Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A robust intelligent audio watermarking scheme using support vector machine*

Mohammad MOSLEH[†‡1], Hadi LATIFPOUR[1], Mohammad KHEYRANDISH[1], Mahdi MOSLEH[2], Najmeh HOSSEINPOUR[3]

([1]*Department of Computer Engineering, Dezfoul Branch, Islamic Azad University, Dezfoul, Iran*)
([2]*Department of Computer Engineering, Andimeshk Branch, Islamic Azad University, Andimeshk, Iran*)
([3]*Young Researchers and Elite Club, Andimeshk Branch, Islamic Azad University, Andimeshk, Iran*)
[†]E-mail: Mosleh@iaud.ac.ir

Received Sept. 10, 2015; Revision accepted Feb. 21, 2016; Crosschecked Nov. 8, 2016

Abstract: Rapid growth in information technology and computer networks has resulted in the universal use of data transmission in the digital domain. However, the major challenge faced by digital data owners is protection of data against unauthorized copying and distribution. Digital watermark technology is starting to be considered a credible protection method to mitigate the potential challenges that undermine the efficiency of the system. Digital audio watermarking should retain the quality of the host signal in a way that remains inaudible to the human hearing system. It should be sufficiently robust to be resistant against potential attacks. One of the major deficiencies of conventional audio watermarking techniques is the use of non-intelligent decoders in which some sets of specific rules are used for watermark extraction. This paper presents a new robust intelligent audio watermarking scheme using a synergistic combination of singular value decomposition (SVD) and support vector machine (SVM). The methodology involves embedding a watermark data by modulating the singular values in the SVD transform domain. In the extraction process, an intelligent detector using SVM is suggested for extracting the watermark data. By learning the destructive effects of noise, the detector in question can effectively retrieve the watermark. Diverse experiments under various conditions have been carried out to verify the performance of the proposed scheme. Experimental results showed better imperceptibility, higher robustness, lower payload, and higher operational efficiency, for the proposed method than for conventional techniques.

Key words: Audio watermarking, Copyright protection, Singular value decomposition (SVD), Machine learning, Support vector machine (SVM)
http://dx.doi.org/10.1631/FITEE.1500297 **CLC number:** TP391

1 Introduction

The rapid development of cybernetics and information technologies, their universal application in various sectors of the postindustrial era, in the form of Internet and Intranet worlds, and the emergence of a knowledge-based environment for the swift and easy transportation of various types of data (such as audio, image, and video) allow people to share their data. One of the consequential legacies of such a technological revolution is the privilege for people of all standards of living and socioeconomic orientations to exchange and share data. These advantages, however, have not been without costs for designers, manufacturers, and users alike. One of such problems is the fact that owners are, more often than not, reluctant to let their works be disseminated without their due sanction. It is therefore crucial to effectively safeguard and protect data with strict copyright. Watermarking is one of such techniques that make it

‡ Corresponding author
* Project supported by the Dezfoul Branch, Islamic Azad University, Dezfoul, Iran
ORCID: Mohammad MOSLEH, http://orcid.org/0000-0002-0991-1623

possible for the owner information (i.e., watermark signal or simply watermark) to be imperceptibly hidden into an original signal as a text, numerical sequence, or image that safeguards owner information against attacks. Various watermarking schemes recently applied principally use images or videos. The audio watermarking approach has attracted the attention of researchers. Because the human auditory system is much more sensitive than its visual system, embedding data within audio signals in an imperceptible way is significantly harder than embedding an image or video.

An effective audio watermarking scheme should generally meet the imperceptibility, robustness, and payload requirements (Hartung and Kutter, 1999; Arnold, 2000; Cox *et al.*, 2007).

The imperceptibility condition implies that the difference between the watermarked and original signals is not understandable. Based on the recommendation by the International Federation of the Phonographic Industry (IFPI), a watermarked signal should function in such a manner as to have a signal-to-noise ratio (SNR) larger than 20 dB. Note that SNR is known as an objective measure that cannot be considered solely to be a good measurement for audio imperceptibility. For this reason, another objective measurement, called the objective difference grade (ODG), is considered. Moreover, a subjective measurement, mean opinion score (MOS), is used.

Robustness denotes the resistance of the watermarked signal against common signal processing and malicious attacks. Payload determines the maximum volume of data that can be placed in the host signal. As far as audio signals are concerned, payload refers to the number of watermark bits that can possibly be inserted within a host signal per unit of time. It is noteworthy that the minimum payload is 20 bits/s.

The watermarking schemes try to optimize the trade-off among the aforementioned requirements. No single and exclusive method has been presented to simultaneously meet all said requirements.

