

Detecting Image Splicing Based on Noise Level Inconsistency

Heng Yao¹ · Shuozhong Wang² · Xinpeng Zhang² ·
Chuan Qin¹ · Jinwei Wang³

Received: 24 November 2015 / Revised: 12 April 2016 / Accepted: 1 June 2016 /
Published online: 11 June 2016
© Springer Science+Business Media New York 2016

Abstract In a spliced image, areas from different origins contain different noise features, which may be exploited as evidence for forgery detection. In this paper, we propose a noise level evaluation method for digital photos, and use the method to detect image splicing. Unlike most noise-based forensic techniques in which an AWGN model is assumed, the noise distribution used in the present work is intensity-dependent. This model can be described with a noise level function (NLF) that better fits the actual noise characteristics. NLF reveals variation in the standard deviation of noise with respect to image intensity. In contrast to denoising problems, noise in forensic applications is generally weak and content-related, and estimation of noise characteristics must be done in small areas. By exploring the relationship between NLF and the camera response function (CRF), we fit the NLF curve under the CRF constraints. We then formulate a Bayesian maximum *a posteriori* (MAP) framework to optimize the NLF estimation, and develop a method for image splicing detection according to noise level inconsistency in image blocks taking from different origins. Experimental results are presented to show effectiveness of the proposed method.

Keywords Digital image forensics · Image splicing detection · Noise level function (NLF) · Bayesian maximum *a posteriori* (MAP)

✉ Heng Yao
hyao@usst.edu.cn

¹ Shanghai Key Lab of Modern Optical System, and Engineering Research Center of Optical Instrument and System, Ministry of Education, University of Shanghai for Science and Technology, Shanghai 200093, China

² School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

³ School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

1 Introduction

Digital image forensics is a technique of photo source identification and integrity authentication. This paper considers the problem of detecting image splicing without any prior knowledge. Existing methods of image splicing detection generally fall into two categories: 1) methods based on image fingerprints such as re-sampling traces [23], double JPEG compression [5], contrast enhancement [24], median filtering [31], etc., and 2) methods based on constraints of certain image attributes such as copy-move [10], shadow matte [14], planar homography [32] and perspective features [30]. The noise feature as an intrinsic fingerprint of a photo provides a useful clue for image forensics. The method proposed in this paper belongs to the second category.

1.1 Related Works

Every image acquired by a camera contains noise that can be classified into two types: photo-response non-uniformity noise (PRNU) and random noise. PRNU is caused by slight errors in sensitivity of the camera's photoelectrical sensors. It has a roughly fixed pattern among pictures taken by the same model of camera. Since this type of non-uniformity is inherent in the camera, it can be used as a fingerprint for camera identification. Lukas et al. [16] developed a method to verify the pattern of noise distribution. They determined the reference noise pattern of a camera by averaging noise extracted from several images. Given an image, they obtained the pattern noise from the image using a smoothing filter, and identified the camera model by comparing with candidate reference patterns. The same approach was also used to develop a method of photo integrity detection [1]. A major limitation of the PRNU-based methods is that it relies on prior knowledge about the camera model, and requires a large number of training images taken by the registered equipment. Furthermore, if a spliced image is composed of two origins taken by the same camera, the method will fail.

Another type of noise, i.e., random noise, include quantum effects, thermal fluctuations and dark current leakage. Although hardware manufacturers have been trying hard to suppress random noise, noise is inevitable. Estimating the noise level from a single image is a significant task in many applications, in particular, image denoising. Most of earlier approaches assumed the additive white Gaussian noise (AWGN) model [4], and described the noise level with its variance or standard deviation. However, the AWGN conjecture may not hold for real-life digital photographs because the actual CMOS/CCD sensor noise is strongly dependent on the light intensity. Based on this consideration, Liu et al. [12] defined a noise level function (NLF) with respect to image intensity. They collected a sample set representing spatial average and variation, and found the lower envelope of the samples. A Bayesian framework was used to optimize the fitting. The method was later extended to a noise removal application [13]. However, the shape of the NLF curve was not well constrained in the Bayesian process. Thus, the incomplete homogenous samples may affect the accuracy of estimation. Yang et al. [27] concerned the problem of insufficient homogenous sample blocks in noise level estimation. They estimated NLF using sparse representation techniques via dictionary recovery. To improve the fitting approximation to the ground-truth functions, a sparse recovery model extracted in discrete cosine domain was further developed [28]. However, a reliable dictionary needed to be trained before estimation. Liu et al. [15] developed a method to estimate signal dependent noise parameters based on the mean and variance of the selected weak textured patches. In their method, the number of noise parameters was set to 3,

and the influence of camera response function was neglected. In the preliminary version of our work [29], the strong relationship between NLF and the camera response function was revealed and NLF can be estimated from relative incomplete samples. For noise inconsistency forensic applications, Mahdian et al. [18] proposed a noise variance estimator based on the median absolute deviation (MAD) in the high-pass Haar wavelet subband, and applied it into image splicing detections. Based on the observation of projection kurtosis concentration in band-pass domain [22], Lyu et al. [17] proposed a blind noise estimation approach to evaluate the noise level using projection kurtosis, and designed a splicing detection method based on blind local variance estimation. However, the blind noise estimation method may not be suitable to relatively complex noise models, and for different intensities in an untampered photo, the noise variance may not be close enough. Kobayashi et al. [9] proposed a stationary-scene video forgery detection method based on inconsistency in NLF. They identified inserted pixels from another video source by using maximum *a posteriori* (MAP) estimation for the noise model when NLF of the region was inconsistent with the rest of the video. However, this method cannot be applied to single image forgery detection since NLF was collected by averaging static-scene video frames in [9], rather than estimated from a single image.

1.2 Our contributions

As mentioned in the above, the method proposed in [13] was applicable to suppressing strong noise in images. In image forensic applications, however, the goal is to evaluate the noise level, no matter high or normal, of the whole images or just in local image blocks. Actually the forged region may be small, that is, the collected sample set available for estimation may be incomplete. In this paper, we consider noise level evaluation from images with moderate noise level using incomplete sample sets. To do so, we first model the noise level based on the relationship between NLF and the corresponding camera response function (CRF), and divide the image into edge and non-edge regions. Since samples taken from non-edge regions are usually inadequate, the NLF profile is constrained by the shape of CRF estimated from the edge regions. A Bayesian MAP framework is used to estimate NLF. The proposed approach of NLF estimation is then applied to detect image splicing. The idea is that, if an image region comes from a different source, there is likely a trace of noise level inconsistency due to distinguishable source camera models or different camera settings.

The rest of the paper is organized as follows. Section 2 discusses the relationship between CRF and NLF. Sections 3 and 4 present the proposed methods for NLF estimation and splicing detection, respectively. Section 5 shows effectiveness of the method, and Section 6 concludes the paper.

2 Noise Level Function Modeling

In this section, we explore the relationship between CRF and the corresponding NLF. An NLF model is then established, which will be used to estimate NLF in Section 3.

2.1 Camera response function

In many image processing algorithms, the recorded image intensity is assumed to be proportional to the scene radiance recorded by the camera sensor. However it is

generally not the case. The camera response function (also termed the radiometric response function) $f(\cdot)$ is defined to describe the non-linear mapping between the scene radiance R and the measured intensity I in an image:

$$I = f(R) \quad (1)$$

Assuming f is continuous and monotonic and neglecting errors due to discretization, the inverse response function, $R = g(I)$, can be obtained. Since only the observed output intensity, I , is available, most CRF estimators attempt to find the inverse CRF g instead of f .

2.2 Noise level function

As stated in [26], noise produced in a digital camera is not simply additive, but strongly dependent on the image intensity. To describe the relationship between noise and image intensity, a noise level function (NLF) is defined as variation of the standard deviation of noise with respect to image intensity [13]. Figure 1 illustrates the imaging process that transforms radiance photons into intensity bits. Scene radiance passes through the lens and is contaminated by several kinds of noise prior to the transformation. Let R_0 , R and I denote noise-free scene radiance, noisy radiance and observed intensity respectively. Intensity of the noise-free image is denoted as I_0 , which cannot be obtained from a single image. There are four main types of noise [7]: photon shot noise N_{PS} , dark current noise N_{DC} , read-out noise N_{RO} , and quantization noise N_Q . Neglecting the interference of camera filter array (CFA) sampling and interpolation, the observed image intensity can be expressed as:

$$I = f(R) + N_Q = f(R_0 + N_{PS} + N_{DC} + N_{RO}) + N_Q \quad (2)$$

where $f(\cdot)$ is CRF. By Taylor expansion, the first-order approximation is [25]:

$$I \approx f(R_0) + f'(R_0) \cdot (N_{PS} + N_{DC} + N_{RO}) + N_Q \quad (3)$$

Assume that all noise sources are zero-mean and independent of each other, and their standard deviations are σ_{PS} , σ_{DC} , σ_{RO} and σ_Q respectively. Variance of R and I can be written as:

$$\sigma_R^2 = \sigma_{PS}^2 + \sigma_{DC}^2 + \sigma_{RO}^2 \quad (4)$$

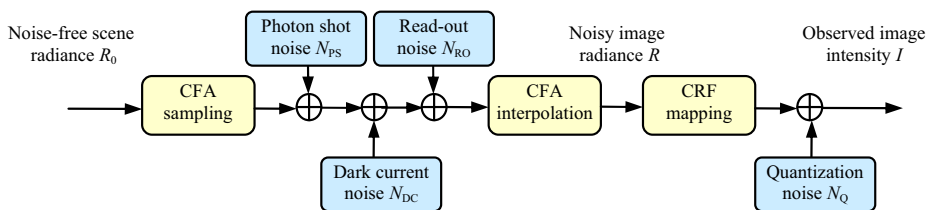


Fig. 1 Imaging process in which radiance photons are transformed into intensity bits. The signal captured by photoelectrical sensors is contaminated by shot noise, dark current noise and read-out noise in the process. After non-linear mapping, the image irradiance is quantized to produce intensity bits

$$\sigma_1^2 \approx f'^2(R_0) \cdot (\sigma_{PS}^2 + \sigma_{DC}^2 + \sigma_{RO}^2) + \sigma_Q^2 \quad (5)$$

$$\sigma_1^2 \approx f'^2(R_0) \cdot (R_0 \sigma_1^2 + \sigma_2^2) \quad (6)$$

where $\sigma_1^2 = \sigma_{PS}^2/R_0$ and $\sigma_2^2 = \sigma_{DC}^2 + \sigma_{RO}^2$. Since NLF is a function of the noise-free image intensity I_0 , and $R_0 = g(I_0)$, Equation (6) can be rewritten as:

$$\sigma_1(I_0) \approx \sqrt{g(I_0)\sigma_1^2 + \sigma_2^2/g'(I_0)} \quad (7)$$

Now we have revealed the relationship between NLF σ_1 and the inverse CRF g . Using (7), NLF can be found from g , σ_1 and σ_2 . The inverse CRF g depends on the camera model, and σ_1 and σ_2 are related to image capture parameters such as the sensitivity to light (i.e., ISO), shutter speed and aperture, and to the camera specifications such as the sensor noise level. We model NLF based on (7) for the following two reasons: 1) unlike the training-based models [12, 13, 27, 28], the parameters used here have physical meanings, and 2) estimation of NLF can well be constrained by the shape of CRF, especially under a low noise condition. Therefore NLF and CRF can be determined simultaneously.

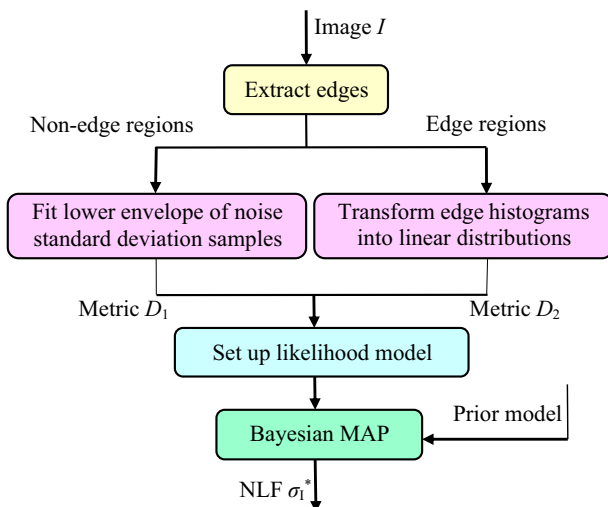
3 Estimation of Noise Level Function

As described in Section 2, NLF and CRF curves are strongly related. For estimating the noise level, we usually face two conditions: we have known the specific CRF or camera model before, or we do not have any prior knowledge of CRF. For the first condition, we can estimate the shape of NLF more accurately using (7). However, the second condition is more common, since it's usually hard to obtain the CRF data in advance. Thus, in this paper, we focus on how to estimate the most probable NLF and CRF simultaneously using Bayesian approach. According to [11, 13], NLF and CRF can be estimated from non-edge regions and edge regions, respectively. Both methods motivate us to develop a method to optimize NLF and CRF simultaneously. In the first step of the proposed method, the image is divided into non-edge and edge regions using edge extraction techniques. For non-edge regions, execute mean-shift segmentation and calculate the mean of smoothed segments and standard deviation of noise residual, respectively, which are collected as NLF estimation sample set Ω_1 . Then, define a metric D_1 to fit the lower envelope of the collected Ω_1 . Next, for edge regions, collect CRF estimation sample set Ω_2 by seeking the edge patches containing two regions with distinct but uniform colors, as well as define a metric D_2 to transform edge histograms into linear distributions. Last, the NLF is optimized using a MAP criterion after setting up a likelihood model with combination of D_1 and D_2 . Figure 2 is a block diagram of the proposed NLF estimation method. More details are introduced in the following subsections.

3.1 Sample sets for NLF estimation and definition of distance metrics

We first use the Canny operator and three-pixel dilation to divide the image I into two groups of regions: edge and non-edge. Assume these regions are independent of

Fig. 2 Block diagram of the proposed NLF estimation method



each other. We group non-edge pixels to form non-overlapping regions based on spatial and intensity similarities using a mean-shift segmentation method as described in [3]. We apply a de-noising filter, denoted $d(\cdot)$, to produce a smoothed version of the image, I_D . Here, $d(\cdot)$ is a wavelet-based adaptive filter [21], which has been shown to be effective for images contaminated by common noise [16]. We define a noise residual N as the difference between I and I_D . Then I_D and N are grouped into segments simultaneously according to the mean-shift segmentation coordinates. Let I_i , I_{Di} and N_i be the i -th segments in I , I_D and N , and \hat{I}_i and \hat{O}_i are the average of I_{Di} and standard deviation of N_i respectively. For each segment, we get a sample pair consisting of \hat{I}_i and \hat{O}_i . If a segment N_i is small enough or \hat{O}_i is larger than the global standard deviation of N , we consider it as improper and discard it. With all proper segments, we can get a sample set $\{(\hat{I}_i \text{ and } \hat{O}_i)\}$ for each color channel, i.e., red, green and blue. Compared to [13], instead of spatial average and variation from each I_i , we can obtain more precise samples for the fitting by collecting samples of \hat{I}_i and \hat{O}_i from I_{Di} and N_i respectively.

However, since no ideal filter can perfectly separate the image content from noise, we cannot estimate the noise level of an image directly from $\{(\hat{I}_i \text{ and } \hat{O}_i)\}$. There are two types of errors in N : image details mistreated as noise, and noise mistreated as image details. Generally, probability of the first type of errors is larger than the second. For this reason, we calculate differences between the individual samples and the lower envelope of all samples, and use the sum of these differences as a measure of the difference between the ground-truth NLF and that of the collected samples. To obtain the lower envelope of $\{(\hat{I}_i \text{ and } \hat{O}_i)\}$ for each color channel, we discretize the entire range of intensity $[0, 1]$ into equal intervals $\{[nh, (n+1)h]\}$ with $n = 0, 1, \dots, (1/h-1)$, where h is the interval and is set to $1/64$ in the present work. For any subset $\Lambda_n = \{(\hat{I}_i \text{ and } \hat{O}_i) | nh \leq \hat{I}_i \leq (n+1)h\}$, we find a pair (\hat{I}_n, \hat{O}_n) with the minimum $\hat{O}_n = \min_{\Lambda_n} \hat{O}_i$. By traversing all possible Λ_n , we can get a sample set $\{(\hat{I}_n, \hat{O}_n) | 0 \leq n \leq 1/h-1\}$ from each channel. We then combine all sets collected from all channels to form a set, denoted Ω_1 . Suppose the ground truth NLF is

known, denoted σ_1 . We use the sum of square error to measure the difference between σ_1 and the samples $\hat{\sigma}_n$ in Ω_1 . Denote this difference metric as D_1 :

$$D_1(\sigma_1; \Omega_1) = \sum_{k=R,G,B} \sum_{n=0}^{1/h-1} [\sigma_{1k}(\hat{I}_{nk}) - \hat{\sigma}_{nk}]^2 \quad (8)$$

where the subscript k is used to distinguish components in different color channels. Following the model derived from Equation (7), Equation (8) can be rewritten as

$$D_1(g, \sigma_1, \sigma_2; \Omega_1) = \sum_{k=R,G,B} \sum_{n=0}^{1/h-1} \left[\frac{\sqrt{g(\hat{I}_{nk})\sigma_{1k}^2 + \sigma_{2k}^2}}{g'(\hat{I}_{nk})} - \hat{\sigma}_{nk} \right]^2 \quad (9)$$

