EURASIP Journal on Information Security
a SpringerOpen Journal

RESEARCH Open Access

An enhanced audio ownership protection scheme based on visual cryptography

Rimba Whidiana Ciptasari[1,2*], Kyung-Hyune Rhee[3] and Kouichi Sakurai[2]

Abstract

Recently, several ownership protection schemes which combine encryption and secret sharing technology have been proposed. To reveal the original message, however, they exploited XOR operation which is similar to a one-time pad. It is fairly losing the reconstruction simplicity due to the human visual system (HVS). It should be noted that it is completely different from the original concept of visual cryptography proposed by Naor and Shamir. To decrypt the secret message, Naor and Shamir's concept stacked k transparencies together. The operation solely does a visual OR of the shares rather than XOR, the way HVS does. In this paper, we, consequently, adopt Naor and Shamir's concept to apply correct theory of visual cryptography. Furthermore, audio copyright protection schemes which exploit chaotic modulation or watermark integration into frequency components have been widely proposed. Nevertheless, security issue against intentional distortions has not been addressed yet. In this paper, we aim to construct a resilient audio ownership protection scheme to enhance the security by integrating the discrete wavelet transform and discrete cosine transform, visual cryptography, and digital timestamps. In the proposed scheme, the watermark does not require to be embedded within the original audio but is used to generate a secret image and a public image. The watermark is then acquired by performing OR between the secret and public image. We can alleviate the trade-off expenses between the capacity of data payload and two other important properties such as imperceptibility and robustness without modifying the original audio signals. The experiments against a variety of audio signals processing provided by StirMark confirm superior robustness of the proposed scheme. We also demonstrate the intentional distortion by modifying the original content via experiments, it reveals comparable reliability. The proposed scheme can be widely applied to the area of audio ownership protection.

Keywords: Digital watermarking; Audio ownership protection; Visual cryptography; Transform domain; Timestamp

1 Introduction

1.1 Background

Protection of an intellectual property has become a major problem in the digital age. It is possible to duplicate digital information a million-fold and distribute it over the entire world in seconds through the Internet. There are various techniques for preventing and/or minimizing the risk of copying, making copying easier to detect, and assisting in proving infringement. One of the technical measures is to embed a 'digital watermark' in the host data. The watermark is regarded as a code, which is impossible or very difficult to detect and/or remove, and it can be used to identify the source of the copied data [1]. This aids users in proving copyright infringement.

Among the development of digital watermarkings in a various multimedia, digital audio watermarking provides a special challenge because the human auditory system (HAS) is extremely more sensitive than a human visual system (HVS) [2]. Most audio watermark algorithms insert the information as a plain-bit or adjusted digital signal using a key-based embedding algorithm. The embedded information is hidden and linked inseparably with the source data structure. For the optimal watermarking application trade-offs among competing criteria such as robustness, non-perceptibility, capacity, non-detectability, and security have to be considered. However, there is always trade-off between capacity and other two important properties, non-perceptibility and robustness. A higher capacity is always obtained at the expense of

*Correspondence: rmb@ittelkom.ac.id
[1]Department of Informatics, Telkom University, Bandung 40257, Indonesia
[2]Graduate School of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
Full list of author information is available at the end of the article

either robustness or non-perceptibility (or both) [3]. Further, some audio quality degradations inevitably occur due to the embedding process.

1.2 Related work

In order to eliminate the trade-offs among competing criteria aforementioned, several audio ownership protection schemes [4-6], which are different from the traditional watermarking, have been proposed. These schemes are referred to as *zero-watermarking*. In the paper [4], three-level discrete wavelet decomposition (DWT) is applied to get the low-frequency subband of the host audio, which is the perceptually significant region of it. To make the scheme resist lossy compression operation such as MP3 compression, discrete cosine transform (DCT) is performed on the obtained low-frequency wavelet coefficients. And by considering the Gaussian signal suppression property of higher-order cumulant, the fourth-order cumulants of the obtained DWT-DCT coefficients are calculated to ensure the robustness of the scheme against various noise addition operations. Finally, the essential features extracted based on DWT, DCT, and higher-order cumulant are used for generating binary pattern. In addition, the scheme introduced the presence of the authentication center to keep the copyright information such as the secret keys, original host audio, and the corresponding digital timestamp used in copyright demonstration.

Wang and Hu [5] proposed the scheme created by selecting some maximum absolute value of low frequency wavelet coefficients of original audio. The construction of the watermark is random by chaotic sequence. After generating the watermark, chaotic inverse search is adopted to get the initial value of another watermark sequence that is identical to the original one. In verification phase, instead of using an original audio, they exploited chaotic modulation to generate the original watermark sequence. In order to reduce the processing time, they cut the watermark into fifty sections. According to our experiment, despite long hours of executing the initial value searching process, we could not achieve the convergence condition. The initialization of its initial value is a somewhat trial-and-error process. The time complexity of each section is O(NM) where N indicates the watermark's size, and M refers to the number of iterations. In this case, we cannot predict the M value. We, therefore, argue that their algorithm is not efficient. Moreover, their scheme indeed requires the length of its original watermark sequence to generate original watermark W in extraction stage. This value was not kept either in secret key K or initial vector H. In other words, their scheme cannot be regarded as a blind watermarking.

The authors also proposed a modification of Chen and Zhu's scheme for generating secret keys in their earlier work [6]. Compared to that of Chen and Zhu's, the key's size is relatively the same as its watermark. The scheme, however, is claimed to have good degree of robustness, imperceptibility, and payload capacity.

Furthermore, some ownership protection schemes which combine encryption and secret sharing technology [7-12] have also been proposed, and they achieved good results. Several works in visual cryptography [7,9,11] were performed in a distinctive way. In order to retrieve the secret image, they exploit XOR operation among shares instead of stacking them. This mechanism is considered as an appropriate way to be employed in ownership protection area. Lou et al. [11] proposed the scheme that extracts the feature from the protected image by utilizing the secret key and the relation between the low and middle sub-band wavelet coefficients. Then, the feature and watermark are used to generate a secret image by the codebook of visual cryptography technique. To provide further protection, the secret image, with the exception of the secret key and codebook, is registered to certification authority (CA). In the verification procedure, public image is first generated from the suspected image. The extracted watermark is obtained by performing XOR operation between secret and public image. However, such an impressive combination has not yet been proposed for audio.

Lee and Chen [10] introduced cryptographic tools into the watermarking process to provide security against malicious attacks. As a first step, a gray-level original image was decomposed by exploiting wavelet transform. Vector quantization was then exploited to generate indices set I that would be signed by the owner with digital signature technique. Lastly, the owner sent signed indices set S to a trusted CA. CA digitally added time and date when it received them. This scheme can protect the indices set from alteration, and everyone can use it to verify the copyright logo corresponding to the test image.

