

一种基于 MPEG-2 AAC 编码的音频水印方法

王敬¹⁾ 杨扬²⁾ 肖蓉³⁾

1) 北京科技大学网络中心, 北京 100083 2) 北京科技大学信息工程学院, 北京 100083

3) 中国农业银行软件开发中心, 北京 100073

摘要 提出了一种针对 MPEG-2 AAC 压缩音频的脆弱水印算法. 该算法利用了 AAC 中 MDCT 量化系数大于 15 的 Huffman 编码特性来嵌入水印, 并依据水印的频率分布和音频信号感知熵 PE 的大小对嵌入算法进行了分析和改进. 实验表明, 该算法具有较高隐藏率和良好的不可感知性, 并且水印的嵌入和提取过程十分方便快捷, 适合进行实时分析.

关键词 音频水印; AAC; 压缩域; Huffman 编码

分类号 TP 309.2

Audio watermarking based on MPEG-2 AAC

WANG Jing¹⁾, YANG Yang²⁾, XIAO Rong³⁾

1) Network Center, University of Science and Technology Beijing, Beijing 100083, China

2) School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China

3) Software Development Center, Agriculture Bank of China, Beijing 100073, China

ABSTRACT A fragile audio watermarking method for MPEG-2 AAC bitstreams was proposed. The algorithm of watermark embedding was carried out by modifying the LSBs (least significant bits) of quantized MDCT coefficient's magnitude that is greater than 15, and was improved by considering the watermark's distribution in frequency and the perceptual entropy of host audio. Experimental result shows the proposed watermarking scheme is not appreciable, and it is possible to insert high hide rate additive data effectively into encoded bitstreams. The embedding and retrieving process is easy and it is adaptive for real-time analysis applications.

KEY WORDS bitstream watermark; AAC; compression; Huffman coding

近年来, 音频压缩技术的不断进步使得数字音频的制作和传播越来越容易, 但同时也对这些数字音频文件的版权保护提出了更高的要求. 通常采用数字水印技术来对数字音频文件的版权进行保护. 由于 MPEG-1 layer 3 (MP3) 音频编码和新一代的 MPEG-2 Advanced Audio Coding (AAC) 是目前使用相当广泛的音频压缩技术, 因此针对这两种压缩算法的数字音频水印技术是近年来学术界研究的热点之一, 已经有了一些研究成果.

针对 MPEG 编码的水印算法可以分为四类: (1) 利用 MP3 的内循环的中止条件嵌入水印. 这种算法对攻击比较敏感, 如著名的 MP3Stego 水印技术^[1]. (2) 在 MDCT 系数上嵌入水印. MP3 压缩和

水印嵌入同时进行, 根据某些 MDCT 系数的小数点后第 1 个非零位置 P 在经过 MP3 编码器压缩后不会发生变化的原理, 通过改变 MDCT 系数的第 1 个非零位置来嵌入水印^[2-3]. 该方法可以抵御多次解压缩/压缩的攻击, 但是不能完全有效地检测嵌入水印的信息, 水印嵌入比例小. (3) 在比例因子嵌入水印. 文献[4-5]均采用了这种方法. 由于压缩音频中所含比例因子本身就很少, 因此嵌入的水印量小, 并且该算法过程复杂, 计算量较大. (4) 通过 Huffman 编码的特性嵌入水印. 如文献[6]通过一种将 MP3 中某些 Huffman 码书替换成和它有相同码长的码书来嵌入水印.

本文将通过 MPEG-2 AAC 的 Huffman 编码的

特性来嵌入水印,利用量化系数大于 15 的 Huffman 码字作为宿主进行嵌入,并在此基础上对嵌入算法进行分析和改进,提高嵌入水印的不可感知性. Kim^[7]对 MP3 文件使用 Huffman 编码嵌入水印的方法为本文提供了思路.

1 MPEG-2 AAC 编码^[8]

1.1 MPEG-2 AAC 简介

为了达到更高的编码质量和效率, MPEG-2 AAC(先进音频编码)^[8]打破了对 MPEG-1 layer 3 (MP3)的向下兼容性,在 MPEG 系统上进一步改进并加入了很多新的功能,大大增强了编码的灵活度,在保证音质的同时更大限度地压缩了码率.

