Markov bidirectional transfer matrix for detecting LSB speech steganography with low embedding rates

Wanxia Yang, Shanyu Tang*, Miaoqi Li, Beibei Zhou, Yijing Jiang

* Corresponding author: Prof. Shanyu Tang (shanyu.tang@gmail.com, shanyu.tang@uwl.ac.uk)

Abstract: Steganalysis with low embedding rates is still a challenge in the field of information hiding. Speech signals are typically processed by wavelet packet decomposition, which is capable of depicting the details of signals with high accuracy. A steganography detection algorithm based on the Markov bidirectional transition matrix (MBTM) of the wavelet packet coefficient (WPC) of the second-order derivative-based speech signal is proposed. On basis of the MBTM feature, which can better express the correlation of WPC, a Support Vector Machine (SVM) classifier is trained by a large number of Least Significant Bit (LSB) hidden data with embedding rates of 1%, 3%, 5%, 8%,10%, 30%, 50%, and 80%. LSB matching steganalysis of speech signals with low embedding rates is achieved. The experimental results show that the proposed method has obvious superiorities in steganalysis with low embedding rates compared with the classic method using histogram moment features in the frequency domain (HMIFD) of the second-order derivative-based WPC and the second-order derivative-based Mel-frequency cepstral coefficients (MFCC). Especially when the embedding rate is only 3%, the accuracy rate improves by 17.8%, reaching 68.5%, in comparison with the method using HMIFD features of the second derivative WPC. The detection accuracy improves as the embedding rate increases.

Keywords: Speech; LSB; Steganalysis; WPC; Markov transition matrix; MFCC

I. Introduction

Steganography is one of the important technologies in information security. It takes advantage of redundancy in digital carriers to embed confidential information without reducing the carrier quality. It not only hides the content of private information but also its existence, thus ensuring secure transmission [1]. Steganography, however, is a "double-edged sword." People with evil intentions or terrorists could use it to organize terrorist attacks or criminal activities. Steganography technologies have not only been the concern of the international academic community but are also highly valued by military and other security sectors in many countries. As a countermeasure technique of steganography, steganalysis is a hot issue and one of the most difficult problems is its particularly low embedding rate.

In recent years, with the rapid development of Voice over Internet Protocol (VoIP) communication technology, annual VoIP traffic has reached more than one hundred billion minutes over the Internet [2-3]. Theoretically, as the duration of a VoIP session is sustainable for any length of time, the secret information of unlimited capacity can be transmitted in VoIP even with a very low embedding rate by only prolonging the session time. For example, even if the embedding rate is only 1%, 80 bit confidential information can still be embedded in G.711 audio data sampled at 8000 times per second, and this hiding capacity is much higher than that of the same length of voice data in the low rate speech coding standard. In fact, this has brought great challenges to speech steganalysis because the current steganalysis mainly utilizes statistical methods to analyze the characteristics changes in the carrier before and after steganography. The alteration of carrier structure caused by steganography with low embedding rate is very small and hard to detect. Moreover, the predominant characteristic of VoIP speech stream is real-time, so it is very difficult to capture small changes in carrier features caused by steganography and conduct correct detection in a short period of time. Studies on steganography detection algorithms at low embedding rates are relatively few; however, it is of vital importance this topic is

thoroughly researched.

Based on the Markov bidirectional transition matrix (MBTM) of wavelet packet coefficients (WPC) of the second-order derivative of speech signals, the detection algorithm of by Least Significant Bit (LSB) matching steganography at low embedding rates as a VoIP carrier is proposed. Experimental results show that the algorithm has achieved good results. This algorithm provides a very good method to solve the difficult problem of steganography detection with a low embedding rate.

This paper is organized as follows. In Section 2, the related work on LSB algorithm of steganography and steganalysis is introduced. In Section 3, the methods for extracting the features of MBTM and the histogram moments in the frequency domain (HMIFD) based on the WPC of the second-order derivative of the speech signal are proposed. In Section 4, the detection algorithms designed in Section 3 are simulated and the experimental results are analyzed. Section 5 concludes the work.

II. Related Work

Changes to the carrier introduced by LSB steganography is normally very small, so the LSB embedding capacity can be large and has been widely used. At the same time, the detection algorithms for LSB steganography were soon presented. For example, the study in [4] theoretically proved that the LSB matching steganography could make the multi-order difference histogram of the image smoother, so the co-occurrence matrix of the adjacent pixels was calculated as features to realize the reliable detection of LSB matching steganography. Zhang et al. [5] proposed an algorithm based on Markov properties of adjacent pixel and differential histogram for testing. Based on a regional random index, three features of the histogram information entropy, special value and origin moment were adapted to detect LSB matching steganography in different embedding rates [6-7]. These methods are all based on single or low dimensional features, so the detection

accuracy needs enhancement, which was not ideal, especially for small embedding rates.

