Steganalysis Based on Double JPEG Compression

Meng Dai
Department of Computer Science & Information engineering
Shanghai Institute of Technology
Shanghai, 200235, China

Yunxiang Liu
Department of Electronic & Communication
East China University of Science & Technology
Shanghai, 200237, China

Abstract*—Due to the stego image double compressed before embedding in some steganographic techniques, double compression could be one of the features of steganography. After JPEG double compressed, the DCT statistic of JPEG image will be changed, which is the focus in this paper. The distributions of DCT coefficients are changed with the double compression quantization tables: q_1, q_2. The difference of the distributions is analyzed in this paper. The different detection methods are proposed based on different q_1, q_2 which are effective proved by the experiments in this paper.

Keywords-steganography*；*steganalysis*；*JPEG double compressing*；*Histogram

I. Introduction

Some steganographic algorithms (F5[1] , OutGuess[2]) always decompress the cover JPEG image into the spatial domain before embedding. During embedding, the image is compressed again, usually with a default quantization matrix (F5 uses default quality factor 80, OutGuess 75). If the quantization matrix used during embedding differs from the original matrix, the resulting stego image is double-compressed.

The statistics of DCT coefficients in double compressed JPEG images may significantly differ from the statistics in single-compressed images. Thus, in some senses, detecting whether the image is double compressed is one of the steganalysis methods.

In paper [3], the histograms of the DCT coefficients are computed. If these histograms contain periodic patterns, then the image is very likely to have been double compressed. But the periodic patterns introduced by double JPEG compression depend on the quality parameters. If the second qualities are less than the fist, we can not find the periodic artifacts. And the histograms of double compressed images satisfy the Laplacian distribution (see the details in section 4).

This paper follows the former research work [4]. In this paper we proposed a new method to detect if the image is double compressed.

II. Procedure of Double Compressed in Steganalysis

JPEG is a standardized image compression procedure proposed by a committee with the same name JPEG (Joint Photographic Experts Committee). To be generally applicable, the JPEG standard specified two compression schemes: a lossless predictive scheme and a lossy scheme based on the Discrete Cosine Transform (DCT). The most popular lossy compression technique is known as the baseline method and encompasses a subset of the DCT-based modes of operation. The encoding of an image involves three basic steps:

- Discrete Cosine Transform (DCT): An image is divided into 8×8 blocks in raster scan order (left to right, top to bottom), shifted from unsigned to signed integers (e.g., from [0; 255] to [_128; 127]), and each block's DCT computed.
- Quantization: The DCT coefficients obtained in the previous step are uniformly quantized, i.e., divided by a quantization step and rounded off to the nearest integer. Since quantization is a non-invertible operation this step represents the main source of information loss.
- Entropy Encoding: This step involves lossless entropy compression that transforms the quantized DCT coefficients into a stream of compressed data. The most frequently used procedure is Huffman coding, although arithmetic coding is also supported.

We say that a JPEG image has been double compressed if the JPEG compression was applied twice, each time with a different quantization matrix and with the same alignment with respect to the grid. Figure 1 shows the procedure of double JPEG compression. In which, q_1 、 q_2 are different quantization matrix, and $q_1 \neq q_2$.

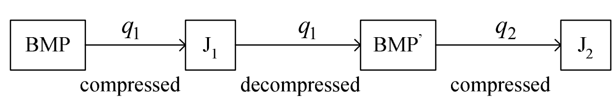


Figure 1. Procedure of double JPEG compression

III. Analysis of Quantization Matrix in JPEG Double Compression

K_1 is defined as AC coefficients that one image firstly compressed with quantization matrix q_1 , K_2 is those AC coefficients which double compressed with quantization matrix q_2 followed by q_1 (without special note, q_1 and

q_2 are proportional to standard quantization matrix in this paper). Thus, K_1, K_2 are satisfied:

$$K_2 = \text{interger\_round}[K_1 \cdot r] \qquad (1)$$

where $r = q_1 / q_2$ and integer_round means rounded to integers.

