COUNTERING ANTI-FORENSICS OF MEDIAN FILTERING*

Hui Zeng, Tengfei Qin, Xiangui Kang, Li Liu*

School of Information Sci. &Tech., Sun Yat-sen University. isskxg@mail.sysu.edu.cn

ABSTRACT

The statistical fingerprints left by median filtering can be a valuable clue for image forensics. However, these fingerprints may be maliciously erased by a forger. Recently, a tricky anti-forensic method has been proposed to remove median filtering traces by restoring images' pixel difference distribution. In this paper, we analyze the traces of this anti-forensic technique and propose a novel counter method. The experimental results show that our method could reveal this anti-forensics effectively at low computation load. According to our best knowledge, it's the first work on countering anti-forensics of median filtering.

Index Terms— Image forensics, median filtering, anti-forensic, pixel difference

1. INTRODUCTION

Image forensics in an adversarial environment has raised more and more attention in recent years. There are two major roles in this area. The forger side aims to find the weakness of traditional forensics usually called anti-forensics [1-5]. The investigator role aims to detect the traces of forger, usually been taken as a patch on traditional forensics [6-7]. Furthermore, some advanced mathematical tools, such as game theory, are resorted to analyze the ultimate limits of the interplay between these two roles [8-9]. Researches of both directions are helpful in improving the security of forensics. In this paper, we focus on improving the security of median filtering detection.

Median filtering detection is important in image forensic and image steganalysis [10-12]. Many forensic methods are developed to identify median filtering with excellent performance in recent years [13-15]. So far works only consider the robustness of median filtering detection with some common processing, such as JPEG compression. However, the existence of some malicious manipulation (anti-forensics) makes the detection task more complex.

A target attack aiming at median filtering detection is proposed in [1]. The image's pixel difference distribution is modified with adding anti-forensic noise. Since a number of median filtering detectors are operating on the statistics of pixel value differences of an image, this anti-forensic method could fool median filtering detectors or significantly reduce their performance. This attack challenges the traditional forensics and urges researcher to find countermeasures.

To address this challenge, we analyze the traces left by anti-forensics and propose a novel counter method. The proposed counter method has low complexity.

The remainder of this paper is organized as follows. In Section 2, we briefly review the anti-forensic method [1] and evaluate its effectiveness against some median filtering detectors. In Section 3, we analyze the options available to the adversary and the traces left by anti-forensics. The counter method is described in Section 4. In Section 5, we report the experimental results and evaluate its counter anti-forensic performance. The paper is concluded in Section 6.

2. ANTI-FORENSICS OF MEDIAN FILTERING

Many median filtering detectors capture features of an image's pixel value difference distribution [10] [11] [13] [14]. As a result, the intuitive idea of anti-forensics against these detectors is to restore the image's pixel difference distribution.

In [1], the pixel difference distribution of an image is modeled with a two dimensional generalized Gaussian distribution [16], which can be determined by a covariance matrix $\sum$ and a shape parameter α.

To restore the image's pixel difference distribution in both horizontal and vertical directions, the forger firstly estimates the pixel difference distribution of an unaltered image $\hat{f}_U$ from that of a median filtered image f_M. Using $\hat{f}_U$ and f_M, the forger can calculate the distribution of the anti-forensic noise to be added to the median filtered image:

$$f_N = IDFT\{DFT\{\hat{f}_U\} / DFT\{f_M\}\}. \quad (1)$$

Once f_N is obtained, the forger can anti-forensically modify the pixel values according to f_N. To avoid introducing large distortion, the forger partitions the image into blocks and selects an anchor point at the center of a

* This work was supported by NSFC (Grant nos. 61379155 and U1135001), the Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20110171110042) and NSF of Guangdong province (Grant no. s2013020012788)

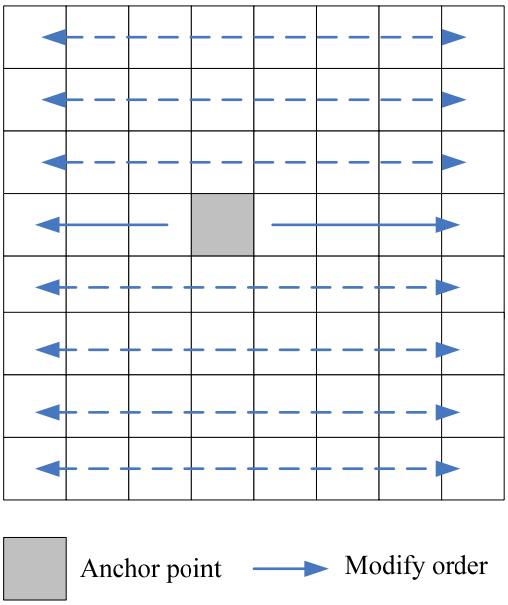

Fig. 1. The procedure of anti-forensics, $B = 8$.

block. Without loss of generality, we assume the block size is B and the anchor point at the location ($B/2$, $B/2$).

