Comparative Histogram Analysis of LSB-based Image Steganography

KI-HYUN JUNG
Department of Cyber Security
Kyungil University
50 Gamasil-gil, Hayang-eup, Gyeongsan-si, Gyeongbuk 38428
REPUBLIC OF KOREA
khanny.jung@gmail.com

Abstract: In image steganography, least significant bits replacement and pixel-value differencing methods have been used. In this paper, data hiding methods based on least significant bits substitution are analyzed and compared on the side of histogram attack. In comparative histogram analysis, histogram of stego-image is displayed considering the embedding capacity and visual image quality. Through histogram analysis, a new image steganography will be designed that has a high embedding capacity and be robust to histogram attack.

Key-Words: data hiding, steganography, steganalysis, least significant bit, histogram attack.

1 Introduction
Information security has been important to keep important data from unauthorized access. Cryptography and data hiding techniques are common used to solve data security such as ownership and copyright protection. In steganography, least significant bit (LSB) replacement and pixel-value differencing (PVD) are common techniques in spatial domain. LSB-based steganography can embed the secret data into least significant bits within a permitted limit on image distortion to human visual system. PVD-based steganography used the difference value of two consecutive pixels to keep high embedding capacity [1-3]. LSB and PVD algorithm are often used in nowadays [4-30].

There are many kinds of algorithm which is based on LSB replacement techniques [4-16]. Least significant bit matching technique is to add or subtract one bit randomly when the embedding secret bit does not match. Revisited least significant bit matching method is to improve the image quality by reducing the number of modification. There are some works to improve embedding capacity and image quality like as optimal least significant bit substitution, the genetic algorithm based least significant bit substitution and the modulus least significant bit replacements.

In this paper, k-bit LSB replacement, LSB matching revisited and XOR-based data hiding methods are described. For LSB-based image steganography, histogram analysis according to embedding capacity is compared. From various histogram analysis, a new image steganography is presented that will be worked in details.

This paper is organized as follows. Section 2 reviews LSB-based image steganography and analyzed histogram of stego-image. In Section 3, comparative histogram analysis is compared. In Section 4, some solution to be secure on histogram analysis is discussed and the conclusions and future work are presented in Section 5.

2 Related Works
In this section, some previous works with 1 bit per pixel (bpp) on the embedding capacity are described. A gray cover image with $W \times H$ can be displayed as Fig 1.

![Fig. 1. Structure of a cover image.](image-url)

A pixel value $p(x, y)$ has a range from 0 to 255 for a gray image as Eq. (1).

$$p(x, y) = \{p(x, y) \mid 0 \leq p(x, y) \leq 255, 0 \leq x \leq W, 0 \leq y \leq H\}$$

(1)
A pixel contained gray value $p(x, y) = (p_7p_6p_5p_4p_3p_2p_1p_0).$ can be displayed as Fig. 2.

$$\begin{array}{c|c|c|c|c|c|c|c|c|c} 
\text{LSB7} & \text{LSB6} & \text{LSB5} & \text{LSB4} & \text{LSB3} & \text{LSB2} & \text{LSB1} & \text{LSB0} \\
\hline 
\hline 
p_7 & p_6 & p_5 & p_4 & p_3 & p_2 & p_1 & p_0 \\
\hline 
\end{array}$$

Fig. 2. Structure of a pixel.

2.1 LSB-based Data Hiding Methods

In $k$-bit LSB replacement method, the $k$ secret bits can be embedded into the rightmost positions of the pixel of a cover image for $k$-bit LSB replacement technique. Assume that $s_i$ are secret bit stream for a pixel $p_i$, new pixel $p_i'$ can be calculated by Eq. (2) in embedding scheme.

$$p'(x, y) = \left( p(x, y) - (p(x, y) \mod 2^k) + \sum_{i=0}^{k-1} s_i \right) \quad (2)$$

In extracting scheme, the secret bits can be extracted from the stego-pixel without any side information.

$$s_i = p'(x, y) \mod 2^k \quad (3)$$

For example, it is called 1-bit LSB replacement technique for $k = 1$ and 2-bit LSB substitution for $k = 2$.

