Detection of LSB Matching Steganography in Decompressed Images

Jun Zhang and Dan Zhang

***Abstract*—Studies show that no detectors for LSB matching have yet proven universally reliable and it is hard to predict which types of image are suitable for a specific steganalyzer. For the kind of decompressed images, this paper presents an efficient steganalyzer that exploits the fact that the noise residuals in the DCT domain are rather concentrated on zero and very sensitive to LSB matching. Experimental results show that it is almost perfect at embedding rate 0.5 bpp and that it is the accuracy of 90.9% at 0.1 bpp against the accuracy of 44.6% with the WAM steganalyzer. However, the proposed detector works only as long as the exact JPEG decompressor is known.**

***Index Terms*—Decompressed image, information hiding, LSB matching, steganalysis.**

I. Introduction

STEGANOGRAPHY seeks to provide a covert communication channel between two parties [1]. A common class of steganographic algorithms embeds the secret message in cover works such as images, video, audio or text. The combination of cover work and secret message is referred to as the stego work and a goal of all steganographic algorithms is to ensure undetectability, i.e. that a third party, referred to as the Warden, is unable to distinguish between a cover work and a stego work.

On the other hand, the detection of a stego work is the goal of steganalysis. Almost all steganalysis algorithms rely on the steganographic algorithm introducing statistical differences between cover and stego works. There are two classes of steganalysis algorithms—blind and targeted. Blind steganalysis algorithms are intended to detect a wide range of steganographic algorithms, including previously unseen algorithms. In contrast, targeted steganalysis algorithms are intended for a specific steganographic algorithm. In the paper, we describe a blind steganalysis algorithm for the detection of LSB matching, which hides a secret data bit by randomly incrementing or decrementing the corresponding pixel of the cover image when the secret data bit does not match the LSB of the pixel, otherwise keeping the pixel unchanged.

Perhaps surprisingly, detection of LSB matching has proved considerably more difficult than for LSB replacement. A number of steganalysis algorithms have been proposed. Harmsen and Pearlman [2] noted that, for images, adding noise in the spatial domain corresponds to low-pass filtering of the intensity/colour histogram. Consequently, the histogram of a stego image has less high-frequency power than the corresponding histogram of the cover image. Thus, the center of gravity of $|F(h)|$, which denotes the Fourier transform of the histogram h, will decrease after LSB matching embedding. This property was used as a feature for distinguishing between cover and stego images. While good results were reported on a small test set using colour histograms, subsequent experiments revealed that this technique performs poorly on LSB matching in grayscale images.

To address this issue, Ker [3] proposed two novel ways of applying the histogram characteristic function (HCF), based on i) calibrating the output using a downsampled image and ii) computing the adjacency histogram instead of the usual intensity histogram, which is referred to as the AD-HCF method. Significant improvements in detection of LSB matching in grayscale images were thereby achieved.

Based on the same observation that LSB matching steganography is equivalent to low-pass filtering the intensity histogram, Zhang *et al.* [4] chose to focus their attention on the local extrema of the histogram. The filtering operation will indeed reduce the amplitude of local extrema. As a result, they rely on the sum of the absolute differences between local extrema and their neighbors in the histogram to distinguish between cover and stego images.

Contemporaneously, Holotyak and Fridrich [5] described a blind steganalysis approach for LSB matching based on classifying higher-order statistical features derived from an estimation of the stego signal in the wavelet domain. Goljan *et al.* [6] presented an improved version by using absolute moments of the noise residual, which is called the Wavelet Absolute Moments method (WAM). The proposed approaches are flexible and enable reliable detection of the presence of secret messages embedded using a wide range of steganographic methods that include LSB matching, LSB replacement, stochastic modulation, and others.

Nevertheless, recent studies show that no detectors for LSB matching have yet proven universally reliable and their performances heavily depend on types of images. For example, a recent study [7] of three LSB matching steganalyzers in three

Manuscript received July 15, 2009; revised October 09, 2009. First published October 30, 2009; current version published November 13, 2009. This work was supported by the National Natural Science Foundation of China (60873198), Guangdong Natural Science Foundation (06023961), and Natural Science Foundation of Department of Education of Guangdong Province (05Z013). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alex C. Kot.

J. Zhang is with the School of Information Science, Guangdong University of Business Studies, Guangzhou, China (e-mail: zhangjundan123@yahoo.com.cn).

