FORENSIC DETECTION OF MEDIAN FILTERING IN DIGITAL IMAGES

Gang Cao Yao Zhao Rongrong Ni Lifang Yu Huawei Tian

Institute of Information Science, Beijing Jiaotong University, Beijing, 100044, China
{06112056, yzhao, rrni, 06112055, 06120416}@bjtu.edu.cn

ABSTRACT

In digital image forensics, prior works are prone to the detection of malicious tampering. However, there is also a need for developing techniques to identify general content-preserved manipulations, which are employed to conceal tampering trails frequently. In this paper, we propose a blind forensic algorithm to detect median filtering (MF), which is applied extensively for signal denoising and digital image enhancement. The probability of zero values on the first order difference map in texture regions can serve as MF statistical fingerprint, which distinguishes MF from other operations. Since anti-forensic techniques enjoy utilizing MF to attack the linearity assumption of existing forensics algorithms, blind detection of the non-linear MF becomes especially significant. Both theoretically reasoning and experimental results verify the effectiveness of our proposed MF forensics scheme.

Keywords—Image Forensics, Median Filtering, Image Difference

1. INTRODUCTION

As digital techniques advance, plenty of powerful media editing softwares appear and make sophisticated photo forgeries created easily and frequently. The integrity and authenticity of digital images can no longer be taken for granted. So there is an increasing need for techniques to detect alterations blindly. Digital image forensics is just such an art.

In general, image alterations can be classified into two types: content-changing tampering and content-preserving manipulations. Correspondingly, prior works fall into two categories. In the first category, forensics methods focus on detecting image tampering such as splicing [1] and copy-move [2, 3], by which the image content is reshaped visually and semantically. In the second category, common digital image manipulations such as resampling [4], compression [5], contrast enhancement [6, 7], blur [8, 9] and sharpening [10] are to be identified passively [11, 12]. These types of operations are also frequently used in the general image processing pipeline.

Although the content-preserving manipulations merely change image perceptual quality but not semantic content, their blind detection is still forensically significant. In one hand, such manipulations are usually applied to conceal visual trail of tampering operations for creating realistic forgeries. The fact can be affirmed is that the detection of different artifical manipulations can certainly throw in doubt the integrity and authenticity of digital images. On the other hand, general manipulations are employed as postprocessing to destroy the forensically significant fingerprints, which are left by previous tampering operations. In such a scenario, existing forensic algorithms are unavoidable to be fooled to some extent.

In this paper, we focus on the blind detection of median filtering, which is commonly used to denoise and smooth images. For forgery makers, MF can be exploited to retouch the crude tampering trace. It is important to note that plenty of current forensic methods, such as resampling and CFA interpolation detection, rely on the assumption of linear correlation between neighboring pixels [4, 13]. As a non-linear operator, median filter can destroy such linear relation and serve as an attack against the interpolation detection algorithms. A new resampling algorithm [14], which can invalidate Popescu and Farid's state-of-the-art resampling detector [4], was designed by using MF to hide interpolation traces. In such a scenario, the forensic detection of MF can help to identify the new resampling operations. In the synchronous work [15], streaking artifacts and subtractive pixel adjacency matrix (SPAM) features are employed to detect MF in bitmap and JPEG post-compressed images, respectively. Different from such metrics, in our proposed scheme, statistical fingerprint feature is devised as the occurrence probability of zero values on the one-order difference map of textured regions. The abnormity of such metric would be considered as evidence for identifying MF operations.

The rest of this paper is organized as follows. In Section 2, we formulate the statistical characteristics of image difference before and after MF, followed by the proposed MF detection scheme in Section 3. The evaluation results of MF forensic detection are reported in Section 4. Finally we draw the conclusions in Section 5.

This paper is supported in part by National 973 program (No. 2006CB303104), National Natural Science Foundation of China (No. 60702013, No. 60776794), Beijing Natural Science Foundation (No. 4073038).

