Source Separation of Steganography Mixed Audio Signal

Yawai Tint, Khin Than Mya

University of Computer Studies,Yangon

yawai.ywt@gmail.com,khinthanmya@gmail.com

Abstract

The research on blind source separation is focus in the community of signal processing and has been developed in recent years. This paper proposes enhance audio steganalysis technique, which adopts Independent Component Analysis (ICA) for steganography detection and extraction process. Steganography can be successfully detected during the Principle Component Analysis (PCA) whitening stage. A nonlinear ICA algorithm, which is able to efficiently extract various temporally correlated sources from their observed linear mixtures, is used for blind steganography extraction.

Keywords: Steganalysis,Independent component analysis (ICA), blind signal separation

1. Introduction

Steganography is the art and science of hiding data in digital images, audios and videos, etc. To the contrary, steganalysis is the art and science of detecting the information-hiding behaviors in these covers. In the past few years, many researchers presented several steganalysis methods to detect the information-hiding behaviors in multiple steganography systems. In audio steganalysis , Ru et al. presented a detection method by measuring the features between the signal and a self-generated reference signal via linear predictive coding[5]; Avcibas designed the content independent distortion measures as features for classifier design [7]. Ozer et al. constructed the detector based on the characteristics of the denoised residuals of the audio file [6]. To detect the information hiding in audios, Johnson et al. set up a statistical model by building a linear basis that captures certain statistical properties of audio signals[2] In this article, propose an Independent Component Analysis (ICA) is the process of extracting unknown independent source signals from sensor that are unknown combinations of the source signals. It is a novel statistical technique that aims at finding linear projection of the data that maximize their mutual independence. The work has been one of the most exciting topics in the fields of neural computation, advanced statistics, signal processing, and communication engineering which find independent components from observing multidimensional data based on higher order statistics.

In order to solve the problem of blind signal separation, many techniques have been proposed. Among them, the independent component analysis methods are based on the assumption of mutual independence of the sources. Blind steganography decoding is used by robust batch ICA algorithm. The motivations and advantages of using ICA technique in detection and extraction could be the following:

- Any of the three source signal, that is, the steganalysis, the key and the original signal, can be recovered by the separation process, which is just like a reverse process of steganography.
- The steganography can be embedded either instantaneously or convolutively, that is, the observed stegano-signal can be either instantaneous or convolutive mixture.

It does not require the original data for decoding, except the secret key. The method is not only a steganography tool but also a synchronization tool between the transmitter and the receiver. Any embedding process can be used, in which the original audio and steganography are statistically independent and linearly mixed.

2. Feature Extraction

In this paper, three types of features are computed from each frame, time domain based features, frequency domain based features and Mel-cepstral domain based features.

2.1. Mel-Cepstral Domain based Features

Mel-frequency cepstral coefficients are non-parametric representations of audio signal, which models the human auditory perception system. The term "mel" is a unit of measurement of the perceived frequency or pitch of a tone. The mapping between the frequency scale (Hz) and the perceived frequency scale (mels) is approximately linear below 1 kHz and logarithmic at higher frequencies. The suggested formula that approximates this relationship is as follows

$$F_{mel} = 2595.\log_{10}\left(1 + \frac{F_{Hz}}{700}\right) \quad (1)$$

where F_{mel} is the perceived frequency in mels and F_{Hz} is the frequency in Hz.

The critical-band filters in the frequency domain (Hz) are illustrated in Figure (1). In the mel-frequency domain, the bandwidth and the spacing of these critical-band filters are invariable values, 300 mels and 150 mels, respectively.

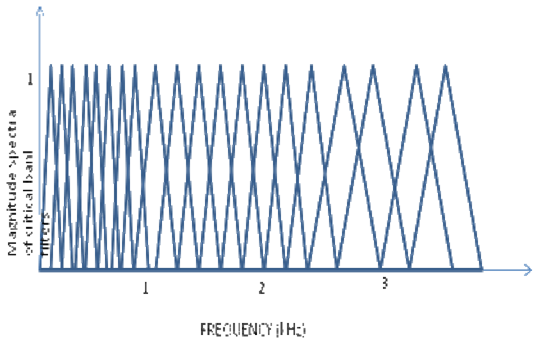


Figure (1) Critical-band filters in the frequency domain.

