Signal Processing 91 (2011) 1693–1708

Contents lists available at ScienceDirect

Signal Processing

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

A robust audio watermarking scheme based on reduced singular value decomposition and distortion removal

Jian Wang*, Ron Healy, Joe Timoney

Department of Computer Science, NUI Maynooth, County Kildare, Ireland

A R T I C L E I N F O

Article history:
Received 26 July 2010
Received in revised form
13 January 2011
Accepted 17 January 2011
Available online 4 February 2011

Keywords:
Audio watermarking
Singular value decomposition
MP3
Distortion removal
Psychoacoustics

A B S T R A C T

This paper presents a blind audio watermarking algorithm based on the reduced singular value decomposition (RSVD). A new observation on one of the resulting unitary matrices is uncovered. The proposed scheme manipulates coefficients based on this observation in order to embed watermark bits. To preserve audio fidelity a threshold-based distortion control technique is applied and this is further supplemented by distortion suppression utilizing psychoacoustic principles. Test results on real music signals show that this watermarking scheme is in the range of imperceptibility for human hearing, is accurate and also robust against MP3 compression at various bit rates as well as other selected attacks. The data payload is comparatively high compared to existing audio watermarking schemes.

1. Introduction

As audio, video and other works have become more widely available in digital form in recent years, the ease with which perfect copies can be made has increased the impact of unauthorized copying to such an extent that it is a major concern of the entertainment industry. The Recorded Industry Association of America (RIAA) states that the cost to the music industry alone is in the region of $12.5 billion annually [5]. It is obvious therefore that the issue undermines the music and film publishing industries. Concerns over protecting copyright have triggered significant research in, amongst other techniques, attempts to hide copyright messages in digital media. Examples of various techniques that have been developed are described in [1,11].

Digital watermarking is a technique whereby data is hidden in a cover or host signal as a form of steganography. Depending on whether the decoding phase of a watermarking scheme requires access to the original signal and/or the actual watermark in order to successfully decode the embedded data, the scheme may be described as ‘informed’ or ‘blind’. Informed decoding requires the original cover or the actual watermark. This presents many limitations for the scheme [4]. Blind decoding [12] does not require any information except private key in order to decode the watermark from the candidate signal. A watermarking scheme that provides for blind decoding is more difficult to achieve but offers a wider variety of potential applications of the scheme.

Depending on the watermark application and purpose, different requirements arise resulting in various design issues. Imperceptibility, or perceptual transparency, of the watermark is a general requirement independent of the application domain [11]. Artefacts introduced through a watermarking process are not only annoying and undesirable but may also reduce or destroy the commercial value of the watermarked data. Robustness, on the other hand, has to be considered in many application domains to prevent potential removal of watermark information by deliberate or accidental attacks on the watermarked signal. For some applications, the accidental attacks that would be

* Corresponding author.
E-mail address: jwang@cs.nuim.ie (J. Wang).

0165-1684/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2011.01.014

expected to be applied to watermarked signals would be known. As far as music is concerned, one of the most common attacks is compression. One very common technique is MP3 compression and surviving MP3 compression has therefore become an essential requirement for a watermarking scheme in the music domain [13]. This is due to the fact that MPEG-1 Layer III (MP3) and its variants have become the *de facto* standard for transmission and storage of compressed audio for both World Wide Web (WWW) and portable media applications [13]. The difficulty when designing a watermark that can withstand MP3 compression is related to the fact that the MP3 algorithm nonlinearly modifies the spectrum based on a psychoacoustic model and the expected bit rate. Thus, any watermark embedded among the audio's spectral components risks being lost after compression. Furthermore, the audio watermarking scheme should be robust against some other common attacks such as filtering [2,4,11].

Most audio watermarking algorithms use either a time domain [6,7] or frequency domain[8–10] masking property to embed a watermark [3,4]. Our previous work has considered a watermarking scheme utilizing the CSPE algorithm to encode the cover signal [60]. Due to the high precision of CSPE in estimating frequency components and thus through manipulating the identified components by CSPE, this scheme [60] achieved an almost 100% precision and extremely high perceptual transparency. However, the approach was found to be unsuitable for audio files that were subject to compression, as the watermark was sometimes lost during the process.

