Exposing Region Splicing Forgeries with Blind Local Noise Estimation

Siwei Lyu, Xunyu Pan & Xing Zhang

International Journal of Computer Vision

ISSN 0920-5691
Volume 110
Number 2

Int J Comput Vis (2014) 110:202-221
DOI 10.1007/s11263-013-0688-y

Springer

Int J Comput Vis (2014) 110:202–221
DOI 10.1007/s11263-013-0688-y

Exposing Region Splicing Forgeries with Blind Local Noise Estimation

Siwei Lyu · Xunyu Pan · Xing Zhang

Received: 28 February 2013 / Accepted: 30 November 2013 / Published online: 25 December 2013

Abstract Region splicing is a simple and common digital image tampering operation, where a chosen region from one image is composited into another image with the aim to modify the original image's content. In this paper, we describe an effective method to expose region splicing by revealing inconsistencies in local noise levels, based on the fact that images of different origins may have different noise characteristics introduced by the sensors or post-processing steps. The basis of our region splicing detection method is a new blind noise estimation algorithm, which exploits a particular regular property of the kurtosis of nature images in band-pass domains and the relationship between noise characteristics and kurtosis. The estimation of noise statistics is formulated as an optimization problem with closed-form solution, and is further extended to an efficient estimation method of local noise statistics. We demonstrate the efficacy of our blind global and local noise estimation methods on natural images, and evaluate the performances and robustness of the region splicing detection method on forged images.

Keywords Blind local noise estimation · Natural image statistics · Digital image forensics

S. Lyu (✉) · X. Pan · X. Zhang
Computer Science Department, University at Albany, State University of New York, Albany, NY, USA
e-mail: slyu@albany.edu

X. Pan
e-mail: xpan@albany.edu

X. Zhang
e-mail: xzhang@albany.edu

1 Introduction

The past decade has produced stunning advances in digital photography. High quality digital camera and digital image processing techniques started in the past three decades are now commonplace in personal electronics. Billions of digital images are uploaded to photography portals like Flickr.com and Instam on a daily basis, and our ability to editing digital images has improved dramatically, thanks to applications such as Adobe Photoshop or GNU Gimp.

Despite the obvious benefits and conveniences these technologies brought to the users, there also comes an unpleasant ramification—digital images have become more vulnerable to malicious tampering, and the ease of digital image manipulation has started to erode the long-held trustfulness of photographs as reliable records of events. Studies in cognitive psychology (Garry and Gerrie 2005; Sacchi et al. 2007) have suggested that doctored photographs can affect the memory of past events, and we are facing with an increasing number and level of sophistication of digitally manipulated images with negative financial, legal, and/or political consequences. This circumstance calls for the study of digital image forensic (Farid 2009; Sencar and Memon 2012), a research field aiming to find effective solutions to questions such as "how to determine a digital image's authenticity?" and "how to expose tampering operations in a digital image?".

In this work, we focus on the detection of a common and simple manipulation of digital image, known as *region splicing*, which creates a forged image by compositing regions extracted from *different* source images.[1] With carefully chosen source images and the aid of sophisticated image edit-

[1] If the regions are from the *same* image, the corresponding tampering operation is known as *region cloning*, which will not be considered in this work.

ing tools, spliced regions can oftentimes easily evade casual visual inspection.

Subsequently, we describe an effective method to detect spliced regions in a digital image that is based on noise in digital images. Such noise is typically introduced during acquisition or subsequent processing of an image. For an untampered image, one can reasonably assume that the noise statistics across different pixels differ only slightly. Thus, spliced regions from an image with significantly different noise statistics can be exposed with the inconsistency of local noise characteristics. As such, our region splicing detection method is based on a simple method to estimate local noise statistics from an image. Specifically, our blind noise estimation method is motivated by an observed statistical regularity of natural images known as *projection kurtosis concentration*—the kurtosis of natural images in band-pass domains (e.g., obtained from the discrete cosine transform, wavelet decomposition or randomly generated band-pass filters) tend to be close to a positive constant (Bethge 2006; Lyu and Simoncelli 2009; Zoran and Weiss 2009).

The major contributions of this work are summarized as follows:

1. We provide empirical evidence and theoretical justification of the projection kurtosis concentration phenomenon;
2. Using this phenomenon, we formulate noise estimation as the optimization of an objective function that is robust to infrequent outlying projection kurtosis values and affords a closed-form solution;
3. We further extend this method to estimate locally varying noise levels that is accelerated with *integral image*;
4. We demonstrate the performances and robustness of our method on several sets of forged images with spliced regions, which compare favorably with those of several state-of-the-art region splicing detection methods.

The rest of the paper is organized as follows. After reviewing relevant work in Sect. 2, we describe the projection kurtosis concentration property of natural images in Sect. 3. Then the blind global noise estimation algorithm based upon this property is introduced in Sect. 4. This algorithm is further extended to the estimation of locally varying noise levels in Sect. 5. In Sect. 6, we elaborate on the splicing detection method using the blind local noise estimation method. Section 7 concludes the paper with discussion and future works. Some preliminary results of this work were reported in Pan et al. (2012c).

2 Related Works

In this section, we briefly review relevant previous works on blind noise estimation and region splicing detection.

2.1 Blind Noise Estimation

Noise can be attributed to numerous factors during the acquisition, transmission and processing of digital images, including thermal effects, sensor saturation, quantization and random interference of the communication channels (Nakamura 2006). Effective estimation of noise statistics from a noise corrupted image, known as blind noise estimation, is an important task in image processing, and many further tasks such as denoising and deblurring predicate on the knowledge of the noise statistics.[2] Blind noise estimation methods differ in the choice of image and noise models. Most existing techniques assume a stationary noise model with spatially invariant statistics, usually a zero-mean white Gaussian process that is additive and independent of the image signal. As such, most existing noise estimation methods focus on the estimation of noise variances.

One commonly (though usually implicitly) used assumption in blind noise estimation methods is that there are "dead zones" in the noise-free images or their transformations where there is no image structure. As variations in such regions are completely due to the noise, they can be used to estimate parameters in the noise model. Early approaches seek such regions in the frequency or multi-scale image representations. For instance, the method of Meer et al. (1990) estimates noise variance in the highest-frequency part of the noisy image's Fourier spectrum. Based on the observation that there are less edges and contours with diagonal orientations in natural images, noise variance is estimated in the high-frequency diagonal subband of an orthonormal wavelet decomposition using the median absolute deviation (MAD) estimator (Donoho 1995). Another family of methods work in the spatial domain to identify regions in the noisy image corresponding to portions of the original image with near constant values. Early approaches use simple thresholding to find such regions (Olsen 1993; Rank et al. 1999). More sophisticated approaches group regions in the noisy image that have similar structures, and estimating noise from the residuals after the common structures are removed (Konstantinides and Natarajan 1997; Ponomarenko et al. 2003; Foi et al. 2007; Danielyan and Foi 2009), or select patches that do not have strong edges and contours (Bilcu and Vehvilainen 2005; Tai and Yang 2008; Liu et al. 2012).

The works in Benedict and Soong (1967), Matzner and Engleberger (1994), Pauluzzi and Beaulieu (2000) represent a different general methodology of noise estimation, which exploits the relationship between noise variance and higher-order statistics, especially kurtosis, of noise free images.

[2] There are also methods that incorporate noise estimation into tasks such as denoising (Portilla 2004) or deblurring (Schmidt et al. 2011). We shall not discuss such methods subsequently as they are usually less efficient when the purpose is to estimate noise statistics.

However, these early methods are not truly blind, as they assume the knowledge of the kurtosis of the noise free signal, or the availability of regions in the noisy signal where the kurtosis of the noise free signal can be estimated. The first blind noise estimation technique along this line of methods is described in Zoran and Weiss (2009), which takes advantage of the projection kurtosis concentration phenomenon (a detailed comparison with our method will be given in Sect. 4.2).

There have also been noise estimation methods that use more realistic noise models incorporating the sensor characteristics and dependencies of noise statistics and pixel intensities. In the work of Förstner (1998), the noise variance is estimated from the gradient of smooth or small textured regions, assuming a signal-dependent noise model. In Stefano et al. (2004), using training samples and the Laplacian model for the marginal statistics of natural images in band-pass filtered domains, a function is learned for estimating noise variances. The method of Liu et al. (2008) uses a piecewise smooth image model and sensor-based noise model to obtain an upper-bound of the noise variance.

Existing noise estimation methods typically assume a stationary noise model, where noise statistics are invariant to spatial locations (though they may be dependent on the pixel intensity values). When a stationary noise model cannot be assumed, for instance for the digitally tampered images with spliced regions, one needs a blind local noise estimation method. Simply applying existing global noise estimation methods to local patches surrounding each pixel location usually leads to inefficient algorithm. An important objective of our work is to develop an efficient blind local noise estimation method based on global blind noise estimation method.

2.2 Region Splicing Detection

Splicing regions in a digital image are usually detected by exposing significant variations of intrinsic characteristics that would be otherwise consistent in an untampered image. Existing region splicing detection methods can be categorized into two classes. Many existing methods focus on determining if an image contains spliced regions, without revealing the actual location of the spliced region. As such, they are typically implemented with a classifier (commonly SVM or boosting classifiers) learned from a training set of original and forged images. Key to the classification based methods is the choice of the feature vectors that are used by the classifier. Recent works have explored feature vectors obtained from camera response functions (Hsu and Chang 2006, 2007; Lin et al. 2005), photo response non-uniformity noise pattern (Chen et al. 2008), consistency of region and object boundaries, (Chen et al. 2007; Qu et al. 2009; Shi et al. 2007), JPEG compression (Wang et al. 2009) or general statistics of natural images (Bayram et al. 2006; Fu et al. 2007; Ng and Chang 2004). Note that though in principle, the classification based methods can be used to locate spliced regions by applying to sliding windows in an image or to any user specified regions, the resulting algorithm is usually inefficient for practical needs.

On the other hand, *location based* methods provide the extent and location of potential spliced regions. Currently, the most effective splicing detection methods in this category (He et al. 2006; Lin et al. 2009) are based on the quantization artifacts due to the double JPEG compression. However, the applicability of such methods is limited by a few factors including (1) the tampered image must be in JPEG format, (2) the image contributing the background must be in JPEG format and must have a different quality factor from the tampered image, and (3) the spliced region has to exhibit no double JPEG artifact.

Local noise levels provide an alternative approach for the detection of spliced regions, without limiting to specific image file format.[3] The method in Popescu and Farid (2004) is based on an estimator of the noise variance from the kurtosis of the original image. The method in Mahdian and Saic (2009) is based on the MAD noise variance estimator in the high-pass band from an orthonormal wavelet decomposition. The method of Chen et al. (2007) reveals region splicing by detecting the presence of photo-response non-uniformity noise (PRNU) in small regions. This method assumes that either the camera that took the image is available to the analyst or at least some other non-tampered images taken by the camera are available, and is therefore not a blind detection method. In this work, we demonstrate that our blind local noise estimation method leads to a general region splicing detection method that achieves the state-of-the-art detection performance.

3 Projection Kurtosis and Noise Variances

This section describes a statistical property of natural images in band-pass domains, *projection kurtosis concentration*, and its relation with the noise variances. These results are the foundation of the blind noise estimation methods introduced in subsequent sections.

3.1 Kurtosis and Projection Kurtosis

For a 1D random variable x, denoting $\mathcal{E}_x\{f\} = \int_x f(x)p(x)dx$ as the expectation of a function $f(\cdot)$ with regards to its distribution, the kurtosis (some works also referred to this as the excessive kurtosis) (Feller 1968) is defined as

[3] The image noise has also been used for in digital image forensics to identify the source camera models or ensuring authenticity of an image (Filler et al. 2008; Lukas et al. 2006; Lukás et al. 2006).

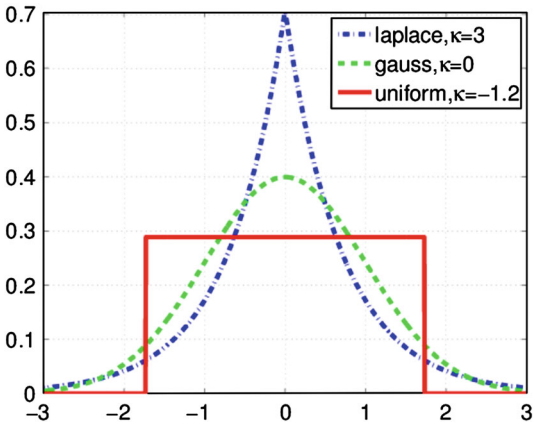

Fig. 1 Kurtosis of three different distribution models with the zero mean and unit variance

$$\kappa(x) = \frac{C_4(x)}{C_2^2(x)}, \tag{1}$$

where $C_2(x) = \mathcal{E}_x\left\{(x - \mathcal{E}_x\{x\})^2\right\}$, which coincides with the variance of x, and $C_4(x) = \mathcal{E}_x\left\{(x - \mathcal{E}_x\{x\})^4\right\} - 3C_2(x)$ are the second and fourth order cumulants of x, respectively (Feller 1968). Kurtosis is invariant to scaling, i.e., $\kappa(sx) = \kappa(x)$ for any $s > 0$, and measures how "peaky" a distribution is. In particular, a Gaussian distribution $p(x) \propto \exp\left(-\frac{1}{2}x^2\right)$ has zero kurtosis, while a super-Gaussian distribution (distributions that are more concentrated around the mean than Gaussian, such as the Laplacian distribution $p(x) \propto \exp(-|x|)$) has positive kurtosis, and a sub-Gaussian distribution (distributions that are flatter than Gaussian, such as the uniform distribution) has negative kurtosis. Figure 1 compares the kurtosis of distributions from three different parametric families (Laplacian, Gaussian and uniform). Each has a mean zero and the parameters have been chosen to result in a variance equal to 1.