Chen and Horng [12] improved their earlier work [10]. In order to resist against geometric distortions, the watermark was first permutated based on two-dimension pseudorandom permutation generated by seed s. Then, the polarity table T was constructed to be used in computing the verification key K. They included digital signature and timestamp to avoid either counterfeit or copy attacks and to make public verification possible. The advantage of their scheme was that it is resistant to blind pattern matching attack.

1.3 Challenge issues

Based on related work, we summarize the following challenge issues:

1. Consider the watermarking scheme proposed by Chen and Zhu [4]. The embedding process takes host

audio A and watermark w as input and generates three secret keys. These keys imply the information of selected frames, extracted feature points, and its watermark, where respectively this information is denoted by K_1, K_2, and K_3. Consider the case when an adversary intends to produce a watermarked file using the same procedure in the paper [4]. The adversary simply extracts the information of selected frames and then applies exclusive-or operation for adversary's watermark like binary image to obtain the K_3. In an extreme case, it is sufficient for the adversary to modify K_3. Thus, K_3 contains the information of watermark. As a result, an adversary can easily produce the information K_1, K_2, and K_3 from an audio file and can claim that the file contains his/her watermark. This situation shows that Chen and Zhu's scheme suffers from security weakness. Referring to the concept which is described in [3,13], the security of watermark algorithms depends on the secret keys used for embedding and recovery process. In contrast to this concept, Chen and Zhu's secret keys are somewhat public knowledge rather than confidential information. The first challenge issue is on how to improve the scheme in order to fulfill an appropriate watermarking concept.

2. As previously mentioned, some image ownership protection schemes [7-12], which combine encryption and secret sharing technology have also been proposed. Regarding original visual cryptography (VC) proposed by Naor and Shamir [14], the ciphertext is supposed to be revealed directly by a HVS. In that case, HVS does a visual OR rather than XOR operation. Unfortunately, most aforementioned existing schemes exploited XOR operation. Hence, the second challenge issue is on how to employ VC correctly in a digital watermarking area.
3. In terms of audio intellectual property protection, both Chen and Zhu [4] and Wang and Hu [5] do not provide any experimental results dealing with security aspects of their scheme against intentional distortions. Although Chen and Zhu [4] registers their secret keys, host original image, and timestamp to CA for copyright demonstration, it reflects that the timestamp is not digitally added by CA. They do not provide a detailed explanation on this issue as well. We argue whether this situation leads to owner's deception. Furthermore, most watermarking algorithms cannot resist against malicious manipulations of the content. Such manipulations may distort audio data as well as readily destroy or even remove the watermark. The last challenge issue is on how to enhance security against intentional distortions.

1.4 Contribution

This paper proposes a novel audio watermarking based on visual cryptography that can be exploited in ownership protection area. Akin to our previous work [6], we extract the feature by performing H-level wavelet decomposition to obtain low-frequency subband of segmented host audio. To make the proposed scheme resistant to lossy compression operation, discrete cosine transform is performed to the obtained low-frequency wavelet coefficients. We use the whole DWT-DCT coefficients rather than a certain part of coefficients to adjust matrix dimension.

In the proposed scheme, the watermark does not require to be embedded into the original audio but is used to generate secret and public share images by using the visual cryptography technique. In a nutshell, feature extraction is first accomplished to obtain digital audio's features by frequency-domain functions. The sharing matrices referred to as *codebook* are then generated in such ways that have two properties: contrast and secrecy. Instead of data embedding, audio's features and binary-valued watermark are integrated to construct secret shares based on generated codebook. In other words, the image shares contain watermark information. In contrast to existing schemes [7-12] that exploit XOR operation, we employ a visual OR of the shares to reveal the original watermark as its original concept stated in [14].

Further, product registration to a trusted authority is a well-established way of protecting intellectual property rights as well as offering indisputable proof of original ownership and legal rights [15]. In order to prevent any intentional distortion, digital timestamping is incorporated in a proposed scheme. Referring to timestamping's mechanism [16], we simplify the protocol by using CA as a trusted party which is responsible for the issuing and verification of timestamps as well as issuing a digital certificate that contains a name of the holder, a serial number, expiration date, and a holder's public key. Therefore, the steps of generating a timestamp are as follows. At first, the owner signs his protected data using his private key and generates a fingerprint by using a digital signature function. Then, the fingerprint is sent to CA. The CA generates a timestamp based on the owner's fingerprint and the date and time obtained from an accurate time source. The timestamp is sent back to the owner. The CA keeps a record of the timestamp for future verification.

The rest of the paper is organized as follows. Section 2 describes the development of an ownership protection scheme. In Section 3, the proposed scheme is investigated against incidental and intentional distortions. Finally, the conclusion is provided in Section 4.

2 Proposed scheme

The proposed scheme comprises two stages: *share image generation stage* and *watermark verification stage*. Host audio is first segmented into several frames, and each frame contains N samples. Next, the sample features are extracted by performing wavelet decomposition to obtain the low-frequency coefficients. Then, DCT is exploited only to the obtained low-frequency wavelet coefficients. Afterward, features of DCT coefficients are calculated. Finally, encoding utilizes these features and binary-valued watermark to generate secret share images according to the concept of Naor and Shamir's scheme [14]. One of the secret share image is then registered to CA for further protection and will be used for watermark verification purpose.

To retrieve the watermark, the received audio is segmented into several frames that contain N samples each. Then, the samples' features are extracted by performing wavelet decomposition to obtain the low-frequency coefficients. Next, DCT is exploited to the obtained low-frequency wavelet coefficients, and the DCT coefficients are calculated. The decoding exploits these features and registered share image to generate a public share image. The watermark is recovered by performing OR operation between secret and public share images and then used to verify the ownership. The following subsections provide more detailed description on each stage.

2.1 Main process in the proposed scheme

2.1.1 Feature extraction

To accomplish feature extraction, the host audio is first segmented into several frames in which each frame contains N samples and T-level wavelet decomposition is performed on each frame. Then, approximated coefficients in the LL_T subband are transformed to DCT coefficients. Let $A^{TC} = DCT(A^T) = \{a^{TC}(n) | n = 1, \ldots, \frac{N}{2^T}\}$ be the obtained DCT coefficients. The output array of DCT coefficients contains real numbers, and they have a range from -1 to 1. The feature type t is then obtained by the following conditions:

$$t = \begin{cases} 1 & \text{-1} \leq a^{HC}(n) \leq 0 \\ 2 & 0 < a^{HC}(n) \leq 1. \end{cases} \tag{1}$$

2.1.2 Encoding and decoding

In principle, encoding is the process of generating secret shares by integrating binary value of the watermark and digital audio's features, while decoding refers to process of revealing the original watermark message by stacking those secret shares.

