Multi-scale Analysis Strategies in PRNU-based Tampering Localization

Paweł Korus, *Member, IEEE,* Jiwu Huang, *Fellow, IEEE*

***Abstract*—Accurate unsupervised tampering localization is one of the most challenging problems in digital image forensics. In this study, we consider photo response non-uniformity (PRNU) analysis and focus on the detection of small forgeries. For this purpose, we adopt a recently proposed paradigm of multi-scale analysis and discuss various strategies for its implementation. Firstly, we consider a multi-scale fusion approach which involves combination of multiple candidate tampering probability maps into a single, more reliable decision map. The candidate maps are obtained with sliding windows of various sizes and thus allow to exploit the benefits of both small and large-scale analysis. We extend this approach by introducing modulated threshold drift and content-dependent neighborhood interactions, leading to improved localization performance with superior shape representation and easier detection of small forgeries. We also discuss two novel alternative strategies: a segmentation-guided approach which contracts the decision statistic to a central segment within each analysis window; and an adaptive-window approach which dynamically chooses analysis window size for each location in the image. We perform extensive experimental evaluation on both synthetic and realistic forgeries and discuss in detail practical aspects of parameter selection. Our evaluation shows that multi-scale analysis leads to significant performance improvement compared with the commonly used single-scale approach. The proposed multi-scale fusion strategy delivers stable results with consistent improvement in various test scenarios.**

***Index Terms*—digital image forensics; tampering localization; decision fusion; multi-scale analysis; photo-response non-uniformity; Markov random fields**

I. Introduction

Analysis of sensor pattern noise signatures, and in particular photo-response non-uniformity (PRNU), constitutes one of the most powerful forensic techniques for digital photographs [1]. Imperfections of imaging sensors introduce consistent noise, characteristic for each device, which enables reasoning about the origin and authenticity of photographs. Identification of signature inconsistencies in various regions of an image leads to a localization map indicating the most likely tampered content and is invaluable for discovering intentions of a forger.

Despite considerable interest of research community, reliable tampering localization with PRNU signatures still poses challenges - especially for small forgeries. The core of PRNU signature verification involves correlation of a known noise pattern with its estimate from the investigated image. The operation is performed in a sliding window manner, which causes problems when the current window is heterogeneous, i.e., contains both original and tampered pixels.

The problem can be mitigated by explicitly guiding the computation with boundaries of image segments [2], or with guided filtering [3, 4]. Both approaches essentially contract the scope of correlation statistics. The first study used manual segmentation, requiring forensic analysts to painstakingly delineate meaningful objects in the image [2]. This method is particularly beneficial for small forgeries involving insertion of a highly contrasting object onto a solid background. However, for subtle object removal forgeries it is often impossible to define meaningful segments for such analysis. Additionally, prospective automation of this approach depends heavily on the success of image segmentation, which is still a challenging problem on its own. Instead, a recent study proposed to use guided filtering, but demonstrated no benefits for forgeries larger than $\approx 128 \times 128$ px [3].

Localization performance can also be improved by incorporating dependencies between neighboring image regions, e.g., by adopting a Markovian prior, which allows to propagate reliable decisions into ambiguous areas [5]. Further improvement can be obtained by better suppression of image content in the PRNU estimate. This can be achieved by: adopting more reliable denoising (e.g., BM3D [5, 6]); equalizing the spectrum of the PRNU [7]; retaining only its phase information [8]; attenuating strong components bleeding from image content [9]; or suppressing color interpolation artifacts based on color filter array's structure [10].

An ability to detect small forgeries is directly connected to the analysis window size, and constitutes one of the key trade-offs in unsupervised tampering localization. On the one hand, it is desirable to use small windows, capable of detecting even small forgeries. On the other hand, larger windows are required to obtain sufficient discriminability of detection statistics. Following the recommendations of Chen et al. [11], the prevailing window size in PRNU analysis is 128×128 px. We have recently proposed a multi-scale analysis approach which addresses the above problem and combines the benefits of small-scale and large-scale analysis [12]. Our method involves generation of separate candidate maps, obtained with different analysis windows, and their subsequent fusion into a single, more reliable map. Our previous work focused on the detection of JPEG splicing forgeries based on supervised learning with support vector machines trained on mode-based first digit features of DCT coefficients [13].

In this work, we study different strategies for performing

This work was partly supported by NSFC (U1636202, 61402295, 61572329) and Shenzhen R&D Program (JCYJ20160328144421330).

P. Korus is with the College of Information Engineering, also with the Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen, China, and also with the Department of Telecommunications, AGH University of Science and Technology, Kraków Poland (e-mail: pkorus@agh.edu.pl).

J. Huang is with the College of Information Engineering, and also with the Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen, China (e-mail: jwhuang@szu.edu.cn).

Supplementary materials are available at http://kt.agh.edu.pl/%7Ekorus
Source code is available at https://github.com/pkorus/multiscale-prnu

multi-scale analysis in PRNU-based tampering localization. Firstly, we adapt our fusion procedure to the forensic detector at hand. We further extend our approach by introducing content-dependent neighborhood interactions and modulated strength of the threshold drift. The resulting fusion method enforces stronger decision propagation within similar image regions leading to superior shape representation and more reliable detection of small forgeries. Hence, the introduced adaptive neighborhood interactions serve the same purpose as the recently proposed method based on guided filtering. However, it is easier to generalize to arbitrary forensic detectors, not necessarily involving 2-dimensional image filtering.

Secondly, we consider two novel alternative strategies. Our *segmentation-guided* strategy computes the correlation over a central segment within each analysis window. Our approach builds upon the main ideas of Chierchia et al. [2, 3] but obtains better results with consistent improvement regardless of the tampering area's size. We also consider a novel *adaptive-window* strategy, which chooses the analysis window size individually for every location in the image. We enhance both strategies, by adopting a conditional random field (CRF) to model neighborhood interactions and make the final decision.

We perform extensive experimental evaluation both on synthetic forgeries with strictly controlled tampering area, and on high-quality realistic forgeries prepared in modern photo-editing software. The considered multi-scale strategies allow to obtain significant performance improvement compared to conventional single-scale analysis. The proposed multi-scale fusion approach is the most versatile approach and delivers consistent benefits in diverse test scenarios.

The main contributions of our work include: (1) detailed discussion and experimental evaluation of various strategies of multi-scale analysis in PRNU-based tampering localization; (2) adaptation of our multi-scale fusion approach to PRNU-based tampering localization and its further extension by introducing content-dependent neighborhood interactions; (3) practical implementation of two alternative adaptive strategies, further augmented with a random field-based decision; our segmentation-guided strategy addresses limitations of a recent similar method based on guided filtering and obtains significant improvement of localization performance regardless of the tampering area's size.

The paper is organized as follows. In Section II, we review the state-of-the-art in tampering localization techniques and introduce the fundamentals of PRNU-based localization. We then extend our multi-scale fusion approach (Section III) and introduce two novel alternative strategies (Section IV). Our experimental evaluation scenario, and the obtained results are presented in Section V. Finally, we discuss the conclusions and limitations of our work (Section VI). The paper comes with supplementary materials and source code available online.

II. Existing Tampering Localization Techniques

Analysis of sensor pattern noise is not the only method of unsupervised tampering localization. A compelling image forgery needs to satisfy both high-level and low-level consistency constraints. High-level aspects include consistency of shadows & lighting [14, 15], depth-of-field & motion blur [16, 17], and perspective & geometry [18, 19]. Similarly, sophistication of the image acquisition pipeline allows to exploit many low-level signatures, starting from image demoisaicing [20, 21], camera response function [22], or local noise levels [23], up to final stages of JPEG compression [13, 24, 25]. When analyzing uncompressed bitmap images, forensic analysis can look for traces of previous compression, and even estimate the utilized JPEG quantization tables [26].

It is also possible to detect traces of popular forgeries, (e.g., splicing [27], copy-move [28], seam carving [29]) or other common image processing that is often used to mask the actual forgery (e.g., median filtering [30], resampling [31]). While many existing detectors are targeted at one specific type of forgery, some studies aim to provide general solutions capable of distinguishing many operations. A popular approach involves adoption of pixel co-occurrence models from image steganalysis (e.g., subtractive pixel adjacency matrix [32] or spatial rich models [33]). Such features can discriminate boundaries of a splicing forgery [34], low-level characteristics of specific camera models [35], or image patches with various post-processing [36, 37]. It has been recently shown that low-level models based on Gaussian mixtures can deliver competitive performance in the latter problem [37].

The final decision can hence include hints of many independent detectors, which can be fused together into a single more reliable assessment. Existing approaches vary from standard supervised or ensemble learning [38, 39] to sophisticated frameworks based on fuzzy logic [40], or the Dempster-Shafer theory of evidence [41]. The latter allow to deal with uncertainty and compatibility of individual detectors by defining both feasible and unfeasible combinations of individual traces. While decision fusion seems to be an emerging trend in image forensics, it has been studied mainly for forgery detection. In tampering localization simple logical rules still prevail [42, 43], and extension of advanced fusion methods is the subject of ongoing research [44]. A recent evaluation of various combination rules in a tampering localization scenario can be found in [43].

A conventional approach to tampering localization involves comparing responses of a forensic detector to a threshold. While the threshold can be chosen based on solid theoretical foundations, e.g., the Neyman-Pearson criterion [11], it is non-trivial to accurately model conditional distributions of the detection statistic. As a result, researchers began to explore other approaches that dispense with explicit distribution modeling. It can be argued that tampered areas should constitute a sparse cluster of outliers in a forensic feature space [45]. An iterative procedure derived from robust principal component analysis (PCA) can estimate the expected low-rank structure of pristine image features and detect non-matching tampering. This method does not require any training and relies solely on the investigated image. Tampering localization can also be performed by analyzing the behavior of a random walker on a graph spanned over the image [46]. Weights in the graph are derived from responses of a forensic detector. Performing the random walk according to a maximal entropy principle possesses a strong localization property, which allows to highlight tampered areas, and attenuate unimportant background.

A. Tampering Localization with PRNU Analysis

This section briefly introduces PRNU-based tampering localization. For more information, readers can refer to [11].

Due to manufacturing imperfections, imaging sensors exhibit minor photo response variations across their pixels. This phenomenon manifests itself as a slight device-specific noise that is consistent for all photographs. Estimation of this noise (with proper care accounting for post-capture geometric distortions like resizing, cropping or lens distortion compensation [47, 48]) yields a unique fingerprint of the camera, enabling reasoning about the origin, processing history and authenticity of digital photographs.

In the following description, we denote 2-dimensional arrays with lowercase bold symbols (e.g., $\mathbf{y}$), and for the sake of notation simplicity address individual elements of the array with a single index (e.g., y_i). We consider a simplified model of the image acquisition pipeline [11]:

$$y_i = (1 + k_i)x_i + n_i \ , \tag{1}$$

where $\mathbf{y}$ is a captured image; $\mathbf{x}$ is its idealized noise-free version; $\mathbf{k}$ is the PRNU pattern; and $\mathbf{n}$ is an additive distortion that accounts for all remaining sources of noise. The PRNU $\mathbf{k}$ can be estimated according to the maximum likelihood principle from residual images $\mathbf{r}$ with suppressed content:

$$r_i = y_i - \hat{x}_i \approx y_i k_i + n'_i \ , \tag{2}$$

where $\mathbf{n}'$ is an aggregated noise component, and $\hat{\mathbf{x}} = \mathcal{D}(\mathbf{y})$ is an estimate of the noise-free image, obtained with a denoising filter, such as wavelet-domain filtering [49] or BM3D [6].

The tampering localization protocol involves sliding window analysis of the investigated photograph. Let $\mathbf{y}[i]^{(\omega)}$ denote contraction of the original image $\mathbf{y}$ to a square window of size $\omega \times \omega$ centered around pixel i. For notation simplicity, we will omit the window size index if it is clear from context. Then, a *correlation field* can be obtained by computing normalized correlation between the residual image $\mathbf{r}[i]$ with an estimate of the PRNU $\hat{\mathbf{k}}[i]$ (multiplied by image content $\mathbf{y}[i]$):

$$q_i = \mathbf{r}[i] \otimes \hat{\mathbf{k}}[i] = \frac{(\mathbf{r}[i] - \overline{\mathbf{r}[i]}) \odot (\hat{\mathbf{k}}[i] - \overline{\hat{\mathbf{k}}[i]})}{||\mathbf{r}[i] - \overline{\mathbf{r}[i]}|| \cdot ||\hat{\mathbf{k}}[i] - \overline{\hat{\mathbf{k}}[i]}||} \ , \tag{3}$$

where $\odot$ denotes element-wise multiplication, and operators $||\mathbf{r}||$ and $\overline{\mathbf{r}}$ correspond to L2 norm and average value of array $\mathbf{r}$. Based on the obtained detection statistic, the detector decides in favor of one of two hypotheses: H_0 - signature is absent (content is tampered with); H_1 - signature is present (content is authentic).

