Digital Audio Blind Watermarking Algorithm Based on Audio Characteristic and Scrambling Encryption

WU Weina
School of mathematics and computer, Guangdong Ocean University
Haida Lu No.1, Mazhang District
zhanjiang 524088, China
Suona140@163.com

Abstract—**In this paper, combining with the zero-crossing rate and short-time energy of the audio frame characteristics, an algorithm based on BTC, Arnold and chaos scrambling encryption for seeking the audio frame suitable for the embedded watermark is proposed. And also, the energy is dynamically adjusted to determine the embedding strength so as to self-adaptively embed watermark in the audio frame. Furthermore, the proposed algorithm, as a zero-watermarking algorithm, doesn't need to modify the original audio signal and can realize blind detection during watermark extraction. The imperceptibility of the algorithm is good, and the security is well due to the Arnold and chaos scrambling encryption. The simulation attack results show that the algorithm has good robustness.**

Keywords—digital audio watermarking; Arnold scrambling; chaos encryption; discrete wavelet transform (DWT); audio characteristic

I. Introduction

The rapid development of digital multimedia technology and Internet technology makes the transmission and download of the multimedia digital production such as image and audio have become extremely convenient. But the phenomena such as illegal copying and spread of piracy make the legal interests of the authors and publishers of the productions to be severely violated [1]. Digital watermark technology is an effective way to solve the problem of copyright [2]. In order to better protect their rights, the digital audio watermarking technology is generated at the right moment, thus providing strong technical support for copyright protection, manipulation detection, etc. During the audio watermark embedding process, the high security of the watermark image should be maintained and it is also requested to better recover the original characteristics of the original audio signal after audio watermark embedding and obtain good imperceptibility. The various audio watermark embedding technologies are proposed [3-6], and among these technologies, DWT is widely employed because of its multi-scale and multi resolution characteristics [7]. And also, the imperceptibility and robustness of audio watermark have been realized. But, the parameters of most algorithms when embedding are stationary, and improper watermark embedding strength may cause severe redundancy to the original carrier signal, resulting in increasing the integration difficulty of the inaudibility and the robustness of the audio watermark. Or, the algorithm is relatively complex, and the computation is large. Or, some audio watermark algorithms must to employ the original audio signals when extracting [8-10], result in increasing the complexity of watermark detection. Or, the security of watermark is relatively low. Blind watermark algorithm is a difficulty of the audio watermark, but it is also the tendency. Therefore, adaptive blind detection watermarking algorithm arises opportunely in the process of digital watermarking applications.

The algorithm present is zero-watermark algorithm due to not need the original audio signal in this paper, and has good imperceptibility. Furthermore, the algorithm can realize blind detection during watermark extraction. In addition, Before watermark embedding, Arnold scrambling is firstly implemented for the watermark image, and then the watermark is converted into one-dimensional sequence, finally the chaos iteration is adopted to further encrypt the scrambled sequence to obtain high security, so the algorithm have high security.

II. Watermarking Embedding

A. Watermarking Preprocessing

The original watermark image is firstly preprocessed with BTC so as to decrease the information quantity of a watermarking, which enhances the carrying capability of an original audio signal. Furthermore, to improve security of original watermark image, after encoded by BTC, the watermarking is encrypted by double encryption, Arnold and chaos scrambling. The watermarking preprocessing is followed by BTC, Arnold scrambling, dimension-decrease and chaos encryption.

Firstly, the original audio signal A is divided into M audio segments, and each segment has $L'=(LA/M)$ sampling points, where LA is the length of the audio signal A.

Secondly, assuming that the size of an original gray-level watermarking image U is MxN. In order to achieve BTC operation of a watermark image, the image is divided into the size of 2x2 without overlapping sub-block, and then the mean value of each sub-block is calculated as a threshold value. The pixels below the threshold value are mapped as “0” while the pixels above the threshold value are mapped as “1” in the sub-blocks, and the mean values of the sub-blocks below the threshold value and above the threshold value are taken and reserved as reconstruction level (key1), and accordingly form

978-1-4673-8979-2/17/$31.00 ©2017 IEEE

bitmap U' composed of bit mapping. Arnold scrambling operation is implemented for the bitmap to obtain U'' which is taken and reserved as key2. Arnold transformation formula is as follows:

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

Suppose that the watermark image after Arnold scrambling operation is expressed as follows:

$$U^{"} = \{u(i,j), \\ 1 \le i \le M, 1 \le j \le N, u(i,j) \in \{0,1\}\} \quad (2)$$

Because the watermarking image U is two-dimension, while the audio signal A is one-dimension, it should be converted into one-dimensional sequence V before the image is embedded into a audio signal A, then the one-dimensional sequence V can be expressed as

$$V = \{v(k) = u(i,j), 1 \le i \le M, 1 \le j \le N, \\ k = (i-1) \times N + j, 1 \le k \le M \times N\} \quad (3)$$

where $U(i,j) \in \{0,1\}$; $M \times N$ is the size of the watermark image.

