This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

PRNU-Based Content Forgery Localization Augmented With Image Segmentation

XUFENG LIN[1] and CHANG-TSUN LI[1] (Senior Member, IEEE)
[1]School of Information Technology, Deakin University, Waurn Ponds Campus, Geelong, VIC 3216, Australia

Corresponding author: Xufeng Lin (e-mail: xufeng.lin@deakin.edu.au).

This work is supported by the EU Horizon 2020 project, entitled Computer Vision Enabled Multimedia Forensics and People Identification (Project no. 690907; Acronym: IDENTITY).

ABSTRACT The advances in image editing and retouching technology have enabled an unskilled person to easily produce visually indistinguishable forged images. To detect and localize such forgeries, many image forensic tools rely on visually imperceptible clues, e.g. the subtle traces or artifacts introduced during image acquisition and processing, and slide a regular, typically square, detection window across the image to search for discrepancies of specific clues. Such a sliding-window paradigm confines the explorable neighborhood to a regular grid and inevitably limits its capability in localizing forgeries in a broad range of shapes. While image segmentation that generates segments adhering to object boundaries might be a promising alternative to the sliding window-based approach, it is generally believed that the potential of the segmentation-based detection scheme is hindered by object removal forgeries where meaningful segments are often unavailable. In this work, we take forgery localization based on photo-response non-uniformity (PRNU) noise as an example and propose a segmentation-based forgery localization scheme that exploits the local homogeneity of visually indiscernible clues to mitigate the limitations of existing segmentation approaches that are merely based on visually perceptible content. We further propose a multi-orientation localization scheme that integrates the forgery probabilities obtained with image segmentation and multi-orientated detection windows. The multi-orientation scheme aggregates the complementary strengths of image segmentation and multi-oriented detection window in localizing the *object insert* and *object removal* forgeries. Experimental results on a public realistic tampering image dataset demonstrate that the proposed segmentation-based and multi-orientation forgery localization schemes outperform existing state-of-the-art PRNU-based forgery localizers in terms of both region and boundary F_1 scores.

INDEX TERMS Digital image forensics, image forgery localization, image segmentation, multimedia security, multi-orientation detection, photo-response non-uniformity noise

I. INTRODUCTION

The ease of manipulating images with increasingly powerful image editing tools has also led to the growing appearance of digitally altered forgeries, which raise serious concerns about the credibility of digital images. Developing image forgery detection and localization techniques, therefore, becomes essential under the circumstances where digital images serve as critical evidence. A variety of approaches have been proposed to detect and localize image forgeries. Active techniques such as digital watermarking [1, 2] are effective in verifying the authenticity of an image, but the requirement of embedding additional information at the creation of the image limits their widespread use. On the contrary, passive or blind techniques rely on the image data itself without requiring any pre-embedded information and thus have attracted great interests over the past two decades. The passive image forgery detection and localization techniques in the literature mainly fall into five categories:

(1) The first category is based on specific statistical properties of natural images such as higher-order wavelet statistics [3, 4], image texture features [5] and residual-based features [6, 7].

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

(2) The second category includes the techniques that seek the traces left by specific image manipulations such as resampling [8, 9], contrast enhancement [10, 11], median filtering [12], copy-move manipulations [13] and JPEG compression [14–16].
(3) The third category relies on the regularities or constraints that make images physically plausible. For instance, anomalies in lighting [17], shadows [18, 19], reflections [20] and perspective constraints [21] have been exploited for exposing image manipulations.
(4) The fourth category exploits the artifacts introduced in the image acquisition pipeline of digital camera. These artifacts can be either caused by specific processing components such as demosaicking artifacts [22], camera response function abnormalities [23, 24], optical aberrations [25] and imaging sensor noise [26–28], or caused by the complex interplay of multiple components, such as local noise levels [29].
(5) Inspired by the success of convolutional neural networks (CNN) in various multimedia tasks, the fifth category comprises the techniques that automatically learn the features for image forgery detection or localization using CNN [30–32].

Besides the above-mentioned techniques, recent years have seen some works on fusing the outputs of multiple forensic detectors, e.g. under the framework of fuzzy logic [33], the Dempster-Shafer theory of evidence (DSTE) [34], simple logical rules [35, 36] or Bayesian inference [37], but the performance of the overall system depends on that of each individual forensic detector. To localize the forgeries in an image, a prevailing paradigm is to move a regular, typically square, detection window across the image and analyze the forensic clues in the detection window at different locations. Such a sliding-window paradigm allows for the convenient and efficient processing of the image but also has two inherent limitations: 1) The regular shape of the detection window hinders the detection of forgeries in various possible shapes. For instance, a square detection window is not suitable for detecting tree branches or human limbs. 2) Due to the object boundary unawareness of the sliding-window detection scheme, pixels of different natures, e.g. some pixels are forged and others are pristine, are often contained in the same window. Jointly processing those 'heterogeneous' pixels without any distinction makes the detection highly error-prone.

The above limitations can be alleviated by image segmentation. Local structure of visual content is typically exploited by segmentation algorithms to divide an image into a number of 'segments' or 'superpixels', each of which consists of a group of connected pixels that are perceptually or statistically similar. The resultant superpixels provide informative clues for inferring the boundaries of the forged regions. However, given the variability of possible forgeries, the limitations of image segmentation are also apparent: 1) There might be no distinct local structure information in homogeneous regions to guide the segmentation. 2) The boundaries of the forged regions do not always align with the object boundaries. For instance, it is not uncommon to copy some collateral background pixels along with the forged object to make it blended into the background more naturally. In such cases, pixels of different natures, i.e. the collateral forged pixels and their neighboring authentic pixels, might be grouped into the same superpixel and cause further uncertainties.

In this paper, we take forgery localization based on photo-response non-uniformity (PRNU) noise as an example and show that it is possible to mitigate the limitations of segmentation merely based on visual content by incorporating the local homogeneity of visually imperceptible clues. The main contributions of our work can be summarized as follows.

- We identify the essential criteria that a superpixel segmentation algorithm should satisfy for the task of forgery localization and propose a multi-scale segmentation-based localization scheme in cooperation with some segmentation algorithm that fulfills the identified criteria.
- We propose a novel superpixel segmentation algorithm that encodes both visual content and visually imperceptible clues for content forgery localization. Different from all previous works that only use visual boundary information for image segmentation, the proposed algorithm incorporates the information from visually imperceptible forensic clues to guide the segmentation for forgery localization.
- We propose a multi-orientation fusion strategy that aggregates the complementary strengths of image segmentation and multi-oriented detection in localizing various types of image forgeries.
- We conduct comprehensive experiments on the effects of various segmentation algorithms on different forgery types, i.e. hard-boundary and soft-boundary forgeries, in terms of both region and boundary F_1 scores.

The rest of this paper is organized as follows. In Section II, we will revisit the background of PRNU-based forgery localization and some related work. In Section III, the details of the proposed segmentation-based and multi-orientation forgery localization schemes will be given. Section IV presents the experimental results and analysis on realistic forgeries. Finally, Section V concludes the work.

II. RELATED WORK

In this section, we will first revisit the background of the PRNU-based forgery localization. PRNU noise mainly arises from the inhomogeneity of silicon wafers introduced during imaging sensor fabrication and manifests itself as the pixel-to-pixel variation in light sensitivity. It is present in every image and practically unique to the sensor that captures the image. Therefore, its absence can be used to signify the forgery in an image under investigation, provided that the reference PRNU $\boldsymbol{z}$ of the source camera is available for comparison. For this purpose, a binary hypothesis test is typically carried out for each pixel i in the image's noise

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

residual $\boldsymbol{\omega}$, i.e. the difference between the original image and its denoised version:

$$\begin{cases} h_0 : \omega_i = \upsilon_i \\ h_1 : \omega_i = z_i + \upsilon_i \end{cases} \tag{1}$$

where υ_i is Gaussian random noise. If pixel i is forged, z_i is supposed to be absent in ω_i, which corresponds to the null hypothesis h_0. Otherwise, the alternative hypothesis h_1 is accepted.

Due to the weak nature of $\boldsymbol{z}$, the above hypothesis testing usually requires joint processing of a large amount of pixels, so it is typically conducted in a sliding-window fashion by calculating the normalized cross-correlation ρ_i over the pixels within a $w \times w$ detection window Ω_i centered at each pixel i:

$$\rho_i = \frac{\sum_{j\in\Omega_i}(\omega_j - \bar{\omega})(z_j - \bar{z})}{\sqrt{\sum_{j\in\Omega_i}(\omega_j - \bar{\omega})^2}\sqrt{\sum_{j\in\Omega_i}(z_j - \bar{z})^2}}, \tag{2}$$

where the overbar stands for the respective arithmetic mean within the detection window. Provided that the distributions of the test statistic ρ_i under h_0 and h_1, i.e. $p(\rho_i|h_0)$ and $p(\rho_i|h_1)$, are available, either Neyman-Person [26] criterion or Bayes' rule [27, 28] can be used to make the final decision on the authenticity of pixel i. We will adopt the Bayes' rule as in [27, 28] for the easier formulation of multi-clue fusion under a uniform probability framework. We model $p(\rho_i|h_0)$ as a zero-mean Gaussian distribution $\mathcal{N}(0, \hat{\sigma}_0^2)$ and estimate the variance using the images captured by the cameras that are different from the source camera. While $p(\rho_i|h_1)$ can also be modeled as a Gaussian distribution $\mathcal{N}(\hat{\rho}_i, \hat{\sigma}_1^2)$, its estimation is more challenging since the expected correlation $\hat{\rho}_i$ is highly dependent on the image content. To address this problem, we adopt the correlation predictor [26] based on local image features, namely the image intensity, texture, signal flattening, texture-intensity features and their second-order terms, to estimate the mean $\hat{\rho}_i$ and variance $\hat{\sigma}_1^2$ of $p(\rho_i|h_1)$. Finally, the forgery probability of pixel i is formulated as

$$\begin{aligned} P_i &= \frac{p(\rho_i|h_0)}{p(\rho_i|h_0) + p(\rho_i|h_1)} \\ &= \left(1 + \exp\left(\frac{\rho_i^2}{2\hat{\sigma}_0^2} - \frac{(\rho_i - \hat{\rho}_i)^2}{2\hat{\sigma}_1^2} - \ln\frac{\hat{\sigma}_1}{\hat{\sigma}_0}\right)\right)^{-1}. \end{aligned} \tag{3}$$

In the above sliding window-based framework, the test statistic ρ_i is calculated over all the pixels within the detection window. It becomes problematic if the detection window falls across the boundary between the pristine and the forged regions. In such case, both the pristine and forged pixels contribute to the calculation of ρ_i, which gives rise to the chance of the pixels near the boundaries being undetected. This problem can be alleviated by means of hard segmentation, as shown in Chierchia *et al.*'s work [38], but its rely on accurate manual segmentation makes it infeasible in most practical scenarios. They further proposed a method [39] based on guided filtering [40] to enforce the calculation of ρ_i more aligned with the boundaries of objects. However, its effectiveness relies on the visually perceptible boundary information of the image and thus is largely compromised for *object removal* forgeries where no meaningful boundary information is available. In light of the fact that meaningful forgeries usually appear in a group of connected pixels, another direction of improvement lies in taking into account pixels' spatial dependencies in the final decisions. For instance, Chierchia *et. al* [27] sought to improve the localization performance by modeling the decision statistic as a Markov Random Field (MRF) and recasting the forgery localization as an optimization problem under the Bayesian framework. This method only considers the neighborhood consistency of the final decisions but ignores the content similarity within the neighborhood.

