Received: 11 October 2021 | Revised: 26 November 2021 | Accepted: 7 January 2022

DOI: 10.1111/1556-4029.14991

JOURNAL OF FORENSIC SCIENCES AAFS

PAPER

Digital & Multimedia Sciences

Source-anchored, trace-anchored, and general match score-based likelihood ratios for camera device identification

Stephanie Reinders PhD[1,2] | Yong Guan PhD[2,3] | Danica Ommen PhD[1,2] | Jennifer Newman PhD[2,4]

[1]Department of Statistics, Iowa State University, Ames, Iowa, USA

[2]Center for Statistics and Applications in Forensic Evidence, Iowa State University, Ames, Iowa, USA

[3]Department of Electrical and Computer Engineering, Iowa State University, Ames, Iowa, USA

[4]Department of Mathematics, Iowa State University, Ames, Iowa, USA

Correspondence
Stephanie Reinders PhD, Department of Statistics, Iowa State University, Ames, IA, USA.
Email: srein@iastate.edu

Funding information
This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, University of Virginia and Duke University. Open access funding provided by the Iowa State University Library

Abstract

Forensic camera device identification addresses the scenario, where an investigator has two pieces of evidence: a digital image from an unknown camera involved in a crime, such as child pornography, and a person of interest's (POI's) camera. The investigator wants to determine whether the image was taken by the POI's camera. Small manufacturing imperfections in the photodiode cause slight variations among pixels in the camera sensor array. These spatial variations, called photo-response non-uniformity (PRNU), provide an identifying characteristic, or fingerprint, of the camera. Most work in camera device identification leverages the PRNU of the questioned image and the POI's camera to make a yes-or-no decision. As in other areas of forensics, there is a need to introduce statistical and probabilistic methods that quantify the strength of evidence in favor of the decision. Score-based likelihood ratios (SLRs) have been proposed in the forensics community to do just that. Several types of SLRs have been studied individually for camera device identification. We introduce a framework for calculating and comparing the performance of three types of SLRs — source-anchored, trace-anchored, and general match. We employ PRNU estimates as camera fingerprints and use correlation distance as a similarity score. Three types of SLRs are calculated for 48 camera devices from four image databases: ALASKA; BOSSbase; Dresden; and StegoAppDB. Experiments show that the trace-anchored SLRs perform the best of these three SLR types on the dataset and the general match SLRs perform the worst.

KEYWORDS
digital cameras, digital evidence, digital images, forensic camera identification, score-based likelihood ratios and SLR

Presented at the International Association for Identification 105th Educational Conference, August, January 7, 2021, in Nashville, TN.

Highlights

- SLRs convey the strength of evidence in favor of a match or non-match for camera identification.
- The three types of SLRs considered achieve low rates of misleading evidence on the dataset.
- Trace-anchored SLRs outperform source-anchored and general match SLRs on the dataset.

1 | INTRODUCTION

Digital image forensics is a branch of forensic science that analyzes digital photographs and videos. Like many other areas of pattern evidence identified in the 2009 landmark report by the National Research Council [1], multimedia analysis was identified as lacking in probabilistic and statistical foundations. This absence of sound scientific methods provides challenges to meet the Daubert standard established by the court case Daubert v. Merrell Dow Pharmaceuticals [2] and can compromise the probative value of the evidence. Evidentiary strength of forensic findings relies on rigorous and peer-reviewed research experiments, where the reliability and validity of the analysis has been tested so it will withstand increasing scrutiny in the courts [3]. *Score-based likelihood ratios* (SLRs) provide one method for quantifying the probative value of a piece of evidence and are an area in which digital pattern evidence is building its repertoire of evidence-based research findings. In this paper, we develop a framework for calculating three types of SLRs to quantify the weight of evidence for the digital image forensic problem of *camera device identification*, where the goal is to identify a particular camera device (as opposed to camera model) that captured a questioned image.

In pattern evidence analysis, the investigator is often confronted with a *source identification* problem. The investigator has two impressions, one impression E_u from the crime scene where E_u is from an *unknown source*. The other impression E_k is directly acquired from a *specific known source*, related to the person of interest (POI). The investigator asks: how likely is it that both impressions originate from the specific known source? How likely is it that the crime scene impression does not originate from the specific known source? These questions are often expressed as two competing, specific-source hypotheses [4].

H_p: the impression E_u originated from the specificknown source that created E_k,
H_d: the impression E_u originated from a different sourcethan the specific known source.

In practice, the prosecution and the defense do not specify these hypotheses, but we will use standard nomenclature and refer to H_p as the prosecution's hypothesis and H_d as the defense's hypothesis.

An investigator faced with a camera device identification problem might have a digital image I_u that contains child pornography and a camera fingerprint F_k estimated from images of innocuous content from a POI's camera. (We use a capital letter *I* to denote images and a capital letter *F* to denote camera fingerprints. The subscripts denote the camera that created the image or fingerprint, the letter u stands for unknown camera and the letter k stands for the POI's camera, which is also called the specific known device.) The prosecution's hypothesis is that the child pornography image originated from the POI's camera that also created the camera fingerprint, while the defense's hypothesis is that the child pornography image did not originate from the POI's camera.

Most camera device identification methods rely on a camera sensor property called photo-response non-uniformity (PRNU) [5]. The measured response of a camera sensor array to incoming photons slightly varies from pixel to pixel due to the manufacturing process and imperfections in the photodiode. In principle, the spatial placement of these variations from the mean response of the array of pixels provides an identifying characteristic, or fingerprint, of the camera. We will use the terms PRNU and camera fingerprint interchangeably.

In Figure 1, we summarize the algorithm presented in [6,7] to estimate the camera fingerprint F_k of device C_k, which is given by equation

$$F_k = \frac{\sum_{j=1}^{n} X_k^j I_k^j}{\sum_{j=1}^{n} \left(I_k^j\right)^2}, \tag{1}$$

where I_k^j $(j = 1, 2, \ldots, n)$ are images from C_k and X_k^j $(j = 1, 2, \ldots, n)$ are noise residuals of the images calculated by subtracting a denoised version of the image from the image itself. The image I_k^J, the noise residual X_k^j, and the fingerprint F_k in Equation (1) can all be represented as matrices of pixel values and multiplication is performed element-wise. Note that we use slightly different notation than that in [6,7]: we use a subscript to denote the camera device to which an image belongs.

Collect a set of n images
For each image I_k^j $(j = 1, 2, \ldots, n)$
  Use denoising filter D to create a denoised version $D(I_k^j)$
  Calculate the noise residual $X_k^j = I_k^j - D(I_k^j)$
Calculate the PRNU F_k using equation (1)

FIGURE 1 Summary of the PRNU estimation algorithm presented in [6,7]

An investigator can estimate the true PRNU of a camera using the above algorithm and a set of images from the POI's camera. If the investigator has one single image, then the PRNU estimate is simply the noise residual of the single image. An assumption made of the extracted PRNU is that it is unique for a camera sensor. A recent paper [8] presents results that show that recent advances in the image processing pipeline such as customized HDR algorithms can cause small similarities in the camera fingerprints from different cameras of the same model. They show that these small similarities can slightly increase the rate of false positives between different cameras of the same model. We invite readers to pursue open questions that arise from this recent publication.

After estimating the camera fingerprint F_k from device C_k a questioned image I_u from unknown device C_u is compared with F_k using a (dis)similarity score Δ. The value of interest is

$$\delta = \Delta\left(X_u I_u F_k\right), \tag{2}$$

where multiplication between I_u and F_k is performed element-wise (See Ref. [6] for an explanation of why the noise residual is compared to the product of the image and the camera fingerprint). Early camera device identification works used the sample correlation as the similarity score [5,6,9,10]. Later the peak-to-correlation energy (PCE) replaced sample correlation because the PCE is robust against a periodic signal called *linear pattern* that created problems for statistical models that used sample correlation [11]. Most previous work in camera device identification focuses on developing what we term a *universal detector* [7,9,11,12,13,14,15] where the authors aim to create a single system that works for any questioned image and camera device. Over the years, universal detectors that use the PRNU have been tested against cropping and scaling [13], gamma correction and denoising [6], compression [14], lens distortion [16], and contrast enhancement, histogram equalization, and white balance [15]. Most previous work determines if a questioned image I_u came from a specific known camera C_k by comparing the value of interest $\delta = \Delta(X_u, I_u F_k)$ from Equation (2) to an ad-hoc decision threshold t, where t is typically chosen based on a constructed set of similarity scores between an image and a camera fingerprint known to be from the same camera (matching) and an image and a camera fingerprint known to be from different cameras (non-matching). In the universal detector approach, the researchers' goal is to create a single detector and decision threshold t that can be applied to any image and any camera device. Of particular interest to us, the universal detector approach addresses the common-source hypotheses, which are less pertinent to the decision-makers in criminal justice trials than the specific-source hypotheses (see Ref. [4] for more information on the common/specific-source problems). Furthermore, the universal detector methods aim to *classify* pairs of images as either matching or non-matching, using pattern recognition and classification methods, such as linear discriminant analysis, to define the decision threshold. This results in a binary decision that gives no information about the strength of the evidence in favor of that decision. These methods differ from the likelihood ratio-style approaches, which aimed to *quantify* the probability of observing the evidence under two competing hypotheses.

