한국컴퓨터정보학회 동계학술대회 논문집 제22권 제1호 (2014. 1)

경험적 모드 분해법을 이용한 오디오 워터마킹

Phuong Nguyen°, 김종면*
°*울산대학교 전기공학부
e-mail: : phuongnguyen.cse°@gmail.com, jmkim07@ulsan.ac.kr*

Audio Watermarking Using Empirical Mode Decomposition

Phuong Nguyen°, Jong-Myon Kim*
°*Dept. of Electrical Engineering, University of Ulsan

● 요 약 ●

This paper presents a secure and blind adaptive audio watermarking algorithm based on Empirical Mode Decomposition (EMD). The audio signal is divided into frames and each one is decomposed adaptively, by EMD, into several Intrinsic Mode Functions (IMFs). The watermark and the synchronization codes are then embedded into the extrema of the last IMF. The experimental results show that the proposed method has good imperceptibility and robustness against signal processing attacks.

키워드: Audio watermarking, empirical mode decomposition, synchronization code

I. Introduction

Recent advances in Internet and digital multimedia technology have allowed transmission and distribution of digital multimedia (audio, image and video) easily and efficiently to distant places. However, this convenience allows unauthorized copying and distribution of multimedia data. Copyright protection of digital data has become an important issue. Digital watermarking [1] technology has received great deal of attention to solve this problem. Digital watermarking is a process of embedding watermark data into the audio signal. Main requirements of digital audio watermarking are imperceptibility, robustness and data capacity [2].Some preliminary results have appeared recently in [3], [4] showing the interest of EMD for audio watermarking. In [3], the EMD is combined with Pulse Code Modulation (PCM) and the watermark is inserted in the final residual of the sub-bands in the transform domain. Another strategy is presented in [4] where the EMD is associated with Hilbert transform and the watermark is embedded into the IMF containing highest energy. However, why the IMF carrying the highest amount of energy is the best candidate mode to hide the watermark has not been addressed. Further, in practice an IMF with highest energy can be a high frequency mode and thus it is not robust to attacks. In this paper, to simultaneously have better resistance against attacks and imperceptibility comparing to [3], [4], we embed the watermark in the extrema of the last IMF.

For embedding and extracting the watermark, we use a watermarking technique in the category of Quantization Index Modulation (QIM) due to its good robustness and blind nature [5]. Parameters of QIM are chosen to guarantee that the embedded watermark in the last IMF is inaudible. The watermark is associated with a synchronization code to facilitate its location. An advantage to use the time domain approach, based on EMD, is the low cost in searching synchronization codes.

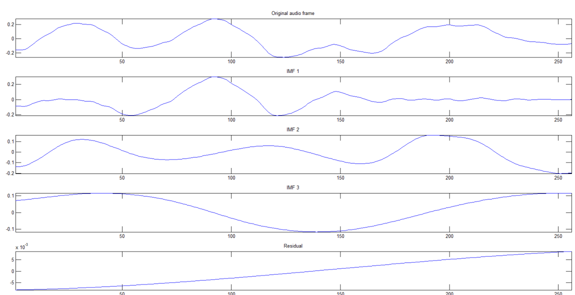


그림 1. 오디오 원형과 내재모드함수
Fig. 1. Original audio frame and its IMFs

* Corresponding Author.

Sync Code	Watermark bits	Sync Code	Watermark bits	…..

그림 2. 워터마크 자료 구조

Fig. 2. Watermark data structure

II. Proposed Method Of Audio Watermarking

The idea of the proposed watermarking method is to hide into the original audio signal a watermark together with a Synchronization Code (SC) in the time domain. The input signal is first segmented into frames and EMD is conducted on every frame to extract the associated IMFs (Fig. 1). Then a binarydata sequence consisted of SCs and informative watermark bits (Fig. 2) is embedded in the extrema of a set of consecutive last-IMFs. A bit (0 or 1) is inserted per extrema.

