

Classification of LSB and MLSB Stego Images via Shift and Pixel Pairs Irrelevance

ZHANG Qian, ZHU Yuefei, GU Chunxiang, LIU Nan

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

qian_zhang2004@126.com

Abstract— For distinguish the LSB (Least significant bit) replacement stego image from MLSB (Multiple least significant bits) stego image, which are two typical kinds of steganographical methods of image spatial domain and have been applied widely, a classification algorithm based on the shift of pixel value and irrelevance of pixel pairs is proposed. In this algorithm, a shift operator is adopted for each pixel value of test image, and then the irrelevance of a kind of special adjoint pixels is extracted as the feature, at last the BP (Back-Propagation) neural network is designed to classify LSB replacement and MLSB replacement images. Results of experiments show that the proposed method can distinguish the MLSB replacement stego images from LSB replacement stego images reliably.

Keywords— Image classification, LSB replacement steganography, MLSB steganography, Shift, Irrelevance

I. INTRODUCTION

Steganography is one of the important research subjects in information security field. As a new art of covert communication, the main purpose of steganography is to convey messages secretly by concealing the existence of messages under digital media files, such as images, audio, or video files.[1] Compared with the Cryptography, modern Steganography not only encrypts messages but also masks the very presence of the communication. Usually, the image steganography includes steganography based on BMP format image and steganography based on JPEG format image. Among the existing steganography methods of BMP images, LSB[2] and MLSB[3][4] are two typical approaches and have been used widely. These two methods embed the message into the multiple significant bit planes of cover image pixels. Unlike LSB, MLSB steganography method not only modify the least significant bit plane of pixel, but also replace other bit planes of pixel, so can defeat many existing steganalysis methods for LSB steganography. Many experts made efforts on the LSB and MLSB steganalysis research these years.

In the aspect of LSB steganalysis, Fridrich et. al.[5] developed a steganalysis method for detection of LSB embedding in 24-bit color images (the Raw Quick Pairs, RQP method). The RQP method is based on analyzing close pairs of colors created by LSB embedding. It works reasonably well as long as the number of unique colors in the

cover image is less than 30% of the number of pixels. The RQP method can only provide a rough estimate of the size of the secret message, and the results become progressively unreliable once the number of unique colors exceeds about 50 percent of the number of pixels. Westfeld [6][7] performed the blind steganalysis based on statistical analysis of PoVs (pairs of values). This method, so-called χ^2 -statistical test, gave a successful result to a sequential LSB embedding steganography. Provos [8] pointed that this idea can also detect random embedding if applied to smaller part of images. Fridrich et al.[9] also proposed a powerful RS (regular and singular groups method). This method makes statistics of the alterations of regular groups and singular groups in the image to estimate the embedding length accurately. It is suitable to color or gray-scale images when the messages are embedded randomly. Dumitrescu et al. [10] proposed SPA, a method to detect LSB steganography via sample pair analysis, which is called Couples Steganalysis in [11]. When the embedding ratio is more than 3%, this method can estimate it with relatively high precision. Ker[11] indicated that in some sense, RS and SPA are the two most reliable detectors of thinly-spread LSB steganography until now. He estimated the reliabilities of Pairs [12], RS and SPA methods through large quantities of experiments, and proposed some good improvements. In addition, a LSB steganalysis method based on weighed-stegoimage is presented in [13].

In the aspect of MLSB steganalysis, some steganalysis methods have been proposed to estimate the embedding rates in higher bit planes. Ref.[14] and [15] have proposed I2Couples, 2Couples and QEM algorithms to estimate the embedding rate of two least significant bits (2LSB) steganography. For the multiple least significant bits (MLSB) replacement steganography, Ref.[16][17][18] have given some methods to estimate the embedding rate of two different embedding paradigms which embed the message into MLSB planes with equivalent rates. In addition, Ref.[19] have applied the results in [16] to estimate the embedding rate of MLSB steganography in each bit plane of images.

From the existing steganalysis methods mentioned above, we can found that although there are some efficient steganalysis for LSB and MLSB stego image, however, these methods can not distinguish the LSB stego images from MLSB stego images. To classify these two kinds of stego

images, in this paper, we give a new classification method based on sensitive features and BP natural network classifier, where the sensitive features are the irrelevances extracted from the adjoin pixel pairs of test images. Experimental results show that the proposed method can distinguish the MLSB stego images from LSB stego images reliably, and the classification accuracy exceeds 93%.

