УЧЕНЫЕ ЗАПИСКИ КАЗАНСКОГО УНИВЕРСИТЕТА.
СЕРИЯ ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

2021, Т. 163, кн. 1
С. 77–89

ISSN 2541-7746 (Print)
ISSN 2500-2198 (Online)

ORIGINAL ARTICLE

UDC 003.26

doi: 10.26907/2541-7746.2021.1.77-89

A NEW DCT FILTERS-BASED METHOD TO IMPROVE THE RESISTANCE OF TERNARY WATERMARKS IN AUDIO FILES AGAINST ATTACKS

R.Kh. Latypov, E.L. Stolov

Kazan Federal University, Kazan, 420008 Russia

Abstract

Watermarks used to verify the authorship of various works of art is a simple tool available to novice authors. This paper proposes a new technique to enhance the resistance of watermarks in audio files. As parts of a melody or image, leveraged watermarks can be recognized by human beings, even if dramatically corrupted. Nevertheless, increasing the quality of extracted watermarks is always helpful. This is achieved by new methods developed for damaged watermark restoration. In our methods, the source media is converted into a ternary code. The suggested technique is based on the restoration of a part of the source's features via its ternary code. Discrete cosine transform matrix rows are used as the final impulse response filters. Some features of the source are extracted by applying a linear combination of these filters to the ternary watermark. We consider the two most frequently occurring attacks: filtering and converting into mp3 format.

Keywords: digital watermarks, ternary coding, audio files, discrete cosine transform

Introduction

Watermarks have received much attention in designing practical solutions to protect the digital media copyright by embedding them in the original audio signal. A survey of the methods used for inserting watermarks in the host file is presented in [1]. Other review papers [2, 3] and books [4, 5] also sum up various sides of audio watermarking. As a rule, a watermark is inserted into the time domain or transform domain. The watermarking techniques proposed in [6–8] are based on the spread spectrum insertion of watermarks in the time domain. In the paper [9], the authors use perceptual temporal and frequency masking to embed the watermark bits into the host signal. In [10], an efficient algorithm of audio watermarking in the frequency domain is described. These examples show that balancing among the required watermark properties is a challenging task. In a recent paper [11], the authors introduce a new watermarking technique for audio using *spikegram* and a two-dictionary method. Their experiments show its good robustness to mp3 compression. In [12], a computationally efficient audio watermarking method is proposed. The discrete cosine transform (DCT) is applied to the host audio signal to obtain a certain audio segment. The experimental results demonstrate the robustness of the method against frequent audio attacks. In the paper [13], the audio signal is transformed into the frequency domain, and singular value decomposition is applied. The robustness of the method against various audio attacks is demonstrated. Using the discrete sine transform method in the study [14] increased the robustness, thereby resulting in a good BER value. There are also hybrid methods

where both domains are transformed, but we shall not discuss them here, because they are a subject of a separate research.

The main features of digital audio watermarking are perceptual transparency, robustness, security, verification reliability, and data capacity. More precisely, watermarks must be inaudible within the host signal to keep the audio quality and robust to signal distortions applied to the host file. Furthermore, watermarks should be both undetectable by unauthorized parties and easy to extract to prove ownership. Finally, high rate is desirable, i.e., a large number of watermark digits must be embedded successfully into the host audio per time unit. To meet these needs, new watermarking techniques should be developed.

Generally, original digital audio products are saved in an uncompressed format (wav 16, 24, or 32 bits). If the author wants to publish their composition on the Internet, they must use a compressed form of work, i.e., the form used in various advertising media. Users of mobile gadgets listen to the compressed version, and often that is all they want. If a scammer tries to take credit for the original audio file, they convert the compressed version using a slight transform to deny complete copying. To prevent it, the author must embed a watermark that can resist a series of the mentioned transformations. Formally, the product's author must predict the exact form of the watermark extracted from the file. Any deviation from the predicted original can be viewed as a subject for challenging the authorship. It is hardly possible that the inserted watermark stays unchanged following various transforms. Therefore, recognizing corrupt watermarks is an urgent problem. We suggest leveraging a human for recognition of a damaged watermark in the file. With this aim, we employed famous melodies or pictures as watermarks. Such objects can be easily identified, even if a considerable part of the watermark is distorted. The focus of our research is on enhancing the quality of the damaged watermark.

Let us briefly describe the methods used to insert a watermark into an audio file. This watermark can be seen as a sequence

$$Watermark = a_1, a_2, \ldots, a_N. \tag{1}$$

Here, a_i can be real numbers or binary numbers or elements of a set. In our paper, we investigate the case where $a_i \in \{-1, 0, 1\}$. We use the notion $trit$ for a variable that can acquire three possible values from the set. We show that such type of coding offers significant advantages. The audio file under protection, also known as a container, is also a sequence

$$File = s_1, s_2, \ldots, s_M. \tag{2}$$

In this paper, we experiment with 16-bit format audio file elements, where all samples are integers from interval $[-32768, 32767]$. Most of the transforms with insertion assume a floating-point format, which means that the downloaded file must be converted into float format. After all the manipulations, the file should be converted back to the original format. The above-described operations result in a loss of accuracy, and this property must be taken into account.

