ANTI-FORENSICS OF JPEG COMPRESSION

Matthew C. Stamm, Steven K. Tjoa, W. Sabrina Lin, and K. J. Ray Liu

Dept. of Electrical and Computer Engineering, University of Maryland, College Park

ABSTRACT

The widespread availability of photo editing software has made it easy to create visually convincing digital image forgeries. To address this problem, there has been much recent work in the field of digital image forensics. There has been little work, however, in the field of anti-forensics, which seeks to develop a set of techniques designed to fool current forensic methodologies. In this work, we present a technique for disguising an image's JPEG compression history. An image's JPEG compression history can be used to provide evidence of image manipulation, supply information about the camera used to generate an image, and identify forged regions within an image. We show how the proper addition of noise to an image's discrete cosine transform coefficients can sufficiently remove quantization artifacts which act as indicators of JPEG compression while introducing an acceptable level of distortion. Simulation results are provided to verify the efficacy of this anti-forensic technique.

Index Terms— Anti-Forensics, Digital Forensics, JPEG Compression

I. INTRODUCTION

The widespread availability of the Internet, coupled with the development of affordable, high quality digital cameras has resulted in an environment where digital images have supplanted traditional film-based photographs as the primary source of visual information in several scenarios. This has proved problematic due to the fact that powerful graphics editing software has enabled forgers to easily manipulate digital images. As a result, a number of researchers have developed computer-based forensic algorithms to detect digital forgeries even when they are visually convincing [1]. Additionally, methods have been developed to perform other forensically significant tasks such as tracing an image's processing history or determining the device used to capture an image.

One image processing operation of particular forensic significance is JPEG compression, which is one of the most popular image compression formats in use today. Prior work has shown that an image's origin can be determined by comparing the quantization tables used during JPEG compression to a database of those employed by specific digital camera models and image editing software [2]. If the quantization table is matched to an image editor, then the authenticity of the image can be questioned. The use of JPEG compression can be detected even if the image is later saved in an uncompressed format and the quantization table used during compression can be estimated directly from the previously compressed image [3]. JPEG compression followed by recompression using a different quantization table can be detected, and the primary quantization table can be estimated [1] [4]. Additionally, localized evidence of double JPEG compression can be used to identify image forgeries [5].

At present, virtually all existing digital image forensic techniques assume that no *anti-forensic* methods are employed by an image forger to disguise evidence of image tampering or alter other forensically significant image properties. This assumption proves to be a rather strong one, given the fact that an image forger may have a digital signal processing background and be well versed in digital forensics literature. To account for this possibility, anti-forensic image processing operations must be developed and studied so that weaknesses in existing image forensic techniques can be made known to researchers. This will allow researchers to know when forensic results can be trusted and to assist researchers in the development of improved digital forensic techniques. The study of anti-forensic operations may also lead to the development of techniques capable of detecting when an anti-forensic operation has been used. Furthermore, anti-forensic operations may be used to provide intellectual property protection by preventing the reverse engineering of proprietary signal processing operations used by digital cameras through digital forensic means. To the best of our knowledge, there are only two existing anti-forensic techniques: a set of operations designed to render image rotation and resizing undetectable and an technique to synthesize color filter array patterns [6] [7].

In this work, we propose an anti-forensic operation capable of disguising key evidence of JPEG compression. It operates by removing the discrete cosine transform (DCT) coefficient quantization artifacts indicative of JPEG compression. The resulting anti-forensically modified image can then be re-compressed using a different quantization table to hide evidence of tampering or to falsify the images origin. Alternatively, further processing can be performed to remove blocking artifacts and the image can be passed off as never-compressed.

II. JPEG COMPRESSION ARTIFACTS

When a grayscale image undergoes JPEG compression, it is first segmented into a series of 8×8 pixel blocks, then the DCT of each block is computed. Next, each DCT coefficient is quantized by dividing it by its corresponding entry in a quantization matrix Q, such that a DCT coefficient X at the block position (i, j) is quantized to the value $\hat{X} = \text{round}(\frac{X}{Q_{i,j}})$. Finally, the quantized DCT coefficients are rearranged using the zigzag scan order and losslessly encoded.

To decompress the image, the sequence of quantized DCT coefficients is losslessly decoded then rearranged into its original ordering. Dequantization is performed by multiplying each quantized coefficient by its corresponding entry in the quantization matrix, resulting in the dequantized coefficient $Y = Q_{i,j}\hat{X}$. Finally, the inverse DCT (IDCT) of each block of DCT coefficients is computed and the resulting pixel values are rounded to the nearest integer. Pixel values greater that 255 or less than 0 are truncated to 255 or 0 respectively, yielding the decompressed image.