IV. Detection of Double Compression in Deferent Domain

Giving the K_1, the value of K_2 is determined by variant r. Due to most non-zero AC coefficients be in the low-frequency band, the focus domain is between 0 and 10 in AC histogram is our study objects.

The equation (1) equivalent to

$$K_2 - 0.5 \le r \cdot K_1 < K_2 + 0.5 \qquad (2)$$

If $K_2 = K_1$, which is equivalent to $integer\_round[r \cdot K_1] = K_1$, the following formula must be established:

$$K_1 - 0.5 \le r \cdot K_1 < K_1 + 0.5 \qquad (3)$$

We study three ranges r_1 r_2 and r_3 ($r_1 \in (0, 0.95)$, $r_2 \in [0.95, 1.05)$ $r_3 \in [1.05, 100)$).

In the next sections, we discuss the detecting method in these three ranges respectively.

A. $r=r_2$

When the value of variant r changed within [0.95, 1.05), $K_1 = K_2$ and the histogram has no modification before and after second compressed. If the steganography softwares or algorithms' compressed mechanism is under this condition, the method of detecting whether the JPEG image double compressed can not be used to decide whether this image is stego or not. Detection methods of the higher order statistic [5] and feature based [6] will be effective to these stego images.

Farid [5] proposes a set of sensitive higher-order statistics derived from the wavelet decomposition of the stego-image. Then, he uses Fisher Linear Discrimination analysis to divide the feature vectors into two linear subspaces, one corresponding to stego images, the other to original images.

In paper [6], the detection method is using linear classifier trained on feature vectors corresponding to cover and stego images. In contrast to previous blind approaches, the features are calculated as the difference between a specific macroscopic functional calculated from the stego image and the same functional obtained from a decompressed, cropped, and recompressed stego image. The formulas are built from marginal and joint statistics of DCT coefficients. Experimental results in this paper reveal new facts about current steganographic methods for JPEGs and new design principles for more secure JPEG steganography.

B. $r=r_1$

In the range (0,0.95), the histogram of K_2 (the histogram of a OutGuess stego image which is double compressed with quantization matrix q_1 followed by customized quantization matrix q_2, denoted by '△' in figure 2) is higher and wider than K_1's(the histogram of a jpeg image which is single compressed with quantization matrix q_2, denoted by '+'). But there are no obvious differences in histograms between K_2 and K_1 (see fig.2, r=0.4).For comparison, we also give the histogram of the image singe compressed by q_2 . The maximum $(K_2)_{\max}$ is 56 and the minimum $(K_2)_{\min}$ is -51. The maximum $(K_1)_{\max}$ is 56 and the minimum $(K_1)_{\min}$ is -51. In figure 2, the profiles of the histograms K1 and K2 are almost similar. So, we can not detect whether or not an image was double compressed by analysis the histogram directly.

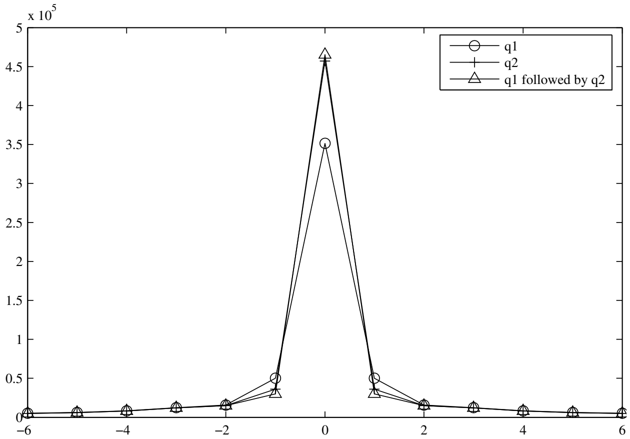


Figure 2. The comparison between single and double compressed image (r=0.4)

1) Features Extraction

Figure 3 shows the procedure of features extraction. In figure.3, the q can be directly extracted from the header of jpeg image. The test image J_{test} is de-compressed to the spatial domain, cropped by 4 pixels in each direction and recompressed with the same quantization table q as J_{test} to obtain J'_{test} . Here, for single compressed image, $q = q_1$, for double compressed, $q = q_2$.

