International Journal of Computational Intelligence Systems, Vol. 7, No. 1 (February 2014), 147-159

A Novel Audio Aggregation Watermarking Algorithm Based on Copyright Protection

Rangding Wang[*], **Qian Li**, **Chao Jin**, **Juan Li**, **Yiqun Xiong**
Faculty of Information Science and Engineering, Ningbo University, 818 Fenghua Road Jiangbei District Ningbo, 315211, China

Received 11 September 2012

Accepted 3 July 2013

Abstract

A novel algorithm of audio aggregation watermarking was proposed in this paper for copyright protection. The algorithm not only protects copyright of a single audio in audio aggregation but also protects copyright of the whole audio aggregation. The experimental results showed the watermark of aggregate audio can resist to some attacks, for example, deleting audio, and substituting audio; at the same time, watermark of every single audio is robust to some attacks such as low-filtering, resampling, requantization, noise addition and mp3 compression. The proposed algorithm has good imperceptibility，strong robustness and low complexity.

Keywords: lifting wavelet; audio aggregation; audio watermarking; random mixing matrix.

1. Introduction

With the rapid development of information technology, the copyright protection of digital products has been paid more and more attentions. As a pioneer of watermarking technology, electronic watermark was firstly proposed by Tirkel et al,[1] and then the concept of digital watermark had be published in their famous paper "A Digital Watermark".[2]

Generally, a watermarking scheme should meet the following criteria.

(i) Imperceptibility (Invisibility).Digital watermark hidden in carrier cannot be perceived. Imperceptibility can be evaluated using both objective and subjective measures. In practice, we commonly use our subjective feeling to test the invisibility of the watermark. When our human senses cannot tell whether it contains a watermark or not, we assume that the watermark is transparent and invisible. On the other hand, According to IFPI (International Federation of the Phonographic Industry) recommendations, a watermarked audio signal should maintain more than 20 dB SNR.

(ii) Robustness. Relatively complete watermark information is still able to be extracted by algorithm when the watermarked carrier subjected to some attacks, which illustrates the watermark can resist these attacks, namely, it has robustness. Furthermore, if it resists much more attacks, the stronger robustness will be.

(iii) Security. When the watermarked carrier intercepted by malicious attacker, and he has known the watermark embedding algorithm, but it still cannot extract the watermark correctly without the right key. In order to ensure the security, digital watermarking technology often used with encryption methods in watermarking algorithm.

With respect to the domain in which the watermark is embedded, the techniques can be classified into two categories: spatial-domain and frequency-domain techniques. Spatial-domain algorithms[3,4] directly embed the watermark into the host carrier, whereas frequency-domain algorithms[5,6] embed the watermark based on modifying the coefficients in a transform domain such as Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), and Discrete Wavelet Transform (DWT).Compared with spatial-domain methods, transform domain methods are more robust against many common attacks. In this paper, in order to enhance the robustness of the scheme, we apply the transform domain technique.

Lifting-based second-generation wavelet transforms have been proposed by Sweldens.[7] Lifting scheme for wavelet is an effective method to construct second-generation wavelet. Integer to integer wavelet transform can be achieved by

[*] Corresponding author: wangrangding@nbu.edu.cn

wavelet lifting scheme. Meanwhile, lifting scheme provides more rapid implementation for the first generation wavelet. It is much simpler than the first generation wavelet in terms of construction, since it does not depend on Fourier transform. And all the first generation wavelet can be achieved by wavelet lifting scheme. Due to the advantages of lifting wavelet transform, it has recently being used for digital watermarking applications.[8] Therefore digital watermarking technology based on lifting wavelet transform has a good application prospect. One-level LWT can decompose an audio into low-frequency components and high-frequency components. In the analysis of multi-level decomposition for general audio, we find that the components at the lowest level collect most energies of the audio. If the audio is so disturbed by noise that a certain degree of distortion is produced, generally it will not lead to a greater impact on this part. It indicates that the audio retains its major components, and also shows that the main components of the audio have a strong anti-interference ability. Therefore, embedding watermark in this part, we can get a better anti-interference. On the other hand, it will do bad impact on perceived audio quality. To resolve this conflict, zero-watermark technology has been utilized in this paper.

Traditional watermarking algorithm embeds the watermark by modifying the information of carrier in spatial domain, frequency domain. This embedding method has led to distortion more or less on original carrier which has bad impact on imperceptibility. Zero-watermarking algorithm[9] is proposed to solve this problem. The main principle is constructing the zero-watermark according to the features from carrier itself or the coefficients of carrier in frequency domain. Zero-watermark technology solves the contradiction between robustness and imperceptible perfectly because it does not has any impact on audio quality perception. Zero-watermarking technology is first applied to the image by many scholars.[10] Later, it also has been used to audio watermarking study.[11] Zero-watermark is mainly constructed in DCT domain,[12] DWT domain[13] and high-order cumulant[14] by setting threshold. Determine whether the similarity between the extracted watermark and the original watermark is more than the threshold to get the copyright information. Zero-watermarking technology also has disadvantages. For instance, zero-watermark is meaningless sequence. In addition, zero-watermarks constructed by different carriers have high similarity which goes against determine the watermark when detected.

In generally, almost all existed watermarking technologies are only used to protect the copyright of a single carrier.[15-19] The protection method of copyright of multimedia works in a form of aggregation has lacked up to now. As an example, audio works are often published in a form of audio album so that every audio work has the same copyright in the album. Thus, the existed audio watermarking just embeds a copyright watermark into every audio in album for protecting all of audio works in the aggregation, respectively. That is to say, every audio work in the album needs to embed the same complete watermark. The robustness and transparency of watermarking cannot be achieved a high balance with this method. Therefore, a watermarking scheme that can deal with an audio aggregation will bring much convenience. Furthermore, pirates are more willing to work on the entire audio aggregation for commercial interests. Thus it can be seen that watermark technology that lets audio aggregation as carrier has a bright future. An audio aggregation watermarking technology (AAWT) was firstly proposed in Ref. 20 for protecting a whole audio aggregation. Liu et al.[21] used the labeled-codeword based on VQ and the middle-frequency DCT coefficients to propose a robust audio watermarking algorithm. In Ref. 22, the codebook based on VQ is divided into subcodebooks by codeword pairing. Then, the watermark is embedded by modulating the middle-frequency coefficients according to the watermarking sequence and the genus bits of subcodebooks after MDCT transform. And Li et al.[23] proposed a robust audio aggregation watermark based on vector quantization.

In this paper, an audio aggregation watermarking algorithm is further presented for implementing the copyright protection of a whole audio aggregation and the copyright protection of a single audio in the aggregation, simultaneously. The basic idea is there are several different audio works in an audio aggregation, which means they have different features respectively. The zero-watermark information for each audio is constructed by extracting the features from the corresponding audio work. Finally, zero-watermark is generated for audio aggregation by some processing.

Compared with the known audio aggregation watermarking algorithms in literature, the advantages of our scheme are as follows.

