Analysis of Seam-Carving-Based Anonymization of Images Against PRNU Noise Pattern-Based Source Attribution

Ahmet Emir Dirik, Hüsrev Taha Sencar, and Nasir Memon

***Abstract*— The availability of sophisticated source attribution techniques raises new concerns about privacy and anonymity of photographers, activists, and human right defenders who need to stay anonymous while spreading their images and videos. Recently, the use of seam-carving, a content-aware resizing method, has been proposed to anonymize the source camera of images against the well-known photoresponse nonuniformity (PRNU)-based source attribution technique. In this paper, we provide an analysis of the seam-carving-based source camera anonymization method by determining the limits of its performance introducing two adversarial models. Our analysis shows that the effectiveness of the deanonymization attacks depend on various factors that include the parameters of the seam-carving method, strength of the PRNU noise pattern of the camera, and an adversary's ability to identify uncarved image blocks in a seam-carved image. Our results show that, for the general case, there should not be many uncarved blocks larger than the size of 50×50 pixels for successful anonymization of the source camera.**

***Index Terms*— PRNU noise pattern, seam-carving, source attribution, anonymization, de-anonymization attacks, counter-forensics.**

I. Introduction

MULTIMEDIA forensics have made tremendous progress that has enabled analysis of images and videos in ways that were not possible before [1]. Among the various research areas that comprise media forensics, source attribution has attracted a great amount of research interest and inquiry [2]–[5]. Source attribution techniques essentially allow obtaining information about the imaging device used for capturing a given media content. Some of these techniques have more recently started to be used as evidence in court of law, with potentially many more to follow in the future. At the same time, however, the capabilities offered by these techniques raise some new questions concerning individual rights to privacy and anonymity. It is especially important for the photographers, activists, and human right defenders to stay anonymous while spreading their images and videos [6].

To date, the most effective source attribution method is based on photo-response non-uniformity (PRNU) noise pattern [2]. PRNU noise is an imperfection caused mainly by the impurities in silicon wafers and affects the light sensitivity of each individual pixel. This noise generates a unique pattern which is present in every image or frame captured by the imaging sensor; therefore, it acts as a fingerprint of the imaging device. This fingerprint can be obtained by image denoising techniques where the difference between the original image and its denoised version is treated as an estimate of the fingerprint. To determine whether a given image or video is captured by a particular device, a fingerprint is estimated from the given media and is matched to the available device fingerprint. The matching is performed either using normalized correlation (NC) [2] or peak to correlation energy (PCE) [7], which determines the sharpness of the correlation peak. A decision is made by comparing the measured statistic to a pre-determined threshold.

Ultimately, the problem of anonymizing images and videos against source attribution relies on the ability to circumvent underlying attribution techniques, and there are two main counter forensics approaches to achieve this goal. The first approach relies on weakening the PRNU fingerprint. This can be realized by subjecting the images to strong filtering or severe compression, so that PRNU fingerprint cannot be reliably extracted. This seemingly straightforward task is difficult to accomplish due to the image and video quality considerations. In [2], Lukáš *et al.* investigated the robustness of PRNU fingerprint to gamma correction and JPEG compression with various quality factors. Their results show that PRNU fingerprint can still be identified from images compressed at JPEG quality factor 50. Later, Rosenfeld *et al.* [8] investigated the robustness of PRNU fingerprint against common signal processing operations like denoising, re-compression and out-of-camera demosaicing. Their results show that even after eight rounds of denoising there is still a significant correlation between the noise pattern of the multiply-denoised image and the PRNU fingerprint of the camera. Similarly, it is shown that re-compression will not be able to eliminate the PRNU noise within tolerable limits of image quality loss, *i.e.,* above 25dB in PSNR. Their results also showed that although some demosaicing algorithms, like patterned pixel

Manuscript received February 17, 2014; revised August 5, 2014 and September 23, 2014; accepted September 30, 2014. Date of publication October 1, 2014; date of current version November 12, 2014. This work was supported by the Center for Interdisciplinary Studies in Security and Privacy, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gwenaël J. Doërr.

A. E. Dirik is with the Department of Electrical and Electronics Engineering, Faculty of Engineering, Uludağ University, Bursa 16120, Turkey, and also with New York University Abu Dhabi, Abu Dhabi 129188, United Arab Emirates (e-mail: edirik@uludag.edu.tr).

H. T. Sencar is with the TOBB University of Economics and Technology, Ankara 06520, Turkey, and also with New York University Abu Dhabi, Abu Dhabi 129188, United Arab Emirates (e-mail: htsencar@etu.edu.tr).

N. Memon is with the Polytechnic Institute New York University, Brooklyn, NY 11201 USA (e-mail: memon@nyu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2014.2361200

grouping and adaptive homogeneity-directed interpolation, are effective in reducing accuracy, they cannot completely remove the PRNU fingerprint. Another method deployed for PRNU noise suppression is based on the use of flat-fielding [9], [10]. One drawback of this approach, however, is that for effective removal of PRNU fingerprint, flat-fielding has to be applied to the photo sensor output right before demosaicing or any color correction. This makes flat-fielding approach inapplicable for most consumer cameras which do not support outputs in raw format [2]. Recently, Dirik *et al.* proposed an adaptive PRNU denoising method, which can remove PRNU fingerprint without compromising image quality significantly in terms of PSNR [11].

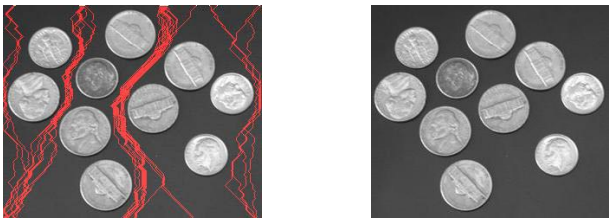
The second anonymization approach focuses on disturbing the alignment, rather than trying to remove the pattern itself, between the two PRNU fingerprints that are being matched. Since each element of the PRNU fingerprint is related to a pixel's light sensitivity, it is important to align the two noise values extracted from the same pixels when computing the matching statistic. Hence, geometric transformations, which involve operations like resizing, rotation, cropping, *etc.*, can very effectively impede the fingerprint matching process. Furthermore, this approach offers a better trade-off between image quality and anonymity as misalignment of two fingerprints can be achieved by introducing less disturbing artifacts as compared to removing the PRNU fingerprint through operations like heavy compression or filtering. However, it is shown that the parameters of such transformations can be determined by a brute-force search, and the PRNU fingerprints can still be aligned if they are of the same camera [12], [13]. Using resampling detection techniques the transform parameters of resizing or rotation can be estimated directly. As a result source identification can be achieved without any brute-force search. These works crucially showed that effective anonymization against image source attribution needs to incorporate irreversible transformations. In [14], Bayram *et al.* proposed a new anonymization technique along this direction based on a tailored version of the seam-carving algorithm to introduce irreversible misalignments while at the same time retaining as much image quality as possible. In this work, we contribute to the research in this direction by providing an analysis of source anonymity and consider a number of strategies an adversary may deploy to attain the source attribution.

In the following section, we provide details of the seam carving based anonymization and PRNU noise based source attribution. This is followed by analysis of matching statistic when small-sized noise patterns are used for source attribution. Section IV investigates the reliability of fingerprint matching using an abstract model of the described approach in [14]. Adversarial strategies and their impact on the effectiveness of source camera anonymization are discussed in Section IV-C. Our conclusions are given in Section V.

II. Background

A. Forced Seam Removal Based Anonymization

Seam carving has been introduced to perform content-aware image resizing and its main idea is based on finding



(a) (b)

Fig. 1. (a) Coins image with 50 vertical seams superposed. (b) Coins image after seam carving.

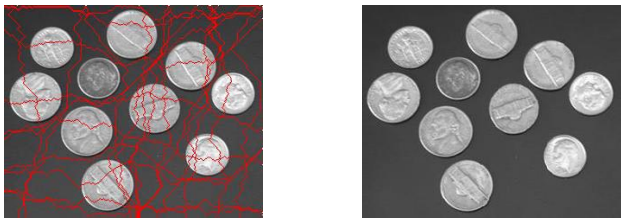
(a) (b)

Fig. 2. (a) Castle image with 100 vertical seams superposed. (b) Castle image after seam carving.

connected paths of pixels, *i.e.,* seams, in the image with the least variation from the surrounding pixels [15]. Seams may run in the horizontal or vertical direction of the image and are obtained from the gradient image by treating the gradient magnitudes associated with each pixel as the energy of that pixel. A dynamic programming method is used to identify the seams that have the minimum total energy. It is shown that seam carving method yields higher perceived image quality than resizing and may cause visual distortions when image has high energy content all across or a lot of regular geometric textures [15]. Coins and castle[1] images in Figures 1 and 2 show implementation of conventional seam carving algorithm by removing vertical seams. Figure 1.a and 2.a show the original images with vertical seams superposed. As can be seen, the seams do not cross the edges of the coins or the castle as the magnitude of gradient on the edges yield relatively high values. In both images, seam-carving has not introduced any perceptual distortion. Seam-carved images in Figs. 1-4 are from the paper [14].

When an image is downsized by seam-carving, pixels along the paths of the selected seams are removed, and all remaining pixels are shifted horizontally or vertically to fill the gap. What is more is that the analysis of the seam-carved image will not trivially yield information about the locations of the seams and the number of removed seams. In the context of source attribution, this loss of information is very critical. The fact that seam carving cannot be reversed assures that the PRNU noise pattern obtained from a seam-carved image will not align with that of the corresponding source camera. One important limitation of conventional seam carving is that, in the high gradient image regions, it will leave large uncarved image blocks to preserve the visual quality. Essentially, such large

[1] http://commons.wikimedia.org/wiki/File:Broadway_tower.jpg



(a) (b)

Fig. 3. (a) Coins image with random vertical and horizontal seams superposed. (b) Coins image after forced seam carving.

(a) (b)

Fig. 4. (a) Castle image with random vertical and horizontal seams superposed. (b) Castle image after forced seam carving.

blocks can be identified relatively easily and used for matching at the expense of a simple search of the noise pattern of the uncarved block in the camera fingerprint.