From (2) and (3) it clearly follows that the decision is highly dependent on image content - stronger response can be expected in bright and flat image regions with reduced contamination of the signature by the remains of the image's texture. This challenge is addressed by predicting the expected correlation response $\hat{q}_i = \mathcal{Q}(\mathbf{y}[i]|\theta)$ based on selected features of the content [11]. The predictor, parametrized by θ, takes into account special situations like saturated image regions where PRNU cannot be detected. In practice, the measured correlation q_i oscillates around the expected values (0 or $\hat{q}_i$ for tampered and authentic regions, respectively). This stochastic process is commonly modeled either as a Gaussian or a generalized Gaussian distribution [5, 11]. For each analysis window, the detector ranks the likelihood of the following hypotheses (assuming a Gaussian noise model):

$$\begin{cases} H_0 & q_i \sim \mathcal{N}(0, \sigma_0) \ , \\ H_1 & q_i \sim \mathcal{N}(\hat{q}_i, \sigma_1) \ . \end{cases} \tag{4}$$

Hence, a typical camera model will contain:

$$\begin{cases} \hat{\mathbf{k}} & \text{PRNU estimate,} \\ \sigma_0 & \text{variance of the detection statistic for } H_0, \\ \sigma_1 & \text{variance of the detection statistic for } H_1, \\ \theta & \text{parameters of the correlation predictor.} \end{cases} \tag{5}$$

Due to signatures' independence under H_0, the variance σ_0 can also be estimated as $\sigma_0 = 1/m$ with m denoting the number of samples (pixels) used in the computation. Finally, the probability of the image window being tampered is:

$$c_i = \mathcal{B}(q_i|\sigma_0, \sigma_1, \hat{q}_i) = \frac{P_{\mathcal{N}(0,\sigma_0)}(q_i)}{P_{\mathcal{N}(\hat{q}_i,\sigma_1)}(q_i) + P_{\mathcal{N}(0,\sigma_0)}(q_i)} \tag{6a}$$

$$= \left(1 + e^{-\log(\sigma_1/\sigma_0) - \frac{(q_i - \hat{q}_i)^2}{2\sigma_1^2}) + \frac{q_i^2}{2\sigma_0^2}}\right)^{-1} \tag{6b}$$

where $P_{\mathcal{X}}$ denotes a probability density function of distribution $\mathcal{X}$. Computation in log domain (6b) allows to avoid numerical precision limitations.

The original localization method [11] used a slightly different formulation where the obtained correlation q_i is validated against two thresholds corresponding to the desired error rates. However, in the considered decision fusion setting it is crucial to exploit full real-valued responses (be it probabilities or other scores). Such *measurement-level* fusion leads to better performance than binary *decision-level* fusion [12, 41]. Adoption of probabilities instead of raw correlation [5] or peak-to-correlation energy (PCE) scores (commonly used in PRNU-based source attribution [1]), makes it more convenient to formulate the fusion problem, and facilitates easier generalization to different forensic detectors where probabilistic measures emerge naturally.

In order to obtain reliable statistics, the correlation should be computed over sufficiently large windows. However, excessively large sizes are discouraged since they are more likely to: (a) miss small forgeries; (b) violate stationarity assumptions [11]. Hence, the choice of analysis window size involves a trade-off between the desired localization resolution and reliability of the detection statistic. Following the original paper [11], virtually all PRNU-based localization schemes described in scientific literature use square windows of size $\omega = 128$. The computed scores can be applied either to a central pixel of the window (*central-pixel attribution*) or to all pixels in the window (*full-window attribution*). Due to large window size, PRNU-based localization schemes typically use the former strategy. The described localization algorithm is summarized as pseudo-code in Alg. 1.

Algorithm 1 Pseudo-code for the standard localization algorithm: $\mathcal{D}$ - denoising; $\mathcal{Q}$ - correlation predictor; $\mathcal{B}$ - tampering probability; $\mathcal{S}$ - sub-sampling & padding; $\mathcal{H}$ - post-processing heuristics.

```
Input: $\mathbf{y}, \hat{\mathbf{k}}$                         ▷ input image, PRNU estimate
Input: $\omega, \Delta\omega, \tau$              ▷ window size, analysis stride, threshold
Input: $\theta, \sigma_0, \sigma_1$                     ▷ camera model parameters
  $\mathbf{r} \leftarrow \mathbf{y} - \mathcal{D}(\mathbf{y})$                          ▷ noise residual
  for $i \in$ analysis locations do
    $\hat{q}_i \leftarrow \mathcal{Q}(\mathbf{y}[i] \mid \theta)$                    ▷ correlation estimate
    $q_i \leftarrow \mathbf{r}[i] \otimes \hat{\mathbf{k}}[i]$                      ▷ correlation, Eq. (3)
    $c_i \leftarrow \mathcal{B}(q_i \mid \hat{q}_i, \sigma_0, \sigma_1)$         ▷ tampering probability, Eq. (6)
  end for
  $\mathbf{c} \leftarrow \mathcal{S}(\mathbf{c}, \Delta\omega)$                     ▷ sub-sampling & padding
  return $\mathbf{t} \leftarrow \mathcal{H}(\mathbf{c} > \tau)$                      ▷ final decision
```

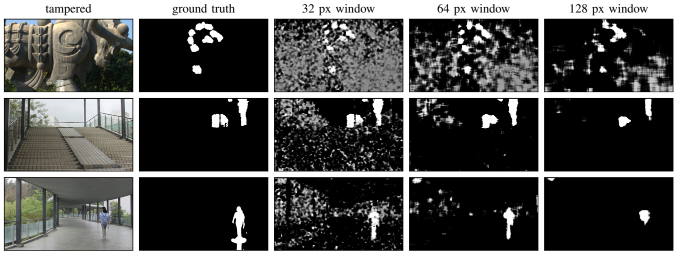


Fig. 1. Example multi-scale tampering probability maps for 3 realistic forgeries; while the forgeries are only roughly detected by the default 128 px window, they are accurately detected by smaller windows; uncertainties (gray regions with scores $c_i \approx 0.5$) will be resolved by a multi-scale fusion procedure by exploiting information available in all scales of analysis.

Based on the obtained tampering probability map, the final decision map is obtained by heuristic post-processing. Firstly, the real-valued map $\mathbf{c}$ is compared to a threshold τ. The resulting binary map is cleaned by removing small connected components having less than a quarter of the total pixel count in the window. Finally, morphological dilation is applied to compensate for the erosion of the detection boundary and fill prospective small holes in the detected objects. In this study, we used a disk-shaped structural element with radius of 10 px.

III. Multi-scale Forensic Analysis

In this section, we introduce multi-scale forensic analysis [12]. We focus on a random-field formulation of the problem where decision fusion resolves to finding an optimal labeling of authentication units (e.g., image blocks / pixels) that minimizes a given energy function. We have adapted the algorithm to specific characteristics of PRNU verification and further extended it with new features. The differences with respect to the original algorithm are as follows: (1) we discard only empty candidate maps that contribute no information for the localization procedure; (2) we do not perform separate detection of unreliable image regions and rely on the correlation predictor to account for such cases; (3) we use real-valued instead of binary reliability maps; (4) we modulate the strength of the threshold drift proportionally to the reliability of individual image regions; (5) we introduce adaptive neighborhood interactions which strengthen decision propagation within similar image regions.

Note that changes (1) and (2) were made specifically for the PRNU detector at hand, and may need to be reconsidered when adopting multi-scale analysis for other detectors. The remaining modifications are more general enhancements, which give the algorithm more flexibility and should improve localization capabilities regardless of the detector. In particular, modification (5) leads to significantly improved shape representation and enhances detection of small forgeries.

A. Fundamentals of Multi-scale Forensic Analysis

Multi-scale tampering localization involves analysis of the investigated image with sliding windows of successively increasing size $\{\omega_s\}$ for $s \in \{1, \ldots, S\}$. Resulting *candidate maps* are then fused together to obtain a single tampering map that combines the benefits of both small-scale and large-scale analysis. Formally, the goal of a fusion procedure is to produce a binary decision map given a set of S candidate maps:

$$\left(\{\mathbf{c}^{(s)}\}, \{\mathbf{p}^{(s)}\}, \mathbf{y}\right) \rightarrow \mathbf{t} \in \{0,1\}^N \ , \tag{7}$$

where $\mathbf{c}^{(s)} \in [0,1]^N$ denotes the s-th input candidate map corresponding to analysis window of size ω_s. We assume identical size of all maps $N = N_x \times N_y$ with elements corresponding to tampering probabilities of individual authentication units. Each candidate map has a corresponding *reliability map* $\mathbf{p}^{(s)} \in [0,1]^N$ which identifies its unreliable regions, e.g., due to pixel saturation or other limitations of the forensic detector. The fusion procedure can also exploit image content $\mathbf{y}$ to guide tampering localization.

Three example sets of multi-scale tampering probability maps are shown in Fig. 1. Note that while the forgeries are only roughly detectable with the commonly used 128 px window, they can be accurately detected in smaller scales. While smaller windows yield more noisy and uncertain maps (many regions with scores $c_i \approx 0.5$), these ambiguities can be resolved by the fusion procedure by incorporating information available on all scales of analysis.

The fusion problem can be formulated in terms of random fields and resolves to finding the optimal labeling of authentication units (with labels $t_i = 1$ corresponding to tampered regions) that minimizes the following energy function [12]:

$$\frac{1}{S}\sum_{i=1}^{N}\sum_{s=1}^{S} E_\tau(c_i^{(s)}, t_i) + \alpha \sum_{i=1}^{N} t_i + \sum_{i=1}^{N}\sum_{j \in \Xi_i} \beta_{ij}|t_i - t_j| \ . \tag{8}$$

The first term (referred to as the *data term*) penalizes differences with respect to the candidate maps (the potentials $E_\tau(c, t)$ will be described in detail later). The second term introduces a penalty α for tampered authentication units, and thus can be used to bias the decision towards either of the hypotheses. The third term penalizes differences in the decisions for neighboring authentication units, and thus encodes a preference towards piecewise-constant solutions.

The above formulation uses a Markovian prior to model interactions between neighboring authentication units. The decision for each unit i depends directly only on its own potentials, and on the decisions for its neighbors $j \in \Xi_i$. In this study, we consider a 2nd-order neighborhood, i.e., Ξ_i contains up to 8 immediate neighbors (pruned accordingly near image borders). In our previous work, the neighborhood interaction penalty β_{ij} was set to a constant value $\beta_{ij} = \beta$. In this study, we allow for adaptive selection of the penalty for each pair of authentication units (i, j). The issue will be discussed in detail in Section III-C.

Potentials of the data term are responsible for maintaining resemblance to the input candidate maps. Since the candidate scores correspond to probabilities of individual hypotheses, the data terms could be obtained as $-\log(c_i)$ and $-\log(1 - c_i)$, respectively. However, we use the following generalization:

$$E_\tau(c, t) = -\log \max(\Psi_{\text{min}}, \Psi_\tau(c, t)), \tag{9}$$

with $\Psi_{\text{min}} \in [0, 1]$ and:

$$\Psi_\tau(c, t) = \begin{cases} 1 - \frac{c}{2\tau} & \text{for } t = 0, \\ 1 + \frac{c}{2(1-\tau)} - \frac{1}{2(1-\tau)} & \text{for } t = 1, \end{cases} \tag{10}$$

where $\tau \in (0, 1)$ is a quasi-threshold that equalizes potentials for both decisions, i.e., $E_\tau(\tau, 0) = E_\tau(\tau, 1)$. Setting a minimal value Ψ_{min} (0.001 in our experiments) prevents the nodes from becoming fixed to certain decisions (due to infinite energy).