LSB matching revisited [8] was proposed to improve the image quality with the same embedding capacity. The proposed method modified the LSB matching to ±1 from the pixel-pair values of a cover image. The function $F(i, j)$ is given as Eq. (4).

$$F(i,j) = f(p(x,y), p(x,y+1))$$

\[ = \text{LSB}\left(\frac{p(x,y)}{2}\right) + p(x,y+1) \quad (4) \]

For two secret bits $s_i$ and $s_{i+1}$, new two pixel values can be calculated by Eq. (5).

\[ \left( \begin{array}{c} p'(x,y) \\ p'(x,y+1) \end{array} \right) = \left( \begin{array}{c} p(x,y) \\ p(x,y+1) \pm 1 \end{array} \right), \]

\[ \text{if } s_i = \text{LSB}(p(x,y)) \text{ and } s_{i+1} \neq F(i,j) \]

\[ = \left( \begin{array}{c} (p(x,y) - 1) \pm 1, p(x,y+1) \end{array} \right), \]

\[ \text{if } s_i \neq \text{LSB}(p(x,y)) \text{ and } s_{i+1} = F(i,j) \]

\[ \left( \begin{array}{c} p(x,y) \pm 1, p(x,y+1) \end{array} \right), \]

\[ \text{if } s_i \neq \text{LSB}(p(x,y)) \text{ and } s_{i+1} \neq F(i,j) \]

The data hiding technique using XOR operation in a sub-block with three pixels was proposed to embed three secret bits [16]. New three pixels can be calculated as follows for three pixels $p(x, y) = (a_i+a_i+6a_i+5a_i+4a_i+3a_i+2a_i+1a_i),$ $p(x, y+1) = (b_i+b_i+6b_i+5b_i+4b_i+3b_i+2b_i+1b_i),$ $p(x, y+2) = (c_i+c_i+6c_i+5c_i+4c_i+3c_i+2c_i+1c_i)$ and three secret bits $s_i, s_{i+1}, s_{i+2}$. First, new three functions $F_i, F_{i+1},$ and $F_{i+2}$ can be calculated by Eq. (6).

$$\begin{align*}
F_i &= a_i \oplus a_{i+1} \oplus b_i \\
F_{i+1} &= b_i \oplus b_{i+1} \oplus c_i \\
F_{i+2} &= c_i \oplus c_{i+1} \oplus a_i
\end{align*} \quad (6)$$

Next, new three pixels $p'(x, y), p'(x, y+1), p'(x, y+2)$ can be obtained by Eq. (7).

\[ (p'(x,y), p'(x,y+1), p'(x,y+2)) = \begin{cases} 
(p(x,y), p(x,y+1), p(x,y+2)) & \text{if } F_i = s_i \text{ and } F_{i+1} = s_{i+1} \text{ and } F_{i+2} = s_{i+2} \\
(p(x,y), p(x,y+1) \pm 1, p(x,y+2)) & \text{if } F_i \neq s_i \text{ and } F_{i+1} = s_{i+1} \text{ and } F_{i+2} = s_{i+2} \\
(p(x,y), p(x,y+1), p(x,y+2) \pm 1) & \text{if } F_i = s_i \text{ and } F_{i+1} \neq s_{i+1} \text{ and } F_{i+2} = s_{i+2} \\
(p(x,y) + 1, p(x,y+1), p(x,y+2)) & \text{if } F_i = s_i \text{ and } F_{i+1} \neq s_{i+1} \text{ and } F_{i+2} \neq s_{i+2} \\
(p(x,y) + 1, p(x,y+1) \pm 1, p(x,y+2)) & \text{if } F_i \neq s_i \text{ and } F_{i+1} = s_{i+1} \text{ and } F_{i+2} \neq s_{i+2} \\
(p(x,y) \pm 1, p(x,y+1), p(x,y+2)) & \text{if } F_i \neq s_i \text{ and } F_{i+1} \neq s_{i+1} \text{ and } F_{i+2} = s_{i+2} \\
(p(x,y) \pm 1, p(x,y+1) \pm 1, p(x,y+2)) & \text{if } F_i \neq s_i \text{ and } F_{i+1} \neq s_{i+1} \text{ and } F_{i+2} \neq s_{i+2} \\
(p(x,y), p(x,y+1), p(x,y+2) \pm 1) & \text{if } F_i = s_i \text{ and } F_{i+1} \neq s_{i+1} \text{ and } F_{i+2} \neq s_{i+2} \text{ and } p(x,y+2) \mod 2 = 0 \\
(p(x,y) \pm 1, p(x,y+1), p(x,y+2)) & \text{if } F_i \neq s_i \text{ and } F_{i+1} = s_{i+1} \text{ and } F_{i+2} \neq s_{i+2} \text{ and } p(x,y) \mod 2 = 0
\end{cases} \quad (7)
These data hiding techniques can embed the secret bit in the pixel, in other words, the embedding capacity is 1 bit per pixels.

