

基于高斯混合模型的 DCT 域水印检测方法

林晓丹¹

摘要 基于音频 DCT 系数的统计特征, 提出了一种水印检测方法. 采用扩频的方法在 DCT 域嵌入水印, 分别对嵌入水印和未包含水印的音频信号在 DCT 域进行统计学习, 得到对应的高斯混合模型 (Gaussian mixture model, GMM). 接收端采用最大似然检测, 判断是否嵌入水印并提取相应的水印信息. 仿真结果表明本文的水印检测算法对常见的信号攻击具有鲁棒性, 与传统的相关检测法相比, 检测可靠性更高.

关键词 水印检测, DCT 域, 高斯混合模型, 最大似然

引用格式 林晓丹. 基于高斯混合模型的 DCT 域水印检测方法. 自动化学报, 2012, 38(9): 1445–1448

DOI 10.3724/SP.J.1004.2012.01445

DCT-domain Watermark Detection Using Gaussian Mixture Model

LIN Xiao-Dan¹

Abstract An audio watermark detector is developed based on the statistical characteristics of audio in DCT domain. Watermarks are embedded in DCT domain on a spread spectrum basis. A Gaussian mixture model (GMM) is introduced to model both the watermarked and unwatermarked audio signals in DCT domain. At the receiver a maximum likelihood detector is employed. Simulation results indicate that the proposed method allows for common signal processing and provides a better detecting accuracy compared with the linear correlation detector.

Key words Watermark detection, DCT domain, Gaussian mixture model (GMM), maximum likelihood

Citation Lin Xiao-Dan. DCT-domain watermark detection using Gaussian mixture model. *Acta Automatica Sinica*, 2012, 38(9): 1445–1448

网络通信技术的飞速发展推动了传统多媒体的数字化进程. 用户可以方便快捷地借助于网络获取数字多媒体资源, 甚至对其进行复制、修改和非法传播. 因此数字媒体的版权问题也成为信息化的瓶颈, 版权保护日益引起社会的关注. 数字水印技术将水印与原始作品紧密结合, 普通攻击很难将其去除, 因而已成为版权保护的主要手段. 用于版权保护的数字水印技术要求嵌入的版权保护信息具有不可察觉性, 且能够经受各种常规的信号处理. 扩频水印^[1]是目前广泛使用的鲁棒水印技术, 将水印信息经过扩频调制后叠加在原始数据上, 具有抗干扰能力强、保密性高的特点. 检测水印时计算含水印信号与伪随机码的线性相关, 与判决阈值进行比较后提取出水印序列. 然而, 线性相关检测法只有在宿主信号具有高斯分布的情况下才是最优的. 且该方法是在不考虑信道噪声, 以及宿主信号和水印信号的相关性的情况下得到的. 如果遭受了信号处理或恶

意攻击, 检测器的性能还将进一步变差. 文献 [2] 采用排除离群点的方法来降低检测错误, 但如何确定离群点比较困难. 文献 [3–4] 提出的基于统计特征的音频水印算法, 并没有对音频统计特性进行建模, 所以算法性能随载体音频的不同差异很大. 因此, 较有效的做法是利用信号的统计特性, 采用局部优化的检测方法. 文献 [5–6] 分别基于图像 DCT 域和小波域的广义高斯模型, 提出了水印的局部优化检测算法. 然而广义高斯模型不具有局部自适应性, 无法准确拟合图像变换域的统计特征, 因而文献 [7–8] 提出了基于高斯混合模型的图像水印检测方法, 但这两种方法只有在原始水印已知的前提下才是可行的. 文献 [9] 在量化索引调制的图像水印基础上采用最大似然检测, 发现较其他检测方法可获得更好的性能. 文献 [10] 基于回归型支持向量机和图像矩特征, 提出了一种强鲁棒性的图像水印检测算法. 上述方法都是针对图像的水印检测器, 目前对音频水印检测器的研究则匮乏得多. 文献 [11] 基于音频直方图统计特征, 提出了水印的嵌入和检测方法. 文献 [12–13] 则特别针对扩频水印, 提出了一种以音频作为载体的水印检测方法, 很大程度上降低了水印的误检率. 但这两种方法都需要设定阈值以确定水印的存在与否. 本文在 DCT 域以扩频方式嵌入水印, 为嵌入水印后的 DCT 交流系数概率分布建立高斯

收稿日期 2011-07-28 录用日期 2011-11-10
Manuscript received July 28, 2011; accepted November 10, 2011
泉州市科技计划项目 (2011G7) 资助
Supported by Quanzhou Municipal Science and Technology Program (2011G7)
本文责任编辑 刘一军
Recommended by Associate Editor LIU Yi-Jun
1. 华侨大学信息科学与工程学院 厦门 361021
1. College of Information Science and Engineering, Huaqiao University, Xiamen 361021

混合模型 (Gaussian mixture model, GMM). 检测时提取待测信号的 DCT 系数, 利用建立好的高斯混合模型, 根据最大似然原则判决水印存在与否并提取对应的水印信息.