Taking a color image sized 512×512 as shown in Fig. 3(a), we explain the process of collecting $\{\hat{\sigma}_{nG}, \hat{\sigma}_{nB}\}$ from the non-edge regions step by step as illustrated in Fig. 3(b)–(e). From Fig. 3(e), we cannot guarantee that all samples \hat{I}_n cover the full range of image intensities, especially for small sized images. As stated in Section 2, NLF strongly depends on the shape of CRF. Therefore, we use an edge-based inverse CRF estimation method proposed in [11] to restrict the profile of NLF. Suppose a small patch in the image containing two regions of slowly varying intensity, denoted R_1 and R_2 . Let \hat{R}_1 and \hat{R}_2 be the mean color of R_1 and R_2 respectively. In addition, assume that a pixel \hat{R}_E is on the boundary between R_1 and R_2 . Here \hat{R}_1 , \hat{R}_2 and \hat{R}_E all have three elements representing intensities in the RGB channels respectively. Radiance of the boundary pixel should be a linear combination of the pixels in R_1 and R_2 before nonlinear CRF mapping [11]. This property is applied to estimate the inverse CRF g . We now seek a function g to map \hat{R}_1 , \hat{R}_2 and \hat{R}_E back to the linear relationship in the color spaces. Let the distance from $g(\hat{R}_E)$ to line $g(\hat{R}_1)g(\hat{R}_2)$ be H , which can be computed as:

$$H = \frac{|[g(\hat{R}_1) - g(\hat{R}_2)] \times [g(\hat{R}_E) - g(\hat{R}_2)]|}{|g(\hat{R}_1) - g(\hat{R}_2)|} \quad (10)$$

where $|\cdot|$ is the Euclidean norm (L^2 -norm) of a vector and \times denote a cross product of two vectors. After scanning all patches along the extracted edges, we select the patches containing two regions with distinct but uniform colors as valid sample patches for estimation. A close-up view of a valid patch and a complete view of all selected edge patches are shown in Fig. 3(g) and (h) respectively. Suppose the total number of valid patches is M , and a subscript m is added to indicate the m -th patch, then we get a sample set $\{\hat{R}_{1m}, \hat{R}_{2m}, \hat{R}_{Em}\}$, denoted Ω_2 . Given an inverse CRF g , we define a total distance metric D_2 to measure the linearity mapping by g :

$$D_2(g; \Omega_2) = \sum_{\Omega_2} H_m^2 \quad (11)$$

Different from [11], we define D_2 as the sum of squares instead of direct accumulation to make the definitions of D_1 and D_2 consistent. Although techniques for CRF estimation are available, we apply an edge-based method for the following

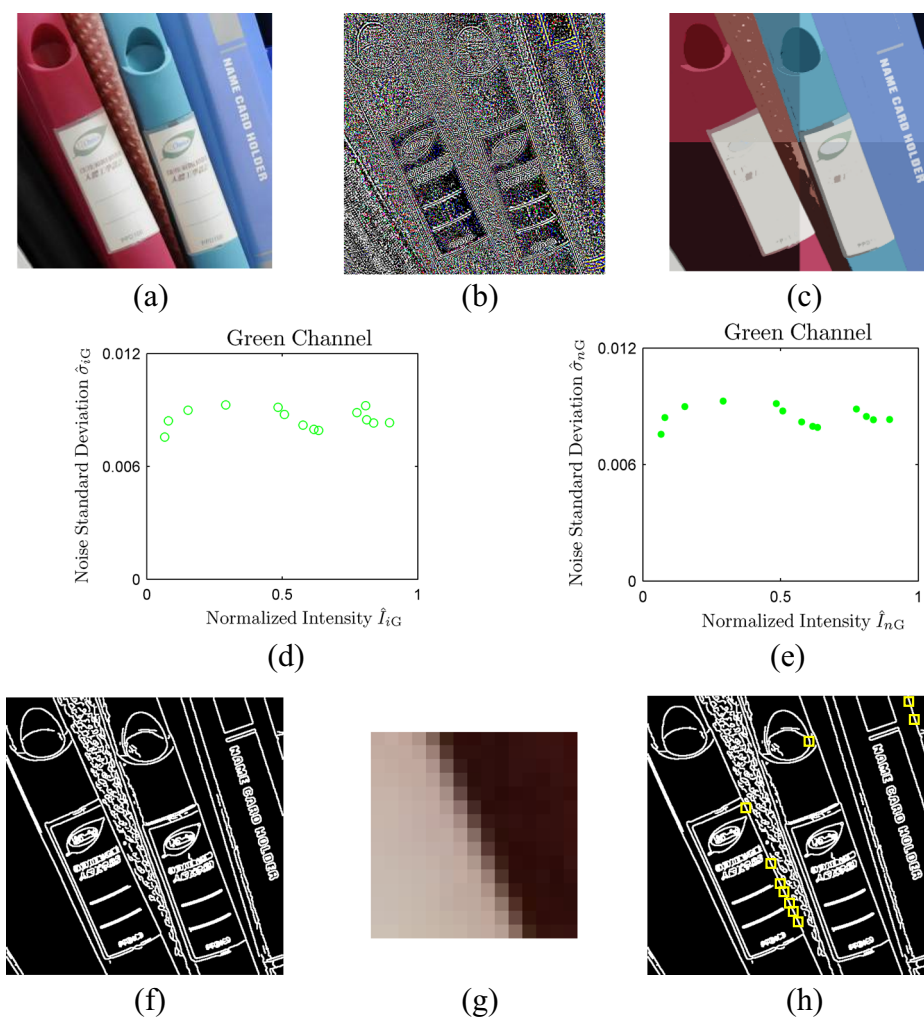


Fig. 3 Sample set collection for NLF estimation. **(a)** A 512×512 image cropped from a photo taken by Nikon D300 with shutter speed 1/15 s and ISO 1600. **(b)** Residual noise extracted using the filter proposed in [21]. **(c)** Mean-shift segmented image in which each segment is represented by the mean color of the region. **(d)** A sample set $\{(\hat{I}_i, \hat{\sigma}_i)\}$ in the green channel, collected by computing the mean of smoothed segments and standard deviation of noise residual. **(e)** Lower envelope of **(d)**. **(f)** Edge map extracted from **(a)**. **(g)** A close-up view of a valid edge patch, in which the blending edge pixel \hat{r}_E is centrally located. **(h)** A complete view of all selected patch samples labeled by yellow boxes

reason: Image edge regions and non-edge regions can be treated as independent components, and this merit can greatly facilitate the construction of a likelihood model in the next step.

So far we have explained how the two independent sample sets Ω_1 and Ω_2 can be collected from non-edge regions and edge regions. In addition, two distance metrics D_1 and D_2 are defined as measures of differences between the samples and the estimated results. In the next step, D_1 and D_2 will be minimized simultaneously using Bayesian MAP inference.

3.2 NLF estimation based on Bayesian MAP

To estimate NLF from insufficient sample sets Ω_1 and Ω_2 , we use a Bayesian maximum MAP inference. Before maximizing the *a posteriori* probability, *a priori* and likelihood models should be set up.

Suppose g , σ_1 and σ_2 are independent of each other. Based on Equation (7), the *a priori* model of NLF can be simplified as:

$$P(\sigma_1) = P(g, \sigma_1, \sigma_2) = P(g) P(\sigma_1) P(\sigma_2) \quad (12)$$

where $P(g)$ is the *a priori* probability of inverse CRF. $P(\sigma_1)$ and $P(\sigma_2)$ are *a priori* probabilities of σ_1 and σ_2 , supposed to obey the uniform distribution. Due to sampling inadequacy, it is impossible to estimate g with superfluous parameters. To use fewer coefficients to represent g as suggested in [6], any g can be concisely represented as

$$g = g_0 + \sum_{j=1}^J \alpha_j g_j \quad (13)$$

where g_0 and g_j ($j = 1, 2, \dots, J$) are the mean vector and eigenvectors of 201 real-world inverse CRF curves, respectively. All these vectors can be downloaded from the database of real-world CRFs (DoRF) [6], which are derived from a principal component analysis (PCA) on the observed CRF data. Equation (13) indicates that a limited number of representation coefficients α_j ($j = 1, 2, \dots, J$) can represent g . Note that the lengths of vectors g_0 and g_j are all 1024 in this model. Following [11], J is set to 5 and the *a priori* probability of g is formulated by a Gaussian mixture model which can be trained from the PCA coefficients in the DoRF database.

The likelihood of σ_1 is the occurrence probability of the observed sample sets Ω_1 and Ω_2 once the parameters for estimation are given. Since Ω_1 and Ω_2 can be regarded as independent sets and the likelihood probability decreases monotonically with the increase of D_1 and D_2 , the likelihood function $L(\sigma_1)$ is defined as a negative joint bivariate exponential distribution:

$$\begin{aligned} L(\sigma_1) &= P(\Omega_1, \Omega_2 | g, \sigma_1, \sigma_2) \\ &= P(\Omega_1 | g, \sigma_1, \sigma_2) P(\Omega_2 | g) \\ &\propto \exp[-\lambda_1 D_1(g, \sigma_1, \sigma_2; \Omega_1) - \lambda_2 D_2(g; \Omega_2)] \end{aligned} \quad (14)$$

where likelihood L is proportional to a product of two negative exponential functions with respect to D_1 and D_2 , respectively. The weights λ_1 and λ_2 control intensity of the constraint on the shape of NLF, and are both set to 20 empirically in the present work.

