Pattern Recognition 74 (2018) 556–567

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Inference of a compact representation of sensor fingerprint for source camera identification

Ruizhe Li[a], Chang-Tsun Li[a,b,*], Yu Guan[c]

[a] Department of Computer Science, University of Warwick, UK
[b] School of Computing and Mathematics, Charles Sturt University, Australia
[c] Open Lab, School of Computing, Newcastle University, UK

ARTICLE INFO

Article history:
Received 30 November 2016
Revised 28 August 2017
Accepted 18 September 2017
Available online 27 September 2017

Keywords:
Image forensics
Source camera identification (SCI)
Sensor pattern noise (SPN)
PCA de-noising

ABSTRACT

Sensor pattern noise (SPN) is an inherent fingerprint of imaging devices, which provides an effective way for source camera identification (SCI). Although SPNs extracted from large image blocks usually yield high identification accuracy, their high dimensionality would incur a high computational cost in the matching stage, consequently hindering many applications that require efficient camera matchings. In this work, we employ and evaluate the concept of principal component analysis (PCA) de-noising in SCI tasks. Based on this concept, we present a framework that formulates a compact SPN representation. To enhance the de-noising effect, we introduce a training set construction procedure that minimizes the impact of various interfering artifacts, which is especially useful in some challenging cases, e.g., when only textured reference images are available. To further boost the SCI performance, a novel approach based on linear discriminant analysis (LDA) is adopted to extract more discriminant SPN features. To evaluate our methods, extensive experiments are conducted on the Dresden image database. The results indicate that the proposed framework can serve as an effective post-processing procedure, which not only boosts the performance, but also greatly reduces the computational cost in the matching phase.

1. Introduction

Nowadays, the use of digital images or videos as evidence in the fight against physical crime and cybercrime is a norm, which makes multimedia forensics crucial. Typically, multimedia forensics includes source camera verification and identification, source-oriented images classification, integrity verification, forgery detection, authentication, etc. Source camera identification, as an important branch of multimedia forensics, is about answering the question: *Which one of the many cameras has taken the image in question?* This is actually a task of matching the camera fingerprint of an image in question to a set of reference fingerprints, each representing a different camera. The size of the reference fingerprint set can be in the order of millions. How to deal with such a task more accurately and efficiently is the focus of this paper.

In order to link digital images to the source cameras, many techniques have been proposed in the last two decades. These techniques can be broadly divided into three categories. The simplest way is to use digital images' metadata that contains the information of the source camera [1].

However, due to the wide prevalence and great user-friendliness of multimedia processing tools nowadays (e.g., Adobe Photoshop and IrfanView), metadata can be easily changed or removed by laymen. Therefore, metadata is no longer regarded as reliable for authentication purposes. Another possible way is to use the digital watermark, which is a signature embedded in the image by a certain type of cameras [2]. This technique is useful in the cases of proving ownership of copyright. Yet it is only applicable to the cameras that have watermarking mechanism [2]. The third category of techniques rely on the intrinsic characteristics of digital cameras left in the captured images. Many traces left in the content by various hardware and software components in the image acquisition pipeline can be exploited to link the image to its source camera. Good examples are sensor pattern noise (SPN) [3–8], lens aberrations [9], color filter array (CFA) interpolation artefacts [10], JPEG compression [1], and the combination of several intrinsic characteristics [11]. Among these modalities, SPN has been proved to be the most effective camera fingerprint as it is capable of differentiating individual cameras of the same model.

Sensor pattern noise is produced by the imaging sensor and primarily caused by the manufacturing imperfections and the in-

* Corresponding author.
E-mail addresses: ruizhe.li@warwick.ac.uk (R. Li), c-t.li@warwick.ac.uk (C.-T. Li), yu.Guan@ncl.ac.uk (Y. Guan).

https://doi.org/10.1016/j.patcog.2017.09.027

homogeneity of silicon wafers. It is essentially the slight variations in the intensity of individual pixels. For instance, even if a sensor takes an image of an evenly lit scene, the resulting image will still exhibit slight changes in intensity between individual pixels [3]. Every image taken by the same sensor would exhibit the same SPN pattern, while two sensors, even made from the same silicon wafer, would exhibit uncorrelated patterns [3].

The dimensionality of SPN is as large as that of the original image. As a result, not only each SPN needs a fairly large amount of space for storage, but memory access would also take considerable amount of time. Moreover, SPN matching involves vector operations and the complexity is proportional to the size of SPNs. Thus, with a large number of reference SPN in the database to be matched, the complexity of matching process would become a critical concern.

In order to address the high complexity issue, many efforts [12–18] have been made in recent years. In [12], Bayram *et al.* embeded reference SPNs in a binary search tree, where the leaf/internal node represents a reference/composite SPN. Based on this structure, the total number of SPN matchings to be performed is substantially reduced. However, errors tend to increase significantly when a large number of reference SPNs are stored in a single binary tree. On the other hand, more methods reduce the computational complexity by compressing the SPN. In [13,14], the authors introduced a SPN digest technique for dimensionality reduction, which preserves the largest elements and their corresponding locations. In [15], Bayram *et al.* binarized SPN, which considerably reduces the storage requirements and speeds up loading of SPN into the memory. However, the binarization process inevitably degrades the matching accuracy due to information loss. In [16,17], Valsesia *et al.* reduced the dimensionality of SPN using random projection. However, since the subspace is randomly selected, the obtained representation is unlikely to be optimal and tends to compromise the matching accuracy.

