Improved LSB Matching Steganography Resisting Histogram Attacks

Ling Xi, Xijian Ping, Tao Zhang
Zhengzhou Information Science and Technology Institute
Zhengzhou , China
bluesunshine_xl@sina.com

Abstract—**This paper presents an improved LSB matching steganography, which complementarily modifies the pairs of pixels with adjacent intensity to embed secrete message. In LSB matching steganography, when adding or subtracting one from the cover image pixel, two adjacent bins of the histogram will be altered---the bin value of the modified pixel's intensity increased by one, and one of its adjacent bin's value decreased by one. Based on the alteration of histogram caused by LSB matching, the improved algorithm embeds two bits in a pair of pixels with adjacent intensity one time so as to minimize the alteration of histogram. Compared to the original LSB matching, the new one preserves the first order statistical property of cover image extremely. Experimental results show the proposed algorithm obviously outperforms LSB matching algorithm in the aspect of resisting first order statistical attacks based on histogram.**

Keywords- steganography;LSB matching; adjacent histogram; complementary; statistical analysis;

I. Introduction

The aim of steganography is to hide information imperceptibly into a cover media, so that the presence of hidden data cannot be diagnosed. LSB replacement steganography is a famous effective method which substitutes the last significant bits of the cover media for secrete bits . But it will produce "pairs of value" on the intensity histogram of the stego image, according as which analysers can detect the presence of secrete message successfully using statistical analyzer tools such as χ^2 [1]、RS[2]. Whereafter LSB matching steganography was put forward to eliminate the phenomenon of "pairs of value" caused by LSB replacement method. But study on LSB matching showed that it reduced to a low-pass filtering of intensity histogram, which offered analyser an opportunity to distinguish stego image from cover one. Such steganalyzers as HCF-COM[3] (the steganalysis based on histogram characteristic function centre of mass) and ALE[4] (the steganalysis based on amplitudes of local extrema) are designed to label the images embedded using LSB matching steganography.

Study on state-of-the-art stegnalyzers, we find that for fighting back attacks based on histogram, the most important thing is to minimize the histogram alteration caused by steganography. From such a viewpoint, we propose a new steganography algorithm which significantly improves LSB matching. Recurring to the complementary effect on the histogram of random ± 1, it makes majority embedding modifications take place at the pixels with adjacent intensity. Therefore the histogram of stego image is very similar as the cover one. Steganalytic experiments with calibrated HCF-COM and 1D ALE of different image database clearly demonstrate that the improved algorithm is predominant over the previous one.

II. Mproved Lsb Matching Algorithm Based On Modification Of Pixels With Adjacent Intensity

A. Effects of LSB matching steganography on histogram

The original LSB matching algorithm can be formally described as follows:

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

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 $\{-1, +1\}$. This process can be applied to all pixels in the image or only for a pseudo-randomly chosen portion, when the embedding rate ρ is less than one, the length of the hidden message is less than the number of pixels in the image.

Let's consider the influence of LSB matching on intensity histogram. For the encrypted secrete bits are pseudorandom, the possibility that the secrete bit equals to the LSB of the selected pixel is 1/2. It means that $1-\rho/2$ pixels of the cover image keep unchanged and the remainders will be modified. Because the distribution of κ is uniform on {-1,+1} ,the $\rho/2$ pixels are modified by +1 or -1 with the same possibility 1/2. Therefore the possibility of a intensity k which depicted as histogram of stego image can be given by[5]:

$$h_s(k) = \frac{\rho}{4} h_c(k-1) + (1-\frac{\rho}{2}) h_c(k) + \frac{\rho}{4} h_c(k+1) \quad (2)$$

Expression(2) means LSB matching induces a low-pass filtering with the kernel $\{\frac{\rho}{4}, \ 1-\frac{\rho}{2}, \ \frac{\rho}{4}\}$ of the intensity histogram. This implies that the histogram of a stego contains less high-frequency power than the histogram of the corresponding cover image. The abnormality was used by steganalyzers to distinguish stego images from cover images. Accordingly this is the key point of our improved work.

