Identifying Shifted Double JPEG Compression Artifacts for Non-intrusive Digital Image Forensics

Zhenhua Qu[1,3], Weiqi Luo[2,⋆], and Jiwu Huang[1]

[1] School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, China
[2] School of Software, Sun Yat-Sen University, Guangzhou, China
[3] Guangdong Research Institute of China Telecom, Guangzhou, China
qzhua3@gmail.com, weiqi.luo@yahoo.com, isshjw@mail.sysu.edu.cn

Abstract. Non-intrusive digital image forensics (NIDIF) aims at authenticating the validity of digital images utilizing their intrinsic characteristics when the active forensic methods, such as digital watermarking or digital signatures, fail or are not present. The NIDIF for lossy JPEG compressed images are of special importance due to its pervasively use in many applications. Recently, researchers showed that certain types of tampering manipulations can be revealed when JPEG re-compress artifacts (JRCA) is found in a suspicious JPEG image. Up to now, most existing works mainly focus on the detection of doubly JPEG compressed images without block shifting. However, they cannot identify another JRCA – the shifted double JPEG (SD-JPEG) compression artifacts which are commonly present in composite JPEG images. In this paper, the SD-JPEG artifacts are modeled as a noisy 2-D convolutive mixing model. A symmetry verification based method and a first digit histogram based remedy method are proposed to form an integral identification framework. It can reliably detect the SD-JPEG artifacts when a critical state is not reached. The experimental results have shown the effectiveness of the proposed framework.

1 Introduction

Digital images have been pervasively used as evidences in many applications. Their credibility has become increasingly important yet also challenging to establish for the abusive use of modern digital image editing software, such as Photoshop and GIMP. Digital signature and digital watermarking are the typically used techniques to ensure the image content trustworthy. These methods need additional processing at the time of data creation, such as signature generation and watermarking embedding, for facilitating tampering detection at a later time. In many forensic cases, however, the provider himself is the fraudster. Image inspectors cannot depend on these active methods since the side information for detection is not available, for instance, most images on the Web. The demands from practical use, therefore, have urged the researchers to reconsider the image authentication problem from a different perspective.

Recently, many non-intrusive digital image forensics (NIDIF) methods have been proposed by researchers. In NIDIF, the image provider is untrusted and the authentication is performed on the inspector side solely based on the image data. By analyzing the

[⋆] Corresponding author.

S.-M. Hu and R.R. Martin (Eds.): CVM 2012, LNCS 7633, pp. 1–8, 2012.

disturbance or violation of some intrinsic characteristics of original images, the NIDIF can detect tampered or fake digital image content in many forms. Many characteristics have been used for the purpose of image forensics, for example, lighting consistency [7], color filter array (CFA) interpolation [1], fixed pattern noise(FPN) [4], and camera response function (CRF) [11]. However, due to the intricate mathematical nature of these problems, each method works under some preconditions and there is currently no universal solution for all tampering conditions.

JPEG re-compression artifacts (JRCA), one kind of such characteristics can be utilized to detect tampering in JPEG images. When tapering a JPEG image, we have to decode it into spatial domain, and then modify some regions within the image, and finally re-compress it as a JPEG file. So once JRCAs were found in a JPEG image, the image is highly suspected to be modified and is untrustworthy. Currently, researches have focused on addressing the so-called *Double JPEG Compression Problem*, which entails identifying images that suffered from lossy JPEG compression twice without block shifting. Lukáŝ et al. [9] use neural network to estimate the primitive quantization table coefficients. Popescu et al. [12] utilized the periodical artifacts of the re-quantized blockwise discrete cosine transform (BDCT) coefficient histograms. Fu et al. [5] contributed a generalized Benford's Law model of the BDCT coefficients. He et al. [6] use the artifact to identify JPEG image splicing. However, when the fraudster simply shift or crop the JPEG image with several rows or columns before recompression, all the mentioned methods would fail. In this paper, therefore, we will focus on the detection of JPEG recompressed images with block shifting.