Most of conventional audio watermarking methods attempt to improve the trade-off among the three mentioned requirements, namely, imperceptibility, robustness, and payload. Such improvements have been achieved by introducing an appropriate embedding scheme, based on new modulation in a special domain, locating optimal insertion regions, capturing the intensity of watermark embedding, or combinations of them. Generally, non-intelligent correlation-based watermark decoders are applied in their extraction due to low complexity and simplicity. The major problems of these detectors are dependency on the decision threshold and the use of a series of specific rules, which undermines their performance. A new intelligence-based watermark decoding scheme, using the support vector machine (SVM), is used to overcome the system deficiencies, by learning the destructive effects of attacks and effectively extracting the watermark data. Moreover, the watermark data is inserted into the host signal according to the modulation of singular values (SVs) in the singular value decomposition (SVD) domain. The main reason for selecting the SVD domain is that the SVs change slightly and do not affect the quality of the audio signal. Moreover, the SVs are not significantly affected by audio-processing attacks.

2 Related works

Typical methods presented for audio watermarking can be categorized as intelligent audio watermarking schemes and non-intelligent ones.

2.1 Intelligent methods

Wang and Lin (2005) proposed an audio watermarking scheme based on discrete wavelet transform (DWT) and SVM. They used SVM as a means for capturing the characteristics of a watermark, embedded in low-frequency coefficients, in the wavelet domain. Wang *et al.* (2008) suggested a new SVM-based scheme, in which SVM was used to locate the optimal embedding positions in the original signal. Then they performed embedding operations on statistical average values of low-frequency components in the wavelet domain. Moreover, the SVM learning machine was applied in image watermarking. Yen and Wang (2006) proposed a new image watermarking scheme based on SVM, in which the SVM was trained to learn how to embed a binary random sequence, called the training sequence, into blue channels of the central and surrounding pixels. Then, with the help of the trained SVM, the owner's signature is embedded. In other words, the trained SVM was used to determine whether the current blue channels of the central and surrounding pixels need to be modified. Tsai and Sun (2007) proposed an SVM-based color image watermarking scheme in the spatial domain, wherein watermark extraction was considered a binary classification problem. Fu and Peng (2007) proposed an image watermarking scheme based on the subsampling technique in the wavelet domain. By using support vector regression (SVR), their scheme could model the relationship between randomly selected coefficients and the corresponding coefficients in other positions. Peng *et al*. (2010) presented an SVM-based image watermarking scheme in the multiwavelet domain, in which the special frequency band and property of the image in the multiwavelet domain were applied in the watermarking algorithm. SVM was used to learn the mean value relationship of multiwavelet coefficients in two approximation subbands.

Lei *et al*. (2013) proposed a novel evolution-based audio watermarking scheme for breath sound, by combining lifting wavelet transform (LWT), discrete cosine transform (DCT), SVD, and dither modulation (DM) quantization techniques. At first, this method applies DCT over LWT high-frequency coefficients. Afterward, the embedding operation was carried out by quantizing the SVs of the LWT–DCT coefficients. The particle swarm optimization algorithm was considered for optimizing the quantization steps.

Peng *et al*. (2013) introduced a new learning-based algorithm using kernel Fisher discriminant analysis (KFDA). In this scheme, each frame is divided into two subframes, using the downsampling technique, and the watermark data is embedded according to the local energy between the subframes. Moreover, the KFDA discriminator is applied to model the local energy relationship hidden in the watermarked audio signal.

Mohsenfar *et al.* (2013) suggested an efficient audio watermarking method based on the genetic algorithm (GA). In their method, after applying QR decomposition on audio blocks, the most appropriate locations in the QR domain are searched by GA to embed the watermark data.

2.2 Non-intelligent methods

Abd El-Samie (2009) presented an efficient SVD-based algorithm, in which the original signal was first segmented in both time and several transform domains. Then, the watermark was inserted by applying two sequential SVD transformations on the segments.

Fan and Wang (2009) proposed an effective method using discrete fractional sine transform (DFRST). In this method, after selecting the middle-frequency coefficients of the audio blocks and performing DFRST on them, the watermark sequence is embedded into the largest amplitude of DFRST coefficients.

Dutta *et al*. (2010) suggested a watermarking method for localizing the high-energy regions in the original signal for embedding the watermark data. The embedding operation was carried out by quantizing the SVs in the wavelet domain.

Tao *et al*. (2010) introduced a new scheme based on LWT. The watermark data was inserted by modifying the statistical average values of the subband coefficients. Moreover, the watermark strength was determined adaptively.

Bhat *et al*. (2010) proposed an adaptive algorithm based on SVD in the wavelet domain. Data embedding was performed by modulating the SVs of low frequencies in the wavelet domin, and then using the quantization index modulation (QIM) technique. The quantization step is characterized by the mean, standard deviation, and norm of the low frequencies. Moreover, Bhat *et al.* (2011) suggested a method using DM quantization of the SVs of the blocks.

Wang *et al*. (2011a) proposed a pseudo-Zernike moment-based audio watermarking scheme to withstand desynchronization attacks. This method embeds watermark data into the average value of the modulus of the low-order pseudo-Zernike moment. Moreover, they proposed an audio watermarking scheme, in which, instead of low-order pseudo-Zernike moments, wavelet moment invariance was used (Wang *et al.*, 2011b). Moreover, because of high performance, the moments were applied for image watermarking with great success (Tsougenis *et al.*, 2012).

Lei *et al*. (2012) presented a robust audio watermarking scheme by combining LWT and SVD. The watermark data was embedded into the SVs of the LWT low-frequency coefficients.

