On Detection of Median Filtering in Digital Images

Matthias Kirchner [a] and Jessica Fridrich [b]

[a] Technische Universität Dresden, Dept. of Computer Science, 01062 Dresden, Germany
[b] SUNY Binghamton, Dept. of Electrical and Computer Engineering, Binghamton, NY 13902

ABSTRACT

In digital image forensics, it is generally accepted that intentional manipulations of the image content are most critical and hence numerous forensic methods focus on the detection of such 'malicious' post-processing. However, it is also beneficial to know as much as possible about the general processing history of an image, including content-preserving operations, since they can affect the reliability of forensic methods in various ways. In this paper, we present a simple yet effective technique to detect median filtering in digital images—a widely used denoising and smoothing operator. As a great variety of forensic methods relies on some kind of a linearity assumption, a detection of non-linear median filtering is of particular interest. The effectiveness of our method is backed with experimental evidence on a large image database.

Keywords: digital forensics, median filter, processing history, image processing

1. INTRODUCTION

Digital image forensics has recently become a widely studied stream of research in multimedia security. Ubiquitous digital imaging devices and sophisticated editing software gave rise to the need for forensic toolboxes that can blindly assess the authenticity of digital images without access to the source image or source device[1,2] or the aid of an auxiliary watermark signal.[3] When reasoning about the authenticity of digital images, it is necessary to have at least a rough working definition of what constitutes a manipulation and what is considered to be a 'legitimate' post-processing.[4] It is generally accepted that intentional manipulations of the image content (e.g., copy & paste operations or image splicing) are more critical and hence numerous forensic methods focus on detection of such 'malicious' post-processing. However, it is also beneficial to know as much as possible about the general processing history of an image, including content-preserving operations, such as compression,[5] contrast enhancement,[6] sharpening,[7] and denoising.

Even though such image processing primitives typically do not harm the authentic value of an image, they are of interest in a forensic examination of an image since they can affect forensic methods in various ways. First, the actual state of an image prior to manipulation may influence the set of tools we are using to analyze the image or our interpretation of the evidence derived from these tools. This is related to the field of steganalysis, where, for instance, the choice of a suitable spatial-domain detector should be made conditional to the cover properties.[8] Second, certain post-processing steps may interfere with or diminish subtle traces of previous manipulations and thus decrease the reliability of forensic methods.

In the course of this paper, we shall focus on the median filter, a well-known denoising and smoothing operator.[9] In the line with what was mentioned above, we believe that a detection of median filtered images is of particular interest since a great variety of image forensic techniques rely on some kind of linearity assumption. Because median filtering is a highly non-linear operation, it is likely to affect the reliability of these methods. A typical example is the detection of resampling,[10] which employs a local linear predictor of pixel intensities and was shown to be vulnerable to median filtering.[11]

The rest of this paper is organized as follows: Starting from a short review of basic properties of the median filter in Sect. 2, we will center on the so-called streaking artifacts in Sect. 3 and show how this characteristic can actually be used to detect median filtering in bitmap images. Since forensic methods are generally desired to be robust against lossy post-compression, Sect. 4 will focus on detection of median filtering after JPEG compression. Both sections are underpinned by detailed experimental results from a large database of images. Finally, Sect. 5 concludes the paper.

Further author information: matthias.kirchner@inf.tu-dresden.de, fridrich@binghamton.edu

2. MEDIAN FILTERED IMAGES

Given a set of random variables $\mathcal{X} = (X_1, X_2, \ldots, X_N)$, the order statistics $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(N)}$ are random variables, defined by sorting the values of X_i in an increasing order. The median value is then given as

$$\mathsf{median}(\mathcal{X}) = \begin{cases} X_{(K+1)} = X_{(m)}\,, & \text{for } N = 2K+1 \\ 1/2\left(X_{(K)} + X_{(K+1)}\right)\,, & \text{for } N = 2K, \end{cases} \tag{1}$$

where $m = 2K + 1$ is the median rank. The median is considered to be a robust estimator of the location parameter of a distribution and has found numerous applications in smoothing and denoising, especially for signals contaminated by impulsive noise.[9]

For a grayscale input image with intensity values $x_{i,j}$, the two-dimensional median filter is defined as

$$y_{i,j} = \underset{(r,s)\in\mathcal{W}}{\mathsf{median}}(x_{i+r,j+s})\,,$$

where $\mathcal{W}$ is a window over which the filter is applied. For the rest of this paper, we assume symmetric square windows of size $M \times M$ with $M = 2L + 1$, i. e., the median rank m equals $m = (M^2 + 1)/2$. This is probably also the most widely used form of this filter.

In order to describe some characteristics of median filtered images and compare the median filter to other filters, it is useful to study the output distribution of the median filter. Due to its non-linearity, however, theoretical analysis of the general relation between the input and output distribution of the median filter is highly non-trivial. For this reason, it is often assumed that the input samples are i.i.d. The general cumulative distribution function (CDF) F_Y for output samples $y_{i,j}$ and i.i.d. input samples $x_{i,j}$ with CDF F_X is given by[12]

$$F_Y(y) = \sum_{k=m}^{M^2} \binom{M^2}{k} [F_X(y)]^k \left[1 - F_X(y)\right]^{M^2-k}\,,$$

A special yet interesting case is the sample median of i.i.d. input samples following a normal distribution, $x_{i,j} \sim \mathcal{N}(\mu, \sigma)$, which was shown to asymptotically (as $M \to \infty$) follow a normal distribution again,[13,14]

$$y_{i,j} \sim \mathcal{N}(\mu, \sigma_m)\,, \qquad \text{where } \sigma_m = \sqrt{\frac{\pi}{2}} \cdot \frac{\sigma}{M}\,.$$

Since, in filtered images, pixels in a close neighborhood originate from overlapping windows, they are correlated to some extent and thus the joint distribution of adjacent pixels is generally of interest. For an $M \times M$ median filter with i.i.d. input $F_X(x)$, Liao et al.[15] derive an expression for the bivariate distribution of two output pixels y_p and y_q (H pixels window overlap), $F_Y(y_p, y_q)$. The formula, which can be found in Appendix A, highlights how cumbersome the theoretical description of median filtered images can become even under the unrealistic assumption of i.i.d. pixel intensities.

