DETECTING LSB MATCHING BY CHARACTERIZING THE AMPLITUDE OF HISTOGRAM

Yunkai Gao, Xiaolong Li, Bin Yang and Yifeng Lu

Institute of Computer Science and Technology, Peking University, Beijing 100871, China

ABSTRACT

In this paper, we present an improved method for detecting LSB matching steganography in gray-scale image. Our improvements focus on three aspects: (1) instead of using the amplitude of local extrema of the image's histogram in the previous work, we turn to considering the sum of the amplitude of each point in the histogram; (2) incorporating the calibration (downsample) technique with the current method; (3) the sum/difference image (which is defined as the sum or difference of two adjacent pixels in the original image) is taken into consideration to provide additional statistical features. Extensive experimental results show that the novel steganalyzer outperforms the previous ones.

Index Terms— Information hiding, steganalysis, LSB matching

1. INTRODUCTION

Steganography is the art and science of secret communication. A steganographic scheme thus embeds secret data in innocuous looking cover data (e.g., digital images) so as not to arouse the third party's suspicion. On the contrary, the goal of steganalysis is to detect whether a given medium has secret data in it. Furthermore, steganalysis can serve as an effective way to judge the security of steganographic techniques.

As two widely-used steganographic schemes for digital images, LSB (least significant bit) replacement [1] and LSB matching (also known as "± 1 embedding") [2] have advantages of high payload, good visual imperceptibility and extreme ease of implementation. The embedding procedure of LSB replacement is rather simple: firstly, convert the secret data into a stream of bits; secondly, choose cover pixels in a pseudo-random order generated by a shared secret key; finally, replace the LSB of each selected cover pixel by the corresponding secret data bit. For LSB matching, it is a minor modification of LSB replacement: when the secret data bit does not match the LSB of the cover image, 1 is either added to or subtracted from the cover pixel value randomly. By exploring the embedding asymmetry in LSB replacement, some recent work has shown that one can effectively detect LSB replacement [3–5], even when the embedding rate (secret data bits embedded per pixel) is rather low. Nevertheless, the study on steganalysis of LSB matching is just in its early stage. In fact, it has been proved that LSB matching is much harder to detect than LSB replacement [6]. In this paper, we study the steganalytic techniques and give an improved method for detecting LSB matching. Some state-of-the-art and recently proposed steganalyzers are briefly reviewed as below.

Existing steganalytic methods can be classified into two categories: targeted and blind (universal). The targeted steganalytic methods aim to identify the presence of secret data embedded by a specific steganographic scheme, while the blind steganalytic methods are independent of the steganographic scheme. Usually, for a specific steganographic scheme, the blind steganalyzers are less efficient than the targeted steganalyzers. In [7], Harmsen *et al.* proposed to use the HCF COM (center of the mass of the histogram characteristic function) for the detection of additive noise based steganography including LSB matching. In [6], incorporating the HCF COM and the calibration (downsample) technique, Ker improved Harmsen *et al.*'s work by proposing some targeted steganalyzers for detecting LSB matching. Recently, Ker's calibration based steganalyzers were further studied by Li *et al.* [8]. The authors suggested calculating the calibration based steganalyzers on the difference image, which is defined as the difference of the adjacent pixels in the original image. In [9], Goljan *et al.* introduced the so-called WAM (wavelet absolute moment) blind steganalyzer, which was reported to outperform Ker's steganalyzers in [6], especially for compressed images. As most blind steganalyzers (e.g., [10, 11]), WAM estimates the stego noise by applying some denoising techniques. In [12], based on the statistics of ALE (amplitude of local extrema) of image's histogram, Zhang *et al.* described a targeted steganalyzer to detect LSB matching. The authors observed that the ALE would decrease after LSB matching embedding. Cancelli *et al.* extended Zhang *et al.*'s work to the two dimensional histogram [13]. They demonstrated experimentally that the novel steganalyzer was much more reliable than the original one described in [12], the WAM, and the calibration based steganalyzers in [6].

