A Method for Detecting JPEG Anti-forensics

Dinesh Bhardwaj(✉), Chothmal Kumawat, and Vinod Pankajakshan

Department of Electronics and Communication Engineering,
Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India
{dinesdec,cmeca.dec2015,vinodfec}@iitr.ac.in

Abstract. In this paper, a new approach is proposed for the detection of JPEG anti-forensic operations. It is based on the fact that when a JPEG anti-forensic operation is applied, the values of DCT coefficients are changed. This change decreases, especially in high frequency subbands, if we apply anti-forensic operation again. Hence, we propose to calculate a normalized difference between absolute values of DCT coefficients in 28 high frequency AC-subbands of the test image and its anti-forensically modified version. Based on this normalized feature, it is possible to differentiate between uncompressed and anti-forensically modified images. Experimental results show the effectiveness of the proposed method.

1 Introduction

The technological advancements have made digital cameras easily available and the increasing use of internet as a communication media made the digital images an easy way of conveying the visual information. But the digital images can be manipulated easily with the help of photo-editing softwares, even by non-professional users. These manipulations are known as image forgery in which a part of a realistic image is tampered. These forgeries have made the authenticity of digital images doubtful or we can say that the old saying "to see is to believe", might not be true in the case of digital images. In most of the cases, these forged images are not harmful, but if the same are used for malicious purposes such as evidences in the courtroom, or with the celebrities images, or as material for propaganda may defame an individual or organization. The doctored images/videos may create a negative environment or influence the public opinion. Many forensic techniques that can detect forgeries have been proposed in recent years. These forensic techniques detect the traces left by signal processing operations associated with forgery or fingerprints of acquisition devices [1].

Among the forensic techniques, JPEG forensics have received considerable research attention. This is due to the fact that the JPEG is the most widely used image compression standard. It introduces blocking artifacts across the boundaries of 8×8 sub-image blocks in spatial domain and DCT histogram artifacts in form of clustering of DCT coefficient around the integer multiple of quantization

R. Rameshan et al. (Eds.): NCVPRIPG 2017, CCIS 841, pp. 190–197, 2018.
https://doi.org/10.1007/978-981-13-0020-2_17

step size in frequency domain. Furthermore, any forgery of a JPEG compressed image results in multiple compression artifacts in forged image. Based on these artifacts, various forensic techniques have been proposed in the literature to detect doctored images [2–4].

Most of the forensic techniques do not account for the possibility of anti-forensics. An adversary, who is familiar with the signal processing operations involved in the forgery, may develop an anti-forensic technique by removing the traces left by these operations. This results in an undetectable image forgery which can deceive forensic detectors. It is necessary for the forensic researchers to study any loophole in the forensic techniques, so that the undetectable image forgeries due to anti-forensic operation can be detected. In [5], Stamm *et al.* introduced the concept of JPEG anti-forensics to remove the JPEG histogram artifacts by adding an anti-forensic dither in the DCT-domain. In [6], the authors extended this work by removing blocking artifacts in addition to the dithering operation. Fan *et al.* [7] proposed another approach for concealing both the blocking and the DCT quantization artifacts with an improved trade-off between the forensic undetectability and perceptual quality of the anti-forensically modified image.

In response to these anti-forensics techniques, there are a few algorithms proposed which can detect the presence of anti-forensic operations. Valenzise *et al.* [8] observed that the addition of dithering noise in the DCT coefficients to remove histogram artifacts introduces grainy noise in the spatial domain. They proposed a technique to detect the presence of anti-forensic dither in a test image by computing the total variation (TV) in different recompressed versions (with different quality factors) of the test image. Lai and Böhme [9] proposed a calibrated feature which is based on the variance of DCT coefficients in high frequency subbands of test image and its calibrated version which is obtained by cropping the test image. Both these methods successfully detect the anti-forensic operation [5]. A machine learning based approach is proposed by Haodong Li *et al.* [10] for detecting various image operations as well as anti-forensic operations. Apart from the technique in [10], which is a computationally expensive machine learning based technique, there exists no technique which can detect the anti-forensic operations in [6,7]. This paper proposes a technique for differentiating uncompressed images from anti-forensically modified images. It is based on the change in the DCT coefficients due to anti-forensic operation.

