Blind image steganalysis based on wavelet coefficient correlation

Han Zong, Fen-lin Liu, Xiang-yang Luo*

Zhengzhou Information Science and Technology Institute, Henan 450002, PR China

ARTICLE INFO

Article history:
Received 25 November 2010
Received in revised form 31 January 2012
Accepted 26 February 2012

Keywords:
Steganalysis
Blind steganalysis
Wavelet decomposition
Joint probability density matrix
Coefficient correlation
Co-occurrence matrix

ABSTRACT

To detect the presence of information in a stego image more reliably, a blind JPEG steganalysis method based on inter- and intra-wavelet subband correlations in the wavelet domain is proposed. First, after two-level wavelet decomposition, the joint probability density of each subband’s difference from neighboring coefficients in the horizontal, vertical, and diagonal directions is calculated, and the entropy and energy are extracted from the joint probability density matrix as features. Then the image is decomposed into three subbands, and the PDF (probability density function) is extracted from each subband’s wavelet coefficient. Finally, the three kinds of features described above are combined to detect the image. In experiments, the proposed method is compared with various other blind steganalysis methods, and the impacts of different feature combinations on detection accuracy are discussed. Experimental results from typical JPEG image stego algorithms such as F5, Jsteg, Outguess, and Jphide show that the proposed method significantly outperforms typical blind steganalysis methods. The proposed method also has some detection capabilities for double-compressed images.

1. Introduction

Steganography is the art of hiding the very act of communicating such things as digital images by embedding secret messages into innocuous visible covers. Nowadays, steganography has become a serious challenge to digital forensics. As the opposite technology of steganography, steganalysis is used mainly to detect stego covers and to extract information. On a more general level, steganalysis forms part of digital forensics technology. Blind steganalysis, which means the identification of suspicious images, is the first step in steganalysis (Fridrich et al., 2005). Blind steganalysis does not try to determine the specific hiding method, but makes correct identifications of detected objects that may contain hidden information. Therefore, compared with specific steganalysis methods, blind steganalysis is more practical (McBride et al., 2005). JPEG images have become the most widely used image format because of their low data transmission requirements and robustness to various kinds of interference. In fact, steganography and steganalysis of JPEG images have attracted much attention.

To some extent, steganalysis can be classified into two classes: signature steganalysis and statistical steganalysis (McBride et al., 2005). In general, statistical steganalysis techniques are better than signature steganalysis techniques, and many popular algorithms belong to the former class (Nissar and Mirb, 2010). A blind steganalysis algorithm based on the high-order statistics of wavelet coefficients and mainly using weak correlations between the coefficients in the wavelet domain was proposed by Farid and Lyu (2002). The specific algorithm uses fourth-order PDF (probability density function) moments, which are first extracted from the three level wavelet subband coefficients in the vertical, horizontal, and diagonal directions, and then the fourth-order PDF moments from each subband coefficient’s log-linear error are combined to yield the feature. This algorithm can effectively detect stego images embedded by Jsteg (Hsu and Wu, 1999), Outguess (Provos, 2001), Steghide (Hetzl, 2003), and other methods. In Lyu and Farid (2004), the method proposed in Farid and Lyu (2002) was improved by extracting the fourth-order PDF...
moments from each color component’s wavelet coefficients and their log-linear forecast errors to detect the stego image. Another algorithm which combines the fourth-order moments of the wavelet coefficients, the fourth-order moments of the wavelet coefficients’ forecast errors, and the phase statistics as features to detect stego images was proposed in Lyu and Farid (2006). In addition, four kinds of feature vectors (edge statistics, error statistics, phase statistics, and combinations of these) were investigated separately. The results showed that a combination of the three features was better than any of the features separately. In Goljan et al. (2006), an improved method based on the WAM (wavelet absolute method) was proposed. Absolute moments of high-order PDFs (probability density functions) were extracted from the wavelet coefficients as features. These features were found to be more sensitive and better for classification.

In Shi et al. (2005) and Xuan et al. (2005), a kind of blind steganalysis algorithm based on the CF (characteristic function) moments extracted from a wavelet coefficient histogram was proposed. The algorithm first transforms an image into two classes to obtain nine subbands (including the original image), then calculates each subband coefficient’s histogram characteristic function, and then calculates both the first- and second-order statistical moments of the function. The method was found to give better detection results than F5, Outguess, and other algorithms. It has been pointed out in Xuan et al. (2005) that the n-th-order moment of a characteristic function is related to the n-th-order moment derived from a wavelet coefficient histogram, and therefore the embedded information will lead to a change in the characteristic function moment. An image is first transformed into three classes to obtain 13 subband coefficients (including the original histogram); then the characteristic function moments are extracted as features for steganalysis. These two features (CF moments and PDF moments) have been analyzed theoretically and experimentally in Wang and Moulin (2007). It can be concluded from the analysis described above that for wavelet coefficients, the characteristic function moments are more sensitive to embedding of information and more conducive to classification than the PDF moments, but that for predicting the errors of wavelet coefficients, the PDF moments are more sensitive than the CF moments.

