Exposing Image Forgery with Blind Noise Estimation

Xunyu Pan
Computer Science Department
University at Albany, SUNY
Albany, NY 12222, USA
xypan@cs.albany.edu

Xing Zhang
Computer Science Department
University at Albany, SUNY
Albany, NY 12222, USA
xz654242@albany.edu

Siwei Lyu
Computer Science Department
University at Albany, SUNY
Albany, NY 12222, USA
lsw@cs.albany.edu

ABSTRACT

Noise is unwanted in high quality images, but it can aid image tampering. For example, noise can be intentionally added in image to conceal tampered regions or to create special visual effects. It may also be introduced unnoticed during camera imaging process, which makes the noise levels inconsistent in splicing images. In this paper, we propose a method to expose such image forgeries by detecting the noise variance differences between original and tampered parts of an image. The noise variance of local image blocks is estimated using a recently developed technique [1], where no prior information about the imaging device or original image is required. The tampered region is segmented from the original image by a two-phase coarse-to-fine clustering of image blocks. Our experimental results demonstrate that the proposed method can effectively detect image forgeries with high detection accuracy and low false positive rate both quantitatively and qualitatively.

Categories and Subject Descriptors

I.4 [**Image Processing**]: Miscellaneous

General Terms

Security

Keywords

Image Forensics, Noise Estimation, Unsupervised Learning

1. INTRODUCTION

With the rapid growth of the Internet and the popularity of digital imaging devices, digital imagery has become our major information source. Meanwhile, the development of image manipulation techniques employed by most image editing software brings new challenges to the credibility of photographic images as the definite records of events. Consequently, forensic tools aiming to verify the integrity of the digital images are in high demand and have hence drawn significant attention.

MM&Sec'11, September 29–30, 2011, Buffalo, New York, USA.

Figure 1: Two image forgeries. (left) A forgery is generated with artificially added noise to create a falling snow/rain effect. (right) The image region of a toy puppy cropped from one source image is spliced with the other source image to make a composite image, where the noise levels in the two source images are distinct.

Image noise is the variation of brightness of pixels intrinsic in the image acquisition and processing processes. Due to the inherent characteristics of each individual camera sensor, the variance of noise in an untampered image is in general uniform across the entire image. When image regions from different images with different intrinsic noise levels are combined together to create a forgery, or noises are intentionally added in forged regions to conceal tampering [2] or to add special visual effects, the inconsistency of noise level in different regions of the image can be used to expose the tampered regions. Fig.1 shows two image forgeries in which the image noise is involved. For the example shown in the left panel, a falling rain/snow appearance is created on the window using artificially added noise for special visual effect. In the right panel, a splicing image is generated from two individual source images with different noise levels, where the image noise is introduced unintentionally.

In this work, we propose a new detection method to effectively locate image forgeries based on inconsistency in image noise levels. The image in question is first segmented into image blocks for initial noise estimation by the technique introduced in [1], and clustered into clean and tampered blocks. The detected suspicious regions are further segmented into smaller blocks for refined noise estimation and classification in the second phase to obtain final detection results. Our experiments show that this coarse-to-fine strategy improves the detection accuracy and reduces the computation complexity.

2. BACKGROUND

Noise is usually artificially added to help make digital image forgeries. Digital cameras may also bring noise automatically during the imaging process. Both artificially added and automatically introduced noise can be utilized to detect image tampering.

One can manipulate digital image by adding noise either to hide the results of image tampering or to create special visual effects. First, after the process of image manipulation, tampered regions usually exhibit suspicious statistical characteristics compared to the surrounding areas. To remove the traces of image manipulation, one easy and direct approach is to conceal the tampered regions with artificially added noise [2]. Second, the development of advanced image editing software helps improve the image viewing quality by conveniently creating special visual effects in images, where image noise plays an important role in many of these tampering operations. For example, with just a few operations using Photoshop, a convincing forgery with a falling rain/snow effect can be generated. In both manipulation cases, the intentionally added noise can be detected and hence serves as a strong evidence of image tampering.