To alleviate the common limitation (i.e., reduced accuracy) of the afore-mentioned SPN compression methods [13–17], in our previous work [19,20], we presented a feature extraction algorithm based on the concept of PCA de-noising [21,22], and promising results were achieved on a small dataset. However, this method is based on the assumption that the training set is well representative of the population so that an effective SPN feature extractor can be learned. Unfortunately, the noise residuals in the training set can be contaminated by many sources of interference, making the training set less representative. To learn a robust SPN feature extractor from the noisy training data, in this work, we further propose a training set construction procedure and provide its theoretical basis. We also provide more detailed discussion of the SPN feature extractors and treat it as a general post-processing framework on other SPN methods. It is evaluated in term of effectiveness and efficiency on a much larger dataset. We also test this framework on some challenging cases, e.g., all the reference SPNs are extracted from images with significant scene details (a form of distortion to the SPN), which are scenarios barely considered by previous works.

The rest of this paper is organized as follows. Section 2 provides a brief review on the three main steps of the SPN-based SCI system. In Section 3, we present the proposed training dataset construction procedure and the feature extraction method in details. In Section 4, the proposed source camera identification method is summarized, which is then followed by extensive experimental evaluations in Section 5. Section 6 concludes the work. Note that, in this manuscript, we use bold upper-case letters to represent matrices, and bold lower-case letters to denote vectors.

2. Background

In order to decide whether a query image is taken by one of the cameras in a large dataset, three main steps are required, i.e., SPN extraction, reference SPN estimation and SPN matching. In this section, techniques for these three steps are briefly reviewed.

2.1. SPN extraction

The most important step of the SPN-based SCI framework is to extract the SPNs from digital images. In [4], Chen *et al.* modeled the output of imaging sensor $\mathbf{I}$ and explained the general idea about how to extract SPN, such as

$$\mathbf{I} = (\mathbf{1}+\mathbf{K})\mathbf{I}^{(0)} + \boldsymbol{\Theta} = \mathbf{I}^{(0)} + \mathbf{I}^{(0)}\mathbf{K} + \boldsymbol{\Theta} \tag{1}$$

In Eq. (1), $\mathbf{I}^{(0)}$ is the noiseless sensor output and $\mathbf{I}^{(0)}\mathbf{K}$ represents the discriminative part of SPN, i.e., PRNU noise, which is a multiplicative noise and the signal of our interest. The matrix $\mathbf{K}$ is the PRNU multiplicative factor, where all the elements in it are typically close to 0. $\boldsymbol{\Theta}$ is a combination of random noise, such as shot noise, read-out noise, and quantization noise. In order to extract the signal of interest $\mathbf{I}^{(0)}\mathbf{K}$ from the observation $\mathbf{I}$, the host signal $\mathbf{I}^{(0)}$ should be removed. Generally, the noiseless image $\mathbf{I}^{(0)}$ is unknown, but we can estimate it by de-noising the observation $\mathbf{I}$, i.e., $\hat{\mathbf{I}}^{(0)} = F(\mathbf{I})$, where F indicates a de-noising algorithm and $\hat{\mathbf{I}}^{(0)}$ is an estimation of the noiseless image $\mathbf{I}^{(0)}$. Then, the signal of interest can be roughly extracted by subtracting the estimation $\hat{\mathbf{I}}^{(0)}$ from the observation $\mathbf{I}$, such as

$$\begin{aligned} \mathbf{X} &= \mathbf{I} - F(\mathbf{I}) = \mathbf{I} - \hat{\mathbf{I}}^{(0)} \\ &= \mathbf{I}^{(0)} + \mathbf{I}^{(0)}\mathbf{K} + \boldsymbol{\Theta} - \hat{\mathbf{I}}^{(0)} \\ &= \mathbf{IK} + \mathbf{I}^{(0)} - \hat{\mathbf{I}}^{(0)} + (\mathbf{I}^{(0)} - \mathbf{I})\mathbf{K} + \boldsymbol{\Theta} \\ &= \mathbf{IK} + \boldsymbol{\Xi} \end{aligned} \tag{2}$$

where $\mathbf{X}$ is the noise residual where the true SPN is present, $\boldsymbol{\Xi}$ is the sum of $\boldsymbol{\Theta}$ and two additional noise terms introduced by the de-noising filter.

From Eq. (2), one can see that the better a de-noising algorithm F is, the closer the de-noised version $\hat{\mathbf{I}}^{(0)}$ is to the noiseless image $\mathbf{I}^{(0)}$, and thus the less noise would be introduced by the de-noising filter and left in the final output $\mathbf{X}$. Therefore, the performance of a SPN extractor is primarily determined by the choice of the de-noising algorithm F. In [3], Lukas *et al.* proposed to transform the noisy image $\mathbf{I}$ into wavelet transform domain and apply the Mihcak filter [23] to extract the SPN components from the high frequency wavelet coefficients of $\mathbf{I}$. In [24], Chierchia *et al.* proposed to replace the Mihcak filter with a more recent technique, namely the sparse 3D transform-domain collaborative filtering [25]. In [26], Kang *et al.* proposed a SPN predictor based on context adaptive interpolation (PCAI), which is to apply the context adaptive interpolator [27] as the de-noising function F to predict the noiseless image $\mathbf{I}^{(0)}$ and extract SPN in the spatial domain.

Also demonstrated in Eq. (2) is the fact that the noise residual $\mathbf{X}$ contains not only the SPN term $\mathbf{IK}$ but also the noise term $\boldsymbol{\Xi}$. This leaves room for further enhancement. In [5], Li demonstrated that the noise residual contains the traces of scene details. Therefore, Li proposed 5 enhancing models to attenuate the impact of scene details. In [28], Li and Li proposed a color-decoupled SPN extraction method to prevent the color interpolation errors from propagating into the noise residual. In [29], Chen *et al.* proposed to suppress the JPEG blocky artifacts by transforming the noise residual into the discrete Fourier transform domain and suppressing the Fourier coefficients with extremely larger magnitude.