Image statistical features were obtained from DCT coefficients in Xuan et al. (2006). The co-occurrence matrices were calculated from DCT or transformed DCT coefficients, and parts of the data in a co-occurrence matrix were used directly as features. In Shi et al. (2007), features extracted from inter- and intra-wavelet blocks of DCT coefficients were able to reflect a certain number of characteristics of the original JPEG image, namely the histogram, the co-occurrence matrix, and other statistical characteristics. Chen and Shi (2008) have shown that the correlations of DCT coefficients in the inter- and intra-wavelet blocks should be considered and that the DCT coefficients should be re-mixed. Finally, the co-occurrence matrix statistics of the changed data in the four directions (horizontal, vertical, 45° diagonal, and 135° diagonal) were used as features. The test results were better than for the tests described earlier.

The algorithms described above still have some problems: existing algorithms for the wavelet domain either consider only the one-dimensional statistical properties of the wavelet coefficients or consider only the local statistical characteristics of the subband wavelet coefficients. In the DCT domain, although the algorithms described above extract two-dimensional statistical features and the correct detection ratio is fairly high, the weak point is that the dimensionality of the extracted features is too high and the calculations are complex. Moreover, the images have a soft relationship between the image rows and image columns (Gul and Kurugollu, 2010).

Although the correlation between the pixels of an image is removed when the image is transformed in a wavelet domain, the scale continuity and clustering of the wavelet coefficients can be fully reflected. To enhance the accuracy of JPEG stego-detection further, this research has used the correlation of the wavelet coefficients from the inter- and intra-wavelet subbands in the wavelet domain to propose a multi-feature blind steganalysis algorithm for JPEG images. First, after decomposition of the two-order wavelets, the joint probability density of each subband’s difference from adjoining coefficients in the horizontal, vertical, and diagonal directions is calculated. The entropy and energy are extracted from the joint probability density matrix as features; then the image is decomposed into three subbands; the PDF is extracted from each subband’s wavelet coefficient; the three features are combined; and finally a backpropagation (BP) neural network classifier is used to detect the image. Experimental results have shown that the detection accuracy of this algorithm is much improved over that of typical existing steganalysis algorithms and that the proposed algorithm also provides detection accuracy for double-compressed images. In addition, a number of digital investigations of the test results have been performed, and various impacts of different feature combinations on the test results have been discussed. This research has also shown that even collecting all the features of an original image may not yield a better detection result because compatibility or repellency still exist between pairs of features. After testing, a combination of the features mentioned in this article, namely the entropy, energy, and PDF moments, has been shown to offer the best detection performance.

2. Source structure for feature extraction

In the source structure for feature extraction, an image is decomposed into two levels in the wavelet domain. Then, according to the inter- and intra-wavelet subband correlation coefficients, the joint probability density of each subband’s difference from adjoining coefficients in the horizontal, vertical, and diagonal directions is calculated. Two kinds of matrix that are sensitive to embedded image information will be used to extract statistical features.

2.1. Wavelet decomposition

Although the wavelet transform was developed based on the Fourier transform, the two exhibit substantial differences (Gul and Kurugollu, 2010). The Fourier transform uses
a separate time domain or frequency domain to represent the signal characteristics, while the wavelet transform uses a combination of the time domain and the frequency domain. For blind steganalysis and other steganography algorithms, the embedded location of the image is unknown. Moreover, after a DCT transform, the image noise gathers at the high-frequency DCT coefficients, while the detail information in the image gathers at the intermediate-frequency DCT coefficients. In a JPEG steganography algorithm, the secret information is always embedded in the intermediate-frequency coefficients, that is to say, embedded in some of the details of the image. However, the high-frequency sub-band coefficients of the wavelet transform also reflect certain details of the image. After the secret information has been embedded, the change in the image may be reflected in the high-frequency wavelet subband coefficients. In addition, after the secret information has been embedded in the intermediate-frequency coefficients, the change in the image is small. Nevertheless, even if only one bit is changed, the change of pixel value in the spatial domain becomes greater.