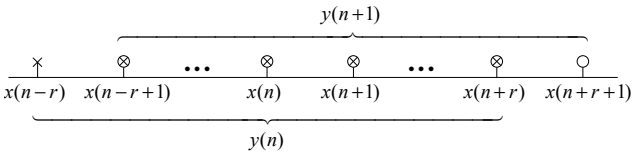


Fig. 1. The original signal $x(n)$ is median-filtered to be $y(n)$.

2. ANALYSIS OF MEDIAN-FILTERED SIGNAL

In this section, we carry out an analytic investigation of MF fingerprint properties from the signal statistic standpoint. Such fingerprint metric will become the identity card to distinguish MF from un-MF and other operations.

Order-statistics filter is a kind of nonlinear spatial filter whose response is based on ranking the digital elements contained in a local range. The best-known example in this category is the median filter, which, as its name implies, replaces the current value by the median of the element values in the neighborhood. Referring to Fig. 1, the median filtering on a one-dimensional digital signal (i.e. one row or column of a digital image) is illustrated. Let x be the original signal,

$$x = \{ x(n) \}, \quad n = 0, 1, 2, ..., N \tag{1}$$

where the integer $x(n) \in [0, 255]$.

Suppose the median-filtered signal is y , which can be written as,

$$y(n) = \underset{i \in [n-r,\, n+r]}{\text{median}} \{ x(i) \}, \quad n = 0, 1, 2, ..., N. \tag{2}$$

Here, width of the filter window is $w = 2r+1,\ r = 1, 2, 3, ...$. Boundary is processed implicitly and the output is supposed to have the same size as the input array. Correspondingly, we have

$$y(n+1) = \underset{i \in [n-r+1,\, n+r+1]}{\text{median}} \{ x(i) \}. \tag{3}$$

From the formula (2) and formula (3), we can see that there exist common elements for computing the median values

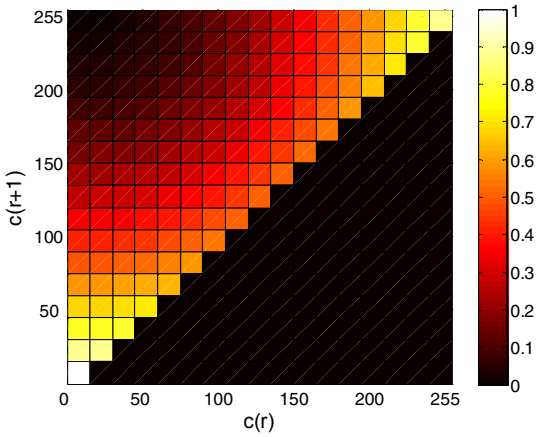


Fig. 2. Probability map for $P\{ y(n) = y(n+1) \}$.

$y(n)$ and $y(n+1)$. They are

$$C = \{ x(i) \,|\, i \in [n-r+1, n+r] \}. \tag{4}$$

It can be concluded that a certain distinct order must exist among the elements of the set C . The sorted sequence can be rewritten as

$$C_o = \{ c(i) \,|\, c(i) \le c(i+1),\ i \in [1, 2r] \}. \tag{5}$$

Then the filtered value $y(n)$ can be determined in terms of the relationship between $x(n-r)$ and the elements within C_o . That is,

$$y(n) = \begin{cases} c(r) & if\ x(n-r) \in [0, c(r)] \\ x(n-r) & if\ x(n-r) \in (c(r), c(r+1)]. \\ c(r+1) & if\ x(n-r) \in (c(r+1), 255] \end{cases} \tag{6}$$