To allow for localizing small-sized forgeries, Korus and Huang [28] proposed a segmentation-based strategy that calculates the test statistic ρ_i only using the pixels with similar intensity levels as the central pixel i, as well as two fusion strategies, i.e. multi-scale and adaptive-window, that combine multiple candidate forgery probability maps obtained with detection windows of different sizes. Additionally, they adopted the Conditional Random Fields (CRF) model to incorporate the content-dependent neighborhood interactions into the final decisions. Although only the intensity difference between individual pixels is considered, the incorporation of image content in the simple segmentation-based strategy and the CRF-based model results in significant performance improvement [16, 28]. Further along this path, it poses an interesting question that whether further improvement can be gained if the image content is exploited in a more sophisticated and comprehensive way. This motivates us to resort to explicit image segmentation that fully exploits the local structure and homogeneity of the image for improving the forgery localization performance.

III. PROPOSED METHODS

A. SEGMENTATION-BASED FORGERY LOCALIZATION SCHEME

For PRNU-based forgery localization, a straightforward way to exploit image content would be segmenting the image into a number of superpixels according to the visual content and calculating the forgery probability for each superpixel. However, compared to the methods based on regular detection windows that are able to generate pixel-wise probability maps, this will result in a low-resolution forgery probability map because the pixels belonging to the same superpixel are assigned with the same probability. Consequently, the chance of mis-detection for *object removal* forgeries will be considerably increased if the size of the superpixel is not appropriately specified. For this reason, we apply image segmentation at different scales and fuse the resultant forgery probability maps to form a single informative probability map. The framework of the proposed segmentation-based multi-scale localization scheme is illustrated in Fig. 1.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

1) Multi-Scale Image Segmentation: Given the variety of superpixel segmentation algorithms, not all of them suffice for our purpose. We identify a few important criteria that the superpixel segmentation algorithms should satisfy for the task of forgery localization:

- Boundary adherence. This criterion measures the agreement between the boundaries of the objects and the resultant superpixels. A superpixel algorithm with good boundary adherence effectively avoids segmenting different objects or different parts of an object into the same superpixel, thus reducing the risk of generating heterogeneous superpixels containing both pristine and forged pixels for *object insert* forgeries.
- Controllability over superpixel number. Some algorithms do not provide direct control over the number of generated superpixels. As the segmentation needs to be carried out at different scales, easy control over the number of generated superpixels is preferable.
- Balanced segmentation granularity. Some superpixel algorithms generate superpixels of heavily unbalanced size. The over-sized superpixels substantially increase the chance of generating heterogeneous superpixels while the under-sized superpixels reduce the reliability of the detection. Thus, an algorithm capable of generating superpixels of balanced size is desirable.

Based on these criteria, we shortlist three superpixel algorithms for multi-scale segmentation:

- Simple Linear Iterative Clustering (SLIC) [41]: SLIC algorithm iteratively updates superpixel centers and assigns pixels to their closest centers in the 5-dimensional pixel color and coordinate space until the algorithm converges. Its simplicity, computational efficiency and high-quality overall segmentation results make it one of the most widely used superpixel algorithms in various applications.
- Entropy Rate Superpixel segmentation (ERS) [42]: ERS algorithm considers each pixel as a vertex in a graph and formulates the segmentation as an optimization problem of graph topology. It incrementally optimizes an object function consisting of the entropy rate of random walks on the graph and a balancing term to achieve homogeneous and similar-sized superpixels. ERS exhibits remarkable boundary adherence, which is desirable for localizing *object insert* forgeries.
- Extended Topology Preserving Segmentation (ETPS) [43, 44]: ETPS algorithm initially partitions the image into the desired number of regular superpixels and continuously modifies the boundary pixels in a coarse-to-fine manner by optimizing an objective function that encodes information about colors, positions, boundaries and topologies. The diversity of information encoded in the objective function and the efficient coarse-to-fine optimization make ETPS one of the state-of-the-art superpixel algorithms in terms of both segmentation quality and efficiency.

For more details about the above superpixel algorithms, we refer the reader to the original papers. Note that the multi-scale framework in Fig. 1 is similar to the work of Zhang *et al.* [45]. However, their analysis is only limited to the SLIC algorithm and falls short in the impact of segmentation on different types of image forgeries.

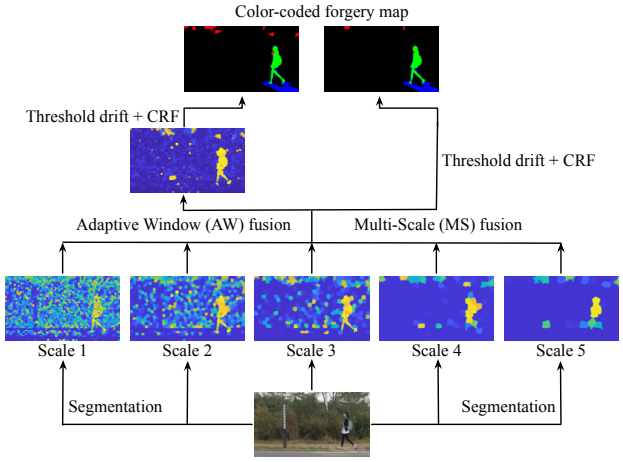


FIGURE 1: The segmentation-based forgery localization scheme. For more details about the Adaptive Window (AW) and Multi-Scale (MS) fusion algorithms, please refer to Section IV-C or [28].

2) Exploiting Homogeneity of PRNU: As mentioned in Section II, only relying on the visual content for image segmentation might become problematic for *object removal* forgeries where no distinct visual information is available. Although this problem can be mitigated by the multi-scale strategy, it would be beneficial if the additional information from PRNU can be incorporated to guide the segmentation at each scale. The perceptually invisible PRNU not only provides useful clues when salient visual information is unavailable but also serves as a supplement to the visual information to eliminate the ambiguity in regions containing complex patterns or structures. In what follows, we will describe how the homogeneity of PRNU can be integrated to guide the segmentation.

Let N be the number of pixels in the image and K be the desired number of superpixels. The partitioning of the image into superpixels can be represented by a set of random variables $\boldsymbol{s} = (s_1, .., s_N)$, where $s_i \in \{1, .., K\}$ denotes the superpixel to which pixel i belongs. Following the ETPS algorithm in [43, 44], we formulate the image segmentation as an optimization problem with the following energy function

$$\begin{aligned}\mathcal{E}(\boldsymbol{s}, \boldsymbol{\mu}, \boldsymbol{c}, \boldsymbol{P}) = &\sum_i \mathcal{E}_{col}(s_i, c_{s_i}) + \lambda_{pos} \sum_i \mathcal{E}_{pos}(s_i, \mu_{s_i}) \\ &+ \lambda_b \sum_i \sum_{j \in \mathcal{N}_8(i)} \mathcal{E}_b(s_i, s_j) + \mathcal{E}_{topo}(\boldsymbol{s}) + \mathcal{E}_{size}(\boldsymbol{s}) \\ &+ \lambda_{etp} \sum_i \mathcal{E}_{etp}(s_i, P_{s_i}), \end{aligned} \tag{4}$$

where $\boldsymbol{c}=(c_1, ..., c_K)$, $\boldsymbol{\mu}=(\mu_1, ..., \mu_K)$, and $\boldsymbol{P}=(P_1, ..., P_K)$ are respectively the set of the mean colors, the mean

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

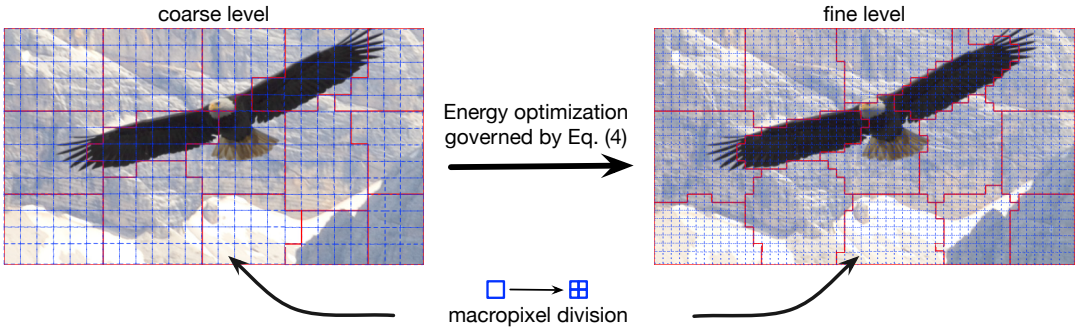


FIGURE 2: Demonstration of macropixel division for the coarse-to-fine optimization. The regions enclosed by red lines are superpixels and the regular grids bounded by blue dotted lines are macropixels.

positions, and the forgery probabilities for K superpixels. The details of different energy terms in Eq. (4) are given below.

Appearance Coherence:

$$\mathcal{E}_{col}(s_i, c_{s_i}) = ||\mathcal{I}(i) - c_{s_i}||_2^2 \tag{5}$$

This term measures the squared Euclidean distance in the CIELAB color space between pixel i and the mean color of the superpixel that i belongs to. It encourages the color homogeneity within superpixels.

Shape Regularity:

$$\mathcal{E}_{pos}(s_i, \mu_{s_i}) = ||\mathcal{L}(i) - \mu_{s_i}||_2^2. \tag{6}$$

It measures the coordinate distance between pixel i and the mean position of the superpixel that i belongs to. This term encourages compact superpixels of regular shapes.

Boundary Length:

$$\mathcal{E}_b(s_i, s_j) = \begin{cases} 1, & \text{if } s_i \neq s_j \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

This term measures the length of borders between superpixels. Irregular boundaries normally have a length longer than regular boundaries, thus this term penalizes local irregularities in the superpixel boundaries. Note that encouraging the shape and boundary regularity of the generated superpixels does not contradict with our aim to localize forgeries of various shapes because forged objects or regions rarely have wiggly shapes and boundaries.

Topology Preservation: The term $\mathcal{E}_{topo}(\boldsymbol{s})$ is used to enforce that each superpixel is composed of a single connected component without holes. This is done by checking if changing the assignment of a boundary pixel will violate connectivity.

Minimal Size: The term $\mathcal{E}_{size}(\boldsymbol{s})$ returns a value of ∞ if the size of each resultant superpixel is less than 1/4 of its initial size. It encourages superpixels with similar sizes.

PRNU Homogeneity:

$$\mathcal{E}_{etp}(s_i, P_{s_i}) = \left(255 \cdot H(P_{s_i})\right)^2, \tag{8}$$

where P_{s_i} is the forgery probability of superpixel s_i and $H(P_{s_i}) = -P_{s_i}\log_2 P_{s_i} - (1-P_{s_i})\log_2(1-P_{s_i})$ is the entropy of a Bernoulli process characterized by P_{s_i}. If a superpixel contains both pristine and forged pixels, the estimated forgery probability P_{s_p} will tend to be closer to 0.5, where $H(P_{s_i})$ attains its maximum value. Thus, $H(P_{s_i})$ is an indicator of the homogeneity of PRNU within a superpixel. Actually, a similar measurement has been used in [28, 46] to assess the confidence or reliability of the estimated forgery probability. The factor 255 is used to promote $\mathcal{E}_{etp}(s_i, P_{s_i})$ to the comparable level as the color energy term $\mathcal{E}_{col}(s_i, c_{s_i})$. Hereafter, we will refer to this algorithm as entropy-guided ETPS (EG-ETPS) algorithm.

