Steganalysis for LSB Matching in Images with High-frequency Noise

Jun Zhang
School of Information Science
Guangdong University of Business Studies
P. R. China
Email: j.zhang@adastral.ucl.ac.uk

Ingemar J. Cox and Gwenaël Doërr
Dept. of Electrical & Electronic Engineering
University College London
Adastral Park Postgraduate Campus
United Kingdom
Email: ingemar@ieee.org, g.doerr@adastral.ucl.ac.uk

***Abstract*—Considerable progress has been made in the detection of steganographic algorithms based on replacement of the least significant bit (LSB) plane. However, if LSB matching, also known as ± 1 embedding, is used, the detection rates are considerably reduced. In particular, since LSB embedding is modeled as an additive noise process, detection is especially poor for images that exhibit high-frequency noise - the high-frequency noise is often incorrectly thought to be indicative of a hidden message. To overcome this, we propose a targeted steganalysis algorithm that exploits the fact that *after* LSB matching, the *local maxima* of an images graylevel or color histogram *decrease* and the *local minima increase*. Consequently, the sum of the absolute differences between local extrema and their neighbors in the intensity histogram of stego images will be smaller than for cover images. Experimental results on two datasets, each of 2000 images, demonstrate that this method has superior results compared with other recently proposed algorithms when the images contain high-frequency noise, e.g. never-compressed imagery such as high-resolution scans of photographs and video. However, the method is inferior to the prior art when applied to decompressed imagery with little or no high-frequency noise.**

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. 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. Typically, they are based on machine learning techniques. In contrast, targeted steganalysis algorithms are intended for a specific steganographic algorithm, known as the *target*. In the paper, we describe a targeted algorithm for the detection of LSB matching or ± 1 embedding.

Perhaps surprisingly, detection of LSB matching has proved considerably more difficult than for LSB replacement. A number of steganalysis algorithms have been proposed, all of which model LSB matching as the addition of noise. 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 propoerty 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 [3].

To address this issue, Ker [3] proposed two novel ways of applying the histogram characteristic function (HCF)[1], based on (i) calibrating the output using a downsampled image, and (ii) computing the adjacency histogram instead of the usual intensity histogram. Significant improvements in detection of LSB matching in grayscale images were thereby achieved.

Contemporaneously, the authors of [4] proposed a method for steganalysis of LSB matching in the spatial domain. The method used a high-pass FIR filter and then recovered an approximate message length using a maximum likelihood estimator. However, they observe that this approach is not effective for never-compressed images derived from a scanner.

Subsequently, Holotyak and Fridrich [5] described a blind steganalysis approach 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 of [5] by using absolute moments of the noise residual. 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, the steganalyzers mentioned above have poor detection performance for LSB matching in grayscale images with high levels of high-frequency noise, such as high-resolution scans of photographs. This is due to the fact that

[1] Essentially the FFT of the intensity/colour histogram.

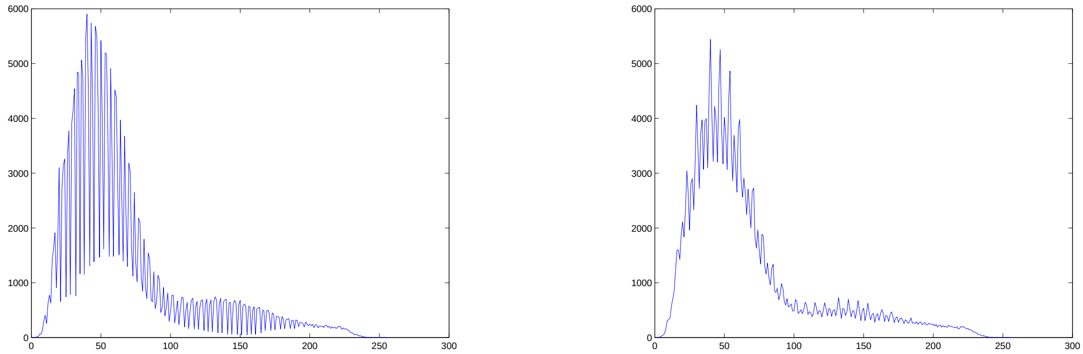


Fig. 1. Histograms of the cover image and stego image using LSB matching.

the image noise masks the additive stego signal. It appears to be very difficult for steganalyzers based on an additive noise model to accurately distinguish between the stego signal and naturally occurring noise in images.

To address this issue, Section II examines the effect of LSB matching on the intensity histogram of graylevel images. We show that the local maxima of the histogram of images will decrease and the local minima will increase after LSB matching. This property can be used to define a feature that can be used to detect LSB matching. This feature is the sum of the absolute differences between each local extremum and its neighbors in the intensity histogram of stego images.