The variance and kurtosis can also be expressed using the raw (un-centered) moments, $\mu_m = \mathcal{E}_x\{x^m\}$, as:

$$\sigma^2 = \mu_2 - \mu_1^2 \text{ and } \kappa = \frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4}{\mu_2^2 - 2\mu_2\mu_1^2 + \mu_1^4} - 3. \tag{2}$$

Furthermore, Eq. (2) suggests practical estimations of variance and kurtosis from samples of x by approximating the raw moments of first to fourth order with averages, as $\mu_m \approx \frac{1}{m}\sum_{k=1}^{m} x_k^m$. This fact is important in deriving our local noise estimation method in Sect. 5.

For a d-dimensional random vector $\mathbf{x}$, we define the kurtosis of the 1D projection of $\mathbf{x}$ onto a unit vector $\mathbf{w}$, $\kappa(\mathbf{w}^T\mathbf{x})$, as its *projection kurtosis* with regards to $\mathbf{w}$. The projection kurtosis provides an effective means to probe the statistical properties of high dimensional variables. For instance, if $\mathbf{x}$ is a Gaussian vector, its projection over any $\mathbf{w}$ has a 1D Gaussian distribution, so its projection kurtosis is always zero. Indeed, several effective algorithms for blind source separation (Cardoso 1999) and independent component analysis (ICA) (Hyvärinen 1999) are based on finding projection directions that maximize the projection kurtosis. On large images, we can compute the projection kurtosis of local pixel patches more efficiently by convolving the image with the corresponding filters, and estimate the projection kurtosis from the filter responses.[4]

3.2 Kurtosis and Noise Variances

Consider a random variable $y = x + n$ as the sum of a random variable x and a Gaussian variable n, the additivity of cumulants of independent variables (Feller 1968) implies $C_4(y) = C_4(x)$, or with Eq. (1), $\kappa(y)(\sigma^2(y))^2 = \kappa(x)(\sigma^2(x))^2$. Replacing $\sigma^2(y)$ with $\sigma^2(y) = \sigma^2(x) + \sigma^2(n)$ and rearranging terms yield a simple but fundamental relation between kurtosis and variance, as:

$$\kappa(y) = \kappa(x) \cdot \left(\frac{\sigma^2(x)}{\sigma^2(y)}\right)^2 = \kappa(x) \cdot \left(\frac{\sigma^2(y) - \sigma^2(n)}{\sigma^2(y)}\right)^2. \tag{3}$$

If we know the kurtosis of the original variable x, Eq. (3) can be used to estimate the unknown variance of n, since the kurtosis and variance of y can both be estimated from samples. On the other hand, if $\kappa(x)$ is unknown, this will not work as Eq. (3) provides one constraint with two unknown quantities.

A similar relation can be obtained for the projections of high dimensional signals. Let $\mathbf{z}$ be a zero-mean i.i.d. white Gaussian noise vector with covariance matrix $\sigma^2 I$ and independent of random vector $\mathbf{x}$ with covariance matrix $\Sigma_{\mathbf{x}}$. The variances of the projections of the $\mathbf{z}$, $\mathbf{x}$ and $\mathbf{y} = \mathbf{x} + \mathbf{z}$ on a unit vector $\mathbf{w}$ are given by

$$\begin{aligned}
\sigma^2(\mathbf{w}^T\mathbf{z}) &= \mathbf{w}^T\mathcal{E}_{\mathbf{z}}\left\{\mathbf{z}\mathbf{z}^T\right\}\mathbf{w} = \sigma^2\mathbf{w}^T\mathbf{w} = \sigma^2 \\
\sigma^2(\mathbf{w}^T\mathbf{x}) &= \mathbf{w}^T\mathcal{E}_{\mathbf{z}}\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{w} = \mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w} \\
\sigma^2(\mathbf{w}^T\mathbf{y}) &= \sigma^2(\mathbf{w}^T\mathbf{x}) + \sigma^2(\mathbf{w}^T\mathbf{z}) = \mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w} + \sigma^2.
\end{aligned}$$

Correspondingly, Eq. (3) for the projections becomes

$$\begin{aligned}
\kappa(\mathbf{w}^T\mathbf{y}) &= \kappa(\mathbf{w}^T\mathbf{x})\left(\frac{\sigma^2(\mathbf{w}^T\mathbf{x})}{\sigma^2(\mathbf{w}^T\mathbf{y})}\right)^2 \\
&= \kappa(\mathbf{w}^T\mathbf{x})\left(\frac{\sigma^2(\mathbf{w}^T\mathbf{y}) - \sigma^2}{\sigma^2(\mathbf{w}^T\mathbf{y})}\right)^2
\end{aligned} \tag{4}$$

In practice, when $\mathbf{x}$ are patches from natural images and the projection directions correspond to band-pass filters, we

[4] The projection directions need to be reflected in both horizontal and vertical directions, and the convolution is equivalent to projection assuming proper boundary handling.

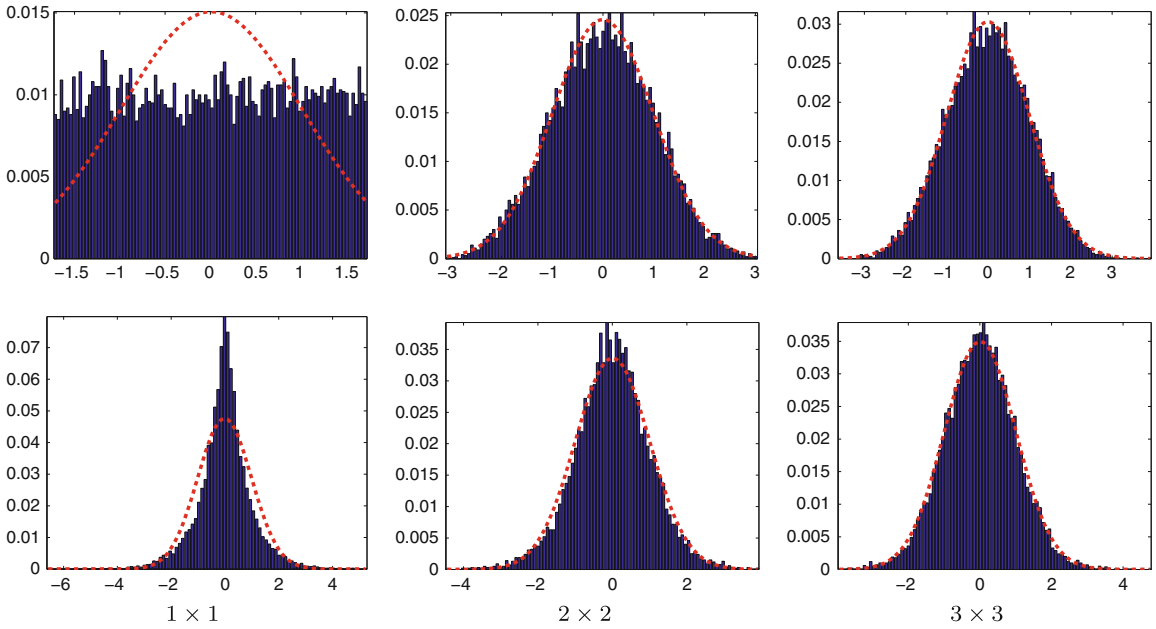

Fig. 2 Histograms of uniform noise (*top*) and Laplacian noise (*bottom*) after convolving with band-pass DCT filters of different sizes (columns). Also shown are the corresponding best fit Gaussian distributions (*dashed lines*)

can take advantage of regular statistical properties of natural images in band-pass domains (Simoncelli and Olshausen 2001).

We would like to point out another important property of Eq. (4), that it is also approximately satisfied when $\mathbf{z}$ follows a i.i.d. non-Gaussian model. To see this, first note that the projection of $\mathbf{w}^T\mathbf{z} = w_1 z_1 + \cdots w_d z_d$ corresponds to a weighted sum of independent variables $z_1, \cdots, z_d$, which are with zero mean and identical and finite variance σ^2. As such, the new random variables $w_1 z_1, \cdots, w_d z_d$ are zero mean and with variances $w_1^2\sigma^2, \cdots, w_d^2\sigma^2$, respectively, and correspondingly the variance of $w_1 z_1 + \cdots w_d z_d$ is $\sigma^2\|\mathbf{w}\|^2 > 0$ and is finite. Further assuming z_i has a symmetric distribution, we have $\mathcal{E}_{z_i}\{z_i^3\} = 0$ and

$$\frac{\sum_{i=1}^{d}\mathcal{E}_{z_i}\{(w_i z_i)^3\}}{\sigma^3\|\mathbf{w}\|^3} = \frac{\sum_{i=1}^{d} w_i^3\mathcal{E}_{z_i}\{z_i^3\}}{\sigma^3\|\mathbf{w}\|^3} = 0,$$

Thus these variables satisfy the Lyapunov condition for $\delta = 1$, and the Lyanunov's central limit theorem (Appendix 1.) (Feller 1968) applies, i.e., $\mathbf{w}^T\mathbf{z}$ asymptotically approaches a Gaussian variable with zero mean and variance $\sigma^2\|\mathbf{w}\|^2$ as $d \to \infty$. In practice, the convergence to the Gaussian distribution is much faster, as illustrated in Fig. 2 where with filter of size 3×3 the empirical distribution of the noise can be well fit with the corresponding Gaussian distribution.

Evaluating Eq. (4) over unit projection directions $\mathbf{w}_1, \cdots, \mathbf{w}_K$ generates K equations with $K + 1$ unknowns, i.e., σ^2 and $\kappa(\mathbf{w}_1^T\mathbf{x}), \cdots, \kappa(\mathbf{w}_K^T\mathbf{x})$, given that projection kurtosis and variances of the noise corrupted vector $\mathbf{y}$ are estimated from the samples. As such, Eq. (4) alone cannot lead to a unique solution to the noise variance, for which we need more information about $\kappa(\mathbf{w}_1^T\mathbf{x}), \cdots, \kappa(\mathbf{w}_K^T\mathbf{x})$.

3.3 Projection Kurtosis Concentration: Empirical Evidence

To this end, we first perform an empirical study on the projection kurtosis of natural images in the band-pass domains. Our experiment is based on 200 images chosen from the Van Hateren database . These images are chosen for their low intrinsic camera noise levels and balanced dynamic ranges without significant over- or under-exposures.

Show in the left panel of Fig. 3 are the projection kurtosis, sorted in a descending order, over bases obtained from different linear transforms. Specifically we use bases from the 2D discrete cosine transform (DCT), the principal component analysis (PCA), the ICA (implemented with the FastICA algorithm Hyvärinen 1999), 2D Haar wavelet (HAAR), and random symmetric bases (RAND), respectively. All bases used in this experiment are of size 8×8 pixels. The random symmetric bases are obtained in the same manner as in Bethge (2006) with *symmetric orthogonalization*. Specifically, from a random matrix $\tilde{V}$ whose elements are independent Gaussian samples of zero-mean and unit variance, random bases are obtained as columns of an orthonormal matrix V, which is generated using as $V = \tilde{V}(\tilde{V}^T\tilde{V})^{-1/2}$. Note that orthogonality of the bases are not essential in these results and similar results can be obtained with over-complete oriented band-pass representations such as the steerable pyramid (Simoncelli and Freeman 1995).

