,划分的音频子带数 $M_1=4$,M_1 越大,篡改检测的概率越高,均匀量化步长 $S=0.02$,调制系数 $\Delta=0.001$,十进制数地址索引转化为二进制比特位数 $m=12$,所采用的混沌系统为 Logistic 混沌映射。

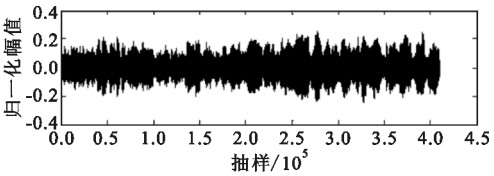

图 3　测试音频信号
Fig. 3 The original audio signal

5.1　检测可靠性

检测可靠性是本取证算法最重要的性能指标。为了证明由测试音频信号生成的辨识水印是唯一的,选取了 57 个音频(包括图 3 所示的测试音频信号)进行测试,检测结果如图 4 和图 5 所示。由图 4 及图 5 的实验结果可以看出,通过本文算法生成的辨识水印与原始测试音频信号是一一对应的。

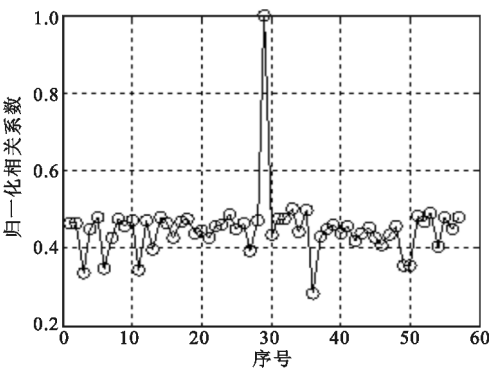

图 4　归一化相关系数结果
Fig. 4 Results of normalized correlation

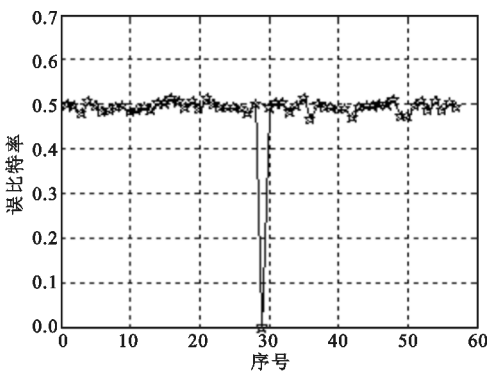

图 5　误比特率结果
Fig. 5 Results of bit error rate

5.2 篡改定位测试

为了评价算法的篡改定位能力,对测试音频信号进行了 3 类攻击。篡改类型 1 是随机地删除部分音频信号,篡改类型 2 是用其他音频信号的内容来替换测试音频信号的部分内容,篡改类型 3 是用测试音频信号的一部分内容替换另一部分内容。图 6(a)所示是删除测试音频信号的前40 000个抽样,图 6(b)所示是测试音频信号第40 001个抽样到第80 000个抽样、第160 001个抽样到第200 000个抽样被其他音频信号替换,图 6(c)所示是测试音频信号第80 001个抽样到第120 000个抽样、第200 001个抽样到第240 000个抽样被测试音频信号的第120 001个抽样到第160 000个抽样、第300 001个抽样到第340 000个抽样替换。图 7 给出了篡改定位结果,其中,$T_A(p)=1$ 表示对应的音频帧内容被篡改,$T_A(p)=0$表示对应的音频帧内容未变。由图 7 可以看出本文算法具有很好的篡改定位能力。