Because of the lossy nature of JPEG compression, two important artifacts will be introduced into the decompressed image. The coupling of the quantization and dequantization operations force the value each DCT coefficient to be an integer multiple of the quantization step size $Q_{i,j}$. Though the process of rounding and truncating the decompressed pixel values to the set $\{0, \ldots, 255\}$ perturbs the DCT coefficient values, the DCT coefficient values typically remain tightly clustered around integer multiples of $Q_{i,j}$. This artifact can clearly be seen in Fig. 1 which shows DCT coefficient histograms of both an uncompressed image and one which has undergone JPEG compression. A second compression artifact is the pixel value discontinuities which occur across block boundaries as a result of the blockwise lossy compression employed by JPEG.

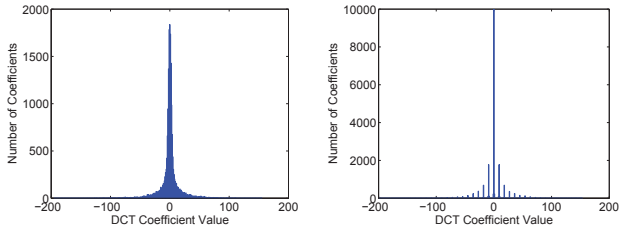

Fig. 1. Left: Histogram of DCT coefficients from an uncompressed image. Right: Histogram of DCT coefficients from the same image after JPEG compression

These discontinuities are commonly referred to as blocking artifacts and are often visually discernible.

When examining an image saved in an uncompressed or losslessly compressed format, both DCT coefficient quantization and blocking artifacts can be used as forensic indicators of previous JPEG compression. Furthermore, DCT coefficient quantization artifacts can be used to estimate the values of the quantization table used during JPEG compression. Because the removal of blocking artifacts has been extensively studied in the past, we concern ourselves with the anti-forensic removal of DCT coefficient quantization artifacts.

III. DCT COEFFICIENT QUANTIZATION ARTIFACT REMOVAL

In order to disguise evidence of previous JPEG compression, all DCT coefficient quantization artifacts must be removed from an image by an anti-forensic image processing operation. It is important to note that the original, unquantized DCT coefficients need not be recovered by the anti-forensic operation. All that is required is that the operation yield an image whose DCT coefficients are free from quantization artifacts and whose DCT coefficient distributions are plausible for an uncompressed image.

To accomplish this, we propose an anti-forensic operation which adds noise to each quantized DCT coefficient so that the values of each DCT coefficient are no longer clustered around integer multiples of $Q_{i,j}$. When doing this, the choice of the additive noise distribution is critical to the performance of the anti-forensic operation. If the additive noise is of insufficient strength, quantization artifacts may remain in the anti-forensically modified image. By contrast, if too much noise is added then unacceptable visual distortions may be introduced into the image.

To prevent these problems, we first estimate the distribution of the unquantized DCT coefficients. Next, we add to each DCT coefficient noise whose distribution is conditionally dependent upon the coefficient value to which it is added. These conditional distributions are chosen to be normalized segments of length $Q_{i,j}$ of the estimated distribution, with each segment being centered about a quantized DCT coefficient value. This is done to exploit the fact that the unquantized DCT coefficient value corresponding to each quantized one must lie in the interval $[Y - \frac{Q_{i,j}}{2}, Y + \frac{Q_{i,j}}{2}]$. By choosing the additive noise distributions in this manner, the marginal distribution of anti-forensically modified DCT coefficients will match the estimated distribution of unquantized DCT coefficients.

III-A. Estimating the Unquantized DCT Coefficient Distribution

We model the unquantized DCT coefficients as being distributed according to the Laplacian distribution

$$P(X = x) = \tfrac{\lambda}{2} e^{-\lambda|x|} \tag{1}$$

for the AC components [8]. After quantization, the AC components of the DCT coefficients will be distributed according to the discrete Laplacian distribution

$$P(Y = y) = \begin{cases} 1 - e^{-\lambda Q_{i,j}/2} & \text{if } y = 0, \\ e^{-\lambda|y|} \sinh(\frac{\lambda Q_{i,j}}{2}) & \text{if } y = kQ_{i,j}, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $k \in \mathbb{Z}$, $k \neq 0$. Assuming that the quantization table is known, a maximum likelihood estimate of the parameter λ can be obtained from the quantized DCT coefficients using the formula