1 水印嵌入

假设隐藏的消息是由 0 和 1 构成的比特串, 水印信号的产生类似于直接序列扩频的方法. 采用 m 序列作为扩频码, 对消息序列以码率 cr 进行扩展, 若嵌入消息比特 0, 则映射成伪随机序列 $PN_0 = \{p_0(n)|n = 1, 2, \dots, cr\}$, 否则映射成 $PN_1 = \{p_1(n)|n = 1, 2, \dots, cr\}$. $p_0(n), p_1(n) \in \{-1, 1\}$. 嵌入水印前, 先对音频作分帧处理, 把音频分成长度为 N 的短时帧. 对每个音频帧执行 DCT 变换, 水印的嵌入强度可以根据音频帧短时能量自适应调整. 音频帧短时能量越大, 则嵌入的水印强度越大. 此外, 为了在水印的鲁棒性和不可察觉性之间取得较好的平衡, 选取 cr 个 DCT 中频系数作为水印的嵌入位置. 对修改后的音频帧执行 DCT 逆变换, 得到含水印的音频. 修改后的系数 $d_m(n)$ 如式 (1) 所示:

$$d_m(n) = d_m^0(n) + \alpha \cdot E_m \cdot p_m(n) \quad (1)$$

式中, d_m^0 为原 DCT 系数, E_m 为第 m 帧的归一化短时能量, $p_m(n)$ 为对应的扩频码第 n 个值, α 为常数值, 水印的嵌入强度由 α 和 E_m 共同确定. α 越大, 鲁棒性越强.

2 水印检测

水印的检测可以简化为一个分类问题. 这里需要把音频帧分成三类, 即没有嵌入水印、嵌入的原始消息比特分别为 1 和 0. 这时可以借助于训练好的高斯混合模型, 利用概率大小作出判决.

2.1 DCT 域高斯混合建模

音频信号的数据量要远大于图像信号, 且具有非平稳特性. 而广义高斯模型是一种全局概率模型, 不具有局部自适应性. 由于音频 DCT 域交流系数大部分都集中在 0 附近, 便于用有限阶的高斯混合模型对其建模. 因此文中引入高斯混合模型以便更精确地描述音频 DCT 域的统计特征, 概率模型中的参数由 DCT 域系数分布确定. 图 1 给出了采用三阶高斯混合模型对 DCT 交流系数的拟合结果.

对于嵌入 1、嵌入 0 和未嵌入水印的音频帧, 分别得到三个不同参数的高斯混合模型.

$$p_e(X_l|\lambda_e) = \sum_{i=1}^M \alpha_{e,i} f_{e,i}(X_l; \mu_{e,i}, \sigma_{e,i}^2) \quad (2)$$

其中, $e = 0, 1, 2$. $e = 2$ 对应没有嵌入水印的情况, X_l 为可观测的样本. 为了得到这三个高斯混合模型, 需要确定高斯混合模型的阶数 M , 参数集 $\lambda_e = \{\alpha_{e,i}, \mu_{e,i}, \sigma_{e,i}^2\}$, 其中 $i = 1, 2, \dots, M$, $\alpha_{e,i}$ 为各维高斯分量的权重, 满足 $\sum_{i=1}^M \alpha_{e,i} = 1$, $\mu_{e,i}$ 和 $\sigma_{e,i}^2$ 分别为均值和方差. 为了使处理简单, 令上述三种情况下的概率密度函数形式相同, 所不同的仅是其参数集.

