An Approach towards Robustness of Digital Audio Watermarking

Aye Thinzar Myint
University of Computer Studies, Yangon
ayethinzarmyint.4th@gmail.com

Abstract

Digital Watermarking has been proposed as a mean to identify the owner or distributor of digital data. It involves the concealment of data within a discrete audio file. Audio Watermarking has been proposed as a possible solution, embeds copyright information into audio files as a proof of their ownership. In this paper, we propose an effective, robust and inaudible Audio Watermarking algorithm. Audio Watermarking scheme has been carried out where a visually recognizable binary image is used as a watermark. The effectiveness of the algorithm is brought by applying a cascade of two powerful mathematical transforms: the discrete wavelet transform (DWT) and the singular value decomposition (SVD). In the proposed algorithm, watermark bits are not embedded directly into DWT's coefficient, but rather on the elements of singular values of the DWT sub-bands of the audio frames. This system is implemented using MATLAB programming language and validated through a set of music records. Experimental results are presented in this paper to demonstrate the effectiveness of the proposed algorithm.

1. Introduction

During the past few years, digital multimedia technology and communication network have made great progress and they are now becoming increasingly important in daily life. Consequently, intellectual property protection is a pressing concern for content owners who are exhibiting digital representation of the photographs, music, video and original artworks through the Internet. The large number of high-speed internet users and the availability of audio recording and editing software and devices, such as CD-writers and mp3 player, have put the music industry in a critical situation. The industry's intellectual properties are becoming in danger of being attacked by so called "music pirates".

Digital watermarking is one of the most popular approaches considered as a tool for providing the copyright protection of digital contents. This technique is based on direct embedding of additional information data (called watermark) into the digital contents. Ideally, there must be no perceptible difference between the watermarked and original digital contents, and the watermark should be easily extractable, reliable and robust against data compression or any signal manipulations [4]. According to the International Federation of the Phonographic Industry (IFPI), audio watermarking should have the following specifications: 1) Audio watermarking should not degrade perception of original signal. 2) Signal to noise ratio (SNR) should be greater than 20 dB and there should be more than 20 bits per-second (bps) data payload for watermark. 3) Watermark should be able to resist most common audio processing operations and attacks. 4) Watermark should be able to prevent unauthorized detection, removal and embedding, unless the quality of audio becomes very poor.

Watermarking is mainly used for copy-protection and copyright protection. Generally, it can be classified as non-blind watermarking and blind watermarking. In non-blind schemes, the cover signal (the original signal) is needed during the detection process to detect the mark with the match of the original. In blind schemes, the original signal is not required. Solely the key, which is typically used to generate some random sequence used during the embedding process, is required.

2. Related Work

Human Auditory System (HAS) is much more sensitive than Human Visual System (HVS) and that in audibility is much more difficult than to achieve than invisibility for images. Audio watermarking classes are: Frequency Domain Audio Watermarking employs human perceptual properties and frequency masking characteristic of the HAS system for watermarking. These techniques usually use DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), and DWT (Discrete Wavelet Transform). A method which uses MPEG 1 Layer 3 compression to determine where and how the embedded watermark is introduced in [6]. In the time-domain watermarking techniques, watermark is directly embedded into audio signal. No domain transform is required in this process [5]. Compressed-Domain techniques operate on compressed, rather than uncompressed audio for watermarking. The watermark is embedded directly into already compressed audio bit stream to prevent the watermark from being removed by compression.

MPEG audio compression is a lossy algorithm and uses the special nature of the HAS. A digital watermarking scheme for mpeg audio Layer 3 audio files that operates directly in the compressed data while manipulating the time and sub-band/channel domain is presented in [3]. Wavelet-Domain can be used to decompose a signal into two parts, high frequencies and low frequencies. Embedding and extraction of watermarks are performed in randomly selected perceptually significant sub-bands and perceptual transparency is obtained using frequency masking property [2]. In this paper, we propose an audio watermarking algorithm that satisfies the requirements of effective audio watermarking; inaudibility and watermark robustness to removal or degradation. The requirements were met by the proposed algorithm by exploiting the attractive properties of two powerful mathematical transforms; the Discrete Wavelet Transform (DWT) and the Singular Value Decomposition (SVD). In the proposed algorithm, watermark bits are not embedded directly into DWT's coefficients, but rather on the elements of singular values of the DWT sub-bands of the audio frames.