Likelihood ratios (LRs) are used in single source DNA analysis [17] and glass fragment analysis [18,19]. An LR for pattern evidence is defined as the ratio of the likelihoods of observing both impressions under hypotheses H_p and H_d. More specifically, for an impression E_u from a crime scene and an impression E_k from a source related to the POI, the LR would be written

$$\mathrm{LR} = \frac{P\left(E_u E_k H_p\right)}{P\left(E_u E_k H_d\right)}, \tag{3}$$

where $P(\cdot)$ is a joint probability density (or mass) function (PDF) [20]. The PDF in the numerator of the LR describes the likelihood of observing both impressions E_u and E_k if they originated from the specific known source and the PDF in the denominator would be the likelihood of observing both impressions E_u and E_k if they were a "random match" and E_u came from a different source than the specific known source. To formulate an LR as in Equation (3), the investigator needs an applicable set of measurements, often called features, where the variability between the features of two impressions from the same source can be distinguished, with high accuracy, from the variability between the features of two impressions from different sources. The representation of many pattern evidence data is often high-dimensional and complex, making such sets of features extremely challenging to identify. Alternate methods for assessing the weight of the evidence are being explored in various forensic fields, including machine learning with paired feature differences [21] and SLRs [22].

To our knowledge, full-fledged likelihood ratios have not yet been implemented in camera device identification, but Qiao et al. [23,24] employed a related method, a likelihood ratio test (LRT), for camera device identification to determine which of two camera devices C_1 or C_2 captured a questioned image I_u from an unknown source. They formulate the problem as a classification problem and consider the following two hypotheses:

$$H_1: I_u \text{ originated from device } C_1$$
$$H_2: I_u \text{ originated from device } C_2.$$

Note that we use our own notation here to be consistent with the rest of this paper. The authors also address the question of which device from a set of n devices $C_1, C_2, \ldots, C_n$ took questioned image I_u by performing an LRT for each possible pair of devices and declaring the device identified by the most LRTs to be the device that captured the I_u [23]. If a case arises in practice where a questioned image is known to have been taken by one of two devices, the likelihood ratio that Qiao et al. treat as the test statistic in their likelihood ratio test could potentially quantify the strength of evidence. Because their intent was to solve a multi-classification problem by determining which of n classes (cameras) the question image I_u came from, this framework does not readily provide a means of quantifying the strength of the evidence when the pool of potential sources

of I_u is larger than two devices. LRs still remain elusive for camera device identification.

SLRs have appeared in a variety of forensic fields when LRs are unavailable: glass fragments and shoe impressions [22]; handwriting [25]; MDMA tablets [26]; fingerprints [27]; speaker recognition [28]; and facial recognition [29]. SLRs use a similarity score Δ to measure the similarity (or dissimilarity) of two pieces of evidence E_u and E_k and then calculate the likelihood of obtaining the score $\Delta(E_u, E_k)$ under hypotheses H_p and H_d. An SLR is then the ratio of these two likelihoods. A reference set of *matching* scores between two pieces of evidence known to come from the same source are used to estimate the PDF in the SLR numerator. The PDF in the denominator is estimated from a reference set of *non-matching* scores between two pieces of evidence known to come from different sources. We consider three methods presented by Hepler et al. [25] – trace-anchored, source-anchored, and general match – for constructing non-matching scores to estimate the SLR denominator for handwriting evidence. Consider the scenario where investigators have a questioned handwritten document E_u and a handwriting sample E_k from a POI. They specify an alternative population of possible writers. Researchers and practitioners are still in disagreement about how alternative populations should be constructed [30,31]. The trace-anchored method considers similarity scores between the questioned document E_u and handwriting samples from writers in the alternative population. Hepler et al. [25] define the trace-anchored SLR as

$$\mathrm{SLR}_{\mathrm{trace}} = \frac{P\left(\Delta\left(E_u E_k\right) E_k H_p\right)}{P\left(\Delta\left(E_u E_k\right) E_u H_d\right)}. \quad (4)$$

Neumann and Ausdemore [32] present an alternative definition of a trace-anchored SLR:

$$\mathrm{SLR}_{\mathrm{trace}} = \frac{P\left(\Delta\left(E_u E_k\right) E_u H_p\right)}{P\left(\Delta\left(E_u E_k\right) E_u H_d\right)}.$$

In this case, both the numerator and denominator condition on the evidence from the unknown source, unlike Equation (4) that anchors on the evidence from the known source in the numerator and the evidence from the unknown source in the denominator. Neumann and Ausdemore state an SLR that anchors on the evidence from the unknown source in the numerator is not useful in practice because the investigator would need other objects from the same unknown source as E_u to estimate the numerator. Additionally, Neumann and Ausdemore point out that because the trace-anchored SLR defined by Hepler et al. anchors on two different objects in the numerator and denominator it is highly unlikely that it will converge to the desired Bayes factor. However, this is only a drawback from a Bayesian perspective and the validity of this method for other statistical frameworks has yet to be explored.

The source-anchored method considers similarity scores between the POI's handwriting sample E_k and handwriting samples from writers in the alternative population. The source-anchored SLR is defined

$$\mathrm{SLR}_{\mathrm{source}} = \frac{P\left(\Delta\left(E_u E_k\right) E_k H_p\right)}{P\left(\Delta\left(E_u E_k\right) E_k H_d\right)}. \quad (5)$$

Neumann and Ausdemore [32] criticize this approach for being incoherent from a Bayesian perspective, but Garton [33] takes a different viewpoint.

The general match method considers similarity scores between handwriting samples from pairs of different writers randomly selected from the alternative population. The general match SLR is defined

$$\mathrm{SLR}_{\mathrm{general}} = \frac{P\left(\Delta\left(E_u E_k\right) E_k H_p\right)}{P\left(\Delta\left(E_u E_k\right) H_d\right)}. \quad (6)$$

Like the trace-anchored SLR, the general match SLR also does not anchor on the same object in the numerator and denominator. Again, this is problematic from the Bayesian perspective, but further research is needed to ascertain whether this is an issue from the perspective of other statistical frameworks. Hepler et al. [25] demonstrate that these three SLR types can lead to different conclusions on the same evidence. Because of this, the method investigators use to build their reference set of non-matching scores is extremely important.

In many situations, SLRs from Equations (4)–(6) are easier to apply than LRs from Equation (3) because similarity scores reduce the dimensionality of the problem. Instead of needing to fit probability distributions in high dimensions, the investigator only needs to fit distributions to lower dimensional similarity scores. One drawback of SLRs is that unlike LRs, they do not account for rarity. For example, if a witness told the police the color, make, and model of the car that fled the crime scene and a person of interest owns a car of the same color, make, and model, an LR would take into account the rarity of that color, make, and model in the general population of cars (e.g., a red Ferrari Portofino is rarer than a black Toyota Corolla). In contrast, an SLR does not consider the rarity of the car involved (e.g., two red Ferrari Portofino cars are just as similar in color, make and model as two black Toyota Corolla cars). Another drawback of SLRs is that when "pairwise comparison" methods are used, dependency is introduced into the resulting score data. Meaning that any two similarity scores that were created using the same object as one item in the pair will be dependent (e.g., $\Delta(I_1, I_2)$ and $\Delta(I_1, I_3)$ are dependent because they both involve item I_1). Unfortunately, no one knows how to fix this problem yet. Because LRs are as of yet unavailable for camera device identification when more than two devices are considered, despite the limitations, SLRs are the only available alternative.

SLRs have been applied to the digital image forensic problem of *camera source identification*, where the source being identified is a particular camera device [34,35]. Nordgaard and Höglund [34]

introduce the framework for calculating source-anchored SLRs, one of the available types of SLRs. They perform simulation studies in the case where a questioned image came from one of two cameras, and they discuss how their method could be applied to a larger set of cameras. van Houten et al. [35] addressed the scenario where an investigator knows the make and model of the camera that took a questioned image and wishes to determine which device out of a set of 9 or 10 devices of that make and model captured the image. They present specific-source hypotheses, but they construct SLRs that address common-source hypotheses. The difference between the numerators of a common-source SLR and a specific-source SLR lies in the construction of the reference sets of matching scores used to estimate the numerators' PDFs [31]. The reference set of matching scores for a common-source SLR consists of matching scores from devices in the alternative device population in addition to matching scores from the specific known device. The specific-source SLR uses matching scores only from the specific known device. The denominator of a common-source SLR is the same as the general match denominator for a specific-source SLR. van Houten et al. performed experiments on two camera models: the Motorola V360 mobile phone camera (10 cameras) and the Sony DSC-S500 camera (nine cameras). Reinders [36] and Reinders et al. [37] adapted the method Hepler et al. used to calculate trace-anchored SLRs for the camera device identification problem. We present an extended investigation of the use of SLRs for camera device identification beyond the current published literature and develop a framework for calculating all three available types of specific-source SLRs with a larger dataset of 48 camera devices from 26 distinct camera models. We do not compare our method of computing source-anchored SLRs to those in the literature since other authors tailored their methods to small database sizes, and we designed our methods to work with large databases.