The forger modifies all the pixels in the $B/2$ row using

$$y_{B/2,B/2+l} = x_{B/2,B/2} - \sum_{k=0}^{l-1} h_{B/2,B/2+k} - \sum_{k=0}^{l-1} n^{h}_{B/2,B/2+k}, \quad (2)$$

where x is the image before anti-forensic modify, y is the image after anti-forensic modify, h is the pixel difference in horizontal direction, and n^h is the noise realization of the one dimensional noise distribution f_N. This process is shown in Fig. 1 with solid line. In this paper, we call the row that contains anchor point as *anchor row*.

Next, the forger modifies the rows below and above the *anchor row* using

$$y_{B/2+k+1,B/2+l} = y_{B/2+k,B/2+l} - v_{B/2+k,B/2+l} - n^{v}_{B/2+k,B/2+l}, \quad (3)$$

where v is the pixel difference in vertical direction and n^v is the noise realization of the conditional distribution f_N. The forger continues this process until all rows in the block are modified, which is shown in Fig. 1 with dashed lines.

The anti-forensics effectiveness on the UCID image database [17] is evaluated. Three median filter detectors, ρ method, SPAM method [10] and MFF method [13], are used in our experiments. For the ρ detector, we report the detection rate under the probability of false alarm being 1%. For other detectors, we use the original images and the median filtered images to train a model, and use this model to classify the forged images. The detection rates are shown in Table I. It is observed that almost all forged images are classified as the original images, which means that the anti-forensic method [1] successfully fools the detectors.

3. THE TRACES LEFT BY ANTI-FORENSICS

Fig. 2. (a) Original, (b) Anti-forensic modified, $B = 8$, $T = 5$, (c) Anti-forensic modified, $B = 16$, $T = 3$.

Table I
Detection rates (%) with different detectors after anti-forensic modification

	$B = 4$, $T = 3$	$B = 8$, $T = 3$
ρ	1.12	0.90
SPAM	0	0.15
MFF	0	0

In this section, we analyze the options available to the anti-forensic adversary and the traces left by them.

A requirement for anti-forensics is not introducing obvious perceptual distortion. From equation (2) and (3), it can be observed that the distortion is accumulated as the pixel is away from the anchor point. Hence, the block size B and the max value of anti-forensic noise T are the key parameters that affect the visual quality of the forged image. In our experiments, we find that the perceptual distortion would become obvious when $B > 8$ or $T > 4$. Fig. 2 shows the cases when improper parameters are chosen. The perceptual distortion is noticeable especially in smooth areas.

Another artifact left by anti-forensics [1] is the different noise adding strategies between the *anchor row* and the other rows. For example, we examine the pixel difference in the 4th row and 5th row of the forged image in Fig. 1.

$$\begin{aligned} y_{5,i} - y_{5,i+1} &= (y_{4,i} - v_{4,i} - n^{v}_{4,i}) - (y_{4,i+1} - v_{4,i+1} - n^{v}_{4,i+1}) \\ &= (y_{4,i} - y_{4,i+1}) - (v_{4,i} - v_{4,i+1}) - n^{v}_{4,i} + n^{v}_{4,i+1}. \end{aligned} \quad (4)$$

According to the property that when two random variables are added together, their distributions convolve with each other [1], the last three items would cause the histogram of the pixel difference in the 5th row smoother than that of the 4th row (here we make an assumption that these items are independent with the pixel difference in the 4th row). Similarly, the histogram of the pixel difference in the 6th row would smoother than that in 5th row, etc. In general, the histogram of the pixel difference becomes smoother as the row away from the *anchor row*. Hence, the histogram of horizontal pixel difference of different rows would present a certain periodicity.