By wavelet transform theory, if the change of pixel value in the spatial domain is greater, the high-frequency wavelet coefficients will undergo greater changes also. Therefore, the information from stego image detection using wavelet decomposition may be able to yield better test results.

In this research, before the feature extraction process, the image was first decomposed into two or three levels using the specific two-level decomposition process shown in Fig. 1.

Fig. 1(a) shows a JPEG format image of size 512 \times 512 called Lena, and Fig. 1(b) shows a two-level wavelet decomposition map of Lena.

2.2. Adjacent wavelet coefficient differences

In the wavelet domain, subtracting two adjacent wavelet coefficients reflects the relationship between pixel blocks constructed of four adjacent pixels up and down in the spatial image domain, as shown in Fig. 2.

\[
\begin{matrix}
  n_{11} & n_{12} \\
  n_{21} & n_{22}
\end{matrix}
\]

\[
\begin{matrix}
  n_{13} & n_{14} \\
  n_{23} & n_{24}
\end{matrix}
\]

\[N_1\]

\[N_2\]

Fig. 2. Relationship between pixel blocks constructed of four adjacent pixels up and down in the spatial image domain.

Suppose that \(N_1\) and \(N_2\) are two adjacent subband wavelet coefficients after the Haar wavelet transform and that they are constructed from the four adjacent pixels \(n_{11}, n_{12}, n_{21}, n_{22}\) and \(n_{13}, n_{14}, n_{23}, n_{24}\) in the spatial domain (as shown in Fig. 2):

\[
N_1 = \frac{1}{2} \left( \frac{1}{2} (n_{11} + n_{12}) + \frac{1}{2} (n_{21} + n_{22}) \right)
\] (1)

\[
N_2 = \frac{1}{2} \left( \frac{1}{2} (n_{13} + n_{14}) + \frac{1}{2} (n_{23} + n_{24}) \right).
\] (2)

Then:

\[
M_1 = N_1 - N_2 = \frac{1}{2} \left( \frac{1}{2} (n_{11} + n_{12}) + \frac{1}{2} (n_{21} + n_{22}) \right)
- \frac{1}{2} \left( \frac{1}{2} (n_{13} + n_{14}) + \frac{1}{2} (n_{23} + n_{24}) \right).
\] (3)

Eqs. (1)–(3) reveal that the wavelet coefficient differences (denoted as ‘\(\bigcirc\)’ in Fig. 3(b)) reflect not only the relationship between the two pixel blocks, each composed of four pixels \(n_{ij}\) (denoted as ‘\(\bigast\)’ in Fig. 3(a)) in the spatial domain, but also the relationship between these two coefficients (denoted as ‘\(\bigbullet\)’ in Fig. 3(b)). In the approach described in this section, these differences are calculated between adjacent coefficients in every wavelet subband, as shown in Fig. 3.

Fig. 3(a) shows the differential direction of any two wavelet subbands, while Fig. 3(b) shows the differential locations of the two wavelet coefficients.

2.3. Co-occurrence matrix of wavelet coefficients

2.3.1. Co-occurrence matrix calculation for the inter-wavelet subband

In the past, wavelet coefficient statistics were based on the histogram method, and each coefficient was dealt with independently. However, a weak correlation still existed among the wavelet coefficients (Orchard and Ramchandran, 1998). Fig. 4 shows the standard JPEG format image called Lena, with a size of 512 \times 512 pixels and a compression rate of 75%. Fig. 5 shows the stego image in which the F5 steganography algorithm has been used to embed an image with an embedding ratio of 50%. The statistics of the joint probability density function of the
two-level wavelet coefficients in the inter-wavelet subband in the 90°/14 direction were calculated. The results and their difference binary images are shown in Fig. 6. Points for which the co-occurrence matrix element is zero are shown in black, while others are shown in white.

On the surface, from Figs. 4 and 5, there seems to be no different between the cover and stego binary images, but from the difference binary images, it can be determined that when the information was embedded, the wavelet coefficient correlation matrix for the cover image was altered. From this, it is possible to reconstruct the co-occurrence matrix used in the wavelet subband differential matrix.

2.3.2. Co-occurrence matrix calculation for the intra-wavelet subband

In the wavelet domain, in addition to the correlation between the inter-wavelet subband coefficients, there exists some weak correlation among the intra-wavelet subband coefficients. The correlation in the intra-wavelet subband reflects the properties of image texture and edge structure, and the process of stego image embedding often leads to changes in these properties (Burrus, 1998).

