PRNU-Based Forgery Localization in a Blind Scenario

Davide Cozzolino, Francesco Marra, Giovanni Poggi, Carlo Sansone(✉), and Luisa Verdoliva

DIETI - University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy
{davide.cozzolino,francesco.marra,giovanni.poggi,
carlo.sansone,luisa.verdoliva}@unina.it

Abstract. The Photo Response Non-Uniformity (PRNU) noise can be regarded as a camera fingerprint and used, accordingly, for source identification, device attribution and forgery localization. To accomplish these tasks, the camera PRNU is typically assumed to be known in advance or reliably estimated. However, there is a growing interest for methods that can work in a real-word scenario, where these hypotheses do not hold anymore. In this paper we analyze a PRNU-based framework for forgery localization in a blind scenario. The framework comprises four main steps: PRNU-based blind image clustering, parameter estimation, device attribution, and forgery localization. Each of these steps impacts on the final outcome of the analysis. The aim of this paper is to assess the overall performance of the proposed framework and how it depends on the individual steps.

1 Introduction

The wide diffusion of powerful image editing tools has made image manipulation very easy. This impacts on many fields of life, and is especially dangerous in the forensic field, where images may be used as crucial evidence in court. Therefore, in the last decade, digital image forensics has grown tremendously, and new methodologies have been developed to track an image source and determine its integrity. In particular, the interest has focused on passive techniques, which detect traces of manipulations from the analysis of the image itself, with no need of collaboration on the part of the user. Some of these techniques rely on intrinsic camera properties, like sensor defects or lens aberration, while other rely on statistical features introduced both in-camera (e.g., demosaicking) and out-camera (e.g., JPEG compression) processing.

Some of the most successful camera-based methods rely on the Photo Response Non-Uniformity (PRNU), a sort of camera fingerprint contained in every image taken by a specific device. Its use was first proposed in [17], both for source identification and forgery localization. In this work we focus on PRNU-based methods for forgery detection and localization.

To extract the PRNU pattern, the high-level image content is removed through some sort of high-pass filtering, obtaining the so-called image residual.

S. Battiato et al. (Eds.): ICIAP 2017, Part II, LNCS 10485, pp. 569–579, 2017.
https://doi.org/10.1007/978-3-319-68548-9_52

Even in the residual, however, the PRNU pattern represents a weak signal overwhelmed by intense noise, both of random origin, and deriving from imperfect image content removal. This makes all PRNU-based analyses quite challenging, to begin with the very same camera PRNU estimation. Several approaches have been proposed in the literature to reduce the influence of the scene content on PRNU estimation. In [15] some forms of enhancement are considered, while in [7] the use of a nonlocal denoising filter has been shown to reduce the scene content in the residual image. A systematic analysis of post-processing methods aimed at improving PRNU estimation has been recently presented in [1]. With reference to the forgery localization task, several improvements have been proposed with respect to the basic method of [17]. In [5] a predictor is estimated which locally adapts the statistical decision test by taking into account image features, such as texture, flatness and intensity, thus reducing the probability of false alarms. In [9], instead, the problem is recast in terms of Bayesian estimation, using a Markov random field (MRF) prior to model the strong spatial dependencies and take decisions jointly rather than individually for each pixel. In [6,8] the problem of small forgery detection is addressed, resorting to image segmentation and guided filtering to improve the decision statistics. Further improvements have been recently proposed by considering the use of discriminative random fields [4] or by introducing multiscale analysis [14].

All these methods rely on the assumption that a large number of images are available, which are known to come from the camera of interest. However, such an hypothesis is not reasonable in a real-world scenario. Therefore, in this paper we propose and analyze a framework for image forgery localization in a blind scenario [10]. We only assume to have a certain number of images, whose origin, however, is unknown. Then we estimate one or more PRNU's by means of a blind source clustering algorithm and use them to establish the integrity of the image under test.

In the following Section we describe the PRNU-based framework for blind forgery localization, while in Sect. 3 present experimental results with reference to various clustering approaches [2,3,18], in order to assess the overall performance of the proposed framework and how it depends on the individual steps. Finally, in Sect. 4 we draw some conclusions.