The rest of this paper is organized as follows. Section 2 briefly discusses different anti-forensic methods. The proposed method is explained in Sect. 3. Section 4 presents the experimental results and performance analysis, and finally Sect. 5 concludes the paper.

2 JPEG Anti-forensic Methods

2.1 Stamm *et al.*'s Anti-forensic Method [6]

The DCT coefficients of an uncompressed image for a given AC subband follow the Laplacian distribution and JPEG compression converts this into discrete

version [11]. In [6], for removing the JPEG histogram artifacts, a specially designed noise (anti-forensic dither) is added in the DCT subbands in such a way that the AC subbands follow the Laplacian distribution. The Laplacian parameter is estimated from the DCT coefficients of the compressed image for each AC subband by using maximum likelihood estimation (MLE). This dithering process adds grainy noise in the spatial domain which degrades the visual quality and disturbs the correlation of DCT coefficients. Further, JPEG blocking artifacts are removed by applying median filtering followed by the addition of zero mean Gaussian noise in spatial domain. The window size of the median filter and the variance of the noise can be adjusted according to quality factor of the JPEG image.

2.2 Fan *et al.*'s Anti-forensic Method [7]

The main disadvantage of Stamm's [6] anti-forensic method is that it degrades the quality of anti-forensically modified image. Fan *et al.*'s anti-forensic algorithm is a four step technique capable of fooling most of the forensic detectors based on JPEG compression artifacts. This approach achieves better visual quality as compared to [6]. In the first step, a deblocking is performed by using constrained TV-based minimization problem which is solved by projected subgradient method. It removes blocking artifacts and partially fills the gaps in the DCT histogram. In the second step, an adaptive local dithering signal model is used to fill the remaining gaps in DCT histogram. It is pointed out that in case of real data, the Laplacian model works well for the DCT coefficients quantized to zero whereas uniform model fits better for the other quantized DCT coefficients. As compared to [6], the Laplacian parameter for each quantization bin is estimated on the basis of weighted least-square fitting on the first round TV-based deblocked image. A second round TV based deblocking is performed to remove the small artifacts introduced during the histogram smoothing. Finally, decalibration is done to fool the calibrated feature based detector [9].

3 Proposed Detector

We propose a technique that can differentiate between uncompressed images and anti-forensically modified images. When an uncompressed image is JPEG compressed with a certain quality factor, the values of DCT coefficients will be clustered around the integer multiples of the corresponding quantization step size. As a result, gaps are introduced in the histogram of DCT coefficients in each subband. The application of anti-forensic operation on the JPEG compressed image tries to restore the DCT histogram similar to that of the uncompressed image. Figure 1(a) shows the histogram of DCT coefficients in a particular subband of an uncompressed image. Figure 1(b) shows the histogram of the DCT coefficients in the same subband after the JPEG compressed image is anti-forensically modified using the method proposed in [7]. Though the gaps in the DCT histogram are eliminated by the anti-forensic operation, there is a noticeable difference in the

DCT histograms of the uncompressed (Fig. 1(a)) and that of the anti-forensically modified image (Fig. 1(b)). If we compress the anti-forensically modified image and then apply the anti-forensic operation, there is difference between the histograms of the DCT coefficients of the resulting image, which has undergone the anti-forensic operation twice, and that of the image which has undergone anti-forensic operation once. Figure 1(c) shows the histogram of the DCT coefficients in the same subband for the image which had undergone anti-forensic operation twice. Figure 1(d) and (e) shows the absolute difference of the DCT histograms in Figs. 1(a), (b) and (b), (c) respectively. It can be observed form Figs. 1(d) and (e) that the change in the DCT histograms is more when the image is anti-forensically modified for the first time as compared to the change when an image is anti-forensically modified for the second time. It is also observed that this change in DCT histograms is more pronounced in the high frequency DCT subbands. Based on this observation, we propose a detector that can differentiate anti-forensically modified images from uncompressed images. The block diagram of the proposed detector is shown in Fig. 2. Let X be the given test image and X_d denotes its block-DCT transform. The test image is JPEG compressed with a quality factor Q_r and modified by the anti-forensic operation [7]. The resulting image is denoted by X_1 and its block-DCT transform is denoted by X_{1d}. For each of the 28 high frequency subbands (in zig-zag scan order), the