Similarly, $y(n+1)$ is related with $x(n+r+1)$ as follows,

$$y(n+1) = \begin{cases} c(r) & if\ x(n+r+1) \in [0, c(r)] \\ x(n+r+1) & if\ x(n+r+1) \in (c(r), c(r+1)]. \\ c(r+1) & if\ x(n+r+1) \in (c(r+1), 255] \end{cases} \tag{7}$$

$$\begin{aligned} P\{ y(n) = y(n+1) \} &= P\{ x(n-r) \in [0, c(r)], x(n+r+1) \in [0, c(r)] \} \\ &+ P\{ x(n-r) \in (c(r), c(r+1)],\ x(n+r+1) \in (c(r), c(r+1)],\ x(n-r) = x(n+r+1) \} \\ &+ P\{ x(n-r) \in (c(r+1), 255],\ x(n+r+1) \in (c(r+1), 255] \}. \end{aligned} \tag{8}$$

$$\begin{aligned} P\{ y(n) = y(n+1) \} &= \left(\frac{c(r)+1}{256} \right)^2 + \left(\frac{c(r+1)-c(r)}{256} \right)^2 \cdot \frac{1}{C^1_{c(r+1)-c(r)}} + \left(\frac{255-c(r+1)}{256} \right)^2 \\ &= \frac{(c(r)+1)^2 + c(r+1) - c(r) + (255-c(r+1))^2}{256^2}. \end{aligned} \tag{9}$$

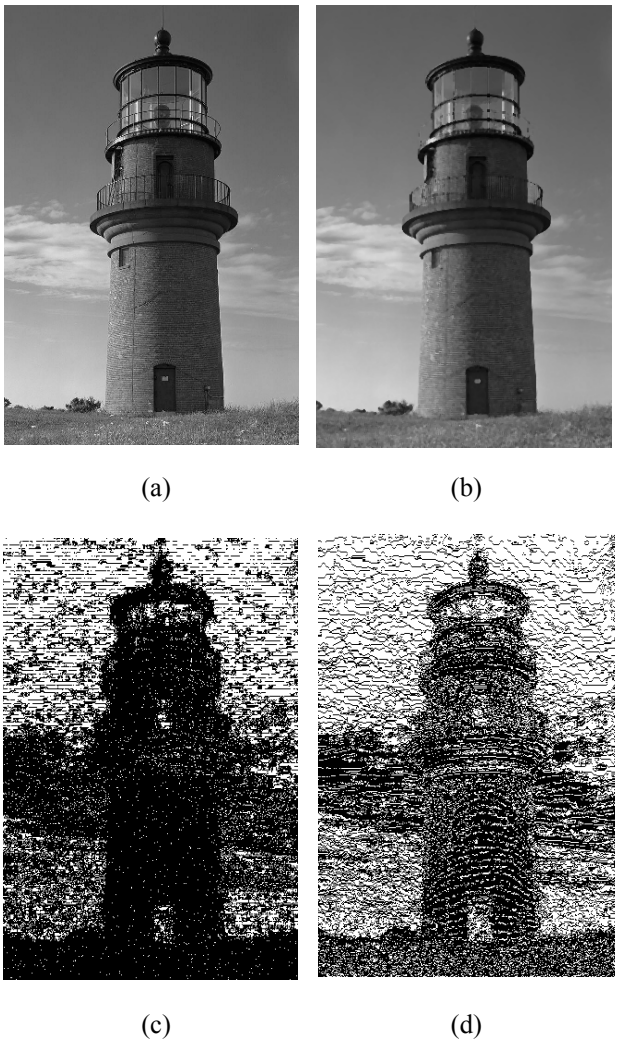
Because the space between $x(n-r)$ and $x(n+r+1)$ is as long as $2r+1$, the correlation between such two variables is so weak that they can be assumed to be independent. Then probability of $y(n)$ equaling to $y(n+1)$ can be deduced as the formula (8). On the assumption that both $x(n-r)$ and $x(n+r+1)$ confirm the uniform distribution $U[0,255]$, the probability formula (8) can be simplified as that in formula (9). From this formula, we can see that the probability $P\{y(n)=y(n+1)\}$ is just a function of the two median elements, namely $c(r)$ and $c(r+1)$. Visualized probability map is indicated in Fig.2, where the probability value at each point is signified by color. From the map, it can be observed that higher probability occurs if $c(r)$ and $c(r+1)$ get closer.