Formally, the basic model of visual secret sharing is denoted as k out of n problem. Given a secret message, we would like to generate n transparencies so that the original message is visible if any k of them are stacked together; otherwise, the message is totally invisible. We exploit original encryption problem proposed by Naor and Shamir [14], that is a 2 out of 2 or (2,2)-secret sharing problem. The watermark is visible if two shares are stacked together; otherwise, it does not provide any information.

In this paper, the watermark consists of a collection of black and white pixels. Each original pixel appears in n shares, one for each transparency. Each share consists of m black and white sub-pixels. The resulting sharing matrices can be represented as two collections of $n \times m$ Boolean matrices $\mathbf{S}$=$\{S_0, S_1\}$. To share either a white or black pixel, one randomly chooses one of the matrices in either S_0 or S_1, respectively. When transparencies $i_1, i_2, \ldots, i_k$ are stacked together, the black subpixels appearing on a combined share are represented by OR operation of rows $i_1, i_2, \ldots, i_k$ in sharing matrices $\mathbf{S}$. The gray level of this combined share is proportional to the Hamming weight H(V) where V is the m-vector of the resulting OR operation [14].

The sharing matrices should satisfy two properties, namely *contrast* and *secrecy*.

1. In case of contrast, the gray level G is deemed valid if the following condition is satisfied.

$$G = \begin{cases} \text{black} & \text{if } \mathbf{H}(\mathbf{V}) \geq d \\ \text{white} & \text{otherwise} \end{cases} \tag{2}$$

   for a threshold $1 \leq d \leq m$. In order to comply with a condition (2), the codebook shown in (3) and (4) is arranged in such a way that H(V) is 2 or 3 in S_0, while it is 4 in S_1.
2. In terms of secrecy, the number of 1's in $\mathbf{S}$ should have same probability distribution, i.e., codebook shown in (3) and (4) has probability $\text{Prob}(S_i =' 1'/0) = \text{Prob}(S_i =' 1'/1) = 0.5$. Let $S = [s_{ij}]$ be a Boolean matrix with a row for each share and a column for each subpixels. For each pixel, the share matrix must be chosen at random and must be known only by the sender (owner) and receiver (CA), while the codebook is publicly known.

The examples of share matrix representations used in our proposed scheme are described as follows.

$$S_0 = \left\{ \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \right\} \tag{3}$$

$$S_1 = \left\{ \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \right\} \tag{4}$$

2.1.3 Watermark reduction

Since it is accomplished by applying four subpixels per pixel, it affects the aspect ratio of original image. In order to compute bit error rate (BER), it is required to have extracted watermark in the same size as its original. Let $W(M \times N)$ be the original watermark image. Note that the extracted watermark W' will be equal to $M \times 4N$. In order to yield the same watermark size as its original one, it is necessary to accomplish the reduction process of extracted watermark. Assume that black pixel is assigned as 1 and white pixel value is 0, the reduction process is performed based on the following conditions:

$$\text{Reduction result} = \begin{cases} 1 & \text{if the number of black pixel} > 3 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

2.2 Share image generation and verification procedure

2.2.1 Share image generation procedure

Figure 1 illustrates the secret share image generation, and the procedure is described as follows.

Input: host original audio $A = \{a(i) | i = 1, \ldots, L_{\text{sample}}\}$, binary image watermark $W(N\times N)$ $=\{w(i,j)|w(i,j)\epsilon\{0,1\}\}$, and codebook C.

Output: secret share images $S_A(N \times mN)$ and $S_B(N \times mN)$ where m is the number of subpixels per pixel.

Step 1. Firstly, A is segmented into F frames, denoted as $Fr = \{fr_i | i = 1, \ldots, F\}$, and each frame contains N samples. Next, T-level wavelet decomposition is performed on each frame fr_i to yield its coarse signal A^T and detail signal $D^T, D^{T-1}, \ldots, D^1$. Then, to take advantage of low-frequency coefficient, which is robust against signal processing manipulations, DCT is only performed on A^T and obtained DCT coefficients are denoted as

$$A_k^{TC} = \mathrm{DCT}(A_k^T) = \left\{ a_k^{HC}(n) \mid n = 1, \cdots, \frac{N}{2^T} \right\}. \tag{6}$$

Step 2. Construct a new sequence $B_n^{TC} = \{b_n^{TC}(n) | n = 1, \ldots, N/2^T\}$ by taking the first n frames of A_k^{TC}.

Step 3. Let x be 1.

a. Obtain the feature type t from B_n^{TC} based on Equation (1).
b. Construct a secret share block $S(x)$ by utilizing a codebook C as described in (3) and (4), feature type t, and a corresponding watermark pixel value $w(i,j)$.
c. Add x to one. If $x \leq \mathrm{N}\times \mathrm{N}$ then go to **a**.

Step 4. The secret share images $S_A(N \times mN)$ and $S_B(N \times mN)$ are generated. Note that the security of our scheme is based on the $S_A(N \times mN)$.

Step 5. The next step is timestamping for the protected data. The owner signs the security parameter by using digital signature scheme:

$$f = DS_{OPK}(S_A, C) \tag{7}$$

where $DS_{OPK}(\boldsymbol{.})$ is a digital signature function by using the owner's private key *OPK*, and f stands for owner's fingerprint. Afterward, owner sends f, S, and C to the CA. CA creates a timestamp TS with the owner's fingerprint f, and the time t and date d obtained from an accurate time source as

$$TS = TS_{CAPK}(f, t, d) \tag{8}$$

where $TS_{CAPK}(\boldsymbol{.})$ is a timestamp function by using CA's private key *CAPK*. After creating the timestamp *TS*, it is