D. Zhang is with the School of Computer Science, Sichuan University, Chengdu, China (e-mail: xf84096029@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2009.2035379

image sets found wide variations in both absolute and relative performance. Ker [8] also demonstrated that one of the best steganalyzers, WAM, is hugely variable in its detection power on images from different sources. This raises a question regarding steganalysis of LSB matching: which types of image are suitable for a specific steganalyzer?

For the kind of decompressed images, which have previously been compressed by JPEG, Fridrich *et al.* [9] presented a steganalysis method based on JPEG compatibility. However, there are some difficulties with it. For example, it becomes computationally infeasible for mildly-compressed JPEG images. To overcome this difficulty, we propose a novel steganalysis algorithm that exploits the fact that the noise residuals in the DCT domain are rather concentrated on zero and very sensitive to LSB matching.

Section II gives a brief overview of JPEG compression and pays particular attention to the noise residuals in the DCT domain. We construct a novel steganalyzer by using the higher-order absolute moments of noise residuals. Throughout the paper, we point out some limitations of the proposed technique. Section III gives experimental results and compares the proposed method with some state-of-the-art steganalyzers. Finally, Section IV concludes the paper and outlines future research directions.

II. Steganalysis Based on Noise Residuals in the DCT Domain

A. *Noise Residuals in the DCT Domain of Decompressed Images*

We focus on decompressed images and would like to analyse noise residuals in the DCT domain. Firstly, we briefly outline the basic process of JPEG.

During JPEG compression, a image is first split into disjoint 8 × 8 pixel blocks. Each block is transformed using the Discrete Cosine Transformation (DCT). Then the DCT coefficients d_{ij} are divided by quantization step q_{ij} and rounded to integers: $D_{ij} = round(d_{ij}/q_{ij})\ i,j \in \{0,\cdots,7\}$. Finally, an entropy coding is applied to the quantized coefficients and the image is said to be JPEG compressed one.

The decompression works in the opposite order. After reading the quantized DCT blocks from the JPEG file, the coefficients $\widetilde{d}_{ij}$ of each block are multiplied by the quantization step q_{ij}, i.e. $\widetilde{d}_{ij} = D_{ij} \times q_{ij}$, and then the Inverse Discrete Cosine Transformation (IDCT) is applied to $\widetilde{d}_{ij}$. The values are finally rounded to integers and truncated to a finite dynamic range (usually [0, 255]) and the bitmap image is said to be decompressed one.

Then, let us analyse the noise residuals in DCT domain. For a given decompressed image, if we divide it into 8 × 8 pixel blocks and each block is transformed by DCT, then we can get the DCT coefficients d'_{ij} again. However, d'_{ij} are not exactly $\widetilde{d}_{ij}$. That means, $d'_{ij} = \widetilde{d}_{ij} + r_{ij}$. We call r_{ij} the noise residuals in DCT domain. r_{ij} are mainly from two sources of errors. Both of them were introduced during the IDCT calculation. First, the pixel values, typically integers, are rounded from real numbers. Secondly, any number greater than 255 or smaller than 0 for a pixel value, which is normally limited to 8 bits, is truncated to 255 or 0, respectively. Although we don't know what $\widetilde{d}_{ij}$ are in reality, we assume the rounding of d'_{ij} are good approximated version of $\widetilde{d}_{ij}$. As a result, we can get the following approximate equality:

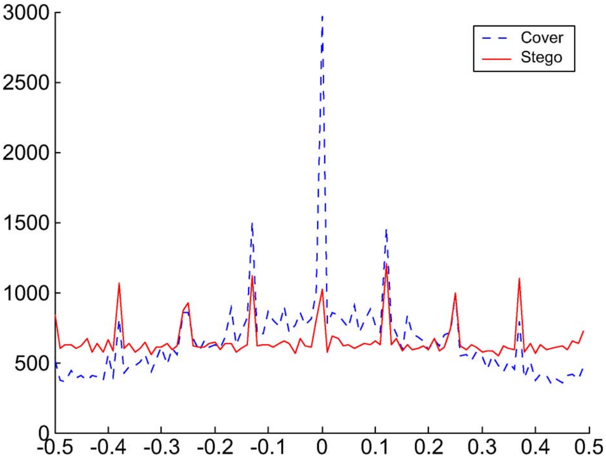

Fig. 1. Histograms of noise residuals of a cover image and the corresponding stego image.