Later, the detection method of multi-feature fusion to form a multi-dimensional training set was suggested for better accuracy. By integrating the transition probability in the wavelet domain with the Markov transition probability in discrete cosine transformation (DCT) expansion domain, a 93.42% image stitching recognition rate was achieved [8-10]. The detection algorithm of the rich model [11-12] was then proposed, since the feature dimension is too large, and the ensemble classifier was chosen to obtain rapid and accurate detection. The inherent characteristics of image content were not given enough attention in the above-stated methods. Research showed that the changes in a carrier after embedding a secret message were closely related to image content. In [13-15], an image was divided into blocks of fixed size according to its steganalysis features; each block was classified, and the final classification results were acquired by voting and better with other method. Recently, various reversible information-hiding methods have been presented, such as multiple histogram modification [16], two-dimensional difference-histogram modification [17], histogram shifting mechanism [18] and vector quantization (VQ) index table with lossless coding and an adaptive switching mechanism [19]. Therefore, further research on the method of steganalysis needs to be conducted.

For steganography and steganalysis in digital carriers, many studies used images as covers; in contrast, few detecting algorithms for steganography in VoIP speech stream were reported. Large structural feature differences exist between image carriers and speech carriers, so the detection algorithm for image steganography cannot be directly applied to speech steganalysis. By exploiting the idea of image steganography detection and analyzing the speech features deeply, researchers have made some progress in speech steganalysis. For example, Huang et al. [20] introduced a sliding window mechanism, which mainly studied the real-time characteristics of a VoIP stream, and the improved RS algorithm could be adapted to real-time detection of VoIP speech. Based on correlation features, the original algorithm was improved in [21]; therefore, real-time detection for LSB

matching steganography in VoIP flow was successful, but previous research can be improved upon. Later, some scholars studied the detection algorithm in compressed speech [22-23]. In [23], the detection method of quantization index modulation (QIM) steganography in a low speed rate encoder G.723.1 was proposed. This encoder extracted the features from the compressed domain and achieved a good detection effect. Histogram flatness, characteristic functions and variance were used to successfully detect the steganography of a fixed code index [24]. By adding a histogram local extremum difference and a 0, 1 distribution probability difference, better detection accuracy was obtained in [25]. However, the aforementioned methods have a poor detection effect on steganography at a low embedding rate.

It can be seen that none of the above methods can describe the details of the signal clearly and cannot solve the steganalysis problem with low embedding rates. However, both theoretical and experimental results have shown that when the derivative order is higher, the signal characteristics at high frequencies are more obvious. Steganography is often regarded as an artificially added high-frequency noise in speech signals. Research results of [26] indicated that the speech changes at high frequencies caused by steganography could not be ignored, and the authors received good detection results by extracting features of MFCC and Markov from the second-order derivative of speech signals.

In view of these issues, by taking advantage of the characteristics that a wavelet packet can focus on any tiny signal details, wavelet packets decompose the second-order derivative of speech signals, and the MBTM feature is extracted from WPC for steganalysis at low embedding rates in this paper. The classic HMIFD of WPC and excellent MFCC based on second-order derivatives of speech signals as features are chosen for comparison. The experimental results demonstrated that the proposed detection algorithm has obvious advantages at different embedding rates.

Since MFCC is most commonly used for speech steganalysis, the extraction method is no longer detailed in

this paper.

The steganalysis method proposed in this paper is shown in Fig. 1.

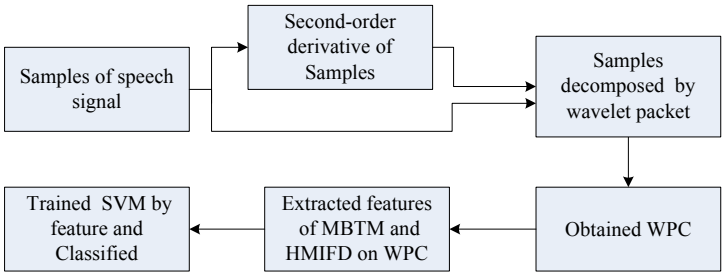


Fig. 1. Steganalysis method

The methodology shown in Fig. 1 is as follows:

(1) speech signal samples are obtained;

(2) the second-order derivative of samples obtained in step (1) is used;

(3) samples in steps (1) and (2) are decomposed by wavelet packets, and two types of WPC are obtained;

(4) the MBTM and HMIFD features of each WPC are extracted; and

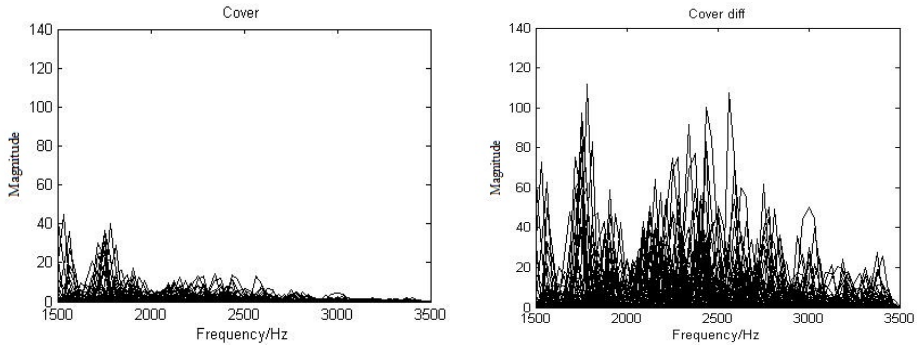
(5) The SVM is trained by features in (4), and the experimental samples are classified.

Ⅲ. Extraction method of the MBTM and HMIFD of WPC

1. Feature analysis of the second-order derivative of the speech signal

Derivatives can clearly depict small signals at high frequency, so the first or second derivative is often used in the singular value and edge detection of an image. In view of this, this study compares a spectrum in the high-frequency range of the second-order derivation and non-second-order derivation of speech, as shown in Fig.