Similarly to our previous work, we use the technique of *threshold drift* [12] - we perform successive adjustment of the quasi-threshold based on hypothetical intermediate decisions from smaller scales. We record an individual threshold for each authentication unit $\tau_i^{(s)}$, and update it as follows:

$$\begin{cases} \tau^{(1)} & \text{if } s = 1 \ , \\ \tau_i^{(s-1)} + \delta p_i^{(s-1)} & \text{if } s > 1 \text{ and } c_i^{(s-1)} \leq \tau_i^{(s-1)}, \\ \tau_i^{(s-1)} - \delta p_i^{(s-1)} & \text{if } s > 1 \text{ and } c_i^{(s-1)} > \tau_i^{(s-1)} \ , \end{cases} \tag{11}$$

where $\delta \in [0, 1]$ is the strength of the drift and $\tau^{(1)}$ is an initial threshold, typically chosen around 0.5. In order not to discard extremely confident scores from larger scales, we do not drift the threshold above 0.95 or below 0.05. In contrast to our previous work, the drift is weighted proportionally to the regions' reliability. Such an approach can further improve the tampering localization performance.

B. Map and Region Reliability in PRNU Analysis

We consider two aspects of map reliability. Firstly, each candidate map has a corresponding reliability map $\mathbf{p}^{(s)}$ which indicates regions with reliable ($p_i = 1$) and unreliable candidate scores ($p_i = 0$). Unreliable regions may benefit from resetting their scores (e.g., to eliminate false positives in saturated areas) either to 0 (a conservative strategy) or close to the initial threshold $\tau^{(1)}$ (to facilitate easier score propagation through neighborhood interactions). Regions with intermediate values will exhibit proportionally attenuated behavior, e.g., by weighting the strength of the threshold drift in (11).

In general, the way of implementing reliability scores depends on individual characteristics of a forensic detector. For the considered PRNU detector, unreliable regions correspond to dark, highly textured or saturated areas. However, the correlation predictor is designed to account for these problems leading to overlapping distributions of the detection statistic and candidate scores close to 0.5. Hence, we do not perform separate detection of such regions, and set the reliability values proportionally to the distance of the scores from 0.5. We used the following reliability expression:

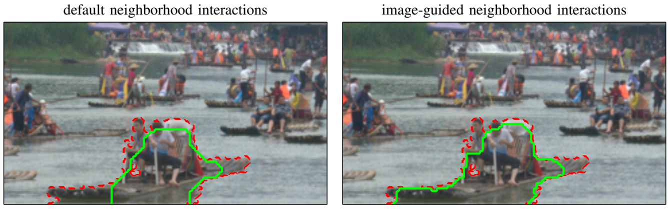

default neighborhood interactions
image-guided neighborhood interactions

Fig. 2. Example multi-scale fusion result with default, content-independent (left) and adaptive, image-guided (right) neighborhood interactions; red color denotes ground truth; green denotes detected tampering.

$$p_i = 1 - e^{-\xi_0 |c_i - \frac{1}{2}|^{\xi_1}}, \tag{12}$$

and found parameters ($\xi_0 = 30, \xi_1 = 2.5$) with a grid search on a small set of diverse test images. For each considered reliability curve, we searched for the best parameters of the CRF (see Section V-C for a detailed description) that maximize the average F_1 score. The best / worst observed F_1 scores were approximately 0.63 and 0.60, respectively. The worst performance was observed for a degenerated curve where $p_i := 1$ which corresponds to the unweighted threshold drift used in our previous work [12]. We found that the shape of the curve is not very important, as long as it assigns high reliability to confident scores and smoothly decreases as $c_i \to 0.5$.

In our previous work [12], we found it very important to assess the utility of whole candidate maps, and reject the ones that do not contribute any useful information for the localization. A similar observation was also made in a multi-modal fusion setting by Ferrara et al. [44] who discarded tampering maps having less than 1/8 tampered blocks. Our previous approach measured similarity of candidate maps to random Gaussian noise. However, in the considered scenario such an approach tended to mix borderline cases and remove some noisy but useful maps. Based on further evaluation, we observed that adverse effect of noisy maps is largely eliminated by the introduced modulation of the threshold drift. Therefore, in this study we discard only empty maps which contribute no useful information for the localization procedure.

C. Exploiting Image Content

The neighborhood interaction penalty β can be computed individually for each pair of neighboring authentication units. We use the tampered image (down-sampled to match the size of candidate maps: $\mathbf{y} \to \mathbf{y}'$) to guide the process. We consider two components of the interaction penalty:

$$\beta_{ij} = \beta_0 + \beta_1 e^{-\frac{1}{2}\phi^{-2} ||y'_i,\ y'_j||^2_{L2}} \ , \tag{13}$$

where the $||y'_i,\ y'_j||_{L2}$ operator denotes L2 distance between two pixels (computed from RGB vectors). The first term encodes a default, content-independent penalty. The second term represents the interactions of only similar authentication units, with similarity attenuation controlled by parameter ϕ (we use an empirically chosen $\phi = 25$). Analogous adaptive terms are used in state-of-the-art image segmentation methods [50].

Fig. 2 shows an example result of multi-scale tampering localization with (right) and without (left) content-dependent penalties. In the former case, the detected area (green) matches the ground truth (red) significantly better. While best results can be obtained for contrasting object insertion forgeries, we did not observe negative effects for more subtle cases with object removal which are challenging for previous methods based on explicit image segmentation. Hence, the introduced mechanism leads to similar benefits as the mentioned guided filtering approach [3]. However, since it does not rely on convolutions, it can be easily adopted to other forensic detectors.

IV. Alternative Adaptive Strategies

Fusion of separate multi-scale maps is not the only way to exploit the benefits of multi-scale analysis. We consider two alternative strategies, including a *segmentation-guided* approach inspired by the work of Chierchia et al. [2, 3] (Section IV-A); and an *adaptive-window* approach where the window size is dynamically chosen for every location in the image (Section IV-B).

For both strategies, we compare two methods of making the final decision. Firstly, we use a conventional post-processing heuristic, which involves explicitly comparing the tampering probability map to a threshold, followed by removal of small connected components (having less than $32^2 = 1024$ px). Note that in contrast to single-scale detectors, adaptive strategies are expected to follow objects' edges much more accurately and are not subjected to morphological dilation.

Secondly, we consider a simplified version of the conditional random field (CRF) described in Section III with the following energy function:

$$E(\mathbf{t}|\mathbf{c}) = \sum_{i=1}^{N} E_\tau(c_i, t_i) + \alpha \sum_{i=1}^{N} t_i + \sum_{i=1}^{N} \sum_{j \in \Xi_i} \beta_{ij} |t_i - t_j| \ . \quad (14)$$

Analogously to multi-scale fusion, this decision is controlled by a quasi-threshold τ and parameterized by tampering penalty α, and two interaction parameters β_0, β_1. As a result, it can also benefit from adaptive neighborhood interactions.

A. Segmentation-Guided PRNU Analysis

Our segmentation-guided detector is fully automatic and uses rough segmentation boundaries to limit the scope of correlation statistics. We start with the default window of size $\omega = 128$ with *central-pixel attribution*. However, the correlation is computed only for pixels that belong to the same segment as the central pixel. Let $\mathbf{v} \in \{0,1\}^{\omega \times \omega}$ denote a binary matrix serving as a segmentation mask, initialized based on pixel distance in the RGB space:

$$v_j = 1 \Leftrightarrow \frac{1}{3} ||y[i]_j, \ y_i||_{L1} < \Delta y \ . \quad (15)$$

We then clean the resulting segmentation with small-size morphological closing (we used a disk-shaped structural element of radius 8 px) and leave only the central segment, i.e., we eliminate connected components that do not overlap the i-th pixel. In order to prevent excessive degradation of the correlation statistic, we assume that at least ω_{min}^2 px are required for the computation. If the segmentation yields a smaller region, we expand it with small-scale morphological dilation. Finally, we adapt the distribution models for the H_0 and H_1 hypotheses by matching their variances σ_0, σ_1 to the actual number of pixels used in the correlation (we used cubic spline interpolation between the estimates $\sigma_0(\omega_s)$ and $\sigma_1(\omega_s)$ available in our multi-scale camera model). The above algorithm is summarized as pseudo-code in Alg. 2.

Algorithm 2 Pseudo-code for the segmentation-guided localization algorithm: $\mathcal{D}$ - denoising; $\mathcal{Q}$ - correlation predictor; $\mathcal{B}$ - tampering probability; $\mathcal{S}$ - sub-sampling (including missing edges completion).

```
Input: $\mathbf{y}, \hat{\mathbf{k}}$                                  ▷ input image, PRNU estimate
Input: $\omega, \Delta\omega, \tau$                         ▷ window size, analysis stride, threshold
Input: $\theta, \{\sigma_0(\omega_s), \sigma_1(\omega_s)\}$            ▷ multi-scale camera model
Input: $\Delta y, \omega_{min}$                 ▷ similarity threshold, minimum window size
  $\mathbf{r} \leftarrow \mathbf{y} - \mathcal{D}(\mathbf{y})$
  for $i \leftarrow$locations do
    $\mathbf{v} \leftarrow \frac{1}{3}||\mathbf{y}[i], \ y_i||_{L1} < \Delta y$                    ▷ rough segmentation
    $\mathbf{v} \leftarrow$ morphological closing of $\mathbf{v}$                           ▷ cleanup
    $\mathbf{v} \leftarrow$ central connected component of $\mathbf{v}$
    while $\omega_{eff} = \sum \mathbf{v} < \omega_{min}^2$ do
      $\mathbf{v} \leftarrow$ morphological dilation of $\mathbf{v}$ ▷ grow region if too small
    end while
    $\hat{q}_i \leftarrow \mathcal{Q}(\mathbf{y}[i]^{(\omega)} \mid \theta)$                    ▷ correlation estimate
    $q_i \leftarrow \{r[i]_j : v_j = 1\} \otimes \{\hat{k}[i]_j : v_j = 1\}$ ▷ correlation, Eq. (3)
    $\hat{c} \leftarrow \mathcal{B}(q_i \mid \hat{q}_i, \sigma_0(\omega_{eff}), \sigma_1(\omega_{eff}))$ ▷ tampering prob., Eq. (6)
  end for
  $\mathbf{c} \leftarrow \mathcal{S}(\mathbf{c}, \Delta\omega)$                              ▷ sub-sampling & padding
  return $\mathbf{t} \leftarrow \underset{\mathbf{t}}{\text{argmin}} \ \text{E}(\mathbf{t}|\mathbf{c})$                  ▷ final decision, Eq. (14)
```

An example localization result is shown in the top row of Fig. 3 which compares correlation fields and tampering probability maps for both the described segmentation-guided and the original fixed-window methods. The segmentation-guided approach can accurately delineate boundaries of contrasting objects, leading to better shape representation, and detection of small forgeries (e.g., the peak of the pagoda). However, if the forgery involves subtle object removal with no obvious boundaries (e.g., inpainting of the electrical wire) no improvement should be expected.

Note that in principle such an approach should use a more sophisticated predictor capable of adapting to arbitrary shapes of image segments. However, the ability to design a better predictor is still an open problem [5] and we leave this aspect for our future work. Note also that the necessity to adapt hypothesis models in the predictor requires control over the actual number of pixels used in the correlation. As a result, although soft-segmentation (e.g., guided filtering) is an inspiring idea, it is likely suboptimal. This argument seems to be supported by the fact that our simple segmentation-guided strategy (both with heuristic and CRF-based decision) obtains better results than the original guided filtering-based approach [3]. In contrast to their method, we observed consistent improvement of localization performance - even for large forgeries (see Section V-D).

B. Adapting Window Size in PRNU Analysis

The *adaptive-window* strategy involves choosing analysis window size individually for each location in the image. In

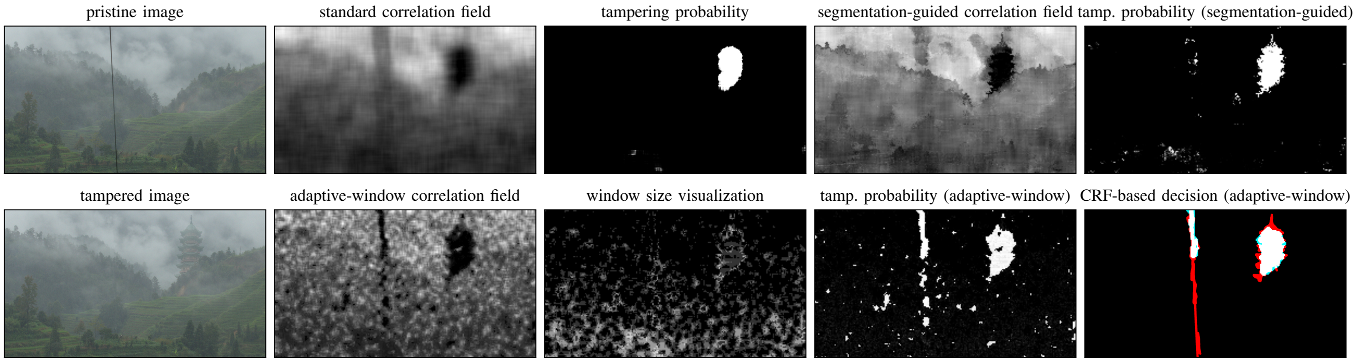

pristine image
standard correlation field
tampering probability
segmentation-guided correlation field
tamp. probability (segmentation-guided)
tampered image
adaptive-window correlation field
window size visualization
tamp. probability (adaptive-window)
CRF-based decision (adaptive-window)

Fig. 3. Example localization result for alternative adaptive localization strategies; the *segmentation-guided* strategy (top right) can accurately delineate contrasting segments which leads to better shape detection and even to the detection of very small forgeries (peak of the pagoda); however, it is not beneficial for subtle object removal forgeries (e.g., the inpainted electrical wire) which are easier to detect with the *adaptive-window* strategy (bottom); window size visualization shows the index s of the chosen scale starting from 32×32 px (black) to 256×256 px (white).

