Exposing Image Splicing with Inconsistent Local Noise Variances

Xunyu Pan Xing Zhang Siwei Lyu*
Computer Science Department
University at Albany, State University of New York
{xypan,xz654242,lsw}@cs.albany.edu

Abstract

Image splicing is a simple and common image tampering operation, where a selected region from an image is pasted into another image with the aim to change its content. In this paper, based on the fact that images from different origins tend to have different amount of noise introduced by the sensors or post-processing steps, we describe an effective method to expose image splicing by detecting inconsistencies in local noise variances. Our method estimates local noise variances based on an observation that kurtosis values of natural images in band-pass filtered domains tend to concentrate around a constant value, and is accelerated by the use of integral image. We demonstrate the efficacy and robustness of our method based on several sets of forged images generated with image splicing.

1. Introduction

The past decade has witnessed remarkable advances in digital image processing and computational photography, resulting in sophisticated image-editing software systems such as `Adobe Photoshop` and `GNU GIMP`. On the other hand, the ease of digital image manipulation has also posed many new challenges. In particular, digital images have become more vulnerable to malicious tampering compared to their non-digital counterparts. This circumstance naturally calls for solutions to important and challenging research questions such as "how can we determine an image's authenticity?" and "how can we detect and locate tampered parts of an image?" [7].

In this work, we focus on the detection of a simple and common image manipulation technique known as *image splicing*, where a forged image is created by compositing regions from *different* source images. Our image splicing detection method is based on the fact that most digital images contain noise introduced either during acquisition (e.g., sensor or quantization noise) or subsequent processing (e.g., compression). For an un-tampered natural image, the noise variances across different regions typically differ only slightly. But with spliced regions from another image with a significantly different intrinsic noise variance, the inconsistency of local noise variances become a telltale evidence of tampering.

The crucial step in our image splicing detection method is to estimate noise variances across different regions in an image. In doing so, we take advantage of a statistical regularity of natural images – the kurtosis values of natural images in general band-pass filtered domains (e.g., DCT or wavelet decomposition) are positive [4] and tend to be close to a constant [3, 21, 34]. Then, approximating kurtosis of natural images across different band-pass filtered channels to be a positive constant, we construct an objective function, using the relationship between the image kurtosis and noise variance in the band-pass filtered domain, to estimate the global noise variance of the whole image. Our objective function is robust to infrequent outlying kurtosis values. More appealingly the objective function has a closed-form optimal solution, the algorithm based on which is further extended to local noise variance estimation and accelerated with *integral image*. Spliced regions are detected by segmenting the estimated local noise variances. The efficacy and robustness of our method are demonstrated on several sets of synthetic and realistic forged images generated with image splicing.

1.1. Related Works

Noise could be introduced in an image by numerous factors during acquisition and processing, such as thermal effects, sensor saturation, quantization errors and transmission errors. Effective estimation of noise statistics, in particular variance, from a single noise-corrupted image is an important task, partly because other common image processing algorithms such as denoising and deblurring are predicated on *a priori* knowledge of the noise variance.

Most noise estimation methods make a simplifying assumption of the noise as a zero-mean white Gaussian pro-

*The first two authors contributed equally to this work.

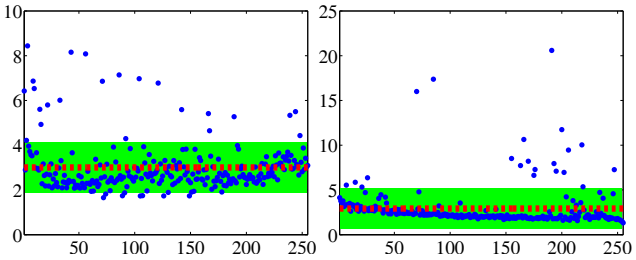


Figure 1. *Demonstration of kurtosis concentration of a natural image for two types of band-pass filtered domains: DCT (left) and FastICA (right). See text for details.*

cess that is additive and independent of the image[1]. As such, image noise can be statistically characterized by their variances. Based on the observation that noise tends to have more energy in the high frequency bands than natural images, estimation of the global noise variance in an image is usually based on the difference image between the noisy image and its low-pass filtered response [15, 25, 29]. For instance, the mean absolute deviation (MAD) of the difference image has been widely used as a reliable noise variance estimator [6]. However, these methods tend to overestimate the noise variance, as frequency alone is not sufficient to differentiate noise from image.

An alternative methodology is to use the relationship between noise variance and certain higher-order statistics of natural images to obviate the low pass filtering step. In [31], 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. Another family of data set independent estimation methods uses the relationship between noise variance and image kurtosis in band-pass filtered domains (see Eq.(1)). However, many works along this direction (e.g., [2, 23, 27]) assume knowing the kurtosis of the original image, which makes them impractical for practical applications. The algorithm of [34] is the first kurtosis-based method that does not require knowing the kurtosis values and achieves state-of-the-art performances (more details given in Section 3).

Image noise have found many uses in digital image forensics. Most notably, the existence or lack of the PRNU (photo response non-uniformity) noise has been used for identifying the source camera model or ensuring authenticity of an image [5, 9, 19, 20]. There have been three major previous works using local noise variances to detect image splicing. The method in [28] is based on an estimator of noise variance from the kurtosis of the original image. The method in [22] is based on the MAD noise variance estimator in the high-pass band from an orthonormal wavelet decomposition. The more recent work in [26] demonstrate detection of spliced regions by extending the noise variance estimator in [34] to local blocks.

