2009 Fifth International Conference on Information Assurance and Security

Universal Steganalysis Based on Statistical Models Using Reorganization of Block-based DCT Coefficients

Shaohui Liu, Lin Ma, Hongxun Yao, Debin Zhao
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
shaohl@vilab.hit.edu.cn

***Abstract*—The goal of stganography is to hide information into media without disclosing the fact of existing communication. Currently, stganography such as least significant bit (LSB), quantization index modulation (QIM) and spread spectrum (SS), has become increasingly widespread. Steganalysis as a counterpart of stganography is to detect the presence of it. In this paper, we present a new universal steganalysis method based on statistical models of the image's discrete cosine transform (DCT) coefficients. In fact, the block-based DCT by proper reorganization of its coefficients can have similar characteristics to wavelet transforms. The presented universal steganalysis method utilizes these characteristics to build statistical models of the image and its prediction-error image. Features extracted from the re-organization DCT blocks of host images and theirs prediction-error images and features extracted from steg images and theirs prediction-error images are used to train the SVM classifier. In the testing, features from those potential images are inputted the trained-well classifier to determine where the potential images are stego images or not. The experiments have shown that the proposed method outperforms in general prior-arts of steganalysis methods based on wavelet transform domain.**

Keywords-Steganalysis;DCT; statistical model

I. INTRODUCTION

Stganography is the art of hiding a message signal in a host signal, such as audio, video, still images and text document without any imperceptible distortion of the host signal for convey information secretly by concealing their existence. Dissimilar to cryptanalysis, the existence of secret information can hardly be detected through the computer analysis of stego-images without original images. This creates a potential problem when this technology is misused for planning criminal activities. For example, some ones can transmit illegal information for avoiding law enforcement via these information hiding technologies. And the difficulty of detecting will slow computer forensics evidence collection when criminals communicate each other via data hiding techniques. So how to determine whether digital media has hidden information become an emergency problem for social security and stability. This new research area is called as steganalysis, which includes detecting existence of secret communication, estimation of message length, and its extraction. Some definitions and several methods of steganalysis were proposed in the literature. Many of stganography tools are available via internet, such as generic LSB method, Chen et al's generic QIM [1] method, Cox et al's SS [2] method, and Chang et al's Side-match [3] method are used frequently. These tools can hide secret messages into original images without perceptual changes. Hence, it is very difficult to distinguish the difference between the original image and the stego image (with hidden message inside). But some statistical changes may be found by analysis in these stego-images in details, no matter how small. Those statistical changes can be used to identify the occurrence of steganography.

To detect the existence of the hiding messages, many steganalysis methods are proposed. All of these methods can be divided into two kinds. One kind is designed to a certain hiding technique. And the other kind is designed to many steganographic methods, named as universal methods.

A universal steganalysis method using higher order statistics was proposed by Farid[4]. The statistics are based on decomposition of images with quadrature mirror filters. The subbands' higher order statistics are calculated for steganalysis. In[5], a steganalysis method based on the mass center (the first order moment) of histogram characteristic function was proposed. The second and third order moments are used for steganalysis. Compared with [4], the success rate has been improved. But it is still not enough good, because it adopted few features extracted from the image for steganalysis. In [6], Shi et al's proposed a universal steganalysis system. The statistical moments of characteristic functions of a test image, its prediction-error image, and their wavelet subbands are selected as features. The method has achieved a good performance. However, when applied to some certain hiding methods, such as Side-match[3], the algorithm doesn't work efficiently.

In this paper, we propose a new universal steganalysis method. The test and the prediction-error images are decomposed using block-based DCT. We show that proper reorganization of block-based DCT coefficients can have similar characteristics to wavelet transform. The statistics of the coefficients of the reorganized DCT subbands are calculated as the features for steganalysis. Extensive experimental results have shown that this steganalysis method outperforms in general prior-arts[4-6] based on DWT, especially when the length of the hidden message is very little.

The remainder of this paper is organized as follows. In section 2, the reorganization strategy of DCT coefficients is addressed. In section 3, the proposed statistical features are

978-0-7695-3744-3/09 $25.00 © 2009 IEEE
DOI 10.1109/IAS.2009.185

discussed. In section 4, the classifier used in our work is briefly described. Experimental results are presented in Section 5. The last section concludes the paper.

II. REORGANIZATION STRATEGY OF DCT COEFFICIENTS

When an image is decomposed using DCT, the characteristic of its DCT domain is of the following two facts:

1. Most of the signal energy is compacted mostly into DC coefficients, thus resulting in little contributions to coefficients in the higher frequency bands.
2. Due to the compact of DCT, the high-frequency components contribute energy to a small number of AC coefficients.