To address this deficiency, Bayram *et al.* proposed incorporating a forced seam removal approach that performs seam carving by forcing seams to pass through selected pixels so that the process leaves no uncarved image blocks that are large enough to yield reliable statistic for successful matching. Figures 3 and 4 demonstrate the forced seam carving process implemented on the same images given above [14]. Although there is not any significant difference in image quality for the two examples, the seams selected on the basis of minimising the size of the largest uncarved block is likely to introduce higher visual distortions. What is more critical is, as noted by the authors, an adversary, who attempts to attribute an image subjected to forced seam carving to its source, might utilise multiple uncarved blocks during matching. The likely success of such adversarial measures and the dependency on the size and the number of the uncarved blocks were not investigated in [14]. To close this gap, in this paper, we describe two specific attacks against forced seam carving based anonymization by utilising sequential hypothesis test and merging multiple blocks together and investigate their effectiveness. It must be noted that, throughout the paper, we use the term attack to refer to attempts to compromise the anonymity of a seam-carved image by reliably attributing the image to its source camera.

B. PRNU Noise Pattern Based Source Attribution

In this section, we present the details of PRNU based source attribution method and provide analytic expressions for decision accuracy and error rates. In the paper, all vectors and matrices are denoted with bold capital letters and scalar values are shown in italic font. For simplicity in writing, we choose to use vectors instead of matrices and all matrices (image, PRNU fingerprint, *etc.*) are transformed to vectors. All vector and matrix operations, such as multiplication, division, and raising to a power, are element-wise unless it is mentioned before its use.

Let $\mathbf{I}$ be a selected color channel of a given image in raw image format. We assume $\mathbf{I}$ has not undergone any post processing such as demosaicing, gamma correction, filtering, color correction, optical correction, and lossy image compression. A simplified model of the sensor output is given by Chen *et al.* in [4] as

$$\mathbf{I} = \mathbf{I_0} + \mathbf{I_0}\mathbf{K} + \Theta, \tag{1}$$

where $\mathbf{I_0}$ is the ideal sensor output without any sensor noise or optical distortion, $\mathbf{K}$ is multiplicative PRNU factor (camera fingerprint), and Θ is the combination of other noises such as dark current, shot noise, read-out noise, and quantization noise. Using a maximum likelihood estimator, the camera fingerprint $\mathbf{K}$ can be estimated from l noise residuals ($\mathbf{W_i} : i = 1, \ldots, l$) obtained from l images ($\mathbf{I_i} : i = 1, \ldots, l$) taken from the same camera. The noise residual is obtained by subtracting the denoised version of the image from itself as $\mathbf{W_i} = \mathbf{I_i} - \Phi(\mathbf{I_i})$, where Φ is a denoising filter. The maximum likelihood estimator of $\mathbf{K}$ is defined in [4] as

$$\mathbf{F} = \frac{\sum_{i=1}^{l} \mathbf{W_i}\mathbf{I_i}}{\sum_{i=1}^{l} \mathbf{I_i}^2}. \tag{2}$$

(1) and (2) do not consider demosaicing or any filtering in the imaging pipeline. If the camera fingerprint (2) is estimated from color-interpolated and post-processed JPEG images, $\mathbf{F}$ contains artifacts of color interpolation which are not unique to the imaging device. To suppress these artifacts, all rows and columns of $\mathbf{F}$ are subtracted by their row and column averages, respectively. Then, the periodic artifacts are filtered in Fourier domain through Wiener filtering [4].

Assume $\mathbf{F_x}$ is the PRNU fingerprint estimate of camera X (C_x) obtained from multiple images taken by the same camera and $\mathbf{F_y}$ is the PRNU estimate of image $\mathbf{I}$. We would like to determine whether the source camera of $\mathbf{I}$ (C_y) is the same as C_x or not. Let $\mathbf{F_x}$ and $\mathbf{F_y}$ be independent and identically distributed (i.i.d.) random variables with zero mean and unit variance. Thus, the PRNU fingerprint estimates can be given by [14]

$$\mathbf{F_x} = \mathbf{K_x} + \theta_\mathbf{x}, \tag{3}$$

$$\mathbf{F_y} = \mathbf{K_y} + \theta_\mathbf{y}. \tag{4}$$

We assume both the PRNU fingerprint $\mathbf{K}$ and the error term θ are zero mean, i.i.d. random variables. Let the variance of the error term θ be σ^2. Hence, the variance of the PRNU fingerprint $\mathbf{K}$ becomes $1 - \sigma^2$. The problem of determining the source of $\mathbf{I}$ can be written as a binary hypothesis test as

$$H_0 : \mathbf{K_x} \neq \mathbf{K_y}, \tag{5}$$

$$H_1 : \mathbf{K_x} = \mathbf{K_y}, \tag{6}$$

where H_0 refers to the case the source camera of $\mathbf{I}$ is different from C_x. Thus, the PRNU fingerprints $\mathbf{K_x}$ and $\mathbf{K_y}$ become different. H_1 refers to the case C_x and C_y are the same.

If H_0 is rejected, $\mathbf{I}$ is assumed to be taken with C_x. The hypothesis tests in (5) and (6) can be verified by measuring the correlation coefficient between the PRNU estimates $\mathbf{F_x}$ and $\mathbf{F_y}$. The correlation coefficient $r(\mathbf{F_x}, \mathbf{F_y})$ is obtained by

$$r(\mathbf{F_x}, \mathbf{F_y}) = \frac{\sum_{i=1}^{s} F_{x_i} F_{y_i}}{\sqrt{\sum_{i=1}^{s} F_{x_i}^2} \sqrt{\sum_{i=1}^{s} F_{y_i}^2}}, \tag{7}$$

where s is the total number of elements of $\mathbf{F_x}$ and $\mathbf{F_y}$. A better similarity metric in terms of determining a fixed decision threshold for all cameras is Peak to Correlation Energy (PCE) [7]. For ease of analysis, we choose to use the correlation coefficient in the rest of the paper, taking into account that the decision threshold may vary for different camera models and brands.

III. Image Source Attribution for Seam-Carved Images

A. Modeling the Correlation Coefficient for Small Blocks

Most of the available cameras today store images in a processed and compressed format. Therefore, distribution of the PRNU noise and matching statistic not only depend on imaging sensor but also the post processing applied in the imaging pipeline. Since the Gaussian assumption on the correlation coefficient also depends on the number pixels in an image, when source camera is to be identified from a small uncarved block of a seam-carved image, this assumption may not hold. Therefore, to compute the decision error rates for typical camera output images, we have to estimate the corresponding probability density function (pdf) for the correlation coefficient r under H_0 and H_1 hypotheses. The Generalized Gaussian distribution can be used to model the correlation coefficient r under H_0 as it is shown in [2] and [4]. The Logistic and the lognormal distributions were suggested to model r in [16] for sensor noise estimates computed by anisotropic diffusion algorithm under H_0 and H_1, respectively.

One challenge here is that the pdf of r under H_1 may vary for different imaging sensors because of different sensor types and post-processing steps. Another challenge concerns finding a parametric model that fits into the distribution of r for small image blocks. From the Central Limit Theorem, we can model the distribution of the correlation coefficient r for large s (total number of pixels) under H_0 [12] as

$$r_{H_0} \sim N(0, \frac{1}{s}). \tag{8}$$

Our experimental observations on JPEG images taken by six digital cameras including DSLR, compact, and mobile phone cameras show that the distribution of the correlation coefficient r for the *non-matching* case fits the Gaussian distribution function (see Figs. 5 and 6). All images used in the experiments were captured at the highest camera quality. All the other settings (like ISO, white balance, *etc.*) were set automatically by the camera. The images were acquired both indoor and outdoor and are of different content. Corresponding PRNU noise estimates were computed by performing wavelet denoising in RGB color channels as it is initially proposed in [17].

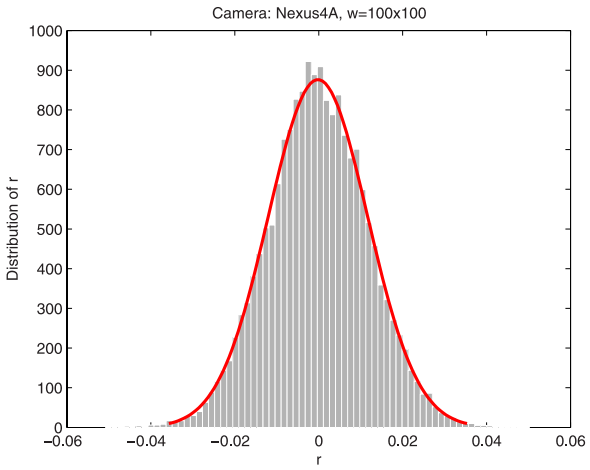


Fig. 5. The correlation coefficient is modeled with the Gaussian distribution for non matching case. (Camera: Nexus; Block size: 100×100 pixels; Number of blocks: 18,050).

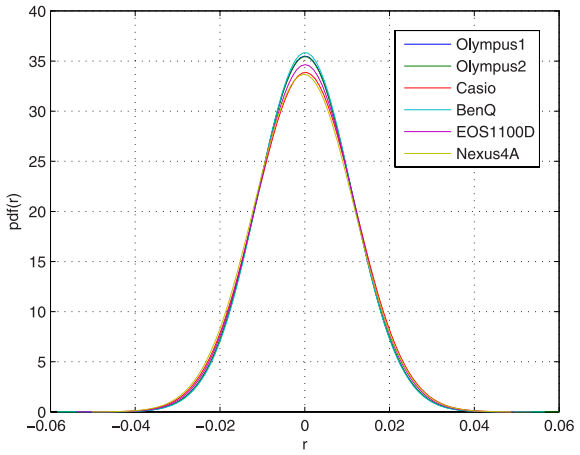


Fig. 6. Gaussian fits for the correlation coefficient r for non matching case (6 cameras); Block size: 100×100 pixels; Num. of blocks: 20,000; Estimated variance: $\sigma^2 = 1.3136 \times 10^{-4}$.