The joint probability density function of the intra-wavelet subband reflected in the spatial domain represents the internal characteristics of the image block. Take the Haar wavelet as an example. As above, suppose that $N_3$ and $N_4$ are two adjacent vertical subband wavelet coefficients after the Haar wavelet transform and that they are constructed from the four adjacent pixels $n_{11}, n_{12}, n_{21}, n_{22}$ and $n_{13}, n_{14}, n_{23}, n_{24}$ in the spatial domain:

$$N_3 = \frac{1}{2} \left( \frac{1}{2} (n_{11} - n_{12}) + \frac{1}{2} (n_{21} - n_{22}) \right)$$

(5)

$$N_4 = \frac{1}{2} \left( \frac{1}{2} (n_{13} - n_{14}) + \frac{1}{2} (n_{23} - n_{24}) \right).$$

(6)

Then:

$$M_2 = N_3 - N_4 = \frac{1}{2} \left( \frac{1}{2} (n_{11} - n_{12}) + \frac{1}{2} (n_{21} - n_{22}) \right)$$

$$- \frac{1}{2} \left( \frac{1}{2} (n_{13} - n_{14}) + \frac{1}{2} (n_{23} - n_{24}) \right).$$

(7)

From Eqs. (3) and (5)–(7), the joint probability density of the intra-wavelet subband coefficients $M_1$ and $M_2$ in the spatial domain reflects the correlation between the two

---

**Fig. 3.** Differences between coefficients in wavelet subbands: 3(a) relationship between two coefficients; 3(b) relationship between the two pixel blocks.

**Fig. 4.** Lena cover image.

**Fig. 5.** Lena stego image.
pixel blocks which are made up of $n_{11}$, $n_{12}$, $n_{21}$, $n_{22}$ and $n_{13}$, $n_{14}$, $n_{23}$, $n_{24}$.

The intra-wavelet subband joint probability density matrix is represented as a binary image, and a comparison of results for cover and stego images is shown in Fig. 8. The experimental image is the standard JPEG format image called Lena. The F5 algorithm was used to embed the information into the image. The calculated results for the intra-wavelet subband and the difference between the cover and the stego image are shown as binary images in Fig. 8. When the co-occurrence matrix element is zero, the corresponding point is represented as black, otherwise as white.

From Fig. 8, it is apparent that according to the joint probability density statistic, there exists a difference between the cover and stego images; the original characteristics of the cover image have been altered.

3. Feature extraction

In the past, a threshold calculated from a co-occurrence matrix or a joint probability density matrix was used in the algorithm for blind detection. The elements that exceed the threshold form a new matrix as features to detect stego images. The method may lead to an increase in feature dimensionality. In the wavelet domain, the sum of the wavelet coefficients is far greater than the coefficients in the DCT domain. If the elements that exceed the threshold are used as features to detect the stego image directly, feature extraction will become more difficult because not all elements are useful. Therefore, it is necessary to filter or revise the elements coming from the co-occurrence or joint probability density matrices.

3.1. Several types of features

In Haralick (1979), 14 kinds of statistical measures, including entropy, energy, moment of inertia, and variance, were extracted from the co-occurrence matrix. In Andrea and Flavio (1995), it was shown that features such as entropy, energy, and moment of inertia have better statistical properties than the others mentioned in this paper. In Farid and Lyu (2002), PDF higher-order statistical moments...
were extracted from the wavelet coefficients as features, an approach which gave good detection results for Jsteg, Outguess, and other algorithms. The statistical algorithms described following are examples of commonly used basic concepts. An introduction to the basic concepts of these statistics will now be provided.

(1) Entropy of the co-occurrence matrix

The entropy of the co-occurrence matrix in a spatial domain is a measure of image clutter. For a grayscale image, if the image has more texture, its entropy value becomes higher; if less, the value becomes lower. The entropy can be expressed as follows:

$$ent_1 = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} g(i,j) \log(g(i,j)), \quad (8)$$

where $g(i,j)$ is a co-occurrence matrix, $i$, $j$ are rows and columns of this matrix, and ranges are ($N = 0,1,\cdots,N_g-1$).

(2) Energy of the co-occurrence matrix

The energy of the co-occurrence matrix in a spatial domain is also called the second angular moment. If the contrast of an image is stronger and the texture is clearer, then the moment of inertia is greater. The energy can be expressed as:

$$ent_2 = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} g^2(i,j), \quad (9)$$

where $g(i,j)$ is a co-occurrence matrix, $i$, $j$ are rows and columns of this matrix, and ranges are ($N = 0,1,\cdots,N_g-1$).

