DETECTABILITY-QUALITY TRADE-OFF IN JPEG COUNTER-FORENSICS

G. Valenzise[†]*, M. Tagliasacchi*[*]*, S. Tubaro*[*]

[†]LTCI Lab. CNRS UMR 5141, Tèlècom ParisTech, Paris, France
[*]Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milan, Italy

ABSTRACT

Removing JPEG quantization footprints from an image inevitably introduces artifacts and traces in the spatial domain. Recently, several robust methods have been proposed to detect footprints of counter-forensics and recover the image's compression history. In this paper we investigate the limitations of these detectors, by proposing an improved counter-forensic attack which adds a post-processing denoising step besides dithering. We consider both a general-purpose denoising algorithm and one targeted to JPEG images. In the latter case, we show that this approach can successfully reduce the accuracy of detectors in the literature to that of a random decision. As a second contribution, we study the trade-off between the detectability of counter-forensics and quality of the tampered image, and show that the loss of quality is not sufficient for the analyst to use available no-reference quality assessment tools as an indicator of an attack.

***Index Terms*—** digital image forensics; counter-forensics; JPEG compression

1. INTRODUCTION

In order to study the weaknesses and limitations of current forensic tools and increase their robustness with respect to a knowledgeable attacker, counter-forensic researchers design techniques able to conceal the tell-tale footprints left by typical image processing operations. Among these traces, compression footprints play an important role, as they can reveal the compression history of an image. This is compelling, since most of the visual content output from point-and-shoot cameras or downloaded from the Internet is only available in compressed lossy formats, among which JPEG is by far the most popular.

Due to quantization, JPEG coding entails characteristic patterns in the first-order statistics (i.e., the histogram) of decoded transform coefficients. Thus, by inspecting the histogram of discrete cosine transform (DCT) coefficients of an image, one can determine whether it has undergone single or double compression [1], and possibly estimate the original quantization matrix [2, 3]. The validity of these methods was first questioned by Stamm et al. [4], who showed that adding a proper *counter-forensic dithering* signal to compressed DCT coefficients renders their histogram indistinguishable from that of a never-compressed picture. This certainly eliminates traces of JPEG in the first-order statistics in the transform domain, but introduces new ones in the spatial domain [5]. Specifically, the dithering signal is an additive noise that alters the fine-scale spatial correlations present in natural images. This observation triggered the development of several accurate detectors of JPEG compression counter-forensics [6][7][8], as well as new, more advanced counter-forensic approaches [9][10].

In spite of the several efforts to characterize the performance of JPEG counter-forensics, it is not clear yet to which extent the JPEG compression history can be accurately reconstructed. In this paper, we aim to advance this comprehension in two ways. First, we propose an improved counter-forensic attack which can deceive current detectors of counter-forensics. To this end, we first remove quantization footprints in the DCT histogram through counter-forensic dithering; then, we post-process the dithered image with a denoising step to smooth out traces of dither noise in the spatial domain. In this way, the accuracy of state-of-the-art detectors of counter-forensics can be reduced down to that of a random decision. We consider both a general-purpose denoising approach based on total-variation minimization, and a method targeted to JPEG compression (shifting-based denoising). In both cases, the strength of denoising can be easily tuned by the attacker through a single scalar parameter.

As a second contribution, we consider the important aspect of the *quality* of the counter-forensically attacked image. We trace detectability-quality curves that show the trade-off between the accuracy of counter-forensic detection and the distortion introduced by counter-forensics. Next, we show that there exists a "sweet spot" region in the parameter space of the proposed attack which enables to lower the detectability of counter-forensics and guarantees a good visual quality. We complement this analysis by taking the point of view of the forensic analyst, who does not have access to the original JPEG images. As a solution, we propose a detector of blurriness based on a popular no-reference blur metric [11], which could increase the overall accuracy when used in conjunction with targeted counter-forensic detectors. Our results show that, at least for the targeted attack with shifting-based denoising, none of these detectors (the universal one based on blur, and the targeted ones proposed in the literature) can be considered reliable against the proposed post-processing attack.