For block-based DCT, an input image is first partitioned into $n \times n$ blocks, where $n = 2^L$, $L > 0$. Each block is then transformed into the DCT domain and can be taken as an L-scale tree of coefficients with $3 \times L + 1$ subbands decomposition. Then, the same subbands for all DCT blocks are grouped and put onto their corresponding positions. We call this reorganization of DCT coefficients[7] into a single DCT clustering entity. Fig. 1 demonstrates an 8×8 DCT with ten-subband decomposition. The reorganization of 8×8 DCT with ten-subband decomposition is illustrated in Fig. 2. In Fig.2, G0 denotes Group of subband 0 and GI denotes Group of subband I.

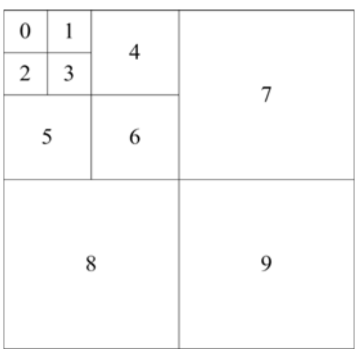


Figure 1. 8×8 DCT block taken as three-scale tree with ten-subband decomposition

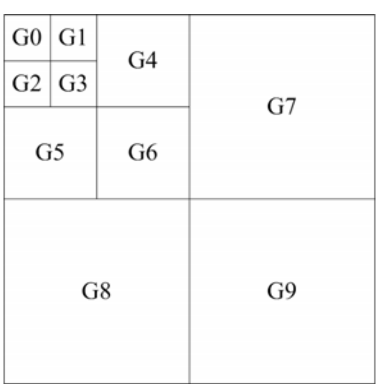


Figure 2. Reorganization of 8×8 DCT blocks into a single DCT clustering entity

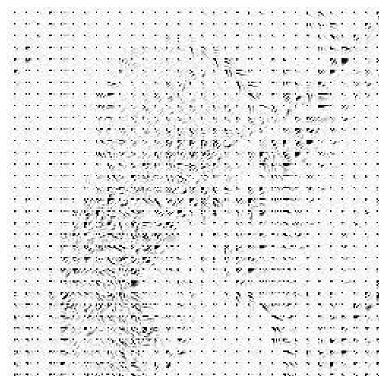
Fig. 3 shows an example of the reorganized DCT coefficients on the Lena image (256×256, 8-bits) with 8×8 DCT. Fig. 3 (a) is 8×8 DCT coefficients, and Fig.3 (b) shows the corresponding reorganized 8×8 DCT coefficients. The gray values (except dc's) in Fig. 3 are obtained by $255 - 5 \times abs(coefficient)$ for better visual presentation.

Analyzing the above images, some facts are presented in the following.

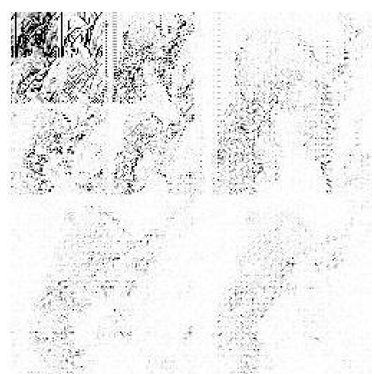
1. Signal energy is compacted mostly into dc coefficients and small number of ac coefficients.

2. Similarity between cross subbands and magnitude decay across subbands can be observed.

Then the above DCT characteristics after reorganization are similar to wavelet transform, which can be further utilized for steganalysis to obtain better performance than Shi et al's method [6].The details will be presented in the next Section.



(a)



(b)

Figure 3. Reorganization of 8×8 DCT coefficients on Lena. (a) DCT coefficients as 8×8 blocks. (b) Reorganization 8×8 DCT coefficients into a single DCT clustering entity.

III. FEATURES FOR STEGANALYSIS

In fact, the basic problem of steganalysis is to decide whether an image has been hidden message or not. Hence, it is just like the problem of classification. Steganalysis can classify images into two classes, one class is host images (without hidden message), and another class is steg-images(with hidden message). So steganalysis is actually a matter of pattern classification in which one key element is to select effective features. Moreover, the features should sensitive to hidden message. Literature has been shown that features extracted from DWT are effective to steganalysis. In this paper, we will extract features from DCT and prove the extracted features are more effective than ones from DWT.