2 Camera-Based Forgery Localization Framework

In both camera identification and forgery localization tasks, the PRNU of the camera of interest is given in advance, or is accurately estimated from a large number of images coming from the camera. However, in many forensic scenarios, and especially in investigation, no information is available on the origin of the images under analysis, neither the probe nor the dataset. Often, however, it is reasonable to assume that the images in the dataset come from just a few different devices. With this assumption, we can pursue PRNU-based forgery localization in a blind scenario, following the framework shown in Fig. 1 and already outlined in [10].

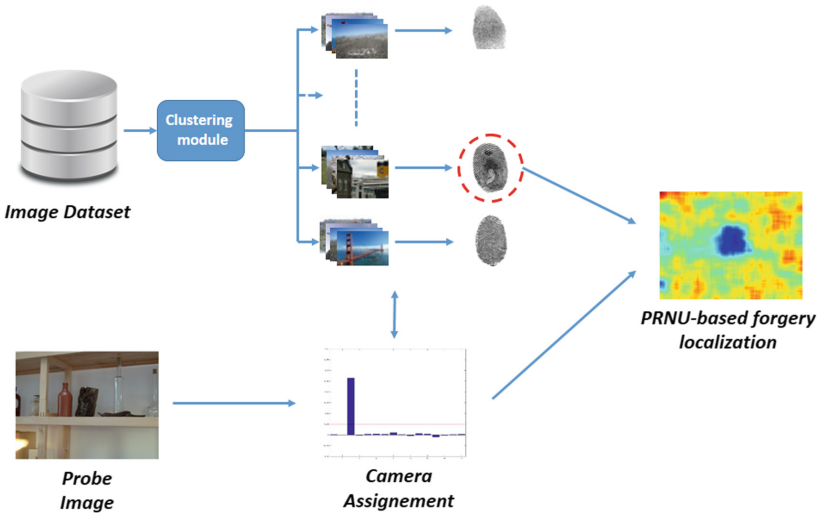

Fig. 1. A framework for PRNU-based forgery localization in a blind scenario.

The considered framework consists of four steps:

1. Residual-based image clustering
2. Cluster PRNU estimation
3. Camera assignment
4. Forgery localization.

The first two steps allow us to group together images coming from the same camera and to estimate their PRNU. Then, in step 3, the test image is associated with one of the clusters (or possibly none) by a PRNU-based correlation test. Finally, the tampered area of the test image is localized by detecting the absence of the selected PRNU. These steps are described in more detail in the following.

2.1 Residual-Based Image Clustering

To perform PRNU-based forgery localization one needs the true PRNU of the camera. Otherwise, it can be estimated by averaging a large number of images taken by the camera of interest. To this end, the first step of the proposed framework aims at grouping together all images of the dataset coming from the same camera. Since these share the same PRNU, they will exhibit a larger correlation than images coming from different cameras. However, before computing correlations, the high-level content of the images, which represents an interference in this context, is removed by high-pass filtering, obtaining the so-called noise residuals.

Let $\mathcal{R} = R_1, R_2, \ldots, R_N$ be the set of all noise residuals in the dataset. We want to partition this set in M distinct clusters, where the number of clusters is not know a priori. Therefore, the output of this step is a partition, P, of the dataset, namely:

$$P = \{C_1, C_2, ..., C_M\} \qquad C_i \cap C_j = \emptyset \; \forall i \neq j, \qquad \bigcup_{i=1}^{M} C_i = \mathcal{R} \tag{1}$$

In the literature, a number of PRNU-based clustering methods have been recently proposed [2,3,10,16,18], some of which will be considered in the experiments. Ideally, we would like to obtain as many clusters as are the source devices in the dataset, $M = M_t$, with M_t the number of devices, and all of them "pure", namely consisting only of images taken by the same device. In practice, the estimated number of clusters may differ from the number of cameras and, even when they coincide, the clusters may not be pure, comprising images coming from different sources. In all cases, the effect is a loss of accuracy in PRNU estimation. When under-partitioning occurs, $M < M_t$, clusters are necessarily "impure", comprising also images coming from other cameras which act as additional noise in the estimation. In case of over-partitioning, $M > M_t$, even pure clusters may comprise only a fraction of all images taken by a camera, leading to a less reliable estimate. The aforementioned effects may both show up in the same clustering experiment. Of course, all deviations from perfect clustering tend to cause a performance loss.

