Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jisa

Blind SVD-based audio watermarking using entropy and log-polar transformation

*Pranab Kumar Dhar**, *Tetsuya Shimamura*

Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan

ARTICLE INFO

Article history:
Available online 11 November 2014

Keywords:
Audio watermarking
Discrete cosine transform
Entropy
Log-polar transformation
Singular value decomposition

ABSTRACT

This paper proposes a blind singular value decomposition (SVD) based audio watermarking scheme using entropy and log-polar transformation (LPT) for copyright protection of audio signal. In our proposed scheme, initially the original audio is segmented into non-overlapping frames and discrete cosine transform (DCT) is applied to each frame. Low frequency DCT coefficients are divided into sub band and entropy of each sub band is calculated. Watermark data is embedded into the Cartesian components of the largest singular value obtained from the DCT sub band with highest entropy value of each frame by quantization. Simulation results indicate that the hidden watermark data is robust against different attacks. The comparison analysis shows that the proposed scheme has high data payload and provides superior performance compared to the state-of-the-art watermarking schemes reported recently.

1. Introduction

In recent years, the rapid growth of multimedia technology has greatly facilitated the transmission and distribution of digital contents over the internet. As a result, the protection of the intellectual property rights of digital media has become an important issue. Digital watermarking has been used extensively for digital right management of multimedia data. It is a process of embedding watermark into the multimedia data to show authenticity and ownership. It has various applications such as copyright protection, data authentication, privacy protection, fingerprinting, broadcast monitoring, and so on. It must successfully satisfy the trade-off among the conflicting requirements of imperceptibility, robustness, and data payload.

A comprehensive survey on watermarking can be found in Cox and Miller (2002) and Cvejic and Seppanen (2007). Most of the watermarking algorithms proposed over the last few years mainly focus on image and video watermarking (Agresteand and Andaloro, 2008; Tsolis et al., 2009; Chan et al., 2005; Noorkami and Mersereau, 2008). Audio watermarking is more challenging than image and video watermarking, because the human auditory system (HAS) is significantly more sensitive than the human visual system (HVS). Most audio watermarking methods utilize either a time domain (Swanson et al., 1998; Lie and Chang, 2006) or a transform domain such as discrete wavelet transform (DWT) (Chen et al., 2013, 2010a, 2010b), lifting wavelet transform (LWT) (Erçelebi and Batakçı, 2009), and fast Fourier transform (FFT) (Megías et al., 2010). Time domain methods are very efficient and easy to implement, however, transform domain methods can provide high robustness. Swanson et al. (Swanson et al., 1998) proposed a watermarking scheme that embeds watermark bits by modifying the audio samples directly. Lie and Chang (Lie and Chang, 2006) introduced a method in which group

* *Corresponding author.*
E-mail addresses: pranab@sie.ics.saitama-u.ac.jp (P.K. Dhar), shima@sie.ics.saitama-u.ac.jp (T. Shimamura).
http://dx.doi.org/10.1016/j.jisa.2014.10.007
2214-2126/

