Wavelet-Based Audio Watermarking Using Adaptive Tabu Search

N. Sriyingyong and K. Attakitmongcol

Signal and Image Processing Research Group
School of Electrical Engineering, Institute of Engineering,
Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand
E-mail: kitti@sut.ac.th

Abstract

In this paper, we present an algorithm for robust audio watermarking in wavelet transform domain. Using Daubechies wavelet decomposition, we perform watermark embedding to wavelet coefficients of host audio signal. In enhance security, a pseudo-random permutation is performed to disperse the spatial relationship of the binary watermark image. In our watermark embedding algorithm, we search for the optimal intensity of watermark by using the adaptive tabu search (ATS). The watermark detection process can be performed without using the original audio signal. The experimental results demonstrate that watermark is inaudible and robust to many digital signal processing, such as resamplig, cropping, low pass filter, additive noise and lossy compression (MP3).

Keywords: Discrete wavelet transform, audio watermarking, Adaptive tabu search

1. Introduction

Over the past decade, internet, PC and other multimedia devices have become very popular. People can download digital multimedia from personal computer, PDA and mobile phone. Thus, digital data can be easily accessed, stored and duplicated. Consequently, the importance of copyright protection becomes very crucial. As a solution to this problem, various digital watermarking techniques have been investigated to address the issue of ownership verification. Watermark can be embedded into various types of media such as audio, still image and image sequence. For audio signal, the goal of audio watermarking is to hide appropriate data into the original audio signal. Ideally, there must be no perceptible difference between the watermarked and original audio signal, and the watermark should be easily extractable, reliable and robust against compression or any signal manipulations. According to the International Federation of the Phonographic Industry (IFPI) [1], audio watermarking should have the following specifications: 1) Audio watermarking should not degrade perception of original signal. 2) SNR should be greater than 20 dB and there should be more than 20 bps data payload for watermark. 3) Watermark should be able to resist most common audio processing operations and attacks. 4) Watermark should be able to prevent unauthorized detection, removal and embedding.

In general, we can classify digital watermarking techniques into two classes depending on the domain of watermark to be embedded: the spatial domain watermarking and the transform domain watermarking. Currently, watermarking techniques based on transform domain are more popular than those based on spatial domain since they provide higher signal quality and much more robust watermark [2, 3]. In [4], Cui *et al.* use cepstrum technique to analyze audio signal. A binary watermark image is embedded in cepstrum domain by using audio masking as a controlling factor. Audio masking has been calculated from minimum masking threshold between power spectrum of audio signal and absolute threshold of hearing in psychoacoustic model. However, the results show that the watermark is not robust against low-pass filtering. X. He *et al.* [5] present spread spectrum technique based on direct spread-spectrum sequence (DSSS) to spread watermark over the highest entropy areas of psychoacoustic model. The watermark is inserted in some special areas of host audio signal to decrease time consuming caused by forward and inverse Fourier transforms. The data rate from spread spectrum technique is lower than other existing methods. In [6], Li *et al.* decompose audio signal into wavelet coefficients and use SNR to determine watermarking embedding intensity. This technique provides watermark which is robust to many attacks but it requires the original audio signal in the detection process. Cui *et al.* [7] use audio compression technology to locate the watermark embedding location in wavelet transform domain by modifying the detail coefficients which are smaller than a selected threshold. The result does not yield watermark which is robust against low-pass filter. The other weak point of this scheme is that this technique requires the original audio signal in the detection process. Tu *et al.* [8] select least significant bit of wavelet coefficients for embedding semi-fragile watermark by quantizing technique. The user-defined quantizing parameter is well designed for watermark to resist against only some attacks.