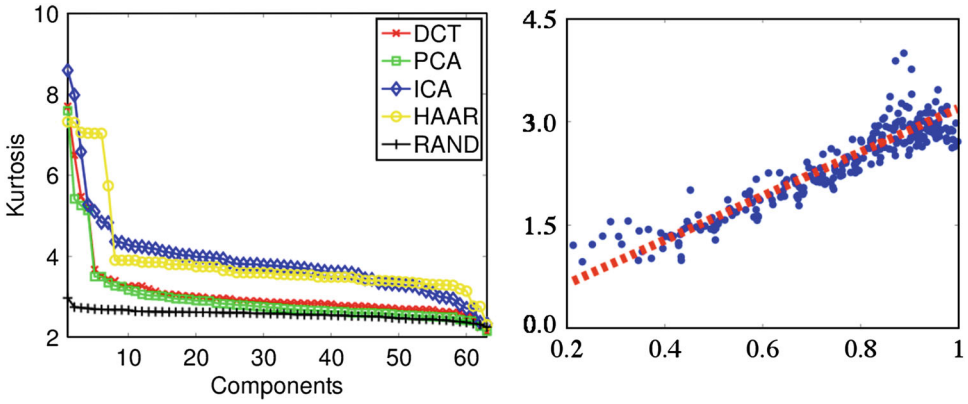

Fig. 3 *Left* Projection kurtosis sorted descendingly for 10, 000 8×8 patches from natural images with projection directions obtained from different linear transforms. *Right* Linear relation between the projection kurtosis of noise corruption of the same set of data (*vertical* axis) and the squared ratio between variances of projected original signal and its noisy corruption (*horizontal* axis). See text for details

As these results show, the projection kurtosis obtained from all types of linear transforms are positive (in comparison, Gaussian noise will have projection kurtosis zero as projection of Gaussian variables are still Gaussian), reflecting the leptokurtic statistics of natural images in these domains (Burt and Adelson 1981; Field 1987). Furthermore, the projection kurtosis obtained with PCA, ICA, DCT and wavelet exhibit relatively large ranges of values. These extreme projection kurtosis values are mainly due to the fact that these representations are designed to better reveal atypical characteristics of natural images. For instance, the top bases from PCA and ICA maximize variance and kurtosis, respectively, and bases in DCT and wavelet have preference over regular spatial frequencies, orientations and scales. Yet, extreme projection kurtosis are relatively rare, with the majority of the projection kurtosis concentrated near to a constant value, indicated by the consistent large stretches of relatively flat regions in the plot. In contrast to the deterministic bases, the projection kurtosis from random bases are more homogeneous. This phenomenon, which in our recent works Pan et al. (2012c) is referred to as *projection kurtosis concentration*, have also been observed in several previous studies (Bethge 2006; Lyu and Simoncelli 2009; Zoran and Weiss 2009).

The concentration tendency of projection kurtosis of natural images can be further corroborated with another experiment, showing that adding Gaussian noise to natural images introduces variations in the projection kurtosis that are dependent on the projection directions. According to Eq. (4), if the noise free signal $\mathbf{x}$ has constant projection kurtosis, there is a linear relation between the projection kurtosis of the noisy signal and the squared ratio of the projection variances of the noise free signal and the noisy signal (the slope in the linear model given by the constant projection kurtosis of the noise free signal). The experimental result shown in the right panel of Fig. 3 is consistent with this prediction. Specifically, this experiment is based on the same image set from the Van Hateren database, with each image corrupted with additive white Gaussian noise ($\sigma = 5$) and analyzed using 300 different random bases. The horizontal coordinates of the blue dots in the plot correspond to the squared ratio of projection variances, $\left(\frac{\sigma^2(\mathbf{w}^T\mathbf{x})}{\sigma^2(\mathbf{w}^T\mathbf{y})}\right)^2$, and the vertical coordinates are the corresponding projection kurtosis, $\kappa(\mathbf{w}^T\mathbf{y})$. The red dash line is the linear model obtained with a least squares fitting. The correlation coefficient of these data is 0.87, suggesting a strong positive linear dependency. Furthermore, the slope of the fitted linear model is very close to the mean projection kurtosis of the noise free image across random bases.

3.4 Projection Kurtosis Concentration: Theoretical Justification

A theoretical justification of the projection kurtosis concentration phenomenon can be obtained from the Gaussian scale mixture (GSM) model of natural images in band-pass domains. Formally, a GSM vector $\mathbf{x} \in \mathcal{R}^d$ with zero mean (Wainwright and Simoncelli 2000) has density function

$$p(\mathbf{x}) = \int_0^\infty \frac{1}{\sqrt{(2\pi z)^d |\det(\Sigma_\mathbf{x})|}} \exp\left(-\frac{\mathbf{x}^T \Sigma_\mathbf{x}^{-1} \mathbf{x}}{2z}\right) p_z(z) dz. \tag{5}$$

where $\Sigma_\mathbf{x}$ is a symmetric and positive definite matrix, and z is a positive random variable (known as the latent scaling factor) with density $p_z(z)$ (Andrews and Mallows 1974). Note that $p(\mathbf{x})$ can be viewed as an infinite mixture of Gaussians with scaled covariance matrices, or equivalently, $\mathbf{x}$ can be understood as the product of two mutually independent random variables as $\mathbf{x} = \mathbf{u} \cdot \sqrt{z}$, where $\mathbf{u}$ is a d-dimensional zero-mean Gaussian vector with covariance matrix $\Sigma_\mathbf{x}$. GSM-based image models have recently received a lot of attention and led to the state-of-the-art performance in image denoising (Portilla et al. 2003).

In the relevance of this work, we present a simple property of the GSM variables that their projection kurtosis is invariant with regards to the projection direction $\mathbf{w}$, which we prove in Appendix 2

Claim 1 *For a GSM random vector* $\mathbf{x}$ *with density function given in Eq. (5) and a unit vector* $\mathbf{w}$*, we have* $\kappa(\mathbf{w}^T\mathbf{x}) =$

$\frac{3\text{var}_z\{z\}}{\mathcal{E}_z\{z\}^2}$, *where* $\mathcal{E}_z\{z\}$ *and* $\text{var}_z\{z\}$ *are the mean and variance of latent variable z, respectively.*

The projection kurtosis concentration of natural sensory signals can thus be explained based on the corresponding GSM models, which have constant projection kurtosis (see Claim 1). As the GSM model accounts for the general statistical properties of natural sensory signals, their similarities could be better brought forth by band-pass representations with no bias or preference on any particular signal structures or characteristics, which explains the more pronounced concentration of projection kurtosis of the random band-pass filter responses.

This result also suggests that when noise is added to a GSM variable, or the local image patches it represents, the projection kurtosis may become deviated from the constant value. let $\mathbf{z}$ be a zero-mean Gaussian noise vector with covariance matrix $\sigma^2 I$ that is independent of $\mathbf{x}$, the projection kurtosis of their sum, $\mathbf{y} = \mathbf{x} + \mathbf{z}$, according to Eq. (4), is given by

$$\kappa(\mathbf{w}^T\mathbf{y}) = \kappa(\mathbf{w}^T\mathbf{x})\left(\frac{\sigma^2(\mathbf{w}^T\mathbf{x})}{\sigma^2(\mathbf{w}^T\mathbf{y})}\right)^2 = \left(\frac{3\cdot\text{var}_z\{z\}}{\mathcal{E}_z\{z\}}\right)\cdot\left(\frac{\mathbf{w}^T\Sigma_\mathbf{x}\mathbf{w}}{\mathbf{w}^T\Sigma_\mathbf{x}\mathbf{w}+\sigma^2}\right)^2. \quad (6)$$

In particular, the last factor in the righthand side of Eq. (6) is invariant to $\mathbf{w}$ only when $\Sigma_\mathbf{x}$ is multiples of the identity matrix, in other words, $\mathbf{x}$ is whitened. Under more general circumstances, $\kappa(\mathbf{w}^T\mathbf{y})$ will be different for different $\mathbf{w}$.

The constant projection kurtosis of a GSM variable and its violation with additive Gaussian noise are illustrated for the 2D case in Fig. 4. In the plot, the projection kurtosis of a two-dimensional GSM variable $\mathbf{x}$ with regards to different projection directions as 2D unit vectors parameterized by their angles are shown as the solid curve. The dashed curve corresponds to the projection kurtosis of $\mathbf{y} = \mathbf{x}+\mathbf{z}$, where $\mathbf{z}$ is a white Gaussian noise. Also shown are the projection kurtosis estimated from 1,000 samples of $\mathbf{x}$ and $\mathbf{y}$ for the same set of projection directions as open and closed circles, respectively. As the results show, for both the theoretical value as calculated in Claim 1 and the estimates from random samples, the projection kurtosis of the 2D GSM variable over 100 uniformly spaced 2D rotation angles between $[0, 2\pi)$ resemble a circle, indicating their invariance to different projection directions. On the other hand, the shape of the curve corresponding to the values of the projection kurtosis of the noisy signal suggests that they change with projection directions.

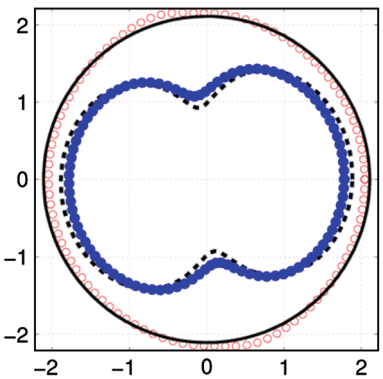

Fig. 4 Plot of projection kurtosis as a function of 2D projection directions (angles) for a GSM vector $\mathbf{x}$ (theoretical value as *solid curve* and sample estimates as *open circles*) and $\mathbf{y} = \mathbf{x} + \mathbf{z}$, where $\mathbf{z}$ is a white Gaussian noise (theoretical value as *dashed curve* and sample estimates as *filled circles*). See texts for more details

4 Blind Global Noise Estimation

We now describe a noise estimation method based on projection kurtosis concentration property. We first describe our method under the Gaussian noise model (Sect. 4.1), and discuss its difference with the work of Zoran and Weiss (2009) (Sect. 4.2). This method is then extended to non-Gaussian and multiplicative noise in Sect. 4.3 and experimental evaluations are reported in Sect. 4.4.

4.1 Estimating Additive Gaussian Noises

Formally, denote $\mathbf{y} = \mathbf{x} + \mathbf{z}$ as the result of contaminating a natural image $\mathbf{x}$ with white Gaussian noise $\mathbf{z}$ of zero mean and unknown variance σ^2, our goal is to estimate σ^2 from the noise corrupted image $\mathbf{y}$, as it is the sole parameter specifying the Gaussian model of $\mathbf{z}$.

To take advantage of projection kurtosis concentration, we project $\mathbf{y}$ into K band-pass channels using K different filters of unit ℓ_2 norm, and denote the kurtosis and variance of the responses of the original and the noisy image in the k^{th} channel as κ_k and $\tilde{\kappa}_k$, respectively. We further denote the variances of the responses of the original and the noisy image in the k^{th} channel as σ_k^2 and $\tilde{\sigma}_k^2$, respectively. According to Eq. (4), these statistics are related as:

$$\tilde{\kappa}_k = \kappa_k\left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2}\right)^2. \quad (7)$$

Due to projection kurtosis concentration, we can approximate the projection kurtosis of the noise-free natural image $\mathbf{x}$ across the K band-pass channels with a constant, or

$$\tilde{\kappa}_k \approx \kappa\left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2}\right)^2. \quad (8)$$

We further note that (1) the band-pass filter responses of natural images tend to have super-Gaussian marginal distributions (Burt and Adelson 1981) with positive kurtosis values (i.e., $\kappa > 0$), and (2) $\sigma_k^2 = \tilde{\sigma}_k^2 - \sigma^2 > 0$. Therefore, we can take square root on both sides of Eq. (7), to obtain

$$\sqrt{\tilde{\kappa}_k} \approx \sqrt{\kappa}\left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2}\right). \tag{9}$$

Equation (9) implies a simple scheme to estimate σ^2: with two different projection directions, $\mathbf{w}_i$ and $\mathbf{w}_j$, one can cancel out the common factor involving $\sqrt{\kappa}$, to obtain

$$\frac{\sqrt{\tilde{\kappa}_i}}{\sqrt{\tilde{\kappa}_j}} = \frac{\tilde{\sigma}_j^2}{\tilde{\sigma}_i^2}\left(\frac{\tilde{\sigma}_i^2 - \sigma^2}{\tilde{\sigma}_j^2 - \sigma^2}\right) \text{ or } \sigma^2 = \frac{\tilde{\sigma}_i^2\tilde{\sigma}_j^2(\sqrt{\tilde{\kappa}_i} - \sqrt{\tilde{\kappa}_j})}{\tilde{\sigma}_i^2\sqrt{\tilde{\kappa}_i} - \tilde{\sigma}_j^2\sqrt{\tilde{\kappa}_j}}.$$

However, projection kurtosis across different band-pass channels are usually not strictly constant in real natural sensory signals, and the estimations of $\tilde{\kappa}_k$ and $\tilde{\sigma}_k^2$ fluctuate due to the sampling effect. These factors can result in no solution to σ^2 and κ that satisfy all constraints in Eq. (8), and hence the simple method cannot be used to reliably estimate σ^2.