Recently, the singular value decomposition (SVD) has been used extensively as an effective technique in digital watermarking [14–18,22–34]. Most existing SVD based watermarking techniques are applied in images [16–18, 22–34]. SVD based audio watermarking techniques, also exist but fewer, some of which can be found in [14,15]. All these SVD based watermarking algorithms can be categorized into two groups.

The first group of SVD based watermarking algorithms are 'informed' [32–34], requiring access to the original signal or the watermark in order to successfully decode the embedded watermark. The scheme proposed in [32] can be used as a typical example to illustrate the idea used for this group. In the embedding stage, a cover image A will be decomposed into three matrices: U, S and V by SVD. The watermark information W then will be linearly added to S, resulting in a new matrix S', then inverse SVD will be applied on S', U and V to get a watermarked image A_w. In the extraction stage, U^T and V will be used to multiply with A_w to get S'. Then the watermark W can be extracted out by a linear subtraction between S' and S.

The second group SVD based schemes are 'blind' and embedding watermark information by manipulating the coefficients in the SVD decomposed matrices, such as U or S. These schemes are based on some observations proposed in [14–18]. Particularly, a digital image watermarking algorithm presented in [16] is based on two observations. Firstly, column-wise modification of the elements of the U matrix will cause less visible distortion than modifying the elements row-wise. Secondly, row-wise modification of the elements of V^T will cause less visible distortion than modifying the elements column-wise. Another particular SVD based digital image watermarking algorithm proposed in [17], is formulated on the observation that the elements in the first column of U and V can be modified without significantly impacting signal integrity, allowing them to survive common attacks including compression. In [14,15], the proposed audio watermarking algorithm is based on the observation that changing S slightly does not affect the quality of the signal much and that the singular values in S are consistent under common signal processing operations. In [18], an image watermarking algorithm based on tuning the coefficients in U was proposed.

Developing robust watermarking techniques for digital audio signals is relatively difficult compared to watermarking digital images. This is due to the high sensitivity of the human ear over a large dynamic range in comparison to the human eye [4,21]. Therefore, alterations have to be made to the SVD based image watermarking approach before it can be applied to audio. Compared with manipulating U, there are three deficiencies of manipulating S [14,15]. The first is that the modification of the largest coefficients in S [14] would cause a greater degradation [18] and it can then be easily detected and/or destroyed as the embedding position is public. The second is that embedding watermark information based on the modification of the less significant coefficients in S risks their loss after compression, which is based on removing less significant singular coefficients in S [19,20]. Finally, only a small number of singular elements in S are available to manipulate, compared with U, which makes it hard to achieve a high data payload [18]. No SVD based audio scheme, which achieves watermarking through manipulating U to embed information, has been proposed yet.

A new observation that the peaks in the second column of U are consistent after different attacks is proposed in this paper. Compared with manipulating the first column of U, modification of the second column introduces much less audible distortion. The proposed watermarking algorithm is based on this new observation. We investigated embedding on other columns and found that the robustness was adversely affected for no apparent gain. As mentioned above, the human auditory system is more sensitive than the human visual system. We therefore need to resolve any specific perceptual issues that arise. Specifically, distortion control is used to control the audible distortions and determine the strength of peaks created to represent watermark information in the embedding stage. Therefore, this scheme is adaptive to specific signal characteristics. Furthermore, psychoacoustic techniques are utilized to further suppress any resulting audible distortion. To the best of our knowledge, control of perceptual distortion has not been used in any SVD based audio watermarking schemes previously. The following sections details how these are achieved.

2. SVD

The SVD is a well-known numerical analysis tool used on matrices. If A is an arbitrary m-by-n matrix, with the full SVD, it can be decomposed as

$$A = USV^T \tag{1}$$