2.2 Camera Fingerprint Estimation

In the second step, each cluster is treated as "pure", and used to estimate both the PRNU and the predictor needed in the localization phase [5].

Given N_m images in the m-th cluster, one can perform a maximum-likelihood (ML) estimate of the PRNU as [5]

$$\widehat{K}_m = \sum_{i=1}^{N_m} \left[\frac{I_i}{\sum_{i=1}^{N^m} I_i^2} \right] R_i \tag{2}$$

In alternative, one can use the simpler sample average

$$\widehat{K}_m = \frac{1}{N_m} \sum_{i=1}^{N_m} R_i \tag{3}$$

which ensures very close performance to the ML case, provided N_m is large enough. On the other hand, when the cluster is too small, both estimates become quite unreliable because the noise residuals, R_i, have a very small signal component overwhelmed by noise. Whatever the estimator, some suitable steps follow to remove non-unique artifacts originated by other camera processes.

Some clustering methods tend to generate a large number of small clusters, even singletons, besides a few large ones. It makes sense to discard such small clusters, due to the ensuing unreliable estimates. Therefore, we introduce a parameter, $N_{\min}$, left to the analyst to set, such that all clusters with $N_m < N_{\min}$ are automatically discarded, avoiding their involvement in the forgery localization process.

Besides the PRNU itself, the localization algorithm proposed in [9] needs a predictor, which establishes the expected value of the correlation for a pristine image. Therefore, for each cluster, we must also estimate the predictor parameters, say Θ_m. To this end, the cluster must be further divided in two subsets, $C_m = C'_m \cup C''_m$. The first one, C'_m, is used to compute an *internal* PRNU, to which images of the second set, C''_m, are correlated. The parameters of the predictor, Θ_m, are then designed to minimize the error between the predicted and observed values of the correlation. Clearly, this further partition of the cluster further stresses the need for it to be large enough. To reduce this problem, we split clusters exactly in half for this task. Note, however, that the final estimate of the cluster PRNU can be carried out from the whole set. Indeed, the test image is completely alien to the cluster, and hence there is no reason to penalize the estimation of the PRNU.

In conclusion, the output of this step is the set of estimated PRNUs and predictor parameters, $\{\widehat{K}_m, \Theta_m, m = 1, \ldots, M\}$.

2.3 Camera Assignment

In this step we try to establish whether the probe image, I_p, is compatible with one of the estimated PRNU's, and which one. This decision is based on the normalized correlation[1]

$$\rho_m = \mathrm{corr}(R_p, I_p \times \widehat{K}_m) \tag{4}$$

between the high-pass image residual, R_p, and each of the scaled fingerprints.

The probe image is assumed to come from the camera with the most correlated PRNU

$$\widehat{K}_{\max} = \arg\max_m \mathrm{corr}(R_p, I_p \times \widehat{K}_m) \tag{5}$$

which is therefore selected to perform forgery localization. However, it is also possible that none of the cameras under analysis originated the probe image, in which case all correlations should be small. To formalize this problem, let us consider the two hypotheses

H_0 : the probe image is alien to the dataset
H_1 : the probe image comes from one of the dataset cameras

To design a statistical test we should know the distribution of ρ under both hypotheses. This is not possible in our blind scenario, therefore we resort to a Neyman-Pearson test, selecting a decision threshold, t, which guarantees a suitably small false alarm probability P_{FA}. Following [13], we assume the normalized correlations to have a Gaussian distribution under H_0

$$\rho \sim N(0, 1/HW) \tag{6}$$

[1] Here, and throughout this work, we assume the images to be perfectly aligned. Otherwise, one can replace normalized correlation with Peak-to-Correlation Energy (PCE) ratio [12], which works correctly also in the presence of image cropping.

where H and W are the image dimensions. Therefore

$$P_{FA} = \Pr(\rho_{\max} > t|H_0) = 1 - (1 - \Pr(\rho_m \leq t|H_0))^M$$
$$= 1 - (1 - Q(t\sqrt{HW}))^M \simeq MQ(t\sqrt{HW}) \quad (7)$$

with the latter approximation holding for small M and $Q(t\sqrt{HW}) \ll 1$. By inverting the above relation the desired threshold is obtained.