3. Background

In this section, we briefly introduce the DWT and SVD transforms, and outline their relevance to the problem of digital watermarking.

3.1. The Discrete Wavelet Transform

The DWT is computed by successive low-pass and high-pass filtering of the discrete time-domain signal as shown in Figure1. Its significance is in the manner it connects the continuous-time multi-resolution to discrete-time filters. In the Figure 1, the signal is denoted by the sequence $x[n]$, where n is an integer. The low pass filter is denoted by G_0 while the high pass filter is denoted by H_0. At each level, the high-pass filter produces detail information; $d[n]$, while the low pass filter associated with scaling function produces coarse approximations, $a[n]$.

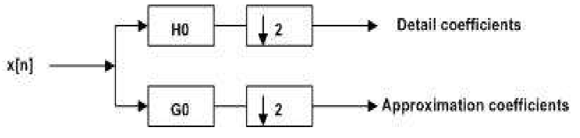

Figure 1. One-level DWT Decomposition

$$H(\omega)=\sum_k h_k \exp(-jk\omega)(high-pass) \quad (1)$$

$$G(\omega)=\sum_k g_k \exp(-jk\omega)(low-pass) \quad (2)$$

The iterative process of the decomposition through high-pass filter H and low-pass filter G to obtain coarse coefficients c and detail coefficients d is,

$$c_{j-1,k}=\sum_n h_n - 2k^{c_{j,n}} \quad (3)$$

$$d_{j-1,k}=\sum_n g_n - 2k^{c_{j,n}} \quad (4)$$

Similarly, the iterative process of the reconstruction is,

$$c_{j,n}=\sum_n h_n - 2k^{c_{j,n}} + \sum_n g_n - 2k^{c_{j,n}} \quad (5)$$

Depending on the application and length of the signal, the low frequencies part might be further decomposed into two parts of high and low frequencies. The original signal **S** can be reconstructed using inverse DWT. Due to its excellent spatial-frequency localization properties, the DWT is very suitable to identify area in an audio signal where a watermark can be embedded effectively. DWT is also applied in speech compression which reduces transmission time in mobile applications [9] and also used for real time audio and video Frequency Multiplexing [8].

3.2 The Singular Value Decomposition

The singular value decomposition is a powerful technique in many matrix computations and analyses. Using the SVD of a matrix in computations, rather than the original matrix, has the advantage of being more robust to numerical error. Additionally, the SVD exposes the geometric structure of a matrix, an important aspect of many matrix calculations. A matrix can be described as a transformation from one vector space to another. The components of the SVD quality the resulting change between the underlying geometry of those vector spaces.

Singular Value Decomposition is said to be a significant topic in linear algebra by many renowned mathematicians. SVD has many practical and theoretical values, other than image compression. One special feature of SVD is that it can be performed on any real ***(m, n)*** matrix. It factors A into three matrices ***U, S, V*** such that, ***$A=USV^T$*** where ***U*** and ***V*** are orthogonal matrices and **S** is a diagonal matrix. The SVD is employed in a variety of applications, from least-squares problems to solving systems of linear equations. The SVD of an ***NxN*** matrix **A** is defined by the operation as in (6).

$$\begin{bmatrix} V_{1,1} & \cdots & \cdots & V_{1,n} \\ V_{2,1} & \cdots & \cdots & V_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ V_{n,1} & \cdots & \cdots & V_{n,n} \end{bmatrix}^T \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \sigma_{nn} \end{bmatrix} \begin{bmatrix} U_{1,1} & \cdots & \cdots & U_{1,n} \\ U_{2,1} & \cdots & \cdots & U_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ U_{n,1} & \cdots & \cdots & U_{n,n} \end{bmatrix}$$