In this paper, in order to further eliminate the correlation of the neighborhood elements in the one-dimensional sequence and improve the robustness and the security of the watermark algorithm, after decreasing audio signal A into one-dimensional sequence, chaos scrambling is employed for image encryption.

Subsequently, following the dimension decrease, Logistic sequence P is generated by chaotic iteration equation. And the appropriate parameters are selected to make it turn into the chaotic state. The parameters are an optional position a, threshold value λ and the initial value x_0. The watermark is processed and reserved as key3, and the P can be express by

$$P = \{p(k), 1 \le k \le M \times N, p(k) \in \{0,1\}\} \quad (4)$$

After watermarking image encryption, watermark sequence W can be expressed as

$$W(k) = V(k) \oplus P(k) \quad (5)$$

where $V(k), P(k) \in \{0,1\}, 1 \le k \le M \times N$.

B. Adaptive Watermarking Embedding

The digital audio watermarking algorithm proposed in this paper do not need modify the original audio, and so it is zero-watermarking algorithm. We employ two local characteristics of an original audio information, zero-crossing rate and short-time energy, to choose the audio frame suitable for the watermarking to be embedded. If watermark embedding strength is too large or too small, it will cause severe redundancy to the original carrier signal. Therefore, in this paper the embedding strength of the watermarking is dynamically adjusted according to wavelet coefficient, improving robustness of the watermarking image.

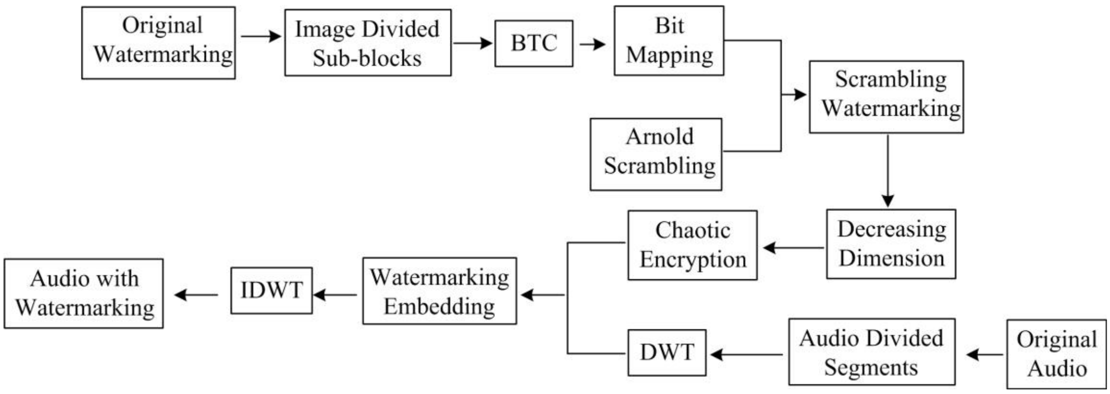


Fig. 1 Schematic diagram for audio watermarking embedding process

According to the characteristic analysis of the audio signal, two threshold values E_0 and Z_0 are set during watermarking embedding, and the audio frame meeting the following conditions: zero-crossing rate $Z \le Z_0$ and short-time energy $E \ge E_0$ are selected, and the watermarking embedding strength is dynamically adjusted according to the wavelet coefficient of the audio frame to realize watermarking embedding. Audio watermarking embedding process is shown in Fig. 1. Specifically, the algorithm steps are as follows:

(1) Supposing audio signal $A = \{a(k), 1 \le k \le L\}$, where L is the length of the audio signal, and then audio signal is framed as follows:

$$a_{kj} = a((k-1) \times F + j), 1 \le k \le K, 1 \le j \le F \quad (6)$$

where a_{kj} is the jth signal value of the kth frame of audio signal; K is the total frame number of the audio information; F is the sample point number of each frame of audio signal.

(2) The audio frames meeting the condition are selected, namely, zero-crossing rate $Z \le Z_0$ and short-time energy $E \ge E_0$. And then, the method of energy dynamic adjustment is used to adaptively embed watermarking.