The previous works on audio watermarking have shown the importance of a parameter called the embedding intensity. If this parameter is high, the watermark is robust but the quality of the watermarked audio signal is poor and vice versa. Previously, this parameter is obtained from the experiments. In this paper, we perform the watermark embedding in the wavelet transform domain. The intensity of watermark is searched by the artificial intelligence technique called the adaptive tabu search. We use the SNR value of the watermarked audio signal and the similarity value of the detected watermark as the objective function. SNR and similarity values represent the quality of the watermarked signal and robustness of the watermark, respectively. This paper is organized as follows: Section 2 reviews basic theory about wavelet transform, adaptive tabu search and data payload. In section 3, watermark embedding algorithm is explained. In section 4, we discuss the watermark detection process. Experiment results and conclusion are given in sections 5 and 6.

2. Preliminaries

2.1 Discrete Wavelet Transform (DWT)

The discrete wavelet transform has received a tremendous amount of interest in many important signal processing applications including audio and image watermarking. It has been developed with the idea of looking at a signal at various scales and analyzing it with various resolutions. The basis functions are obtained from a single prototype wavelet by dilations, contractions and shifts. The principle objective of the wavelet transform is to hierarchically decompose an input signal into a series of successively lower frequency approximation subband and their associated detail subbands. For the dyadic wavelet decomposition, at each level, the low frequency approximation subband and detail subband (or subbands for multidimensional case) contain the information needed to reconstruct the low frequency approximation signal at the next higher resolution level.

2.2 Adaptive Tabu Search (ATS)

ATS is a newer version of AI searching technique than the conventional tabu search (TS). ATS [9] is faster and more efficient searching algorithm than the TS method. The important feature behind this algorithm is the tabu list. It keeps the history of movements from the iterative searching process toward the solution. This search is moved from current solution to find the best solution repeatedly. The search direction may lead to a local minimum problem. The tabu list gives new direction that prevents this problem. ATS is different form TS because of two additional tools: back-tracking and adaptive radius.

The back-tracking process is one way to escape from the local optimum. This tool gives new direction to move for the next search when the number of solution is repeated until it reaches the maximum allowance. The new direction is selected from the tabu list.

The adaptive radius process decreases the search area when the searching process is near global solution. The key of success is a set of decreasing radius which should be appropriately chosen. The large radius (coarser resolution) may lead the process to overlook the best solution but the small radius may consume more computational time.

2.3 Data Payload

The data payload refers to the number of bits that are embedded into original audio within a unit of time. It is measured by *bps* (bit per second). The length of host audio is S second. The watermark data is M bit. The data payload B of this algorithm is defined as:

$$B = M/S \quad \text{bps} \tag{1}$$

3. Embedding Algorithm

In this paper, the watermark data is a binary logo image. The embedding algorithm is performed to the wavelet coefficients obtained from 5-level wavelet decomposition. The security of the algorithm is enhanced by performing a random permutation to the watermark image. The watermarking embedding algorithm is as follows:

1. The watermark data, which is a $M_1 \times M_2$ binary logo image, is transformed into a unidimentional antipodal sequence $w(i) \in \{+1, -1\}$ where M_1 and M_2 are the number of rows and columns of the binary watermark image. Then, we generate the random sequence $r(i)$ which is used to encrypt watermark to ensure security.

2. The input audio signal sampled at 44100 Hz is decomposed into five levels using 4-coefficient Daubecies wavelet (Db4). Next, all obtained wavelet coefficients at coarsest approximation subband are divided into k segments where $k = M_1 M_2$. After this step, we have segmented wavelet coefficients, $C_k(i)$.

3. The average value of each segment, $m_k(i)$, is calculated and removed from all wavelet coefficients in $C_k(i)$ at the coarsest approximation subband to facilitate the embedding process. Let $C_k'(i)$ be the modified $C_k(i)$

4. Then, each bit of watermark data is embedded into each $C_k'(i)$ using the following method: if $w(i)=1$, all coefficients in selected segment are added by $\alpha \cdot \beta$. If $w(i) = -1$, they are subtracted by $\alpha \cdot \beta$, where β is the magnitude of $m_k(i)$ and α is embedding intensity. The values of α is searched by using the ATS when the SNR value of the watermarked audio signal and the similarity value of the detected watermark are the objective function. Then, the watermarked wavelet coefficients $Y_k(i)$ are obtained from the following equation:

$$Y_k(i) = C_k^{'}(i) + \alpha\beta w(i) \quad (2)$$

5. IDWT is applied to the modified wavelet coefficients to transform them back to the audio signal in time domain.

The block diagram of the watermark embedding algorithm is shown in Figure 1. Figure 2 illustrates the diagram inside the watermark insertion block.