Figure 2. The SVD operation

$$SVD(A)=USV^T \quad (6)$$

The diagonal entries of S are called singular value decomposition of **A** and assumed to be arranged in decreasing order **σ>σ+1.**The columns of **U** matrix are called left singular vector of **A.** The columns of **V** matrix are called right singular vector of **A**. SVD-based audio watermarking algorithm add the watermark information to the singular value of the diagonal matrix **S** in such away to meet the imperceptibility and robustness requirements of effective digital audio watermarking. SVD is used for solving many approximation problems in model reduction [1], data compression, system identification and signal processing [7].

4. Proposed Method

In this paper, an audio watermarking system is developed in DWT domain. The embedding algorithm is performed to the wavelet coefficient obtained from 4-level wavelet decomposition. As the watermarks, binary images of 32×32 are used. The proposed algorithm employs a cascade of two transforms; the discrete wavelet transform and singular value decomposition transform.

4.1 Watermark embedding procedure

The watermark embedding algorithm is as follows:

1. Firstly, the binary image watermark is converted into one dimensional vector.
2. The original audio signal should be sampled at a rate of 44100 samples per second.
3. The sampled audio file is segmented into frames of 10,000 samples.
4. A four-level DWT transformation is carried out on each partitioned frame producing five multi-resolution sub-bands: detail sub-bands, ***D1, D2, D3, D4*** and approximation sub-bands, ***A4.***
5. Embedding process could then be performed in detail sub-bands by arranging the above four detail sub-bands into a matrix form using a matrix formulation as follows.

<table>
<tr><td colspan="8">D1</td></tr>
<tr><td colspan="4">D2</td><td colspan="4">D2</td></tr>
<tr><td colspan="2">D3</td><td colspan="2">D3</td><td colspan="2">D3</td><td colspan="2">D3</td></tr>
<tr><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td></tr>
</table>

Figure 3. Matrix Formation of the details sub-bands

6. Using Singular Value Decomposition on the formed matrix, three orthogonal matrixes: ***S, U,*** V^T is produced. ***S*** is a 4×4 diagonal matrix. *S* (1, 1) is used for embedding the watermark and it is needed to store for watermark extraction procedure.
7. The watermark bits of binary image are embedded into the DWT-SVD-transformed audio signal using (7).

$$S(1,1)_w = S(1,1)\times 1+\alpha\times w(n)) \qquad (7)$$

where *w(n)* = watermarked bits
S (1, 1) = top left value in the *S* matrix
$S(1,1)_w$ = watermarked *S* (1, 1)
α = small constant

8. Apply inverse SVD operation on ***U,*** V^T obtained from Step 6 and new ***S*** matrix which has been produced in Step 7 to obtain watermarked matrix.
9. Finally, wavelet reconstruction is conducted using inverse DWT for each watermarked audio frame. Summing all watermarked frames, the overall watermarked audio signal is obtained.

4.2 Watermark Extraction Procedure

The watermarked audio signal and the singular value on each frame are required in watermark extraction process.

1. Extraction process needs to perform step 1 to 6 as in embedding process until new ***S*** matrix is obtained from all frames of watermarked audio.
2. Once new ***S*** matrix or new singular value is obtained, the comparison of this S matrix with corresponding original singular value is needed to extract watermark bits. In the extraction, if the S(1,1) of new singular value matrix is less than original S(1,1) of the S matrix, the extracted watermark bit is assigned as ‘0’ otherwise the extracted watermark bit for the particular frame is ‘1’.
3. Rearrange the extracted bits from the individual frames and the original binary image can be reconstructed.

5. Performance Evaluation

In measuring performance, watermarking algorithms are usually evaluated with fidelity, imperceptibility and robustness. In this evaluation, music signals of length about 4 minutes are needed to cover the watermark bits of 1024. The four audio genres are used: Pop Music, Rock Music, Classical Music and Instrumental Music.

