Image Manipulation Detection Using Sensor Linear Pattern

Miroslav Goljan, Jessica Fridrich, and Matthias Kirchner, Department of ECE, SUNY Binghamton, NY, USA
{mgoljan,fridrich,kirchner}@binghamton.edu

Abstract

In this work, we introduce a new method for localizing image manipulations in a single digital image, such as identifying added, removed (spliced or in-painted), or deformed objects. The method utilizes the so-called Linear Pattern (LP) of digital images as a global template whose integrity can be assessed in a localized manner. The consistency of the linear pattern estimated from the image noise residual is evaluated in overlapping blocks of pixels. The manipulated region is identified by the lack of similarity in terms of the correlation coefficient computed between the power spectral density (PSD) of the LP in that region and the PSD averaged over the entire image. The method is potentially applicable to all images of sufficient resolution as long as the LP in the unmodified parts of the image has different spectral properties from that in the tampered area. No side information, such as the EXIF header or the camera model, is needed to make the method work. Experiments show the capability and limitations of the proposed method, which is robust to mild JPEG compression.

Introduction

The sophistication, availability, and ease of use of advanced image editing software coupled with increasingly more powerful multi-core processors available already in mobile imaging devices mean that digital content is nowadays easy to alter even for casual users of technology. Digital image forensics aims to reestablish trust in digital content by designing techniques capable of identifying regions in images that have undergone an alteration [3]. Early methods to accurately localize tampered regions were based on a pre-embedded fragile watermark [4, 13, 15, 17, 20]. Watermarking, however, is unlikely to be widely applicable for forensic applications for a number of reasons. Most images are not protected by a watermark, the watermark inevitably degrades image quality, and watermarking millions of images is expensive. Lukas et al. [16] showed that digital images contain a "natural" watermark, an intrinsic global signal introduced by the imaging sensor known as "fixed pattern noise" or "camera fingerprint" with its major component, the photo-response non-uniformity (PRNU) noise. The lack of the camera fingerprint in a region indicates that it has been manipulated [7, 8, 5]. Testing for the presence of a camera fingerprint requires, just as it is the case with fragile watermarks, access to (an estimate of) the actual fingerprint. This can be problematic in practical situations, especially when images of unknown provenance ought to be analyzed in large scale. Testing for the presence and consistency of other intrinsic signals or patterns that are naturally present in digital images can mitigate this sometimes limiting constraint. Color filter array artifacts [11], the desynchronization of color channels due to optical defects [18, 21], (inconsistencies in) general noise properties [10], or JPEG compression artifacts [2, 1] can be analyzed without the need for a camera-specific reference signal, for instance.

In the same general context, the method described in this paper makes use of a subtle signal intrinsically present in digital images, the so-called *linear pattern*. Cameras leave the linear pattern (LP) in images during sensor signal readout, color interpolation, and subsequent compression. While it has been previously discussed as a nuisance signal in the framework of camera identification [6] (where its removal is instrumental to control false alarms), we revisit the characteristic linear pattern here for manipulation localization from a single image without access to the camera that took the host image or other images taken by that camera.

We wish to point out that since no single digital forensic method will work universally, the only hope to achieve a reliable *automated* manipulation detection will require fusing the outputs of many forensic tools based on different assumptions [12]. The technique proposed in this paper is thus another tool to be added to the "forensic toolbox". Because regions falsely detected as manipulated can render a method unreliable and difficult to use, we pay close attention to control the false-alarm rate of our algorithm.

The rest of this paper is organized as follows. In the next section, we introduce notation and define the concept of a linear pattern for color and grayscale images while pointing out some of its properties. The forgery detection method itself is described in the third section with a few illustrative examples of forgery localization. Numerous experiments are presented in the fourth section also dealing with the important issue of controlling the false alarms and identifying when the proposed method is not applicable. The paper is closed in the final fifth section, where we also discuss possible future directions.

Preliminaries

A true-color $m \times n$ image, whose integrity is in question, will be represented with three $m \times n$ matrices $\mathbf{I}^{(1)}$, $\mathbf{I}^{(2)}$, and $\mathbf{I}^{(3)}$, $\mathbf{I}^{(k)} \in \{0,\dots,255\}^{m\times n}$, $k = 1,2,3$, corresponding to the red, green, and blue channel. Its noise residual is defined as

$$\mathbf{W}^{(k)} = \mathbf{W}(\mathbf{I}^{(k)}) = \mathbf{I}^{(k)} - F(\mathbf{I}^{(k)}), \tag{1}$$

where F is a denoising filter applied to each color channel. In this work, we use the Wavelet-based Daubechies 8-tap denoising method described in [19]. More on the choice of the filter appears below. Without changing notation, we assume that the mean $\overline{\mathbf{W}}^{(k)} = \frac{1}{mn}\sum_{i,j=1}^{m,n} w_{ij}^{(k)}$ is already subtracted from $\mathbf{W}^{(k)}$, $k = 1,2,3$.

Linear pattern

The *linear pattern* of image $\mathbf{I}$ is formed by three $m \times n$ matrices $\mathbf{L}^{(k)}$,

$$L_{ij}^{(k)} = r_i^{(k)} + c_j^{(k)},\ k = 1,2,3, \tag{2}$$

where $r_i^{(k)}$ and $c_j^{(k)}$ are the averages of the ith row and jth column of $\mathbf{W}^{(k)}$, respectively,

$$r_i^{(k)} = \frac{1}{n}\sum_{j=1}^{n} w_{ij}^{(k)}, \quad c_j^{(k)} = \frac{1}{m}\sum_{i=1}^{m} w_{ij}^{(k)}. \tag{3}$$

Notice that the row and column averages of $\mathbf{L}$ are the same as row and column averages of $\mathbf{W}$. This definition is consistent with the linear pattern first recognized in [6] as $\mathbf{L} = \mathbf{W} - Z(\mathbf{W})$, where $Z(\mathbf{W})$ is the residual $\mathbf{W}$ after "zero-meaning", i.e., making sure that its LP is all zeros.

