SOURCE ANONYMIZATION OF DIGITAL IMAGES: A COUNTER-FORENSIC ATTACK ON PRNU BASED SOURCE IDENTIFICATION TECHNIQUES

Prithviraj Sengupta, Venkata Udaya Sameer, Ruchira Naskar
Department of Computer Science and Engineering
National Institute of Technology Rourkela
Spyridon Dosis
dosis@dsv.su.se
Odisha-769008, India
prithvi1096@gmail.com, {214cs2156,naskarr}@nitrkl.ac.in

Ezhil Kalaimannan
Department of Computer Science
University of West Florida
Pensacola, FL 32514, USA
ekalaimannan@uwf.edu

ABSTRACT

A lot of photographers and human rights advocates need to hide their identity while sharing their images on the internet. Hence, source–anonymization of digital images has become a critical issue in the present digital age. The current literature contains a few digital forensic techniques for “source-identification” of digital images, one of the most efficient of them being Photo–Response Non–Uniformity (PRNU) sensor noise pattern based source detection. PRNU noise pattern being unique to every digital camera, such techniques prove to be highly robust way of source–identification. In this paper, we propose a counter–forensic technique to mislead this PRNU sensor noise pattern based source–identification, by using a median filter to suppress PRNU noise in an image, iteratively. Our experimental results prove that the proposed method achieves considerably higher degree of source anonymity, measured as an inverse of Peak-to-Correlation Energy (PCE) ratio, as compared to the state-of-the-art.

Keywords: Counter Forensics, Digital Forensics, Median Filter, Photo Response Non-Uniformity, Source-Anonymization

1. INTRODUCTION

In today’s cyber world, digital images and videos are used as means of communication in most tenets of life, ranging from media houses, businesses, to even the court of law, where they act as the primary sources of evidence towards any event. They represent the primary source of evidence to be able to present, process and store information. However, with the present rapid advancement of technology, it has become a trivial affair to manipulate and edit authentic digital images, with the use of
low-cost user-friendly, yet versatile image and video processing software and tools, with minimal effort and expertise. In this regard, digital forensics refers to the collection of scientific methods, specifically involving investigation of evidences extracted a-posteriori from digital devices, to systematically infer particulars of an unknown image/video, its origin and generation process.

Source Camera Identification (SCI) is the process of mapping an image back to its source device which is completely based on post-processing of data and without involving any form of data pre-processing such as watermarking or fingerprinting techniques. With a wide availability of various forms of digital cameras, ranging from Digital Single Lens Reflex (DSLR) to cheap mobile phone cameras, the Source Camera Identification problem poses to be a major challenge. Reliable methods to correctly identify the source camera help greatly in cases such as espionage and movie piracy. One such method to identify the source camera was proposed by Lukas, Fridrich and Goljan [3], which uses the presence of a form of noise called Photo-Response Non-Uniformity (PRNU), caused by varying sensitivity of pixel sensors to light. Imperfections during the sensor manufacturing and non-homogeneity of silicon wafers are the primary causes behind formation of PRNU noise. Varying PRNU noise patterns are related to varying number of pixels, depending on the imaging sensors; hence it would be highly unlikely that patterns from different cameras have the same PRNU noise. Given the fact that every digital camera available in the market has its unique imaging sensors, the PRNU sensor pattern noise is unique to every camera, and this feature can be used to differentiate between different makes and models of digital cameras. This makes source camera identification through PRNU noise a highly reliable method.

However, in many cases, unique image source identification is absolutely undesirable. For example, many times photographers, activists and human–right defenders desire to stay anonymous while spreading their images and videos. This calls for the need of image source anonymization techniques. Such source anonymization techniques have a counter-forensic [1] perspective in the sense that they are needed to evaluate and establish the reliability of existing source identification forensic methods.

