GUIDED FILTERING FOR PRNU-BASED LOCALIZATION OF SMALL-SIZE IMAGE FORGERIES

Giovanni Chierchia[†], *Davide Cozzolino*[⋆], *Giovanni Poggi*[⋆], *Carlo Sansone*[⋆], *Luisa Verdoliva*[⋆]

[⋆] Universitá Federico II di Napoli, DIETI, 80125 Naples Italy
[†] Télécom ParisTech/Institut Télécom, LTCI, 75014 Paris France

ABSTRACT

PRNU-based techniques guarantee a good forgery detection performance irrespective of the specific type of forgery. The presence or absence of the camera PRNU pattern is detected by a correlation test. Given the very low power of the PRNU signal, however, the correlation must be averaged over a pretty large window, reducing the algorithm's ability to reveal small forgeries. To improve resolution, we estimate correlation with a spatially adaptive filtering technique, with weights computed over a suitable pilot image. Implementation efficiency is achieved by resorting to the recently proposed guided filters. Experiments prove that the proposed filtering strategy allows for a much better detection performance in the case of small forgeries.

Index Terms— Digital forensics, forgery detection, photo response non-uniformity, guided filters.

1. INTRODUCTION

Image forgery detection and localization is a very challenging task due to the large variety of manipulations a malicious user can perform by means of more and more sophisticated image editing tools [1]. In recent years, research has focused especially on passive techniques which retrieve traces of manipulations from the sole analysis of the image content. The image acquisition phase, in particular, is a valuable source of information as it often leaves peculiar traces, related to characteristics of the lens [2, 3], the color filter array (CFA) pattern [4, 5], or the sensor array [6, 7]. Indeed, one of the most promising approaches to date relies on the photo response non-uniformity (PRNU) noise. The PRNU arises from tiny imperfections in the silicon wafer used to manufacture the imaging sensor [8]. These physical differences generate a unique sensor pattern, specific of each individual camera, constant in time, independent of the scene, and fairly robust to several forms of image processing. Therefore, this pattern can be considered as a sort of camera fingerprint and used as such to accomplish forgery detection or image identification tasks. Different types of tampering, like copy-move, splicing, retouching, all remove the original PRNU from the target area, enabling the detection of the forgery irrespective of the type of attack.

An intense research activity began as soon as the potential of this approach was recognized. In the first PRNU-based technique, proposed in [6] in 2006, blocks extracted from the estimated PRNU of the target image are compared with homologous blocks of the camera PRNU (estimated in advance from a set of sample images) and a tampering is declared whenever the normalized correlation falls below a given threshold. However, since the PRNU is a very weak signal, estimated by means of imperfect tools, its traces can be easily overwhelmed by noise in some regions of the image characterized by saturation or strong textures, leading to false alarms. Therefore, the authors of [6] proposed themselves a new version in [7] to reduce the false alarms by identifying the potentially troublesome regions (through a predictor) and declaring them as genuine irrespective of the observed correlation index. Similar considerations guide the algorithm proposed in [9], where only regions with high signal quality are used, discarding those heavily deteriorated by irrelevant noise. In [10] a strategy to reduce the interference of scene details on the PRNU is proposed, while in [11, 12, 13, 14] the suppression of non-unique artifacts is considered. In [15], canonical correlation analysis is used to increase the reliability of the decision variables. We ourselves proposed several improvements to the basic algorithm of [6, 7] concerning a better method for PRNU estimation based on nonlocal filtering [16], the adoption of a variable-size analysis window to improve resolution [17] and, more recently, the reformulation of PRNU-based forgery detection as a Bayesian estimation problem [18, 19].

This work, following the path initiated in [17], aims at improving the resolution of PRNU-based algorithms. In fact, since the PRNU pattern is a very weak signal, it can be reliably detected only by jointly processing a large number of image samples, through a sliding-window analysis. The size of the sliding-window dictates therefore the effective resolution of the algorithm, causing forgeries smaller than the analysis window to remain often undetected. In [17] we resorted to a preliminary image segmentation to adapt the analysis window to the shape of candidate forgeries. Segmentation, however, is itself a source of errors, and the experimental analysis proved the heavy impact of such errors on performance. Here, we replace hard segmentation with a more flexible soft-segmentation strategy, using adaptive weights in the analysis window, computed on the basis of image content. A fast and effective implementation of this concept is obtained by resorting to guided filters [20]. Experiments prove the proposed algorithm to provide much better results on critical small-size forgeries, with a negligible increase in complexity. In the following, Section II provides the necessary background material, Section III describes the proposed algorithm and Section IV analyzes its performance by numerical experiments.

