Steganalytic methods for the detection of histogram shifting data-hiding schemes

Daniel Lerch and David Megías
Universitat Oberta de Catalunya, Spain.

ABSTRACT

In this paper, some steganalytic techniques designed to detect the existence of hidden messages using histogram shifting methods are presented. Firstly, some techniques to identify specific methods of histogram shifting, based on visible marks on the histogram or abnormal statistical distributions are suggested. Then, we present a general technique capable of detecting all histogram shifting techniques analyzed. This technique is based on the effect of histogram shifting methods on the "volatility" of the histogram of differences and the study of its reduction whenever new data are hidden.

1. INTRODUCTION

Data hiding [13] is a collection of techniques to embed secret data into digital media so that its existence becomes undetectable by some attacking party. Data hiding can be applied to secret communications, copyright protection, authentication of digital contents, etc. The most common carriers used for data hiding are images because their widespread use in the Internet.

To hide data into a cover image, pixel values are changed and, therefore, image distortion occurs. Usually, the distortion due to data hiding is not reversible and the original image can not be recovered. However, there are techniques that have the ability to restore the original image. These techniques are known as reversible data hiding [1, 3, 4, 5, 6, 7, 8, 10].

The simplest non-reversible data hiding method consists of modifying the least significant bit (LSB) of some (or all) pixel values, which is often referred to as LSB steganography. In [12], several attacks on LSB steganography are described. Later, in [2], the RS attack is introduced, which can reliably detect messages even for embedding capacious as low as 0.03 bpp.

A reversible data hiding method based on histogram shifting was proposed in [10], which uses the information about peaks and zeros of the cover image histogram to perform a partial shift, leaving a gap to hide data. In the Ni et al.'s [10] method, the embedded secret data can not be recovered when the knowledge of peak and zero point of histogram are not transmitted to the receiver. In order to overcome the above scenario, Hwang et al. [6] proposed a robust reversible data hiding scheme based on the histogram shifting method. In this method, they proposed the use of a location map that stored the information needed to reverse the process when the minimum point of histogram was non-zero.

Later in [4], Hong et al. presented a method that performs a shift of the histogram of prediction errors. This method is based on Ni et al. [10] but has greater capacity. Hong et al. [4] in their paper use the median edge detector (MED) to predict pixel values (see 2.3 section), Since the histogram of prediction errors is sharply centered at zero, we can use the concept of histogram shifting to hide information without determining the peak and zero points, unlike Ni et al. [10] method.

Although the histogram shifting technique is commonly used in reversible data hiding, several methods have recently emerged and are used as non-reversible ones [9].

There is some work on steganalysis applied to histogram shifting methods. Particularly, the detection based on changes in the shape of the histogram. In Huong et al. [5] a technique to attack the DIH method [8] is presented. This technique is based on an unusual shape in the histogram, similar to the attack we present in Section 2.1. However, this technique is not applicable to Ni et al. [10]. In Kuo et al. [7] a technique to attack the method HKC [6] is presented. As in the previous case, this technique is based on an unusual shape in the histogram. However, this method is not

applicable to Ni et al. [10] either. In both cases, this irregularity affects seven of the histogram bins. Therefore, it hardly ever occurs in cover (unmarked) images. In Ni et a. [10] the irregularity affects only four bins, making it harder to detect.

2.PROPOSED STEGANALYTIC METHODS

In this section, we present four steganalytic techniques. The first steganalytic technique detects Ni et al.'s [10] method, through abnormal shapes in the pixel intensity histogram. The second steganalytic technique detects Mohsenzadeh et al.'s [9] method using an unusual statistical distribution introduced by the algorithm. The third steganalytic technique detects HSPE [4] methods (Histogram Shifting of Prediction Errors) studying the volatility of the histogram. The fourth technique extends the third technique, to create a generic detection scheme which can be applied to detect all the above.

