

基于音频特征和低频系数较小值的水印算法

杨得国, 李 智, 姜金娣

(西北师范大学计算机科学与工程学院, 兰州 730070)

摘 要: 提出一种基于音频特征和 DWT 低频系数相对较小值的数字水印算法。分析音频帧的过零率及短时能量, 选取适当的阈值初步舍弃表明信号中高频信号成分的音频帧, 筛选出待处理的音频帧。将选定的音频帧拼接在一起进行小波变换, 选取低频系数并分段, 通过相邻 3 个水印比特组合的二进制之和确定水印在该段中的嵌入位置, 将系数修改为相邻系数中较小的值。实验结果表明, 该算法通过对音频特征的分析, 能降低提取低频分量的时间复杂度, 实现水印信息的盲检测, 提高水印的鲁棒性。

关键词: 音频数字水印; 音频特征; 短时能量; 过零率; 相对较小值

Watermarking Algorithm Based on Audio Features and Smaller Values of Low Frequency Coefficients

YANG De-guo, LI Zhi, JIANG Jin-di

(College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China)

【Abstract】 This paper presents a digital watermarking algorithm based on audio features and the relatively smaller values of the low frequency of DWT coefficients. The proposed algorithm analyzes the zero-cross ratio and the short-time energy of frame to select the appropriate threshold for initial discarding the high-frequency signal components of the audio frame, selecting the audio frame to be processed. The selected audio frames are stitched together and perform discrete wavelet transform, choosing the low-frequency coefficients and segmenting, using the sum binary value of the neighboring three watermark bit combination to determine the location of the watermark embedding, and make its value smaller. The proposed algorithm for blind watermark detection reduces the extraction time complexity of low-frequency components through the analysis of audio features, and improves the robustness of the watermark.

【Key words】 audio digital watermark; audio feature; short-time energy; zero-cross ratio; relatively smaller value

DOI: 10.3969/j.issn.1000-3428.2012.21.003

1 概述

数字水印技术因其在版权保护和内容认证等方面的重要应用价值, 已成为信号处理领域最热门的研究课题之一。音频数字水印技术^[1]是数字音频作品版权保护的一种非常有效的方法。由于数字音频在多媒体产品中占有较大的比重, 选取数字音频水印作为研究对象, 具有较强的现实意义。数字音频单位时间内的采样点较少, 可嵌入的水印容量相对较小; 同时人类听觉系统(Human Auditory System, HAS)比人类视觉系统(Human Visual System, HVS)更为敏感, 也使得嵌入水印的难度增大。数字音频水印是在研究音频信号的客观特性和人对音频信号的主观感知特性的基础上, 利用载体信号中的一些冗余特性, 实现音频数字作品的版权保护和安全认证等^[2]。

目前提出的数字音频水印算法主要分为时域算法和频域算法^[2-3], 时域算法计算复杂度低, 但水印的鲁棒性

差。文献[4]提出了第一个频域算法, 其良好的性能得到了体现, 较时域算法鲁棒性有了很大的提高, 使得更多的数字水印技术的研究集中于变换域。

根据 IFPI(International Federation of the Phonographic Industry)关于音频水印的鲁棒性定义^[5-6], 音频水印需要满足以下一些基本要求:

(1) 水印具有不可听性, 即嵌入水印不能影响音频的质量。

(2) 嵌入水印后的音频可以提供至少 20 dB 或更高的信噪比, 并且嵌入水印的数据信道至少要有 20 db 的带宽。

(3) 水印必须具有一定的鲁棒性, 能抵抗宿主音频信号上的加性或乘性噪声、MP3 压缩、连续的 D/A 和 A/D 转换、重采样、重量化、滤波、时间伸缩、音频的裁剪与丢失等一般信号处理。

基金项目: 国家自然科学基金资助项目(61165002)