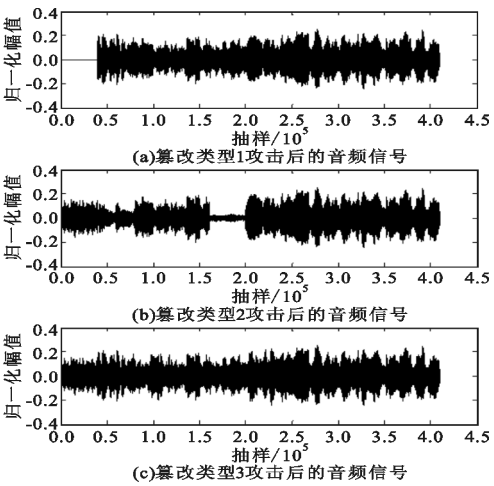

图 6 篡改攻击后的音频信号

Fig. 6 The attacked audio signals

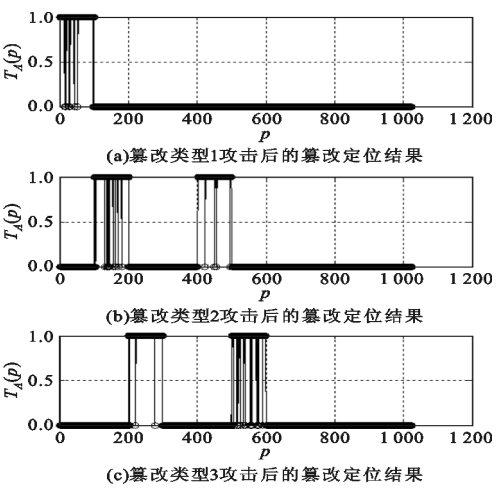

图 7 篡改定位结果

Fig. 7 The tamper location results

5.3 对常规信号处理操作的鲁棒性测试

为了进一步说明本文算法生成辨识水印抵抗常规信号处理操作的能力,进行了一系列典型的信号处理实验。测试音频信号遭受了添加噪声、低通滤波、重采样、重量化、降低噪声、添加回声、MP3 压缩等常规信号处理操作,用误比特率(Bit Error Rate, BER)衡量抵抗常规信号处理操作的鲁棒性,其定义如下:

$$BER=\frac{E}{M\times M_1} \tag{13}$$

其中,E 表示检测的错误比特数。BER 值越小,说明抵抗常规信号处理操作的能力越强。表 1 给出了实验的结果,并与文献[10-11]进行了比较。由表 1 可以看出,本文算法生成的辨识水印对常规音频信号处理操作的鲁棒性较强。

表 1 抗常规信号处理操作的实验结果

Table 1 Experimental results of robustness against common signal processing operations

常规信号处理	*BER*	
	本文算法	文献[10-11]
无攻击	0.000 0	0.000 0
添加噪声 (40 dB)	0.041 3	0.366 7
低通滤波(10 kHz)	0.085 4	0.350 0
重采样 (44.1~176.4~44.1 kHz)	$4.882\,8\times10^{-4}$	0.000 0
重采样(44.1~88.2~44.1 kHz)	$4.882\,8\times10^{-4}$	0.000 0
重采样(44.1~22.05~44.1 kHz)	$2.441\,4\times10^{-4}$	0.000 0
重采样(44.1~11.025~44.1 kHz)	0.001 5	0.133 3
重量化 (16~8~16 b)	0.004 2	0.400 0
降低噪声	0.002 7	0.116 7
添加回声 (100 ms, 50%)	0.000 0	0.000 0
MP3 压缩 (128 kb/s)	0.008 5	0.233 3
MP3 压缩(112 kb/s)	0.008 8	0.316 7
MP3 压缩(96 kb/s)	0.012 0	0.316 7
MP3 压缩(80 kb/s)	0.012 7	0.350 0
MP3 压缩(64 kb/s)	0.017 3	0.350 0
MP3 压缩(56 kb/s)	0.023 7	0.350 0
MP3 压缩(48 kb/s)	0.024 2	0.383 3

6 结 论

基于当前多媒体领域对音频数据真实性、完整性的取证需求,本文通过对音频信号时域统计平均值进行非均匀量化生成辨识水印,提出了一种基于时域统计特征的音频内容取证算法。该方案相比现

有算法而言,具有如下优点:

(1)辨识水印存储于认证中心,无需嵌入到原始音频信号中,确保了音频信号的保真度,特别适用于对音频信号保真度要求很高的场合;

(2)生成的辨识水印具有很好的检测可靠性和篡改定位能力;

(3)生成的辨识水印对常规音频信号处理操作的鲁棒性强。

此外,混沌系统的应用增强了本文算法的安全性。该取证方法计算简单,容易实现,对不同类型的音频均适用和有效。下一步的研究重点是如何实现对低码率下音频压缩数据流的内容认证。

参考文献:

[1] Nishimura R. Audio Watermarking Using Spatial Masking and Ambisonics[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(9):2461-2469.