2. It can be seen that the speech signal details in the high-frequency spectrum are clearly described for its second-order derivative. Although the high-frequency components of speech signals are small, they contain a significant amount of important information about speech features. Therefore, the second-order derivative of the speech signal decomposed by wavelet packets in this study is used to obtain more signal details to enhance the detection accuracy of steganography at low embedding rates.



(a) (b)

Fig. 2.(a)(b) High-frequency spectrum of speech signals before and after the second derivative

2. Feature extraction of WPC

A speech signal is a kind of typical non-stationary signal. Although Fourier transform is the most classic and commonly used method to study speech signals, it is only suitable for the analysis and processing of linear stationary signals. Wavelet packet decomposition has good time-frequency localizing characteristics and can focus on any tiny detail of a signal without redundancy or careless omissions. This type of decomposition also includes the medium and high-frequency information of the signal, which indicates that the wavelet only decomposes signals in low-frequency ranges but not in high-frequency ranges.

Wavelet packet transform is an improved analysis method for Multi-Resolution Analysis (MRA), and its objective is to construct an orthogonal wavelet basis to highly approach $L^2(\mathrm{R})$ space in frequency. According to

the concept of multi-resolution analysis, $\phi(t)$ and $\psi(t)$ are the orthogonal basis functions in scale space V_0 and wavelet space W_0, respectively, and wavelet packet decomposition uses the functions of scale and wavelet to further expand at subspace V_1. Its common recurrence relation is as follows [27]:

$$\varphi_{2n}(t) = \sqrt{2}\sum_{k\in z} h_k \varphi_n(2t-k)$$

$$\varphi_{2n+1}(t) = \sqrt{2}\sum_{k\in z} g_k \varphi_n(2t-k) \qquad (1)$$

where h_k and g_k are the filter coefficients in the following multi-resolution analysis: $g(\kappa) = (-1)^{\kappa} h(1-\kappa)$ When n = 0, $\varphi_0(t) = \phi(t)$ is the scaling function and $\varphi_1(t) = \psi(t)$ is the wavelet function. The set $\{\varphi_n(t)\}_{n\in z}$ of functions recursively defined by (1) is the orthogonal wavelet packet determined by $\varphi_0(t) = \phi(t)$. The recursive formula of WPC is deduced by the wavelet packet definition in (1). The calculation formula is as follows:

$$d_k^{j+1,2n} = \sum_{\mathrm{L}} h_{(2L-k)} d_L^{j,n} \qquad (2)$$

$$d_k^{j+1,2n+1} = \sum_{\mathrm{L}} g_{(2L-k)} d_L^{j,n}$$

The above wavelet packet decomposition process can be represented by the two-fork tree shown in Fig. 3.

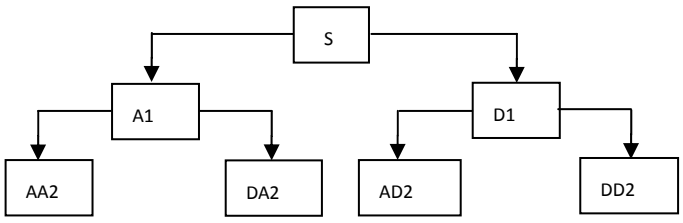


Fig. 3. Tree diagram of two-layer wavelet packet decomposition

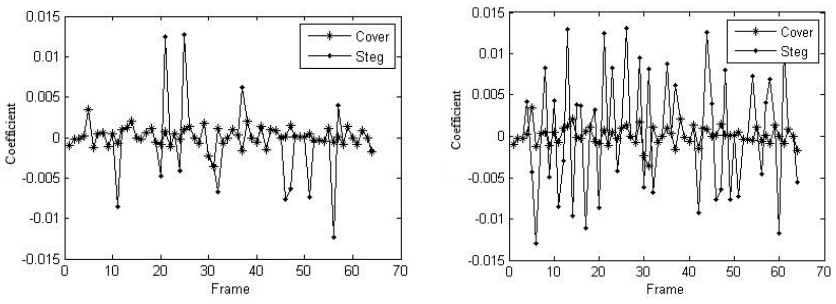
Fig. 3 is a tree graph with two layers of wavelet packet decomposition, and the decomposition relation in the

theory space is as follows:

If $U_j^n = span\left\{2^{\frac{j}{2}} w_n(2^j t-k)_{k\in z}\right\} \quad j,n \in z$

Then, $U_j^0 = V_j \quad U_j^1 = \mathrm{W}_j$, where W_j is an orthogonal complement of V_j in V_{j+1}, that is, $V_j \oplus W_j = V_{j+1} \quad j \in Z$, can be rewritten as $U_{j+1}^0 = U_j^0 \oplus U_j^1$ and further promoted as $U_{j+1}^n = U_j^n \oplus U_j^{n+1}$.

Based on the above theory, considering the tradeoff between filter length and the number of vanishing moments, a Daubechies wavelet, "db6", is applied to decompose the 16-bit mono channel PCM original and stego speech streams by a two-layer wavelet packet. The information entropy of wavelet packet decomposition by anything more than three-layer decomposition is too low to be effective in steganalysis, two-layer wavelet packet decomposition is selected and the approximate contour and detail coefficients of the second layers are extracted, respectively. Based on those coefficients, the MBTM and HMIFD are designed for comparative testing for LSB matching steganography. Experiments prove that the WPC is more sensitive to steganography. As shown in Fig. 4(a) and (b) compares the cover WPC with the stego WPC at the embedding rates of 3% and 30%, respectively. The change in WPC caused by steganography is much larger, so the statistical characteristics of WPC can be fully used as the basis for steganography detection.