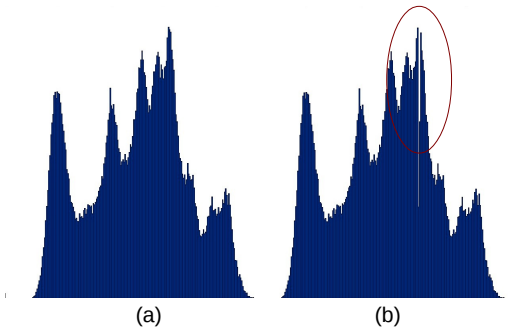
2.1 Ni et al.'s Method

In 2003, Ni et al. [10] presented a reversible data hiding method which consists in shifting the histogram of the image in order to create space to hide data. Their method uses a simple but effective algorithm:

1. Find the maximum (or peak) of the histogram, which corresponds to a pixel value P, and then find a zero, which corresponds to the pixel value Z.

2. Shift the histogram to the right, from peak to zero point. To do this we will add 1 to each pixel of the image with value between P and Z.

3. To insert the message, it is necessary to scan the entire image looking for all pixels with value P. $P+1$ is used to embed '1' and P to embed '0'.



(a) (b)

Fig. 2. (a) Histogram of the original image, (b) Histogram of the marked image

In Fig. 1, we can see the pixel intensity histogram of an example image (Lena), before and after data has been hidden.

If we compare the original histogram with the histogram of the marked image, there is a difference caused by histogram shifting and hiding data.

The abnormal shape in the histogram of the marked image can be detected with some reliability by applying the following observations:

Let a, b, c and d, four consecutive bins of the histogram:

1. $b+c$ is the maximum value of the histogram.
2. b and c have an approximately equal size.
3. a or d are not much smaller than $b+c$.

These three conditions require thresholds to work properly. In Section 3, the chosen values are shown.

2.2 Mohsenzadeh et al.'s Method

In 2009, Mohsenzadeh et al.[9] presented a steganographic method which is able to thwart histogram based steganalysis. Their method uses histogram shifting techniques to hide non reversible data with the following algorithm (we will name it Algorithm 1):

1. Find the maximum bin or peak of the histogram, which corresponds to a pixel value P, and then find the zero of the left (Z_l) and the zero of the right (Z_r) of the peak.

2. Shift the histogram to the right, from peak to zero, and do the same to the left. To do this, 1 is added to each pixel of the image with value between P and Z_r and 1 is subtracted from each pixel between P and Z_l.

3. To embed the message, it is necessary to scan the entire image looking for all pixels with values $P+2$ or $P-2$. To embed '0' we set $I(i-1,j)$ to $P+1$ (or $P-1$) and, to embed '1', we set $I(i+1,j)$ to $P+1$ (or $P-1$).

Mohsenzadeh et al. [9] also present a second method that uses a key and inserts the message in its eight neighboring pixels, rather than just in the right and left ones. It is referred to as Algorithm 2.

Algorithm 1 produces a significant statistical anomaly. There is always a $P+1$ / $P-1$ value next to $P+2$ / $P-2$. For this reason, we can detect hidden data with this algorithm, counting those occurrences next to all the pixels. However, this technique is not very reliable with Algorithm 2.

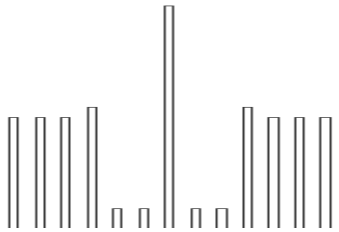


Fig. 3. Frequency of $P\pm1$ together with $P\pm2$

Fig 3. shows the distribution of $P\pm1$ values together with $P\pm2$ values in a marked image. As can be seen, there is an anomaly in the values around the peak.

2.3 Histogram Shifting of Prediction Errors Methods

Histogram Shifting of Prediction Errors (HSPE) methods were presented in [4]. There are many different methods based on HSPE, and the most popular ones have been selected for the experiments presented in this paper..