In this definition, signal $\mathbf{L}$ depends on our choice of the filter F. The main purpose of the filter is to separate the content from noise including the Fixed Pattern Noise (FPN) and random noise while keeping the LP in the noise residual $\mathbf{W}$. For simplicity, the rest of this section refers to grayscale images, for which we drop the index k. The generalization to color images is straightforward by working with each color channel as a grayscale image. As the two vectors, $\mathbf{r} = (r_1, r_2, \ldots, r_m)$ and $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ fully define the linear pattern $\mathbf{L}$, the ordered pair $(\mathbf{r}, \mathbf{c})$ is a one-dimensional representation of the LP.

Energy and normalized energy

We define the *energy of the LP* (in one color channel or a grayscale image) as a pair of quantities,

$$E(\mathbf{L}) = \left(\sum_{i=1}^{m} r_i^2, \sum_{j=1}^{n} c_j^2\right). \tag{4}$$

Assuming $\mathbf{W}$ are i.i.d. realizations of a Gaussian random variable with zero mean and unit variance, the expected energy of its LP is $\mathrm{E}[E(\mathbf{L})] = (m/n, n/m)$. For convenience, we normalize residual $\mathbf{W}$ (as a vector) to unit variance, $\mathbf{W} \leftarrow \mathbf{W}/\sqrt{\mathrm{Var}\{\mathbf{W}\}}$, $\mathbf{L} = \mathbf{W} - Z(\mathbf{W})$, and compute the ratio $E(\mathbf{L})/\left(\frac{m}{n}, \frac{n}{m}\right)$, (the division is element-wise) that can be compared to (1,1) for comparison with the energy ratio of random noise. The corresponding *normalized LP energy* is

$$e(\mathbf{L}) = \left(\frac{n}{m}\sum_{i=1}^{m} r_i^2, \frac{m}{n}\sum_{j=1}^{n} c_j^2\right), \tag{5}$$

where r_i, and c_j are now computed from the residual normalized to unit variance.

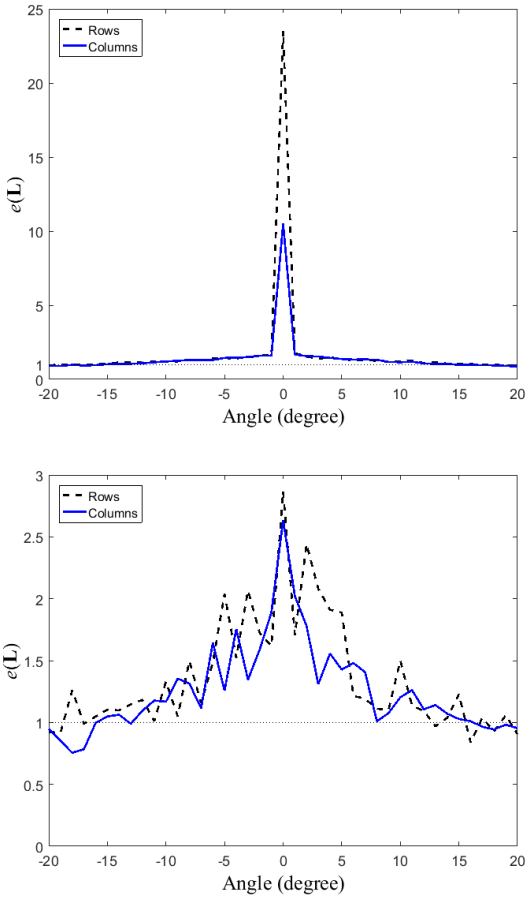


Figure 1. *LP energy e vs. rotation angle of an entire image (top), 200×200 image block (bottom).*

Properties, examples

A linear pattern is introduced into the image during its acquisition (signal processing), demosaicking, and then is "shaped" by subsequent processing and JPEG compression. It typically exhibits strong periodicities that depend on the imaging sensor and the processing pipeline (see examples in Figure 2). Unlike PRNU, linear patterns found in images from different cameras may or may not be correlated.

The LP is particularly useful for detecting forgeries if at least one component of $e(\mathbf{L})$ is larger than 1. Certain image manipulations, such as rotation, tend to decrease the LP energy towards or even below (1,1), see Figure 1, which is a sign of losing the original LP. This opens up the possibility to detect manipulated regions in images by their lack of the original LP.

Figure 2 shows different examples of LPs (close-ups of 100×150 pixel sections in grayscale) obtained for test images from the FAU dataset[1]. The toy examples in Figure 3 demonstrate how various manipulations applied to a small circular region (50 pixels in diameter) introduce local inconsistencies in an image's linear pattern.