Our main contributions in this paper are discussed as follows. We present a counter-forensic technique for digital image source anonymization. The proposed technique operates by suppressing the PRNU noise produced by a camera, effectively. Here, we use median filtering to achieve the above goal. Finally, we compare the performance of the proposed method with a very recent state-of-the-art source-anonymization technique. The results prove that the proposed method succeeds to achieve a considerably higher degree of source-anonymization.

Rest of the paper is organized as follows. In Section 2, we present an overview of the state-of-the-art source identification techniques. We also discuss the existing counter-forensic techniques for source anonymization and discuss their merits and demerits, in Section 2. In Section 3, we discuss in detail the Source Camera Identification method utilizing PRNU sensor pattern noise, and present relevant similarity metrics used in this paper. In Section 4, we lay down the details of the proposed counter-forensic method for image source anonymization. In Section 5, we present our experimental results along with comparison with a very recent scheme. Finally, we conclude the paper with directions for future research in Section 6.
2. RELATED WORK

After a scene has been captured, a number of post-processing operations are performed inside the camera to produce the final digital image. These post processing operations leave traces/fingerprints which can be analyzed by forensic investigators to identify the camera from which the image in question has originated. Following this principle, a major breakthrough in Source Camera Identification happened with the discovery of Sensor Pattern Noise [3] as a fingerprint to identify image source. Sensor pattern noise is generated mainly due to impurities in the camera’s sensor which converts the incident light to digital form. The more recent works in this direction are aiming to strengthen the technique by enhancing the sensor pattern noise through attenuation of scene details [15], and by preprocessing the sensor pattern noise by spectrum equalization [14]. Currently, a number of researchers are aiming to make the fingerprint matching more efficient by using compressed fingerprints [16] and composite fingerprints with group testing strategies [17].

On the contrary, counter-forensics or anti-forensics is a branch of science and technology that deals with misleading or bypassing the existent forensic analyses to detect the presence of forgeries in a given image. Counter-forensics is of particular importance because it challenges the existing methods of forgery detection and assesses their limitations. This further helps in improving and strengthening the existing forensic techniques against intelligent counterfeiter.

In this work, we deal with the anonymization of source camera by which a given digital image was captured. A highly effective and robust method for source camera identification is through utilization of PRNU noise pattern which is unique to every digital camera (make and model). One of the pioneer works in this direction was proposed by Lukas, Fridrich and Goljan [3], as discussed previously. PRNU based source camera identification is carried out by first estimating the PRNU noise or fingerprint of a given digital camera, and then comparing it to the test image through Normalized Cross Correlation, or by calculating the Peak-to-Correlation Energy (PCE) ratio. This method has been explained in detail in the next section.

Source anonymization of digital images has been previously achieved by a different technique, such as flat-fielding [4], Seam-Carving [5], adaptive fingerprint removal [6] and adaptive PRNU denoising, called as APD-1 [7] and APD-2 [8]. These methods have been successful in anonymizing digital images up to a considerable extent; however, most of these techniques have their own limitations. Flat-field images are specifically difficult to capture because it needs dark field and flat frames. Also, these images are ineffective in digital image source anonymization when subjected to JPEG compression [7]. Seam-carving is also another method which results in source anonymization by deleting the low-energy pixels of an image in a particular path or seam. Since this method destroys the lesser significant low-energy pixels from the original images, the PRNU pattern of the given image changes which results in source anonymization. Although this is an effective method, this results into image resizing which is not desirable in many cases [9]. Also, seam-carving has certain limitations such as it cannot have uncarved blocks larger than the size of 50×50 pixels for successful anonymization.
3. PRNU BASED SOURCE CAMERA IDENTIFICATION

As the make and model of digital cameras vary, so do their sensors and the sensor patterns. Every camera has its own unique PRNU pattern (fingerprint) as different sensors produce different reactions to the same level of light intensity. The imaging output can be written as:

\[ P_x = P_0 + (P_0F + \varphi_1) \]  