The rest of the paper is organized as follows. In the next section, we review related work with a particular attention on JPEG counter-forensic detection. In Section 3, we describe the proposed counter-forensic attack. Section 4 presents the detection accuracy/image quality trade-off for the proposed counter-forensic attack, and Section 5 concludes the paper.

2. RELATED WORK

Examples of counter-forensic techniques presented in the recent literature include removing acquisition footprints left by the distinctive camera sensor noise [12]; hiding traces of processing such as resampling [13][14], median filtering [15]; JPEG compression [4][16][17]; general histogram processing [18][19]; or attacks to detectors based on local features [20][21]. In the seminal work [12], the authors

The project REWIND acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 268478.

proposed a taxonomy of counter-forensic techniques, distinguishing between *targeted* or *universal*, and *integrated* or *post-processing*, according to whether the attack aims to fool a specific detector, and if it is integrated in the tampering process or is applied on already forged images, respectively. The JPEG counter-forensic attack proposed in this paper extends the method based on the addition of *counter-forensic dithering* by Stamm et al. [4] with a post-processing step of denoising. We consider both a universal denoising technique (total-variation denoising), and a targeted one (shifting-based denoising). Our approach is somehow similar to the variational method described in [9], which also employs total-variation as a cost function, although the target in that case is eliminating blocking traces. This method has been extended in [10] with a non-parametric DCT histogram smoothing integrated into the attack, in contrast with our post-processing approach based on the Laplacian parametric model of Stamm et al. Both [9] and [10] do not analyze the possible visual quality loss introduced by the attack.

The existence of a trade-off between image quality and detectability has been pointed out for specific instances of counter-forensics in the early work [13], and discussed at a more general level in [22]. For the case of double JPEG compression counter-forensics the authors of [23] trace concealability-rate-distortion surfaces to analyze the trade-off between the strength of counter-forensic dithering and its effects on image distortion and rate. With respect to [23], we consider new attack methodologies and more advanced detectors of JPEG counter-forensics. Also, while in [23] quality is measured in a full-reference fashion, we adopt instead a no-reference approach, which is closer to the point of view of a forensic analyst who may want to consider quality degradations as a forensic clue.

The first detector of JPEG quantization footprints – the comb-like patterns in the histogram of dequantized DCT coefficients – has been proposed in [2], that also derives a maximum-likelihood estimator of the JPEG quantization matrix. Counter-forensic attacks aimed at deleting JPEG quantization footprints generally try to fill the gaps in the comb-shaped histogram of DCT coefficients, and to eliminate JPEG blocking artifacts. The detectors of JPEG counter-forensics are based on the observation that these two operations may alter the spatial correlations between pixels in the image. In this paper we consider the following three detectors:

a) Detector based on total variation [24][6]: This detector targets the counter-forensic dithering attack [4], which appears as an additive noise in the spatial domain. By taking advantage of properties of dithering probability distribution and of the idempotence of quantization, the forensic analyst can detect counter-forensics by re-compressing the doubted image with several JPEG quality factors Q, and compute for each Q a measure of the noisiness of the re-compressed image by means of the total variation (TV) functional[1]. The $\mathrm{TV}(Q)$ function exhibits a different behaviour depending on whether the image was originally compressed or not, and therefore enables to detect images which were counter-forensically dithered, as well as their original JPEG quality factor. The method applies as well to implementations with arbitrary quantization matrices [6].

b) Detector based on calibrated features [7]: The concept of calibrated features was originally proposed in the field of steganalysis for JPEG-compressed images [25], and denotes distinguishing features computable from the stego image which enable to estimate the same quantity in the cover image, thus revealing possible stego signals. In the case of JPEG, calibration is obtained by cropping by 4 pixels horizontally and vertically the image to analyze, and re-compressing it using the same quantization table, in such a way that the blocking structure of JPEG is desynchronized. The distinguishing feature extracted by [7] consists in the energy of DCT coefficients from a set of 28 high-frequency subbands. To detect counter-forensics, the analyst extracts these features from the doubted image $\mathbf{X}$ and from its calibrated version $\mathbf{Y}$, and then computes the average normalized difference F between features of $\mathbf{X}$ and calibrated features of $\mathbf{Y}$. If F is lower than a threshold, the image is considered to be an uncompressed original.