[1] `https://www5.cs.fau.de/research/data/image-manipulation`

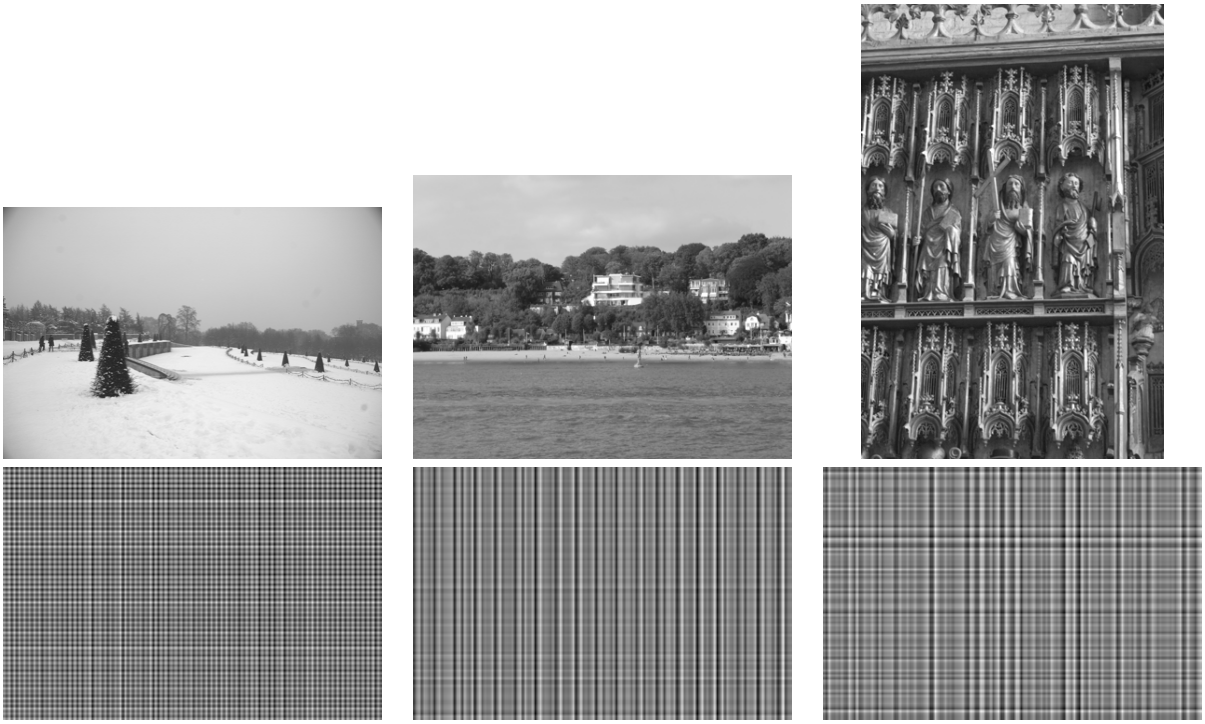


Figure 2. *Linear pattern (close-ups of size* 100×150*) and its periodic structure with periodicity a) two-pixel column-wise and row-wise (test image 'christmas_hedge.png',* 2014×3039*), b) strong 8-pixel periodicity in the column LP* $\mathbf{c}$ *('beach_wood.png',* 2448×3264*), c) LP with less noticable two-pixel periodicity in the row LP* $\mathbf{r}$ *('wood_carvings.png',* 3888×2592*).*

Towards localized manipulation detection

The idea of using the LP for detection of localized image manipulation is simple. If a region in an image has been processed strongly or replaced with content from another image with different pedigree, it is likely that the LP of this region will be incompatible with the LP from the rest of the host image. Let $(\mathbf{r}, \mathbf{c})$ now represent the LP $\mathbf{B} - Z(\mathbf{B})$ of a square block $\mathbf{B} \subset \mathbf{W}$ of size $w \times w$. Evaluating the similarity of this block $\mathbf{L}$ in a sliding window $\mathbf{B}$ with the LP estimated from the rest of the image should reveal the modified areas. However, establishing the presence of a modified LP within the image is challenging for the following reasons:

1. The original LP is not fully known once a part of the image is modified.
2. LP is a weak signal in comparison with the image content and random noise present in images. Its energy is comparable to the energy of the PRNU.
3. LP is not always homogeneous throughout the image also because of the denoising filter performing not as well in textured or noisy regions.
4. Subsequent JPEG compression and processing can suppress or modify the LP.

Attempts to evaluate the similarity of the LP in a block-wise fashion by correlating the LP in blocks $\mathbf{B}$ with the LP averaged over all such blocks resulted in a high false detection or very low overall positive detection. For this reason, we consider a transformed representation, the sample-based power spectral density (PSD) to capture the spectral properties of the LP. The transform is implemented as the Discrete Fourier Transform of circular cross-correlations $\mathbf{y}(\mathbf{r})$ and $\mathbf{y}(\mathbf{c})$ of $\mathbf{r}$ and $\mathbf{c}$,

$$y_\tau(\mathbf{x}) = \frac{1}{w} \sum_{i=1}^{w} x_i x_{i+\tau}, \quad \tau = 1, \ldots, w, \tag{6}$$

$$s_k(\mathbf{x}) = \mathcal{F}(\mathbf{y}(\mathbf{x})) = \sum_{i=1}^{w} y_i(\mathbf{x}) e^{(-2\pi j/w)(k-1)(i-1)}, \tag{7}$$

and will be denoted as $\mathbf{s}(\mathbf{r})$ and $\mathbf{s}(\mathbf{c})$, respectively, $k = 1, \ldots, w$. Note that in (6), $i + \tau \triangleq i + \tau - w$ when $i + \tau > w$.

The pair $\mathbf{s}^{(u)} = (|\mathbf{s}(\mathbf{r})|, |\mathbf{s}(\mathbf{c})|)$ will be called *block signature*, u being the block index and $|\mathbf{s}|$ denotes the absolute value applied to each element of $\mathbf{s}$. Vectors $\mathbf{s}(\mathbf{r})$ and $\mathbf{s}(\mathbf{c})$ are invariant to circular shifts of $\mathbf{r}$ and $\mathbf{c}$. Therefore, the PSD of a windowed periodic signal (with the window size w equal to a multiple of the period length) does not change after the window is shifted. This property is crucial for estimating block signatures that one might more or less expect in every window of a pristine image. We refer to it as the *expected signature*, denoted by $\mathbf{s} \in \mathbb{R}^{2w}$, computed as the average over a suitable set of K image blocks,