2. BACKGROUND

Let $y \in \mathbb{R}^N$ be a digital image observed at the camera output, where y_i indicates the value at site i, either as a single color band or the composition of multiple color bands. Let us assume, in a simplified model [7, 8], that y can be written as

$$y_i = (1 + k_i)x_i + \theta_i = x_i k_i + x_i + \theta_i \qquad (1)$$

where x is the ideal noise-free image, k the camera PRNU, and θ an additive noise term which accounts for all types of disturbances.

The PRNU k is by now our signal of interest, very weak w.r.t. both additive noise θ and ideal image x. To increase the signal-to-noise ratio, we subtract from (1) an estimate of the ideal image, $\widehat{x} = f(y)$, obtained by means of a denoising algorithm, obtaining the so-called noise residual

$$\begin{aligned} r_i &= y_i - \widehat{x}_i = y_i k_i + (x_i - y_i)k_i + (x_i - \widehat{x}_i) + \theta_i \\ &= y_i k_i + n_i \end{aligned} \tag{2}$$

where, for convenience, k multiplies the observed image y rather than the unknown original x. and all disturbances have been collected in a single noise term n.

When a section of the image is tampered with, for example by replacing it with material drawn from other regions, the PRNU term is cancelled. Therefore, to decide about a possible forgery, PRNU-based techniques try to discover whether the PRNU term is present or not. In the following we briefly describe the technique proposed by Chen *et al.* [7], based on sliding-window analysis, referring the reader to the original paper for more detail.

As a preliminary step, the true camera PRNU pattern, k, must be reliably estimated, which requires that either the target camera, or a large enough number of photos taken by it, are available. Note that such an hypothesis is not always satisfied in practice, representing the main limitation of this approach. Given k, the detection is formulated as a binary test between hypothesis H_0 that the camera PRNU is absent (i.e. the pixel has been tampered with) and hypothesis H_1 that the PRNU is present (i.e. the pixel is genuine):

$$\begin{cases} H_0: & r_i = n_i \\ H_1: & r_i = z_i + n_i \end{cases} \tag{3}$$

with $z_i = y_i k_i$. The decision is based on the normalized correlation between r_{W_i} and z_{W_i}, namely, the restrictions of r and z, respectively, to a window W_i centered on the target pixel:

$$\rho_i = \mathrm{corr}(r_{W_i}, z_{W_i}) = \frac{(r_{W_i} - \overline{r}_{W_i}) \odot (z_{W_i} - \overline{z}_{W_i})}{\|r_{W_i} - \overline{r}_{W_i}\| \cdot \|z_{W_i} - \overline{z}_{W_i}\|} \tag{4}$$

where $\odot$ denotes inner product, and $\overline{x}$ indicates mean of x. The algorithm then compares the correlation with a threshold γ_1

$$\widehat{u}_i = \begin{cases} 0 & \rho_i < \gamma_1 \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

where $\widehat{u}_i \in \{0, 1\}$ is the algorithm output, 0 for forgery and 1 for genuine pixel. The threshold is selected according to the Neyman-Pearson criterion so as to guarantee a suitably small false acceptance rate (FAR) $\Pr(\widehat{u}_i = 1 \mid u_i = 0)$, with $u_i \in \{0, 1\}$ the true pixel class. Once fixed the FAR, however, there is no guarantee that the other type of error, the false rejection rate (FRR), remain reasonably small. In fact, under hypothesis H_1, the decision statistic is influenced by the image content. Even in the absence of forgery, the correlation might happen to be very low when the image is dark, saturated, or heavily textured. In [7] this problem is addresses by means of a predictor which, based on local images features, computes the expected value $\widehat{\rho}_i$ of the correlation index under hypothesis H_1. When $\widehat{\rho}_i$ is too low, indicating that no reliable decision can be made, the pixel is always labeled as genuine, the less risky decision, irrespective of the value of ρ_i. Therefore, the test becomes

$$\widehat{u}_i = \begin{cases} 0 & \rho_i < \gamma_1 \ \text{ AND } \ \widehat{\rho}_i > \gamma_2 \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

with γ_2 chosen heuristically by the user. Better strategies are considered in [18] and [19] where decisions are made jointly on all pixels based on a Bayesian/MRF modeling.