c) Detector based on SPAM features [26]: The rationale of this detector is to classify original and counter-forensically attacked images based on the alteration of inter-pixel correlations, modeled as a Markov process through the SPAM (Subtractive Pixel Adjacency Matrix) features. Although originally designed for detecting steganographic embedding of low-amplitude independent stego signals, SPAM features have been successfully employed as input to a support vector machine (SVM) classifier to detect JPEG counter-forensics by considering the counter-forensic dither as the stego signal [6].

The accuracy of these three methods in detecting the counter-forensic dither of [4], averaged over a wide range of JPEG quality factors, is 0.93, 0.97 and 0.95, respectively [6].

3. IMPROVED JPEG COMPRESSION COUNTER-FORENSICS

JPEG counter-forensic approaches proposed in the literature are based on the noisy characteristics of the counter-forensic dithering signal in the pixel domain. This suggests that these detectors could somewhat be fooled by removing these traces, e.g., by adding a denoising step as a post-processing to dithering insertion. We consider here two different forms of denoising, a universal (total variation denoising) and a targeted one (shifting-based denoising).

3.1. Total variation denoising

The total variation (TV) functional has been widely used as a regularization term in image denoising problems to smooth noisy images while preserving to some extent discontinuities due to edges [27]. The TV denoising problem can be written as the following variational problem:

$$\bar{\mathbf{Y}} = \underset{\mathbf{Z}\in[0,255]^M}{\operatorname{argmin}} \|\mathbf{Y}-\mathbf{Z}\|^2 + \gamma\,\mathrm{TV}\{\mathbf{Z}\}, \tag{1}$$

where the first term is the fidelity with respect to the dithered M-pixel image $\mathbf{Y}$. The problem is convex and can be efficiently solved using, e.g., the method proposed in [28]. In our experiments we make use of the implementation available at [29]. The attacker can adjust the parameter γ to trade smoothness of the reconstructed image $\bar{\mathbf{Y}}$ for the fidelity with respect to the input dithered image $\mathbf{Y}$. For $\gamma = 0$ the attack is equivalent to [4], while for higher values of γ the image will be smoother. This attack is expected to be especially effective against the TV detector discussed in Section 2, but does not cancel those traces due to JPEG blockiness to which calibrated features are more sensitive.

3.2. Shifting-based denoising

In alternative to a universal denoising approach, we consider also a targeted attack which aims at removing traces of dithering in the spatial domain by properly combining together shifted and compressed

[1] The TV functional is defined as the ℓ_1-norm of the first-order derivatives of the image.

versions of the dithered image $\mathbf{Y}$. In this way, we average out several copies of the image, thus suppressing the counter-forensic footprint, which is a high-frequency signal. By shifting the different copies, we also reduce the presence of JPEG blocking artifacts, to which the detector based on calibrated features is sensitive. The proposed algorithm proceeds as follows:

- Compute $\mathbf{Y}_P$, obtained by adding 7-pixel padding to $\mathbf{Y}$, replicating the pixel values at the right and bottom image boundaries.
- Compute $\mathbf{Y}_{i,j}$, $0 \leq i,j < 7$, obtained by cropping $\mathbf{Y}_P$ with (i,j) as top-left pixel location.
- JPEG compress $\mathbf{Y}_{i,j}$, $0 \leq i,j < 7$, to obtain $\tilde{\mathbf{Y}}_{i,j}$, with a quality factor $Q_S = Q + \Delta Q$, where Q is the quality factor of the JPEG compressed image available to the adversary.
- Add a counter-forensic dithering signal to $\tilde{\mathbf{Y}}_{i,j}$, to obtain $\bar{\mathbf{Y}}_{i,j}$.
- Compute $\bar{\mathbf{Y}}$, as follows

$$\bar{\mathbf{Y}} = \frac{1}{49} \sum_{i,j \neq 0} \bar{\mathbf{Y}}_{i,j}. \tag{2}$$

  That is, only the $64 - 15 = 49$ images obtained from shifted copies not aligned to the original JPEG grid are averaged together.

