Steganalysis of LSB Matching in Grayscale Images

Andrew D. Ker

***Abstract*—We consider the problem of detecting spatial domain least significant bit (LSB) matching steganography in grayscale images, which has proved much harder than for its counterpart, LSB replacement. We use the *histogram characteristic function* (HCF), introduced by Harmsen for the detection of steganography in color images but ineffective on grayscale images. Two novel ways of applying the HCF are introduced: *calibrating* the output using a downsampled image and computing the *adjacency histogram* instead of the usual histogram. Extensive experimental results show that the new detectors are reliable, vastly more so than those previously known.**

***Index Terms*—Communication systems, computer security, information hiding, signal analysis.**

I. Introduction

THE aim of steganography is to hide information imperceptibly into a cover, so that the presence of hidden data cannot be diagnosed. Steganalysis aims to expose the presence of hidden data. In this letter, we present ways to detect a simple—but particularly difficult to uncover—embedding method for data in bitmap images; to our knowledge, this is the first reliable detector of its kind.

In Section II we include a description of the least significant bit (LSB) matching embedding algorithm and compare it with its more common counterpart, LSB replacement. In Section III, we define the histogram characteristic function (HCF), first used by Harmsen *et al.* [1] to detect additive noise steganography in color images; it fails in the case of grayscale images. In Sections IV and V, we describe two new ways to apply the HCF, and in Section VI, we give experimental results showing that the new techniques, separately or in combination, provide vastly more reliable diagnosis of LSB matching steganography than heretofore known. These results come from a distributed steganalysis project that allows the rapid evaluation of many variants of steganalysis methods against a database of tens of thousands of natural images. Finally, we outline some further directions for research.

II. LSB Matching Steganography

LSB steganography, in which the lowest bit plane of a bitmap image is used to convey the secret data, has long been known to steganographers. Because the eye cannot detect the very small perturbations it introduces into an image and because it is extremely simple to implement,[1] LSB methods are commonly used among the many free steganography tools available on the internet. There are two types of LSB steganography: *LSB replacement* can be uncovered relatively easily and is thoroughly examined in [2]–[4], but fewer and weaker detectors have been proposed for *LSB matching*. It is the latter we consider here, in the particular case when the covers are grayscale images. The LSB matching embedding algorithm is as follows.

Convert the secret data into a stream of bits. Take each pixel of the cover image (possibly in a pseudo-random order generated by a shared secret key): if the LSB of the next cover pixel matches the next bit of secret data, do nothing; otherwise, choose to add or subtract one from the cover pixel value, at random. When the secret message is fewer bits in length than the number of pixels in the cover image, the pseudo-random permutation ensures that changes are spread uniformly throughout the image. The allowable range of pixel values will force the decision of whether to increment or decrement, when the cover pixel is saturated.

LSB replacement is very similar, except that the LSBs of the cover pixels are simply overwritten by the secret bit stream. In either case, the decoding method for the recipient is simply to read back the LSBs of the stego image, according to the order specified by the secret key, if needed; the original cover image is not needed by the recipient and should be discarded by the sender.

Success in finding reliable and sensitive detectors for LSB replacement steganography has not been equalled for LSB matching. It is not immediately obvious why LSB matching should be any harder to detect, as both types of embedding add noise at the same level into the covers. The distinction is in the asymmetries inherent in LSB replacement—a cover pixel with an even value might be left alone or might be incremented by one but never decremented; the converse is true for odd-valued cover pixels. The detectors of [2]–[4] are exploiting this asymmetry, which is not induced by LSB matching. Indeed, these detectors are completely ineffective in exposing LSB matching.