2.4 PRNU-Based Forgery Localization

In the last step of the framework, a PRNU-based forgery localization technique is applied. Several such methods have been proposed in the last few years, and they all share the same basic idea. When the image is tampered with, for example through the splicing of some alien material, its PRNU is locally removed. Therefore, a sliding-window correlation test is performed, and when the local correlation index falls below a given threshold, a forgery is declared. Since the correlation may also depend on the image content, the threshold must be adapted locally by using the predictor with parameters $\Theta_{\max}$ estimated in step 2.

The output of this localization step is a binary decision mask that highlights the pixels that are considered as tampered. Given such a mask, and the corresponding ground truth mask, one can compute a number of performance indicators. However, it is worth pointing out that the output mask should be always analyzed by a human interpreter. In fact, real-life image forgeries are performed with a purpose, and they possess a semantics that is not easily captured by algorithms. The localization mask should be therefore regarded as a diagnostic tool to support the expert decision.

3 Experimental Results

In this section we evaluate the performance of the proposed PRNU-based framework for blind forgery localization. Experiments are carried out on six cameras: Canon EOS-10D, Canon EOS-450D, Canon Ixus 95IS, Nikon D200, Nikon Coolpix S5100, Sony DSC S780. For each camera we use 50 images as training set to perform the PRNU-based clustering and to estimate the cluster PRNUs. Performance is assessed on 50 more images per camera, different from those of the training set. All images have the same size of 768×1024 pixels, and are cropped from the same region of the full-size images. To study forgery localization, we generate forged versions of the test images by pasting on them, at the center, a square region of 128×128 or 256×256 sampled randomly from another image. In addition, we repeat the experiments using JPEG compressed images with a quality factor of 90. All the noise residuals are extracted by using the BM3D denoising filter [11], and removing non-unique artifacts caused by demosaicing and lens distortions as proposed in [5].

Localization results are given in terms of ROC curves, giving pixel-wise probability of detection, P_D, and probability of false alarm, P_{FA}, as a function of

the decision threshold. As a synthetic measure, the area under the ROC curve (AUC) is also computed. Before considering localization, however, we study the performance of previous steps, to understand their impact on the accuracy.

3.1 Image Clustering and PRNU Estimation

We implemented three clustering algorithms, based on Normalized Cuts (NCut) [2], on pairwise nearest neighbor (PCE-PNN) clustering [3,10], and on correlation clustering [18], denoted as Marra2017. Note that NCut requires a threshold parameter to be estimated on a training set, so we consider here an *oracle* version, selecting *a posteriori* the best parameter. For PCE-PNN we used the threshold used by the authors in the original paper. Other PRNU-based clustering methods [3,16] are not considered here because they have been shown in [2,18] to provide a generally worse performance.

Table 1. Performance of clustering algorithms.

	NCut-oracle			PCE-PNN			Marra2017		
Set	ARI	TPR	FPR	ARI	TPR	FPR	ARI	TPR	FPR
Original	0.872	84.31	1.21	0.839	75.74	0.00	0.960	94.79	0.26
JPEG (QF = 90)	0.647	61.07	2.77	0.819	79.02	1.50	0.921	93.58	1.33

Table 1 shows results of clustering algorithms on both original and JPEG compressed images in terms of adjusted rand index (ARI), true positive rate (TPR) and false positive rate (FPR). Marra2017 provides clearly the best results, even better than the oracle version of NCut, with ARI always very close to 1 (perfect clustering).

In Fig. 2 we show a graphical representation of the results. For uncompressed images (left) Marra2017 provides near-perfect results, with just a few extra clusters for the Sony camera, removed because too small ($N_m < N_{\min}$). In this condition, almost all available images can be used to estimate the PRNU's. The other methods show a higher fragmentation, but clusters are large and pure enough to provide good estimations. Using JPEG compressed images, performance impairs for all methods, but only slightly for Marra2017. On the contrary PCE-PNN and NCut-oracle suffer more on this dataset, especially for the Nikon D200 images, that will not allow a good PRNU estimate.