This work is originated from our previous studies [10,13]. In this paper, we analyze the SD-JPEG problem, and then proposed an identification framework. We found that the SD-JPEG problem can be modeled as a noisy 2-D convolutive mixing model (CMM) and the solution has a blind source separation (BSS) essence. However, it cannot be well handled with conventional BSS methods, such as independent component analysis (ICA), since the noisy here is often too strong. By utilizing a cyclosymmetry property of the independent value maps (IVM) of ordinary JPEG (Ord-JPEG, i.e. compressed only once) images , we managed to overcome this problem. But it also results a few undetectable conditions (UDC). They need to be further treated with a remedy scheme. With the shifted distance(s-dist), obtained as a by-product of this identification framework, one can also reveal some image manipulation histories, such as cropping, copy-paste or both, of a suspicious JPEG. The effectiveness of this method is evidenced with extensive experimental results.

2 Modeling the SD-JPEG Compression

Illustrated in Fig. 1, the SD-JPEG compression process generates an SD-JPEG image in three steps: 1) Decompress an Ord-JPEG into the spatial domain; 2) Crop/shift the resulting image with $\varDelta x$ columns and or $\varDelta y$ rows; 3) Re-compress it into a JPEG file.

It can be formulated as a 2-D CMM as follows:

$$\widehat{\mathbf{S}}_{m,n} = \sum_{i=0}^{1}\sum_{j=0}^{1} \mathbf{A}_{\varDelta y,i}\mathbf{S}_{m-i,n-j}\mathbf{A}^{T}_{\varDelta x,j} + \hat{\mathbf{E}}_{m,n} \tag{1}$$

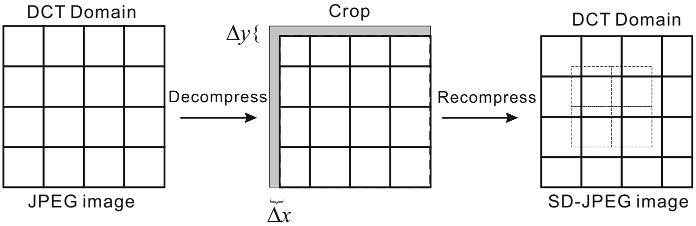


Fig. 1. Gernerating an SD-JPEG image with s-dist=$(\Delta x, \Delta y)$

where $\mathbf{S}_{m,n}$ and $\widehat{\mathbf{S}}_{m,n}$ are the input and output $M \times M$ BDCT coefficient blocks respectively. $\hat{\mathbf{E}}_{m,n}$ is the quantization noise of the secondary JPEG compression.

$$\{\mathbf{A}_{\Delta x,0}, \mathbf{A}_{\Delta x,1}, \mathbf{A}_{\Delta y,0}, \mathbf{A}_{\Delta y,1}\}$$

are called a set of mixing matrices determined by the s-dist $(\Delta x, \Delta y)$ and the BDCT transform matrix. In practice, the multiplications of these matrices are done implicitly by the JPEG compression. Equation (1) indicates that an output block $\widehat{\mathbf{S}}_{m,n}$ is a linear mixture of four input blocks $\{\mathbf{S}_{m,n}, \mathbf{S}_{m,n+1}, \mathbf{S}_{m+1,n}, \mathbf{S}_{m+1,n+1}\}$ overlapped by it. Then the identification of SD-JPEG image can be defined as

Definition 1 (SD-JPEG). *A JPEG image is identified as SD-JPEG if and only if its s-dist is not* $(0, 0)$

As to identify SD-JPEG, one needs to reveal the s-dist or more specifically de-mix the mixture by estimating its mixing matrices. This is the essence of BSS.

3 A Framework for Identifying SD-JPEG Images

The proposed identification framework includes two complementary methods. Firstly, a questionable JPEG image is examined with an ICA-based identification method. If the image is identified as SD-JPEG, the identification process is over and one can further estimate the s-dist to identify the type of image editing. Otherwise, the image is treated with a first digit histogram based method to check out if the s-dist is in one of the three non-trivial UDCs.