Image splicing can also be detected using other features. For instance, the existence of spliced regions in an image can be detected using a classifier of image statistics-based features learned from a training set of authentic and forged images [1, 11, 24], however, these methods do not locate the spliced regions. Other methods (e.g., [13, 17]) extract features from camera response functions (CRF). However, these semi-automatic methods are typically very slow and not suitable for practical analysis due to the complicated nonlinear optimization in estimating CRF [17]. Another family of detection methods (e.g., [12, 16]) use double JPEG artifacts to detect image splicing. However, since these methods rely on the JPEG image format, they cannot be applied to images that are not in the JPEG format.

In contrast, the proposed method has the following advantages:

1. it does not explicitly rely on the knowledge of image format or tampering procedure;
2. it has detection accuracy at the level of pixels yet the overall algorithm is still efficient;
3. its detection performance is extensively evaluated on several different sets of synthesized and actual image splicing forgeries with different formats and level of sophistication.

We will compare existing image splicing detection methods with our method in detail in Section 5.2.

2. Kurtosis Concentration

For a random variable x, we define its kurtosis as $\kappa = \frac{\tilde{\mu}_4}{(\sigma^2)^2} - 3$, where $\sigma^2 = \mathcal{E}_x\left\{(x - \mathcal{E}_x\{x\})^2\right\}$ and $\tilde{\mu}_4 = \mathcal{E}_x\left\{(x - \mathcal{E}_x\{x\})^4\right\}$ are the variance and fourth order central moment of x, respectively. In particular, by this definition, a Gaussian variable has kurtosis zero.

It has been observed in several recent works (e.g., [3, 21, 34]) that for natural images, kurtosis values across different band-pass filter channels tend to be close to a constant, a phenomenon we call *kurtosis concentration*. One probable cause of kurtosis concentration is that natural images tend to have spherically symmetric distributions [21], and therefore have isotropic statistics (including kurtosis), in band-pass filtered domains from whitening transforms including those from DCT and FastICA.

We demonstrate kurtosis concentration in Fig. 1 using a natural image and two different types of band-pass filter transforms, as:

1. 255 band-pass AC filters from a 16×16 DCT decomposition, in ascending order of their basic frequency,
2. and 255 band-pass filters obtained from the FastICA decomposition [14] learned from overlapping patches of 16×16 pixels in size collected from the image.

[1] A more realistic model where image noise variances depend on pixel intensities are used in [18] to obtain a lower-bound on the global noise level.

The blue dots are kurtosis values obtained from the responses of the image convolved with the band-pass filters corresponding to the order given in the horizontal axes. The red dashed line is the mean of these kurtosis values with the shaded region corresponding to $\pm$ standard deviation. One notable characteristics of these results is that the kurtosis values are all positive, reflecting the super-Gaussian marginal statistics of natural images in the band-pass filtered domains [4]. More importantly, except for a few outliers, the majority of the kurtosis values are found not far from their mean value, confirming the concentration behavior of the kurtosis values.

3. Global Noise Variance Estimation

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 $\mathbf{y}$ alone with some general statistical knowledge of $\mathbf{x}$. We would like to point out that the Gaussian assumption is not as restrictive as it looks, in particular when the image is transformed from pixel domains to band-pass filtered domains such as discrete cosine transforms (DCT) and wavelet. This is because the non-Gaussian independent noise in the pixel domain approaches Gaussian after being linearly mixed by the band-pass filters, a direct result of the central limit theorem [8].

We first transform the image into a band-pass filtered domain. We find that the particular form of the band-pass filter transforms does not affect the algorithm and performance, and use the AC filters from a fixed-point DCT decomposition for its efficiency. Further, if the band-pass filters are normalized to have unit one, the variance of the noise is same as that in the original pixel domain.

For a band-pass filtered domain of K channels (i.e., the responses of image convolved with K different band-pass filters), we denote the kurtosis of the original and the noisy image in the k^{th} channel as κ_k and $\tilde{\kappa}_k$, respectively. We use $\tilde{\sigma}_k^2$ for the variance of the k^{th} channel of the noisy image $\mathbf{y}$. These statistics are related [2], as:

$$\tilde{\kappa}_k = \kappa_k \left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2} \right)^2 . \tag{1}$$

For completeness, we provide a derivation of Eq.(1) in Appendix A. Eq.(1) can be further simplified if we consider the statistical regularity of natural images in the band-pass filtered domains – they tend to have super-Gaussian marginal distributions [4], or equivalently, have positive kurtosis values ($\kappa_k > 0$). Also, $\tilde{\sigma}_k^2 - \sigma^2$ is positive as it is the variance of the noise free image in the k^{th} channel. Therefore, we can take square root on both sides of Eq.(1) to yield

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

Now we take advantage of the kurtosis concentration behavior of natural images in the band-pass filtered domains (Section 2). In particular, this suggests that the kurtosis of the noise-free natural image $\mathbf{x}$ across the K band-pass filtered channels can be approximated by a constant, or $\kappa_k \approx \kappa$ ($k = 1, \cdots, K$). We then form an objective function of minimizing the difference of the two sides of Eq.(2), as:

$$\min_{\sqrt{\kappa},\sigma^2} \sum_{k=1}^{K} \left[\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} \left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2} \right) \right]^2 , \tag{3}$$

whose optimal solution provides an estimation to the noise variance. It turns out that this optimal solution is in closed-form, as:

$$\begin{aligned} \sqrt{\kappa} &= \frac{\left\langle \sqrt{\tilde{\kappa}_k} \right\rangle_k \left\langle \frac{1}{(\tilde{\sigma}_k^2)^2} \right\rangle_k - \left\langle \frac{\sqrt{\tilde{\kappa}_k}}{\tilde{\sigma}_k^2} \right\rangle_k \left\langle \frac{1}{\tilde{\sigma}_k^2} \right\rangle_k}{\left\langle \frac{1}{(\tilde{\sigma}_k^2)^2} \right\rangle_k - \left\langle \frac{1}{\tilde{\sigma}_k^2} \right\rangle_k^2} \\ \sigma^2 &= \frac{1}{\left\langle \frac{1}{\tilde{\sigma}_k^2} \right\rangle_k} - \frac{1}{\sqrt{\kappa}} \frac{\left\langle \sqrt{\tilde{\kappa}_k} \right\rangle_k}{\left\langle \frac{1}{\tilde{\sigma}_k^2} \right\rangle_k} , \end{aligned} \tag{4}$$

where $\langle \cdot \rangle_k$ is a shorthand notation for averaging over the K band-pass filtered channels. A detailed derivation of Eq.(4) is given in Appendix B. We should point out that Eq.(4) is key to the extension of the global noise variance estimation to efficient local noise variance estimation.