$$r_{ij} \approx d'_{ij} - round\left(d'_{ij}\right) \tag{1}$$

where the function round makes a real number its nearest integer. Formula (1) shows the noise residuals are obtained by removing the approximated version of a cover image. So, we can expect that the features calculated from the noise residuals are more sensitive to embedding modifications and less sensitive to the image content. Fig. 1 shows the histograms of noise residuals of a cover image (previously compressed by JPEG) and the corresponding stego image, in which 0.5 bpp messages are embedded by LSB matching method. We can see the noise residuals are very sensitive to LSB embedding. Fig. 2 shows the standard deviations of noise residuals of 100 cover images and the corresponding stego images, in which the symbols '$*$' and 'o' stand for that of the cover images and the stego images respectively. This figure demonstrates that the standard deviations of noise residuals of cover images are smaller than that of setgo images. So we can distinguish cover and stego images by the statistical features of noise residuals.

Theoretically, the DCT coefficients d'_{ij} are equal to $\widetilde{d}_{ij}$ due to the properties of the DCT. That means d'_{ij} are multiple of quantization levels. So, all d'_{ij} are integers. As a result, all the noise residuals of d'_{ij} should be zero. However, this is not true in reality due to two sources of errors mentioned above. Nonetheless, from Fig. 1 we can find the fact that the noise residuals of decompressed images are rather concentrated on zero. We believe that it is commonly an inherited character of any decompressed images since it is introduced by the rounding and truncating operators during the IDCT. So, any setganographic algorithms in spatial domain including variants of the LSB matching [10] will destroy this character of decompressed images.

B. *Steganalysis Algorithm*

We construct a 10-D feature vector by using the higher-order absolute moments of noise residuals. The algorithm of the proposed detector is as follows.

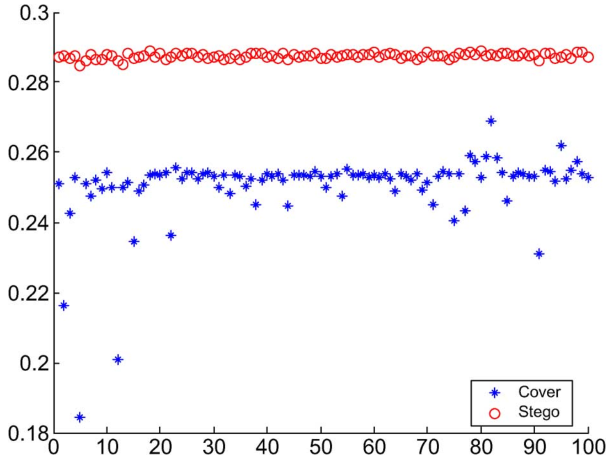

Fig. 2. Standard deviations of noise residuals of 100 cover images and the corresponding stego images.

Step 1) Given a decompressed image I, divide it into 8×8 pixel blocks and each block is transformed by DCT.

Step 2) Compute the noise residuals r_{ij} by formula (1).

Step 3) Work out the central absolute moments of noise residuals,

$$m^p = \frac{1}{|I|} \sum_{(i,j)\in I} |r_{ij} - \overline{r}|^p \quad p = 1, \ldots, 10 \tag{2}$$

where $\overline{r}$ is the mean value of noise residuals.

The 10-D feature vector is formed by the collection of the central absolute moments. Then, one of the simplest classifiers—the Fisher linear discriminator (FLD) is introduced to classify cover and stego images. In the FLD, the feature space is projected on a 1-D space, where various decision rules can be applied for determining the classification thresholds.

According to the algorithm, our scheme has overcome the computational difficulty of Fridrich's method. However, it still has some limitations in practice. First, it works only when the steganalyst knows that the tested images come from JPEG decompressed covers and are not subject to any further processing during and after decompression. Second, it works only as long as the exact JPEG decompressor is known. Since different JPEG decompressors give rise to different outputs, the noise residuals are likely to be subtly different.

III. Experimental Results

We use receiver operating characteristic curves (ROC) to evaluate the performance of steganalyzers. ROC can show how the false positives and true positives vary as the detection threshold is adjusted. Furthermore, we can obtain the accuracy of steganalyzers, which is the area between the ROC curve and the diagonal normalized so that a perfect detection has 100% of accuracy.