Section III then compares the histogram extrema method with the recent algorithm of Ker [3] and Goljan *et al.* [6]. We refer to the latter algorithm as GFH. Experimental results are reported on two datasets, each of 2000 images, derived from the Corel Image Database. Both datasets contain never-compressed images that possess high frequency noise. The experimental results demonstrate that the histogram extrema method has substantially better performance. However, if the datasets are JPEG compressed with a quality factor of 80, the high frequency noise is removed and the histogram extrema method performs worse.

Section IV discusses directions for future work.

II. Analysis for LSB matching

We assume that the cover and stego images are grayscale images with pixels values in the range $0 \ldots 255$. LSB steganography modifies the least significant bits of the pixel values so that they match the corresponding bits of the message to be hidden.

There are two common methods of LSB steganography. The earliest, and simplest method, simply replaces the LSB bitplane of the cover image with the corresponding bits of the message. This can be done for all pixels in the image or only for a pseudo-randomly chosen portion, when the embedding rate, ρ, is less than one, i.e. the length of the hidden message is less than the number of pixels in the image. However, a number of papers have reported very successful steganalysis of LSB replacement [7]–[9]. This success is credited to the fact that LSB replacement is inherently asymmetric, i.e. an even valued pixel will either retain its value or be incremented by one. However, it will never be decremented. The converse is true for odd-valued pixels. This asymmetry is exploited for steganalysis purposes.

The second method of LSB steganography is known as LSB matching. Rather than simple replace the LSB with the desired message bit, the corresponding pixel value is randomly incremented or decremented, thereby removing the asymmetry of odd and even pixels. Specifically, LSB matching can be described by:

$$p_s = \begin{cases} p_c + 1, & \text{if } b \neq \mathrm{LSB}(p_c) \text{ and } \big(\kappa > 0 \text{ or } p_c = 0\big) \\ p_c - 1, & \text{if } b \neq \mathrm{LSB}(p_c) \text{ and } \big(\kappa < 0 \text{ or } p_c = 255\big) \\ p_c, & \text{if } b = \mathrm{LSB}(p_c) \end{cases} \tag{1}$$

where p_s (resp. p_c) denotes a pixel value in the stego image (resp. cover image), b is the message bit to be hidden, and κ is an i.i.d. random variable with uniform distribution on $\{-1, +1\}$[2]. Detection of LSB matching is known to be much more difficult than detecting LSB replacement.

A. Effects of LSB matching steganography on histogram

Let $p_c(i,j)$ denote the pixel value at location (i,j) in the cover image. The intensity histogram is then defined as:

$$h_c(n) = |\{(i,j)|p_c(i,j) = n\}| \tag{2}$$

where n is a grayscale level in the range $0 \ldots 255$. In other words, $h_c(n)$ indicates the number of pixels in the cover image with grayscale value n.

Let us now consider the effect of LSB matching, with an embedding rate ρ, on the cover image histogram. First, there is a 50% chance that the pixel values at selected locations will already have the desired LSB value. Hence, a proportion $(1 - \rho/2)$ of the pixels will not be modified. The remaining pixels are incremented or decremented with equal probability. Assuming that the embedding locations are uniformly distributed and independent of the pixel values, the histogram of the stego-image is given by:

$$h_s(n) = \left(1 - \frac{\rho}{2}\right) h_c(n) + \frac{\rho}{4}\big(h_c(n-1) + h_c(n+1)\big) \tag{3}$$

[2]Note that this strategy may affect bit-planes other than the LSB plane. For example, if the secret bit is a "0", and the original 8-bit pixel value is 01111111, then incrementing this value results in 10000000.

In other words, LSB matching reduces to low pass filtering the intensity histogram with the kernel $[\rho/4, 1-\rho/2, \rho/4]$.

The histograms shown in Figure 1 clearly illustrate this phenomenon. On the left is the histogram of a cover image and on the right the histogram of the corresponding stego image after LSB matching with an embedding rate of $\rho = 1$. It is evident that the histogram of the stego-image is smoother than that of the cover image. This low pass filtering attenuates the energy in the high frequencies, and in particular the amplitude of local extrema.

A local extremum, n^*, in a histogram, $h()$, is defined by:

$$(h(n^*) - h(n^*-1))\,(h(n^*) - h(n^*+1)) > 0 \qquad (4)$$

According to Equation 3, for any local maximum, n^*, we have

$$\begin{aligned} h_s(n^*) &= \left(1-\frac{\rho}{2}\right)h_c(n^*) + \frac{\rho}{4}\big(h_c(n^*-1)+h_c(n^*+1)\big) \\ &= h_c(n^*) - \frac{\rho}{4}\Big[\big(h_c(n^*)-h_c(n^*-1)\big) \\ &\qquad + \big(h_c(n^*)-h_c(n^*+1)\big)\Big] \\ &< h_c(n^*) \end{aligned} \qquad (5)$$

Similarly, for any local minimum point, n^*, we have $h_s(n^*) > h_c(n^*)$. Thus, after LSB matching, the local maxima of an image histogram decrease and the local minima increase.