$$\mathbf{s} = \frac{1}{K} \sum_{u=1}^{K} \mathbf{s}^{(u)}. \tag{8}$$

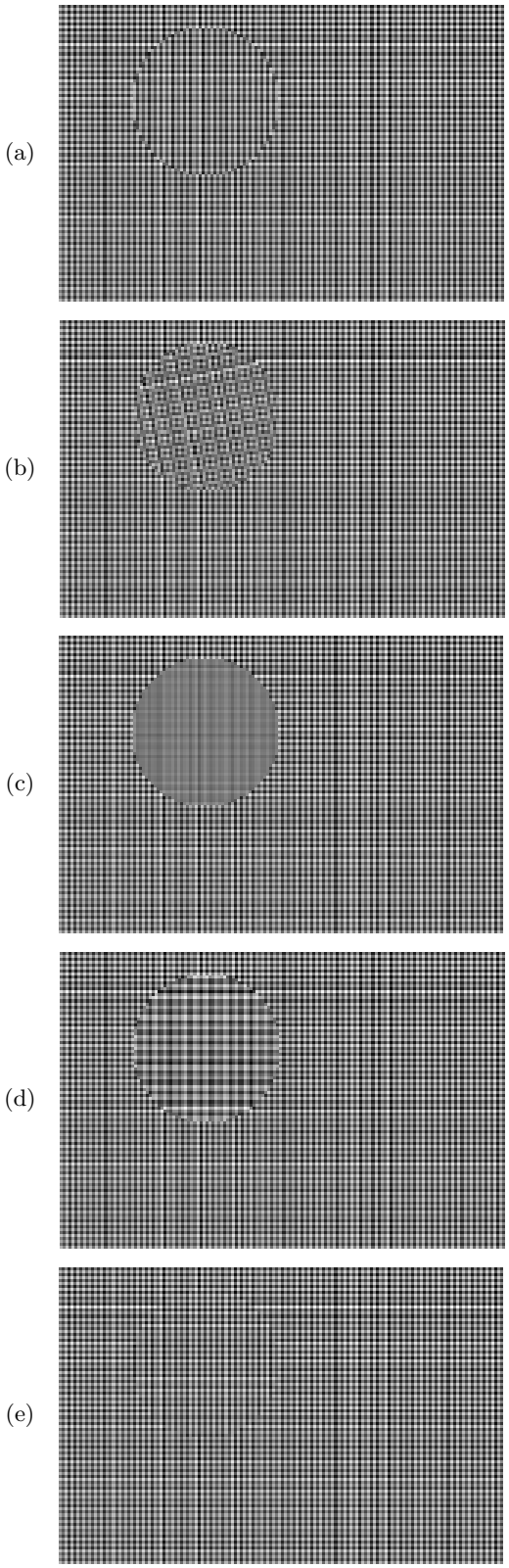


Figure 3. *The effect of a circular area modification on the LP from image (a) in Figure 2: a) copy-paste shifted by an odd number of pixels, b) rotation, c) downsampling, d) upsampling, e) copy-paste shifted by an even number of columns.*

An example of a block signature and the expected signature is shown in Figure 4. The test image is the image (c) from Figure 2, the window size $w = 200$ and the block index $u = 1000$. Notice the pronounced two-pixel period in $\mathbf{r}$ that manifests itself as the peak at $f = 200/2$ and a three-pixel period as peaks at $f = 200/3$ and $f = 400/3$.

The similarity between a block signature and the expected signature $\mathbf{s}$ is the basis for a two-dimensional output mask that reveals manipulated areas as dark (closer to zero rather than to 1), under the assumption that the manipulation is small compared to the image size. The output mask is of the same dimensions as the analyzed image. We take the standard correlation coefficient ρ as a similarity measure, defined for two vectors $\mathbf{a}, \mathbf{b}$ of equal length w as

$$\rho = \mathrm{corr}(\mathbf{a}, \mathbf{b}) = \frac{\left\langle \mathbf{a} - \overline{\mathbf{a}}, \mathbf{b} - \overline{\mathbf{b}} \right\rangle}{\|\mathbf{a} - \overline{\mathbf{a}}\| \left\| \mathbf{b} - \overline{\mathbf{b}} \right\|}, \tag{9}$$

where the bar denotes the sample mean, $\langle \cdot, \cdot \rangle$ denotes the dot product and $\|\cdot\|$ the Euclidian norm.

Before presenting the entire method in detailed steps, we wish to point out a problem that most forgery detection methods must face – the color saturation problem. We say that the pixel is saturated in an 8-bit grayscale image if its value is either 255 (white), 0 (black), or equal to the maximum or minimum value within the image, and one of its four neighboring pixels has the same value. The pixel in a color image is said to be saturated if it is saturated in at least two color channels. Saturation in an entire block makes the LP of that block equal to zero (or very close to zero). Such blocks need to be excluded from calculations. A larger portion of saturated pixels in a block means that the obtained signature will likely differ more strongly from the expected signature. Since a small correlation between signatures can be interpreted as a forged region, we prefer to adjust the correlation for partially saturated blocks towards 1 proportionally to the ratio ψ of saturated pixels in the block.

Description of the method

First, assume that $\mathbf{I}$ is a representation of a grayscale image. The proposed method is described in steps in Algorithm 1.

Notice that the output mask is conveniently bounded, $0 \leq H_{ij} \leq 1$. The last step, adjusting the output mask, is equivalent to adjusting the detection threshold T (introduced later in the experimental section) by multiplying it with $\max_{i,j} H_{ij}$. The threshold t for issuing the warning may be set to $t = 3$ based on observing Figure 11, where we found that the detection ability is mostly lost when $\max(e_1(\mathbf{L}), e_2(\mathbf{L})) < 3$. We settled on the following choice of the parameters in experiments: $w = 200$, the blocks overlap by $200 - 32 = 168$ pixels, which determines the number of blocks $M \times N$.

We also tested a "color version" of the algorithm implementation that requires executing Steps 1 to 4 separately for each color channel and concatenating the resulting 6 PSDs in a modified Step 6 to obtain block signatures $\mathbf{s}^{(u)} = (|\mathbf{s}^{(u)}(\mathbf{r}_R)|, |\mathbf{s}^{(u)}(\mathbf{r}_G)|, |\mathbf{s}^{(u)}(\mathbf{r}_B)|,$

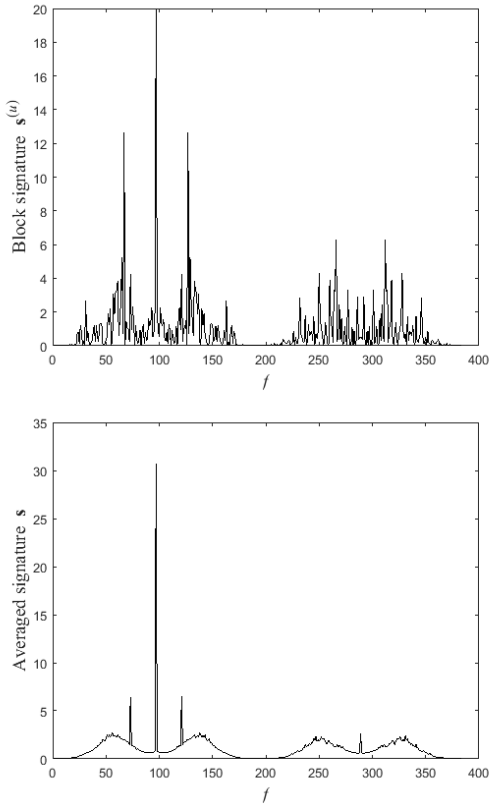


Figure 4. *One block (window) PSD $\mathbf{s}^{(u)}$ (top) and the PSD $\mathbf{s}$ averaged over all blocks 16 pixels apart (bottom). The left ($0 \le f \le 200$) and right ($200 \le f \le 400$) half of each plot are the sample PSDs of $\mathbf{r}$ and $\mathbf{c}$, respectively.*