Hu *et al*. (2014) proposed a new scheme inspired from human auditory perception. At first, they decomposed the audio signal into critical bands, by using discrete wavelet packet transform (DWPT). Then, DCT was applied to analyze the content of the critical bands in DWPT. Subsequently, the watermark data was embedded in the DCT coefficients by the QIM technique.

Dhar and Shimamura (2015) presented a new method based on entropy and log-polar transformation (LPT). At first, low-frequency DCT coefficients of the audio blocks were calculated. After dividing these coefficients into sets of subbands, the entropy of each subband was computed. The watermark data was inserted into the Cartesian components of the largest SV of the DCT subband with the highest entropy value of each block.

3 Preliminaries

3.1 SVD decomposition

SVD, a common analytical framework for numerical analysis, has been used in many signal processing applications. Because of its unique features, SVD has emerged as a universally used technique in the watermarking domain. Given its main technical superiority, any minor modification in large SVs is unlikely to alter the transparency of the cover object. A typical SVD with an $m\times n$ matrix, $\boldsymbol{A}$, is formulated as follows:

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}=\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1r} \\ u_{21} & u_{22} & \cdots & u_{2r} \\ \vdots & \vdots & & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mr} \end{bmatrix}\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1r} \\ s_{21} & s_{22} & \cdots & s_{2r} \\ \vdots & \vdots & & \vdots \\ s_{r1} & s_{r2} & \cdots & s_{rr} \end{bmatrix} \cdot\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1r} \\ v_{21} & v_{22} & \cdots & v_{2r} \\ \vdots & \vdots & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nr} \end{bmatrix}^{\mathrm{T}}, \tag{1}$$

where $\boldsymbol{U}$ is an $m\times r$ orthogonal matrix, $\boldsymbol{S}$ is an $r\times r$ diagonal matrix with nonnegative elements on the diagonal, and $\boldsymbol{V}$ is an $n\times r$ unitary matrix. The nonnegative diagonal elements of $\boldsymbol{S}$, known as SVs of $\boldsymbol{A}$, are arranged in descending order. SVD is generally used in applications such as image processing, least-squares fitting of data, matrix approximation, as well as determination of the matrix rank, range, and null space (Trefethen and Bau, 1997).

3.2 Support vector machine

SVM is known as a powerful binary classifier in the machine learning domain. It can effectively and accurately analyze data and recognize patterns. Given a set of training data containing samples labeled as belonging to one of two determined classes, an SVM training algorithm builds a model that categorizes new samples into one class or the other. This is therefore referred to as a non-probabilistic binary linear classifier. An SVM is a representation of samples that can be imagined as points in space. It maps the samples belonging to isolated categories into regional spaces, divided by a clear gap that is expected to be as wide as possible. New samples are then mapped into the same space and anticipated to belong to one of the categories, based on whichever sides of the gap they fall on (Cortes and Vapnik, 1995).

4 The proposed method

As previously noted, several audio watermarking schemes have been presented. Most of them focus on the embedding stage in an attempt to insert the watermark data in such a manner as to decrease the effect of audio-processing attacks. Their extraction phase, on the other hand, can be performed without any intelligence and solely by a set of specific rules. In this study we propose an intelligent watermark decoder, based on SVM, which is able to learn the disruptive effects of processing attacks and effectively extract the watermark data. Moreover, watermark embedding is performed based on special modulation of the SVs, using the SVD decomposition method. Details of the proposed technique are described in the following sections.

4.1 Watermark embedding

Given its robustness against signal processing attacks, the SVD decomposition method is used to insert the watermark data. The watermark data, **WD**, includes two parts: one is a random binary sequence, known as the training sequence and indicated by **TS**, and the other is a watermark image, of size $M=t_1\times t_2$, denoted by **WI**. Therefore, the length of the watermark data is $P=N+M$. Steps of the embedding procedure are shown in Algorithm 1.

Algorithm 1 Procedure for watermark embedding

Input: host signal $\boldsymbol{Z}$, watermark data **WD**, and key K.

Output: watermarked signal **WS**.

Step 1: Divide the host signal $\boldsymbol{Z}$ into non-overlapping frames, each containing L^Z samples, where

$$L=\left\lfloor\sqrt{\frac{\text{length}(\boldsymbol{Z})}{\text{length}(\mathbf{WD})}}\right\rfloor,$$

where $\lfloor\cdot\rfloor$ returns the largest previous integer.

Step 2: Segment the calculated frames into blocks, represented by $L\times L$ matrices.

Step 3: Select P blocks for embedding watermark data, **WD**, based on a pseudorandom generator, with the key K.

Step 4: Apply SVD over the selected blocks:

$$\boldsymbol{A}_p=\boldsymbol{U}_p\boldsymbol{S}_p\boldsymbol{V}_p^{\mathrm{T}},\ p=1,2,...,P.$$

Step 5: Calculate the norm of the diagonal matrix $\boldsymbol{S}_p$ as follows:

$$X_p=\left\|\boldsymbol{S}_p\right\|=\sqrt{\sum_{i=1}^{L}(S_{pi})^2},\ p=1,2,...,P.$$

Step 6: Quantize the norm values $\boldsymbol{X}_p$ as follows:

$$\varphi(p)=\left\lfloor\frac{X_p}{\varDelta}\right\rfloor,\ p=1,2,...,P,$$

where $\varDelta$ indicates the quantization step.

