IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 14, 2019, accepted August 23, 2019, date of publication August 29, 2019, date of current version September 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938467

An Improved Algorithm for Digital Image Authentication and Forgery Localization Using Demosaicing Artifacts

NHAN LE AND FLORENT RETRAINT

Laboratory of System Modeling and Dependability (LM2S), ICD, CNRS FRE 2019, Troyes University of Technology, 10004 Troyes, France

Corresponding author: Nhan Le (nhan.le@utt.fr)

This work was supported by the National Research Agency (ANR) Project DEFACTO under Project ANR-16-DEFA-0002.

ABSTRACT This paper focuses on the digital image authentication and forgery localization using demosaicing artifacts. The aim is to build an algorithm allowing a bridge between the color filter array pattern and demosaicing algorithm estimation, and the statistical analysis of demosaicing artifacts in spatial domain to improve the authentication and localization performance. After analyzing the evolution of demosaicing traces in camera acquisition pipeline, a robust feature statistic characterizing demosaiced digital images is first developed on the basis of the noise residue of green channel. Such a feature statistic is less sensitive to the edges problem because only the smooth region of green channel is used in the development. Next, a single normal mixture model is proposed to describe the probability distribution of feature statistics for both original and tampered images. Therefore, normality tests can be used to authenticate automatically digital images. The authentication performance can be further improved by human interpretation of supported graphic tools. Finally, a penalized expectation-maximization algorithm is used to localize forged areas in tampered images. Numerous comparative studies on four well-known datasets show that the developed algorithm yields better performance and robustness than existing forensics algorithms of the same kind.

INDEX TERMS Demosaicing traces, digital image authentication, forgery localization, normal mixture model, penalized expectation-maximization algorithm.

I. INTRODUCTION

Images traditionally describe the truth of what has happened in real-world. Nevertheless, in today's digital age, the trustworthiness of such content is of great concern due to the dissemination of easy-to-use and low-cost image editing tools. As a consequence, digital forensics have emerged as an indispensable research field to restore some trust to digital images. Indeed, a large number of techniques for digital image forensics have been developed over the last two decades (see e.g., [1]–[5] for some recent reviews). Generally speaking, these techniques are classified into two major categories [6]: active (non-blind) approach and passive (blind) approach. With ''active'', we mean that some preset authentic information (e.g., watermark, signature) embedded in digital images is required to examine their truthfulness. Whereas, with ''passive'', certain of intrinsic traces in the image acquisition or some specific traces left by forgeries are exploited to distinguish between tampered and natural images. As such, the passive approach does not rely on any prior information, and hence having broader applications than the active approach [7].

The present paper addresses a passive algorithm for digital image authentication and forgery localization using demosaicing traces. In a digital camera acquisition pipeline, demosaicing (also known as color filter array (CFA) interpolation) is an upstream operation for reconstructing a full color image from the sampled data overlaid with a CFA (see Fig. 1). Although such an acquisition pipeline is common for most of commercial camera devices, each step therein is personalized according to manufacturer choices. Traces left by demosaicing are thus different but specific for different camera brands and/or models. When these traces are totally missing in an image, or when there exists any inconsistency in traces for different image regions, the photographic image is likely to be tampered. Therefore, demosaicing traces can

The associate editor coordinating the review of this article and approving it for publication was Kuo-Hui Yeh.

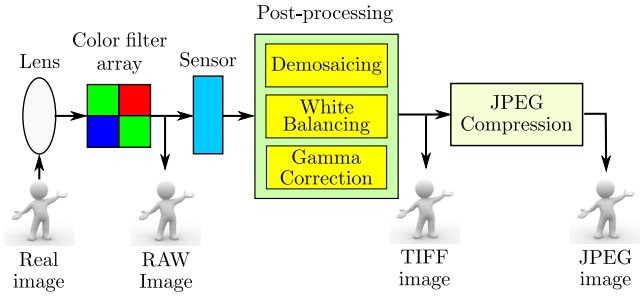

FIGURE 1. Illustration of a typical acquisition pipeline in a digital camera.

be taken as evidence to assess the credibility of a digital image.

Hereinafter, to better understand the novelty of the algorithm proposed in this paper, we first analyze previous works on image forensics considering demosaicing traces. Next, we state the paper scope and its contributions.

A. ANALYSIS OF PREVIOUS WORKS

According to the nature of demosaicing traces, we can categorize previous works on image forgery detection and localization into two main classes. The first class consists of algorithms aiming at estimating from a digital image the CFA pattern and/or the demosaicing algorithm inherent in camera devices. The second class characterizes algorithms used to evaluate the presence/absence of artifacts generated by demosaicing operations.

1) FIRST CLASS

As to the first stream in the class, the works [8]–[12] represent key methods for CFA pattern identification. In [8], the Bayer CFA pattern is identified by minimizing the difference between the raw sensor signal and the inverse demosaiced signal. In [9], the identification is performed via an intermediate value counting algorithm developed from the observation that the value of interpolated color samples is always between the minimum and maximum values of their neighbors. Other method is to compute the ratio between the average noise variance of interpolated pixels and of acquired pixels for all possible candidate CFA patterns of a digital image; the pattern providing the largest ratio is considered as the true one [10]. More recently, color difference blocks are proposed as a means for estimating the CFA configuration in [11], [12].

Beyond the CFA pattern identification, works within the second stream focus more especially on estimating demosaicing algorithms. In [13], an expectation-maximization (EM) algorithm is employed to estimate the coefficients of linear interpolation kernels. An improvement is made in [14] by combining the EM algorithm with average second-order derivative spectrum [15] to obtain interpolation coefficients from smooth and non-smooth regions of images separately. Also partition an image into smooth, horizontal and vertical non-smooth regions, a two-step estimation process is proposed in [16]: (*i*) linear interpolation coefficients associated with each of candidate CFA patterns are first derived using the singular value decomposition, (*ii*) a minimum interpolation error criterion is next used to jointly identify the correct CFA pattern and demosaicing algorithm for separate image regions. In [17], an accurate method based on the partial second-order image derivative correlation models is proposed to recover demosaicing formulas. The method allows to take into account the correlation between three color channels of the image.

Since the CFA pattern and/or the demosaicing algorithm are specific for each class/branch/model of camera devices, they can be used as proofs for image forgery detection and localization. For instance, in [18], [19], the inconsistencies among the estimated demosaicing coefficients are exploited to check if an image has undergone any form of subsequent processing. By remarking that the CFA pattern is changed if the image color is modified, the authors of [20] have designed an advanced intermediate value counting algorithm for measuring the change in the CFA pattern, and hence localizing the extent of color modification in digital images. We also note that estimated CFA pattern and/or demosaicing algorithm are used not only for assessing the credibility of images content, but also for camera source identification (see e.g., [21]). However, the latter is out of the paper scope.

2) SECOND CLASS

The first stream in the second class relies on periodic artifacts caused by the demosaicing. In fact, the image sensors in a CFA are usually organized periodically [22]. Besides, many demosaicing algorithms behave as a filtering process where missing signals are interpolated by periodically applying an interpolation kernel to acquired signals (see e.g., [13]). Therefore, periodicity in demosaicing artifacts is intrinsic to digital photographs, and may help to authenticate images. Inspired by this idea, Popescu and Farid build in [13] a probability map to expose the periodic pattern of correlated pixels. When interpolated pixels are present, the periodicity of the map is clearly visible in the Fourier domain. Such an analysis can be applied to different areas of the tested image to detect the presence of local tampering, however the area size should not be smaller than 256×256 to assure the accuracy of the results. Observing that the variance of the second derivative of interpolated images is periodic [15], Gallagher and Chen use in [23] high pass filtering and Fourier analysis to compute periodic frequency locally. Forged regions of an image will have a different periodicity than the rest. This method is applicable for image blocks with size 64×64 or more. Another method based on periodicity analysis of prediction errors variance is recently proposed by Li *et al.* in [24]. Assuming that interpolation errors are Gaussian distributed, a posterior probability map is derived according to Bayes' rule. Such a map shows the periodic interpolation of the tested image in an intuitive way. An analysis with two dimensional discrete Fourier transform allows to capture the periodicity. By experiments, Li *et al.* show that this method can be applied for local tampering detection with blocks size up to 32×32.

For the second stream, methods for digital image forgery detection are based on the differences in the distribution of acquired and interpolated pixels. In [25], Dirik and Memon recognize that the low pass nature of CFA demosaicing make the variance of the sensor noise in interpolated pixels significantly lower than acquired pixels. As a result, demosaicing artifacts can be measured by a ratio of noise variances between interpolated and acquired pixels. If this ratio is close to 1, tampering has been performed on the image. The method presents a good performance for image blocks with size greater than 96×96. Sharing the same idea, Ferrara *et al.* [26] carry out fine-grained analysis of CFA artifacts and propose a feature to measure the presence of demosaicing even at the smallest 2×2 block level. However, numerical experiments show that the tampering localization yields the best performance at 8×8 block size. While the above works consider spatial features of demosaicing artifacts, González-Fernández *et al.* [27] are rather interested in their spectrum. Indeed, by computing the probability of each interpolated pixel and then applying the discrete cosine transform (DCT) on small blocks of the probability map, the presence/absence of the demosaicing artifacts within a block could be verified via the DCT coefficient at the highest frequency. Experiments show that the method is reliable for blocks of size 16×16.

Compared to the first stream, methods within the second stream brings out better localization resolution and higher fidelity. However, their performance seems more sensitive to JPEG compression than the first stream, especially when the compression quality is significantly low.

3) COMMON REMARKS

Since the demosaicing is an upstream operation in the acquisition pipeline of digital cameras, its traces are independent of forgeries such as cloning, splicing, inpainting, resizing, etc. Therefore, methods based on demosaicing traces do not target any specific forgery operation, but are rather applicable to a variety of operations. Despite this advantage, demosaicing traces are easily destroyed by JPEG compression, even with very high quality levels. This is why these methods is suitable to uncompressed or less-compressed photographs. This is a common and almost unavoidable limitation of demosaicing traces-based methods. Moreover, very often a comparison threshold is required to detect or localize forgeries (see e.g., [24]–[26]). In practice, the choice of such a threshold is not easy and may be very influential in the robustness of the methods. Besides, employed demosaicing traces (i.e., CFA pattern, demosaicing algorithm, periodicity of demosaicing artifacts, differences in the distribution of acquired and interpolated pixels) are closely correlated. However, the above classification of related works implies that they are treated separately in most existing algorithms. Jointly use these traces could improve the performance of image forgery detection and localization.

B. SCOPE, CONTRIBUTIONS AND ORGANIZATION

To authenticate digital photographs and localize forgeries of tampered images efficiently, we develop in the present paper a hybrid algorithm consisting of CFA pattern identification, demosaicing algorithm estimation, and artifacts analysis. According to the above classification of demosaicing traces-based works, the developed algorithm allows a bridge between the first class and the second class. Aiming at fine-grained detection, the identification scheme proposed in [16] is adopted to reveal the CFA pattern and interpolation kernel, and a local analysis similar to [26] is used to expose demosaicing artifacts from prediction residues. Their combination allows to build robust feature statistics characterizing the presence/absence of demosaicing artifacts. Despite these similarities, several improvements have been made.