A. Prediction-errorImage

Steganalysis usually utilizes the changes caused by the hiding to achieve its goal. However, the stego images are imperceptible. Especially, some newly developed steganographys are sophisticated and leave very limited space to steganalysis. Moreover, those limited spaces are very possible to be covered by the transmitted noises. Even those changes also can be thought as noise. Hence, changes caused by hiding should be enhanced to advance the design of steganalysis. In this paper, a predicted image whose pixels are predicted by its neighborhood pixels is adopted to achieve this goal. By subtracting the predicted image from the test image, we can obtain the prediction-error image. It is expected that the prediction-error image could remove all the other noises other than the ones introduced by hiding. Thus, it will make the steganalysis more efficient. The prediction algorithm[7] is expressed as:

$$\hat{x} = \begin{cases} \max(a,b) & c \le \min(a,b) \\ \min(a,b) & c \ge \max(a,b) \\ a+b-c & otherwise \end{cases} \qquad (1)$$

Where a,b,c are the context of the pixel x, and $\hat{x}$ is the prediction value of x. The location can be illustrated in Fig.4.

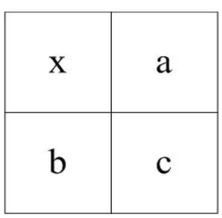


Figure 4. The context of prediction

B. Constructed Statistical Features

The discrete Fourier transform (DFT) of the histogram [8] is chosen as the characteristic function (CF). Then we calculate the statistics of characteristic function which generated from the given test image and its prediction-error image.

First, a test image is decomposed by block-based DCT. After that, the DCT coefficients are reorganized into ten subbands. Compared with wavelet transform, the reorganized DCT coefficients can be viewed as three level wavelet transform. For each level the image is decomposed into 4 subbands, with the whole image, an image includes 13 subbands. As some steganographic methods embed messages into the horizontal or vertical edges of the image to avoid perceptual distortion, we decompose the subbands G7, G8 and G9, to emphasize the statistics during these subbands for further steganalysis. Thus another 12 subbands are generated. Together with previous 13 subbands, total 25 subbands can be used to extract statistical features, as shown in Fig.5. Next step, features are extracted from those 25 subbands.

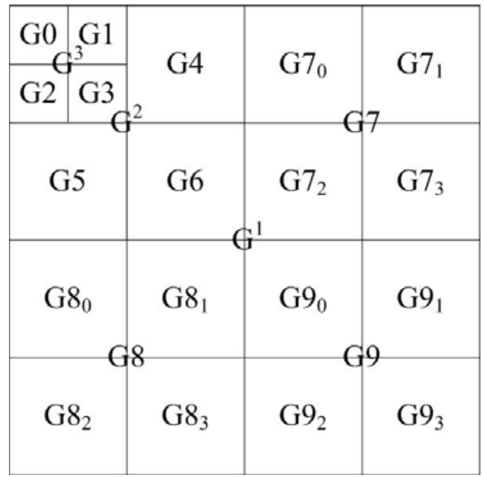


Figure 5. Subbands of the reorganized DCT coefficients

First, the statistical moments in (2) that Shi et al[7] proposed are used as one part of our statistical features.

$$M_n = \frac{\sum_{j=1}^{N/2} f_j^n \left|H(f_j)\right|}{\sum_{j=1}^{N/2} \left|H(f_j)\right|} \quad n = 1,2,3 \quad (2)$$

where $H(f_j)$ is the CF component at frequency f_j and N is the total number of different value levels of coefficients in one subband.

Second, features in Y. Wang et al[10] form the other part of our statistical features. The features are defined in (3). It is clear that the first two features emphasize the information about the high-frequency components and the other one conveys the information about the low- to mid-frequency components. Actually the features defined in (2) are obtained by essentially high-pass filtering the histogram. Experiments indicate that the largest difference between the characteristic function of most original and stego images lies at mid-frequencies, especially for some images whose histograms do not have strong components at high frequencies. This is the reason why those features are chosen.

$$\begin{aligned} g_n &= \sum_{j=0}^{N/2} \left|H(f_j)\right| \sin^n\left(\frac{\pi j}{N}\right) \quad n = 1,2 \\ g_3 &= \sum_{j=0}^{N/4} \left|H(f_j)\right| \sin\left(\frac{4\pi j}{N}\right) \end{aligned} \qquad (3)$$

In a word, all the 25 subbands shown in Fig.5 of the test image according to (1) and (2) can provide a set of $25 \times 6 = 150$ features. However to the prediction-error image, the 12 subbands: $G7_0 \sim G7_3$, $G8_0 \sim G8_3$,

$G9_0 \sim G9_3$ are excluded. Thus, the number of statistical features of remaining 13 subbands of the prediction-error image by using (1) and (2) is $13 \times 6 = 78$ features. Combining the features together, the feature vector formed by total 228 features is utilized for steganalysis.

IV. CLASSIFIER

The design of classifier is another key element in steganalysis. In our work, the support vector machine (SVM) is used as classifier, because of its comparable and efficient classification performance. Due to the constraint of paper length, we do not provide the SVM in detail. Further more information is available in [11]. In our experiments, the tool LIBSVM [12] is utilized as the classifier.