The NRCS Photo Gallery (photogallery.nrcs.usda.gov) is used as the image source in our experiment. For convenience, we crop the original color images into 256 $*$ 256 pixels and convert them to grayscales. Here, cropping was preferred over resizing in order to avoid introducing artifacts due to resampling with interpolation. Then, these grayscale images are selected to form the image set D, which include 2375 images.

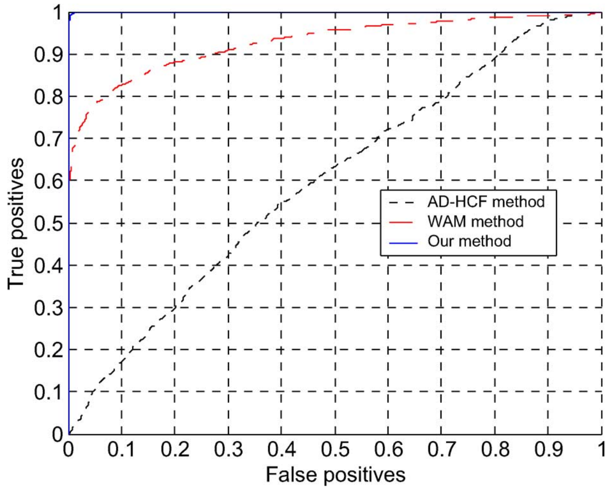

Fig. 3. ROCs of the three steganalyzers at embedding rate 0.5.

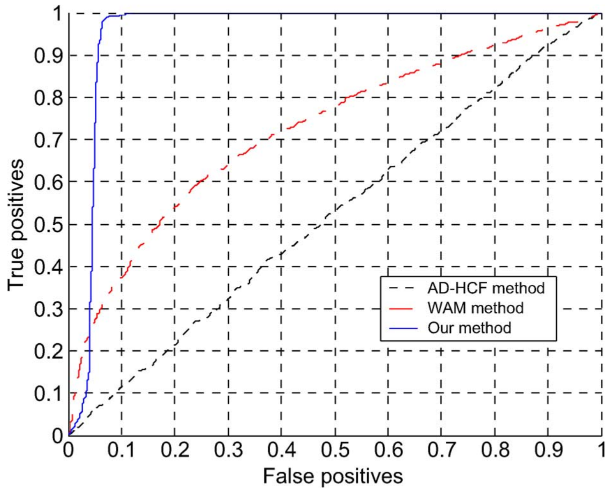

Fig. 4. ROCs of the three steganalyzers at embedding rate 0.1.

We compare our method with the AD-HCF and the WAM on the kind of decompressed images. We firstly compress all images in the set D by JPEG with quality factor 80, and then decompress them to form the image source D'. Finally, we apply LSB matching steganography with embedding rate ρ to all images in the set D' to obtain the set of stego images D''. The training set includes 1000 image pairs (from D' and D'') and the test set includes 1375 image pairs. At embedding rate $\rho = 0.5$, the experimental results are shown in Fig. 3. As we can see, our method is almost perfect, with the accuracy of 99.7%. Meanwhile, the accuracy of WAM and AD-HCF are 85.5% and 19.5% respectively. Fig. 4 shows the experimental results at embedding rate $\rho = 0.1$. The accuracy of our method, WAM and AD-HCF are 90.9%, 44.6%, and 4.2% respectively. It means that our method is still reliable while the AD-HCF method is almost a random guess.

Then, we investigate the accuracy of steganalyzers on the combination of different image sources, which have previously been compressed with quality factors 90, 70, and 50. Appling

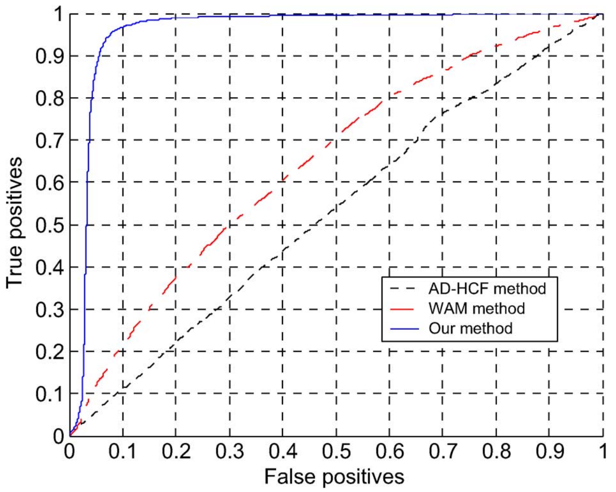

Fig. 5. ROCs of the three steganalyzers on the combination of different image sources at embedding rate 0.1.