On the other hand, we can formulate variance estimation as an optimization problem to minimize the squared difference between the two sides of Eq. (8), as:

$$L\left(\sqrt{\kappa}, \sigma^2\right) = \sum_{k=1}^{K}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa}\left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2}\right)\right)^2. \tag{10}$$

It turns out that the optimal solution minimizing $L\left(\sqrt{\kappa}, \sigma^2\right)$ is unique, and affords a closed-form solution, which, if we denote the average over the K band-pass channels as $\langle\cdot\rangle_k$ and $(x)_+ = \max(x, 0)$, has the following form:

$$\sqrt{\kappa} = \frac{\langle\sqrt{\tilde{\kappa}_k}\rangle_k\left\langle\frac{1}{(\tilde{\sigma}_k^2)^2}\right\rangle_k - \left\langle\frac{\sqrt{\tilde{\kappa}_k}}{\tilde{\sigma}_k^2}\right\rangle_k\left\langle\frac{1}{\tilde{\sigma}_k^2}\right\rangle_k}{\left\langle\frac{1}{(\tilde{\sigma}_k^2)^2}\right\rangle_k - \left\langle\frac{1}{\tilde{\sigma}_k^2}\right\rangle_k^2},$$

$$\sigma^2 = \frac{1}{\left\langle\frac{1}{\tilde{\sigma}_k^2}\right\rangle_k}\left(1 - \frac{\langle\sqrt{\tilde{\kappa}_k}\rangle_k}{\sqrt{\kappa}}\right)_+. \tag{11}$$

The non-negativity of the estimated $\sqrt{\kappa}$ is ensured by the Cauchy-Schwartz inequality. Note that σ^2 is estimated from the harmonic mean of the projection variances of the noisy signal, modulated by a factor determined by the projection kurtosis. Compared to the arithmetic mean, the harmonic mean can mitigate the influence of large outliers and increase the influence of small values, and can improve the robustness of the estimator in the presence of large outliers. A detailed derivation of Eq. (11) can be found in Appendix 3

The choice of the front-end linear band-pass filters in implementing Eq. (11) is important to the estimation performance, as we would like to use a representation where the resulting projection kurtosis are more likely to exhibit strong concentration behaviors. In the preliminary version of our work (Pan et al. 2012c), we have tested using the 2D DCT AC filters. However, as have been shown in Sect. 3.3, it seems that the projection kurtosis of natural images show a much stronger tendency to concentrate around a constant when analyzed in channels from random band-pass filters. The advantage of using random band-pass filters with the solution given in Eq. (11) will be demonstrated experimentally in Sect. 4.4.

4.2 Comparison with Method of Zoran and Weiss (2009)

Based on the empirically observed projection kurtosis concentration in the DCT domain, Zoran and Weiss (2009) estimated the noise variance by minimizing a different objective function corresponding to the squared difference of the two sides of Eq. (8), as:

$$L_{\text{Z\&W}}\left(\sqrt{\kappa}, \sigma^2\right) = \sum_{k=1}^{K}\left[\tilde{\kappa}_k - \kappa\left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2}\right)^2\right]^2. \tag{12}$$

They showed that numerically minimizing this objective function leads to significant improvements over the state-of-the-art methods. We would like to point out three significant differences between this and that of Zoran and Weiss (2009):

1. Projection kurtosis concentration was described as an empirical observation in Zoran and Weiss (2009), while in our work, we provide an extensive empirical study (Sect. 3.3) and a theoretical justification (Sect. 3.4). Our analyses also suggest that random band-pass linear transformations can lead to better projection kurtosis concentration than deterministic representations based on DCT.
2. The objective function Eq. (12) does not seem to have a closed-form—the optimal solution requires solving two cubic equations simultaneously, and in practice has to be optimized numerically. In contrast, the closed-form solution of our method makes it more efficient and facilitates its extension to blind local noise estimation (Sect. 5).
3. Projection kurtosis concentration does not exclude the possibility of outlying projection kurtosis values that are significantly different from their mean value across different band-pass filtered channels (c.f. Sect. 3.3 and the left plot in Fig. 3). These outliers can have a more significant effect on Eq. (12) than Eq. (10). This can be better seen if we denote the two sides of Eq. (8) as a and b, we have $L\left(\sqrt{\kappa}, \sigma^2\right) = (\sqrt{a} - \sqrt{b})^2 \leq |a - b|$ for any $a, b \geq 0$, showing that $L\left(\sqrt{\kappa}, \sigma^2\right)$ is dominated by the ℓ_1 loss, which is more robust to large outliers than the ℓ_2 loss used in $L_{\text{Z\&W}}\left(\sqrt{\kappa}, \sigma^2\right) = (a - b)^2$.

4.3 Non-Gaussian and Multiplicative Noises

Because our method operates in the band-pass domains, the Gaussian assumption on the noise in the pixel domain can be significantly relaxed, because any type of i.i.d. noise in the pixel domain will be similar to Gaussian noise after being

linearly mixed by the filters, a direct result of the central limit theorem (Feller 1968). Therefore, the algorithms based on Eq. (11) can also be used to estimate the variance of additive non-Gaussian i.i.d. noise in the pixel domain, and if the parameters in the non-Gaussian noise model can be derived from the variance, it can be determined with our algorithm. For instance, if the noise has a zero-mean Laplacian model, $p(z) \propto e^{-|\frac{z}{s}|}$, determined with the scale parameter s, it can be estimated using the relation that $s = \sqrt{\sigma^2}$.

Furthermore, this method can also be used to estimate certain multiplicative noise, under which the i.i.d. noise $\mathbf{z}$ is multiplied by (not added to) the original image $\mathbf{x}$ to generate a noise corrupted image as

$$\mathbf{y} = \mathbf{x} \circ \mathbf{z}, \tag{13}$$

where $\circ$ is the point-wise multiplication (Dirac product), assuming positivity of components of $\mathbf{x}$ and $\mathbf{z}$. Multiplicative noise models are more appropriate for realistic image noises in certain signal-to-noise ratio range (Withagen et al. 2005; Liu et al. 2008), and special purpose imaging systems such as ultrasound, radar, sonar (SAS), and laser imaging (Rudin et al. 2003). A common model of multiplicative noise is the Gamma law (Farebrother 1990),

$$p(z) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z}, \text{ for } z > 0, \tag{14}$$

where $\alpha > 0$ and $\beta > 0$ are the the shape and rate parameters, respectively, and $\Gamma(\cdot)$ is the standard Gamma function. Given an image, the scale of multiplicative noise is underdetermined, so it is conventionally assumed without loss of generality that $\mathcal{E}_z\{\log z\} = 0$.

To estimate multiplicative i.i.d. Gamma noise, we first convert it to an additive model by taking logarithm of Eq. (13) to obtain

$$\underbrace{\log \mathbf{y}}_{\tilde{\mathbf{y}}} = \underbrace{\log \mathbf{x}}_{\tilde{\mathbf{x}}} + \underbrace{\log \mathbf{z}}_{\tilde{\mathbf{z}}}. \tag{15}$$

The resulting noise process $\tilde{\mathbf{z}}$ are still i.i.d., and log transformed natural images also have the projection kurtosis concentration property (Bethge 2006). We can then proceed with the algorithm given in Eq. (11) to estimate the variance of $\tilde{\mathbf{z}}$, σ^2, from which, we can further determine α and β using Farebrother (1990):

$$\begin{aligned} &\mathcal{E}_{\tilde{z}}\{\tilde{z}\} = \mathcal{E}_z\{\log z\} = \Psi(\alpha) - \log\beta \\ &\mathrm{var}_{\tilde{z}}\{\tilde{z}\} = \mathcal{E}_z\left\{(\log z)^2\right\} = \Psi_1(\alpha), \end{aligned} \tag{16}$$

where $\Psi(x) = \frac{d}{dx}\Gamma(x)$ and $\Psi_1(x) = \frac{d^2}{dx^2}\Gamma(x)$ are the digamma function trigamma functions, respectively. With the assumption that $\mathcal{E}_{\tilde{z}}\{\tilde{z}\} = 0$, we further have $\alpha = \Psi_1^{-1}(\sigma^2)$ and $\beta = e^{\Psi(\alpha)}$, where both the digamma function and the inverse trigamma function are evaluated numerically.

4.4 Experimental Evaluations

We test the performance of the blind global noise variance estimation method using three sets of grayscale images in uncompressed `PNG` or `TIFF` formats:

1. 25 8-bit images of the Kodak database (Franzen 1999) (KD),
2. 200 8-bit images chosen from the UCID image set (Schaefer and Stich 2004) (UD),
3. 200 16-bit images chosen from the Van Hateren image set (van Hateren and van der Schaaf 1998) (VH).

These images have low intrinsic noise levels and balanced dynamic ranges. To further reduce the effect of intrinsic noise, we perform gentle low-pass filtering to the these images to generate the test images.[5]

4.4.1 Gaussian Noise

In the first set of experiments, we create noisy images by adding white Gaussian noise of different variances to the test images. To accommodate the different value ranges of the test images, we use the peak-signal-to-noise ratio (PSNR), which is defined as $10\log_{10}\left(\frac{I_{\max}^2}{\sigma^2}\right)$ (in the unit of dB), where $I_{\max}$ is the maximum pixel value (255 for 8-bit images and 65,536 for 16-bit images), and σ^2 is the estimated noise variance. Shown in Table 1 are the estimation performances of our algorithm, shown as the mean estimated PSNRs with the standard deviations (in parenthesis). These results are obtained with 63 random band-pass filters of size 8×8 created a priori.

By way of comparison, we also include performances of three state-of-the-art noise estimation methods applied to the same set of noisy images,

1. the method of Liu et al. (2012), which is based on identifying local patches with no significant image structures to collectively estimate the variance of Gaussian noise (corresponding to rows labeled as LTO),
2. the method of Zoran and Weiss (2009), which uses projection kurtosis from the 2D DCT decomposition and objective function Eq. (12) (corresponding to rows labeled as Z&W),
3. the widely used method based on the MAD estimator in the highest frequency diagonal subband of a 2D Haar

[5] We evaluated the noise levels of the raw images from the three image sets using our method, though there is no ground truth to compare. Images from the Van Hateren database have significantly lower noise levels (averaged noise standard deviation 0.25) due to its higher bit-depth and quality, while images from UCID and Kodak data bases have average estimated noise levels around 0.44 and 0.78, respectively.

Table 1 Comparison of noise estimation performances for additive Gaussian noises as the average PSNRs (standard deviations shown in parenthesis)

PSNR		50 dB	40 dB	30 dB	25 dB	20 dB
KD	Ours	**49.61** (0.82)	**40.08** (0.61)	**30.16** (0.53)	25.14 (0.24)	20.07 (0.12)
	LTO	47.33 (1.22)	39.26 (0.73)	30.33 (0.58)	**25.04** (0.24)	**20.03** (0.12)
	Z&W	45.64 (2.01)	37.53 (1.22)	30.54 (0.95)	25.38 (0.55)	20.29 (0.32)
	MAD	39.70 (1.24)	36.08 (0.97)	28.97 (0.72)	24.55 (0.36)	19.83 (0.15)
UD	Ours	49.53 (1.68)	**39.76** (1.43)	30.14 (0.41)	25.12 (0.11)	20.02 (0.03)
	LTO	**49.64** (1.97)	39.55 (1.94)	**30.02** (0.37)	**25.01** (0.09)	**20.00** (0.05)
	Z&W	46.45 (2.04)	39.11 (2.31)	30.32 (0.61)	25.24 (0.28)	20.07 (0.09)
	MAD	45.34 (1.92)	38.44 (1.19)	29.35 (0.47)	25.76 (0.89)	19.99 (0.01)
VH	Ours	**49.70** (1.07)	**40.46** (0.96)	30.10 (0.21)	25.05 (0.09)	20.03 (0.04)
	LTO	48.53 (2.08)	40.85 (1.34)	**30.01** (0.17)	**25.00** (0.07)	**20.00** (0.04)
	Z&W	45.91 (2.20)	39.17 (1.31)	30.19 (0.48)	25.12 (0.33)	20.10 (0.22)
	MAD	45.13 (1.84)	38.75 (0.80)	29.73 (0.26)	24.89 (0.12)	19.96 (0.05)

Bold faces correspond to mean performances closest to the ground truth

wavelet decomposition Donoho (1995) (corresponding to rows labeled as MAD).

These results that on estimating Gaussian noises, the performances of our method are comparable with those of Liu et al. (2012), the current state-of-the-art. We note that the method of Liu et al. (2012) relies more on a stationary Gaussian noise model, in contrast, our method can be extended to non-Gaussian and non-stationary noise models as shall be shown in the following. Compared with the method of Zoran and Weiss (2009), the improvement achieved with our method is significant, particularly for low noise levels (corresponding to high PSNRs). This is likely due to the difference in the objective functions as discussed in Sect. 4.2.

In terms of running time, for an 800 × 600 image, our method and the method of Liu et al. (2012) take on average about 3 s, the method of Zoran and Weiss (2009) takes about 10 s, whereas it takes the MAD estimator about 0.5 s.[6] Yet, it should also be pointed out that most running time of our method is spent on the convolutions with the 63 filters, where the MAD estimator essentially uses only two filters. However, when reduced to two random band-pass filters, the proposed blind noise variance estimation method achieves comparable running time and a slightly better performance over the MAD estimator.