Recall that intense correlation lies between neighboring pixels in original digital images. That attributes to many factors, including the continuity of scene illumination and the imaging properties, e.g., CFA interpolation. Distribution of the inherent luminance and color of practical objects is also locally smooth. As a result, we can confirm that $c(r)$ and $c(r+1)$ are indeed adjacent in practical pixel sequences. Then the median-filtered sequence would possess a high but not low probability for the occurrence of equal neighboring elements. While considering the counterpart probability for the unfiltered signal, $P\{x(n)=x(n+1)\}$, there is a lack of distinct verdict due to the complexity of image statistical model. The theory studied in this section can be analogously extended to multidimensional cases.

3. PROPOSED MF DETECTION SCHEME

Although the occurrence probability for equal neighboring pixels is uncertain in the original image, an interesting statistical phenomenon has been found in texture regions. In such regions, the fluctuation of pixel graylevels is more severe than that in smooth areas. The correlation between neighboring pixels is also comparatively weak. Hence the case of adjoining pixels with identical graylevel happens sparsely, which can be seen from an example natural image shown in Fig. 3 (a) and the corresponding statistical result in Fig. 3 (c). The probability for equal neighboring pixels in original images becomes much low if the statistic scope is limited in texture regions.

Contrastively, according to the analysis in Section 2, the corresponding probability in the texture regions of median-filtered images still keeps at certain level, moreover, usually higher than that of original image. The comparison between Fig.3 (c) and (d) can verify such a conclusion illustratively. The reason for excluding untextured pixels is that the definite statistical discrepancy can't be assured in smooth regions, especially those among different images. Such a motivation can be justified by the sky region of the example image in Fig.3.



(a) (b) (c) (d)

Fig. 3. Statistical discrepancy caused by MF. (a) Original image; (b) 3x3 median-filtered version; (c) Map for original image, where equal (horizontal) neighboring pixels are marked white; (d) Map for the filtered image.

Based on such considerations, we propose MF detection scheme by capturing the unique probability finger-print. For a test image, the intensity channel is denoted by $I(i,j)$, where i=1, 2, ..., M and j=1, 2, ..., N. Firstly, its first order row-based difference is binarized as,

$$\nabla I_r(i,j) = \begin{cases} 1 & if \;\; I(i+1,j)-I(i,j)=0 \\ 0 & if \;\; I(i+1,j)-I(i,j)\neq 0 \end{cases}. \qquad (10)$$

To measure the texture characteristic of each pixel in the test image, variance of local surrounding pixels is calculated as follows,

$$\sigma(i,j) = \underset{\substack{m\in[i-\lceil d/2\rceil,\; i+\lceil d/2\rceil] \\ n\in[j-\lceil d/2\rceil,\; j+\lceil d/2\rceil]}}{\mathrm{Var}} \{I(m,n)\}. \qquad (11)$$