Step 7: Embed the watermark data, **WD**, based on the following modulation:

$$\varphi'(p)=\begin{cases}\varphi(p)+d-[\varphi'(p)\bmod(2d)], & \mathrm{WD}(p)=1,\\ \varphi(p)+d-[(\varphi'(p)+d)\bmod(2d)], & \mathrm{WD}(p)=0,\end{cases}$$

where d is a positive value used to maximize the robustness.

Step 8: Modify the singular values by

$$X'_p=\lfloor\Delta\varphi'(p)\rfloor+\left\lfloor\frac{\varDelta}{2}\right\rfloor,\ p=1,2,...,P,$$

$$\boldsymbol{S}'_p=\boldsymbol{S}_p\left\lfloor\frac{X'_p}{X_p}\right\rfloor,\ p=1,2,...,P.$$

Step 9: Perform the SVD inverse:

$$\boldsymbol{A}'_p=\boldsymbol{U}_p\boldsymbol{S}'_p\boldsymbol{V}_p^{\mathrm{T}},\ p=1,2,...,P.$$

Step 10: Reconstruct watermarked signal **WS** via modified $\boldsymbol{A}'_p$ blocks.

4.2 Watermark extraction

Different signal processing attacks, such as noise addition, filtering, re-sampling, re-quantization, and MP3 compression, can corrupt the watermarked signal and, consequently, undermine the performance efficiency of the watermark decoder. Most audio watermarking schemes generally apply correlation-based watermark decoders. The major challenges posed by these decoders include dependency on the threshold value and the use of a set of non-intelligent specific rules for watermark extraction. In this study we present a novel intelligence-based watermark detector using SVM. It has some technical features that are unique in extracting the watermark data by learning the destructive effects of attacks on SVs of audio blocks. The proposed extraction phase includes two parts: training the SVM learning machine and extracting the owner signature. In the first part, pre-processing operations, including framing, blocking, and computing SVs of blocks, are initially performed on the received signal. Then, N labeled training patterns are generated using the computed SVs of the first N blocks and training sequence, denoting the labels, according to

$$\text{training\_patterns}=(s'_{i1},s'_{i2},...,s'_{iL},\mathrm{TS}(i)),1\le i\le N,\tag{2}$$

where s'_i are the SVs of the ith block of the received signal and TS(i)={−1 or +1} are the labels. Finally, SVM is trained using the labeled training patterns.

In the second part, the main goal includes retrieving the owner signature via the trained SVM machine. For this purpose, at first, M testing patterns are extracted via the SVs of the next M blocks:

$$\text{testing\_patterns}=(s'_{i1},s'_{i2},...,s'_{iL}),\ N+1\le i\le N+M.\tag{3}$$

Subsequently, labels of the extracted testing set are identified via the trained SVM, as follows:

$$\text{owner\_signatuer}(i)=\begin{cases}0, & \text{if returned label}(i)=-1,\\ 1, & \text{if returned label}(i)=+1,\end{cases}\quad N+1\le i\le N+M.\tag{4}$$

The pseudocode of this method is shown in Algorithm 2.

Algorithm 2 Procedure for watermark extraction
Input: received signal $\boldsymbol{Z}'$, training sequence **TS**, and key K.
Output: owner signature.

Step 1: Divide the received signal $\boldsymbol{Z}'$ into non-overlapping frames containing L^2 samples, similar to the same in the embedding phase.

Step 2: Segment the calculated frames into $L\times L$ square blocks.

Step 3: Select P blocks according to the pseudorandom generator with key K.

Step 4: Apply SVD over P selected blocks:

$$\boldsymbol{A}'_p = \boldsymbol{U}'_p \boldsymbol{S}'_p \boldsymbol{V}'^{\mathrm{T}}_p,\ \ p = 1,2,...,P.$$

Step 5: Generate N training patterns via the first N singular value matrices $\boldsymbol{S}'_p$ using Eq. (2).

Step 6: Train SVM by using the training patterns.

Step 7: Generate M testing patterns via the next M singular value matrices $\boldsymbol{S}'_p$ using Eq. (3).

Step 8: Decode labels of M test patterns via the trained SVM. That is, the watermark bit is 0 if the output of SVM is −1, and 1 otherwise.

Step 9: Reconstruct the owner signature via the obtained watermark bits.

5 Experimental results

To evaluate the proposed scheme, five different audio files were used (Table 1). A picture of size M=40×40 (Fig. 1) was used as a watermark image. The training sequence was a random binary sequence with N=1225, zero mean, and unit variance. Hence, the final length of the watermarked data was P=2825.