The literature does contain a few detectors for LSB matching steganography. Westfeld [5] gives a detector for color images, which has its basis in the effect that the embedding algorithm has on the occurrences of close pairs of colors. Westfeld comments that the detector can be applied to grayscale images by converting triplets of gray pixels into the red, green, and blue (RGB) components of a single color pixel, but in practice, this detector performs barely better than a random decision—we give some experimental evidence in Section VI. It is also likely that the blind detector of Lyu and Farid [6] will be somewhat effective, but it has not been tested against a pure LSB matching

Manuscript received July 5, 2004; revised October 5, 2004. This work was supported by the Royal Society. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Petr Tichavsky.

The author is with the Oxford University Computing Laboratory, Oxford OX1 3QD, U.K. (e-mail: adk@comlab.ox.ac.uk).

Digital Object Identifier 10.1109/LSP.2005.847889

[1]In [2] is a program for simple LSB replacement, in the Perl scripting language, that is only 80 characters long.

steganography scheme. The other detector applicable to LSB matching is due to Harmsen [1]. Although the technique was presented for detection of LSB replacement (and some other types of steganography) in RGB color bitmaps, it is equally applicable to LSB matching in grayscale images. We turn to it next.

III. Histogram Characteristic Function (HCF)

Consider a grayscale cover image, made up of pixels with intensity in the range $0,\ldots,N-1$ (usually $N=256$). Write $p_c(i,j)$ for the intensity of the cover image at location (i,j). We imagine that a maximal-length hidden message (one secret bit per cover pixel) is embedded using LSB matching to form a stego image $p_s(i,j)$.

Harmsen's detector uses only the relative frequency of occurrence of each intensity, i.e., the histogram. Write

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

for the histogram of the cover image and similarly $h_s(n)$ for the histogram of the stego image after embedding. We model steganographic embedding as independent additive noise (mod N) and write f_Δ for the mass function of the noise random variable, which takes value n with probability proportional to $|\{(i,j)|p_s(i,j)-p_c(i,j)=n\ (\mathrm{mod}N)\ \}|$.

Because the addition of integer random variables corresponds to the convolution of their mass functions, $h_s=h_c*f_\Delta$. Let $H_c[k]$, $H_s[k]$, and $F_\Delta[k]$ be the N-element discrete Fourier transforms (DFTs) of $h_c(n)$, $h_s(n)$, and $f_\Delta(n)$, respectively; then, we have

$$H_s[k]=H_c[k]F_\Delta[k].$$

Harmsen calls $H_s[k]$ the HCF of the stego image. In view of the symmetry of DFTs of real signals, we need consider only $k=0,\ldots,N/2$. The distribution of the added noise in the case of LSB matching is just $f_\Delta(0)=0.5$, $f_\Delta(\pm1)=0.25$. Elementary calculation gives that $F_\Delta[k]=\cos^2(\pi k/N)$; this monotone function, always no greater than 1, drops to zero as k reaches $N/2$. Therefore, $H_s[k]$ will be no larger than $H_c[k]$ and for large k will be appreciably smaller.

The types of steganography considered by Harmsen also have this property. Thus, to diagnose their presence, Harmsen uses the *center of mass (COM)* of the HCF

$$\mathcal{C}\left(H[k]\right)=\frac{\sum_{i=0}^{n} i\,|H[i]|}{\sum_{i=0}^{n}|H[i]|}$$

where $n=N/2$ to avoid the redundant parts of the DFT. After steganographic embedding

$$\mathcal{C}\left(H_s[k]\right)<\mathcal{C}\left(H_c[k]\right). \tag{1}$$

In [1], Harmsen generalizes this to multidimensional signals, using multidimensional DFTs and producing a multidimensional COM. It is simply this COM that is the discriminator for detecting steganography in RGB color images. Tested against a number of additive-noise steganography methods in color bitmaps (not including LSB matching), quite reliable performance is observed. We have performed tests against LSB matching and found quite good performance here, too, although not as good as quoted in [1]. This is probably because Harmsen used a very limited and homogeneous set of test covers, and this tends to inflate the performance results.

