An Audio Zero-watermark Algorithm Combined DCT with Zernike Moments

XIONG Yi-qun, WANG Rang-ding
(*CKC software lab, Ningbo University, Ningbo, 315211, China)*
Email: xiongyiqundcm@yahoo.com.cn

Abstract

This work proposes an audio zero-watermark algorithm which combined DCT and Zernike Moments. The proposed algorithm combines a part of samples and low-frequency DCT coefficients of the same volume to perform low-order Zernike transform, and then construct watermark based on Zernike moments. Experimental results show that the proposed algorithm can resist both common signal processing and geometrical attacks effectively.

1. Introduction

Audio digital watermarking technology can be used for various applications such as copyright protection, copy control, data authentication, broadcast monitoring, data indexing and so forth[1−4]. An effective audio digital watermarking system must satisfy robustness and the imperceptibility and Zero-watermarking technology solves perceptibility basically. Recent years, available studies on audio watermarking is far less than that of image watermarking or video watermarking. In particular, robust audio watermarking algorithm which aims at effectively resisting desynchronization attacks is extremely scarce.

In[5], Darko Kirovski and Hagai Attias explore two technologies, beat detection and block redundant coding, to combat de-synchronization and watermark estimation as two attacks that have demonstrated superiore effectiveness in preventing watermark detectors from reliably accomplishing their goal. In[6], Robert B¨auml, etc present a channel model for watermark reception with nonperfect synchronization. In[7], By analyzing and deducting the linear relationship between the audio amplitude and Zernike moments, watermarking the low-order moments is achieved in time domain by scaling sample values directly, the degradation in audio reconstruction is avoided. However, it will bring signal distortion and sample overflowing to some extent after scaling sample values.

Zernike moments have characteristics of rotation invariance and not sensitive to noise, which has been widely used in image watermarking[8−11], while used infrequently in audio watermarking. In this paper, we propose an effective audio zero-watermark algorithm based on combination of DCT and Zernike moments, select a part of samples and low-frequency DCT coefficients of the same volume to perform low-order Zernike transform, and then construct watermark based on Zernike moments. Both theoretical analysis and experimental results demonstrate the superiority of the proposed algorithm.

2. Zernike Moments

Zernike introduced a set of complex polynomials which form a complete orthogonal set over the interior of the unit circle, i.e., $x2+y2=1$. Let the set of these polynomials be denoted by $V_{nm}(x,y)$

$$V_{nm}(x,y)=V_{nm}(\rho,\theta)=R_{nm}(\rho)\cdot\exp(jm\theta) \tag{1}$$

Where n is positive integer or zero, m is an integer subject to constraints $(n-|m|)$ is nonnegative and even, ρ is the length of vector from origin to (x,y) pixel, θ is the angle between vector ρ and x axis in counterclockwise. $R_{nm}(\rho)$ is radial polynomial, defined as:

$$R_{nm}(\rho)=\sum_{s}^{n-|m|/2}(-1)^s\cdot\frac{(n-s)!}{s!\cdot\left(\frac{n+|m|}{2}+s\right)!\cdot\left(\frac{n-|m|}{2}+s\right)!}\cdot\rho^{n-2s} \tag{2}$$

Note that, $R_{n,m}(\rho)=R_{n,-m}(\rho)$, So

$$V_{n,-m}(\rho,\theta)=R_{n,-m}(\rho)\exp(-jm\theta)=R_{nm}(\rho)\exp(-jm\theta)=V^*_{nm}(\rho,\theta) \tag{3}$$

These polynomials are orthogonal and satisfy:

$$\iint_{x^2+y^2\le 1}\left[V_{nm}(x,y)\right]^*\cdot V_{pq}(x,y)\,dxdy=\frac{\pi}{n+1}\delta_{np}\delta_{mq} \tag{4}$$

978-0-7695-3381-0/08 $25.00 © 2008 IEEE
DOI 10.1109/CW.2008.24

With $\delta_{np} = \begin{cases} 1 & n = p \\ 0 & n \neq p \end{cases}$ Zernike moments are the projection of the function onto these orthogonal basis functions. The Zernike moment of order n with repetition m for a continuous 2-dimensional function $f(x, y)$ that vanishes outside the unit circle is

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y) \cdot V^*_{nm}(x, y)\, dx dy \tag{5}$$

For a 2-dimensional digital signal, like digital image, usually serve the image center as origin of unit circle, map image pixels to a unit circle, which is $x^2 + y^2 \leq 1$, while vanishes outside the unit circle. The moments are replaced by summations to

$$A_{nm} = \frac{n+1}{\pi} \sum_{n=0}^{+\infty} \sum_{m} f(x, y) \cdot V^*_{nm}(x, y) \tag{6}$$

3. Watermark forming

Audio low-order Zernike moments[7,12] capture basic shape of audio signal and represent low-frequency components, which is robust to many common signal processing such as resampling, low-pass filtering, etc. While high-order Zernike moments indicate high-frequency components, in this paper order is given as 5, audio zero-watermark forming as follows:

- Split original audio x into N segments data according to Figure.1, each segment data has L samples. And then DCT is performed on each audio segment, generating DCT coefficients x' with length NL as shown in Figure.2.