作者简介: 杨得国(1971—), 男, 教授, 主研方向: 网络技术, 多媒体技术; 李 智、姜金娣, 硕士研究生

收稿日期: 2011-12-14 **修回日期:** 2012-02-23 **E-mail:** yangdeguo@nwnu.edu.cn

(4)水印难以删除, 除非声音质量已差到不能容忍的程度。

(5)原则上水印的检测不应该需要原始音频, 即实现盲检测, 因为寻找原始音频是十分困难的。

本文利用混沌扩频技术^[7]对原始水印信息进行加密, 使水印信息满足自相关性强、互相关性弱的特性, 增强水印的安全性; 混沌序列对初值的敏感性高、安全性强, 具有普通伪随机序列所没有的低通特性, 以抵抗低通滤波等攻击。

2 原始水印加密处理

嵌入的水印信息可以是音频、二值图像、彩色图像等, 本文以自制的 52×52 像素的 bmp 二值图像为例。要将二维的图像嵌入一维的音频中, 首先对水印图像进行降维, 为消除元素间的相关性, 增强算法的鲁棒性, 提高水印的安全性, 本文算法利用 Logistic 混沌二进制序列对水印进行混沌加密^[8]。将产生混沌二进制序列时的参数 λ 、初值 x_0 、起始位置 a (即从序列的任意位置 a 开始截取新序列)、扩频因子 C 共同作为用户的私钥 $key(\lambda, x_0, a, C)$, 在私钥未知的情况下, 即使提取出水印信息也无法恢复原始水印图像。

原始水印图像及置乱后的水印图像如图 1 所示。



图 1 水印图像

3 音频帧的过零率与短时能量特征分析

原始音频为非平稳信号, 对其进行分帧加窗分成若干个短时平稳信号, 通常 $10\text{ ms} \sim 20\text{ ms}$ 可认为是短时平稳信号, 其频谱特性和某些物理特性参量可近似地看作是不变的, 相邻两帧之间有 $1/3 \sim 2/3$ 的重叠。对原始音频信号进行分帧加窗, 可以减少信号处理的计算量, 减小频率泄漏误差(减少变换窗的边界效应), 同时也可以降低裁剪等攻击造成的误差扩散。

在本文方法中, 根据音频信号的局部特征来选择合适的待处理的音频帧。这些音频帧在遭受常见信号处理攻击后, 其低频系数值发生变化较小, 提高了水印方法的鲁棒性。本文采用音频信号的 2 个局部特征分别是过零率和短时能量^[9]。

设第 n 帧音频信号的短时能量用 E_n 表示, 则其定义为 $E_n = \sum_{j=1}^F x_n^2(j)$ 。其中, F 为信号帧长。音频帧的短时能量是指一帧内所有采样点的幅值的平方和, 直接反映了一段时间内音频信号的强度, 短时能量越强的音频帧具有更好

的掩蔽特性。

设第 n 帧音频信号的过零率用 Z_n 表示, 其定义为:

$$Z_n = \frac{1}{2} \sum_{j=1}^F |\text{sgn}[x_n(j)] - \text{sgn}[x_n(j-1)]| \quad (1)$$

其中, $\text{sgn}[\]$ 是符号函数, 即 $\text{sgn}[x] = \begin{cases} 1 & (x \geq 0) \\ -1 & (x < 0) \end{cases}$ 。过零率

是指一帧内相邻的采样点改变符号的次数, 反映了音频信号在一帧中幅值变化的剧烈程度, 清音的过零率大, 浊音的过零率小。过零率高表明音频信号中有高频成分, 过零率越小的音频帧具有更好的稳定性。

采用这 2 种音频特征作为选择音频帧的标准是基于对不同类型的音乐和语音的大量实验结果, 本文通过实验来确定音频帧所需满足的短时能量和过零率的条件, 从而舍弃表明信号中高频信号成分的音频帧, 筛选出待处理的音频帧。

实验设计步骤如下:

(1)选取多种不同类型的音频段各 20 s, 对每一种类型的音频信号首先按长度 F 分帧, 得到候选的待嵌入水印的音频帧, 计算每一个音频帧的短时能量和过零率, 对每一个音频帧进行 3 级 db4 小波分解, 提取低频信号的小波系数。

(2)使用顽健性标准测试工具 StirMark for Audio^[10]对各种类型的原始音频信号进行攻击, 对遭受攻击后的各个类型的音频信号依次重复步骤(1)的操作, 分别比较同一类型的音频帧的短时能量、过零率以及低频信号的小波系数的改变程度。