V. EXPERIMENTAL RESULTS

The CorelDraw image database is used for experiments to evaluate the presented steganalysis algorithm. And each image was cropped to the size of 256×256 (before hiding). For each data hiding method, we randomly select 1000 images from the database, and hide the messages into the images according to the corresponding hiding method. In experiments, 1000 images and theirs 1000 steg-images are divided evenly into two parts respectively. 500 host-images and corresponding steg-images are utilized as training samples, and another 500 host-images and corresponding steg-images are used as testing sample. The detection rates are shown in Table 1 and Fig. 6. Fig.6 has shown the different success classifying rate to the different lengths of the message hidden in the images using the generic LSB method.

TABLE I. EXPERIMENTAL SUCCESS CLASSIFYING RATE

Hiding algorithm	**Shi et al's**	**Farid's**	**proposed**
Generic LSB(0.3 bpp)	98.9%	82.4%	99.85%
Generic QIM(0.1 bpp)	99.0%	99.0%	100.00%
Cox et al's SS(α =0.1)	98.1%	77.8%	99.95%
Side-match (0.2 bpp)	88.1%	83.5%	99.85%
4 method combined	89.6%	57.7%	99.81%

Farid's method is downloaded from his website. In Table 1, the classifying rates of LSB, QIM and SS are announced in Shi et al's paper[6]. The other tests listed in Table 1, including the success classifying rates shown in Fig.6, are implemented in Matlab.

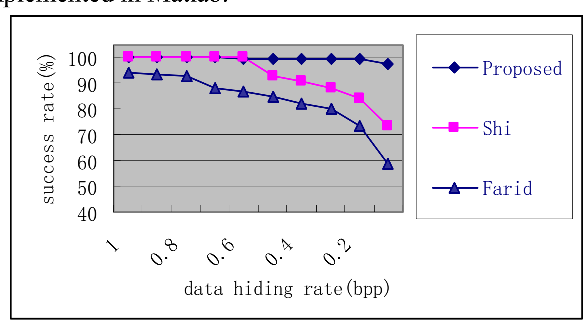


Figure 6. The comparison of our proposed method, Shi et al's and Farid's method to the different lengths of the hidden message using the generic LSB method

VI. DISCUSSIONS AND CONCLUSIONS

In this paper, we proposed a universal steganalysis method which performs very well to detect the fact of hiding. Proposed method is based on the reorganized DCT coefficients of the image, which significantly improve the steganalysis rate comparing with wavelet transform. And the proposed statistical features emphasize not only the high-frequency information, but also the low- and mid-frequency information. Also, experimental results indicates the decompositions of the horizontal-like, vertical-like and the diagonal-like DCT subbands have improved the steganalysis success rate.

ACKNOWLEDGMENT

This work is partly supported by PhD Program Foundation of the Ministry of education of China under Grant 200802131023 and 20060213052, Natural Science Foundation of China under Grant 60803147 and Development Program for Outstanding Young Teachers in Harbin Institute of Technology under Grant HITQNJS.2008.048.

REFERENCES

[1] B. Chen and G. W. Wornell, "Digital watermarking and information embedding using dither modulation," Proceedings of IEEE MMSP, pp. 273-278, 1998.

[2] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia," IEEE Trans On Image Processing, vol. 6, no. 12, pp. 1673-1687, 1997.

[3] C. C. Chang and H. W. Tseng, "A steganographic method for digital images using side-match," Pattern Recognition Letter, vol. 25, no. 12, pp. 1431-1437, 2004.

[4] H. Farid, "Detecting hidden messages using higher-order statistical models," International Conference on Image Processing, Rochester, NY, USA, 2002

[5] J. Harmsen, Steganalysis of Additive Noise Modelable Information Hiding, Rensselaer Polytechnic Institute, NY, April 2003.

[6] Y. Q. Shi, G. Xuan, D. Zou, J. Gao, C. Yang, Z. Zhang, P. Chai, W. Chen, C. Chen, "Steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network," International Conference on Multimedia and Expo, Amsterdam, Netherlands, 2005.

[7] D. Zhao, W. Gao and Y. K. Chan, "Morphological representation of DCT coefficients for image compression," IEEE trans. Circuits Syst. Video Technol., vol. 12, no. 9, September 2002.

[8] A. Leon-Garcia, Probability and Random Processes for Electrical Engineering, Addision-Wesley Publishing Company, 1994

[9] Y. Wang and P. Moulin, "Statistical modeling and steganalysis of DFT-based image steganography," Proc. SPIE: Security, Steganography and Watermarking of Multimedia Contents Ⅷ, vol. 6072, 607202, 2006.

[10] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, 2(2), 12-167, 1998.

[11] C. C. Chang and C. J. Lin, LIBSVM: a library for support vector machines, 2001.