AAC 编码可以分为频谱变换、量化、熵编码和比特流封装四步. 首先,对输入的 PCM 信号分段,每帧信号通过改进的离散余弦变换(MDCT),输出 1024 个频谱分量. 再将这些频谱分量依据不同的采样率和变换块类型划分成数 10 个不同带宽的比

例因子频带,比例因子频带的划分尽可能地拟合人耳的临界频带. 心理声学模型用来估计每一个比例因子频带的最大可允许失真. 通过内外双层循环迭代计算出每个比例因子频带的量化步长,以此对 MDCT 系数进行量化. 量化步长一方面要满足编码使用的比特数小于可分配的比特数,另一方面还要使量化噪声在心理声学模型允许范围之内,然后用 Huffman 编码对量化后的系数进行编码,最后进行比特流封装,得到压缩后的码流.

1.2 Huffman 编码

在 AAC 中的 Huffman 编码是一种无损编码. 需要进行 Huffman 编码的参数包括各子带的比例因子和经过量化的 MDCT 谱线,对比例因子的编码比较简单. 对谱系数进行 Huffman 编码时, AAC 编码一共使用了 12 本码书. 每个 Huffman 码书所能表示的量化频谱系数的最大绝对值、不同码书中 n 重组的个数、是否有符号等情况如表 1 所示.

表 1 Huffman 码书
Table 1 Huffman codebook

码书序号	n 重组	最大绝对值	有符号值	码书序号	n 重组	最大绝对值	有符号值
0	—	0	—	6	2	4	是
1	4	1	是	7	2	7	否
2	4	1	是	8	2	7	否
3	4	2	否	9	2	12	否
4	4	2	否	10	2	12	否
5	2	4	是	11	2	16	否

有两个码书需要特别说明: 码书 0 专门针对在一个编码区内的各系数都为 0 的情况; 码书 11 能够表示绝对值大于等于 16 的量化值. 在量化值大于或等于 16 时,使用一种称为 escape coding 的机制来实现无损编码. 这样量化系数可以用 escape sequence 来表示, escape sequence 由以下几部分组成:

$$\text{escape sequence} =$$

$\langle \text{escape-prefix} \rangle \langle \text{escape-separator} \rangle \langle \text{escape-word} \rangle$.

其中, $\langle \text{escape-prefix} \rangle$ 是一个由 N 位“1”来表示的序列; $\langle \text{escape-separator} \rangle$ 只有一位,固定用“0”来表示,用来分隔 $\langle \text{escape-prefix} \rangle$ 和 $\langle \text{escape-word} \rangle$; $\langle \text{escape-word} \rangle$ 是一个 $N+4$ 位的无符号整数, N 保证下面的式子成立:

$$\text{量化系数} = 2^{N+4} + \langle \text{escape-word} \rangle \quad (1)$$

根据上面的式子可以知当量化系数为 16 时,

$N=0$,得到的 escape-sequence 是“00000”. 同理可知“1011111”的 escape-sequence 代表值为 63 的量化系数. escape-sequence 反映了量化值的真实大小,并且编码后的 escape-sequence 在解压端能够毫无误差地被还原,这给水印的嵌入和提取提供条件.

2 水印嵌入

本文的水印嵌入是对 Huffman 码字进行修改,因此水印可以在音频压缩过程中嵌入,也可以在压缩好的 AAC 音频上直接嵌入. 本文在压缩过程中嵌入水印,采用了视觉可辨的二值图像,这种水印在用于数字音频版权保护或对其内容的原始性进行校验时具有视觉的直观性.

2.1 数字水印预处理

为了消除二维水印图像 W 的像素空间相关性,提高数字水印算法的鲁棒性,确保水印图像某一部分受到破坏后仍能全部或部分地恢复水印,首先

对二维水印图像进行置乱变换. 本文采用了 Arnold 变换^[9]对二维水印图像 W (大小为 $N \times N$) 进行置乱变换. 数字化后的图像可以看作一个矩阵, 令矩阵的坐标 $x, y \in \{0, 1, 2, 3, \dots, N-1\}$, 于是 Arnold 变换可以表示为:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \pmod{N} \quad (2)$$

