Security Analysis on Spatial ±1 Steganography for JPEG Decompressed Images

Weiqi Luo, *Member, IEEE*, Yuangen Wang, *Student Member, IEEE*, and Jiwu Huang, *Senior Member, IEEE*

***Abstract*—Although many existing steganalysis works have shown that the spatial ±1 steganography on JPEG pre-compressed images is relatively easier to be detected compared with that on the never-compressed images, most experimental results seem not very convincing since these methods usually assume that the quantization table of the JPEG stegos previously used is known before detection and/or the length of embedded message is fixed. Furthermore, there are just few effective quantitative algorithms for further estimating the spatial modifications. In this letter, we firstly propose an effective method to detect the quantization table from the contaminated digital images which are originally stored as JPEG format based on our recently developed work about JPEG compression error analysis [1], and then we present a quantitative method to reliably estimate the length of spatial modifications in those gray-scale JPEG stegos by using data fitting technology. The extensive experimental results show that our estimators are very effective, and the order of magnitude of prediction error can remain around 10^{-3} measured by the mean absolute difference.**

***Index Terms*—JPEG images, LSB matching, LSB replacement, quantitative steganalysis.**

I. Introduction

STEGANOGRAPHY aims to hide secret messages into a cover media, such as text, audio, image and video, without drawing suspicious. In many applications, the most important requirement for steganography is the security, which means that the stegos should be visually and statistically similar to their corresponding covers. Usually, there are several factors to influence the security performances significantly, such as the hiding scheme, the length of the secret message to be embedded, and the selection of the cover medias. On the other side, steganalysis firstly tries to differentiate stego medias from covers, then further estimate the length of the secret message, and finally extract and decipher the secret message as the ultimate goal. In this letter, we will analyze the security for performing spatial ±1 steganograhy on the JPEG decompressed images. Here, we call the original JPEG decompressed images as JPEG covers, and call the resulting images after spatial-based data hiding as JPEG stegos.

It is well-known that JPEG is one of the most commonly used image format and has been found in many applications. For instance, various digital cameras export this file format, and most popular image editing software packages such as Adobe Photoshop support the operation of JPEG compression. Up to now, there are many JPEG steganography algorithms, such as Jsteg [2], Outguess [3], Model-based steganography [4], JPHide [5], and F5 [6], have been proposed via adjusting the nonzero DCT coefficient of JPEG images. Typically, the embedding capacity of these algorithms is much lower than that of the spatial steganography algorithms since most DCT coefficients of an image would be quantized to zeros after JPEG compression. Furthermore, the objective quality (e.g,. peak signal to noise ratio) of the DCT-based stego images usually is poorer than those spatial-based stegos if embedding the same secret message, especially for the JPEG images compressed with larger quantization steps. Therefore, some users may employ JPEG decompressed images as cover images, and then perform spatial steganogrpahy algorithms on them directly.

Although many existing steganalysis studies such as [7]–[10] have shown that spatial ±1 steganography algorithms such as LSB Matching on JPEG covers are relatively easier to be detected compared with those on the never-compressed images, the experimental results seem not very convincing since they usually assume that the quantization table of the JPEG stego is known before detection and/or the embedding rate is fixed in their experiments. Moreover, when the quality factor is high (e.g. larger than 85) or the embedding rate r is low (e.g. less than 0.1 bit per pixel), the performances of these algorithms would drop significantly and become poor.

A potential and powerful method for detecting and estimating spatial modifications in JPEG stegos was proposed by Fridrich *et al.* [11]. For each unsaturated 8×8 block B in an image, the method tries all possible de-quantized coefficients $DQ_{u,v}$ that are close to integer multiples of $Q_{u,v}$, where $Q_{u,v}$, $0 \leq u, v \leq 7$, is the quantization steps.[1] Then it checks whether or not there is a set of coefficients DQ which satisfies that $B = [IDCT(DQ)]$. If exists, B is compatible with the quantization table Q. Otherwise, B may be modified by some postprocessing. This method is a powerful way to expose JPEG stegos. However, it may not be a good choice for estimating secret message length effectively because we can not conclude that any compatible block must be original un-tampered one [12]. Furthermore, as mentioned in [13], different JPEG decoder may lead to different JPEG decompressed images and thus, it may

Manuscript received August 02, 2010; revised October 23, 2010; accepted October 30, 2010. Date of publication November 09, 2010; date of current version December 02, 2010. This work was supported by NSFC Grants 61003243 and 60633030, 973 Program (2011CB302204), the China Postdoctoral Science Foundation (20080440795) and by China Postdoctoral Science Special Foundation (201003376). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alex C. Kot.