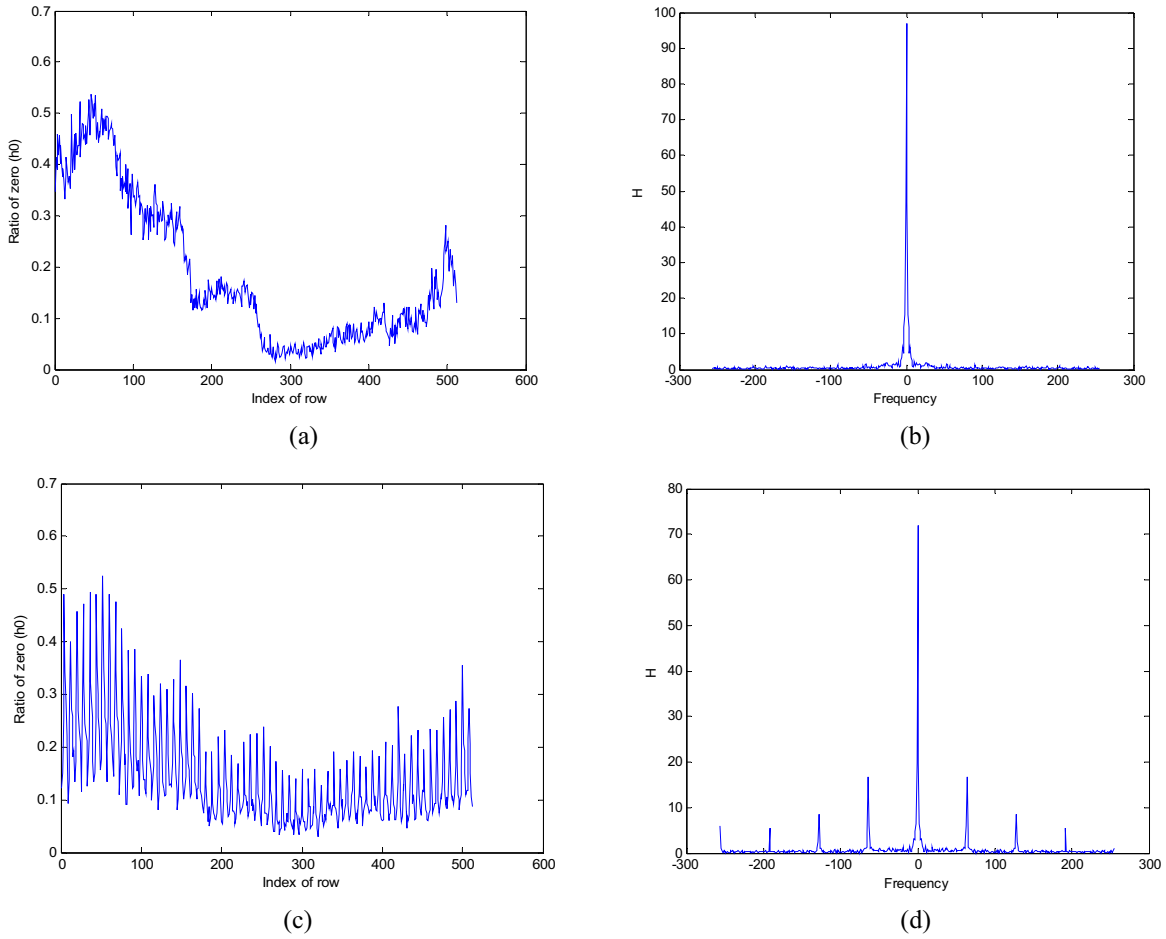

Fig. 3. (a) h^0 of original image, (b) DFT of (a), (c) h^0 of forged image, (d) DFT of (c).

To verify the assumption above, we examine the ratio of zero in each row. In this paper, we denote the ratio of zero of horizontal pixel difference in the i^{th} row as $h^0(i)$, $i \in 1, ..., m$. m is the number of rows of the test image. Fig. 3 plots h^0 and its DFT transform (H) of an original image and a forged image. It is observed that h^0 of a forged image present an obvious periodicity, where the local maximums correspond to the *anchor row*s. This periodicity, which could be used to identify the forgery, becomes more clearly in frequency domain. The detailed procedure of our counter method is described in Section 4.

4. COUNTERMEASURE OF ANTI-FORENSICS

To detect the anti-forensic traces, we propose a countermeasure as shown in Fig. 4. We first calculate a test image's pixel difference in horizontal direction. Then, we calculate the ratio of zero in each row and get h^0. Next, we calculate its DFT transform $H = \text{DFT}\{h^0\}$. If the test image has been anti-forensic modified, there are peaks in the H, which corresponding to the periodicity of h^0, otherwise there is no such phenomenon.