As mentioned before, SLRs quantify the strength of evidence in favor of one of the hypotheses over the other. More specifically, if an SLR is greater than 1 it shows support for H_p rather than H_d and the larger the SLR the stronger the support for H_p. On the other hand, an SLR value less than or equal to 1 supports H_d rather than H_p and the smaller the SLR the stronger the support for H_d. Figure 2 illustrates this.

Section 2 describes the proposed framework for calculating the three types of SLRs for camera device identification. Section 3 explains the results obtained from applying the proposed methods to a dataset of 48 camera devices. Section 4 discusses the strengths and limitations of the proposed methods and the implications of the findings for future work.

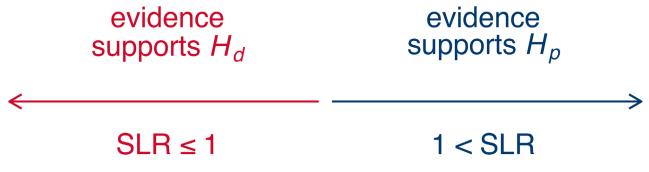


FIGURE 2 Basic interpretation of SLR values [Color figure can be viewed at wileyonlinelibrary.com]

2 | METHODS

We present a framework for calculating three types of specific-source SLRs [25] – trace-anchored, source-anchored, and general match – for the camera device identification problem. We demonstrate this framework on an image dataset from 48 camera devices representing 26 distinct camera models. Building upon previous work, our framework offers a more comprehensive application of SLRs for camera device identification.

We start by describing the scenario we consider. Then we formulate two competing hypotheses H_p and H_d for this scenario. We specify the alternative camera devices and image data that we use in our analyses. We estimate camera fingerprints from the POI's camera and each of the alternative devices. Then we calculate a similarity score δ between the questioned image and the camera fingerprints from the POI's camera. To estimate the probabilities of observing the score δ under each hypothesis, we build reference sets of matching and non-matching scores using image data from the alternative camera devices. Finally, we construct three types of SLRs as the ratio between probability of observing δ under H_p (same source) and the probability of observing δ under H_d (different source).

We consider the scenario where an investigator has two pieces of evidence: a digital image from an unknown camera device that was involved in a crime; and a camera fingerprint from a POI's camera device. Our method requires that the investigator has access to one or more images that are known to have originated from the POI's camera that can be used to estimate a camera fingerprint. Instead of using the generic labels E_u and E_k for evidence as in the previous section, here we denote a questioned image of unknown source as I_u to make clear that the evidence is an image. We use F_k to denote a camera fingerprint from the specific known source, the POI's camera, and we use C_k to refer to the POI's camera. We construct the following two competing hypotheses:

H_p: questioned image I_u and camera fingerprint F_k both originated from camera device C_k
H_d: camera fingerprint F_k originated from camera device C_k, but questioned image I_u did not.

The goal then is to construct score-base likelihood ratios to evaluate the strength of the evidence regarding these specific-source hypotheses.

Our experiments use image data from four digital image databases: Alaska version 1 (R. Cogranne, Q. Giboulot, P. Bas, personal communication, December 1, 2019), BOSSbase [38], Dresden (Dressden Image Database, T. Gloe, R. Bohme, personal communication, December 1, 2019), which is described in [39], and StegoAppDB [40,41]. Because this work is the first implementation of our proposed framework, we chose to restrict our experiments to RAW, auto-exposure images that we converted to TIFF format using Adobe Photoshop's Image Processor without LZW compression or resizing. We use photos taken in landscape orientation only and ignore devices with fewer than 100 such images. Restricting to auto-exposure eliminates the effect of manual camera exposure, which may affect the PRNU accuracy. Using the RAW image

data and converting to TIFF eliminates the effect of compression quality from JPEG images that can also affect accuracy. Landscape photos avoid calculation of rotation of the device to perform the "best" fit, another computational issue we put aside. Finally, we use 512 × 512 sub-images cropped from the center of each photo, rather than the entire photo itself. This avoids the computationally expensive alignment process to compare images of different sizes, as would be necessary in real-world case scenarios. These choices limit the effect of other complicating factors that potentially affect accuracy, so that analysis of SLR scores can avoid complicating factors. Future work should investigate these and other factors for their impact to accuracies.

A total of 48 devices from all four databases had at least 100 RAW, auto-exposure, landscape-oriented images, so these are the devices that we use in our experiments. Of those 48 devices, 23 are digital still cameras, 24 are mobile phones, and one is a tablet. The 48 devices represent 26 distinct camera models and 16 of the models have at least two devices. Ideally, we would have a much larger set of devices, but this is the largest set of images that we could find where the ground truth of the camera device has been authenticated. We randomly select 100 images from each of the 48 devices and pre-process the images by converting the RAW images to TIFF in Photoshop using the Image Processor with no LZW compression or resizing. Then we center-crop the images to 512 × 512 and save them as PNG in MATLAB. (Images can be cropped in Python using the Python Imaging Library or similar libraries.) We split the sample of 100 images from each of the 48 devices into a training set of 80 images and a set of 20 testing images, which serve as our questioned images. We have 20 × 48 = 960 questioned images in total.

Camera fingerprints are estimated from each of the 48 camera devices. (MATLAB and Python implementations of fingerprint estimation are available at [42]. We used the MATLAB code.) Each device C_i has $n = 80$ training images $I_i^1, I_i^2, \ldots, I_i^n$. (The subscript denotes the camera, and the images are numbered in the superscript.) A denoising filter D is used to extract a noise residual from each image: $X_i^j = I_i^j - D\left(I_i^j\right)$ for $j = 1, 2, \ldots, n$. The $n = 80$ training images are divided into 8 folds of 10 images each. Camera fingerprint F_i^1 calculated with Equation (1) and all training images except those in fold 1, camera fingerprint F_i^2 is estimated from all training images except those in fold 2, and so on. This results in 8 camera fingerprints from device C_i.

We need a way to measure the similarity (or dissimilarity) between the questioned image I_u and the camera fingerprint F_k from the POI's device. In other words, we need to choose which similarity score to use in Equation (2). We used peak-to-correlation energy (PCE) in our initial experiments, but we found that the large variance (on the order of 10^6) of the observed PCE scores produced many unstable SLR values where the numerator of the SLR is tiny and the denominator is zero. We found that for our dataset the sample correlation has much smaller variance (on the order of 10^{-3}) and thus produced more stable results. We chose to use the correlation distance, which is defined as one minus the sample correlation. For $X, Y \in \mathbb{R}^n$ the correlation distance is

$$1 - \frac{(X-\overline{X})'(Y-\overline{Y})}{\sqrt{(X-\overline{X})'(X-\overline{X})}\sqrt{(Y-\overline{Y})'(Y-\overline{Y})}},$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. The correlation distance between two images X and Y with dimensions $m \times n$ can be calculated by first converting the images to vectors of length mn. The noise residual $X_u = I_u - D(I_u)$ is obtained by subtracting a denoised version of the image, created with denoising filter D, from the image itself. Then the investigator calculates the correlation distance between the questioned image I_u and the camera fingerprint F_k from the POI's camera using Equation (2) where multiplication between I_u and F_k is performed element-wise and X_u and $I_u F_k$ are first converted to vectors.

A trace-anchored, source-anchored, and general match SLR is calculated for each questioned image I_u and each of the 48 devices set as the specific known device C_k in turn. For a given questioned image I_u and a given specific known device C_k the correlation distance is calculated between a noise residual X_u of I_u and the product (element-wise) $I_u F_k^j$ (for $j = 1, \ldots, 8$) and the results are averaged:

$$\delta = \frac{1}{8}\sum_{j=1}^{8} \Delta\left(X_u I_u F_k^j\right). \tag{7}$$

By taking the average score over the eight fingerprints, we are adapting the subsampling algorithm used by Hepler et al. [25] for estimating the numerator distribution for the device identification problem by creating "pseudo camera fingerprints."

