Blind Forensics of Median Filtering Based on Markov Statistics in Median-Filtered Residual Domain

Yujin Zhang, Chenglin Zhao, Feng Zhao, and Shenghong Li

Abstract Revealing the processing history of a digital image has received a great deal of attention from forensic analyzers in recent years. Median filtering is a non-linear operation and has been used widely for noise removal and image enhancement. Therefore, exposing the traces introduced by such operation is helpful to forensic analyzers. In this paper, a passive forensic method to detect median filtering in digital images is proposed. Since overlapped window filtering introduces the correlation among the elements of the median-filtered residual (MFR) which is referred to as the difference between a test image and its corresponding median-filtered version, the transition probability matrices along the horizontal, vertical, main diagonal and minor diagonal directions are calculated from the MFR to characterize the correlation among the elements of the MFR. All elements of these transition probability matrices are served as discriminative features for median filtering detection. Experiment results demonstrate the effectiveness of the proposed method.

Keywords Image forensics • Median filtering • Median-filtered residual

Y. Zhang • S. Li (✉)
Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: yjzhang82@sjtu.edu.cn; shli@sjtu.edu.cn

C. Zhao
School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
e-mail: clzhao@bupt.edu.cn

F. Zhao
Department of Science and Technology, Guilin University of Electronic Technology, Guilin 541004, China
e-mail: zhaofeng@guet.edu.cn

1 Introduction

Nowadays, with the rapid development of smart image acquisition devices and image processing tools with friendly interface, forensic analyzers pay more and more attention to the authenticity of digital images. Existing approaches for digital image forensics can be roughly divided into two categories: active [1] and passive [2]. The active approaches mainly insert watermarks or signatures to digital images at the time of recording. Compared with the active methods, the passive ones only need the intrinsic characteristics of digital images rather than any watermarks or signatures, making them more practical. There are many promising passive forensic methods for the detection of malicious forgeries [3]. However, the subtle traces of tampering may be diminished by using some content-preserved manipulations, such as median filtering, blurring and contrast enhancement, which do not change the image content in general, but may decrease the reliability of forensics techniques [4]. Therefore, detecting such operations is helpful to forensic analyzers. We focus on the forensics of median filtering in this paper because it has been used widely for noise removal and image enhancement.

Recently, some passive techniques for detecting median filtering have been developed. In [5], Kirchner and Fridrich employed the ratio of histogram bins h_0 and h_1, and subtractive pixel adjacency matrix (SPAM) features in the first-order difference domain to capture the median filtering artifacts for the uncompressed and JPEG post-compressed images, respectively. In [6], Cao *et al.* used the probability of zero values on the first-order difference image in textured regions as statistical fingerprint to distinguish the median-filtered images from the non-median-filtered images. Yuan in [7] constructed the median filtering forensics (MFF) and scalar merged features to detect median filtering based on the observation that median filtering significantly affects either the order or the quantity of the gray levels contained in the image area encompassed by the filter window. Chen *et al.* in [8] formed the distinguishing fingerprints by combining global probability features with correlation features in the difference domain for median filtering detection. Kang *et al.* in [9] stated that overlapped window filtering introduces the correlation among the elements of the median filtered residual (MFR) and used the autoregressive model to model the MFR for identifying median filtering, achieving better performance than the methods in [5] and [7] in the JPEG post-compressed scenario.

It has been shown in [3, 5] that Markov transition probability matrix (TPM) can be used to model the correlation between the adjacent elements effectively. The above discussion motivates us to investigate the effectiveness of Markov features in the MFR domain for median filtering detection. We use transition probability matrices along the horizontal, vertical, main diagonal and minor diagonal directions generated from the MFR to characterize the correlation among the elements of the MFR to capture the traces introduced by median filtering in this paper.

The rest of this paper is organized as follows. The proposed method is described in Sect. 2. The Experimental results are reported in Sect. 3. Finally, the conclusions are drawn in Sect. 4.