The attenuation of local extrema by LSB matching motivated us to consider the sum of absolute differences between each local extremum and its neighbors in the histogram. These sums are denoted D_c and D_s for the cover and stego images respectively. That is,

$$D_c = \sum_{n^*} |2.h_c(n^*) - h_c(n^*-1) - h_c(n^*+1)| \qquad (6)$$

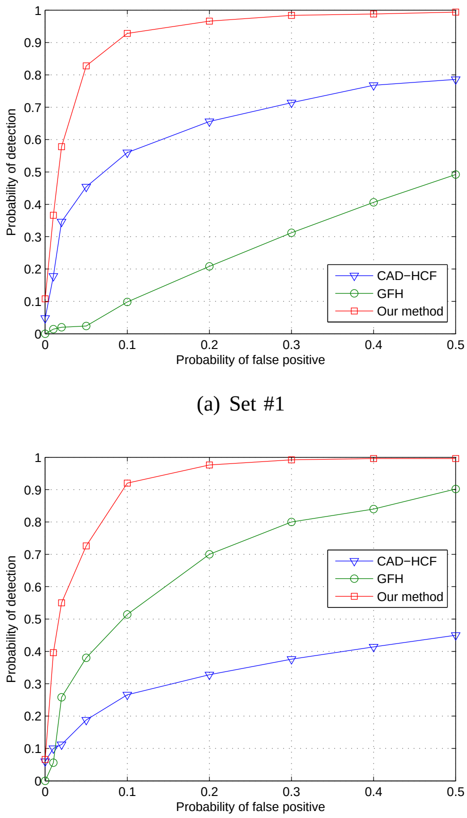
$$D_s = \sum_{n^*} |2.h_s(n^*) - h_s(n^*-1) - h_s(n^*+1)| \qquad (7)$$

It is expected that $D_c > D_s$ for any image after LSB matching steganography and experimental results support this claim. In fact, for extrema in a cover histogram and corresponding extrema in the stego histogram, it can be shown that the corresponding value of the maxima in the stego histogtram is less than in the cover histogram. Similarly, the local minima in the stego histogram increase in value compared to their corresponding minima in the cover histogram.

In the next Section we compare this discriminant to the previous work of Ker [3] and Goljan *et al.* [6].

III. Comparison with Ker's and Goljan *et al*'s Algorithms

All experimental results are reported on two image tests sets derived from the Corel Image Database. Each set consists of 2000 never-compressed images. Set #1 includes 2000 color images of artwork. The original images are 24-bit, with dimensions 512×768 pixels. For convenience, we crop the original color images to 512×512 and covert them to 8-bit grayscale. However, we do not resample the original images. Set #2 also includes 2000 color images of various topics including natural landscapes, people, animals, and cars. Once again, we crop the images to 512×512 and convert them to 8-bit grayscale. Since the images have not been compressed, they typically exhibit high frequency noise. Each image was embedded with a randomly generated message. The message embedding rate $\rho = 0.5$.



(a) Set #1

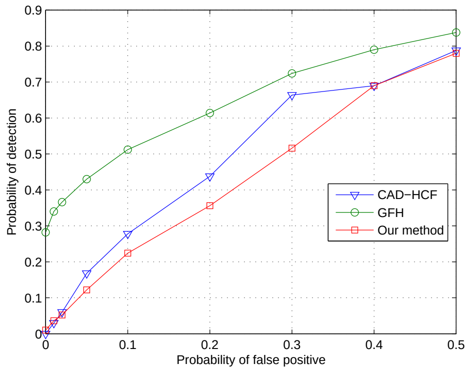
(b) Set #2

Fig. 2. ROC curves comparing our method to that of Ker and GFH, with an embedding rate of $\rho = 0.5$. Both datasets have never been compressed.

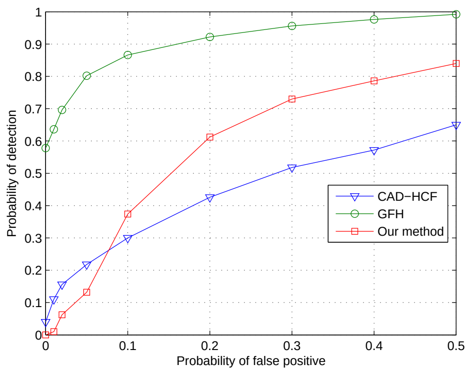
We compare our method with (i) the calibrated adjacency, HCF-COM, version of Ker's method, which is a tageted steganalysis method, and (ii) the blind steganalysis method of GFH. Due to space limitations, the reader is directed to [3] for further details.