Based on the methods proposed in [12, 13] and the calibration technique originally introduced in [6], this paper presents an improved method for detecting LSB matching, which can remarkably increase the detection performance. In Section 2, we give a brief description of Zhang *et al.*'s work [12] and Cancelli *et al.*'s work [13]. Then in Section 3, we present our improvements on the ALE based steganalyzers. Other than using the sum of histogram's ALE in [12,13], we consider the sum of the amplitude of each point in the histogram (one dimensional and two dimensional) instead. Moreover, the calibration technique and the utilization of sum/difference image are taken into consideration. Extensive experimental results are shown in Section 4. The final conclusions are drawn in Section 5.

2. THE ALE BASED STEGANALYZERS

Let I_c be a gray-scale image, I_s be its stego image by LSB matching with embedding rate α, h_c and h_s be the histograms of I_c and I_s, respectively. As a consequence of LSB matching embedding, we know that h_s is a regularization of h_c: $h_s = f_\alpha * h_c$, where the convolutional kernel f_α is the distribution of embedding noise: $f_\alpha(0) = 1 - \alpha/2$, $f_\alpha(1) = f_\alpha(-1) = \alpha/4$. In [12], the authors proved that $h_s(n) < h_c(n)$ when n is a local maximum of h_c, and the inverse is true for local minimum, where the local extremum n is defined by: $h_c(n) > h_c(n \pm 1)$ (for local maximum) or $h_c(n) < h_c(n \pm 1)$ (for local minimum). That is to say, after LSB matching embedding, the local maxima of an image's histogram will decrease and the local minima will increase. Based on this obser-

Corresponding author: Bin Yang, e-mail: yangbin@icst.pku.edu.cn

vation, Zhang *et al.* pointed out that the ALE would reduce after LSB matching embedding. Then they proposed to use the following quantity

$$D(I) = \sum_{n \in E_1} |2h_1(n) - h_1(n+1) - h_1(n-1)| \qquad (1)$$

as a discriminant to classify images as cover or stego, where E_1 is the set of local extrema of h_1, h_1 is the histogram of the observed image I. Here we use subscript "1" to indicate the one dimensional (1-D, in brief) case. As expected, experimental results have shown that $D(I_s) < D(I_c)$ holds for most images. In [13], Cancelli *et al.* improved Zhang *et al.*'s work, they slightly modified the discriminant $D(I)$ and extended the ALE to the two dimensional (2-D, in brief) histogram. Firstly, to remove the border effect (note that the 0-valued pixels will always increase and the 255-valued pixels will always decrease), $D(I)$ was partitioned into two parts:

$$A_1(I) = \sum_{n \in E_1 \cap \{3,4,\ldots,252\}} |2h_1(n) - h_1(n+1) - h_1(n-1)|$$

and

$$d_1(I) = \sum_{n \in E_1 \cap \{1,2,253,254\}} |2h_1(n) - h_1(n+1) - h_1(n-1)|.$$

Secondly, noting that the change of 2-D histogram can also be modeled as a convolution procedure (from cover to stego), the authors considered the 2-D histogram h_2 and defined the following features:

$$A_2(I) = \sum_{n \in E_2} \left| 4h_2(n) - \sum_{p \in N} h_2(n+p) \right|$$

and

$$d_2(I) = \sum_{k=0}^{255} h_2(k,k),$$

where $N = \{(0, \pm 1), (\pm 1, 0)\}$ and $E_2 \subset \{1, 2, ..., 254\}^2$ is the subset of local extrema of h_2 which satisfies the symmetrical property: $(n_1, n_2) \in E_2 \Leftrightarrow (n_2, n_1) \in E_2$. Here, similarly to the 1-D case, the local extremum of h_2 is defined by: $h_2(n) > h_2(n+p)$, $\forall p \in N$ (for local maximum) or $h_2(n) < h_2(n+p)$, $\forall p \in N$ (for local minimum). Finally, by considering the above mentioned 1-D histogram based features $\{A_1(I), d_1(I)\}$ and the 2-D histogram based features $\{A_2(I), d_2(I)\}$ with four directions (horizontal, vertical, main diagonal, and minor diagonal), the authors used the $2 + 2 \times 4 = 10$ features and the FLD (Fisher linear discriminant) to build a two-class classifier to distinguish cover images from stego images.