We consider only watermarks inserted into the time domain. The container is divided into non-overlapping fragments $Fragm = s_k, s_{k+1}, \ldots, s_{k+L-1}$. Any $trit$ a_i in (1) corresponds to a fragment $Fragm_j$. The length L of the fragment and the number of $\{a_i\}$ destined to the same piece define the method's payload. It should be noted that the length of the fragment may vary.

The assigned parts change depending on the elements selected from (1), and the modified version of the piece takes its former place (time-domain transform). There are known insertion methods utilizing particular transforms. Most of them have a better payload, but they require more computations to detect and select appropriate fragments for embedding, because the transformed fragment must begin and end with silence or noise. It is sometimes challenging to meet these conditions.

The issue boils down to the technology's choice for a more compact form of storing the chosen watermark features and how to store them in the container. In the classical approach, the watermark is represented as a binary sequence [1]. We use ternary coding for this purpose. A technique was developed by us for suboptimal presentation of an audio signal or a gray-level image as a ternary sequence (see [15, 16]). In our paper, these techniques are further refined, theoretical justification is carried out, and experimental results give numerical estimates of the developed method.

The amplitude modulation follows the B-splines theory [17], and the technique based on echo hiding is selected for insertion [18, 19]. Here, most attention is paid to improving the methods of watermark extraction from the container after an attack.

The following notations are used. Vector $\mathbf{A} = \langle a_1, a_2, \ldots, a_n \rangle$. If $\mathbf{A}$, $\mathbf{B}$ are two vectors of the length n, then $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} \cdot \mathbf{B}$ are the same lengths obtained by entrywise addition and multiplication of the vector components, and $(\mathbf{A}, \mathbf{B})$ is the dot product of these vectors. The symbol $A[k]$ denotes the k-th item in sequence A, $Matr[i, j]$ is the item of the matrix in i-th row and j-th column, $Matr[*, j]$ is the j-th column of the matrix. The power of the vector is the sum of squares of its items. If G is a function or sequence, then FG results from its Fourier transform or discrete Fourier transform. We often have to compare a vector $\mathbf{A}$ with its transformation $\mathbf{B}$. To do this, we use the signal-to-noise ratio (SNR) in the following two forms:

$$SNR = 10 \log_{10} \left(\frac{\sigma^2(\mathbf{A})}{\sigma^2(\mathbf{A} - \mathbf{B})} \right) \tag{3}$$

or

$$SNR = 10 \log_{10} \left(\frac{\sigma^2(\mathbf{FA})}{\sigma^2(\mathbf{FA} - \mathbf{FB})} \right). \tag{4}$$

Here, $\sigma^2(G)$ represents a variance of the function or sequence. The former formula fits the situation where the transformation is a point operation (the result depends on one point of the source). In contrast, the latter formula is suitable for spatial operations. The formulas give correct results if both vectors have the same variance. Hereinafter, we use denotation $SNR(\mathbf{A}/\mathbf{B})$ instead of the formulas.

1. Fast ternary coding watermarks

In this section, we consider some fast algorithms for suboptimal ternary coding of audio signals and gray-level images.

1.1. Audio signal. A snippet of the regular sound signal and its step-version appear as shown in Fig. 1.

It follows from Fig. 1 that the sound signal is well balanced, which means that positive and negative values have approximately the same power. Let us consider a fragment $Fragm$ of (2). The most evident ternary code of the sample is as in (5). We selected a threshold Thr and created a new fragment $Appr$

$$Appr[k] = \begin{cases} 0 & \text{if } |Fragm[k]| < Thr \\ \mathbf{sign}(Fragm[k]) & \text{otherwise.} \end{cases} \tag{5}$$

The criterion for optimal approximation of Arr by $Appr$ is (6)

$$\|Arr - C \cdot Appr\| \to \min, \quad \|Arr\| = \|C \cdot Appr\| \tag{6}$$

This criterion is equivalent to SNR maximization by (3). The suboptimal value of Thr meeting criterion (6) can be obtained by changing Thr with a small step. That calculation must be performed for each fragment used for insertion. Such a solution can not

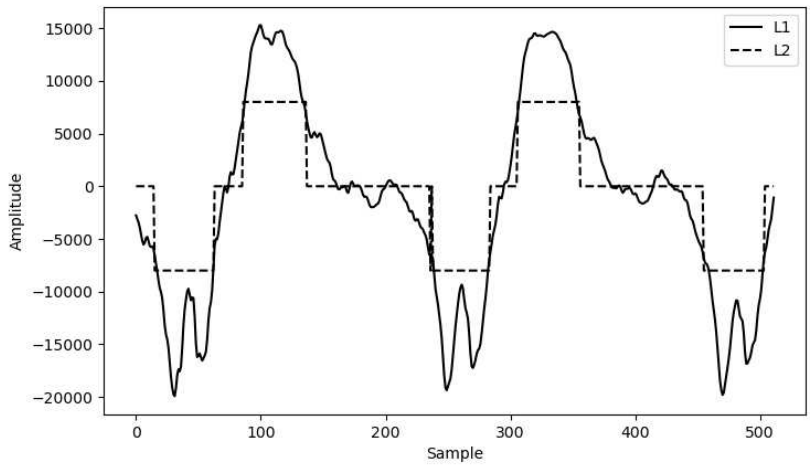


Fig. 1. L1 – a snippet of the regular sound file; L2 – a step-wise version of the snippet