On the other hand, the image noise may be introduced automatically by cameras themselves. ISO speed, a term referring to the sensitivity of a film to light, was widely used in the age of film cameras. Nowadays, as digital cameras dominate the market, ISO speed is still a critical factor affecting the quality of output photographs. ISO speed measures the sensitivity of imaging sensor, which is a important factor in determining the shutter speed and aperture in order to set appropriate exposure time. To obtain high quality photos, ISO speed is set to proper values under various light conditions. Generally, high ISO mode is typically used in low light conditions where the imaging sensor is set to more sensitive to light, which however makes the output photos noisier. On the other hand, low ISO mode is used when the light is strong, which results in finer grains in the output photos due to the lower sensitivity of the sensor to the light. A splicing forgery is usually generated from two individual source images captured by different cameras in different time, where the shooting scenarios are typically not exactly same. The inconsistency of noise level within the image hence can be used as evidence to identify the tampered regions. As the noise level difference between source images is usually unnoticed from the tampering perspective, this technique can be very effective for image splicing forgery detection.

3. PREVIOUS WORK

In digital image forensics, image noise has been widely used for source identification and manipulation detection. Related prior work falls roughly into three major categories.

In the first category, noise is used as a distinct feature for camera model identification. In [3], the photo-response nonuniformity (PRNU), which is a unique stochastic characteristic of imaging sensors, is employed as an intrinsic fingerprint to identify the source camera for a given image by pattern correlation. The method is improved in [4] for PRNU estimation with fewer training images and further used for image tampering detection. The major limitation of this type of methods is that they depend on the knowledge of specific camera models.

Recently, another category of methods are developed, which use extraction of additional noise features plus the aid of supervised learning algorithm. In [5], demosaicing characteristics are combined with PRNU in a two-round learning process to identify camera model. The method is further integrated into a single classification model [6] using support vector machine with a radial basis function. This class of techniques is extended for digital scanner identification in [7, 8]. By extracting the statistical noise features from image denoising operations, wavelet analysis, and neighborhood predication, another feature based approach [9] is proposed to detect image tampering. However, the supervised learning method does not provide the exact extent and location of the tampered regions. Another limitation is that only several specific camera models are examined by the learning algorithm, while the detection performance on other random selected camera models is unknown.

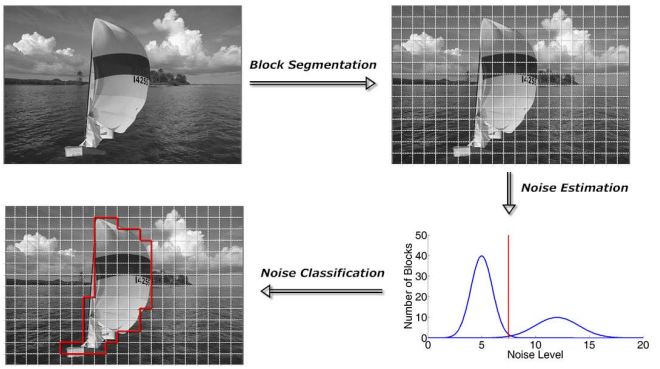

Figure 2: Illustration of the proposed image forgery detection method.

In the last category, image noise variance is estimated at local image blocks to locate suspicious regions. In [2], noise variance is estimated by computing the second and fourth moments at each local image block. But the method assumes that the kurtosis of the original signal and of the noise is known. Another recently developed method [10] uses a median-based estimator to compute the variance of noise at each image block in high frequency sub-band of the wavelet transformed image. These image blocks are then merged by examining noise difference between neighboring blocks to form various homogenous regions. Although this method can locate tampered regions, the detection accuracy and the false positive rate are not extensively evaluated. Meanwhile, the computation of the proposed algorithm can be expensive due to the inefficient region merging algorithm.

4. METHOD

In our method, an image is first segmented into non-overlapping image blocks. The noise variance at each local image block is computed using an effective noise estimation method [1]. A clustering step is then employed to separate these image blocks into clusters based on the similarity of their estimated noise. The primary steps involved in the proposed method is summarized in Fig.2.

4.1 Image Noise Estimation

It has been widely observed that the second order statistics for natural images are invariant with regards to scale [11]. Empirical observations suggest that the kurtosis (related with fourth-order cumulant) of natural images is also constant across different scales. Based on this assumption, a method is proposed in [1] to estimate the level of noise added to a clean image at various stages of production.