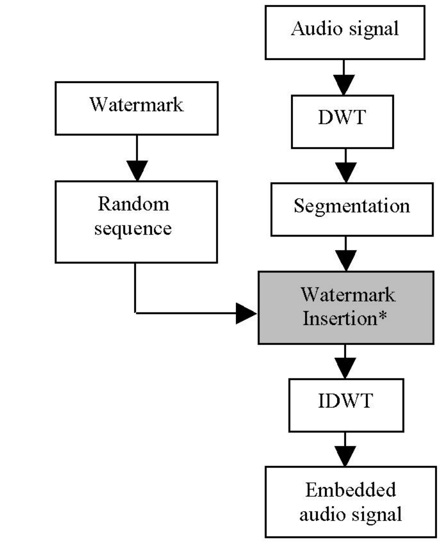

Fig. 1 Diagram of watermark embedding algorithm

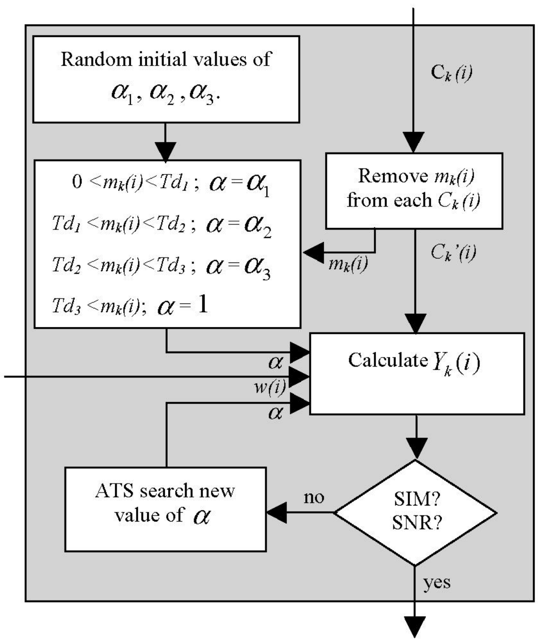

Fig. 2 Diagram inside "Watermark insertion*" block

In Figure 2, Td_1, Td_2 and Td_3 are threshold values. The $m_k(i)$ from each segment is compared with these threshold values to select an appropriate α. In this paper, these threshold values are obtained from the experiment. We suggest that these values should be lower than average value of all m_k.

According to the International Federation of the Phonographic Industry (IFPI), the SNR of watermarked audio signal should be greater than 20 dB. From the experiment, we found that if the SNR of the watermarked audio signal is 23 dB, it is hard to distinguish the watermarked signal from the original one. We select SNR value (SNR=23 dB) and similarity (SIM=1) between extracted watermark and original watermark to be the objective functions of ATS to search proper value of embedded intensity. The SNR value can be set higher for applications that require very high quality watermarked audio signal.

4. Detection Algorithm

The detection algorithm is performed without using the original audio signal. We first decompose watermarked audio signal into 5-level wavelet decomposition. Then, we segment the coefficients at the coarsest approximation subband as in the embedding process and calculate the mean value of each segment of wavelet coefficients, If the mean is larger than zero, a bit "1" is detected. If the mean is lower than zero, a bit "-1" is detected. This step is repeated until all embedded bits are detected. Then, we decrypt the watermark by using the same random sequence used in embedding procedure. Finally, all detected bits are rearranged to form a binary image as a detected watermark.