Z_n and E_n of K frames are orderly calculated. For each kind of audio signal, the audio frames meeting the condition of $Z_n \le Z_0$ and $E_n \ge E_0$ are selected, and spliced to obtain the audio frame Ae, which is embedded watermarking image, and then Ae is divided into $M \times N$ sections. Therefore, each frame audio has Q=$\lfloor \frac{L}{M \times N} \rfloor$ audio data. The audio frame after spliced can be express as follows:

$$Ae = \{Ae(k), 1 \le k \le Q\} \tag{7}$$

The aim of an audio divided into sections is to make the watermarking signal uniform distribution in the audio carrier, and then locate conveniently the position of the watermarking when embedding.

For the each audio signal section $Ae(k)$ after divided, H-layer discrete wavelet transform (DWT) is respectively carried out. For example, after DWT operation is implemented for the Kth segment of the audio signal, the wavelet coefficients obtained thereby are as follows: Ae_K^H, De_K^H, De_K^{H-1} z $\cdots$ De_K^1, wherein Ae_K^H is the low-frequency coefficient of H-layer DWT. Because the high frequency coefficients of DWT describe the detail components of an audio signal, while the low frequency coefficients describe the rough components of the audio signal, the watermarking information will be embedded into the low frequency coefficients in the algorithm proposed.

$$Ae_K^{H\prime} = Ae_K^H\left[1+\alpha_k m(i)\right] \tag{8}$$

where α_k is the strength of a watermarking embedding. After DWT, by calculation of the average energies of the approximate component coefficients of each audio signal, combining the basic embedding strength β, the embedding strength of the audio data segments is dynamically adjusted to adaptively determine the embedding strength for each audio data segment. So, the α_k can be expressed by

$$\alpha_k = (\sum_{k=1}^{Q} Ae(k)^2)/Q \times \beta \tag{9}$$

According to the equation to obtain the modified wavelet coefficients, the wavelet transform is carried out, and then the digital audio embedded with watermarking are obtained. Subsequently, the audio with watermarking is integrated to that without watermarking, so it is concluded that the final digital audio signal containing watermark.

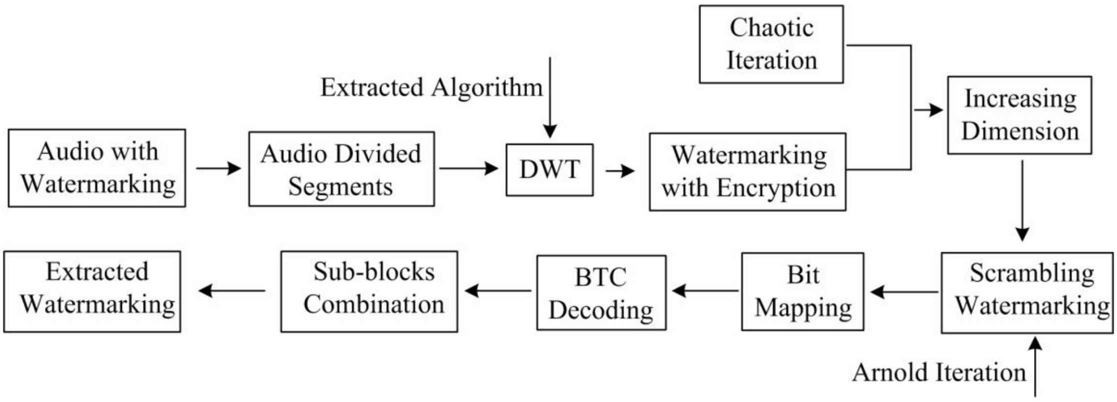


Fig. 2 Schematic diagram for audio watermarking extraction process

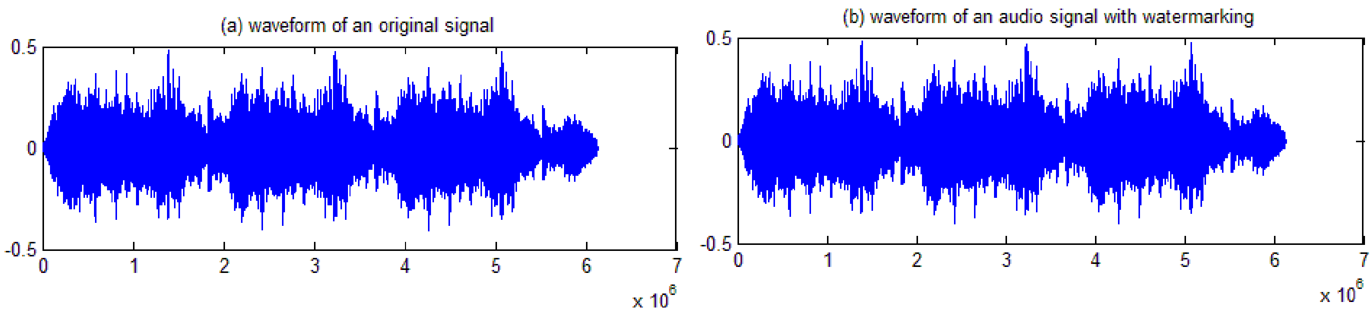


Fig. 3 Waveform of the original audio signal and the audio signal with watermarking