We also study the effect of different choice of linear band-pass filters in our noise estimation method. We compare results obtained with 63 random symmetric random band-pass filters of size 8 × 8 and 63 AC filters from the 2D DCT decomposition of the same size. Shown in Table 2 are the performances of using random band-pass filters (row RND) and DCT filters (row DCT) on the Van Hateren image set for additive Gaussian noises. These results suggest that using random band-pass filters can lead to better average performances over the deterministic DCT filters, albeit with higher variations in the estimations.

4.4.2 Non-Gaussian and Multiplicative Noise

In the next set of experiments, we test our method to estimate statistics of (1) additive zero mean Laplacian noise (determined by the scale parameter s) and (2) multiplicative Gamma noise specified with parameter α.[7] Random noises of the two types are applied to 100 images from the Van Hateren set, to which our method is applied to estimate the noise parameters. For the Laplacian noise, estimation of s is obtained using the square root of the estimated noise variance. For the Gamma noise, estimations of α and β are obtained from the estimated noise variance of the logarithm transformed image using Eq. (16) with numerical evaluation of the di-gamma and tri-gamma functions, see Sect. 4.3. We present the experimental results in Tables 3 and 4.

We also applied the three other methods as comparisons. In both cases, estimation methods that have a stronger reliance on the Gaussian noise model, i.e., Liu et al. (2012) and Donoho (1995), perform worse compared to the kurtosis based methods, with our method outperforming the method of Zoran and Weiss (2009) in general.

[6] All results are based on unoptimized MATLAB code running on a machine of 2.4 GHz and 4 GB RAM. This improves on our early results (Pan et al. 2012c), as we use separable random filters, and can be implemented as two consecutive 1D convolutions. This gives it an advantage in running time compared to the 2D random filters that have to be implemented as one 2D convolution step.

[7] β can be computed from α assuming the log noise has mean zero, c.f. Eq. (16).

Table 2 Comparison of noise estimation performances using random band-pass filters (RND) and deterministic DCT filters (DCT)

PSNR	50 dB	40 dB	30 dB	25 dB	20 dB
RND	49.70 (1.07)	40.46 (0.96)	30.10 (0.21)	25.05 (0.09)	20.03 (0.04)
DCT	49.64 (0.84)	40.98 (0.38)	30.12 (0.08)	25.06 (0.03)	20.04 (0.01)

Table 3 Comparison of noise estimation performances for additive Laplacian noises as the average estimation of scale parameter (standard deviations shown in parenthesis)

	s = 1	s = 5	s = 10	s = 25
Ours	**1.07** (0.72)	**5.12** (1.38)	**10.29** (1.42)	**25.44** (3.24)
LTO	2.44 (1.23)	3.42 (2.61)	9.03 (2.82)	23.15 (1.21)
Z&W	1.94 (1.02)	5.42 (1.68)	10.86 (1.75)	26.95 (1.21)
MAD	3.68 (0.92)	7.84 (1.94)	14.62 (2.63)	21.35 (4.58)

Bold faces correspond to mean performances closest to the ground truth

Table 4 Comparison of noise estimation performances for multiplicative Gamma noises as the average estimation of parameter α (standard deviations shown in parenthesis)

	$\alpha = 1$	$\alpha = 5$	$\alpha = 10$	$\alpha = 25$
Ours	**1.45 (1.32)**	**6.16 (2.04)**	**11.97 (3.21)**	**28.44 (4.52)**
LTO	5.21 (3.19)	8.32 (2.89)	18.37 (7.06)	30.22 (8.23)
Z&W	2.22 (1.08)	7.03 (3.25)	13.88 (6.39)	29.72 (5.77)
MAD	5.49 (4.25)	8.28 (4.70)	19.38 (4.46)	30.32 (3.48)

Bold faces correspond to mean performances closest to the ground truth

5 Local Noise Variance Estimation

In this section, we describe an efficient blind local noise estimation method based on the estimator developed in the previous section. The goal is to obtain noise variance $\sigma^2(i, j)$ at each pixel location (i, j) using statistics collected from corresponding rectangular windows in the K band-pass channels, $\Omega^k_{(i,j)}$. The size of the window controls the trade-off between the precision and variance of the estimation. In general, smaller window sizes lead to finer spatial resolution, but the estimated statistics may suffer from higher variance due to the smaller number of samples. Estimations based on larger window sizes are more stable, but rapid changes in the underlying statistics will not be captured precisely.

Though in principle any global noise estimation method can be used by simp lying applying it locally, we take advantage of the closed-form estimator developed in the previous section for a more efficient non-iterative solution. Note that both variance and kurtosis can be computed from the raw (un-centered) moments, Eq. (2), which is estimated in the local window with sample averages, as:

$$\mu_m\left(\Omega^k_{(i,j)}\right) \approx \frac{1}{|\Omega^k_{(i,j)}|} \sum_{(i',j')\in\Omega^k_{(i,j)}} x(i', j', k)^m,$$

where we denote $x(i', j', k)$ the response at (i', j') in the kth band-pass channel. A straightforward implementation then uses the estimated local statistics to compute kurtosis and variance and then evaluate Eq. (11) for each local window $\Omega^k_{(i,j)}$. This leads to an overall running time of $\mathcal{O}(MNK)$, where N and M are the sizes of the image and local windows, respectively, and K is the number of band-pass channel used. We will show that this implementation can be accelerated to a running time of $\mathcal{O}(NK)$ using *integral image* (Viola and Jones 2002).

Integral image (also known as sum area table Crow 1984) is a data structure for efficient computation of sum values in rectangular regions in an image (or in our case, one channel in a band-pass filtered domain). In particular, denote the integral image constructed from an image $\mathbf{x}$ as $\mathcal{I}(\mathbf{x})$, each pixel of $\mathcal{I}(\mathbf{x})$ corresponds to the sum of all pixels of $\mathbf{x}$ in the rectangular region defined by $[1, i] \times [1, j]$. The integral image can be efficiently constructed in linear time of the dimension of $\mathbf{x}$ (Viola and Jones 2002). Summation in any rectangular window specified with $[i, i + I] \times [j, j + J]$ in $\mathbf{x}$ can then be evaluated with just three addition/subtraction operations on the corresponding integral image, as:

$$\mathcal{I}(\mathbf{x})_{i+I,j+J} - \mathcal{I}(\mathbf{x})_{i,j+J} - \mathcal{I}(\mathbf{x})_{i+I,j} + \mathcal{I}(\mathbf{x})_{i,j}.$$

In particular, the mth order raw moment of the rectangular window $[i, i + I] \times [j, j + J]$ can be computed as

$$\frac{1}{IJ}\Big[\mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m\text{ times}})_{i+I,j+J} - \mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m\text{ times}})_{i,j+J} - \mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m\text{ times}})_{i+I,j} + \mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m\text{ times}})_{i,j}\Big], \quad (17)$$

where $\circ$ is the point-wise multiplication (Dirac product). Subsequently, the local statistics of all overlapping windows in a subbband can be computed in time that is agnostic to the local window sizes, which also facilitates efficient selection of the size of local windows for an optimal tradeoff of estimation precision (large windows) and location accuracy (small windows).

We summarize the basic steps of our local noise estimation algorithm as the follows.

1. Decompose the image into K band-pass filtered channels using AC filters from the DCT decomposition;

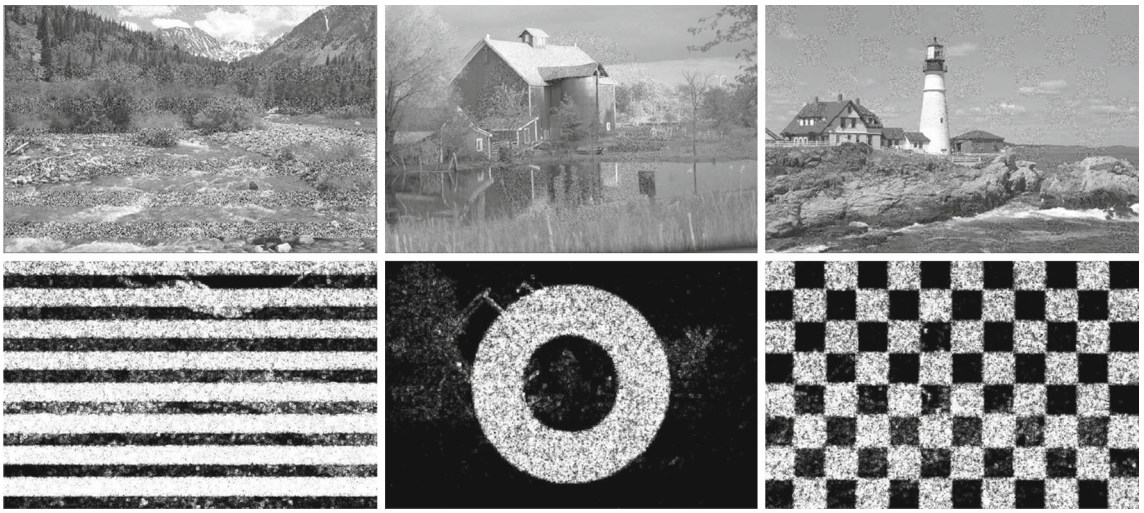

Fig. 5 Local noise variance estimation for three images with different additive white Gaussian noise patterns

2. Compute the integral images of the first to the fourth order raw moments in each of the K band-pass filtered channels with Eq. (17);
3. Compute variance and kurtosis for each local window in each band-pass filtered channel with Eq. (2);
4. Estimate noise variance by evaluating Eq. (11) for each local window across all band-pass filtered channels.

5.1 Experimental Evaluation

The performances of the local noise variance estimator are evaluated on a set of noisy images generated from the Kodak image set, on each of which spatially varying Gaussian noises are added. In particular, we use three types of spatial patterns for the additive noise, (1) horizontal stripes of Gaussian noises of uniformly increasing variances from $\sigma = 10$ (top) to $\sigma = 20$ (bottom), (2) Gaussian noises of $\sigma = 20$ with annular patterns and (3) Gaussian noises of $\sigma = 20$ with checkerboard patterns. Three examples of the noise corrupted images are shown in the first row of Fig. 5. The implemented algorithm uses 63 8×8 random band-pass filters, with 5×5 local windows.

The second row of Fig. 5 shows the estimated local noise variances of each corresponding image. These results suggest that our method can recover the spatial patterns of these locally varying noises. On the other hand, it should also be noted that different local noise variances intrinsic to the original image are also detected, which are most clearly visible for the middle image that has extensive smooth regions (e.g., sky) and textures (e.g., trees).

We also perform a qualitative evaluation of our method. Table 5 shows the average and standard deviation (in parenthesis) of the root mean squared difference between the estimated noise standard deviations and the ground truth values averaged over 100 images generated for each noise pattern. As a comparison, in the same table, performances of a local noise estimation method based on the MAD estimator (Mahdian and Saic 2009) are also reported. As these results show, our method achieves a significant margin in estimation performance over the alternative method. Last, the dynamic programming implementation also leads to a significant speedup—analyzing an image of size 600×800 pixels takes 8 s, while the algorithm without dynamic programming takes more than 600 s.[8]

Table 5 Quantitative evaluation with RMSE of our local noise estimation method on three different data sets

	Stripes	Annulus	Checkerboard
Our method	**3.17 (4.22)**	**6.18 (5.23)**	**4.12 (4.46)**
Method of Mahdian and Saic (2009)	7.69 (10.25)	12.76 (7.74)	9.89 (6.82)

The reported performances are the average and standard deviation (in parenthesis) of the RMSE averaged over 100 images generated for each noise pattern
Bold values indicate better performance

6 Region Splicing Detection

We use the blind local noise estimation method developed in the previous section to detect region splicing. Assuming untampered images have spatially homogeneous noise statistics, a composite image with regions from other images with different noise characteristics can be exposed by the inconsistencies of local noise statistics.

[8] All results are based on unoptimized MATLAB code running on a machine of 2.4 GHz and 4GB RAM.