1) We theoretically analyze how the mean and variance of prediction residues in interpolated and acquired signals evolve in a digital camera acquisition pipeline from the RAW format to the JPEG format. This is the basis to build feature statistics of interest. Besides, the analysis also helps to explain in part why the demosaicing traces-based approach is less effective with JPEG compressed images.
2) In most demosaicing traces-based algorithms (see e.g., [13], [24], [26], [28]), prediction residues are given from entire pixels. Here, by further partitioning a pixel into content and noise parts, we realize that the demosaicing behaves in the same manner for content, noise, as well as entire pixel. However, due to the weak energy of noise, demosaicing traces are most visible in noise part. This is why feature statistics used in the algorithm are derived from noise residues rather than the residues of content or entire pixels.
3) As mentioned in [26], the presence of sharp edges in images reduces the quality of feature statistics, because it may disrupt the correlation between interpolated and acquired residues. To overcome this obstacle, we build our feature statistics from smooth region of the green channel instead of the entire channel. As such, the edges problem is no longer a great concern.
4) Regarding the authentication, after verifying the standard normal distribution of feature statistics in natural images, we adopt normality tests (i.e., Anderson-Darling test, one-sample Kolmogorov-Smirnov test, Jarque-Bera test, and Lilliefors test) to verify their integrity automatically. Such an automatic detection is somewhat scattered in demosaicing traces-based works. Besides, we also provide other tools (i.e., Q-Q plot diagram, probability distribution curves, and localization map) which can help to improve authentication performance by human interpretation. It is noticed that the image authentication is more general than the content integrity verification, because it encompasses also the source attribution, as well as the context verification [4]. Nevertheless, for the sake of consistency, we follow the common

convention in the Multimedia Forensics community, and consider the terms "*authentication*" and "*integrity verification*" having the same meaning in this paper.

5) By modeling the distribution of feature statistics in tampered images as a normal mixture, we applied a penalized EM algorithm to localize forged regions. Since the algorithm does not require a comparison threshold, it provides robuster localization results than traditional threshold-based methods [25], [26].

Numerous numerical experiments also confirm that the proposed algorithm yields better performance than existing algorithms of the same kind.

The remainder of the paper is organized as follows. Section II focuses on analyzing statistical properties of demosaicing traces when they evolve in a digital camera acquisition pipeline. The analysis is next extended to develop feature statistics measuring the unbalance between the local variances of prediction residues for the green channel of demosaiced images in Section III. A robust feature statistic is eventually identified. Section IV develops a single statistical model for the proposed feature statistic. Such a model is next used to authenticate digital images and to localize the tampered regions. Numerous numerical experiments are provided in Section V. Finally, some conclusions and perspectives are discussed in Section VI.

II. STATISTICAL ANALYSIS OF DEMOSAICING TRACES

Analyzing statistical properties of demosaicing traces is a key step to define a relevant feature statistic for tampering detection and localization. In this section, we are interest more especially in the mean and variance of prediction residues in both interpolated and acquired signals. In ideal cases, we prove that the residues variance in acquired signals is greater than in interpolated signals, while their mean values are always 0. However, such a difference weakens under the impact of JPEG compression. Hereinafter, as in [26], only the green channel is considered, because it is upsampled by a factor 2. The same number of acquired and interpolated pixels in each generic square block leads to the same estimation reliability for both classes of pixels. Moreover, for an easier representation, the analysis is just done for one dimensional signals (i.e., a given row in the green channel of digital image). The results for two dimensional signals can be interpreted in a similar way.

A. DEMOSAICING TRACES AT PIXEL LEVEL

Let consider a row in the green color channel of an image obtained by using a Bayer filter and a linear demosaicing algorithm. Without loss of generality, we assume that the Bayer CFA pattern is arranged in the manner that acquired pixels are at even positions of the row (interpolated pixels will be at odd positions). Each acquired pixel consists of two parts: *real scene content* and *noise*. The real scene content is the true image information, while noise is generated everywhere during camera imaging. Mathematically, an acquired pixel $p_A(x)$ at position x on the row can be expressed as [10]

$$p_A(x) = c_A(x) + n_A(x), \quad \text{if } x \text{ even} \tag{1}$$

where $c_A(x)$ and $n_A(x)$ are respectively the content and noise of $p_A(x)$. Applying a linear demosaicing algorithm with interpolation kernel h_u to acquired pixels, we obtain the value of an interpolated pixel at position x on the row as

$$p_I(x) = \sum_{u \neq 0} h_u p_A(x+u), \quad \text{if } x \text{ odd}, \tag{2}$$

where $\sum_{u\neq 0} h_u = 1$, and u takes odd values. Substituting (1) in (2), we obtain

$$p_I(x) = c_I(x) + n_I(x), \quad \text{if } x \text{ odd}, \tag{3}$$

where $c_I(x)$ and $n_I(x)$ are the content and noise of $p_I(x)$

$$c_I(x) = \sum_{u \neq 0} h_u c_A(x+u), \tag{4}$$

and

$$n_I(x) = \sum_{u \neq 0} h_u n_A(x+u). \tag{5}$$

The similarity of (2), (4) and (5) implies that the content part and noise part are altered in the same way as the entire pixel after the demosaicing. In other words, we can find traces left by the demosaicing in the content, in the noise and in the entire pixel at odd positions of the considered row.

Hereinafter, we use s to represent a signal which may be either the entire pixel, its content part or its noise part. Accordingly, s_A and s_I denote respectively the acquired and interpolated signals. The resulting signal s_R is either an acquired signal or an interpolated signal

$$s_R(x) = \begin{cases} s_A(x) & \text{if } x \text{ even}, \\ s_I(x) = \sum_{u\neq 0} h_u s_A(x+u) & \text{if } x \text{ odd}. \end{cases} \tag{6}$$

We will analyze the mean and variance of signal residue to find out some useful demosaicing artifacts when the digital image is in TIFF format and in JPEG format.

B. DEMOSAICING ARTIFACTS IN TIFF IMAGES

The RAW image, after demosaicing, white balancing and gamma correction, becomes an uncompressed high-quality image in TIFF format. Since, the white balancing and the gamma correction are lossless operation in terms of information [29], we expect that the characteristics of demosaicing artifacts do not much changed. Let k_u, with $\sum_{u\neq 0} k_u = 1$, is an estimate of the interpolation kernel h_u when the considered image is in TIFF format, then the predicted signal is computed as

$$s_P^{\text{tiff}}(x) = \sum_{u \neq 0} k_u s_R^{\text{tiff}}(x+u). \tag{7}$$

When u is odd, then $x+u$ is odd if x even, and is even otherwise. Using (6), $s_P(x)$ can be rewritten by

$$s_P^{\text{tiff}}(x) = \sum_{u\neq 0} k_u s_R^{\text{tiff}}(x+u) = \begin{cases} \sum_{u\neq 0} k_u \sum_{v\neq 0} h_v s_A^{\text{tiff}}(x+u+v) & \text{if } x \text{ even}, \\ \sum_{u\neq 0} k_u s_A^{\text{tiff}}(x+u) & \text{if } x \text{ odd}. \end{cases} \tag{8}$$

The residue, which is the difference between resulting and predictive signals, is thus expressed as

$$e^{\text{tiff}}(x) = s_R^{\text{tiff}}(x) - s_P^{\text{tiff}}(x) = \begin{cases} e_A^{\text{tiff}}(x) & \text{if } x \text{ even}, \\ e_I^{\text{tiff}}(x) & \text{if } x \text{ odd}, \end{cases} \tag{9}$$

where

$$e_A^{\text{tiff}}(x) = s_A^{\text{tiff}}(x) - \sum_{u\neq 0} k_u \sum_{v\neq 0} h_v s_A^{\text{tiff}}(x+u+v), \tag{10}$$

and

$$\begin{aligned} e_I^{\text{tiff}}(x) &= \sum_{u\neq 0} h_u s_A^{\text{tiff}}(x+u) - \sum_{u\neq 0} k_u s_A^{\text{tiff}}(x+u) \\ &= \sum_{u\neq 0} (h_u - k_u)\, s_A^{\text{tiff}}(x+u). \end{aligned} \tag{11}$$

Since the size of kernel window is usually small, acquired signals in such a window can be assumed identical independent distributed (i.i.d.) with mean μ and variance σ^2. Consequently, the mean $E\left[e^{\text{tiff}}(x)\right]$ and variance $\text{var}\left[e^{\text{tiff}}(x)\right]$ of the residue of TIFF image are expressed as

- if x is even, then

$$E\left[e_A^{\text{tiff}}(x)\right] = \mu - \sum_{u\neq 0} k_u \sum_{v\neq 0} h_v \mu = 0, \tag{12}$$

and

$$\text{var}\left[e_A^{\text{tiff}}(x)\right] = \sigma^2 \left(1 + \sum_{u\neq 0} k_u^2 \sum_{v\neq 0} h_v^2\right). \tag{13}$$

- if x is odd, then

$$E\left[e_I^{\text{tiff}}(x)\right] = \sum_{u\neq 0} h_u \mu - \sum_{u\neq 0} k_u \mu = 0. \tag{14}$$

and

$$\text{var}\left[e_I^{\text{tiff}}(x)\right] = \sigma^2 \sum_{u\neq 0} (h_u - k_u)^2. \tag{15}$$

These results are obtained following similar steps in [26].

We find that $E\left[e^{\text{tiff}}(x)\right]$ is always equal to 0 for whatever position of x. Meanwhile, when the estimate of the interpolation kernel k_u is close to the original one h_u, $\text{var}\left[e^{\text{tiff}}(x)\right]$ is close to 0 at the positions of interpolated signal (i.e., $\text{var}\left[e_I^{\text{tiff}}(x)\right] \to 0$), while it is greater than σ^2 at the positions of acquired signal (i.e., $\sigma^2 < \text{var}\left[e_A^{\text{tiff}}(x)\right] \leq 2\sigma^2$). As such, the difference between variances of residues in acquired and interpolated signals can be seen as a useful demosaicing artifact for TIFF images.

C. DEMOSAICING ARTIFACTS IN JPEG IMAGES

For storage, the high-quality TIFF image is compressed into JPEG format (see e.g., [30] for the detail of compression process). If a lossy JPEG compression is applied, the high-frequency components of 8×8 blocks in DCT domain are weakened by quantization which cannot be restored. This results in the local homogenization of 8×8 blocks in spatial domain. As in [24], we can model such a phenomenon by mixing the features of the acquired and interpolated signal residues of TIFF image. Let e_A^{tiff} and e_I^{tiff} be the representatives of acquired and interpolated signal residues in a certain row of 8×8 blocks of the TIFF image, the associated signal residues in 8×8 blocks of the JPEG image can be expressed by

$$\begin{cases} e_A^{\text{jpeg}} = \alpha e_A^{\text{tiff}} + (1-\alpha)\, e_I^{\text{tiff}}, \\ e_I^{\text{jpeg}} = \alpha e_I^{\text{tiff}} + (1-\alpha)\, e_A^{\text{tiff}}, \end{cases} \tag{16}$$

where $\alpha \in [0.5, 1]$ is a weighting factor related to the compression quality Q. When $\alpha = 0.5$, $e_A^{\text{jpeg}} = e_I^{\text{jpeg}} = \frac{1}{2}\left(e_A^{\text{tiff}} + e_I^{\text{tiff}}\right)$; when $\alpha = 1$, $e_A^{\text{jpeg}} = e_A^{\text{tiff}}$ and $e_I^{\text{jpeg}} = e_I^{\text{tiff}}$. These two configurations correspond respectively to a very small value ($Q < 10$) and a very high value ($Q = 100$) of compression quality.