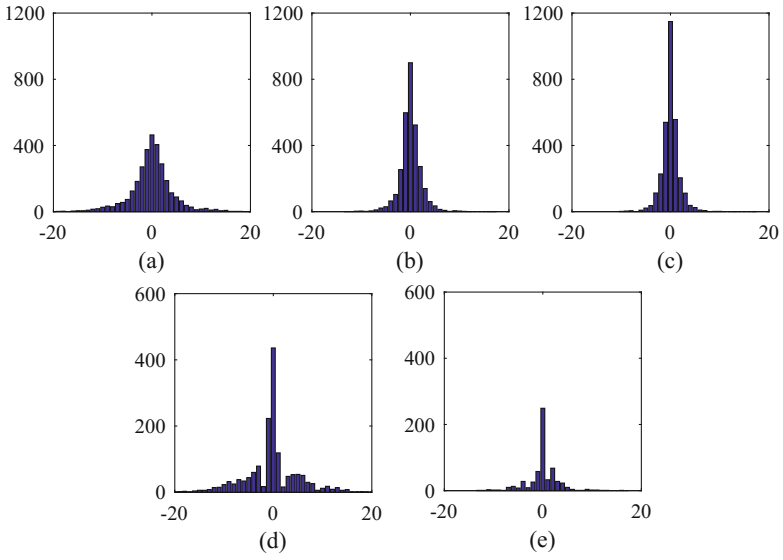

Fig. 1. DCT histograms of 51-th (in zig-zag scan order) AC subband for (a) uncompressed image. (b) image obtained by applying anti-forensic operation [7] on the compressed version of (a). (c) image obtained by applying anti-forensic operation [7] on the compressed version of (b). (d) the absolute difference of DCT histograms of (a) and (b). (e) the absolute difference of DCT histograms of (b) and (c).

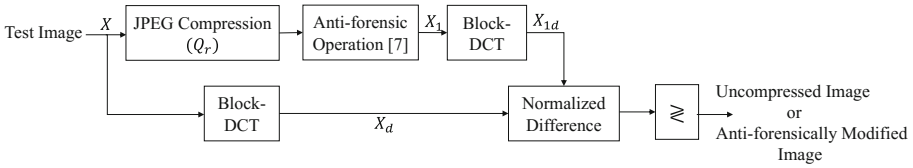

Fig. 2. Block diagram of proposed method

sum of absolute values of the DCT coefficients is computed for both X_d and X_{1d}, denoted by S_d^i and S_{1d}^i, respectively. Here, i is the subband index which varies from 37 to 64. Then the normalized difference (D) is calculated as

$$D = \sum_i \left| \frac{S_d^i - S_{1d}^i}{S_{1d}^i} \right|. \tag{1}$$

The given test image may be uncompressed or anti-forensically modified image. Hence, if the test image is uncompressed, the JPEG compression and anti-forensic operations are applied only once and the difference is calculated between DCT coefficients of uncompressed image and anti-forensically modified image. On the other hand, if the given test image had already undergone an anti-forensic operation, it is JPEG compressed and modified again by an anti-forensic operation. As a result, the difference is calculated between the given anti-forensically modified image and its anti-forensic version. Hence, the value of the proposed normalized difference (D) is small for anti-forensically modified image as compared to that of uncompressed image.

Now, the question arises, at what quality factor, the test image should be compressed and anti-forensically modified. If the test image is anti-forensically modified image, it had already been JPEG compressed by a certain quality factor (Q) and modified by an anti-forensic operation. It is observed that the normalized difference (D) will be less if we compress and anti-forensically modify the image at same quality factor i.e. $Q_r = Q$. But, the value of quality factor Q is not available to the forensic detector. Hence, we calculate a normalized feature (D) using a set of quality factors (Q_r) and the average normalized feature value $\overline{D}$ is used. In the experiments, we have used $Q_r = \{40, 50, ..., 80\}$. Finally, by applying a suitable threshold on $\overline{D}$, it is possible to differentiate between uncompressed and anti-forensically modified images.