We adopt the coarse-to-fine optimization framework in [43] to minimize the objective function in Eq. (4). As shown in Algorithm 1, the algorithm starts by equally partitioning the image into K non-overlapping square grids of size $\lfloor\sqrt{N/K}\rfloor \times \lfloor\sqrt{N/K}\rfloor$ px, where N is the total number of pixels and $\lfloor\cdot\rfloor$ is the floor function. Each grid is considered as an initial superpixel and further divided into a number of macropixels of size $M \times M$ pixels. M is initially set to 64 but will be reset to $\lfloor\sqrt{N/K}\rfloor/2$ if $\lfloor\sqrt{N/K}\rfloor < 64$ to allow for macropixel division at the beginning of the algorithm. As demonstrated in Fig. 2, a macropixel is a square image block that will be jointly considered to evaluate the objective energy function, i.e. what pixel i is referred to in Eq. (4), and at the finest level, each macropixel only consists of one pixel. Initially, the statistics of mean color $\boldsymbol{\mu}$, mean position $\boldsymbol{c}$ and forgery probability $\boldsymbol{P}$ are computed for each superpixel based on the corresponding statistics of the macropixels comprising the superpixel. Then, the following coarse-to-fine optimization iteratively updates the segmentation by proposing small local changes at boundaries.

First, all the boundary macropixels, i.e. those with at least one adjacent macropixel belonging to a different superpixel, are put into a FIFO priority list and popped out one by one to check if changing the label of the popped macropixel i will

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

Algorithm 1 Entropy-Guided ETPS Algorithm

Input:

$\boldsymbol{\mathcal{I}}$: image to be segmented;
$\boldsymbol{r}$: camera reference PRNU;
$\boldsymbol{n}$: noise residual extracted from $\boldsymbol{\mathcal{I}}$;
$\widetilde{\boldsymbol{I}}$, $\widetilde{\boldsymbol{T}}$: intensity and texture feature maps;
$\widetilde{\boldsymbol{S}}$, $\widetilde{\boldsymbol{\Upsilon}}$: signal-flattening and texture-intensity feature maps;
$\boldsymbol{\Psi}$: camera parameters incl. predictors under h_0 and h_1;
$\boldsymbol{\lambda}$: weighting factors in the objective energy function;
K: the desired number of superpixels;

Output:

$\boldsymbol{s}$: resultant segmentation of the image;

Partition $\boldsymbol{\mathcal{I}}$ into K non-overlapping square superpixels;
Partition each superpixel into regular macropixels of size 64×64 and set current maximal macropixel size $m = 64$;
Calculate the initial energy, Eq. (4), for each superpixel;
do
  if $m < 16$ **then**
    set $\lambda_{etp} = 0$
  end if
  Initialize a FIFO list with all boundary macropixels;
  while list is not empty **do**
    Pop out boundary macropixel i from list;
    if invalid_connectivity(i) **then**
      continue
    end if
    $\hat{s}_i = \arg\min_{s_i} \mathcal{E}(\boldsymbol{s}, \boldsymbol{\mu}, \boldsymbol{c}, \boldsymbol{P})$
    if $\hat{s}_i$ is updated **then**
      Incrementally update $\boldsymbol{\mu}$, $\boldsymbol{c}$ and $\boldsymbol{P}$ for the two superpixels involved;
      Append any boundary macropixel j in the 4-connected neighborhood of i to the list;
    end if
  end while
  if $m > 1$ **then**
    Bisect or quadrisect any divisible macropixel;
    Update maximal macropixel size $m \leftarrow \lceil \frac{m}{2} \rceil$;
  else
    break
  end if
while

violate connectivity. If the label change is admissible and results in an energy decrease, we assign macropixel i to one of its neighboring superpixels that gives the lowest energy. Meanwhile, we incrementally update the mean colors, mean positions and forgery probabilities of the two involved superpixels (i.e. the superpixel that macropixel i belonged to and the superpixel that macropixel i is newly assigned to) and append the boundary macropixels adjacent to i at the end of the list. This process is repeated until the list is empty, at which point the optimization will proceed to next finer level by bisecting or quadrisecting each macropixel to a smaller size.

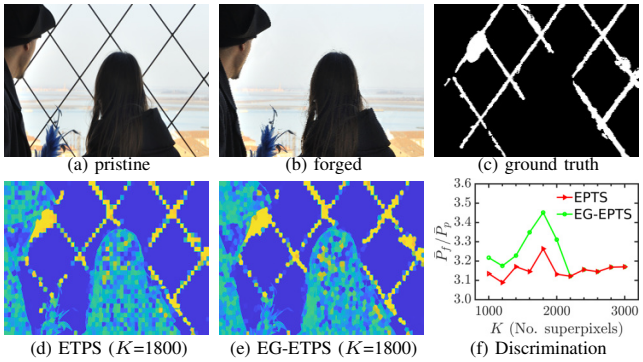


FIGURE 3: The effect of PRNU homogeneity term on the segmentation results. For ETPS and EG-ETPS, we used the same parameters except for λ_{etp}, which was set to 2 for the results shown above.

Unlike other terms in Eq. (4), the PRNU homogeneity term can only be reliably evaluated when the size of macropixel is sufficiently large due to the weak nature of PRNU. The above coarse-to-fine framework makes it possible to reliably integrate the PRNU homogeneity information at coarse levels. We set $\lambda_{etp} = 0$ when the maximal macropixel size $m<16$ when the information from PRNU is no longer reliable. One of the key steps in Algorithm 1 is the incremental update of the mean colors, mean positions and forgery probability for calculating the energy value for each superpixel. Just as other terms, the PRNU homogeneity term can also be efficiently updated in an incremental manner as the assignment of a macropixel changes. The reader is referred to the Appendix for more details.

To see how the PRNU homogeneity term affects the segmentation, we show an example of *object removal* forgery in Fig. 3. We use the ratio of the average forgery probability $\bar{P}_f$ in the forged regions to the average forgery probability $\bar{P}_p$ in the pristine regions as the indicator to differentiate the forged regions from the pristine ones. For the image of size 1920×1080 px, we vary the superpixel number K from 1000 to 3000 and show the result in Fig. 3f. We can observe a higher ratio EG-ETPS when the superpixel number $K < 2200$, which indicates an easier separation of the forged and the pristine regions. The forgery probability maps when $K = 1800$ are shown in Fig. 3e. It can be observed that EG-ETPS outputs a more homogeneous and coherent probability map.

B. MULTI-ORIENTATION FORGERY LOCALIZATION SCHEME

1) Multi-Orientation Forgery Detection: Most existing image forgery detectors apply square detection windows. Such an isotropic detection scheme inherently limits the capability

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

to detect arbitrary-shaped and arbitrary-oriented forgeries. For instance, subtle forgeries such as human body limbs or tree branches might be undetectable with a square detection window. Although this issue can be mitigated by the use of detection windows of smaller size, the reliability will also be compromised at smaller scales. To allow for more accurate forgery localization, the various shapes and orientations of the forged regions need to be taken into consideration in the design of the localization framework. Inspired by the great success of faster R-CNN [47] in detecting objects based on anchor boxes, which are a set of predefined bounding boxes of certain scales and aspect ratios, we extend the multi-scale forgery localization scheme in [28] by adopting detection windows of various aspect ratios and orientations at each scale.

Based on the multi-scale framework in [28], we replace the square detection window at each scale with detection windows of multiple aspect ratios and orientations, as illustrated in Fig. 4. To reduce the computation, we only use 5 scales, i.e. $w{\in}\{32, 48, 64, 96, 128\}$, rather than the 7 scales in [28]. For each scale, we use detection windows of 3 aspect ratios (1:1, 1:2 and 1:3) and rotate the window with a specific aspect ratio by a specific set of orientations. Specifically, we consider 11 orientations: 1 orientation $\{0\}$ for aspect ratio 1:1, 4 orientations $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ for aspect ratio 1:2, and 6 orientations $\{0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}\}$ for aspect ratio 1:3. Such configuration ensures that each window does not have too much overlap with its rotated versions. Note that the numbers of pixels within the detection windows at the same scale are approximately the same. Taking the scale w=32 as an example, the sizes of detection windows of the 3 aspect ratios are configured to $32{\times}32$, $22{\times}44$ and $18{\times}54$ px. For this reason, the offline-trained correlation predictor at one scale can be used to predict the intra-camera correlations, i.e. $p(\rho_i|h_1)$ in Eq. (3), for detection windows of different orientations at the same scale.

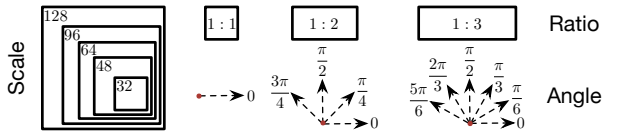


FIGURE 4: Detection windows used in our multi-orientation forgery localization scheme.

2) Multi-Orientation Fusion: Having obtained 11 candidate forgery probability maps at each scale (corresponding to each of the 11 multi-oriented detection windows), we need a fusion scheme to form a single informative probability map. At first thought, a simple pixel-wise maximum fusion rule, i.e. selecting the largest value of the candidate probabilities for each pixel, will suffice for the task since we aim to detect any possible forgery, but this will also introduce substantial false positives, i.e. mis-labeling pristine pixels as forged. Ideally, the best detection result can be obtained if the detection window is perfectly aligned with the forged region. Thus, it is reasonable to accept the forgery probability calculated with the detection window that agrees best with the segmentation result. Suppose P_b, $b \in \{1, ..., 11\}$ are the candidate forgery probabilities obtained with 11 detection windows centered at pixel i. We adopt the following fusion strategy for pixel i:

$$P_i = rP_{b^\star} + (1-r)P_{s_i}, \tag{9}$$

where P_{s_i} is the forgery probability of the superpixel s_i that pixel i belongs to, $b^\star$ is the index of the detection window that has the best agreement with s_i, i.e.

$$b^\star = \underset{b\in\{1,...,11\}}{\arg\max}\ W_b \cap s_i. \tag{10}$$

r is the proportion of the detection window $W_{b^\star}$ that overlaps with s_i. It measures how much of the probability obtained with window $b^\star$ contributes to the final result. $r = 1$ when the detection window completely falls within a superpixel, which means we only rely on the result obtained with one of the detection windows and is similar to the traditional detector merely based on sliding windows. $r < 1$ when the detection window $W_{b^\star}$ falls across two or more superpixels and $(1-r)P_{s_i}$ will be used to compensate for the pixels falling outside of the superpixel that pixel i belongs to. Note that the above fusion is performed at each scale based on the superpixel segmentation and forgery probability maps obtained with 11 detection windows. We introduce a parameter ξ to control the average size of superpixels relative to the size of detection window at each scale. Specifically, if the size of the detection window is $w{\times}w$ px, the number of superpixels K is set to

$$K = \lfloor \frac{N}{\xi w^2} \rceil, \tag{11}$$

where N is the total number of pixels in the image and $\lfloor\cdot\rceil$ is the rounding operator. To see how the fusion strategy affects the final forgery probability when $r < 1$, we analyze the following simplified scenarios as shown in Fig. 5:

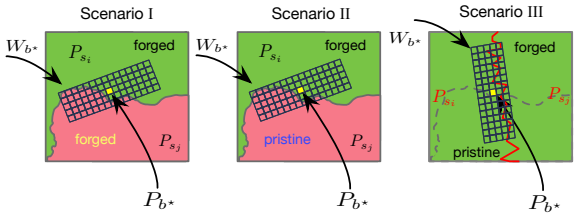


FIGURE 5: Three simplified scenarios for multi-orientation forgery probability fusion. The pixel of interest is highlighted in yellow and different color appearances are highlighted in different colors.