B. The Improved Algorithm Based on Modification of Pixels with Adjacent Intensity

978-1-4244-5540-9/10/$26.00 ©2010 IEEE

The intensity histogram is a statistical characteristic of an image which pictures the appearance frequency of each intensity. The spatial steganography always embeds secrete bits by modifying the intensity of pixels, therefore the histogram of a stego image usually is distinguishable from the cover. LSB matching embeds secrete bits by ± 1 randomly. The randomicity of the modification direction offers us opportunity to improve LSB matching. Let's study the impact on the histogram when the i th modification occurs. Assume the intensity of current pixel is k . $h^{(i)}(k)$ denotes the appearance frequency of intensity k after the i th modification. Whichever modification directions we take, +1 or -1, the appearance frequency of intensity k will decrease by 1. It means the embedding operation removes 1 from bin k of the histogram which valued as $h^{(i)}(k)$. At the same time, the value of its adjacent bin ($h^{(i)}(k+1)$ or $h^{(i)}(k-1)$ will increase by 1 according to +1 or -1 modification . The following expression tells us the change of the histogram after the i th modification:

$$\begin{cases} h^{(i)}(k) = h^{(i-1)}(k) - 1 \\ h^{(i)}(k+1) = h^{(i-1)}(k+1) + 1 \ \ \textit{(add 1)} \\ h^{(i)}(k-1) = h^{(i-1)}(k-1) + 1 \ \ \textit{(minus 1)} \end{cases} \tag{3}$$

Assume the embedding modification occurs M times all together, then $h^{(0)}(k)$ is the bin k value of cover image histogram and $h^{(M)}(k)$ is that of stego image. Therefore we gets the boundary condition of (3) :

$$h^{(0)}(k) = h_c(k) \tag{4}$$

$$h^{(M)}(k) = h_s(k) \tag{5}$$

(3)~(5) depict the change of histogram during the whole embedding process. It is easy to find that the modification of one pixel causes change of two adjacent bins of histogram, and the modification direction is adverse.

Assume A and B are two pixels selected to bear secrete message using LSB matching. The intensity of $A(x,y)$ is k , while intensity of $B(x',y')$ is $k+1$. And the LSB of A and B are all different from their respective secrete bits to be embedded. That means to embed secret bits, the two pixel's intensity must be add or subtract 1 randomly. If we choose +1 to A and -1 to B ,then two bits of secrete message is embedded, While the +1 operation of A alters the histogram by

$$\begin{cases} h^{(i)}(k) = h^{(i-1)}(k) - 1 \\ h^{(i)}(k+1) = h^{(i-1)}(k+1) + 1 \end{cases} \tag{6}$$

And the -1 of B by:

$$\begin{cases} h^{(i+1)}(k+1) = h^{(i)}(k+1) - 1 \\ h^{(i+1)}(k) = h^{(i)}(k) + 1 \end{cases} \tag{7}$$

Thus the i th and the $i+1$ th modification is expressed as follows dependently:

$$\begin{cases} h^{(i+1)}(k+1) = h^{(i)}(k+1) - 1 \\ h^{(i+1)}(k) = h^{(i)}(k) + 1 \end{cases} \tag{8}$$

Take (8) into (7),we gets the synthetic influence on the histogram of the two times' modification:

$$\begin{cases} h^{(i+1)}(k) = h^{(i-1)}(k) \\ h^{(i+1)}(k+1) = h^{(i-1)}(k+1) \end{cases} \tag{9}$$

It means when we introduce some constraint to the modification mode of A and B , the histogram will keep unchanged after embedding the two bits.

To the convenience of description, we name the two pixel A and B as a pair of complementary pixels when they satisfy:

$$LSB(p_c(A)) \neq b_1 \ \text{ and } \ LSB(p_c(B)) \neq b_2$$

$$p_c(A) = p_c(B) + 1 \ \text{ or } \ p_c(A) = p_c(B) - 1$$

Whereas $p_c(A)$ and $p_c(B)$ denote the intensity of pixel A and B in cover image while b_1 and b_2 denote the corresponding secrete bits.

According to above discussion, we know that when embedding two bits in a pair of complementary pixels by +1 to the lower intensity one and -1 to the higher, the histogram will not change at all. Thus it can be seen that when all the embedding modifications occur on pairs of complementary pixels, the intensity histogram of cover image and that of stego will be just the same. Whereas not any pixel can find its complementary one in an image. We call those pixels who need to be modified to embedding secret bits and can't find its complementary pixel as an isolated embedding pixel. So the isolated pixels still will make the stego histograms different. Thus the number of isolated pixels is an important parameter to the performance of the algorithm. We usually choose nature image as cover image, and the continuity of its intensity promises the number of the complementary pairs will be far more than that of isolated embedding ones. Then the validity of the improved algorithm can be promised too.

