ICSP2008 Proceedings

Audio Watermarking in DCT：Embedding Strategy and Algorithm

Gaorong Zeng Zhengding Qiu
Institute of Information Science, Beijing Jiaotong University, Beijing, P. R. China
Email: loch320@163.com

Abstract

The position of watermarking plays an important role when the tradeoff between robustness and inaudibility is considered. In the paper we analyze the mathematical relationship between the embeddable position and the noise of the audio watermarking in the DCT (discrete cosine transform) domain. We consider DC (direct current) coefficient to tradeoff robustness and inaudibility. And a blind watermarking scheme is implemented through Quantization index modulation (QIM) technology. Experimental results show watermarking is robust to common signal processing operation and mp3 compressing, and inaudibility is satisfaction.

1. Introduction

Internet application widely makes it easier for music pirates to copy and distribute music and songs around the world illegally. As a supplementary means, many digital watermarking techniques are intentionally designed for copyright protection of the media work (such as image, audio and video, etc) [1]. Compared with watermarking in the time domain, Watermarking in the transform domain becomes more feasible because of better robustness. Audio watermarking based on DCT is one of many common techniques.

Many strategies were proposed to implement the watermarking such as modifying low frequency, middle frequency, or high frequency DCT coefficient of an audio signal. These strategies have their own advantages and flaws. Some obtain more robustness, and others have more audibility. However, the tradeoff between robustness and inaudibility is seldom considered. Ma Yiping et al [2] have ever analyzed the noise sensitivity of DCT coefficients, but his noise sensitivity vector was constant according to calculation.

In this paper, the relation between the embeddable location and the noise caused by embedding watermark is discussed and a mathematic expression is given in section 2. An embedding strategy is proposed to implement the watermarking based on QIM technology in section 3. Experiments and simulations are realized in section 4. And a simple conclusion is drawn in section 5.

2. Noise sensitivity of DCT coefficients

As we know, it is very good for acoustic quality to embed message in high frequency coefficients, however the hiding message is so vulnerable that it often fails to be extracted after some simple attacks. To strengthen robustness, some researchers embed message in low frequency coefficients [3], but this can cause audibility to decline. The influence of modifying DCT coefficients is analyzed in this section.

2.1. DCT coefficients noise model and related work

Watermark message is embedded into the host audio by modifying DCT coefficients, which can be regarded as adding a noise to the original audio signal.

DCT and IDCT expressions are as followed.

$$F(k)=c(k)\sum_{n=0}^{N-1} f(n)\cos[\frac{(2n+1)k\pi}{2N}] \quad (1)$$

$$f(n)=\sum_{k=0}^{N-1} c(k)F(k)\cos[\frac{(2n+1)k\pi}{2N}] \quad (2)$$

Where $f(n)$ are time audio series, and $F(k)$ are DCT coefficients series, N is the number of samples when DCT is done, $0 \le n,k < N$.

$c(k)$ is the coefficient, which is below:

This work was supported by a grant from Science and Technology Foundation of Beijing Jiaotong University(2005SM009), and a grant from the National High Technology Research and Development Program of China (2007AA01Z460)

978-1-4244-2179-4/08/$25.00 ©2008 IEEE

$$c(k)=\begin{cases}\frac{1}{\sqrt{N}}, & k=0\\ \sqrt{\frac{2}{N}}, & k\neq 0\end{cases} \qquad (3)$$

In order to embed watermark, we need to modify the DCT coefficients $F(k)$. While a signal $E(i)$ (denoting watermark information) is added to one of these coefficients, $F(i)$ (where $0\le i<N$), i.e. the i th coefficient became $F^{'}(i)$:

$$F^{'}(i)=F(i)+E(i) \qquad (4)$$

The corresponding time series are calculated by IDCT:

$$\begin{aligned} f^{'}(n) &= \sum_{k=0}^{N-1} c(k)F^{'}(k)\cos[\frac{(2n+1)k\pi}{2N}] \\ &= \sum_{k=0}^{N-1} c(k)F(k)\cos[\frac{(2n+1)k\pi}{2N}] + c(i)E(i)\cos[\frac{(2n+1)i\pi}{2N}] \\ &= f(n)+e(i,n) \end{aligned} \qquad (5)$$

Where $e(i,n)=c(i)E(i)\cos[\frac{(2n+1)i\pi}{2N}]$ (6)

$e(i,n)$ is the noise on the n th sample in time domain since the i th DCT coefficient $F(i)$ has been altered. The total noise energy is $\sum_{n=0}^{N-1} e(i,n)$.