In this case, the parameter ΔQ can be adjusted by the adversary to trade-off image quality with detection accuracy, with smaller (negative) value of ΔQ producing the stronger denoising effect and thus stronger concealment of dithering.

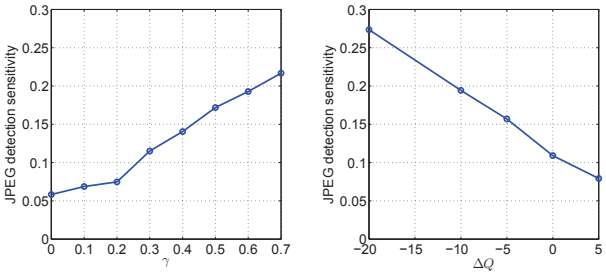
Notice that this method is similar to the denoising algorithm proposed in [30], which is also based on re-compressing shifted versions of the JPEG image. In our context, we restricted the averaging step to not aligned shifted copies, and we performed the insertion of the counter-forensic dithering signal twice: a first time to obtain $\mathbf{Y}$ from $\tilde{\mathbf{X}}$; and a second time on each of the shifted and compressed cropped images $\tilde{\mathbf{Y}}_{i,j}$ to obtain $\bar{\mathbf{Y}}_{i,j}$. Indeed, both steps are necessary. If the first one is omitted, comb-shaped artifacts appear in the histogram of the DCT coefficients. Conversely, if the second step is omitted, the counter-forensic method would be identified by a detector based on calibrated features. As demonstrated in Section 4, the joint use of both makes the approach undetectable to all tested approaches.

4. EXPERIMENTAL RESULTS

In this section we evaluate the performance of the proposed counter-forensic attack, in terms of detectability of JPEG quantization footprints after counter-forensics, and of the trade-off detectability-quality degradation of JPEG counter-forensics when using the detectors described in Section 2. The experiments were conducted using as test material the 1338 images of the UCID dataset [31]. We prepared a separate dataset for each value of $\gamma \in \{0.0, 0.1, 0.2, \ldots, 0.7\}$ and $\Delta Q \in \{-20, -10, -5, 0, +5\}$. In each dataset, half of the images were JPEG-compressed at a random quality factor $Q \in [30, 95]$. Then, for each image, we added a counter-forensic dithering signal and we processed the image according to one of the denoising methods described in Section 3.

4.1. Detectability of JPEG compression after counter-forensics

In order to test the effectiveness of the proposed counter-forensic attack in removing JPEG quantization footprints, we implemented the maximum-likelihood (ML) estimator of JPEG quantization matrix described in [2] and used it to build a detector of JPEG compression similarly to what is done in [4]. Specifically, an image was considered as never compressed if all the entries of the estimated quantization matrix except for the DC coefficient are equal to one. Since we observed that, especially for high-frequency subbands, the result of the ML estimation could be a valid quantization step even in case of non-compressed image, we only considered a subset of DCT frequencies (10 low-to-medium subbands) and declare an image as compressed if all of their quantization steps are estimated to be one, or if at least one of these quantization matrix entries could not be determined (e.g., because all coefficients in the subband are equal to zero). This setting guarantees to detect as many JPEG compressed images as possible, although it implies a higher false alarm rate.



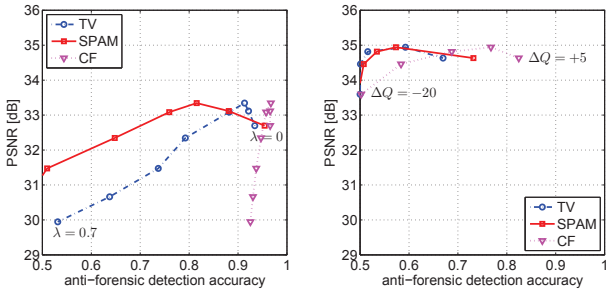
(a) Total-variation denoising (b) Shifting-based denoising

Fig. 1. Fraction of images attacked with counter-forensic plus denoising which are recognized to be JPEG compressed using the detector based on [2]. Results are the average over several quality factors in $[30, 95]$.