Table 1 Characteristics of audio files

Genre	Playing time (s)	Sampling rate (kHz)	Format	Coding
Jazz	9.98			
Blue	9.28			
Classic	10.00	44.1	Wave	16-bit Mono
Rock	9.65			
Electronic	10.00			

Fig. 1 The watermark image

In addition, the proposed method was implemented using MATLAB. To train the SVM classifier, values of some basic parameters are listed in Table 2 and other parameters were set to their default values.

Table 2 Values of some basic parameters for training the SVM classifier

Parameter	Value	Description
Kernel	RBF	Kernel function used to map the training data into the kernel space
Method	LS	Method used to find the hyper-plane

Generally, there are two criteria for defining the imperceptibility for audio watermarking schemes, namely ODG and SNR.

ODG is an objective test, the result of which is calculated by perceptual evaluation of the audio quality algorithm (PEAQ), characterized in ITU BS.1387-1 (Kabal, 2002). To measure the ODG values, we have applied the OPERA software based on PEAQ Advanced, which has been introduced by OPTICOM. ODG ranged from −4 to 0 (Table 3). Because the final judgment on the quality of the watermarked audio signal was performed by the human auditory system, we have used a subjective test. The MOS measure is considered a subjective criterion (Table 3).

Table 3 ODG and MOS descriptive values

Description	ODG	MOS
Imperceptible	0	5.0
Perceptible, but not annoying	−1	4.0
Slightly annoying	−2	3.0
Annoying	−3	2.0
Very annoying	−4	1.0

SNR is a measure applied in signal processing science and expresses the power ratio of the original signal with respect to the power of background noise. It can be calculated as follows:

$$\mathrm{SNR} = 10\lg\left(\frac{\sum_n \boldsymbol{S}^2[n]}{\sum_n [\boldsymbol{S}[n] - \boldsymbol{S}'[n]]^2}\right)\ \mathrm{dB}, \quad (5)$$

where $\boldsymbol{S}[\cdot]$ is the original signal and $\boldsymbol{S}'[\cdot]$ is the watermarked signal.

The results obtained by evaluating the transparency of the proposed method over different audio files are illustrated in Table 4.

Table 4 SNR and ODG values of the proposed method

Audio file	SNR (dB)	ODG	MOS
Jazz	42.04	−0.42	4.91
Classic music	42.35	−0.34	4.91
Blues	41.31	−0.47	4.89
Rock-and-roll	41.92	−0.30	4.90
Electronic	42.76	−0.14	4.91
Average	42.07	−0.33	4.90

The average values SNR=42.0779 dB and ODG=−0.33 indicated a high imperceptibility obtained by using the proposed method.

Moreover, subjective evaluation has been done using an ABX test (Acevedo, 2006). In each test we considered the original audio signal A, the watermark audio signal B, and the undefined audio signal X, which could be either A or B. For this purpose, five listeners were employed and asked to report the MOS scores, according to Table 3. All the reported MOS scores were very close to five, indicating that our proposed method presents a very good imperceptibility.

Table 5 shows a comparison between the proposed method and former works in terms of transparency.

Table 5 Comparison of transparency between the proposed method and existing methods

Reference	SNR (dB)
Abd El-Samie (2009)	27.13
Fan and Wang (2009)	40.70
Bhat *et al.* (2010)	24.37
Lei *et al.* (2012)	40.00
Mohsenfar *et al.* (2013)	25.89
Hu *et al.* (2014)	20.95
Dhar and Shimamura (2014)	37.20
Proposed method	42.07

To evaluate the robustness of the proposed method against different signal processing attacks, two important measures, normalized correlation (NC) and bit error rate (BER), were applied.

The NC coefficient was used to compare the similarity between the original and the extracted watermarks, expressed by

$$\mathrm{NC}(\boldsymbol{W}_a,\boldsymbol{W}_b)=\frac{\sum_{i=1}^{P}\boldsymbol{W}_a(i)\cdot\boldsymbol{W}_b(i)}{\sqrt{\sum_{i=1}^{P}\boldsymbol{W}_a^2(i)}\cdot\sqrt{\sum_{i=1}^{P}\boldsymbol{W}_b^2(i)}},\tag{6}$$

where $\boldsymbol{W}_a$ and $\boldsymbol{W}_b$ are the original and extracted watermarks respectively, and P denotes the watermark length.

The BER measure was applied to show the accuracy of the watermark decoder, after signal processing attacks. It is defined as follows:

$$\mathrm{BER}(\boldsymbol{W}_a,\boldsymbol{W}_b)=\frac{1}{P}\sum_{i=1}^{P}\boldsymbol{W}_a(i)\oplus\boldsymbol{W}_b(i),\tag{7}$$

where '⊕' is the exclusive OR (XOR) operator.

Stirmark attacks (Lang, 2005) and also common signal processing attacks, including noise addition, low-pass filtering, MP3 compression, re-quantization, and re-sampling, are presented as follows:

Noise addition: The white Gaussian noise is added to the watermarked signal.

Low-pass filtering: A low-pass filter with a 4 kHz cut-off frequency is applied to the watermarked signal.

MP3 compression: The watermarked signal is compressed at the bit rate of 64 kb/s and then decompressed.