(a) (b)

Fig. 4.(a)(b) Contrast diagram of cover WPC and stego WPC with the embedding rates of 3% and 30%

3. Calculation of WPC MBTM

With regard to image steganalysis, many detection algorithms take the Markov transfer probability matrix of the adjacent DCT coefficients as sensitive characteristics and obtain high detection precision. These high-order statistics can better reflect the correlation between the image pixels or coefficients, so it is more sensitive in capturing the damage to this correlation caused by steganography. Furthermore, speech signals have a strong correlation as well, which makes compression coding for linear prediction techniques through the correlations between speech sampling points a possibility. Therefore, the correlation of speech feature parameters can be exactly described by the high-order Markov probability transition matrix and used to detect steganography. The traditional second-order Markov correlation model assumes that the current speech signal is related to the first two speech signals, which are as follows:

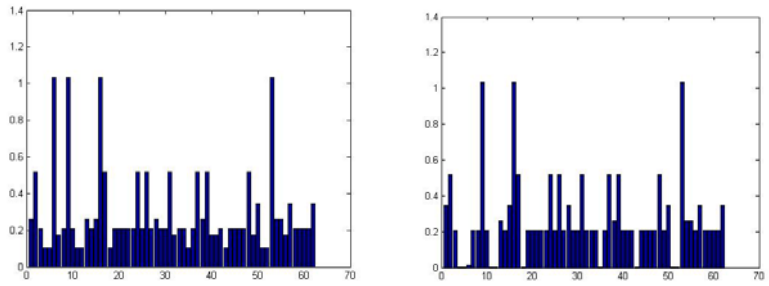
$$\begin{aligned} M_{\lambda_1\lambda_2\lambda_3} &= p(\mathrm{a(i+2)} = \lambda_3 / a(\mathrm{i+1}) = \lambda_2, a(\mathrm{i}) = \lambda_1) \\ &= p(\mathrm{a(i+2)} = \lambda_3, a(\mathrm{i+1}) = \lambda_2, a(\mathrm{i}) = \lambda_1) / p(a(\mathrm{i+1}) = \lambda_2, a(\mathrm{i}) = \lambda_1) \end{aligned} \tag{3}$$

where $p(a(i)) = \lambda_i$ is the parameter probability, and $M_{\lambda_1\lambda_2\lambda_3}$ is the transition probability.

This paper analyzes the correlation of adjacent wavelet coefficients before and after steganography, which is stronger than that of the first two coefficients, as proven in this study and shown in Fig. 5(a)(b). Therefore, the traditional second-order Markov is improved; on the basis of the intermediate-coefficients probability of three adjacent coefficients, the joint probability distribution matrix of the adjacent three coefficients is computed in terms of features, and its calculation expression is shown below:

$$\begin{aligned} M_{\lambda_1\lambda_2\lambda_3} &= p(\mathrm{a(i+2)} = \lambda_3, a(\mathrm{i}) = \lambda_1 / a(\mathrm{i+1}) = \lambda_2) \\ &= p(\mathrm{a(i+2)} = \lambda_3, a(\mathrm{i+1}) = \lambda_2, a(\mathrm{i}) = \lambda_1) / p(a(\mathrm{i+1}) = \lambda_2) \end{aligned} \tag{4}$$

where $p(a(i)) = \lambda_i$ is the probability of the parameter, and $M_{\lambda_1\lambda_2\lambda_3}$ is the transition probability.



(a) (b)

Fig. 5.(a)(b) $M_{\lambda_1\lambda_2\lambda_3}$ **values of the improved method and the conventional method**

The specific steps of this method are as follows:

(1) the second derivative for speech signal is obtained;

(2) the second-order derivative speech signal is segmented into frames according to the short-time stability of speech with 256 points in a frame as not to overflow the Markov transition probability matrix calculation due to excessive dimensions;

(3) a Daubechies wavelet “db6” is applied to decompose each speech signal frame in step(2) by two-layer wavelet packets, and the contour and detail coefficients of the second layers are extracted (as shown in Fig. 3, AA2, DA2, AD2, and DD2);

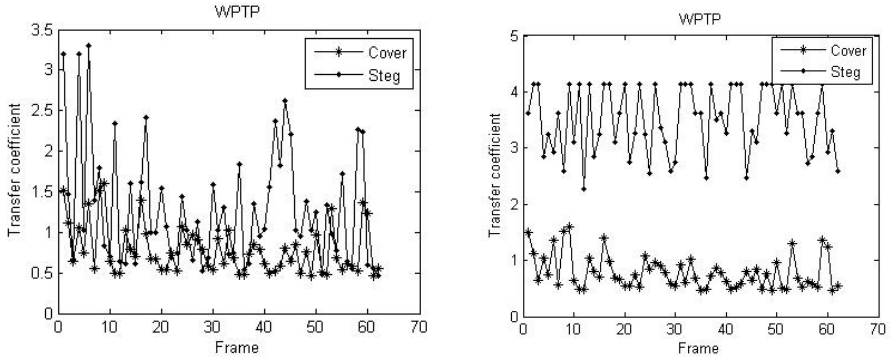
(4) the joint probability matrix of three adjacent coefficients of each wavelet sub-band is calculated, and on the basis of the intermediate-coefficients probability, the transfer probability matrix of the three adjacent coefficients is determined as shown in formula (4);