3.2 Image to Cluster Assignment

After clustering the images and estimating the cluster fingerprints, the probe image is correlated with all PRNU's. If the maximum correlation exceeds the decision threshold, t, forgery localization is performed. Together with the 600 test images coming from the selected cameras, we use 600 (negative) images

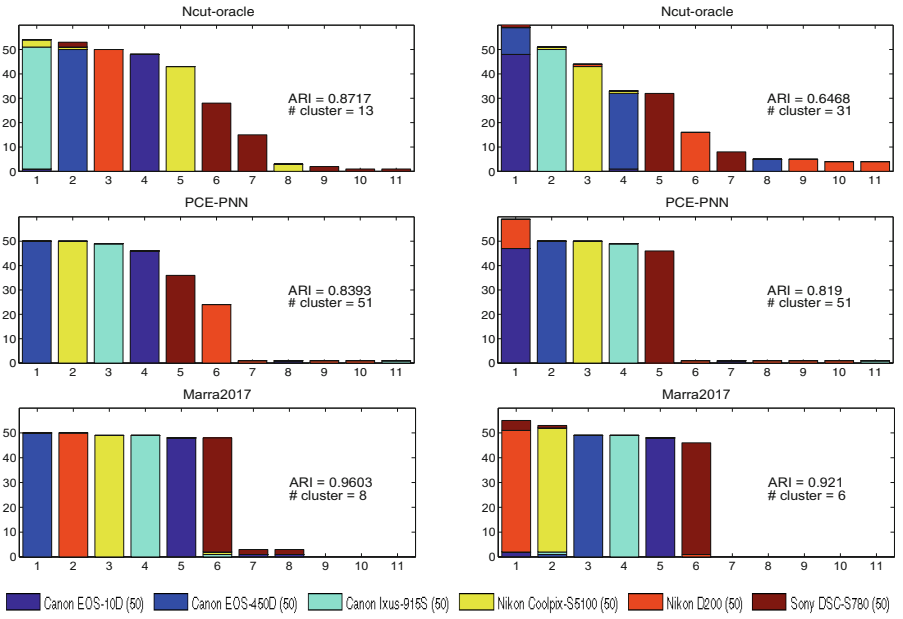

Fig. 2. Clustering results on original images (left) and JPEG compressed images (right) for NCut-oracle, PNN-PCE and Marra2017. Colors refer to the devices (see legend) while bar height indicate number of images in a cluster. (Color figure online)

taken from other sources, and cropped to the same size. Table 2 shows the detection performance for a threshold, t, set so as to obtain a theoretical false alarm probability $P_{FA} = 10^{-3}$. In detail, the FPR is the fraction of negative images that pass the test, while the TPR is the fraction of positive images (taken by one of the cameras in the dataset) recognized as such. The FPR is always very small, compatible with the theoretical level. The TPR is also quite large, but almost 6% of the positives are rejected, a fraction that grows above 10% with JPEG compressed images (almost 20% for PCE-PNN). Considering that Marra2017 provides near-perfect clustering, these errors must be attributed to the intrinsic problems of PRNU estimation. After correct detection, we could still

Table 2. Detection performance on original and JPEG compressed images.

	Original		JPEG (Qf = 90)	
Set	TPR	FPR	TPR	FPR
NCut-oracle	94.3%	0%	89.2%	0%
PCE-PNN	94.0%	0.3%	81.0%	0.7%
Marra2017	93.9%	0%	89.6%	1.5%