W. Luo is with School of Software, Sun Yat-Sen University, Guangzhou 510275, China, and also with the Guangdong Key Laboratory of Information Security and Technology, Guangzhou 510275, China (e-mail: weiqi.luo@yahoo.com).

Y. Wang and J. Huang are with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510275, China, 510275 (e-mail: isshjw@mail.sysu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2010.2091127

[1]This method also assumes that all 64 quantization steps $Q_{u,v}$ can be extracted correctly from JPEG stegos.

confuse the compatibility analysis significantly. Besides, the time complexity of the method will rapidly increase with increasing quality factor. Taking $QF = 95$ for example, the total number of combinations of all 64 coefficients become too large to handle.

In this letter, we propose an integrated and computationally feasible method for analyzing the security of spatial ± 1 steganography for JPEG decompressed images. The proposed method includes the following two steps. First, we present a simple yet effective way to detect the quantization table from JPEG Stegos. Second, according to the detected quantization table, we present a quantitative steganalysis for further estimating the embedding rate reliably. The experimental results evaluated on thousands natural images show the effectiveness of the proposed method.

The rest of the letter is arranged as follows. Section II presents our methods for detecting quantization table and estimating the secret message length. Section III shows the corresponding experimental results. The concluding remarks are given in Section IV.

II. Proposed Methods

In order to estimate the length of secret message from a JPEG stego, we firstly estimate the quantization table from the JPEG stego, and then we can obtain a reliable quantitative method based on the estimated quantization table. The proposed methods are shown as follows.

A. Detecting Quantization Table From JPEG Stegos

In this subsection, we are going to estimate the quantization table from a questionable image, which is a JPEG stego with an unknown embedding rate ranging from 0 to 1. Here, we assume that the range (rather than the quantization table in most existing works) of the quantization tables is known. This is an reasonable assumption in some scenarios [1], [14]. In our experiments, we assume that quantization tables are standard tables with quality factors from 1 to 95 as widely used in many other steganalysis and forensics works, such as [14]. And the detection method is very similar to our previous method [1]. We define a new similarity measure Sim between two images I and J with the same size of $m \times n$ as follows:

$$Sim(I, J) = \frac{|E|}{m \times n} \tag{1}$$

where $E = \{(x, y)|\ |I(x, y) - J(x, y)| \leq d, 1 \leq x \leq m, 1 \leq y \leq n\}$, and d is set as 1 for ± 1 spatial modifications, e.g. LSB replacement and LSB matching.

For a given JPEG stego JS, we firstly recompress it with all candidate quantization tables, i.e. quality factors ranging from 1 to 95, and obtain the corresponding decompressed images JS_i, and then we calculate $Sim(JS, JS_i)$ respectively. Finally, the estimated quality factor $\hat{QF}$ can be obtained by

$$\hat{QF} = \arg\max_i \left(Sim(JS, JS_i), i = 1, 2, \ldots 95\right). \tag{2}$$

B. Estimating Message Length From JPEG Stegos

After obtaining the quantization table from a JPEG stego, we will further estimate the secret message length. The key issue of the proposed method is to recover the original JPEG cover JC from the JPEG stego JS. To this end, we recompress the JS with the quantization table Q as mentioned above, and the estimated JPEG cover $\hat{JC}$ can be obtained by

$$\hat{JC} = \mathcal{F}(JS, Q) = JC + \mathcal{F}\left(R + N(r), Q\right) \tag{3}$$

where $\mathcal{F}(JS, Q)$ denotes the JPEG recompression operator on JS with the same quantization table Q, R is the rounding error in previous JPEG decompression, and $N(r)$ is an $i.i.d$ additive random noise introduced by ± 1 spatial modifications with an embedding rate r to be estimated.

