COUNTERING ANTI-JPEG COMPRESSION FORENSICS

Haodong Li[‡]*, Weiqi Luo*[†]*, Jiwu Huang*[‡]

[†] School of Software, Sun Yat-sen University, Guangzhou, P.R. China, 510006
[‡] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, P.R. China, 510006

ABSTRACT

The quantization artifacts and blocking artifacts are the two significant properties in the JPEG compressed images. Most relative forensic techniques usually use such inherent properties to provide some evidences on how image data is acquired and/or processed. A wise attacker, however, may perform some post-operations to confuse the two artifacts to fool current forensic techniques. Recently, Stamm *et al.* in [1] propose a novel anti-JPEG compression method via adding anti-forensic dither to the DCT coefficients and further reducing the blocking artifacts. In this paper, we found that the dithering operation will inevitably destroy the statistical correlations among the 8×8 intrablock and interblock within an image. In the view of JPEG steganalysis, we employ the transition probability matrix of the DCT coefficients to measure such modifications for identifying the forged images from those original JPEG decompressed images and uncompressed ones. On average, we can obtain a detection accuracy as high as 99% on the image database of UCID [2].

***Index Terms*—** JPEG Compression; Anti-Forensics; JPEG Steganalysis

1. INTRODUCTION

JPEG is one of the most commonly used compression schemes in many practical applications. Therefore JPEG image forensics have attracted increasing attention recently. Typically, there are two significant properties available for forensic analysis. The first and the most obvious property is the blocking artifacts in the spatial domain. Due to the block-based processing in the lossy JPEG compression, the discontinuous pixels usually occur in the boundary between two adjacent 8×8 blocks. Such a blocking signature can served as an evidence of JPEG compression [3], and some tampering operations [4]. Another important property is the quantization artifacts in the DCT frequency domain. During JPEG compression, each DCT frequency component in the 8×8 block is quantized by a quantization step. It will lead to a specific shape of the corresponding DCT histogram. That is, those dequantized coefficients will just cluster in the multiples of the quantization step. Combined with the Laplacian property for the DCT AC components in natural images, it is possible to estimate the quantization parameters in the previous JPEG compression [5, 6], detect double JPEG compression [7], and locate the tampered regions in JPEG composite images [8, 9].

So far, most existing literature mainly focuses on JPEG forensics via analyzing the two artifacts as described above, and only a few works [1, 10, 11] have been reported for the purpose of anti-JPEG compression forensics. Improved by their previous works [10, 11], Stamm *et al.* combine two post-operations to erase the traces left by JPEG compression [1]. The first operation tries to confuse the quantization artifacts by adding dither to the DCT coefficients of a compressed image so that the histograms are similar to the original uncompressed ones. The second one is further to reduce the blocking artifacts via boundary blurring. In [12], Valenzise *et al.* proposed a countering technique by measuring the noisiness of images obtained by re-compressing the forged image at different quality factors, and show that it can effectively identify the forged images with the dithering operation. Based on our experiments, however, the performance of the countering technique [12] will decrease significantly when the blurring operation is applied, please refer to Table 1 and 2 in Section 4 for more details.

Via analysis, we found that the anti-forensic schenes in [1] have to modify the DCT coefficients of the JPEG compressed images in order to preserve the DCT distributions of uncompressed ones. In the view of JPEG steganalysis, we regard such modifications as a process of data hiding. Based on the Markov random process, we then employ the transition probability matrix of the DCT coefficients introduced in [13] to denote the statistical correlations among the 8×8 intrablock and interblock in an image. The experimental results show that our method can detect the operations introduced in [1] with an average accuracy as high as 99%, which outperforms the existing forensics techniques [3, 6, 12] significantly.

The rest of the paper is arranged as follows. Section 2 describes the anti-JPEG compression methods used in [1]. Section 3 proposes the corresponding countering technique. Section 4 demonstrates the experimental results and discussions. Finally, the concluding remarks will be given in Section 5.

This work is supported by the 973 Program (2011CB302204), NSFC (61003243), China Postdoctoral Science Special Foundation (201003376), and the funding of Zhujiang Science & technology (2011J2200091)