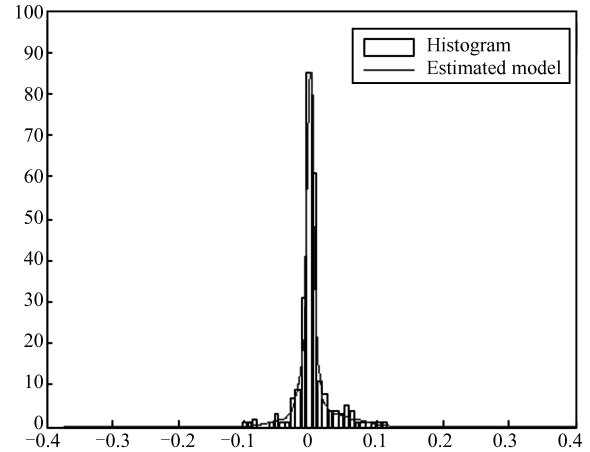


图 1 DCT 交流系数的高斯混合模型拟合

Fig. 1 AC coefficients in DCT domain fitted by GMM

2.2 高斯混合模型参数估计

期望最大化 (Expectation-maximization, EM) 算法^[4] 适用于在数据不完整的情况下最优地估计概率模型中的隐含参数, 大大降低了最大似然估计的计算复杂度, 性能却与最大似然估计相近, 常被用于估计高斯混合模型参数. 定义完备数据集 $Z = (X, Y)$, X 为可观察空间, 由实际样本构成, 即嵌入水印的 DCT 系数; Y 为不可观察空间, 对应于某个高斯模型中的点. 定义对数似然函数 $L(\lambda|Z) = \log p(X, Y|\lambda)$. EM 算法是一种迭代算法, 每次迭代都包括以下两个步骤:

1) E-step: 在已知观测数据 X 及当前参数 $\lambda^{(k-1)}$ 的条件下, 获得似然函数关于未知数据的期望值估计.

$$Q(\lambda, \lambda^{(k-1)}) = E_Y [\log p(X, Y|\lambda)|X, \lambda^{(k-1)}] \quad (3)$$

其中, $\lambda^{(k-1)}$ 是当前参数估计, λ 是新的参数, 用于下一轮迭代. $Q(\lambda, \lambda^{(k-1)})$ 是完备数据集 Z 的对数似然函数关于未知数据 Y 的期望值.

2) M-step: 最大化式 (3) 中计算的期望值.

$$Q^{(k)} = \arg \max_{\lambda} Q(\lambda, \lambda^{(k-1)}) \quad (4)$$

以上两个步骤不断重复, 直到 $Q(\lambda, \lambda^{(k-1)})$ 的当前值相对上一次迭代时的增幅小于设定的阈值时结束迭代. 按照拉格朗日法求带约束条件的极值问题, 通

过迭代为上述三种情况分别估计得到一组 GMM 参数如式 (5)~(7) 所示^[14].

$$\alpha^k = \frac{1}{T} \sum_{t=1}^T p^{(k-1)}(i|\mathbf{x}_t, \lambda) \quad (5)$$

$$\mu_i^k = \frac{\sum_{t=1}^T p^{(k-1)}(i|\mathbf{x}_t, \lambda) \mathbf{x}_t}{\sum_{t=1}^T p^{(k-1)}(i|\mathbf{x}_t, \lambda)} \quad (6)$$

$$\sigma_{i,k}^2 = \frac{\sum_{t=1}^T p^{(k-1)}(i|\mathbf{x}_t, \lambda) (\mathbf{x}_t - \mu_i^k)^2}{\sum_{t=1}^T p^{(k-1)}(i|\mathbf{x}_t, \lambda)} \quad (7)$$

其中, $p(i|\mathbf{x}_t, \lambda)$ 表示特征向量 \mathbf{x}_t 落入隐状态 i (i 为 M 个状态之一) 的概率, k 为迭代次数, T 为可观空间 X 的样本个数. 训练集中的音频为包含水印的音频和未添加水印的音频. 在 DCT 域确定水印嵌入位置, 即 DCT 系数 $D_m = \{d_m(n), d_m(n+1), \dots, d_m(n+cr-1)\}$, m 为帧号. 选取嵌入 1、嵌入 0 和未嵌入水印的三类音频帧对应的 D_m 作为样本序列分别训练三类高斯混合模型, 得到其权重、均值以及协方差参数. 在 EM 迭代前, 需要确定高斯混合模型的阶数以及初始参数 $\lambda_{e,0} = \{\alpha_{e,0}, \mu_{e,0}, \sigma_{e,0}^2\}$. 这里采用 K-均值聚类算法得到初始参数, M 由实验确定. M 越大, 拟合效果越好, 然而计算复杂度也越大.