Figure 1 shows the fraction of images attacked with the proposed method which are recognized as being JPEG compressed. This corresponds to the *sensitivity* of the ML detector (also known as recall in information retrieval terms). Results are given for the two denoising strategies and the different values of parameters that control the denoising strength. We can make two observations on these results. On one hand, the detector of JPEG quantization footprints is not able to detect most of the images attacked with counter-forensics, and the fraction of detected images is always lower than the false alarm rate of our implementation of the ML detector (which is slightly lower than 0.3). On the other hand, the sensitivity to JPEG quantization footprints is slightly higher when using shifting-based denoising. The reason for that is linked to the denoising mechanism, which involves several JPEG re-compressions in the case of shifting-based denoising (that are not used instead for TV denoising), which are more likely to leave residual traces of quantization footprints.

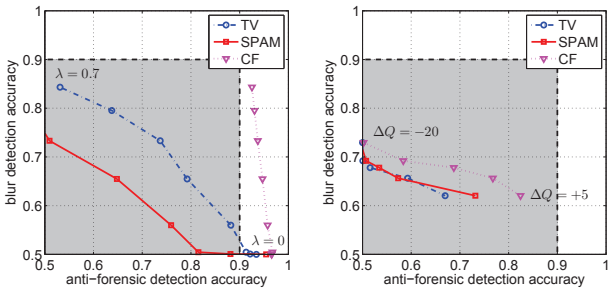
4.2. Detectability-quality trade-off of JPEG counter-forensics

We tested the detectability of the proposed JPEG counter-forensic attack with the three detectors described in Section 2: the total-variation detector (TV), the detector based on calibrated features (CF) and that based on SPAM features. The performance criterion considered here is the accuracy, defined as the number of correct classifications. The choice of the accuracy metric is justified by the fact that we employ a balanced dataset, i.e., both attacked and uncompressed original images are present in equal proportion. In addition to the counter-forensics detection accuracy, we measure the effect of denoising on the quality of the attacked image. To this end,



(a) Total-variation denoising (b) Shifting-based denoising

Fig. 2. Accuracy-quality curves. The PSNR is computed between the input (JPEG compressed image) and the output of the counter-forensic attack. The rightmost points ($\lambda = 0$) in part (a) of the figure correspond to counter-forensic method in [4].



(a) Total-variation denoising (b) Shifting-based denoising

Fig. 3. Combining the detection of counter-forensics with that of blur. For a given target accuracy of the analyst, the shaded area represents the points where the attacker "wins".

we compute the PSNR between the attacked (dithered and denoised) image, and the original JPEG image, i.e., the distortion between the output and the input of the proposed counter-forensic algorithm.

The accuracy-quality curves are traced in Figure 2. We can make three observations. The first one is that the post-processing denoising step in the proposed attack can effectively reduce the performance of counter-forensic detectors to that of a random decision in most of the cases. A remarkable exception is that of TV denoising and CF, which can be somehow expected since TV denoising is not especially effective in removing the blocking structure of JPEG, which is then detected through calibration. The opposite is true for the case of shifting-based denoised, and again this is in line with our hypotheses as this method is targeted to the specific distortion introduced by JPEG. The second remark is related to quality, which is notably higher in the case of the targeted attack (shifting-based denoising). We validated this observation by computing the PSNR between the attacked images and the original pictures *before* JPEG compression. We found that, for $\Delta Q = 0$, the attacked images have a lower average distortion with respect to the original non-compressed pictures than the compressed JPEG images. Finally, the third and most interesting comment concerns the trade-off that exists between quality and counter-forensic detectability. The attacker can tune the strength of counter-forensics in such a way to conceal as much as he wants his traces, but if he pushes this concealment too far, the resulting image quality can be degraded. What is interesting is that, with shifting denoising, he can keep the PSNR of attacked images higher than 34 dB, and reduce the accuracy of the most powerful detector (CF) available in the literature below 0.6.