3. THE IMPROVED STEGANALYZER

Before introducing our improvements on the ALE based steganalyzers, we first point out that the local extrema of h_c would change after data embedding, i.e., the set E_1 and E_2 would vary when the cover image turns into stego. Here is an example. The 512×512 gray-scale image "Lena" has 67 local extrema for 1-D histogram. The number of local extrema changes to 69, 66, 76 and 72 after data embedded by LSB matching with embedding rate 0.1, 0.2, 0.5 and 1. The number of local extrema shared by the cover image and the stego image is 59, 51, 50 and 26, respectively. Moreover, the same phenomenon occurs for 2-D histogram. Thus the uncertainty of local extrema of histogram can affect the detection performance. Since LSB matching embedding leads to low pass filtering the intensity histogram,

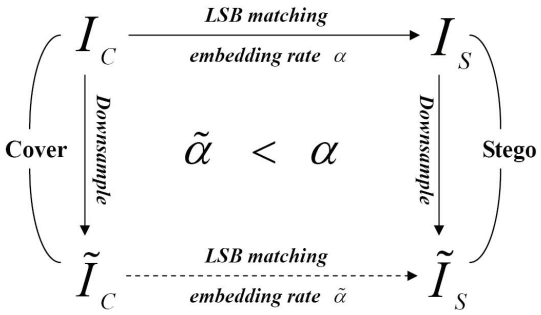


Fig. 1. The calibration (downsample) technique involves in LSB matching steganography.

the filtering operation will reduce the amplitude of each point of the histogram, not only the extrema. Hence, based on the above discussion, one natural modification of the discriminant $D(I)$ defined by Eq.(1) can be made as follows, in which we sum the amplitude of each point in the histogram,

$$D_1(I) = \sum_{n=1}^{254} |2h_1(n) - h_1(n+1) - h_1(n-1)|. \qquad (2)$$

The comparison of the detection performance of $D(I)$ with $D_1(I)$ will be reported later.

As mentioned in Section 1, the calibration (downsample) technique originally proposed by Ker is very useful for detecting LSB matching [6]. The downsampled stego image can be regarded as the stego image of the downsampled cover image by LSB matching with a reduced embedding rate, thus the procedure of downsample can reduce the embedding noise (see Fig.1). The fact was theoretically illustrated by Ker [14] and Li *et al.* [8]. We omit the details here due to the limitation of the space. Indeed, this technique can also make contribution when combining with $D_1(I)$, i.e., we propose to consider the following dimensionless quantity:

$$D_2(I) = \frac{\sum_{n=1}^{254} |2h_1(n) - h_1(n+1) - h_1(n-1)|}{\sum_{n=1}^{254} |2\widetilde{h}_1(n) - \widetilde{h}_1(n+1) - \widetilde{h}_1(n-1)|}, \qquad (3)$$

where $\widetilde{h}_1$ is the histogram of the downsampled image $\widetilde{I}$, whose pixel value is given by

$$\widetilde{I}_{i,j} = \left\lfloor (I_{2i,2j} + I_{2i+1,2j} + I_{2i,2j+1} + I_{2i+1,2j+1})/4 \right\rfloor. \qquad (4)$$

Fig.2 shows the comparison of the ROC (receiver operating characteristics) curves for four different steganalyzers: (1) $D(I)$, (2) $D_1(I)$, (3) $D_2(I)$, and (4) $\{D_1(I), D_2(I)\}$. In the case (4), we use $\{D_1(I), D_2(I)\}$ as statistical features and put them in SVM (support vector machine) to build a two-class classifier. From this figure, we can see that $D_1(I)$ performs a little better than $D(I)$, and the calibration based steganalyzer $D_2(I)$ can significantly improve the detection performance as compared with its original (non-calibrated) form $D_1(I)$. Finally, we can get the best detection performance when combining $D_1(I)$ with $D_2(I)$. This primary experimental result verifies our aforementioned discussion. Here, we use 3000 uncompressed images for testing, and the cover images are embedded with maximal-length random data bits.