For this reason, many studies in the literature have focused on more specific features of interest when analyzing the median filter. As such, the median filter was found to preserve edges better than, for instance, the moving average filter.[16] It is also known that median filtered images exhibit regions of constant or nearly constant intensities.[17] A further stream of research addresses the so-called roots of the median—signals which are invariant to median filtering—as well as the convergence of arbitrary signals to such roots.[18]

3. STREAKING ARTIFACTS

One of the main differences between the median filter and other types of linear and non-linear filters is that, for an odd filter dimension, its output samples are directly drawn from the set of input samples, cf. Eq. (1). For discrete-valued signals, this means, in particular, that no rounding to integers has to be performed after filtering. Because of overlapping filter windows, there exists a non-zero probability that the output pixels in a certain neighborhood originate from the same position of the input image. This effect is called *streaking* and

was quantitatively analyzed by Bovik.[17] For continuous-valued i.i.d. input samples, he derived expressions for the probability that two pixels with a certain distance have equal intensity. While being a function of the filter size, it turns out that these probabilities are independent of the actual distribution of the input. Tables with probabilities for different filter sizes and pixel distances can be found in the original publication.[17]

Obviously, the presence of such a specific 'probability pattern' would be a very strong indication of previous median filtering. However, while the reported distribution-independence of streaking artifacts in continuous-valued i.i.d. signals is based on the zero probability of two input samples being equal, typical digital images have discrete-valued pixel intensities drawn from a finite alphabet. Here, the streaking probabilities become distribution-dependent because the quantized intensities can *a priori* be equal-valued. The probability that two integer grayscale output pixels y_p, y_q have equal intensity can generally be written as

$$P_0 = \Pr(y_p = y_q) = \sum_i F_Y(i,i) - \sum_i F_Y(i-1,i) + F_Y(i,i-1) - F_Y(i-1,i-1)\,, \tag{2}$$

where, for i.i.d. input samples, F_Y is the joint distribution given in Eq. (10) in the appendix. Figure 1 demonstrates the distribution-dependence by plotting the probability P_0 for two adjacent output pixels (for instance the horizontal neighbors $y_{i,j}$ and $y_{i,j+1}$) as a function of the standard deviation of quantized i.i.d. Gaussian input samples. As to be expected, P_0 considerably increases for median filtered images. Due to its larger support, the 5×5 filter results in stronger artifacts than the 3×3 filter. Similar graphs can be obtained for two output pixels that are not directly adjacent but still originate from overlapping windows.

3.1. Measuring Streaking Artifacts in Digital Images

Streaking artifacts with respect to groups of two pixels can be well analyzed by means of their first-order difference. In the following, denote $d_{i,j}^{(k,l)}$ as the first-order difference image with lag (k,l), i. e.,

$$d_{i,j}^{(k,l)} = y_{i,j} - y_{i+k,j+l}\,, \tag{3}$$

and $\mathcal{H}^{(k,l)} = \{\dots, h_{-1}^{(k,l)}, h_0^{(k,l)}, h_1^{(k,l)}, \dots\}$ as the corresponding histogram of differences. (For notational convenience, we will omit the superscript (k,l) whenever the concrete values of the vertical and horizontal lag are not of prior interest.) Except for almost flat regions ($\sigma \approx 0$), the i.i.d. input example from Fig. 1 suggests that a clear distinction between filtered and non-filtered images can be obtained by simply determining the relative frequency $\tilde{h}_0 = h_0/|\mathbf{y}|$ as an estimate of P_0, where $|\mathbf{y}|$ is the number of pixels in image $\mathbf{y}$.

In real images, of course, pixels are generally neither i.i.d. nor Gaussian. Simple tests show that $\tilde{h}_0$ greatly varies for different images, depending on image content and characteristics. Figure 2 depicts density estimates of empirically determined relative frequencies $\tilde{h}_0^{(1,0)}$ from 6500 original never-compressed images and their 3×3 and 5×5 median filtered versions, respectively.* Although we can observe the principal effect of median filtering in terms of a shifted distribution, a general conclusion can not be drawn. A long right tail makes smooth originals indistinguishable from filtered images. Without concrete knowledge of the input distribution, $\tilde{h}_0$ is ultimately only some measure of (local) smoothness. Even though Liao et al.[15] also provide more general expressions of F_Y for arbitrary input distributions, it is hardly possible to make general assumptions on the distribution of the input pixels.

Except for saturated regions, smooth original images will however not only exhibit a large bin $\tilde{h}_0$, but generally a predominance of relatively small first-order differences. As a result, $\tilde{h}_{+1/-1}$ is typically comparable in height to $\tilde{h}_0$. Streaking due to median filtering, on the other hand, tends to increase particularly the bin $\tilde{h}_0$ relative to $\tilde{h}_{+1/-1}$. For this reason, we explore the ratio

$$\varrho^{(k,l)} = h_0^{(k,l)} \big/ h_1^{(k,l)} \tag{4}$$

as a detection statistic to distinguish filtered from non-filtered images. While we expect median filtered images to result in ratios $\varrho \gg 1$, non-saturated originals will yield rather small ratios $\varrho \approx 1$. Note that a strong relative

*A description of the image database is given in Sect. 3.2.

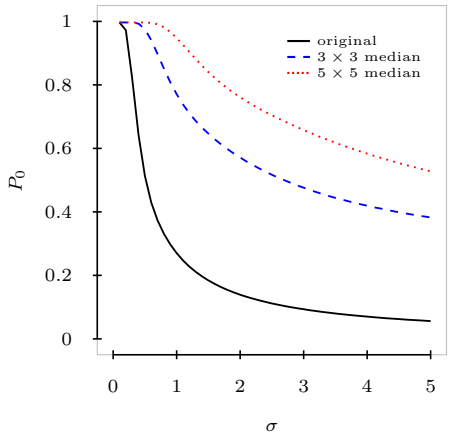


Figure 1. Streaking probabilities P_0 (direct vertical or horizontal neighbors) for quantized i.i.d. Gaussian samples input samples with variance σ^2.