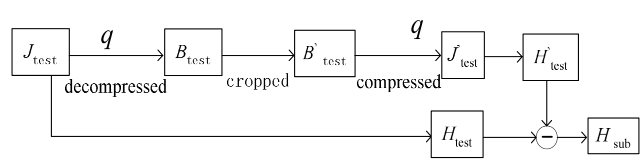


Figure .3 Features extraction diagram

The cropping and recompression should produce a "calibrated" image with most macroscopic features similar to the single compressed image. Thus its DCT coefficients should have approximately the same statistical properties as the cover image [6]. H'_{test} is the histogram of J'_{test} , and H_{test} is J_{test}. The feature H_{sub} is obtained as the absolute value of the difference

$$H_{\text{sub}} = \text{abs}(H_{\text{test}} - H'_{\text{test}}) \qquad (4)$$

and abs is absolute value operator. In range (0, 0.75). H^0 and H^1 have some differences from the single-compressed image(r=1). In range $r \in [0.75, 0.95)$, H^0 and H^1 have no differences. So, in the range $r \in (0, 0.75)$, the feature is

$$f_0 = H^0_{\text{sub}} \Big/ H^1_{\text{sub}} \tag{5}$$

In the range $r \in [0.75, 0.95)$, the feature is

$$f_i = H^i_{\text{sub}} \Big/ H^{i+1}_{\text{sub}}, i = 1, 2... \tag{6}$$

2) Classification and Detection

In experiment, our image database contained approximately 724 images. Training database contains 400 JPEG images, including 200 cover images (BMP format from scanner single compressed with customize quantized factor), 100 F5 stego images, 100 OutGuess images. Testing database contains 324 JPEG images, including 162 cover images and 162 stego images. The result of classification as followed:

TABLE.1 CLASSIFICATION OF IMAGES IN $r \in (0, 0.95)$

	r	Detection (%)
One compressed		89.51
Double compressed	(0,0.75)	82.10
One compressed		88.27
Double compressed	(0.75,0.95)	98.77

C. $r = r_3$

Let $k_1^{d_1}, k_1^{d_2} \in K_1$ and satisfied $k_1^{d_2} = k_1^{d_1} + 1$. After double compressed, the $k_1^{d_1}$ and $k_1^{d_2}$ are modified to $k_2^{d_1}$ and $k_2^{d_2}$, respectively. Namely, $k_2^{d_1} = \text{integer\_round}[k_1^{d_1}]$, $k_2^{d_2} = \text{integer\_round}[k_1^{d_2}]$ .Thus the difference between $k_2^{d_1}$ and $k_2^{d_2}$ is

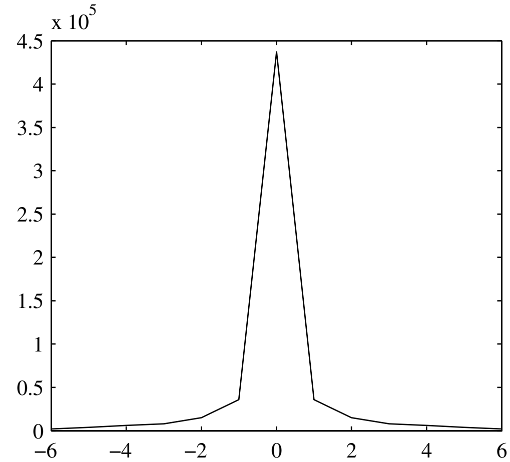
$$\begin{aligned} k_2^{d_2} - k_2^{d_1} &= \text{integer\_round}[r \cdot k_1^{d_2}] - \text{integer\_round}[r \cdot k_1^{d_1}] \\ &= \text{integer\_round}[r \cdot k_1^{d_1} + r] - \text{integer\_round}[r \cdot k_1^{d_1}] \\ &= r \cdot k_1^{d_1} + r + p_1 - r \cdot k_1^{d_1} - p_2 \\ &= r + (p_1 - p_2) \end{aligned}$$