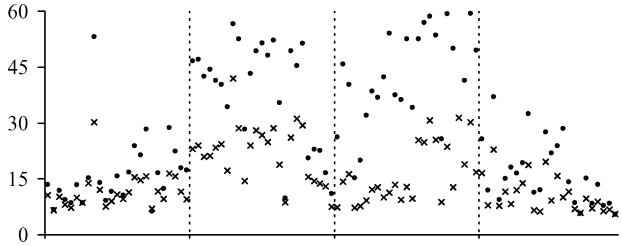

Fig. 1. Values of $\mathcal{C}(H[k])$ (*circles*) before and (*crosses*) after embedding for 100 images from four different sources.

Harmsen does not try the detector on grayscale images, although the theory suggests that it should work in the same way. The one-dimensional analogue is to give a positive diagnosis of steganography when $\mathcal{C}(H[k])$ is greater than a threshold. It turns out that such a detector performs very poorly indeed.

The reason is illustrated in Fig. 1, plotting the classifier $\mathcal{C}(H[k])$ for 100 grayscale images, before and after embedding. The images came from four different sources, and the chart is divided according to the image source. As can be seen, (1) generally holds (the crosses are almost all lower than the corresponding circles). However, the values of $\mathcal{C}(H_c[k])$ depend heavily on the source of cover image, and worse, there is high variability amongst $\mathcal{C}(H_c[k])$, often far more than the typical differences between $\mathcal{C}(H_c[k])$ and $\mathcal{C}(H_s[k])$. The significant weakness of this method is that the detector does not see the cover image and so does not know $\mathcal{C}(H_c[k])$.

There is an essential difference between the color and grayscale bitmaps. In the former, the histograms are fairly sparse, with colors occurring in clusters. Because there are only N (as opposed to N^3) possibilities for the pixels in grayscale images, such clustering does not happen. The result is noise in the HCF COM, which swamps its discriminating ability. In Section VI, we give evidence that it makes a poor detector of LSB matching steganography when the cover images are grayscale.

IV. Calibration by Downsampled Image

We first address the problem of the high variability of $\mathcal{C}(H_c[k])$. Consider downsampling an image by a factor of two in both dimensions using a straightforward averaging filter. Precisely, let $p'_c(i,j)$ be the pixel intensities of the downsampled cover image given by

$$p'_c(i,j)=\left\lfloor\sum_{u=0}^{1}\sum_{v=0}^{1}\frac{p_c(2i+u,2j+v)}{4}\right\rfloor$$

and $p'_s(i,j)$ the similarly downsampled version of the stego image. We divide the summed pixel intensities by four and take the integer part to reach images with the same range of values as the originals. We compute the HCF and COM of these two downsampled images $\mathcal{C}(H'_c[k])$ and $\mathcal{C}(H'_s[k])$.

Our observation is that $\mathcal{C}(H_c[k])$ and $\mathcal{C}(H'_c[k])$ are usually very close—the downsampling operation does not greatly affect the COM of the HCF of cover images. Fig. 2 plots these

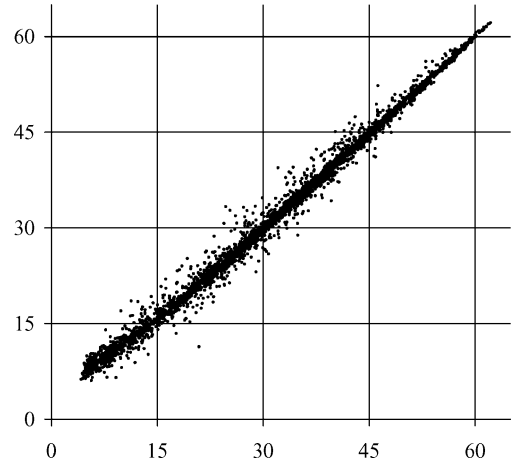

Fig. 2. (x-*axis*) $\mathcal{C}(H_c'[k])$ is close to (y-*axis*) $\mathcal{C}(H_c[k])$. Scatterplot from 2000 JPEG images.