3.1 ICA-Based Identification Method

In our previous work [13], we proposed an ICA-based identification method which is able to detect SD-JPEG in most conditions. It works by firstly generating an IVM from the BDCT coefficients of the image, and then calculating a RAVM (relative asymmetric value map) from the IVM's cyclosymmetricity , and finally 13 discriminative features are extracted from the RAVM to train an SVM classifier to automate the identification process. For details of this method please refer to [13].

The above method [13] adopted kurtosis as the objection function. We found that the selection of the objection function is a key issue that influences the identification accuracy of our method. As the fourth order statistics, kurtosis is sensitive to out lies, and thus is not a robust measure of the non-Gaussianity of the distribution. In this work, we further improved the method by adopting a more robust entropy-based objective function. Inspired by the famous InfoMax algorithm [2], we use the entropy objective function here as to more distinctively captures the sparse histogram. The entropy of a BDCT subband s is defined as:

$$J(\mathbf{s}) = \sum_{i} h(i) \log h(i) \tag{2}$$

where $h(i)$ is the i-th bin of the hitogram of BDCT coefficients. Here the discrete entropy is used to approximately calculate the continuous entropy.

We calculate the IVM with this new objective function, and keep the other steps the same as before. With some intrinsic characteristics of natural images, we can prove that the cyclosymmetry property of IVM will hold as well. Due to page limitation, the detailed derivations will not be presented in this paper. This new implementation can significantly improve the performance of the ICA-based method, and the experimental results are given in section 4.

For the same reason, the formulas used for estimating the s-dist would usually work better than the method [13]. By comparing the s-dist estimated from the image regions that correspond to a foreground object under suspicion and its background respectively, we can identify at least three major types of image manipulation methods. For example, if the foreground and the background have the same s-dist other than $(0, 0)$, the image can be generally judged as a "cropping" manipulation, otherwise, when the two s-dists are inconsistent, a "splicing" manipulation is detected. In particular, the condition in which either of the two inconsistent s-dists is not $(0, 0)$ cannot be well handled by most of existing methods.

3.2 Handling the Undetectable Conditions

There are a totally four UDCs as indicated by their s-dist, that is $(0, 0)$, $(0, 4)$, $(4, 0)$ and $(4, 4)$. Because if an Ord-JPEG image is re-compressed with one of these special s-dists, the cyclosymmetry of the IVM will not be violated. It will be identified as Ord-JPEG in the symmetry verification scheme and need special treatments here. The UDCs can be classified into two types. Firstly, when the s-dist is $(0, 0)$, this refers to double JPEG mentioned in Section 1. It can be identified by several existing methods [5,12]. Thus, it is a trivial UDC and will not be discussed here. The other three conditions are non-trivial cases. Each of them occurs with a probability of $1/64 = 1.56\%$ if the shifting was performed randomly. However, they demand more effects than the detectable s-dists. Here we proposed a learning-based approach to detect SD-JPEG in the non-trivial undetectable conditions. By training a classifier with the histograms of Ord-JPEG and SD-JPEG images, we can use the classifier for judgment. However, doing this will result a very high dimension feature set by concatenating the all 63 AC BDCT coefficient histograms each with approximately 200 bins which is also computationally intractable with many modern classifiers.

To bring down the feature dimension while maintaining its discriminative power, we adopt a learning-based approach with the FDH [8] derived from the famous *First Digit Law* or "Benford's Law" [5]. It simply counts the coefficients with their first digits to form a histogram with only nine bins. It has been shown that for a wide variety of "ordinary" distributions, e.g., the exponential family, the resulting FDH can be fitted with a generalized function [5]. On the contrary, an "abnormal" distribution with multiple plumbs resulting by SD-JPEG, will cause irregular rises and falls in the FDH.