(1). It has low computational complexity, because that zero-watermarking technology was adopted to avoid complex watermark embedding process; lifting wavelet transform was used to increase computing velocity; and the mixing matrix is the random mixing matrix which is generated easily.

(2). It has good imperceptibility and strong robustness. Our algorithm was proposed based on the zero-watermarking technology which has been struck a better balance between algorithm's robustness and cover's invisibility. Experimental results confirmed that our approach keeps the merits of zero-watermarking technology, and resolves the contradiction between imperceptibility and robustness.

(3). In terms of characteristics of audio aggregation, some attacks would be taken on audio aggregation are proposed.

The paper is organized as follows. In section 2, basic theory have been presented, including AAWT, the theory of lifting wavelet transform and Euclid method, the wavelet decomposed result is deduced and used in the proposed algorithm and Chaos Theory. Section 3 presents the total framework of the algorithm, i.e., the scheme of aggregate Zero-Watermark (ZW). The simulation results and analysis of copyright protection are illustrated in section 4. Finally, conclusions are drawn in section 5.

2. Basic Knowledge

2.1. *Audio aggregation watermark technology*

Allow me introduce the main idea of the audio aggregation watermark technology (AAWT) first. In AAWT,[24] audio aggregation as a whole is taken as a carrier for watermark embedding and extraction. The watermark is relative to the whole audio aggregation. In this technology, we can achieve copyright protection and authentication for audio aggregation by just completing embedding and extraction one time.

The technology of copyright protection and authentication for audio aggregation is more complicate than the traditional technology which just takes single audio as carrier. In AAWT, many factors must be considered because it is based on an audio aggregation which may not be appeared in traditional audio watermarking technology. For example, how to embed a single watermark in the whole audio aggregation and how to extract it, several attack methods for audio aggregation must be proposed besides traditional attacks. These factors are different from the traditional audio technology. Based on this main ideology, the existence of watermark is changed accordingly because the carrier is audio aggregation, not the single audio. Firstly, the main change is embedding method. The most critical function is to link the various audio works in the audio aggregation organically. Then the audio works which have relation between each other in logic will form a copyright information related aggregation. From the basic idea of AAWT, it requires the watermark must be extracted on the whole audio aggregation, not single audio work. That is to say, the single audio which contained watermark information in aggregation after watermark embedding cannot declare the copyright for aggregation. We can declare the copyright when extract the watermark information from a relatively complete audio aggregation.

From the perspective of audio aggregation, an algorithm for copyright protection has been proposed in this paper. The overall model of the algorithm is shown in Fig.1. Firstly, using fast lifting wavelet decomposition technique to extract features from every audio work in aggregation. Secondly, construct zero-watermarks (ZWs) for every audio according to their features. Zero-watermark (ZW) of single audio work is used to implement copyright protection of corresponding audio. Thirdly, compose all zero-watermarks (ZWs) of all audio works into a zero-watermark matrix (ZWM), where every zero-watermark (ZW) is row vector of the ZWM; Finally, generate a random mixing matrix to mix with the ZWM, and then get aggregate zero-watermark, which is copyright of audio aggregation, is used to implement copyright protection of the audio aggregation.

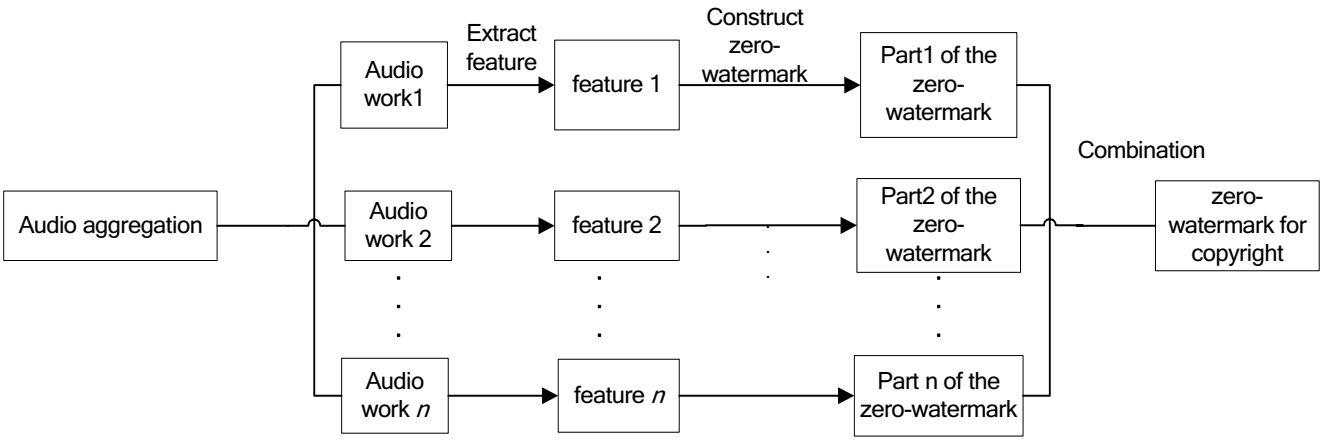

Fig.1. The method of audio aggregation zero-watermark generation based on watermark distribution

2.2. *Lifting wavelet transform*

Traditional wavelet transform[25] is a convolution of the signal and the filter, which has large amount of calculation and need additional data space for storing the convolution results. Wavelet basis of the traditional and first-generation wavelet transform is constructed by translating and flexing the base in Euclidean space. So it is not suitable for the applications of non-Euclidean space. Lifting Wavelet transform gives the full explanation for the wavelet in all spatial domains, which has many excellent features, such as simple structure, low computation, reversible transform for integer to integer and so on. Therefore, lifting wavelet scheme came into being as an ideal method for constructing the second-generation wavelet transform. Therefore, in this paper, we used lifting wavelet[26, 27] to finish wavelet decomposition rapidly, which does not depend on Fourier transform, fully complete construction of orthogonal wavelet filter in time domain.

Provided that finite impulse response filter (FIR) $h = \{h_k\}_{k=k_1}^{k_2}$, its Z transform is a Laurent polynomial $h(z) = \sum_{k=k_1}^{k_2} h_k z^{-k}$. In the bi-orthogonal wavelet transform, the decomposition filters ($\tilde{h}$ and $\tilde{g}$) and synthesis filters (h and g) must satisfy following conditions for ensuring perfect reconstruction of original signal, that is

$$h(z)\tilde{h}(z^{-1}) + g(z)\tilde{g}(z^{-1}) = 2 \tag{1}$$

$$h(z)\tilde{h}(-z^{-1}) + g(z)\tilde{g}(-z^{-1}) = 0 \tag{2}$$

However, multiphase expression of filter h is $h(z) = h_e(z^2) + z^{-1}h_0(z^2)$, where h_e contains even coefficients $h_e(z) = \sum_k h_{2k} z^{-k}$, h_0 contains odd coefficients $h_o(z) = \sum_k h_{2k+1} z^{-k}$, so a multiphase matrix can be used to express the filter pair (h, g), which is given as

$$p(z) = \begin{bmatrix} h_e(z) & g_e(z) \\ h_o(z) & g_o(z) \end{bmatrix} \tag{3}$$

There are three processes: split, predict and update in the forward process of lifting wavelet transform, which is shown in the following Fig.2.