Although the above description remains necessarily at a conceptual level, it is worth going into some more detail for what concerns the decision statistic of equation (4). Given the low, and spatially varying, signal-to-noise ratio characterizing this problem, the two conditional pdf's $p_{\rho|H_0}(\cdot)$ and $p_{\rho|H_1}(\cdot)$ can overlap significantly, causing large probabilities of error. To obtain a reasonable separation between them, one is forced to compute the correlation over a large window, for example, 128×128 pixels, as done in [7]. By so doing, however, one is implicitly renouncing to detect forgeries much smaller than the window size (or just much thinner). In these cases, in fact, the analysis window comprises pixels of both types, forged and genuine, providing a highly unreliable decision statistic. In the original algorithm, in fact, detected forged regions smaller than 64×64 pixels (one fourth of the window size) are canceled right away, as they are more easily generated by random errors than by actual forgeries. Low resolution is therefore a major problem of this algorithm.

3. PROPOSED METHOD

To gain a better insight into our estimation problem let us elaborate some more on equation (4) introducing some simplifications. First of all, we neglect the means (which are typically negligible) and, considering that the terms at the denominator serve only to normalize the correlation, focus on the scalar product on the numerator. Remember that $z = yk$ is the camera PRNU multiplied point-wise by the input image and, likewise, $r = hy + n$ is the noise residual, with h the observed PRNU which might or might not coincide with k. Therefore, if we divide all terms point-wise by y, we obtain eventually the quantity

$$\tau_i = \frac{1}{|W_i|} \sum_{j \in W_i} \frac{r_j}{y_j} \frac{z_j}{y_j} = \frac{1}{|W_i|} \sum_{j \in W_i} (h_j + \frac{n_j}{y_j}) k_j \tag{7}$$

By defining a new noise field $\eta = nk/y$, and introducing generic weights ω_{ij}, eq.(7) becomes

$$\tau_i = \sum_{j \in W_i} \omega_{ij} (h_j k_j + \eta_j) \tag{8}$$

which can be interpreted as the linear filtering of the image hk affected by the additive noise η. In [7] the weights are all equals to one $1/|W_i|$, hence, a simple boxcar filtering is carried out.

Assuming that the whole analysis window is homogeneous, either genuine ($h = k$) or forged ($h \neq k$) and, for the sake of simplicity, that y is constant over the window, so that $E[\eta_i] = \sigma_\eta^2$, we can characterize the random variable τ

$$E[\tau] = \begin{cases} \langle k^2 \rangle_i & h = k \\ 0 & h \neq k \end{cases} \tag{9}$$

$$\mathrm{VAR}[\tau] = \sigma_\eta^2 \sum_j \omega_{ij}^2 \tag{10}$$

where $\langle k^2 \rangle$ is the power of the camera PRNU estimated over W_i. In this condition, using uniform weights $\omega_{ij} = 1/|W_i|$ is indeed optimal, as it minimizes the variance of the estimate, and maximizes the probability of deciding correctly. However, if some of the predictor pixels are not homogeneous with the target, that is, forged instead of genuine or viceversa, the estimate will suffer a systematic bias,