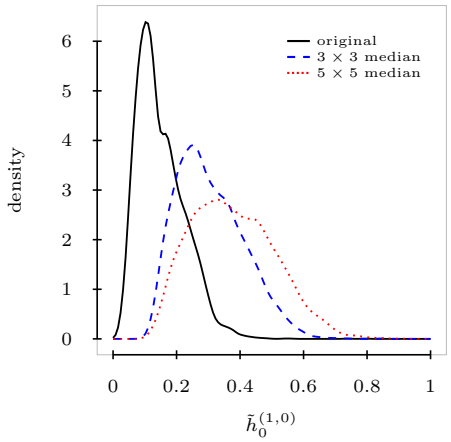


Figure 2. Density estimates for relative frequencies $\tilde{h}_0^{(1,0)}$ from 6500 original images and their 3×3 and 5×5 median filtered versions, respectively.

increase is indeed characteristic for the median filter, since, in contrast to conventional linear and non-linear smoothers, no rounding to integer values has to be performed after filtering.

Apparently, strong saturation effects in the original image will render the detection of median filtered images by means of ϱ unreliable. To obtain a more robust discriminating feature, we divide the image under investigation into the set $\mathcal{B}$ of non-overlapping blocks of dimension $B \times B$. By determining ϱ_b as the ratio of histogram bins h_0 and h_1 from the b-th block, the influence of saturated image blocks can be reduced by taking the weighted median

$$\hat{\varrho} = \underset{b \in \mathcal{B}}{\mathsf{median}} \left(w_b \varrho_b \right) , \tag{5}$$

as a robust detection feature. Here, the weights w_b function as an attenuation factor for saturation effects. In the course of this paper, we set w_b to

$$w_b = 1 - \left(\frac{h_0}{B^2 - B} \right) , \tag{6}$$

giving less weight to strongly saturated blocks.

3.2. Experimental Results

For an experimental evaluation of the proposed detection feature, we make use of a database of altogether 6500 never-compressed RGB images, stemming from 22 different digital cameras with resolutions in the range of 1000×1500 to 2000×3000. All tests are carried out after converting each image to grayscale prior to any further processing or analysis. Whenever not made explicit, our detectors were run on full-size images with first-order differences with lag $(k, l) = (1, 0)$, i.e., the differences of direct vertical neighbors.

To demonstrate the principal effectiveness of our relatively simple discriminating feature, Fig. 3 depicts ROC curves compiled from each 6500 original and 3×3 median filtered images. More specifically, graphs obtained using ϱ, i.e., ratios of difference-histogram bins from the full-sized images, and $\hat{\varrho}$, i.e., the median of ratios of difference-histogram bins from $|\mathcal{B}|$ 64×64 blocks, are displayed. The ROC curves indicate that the block-based approach is indeed more robust to false alarms, giving a perfect detection for a false positive rate of $< 1.8\,\%$.

While Fig. 3 indicates that the weighted median of the block measures ϱ_b is superior to the full-image measure ϱ, it is important to note that ϱ_b itself is more sensitive to local variations throughout the image. This is important whenever a *per*-block decision is required. Figure 4 reports ROC curves for a block-wise detection of 3×3 median filtering for different block sizes. The curves were obtained by determining ϱ_b for all non-overlapping blocks of all images in the database. In general, it can be observed that the overall performance increases with increasing block sizes.

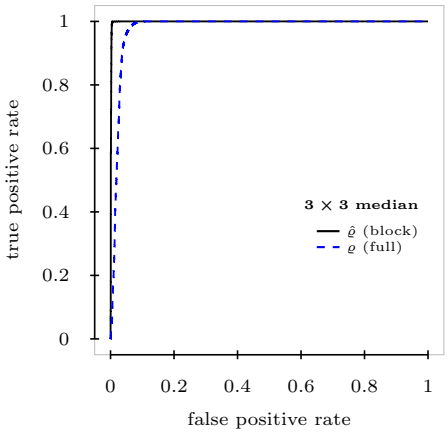


Figure 3. Detection results for 3×3 median filtering. ROC curves for ϱ and $\hat{\varrho}$, ($B = 64$). The block based approach is more robust to false alarms.

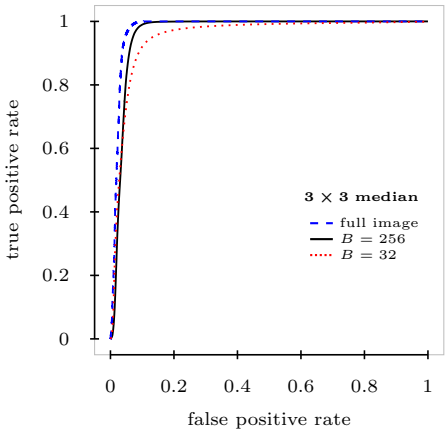
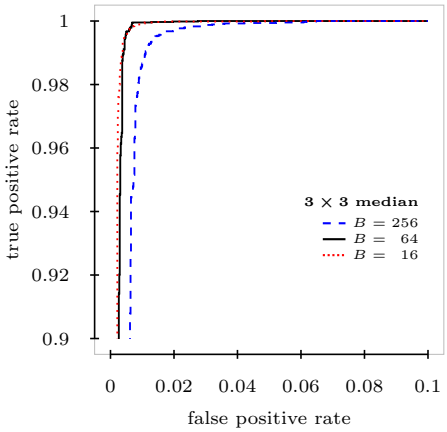


Figure 4. Detection results for 3×3 median filtering. ROC curves for ϱ and ϱ_b for varying block sizes. Smaller blocks are less robust to local image characteristics.



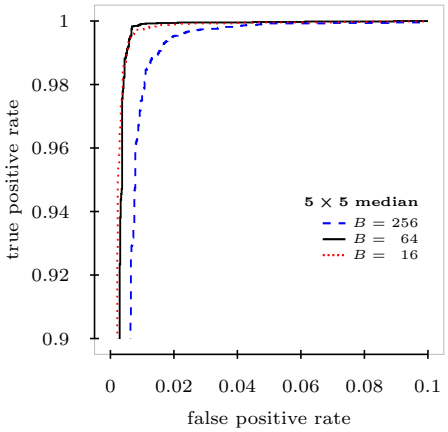


Figure 5. Detection results for 3×3 (left) and 5×5 (right) median filtering, respectively. ROC curves for different block sizes B. Vertical differences with lag $(k, l) = (1, 0)$. Smaller block sizes increase detection performance. (Note the different scaling of the axes compared to all other figures.)