It is easy to find that there is another advantage of lifting wavelet transform. Once we know the forward transform, the inverse transform will be got easily. So the inverse process of lifting wavelet transform is shown in Fig.3.

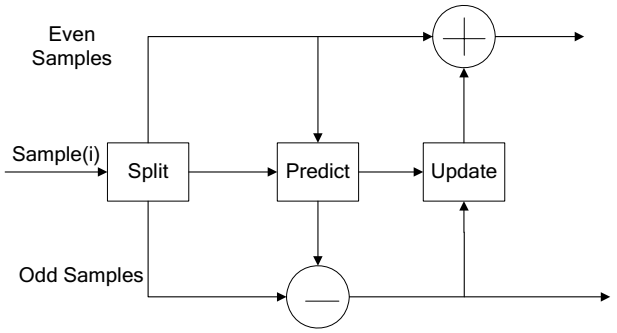

Fig.2. The forward process of lifting wavelet transform

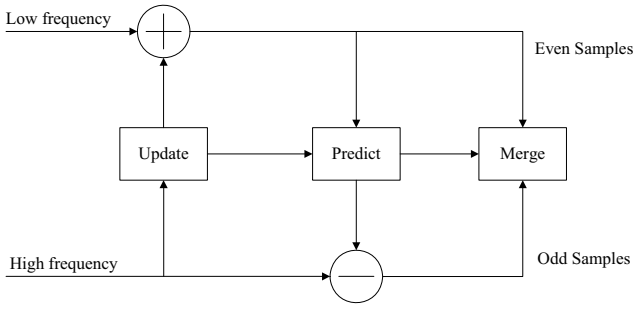

Fig.3. The inverse process of lifting wavelet transform

From Fig.3, we conclude the inverse process as follow. Firstly, the even sequences are recovered by completing the inverse operation of update. Secondly, the odd sequences are recovered by completing the inverse operation of predict. Finally, the original signal will be obtained by merging the even sequence and odd sequence.

The specific process of lifting wavelet transform is illustrated as following.

Lazy wavelet transform is firstly adopted to decompose original signal into the composition of even sequence and odd sequence. Then, in terms of the original filter h and g, a new high-pass filter g^{new} is attained as

$$g^{new} = g(z) + h(z)s(z^2) \tag{4}$$

And then based on the original filter h and filter g^{new}, a new low-pass filter h^{new} is attained as

$$h^{new} = h(z) + g(z)t(z^2) \tag{5}$$

where $s(z)$ and $t(z)$ are both Laurent polynomials. Evidently, multiphase matrix (3), (4) and (5) can be expressed as following

$$p^{new}(z) = p(z)\begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix}, p^{new}(z) = p(z)\begin{bmatrix} 1 & 0 \\ t(z) & 1 \end{bmatrix} \quad (6)$$

According to Euclid method,[28, 29] corresponding to a given complementary filter pair (*h*, *g*), *p*(*z*) is formed as

$$p(z) = \prod_{i=1}^{m}\begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix}\begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix} (1 \le i \le m) \quad (7)$$

where $s_i(z)$ and $t_i(z)$ are Laurent polynomials and *K* is a non-zero constant.

Thus wavelet transforms of the finite impulse response filters can all be achieved by repeating above processes. In this paper, we choose traditional compact wavelet base 'db2', then low-pass filter *h* and high-pass filter *g* are

$$\begin{aligned} h(z) &= h_0 + h_1 z^{-1} + h_2 z^{-2} + h_3 z^{-3} \\ g(z) &= -h_3 z^2 + h_2 z - h_1 + h_0 z^{-1} \end{aligned} \quad (8)$$

where $h_0 = (1+\sqrt{3})/(4\sqrt{2})$, $h_1 = (3+\sqrt{3})/(4\sqrt{2})$, $h_2 = (3-\sqrt{3})/(4\sqrt{2})$, $h_3 = (1-\sqrt{3})/(4\sqrt{2})$, multiphase matrix is

$$p(z) = \tilde{p}(z) = \begin{bmatrix} h_0 + h_2 z^{-1} & -h_3 z - h_1 \\ h_1 + h_3 z^{-1} & h_2 z + h_0 \end{bmatrix} \quad (9)$$

By Euclid method, decomposed results is

$$p(z) = \tilde{p}(z) = \begin{bmatrix} 1 & -\sqrt{3} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \frac{\sqrt{3}}{4} + \frac{\sqrt{3}-2}{4} z^{-1} & 1 \end{bmatrix}\begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix}\begin{bmatrix} (\sqrt{3}+1)/\sqrt{2} & 0 \\ 0 & (\sqrt{3}-1)/\sqrt{2} \end{bmatrix} \quad (10)$$

2.3. *Chaos theory*

Chaotic sequence is a pseudo-random sequence which can be generated easily by the iterative equations, nonlinear equations or partial differential equations.[30,31] Therefore, many literatures use chaotic sequence to realize scrambling, encrypting and hiding information because it has following properties: easy generation, strong sensitivity to initial conditions and completely reproducible. At present, the commonly used methods of one-dimensional chaotic map are mainly as: Logistic map, Chebshev map, Reny map, Torus automorphism and so on.

In this paper, we use Logistic map which defined as:

$$x_{k+1} = \mu x_k (1 - x_k) \quad (11)$$

Where μ usually takes a real number close to 4. Research shows when $\mu \in (3.56994567, 4]$, Logistic map is in the chaotic state. In the experiment, take $\mu = 4$, initial value $x_0 = 0.39$. As chaotic Logistic map's Chaos and the strong sensitivity to initial conditions, the attackers could not predict the watermark when using it to control digital watermark sequence generated which can enhance the security of the system.

3. Proposed Algorithm

Fig.4 shows the framework of the proposed algorithm. Firstly, zero-watermark of each audio work is constructed, denoted as $w(k)$ $(1 \le k \le n)$. The copyright information of single audio can be determined by computing correlation value between $w(k)$ and $w'(k)$. $w'(k)$ is the extracted zero-watermark of the *kth* corresponding audio work. Secondly, the aggregate zero-watermark is formed by composing zero-watermark of each audio work and mixing with ICA matrix. The copyright information of audio aggregation can be determined by computing the correlation value between the original aggregate zero-watermark W and the extracted aggregate zero-watermark W'.

3.1. *Watermark forming process*

Fig.5 shows the process of watermark forming, which contains forming of aggregate zero-watermark (ZW) and forming of every single ZW. Detailed steps are as follows.