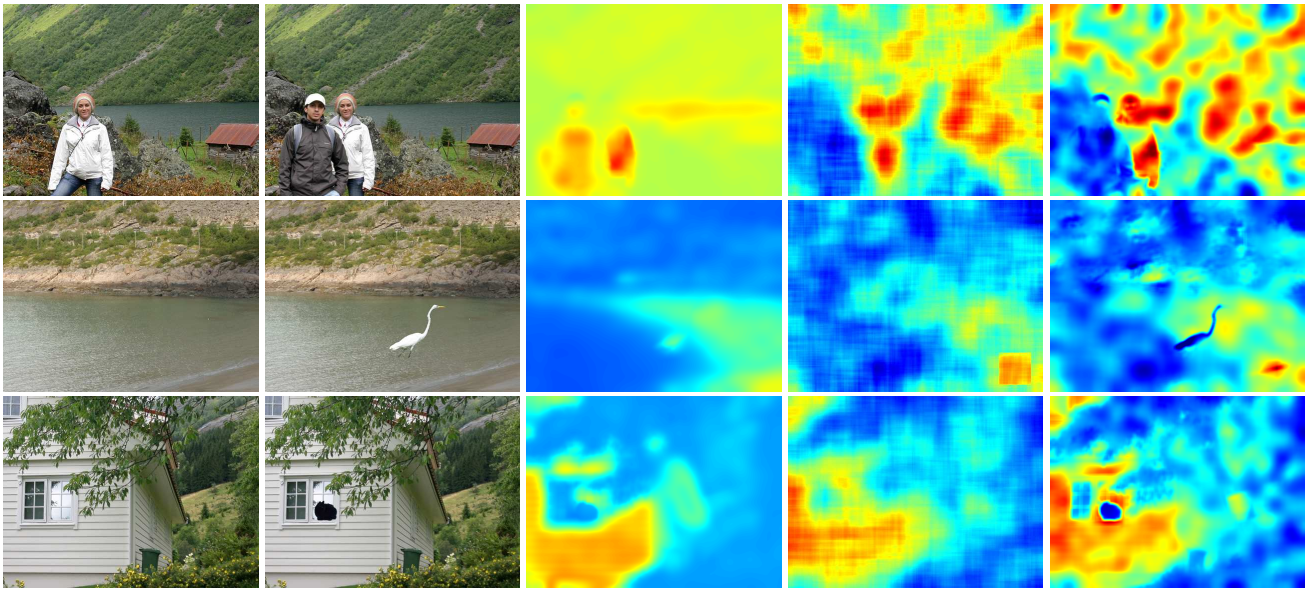

Fig. 1. Sample results. From left to right, original and forged image, correlation field predicted, and computed by boxcar and guided filtering.

namely, the means will not be 0 or $\langle k^2 \rangle$ anymore, but some intermediate values, heavily affecting the decision performance. In this case, the uniform weights are no more optimal, in general, and one should instead reduce the influence of heterogeneous pixels by associating a small or even null weight with them.

This is exactly the problem of small-size forgeries. By using a large analysis window with fixed weights we happen to include pixels of different nature, and the decision variable becomes strongly biased and basically useless, even in favourable (bright, smooth, unsaturated) areas of the image. If we could find and include in the estimation only predictors homogeneous with the target, all biases would disappear, at the cost of an increased estimation variance.

The bias / variance trade-off is indeed well-known in the denoising literature. This problem has received a great deal of attention, recently, in the context of nonlocal filtering, the current state of the art in denoising [21, 22], where predictor pixels are weighted based on their expected similarity with the target. The similarity, in its turn, is typically computed by comparing patches of pixels centered on the target and the predictor pixels respectively. This approach cannot work with our noise-like input image, rz, as it lacks the structures necessary to compute a meaningful similarity measure. However, we can take advantage of the original observed image y, using it as a "pilot" (again a well-known concept in denoising) to compute similarities, and applying the resulting weights in the actual filtering of the rz field.

Interestingly, this conceptual path has lead to an approach pretty similar to that followed in [17]. In both cases we use the original image to drive the filtering process emphasizing predictors that are likely to belong to the same object as the target. This happens through a preliminary segmentation in [17], by means of a more flexible adaptive filtering, here. It is worth underlining that our changes will concern only the correlation computation, while the decision process remains the same as in [7].

3.1. Implementation by guided filtering

Adaptive space-variant filters are typically characterized by high computational complexity and this is certainly the case with nonlocal filtering, where intensive patch-based processing is required. Considering, in addition, that the weak PRNU signal calls for large filtering windows, conventional nonlocal filters [21, 22] become unacceptably complex for this application. We resort therefore to guided filters, a recently proposed technique, which implements nonlocal filtering by leveraging heavily on the use of a pilot image associated with the target image.

Here, we follow closely the development and notation used in [20], referring the reader to the original paper for a more detailed treatment. Let p be the image to be filtered, q the filter output, and I a pilot image assumed to bear valuable information on p. We consider linear filtering, in the form

$$q_i = \sum_j \omega_{ij} p_j \tag{11}$$

Then, we assume that, locally to each pixel i, q depends linearly on I, that is

$$q_j = a_i I_j + b_i, \quad \forall j \in \Omega_i \tag{12}$$

where Ω_i is a square window of radius r centered on i. The parameters a_i and b_i are chosen to minimize over Ω_i the squared error between observed image and model

$$(a_i, b_i) = \arg\min_{(a,b)} \sum_{j \in \Omega_i} [(a_i I_j + b_i - p_j)^2 + \varepsilon a_i^2] \tag{13}$$

with ε a regularizing parameter that penalizes large values of a. The optimal values are

$$a_i = \frac{1}{|\Omega_i|} \sum_{j \in \Omega_i} \frac{I_j\, p_j - \bar{I}_i\, \bar{p}_i}{\sigma_i^2 + \varepsilon} \tag{14}$$