We build reference sets of known matching scores and three sets of non-matching scores – trace-anchored, source-anchored, and general match – to estimate the probability of obtaining the score δ under each hypothesis. The reference set of matching scores is used to estimate the probability of obtaining the score δ if I_u and F_k originated from the same camera. A matching score is calculated as the correlation distance between the j-th fingerprint F_k^j and the noise residual $X_k^{\ell,j}$ of image $I_k^{\ell,j}$, which is the ℓ-th image in fold j.

$$\text{Matching scores:} \quad \Delta\left(X_k^{\ell,j} I_k^{\ell,j} F_k^j\right) \text{ for } \ell = 1, \ldots, 10 \text{ and } j = 1, \ldots, 8.$$

Figure 3 illustrates the calculation of matching scores. In total, we calculate 8 × 10 = 80 matching scores for device C_k. Note that scores calculated using the same fingerprint or the same noise residual and image are dependent, so we do not have 80 independent scores. Also, note that matching scores do not use the question image I_u.

Non-matching scores require a set of alternative camera devices. In our experiments, we set aside specific known device C_k and treat the other 47 devices from our dataset as the set of alternative camera devices. To calculate a trace-anchored non-matching score, we fix the questioned image I_u and calculate the correlation distance between its noise residual X_u and a camera fingerprint F_a from an alternative device C_a.

$$\text{Trace} - \text{anchored non} - \text{matching score:} \quad \Delta\left(X_u I_u F_a\right).$$

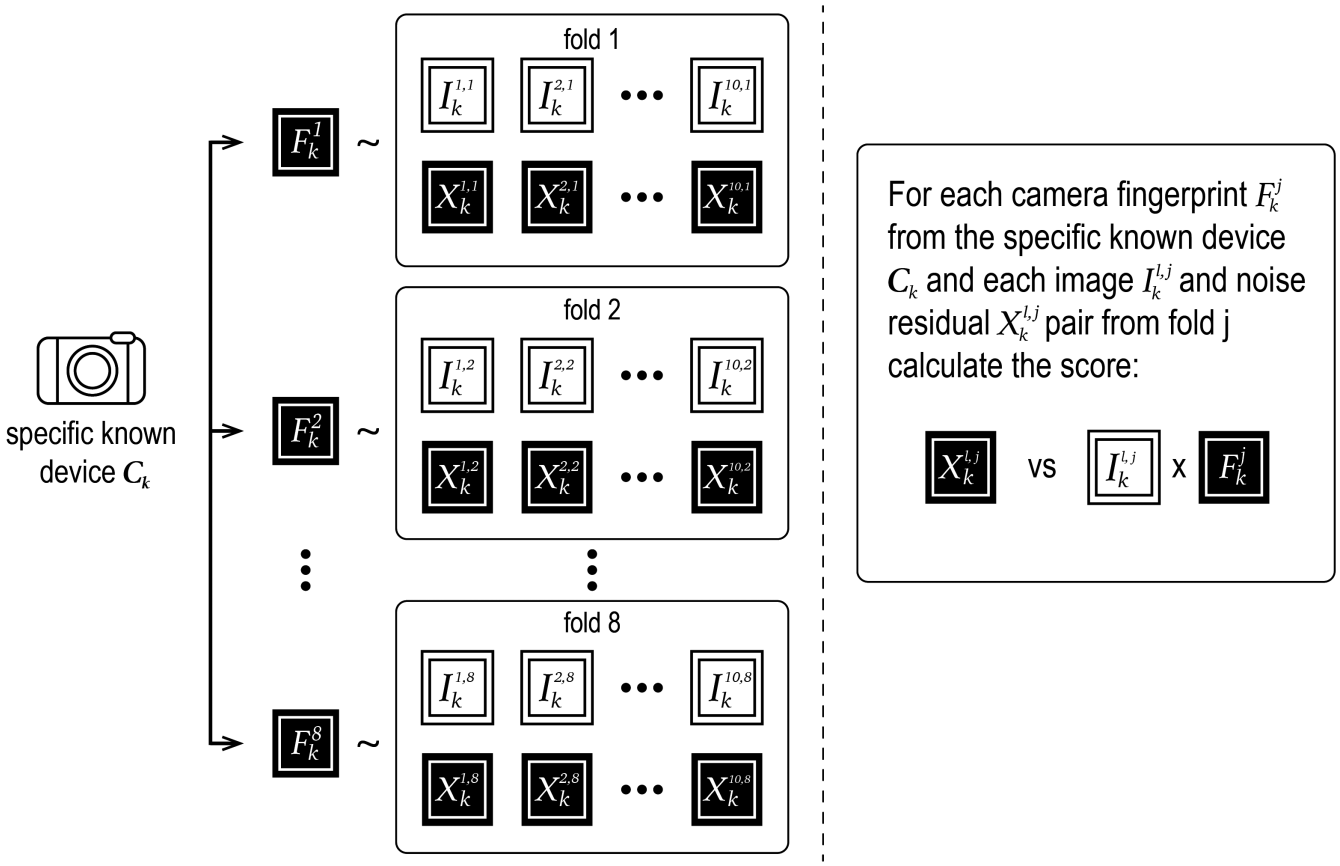
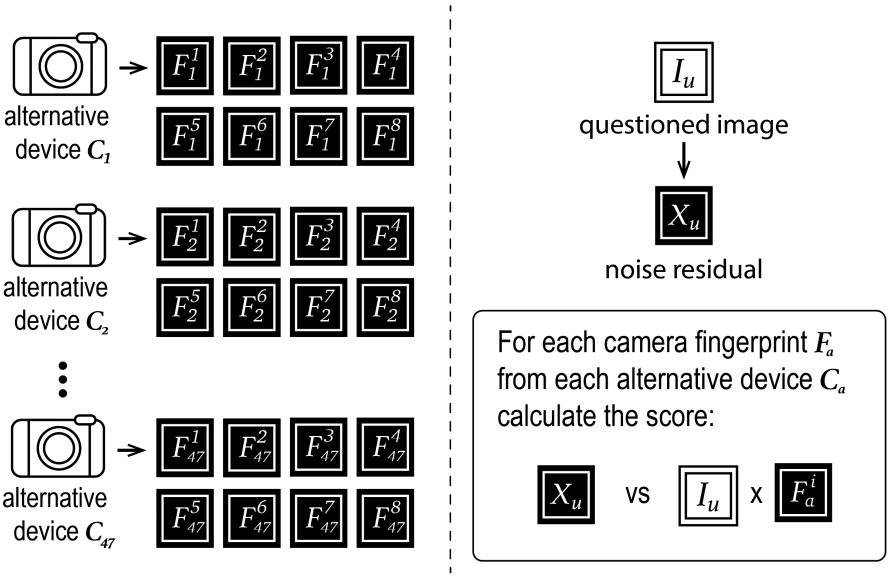


FIGURE 3 Calculating matching scores for specific known device C_k. Each camera fingerprint F_k^j was estimated from all training images except those in fold *j*. because the images in fold *j* were not used to estimate fingerprint F_k^j, matching scores are calculated between fingerprint F_k^j and the images and corresponding noise residuals in fold *j*

FIGURE 4 Calculating trace-anchored non-matching scores for questioned image I_u and specific known device C_k



As illustrated in Figure 4, we calculate trace-anchored non-matching scores between questioned image I_u and each of the eight fingerprints from each of the 47 alternative devices. This results in $8 \times 47 = 376$ trace-anchored non-matching scores for questioned image I_u. Because some of these scores were calculated with either the same fingerprint or the same noise residual and image these scores are not independent. Notice that trace-anchored scores do not use any information from the POI's device C_k. Also, while we

know that C_a and the POI's camera C_k are not the same device, the questioned image I_u could have originated from C_a. However, our hypotheses ask whether I_u originated from a device other than C_k but do not identify the alternative source of the image. If the investigator wants to evaluate the probability that I_u originated from device C_a, then new hypotheses should be constructed and a new SLR with C_a in place of C_k as the specific known device will be calculated.

A source-anchored non-matching score is calculated between a fingerprint F_k of specific known device C_k and the noise residual X_a of a training image I_a from alternative device C_a.

$$\text{Source} - \text{anchored non} - \text{matching score:} \quad \Delta\left(X_a I_a F_k\right).$$

We calculate source-anchored non-matching scores between each of the 8 camera fingerprints from C_k and each of the 80 training images from each of the 47 alternative devices for a total of $8 \times 80 \times 47 = 30{,}080$ scores. Many of these scores are dependent because they are calculated from either the same fingerprint or the same noise residual and image. Figure 5 illustrates the calculation of these scores. Note that the questioned image I_u is not considered in the source-anchored scores.