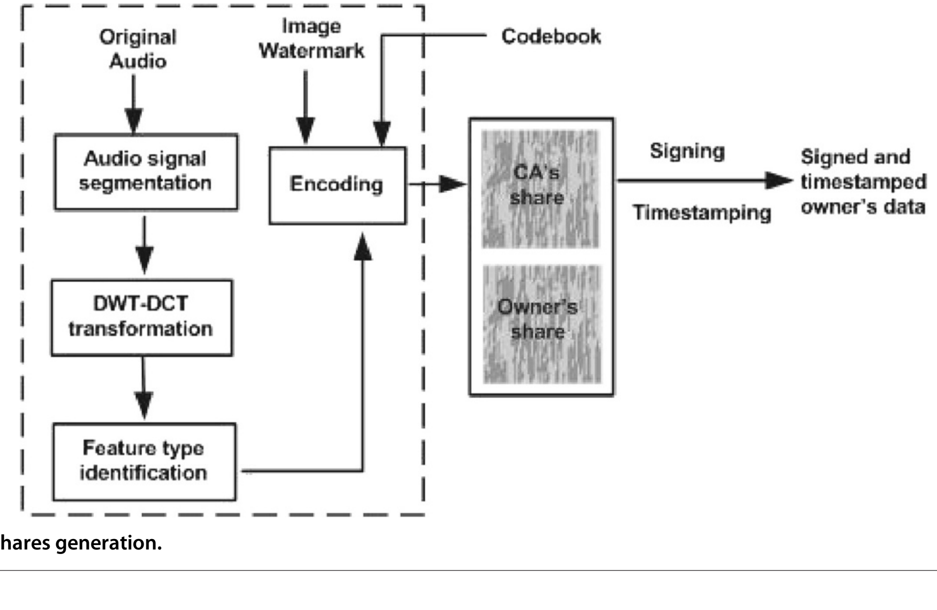


Figure 1 Secret shares generation.

sent back to the owner and kept as an archive by CA as well. Subsequently, *f*, *TS*, S_A, and *C* are used by CA in verification purpose when the dispute arises. Note that timestamping mechanism is completed by CA so that detail discussion of digital signature is beyond the scope of this paper.

2.2.2 Watermark verification and extraction procedure

The presence of original audio is not required in verification and extraction phase. In order to verify the copyright of an audio, anyone can use CA's public key to validate the timestamp *TS* and owner's public key to validate the signature *f*. When a dispute arises or multiple claims occur, the earlier registered data will be regarded as the original one. In the meantime, S_A and *C* are used to verify copyright watermark's logo corresponding to the received audio.

As depicted in Figure 2, the extraction procedure is similar to share image generation procedure and is illustrated as follows:

Input: a received audio $\{A' = a(s)|s = 1, \ldots, L_{sample}\}$, a secret share image $S(N\times N)$, and a codebook *C*.

Output: an extracted watermark logo $EW(N\times N)$

Step 1. A' is segmented into *F* frames, denoted as $Fr = \{fr_i|i = 1, \ldots, F\}$, and each frame contains *N* samples. Next, *T*-level wavelet decomposition is performed on each frame fr_i to yield its coarse signal A^T and detail signal $D^T, D^{T-1}, \ldots, D^1$. Then, DCT is on A^T and obtained DCT coefficients are denoted as $A_k^{TC} = \mathrm{DCT}(A_k^T) = \{a_k^{HC}(n)|n = 1, \cdots, N/2^T\}$.

Step 2. Construct a new sequence $B_n^{TC} = \{b_n^{TC}(n)|n = 1, \ldots, N/2^T\}$ by taking the first *n* frames of A_k^{TC}.

Step 3. Let *x* be 1.

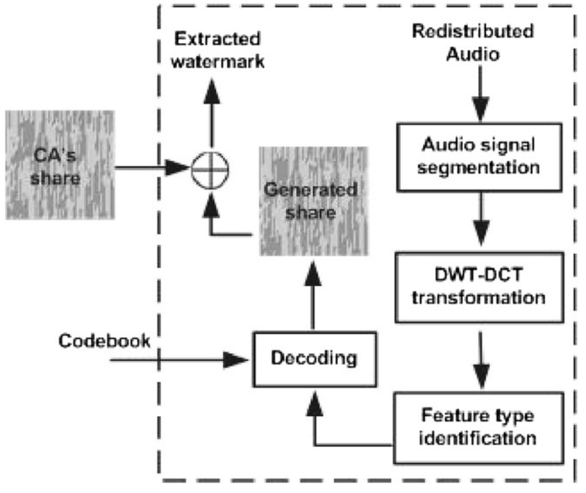


Figure 2 Watermark extraction procedure performed by CA.

a. Obtain the feature type *t* from B_n^{TC} based on Equation (1).
b. Construct a public share block S_Bx by utilizing a codebook *C* as described in (3) and (4) and feature type *t*.
c. Add *x* to one. If $x \le N \times N$ then go to **a**.

Step 4. A public share image $S_BN \times mN$ is yielded. An extracted watermark $W'(N \times mN)$ is obtained by

$$W' = S_A \mathrm{OR} S_B. \tag{9}$$

Step 5. Afterward, watermark reduction process is performed according to Equation 5 to obtain the recovered watermark $EW(N \times N)$.

3 Experimental results

To demonstrate the feasibility of the proposed scheme in terms of ownership protection requirements, some experiments are conducted. Bit error rate is employed to measure robustness of the zero-watermarking system,

$$\mathrm{BER} = \frac{B}{MN}100\% \tag{10}$$

where *B* is the number of erroneously extracted bits. Signal-to-noise ratio (SNR) is the ratio of quality sound to noise. The higher the decibel (dB) value, the better is the quality of the sound. For instance, a signal-to-noise ratio of 90 or 100 decibels is considered high fidelity. In this paper, SNR

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{i=0}^{N-1} f^2(n)}{\sum_{i=0}^{N-1}\left(g(n) - f(n)\right)^2}\right) \tag{11}$$

is applied to evaluate the quality comparison between the attacked audio and original audio. Where *f(n)* is an original audio sample, and *g(n)* is an attacked audio sample. SNR value is getting larger, thus leading to better audio quality.

Pearson's correlation, denoted as $\rho(x,y)$,

$$\rho(x,y) = \frac{K\sum_{i=1}^{K} X_iY_i - (\sum_{i=1}^{K} X_i)(\sum_{i=1}^{K} Y_i)}{\sqrt{[K\sum_{i=1}^{K} X_i^2 - (\sum_{i=1}^{K} X_i)^2]\,[K\sum_{i=1}^{K} Y_i^2 - (\sum_{i=1}^{K} Y_i)^2]}} \tag{12}$$

is employed to represent correlation between two images where $\rho(x,y)$ is a correlation coefficient (CC) between *x* and *y*, *X* is an image 1, *Y* is an image 2, and *K* is the number of image bits.