Now, based on the *a priori* and likelihood models, we solve the problem with Bayesian MAP inference. In general, for a color image the three CRF curves corresponding to the R, G, and B channels are different. However, since the differences are not large, we consider that the inverse CRFs of the three channels are the same to make things easier. With this simplification, we can estimate all three NLFs using 11 parameters: $\{(\alpha_1, \dots, \alpha_5, \sigma_{1R}, \sigma_{2R}, \sigma_{1G}, \sigma_{2G}, \sigma_{1B}, \sigma_{2B})\}$. Taking logarithm

of the *a posteriori* function, the Bayesian inference turns out to be a minimization problem:

$$\begin{aligned}\sigma_1^* &= \underset{\{\alpha_l, \sigma_{1k}, \sigma_{2k}\}}{\operatorname{argmin}} [\lambda_1 D_1(g, \sigma_1, \sigma_2; \Omega_1) + \lambda_2 D_2(g; \Omega_2) - \log(P(g))] \\ &= \underset{\{\alpha_l, \sigma_{1k}, \sigma_{2k}\}}{\operatorname{argmin}} \lambda_1 \sum_{k=\text{RGB}} \sum_{n=0}^{1/h-1} \left[\frac{\sqrt{g(\hat{I}_{nk})\sigma_{1k}^2 + \sigma_{2k}^2}}{g'(\hat{I}_{nk})} - \sigma_{nk} \right]^2 \\ &\quad + \lambda_2 \sum_{m=1}^M \left\{ \left| \frac{[g(\hat{R}_{1m}) - g(\hat{R}_{2m})] \times [g(\hat{R}_{Em}) - g(\hat{R}_{2m})]}{|g(\hat{R}_{1m}) - g(\hat{R}_{2m})|} \right| \right\}^2 - \log(P(g))\end{aligned}\quad (15)$$

To obtain the minimum of Equation (15), we seek local minimum using the non-linear Levenberg–Marquardt–Fletcher algorithm [19] with 40 groups of different initial values. To reduce computation complexity, the maximum number of iterations for each group is limited to 200. Finally NLF is determined by selecting one group of optimal values from all candidate minima followed by low-pass filtering to refine the shape of NLF.

4 Image Splicing Detection

A spliced image is formed with two or more parts taken from different origins that may contain inconsistent noise characteristics. In principle, we can separate the global image into several local segments and estimate all their NLFs exhaustively using the method presented in the previous section. However, the computational complexity is relatively high, especially for some images with larger sizes. As a trade-off between complexity and precision, we propose a suspicious region based splicing detection method. To start with, we can manually select three or more than three suspicious regions and one background from the image for detection. This manipulation is feasible for the occasion that the number of images for detection is relatively small. Once the number turns to be high, we suggest applying saliency detection method such as [8] to locate suspicious regions automatically. Suppose we select S suspicious regions, denoted SR_1, SR_2, \dots, SR_S , as likely candidates for forgery identification and a background region BG . Next, we estimate NLFs from the background BG and candidate regions SR_1, SR_2, \dots, SR_S respectively using the method described in Section 3. Let $\sigma_{BG}^*, \sigma_{SR1}^*, \sigma_{SR2}^*, \dots, \sigma_{SRs}^*$ denote the recovered NLFs from $BG, SR_1, SR_2, \dots, SR_S$ respectively. Set σ_{BG}^* as a reference pattern and define a factor C_{B-Rs} , $s = 1, 2, \dots, S$ to measure the inconsistency between background and each salient region:

$$C_{B-Rs} = \sum_{k=\text{RGB}} \sum_{l'=0}^1 (\sigma_{SRsk}^*(I') - \sigma_{BGk}^*(I'))^2 \quad S = 1, 2, \dots, S \quad (16)$$

where the sum of squared errors of prediction (SSE) for all color channels is calculated to measure the discrepancy. The maximum of C_{B-Rs} , denoted $C_{B-R\max}$, indicates the most

probable forgery region. To determine whether the block is spliced or not, we calculate the ratio of C_{B-Rmax} to the average of remaining C_{B-Rs} , denoted C_{B-Rave} , as:

$$\eta = C_{B-Rmax} / C_{B-Rave} \quad (17)$$

where η is a forgery detector to evaluate deviation of the likely candidate from other regions. When the ratio is smaller than a predefined threshold T , (which will be discussed in subsection 5.2) the image for detection can be grouped into noise consistency images. Otherwise, a forged region is identified and the second maximum of C_{B-Rs} is determined according to (17). More splicing regions can be identified until the number of remaining C_{B-Rs} equals two. Thus, we can locate at most $N-2$ forged regions theoretically and N is usually set to 3, 4, or 5. However, to improve the accuracy of spliced region detection, a boundary based object segmentation method [2] is applied to generate a refined forgery detection map.

5 EXPERIMENTAL RESULTS

5.1 NLF estimation

To evaluate the proposed method, we compare the estimated NLF with the ground-truth NLF curves both for synthetic and captured real photos.

First, we evaluate our method on images contaminated by artificial noise. To compare with the existing NLF estimation methods, we use the same 17 synthetic test images as used in [13] taking from the Berkeley image segmentation database [20]. According to (7), the ground truth NLF can be generated by setting a specified inverse CRF function g and two noise parameters σ_1 and σ_2 . In this experiment, we select the 60th function in the CAVE database [6], $\sigma_{1G} = \sigma_{1R} = \sigma_{1B} = 0.10$, and $\sigma_{2G} = \sigma_{2R} = \sigma_{2B} = 0.02$. Some synthesized noisy images for testing are shown in Fig. 4(a)–(d) and their corresponding grayscale histograms are given in Fig. 4(e)–(h). The histograms show that most pixels in Fig. 4(a) are distributed at both ends of

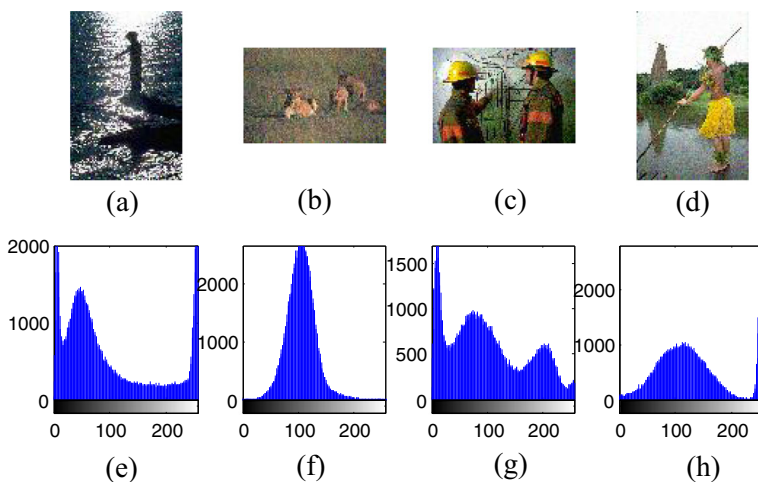


Fig. 4 Synthesized noisy images and their corresponding grayscale histograms. (a)–(d): test images, and (e)–(h): corresponding histograms of (a)–(d)

the histogram. In other words, few samples can be found in the middle range of intensity for estimation. In Fig. 4(b), the image intensity has a relatively narrow range, concentrated in the middle part of the histogram. Obviously, it is challenging to estimate low-end and high-end NLF curve from samples with intensity mainly in the middle range. Images in Fig. 4(c)–(d) have relatively wide intensity ranges and few texture contents. Thus, enough samples can be extracted to fit the whole NLF curves.

Fig. 5(a)–(d) are estimated results from the test images in Fig. 4 (a)–(d) respectively. From left to right, each column provides estimated NLFs (color solid lines), observed sample sets Ω_1 (color dots) and the reference ground-truth NLFs (gray dashed lines) of red, green and blue components, respectively. For comparison with the up-to-date methods, the results estimated by the Bayesian inference based method [13] (black dotted lines), the sparse representation based method [27] (cyan dash-dot lines) and DCT domain based method (yellow dash-dot lines) [28] are also shown in Fig. 5. As [15] assumes a noise model very different from [13, 28] and ours, comparison with [15] on synthetic photos is not made in the experiments.

We observe that, for test images with wide range of intensity such as Fig. 4 (c) and Fig. 4 (d), NLF estimates with all methods coincide with the ground-truth NLFs. However, for some images with relatively narrow intensity range like Fig. 4 (a) and Fig. 4 (b), the proposed method and the method of [28] outperform [13] because [13] assumes all samples are well distributed. If the samples do not span a sufficient range, there must be notable deviation from the ground-truth.