2   Measuring the Median Filtering Artifacts Using Markov Statistics of the Median Filtered Residual
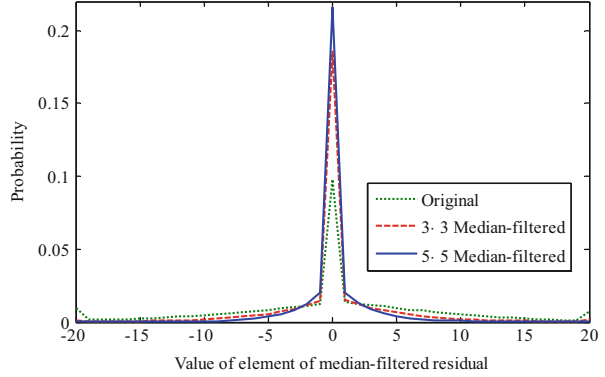
The main idea of the median filter is to sort the gray levels encompassed by the filter window and replace the value of the center pixel with the median of the gray levels. The most extensively used median filters are with the square windows of odd sizes (i.e. 3×3, 5×5, $\cdots$). Therefore, we concentrate on these forms of median filtering in this paper.

It is nonlinearity of median filtering that makes the theoretical analysis of general relationship between input and output distributions of median filter complicated [10]. Therefore, the forensics of median filtering is mainly focused on specific features extraction [5–9]. The first-order difference has been proven to be effective in capturing median filtering artifacts [5]. However, the performance of statistical features derived from the first-order difference domain is susceptible to the interference of the image content. It has been shown in [9] that the median filtered residual (MFR) can reduce the effects caused by the diversity of image content effectively. This motivates us to further investigate the statistical characteristics of the MFR. We first crop 1,000 images with size 256×256 from NRCS database [11] and convert them to the 8-bit grayscale images. Then, we perform 3×3 and 5×5 median filtering on these images to obtain their corresponding median-filtered versions, respectively. Figure 1 shows the arithmetic average of the histograms ranging from -20 to 20 calculated from the MFRs of the original images and their corresponding median-filtered versions, respectively.

From Fig. 1, it is observed that the elements of the MFRs are concentrated around zeros and quickly fall off. In addition, the distribution of the elements of the MFRs generated from the median-filtered images has a sharper peak than that generated from the original images. This means that overlapped window filtering introduces the correlation among the elements of the MFR. Markov transition probability matrix (TPM) [3, 5] has also been proven to be effective in characterizing the correlation among the elements. Motivated by the Markov TPM, the proposed discriminative features for median filtering detection are constructed as follows:

1) Calculate the difference between the given image with size $M \times N$ and its corresponding 3×3 median-filtered version to obtain the MFR array E similar to that in [9].
2) Calculate the transition probability matrices along the horizontal, vertical, main diagonal and minor diagonal directions of the MFR array E in the range [-T,T] using

Fig. 1 The arithmetic
average of the histograms
ranging from -20 to
20 calculated from the
MFRs of 1,000 original
images and their
corresponding median-
filtered versions,
respectively

$$P_h(u,v) = \frac{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i,j)-u)\delta(E(i+1,j)-v)}{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i,j)-u)}$$

$$P_v(u,v) = \frac{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i,j)-u)\delta(E(i,j+1)-v)}{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i,j)-u)}$$

$$P_d(u,v) = \frac{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i,j)-u)\delta(E(i+1,j+1)-v)}{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i,j)-u)} \tag{1}$$

$$P_m(u,v) = \frac{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i+1,j)-u)\delta(E(i,j+1)-v)}{\sum\limits_{i=0}^{M-2}\sum\limits_{j=0}^{N-2}\delta(E(i+1,j)-u)}$$

where the subscripts h, v, d, and m denote the horizontal, vertical, main diagonal and minor diagonal directions, respectively; $u, v \in \{-T, -T+1, \cdots, 0, \cdots, T\}$ and

$$\delta(n-n_0) = \begin{cases} 1, & n = n_0 \\ 0, & n \neq n_0 \end{cases} \tag{2}$$

By doing so, each of the transition probability matrices is of size $(2T+1) \times (2T+1)$.

3) Average P_h and P_v, and then P_d and P_m to form the matrices P_1 and P_2 as

$$P_1 = (P_h + P_v)/2; \quad P_2 = (P_d + P_m)/2 \qquad (3)$$

4) Take all elements of the matrices P_1 and P_2 as discriminative features for median filtering detection.

Note that since the mask of the median filtering is symmetric about the origin, the averaging procedure reduces the feature dimension without affecting the detection performance of the proposed method. Based on the experimental dataset prepared in Sect. 3.1, we empirically choose the threshold $T = 7$ for a compromise between the detection performance and computing complexity when constructing the transition probability matrices. Therefore, there are $2(2T + 1)^2 = 450$ elements in the developed feature set for the given image.