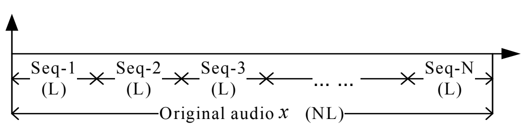


Figure.1 N consecutive audio segments

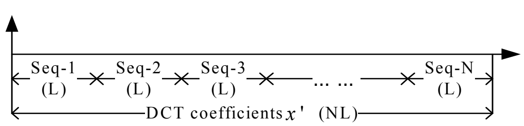


Figure.2 N consecutive DCT coefficients segments

Select the first segment data in digital audio, namely $x(0:1L)$, at the same time select the first segment data in DCT coefficients, namely $x'(0:1L)$, compute h by combining the two, that is

$$h = x'(0:1L) + 2 * x(0:1L) \tag{7}$$

Where h is a 1-dimensional sequence with length L.

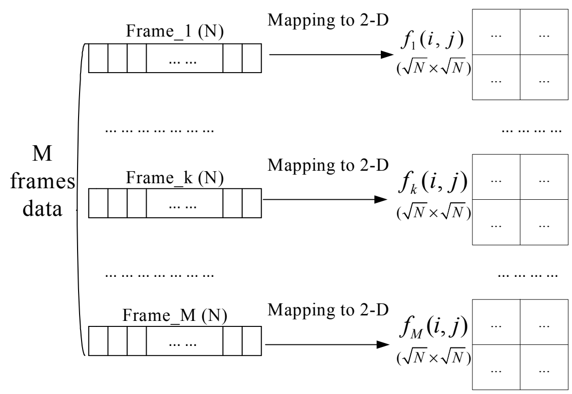
- h is divided into M frames with length N, regulation of dividing frames as follows:

$$L = M \times N + R \tag{8}$$

Where M is frame number、 N is the length of each frame subject to constraints $(l \times l = N)$ with l is positive integer, R is the remaining length of the audio as much as possible to zero.

- The M frames data is performed Zernike transform respectively, however, Zernike moments work only for 2-dimensional data, in this paper, M frames 1-dimensional sequences are mapped to corresponding 2-dimensional signals respectively, shown as follows:

Figure.3 2-dimensional mapping of M frames data



As shown in Figure.3, Frame-k is the k^{th} frame data, $f_k(i, j)$ is 2-dimensional signals by mapping the k^{th} frame data, its size satisfy $(\sqrt{N} \times \sqrt{N})$, with $(1 \leq k \leq M$, $1 \leq i \leq \sqrt{N}$, $1 \leq j \leq \sqrt{N})$.

- Calculate Zernike moment of each frame respectively, The formula as follows

$$A_{nm(k)}(l) = \frac{n+1}{\pi} \sum_{n=0}^{+\infty} \sum_{m} f_k(i, j) \cdot V^*_{nm(k)}(i, j)$$
$$(1 \leq k \leq M, 1 \leq l \leq N') \tag{9}$$

Where $A_{nm(k)}(l)$ is Zernike moments of the k^{th} frame, it is a 1-dimensional vector. the length of l is related with choices of repetition m and order n, in this paper, N'—the length of l is 100.

- A matrix with the size $(M \times 100)$ is combined by Zernike moments of M frames data, as follows

$$A_{nm} = \begin{bmatrix} A_{nm(1)}(1, 2, \cdots l \cdots 100) \\ A_{nm(2)}(1, 2, \cdots l \cdots 100) \\ \ldots\ldots \\ A_{nm(k)}(1, 2, \cdots l \cdots 100) \\ \ldots\ldots \\ A_{nm(M)}(1, 2, \cdots l \cdots 100) \end{bmatrix} \tag{10}$$

And then transpose the matrix A_{nm} to gain a matrix A_{nm}^T with size $(100 \times M)$, respectively calculate the sum of each row vector's modulus in A_{nm}^T, that is

$$s(l) = sum(abs(A_{nm}^{T}(l))) \tag{11}$$

Where $A_{nm}^{T}(l)$ is the l^{th} row vector in A_{nm}^{T}, $s(l)$ is the sum of the l^{th} row vector' modulus in A_{nm}^{T}.