The opposite is true when we consider the robust estimate $\hat{\varrho}$ over all $B \times B$ blocks of the image. The weighted median effectively attenuates too strong influences of local image characteristics, such as saturation. Figure 5 compares ROC curves for $\hat{\varrho}$ for the detection of 3×3 (left) and 5×5 (right) median filtering, respectively, for varying block sizes. Observe how the detectability grows with the decreasing block size B. However, while the advantage of using blocks with $B = 64$ over those of dimension 256×256 is considerable, the performance gain for smaller block sizes is diminishing. Our experiments showed that $B = 64$ is a reasonable compromise between detectability and the computing time needed to determine the median over all blocks.

In general, we found that there is a fair amount of variation in the values of $\hat{\varrho}$, mainly reflecting the different sources of our test images. Besides the already mentioned saturation effects, for instance, the noise level of the digital camera can be a very influential factor. Figure 6 gives an idea of how $\hat{\varrho}$ is distributed over our test database by plotting the values for all original and equivalent 3×3 median filtered images. The best distinction between originals and filtered images was found for the approximately 1600 images corresponding to the rightmost part of Fig. 6. These images are from Ker's 'gold standard' image set,[8] which was built from RAW camera images by switching any denoising off. As a consequence, $\hat{\varrho}$ is particularly low for the originals. A visual inspection of those originals with a relatively large $\hat{\varrho}$-value revealed the presence of large homogeneous regions. Despite the

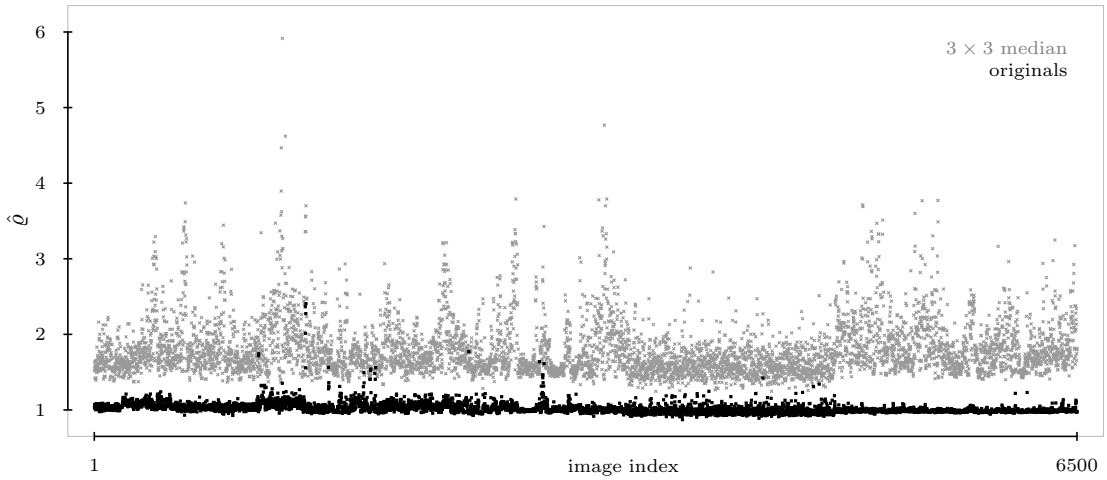


Figure 6. $\hat{\varrho}$-values for each 6500 original (solid squares) and 3×3 median filtered images (gray crosses), ($B = 64$). The variation in the measured values reflects the different sources of the images.

Table 1. Minimum average decision error P_e of median filtering detectors ϱ and $\hat{\varrho}$ for filter sizes 3×3 and 5×5. The results were obtained for 6500 never-compressed images and $(k, l) = (1, 0)$.

	ϱ	$\hat{\varrho}$ with block size B				
	(full)	256	128	64	32	16
3×3 median	0.039	0.009	0.005	0.004	0.004	0.004
5×5 median	0.040	0.012	0.006	0.004	0.004	0.005

generally very good detection performance, this might indicate that the weighting factors w_b could be further refined in future work.

Since streaking artifacts are believed to be a specific characteristic of median filtered images, our detection feature should be well able to distinguish median filtered images from those processed with alternative smoothers. Figure 7 reports exemplarily detection results for this scenario. More specifically, the ROC curves for the discrimination between 3×3 median filtered images and images processed with a 3×3 mean filter, as well as a 3×3 binomial filter are depicted. The curves indicate that streaking is indeed characteristic for the median filter, whereas a comparison with the mean filter gives slightly worse performance.

Note that we also investigated alternative lags in the computation of the first-order differences. Here, differences between pixels from a direct neighborhood, i.e., $|k|, |l| \leq 1$, generally discriminated better than larger lags, whereas horizontal/vertical differences always yielded preferable detection results. As to be expected, we could not find a considerable deviation from the above reported results among all of the direct horizontal and vertical differences $(k, l) \in \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$.

In general, all of the above reported results indicate that our relatively simple detection feature based on the ratio of first-order difference-histogram bins is a very reliable measure to detect median filtering in bitmap images. Table 1 summarizes our findings in terms of the minimum average decision error under the assumption of equal priors and equal costs,

$$P_e = \min {}^1\!/_2\big(P_{\mathrm{FP}} + (1 - P_{\mathrm{TP}})\big) \;, \tag{7}$$

where P_{FP} and P_{TP} denote the false positive and true positive rates, respectively. However, we admittedly have to say that bitmap images are only half of the overall picture. In the following, we will therefore discuss why $\hat{\varrho}$ is not robust to JPEG compression and present an alternative approach.

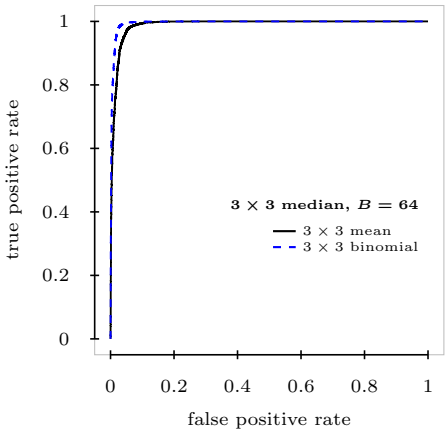


Figure 7. Detection results for 3×3 median filtering, taking images obtained by 3×3 mean filtering and 3×3 binomial filtering as 'originals' ($B = 64$). Strong streaking artifacts are not present in linearly smoothed images.