The distribution of a clean nature image x is non-Gaussian in the band pass filtered domain such as wavelet and DCT. They can be well fitted with a Generalized Laplacian Model with the density function:

$$f(x) = \frac{\alpha}{2\sigma_x \Gamma(1/\alpha)} e^{-(|x|/\sigma_x)^{\alpha}}, \quad (1)$$

where α is the shape parameter and σ_x is the scale parameter. Hence, the kurtosis of image x is represented as:

$$\kappa_x = \frac{\Gamma(1/\alpha)\Gamma(5/\alpha)}{\Gamma(3/\alpha)^2}. \quad (2)$$

Now suppose a white Gaussian noise η of zero mean and unknown variance σ_n^2 is added to the image x to obtain an image y, denote as:

$y = x + \eta$. The kurtosis of y is then computed [2] as:

$$\kappa_y = \frac{\kappa_x - 3}{\left(1 + \frac{\sigma_n^2}{\sigma_x^2}\right)^2} + 3. \quad (3)$$

The kurtosis k_y and variance σ_y^2 in Eq.(3) can be estimated by image y. As $\sigma_x^2 = \sigma_y^2 - \sigma_n^2$, so if k_x is known, we can solve a nonlinear equation to obtain σ_n^2. However, typically we don't know k_x in practice. But if assuming that the kurtosis satisfies scale invariance (i.e., be constant across different frequency bands), we can solve a nonlinear optimization problem:

$$\hat{\kappa}_x, \hat{\sigma}_n^2 = \underset{\kappa_x, \sigma_n^2}{\operatorname{argmin}} \sum_{i=2}^{N^2} \left| \frac{\kappa_x - 3}{\left(1 + \frac{\sigma_n^2}{\sigma_{y_i}^2 - \sigma_n^2}\right)} + 3 - \kappa_{y_i} \right|^2, \quad (4)$$

where the response image y_i is produced by the convolution of y with the ith filter from the $N \times N$ DCT basis.

The noise estimation algorithm over the image y can be summarized as the following major steps:

1. **Conversion to DCT domain**: Produce the response image y_i by the convolution of y with each filter i from the 8×8 DCT basis.

2. **Computation on response images**: Compute variance $\sigma_{y_i}^2$ and kurtosis κ_{y_i} for each response image y_i.

3. **Noise estimation**: The variance of the added noise and the kurtosis of the original clean image can be estimated by minimizing Eq.(4) using the MATLAB fminsearch function.

It should be noted that even though we assume the additive white Gaussian noise in the pixel domain, this is not as restricted as it seems, as very non-Gaussian independent noise in the pixel domain will mix in to be Gaussian noise in the filter domain due to the central limit theorem and noise independence.

4.2 Block Size Evaluation

To locate suspicious image regions, the tampered image is segmented into non-overlapping square image blocks for local noise estimation. Generally, the accuracy of noise estimation relies on the size of image block. To select an appropriate block size for the image segmentation, we evaluate the estimation performance by applying the noise estimation algorithm [1] on randomly selected image blocks with various sizes.

More specifically, we randomly crop square regions from the 25 sample images in KODAK dataset[1] to produce the image blocks for size evaluation. The source images in the KODAK dataset are typical nature images where both homogeneous and texture areas are included. The image blocks generated are of size 16×16, 32×32, 64×64, 128×128 and 256×256 pixels respectively. For each type of block size, we produce 100 image blocks which are randomly cropped at different locations from the images in KODAK dataset. The cropped image blocks are processed by adding zero mean white Gaussian noise with standard deviation $\sigma = 25$. The means and standard deviations of the estimated noise level at each of the 100 image blocks with various sizes are shown in Fig.3, where the green horizontal line is the ground truth level of the adding Gaussian noise. The evaluation results demonstrate that the noise estimated for large image blocks is generally more accurate and stable than for smaller ones as more statistical information is utilized.

[1]image source: `http://r0k.us/graphics/kodak/`.