- Scenario I: Two forged regions with different color appearances are segmented into two different superpixels s_i and s_j, which often occurs inside a forged region. Suppose the forgery probability obtained with the detection window $W_{b^\star}$ is $P_{b^\star}$ and the forgery probabilities

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

for s_i and s_j are P_{s_i} and P_{s_j}, respectively. The pixels considered in this scenario are all forged pixels, so we can simply assume that $P_{s_i} \approx P_{s_j} \approx P_{b^\star}$ since the correlation predictor has been designed to account for different color appearances. Therefore, the final forgery probability $P_i \approx P_{b^\star}$, which means that the fusion is equivalent to selecting the probability corresponding to the detection window that agrees best with the segmentation results.

- Scenario II: Two neighboring regions, with one forged and the other pristine, have different color appearances and are segmented into two superpixels s_i and s_j. This case usually occurs for *object insert* forgeries, where the inserted object usually has a different color appearance from the background. In such case, the pixels within the detection window falling outside the forged region lead to an attenuated $P_{b^\star}$. The above fusion strategy compensates the attenuated $P_{b^\star}$ by adding a term $(1-r)P_{s_i}$. As P_{s_i} is calculated over forged pixels and is expected to be $> P_{b^\star}$, it results in a $P_i > P_{b^\star}$ compensating for the attenuation caused by the pixels falling outside the forged region.
- Scenario III: In this scenario, the forged and pristine regions have the same or similar color appearance, which often occurs in the case of *object removal* forgery. Due to the lack of distinguishable color appearance, some parts of the two regions are quite likely to be segmented into the same superpixel. For instance, the segmentation may end up with two superpixels s_i and s_j separated by the red line. If we assume that s_i and window $W_{b^\star}$ are equally possible to contain heterogeneous pixels, this scenario is similar to the detection merely based on regular detection windows. In practice, because most superpixel algorithms will try to utilize as much local information as possible to perform the segmentation, the above fusion is expected to deliver comparable or even better performance than the methods merely based on regular detection windows.

Finally, we apply the multi-scale fusion strategies as proposed in [28] for the fused forgery probability maps obtained at different scales. The framework of our proposed multi-orientation forgery localization scheme is demonstrated in Fig. 6.

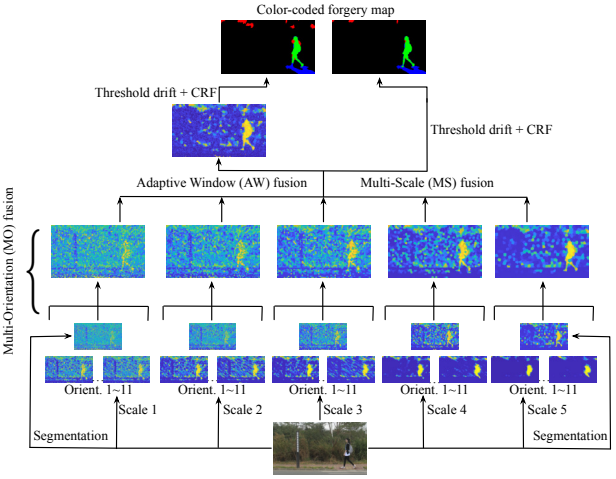


FIGURE 6: The multi-orientation forgery localization scheme. For more details about the Adaptive Window (AW) and Multi-Scale (MS) fusion algorithms, please refer to Section IV-C or [28].

IV. EXPERIMENTS

A. DATASETS

Our experiments were conducted on the realistic image tampering dataset (RTD) [28, 48], which contains 220 realistic forgeries captured by 4 different cameras: Canon 60D, Nikon D90, Nikon D7000, Sony α57 (each camera responsible for 55 forgery images). The images in the RTD dataset are 1920×1080 px RGB images stored in the TIFF format and cover various types of forgeries including *object insert* and *object removal* forgeries. The RTD dataset provides a good benchmark to evaluate the performance of camera-based content authentication techniques. Considering the nature of our proposed methods, it is reasonable to conduct evaluations separately on the *object insert* and *object removal* forgeries. However, the wide variety of forgeries in the RTD dataset makes it hard to simply divide them into two categories. Therefore, we divided the RTD dataset into three subsets:

(i) *Hard-boundary* Forgeries (HBF): This subset contains the forged images with visually distinguishable boundaries between the pristine and the forged regions. It mainly consists of the forgeries created by object insert, object replacement, color altering, etc.

(ii) *Soft-boundary* Forgeries (SBF): This subset contains the forged images with visually smooth and indistinguishable boundaries between the pristine and the forged regions. It mainly consists of the forgeries created by background texture synthesis or content-aware filling.

(iii) *Mixed-boundary* Forgeries (MBF): This subset contains the images with both hard and soft boundaries between the pristine and the forged regions.

The dataset division was done by visually inspecting each image in the RTD dataset. This results in 100, 80 and 40 images in the HBF, SBF and MBF subsets, respectively. Some examples in these three subsets are shown in Fig. 7. Considering the relatively small size of the MBF subset, we merged MBF with both HBF and SBF to create two subsets, i.e. HBF+MBF (140 images) and SBF+MBF (120 images), for evaluating the localization performance on forgeries of two boundary types.

B. EVALUATION PROTOCOLS

The localization performance is commonly evaluated by the *region* F_1 score. For an image $I_i, 1 \leq i \leq \mathcal{S}$, let G_i be its binarized ground truth forgery map and $L_i(\tau)$ be the binary forgery map output by a forgery localization algorithm with

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

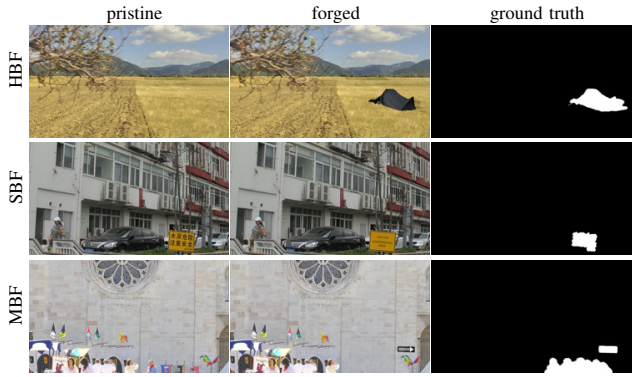


FIGURE 7: Examples of forged images in *hard-boundary* forgeries (HBF), *soft-boundary* forgeries (SBF) and *mixed-boundary* forgeries (MBF).

a threshold τ. The *region* F_1 score is defined as

$$F_r(G_i, L_i(\tau)) = \frac{2 \cdot \mathcal{P}_r(G_i, L_i(\tau)) \cdot \mathcal{R}_r(G_i, L_i(\tau))}{\mathcal{P}_r(G_i, L_i(\tau)) + \mathcal{R}_r(G_i, L_i(\tau))}, \quad (12)$$

where $\mathcal{P}_r(G_i, L_i(\tau))$ and $\mathcal{R}_r(G_i, L_i(\tau))$ are the precision (the fraction of correctly detected forgery pixels) and the recall (the fraction of ground-truth forgery pixels detected), respectively. Like in [28], we measured the overall performance with a curve of average $\bar{F}_r(\tau)$ score (each point in the curve is the F_r score averaged over $\mathcal{S}$ images for a given τ) and an average peak $\hat{F}_r$ score (i.e. the average of the highest achievable F_r score over all possible thresholds for $\mathcal{S}$ images):

$$\begin{cases} \bar{F}_r(\tau) = \dfrac{1}{\mathcal{S}} \displaystyle\sum_{i=1}^{\mathcal{S}} F_r(G_i, L_i(\tau)) \\ \hat{F}_r = \dfrac{1}{\mathcal{S}} \displaystyle\sum_{i=1}^{\mathcal{S}} \max_{\tau \in (0,1)} F_r(G_i, L_i(\tau)). \end{cases} \quad (13)$$

F_r score measures the localization performance by considering the overlapping area (in terms of pixel counts) between G_i and $L_i(\tau)$. However, as pointed out in [48], the *region* F_1 score is insufficient to accurately assess the contour adherence between the ground-truth and localized forgery regions. In fact, the contour conveys key information about the object shape and provides inferable clues to uncover the potential forged object, which is particularly desirable in the scenario of *soft-boundary* forgeries. We thus also evaluated the localization performance with *boundary* F_1 score [49], which has been widely used to assess the boundary adherence for image segmentation. For an image $I_i, 1 \leq i \leq \mathcal{S}$, let B_i be the contour map of the binarized ground-truth forgery map and $D_i(\tau)$ be the contour map of the binary forgery map output by a forgery localization algorithm given a threshold τ. The *boundary* F_1 score is defined as

$$F_b\left(B_i, D_i(\tau)\right) = \frac{2 \cdot \mathcal{P}_b\left(B_i, D_i(\tau)\right) \cdot \mathcal{R}_b\left(B_i, D_i(\tau)\right)}{\mathcal{P}_b\left(B_i, D_i(\tau)\right) + \mathcal{R}_b\left(B_i, D_i(\tau)\right)}, \quad (14)$$

where $\mathcal{P}_b(B_i, D_i(\tau))$ and $\mathcal{R}_b(B_i, D_i(\tau))$ are the boundary precision and the boundary recall given a *distance tolerance* θ:

$$\begin{cases} \mathcal{P}_b(B_i, D_i(\tau)) = \dfrac{1}{|D_i(\tau)|} \displaystyle\sum_{z \in D_i(\tau)} [\![d(z, B_i) < \theta]\!] \\ \mathcal{R}_b(B_i, D_i(\tau)) = \dfrac{1}{|B_i|} \displaystyle\sum_{z \in B_i} [\![d(z, D_i(\tau)) < \theta]\!]. \end{cases} \quad (15)$$

Here, $[\![z]\!]$ is the Iversons bracket notation, i.e. $[\![z]\!] = 1$ if $z =$ true and 0 otherwise, $d(\cdot)$ is the Euclidean distance, and θ determines whether a boundary point has a match or not and is set to 0.75% of the image diagonal throughout our evaluation. Similarly to the *region* F_1 score, we evaluate the overall boundary quality with a curve of average $\bar{F}_b(\tau)$ score and an average peak $\hat{F}_b$ score:

$$\begin{cases} \bar{F}_b(\tau) = \dfrac{1}{\mathcal{S}} \displaystyle\sum_{i=1}^{\mathcal{S}} F_b(B_i, D_i(\tau)) \\ \hat{F}_b = \dfrac{1}{\mathcal{S}} \displaystyle\sum_{i=1}^{\mathcal{S}} \max_{\tau \in (0,1)} F_b(B_i, D_i(\tau)). \end{cases} \quad (16)$$

C. EVALUATED ALGORITHMS AND PARAMETER SETTINGS

We considered the following single-orientation forgery localization algorithms for the comparison with our proposed schemes:

- Simple Thresholding (ST) based algorithm [26]: For ST, we first generated a binarized forgery map by comparing the forgery probability map obtained by a detection window of 128×128 px with a threshold $\tau \in [0, 1]$. Then we removed the connected forged regions with pixels fewer than 64×64 px and applied image dilation with a disk kernel with a radius of 16 px to generate the final decision map. Note that ST algorithm is a single-scale detector with no image content involved in the final decision-making process.
- Single-Orientation Multi-Scale (SO+MS) fusion based algorithm [28]: SO+MS formulates the forgery localization as an image labeling problem and solves it by the conditional random field (CRF) model. The data term of the CRF model is the average of the threshold-drifted [16, 28] forgery probability maps obtained with single-oriented detection windows at 7 different scales, and the neighborhood interaction of image content is encoded in the regularization term. The final binary decision map is obtained by optimizing the CRF model.
- Single-Orientation Adaptive Window (SO+AW) fusion based algorithm [28]: SO+AW aims to fuse the forgery probability maps obtained with single-oriented detection windows of 7 scales. Starting from the smallest scale, it looks for a sufficiently confident decision (i.e. forgery probability that is far from 0.5) for each location in the image. If the decision at a smaller scale is not confident enough, it proceeds to the next larger scale

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

until a sufficiently confident decision and an agreement between two consecutive scales are reached. Finally, the final binary decision map is obtained by applying the threshold drift strategy and optimizing the CRF model.