All the audio signals used in this test are audio with 16 bits/sample, 44.1 KHz sample rate, and 15 s long. We take various audio data files with the most commonly related to copyright protection issue. Therefore, three types of audio, including classical (violin and bass), jazz (singer and band), and instrumental (solo piano, solo guitar), are used in the experiments. The watermark to be embedded is a

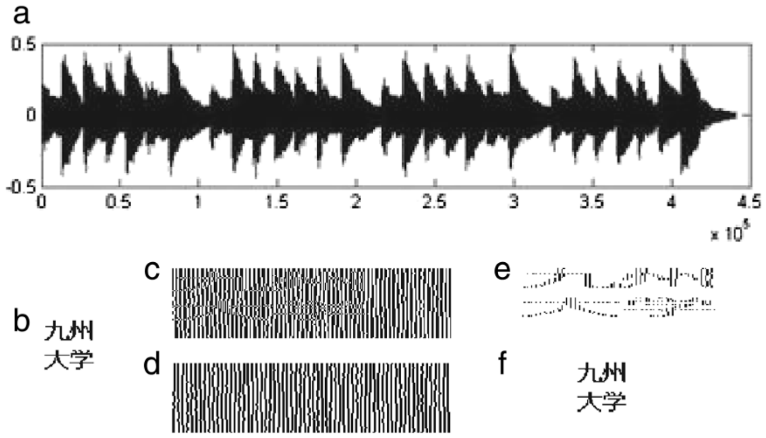


Figure 3 Watermark extraction result without being attacked. (a) Original audio signal, **(b)** original watermark, **(c)** secret share, **(d)** public share, **(e)** extracted watermark, and **(f)** reduced watermark.

visually recognizable binary image of size 64 × 64. Three-level wavelet decomposition is performed, and the frame length is 512 samples.

3.1 Watermark extraction

We first investigate our proposed scheme in recovering the watermark without being attacked. According to the experimental results described in Figure 3, BER and correlation coefficient values of all types of audio files are respectively 0% and 1. It demonstrates that each bit of watermark data is completely extracted and identical to the original one.

On the other hand, an erroneous condition is discovered in embedding phase of Chen and Zhu's scheme [4]. Consider the binary image watermark $W = \{w_{i,j} | w_{i,j} \epsilon \{0, 1\}, i = 0, \ldots, M-1; j = 0, \ldots, N-1\}$.

Table 1 BER and correlation coefficient of extracted watermark attacked by StirMark

Attacks	Instrumental		Jazz		Classical		CC in
	BER	CC	BER	CC	BER	CC	[4]
a. AddBrumm	1.90%	0.92	0.85%	0.96	2.24%	0.91	0.98
b. AddDynNoise	5.57%	0.80	2.25%	0.91	3.17%	0.87	-
c. AddNoise	5.25%	0.81	1.81%	0.92	3.66%	0.86	0.99
d. *AddSinus*	*0.098%*	*0.99*	*0.44%*	*0.98*	*1.17%*	*0.95*	*0.92*
e. Amplify	0%	1	0%	1	0%	1	1
f. BassBoost	2.39%	0.90	7.67%	0.75	2.05%	0.91	-
g. BitChanger	0%	1	0%	1	0%	1	-
h. *Compressor*	*0%*	*1*	*0.68%*	*0.97*	*0%*	*1*	*0.99*
i. Echo	26.44%	0.47	18.70%	0.55	23.34%	0.50	1
j. ExtraStereo	5.91%	0.80	1.54%	0.93	4.27%	0.84	-
k. FlippSample (2000)	33.86%	0.4	17.11%	0.57	27.61%	0.45	-
FlippSample (100)	10.84%	0.68	3.34%	0.87	8.54%	0.72	-
l. LSBZero	0%	1	0%	1	0%	1	-
m. NoiseMax	4.79%	0.82	1.61%	0.93	3.91%	0.85	-
n. RCHighPass	3.49%	0.86	8.94%	0.72	3.10%	0.87	-
o. RCLowPass	1.22%	0.95	0.51%	0.98	1.001%	0.96	1
p. ReplaceSamples	13.14%	0.64	5.03%	0.81	15.73%	0.59	-
q.Smooth	4.37%	0.83	0.81%	0.96	2.37%	0.90	0.99
r. Smooth2	3.10%	0.87	0.71%	0.97	2.51%	0.90	-
s. Stat1	4.76%	0.82	2.29%	0.90	4.42%	0.83	0.94
t. Stat2	1.68%	0.93	0.29%	0.99	1.05%	0.95	1
u. ZeroCross	3.78%	0.85	1.34%	0.94	3.83%	0.86	0.97
v. Resampling	0%	1	0%	1	0%	1	1

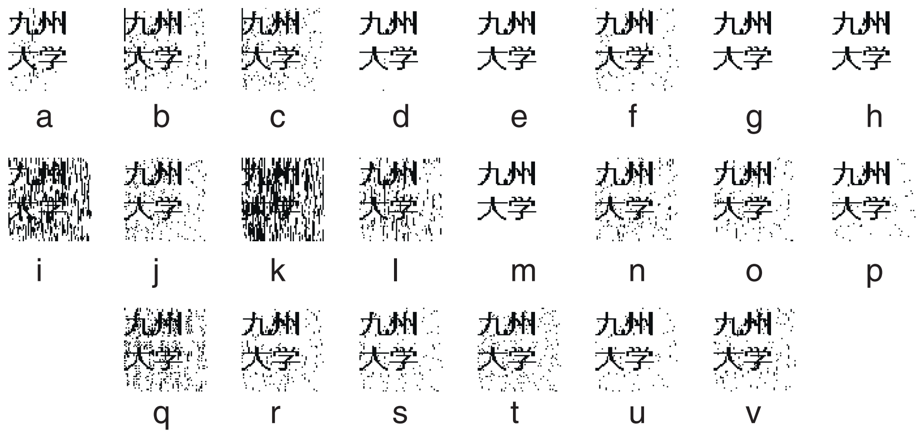


Figure 4 Examples of extracted watermark from attacked audio by StirMark. The music type of Instrumental is taken as an example. **(a to v)** The extracted watermark from StirMark attacks.

To generate watermark key, they first constructed binary pattern matrix $B = \{b_{t,p} | b_{t,p} \epsilon \{0, 1\}, t = 0, \ldots, T-1; p = 0, \ldots, P-1\}$ where T is the number of selected frame and P is the number of selected coefficient cumulants on all selected frame. Then, the watermark key K_3 was generated by performing XOR operation between binary pattern matrix B and image watermark W. Notice that matrix dimension of K_3 will be equal to B. It is reflected by the provided formula in [4] on how to find each pixel position in W that corresponds to B. In the extraction phase, the extracted watermark W' is revealed by conducting XOR operation between K_3 and B. The dimension between W' and W is different, thus causing the extracted watermark to be unrecognizable and unusable for verification purpose. To improve the problem, we simply utilize the entire of the obtained DWT-DCT coefficients rather than employ certai coefficients.