$$b_i = \bar{p}_i - a_i \bar{I}_i \tag{15}$$

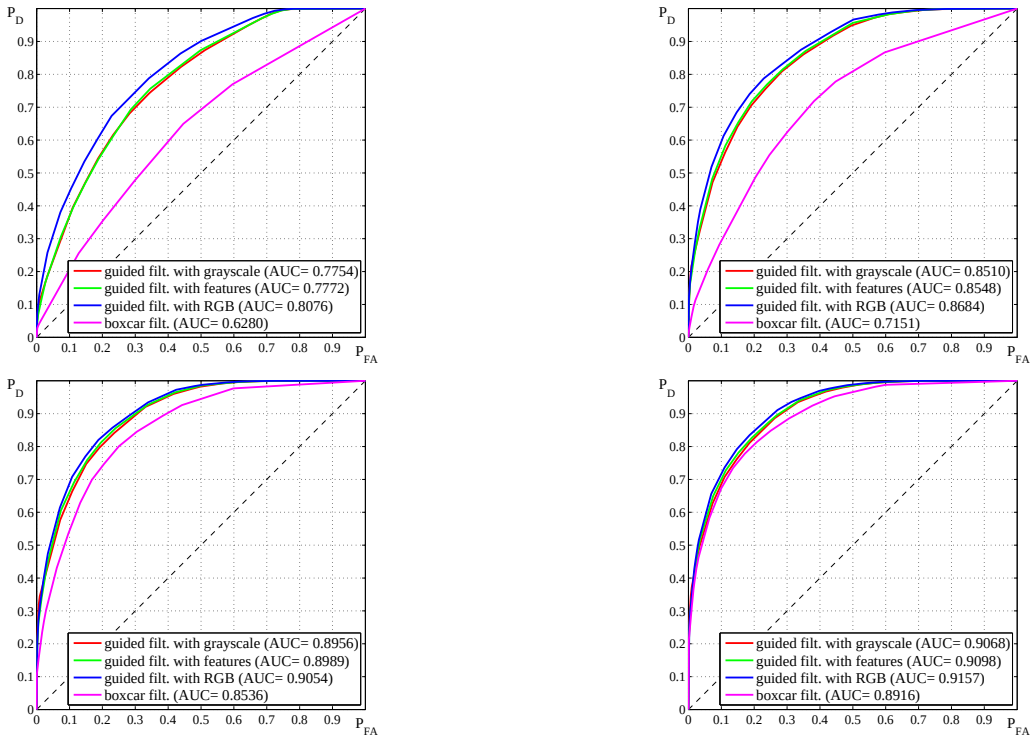

Fig. 2. ROCs obtained with boxcar and guided filtering with forgeries of size: 48×48, 64×64, 96×96, and 128×128 pixels.

where $\bar{x}_i$ indicated average of x over Ω_i and σ_i^2 is the variance of I over Ω_i. By substituting the optimal values back into (12) we obtain an estimate of q_j for all output pixels in the window Ω_i. Each of these pixels, however, falls in several such windows, and hence, to obtain the final filtered value, we average all such estimates

$$q_j = \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} (a_i I_j + b_i) = \bar{a}\, I_j + \bar{b} \tag{16}$$

which is the final expression of the linear filtering process of p guided by the pilot image I under the local linear model (12). The main reason for reporting all intermediate expression is to point out that all computations amount to a few boxcar filtering, applied to p, I, I^2, a, and b, carried out by integral image techniques with negligible complexity.

For our algorithm, of course, the input image is the product rz, the output is the decision statistic ρ, while the pilot (scalar) image can be a combination of the color bands of the original image y, its denoised version x, or any suitable field of features extracted from these images. By tuning the two parameters of the filter, the window radius r and the regularizing parameter ϵ, the influence of the pilot image in the filtering process can be modulated at will.

4. EXPERIMENTAL RESULTS

To prove the potential of the proposed approach we begin by showing, in Fig.1, a few sample images and the corresponding correlation fields. The image on the first row presents a large forgery, easily detectable in both the correlation fields (last two columns) as the region is much darker than in the predicted field (middle column). On the second and third row, instead, we have quite small forgeries, which leave little or no trace in the field computed by boxcar filtering, while are clearly detectable in the field obtained by guided filtering. Although these last examples are very favourable for the guided filtering approach, due to the high contrast between forgeries and background, they make clear that the original image can help making a better decision.

