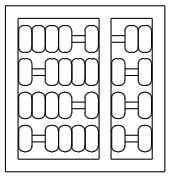


INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

OPEN-SET CAMERA BALLISTICS: MATCHING AN IMAGE TO A CAMERA WITH LITTLE KNOWLEDGE OF THE UNKNOWN

F. O. Costa *M. Eckmann* *A. Rocha*

Technical Report - IC-12-11 - Relatório Técnico

April - 2012 - Abril

OPEN-SET CAMERA BALLISTICS: MATCHING AN IMAGE TO A CAMERA WITH LITTLE KNOWLEDGE OF THE UNKNOWN

Filipe O. Costa* Michael Eckmann† Anderson Rocha‡

Abstract

Similar to ballistic tests in which we match a gun to its bullets, we can identify a given digital camera that acquired an image under investigation. In this paper, we introduce a method for identifying whether or not an image was captured by a specific digital camera. The method relies on well-known and established noise residual features related to the images under investigation. The novelty of our approach is in the extension of such features considering an "open set" recognition scenario, under which we can not rely on the assumption of full access to all of the potential source cameras. In this case, we model the decision space to take advantage of a few known cameras and carve the decision boundaries to decrease false matches increasing the reliability of image source attribution as an aid for digital forensics in the court of law.

1 Introduction

As a way to represent a unique moment in space-time, digital images are often taken as silent witnesses in the court of law and are a crucial piece of crime evidence (e.g., in child pornography, movie piracy cases, or insurance claims). Verifying a digital image's integrity and authenticity is an important task in forensics especially considering that the images can be digitally modified easily [1].

The authenticity of an image under investigation can be enforced by identifying its source. In the same manner bullet scratches allow forensic examiners to match a bullet to a particular gun with reliability high enough to be accepted in courts, source attribution techniques aim at looking for "scratches" left in an image by the source camera. These marks can be caused by factory defects, interaction between device components and the light, and others [2].

Currently, our community has put some effort into the identification of image sources generated by a scanner, printer, or camera. A simple way to identify an image's source is by its EXIF header, which contains textual information about the digital camera type and the

*Instituto de Computação, Universidade Estadual de Campinas, 13081-970 Campinas, SP.

†Skidmore College, Saratoga Springs, NY, USA

‡Instituto de Computação, Universidade Estadual de Campinas, 13081-970 Campinas, SP. Pesquisa desenvolvida com suporte financeiro da FAPESP, processo 2011/03808-3

conditions under which the image was taken (exposure, date and time, etc.). In the case of JPEG encoded images, additional information about the source can be gathered from the quantization table in the JPEG header. However, we cannot rely on such headers because their information can be easily destroyed or replaced [1].

Ruling out the EXIF headers, the problem of digital image source attribution may still be approached in other ways. Some approaches aim at identifying the brand/model of the source camera. For this, approaches generally analyze color interpolation algorithms. Nevertheless, many camera brand/models use components by only a few factories, and the color interpolation algorithm is the same (or similar) among different models of the same brand. More details can be found in [2][1].

Most source attribution approaches aim at identifying the specific camera, not just the make and model that generated an image. Some work focuses on device defects [3, 4]. However, some cameras do not contain any defects and other cameras eliminate defective pixels by post-processing their images on-board.

Methods based on sensor pattern noise (SPN) have drawn positive attention from the community due to the fact that they can identify not only camera models of the same make, but also individual instances of the same model. The deterministic component of SPN is caused by many factors such as imperfections during the sensor manufacturing process, different sensitivity of pixels with respect to light due to the inhomogeneity of silicon wafers, variable sensitivity of each sensel to light, and the uniqueness of manufacturing imperfections that even sensors of the same model would possess. These factors make SPN a robust fingerprint for identifying and linking source cameras and verifying the integrity of images [1, 2, 5].