Here, $p_1, p_2 \in [-0.5, 0.5)$, $r \geq 1.05$. So, there muse be exist $k_1^{d_1}$ and $k_1^{d_2}$ satisfy the formula $p_1 - p_2 \geq 0.95$. Namely:

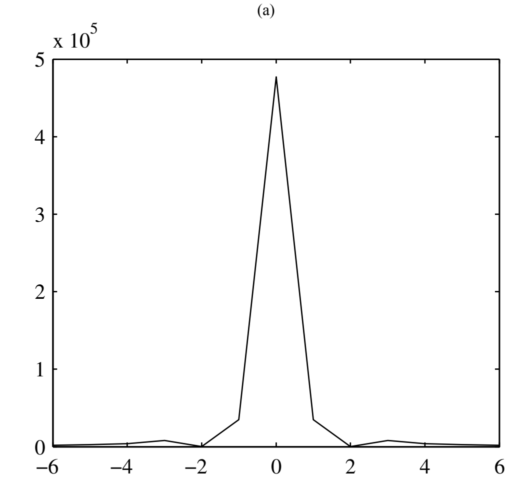
$$k_2^{d_2} - k_2^{d_1} \geq 2 \tag{7}$$

From the above formula, we can draw the conclusion: the number of coefficients whose AC value equals 1 is zero. So, in the range r_3 , zero point will exist in the histogram of the testing stego image. We can search the zero point in the histogram to detect whether the testing image is steganography or not.

Under the condition of $r \in [1.05, 100)$, the following experiment confirms the above conclusion. We test 600 stego images, including 300 embedded messages using Outguess and 300 using F5. The accuracy of detection is 100%. The figure 4 shows the comparison of the histogram between one compressed and double compressed（Here, we only list the condition of r=1.4）



(a)



(b)

Figure 4. Comparison of the histogram between one compressed and double compressed（r=1.4）

(a. one compressed b. double compressed)

V. CONCLUSION

In this paper, we proposed a method to detect the stego image using Outguess or F5 software based on the double-compressed features. In the practical application of the picture, the more conditions are the range $r \in (0, 0.75)$ and $r \in [1.05, 100)$. And in the range of $r \in [0.75, 0.95)$, the r closer 0.95 the more features of testing we used.

As the experiments proved, the method we proposed is effective. And this method is also suit for judging the images whether tempered or not.

ACKNOWLEDGMENT

The work on this paper was supported by Shanghai Education Commission Foundation for Excellent Young High Education Teacher #YYY-07007 and Open Research Foundation of Shanghai Institute of Technology #YJ2007-25, China.

REFERENCE

[1]. Westfeld A. F5-a steganographic algorithm: high capacity despite better steganalysis. Proceeding of 4^{th} International Workshop on Information Hiding. New York: Springer-Verlag ,2001. 289-302.

[2]. N.Provos. Defending Against Statistical Steganalysis. Proceedings of the 10th USENIX Security Symposium, Washington, DC: USENIX press, 2001. 323-335.

[3]. Popescu Alin C, Farid Hany. Statistical Tools for Digital Forensics. Proceeding 6^{th} International Workshop on Information Hiding. Berlin: Springer, 2004.128-147.

[4]. Dai Meng. The Analysis and Detection of Double JPEG Compression. Journal of Image and Graphics, 11(11)2006:1619-1622.

[5]. arid Hany and L.Siwei. Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines. Proceeding 5^{th} International Workshop on Information hiding. Noordwijkerhout: Springer, 2002.340-354

[6]. J. Fridrich. "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. Proceeding 6th Information Hiding Workshop. Berlin: Springer, 2004.67-891.