2.3 水印检测

利用密钥确定水印嵌入位置, 提取待测信号音频帧的 DCT 系数 D_m (m 为帧号) 作为式 (2) 的输入 X_l , 计算出对应概率 $p_0(D_m)$, $p_1(D_m)$ 和 $p_2(D_m)$. 检测准则由式 (8) 给出, e_m 即从第 m 帧提取出的水印比特.

$$e_m = \arg \max_{e \in \{0,1,2\}} p_e(D_m|\lambda_e) \quad (8)$$

3 仿真实验及分析

实验中采用采样率 44.1 kHz, 16 比特量化, 长度约 20 s 的单声道音频用于训练高斯混合模型参数. 选取其中的 1000 帧嵌入消息比特 0、1000 帧嵌入比特 1, 分别用于 GMM 模型 ($e = 0$ 和 $e = 1$) 的训练. 此外, 选取未嵌入水印的 1000 帧用于 GMM 模型 ($e = 2$) 的训练. 实验中 $cr = 15$, 音频帧长为 256. 图 2 给出了混合阶数 M 和似然函数值的关系曲线. 可见 $M \geq 6$ 时, 拟合效果较好. 考虑到计算代

价问题, 本文选取 6 作为高斯混合阶数. 选取十段单声道音频添加水印后作为测试集 (每段音频的最后 100 帧不嵌入水印), 每帧嵌入 1 比特消息, 采样率 44.1 kHz, 16 比特量化, 总长度约 150 秒. 定义误码率 $p_e = L_e/L$, 其中 L_e 为从所有音频中提取的水印错误比特总数, 误检的情况 ($e = 2$ 误判为其他两种情况或其他两种情况被误判为 $e = 2$) 也视为比特错误. L 是所有测试音频段中嵌入消息序列的总长度. 对测试音频执行几种常见攻击: 1) 低通滤波: 5 阶巴特沃思低通滤波器, 截止频率为 11.025 kHz; 2) 下采样到 22.05 kHz 再还原成 44.1 kHz; 3) 重量化: 从 16 比特量化成 8 比特再量化成 16 比特; 4) 码率为 64 kbps 的 MP3 压缩. 表 1 给出了在不同的水印嵌入强度下, 本文方法对于音频经受上述常规攻击后的水印检测结果. 由表 1 结果可知, 本文方法可以在尽量保持水印不可感知性的前提下改善其抵抗攻击的性能.

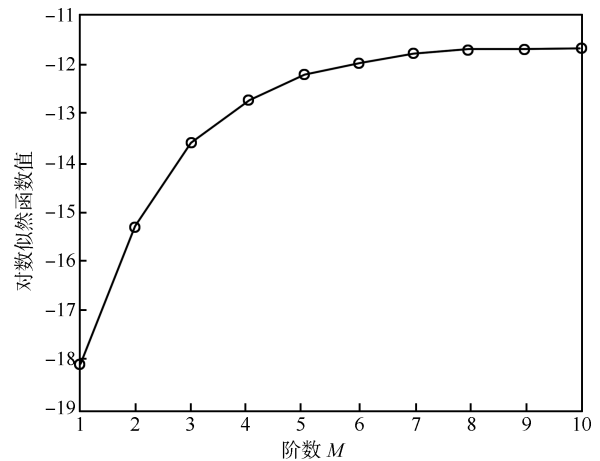


图 2 高斯混合阶数与对数似然值的关系

Fig. 2 Effect of the number of GMM components on log likelihood

表 1 抵抗常规信号处理的性能 ($p_e(\%)$)

Table 1 Performance against common signal processing

α	低通滤波	重采样	重量化	MP3 压缩
0.01	0.36	1.24	3.23	5.62
0.02	0.34	1.24	3.2	5.48
0.03	0.34	1.21	3.14	5.4

此外, 为了测试本文算法对其他攻击的鲁棒性, 实验中还采用 Stirmark 音频水印测试软件^[15] 对音频信号实施如表 2 所示的各种攻击, 分别采用本文方法与相关法对水印进行检测并给出相应的误检率. 实验结果表明, 对于大多数攻击, 本文算法比相关检测法在检测性能上有了很大的提升.

表 2 对恶意攻击的鲁棒性比较 ($p_e(\%)$)

Table 2 Robustness comparison against malicious attacks

Attack	本文方法	相关检测	Attack	本文方法	相关检测
Addbrumm_10100	2.4	33.5	FFT_hlpass	3.9	28.7
Addnoise_900	10.5	29.6	FFT_invert	48.3	41.9
Compressor	1.9	26.8	Flippsample	15.3	47.2
Cutsamples	14.9	41.8	Lsbzero	0.2	22.8
Dynnoise	5.6	27.1	Normalize	0	14.5
Echo	4.2	33.4	Stat1	1.5	22.6
Exchange	0.4	26.0	Zeroremove	2.8	24.3

4 结论

GMM 模型是对非高斯数据进行建模的最有效的方法之一, 本文研究了基于 GMM 模型的音频水印盲检测器. 对 DCT 域交流系数统计分布的拟合, 得到不同的 GMM 模型. 检测时根据最大似然原则检测并提取水印. 实验结果表明本文的方法在提高检测可靠性方面取得了较好的效果, 且能够在尽量减少音频失真的同时保证水印的鲁棒性. 特别是在音频信号遭受常规信号攻击和恶意攻击后, 相对相关检测器, 在误检率方面本文的算法取得了显著改善.