III. Watermarking Extraction

The watermarking extraction process is actually the inverse process of watermarking embedding process. Audio watermarking extraction process is shown in Fig. 2. The watermarking algorithm proposed in this paper is a blind watermarking extraction algorithm because of without the original audio signal when extracting the watermarking image. The specific implementation steps are described as follows:

(1) The H-layer of the DWT of the audio signal with watermarking is operated, and the low frequency coefficient of each section $A_{e_K}^{H}{}'$ is obtained.

(2) Calculating the embedding strength α_k of each audio data section to extract the watermarking image information, the extraction equation can be expressed by

$$m(i)=\left|A_{e_K}^{H}{}'-A_{e_K}^{H}\right|/\alpha_k \tag{10}$$

(3) The initial position, threshold value and initial value of chaotic encryption are employed to generate Logistic sequence, and then key23 is used for binarization processing to get $P(k)$. So, one-dimensional sequence is

$$V(k)=W(k)\oplus P(k) \tag{11}$$

(4) The one-dimensional sequence V(k) is increase to two-dimensional watermarking image, and then key1 is employed for Arnold inverse conversion to obtain the bitmap after BTC operation, and then BTC decoding operation is implemented for this bitmap to reconstruct the image to obtain the original watermarking image U.

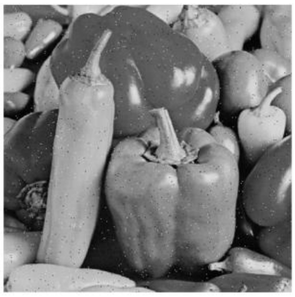
IV. Simulation Experiment and Results Analysis

Experiment simulation test is carried out based on Matlab 7.0. Fig.3 shows audio signal before and after embedded watermarking respectively. It is seen from Fig. 3 that there is no difference in the audio signal before and after the watermarking embedding, and the original audio signal quality isn't affected. Fig.4 illustrates the original watermarking image and extracted watermarking image. The PSNR is 35.2606, and NC is 1.

(a) original watermarking image (b) watermarking image extracted

Fig. 4 Original and extracted watermarking image

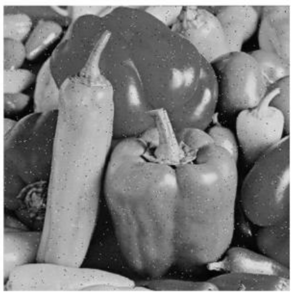
NC=1
BER=0
PSNR=35.2606
(a) original watermark



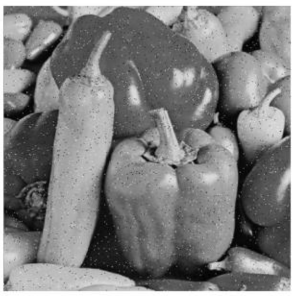
NC=0.958900
BER=0
PSNR=31.4312
(b) re-sampling

NC=0.968600
BER=0
PSNR=32.9822
(c) re-quantization

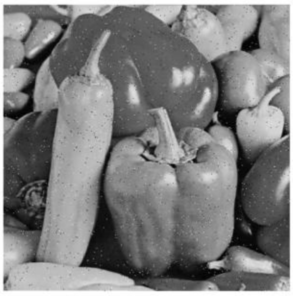
NC=0.960800
BER=0
PSNR=31.0268
(d) low-pass filter



NC=0.951800
BER=0.190270
PSNR=33.5268
(e) echo interference



NC=0.819620
BER=1.247260
PSNR=19.9332
(f) adding noise



NC=0.827360
BER=9.860380
PSNR=26.0326
(g) MP3 compression

Fig. 5 Simulation results of the various attacks and the calculated results of the relevant parameters

Audio signal is the watermarking image extracted after various attacks. In order to test the robustness of the proposed algorithm, the watermarking system is attacked by the following attacks: echo interference, low-pass filter, re-sampling, re-quantization, adding noise, and MP3 compression etc. Among these attacks, echo interference is 50ms echo; the cut-off frequency and the maximum attenuation of the low-pass filter is respectively 8kHz and 2dB; the re-sampling is once extraction and interpolation with coefficient of 2; 16bit audio signal is firstly quantized into 8bit audio signal and then re-quantized into 16bit audio signal; the added noise is Gaussian white noise, and its mean value and the variance is respectively 0 and 0.01; during MP3 compression, the audio signal is MP3 compressed into 128kb/s, and then converted WAV format. Fig. 5 is the extracted watermark image under the various attacks as well as the calculated value of the relevant coefficient, normalization correlation coefficient NC, bit error rate BER, the peak signal-to-noise ratio PSNR.