two quantities computed from a library of 2000 JPEG cover images; the correlation coefficient is 0.9967, and the quantities are within 20% of each other for 90% of these images. Even scanned images that have never been subject to JPEG compression, which are usually very noisy and difficult to predict, tend to have this property (we observed a correlation coefficient of 0.9737 from a set of 3000 images never subject to JPEG compression).[2] We conclude that, for images without hidden data

$$\mathcal{C}\left(H_c'[k]\right) \approx \mathcal{C}\left(H_c[k]\right). \tag{2}$$

Second, although the LSB matching process does introduce noise into the downsampled cover image and consequently reduces the COM of the HCF, we have observed that it does so to a lesser extent than in the full-sized images, i.e.,

$$\mathcal{C}\left(H_c[k]\right) - \mathcal{C}\left(H_s[k]\right) > \mathcal{C}\left(H_c'[k]\right) - \mathcal{C}\left(H_s'[k]\right). \tag{3}$$

The inequality is justified by experimental evidence: Fig. 3 plots these two quantities for 2000 JPEG images to demonstrate that (3) usually holds; it is true for 97.5% of these cover images. We attribute this effect to the adding and rounding operations in the downsampling process, which tend to even out added noise. Combining (2) and (3) gives a new detector for the presence of LSB matching steganography: $\mathcal{C}(H[k]) < \mathcal{C}(H'[k])$ in most images that have been subject to steganography, with approximate equality in other images. In view of the variation between the magnitudes of such values, we use $\mathcal{C}(H[k])/\mathcal{C}(H'[k])$ as a dimensionless discriminator; the downsampled image is *calibrating* the COM of the full-sized image.[3] The reliability of this detector is tested in Section VI.

V. HCF of Adjacency Histogram

We now turn to the essential difference between color and grayscale images, namely, that the histograms of the former

[2]The 2000 images used to construct the scatterplot are taken at random from the set of 20 000 images described in Section VI. The 3000 uncompressed images refer to the set described in Section VI.

[3]Another form of calibration was developed independently by Fridrich in [7], although this technique is quite different and only applicable in DCT-domain steganography.

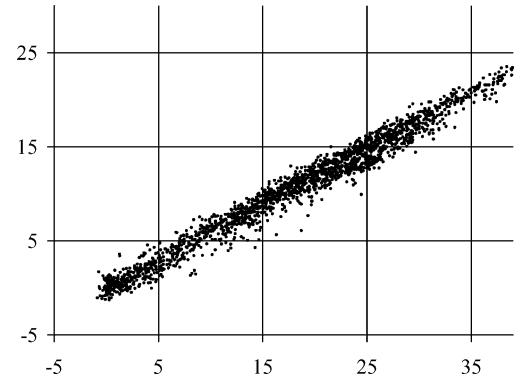

Fig. 3. (x-*axis*) $\mathcal{C}(H_c[k]) - \mathcal{C}(H_s[k])$ is usually greater than (y-*axis*) $\mathcal{C}(H_c'[k]) - \mathcal{C}(H_s'[k])$. Scatterplot from 2000 JPEG images.

tend to be sparse. One way to make the histogram artificially more sparse is to consider the two-dimensional *adjacency histogram*, expressing how often each pixel intensity is observed horizontally next to each other: $h_c^2(m,n) = |\{(i,j)|p_c(i,j) = m, p_c(i,j+1) = n\}|$. Because adjacent pixels tend to have close intensities, this histogram is sparse off the diagonal. It is affected by steganography, *mutatis mutandis*, in the same way as the standard histogram.