Algorithm 3 Pseudo-code for the adaptive-window localization algorithm: $\mathcal{D}$ - denoising; $\mathcal{Q}$ - correlation predictor; $\mathcal{B}$ - tampering probability; $\mathcal{S}$ - sub-sampling (including missing edges completion).

```
Input: $\mathbf{y}, \hat{\mathbf{k}}$    ▷ input image, PRNU estimate
Input: $\{\omega_s\}, \Delta\omega, \tau$    ▷ window sizes, analysis stride, threshold
Input: $\{\theta(\omega_s), \sigma_0(\omega_s), \sigma_1(\omega_s)\}$    ▷ multi-scale camera model
Input: $\Delta c_1, \Delta c_2$    ▷ score thresholds (stopping criteria)
  $\mathbf{r} \leftarrow \mathbf{y} - \mathcal{D}(\mathbf{y})$
  for $i \leftarrow$locations do
    $s \leftarrow 1$    ▷ start with the smallest window
    $\tilde{c} \leftarrow 0.5$    ▷ buffer for last score
    $s_{max} \leftarrow$ max window for location $i$
    while $s \leq s_{max}$ and $|\tilde{c} - 0.5| < 0.5 - \Delta c_1$ do
      $\hat{q}_i \leftarrow \mathcal{Q}(\mathbf{y}[i]^{(\omega_s)} \mid \theta(\omega_s))$    ▷ correlation estimate
      $q_i \leftarrow \mathbf{r}[i]^{(\omega_s)} \otimes \hat{\mathbf{k}}[i]^{(\omega_s)}$    ▷ correlation, Eq. (3)
      $\hat{c} \leftarrow \mathcal{B}(q_i \mid \hat{q}_i, \sigma_0(\omega_s), \sigma_1(\omega_s))$    ▷ tampering prob., Eq. (6)
      if $|\hat{c} - 0.5| > |\tilde{c} - 0.5|$ then
        $c_i \leftarrow \hat{c}$    ▷ use new score if more confident
        if $|\tilde{c} - 0.5| > \Delta c_2$ and $(\tilde{c} - 0.5)(\hat{c} - 0.5) > 0$ then
          $s \leftarrow S$    ▷ if confident enough and scores agree, stop
        end if
        $\tilde{c} \leftarrow \hat{c}$
      end if
      $s \leftarrow s + 1$    ▷ increase window size
    end while
  end for
  $\mathbf{c} \leftarrow \mathcal{S}(\mathbf{c}, \Delta\omega)$    ▷ sub-sampling & padding
  return $\mathbf{t} \leftarrow \operatorname{argmin}_{\mathbf{t}} \mathrm{E}(\mathbf{t}|\mathbf{c})$    ▷ final decision, Eq. (14)
```

our experiments, we used a small set of candidate scales $\{w_s\}$. The analysis starts by evaluating the tampering probability according to (6b) for the smallest window (in our case, the $\omega_1 = 32$ window). If the window is too small and a confident decision cannot be reached, the window size is increased to the next available scale ω_{s+1}. Such an approach will use smaller windows in more confident, bright and flat areas, and larger windows in darker, more textured regions of the image. In our experiments, we proceed to the next window size if $|c_i - 0.5| < 0.5 - \Delta c_1$. The new tampering probability estimate is accepted if it is more confident than the previous one.

Since unreliably detected small forgeries are likely to be replaced by a confident contrary decision as soon as the window size gets large enough, we use an additional rule which stops increasing the window size if the next (larger) window reinforces a previous, reasonably confident detection ($|c_i - 0.5| > \Delta c_2$). We stop increasing the window size after the largest possible scale is reached - either $\omega_S = 256$ or a smaller one in the vicinity of image borders. The described algorithm is summarized as pseudo-code in Alg. 3.

An example localization result is shown in the bottom row of Fig. 3. Note that the correlation field, although more noisy, represents tampered objects more accurately. It allowed to detect a small inpainting forgery that was missed by other detectors. Just as expected, analysis windows are larger in darker, more textured areas, but quickly get smaller in favorable conditions. Note that the described approach is prone to generate small false positives stemming from inaccurate predictions on the smaller scales. However, we found that introduction of neighborhood interactions by means of a CRF can effectively remove most of the artifacts. The final decision map in Fig. 3 is free of false positives.

V. Experimental Evaluation

Our experimental evaluation covers both synthetic (Section V-D) and realistic (Section V-E) forgeries. Our primary evaluation criterion is the F_1 score:

$$F_1 = \frac{2 \cdot tp}{2 \cdot tp + fn + fp} , \tag{16}$$

where tp, fn, fp denote statistics of the observed true positives, false negatives, and false positives. We summarize localization performance as an average F_1 score (averaged over test images for a given decision threshold). We also consider *peak* F_1 scores which correspond to the maximal F_1 score for each test image (over all possible thresholds τ). For the sake of discussion completeness, we also generate the corresponding receiver operation characteristics (ROC) by sweeping the decision threshold τ over 49 values uniformly distributed in (0,1). Note however, that we do not perform any explicit optimization of ROC performance.

A. Evaluated Detectors

Our evaluation includes four basic variants of PRNU-based localization (described in detail in Sections III and IV).

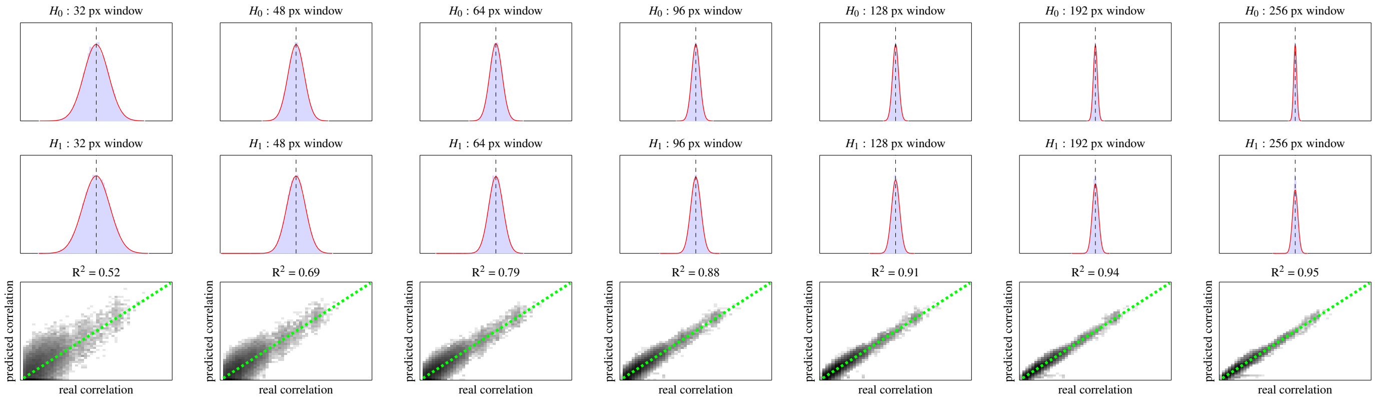


Fig. 4. Visualization of a multi-scale camera model of a Sony SLT $\alpha 57$ camera: (top) empirical distribution of the correlations for the H_0 hypothesis during predictor training and the corresponding Gaussian fit; (middle) empirical distribution of the correlations for the H_1 hypothesis; (bottom) correlation prediction accuracy visualized as a 2-dimensional histogram of the *real vs. prediction* scatter plot; all histograms share the same range for the X axis.

First, we consider a series of conventional single-scale detectors (for various window sizes) with heuristic map post-processing. Secondly, we consider a multi-scale fusion approach (abbreviated as *MSF*) which combines multiple candidate maps obtained with windows $\omega_s \in \Omega = \{32, 48, 64, 96, 128, 192, 256\}$. Finally, we consider two variants of the adaptive-window (*AW*) and segmentation-guided (*SG*) strategies: with heuristic map cleaning and with a CRF-based decision (*AW+* and *SG+*). Having only 2 decision labels, and guaranteed sub-modular potentials ($\beta_0, \beta_1 \geq 0$), we used a graph cuts-based solver [51, 52] from UGM toolbox [53] to quickly find the optimal tampering map minimizing (8). Parameter choice is discussed in Section V-C.

All schemes use central-pixel attribution, which requires special handling near image borders. The remaining unclassified pixels could be filled by padding with repetition. However, we found that better results can be obtained by padding the input image instead (by $\omega/2$ with a mirror reflection in each direction). This allows to populate all pixels within the map and prevents the *adaptive-window* strategy from forcibly using only smallest windows near image borders. In practice, it is unnecessary to move the analysis window by 1 px as such localization resolution is beyond capabilities of PRNU analysis. We used 8 px stride for the considered multi-scale strategies and 4 px stride for the single-scale detectors. For comparison with the ground-truth, all tampering maps are up-sampled to full image size.

B. Data Set Composition

Our data set contains both synthetic and realistic forgeries from four digital cameras: Sony $\alpha 57$, Canon 60D, Nikon D7000, and Nikon D90 (Tab. I). All images were cropped to the middle fragment of size 1920×1080 px (2 Mpx). For all cameras the correlation predictor was trained on 25,000 randomly chosen patches from 50 diverse images. Another 250 diverse photographs were used to generate the synthetic and perform realistic forgeries. Images from the Sony $\alpha 57$ camera originate from a personal photo collection, and include diverse images, taken in various lighting conditions, at various ISO settings. The images were acquired in RAW format, and were converted to TIFF format using *dcraw* software with default settings. The PRNU was obtained from 90 dedicated out-of-focus flat images using the original MLE estimator with wavelet-based denoising [54]. Images from the Canon 60D camera come from a personal photo collection. They were acquired in RAW format and converted to TIFF with Canon's default software. PRNU was estimated from 200 favorable (bright, low-texture) natural images. The remaining photographs (Nikon cameras) were taken from the RAISE dataset [55], carefully cleaned of duplicated photos. We directly obtained TIFF images and estimated the PRNU from 200 favorable images.

TABLE I
SUMMARY OF INCLUDED DIGITAL CAMERAS

Camera	# Images		Source	Predictor
	Tampered	PRNU est.		quality R^2
Sony $\alpha 57$	52	90 flat	own	0.52 - 0.95
Canon 60D	27	200 natural	own	0.65 - 0.86
Nikon D7000	26	200 natural	RAISE	0.47 - 0.80
Nikon D90	31	200 natural	RAISE	0.56 - 0.83

In our multi-scale analysis setting, camera models contain individual $(\sigma_0, \sigma_1, \theta)$ for every considered scale of analysis[1]. A visualization of a multi-scale model for Sony $\alpha 57$ is shown in Fig. 4. The top and middle rows show histograms of correlation scores from predictor training for H_0 and H_1, respectively. The bottom row shows the quality of the obtained predictor, visualized as a 2-dimensional density plot of observed vs. predicted correlations and measured by the R^2 coefficient.

Synthetic forgeries were generated by replacing a randomly located (with pixel-wise accuracy) square region of the input image with a randomly chosen patch from a different camera. In total, we generated 250 forged images per camera for each of the following tampering sizes: 48, 64, 96, 128, 192, and

[1] In principle one could also extrapolate the parameters of the model from a single-scale setting. For three considered cameras, we found the minimum correlation between the parameters of the predictor to be 0.80 (Nikon D7000), 0.90 (Nikon D90) and 0.96 (Sony $\alpha 57$) among 7 considered window sizes. However, in preliminary evaluation we obtained better results by recording individual parameters for each analysis window size.