- Segmentation-Guided (SG) based algorithm [28]: SG algorithm calculates the forgery probability by only considering the pixels with an intensity value close to the central pixel (the average L_1 distance in RGB space less than 15) within a detection window of 128×128 px. It implements the idea of image segmentation but only exploits the intensity difference between individual pixels. Similarly to SO+AW and SO+MS, the threshold drift strategy and CRF are applied to obtain the final binary decision map.

Note that the notations 'SO+MS', 'SO+AW', 'SG' used in this paper correspond to the 'MSF', 'AW+' and 'SG+' algorithms in [28]. We use the notations 'AW' and 'MS' to denote the adaptive window and multi-scale strategies proposed in [28] for multi-scale fusion. For our proposed segmentation-based schemes, we will use the notation 'SEG+FUSION' to represent the FUSION$\in$\{AW, MS\} fusion of the probability maps calculated based on the superpixels generated by SEG$\in$\{SLIC, ERS, ETPS, EG-ETPS\} algorithm. Similarly, for our proposed multi-orientation forgery localization scheme, we will use the notation 'SEG+MO+FUSION' to represent the FUSION$\in$\{AW, MS\} fusion of the integrated probability maps obtained with multi-oriented detection windows and the superpixel algorithm SEG$\in$\{SLIC, ERS, ETPS, EG-ETPS\}.

For SO+MS, SO+AW and SG, we used exactly the same parameters as summarized in Table II of [28]. For the segmentation algorithms used in this work, their parameters are given as follows:

- SLIC [41]: We set both the compactness parameter and the iteration number of pixel assignment and centroid updating to 10.
- ERS [42]: As suggested in [42], we set the weighting factor of the balancing term λ'=0.5 and the Gaussian kernel bandwidth σ=5.0 for calculating the pixel similarities.
- ETPS [43]: We set the weighting factors of both the shape regularity term and the boundary length term to 0.2, i.e. $\lambda_{pos} = 0.2$ and $\lambda_b = 0.2$.
- EG-ETPS: We used exactly the same parameters as ETPS except for the weighting factor λ_{etp} of the PRNU homogeneity term. As can be expected, the setting of λ_{etp} will depend on the quality of PRNU. Thus we empirically set $\lambda_{etp} = 2.5 \cdot \exp(-(R^2 - 1)^2/0.3)$. R^2 is the adjusted R-squared coefficient for the correlation predictor trained at the scale of 64×64 px, which is a good indicator of the quality of PRNU. This results in a $\lambda_{etp} \in [1.8, 2.2]$ for the four cameras in the RTD dataset.

For our proposed multi-orientation scheme, we set $\xi = 0.2$, which controls the average size of superpixels relative to the window size at each scale.

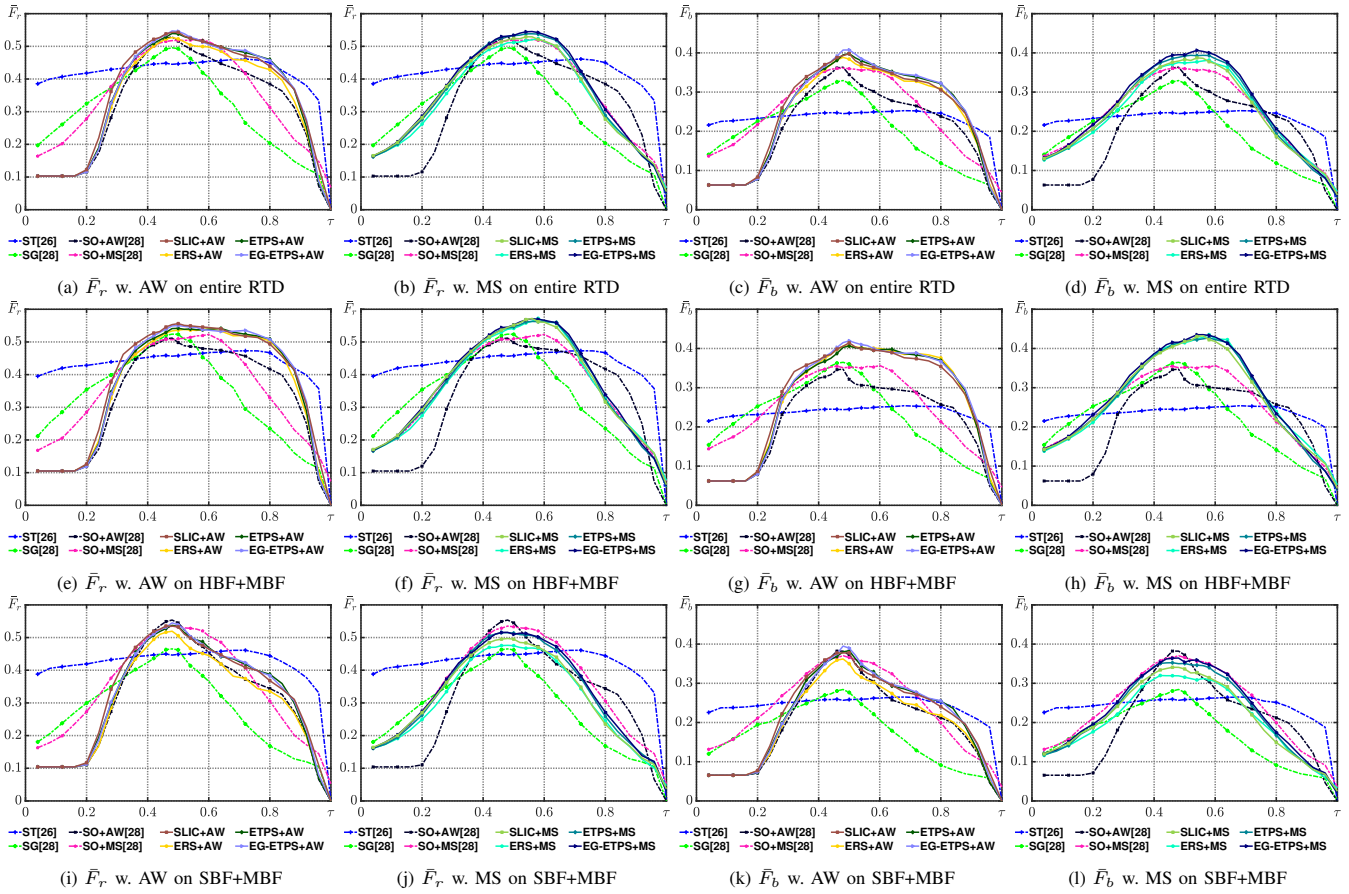
1) Results for Segmentation-based Localization Schemes: In this experiment, we applied the adaptive window and multi-scale fusion strategies directly on the probability maps obtained by calculating the forgery probability on each generated superpixel. To make the average superpixel size consistent with the detection window size at each scale, we specified the desired superpixel number K of each segmentation algorithm as $\lfloor \frac{N}{w^2} \rceil$, where N is the number of pixels in the image, $w \times w$ is the size of the detection window at a specific scale and $\lfloor \cdot \rceil$ is the rounding operator. We only used 5 scales, i.e. $w \in \{32, 48, 64, 96, 128\}$, for our segmentation-based schemes, while for the compared methods SO+AW and SO+MS proposed in [28], we still used 7 scales.

we show the average score curves $\bar{F}_r$ and $\bar{F}_b$ in Fig. 9 as well as the average peak scores $\hat{F}_r$ and $\hat{F}_b$ in Fig. 8, respectively. For easy comparison, we also summarized the highest average $\bar{F}_r$, $\bar{F}_b$ and the performance gain of the best segmentation-based scheme relative to the corresponding single-orientation scheme in Table 1. Note that ST and SG are not included in Table 1 because they have inferior performance than multi-scale methods as clearly shown in Fig. 9. We can see that even with only 5 scales, the segmentation-based forgery localization schemes are able to deliver better performance than SO+AW and SO+MS on the entire RTD dataset. The advantage of the segmentation-based schemes is more evident if the performance is measured by $\bar{F}_b$ (the 3rd and 4th columns of Fig. 9), which is increased by 12.4% and 11.8% for AW and MS multi-scale fusions, respectively. Regarding the localization performance on two different boundary types, it is not surprising to see that segmentation-based schemes outperform the single-orientation schemes by large margins on HBF+MBF, with 9.6% increase in $\bar{F}_r$ and 21.8% increase in $\bar{F}_b$ when MS is applied. However, due to the lack of distinct local information to guide the segmentation, segmentation-based schemes perform slightly worse than those based on detection windows when localizing the forgeries in SBF+MBF.

As for the comparison between different segmentation algorithms, we can observe that EG-ETPS benefits from the integration of PRNU homogeneity information and slightly outperforms other segmentation algorithms. It is worth mentioning that the benefits of EG-ETPS are mainly reflected in detecting the boundaries of the forgeries, so its overall performance gain over other segmentation algorithms depends on the amount of *soft boundaries* and is more evident when the performance is measured in terms of boundary measurement F_b. For the other three segmentation algorithms, ETPS delivers generally better performance than SLIC and ERS. An interesting observation is that the superpixels generated by ERS exhibit excellent boundary adherence and provide the best $\hat{F}_b$ performance among the four segmentation algorithms on HBF+MBF (see Fig. 8), but the trade-off is that its capability in detecting forgeries in SBF+MBF is greatly compromised.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access



(a) $\bar{F}_r$ w. AW on entire RTD (b) $\bar{F}_r$ w. MS on entire RTD (c) $\bar{F}_b$ w. AW on entire RTD (d) $\bar{F}_b$ w. MS on entire RTD

(e) $\bar{F}_r$ w. AW on HBF+MBF (f) $\bar{F}_r$ w. MS on HBF+MBF (g) $\bar{F}_b$ w. AW on HBF+MBF (h) $\bar{F}_b$ w. MS on HBF+MBF

(i) $\bar{F}_r$ w. AW on SBF+MBF (j) $\bar{F}_r$ w. MS on SBF+MBF (k) $\bar{F}_b$ w. AW on SBF+MBF (l) $\bar{F}_b$ w. MS on SBF+MBF

FIGURE 8: Segmentation-based vs. single-orientation forgery localization schemes in terms of average score curves. The $\bar{F}_r$ and $\bar{F}_b$ on the RTD dataset (220 images), HBS+MBF (140 images) and SBF+MBF (120 images) subsets are shown in the 1st, 2nd and 3rd row, respectively.