amplitudes are modified to achieve high robustness. However, both methods have low data payload. In Chen et al. (2013), authors presented an adaptive method using wavelet based entropy, but robustness to resampling and low-pass filtering attacks are quite low. Chen et al. (Chen et al., 2010a) proposed an algorithm that embeds watermark information by energy-proportion scheme. However, the SNR results of this algorithm are not satisfactory. In Chen et al. (2010b), authors introduced an optimization-based watermarking scheme which embeds watermark in the lowest-frequency coefficients of DWT. However, the subjective evaluation of watermarked audio signals has not been conducted in this scheme. Erçelebi and Batakçı (2009) proposed a watermarking method based on LWT in which a binary image is embedded as watermark. However, from the reported result, robustness to attacks of this method is quite low. Megías et al. (2010) suggested a watermarking method that embeds watermark in FFT domain, but it has low data payload. Recently, the singular value decomposition (SVD) has been used as an effective technique in digital watermarking (El-Samie, 2009; Al-Nuaimy et al., 2011; Ali and Ahmad, 2010; Bhat et al., 2010; Lei et al., 2011). El-Samie (2009) and Al-Nuaimy et al. (2011) proposed an efficient SVD based audio watermarking algorithm which is domain adaptive. Moreover, the proposed segment-by-segment approach enhanced the detectability compared to the simple approach utilizing the whole original audio signal directly. However, the detection scheme is non-blind and robustness needs further improvement. In Ali and Ahmad (2010), authors proposed a method based on DWT and SVD. But this method is also non-blind and has low data payload. The most recent SVD based blind audio watermarking methods are proposed by Bhat et al. (2010) and Lei et al. (2011). These methods provide high robustness; however the data payload of these methods is quite low. Moreover, some other techniques such as empirical mode decomposition (EMD) (Khaldi and Boudraa, 2013), time spread (TS) echo method (Xiang et al., 2011; Erfani and Siahpoush, 2009), and audio histogram technique (Xiang and Huang, 2007; Xiang et al., 2008) are becoming popular in audio watermarking field. The main limitation of the existing audio watermarking techniques is the difficulty to obtain a favorable trade-off among imperceptibility, robustness, and data payload. To overcome this limitation, in this paper, we propose a blind SVD-based audio watermarking scheme in DCT domain using entropy and log-polar transformation (LPT). The main features of the proposed scheme include (*i*) it utilizes the entropy, DCT, SVD, LPT, and quantization jointly, (*ii*) it uses a tent map which contains the chaotic characteristic to enhance the confidentiality of the proposed scheme, (*iii*) watermark is embedded by quantizing the Cartesian component of highest singular value obtained from the DCT sub band with highest entropy value, (*iv*) watermark extraction process is blind, and (*v*) it achieves a good trade-off among imperceptibility, robustness, and data payload. Experimental results demonstrate that the proposed watermarking scheme shows high robustness against various attacks such as noise addition, cropping, resampling, requantization, signal addition, signal subtraction, and MP3 compression. Moreover, it outperforms state-of-the-art methods (Chen et al., 2013, 2010a; Erçelebi and Batakçı, 2009; Al-Nuaimy et al., 2011; Ali and Ahmad, 2010; Bhat et al., 2010; Lei et al., 2011; Khaldi and Boudraa, 2013; Xiang et al., 2011; Erfani and Siahpoush, 2009; Xiang and Huang, 2007; Xiang et al., 2008) in terms of imperceptibility, robustness, and data payload. This is because watermark bits are embedded into the Cartesian components of the largest singular value obtained from the DCT sub band with highest entropy value of each frame. The data payload of the proposed scheme is 172.39 bps which is relatively higher than that of the state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 provides background information including DCT, SVD, and LPT. Section 3 introduces our proposed watermarking scheme. Section 4 compares the proposed scheme with the state-of-the-art methods in terms of imperceptibility and robustness. Section 5 provides performance analysis of the proposed scheme. Lastly, the conclusion of this paper is presented in Section 6.

2. Background information

2.1. Discrete cosine transform

The DCT is widely used in signal and image processing, especially for lossy data compression. It expresses the sequence of many data points in terms of a sum of cosine functions oscillating at different frequencies. It can be written as.

$$X(k) = c(k)\sum_{n=0}^{N-1} x(n)\cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right), \quad k = 0, 1, \ldots, N-1 \tag{1}$$

where $x(n)$ is the audio signal with length N samples and

$$c(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k = 1,\ 2, \cdots N-1 \end{cases} \tag{2}$$

The DCT has a strong ‘energy compaction’ property, i.e., most of the signal information tend to be concentrated in a few low-frequency components of the DCT coefficients. This property can be utilized in audio watermarking to reduce the deterioration of watermarked signal.

2.2. Singular value decomposition

Let $H = \{H_{ij}\}_{p\times p}$ be an arbitrary matrix with SVD of the form $H = USV^T$, where U and V are orthogonal $p \times p$ matrices and S is a $p \times p$ diagonal matrix with non negative elements. The diagonal entries of S are called the singular values (SVs) of H where $S = \text{diag}(\sigma_1, \sigma_2, \ldots \sigma_p)$, the columns of U are called the left singular vectors of H, and the columns of V are called the right singular vectors of H.

2.3. Log-polar transformation

The LPT is a conformal transformation from the Cartesian coordinate system to the log-polar coordinate system.

Consider the log-polar coordinate system (r, θ), where r denotes the logarithm of the distance between a given point