Step1. Every audio work in original aggregation $\{x_1, x_2, \cdots, x_k, \cdots x_n\}$ is putted into the 4 level lifting wavelet decomposition, respectively. Then corresponding low-frequency coefficients of each audio are gotten, an aggregation composed by these coefficients is $\{A_4(1), \cdots, A_4(k), \cdots, A_4(n)\}$, where *n* is number of audio works in original audio aggregation, $A_4(k)$ is the low-frequency coefficients of *kth* audio in audio aggregation.

Step2. Chaotic sequence is used to modulate $A_4(k)$, the initial value of the chaotic sequence is selected as 0.39 which is saved as a Key. The modulated aggregation is $\{A_4^{zl}(1), \cdots, A_4^{zl}(k), \cdots, A_4^{zl}(n)\}$, where $A_4^{zl}(k)$ is modulated low-frequency coefficients of *kth* audio in audio aggregation.

Step3. ZW of corresponding audio is constructed with the modulated low-frequency coefficients of every audio in audio aggregation and *n* ZWs of *n* audio works are formed. And then a zero-watermark matrix (ZWM) *M* is composed with these n ZWs, where every ZW is row vector of the ZWM.

The forming process of ZW of every single audio is as following: given *k* is positive integer which is confined to [1-n], ZW *w*(*k*) of *kth* original audio work is calculated, that is

$$w(k)(q)=\begin{cases}1 & A_4^{zl}(k)(q) > A_4^{zl}(k)(q+1)\\ 0 & A_4^{zl}(k)(q) \le A_4^{zl}(k)(q+1)\end{cases} \quad (12)$$

where $w(k)(q)$ is *qth* coefficient of ZW $w(k)$, $q \in [1, t-1]$, $t = length(A_4^{zl}(k))$, $A_4^{zl}(k)(q)$ is *qth* coefficient of $A_4^{zl}(k)$.

Step4. A mixing matrix is generated to mix with above ZWM *M*, rank of *M* is *n*, here, the mixing matrix is a random mixing matrix *A*, the element value of which is confined to [0-1], the matrix *A* is saved as a Key. We also use other mixing matrix to mix with above ZWM *M* such as ICA mixing matrix $B^{32,33}$ and random matrix *C* which is conformed to Gaussian distribution. According to experimental result in Section IV, the robustness of aggregate ZW which is gotten by using *B* or *C* to mix with *M* is poorer than the robustness of aggregate ZW which is gotten by using *A* to mix with *M*. So the random mixing matrix *A* is adopted in this paper.

Step5. Matrix *A* and matrix *M* are multiplied, and then get a matrix *W*, which is an aggregate ZW of the whole audio aggregation.

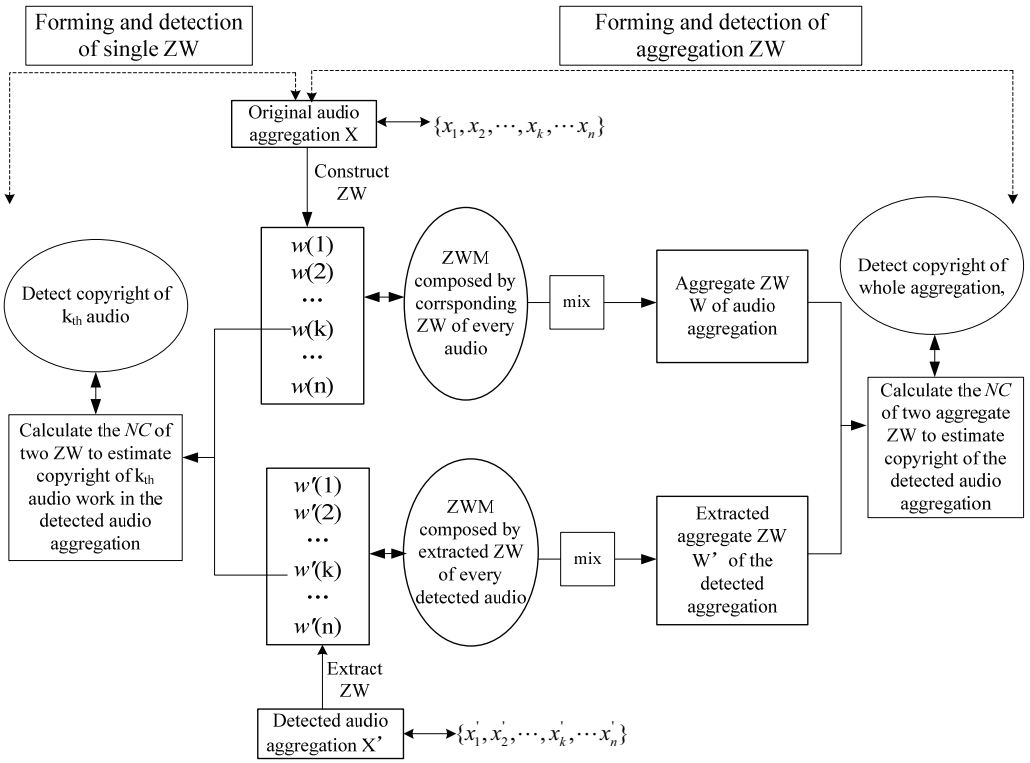

Fig.4. The total framework of the proposed algorithm

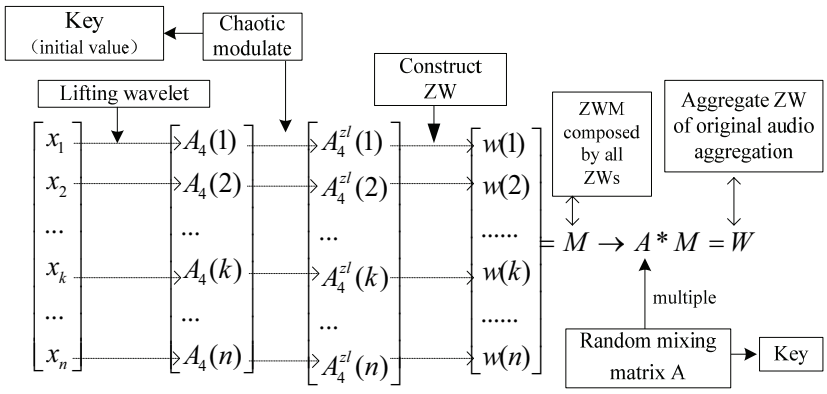

Fig.5. Zero-watermark forming process

3.2. *Watermark extraction process*

During watermark extraction process, the Key which is saved in watermark forming process and the order of all audio works in audio aggregation is needed. Fig.6 shows the process of watermark extraction, which contains the extraction of every single ZW from every detected audio work in the detected audio aggregation and the extraction of aggregate ZW from the detected audio aggregation.