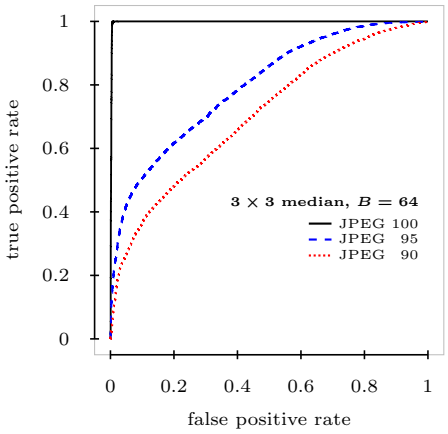


Figure 8. Detection results for 3×3 median filtering after JPEG post-compression with varying compression qualities ($B = 64$). JPEG qualities below 100 render median filtering undetectable.

4. MEDIAN DETECTION AFTER JPEG POST-COMPRESSION

As digital images are often stored in the JPEG format after processing, a forensic method is generally desired to be robust against lossy JPEG compression. It is easy to see that this is not the case for our detection feature $\hat{\varrho}$. Depending on the JPEG compression quality, quantization smooths out the subtle inter-pixel relations on which we based our decision. Moreover, even moderate JPEG compression might already introduce pixel neighborhoods of constant intensity leading to a high number of false alarms.

To give an impression of how much $\hat{\varrho}$ fails to detect median filtering after post-compression, Fig. 8 depicts ROC curves for the 3×3 filter and varying JPEG post-compression qualities. While the quality of 100 has virtually no effect, qualities of 95 or below make the detection impossible. These results apparently call for an alternative detection procedure for JPEG images.

4.1. SPAM Features

While the discrimination based on the ratio h_0/h_1 is generally not possible under JPEG post-compression, first-order differences still form a useful basis for distinguishing filtered from non-filtered images. Since the median filter is a smoothing operator, it affects the distribution of first-order differences not only by increasing h_0 but also by making the histogram of differences more peaky in general. This effect is further enhanced by quantization during JPEG post-compression, which indicates that median filtering should be well detectable after JPEG compression.

In a recent paper, Pevný et al.[19] introduced the *subtractive pixel adjacency matrix* (SPAM) features as a means to analyze the conditional joint distribution of first-order difference images. While the main application in Ref. 19 was to measure the effects of ± 1 steganography (which leads to more noisy images and thus a wider distribution), we believe that the SPAM features are also a valuable resource to detect median filtering. As described above, the effect that is to be captured here is exactly the opposite, i.e., a sharper distribution shape due to median filtering.

The SPAM features are obtained by modeling the horizontal, vertical, and diagonal first-order differences $d_{i,j}^{(k,l)}$ with lags $(k,l) \in \{-1,0,1\}^2$ as n-th order Markov chains. More specifically, the transition matrices $\mathbf{M}^{(k,l)}$ are computed for each of the 8 possible lags, where the matrix elements are given by the transition probabilities:

$$M_{\delta_n,\ldots,\delta_0}^{(k,l)} = P\left(d_{i+kn,j+ln}^{(k,l)} = \delta_n \,\middle|\, d_{i+k(n-1),j+l(n-1)}^{(k,l)} = \delta_{n-1}, \ldots, d_{i,j}^{(k,l)} = \delta_0\right). \tag{8}$$

Since median filtering affects, in particular, small absolute differences, it is feasible to limit the range of considered differences to $|\delta_n|, \ldots, |\delta_0| \leq T$. This also helps keep the dimensionality of the model low, which is important for

practical implementation.[19] These transition probabilities are then taken to form a D-dimensional feature vector $\mathbf{F}$, $D = 2 \cdot (2T+1)^{n+1}$, by averaging the four horizontal/vertical and the four diagonal matrices, respectively:

$$\mathbf{F} = (\mathbf{F}^{(\mathsf{h/v})}, \mathbf{F}^{(\mathsf{d})}) \quad \text{with elements} \quad \begin{cases} F^{(\mathsf{h/v})}_{\delta_n,\ldots,\delta_0} = 1/4 \left(M^{(1,0)}_{\delta_n,\ldots,\delta_0} + M^{(-1,0)}_{\delta_n,\ldots,\delta_0} + M^{(0,1)}_{\delta_n,\ldots,\delta_0} + M^{(0,-1)}_{\delta_n,\ldots,\delta_0} \right) \\ F^{(\mathsf{d})}_{\delta_n,\ldots,\delta_0} = 1/4 \left(M^{(1,1)}_{\delta_n,\ldots,\delta_0} + M^{(-1,-1)}_{\delta_n,\ldots,\delta_0} + M^{(1,-1)}_{\delta_n,\ldots,\delta_0} + M^{(-1,1)}_{\delta_n,\ldots,\delta_0} \right) \end{cases} . \quad (9)$$

Note that, in some sense, the SPAM features can be seen as a generalization of the ϱ-based approach from the previous section, where first-order differences are modeled as a zeroth-order Markov chain.

Once the features have been calculated, they are fed into a suitable classifier trained on a set of original (JPEG) images and median filtered images.

4.2. Experimental Results

To evaluate the suitability of the SPAM features for detecting median filtering, we use the same database of 6500 images as in Sect. 3.2. Throughout all experiments, we employ as a classifier *soft-margin support vector machines* (SVMs)[20] with Gaussian kernel $k(x,y) = \exp(-\gamma \|x-y\|^2)$ trained and tested on the $N \times N$ center region of the grayscale images.[†] One classifier is trained for each analyzed filter and JPEG post-compression quality by randomly bisecting the corresponding set of original (JPEG) images and median filtered images into a training and testing subset. The classifiers are parametrized by two hyper-parameters, γ and C,[20] determined from the training set by a search over all possible pairs on the following multiplicative grid:

$$C \in \left\{ 10^c \,\middle|\, c \in \{-3,-2,\ldots,4\} \right\} \;\times\; \gamma \in \left\{ 2^{-\nu} \,\middle|\, \nu \in \log_2 D + \{-4,-3,\ldots,4\} \right\}$$

using five-fold cross-validation. In the standard setup, we use SPAM features corresponding to a second-order Markov model, extracted from images of size 512×512 with a threshold $T = 3$.

Figure 9 shows the ROC curves for the detection of 3×3 median filtering using SPAM features and soft-margin SVMs as described above. More specifically, the top left graph depicts the results for JPEG post-compression qualities in the range of $\{90, 80, 70\}$. The detection is relatively very reliable for moderate JPEG compression. As expected, the detection performance decreases when lowering the JPEG quality because stronger quantization increases the number of small absolute first-order differences in the original images.