实验结果分析表明, 过零率限制在某一阈值下的帧具有较强的抗攻击能力, 并且在这一过零率阈值的控制下, 能量较大的音频帧抗攻击能力越强, 其逼近信号的小波系数平均值的改变程度越小, 因此, 这些音频帧将具有更强的抗攻击能力, 这种选择方案更好地体现了时域掩蔽特性。可见在选择待处理的音频帧时, 应将音频帧的过零率与短时能量结合考虑。

4 水印方法

4.1 水印嵌入算法

由于音频信号是一维的非平稳信号, 而小波变换具有多分辨率特性, 它在时域和频域的局部化定位观测与人耳的听觉分辨特性类似, 因此小波变换尤其适用于音频信号的分析与处理。

根据前面对音频信号的特征分析可知, 在水印嵌入时, 设置 2 个阈值 E_0 和 Z_0 , 选取满足条件: 过零率 $Z_n \leq Z_0$ 且短时能量 $E_n \geq E_0$ 的音频帧, 将选定的音频帧拼接在一起并进行小波变换。小波变换后的低频分量集中了音频的绝大部分能量, 代表音频信号感知上最重要的部分, 对一般的信号处理具有稳定性。将系数修改为相邻系数中较小的值。因此, 本文算法中选取低频系数并分段, 每段嵌入一个水印比特, 通过相邻 3 个水印比特组合的二进制之和确定水印在该段中的嵌入位置, 将系数修改为相邻系数

中较小的值。

本文的水印嵌入算法步骤如下:

(1) 设 S 是长度为 L 的原始音频信号:

$$S = \{s(i), 1 \leq i \leq Len\} \quad s(i) \in \{0, 1\} \quad (2)$$

将 S 进行分帧后表示为:

$$s_{kj} = s((k-1) \times F + j) \quad 1 \leq k \leq K, 1 \leq j \leq F \quad (3)$$

其中, s_{kj} 表示第 k 帧音频的第 j 个信号值; K 表示音频信息的总帧数; F 表示每帧音频的样点数。

(2) 依次计算 K 帧的过零率 Z_n 和短时能量 E_n , 选取满足条件: 过零率 $Z_n \leq Z_0$ 且短时能量 $E_n \geq E_0$ 的音频帧。

(3) 将选定的音频帧拼接后进行 3 级 “db4” 小波变换并提取低频系数: $ca3 = \{ca3(i), 1 \leq i \leq L\}$, 其中, L 为其长度。

将低频系数 8 个分为一段, 分段后第 t 段的低频系数为: $ca3(t) = ca3((t-1) \times 8 + m)$, 其中, $1 \leq m \leq 8$, 利用相邻 3 个水印比特组合的二进制之和确定水印的嵌入位置。

设水印比特表示为 $w(v)$, 依据下式确定第 v 个水印比特在第 v 段中的嵌入位置 $q(v)$:

$$q(v) = w(v-2) \times 4 + w(v-1) \times 2 + w(v) \times 1 + 1 \quad (4)$$

其中, $w(v-2)$ 、 $w(v-1)$ 、 $w(v)$ 表示相连的 3 个水印比特, $1 \leq q(v) \leq 8$ 。

(4) 对 $q(v)$ 的系数值进行修改, 使之成为相邻位置系数中较小的值, 具体修改过程按下式进行处理:

$$\begin{aligned} & \text{if } w(v) = 0 \\ & \quad \begin{cases} \text{if } ca3((k-1) \times 8 + q(v)) \geq ca3((k-1) \times 8 + q(v) + 1) \\ ca3((k-1) \times 8 + q(v)) = ca3((k-1) \times 8 + q(v) + 1) - \alpha \end{cases} \\ & \text{if } w(v) = 1 \\ & \quad \begin{cases} \text{if } ca3((k-1) \times 8 + q(v)) \geq ca3((k-1) \times 8 + q(v) - 1) \\ ca3((k-1) \times 8 + q(v)) = ca3((k-1) \times 8 + q(v) - 1) - \alpha \end{cases} \end{aligned} \quad (5)$$