be treated as an acceptable one, because it requires significant resources. In [20], it is shown that a solution that is close to an optimal one can be obtained by calculation employing

$$Thr = Mx \cdot (U \cdot Std/Mx + V), \tag{7}$$

where $Mx = \max(Fragm)$ and $Std = std(|Fragm|)$ is standard deviation. The coefficients U, V depend on the sampling frequency and the length of the fragment. In [20], a method for obtaining these values is proposed. For example, the values $U = 1.2491$, $V = -0.05047$ provide satisfactory results for the sampling frequency of 44100 Hz and the length of the fragments equal to 512.

1.2. Gray-level image. Gray-level image is presented as $M \times N$ matrix $Matr$ with items in $[0, 255]$. Ternary coding is realized analogous to the audio signal. To make the matrix balanced, the source matrix is changed to a new $MatrB = Matr - \mathbf{mean}(Matr)$ by subtracting the mean value. Then, $Appr$ is built using a Thr by means (5), replacing $Fragm$ by $MatrB$. As before, the suboptimal value under the criterion can be obtained by changing Thr with a small step. Nevertheless, in this case, an analog of (7) exists. To this end, let us select some sizes of matrices, for example, 28×28, which is an informal standard for the size of pictures in image recognition. The calculation is performed by Algorithm 1.

Algorithm 1

1: Create some matrices of the size 28×28, dividing various pictures into non-overlapping blocks of that size.
2: Convert each block into a balanced one and obtain a suboptimal threshold by exhaustive search of Thr.
3: Calculate coefficients U, V in (7) using linear regression.

This paper uses $U = 1.31525$, $V = -0.04230$ while calculating the threshold for image processing. A comparison of SNRs obtained by the exhaustive search and using (7) is shown in Fig. 2. One can see that the graphs differ very slightly.

An example of conversion of the gray-level image into a ternary one is shown in Fig. 3.

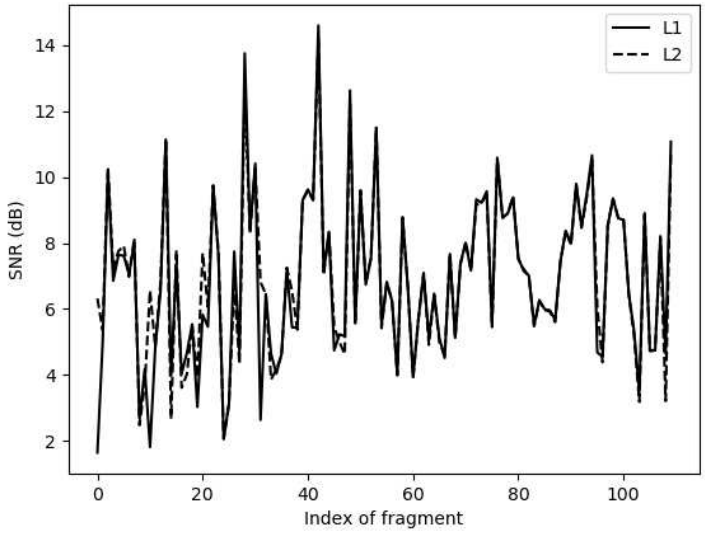
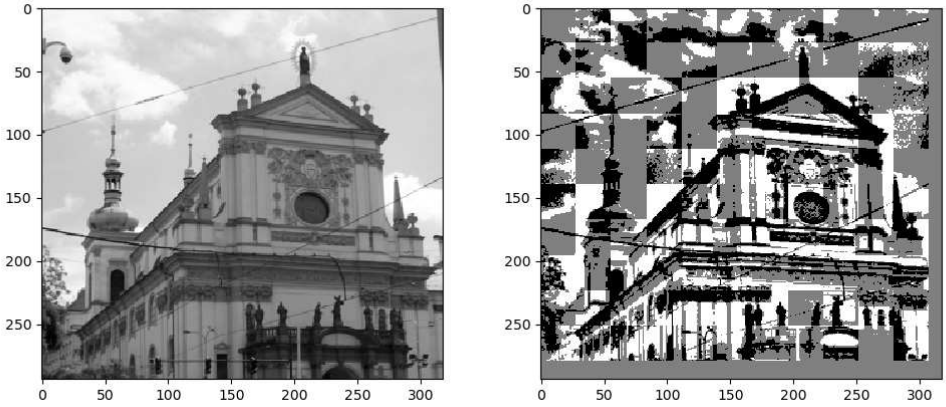


Fig. 2. Comparison of the SNRs obtained by exhaustive search and regression. L1 – exhaustive search, L2 – linear regression



(a) (b)

Fig. 3. Original gray-level image and its ternary representation by 28×28 fragments

2. Embedding ternary watermark into an audio file

Both the melody and image used as watermarks are taken as the ternary sequences while embedding, so we do not consider these watermark embedding procedures separately.