Re-quantization: Quantize down the watermarked signal to 8-bit, followed by re-quantization to 16-bit.

Re-sampling: Down-sample the original watermarked signal from 44.1 kHz to 22.05 kHz, which is then restored for sampling again, at 44.1 kHz.

The robustness of the proposed method against several Stirmark attacks on diverse music files was evaluated. Results are shown in Table 6.

Table 7 illustrates the NC and BER averages of the proposed method against Stirmark attacks for five different music files.

The proposed method was also tested with common signal processing attacks in MATLAB for exhaustive evaluation of robustness. The results are

Table 6 NC and BER for the proposed method against Stirmark attacks on different music files

Attack	NC					BER (%)				
	Electronic	Blues	Rock	Classic	Jazz	Electronic	Blues	Rock	Classic	Jazz
No attack	1	1	1	1	1	0	0	0	0	0
AddBrumm	1	1	1	1	1	0	0	0	0	0
Addnoise	1	1	1	1	1	0	0	0	0	0
Compressor	1	1	1	1	1	0	0	0	0	0
Dynnoise	1	1	1	1	1	0	0	0	0	0
Exchange	1	1	1	1	1	0	0	0	0	0
ExtraStereo	1	1	1	1	1	0	0	0	0	0
Echo	0.9397	0.9435	0.9320	0.9135	0.9312	9	7	8	10	8
FFT_Invert	1	1	1	1	1	0	0	0	0	0
FFT_Real_Reverse	1	1	1	1	1	0	0	0	0	0
Invert	1	1	1	1	1	0	0	0	0	0
Lsbzero	1	1	1	1	1	0	0	0	0	0
Rc_Lowpass	1	1	1	1	1	0	0	0	0	0
State1	1	1	1	1	1	0	0	0	0	0
Zerocross	1	1	1	0.9992	0.9946	0	0	0	0	1
State2	0.9949	0.9966	1	0.9966	0.9916	0	0	0	0	1
Smooth2	0.9933	0.9966	0.9983	0.9946	0.9874	1	0	0	1	2
Normalize	0.9931	0.9950	0.9966	0.9865	0.9865	0	0	0	2	2
Zeroremove	0.9763	0.9772	0.9899	0.9792	0.9698	3	3	1	3	5
Smooth	0.9712	0.9715	0.9636	0.9752	0.9632	4	4	4	4	6
Average	0.9934	0.9940	0.9940	0.9922	0.9912	0.85	0.70	0.65	1	1.25

shown in Table 8. The average BER in the presence of mentioned common signal processing attacks was 0.13%, and the average NC was very close to 1. This is a clear evidence of relatively high robustness of the proposed method, compared with other models. The percentages of NC and BER criteria for noise addition, low-pass filtering, MP3 compression, re-quantization, and re-sampling were presented. The comparative results are provided in Tables 9 and 10.

The watermark payload was determined as the number of bits that can be embedded in and extracted from the audio signal. This can be calculated in the number of bits per second (bit/s). Considering an L-second audio signal and a P-bit watermark data, the payload can be defined as follows (Bhat *et al*., 2010):

$$\text{Data payload} = P / L \text{ bits/s}. \tag{8}$$

So, considering that the length of useful watermark data is 1600 bits, the proposed method can provide a payload of 172.41 bits/s at the maximum.

The results of the proposed method in comparison with other methods, in terms of capacity, SNR, and BER, are shown in Table 11.

Table 7 Averages of NC and BER for the proposed method against Stirmark attacks on the music files

Attack type	NC	BER (%)
No attack	1	0
AddBrumm	1	0
Addnoise	1	0
Compressor	1	0
Dynnoise	1	0
Exchange	1	0
ExtraStereo	1	0
Echo	0.9320	8.4
FFT_Invert	1	0
FFT_Real_Reverse	1	0
Invert	1	0
Lsbzero	1	0
Rc_Lowpass	1	0
State1	1	0
Zerocross	0.9987	0.2
State2	0.9959	0.2
Smooth2	0.9940	0.8
Normalize	0.9915	0.8
Zeroremove	0.9784	3.0
Smooth	0.9689	4.4
Average	0.9929	0.89

Table 8 NC and BER for the proposed scheme against common attacks in MATLAB on different music files

Attack	NC					BER (%)				
	Electronic	Blues	Rock	Classic	Jazz	Electronic	Blues	Rock	Classic	Jazz
No attack	1	1	1	1	1	0	0	0	0	0
Noise addition	1	1	1	1	1	0	0	0	0	0
Low-pass filtering (4 kHz)	1	1	1	1	1	0	0	0	0	0
MP3 (64 kbps)	1	1	1	1	1	0	0	0	0	0
Re-quantization	1	0.9821	0.9754	0.9853	0.9978	0	1	2	1	0
Re-sampling (22.05 kHz)	0.9946	1	1	1	0.9991	0	0	0	0	0
Average	0.9991	0.9970	0.9959	0.9975	0.9994	0	0.1666	0.3333	0.1666	0