由此作迭代程序, 直到图像变得杂乱无章. 接下来, 对置乱后的二维水印图像进行降维处理 (即将置乱后的二维水印图像转换为二维的数字水印序列 V , 以便将二维的灰度图像嵌入到一维的数字音频信号中), 即:

$$V = \{v(k) = w(i, j), 0 \leq i < N,$$

$$0 \leq j < N, k = i \cdot N + j\} \quad (3)$$

2.2 基于 escape coding 的水印嵌入算法

通过修改 escape-word 的最低有效位 (LSB) 来嵌入水印. MPEG-2 AAC 中规定, 量化后的系数最大值不能超过 8 191, 因此 escape-word 的最低有效位数可以从 4 bit 到 12 bit. 为了减小嵌入水印产生的听觉失真, 只使用 1 bit 和 2 bit 的最低有效位来嵌入水印. 当 escape-prefix 中“1”的个数为零即量化系数在 16 到 31 时, 只取 1 bit 最低有效位进行嵌入; 当 escape-prefix 中“1”的个数大于零即量化系数大于或等于 32 时, 取 2 bit 最低有效位进行嵌入. 具体的嵌入方法是根据当前的水印象素值来对 escape-word 的最低有效位进行修改 (如表 2).

表 2 嵌入方法

Table 2 Embedding method

条件	最低有效位 (LSB)/bit	水印象素值	嵌入水印
escape-prefix 中“1”的个数=0	1	255	'1'
		0	'0'
escape-prefix 中“1”的个数>0	2	255, 255	'11'
		255, 0	'10'
		0, 255	'01'
		0, 0	'00'

为了分析嵌入水印的结果, 需要对水印嵌入后的音频的听觉质量进行评价. 国内经常使用的信噪比 (signal to noise ratio, SNR) 方法没有考虑到人类听觉系统的感知特性, 对于高质量的音频编解码系统而言这种方法并不实用. 因此本文采用 ITU-R 提出的感知音频质量评价 PEAQ (perceived evaluation of audio quality) 模型^[10]来评价音频听觉质量.

PEAQ 模型是一种音频质量客观评价模型. 它将参考信号和失真信号通过感知声学模型模拟人对音频信号的感知, 认知模型将感知声学模型输出值在频域和时域进行综合产生一系列模型输出变量 (MOV), 通过一个神经网络, 计算出最终的客观失真等级 ODG (objective difference grade). ODG 有五个等级, 见表 3. 实验结果如表 4 所示.

表 3 ODG 的五个等级

Table 3 Five classes of ODG

客观失真等级 (ODG)	0	-1	-2	-3	-4
感知效果	不可感知	可感知	有些恼人	恼人	非常恼人

表 4 水印嵌入结果

Table 4 Results of watermark embedding

音乐名称 (48 000 Hz, 133 kbps)	文件长度/ s	嵌入水印/ bit	每秒嵌入水印/ (bit · s ⁻¹)	感知音频质量评价: 客观失真等级, PEAQ: ODG
流行音乐	19.63	18 075	920.8	-0.312
爵士乐	20.73	30 437	1 468.3	-0.390
经典音乐	19.71	42 879	2 175.5	-0.773
民歌	20.28	18 447	909.6	-0.744
鼓乐	20.40	23 833	1 168.2	-0.146
摇滚乐	20.06	11 261	561.7	-0.247
演唱	20.10	7 968	396.4	-0.245
乡村音乐	20.20	30 139	1 492.0	-0.757

2.3 嵌入算法改进

不难发现,虽然 escape coding 在各类音乐压缩时都会发生,但发生的频率却各不相同,嵌入水印的码率会随着宿主不同有较大不同,同时对宿主音频音质的改变也不相同.

通过比较这些音乐的频谱分布可以发现,

escape coding 的发生频率与该音乐在频率上的分布密切相关,如图 1 所示. 经典音乐的水印嵌入量大,它的频率分布主要集中在低频区;演唱的水印嵌入量较小,它的频率分布则比较均匀. 也就是说,能量主要集中在低频区时的音乐发生 escape coding 的频率要比能量在各个频率分布均匀的音乐高得多.