The derivation of MFCCs is based on the powers of the theses critical-band filters. Let X(m) denote the power spectrum of an audio stream, S[k] denote the power in k-th critical band and M represent the number of the critical bands in mel scale, ranging usually from 20 to 24. Then

$$S[k] = \sum_{j=0}^{f/2-1} W_k(j).X(j),$$

$$k = 1, \dots \dots, M, \quad (2)$$

where Wk is the critical-band filter.

Let L denote the desired order of the MFCC. Then we can find the MFCCs from logarithm and cosine transforms as follows

$$C[n] = \sum_{k=1}^{M} \log(S[k]) \cos\left[(k-0.5)\frac{n\pi}{M}\right], n = 1, \ldots, L. \quad (3)$$

2.2. Time Domain based Features

The well known short-term energy and zero-crossing rate (ZCR) are two popular choices in this category. ZCR measures the number of time domain zero crossings (divided by the frame's length).

$If\ \{x(0), x(1), \ldots, x(N-1)\}$ is the short term frame, then two feature are given by

(i) Short-term energy

$$E = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n) \quad (4)$$

(ii) Short-term zero crossing rate (ZCR)

$$ZCR = \frac{1}{N} \sum_{n=1}^{N} \frac{|sgn[x(n)] - sgn[x(n-1)]|}{2} \quad (5)$$

2.3. Frequency Domain based Features

The spectral features including spectral centroid, spectral rolloff, spectral flux and spectral entropy are derived as follows

(i) Spectral centroid : A measure of the spectral shapes with high values corresponding to brighter sounds. The spectral controid, C_t , of the t*th* frame is defined as the center of "gravity" of its spectrum, i.e.

$$C_t = \frac{\sum_{k=0}^{N-1}(k+1)X_t(k)}{\sum_{k=0}^{N-1} X_t(k)} \quad (6)$$

(ii) Spectral roll-off: The frequency below which certain percentage (usually 85% or 90%) of the magnitude distribution of the spectrum is concentrated. If the m*th* DFT coefficient corresponds to the spectral rolloff of the t*th* frame, then the following equation holds.

$$\sum_{k=0}^{M} X_t(k) = C \sum_{k=0}^{N-1} X_t(k) \quad (7)$$

Where C is the adopted percentage

(iii) Spectral flux: A measure of the local spectral change between successive frames. It is defined as the squared difference between this normalized magnitudes of the spectra of two successive frames:

$$FL(t, t-1) = \sum_{k=0}^{N-1} (N_t(k) - N_{t-1}(k))^2 \quad (8)$$

Where

$$N_t(k) = \frac{X_t(k)}{\sum_{l=0}^{N-1} X_t(l)}, N_t(k) \quad (9)$$

is the k*th* normalized DFT coefficient at the t*th* frame

(iv) Spectral entropy: Entropy is a measure of the uncertainty or disorder in a given distribution. In order to compute spectral entropy, the spectrum of the short-term frame is first divided into L sub-bands (bins). The energy E*i* of the *ith* sub-bands i=0, ……………, L-1, is then normalized by the total spectral energy, yielding $n_i = \frac{E_i}{\sum_{i=0}^{L-1} E_i}, i = 0, \dots \dots, L-1$ The entropy of the normalized spectral energy is then computed by the equation.

$$H = -\sum_{i=0}^{L-1} n_i . \log_2(n_i) \qquad (10)$$

3. Dimensional Reduction Based on ICA

Each frame is represented as a vector x_{m*1}, a audio file contains *n* frame is represented as a matrix $X = (x_1, x_2, \dots, x_n)^T$. That refers the audio features to a mixing of various independent source features. In the formulation of ICA, the source matrix is assumed to be an unknown mixture of unknown sources.