$|\mathbf{s}^{(u)}(\mathbf{c}_R)|, |\mathbf{s}^{(u)}(\mathbf{c}_G)|, |\mathbf{s}^{(u)}(\mathbf{c}_B)|)$, $u = 1, \ldots, MN$, where $(\mathbf{r}_R, \mathbf{c}_R)$, $(\mathbf{r}_G, \mathbf{c}_G)$, $(\mathbf{r}_B, \mathbf{c}_B)$ are linear patterns in R, G, B channels, respectively. The resulting block signature is now of length $6w$. However, the detection performance of this (slower) version of the algorithm is not better than converting the inspected image to grayscale and running Algorithm 1. Moreover, in this "color version" the resulting output mask cannot reveal in which color channel the forgery occurred if not in all three.

Illustrative example

Unlike forensic methods that use the sensor noise fingerprint, the proposed method does not rely on external knowledge of the LP associated with the image source (typically a digital camera). The assumption instead is the existence of a similarity between spectral properties of block LPs, or in other words, the presence of some form of periodicity in the LP. This periodicity allows us to predict what properties the LP should have in individual unmodified image blocks.

To demonstrate how certain forgeries disrupt the LP, we used the 6 Mpixel test image 'christmas_hedge.png' from the FAU dataset and created our own naïve forgery.

The modified area is circular and the donor content is taken from a nearby region within the same image. In this example, we opted not to add or remove an easily spotted object from the image in order to make it more convincing

Algorithm 1 Forgery detection for grayscale images using the LP.

1. Compute the noise residual as $\mathbf{W} = \mathbf{W}(\mathbf{I}) = \mathbf{I} - F(\mathbf{I})$, where F is the Daubechies 8-tap wavelet denoising filter [19].
2. If $\max(e(\mathbf{L})) < t$, issue a warning "*The test is likely to fail*".
3. Divide $\mathbf{W}$ into $M \times N$ overlapping blocks (windows) $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_{MN}$ of size $w \times w$.
4. Compute the one dimensional LP $\mathbf{l} = (\mathbf{r}, \mathbf{c})$ of length $2w$ for each block u, $u = 1, \ldots, MN$.
5. Compute the power spectral densities $\mathbf{s}^{(u)}(\mathbf{r})$ and $\mathbf{s}^{(u)}(\mathbf{c})$ as the Fourier transform of the sample autocorrelation (PSD Eq. 7) of $\mathbf{r}, \mathbf{c}$, respectively.
6. Form the "block signature" $\mathbf{s}^{(u)} = (|\mathbf{s}^{(u)}(\mathbf{r})|, |\mathbf{s}^{(u)}(\mathbf{c})|)$ of length $2w$ for uth block.
7. Determine the "expected signature" of the block LP as the average $\mathbf{s} = \frac{1}{MN} \sum_{u=1}^{MN} \mathbf{s}^{(u)}$.
8. Compute the similarity between each block signature and the expected signature as the correlation coefficient $\rho_u = \mathrm{corr}(\mathbf{s}^{(u)}, \mathbf{s})$, $u = 1, \ldots, MN$.
9. Adjust for the block saturation ψ_u, $\rho_u = \rho_u + (1 - \rho_u) \cdot \psi_u$.
10. Compute the output mask $\mathbf{H} \in \mathbb{R}^{m \times n}$, $H_{ij} = \mathrm{mean}\{\rho_u; \mathbf{B}_u \text{ contains pixel } (i,j)\}$.
11. Rescale $\mathbf{H}$ from interval $[0, \max_{i,j} H_{ij}]$ to $[0, 1]$.

that the algorithm is not detecting the modification due to other effects, such as harsh discontinuities in luminance, edges, or colors. If the source of the replacement part is taken from a different image the detection should typically have a better chance of success because the LP tends to differ more. Depending on the periodicity of the LP, a simple copy-paste forgery may be detected. However, when the LP of the modified region matches the original LP in terms of phase, the algorithm detects only the boundary of the pasted area (Figure 5(a)). Manipulations, such as rotation or scaling of the pasted area, result in a positive detection that shows up as a dark solid region in the output mask (cases (b–d) in Figure 5).

Experiments

In this section, we present examples of positive detection on test images from the FAU dataset and assess the robustness of the method to JPEG compression. The results quantified by an output mask metric are presented on a larger database of forgeries.

Positive localization

Fourty eight realistic forgery examples are included in the FAU dataset. In each, a few versions of the same forgery with different processing applied to the pasted region is provided, including noise addition, rotation by a set of small angles, upsampling and downsampling by a few percent. The parameters of the added noise were not specified. The ground truth for forgeries in FAU dataset was

Figure 5. *Detection output of Algorithm 1 applied to four types of area processing. The forged image (left), the output mask (right) for a) copy-paste forgery shifted by an odd number of pixels), b) rotation by 12 degrees, c) downsampling, d) upsampling.*

prepared for testing common copy-paste detection methods that search for copied objects within one image [9]. Such methods cannot tell which of the two or more similar objects are at their original location and which are pasted and replaced original content. Therefore, both occurences are marked as forged in the ground-truth binary image provided at `https://www5.cs.fau.de/research/data/image-manipulation/`. On the other hand, the proposed LP-based method only identifies the forged area. For this reason, we opted not to compute a formal detection score for this dataset and only show selected insightful examples (Figures 6,7, and 8).

JPEG compressed forgery

The previous examples did not involve JPEG compression that may likely be applied when saving the forged image. Unfortunately, such compression can suppress the LP needed for the proposed method to work well. The performance naturally degrades with decreasing JPEG quality factor. Typically JPEG compression with quality factor 95 or lower prevents the proposed method from working. Even compression with quality factor 100 may cause missed detection, depending on the detection threshold T for the output mask. A typical example of the effect of JPEG compression is presented in Figure 8.