Formula (3) shows that we can approximate the original JPEG cover JC by recompressing the JS with the same quantization table Q, and the difference between $\mathcal{F}(JS, Q)$ and JC is mainly dependent on $\mathcal{F}(R + N(r), Q)$. Please note that the distribution of the random values $R + N(r)$ has mean 0 and variance $1/12 + r/2$ based on the error analysis in [14]. Therefore, the larger the quantization steps employed in Q, and/or the less secret message r embedded into JC, the more random values $R + N(r)$ will quantized to zeros after JPEG recompression $\mathcal{F}$ with high probability, meaning that we can approximate the JC more accurately.

To estimate the secret message length from JS, we firstly define the observed modification rate r_{om} after JPEG recompression operation as follows

$$r_{om} = \frac{|D|}{m \times n} \tag{4}$$

where $D = \{(x, y)|JS(x, y) \neq \hat{JC}(x, y), 1 \leq x \leq m, 1 \leq y \leq n\}$. Please note that we just consider all the 8×8 unsaturated blocks within an image to remove the effect of truncation error in JPEG decompression.

It is expected that r_{om} will be statistically equal to the half of the true embedding rate r, i.e. $r = 2 \times r_{om}$, if $\hat{JC} = \mathcal{F}(JS, Q) = JC$. However, this formula does not usually hold especially for the quantization table Q with smaller quantization steps, e.g., quality factors larger than 85. The reason is that most $\mathcal{F}(R + N(r), Q)$ is not equal to zero after slight quantization, and thus $\hat{JC} \neq JC$ for most 8×8 blocks within the images in such cases. Therefore, for a given Q, we use K order polynomial to bridge the relationship between the observed modification rate r_{om} and the true embedding rate r as follows, and achieve satisfactory results based on our extensive experiments:

$$\hat{r} = \sum_{j=0}^{K} w_j \times r_{om}^j \tag{5}$$

where $\hat{r}$ is the estimated version of r, and w_j are the coefficients of a polynomial of degree K, $j = 0, 1, 2, \ldots K$.

III. Experimental Results

In our experiments, we randomly select 3000 images from the image databases, including **Corel**, **NRCS**, and **UCID**. Besides that, we take 4000 images which are stored in raw/TIFF format using different cameras. In all, there are 7000 uncompressed images including (but not limited to) landscapes, people, plants, animals and buildings. All the color images are firstly converted into gray-scale images with sizes ranging from 384×512 to 768×512.

TABLE I
ACCURACIES FOR QUANTIZATION TABLE DETECTION FROM JPEG STEGOS WITH DIFFERENT QUANTIZATION TABLES AND/OR EMBEDDING RATES
THE VALUES WITH AN ASTERISK (∗) INDICATE THE BETTER PERFORMANCES BETWEEN THE TWO METHODS

		$r=0$	$r=0.2$	$r=0.4$	$r=0.6$	$r=0.8$	$r=1.0$
QF=95	JPEG WS	98.02	97.99	98.06	98.07	98.07	98.07
	Proposed	99.97 *	99.99 *	99.99 *	99.99 *	99.99 *	99.99 *
QF=90	JPEG WS	89.93	97.14	97.86	97.94	98.04 *	98.07 *
	Proposed	99.97 *	99.99*	99.91 *	98.90 *	94.96	86.37
QF=80	JPEG WS	36.77	82.00	94.30	96.60	97.49	97.70
	Proposed	99.99 *	99.99 *	99.99 *	99.99 *	99.99 *	99.94 *
QF=70	JPEG WS	5.90	9.66	17.36	35.26	63.26	82.51
	Proposed	99.99 *	99.99 *	99.99 *	99.99 *	99.99 *	99.99 *
QF=60	JPEG WS	30.69	50.81	71.30	86.03	93.03	94.74
	Proposed	99.87 *	99.99 *	99.99 *	99.99 *	99.99 *	99.99 *
QF=50	JPEG WS	26.56	48.66	68.95	81.90	89.40	93.27
	Proposed	85.33 *	99.81 *	99.86 *	99.87 *	99.89 *	99.89 *

A. Detecting Quantization Table From JPEG Stegos

In this experiment, all original images are firstly JPEG compressed with different quality factors. And then we embed secret messages with different rates ranging from 0 (JPEG cover) to 1 using spatial ±1 modifications on the JPEG decompressed images to obtain the JPEG stegos. Therefore, we have 7000 JPEG stegos in each case. To show the effectiveness of our method, we compare it with the existing one [14] and show the experimental results in Table I.