3   Experiments and Results

3.1   Image Database

In our experiments, NRCS image database [11] is used to evaluate the performance of the propose method. The original color images in above database are of size $1,500 \times 2,100$ in general. We first pick out 1,000 images in TIFF formats from the above database randomly and convert them to the 8-bit grayscale images. Then, to investigate the performance of the proposed method for the varying image size, we crop 1,000 image blocks with size 256×256, 128×128 and 64×64 from central region of each grayscale image to construct three sizes of the image databases, respectively. After that, for each of three sizes of the image databases prepared above, nine training-testing pairs are constructed similar to that in [7] as follows:

1) Perform 3×3 and 5×5 median filtering, 3×3 average filtering, 3×3 Gaussian low-pass filtering with the standard deviation $\sigma = 0.5$, and rescale operation which is randomly generated by nearest or bilinear interpolation and scaling factors 1.1 on all images in the original database, denoted as S^{ORI}, to obtain the corresponding databases S^{MF3}, S^{MF5}, S^{AVE}, S^{GAU} and S^{RES}, respectively.
2) Build eight training-testing pairs using the image databases S^{MF3}, S^{MF5}, S^{ORI}, S^{GAU}, S^{RES} and S^{AVE}, i.e. MF3 versus ORI, MF5 versus ORI, MF3 versus GAU, MF5 versus GAU, MF3 versus RES, MF5 versus RES, MF3 versus AVE, and MF5 versus AVE.
3) Group each 50 % (randomly selected) of the median-filtered databases S^{MF3} and S^{MF5} into the database S^{MF35}, and group each 25 % (randomly selected) of the non-median-filtered databases S^{ORI}, S^{GAU}, S^{RES} and S^{AVE} into the database

Fig. 2 Detection accuracies achieved by using the proposed, SPAM-based, and MFF-based methods for varying training-testing pairs built in Sect. 3.1
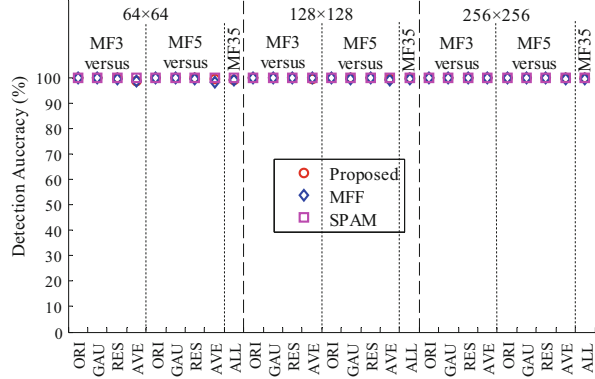
S^{ALL}. Build one training-testing pair using the databases S^{MF35} and S^{ALL}, i.e. MF35 versus ALL.

3.2 Classification

LIBSVM [12] is used as the classifier in our experiments. The RBF kernel function is selected for classification. For each of training-testing pairs constructed above, 2/5 of the positive samples and 2/5 of the negative ones are randomly picked out to train the SVM classifier, and the remaining 3/5 of the positive samples and 3/5 of the negative ones are used to test the trained SVM classifier. Note that all the median-filtered images are considered as the positive samples, while all the non-median-filtered images are defined as negative samples. The optimal parameters (C,γ) for the SVM classifier are achieved in the multiplicative grid $(C,\gamma) \in \{(2^i, 2^j) | i \in \{0, 0.5, \ldots, 8\}, j \in \{-5, -4.5, \ldots, 5\}\}$ by a fivefold cross-validation on the training set. The above procedure is repeated thirty times for reducing the effect of randomness caused by image selection for training and testing. The detection accuracy, which is the arithmetic average of true positive rate (TPR) and true negative rate (TNR), is averaged over thirty times random experiments.

3.3 Detection of Median Filtering

To evaluate the effectiveness of the proposed method, a series of experiments are carried out. Two state-of-the-art methods (i.e. the SPAM-based [5] and MFF-based methods [7]) have also been investigated for comparison on the same dataset. For the SPAM detector, the threshold T in [5] is set to 3, which leads to the 686-D SPAM features. Figure 2 shows the detection performance of the proposed, SPAM-based, and MFF-based methods for varying training-testing pairs built in Sect. 3.1. As seen in Fig. 2, all these three methods perform well in detecting median filtering