In Fig. 6, the pdf estimates of the correlation coefficient for the non-matching case are depicted for six cameras. The cameras used in the experiment are: Canon EOS1100D, LG Nexus, Olympus D745 (2), Casio QV-R200, and BenQ AE100. In the experiment, the correlation coefficients were measured from 100×100 pixels size blocks of the test images (361 blocks/image) using the camera fingerprint estimates. For each camera, there are 50 images taken with the highest camera quality in the test set with varying content. The camera fingerprints were estimated from 200 images not used during the tests. The average variance of the distributions in Fig. 6 is $1.3136 \times 10^{-4} \approx \frac{1}{100 \times 100} = \frac{1}{s}$. This measurement is in good agreement with the correlation model in (8). On the other hand, the distribution of the correlation coefficient for the *matching* case was found to be asymmetric and skewed to the right. As an example, the distribution of the correlation coefficient for Nexus under H_1 is depicted in Fig. 7.

Our experimental analysis on six cameras indicate that the generalized extreme value (GEV) distribution [18] fits well to the pdf of r computed from small image sub-blocks

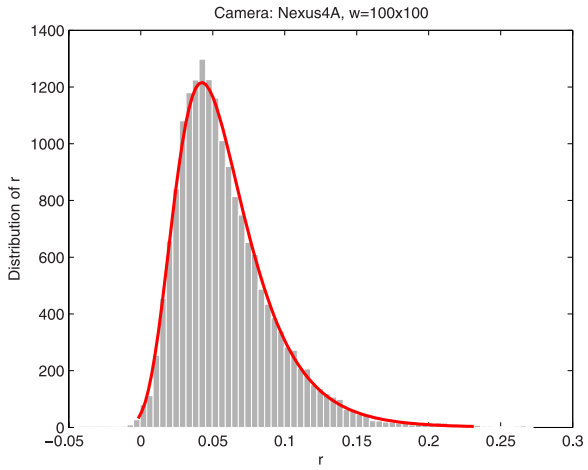


Fig. 7. The correlation coefficient is modeled with the Generalized Extreme Value (GEV) distribution for the matching case. (Camera: Nexus; Block size: 100 × 100 pixels; Number of blocks: 18,050).

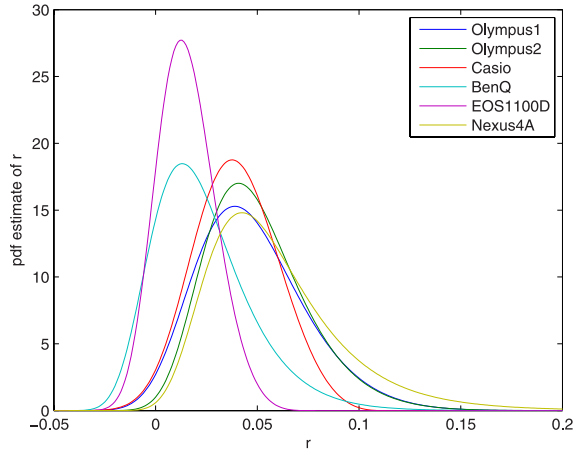


Fig. 8. The pdf (generalized extreme value distribution) estimates of the correlation coefficient under H_1 for six cameras including mobile, DSLR, and compact cameras when the block size is set to 100 × 100 pixels.

(see Figs. 7 and 8). The GEV distribution combines three types of extreme value distributions into a single mathematical form; therefore, it is extremely versatile to model data [19]. The GEV distribution has three parameters: μ (location), σ (scale), and ξ (shape):

$$\rho_{match} \sim GEV(\mu, \sigma, \xi). \tag{9}$$

The cumulative distribution function (cdf) of GEV is given by

$$F(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}. \tag{10}$$

The pdf of GEV is:

$$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma}e^{-t(x)}t(x)^{\xi+1}, \tag{11}$$

$$t(x) = -\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}. \tag{12}$$

Fig. 8 and Table I show that pdf parameters of GEV distribution depend on sub-image dimensions and camera models. To investigate the effect of various image dimensions

TABLE I
THE PARAMETERS OF THE GEV DISTRIBUTION MODEL FOR VARIOUS IMAGE DIMENSIONS AND CAMERAS

GEV model parameters	Nexus 4 100 × 100	Canon EOS 100 × 100	Nexus 4 50 × 50	Canon EOS 50 × 50
μ	0.043430	0.009644	0.028779	0.004418
σ	0.024866	0.013545	0.025656	0.019649
ξ	0.039763	-0.196665	-0.001430	-0.193134

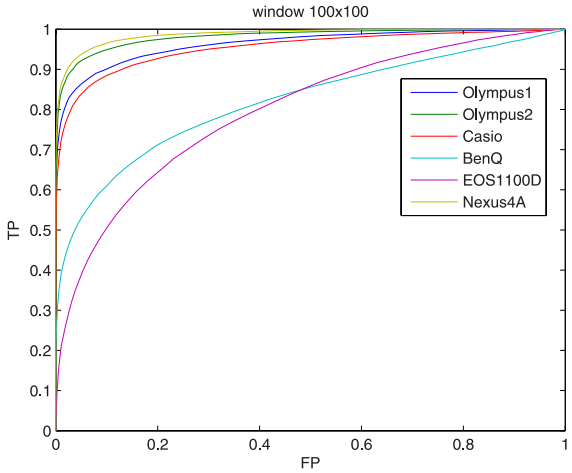


Fig. 9. The ROC curves for 100 × 100 pixels sub-blocks.

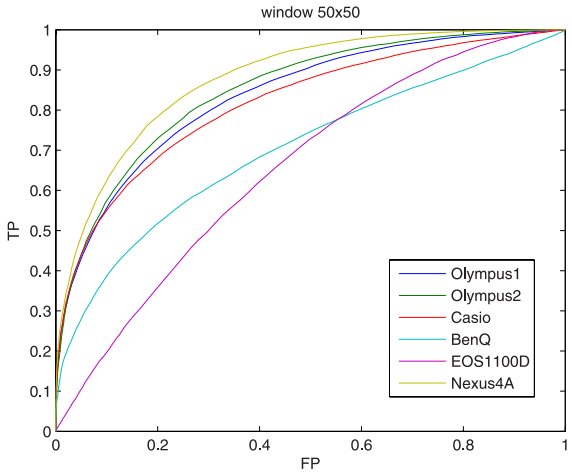


Fig. 10. The ROC curves for 50 × 50 pixels sub-blocks.

on source camera identification, we measured the correlation coefficient for smaller blocks (of size 50 × 50 pixels). For each camera, we measured the correlation coefficient using the corresponding camera fingerprint estimate from 20,000 sub-blocks of 50 test images. The receiver operating characteristic (ROC) curves of the six cameras for 100 × 100 pixels and 50 × 50 pixels size block dimensions are given in Figs. 9 and 10, respectively.

The ROC curves show that it is hard to make generalizations on the performance of source camera identification for small image dimensions. Nevertheless, it is clearly seen in Fig. 10 that high detection accuracy cannot be achieved for smaller block sizes (50 × 50 pixels or less) for all cameras.

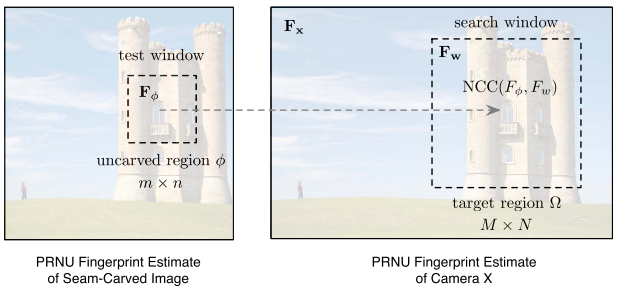


Fig. 11. Source attribution by NCC from uncarved region.

In Figs. 9 and 10, there are big gaps between the ROC curves of Nexus and Canon EOS1100D wherein the most reliable attribution is achieved with the Nexus while the worst with the Canon EOS1100D. This is mostly because the DSLR camera EOS1100D has significantly larger sensor dimensions and less noise as compared to the other camera sensors. Since it is impractical to test on all digital cameras and phones in the market, we choose the ROC curves of Nexus and Canon EOS1100D as the rough upper and lower bounds of the source camera identification for small image dimensions. The estimated GEV model parameters of the Nexus 4 and Canon EOS1100D for 50×50 pixels and 100×100 pixels windows are given in Table I. In the rest of the paper, all analysis will be given using these two cameras and their estimated GEV distribution models.

B. Fingerprint Verification for Seam-Carved Images

Consider a scenario that an image $\mathbf{I}$ has been subjected to seam-carving as described in Section II-A. Let the seam-carved image be $\mathbf{I_{sc}}$. We would like to determine whether $\mathbf{I_{sc}}$ is taken by camera X or not. Assume we have access to the PRNU noise pattern based fingerprint, $\mathbf{F_x}$, of camera X. Since seam-carving destroys the pixel-to-pixel relations between $\mathbf{I}$ and $\mathbf{I_{sc}}$, the original fingerprint estimate of $\mathbf{I}$ cannot be obtained from the image $\mathbf{I_{sc}}$. On-the-other-hand, some regions in $\mathbf{I_{sc}}$, like the high gradient energy regions, will not be subjected to seam-carving. These uncarved regions contain some parts of the original PRNU fingerprint estimate $\mathbf{F_y}$, and if these regions are identified, source camera identification can be realized.

Assume there is an uncarved $m \times n$ pixels size region ϕ in a seam-carved image and we would like to search for the PRNU fingerprint estimate associated with the region ϕ ($\mathbf{F}_\phi$) within the camera fingerprint $\mathbf{F_x}$ (as depicted in Fig. 11). Searching a match between $\mathbf{F}_\phi$ and $\mathbf{F_x}$ can be achieved by normalized cross correlation (NCC) [20]. NCC computation time can be significantly narrowed by estimating the relative position of $\mathbf{F}_\phi$ in $\mathbf{F_x}$. Assume that we search $\mathbf{F}_\phi$ in a larger $M \times N$ pixels size region $\mathbf{\Omega}$ of $\mathbf{F_x}$ using NCC. Let the PRNU fingerprint estimate of the region $\mathbf{\Omega}$ be $\mathbf{F_w}$ such that $\mathbf{F_w} \subset \mathbf{F_x}$. The output of NCC between $\mathbf{F}_\phi$ and $\mathbf{F_w}$ is a $(M-m+1) \times (N-n+1)$ matrix comprising the correlation coefficients computed from all shifts of $\mathbf{F}_\phi$ over $\mathbf{F_w}$. Let the total number of elements of the NCC output be L. To identify the source camera, we pick the maximum NCC value (ρ_{max}) and check whether ρ_{max} is above the decision threshold τ or not.