4 Experimental Results

The performance of the proposed detector is evaluated on three standard datasets: UCID [12], Dresden [13] and Boss Base 1.01 [14]. There are total of 3691 images in which 1338 images are from UCID database, 353 images are from Dresden database and remaining 2000 images are taken from the Boss Base 1.01 database. Each image is converted into gray scale and are then JPEG

compressed with 5 different quality factors: $Q = \{40, 50, ..., 80\}$. Each of these JPEG compressed images is anti-forensically modified by [7] which form a set of anti-forensic images. The average normalized difference ($\overline{D}$) feature is calculated for both uncompressed and anti-forensically modified images. Figure 3 shows the scatter plot of the average normalized difference for uncompressed images and anti-forensically modified images for quality factor $Q = 60$. It is observed that with a suitable threshold, we can effectively classify uncompressed images from anti-forensically modified images.

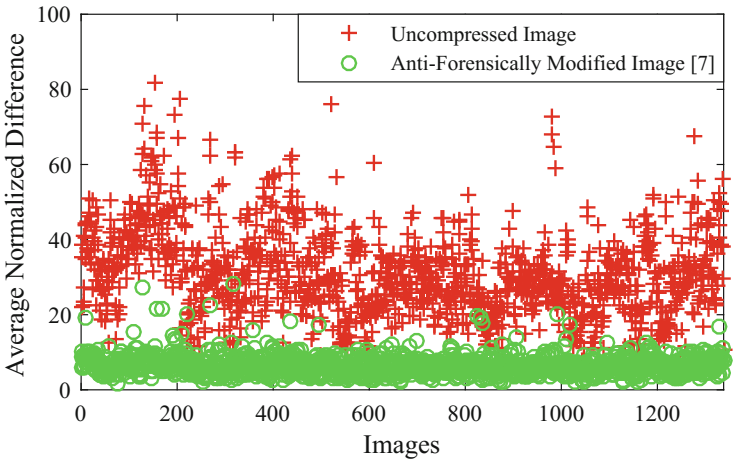

Fig. 3. Scatter plot of proposed feature at $Q = 60$ (UCID database)

The receiver operating characteristics (ROC) is used as the performance measure. The ROC curve gives the classification ability of the proposed feature: the closer the apex of the curve towards the top left corner, the higher is the classification ability i.e. high true-positive rate (TPR) and low false-positive rate (FPR). Figure 4 shows the ROC curves for the proposed detection method for different values of quality factor i.e. $Q = 40, 50, ..., 80$. We can observe that the performance degrades as the quality factor increases because higher the quality factor, lower is the quantization step size and lesser is the amount of noise added by the anti-forensic operation. We also tested proposed feature for the detection of anti-forensic operation [6]. We calculated the average normalized difference ($\overline{D}$) for uncompressed and anti-forensically modified images by [6] for three different quality factors i.e. $Q = 40, 60, 80$ and the corresponding ROC curves are shown in Fig. 5. The area under curve (AUC) for proposed detector at different quality factors are reported in Table 1. As expected, the AUC decreases with increase in the value of quality factor for both the anti-forensic techniques. These results show that the proposed detector can detect both the anti-forensic operations with a good accuracy.