The last type of non-matching scores are general match non-matching scores. To calculate one of these scores we randomly select two different devices C_1 and C_2 from the set of 47 alternative devices. Then we calculate the correlation distance between the noise residual X_1 of a training image I_1 from one of the devices with a camera fingerprint F_2 from the other device.

$$\text{General match non} - \text{matching scores:} \quad \Delta\left(X_1 I_1 F_2\right) \text{ and } \Delta\left(X_2 I_2 F_1\right).$$

We also calculate $\Delta(X_2, I_2F_1)$ where X_2 is the noise residual of a training image I_2 from device C_2 and F_1 is a camera fingerprint from device C_1 because $\Delta(X_1, I_1F_2)$ is not generally equal to $\Delta(X_2, I_2F_1)$. Figure 6 shows that we calculate a general match non-matching score between the 80 training images of one device and the 8 camera fingerprints of the other device for each pair of alternative devices. This results in $47 \times 46 \times 80 \times 8 = 1{,}383{,}680$ general match non-matching scores. Again, many of these scores are dependent because they were calculated from either the same fingerprint or the same noise residual. Notice that the general match non-matching scores do not use the questioned image or the specific known device.

We do not know the true PDFs of matching and non-matching scores, so we fit PDF estimates to each set of scores and use these estimates to construct the SLR. We acknowledge that there are several possible methods for estimating these PDFs, including both parametric and non-parametric options. Nordgaard and Höglund [34] use a parametric method due to the relatively low amount of background information (they only have two camera devices to use for comparison). In contrast, we have much more background information (hundreds of images from 48 camera devices), so we chose to explore a non-parametric method. We use kernel density estimation to fit PDFs f_m, f_{trace}, $f_{source,}$ and $f_{general}$ to the reference sets of matching, trace-anchored, source-anchored, and general match non-matching scores, respectively. Similar methods were employed in [43]. We use the MATLAB fitdist function to perform kernel density estimation. (Kernel density estimation can be performed in Python with Scikit-Learn's Nearest Neighbors library.) The kernel density estimator $\hat{f}_h: \mathbb{R} \to [0\,\infty)$ applied by fitdist is defined as

$$\hat{f}_h(y) = \frac{1}{nh}\sum_{i=1}^{n}\frac{K(y - y_i)}{h}, \tag{8}$$

where n is the sample size, y_1, ..., y_n are random samples from the unknown distribution, K is the kernel smoothing function, and h is the bandwidth. In our case, y_1, ..., y_n are the scores, K is the normal kernel function [44]. We allow fitdist to choose the optimal bandwidth h. Generally, this method of naively selecting the bandwidth will undersmooth the estimated PDFs because this method relies on the assumption of independent data, which we know we do not have due to the pairwise nature of creating the matching and non-matching scores.

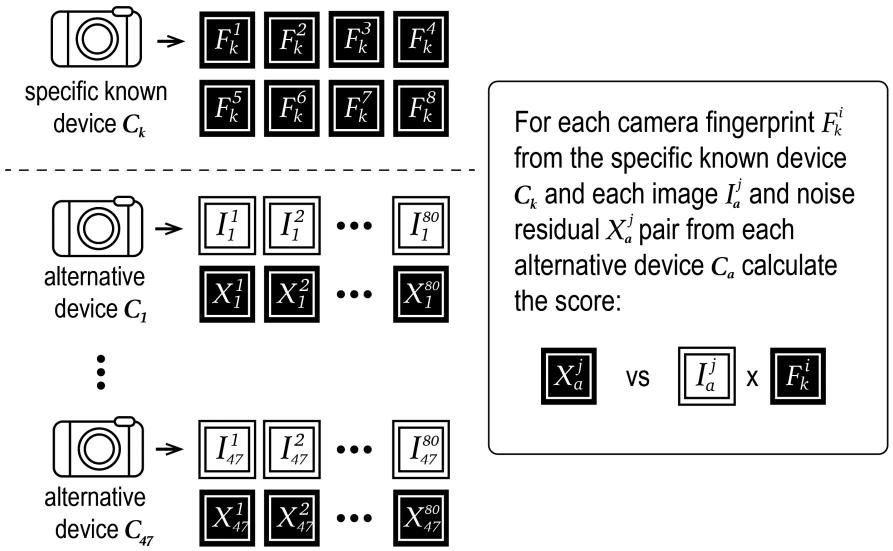


FIGURE 5 Calculating source-anchored non-matching scores for specific known device C_k

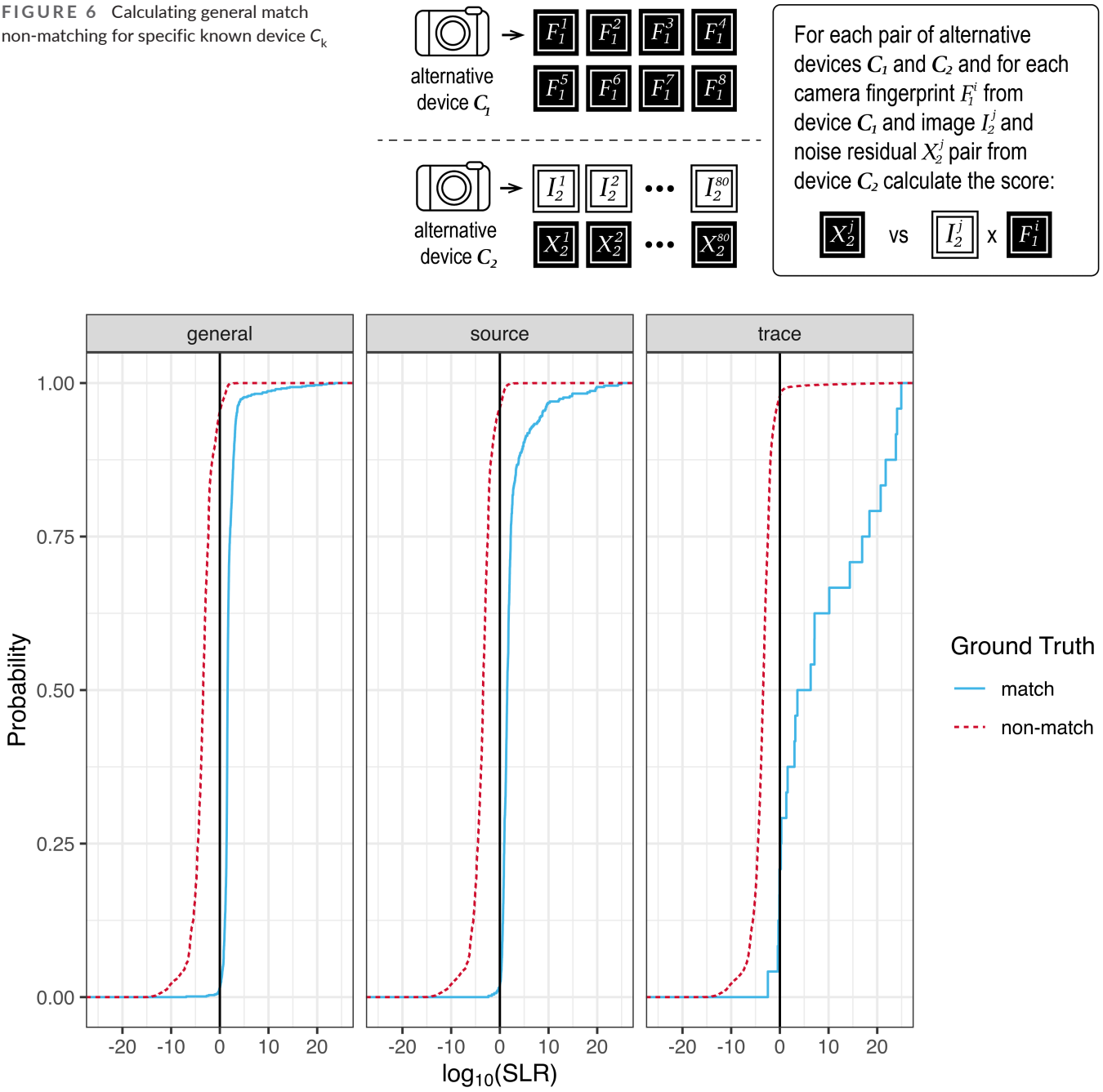


FIGURE 6 Calculating general match non-matching for specific known device C_k

FIGURE 7 Tippet plots of general match, source-anchored, and trace-anchored SLRs under two scenarios: Match (H_p is true) and non-match (H_d is true) [Color figure can be viewed at wileyonlinelibrary.com]

However, there is currently no method of fitting better PDF estimates for pairwise dependent data.

All the pieces are in place for us to calculate the trace-anchored, source-anchored, and general match SLRs. The numerator of all three SLR types shown in Equations (4)–(6) are estimated by fitting a PDF f_m to the matching scores matching scores reference set using Equation (8) and evaluating f_m at the mean similarity score δ from Equation (7). The denominator of the trace-anchored SLR is the PDF f_{trace} fit to the set of trace-anchored non-matching scores using Equation (8) evaluated at δ from Equation (7). The trace-anchored SLR is defined

$$\mathrm{SLR}_{\mathrm{trace}} = \frac{f_{\mathrm{m}}(\delta)}{f_{\mathrm{trace}}(\delta)}.$$

Similarly, the source-anchored SLR is defined

$$\mathrm{SLR}_{\mathrm{source}} = \frac{f_{\mathrm{m}}(\delta)}{f_{\mathrm{source}}(\delta)},$$

where f_{source} is the PDF fit to the source-anchored non-matching scores. Lastly the general match SLR is defined

$$\mathrm{SLR}_{\mathrm{general}} = \frac{f_{\mathrm{m}}(\delta)}{f_{\mathrm{general}}(\delta)},$$

where f_{general} is the PDF fit to the general match non-matching scores.