2.1. Amplitude modulation. We use B-splines for amplitude modulation. The only property of B-splines considered in this paper is the graph of the function shown in Fig. 4. The function and its derivatives are zeros at endpoints. Our approach to the problem is presented in [16, 21, 22]. Here, we clarify the main idea. Let $Fragm$ be a fragment of the container destined for embedding a ternary symbol a. We change

$$\mathbf{Fragm} \rightarrow \begin{cases} \mathbf{Fragm_A} = \mathbf{Fragm} + \mathbf{Bspline} \cdot Coef \cdot a \\ \mathbf{Fragm_M} = \mathbf{Fragm} \cdot (\mathbf{1} + \mathbf{Bspline} \cdot Coef \cdot a) \end{cases}$$

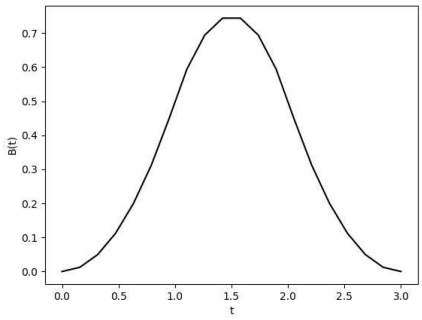


Fig. 4. Graph of B-spline

We call $\mathbf{Fragm_A}$ and $\mathbf{Fragm_M}$ the additive and multiplicative insertion results, respectively. $Coef$ is a coefficient influencing the transparency of the watermark. Choosing $Coef$ is always a trade-off between transparency and resistance to attack. To extract the watermark from $Fragm_A$ or $Fragm_M$, we have to hold $Fragm$. It means that we must keep the pure file within reach during the extraction process. We suggest calling it the non-blind method. This circumstance is not convenient. In [21], a method to avoid this drawback is proposed. The simplest version of this method is as follows. The source fragment $Fragm$ is divided into $Fragm_e$ and $Fragm_o$ consisting of $Fragm$ with even and odd indices. The watermark is embedded into $Fragm_o$, while $Fragm_e$ is used as a pure version of $Fragm_o$.

2.2. Ternary echo hiding. The idea of ternary echo hiding is presented in [22]. The simplest version of the procedure is as follows. Let $Fragm$ be a fragment of length L, destined for embedding a ternary symbol a, belonging to the container. Consider an infinite impulse response (IIR) filter with the transfer function

$$Tr(n,p,Coef,L) = \frac{1 + Coef \cdot a \cdot \exp(-w \cdot p \cdot n/L)}{1 - Coef \cdot a \cdot \exp(-w \cdot p \cdot n/L)}, \quad w = 2 \cdot \pi \cdot i.$$

Inverse discrete Fourier transform (IDFT) of $\mathbf{Tr}$ has a splash at point p. Let $\mathbf{FFragam}$ be a discrete Fourier transform of $\mathbf{Fragm}$. The embedding procedure is reduced to replacement of $\mathbf{Fragm}$ for $\mathbf{MFragm} = IDFT(\mathbf{FFragm} \cdot \mathbf{Tr})$. For extraction, one calculates cepstrum

$$\mathbf{Ceps} = IDFT(\log|\mathbf{FFragm} \cdot \mathbf{Tr}|).$$

If $a \neq 0$, then $Ceps$ has a splash at position p. The signs of the splash and a are the same.

3. Partial restoration of the signal from its ternary code

This section presents a method for partial restoration of the signal through its ternary code (5), which is possible only if a piece of additional information about the signal is utilized. If we know the method for enhancing the perception of a media object extracted from the container, we increase the resistance of the embedding procedure to attacks that corrupt the containers.

Here, we present the exact definition of the enhancing procedure. Let $Sign$ be a signal and $Appr$ its ternary approximation. Let us calculate $Val = SNR(\mathbf{Sign}/\mathbf{Appr})$. We convert $Appr$ into a new sequence $NewAppr$ by applying a certain algorithm. If $Val < SNR(\mathbf{Sign}/\mathbf{NewAppr})$ is fulfilled, the goal is achieved.