NIMBLE Challenge

NIMBLE Challenge is a platform for testing and evaluating systems for image forgery detection organized by NIST.[2] For self-evaluation, both forged images as well as the ground truth binary masks are available. Among the test datasets, the set denoted as NC2016 contains high quality (HQ) and low quality (LQ) images. The LQ images were first upsampled and then downsampled again to their original size, before being compressed with the standard JPEG 75% quantization table, which is too harsh for the proposed detector. The HQ set of 282 images was compressed with a non-standard quantization table, as given below,

$$\begin{pmatrix} 12 & 8 & 8 & 12 & 17 & 21 & 24 & 17 \\ 8 & 9 & 9 & 11 & 15 & 19 & 12 & 12 \\ 8 & 9 & 10 & 12 & 19 & 12 & 12 & 12 \\ 12 & 11 & 12 & 21 & 12 & 12 & 12 & 12 \\ 17 & 15 & 19 & 12 & 12 & 12 & 12 & 12 \\ 21 & 19 & 12 & 12 & 12 & 12 & 12 & 12 \\ 24 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\ 17 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \end{pmatrix}.$$

Evaluating the manipulation localization performance of the proposed method requires setting a detection threshold that converts the grayscale output mask to a binary mask. We fixed this threshold for all tests at $T = 0.5$. The accuracy of localization is evaluated with the Matthews Correlation Coefficient (MCC):

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \qquad (10)$$

where TP is the true positive area, TN is the true negative area, FP is the false positive area, and FN is the false negative area, all computed by comparing the output binary mask to the ground truth mask. If $\mathrm{MCC} = 1$, there is perfect correlation between the ground truth and the algorithm output mask. If $\mathrm{MCC} = 0$, there is no correlation. If $\mathrm{MCC} = -1$, there is perfect anti-correlation. Negative score means that we switched the original content for manipulated and vice versa. In fact, if two images are spliced together,

[2] `https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation`

(a) Pasted region contains mild noise

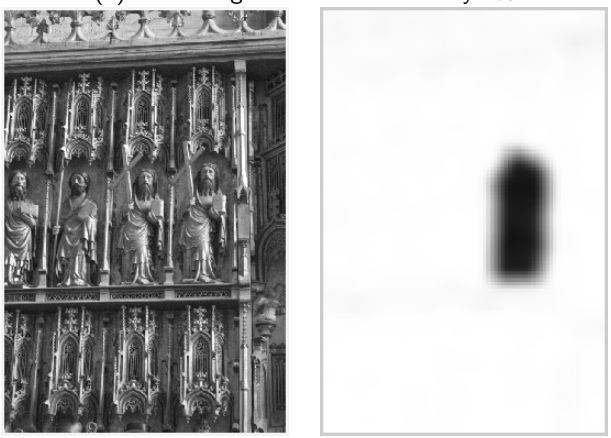
(b) Pasted region contains strong noise

(c) Pasted region was rotated by 2 degrees

Figure 6. *Example of realistic forgery detections when the pasted region (a) contains a weak added noise, (b) strong noise, (c) was rotated by 2 degrees.*

(a) Pasted region was rotated by 10 degrees

(b) Pasted region was downscaled by 3%



(c) Pasted region was was upsampled by 3%

Figure 7. *Example of realistic forgery detections when the pasted region was (a) rotated by 10 degrees, (b) downscaled by 3%, (c) upsampled by 3%. The resizing was done in Matlab using bi-cubic interpolation.*

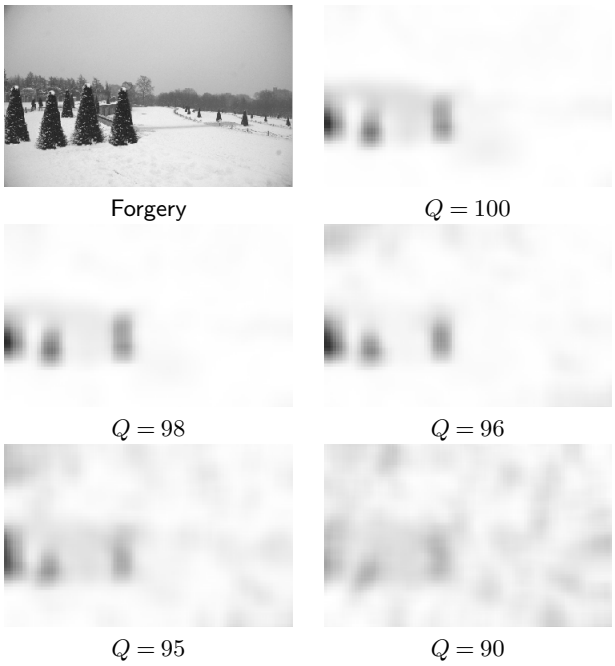


Figure 8. *Output masks after compressing the forged image with quality factor Q. The added trees (upper left) were downscaled by 5%.*

there is an inevitable ambiguity in what is original and what is not. Therefore, we decided to compute the *absolute score* |MCC| in order to make sense of the score averaged over the whole set of test images.

The proposed Algorithm 1 performs reasonably well with varying accuracy for about a quarter of the NC2016-HQ test set. Figure 9 depicts the distribution of the absolute scores. The mean absolute MCC score was 0.1894 for the grayscale and 0.1843 for the color version of the algorithm, respectively. The mean MCC of the best 70 images out of the 282 was 0.5640.

We wish to point out that some of the scores would be higher if the ground truth binary map at the input of our detection evaluation was always correct. In some cases, this binary map is (perhaps mistakenly) shifted by a not negligible number of pixels. For an ease of future comparison, all MCC scores reported in this work are based on the ground truth provided by NIST despite its occasional incorrectness.

One example of a successful forgery localization is shown in Figure 10. Note that parts of the forged area that are saturated blacks would go undetected. The output mask would show them as pure white, the same way as saturation at the white end of the gray scale would be shown. Luckily, near-black colors in this forged image contain small amount of noise and thus do not qualify as saturated.