(5) first, the sum of the transition probability matrix of four sub-and wavelet coefficients is computed, and the sum of the transition probability matrix of each frame is calculated using formula (5); the experiment then compares the feature of the sum of four sub-band wavelet coefficients at embedding rates of 3% and 30% before and after steganography, as shown in Fig. 6 (a) (b); and the change aroused by steganography is obvious:

$$M_{\lambda_1\lambda_2\lambda_3}^{\mathrm{mn}} = \frac{\sum_{k=1}^{n}\sum_{l=1}^{m} p(\mathrm{a(i+2)} = \lambda_3, a(\mathrm{i+1}) = \lambda_2, a(\mathrm{i}) = \lambda_1)}{\sum_{k=1}^{n}\sum_{l=1}^{m} p(a(\mathrm{i+1}) = \lambda_2)} \tag{5}$$

where l is the sub-band number, k is the frame number, $p(a(i)) = \lambda_i$ is the parameter probability, and $M_{\lambda_1\lambda_2\lambda_3}^{mn}$ is the transition probability of N frames of the sub-band coefficients; and

(6) a similar method is used to calculate the corresponding features of non-second derivative speech signals.



(a) (b)

Fig. 6.(a)(b) Contrast diagram of cover and stego MBTMs at embedding rates of 3% and 30%

4. Calculation of WPC HMIFD

In the frequency domain, information steganography is equivalent to a low-pass filter histogram. Even if little information is hidden in a carrier, it still makes the histogram peak variations very obvious, which has been experimentally proven in this study, as shown in Fig. 7. Thus, multi-order statistical histogram moments in the frequency domain are chosen as steganalysis features. The calculation is shown in (6):

$$m_n = \sum_{k=1}^{N} \left| f_k \right|^n p(f_k) \tag{6}$$

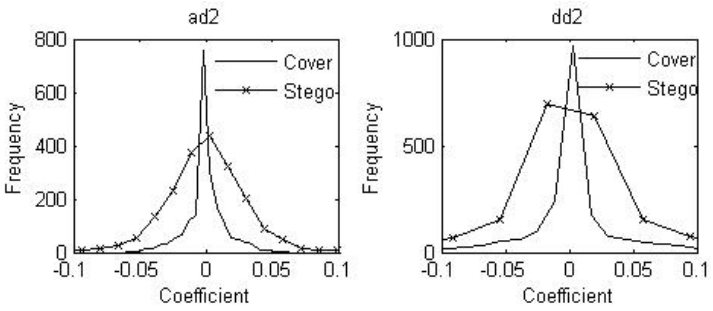


Fig. 7. Histogram of the wavelet coefficient at low and high frequencies before and after steganography

where *n* is the order number of the statistical moments, *N* is the number of horizontal axis variables, f_k is the k-th frequency of DFT, and $p(f_k)$ is the amplitude distribution of the DFT for the histogram.

$$p(f_k) = |H(f_k)| \Big/ \left(\sum_{k=1}^{N} |H(f_k)|\right) \tag{7}$$

where $\left|H(f_k)\right|$ is the magnitude of the *k*-th frequency of DFT for $h(x_k)$. The formula for DFT is as follows:

$$H(f) = \frac{1}{N} \sum_{k=0}^{N-1} h(k) e^{-j2\pi fk/N} \tag{8}$$

where $h(k)$ is the histogram.

The specific calculation processes are designed as follows:

(1) the second derivative is determined for the speech signal;

(2) a Daubechies wavelet, "db6," is applied to decompose the second-order derivative speech signal by two-level wavelet packets, and then 4 (AA2, DA2, AD2, and DD2) sub-bands of the contour and details are obtained, as shown in Fig.3;

(3) four sub-band coefficients (CA_3, CD_3, CD_2, and CD_1) are extracted from Step (2), and first-order, second-order, and third-order absolute HMIFDs of all band coefficients are computed to form 12 dimensional features; and

(4) a similar method is used to calculate the corresponding features of the non-second derivative speech

signal.

Ⅳ. Training method of SVM classifier and experimental results analysis

1. Training method of SVM classifier

To verify the accuracy of the proposed method for the detection of LSB matching steganography with different embedding rates, especially low embedding rates, as well as meet the real-time requirements of VoIP voice flow, 160 mono 16-bit coding PCM audio files sampled at 8000 Hz were collected as covers. Each audio has the duration of only one second. The same amount of stegos were produced by LSB matching steganography with embedding rates of 1%, 3%, 5%, 8%, 10%, 30%, 50%, and 80%, respectively. In each cover and stego file, 100 audio samples are randomly assigned to the training set; the remaining 60 audio samples constitute the testing set. LIBSVM3.1 version was used for classification. The parameter settings were all default settings of LIBSVM. The kernel function of Basis Function Radial (RBF) was selected. The results consisted of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) classifiers. The classification accuracy is calculated as (TP +TN) / (TP + FN + FP + TN) [28]. To display the superiorities of the proposed algorithm based on the WPC MBTM of the second-order derivative speech signal in this study for steganalysis with low embedding rate, the proposed algorithm based on HMIFD of WPC and the MFCC of the second-order derivative speech signal were first compared. Then, the detection results of three corresponding features of the second derivative and non-second derivative speech signal were contrasted. The specific steps were as follows:

(1) 160 cover samples were chosen to generate 160 stego samples by LSB matching steganography;

(2) the method in Section Ⅲ was applied to extract features of the WPC MBTM from 320 experimental

samples (cover and stego) of the second-order derivative and non-second-order derivative, respectively, and form 62 dimensional feature vector sets, L and L_1, for each;

(3) the method in Section Ⅲ was used to extract features of the HMIFM from 320 experimental samples of the second-order derivative and non-second-order derivative, respectively, and form 12 dimensional feature vector sets, M and M_1, for each;

(4) the MFCC feature vectors were extracted from 320 experimental samples of the second-order derivative and non-second-order derivative, respectively, and formed 36 dimensional feature vector sets, N and N_1, for each; and

(5) all features of the 200 samples were taken out to constitute the training set, the remaining samples composed the testing set, and the SVM classifier was trained.

2. Experimental results analysis

Based on the above steps, the simulation experiments on the detection performance of LSB matching steganography were carried out with the feature set of the second-order derivative (L, M and N) and non-second derivative (L_1, M_1 and N_1), respectively, and the testing results are presented in Table 1 and Table 2. Table 1 (second derivative) shows that the MBTM method proposed in this study was superior to the HMIFD and MFCC solutions at different embedding rates except for 1%; however, the latter two features showed good detection results. From the comparison of Table 1 with Table 2, it was obvious that the detection accuracy of the first two features of the second-order derivative outperforms that of corresponding features of the non-second-order derivative at different embedding rates. However, the detection result of the MFCC feature of the second-order derivative was not superior to that of the non-second-order derivative. This might occur because MFCC is the characteristic parameter of speech signals based on the special perceptual properties of the human ear rather than

on statistics.

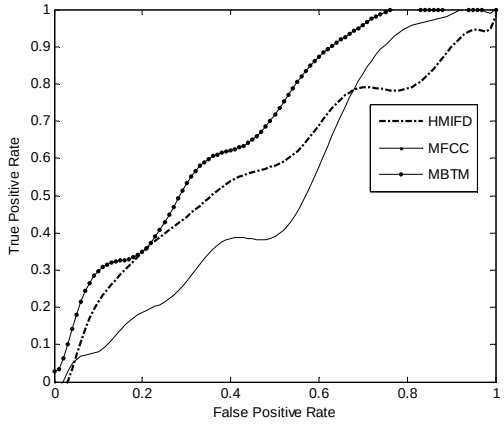
Table 1 Detection results of LSB steganography of the second derivative signals in different embedding rates

Embedding rates / Features	1%	3%	5%	8%	10%	30%	50%	80%
MBTM	55%	68.5%	75%	78.3%	81.7%	89.2%	89.5%	95.8%
HMIFD	50%	56.3%	55.2%	58.3%	59.4%	71.8%	74%	78%
MFCC	58.3%	57.3%	65%	62.5%	62.5%	59.4%	56%	75%

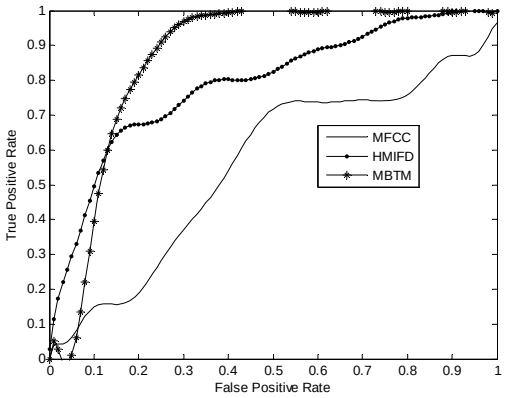
Table 2 Detection results of LSB steganography of non-second derivative signals in different embedding rates

Embedding rates / Features	1%	3%	5%	8%	10%	30%	50%	80%
MBTM	46.8%	59.2%	69.2%	72.5%	77.5%	82.5%	85.8%	89.5%
HMIFD	45.8%	45.8%	52.1%	63.5%	54.2%	69.7%	71%	75%
MFCC	45.8%	66.7%	69.2%	67.7%	63.5%	72.9%	68.8%	69.8%

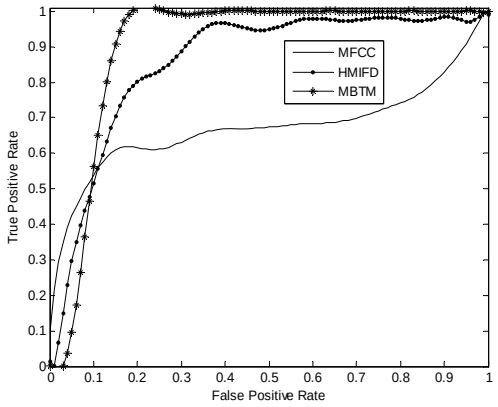
In addition, the ROC curves at the embedding rates of 3%, 30% and 80% were performed to compare the detection results of the three kinds of features in Table 1. As shown in Fig. 8 (a) (b) (c).