It is observed that our method works better in most cases. Note that the detection accuracy of method [14] is rather poor in some cases. For instance, it usually wrongly estimates the JPEG stegos which are previously compressed with a quality factor 70 as a quality factor 38. The reason is that the local minimum of mean square error as mentioned in [14] will usually be presented when the quantization steps in the second quantization table (e.g., $QF = 38$) are around multiples of the corresponding steps in the primary quantization table (e.g. $QF = 70$), which will lead to the false estimation. However, our proposed method still performs well in such cases. Note that our accuracy is relatively lower when $\mathrm{QF} = 50$. The reason is that the quantization tables with $\mathrm{QF} = 49$, 50, and 51 are close enough, and they are easily confused.

B. Estimating Message Length From JPEG Stegos

In the experiment, we firstly randomly divide the 7000 original images into two parts, one for training and the other for testing. Here, we assume that the quantization tables previously used in those JPEG stegos can be estimated correctly using the method as mentioned in previous subsection.

First of all, we need to determine the order K of the polynomial in formula (5). To this end, we firstly JPEG compress original images with a quality factor from 50 to 95, and perform ±1 modification on the JPEG decompressed images with 5 random embedding rates $r \in (0, 1]$. And then we calculate the observed modification rate r_{om} for each JPEG stego according to formula (4). Finally we can get the coefficients w_j for a given order K, where $j = 0, 1, 2, \ldots K$, by minimizing the sum of the squares of the deviations between r_{om} and r for the training data. These coefficients are then used to estimate r for the testing data. We show the mean absolute difference (MAD) as a function of the order K for different quality factors in Fig. 1. It is observed that K in the range of $3 \leq K \leq 9$ gives small values of prediction errors. Therefore, we set $K = 3$ in the following. The fitting curves and corresponding coefficients of the 3rd-order polynomials are shown in Fig. 2 and Table II, respectively.

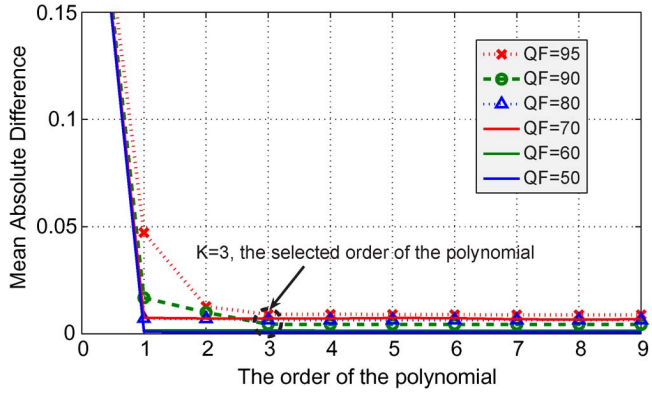

Fig. 1. MAD as a function of the order of the polynomial K.

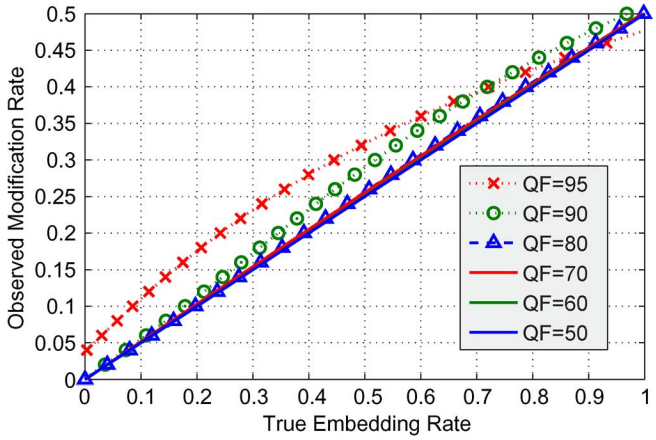

Fig. 2. Fitting curves for different quality factors. The x-coordinate is the true embedding rate r, the y-coordinate is the observed modification rate r_{om}.