5. Experimental Results

The algorithm is applied to a set of audio signal including rock, pop, dance, country, and classical (instrument) music. Each music is about 30 seconds in length, 16-bit mono and sampled at 44100 Hz sampling rate.

Watermark image is a binary image with the size of 25×25 pixels. We compute similarity between extracted watermark and original watermark by the following correlation equation:

$$Sim = \frac{\sum_{i=1}^{M1}\sum_{j=1}^{M2} W(i,j)W^*(i,j)}{\sqrt{\sum_{i=1}^{M1}\sum_{j=1}^{M2} W(i,j)^2}\sqrt{\sum_{i=1}^{M1}\sum_{j=1}^{M2} W^*(i,j)^2}} \quad (3)$$

where W and W^* are original and extracted watermarks, respectively, i and j are indexes of the binary watermark image.

The threshold values Td_1, Td_2 and Td_3 are 0.001, 0.01 and 0.1, respectively, and data payload B is approximately 21 bps. We use MATLAB 6.5 as a simulation program. A blind listening test was performed to confirm transparency of the watermark. Most listeners could not distinguish the watermarked signal from the original one.

Figure 3 illustrates the convergence of the ATS optimization for the example of classical music (classic2). The objective value S can be computed from the following equation:

$$S = \sigma_1 DSNR + \sigma_2 DSIM \qquad (4)$$

where *DSNR* is the difference between obtained SNR from each iteration and desired SNR (SNR=23 dB) , *DSIM* is the difference between obtained similarity from each iteration and desired similarity (SIM=1). σ_1 and σ_2 are weighting factors of *DSNR* and *DSIM* indexes, respectively. Each weighting factor represents how important each index is during the searching process. In this experiment, both weighting factors are set to 0.5 because both indexes are equally important.

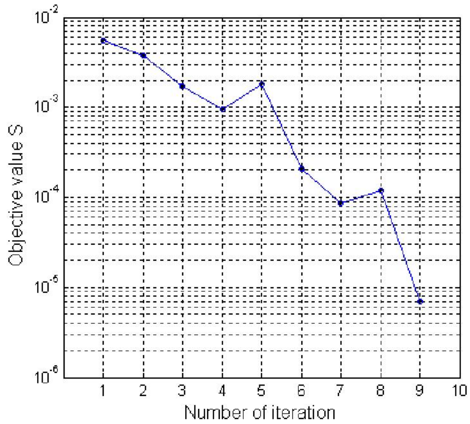

Fig. 3 Convergence of the ATS for the example of classical music

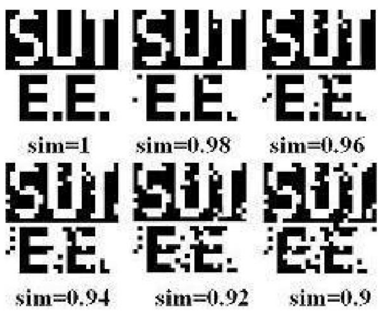

Fig. 4 Extracted watermark image with various similarity values

Extracted watermark images with difference similarities are shown in Figure 4. The comparison of our algorithm is made with the algorithm using the error correction codes. Table 1 shows the similarities of watermarked signal from both algorithms after various attacks. It can be seen that the watermark using ATS is more robust than the one using the error correction codes.

In order to see the effect of using different wavelet, we perform the experiment using other wavelet families such as Haar, Coiflets and Symlets. We found that different wavelet has little effect to the performance of the algorithm.

6. Conclusions

In this paper, a novel robust audio watermarking algorithm using ATS was described. Watermark insertion and watermark detection processes are performed in wavelet transformed domain. The embedding intensity is searched by using an artificial intelligent technique called the adaptive tabu search (ATS). The results show that the embedding algorithm based on ATS achieves more robust watermark than the conventional error correction code algorithm.

7. Acknowledgment

The authors would like to thank Flt.Lt. Prayoth Kumsawat for valuable suggestion and comments, and Suranaree University of Technology for all its support.