[2] Xiang Yong, Natgunanathan I, Peng Dezhong, et al. A Dual-channel Time-spread Echo Method for Audio Watermarking[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(2):383-392.

[3] Khan M K, Xie Ling, Zhang Jiashu. Chaos and NDFT-based Spread Spectrum Concealing of Fingerprint-biometric Data into Audio Signals[J]. Digital Signal Processing, 2010, 20(1):179-190.

[4] 王向阳,祁薇. 用于版权保护与内容认证的半脆弱音频水印算法[J]. 自动化学报,2007,33(9):936-940.
WANG Xiang-yang, QI Wei. A Semi-fragile Audio Watermarking for Copyright Protection and Content Authentication[J]. Acta Automatica Sinica, 2007, 33(9): 936-940. (in Chinese)

[5] Chen Ning, Zhu Jie. A Multipurpose Audio Watermarking Scheme for Copyright Protection and Content Authentication[C]//Proceedings of 2008 IEEE International Conference on Multimedia and Expo. Hannover: IEEE, 2008:221-224.

[6] 范明泉,王宏霞. 基于音频内容的混合域脆弱水印算法[J]. 铁道学报,2010,32(1):118-122.
FAN Ming-quan, WANG Hong-xia. Content-based Fragile Audio Watermarking in Hybrid Domain[J]. Journal of the China Railway Society, 2010, 32(1): 118-122. (in Chinese)

[7] Chen Fan, He Hongjie, Wang Hongxia. A Fragile Watermarking Scheme for Audio Detection and Recovery[C]//Proceedings of 2008 IEEE International Conference on Image and Signal Processing. Sanya:IEEE,2008:135-138.

[8] Gulbis M, Muller E, Steinebach M. Content-based Authentication Watermarking with Improved Audio Content Feature Extraction[C]//Proceedings of 2008 IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Harbin:IEEE, 2008:620-623.

[9] 王宏霞,范明泉. 基于质心的混合域半脆弱音频水印算法[J]. 中国科学:信息科学,2010,40(2):313-326.
WANG Hong-xia, FAN Ming-quan. Centroid-based Semi-fragile Audio Watermarking in Hybrid Domain[J]. Science in China: Information Science, 2010, 40(2): 313-326. (in Chinese)

[10] Xiang Shijun, Huang Jiwu. Time-scale Invariant Audio Watermarking Based on the Statistical Features in Time Domain[C]//Proceedings of 2006 International Conference on Information Hiding. Virginia:Springer,2006:1-16.

[11] Xiang Shijun, Huang Jiwu. Histogram-based Audio Watermarking Against Time-scale Modifications and Cropping Attacks [J]. IEEE Transactions on Multimedia, 2007, 9(7):1357-1372.

作者简介:

谢　玲(1981—),女,云南昆明人,分别于 2004 年和 2007 年获西南交通大学学士学位和硕士学位,现为工程师,主要研究方向为雷达信号处理、多媒体信号处理;

XIE Ling was born in Kunming, Yunnan Province, in 1981. She received the B. S. degree and the M. S. degree from Southwest Jiaotong University in 2004 and 2007, respectively. She is now an engineer. Her research interests include radar signal processing and multimedia signal processing.

Email:wangyangxie@126.com

范明泉(1982—),男,江苏南通人,分别于 2004 年和 2010 年获西南交通大学学士学位和博士学位,现为助理研究员,主要研究方向为信息安全。

FAN Ming-quan was born in Nantong, Jiangsu Province, in 1982. He received the B. S. degree and the Ph. D. degree from Southwest Jiaotong University in 2004 and 2010, respectively. He is now an assistant researcher. His research direction is information security.

Email:mqfan_sc@163.com