Let ρ_{match} be the matching NCC value and ρ_{no_i} be the non-matching NCC value, respectively. If $\mathbf{F_w}$ contains a matching sub-window with $\mathbf{F}_\phi$, then there would be one matching correlation (ρ_{match}), and $L-1$ number of non-matching correlation values ($\rho_{no_i} \in \rho_{\mathbf{no}} : i = 1, \ldots, L-1$) in the NCC output. $\rho_{\mathbf{no}}$ is a vector containing all non-matching correlation values of NCC matrix. The correlation coefficient (ρ_{no_i}) for the non-matching case can be modeled using (8):

$$\rho_{no_i} \sim N(0, \frac{1}{mn}). \tag{13}$$

To model the matching case, we will use the GEV distribution estimates of Nexus (for the best performance case), and Canon EOS1100D (for the worst performance case). Hence, the upper and the lower bounds for source camera identification errors can be roughly estimated. The probability of missing a correct match can be written as follows:

$$P_{miss} = Pr(\rho_{match} \leq \tau) + Pr(max(\rho_{\mathbf{no}}) > \rho_{match} \geq \tau). \tag{14}$$

The probability of matching correlation, ρ_{match}, being lower than the decision threshold, τ, can be expressed with the cumulative distribution function of GEV:

$$Pr(\rho_{match} \leq \tau) = F_{\rho_{match}}(\tau; \mu, \sigma, \xi). \tag{15}$$

Let y be the maximum value of a random vector $\mathbf{x}$ involving k i.i.d. normal random variables. The cumulative distribution function of y ($F_y(\tau)$) is given by

$$F_y(\tau) = Pr(max(\mathbf{x}) \leq \tau), \tag{16}$$

$$= Pr(x \leq \tau)^k. \tag{17}$$

Hence, $Pr(max(\rho_{\mathbf{no}}) > \tau)$ can be written as

$$Pr(max(\rho_{\mathbf{no}}) > \tau), = 1 - Pr(max(\rho_{\mathbf{no}}) \leq \tau), \tag{18}$$

$$= 1 - Pr(\rho_{no_i} \leq \tau)^{L-1}, \tag{19}$$

$$= 1 - [1 - Q(\tau\sqrt{mn})]^{L-1}, \tag{20}$$

where $Q(\cdot)$ denotes the Q-function used for computing the tail probability of the standard normal distribution. $Pr(max(\rho_{\mathbf{no}}) > \rho_{match} \geq \tau)$ equation can be computed by integrating ρ_{match} from τ to ∞. That is,

$$Pr(max(\rho_{\mathbf{no}}) > \rho_{match} \geq \tau) = \int_{z=\tau}^{\infty} Pr(max(\rho_{\mathbf{no}}) > z \mid \rho_{match} = z)\ f_{\rho_{match}}(z)dz, \tag{21}$$

$$= \int_{z=\tau}^{\infty} \{1 - [1 - Q(z\sqrt{mn})]^{L-1}\}\ f_{\rho_{match}}(z)dz, \tag{22}$$

where $f_{\rho_{match}}(z)$ is the GEV of ρ_{match}. Hence, the probability of miss can be obtained by substituting (15) and (22) into (14). The probability of finding a correct match (true positive) can be obtained as

$$P_{TP} = 1 - P_{miss}. \tag{23}$$

Similarly, the probability of finding a false match (P_{FM}) at a wrong position under H_1 hypothesis can be expressed as

$$P_{FM} = Pr(max(\rho_{\mathbf{no}}) > \rho_{match} \text{ and } max(\rho_{\mathbf{no}}) > \tau). \tag{24}$$

Equation (24) can be split into two terms:

$$P_{FM} = Pr(max(\rho_{\mathbf{no}}) > \rho_{match} \geq \tau) + Pr(max(\rho_{\mathbf{no}}) > \tau \geq \rho_{match}). \quad (25)$$

The first term in (25) is the same as (21), and the second one can be re-arranged as

$$Pr(max(\rho_{\mathbf{no}}) > \tau \geq \rho_{match}) = Pr(max(\rho_{\mathbf{no}}) > \tau)\ Pr(\rho_{match} \leq \tau), \quad (26)$$

and evaluated using (18) and (15). It should be noted that the probability of false match in (24) is not the false positive rate. P_{FM} refers to the case where the seam-carved image is taken by the source camera in question (under H_1 hypothesis) but the correlation peak is detected at an incorrect location.

Taking into account (18), the probability of false detection P_{FP} under H_0 hypothesis (*i.e.*, seam-carved image is not taken by the given source camera) is obtained as

$$P_{FP} = 1 - Pr(\rho_{no_i} \leq \tau)^L, = 1 - [1 - Q(\tau\sqrt{mn})]^L. \quad (27)$$

Hence, given the (GEV) distribution of the correlation coefficient under H_1 hypothesis, one could compute the probability of detection (P_{TP}) and the probability of false positive (P_{FP}) using (23) and (27), respectively.

Assume we would like to determine the source camera of a seam-carved image from $m \times n$ pixels size test window where $m \times n$ is set to 100×100 pixels and the dimension of the search window $\mathbf{F_w}$ is denoted by $M \times N$ as before. From the estimated GEV model parameters given in Table I, we can compute the ROC curves corresponding to the Nexus 4 and Canon EOS1100D cameras for various sizes of search windows as presented in Figs. 13 and 15. The blue curves in the figures refer to the case where $M \times N = m \times n$, and they are computed using the correlation measurements for the matching and non-matching cases. The ROC curves in Figs. 13 and 15 show that the probability of detection decreases as the search window $\mathbf{F_w}$ gets larger in size. Similarly, Tables II and III provide the true positive rates for the two cameras when the largest uncarved region in a seam-carved image is assumed to be 100×100 and 50×50 pixels, respectively, at a fixed false alarm rate of 10^{-4}.

When the size of the test window is fixed and the size of the search window is increased under H_1, the maximum of the non-matching cross correlations ($max(\rho_{\mathbf{no}})$) increases as well. On the contrary, the matching correlation ρ_{match} in NCC vector is not affected from this change. As a result, both the false positive rate and the camera identification accuracy decrease as the search windows get larger. It can also be seen that when the uncarved block size is 100×100 pixels, probability of identification for the Nexus 4 varies in the range of $0.3659 - 0.5663$ (depending on the search window size), whereas for Canon EOS1100D it varies in the range of $0.0005 - 0.0160$. When the uncarved block size is reduced to 50×50, however, measured probability values for both cameras become sufficiently small.

C. Experimental Evaluation

To evaluate the validity of the simulation results, we conducted new experiments considering the H_1 hypothesis, where the search window contains the test block, on images taken by the Nexus 4 and Canon EOS1100D cameras. The camera PRNU fingerprints were estimated from 100 images and the source camera identification tests were conducted on 20 test images not used for camera fingerprint estimation. For the experiments, we randomly picked a test window of size 100×100 pixels and searched it within a larger region (with varying sizes of 110×110, 150×150, 250×250, 250×250 pixels) by normalized cross correlation. We repeated the block search 10,000 and 100,000 times for the Nexus 4 and Canon 1100D cameras, respectively.

Fig. 12 compares the probability of false positive (P_{FP}) measured with the experiments and simulations based on the model. To obtain the empirical P_{FP} curve, a 50×50 pixel-sized test window was searched within a 150×150 pixels window in 20 test images taken by the Nexus 4. (This experiment was repeated 10,000 times to obtain a smockoth curve.) It is found that, the empirically obtained P_{FP} values are higher than the ones estimated by the model. This mismatch between the model and the empirical results can be mainly attributed to interference from the content, which becomes predominant at smaller scales and effectively degrades the block search accuracy by increasing P_{FP}. Therefore, the presented model for P_{FP} based on i.i.d. assumption will yields to performance upper bound for the adversary, since in practice he would observe a higher rate of false-positives.

Fig. 14 depicts the empirical ROC curves for the Nexus 4 considering various sizes of search windows. It can be seen in Fig. 14 that the empirical ROC curves (for search windows of 110×110, 150×150, 250×250) are slightly below the curves obtained by the simulations given in Fig. 13. Similarly, Fig. 16 depicts the empirical ROC curves of Canon EOS1100D for search window sizes of 110×110, 150×150, and 250×250 pixels. Again, the empirical ROC curves of Canon EOS1100D are below the curves obtained through simulations using the model that are presented in Fig. 15. As it is shown in Fig. 12, given a fixed decision threshold, the measured false positive rate is greater than the false positive rate estimated by the model for the non matching case (H_0). For the matching case (H_1), we compute the GEV parameters from the actual correlation measurements so we can estimate the true positive rate accurately. As a result, the estimated ROC curves by the model indicate better detection performance and establish an upper bound on ROC for searching a small block in a larger window.

For the no-search case, where the sizes of test and search windows are equal, it can be observed that this characteristic still holds for the Nexus 4 camera (*i.e.,* the ROC curve estimated by the model in Fig. 13 is below the corresponding one given in Fig. 14) For the Canon EOS1100D camera, however, the two curves associated with the no-search cases, given in Figs. 15 and 16, are observed to closely follow each other.