(1)

where \( P_x \) is the image output which consists of both the PRNU noise (camera fingerprint) and other noises such as shot noise and dark current. In the above equation, \( P_0 \) is the amount of incident light, \( F \) is the camera fingerprint or the PRNU noise and \( \varphi_1 \) is the shot noise or Poisson noise. If we are given a set of images which are said to be generated from the same camera, we can calculate the PRNU pattern or camera fingerprint \( F \) from it. The Noise Residual (NR) of a single \((i)\) image can be calculated as:

\[ NR_x^{(i)} = P_x^{(i)} - DF(p_x^{(i)}) \]

(2)

where, the original image is passed through a Denoising Filter (DF) to produce a denoised image. The denoised image is then subtracted from the original image to generate the Noise Residual \( NR_x^{(i)} \). The PRNU noise pattern can then be calculated as:

\[ \frac{\sum_{x=1}^{n} NR_x^{(i)} p_x^{(i)}}{\sum_{x=1}^{n} (p_x^{(i)})^2} \]

(3)

where, \( n \) is the number of images used to calculate the fingerprint \( F \). The accuracy of the estimated value of the fingerprint \( F \) is directly proportional to the number of images used to calculate \( F \) i.e., higher the number of training images \( n \), better the estimated value of the PRNU pattern noise.

Now, to measure the similarity between the Noise Residual and the PRNU noise pattern of a camera, we calculate the Normalized Cross-Correlation (\( \rho \)) between the Noise Residual (NRx) and the Camera Fingerprint (F) as:

\[ \rho = \frac{\sum_{k=1}^{K}(NR_x[k] - \bar{NR}_x)(F[k+c] - \bar{F})}{|\|NR_x - \bar{NR}_x\|||\|F - \bar{F}\||} \]

(4)

where, \( c \) is the number of circular shifts, \(|x|\) is the L2 norm, and \( K \) is the total number of pixels of the image output.

The cross-correlation value gives an idea about the similarity between the Noise Residual and the camera fingerprint of an image. If the image is not captured by the camera whose fingerprint we have, the cross-correlation value would be very close to zero; whereas if the image is taken by the same camera we have, the cross-correlation value would be significantly higher. Although, using normalized cross-correlation is an efficient method to differentiate whether an image has been taken from a given camera or not, we do not have a common threshold for each camera to decide whether an image has been sufficiently anonymized, specific to a particular application. To counter this problem, we measure the level of source anonymity through another parameter called the Peak-to-Correlation Energy (PCE) ratio. This can be calculated as:

\[ PCE = \frac{\rho^2_{peak}}{|\|\rho\||\sum_{r \in \mathcal{E}} r^2|} \]

(5)

where, \( \rho \) is the normalized cross correlation between the Noise Residual and PRNU noise. \( \rho_{peak} \) is the smallest \( \rho \) that is greater than or equal to each of the cross-correlation values. The letter ‘r’ represents the set of all entries of the cross correlation and \( \mathcal{E} \) represents a small area near the peak height which is removed in order to calculate the PCE ratio. Symbol ‘\( \rho \)’ represents the values of the cross-correlations corresponding to the entries in \( r \), but not belonging to \( \mathcal{E} \). Previous works on PCE ratio calculation has shown that the PCE threshold
can be set at 50 [13]. This means if the value of the PCE ratio for the original given image and the camera fingerprint is greater than 50, then source identification is possible whereas any value less than 50 makes source identification impossible for the given images.

4. PROPOSED SOURCE ANONYMIZATION THROUGH IMAGE MEDIAN FILTERING

In this section, we present the details for the operation of the proposed method to impede source identification of digital images by removing the PRNU noise based camera specific (unique) traces or fingerprints from the images. Our main objective here is to lower the PCE value of the unanonymized image below the detection threshold, in order to achieve source anonymization. We do so by denoising the PRNU pattern of the original image using a median filter [2]. The proposed method has been explained in detail next.