II. ALGORITHM OF CLASSIFICATION

The main steps of proposed algorithm can be described as follows:

- (1) Pixel value shifting. For the training images and test images, right shift all pixel values of image 1 bit, and obtain the shifted pixel values.
- (2) Features extraction from shifted pixel values. From the training images and test images, extract the irrelevance of adjoin pixel pairs of images as features.
- (3) Features preprocessing of classification. For example, eliminate ill-conditioning features, and normalize the extracted features.
- (4) Selection and design of classifier. According to the feature vectors obtained from (3), select a suitable classifier and design the parameters of classifier.
- (5) Training of classifier. Use the features of image whose class has been known to train the classifier, and get a trained classifier with good performance.
- (6) Image examining. Use the classifier from (5) to detect test images, and obtain some outputs of classifier.
- (7) Classification. According to a threshold value, classify the outputs of (6) to the type of LSB stego images or MLSB stego images.

III. FEATURES EXTRACTION

A key issue in the classification of stego images is extracting some effective features. The features should be sensitive to the hidden message. In this paper, we first right shift each pixel value 1 bit, and then from the shifted pixels, calculate the irrelevance of adjoin pixel pairs as feature.

Image marginal information is one of important characteristics. Differential calculus reflecting the change of signal can boost up image marginal information, of which first-order calculus is to compute gradient energy. Hence, measuring the variation of gradient energy is promising as a feature to blindly determine the existence of hidden messages. According to this fact, Ref. [20] extracted gradient energy of image as a feature. A brief description of gradient energy is given as follows.

With regard to an image, $f(i, j)$ represents pixel at (i, j) , its gradient energy can be defined by formula (1),

$$G[f(i, j)] = \left[\left(\frac{\partial f}{\partial i} \right)^2 + \left(\frac{\partial f}{\partial j} \right)^2 \right]^{1/2} \quad (1)$$

As for digital image, formula (1) can be replaced by mathematic difference,

$$G[f(i, j)] = \left\{ [f(i, j) - f(i-1, j)]^2 + [f(i, j) - f(i, j-1)]^2 \right\}^{1/2} \quad (2)$$

Ref. [20] defined the total gradient energy of an image as

$$F_1 = \sum_{i=1}^M \sum_{j=1}^N G(f(i, j))^2 \quad (3)$$

where M, N denote width and height of an image respectively. The gradient energy of a signal will increase after data embedding. Here, we take LSB steganography for examples to show the ratios of features between stego images and corresponding original images. Figure 1(a) shows the gradient energies ratios of 1038 stego images to original images.

This method based on gradient energy features just detects image through checking the change of spatial smoothness between original image and stego image, essentially. To improve the detection accuracy of algorithm, we extract the irrelevances of adjoin pixel pairs as features, and classify images through measuring the changes of these irrelevances.

Assuming that the pixel value of an image is represented by the succession of samples $s_1, s_2, \dots, s_N$ (the index represents the location of a sample in the image), a sample pair means a two-tuple (s_i, s_j) , $1 \leq i, j \leq N$. Let P be a set of sample pairs drawn from an image, then P can be seen as a multiset of two-tuples (u, v) , where u and v are the values of two adjacent samples, $0 \leq u \leq 2^b - 1$, $0 \leq v \leq 2^b - 1$, and b is the number of bits to represent each sample value. For example, for gray image, the scope of pixel value is $[0, 255]$, so b is 8. There are two kinds submultisets of P , X_{2m+1} and Y_{2m+1} . Both X_{2m+1} and Y_{2m+1} contain pairs (u, v) that differ by $2m+1$, (i.e. $|u - v| = 2m+1$). Those pairs in which the even component is larger belong to X_{2m+1} , whereas those pairs in which the odd component is larger belong to Y_{2m+1} . For natural images (original or original image), the probabilities of a sample pair in X_{2m+1} and Y_{2m+1} are almost equal, that's to say:

$$E\{|X_{2m+1}|\} = E\{|Y_{2m+1}|\} \quad (4)$$

where $|\bullet|$ represents the cardinality of the set $\bullet$, $E\{\bullet\}$ means the mathematical expectation of $\bullet$. After embedding message, this equation will be broken. Based on this phenomenon Ref. [10] proposed a powerful steganalysis method for LSB steganography.