TABLE II

CAMERA IDENTIFICATION PERFORMANCE USING 100 × 100 PIXELS SIZE TEST WINDOW ($P_{FP} = 10^{-4}$)

Search window (M × N)	True Positive (Nexus 4)	True Positive (Canon EOS)
110 × 110	0.5663	0.0160
150 × 150	0.4846	0.0055
250 × 250	0.4350	0.0024
500 × 500	0.3954	0.0010
1000 × 1000	0.3659	0.0005

TABLE III

CAMERA IDENTIFICATION PERFORMANCE USING 50 × 50 PIXELS SIZE TEST WINDOW ($P_{FP} = 10^{-4}$)

Search window (M × N)	True Positive (Nexus 4)	True Positive (Canon EOS)
110 × 110	0.0430	3.4046^{-14}
150 × 150	0.0375	8.2361^{-17}
250 × 250	0.0314	2.1401^{-21}
500 × 500	0.0256	1.3682^{-24}
1000 × 1000	0.0214	1.9190^{-23}

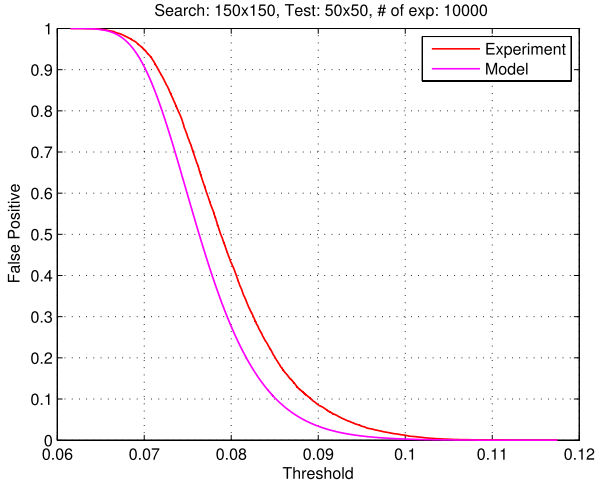


Fig. 12. False positive probability (P_{FP}) associated with using ρ_{max} as the matching statistic under H_0. The model is compared with the real measurements for the Nexus 4 camera.

IV. ATTACKING SEAM-CARVING BASED ANONYMIZATION

The results of the previous section show that reliable source identification is not possible when the size of the largest uncarved block is as small as 50 × 50 pixels. However, an adversary may attempt to find several such uncarved blocks as depicted in Fig. 17 and try to utilise them when performing source attribution. It is intuitive that utilising more blocks will lead to more reliable decisions. In this section, we introduce two adversarial scenarios against seam-carving based image anonymization and analyze the performance by both simulations and experiments.

A. Sequential Likelihood Ratio Test

Let k be the number of uncarved blocks to be tested in a seam-carved image. For more reliable source identification, an adversary may try to identify multiple uncarved blocks by checking whether or not they yielded a matching statistic that exceeded the decision threshold, as described in Section III-B. However, instead of evaluating the matching results individu-

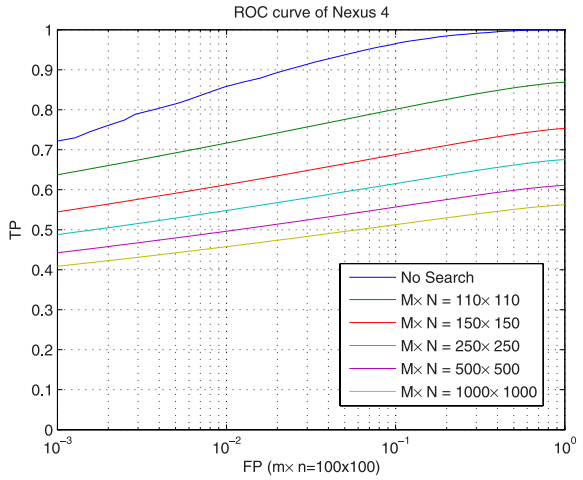


Fig. 13. The ROC curves estimated by the *model* for seam-carved images. A small region (m×n pixels) in the seam-carved image is searched within the camera fingerprint of the Nexus 4 camera.

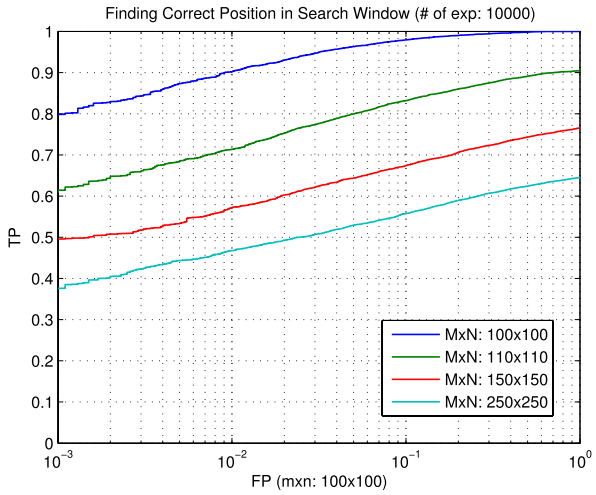


Fig. 14. The ROC curves estimated by the experiments for seam-carved images. A small region (m×n pixels) in the seam-carved image is searched within the camera fingerprint of the Nexus 4 camera.

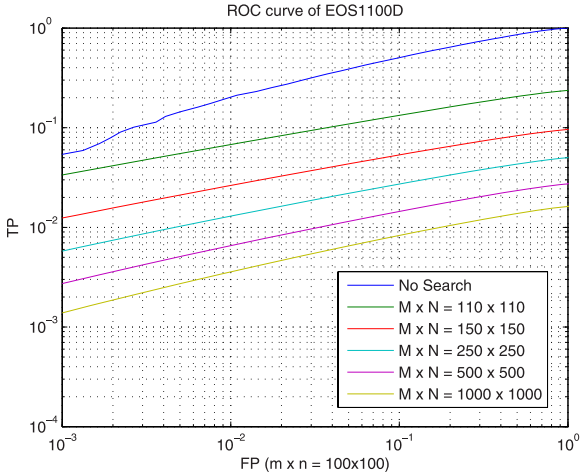


Fig. 15. The ROC curves estimated by the *model* for seam-carved images. A small region (m×n pixels) in the seam-carved image is searched within the camera fingerprint of the Canon EOS1100D camera.

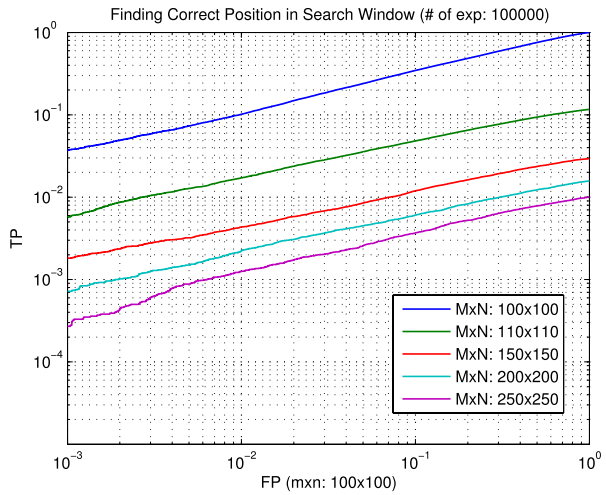


Fig. 16. The ROC curves estimated by the experiments for seam-carved images. A small region (m×n pixels) in the seam-carved image is searched within the camera fingerprint of the Canon EOS1100D camera.

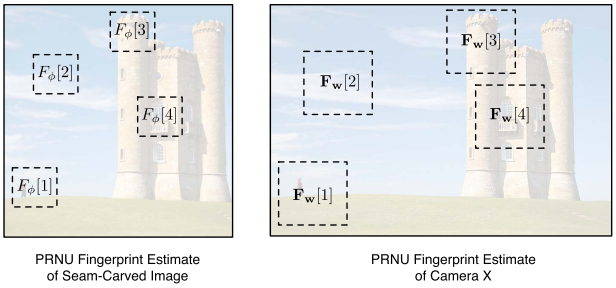


Fig. 17. Source attribution based on multiple uncarved blocks.

ally, all obtained matching statistic can be mapped to a single statistic and source camera identification can be realized using this new variable. Such an attack can be formulated as a sequential likelihood ratio test [21], [22].

Noting that under H_1 hypothesis, the NCC vector $\rho_{\mathbf{ncc}}$ obtained by searching an $m \times n$ pixels size PRNU fingerprint sub-block in a $M \times N$ pixel sized block will include one matching correlation coefficient (ρ_{match}) and $L-1$ non-matching correlation coefficients ρ_{no_j} ($j = 1, ..., L-1$), where $L = (M-m+1)\cdot(N-n+1)$ is the total number of elements in $\rho_{\mathbf{ncc}}$. Although it is very optimistic, we assume that the search window always contains a correct match within the searched fingerprint sub-block. (That is, all the $M \times N$ sized blocks picked by the adversary contain an uncarved block of size $m \times n$ pixels.) Hence, the maximum of the $\rho_{\mathbf{ncc}}$ vector can be defined as

$$max(\rho_{\mathbf{ncc}}) = \delta \cdot \rho_{match} + (\delta - 1) \cdot max(\rho_{\mathbf{no}}), \tag{28}$$

where δ is the indicator variable defined as

$$\delta = \begin{cases} 1 & \text{if } max(\rho_{\mathbf{no}}) \leq \rho_{match}, \\ 0 & \text{if } max(\rho_{\mathbf{no}}) > \rho_{match}. \end{cases} \tag{29}$$