The detection results obtained from the SPAM classifier are subject to a number of different parameter choices. The threshold T controls the magnitude of the maximum absolute differences which are to be taken into consideration. To see how this parameter influences the detection performance, the top right graph of Fig. 9 shows the ROC curves for a fixed JPEG compression with quality 80 and $T \in \{1,2,3\}$. While the results for $T = 3$ and $T = 2$ are comparable (with $T = 2$ giving slightly worse results), the detectability considerably drops when setting $T = 1$. This is again expected because a lower threshold limits the classifier's models of original and filtered images, giving rise to ambiguities. Similar results were obtained for different post-compression qualities as well.

Since median filtering is expected to affect horizontal/vertical and diagonal differences in a similar manner, we further tested whether the diagonal features $\mathbf{F}^{(\mathsf{d})}$ provide some additional information beyond what is already captured by the horizontal/vertical features $\mathbf{F}^{(\mathsf{h/v})}$. As indicated by the bottom left graph of Fig. 9, for a fixed JPEG post-compression quality 80, the detectability is indeed more or less invariant to keeping/removing the diagonal features. Table 2 presents the condensed results in terms of the minimum decision error P_e for varying JPEG qualities and suggests that analyzing the averaged horizontal and vertical transition probabilities is sufficient as long as $T > 1$.

In our last experiment, we investigate the influence of the size of the analyzed image (region). Because the SPAM features are basically an estimate of the transition matrix of first-order differences, it is to be expected

[†]The cropping is done for reasons of keeping the computing time for the feature extraction practicable. Since the images are cropped from the center part of the full size image, there might be a bias towards less saturated images (assuming that saturation mostly occurs in the uppermost part of an image). On the other hand, the relatively high dimensionality of the SPAM features should be generally well able to cope with saturation effects.

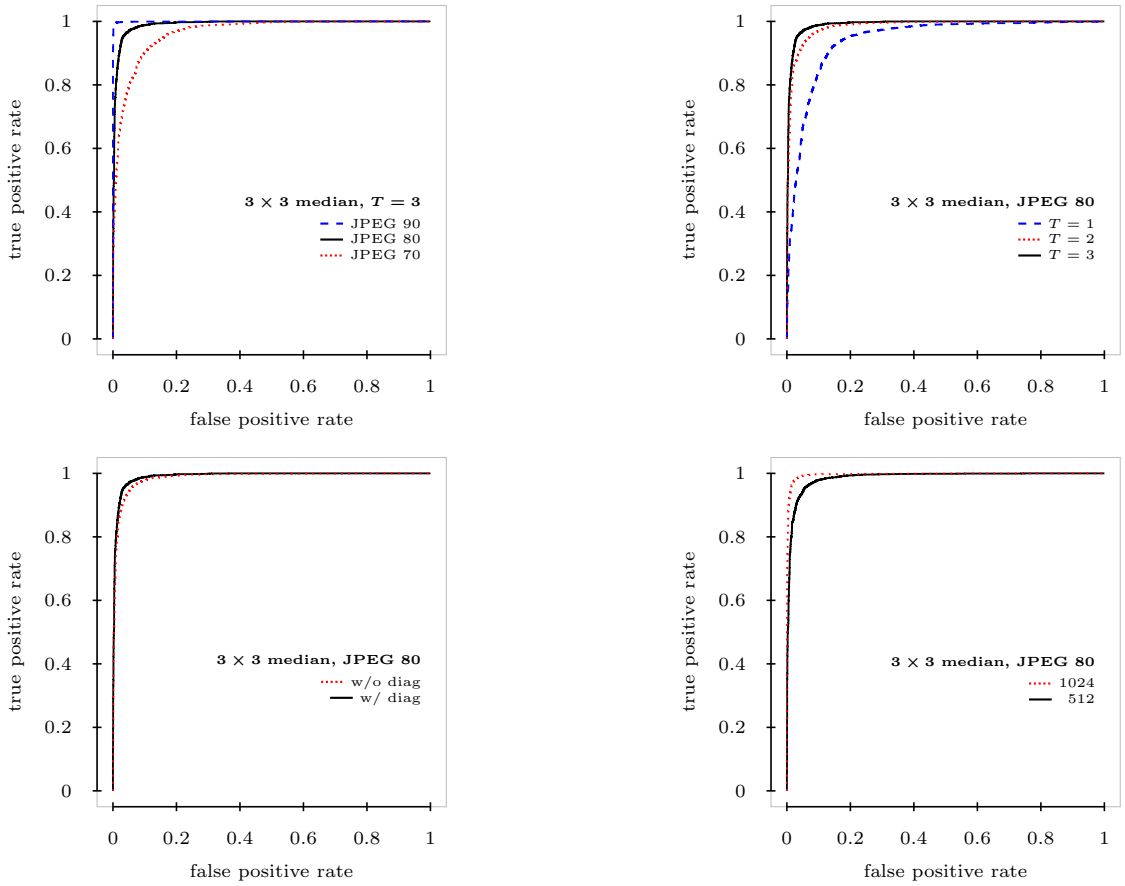


Figure 9. SPAM results for the detection of 3×3 median filtering with JPEG post-compression. ROC curves for different post-compression qualities (top left; $T = 3$, 512×512 images, horizontal/vertical and diagonal features); as well as curves for a fixed JPEG quality 80 and varying parameter settings: threshold T (top right), with or without diagonal features (bottom left), and image size (bottom right). Detection performance decreases with decreasing JPEG quality, SPAM threshold T and image size. It is however more or less invariant to ignoring the diagonal SPAM features.

that these features become more robust to local variations as the image size increases (also see the discussion of the block size in Sect. 3.2). Such local variations might be, for instance, caused by saturated or highly textured regions. The bottom right graph of Fig. 9 shows the ROC curves for a JPEG compression quality 80 and images of size 512×512 and 1024×1024, respectively. In line with our expectations, a better performance is obtained for the larger images. Again, corresponding results were obtained for different JPEG compression qualities.