In our paper, we let the irrelevance between X_{2m+1} and Y_{2m+1} as δ at $m=0$, and define δ as follows,

$$\delta = \frac{||Y_1| - |X_1||}{|Y_1| + |X_1|} \quad (5)$$

Obviously, the irrelevance will be changed remarkably after embedding some message into natural image. So, we can extract the irrelevance as feature of spatial domain. The process of extraction is as following,

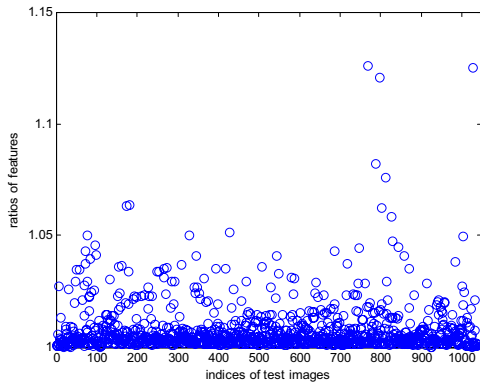
Step 1: For each testing image, count the number of pairs of adjoin pixel that differences of pixels values are 1. Among these two-tuple pairs, if the bigger value is even, then the cardinality of the multiset X adds 1, otherwise,

the cardinality of the multiset Y adds 1.

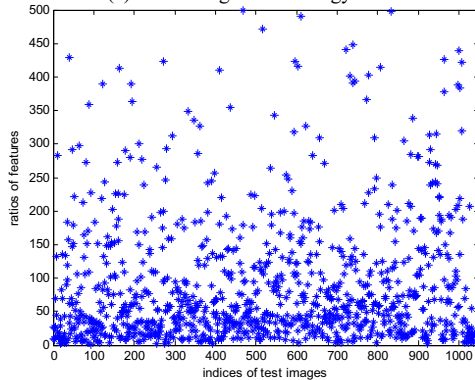
Step 2: According to the formula (5), calculate the irrelevance δ for each test image, thus we can obtain the spatial features of all testing images, and these features can be set as a feature vector F_{SPA} .

For LSB steganography, Figure 1(a) shows the gradient energy ratios of 1038 stego images to original images, and Figure 1(b) shows the irrelevances of 1038 stego images to original images. Note that the ordinate of Figure 1(b) is different to 1(a). For the shifted pixels of images, Figure 2(a) and 2(b) shows the ratios of irrelevance features between stego and original images respectively, where the steganography method is LSB and MLSB (M=2), the embedding ratio is 100%.

From the Figure 1, we can conclude that ratios of irrelevances are remarkably bigger than those of gradient energies, and from the Figure 2, we can found that ratios of irrelevances are distinct different for LSB steganography and MLSB steganography if the pixels values of image are right shifted 1 bit. This indicates that the irrelevances extracted from shifted pixels have different sensitivity for LSB steganography and MLSB steganography, so they may be a kind of good feature for classifying the LSB stego images and MLSB stego images.

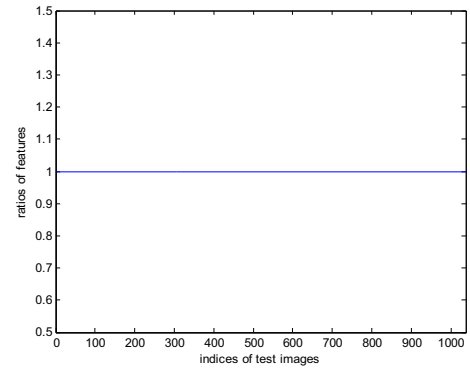


1(a) Ratios of gradient energy features

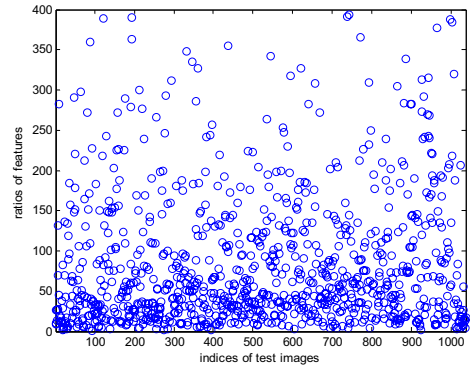


1(b) Ratios of irrelevance features

Figure 1. Ratios of spatial features between stego and original images, where the steganography method is LSB replacement, and the embedding ratio is 100%.



2(a) Ratios of irrelevance features for shifted pixels (LSB)



2(b) Ratios of irrelevance features for shifted pixels (MLSB)

Figure 2. Ratios of irrelevance features between stego and original shifted images, where the steganography method is LSB and MLSB (M=2), the embedding ratio is 100%.