Our approach is based on discrete cosine transform (DCT), although any other orthogonal transform can be used. Let us recall the matrix of DCT of order P:

$$Matr[i,j] = \cos(i \cdot (j+0.5) \cdot \pi/P), \quad i,\, j = 0,1,\ldots,P-1. \tag{8}$$

Formally, $Matr$ becomes orthogonal only after the normalization of rows, but we use the standard definition (8). It is clear that $Matr[0,j] = 1,\ j = 0,\ldots,P-1$ so

$$\sum_j Matr[i,j] = 0, \quad i \neq 0. \tag{9}$$

Let $Sign = s_0, s_1, \ldots, s_{N-1}$ be a sequence. Let

$$F_j(k) = \sum_{t=0}^{P-1} Matr[t,j] \cdot s_{k+t}. \tag{10}$$

The values $F_j(k)$ have the following sense. A window of the length P slides along $Sign$, and k denotes the window's start position at $Sign$. For any k, the values of $F_j(k)$, $j = 0,1,\ldots,P-1$ up to factor are coefficients of the inverse DCT of the sequence inside of the window. It follows from (9), (10) that

$$\sum_j F_j(k) = P \cdot s_k. \tag{11}$$

For our future goals, it is more convenient to utilize another form of (11). Let B_j be the $M[*,j]$ record in inverse order and be $Filter_j$ the final impulse response (FIR) filter with coefficients B_j. In these notations,

$$F_j(k) = Filter_j(s_0, s_1, \ldots, s_{N-1})[k]$$

and

$$\sum_j Filter_j(s_k, s_{k+1}, \ldots, s_{k+P-1})[0] = P \cdot s_k.$$

Consider the following problem. Let $Tern = d_0, d_1, \ldots, d_{N-1}$ be a ternary version of the signal $Sign$. Let $C_0, C_1, \ldots, C_{R-1}$, $R \leq P$ be coefficients and

$$Filter = \sum_{j=0}^{R-1} C_j \cdot Filter_j$$

Let us create a sequence

$$Rest = f_0, f_1, \ldots, f_{N-1} = Filter(d_0, d_1, \ldots, d_{N-1}).$$

and solve the task

$$SNR(Sign/Rest) \to \max. \tag{12}$$

We use only R first column of the matrix (8), because the corresponding filters save the most informative low part of spectra. A rough solution of (12) can be found. We use the package [23] to this end.

The next step is the choice of parameter P in (8). There are principal drawbacks of the restoration sources by the described schema. It follows from Fig. 1 that the step-function corresponding to the source signal can have constant values on the interval of significant length. It means that the window sliding along the step function has the form

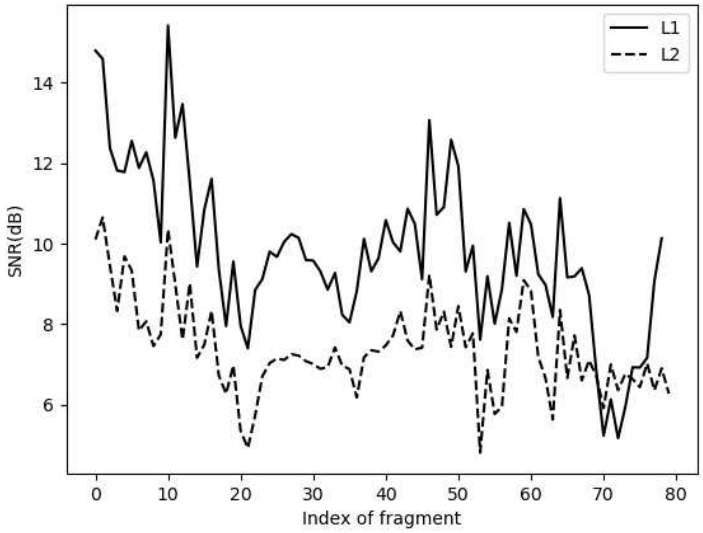


Fig. 5. Wave file as a watermark. Compare of SNRs of the approximation speech file by step function and by filtered step function. L1 – filtered step function, L2– step function. Fragment length = 512, $P = 31$, $R = 16$

$s, s, \ldots, s$, where $s \in \{-1, 0, 1\}$ and the restored sequence *Rest* also has constant values at points corresponding to such intervals. One possible solution is to choose P higher than the maximal length of the constant intervals. This choice makes the optimization procedures require computer resources. Instead, we propose another solution consisting in changing the FIR filters for the IIR filter by a serial connection of the FIR filter and IIR filter of the first order having the transfer function

$$Trans_1[w] = \frac{1}{1 + Q \cdot \exp(-2 \cdot \pi \cdot i \cdot w)}.$$

The best values of Q and parameters of the filters must be selected depending on the object under research. One example of our technique's effectiveness is the sound file used as a watermark and shown in Fig. 5. Following the filtering, the quality of approximation increases by about 4 dB.

4. Experiments

In this section, we present some experimental results substantiating the suggested technique. We use a picture as a watermark and a speech file sampled with the frequency 44100 Hz as a container. The source picture is converted into a ternary form inserted into the container by amplitude modulation or echo hiding, extracted, and filtered by the filter based on the DCT matrix rows. All produced pictures are compared with the original by SNRs. Fig. 6 shows an example of watermark insertion in ideal conditions without an attack. One can see that some error occur during the extraction. In the restored image, most errors are suppressed. Here, we investigate the watermark resistance to two kinds of attacks: container filtering, container conversion into mp3 format.