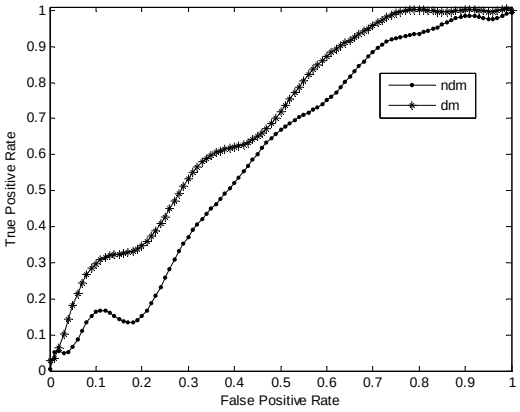
(a)



(b)



(c)



(d)

Fig. 8(a)(b)(c) ROC comparison curve of second derivative three types features at the embedding rates of 3%, 30% and 80%, (d) ROC comparison curve of second derivative and non-second derivative MBTM **features at the embedding rate of 3%**

The results of the MBTM features were optimal, followed by those of the HMIFD feature, while the results of the MFCC features were relatively weak. ROC curves of the detection results of MBTM features were constructed for comparing Table 1 (second derivative) and Table 2 (non-second derivative) at the embedding rate of 3%, as shown in Fig. 8 (d): dm indicates MBTM in Table 1 (second derivative), and ndm stands for the MBTM in Table 2 (non-second derivative). It is obvious that the detection accuracy of the dm feature is higher

than that of the ndm feature.

Ⅴ Conclusions and Innovation

Aiming at the difficulty of steganography detection with a low embedding rate, the detection algorithm for LSB matching steganography with low embedding rate was proposed on MBTM features by taking advantage of the second-order derivative speech signal WPC that can describe the signal details more accurately and combined MBTM that can more exactly express the correlation of WPC. First, the proposed algorithm was compared with the classic algorithm based on the HMIFD of WPC and the MFCC of the second derivative speech signal. Second, the detection results of three corresponding features of the second-order derivative and the non-second-order derivative speech signal were contrasted. Then, the SVM classifier was trained by a large number of VoIP voice streaming data. The accuracy of each feature of LSB matching steganography was tested at the embedding rates of 1%, 3%, 5%, 10%, 30%, 50%, and 80%. The experimental results show the following: (1) the algorithms with three features (the second derivative) have better detection for LSB matching steganography; (2) the detection accuracy based on the MBTM of the second derivative WPC is optimal and prominent especially at low embedding rates, and its execution time is longer than the latter two; (3) the detection accuracy of the first two features (the second derivative) is higher than that of the corresponding features (the non-second derivative), the accuracy is improved with the increased embedding rate, and the performance is more stable; and (4) the detection performance of the MFCC feature (second derivative) for each embedding rate is not better than that of the corresponding features (non-second derivative), and the detection accuracy is not improved with the increase of the embedding rate because the MFCC features of the speech signal are extracted by making use of the special perceptual properties of the human ear and not statistics.

Innovation 1: The Markov bidirectional transfer matrix was proposed as a matrix of speech signal detection

features, and the frame (256 samples or 32 ms) speech signal was used to reduce the feature dimension (62 dimension), which greatly decreased the computational complexity and improved the real-time detection.

Innovation 2: By combining the wavelet packet decomposition of the second derivative speech signal and the Markov features, the short-time correlation of the speech signals can be better described and the detection accuracy of PCM codes at low embedding rates can be improved greatly.

Acknowledgements

This work was supported in part by grants from the National Natural Science Foundation of China (No. 61402115). The authors would like to thank anonymous reviewers for their valuable suggestions.

Reference

[1] Huang Yongfeng, Shanyu Tang (2016) Covert voice over internet protocol communications based on spatial model. Science China Technological Sciences, No. 1, Vol. 59, pp. 117–128.

[2] Huang Yongfeng, Shanyu Tang, Yuan Jian (2011) Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec. IEEE Transactions on Information Forensics and Security, Vol. 6, No. 2, pp. 296–306.

[3] Huang Yongfeng, Liu Chenghao, Shanyu Tang (2012) Steganography Integration into a Low-Bit Rate Speech Codec. IEEE Transactions on Information Forensics and Security, Vol. 7, No. 6, pp. 1865–1875.

[4] Zhihua Xia, Xinhui Wang, Xingming Sun, and Baowei Wang (2014) Steganalysis of least significant bit matching using multi-order differences. Security and Communication Networks, Vol. 7, No. 8, pp. 1283–1291.

[5] Zhang Tao, Zhang Yan, Li Wenxiang, et al. (2009) Steganalysis of LSB matching exploiting correlations between pixel differences. Journal of Computer Research and development, Vol. 46 (Suppl), pp. 143–146 (in Chinese).

[6] Xiong Gang, Ping Xijian, Zhang Tao, and Li Kan (2013) Steganalysis of LSB Matching Based on the Measurement or the

Region Randomness. Journal of Computer Research and Development, Vol. 50, No. 5, pp. 942–950.

[7] Chen Beijing, Shu Huazhong, Coatrieux Gouenou, Chen Gang, Sun Xingming, Coatrieux Jean-Louis (2015) Color image analysis by quaternion-type moments. Journal of Mathematical Imaging and Vision, Vol. 51, No. 1, pp. 124–144.