2.2 Histogram Analysis of LSB Layers

For LSB layers, histogram was analyzed for Airplane and Baboon images. 512 x 512 gray images were used and the embedding secret data was generated by pseudo-random number. The Airplane image and histogram are shown in Fig. 3.

![Airplane](image1)
![Histogram of Airplane](image2)

![Baboon](image3)
![Histogram of Baboon](image4)

Fig. 3. Airplane image and histogram.

To understand the distortion for layer level, the results of Airplane image by inserting one bit for each pixel is tested as shown in Fig. 4.

![LSB0](image5)
![LSB1](image6)

![LSB2](image7)
![LSB3](image8)

![LSB4](image9)
![LSB5](image10)

![LSB6](image11)
![LSB7](image12)

Fig. 4. Airplane images for layer level.

Histogram analysis of pixel layer is tested as shown in Fig. 5. For Airplane image, histogram of stego-image is different as the layer is higher. In especial, histogram graph of Fig. 5(h) corresponding to Fig. 4(h) has more distortion than any other histogram graphs.
For Baboon image, result images of each LSB layer are displayed in Fig. 6.

Even though the number of changed bits is similar except the different layer in 512 x 512 gray images, the value of PSNR and Q index are different as shown in Table 1.
### Table 1. Comparison of visual quality

<table>
<thead>
<tr>
<th>Layer</th>
<th>Airplane</th>
<th>Baboon</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>Q index</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td>LSB0</td>
<td>51.13</td>
<td>0.9602</td>
</tr>
<tr>
<td>LSB1</td>
<td>45.12</td>
<td>0.8894</td>
</tr>
<tr>
<td>LSB2</td>
<td>39.10</td>
<td>0.7446</td>
</tr>
<tr>
<td>LSB3</td>
<td>33.08</td>
<td>0.5803</td>
</tr>
<tr>
<td>LSB4</td>
<td>27.06</td>
<td>0.4115</td>
</tr>
<tr>
<td>LSB5</td>
<td>21.04</td>
<td>0.3143</td>
</tr>
<tr>
<td>LSB6</td>
<td>15.01</td>
<td>0.1422</td>
</tr>
<tr>
<td>LSB7</td>
<td>08.98</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

In Table 1, the value of PSNR from LSB0 to LSB3 maintained 30 dB above which means it is difficult to discriminate the distortion by human visual system. As a result, 1-bit substitution can keep the shape of histogram even though the layer is higher. In next section, histogram analysis is tested on previous image steganography.

### 3 Comparative Histogram Analysis

In experimental results, six gray images were used as cover images sized with 512 x 512 as shown in Fig. 8 and the secret data was generated by pseudo-random number. Test program was developed by visual studio and tested on Intel i5-4590 CPU 3.30GHz with 4GB memory.

Histogram of cover images is displayed in Fig. 9 to compare with LSB-based image steganography.

Histogram analysis is described for two categories – $k$-bit LSB and LSB-based data hiding methods.

#### 3.1 $k$-bit LSB Data Hiding

In this subsection, LSB substitution methods for $k = 1, 2, 3$ are compared. For Airplane image, histogram is displayed in Fig. 10, where the embedding capacity is 262,144 bits and PSNR is 51.13 dB for $k = 1$, 524,288 bits and 43.99 dB for $k = 2$, and 786,432 bits and 35.75 dB for $k = 3$. As a result, histogram of 1-bit LSB is similar with that of
a cover image, but 2-bit LSB and 3-bit LSB are easy to discriminate each other.

For Baboon image, the embedding capacity and PSNR are 262,144 bits and 51.16 dB for $k = 1$, 524,288 bits and 44.43 dB for $k = 2$, and 786,432 bits and 35.70 dB for $k = 3$ respectively.

As a result, 1-bit LSB technique can be applied with other image steganography considering histogram.

### 3.2 LSB-based Data Hiding

LSB matching, LSB matching revisited and LSB using 3 pixels sub-block are tested since the embedding capacity is similar [5, 7, 13]. For Airplane image, the embedding capacity is 262,144 bits and PSNR is 51.13 dB for LSB matching, 262,144 bits and 51.14 dB for LSBM revisited, and 233,024 bits and 52.76 dB for LSB using 3 pixels sub-block.

For Baboon image, test results of embedding capacity and PSNR are 262,144 bits and 51.16 dB, 262,144 bits and 44.18 dB, and 233,024 bits and 40.68 dB respectively.

Next, LSB-based methods based on 2-bit XOR embedding and modulo three strategies are tested and analyzed [12, 15].
For Airplane image, the embedding capacity is 524,288 bits and PSNR is 43.97 dB in 2-bit XOR method, 960,100 bits and 37.63 dB in LSB method based on modulo three strategies.