Let consider now the mean and variance of e_A^{jpeg} and e_I^{jpeg}. From (16), we derive, by using (12) and (14), that

$$E\left[e_A^{\text{jpeg}}\right] = E\left[e_I^{\text{jpeg}}\right] = 0, \tag{17}$$

and that

$$\begin{aligned} \text{var}\left[e_A^{\text{jpeg}}\right] = \alpha^2 \text{var}\left[e_A^{\text{tiff}}\right] + (1-\alpha)^2 \text{var}\left[e_I^{\text{tiff}}\right] \\ + 2\alpha(1-\alpha)\,\text{cov}\left[e_A^{\text{tiff}}, e_I^{\text{tiff}}\right], \end{aligned} \tag{18}$$

and

$$\begin{aligned} \text{var}\left[e_I^{\text{jpeg}}\right] = \alpha^2 \text{var}\left[e_I^{\text{tiff}}\right] + (1-\alpha)^2 \text{var}\left[e_A^{\text{tiff}}\right] \\ + 2\alpha(1-\alpha)\,\text{cov}\left[e_A^{\text{tiff}}, e_I^{\text{tiff}}\right]. \end{aligned} \tag{19}$$

Since $\text{var}\left[e_A^{\text{tiff}}\right] \geq \text{var}\left[e_I^{\text{tiff}}\right]$ (see Subsection II-B), we obtain

$$\text{var}\left[e_A^{\text{jpeg}}\right] \geq \text{var}\left[e_I^{\text{jpeg}}\right]. \tag{20}$$

Thus, for the image in JPEG format, the difference between residues variances in acquired and interpolated signals is still a potential demosaicing artifact. But, such a difference is much weaker when Q decreases, and even disappeared when $Q < 10$ (i.e., $\text{var}\left[e_A^{\text{jpeg}}\right] = \text{var}\left[e_I^{\text{jpeg}}\right]$ when $\alpha = 0.5$). This phenomenon is depicted clearly in Fig. 3 of [24].

From the above analyses, we can conclude that, except some special configurations, the unbalance between the variances of signal residues at acquired and interpolated positions is an inherent property of digital images in TIFF and JPEG formats with high compression quality. When the image is strongly compressed, the unbalance is less clear. This explains in part why the demosaicing traces-based approach does not work well with JPEG compressed images.

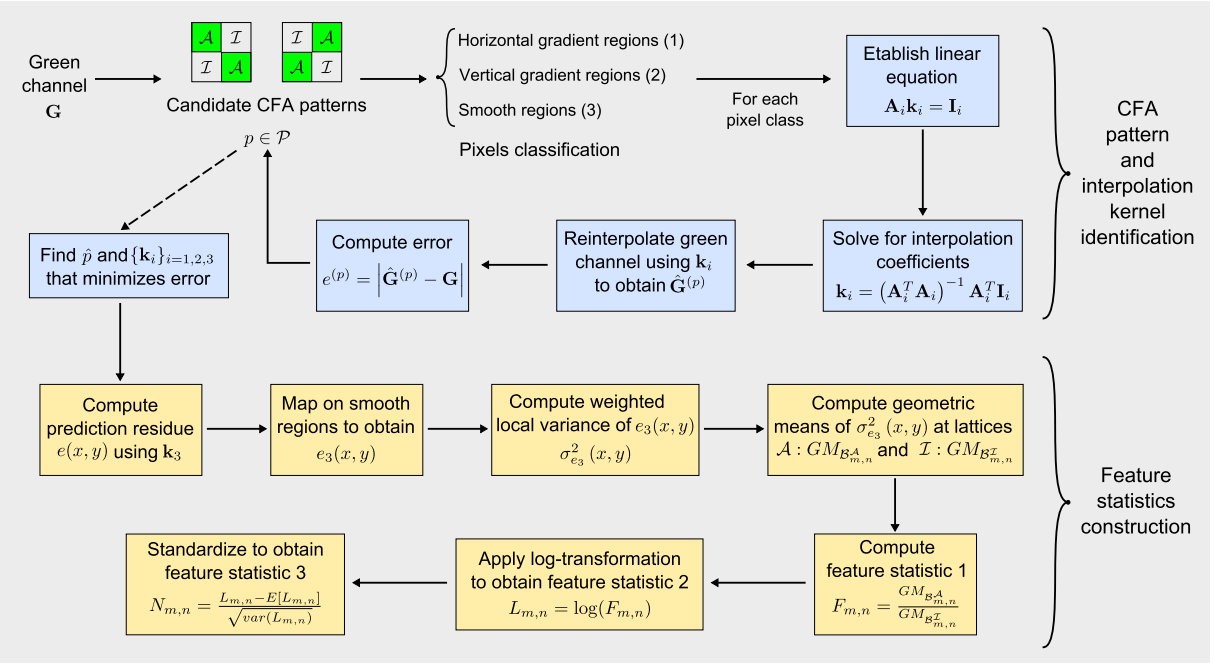

FIGURE 2. Proposed process for Bayer CFA pattern identification, interpolation kernel estimation, and feature statistics construction.

III. FEATURE STATISTICS FOR DEMOSAICED IMAGES

This section aims at extending the analysis in Section II to build feature statistics characterizing demosaiced images. To this end, we develop a complete process consisting of the identification of Bayer CFA pattern, the estimation of interpolation kernel, and the construction of feature statistics measuring the unbalance between the local variances of prediction residues. Especially, only smooth region of the green channel is considered to avoid perturbations due to edges. The main steps of the process is schematically illustrated Fig. 2. In the following, we provide the detailed developments for the process.

A. CFA PATTERN AND INTERPOLATION KERNEL

To avoid the edges effects, we adapt the identification scheme proposed by Swaminathan *et al.* in [16] to reveal the CFA pattern and interpolation kernel for the green channel of the considered image. With a given CFA pattern $p \in \mathcal{P}$ (see Fig. 2), we divide the green channel $\mathbf{G}$ into three kinds of regions based on the gradient features in a local neighborhood as follows. Let $s_R(x, y)$ be the resulting signal value at location (x, y), local gradient profiles along the horizontal and vertical directions are respectively defined as

$$H(x, y) = |s_R(x, y-2) + s_R(x, y+2) - 2s_R(x, y)|,$$

and

$$V(x, y) = |s_R(x-2, y) + s_R(x+2, y) - 2s_R(x, y)|.$$

Let T be a predetermined threshold, the partition of gradient profile plane is summarized as in Table 1. For each region,

TABLE 1. Regions and meanings.

Region	Gradient relation	Meaning
$\mathbf{R}_1$	$H(x,y) - V(x,y) \geq T$	significant horizontal
$\mathbf{R}_2$	$V(x,y) - H(x,y) \geq T$	significant vertical
$\mathbf{R}_3$	$\|V(x,y) - H(x,y)\| < T$	mostly smooth

we approximate interpolated signals with a set of linear equations of acquired signals, such that

$$\mathbf{A}_i\mathbf{k}_i = \mathbf{I}_i, \tag{21}$$

where $\mathbf{A}_i$, $\mathbf{I}_i$ and $\mathbf{k}_i$, $i \in \{1, 2, 3\}$, are respectively the matrix of acquired signals, the vector of interpolated signals, and the linear interpolation kernel of the region R_i. Solving (21) by the well-known least-squares method gives the following kernel $\mathbf{k}_i$

$$\mathbf{k}_i = \left(\mathbf{A}_i^T\mathbf{A}_i\right)^{-1}\mathbf{A}_i^T\mathbf{I}_i, \tag{22}$$

where $\mathbf{A}_i^T$ and $\mathbf{A}_i^{-1}$ denote respectively the transpose and inverse of matrix $\mathbf{A}_i$. Next, the obtained interpolation kernels are then used to reconstruct an estimation $\hat{\mathbf{G}}^{(p)}$ of the green channel $\mathbf{G}$. We repeat the above process to derive the reconstruction error for each CFA pattern. The optimal CFA pattern and interpolation kernel are jointly selected as the combination that yields the lowest reconstruction error.

B. LOCAL WEIGHTED VARIANCE OF RESIDUES

The analysis in Section II is done under the assumption of i.i.d. acquired signals. To adapt its results to the two dimensional green channel of the digital image, we should evaluate

the variance of prediction residues over small $(2K+1) \times (2K+1)$ windows, in which signal values are expected stationary. When sharp edges are present in the tested image, they may disturb this stationary property even for small windows. To overcome this obstacle, we propose using only the smooth region $\mathbf{R}_3$ in the computation of the local variance of residues. By this way, the condition of i.i.d. acquired signals is guaranteed.

More precisely, we first compute the residue of the two dimensional green channel using the interpolated kernel $\mathbf{k}_3$ of the region $\mathbf{R}_3$

$$e(x,y) = s_R(x,y) - \sum_{u,v\neq 0} k_{3,u,v} s_A(x+u, y+v), \quad (23)$$

where $s_R(x,y)$ and $s_A(x,y)$ denote respectively the resulting and acquired signals at the location (x,y) of the green channel, and $k_{3,u,v}$ is an element in $\mathbf{k}_3$. Then, mapping the residue $e(x,y)$ onto the smooth region R_3, we obtain

$$e_3(x,y) = \begin{cases} e(x,y) & \text{if } (x,y) \in \mathbf{R}_3, \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

Following [31], the local weighted variance of $e_3(x,y)$ within $(2K+1) \times (2K+1)$ windows is computed by

$$\sigma^2_{e_3}(x,y) = \frac{1}{c}\left(\sum_{i,j=-K}^{K} \alpha_{i,j} e_3^2(x+i, y+j) - \mu^2_{e_3}(x,y) \right), \quad (25)$$

where $\mu_{e_3}(x,y)$ denotes the local weighted mean

$$\mu_{e_3}(x,y) = \sum_{i,j=-K}^{K} \alpha_{i,j} e_3(x+i, y+j), \quad (26)$$

$c = 1 - \sum_{ij=-K}^{K} \alpha_{ij}^2$ is a scale factor, and α_{ij} are suitable weights given by

$$\alpha_{ij} = \frac{\alpha'_{ij}}{\sum_{i,j} \alpha'_{ij}}, \quad (27)$$

in which $\alpha'_{ij} = W(i,j)$, if $e_3(x+i, y+j)$ belongs to the same class of $e_3(x,y)$; and $\alpha'_{ij} = 0$, otherwise. $W(i,j)$ is a $(2K+1) \times (2K+1)$ Gaussian windows centered at (i,j).