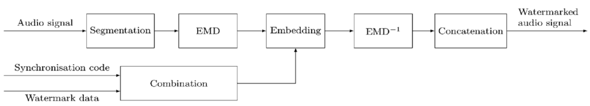


그림 3. 오디오 원형과 내재모드함수

Fig. 3. Original audio frame and its IMFs

Since the number of IMFs and then their number of extrema depend on the amount of data of each frame, the number of bits to be embedded varies from last-IMF of one frame to the following. Watermark and SCs are not all embedded in extrema of last IMF of only one frame. In general the number of extrema per last-IMF (one frame) is very small compared to length of the binary sequence to be embedded. This also depends on the length of the frame. If we design by N1 the numbers of bits of SC respectively, and the original watermark data will be splitted into P watermark frames, each frame has N2 bits , then the length of binary sequence to be embedded is equal to *P*(N1+N2)*. Thus, these *P*(N1+N2)* bits are spread out on several last-IMFs (extrema) of the consecutive frames. Finally, inverse transformation (EMD-1) is applied to the modified extrema to recover the watermarked audio signal by superposition of the IMFs of each frame followed by the concatenation of the frames (Fig. 3). For data extraction, the watermarked audio signal is split into frames and EMD applied to each frame (Fig. 4). Binary data sequences are extracted from each last IMF by searching for SCs (Fig. 5). We show in Fig. 6 the last IMF before and after watermarking. This figure shows that there is little difference in terms of amplitudes between the two modes. The proposed watermarking scheme is blind, that is, the host signal is not required for watermark extraction.

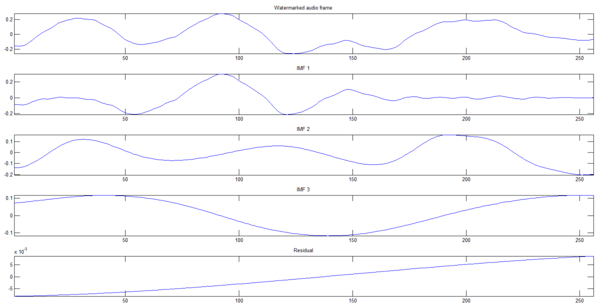


그림 4. 워터마킹된 오디오와 내재모드함수

Fig. 4. Watermarked audio frame and its IMFs

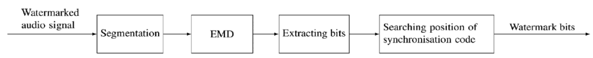


그림 5. 워터마크 추출자

Fig. 5. Watermark Extractor

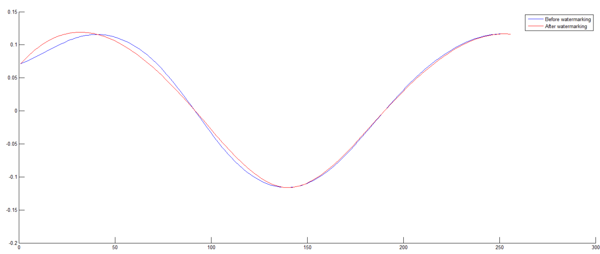


그림 6. 마지막 내재모드함수 전과 워터마킹 후

Fig. 6. Last IMF before and after watermarking

2.1 Watermark Embedding

Before embedding, SCs are combined with watermark bits to form a binary sequence denoted by $\{m_i\}$ i-th bit of watermark (Fig. 2). Basics of our watermark embedding are shown in Fig. 3 and detailed as follows:

Step 1) Split original audio signal into frames.

Step 2) Decompose each frame into IMFs by using EMD.

Step 3) Embed times the binary sequence into extrema of the last IMF by QIM [13]:

$$e_i^* = \begin{cases} \lfloor e_i/S \rfloor \cdot S + sgn(3S/4) & \text{if } m_i = 1 \\ \lfloor e_i/S \rfloor \cdot S + sgn(S/4) & \text{if } m_i = 0 \end{cases}$$

where e_i and e_i^* are the extrema of the last IMF of the host audio signal and the watermarked signal respectively. sgn function is equal to "+" if is a maxima, and "-" if it is a minima. $\lfloor\ \rfloor$ denotes the floor function, and S denotes the embedding strength chosen to maintain the inaudibility constraint.

Step 4) Reconstruct the frame (EMD-1) using modified and concatenate the watermarked frames to retrieve the watermarked signal.

2.2 Watermark Extraction

For watermark extraction, host signal is split into frames and EMD is performed on each one as in embedding. We extract binary data. We then search for SCs in the extracted data. This procedure is repeated by shifting the selected segment (window) one sample at time until a SC is found. With the position of SC determined, we can then extract the hidden information bits, which follows the SC. Let y = $\{m_i^*\}$ denote the binary data to be extracted and denote the original SC. To locate the embedded watermark we search the SCs in the sequence $\{m_i^*\}$bit by bit. The extraction is performed without using the original audio signal. Basic steps involved in the watermarking extraction, shown in Fig. 5, are given as follows:

Step 1) Split the watermarked signal into frames.
Step 2) Decompose each frame into IMFs using EMD.
Step 3) Extract the extrema $\{e_i^*\}$ of the last IMF.
Step 4) Extract from using the following rule [13]:

$$m_i^* = \begin{cases} 1 \ \text{ if } \ e_i^* - \lfloor e_i/S \rfloor \geq sgn(S/2) \\ 1 \ \text{ if } \ e_i^* - \lfloor e_i/S \rfloor < sgn(S/2) \end{cases}$$

Step 5) Set the start index of the extracted data, y, to $I = 1$ and select $L = N_1$ samples (sliding window size).
Step 6) Evaluate the similarity between the extracted segment $V = y(I : L)$ and U bit by bit. If the similarity value is $\geq \tau$, then is taken as the SC and go to Step 8. Otherwise proceed to the Step 7.
Step 7) Increase I by 1 and slide the window to the next L = N_1 samples and repeat Step 6.
Step 8) Take the sequence $y(I + N_1 : I+N_1+N_2-1)$ as the watermark data frame.
Step 9) $I \leftarrow I + N_1+N_2$, of the new I value is equal to sequence length of bits, go to Step 10 else slide the window to the next $L = N_1$ samples and repeat Step 6.
Step 10) Extract the P watermark frames and make comparison bit by bit between these marks, for correction, and finally extract the desired watermark.

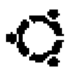
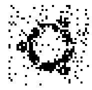


그림 7. 바이너리 워터마크
Fig. 7. Binary Watermark

III. Experimental Result

We tested our algorithm on 16 bit signed stereo audio signals, including classic, pop, speech, sampled at 44.1 kHz with the length of about 15 seconds in the WAVE format. The embedded watermark, W, is a binary logo image of size M × N = 40 × 40 = 1600 bits (Fig.7). We convert this 2D binary image into 1D sequence in order to embed it into the audio signal. The SC used is the 16 bit Barker sequence 1111100110101110. Each audio signal is divided into frames of size 64 samples and the threshold τ is set to 7. The S value is fixed to 0.000437. These parameters have been chosen to have a good compromise between imperceptibility of the watermarked signal and robustness.

표 1 추출된 워터마크의 BER과 NC
Table 1 BER and NC of Extracted Watermerk

Audio Signal	BER %	NC	Extracted watermarked
Classic	18.375	0.88457	
Pop	11.5625	0.92881	
Speech	16	0.9002	

IV. Conclusions

In this paper a new adaptive watermarking scheme based on the EMD is proposed. Watermark is embedded in very low frequency mode (last IMF), thus achieving good performance against various attacks. Watermark is associated with synchronization codes and thus the synchronized watermark has the ability to resist shifting and cropping. Data bits of the synchronized watermark are embedded in the extrema of the last IMF of the audio signal based on QIM. In all audio test signals, the watermark introduced no audible distortion, and the watermarked audio signal are indistinguishable from original ones. Furthermore, the experimental result shows that out proposed scheme achieves NC values are all above 0.88457 and BER values are below 18.375%.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. NRF-2013R1A2A2A05004566).

References

[1] I. Cox, M. Miller, J. Bloom, J. Fridrich, T. Kalke, “Digital Watermarking and Steganography”, second ed., Morgan Kaufmann, Burlington, MA, 2007.

[2] K. Khaldi, A. O. Boudraa, "Audio Watermarking via EMD", IEEE Trans. Audio, Speech, and Language Processing, 21 (3) (2013) 675 - 680.

[3] L. Wang, S. Emmanuel, and M. S. Kankanhalli, "EMD and Psychoacoustic Model Based Watermarking for Audio," in Proc. IEEE ICME, 2010, pp. 1427-1432.

[4] A. N. K. Zaman, K. M. I. Khalilullah, Md. W. Islam, and Md. K. I. Molla, "A Robust Digital Audio Watermarking Algorithm Using Empirical Mode Decomposition," in Proc. IEEE CCECE, 2010, pp. 1-4.

[5] B. Chen and G. W. Wornell, "Quantization Index Modulation Methods for Digital Watermarking and Information Embedding of Multimedia," J. VLSI Signal Process. Syst., vol. 27, pp. 7-33, 2001.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shin, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis," Proc. R. Soc. Lond. A, Math. Phys. Sci., vol. 454, no. 1971, pp. 903-995, Mar. 1998.

[7] V. Bhat, K. I. Sengupta, and A. Das, "An Adaptive Audio Watermarking Based on the Singular Value Decomposition in the Wavelet Domain," Digital Signal Process., vol. 2010, no. 20, pp. 1547-1558, 2010.