(3) Moment of inertia of the co-occurrence matrix

The moment of inertia of the co-occurrence matrix in a spatial domain reflects the clarity of the image. In other words, if the contrast of an image is stronger and the texture is clearer, then the moment of inertia is greater. The moment of inertia can be expressed as:

$$ent_3 = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i-j)^2 g(i,j), \quad (10)$$

(4) PDF moment of the wavelet coefficient

Suppose that $X = (x_1,x_2,\cdots,x_N)$ is an independent and identically distributed sequence, its probability density function is $P(x)$, and its nth-order PDF moment is:

$$m_n = EX^n = \int_{-\infty}^{\infty} p(x)x^n dx. \quad (11)$$

For an independent and identically distributed random sequence $X = (x_1,x_2,\cdots,x_N)$ of probability density functions $p(x)$, its nth-order PDF moment can be calculated using an unbiased estimation method as follows:

$$\hat{m}_n = \frac{1}{N} \sum_{i=1}^{N} x_i^n, \quad n \geq 1. \quad (12)$$

The four kinds of features described above can detect images effectively. However, after further experiments, it was found that not all combinations of features led to a better detection result because compatibility and repulsion exist among the features. The combination of features chosen in this research (entropy, energy, and PDF moments) has better detection performance than other combinations studied.

3.2. Feature extraction algorithm

(1) Feature extraction algorithm using the inter-wavelet subband coefficient

**Input:** The group of pictures to be detected.

**Output:** Detection results for the pictures and the degree of accuracy for each.

**Step:**

1. Training image sample collection. The collection of image samples should refer to the detection of image formats (BMP, JPEG, etc.), content, texture, color and other characteristics; similar image samples should be selected and tested as training samples.
II. Image decomposition. The training image and testing image should be decomposed into three levels in a wavelet domain. In addition, all 12-wavelet subbands including the original image should be obtained.

III. Feature extraction.

a) Calculate the adjacent coefficients’ difference in the first- and second-level subbands (including the original) in the horizontal, vertical, and diagonal directions.

b) Calculate the co-occurrence matrix of the differential coefficients for adjacent inter-wavelet subbands.

c) Calculate the entropy and energy of the co-occurrence matrix as a feature, obtaining a 54-D feature.

d) Calculate PDF moment statistics for all three level high-frequency subband wavelet coefficients, thus obtaining a 36-D feature.

(2) Feature extraction algorithm using intra-wavelet subband coefficients

**Input:** The group of pictures to be detected.

**Output:** The results for the pictures and the degree of accuracy for each.

**Step:**

I. Training image sample collection. The collection of image samples should refer to the detection of image formats (BMP, JPEG, etc.), content, texture, color, and other characteristics; similar image samples should be selected and tested as training samples.

II. Image decomposition. The training image and the testing image should be decomposed into three levels in a wavelet domain. In addition, all 12-wavelet subbands including the original image should be obtained.

III. Feature extraction.

a) Calculate the difference between each pair of adjacent coefficients in the first- and second-level subbands (including the original) in the horizontal, vertical, and diagonal directions.

b) Calculate the joint probability density statistic for the differential matrix for the adjacent intra-wavelet subband coefficients.

c) Calculate the entropy and energy for the co-occurrence matrix, obtaining a 36-D feature.

Combine the extracted features described above to obtain a 126-D feature for use in classification.

### 3.3. Pretreatment of classified features

To improve classification accuracy, feature pretreatment should be performed before classification; the method is mentioned in Wang (2007). For some feature \( f \), its normalized features are \( \tilde{f} \):

\[
\tilde{f} = (f - f_{\text{min}}) / (f_{\text{max}} - f_{\text{min}}),
\]

where \( f_{\text{max}} \) and \( f_{\text{min}} \) are the maximum and minimum results for the training image. In essence, this normalization operation can avoid the excessive influence of outlier feature values, not only improving the performance of an FLD classifier significantly, but also effectively improving the classification performance of a BP neural network classifier.

### 4. Neural network classifier design

A backpropagation (BP) neural network is a theoretical mathematical model of a human neural network. A BP neural network imitates the neural network structure of a human brain to establish a signal processing system. In addition, a BP neural network connected by a large number of neurons forms a complex network with a high degree of nonlinearity. The input node is the first layer of neurons, the output node is the last layer of neurons, and the middle layers (one or more) between the first and last layers are called hidden-layer neurons. A BP network model can be divided into learning and classification stages.

The BP neural network used in this research has three layers, with the number of input layer neurons equal to 126, corresponding to the 126-dimensional features. There are five hidden-layer neurons plus the output layer with one neuron. The output range of the output layer neuron is [0, 1]. The decision threshold is \( \text{Thd} = 0.5 \). If the output value is less than \( \text{Thd} \), then the output value can be classified as a cover image; otherwise, the output is judged to contain stego images. The network-based expected minimum error is set to 0.01, and the maximum number of iterations is set to 3000.