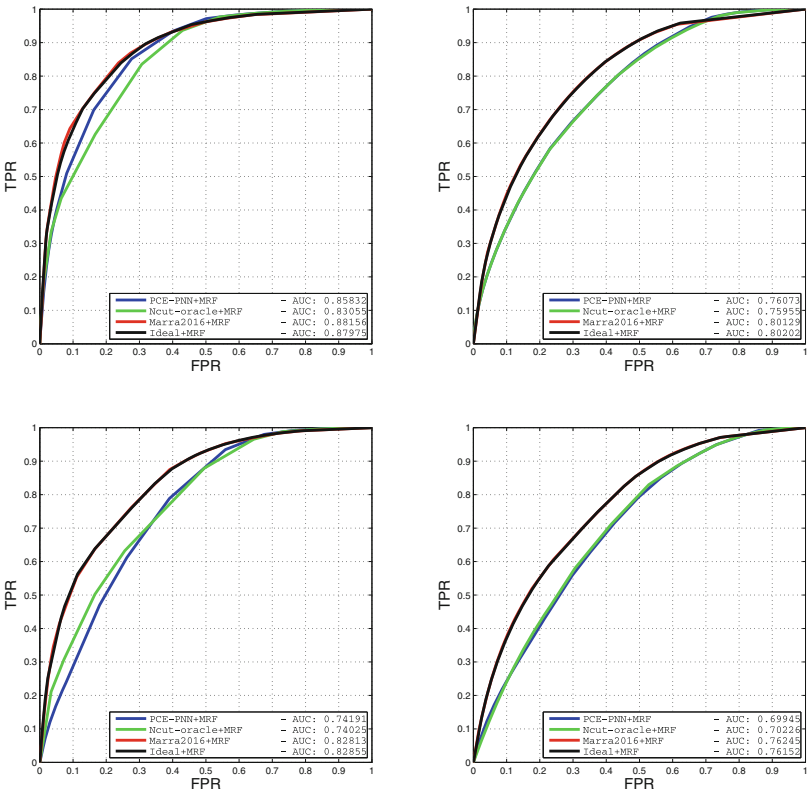

Fig. 3. Forgery localization results on original (top) and JPEG compressed images (down) with forgeries of 256 × 256 (left), and 128 × 128 pixels (right).

have a wrong assignment, that is, the probe image could be associated with a wrong camera/PRNU. However, our experiments show this event to be extremely unlikely, with probabilities lower than 0.1% in all cases and not reported in detail for the sake of brevity.

3.3 Forgery Localization

We conclude this analysis by studying forgery localization performance. Localization is carried out by the algorithm proposed in [9], based on a MRF prior and on the predictor of [5]. Together with the ideal case where the PRNU's are estimated from all available images, the case of real-world imperfect clustering is also considered, with all methods discussed before.

Figure 3 shows the ROC curves for original (top) and JPEG compressed images (down) with the two different forgery sizes. With large forgeries on uncompressed images results are very good. The AUC's are close to 0.9 with both ideal and Marra2017 clustering, and only slightly smaller with the other

clustering methods. Surprisingly, Marra2017 provides even a small improvement with respect to ideal clustering, maybe because the discarded images are outliers that impact negatively on the PRNU estimation. As expected, all results impair somewhat when considering smaller forgeries and JPEG compressed images. However, the performance obtained with blind clustering keep being very close (equal for Marra2017) to those of ideal clustering.

Finally, we assess the performance when we renounce clustering altogether, computing a single PRNU estimated by averaging all images in the dataset. This "naive" approach makes sense, since the estimated PRNU will bear traces of all camera fingerprints, although attenuated due to the large number of unrelated images averaged together. Figure 4 shows a significant performance drop with respect to the best clustering-based solution, both with original and JPEG compressed images (only 256×256 pixel forgeries, for brevity) which fully supports our findings.

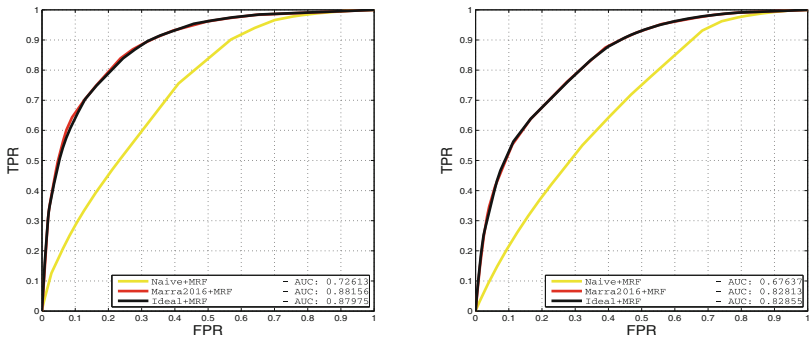

Fig. 4. Results for clustering-based and "naive" solutions on original (left) and JPEG compressed images (right) with 256×256 pixel forgeries.