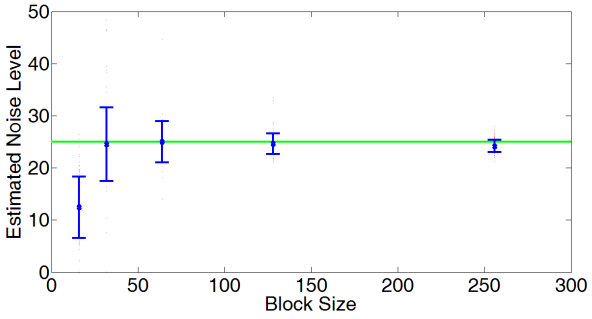

Figure 3: The means and standard deviations of the noise levels estimated for image blocks with various sizes.

Based on the experimental results on block size evaluation, a two-phase coarse-to-fine strategy is implemented to detect image forgeries. We first segment the noise tampered image into 64×64 pixel blocks for initial detection. In the second phase, only suspicious image regions revealed by the initial detection are further segmented into 32×32 pixel blocks for refined detection.

4.3 Initial Detection

Our method begins by segmenting the image into 64×64 pixel image blocks for initial detection because the blocks of this size can reach the best tradeoff between the stability of the noise estimation and the precision of locality. The noise variance for each of these blocks is then computed using the estimation algorithm of [1] that is briefly described in section 4.1. Due to the noise introduced during the tampering process, the image blocks located within the tampered regions generally have distinct noise variances compared with the uniform noise variance across the image. In case of image splicing, the noise variance of the image region cropped from one source image is typically different from that of the remaining part of that image which is obtained from the other source image.

We then apply the k-means algorithm to classify the estimated noise levels and group all image blocks into two clusters. The cluster with fewer blocks is treated as tampered region, assuming that the area of the tampered region is usually smaller than their authentic counterparts. Note that the noise level of image regions with complex textures or edges may also have different noise levels, and become false detections. Consequently, we apply a refined detection in the second step to further improve overall detection accuracy by reducing false positives.

4.4 Refined Detection

In the second phase of our tampering detection method, we process a refined detection using smaller image blocks to further improve overall detection accuracy. More specifically, for each of those 64×64 pixel blocks within the suspicious image regions located by the initial detection, we further segment them into 4 non-overlapping 32×32 pixel image blocks and estimate the noise variance for each block using the same algorithm specified in section 4.1. Due to the noise variance in 32×32 pixel blocks being related to the enclosing 64×64 blocks, our estimation is a weighted sum of the two estimations, which is 20% of the 32×32 pixel blocks and 80% 64×64 pixel blocks.

We further classify these 32×32 pixel blocks using a k-means algorithm based on both their noise variance and image coordinates, where neighboring blocks have higher probability to be classified as belonging to a same cluster. To avoid parts of tampered region

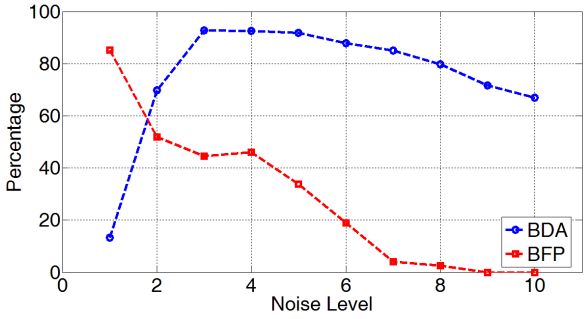

Figure 4: The BDA/BFP rates for image forgery tampered at noise level with standard deviations from 1 to 10.

being removed falsely, we process an additional step to examine the two clusters of image blocks classified by the k-means algorithm. More specifically, denote the c_1 and c_2 as the mean value of the estimated noise levels of the blocks in these two clusters respectively, we keep the two clusters when $|c_1 - c_2| \leq \sigma_l$, where σ_l is the noise standard deviation of the cluster with more image blocks. The 32×32 pixel blocks with sufficiently different noise variance and spatial distance from the bulk of the suspicious image blocks are treated as non-tampered and removed. The image regions covered by the remaining image blocks are the final detected noise-tampered regions.

For images with multiple regions tampered by noise with different variances, we run our detection algorithm several rounds with the detected tampered regions masked out from the next round search. The whole algorithm stops when no tampered region is found larger than a preset area threshold.