The $\mathrm{SLR}_{\mathrm{trace}}$, $\mathrm{SLR}_{\mathrm{source}}$, and $\mathrm{SLR}_{\mathrm{general}}$ use the same numerator, which gives the likelihood of observing the score δ if H_{p} is true. The denominators of the three SLRs give the likelihood of observing the score δ if H_{d} is true, each SLR type using a different definition of non-matching scores. An SLR value is the ratio of these two likelihoods. sss 2 shows how SLR values are commonly interpreted based on whether they are less than or greater than 1.

3 | RESULTS AND DISCUSSION

We calculate the trace-anchored, source-anchored, and general match SLRs between 960 questioned images (20 images from each of the 48 camera devices in the dataset) and each of the 48 devices set as the specific known device in turn, resulting in $3 \times 960 \times 48 = 138{,}240$ SLRs. The prosecution hypothesis H_{p} is true (the questioned image I_{u} and the camera fingerprint F_{k} both originated from the person of interest's camera C_{k}) for $3 \times 48 \times 20 = 2880$ of these SLRs, and the defense hypothesis H_{d} is true (the fingerprint F_{k} originated from the POI's camera C_{k} but the questioned image I_{u} did not) for the other 135,360 SLRs.

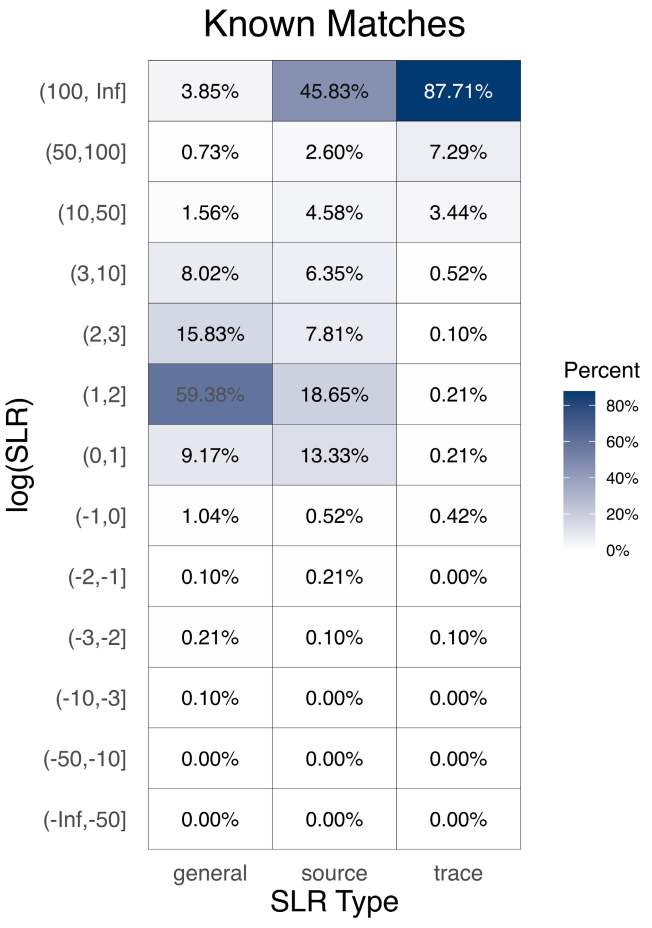


FIGURE 8 Each tile shows the percentage of known matching $\log_{10}$(SLR) values that fall into a particular interval. values greater than 0 correctly support H_{p} relative to H_{d} and values less than or equal to 0 are misleading evidence in favor of H_{d}. Values closer to 0 show weaker support and values farther for 0 show stronger support [Color figure can be viewed at wileyonlinelibrary.com]

3.1 | Overall performance

The Tippet plots in Figure 7 (following the convention used in [45]) show the empirical cumulative distribution function of SLR scores

TABLE 1 Rates of misleading evidence in favor of H_{p}

General match	Source-anchored	Trace-anchored
0.0466	0.0409	0.0267

TABLE 2 Rates of misleading evidence in favor of H_{d}

General match	Source-anchored	Trace-anchored
0.0146	0.00833	0.00521

TABLE 3 Rates of misleading evidence in favor of H_{p} by the model of the questioned image

Model of questioned image	General match	Source-anchored	Trace-anchored
Canon EOS 100D Rebel SL1	0.0277	0.0399	0.0447
Canon EOS 20D	0	0	0
Canon EOS 400D	0	0	0.0011
Canon EOS 60D	0.034	0.0394	0.0404
Canon Rebel XSi	0	0	0.0043
iPad pro 7.1 13 inch	0.183	0.1064	0.0447
iPhone 6s	0.0899	0.0404	0.0149
iPhone 6s Plus	0.0383	0.0117	0.0064
iPhone 7	0.0316	0.0133	0.0106
iPhone 7 Plus	0.05	0.0229	0.0096
iPhone 8	0.0787	0.0489	0.0229
iPhone X	0.0718	0.0495	0.0191
Nikon 1 AW	0	0	0.0011
Nikon D200	0.0191	0.0213	0.0234
Nikon D5200	0.0848	0.0851	0.067
Nikon D70	0.077	0.0848	0.0365
Nikon D70S	0.0819	0.0846	0.0473
Nikon D7100	0.084	0.083	0.0287
OnePlus 5	0.0154	0.0186	0.0218
Panasonic Lumix DMC FZ28	0	0	0.0011
Panasonic Lumix DMC GM1	0	0	0.0011
Pentex K50	0.0213	0.0213	0.0223
Pixel 1	0.0362	0.0574	0.0314
Pixel 2	0.0144	0.0197	0.0202
Samsung Galaxy S8	0.0468	0.0521	0.0617
Sony ILCE alpha 6000	0.0681	0.0809	0.0532

TABLE 4 Rates of misleading evidence in favor of H_d by the model of the questioned image

Model of questioned image	General match	Source-anchored	Trace-anchored
Canon EOS 100D Rebel SL1	0.025	0	0
Canon EOS 20D	0	0	0
Canon EOS 400D	0	0	0
Canon EOS 60D	0.05	0.05	0.05
Canon Rebel XSi	0	0	0
iPad pro 7.1 13 inch	0	0	0
iPhone 6s	0	0	0
iPhone 6s Plus	0	0	0
iPhone 7	0.0375	0.0625	0
iPhone 7 Plus	0	0	0
iPhone 8	0.025	0.05	0
iPhone X	0.025	0	0
Nikon 1 AW	0	0	0
Nikon D200	0	0	0
Nikon D5200	0	0	0
Nikon D70	0.05	0	0
Nikon D70S	0	0	0
Nikon D7100	0	0	0
OnePlus 5	0	0	0
Panasonic Lumix DMC FZ28	0	0	0
Panasonic Lumix DMC GM1	0	0	0
Pentex K50	0	0	0
Pixel 1	0.0125	0	0
Pixel 2	0.075	0	0
Samsung Galaxy S8	0	0	0
Sony ILCE alpha 6000	0	0	0.2

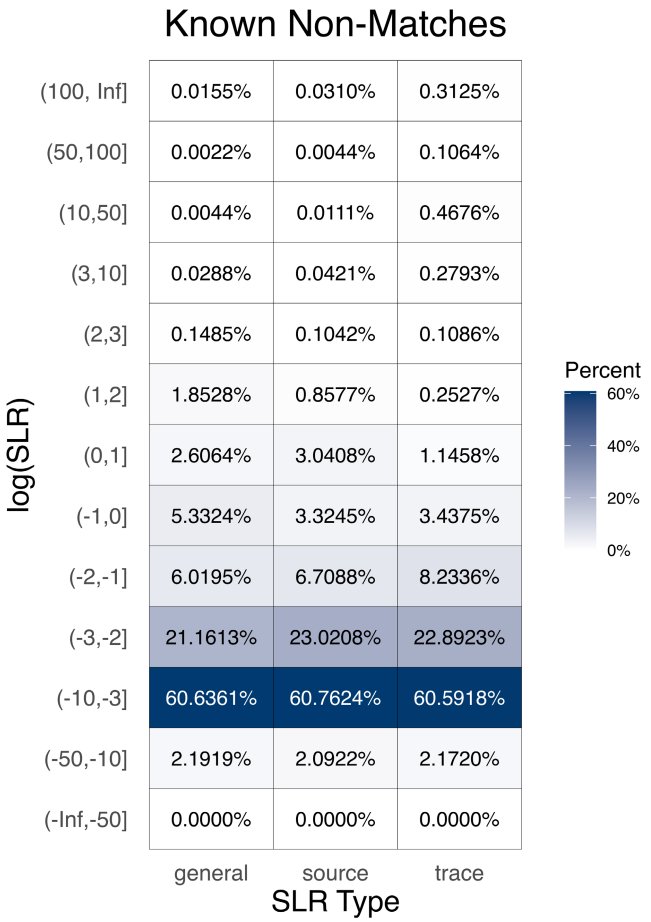


FIGURE 9 Each tile shows the percentage of known non-matching $\log_{10}$(SLR) values that fall into a particular interval. values less than or equal to 0 correctly support H_d relative to H_p and values greater than 0 are misleading evidence in favor of H_p. Values closer to 0 show weaker support and values farther for 0 show stronger support [Color figure can be viewed at wileyonlinelibrary.com]