Correspondingly, the pdf of $max(\rho_{\mathbf{ncc}})$ under hypothesis H_1 can be obtained as

$$\begin{aligned} f_{max(\rho_{\mathbf{ncc}})}(x \mid H_1) = {} & Pr(\delta = 1) \cdot f_{\rho_{match}}(x) \\ & + Pr(\delta = 0) \cdot f_{max(\rho_{\mathbf{no}})}(x), \end{aligned} \tag{30}$$

where $f_{\rho_{match}}(x)$ and $f_{max(\rho_{\mathbf{no}})}(x)$ are the probability density functions of ρ_{match} and $max(\rho_{\mathbf{no}})$, respectively. From the definition of the probability density function, $f_{max(\rho_{\mathbf{no}})}(x)$ can be obtained as

$$f_{max(\rho_{\mathbf{no}})}(x) = \frac{\partial}{\partial x} F_{max(\rho_{\mathbf{no}})}(x), \tag{31}$$

$$= \frac{\partial}{\partial x} F_{\rho_{no}}(x)^{L-1}, \tag{32}$$

$$= (L-1) F_{\rho_{no}}(x)^{L-2} f_{\rho_{no}}(x), \tag{33}$$

where $F(x)$ is the cumulative distribution function and $F_{\rho_{no}}(x)$ can be computed as $1 - Q(x\sqrt{mn})$ since $\rho_{no} \sim N(0, 1/(mn))$. To obtain the bimodal distribution of $max(\rho_{\mathbf{ncc}})$, $Pr(\delta = 1)$ and $Pr(\delta = 0)$ values should also be obtained. $Pr(\delta = 0)$ can be written as

$$\begin{aligned} Pr(\delta = 0) &= Pr(max(\rho_{\mathbf{no}}) > \rho_{match}), \\ &= \int_{z=-\infty}^{\infty} Pr(max(\rho_{\mathbf{no}}) > z \mid \rho_{match} = z)\ f_{\rho_{match}}(z) dz, \end{aligned} \tag{34}$$

$$= \int_{z=-\infty}^{\infty} \{1 - [1 - Q(z\sqrt{mn})]^{L-1}\}\ f_{\rho_{match}}(z) dz, \tag{35}$$

and $Pr(\delta = 1)$ can be obtained by $Pr(\delta = 1) = 1 - Pr(\delta = 0)$. Under H_0 hypothesis, there won't be any matching correlation coefficient in the $\rho_{\mathbf{ncc}}$ vector, *i.e.*, $max(\rho_{\mathbf{ncc}}) = max(\rho_{\mathbf{no}})$. Hence, the probability density function of $max(\rho_{\mathbf{ncc}})$ under H_0 can be obtained as

$$f_{max(\rho_{\mathbf{ncc}})}(x \mid H_0) = \frac{\partial}{\partial x} F_{max(\rho_{\mathbf{no}})}(x), \tag{36}$$

$$= \frac{\partial}{\partial x} F_{\rho_{no}}(x)^{L}, \tag{37}$$

$$= L \cdot F_{\rho_{no}}(x)^{L-1} \cdot f_{\rho_{no}}(x). \tag{38}$$

The likelihood ratio test tries to determine which hypothesis, H_0 or H_1, is more likely to explain the observed correlation value $x_i = max(\rho_{\mathbf{ncc}})$ associated with the search of uncarved block i and is defined as

$$\Lambda_i(x_i) = \frac{f_{max(\rho_{\mathbf{ncc}})}(x_i \mid H_1)}{f_{max(\rho_{\mathbf{ncc}})}(x_i \mid H_0)}. \tag{39}$$

In a similar manner, sequential log-likelihood ratio test decides under which hypothesis the observed values x_i, $i = 1, ..., k$, are more likely by computing a test statistic as follows

$$log\Lambda(\mathbf{x}) = \sum_{i=1}^{k} log \frac{f_{max(\rho_{\mathbf{ncc}})}(x_i \mid H_1)}{f_{max(\rho_{\mathbf{ncc}})}(x_i \mid H_0)}. \tag{40}$$

Essentially, with this test, instead of making k hard-decisions concerning the match of each uncarved block, we map k measurements into a single statistic to improve the identification performance.

To test the effectiveness of the attack, we used the correlation coefficient models devised for the Nexus 4 and Canon EOS 1100D cameras given in Sec. III-A. As before, Canon EOS is selected to represent a rough lower bound and the Nexus 4 represents a rough upper bound for the identification accuracy. For a given k, we computed the log-likelihood ratio $log\Lambda(\mathbf{x})$ by evaluating (30) and (38) using the

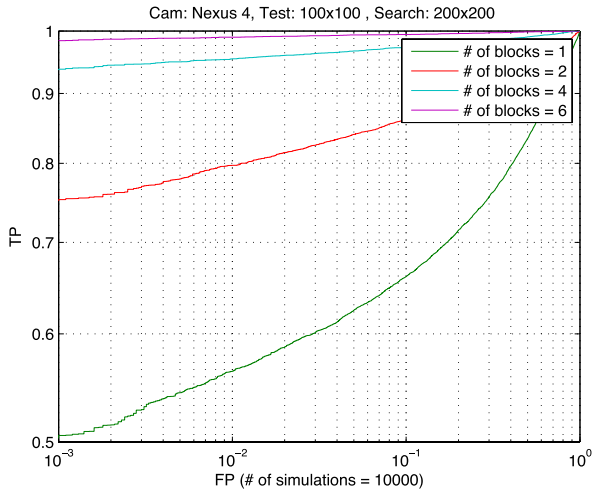


Fig. 18. The ROC curves estimated by the *model* for the sequential log-likelihood ratio test for the Nexus 4 camera when the uncarved block is 100×100 pixels and the search window is 200×200 pixels in size.

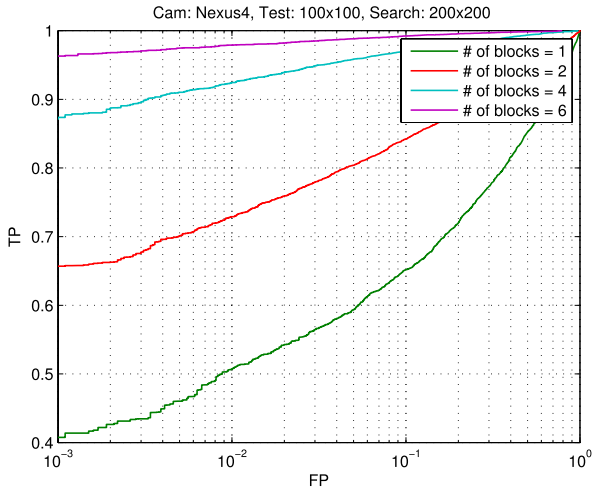


Fig. 19. The ROC curves estimated by *experiments* for the sequential log-likelihood ratio test for the Nexus 4 camera when the uncarved block is 100×100 pixels and the search window is 200×200 pixels in size.

corresponding GEV distribution parameters under matching and non-matching cases. Then, ρ_{match} and ρ_{no} values were generated randomly using their distribution models based on H_0 and H_1 hypotheses. To estimate the distribution of $log\Lambda(\mathbf{x})$, we repeated the experiment 10, 000 times where at each run we computed a single log-likelihood ratio. Finally, we estimated true positive (TP) and false positive (FP) values for various decision thresholds.

In the experiments, we first set m and n to 100 and M and N to 200, and computed the log likelihood ratios for the Nexus 4 and Canon EOS1100D cameras by simulation. We repeated the experiments for different values of k and obtained the corresponding ROC curves, which are given in Figs. 18 and 20. Figure 18 shows that the source camera of a seam-carved image can be identified with probability 0.9796 at a false positive probability of 0.001 when $k = 6$. For larger values of k, the probability of successful matching approaches to the value of 1.0, making seam-carving based anonymization completely ineffective. The ROC curves given in Fig. 20 show that for Canon EOS1100D, however, the attack cannot yield an identification probability better than 0.3 even for very high values of k, *i.e.*, for $k = 500$. Next, we set m and n to 50 and M and N to 150, and repeated the same experiment. Corresponding ROC curves are given in Fig. 22. The ROC curves show that the Nexus 4 can be successfully identified only when k is in the order of 100 and Canon EOS 1100D cannot be identified even when k is selected to be 500 (Fig. 22).

For experimental evaluation, we employed the sequential likelihood attack with various number of small blocks on 20 test images taken by the two cameras. As before, the camera fingerprints used in the experiments were obtained from 100 images not used in the sequential attack. The ROC curves were computed by 10,000 random measurements on 20 test images. In the experiments, we randomly selected the search and the test windows with no restriction on block overlapping. The ROC curves given in Figs. 18-19, 20-21, and 22-23 show that the simulation results obtained from the proposed model are in good agreement with the experimental results.

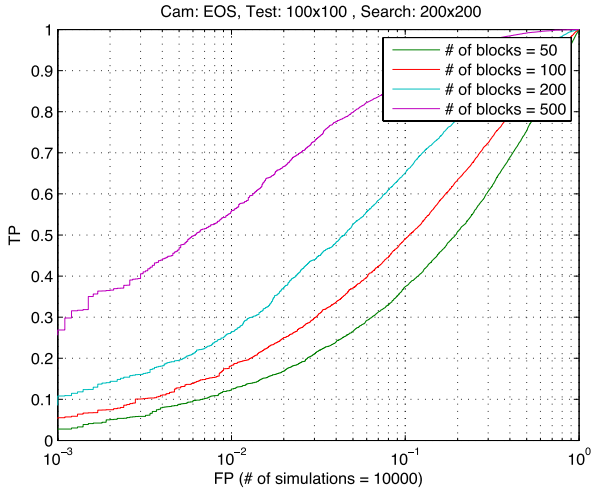


Fig. 20. The ROC curves estimated by the *model* for the sequential log-likelihood ratio for the Canon EOS1100D camera when the uncarved block is 100×100 pixels and the search window is 200×200 pixels in size.

B. Block Merging Attack

In this scenario, the adversary searches for image blocks that yield NCC values greater than the decision threshold τ and merges those blocks to obtain a larger matching region. Since the reliability of correlation statistic increases with the dimension of the data, by combining smaller matching regions together a higher identification accuracy can be achieved. That is, rather than performing individual matches between the fingerprint estimates extracted from the regions $F_\phi[i]$, $i = 1, 2, 3, \ldots$, as shown in Fig. 17, and the corresponding regions in the camera fingerprint, a larger fingerprint sub-block is obtained by combining $F_\phi[i]$ regions to form the region F_Ψ. Once F_Ψ is identified, the corresponding fingerprint estimate is compared with the original camera fingerprint by computing the correlation coefficient ρ_Ψ. If ρ_Ψ is higher than the decision

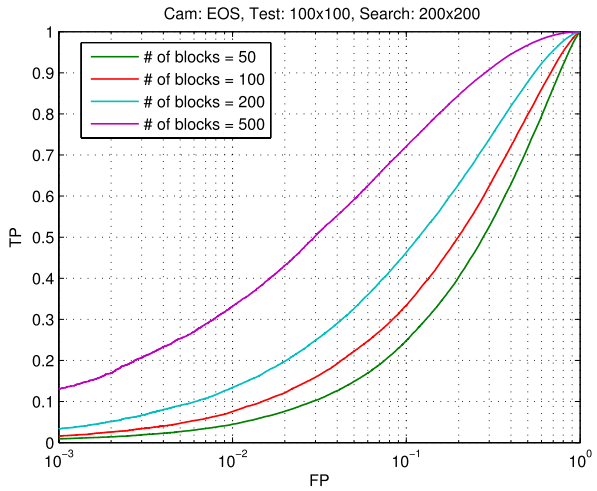


Fig. 21. The ROC curves estimated by *experiments* for the sequential log-likelihood ratio for the Canon EOS1100D camera when the uncarved block is 100×100 pixels and the search window is 200×200 pixels in size.