For a given JPEG image, we try every UDC s-dist to de-mix it. With each of the demixed AC components, we obtain one FDH. And then these 63 FDHs are concatenated as one feature vector that is fed to a classifier. Non-linear regression based statistical classifiers, such as the RBF kernel SVM [3] used here, can implicitly fitted these histograms to a generalized function. In addition, a binary judgment can be given by adaptively weighting these fitting errors. The benefit of this method is that one does not need to have any explicit knowledge of the distribution of the SD-JPEG BDCT coefficients.

4 Experimental Results

In our experiments, the test image database [10] includes 1000 uncompressed TIFF images taken by a Panasonic DMZ-FZ30 digital camera with indoor and outdoor scenes. The performance in detectable and undetectable conditions is evaluated separately with the two methods mentioned in Section 3.1 and Section 3.2. The image sizes are ranging from 640×480 to 600×1200, and $QF_1, QF_2 \in [60, 65, ..., 95]$. A pair of Ord-JPEG and SD-JPEG is generated from each image with strict consistency in their image contents. In the experiments, the primary quality factor QF_1 and and s-dist are chosen with uniform distributions. The SVM classifier is trained with half of them and tested with the other half. This process is repeated 5 times to obtain the average results.

Figure 2(a) shows the performance of detectable conditions with the symmetry verification scheme. Figure 2 (b–d) shows the classification accuracy for three UDCs. It is observed that for those images with a fixed size, the identification performance of SD-JPEG is mainly depended on the three parameters: the s-dist $(\varDelta x, \varDelta y)$, the primary quality factor QF_1, and the secondary quality factor QF_2. And the detection accuracy will become better with increasing image sizes.

A fraudster will always prefer to "wipe out" the artifacts caused by the primary compression so as not to raise any suspicion. Intuitively, there should be a critical state for QF_2. For example, if the secondary compression is weaker than the primary one, say $QF_2 > QF_1$, more traces will be preserved and it will be fully detectable. In contrast, because the secondary compression is stronger than the former one, it might have completely removed the traces of the first compression and make the re-compressed image unidentifiable in any sense. Therefore, when a QF_1 is specified, we are greatly concerned about whether there exist some QF_2, in which the re-compressed image will be identified as an Ord-JPEG image.

Figure 3 shows how the classification performance would vary with different combination of QF_1 and QF_2 both for detectable and undetectable condition. It is evident that there is a "cliff" where the performances begin to deteriorate quickly. This figure also shows, however, that the critical state is *NOT* at $QF_1 = QF_2$. For example, in Fig. 3(a),