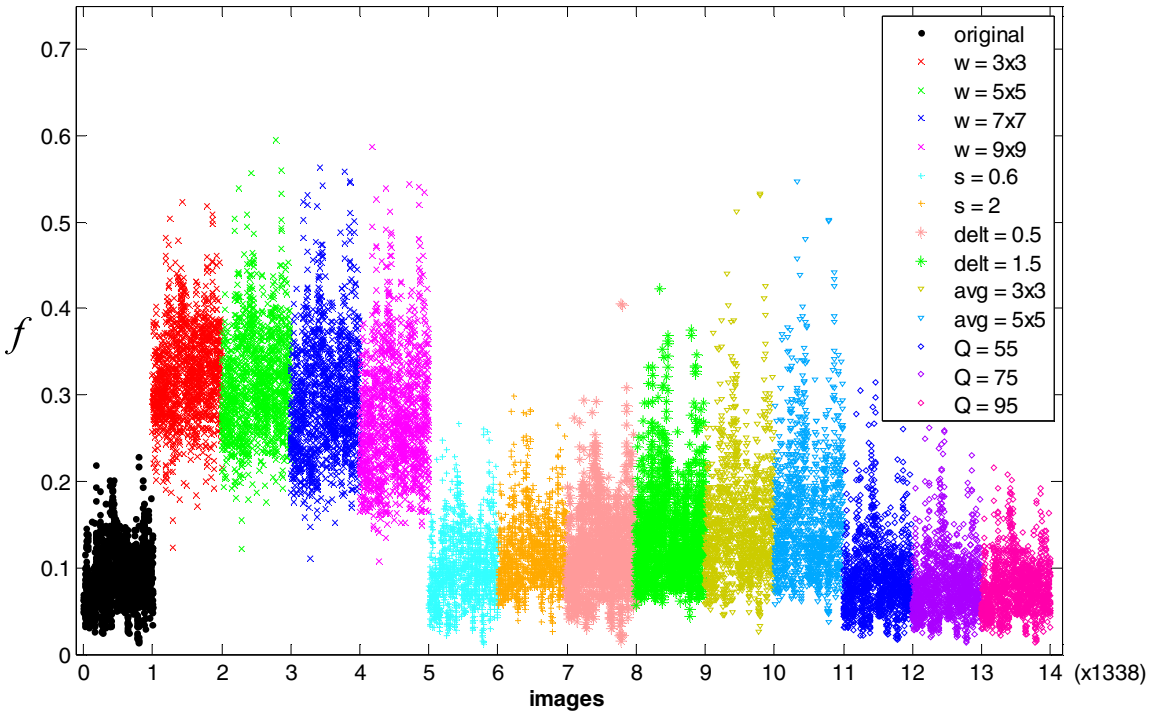


Fig. 4. Distribution of the features extracted from different types of sample images: original, median filtered (window size w=3x3, 5x5, 7x7, 9x9), bilinear scaled (scaling factor s=0.6, 2), Gaussian filtered (delt=0.5, 1.5), average filtered (window size avg=3x3, 5x5), JPEG compressed (Q=55, 75, 95).

Here, side width of the square statistic region is designated by d . Then such a variance matrix can be binarized by the following thresholding constraints:

$$V(i,j)=\begin{cases}1 & if\ \ \sigma(i,j)\geq\tau\\ 0 & if\ \ \sigma(i,j)<\tau\end{cases}. \tag{12}$$

where τ is the threshold chosen for determining if a pixel is textured. Then the metric, which measures the frequency of zero values in the first-order difference of texture regions, is designed to be as

$$f_r=\frac{\sum_i\sum_j \nabla I_r(i,j)\cdot V(i,j)}{\sum_i\sum_j V(i,j)}. \tag{13}$$

The metric f_r is the fingerprint for identifying MF operation. Correspondingly, alike metric f_c can be generated from the column-based difference. For the convenience of making decision, such two metric are to be fused simply as follows,

$$f=[f_r,f_c]\bullet[\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}] \tag{14}$$

where $\bullet$ denotes dot product between two vectors.

At last, the detection result of MF can be obtained by thresholding classification in the light of the feature f. The decision threshold can be selected according to the user's expectation of accepted false alarm rate. Images with metric of f greater than the decision threshold are to be judged as median-filtered ones.

4. EXPERIMENTS AND DISCUSSIONS

To evaluate the performance of the MF detection algorithm, the popular image dataset UCID [16] is introduced for test. The UCID dataset consists of 1338 uncompressed TIFF images on a variety of topics including natural scenes and man-made objects, both indoors and outdoors. Note that all images are color images and we convert them to grayscale in the standard manner. Results furnished here are obtained from operating the intensity images. In all the following experiments, the parameters used for determining textured pixels are set: $d=7$, $\tau=100$.