C. DEFINITION OF FEATURE STATISTICS

As shown in Section II, the unbalance between residues variances of acquired and interpolated signals is inherent in demosaiced images. Here, the aim is to develop some feature statistics to expose this unbalance on the green channel locally. To this end, we first divide the variance map of residues with size $N \times N$ into $B \times B$ non-overlapping blocks, where B is a multiple of the length of Bayer's filter. Each $B \times B$ block is composed of $\frac{B^2}{2}$ acquired positions at quincunx lattices $\mathcal{A}$, and $\frac{B^2}{2}$ interpolated positions at complementary quincunx lattices $\mathcal{I}$. Accordingly, we distinguish in a given (m,n) block $\mathcal{B}_{m,n}$, $m,n = 0, \ldots, \frac{N}{B} - 1$, two sets: *acquired variances* $\mathcal{B}^{\mathcal{A}}_{m,n}$ and *interpolated variances* $\mathcal{B}^{\mathcal{I}}_{m,n}$. As in [26], each of these sets can be characterized by the *geometric mean* of their elements

$$GM_{\mathcal{B}^{\mathcal{A}}_{m,n}} = \left(\prod_{m,n\in\mathcal{B}^{\mathcal{A}}_{m,n}} \sigma^2_{e_3}(m,n) \right)^{\frac{1}{\frac{B^2}{2}}}, \quad (28)$$

and

$$GM_{\mathcal{B}^{\mathcal{I}}_{m,n}} = \left(\prod_{m,n\in\mathcal{B}^{\mathcal{I}}_{m,n}} \sigma^2_{e_3}(m,n) \right)^{\frac{1}{\frac{B^2}{2}}}. \quad (29)$$

Note that the *geometric mean* is used instead of the well-known *arithmetic mean* because it is less sensitive to extreme values. We can therefore define the unbalance between the local variance of signal residues at lattices $\mathcal{A}$ and $\mathcal{I}$ in the block $\mathcal{B}_{m,n}$ by the fraction

$$F_{m,n} = \frac{GM_{\mathcal{B}^{\mathcal{A}}_{m,n}}}{GM_{\mathcal{B}^{\mathcal{I}}_{m,n}}}. \quad (30)$$

By numerical experiments, we find that the probability density function (pdf) of $F_{m,n}$, $m,n = 0, \ldots, \frac{N}{B} - 1$, is positively skewed with long tail on the right. Such a form does not allows feasible pdf fitting. This is why we apply the log-transformation to $F_{m,n}$ to favor the normality

$$L_{m,n} = \frac{1}{\frac{B^2}{2}} \sum_{m,n\in\mathcal{B}^{\mathcal{A}}_{m,n}} \log\left(\sigma^2_{e_3}(m,n)\right) - \frac{1}{\frac{B^2}{2}} \sum_{m,n\in\mathcal{B}^{\mathcal{I}}_{m,n}} \log\left(\sigma^2_{e_3}(m,n)\right). \quad (31)$$

Clearly, the feature $L_{m,n}$ represents the difference between the *arithmetic mean* of the logarithm of variances in the set $\mathcal{B}^{\mathcal{A}}_{m,n}$ and $\mathcal{B}^{\mathcal{I}}_{m,n}$. The mean and variance of $L_{m,n}$ vary from image to image. To have a feature statistic independent of image content, $L_{m,n}$ is standardized

$$N_{m,n} = \frac{L_{m,n} - E\left[L_{m,n}\right]}{\sqrt{\text{var}\left[L_{m,n}\right]}}, \quad (32)$$

so that the mean and variance of $N_{m,n}$ are fixed to 0 and 1.

D. ROBUSTNESS OF FEATURE STATISTICS

Since all the signals such as entire pixel, content part, and noise part could be used to compute the feature statistics $F_{m,n}$, $L_{m,n}$ and $N_{m,n}$, two questions arise naturally.

1) Which one among the three feature statistics $F_{m,n}$, $L_{m,n}$ and $N_{m,n}$ is more relevant to characterize demosaiced images?
2) Which signal allows the robustest feature statistic?

The following experiments seek answers to these questions.

Given a natural digital camera image, we consider at first the green channel, its content part and its noise part (see the first line of Fig. 3 for an example given from Image

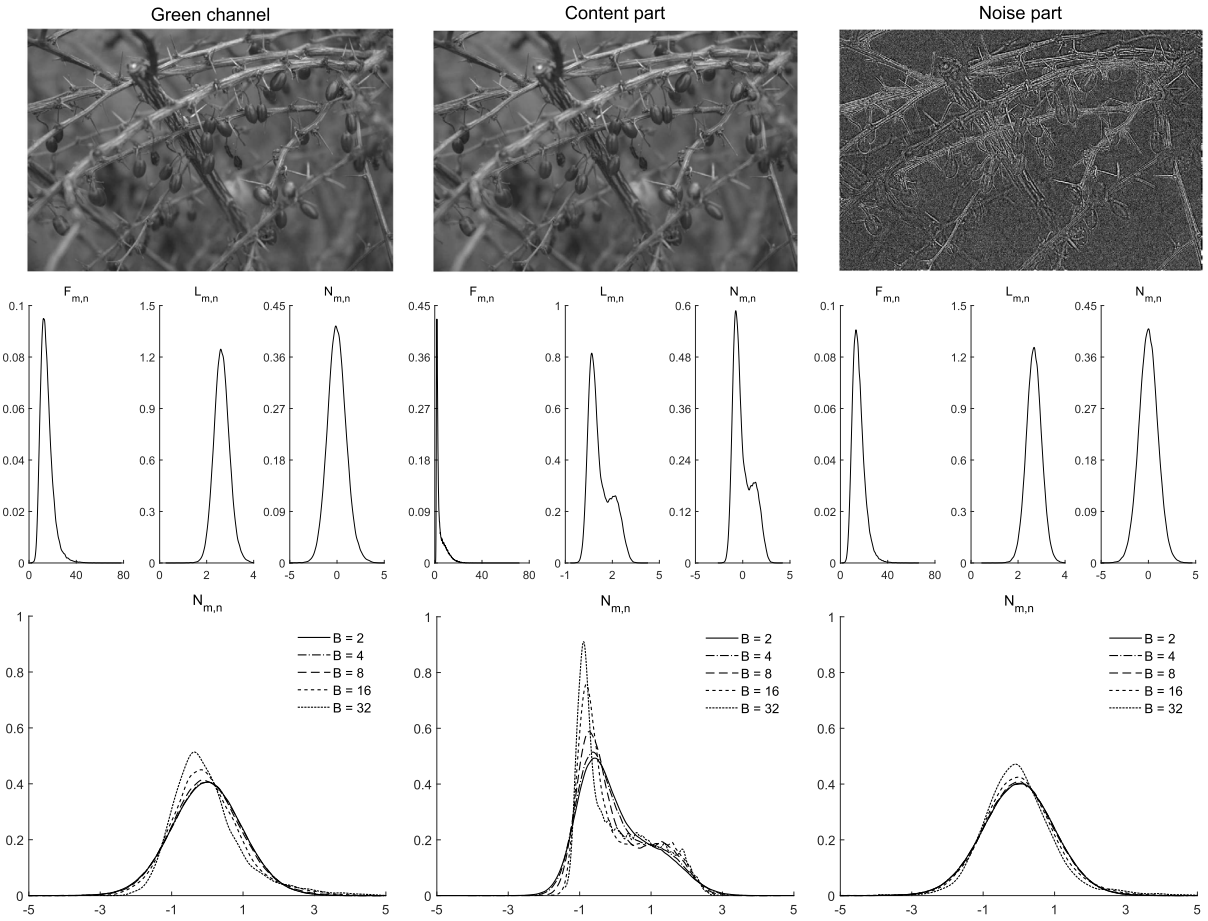

FIGURE 3. Used signals and feature statistics pdfs.

Manipulation Dataset [32]). For each kind of signals, we compute the local weighted variance map of prediction residues (with $K = 3$), and derive the three feature statistics of interest (with $B = 8$). Using kernel density estimation (KDE) method [33], the associated pdf of $F_{m,n}$, $L_{m,n}$ and $N_{m,n}$ are given in the second line of Fig. 3. Clearly, the pdf forms of $F_{m,n}$ are uncontrollable for all kinds of signals. For the pdf forms of $L_{m,n}$ and $N_{m,n}$, the ones obtained from the green channel or its noise part have a nice bell shape as expected, while the ones computed from the content part have not. The reason is that the weak demosaicing traces in the content part is easily covered and/or strongly affected by the real scene data. Meanwhile, for the noise part, although the demosaicing traces is still weak, the energy of noise is relatively small. So that the portion of demosaicing traces in noise is remarkable. The energy of pixels in green channel is high, but demosaicing traces therein are also much stronger. Consequently, the demosaicing artifacts given from the green channel are still significant. Compared to $L_{m,n}$, the advantage of $N_{m,n}$ is that its pdf is independent of image content, and hence more suitable for detection goal. To find out which one between the green channel and its noise part can provide robuster $N_{m,n}$, we vary the size of $B \times B$ blocks and observe the variation in shape of $N_{m,n}$ pdfs (see the third line of Fig. 3). Obviously, the $N_{m,n}$ pdfs given from the noise part is less sensitive to the value of B (especially since $B = 8$). Similar results given by repeating this experiment for various images allow us to confirm that the feature statistic $N_{m,n}$ computed from the noise part is robustest.

In the remainder of this paper, we use $N_{m,n}$ as the default feature statistic. Besides, to simplify the notation, we use a single index instead of double of indices to indicate the feature statistic (i.e., using N_i to denote the feature statistic associated with the i-th $B \times B$ block).

IV. FEATURE STATISTIC MODELING FOR IMAGE AUTHENTICATION AND FORGERY LOCALIZATION

This section aims to develop a single statistical model based on the feature statistic N_i for both natural and tampered digital images. Such a model is next used to authenticate digital images and to localize the tampered regions if any.

A. STATISTICAL MODEL OF FEATURE STATISTIC

As mentioned in Subsection III-D, when the tested image is authentic, N_i is consistent and distributed following a standard normal distribution $\mathcal{N}(0, 1)$. When some regions in

the image have been manipulated by a new content coming either from other regions of the same image (e.g., copy-move forgery, inpainting forgery), or from another images (e.g., splicing forgery), the demosaicing traces in these regions are normally different than the remainder of the image. The feature statistic N_i of such a tampered image is no longer consistent. In this case, we expect that values of N_i come from two different populations $\mathcal{P}_1$ and $\mathcal{P}_2$ corresponding to the untampered and tampered regions respectively. Each region is part of demosaiced images, so $\mathcal{P}_k$, $k \in \{1, 2\}$, is a normal distribution with mean μ_k and variance σ_k. For simplicity, we also set $\sigma_1 = \sigma_2 = \sigma$, because the mean parameters contribute to the separation between $\mathcal{P}_1$ and $\mathcal{P}_2$ in most situations. As a result, the population of N_i for an entire tampered image can be modeled as a normal mixture distribution $\mathcal{NM}(\gamma, \mu_1, \mu_2, \sigma)$ with pdf

$$f(N_i; \gamma, \mu_1, \mu_2, \sigma) = (1-\gamma) \cdot f(N_i; \mu_1, \sigma) + \gamma \cdot f(N_i; \mu_2, \sigma), \tag{33}$$

where γ, $0 \leq \gamma \leq 1$, denotes the proportion of the population 2, and $f(x; \mu_k, \sigma)$ standards for the normal p.d.f $\mathcal{N}(\mu_k, \sigma)$

$$f(N_i; \mu_k, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2\sigma^2}(N_i-\mu_k)^2}. \tag{34}$$

As in [34], to avoid the nonidentifiability of (33), we set $0 \leq \gamma \leq 0.5$. This implies that the population 2 of the model (33) is the tampered population $\mathcal{P}_2$ under the assumption that tampered regions are smaller than the remainder of the image. When $\mu_1 = \mu_2$ or $\gamma = 0$, the model (33) degenerates into a single normal distribution. By this way, we can also use (33) as a model for both authentic and tampered images.