### 5. Experimental results and analysis

#### 5.1. Experimental setup

The original images chosen for the experiment came from the NRCS image database (NRCS Photo Gallery, April 2004) plus some common standard images. There were 2056 images in total, of which 1028 were JPEG images (the standard images included Lena, Baboo, Peppers, and 24

<table>
<thead>
<tr>
<th>Embedding method</th>
<th>Number of cover images</th>
<th>Number of stego images</th>
<th>Image format</th>
<th>Type of embedding ratio</th>
<th>Size or percentage of embedded information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outguess</td>
<td>1028</td>
<td>1028 \times 3 = 3084</td>
<td>JPEG</td>
<td>3</td>
<td>256 \times 256; 64 \times 64; 16 \times 16</td>
</tr>
<tr>
<td>Jsteg</td>
<td>1028</td>
<td>3084</td>
<td>JPEG</td>
<td>3</td>
<td>256 \times 256; 64 \times 64; 16 \times 16</td>
</tr>
<tr>
<td>F5</td>
<td>1028</td>
<td>3084</td>
<td>JPEG</td>
<td>3</td>
<td>200 \times 200; 64 \times 64; 16 \times 16</td>
</tr>
<tr>
<td>Jhhide</td>
<td>1028</td>
<td>3084</td>
<td>JPEG</td>
<td>3</td>
<td>256 \times 256; 64 \times 64; 16 \times 16</td>
</tr>
<tr>
<td>F5</td>
<td>1100</td>
<td>1100 \times 3 = 3300</td>
<td>JPEG double compress</td>
<td>3</td>
<td>30%; 50%; 100%</td>
</tr>
</tbody>
</table>
other JPEG images). The image sizes included 68 × 512, 640 × 480, 512 × 512, and three other sizes. The double-compressed JPEG images chosen for this experiment were photographic images, 1100 images in total, with an image size of 800 × 600. The images cover a wide range of subjects, including natural scenery, manufacturing facilities, and human portraits.

Based on the original image group and the five kinds of steganographic methods described earlier, the corresponding group containing secret images was generated according to the image format that was suitable for the different steganographic methods. Jsteg, Outguess, Jphide (Latham, 2003), and F5 (Wesfeld, 2001) steganographic methods were used to embed different percentages of information in 1028 JPEG images. Three kinds of images with sizes of 256 × 256, 64 × 64, and 16 × 16 were embedded in the image using the Jsteg, Outguess, and Jphide methods respectively: another group of 200 × 200, 64 × 64, and 16 × 16 images was embedded in the image using the F5 embedding method. Certain percentages of random information were also embedded in the 1100 double-compressed JPEG images using the F5 steganographic method.

The reason for embedding different percentages of information was on the one hand, to take into account the variety of steganographic methods and the hiding capacity of the image itself, and on the other hand, to test the performance of the detection algorithm with different percentages of information embedded. Results for 14 experimental groups were obtained: two groups of cover images and one 24-group of stego images, with each group including 1028 images. The experimental images were as given in Table 1. Haar wavelets were used to decompose the experimental images. The experiment was divided into three parts: (1) testing results for different feature combinations; (2) testing results for the four typical steganographic algorithms and comparison with the results obtained by existing blind steganalysis methods; (3) testing for double-compressed images.

This algorithm is comparable to that used in Farid and Lyu (2002), Goljan et al. (2006), Xuan et al. (2005), and Chen and Shi (2008). For classification, support vector machines (SVM) were used to train and test the samples in Farid and Lyu (2002), Goljan et al. (2006), Xuan et al. (2005), and Chen and Shi (2008); to determine the RBF inner dot product, the $r$ and $c$ parameters were set to $r = 0.5$, $c = 512$. This algorithm uses a BP neural network as a classifier and specific sets of parameters, which are presented in Table 1. The SVM was used for training and testing with double-compressed images. The images were selected from the NRCS Image Database and from photographs. The experimental environment is described in Table 2, and the test results are shown in Table 3. The test results for double-compressed images are shown by the ROC curve in Fig. 9.

From these experimental results, it is clear that the correct detection ratio obtained by the proposed algorithm was much higher than those indicated in Table 3. The chosen features were also more sensitive to embedded information. Table 3 shows that the 126-D features mentioned in this paper achieved detection rates greater than or equal to 90%, which compares favorably to F5, Jphide, Outguess, Jsteg, etc., with a false positive rate of 4.67%. However, as the percentage of embedded information decreases, the correct detection ratio will also decline.