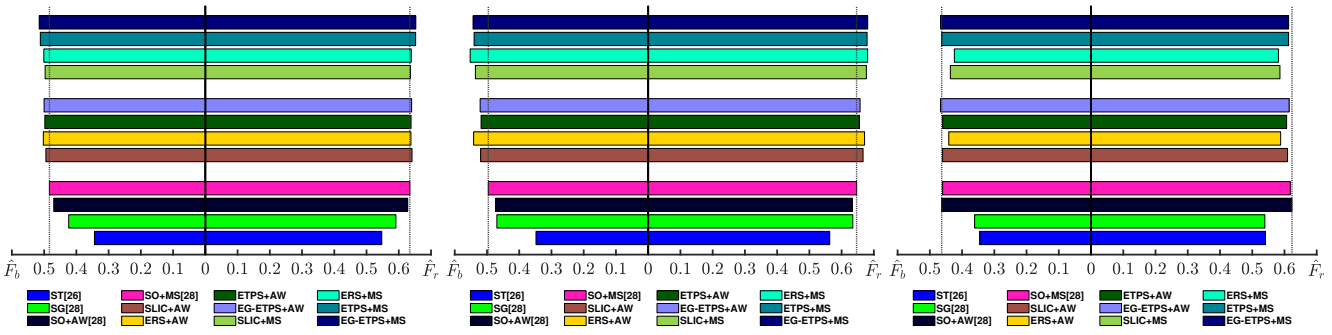
2) Results for Multi-Orientation Localization Schemes: In this experiment, we aim to compare the performance of our proposed multi-orientation localization schemes and the methods based on single-oriented detection windows. The comparison results are shown in Fig. 10 and Fig. 11. Similarly, for easy comparison, we also summarized the highest average measures $\bar{F}_r$ and $\bar{F}_b$ across all thresholds and the performance gain of the best multi-orientation scheme relative to the corresponding single-orientation scheme in Table 2.

We can see that the MS fusion strategy achieves considerably better performance than the AW fusion for our proposed multi-orientation schemes. A closer inspection revealed that AW is more likely to introduce false positives especially in the regions where PRNU is substantially attenuated. In addition, compared to the single-orientation schemes, significant performance improvement can be observed for the multi-orientation localization schemes when the MS fusion strategy is applied, with the highest $\bar{F}_r$ and $\bar{F}_b$ increased by 7.5% and 18.7% respectively on the entire RTD dataset. The multi-orientation schemes are even more advantageous on the HBF+MBF subset as evidenced by the performance gains of 10.9% in $\bar{F}_r$ and 28.6% in $\bar{F}_b$. While on the SBF+MBF subset, the multi-orientation schemes still deliver comparable performance as the best two single-orientation schemes SO+AW and SO+MS.

Another important observation for the proposed multi-orientation schemes is that the performance gap between the four segmentation algorithms becomes very small. For the segmentation-based localization schemes, the difference between the segmentation algorithms are much more noticeable when localizing the forgeries in the SBF+MBF subset (see the fourth row of Fig. 9). However, for the multi-orientation localization schemes, the capability of localizing *soft-boundary* forgeries mainly stems from the detection based on multi-oriented detection windows rather than the segmentation algorithms, which narrows the performance gaps between different segmentation algorithms. Some examples of forgery localization can be found in Fig. 12, where we only show the results of EG-ETPS and MS for the proposed multi-orientation forgery localization schemes. Note that for MS, we show the average of the probability maps across different scales to approximate the fused probability map in Fig. 12.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access



(a) $\hat{F}_r$&$\hat{F}_b$ on entire RTD (b) $\hat{F}_r$&$\hat{F}_b$ on SBF+MBF (c) $\hat{F}_r$&$\hat{F}_b$ on SBF+MBF

FIGURE 9: Segmentation-based vs. single-orientation forgery localization schemes in terms of average peak scores. The $\hat{F}_r$ and $\hat{F}_b$ on the RTD dataset (220 images), HBS+MBF (140 images) and SBF+MBF (120 images) subsets are shown in the 1st, 2nd and 3rd column, respectively. The vertical dash lines indicate the best performance of single-orientation schemes, i.e. ST, SG, SO+AW and SO+MS.

Dataset		Method	highest $\bar{F}_r$		highest $\bar{F}_b$	
			AW	MS	AW	MS
Entire RTD		SO [28]	0.528	0.523	0.363	0.363
	Proposed	SLIC [41]	0.543	0.529	0.398	0.387
	Proposed	ERS [42]	0.527	0.524	0.388	0.382
	Proposed	ETPS [43]	0.539	0.538	0.398	0.394
	Proposed	EG-ETPS	**0.547**	**0.545**	**0.408**	**0.406**
		Increased by	3.6%	4.2%	12.4%	11.8%
HBF + MBF		SO [28]	0.511	0.522	0.346	0.357
	Proposed	SLIC [41]	**0.556**	**0.572**	0.414	0.432
	Proposed	ERS [42]	0.540	**0.572**	0.406	0.435
	Proposed	ETPS [43]	0.542	0.563	0.406	0.427
	Proposed	EG-ETPS	0.550	0.571	**0.420**	**0.435**
		Increased by	8.8%	9.6%	21.4%	21.8%
SBF + MBF		SO [28]	**0.554**	**0.537**	0.384	**0.374**
	Proposed	SLIC [41]	0.535	0.497	0.378	0.341
	Proposed	ERS [42]	0.519	0.476	0.362	0.320
	Proposed	ETPS [43]	0.537	0.515	0.383	0.353
	Proposed	EG-ETPS	0.545	0.516	**0.395**	0.365
		Increased by	-1.6%	-3.9%	2.9%	-2.4%

TABLE 1: Segmentation-based vs. single-orientation forgery localization schemes in terms of the highest average *region* $\bar{F}_r$ scores and *boundary* $\bar{F}_b$ scores across all thresholds. 'SO' stands for the fusion of probability maps obtained with single-oriented detection windows as proposed in [28]. The best performance on each dataset is highlighted in bold.

Dataset		Method	highest $\bar{F}_r$		highest $\bar{F}_b$	
			AW	MS	AW	MS
Entire RTD		SO [28]	0.528	0.523	0.363	0.363
	Proposed	SLIC+MO	0.521	0.555	0.385	0.423
	Proposed	ERS+MO	0.523	**0.562**	0.381	**0.435**
	Proposed	ETPS+MO	0.527	0.560	0.391	0.428
	Proposed	EG-ETPS+MO	**0.534**	**0.562**	**0.395**	0.431
		Increased by	1.1%	7.5%	8.8%	18.7%
HBF + MBF		SO [28]	0.511	0.522	0.346	0.357
	Proposed	SLIC+MO	0.549	0.564	0.399	0.435
	Proposed	ERS+MO	**0.559**	**0.579**	**0.424**	**0.459**
	Proposed	ETPS+MO	0.554	0.574	0.411	0.447
	Proposed	EG-ETPS+MO	0.556	0.575	0.414	0.449
		Increased by	9.4%	10.9%	22.5%	28.6%
SBF + MBF		SO [28]	**0.554**	0.537	**0.384**	0.374
	Proposed	SLIC+MO	0.516	**0.548**	0.366	**0.403**
	Proposed	ERS+MO	0.490	0.537	0.345	0.393
	Proposed	ETPS+MO	0.513	0.544	0.363	0.399
	Proposed	EG-ETPS+MO	0.517	0.544	0.376	0.401
		Increased by	-6.8%	2%	-2.1%	7.8%

TABLE 2: Multi-orientation vs. single-orientation forgery localization schemes in terms of the highest average *region* $\bar{F}_r$ scores and *boundary* $\bar{F}_b$ scores across all thresholds.

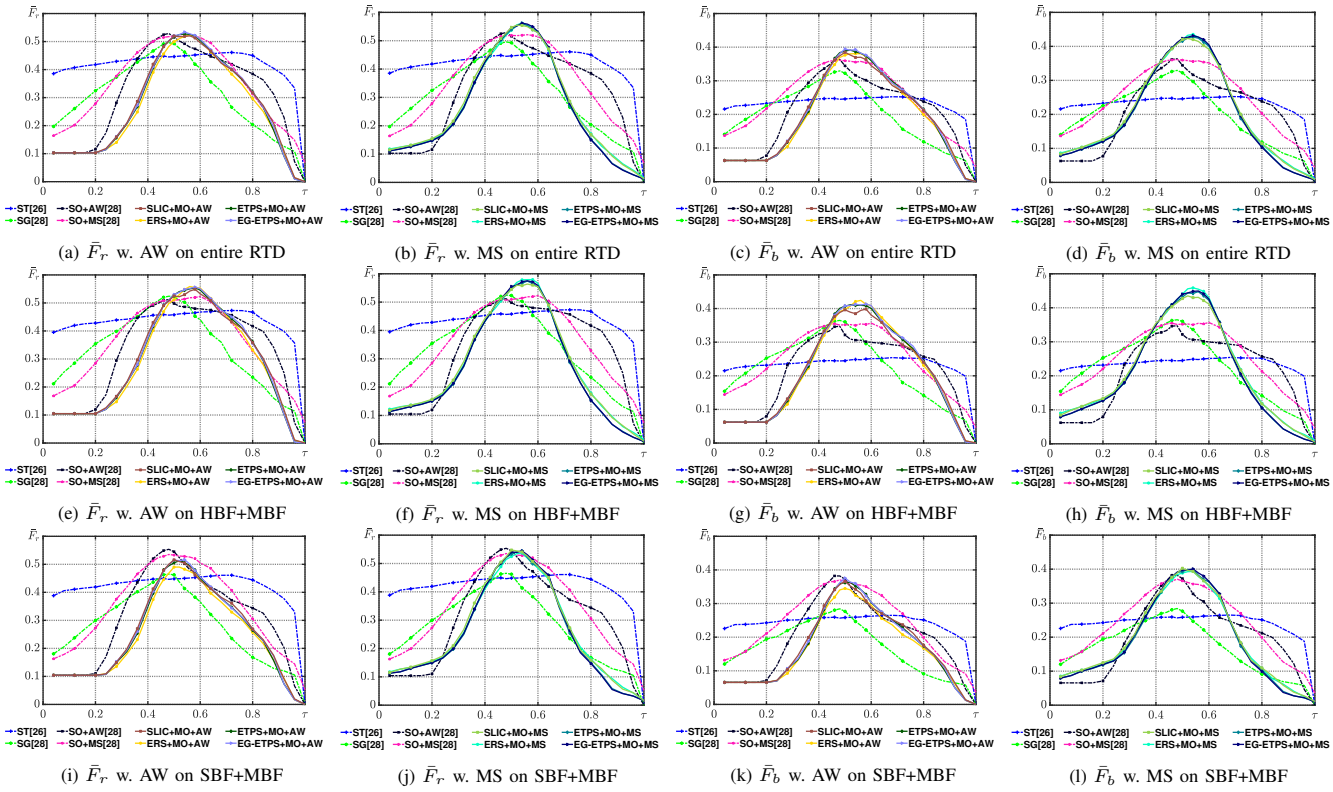
D. ROBUSTNESS AGAINST JPEG COMPRESSION

One of the main threats to PRNU-based forgery localization is JPEG compression. Thus, another experiment was conducted to evaluate the robustness against JPEG compression. We generated 6 new versions of each image by re-saving the corresponding TIFF images with JPEG quality factors of 100, 95, 90, 85, 80 and 75. We then ran all the localization algorithms on the image set of each version and calculated the performance statistics. To calculate the corresponding forgery probabilities, we used the reference PRNU and correlation predictors trained with TIFF images for JPEG images of different quality levels.