IV. DESIGN OF CLASSIFIER

A three-layer Back-propagation (BP) Neural Network is adopted in the paper to classify the extracted features. The input layer contains only one neurons corresponding to 1-D feature of each test image, the hidden layer is composed of 2 neurons, and the output layer contains one neuron. The logistic sigmoid function is given as $f(x) = 1/(1 + \exp(-x))$. The output value is within the range $[0, 1]$. If the value is close to 1, the classifier indicates the existence of hidden message; if the value is near 0, the classifier reveals no embedding in a considered image. The most iterative times is set 2000 epochs, and the output error of the validation set is less than 0.01. The model of adopted BP neural network classifier is shown as Figure 3.

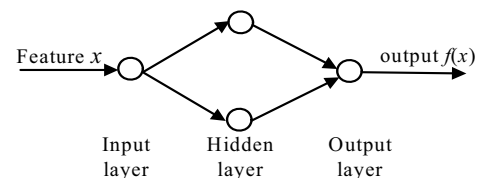
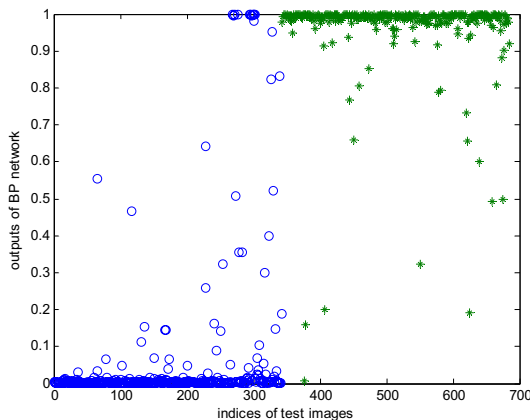


Figure 3. Model of BP neural network classifier with three layers

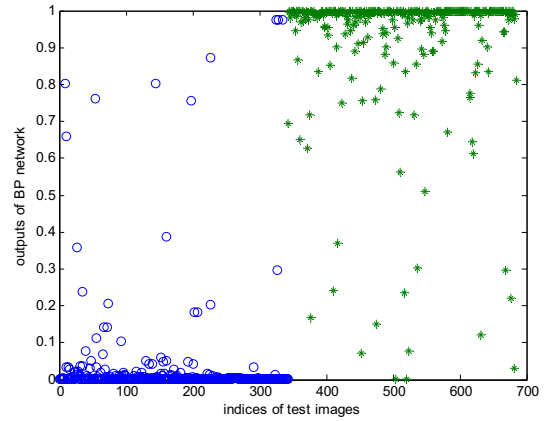
V. EXPERIMENTAL RESULTS

All the experiments are implemented in Matlab 7.0. We downloaded 1038 natural images from NRCS^[21]. Those images are with same sizes of 512×512 . They span an extensive range of landscape, architecture, animal, and people. For the set of 1000 natural images, we use LSB and MLSB (M=2 and 3) steganography to produce three sets of stego images. Among all sets of stego images, messages are embedded with the ratio of 100%. In all the results presented below, for all stego images sets and corresponding cover images sets, 692 images per set are chosen to train BP neural network, and the remaining 346 images in each set are used to test.

Figure 4 describes the output results of BP neural network classifier, where the figure (a) and (b) are corresponding with the case of the correct classification ratio of LSB stego images and MLSB(M=2), and the correct classification ratio of LSB stego images and MLSB(M=3), the embedding ratios are 100%. In these figures, the circle "o" means the output of LSB stego image, and the snowflake "*" mean the output of MLSB stego image. We can found that almost all the outputs of BP neural network classifier are fasten on the port of value 0 or 1, this is in favor of classification. Table.1 shows the average accuracy of proposed algorithm for LSB stego images and MLSB(M=2) stego images under three embedding ratios (100%, 50% and 25%). From Figure 4 and Table 1, we can see that the proposed method can distinguish the LSB stego images from MLSB stego images reliably, and the average classification correct ratio exceeds 93%.



4(a) Outputs of BP classifier for LSB stego images and MLSB (M=2) stego images



4(b) Outputs of BP classifier for LSB stego images and MLSB (M=3) stego images

Figure 4. Outputs results of classifiers for LSB stego images and MLSB images

TABLE 1. CLASSIFICATION ACCURACY OF PROPOSED METHOD

Embedding ratios	LSB stego images	MLSB stego images
100%	96.4%	95.6%
50%	93.6%	92.1%
25%	91.1%	90.2%
Average accuracy	93.2%	

VI. CONCLUSION

In this paper, the classification problem of LSB stego images and MLSB stego images is discussed, and a classification algorithm based on pixel shift and irrelevance of pixel pairs is presented. Experimental results show that the proposed method can distinguish the MLSB stego images from LSB stego images reliably, and the classification accuracy exceeds 93%. Next step, we will study the method of features selection and optimization to improve the detection accuracy of proposed algorithm.