Table 2 Performance over various durations

Attacks	00:01:38		00:02:20		00:04:08	
	BER	CC	BER	CC	BER	CC
a. AddBrumm	2.44%	0.90	0.90%	0.96	3.88%	0.85
b. AddDynNoise	3.15%	0.87	3.22%	0.87	4.49%	0.83
c. AddNoise	6.61%	0.77	2.73%	0.89	7.91%	0.74
d. AddSinus	0.46%	0.98	0.31%	0.99	1.83%	0.92
e. Amplify	0%	1	0%	1	0.02%	0.99
f. BassBoost	5.81%	0.79	10.08%	0.69	13.75%	0.63
g. BitChanger	0%	1	0%	1	0%	1
h. Compressor	0%	1	0.6836%	0.9687	0%	1
i. Echo	18.43%	0.56	21.48%	0.52	24.80%	0.48
j. ExtraStereo	15.67%	0.60	17.26%	0.57	10.52%	0.68
k. FlippSample (2000)	21.34%	0.52	11.35%	0.67	30.44%	0.43
FlippSample (100)	5.98%	0.79	3.10%	0.87	8.30%	0.73
l. LSBZero	0%	1	0.02%	0.99	0.02%	0.99
m. NoiseMax	6.60%	0.77	1.78%	0.92	6.23%	0.78
n. RCHighPass	9.59%	0.70	17.79%	0.57	19.09%	0.55
o. RCLowPass	0.42%	0.98	0.27%	0.99	0.66%	0.97
p. ReplaceSamples	16.77%	0.58	6.13%	0.78	0%	1
q.Smooth	0.49%	0.98	0.32%	0.99	0.68%	0.97
r. Smooth2	0.85%	0.96	0.44%	0.98	1.05%	0.95
s. Stat1	1.83%	0.92	1.10%	0.95	2.10%	0.91
t. Stat2	0.37%	0.98	0.12%	0.99	0.29%	0.99
u. ZeroCross	8.74%	0.72	1.78%	0.92	8.50%	0.73

3.2 Robustness against incidental distortions

Incidental distortion refers to the distortions introduced from real applications which do not change the content of the multimedia data [17]. To evaluate the robustness to such distortions, the scheme is tested by performing various attacks of audio signal processing provided by StirMark for Audio (SMFA) version 1.03 [18] as well as exploiting their default values. The aim of SMFA is to delete, remove, or destroy the digital watermark by modifying the signal of the audio file. According to the Table 1, the minimum acceptable value of BER and CC are located on FlippSample attack, which are approximately 26.19% and 0.47, respectively, and the extracted watermark is still visually recognizable. This attack flips 2,000 samples every 10,000 with sample 6,000 ahead. However, when the attack only flips 100 samples, the average of BER and CC have both improved to approximately 7.57% and 0.76, respectively. Thus, it leads to assertion that in general the proposed scheme has a satisfactory performance against StirMark attacks, especially BitChanger, Compressor, and LSBZero as depicted in Figure 4.

The next attack conducted is downsampling generated by Cool Edit Pro 2.1. The sample of audio rate is adjusted from 44,100 to 22,050 Hz, and then, its sample rate is readjusted to 44,100 Hz. This process might cause an alteration in some parts of audio data. Consequently, the watermark data cannot be completely extracted. However, the BER and correlation coefficient value as shown in Table 1, which are 0% and 1, respectively, indicate that the proposed scheme resists to such attack.

To evaluate the robustness of proposed scheme, we draw a comparison to earlier method [4] subjected to StirMark attacks as well as short duration. In more detail, all the audio signals used in [4] were audio with 16 bits/sample, 44.1 KHz sample rate, and 28.73 s long. The music styles used throughout their experiment were not explicitly reported. In order to properly compare the schemes, we deliberately exploit various music styles. We expect the music styles used in [4] to be any of ours. BER and correlation coefficient values are reported in Table 1, and the extracted watermark against those attacks is illustrated in Figure 4. The results indicate that our proposed scheme outperforms Chen's scheme [4] on AddSinus and Compressor attacks. In case of other attacks, we still achieve considerable results compared to Chen's scheme.

Furthermore, to verify the efficacy of the proposed scheme, evaluation over various durations is conducted as well. The duration is ranging from 1 to 4 min. However, we did not perform comparative experiments because the duration either in [4] or [5] is approximately below 60 s. The experimental results against SMFA are reported in Table 2. In general, the findings show that longer duration provides fairly the same performance as short duration. For example, BitChanger attack indicates exactly the same results, while amplify and LSBZero attacks demonstrate that the number of error bit is only one. To confirm the findings, the resulting extracted watermarks are provided in Figure 5.

3.3 Robustness against intentional distortions

Intentional distortion refers to distortions conducted by deliberately modifying the host content [17]. It can be performed by overwriting or removing the watermark. In the following subsection, we address two types of intentional distortions: *counterfeit attack* and *multiple claims situation*.

3.3.1 Counterfeit attack

In some cases, the adversary tries to confuse ownership by creating a faked original or faked watermarked audio. In this case, an adversary performs a distortion by modification of a set of features of received audio A' so-called faked original audio A^f. By doing so, it is expected that the original watermark will be destroyed. One simple way to alter the features is to modify the sample data in such a way that the SNR is still acceptable. Figure 6 demonstrates spectrogram of original audio signal and its faked version due to sample data alteration. The vertical axis represents frequencies up to 20,000 Hz, the horizontal axis shows positive time toward the right, and the colors represent the most important acoustic peaks for a given time frame, with red representing the highest energies, then in decreasing order of importance, orange, yellow, green, cyan, blue, and magenta.

Once the faked signal is constructed, the adversary may embed his watermark onto it and produce another watermarked audio. In the verification phase, the adversary's audio signal is verified by using registered secret share image. As shown in Table 3, the number of error bits is approximately in ranges 49 to 118 bits from 4,096 bits, and the owner's watermark is completely extracted. It indicates that the proposed scheme performs well in

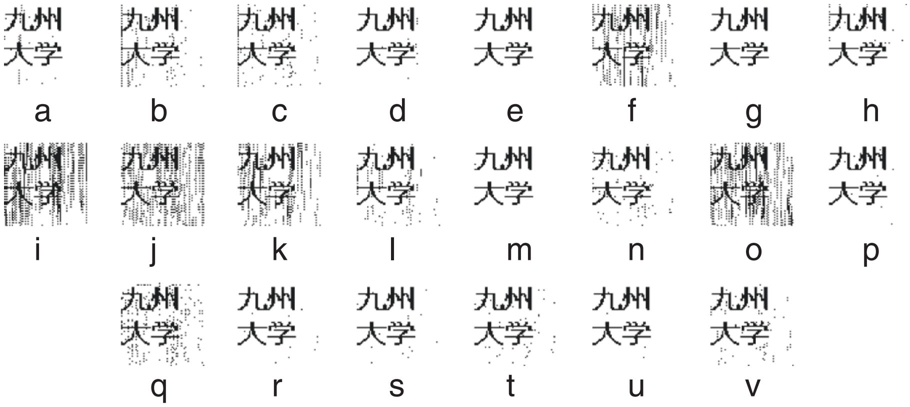


Figure 5 Examples of extracted watermark over various durations. We take audio with duration 00:02:20 as a sample. **(a to v)** The extracted watermark from StirMark attacks reported in Table 2.

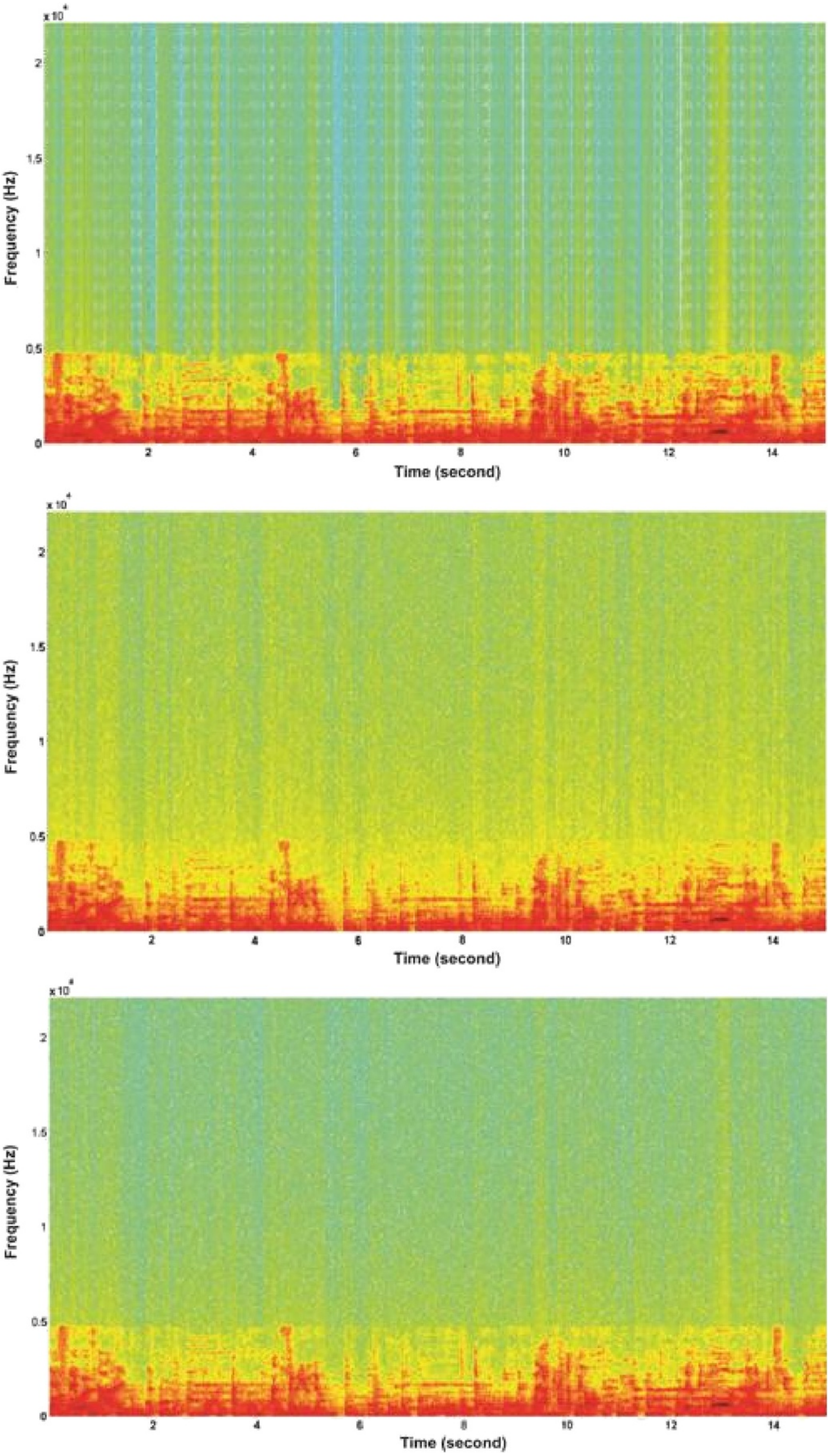
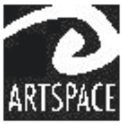
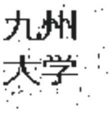
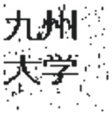


Figure 6 Spectrogram of Jazz and its faked signals due to intentional distortion. (Top) Original audio signal. (Middle) Faked original signal with SNR = 20.9949 dB. (Bottom) Faked original signal with SNR = 27.0155 dB. The figure is intended for color reproduction on the Web and in print.

Table 3 Watermark extraction performance against intentional distortion

Adversary's watermark	Extracted watermark	SNR	BER	CC
		26.2626 dB	1.1963%	0.9467
		20.242 dB	2.8809%	0.8824

watermark verification phase and possesses an unambiguous property.

3.3.2 *Multiple claims*

In this situation, the adversary attempts to provoke a dispute by embedding his/her own message. The following is the model of the proposed scheme. In such a scheme, let $\mathbf{x} = (x(1)\ldots x(N_f))^T$ be a feature vector extracted from the audio content with length-N_f. The message to be hidden is a binary matrix $\mathcal{W}$ of size $N \times N$. The scheme exploits (2,2)-secret sharing. The codebook C comprises two $2 \times n$ boolean matrices $(\mathcal{C}_i^0, \mathcal{C}_i^1)$ with:

- $i = (1\ldots f), f$ is the number of feature type.
- $\mathcal{C}_i^0$ and $\mathcal{C}_i^1$ are the base matrices for black and white pixel, respectively.

The scheme is defined as the four-tuple $(\mathcal{W}, \mathcal{E}, \mathcal{D}, C)$, where:

- $\mathcal{E}$: $\mathbf{x}\times\mathcal{W} \times C \to \mathcal{S}$ is the encoder mapping a sequence $\mathbf{x}$, a hidden message $\mathcal{W}$ using codebook C to a secret share image $\mathcal{S}$.
- $\mathcal{D}$: $\mathbf{x}\times C \to \mathcal{P}$ is the decoder mapping a sequence $\mathbf{x}$ using codebook C to a public share image $\mathcal{P}$.