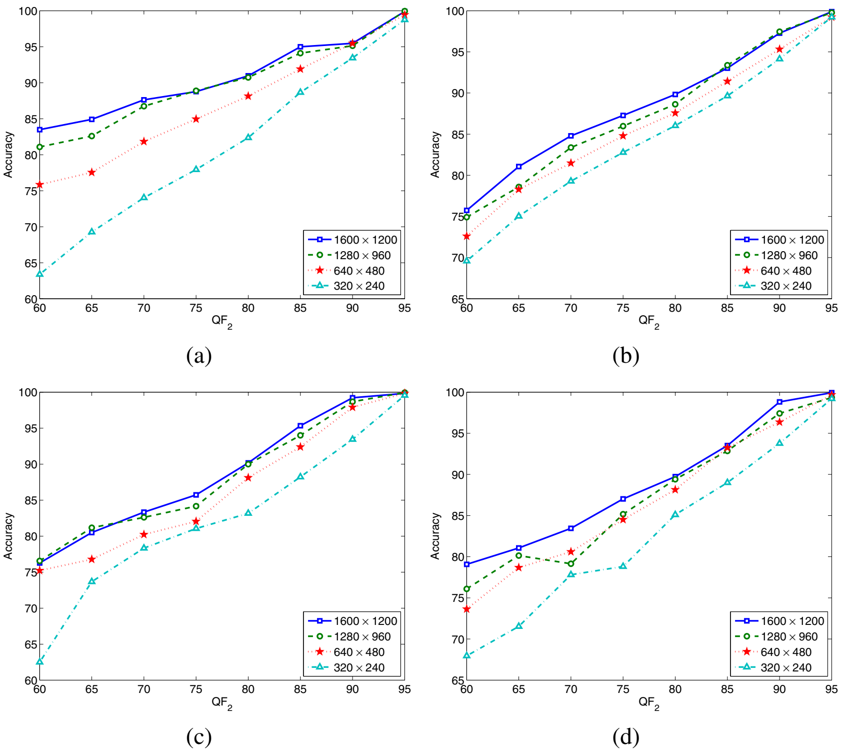


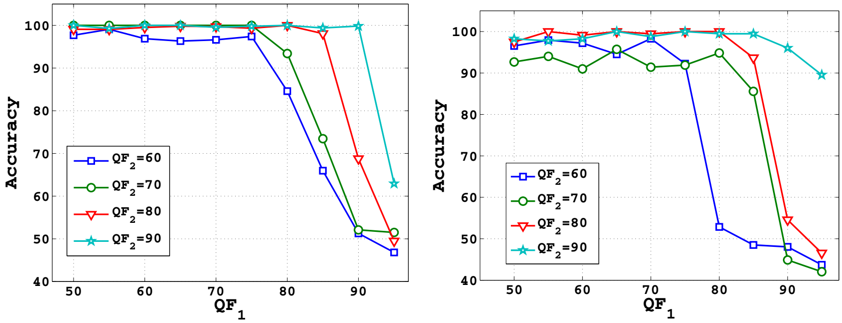
Fig. 2. The classification performance in detectable and three non-trivial UDCs. (a) symmetric verification method (b–d) FDH method with three UDCs $(4,0)$, $(0,4)$, $(4,4)$.

when $QF_2 = 60$ the performance does not begin to deteriorate until $QF_1 = 75$. In addition, this bias will be smaller with a larger QF_2. A similar condition is also observed in Fig. 3(b). This indicates that an SD-JPEG image is still identifiable even if the secondary compression is slightly stronger than the primary one.

We also measured the goodness of the estimation of s-dist. It is measured by the *average shifted distance error* (ASDE).

$$ASDE = mean\left(\sqrt{(\hat{x} - \Delta x)^2 + (\hat{y} - \Delta y)^2}\right) \tag{3}$$

where $(\hat{x}, \hat{y})$ and $(\Delta x, \Delta y)$ are the estimated and ground truth s-dist, respectively. Table 1(a) and Table 1(b) show the ASDE of the condition when $QF_1 \leq QF_2$ and $QF_1 > QF_2$ respectively. As shown, when $QF_1 \leq QF_2$ nearly all the s-dists are correctly estimated. This suggests that the s-dist can be revealed without error when the critical state is not reached. However, when $QF_1 > QF_2$, and there would be estimation errors. The estimation errors are relatively large in the location near the four UDCs.



(a)

(b)

Fig. 3. The classification accuracy curves decline suddenly with the increase of QF_1 at a specified QF_2. (a) ICA-based method for detectable condition. (b) FDH-based method for the UDC with s-dist=$(0, 4)$.

Table 1. (a)The ASDE of $QF_1 \in [50 \dots 80]$ and $QF_2 = 75$. (b)The ASDE of $QF_1 \in [80 \dots 95]$ and $QF_2 = 75$.

(a)

Δx Δy	0	1	2	3	4	5	6	7
0	NA	0.00	0.00	0.03	NA	0.01	0.00	0.00
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	NA	0.00	0.00	0.00	NA	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(b)

Δx Δy	0	1	2	3	4	5	6	7
0	NA	1.63	2.00	2.52	NA	1.85	1.45	1.50
1	1.42	1.16	1.43	1.93	2.01	1.52	1.17	0.91
2	1.60	1.48	1.50	2.08	1.91	1.51	1.34	1.06
3	2.55	1.70	1.90	2.24	2.82	1.77	1.75	1.74
4	NA	2.34	2.41	2.92	NA	2.50	2.22	2.24
5	2.70	2.04	2.02	2.33	2.88	2.21	1.97	1.64
6	1.90	1.54	1.57	2.50	2.22	1.61	1.38	1.24
7	1.50	1.18	1.37	2.09	1.94	1.56	1.18	0.87