The alterations of histograms of prediction errors are more difficult to detect than those of pixel intensity histograms since histograms of prediction errors can be generated from different prediction formulas. However, neighboring pixels are often used. For example, Hong et al. [4] use the median edge detector (MED) prediction to calculate the predicted value of p:

$$p = \begin{cases} \min(b,c), & \text{if } a \geqslant \max(b,c), \\ \max(b,c), & \text{if } a \leqslant \min(b,c), \\ b+c-a, & \text{otherwise.} \end{cases}$$

where a, b and c are three neighbors of the corresponding pixel, as shown in Fig. 3:

a	b
c	x

Fig. 3. 2x2 adjacent pixels

Although there are simpler methods such as horizontal prediction:

$$p=c$$

vertical prediction:

$$p=b$$

diagonal prediction:

$$p=a$$

and others even more sophisticated, such as a causal template prediction, for example:

$$p=\frac{a+b+c+d}{4}$$

where a, b, c and d are shown in Fig. 4 with respect to the pixel *x* to be predicted.

a	b	d
c	x	

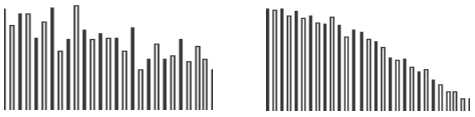
Fig. 4. 2x3 adjacent pixels

HSPE based methods create a histogram from the differences between each of the pixels in the image and its prediction p. This histogram can be used to embed a message using techniques such as Ni et al.'s [10].

We can not analyze the histogram of prediction errors, since we do not know what prediction formula was used. It is necessary to take another approach.

One of the common traits of the HSPE methods is that they modify areas where the pixels are similar. When similar pixels are

modified by adding one, these pixels will advance to the next bin of the histogram. As this occurs throughout the histogram, all the bins will now have the pixels of its neighbors, so we obtain a less volatile histogram.



(a) (b)

Fig. 2. (a) original image histogram
(b) HSPE marked image histogram

We can measure the volatility of the histogram comparing the value of each bin with the average of its neighbors.

$$V=\sum_{i=1}^{255}\left|\frac{H[i-1]-2\mathrm{H}[i]+H[i+1]}{3}\right|$$

The formula above is presented for clarity. However, normalizing the value of *V* will provide better results.

The experiments show that the volatility of the image histogram is significantly reduced when a message is embedded into a cover image. However, when it is embedded into a stego image, the volatility is reduced to a smaller extent. This provides a detection mechanism: the volatility of the analyzed image can be compared with the volatility after embedding a new message into it. If this process significantly reduces the volatility, the image is cover, otherwise it is stego.

2.4 Generic staganalytic method

The generic technique presented above is useful because it uses common characteristics of different data hiding systems. However, it does not detect the methods introduced by Ni et al. [10] or Mohsenzadeh et al. [9]. The reason is that these methods affect only a specific group of pixels, and therefore, only a few specific bins of the histogram. Thus, it is necessary to use a histogram in which all methods affect all the bins. We may use, for example, a histogram of differences.

We have used the following prediction:

$$p=\frac{a+b+c}{3}\quad ,$$

where a , b and c are:

a	b
c	x

Fig. 5. 2x2 adjacent pixels

And create a histogram based on the value of the differences.

$$H[i]=|x-p|$$

Now, the method proceeds as in the previous section. Firstly, the volatility is calculated, then a new message is embedded and, finally, the volatility is computed again. If it significantly increases the volatility, the image is cover, otherwise it is stego.

In the histogram of differences, after hiding new data, the volatility increases. However, after embedding a new message, the volatility remains almost unchanged.

3. EXPERIMENTAL RESULTS

In the experiments presented in this section, the database of 1371 images of NCRS [11] has been used. These images are marked with the different methods described above.

In this section, we present our experimental results obtained with all the proposed algorithms.

In the generic algorithm, we have used a threshold of 15%. This means that an image is considered stego if its volatility increases less than 15% when inserting a new message, otherwise it is considered cover. Experiments show that this threshold is appropriate.

For detecting HSPE methods the specific algorithm presented in Section 2.3 with a threshold of 15% has been used. This means that an image is considered stego if its volatility decreases less than 15% when inserting a new message, otherwise it is

considered cover. Experiments show that this threshold is appropriate.

3.1 Ni et al's method

The specific algorithm requires two thresholds. The first threshold is to verify that a and b have a similar size. We have used a maximum difference of 10%. The second threshold is to verify that a and d are not much smaller than $b+c$. A maximum difference of 30% has been used. The experiments show that these thresholds are appropriate.