$$\lambda_{ML} = -\frac{2}{Q_{i,j}} \ln(\gamma), \tag{3}$$

where γ is defined as

$$\gamma = \frac{-N_0 Q_{i,j}}{2NQ_{i,j} + 4S} + \frac{\sqrt{N_0^2 Q_{i,j}^2 - (2N_1 Q_{i,j} - 4S)(2NQ_{i,j} + 4S)}}{2NQ + 4S}, \tag{4}$$

and where $S = \sum_{k=1}^{N} |y_k|$, N is the total number of observations of the quantized (i, j) DCT coefficent, N_0 is the number of observations taking the value zero, and N_1 is the number of nonzero observations [9]. The quantized DCT coefficients can be obtained from the perturbed DCT coefficients, denoted by Y', through requantization according to the formula

$$Y = Q_{i,j} \operatorname{round}\left(\frac{Y'}{Q_{i,j}}\right). \tag{5}$$

Unfortunately, no accurate model of the distribution of the DC component exists. To compensate for this, we modify DC coefficient values in a different manner than AC coefficient values. This is discussed in greater detail in the subsequent section.

III-B. Additive Noise Distribution

As was previously stated, each anti-forensically modified DCT coefficient Z is obtained according to the equation

$$Z = Y + N \tag{6}$$

where N is additive noise whose distribution is conditionally dependent on the value of Y.

When modifying AC components, the choice of the conditional noise distribution is dictated by the estimated model distribution. For quantized DCT coefficients taking the value zero, the additive noise distribution is given by

$$P(N = n|Y = 0) = \begin{cases} \frac{1}{c_0} e^{-\lambda_{ML}|n|} & \text{if } \frac{-Q_{i,j}}{2} \geq n > \frac{Q_{i,j}}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $c_0 = 1 - e^{-\lambda_{ML} Q_{i,j}/2}$. For quantized DCT coefficients taking the nonzero value y, we use the noise distribution

$$P(N = n|Y = y) = \begin{cases} \frac{1}{c_1} e^{-\operatorname{sgn}(y)\lambda_{ML}(n+q/2)} & \text{if } \frac{-Q_{i,j}}{2} \geq n > \frac{Q_{i,j}}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where $c_1 = \frac{1}{\lambda_{ML}}(1 - e^{-\lambda_{ML} Q_{i,j}})$.

Assuming that the model distribution is accurate and that $\lambda_{ML} = \lambda$, this choice of conditional noise distributions ensures that the distribution of anti-forensically modified DCT coefficients will exactly match the model distribution of unmodified DCT

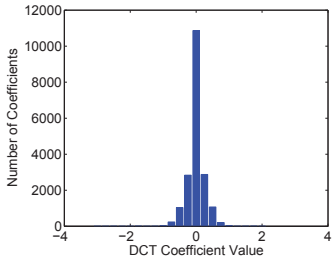

Fig. 2. Histogram of perturbed DCT coefficients which were previously quantized to zero.

coefficients. To see this, we can use the law of total probability to write

$$\begin{aligned} P(Z=z) &= \sum_{y} P(Z=z|Y=y)P(Y=y) \\ &= \sum_{y\neq 0} \tfrac{1}{c_1} e^{-\operatorname{sgn}(y)\lambda(z-y+Q_{i,j}/2)} e^{-\lambda|y|} \sinh(\tfrac{\lambda Q_{i,j}}{2}) \\ &\quad + \tfrac{1}{c_0} e^{-\lambda|n|}(1-e^{-\lambda Q_{i,j}/2}) \\ &= \tfrac{\lambda}{2} e^{-\lambda|z|} \end{aligned} \tag{9}$$

It has been previously observed that the variance of each DCT coefficient decreases as one moves toward higher frequencies [8]. For many image and quantization table pairings, all values of certain high frequency DCT coefficients are quantized to zero. In such cases, the parameter λ_{ML} cannot be estimated. Though this may initially seem problematic, we can allow these coefficients to remain unmodified. This is because the perturbations to each DCT coefficient caused by mapping all decompressed pixel values to the set $\{0,\ldots,255\}$ will result in a plausible DCT coefficient distribution. This can be seen in Fig. 2, which shows a typical histogram of perturbed DCT coefficients which were previously quantized to zero.