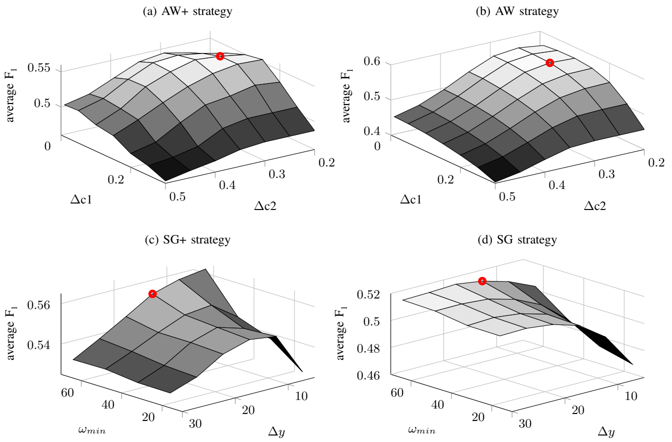


Fig. 5. Results of grid search for the parameters of the alternative adaptive-window and segmentation-guided strategies with both heuristic and CRF decision; the chosen configuration is indicated by a red marker.

256 px. For realistic evaluation, we prepared 136 diverse high-quality forgeries using popular photo editing software (*GIMP* and *Affinity Photo*). The forgeries are of various size and character and include object insertion, object removal and also more subtle changes to existing content, like subtle shadows or reflections, that are unlikely to be detected with PRNU analysis. Example forgeries are shown in Fig. 11 (please refer to supplementary materials for more examples). Ground truth maps were generated as a pixel-wise difference between the pristine and doctored images (cleaned using small-scale morphological filtering, and corrected manually if necessary).

C. Parameter Selection

Parametrization of the considered multi-scale strategies is summarized in Table II along with the values used in our experiments. This section introduces the protocol that we followed to choose the most important parameters.

The adaptive-window approach is controlled with two main parameters $\Delta c_1, \Delta c_2$. We used grid search to assess their impact on the average F_1 score on a randomly chosen subset from our realistic forgeries. To speed up the computations, we measured and averaged the F_1 score for three thresholds $\tau = 0.4, 0.5, 0.6$. We used both variants with heuristic post-processing and with CRF-based decision. The obtained results (top row in Fig. 5) shows very similar behavior in both cases. Finally, we chose a configuration which performs best in both cases (with negligible loss with respect to the individual optima). It is marked on the surface plot with a red circle.

We followed the same protocol for the segmentation-guided strategy to asses the impact of ω_{min} and Δy. In this case, the heuristic post-processing and CRF-based decision reveal somewhat different behavior (bottom row in Fig. 5). The former prefers values of Δy around 20-25 and slightly improves for larger ω_{min}. The latter also improves along with ω_{min}, but prefers smaller pixel similarity thresholds (from 15 down to 5 for larger ω_{min}). Finally, we chose $\Delta y = 15$ which seems to be a good choice for various conditions and incurs a penalty of only 0.006 with respect to the best observed F_1 score.

TABLE II
PARAMETER SUMMARY FOR ALL CONSIDERED LOCALIZATION SCHEMES

Method	Symbol	Parameter	Value
MSF	α	decision bias	-1.00
	β_0	default interaction strength	0.55
	β_1	adaptive interaction strength	5.60
	ϕ	color similarity attenuation	25
	δ	threshold drift	0.18
	Ψ_{min}	minimum data term potential	0.001
	δ_{sep}	threshold drift margin	0.05
	ξ_1	reliability curve parameter	30
	ξ_2	reliability curve parameter	2.5
AW+	α	decision bias	-0.90
	β_0	default interaction strength	0.25
	β_1	adaptive interaction strength	3.50
	ϕ	color similarity attenuation	25
	Ψ_{min}	minimum data term potential	0.001
	Δc_1	immediate score acceptance threshold	0.1
	Δc_2	secondary score acceptance threshold	0.25
SG+	α	decision bias	-0.5
	β_0	default interaction strength	2.0
	β_1	adaptive interaction strength	1.15
	ϕ	color similarity attenuation	25
	Ψ_{min}	minimum data term potential	0.001
	ω_{min}	minimum window / segment size	64
	Δy	pixel similarity threshold	15
	–	segmentation cleaning struct. element	8-px disk

We then focused on the selection of the parameters for the conditional random fields: tampering penalty (α); content independent and adaptive neighborhood interaction penalties (β_0, β_1); and threshold drift strength (δ). While sound interpretation allows forensic analysts to fine-tune these parameters manually, we aim to measure typical localization performance in a fixed setting that does not require manual intervention. In order to choose reasonable parameters, we aim to maximize the average F_1 score on a small set of example forgeries. For best results, this training set should include diverse examples with varied content affected by both object insertion and removal forgeries. In our evaluation, we used the realistic forgeries from the Sony $\alpha 57$ camera.

The number of parameters makes it impractical to perform grid search with reasonable accuracy. Adoption of gradient methods is limited due to non-trivial behavior of the threshold drift (δ), intractable exact optimization of the problem (due to combinatorial complexity of the partition function of the random field), and insufficient correlation of approximations of formal probabilistic optimization criteria with practical localization performance measures (e.g., F_1 score). We obtained limited success with saddle-point approximation of a likelihood objective [56], however, we could not guarantee convergence to optimal parameters (with respect to either F_1 score or classification accuracy).

As a result, we resort to random parameter search which relatively quickly chooses reasonable values, and still allows us to assess the impact of individual parameters. In order to speed up processing, we used a fixed decision threshold $\tau = 0.5$. We sampled parameters' values from the following search space (chosen empirically based on preliminary experiments): $\alpha \leftarrow \mathcal{U}(-5, 5)$, $\beta_0, \beta_1 \leftarrow |\mathcal{U}(-0.5, 7)|_+$, $\delta \leftarrow |\mathcal{U}(-0.025, 0.425)|_+$

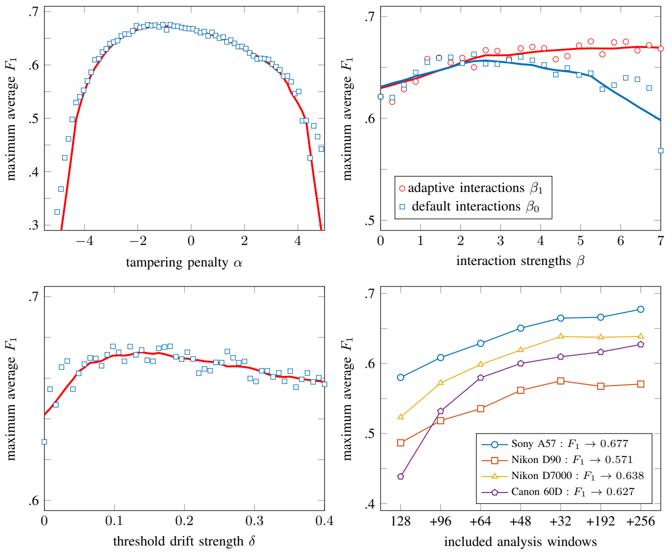


Fig. 6. Impact of individual parameters on the maximum achievable F_1 score for $\tau = 0.5$ for the multi-scale fusion approach: tampering penalty α (top left); neighborhood interactions β (top right, isolated impact); threshold drift δ (bottom left); included analysis windows (bottom right).

where the $|x|_+ := \max(0, x)$ operator truncates numbers to positive values. Setting the lower bound of the search range slightly below zero allowed us to obtain configurations where a given parameter is exactly 0.

In order to validate the improvement of multi-scale analysis, we estimate the maximum average F_1 score for different numbers of available analysis windows. We start with the commonly used 128×128 px window and successively include smaller, and finally larger windows (see Fig. 6). Note that in this experiment parameter selection was performed separately for each camera. Later on, this will allow us to assess the penalty for using parameters chosen for the Sony $\alpha57$ camera. For each camera and each set of analysis windows, we sampled 1,500 random parameter configurations. The obtained results are shown in Fig. 6. It can be observed that the performance consistently improves as successive analysis scales are included. In case of Nikon cameras, we observed saturation of the F_1 scores, which no longer improve after including windows larger than $\omega = 128$. However, larger windows proved beneficial for the remaining cameras and therefore we decided to use the whole set in the remaining experiments.

In order to assess the impact of individual parameters, we sampled 10,000 random parameter configurations (Sony $\alpha57$, all analysis windows). For each parameter, we quantize its values and measure the maximum achievable F_1 score over all remaining parameters, e.g., for quantized value α_c :

$$F_1(\alpha_c) = \max_{\beta_0,\beta_1,\delta} \{F_1(\alpha, \beta_0, \beta_1, \delta) : \mathcal{K}(\alpha) = \alpha_c\} \ , \qquad (17)$$

where $\mathcal{K}$ denotes the utilized quantizer. In order to better visualize the behavior, we plot a moving average of the obtained samples. For neighborhood interaction strengths β, we show their isolated impact, i.e., when the other type of interactions in minimized ($\beta < 0.25$). The obtained results (Fig. 6) show clear preference of the tampering penalty α towards -1.25 and the threshold drift towards 0.15. The neighborhood interaction strengths exhibit slightly more complex behavior. Firstly, both types interaction are advantageous. However, we can observe that the default interaction strength prefers smaller values (up to approx. 2.5). The adaptive interaction strength not only prefers larger values (greater than 4) but also allows to reach better results. Finally, the interaction strengths are negatively correlated, and setting larger value for one of them will typically require making the other one smaller. Hence, the obtained results confirm our previous intuition that introduction of content-dependent neighborhood interactions should be beneficial (see Fig. 2) as it will allow to match the shape of the inserted objects more accurately[2]. We did not observe adverse impact for more subtle object removal forgeries.

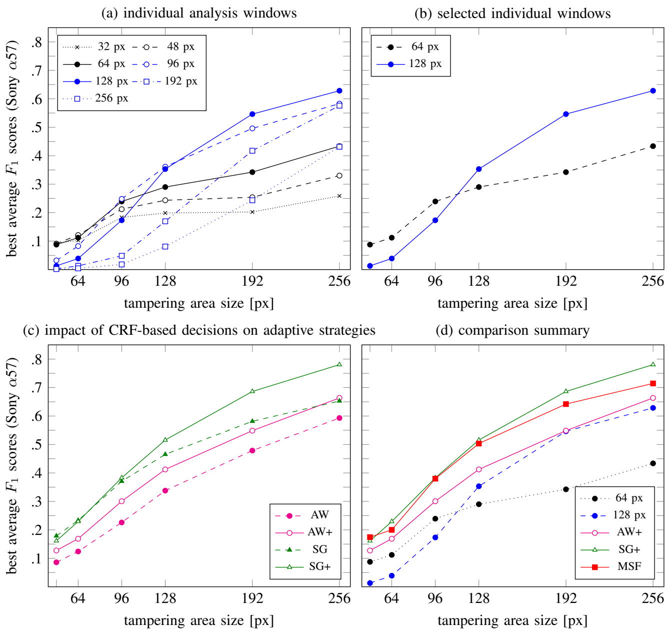


Fig. 7. Localization performance for synthetic forgeries: (a) compares individual single-scale detectors; (b) compares two single-scale detectors with best performance for small and for large forgeries; (c) shows the impact of the CRF-based decision for the segmentation-guided and adaptive-window strategies; (d) compares the multi-scale and single-scale detectors.

The above description addressed CRF parameter selection for the multi-scale fusion approach. We followed an analogous procedure for the remaining strategies. The obtained final parameters (Table II) will be subsequently used for evaluation on both synthetic and realistic forgeries for all cameras.

D. Localization Performance for Synthetic Forgeries

Evaluation on synthetic forgeries reveals the limits of the localization capability for tampering of a given size. Fig. 7a shows how the localization performance changes with the analysis window size for the Sony $\alpha57$ camera (results for the remaining cameras are similar and are included in supplementary materials). As a baseline for comparison, we show

[2] Note that even for the object insertion forgeries, the actual ground truth mask typically does not follow exactly the edges of the object. In most cases, an additional space around the object needed some post-processing to make the forgery more visually appealing.