4.1. Evaluate Baseline Detection Performance

The basic capability for MF detection is evaluated under two cases: 1) without the disturbance of other operations; 2) with other operations occurred before MF.

In the first case, untouched original images and their median-filtered versions are taken as negative (N) and positive samples (P), respectively. MF with different filter window sizes are tested, w=3x3, 5x5, 7x7, 9x9, respectively.

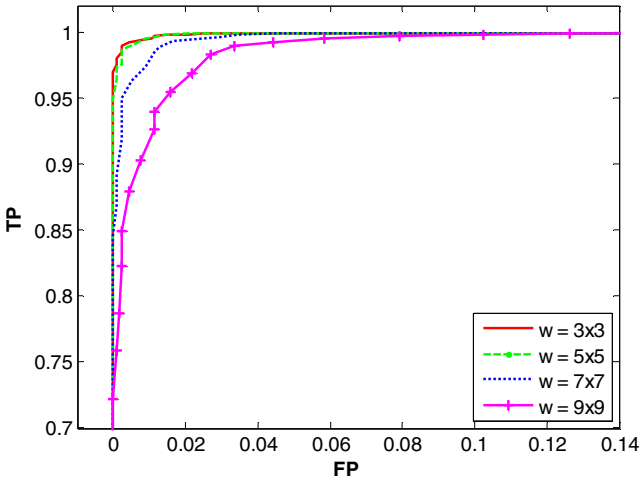


Fig. 5. ROC curve for the classification between original images and their median-filtered (5x5) versions.

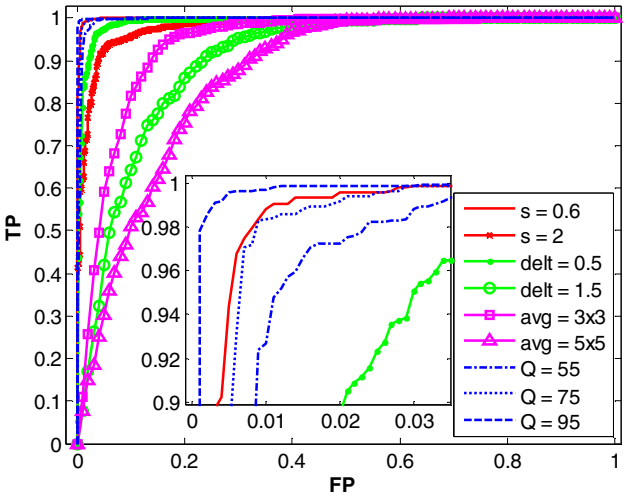


Fig. 6. ROC curve for the classification between the manipulated images and their median-filtered (5x5) versions.

The feature metric *f* extracted from each sample is indicated in Fig. 4. We can see that feature values of median filtered images are intuitively higher than those of original ones.

The probabilities of true positive (TP) and false positive (FP) are determined at a given threshold (ranges from 0 to 1) by calculating the percent of correctly classified median-filtered images and that of incorrectly classified unfiltered images respectively. The classification results are reported by receiver operating characteristics (ROC) curves, shown in Fig. 5. Perfect performance has been achieved, especially in the case of small window sizes, i.e. w=3x3, which does not incur obvious distortion on image quality.

In the second case, classification is performed between previously manipulated images (N) and their 5x5 median-filtered versions (P). Manipulations such as bilinear scaling, Gaussian filtering, average filtering and JPEG compression