The results on the entire RTD dataset in terms of $\hat{F}_r$ and $\hat{F}_b$ are illustrated in Fig. 13. For the readability of the figure, we did not show the adaptive-window fusion results for the segmentation-based and multi-orientation localization schemes. As can be observed, when the image quality is above 90, both the segmentation-based and the multi-orientation schemes consistently outperform the single-orientation methods ST, SG, SO+AW and SO+MS on the entire RTD dataset. We can also observe that the $\hat{F}_r$ of SG deteriorates more slowly than that of other methods as the compression becomes more severe (see Fig. 13a and 13c), but this phenomenon was not observed if the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

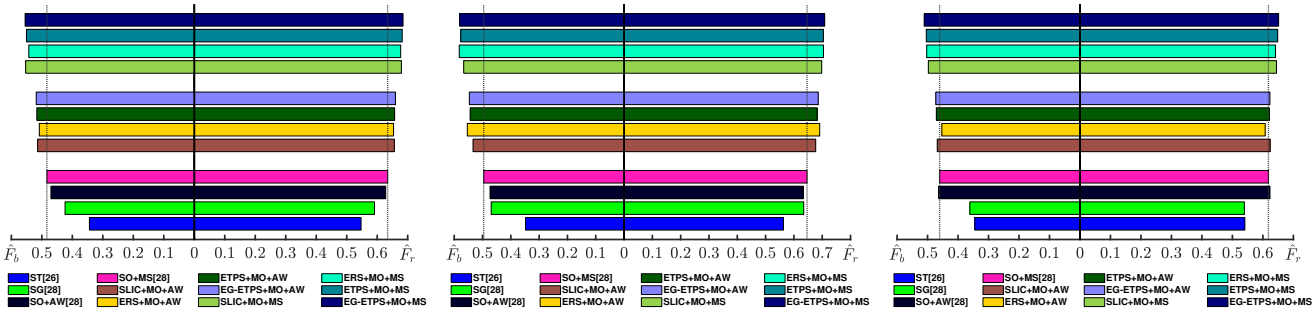


(a) $\bar{F}_r$ w. AW on entire RTD (b) $\bar{F}_r$ w. MS on entire RTD (c) $\bar{F}_b$ w. AW on entire RTD (d) $\bar{F}_b$ w. MS on entire RTD

(e) $\bar{F}_r$ w. AW on HBF+MBF (f) $\bar{F}_r$ w. MS on HBF+MBF (g) $\bar{F}_b$ w. AW on HBF+MBF (h) $\bar{F}_b$ w. MS on HBF+MBF

(i) $\bar{F}_r$ w. AW on SBF+MBF (j) $\bar{F}_r$ w. MS on SBF+MBF (k) $\bar{F}_b$ w. AW on SBF+MBF (l) $\bar{F}_b$ w. MS on SBF+MBF

FIGURE 10: Multi-orientation vs. single-orientation forgery localization schemes. The $\bar{F}_r$ and $\bar{F}_b$ on the RTD dataset (220 images), HBS+MBF (140 images) and SBF+MBF (120 images) subsets are shown in the 1st, 2nd and 3rd row, respectively.



(a) $\hat{F}_r$&$\hat{F}_b$ on entire RTD (b) $\hat{F}_r$&$\hat{F}_b$ on HBF+MBF (c) $\hat{F}_r$&$\hat{F}_b$ on SBF+MBF

FIGURE 11: Multi-orientation vs. single-orientation forgery localization schemes in terms of average peak scores. $\hat{F}_r$ and $\hat{F}_b$ on the RTD dataset (220 images), HBS+MBF (140 images) and SBF+MBF (120 images) subsets are shown in the 1st, 2nd and 3rd column, respectively. The vertical dash lines indicate the best performance of single-orientation schemes, i.e. ST, SG, SO+AW and SO+MS.

performance is measured by $\hat{F}_b$.

V. CONCLUSIONS

In this work, we investigated the potential of explicit image segmentation for content forgery localization. We have shown that image segmentation by exploiting the visual content is beneficial for improving the performance of forgery localization based on imperceptible forensic clues, especially for *hard-boundary* forgeries. While the effectiveness of segmentation merely based on visual content can be compromised for *soft-boundary* forgeries, such limitation can be mitigated by further integrating the local homogeneity of imperceptible forensic clues to guide the segmentation. To better resolve the issue of detecting *soft-boundary* forgeries, we also proposed a localization scheme based on the multi-orientation fusion of the forgery probability maps obtained by multi-orientation detection and image segmentation. With the aid of the multi-scale fusion, the multi-orientation detection is effective in detecting *soft-boundary* forgeries and the segmentation is particularly good

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

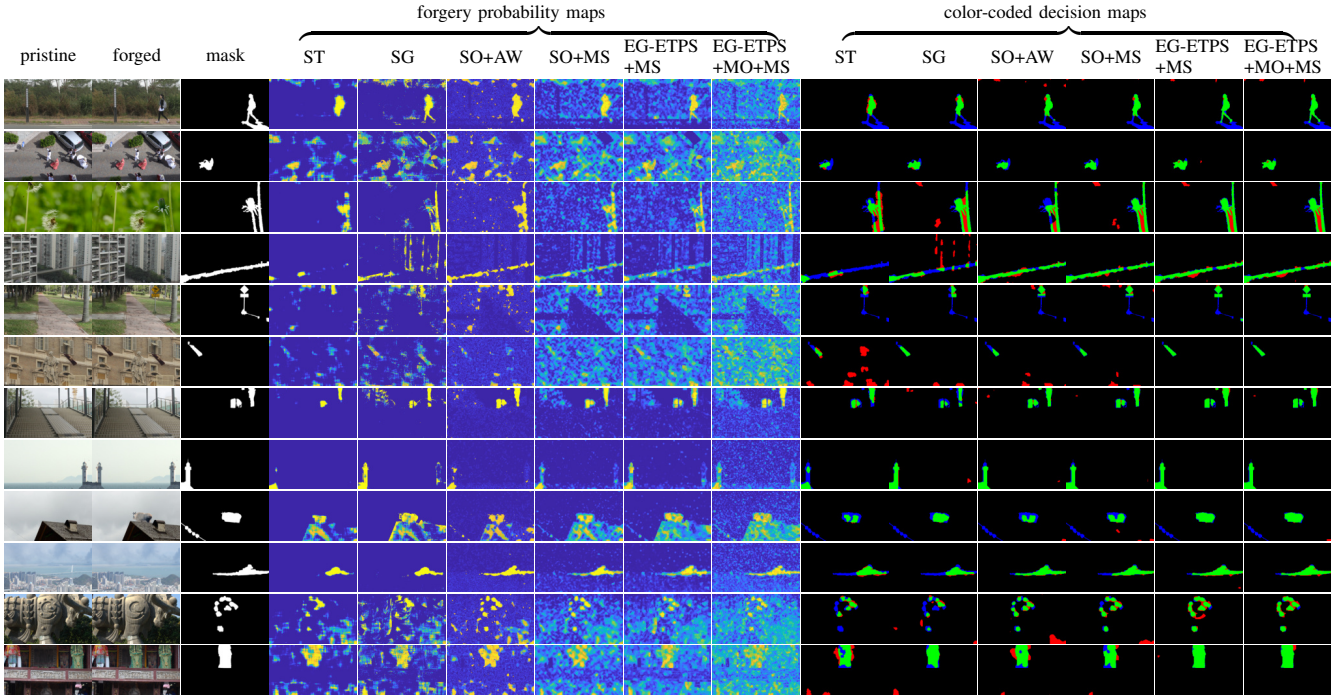


FIGURE 12: Example forgery localization results. Color coding: *green*: detected forged regions (true positives); *red*: detected pristine regions (false positives); *blue*: mis-detected forged regions (false negatives).

at identifying *hard-boundary* forgeries. Integrating them in a complementary way leads to the superior localization performance for our proposed multi-orientation schemes at the expense of extra computation complexity. Although we used PRNU-based forgery localization as an example in this paper, we believe that similar ideas can also apply to forgery detectors based on other forensic clues. Further investigations on the potential of image segmentation in other forensics detectors as well as the combination of them will be conducted in our future work.

APPENDIX: INCREMENTAL UPDATE OF FORGERY PROBABILITY

We calculate the forgery probability P_{s_i} of superpixel s_i using Eq. (3):

$$P_{s_i} = \left(1 + \exp\left(\frac{\rho_{s_i}^2}{2\hat{\sigma}_0^2} - \frac{(\rho_{s_i} - \hat{\rho}_{s_i})^2}{2\hat{\sigma}_1^2} - \ln\frac{\hat{\sigma}_1}{\hat{\sigma}_0}\right)\right)^{-1}, \tag{17}$$

where ρ_{s_i} and $\hat{\rho}_{s_i}$ are respectively the real and the expected correlations between the reference PRNU and the noise residual of superpixel s_i, and $\hat{\sigma}_0$ and $\hat{\sigma}_1$ are the standard deviations of the correlation distributions under hypotheses h_0 and h_1. Like in [28], we use spline interpolation to dynamically obtain $\hat{\sigma}_0$ and $\hat{\sigma}_1$ based on the number of pixels in s_i and the standard deviations of the correlation distributions for scales $\{32, 48, ..., 256\}$ estimated from training data. By assuming that PRNU noise is locally zero-mean, we simplify Eq. (2) as

$$\rho_{s_i} = \frac{||\boldsymbol{r}_{s_i} \cdot \boldsymbol{n}_{s_i}||_1}{||\boldsymbol{r}_{s_i}||_2 \cdot ||\boldsymbol{n}_{s_i}||_2} \tag{18}$$

to allow for the efficient incremental update for each superpixel s_i. For the expected correlation $\hat{\rho}_{s_i}$, we adopt the correlation predictor based on least-square estimator [26] :

$$\hat{\rho}_{s_i} = \boldsymbol{f}_{s_i}\hat{\boldsymbol{\theta}}_{s_i}, \tag{19}$$

where $\hat{\boldsymbol{\theta}}_{s_i}$ is a 15×1 least-square parameter vector corresponding to the scale closest to the number of pixels in s_i, and $\boldsymbol{f}_{s_i}$ is a 1×15 vector composed of four image features (i.e. the intensity, texture, signal-flattening and texture-intensity features) and their second-order terms for superpixel s_i. Since the calculation of the image feature is a local operation, e.g. the intensity feature is the mean of the attenuated intensity of an image block, $\boldsymbol{f}_{s_i}$ can be updated incrementally if a macropixel j is removed from or added to s_i:

$$\boldsymbol{f}_{s_i}^{(t)} = \boldsymbol{f}_{s_i}^{(t-1)} - \frac{N_j}{N_{s_i}}\left(\boldsymbol{f}_{s_i}^{(t-1)} \pm \boldsymbol{f}_j\right), \tag{20}$$

where $\boldsymbol{f}_{s_i}^{(t-1)}$ and $\boldsymbol{f}_{s_i}^{(t)}$ are the image features of superpixel s_i before and after the update, $\boldsymbol{f}_j$ and N_j are respectively the image features and the number of pixels in macropixel j, and N_{s_i} is the number of pixels in s_i after the update.

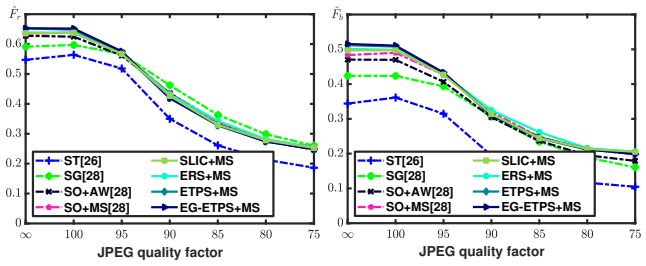
REFERENCES

[1] P. W. Wong and N. Memon. Secret and public key image watermarking schemes for image authentication and ownership verification. *IEEE Trans. Image Process.*, 10(10):1593–1601, 2001.