False alarm control

Achieving a low rate of falsely identified areas as tampered is crucial for an automatized detector of image forgery. The first measure to keep FPs low is to check if the image under inspection satisfies the assumptions required by the

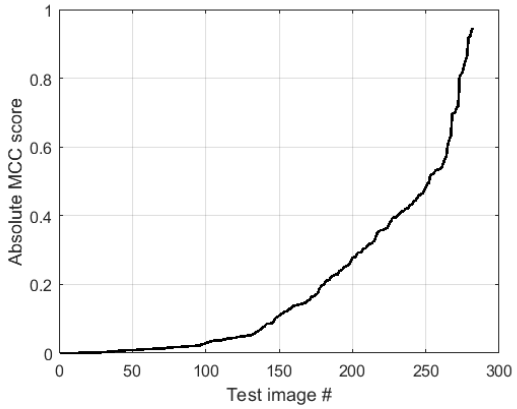


Figure 9. *Distribution of absolute MCC score for NC2016-HQ database (sorted by the score).*

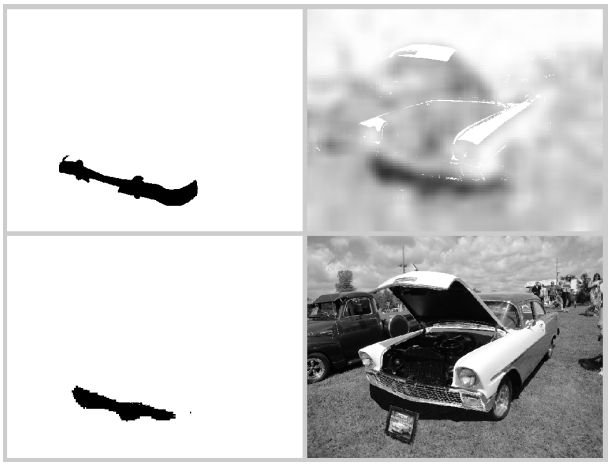


Figure 10. *Example of a successful localization. Ground truth (upper left), output mask (upper right), output mask (lower left), forgery (lower right). Saturated pixels display themselves as white in the output mask.*

detection method. For an LP-based method, the linear pattern present in the original parts of the image must not be "overwhelmed" by excessive noise or JPEG quantization. JPEG images compressed with quality factor lower than 95 in most cases resulted in a missed detection. To keep error rates low, we suggest to reject JPEG images compressed below 95% quality before testing. More research is needed to quantify the role of the quantization table and particular DCT frequencies in order to gather statistical data about the error rates for images compressed with non-standard quantization tables. We hypothesize that high-frequency DCT coefficients are important for preserving the LP during compression.

While JPEG compression may lead to missed detection, it did not increase the FP rate in the following test with all 48 images from the FAU dataset. Each copy-move forgery, without any rotation, scaling, or other processing of the pasted region, was compressed with the standard quality factor $q = 100, 90, 80, \ldots, 20$ resulting in $9 \times 48 = 432$ test images. The false positive ratio in the output masks

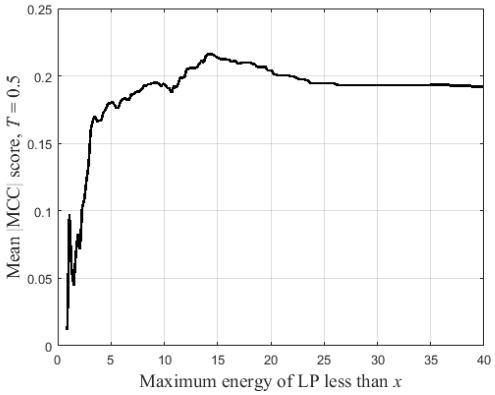


Figure 11. *Mean absolute MCC score vs. maximum energy,* $\max e(\mathbf{L}) < x$.

Figure 12. *Falsely identified areas due to strong noise caused by high ISO during exposure.*

averaged over this set of 438 tests at fixed theshold $T = 0.5$ was $FP = 0.0114$. On the other hand, on images from the NC2016-HQ test set used above, we found $FP = 0.0405$.

It is also important to consider the image dimensions when interpreting the output of the proposed method. Larger images enable a better estimation of the expected signature $\mathbf{s}$. Both the image height and width should be at least several times larger than the window size w. Our experiments with $w < 200$ gave an increased FP rate.

Computing the normalized energy $e(\mathbf{L})$ of the image's LP can be an early indication of whether the detection method is applicable. Experimental evidence in Figure 11 suggests that a low total energy of the LP impedes the proposed method. The plot shows a sharp performance drop (in terms of the absolute MCC score) for images with $\max(e_1(\mathbf{L}), e_2(\mathbf{L})) < 3$. Note that the assumption of a sufficiently strong LP is emphasized in Step 2 of Algorithm 1.

Probably the most challenging limitation comes from images with certain textures or a high level of noise, such as when a grainy texture is mistaken by the denoising filter F for noise or when the image is taken at a high ISO. One example is shown in Figure 12. How to eliminate this type of false alarm remains an open problem. If the information about high ISO is available (for example from the EXIF header), then the test can be rejected in advance.

The very last type of failure is due to lens distortion (LD) correction. This geometric distortion prevents Algorithm 1 from working correctly because the LP remains only in the optical center of the image, the area lest distorted by the lens. Thus, the distortion would have to be removed prior to applying the algorithm. We consider this problem as a future direction that can be addressed by identifying images corrected for LD to prevent increasing the FP rate. If the optical center was in the geometrical center of the image, then the LD correction can possibly be inverted [14].

Conclusions

The linear pattern as an intrinsic signal present in most digital images has been overlooked for applications in digital forensics. In this paper, we proposed using the LP for detection and, mainly, for localization of image splicing type of forgeries. The method can detect and localize image splicing and certain copy-move forgeries by checking the integrity of the power spectral density of the LP computed on sliding blocks. Unlike most other methods that are based on detecting signs of processing associated with the forgery operation, this LP-based method works with a signal that had been present in the original image and uses it as a type of "natural watermark". The method has its shortcomings as it is limited to uncompressed and high quality JPEG images at high resolution. Occasionally, certain textures may make the locally extracted LP differ from the rest of the image, which can introduce false alarms. As a future direction, characterization of such textures and a proper adjustment of he proposed algorithm may resolve this issue.