By above analysis, we achieve the procedure of the improved LSB matching algorithm which try out to find pairs of complementary pixels to embed data. Following is the flow chart:

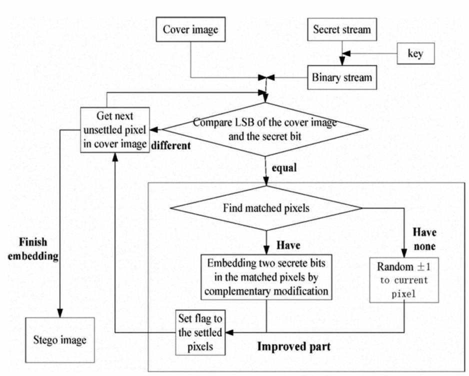


Figure 1. Flow char of the improved steganography algorithm

C. *Analysis for the performance of the Improved Algorithm*

Refer to part B, we know it is the ratio of the number of the isolated embedding pixels to that of complementary ones that decides the performance of the new algorithm. And the ratio has a close relationship with such propers of the cover image as size and content. Then we will discuss it in more detail. At first define

$$\gamma = \frac{number\ of\ isolated\ embedding\ pixels}{number\ of\ modified\ pixels} \qquad \gamma \le 1$$

Then let's consider the influence to the intensity histogram of the improved algorithm under the definition of γ.

$$\begin{aligned} & P\{p_s = p_c + 1\} \\ & = \rho\ P\{b \ne LSB(p_c)\}\ P\{\kappa > 0\} \\ & \ P\{isolated\ embedding\ pixels \mid modify\ pixels\} \\ & = \frac{\rho\gamma}{4} \end{aligned} \tag{10}$$

where $P\{p_s = p_c + 1\}$ denotes the probability of +1 embedding, $\rho \bullet P\{b \ne LSB(p_c)\} \bullet P\{\kappa > 0\}$ denotes the probability of +1 modification, while $P\{isolated\ embedding\ pixels \mid modified\ pixels\}$ which equals to γ denotes the possibility of embedding in isolated embedding pixels. In the same way, we can also get the other two formulas as:

$$\begin{aligned} P\{p_s = p_c - 1\} &= \frac{\rho\gamma}{4} \\ P\{p_s = p_c\} &= 1 - \frac{\rho\gamma}{2} \end{aligned} \tag{11}$$

Considering the relationship between histogram and the appearance frequency of intensity, we can conclude that with the new algorithm:

$$h_s(k) = \frac{\rho\gamma}{4} h_c(k-1) + (1 - \frac{\rho\gamma}{2}) h_c(k) + \frac{\rho\gamma}{4} h_c(k+1) \tag{12}$$

From (12), we know that as long as $\gamma \le 1$, the histogram of stego image and that of cover image will be more alike gotten by the improved method than by LSB matching. Iff all the modified pixel is isolated, that is when $\gamma = 1$,the improved algorithm is just the precious LSB matching . That is to say, on most conditions, the improved algorithm will do better in the aspect of histogram keeping than LSB matching.

The key factor of the new algorithm is γ .It is closely related to the size and content of cover image. To show the relationship, we calculate γ of 200 pieces of cover images with different size separately, carry out the mean, maximum and minimum, and list the result in the following table:

TABLE I. THE INFLUCE OF IMAGE SIZE TO γ

	γ_min	γ_max	γ_mean
128×128	0.009569	0.258353	0.063309
256×256	0.003043	0.098913	0.013693
512×512	0.002402	0.040562	0.006325

From the table, we can draw a conclusion that to general nature images, $\gamma << 1$, and the larger the cover image is, the smaller the γ will be. It means the improved algorithm get better effect for large cover image.

With a view to PNSR and visual impercipience, there is no difference between the new algorithm and LSB matching, for not only the number of modified pixels but also the modification amplitude of the two algorithms is just the same. That is to say, the well visual invisibility and good image quality are kept alive in the improved algorithm. Furthermore, the number of modified pixels of the two algorithms is same too. Then the embedding capacity of them is no difference.

III. Experimental Results

A. *The Experimental Results for Typical Image*

Using encrypted binary bit stream as secrete message, we hiding data in typical 512×512 cover image 'man.bmp' with the rate of 100% by LSB matching and our improved algorithm separately, the results showed as follows:

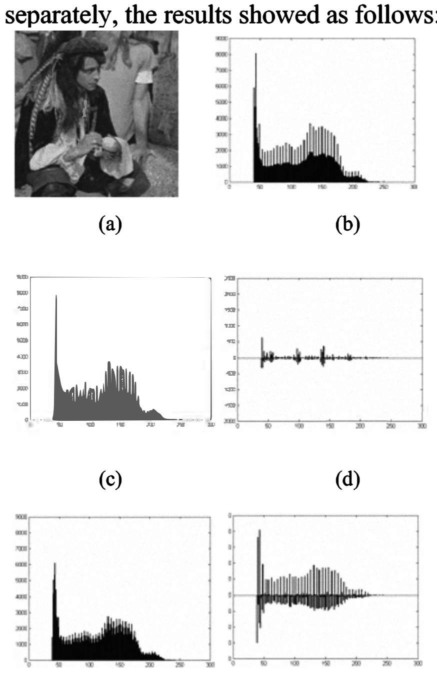


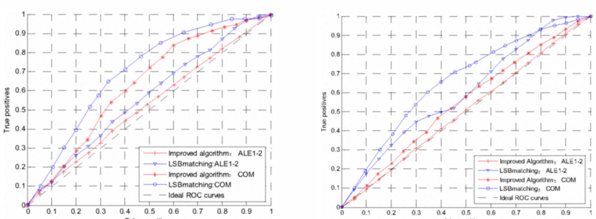
Figure 3. (a) 512×512 stantard image man.bmp as cover image (b)The histogram of cover image(c)Stego image histogram with improved method of 100% rate (c) Difference between stego image and cover image by improved algorithm (d) Stego image histogram by LSB matching (e) Difference between stego image and cover image with LSB matching

Fig 3 (c)(d) gives the stego image's histogram and the histogram difference between the stego and cover image using improved algorithm, and as a comparison (e) (f) shows that of LSB matching. From those figures, we can find the stego image histogram gotten by new method is very similar as the cover image, and the low pass filtering effect is not so obvious compared to the previous method.

B. The Experimental Results for Anti-histogram Statistical Analysis

Nowadays, HCF-COM proposed by Harmsen and ALE proposed by J Zhang *et al.* are two prevalent steganalysis methods for LSB matching. For a better partition, Ker improved Harmsen's HCF-COM as a calibrated HCF-COM algorithm. And Ker *et al* modified the design of the ALE steganalyzer by removing interferences at the histogram borders and incorporating additional complementary 2-dimention features. In these method, they all made use of the low pass filtering effect of LSB matching to steganalyse. So it can be deduced that our improved stegonagraphy will be more steganalysis-resistance than LSB matching.

In order to value of histogram statistical analysis resistant property of our improved algorithm, we do experiments on a database composed of images originating from two different sources. 1,000 images from the Camera database used in [6] and the other 1,000 images from the NRCS Photo Gallery. To each of these images, we used LSB matching and the improved algorithm to embed encrypted secrete data with rate 100%, then detected them using calibrated HCF COM steganalyzer and the improved 1D ALE steganalyzer. The following ROC curves is the results of our experiments.



(a) (b)

Figure 4. (a) ROC Curves of Camera database (b)ROC Curves of NRCS database

When detected by ALE, we randomly choos e 50% images as training set, while remainders as test set, repeated 10 times. And the ROC curves are vertically averaged by the fixed false positives to obtain the mean performance of the system. From fig 4, we know no matter which database is used, the true positives for improved algorithm based on our complementary modification are lower than the LSB matching algorithm under the same false positives.

IV. CONCLUSION AND FUTURE WORK

In order to eliminate the influence of histogram of LSB matching steganography and intensify the capability of statistical analysis resistance, we proposed an improved algorithm based on complementary modification. The algorithm embeds as many secrete bits in pairs of complementary pixels as possible, so as to reduce the modification which will make histogram be filtered by a low-pass filtering. Because the modification amplitude and the number of modified pixels are all same as LSB matching steganography, the improved algorithm keeps the good visual invisibility and high embedding capacity of LSB matching.

Experimental results show that the improved algorithm outperforms LSB matching in resisting statistic attack of 1D histogram. But the 2D features are not considered in the algorithm, which make it can not resist the 2D steganalyzer. This issue needs to be studied further.

AKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China (No. 60903221).

REFERENCES

[1] WESTFELD A , TZMANN A P . Attacks on steganographic systems[C]. 3 Inlernalional Workshop 011 Inybrmation Hiding, 1999.

[2] FRIDRICH J, GOLJAN M, DU R. Detecting lsb steganography in color and gray scale images[C]. IEEE Multime–dia Special Issue on SP c'urity。 2001, 22 23

[3] J.Harmsen and W.Pearlman. "Steganalysis of additive noise modelable information hiding" in Security and Watermarking of Multimedia Contents V, January 2003.vol.5020 of Proceedings of SPIE.pp.131-142

[4] J. Zhang, I. J. Cox, and G. Do ¨ err, "Steganalysis for LSB matching in images with high-frequency noise," [C] in Proceedings of the IEEE Workshop on Multimedia Signal Processing, October 2007 pp385-388.

[5] J. Harmsen and W. Pearlman, "Steganalysis of additive noise information hiding" in Security and Watermaking of Multimedia Contents V. January 2003, vol. 5020 of Proceedings of SPIE pp 131-142.

[6] Giacomo Cancelli , Gwena ¨ el Do ¨ err , Mauro Barni , and Ingemar J. Cox A Comparative Study of ± 1 Steganalyzers [C], in Proceedings of the IEEE International Workshop on Multimedia Signal Processing, pp. 791-796, 2008.