for each SLR type when H_p is true and when H_d is true. We see that the three SLR types under consideration perform well but imperfectly. Misleading evidence in favor of H_d occurs when H_p is true but the SLR score is less than zero. Misleading evidence in favor of H_p occurs when H_d is true but the SLR score is greater than zero. The Tippet plots show that both types of misleading evidence occur for all three SLR types. The trace-anchored SLRs might appear on first glance to perform poorly when H_p is true, but that is not the case. The trace-anchored SLR curve under H_p rises more slowly than the other two SLR types because the majority of trace-anchored SLR scores are larger than the source-anchored and general match scores (see Figure 8). This behavior is precisely what we wish to see because it means that when H_p is true many of the trace-anchored SLRs correctly show strong support for H_p relative to H_d.

The exact rates of misleading evidence in favor of the prosecution (RMEP) and in favor of the defense (RMED) are show in Tables 1 and 2, respectively. Both the RMEP and RMED are lowest for the trace-anchored SLRs. Tables 3 and 4 show the RMEP and RMED for each camera model. It is interesting to note that a particular model might have its highest RMEP under one SLR type while a different model might have its highest RMEP under a different SLR type. The RMEP is high for some camera models and work should be done to attempt to lower the RMEP. In particular, we propose researching how incorporating an *inconclusive zone* where $\log_{10}$(SLR) values close to 0 are considered inconclusive and using *close non-matches*, cameras of the same model, as the alternative device population might decrease the rates of misleading evidence (see Section 4 for more discussion of future work).

3.2 | Evaluating the strength of the evidence

SLRs allow us to quantify the strength of the evidential support for H_p or H_d. Figures 8 and 9 group all 138,240 general match, source-anchored, and trace-anchored $\log_{10}$(SLR) scores into intervals based on their values. Figure 8 shows $\log_{10}$(SLR) values from known matches and Figure 9 shows known non-matches. Each cell displays the percentage

of general match, source-anchored, or trace-anchored $\log_{10}$(SLR) values that fall into a particular interval. Almost 69% of general match SLRs for known matches correctly support H_p relative to H_d but the strength of the evidence is rather weak with these values between 0 and 2. On the other hand, 87.71% of the trace-anchored SLRs for known matches correctly show stronger support for H_p with $\log_{10}$(SLR) values greater than 100, while only 41.83% of source-anchored SLRs and 3.85% of general match SLRs show the same strength of support for H_p relative to H_d. In this respect, the trace-anchored SLRs perform better than the other two SLR types when H_p is true. It is worth noting that 61.5% of the trace-anchored SLRs for known matches are infinite because the denominator is zero. This is an artifact caused by KDE. All three SLR types perform well and similarly on known non-matches. For roughly 10% of the $\log_{10}$(SLR) values the evidence in favor of H_d is rather weak with the values falling between −2 and 0. Over 80% of the $\log_{10}$(SLR) values show stronger support for H_d relative to H_p with values less than or equal to −2.

4 | CONCLUSIONS

This paper presents a framework for calculating general match, source-anchored, and trace-anchored SLRs to not only address the prosecution's and defense's hypotheses in camera device identification problem, but also to provide a means of quantifying the strength of the evidence in favor of one hypothesis over the other. The dataset consists of digital images from 48 camera devices representing 26 distinct models. It includes digital still cameras, mobile phones, and one tablet. To the best of our knowledge, this is the first time camera device identification experiments using all three types of SLRs have been performed. General match SLRs performed poorly on the dataset, while both the source-anchored and trace-anchored SLRs performed well, with the trace-anchored SLRs performing the best of the three. If this dataset were to be used as the reference dataset for camera device identification SLRs in practice, trace-anchored SLRs would be recommended for use. One large caveat, however, is that it is unknown if similar results would occur on different datasets. Before applying SLRs in practice with a new dataset, it would be a good idea to replicate the study presented in this paper on the new dataset.

We only used RAW, center-cropped, auto-exposure, landscape-oriented images in our dataset. Future research should explore SLRs on a much wider variety of image data. Additionally, we only considered the *closed set* scenario where the questioned image's camera was present in the dataset. Future work should explore the *open set* scenario where the questioned image's camera is not in the dataset.

Previous work has shown that it is possible to distinguish between different models based on image artifacts created by the color filter array and the image processing pipeline [46]. We plan to leverage this in future work with SLRs where we restrict the reference dataset to *close non-matches* where the cameras in the dataset are the same model or brand as the POI's camera.

Several researchers [25,47] have employed an *inconclusive zone* where $\log_{10}$(SLR) values are considered inconclusive if $t_1 < \log_{10}(\text{SLR}) < t_2$ for a real-valued constants t_1 and t_2 where $t_1 < t_2$. These researchers considered $t_1 = -t$ and $t_2 = t$ for a constant $t > 0$. This type of inconclusive zone gives equal weight to both hypotheses. Future work could explore methods for choosing the best t_1 and t_2 to minimize the rates of misleading evidence. Additionally, a *defense biased inconclusive zone* with $t_1 = 0$ and $t_2 = t$ for a constant $t > 0$, which places a higher burden of proof on the prosecution while granting the benefit of the doubt to the defense could also be explored.

ACKNOWLEDGEMENTS

Open access funding provided by the Iowa State University Library.

ORCID

Stephanie Reinders https://orcid.org/0000-0002-9411-3043
Danica Ommen https://orcid.org/0000-0001-9955-3817

REFERENCES

1. National Research Council. Strengthening forensic science in the United States: a path forward. Washington, DC: The National Academies Press; 2009.
2. Daubert v. Merrell Dow Pharmaceuticals, Inc. 509 U.S. 579; 1993.
3. Stern HS. Statistical issues in forensic science. Annu Rev Stat Appl. 2017;4:225–44. https://doi.org/10.1146/annurev-statistics-041715-033554
4. Ommen DM, Saunders CP. A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio. Stat Sci. 2021;36(3):344–59. https://doi.org/10.1214/20-STS805
5. Lukas J, Fridrich J, Goljan M. Determining digital image origin using sensor imperfections. Proceedings SPIE 5685, image and video communications and processing; 2005 Jan 16–20; San Jose, CA. Bellingham, WA: SPIE; 2005. p. 249–60. https://doi.org/10.1117/12.587105
6. Chen M, Fridrich J, Goljan M, Lukas J. Determining image origin and integrity using sensor noise. IEEE Trans Inf Forensics Secur. 2008;3(1):74–90. https://doi.org/10.1109/TIFS.2007.916285
7. Goljan M, Fridrich J, Filler T. Large scale test of sensor fingerprint camera identification. Proceedings SPIE 7254, media forensics and security; 2009 Jan 19–21; San Jose, CA. Bellingham, WA: SPIE; 2009. p. 72540I. https://doi.org/10.1117/12.805701
8. Massimo L, Fontani M, Piva A. A leak in PRNU based source identification—questioning fingerprint uniqueness. IEEE Access. 2021;9:52455–63. https://doi.org/10.1109/ACCESS.2021.3070478
9. Lukas J, Fridrich J, Goljan M. Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensics Secur. 2006;1(2):205–14. https://doi.org/10.1109/TIFS.2006.873602
10. Chen M, Fridrich J, Goljan M. Digital imaging sensor identification (further study). Proceedings SPIE 6505, security, steganography, and watermarking of multimedia contents IX; 2007 Jan 29-Feb 1; San Jose, CA. Bellingham, WA: SPIE; 2007. p. 65050P. https://doi.org/10.1117/12.703370
11. Goljan M. Digital camera identification from images – estimating false acceptance probability. In: Kim HJ, Katzenbeisser S, ATS H, editors. Digital watermarking IWDW 2008. Leture notes in computer science, vol 5450; 2008 Nov 10–12; Busan, Korea.