Ma Yiping et al [1] have ever defined a noise sensitivity coefficient $\beta(i)$, which is follow:

$$\text{Let } t(i,n)=\begin{cases}\cos\frac{\pi}{4}, & i=0\\ \cos\frac{(2n+1)i\pi}{2N}, & i\neq 0\end{cases} \qquad (7)$$

$$\beta(i)=\frac{\sum_{n=0}^{N-1} t^2(i,n)}{N} \qquad (8)$$

According to (7), we can obtain:

$$i=0\,,\ \sum_{n=0}^{N-1} t^2(i,n)=\sum_{n=0}^{N-1}\cos^2\frac{\pi}{4}=\frac{N}{2} \qquad (9)$$

$i\neq 0$,

$$\begin{aligned} \sum_{n=0}^{N-1} t^2(i,n) &= \sum_{n=0}^{N-1}\cos^2\frac{(2n+1)i\pi}{2N}=\sum_{n=0}^{N-1}\frac{1}{2}(1+\cos\frac{(2n+1)i\pi}{N}) \\ &= \frac{N}{2}+\frac{1}{2}\sum_{n=0}^{N-1}(\cos\frac{(2n+1)i\pi}{N})=\frac{N}{2}+0=\frac{N}{2} \end{aligned} \qquad (10)$$

$$\text{So } \beta(i)=\frac{\sum_{n=0}^{N-1} t^2(i,n)}{N}=\frac{N/2}{N}=0.5 \qquad (11)$$

i.e. $\beta(i)$ is a constant number. Obviously, we can't select the valid location based on $\beta(i)$, because the contribution of every DCT coefficient position is equivalent.

2.2. DCT coefficients noise sensitivity

The distortion of audio signal is a considerable factor while embedding a watermark. Compared with original audio signal, the quality of watermarked audio signal will decline more or less. Usually difference distortion measures are used to evaluate the aural quality. Signal to Noise Ratio (SNR) and Peak Signal to Noise Ratio (PSNR) are two classic ways.

Assume that original audio signal has N samples, $f_o(n)$ and $f_w(n)$ denote the original audio signal and the watermarked audio signal, respectively. SNR and PSNR can be calculated as followed:

$$SNR=10\log_{10}\frac{\sum_{n=0}^{N-1} f_o^{\,2}(n)}{\sum_{n=0}^{N-1}(f_w(n)-f_O(n))^2} \qquad (12)$$

$$PSNR=10\log_{10}\frac{Nf_w^{\,2}(m)}{\sum_{n=0}^{N-1}(f_w(n)-f_O(n))^2} \qquad (13)$$

Where $f_w(m)$ denote the maximum amplitude of the watermarked audio signal.

According to (5), (6), (7), we can obtain:

$$SNR=10\log_{10}\frac{N\sum_{n=0}^{N-1} f_o^{\,2}(n)}{2E^2(i)\sum_{n=0}^{N-1} t^2(i,n)}=10\log_{10}\frac{\sum_{n=0}^{N-1} f_o^{\,2}(n)}{E^2(i)} \qquad (14)$$

$$PSNR=10\log_{10}\frac{N^2 f_w^{\,2}(m)}{2E^2(i)\sum_{n=0}^{N-1} t^2(i,n)}=10\log_{10}\frac{Nf_w^{\,2}(m)}{E^2(i)} \qquad (15)$$

We can see that the SNR and PSNR are identical when modification quantity $E(i)$ is fixed.

Additionally, the Human Auditory System (HAS) is insensitive to small amplitude changes, and some ‘quiet’ audio signal, i.e. smaller amplitude ones tend to be ignored in the presence of loud sounds [4, 5]. The larger amplitude of the audio signal is, the more distortion is tolerated. In the paper, we choose the DC coefficient as the embedding position because it is the most important component with the largest amplitude.

3. Embedding strategy and algorithm

We choose binary image (40*20) as watermark message, audio (.wav) as host audio signal. According to the analysis above, embedding strategy is following.

3.1. Preprocessing of original audio signal and watermark

The original audio signal is divided into many frames. Every frame includes L samples except the last frame. L is the length of a frame. Through the formula (16) of signal energy:

$$E = \sum_n f^2(n) \tag{16}$$

We can calculate the energy of every frame. If the energy of one frame is higher than the average energy of all the frames, the frame belongs to high energy section, otherwise, it will fall into low energy and can't be used to embed watermark.