Table 9 Comparison of NC between the proposed scheme and existing methods

Reference	NC					
	No attack	Noise addition	Low-pass filtering (4 kHz)	MP3 (64 kb/s)	Re-quantization	Re-sampling (22.05 kHz)
Fan and Wang (2009)	1	0.9729	0.9762	0.9974	NA	1
Tao *et al.* (2010)	1	0.9863	NA	0.9621	1	0.9634
Bhat *et al.* (2010)	1	1	0.9992 (11.025 kHz)	0.9996	1	0.9955
Bhat *et al.* (2011)	1	1	1 (11.025 kHz)	0.9969	1	0.9982
Peng *et al.* (2013)	1	0.9645	0.9517	0.9457	1	0.9295
Dhar and Shimamura (2014)	1	0.9931	NA	0.9695 (128 kb/s)	1	0.9882
Proposed method	1	1	1	1	0.9881	0.9987

Table 10 Comparison of the BER obtained using the proposed scheme and existing methods

Reference	BER (%)					
	No attack	Noise addition	Low-pass filtering (4 kHz)	MP3 (64 kb/s)	Re-quantization	Re-sampling (22.05 kHz)
Wang *et al.* (2008)	0	0	0.0002	0	0	0
Fan and Wang (2009)	0	0.0422	0.0371	0.0042	NA	0
Tao *et al.* (2010)	0	1.32	NA	2.1	0	6.8
Bhat *et al.* (2010)	0	0	0	0	0	0.6
Bhat *et al.* (2011)	0	0	0 (11.025 kHz)	0.2	0	0.2
Wang *et al.* (2011a)	0	0.72	0.03	0.24	0.23	0.14
Lei *et al.* (2012)	0	0	0	0	0	0.145
Mohsenfar *et al.* (2013)	0	0	0 (9 kHz)	7	NA	0
Hu *et al.* (2014)	0	0.06	0	1.08	0	0
Dhar and Shimamura (2014)	0	0.830	NA	3.66 (128 kb/s)	0	1.416
Proposed method	0	0	0	0	0.8	0

Table 11 Comparison of the proposed method with existing methods in terms of capacity, SNR, and BER

Reference	Capacity (bit/s)	SNR (dB)	Average of BER (%)
Fan and Wang (2009)	86.00	40.70	0.01
Bhat *et al.* (2010)	45.90	24.37	0.10
Wang *et al.* (2011a)	27.20	NA	0.22
Lei *et al.* (2012)	170.67	40.00	0.03
Mohsenfar *et al.* (2013)	159.00	25.89	1.40
Hu *et al.* (2014)	150.73	20.95	0.19
Dhar and Shimamura (2014)	172.39	37.20	1.18
Proposed method	172.41	42.07	0.13

As can be observed, the comprehensive evaluation of the proposed scheme includes performing several transparency and robustness tests on different music files: electronic, Blues, Rock, classic, and jazz. The results showed that despite having a high payload, the proposed method has maintained a higher transparency compared to other methods. Moreover, the results of the test on the proposed method in terms of robustness against common signal processing attacks showed a high resistance quality.

6 Conclusions

A new robust intelligent audio watermarking scheme using SVM has been proposed. The system is configured in such a manner that the watermark data is inserted in the original signal, according to modulations of the singular values in the SVD domain. An intelligence-based watermark decoder capable of extracting the watermark data using SVM is suggested. SVM can effectively learn the disruptive effects of attacks on singular values. The results obtained by the tests demonstrate high transparency and robustness against signal processing attacks. Although in SVM training the complexity of the proposed scheme is greater than that of correlation-based conventional audio watermarking schemes, our proposed method has inbuilt unique features which optimize the trade-off among the three requirements of imperceptibility, robustness, and payload.

Acknowledgements

The presented paper is the result of a research entitled 'Presenting an adaptive audio watermarking scheme for increasing robustness against attacks using optimization algorithms and computational intelligent techniques', which was sponsored by Islamic Azad University of Dezfoul and the authors hereby express their thanks.

References

Abd El-Samie, F.E., 2009. An efficient singular value decomposition algorithm for digital audio watermarking. *Int. J. Speech Technol.*, **12**(1):27-45. http://dx.doi.org/10.1007/s10772-009-9056-2

Acevedo, A.G., 2006. Audio watermarking quality evaluation. *In*: E-business and Telecommunication Networks. Springer, p.272-283.

Arnold, M., 2000. Audio watermarking: features, applications, and algorithms. IEEE Int. Conf. on Multimedia and Expo, p.1013-1016.

Bhat, K.V., Sengupta, I., Das, A., 2010. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. *Dig. Signal Process.*, **20**:1547-1558. http://dx.doi.org/10.1016/j.dsp.2010.02.006

Bhat, K.V., Sengupta, I., Das, A., 2011. An audio watermarking scheme using singular value decomposition and dither-modulation quantization. *Multim. Tools Appl.*, **52**(2):369-383. http://dx.doi.org/10.1007/s11042-010-0515-1

Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.*, **20**(3):273-297. http://dx.doi.org/10.1007/BF00994018

Cox, I., Miller, M., Bloom, J., *et al.*, 2007. Digital Watermarking and Steganography. Morgan Kaufmann, USA.