5. EXPERIMENTS

In this section, we evaluate the proposed detection method experimentally. We first generate a set of forged images based on one uncompressed true color image in KODAK dataset. A randomly chosen square region of size 192×192 pixels corresponding to 3.12% of total image area is tampered with white Gaussian noise, with variances in the range of [1, 10] with step size 1. For each of the 10 noise levels, we generate 100 tampered images, resulting in a total 1000 forged images.

We use two quantitative measures to evaluate the performance of our forgery detection method. Denote Ω as the number of image blocks in the true noise tampered region, and $\tilde{\Omega}$ as the number of image blocks in the detected region. We define the *block detection accuracy* (BDA) rate as the fraction of image blocks in noise-tampered region that are correctly identified, i.e., BDA $= \frac{|\tilde{\Omega} \cap \Omega|}{|\Omega|}$ and the *block false positive* (BFP) rate as the fraction of image blocks in untampered region that are detected falsely as from noise-tampered region, i.e., BFP $= \frac{|\tilde{\Omega} - \Omega|}{|\tilde{\Omega}|}$.

To reduce the effect of random samples, each pair of BDA/BFP rates is computed as the averages over all 100 forged images tampered at each noise level. The resulting BDA/BFP curves for various added noise levels are shown in Fig.4. It can be observed from the figure that the BDA/BFP rates are usually sufficient to identify the tampered regions visually for noise standard deviation larger than 2. The BFP rate drops sharply to 0% as the variance of added noise becomes distinct and hence easy to detect. On the other hand, the BDA also drops slowly but keeps relatively stable around 70% as the added noise level increases. This is expected, as the absolute noise estimation error becomes large when the noises with higher

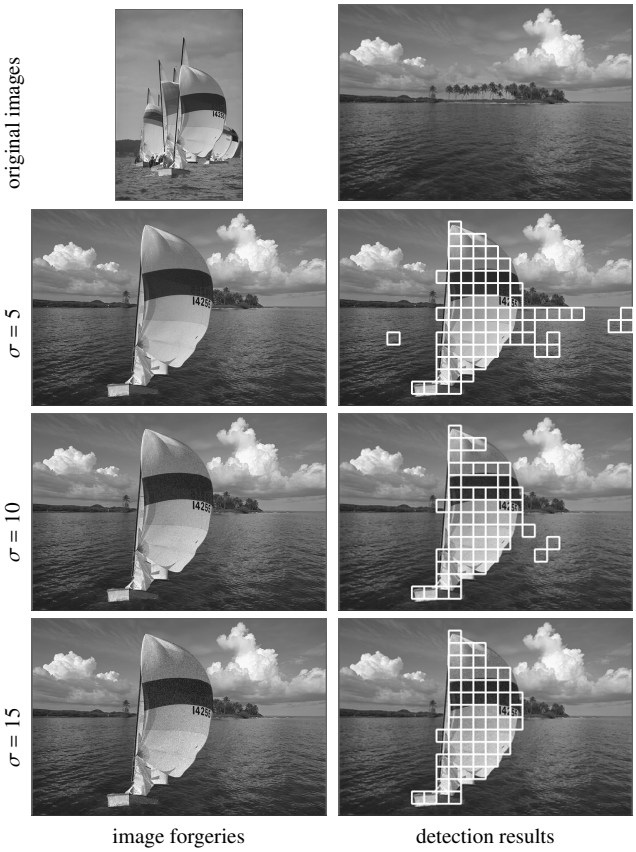

Figure 5: Image forgeries tampered with various levels of added noise and the detection results using our method. See text for details.

variance are added, which makes the partition of the tampered and untampered region more difficult.

Furthermore, we evaluate our forgery detection method qualitatively on a set of forgery created by the splicing of two individual images. As shown in the left column of Fig.5, one authentic image in the Kodak dataset is spliced with an image region cropped from another authentic image in the same dataset to generate forged images. During the splicing, the cropped region is processed by adding zero mean white Gaussian noise with standard deviation σ = 5, 10 and 15 respectively. Shown in the right column of Fig.5 are the detection results of the proposed method on these forged images. It can be observed that the overall detection accuracy is improved for the added noises with higher variance because they are relatively easy to detect.