As shown in Fig. 2, when the quality factors are high, e.g. larger than 90, the observed modification rate r_{om} is usually greater than 0 even the embedding rate $r = 0$. The reason is that most quantization steps are 1 or 2 in such cases, if performing recompression on the JPEG stego (JS) as shown in formula (3), the item $\mathcal{F}(R + N(r), Q) = \mathcal{F}(R + 0, Q)$ is not always equal to zero due to the slight quantization, which means that we can not recover the JPEG cover JC exactly, namely, $\hat{JC} \neq JC$ for some 8×8 blocks within the image.

TABLE II
THE COEFFICIENTS FOR DIFFERENT QUALITY FACTORS

	QF=95	QF=90	QF=80	QF=70	QF=60	QF=50
w_3	4.2865	3.5654	0.7710	0.0350	0.3033	-0.0023
w_2	-0.2312	-1.8278	-0.4007	0.1243	-0.2039	0.0010
w_1	1.3343	1.9638	2.0042	1.9178	2.0272	2.0000
w_0	-0.0500	-0.0027	0.0001	0.0012	-0.0003	0.0000

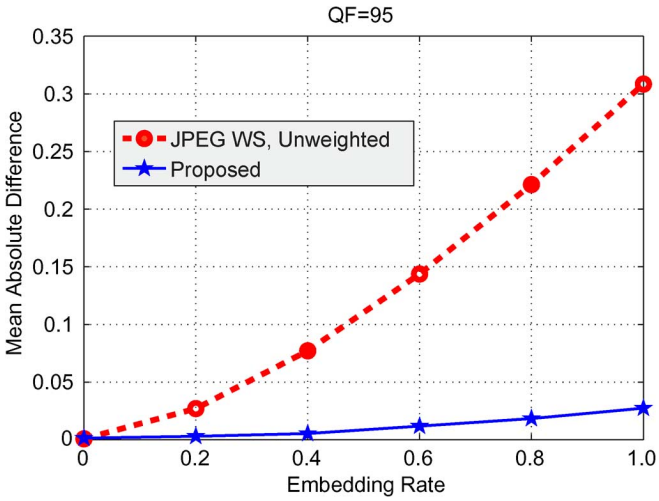

Fig. 3. MAD of the two methods for JPEG stegos with $QF = 95$.

TABLE III
THE AVERAGE MAD OF THE ESTIMATOR [14] AND OUR ESTIMATOR FOR JPEG STEGO USING LSB REPLACEMENT (LSBR)

LSBR	QF=95	QF=90	QF=80	QF=70	QF=60	QF=50
JPEG WS	0.130	0.042	0.008	0.010	0.003	0.007
Proposed	**0.011**	**0.006**	**0.007**	**0.007**	**0.002**	**0.001**

When the quality factors are low, e.g., less than 80 (all the quantization steps are equal to or larger than 4), then the random values $R + N(r)$ in formula (3) will be quantized to zero with very high probability, which means that we obtain $\hat{JC} = JC$ for almost all 8×8 blocks within the image. In such cases, $r \approx 2 \times r_{om}$, like the fitting curves of $QF = 80$, 70, 60, and 50 shown in Fig. 2.

To show the effectiveness of our estimators, we create the new JPEG stegos using LSB replacement with six embedding rates of $r = 0$ (JPEG cover), 0.2, 0.4, 0.6, 0.8, and 1.0. Our estimators with the coefficients in Table II are compared with the best currently-known estimators [14] for LSB replacement in JPEG stego. Fig. 3 illustrates the mean absolute difference as a function of the embedding rates for $QF = 95$. It is observed that the predication error increases with increasing the embedding rates for the method [14] in such case, while our method will remain similar for all the embedding rates. For a given quality factor, Table III shows the average predication error for the six embedding rates. It is also observed that our method outperforms the method [14] in all cases, especially for JPEG stegos with higher quality factors, e.g., 90 and 95.