We use root-mean-square error (RMSE) and infinity norm (L^∞ -norm) to measure the difference between the estimated and reference NLFs. L^∞ -norm of a vector is the absolute maximum among the elements. Table 1 compares error statistics of the proposed method with [13, 27]. Means and

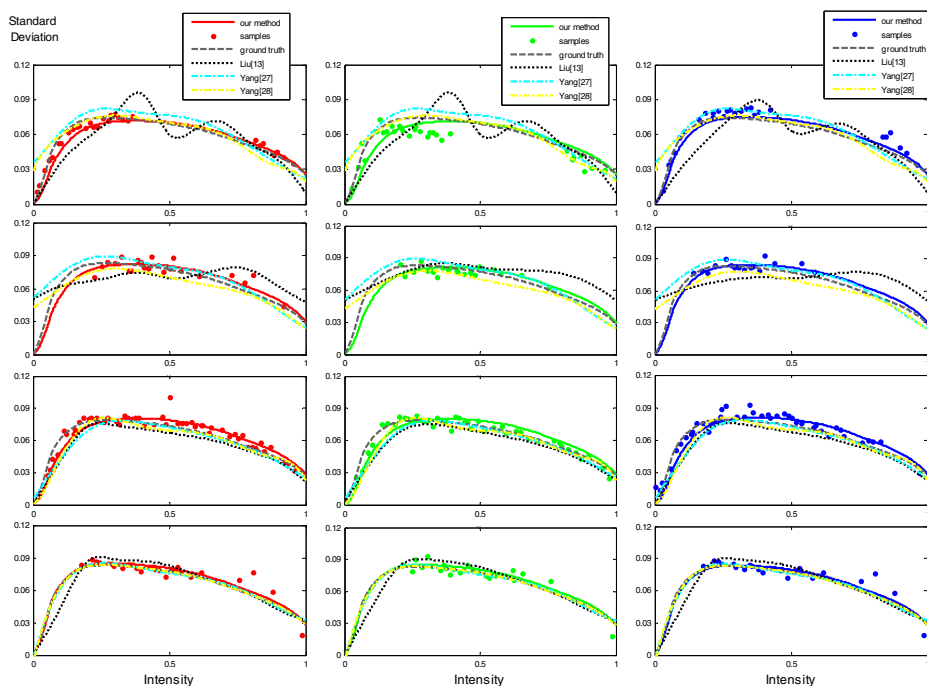


Fig. 5 Estimated NLFs compared with the ground-truth values and results of the recent methods [13, 27, 28]. From top to bottom: estimation results of Fig. 4 (a) to (d), respectively

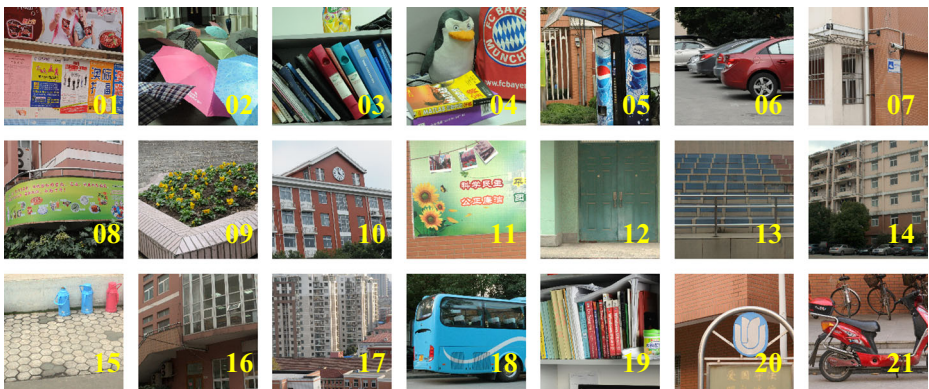
Table 1 Error statistics of RMSE and L^∞ -norm between estimated NLF and ground truth

Error Metrics	Proposed		Liu et al. [13]		Yang et al. [27]	
	mean	variance	mean	variance	mean	variance
RMSE	0.0029	4.53×10^{-7}	0.0048	1.09×10^{-5}	0.0035	5.08×10^{-7}
L^∞ -norm	0.0097	1.16×10^{-5}	0.0110	1.44×10^{-4}	0.0113	1.07×10^{-5}

variances of the two metrics are given. From the table we observe that the proposed method has a smaller statistical error between the estimates and the ground truth for synthesized images.

Next, we compare the NLF estimates with the ground-truth NLF curves from captured real photos. Following [13], the reference ground-truth NLF is obtained by taking 30 photographs of a stationary scene with the same camera and the same settings, and calculating a mean image. We collect 21 groups of test photos as our dataset, shown in Fig. 6. The photos are numbered consecutively from top left to bottom right, where 1–10 are captured with Nikon D300, 11–16 Nikon D3100, 17–19 Canon EOS 400D, and 20–21 Sony A350. Photos are saved either in RAW or fine JPEG formats, and cropped or down-sampled to 1024×1024 to reduce computation complexity. Examples of the estimation results with the four cameras are given in Fig. 7. The left column shows the test images, the other three columns give the estimated NLFs (solid color lines), the observed sample sets Ω_1 (dotted color lines) and the reference ground-truth NLFs (dashed gray lines) of red, green and blue components, respectively. From the depicted ground-truth NLFs, we observe that the noise standard deviation exposes the strong relationship with respect to intensity. From another point of view, this observation also demonstrates the reasonability of our assumed noise model.

For comparison with other recent methods, NLFs recovered by sparse presentation based method [27] (dash-dotted cyan lines), DCT domain based method [28] (dash-dotted yellow lines) and gradient-descent base method [15] (dotted black lines) are also shown in Fig. 7. We observe that most NLF estimates are in good agreement with the ground-truth, and the proposed method performs better than the method [15]. Note that the assumed noise model in [15] is very different from ours. Deviation exists at the high intensity end in [15] due to the effect of neglecting camera response functions. The main difference between [27, 28] and the present work is that, in our case, NLFs of all channels can be estimated from the corresponding color samples, while in

**Fig. 6** Test scenes for evaluation of noise level function

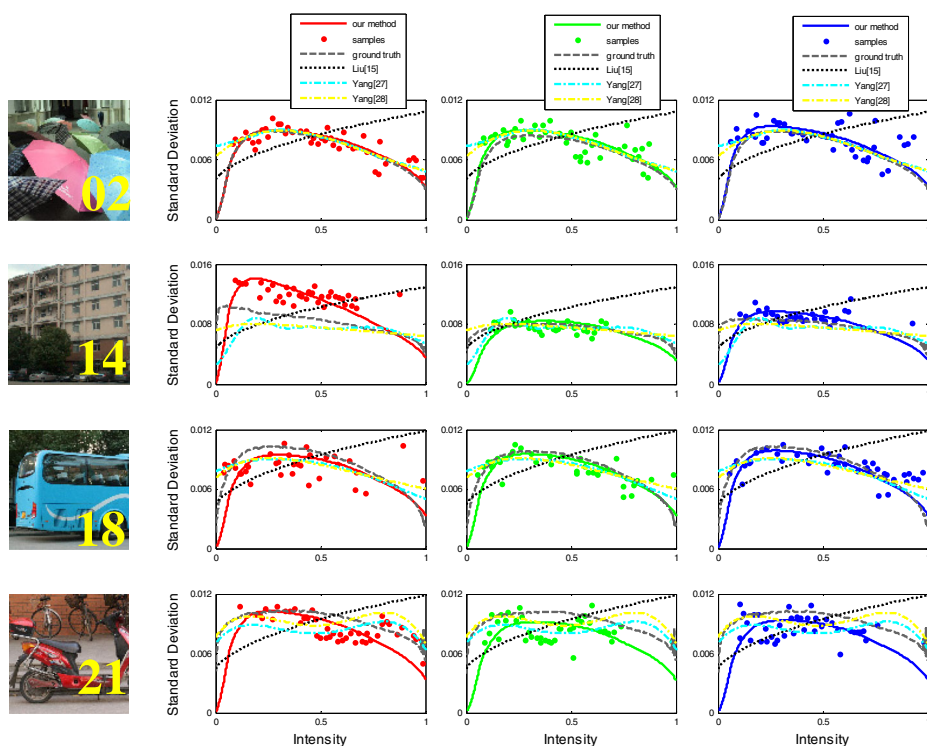


Fig. 7 Examples of NLF estimation from single images compared with the ground-truth NLFs obtained from multiple images and the estimation results by [15, 27, 28]

[27, 28], NLF is estimated from the green samples or gray intensity for all channels. Figure 7 shows distinct differences between color channels in almost all test images.

The estimation result from the red component of test image No. 14 shows clear disagreement with the ground-truth. This error may be caused by a large area of complicated texture in the lower left part of the image. Figure 8 presents RMSE of all test images in all color channels, indicating that the proposed method has a relative low statistical error between the estimates and the ground truth.