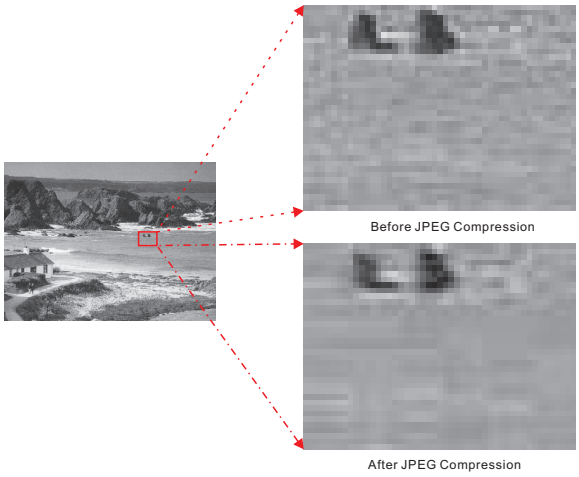
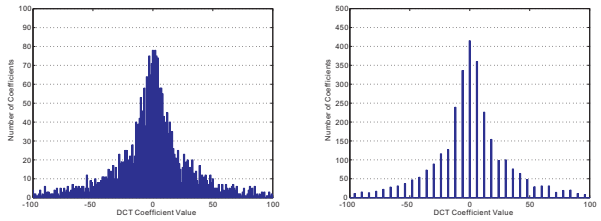


Fig. 1. Illustration of the blocking artifacts



(a) Before JPEG Compression (b) After JPEG Compression

Fig. 2. Illustration of the quantization artifacts: the distributions of DCT coefficients (at the position of (2,2)) for the uncompressed and JPEG compressed images in Fig. 1. Note that the quantization step used here is 6.

2. ANTI-JPEG COMPRESSION FORENSICS

Given a bitmap image that has been JPEG compressed before, anti-JPEG compression forensics aim to erase the statistical traces left by previous JPEG compression so that the current forensic techniques fail to detect those resultant images. As described in the introduction, the two obvious traces introduced by lossy JPEG compression are the blocking artifacts presented in the spatial domain and the quantization artifacts presented in the DCT frequency domain, just as illustrated in Fig. 1 and Fig. 2, respectively.

As illustrated in Fig. 2(b), the dequantized DCT coefficients after JPEG compression will just appear at the multiples of the quantization step. To remove such quantization artifacts, the method [1] firstly estimates the distribution of an image's transform coefficients before compression (*i.e.* the distribution in Fig. 2(a)), and then adds anti-forensic dither to the transform coefficients of a compressed image so that their distribution matches the estimated one. By doing so, those DCT coefficients will spread over the integers rather than just occur at the multiples of the quantization step, which means that the quantization artifacts will be reduced.

In order to remove the blocking artifacts in Fig. 1, the authors in [1] proposed a new deblocking algorithm aiming to resist the existing forensic work [3]. The algorithm employs a median filter on an image, and then applies a low-power zero-mean Gaussian noise to each of its pixel values. The experimental results in [1] show that it can significantly decrease the detection performance of the forensic work [3].

In [1], the authors proposed two anti-JPEG compression methods. The first method (called 'dither' for short) just adds anti-forensic dither to the DCT coefficients for removing quantization artifacts; the second one (call 'dither+blur' for short) combines the 'dither' operation and the technique of removing blocking artifacts with the purpose of concealing both JPEG compression artifacts.

3. COUNTERING TECHNIQUE

As described in Section 2, both anti-forensic methods (*i.e.* 'dither' and 'dither+blur') in [1] have to add an anti-forensic noise to the DCT coefficients of the JPEG compressed images in order to preserve the corresponding distributions of the DCT coefficients in the original uncompressed images at each frequency components (x, y), where $1 \leq x \leq 8, 1 \leq y \leq 8$, respectively. In such a way, almost all DCT coefficients of the JPEG compressed images will be changed. Therefore, the quality of the anti-forensic modified images will decrease inevitably. What is more, other statistical artifacts as follows will be introduced at the same time.

Due to the strong correlation among those adjacent pixel values in natural images, those quantized DCT frequency components at each position (x, y), are not independent of each other, especially for those DCT coefficients within the 8×8 block and the corresponding frequency components in the adjacent blocks, as illustrated in Fig. 3. If modifying the DCT coefficients randomly, their correlations among the 8×8 intrablock and interblock will be destroyed inevitably. By detecting such correlations, it is possible to differentiate the forged images with the two methods in [1] from those uncompressed and/or original JPEG compressed images. Thus, the key issue is to measure the correlations among the adjacent DCT coefficients for a given image.

Furthermore, in the view of JPEG steganalysis, the dithering operation can be regarded as a process of data embedding since both steganography and dithering will modify the DCT coefficients and change the correlations as mentioned above. Therefore, some universal JPEG steganalyzers may be available for exposing the two anti-JPEG compression detection methods in [1]. In this paper, we employ the transition probability matrix introduced in [13] as a measurement of the DCT coefficients correlation.