For imperceptibility, watermarked songs (music) or audio could be listened to determine the performance. For robustness issues, we designed the experiments with various attacks such as - zeroing, noise adding, cropping and duplicating.

6. Experimental Results

In evaluating this implemented watermark algorithm, the similarity between extracted watermark and original watermark is used. The similarity is defined as

$$Sim = \frac{\sum_{i=1}^{n} W(i)W'(i)}{\sqrt{\sum_{i=1}^{n} W(i)^2}\sqrt{\sum_{i=1}^{n} W'(i)^2}} \qquad (8)$$

where *W* and *W'* are original watermark and extracted watermark, *i* is the index of binary watermark sequence and *n* is number of the watermark sequence bits. In order to examine the effectiveness of the presented algorithm, the experiments were performed in two ways: the similarity test and the robustness test.

6.1 The Similarity Test

In these experiments, 20 music files were grouped into four genres. Each music file was tested with 3 watermark images. There is no error in conducting the experiments with pop and rock music genres so that similarity values are exactly 1. Hence the original watermark information can be accurately recovered in these cases. But there exist differences between the extracted watermark and original one when the similarity test was performed on classical music and instrumental music. The resulted average correctness is 0.9994 for classical genre and 0.9992 for instrumental genre, respectively.

6.2 The Robustness Test

The robustness test was conducted under four different conditions. In this experimental framework, each of the music pieces employed in the similarity test were again accessed with three different watermark images.

Zeroing Attack

In this attack, the samples located at multiples of 20^{th} or 40^{th} or 60^{th} of the whole signal were replaced with zeros. When zeroing attacks are tested on the audio signals, the difference between the original and extracted watermarks is of small amount as shown in Table.1.

Table 1: Performance on Zeroing Attack

No	Songs	Similarity		
	zero position	20th	40^{th}	60th
1.	Pop	0.9735	0.9719	0.9781
2.	Rock	0.9885	0.9833	0.9915
3.	Classical	0.9910	0.9892	0.9930
4.	Instrumental	0.9556	0.9516	0.9619

Noise Attack

White noise value of 15dB and 20dB SNRs were added to the watermarked audio signals with the following formula.

$$SNR = 20\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \qquad (9)$$

P_{signal} = power of the original signal
P_{noise}=power of the noise added to signal

According to the result data shown in Table 2, the recovered watermarks have average similarity values of 0.8406 for 4 audio genres when adding the noise values at 15dB. However, after adding white noise at 20dB SNR, the percentage of similarity is almost 90%.

Table 2: Performance on Noise Attack

No	Songs	Similarity	
	SNR	15dB	20dB
1.	Pop	0.8510	0.9063
2.	Rock	0.8419	0.9295
3.	Classical	0.8369	0.9087
4.	Instrumental	0.8324	0.8779

Cropping Attack

In cropping attack, three cropping percentage was used. The last 10%, 20%, 50% samples of every frame were cropped so the first samples of the each frame move towards the previous frame. Table 3 shows the resemblance between original and detected watermarks on each genre group. It is obvious that this method is robust against cropping attack except for high cropping percentage of 50%.

Table 3: Performance on Cropping Attack

No	Songs	Similarity		
	crop percentage	10%	20%	50%
1.	Pop	0.9255	0.9006	0.7253
2.	Rock	0.9131	0.8914	0.8215
3.	Classical	0.8985	0.8995	0.7271
4.	Instrumental	0.9322	0.9021	0.7264

Duplicating Attack

The samples located at multiples of every 20, 40 and 60 of the whole signal were replaced by previous sample in this duplicating test. Most of the extracted watermarks are similar with the original watermark images with similarity values greater than 0.9 as listed in Table 4.