More convincing results are presented in Fig.2, showing the receiver operating curves (ROC) obtained using the original algorithm [7] and three implementations of the proposed method using different pilots (grayscale image, image in RGB, and some features extracted from the image). We use a test set of 200 768×1024-pixel images with a square forgery at the center, drawn at random from a different image. The camera (a Canon EOS-450D) PRNU is estimated off-line on a separate training set, used also to design the predictor. Each ROC is the upper envelope of pixel-level (P_D, P_{FA}) points obtained as the algorithm parameters vary. For guided filtering we used $r = 32$ and $\varepsilon = 0.16$, while the usual 128×128 window ($r = 64$) is used for boxcar filtering, and in all cases, to allow a fair comparison, the minimum size of acceptable detected forgeries was lowered to 32×32 pixels. Comparison is carried out separately for very-small, small, medium and large forgeries. With forgeries of dimension 48×48 pixels and 64×64 pixels (first two graphs), guided filtering guarantees a large performance improvement over boxcar filtering, synthesized by the area under curve (AUC) figure which grows from 0.63 to 0.78 in the first case and from 0.71 to over 0.85 in the second. With medium-size forgeries, 96×96 pixels, the performance gain is much more limited, with the AUC growing from 0.85 to 0.90, and becomes almost negligible, as expected, with larger 128×128 forgeries. No significant difference is observed, instead, as the pilot image changes, with the RGB pilot only slightly preferable to the others.

5. REFERENCES

[1] "Photo tampering throughout history," http://www.fourandsix.com/photo-tampering-history/

[2] I. Yerushalmy, and H. Hel-Or, "Digital Image Forgery Detection Based on Lens and Sensor Aberration," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 71–91, 2011.

[3] H. Fu, and X. Cao, "Forgery Authentication in Extreme Wide-Angle Lens Using Distortion Cue and Fake Saliency Map," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1301–1314, Aug. 2012.

[4] A.C. Popescu, and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, Oct. 2005.

[5] A.C. Kot, "Accurate Detection of Demosaicing Regularity for Digital Image Forensics," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 899–910, Dec. 2009.

[6] J. Lukas, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," *Proceedings of the SPIE*, vol. 6072, pp. 362–372, 2006.

[7] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining Image Origin and Integrity Using Sensor Noise" *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74–90, 2008.

[8] G.E. Healey, and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 267–276, Mar. 1994.

[9] B.B. Liu, Y. Hu, and H.K. Lee, "Source camera identification from significant noise residual regions," in *IEEE International Conference on Image Processing*, pp. 1749–1752, 2010.

[10] C.T. Li, "Source Camera Identification Using Enhanced Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, June 2010.

[11] C.T. Li and Y. Li, "Color-Decoupled Photo Response Non-Uniformity for Digital Image Forensics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 2, pp. 260–271, Feb. 2012.

[12] J. Fridrich, "Sensor Defects in Digital Image Forensic," *Digital Image Forensics*, pp. 179–218, 2012.

[13] M. Goljan and J. Fridrich, "Sensor-fingerprint based identification of images corrected for lens distortion," in *Proc. of SPIE: Media Watermarking, Security, and Forensics*, vol. 8303, 2012.

[14] T. Gloe, S. Pfennig, and M. Kirchner, "Unexpected Artefacts in PRNU-Based Camera Identification: A 'Dresden Image Database' Case-Study," in *ACM Workshop on Multimedia and Security*, pp. 109–114, 2012.

[15] C. Zhang and H. Zhang, "Exposing Digital Image Forgeries by Using Canonical Correlation Analysis," *20th International Conference on Pattern Recognition*, pp. 838–841, 2010.

[16] G. Chierchia, S. Parrilli, G. Poggi, C. Sansone, and L. Verdoliva, "On the influence of denoising in PRNU based forgery detection," in *ACM workshop on Multimedia in Forensics, Security and Intelligence* pp. 117–122, 2010.

[17] G. Chierchia, S. Parrilli, G. Poggi, L. Verdoliva, and C. Sansone, "PRNU-based detection of small-size image forgeries," *International Conference on Digital Signal Processing (DSP)*, pp. 1–6, 2011.

[18] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "PRNU-based forgery detection with regularity constraints and global optimization," *International Multimedia Signal Processing Conference (MMSP)*, 2013.

[19] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A Bayesian-MRF Approach for PRNU-based Image Forgery Detection," *IEEE Transactions on Digital Forensics and Security*, submitted, 2013.

[20] K. He, J. Sun, and X. Tang, "Guided Image Filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, June 2013.

[21] A. Buades, B. Coll, J.M. Morel, "A Review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[22] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.