Step1. Every detected audio work of the detected aggregation $\{x_1^{'}, x_2^{'}, \cdots, x_k^{'}, \cdots x_n^{'}\}$ is putted into the 4 level lifting wavelet decomposition, respectively. Then corresponding low-frequency coefficients of each detected audio are gotten, an aggregation composed by these coefficients is $\{A_4^{'}(1), \cdots, A_4^{'}(k), \cdots, A_4^{'}(n)\}$, where $A_4^{'}(k)$ is the low-frequency coefficients of *kth* detected audio in the detected audio aggregation.

Step2. Using chaotic sequence to modulate $A_4^{'}(k)$, the initial value of the chaotic sequence is the Key 0.39 which was saved in watermark forming process. The modulated aggregation is $\{A_4^{'zl}(1), \cdots, A_4^{'zl}(k), \cdots, A_4^{'zl}(n)\}$, where $A_4^{zl}(k)$ is the modulated low-frequency coefficients of *kth* detected audio in the detected audio aggregation.

Step3. ZW of corresponding detected audio is extracted with the modulated low-frequency coefficients of every detected audio in the detected audio aggregation and *n* ZWs of *n* detected audio works are extracted. And then a ZWM *M'* are composed with *n* ZWs, where every ZW is row vector of the *M'*.

Step4. Matrix *A* which was saved in watermark forming process multiplied with the matrix *M'*, and then get a matrix *W'*, which is the extracted aggregate ZW from the whole detected audio aggregation.

Step5. In order to judge whether the detected audio aggregation has copyright information, it is necessarily to calculate the correlation value *NC*(*W*, *W'*) of original aggregate ZW *W* and the extracted aggregate ZW *W'*,

$$NC(W, W') = \frac{\sum_{i=1}^{n}\sum_{j=1}^{s} W(i,j)W'(i,j)}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{s} W^2(i,j)}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{s} W'^2(i,j)}} \tag{13}$$

where *W*(i, j) and *W'(i, j)* are corresponding coefficients. If the *NC* is larger than the first threshold which was set in advance, the detected audio aggregation has the copyright information; or else, the detected audio aggregation does not have the copyright information.

Step6. In order to judge whether the single detected audio has copyright information, computing correlation value *NC* of ZW of every original audio and the extracted ZW of every corresponding detected audio, and calculating *BER* of the extracted ZW of every detected audio,

$$NC(w(k), w'(k)) = \frac{\sum_{q=1}^{s} w(k)(q)w'(k)(q)}{\sqrt{\sum_{q=1}^{s} w(k)(q)^2 \times \sum_{q=1}^{s} w'(k)(q)^2}} \tag{14}$$

where *NC*(*w*(*k*),*w'*(*k*)) is correlation value of between ZW of *kth* audio in original audio aggregation and the extracted ZW of *kth* audio in detected audio aggregation, *w*(*k*)(*q*) and *w'*(*k*)(*q*)is corresponding coefficients.

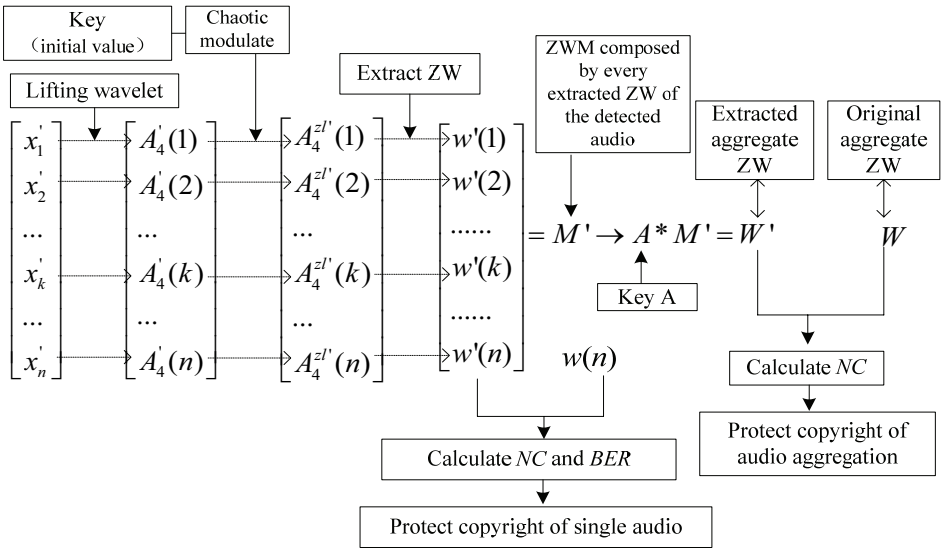

Fig.6 Watermark extraction process

$$BER(w(k), w'(k)) = \frac{1}{m}\sum_{q=1}^{m}\begin{cases}1 & w(k)(q) \neq w'(k)(q)\\ 0 & w(k)(q) = w'(k)(q)\end{cases} \quad (15)$$

where m is length of ZW $w(k)$.

If the *NC* is larger than the second threshold which was set in advance and the *BER* is less than the third threshold which was set in advance too, the corresponding detected audio has the copyright information, or else, it doesn't have the copyright information.

4. Experimental Analysis

In our experiment, five types of music (20s, mono, 16bits/sample, 44.1 KHz and WAVE format) such as blues, classical, country, folk and pop, is chosen to compose an audio aggregation. And 2 audio works are selected from each kind of music, so an audio aggregation has n=10 audio works. The first threshold and the second threshold are equal and set as 85%, the third threshold is set as 30% by a mount of experimental tests. In the following, the robustness of the whole audio aggregation and single audio in audio aggregation are analyzed respectively. Moreover, the elapsed times of watermark forming and watermark extraction which compose the proposed algorithm were given at the end of this section.

4.1. Analysis of the robustness for the whole audio aggregation

Usually traditional watermarking algorithm selects a single audio as carrier to be researched, and adopts some attacks such as low-pass filtering, resampling, requantization, noise addition, mp3 compression, cropping to analyze the robustness of the algorithm. In this paper, we attempt to propose following five attacks to analyze the robustness of the audio aggregation watermarking algorithm.

(i) Traditional Attack

All the detected audio works in the detected audio aggregation are attacked by traditional attacks which are shown in Table 1. Table 1 shows the *NC* of original aggregate ZW which is formed with *A*, *B* or *C* matrix and the extracted aggregate ZW which is extracted with corresponding matrix *A*, *B* or *C*. The robustness of aggregate ZW, which is generated by using *B* (ICA mixing matrix) or *C* (Gaussian distribution matrix) to mix with *M*, is compared with the robustness of aggregate ZW, which is produced by using *A* (random mixing matrix) to mix with M, the NC is larger, the robustness is stronger.

According to Table 1, the *NC* of extracted aggregate ZW and original aggregate ZW when ZW is produced by using random mixing matrix *A* are larger than the *NC* of extracted aggregate ZW and original aggregate ZW when by using the matrix *B* or *C*, which indicates that the matrix *A* is superior to *B* and *C*, so *A* is adopted in this paper. And by Table 1 when using matrix *A* the *NC* of original ZW and the corresponding extracted ZW from the detected audio aggregation are larger than the first threshold 85% after the detected audio aggregation suffering from every traditional attack, which implies the proposed algorithm can resist to traditional attacks.