3.1. Comparison with Method of [34]

Our method is similar to the method in [34], which estimate noise variance by minimizing the squared difference of the two sides of Eq.(1), as:

$$\min_{\kappa,\sigma^2} \sum_{k=1}^{K} \left[\tilde{\kappa}_k - \kappa \left(\frac{\tilde{\sigma}_k^2 - \sigma^2}{\tilde{\sigma}_k^2} \right)^2 \right]^2 . \tag{5}$$

We would like to point out two important advantages of our method over Eq.(5). First, the solution optimizing Eq.(5) does not afford a closed-form, which has to be solved numerically. This makes it inefficient to extend it to local noise variance estimation for detecting image splicing [26]. We also find that the numerical implementation of Eq.(5) sometimes becomes diverging for low noise variances, which are reflected by the N/A entries in Table 1.

Furthermore, the kurtosis concentration does not exclude the possibility of outlying kurtosis values that are significantly different from their mean value across different band-pass filtered channels (see Fig. 1). Such outliers have significant effect on the squared loss between the two sides of Eq.(1). On the other hand, if we denote the two sides of Eq.(1) as a and b, we have $(\sqrt{a} - \sqrt{b})^2 \leq |a - b|$ for any $a, b \geq 0$. Since the objective function in Eq.(3) grows slower than a l_1 loss, which is more robust to outliers compared with the l_2 loss incorporated in Eq.(5), our method is expected to be more robust in the presence of outliers.

PSNR	Kodak		UCID		Van Hateren	
	$\tilde{\sigma}_0$	$\hat{\sigma}_1$	$\tilde{\sigma}_0$	$\hat{\sigma}_1$	$\tilde{\sigma}_0$	$\hat{\sigma}_1$
20dB	20.03 (0.12)	20.29 (0.32)	20.06 (0.21)	20.37 (0.39)	20.00 (0.04)	20.10 (0.22)
25dB	25.04 (0.24)	25.38 (0.55)	25.19 (0.50)	25.55 (0.76)	25.00 (0.07)	25.12 (0.33)
30dB	30.16 (0.53)	30.54 (0.95)	30.46 (1.16)	31.15 (6.06)	30.01 (0.17)	30.19 (0.48)
40dB	41.08 (3.61)	N/A (N/A)	41.71 (4.33)	N/A (N/A)	40.46 (1.96)	43.24 (17.68)
50dB	46.01 (4.82)	N/A (N/A)	45.93 (5.80)	N/A (N/A)	48.70 (3.70)	N/A (N/A)

Table 1. *The average performance of our global noise variance estimation method ($\tilde{\sigma}_0$), with comparisons of the performance of the work in [34] ($\hat{\sigma}_1$) on several different data sets. We show both the mean and standard deviations (in parenthesis) of the estimations. N/A entries correspond to cases when some running of the estimator in [34] diverges.*

3.2. Experimental Results

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

1. 25 8-bit images from the Kodak database [10],
2. 200 8-bit images from the uncompressed image database (UCID) [30],
3. 200 16-bit images from the Van Hateren database [32].

We choose images that have low intrinsic noise variances and are taken under good illuminations. We then add white Gaussian noise of different variances to these images with designated peak-signal-to-noise ratio (PSNR) to accommodate the different dynamic ranges of these images. The PSNR, in unit of deci-Bell (dB), is defined as $20\log_{10}\left(\frac{I_{\max}}{\sigma}\right)$, where $I_{\max}$ is the maximum pixel value of the image, and σ is the standard deviation of the additive white Gaussian noise. Table 1 summarizes the average performance of our method, showing the mean and standard deviation of the estimated noise variances in PSNR. We compare the performance of our method with the state-of-the-art work of [34]. Both sets of results are obtained based on 63 channels from the AC filters in a 8×8 DCT decomposition. Overall, these results indicate that our method achieves more accurate estimations, in particular for low noise variances (high PSNRs). One reason is the robustness of our method to outliers of kurtosis. However, the difference in the estimations of the two methods reduces for low PSNRs, with high noise overwhelming the image.

4. Local Noise Variance Estimation

The global noise variance estimation method assumes that the noise variance σ^2 is a constant across the image. A more general scenario is for the noise variance to change spatially. In this section, we extend the global noise variance estimator given in Eq.(4) to an efficient local noise variance estimation method.

Specifically, our goal is to estimate noise variance $\sigma^2(i,j)$ at each pixel location (i,j) using Eq.(4), with statistics collected from all surrounding rectangular windows of (i,j), $\Omega^k_{(i,j)}, k = 1, \cdots K$, in all band-pass filtered channels. In doing so, we first express the variance and kurtosis using the raw (un-centered) moments, $\mu_m = \mathcal{E}_x\{x^m\}$, as:

$$\begin{aligned} \sigma^2 &= \mu_2 - \mu_1^2 \\ \kappa &= \frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4}{\mu_2^2 - 2\mu_2\mu_1^2 + \mu_1^4} - 3. \end{aligned} \tag{6}$$

The raw moments are estimated for each local window by spatial averaging,

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

where we denote $x(i',j',k)$ the response at (i',j') in the k^{th} band-pass filtered channel. A naive implementation, applying Eq.(6) and then Eq.(4) to each local window $\Omega^k_{(i,j)}$, will lead to an overall running time of $\mathcal{O}(MNK)$, where N and M are the sizes of the image and local window in pixel. This will become inefficient when M is relatively large. On the other hand, we can use *integral image* [33] to accelerate the overall algorithm to a running time of $\mathcal{O}(NK)$.