As before, we form the HCF $H^2[k,l]$ (this time using a two-dimensional DFT) and the two-dimensional COM. In order to produce a one-dimensional discriminator, it is necessary to combine the two-dimensional COM somehow. In view of the symmetry between the two components, we use only one quadrant of the DFT and simply take the components' sum, so the discriminator used is equal to

$$\mathcal{C}^2\left(H^2[k,l]\right) = \frac{\sum_{i,j=0}^{n}(i+j)\left|H^2[i,j]\right|}{\sum_{i,j=0}^{n}\left|H^2[i,j]\right|}.$$

As before, we expect this to be low in the presence of steganography. We can also combine this technique with the calibration of Section IV.

VI. Experimental Results

We compare the standard HCF COM detector against the adjacency histogram version and calibrated versions of both. We also include the statistic outlined in [5]. These results were produced by a distributed steganalysis database, outlined in [3], that uses a dedicated cluster of machines rapidly to evaluate the reliability of detectors over very large sets of cover images. We test here against two sets of images.

3000 Uncompressed Images. The entirety of the USDA NRCS Photo Gallery http://photogallery. nrcs.usda.gov: The images were downloaded as very high resolution TIF files (mostly 2100 × 1500) and appear to be scanned from a variety of film and paper sources. For testing, the images were resampled to 640 × 418 and converted to grayscale.

20 000 JPEG Images. From a stock photo CD [8]: These images are stored in JPEG format at quality factor 58 (fairly harsh compression) and are approximately sized 640 × 400. They were converted to grayscale before use.

In Fig. 4, we give receiver operating characteristic (ROC) curves, showing how false-positive and false-negative errors

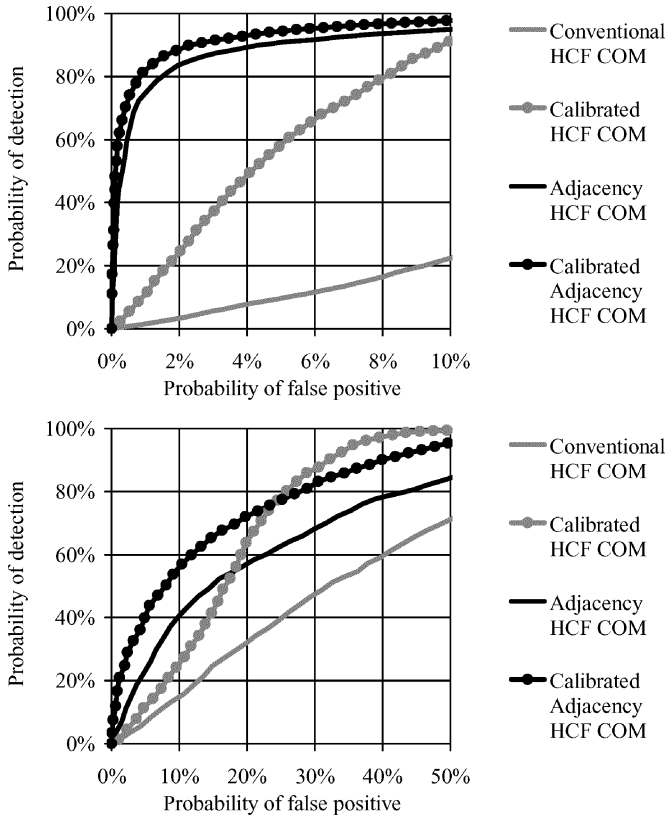

Fig. 4. ROC curves. (Top) Generated from 20 000 images that have been subject to fairly harsh JPEG compression. (Bottom) From 3000 uncompressed bitmaps.

tradeoff as the detection threshold is varied, for the two sets of cover images embedded with maximal-length random messages. Note that the x-axes have been scaled to focus on regions of interest. In [3], we argue that a reasonable one-dimensional measure of performance is the false positive rate when the false negative rate is 50%; this is shown in Table I. In order to simulate the fairly mild JPEG compression effects caused by typical storage in a digital camera, we have repeated this experiment with the 3000 uncompressed images first subject to JPEG compression at quality factor 80. In the case of maximal-length embedded messages, it is apparent that the new techniques give rise to detectors that are much more reliable: Detection in uncompressed covers (which contain a lot of noise even before embedding) is now possible, and detection in JPEG compressed covers becomes extremely reliable.