Note that the questionable images under investigation are bitmap images rather than JPEG images, so those JPEG steganalyzers such as [13] can not be used directly. In

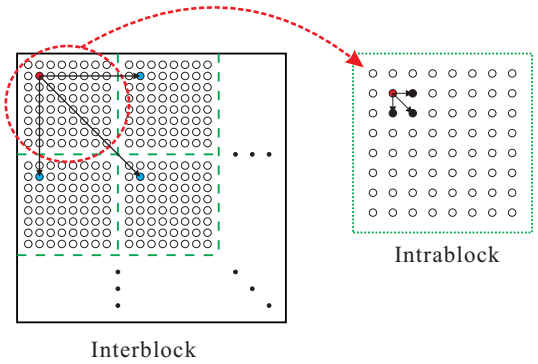


Fig. 3. Illustration of the correlations among the 8×8 Intrablock and Interblock within an image

such a case, we need to divide the bitmap image into non-overlapping 8×8 blocks, then apply discrete cosine transform to each block and obtain the DCT coefficients. Finally those DCT frequency values are rounded to the nearest integers, which are regarded as the JPEG DCT coefficients for the given bitmap image.

Based on the properties of dithering operation used in [1], we just consider the horizontal and vertical correlations among the 8×8 intrablock and interblock coefficients. Firstly, we calculate the differential coefficients arrays among intrablock and interblock in both directions. Then, these four arrays are modeled by Markov random process and the one-step transition probability matrices are computed to characterize them. In order to reduce complexity, we also utilize a threshold technique as referred in [13] and set both the thresholds T_1 and T_2 as 2, that is to say, every transition probability matrix has $(2 \times 2 + 1)^2 = 25$ dimensions. In this way, we can significantly reduce the dimension of the feature vector in [13] from 486 dimensions to 100 dimensions, and still obtain very good results based on the extensive experiments in the following section.

4. EXPERIMENTAL RESULTS

The 1,338 uncompressed images coming from UCID[2] are used in our experiments. These images are firstly JPEG compressed with a given quality factor, then both anti-JPEG compression techniques as described in Section 2 (*i.e.* the 'dither' and 'dither+blur') are applied on the JPEG decompressed image to 'erase' the traces left by JPEG compression. Thus, for each image and each quality factor, we obtain four test images: the original image, the JPEG decompressed image, and two forged counterparts with the two anti-forensic techniques. All images here are stored as bitmaps. We aim to identify the four different kinds of images with the proposed 100-D feature vector as described in Section 3. Besides, three other forensic methods are also involved in our comparative studies, including the edge-based method [3], our previous work [6] for identifying JPEG images, and the total variation (TV-based) method [12].

Table 3. The confusion matrix with the proposed feature. The asterisk ($*$) here denotes the value less than 1%.

	Original	JPEG	Dither	Dither+Blur
Original	**99.58**	0	$*$	$*$
JPEG	$*$	**99.94**	0	0
Dither	$*$	0	**99.76**	0
Dither+Blur	$*$	0	$*$	**99.82**

In the three following experiments, all test images are randomly divided into two parts: 75% is used in the training stage and the remaining 25% is used for testing. The support vector machines (SVM) classifier [14] is used for classification. Please note that all the results are averaged over five times for splitting the testing data and training data alternately.

In the first experiment, we try to identify the uncompressed images from the images using the 'dither' operation. The experimental results are shown in Table 1. It is observed that the proposed feature and TV-based method [12] performs better than the other two methods. On average, the edge-based method [3] still achieves over 83% accuracy.

In the second experiment. we try to differentiate the uncompressed images from the images using both 'dither' and 'blur' operations. The experimental results are shown in Table 2. Compared with the results in Table 1, the performance of the edge-based method [3] will decrease to random guessing (around 50%), since the blocking artifacts have been removed significantly after the 'blur' operation. It is also observed that the average detection accuracy of the existing countering technique [12] will decrease about 20%, which means that the method [12] is not suitable for detecting the images with 'dither+blur' operation. From Table 1 and Table 2, we conclude that the proposed feature performs the best (over 99% in both experiments) among the four forensic techniques. And our previous work [6] can still obtain satisfying results with an average accuracies of 81% and 97%, respectively.

In the third experiment, we further identify the four kinds of images at random quality factors using the proposed feature. A four-kinds SVM classifier is generated by the training instances and its detection performance is evaluated by the testing instances. The confusion matrix in Table 3 shows that we can still obtain an average detection accuracy over 99% in such a case.

5. CONCLUDING REMARKS

In this paper, we propose a countering anti-JPEG compression method based on the correlations among intrablock and

Table 1. Average detection accuracies for identifying uncompressed images and those images after the 'dither' operation. The random QF denotes the quality factor is randomly selected from 50 to 95, and the number with an asterisk (∗) denotes the best performance among the four detection methods.