While all of the above findings for the 3×3 median filter also hold for larger filter dimensions, the detection performance generally increases considerably. This is so because larger filters have a stronger smoothing effect and thus affect the distribution of small absolute differences more severely. To demonstrate the gain in detectability for a larger filter, Fig. 10 reports some sample results for a 5×5 filter. An almost perfect discrimination between original (JPEG) images and filtered images was obtained for different JPEG compression qualities in the range $\{90, 80, 70\}$. The zoomed-in part reveals that the detection performance decreases only slightly when lowering the JPEG quality (also see Tab. 2). In agreement with the results from Fig. 9, only the horizontal/vertical features where used for classification. However, as indicated by the right graph of Fig. 10, T should still be chosen to be at least $T = 2$. Otherwise, feature vectors with a dimensionality that is too low make the detection unreliable.

Finally, it is worth noting that the first-order SPAM features by and large yield results in line with the presented findings for the second-order features, however at a generally reduced detectability. For the sake of brevity, we refrain from reporting them separately.

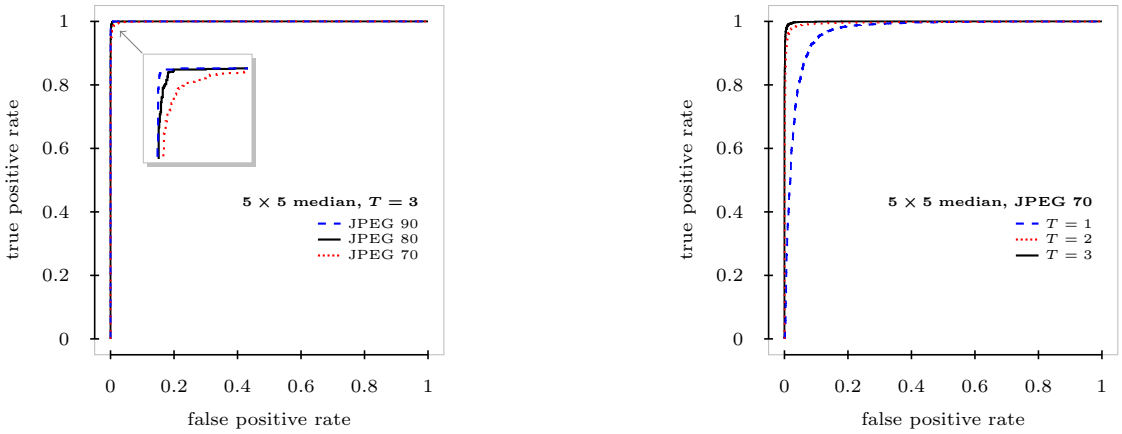


Figure 10. SPAM results for the detection of 5×5 median filtering with JPEG post-compression. ROC curves for different post-compression qualities (left; $T = 3$, 512×512 images, horizontal/vertical features only) and for a fixed JPEG quality 70 and varying threshold settings (right). Detection performance is generally better compared to 3×3 filtering.

On a more general level, it has to be mentioned that—contrary to the streaking-based approach of Sect. 3.1—the SPAM features do *not* solely capture the specific effects of median filtering. After JPEG compression, other linear and non-linear smoothers will have a similar (more or less strong) impact on the distribution of the first-order differences. While this proliferates the ambiguities in the determination of the concrete pre-processing history, especially when the JPEG quality becomes lower, it might be argued that after strong-enough compression it is sufficient to know that an image has been smoothed before because typical filter characteristics are suppressed by JPEG artifacts anyway. On the other hand, offering a more positive focus, SPAM features may evolve to become a general-purpose smoothing detector—a subject of future research.

5. CONCLUDING REMARKS

In this paper, we have investigated the detection of median filtering in digital images. In the broader framework of digital image forensics, we see this endeavor as a contribution to the problem of determining the general processing history of digital images. While the application of 'classical' image processing primitives for denoising, sharpening, or contrast enhancement does typically not *per se* harm the authentic value of an image, it is still of high interest to learn as much as possible about what exactly has happened to an image and to make informed decisions based on this knowledge. As such knowledge is desirable not only in forensics but also in steganalysis and watermarking, we deem our methods as valuable instruments in various fields of multimedia security.

The presented findings of this paper are twofold. For *uncompressed images*, the analysis of the so-called streaking artifacts[17] in median filtered images has proven to be a reliable measure for discriminating between filtered and non-filtered images. A perfect detection was achieved for false positive rates as low as 1.8 % (3×3 median, $B = 64$). While the detector is of splendid simplicity—relying on a single feature derived from histogram bins of the first-order difference image—it turned out that it is not applicable to images that are JPEG compressed after filtering. In the *post-compression scenario*, we therefore turned to a more complex detector based on the recently introduced SPAM features,[19] combined with support vector machines. Here, depending on the size of the median filter, a very reliable detection was possible even for JPEG qualities as low as 70 ($P_e = 1.1\,\%$ for the 5×5 median, $T = 3$).

In the case of median pre-compression (i.e., median filtering of already JPEG compressed images), which was not explicitly discussed in the paper, we found that the variation of the $\hat{\varrho}$-values between different original images is reduced to some extent. After the second compression, the SPAM features are still able to detect median filtering reliably. In fact, a low pre-compression quality can even increase the detector's performance.

As to the limitations, we have to note that JPEG compression does a good job in obfuscating the actual type of smoothing applied to the image before compression. While being generally well-detectable with the SPAM

Table 2. Minimum average decision error P_e of median filtering detectors based on SPAM for filter sizes 3×3 and 5×5. All results were obtained on a set of approximately 3250 JPEG images of dimension 512×512. The detectors' performance is broken down by the JPEG post-compression quality and the feature vector dimension (the threshold T, with or without diagonal features $\mathbf{F}^{(\mathsf{d})}$).

	JPEG 90			JPEG 80			JPEG 70		
	$T=3$	$T=2$	$T=1$	$T=3$	$T=2$	$T=1$	$T=3$	$T=2$	$T=1$
3×3 median									
$\mathbf{F}^{(\mathsf{h}/v)}$	0.010	0.021	0.068	0.051	0.078	0.146	0.105	0.133	0.249
$\mathbf{F}^{(\mathsf{h}/v)}, \mathbf{F}^{(\mathsf{d})}$	0.006	0.015	0.047	0.038	0.059	0.111	0.097	0.125	0.205
5×5 median									
$\mathbf{F}^{(\mathsf{h}/v)}$	0.002	0.003	0.008	0.004	0.009	0.025	0.011	0.021	0.077
$\mathbf{F}^{(\mathsf{h}/v)}, \mathbf{F}^{(\mathsf{d})}$	0.001	0.002	0.008	0.006	0.007	0.020	0.008	0.017	0.057

features, experiments showed that, contrary to the analysis of streaking artifacts in uncompressed images, it is not possible to distinguish between the median filter and other smoothers. While this could also be turned into an advantage by considering the SPAM features as a general purpose smoothing detector, alternative or additional features should be explored that allow to track down further particularities of the median filter. A possible candidate is, for instance, the median filter's relatively good edge-preserving property compared to linear smoothers.