REFERENCES

- [1] Luo X, Yang C, Liu F. Equivalence Analysis among DIH, SPA, and RS steganalysis methods, the 10th IFIP conference of Communication and Multimedia Security, LNCS 4237, 2006, 161-172.
- [2] Petitcolas F A P, Anderson R J, Kuhn M G. Information hiding – a survey. Proceedings of IEEE, 1999, 87(7): 1062-1078.
- [3] B. Nguyen, S. Yoon, and H. Lee. Multi bit plane image steganography. In Proc. International Workshop on Digital Watermarking, Y.Q. Shi and B. Jeon, Eds., 2006, LNCS, vol. 4283, 61-70.
- [4] N. Wu and M. Hwang. Data hiding: current status and key issues. International Journal of Network Security, 2007, 4(1): 1-9.
- [5] Fridrich J, Goljan M. Practical steganalysis of digital images – state of the art. In: Proceedings of SPIE, Security and Watermarking of Multimedia Contents IV, 2002, vol. 4675, 1-13.
- [6] Westfeld A, Pfitzmann A. Attacks on steganographic systems. In: Proceedings of 3rd International Workshop on Information Hiding, Lecture Notes in Computer Science, 2000, vol. 1768, 61-75.
- [7] Lee K, Westfeld A, Lee S. Category attack for LSB embedding of JPEG images. In: Proceedings of International Workshop on Digital Watermarking, Lecture Notes in Computer Science, 2006, vol. 4283,

- [8] Provos N. Defending against statistical steganalysis. In: Proceedings of 10th USENIX Security Symposium, 2001, 323-335.
- [9] Fridrich J, Goljan M, Du R. Reliable detection of LSB steganography in color and grayscale images. In: Proceedings of the Workshop on Multimedia and Security, 2001, 27-30.
- [10] Dumitrescu S, Wu X L, Wang Z. Detection of LSB steganography via sample pair analysis. IEEE Transactions on Signal Processing, 2003, 51(7): 1995-2007.
- [11] Ker A. Improved detection of LSB steganography in grayscale images. In: Proceedings of 6th International Workshop on Information Hiding, Lecture Notes in Computer Science, 2004, vol. 3200, 97-115.
- [12] Fridrich, J., Goljan, M., Soukal, D.. Higher-order statistical steganalysis of palette images. In: Proceedings of SPIE: Security and Watermarking of Multimedia Contents V, 2003, 5020: 178~190.
- [13] Fridrich J, Goljan M. On Estimation of secret message length in LSB steganography in spatial domain. Proceedings of SPIE Security, Steganography and Watermarking of Multimedia Contents VI. 2004, 5306: 23-24.
- [14] Ker A D. Steganalysis of embedding in two least-significant bits. IEEE Transactions on Information Forensics and Security, 2007, 2(1): 46-54.
- [15] Luo X Y, Yang C F, Wang D S, Liu F L. LTSB steganalysis based on quartic equation. LNCS Transactions on Data Hiding and Multimedia Security, 2007, II: 68-90.
- [16] Yang C F, Liu F L, Luo X Y, Liu B. Steganalysis frameworks of embedding in multiple least-significant bits. IEEE Transactions on Information Forensics and Security, 2008, 3(4): 662-672.
- [17] Yu X Y, Noboru B. A fast and effective method to detect multiple least significant bits steganography. In: Proceedings of 2008 ACM symposium on Applied computing, 2008, 1443-1447.
- [18] Luo X Y, Wang Q X, Yang C F, Liu F L. Detection of LTSB steganography based on quartic equation. In: Proceedings of IEEE the 8th International Conference on Advanced Communication Technology, 2006, 1199-1204.
- [19] Yang C F, Luo X Y, Liu F L. Embedding ratio estimating for each bit plane of image. In: Proceedings of 11th International Workshop on Information Hiding, Lecture Notes in Computer Science, 2009, vol. 5806, 59-72.
- [20] Lie W N, Lin G S. A feature-based classification technique for blind image steganalysis. IEEE Transactions on Multimedia, 2005, 7(6): 1007-1020.
- [21] NRCS Photo Gallery. <http://photogallery.nrcs.usda.gov/>