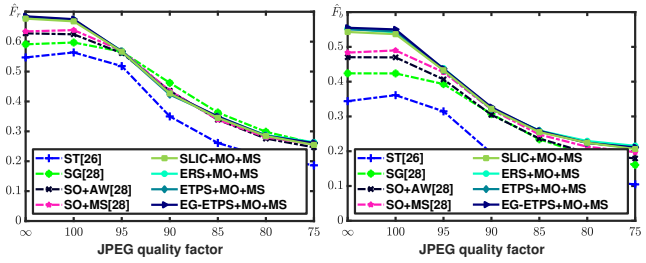
[2] M. U. Celik, G. Sharma, and A M. Tekalp. Lossless watermarking for image authentication: a new framework and an implementation. *IEEE Trans. Image Process.*, 15(4):1042–1049, 2006.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access



(a) $\hat{F}_r$ on entire RTD (b) $\hat{F}_b$ on entire RTD



(c) $\hat{F}_r$ on entire RTD (d) $\hat{F}_b$ on entire RTD

FIGURE 13: Impact of JPEG compression on localization performance. '∞' in the x-axis corresponds to uncompressed images. (a): segmentation-based vs. single-orientation w.r.t. $\hat{F}_r$; (b): segmentation-based vs. single-orientation w.r.t. $\hat{F}_b$; (c): multi-orientation vs. single-orientation w.r.t. $\hat{F}_r$; (d): multi-orientation vs. single-orientation w.r.t. $\hat{F}_b$.

[3] H. Farid and S. Lyu. Higher-order wavelet statistics and their application to digital forensics. In *Proc. IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR)*, volume 8, pages 94–94, 2003.

[4] J. Wang, T. Li, Y.-Q. Shi, S. Lian, and J. Ye. Forensics feature analysis in quaternion wavelet domain for distinguishing photographic images and computer graphics. *Multimedia Tools Appl.*, 76(22):23721–23737, 2017.

[5] X. Shen, Z. Shi, and H. Chen. Splicing image forgery detection using textural features based on the grey level co-occurrence matrices. *IET Image Process.*, 11(1):44–53, 2016.

[6] L. Verdoliva, D. Cozzolino, and G. Poggi. A feature-based approach for image tampering detection and localization. In *Proc. IEEE Int. Workshop Inf. Forensics Security*, pages 149–154, 2014.

[7] H. Li, W. Luo, X. Qiu, and J. Huang. Identification of various image operations using residual-based features. *IEEE Trans. Circuits Syst. Video Technol.*, 2016.

[8] M. Kirchner. Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue. In *Proc. ACM Workshop Multimedia Security*, pages 11–20, 2008.

[9] X. Liu, W. Lu, Q. Zhang, J. Huang, and Y.-Q. Shi. Downscaling factor estimation on pre-jpeg compressed images. *IEEE Trans. Circuits Syst. Video Technol.*, 2019.

[10] M. C Stamm and KJ Ray Liu. Forensic detection of image manipulation using statistical intrinsic fingerprints. *IEEE Trans. Inf. Forensics Security*, 5(3):492–506, 2010.

[11] X. Lin, C.-T. Li, and Y. Hu. Exposing image forgery through the detection of contrast enhancement. In *Proc. Int. Conf. Image Process.*, pages 4467–4471, 2013.

[12] X. Kang, M. C Stamm, A. Peng, and KJ Ray Liu. Robust median filtering forensics using an autoregressive model. *IEEE Trans. Inf. Forensics Security*, 8(9):1456–1468, 2013.

[13] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra. A sift-based forensic method for copy–move attack detection and transformation recovery. *IEEE Trans. Inf. Forensics Security*, 6(3):1099–1110, 2011.

[14] Irene Amerini, Rudy Becarelli, Roberto Caldelli, and Andrea Del Mastio. Splicing forgeries localization through the use of first digit features. In *Proc. IEEE Int. Workshop Inf. Forensics Security*, pages 143–148, 2014.

[15] W. Wang, J. Dong, and T. Tan. Exploring dct coefficient quantization effects for local tampering detection. *IEEE Trans. Inf. Forensics Security*, 9(10):1653–1666, 2014.

[16] P. Korus and J. Huang. Multi-scale fusion for improved localization of malicious tampering in digital images. *IEEE Trans. Image Process.*, 25(3):1312–1326, 2016.

[17] E. Kee and H. Farid. Exposing digital forgeries from 3-D lighting environments. In *Proc. Int. Workshop Inf. Forensics Security*, pages 1–6, 2010.

[18] W. Zhang, X. Cao, J. Zhang, J. Zhu, and P. Wang. Detecting photographic composites using shadows. In *Proc. Int. Conf. Multimedia and Expo*, pages 1042–1045, 2009.

[19] E. Kee, J. F O'brien, and H. Farid. Exposing photo manipulation from shading and shadows. *ACM Trans. Graph.*, 33 (5):165–1, 2014.

[20] J. F O'Brien and H. Farid. Exposing photo manipulation with inconsistent reflections. *ACM Trans. Graph.*, 31(1):4–1, 2012.

[21] H. Yao, S. Wang, Y. Zhao, and X. Zhang. Detecting image forgery using perspective constraints. *IEEE Signal Process. Lett.*, 19(3):123–126, 2012.

[22] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva. Image forgery localization via fine-grained analysis of CFA artifacts. *IEEE Trans. Inf. Forensics Security*, 7(5):1566–1577, 2012.

[23] T.-T. Ng, S.-F. Chang, and M.-P. Tsui. Using geometry invariants for camera response function estimation. In *Proc. Conf. Computer Vision Pattern Recognition*, pages 1–8, 2007.

[24] Y.-F. Hsu and S.-F. Chang. Camera response functions for image forensics: an automatic algorithm for splicing detection. *IEEE Trans. Inf. Forensics Security*, 5(4):816–825, 2010.

[25] I. Yerushalmy and H. Hel-Or. Digital image forgery detection based on lens and sensor aberration. *Int. J. Computer Vision*, 92(1):71–91, 2011.

[26] M. Chen, J. Fridrich, M. Goljan, and J. Lukás. Determining image origin and integrity using sensor noise. *IEEE Trans. Inf. Forensics Security*, 3(1):74–90, 2008.

[27] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva. A Bayesian-MRF Approach for PRNU-Based Image Forgery Detection. *IEEE Trans. Inf. Forensics Security*, 9(4):554–567, 2014.

[28] P. Korus and J. Huang. Multi-Scale Analysis Strategies in PRNU-Based Tampering Localization. *IEEE Trans. Inf. Forensics Security*, 12(4):809–824, 2017.

[29] X. Pan, X. Zhang, and S. Lyu. Exposing image splicing with inconsistent local noise variances. In *Proc. Int. Conf. Computational Photography*, pages 1–10, 2012.

[30] L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. J Delp, and S. Tubaro. Tampering Detection and Localization through Clustering of Camera-Based CNN Features. In *Proc. Conf. Computer Vision Pattern Recognition Workshops*, pages 1855–1864, 2017.

[31] I. Amerini, T. Uricchio, L. Ballan, and R. Caldelli. Localization of jpeg double compression through multi-domain convolutional neural networks. In *Proc. IEEE CVPR Workshop on Media Forensics*, 2017.

[32] D. Cozzolino and L. Verdoliva. Noiseprint: a cnn-based camera model fingerprint. *IEEE Trans. Inf. Forensics Security*, 2019.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042780, IEEE Access

[33] M. Barni and A. Costanzo. Dealing with uncertainty in image forensics: A fuzzy approach. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, pages 1753–1756, 2012.

[34] 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*, 8(4):593–607, 2013.

[35] L. Gaborini, P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro. Multi-clue image tampering localization. In *Proc. IEEE Int. Workshop Inf. Forensics Security*, pages 125–130, 2014.

[36] D. Cozzolino, D. Gragnaniello, and L. Verdoliva. Image forgery detection through residual-based local descriptors and block-matching. In *Proc. IEEE Int. Conf. Image Process.*, pages 5297–5301, 2014.

[37] A. Ferreira, S. C Felipussi, C. Alfaro, P. Fonseca, J. E Vargas-Muñoz, J. A dos Santos, and A. Rocha. Behavior knowledge space-based fusion for copy–move forgery detection. *IEEE Trans. Image Process.*, 25(10):4729–4742, 2016.

[38] G. Chierchia, S. Parrilli, G. Poggi, L. Verdoliva, and C. Sansone. PRNU-based detection of small-size image forgeries. In *Proc. Int. Conf. Digit. Signal Process.*, pages 1–6, Jul 2011.

[39] G. Chierchia, D. Cozzolino, G. Poggi, C. Sansone, and L. Verdoliva. Guided filtering for PRNU-based localization of small-size image forgeries. In *Proc. Int. Conf. Acoust., Speech and Signal Process.*, pages 6231–6235, May 2014.

[40] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(6):1397–1409, 2012.

[41] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2274–2282, 2012.

[42] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy rate superpixel segmentation. In *Proc. Conf. Computer Vision Pattern Recognition*, pages 2097–2104, 2011.

[43] J. Yao, M. Boben, S. Fidler, and R. Urtasun. Real-time coarse-to-fine topologically preserving segmentation. In *Proc. Conf. Computer Vision Pattern Recognition*, pages 2947–2955, 2015.

[44] K. Yamaguchi, D. McAllester, and R. Urtasun. Efficient joint segmentation, occlusion labeling, stereo and flow estimation. In *European Conf. Computer Vision*, pages 756–771, 2014.

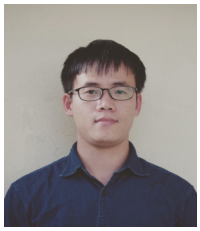
[45] Weiwei Zhang, Xinhua Tang, Zhenghong Yang, and Shaozhang Niu. Multi-scale segmentation strategies in prnu-based image tampering localization. *Multimedia Tools Appl.*, 78(14):20113–20132, 2019.

[46] Y. Liu, Q. Guan, X. Zhao, and Y. Cao. Image forgery localization based on multi-scale convolutional neural networks. In *Proc. ACM Workshop Inform. Hiding Multimedia Security*, pages 85–90, 2018.

[47] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Proc. Advances in Neural Inform. Process. Systems*, pages 91–99, 2015.

[48] P. Korus and J. Huang. Evaluation of random field models in multi-modal unsupervised tampering localization. In *IEEE Int. Workshop Inf. Forensics Security*, pages 1–6, 2016.

[49] G. Csurka, D. Larlus, F. Perronnin, and F. Meylan. What is a good evaluation measure for semantic segmentation? In *Proc. British Mach. Vis. Conf.*, volume 27, 2013.



XUFENG LIN received the B.E. degree in electronic and information engineering from the Hefei University of Technology, Hefei, China, in 2009, the M.E. degree in signal and information processing from the South China University of Technology, Guangzhou, China, in 2012, and the Ph.D. degree in computer science from the University of Warwick, Coventry, U.K., in 2017. He is currently a Research Fellow with the School of Information and Technology, Deakin University, Australia. His research interests include digital forensics, multimedia security, machine learning, and computer vision.

CHANG-TSUN LI (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from National Defence University (NDU), Taiwan, in 1987, the M.Sc. degree in computer science from the U.S. Naval Postgraduate School, USA, in 1992, and the Ph.D. degree in computer science from the University of Warwick, U.K., in 1998. From 1998 to 2002, he was an Associate Professor with the Department of Electrical Engineering, NDU. He was a Visiting Professor with the Department of Computer Science, U.S. Naval Postgraduate School, in 2001. He was a Professor with the Department of Computer Science, University of Warwick, in January 2017. From January 2017 to February 2019, he was a Professor with Charles Sturt University, Australia. He is currently a Professor with the School of Information Technology, Deakin University, Australia. His research interests include multimedia forensics and security, biometrics, data mining, machine learning, data analytics, computer vision, image processing, pattern recognition, bioinformatics, and content-based image retrieval. The outcomes of his multimedia forensics research have been translated into award-winning commercial products protected by a series of international patents and have been used by a number of police forces and courts of law around the world. He also served as a member of the international program committees for several international conferences. He involved in the organisation of many international conferences and workshops. He is also actively contributing keynote speeches and talks at various international events. He is also the EURASIP Journal of Image and Video Processing (JIVP) and an Associate Editor of IET Biometrics.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/