In addition, we make use of a common robustness evaluation tool -- "Stirmark Benchmark for Audio" to attack the watermarked audio aggregation. Default parameters are accepted in our experiments and Ref. 34 gives the description of all attacks in this system. All audio works in the audio aggregation would be attacked when evaluating the robustness. Table 2 shows the detection results of the proposed method, scheme [21], scheme [22] and scheme [23]. It can be seen that the four algorithms are all robust to some common signal processing attacks including noise addition, lossless compression and so on. But the ability of resisting attacks su-

Table 1. Test parameters under traditional attacks

Attack type	*NC*			Attack type	*NC*		
	Mixing matrix				Mixing matrix		
	A	*B*	*C*		*A*	*B*	*C*
No attack	1.0	1.0	1.0	Mp3compression (128kbps)	0.9055	0.5418	0.5123
Low-filtering (6 order butterworth 22.05kHz)	1.0	0.9729	0.9999	Cropping (front 10%)	0.8987	0.5008	0.4735
Resampling (22.05kHz)	0.9840	0.9252	0.9230	Cropping (middle 10%)	1.0	0.9999	0.9925
Requantization (8bit)	0.9968	0.9853	0.9812	Cropping (behind 10%)	1.0	1.0	0.9873
requantization (32bit)	1.0	0.9987	0.9842	Noise addition (26dB)	0.9811	0.9246	0.9321

Table 2. Robustness comparison under attacks in Stirmark Benchmark for Audio

Attack type	*NC*			
	Ref. 21	Ref. 22	Ref. 23	This paper
No attack	0.9921	1.0000	1.0000	1.0000
Addbrumm_100	0.9915	1.0000	1.0000	1.0000
Addnoise_100	0.6386	0.8071	0.8661	1.0000
LSBzero	0.9915	1.0000	1.0000	0.9891
Extrastereo_30	0.9921	1.0000	1.0000	0.9945
Normalize	0.9921	1.0000	1.0000	1.0000
FFT_real_reverse	0.9915	1.0000	1.0000	0.9956
Compressor	0.9890	0.9994	0.9970	1.0000
Zerocross	0.6310	0.7710	0.8251	0.9114
Flippsample	0.7687	0.8874	0.8619	0.8948
Smooth	0.4523	0.5907	0.6736	0.7366
RC_lowpass	0.5228	0.7141	0.6742	0.8532

Table 3. Test parameters under joint attacks

Attack type	*NC*
Low-filtering(6 order butterworth, 22.05kHz) and resampling(22.05kHz)	0.9791
Low-filtering(6 order butterworth, 22.05kHz) and requantization(8bit)	0.9893
Mp3 compression (128kbps) and cropping (front 10%)	0.8879
Noise addition(26dB) and mp3comression(128kbps)	0.9088
Low-filtering(6 order butterworth, 22.05kHz) and resampling(22.05kHz) and requantization 8bit	0.9791
Resampling(22.05kHz) and requantization(8bit) and cropping (middle 10%)	0.9830
Resampling(22.05kHz) and requantization(8bit) and cropping (front 10%)	0.8946

ch as low pass filtering, flippsample and smooth is poor. In conclusion, comparing with other three algorithms, our method can achieve better robustness under most traditional attacks.

(ii) Joint Attack

Every joint attack shown in Table 3 is taken on all the detected audio works in detected audio aggregation simultaneously, and aggregate ZW is extracted, and then the *NC* of the extracted aggregate ZW and original aggregate ZW is calculated.

According to Table 3, the *NC* shown is Table 3 are larger than 85%, which indicates the proposed algorithm can resist to above joint attacks.

(iii) Dividing Aggregation Attack.

The detected audio aggregation is divided into p subsets, and different attacks shown in Table 4 and Table 5 are taken on each subset, respectively. The number of subsets p is optional, and the number of the detected audio works in a subset is also optional, Table 4 and Table 5 show the *NC* of original aggregate ZW and the extracted aggregate ZW which is extracted under conditions that p select five different value, p equals to 5, 4, 4, 3, 3 or 3, respectively, and every subset is suffered from different attack.

According to Fig.7, after dividing aggregation into above five situations, the *NC* of the extracted aggregate ZW and original aggregate ZW are larger than 85%, which indicates the proposed algorithm can resist to above dividing aggregation attacks.

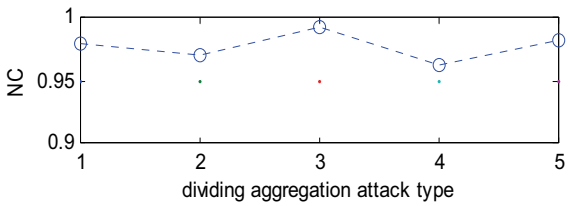

Fig.7 NC under every type of dividing aggregation attack

(iv) Deleting Audio Attack

During watermark extraction process, the order of every audio works in original audio aggregation is known, the deleted audio can be searched after the detected audio aggregation are suffered from deleting audio attacks and substituted with zero sequence, and then the aggregate ZW of

Table 4. Test parameters under dividing aggregation attacks(Audio1-Audio6)

	Audio1& Audio2	Audio3	Audio4	Audio5	Audio6
Attack type1	Low-filtering (6 order butterworth, 22.05kHz)	Resampling (22.05kHz)	Resampling (22.05kHz)	Requantization (8bit)	Requantization (8bit)
Attack type2	Low-filtering (6 order butterworth, 22.05kHz)	Resampling	Mp3 compression (128kbps)	Mp3 compression (128kbps)	Mp3 compression (128kbps)
Attack type3	Low-filtering (6 order butterworth, 22.05kHz)	Low-filtering (6 order butterworth, 22.05kHz)	Requantization (8bit)	Requantization (8bit)	Noise addition (26dB)
Attack type4	Resampling (22.05kHz)	Resampling (22.05kHz)	Resampling (22.05kHz)	Mp3 compression (128kbps)	Mp3 compression (128kbps)
Attack type5	Low-filtering (6 order butterworth, 22.05kHz)	Low-filtering (6 order butterworth, 22.05kHz)	Resampling (22.05kHz)	Resampling (22.05kHz)	Resampling (22.05kHz)
Attack type6	Cropping (front 10%)	Cropping (front 10%)	Noise addition (26dB)	Noise addition (26dB)	Noise addition (26dB)

Table 5. Test parameters under dividing aggregation attacks(Audio7-Audio10)

	Audio7	Audio8	Audio9	Audio10
Attack type1	Noise addition (26dB)	Noise addition (26dB)	Mp3 compression (128kbps)	Mp3 compression (128kbps)
Attack type2	Mp3 compression (128kbps)	Cropping (middle 10%)	Cropping (middle 10%)	Cropping (middle 10%)
Attack type3	Noise addition (26dB)	Noise addition (26dB)	Cropping (behind 10%)	Cropping (behind 10%)
Attack type4	Mp3 compression (128kbps)	Mp3 compression (128kbps)	Noise addition (26dB)	Noise addition (26dB)
Attack type5	Requantization (8bit)	Requantization (8bit)	Requantization (8bit)	Requantization (8bit)
Attack type6	Cropping (front 10%)	Mp3 compression (128kbps)	Mp3 compression (128kbps)	Mp3 compression (128kbps)

Table 6. The robust watermark detection results under deleting audio attacks

Attack type	*NC*			
	Ref. 21	Ref. 22	Ref. 23	This paper
Delete 1 audio from aggregation	0.9921	0.9993	0.9993	0.9957
Delete 2 audio from aggregation	0.9822	0.9944	0.9397	0.9871
Delete 3 audio from aggregation	0.9103	0.9609	0.8720	0.9802
Delete 4 audio from aggregation	0.6851	0.7198	0.8022	0.9668

the new aggregation is extracted. Table 6 shows the *NC* of original aggregate ZW and the extracted aggregate ZW under deleting audio.