TABLE III

BEST AVERAGE F_1 SCORES [0-100] OF INDIVIDUAL DETECTORS FOR SYNTHETIC SQUARE FORGERIES OF VARIOUS SIZE.

detector / window size	**Tampered area size [px]**																							
	Sony A57						**Canon 60D**						**Nikon D90**						**Nikon D7000**					
	48	**64**	**96**	**128**	**192**	**256**	**48**	**64**	**96**	**128**	**192**	**256**	**48**	**64**	**96**	**128**	**192**	**256**	**48**	**64**	**96**	**128**	**192**	**256**
32 px window	9	10	18	20	20	26	10	14	20	21	31	34	14	14	23	26	33	33	7	11	15	19	26	31
48 px window	9	12	21	24	25	33	11	18	27	31	41	47	11	16	28	33	38	41	7	15	22	27	37	42
64 px window	9	11	24	29	34	43	9	15	32	38	49	56	8	13	29	37	44	48	5	15	27	34	44	49
96 px window	3	8	25	36	50	58	3	8	22	35	57	66	2	8	24	38	50	55	2	7	18	32	49	59
128 px window	1	4	17	35	55	63	1	3	14	23	56	67	1	2	13	28	48	57	1	3	12	25	47	61
192 px window	0	1	5	17	42	58	0	1	2	8	30	53	0	1	3	10	30	44	0	1	3	11	26	48
256 px window	0	0	2	8	24	43	0	0	1	2	10	28	0	0	1	4	17	30	0	0	1	4	16	31
multi-scale fusion	17	20	38	50	64	71	14	21	34	43	65	74	10	17	33	45	56	62	11	18	27	39	57	67
segmentation-guided (heuristic)	10	20	36	46	58	66	8	15	30	38	58	68	5	11	25	40	51	58	5	16	32	45	57	65
segmentation-guided (CRF)	10	21	39	52	70	80	8	15	31	40	66	78	2	7	20	35	56	70	3	14	30	46	69	80
adaptive-window (heuristic)	9	12	23	34	48	59	11	16	28	34	49	59	7	12	24	33	44	51	6	12	21	29	41	53
adaptive-window (CRF)	13	17	30	41	55	66	12	16	31	39	57	67	7	10	22	33	42	52	7	12	22	31	47	61

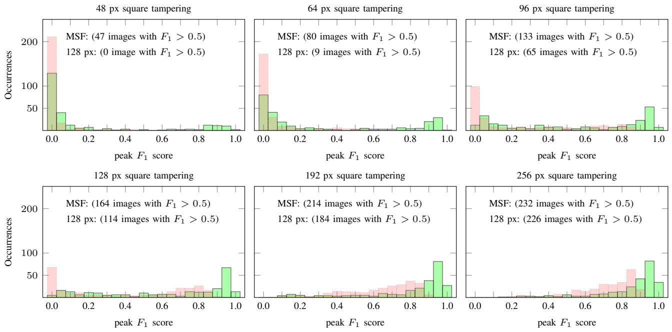


Fig. 8. Histograms of peak F_1 scores for individual images from the Nikon D7000 camera in the synthetic forgery test for the multi-scale fusion approach (green) and the conventional single-scale detector with a 128 px window (red).

the 128 px and 64 px windows which obtain best results for large and small forgeries, respectively. While small windows give a chance to detect smaller forgeries, they are less reliable and perform poorly for large tampering. Complete numerical results for all cameras are collected in Table III.

The considered multi-scale strategies combine the benefits of small-scale and large-scale analysis and consistently outperform the conventional single-scale method across all tampering sizes. The *adaptive-window* and *segmentation-guided* strategies clearly benefit from replacing standard heuristic post-processing with a CRF (Fig. 7c). Introduction of neighborhood dependencies offsets the problems stemming from inaccurate segmentation or scattered small false positives.

Overall, we observed the most stable improvement for the multi-scale fusion strategy, which nearly always performed better than any single-scale detector. While for large forgeries the best results were obtained by the segmentation-guided strategy, it tended to deteriorate in some small forgery cases (e.g., small forgeries for Nikon D90). Compared to the standard $\omega = 128$ window, the above multi-scale strategies always delivered superior performance. Overall, the smallest improvement was observed for the adaptive-window strategy. It typically yielded similar performance as the best single-scale detector. Hence, it also brings some advantages over conventional single-scale analysis.

In order to clearly illustrate potential detectability of small forgeries, we show a histogram of the peak F_1 scores for both the conventional single-scale 128 px window and for the proposed multi-scale fusion approach (Fig. 8). In favorable conditions (bright, low-texture images), thanks to incorporation of small-scale windows, the multi-scale approach can reliably detect ($F_1 \approx 0.9$) forgeries as small as 64×64 px. However, it brings significant benefits for all tampering sizes.

The benefits of all multi-scale strategies can also be observed in the receiver operation characteristics (Fig. 9). For the sake of presentation clarity, we show only the curves for two single-scale detectors (64×64 px and 128×128 px windows). Note that our parameter selection procedure did not involve explicit optimization of ROC performance. While the F_1 score is obviously correlated with classification accuracy, it should be possible to further improve the curves by properly adapting the optimization criterion.

E. Localization Performance for Realistic Forgeries

The obtained localization performance for realistic forgeries is summarized in Fig. 10. The best single-scale detectors used the $\omega = 64$ window (Canon 60D) or the $\omega = 96$ window (other cameras). Improvement over the commonly used $\omega = 128$ ranges from minor to significant (Canon 60D). Fig. 10 shows both aggregated performance statistics (ROC curves, average and peak F_1 scores) and detailed image-level comparison (scatter plots of peak F_1 scores). The scatter plots always use the best single-scale detector as a baseline.

Similarly to synthetic evaluation in Section V-D, the proposed multi-scale strategies deliver considerable benefits. The most stable improvement can be observed for the multi-scale fusion approach. Inspection of scatter plots (4th column) clearly shows that the localization performance improves for most cases (note that the improvement with respect to the commonly used $\omega = 128$ window is even greater). We observed only a few cases with considerably worse results (images *DPP0122*, *DPP0251*, and *DPP0445* from the Canon 60D camera). Further inspection revealed that they were caused by confident false positives in medium scale candidate maps. As expected, the errors were observed in highly-textured areas of low or medium brightness and can be attributed to inadequate performance of the correlation predictor.

Similarly to synthetic evaluation, the best results were often provided by the segmentation-guided strategy, especially for

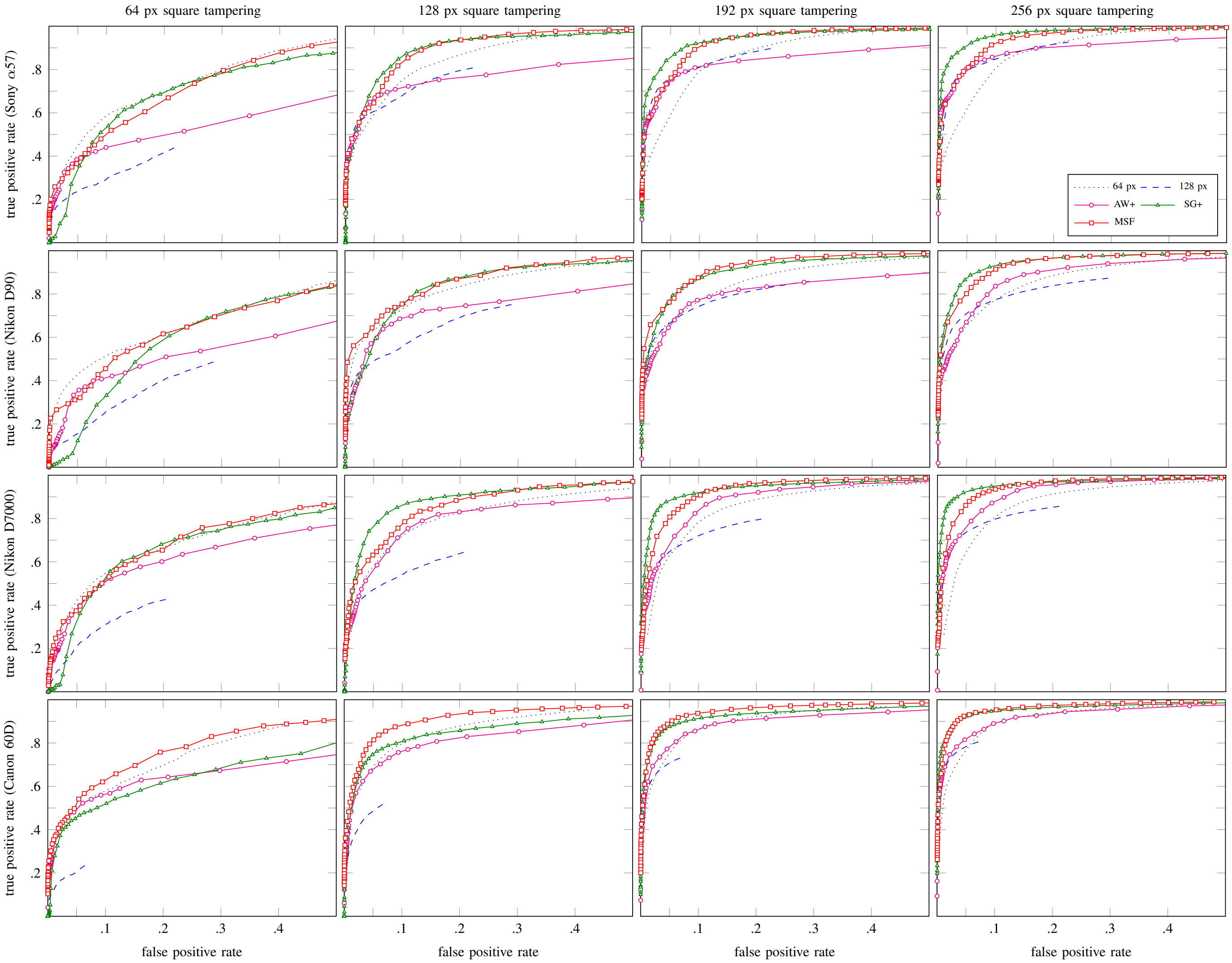


Fig. 9. Receiver operation characteristics from the detection of synthetic square forgeries by all of the considered multi-scale strategies and two representative single-scale detectors with the 64×64 px and the 128×128 px windows; for more detailed results, please refer to supplementary materials.

highly contrasting object insertion forgeries. However, this approach has clearly suffered in more challenging conditions. Inspection of scatter plots (5th column) reveals significantly scattered results with many cases of performance deterioration. We observed that poor results are often caused by highly detailed objects with small and dark areas of irregular shape. Similar problem occur in irregularly shaped saturated areas. Such cases are difficult to handle for both segmentation algorithms and the correlation predictor. However, we expect that better results could be obtained with a predictor capable of handling arbitrary irregular shapes of the segments.

Compared to synthetic evaluation, we observed somewhat better performance of the adaptive-window strategy. In this experiment, it delivered more competitive results with considerable improvement over the standard single-scale approach (scatter plots are available in supplementary materials). Similarly to the multi-scale fusion approach, we observed more consistent results than for the segmentation-guided strategy.

In order to assess sensitivity of the localization performance to the choice of CRF parameters, we compare the obtained scores to the results from analysis window size selection (Fig 6) where we sought the best performance for each camera separately. Table IV compares the best average F_1 score for the adopted parameter choice with the best configuration observed during our random search, and with a configuration where all parameters are zeroed. While we can observe significant performance improvement with respect to the zeroed configuration, the gap from the best parameters is much smaller. While this demonstrates that further improvement can be achieved by fine-tuning the parameters individually, good performance can be expected from a universal parameter configuration.