8. References

[1] S. Wu, J. Huang, D. Huang and Y. Q. Shi, "Efficiently Self-Synchronized Audio watermarking for Assured Audio Datat Transmittion," *IEEE Trans. on Broadcasting*, vol. 51, no. 1, pp. 69-76, Mar. 2005.

[2] C. Xu, J. Wu, Q. Sun, "Digital audio watermarking and its application in multimedia database," *Fifth International Symposium on Signal Processing and its Applications (ISSPA '99)*, pp. 91-94, Aug. 22-25, 1999.

[3] Y. Shao, G. Wu and X. Lin, "A wavelet based adaptive watermarking algorithm," *Inter. Conf. Info-tech and Info-net (ICII'01)*, vol. 3, pp. 384 - 389, Beijing, 2001.

[4] L. Cui, S. Wang and T. Sun, "The application of binary image in digital audio watermarking," *Neural Networks and Signal Processing*, vol 2, pp. 1497 – 1500, 2003.

[5] X. He and M.S. Scordilis, "Improved spread spectrum digital audio watermarking based on a modified perceptual entropy psychoacoustic model," *IEEE Southeast-Con.*, pp. 283 – 286, Apr. 2005.

[6] X. Li, M. Zhang and S. Sun, "Adaptive audio watermarking algorithm based on SNR in wavelet domain," *Natural Language Processing and Knowledge Engineering*, pp. 287 – 292, 2003.

[7] L. Cui, S. Wang and T. Sun, "The application of wavelet analysis and audio compression technology in digital audio watermarking," *Neural Networks and Signal Processing*, vol 2, pp. 1533 - 1537, 2003.

[8] R.Tu and J. Zhao, "A novel semi-fragile audio watermarking scheme," *Haptic, Audio and Visual Environments and Their Applications (HAVE'03)*, pp. 89 – 94, 2003.

[9] K-N. Areerak and S. Sujitjorn, "Performance comparison between genetic algorithm and Tabu search methods," *Suranaree J. Sci. Technol.*, vol. 9, no.1, (in Thai), pp. 61-68, 2002.

Attacks	Rock1	Rock2	Pop1	Pop2	Dance1	Dance2	Country1	Country2	Classic1	Classic2
No attack	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1
Down-sampling (44kHz to 22kHz)	1	1	1	1	1	1	1	1	1	1
	0.996	1	0.998	1	0.998	0.996	1	1	1	1
Low pass filter (cutoff frequency 3kHz)	1	1	0.997	0.995	0.998	0.978	0.998	0.997	1	1
	0.961	0.968	0.976	0.961	0.949	0.949	0.947	0.944	0.993	1
Random noise (power 1% of original audio signal)	1	0.993	0.993	0.997	0.998	0.992	0.997	0.998	0.998	1
	0.982	0.930	0.955	0.961	0.970	0.982	0.982	0.961	0.989	1
White Gaussian noise (power 1% of original audio signal)	1	0.998	0.998	0.995	0.998	0.997	1	0.998	0.998	1
	1	1	0.964	0.968	0.979	0.975	0.990	0.996	1	1
MP3: 56kbps	0.990	0.992	0.995	1	0.993	0.998	0.997	1.000	0.998	1
	0.951	0.954	0.978	0.970	0.975	0.954	0.982	0.989	1	1
Jitter 1%	1	1	0.998	0.997	1	0.998	1	0.998	1	1
	0.998	0.998	0.998	0.982	0.982	0.979	0.996	0.996	1	1
Cropping 10000 samples (7.6%) at 10 random position	0.962	0.966	0.971	0.966	0.962	0.966	0.962	0.966	0.971	0.971
	0.949	0.949	0.967	0.944	0.951	0.944	0.961	0.961	0.951	0.948

Table. 1 Similarity of watermarked audio signal with various attacks (White lines are the results using ATS and shaded lines are the results using error correction code).