Fig. 8 RMSE of the test images for each channel

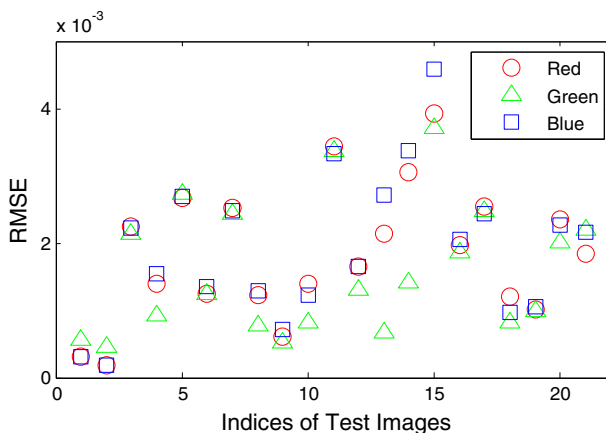
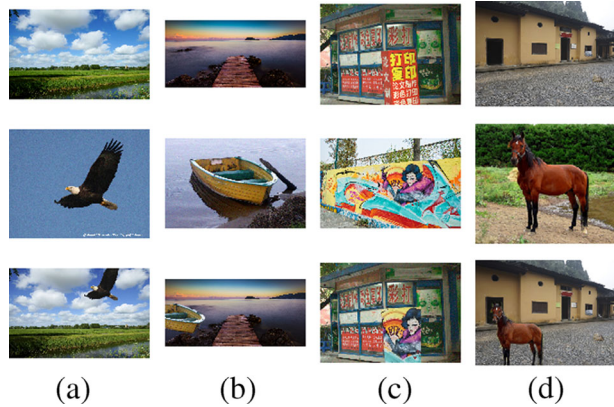


Fig. 9 Examples of splicing detections. From top to bottom: original photos used as background, original photos from which the inserted objects are taken, and synthesized fake photos



5.2 Detection of image splicing

We now apply the proposed method described in Section 4 to detect image splicing. In most cases of image forgery, attractive objects are inserted into background photos. We download some images with background and attractive objects from the internet, and make forged photos using Adobe Photoshop. On the bottom row of Fig. 9 are

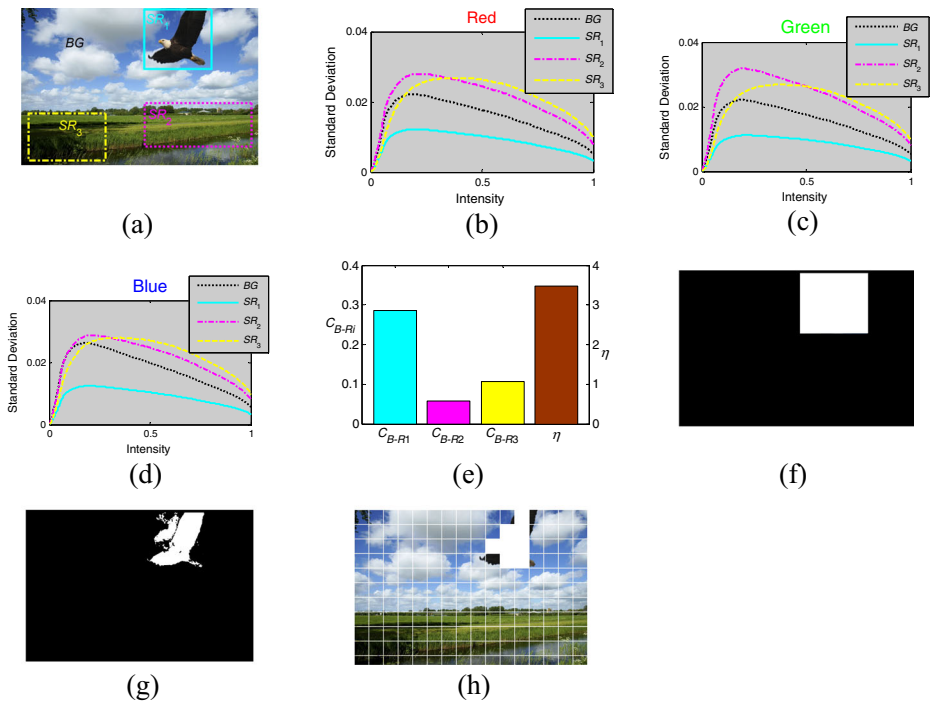


Fig. 10 Forgery detection results of Fig. 9(a): **a** salient regions extraction; **b-d** NLF estimates from all salient regions and background in red, green and blue channels, respectively; **e** inconsistency metric of $CB-R_i$, and candidate deviation evaluator η as indicated by the vertical axis on the right-side; **f** preliminary detection results; **g** refined detection results based on [2]; and **h** detection result using the method of [22]

Table 2 Detection rates on 100 forged and authentic photos with different thresholds

Detection rate	Proposed					Pan et al. [22]
	T=2	T=2.5	T=3	T3.5=	T=4	
True detection rate (TDR)	0.90	0.87	0.85	0.81	0.72	0.84
False detection rate (FDR)	0.25	0.21	0.13	0.10	0.08	0.20

spliced photos, with corresponding background pictures on the top row, and inserted objects in the middle. The images are captured with different camera models using different exposure and ISO settings, therefore have different noise levels.

The four spliced images in the bottom row of Fig. 9 are detected, with the results given in Fig. 10, 11, 12 and 13. In each figure, (a) shows the selected suspicious regions, annotated SR_1 , SR_2 and SR_3 , where the number of suspicious regions is set to 3; (b) through (d) give estimated NLFs from all saliency and background regions in the corresponding red, green and blue channels, respectively. The inconsistency metric of each saliency region and background (viz. C_{B-Ri} , $i = 1, 2, 3$) and their corresponding candidate deviation evaluator η are calculated and illustrated in (e). If η is larger than a predefined threshold T , which was set to 3 here, the candidate saliency region is marked as spliced region. Sub-figures (f) and (g) in Fig. 10, 11, 12 and 13 provide our preliminary and refined detection results, with white regions indicating forged areas. For comparison, (h) shows detection results obtained using another method [22]

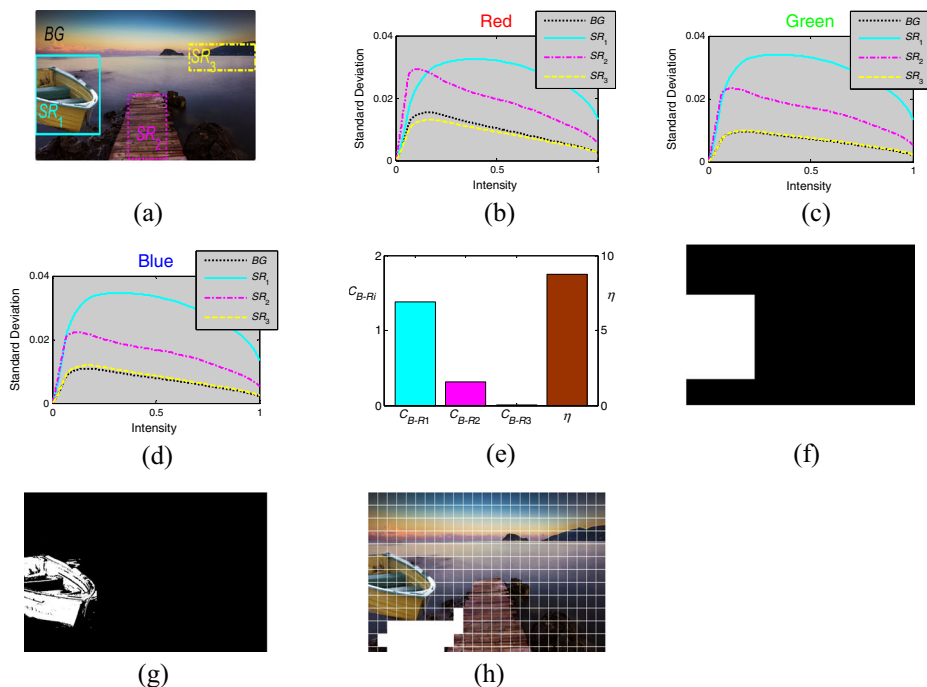


Fig. 11 Forgery detection results of Fig. 9(b): **a** salient regions extraction; **b–d** NLF estimates from all salient regions and background in red, green and blue channels, respectively; **e** inconsistency metric of $CB-Ri$, and candidate deviation evaluator η as indicated by the vertical axis on the right-side; **f** preliminary detection results; **g** refined detection results based on [2]; and **h** detection result using the method of [22]

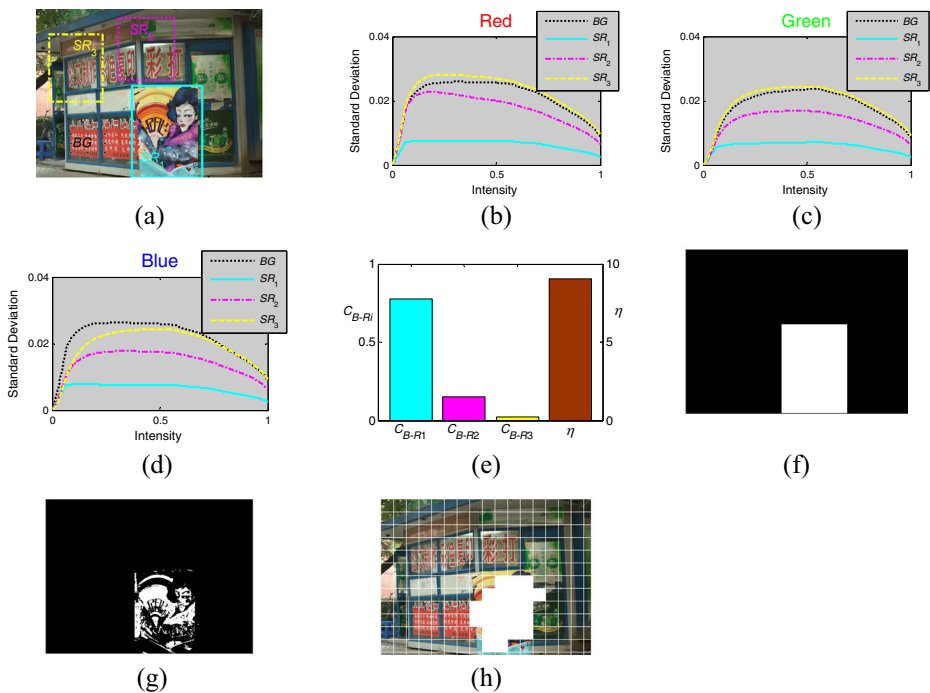


Fig. 12 Forgery detection results of Fig. 9(c): **a** salient regions extraction; **b-d** NLF estimates from all salient regions and background in red, green and blue channels, respectively; **e** inconsistency metric of $CB-Ri$, and candidate deviation evaluator η as indicated by the vertical axis on the right-side; **f** preliminary detection results; **g** refined detection results based on [2]; and **h** detection result using the method of [22]