B. MODEL PARAMETERS ESTIMATION

Given the model (33), the next issue is to estimate the parameters γ, μ_1, μ_2 and σ from the set of feature statistics N_i. A penalized EM algorithm has been developed for this issue. Let $\mathbf{N} = \{N_1, \ldots, N_q\}$ be a set of q feature statistics sampled from a normal mixture population $\mathcal{NM}(\gamma, \mu_1, \mu_2, \sigma)$, its ordinary log-likelihood function is given by

$$l_q(\gamma, \mu_1, \mu_2, \sigma) = \sum_{i=1}^{q} \log\left((1-\gamma) f(N_i; \mu_1, \sigma) + \gamma f(N_i; \mu_2, \sigma)\right). \tag{35}$$

As proved in [35], $l_q(\gamma, \mu_1, \mu_2, \sigma) \to \infty$ if $\mu_k \to N_i$ and $\sigma \to 0$ with the other parameters fixed. This implies that ordinary maximum-likelihood estimator of $(\gamma, \mu_1, \mu_2, \sigma)$ is not well-defined [36], [37]. To remedy, Chen *et al.* [35] propose adding penalty term to the ordinary log-likelihood function. Such an approach has been proved efficient because of the strong consistency of maximum likelihood estimators for various penalties on σ [35], [38]. Therefore, we can define a penalized log-likelihood function as

$$pl_q(\gamma, \mu_1, \mu_2, \sigma) = l_q(\gamma, \mu_1, \mu_2, \sigma) + p_q(\sigma), \tag{36}$$

where $p_q(\sigma)$ is the penalty function on σ. To compensate the aforementioned undesirable configuration, we should select $p_q(\sigma)$ such that it is bounded when σ is large, but goes to $-\infty$ as $\sigma \to 0$. Chen *et al.* [39] have recommended

$$p_q(\sigma) = -a_q \cdot \left(\frac{s_q^2}{\sigma^2} + \log\left(\frac{\sigma^2}{s_q^2}\right)\right), \tag{37}$$

where $s_q^2 = \frac{1}{q}\sum_{i=1}^{q} N_i^2$ denotes the sample variance (the sample mean $\bar{N} = \frac{1}{q}\sum_{i=1}^{q} N_i$ is always 0), and a_q is a positive tuning parameter. A large value of a_q implies a strong conviction in the prior estimate of σ [35]. As in the works [40], [41], we choose $a_q = \frac{1}{q}$.

Until now, the estimation problem returns to find the tuple $(\hat{\gamma}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma})$ that maximizes penalized log-likelihood function (36). Moreover, we would like to perform a population clustering for the set of feature statistics $\mathbf{N}$. This is why we try to introduce a vector of binary latent variables $\mathbf{Z} = \{Z_1, \ldots, Z_q\}$ with $Z_i = (Z_{i1}, Z_{i2})$ indicating the cluster of a sample N_i in $\mathbf{N}$. The variable Z_{ik}, $k \in \{1, 2\}$, is defined as follows

$$Z_{ik} = \begin{cases} 1 & \text{if } N_i \text{ is from the population } k \\ 0 & \text{otherwise} \end{cases}, \tag{38}$$

and $\sum_{k=1}^{2} Z_{ik} = 1$. The maximizing the likelihood estimation is now done on the complete data $\{\mathbf{N}, \mathbf{Z}\}$ rather than on the incomplete data only $\mathbf{N}$. As pointed out in [42, page 431], the joint pdfs of Z_i and $N_i \mid Z_i$ are respectively

$$f(Z_i) = \prod_{k=1}^{2} \gamma_k^{Z_{ik}}, \tag{39}$$

and

$$f(N_i \mid Z_i) = \prod_{i=1}^{2} f^{Z_{ik}}(N_i; \mu_k, \sigma), \tag{40}$$

where $\gamma_1 = 1 - \gamma$, $\gamma_2 = \gamma$, and $f(N_i; \mu_k, \sigma)$ is given from (34). Therefore, the complete data likelihood function takes the form

$$f(\mathbf{N}, \mathbf{Z}) = \prod_{i=1}^{q} \prod_{k=1}^{2} \gamma_k^{Z_{ik}} f^{Z_{ik}}(N_i; \mu_k, \sigma). \tag{41}$$

Taking the logarithm, we obtain the complete data log-likelihood function

$$l_c(\gamma, \mu_1, \mu_2, \sigma) = \sum_{i=1}^{q} \sum_{k=1}^{2} Z_{ik} \times \left(\log \gamma_k - \log \sigma - \frac{1}{2\sigma^2}(x - \mu_k)^2 - \frac{1}{2}\log 2\pi\right). \tag{42}$$

Adding $p_q(\sigma)$, we obtain the complete data penalized log-likelihood function

$$pl_c(\gamma, \mu_1, \mu_2, \sigma) = l_c(\gamma, \mu_1, \mu_2, \sigma) + p_q(\sigma). \tag{43}$$

Algorithm 1 Penalized EM Algorithm

I-step: For $k = 1, 2$, get $\gamma^{(0)}$ from the k-means algorithm [42, chapter 9], then cluster $\mathbf{N}$, and compute $\mu_k^{(0)}$ and $\sigma_k^{(0)}$ as the maximum likelihood estimates of $\mathcal{N}(\mu_k, \sigma_k)$

$$\mu_k^{(0)} = \frac{1}{q_k}\sum_{i=1}^{q_k} N_i,\ \sigma_k^{(0)} = \sqrt{\frac{1}{q_k}\sum_{i=1}^{q_k}\left(N_i - \bar{N}_k\right)^2}, \quad (44)$$

where q_k is the number of elements of the population k. Set $\sigma^{(0)} = \sigma_k^{(0)}$ if k is the larger population.

E-step: Given $\mathbf{N}$ and the vector of parameters $\Theta^{(t)} = \left(\gamma^{(t)}, \mu_1^{(t)}, \mu_2^{(t)}, \sigma^{(t)}\right)$ at the t-th current iteration, compute

$$\gamma_{i1}^{(t+1)} = \frac{\left(1-\gamma^{(t)}\right) f\left(N_i; \mu_1^{(t)}, \sigma^{(t)}\right)}{f\left(N_i; \gamma^{(t)}, \mu_1^{(t)}, \mu_2^{(t)}, \sigma^{(t)}\right)}, \quad (45)$$

and

$$\gamma_{i2}^{(t+1)} = \frac{\gamma^{(t)} f\left(N_i; \mu_2^{(t)}, \sigma^{(t)}\right)}{f\left(N_i; \gamma^{(t)}, \mu_1^{(t)}, \mu_2^{(t)}, \sigma^{(t)}\right)}, \quad (46)$$

where $f\left(N_i; \gamma^{(t)}, \mu_1^{(t)}, \mu_2^{(t)}, \sigma^{(t)}\right)$ is given from (33).

M-step: Update the set of parameter $\Theta^{(t)}$ as follows.

1) Update $\gamma^{(t)}$ by

$$\gamma^{(t+1)} = \frac{\sum_{i=1}^{q}\gamma_{i2}^{(t+1)}}{q}. \quad (47)$$

2) Update $\mu_k^{(t)}, k \in \{1, 2\}$ by

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^{q}\gamma_{ik}^{(t+1)} N_i}{\sum_{i=1}^{q}\gamma_{ik}^{(t+1)}}. \quad (48)$$

3) Update $\sigma^{(t)}$ by

$$\sigma^{(t+1)} = \sqrt{\frac{S^{(t+1)} + 2a_q s_q^2}{q + 2a_q}}. \quad (49)$$

where

$$S^{(t+1)} = \sum_{i=1}^{q}\sum_{k=1}^{2}\gamma_{ik}^{(t+1)}\left(N_i - \mu_k^{(t+1)}\right)^2. \quad (50)$$

S-step: Stop the algorithm whenever t exceeds a limited iteration number, or the norm $\left\|\Theta^{(t+1)} - \Theta^{(t)}\right\|$ is smaller a threshold. Set $\left(\hat{\gamma}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}\right)$ as the final $\Theta^{(t+1)}$, and stock the final $\gamma_{i1}^{(t+1)}$ and $\gamma_{i2}^{(t+1)}$ for feature statistics clustering.

We find that $pl_c(\gamma, \mu_1, \mu_2, \sigma)$ can be trivially maximized in closed form. Unfortunately, we do not have values for the latent variables $\mathbf{Z}$, thus we cannot use directly $pl_c(\gamma, \mu_1, \mu_2, \sigma)$. Instead, we consider firstly its expected value under the posterior distribution of the latent variables (i.e., E-step of the EM algorithm). Next, we maximize this expectation (i.e., M-step of the EM algorithm). Such a procedure can be represented by Algorithm 1.

C. AUTHENTICATION AND LOCALIZATION

Following the above analyses, we can model the digital images authentication as a decision problem between two hypotheses

$$\begin{cases}\mathcal{H}_0 : N_i \sim \mathcal{N}(0, 1) & \text{authentic image,}\\ \mathcal{H}_1 : N_i \nsim \mathcal{N}(0, 1) & \text{no conclusion.}\end{cases} \quad (51)$$

In reality, dazzling areas or missing color ranges existing on digital images may distort the distribution of N_i. Therefore, the distribution of N_i does not always have a perfect Gaussian form even for authentic images. To partially weaken the distortion impacts, we propose regenerating random samples $\hat{N}_i$ from the model (33) using the estimated parameters obtained by Algorithm 1 (see e.g., [43, page 53] for a simulation algorithm). Then, we work with $\hat{N}_i$ instead of N_i. As such, the problem (51) becomes

$$\begin{cases}\mathcal{H}_0 : \hat{N}_i \sim \mathcal{N}(0, 1) & \text{authentic image,}\\ \mathcal{H}_1 : \hat{N}_i \nsim \mathcal{N}(0, 1) & \text{no conclusion.}\end{cases} \quad (52)$$

Normality tests are next carried out to decide if the tested image is authentic or not. In this paper, well-known normality tests [44] (i.e., Anderson-Darling test, one-sample Kolmogorov-Smirnov test, Jarque-Bera test and Lilliefors test) are jointly used to achieve a reliable decision. If all these tests return $\mathcal{H}_0$, the image is decided to be authentic automatically. Additionally, the authentication can be done by human interpretation thanks to graphical tools such Q-Q plot, probability distribution curves, and localization map.

Regarding the forgery localization, we can decide the belonging of N_i to $\mathcal{P}_1$ and $\mathcal{P}_2$ from the set $\{\gamma_{i1}\}_{i=1,\ldots,q}$ and $\{\gamma_{i2}\}_{i=1,\ldots,q}$ (see the output of Algorithm 1) as follow

$$N_i \in \begin{cases}\mathcal{P}_1 : \text{ authentic portion} & \text{if } \gamma_{i1} > \gamma_{i2},\\ \mathcal{P}_2 : \text{ tampered portion} & \text{if } \gamma_{i1} \le \gamma_{i2}.\end{cases}, \quad (53)$$

Since each N_i, $i = 1, \ldots, q$, corresponds to the i-th $B \times B$ block of the tested image, we can therefore derive a binary map indicating tampered region of the image.