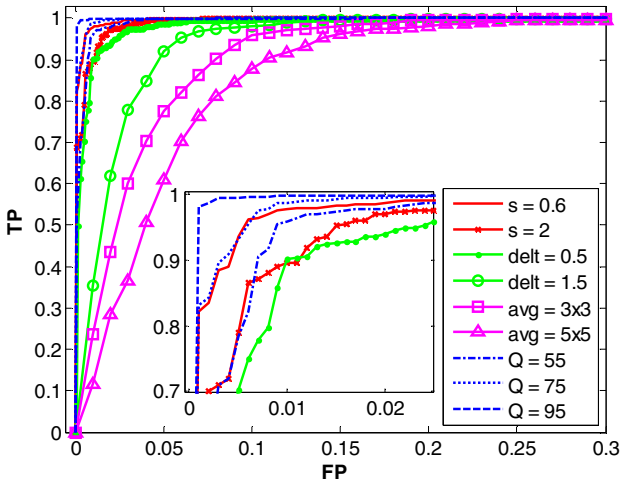


Fig. 7. ROC curve for the classification between median-filtered (5x5) images and the images processed by other manipulations.

are considered to simulate the various processing that one image may be suffered practically. The classification results are shown in Fig. 6. It can be seen that MF is harder to be detected on heavily Gaussian filtered and average filtered images than that on other manipulated images. However, not bad results can still be gained. For example, MF on Gaussian blurred (delt=1.5) images was detected accurately with probability TP=0.85 with a false positive probability of FP=0.2. Despite of such two types of manipulations, others nearly have none serious disturbance on MF detection, which can be verified by the high TP values under low FP.

It should be necessary to point out that MF detection on resampled images behaves well, which can be observed from the ROC curves corresponding to s=0.6 and s=2. Such results indeed verify the validity of our proposed forensics method on circumventing the recent resampling technique, which attacks the previous resampling detector by carrying out MF on interpolated images [14]. The success of MF detection in such situation is to raise new challenges for the anti-forensics techniques.

4.2. Distinguish MF from Other Manipulations

As a manipulation forensics, performance on differentiating MF from the other manipulations is required to be appraised. Features extracted from all manipulated sample images are displayed in Fig. 4, from which we can find the discrepancy intuitively. In this test, manipulated images (N) and 5x5 median-filtered images (P) are classified and the results are shown in Fig. 7. Satisfied effect for MF discrimination has been obtained. Although distinguishing MF from Gaussian and average filtering are more difficult than that from other manipulations, but the fingerprint feature still behaves well. For example, when FP=0.1, even for the worst case, average filtering avg=2, TP value can arrive at 0.87. In the other cases, TP always keeps higher than 0.95.

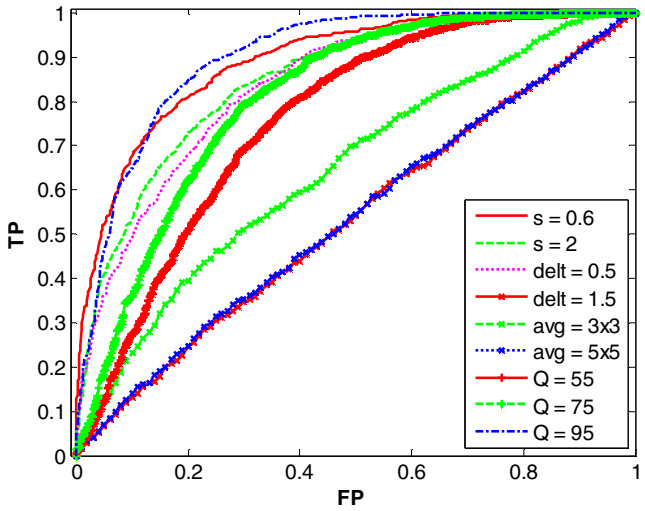


Fig. 8. ROC curve for the classification between the manipulated images and the post-manipulated median-filtered (5x5) images.

4.3. Assess Robustness of MF Detection

For assessing the robustness of our proposed MF detection method, the median-filtered (5x5) sample images' different post-manipulated versions (P) and the manipulated original images (N) are classified. Such a test is to investigate the performance of our proposed MF detection algorithm when different post-processing occurs.