Berlin/Heidelberg: Springer; 2008. p. 454–68. https://doi.org/10.1007/978-3-642-04438-0_38
12. Goljan M, Chen M, Fridrich J. Identifying common source digital camera from image pairs. Proceedings of the 2007 IEEE international conference on image processing; 2007 Sep 16–19; San Antonio, TX. Piscataway, NJ: IEEE; 2007. p. VI-125–8. https://doi.org/10.1109/ICIP.2007.4379537
13. Goljan M, Fridrich J. Camera identification from cropped and scaled images. Proceedings SPIE 6819, security, forensics, steganography, and watermarking of multimedia contents X. International Society for Optics and Photonics; 2008 Jan 28–30; San Jose, CA. Bellingham, WA: SPIE; 2008. p. 68190E. https://doi.org/10.1117/12.766732
14. Goljan M, Chen M, Pedro C, Fridrich J. Effect of compression on sensor-fingerprint based camera identification. Proceedings of the IS&T international symposium on electronic imaging science and technology; 2016 Feb 14–18; San Francisco, CA. Springfield, VA: Society for Imaging Science and Technology; 2016. p. 1–10. https://doi.org/10.2352/ISSN.2470-1173.2016.8.MWSF-086
15. Samaras S, Mygdalis V, Pitas I. Robustness in blind camera identification. Proceedings of the 23rd international conference on pattern recognition; 2016 Dec 4–8; Cancun, Mexico. Piscataway, NJ: IEEE; 2016. p. 3874–9. https://doi.org/10.1109/ICPR.2016.7900239
16. Goljan M, Fridrich J. Sensor-fingerprint based identification of images corrected for lens distortion. Proceedings SPIE 8303, media watermarking, security, and forensics; 2012 Aug 11–16; Burlingame, CA. Bellingham, WA: SPIE; 2012. p. 83030H. https://doi.org/10.1117/12.909659
17. Steele CD, Balding DJ. Statistical evaluation of forensic DNA profile evidence. Annu Rev Stat Appl. 2014;1(1):361–84. https://doi.org/10.1146/annurev-statistics-022513-115602
18. Lucy D, Aitken CGD. Evaluation of trace evidence in the form of multivariate data. J R Stat Soc Ser C Appl Stat. 2004;53(1):109–22. https://doi.org/10.1046/j.0035-9254.2003.05271.x
19. Ommen DM, Saunders CP, Neumann C. The characterization of Monte Carlo errors for the quantification of the value of forensic evidence. J Stat Comput Simul. 2017;87(8):1608–43. https://doi.org/10.1080/00949655.2017.1280036
20. Aitken C, Taroni F. Statistics and the evaluation of evidence for forensic scientists. Hoboken, NJ: John Wiley & Sons; 2004. p. 95–6.
21. Park S, Carriquiry A. Learning algorithms to evaluate forensic glass evidence. Ann Appl Stat. 2019;13(2):1068–102. https://doi.org/10.1214/18-AOAS1211
22. Park S. Learning algorithms for forensic science applications [dissertation]. Ames, IA: Iowa State University; 2018.
23. Qiao T, Retraint F. Identifying individual camera device from raw images. IEEE Access. 2018;6:78038–54. https://doi.org/10.1109/ACCESS.2018.2884710
24. Qiao T, Retraint F, Cogranne R, Thai TH. Individual camera device identification from JPEG images. Signal Process Image Commun. 2017;52:74–86. https://doi.org/10.1016/j.image.2016.12.011
25. Hepler AB, Saunders CP, Davis LJ, Buscaglia J. Score-based likelihood ratios for handwriting evidence. Forensic Sci Int. 2012;219(1–3):129–40. https://doi.org/10.1016/j.forsciint.2011.12.009
26. Bolck A, Weyermann C, Dujourdy L, Esseiva P, van der Berg J. Different likelihood ratio approaches to evaluate the strength of evidence of MDMA tablet comparisons. Forensic Sci Int. 2009;191(1–3):42–51. https://doi.org/10.1016/j.forsciint.2009.06.006
27. Egli NM, Champod C, Margot P. Evidence evaluation in fingerprint comparison and automated fingerprint identification systems – modelling within finger variability. Forensic Sci Int. 2007;167(2–3):189–95. https://doi.org/10.1016/j.forsciint.2006.06.054
28. Gonzalez-Rodriguez J, Drygajlo A, Ramos-Castro D, Garcia-Gomar M, Ortega-Garcia J. Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. Comput Speech Lang. 2006;20(2–3):331–55. https://doi.org/10.1016/j.csl.2005.08.005
29. Gonzalez-Rodriguez J, Fierrez-Aguilar J, Ramos-Castro D, Ortega-Garcia J. Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. Forensic Sci Int. 2005;155(2–3):126–40. https://doi.org/10.1016/j.forsciint.2004.11.007
30. Champod C, Evett IW, Jackson G. Establishing the most appropriate databases for addressing source level propositions. Sci Justice. 2004;44(3):153–64. https://doi.org/10.1016/s1355-0306(04)71708-6
31. Ommen DM, Saunders CP. Building a unified statistical framework for the forensic identification of source problems. Law Probab Risk. 2018;17(2):179–97. https://doi.org/10.1093/lpr/mgy008
32. Neumann C, Ausdemore MA. Defence against the modern arts: the curse of statistics 'score-based likelihood ratios'. arXiv preprint. arXiv:1910.05240 2019.
33. Garton N. Score-based likelihood ratios and sparse gaussian processes [dissertation]. Ames, IA: Iowa State University; 2020.
34. Nordgaard A, Höglund T. Assessment of approximate likelihood ratios from continuous distributions: a case study of digital camera identification. J Forensic Sci. 2011;56(2):390–402. https://doi.org/10.1111/j.1556-4029.2010.01665.x
35. van Houten W, Alberink I, Geradts Z. Implementation of the likelihood ratio framework for camera identification based on sensor noise patterns. Law Probab Risk. 2011;10(2):149–59. https://doi.org/10.1093/lpr/mgr006
36. Reinders S. Statistical methods for digital image forensics: algorithm mismatch for blind spatial steganalysis and score-based likelihood ratios for camera device identification [dissertation]. Ames, IA: Iowa State University; 2020.
37. Reinders S, Lin L, Chen W, Guan Y, Newman J. Score-based likelihood ratios for camera device identification. Proceedings of the IS&T international symposium on electronic imaging science and technology; 2020 Jan 26–30; Burlingame, CA. Springfield, VA: Society for Imaging Science and Technology; 2020. p. 215. https://doi.org/10.2352/ISSN.2470-1173.2020.4.MWSF-215
38. BOSSbase [cited 2022 Jan 7]. Available from: http://agents.fel.cvut.cz/stegodata/
39. Gloe T, Bohme R. The 'Dresden image Database' for benchmarking digital image forensics. Proceedings of the 2010 ACM symposium on applied computing; 2010 Mar 22–26; New York, NY. New York, NY: ACM; 2010. p. 1584–90. https://doi.org/10.1145/1774088.1774427
40. Center for Statistics and Applications in Forensic Evidence. StegoAppDB [cited 2022 Jan 7]. Available from: https://forensicstats.org/stegoappdb/
41. Newman J, Lin L, Chen W, Reinders S, Wang Y, Guan Y, et al. StegoAppDB: a steganography apps forensics image database. Proceedings of the IS&T international symposium on electronic imaging science and technology; 2019 Jan 13–17; Burlingame, CA. Springfield, VA: Society for Imaging Science and Technology; 2019. p. 536.1–12. https://doi.org/10.2352/ISSN.2470-1173.2019.5.MWSF-536
42. Digital Data Embedding Laboratory. Camera fingerprint – MATLAB & Python implementation [cited 2021 Nov 15]. Available from: http://dde.binghamton.edu/download/camera_fingerprint/
43. van Es A, Wiarda W, Hordijk M, Alberink I, Vergeer P. Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis. Sci Justice. 2017;57(3):181–92. https://doi.org/10.1016/j.scijus.2017.03.002
44. Mathworks. Kernel distribution [cited 2020 Jun 30]. Available from: https://www.mathworks.com/help/stats/kernel-distribution.html

45. Gill P, Curran J, Neumann C. Interpretation of complex DNA profiles using Tippett plots. Forensic Sci Int Genet Suppl Ser. 2008;1(1):646–8.
46. Filler T, Fridrich J, Goljan M. Using sensor pattern noise for camera model identification. Proceedings of the 15th IEEE international conference on image processing; 2008 oct 12–15; San Diego, CA. Piscataway, NJ: IEEE; 2008. p. 1296–9. https://doi.org/10.1109/ICIP.2008.4712000
47. Veneri FA, Ommen DM. An evaluation of score-based likelihood ratios for glass data [thesis]. Ames, IA: Iowa State University; 2021.

How to cite this article: Reinders S, Guan Y, Ommen D, Newman J. Source-anchored, trace-anchored, and general match score-based likelihood ratios for camera device identification. J Forensic Sci. 2022;67:975–988. https://doi.org/10.1111/1556-4029.14991