Integral image is a dynamic programming technique for efficient computation of sum values in rectangular regions in an image (or one channel in a band-pass filtered domain). In particular, each pixel in an integral image constructed from an image $\mathbf{x}$, $\mathcal{I}(\mathbf{x})_{i,j}$, corresponds to the sum of all pixels of $\mathbf{x}$ in $[1,i]\times[1,j]$. Summation in any $[i,i+I]\times[j,j+J]$ in $\mathbf{x}$ can be evaluated with just three addition/subtraction operations on the corresponding integral image, as:

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

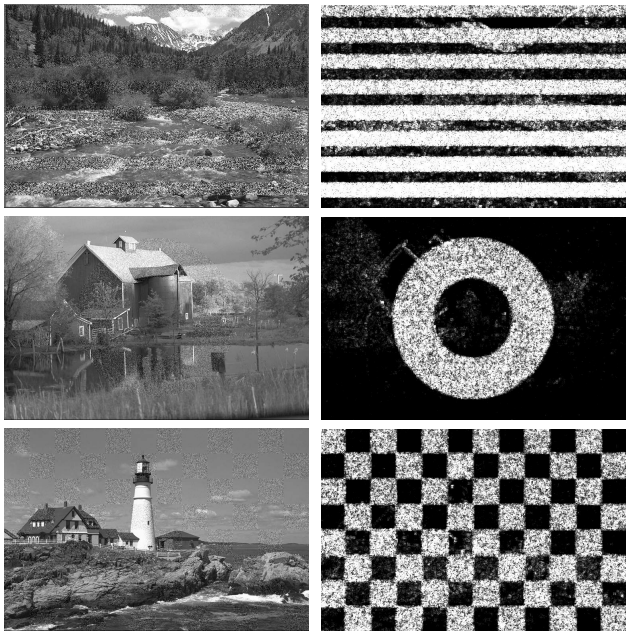
Furthermore, an integral image can be efficiently constructed in linear time of the dimension of $\mathbf{x}$.

Using integral image, we can efficiently compute raw moments in overlapping rectangular regions. Specifically, denote $\circ$ as point-wise multiplication, the m^{th} order raw moment in $[i,i+I]\times[j,j+J]$ can be computed, as:

$$\begin{aligned} \frac{1}{IJ}\Big[&\mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m \text{ times}})_{i+I,j+J} - \mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m \text{ times}})_{i,j+J} \\ &- \mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m \text{ times}})_{i+I,j} + \mathcal{I}(\underbrace{\mathbf{x}\circ\cdots\circ\mathbf{x}}_{m \text{ times}})_{i,j}\Big]. \end{aligned} \tag{7}$$

The basic steps of our local noise variance estimation algorithm are then summarized as:

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



noise-corrupted images detection results

Figure 2. *Local noise variance estimation for three different images with different additive white Gaussian noise patterns.*

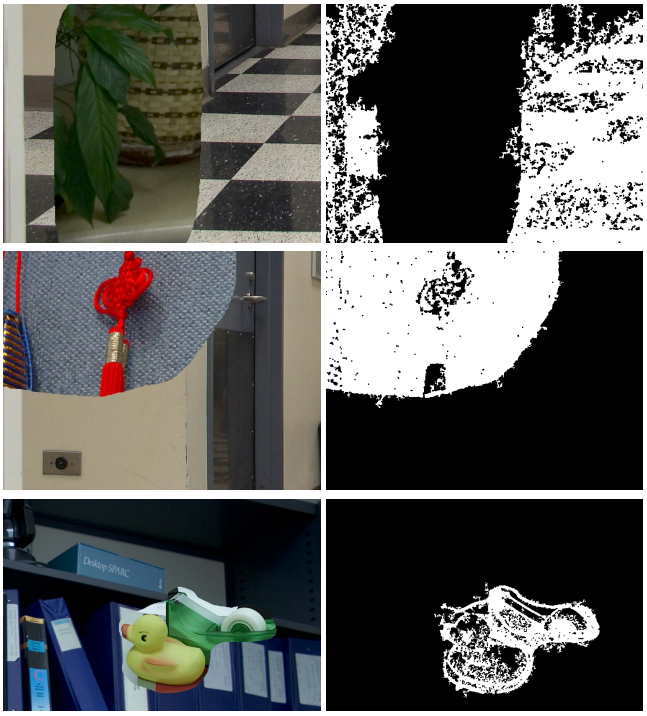
2. Compute the integral images of the first to the fourth order raw moments in each of the K band-pass filtered channels with Eq.(7);
3. Compute variance and kurtosis for each local window in each band-pass filtered channel with Eq.(6);
4. Estimate noise variance by evaluating Eq.(4) for each local window across all band-pass filtered channels.

4.1. Experimental Results

We test the local noise variance estimation method using a set of images corrupted with spatially varying white Gaussian noise of certain patterns. The left column of Fig. 2 shows three examples of the synthesized images, the top image was generated with horizontal stripes of white Gaussian noise, each stripe with different noise variances – PSNR ranging from 13dB at the bottom to 20dB[2] at the top with 1dB increment. The middle and bottom images were created by adding white Gaussian noise of PSNR 20dB, of an annular and a checkerboard structure, respectively. The right column of Fig. 2 shows the estimated local noise variances of the corresponding images using our method, with the intensities proportional to the estimated standard deviations. These results were obtained based on 63 8×8 DCT AC filters, with overlapping sliding windows of 32×32 pixels in each channel.