Test results are shown in Fig. 8. The conclusion can be made from the ROC curves is that the MF detection scheme can resist the post operations to some extent, except the heavily Gaussian filtering and average filtering. Obvious interfere on the local pixel configuration may be the reason for the invalidation on filtering post-processing. On the contrary, the robustness resisting scaling is validated to be better than other operations, no matter for downscaling or upscaling. Besides, an inherent regulation which can be concluded is that the weaker the post-operation strength is, the better the detection accuracy can be achieved. Besides, it should be admitted that the proposed MF forensic algorithm is fragile to additional noise. However, post-processing is rarely applied after median filtering operation in practical applications, especially in the case of sophisticated counter-forensic techniques.

5. CONCLUSION

In this paper, we present an effective forensic algorithm to detect the median filtering manipulation, which is usually applied to erase the forensically significant fingerprints. Statistical characteristics of the median-filtered signal is analyzed and measured by the probability of zero value on the difference map of textured pixels. The median-filtered image is identified by perform thresholding adjudication on the fingerprint metric. A series of experiments are designed to test MF detection performance, differentiation capability and robustness respectively. Effectiveness of the proposed MF forensic detection scheme has been verified extensively.

6. REFERENCES

[1] Y.-F. Hsu and S.-F. Chang, "Image splicing detection using camera response function consistency and automatic segmentation," in International Conference on Multimedia and Expo, Beijing, 2007.

[2] W. Chen, Y. Q. Shi and W. Su, "Image splicing detection using 2-D phase congruency and statistical moments of characteristic function," SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA, USA, 2007.

[3] S. Bayram, H. T. Sencar and N. Memon, "An efficient and robust method for detecting copy-move forgery," in International Conf. on Acoustics, Speech and Signal Processing, Taipei, 2009.

[4] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," IEEE Transactions on Signal Processing, vol. 53, no. 2, pp.758-767, 2005.

[5] H. Farid, "Exposing digital forgeries from JPEG ghosts," IEEE Transactions on Information Forensics and Security, vol. 4, no. 1, pp.154-160, 2009.

[6] M. Stamm and K. J. R. Liu, "Blind forensics of contrast enhancement in digital images," in International Conference Proceedings on Image Processing, San Diego, 2008.

[7] M. Stamm and K. J. R. Liu, "Forensic estimation and reconstruction of a contrast enhancement mapping," in International Conf. on Acoustics, Speech and Signal Processing, Dallas, Texas, USA, 2010.

[8] D. Hsiao and S. Pei, "Detecting digital tampering by blur estimation," 1st International Workshop on Systematic Approaches to Digital Forensic Engineering, Washington, 2005.

[9] G. Cao, Y. Zhao and R. Ni, "Edge-based blur metric for tamper detection," Journal of Information Hiding and Multimedia Signal Processing, vol. 1, no. 1, pp.20-27, 2010.

[10] G. Cao, Y. Zhao and R. Ni, "Detection of image sharpening based on histogram aberration and ringing artifacts," in International Conference Proceedings on Multimedia and Expo, New York, 2009.

[11] S. Bayram, I. Avcubas, B. Sankur and N. Memon, "Image manipulation detection," Journal of Electronic Imaging, vol. 15, no. 4, pp. 04110201-04110217, 2006.

[12] W.-H. Chuang, A. Swaminathan and M. Wu, "Tampering identifycation using empirical frequency response," in International Conference Proceedings on Acoustics, Speech and Signal Processing, Taipei, 2009.

[13] A. E. Dirik and N. Memon, "Image tamper detection based on demosaicing artifacts," in International Conference Proceedings on Image Processing, Cairo, 2009.

[14] M. Kirchner and R. Böhme, "Hiding traces of resampling in digital images," IEEE Transactions on Information Forensics and Security, vol. 3, no. 4, pp.582-592, 2008.

[15] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA, USA, 2010.

[16] G. Schaefer and M. Stich, "UCID - An uncompressed colour image database," in Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia, San Jose, 2004.