According to our scheme, $\mathcal{S}$ is kept by CA while $\mathcal{P}$ as well as codebook are publicly known. Suppose the adversary intends to rewrite the content with his hidden message. We would like to show that all his efforts are fairly unworthy.

Suppose $\mathbf{x}^*$, C^*, and $\mathcal{W}^*$ are the feature vector extracted from the retrieved audio content, adversary's codebook, and adversary's hidden message, respectively. Based on aforementioned statement, we might convey that $C^* \equiv C$ such that

- $\mathcal{E}$: $\mathbf{x}^*\times\mathcal{W}^* \times C^* \to \{\mathcal{S}^*, \mathcal{P}^*\}$ where $\mathcal{S}^*$ is the adversary's secret share. Note that $\mathcal{S}^*$ is not required since the original $\mathcal{S}$ have been registered by the owner in advance.
- $\mathcal{D}$: $\mathbf{x}^*\times C^* \to \mathcal{P}^*$ where $\mathcal{P}^*$ is the adversary's public share. Due to the property of our scheme, it is obvious that $\mathbf{x}^* \equiv \mathbf{x}$ which implies that $\mathcal{P}^* \equiv \mathcal{P}$. Thus the adversary's hidden message will never be extracted. □

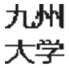


Figure 7 The example of extracted watermark of multiple claims condition.

To evoke multiple claims situation, the adversary embeds his watermark, which is depicted in Table 3, onto the $\mathbf{x}^*$. Figure 7 shows that original's watermark remains extracted.

4 Conclusions

This paper investigates the problem of constructing an audio ownership protection scheme in order to resist against both intentional and incidental distortions. To achieve these goals, we have integrated wavelet transform, visual cryptography, and digital timestamp into an ownership protection scheme. The trade-off between data payload and two other properties, imperceptibility and/or robustness, can be reduced, while preserving its audio signal quality. According to experimental results, the proposed scheme fulfills several properties of ownership protection including perceptual transparency, blindness, robustness, security, and unambiguousness. In terms of security, it is achieved by means of visual cryptography method. Without possessing both shares, it is infeasible for anyone to retrieve the secret image from each share. The integrity of codebook and its secret share image is guaranteed by certification authority through timestamp mechanism. It indicates that audio ownership protection can take advantage from the combination of visual cryptography and watermarking and proposed scheme can be widely applied to the area of audio ownership protection.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
The authors wish to thank the anonymous reviewers for their very constructive and helpful comments. Research support for the first author was provided by the Directorate General of Higher Education, Ministry of National Education, Indonesia. The second author acknowledges support provided by Grant NRF-2011-013-D00121 from the National Research Foundation of Korea.
This is an expanded version of a paper [6] presented at the Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2011) with further analysis and some new simulation experiments.

Author details
[1]Department of Informatics, Telkom University, Bandung 40257, Indonesia. [2]Graduate School of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan. [3]Department of IT

Convergence and Application Engineering, Pukyong National University, 599-1, Daeyeon 3-Dong, Nam-Gu, Busan 608-737, Korea.

Received: 6 August 2012 Accepted: 18 December 2013
Published: 4 January 2014

References

1. S Stokes, *Digital Copyright: Law and Practice*, 2nd edn. (Hart Publishing, New York, 2005), p. 120
2. X-Y Wang, Y-R Cui, H-Y Yang, H Zhao, A new content-based digital audio watermarking algorithm for copyright protection, in *Proceedings of the 3rd International Conference on Information Security (SEC)*, vol. 85 (ACM, New York, 2004), pp. 62–68
3. M Barni, F Bartolini, *Watermarking systems engineering: enabling digital assets security and other applications,* (Marcel Decker, New York, USA, 2004), pp. 6–11
4. N Chen, J Zhu, A robust zero-watermarking algorithm for audio. EURASIP J. Adv. Signal Process. **2008**, 453580 (2008)
5. R Wang, W Hu, Robust audio zero-watermark based on LWT and chaotic modulation, in *Proceeding International Workshop Digital Watermarking (IWDW),* (Springer, Heidelberg, 2007), pp. 373–381
6. RW Ciptasari, A Fajar, FA Yulianto, K Sakurai, An efficient key generation method in zero-watermarking for audio, in *Proceeding of 7th IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP),* (IEEE, Washington DC, 2011), pp. 336–339
7. Z Wang, CC Chang, HN Tu, MC Li, Sharing a secret image in binary images with verifcation. J. Inform. Hiding Multimedia Signal Process. **2**(1) (2011)
8. CC Chang, JC Chuang, An image intellectual property protection scheme for gray-level images using visual secret sharing strategy. Pattern Recognit. Lett. **23**, 931–941 (2001)
9. SL Hsieh, LY Hsu, IJ Tsai, A copyright protection scheme for color images using secret sharing and wavelet transform, in *Proceedings of World Academy of Science, Engineering And Technology,* (World Academy of Science, Engineering and Technology, Egypt, 2005), pp. 17–23
10. WB Lee, TH Chen, A public verifiable copy protection technique for still images. J Syst. Software, 195–204 (2002)
11. DC Lou, HK Tso, JL Liu, A copyright protection scheme for digital images using visual cryptography technique. J. Comput. Stand. Interfaces **29**, 125–131 (2006)
12. TH Chen, GB Horng, WB Lee, A publicly verifiable copyright-proving scheme resistant to malicious attacks. IEEE Trans. Ind. Electron. **52** (1), (2005)
13. IJ Cox, ML Miller, JA Bloom, *Digital Watermarking* (Morgan Kauffman Publisher, San Francisco, 2002), pp. 45–47
14. N Naor, A Shamir, Visual cryptography. Advances in Cryptology: Eurocrypt. **94**, 1–12 (1995)
15. G Voyatzis, I Pitas, Protecting digital image copyrights: a framework. IEEE Comput. Graph. Appl. **19**(1), 18–24 (1999)
16. Electronic Time-stamping. https://www.digistamp.com/technical/how-a-digital-time-stamp-works/. Accessed 25 Oct 2011.
17. D He, Q Sun, Multimedia authentication, in *Multimedia Security Technologies for Digital Rights Management*, ed. by W Zeng, H Yu, and CY Lin (Academic Press, London, 2006), pp. 111-138
18. A Lang, StirMark benchmark for audio, http://sourceforge.net/projects/stirmark/files/. Accessed 17 November 2011

doi:10.1186/1687-417X-2014-2
Cite this article as: Ciptasari *et al.*: An enhanced audio ownership protection scheme based on visual cryptography. *EURASIP Journal on Information Security* 2014 **2014**:2.