For Baboon image, results of embedding capacity and PSNR are 524,288 bits and 44.86 dB, 960,100 bits and 37.66 dB in respective. Considering the results of histogram, a new embedding scheme is required to be robust to histogram attack.

Table 2. Comparison of performance

<table>
<thead>
<tr>
<th>Cover images</th>
<th>1-bit LSB</th>
<th>2-bit LSB</th>
<th>3-bit LSB</th>
<th>LSB matching</th>
<th>LSBM revisited</th>
<th>LSB using 3-pixels sub-block</th>
<th>LSB using Modulo Three Strategy</th>
<th>LSB using 2-bit XOR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
</tr>
<tr>
<td>Airplane</td>
<td>262,144</td>
<td>51.13</td>
<td>524,288</td>
<td>35.75</td>
<td>262,144</td>
<td>51.13</td>
<td>529,416</td>
<td>35.75</td>
</tr>
<tr>
<td>Animal</td>
<td>262,144</td>
<td>51.16</td>
<td>524,288</td>
<td>35.67</td>
<td>262,144</td>
<td>51.16</td>
<td>541,120</td>
<td>35.67</td>
</tr>
<tr>
<td>Baboon</td>
<td>262,144</td>
<td>51.16</td>
<td>524,288</td>
<td>35.70</td>
<td>262,144</td>
<td>51.16</td>
<td>544,976</td>
<td>35.70</td>
</tr>
<tr>
<td>Lena</td>
<td>262,144</td>
<td>51.14</td>
<td>524,288</td>
<td>35.70</td>
<td>262,144</td>
<td>51.14</td>
<td>525,892</td>
<td>35.70</td>
</tr>
<tr>
<td>Man</td>
<td>262,144</td>
<td>51.14</td>
<td>524,288</td>
<td>35.66</td>
<td>262,144</td>
<td>51.14</td>
<td>529,944</td>
<td>35.66</td>
</tr>
<tr>
<td>Peppers</td>
<td>262,144</td>
<td>51.14</td>
<td>524,288</td>
<td>35.69</td>
<td>262,144</td>
<td>51.14</td>
<td>527,564</td>
<td>35.69</td>
</tr>
<tr>
<td>Average</td>
<td>262,144</td>
<td>51.15</td>
<td>524,288</td>
<td>35.70</td>
<td>262,144</td>
<td>51.15</td>
<td>533,152</td>
<td>35.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cover images</th>
<th>1-bit LSB</th>
<th>2-bit LSB</th>
<th>3-bit LSB</th>
<th>LSB matching</th>
<th>LSBM revisited</th>
<th>LSB using 3-pixels sub-block</th>
<th>LSB using Modulo Three Strategy</th>
<th>LSB using 2-bit XOR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
<td>EC (bits)</td>
<td>PNSR (dB)</td>
</tr>
<tr>
<td>Airplane</td>
<td>262,144</td>
<td>51.13</td>
<td>524,288</td>
<td>35.75</td>
<td>262,144</td>
<td>51.13</td>
<td>529,416</td>
<td>35.75</td>
</tr>
<tr>
<td>Animal</td>
<td>262,144</td>
<td>51.16</td>
<td>524,288</td>
<td>35.67</td>
<td>262,144</td>
<td>51.16</td>
<td>541,120</td>
<td>35.67</td>
</tr>
<tr>
<td>Baboon</td>
<td>262,144</td>
<td>51.16</td>
<td>524,288</td>
<td>35.70</td>
<td>262,144</td>
<td>51.16</td>
<td>544,976</td>
<td>35.70</td>
</tr>
<tr>
<td>Lena</td>
<td>262,144</td>
<td>51.14</td>
<td>524,288</td>
<td>35.70</td>
<td>262,144</td>
<td>51.14</td>
<td>525,892</td>
<td>35.70</td>
</tr>
<tr>
<td>Man</td>
<td>262,144</td>
<td>51.14</td>
<td>524,288</td>
<td>35.66</td>
<td>262,144</td>
<td>51.14</td>
<td>529,944</td>
<td>35.66</td>
</tr>
<tr>
<td>Peppers</td>
<td>262,144</td>
<td>51.14</td>
<td>524,288</td>
<td>35.69</td>
<td>262,144</td>
<td>51.14</td>
<td>527,564</td>
<td>35.69</td>
</tr>
<tr>
<td>Average</td>
<td>262,144</td>
<td>51.15</td>
<td>524,288</td>
<td>35.70</td>
<td>262,144</td>
<td>51.15</td>
<td>533,152</td>
<td>35.70</td>
</tr>
</tbody>
</table>
In experimental results, previous works that have 1 bit per pixel embedding capacity are robust relatively to histogram attacks. As increasing the embedding capacity, the results of histogram is distinguished with that of a cover image. Although the LSB-based image steganography has a high embedding capacity, histogram and PSNR are weak to human visual system.