Algorithm:	Specific	Generic
Successful	85.19%	85.22%
Positive	40.29%	44.93%
Negative	44.89%	40.29%
False positive	5.10%	9.70%
False negative	9.70%	5.06%

Table 1. Experimental results for Ni et al.'s method

The results are shown in Table 1. 1000 images have been used in the experiments, 500 of which are cover (unmarked) and the other 500 have been marked with Ni et al.'s method. The row “Successful” refers to the percentage of correctly identified images (either as cover or stego), the row “Positive” reports the percentage of the correctly identified stego images (the maximum is 50%), “Negative” reports the number of correctly identified cover (unmarked) images (the maximum is again 50%), “False positive” reports the percentage of unmarked images incorrectly identified as stego and, finally, “False negative” is the percentage of stego images which are not correctly detected by the technique. Note that the number of positives plus false negatives equals 50%. Analogously, the number of negatives plus false positives also equals 50% of the total number of images.

Note that both the specific and the generic method correctly identify more than 85% of the images. The specific scheme has a higher percentage of false negatives, whereas the generic one has a higher rate of false positives.

3.2 Mohsenzadeh et al.'s Algo 1 method

The specific algorithm described in Section 2.2 does not need threshold, just finds the shape, see Fig. 3.

Algorithm:	Specific	Generic
Successful	90.99%	81.65%
Positive	42.19%	41.35%
Negative	48.79%	40.29%
False Positive	01.23%	09.70%
False Negative	07.76%	08.64%

Table 2. Experimental results for Mohsenzadeh et al.'s Algo 1 method

The results shown in Table 2 indicate higher scores in the specific algorithm, but the generic algorithm also has remarkable results.

3.3 Mohsenzadeh et al.'s Algo 2 method

In this case, only the generic algorithm has been applied.

Algorithm:	Specific	Generic
Successful	-	90.18%
Positive	-	49.89%
Negative	-	40.29%
False Positive	-	9.70%
False Negative	-	0.10%

Table 3. Experimental results for Mohsenzadeh et al.'s Algo 2 method

The results shown in Table 3 indicate a large amount of successes using the generic algorithm. It is remarkable the highly reliable detection of positives, with a 49.89% for a maximum of 50%.

3.4 Horizontal HSPE

In this case, the thresholds used for the specific method for all HSPE techniques (Sections 3.4 to 3.8) are given at the beginning of the section.

Algorithm:	Specific	Generic
Successful	86.94%	87.16%
Positive	43.47%	46.86%
Negative	43.47%	40.29%
False Positive	6.52%	9.70%
False Negative	6.52%	3.13%

Table 4. Experimental results for horizontal prediction errors

Table 4 shows similar successes in both algorithms, which are slightly higher with the generic algorithm.

3.5 Vertical HSPE

Algorithm:	Specific	Generic
Successful	88.84%	87.19%
Positive	45.36%	46.90%
Negative	43.47%	40.29%
False positive	06.52%	09.70%
False negative	04.63%	03.09%

Table 5. Experimental results for vertical prediction errors

Table 5 shows similar successes in both algorithms, which are slightly higher with the specific algorithm.

3.6 Diagonal HSPE

Algorithm:	Specific	Generic
Successful	87.52%	88.65%
Positive	44.05%	48.35%
Negative	43.47%	40.29%
False positive	6.52%	9.70%
False negative	5.94%	1.64%

Table 6. Experimental results for diagonal prediction errors

Table 6 shows similar successes in both algorithms, which, again, are slightly higher with the generic algorithm.

3.7 Causal HSPE

Algorithm:	Specific	Generic
Successful	61.19%	86.10%
Positive	17.72%	45.80%
Negative	43.47%	40.29%
False positive	6.52%	9.70%
False negative	32.27%	4.19%

Table 7. Experimental results for causal prediction errors

The results of Table 7 show that the specific method is particularly unsuitable for this embedding technique using the same thresholds as for the other HSPE methods**.** Slightly modifying the threshold of the specific algorithm leads to rates above 80% of successful identification, but this also increases the number of false positives.