Figure 2 demonstrates a significant improvement in performance over that of Ker [3] and GFH [6]. We note, that the performance of both algorithms varies across the two datasets. However, the variation in performance of our histogram extrema method is much less. For example, at a false positive rate of 50% the detection rates of Ker's method are 79% in Set #1 and 45% for Set #2, and the detection rates for GFH method are 50% and 90%, respectively. In comparison, our method has detection rates of almost 100%.

Figure 3 compares the performance of the two algorithms when the two datasets have been JPEG compressed with a quality factor of 80, prior to LSB matching. In this case we see that the histogram extrema method is inferior to Ker's and



(a) Set #1



(b) Set #2

Fig. 3. ROC curves comparing our method to that of Ker and GFH, with an embedding rate of $\rho = 0.5$. Here both datasets have been JPEG compressed with quality factor 80.

GFH algorithms. This is expected as the histogram extrema algorithm is not designed for the case where high-frequency noise is absent. Interestingly, while the GFH method performs better on both datasets after compression, the performance of Kers method actually decreases for dataset #1. We currently do not understand the reason for this.

IV. Conclusion

Detection of LSB matching in cover images that exhibit high-frequency noise, such as never-compressed grayscale images and scans of photographs, is difficult for current steganalysis algorithms. This is because LSB matching is often modeled as an additive noise process and the high-frequency noise confusing current steganalysis algorithms.

It has previously been noted that the addition of noise to an image has a low-pass filtering effect on the intensity histogram of the image. We showed that LSB matching also manifests itself on the extrema of the histogram. Specifically, the local maxima of the intensity histogram of cover images will decrease, while its local minima will increase after LSB matching. We used this property to construct a new discriminant feature, the sum of absolute differences between each local extremum and its neighbors in the histogram. A key distinction from prior work, is that our features are local rather than global properties of the intensity histogram.

Experimental comparisons were performed between our histogram extrema method and the algorithms of Ker and Goljan *et al.*. Two datasets of 2000 never-compressed images were used. The experiments clearly demonstrated the improved performance of the histogram extrema method on datasets that contain high-frequency noise. However, when these datasets are JPEG compressed with a quality factor or 80, the histogram extrema method performs worse.

The complimentary nature of the histogram extrema method and that of Ker and Goljan *et al.* suggests a future research direction to develop a hybrid method that combines the advantages of all three methods.

It is well-known that the performance of steganalysis methods can vary greatly depending on the datasets. And we observed this in our experimental results. Further work is needed to understand this variability and to characterize it for particular algorithms.

Acknowledgements

Jun Zhang is supported by Guangdong Natural Science Foundation (NO:06023961),Department of Education of Guangdong Province(NO:05Z013) and China Scholarship Council.

References

[1] I. J. Cox, T. Kalker, G. Pakura, and M. Scheel, “Information transmission and steganography,” in *Proceedings of the 4th International Workshop on Digital Watermarking*, ser. LNCS, vol. 3710, September 2005, pp. 15–29.

[2] J. Harmsen and W. Pearlman, “Steganalysis of additive noise modelable information hiding,” in *Security and Watermarking of Multimedia Contents V*, ser. Proceedings of SPIE, vol. 5020, January 2003, pp. 131–142.

[3] A. D. Ker, “Steganalysis of LSB matching in grayscale images,” *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 441–444, June 2005.

[4] J. Fridrich, D. Soukal, and M. Goljan, “Maximum likelihood estimation of length of secret message embedded using $\pm k$ steganography in spatial domain,” in *Security and Watermarking of Multimedia Contents V*, ser. Proceedings of SPIE, vol. 5681, January 2005, pp. 595–606.

[5] T. Holotyak, J. Fridrich, and S. Voloshynovskiy, “Blind statistical steganalysis of additive steganography using wavelet higher order statistics,” in *Proceedings of the 9th IFIP TC-6 TC-11 Conference on Communications and Multimedia Security*, ser. LNCS, vol. 3677, September 2005, pp. 273–274.

[6] M. Goljan, J. Fridrich, and T. Holotyak, “New blind steganalysis and its implications,” in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, ser. Proceedings of SPIE, vol. 6072, January 2006, pp. 1–13.

[7] J. Fridrich, M. Goljan, and D. Soukal, “Higher-order statistical steganalysis of palette images,” in *Security and Watermarking of Multimedia Contents V*, ser. Proceedings of SPIE, vol. 5020, January 2003, pp. 178–190.

[8] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, “Detection of hiding in the least significant bit,” *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 3046–3058, October 2004.

[9] A. D. Ker, “A general framework for structural analysis of LSB replacement,” in *Proceedings of the 7th Information Hiding Workshop*, ser. Lecture Notes in Computer Science, vol. 3727, June 2005, pp. 296–311.