For an illustration, we show in Fig. 4 the authentication and localization results for an original image and a tampered image respectively. For the original image, the distribution of feature statistics is standard normal. Accordingly, with a type 1 error $\alpha = 0.05$, all the considered normality tests return $\mathcal{H}_0$, while the points in Q-Q plot diagram follow a linear pattern. This confirms the authenticity of the image. Looking at the localization map, black and white blocks are equally mixed overall the image because $\gamma = 0.49$, and no concrete form is appeared. For the tampered image, the distribution of feature statistics is no longer standard normal. All the normality tests return then $\mathcal{H}_1$, and the points in Q-Q plot diagram follow a strongly nonlinear pattern. The outliers of dash red line in Q-Q plot diagram correspond to the smaller population in the mixture pdf of feature statistics. The location map now shows

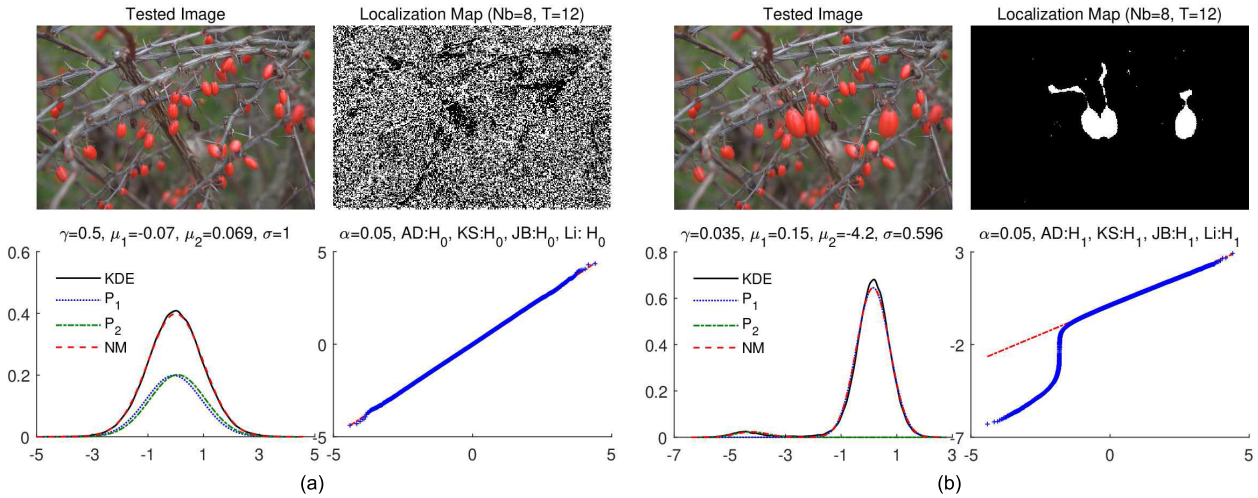

FIGURE 4. Authentication and localization results for digital images given from Image Manipulation Dataset [32]. (a) Original image. (b) Tampered image.

clearly the tampered regions in white and authentic region in black.

V. NUMERICAL EXPERIMENTS

Four well-known datasets: Image Manipulation [32], MICC-F600 [45], Realistic Tampering [46], and CUISDE [47] are used for numerical experiments. Their detailed description can be found in [4]. Through these datasets, we aim at evaluating the images authentication and forgeries localization ability of the proposed algorithm.

A. EVALUATION CRITERIA

Due to highly imbalanced datasets[1], Precision (P), Recall (R) and F_1-Score (F_1) are chosen as criteria for performance evaluation [48]. Precision and Recall are computed from the confusion matrix of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) as

$$P = \frac{TP}{TP + FP}, \quad \text{and} \quad R = \frac{TP}{TP + FN}. \tag{54}$$

High Precision implies a high probability that detected results are relevant, while high Recall means a high probability that relevant results are detected. Obviously, using separately either Precision or Recall is not enough to evaluate the performance of a algorithm. F_1-Score, which is the harmonic mean of Precision and Recall, takes a high value when Precision and Recall are both important

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP}. \tag{55}$$

So, it might be a better measure for performance evaluation. The higher F_1-Score, the more the algorithm is efficient. Moreover, depending on the authentication or localization goals, the above measures has their own meanings (see Table 2 for the details).

[1]Datasets here refer to authentic and tampered images in the case of image authentication, and to untampered and tampered pixels in the case of forgery localization

TABLE 2. Meaning of measures.

	Images authentication (image level)
TP	number of authentic images correctly detected
FN	number of undetected authentic images
FP	number of tampered images wrongly detected
TN	number of undetected tampered images
P	probability that detected images are authentic
R	probability that authentic images are detected

	Forgery localization (pixel level)
TP	number of tampered pixels correctly localized
FN	number of unlocalized tampered pixels
FP	number of authentic pixels wrongly localized
TN	number of unlocalized authentic pixels
P	probability that localized pixels are tampered
R	probability that tampered pixels are localized

B. BENCHMARK ALGORITHMS

The performance and robustness of the developed algorithm are assessed though comparative studies with algorithms proposed by Dirik and Menon in [25], and by Ferrara *et al.* in [26]. These two benchmarks are threshold-based algorithms, so their performance depend closely on a threshold τ used to distinguish between tampered and authentic regions in an image. Meanwhile, the performance of the developed algorithm is dependent on the threshold T used to determine the smooth region $\mathbf{R}_3$. Besides, the blocks size B is influential in the performance of all the algorithms. So, sensitivity studies to these factors are necessary. We also note that the two benchmark algorithms focuses on images forgery localization rather than on images authentication.

C. AUTHENTICATION PERFORMANCE

To assess the authentication performance of the proposed algorithm, we first randomly choose a set of 150 authentic images and 200 tampered images from the 4 above datasets. Next, we apply the algorithm to obtain TP, FN, FP and TN, and thence compute the measures P, R and F_1. Such a process is applied to the 2 following case studies.

- *Case study 1*: B takes respectively the value 2, 4, 8, 16 and 32, T is fixed at 20.
- *Case study 2*: B is fixed at 16, T varies from 10 to 30 with step 5.

Tables 3 and 4 show the results when the authentication is done automatically and is interpreted by human support.

TABLE 3. Results of automatic authentication.

	Case study 1 ($T = 20$)				
B	2	4	8	16	32
P (%)	46.02	42.59	37.50	45.59	57.63
R (%)	34.67	30.67	22.00	20.67	22.67
F_1 (%)	39.54	35.66	27.73	28.44	32.54

	Case study 2 ($B = 16$)				
T	10	15	20	25	30
P (%)	45.33	43.94	45.59	43.42	39.39
R (%)	22.67	19.33	20.67	22.00	17.33
F_1 (%)	30.22	26.85	28.44	29.20	24.07

TABLE 4. Authentication results given by human interpretation.

	Case study 1 ($T = 20$)				
B	2	4	8	16	32
P (%)	59.82	59.45	64.67	61.59	63.64
R (%)	87.33	86.00	79.33	62.00	46.67
F_1 (%)	71.00	70.30	71.26	61.79	53.85

	Case study 2 ($B = 16$)				
T	10	15	20	25	30
P (%)	63.80	64.15	61.59	62.58	59.73
R (%)	69.33	68.00	62.00	64.67	59.33
F_1 (%)	66.45	66.02	61.79	63.61	59.53

Clearly, the performance of the automatic authentication is relatively weak. Indeed, despite working on the smooth region $\mathbf{R}_3$ instead of on the entire green channel $\mathbf{G}$, and using $\hat{N}_i$ rather than N_i, shape edges and/or strong dazzling areas existing on realistically authentic digital images still distort the standard Gaussian form of $\hat{N}_i$ pdf. As a results, the automatic authentication via normality tests is less efficient. However, the performance is much more improved thanks to human interpretation of graphic tools (i.e., pdf curves, Q-Q plot, and localization map). To see how the human interpretation can help, we introduce an example as in Fig. 5. Since not all the considered normality tests (i.e., three among four) result in $\mathcal{H}_0$, the automatic authentication

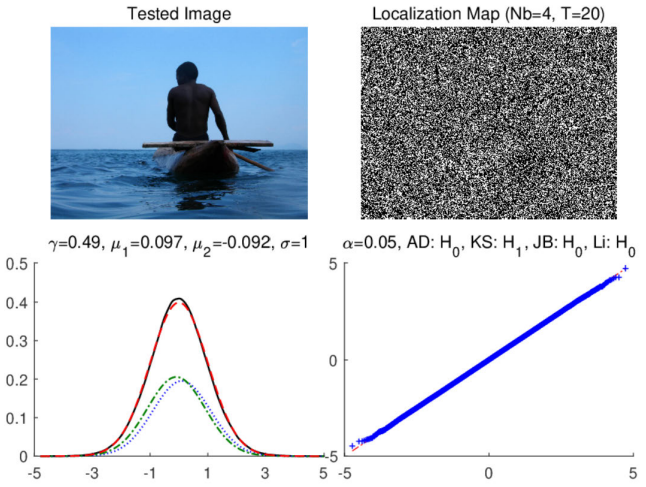

FIGURE 5. Example of image authentication by human interpretation (original image given from Image Manipulation Dataset [32]).

returns "no conclusion" for the tested image. Nevertheless, looking at the pdf curves and the Q-Q plot, we find that the Gaussian assumption of $\hat{N}_i$ is almost satisfied. Moreover, there is no concrete shape found in the localization map. All these elements confirm the credibility of image content. Thus, the human interpretation decides that the image is authentic, which is actually true.

Observing the values of F_1-Score in Table 3 and 4, we also find the importance of B and T. The increasing of blocks size B leads to smaller samples set in constructing pdf of $\hat{N}_i$. Meanwhile, by setting T at a high value, more edges are allowed in the computation of N_i. So, it is not surprising that the authentication performance decreases in these configurations of T and B.

D. LOCALIZATION PERFORMANCE

Regarding the localization, the aim is to find out (*i*) which kinds of forgeries could be localized by the proposed algorithm, and (*ii*) how good is the proposed algorithm compared to algorithms of the same kind. Tampered images in Image Manipulation Dataset [32] and Realistic Tampering Dataset [46] are used in numerical experiments because of various kinds of forgeries therein. Moreover, binary ground truths are also provided.

1) FIRST LOCALIZATION ISSUE

To reply to the first issue, we use the proposed algorithm (with $T = 20$ and blocks size $B = 8$) to derive location maps, and thence compare them with the associated ground truths. We also apply the algorithms proposed by Dirik and Menon [25], and by Ferrara *et al.* [26] to obtain benchmark location maps. As illustrated in Fig. 6, various kinds of forgeries are successfully localized by our algorithm.

Clearly, localization maps returned by our algorithm and Ferrara *et al.*'s algorithm are closer to ground truths, and more confident than Dirik and Menon's algorithm. When zooming in the localized areas, our algorithm allows better resolution