As a more realistic test, we applied our method to detect image splicing where the forgery is created with individual source images captured in distinct shooting scenarios. Shown in Fig.6 is the detection result of a forged image obtained by the splicing of two individual images downloaded from image hosting website Flickr.com, where the original image 1 was captured by a SONY DSC-H20 digital camera with ISO speed 400 and the original image 2 was captured by a Canon EOS-60D digital camera with ISO speed 1600. The forgery is generated by the splicing of the image region of a toy puppy cropped from the image 2 with the general background of the image 1. The size of the cropped image region is also scaled down to fit the background better. The detection result shows that our detection method can accurately expose the extent and location of the tampered region based on the inconsistency of noise level in different portions of the forged image. Note that the image noise

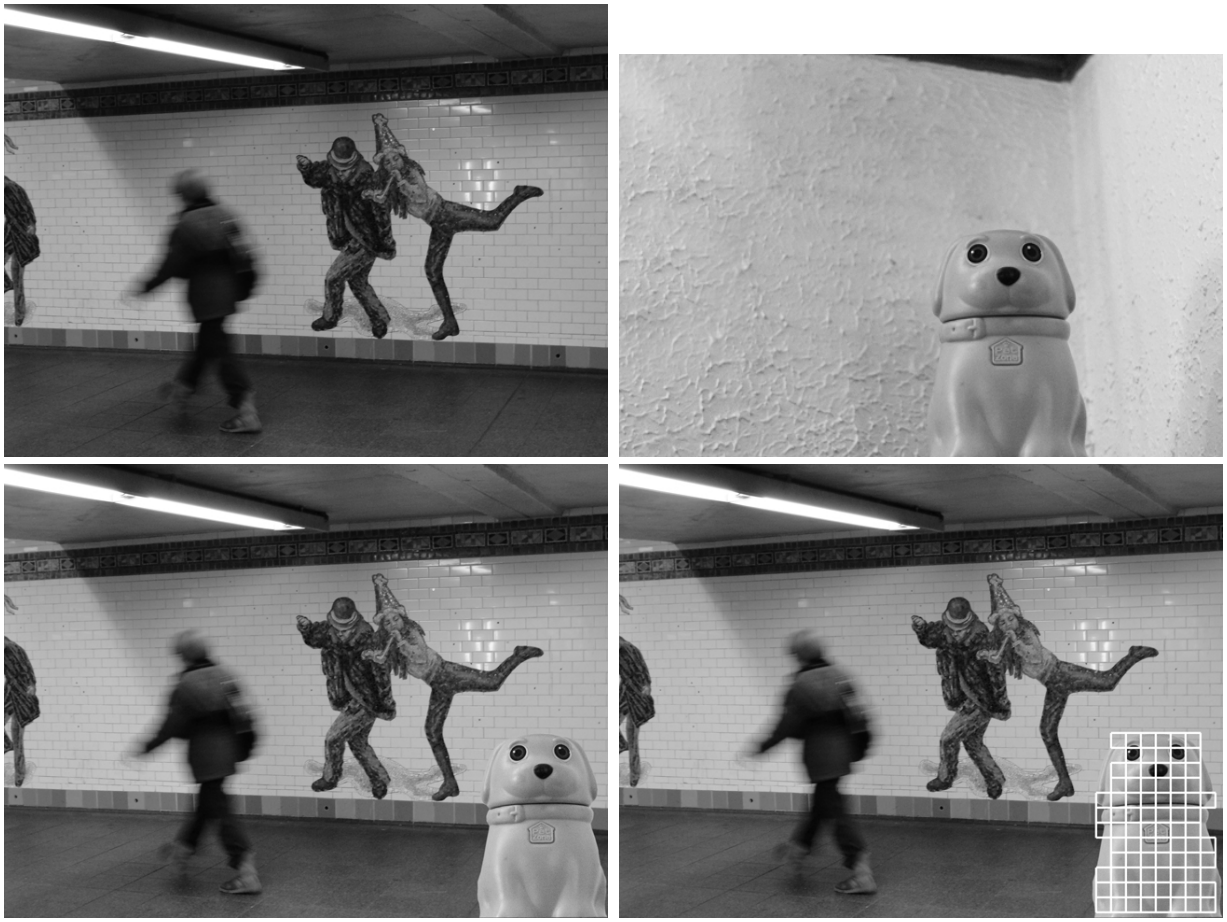

Figure 6: Detection result of our method on one realistic image splicing forgery. (top left) Original image 1 captured by a SONY DSC-H20 digital camera with ISO speed 400. (top right) Original image 2 captured by a Canon EOS-60D digital camera with ISO speed 1600. (bottom left) A spliced image generated from part of the image 1 and of the image 2. (bottom right) Detection result. See text for details.