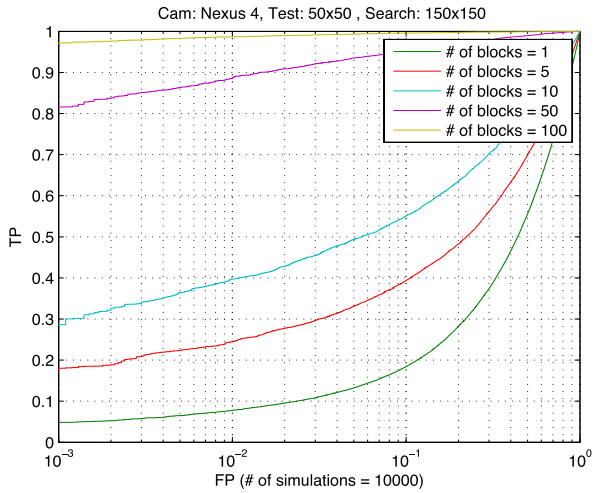


Fig. 22. The ROC curves estimated by the *model* for the sequential log-likelihood ratio for the Nexus 4 camera when the uncarved block is 50×50 pixels and the search window is 150×150 pixels in size.

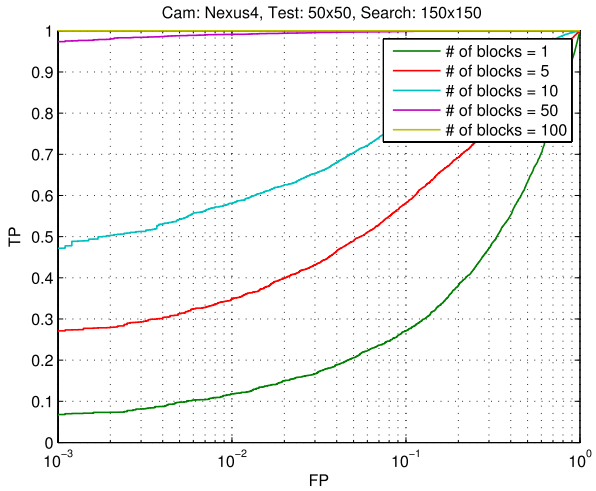


Fig. 23. The ROC curves estimated by *experiments* for the sequential log-likelihood ratio for the Nexus 4 camera when the uncarved block is 50×50 pixels and the search window is 150×150 pixels in size.

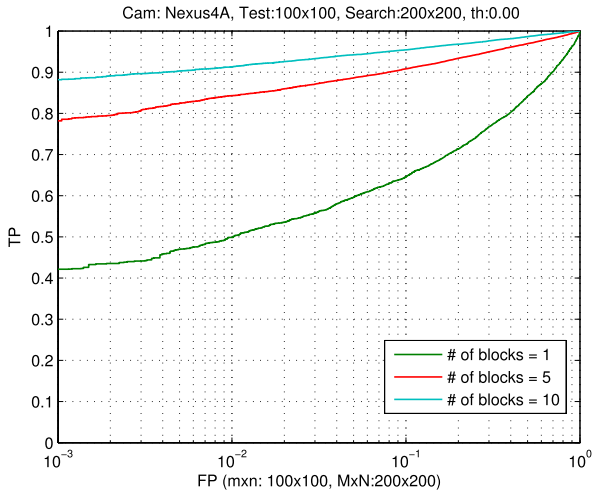


Fig. 24. *Block merging attack*: The ROC curves for the Nexus 4 camera when $\tau = 0.00$, the uncarved block size is 100×100 pixels, and the search window size is 200×200 pixels.

threshold τ_Ψ, the seam-carved image is deemed to be taken by the camera in question. It must be noted here that some of the sub-blocks that comprise Ψ will be due to false matches, and if the number of such blocks is relatively high, the probability of identification may not increase significantly with this attack.

Given a decision threshold τ, $P_{TP}(\tau)$ and $P_{FM}(\tau)$ for $m \times n$ pixel sized blocks can be obtained by evaluating (23) and (24), respectively. Let k be the total number of NCC measurements associated with different image blocks that yielded NCC values higher than the decision threshold τ. Out of these k blocks, let the number of correctly matching ones be k_{true} and the number of incorrectly identified blocks be k_{false}. From the law of large numbers, k_{true} and k_{false} can be approximately calculated as

$$k_{true} \approx k \cdot P_{TP}(\tau), \tag{41}$$

$$k_{false} \approx k \cdot P_{FM}(\tau). \tag{42}$$

The correlation coefficient ρ_Ψ is computed between the fingerprint estimate from the merged blocks $\mathbf{F}_{\Psi_\mathbf{x}}$ and the corresponding original camera fingerprint estimate $\mathbf{F}_{\Psi_\mathbf{y}}$ as

$$\rho_\Psi = corr(\mathbf{F}_{\Psi_\mathbf{x}}, \mathbf{F}_{\Psi_\mathbf{y}}). \tag{43}$$

Under H_0 hypothesis (source cameras don't match), all measured NCC values greater than the decision threshold would be false positives. Hence, the mean value and the variance of ρ_Ψ are obtained as

$$E(\rho_\Psi | H_0) = 0\ ,\ var(\rho_\Psi | H_0) = \frac{1}{kmn}. \tag{44}$$

Ultimately, the false positive rate can be estimated as a function of decision threshold τ_Ψ as

$$P_{FP}(\tau_\Psi) = Q\left(\tau_\Psi \cdot \sqrt{kmn}\right) \tag{45}$$

For experimental evaluation, we employed the block merging attack with various number of small blocks on 20 test

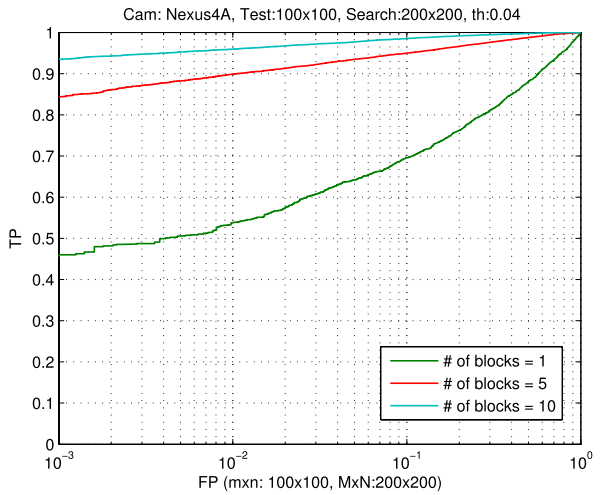


Fig. 25. *Block merging attack*: The ROC curves for the Nexus 4 camera when $tau = 0.04$, the uncarved block size is 100×100 pixels, and the search window size is 200×200 pixels.

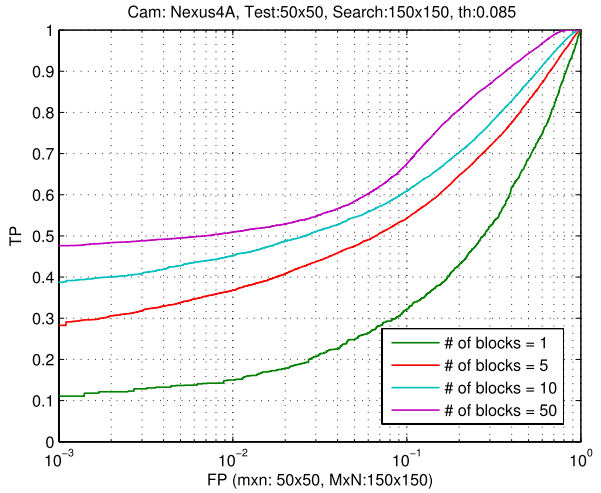


Fig. 26. *Block merging attack*: The ROC curves for the Nexus 4 camera when $\tau = 0.085$, the uncarved block size is 50×50 pixels, and the search window size is 150×150 pixels.

images taken by the Nexus 4 and Canon EOS1100D. Please note that k is the number of blocks that yield higher cross correlations than τ. Once we find k proper blocks, we merge them together to have a larger PRNU estimate and compare it with the corresponding region in the camera fingerprint. The camera fingerprints are the same with the ones used in the sequential attack experiments. The new empirical results for the block merging attack are depicted in Figures 24-28. The ROC curves were computed by 20,000 random measurements on 20 test images. In the experiments, we randomly selected the search and the test windows with no restriction on block overlapping. For Figures 24, 25, 27, and 28, the size of the test window and the search window were set to 100×100 pixels and 200×200 pixels, respectively.