Because no general model accurately represents the DC coefficient distribution, we add noise with the distribution

$$P(N=n) = \begin{cases} \frac{1}{Q_{i,j}} & \text{if } \frac{-Q_{i,j}}{2} \leq n < \frac{Q_{i,j}}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

to each value. The resulting distribution of the anti-forensically modified DC coefficient approximates the distribution of the unquantized DC coefficient as constant over each quantization interval. Though this may lead to step discontinuities at the boundary between intervals, we have observed that typically very few image blocks have DC terms of the DCT that are quantized to the same value. As a result, step discontinuities are not discernible when examining a histogram of an image's DC DCT coefficient values.

One advantage of choosing the additive noise distributions in this manner is that a bound can be placed on the error between the anti-forensically modified DCT coefficients and their unquantized counterparts. The absolute error between an unquantized DCT coefficient in the (i,j) position and its quantized counterpart can be bounded by

$$|X-Y| \leq \tfrac{Q_{i,j}}{2}. \tag{11}$$

Because the support of each additive noise distribution is $[\frac{-Q_{i,j}}{2}, \frac{Q_{i,j}}{2}]$, the following bound can be placed on the absolute error between an anti-forensically modified DCT coefficient and its unquantized counterpart.

$$|X-Z| \leq Q_{i,j}. \tag{12}$$

Fig. 3. Top: JPEG compressed image using a quality factor of 65. Bottom: Anti-forensically modified image.

If $Q_{i,j}$ is sufficiently small for all DCT coefficients, the image distortions which result from adding noise to the quantized DCT coefficients will be visually undetectable.

Additionally, this choice of noise distributions leads to a fairly simple and efficient implementation. Only one parameter per DCT subband must be estimated to generate the noise distributions for the AC components, and no parameters must be estimated for the DC component. Furthermore, the noise distribution used for nonzero AC components depends only on the sign of the DCT coefficient value. Noise to be added to negative DCT values can be generated from noise realizations intended for positive DCT values simply by multiplying by negative one. Because of this, only two noise distributions must be generated per AC component and only one for the DC component.

IV. SIMULATIONS AND RESULTS

Results obtained using our anti-forensic DCT quantization artifact removal operation are shown in Fig. 3, which displays a JPEG compressed image using a quality factor of 65 before and after our anti-forensic operation is applied. As can be seen from these images, very little visual distortion is introduced by our anti-forensic operation. Furthermore, the PSNR between the anti-forensically modified image and the JPEG compressed image is PSNR= 41.63 dB. This indicates that our proposed anti-forensic operation introduces an acceptable level of distortion while producing an image that can plausibly be passed off as never-compressed.

Figs. 4 and 5 show DC and AC DCT coefficient histograms of the image displayed in Fig. 3 before and after JPEG compression as well as after our anti-forensic operation has been applied. DCT quantization artifacts are clearly visible in the histograms obtained from the JPEG compressed image. These artifacts are absent from the histograms of anti-forensically modified DCT coefficients. Additionally, the histograms of the anti-forensically modified coefficients closely match those obtained from the uncompressed image. This reinforces the assertion that the anti-forensically modified image can be passed off as never-compressed.

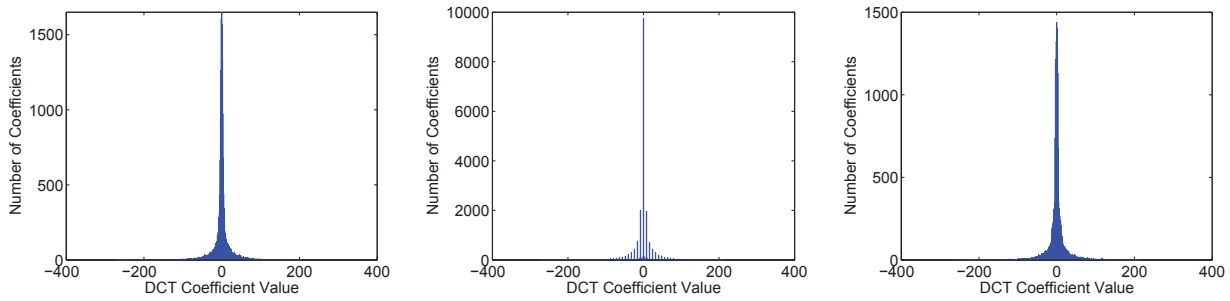

Fig. 4. Histogram of (2,2) DCT coefficients taken from an uncompressed version of the image shown in Fig. 3 (left), the same image after JPEG compression (center), and an anti-forensically modified copy of the JPEG compressed image(right).