5 Conclusion

In this paper, we proposed a method to identify the shifted and re-compressed JPEG image blocks. We extend our previous work [13] in two aspects. Firstly, an entropy-based function is used to more distinctively capture the abnormal BDCT coefficient histograms of SD-JPEG image. Secondly, an FDH based method is provided as a remedy for the three non-trivial UDCs that were undetectable in our previous approach.

One limitation of the proposed method is that the exhaustive search scheme can only deal with pixel-wise shifting. However in some image editing software, the shifting can be conducted in sub-pixel level which would result in a continuous space for searching the de-mixing matrices. Based on the same reason, rotated or scaled JPEG re-compression is also not detectable here. It would be interesting but challenging to determine whether there exist continuous optimization based methods that can learn these parameters without the restriction of limit number of s-dist. These topics will be addressed in future works.

Acknowledgment. This work is supported in part by the 973 Program (2011CB302204), NSFC (61003243, 61272191), and the funding of Zhujiang Science and technology (2011J2200091).

References

1. Bayram, S., Sencar, H., Memon, N.: Identifying digital cameras using cfa interpolation. In: Advances in Digital Forensics II, vol. 222, pp. 289–299 (2006)
2. Bell, J.A., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159 (1995)
3. Chang, C.C., Lin, C.J.: LIBSVM:a library for support vector machines (2001), http://www.csie.ntu.edu.tw/~cjlin/libsvm
4. Chen, M., Fridrich, J., Lukáš, J., Goljan, M.: Imaging Sensor Noise as Digital X-Ray for Revealing Forgeries. In: Furon, T., Cayre, F., Doërr, G., Bas, P. (eds.) IH 2007. LNCS, vol. 4567, pp. 342–358. Springer, Heidelberg (2008)
5. Fu, D.D., Shi, Y.Q., Su, W.: A generalized benford's law for jpeg coefficients and its applications in image forensics - art. no. 65051l. In: Security, Steganography, and Watermarking of Multimedia Contents IX, vol. 6505, p. L5051 (2007)
6. He, J., Lin, Z., Wang, L., Tang, X.: Detecting Doctored JPEG Images Via DCT Coefficient Analysis. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part III. LNCS, vol. 3953, pp. 423–435. Springer, Heidelberg (2006)
7. Johnson, M.K., Farid, H.: Exposing digital forgeries in complex lighting environments. IEEE Trans. Inf. Forensics Security 2(3), 450–461 (2007)
8. Li, B., Shi, Y.Q., Huang, J.W.: Detecting doubly compressed jpeg images by using mode based first digit features. In: IEEE Workshop on Multimedia Signal Processing, pp. 730–735 (2008)
9. Lukas, J., Fridrich, J.: Estimation of primary quantization matrix in double compressed jpeg images. In: Proc. of DFRWS, Cleveland, OH, USA (2003)
10. Luo, W., Qu, Z., Huang, J., Qiu, G.: A novel method for detecting cropped and recompressed image block. In: IEEE Int. Conf. on Acoustics Speech and Signal Processing, April 15-20, vol. 2, pp. II-217–II-220 (2007)
11. Ng, T.T., Chang, S.F., Tsui, M.P.: Using geometry invariants for camera response function estimation. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, June 17-22, pp. 1–8 (2007)
12. Popescu, A.: Statistical Tools for Digital Image Forensics. Ph.D. thesis, Department of Computer Science,Dartmouth College (2005)
13. Qu, Z., Luo, W., Huang, J.: A convolutive mixing model for shifted double jpeg compression with application to passive image authentication. In: IEEE Int. Conf. on Acoustics Speech and Signal Processing, March 31-April 4, pp. 1661–1664 (2008)