4.1. Amplitude modulation. Since the container must retain its customer properties, we consider filtering the container with the help of the low-pass FIR filters

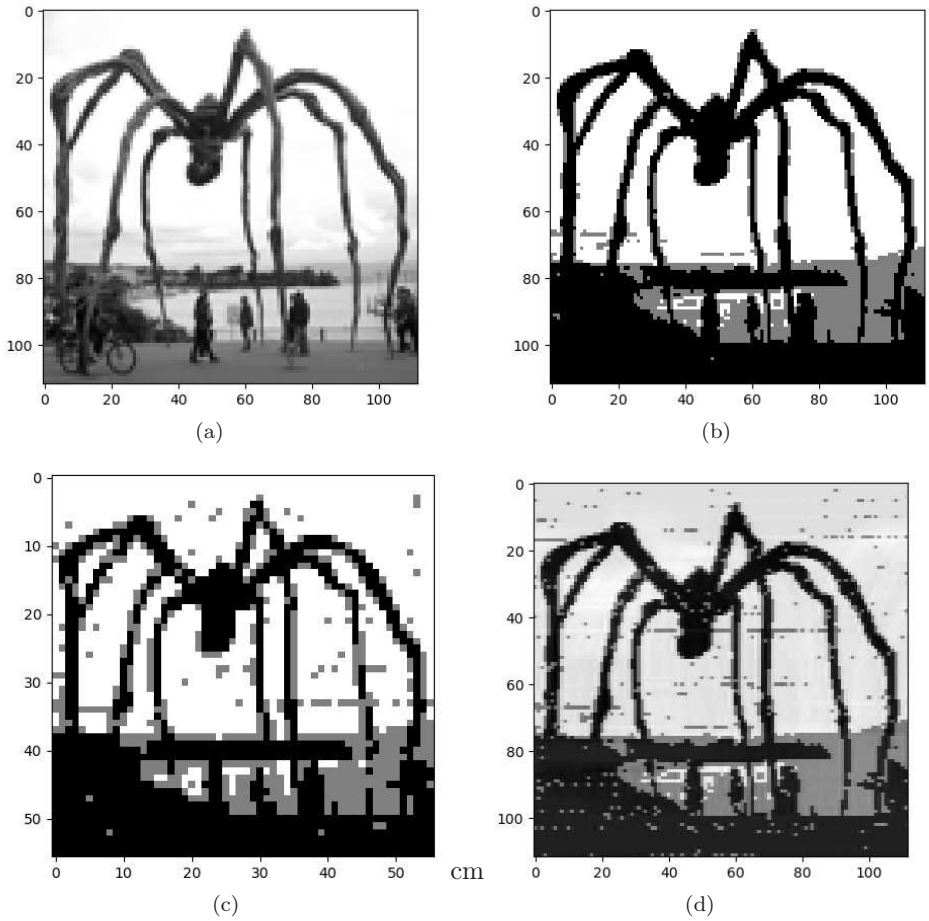


Fig. 6. Original gray-level image (a); its ternary representation (b), $SNR = 9.01$; its extracted image (c), $SNR = 8.32$; restored by filter image (d), $SNR = 9.70$

Table 1. Degradation (in dB) of the restored watermark depending on the FIR filter length. Non-blind version

FIR filter length	Extracted step function	Enhanced image
3	6.70	8.38
5	6.62	8.36
7	6.41	8.28
9	6.16	8.18
19	5.07	8.09

of different lengths having a stopband frequency of 19 kHz [23]. The results are given in Table 1.

The resistance to *mp*3 attack is investigated by employing the package PyDub [24]. The *wav* format container is transformed into *mp*3 format with various bitrate and then converted into *wav* format again. The initial bitrate of the container is 704 kbps. A blind version of the amplitude modulation can not be resistant to *mp*3 attack, whereas a non-blind version resists it, and SNRs are not determined by the bitrate. The extracted step function has $SNR = 6.83$ and the enhanced version has $SNR = 8.35$.

Table 2. Degradation (in dB) of the restored watermark depending on the FIR filter length. Ternary echo hiding version

FIR filter length	Extracted step function	Enhanced image
3	–0.4	3.85
5	–0.23	3.76
7	–0.15	3.54
9	–0.12	3.53

Table 3. Degradation (in dB) of the restored watermark depending on the bitrate used by transformation into mp3 format. Ternary echo hiding version

Bitrate (kbps)	Extracted step function	Enhanced image
704	5.93	8.21
500	5.93	8.21
400	5.93	8.21
200	5.71	8.20
100	4.62	7.53
50	3.26	6.94

4.2. Ternary echo hiding. We investigate the simplest version of ternary echo hiding, i.e., when only a single *trit* is inserted into the fragment at the position equal to half of the fragment length. The resistance to the filtering attack is shown in Table 2. The properties of the enhanced image after mp3 attack are provided in Table 3.

Acknowledgments. This study was supported by the Kazan Federal University Strategic Academic Leadership Program.