Although previous approaches have been effective for image source attribution, they were investigated in a closed-set scenario, with the assumption that an image under investigation was generated by one of n known cameras available during training. Unfortunately, we cannot always be sure that an image was generated by one of the cameras under suspicion. Hence, it is important to model the source camera attribution problem in an open-set scenario, in which we only have access to a limited set of suspect cameras. An open-set scenario mimics a realistic situation better than a closed-set one. We need a classification model according to the few available classes while trying to take the unknown variables into consideration.

In this paper, we propose a technique to attribute an image to its specific source by SPN in an open-set scenario, where we have access to a limited set of cameras for training, and an image can be generated by any camera, including cameras to which we never had access. This work is an extension of Lukas et al.'s approach [5] based on SPN source camera attribution.

This is a first step to robust source camera attribution approaches, analyzing images with different resolutions and acquisition conditions with high precision results through machine learning techniques.

2 SPN CAMERA ATTRIBUTION

Lukas et al. [5] have proposed an approach to identify the specific source of one image using SPN estimates. The authors formulate the problem as a detection of the camera sensor pattern noise (SPN).

For each image I_i contained in a set of $\mathbf{K}$ images, we calculate the residual noise R_i using a filter F based on the Discrete Wavelet Transform (DWT) [6].

$$R_i = I_i - F(I_i), \text{where } I_i \in \mathbf{K}. \tag{1}$$

We then calculate the reference pattern P_c of SPN as the average of residual noise of the set. The residual noise is used in this step to reduce the influence of scene details

$$P_c = \frac{1}{K}\sum_{i=1}^{K} R_i. \tag{2}$$

Finally, we calculate the correlation ρ_c between the residual noise R of one image J under investigation and the SPN P_c

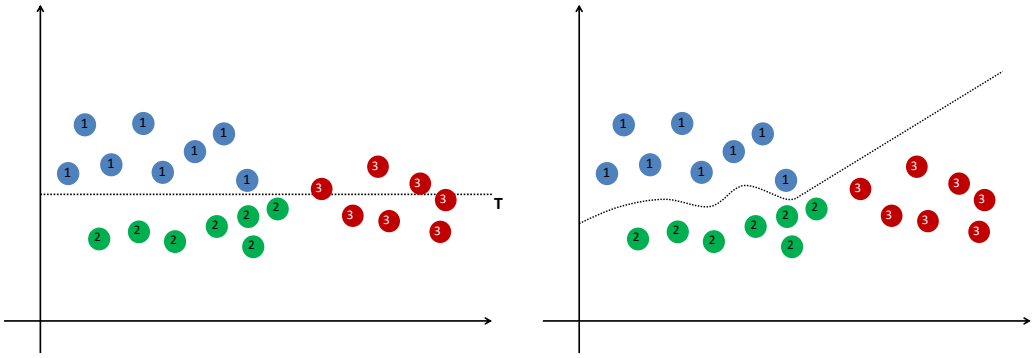
$$\rho_c(J) = corr(R, P_c) = \frac{(R - \bar{R}).(P_c - \bar{P}_c)}{||R - \bar{R}||.||P_c - \bar{P}_c||}, \tag{3}$$

where the bar above the symbol denotes a mean value. A threshold T is calculated using the Neyman-Pearson approach to minimize the false rejection rate (FRR) while imposing a bound on the false acceptance rate (FAR). If the value of this correlation is higher than T, the authors consider that the suspect image was generated by the camera under investigation. High accuracy rates were reported in [5] while testing with nine cameras, in a closed-set scenario.

3 OPEN-SET SOURCE CAMERA ATTRIBUTION

Although the approach presented in [5] is effective for source camera attribution, it is important to note that, for estimating the threshold T, the authors assumed they had examples from all cameras, and have subsequently labelled the entire space in binary fashion as either positive (generated by the camera under investigation) or negative (otherwise). Considering that T is linear, this approach may not be so effective if we need to analyze images generated by an unknown camera at training time.