[8] Zhao Yanli, Wang Xing (2013) Steganalysis of JPEG images based on bilateral transition probability matrix. Journal of Computer Applications, Vol. 33, No. 4, pp. 1074 -1076.

[9] Cho S, Cha B, Wang J (2010) Block-based image steganalysis: algorithm and performance evaluation. Proceedings of IEEE Int. Symp. Circuits and Systems. Piscataway, NJ. May 30- June 2, 2010, pp. 1679–1682.

[10] Pevný T, Fridrich J (2007) Merging Markov and DCT features for multi-class JPEG steganalysis. Proceedings of Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA. January 28, pp. 1–13

[11] Kodovský J, Fridrich J (2012) Steganalysis of JPEG images using rich models. Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIV, vol. 8303, San Francisco, CA. January 22, pp. 0A 1–13.

[12] Fridrich J, Kodovský J (2012) Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, Vol. 7, No. 3, pp. 868–882.

[13] He Z, Lu W, Sun W, et al (2012) Digital image splicing detection based on Markov features in DCT and DWT domain. Pattern Recognition, Vol. 45, No. 12, pp. 4292–4299.

[14] Xia Zhihua, Wang Xinhui, Sun Xingming, Liu Quansheng, Xiong Naixue (2016) Steganalysis of LSB matching using differences between nonadjacent pixels. Multimedia Tools and Applications, Vol. 75, No. 4, pp. 1947–1962.

[15] Tang Weixuan, Li Haodong, Luo Weiqi, Huang Jiwu (2016) Adaptive Steganalysis Based on Embedding Probabilities of Pixels, IEEE Transactions on Information Forensics and Security, Vol. 11, No. 4, pp: 734–745.

[16] Li Xiaolong, Zhang Weiming, Gui Xinlu (2015) Efficient reversible data hiding based on multiple histograms modification, IEEE Transactions on Information Forensics and Security, Vol. 10, No. 9, pp. 2016–2027.

[17] Li X, Zhang W, Gui X, Yang B (2013) A novel reversible data hiding scheme based on two-dimensional difference-histogram

modification. IEEE Transactions on Information Forensics and Security, Vol. 8, No. 7, pp. 1091–1100.

[18] Qin Chuan, Chang Chin-Chen, Huang Ying-Hsuan (2013) An inpainting-assisted reversible steganographic scheme using a histogram shifting mechanism. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 23, No. 7, pp. 1109–1118.

[19] Qin C, Hu YC (2016) Reversible data hiding in VQ index table with lossless coding and adaptive switching mechanism. Signal Processing, Vol. 129, pp. 48–55.

[20] Huang YF, Tang S, Zhang, Y (2011) Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis. IET Communications, Vol. 5, No. 7, pp. 929–936.

[21] Tao H, Sun D, Huang Y (2014) A detection method of subliminal channel based on VoIP communication. Proceedings of the 1st International Workshop on Information Hiding and its Criteria for Evaluation. New York, USA, June, 2014, pp. 37–41.

[22] Huang Y, Tang S, Bao C, Yip Y J (2011) Steganalysis of compressed speech to detect covert voice over Internet protocol channels. IET Information Security, Vol. 5, No. 1, pp. 26–32.

[23] Li S, Tao H, Huang Y (2012). Detection of QIM steganography in G.723.1 bit stream based on quantization index sequence analysis. Journal of Zhejiang University-SCIENCE C (Computers & Electronics), Vol. 13, No. 8, pp. 624–634

[24] Yan Diqun and Wang Rangding (2014) Detection of m-p3stego exploiting recompression calibration-based feature. Multimedia Tools & Applications, Vol. 72, No. 1, pp. 865–878.

[25] Guo Honggang, Yan Diqun, Wang Rangding, Wang Zhangyan, Wang Lin, Linqiang T. U. (2015) MP3 steganalysis based on difference statistics. Computer Engineering and Applications, Vol. 51, No. 7, pp. 88–92.

[26] Liu Q, Sung AH, Qiao M (2011) Derivative-based audio steganalysis. ACM Trans. Multimedia Comput. Commun. Vol. 7, Iss. 3, Article No. 18, pp. 1-18.

[27] Zhang Mining, Yang Gang, Zhang Zhen (2014) Noise Endurance Steganalysis Algorithm Based on WPD. Journal of Chinese Computer Systems. Vol. 35, No. 4, pp. 941–944.

[28] Gu Bin, Sheng Victor S., Wang Zhijie, Ho Derek, Osman Said, Li Shuo (2015) Incremental learning for ν-Support Vector

Regression. Neural Networks, Vol. 67, pp. 140–150.

Wanxia Yang's PhD Principal Supervisor is Professor Shanyu Tang who is Chair Professor of Information Security at the University of West London, United Kingdom.

Shanyu Tang (A'08–M'08–SM'10) received the Ph.D. degree from Imperial College London, United Kingdom in 1995.

Professor Shanyu Tang is currently Chair Professor of Information Security at the University of West London, UK. He was Distinguished Professor of Information Security in the School of Computer Science at China University of Geosciences from 2012-2016. Professor Tang is dedicated to adventurous research in fractal computing methods for covert communications, multimedia security, and digital steganography. He is the principal grant holder of eight externally funded research projects including three research grants from the UK government. He has contributed to 99 scientific publications — 60 refereed journal papers including IEEE/ACM TRANSACTIONS and IET journal papers, and held one patent.