based on a kurtosis estimator for revealing noise variances inconsistency in DCT domain. From the detection results, we observe that, for photos with low-texture background and high-texture spliced region such as Fig. 9(a), (c) and (d), both methods can provide satisfactory results. However, our method can provide better accuracy since a saliency segmentation technique [2] is applied. Besides, for photos with high-texture background such as Fig. 9(b), the method of [22] produce some false detection as shown in Fig. 11(h), while the proposed method can accurately locate the spliced region as shown in Fig. 11(g).

To seek an optimal threshold T , we collect 100 authentic photos and make 100 forged photos. We then study statistics of the true detection rates (TDR) and false detection rates (FDR) with different thresholds from the forged and authentic photos. TDR is the percentage of correctly identified forged regions that are indeed spliced, while FDR is the percentage of falsely detected tampered regions that are actually authentic. Table 2 lists TDRs and FDRs obtained using the proposed method with thresholds $T = 2, 2.5, 3, 3.5$ and 4, respectively. We take a trade-off between high TDR and low FDR and recommend to use a threshold $T = 3$. In this case, $TDR = 0.85$ and $FPR = 0.13$, both considered acceptable. Besides, to compare the detection rates of the proposed method with [22], the corresponding detection rates of [22] are listed in the right column of Table 2. The table indicates that, with a proper threshold setting, the proposed method has a higher true detection rate and a lower false detection rate.

Finally, we test our methods on some synthesized photos downloaded from an image manipulation and contest website www.worth1000.com, and some authentic photos, as shown in Fig. 14(a)-(b) and (c)-(e) respectively. The left column shows the test images, the two

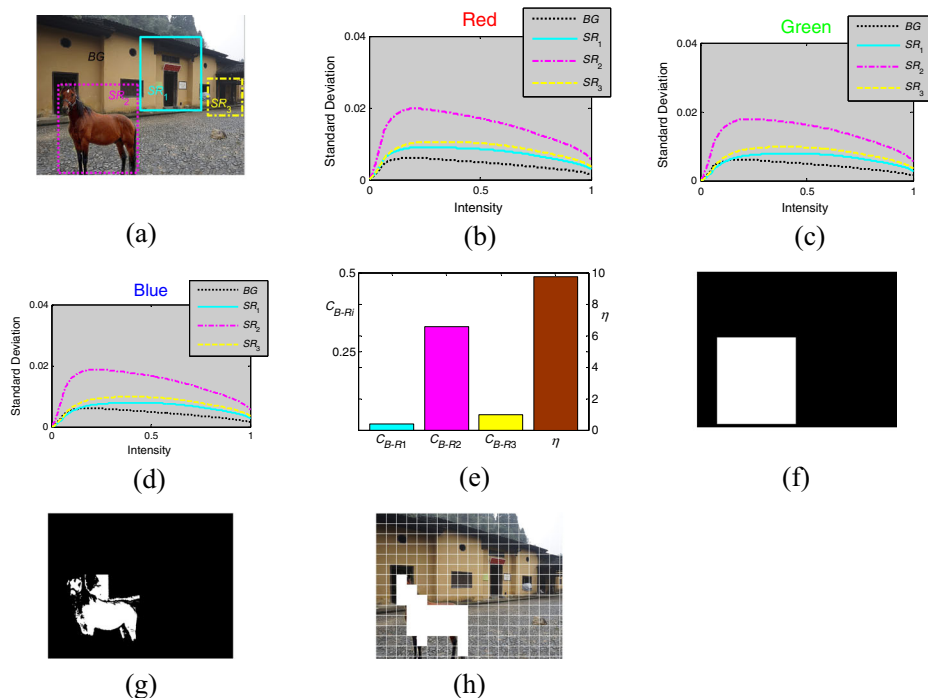


Fig. 13 Forgery detection results of Fig. 9(d): **a** salient regions extraction; **b–d** NLF estimates from all salient regions and background in red, green and blue channels, respectively; **e** inconsistency metric of C_{B-Ri} , and candidate deviation evaluator η as indicated by the vertical axis on the right-side; **f** preliminary detection results; **g** refined detection results based on [2]; and **h** detection result using the method of [22]

middle columns show extracted salient regions and the corresponding C_{B-Ri} and η values respectively. The η values are compared with a given threshold 3 to identify splicing manipulations and locate forged regions. The results are shown in the right-most column. For the tested synthesized photos in (a) and (b), both inserted irregular objects are correctly located as the η values are greater than 3. For most photos with moderate noise level and well-distributed textures such as (c) and (d), all η values are below 3 and therefore correctly identified as authentic. For low noise photos containing strong texture contents, however, the high texture regions may sometimes be judged incorrectly since NLF estimation in these regions is less reliable. In Fig. 14(e), for example, a high-texture region is incorrectly judged as forgery. For good detection performance, the proposed method requires good saliency detection and segmentation algorithms. We expect that splicing detection accuracy can be improved further when better techniques for saliency detection and segmentation become available.

6 Conclusion and Discussion

In this paper, we propose a scheme to estimate the noise level function (NLF) from a single color image, and a method for detecting image splicing by revealing noise level inconsistency in different regions of the test image. Although noise is assumed independent of signals in many researches, it is often unrealistic. In actual imaging systems, photon noise is related to

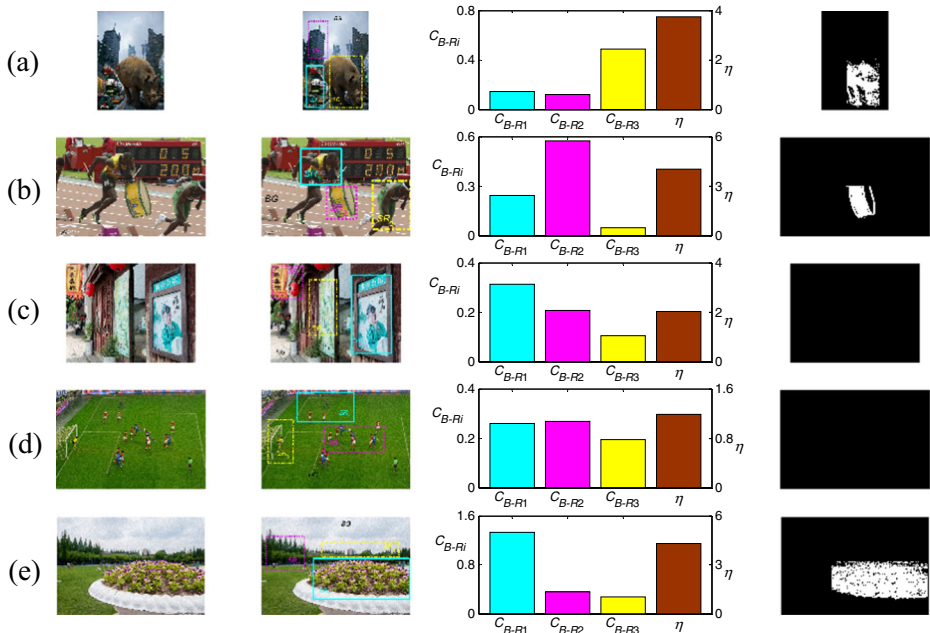


Fig. 14 Experiments on (a)–(b) synthesized photos downloaded from www.worth1000.com, and (c)–(e) authentic photos. Left column: test photos; middle left column: three most saliency regions in each photo; middle right column: inconsistency metric of $CB-Ri$ and candidate deviation evaluator η ; and right column: detection results. Values of η in (a)–(b) are above the given threshold, while that of (c)–(d) are below. These are correctly judged as forged and authentic, respectively. But the authentic photo (e) is falsely judged as a fake as η is above the threshold

brightness levels in a complicated way. Therefore, a radiance dependent noise model is used in the present work. Following a definition of the noise level function, statistical samples of noise are collected from image blocks. To deal with the problem of insufficient sample set, a model is developed for the NLF based on the first-order Taylor expansion. According to this model, the profile of NLF strongly depends on its corresponding camera response function (CRF). To constrain the shape of NLF, we use an edge-based CRF estimator and integrate it into the proposed estimation method within a Bayesian framework. NLF estimation is then applied to image forensics, specifically splicing detection. Experimental results show effectiveness of the NLF estimation and the image splicing detection methods.