According to Table 6, the *NC* of original ZW and the ZW extracted by the proposed algorithm is larger than the first threshold 85% if the number of the deleted audio is less than 4, the proposed algorithm also can extract aggregate ZW and testify the copyright information, which indicates the proposed algorithm can resist to deleting audio attack better.

(v) Substitute Audio Attack

When the detected audio aggregation is suffered from substituting audio attack, the substituted audio can be searched and substituted with zero sequence, and then aggregate ZW of the new aggregation is extracted. Table 7 shows the *NC* of original aggregate ZW and the extracted aggregate ZW under substituting audio attack.

According to Table 7, when substituted 4 audios from the aggregation, the *NC* of original ZW and the ZW extracted by the proposed algorithm is larger than the first threshold 85%. It indicates that our method can resist to substituting audio attack better.

4.2. Analysis of robustness for single audio in audio aggregation

Above the copyright protection of the whole audio aggregation has been discussed, however, aggregate ZW can not protect copyright of single audio in audio aggregation. To protect copyright of a single audio which is pirated, ZW of every audio works in audio aggregation is constructed, Table 8 shows the *NC* of the extracted ZW from corresponding attacked detected audio and original ZW of an optional selected original audio, and the *BER* of the extracted ZW of corresponding detected audio after attacking.

According to Table 8, all of the *NC* are larger than the second threshold 85%, all of the *BER* are less than the third threshold 30%, which indicates the optional selected detected audio has copyright information, at the same time the *NC* of original ZW of other audio works in audio aggregation and the extracted ZW of corresponding detected audio, and the *BER* of the extracted ZW of corresponding detected audio after being suffered from every kind of attack in Table 8 are calculated. They are very close to the figure in Table 8. Therefore, the single ZW of every audio in the audio aggregation in this paper has strong robustness.

4.3. Analysis of the proposed algorithm's complexity

We have tested the elapsed times of the processes of watermark forming (or watermark embedding) and watermark extraction of algorithm [21], [22], [23] and our algorithm. The experiment has been repeated 50 times for the audio aggregation and any one of it respectively. And the testing environments are as follows: a computer of 3.19GHz CPU and 1.74GB RAM, an IDE of MATLAB R2007a. Table 9 shows the mean elapsed times for the two procedures of each method.

As can be seen, the proposed algorithm has the shortest elapsed times, whether for the watermark forming or watermark extraction. According to section 3 we know that the proposed method adopts zero watermark technology and this can avoid the complicated process of watermark embedding. Besides, lifting wavelet transform which applied to our method has fast computational velocity because its transforming is independent of Fourier Transform. In addition, the matrix which we choose for processing the zero-watermark matrix is generated easily. The points mentioned above have contributed to reducing the complexity of the proposed algorithm.

Table 7. The robust watermark detection results under substituting audio attacks

Attack type	*NC*			
	Ref. 21	Ref. 22	Ref. 23	This paper
Substitute 1 audio from aggregation	0.9909	0.9993	0.9993	0.9957
Substitute 2 audio from aggregation	0.9799	0.9335	0.9725	0.9871
Substitute 3 audio from aggregation	0.7073	0.8137	0.9120	0.9802
Substitute 4 audio from aggregation	0.6497	0.5878	0.8452	0.9668

Table 8. Test parameters of single audio under several attacks

Attack type	*NC*	*BER*	Attack type	*NC*	*BER*
No attack	1.0	0.0	Mp3compression(128kbps)	0.8813	25.2
Low-filtering(6 order butterworth, 22.05kHz)	1.0	0.0	Cropping(front 10%)	0.8912	22.15
Resampling(22.05kHz)	0.9634	3.72	Cropping(middle10%)	1.0	0.0
Requantization(8bit)	0.9953	0.48	Cropping(behind10%)	1.0	0.0
requantization(32bit)	0.9987	0.25	Noise addition (26dB)	0.9421	5.88

Table 9. Comparison of different algorithm's elapsed time

Algorithm process	*Elapsed Time(s)*							
	An audio				An aggregation			
	Ref. 21	Ref. 22	Ref. 22	This paper	Ref. 21	Ref. 22	Ref. 23	This paper
Watermark forming (embedding)	20.4	18.57	34.4	1.44	1639.79	1375.71	3713.93	10.75
Watermark extraction	1.72	1.22	3.41	1.42	25.11	14.90	43.26	10.43

5. Conclusion

Most of pirates aim to the audio aggregation, the copyright protection for audio aggregation is very important, if traditional watermarking algorithm was adopted to protect an audio aggregation, the common watermark would be embedded into every audio in the audio aggregation, respectively, the watermarks must be extracted from every audio in the detected audio aggregation respectively when detecting watermark. The watermark embedding process and the watermark extracting process both have been taken n times, so this algorithm has high computational complex, it only protects copyright of every audio in the audio aggregation, respectively, so if one of the audio works in audio aggregation can not extract watermark, the algorithm can not protect the copyright of the whole audio aggregation. The proposed algorithm in this paper uses AAWA technology to generate the common watermark of the audio aggregation directly, and also extract the watermark from the detected audio aggregation directly.

The watermark forming process and the watermark extracting process both have been taken only one time, and even one or more audio are suffered from deleted or substituted, the algorithm can protect the copyright of the whole audio aggregation. The algorithm not only protects copyright of every single audio in audio aggregation but also protects copyright of the whole audio aggregation. The experimental results show the proposed algorithm has good imperceptibility, strong robustness, and novelty. Several points are shown as follows to discuss for later research. 1) The proposed scheme doesn't overcome shortcomings of zero-watermark. 2) When the audio aggregation has been attacked, the proposed scheme cannot tamper it correctly and also cannot recover it.

Acknowledgements

This work is sponsored by K.C. Wong Magna Fund in Ningbo University, National Natural Science Foundation of China (NSFC: 61170137, 61300055, 61301247), Doctoral Fund of Ministry of Education of China(20103305110002), Zhejiang natural science foundation of China (ZJNSF: LY13F020013), Ningbo natural science foundation of China (2013A610057), Open Fund of Zhejiang Provincial Top Key Discipline of Information and Communication Engineering (XKXL1310).