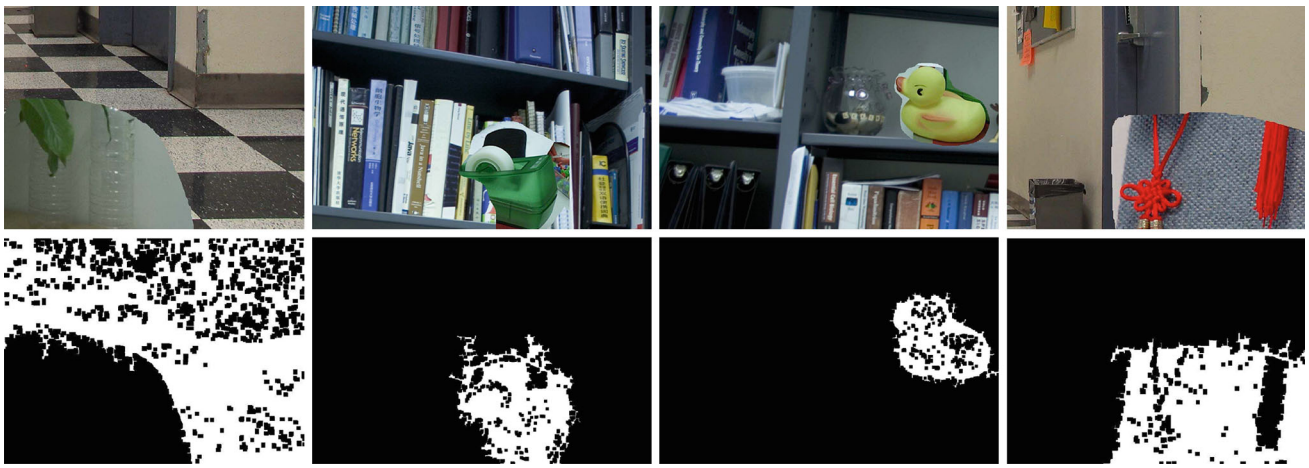

Fig. 6 Examples of region splicing detection results for forged images from the Columbia Uncompressed Image Splicing Detection Evaluation Dataset (Hsu and Chang 2006). See text for details

To this end, we first apply the blind local noise estimation method to compute the local variances at all pixel locations. Specifically, these local noise variances are the mean of the estimations obtained with local window of different sizes ($n \times n$ with n taking 2, 4, 8 10 and 12), based on 63 random band-pass 8×8 filters. The use of averages over different window sizes is to accommodate the precisions in estimating statistics (larger windows) and recovering changes in local statistics (smaller windows). We then segment the estimated local noise statistics into distinct regions with significantly different values and remove small isolated regions. Regions with similar noise levels are further connected and merged using mathematical morphological operations (Serra 1988).

It should be pointed out that due to the complex nature of real camera noise, the estimations obtained from our algorithm are usually approximations to the actual of local noise variances. However, for the purpose of detecting region splicing, it is sufficient to expose significant *inconsistencies* in the local noise variances, which is what our algorithm is used for.

6.1 Qualitative Evaluations and Comparisons

We first test our region splicing detection method on 180 forged images from the Columbia Uncompressed Image Splicing Detection Evaluation Dataset (Hsu and Chang 2006). These forged images were created from digital photographs captured with a wide range of camera models and different lighting conditions (indoor, outdoor, cityscape and natural environment). The spliced regions are manually selected from the source images and randomly pasted into different images to create the forgeries. The forged images are in high resolution and uncompressed TIFF format. Several examples of the forged images, together with the detection results using our method, are shown in Fig. 6. We further apply our method to another set of forged images with spliced regions, several examples, which are submissions to the contest of realistic image manipulation on the web site http://www.Worth1000.com. and the corresponding detections of which are shown in Fig. 7. Compared to the images from the Columbia data set, these spliced regions in these images are better merged with the original image and more challenging to detect.

In general, these results suggest that inconsistencies of local noise variances, which can be revealed effectively with our method, are useful in exposing spliced regions. However, these results also suggest some limitations of all noise based splicing detection method. In particular, the basic assumption that local noise variances are consistent within image regions of the same origin may not be true, this is particularly the case for images with a lot of small scale textures. We find that such is a major cause of false detections and miss detections. In particular, when applied to the 180 original images from the Columbia data set, our method found spliced regions of sufficient sizes in 5 images, mainly due to the variation of local noise variances within these untampered images.

We also compare the performance of our method with several existing region splicing detection methods that can automatically locate the spliced region in an image.[9] We first compare with two noise-based region splicing detection methods (Mahdian and Saic 2009; Pan et al. 2011). The former is based on the MAD estimator of local noise variances in the high-pass Haar wavelet subband. The latter directly applies the global noise variance estimation method in Zoran and Weiss (2009) to local pixel patches. The original algorithms in both

[9] This precludes methods that only classify whole image as containing spliced regions (e.g., Bayram et al. 2006; Fu et al. 2007; Ng and Chang 2004), or require initial user input for possible spliced regions (Popescu and Farid 2004; Hsu and Chang 2006; Lin et al. 2005, or predicate on more detailed knowledge of the imaging processes (Chen et al. 2007).

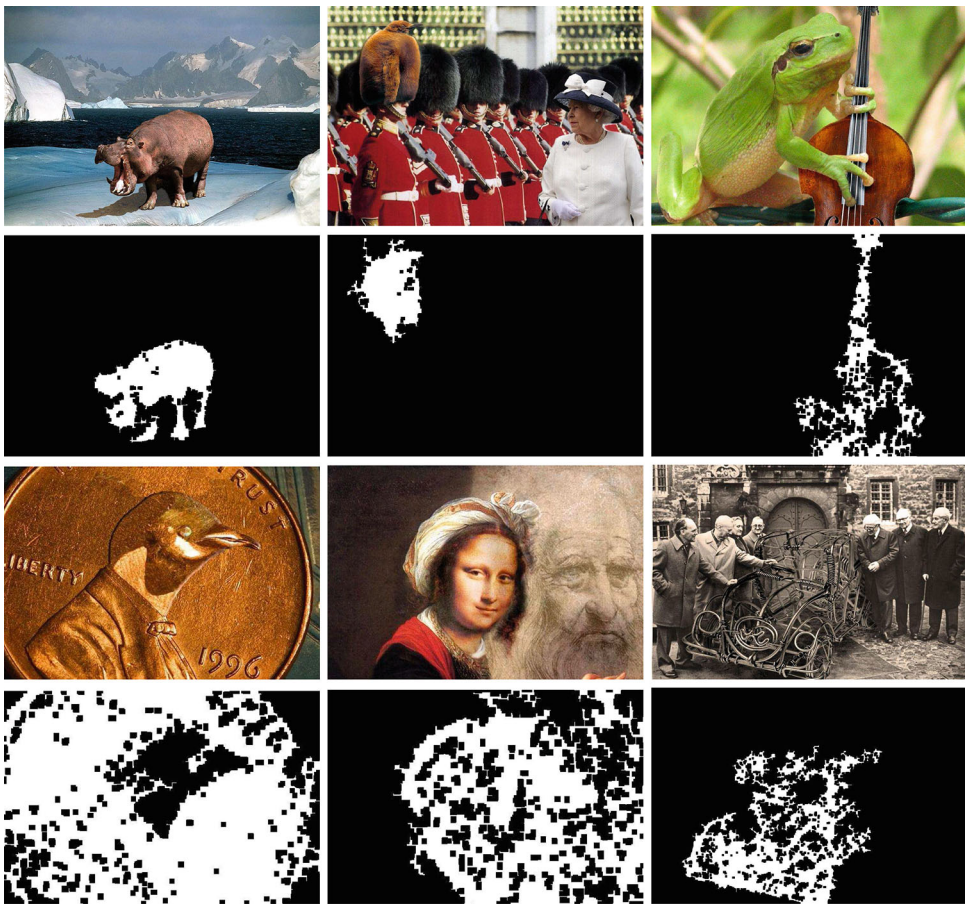

Fig. 7 Detection results on splicing image forgeries from *Worth1000.com*. Images are used with the courtesy and permission of the original contributors

methods examine only non-overlapping blocks (40×40 and 64×64, respectively).

We created two tampered images with the aid of `Adobe Photoshop` using region splicing. The top three rows of Fig. 8 show the original images contributing the background, the original images contributing the spliced regions, and the tampered images, respectively. The original images (in uncompressed TIFF format) are obtained from Flickr.com, and they are taken with different cameras: the original images in the top row were taken with a Canon 400D camera and a Canon Digital Rebel 100 camera, respectively. The two original images in the second row were taken with a Canon EOS-60D digital camera and a SONY DSC-H20 digital camera, respectively. The differences in sensor types, exposure, and ISO speeds in these images are the main causes of the different noise variances in the forged images. Comparisons of detected spliced regions are shown in the last two rows of Fig. 8. For the two comparable methods, the settings were identical as given in each paper. Though all three methods seem to locate the spliced region, our method provides a higher level of accuracy because the region is determined to the pixel level.

We further compare with the detection method of Lin et al. (2009), which exploits some particular properties of tampered images in the JPEG format. Specifically, if we assume the tampered JPEG image is a composite of regions from multiple different original JPEG images, each of which has different JPEG compression qualities, then during its generation, most of the image blocks undergo at least twice of JPEG compression, one from the original image and the other from the process of saving the tampered image in JPEG format. It is known that double JPEG compressed images have distinct properties in the histograms of DCT coefficients (Popescu and Farid 2004).

We compare the detection results of our method and the method of Lin et al. (2009), which exposes spliced regions by detecting each 8×8 image blocks for double JPEG characteristics, on two example images (Fig. 9). As the results show, both methods seem to be able to identify similar tampered regions, but our method fails to detect the entire tampered regions. One cause may be the low JPEG quality (Q = 65) of these images. Heavy JPEG compression affects most the high-frequency bands of the image, where the noise resides. As such, the difference in noise levels between the original images and the spliced regions are also reduced.

On the other hand, the detection of Lin et al. (2009) is only to the level of 8×8 JPEG blocks, while our method can provide more accurate demarcation of the spliced region. In addition, we would like to point out that for the method of Lin et al. (2009) to be effective, several restrictive assumptions

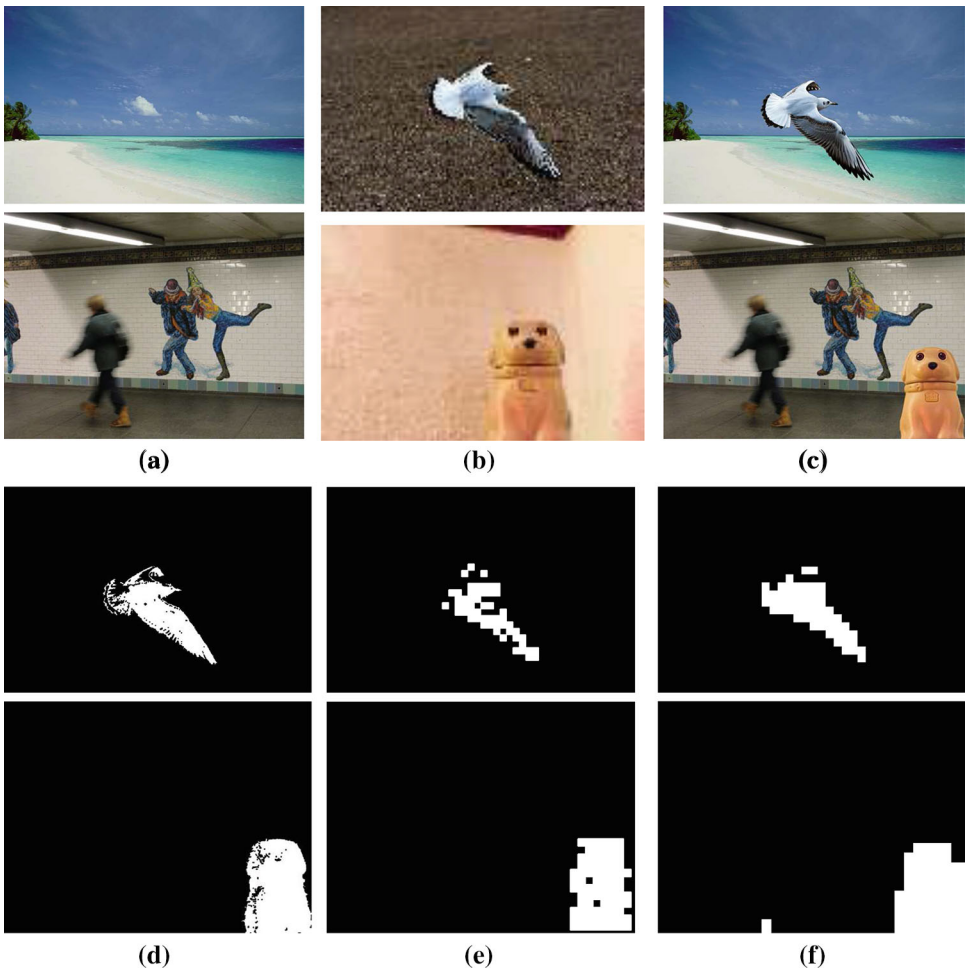

Fig. 8 (**a**,**b**) Two pairs of original JPEG images from Flickr.com. (**c**) Forgeries generated with image splicing using the original images with `Adobe Photoshop`. (**d**) Splicing detection results using our method. (**e**) Splicing detection results using the method in Pan et al. (2011). (**f**) Splicing detection results using the method in Mahdian and Saic (2009)