The Peak-to-Correlation Energy (PCE) value of any unanonymized image with respect to its source camera has been observed to be always much higher than the detection threshold of 50, i.e., PCE(P,F) >> 50. Generally, the original image contains a number of noises of different kinds, such as salt-and-pepper noise, Poisson noise, PRNU sensor pattern noise etc. Now, to remove the traces of these different noise forms, we denoise the given images using a median filter [2], which effectively suppresses the noise effects to a considerable extent and has a smoothening effect on the images. To compute the median of n integer pixel values \( i_1, i_2, \ldots, i_n \), the integers are first sorted into the sequence \( i_{(1)}, i_{(2)}, \ldots, i_{(n)} \), such that \( i_{(1)} \leq i_{(2)} \leq \cdots \leq i_{(n)} \). Then, their median is computed as:

\[
\text{median} = \begin{cases} 
\frac{i_{\left\lceil \frac{n+1}{2} \right\rceil}}{2} & \text{if } n \text{ is odd} \\
\frac{i_{\left\lceil \frac{n}{2} \right\rceil} + i_{\left\lceil \frac{n}{2} \right\rceil + 1}}{2} & \text{otherwise}
\end{cases}
\]

Median filtering applied to a set of n pixels \( p_1, p_2, \ldots, p_n \) of an image, using a filter window size of 3, produces n median-filtered pixels \( p_1', p_2', \ldots, p_n' \), which are computed as:

\[
p_1' = \text{median}(p_1, p_1, p_2) \\
p_2' = \text{median}(p_1, p_2, p_3) \\
\vdots \\
p_{n-1}' = \text{median}(p_{n-2}, p_{n-1}, p_n) \\
p_n' = \text{median}(p_{n-1}, p_n, p_n)
\]

In the proposed method we have used a 2-dimensional median filter with a 3×3 window. As discussed earlier, the noise residual of the original images can be calculated according to Eq. (2). We suppress the original noise residual by applying a 3×3 median filter to the given unanonymized image. Next, we subtract the new noise residual from the original noise residual by multiplying it with a factor ‘a’, so that the PRNU terms in both the noise residuals become equal and cancel out, leaving no trace of the original image source. The procedure is described below.

I. We use the median filter to suppress both the PRNU noise F and the shot noise term \( \Phi_1 \). After applying median filter we are left with a new noise residual \( NR' \) which is equal to:

\[
NR' = mFP_0 + \Phi_2
\]

where \( m < 1 \), \( P_0 \) is the amount of incident light, F is the Camera Fingerprint, \( \Phi_2 \) is the suppressed shot noise (or Poisson noise) and variance(\( \Phi_2 \)) < variance(\( \Phi_1 \)).

II. To remove the PRNU term from the given image, we multiply the suppressed noise residual with the factor \( \alpha = 1/m \), and then subtract the
resultant noise residual from the original image.

\[ P'_x = P_x - \alpha NR' \]  \hspace{1cm} (9)

From Eq. (1) and Eq. (9), we obtain:

\[ P'_x = P_0 + P_0 F + \phi_1 - \alpha (mF P_0 + \phi_2) \]  \hspace{1cm} (10)

which results in

\[ P'_x = P_0 + (\phi_1 - \alpha \phi_2) \]  \hspace{1cm} (11)

Thus, the output image \( P_x \) is free from the PRNU pattern noise. The next step is to calculate the value of \( \alpha \) accurately to efficiently remove the camera fingerprint from the given image.

III. To find the optimal value of \( \alpha \) (best suited for a given dataset), we do an iterative search while trying to obtain the minimal PCE value of the given image. The conventional decision threshold for PCE value according to current literature is 50, as discussed previously in Section 3. Hence, any PCE value less than 50 is good enough to ensure the anonymity of the image; however, we try to find the minimal value, because lesser the value, greater is the degree of source anonymization achieved. We calculate the value of \( \alpha \) as follows:

\[ \alpha = \arg_{\alpha \in [1, \infty)} \min(PCE((P'_x), F)) \]  \hspace{1cm} (12)

Equation 12 suggests that the PCE value (in every iteration) is calculated as a function of \( P'_x \) and \( F \), which is in accordance with our discussion in Section 3. We initialize with \( \alpha = 1 \) in the first iteration, and keep incrementing \( \alpha \) as \( \alpha_i = \alpha_{i-1} + (\alpha_{i-1}/10) \) in subsequent iterations \( 1 \leq i < \infty \).