Further research is also needed to investigate the exact genesis of the LP, how it is affected by various processing, and its relationship to other entities proposed for forensic applications. Since color filter array interpolation (and possibly other type of resampling) contribute to the energy of the LP, there likely exists a relationship between the proposed method and methods that use interpolation artifacts. Another related forensic entity are "JPEG dimples" [1], which are dots separated by 8 pixels due to one-sided quantization in the discrete cosine transform (DCT) used for JPEG compression. The dimples contribute to a strong(er) period 8 in LPs from cameras that exhibit them.[3]

Acknowledgments

This material is based on research sponsored by DARPA and Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA and Air Force Research Laboratory (AFRL) or the U.S. Government.

Author Biography

Miroslav Goljan received the Ph.D. degree in Electrical Engineering from Binghamton University in 2002 and the M.S. in Mathematical Informatics from Charles University in Prague, Czech Republic, in 1984. He is Research Scientist at the Dept. of Electrical and Computer Engineering at Binghamton University. His research

[3] According to [1], approximately half of all cameras available on the market today exhibit the dimples.

focuses on digital image and digital camera forensics, steganography, steganalysis, and reversible data hiding in digital media.

Jessica Fridrich is Distinguished Professor of Electrical and Computer Engineering at Binghamton University. She received her PhD in Systems Science from Binghamton University in 1995 and MS in Applied Mathematics from Czech Technical University in Prague in 1987. Her main interests are in steganography, steganalysis, and digital image forensics. Since 1995, she has received 20 research grants totaling over $11 mil that lead to more than 180 papers and 7 US patents.

Matthias Kirchner is an Assistant Professor of Electrical and Computer Engineering at Binghamton University. He received his PhD in Computer Science from Technical University of Dresden in Germany. His research focuses on multimedia security, and particularly on a variety of practical and theoretical problems in the area of (multi)media forensics and counter-forensics.

References

[1] S. Agarwal and H. Farid. Photo forensics from JPEG dimples. In *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2017.

[2] T. Bianchi and A. Piva. Image forgery localization via block-grained analysis of JPEG artifacts. *IEEE Transactions on Information Forensics and Security*, 7(3):1003–1017, 2012.

[3] R. Böhme and M. Kirchner. Media forensics. In Stefan Katzenbeisser and Fabien Petitcolas, editors, *Information Hiding*, chapter 9, pages 231–259. Artech House, Norwood, MA, 2016.

[4] M. U. Celik, G. Sharma, and E. Saber. A hierarchal image authentication watermark with improved localization and security. In *Proceedings IEEE, International Conference on Image Processing, ICIP 2001*, number ID 3532, Thessaloniki, Greece, October 7–10, 2001. CD ROM version.

[5] S. Chakraborty and M. Kirchner. PRNU-based image manipulation localization with discriminative random fields. In *Media Watermarking, Security, and Forensics 2017*, pages 113–120, 2017.

[6] M. Chen, J. Fridrich, and M. Goljan. Digital imaging sensor identification (further study). In E.J. Delp and P.W. Wong, editors, *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, pages 0P 1–12, San Jose, CA, January 2007.

[7] M. Chen, J. Fridrich, and M. Goljan. Imaging sensor noise as digital x-ray for revealing forgeries. In T. Furon et al., editor, *Proc. 9th Information Hiding Workshop, Saint Malo, France*, volume 4567 of *LNCS*, pages 342–358. Springer-Verlag, June 2007.

[8] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva. A Bayesian-MRF approach for PRNU-based image forgery detection. *IEEE Transactions on Information Forensics and Security*, 9(4):554–567, 2014.

[9] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou. An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on Information Forensics and Security*, 7(6):1841–1854, 2012.

[10] D. Cozzolino, G. Poggi, and L. Verdoliva. Splicebuster: a new blind image splicing detector. In *IEEE International Workshop on Information Forensics and Security*, Rome, Italy, November 16–19 2015.

[11] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva. Image forgery localization via fine-grained analysis of CFA artifacts. *IEEE Transactions on Information Forensics and Security*, 7(5):1566–1577, 2012.

[12] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni. A framework for decision fusion in image forensics based on Dempster–Shafer theory of evidence. *IEEE Transactions on Information Forensics and Security*, 8(4):593–607, April 2013.

[13] J. Fridrich, M. Goljan, and R. Du. Invertible authentication watermark for JPEG images. In *International Symposium on Information Technology, ITCC 2001*, pages 223–227, Las Vegas, NV, 2001.

[14] M. Goljan and J. Fridrich. Estimation of lens distortion correction from single images. In A. Alattar, N. D. Memon, and C. Heitzenrater, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2014*, volume 9028, pages 0N1–13, San Francisco, CA, February 3–5, 2014.

[15] D. Kundur and D. Hatzinakos. Towards a telltale watermarking technique for tamper proofing. In *Proceedings IEEE, International Conference on Image Processing, ICIP 1998*, volume 2, Chicago, IL, October 4–7, 1998.

[16] J. Lukáš, J. Fridrich, and M. Goljan. Detecting digital image forgeries using sensor pattern noise. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII*, volume 6072, page 20Y, San Jose, CA, January 16–19, 2006.

[17] L. M. Marvel, G. W. Hartwig, and C. Boncelet Jr. Compression-compatible fragile and semi-fragile tamper detection. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents II*, volume 3971, pages 131–139, San Jose, CA, January 24–25, 2000.

[18] O. Mayer and M. Stamm. Improved forgery detection with lateral chromatic aberration. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2024–2028, March 2016.

[19] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, 6(12):300–303, 1999.

[20] M. Wu and B. Liu. Watermarking for image authentication. In *Proceedings IEEE, International Conference on Image Processing, ICIP 1998*, Chicago, IL, October 4–7, 1998.

[21] I. Yerushalmy and H. Hel-Or. Digital image forgery detection based on lens and sensor aberration. *International Journal of Computer Vision*, 92(1):71–91, 2011.