References

1. Hua G., Huang J., Shi Y.Q., Goh J., Thing V.L.L. Twenty years of digital audio watermarking – a comprehensive review. *Signal Process.*, 2016, vol. 128, pp. 222–242. doi: 10.1016/j.sigpro.2016.04.005.
2. Chauhan S., Rizvi S. A survey: Digital audio watermarking techniques and applications. *Proc. 2013 4th Int. Conf. on Computer and Communication Technology (ICCCT), 20–22 Sept., 2013.* Allahabad, India, IEEE, 2013, pp. 185–192. doi: 10.1109/ICCCT.2013.6749625.
3. Bajpai J., Kaur A. A literature survey – various audio watermarking techniques and their challenges. *Proc. 2016 6th Int. Conf. – Cloud System and Big Data Engineering (Confluence), 14–15 Jan., 2016.* Noida, India, IEEE, 2016, pp. 451–457. doi: 10.1109/CONFLUENCE.2016.7508162.
4. Xiang Y., Hua G., Yan B. *Digital Audio Watermarking: Fundamentals, Techniques, and Challenges.* Springer Singapore, 2017. xii, 90 p. doi: 10.1007/978-981-10-4289-8.
5. Thanki R. *Advanced Techniques for Audio Watermarking.* Springer Int. Publ., 2020. xv, 101 p. doi: 10.1007/978-3-030-24186-5.
6. Bassia P., Pitas I., Nikolaidis N. Robust audio watermarking in the time domain. *IEEE Trans. Multimedia*, 2001, vol. 3, no. 2, pp. 232–241. doi: 10.1109/6046.923822.
7. Xiang Y., Natgunanathan I., Peng D., Hua G., Liu B. Spread spectrum audio watermarking using multiple orthogonal PN sequences and variable embedding strengths and

polarities. *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 2018, vol. 26, no. 3, pp. 529–539. doi: 10.1109/TASLP.2017.2782487.

8. Wang S., Yuan W., Wang J., Unoki M. Inaudible speech watermarking based on self-compensated echo-hiding and sparse subspace clustering. *Proc. 2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP).* Brighton, U. K., IEEE, 2019, pp. 2632–2636. doi: 10.1109/ICASSP.2019.8682352.
9. Swanson M.,D., Zhu B., Tewfik A.H., Boney L. Robust audio watermarking using perceptual masking. *Signal Process.*, 2017, vol. 66, no. 3, pp. 337–355. doi: 10.1016/S0165-1684(98)00014-0.
10. Pardhu T., Perly B. Digital image watermarking in frequency domain. *Proc. 2016 Int. Conf. on Communication and Signal Processing (ICCSP), 6–8 Apr., 2016.* Melmaruvathur, India, IEEE, 2016, pp. 208–211. doi: 10.1109/ICCSP.2016.7754123.
11. Erfani Y., Pichevar R., Rouat J. Audio watermarking using spikegram and a two-dictionary approach. *IEEE Trans. Inf. Forensics Secur.*, 2017, vol. 12, no. 4, pp. 840–852. doi: 10.1109/TIFS.2016.2636094.
12. Subir, Joshi A.M. DWT-DCT based blind audio watermarking using Arnold scrambling and Cyclic codes. *Proc. 2016 3rd Int. Conf. on Signal Processing and Integrated Networks (SPIN), 11–12 Feb., 2016.* Noida, India, IEEE, 2016, pp. 79–84. doi: 10.1109/SPIN.2016.7566666.
13. Hwang M.J., Lee J., Lee M., Kang H.G. SVD-based adaptive QIM watermarking on stereo audio signals. *IEEE Trans. Multimedia*, 2018, vol. 20, no. 1, pp. 45–54. doi: 10.1109/TMM.2017.2721642.
14. Budiman G., Suksmono A., Danudirdjo D., Pawellang S. QIM-based audio watermarking with combined techniques of SWT-DST-QR-CPT using SS-based synchronization. *2018 Proc. 6th Int. Conf. on Information and Communication Technology (ICoICT), 3–5 May 2018.* Bandung, Indones., IEEE, 2018, pp. 286–292. doi: 10.1109/ICoICT.2018.8528727.
15. Absalyamova K.S., Latypov R.Kh., Stolov E.L. Ternary code of melody and reliable audio watermarking. *Proc. 2019 27th Telecommun. Forum (TELFOR), 26–27 Nov. 2019.* Belgrad, Serbia, IEEE, 2019, pp. 1–4. doi: 10.1109/TELFOR48224.2019.8971187.
16. Latypov R.Kh., Stolov E.L. Ternary picture as watermark for audio files. *Proc. 2020 3rd Int. Conf. on Computer Applications & Information Security (ICCAIS), 19–21 March, 2020.* Er-Riyadh, Saudi Arabia, IEEE, 2020, pp. 1–6. doi: 10.1109/ICCAIS48893.2020.9096786.
17. Unser A. Splines: A perfect fit for signal and image processing. *IEEE Signal Process. Mag.*, 1999, vol. 16, no. 6, pp. 22–38. doi: 10.1109/79.799930.
18. Bender W., Gruhl D., Morimoto N., Lu A. Techniques for data hiding. *IBM Syst. J.,* 1996, vol. 35, no. 3.4, pp. 313–336. doi: 10.1147/sj.353.0313.
19. Hua G., Goh J., Thing V.L.L. Cepstral analysis for the application of echo-based audio watermark detection. *IEEE Trans. Inf. Forensics Secur.*, 2015, vol. 10, no. 9, pp. 1850–1861. doi: 10.1109/TIFS.2015.2431997.
20. Latypov R., Stolov E. Speaker diarization based on speech signal approximation by step-function. *Proc. 28th IEEE Conf. of Open Innovations Association FRUCT (FRUCT28)*, 2021, pp. 598–604. doi: 10.5281/zenodo.4514965.
21. Latypov R.Kh., Stolov E.L. Ternary coded melody as blind audio watermark. *Telfor J.*, 2020, vol. 12, no. 1, pp. 28–33. doi: 10.5937/telfor2001028L.
22. Latypov R.Kh., Stolov E.L. Ternary echo hiding in audio files. *Proc. 2020 28th Telecommun. Forum (TELFOR), 24–25 Nov., 2020.* Belgrad, Serbia, 2020, pp. 1–4. doi: 10.1109/TELFOR51502.2020.9306575.