These results show that our method is very effective in revealing the spatially varying noise variances in these images. On the other hand, note 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). This potentially can cause false detections when this method is used for splicing detection (see Discussion).

[2]We pick high noise variance so that noise will be visible in Fig. 2.



forged images detection

Figure 3. *Detection results of our method for three forged images with image splicing from the* Columbia uncompressed image splicing detection evaluation dataset *[13]. See text for details.*

5. Image Splicing Detection

One important application of our local noise variance estimation algorithm is to detect image splicing. As most digital images contain intrinsic noise, an un-tampered natural image is expected to have similar noise variances across different locations. However, for a forged image with regions from other images, the spliced region can be exposed with the inconsistency in local image noise variances.

We further process the estimated local noise variances by segmenting the image into regions with significantly different noise variances. While more sophisticated image segmentation algorithms can be used, we find that the simple k-means algorithm usually achieves satisfactory performance. We then perform post-processing steps on the detected regions: first we use an area threshold to remove small isolated regions, then we use mathematical morphological operations to smooth and connect the detected regions.

We test this algorithm on two sets of forged images created with image splicing. The first experiment is based on a set of 180 forged images obtained from the *Columbia*

forged images detection

Figure 4. *Detection results of our method for a set of image splicing image forgeries from Worth1000.com. See text for details.*

uncompressed image splicing detection evaluation dataset [13], which are in high resolution and uncompressed. The original images providing the spliced regions are taken with different camera models. Three examples of the forged images, together with their detection results, are shown in Fig. 3. As these results show, local noise variances provide reliable information for differentiating spliced regions from those in the original images.

We further test our splicing detection method on a second set of forged images created with image splicing and collected from 50 candidates of the image manipulation

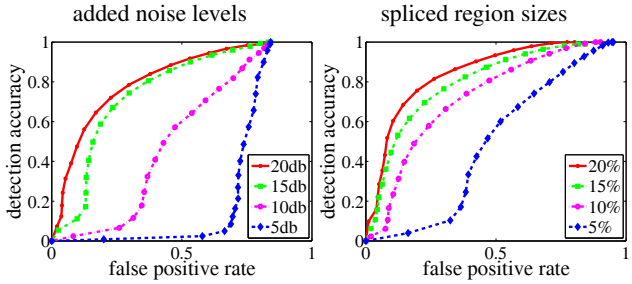


Figure 5. *ROC curves of detection accuracy v.s. false positive rates for (***left***) different noise variances and (***right***) different sizes of the spliced region in percentage of the size of the original image.*

contest on Worth1000.com. Shown in Fig. 4 are several randomly chosen examples of the forged images in this data set and their detection results using our method. Note that these forgeries are carefully manipulated and processed and have realistic appearance, many can only be exposed based on conceptual knowledge of the physical world (e.g., hippopotamus is unlikely to be found in arctic regions). On the other hand, detections based on our method are effective in revealing spliced regions using only low level image statistics, albeit occasional false detections or mis-detections also occur (see Discussion).

5.1. Quantitative Evaluation and Robustness

We further perform a quantitative evaluation of the performance of our image splicing detection. In particular, 100 images were chosen from the UCID database [30]. For each image, we added white Gaussian noise to a randomly chosen rectangular region to simulate a spliced region with different noise variance. 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 un-tampered regions that are detected as tampered ones.

All results are based on 63 AC channels from 8×8 DCT decomposition with a fixed local window size. We use the receiver-operator characteristics (ROC) curve as a comprehensive evaluation of the detection accuracy and false positive rates shown in Fig. 5. 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 first panel in Fig. 5 corresponds to ROC curves for a fixed size of the spliced region (16% of the original image) and with different noise levels. We measure the added noise strength with the local signal-to-noise ratio defined as $\log_{10}\left(\frac{\sigma_0^2+\sigma^2}{\sigma^2}\right)$, where σ_0 and σ correspond to the variances of the spliced region in the original image (estimated with the noise variance estimation method of Section 3) and the added noise, respectively. The second panel in Fig. 5 shows ROC curves for a fixed noise variance corresponding

	our method	method in [22]	method in [26]
image A	47	83	490
image B	38	53	333

Table 2. *Running time (in seconds) comparison of noise-based splicing detection methods based on the two images in Fig. 6.*

to local SNR of 20dB and with different sizes of the spliced regions in the percentage of the size of the original images.

These results show that the overall detection accuracies for higher noise variances or larger regions are better. However, for low noise variances and smaller regions, higher detection accuracy usually suggests higher false detections, which can be attributed to the fact that natural images may have different local noise variances due to the co-existence of texture and smooth regions (see Section 4.1).

5.2. Comparison with Related Works

Next, we compare our method's performance in splicing detection with that of the previous methods. To make the results impartial, we do not compare with methods that only detect the occurrence of splicing without localizing the spliced regions, e.g., [1, 11, 24]. Furthermore, we exclude comparison with methods that are only tested on individual pixel blocks picked by the user, without comparable results showing the locations of detected spliced regions, e.g., [13, 17].

We first compare our method with two other recent noise-based splicing detection methods [22, 26][3]. The former is based on the maximum absolute deviation (MAD) estimator of local noise variances in the high-pass band obtained with an orthogonal wavelet decomposition, while the latter directly applies the global noise variance estimation method in [34] to local noise variance estimation. Both methods examine only non-overlapping blocks (40×40 and 64×64, respectively) to balance performance and running time (see Table 2).