Next, we present the pseudo-code representation of the proposed source anonymization technique:

**PROPOSED SOURCE ANONYMIZATION**

**INPUT:** Training Images, Test Image.

**OUTPUT:** Anonymized Test Image \( I_{ANON} \).

1. Read Training Images
2. CameraFingerprint = getFingerprint( Training Images )
3. Read Test Image
4. \( I_{ANON} = \) MedianFilter( Test Image )
5. Compute \( V_{PCE} = PCE( \text{Test Image, Fingerprint} ) \)
6. Set \( x_0 = 0, x_1 = 1 \) and \( x_2 = 0 \)
7. While \( (V_{PCE} > 0) \)
8. \( NR = \) Test Image – \( I_{ANON} \)
9. \( x_2 = x_1 + ((x_1-x_0)/10) \)
10. \( x_0 = x_1 \) and \( x_1 = x_2 \)
11. \( NR' = x_2 * NR \)
12. \( I_{ANON} = \) Test Image – \( NR' \)
13. \( V_{PCE} = PCE( I_{ANON}, \text{CameraFingerprint} ) \)
14. End Loop
15. Return ( \( I_{ANON} \) )

The functions used in the above algorithm, along with their input and output parameters are listed below:

1. Function getFingerprint:
   **INPUT:** Training Images
   **OUTPUT:** CameraFingerprint or PRNU estimate

2. Function MedianFilter
   **INPUT:** Image
   **OUTPUT:** Denoised Image

3. Function PCE
   **INPUT:** Image, CameraFingerprint
   **OUTPUT:** PCE Value

We initially checked the PCE value of the original image with respect to the camera fingerprint, (which was observed to be in the range of 125 to 850 for the present application). We then decrease the PCE value gradually through the above iterative search process, where we keep on subtracting the suppressed noise residual from the original image. The \( \alpha \) value corresponding to the minimal value of PCE found, is decided to be the optimal one.
5. EXPERIMENTAL RESULTS AND DISCUSSION

Our proposed method has been implemented in MATLAB using its Image Processing Toolbox. For our experiments, we have used images captured by cameras of four different makes and models, the makes being Sony, Canon, Kodak and Ricoh. The images have been collected from the Dresden Image Database [18], which is a standard database used by forensic researchers worldwide.

The Dresden Image Database [18] is adopted as a standard benchmark for evaluation of forensic techniques, by researchers world-wide. This database has been widely used for the purpose of benchmarking camera-based digital forensic techniques. The database consists of more than 14,000 images of various indoor and outdoor scenes, acquired under controlled and comparable environmental conditions using altogether 73 digital cameras from 25 different models, to ensure that device-specific and model-specific characteristics could be disentangled and studied separately. To study the device-specific sensor noise pattern of all cameras in the image database, 50 dark frame images and 50 flatfield images were acquired for each device. The lens was covered to acquire the dark frames and a homogeneously backlit screen was used to acquire the flatfield frames. These auxiliary images were made ready to analyze the stable parts of the sensor noise pattern separately: fixed pattern noise (FPU) and photo-response non-uniformity (PRNU). So, this enables, among others, research on ways to suppress or to forge the device-specific sensor noise pattern.

In this paper, we compare the performance of proposed method with the very recent state-of-the-art technique proposed by Dirik and Karakucuk, called the Adaptive PRNU Denoising or simply APD-1 [7]. First, we estimated a value of the PRNU camera fingerprint $F_x$, using 80 images (training set) of each camera model, so that now, we have the camera fingerprints of all four camera models used in the experiment. Next, we calculate the PCE value of each (original) image with respect to the fingerprint of its authentic source. This gives us the PCE of the unanonymized images. We conducted the experiment with another 40 different images (test set) from each of the four camera models. The PCE values of the original unanonymized images are shown in Table 1.