Now, we consider the 2-D histogram. We know that, the 2-D histogram can reflect some relationship between two adjacent pixels, but the data embedding procedure may destroy this relationship in certain sense. The fact gives a clue to design steganalyzers, e.g., the

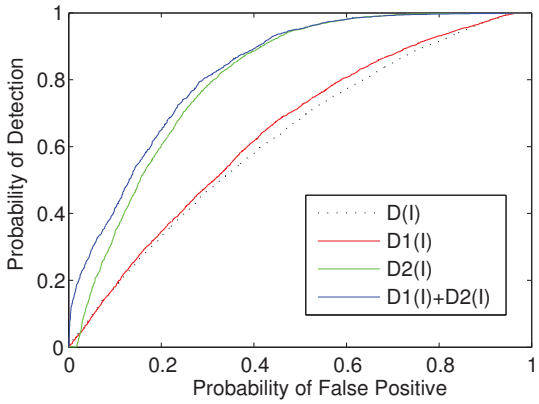


Fig. 2. Comparison of ROC curves for 3000 uncompressed images. The different curves stand for: $D(I)$ (black, dotted), $D_1(I)$ (red, solid), $D_2(I)$ (green, solid), and $\{D_1(I), D_2(I)\}$ (blue, solid).

adjacency HCF COM [6]. Following the work of Cancelli *et al.*, we extend the 1-D histogram based statistical features $D_1(I)$ and $D_2(I)$ to the 2-D case. We then define the additional second order statistical features as below, in which we sum the amplitude of each point in the 2-D histogram and incorporate the calibration technique:

$$D_3(I) = \sum_{n\in M} \lambda_n \left| 4h_2(n) - \sum_{p\in N} h_2(n+p) \right| \tag{5}$$

and

$$D_4(I) = \frac{\sum_{n\in M} \lambda_n |4h_2(n) - \sum_{p\in N} h_2(n+p)|}{\sum_{n\in M} \lambda_n |4\widetilde{h}_2(n) - \sum_{p\in N} \widetilde{h}_2(n+p)|}, \tag{6}$$

where $M = \{1, 2, ..., 254\}^2$ is the index set,

$$\lambda_n = \frac{1}{1+(n_1-n_2)^2} \qquad n = (n_1, n_2)$$

are weighted parameters which reflect the sparse intensity of 2-D histogram, $\widetilde{h}_2$ is the 2-D histogram of the downsampled image $\widetilde{I}$ which is defined by Eq.(4). The weighted parameters have been experimentally proved effective, and they can also be replaced by other similar functional expressions.

Furthermore, we introduce the concepts of sum image and difference image. For an image I, the horizontal sum image I^s is defined by: $I^s_{i,j} = I_{i,2j} + I_{i,2j+1}$, and the horizontal difference image I^d is defined by: $I^d_{i,j} = I_{i,2j} - I_{i,2j+1} + 255$, thus the pixel values of I^s and I^d vary from 0 to 510. Similarly, one can define the vertical sum and difference image. In [8], Li *et al.* suggested calculating the calibration based steganalyzers on difference image. The reason is that the difference image can better present the embedding noise as compared with the original image when wrapped by LSB matching, since the distribution of pixel value of difference image is rather concentrated and the maximal modification changes from 1 (for original image) to 2 (for difference image) after data embedding. Sum and difference images have also been proved experimentally to be obviously useful, so we take features from sum and difference image into account.

According to the afore discussion, our novel steganalytic algorithm can be described as follows.

Step 1: For a given image, calculate the sum image and the difference image in two directions (horizontal, vertical). Then we get totally five images.

Step 2: For each image derived from Step 1, calculate its downsampled image by Eq.(4), and get 1-D histogram and 2-D histogram in four directions (horizontal, vertical, main diagonal, and minor diagonal) for the image and its downsampled image, then calculate the features $D_1(I)$, $D_2(I)$, $D_3(I)$, and $D_4(I)$ according to Eq.(2), Eq.(3), Eq.(5), and Eq.(6) for 1-D or 2-D histogram. Therefore we arrive at a total of $5 \times (2 + 2 \times 4) = 50$ features.

Step 3: Put the 50 features into the SVM classifier to get the result.

4. EXPERIMENTAL RESULTS

In this section, we present experimental results of the proposed steganalyzer. First, we describe the image sets used in our experiments. (1) Image Set 1 (IS-1): same as [6], we downloaded 3000 images from the USDA NRCS Photo Gallery[1]. For testing, we resampled each of them to the $1/3$ of the original size (the size of the result images are about 700×500) and converted each image to gray-scale. (2) Image Set 2 (IS-2): JPEG version of Image Set 1 with quality factor 90. Then, we use SVM to train and test. We randomly select $1/4$ of the image set (for cover and stego) to train, and use the rest to test. The procedures are repeated 10 times for cross-validation and the ROC curves are vertically averaged to obtain the mean performance of the scheme. The overall performance of the steganalyzer is then measured by computing the AUC (area under the ROC curve) value. An AUC value close to 1 indicates excellent discrimination, while a value close to 0.5 indicates poor discrimination. For experiments, besides our steganalyzer (noted by S_8), six targeted (noted by $S_1, ..., S_6$) and one blind (noted by S_7) steganalyzers are taken into consideration, thus we consider the following steganalyzers:

(1) S_1: Calibrated HCF COM [6],
(2) S_2: Calibrated Adjacency HCF COM [6],
(3) S_3: ALE based steganalyzer for 1-D histogram [12],
(4) S_4: ALE based steganalyzer for 1-D and 2-D histogram [13],
(5) S_5: the targeted steganalyzer proposed in [15],
(6) S_6: Calibrated HCF COM calculated on difference image, i.e., the steganalyzer G^{ps}_{64} introduced in [8],
(7) S_7: WAM [9],
(8) S_8: our novel steganalyzer.

As shown in Table.1, each column (except the first column) shows the AUC values of the steganalyzers $S_1 \sim S_8$ for a given image set (the first row) at a given embedding rate (the second row). From these experimental results, we can see that: (1) for uncompressed and compressed images, our novel steganalyzer outperforms all the state-of-the-art steganalyzers $S_1 \sim S_7$; (2) especially, for compressed images, the novel steganalyzer performs rather well even when the embedding rate is low, e.g., we arrive an AUC value of 0.95 at an embedding rate 0.1.

Finally, to well illustrate the excellent performance of the novel steganalyzer, we present the comparisons of the ROC curves in Fig.3 for the steganalyzers S_4, S_6, S_7 and S_8. It is obvious that the novel one performs well while some other steganalyzers are ineffective.

5. CONCLUSION

In this paper, we investigated the ALE based steganalyzers proposed in [12, 13], and gave an improved steganalytic method for detecting LSB matching steganography in gray-scale image. By considering the sum and difference image, summing the amplitude of each point of histogram (1-D, 2-D) and employing the calibration technique,

[1] http://photogallery.nrcs.usda.gov

Table 1. Comparison of the AUC values for different steganalyzers on two image sets. Here, E-R means embedding rate.

Image Set	IS-1	IS-1	IS-1	IS-2	IS-2	IS-2
E-R	1.0	0.5	0.2	0.5	0.2	0.1
S_1 [6]	0.760	0.593	0.523	0.596	0.520	0.508
S_2 [6]	0.758	0.604	0.525	0.702	0.538	0.507
S_3 [12]	0.602	0.557	0.522	0.575	0.534	0.515
S_4 [13]	0.827	0.673	0.555	0.922	0.763	0.619
S_5 [15]	0.842	0.693	0.574	0.879	0.719	0.609
S_6 [8]	0.833	0.725	0.584	0.962	0.858	0.714
S_7 [9]	0.792	0.661	0.528	0.990	0.942	0.795
S_8	0.907	0.789	0.639	0.995	0.982	0.950

the novel steganalyzer thus obtained outperforms the old ones. Extensive experimental results have shown that LSB matching can be detected substantially with an embedding rate of 0.1 for compressed images and 0.5 for uncompressed images. The previous work can hardly reach such detection performance. Though LSB matching for compressed cover can be easily detected even for a low embedding rate, the detection for uncompressed cover is still a challenge for steganalysts. For instance, the current steganalyzers can not give an acceptable detection performance of LSB matching for the uncompressed USDA NRCS Photo Gallery even when the embedding rate is 1. Moreover, the proposed steganalyzer can be evidently applied to the additive noise based steganography, then the further experimental results are expected to verify its universality.

6. REFERENCES

[1] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding - A survey," *Proc. of the IEEE*, vol. 87, no. 7, pp. 1062–1078, July 1999.

[2] T. Sharp, "An implementation of key-based digital signal steganography," in *Proc. of the 4th International Workshop on Information Hiding*, 2001, vol. 2137 of *LNCS*, pp. 13–26.

[3] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color and gray-scale images," *IEEE Multimedia*, vol. 8, no. 4, pp. 22–28, October–December 2001.