automatically introduced by the imaging device of digital camera is usually non-Gaussian and not simply additive. Fortunately, our method makes no assumption on the form of the underlying distribution of noise in the pixel domain and hence can detect the splicing forgery simply based on the noise level differences, where the ISO speed plays an important role.

With the aid of sophisticated photo editing software Photoshop, we may create more interesting forgery in single image by some latest tampering techniques widely used in Photoshop user groups. An example is shown in Fig.7, where a falling rain/snow appearance is created on a kitchen window. As one of the most important step, the image noise is artificially added to the original image during the manipulation process. The detection results demonstrate that the proposed method can also find the tampered region with reasonable accuracy.

6. DISCUSSION AND FUTURE WORK

In this paper, we describe a novel method for image forgery detection based on the clustering of image blocks with different noise variances. The variance of image noise at each local image block is estimated blindly using properties of natural images [1]. Our method finds both the extent and location of the tampered region by clustering image blocks according to their estimated noise levels. The advantage of the proposed method is that it requires no prior knowledge of the imaging device or of kurtosis of the original image. Experimental results show that the proposed method can expose tampered regions concealed by image noise or forgeries created using artificially added noise for special visual effects. The forgeries generated by image splicing can also be effectively identified using our method based on the inconsistency of noise level in the spliced image regions.

For future works, one direction we would like to further study is if our current detection method is applicable to detect noise inconsistency due to different JPEG compression qualities, as this is usually the case when two image regions are spliced together. We would also like to study the robustness of the proposed algorithm with regards to several imaging conditions. Finally, we are also interested in combining the current detection method with other noise based features to increase the detection performance and the range of the forensic applications.

7. ACKNOWLEDGMENTS

This work was supported by an NSF CAREER Award (IIS09-53373) and by the University at Albany Faculty Research Awards Program (FRAP)-Category A.

8. REFERENCES

[1] Daniel Zoran and Yair Weiss, "Scale invariance and noise in nature image," in *IEEE International Conference on Computer Vision, Kyoto, Japan*, 2009.