Example tampering localization results for all strategies are shown in Fig. 11. Successive rows are labeled vertically with filenames of individual forgeries. The figure shows both tampering probability maps and color-coded decision maps. In order to clearly illustrate the best localization potential, the thresholds were chosen individually for each image and correspond to the best achievable F_1 score. The examples are ordered by their peak F_1 score improvement of the MSF strategy with respect to the single scale detector with

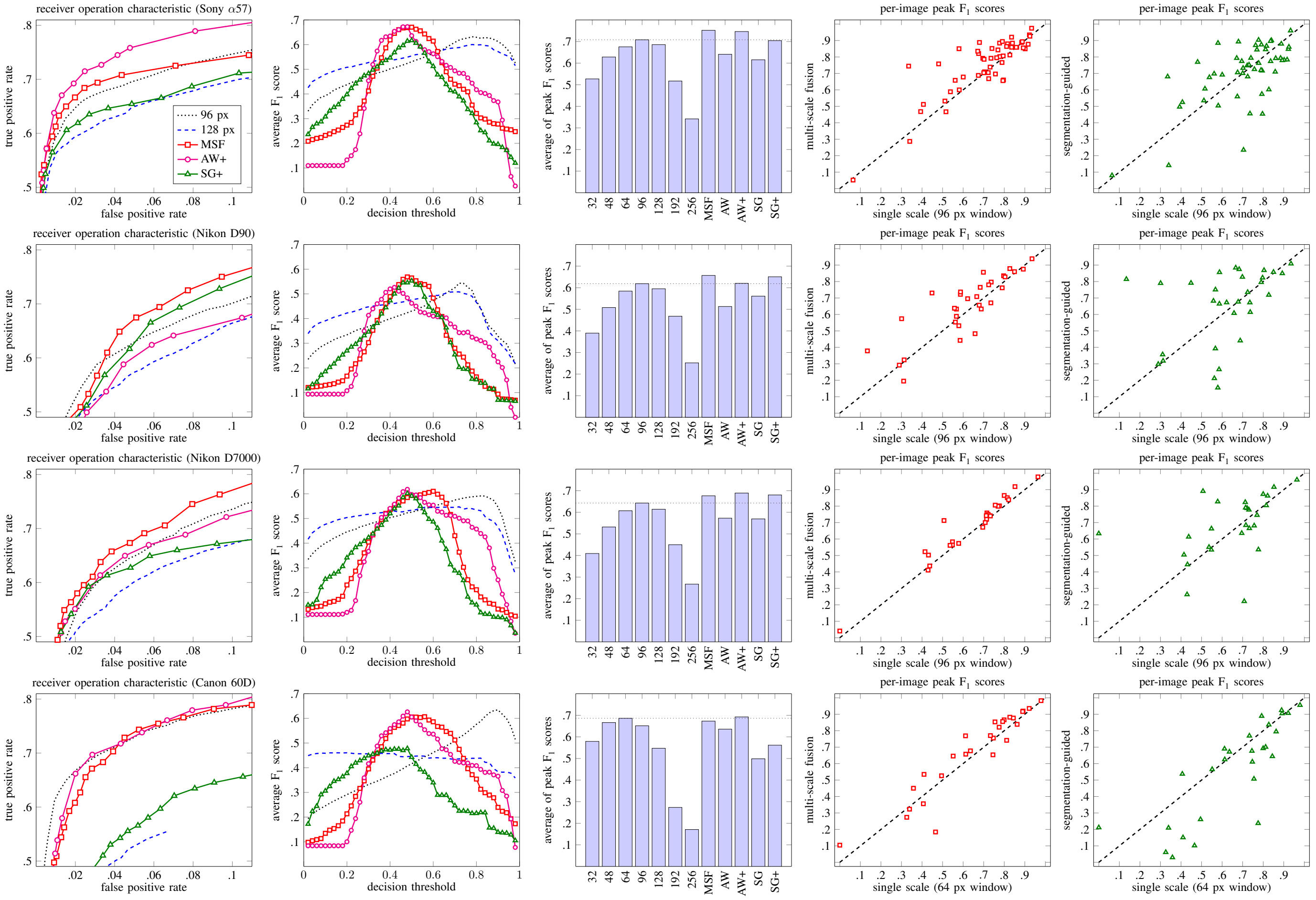


Fig. 10. Tampering localization performance for realistic forgeries; notation: *MSF* - multi-scale fusion; *AW* (*AW+*)- adaptive-window strategy with heuristic (CRF-based) decision; *SG* (*SG+*) - segmentation-guided strategy with heuristic (CRF-based) decision.

TABLE IV
LOCALIZATION PERFORMANCE OF THE MULTI-SCALE FUSION STRATEGY FOR VARIOUS CHOICES OF CRF PARAMETERS.

Camera	**Average F_1 scores for:**		
	zeroed param.	**chosen param.**	**best param.**
Sony α57	0.612	0.671	0.677
Canon 60D	0.539	0.607	0.627
Nikon D90	0.440	0.569	0.571
Nikon D7000	0.537	0.610	0.638

$\omega = 96$ which performed best overall. Images with the greatest performance gain are shown on top. Four examples at the bottom correspond to deteriorated performance. The included examples show all of the mentioned phenomena: superior shape detection of the segmentation-guided strategy in favorable conditions (e.g., *DSC07311*); problems of the segmentation-guided strategy with subtle object removal and high-detail areas (*DSC07004* or *DPP0086*); lack of improvement from segmentation-guidance for object removal forgeries (*DSC05748*); improved detection of small forgeries by the adaptive window strategy (*DSC05810*); content-guided propagation for the multi-scale fusion strategy (*DSC06083*). For more examples please refer to supplementary materials.

F. Robustness Evaluation

In this experiment, we measure the impact of lossy JPEG compression on the best average F_1 score. We consider a joint data set of realistic forgeries from all considered cameras (136 images in total). We generated 6 new versions of every image for JPEG quality factors 75, 80, 85, 90, 95, and 100. We used a common setting with 4:2:0 chroma sub-sampling (horizontal and vertical resolutions of Cb and Cr channels are halved). We used the same detectors (with the same settings) as in previous experiments. Camera models were adjusted by training separate predictors for different quality levels. The PRNU estimate $\hat{\mathbf{k}}$ was left unchanged (trained on TIFF images). During localization, the JPEG quality level was read from meta-data and used to choose the relevant predictor.

The obtained results are shown in Fig. 12. Individual single-scale detectors (for different window sizes ω) are compared in (b). Selected best detectors ($\omega = 96$ and $\omega = 128$) are compared to multi-scale strategies in (a). We can observe that the standard $\omega = 128$ detector achieves the best overall single-scale performance. While the $\omega = 96$ window yielded better results on TIFF and JPEG 100 images, it then quickly deteriorated with increasing compression strength. The smallest scales are no longer useful below quality 90-95 (depending on image content). Interestingly, we observed improvement of

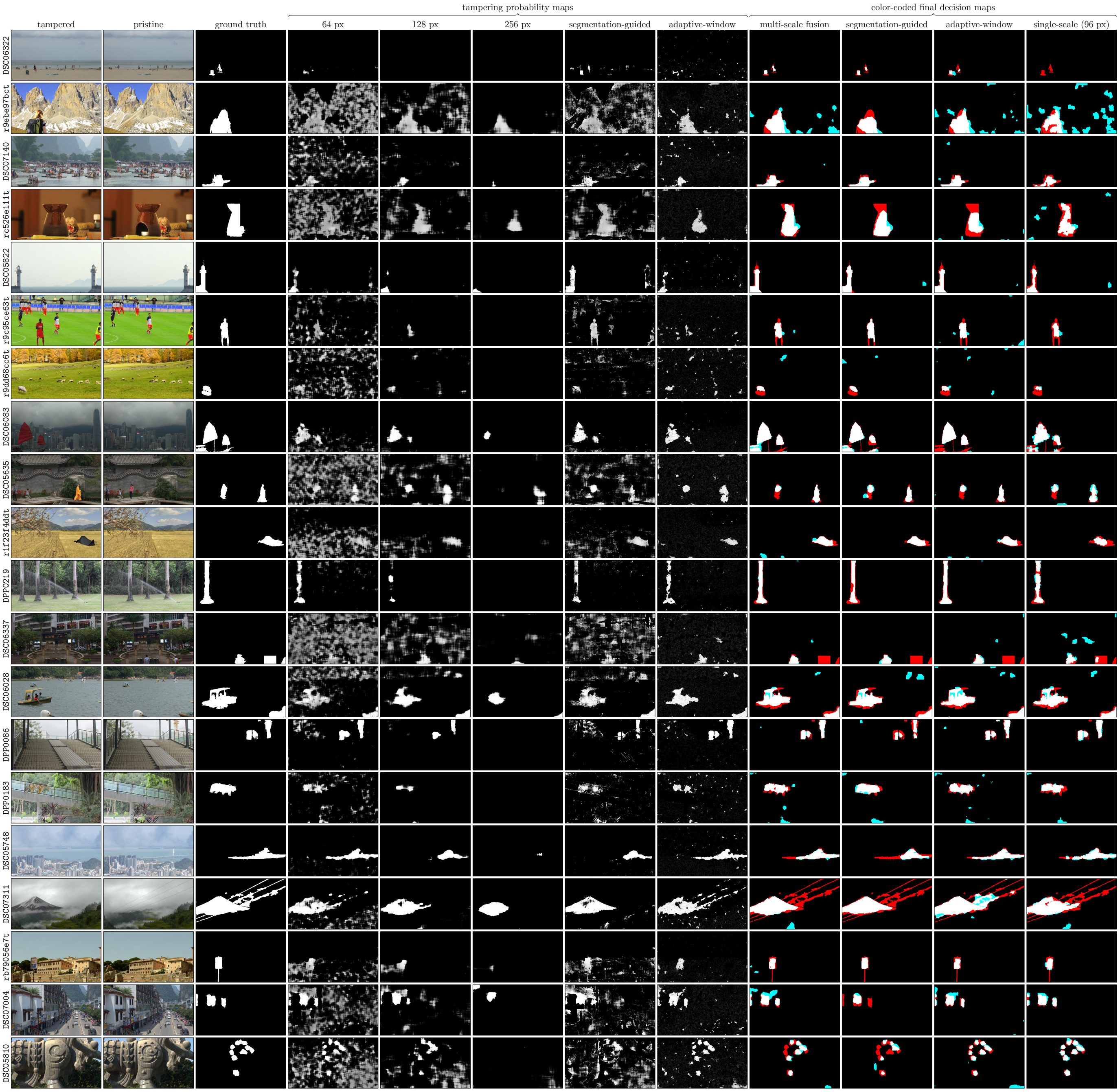


Fig. 11. Example tampering localization results; color coding: *white* - detected tampered regions (*tp*); *red* - undetected tampered regions (*fn*); *cyan* - detected authentic regions (*fp*); *black* - undetected authentic regions (*tn*); for more examples, please refer to supplementary materials.

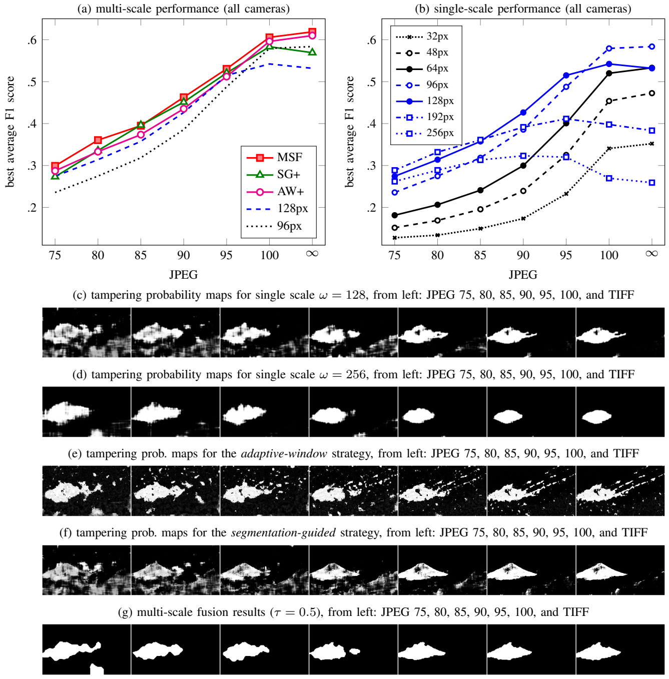
localization performance for the largest windows. An example result illustrating this phenomenon for $\omega = 256$ is shown in (d). Hence, once small-scale windows loose reliability, the multi-scale fusion approach can still extract useful information from windows larger than $\omega = 128$.

It can be observed that our multi-scale fusion strategy delivers the best results with consistent benefits for all JPEG quality levels. While the greatest improvement can be expected for high-quality JPEGs or uncompressed TIFF images, it remains beneficial also for higher compression strengths. The remaining *adaptive-window* and *segmentation-guided* strategies delivered worse robustness with smaller gains and less consistent results. Fig. 12(c)-(g) show changes of tampering localization results for an example forgery (image *DSC07311* from Fig. 11) for all of the considered JPEG quality levels.

VI. Conclusions

In this study, we evaluated 3 strategies for multi-scale analysis in PRNU-based tampering localization. We considered dynamic window size selection, computation of the correlation over coherent image segments, and fusion of separate response maps obtained with various window sizes.