However, the calibration technique presented here is not so useful when the amount of hidden data is less than the maximum possible, as can be seen from the lower half of Table I. The adjacency histogram still leads to improved detectors, but calibration brings no substantial benefits and, in one case, a moderate worsening of performance. Further inspection shows the cause: Equation (3) does not hold for shorter messages. This issue remains to be addressed.

VII. Conclusion and Future Work

We have given two novel ways to apply the HCF in the diagnosis of steganography and demonstrated that they produce reliable detectors for LSB matching steganography in grayscale images. We should transfer these improvements into the full-color case. While the adjacency histogram may not be so useful, it is highly likely that calibration will improve the reliability of detection. We also aim to find a more refined method of calibration, from further study of $\mathcal{C}(H'_s[k])$, that degrades gracefully with shorter messages.

Finally, it is apparent from Harmsen's work that the detectors given here will also detect other types of steganography. While most unlikely to match the performance of dedicated asymmetry-exploiting statistics for LSB replacement, such as in [2], there may be applications in spread-spectrum spatial-domain steganography or possibly even in the DCT domain.

TABLE I
False-Positive Rate at Which False-Negative Rate Is 50%

	3000 bitmaps		20000
	unaltered	compressed	JPEGs
Maximal-length messages			
Westfeld's detector [5]	46.4%	44.8%	43.1%
HCF COM	31.8%	30.7%	19.1%
Calibrated HCF COM	16.8%	18.1%	4.1%
Adjacency HCF COM	14.5%	2.0%	0.3%
Calibrated Adjacency HCF COM	8.0%	0.3%	0.1%
Messages at 50% of the maximum			
Westfeld's detector [5]	46.9%	46.9%	43.0%
HCF COM	41.1%	39.2%	34.8%
Calibrated HCF COM	35.0%	38.1%	32.7%
Adjacency HCF COM	26.9%	8.8%	7.8%
Calibrated Adjacency HCF COM	27.5%	13.7%	7.3%

References

[1] J. Harmsen and W. Pearlman, "Higher-order statistical steganalysis of palette images," in *Proc. SPIE Security Watermarking Multimedia Contents*, vol. 5020, E. J. Delp III and P. W. Wong, Eds., 2003, pp. 131–142.
[2] A. Ker, "Improved detection of LSB steganography in grayscale images," in *Proc. Inf. Hiding Workshop, Springer LNCS*, vol. 3200, 2004, pp. 97–115.
[3] ——, "Quantitive evaluation of pairs and RS steganalysis," in *Proc. SPIE Security, Steganography, Watermarking Multimedia Contents*, vol. 5306, E. J. Delp III and P. W. Wong, Eds., 2004, pp. 83–97.
[4] J. Fridrich and M. Goljan, "Practical steganalysis of digital images—State of the art," in *Proc. SPIE Security Watermarking Multimedia Contents*, vol. 4675, E. J. Delp III and P. W. Wong, Eds., 2002, pp. 1–13.
[5] A. Westfeld, "Detecting low embedding rates," in *Proc. Inf. Hiding Workshop, Springer LNCS*, vol. 2578, 2002, pp. 324–339.
[6] S. Lyu and H. Farid, "Steganalysis using color wavelet statistics and one-class vector support machines," in *Proc. SPIE Security, Steganography, Watermarking Multimedia Contents*, vol. 5306, E. J. Delp III and P. W. Wong, Eds., 2004, pp. 35–45.
[7] J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," in *Proc. Inf. Hiding Workshop, Springer LNCS*, vol. 3200, 2004, pp. 67–81.
[8] *20,000 photos*, Focus Multimedia, Staffs, U.K., 2002.