The closed-set scenario is when we assume that the camera that generated the image under investigation is among the set of cameras available during training. The open-set approach is one in which we do not assume that the image under investigation was generated by an available camera. Some available cameras are considered, but not all images come from these cameras, thereby optimizing the solution for the unknowns as well as the knowns. The open-set approach is closer to a real scenario than a closed-set approach. To solve the problem of linearity of T, we believe that machine learning techniques are suited better to calculate a hyperplane to separate the positive and negative classes, according to Figure 3.



(a) (b)

Figure 1: Lukas et al. [5] approach (a) and ours (b) in an open-set scenario, considering classes one (blue) and two (green) as known cameras during training and class three (red) as an unknown camera at training time.

In this approach, for source camera attribution in an open-set scenario, we take, for each image, nine regions of interest (ROI) of 512×512 pixels of size according to Figure 2. For ROIs 1-5 (in the center), we are assuming that these regions coincide with the principal axis of the lens and should have more scene details because amateur photographers usually focus the object of interest in the center of the lens. These regions tend to have more scene details and, consequently, may have more noise information. The ROIs 6-9 (in the periphery of the picture) are also important because some cameras have effects caused by vignetting, that is a radial falloff of intensity from the center of the image, causing a reduction of an image's brightness or saturation at the periphery [7].

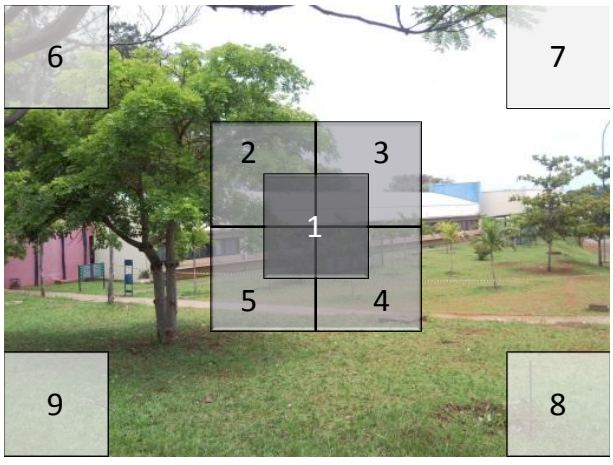


Figure 2: Regions of interest of 512×512 pixels each.

For each region shown in Figure 2, we calculate the noise pattern as discussed in [5]. Lukas et al. calculate the noise pattern considering images in gray-scale, but this can be trivially expanded to other color spaces. In this article, we calculated the SPN considering the channels R (red), G (green) and B (blue). We also calculated the SPN considering the Y channel (luminance, from YCbCr color space) which is a combination of R, G and B channels (as a gray-scale version of the image) [8]. With this, we end up with 36 reference noise patterns to represent one camera, where, for each region, we calculated one SPN for each color channel, as shown in Figure 3.

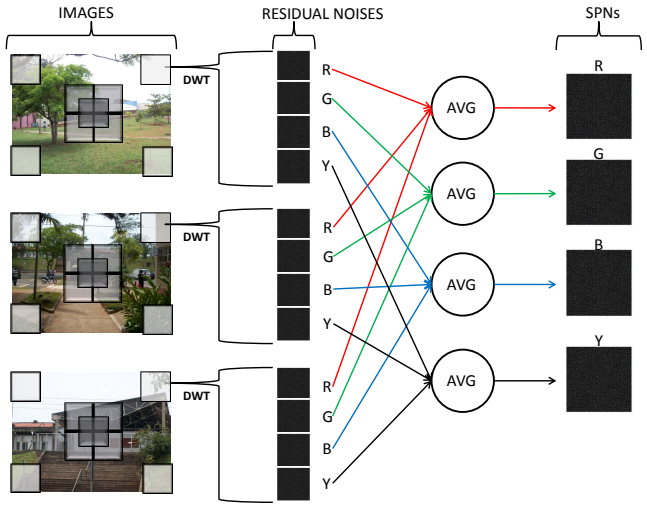


Figure 3: Calculating SPN for one region, considering R, G, B and Y color channels. For each ROI, we extract the noise residuals using a filter based on DWT, generating one noise residual for each channel. Then, we calculate the average between noises of the same color channel from many images, generating the reference noise pattern for each color channel that represents the camera under investigation. The process is performed for all nine marked regions.

It is important to note that this type of region characterization allows us to compare images with different resolutions without color interpolation artifacts, and it is not necessary to do zero-padding, for instance, when comparing images of different resolutions.

For each image, we calculate its noise and form a feature vector considering the correlation between each ROI of an image and the corresponding noise pattern for each camera. With these correlation values we have 36 features for each image, considering one camera, labelling images taken by the camera under investigation as the positive class and the remaining available cameras as the negative class. Observe that some of these images will be considered as our unknown negative class.

To accomplish this, we use a Two-Class Support Vector Machine classifier [9] for camera source attribution. First, we perform the training considering all positive and the available negative samples. For this, we use an RBF kernel and grid search during training to choose the best classification model.

After the model calculation, we analyze the classification values of all images. Let S_P and S_N be the samples of positive and negative classes respectively, used in the training step and M_P and M_N be the number of misclassified samples of positive and negative classes respectively, considering the best model of classification generated in the training step. After training, we move the generated decision hyperplane to best separate between these two classes, optimizing the decision boundaries for the available cameras while trying to account for the unknown by minimizing the training error ERR_{tr} calculated by

$$ERR_{tr} = \frac{\left(\frac{M_P}{S_P} + \frac{M_N}{S_N}\right)}{2}. \tag{4}$$

We call this process Decision Boundary Carving (DBC). The use of DBC increases the specificity and sensitivity of the chosen model. In this work, we aim at increasing the best relative accuracy results.

For testing, we calculate the accuracy of considering the best separation hyperplane obtained in the training step after DBC. The testing set contains samples of positive and negative classes (with non-overlapping samples with the training set). The testing set also has samples of images for which the source cameras were not available during training (we consider that their sources are unknown and they were not available as suspect cameras during investigation).

4 EXPERIMENTS AND RESULTS

For validation, we created a dataset with 25 digital cameras. Table 1 shows the cameras' details. For each camera, we generated 150 images with different configurations of light (indoor and outdoor), zoom and flash[1]. All images were taken in native resolution and JPEG quality compression. These images were randomly separated into five sets to perform a 5-fold cross-validation [10]. For each run, we consider three of these sets to generate the camera sensor pattern noise, one for the SVM training (considering only images for the available cameras for training) and the last one for testing (considering images from all cameras). The process is repeated five times, changing the sets.

We use the LibSVM library [11] for SVM classification. Because of the unbalanced number of samples between positive and negative classes in each test, the accuracy for each camera is calculated considering a relative classification accuracy

$$Acc_R = \frac{Acc_P + Acc_N}{2}, \tag{5}$$

where Acc_P and Acc_N are the accuracy for positive and negative classes, respectively. Note that the unknown cameras should be classified as negative class, as we do not have access to them. The average accuracy Acc_M for each camera is calculated as

$$Acc_M = \frac{1}{K}\sum_{i=1}^{K} Acc_R^i, \text{where } K = 5. \tag{6}$$

[1]The dataset will be freely available at `http://www.ic.unicamp.br/~rocha/pub/communications.html`

Table 1: Cameras used in experiments.

	Camera	Native Resolution
1	Canon PowerShot SX1-LS	3840×2160
2	Kodak EasyShare c743	3072×2304
3	Sony Cybershot DSC-H55	4320×3240
4	Sony Cybershot DSC-S730	2592×1944
5	Sony Cybershot DSC-W50	2816×2112
6	Sony Cybershot DSC-W125	3072×2304
7	Samsung Omnia	2560×1920
8	Apple iPhone 4 (1)	2592×1936
9	Kodak EasyShare M340	3664×2748
10	Sony Cybershot DSC-H20	3648×2736
11	HP PhotoSmart R727	2048×2144
12	Canon EOS 50d	4752×3168
13	Kodak EasyShare Z981	4288×3216
14	Nikon D40	3008×2000
15	Olympus SP570UZ	3968×2976
16	Panasonic Lumix DMC-FZ35	4000×3000
17	Sony Alpha DSLRA 500L	4272×2848
18	Olympus Camedia D395	2048×1536
19	Sony Cybershot DSC-W120	3072×2304
20	Nikon Collpix S8100	4000×3000
21	Sony Cybershot DSC-W330	4320×3240
22	Apple iPhone 4(2)	2592×1936
23	Cannon Powershot A520	1600×1200
24	Apple iPhone 3	1600×1200
25	Samsung Star	2048×1536

The results we report correspond to the final accuracy Acc_F, calculated as the average over all cameras

$$Acc_F = \frac{1}{N} \sum_{i=1}^{N} Acc_M^i, \tag{7}$$

where N is the number of available cameras during training.

We analyze the open-set image source attribution considering that we have access to 15, 10, 5 and 2 suspect cameras, but the images can be generated by any of the 25 cameras shown in Table 1. In these scenarios, we consider that we never have access to cameras 16–25. In the first case, we consider that we have access to cameras 1–15 which means we train with cameras 1–15 as suspect cameras but the images under investigation can come from any of the 25 cameras of Table 1. Two experiments with 10 cameras were performed (cameras 1–10 and cameras 6–15). The experiments with five cameras were performed considering three different combinations of five cameras (1–5, 6–10, 11–15). The experiments with two available cameras were performed with seven different combinations (1–2, 3–4, and so forth). The result, for each case, is the average of the results for tests considering each combination of cameras. Table 2 compares the proposed methods to Lukas et al.'s approach [5] in an open-set scenario.

Table 3 shows results for the experiments considering we have two available cameras with the same brand/model (iPhone 4; cameras 8 and 22), but an image can be generated

Table 2: Results ($ACC_F \pm$ stdev, in (%)), and reduction in error (RE) for 15, 10, 5, and 2 available cameras during training. For example, an open set with 15/25 cameras consists of training on 15 cameras but testing on images that can come from any of the 15 cameras as well as 10 other unknown cameras (450 + 300 test images per round).

	Open-set Cameras			
	15	**10**	**5**	**2**
LUKAS ET AL. [5]	89.97 ± 2.04	91.63 ± 2.42	88.81 ± 4.44	89.43 ± 4.49
OURS	**95.89** ± 1.97	**96.63** ± 1.38	**95.65** ± 1.76	**96.43** ± 2.16
OURS (DBC)	**97.77** ± 1.15	**97.20** ± 0.47	**96.42** ± 0.89	**93.94** ± 2.76
RE (%)	59.02	59, 63	61.12	66.22
RE (DBC) (%)	79.78	66.54	68.29	42.66

by any of the 25 cameras. The results show the average of five tests per camera (5-fold).

Table 3: Results considering cameras with same brand and model.

LUKAS ET AL. [5]	88.89 ± 1.56
OURS	**95.54** ± 0.72
OURS (DBC)	**95.17** ± 1.17
RE (%)	59.85
RE (DBC) (%)	56.52

Tables 2 and 3 show a statistically significant improvement in the overall performance when comparing the methods we propose and the baseline of Lukas et.al. [5]. They also show that it is possible to reliably identify image sources in an open-set scenario. The results show that the decision boundary carving DBC is not necessary when we have access to only two suspect cameras, but it can be useful when we have more suspect cameras.

Table 4 shows a breakdown of the results for the closed (C) and open-set (O) case with 15 cameras. A closed set of 15 cameras consists of training on 15 cameras and testing on images only from these 15 cameras (450 test images per round). It shows the true positives, as well as the true negatives with results in (%). Note that Lukas et al.'s method [5] behaves similarly in the open and closed set. Our proposed methods show higher performance in both the closed and open-set scenarios (up to $\cong 80\%$ reduction of false positives and false negatives).

5 CONCLUSION

In this work, we explained that solving the image source attribution problem in an open-set scenario is important because it is closer to a real environment, where an image can be taken by any unknown camera unavailable in the seized set of cameras during an investigation. This is just a first step to robust source camera attribution techniques. With our approach,

Table 4: Breakdown for the closed (C) setup with 15 cameras and and open-set (O) with 15 cameras for training and 25 for testing.

	[5] (C)	[5] (O)	OURS (C)	OURS (O)	OURS-DBC (C)	OURS-DBC (O)
TP	87.9 ± 1.79	87.9 ± 1.79	92.0 ± 1.69	92.0 ± 1.69	97.2 ± 0.84	97.2 ± 0.84
TN	92.6 ± 1.14	92.0 ± 0.99	99.8 ± 0.04	99.7 ± 0.1	99.8 ± 0.05	98.3 ± 0.24

it is possible to analyze images with different resolutions. Furthermore, we can identify source cameras considering complementary characterization methods taking advantage of all the potential of machine learning classification techniques.

Expanding on the work of Lukas et al. [5], our experiments report high accuracy results. The next step of this work is tuning the classification model for one class classification, where we train the classifier with a given class of interest only. This can be useful in an open-set scenario, when we have access to only one camera.

Furthermore, this work can be improved to help the combat against counter forensic approaches, as presented in [12]. Possible future work is to analyze some counter forensic techniques to this work, considering the techniques and features proposed.

Finally, we believe that efforts like the one presented by Lukas et al. [5] and further expanded upon in this paper will move source attribution approaches toward meeting the strong standards of the Daubert trilogy [13] which establishes a high bar for acceptance of forensic evidence (analog and digital) in US courts and, possibly, in other countries.

References

[1] A. Rocha, W. Scheirer, T. E. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensic," *ACM CSUR*, vol. 42, pp. 26:1–26:42, October 2011.

[2] A. Swaminathan, M. Wu, and K.J.R. Liu, "Component forensic - theory, methodologies e applications," *IEEE SPM*, vol. 26, pp. 38–48, March 2009.

[3] K. Kurosawa, K. Kuroki, and N. Saitoh, "CCD fingerprint method – identification of a video camera from videotaped images," in *IEEE ICIP*, 1999, pp. 537–540.

[4] Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh, "Methods for identification of images acquired with digital cameras," *Enabling Technologies for Law Enforcement and Security*, vol. 4232, pp. 505–512, 2001.

[5] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE TIFS*, vol. 1, no. 2, pp. 205–214, 2006.

[6] S. Lyu, *Natural Image Statistics for Digital Image Forensics*, Phd thesis, Dartmouth College, August 2005.

[7] D. B. Goldman and J. H. Chen, "Vignette and exposure calibration and compensation," in *IEEE ICCV*, 2005, pp. 899–906.

[8] X. Wang and Z. Weng, "Scene abrupt change detection," in *Canadian Conf. on Electrical and Computing Engineering*, 2000, pp. 880–883.

[9] C. Cortes and V. Vapnik, *Machine Learning*, chapter Support-Vector Networks, pp. 273–297, Kluwer Pub., 20 edition, 1995.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[11] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM TIST*, vol. 2, pp. 27:1–27:27, 2011.

[12] M. Goljan, J. Fridrich, and M. Chen, "Defending against fingerprint-copy attack in sensor-based camera identification," in *IEEE TIFS*, 2011, pp. 227–236.

[13] Donald E. Shelton, *Forensic Science in Court - Challenges in the 21st Century*, Rowman & Littlefield Publishers, 2011.