In order to strengthen the security, the watermark image is permuted through performing Arnold transform then turned into one-dimension sequence. As follow:

$$V = \{v(k) = w(i,j), 0 \le i < M_1, 0 \le j < M_2, k = i \times M_2 + j\} \tag{17}$$

Where M_1 is the number of rows, M_2 is the number of columns, $w(i,j) \in \{0,1\}$.

3.2. The embedding phase

(1) The original audio is divided into two families of frames in term of energy, high energy section and low energy section. Refer to section 3.1.

(2) For low energy section, all frames keep invariable. For high energy section, every frame with L samples is decomposed into non-overlapping $L/8$ blocks and the DCT is performed independently for every block. We choose the DC coefficients $F(0)$ as the embedding positions.

(3) Embedding is based on QIM method [6,7]. Quantization process is in the following:

Quantizing the DC coefficient $F(0)$ to the closest A_k or B_k will finish the embedding of watermark bit '0' or '1'.

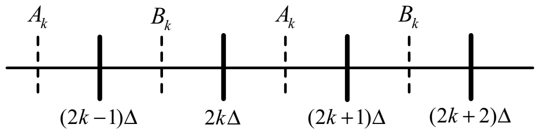


Figure1. Quantization DC coefficient $F(0)$

$$F^{'}(0) = \begin{cases} A_k, if\ w=1 & and \quad \arg\min\left|F(0)-A_k\right| \\ B_k, if\ w=0 & and \quad \arg\min\left|F(0)-B_k\right| \end{cases} \tag{18}$$

$$A_k = (2k+\frac{1}{2})\Delta,\ B_k = (2k-\frac{1}{2})\Delta,\ k = 0,\pm1,\pm2... \tag{19}$$

Where Δ is quantization step, w is a watermark bit, $F^{'}(0)$ is the quantization result of $F(0)$. The largest quantization error is $\left|F(0)-F^{'}(0)\right| \le \Delta$.

Replacing $F(0)$ with $F^{'}(0)$, Inverse DCT is applied to the modified DCT coefficients of each embedding block to rebuild the waveform in the time domain.

Repeat step (2), (3), until all watermark bits are embedded.

3.3. The extracting phase

The watermark extraction process is to obtain an estimative watermark from a possibly destroyed watermarked audio signal. The extracting phase can be performed without the original audio signal and described as follows.

(1) For the possibly destroyed watermarked audio signal, we firstly perform the same procedure as (1) and (2) of the embedding phase and extract the DC coefficient $\tilde{F}(0)$ of every block.

(2) Judge the embedded information $\tilde{v}(0)$ of each block by (20).

$$\tilde{v}(k) = \mathrm{mod}(ceil(\frac{\tilde{F}(0)}{\Delta}), 2) \tag{20}$$

(3) Repeat Steps (1) to (2) until all the watermark bits are extracted.

(4) Revert the one-dimension bit series $\tilde{v}(k)$ ($0 \le k < M_1 \times M_2$) to the two-dimension image information, and get the estimative watermark image through the Arnold back-transforming process.

After the extracting phase is completed, the watermark is extracted. The similarity measure between the given watermarks and the extracted ones, W and $\tilde{W}$, is computed according to (21).

$$\rho(W,\tilde{W}) = \frac{\sum_{i=0}^{M_1-1}\sum_{j=0}^{M_2-1} w(i,j)\tilde{w}(i,j)}{\sqrt{\sum_{i=0}^{M_1-1}\sum_{j=0}^{M_2-1} w^2(i,j)}\sqrt{\sum_{i=0}^{M_1-1}\sum_{j=0}^{M_2-1} \tilde{w}^2(i,j)}} \tag{21}$$

4. Experimental results and discussions

To evaluate the performance of the proposed method, a mono audio signal (sampling frequency of 44100 Hz and 16 bits precision, 2 seconds length) is used as test host audio. The watermark is a 40*20 binary image as shown in Figure 3(a). To tradeoff robustness and inaudibility we choose $\Delta = 0.03$. In contrast, we also choose another two positions, 3rd AC coefficient, 4th AC coefficient to test. The test results are also showed in Table 1, Table 2 and Figure 3. As shown in Table 1, the 'SNR' of watermarked audio in three positions are very close and are more than 20db of SDMI (Secure Digital Music Initiative), and human's hearing is satisfaction. Figure 2 is a contrast between original audio signal and watermarked audio signal is shown in Figure 2 when embedding works in DC coefficient.