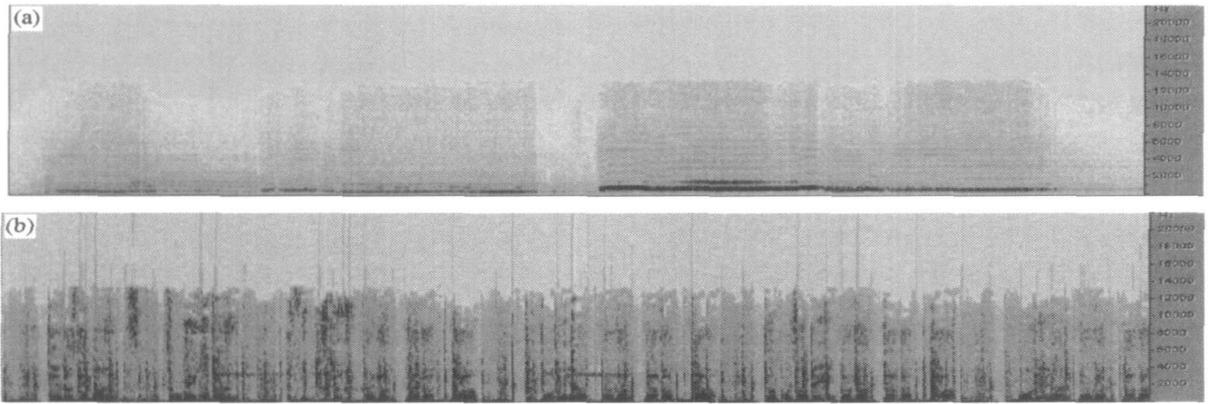


图 1 不同嵌入码率的频率分布. (a) 经典音乐的频率分布; (b) 演唱的频率分布

Fig. 1 Frequency distributions at different embedded rates: (a) classic frequency distribution; (b) singing frequency distribution

由于人耳对于低频声音敏感,为了进一步提高嵌入水印后的听觉质量,应该减少水印在低频段的嵌入量. 在 MPEG-2 AAC 中, MDCT 系数根据心理声学模型-II 从低频到高频被划分成了多个比例因子频带,因此可由 escape coding 发生所处的比例因子带号来判断当前的频率段,从而控制各频段上的水印嵌入量. 但通过实验发现,由于大部分的 escape coding 发生在低频区域,如果一味的控制低频区的水印嵌入,提高了听觉质量,却使水印的嵌入码率变得很小.

为了平衡水印的嵌入码率和不可感知性,本文利用 AAC 中根据心理声学模型计算出的感知熵 PE (perceptual entropy)^[11] 作为另一嵌入条件. PE 表示信号对噪声的容忍度. 在 AAC 中,当 PE 值大于 1800 时表明该段音乐变化较大,为避免失真 AAC 给它分配较多的比特数,使用短窗(256 个样本)来进行编码;当 PE 小于 1800 时,为提高编码效率使用长窗(2048 个样本)给信号编码. 根据刘伟等^[12] 的研究结果表明,MP3 压缩后的信噪比与 PE 值的变化趋势一致,即对 PE 较大的帧 MP3 所产生的影响较小,因此在短窗信号中嵌入水印引起的失真较小,这一点同样适用于 AAC.

基于 escape coding 发生的频段和 PE 的大小,本文对水印嵌入算法进行了改进:对 PE 小于 1800 的短窗信号不限制水印嵌入量,而对长窗信号则由比例因子带号决定水印嵌入量(表 5).

表 5 长窗信号水印嵌入量控制

Table 5 Watermark embedding control by long window signal			
比例因子带号, sb	sb ≤ 5	5 < sb < 20	20 < sb
嵌入量/bit	0	10	无限制

经过码率控制后的水印嵌入结果如表 6 所示. 所有嵌入水印的音乐在听觉质量上都有了提高,同时不同音频水印嵌入码率的差异减小,说明嵌入码率更稳定.

3 水印提取

水印提取即水印嵌入的逆过程. 在 Huffman 解码之前,找到所有使用码书号为 11 的 Huffman 码字,并判断该编码是否使用了 escape coding; 根据嵌入算法可以得到嵌入的水印值,再将提取的水印通过 Anorld 反变换^[9],恢复成原来的图像.

4 实验结果

以 folk 文件为例,水印图像(大小 70 × 70)如图 2(a)所示,实际嵌入的水印量是 6 495 bit (图 2(b)),在音频文件没有受到攻击时,水印可以 100% 被提取出来(图 2(c)). 由于嵌入和提取算法复杂度小,嵌入和提取的速度很快,几乎不会增加音频文件压缩和解压缩的时间. 虽然这种水印嵌入量较大,但由于只是对 Huffman 的 LSB 进行修改,原压缩音频的比特率不会增加.