### 4 Design of Robust Steganography

In this section, a basic algorithm to be robust to histogram attack is deduced.

First, consider that the pixel-value differencing algorithm [19] is robust to histogram attack. For Airplane image as shown in Fig. 16, the histogram is similar even though the embedding capacity is 409,778 bits and the PSNR is 40.06 dB which is larger than 1-bit LSB replacement. The histogram of Baboon image has also similar result as in Fig. 16(b). The embedding bits are 457,087 bits and the PSNR is 37.00 dB. Many previous works related to pixel-value differencing method have similar results and it can be useful to design a new data hiding method to be robust to histogram attack.

![Fig. 16. Histogram of pixel-value differencing.](image)

In histogram analysis, it is difficult to propose a new LSB-based image steganography to be robust to histogram attack. As a result, hybrid image steganography can solve the histogram problem.

#### Algorithm 1. Embedding Algorithm

<table>
<thead>
<tr>
<th>Input</th>
<th>a cover image and secret data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>a stego-image</td>
</tr>
<tr>
<td>Step 1</td>
<td>Divide into two non-overlapping pixel-pair for a cover image</td>
</tr>
<tr>
<td>Step 2</td>
<td>Select MAX and MIN value in the pixel-pair</td>
</tr>
<tr>
<td>Step 3</td>
<td>Apply minus operation to the MAX pixel</td>
</tr>
<tr>
<td>Step 4</td>
<td>Apply plus operation to the MIN pixel</td>
</tr>
<tr>
<td>Step 5</td>
<td>Repeat Step 1 to Step 4</td>
</tr>
</tbody>
</table>

The pseudo code of the proposed embedding algorithm is described in Algorithm 1 by considering histogram analysis.

In Algorithm 1, the proposed embedding algorithm will use a pixel block not a pixel only since least significant bits replacements are weak on histogram attack by using a pixel only. By using pixel-pair, it make difficult to analyze whether there has secret data in image itself. The MIN and MAX value are useful not to be uniform and it can solve the fall-off the boundary problem. The minus and plus operation are considered to use the gap between two pixels in the sub-block.

The proposed embedding algorithm can hide two bits of the secret data. In the future work, the proposed algorithm will be implemented and tested. Next, the pseudo code of extracting algorithm is described in Algorithm 2.

#### Algorithm 2. Extracting Algorithm

<table>
<thead>
<tr>
<th>Input</th>
<th>a stego-image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>the secret data</td>
</tr>
<tr>
<td>Step 1</td>
<td>Divide into two non-overlapping pixel-pair for a stego-image</td>
</tr>
<tr>
<td>Step 2</td>
<td>Extract 2-bit embedding data from MAX pixel</td>
</tr>
<tr>
<td>Step 3</td>
<td>Extract 2-bit embedding data from MIN pixel</td>
</tr>
<tr>
<td>Step 4</td>
<td>Accumulate the extracted secret bits</td>
</tr>
<tr>
<td>Step 5</td>
<td>Repeat Step 1 to Step 4</td>
</tr>
</tbody>
</table>

The proposed method was motivated at differentiated the embedding bits on each pixel. To do this, the maximum and minimum value can be used to be robust.

#### 5 Conclusion and Discussions

In this paper, image steganography based on least significant bits has been comparative and analyzed on the side of histogram. LSB-based image steganography has been used in data security. In the comparative analysis, histogram of a stego-image was similar with that of a cover image on the low
embedding capacity. As increasing the embedding capacity, image steganography could be discriminated by human visual system. After histogram analysis, the design of a new data hiding method to be robust to histogram attack was introduced and described advantages of the proposed embedding and extraction method. Main ideas to propose a data hiding method to be robust to histogram attack are (a) a method of embedding and extracting the secret data using a sub-block with multiple pixels should be proposed without using only one pixel and (b) there is required to embed and extract the secret data in each pixel in different ways, without hiding the secret data uniformly in one pixel within the sub-block. In the future, a proposed pseudo algorithm will be implemented and compared with previous works. And the proposed work will be robust to histogram attack.

Acknowledgement
This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01058019).

Disclosure Statement
This paper is recommended to publish by extending the conference paper that has presented at IARAS conference 2017 in Athens [16].

References:


