Pre-print of article that will appear at SIBGRAPI 2012.

Open Set Source Camera Attribution

Filipe de O. Costa
Institute of Computing
University of Campinas (UNICAMP)
Campinas, São Paulo, Brazil
filipe.costa@students.ic.unicamp.br

Michael Eckmann
Dept. of Mathematics and Computer Science
Skidmore College
Saratoga Springs, NY, USA
meckmann@skidmore.edu

Walter J. Scheirer
Securics Inc.
University of Colorado
Colorado Springs, CO, USA
wjs3@vast.uccs.edu

Anderson Rocha
Institute of Computing
University of Campinas (UNICAMP)
Campinas, São Paulo, Brazil
anderson.rocha@ic.unicamp.br

***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 discuss a method for identifying whether or not an image was captured by a specific digital camera. The method relies on noise residual features related to the images under investigation. Our approach considers an "open set" recognition scenario, under which we can not rely on the assumption of full access to all of the potential source cameras. This is the only scenario investigators are faced with in the real world. 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. This approach performs favorably vs. the state-of-the-art.**

***Keywords*-Digital Forensics; Open Set Recognition; Camera Attribution; Sensor Fingerprinting.**

I. 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 that 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, the forensic community has put some effort into the identification of image sources generated by a scanner [3, 4], printer [5, 6], or camera [7, 8, 9, 10]. A simple way to identify an image's source is by its EXIF header when available for a format (e.g., JPEG and TIFF), which contains textual information about the digital camera type and the 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 EXIF 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 have the objective of identifying the brand/model of the source camera [11, 12]. For this, approaches generally analyze color interpolation algorithms. Nevertheless, many camera brands/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 [1, 2].

Most source attribution approaches aim at identifying the specific camera, not just the make and model that generated an image. This generally can be done by analyzing image artifacts caused by factoring defects. Methods based on sensor pattern noise (SPN) have drawn positive attention from the forensic 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, 9].

Although previous approaches have been effective for image source attribution, many of them 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 investigation. 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 much 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.

Contributions:

- We discuss our technique to match an image to its specific source by using SPN features in an Open Set scenario, in which 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.
- We account for the unknown cameras by optimizing the decision boundary hyperplane found by a traditional

Support Vector Machine (SVM) classifier and minimizing the training data error associated with it.
- In addition, we also have a minor contribution in the feature characterization part of the problem since we extend upon Lukas et al.'s approach [9] based on SPN source camera attribution.

To our knowledge, this is a first step towards robust source camera attribution approaches, analyzing images with different resolutions and acquisition conditions with high classification results through machine learning techniques.

Finally, we organized this paper in five sections: Section II presents some related work about camera source attribution. Section III shows details about the Open Set recognition problem. Section IV presents our approach for the problem of source attribution. Section V presents the experiments and results of this work.

II. Forensic Related Work

For specific device source attribution, we aim at identifying the exact camera that produced the image in question. It can be done considering hardware and component imperfections, effects of operational conditions, environment, noise, sensor dust on lens, etc. It is important to observe that these features may be temporal by nature, and thus, not reliable in certain circumstances [1].

There is some previous work that analyzes device defects for image source identification. Kurosawa et al. [7] propose an approach to identify an image's source by using fixed pattern noise (FPN). FPN is caused by dark currents, which refers to the accumulation of electrons in each sensel of the device due to thermal action. The approach proposed in [7] is limited because not all cameras have these defects. Furthermore, this kind of noise is not robust and can be destroyed when the final image is being generated, considering modern cameras.

The approach proposed by Geradts et al. [13] aims at identifying the specific camera that generated one image by analyzing pixel defects. The authors considered hot pixels (individual pixels on the sensors with higher than normal charge leakage), cold pixels (pixels with lower than normal charge) and pixel traps (clusters of hot or cold pixels). The authors did not report quantitative results about the approach's effectiveness. The major problem of this technique is the fact that some cameras do not contain any defects and other cameras eliminate defective pixels by post-processing their images on-board.

Dirik et al. [8] proposed a method that analyzed artifacts caused by dust on the lens at the time the image was taken. The authors consider that the dust on the lens generates a pattern of artifacts that can be extracted from images. They report visual results, considering a scenario with two different cameras. However, this approach is also limited, because these artifacts are temporal by nature and can be easily destroyed (the lens may be cleaned, for instance).

Approaches based on sensor pattern noise (SPN) for image source attribution have drawn positive attention from the forensic community due to the fact that they are a robust way to identify the specific camera, including individual instances of the same model, and not just the brand/model of the device. As Fig. 1 shows, we can consider two types of noise patterns: Fixed Pattern Noise (FPN) and Photo Response Non-Uniformity Noise (PRNU). FPN is caused by dark currents, as discussed by Kurosawa et al. [7]. The PRNU is divided into low-frequency defects noise (LFD) and pixel non-uniformity noise (PNU). LFD is usually caused by light refraction on particles near the camera and zoom configurations. This kind of noise is not considered for camera attribution because of its unstable nature. PNU is caused by the interaction between the light and each sensel of the sensor array.

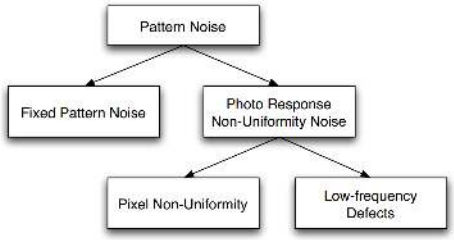


Fig. 1. Sensor pattern noise hierarchy [1].

Lukas et al. [9] have proposed an approach to identify the specific source of one image using PNU. The authors formulate the problem as a detection of the camera sensor pattern noise. The approach works as follows: for each image I_j contained in a set of images $\mathbf{K}$, calculate the residual noise R_{I_j} using a filter F based on the Discrete Wavelet Transform (DWT) [14].

$$R_{I_j} = I_j - F(I_j) \tag{1}$$

Then, they calculate the reference pattern P_c of sensor pattern noise as the average of residual noise of the set. The residual noise is used in this step to reduce the influence of scene details.

$$P_c = \frac{1}{k}\sum_{i=1}^{k} R_{I_i}, \text{ where } k = |\mathbf{K}|. \tag{2}$$

Finally, they calculate the correlation value ρ_c between the residual noise R_J of one image J under investigation and the SPN P_c of a set of images of a given camera.

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

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 [9] while testing with nine cameras, and the results are confirmed in [15, 16]. Fig. 2 depicts a representation of Lukas et al.'s [9] approach.

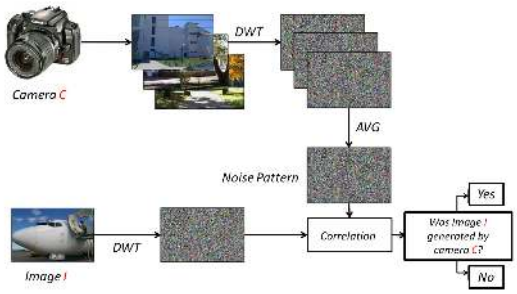
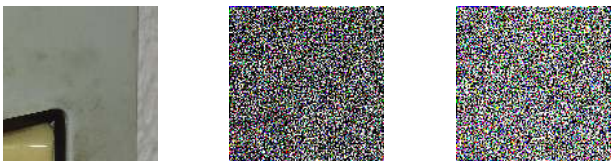


Fig. 2. Lukas et al. [9] approach.

Li [10] proposed an enhancement for the method of Lukas et. al. [9]. The author examined the influence of scene details in the reference pattern noise. According to the author, the high frequencies (e.g., object edges) existing in an image can contaminate its PRNU component, and lead to unsatisfactory camera identification results through sensor pattern noise. The author proposed a sensor pattern noise enhancement method to reduce the influence of the scene content in the noise component. Considering one image $I_p \in \mathbf{I}$, after extracting its noise $n = R_{I_p}$ according to Eq. 1, the authors applied a normalization in each pixel $n(x, y)$, generating the enhanced noise $n_e(x, y)$. The model which yielded the best results is defined by

$$n_e(x,y) = \begin{cases} e^{-0.5n^2(x,y)/\alpha^2}, & \text{if } 0 \leq n^2(x,y); \\ -e^{-0.5n^2(x,y)/\alpha^2}, & \text{otherwise;} \end{cases} \quad (4)$$

where α is defined by the user. The best value reported in that work for this normalization is $\alpha = 7$. Fig. 3 shows the original image (a), its sensor pattern noise (b) and its enhanced sensor pattern noise (c). The author reports accuracy of 94% in a scenario with six cameras, considering a center 512×512 region of the image.



(a) (b) (c)

Fig. 3. (a) An input image. (b) Its noise residual calculated as in [9]. (c) Noise residual enhancement calculated as in Li's work [10].

The approach presented by Lukas et al. [9] and its enhancement proposed by Li [10] allowed the development of other approaches based on their concepts, as approaches that aim at identifying the common source of image pairs [17] or clustering of image sets [18, 19]. Considering source camera identification by sensor noise, there are some approaches whose objective is to discover inconsistencies in camera identification methods and explore how these inconsistencies can make the source camera identification task difficult [20, 21]. These approaches are called counter-forensic techniques, and are also important in a forensic research field because they can help improve the resilience of existing forensic methods. However, we do not consider the existence of counter-forensic techniques in the present work.

Although the approaches presented in [9] and [10] are effective for source camera attribution, it is important to note that, for estimating the threshold T, the authors assumed they had examples from all the cameras, and have subsequently labelled the entire space in a 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. When we do not have access to all cameras in an investigation, we believe (and give evidence supporting our belief) that machine learning techniques are better suited to calculate a hyperplane to separate the positive and negative classes in such a scenario, and that is the main subject of this paper.

III. Open Set Classification and Related Work

A Closed Set scenario assumes that the camera that generated the image under investigation is among the set of cameras available during training. The Open Set approach, on the other hand, does 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 known. The important difference is that all positive examples are similar, but each negative example has its own particularities [22]. Fig. 4 depicts an example of Open Set classification.

In machine learning, most of the time we do not need, do not have access to, or do not know all possible classes to consider. For instance, when classifying whether or not an image contains a hidden message [1] we might have training examples of only pristine images (with no hidden messages) and perhaps some images of only one or two algorithms for hiding messages. A robust classifier must consider all other possible types algorithms for hiding messages as relevant features as a negative class. In many cases, to model this negative class is non-viable or impossible (for instance, considering all existent algorithms for hiding messages).

Open Set recognition has received only limited treatment in the pattern recognition literature. For instance, in a study of face recognition evaluation methods outlined by Phillips et al. [23], the authors define a threshold T where all face identifications necessarily must be higher than this value to be considered a match. However, being greater than T is not sufficient to be considered a match therefore possible unknown impostors are considered even though the system is not trained with all possible impostors.

Considering problems with source attribution, Wang et al. [24] perform the camera model identification that generated one image through Color Filter Array (CFA) coefficients estimation as is done in [12, 11], and use a combination of two classification approaches: Two-class SVMs (TC-SVM) [25]

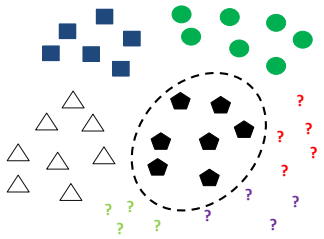


Fig. 4. Example of Open Set classification. In essence, open set recognition explicitly presumes not all classes are known *a priori*. The above diagram shows a known class of interest ("pentagon"), surrounded by other classes that are not of interest, which can be known ("triangle", "circle", "square"), or unknown ("?").

and One-class SVMs (OC-SVM) [26]. The second approach might be considered a solution for Open Set as it uses one-class classifiers. For that, the authors use only two of 17 available cameras for training (one class of interest and one for outlier definition, which can be seen as a form of accounting for the unknown) and all 17 cameras for testing. The work reported results of approximately 91% correct matches. The disadvantage of this approach is the fact that, considering CFA coefficients, we can identify the brand/model of the camera that generated one image, but we can not identify the specific device, considering that different cameras of the same brand/model probably have the same CFA components. Their solution is also different than ours given that we propose a way to automatically find a new position for the decision hyperplane based on minimizing the training data error in the case of a binary SVM classifier.

Li [18] and Caldelli et al. [19] proposed different approaches to separate a set of images into clusters according to their source devices. In both works, the authors consider that they do not have any information about the cameras that generated such images. Although they use unsupervised classification for the cluster definition, in this case, the works do not consider an Open Set recognition application because the authors used a Closed Set of cameras for validation, and one image necessarily was generated by one of the cameras (that is, they do not consider any unknown class during training).

IV. Our Approach

Our approach for source camera attribution considering an Open Set scenario works as follows:

A. Definition of Regions of Interest;
B. Feature Characterization;
C. Source Camera Attribution in an Open Set scenario.

A. Definition of Regions of Interest

Lukas et al. [9] consider a central region of the image to determine the source of an image. Li's [10] approach is also performed considering a central region and in other experiments, the whole image. However, Li [10] performed the experiments in a scenario where the author has six suspect cameras with the same native resolutions (that is, all the images used in those experiments have the same dimensions). When we have images with different sizes, to consider a common region of all images (for example, the central region) may be better for image source camera attribution.

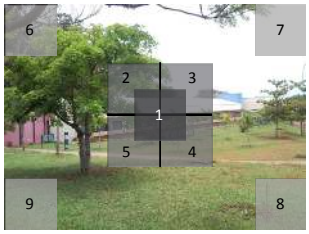


Fig. 5. Regions of interest of 512×512 pixels each.

According to [27], different regions of the image can have different information about the source camera fingerprint. In our approach for source camera attribution in an Open Set scenario, we aim at considering many regions of one image instead of using just the central region as is done in [9] [10]. We take, for each image, nine regions of interest (ROI) of 512×512 pixels according to Fig. 5. 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 [27, 28].

B. Feature Characterization

As we discussed on Section II, defining a linear threshold to separate positive and negative samples may be not so effective if we need to analyze images in an Open Set scenario, when an image can be generated by an unknown device. In our approach, we aim at performing the source camera attribution by machine learning techniques. One contribution of our approach is the definition of some features that can well represent the source camera fingerprint.

For each region shown in Fig. 5, we calculate the noise pattern as discussed in [9]. 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 as defined in Eqs. 1 and 2 considering the channels R (red), G (green) and B (blue), separately. 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) [29]. 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 Fig. 6.

It is important to note that this type of region characterization allows us to compare images with different

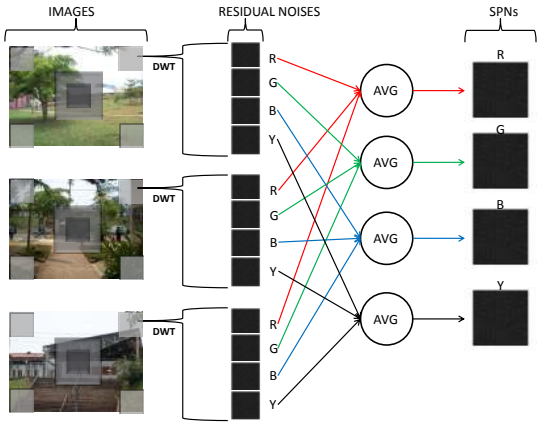


Fig. 6. Calculating SPN for one region, considering R, G, B and Y color channels. For each ROI, we extract the noise residuals using DWT-based filter, generating one noise residual for each channel. Then, we calculate the average between noises of the same color channel from many images, generating the reference noise pattern for each color channel that represents the camera under investigation. The process comprises the nine marked regions.

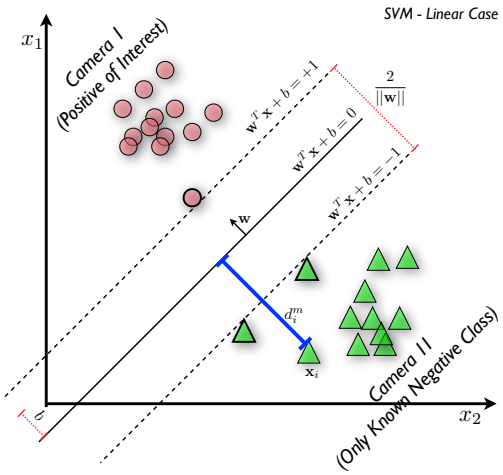


Fig. 7. Example of an SVM classifier considering a linear case.

resolutions without color interpolation artifacts, and it is not necessary to do zero-padding, for instance, when comparing images of different sizes.

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, according to Eq. 3. With these correlation values we have 36 features for each image, considering one camera, labeling images taken by the camera under investigation as the positive class and the remaining available cameras as the negative class.

C. Source Camera Attribution in an Open Set Scenario

The main contribution of this paper is the use of machine learning techniques to solve the source attribution problem in an Open Set scenario. To solve our problem, first, we find a classifier from the training set of examples considering the positive and the available negative samples. Formally, given training data $(\mathbf{x}_i, y_i)$ for $i = 1 \ldots N$, with $\mathbf{x}_i \in \Re^d$ and $y_i \in \{-1, 1\}$, we need to learn a classifier $f(\mathbf{x})$ such that

$$f(\mathbf{x}_i) = \begin{cases} \geq 0, & y_i = +1 \\ < 0, & y_i = -1. \end{cases} \tag{5}$$

Let $\mathbf{X}$ be our training data matrix in which the n^{th} row of $\mathbf{X}$ corresponds to the row vector $\mathbf{x}_i^T$. Consider that the positive training class consists of feature vectors $\mathcal{P} = \{\mathbf{x}_1^p, \mathbf{x}_2^p, \ldots \mathbf{x}_{n_{pos}}^p\}$ and the negative class(es) consists of $\mathcal{N} = \{\mathbf{x}_1^n, \mathbf{x}_2^n, \ldots \mathbf{x}_{n_{neg}}^n\}$ where $N = n_{pos} + n_{neg}$ is the total number of training examples.

We can find a maximum margin separation hyperplane $w^T\mathbf{x}+b = 0$ (linear case) or $w^T\phi(\mathbf{x})+b = 0$ (nonlinear case) by means of the classical support vector machine classification algorithm [25, 30] which aims at finding a classifier able to separate the data points from $\mathcal{P}$ and $\mathcal{N}$, where $\mathbf{w}$ is the normal to the hyperplane, b is the bias of the hyperplane such that $|b|/||\mathbf{w}||$ is the perpendicular distance from the origin to the hyperplane, and ϕ is a mapping function from original feature space to a higher dimensional space by means of the kernel trick [30].

After finding a maximum margin separation hyperplane (classifier $f(\cdot)$) from the training data points X, we have a situation as the one depicted (only for the linear case above) in Fig. 7 in which we have one class of interest as positive class (consisting of data points from one camera) and only one negative class (consisting of data points from another known camera). According to this model, each data point $\mathbf{x}_i$ during training is at a distance d_i^m to the decision boundary given the SVM model and can be classified as of class $+1$ if $w^T\mathbf{x_i} + b \geq 0$ (linear case) and as -1, otherwise.

SVM uses structural risk minimization (SRM) [30] which is an inductive principle for model selection used for learning from finite training data sets to solve the problem of finding the maximum margin separation hyperplane. However, it turns out that SVM can only minimize the risk in this case based on what it knows from the training data. In the Open Set case, many more classes can appear as being a negative class which could damage the operation of the classifier during tests.

Therefore, in this paper we define a policy of minimizing the risk for the unknown for the Open Set case by minimizing the data error $\mathcal{D}$ during training after the hyperplane is calculated by SVM. We define the data error $\mathcal{D}$ as the inverse of the normalized classification accuracy $A(\mathbf{X})$ during training

$$A(\mathbf{X}) = \frac{\left(\frac{\sum_{i=1}^{n_{pos}} \theta(\mathbf{p_i})}{n_{pos}} + \frac{\sum_{j=1}^{n_{neg}} \omega(\mathbf{n_j})}{n_{neg}}\right)}{2} \tag{6}$$

where

$$\theta(\mathbf{p_n}) = \begin{cases} 1, & \text{if } f(\mathbf{p_i}) \geq \mathbf{0} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

$$\omega(\mathbf{k_l}) = \begin{cases} 1, & \text{if } f(\mathbf{k_j}) < \mathbf{0} \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Equation 6 means we analyze the classification values of all training samples to find its classification accuracy $A(\mathbf{X})$.

Considering the calculated hyperplane in the training step, we propose to account for unknown classes by moving the decision hyperplane by a value ε inwards the positive class or outwards in the direction of the negative known class(es). The rationale is that by moving the plane we can be more strict to what we know as positive examples and therefore classify any other "too different" data point as negative or we can be less strict about what we know with respect to the positive class and accept more distant data points as possible positive ones. As a first step towards solving the camera attribution problem in an Open Set scenario, we consider ε to move in the interval given by the most positive example (farthest from the decision hyperplane) and the most negative example (farthest from the decision hyperplane). For simplification, we might constrain the interval, as we do in this paper, to be tighter such as $\in [-1, 1]$ to do not drastically change the initial hyperplane found by SVM.

The ε value represents a movement on the decision hyperplane $w^T\mathbf{x}+b+\varepsilon = 0$ (linear case) or $w^T\phi(\mathbf{x})+b+\varepsilon = 0$ (nonlinear case). Fig. 8 depicts an example for a nonlinear case.

In this paper, we loosely call this process as Decision Boundary Carving (DBC). The value of ε is defined by an exhaustive search to minimize the training data error, which we accomplish by minimizing $\frac{1}{A(\mathbf{X})}$ and altering Equations 7 and 8 to:

$$\theta(\mathbf{p_n}) = \begin{cases} 1, & \text{if } f(\mathbf{p_i}) \geq \varepsilon \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

$$\omega(\mathbf{k_l}) = \begin{cases} 1, & \text{if } f(\mathbf{k_j}) < \varepsilon \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Given any data point $\mathbf{z}$ during testing, it is classified as a positive example if $f(\mathbf{z}) > \varepsilon$.

V. Experiments and Results

To validate the ideas discussed in this paper, we created a dataset with 25 digital cameras. Table I 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 [30]. 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 [31] for SVM classification and only consider the nonlinear case here with a radial basis function kernel (RBF).

[1]The dataset will be freely available upon acceptance.

TABLE I
Cameras used in the 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 Coolpix S8100	4000×3000
21	Sony Cybershot DSC-W330	4320×3240
22	Apple iPhone 4(2)	2592×1936
23	Canon Powershot A520	1600×1200
24	Apple iPhone 3	1600×1200
25	Samsung Star	2048×1536

After calculating the relative accuracy for each camera according to

$$A_R = \frac{A_P + A_N}{2}, \tag{11}$$

which is the number of correct classifications during testing for positive (A_P) and negative (A_N) data points for a given camera, the average accuracy A_M for each camera is calculated as

$$A_M = \frac{1}{z}\sum_{i=1}^{z} A_R^i, \tag{12}$$

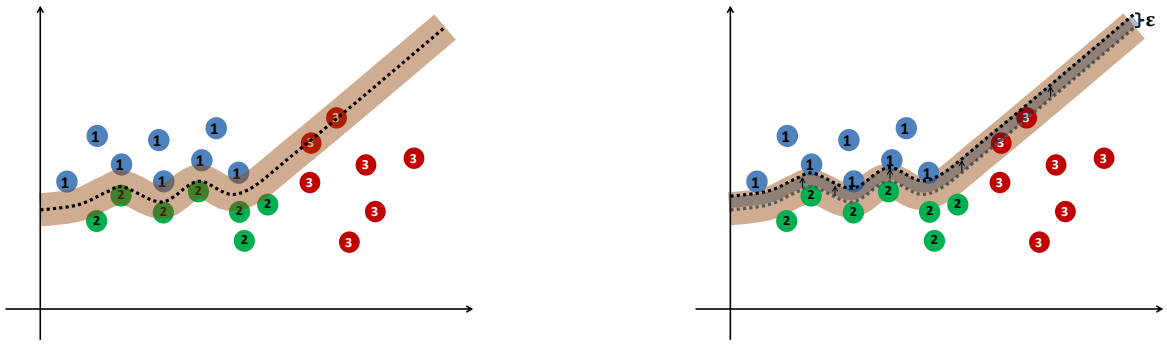
where $z = 5$ runs of the cross-validation.

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

$$A_F = \frac{1}{N_C}\sum_{i=1}^{N_C} Acc_M^i, \tag{13}$$

where N_C 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 except during testing. 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).



(a) (b)

Fig. 8. Our Open Set implementation for source camera attribution using Decision Boundary Carving (DBC). (a) shows the calculated separation hyperplane, considering the blue and green data points as the known positive (1) and negative (2) classes, respectively, and the red data points represent the unknown classes (3). The orange region represents the distance between the margins of the positive and negative support vectors. (b) shows the DBC over the calculated hyperplane, represented by the blue region. Note that the process of carving the decision boundary seeks the minimization of the risk of the unknown via minimizing the data error $\mathcal{D}$ which is implemented as the misclassification during training ($1/A(x)$).

To analyze the effectiveness of our implementation for Open Set source camera attribution, we performed some experiments with and without this technique. To analyze the importance of the choice of ROIs shown in Fig. 5, we performed experiments with our approach considering just the central ROI (ROI #1) similar to existing techniques in the literature as well as experiments considering all of the ROIs.

The result, for each case, is the average of the results for tests considering each combination of cameras. Table II shows the comparison of the proposed methods to Lukas et al.'s [9] and Li's [10] approaches in an Open Set scenario. We refer to our approach considering only the ROI #1 as $T1$, with ROI #1 plus the Open Set decision boundary carving solution as $T2$, our approach considering all ROIs without DBC as $T3$ and the complete solution with all regions plus DBC as $T4$.

TABLE II
RESULTS ($A_F \pm \sigma$, IN (%)) 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 – Results in Percentage			
	15	**10**	**5**	**2**
LUKAS ET AL. [9]	94.54 ± 2.10	93.93 ± 2.09	94.45 ± 2.17	93.08 ± 2.56
LI. [10]	94.07 ± 2.31	93.49 ± 2.35	93.94 ± 2.35	92.82 ± 2.94
OURS – T1 Only central ROI	91.01 ± 3.13	91.22 ± 2.61	93.40 ± 2.58	**94.16** ± 2.66
OURS – T2 Central ROI + DBC	**95.57** ± 1.52	**94.95** ± 1.78	**95.11** ± 1.62	**94.34** ± 2.05
OURS – T3 All ROIs without DBC	**95.89** ± 1.97	**96.63** ± 1.38	**95.65** ± 1.76	**96.43** ± 2.16
OURS – T4 All ROIs + DBC	**98.10** ± 1.15	**97.53** ± 0.47	**96.77** ± 0.89	**94.49** ± 2.76

Table II shows a statistically significant improvement in the overall performance when comparing the methods we propose and the baseline of Lukas et al. [9] and Li [10]. It also shows that it is possible to reliably identify image sources in an Open Set scenario. The results show that our implementation of the Open Set by means of Decision Boundary Carving is not worth employing when we have access to only two suspect cameras, but it can be useful when we have more suspect cameras. Furthermore, it is easy to see the improvement in results when we consider more ROIs for this identification. The approach proposed by Li [10] does not statistically improve the characterization part (considering the dataset used in this work).

Table III 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 by any of the 25 cameras. In this experiment we consider our approach in all ROIs. The results show the average of five tests per camera (5-fold).

TABLE III
RESULTS CONSIDERING CAMERAS WITH SAME BRAND AND MODEL (APPLE IPHONE 4).

Lukas et al. [9]	94.29 ± 2.20
Li [10]	93.90 ± 2.55
OURS – T3	95.54 ± 0.72
OURS – T4	95.17 ± 1.17

Table III shows that our approach is also effective in scenarios in which we have cameras of the same brand/model (in this case an Apple iPhone 4). Interestingly, in this case, the accounting for unknown via decision boundary carving, does not provide a result statistically different than the one without the optimization.

Table IV shows a breakdown for the case with 15 known cameras and 25 for testing (10 unknown). It shows the true positives, as well as the true negatives with results in $X\%$ $\pm$ σ (standard deviation), and in raw numbers considering the average of a 5-fold cross validation protocol. Note that the proposed method shows higher performance than Lukas et al.'s [9] and Li's [10] approaches considerably reducing the

risk for the unknown as we can see in the high number of true negatives (consequently low false positives) with a very low standard deviation and an increase in the true positives.

TABLE IV
BREAKDOWN FOR THE OPEN SET SETUP WITH 15 CAMERAS FOR TRAINING AND 25 FOR TESTING. RESULTS CONSIDERING THE AVERAGE OF A 5-FOLD CROSS-VALIDATION PROTOCOL. 450 + 300 TEST IMAGES PER ROUND.

	Lukas et al. [9]	**Li [10]**	**Ours – T4**
TP	92.5% $\pm$ 5.15 (27.75 / 30)	91.6% $\pm$ 5.70 (27.48 / 30)	**97.9%** $\pm$ 0.84 (29.37 / 30)
TN	96.5% $\pm$ 2.25 (694.8 / 720)	96.5% $\pm$ 2.41 (694.8 / 720)	**98.3**% $\pm$ 0.24 (707.8 / 720)

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

Expanding upon the work of Lukas et al. [9], our experiments report high accuracy results. The next step of this work is tuning the classification model for one class classification, in which 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 to combat against counter-forensic approaches, as presented in [32]. Future work includes analyzing some counter-forensic techniques to this work and the application of this technique to other pattern recognition and vision problems as it is general enough and we envision other applications for it.

Finally, we believe that efforts like the ones presented by [9], [10] and the one in this paper will move source attribution approaches toward meeting the strong standards of the Daubert trilogy [33] which establishes a high bar for acceptance of forensic evidence (analog and digital) in courts in the US and possibly in other countries.

ACKNOWLEDGEMENT

We would like to thank Microsoft Research and the São Paulo Research Foundation (FAPESP) for the financial support.

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. Liu, "Component forensic - theory, methodologies e applications," *IEEE SPM*, vol. 26, pp. 38–48, 2009.

[3] N. Khanna, A. K. Mikkilineni, G. T. C. Chiu, J. P. Allebach, and E. J. Delp, "Scanner identification using sensor pattern noise," *SPIE SSWMC*, vol. 6505, 2007.

[4] N. Khanna, A. K. Mikkilineni, and E. J. Delp, "Scanner identification using feature-based processing and analysis," *IEEE TIFS*, vol. 4, no. 1, pp. 123–139, 2009.

[5] P. Chiang, N.Khana, A. K. Mikkilineni, M. V. O. Segovia, S. Suh, J. P. Allebach, G. T. C. Chiu, and E. J. Delp, "Printer and scanner forensics: Examining the security mechanisms for a unique interface," *IEEE SPM*, vol. 72, 2009.

[6] E. Kee and H. Farid, "Printer profiling for forensic and balistic," in *ACM MM&Sec*, vol. 10, 2008.

[7] 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.

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

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

[10] C.-T. Li, "Source camera identification using enhanced sensor pattern noise," *IEEE TIFS*, vol. 5, no. 2, pp. 280–287, June 2010.

[11] A. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE TSP*, vol. 53, pp. 3948–3959, 2005.

[12] M. Kharrazi, H. Sencar, and N. Memon, "Blind source camera identification," in *IEEE ICIP*, Singapore, 2004.

[13] 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.

[14] S. Lyu, "Natural image statistics for digital image forensics," PhD Thesis, Dartmouth College, August 2005.

[15] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," in *SPIE*, vol. 7254, January 2009.

[16] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," *IEEE TIFS*, vol. 3, pp. 74–90, 2008.

[17] M. Goljan and J. Fridrich, "Identifying common source digital camera from image pairs," in *IEEE ICIP*, 2007, pp. 14–19.

[18] C.-T. Li, "Unsupervised classification of digital images using enhanced sensor pattern noise," in *IEEE ISCAS*, 2010, pp. 3429–3432.

[19] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti, "Fast image clustering of unknown source images," in *IEEE WIFS*, 2010, pp. 1–5.

[20] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, "Can we trust digital image forensics?" in *ACM Multimedia*, 2007, pp. 78–86.

[21] R. Caldelli, I. Amerini, and A. Novi, "An analysis on attacker actions in fingerprint–copy attack in source camera identification," in *IEEE WIFS*, 2011.

[22] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, pp. 536–544, 2003.

[23] P. Phillips, P. Grother, and R. Micheals, *Handbook of Face Recognition*. Springer, 2005, ch. Evaluation Methods on Face Recognition, pp. 329–348.

[24] B. Wang, X. Kong, and X. You, "Source camera identification using support vector machines," in *Advances in Digital Forensics V*, 2009, vol. 306, pp. 107–118.

[25] C. Cortes and V. Vapnik, *Machine Learning*, 20th ed. Kluwer Pub., 1995, ch. Support-Vector Networks, pp. 273–297.

[26] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution." *Neural Computation*, vol. 13, pp. 1443–1471, 2001.

[27] C.-T. Li and R. Sata, "On the location-dependent quality of the sensor pattern noise and its implication in multimedia forensics," in *ICDP*, 2011.

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

[29] X. Wang and Z. Weng, "Scene abrupt change detection," in *CCECE*, 2000, pp. 880–883.

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

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

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

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