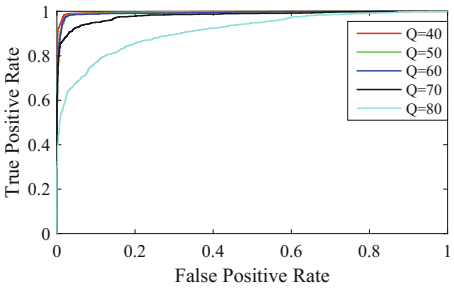

Fig. 4. Receiver operating characteristic (ROC) curve for detecting [7] (UCID database)

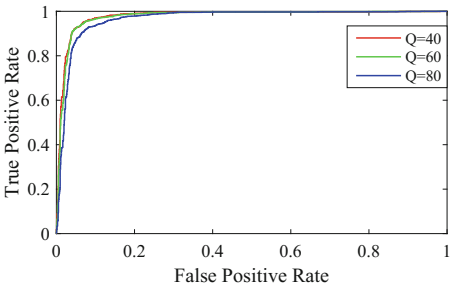

Fig. 5. Receiver operating characteristic (ROC) curve for detecting [6] (UCID database)

Table 1. Performance of the proposed method in terms of AUC

Dataset	[7]			[6]		
	Q = 40	Q = 60	Q = 80	Q = 40	Q = 60	Q = 80
UCID [12]	0.99	0.99	0.91	0.98	0.98	0.97
Dresden [13]	0.98	0.96	0.93	0.99	0.99	0.99
Boss base 1.01 [14]	0.98	0.96	0.90	0.98	0.98	0.97

5 Conclusions

This paper proposed a technique for differentiating between uncompressed images and anti-forensically modified JPEG images. The technique is based on the change in values of DCT coefficients by an anti-forensic operation. A normalized difference feature is calculated for measuring these changes. Experimental results show that the proposed method effectively classifies uncompressed images from anti-forensically modified images. Future research includes the detection of quality factor of the anti-forensically modified JPEG image.

References

1. Farid, H.: Image forgery detection. IEEE Sig. Process. Mag. **26**(2), 16–25 (2009)
2. Huang, F., Huang, J., Shi, Y.Q.: Detecting double JPEG compression with the same quantization matrix. IEEE Trans. Inf. Forensics Secur. **5**(4), 848–856 (2010)
3. Chen, Y.L., Hsu, C.T.: Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection. IEEE Trans. Inf. Forensics Secur. **6**(2), 396–406 (2011)
4. Bianchi, T., Piva, A.: Image forgery localization via block-grained analysis of JPEG artifacts. IEEE Trans. Inf. Forensics Secur. **7**(3), 1003–1017 (2012)
5. Stamm, M.C., Tjoa, S.K., Lin, W.S., Liu, K.J.R.: Anti-forensics of JPEG compression. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1694–1697, March 2010

6. Stamm, M.C., Liu, K.J.R.: Anti-forensics of digital image compression. IEEE Trans. Inf. Forensics Secur. **6**(3), 1050–1065 (2011)
7. Fan, W., Wang, K., Cayre, F., Xiong, Z.: JPEG anti-forensics with improved tradeoff between forensic undetectability and image quality. IEEE Trans. Inf. Forensics Secur. **9**(8), 1211–1226 (2014)
8. Valenzise, G., Tagliasacchi, M., Tubaro, S.: Revealing the traces of JPEG compression anti-forensics. IEEE Trans. Inf. Forensics Secur. **8**(2), 335–349 (2013)
9. Lai, S., Böhme, R.: Countering counter-forensics: the case of JPEG compression. In: Proceedings of International Conference on Information Hiding, pp. 285–298 (2011)
10. Li, H., Luo, W., Qiu, X., Huang, J.: Identification of various image operations using residual-based features. IEEE Trans. Circ. Syst. Video Technol. **PP**(99), 1 (2016)
11. Lam, E.Y., Goodman, J.W.: A mathematical analysis of the DCT coefficient distributions for images. IEEE Trans. Image Process. **9**(10), 1661–1666 (2000)
12. Schaefer, G., Stich, M.: UCID - an uncompressed colour image database. In: Proceedings of SPIE, pp. 472–480, March 2004
13. Gloe, T., Bhme, R.: The 'Dresden Image Database' for benchmarking digital image forensics. In: Proceedings of the 25th Symposium on Applied Computing (ACM SAC 2010), vol. 2, pp. 1585–1591 (2010)
14. Bas, P., Filler, T., Pevný, T.: Break our steganographic system: the ins and outs of organizing boss. In: Proceedings of 13th International Conference on Information Hiding, pp. 59–70 (2011)