Figures 24 and 25 show the corresponding ROC curves for the Nexus 4 camera. It can be seen in Fig. 25 that a true positive rate greater than 0.90 at a false positive rate of 0.001

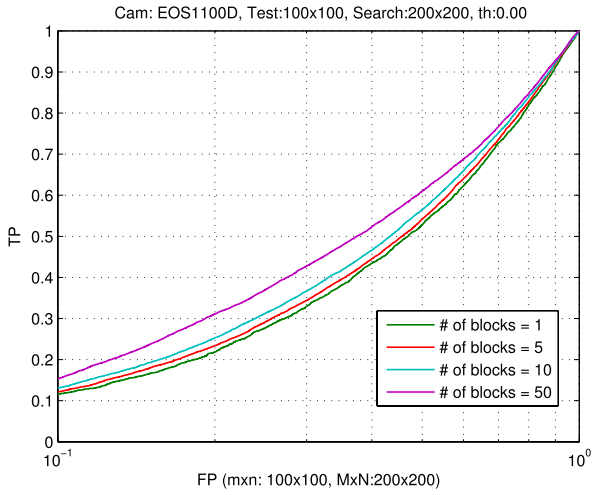


Fig. 27. *Block merging attack*: The ROC curves for the EOS1100D camera when $\tau = 0.00$, the uncarved block size is 100×100 pixels, and the search window size is 200×200 pixels.

can be achieved by merging 10 blocks of 100×100 pixels size. When the τ is set to zero, a small decrease in detection rate is observed (see Fig. 24). However, as shown in Fig. 26, if we reduce the size of the test window to 50×50 pixels and the size of the search window to 150×150 pixels, the detection rate drops drastically to 0.4 for the same false positive rate using the same number of blocks. Fig. 26 also shows that if we increase the number of blocks to 50, the true positive only gets closer to 0.5 for $P_{FP} = 0.001$. Hence, we can infer from these results that block merging attack becomes less effective when the size of test window is set to 50×50 pixels or less. Figures 27 and 28 similarly provide the results for the Canon EOS1100D camera when τ is, respectively, set to 0.00 and 0.04. In both cases, it is observed that the block merging attack has not been effective for the Canon EOS1100D camera.

C. Discussion on the Effectiveness of Attacks

Our analysis reflects the performance of the attacks on two cameras that are at the two extreme in terms of the strength of their PRNU noise patterns. DSLR cameras are known to have better noise immunity due to the large imaging sensors they deploy. This is contrary to embedded devices such as smartphones which typically deploy low-end sensors. Our results show that if many sub-blocks about the size of 100×100 pixels or larger are left uncarved, the forced seam-carving technique will not be able to provide sufficient anonymity regardless of what camera is used. We must note that our attack scenarios reflect the best possible setting for the adversary. In this regard, we assume that the search window selected by the adversary always contain an uncarved image block, and we consider relatively small sized search windows (in the order of 200×200 pixels) when performing the attack. In practice, however, the adversary will need to select larger search windows which will reduce the effectiveness of the attacks.

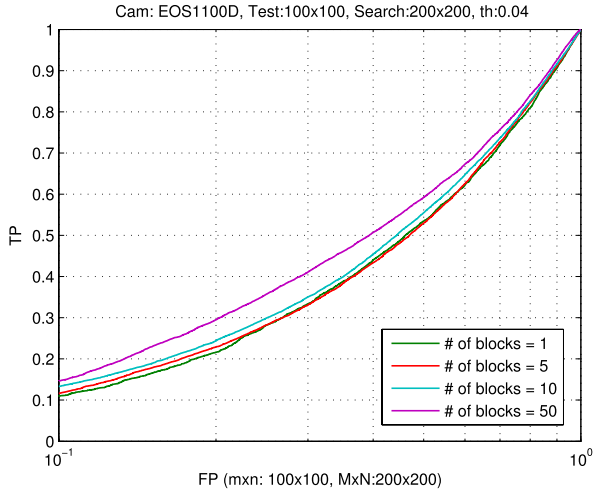


Fig. 28. *Block merging attack*: The ROC curves for the EOS1100D camera when $\tau = 0.04$, the uncarved block size is 100×100 pixels, and the search window size is 200×200 pixels.

V. Conclusion

In this study, we investigated the performance of seam-carving based source anonymization proposing two de-anonymization attacks and analysed their effectiveness. The idea behind the proposed de-anonymization attacks is based on finding uncarved image blocks and using them collectively for source camera attribution. In this regard, the first attack utilises sequential likelihood ratio test and the second one is based on merging multiple image blocks to obtain a larger block yielding higher matching statistic.

We showed that correlation based matching statistic can be better modeled by generalized extreme value (GEV) distribution when small blocks are used for source identification. Analytical results derived using the GEV model parameters from six different cameras, as well as the simulation results, show that effectiveness of forced seam carving as an anonymization technique highly depends on the strength of the PRNU noise pattern associated with a camera. Our results show that the effectiveness of the attacks increase with the size and number of uncarved image blocks and decrease with the size of the search window. To conclude, our analysis show that seam-carving based anonymization has to be customised for each camera by adjusting the maximum size of the blocks that can be left uncarved with respect to the strength of the PRNU noise pattern of the given camera. Incorporating forced seam-carving with geometrical transformations like barrel distortion, affine transforms, or image resizing prior to application of seam carving will help ensure strong anonymity against attacks. In such a case, to identify the source camera, the adversary should consider determining the parameters of resizing [12] and optical correction [23] operations and then apply the attack for each setting, which would require excessive amount of time and power. For stronger source anonymity, forced seam-carving can also be applied after performing adaptive PRNU denoising [11] which cannot be reversed by parametric search.

Acknowledgment

The authors would like to thank Ahmet Karaküçük for his help on experimental analysis.

References

[1] H. T. Sencar and N. Memon, *Digital Image Forensics: There is More to a Picture Than Meets the Eye*. New York, NY, USA: Springer-Verlag, 2013.

[2] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.

[3] A. E. Dirik, H. T. Sencar, and N. Memon, "Digital single lens reflex camera identification from traces of sensor dust," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 3, pp. 539–552, Sep. 2008.

[4] M. Chen, J. Fridrich, M. Goljan, and J. Lukás, "Determining image origin and integrity using sensor noise," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 74–90, Mar. 2008.

[5] A. E. Dirik, H. Sencar, and N. Memon, "Flatbed scanner identification based on dust and scratches over scanner platen," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2009, pp. 1385–1388.

[6] S. Nagaraja, P. Schaffer, and D. Aouada, "Who clicks there: Anonymising the photographer in a camera saturated society," in *Proc. 10th Annu. ACM Workshop Privacy Electron. Soc.*, 2011, pp. 13–22.

[7] M. Goljan, "Digital camera identification from images—Estimating false acceptance probability," in *Digital Watermarking* (Lecture Notes in Computer Science), vol. 5450, H.-J. Kim, S. Katzenbeisser, and A. Ho, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 454–468.

[8] K. Rosenfeld, H. T. Sencar, and N. Memon, "A study of the robustness of PRNU-based camera identification," *Proc. SPIE*, vol. 7254, p. 72540, Jan. 2009.

[9] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, "Can we trust digital image forensics?" in *Proc. ACM 15th Int. Conf. Multimedia (MULTIMEDIA)*, Sep. 2007, pp. 78–86.

[10] R. Böhme and M. Kirchner, "Counter-forensics: Attacking image forensics," in *Digital Image Forensics*, H. T. Sencar and N. Memon, Eds. New York, NY, USA: Springer-Verlag, 2013, pp. 327–366.

[11] A. E. Dirik and A. Karaküçük, "Forensic use of photo response nonuniformity of imaging sensors and a counter method," *Opt. Exp.*, vol. 22, no. 1, pp. 470–482, 2014.

[12] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images," *Proc. SPIE*, vol. 6819, pp. 68190E1–68190E13, Feb. 2008.

[13] M. Goljan and J. Fridrich,, "Sensor fingerprint digests for fast camera identification from geometrically distorted images," *Proc. SPIE*, vol. 8665, p. 86650B, Mar. 2013.

[14] S. Bayram, H. T. Sencar, and N. D. Memon, "Seam-carving based anonymization against image & video source attribution," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep./Oct. 2013, pp. 272–277.

[15] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. ID 10.

[16] W. van Houten and Z. Geradts, "Using anisotropic diffusion for efficient extraction of sensor noise in camera identification," *J. Forensic Sci.*, vol. 57, no. 2, pp. 521–527, 2012.

[17] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," *Proc. SPIE*, vol. 7254, p. 72540I, Feb. 2009.

[18] S. Kotz and S. Nadarajah, *Extreme Value Distributions*, vol. 31. Singapore: World Scientific, 2000,

[19] I. Alves and C. Neves, "Extreme value distributions," in *International Encyclopedia of Statistical Science*, M. Lovric, Ed. Berlin, Germany: Springer-Verlag, 2011, pp. 493–496.

[20] J. Lewis, "Fast normalized cross-correlation," *Vis. Inter.*, vol. 10, no. 1, pp. 120–123, 1995.

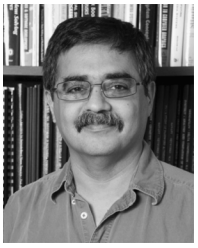
[21] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, 1945.

[22] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York, NY, USA: Springer-Verlag, 1985.

[23] M. Goljan and J. Fridrich, "Sensor-fingerprint based identification of images corrected for lens distortion," *Proc. SPIE*, vol. 8303, p. 83030H, Feb. 2012.

Ahmet Emir Dirik received the B.Sc. and M.Sc. degrees in electronics engineering from Uludağ University, Bursa, Turkey, and the Ph.D. degree in electrical engineering from the Polytechnic Institute of New York University, Brooklyn, NY, USA, in 2010. He is currently an Assistant Professor with the Department of Electrical and Electronic Engineering, Uludağ University. His research interests include multimedia forensics, security, and image processing.

Hüsrev Taha Sencar received the B.Sc. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, the M.Sc. degree in electrical and electronics engineering from Başkent University, Ankara, and the Ph.D. degree in electrical engineering from the New Jersey Institute of Technology, Newark, in 2004. He is currently an Assistant Professor with the Department of Computer Engineering, TOBB University of Economics and Technology, Ankara. His research interests include digital forensics and security problems in multimedia applications.



Nasir Memon received the B.Eng. degree in chemical engineering and the M.Sc. degree in mathematics from the Birla Institute of Technology and Science, Pilani, India, and the M.Sc. and Ph.D. degrees in computer science from the University of Nebraska, Lincoln, NE, USA. He is currently a Professor with the Department of Computer Science and Engineering and the Director of the Information Systems and Internet Security Laboratory with the Polytechnic Institute New York University, Brooklyn, NY, USA. His research interests include digital forensics, data compression, and multimedia computing and security.