In comparison with most state-of-the-art NLF estimation methods, the proposed scheme has higher accuracy in data fitting. Besides, the proposed scheme avoids training of a large amount of data, and optimizes parameters of NLF and CRF simultaneously in the Bayesian inference. In image forensic applications, as far as we know, there is no existing method to explore possible noise level inconsistency between any two blocks from different origins even if they are taken with the same camera. The main contribution of this paper is to reveal this kind of inconsistency and apply it to image splicing detection.

Acknowledgments This work was supported by the National Natural Science Foundation of China (61303203, 61525203, 61472235), the Natural Science Foundation of Shanghai, China (13ZR1428400), the Innovation Program of Shanghai Municipal Education Commission (14YZ087), the Program of Shanghai Dawn Scholar (14SG36), Shanghai Academic Research Leader (16XD1401200), the Open Project Program of the National Laboratory of Pattern Recognition (201600003), Shanghai Engineering Center Project of Massive Internet of Things Technology for Smart Home (GCZX14014), the PAPD Fund, and the CICAET Fund.

References

- Chen M, Fridrich J, Goljan M, Lukas J (2008) Determining image origin and integrity using sensor noise. *IEEE Trans Inf Forensics Security* 3(1):74–90
- Cheng M, Warrell J, Lin W, Zheng S, Vineet V, Crook N (2013) Efficient salient region detection with soft image abstraction. In: *Proc. IEEE Int. Conf. Computer Vision*, Sydney, Australia, p 1529–1536
- Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24(5):603–619
- Donoho D (1995) De-noising by soft-thresholding. *IEEE Trans Inf Theory* 41(3):613–627
- Farid H (2009) Exposing digital forgeries from JPEG ghosts. *IEEE Trans Inf Forensics Security* 4(1):154–160
- Grossberg MD, Naya SK (2004) Modeling the space of camera response functions. *IEEE Trans Pattern Anal Mach Intell* 26(10):1272–1282
- Healey GE, Kondepudy R (1994) Radiometric CCD camera calibration and noise estimation. *IEEE Trans Pattern Anal Mach Intell* 16(3):267–276
- Hou X, Harel J, Koch C (2012) Image signature: highlighting sparse salient regions. *IEEE Trans Pattern Anal Mach Intell* 34(1):194–201
- Kobayashi M, Okabe T, Sato Y (2010) Detecting forgery from static-scene video based on inconsistency in noise level functions. *IEEE Trans Inf Forensics Security* 5(4):883–892
- Li J, Li X, Yang B, Sun X (2015) Segmentation-based Image Copy-move Forgery Detection Scheme. *IEEE Trans Inf Forensics Security* 10(3):507–518
- Lin S, Gu J, Yamazaki S, Shum HY (2004) Radiometric calibration from a single image. In: *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Washington, USA 2:938–945
- Liu C, Freeman WT, Szeliski R, Kang SB (2006) Noise estimation from a single image. In: *Proc IEEE Int Conf Computer Vision and Pattern Recognition*, New York, USA, vol 1, pp 901–908
- Liu C, Szeliski R, Kang SB, Zitnick CL, Freeman WT (2008) Automatic estimation and removal of noise from a single image. *IEEE Trans Pattern Anal Mach Intell* 30(2):299–314
- Liu Q, Cao X, Deng C, Guo X (2011) Identifying image composites through shadow matte consistency. *IEEE Trans Inf Forensics Security* 6(3):1111–1122
- Liu X, Tanaka M, Okutomi M (2013) Estimation of signal dependent noise parameters from a single image. In: *Proc. IEEE Int. Conf. Image Processing*, Melbourne, Australia, pp 79–82
- Lukas J, Fridrich J, Goljan M (2006) Digital camera identification from sensor pattern noise. *IEEE Trans Inf Forensics Security* 1(2):205–214
- Lyu S, Pan X, Zhang X (2014) Exposing Region Splicing Forgeries with Blind Local Noise Estimation. *Int J Comput Vis* 110(2):202–221
- Mahdian B, Saic S (2009) Using noise inconsistencies for blind image forensics. *Image Vis Comput* 27(10):1497–1503
- Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM J Appl Math* 11(2):431–441
- Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating Segmentation Algorithms and Measuring Ecological Statistics. In: *Proc. IEEE Int. Conf. Computer Vision*, Vancouver, Canada, 2, pp 416–423
- Mihcak MK, Kozintsev I, Ramchandran K (1999) Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising. In: *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Phoenix, USA, 6:3253–3256
- Pan X, Zhang X, Lyu S (2011) Exposing image forgery with blind noise estimation. In: *Proc. 13th ACM multimedia workshop on Multimedia and security (MM&Sec'11)*, Buffalo, USA, pp 15–20
- Popescu AC, Farid H (2005) Exposing digital forgeries by detecting traces of re-sampling. *IEEE Trans Signal Process* 53(2):758–767
- Stamm MC, Liu KJR (2010) Forensic detection of image manipulation using statistical intrinsic fingerprints. *IEEE Trans Inf Forensics Security* 5(3):492–506
- Takamatsu J, Matsushita Y, Ikeuchi K (2008) Estimating radiometric response functions from image noise variance. In: *Proc. 10th European Conf. Computer Vision*, Marseille, France vol 4, pp 623–637
- Tsin Y, Ramesh V, Kanade T (2001) Statistical calibration of CCD imaging process. In: *Proc. 8th IEEE Int. Conf. Computer Vision*, Vancouver, Canada, 1, pp 480–487
- Yang J, Wu Z, Hou C (2012) Estimation of signal-dependent sensor noise via sparse representation of noise level functions. In: *Proc. IEEE Int. Conf. Image Processing*, Orlando, pp 673–676
- Yang J, Gan Z, Wu Z, Hou C (2015) Estimation of signal-dependent noise level function in transform domain via a sparse recovery model. *IEEE Trans Image Process* 24(5):1561–1572

29. Yao H (2016) A novel image noise level function estimation approach using camera response function constraint. *MATEC Web Conferences* 42:06004
30. Yao H, Wang S, Zhao Y, Zhang X (2012) Detecting image forgery using perspective constraints. *IEEE Signal Process Lett* 19(3):123–126
31. Yuan H (2011) Blind forensics of median filtering in digital images. *IEEE Trans Inf Forensics Security* 6(4): 1335–1345
32. Zhang W, Cao X, Qu Y, Hou Y, Zhao H, Zhang C (2010) Detecting and extracting the photo composites using planar homography and graph cut. *IEEE Trans Inf Forensics Security* 5(3):544–555



Heng Yao received the B. Sc. degree from Hefei University of Technology, China, in 2004, the M. Eng. degree from Shanghai Normal University, China, in 2008, and the Ph. D. degree from Shanghai University, China, in 2012. Currently, he is with School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, China. His research interests include digital forensics, image processing, and pattern recognition



Shuozhong Wang received the B.S. degree from Peking University, Beijing, China, in 1966, and the Ph.D. degree from University of Birmingham, Birmingham, U.K., in 1982. Currently, he is a Professor with Shanghai University, Shanghai, China. His research interests include image processing, audio processing, and information hiding



Xinpeng Zhang received the B.S. degree in computational mathematics from Jilin University, China, in 1995, and the M.E. and Ph.D. degrees in communication and information system from Shanghai University, China, in 2001 and 2004, respectively. Since 2004, he has been with the faculty of the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. He was with the State University of New York at Binghamton as a visiting scholar from January 2010 to January 2011, and Konstanz University as an experienced researcher sponsored by the Alexander von Humboldt Foundation from March 2011 to May 2012. He is an Associate Editor for IEEE Transactions on Information Forensics and Security. His research interests include multimedia security, image processing, and digital forensics. He has published more than 200 papers in these areas.



Chuan Qin received the B.S. degree in electronic engineering and the M.S. degree in signal and information processing from Hefei University of Technology, Anhui, China, in 2002 and 2005, respectively, and the Ph.D. degree in signal and information processing from Shanghai University, Shanghai, China, in 2008. Since 2008, he has been with the faculty of the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, where he is currently an Associate Professor. He was with Feng Chia University at Taiwan as a Postdoctoral Researcher and Adjunct Assistant Professor from July 2010 to July 2012. His research interests include image processing and multimedia security. He has published more than 60 papers in these research areas



Jinwei Wang received the Ph.D. degree in information security at Nanjing University of Science & Technology in 2007 and was a visiting scholar in Service Anticipation Multimedia Innovation (SAMI) Lab of France Telecom R&D Center (Beijing) in 2006. He worked as a senior engineer at the 28th research institute, CETC from 2007 to 2010. He worked as a visiting scholar at New Jersey Institute of Technology, NJ, USA from 2014 to 2015. Now he works as an associate professor at Nanjing University of Information Science and Technology. His research interests include multimedia copyright protection, image forensics, image encryption and data authentication. He has published more than 30 papers, hosted and participated in more than 10 research projects