We use the peak detection method of [18] to detect the peak. If the value of $H(i)$ is greater than the mean in a window $[i\text{-}W, i+W]$ by a threshold T_h, the peak location i except DC location is recorded as p. In our experiments, we choose $W = 10$, and $T_h = 1.4$ empirically. If at least one peak is detected in H , the test image is deemed as a forged one. Otherwise, the tested image is recognized as an original one. Moreover, the first peak location $p(1)$ corresponds to the block size B:

$$B = m / p(1). \tag{5}$$

5. EXPERIMENTAL RESULTS

We evaluate the performance of our counter anti-forensic method using the UCID database which consists of 1338 color images. For anti-forensics, we follow the settings of [1] and convert all the images to gray scale before any further processing. The gray scale images were used as the

unaltered images. For the median filtered image database, the grayscale images were processed using a median filter with support 3. For the anti-forensic image database, the median filter images were processed as [1] with the parameter $B = \{3, 4, 5, 8\}$ and $T = \{2, 3, 4\}$. We don't focus on the case that $B > 8$ or $T > 4$ because they would introduce obvious perceptual distortion.

The experimental results are reported in Table II. Some examples of false rejected images are show in Fig. 5. It can be observed that the textures of these images are very complex which can affect the periodicity used in our algorithm. As stated in Section 4, the proposed method can also estimate the block size used in anti-forensic. The

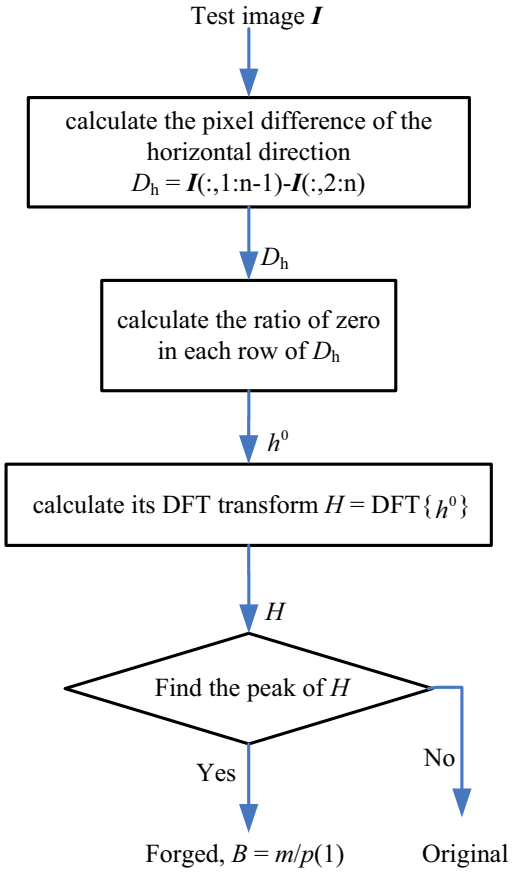

Fig. 4. Flowchart of the proposed method

Table II
Detection rates (%) at the false alarm rate being zero

T \ B	3	4	5	8
2	99.3	99.3	99.6	99.8
3	99.1	99.0	99.4	99.7
4	99.3	99.0	99.6	99.6

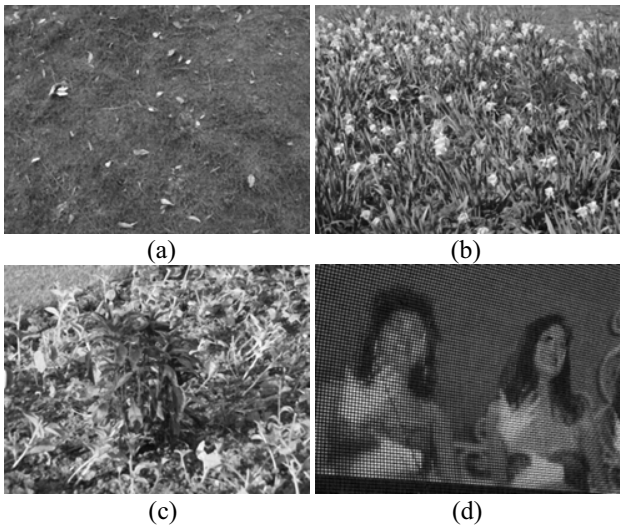

Fig. 5. Examples of images falsely rejected by the proposed method.

Table III
Success rates (%) of obtaining correct block size B

T \ B	3	4	5	8
2	99.3	99.3	99.6	99.8
3	99.1	99.0	99.4	99.7
4	99.3	99.0	99.6	99.6

success rates (%) of obtaining correct block size B are reported in Table III. In general, the proposed method can detect the forgery accurately where wide range of parameters are used by the forger in anti-forensics from Table II. Note that the proposed method does not need complex training process and the average run time to test an image from the UCID database is less than 0.003s. All the tests are performed on a computer with a 3.1 GHz processor and 4 GB RAM.