In this paper, the unsmooth and instable audio signal is divided into several short-time stable signals, and the audio frame meeting relevant zero-crossing rate and short-time energy conditions is selected, and the embedding strength of the watermarking to be embedded is adaptively adjusted through dynamic energy adjustment to improve algorithm robustness, without influencing the original audio signal quality.

Therefore, the simulation results and the corresponding calculating show that the proposed algorithm has the following advantages: the carrying capability of the original audio signal has been improved due to employing BTC during watermarking preprocessing; because the Arnold scrambling and chaotic encryption are used during watermarking preprocessing, the security of the algorithm has been greatly improved; the algorithm has good imperceptibility according to the Fig. 3; from the experiment results of various attacks in Fig. 5, the algorithm has good robustness.

V. Conclusions

Based on BTC, Arnold and chaos scrambling encryption, and combining the zero-passing rate and short-time energy characteristics of the audio frame, a method for determining the audio frame suitable for the blind embedded watermarking is proposed in this paper. The preprocessing of watermarking image eliminates the relevant information of the watermark image, result in realizing the hidden of the watermark information. The associated information of the watermarking image is eliminated through the preprocessing of the watermarking image to hide the watermarking information. The masking characteristics of the human auditory system are comprehensively adopted in this algorithm to adaptively determine the watermarking embedding strength, and the average energy of each segment is adopted to dynamically adjust the embedding strength to adaptively embed the watermarking, and meanwhile the blind watermarking extraction is realized in this algorithm, and the original audio signal is not needed for watermarking information extraction. By calculating the average energy of each segment so as to dynamically adjust the embedding strength, the watermarking can be adaptively embedded. Furthermore, the algorithm realizes the blind extraction of the watermark, and the watermarking extraction does not require the original audio signal. The simulation attack results show that the algorithm has good imperceptibility and good robustness, and the security is well due to the Arnold and chaos scrambling encryption.

References

[1] Li Shanshan, Liu Tingting, Bao Kun, Qiao Ruijuan, "The study on the validity and reliability of digital audio watermarking," Computer security, issue 2, pp. 34-36, 2014.

[2] Kob S, Nishimura R, Suzuki Y. "Time-spread echo method for digital audio watermarkring," IEEE Trans on multimedia, vol. 7, issue 2, pp. 212-221, 2005

[3] Cvejic N, Sepanen T, "Audio watermarking using attack characterization, "Electronics Letters, vol. 39, issue 13, pp. 1020-1021, 2003.

[4] Basia P, Pitas I, "Nikolaidis N. Robust Audio Watermarking in the Time Domain," IEEE Trans on Multimedia, vol. 3, issue 2 pp.232-241, 2001.

[5] Cox I J, Kilian J, Leighton F T, et al, "Secure Spread SpectrumWatermarking for Images, Audio and Video," IEEE Trans on Image Processing, vol. 6, issue 12, pp.1673-1687, 1997.

[6] WANG Xiangyang, ZHAO Hong, "A novel synchronization invariant audio watermarking scheme based on DWT and DCT," IEEE transactions on signal processing, vol. 54, issue 12, pp. 4835-4840, 2006.

[7] Mohidul Islam M, Debnath R, Alamgir Hossian S K, DWT digital watermarking technique and its robustness on image rotation , scaling, JPEG compression, cropping and multiple watermarking," International conference on information and communication technology, pp. 246-249, 2007.

[8] Rodrigueag C, Miyatake M N, Meanahm P, "Analysis of audio watermarking schemes," IEEE Proceeding of the 2nd International Conference on Electrical and Electronics, pp. 17-20,2005.

[9] Chen Licong, Yao Zhiqiang, "Blind Audio Watermarking Algorithm Based on Local Extremum," COMPUTER ENGINEERING AND APPLICATIONS, vol. 41, issue 19, pp. 41-43, 2005.

[10] Xu Dacheng, Shi Minghui, "Digital Audio Watermarking Technology Based on DWT," COMMUNICATIONS TECHNOLOGY, issue 11, pp. 86-88, 2002.