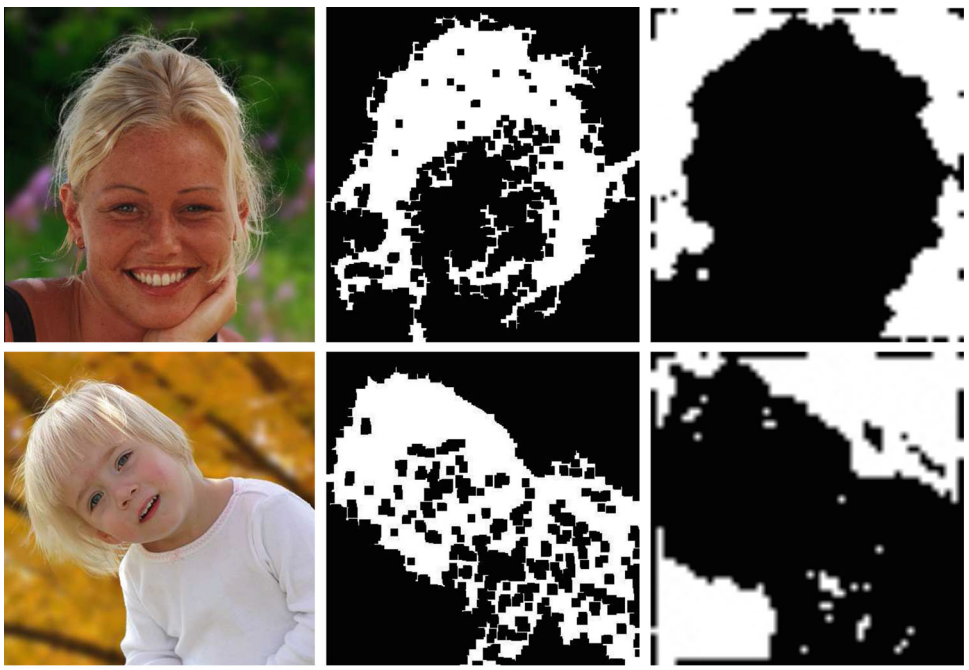

Fig. 9 Comparison of splicing detection results of our method with the method in Lin et al. (2009). The forged images and detection results are provided with permission by the authors of Lin et al. (2009)

have to be met: (1) the tampered image must be in JPEG format, (2) the image contributing the background must be in JPEG format and must have a different quality factor from the tampered image, and (3) the spliced region has to exhibit no double JPEG artifact (Lin et al. 2009). In contrast, noise-based detection methods are based on more general image statistics and can be applied to forged images created from original images of different formats.

6.2 Quantitative Evaluations and Sensitivity Analyses

We perform a quantitative evaluation of our region splicing detection method. We use 100 untampered images from the UCID database (Schaefer and Stich 2004). We generate a forged image by (1) randomly choosing a pair of untampered images, (2) randomly selecting a rectangular region from one image, (3) adding white Gaussian noise of a given level to the selected region, and (4) pasting it into a random region in the other image. This simple process aims to simulate the process of region splicing, while the use of rectangular region is to facilitate calculating the qualitative performance measures. We quantify the detection accuracy with the fraction of pixels in tampered regions that are correctly identified, and define the false positive rate as the fraction of pixels in untampered regions that are included in the detected spliced regions. We use the receiver-operator characteristics (ROC) curve as a comprehensive evaluation of the detection accuracy and false positive rates shown in Fig. 10. Performance levels on each ROC curve are obtained by adjusting the segmentation threshold of local noise variances. The reported ROC curves are results of averaging over 100 forged images.

The left panel in Fig. 10 corresponds to ROC curves for a fixed size of the spliced region (16 % of the original image) and different noise levels. We measure the added noise strength with the local signal-to-noise ratio defined as $\log_{10}(\frac{\sigma_0^2+\sigma^2}{\sigma^2})$, where σ_0 and σ correspond to the variances of the spliced region in the original image (estimated with the global noise variance estimation method of Sect. 4) and the simulated Gaussian noise, respectively. The middle panel shows ROC curves for a fixed noise variance corresponding to local SNR of 20 dB and different sizes of the spliced regions (measured as the percentage of the size of the original images). The right panel of Fig. 10 shows ROC curves for fixed noise variance and splicing region sizes, but the forged image undergoes JPEG compression of different quality rates. These results show that the overall detection accuracies for higher noise variances or larger regions are better. However, the performances deteriorate with lower noise variances, smaller regions, or low JPEG qualities, as under such circumstances, the difference between spliced region and the original image cannot be effectively reflected by local noise variances.

7 Discussion

In this work, we describe an effective and efficient method of blind (local) image noise estimation and its application in detecting region splicing. Our method is based on the observed projection kurtosis concentration phenomenon. The blind noise estimation problem is then formulated as an optimization problem, whose closed-form solution is the basis of an efficient blind local noise estimation method. We provide experimental evaluations showing the effectiveness and robustness of our method.

Though we have demonstrated promising performances, we would like to point out that the technique described in this work is based on some very simple and specific statistical aberration caused by additive noise. Unquestionably, there exist much more statistical characteristics that distinguish natural image patches from noise patches, and one can expect more sophisticated methods based on contrasting the two types of patches may become more effective in estimating the noise.

We are also aware of several limitations of our region splicing detection method. First, our method relies on the assumption that the spliced region and the original image have different intrinsic noise variances. Therefore, wherever their difference in noise variances is not significant, our method may fail to locate the spliced region. One case in point is when the tampered image underwent heavy JPEG compression. Furthermore, we assume that the intrinsic noise variances are similar across different pixels of the untampered

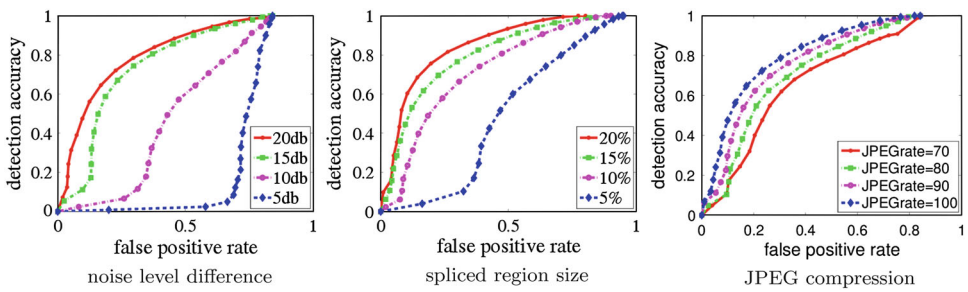

Fig. 10 ROC curves of detection accuracy vs false positive rates for different settings of the tampering operation

image. This may not hold for images with large regions with distinct texture and smooth regions (e.g., a forest in the background of sky) or large regions of saturated pixels. The existence of such regions is the major source of false detections of our current method. Thirdly, inconsistency in local noise statistics is only one aspect that can be used to expose spliced regions. In particular, the blind noise estimation method provides an easy approach for forged images to avoid being detected by such methods, by first estimating and compensating the difference in noise variances of the original and spliced regions. To be more effective in detection, these methods should be combined with other detection methods that are based on alternative regular aspects of untampered images.

There are several important directions that we would like to further extend the current work. The blind local noise estimation method assume a simple noise model that is (1) additive, (2) signal independent, (3) independent, and (4) Gaussian distributed. As we have shown both theoretically (Sect. 4.3) and empirically (Sect. 4.4), the additivity and Gaussianity assumptions of the noise model can be relaxed to certain extend in our method. An important future work is to further extend our method to correlated noise, where instead of local noise variance, we would like to estimate the local noise *covariance* matrix from a single noise corrupted image. Such an algorithm can also help to reduce false detections of spliced regions, as noise typically has much weaker inter-pixel correlations compared to structures in images, and features from the estimated local noise covariance matrix can be used to further differentiate image structures from noise.

Second, we have demonstrated a 1D variant of our region splicing detection method is also effective for audio forgeries (Pan et al. 2012b). Along this direction, we will further extend our method for the detection of video splicing by identifying significant difference in spatial-temporal local noise variances, so it can be applied to detect region splicing in videos. We would also study local noise statistics as a characterizing "fingerprint" of the acquisition devices (e.g., camera or scanner) to prove or disprove an image's authenticity.

Last, the ubiquity of noise suggests wider applicability of our noise variance estimation algorithms. In particular, we have recently also studied the use of a variant of our local noise variance estimation algorithm on scanning pattern estimation from sensitivity coding magnetic resonance images in medical imaging applications (Pan et al. 2012a).

Acknowledgments We would like to thank Daniel Zoran, Zhouchen Lin and Babak Mahdian for kindly sharing the images, codes and results of their works with us. We would also like to thank the two anonymous reviewers for their constructive comments that helped us improve this work. This work is supported in part by the National Science Foundation under Grant Nos. IIS-0953373, IIS-1208463 and CCF-1319800.

Appendix 1: Lyapunov Central Limit Theorem

Suppose $\{X_1, \cdots, X_n\}$ is a sequence of independent random variables, each with finite expected value μ_i and variance σ_i^2. Define $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$, If for some $\delta > 0$, the "Lyapunovs condition"

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}\left[\, |X_i - \mu_i|^{2+\delta} \,\right] = 0$$

is satisfied, then a sum of $(X_i - \mu_i)/s_n$ converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \xrightarrow{d} \mathcal{N}(0,\ 1).$$

Appendix 2: Derivation of Claim 1

Based on the joint density function of the GSM variable $\mathbf{x}$, Eq. (5), we can obtain the marginal distribution of its projection on a non-zero vector $\mathbf{w}$ as:

$$p_{\mathbf{w}}(t) = \int_{\mathbf{x}:\mathbf{w}^T\mathbf{x}=t} p(\mathbf{x})d\mathbf{x} =$$
$$\int_{\mathrm{z}} p_{\mathrm{z}}(\mathrm{z})d\mathrm{z} \int_{\mathbf{x}:\mathbf{w}^T\mathbf{x}=t} \frac{1}{\sqrt{(2\pi\,\mathrm{z})^d |\det(\Sigma_{\mathbf{x}})|}} \exp\left(-\frac{\mathbf{x}^T\Sigma_{\mathbf{x}}^{-1}\mathbf{x}}{2\mathrm{z}}\right) d\mathbf{x}.$$

The marginalization in the inside integral reduces to a Gaussian distribution with zero mean and variance $\mathrm{z}\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w}$, based on the property of Gaussian distributions. Therefore, $p_{\mathbf{w}}(t)$ is a 1D GSM model with mixing density $p_{\mathrm{z}}(\mathrm{z})$, as:

$$p_{\mathbf{w}}(t) = \int_{\mathrm{z}} \mathcal{N}_t(0, \mathrm{z}\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w})\, p_{\mathrm{z}}(\mathrm{z})d\mathrm{z} =$$
$$\int_{\mathrm{z}} \frac{1}{\sqrt{2\pi\,\mathrm{z}\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w}}} \exp\left(-\frac{t^2}{2\mathrm{z}\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w}}\right) p_{\mathrm{z}}(\mathrm{z})d\mathrm{z}$$

Now, the variance of $\mathbf{w}^T\mathbf{x}$ is computed as

$$\mathcal{E}_t\left\{t^2\right\} = \int_{\mathrm{z}} p_{\mathrm{z}}(\mathrm{z})d\mathrm{z} \int_t t^2 \mathcal{N}_t(0, \mathrm{z}\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w}) =$$
$$\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w} \int_{\mathrm{z}} \mathrm{z}\, p_{\mathrm{z}}(\mathrm{z})d\mathrm{z} = \mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w}\mathcal{E}_{\mathrm{z}}\left\{\mathrm{z}\right\}.$$

Furthermore, the fourth order moment of $\mathbf{w}^T\mathbf{x}$ is given as

$$\mathcal{E}_t\left\{t^4\right\} = \int_{\mathrm{z}} p_{\mathrm{z}}(\mathrm{z})d\mathrm{z} \int_t t^4 \mathcal{N}_t(0, \mathrm{z}\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w}) =$$
$$3\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w} \int_{\mathrm{z}} \mathrm{z}^2 p_{\mathrm{z}}(\mathrm{z})d\mathrm{z} = 3\mathbf{w}^T\Sigma_{\mathbf{x}}\mathbf{w}\mathcal{E}_{\mathrm{z}}\left\{\mathrm{z}^2\right\},$$

where we use the fact that for a Gaussian distribution $\mathcal{N}_t(0, \sigma^2)$, its fourth order moment is $3\sigma^4$. Putting all results

together in the definition of kurtosis, we have

$$\kappa(\mathbf{w}) = \frac{\mathcal{E}_t\{t^4\}}{\mathcal{E}_t\{t^2\}^2} - 3 = \frac{3\mathcal{E}_z\{z^2\}}{\mathcal{E}_z\{z\}^2} - 3$$
$$= \frac{3(\mathcal{E}_z\{z^2\} - \mathcal{E}_z\{z\}^2)}{\mathcal{E}_z\{z\}^2} = \frac{3\mathrm{var}_z\{z\}}{\mathcal{E}_z\{z\}}.$$

Appendix 3: Derivation of Eq. (11)

First we expand the objective function in (7) as

$$L\left(\sqrt{\kappa}, \sigma^2\right) = \sum_{k=1}^{K}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right)^2,$$

The gradient of $L\left(\sqrt{\kappa}, \sigma^2\right)$ with regards to the two parameters are computed as, as:

$$\frac{\partial L}{\partial \sigma^2} = 2\sum_{k=1}^{K}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right)\frac{\sqrt{\kappa}}{\tilde{\sigma}_k^2}. \quad (18)$$

$$\frac{\partial L}{\partial \sqrt{\kappa}} = 2\sum_{k=1}^{K}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right)\left(\frac{\sigma^2}{\tilde{\sigma}_k^2} - 1\right) \quad (19)$$

Setting Eq. (18) to zero, and considering $\sqrt{\kappa} > 0$, we have

$$\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right) = 0, \quad (20)$$

Setting Eq. (19) to zero and substituting with Eq. (20) yield

$$\sum_{k=1}^{K}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right) = 0,$$

from which we can obtain

$$\sigma^2 = \frac{1}{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}} - \frac{1}{\sqrt{\kappa}}\frac{\sum_{k=1}^{K}\sqrt{\tilde{\kappa}_k}}{\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}}. \quad (21)$$

Next, substituting Eq. (21) back into Eq. (20), we have

$$\sqrt{\kappa}\left(\frac{1}{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}} - \frac{1}{\sqrt{\kappa}}\frac{\sum_{k=1}^{K}\sqrt{\tilde{\kappa}_k}}{\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}}\right)\sum_{k=1}^{K}\frac{1}{(\tilde{\sigma}_k^2)^2}$$
$$+\sum_{k=1}^{K}\frac{\sqrt{\tilde{\kappa}_k}}{\tilde{\sigma}_k^2} - \sqrt{\kappa}\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2} = 0. \quad (22)$$

Further arranging terms and replacing average over different channels with $\langle\cdot\rangle_k$ yield Eq. (11). Further checking the second-order conditions ensures that the solution is the unique global minimizer of Eq. (7).