Table IV shows the experimental results for JPEG stegos using LSB matching. We can see that the method [14] is totally ineffective. However, our method still works and gives similar results (the order of magnitude of prediction error remains 10^{-3} even for the quality factors as high as 95) as shown in Table III for the LSB replacement.

TABLE IV
THE AVERAGE MAD OF THE ESTIMATOR [14] AND OUR ESTIMATOR FOR JPEG STEGO USING LSB MATCHING (LSBM)

LSBM	QF=95	QF=90	QF=80	QF=70	QF=60	QF=50
JPEG WS	0.493	0.492	0.492	0.489	0.491	0.4880
Proposed	**0.008**	**0.004**	**0.006**	**0.007**	**0.002**	**0.001**

IV. CONCLUDING REMARKS

Most existing steganalysis methods for JPEG stegos usually assume the quantization table previously used is known and/or the length of hidden message is fixed, and thus their conclusion about the weakness of JPEG decompressed images used as spatial-based covers seems not very convincing. In this letter, we propose an integrated and computationally feasible method for analyzing the security of spatial ± 1 steganography on JPEG decompressed images. We firstly present an effective method to detect the quantization table form the JPEG Stegos, we then further estimate the secret message length based on the detected quantization table. The experimental results have shown the effectiveness of our method, which indicates that spatial domain embedding in JPEG covers is highly insecure.

REFERENCES

[1] W. Luo, F. Huang, and J. Huang, "JPEG error analysis and its applications to digital image forensics," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 3, pp. 480–491, Sep. 2010.
[2] Jsteg [Online]. Available: http://zooid.org/~paul/crypto/jsteg/
[3] N. Provos, "Defending against statistical steganalysis," in *Proc. 10th Conf. USENIX Security Symp.*, 2001, pp. 323–335.
[4] P. Sallee, "Model based steganography," in *Proc. Int. Workshop on Digital Watermarking*, 2003, pp. 174–188.
[5] S. Hetzl and P. Mutzel, "A graph theoretic approach to steganography," in *Proc. 9th IFIP TC-6 TC-11 Int. Conf., Communications and Multimedia Security*, 2005, vol. 3677, pp. 119–128.
[6] A. Westfeld, "F5—A steganographic algorithm high capacity despite better steganalysis," in *Proc. 4th Int. Workshop on Information Hiding*, 2001, vol. 1768, pp. 289–302.
[7] A. D. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Process. Lett.*, vol. 12, no. 6, pp. 441–444, Jun. 2005.
[8] J. Fridrich, D. Soukal, and M. Goljan, "Maximum likelihood estimation of length of secret message embedded using $\pm$k steganography in spatial domain," in *Proc. SPIE, Security, Steganography, and Watermarking of Multimedia Contents VII*, 2005, vol. 5681, pp. 595–606.
[9] X. Li, T. Zeng, and B. Yang, "Detecting LSB matching by applying calibration technique for difference image," in *Proc. 10th ACM Workshop on Multimedia and security*, Oxford, U.K., 2008, pp. 133–138.
[10] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 2, pp. 215–224, Jun. 2010.
[11] J. Fridrich, M. Goljan, and R. Du, "Steganalysis based on JPEG compatibility," in *Special Session on Theoretical and Practical Issues in Digital Watermarking and Data Hiding, Multimedia Systems and Applications IV*, Denver, CO, Aug. 19–24, 2001, pp. 275–280.
[12] R. E. Newman, I. S. Moskowitz, L. Chang, and M. M. Brahmadesam, "A steganographic embedding undetectable by JPEG compatibility steganalysis," in *Proc. 6th Int. Workshop on Information Hiding*, 2003, vol. 2578, pp. 258–277.
[13] A. D. Ker, "Resampling and the detection of LSB matching in color bitmaps," in *Proc. SPIE on Security, Steganography, and Watermarking of Multimedia Contents VII*, 2005, vol. 5681, pp. 1–15.
[14] R. Bohme, "Weighted stego-image steganalysis for JPEG covers," in *Proc. 10th Int. Workshop on Information Hiding*, 2008, vol. 4567, pp. 204–219.