- According to equation (11), s is a 1-dimensional sequence, which is used as reference to construct watermark as follows:

$$w(l)=\begin{cases}1, s(2\times l-1)-s(2\times l)\geq 0\\ 0, s(2\times l-1)-s(2\times l)<0\end{cases} \quad (1\leq l\leq N'/2) \tag{12}$$

4. Watermark detecting

Watermark detected scheme, is very similar with formed scheme, simple steps as follows:

- Detected audio y is performed DCT to get DCT coefficients y',select L samples in y, select DCT coefficients with length L in y' as well, get H,

$$H = y'(0:1L) + 2 * y(0:1L) \tag{13}$$

- H is divided into M frames with length N ,each frame is mapped into corresponding 2-dimensional data, M segments 2-dimensional data is performed Zernike transform respectively, we will get a Zernike matrix, and then transpose, respectively calculate the sum of each row vector's modulus in $A1_{nm}^{T}$, as follows:

$$s1(l) = sum(abs(A1_{nm}^{T}(l))) \tag{14}$$

Where $s1(l)$ is the sum of the l^{th} row vector' modulus in $A1_{nm}^{T}$.

- Zero-watermark extracting as follows:

$$w1=\begin{cases}1, s1(2\times l-1)-s1(2\times l)\geq 0\\ 0, s1(2\times l-1)-s1(2\times l)<0\end{cases} \quad (1\leq l\leq N'/2) \tag{15}$$

- Make use of correlation value (NC) to detect watermark and evaluate the similarity between original watermark and the extracted watermark, defined as

$$NC(w, w1) = \frac{\sum_{i=1}^{n} w(i) w1(i)}{\sqrt{\sum_{i=1}^{n} w(i)^2 \times \sum_{i=1}^{n} w1(i)^2}} \tag{16}$$

Where w is original watermark, $w1$ is extracted watermark.

5. Experimental results and analysis

This experiment adopts a set of diverse types of audio signals (20s, mono, 16bits/sample, 44.1 KHz and WAVE format) including blues, classical, country, folk, pop. According to proposed watermark formed scheme, in this experiment, we adopt $L = 22500$, divide the audio into 100 frames with length 255, each frame is mapped into (15×15) 2-dimensional data, and then, Zernike transform is performed on each 2-dimensional data segment with order 5.

Set the threshold of correlation value NC as 0.75. In order to justify the value 0.75, choose five types of audio works randomly to construct watermark, and then compute correlation value NC each other, the results as shown in Figure.4. From Figure.4, NC are all less than 0.55, lager than 0.4, indicates threshold value NC set reasonably. If the correlation value in the watermark detected scheme is greater than the threshold value 0.75, indicate the watermark is right.

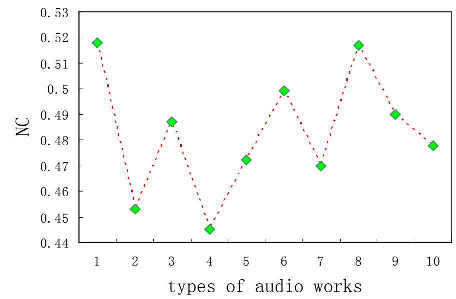


Figure.4 Similarity of diverse audio works

In order to test effectiveness of the proposed algorithm, select blues format audio, a series of following attacks tests are implemented in terms of the robustness. The audio editing and attacking tools adopted are CoolEditPro v2.0, GoldWave v4.25 and Stirmark Benchmark for Audio v0.2.

1). Noise addition. White Gaussian Noise is added to the audio signal to give 26 dB SNR.
2). Low-pass filtering. Adopt a six order Butterworth filter with cut-off frequencies 22.05 KHz.
3). Mp3 compression. The audio signals are compressed into MP3 format by MPEG-1 Layer-3 (LAME 3.92) encoder and decoded, compression rate is 128kbps.
4). Resampling. Consist of subsequent down and up sampling to 22.05 kHz and 44.1 kHz, respectively.
5). Requantization. Audio signal is quantified from the 16 bit to 8 bit or 32 bit respectively, and then quantified to 16 bit again.
6). Cropping. Cut 10%, 30% section of digital audio in the front (or behind).
7). Volume. Volume increased by 50% -150%.
8). AddBrumm. Add buzz as sinus tone to the audio signal. The parameter values are: Frequency = 55 and Strength = 2500.
9). Addsinus. Add a sinus tone to the audio signal. The parameter values are: Frequency = 3000 and Strength = 120.
10). FFt-stat1. Compute a mathematical signal modification in the frequency domain and have no parameteters.
11). FFt-test. Add test attacks to signal in the

frequency domain and have no parameters.