To test our method on more image databases, we also evaluate the proposed method on the DID database [19] and similar performance is achieved. We omit the detailed results here due to the length limitation.

6. CONCLUSIONS

In this paper, we analyze the traces left by anti-forensics of median filtering, and propose a novel countermeasure. The proposed method shows excellent performance in detecting the anti-forensic forgery with low complexity. To the best of our knowledge, it's the first work on countering anti-forensics of median filtering. The proposed method can not end up the cat-and-mouth game between forensics and anti-forensics of median filtering. However, it would help to improve the security of forensics tremendously.

7. REFERENCES

[1] Z.-H. Wu, M. C. Stamm, K. J. R. Liu, "Anti-Forensics of Median Filtering", *IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, Vancouver, Canada, pp. 3043-3047, May 2013.

[2] T. Gloe, M. Kirchner, A. Winkler, and R. Bohme, "Can we trust digital image forensics?" *Multimedia '07*, Proceedings of the 15th international conference on Multimedia, pp. 78–86, Sept. 2007.

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

[4] C. Kwok, O. C. Au, and S. Chui. "Alternative anti-forensics method for contrast enhancement," *In Proceedings of the 10th international conference on Digital-Forensics and Watermarking*, Y. Shi, H. Kim, and F. Perez-Gonzalez (Eds.). *Lecture Notes In Computer Science*, Vol. 7128. Springer Berlin Heidelberg, pp. 398-410, 2011.

[5] M. Fontani, M. Barni, "Hiding traces of median filtering in digital images," *Signal Processing Conference* (*EUSIPCO*), 2012 Proceedings of the 20th European, vol. , no. , pp. 1239-1243, Aug. 2012.

[6] M. Goljan, J. Fridrich, and M. Chen, "Defending against fingerprint-copy attack in sensor-based camera identification," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 227–236, 2011.

[7] S. Lai, R. Böhme, "Countering counter-forensics: The case of JPEG compression," *Information Hiding. Springer Berlin Heidelberg*, 2011, pp. 285-298.

[8] M. C. Stamm, W. S. Lin, and K. J. R. Liu, "Forensics vs. anti-forensics: A decision and game theoretic framework," *IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, pp. 1749–1752, Kyoto, 2012.

[9] 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.

[10] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Proc. SPIE*, *Electronic Imaging, Media Forensics and Security II*, 2010, vol. 7541, pp. 1–12.

[11] G. Cao, Y. Zhao, R. R. Ni, L. F.Yu, and H.W. Tian, "Forensic detection of median filtering in digital images," in *Proc. 2010 IEEE Int. Conf. Multimedia and EXPO*, 2010, pp. 89–94.

[12] A. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," in *Proc. SPIE*, *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, 2008, vol. 6819, pp. 501–517.

[13] H. Yuan, "Blind forensics of median filtering in digital images," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp.1335-1345, Dec. 2011.

[14] C. Chen, J. Ni, R. Huang, J. Huang, "Blind median filtering detection using statistics in difference domain," in *Proc. of information Hiding 2012*, Berkerly, USA, May 2012.

[15] X. Kang, M. C. Stamm, A. Peng, K. J. R. Liu, "Robust Median Filtering Forensics Using an Autoregressive Model," *IEEE Transactions on Information Forensics and Security*, vol.8, no.9, pp.1456,1468, Sept. 2013

[16] L. Boubchir and J. M. Fadili, "Multivariate statistical modeling of images with the curvelet transform," in *Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on*, 28-31, 2005, vol. 2, pp. 747 – 750.

[17] G. Schaefer and M. Stich, "UCID-An uncompressed color image database," In *Proc. of SPIE, Storage and Retrieval Methods and Applications for Multimedia*, 2004, pp. 472–480.

[18] W. Wei, S. Wang, X. Zhang, Z. Tang, "Estimation of Image Rotation Angle Using Interpolation-Related Spectral Signatures With Application to Blind Detection of Image Forgery," *IEEE Transactions on Information Forensics and Security*, vol.5, no.3, pp.507,517, Sept. 2010

[19] T. Gloe and R. Bohme, "The dresden image database for benchmarking digital image forensics," *Journal of Digital Forensic Practice*, vol. 3, no. 2-4, pp. 150-159, 2010.