Our *segmentation-guided* strategy was inspired by recent works based on manual segmentation and guided filtering [2,


(a) multi-scale performance (all cameras)

(b) single-scale performance (all cameras)

(c) tampering probability maps for single scale $\omega = 128$, from left: JPEG 75, 80, 85, 90, 95, 100, and TIFF

(d) tampering probability maps for single scale $\omega = 256$, from left: JPEG 75, 80, 85, 90, 95, 100, and TIFF

(e) tampering prob. maps for the *adaptive-window* strategy, from left: JPEG 75, 80, 85, 90, 95, 100, and TIFF

(f) tampering prob. maps for the *segmentation-guided* strategy, from left: JPEG 75, 80, 85, 90, 95, 100, and TIFF

(g) multi-scale fusion results ($\tau = 0.5$), from left: JPEG 75, 80, 85, 90, 95, 100, and TIFF

Fig. 12. Impact of lossy JPEG compression on tampering localization performance (marker at ∞ corresponds to uncompressed TIFF images: (a) multi-scale strategies vs. selected single-scale detectors; (b) single-scale strategies on various scales of analysis; (c) - (g) tampering probability maps for an example forgery (image *DSC07311*).

3]. The proposed approach is fully automatic and uses a CRF to introduce dependencies between neighboring regions of the image. Such content-guided approaches are particularly beneficial when the forgery involves insertion of a highly contrasting object, especially on a bright and flat background. In such conditions, we can detect even small objects, and precisely delineate their boundaries. However, problems occur for more complex and subtle forgeries (e.g., involving object removal), especially in highly textured areas where presence of many edges disrupts content segmentation.

In contrast to a similar evaluation of the guided-filtering approach [3], we obtained consistent performance improvement for all tampering sizes. The improvement may partly stem from more precise control over the scope of the correlation statistics and adaptation of conditional distribution models. However, further work is needed. While variances of the distributions can be adjusted, it remains an open problem to generate reliable predictions for irregular image segments. We also observed that despite explicit care for potentially unreliable regions in the predictor, confident false positives still occur. At the moment, it may be beneficial to use a separate detector and either cross-reference the results or include this information directly in the localization algorithm, e.g., as reliability maps, like in our multi-scale fusion approach.

Overall, our *multi-scale fusion* approach proved to be the most versatile. It yielded the most stable improvement in various conditions and always provided significant improvement over the commonly used single-scale detector with $\omega = 128$. While the alternative adaptive-window approach could also improve localization performance, it performed worse in general and proved to be more vulnerable to correlation modeling errors on smaller scales. Availability of candidate maps from multiple-scales gives the fusion approach more information and flexibility. Finally, introduction of content-dependent neighborhood interactions can yield similar benefits as explicit use of image segmentation. It not only delivers superior shape representation, but can also be easily used with arbitrary forensic detectors. In our experiments, it proved to be beneficial for all considered localization strategies.

References

[1] J. Fridrich, “Digital image forensics,” *Signal Processing Magazine, IEEE*, vol. 26, no. 2, pp. 26–37, 2009.

[2] G. Chierchia, S. Parrilli, G. Poggi, L. Verdoliva, and C. Sansone, “PRNU-based detection of small-size image forgeries,” in *Proc. of Int. Conf. on Digital Signal Processing*, 2011.

[3] G. Chierchia, D. Cozzolino, G. Poggi, C. Sansone, and L. Verdoliva, “Guided filtering for PRNU-based localization of small-size image forgeries,” in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2014, pp. 6231–6235.

[4] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[5] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, “A Bayesian-MRF approach for PRNU-based image forgery detection,” *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 554–567, 2014.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[7] X. Lin and C. T. Li, “Preprocessing reference sensor pattern noise via spectrum equalization,” *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 126–140, 2016.

[8] X. Kang, Y. Li, Z. Qu, and J. Huang, “Enhancing source camera identification performance with a camera reference phase sensor pattern noise,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 393–402, 2012.

[9] C. T. Li, “Source camera identification using enhanced sensor pattern noise,” *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 280–287, 2010.

[10] Y. Hu, C. Jian, and C. T. Li, “Using improved imaging sensor pattern noise for source camera identification,” in *Proc. of IEEE Int. Conf. on Multimedia & Expo*, 2010, pp. 1481–1486.

[11] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, “Determining image origin and integrity using sensor noise,” *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 74–90, 2008.

[12] P. Korus and J. Huang, “Multi-scale fusion for improved localization of malicious tampering in digital images,” *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1312–1326, 2016.

[13] B. Li, Y.Q. Shi, and J. Huang, “Detecting doubly compressed JPEG images by using mode based first digit features,” in *Proc. of IEEE Workshop on Multimedia Signal Processing*, 2008, pp. 730–735.

[14] M. K. Johnson and H. Farid, “Exposing digital forgeries in complex lighting environments,” *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 450–461, 2007.

[15] T. J. d. Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. d. R. Rocha, “Exposing digital image forgeries by illumination color classification,” *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 7, pp. 1182–1194, 2013.

[16] K. Bahrami, A. C. Kot, L. Li, and H. Li, “Blurred image splicing localization by exposing blur type inconsistency,” *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 999–1009, 2015.

[17] M. P. Rao, A. N. Rajagopalan, and G. Seetharaman, “Harnessing motion blur to unveil splicing,” *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 583–595, 2014.

[18] V. Conotter, G. Boato, and H. Farid, “Detecting photo manipulation on signs and billboards,” in *2010 IEEE International Conference on Image Processing*, 2010, pp. 1741–1744.

[19] W. Zhang, X. Cao, Y. Qu, Y. Hou, H. Zhao, and C. Zhang, “Detecting and extracting the photo composites using planar homography and graph cut,” *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 544–555, 2010.

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

[21] A.E. Dirik and N. Memon, "Image tamper detection based on demosaicing artifacts," in *Proc. of IEEE Int. Conf. on Image Processing*, 2009, pp. 1497–1500.

[22] Y. F. Hsu and S. F. Chang, "Camera response functions for image forensics: An automatic algorithm for splicing detection," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 816–825, 2010.

[23] H. Yao, S. Wang, X. Zhang, C. Qin, and J. Wang, "Detecting image splicing based on noise level inconsistency," *Multimedia Tools and Applications*, pp. 1–23, 2016.

[24] Z. Lin, J. He, X. Tang, and C. K. Tang, "Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis," *Pattern Recognition*, vol. 42, no. 11, pp. 2492 – 2501, 2009.

[25] T. Bianchi and A. Piva, "Image forgery localization via block-grained analysis of jpeg artifacts," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 1003–1017, 2012.

[26] B. Li, T.-T. Ng, X. Li, S. Tan, and J. Huang, "Statistical model of JPEG noises and its application in quantization step estimation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1471–1484, 2015.

[27] D. Cozzolino, G. Poggi, and L. Verdoliva, "Splicebuster: a new blind image splicing detector," in *Proc. of IEEE Int. Workshop on Inf. Forensics and Security*, 2015.

[28] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1841–1854, 2012.

[29] K. Wattanachote, T. K. Shih, W. L. Chang, and H. H. Chang, "Tamper detection of jpeg image due to seam modifications," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2477–2491, 2015.

[30] H. D. Yuan, "Blind forensics of median filtering in digital images," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1335–1345, 2011.

[31] S. Prasad and K. R. Ramakrishnan, "On resampling detection and its application to detect image tampering," in *Proc. of IEEE Int. Conf. on Multimedia & Expo*, 2006, pp. 1325–1328.

[32] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 215–224, 2010.

[33] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, 2012.

[34] D. Cozzolino, D. Gragnaniello, and L. Verdoliva, "Image forgery localization through the fusion of camera-based, feature-based and pixel-based techniques," in *Proc. of IEEE Int. Conf. on Image Processing*, 2014, pp. 5302–5306.

[35] L. Verdoliva, D. Cozzolino, and G. Poggi, "A feature-based approach for image tampering detection and localization," in *Proc. of IEEE Int. Workshop on Inf. Forensics and Security*, 2014.

[36] X. Qiu, H. Li, W. Luo, and J. Huang, "A universal image forensic strategy based on steganalytic model," in *ACM Information Hiding and Multimedia Security Workshop*, 2014, pp. 165–170.

[37] K. Wang W. Fan and F. Cayre, "General-purpose image forensics using patch likelihood under image statistical models," in *Proc. of IEEE Int. Workshop on Inf. Forensics and Security*, 2015.

[38] S. Bayram, İ. Avcıbaş, B. Sankur, and N. Memon, "Image manipulation detection," *Journal of Electronic Imaging*, vol. 15, no. 4, pp. 041102–041102–17, 2006.

[39] H. Cao and A. C. Kot, "Manipulation detection on image patches using fusionboost," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 992–1002, 2012.

[40] M. Barni and A. Costanzo, "Dealing with uncertainty in image forensics: a fuzzy approach," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2012, pp. 1753–1756.

[41] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A Framework for Decision Fusion in Image Forensics Based on Dempster-Shafer Theory of Evidence," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 4, pp. 593–607, 2013.

[42] L. Gaborini, P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Multi-clue image tampering localization," in *Proc. of IEEE Int. Workshop on Inf. Forensics and Security*, 2014, pp. 125–130.

[43] D. Cozzolino, F. Gargiulo, C. Sansone, and L. Verdoliva, "Multiple classifier systems for image forgery detection," in *Image Analysis and Processing*, vol. 8157 of *LNCS*, pp. 259–268. 2013.

[44] P. Ferrara, M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "Unsupervised fusion for forgery localization exploiting background information," in *Proc. of IEEE Int. Conf. on Multimedia & Expo Workshops*, 2015.

[45] Y. L. Chen and C. T. Hsu, "What has been tampered? from a sparse manipulation perspective," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing*, 2013, pp. 123–128.

[46] P. Korus and J. Huang, "Improved tampering localization in digital image forensics based on maximal entropy random walk," *IEEE Signal Processing Letters*, vol. 23, no. 1, 2016.

[47] J. Fridrich M. Goljan, "Camera identification from scaled and cropped images," in *SPIE - Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents*, 2008.

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

[49] M. K. Mihcak, I. K., and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, 1999, vol. 6, pp. 3253–3256 vol.6.

[50] M.-M. Cheng, V. A. Prisacariu, S. Zheng, P. H. S. Torr, and C. Rother, "DenseCut: Densely connected CRFs for realtime GrabCut," *Computer Graphics Forum*, vol. 34, no. 7, 2015.

[51] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

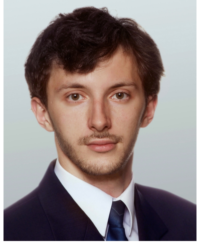
[52] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, 2004.

[53] M. Schmidt, "UGM: A matlab toolbox for probabilistic undirected graphical models," http://www.cs.ubc.ca/~schmidtm/Software/UGM.html, 2011 version.

[54] "DDE laboratory," http://dde.binghamton.edu/, visited in Sept. 2015.

[55] D. T. Dang-Nguyen, C Pasquini, V. Conotter, and G. Boato, "RAISE - a raw images dataset for digital image forensics," in *Proc. of ACM Multimedia Systems*, 2015.

[56] S. Kumar, J. August, and M. Hebert, "Exploiting inference for approximate parameter learning in discriminative fields: An empirical study," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 3757 of *LNCS*, pp. 153–168. 2005.



Paweł Korus (S'09-M'13) received his M.Sc. and Ph.D. degrees in telecommunications (both with honors) from the AGH University of Science and Technology in 2008, and in 2013, respectively. Since 2014 he has been an assistant professor with the Department of Telecommunications, AGH University of Science and Technology, Krakow, Poland. He is currently a postdoctoral researcher with the College of Information Engineering, Shenzhen University, Shenzhen, China.

His research interests include various aspects of multimedia security & image processing, with particular focus on digital image forensics, content authentication, digital watermarking & information hiding. In 2015 he received a scholarship for outstanding young scientists from the Polish Ministry of Science and Higher Education.

Jiwu Huang (M'98–SM'00-F'16) received the B.S. degree from Xidian University, Xi'an, China, in 1982, the M.S. degree from Tsinghua University, Beijing, China, in 1987, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 1998. He was with the School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China. He is currently a Professor with the College of Information Engineering, Shenzhen University, Shenzhen, China.

His current research interests include multimedia forensics and security. He is also a member of the IEEE Circuits and Systems Society Multimedia Systems and Applications Technical Committee and the IEEE Signal Processing Society Information Forensics and Security Technical Committee. He served as an Associate Editor of the IEEE Transactions on Information Forensics and Security from 2010 to 2014. He was a General Co-Chair of the IEEE Workshop on Information Forensics and Security in 2013. He is a Fellow of IEEE.