LSB matching with embedding rate $\rho = 0.1$, we get the corresponding stego images. Let the training set and the test set include 3000 and 4125 image pairs respectively. We obtain the ROC curves of three steganalyzers shown in Fig. 5. The accuracy of our method, WAM and AD-HCF are 91.6%, 29.1%, and 5.8% respectively. Compared with Fig. 4, the performance of our method is not affected by these different image sources, but the WAM steganalysis is variable in its detection power. It means that our detector is effective for any decompressed images.

IV. Conclusion

It is well known that the performance of current state-of-the-art steganalyzers for detection of LSB matching is highly sensitive to the datasets from different sources and it is hard to predict which types of image are suitable for a specific steganalyzer. This paper proposes a good solution on the kind of decompressed images. The proposed method makes use of the fact that the noise residuals of DCT coefficients of decompressed images are rather concentrated on zero and very sensitive to LSB matching. Therefore, a 10-D feature vector is constructed by using the higher-order absolute moments of noise residuals and the FLD is introduced to classify cover and stego images. The experimental results show that the scheme is almost perfect at embedding rate 0.5 bpp and that it is the accuracy of 90.9% at 0.1 bpp superior to the AD-HCF and the WAM methods. Furthermore, the accuracy of the proposed method is not affected by different image sources. It means that it is reliable for any decompressed images.

The proposed method has some limitations that we would like to address in the future. First, it is only suitable for the kind of decompressed grayscale images that are not processed by other image operators during and after decompression. The future work is to design an identifier that can detect the bitmap compression history. When the bitmap is identified as the decompressed image, the method begins to work. Another limitation of the technique is that it works only as long as the exact JPEG decompressor is known. Therefore, we will test how much detection accuracy is lost under different decompression schemes in the future work.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments on this paper and thank Dr. G. Doerr for providing the source code of the WAM steganalyzer.

References

[1] I. J. Cox, T. Kalker, G. Pakura, and M. Scheel, "Information transmission and *steganography*," *Lecture Notes in Computer Science*, vol. 3710, pp. 15–29, 2005.

[2] J. Harmsen and W. Pearlman, "*Steganalysis* of additive noise modelable information hiding," in *Security and Watermarking of Multimedia Contents V, Ser. Proc. SPIE 5020*, 2003, pp. 131–142.

[3] A. D. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Process. Lett.*, vol. 12, no. 6, pp. 441–444, 2005.

[4] J. Zhang, I. J. Cox, and G. Doerr, "Steganalysis for LSB matching in images with high-frequency noise," in *Proc. IEEE Workshop on Multimedia Signal Processing*, 2007, pp. 385–388.

[5] T. Holotyak, J. Fridrich, and S. Voloshynovskiy, "Blind statistical steganalysis of additive steganography using wavelet higher order statistics," in *9th IFIP TC-6 TC-11 Conf. Communications and Multimedia Security*, 2005, vol. 3677, Lecture Notes in Computer Science, pp. 273–274.

[6] M. Goljan, J. Fridrich, and T. Holotyak, "New blind steganalysis and its implications," in *Proc. SPIE 6072*, 2006, pp. 1–13.

[7] G. Cancelli, G. Doerr, I. Cox, and M. Barni, "A omparative study of $+/-1$ steganalyzers," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, 2008, pp. 791–796.

[8] A. D. Ker and I. Lubenko, "Feature reduction and payload location with WAM steganalysis," in *Proc. SPIE 7254*, 2009, pp. 0A01–0A13.

[9] J. Fridrich, M. Goljan, and R. Du, "Steganalysis based on JPEG compatability," in *Multimedia Systems and Applications IV, Ser. Proc. SPIE 4518*, 2002, pp. 275–280.

[10] J. Mielikainen, "LSB matching revisited," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 285–287, 2006.