We compare the performances of these methods with our method based on two tampered images created using `Adobe Photoshop` with spliced regions. The top three rows of Fig. 6 show the original images contributing the background, the original images contributing the spliced regions, and the tampered images, respectively. The original images used in making the forgeries are obtained from `Flickr.com`: the first original image in the top row was taken with a Canon 400D digital camera with an ISO speed 1600, while the source camera for the second original image in the top row is unknown. The two original images in the second row were taken with a Canon EOS-60D digital camera with an ISO speed 1600 and a SONY DSC-H20 digital camera with an ISO speed 400, respectively. The different sensor types and ISO speeds in these images lead to different noise variances, even though the actual sensor noise is usually non-Gaussian and not simply additive [18].

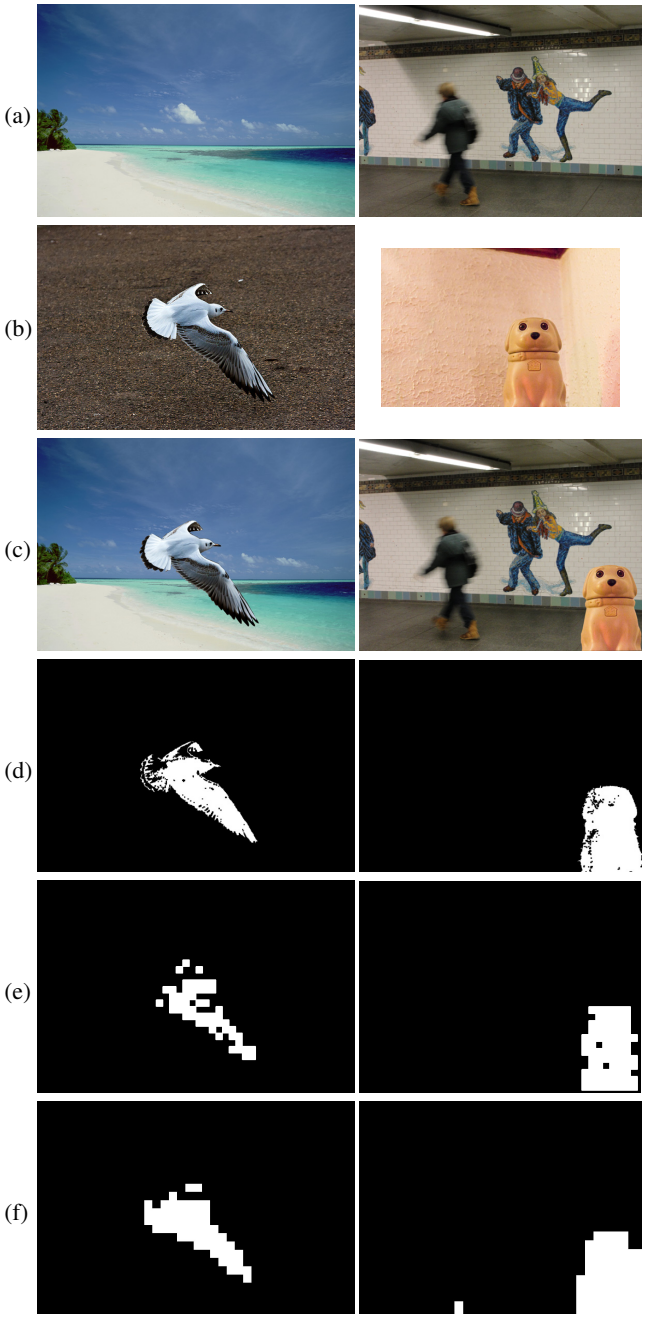
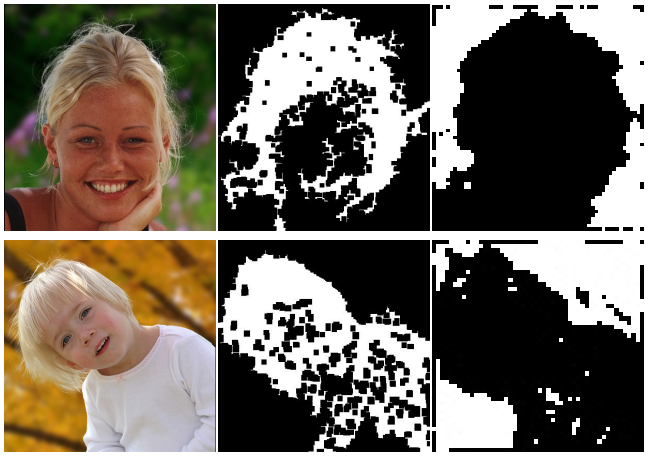


Figure 6. **(a,b)** *Two pairs of original JPEG images from `Flickr.com`.* **(c)** *Forgeries generated with image splicing using the original images with `Adobe Photoshop`. Image courtesy of authors of [26].* **(d)** *Splicing detection results using our method.* **(e)** *Splicing detection results using the method in [26].* **(f)** *Splicing detection results using the method in [22].*

Comparisons in detection accuracy and running time are

[3] We did not compare with the remaining noise-based splicing detection method in [28] because it requires user to select a region with less noise to determine kurtosis of the original image, and thus cannot be implemented as an automatic detection method.



tampered image our method method of [16]

Figure 7. *Comparison of splicing detection results of our method with the method in [16]. Image courtesy of authors of [16].*

shown in the last three rows of Fig. 6 and Table 2, respectively. These results are obtained based on running the three algorithms with unoptimized MATLAB code and a machine with a 2.4GHz Intel CPU and 2GB RAM. For our method, we use the same setting as described in Section 5.1. 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 level of pixels. Furthermore, Table 2 shows that our method is relatively more efficient than the two other alternatives in terms of running time.