4 Conclusion

In this paper we analyze a Camera-based framework for forgery localization in a blind scenario using a controlled dataset. The framework is composed of different steps, each of which is a possible source of error. The aim of our experiment is to show the performance of each single step due to the errors of the previous steps. As we see, for the original images the performance of all clustering algorithm are high enough to create cluster with a low FPR that assure to estimate pure PRNUs (all the images coming from the same camera device). This allow the forgery detector to perform as well as in the ideal case. In the JPEG compressed dataset, we note a performance drop when the clustering become less accurate and fragmented. The comparison with the *naive* solution say us that the effort in having a good clustering algorithm and pure PRNUs estimation is not pointless.

References

1. Al-Ani, M., Khelifi, F.: On the SPN estimation in image forensics: a systematic empirical evaluation. IEEE Trans. Inf. Forensics Secur. **12**(5), 1067–1081 (2017)
2. Amerini, I., Caldelli, R., Crescenzi, P., Mastio, A.D., Marino, A.: Blind image clustering based on the normalized cuts criterion for camera identification. Sig. Process. Image Commun. **29**(8), 831–843 (2014)
3. Bloy, G.J.: Blind camera fingerprinting and image clustering. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 532–534 (2008)
4. Chakraborty, S., Kirchner, M.: PRNU-based forgery detection with discriminative random fields. In: International Symposium on Electronic Imaging: Media Watermarking, Security, and Forensics, February 2017
5. Chen, M., Fridrich, J., Goljan, M., Lukás, J.: Determining image origin and integrity using sensor noise. IEEE Trans. Inf. Forensics Secur. **3**(1), 74–90 (2008)
6. Chierchia, G., Cozzolino, D., Poggi, G., Sansone, C., Verdoliva, L.: Guided filtering for PRNU-based localization of small-size image forgeries. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6231–6235. May 2014
7. Chierchia, G., Parrilli, S., Poggi, G., Sansone, C., Verdoliva, L.: On the influence of denoising in PRNU based forgery detection. In: 2nd ACM workshop on Multimedia in Forensics, Security and Intelligence, pp. 117–122 (2010)
8. Chierchia, G., Parrilli, S., Poggi, G., Sansone, C., Verdoliva, L.: PRNU-based detection of small size image forgeries. In: International Conference on Digital Signal Processing, pp. 1–6. July 2011
9. Chierchia, G., Poggi, G., Sansone, C., Verdoliva, L.: A Bayesian-MRF approach for PRNU-based image forgery detection. IEEE Trans. Inf. Forensics Secur. **9**(4), 554–567 (2014)
10. Cozzolino, D., Gragnaniello, D., Verdoliva, L.: Image forgery localization through the fusion of camera-based, feature-based and pixel-based techniques. In: IEEE Conference on Image Processing, pp. 5237–5241. October 2014
11. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
12. Goljan, M.: Digital camera identification from images – estimating false acceptance probability. In: Kim, H.-J., Katzenbeisser, S., Ho, A.T.S. (eds.) IWDW 2008. LNCS, vol. 5450, pp. 454–468. Springer, Heidelberg (2009). doi:10.1007/978-3-642-04438-0_38
13. Goljan, M., Fridrich, J., Filler, T.: Large scale test of sensor fingerprint camera identification. In: Proceedings SPIE, vol. 7254, pp. 72540I–72540I-12 (2009)
14. Korus, P., Huang, J.: Multi-scale analysis strategies in PRNU-based tampering localization. IEEE Trans. Inf. Forensics Secur. **12**(4), 809–824 (2017)
15. Li, C.T.: Source camera identification using enhanced sensor pattern noise. IEEE Trans. Inf. Forensics Secur. **5**(2), 280–287 (2010)
16. Li, C.T.: Unsupervised classification of digital images using enhanced sensor pattern noise. In: IEEE Int. Symp. on Circuits and Systems, pp. 3429–3432 (2010)
17. Lukàš, J., Fridrich, J., Goljan, M.: Digital camera identification from sensor pattern noise. IEEE Trans. Inf. Forensics Secur. **1**(2), 205–214 (2006)
18. Marra, F., Poggi, G., Sansone, C., Verdoliva, L.: Blind PRNU-based image clustering for source identification. IEEE Trans. Inf. Forensics Secur. **12**(9), 2197–2211 (2017)