3.8 HSPE with MED prediction

Algorithm:	Specific	Generic
Successful	63.78%	85.88%
Positive	20.31%	45.58%
Negative	43.47%	40.29%
False positive	6.52%	9.70%
False negative	29.68%	4.41%

Table 8. Experimental results for MED prediction errors

The results shown in Table 8 are analogous to those of Section 3.7. ,Again, the successful identification rate with the specific technique can be improved over 80% by modifying the thresholds.

3.9 Generic experiments

In this section, an experiment was performed with 1000 cover images and 1000 stego images. The set of stego images contains a mixture of all methods presented in equal parts.

Algorithm:	Generic
Successful	86.05%
Positive	46.15%
Negative	39.90%
False positive	10.10%
False negative	3.85%

Table 9. Experimental results for different histogram shifting data-hiding methods

As shown in Table 9, the results using the generic algorithms for all methods presented are quite reliable.

4. CONCLUSIONS

In this paper, we have shown that histogram shifting based methods cause alterations in the image histogram and that these alterations can be detected. We have introduced an algorithm based on the analysis of the histogram volatility, which can be applied to several methods of data hiding.

The experimental results show that the analysis of the histogram volatility can be used to detect changes in the histogram, and that the difference histogram analysis provides remarkable results, being able to identify between 80% and 90% of the images correctly as cover (unmarked) or stego.

-As future work, it would be advisable to study the use of other histograms to estimate volatility, as well as the analysis application of this steganalytic technique to other methods of data hiding.

REFERENCES

[1] C. C. Chang, W. L. Tai. and C. C. Lin. A Reversible Data Hiding Scheme Based on Side Match Vector Quantization. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, No. 10, pp.1301-1308, 2006.

[2] J. Fridrich, M.Goljan, R. Du. Detecting LSB steganography in color and grays-cale Images. Proceedings of the ACM Workshop on Multimedia and Security. Ottawa, Canada. October 05, 2001.

[3] C. W. Honsinger, P. Jones, M. Rabbani, and J. C. Stoffel. Lossless recovery of an original image containing embedded data. US Patent application, Docket No: 77102/E-D, 2001.

[4] Wien Hong, Tung-Shou Chen, Chih-Wei Shiu. Reversible Data Hiding Based on Histogram Shifting of Prediction Errors. International Symposium on Intelligent Information Technology Application Workshops.

[5] Ho Thi Huong, Ho Van Canh, Trinh Nha Tien. Steganalysis for Reversible Data Hiding. International Journal of Database Theory and Application. Vo. 3, No. 2, June, 2010.

[6] J. H. Hwang, J. W. Kim, and J. U. Choi. A Reversible Watermarking Based on Histogram Shifting. IWDW, LNCS 4283, pp. 348-361, 2006.

[7] Wen-Chung Kuo, Yan-Hung Lin. On the Security of Reversible Data Hiding Based-on Histogram Shift. The 3rd International Conference on Innovative Computing Information. 2008.

[8] Sang-Kwang Lee, Young-Ho Suh, Yo-Sung Ho. Lossless Data Hiding Based on Histogram Modification of Difference Images. PCM 2004, LNCS 3333 (2004), pp340-347.

[9] Yalda Mohsenzadeh, Javad Mohjeri, and Shahrokh Ghaemmaghami. Histogram Shift Steganography: A technique to Thwart Histogram Based Steganalysis. 2009 Second International Workshop on Computer Science and Engineering.

[10] Zhicheng Ni, Yun-Qing Shi, Nirwan Ansari, Wei Su. Reversible Data Hiding. International Symposium on Circuits and Systems. Vol 2. Thailand. May 2003.

[11] NRCS (Natural Resources Conservation Service). Photo Gallery http://photogallery.nrcs.usda.gov/ (Last accessed on June 19th, 2011).

[12] Andreas Westfeld, Andreas Pfitzmann. Attacks on Steganographic Systems. Breaking the Steganographic Utilities EzStego, Jsteg, Steganos, and S- Tools and Some Lessons Learned. 3rd International Workshop on Information Hiding. Dresdenn, Germany. 29 September – 01 October 1999.

[13] W. Zeng. Digital watermarking and data hiding: technologies and applications. Proc. Int. Conf. Inf. Syst., Anal. Syhnth., vol 3, 1998, pp. 223-229.