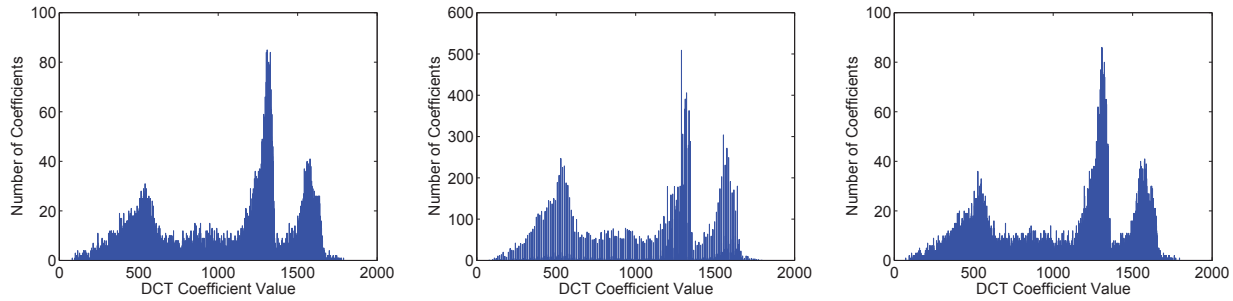

Fig. 5. Histogram of DC DCT coefficients taken from an uncompressed version of the image shown in Fig. 3 (left), the same image after JPEG compression (center), and an anti-forensically modified copy of the JPEG compressed image(right).

To test the effectiveness of our anti-forensic operation on a larger scale, we compressed then anti-forensically modified a set of 244 images taken from the Uncompressed Colour Image Database [10]. These images were compressed using quality factors of 90, 70, and 50. After each image was anti-forensically modified, we used the algorithm described in [3] to estimate the quantization table used during compression and classify each image as never-compressed or previously JPEG compressed. Images were only classified as never-compressed if every quantization table entry was estimated as one or if no estimate could be obtained. We should note that performing classification in this manner significantly biases the output towards deciding that an image was previously JPEG compressed. Despite this, the classifier was unable to detect previous JPEG compression in 95.90% of the anti-forensically modified images previous compressed with a quality factor of 90, 92.62% of those previously compressed with a quality factor of 70, and 81.56% of images previously compressed with a quality factor of 50. Furthermore, none of the quantization table estimates were correct. When the classification constraint was relaxed so that an estimated quantization table with two or fewer non-unity entries resulted in a decision of 'never-compressed', the classifier was unable to detect evidence of JPEG compression in any of the images compressed with quality factors of 90 or 70, and only correctly identified two images compressed with a quality factor of 50 as previously JPEG compressed.

V. CONCLUSIONS

In this paper, we have proposed an anti-forensic operation designed to remove the DCT coefficient quantization artifacts from a JPEG compressed image which several image forensic techniques make use of. This is accomplished by adding noise to the set of quantized DCT coefficients from a JPEG compressed image so that the distribution of anti-forensically modified coefficients matches an estimate of the distribution of unquantized DCT coefficients. Simulations show that when modifying a JPEG compressed image with a quality factor of 90, our proposed anti-forensic operation is capable of fooling a DCT quantization artifact based classifier 95.90% of the time.

VI. REFERENCES

[1] A.C. Popescu and H. Farid, "Statistical tools for digital forensics," in *6th International Workshop on Information Hiding*, Toronto, Canada, 2004.

[2] H. Farid, "Digital image ballistics from JPEG quantization," Tech. Rep. TR2006-583, Dept. of Computer Science, Dartmouth College, 2006.

[3] Z. Fan and R. 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] T. Pevny and J. Fridrich, "Detection of double-compression in JPEG images for applications in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, June 2008.

[5] J. He, Z. Lin, L. Wang, and X. Tang, "Detecting doctored JPEG images via dct coefficient analysis," in *Proc. of ECCV*, 2006, vol. 3593, pp. 423–435.

[6] M. Kirchner and R. Bohme, "Hiding traces of resampling in digital images," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 582–592, Dec. 2008.

[7] M. Kirchner and R. Böhme, "Synthesis of color filter array pattern in digital images," in *Proc. SPIE-IS&T Electronic Imaging: Media Forensics and Security*, Feb. 2009, vol. 7254.

[8] E.Y. Lam and J.W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1661–1666, Oct 2000.

[9] J.R. Price and M. Rabbani, "Biased reconstruction for JPEG decoding," *Signal Processing Letters, IEEE*, vol. 6, no. 12, pp. 297–299, Dec 1999.

[10] G. Schaefer and M. Stich, "UCID: an uncompressed color image database," in *Proc. SPIE: Storage and Retrieval Methods and Applications for Multimedia*, 2003, vol. 5307, pp. 472–480.