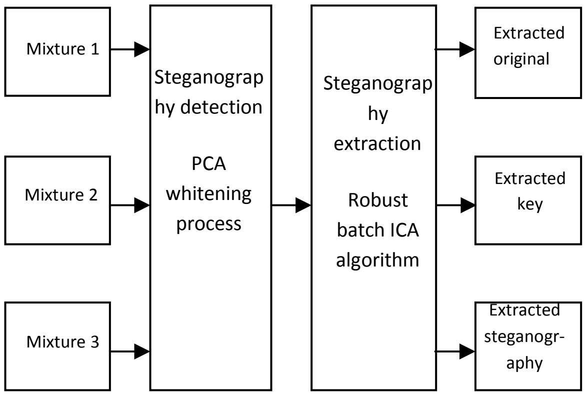
$$x = As = \sum_{i=1}^{n} a_i s_i \qquad (11)$$

Here, components s_i and s_j are unknown sources that are independent from each other for $i \neq j$.

The purpose is to learn the de-mixing matrix W so that the components of the estimated source vectors S are independent from each other in the form:

$$S = WZ = WVX \qquad (12)$$

After independent component analysis that get the independent components $S = (s_1, s_2, \dots, s_m)$ from audio frame.



(2) Proposed steganography detection and extraction scheme

To assure the identifiability of ICA model, it is required that the number of observed linear mixture inputs is at least equal to or larger than the number of independent sources. For the above-proposed detection and extraction scheme, at least three linear mixtures of the three independent sources are needed. Using the key K and with the help of original signal I, two more mixed signal are generated by adding them into the stegano-signal X

$$X_1 = X \quad X_2 = X + cK \quad X_3 = X + dI \qquad (13)$$

where c and d are arbitrary real numbers.

Proposed steganography detection and extraction scheme shown in Figure (2). Firstly, three mixtures input are detected by PCA whitening process and then features are extracted from robust algorithm.

(1) Whitening preprocessing: The most basic and necessary preprocessing is to center x, that is $x = x - E(x)$, so as to make x a zero-mean variable, this processing is made solely to simplify the ICA algorithm. Another useful preprocessing is to whiten the observation signals which make the components are uncorrelated and their variances equal unity to provide uncorrelated components

$$\tilde{x}(t) = Ux(t) \qquad (14)$$

Where the whitening matrix U is usually computed after singular or eigen-value decomposition of the covariance matrix of x (t)

$$U = VD^{\frac{-1}{2}}V^T \qquad (15)$$

Where V is the orthogonal matrix composed eigenvector of covariance of x and $D = diag(d1, \ldots, dn)$ is the diagonal matrix of its eigenvalues. Then the aim of ICA is to estimate the separation matrix W, and the separation signal

$$y(t) = W\tilde{x}(t) \qquad (16)$$

is the estimation of source signal.

(2) Algorithm of ICA: Kurtosis has widely used as a measure of non-Gaussinity in ICA and related fields, which can be estimated simply by using the fourth moment of the sample data. Kurtosis is defined as follows:

$$kurt(s_i) = E[s_i^4] - 3(E[s_i^2])^2 \quad (17)$$

We erect adjective function:

$$kutr(w^T\hat{x}_i) = E[(w^T\tilde{x}_i)^4] - 3[E\{(w^T\tilde{x}_i)^2\}]^2 \qquad (18)$$

Since the observation signal has been pre-whitening. Thus eq (17) can be simplified as:

$$\Delta w\alpha\ E[\tilde{x}_i(w_i(k)^T\tilde{x}_i)^3] - 3\,\|w_i(k)\|^2 w_i(k) \ (19)$$

Using the fixed-point algorithm, the iteration of fixed-point algorithm can be expressed:

$$w_i(k) = E[\tilde{x}_i(w_i(k-1)^T\tilde{x}_i)^3] - 3w_i(k-1) \qquad (20)$$

Thus, obtain the robust batch ICA algorithm as follows:

(1) Center the data to make its mean zero;

(2) Whiten the data to get $\tilde{x}(t)$;

(3) Make $i = 1$;

(4) Choose an initial orthogonal matrix for W and make $k = 1$;

(5) Make $w_i(k) = E[\tilde{x}_i(w_i(k-1)^T x_i)^3] - 3w_i(k-1)$

(6) Make $w_i(k) = \frac{w_i(k)}{\|w_i(k)\|}$

(7) If not converged, make $k = k + 1$ and go back to step (5)

(8) Make $i = i + 1$

(9) When i < number of original signals, go back to step (4)

Until $|w_i(k)^T w_i(k-1)|$ is equal or close to 1, the iteration finished.