Fig. 3 Detection accuracies achieved by using the proposed, SPAM-based, and MFF-based methods for varying training-testing pairs with the quality factor $Q = 90$
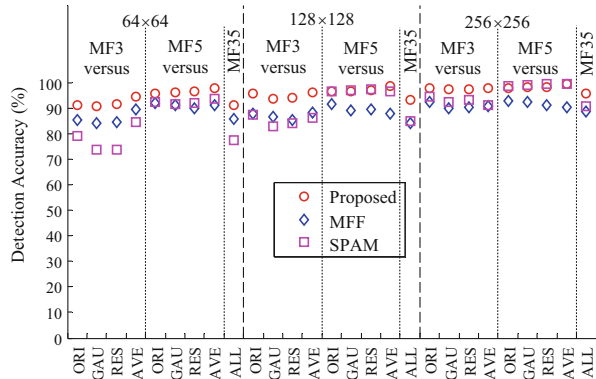
even when the test images are small (i.e. 64×64). In addition, it is easily observed that the SPAM detector achieves nearly perfect detection performance for all the training-testing pairs built in Sect. 3.1.

JPEG is one of the most widely used formats in many digital devices and image processing software. It is expected that a forensic scheme is robust against JPEG compression to certain extent. To assess the robustness of the proposed method against JPEG compression, we use the JPEG versions of the training-testing pairs constructed in Sect. 3.1 with the JPEG quality factor $Q = 90$, resulting in $9 \times 3 = 27$ training-testing pairs. Figure 3 shows the classification performance of the proposed, SPAM, and MFF detectors for JPEG post-compressed training-testing pairs with the quality factor $Q = 90$. From Fig. 3, it is observed that the proposed detector is more robust than the MFF and SPAM detectors in general. Besides, The MFF detector performs better than the SPAM detector in distinguishing the 3×3 median-filtered images from the non-median-filtered ones for lower image resolution (i.e. 128×128 and 64×64), whereas the SPAM detector achieves better performance than the MFF detector for 5×5 median filtering detection.

4 Conclusions

In this paper, Markov statistics in median-filtered residual domain have been investigated for median filtering detection. Overlapped window filtering introduces the correlation among the elements of the MFR, leaving detectable traces. The transition probability matrices along the horizontal, vertical, main diagonal and minor diagonal directions are calculated from the MFR to characterize the correlation among the elements of the MFR. All elements of these transition probability matrices are utilized as discriminative features for median filtering detection. Experiment results have shown that the proposed method can detect the median filtering effectively and perform better than several state-of-the-art methods. Future

work will be devoted to performance improvement of the proposed method and localization of the median-filtered regions in image forgeries.

References

1. Cox IJ, Kilian J, Leighton FT, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process 6(12):1673–1687
2. Luo W, Huang J, Qiu G (2010) JPEG error analysis and its applications to digital image forensics. IEEE Trans Inform Forensics Secur 5(3):480–491
3. Shi YQ, Chen C, Chen W (2007) A natural image model approach to splicing detection. In: Proceedings of ACM Multimedia and Security Workshop, Dallas, TX, 20–21 September, 2007, pp. 51–62
4. Kirchner M, Böhme RJ (2008) Hiding traces of resampling in digital images. IEEE Trans Inform Forensics Secur 3(4):582–592
5. Kirchner M, Fridrich J (2010) On detection of median filtering in digital images. Proc SPIE Electron Imag Med Forensics Secur II 7541:1–12
6. Cao G, Zhao Y, Ni R, Yu L, Tian H (2010) Forensic detection of median filtering in digital images. In: Proceedings of the 2010 I.E. international conference on multimedia and expo, Suntec City, 19–23 July, 2010, pp. 89–94
7. Yuan HD (2011) Blind forensics of median filtering in digital images. IEEE Trans Inform Forensics Secur 6(4):1335–1345
8. Chen C, Ni J, Huang R, Huang J (2013) Blind median filtering detection using statistics in difference domain. In: Proceedings of 14th information hiding. LNCS 7692:1–15
9. Kang X, Stamm MC, Peng A, Liu KJR (2012) Robust median filtering forensics based on the autoregressive model of median filtered residual. In: Proceedings of signal & information processing association annual summit and conference, Hollywood, CA, 3–6 December, 2012, pp. 1–9
10. Bovik AC (1987) Streaking in median filtered images. IEEE Trans Acous Speech Signal Process 35(4):493–503
11. NRCS Photo Gallery [Online]: http://photogallery.nrcs.usda.gov/res/sites/photogallery/
12. Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines [Online]. http://www.csie.ntu.edu.tw/cjlin/libsvm