<table>
<thead>
<tr>
<th>CAMERA</th>
<th>AVERAGE PCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SONY</td>
<td>491.8482</td>
</tr>
<tr>
<td>CANON</td>
<td>129.0365</td>
</tr>
<tr>
<td>KODAK</td>
<td>144.1907</td>
</tr>
<tr>
<td>RICOH</td>
<td>127.1017</td>
</tr>
</tbody>
</table>

Now that we have the PCE values of the original test images, we apply the proposed method to remove the PRNU patterns from the images. Lowering the PCE values of an image to less than the decision threshold would effectively anonymize the image, rendering source identification impossible. A decision threshold of 50 for PCE has been adopted in our experiments.

In order to achieve optimal performance by the proposed technique, we try to lower down the PCE of the images to their minimum values. This is because smaller the PCE value, greater is the degree of anonymization. We applied the proposed iterative search process to estimate the value of $\alpha$ and minimize the PCE values of the original images. The PCE values of the anonymized images obtained by the proposed technique are presented in Table 1.
model, obtained by applying the proposed method. It is evident from
Fig. 1 that even after source anonymization, the degradation in quality of
the images produced by the proposed method is insignificant; hence, this would prevent an
outsider to have any hint of source anonymization carried on the images.

We compare the performance of the proposed method with APD-1 in terms of
source anonymization, which we measure using
Peak-to-Correlation Energy (PCE) ratio,
defined in Section 3. For implementing APD-1,
we used a spatial domain 2D-Wiener filter [7]
to first de-noise the image and then calculate
the Noise Residual. The comparison results
have been shown in Table 2, where we have
presented the PCE values of the anonymized
images, averaged over the entire test set from
each camera model, corresponding to the
proposed method vis-à-vis APD-1. In our
work, we have used APD-1 as comparison
benchmark because of its efficiency in source
camera anonymization (as shown in Table 2).
However, the proposed method outperforms
APD-1, providing a higher degree of
anonymization. As evident from Table 2, the
PCE values obtained by the proposed method
is considerably lower for each camera model, as
compared to that of APD-1. So, we can infer
that the proposed technique succeeds to
achieve a better degree of source anonymization.

### 6. CONCLUSION

In this paper, we have dealt with the problem
of source anonymization of digital images by
suppressing PRNU noise based camera
fingerprints acquired from the images. For this
purpose, we used a median filter for denoising
the source images and the PRNU noise was
removed iteratively. We considered digital

---

Table 2

<table>
<thead>
<tr>
<th>CAMERA MODEL</th>
<th>PCE VALUE</th>
<th>PROPOSED METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SONY</td>
<td>127.1017</td>
<td>1.0823</td>
</tr>
<tr>
<td>CANON</td>
<td>0.0104</td>
<td>0.0027</td>
</tr>
<tr>
<td>KODAK</td>
<td>0.0974</td>
<td>0.0073</td>
</tr>
<tr>
<td>RICOH</td>
<td>0.5369</td>
<td>0.0234</td>
</tr>
<tr>
<td>Average</td>
<td>0.43685</td>
<td>0.27895</td>
</tr>
</tbody>
</table>

In Fig. 1, we have presented four different
anonymized test images, one from each camera
cameras of four different make and model for our experiments and applied the proposed method to the test images obtained from these source cameras. Our experimental results prove that the proposed method outperforms the state-of-the-art Adaptive PRNU Denoising–1 [7] source anonymization technique. We found that while APD–1 could decrease the PCE value only up to 0.43685 on an average, the proposed method succeeded to lower the PCE value down to 0.27895 on an average.

Future research directions include utilizing different other appropriate filters in order to achieve better degrees of source anonymity. Future research in this direction would also include investigation of more recent and efficient state-of-the-art source anonymization techniques, and hence comparison of the proposed technique with those. Along with source–anonymity, future research would also involve investigation of image quality, so as to ensure that there is no significant degradation in the quality of images due to source anonymization.
REFERENCES