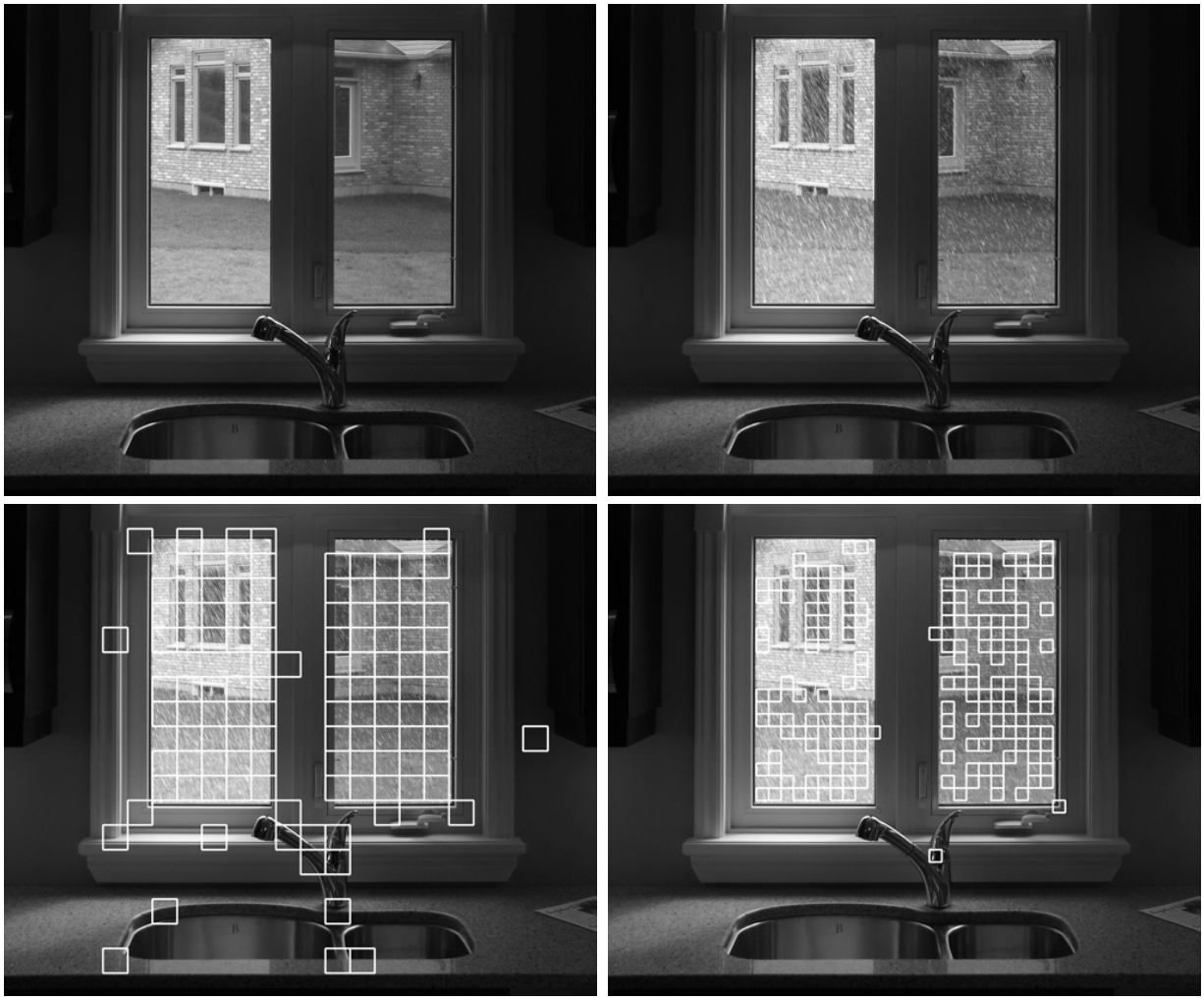

Figure 7: Detection results of our method on one forged image with special visual effects. (top left) The original image. (top right) Falling rain/snow appearance created using Photoshop. (bottom left) Initial detection using 64×64 pixel segmentation blocks. (bottom right) Refined detection using 32×32 pixel segmentation blocks. See text for details.

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

[3] Jan Lukás, Jessica J. Fridrich, and Miroslav Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[4] Mo Chen, Jessica J. Fridrich, Miroslav Goljan, and Jan Lukás, "Determining image origin and integrity using sensor noise," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74–90, 2008.

[5] Yagiz Sutcu, Sevinc Bayram, Husrev T. Sencar, and Nasir D. Memon, "Improvements on sensor noise based source camera identification," in *IEEE International Conference on Multimedia and Expo, Beijing, China*, 2007.

[6] Tomás Filler, Jessica J. Fridrich, and Miroslav Goljan, "Using sensor pattern noise for camera model identification," in *IEEE International Conference on Image Processing, San Diego, California*, 2008.

[7] Hongmei Gou, Ashwin Swaminathan, and Min Wu, "Intrinsic sensor noise features for forensic analysis on scanners and scanned images," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 476–491, 2009.

[8] Nitin Khanna, Aravind K. Mikkilineni, George T. C. Chiu, Jan P. Allebach, and Edward J. Delp, "Scanner identification using sensor pattern noise," in *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, California*, 2007.

[9] Hongmei Gou, Ashwin Swaminathan, and Min Wu, "Noise features for image tampering detection and steganalysis," in *IEEE International Conference on Image Processing, San Antonio, Texas*, 2007.

[10] Babak Mahdian and Stanislav Saic, "Using noise inconsistencies for blind image forensics," *Image and Vision Computing*, vol. 27, no. 10, pp. 1497–1503, 2009.

[11] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communication*, vol. 31, no. 4, pp. 532–540, 1981.