由式(5)可知, 在最理想的情况下, 本文算法是零水印算法, 即不需要对低频系数做任何修改; 当低频系数需要修改时, 将其修改为较小值即可, α 为嵌入强度因子。由于在确定第 v 个水印比特 $w(v)$ 在第 v 段中的嵌入位置时, 需要用到 $w(v-2)$ 、 $w(v-1)$, 当确定第 1 个和第 2 个水印比特嵌入位置时, $w(-2)$ 、 $w(-1)$ 不存在, 因此本文算法中将 $w(1)$ 、 $w(2)$ 作为密钥 $key2$ 传送给检测端, 基于此得出其余的水印比特信息。

4.2 水印检测方案

本文算法实现了盲检测, 具体提取水印的方法如下:

使用与嵌入水印时相同的方法步骤, 提取 3 级小波分解后的低频系数, 每 8 个分为一段, 根据密钥 $key2$ 及下式提取出每段嵌入的水印比特:

$$\begin{cases} \text{if } ca3((k-1) \times 8 + T) > ca3((k-1) \times 8 + T + 1) \\ w(v) = 1 \\ \text{else } w(v) = 0 \end{cases} \quad (6)$$

其中, $T = w(v-2) \times 4 + w(v-1) \times 2 + 1$ 。

利用密钥 $key(\lambda, x_0, a, C)$ 生成 Logistic 混沌二进制序列, 作用于提取出的水印信息, 最终恢复出水印信息 W 。

5 仿真实验与结果分析

本文利用 MatlabR2007b 进行仿真实验, 选取一段长度为 20 s、采样频率为 44.1 kHz、16 位量化精度的单声道音乐进行测试, 原始音频信号如图 2 所示。

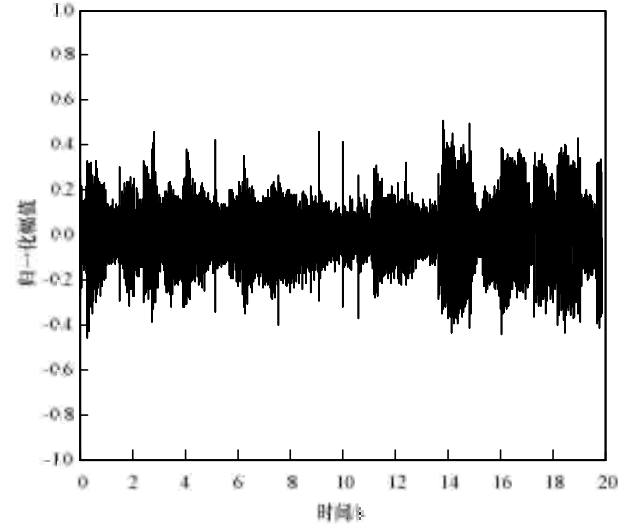


图 2 原始音频信号

本文以 20 ms 为一段, 相邻两帧之间有 2/3 的重叠对音频信号进行分帧, 即 882 点为帧长、294 点为帧移进行分帧, 采用 db4 对各段音频信号进行 3 级小波变换。待嵌入水印为 bmp 二值图像, 大小为 52×52 像素。混沌扩频序列的密钥 $key(3.771\ 3, 0.6, 200, 5)$ 。经计算得: $Z_0 = 36.586\ 4$, $E_0 = 2.869\ 4$ 。

为检测本文算法的鲁棒性, 对原始音频信号进行下面的信号攻击处理, 采用顽健性标准测试工具 StirMark for Audio 实现^[10]。其中, 回声干扰引入 50 ms 的回声; 低通滤波采用截止频率为 8 kHz 的 CHBYSHEV 滤波器; 重采样先将原始信号由原来的 44.1 kHz 下采样到 16.0 kHz, 再恢复采样率至 44.1 kHz; 重量化先将 16 bit 的音频信号量化为 8 bit, 再重新量化到 16 bit; 添加噪声是在音频信号在中加入均值为 0、方差为 0.01 的高斯白噪声; MP3 压缩先将音频信号压缩至 128 kb/s, 再加压缩为 WAV 格式。