[4] A. D. Ker, "A general framework for structural steganalysis of LSB replacement," in *Proc. of the 7th International Workshop on Information Hiding*, 2005, vol. 3727 of *LNCS*, pp. 296–311.

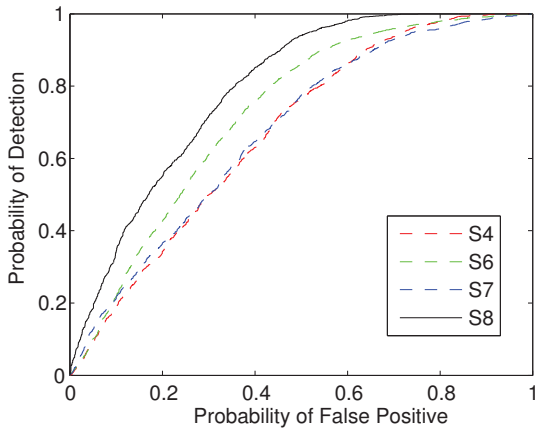
[5] A. D. Ker, "A fusion of maximum likelihood and structural steganalysis," in *Proc. of the 9th International Workshop on Information Hiding*, 2007, vol. 4567 of *LNCS*, pp. 204–219.

[6] A. D. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Process. Lett.*, vol. 12, no. 6, pp. 441–444, June 2005.

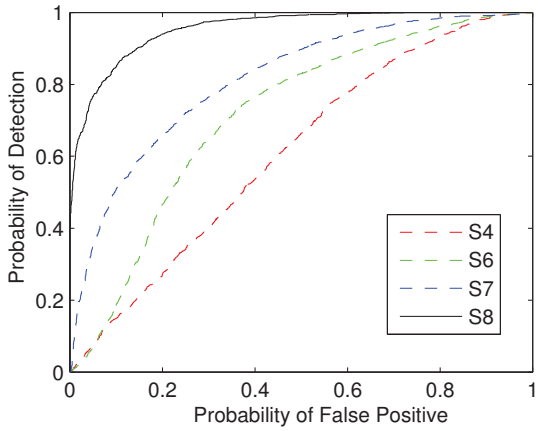
[7] J. J. Harmsen and W. A. Pearlman, "Steganalysis of additive noise modelable information hiding," in *Security and Watermarking of Multimedia Contents V*, 2003, vol. 5020 of *SPIE*, pp. 131–142.

[8] X. Li, T. Zeng, and B. Yang, "Detecting LSB matching by applying calibration technique for difference image," in *Proc. of the 10th Workshop on Multimedia & Security*, 2008, pp. 133–138.

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



(a) IS-1, the embedding rate is 0.5



(b) IS-2, the embedding rate is 0.1

Fig. 3. Comparison of ROC curves for different steganalyzers on two image sets. The different curves stand for, S_4: ALE based steganalyzer for 1-D and 2-D histogram [13] (red, dashed); S_6: Calibrated HCF COM calculated on difference image, i.e., the steganalyzer G_{64}^{ps} introduced in [8] (green, dashed); S_7: WAM [9] (blue, dashed); S_8: our novel steganalyzer (black, solid).

[10] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Inf. Forens. Security*, vol. 1, no. 1, pp. 111–119, March 2006.

[11] H. Gou, A. Swaminathan, and M. Wu, "Noise features for image tampering detection and steganalysis," in *Proc. of the IEEE ICIP*, 2007, vol. 6, pp. 97–100.

[12] J. Zhang, I. J. Cox, and G. Doërr, "Steganalysis for LSB matching in images with high-frequency noise," in *Proc. of the IEEE MMSP*, 2007, pp. 385–388.

[13] G. Cancelli, G. Doërr, I. J. Cox, and M. Barni, "Detection of +-1 LSB steganography based on the amplitude of histogram local extrema," in *Proc. of the IEEE ICIP*, 2008, pp. 1288–1291.

[14] A. D. Ker, "Resampling and the detection of LSB matching in color bitmaps," in *Security, Steganography, and Watermarking of Multimedia Contents VII*, 2005, vol. 5681 of *SPIE*, pp. 1–15.

[15] F. Huang, B. Li, and J. Huang, "Attack LSB matching steganography by counting alteration rate of the number of neighbourhood gray levels," in *Proc. of the IEEE ICIP*, 2007, vol. 1, pp. 401–404.