We further compare the performance of our method with the splicing detection method based on locating the double JPEG artifacts due to the tampering process [16]. The results on two example images are shown in Fig. 7. Both methods seem to be able to identify similar tampered regions, but our method is not able to detect the entire tampered regions in these two images. One cause of this is that the example images are heavily JPEG compressed, which smoothes the image and destroys most of the noise characteristics, and hence the difference in noise variances between the original images and the spliced regions. On the other hand, the method of [16] only detects 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 [16] to be effective, we have to ensure that (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 [16]. These limitations affect the applicability of the JPEG based detection methods, for instance, they cannot be applied to tampered images in uncompressed format, or generated from original images of uncompressed format, so even obvious spliced regions in simple examples as shown in Fig. 3 will not be detected using this method[4]. On the other hand, our method has relatively fewer restrictive assumptions and can be applied in a wider range of detection situations.

6. Discussion

In this work, based on the kurtosis concentration property of natural images in band-pass filtered domains, we describe an effective method for estimating image noise variance. Our method is further extended to an efficient method for the estimation of local noise variances, which is applied to the detection and locating of spliced regions for digital image forensics.

Though we have demonstrated some satisfactory detection performances, we are also aware of some limitations of our 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 pixel locations within the original image. This may not hold for images with distinct texture and smooth regions (e.g., a tree in the background of sky), or those with large regions of saturated pixels. Such inhomogeneous local noise variance can cause our method to make false detections.

There are several important directions that we would like to further extend the current work. One solution to reduce false detections is to estimate the full second-order correlations among local pixels. As noise typically has much weaker inter-pixel correlations compared to textures in images, this can further help our method to differentiate image structures from random noise of various forms. Second, we would also like to further extend our method for the detection of video splicing by identifying significant difference in spatial-temporal local noise variances. Last, we would also like to apply our local noise variance estimation method to the removal of spatially varying sensor noise to improve the visual quality of digital images.

Acknowledgement

We would like to thank Daniel Zoran and Yair Weiss for sharing their code of their work [34] with us. We would also like to thank Dr. Zhouchen Lin for sharing with us the

[4]The authors of [16] also mentioned that their method cannot be tested "on images downloaded from the web because we are not sure of whether the original images were in JPEG" and therefore was only tested on an image database built by the authors [16].

images and results from his work [16], and and Dr. Babak Mahdian for sharing with us the MATLAB code of his work [22]. We would also like to thank the anonymous reviewers and the shepherd for constructive comments that helped us improve the current work. This work is supported by an NSF CAREER Award (IIS-0953373).

Appendix A. Derivation of Eq.(1)

In this appendix, we derive the general relationship between kurtosis and variance in noise corrupted signals. First, with signal $\mathbf{x}$ and white Gaussian noise $\mathbf{z}$ of zero mean and variance σ^2, we denote $\mathbf{y} = \mathbf{x}+\mathbf{z}$ as the noise corrupted signal. Assuming independence of $\mathbf{z}$ to $\mathbf{x}$ and the additivity of the fourth order cumulants (computed as $\mu_4 - 3\sigma^4$) of independent random variables [8], we have

$$\begin{aligned}\mu_4(\mathbf{y}) - 3\sigma^4(\mathbf{y}) &= \mu_4(\mathbf{x}) - 3\sigma^4(\mathbf{x}) + \mu_4(\mathbf{z}) - 3\sigma^4(\mathbf{z}) \\ \sigma^2(\mathbf{y}) &= \sigma^2(\mathbf{x}) + \sigma^2(\mathbf{z}).\end{aligned}$$

Further, because $\mathbf{z}$ is a Gaussian variable, we have $\mu_4(\mathbf{z}) = 3\sigma^4(\mathbf{z})$. Expressing the fourth central moment as $\mu_4 = (\kappa+3)\sigma^4$, we have

$$\begin{aligned}&(\kappa(\mathbf{y}) + 3)\sigma^4(\mathbf{y}) - 3\sigma^4(\mathbf{y}) \\ &= (\kappa(\mathbf{x}) + 3)(\sigma^2(\mathbf{y}) - \sigma^2(\mathbf{z}))^2 - 3(\sigma^2(\mathbf{y}) - \sigma^2(\mathbf{z}))^2.\end{aligned}$$

Rearranging the terms yields Eq.(1).

Appendix B. Derivation of Eq.(4)

First we expand the objective function in (1) 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:

$$\frac{\partial L}{\partial \sigma^2} = 2\sum_{k=1}^{K}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right)\frac{\sqrt{\kappa}}{\tilde{\sigma}_k^2}. \quad (8)$$

$$\frac{\partial L}{\partial \sqrt{\kappa}} = 2\sum_{k=1}^{K}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right)\left(\frac{\sigma^2}{\tilde{\sigma}_k^2} - 1\right) \quad (9)$$

Setting Eq.(8) to zero, and considering $\sqrt{\kappa} > 0$, we have

$$\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}\left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa}\sigma^2}{\tilde{\sigma}_k^2}\right) = 0, \quad (10)$$

Setting Eq.(9) to zero and substituting with Eq.(10) 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

$$\sigma^2 = \frac{1}{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}} - \frac{1}{\sqrt{\kappa}}\frac{\sum_{k=1}^{K}\sqrt{\tilde{\kappa}_k}}{\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}}. \quad (11)$$

Next, substituting Eq.(11) back into Eq.(10), we have

$$\begin{aligned}\sqrt{\kappa}\left(\frac{1}{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}} - \frac{1}{\sqrt{\kappa}}\frac{\sum_{k=1}^{K}\sqrt{\tilde{\kappa}_k}}{\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2}}\right)\sum_{k=1}^{K}\frac{1}{(\tilde{\sigma}_k^2)^2} \\ + \sum_{k=1}^{K}\frac{\sqrt{\tilde{\kappa}_k}}{\tilde{\sigma}_k^2} - \sqrt{\kappa}\sum_{k=1}^{K}\frac{1}{\tilde{\sigma}_k^2} = 0.\end{aligned} \quad (12)$$

Further arranging terms and replacing average over different channels with $\langle\cdot\rangle_k$ yield Eq.(4). Further checking the second-order conditions ensures that the solution is the unique global minimizer of Eq.(1).