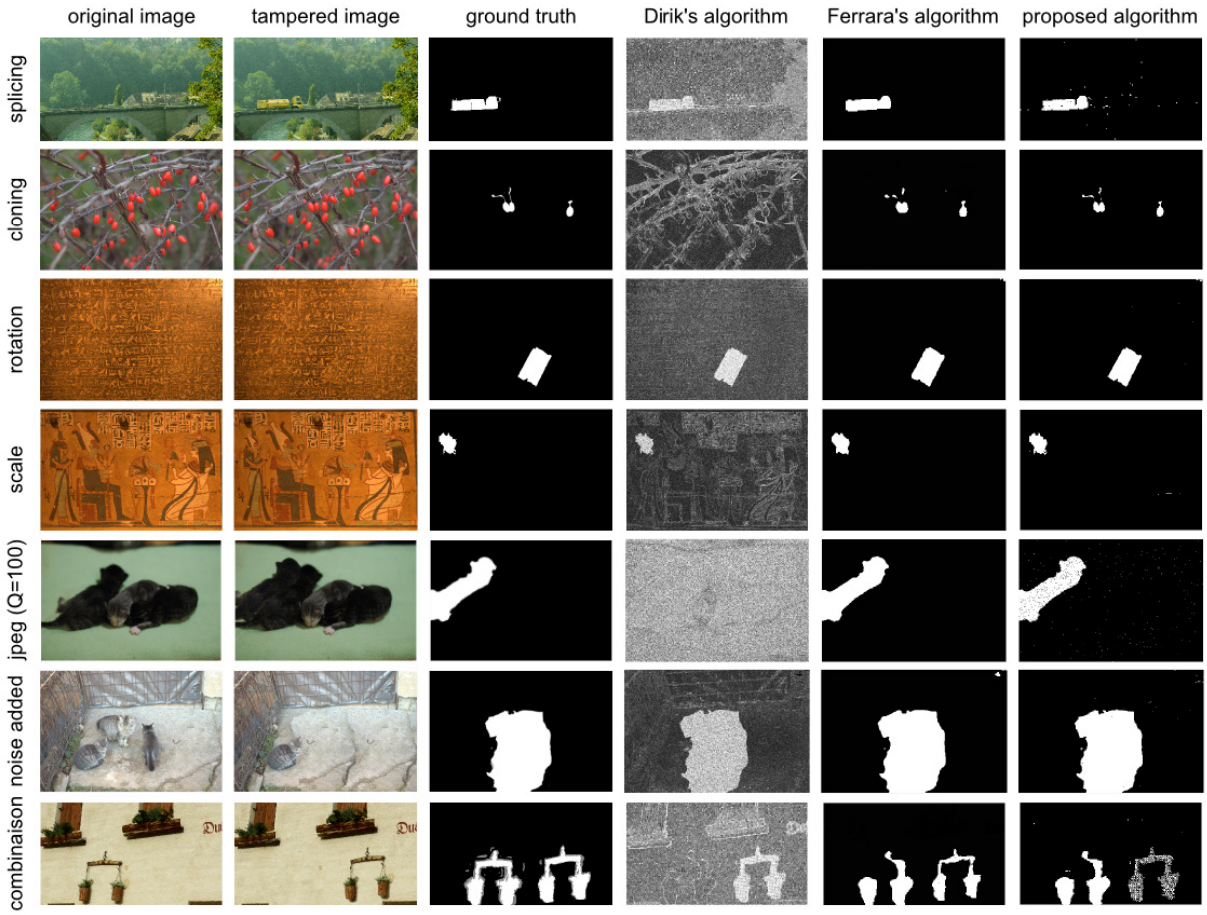

FIGURE 6. Examples of successful forgery localization by the proposed algorithm.

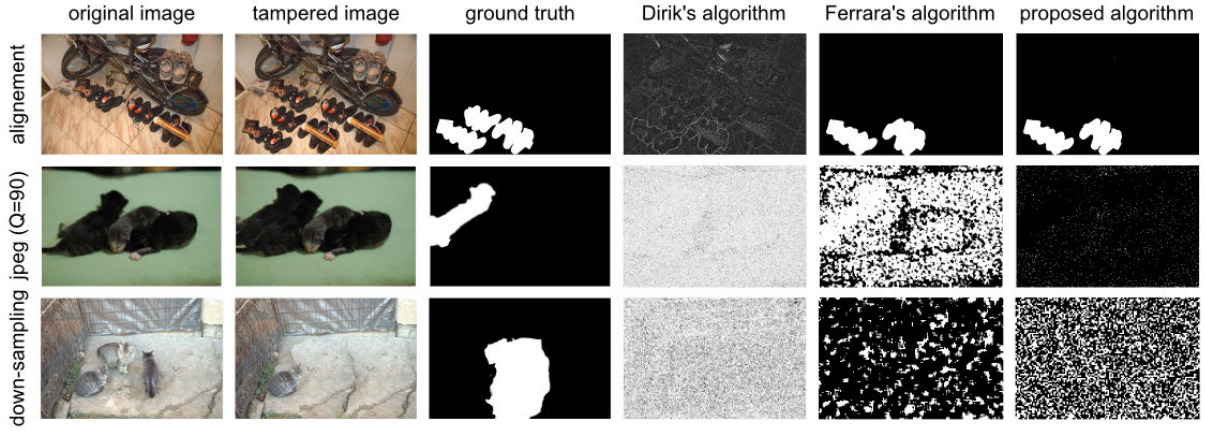

FIGURE 7. Examples of unsuccessful forgery localization by the proposed algorithm.

and higher fidelity. We will understand more deeply this issue through a quantitative study in Subsection V-D.2. Moreover, the use of Ferrara *et al.*'s algorithm is more complicated in practice because a subjective threshold is required to distinguish tampered and authentic region. Fig. 7 shows some configurations that forgeries are not successfully localized. For cloning forgery (first line of Fig. 7), a part of tampered region is missing because it have the same alignment of CFA pattern as the original image. Especially, the localization is completely failed (see second line and third line of Fig. 7), if tampered images undergo additional post-processing (e.g., JPEG compression with low quality, down-sampling, etc.).

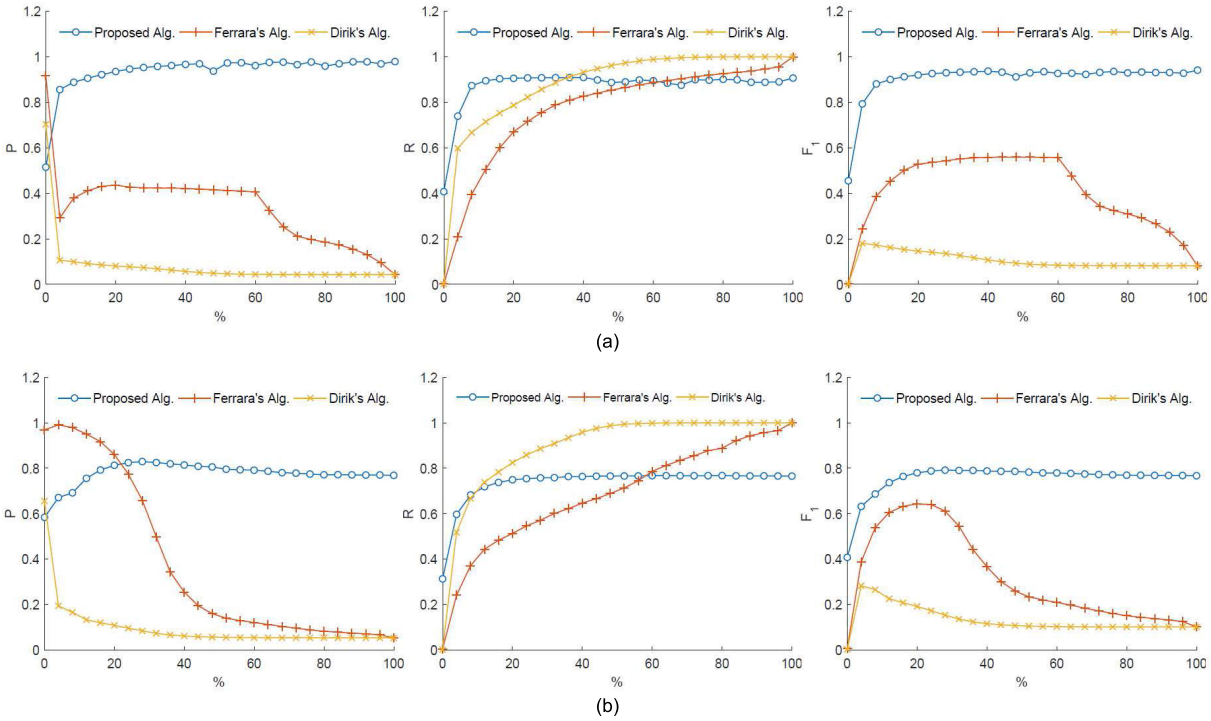

FIGURE 8. Precision, Recall and F_1-Score for the case study 1. (a) Image Manipulation Dataset [32]. (b) Realistic Tampering Dataset [46].

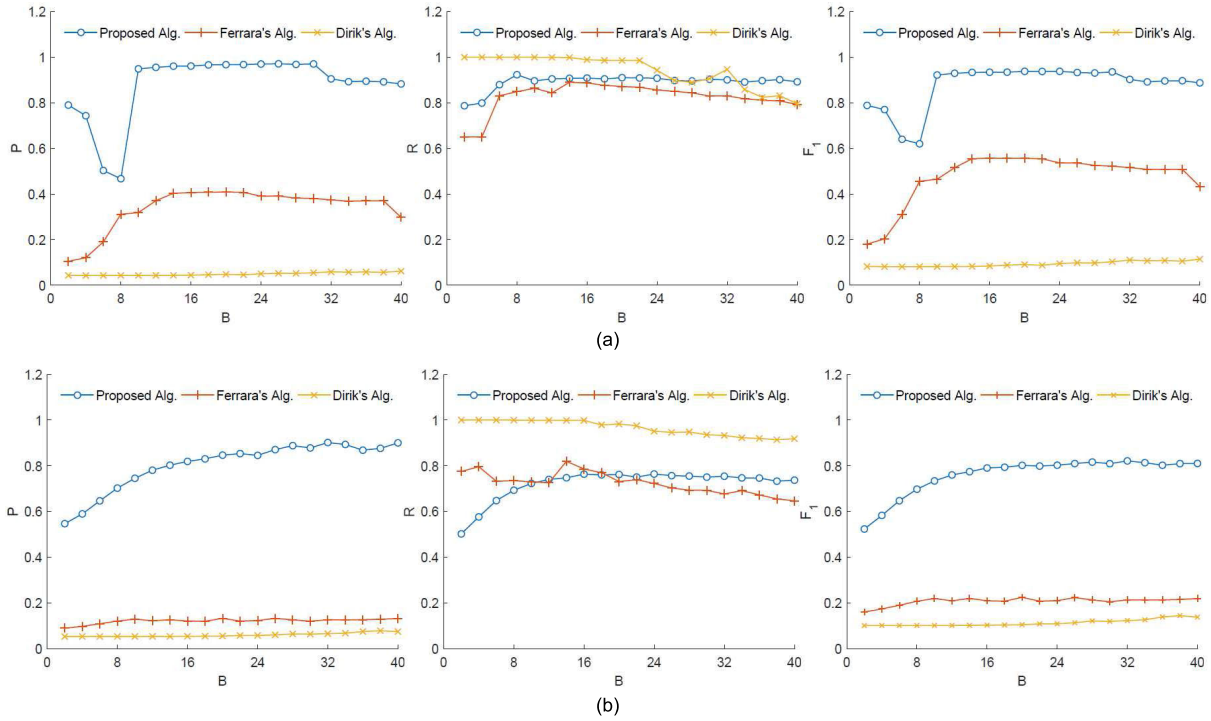

FIGURE 9. Precision, Recall and F_1-Score for the case study 2. (a) Image Manipulation Dataset [32]. (b) Realistic Tampering Dataset [46].

2) SECOND LOCALIZATION ISSUE

To reply to the second issue, we first apply the 3 considered algorithm to compute localization maps for two image sets associated with Image Manipulation Dataset [32] and the Realistic Tampering Dataset [46] respectively. Each set consists of 20 random tampered images. From the localization

maps and corresponding ground truths, we can derive TP, FN, FP, TN, and therefore compute P, R and F_1.

A comparative study on F_1-Scores allows to evaluate the performance of the proposed algorithm, as well as of the two benchmarks. Similarly to Section V-C, two case studies are considered.

- *Case study 1*: B is fixed at 16, the values of τ and T vary. The threshold T in our algorithm takes value in the interval [2; 52], while the threshold τ in the benchmark algorithms takes value from the minimum to the maximum of variance maps. Note that, different variance maps of different images yield different ranges of values for τ. For comparison purpose, the thresholds value is further translated into percentage of associated intervals. Note that 0% and 100% correspond respectively to the minimum and maximum of each interval.
- *Case study 2*: The thresholds τ and T are fixed at 60% of ranges of values, B varies from 2 to 40 with step 2.