References

Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, *36*(1), 99–102.

Bayram, S., Avcibas, I., Sankur, B., & Memon, N. D. (2006). Image manipulation detection. *Journal of Electronic Imaging*, *15*(4), 1–17.

Benedict, T. R., & Soong, T. T. (1967). The joint estimation of signal and noise from the sum envelope. *IEEE Transactions on Information Theory*, *13*(3), 447–454.

Bethge, M. (2006). Factorial coding of natural images: How effective are linear models in removing higher-order dependencies? *The Journal of the Optical Society of America A*, *23*(6), 1253–1268.

Bilcu, R. C., & Vehvilainen, M. (2005). A new method for noise estimation in images. In *IEEE EURASIP International workshop on nonlinear signal and image processing*.

Burt, P., & Adelson, E. (1981). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, *31*(4), 532–540.

Cardoso, J. F. (1999). High-order contrasts for independent component analysis. *Neural Computation*, *11*(1), 157–192.

Chen, M., Fridrich, J. J., Lukás, J., & Goljan, M. (2007). Imaging sensor noise as digital X-ray for revealing forgeries. In *Information hiding* (pp. 342–358).

Chen, W., Shi, Y. Q., & Su, W. (2007). Image splicing detection using 2-D phase congruency and statistical moments of characteristic function. In *Society of photo-optical instrumentation engineers (SPIE) conference series* (Vol. 6505). doi:10.1117/12.704321.

Chen, M., Fridrich, J., Goljan, M., & Lukas, J. (2008). Determining image origin and integrity using sensor noise. *IEEE Transactions on Information Forensics and Security*, *3*(1), 74–90. doi:10.1109/TIFS.2007.916285.

Crow, F. (1984). Summed-area tables for texture mapping. In *ACM SIGGRAPH*.

Danielyan, A., & Foi, A. (2009). Noise variance estimation in nonlocal transform domain. In *International workshop on local and non-local approximation in image processing*.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613–627.

Farebrother, R. W. (1990). The cumulants of the logarithm of a gamma variable the cumulants of the logarithm of a gamma variable. *Journal of Statistical Computation and Simulation*, *36*(4), 243–245.

Farid, H. (2009). Photo fakery and forensics. *Advances in Computers*, *77*, 1–55.

Feller, W. (1968). *An introduction to probability theory and its applications*. New York: Wiley.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *The Journal of the Optical Society of America A*, *4*(12), 2379–2394.

Filler, T., Fridrich, J. J., & Goljan, M. (2008). Using sensor pattern noise for camera model identification. In *IEEE International conference on image processing*. San Diego, CA.

Foi, A., Alenius, S., Katkovnik, V., & Egiazarian, K. (2007). Noise measurement for raw-data of digital imaging sensors by automatic segmentation of non-uniform targets. *IEEE Sensors Journal*, *7*(10), 1456–1461.

Forstner, W. (1998). Image preprocessing for feature extraction in digital intensity, color and range images. In: Springer lecture notes on Earth siences.

Franzen, R. (1999). Kodak lossless true color image suite. http://r0k.us/graphics/kodak. Accessed 17 Dec 2013.

Fu, D., Shi, Y. Q., & Su, W. (2007). Image splicing detection using 2-d phase congruency and statistical moments of characteristic function. In *Proceedings of SPIE security, steganography, and watermarking of multimedia contents IX*.

Garry, M., & Gerrie, M. P. (2005). When photographs create false memories. *Current Directions in Psychological Science*, *14*, 326–330.

He, J., Lin, Z., Wang, L., & Tang., X. (2006). Detecting doctored JPEG images via DCT coefficient analysis. In *ECCV*.

Hsu, Y. F., & Chang, S. F. (2006). Detecting image splicing using geometry invariants and camera characteristics consistency. In *IEEE International Conference on Multimedia and Expo*.

Hsu, Y. F., & Chang, S. F. (2007). Image splicing detection using camera response function consistency and automatic segmentation. In *IEEE International Conference on Multimedia and Expo*.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634.

Konstantinides, G. S. Y. K., & Natarajan, B. K. (1997). Noise estimation and filtering using block-based singular value decomposition. *IEEE Transactions on Image Processing*, *6*(3), 479–483.

Lin, Z., Wang, R., Tang, X., & Shum, H. (2005). Detecting doctored images using camera response normality and consistency. In *CVPR*.

Lin, Z., He, J., Tang, X., & Tang, C. (2009). Fast, automatic and fine-grained tampered JPEG images detection via DCT coefficient analysis. *Pattern Recognition*, *42*(11), 2492–2501.

Liu, C., Szeliski, R., Kang, S. B., Zitnick, C. L., & Freeman, W. T. (2008). Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 299–314. http://doi.ieeecomputersociety.org/10.1109/TPAMI.2007.1176.

Liu, X., Tanaka, M., Okutomi, M. (2012). Noise level estimation using weak textured patches of a single noisy image. In *IEEE International conference on image processing*.

Lukas, J., Fridrich, J., & Goljan, M. (2006). Detecting digital image forgeries using sensor pattern noise. In *Proceedings of SPIE security, steganography, and watermarking of multimedia contents VIII*.

Lukás, J., Fridrich, J. J., & Goljan, M. (2006). Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, *1*(2), 205–214.

Lyu, S., & Simoncelli, E. P. (2009). Nonlinear extraction of 'independent components' of natural images using radial Gaussianization. *Neural Computation*, *18*(6), 1–35.

Mahdian, B., & Saic, S. (2009). Using noise inconsistencies for blind image forensics. *Image and Vision Computing*, *27*(10), 1497–1503.

Matzner, R., & Engleberger, F. (1994). An SNR estimation algorithm using fourth-order moments. In *IEEE International symposium on information theory*.

Meer, P., Jolion, J., & Rosenfeld, A. (1990). A fast parallel algorithm for blind estimation of noise variance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(2), 216–223.

Nakamura, J. (Ed.). (2006). Image sensors and signal processing for digital still cameras. Boca Raton: Taylor and Francis.

Ng, T. T., & Chang, S. F. (2004). A model for image splicing. In *IEEE International conference on image processing (ICIP)*. Singapore.

Olsen, S. I. (1993). Estimation of noise in images: An evaluation. *Computer Vision Graphics Image Process. Graphic Models and Image Process*, *55*(4), 319–323.

Pan, X., Zhang, X., & Lyu, S. (2011). Exposing image forgery with blind noise estimation. In *The 13th ACM workshop on multimedia and security*. Buffalo, NY.

Pan, X., Zhang, X., & Lyu, S. (2012). Blind local noise estimation for medical images reconstructed from rapid acquisition. In *SPIE Symposium on medical imaging*. San Diego, CA.

Pan, X., Zhang, X., & Lyu, S. (2012). Detecting splicing in digital audios using local noise level estimation. In *IEEE International conference on acoustics, speech, and signal processing (ICASSP)*. Kyoto, Japan.

Pan, X., Zhang, X., & Lyu, S. (2012). Exposing image splicing with inconsistent local noise variances. In *IEEE International conference on computational photography*. Seattle, WA.

Pauluzzi, D. R., & Beaulieu, N. C. (2000). A comparison of SNR estimation techniques for the AGWN channel. *IEEE Transactions on Communications*, *48*(10), 1681–1691.

Ponomarenko, N. N., Lukin, V. V., Abramov, S. K., Egiazarian, K. O., & Astola, J. T. (2003). Blind evaluation of additive noise variance in textured images by nonlinear processing of block DCT coefficients. In *SPIE Proceedings* (Vol. 5014, pp. 178–189).

Popescu, A., & Farid, H. (2004). Statistical tools for digital forensics. In *6th International workshop on information hiding*. Toronto, Canada.

Portilla, J. (2004). Full blind denoising through noise covariance estimation using Gaussian scale mixtures in the wavelet domain. In *International conference on image processing*. doi:10.1109/ICIP.2004.1419524.

Portilla, J., Strela, V., Wainwright, M. J., & Simoncelli, E. P. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, *12*(11), 1338–1351.

Qu, Z., Qiu, G., & Huang, J. (2009). Detect digital image splicing with visual cues. In: S. Katzenbeisser, & A. R. Sadeghi (Eds.) *International workshop on information hiding* (pp. 247–261).

Rank, K., Lendl, M., & Unbehauen, R. (1999). Estimation of image noise variance. In *IEE Proceedings. Vision, Image and Signal Processing* (Vol. 146, pp. 80–84).

Rudin, L., Lions, P., & Osher, S. (2003). Multiplicative denoising and deblurring: Theory and algorithms. In S. Osher & N. Paragios (Eds.), *Geometric level set methods in imaging, vision, and graphics*. New York: Springer.

Sacchi, D. L. M., Agnoli, F., & Loftus, E. F. (2007). Changing history: Doctored photographs affect memory for past public events. *Applied Cognitive Psychology*, *21*, 1005–1022.

Schaefer, G., & Stich, M. (2004). UCID—an uncompressed colour image database. In *Proc. SPIE, storage and retrieval methods and applications for multimedia*.

Schmidt, U., Schelten, K., & Roth, S. (2011). Bayesian deblurring with integrated noise estimation. In *IEEE International conference on computer vision*. Colorado Springs, CO.

Sencar, H. T., & Memon, N. (Eds.). (2012). Digital image forensics: There is more to a picture than meets the eye. Dordrecht: Springer.

Serra, J. (1988). *Image analysis and mathematical morphology: Theoretical advances. Image Analysis and Mathematical Morphology*. London: Academic Press.

Shi, Y. Q., Chen, C., & Chen, W. (2007). A natural image model approach to splicing detection. In *Proceedings of the 9th workshop on multimedia and security* (pp. 51–62). New York, NY: ACM. doi:10.1145/1288869.1288878.

Simoncelli, E. P., & Freeman, W. T. (1995). The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *IEEE Second international conference on image processing*.

Simoncelli, E. P., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216. doi:10.1146/annurev.neuro.24.1.1193.

Stefano, A., White, P., & Collis, W. (2004). Training methods for image noise level estimation on wavelet components. *EURASIP Journal on Applied Signal Processing*, *16*, 2400–2407.

Tai, S. C., & Yang, S. M. (2008). A fast method for image noise estimation using Laplacian operator and adaptive edge detection. In *International symposium on communications, control and signal processing*.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1394), 359–366.

Viola, P., & Jones, M. (2002). Robust real-time object detection. *International Journal of Computer Vision*, *57*(2), 137–154.

Wainwright, M. J., & Simoncelli, E. P. (2000). *Scale mixtures of Gaussians and the statistics of natural images*. Cambridge, MA: MIT Press.

Wang, W., Dong, J., & Tan, T. (2009). Effective image splicing detection based on image chroma. In *IEEE International conference on image processing*.

Withagen, P., Groen, F., & Schutte, K. (2005). CCD characterization for a range of color cameras. In *Instrumentation and measurement technology conference, 2005* (Vol 3, pp. 2232–2235). doi:10.1109/IMTC.2005.1604573.

Zoran, D., & Weiss, Y. (2009). Scale invariance and noise in nature image. In *IEEE International conference on computer vision*. Kyoto, Japan.