The analysis in Figure 2 is useful for an attacker to understand when it is convenient for him to stop canceling his traces before making the tampered image practically uninteresting due to excessively loss of quality. In a practical scenario, however, this information does not help very much a forensic analyst, who instead does not have access to the original content, nor to its JPEG-compressed version. Instead, the forensic analyst could quantify the amount of distortion – blur in this case – based *only* on the characteristics of the analyzed image, and use this feature to guess whether the image has been attacked. To this end, the analyst might use a no-reference (NR) blur metric, such as the one described in [11]. For a given image, the metric compares the average variation between its gradient and that of a blurred version of the image (obtained with a 9×9 averaging filter). This difference is normalized between 0 and 1, and is smaller for images that were already blurred, and bigger for images that were originally sharper [2]. We use the NR blur metric to build a detector of the proposed JPEG counter-forensic attack. We employ a separate set of 978 images [33] for finding the optimal threshold of the detector, obtained by maximizing his accuracy over a wide assortment of coding conditions.

The forensic analyst can use the NR blur detector to complement the detectors described in Section 2 and increase his chances to detect counter-forensics. Figure 3 shows accuracy curves with the two kinds of detectors – the ones targeted to JPEG counter-forensics and the universal blurriness metric. In the accuracy plane, one can fix a minimum accuracy level for the joint use of the two kinds of detectors. If this accuracy is not achieved by at least one of the detectors, the attacker "wins", in the sense that his possibilities to fool the analyst are above an acceptable rate. For instance, if the minimum accuracy of the analyst is required to be 90%, this region is represented by the shaded area in the picture. We notice that for shifting-based denoising, all the strategies adopted by the adversary are successful in making the analyst's responses unreliable.

5. CONCLUSIONS

In this paper we propose an improved JPEG counter-forensic attack based on counter-forensic dithering, which augments it with a post-processing denoising step able to remove the traces left by dithering in the spatial domain. We consider two kinds of denoising, a universal technique based on TV and a targeted one designed to reduce JPEG-specific distortion. The targeted attack results to be effective against all the detectors of JPEG counter-forensics proposed in the literature. At the same time, it preserves or even improve the quality of the tampered image, which gives a larger degree of freedom to the attacker and reduces the possibilities that the forensic analyst can detect an attack based on a no-reference evaluation of image blur.

The detectability-quality trade-off analyzed in this paper suggests that effectively removing JPEG compression footprints can be done without necessarily producing significant quality degradations. The results of this paper could also be a useful contribution for game-theoretic approaches that study the optimal strategies for the analyst and the attacker in the presence of quality constraints [34, 35].

[2] We validated the results using also another popular NR blur metric presented in [32], and we found equivalent results to the metric in [11], which is therefore the only one used for the results shown in this paper.

6. REFERENCES

[1] H. Farid, "Digital image ballistics from JPEG quantization," *Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2006-583*, 2006.

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

[3] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," *Proc. of Digital Forensic Research Workshop*, 2003.

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

[5] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "The cost of JPEG compression anti-forensics," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic, May 2011.

[6] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Revealing the traces of jpeg compression anti-forensics," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 335–349, 2013.

[7] S.Y. Lai and R. Böhme, "Countering counter-forensics: the case of JPEG compression," in *Information Hiding*. Springer, 2011, pp. 285–298.

[8] H. Li, W. Luo, and J. Huang, "Countering anti-JPEG compression forensics," in *Proceedings of the International Conference on Image Processing*, Orlando, FL, USA, September 2012.

[9] W. Fan, Wang. K., F. Cayre, and Z. Xiong, "A variational approach to JPEG anti-forensics," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 2013, pp. 3058–3062.

[10] W. Fan, K. Wang, F. Cayre, and Z. Xiong, "JPEG anti-forensics using non-parametric DCT quantization noise estimation and natural image statistics," in *Proc. of the 1st ACM Workshop on Information Hiding and Multimedia Security*, 2013, pp. 117–122.