6. References

1. A. Z. Tirkel, G. A. Rankin, R. M.van Schyndel. Electronic Water- mark, Digital image computing, Technology and Applications (DICTA,1993), pp.666-673.
2. R. G.van Schyndel, A. Z. Tirkel, C. F. Osborne. A Digital Watermark, *First IEEE International Image Processing Conference*, (1994) pp.86-90.
3. M. Celik, G. Sharma, A. M. Tekalp, and E. Saber. Lossless generalized LSB data embedding, *IEEE Trans. on Image Processing.* 14(2)(2005):253-266.
4. C. H. Chang, Z. Ye and M. Zhang. Fuzzy-ART based adaptive digital watermarking scheme, *IEEE Transactions on Circuits and Systems for Video Technology*. 15(1) (2005):65-81.
5. X. Y. Wang, H. Zhao. A novel synchronization invariant audio watermarking scheme based on DWT and DCT, *IEEE Trans. Signal Process*. 54 (12) (2006) 4835–4840.
6. S. Agreste，Guido Andaloro. A new approach to pre-processing digital image for wavelet-based watermark. *Journal of Computational and Applied Mathematics*, 221(2) (2008)274-283.
7. W. Sweldens. The lifting scheme: a construction of second generation wavelets. *SIAM J Math Anal*. 29(2) (1998) 511–46.
8. Amit Bohra,Omar Farooq,Izharuddin, Blind self-authentication of images for robust watermarking using integer wavelet transform,*AEU - International Journal of Electronics and Communications*.63(8) (2009)703-707.
9. C. R. Wang, D. Li. Image zero-watermarking utilizing wavelet zerotree structure and PCA, *Opto-Electronic Engineering*. 32(4)(2005)75-85.
10. J. H. Ma, J. X. He. A Wavelet-Based Method of Zero-watermark, *Journal of Image and Graphics*. 12(4) (2007)581-585.
11. J. Zuo, D. Cui. Zero-watermark resistant to MP3 compression, *Advanced Materials Research*.121-122(2010)254-259.
12. T. Y. Ye. A robust zero-watermarking algorithm against dual print-and-scan process based on discrete cosine transformation**.** *Guangzi Xuebao/Acta Photonica Sinica*.40(1) (2011)142-148.
13. N. Wang, X. Li. RST invariant zero-watermarking scheme based on matching pursuit. *Chinese Journal of Electronics*.15(2) (2006)269-272.
14. Q. Wen, T.F. Sun,S.X. Wang. Concept and Application of Zero-Watermark.*Acta Electronica Sinica*.31(2)(2003)214-216.

15. M. David, H. J. Jordi, M. Julia. A robust frequency domain audio watermarking scheme for monophonic and stereophonic PCM formats. *In Proceeding of the 30th EUROMICRO Conference.* (2004),pp.39-43.
16. J. K. Hyoung, K. Taehoon, Y. In-Kwon. A robust audio watermarking scheme. *In Proceedings of the 2004 International Symposium Circuits and Systems (ISCAS)*, (2004), pp. 696-699.
17. V. Bhat K, I. Sengupta, A. Das. A New Audio Watermarking Scheme Based on Singular Value Decomposition and Quantization. *Circuits Syst Signal Process* (2011). 30: 915–927.
18. T. K. Tewarl, V. Saxena, J. P. Gupta. A Novel Approach to Generate Watermarks Using Auditory Features for Audio Watermarking. *Journal of Theoretical and Applied Information Technology.* Vol. 35 No.2, pp. 156-162
19. M. I. Khan, M. I. H. Sarker, K. Deb, M.H. Furhad. A New Audio Watermarking Method Based on Discrete Cosine Transform with A Gray Image. *International Journal of Computer Science & Information Technology (IJCSIT).* Vol 4, No 4, pp. 119-128.
20. Y. Q. Xiong, R. D. Wang. A Robust Audio Aggregate Zero-Watermark algorithm. *In proceedings of the sixth International Conference on Information Technology: New Generations (ITNG2009).* (Las Vegas, Nevada, USA,2009), pp. 366-370.
21. J. X. Liu, Z. M. Lu, J. S. Pan. A Robust Audio Watermarking Algorithm Based on DCT and Vector Quantization. *In Proceeding of the 8th International Conference on Intelligent Systems Design and Applications(ISDA)*, (2008) pp. 541–544.
22. Y. Zhou, R. D. Wang, D. Q. Yan. An audio watermarking scheme based on VQ codebook pairing. *In Proceeding of the 3rd International Congress on Image and Signal Processing(CISP),* (2010) Vol.8, pp. 4011–4015.
23. J. Li, R. D. Wang, J. Zhu. A Robust Audio Aggregation Watermark Based on Vector Quantization. *Lecture Notes in Electrical Engineering(LNEE),* (2011) vol.102,pp.551-559.
24. J. Li, R. D. Wang, J. Zhu. A Watermark for Authenticating the Integrity of Audio Aggregation Based on Vector Sharing Scheme. *Information Technology Journal*.10(2011)1001-1008.
25. J.A. Antonino-Daviu, M. Riera-Guasp. Feature extraction for the prognosis of electromechanical faults in electrical machines through the DWT, *International Journal of Computational Intelligence Systems*. 2 (2) (2009)158-167.
26. W. Sweldens, P. Schroder. Building your own wavelets at home. In Wavelets in Computer Graphics, *ACM SIGGRAPH Course Notes*, (1996),pp.l5- 87.
27. I. Daubechies. W. Sweldens, Factoring wavelet transforms into lifting steps, *Journal of Fourier Analysis and Applications*. 4(3)(1998)245 -267.
28. Y. B. Xu. The first Chinese translation of the last nine books of Euclid's Elements and its source. *Historia Mathematica*. (2005) vol.32,pp.4-32.
29. Y. V. Genin, Euclid algorithm, orthogonal polynomials, and generalized Routh-Hurwitz algorithm, *Linear Algebra and its Applications*. (1996) vol.246, pp.131-158.
30. S. Ghosh, S. Mishra, P. Saha, Chaos based encryption technique for digital images**,** *International Conference and Workshop on Emerging Trends in Technology,* (2011),pp. 300-303.
31. E. Ott, C. Grebogi, J.A. Yorke. Controlling chaos. *Physical Review Letters*.64(11)(1990)1196-1199.
32. A. Hyvärinen, Survey on independent component analysis, *Neural Computing Surveys*. 2(1999)94-128.
33. P. Comon. Independent component analysis—a new concept. *Signal Processing*, 36(1994) 287-314.
34. J. Dittmann, C. Kraetzer. Audio benchmarking tools and steganalysis(2006). Revision 1.1. Online: http://www.ecrypt.eu.org/documents/D.WVL.10-1.1.pdf