本文使用峰值信噪比(PSNR)、比特误码率(BER)和归一化相关系数(NC)来衡量算法的性能^[2]。PSNR 用来评价原始音频和嵌入水印后的音频之间的差别; 峰值信噪比值越大, 说明图像处理效果越好, 鲁棒性越强。BER 用来衡量提取的水印与原始水印的误差程度; NC 则是对提取的水印与原始水印的相似程度进行度量, 归一化相关系数值越大越好。

音频信号受各种攻击后提取的水印图像以及峰值信噪比 PSNR、相关系数 NC 和误码率 BER 的值见表 1。由实验结果可知, 本文中的算法对常见的攻击, 如回声干扰、低通滤波、重采样、重量化、添加噪声和 MP3 压缩等都具有很好的鲁棒性, 比文献[7]中的方法在各方面都有所提高。

表 1 不同类型攻击下提出的水印图像及其实验结果

各类攻击	本文算法				文献[8]算法			
	PSNR/dB	NC	BER	提取水印图像	PSNR/dB	NC	BER	提取水印图像
无攻击	45.717 60	1.000 00	0.000 0		39.387 8	1.000 0	0.000 0	
回声干扰	41.849 87	0.995 25	0.005 2		39.387 8	0.953 6	0.053 4	
低通滤波	41.467 20	1.000 00	0.000 0		40.275 9	1.000 0	0.000 0	
重采样	40.099 70	1.000 00	0.000 0		40.275 9	0.964 6	0.047 8	
重量化	41.222 20	1.000 00	0.000 0		38.940 1	1.000 0	0.000 0	
添加噪声	43.065 00	0.993 60	0.007 0		31.074 3	0.916 1	0.092 1	
MP3 压缩	30.474 70	0.877 10	0.123 9		29.844 7	0.842 3	0.170 1	

6 结束语

本文提出一种根据 DWT 低频系数相对较小值来嵌入水印的算法, 结合音频帧的过零率及短时能量这 2 个音频特征初步舍弃表明信号中高频信号成分的音频帧, 筛选出待处理的音频帧, 此方法减少了提取低频信号的计算量, 同时该算法利用混沌扩频技术对水印进行处理, 消除元素间的相关性, 增强算法的鲁棒性, 提高水印的安全性。实验结果表明, 该算法能够实现水印信息的盲检测, 降低计算量, 并提高水印的鲁棒性。

参考文献

[1] Petitcolas F A P, Anderson R J, Kuhn M G. Information Hiding: A Survey[J]. Proc. of the IEEE, 1999, 187(7): 1062-1078.

[2] 李 伟. 鲁棒性数字音频水印算法研究[D]. 上海: 复旦大学, 2004.

[3] Zanetakis T G, Essl G, Cook P. Audio Analysis Using the Discrete Wavelet Transform[C]//Proc. of International Conference on Acoustics and Music: Theory and Applications. Skiathos, Greece: [s. n.], 2001.

[4] Cox I J, Kilian J, Leighton T, et al. Secure Spread Spectrum Watermarking for Multimedia[J]. IEEE Transactions on Image Processing, 1997, 6(12): 1673-1687.

[5] Huang Jiwu, Wang Yong, Shi Y Q. A Blind Audio Watermarking Algorithm with Self-synchronization[C]//Proc. of IEEE International Symposium on Circuits and Systems. Arizona, USA: [s. n.], 2002: 627-630.

[6] Swanson M D, Zhu Bin, Tewfik A H. Robust Audio Watermarking Using Perceptual Masking[J]. IEEE Transactions on Signal Processing, 1988, 66(3): 337-355.

[7] 纪 震, 李慧慧, 肖薇薇. 基于混沌序列的数字水印信号研究[J]. 电子学报, 2004, 32(7): 1131-1134.

[8] 朱有忠, 向 俊. 基于 DWT 和相对较小值的音频盲水印新算法[J]. 吉林大学学报, 2010, 28(2): 203-207.

[9] 蔡莲红, 黄德智, 蔡 锐. 现代语音技术基础与应用[M]. 北京: 清华大学出版社. 2003.

[10] Cox I J, Kilian J, Leighton T, et al. A Secure Robust Watermark for Multimedia[C]//Proc. of Workshop on Information Hiding. Cambridge, UK: Spring-Verlag, 1996: 185-206.

编辑 任吉慧