[11] F. Crété-Roffet, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *Proceedings of SPIE*, 2007, vol. 6492, p. 64920I.

[12] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, "Can we trust digital image forensics?," in *Proc. of the 15th ACM Int. Conf. on Multimedia*, 2007, pp. 78–86.

[13] M. Kirchner and R. Böhme, "Hiding traces of resampling in digital images," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 4, pp. 582 –592, dec. 2008.

[14] H. Muammar and P.L. Dragotti, "An investigation into aliasing in images recaptured from an LCD monitor using a digital camera," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2013, pp. 2242–2246.

[15] M. Fontani and M. Barni, "Hiding traces of median filtering in digital images," in *EURASIP European Signal Processing Conference (EUSIPCO12)*, August 2012.

[16] S. Milani, M. Tagliasacchi, and S. Tubaro, "Antiforensics attacks to Benford's law for the detection of double compressed images," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2013, pp. 3053–3057.

[17] C. Pasquini and G. Boato, "JPEG compression anti-forensics based on first significant digit distribution," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2013.

[18] M. Barni, M. Fontani, and B. Tondi, "A universal technique to hide traces of histogram-based image manipulations," in *Proceedings of the ACM Int. Workshop on Multimedia and security*, New York, NY, USA, 2012, MM&Sec '12, pp. 97–104, ACM.

[19] P. Comesaña Alfaro and F. Pérez-González, "Optimal counterforensics for histogram-based forensics," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2013, pp. 3048–3052.

[20] I. Amerini, M. Barni, R. Caldelli, and A. Costanzo, "Counter-forensics of SIFT-based copy-move detection by means of keypoint classification," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–17, 2013.

[21] A. Melloni, P. Bestagini, A. Costanzo, M. Barni, M. Tagliasacchi, and S. Tubaro, "Attacking image classification based on bag-of-visual-words," in *To be presented at IEEE International Workshop on Information Forensics and Security (WIFS)*, 2013.

[22] R. Böhme and M. Kirchner, "Counter-forensics: Attacking image forensics," in *Digital Image Forensics*, H.T. Sencar and N. Memon, Eds., pp. 327–366. Springer New York, 2013.

[23] X. Chu, M.C. Stamm, Y. Chen, and K.J.R. Liu, "Concealability-rate-distortion tradeoff in image compression anti-forensics," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 2013, pp. 3063–3067.

[24] G. Valenzise, V. Nobile, M. Tagliasacchi, and S. Tubaro, "Countering JPEG anti-forensics," in *Proceedings of the International Conference on Image Processing*, Bruxelles, Belgium, September 2011.

[25] J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," in *Proc. Inf. Hiding Workshop, Springer LNCS*, 2004, pp. 67–81.

[26] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215 –224, june 2010.

[27] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, November 1992.

[28] P. Rodríguez and B. Wohlberg, "Efficient minimization method for a generalized total variation functional," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 322–332, 2009.

[29] "Numerical methods for inverse problems and adaptive decomposition," `http://numipad.sourceforge.net/`.

[30] A. Nosratinia, "Enhancement of JPEG-compressed images by re-application of JPEG," *J. VLSI Signal Process. Syst.*, vol. 27, no. 1/2, pp. 69–79, February 2001.

[31] G. Schaefer and M. Stich, "UCID: an uncompressed colour image database," in *Proc. SPIE: Storage and Retrieval Methods and Applications for Multimedia*, 2004, vol. 5307, pp. 472–480.

[32] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proceedings of the International Conference on Image Processing*, Rochester, NY, 2002, vol. 3, pp. 57–60.

[33] "NRCS photo gallery," Available: http://photogallery.nrcs.usda.gov/.

[34] M.C. Stamm, W.S. Lin, and K.J.R. Liu, "Temporal forensics and anti-forensics for motion compensated video," *IEEE Trans. Inf. Forensics and Security*, vol. 7, no. 4, pp. 1315 –1329, August 2012.

[35] M. Barni and B. Tondi, "The source identification game: An information-theoretic perspective," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 450–463, 2013.