	QF=95	QF=90	QF=85	QF=80	QF=75	QF=70	QF=65	QF=60	QF=50	Random QF
Edge-based [3]	75.36	76.71	79.13	81.23	83.17	84.58	86.92	89.43	92.43	83.29
Method [6]	80.33	82.54	80.27	72.63	74.64	79.73	85.66	89.07	94.10	73.47
TV-based [12]	96.86	97.46	98.05	98.05	98.50	98.05	97.01	96.11	96.41	97.16
Proposed	**99.25**∗	**99.67**∗	**99.85**∗	**99.85**∗	**99.97**∗	**99.88**∗	**99.94**∗	**99.97**∗	**99.82**∗	**99.73**∗

Table 2. Average detection accuracies for identifying uncompressed images and those images after the 'dither+blur' operation.

	QF=95	QF=90	QF=85	QF=80	QF=75	QF=70	QF=65	QF=60	QF=50	Random QF
Edge-based [3]	51.02	50.75	50.45	50.93	50.87	50.66	51.02	51.29	52.66	50.69
Method [6]	97.10	96.86	97.72	97.75	98.05	98.05	98.17	98.17	97.54	97.78
TV-based [12]	71.26	71.71	75.90	76.50	79.79	76.80	79.94	79.34	78.44	77.10
Proposed	**99.40**∗	**99.40**∗	**99.61**∗	**99.46**∗	**99.70**∗	**99.49**∗	**99.82**∗	**99.76**∗	**99.67**∗	**99.55**∗

interblock within an image. The extensive experimental results have shown the effectiveness of the proposed method.

As a conclusion, the authors suggest that two restrictions should be carefully considered when designing anti-forensic techniques to remove and/or confuse the statistical artifacts left by previous operations to fool current forensic works. Firstly, the anti-forensic operations should not decrease the quality of images significantly. Secondly and most importantly, the operations should not introduce other new artifacts that can be easily detected by other possible forensic techniques.

6. REFERENCES

[1] M.C. Stamm and K.J.R. Liu, "Anti-forensics of digital image compression," *IEEE Transactions on Information Forensics Security*, vol. 6, no. 3, pp. 1050 –1065, Sept. 2011.

[2] G. Schaefer and M. Stich, "Ucid: an uncompressed color image database," in *Proceedings of SPIE: Storage and Retrieval Methods and Applications for Multimedia*, 2004, vol. 5307, pp. 472–480.

[3] Z. Fan and R.L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 230–235, Feb. 2003.

[4] W. Luo, Z. Qu, J. Huang, and G. Qiu, "A novel method for detecting cropped and recompressed image block," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, april 2007, vol. 2, pp. 217–220.

[5] D. Fu, Y. Q. Shi, and W. Su, "A generalized benford's law for JPEG coefficients and its applications in image forensics," in *Proceedings of SPIE: Security, Steganography, and Watermarking of Multimedia*. 2007, vol. 6505, p. 65051L, SPIE.

[6] W. Luo, J. Huang, and G. Qiu, "JPEG error analysis and its applications to digital image forensics," *IEEE Transactions on Information Forensics Security*, vol. 5, no. 3, pp. 480–491, Sept. 2010.

[7] T. Pevny and J. Fridrich, "Estimation of primary quantization matrix for steganalysis of double-compressed JPEG images," in *Proceedings of SPIE: Security, Forensics, Steganography, and Watermarking of Multimedia*. 2008, vol. 6819, p. 681911, SPIE.

[8] J. He, Z. Lin, L. Wang, and X. Tang, "Detecting doctored JPEG images via dct coefficient analysis," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2006, vol. 3953, pp. 423–435.

[9] H. Farid, "Exposing digital forgeries from JPEG ghosts," *IEEE Transactions on Information Forensics Security*, vol. 4, no. 1, pp. 154 –160, Mar. 2009.

[10] M.C. Stamm, S.K. Tjoa, W.S. Lin, and K.J.R. Liu, "Anti-forensics of JPEG compression," in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, Mar. 2010, pp. 1694–1697.

[11] M.C. Stamm, S.K. Tjoa, W.S. Lin, and K.J.R. Liu, "Undetectable image tampering through JPEG compression anti-forensics," in *Proceedings of IEEE International Conference on Image Processing*, Sept. 2010, pp. 2109–2112.

[12] G. Valenzise, V. Nobile, M. Tagliasacchi, and S. Tubaro, "Countering JPEG anti-forensics," in *Proceedings of IEEE International Conference on Image Processing*, Sept. 2011, pp. 1949 –1952.

[13] C. Chen and Y.Q. Shi, "JPEG image steganalysis utilizing both intrablock and interblock correlations," in *Proceedings of IEEE International Symposium on Circuits and Systems*, May 2008, pp. 3029 –3032.

[14] C. Chang and C. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at http://www.csie.ntu.edu.tw/˜cjlin/libsvm.