Table 1. The SNR of watermarked audio

	DC	3th AC	4th AC
SNR(db)	27.52	28.09	28.69

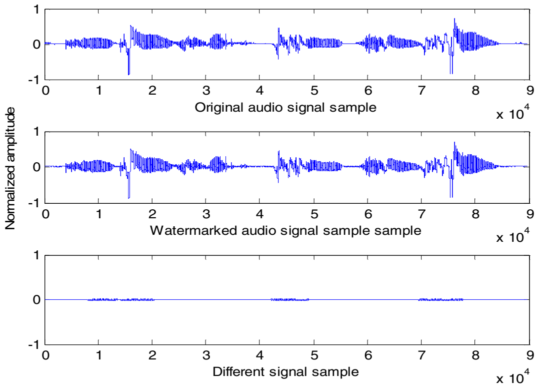
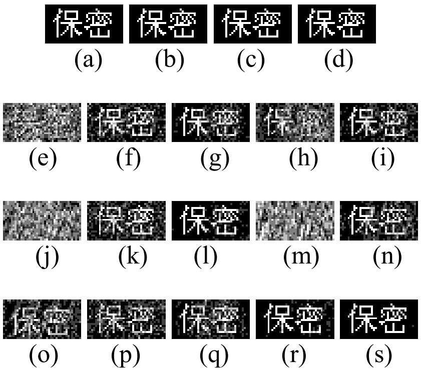


Figure 2. A contrast between original audio signal and watermarked audio signal (DC coefficient as embedding position).

Figure 3 shows extraction watermark in the position DC, 3rd AC, 4th AC respectively under various attacks.



(a) (b) (c) (d)

(e) (f) (g) (h) (i)

(j) (k) (l) (m) (n)

(o) (p) (q) (r) (s)

Fig.3 Extraction watermark under various attacks

Where (a) denotes the original watermark 'image', (b), (c), (d) denote the extraction watermarks without attack, in three positions, DC, 3rd AC, and 4th AC coefficient, respectively. (e), (f), (g), (h), (i) denote the extraction watermarks under low pass filter (Chebyshev(I) with Cut-off frequency 4000Hz, order 5), adding gauss noise($E=0,\sigma=0.01$), MP3 compression (compression ratio is about 10:1), smoothing(span 5 and 3), respectively in the 4th AC coefficient. (j), (k), (l), (m), (n) denote the extraction watermarks under all the attacks above in the 3rd AC coefficient, respectively. (o), (p), (q), (r), (s) denote the extraction watermarks under all the attacks above in DC coefficient, respectively.

Normalization of similarity coefficients is showed in Table 2.

Table 2. The Assessments of several attacks

	DC	3rd AC	4th AC
Cheby1	0.3812	0.1989	0.1858
Gauss	0.8606	0.8830	0.8669
Mp3	0.9046	0.9202	0.9326
Smooth(5)	0.9813	0.6850	0.2608
Smooth(3)	0.9997	0.8970	0.8903
No attack	1	1	1

As shown in Figure 3 and Table 2, better robustness can be obtained in the DC coefficient position. The extraction watermark is blurry under low pass filter, but the audibility is undesirable.

5. Conclusions

Through analyzing the mathematical relationship between the embeddable position and the noise of the audio watermarking signal in the DCT domain, we find that the SNR and PSNR are identical when modification quantity $E(i)$ is fixed. Consequently, we choose the DC coefficient with the largest amplitude as the embedding position. Experimental results show watermarking is robust to common signal processing operation and mp3 compressing, and audibility is satisfaction.

References

[1] Pierre Moulin, Ralf Koetter, Data-hiding codes [J], Proceedings of the IEEE，2005，93(12)2083-2126.

[2] Ma Yiping, Han Jiqin, "Audio Watermark In DCT Domain: Strategy And Algorithm" [J], Chinese Journal Electronics, 2006;34(7): 1260-1264.

[3] I. J. Cox， J. Kilian， T. Leighton， T Shamoon . "Secure spread spectrum watermarking for multimedia" [J], IEEE Transaction on Image Processing，1997,6(12) ：1673-1687.

[4] T. Painter, A. Spanias, "Perceptual coding of digital audio", Proc. IEEE 88 (4) (2000) 451-513.

[5] D. Pan, "A tutorial on mpeg/audio compression", IEEE Multimedia 1995,2 (2) : 60-74.

[6] Chen B, Wornell G W. Quantization index modulation methods: a class of provably good methods for digital watermarking and information embedding, IEEE Trans. Inf. Theory, vol. 47, no. 4, pp. 1423-1443, May 2001.

[7] Liu Tong, QIU Zhengding, A quantization-based image watermarking algorithm[J], Journal of China institute of Communications, 2002,23(10):89-93.