5.2. Detection experiments on different feature combinations

The experiment included two main steps:

First, the comprehensive effectiveness of different feature combinations was compared (comprehensive training and testing): energy, entropy, and moment of inertia; energy, entropy, moment of inertia, and PDF; energy, entropy, and PDF moments; entropy and moment of inertia; energy and entropy; energy and moment of inertia. Comprehensive training and testing means training with all kinds of features that have been extracted from the image, just as for the cover image and the steg image including both F5 and Jsteg images, to test the other part of the image.

Second, the testing effectiveness of different feature combinations was compared separately (individual training and testing): energy, entropy, and moment of inertia; energy, entropy, moment of inertia, and PDF; energy, entropy, and PDF moments; entropy and moment of inertia; energy and entropy; energy and moment of inertia. Individual training and testing means training with one kind of features that has been extracted from the image, just as for the cover image and the steg image including both F5 and Jsteg images, to test the other part of the image.

In Table 4, the average correct detection ratio of energy, entropy, and PDF moments taken jointly was

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Experimental parameters of some typical blind steganalysis algorithms and proposed algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features: own domain</td>
<td>DWT</td>
</tr>
<tr>
<td>Classifier</td>
<td>SVM</td>
</tr>
<tr>
<td>Number of training samples</td>
<td>6168</td>
</tr>
<tr>
<td>Number of testing samples</td>
<td>6168</td>
</tr>
<tr>
<td>Experimental images NRCS 1028 images</td>
<td></td>
</tr>
<tr>
<td>Test algorithm F5, Jsteg, Jphide, Outguess</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Comparison with other steganalysis methods.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F5</td>
<td>77.71%</td>
</tr>
<tr>
<td>Jphide</td>
<td>84.89%</td>
</tr>
<tr>
<td>Jsteg</td>
<td>95.81%</td>
</tr>
<tr>
<td>Outguess</td>
<td>75.33%</td>
</tr>
<tr>
<td>Cover image</td>
<td>94.20%</td>
</tr>
<tr>
<td>False positive</td>
<td>5.8%</td>
</tr>
</tbody>
</table>
96.86%; this classification result is the best of all those shown in Table 4. However, Table 5 also shows that compared with several other feature combinations, entropy, energy, and PDF moments are the ideal combination. Table 5 also shows that compatibility and repulsion exist among the features. Combining all the features to detect the image may not obtain the most effective detection results.

**Fig. 9.** ROC curves for detecting F5 stego images using Goljan’s algorithm and the proposed algorithm. In 9(a), 9(b), and 9(c), the embedding ratio is 30%, 50%, and 100% respectively.
Combination of several features with separate detection rates.

<table>
<thead>
<tr>
<th>Feature combination/ embedding method</th>
<th>F5</th>
<th>Jsteg</th>
<th>Jphide</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy, entropy, moment of inertia, PDF</td>
<td>95.03%</td>
<td>96.37%</td>
<td>95.20%</td>
<td>95.53%</td>
</tr>
<tr>
<td>Energy, entropy, moment of inertia</td>
<td>92.57%</td>
<td>69.18%</td>
<td>84.87%</td>
<td>82.21%</td>
</tr>
<tr>
<td>Energy, entropy, PDF</td>
<td>97.39%</td>
<td>97.95%</td>
<td>95.24%</td>
<td>96.86%</td>
</tr>
<tr>
<td>Energy, entropy</td>
<td>95.96%</td>
<td>75.83%</td>
<td>72.65%</td>
<td>81.48%</td>
</tr>
<tr>
<td>Entropy, moment of inertia</td>
<td>84.87%</td>
<td>72.65%</td>
<td>97.44%</td>
<td>84.57%</td>
</tr>
<tr>
<td>Energy, moment of inertia</td>
<td>92.31%</td>
<td>60.66%</td>
<td>84.99%</td>
<td>79.33%</td>
</tr>
</tbody>
</table>

The testing results for the F5, Jsteg, Jphide, and Outguess steganographic algorithms with different embedding rates are shown in Table 6.

From Table 6, it can be seen that the features extracted by the proposed algorithm produced better detection results on stego images than the F5, Jphide, Outguess, and Jsteg steganographic algorithms. The observed accuracy was greater than 90%, and the correct detection ratio of original images was 95.33%, giving a false positive rate of 4.67%. However, when the percentage of embedded information decreases, the correct detection ratio will also decline.