Fig. 8 and 9 show respectively the results of these two case studies. In all cases, the F_1-Score of the proposed algorithm is more stable at higher value than the benchmark algorithms. This implies that our algorithm is more efficient and robuster. Looking at the diagrams of P and R, we find that the high F_1-Score of our algorithm mostly comes from the high value of P, because the values of R are more or less similar in the three algorithms. This results also mean that our algorithm allows a high probability that localized pixels are tampered, while the benchmark algorithms do not. This is totally consistent with the nature of non-threshold-based and threshold-based localization algorithms.

VI. CONCLUSION AND PERSPECTIVES

We develop in this paper an improved algorithm for digital image authentication and forgery localization by jointly use the color filter array pattern identification, demosaicing algorithm estimation, and the local statistical analysis of demosaicing artifacts in spatial domain. A new feature statistic less sensitive to the edges problem is thus built to characterize demosaiced images. By modeling such feature statistics by a single normal mixture model for both tampered and untampered images, four well-known normality tests (i.e., Anderson-Darling test, one-sample Kolmogorov-Smirnov test, Jarque-Bera test and Lilliefors test) are employed to automatically authenticate digital images. Numerical experiments on the four well-known datasets (i.e. Image Manipulation, MICC-F600, Realistic Tampering, and CUISDE) shows that the performance of automatic authentication is relatively low, but can be much more improved thanks to human interpretation of supported graphic tools (i.e., Q-Q plot diagram, probability distribution curves, and localization map). Regarding the forgery localization, we propose a penalized EM algorithm to automatically distinguish between authentic and forged regions of a tampered image without any requirement on comparison thresholds as in most existing localization algorithm. Such a method is proved to be more effective and robuster by numerical examples.

Even if the developed algorithm yields very encouraging results, we find that the automatic authentication of digital images is still at low performance (see Table 3), and that the forgery localization is limited to uncompressed or less-compressed images (see Fig. 6 and 7). The focus of our near future work is to further improve these two points. Building a new feature statistic taking into account the periodicity of demosaicing artifacts in DCT domain as in [24] seem to be a key step for this work. Another perspective is to extend the developed methods to the videos forensics. The work of Singh and Aggarwal in [49] could be a good orientation.

REFERENCES

[1] J. A. Redi, W. Taktak, and J.-L. Dugelay, "Digital image forensics: A booklet for beginners," *Multimedia Tools Appl.*, vol. 51, no. 1, pp. 133–162, Jan. 2011.
[2] A. Piva, "An overview on image forensics," *ISRN Signal Process.*, vol. 2013, Nov. 2013, Art. no. 496701.
[3] M. A. Qureshi and M. Deriche, "A bibliography of pixel-based blind image forgery detection techniques," *Signal Process., Image Commun.*, vol. 39, pp. 46–74, Nov. 2015.
[4] P. Korus, "Digital image integrity—A survey of protection and verification techniques," *Digit. Signal Process.*, vol. 71, pp. 1–26, Dec. 2017.
[5] S. Teerakanok and T. Uehara, "Copy-move forgery detection: A state-of-the-art technical review and analysis," *IEEE Access*, vol. 7, pp. 40550–40568, 2019.
[6] H. Farid, "Image forgery detection," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 16–25, Mar. 2009.
[7] X. Lin, J.-H. Li, S.-L. Wang, A.-W.-C. Liew, F. Cheng, and X.-S. Huang, "Recent advances in passive digital image security forensics: A brief review," *Engineering*, vol. 4, pp. 29–39, Feb. 2018.
[8] M. Kirchner, "Efficient estimation of CFA pattern configuration in digital camera images," *Proc. SPIE*, vol. 7541, Jan. 2010, Art. no. 754111.
[9] C.-H. Choi, J.-H. Choi, and H.-K. Lee, "CFA pattern identification of digital cameras using intermediate value counting," in *Proc. 13th ACM Multimedia Workshop Multimedia Secur.*, 2011, pp. 21–26.
[10] J. Takamatsu, Y. Matsushita, T. Ogasawara, and K. Ikeuchi, "Estimating demosaicing algorithms using image noise variance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 279–286.
[11] J. J. Jeon, H. J. Shin, and I. K. Eom, "Estimation of Bayer CFA pattern configuration based on singular value decomposition," *EURASIP J. Image Video Process.*, vol. 47, p. 47, Dec. 2017.
[12] H. J. Shin, J. J. Jeon, and I. K. Eom, "Color filter array pattern identification using variance of color difference image," *J. Electron. Imag.*, vol. 26, no. 4, 2017, Art. no. 043015.
[13] A. C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3948–3959, Oct. 2005.
[14] S. Bayram, H. T. Sencar, and N. Memon, "Classification of digital camera-models based on demosaicing artifacts," *Digit. Invest.*, vol. 5, nos. 1–2, pp. 49–59, 2008.
[15] A. C. Gallagher, "Detection of linear and cubic interpolation in JPEG compressed images," in *Proc. 2nd Can. Conf. Comput. Robot Vis.*, May 2005, pp. 65–72.
[16] A. Swaminathan, M. Wu, and K. J. R. Liu, "Nonintrusive component forensics of visual sensors using output images," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 1, pp. 91–106, Mar. 2007.
[17] H. Cao and A. C. Kot, "Accurate detection of demosaicing regularity for digital image forensics," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 4, pp. 899–910, Dec. 2009.
[18] A. Swaminathan, M. Wu, and K. J. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 101–117, Mar. 2008.
[19] 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, Jun. 2012.
[20] C.-H. Choi, H.-Y. Lee, and H.-K. Lee, "Estimation of color modification in digital images by CFA pattern change," *Forensic Sci. Int.*, vol. 226, pp. 94–105, Mar. 2013.

[21] S. Gao, G. Xu, and R.-M. Hu, "Camera model identification based on the characteristic of CFA and interpolation," in *Proc. Int. Workshop Digit. Watermarking*, 2011, pp. 268–280.
[22] D. Menon and G. Calvagno, "Color image demosaicking: An overview," *Signal Process., Image Commun.*, vol. 26, nos. 8–9, pp. 518–533, Oct. 2011.
[23] A. C. Gallagher and T. Chen, "Image authentication by detecting traces of demosaicing," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2008, pp. 1–8.
[24] L. Li, J. Xue, X. Wang, and L. Tian, "A robust approach to detect digital forgeries by exploring correlation patterns," *Pattern Anal. Appl.*, vol. 18, no. 2, pp. 351–365, 2015.
[25] A. E. Dirik and N. Memon, "Image tamper detection based on demosaicing artifacts," in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 1497–1500.
[26] P. Ferrara, T. Bianchi, A. D. Rosa, and A. Piva, "Image forgery localization via fine-grained analysis of CFA artifacts," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1566–1577, Oct. 2012.
[27] E. G. Fernández, A. L. S. Orozco, L. J. G. Villalba, and J. Hernandez-Castro, "Digital image tamper detection technique based on spectrum analysis of CFA artifacts," *Sensors*, vol. 18, no. 9, p. 2804, 2018.
[28] A. Singh, G. Singh, and K. Singh, "A Markov based image forgery detection approach by analyzing CFA artifacts," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 28949–28968, 2018.
[29] T. H. Thai, F. Retraint, and R. Cogranne, "Generalized signal-dependent noise model and parameter estimation for natural images," *Signal Process.*, vol. 114, pp. 164–170, Sep. 2015.
[30] B. C. Smith and L. A. Rowe, "Algorithms for manipulating compressed images," *IEEE Comput. Graph. Appl.*, vol. 13, no. 5, pp. 34–42, Sep. 1993.
[31] P. J. Burt, "Fast algorithms for estimating local image properties," *Comput. Vis., Graph., Image Process.*, vol. 21, no. 3, pp. 368–382, 1983.
[32] 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, Dec. 2012.
[33] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Ann. Statist.*, vol. 38, no. 5, pp. 2916–2957, 2010.
[34] Y. S. Qin and B. Smith, "Likelihood ratio test for homogeneity in normal mixtures in the presence of a structural parameter," *Stat. Sinica*, vol. 14, no. 4, pp. 1165–1177, 2004.
[35] J. Chen, X. Tan, and R. Zhang, "Inference for normal mixtures in mean and variance," *Stat. Sinica*, vol. 18, no. 2, pp. 443–465, 2008.
[36] J. Kiefer and J. Wolfowitz, "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters," *Ann. Math. Statist.*, vol. 27, no. 4, pp. 887–906, 1956.
[37] N. E. Day, "Estimating the components of a mixture of normal distributions," *Biometrika*, vol. 56, no. 3, pp. 463–474, 1969.
[38] G. Ciuperca, A. Ridolfi, and J. Idier, "Penalized maximum likelihood estimator for normal mixtures," *Scand. J. Statist.*, vol. 30, no. 1, pp. 45–59, 2003.
[39] J. Chen, P. Li, and Y. Fu, "Inference on the order of a normal mixture," *J. Amer. Stat. Assoc.*, vol. 107, no. 499, pp. 1096–1105, 2012.
[40] L. Jin, W. Xu, L. Zhu, and L. Zhu, "Penalized maximum likelihood estimator for skew normal mixtures," 2016, *arXiv:1608.01513*. [Online]. Available: https://arxiv.org/abs/1608.01513
[41] J. Chen, "Consistency of the MLE under mixture models," *Stat. Sci.*, vol. 32, no. 1, pp. 47–63, 2017.
[42] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
[43] D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of Monte Carlo Methods* (Wiley Series in Probability and Statistics), vol. 706. Hoboken, NJ, USA: Wiley, 2013.
[44] B. W. Yap and C. H. Sim, "Comparisons of various types of normality tests," *J. Statist. Comput. Simul.*, vol. 81, no. 12, pp. 2141–2155, Dec. 2011.
[45] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, L. Del Tongo, and G. Serra, "Copy-move forgery detection and localization by means of robust clustering with J-Linkage," *Signal Process., Image Commun.*, vol. 28, no. 6, pp. 659–669, 2013.
[46] P. Korus and J. Huang, "Multi-scale analysis strategies in PRNU-based tampering localization," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 4, pp. 809–824, Apr. 2016.
[47] Y.-F. Hsu and S.-F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 549–552.
[48] O. M. Al-Qershi and B. E. Khoo, "Evaluation of copy-move forgery detection: Datasets and evaluation metrics," *Multimedia Tools Appl.*, vol. 77, no. 24, pp. 31807–31833, 2018.
[49] R. D. Singh and N. Aggarwal, "Detection and localization of copy-paste forgeries in digital videos," *Forensic Sci. Int.*, vol. 281, pp. 75–91, Dec. 2017.

NHAN LE received the Engineering degree in industrial engineering and management from the Danang University of Science and Technology, Vietnam, in 2016, and the M.S. degree in systems optimization and security from the Troyes University of Technology, France, in 2018, where she is currently pursuing the Ph.D. degree in system optimization and dependability with the Laboratory of System Modeling and Dependability (LM2S). Her research interests include statistical image processing, digital image forensics, and hypothesis testing theory.

FLORENT RETRAINT received the Engineering Diploma degree in computer science from the University of Technology of Compiegne, in 1993, the M.S. degree in applied mathematics from the National School of Computer Science and Applied Mathematics of Grenoble (ENSIMAG), in 1994, and the Ph.D. degree in applied mathematics from the National Institute of Applied Sciences of Lyon, France, in 1998. He held a postdoctoral position at CEA Grenoble, for one year. He was a Research Engineer with Thomson CSF, for two years. Since 2002, he has been with the Laboratory of System Modeling and Dependability (LM2S), Troyes University of Technology, where he is currently a Full Professor. His research interests include image modeling, statistical image processing, hypothesis testing theory, anomaly detection, and localization.