23. Virtanen P., Gommers R., Oliphant T.E., Haberland M., Reddy T., Cournapeau D., Burovski E., Peterson P., Weckesser W., Bright J., van der Walt S.J., Brett M., Wilson J., Millman K.J., Mayorov N., Nelson A.R.J., Jones E., Kern R., Larson E., Carey C.J., Polat I., Feng Y., Moore E.W., VanderPlas J., Laxalde D., Perktold J., Cimrman R., Henriksen I., Quintero E.A., Harris Ch.R., Archibald A.M., Ribeiro A.H., Pedregosa F., van Mulbregt P., SciPy 1.0 Contributors. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods*, 2020, vol. 17, pp. 261–272. doi: 10.1038/s41592-019-0686-2.

24. Robert J., Webbie M. et al. PyDub. 2018. Available at: http://pydub.com/.

Received
January 19, 2021

Latypov Roustam Khafizovich, Doctor of Technical Sciences, Professor, Head of Department of System Analysis and Information Technologies

Kazan Federal University
ul. Kremlevskaya 18, Kazan, 420008 Russia
E-mail: *roustam.latypov@kpfu.ru*

Stolov Evgeny L'vovich, Doctor of Technical Sciences, Professor, Leading Research Fellow of Research Laboratory of Computational Technologies and Computer Modeling

Kazan Federal University
ul. Kremlevskaya, 18, Kazan, 420008 Russia
E-mail: *ystolov@list.ru*

ОРИГИНАЛЬНАЯ СТАТЬЯ

УДК 003.26 doi: 10.26907/2541-7746.2021.1.77-89

Новый метод увеличения устойчивости троичных водяных знаков в аудио-файлах с помощью фильтров на основе дискретного косинус-преобразования

Р.Х. Латыпов, Е.Л. Столов

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

Аннотация

Использование цифровых водяных знаков (ЦВЗ) для доказательства авторства различных произведений – это простое средство, доступное начинающим авторам. В этой статье мы предлагаем новый подход для повышения устойчивости водяных знаков в аудио-файлах к различным атакам. Мы используем в качестве водяных знаков отрывки хорошо известных мелодий или фрагменты фотографий. Такие водяные знаки могут быть легко распознаны человеком даже после существенной деградации. Увеличивая качество извлеченного ЦВЗ, мы облегчаем задачу распознавания. Исходный водяной знак кодируется троичным кодом, в результате чего происходит его деградация. После атаки на контейнер качество еще больше ухудшается. Используя линейную комбинацию фильтров, построенных из строк матрицы дискретного косинус-преобразования, удается существенно улучшить качество водяного знака. Мы рассматриваем два вида наиболее распространенных атак на аудио-файлы: фильтрация и перевод файла в mp3-формат.

Ключевые слова: цифровые водяные знаки, троичное кодирование, аудио файлы, дискретное косинус преобразование

Поступила в редакцию
19.01.2021

Латыпов Рустам Хафизович, доктор технических наук, профессор, заведующий кафедрой системного анализа и информационных технологий

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *roustam.latypov@kpfu.ru*

Столов Евгений Львович, доктор технических наук, профессор, ведущий научный сотрудник НИЛ «Вычислительные технологии и компьютерное моделирование»

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *ystolov@list.ru*

⟨ ***For citation*:** Latypov R.Kh., Stolov E.L. A new DCT filters-based method to improve the resistance of ternary watermarks in audio files against attacks. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2021, vol. 163, no. 1, pp. 77–89. doi: 10.26907/2541-7746.2021.1.77-89. ⟩

⟨ ***Для цитирования*:** *Latypov R.Kh., Stolov E.L.* A new DCT filters-based method to improve the resistance of ternary watermarks in audio files against attacks // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2021. – Т. 163, кн. 1. – С. 77–89. – doi: 10.26907/2541-7746.2021.1.77-89. ⟩