References

[1] S. Bayram, I. Avcibas, B. Sankur, and N. D. Memon. Image manipulation detection. *J. Electronic Imaging*, 15(4):1–17, 2006. 2, 7

[2] T. R. Benedict and T. T. Soong. The joint estimation of signal and noise from the sum envelope. *IEEE Trans. Inform. Theory*, 13:447–454, 1967. 2, 3

[3] M. Bethge. Factorial coding of natural images: how effective are linear models in removing higher-order dependencies? *J. Opt. Soc. Am. A*, 23(6):1253–1268, 2006. 1, 2

[4] P. Burt and E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, 31(4):532–540, 1981. 1, 3

[5] M. Chen, J. Fridrich, M. Goljan, and J. Lukas. Determining image origin and integrity using sensor noise. *IEEE Transactions on Information Forensics and Security*, 3(1):74–90, March 2008. 2

[6] D.L.Donoho. De-noising by soft-thresholding. *IEEE Trans. on Info. Theo.*, 41(3):613–627, 1995. 2

[7] H. Farid. Photo fakery and forensics. In *Advances in Computers*. 2009. 1

[8] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, January 1968. 3, 9

[9] T. Filler, J. J. Fridrich, and M. Goljan. Using sensor pattern noise for camera model identification. In *IEEE International Conference on Image Processing, San Diego, California*, 2008. 2

[10] R. Franzen. Kodak lossless true color image suite. source: http://r0k.us/graphics/kodak, 1999. 4

[11] D. Fu, Y. Q. Shi, and W. Su. 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*, 2007. 2, 7

[12] J. He, Z. Lin, L. Wang, and X. Tang. Detecting doctored JPEG images via DCT coefficient analysis. In *ECCV*, 2006. 2

[13] Y.-F. Hsu and S.-F. Chang. Detecting image splicing using geometry invariants and camera characteristics consistency. In *International Conference on Multimedia and Expo*, 2006. 2, 5, 6, 7

[14] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999. 2

[15] J. Immerkær. Fast noise variance estimation. *Computer Vision and Image Understanding*, 64(2):300–302, 1996. 2

[16] Z. Lin, J. He, X. Tang, and C. Tang. Fast, automatic and fine-grained tampered JPEG images detection via DCT coefficient analysis. *Pattern Recognition*, 42(11):2492–2501, 2009. 2, 8, 9

[17] Z. Lin, R. Wang, X. Tang, and H. Shum. Detecting doctored images using camera response normality and consistency. In *CVPR*, 2005. 2, 7

[18] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:299–314, 2008. 2, 7

[19] J. Lukas, J. Fridrich, and M. Goljan. Detecting digital image forgeries using sensor pattern noise. In *Proceedings of SPIE Security, Steganography, and Watermarking of Multimedia Contents VIII*, 2006. 2

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

[21] S. Lyu and E. P. Simoncelli. Nonlinear extraction of 'independent components' of natural images using radial Gaussianization. *Neural Computation*, 18(6):1–35, 2009. 1, 2

[22] B. Mahdian and S. Saic. Using noise inconsistencies for blind image forensics. *Image and Vision Computing*, 27(10):1497–1503, 2009. 2, 7, 9

[23] R. Matzner and F. Engleberger. An snr estimation algorithm using fourth-order moments. In *IEEE Int. Symp. Information Theory*, 1994. 2

[24] T.-T. Ng and S.-F. Chang. A model for image splicing. In *IEEE International Conference on Image Processing (ICIP)*, Singapore, 2004. 2, 7

[25] S. I. Olsen. Estimation of noise in images: An evaluation. *Comput. Vision Graphics Image Process. Graphic Models and Image Process*, 55(4):319–323, 1993. 2

[26] X. Pan, X. Zhang, and S. Lyu. Exposing image forgery with blind noise estimation. In *The 13th ACM Workshop on Multimedia and Security*, Buffalo, NY, 2011. 2, 3, 7

[27] D. R. Pauluzzi and N. C. Beaulieu. A comparison of snr estimation techniques for the AGWN channel. *IEEE Transactions on Communications*, 48(10):1681–1691, 2000. 2

[28] A. Popescu and H. Farid. Statistical tools for digital forensics. In *6th International Workshop on Information Hiding*, Toronto, Canada, 2004. 2, 7

[29] K. Rank, M. Lendl, and R. Unbehauen. Estimation of image noise variance. In *IEE Vis. Image Signal Process*, volume 146, pages 80–84, 1999. 2

[30] G. Schaefer and M. Stich. UCID - an uncompressed colour image database. In *Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia*, 2004. 4, 6

[31] A. Stefano, P. White, and W. Collis. Training methods for image noise level estimation on wavelet components. *EURASIP Journal on Applied Signal Processing*, 16:2400– 2407, 2004. 2

[32] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings: Biological Sciences*, 265(1394):359–366, Mar 1998. 4

[33] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 2002. 4

[34] D. Zoran and Y. Weiss. Scale invariance and noise in nature image. In *IEEE International Conference on Computer Vision, Kyoto, Japan*, 2009. 1, 2, 3, 4, 7, 8