References

- Bassia P, Pitas I, Nikolaidis N. Robust audio watermarking in the time domain. *IEEE Transactions on Multimedia*, 2001, **3**(2): 232–241
- Kim H J, Kim T, Yeo I K. A robust audio watermarking scheme. In: Proceedings of the 2004 International Symposium on Circuits and Systems. Vancouver, Canada: IEEE, 2004. 696–699
- Wang Xiang-Yang, Niu Pan-Pan. A content-based digital audio watermarking scheme. *Acta Automatica Sinica*, 2008, **34**(8): 1001–1003
(王向阳, 牛盼盼. 基于音频统计特性的数字水印嵌入算法. 自动化学报, 2008, **34**(8): 1001–1003)
- Zhang L, Chen L M, Qian G B. Self-synchronization adaptive blind audio watermarking. In: Proceedings of the 12th International Multi-Media Modelling Conference. Beijing, China: IEEE, 2006. 381–384
- Sun Zhong-Wei, Zhu Yan, Feng Deng-Guo. Performance analysis of DCT-domain watermark detection based on local optimum detection. *Acta Electronica Sinica*, 2005, **33**(5): 864–867
(孙中伟, 朱岩, 冯登国. DCT 域图像水印的局部优化检测性能研究. 电子学报, 2005, **33**(5): 864–867)
- Fan Jun-Feng, Zhao You-Jun, Di Lan-Zhen. An adaptive watermarking scheme based on general Gauss distribution model of wavelet coefficients. *Microelectronics and Computer*, 2003, **23**(5): 160–162
(范俊峰, 赵友军, 邸兰振. 基于广义高斯分布模型的小波域自适应盲水印方案. 微电子学与计算机, 2003, **23**(5): 160–162)
- Duan F B, Wang Y. A nonlinear detector for optimal DCT-domain watermark detection. In: Proceedings of the 2010 International Conference on Electrical and Control Engineering. Wuhan, China: IEEE, 2010. 176–178
- Sun Z W, Ma J. DWT-domain watermark detection using Gaussian mixture model with automated model selection. In: Proceedings of the 2009 International Symposium on Computer Network and Multimedia Technology. Wuhan, China: IEEE, 2009. 1–4
- Lu W M, Li W Q, Safavi-Naini R, Ogunbona P. Optimal image watermark decoding. *Lecture Notes in Computer Science*. New York: Springer, 2006. 141–149
- Xu Zi-Han, Wang Xiang-Yang. An SVR based image watermarking detection algorithm against geometric attacks. *Acta Automatica Sinica*, 2009, **35**(1): 23–27
(徐紫涵, 王向阳. 可有效抵抗一般性几何攻击的数字水印检测方法. 自动化学报, 2009, **35**(1): 23–27)
- Zhang X M. Segmenting histogram-based robust audio watermarking approach. *Journal of Software*, 2008, **3**(9): 3–11
- Malik H, Khokhar A, Ansari R. Improved watermark detection for spread-spectrum based watermarking using independent component analysis. In: Proceedings of the 5th ACM Workshop on Digital Rights Management. Alexandria, USA: ACM, 2005. 102–111
- Li L, Fang X. New detection scheme for spread spectrum audio watermarking. In: Proceedings of the 2010 IEEE International Conference on Wireless Communications, Networking and Information Security. Beijing, China: IEEE, 2010. 330–334
- Dempster A P, Land N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 1977, **39**(1): 1–38
- Steinebach M, Petitcolas F A P, Raynal F, Dittmann J, Fontaine C, Seibel S, Fontaine C, Seibel S, Fates N, Ferri L C. StirMark benchmark: audio watermarking attacks. In: Proceedings of the 2001 International Conference on Information Technology: Coding and Computing. Las Vegas, USA: IEEE, 2001. 49–54



林晓丹 华侨大学信息科学与工程学院讲师. 2008 年获上海交通大学硕士学位. 主要研究方向为多媒体信号处理和数字水印. E-mail: xd_lin@hqu.edu.cn
(LIN Xiao-Dan Lecturer at the College of Information Science and Engineering, Huaqiao University. She received her master degree from Shanghai

Jiao Tong University in 2008. Her research interest covers signal processing and digital watermarking.)