ACKNOWLEDGMENTS

The authors want to thank Jan Kodovský for his help with setting up the SVMs. Matthias Kirchner gratefully receives a doctorate scholarship from Deutsche Telekom Stiftung, Bonn, Germany. Part of this research has been accomplished while the first author was a visiting scholar at SUNY Binghamton. Jessica Fridrich was supported by an NSF award CNF-0830528.

APPENDIX A. BIVARIATE OUTPUT DISTRIBUTION OF THE MEDIAN FILTER

For an $M \times M$ median filter with i.i.d. input $F_X(x)$, the joint distribution of two output pixels y_p and y_q (H pixels window overlap), $F_Y(y_p, y_q)$, is given by[15]

$$F_Y(y_p, y_q) = \begin{cases} \displaystyle\sum_{\nu=0}^{H} S_{\nu,H}(y_q) \sum_{\kappa=m-\nu}^{M^2} S_{\kappa,M^2-H}(y_q) \sum_{\lambda=m}^{M^2-H-\nu\nu} \sum_{\gamma=0} S_{\lambda-\gamma,M^2-H}(y_p) S_{\gamma,\nu}(y_p \mid y_p \le y_q)\,, & \text{for } y_p \le y_q \\ \displaystyle\sum_{\nu=0}^{H} S_{\nu,H}(y_q) \sum_{\kappa=m-\nu}^{M^2} S_{\kappa,M^2-H}(y_q) \sum_{\lambda=m}^{M^2} \sum_{\gamma=\nu}^{H} S_{\lambda-\gamma,M^2-H}(y_p) S_{\gamma-\nu,H-\nu}(y_p \mid y_p > y_q)\,, & \text{else,} \end{cases} \tag{10}$$

where $m = (M^2+1)/2$ and

$$S_{a,b}(y) = \binom{b}{a} [F_X(y)]^a [1 - F_X(y)]^{b-a}\,,$$

$$S_{a,b}(y_p \mid y_p \le y_q) = \binom{b}{a} [F_{X|X\le y_q}(y_p)]^a [1 - F_{X|X\le y_q}(y_p)]^{b-a}\,,$$

$$S_{a,b}(y_p \mid y_p > y_q) = \binom{b}{a} [F_{X|X> y_q}(y_p)]^a [1 - F_{X|X> y_q}(y_p)]^{b-a}\,.$$

REFERENCES

1. H. T. Sencar and N. Memon, "Overview of state-of-the-art in digital image forensics," in *Algorithms, Architectures and Information Systems Security*, B. B. Bhattacharya, S. Sur-Kolay, S. C. Nandy, and A. Bagchi, eds., *Statistical Science and Interdisciplinary Research* **3**, ch. 15, pp. 325–348, World Scientific Press, 2008.
2. H. Farid, "Image forgery detection," *IEEE Signal Processing Magazine* **26**(2), pp. 16–25, 2009.
3. I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, Second edition, Morgan Kaufmann, New York, 2008.
4. R. Böhme, F. Freiling, T. Gloe, and M. Kirchner, "Multimedia forensics is not computer forensics," in *Computational Forensics, Third International Workshop, IWCF 2009, The Hague, Netherlands, August 2009, Proceedings*, Z. J. Geradts, K. Y. Franke, and C. J. Veenman, eds., *Lecture Notes in Computer Science* **LNCS 5718**, pp. 90–103, Springer Verlag, (Berlin, Heidelberg), 2009.
5. R. Neelamani, R. de Queiroz, Z. Fan, S. Dash, and R. G. Baraniuk, "JPEG compression history estimation for color images," *IEEE Transactions on Image Processing* **15**(6), pp. 1365–1378, 2006.
6. M. Stamm and K. J. R. Liu, "Blind forensics of contrast enhancement in digital images," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP 2008)*, pp. 3112–3115, 2008.
7. G. Cao, Y. Zhao, and R. Ni, "Detection of image sharpening based on histogram aberration and ringing artifacts," in *Proceedings of the 2009 IEEE International Conference on Multimedia and Expo (ICME 2009)*, pp. 1026–1029, 2009.
8. A. D. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," in *Proceedings of SPIE-IS&T Electronic Imaging: Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, E. J. Delp, P. W. Wong, J. Dittmann, and N. Memon, eds., **6819**, p. 681905, 2008.
9. I. Pitas and A. N. Venetsanopoulos, "Order statistics in digital image processing," *Proceedings of the IEEE* **80**(12), pp. 1893–1921, 1992.
10. A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Transactions on Signal Processing* **53**(2), pp. 758–767, 2005.
11. M. Kirchner and R. Böhme, "Hiding traces of resampling in digital images," *IEEE Transactions on Information Forensics and Security* **3**(4), pp. 582–592, 2008.
12. H. A. David, *Order Statistics*, Wiley, New York, 1970.
13. H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
14. J. T. Chu, "On the distribution of the sample median," *The Annals of Mathematical Statistics* **26**(1), pp. 112–116, 1955.
15. G.-Y. Liao, T. A. Nodes, and N. C. Gallagher, "Output distributions of two-dimensional median filters," *IEEE Transactions on Acoustics, Speech and Signal Processing* **33**(5), pp. 1280–1295, 1985.
16. A. C. Bovik, T. Huang, and D. Munson, "The effect of median filtering on edge estimation and detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**(2), pp. 181–194, 1987.
17. A. C. Bovik, "Streaking in median filtered images," *IEEE Transactions on Acoustics, Speech and Signal Processing* **35**(4), pp. 493–503, 1987.
18. J. Astola, P. Heinonen, and Y. Neuvo, "On root structures of median and median-type filters," *IEEE Transactions on Acoustics, Speech and Signal Processing* **35**(8), pp. 1199–1201, 1987.
19. T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," in *MM&Sec'09, Proceedings of the Multimedia and Security Workshop 2009, September 7-8, 2009, Princeton, NJ, USA*, pp. 75–84, ACM Press, (New York, NY, USA), 2009.
20. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007.