5.3. Detection experiments on double-compressed images

To investigate performance on double-compressed images, the ROC curve was used. The results of the proposed algorithm were compared with those from Goljan’s algorithm. In addition, the SVM classifier with 3 × 3 × 9 = 81 features (Goljan’s algorithm) was trained separately for 1100 images divided into 550 cover and 550 stego images. All 550 original images were double-compressed colored images taken with a camera. The SVM classifier with 126 features (proposed method) was also trained separately for 1100 images divided into 550 cover and 550 stego images. All 550 original images were double-compressed colored images taken with a camera. F5 steganographic methods were then used to embed information into 1100 double-compressed JPEG images at different rates. The results are shown in Fig. 9, which presents the ROC curves of the detection results for the F5 double-compressed stego images using Goljan’s algorithm and the proposed algorithm. In Fig. 9(a)–(c), the embedding ratio was 30%, 50%, and 100% respectively. It is apparent from Fig. 9 that both methods provided effective detection. However, the proposed algorithm had better performance than Goljan’s method.

Table 5

<table>
<thead>
<tr>
<th>Feature combination/ embedding method</th>
<th>F5</th>
<th>Jsteg</th>
<th>Jphide</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy, entropy, moment of inertia, PDF</td>
<td>90.53%</td>
<td>92.41%</td>
<td>94.68%</td>
<td>71.07%</td>
</tr>
<tr>
<td>Energy, entropy, moment of inertia</td>
<td>95.66%</td>
<td>70.13%</td>
<td>91.41%</td>
<td>84.87%</td>
</tr>
<tr>
<td>Energy, entropy, PDF</td>
<td>97.39%</td>
<td>97.95%</td>
<td>95.34%</td>
<td>95.33%</td>
</tr>
<tr>
<td>Energy, entropy</td>
<td>90.01%</td>
<td>66.77%</td>
<td>93.97%</td>
<td>84.63%</td>
</tr>
<tr>
<td>Entropy, moment of inertia</td>
<td>84.84%</td>
<td>63.28%</td>
<td>77.34%</td>
<td>97.44%</td>
</tr>
<tr>
<td>Energy, moment of inertia</td>
<td>78.91%</td>
<td>55.13%</td>
<td>62.38%</td>
<td>84.99%</td>
</tr>
</tbody>
</table>

Integrated testing results for four kinds of steganographic algorithms.

<table>
<thead>
<tr>
<th>Embedding method/image size</th>
<th>256 × 256/200 × 200</th>
<th>128 × 1828 64 × 128</th>
</tr>
</thead>
<tbody>
<tr>
<td>F5</td>
<td>97.47%</td>
<td>97.28%</td>
</tr>
<tr>
<td>Jsteg</td>
<td>98.83%</td>
<td>98.25%</td>
</tr>
<tr>
<td>Jphide</td>
<td>99.61%</td>
<td>94.34%</td>
</tr>
<tr>
<td>Outguess</td>
<td>96.89%</td>
<td>95.91%</td>
</tr>
<tr>
<td>JPEG (original)</td>
<td>95.33%</td>
<td></td>
</tr>
</tbody>
</table>

6. Conclusions

A blind JPEG steganalysis method is proposed here based on the correlation of inter- and intra-wavelet sub-bands in the wavelet domain and feature extraction from the co-occurrence matrix. First, after two-order wavelet decomposition, the joint probability density of each sub-band’s difference coefficients with adjoining coefficients in the horizontal, vertical, and diagonal directions is calculated, and the entropy and energy are extracted from the joint probability density matrix as features. Then, the image is decomposed into three subbands, and the PDF is extracted from each subband’s wavelet coefficient. Finally, the three features are combined to detect the image. Experimental results from typical JPEG image stego algorithms such as F5, Jsteg, Outguess, and Jphide show that the proposed algorithm has much better detection accuracy than typical existing steganalysis algorithms and offers good detection accuracy as well for double-compressed images. In addition, a number of digital analyses of the test results have been performed, and the impacts of different feature combinations on the test results have been discussed. The conclusion can be drawn that compatibility and repulsion exist between the features and that combining all features may not provide the most effective detection. The features chosen in this research, entropy, energy, and the combination of PDF moments, had better detection performance than the other sets of features studied. Future research will explore further the optimal selection of features and the use of a feature extraction algorithm with more sensitivity to other steganographic methods such as MB (Model-Based) and PQ (Perturbed Quantization) steganography.

Acknowledgments

This work has been edited by International Science Editing Company which provided a language editing service. And this work was supported by the National Natural Science Foundation of China (Grant Nos. 60970141, 60902102 and 61170032), China Postdoctoral Science Foundation (Grant No. 201104911838), and by the Doctoral Dissertation Innovation Fund of the Information Science and Technology Institute (Grant No. BSLWCX200804).

References