表 6 改进后的水印嵌入结果

Table 6 Results of improved watermark embedding

音乐名称 (48 000 Hz, 133 kbps)	文件长度/ s	嵌入水印/ bit	每秒嵌入水印/ (bit · s ⁻¹)	感知音频质量评价: 客观失真等级, PEAQ:ODG
流行音乐	19.63	9 054	461.2	-0.270
爵士乐	20.73	12 548	605.3	-0.325
经典音乐	19.71	21 245	1 077.9	-0.732
民歌	20.28	6 495	320.3	-0.690
鼓乐	20.40	13 210	647.5	-0.132
摇滚乐	20.06	4 900	244.3	-0.217
演唱	20.10	5 129	255.2	-0.214
乡村音乐	20.20	12 680	627.8	-0.678



图 2 水印图像比较. (a) 水印图像; (b) 实际嵌入水印; (c) 实际提取水印

Fig. 2 Comparison of watermark images: (a) watermark image; (b) actually embedded watermark; (c) actually retrieved watermark

5 结论

本文通过修改 MPEG-2 AAC 帧数据中的 Huffman 码字设计了一种基于 MPEG-2 AAC 的隐写算法. 该算法提取水印时不需要原始音频参与, 嵌入水印后的音频在听觉上与原始音频无任何差异, 并且计算简单, 嵌入速度快, 嵌入量大, 嵌入和提取可以与压缩和解压缩同步完成, 也可在码流上直接嵌入与提取, 具有较强的实用性. 由于该水印是一种脆弱水印, 对各种信号攻击都很敏感, 因此可以用在数字产品的完整性认证中. 该算法还存在一些待改进的地方, 如可以在水印嵌入和提取时增加同步机制来抵抗同步攻击, 这将是今后的研究方向.

参 考 文 献

- [1] Fabien Petitcolas: *mp3stego* [EB/OL]. [2006-05-01]. <http://www.cl.cam.ac.uk/~fapp2/steganography/mp3stego>
- [2] Wang C T, Chen T S, Chao W H. A new audio watermarking based on modified discrete cosine transform of MPEG/Audio layer III // *Proceedings of the 2004 IEEE International Conference on Networking, Sensing & Control*. Taipei, 2004: 984
- [3] Moghadam N, Sadeghi H. Genetic content-based MP3 audio watermarking in MDCT domain [J/OL]. *Proc World Acad Sci Eng Technol*, 2005, 7: 248 [2006-06-08]. <http://www.waset.org/pwaset/v7/v7-69.pdf>
- [4] Koukopoulos D, Stamatou Y. A watermarking scheme for MP3 audio files [J/OL]. *Int J Signal Process*, 2006, 2/3: 206 [2006-05-01]. <http://www.waset.org/ijsp/v2/v2-3-30.pdf>
- [5] Koukopoulos D, Stamatou Y. An efficient watermarking method for MP3 Audio files [J/OL]. *Proc World Acad Sci Eng Technol*, 2005, 7: 154 [2006-05-01]. <http://www.waset.org/pwaset/v7/v7-29.pdf>
- [6] Torrubia A, Mora F. Perceptual cryptography on MPEG layer III bit streams. *IEEE Trans Consumer Electron*, 2002, 48(4): 1046
- [7] Kim D H, Yang S J, Chung J H. Additive data insertion into MP3 bitstream using linbits characteristics // *Proceeding on the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Montreal, 2004
- [8] ISO/IEC 13818-7. *Information Technology-Generic Coding of Moving Picture and Associated Audio, Part 7: Advanced Audio Coding*. 1997
- [9] Tian Y K, Jia C Y, Wang Q W. Image scrambling and restoring algorithm based on Arnold transform. *J Dalian Maritime Univ*, 2006, 32(4): 107
(田云凯, 贾传炎, 王庆武. 基于 Arnold 变换的图像置乱及其恢复. 大连海事大学学报, 2006, 32(4): 107)
- [10] ITU-R Recommendation BS 1387-1, *Method for Objective Measurements of Perceived Audio Quality*. Geneva: International Telecommunications Union, 2001
- [11] Jonhston J D. Transform coding of audio signal using perceptual noise criteria. *IEEE J Selected Areas Commun*, 1988, 6: 314
- [12] Liu W, Wang S Z, Zhang X P. Frequency domain audio watermark embedding capable of resisting MP3 encoding. *J Appl Sci*, 2005, 23(4): 341
(刘伟, 王朔中, 张新鹏. 抗 MP3 编码的音频水印频域嵌入方案. 应用科学学报, 2005, 23(4): 341)