Table 4: Performance on Duplicating Attack

No	Songs	Similarity		
	duplicate position	20th	40th	60th
1.	Pop	0.9941	0.9971	0.9967
2.	Rock	0.9988	0.9993	0.9994
3.	Classical	0.9988	0.9992	0.9983
4.	Instrumental	0.9914	0.9978	0.9970

7. Discussions

The results of robustness against attacks show that the proposed audio watermarking scheme needs some improvements to get better robustness, especially in noise and cropping attacks. When zeroing and duplicating attacks are tested on the audio signals, it is found that the difference between similarity measurement before and after attack is trivial and less sensitive to this attack. When zeros are replaced in watermarked audio, similarity value is always greater than 0.95 whereas for duplicating attacks, at least 96% of extracted watermark bits are the same as that of the original watermark. For all tested genres of audio, there exists degradation in perceptibility of the embedded watermark audio in instrumental music. In cropping attack, similarity values gradually decreases to 0.7 when cropping level reaches to 50%. Among the tested distortion experiments, the proposed system is most robust as well as imperceptible to duplicating attacks. However, it is noted that the similarities could vary according to the embedded watermark images and host audio signals.

8. Conclusions

An imperceptible and robust audio watermarking technique based on cascading the discrete wavelet transform (DWT) and the singular value decomposition (SVD) is presented. The elements of singular value of the DWT sub-bands of the audio frame are used to embed the watermark bits. The proposed scheme is designed as a non-blind watermarking scheme as they are more secure than the blind techniques. However, only one bit of the original singular value is needed in the watermark extraction process. As objectives tests, four attacks are performed on the watermarked audio signals. The results indicate that the proposed system is most resistant to duplicating attack while cropping is least robust. Among all test genres, rock music is most inaudible and best imperceptible in watermark embedding as opposed to instrumental music. We can verify the effectiveness of this audio watermarking as a reliable solution to the copyright protection which is facing the music industry.

References

[1] F. V. Belzen, S. Weiland and J. D. Graaf, "Singular value decompositions and low rank approximations of multi-linear functional", *Proceedings of the 46th IEEE Conference on Decision and Control New Orleans, LA, USA, Dec.12-14, 2007*, pp.3751-3756, 2007.

[2] N. Cvejic, T. Seppanen, "Robust Audio Watermarking in Wavelet Domain Using Frequency Hopping and Patchwork Method", *Proceedings of the 3rd International Symposium on Image and Signal Proceeding and Analysis,* pp.251-255, 2003.

[3] D. K. Koukopoulos, Y. C. Stamatiou, "A Compressed-Domain Watermarking Algorithm for MPEG Audio Layer 3" *Multimedia and Security Ottawa Canada,* pp.7-10, 2001.

[4] P. Kumasawat, K. Attakitmongol, A. Srikaew. "Digital Audio Watermarking for copyright protection based on Multi-wavelet Transform", *EUROISI 2008, LNCS 5376*, pp.155-164, 2008.

[5] W. N. Lie, L. C. Chang,"Robust and High-quality Time-domain audio watermarking based on Low Frequency amplitude modification" *IEEE on Transactions Multimedia*, vol. 8, pp. 46-59, 2006.

[6] D. Megias, J. H. Joancomarti. J. Manguillon, "A Robust- Frequency Domain Audio Watermarking Scheme for Monophonic and Stereophonic PCM Formats", *Proceedings of the 30th EUROMICRO Conference,* pp.449-452, 2004.

[7] H. Özer & B. Sankur, "An SVD based audio watermarking technique", *In Proceedings of the IEEE 13th conference on signal processing and communications applications*, (pp. 452-455), 2005.

[9] Y. J. Wu, S. Shimamoto, "A Study on DWT-Based Digital Audio Watermarking for Mobile Ad Hoc Network", *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC'06),* pp. 1-5, 2006.

[8] S. A M. A. Taweel and P.Sumari, "Digital Video Watermarking in the Discrete Wavelet Transform Domain

", *2009 Sixth International Conference on Computer Graphics, Imaging and Visualization,* pp.133-137,2009.