Table.1 robustness test: detection results for various attacks (BER (%))

Types of attack	NC	BER	Types of attack	NC	BER
Noise addition	1.0000	0	Volume	1.0000	0
Low-pass filering	0.9789	2	AddBrumm_100	1.0000	0
Mp3 compression	0.9106	8	AddBrumm_1100	0.9892	1.6
Resampling	1.0000	0	AddBrumm_2100	0.9100	8
Requantization(8bit、32bit)	1.0000	0	Addsinus	0.9336	8
Cropping（10%）	1.0000	0	FFt_stat1	0.8956	10
Cropping（30%）	1.0000	0	FFt_test	0.8759	12

Table 1 shows the robustness test's detection results for the various attacks. We obtained a perfect detection result for most of attacks. We also get robustness by testing another four types of audio signals. Firstly, Zernike moments is complete orthogonal set, which is greatly robust to geometric attacks such as cropping and can also better resist a number of attacks in frequency domain. Secondly, the proposed algorithm adopts low-order Zernike moments to construct watermark, which represent low-frequency components, can resist effectively attacks such as low-pass filtering、resampling、requantization、addsinus 、AddBumm, etc.

6. Conclusion

In this paper, we propose a superior audio zero-watermark algorithm combined DCT with Zernike moments. select both low-frequency DCT coefficients and the original audio sample values, calculate summation of the two, divide frames, perform Zernike transform on each frame respectively, robust watermark is constructed based on Zernike moments. Experiment results show that the proposed algorithm is strongly robust to many attacks such as low-pass filtering、resampling、requantization、noise addition 、cropping, etc, at the same time, Zero-watermarking technology offers a good solution for the contradiction between robustness and transparency. Otherwise, the proposed algorithm has characteristics of lower computational complexity, easy implement.

Acknowledgements.

This work is supported by the National Natural Science Foundation of China(No. 60672070) and Ningbo Science & Technology Project of China（2006B100067）.

References

[1] M. Mansour, A. Tewfik. Time-scale invariant audio data embedding [A]. *Proceedings of IEEE International Conference on Multimedia and Expo*, ICME[C]. 2001. 76-79.

[2] Petitcolas F A P, Anderson RJ. Information hiding-A survey J. *Proceedings of the IEEE*, 1999, 87(7):1062-1077.

[3] A N Lemma，et a1．A Temporal Domain Audio Watermarking Technique [J．*IEEE Trans．on Signal Processing*，2003，51 (4)：1088— 1097．

[4] B. Chen and G. W. Wornell. Provably robust digital watermarking. *In Proceedings of SPIE*: Multimedia Systems and ApplicationsII (part of Photonics East'99), volume 3845, pages 43–54, Boston, MA, USA, September1999.

[5] Darko Kirovski, Hagai Attias . Audio watermark robustness to desynchronization via beat detection[J],*computer science ISSN 0302-9743, International workshop on information hiding* NO.5, 2003, vol. 2578, pp. 160-176.

[6] Robert B¨auml，Joachim J. Eggers and Johannes Huber. A Channel Model for Watermarks Subject to Desynchronization Attacks [A]. *4th Intl. ITG Conferenceon Source and Channel Coding Berlin*[C], Jan. 28-30, 2002. Vol.4675，pp.281-292.

[7] Shijun Xiang, Jiwu Huang. Robust audio watermarking based on low-order Zernike moments [A]. *IWDW* 2006, LNCS 4283[C], 2006, pp. 226-240.

[8] H.S.Kim and H.K.Lee:Invariant Image Watermark Using Zernike Moments. *IEEE Transaction on Circuits and Systems for Video Technology*,Vol.13,No.8,(2003)766-775.

[9] Y.Q.Xin,Simon Liao and Miroslaw Pawlak: A Multibit Geometrically Robust Image Watermark Based on Zernike Moments. *Proc. of the 17th International Conference on Pattern Recognition*,(2004) 861-864.

[10] J.Chen, H.X.Yao, W.G. and S.H.Liu: A Robust Watermarking Method Based on Wavelet and Zernike Transform. *Proc. of the 2004 International Symposium on Circuits and Systems*,Vol.2 (2004) 23-26.

[11] M. Farzam, S.Shahram Shirani. A robust multimedia watermarking technique using Zernike Transform. *Proc. of IEEE International Workshop Multimedia Signal Processing*, 2001, 529–534.

[12] Xiang Shijun. Robust Audio Watermarking Algorithms. *Sun Yat-sen University doctoral dissertation*, 2006.