4. Proposed Framework

Signals were used for simulations; they are sampled at 44100Hz.

The original audio signals are shown in Figure (3)

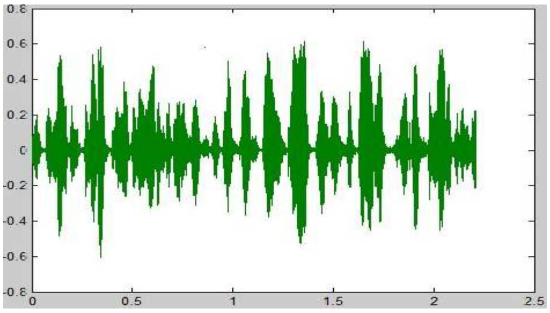


Figure (3) The original signals

Steganography signal depicted by Figure (4)

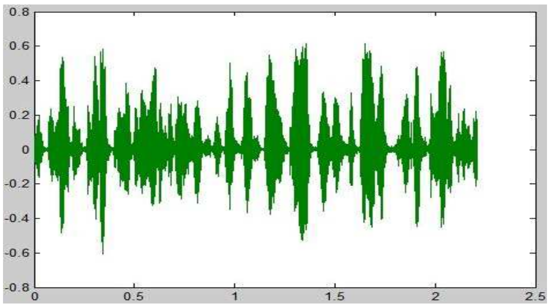


Figure (4) The steganography signals

This paper selected ICA as BSS algorithm and robust batch as an implementation of ICA. Using robust ICA for steganalysis, that has simple and fewer numbers of features. ICA gives s a representation, or transformation, of multidimensional data that seems to be well suited for subsequent information processing. This algorithm estimates all the independent components at the same time, and estimating only a subset of them. ICA is a very general purpose statistical in which observed random data are linearly transformed into components that are maximally independent from each other, and simultaneously have "interesting" distributions. The robust algorithm has most of the advantages of neural algorithms: It is parallel, distributed, computationally simple, and requires little memory space.

5. Conclusion

In this paper, enhance audio steganalysis technique based on Independent Component Analysis (ICA) has been proposed. This study also presents frequency domain based features, time domain based features and Mel-cepstral domain based features. The audio steganography is readily detected by Principle Component Analysis (PCA) whitening process. The robust batch ICA algorithm is an effective blind source separation approach and that can extract steganography exactly.

References

[1] T. Pevny and J. Fridrich Merging Markov and DCT features for Multi-class JPEG Steganalysis. Contents IX, vol.6505, 2007.

[2] A .Hyvarinen. "A family of fixed point algorithms for independent component analysis, Neural Computation, 1997, 9 (7) 1483~1492

[3] T. W. Lee, Independent Component Analysis – Theory and Applications Kluwer Academic Publishers, 1998.

[4] D. Yu, F. Satter " Watermark Detection and Extraction Using Independent Component Analysis Metho" EURASIP Journal on Applied Signal Processing 2002:1,92-104

[5] Q. Liu, A. Sung, Feature Mining the Nuero-fuzzy Inference System for Steganalysis of LSB Matching Steganography in Grayscale Images. Proc. Of 20th International Joint Conference on Artificial Intelligence, pp.2808-2813, 2007.

[6] T. Le-Tien, D. Vo-Ngoc, "Independent Component Analysis applied to watermark extraction and its

implemented model on FPGAs" IEEE Signal Processing Letters, 2010.

[7] X. Ru, Y.Zhang and F.Wu . Audio Steganalysis Based on " Negative Resonance Phenomenon" Caused by Steganographic Tools. Journal of Zhejiang University SCIENCE A,7(4):577-583,2006.