Dhar, P.K., Shimamura, T., 2015. Blind SVD-based audio watermarking using entropy and log-polar transformation. *J. Inform. Secur. Appl.*, **20**:74-83. http://dx.doi.org/10.1016/j.jisa.2014.10.007

Dutta, M.K., Pathak, V.K., Gupta, P., 2010. A robust watermarking algorithm for audio signals using SVD. Int. Conf. on Cotemporary Computing, p.84-93. http://dx.doi.org/10.1007/978-3-642-14834-7_9

Fan, M., Wang, H., 2009. Chaos-based discrete fractional Sine transform domain audio watermarking scheme. *Comput. Electr. Eng.*, **35**(3):506-516. http://dx.doi.org/10.1016/j.compeleceng.2008.12.004

Fu, G., Peng, H., 2007. Subsampling-based wavelet watermarking algorithm using support vector regression. Int. Conf. on "Computer as a Tool", p.138-141. http://dx.doi.org/10.1109/EURCON.2007.4400269

Hartung, F., Kutter, M., 1999. Multimedia watermarking techniques. *Proc. IEEE*, **87**(7):1079-1107. http://dx.doi.org/10.1109/5.771066

Hu, H.T., Hsu, L.Y., Chou, H.H., 2014. Perceptual-based DWPT-DCT framework for selective blind audio watermarking. *Signal Process.*, **105**:316-327. http://dx.doi.org/10.1016/j.sigpro.2014.05.003

Kabal, P., 2002. An Examination and Interpretation of ITU-R BS. 1387: Perceptual Evaluation of Audio Quality. TSP Lab Technical Report, Department of Electrical & Computer Engineering, McGill University.

Lang, I.A., 2005. Stirmark Benchmark for Audio (SMBA): Evaluation of Watermarking Schemes for Audio. Version 1.3.1.

Lei, B., Soon, I., Zhou, F., *et al.*, 2012. A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition. *Signal Process.*, **92**(9): 1985-2001. http://dx.doi.org/10.1016/j.sigpro.2011.12.021

Lei, B., Song, I., Rahman, S.A., 2013. Robust and secure watermarking scheme for breath sound. *J. Syst. Softw.*, **86**(6):1638-1649. http://dx.doi.org/10.1016/j.jss.2013.02.022

Mohsenfar, S.M., Mosleh, M., Barati, A., 2013. Audio watermarking method using QR decomposition and genetic algorithm. *Multim. Tools Appl.*, **74**(3):759-779. http://dx.doi.org/10.1007/s11042-013-1694-3

Peng, H., Wang, J., Wang, W., 2010. Image watermarking method in multiwavelet domain based on support vector machines. *J. Syst. Softw.*, **83**(8):1470-1477. http://dx.doi.org/10.1016/j.jss.2010.03.006

Peng, H., Li, B., Luo, X., *et al.*, 2013. A learning-based audio watermarking scheme using kernel Fisher discriminant analysis. *Dig. Signal Process.*, **23**(1):382-389. http://dx.doi.org/10.1016/j.dsp.2012.08.006

Tao, Z., Zhao, H.M., Wu, J., *et al.*, 2010. A lifting wavelet domain audio watermarking algorithm based on the statistical characteristics of sub-band coefficients. *Arch. Acoust.*, **35**(4):481-491.

Trefethen, L.N., Bau, D.III, 1997. Numerical Linear Algebra. SIAM.

Tsai, H.H., Sun, D.W., 2007. Color image watermark extraction based on support vector machines. *Inform. Sci.*, **177**(2): 550-569. http://dx.doi.org/10.1016/j.ins.2006.05.002

Tsougenis, E., Papakostas, G.A., Koulouriotis, D.E., *et al.*, 2012. Performance evaluation of moment-based watermarking methods: a review. *J. Syst. Softw.*, **85**(8):1864-1884. http://dx.doi.org/10.1016/j.jss.2012.02.045

Wang, J., Lin, F.Z., 2005. Digital audio watermarking based on support vector machine. *J. Comput. Res. Dev.*, **42**(9): 1605-1611 (in Chinese).

Wang, X.Y., Niu, P.P., Qi, W., 2008. A new adaptive digital audio watermarking based on support vector machine. *J. Netw. Comput. Appl.*, **31**(4):735-749. http://dx.doi.org/10.1016/j.jnca.2007.10.001

Wang, X.Y., Ma, T.X., Niu, P.P., 2011a. A pseudo-Zernike moment based audio watermarking scheme robust against desynchronization attacks. *Comput. Electr. Eng.*, **37**(4): 425-443. http://dx.doi.org/10.1016/j.compeleceng.2011.05.011

Wang, X.Y., Niu, P.P., Lu, M.Y., 2011b. A robust digital audio watermarking scheme using wavelet moment invariance. *J. Syst. Softw.*, **84**(8):1408-1421. http://dx.doi.org/10.1016/j.jss.2011.03.033

Yen, S.H., Wang, C.J., 2006. SVM based watermarking technique. *Tamkang J. Sci. Eng.*, **9**(2):141-150.