

Detecting Image Forgery Based on Noise Estimation

Yongzhen Ke^{1*}, Qiang Zhang¹, Weidong Min¹ and Shuguang Zhang²

¹*School of Computer Science and Software Engineering, Tianjin Polytechnic University, China*

²*JiangSu Wisedu Information Technology Co., Ltd, Jiangsu, China*

keyongzhen@tjpu.edu.cn

Abstract

With the advent of the Internet and low-price digital cameras, as well as powerful image editing software, the authenticity of digital images can no longer be taken for granted. Image noises are often introduced into the tampered region during image manipulation process. In this paper, we propose a detection method to locate image forgeries based on noise estimation on HSV color space and hybrid clustering method combined with unsupervised clustering and supervised clustering. A suspicious image is first converted into HSV color space and segmented into non-overlapping image blocks. Then the noise variance at each local image block is estimated as input of unsupervised clustering. Finally, a supervised clustering method based on SVM is used to further improve the detection accuracy. Our experimental results demonstrate that the proposed method can effectively expose tampered regions from tampered images.

Keywords: *Image forensic, Image manipulation, Tampering detection, Noise estimation, SVM*

1. Introduction

Due to the advent of low-cost digital cameras and the availability of many powerful photo editing tools, it becomes easier to create forgery images by non-professional users. Forgery images are made in such a way that it is very difficult to identify the forgery part of the manipulated images at the first sight. Because image authenticity is significant in many social areas and plays a crucial role in people's lives, there is the need to identify the forgery portions of the manipulated images. A commonly used method to conceal the traces of tampering is the addition of locally random noise to the altered image regions. Typically, the amount of noise is uniform across the entire authentic image. Due to the noise introduced during the tampering process, the image blocks located within the tampered regions generally have distinct noise variance compared with the uniform noise variance across the image. This may cause inconsistencies in the image's noise variance. The inconsistency of the noise variance within the image can hence be used as evidence to identify the tampered regions. Furthermore, when two or more images from different sources are spliced together, the forged image may then contain several regions with various noise variances as well. Therefore, the detection of various noise variances in an image may be very effective for image forgery detection [1-4].

In this paper, we propose a detection method to effectively locate image forgeries by detecting inconsistency of image noise variance on the saturation component of HSV color space. The image is first converted to HSV color space from RGB color space. Then it is

segmented into non-overlapping image blocks. The noise variance at each local image block is computed using an effective noise estimation method [5]. An unsupervised clustering step is then employed to separate these image blocks into clusters based on the similarity of their estimated noise. Finally, a supervised clustering method based on SVM is used to further improve the detection accuracy.

This paper is organized as follows. Section 2 presents the related works. The proposed methods are described in detail in Section 3. Experimental procedure and results are discussed in Section 4. Finally, conclusions and future work are drawn in Section 5.

2. Related Work

Image noise detection has been widely used for image source identification and forgery detection. The photo-response non-uniformity (PRNU), which is a unique stochastic characteristic of imaging sensors, was employed as an intrinsic fingerprint to identify the source camera for a given image [6]. Furthermore, Mo Chen *et al.*, [7] described a unified framework for both device identification and integrity verification using pixel PRNU. Y. Sutcu *et al.*, [8] proposed a method that demosaicing characteristics were combined with PRNU in a two-round learning process to identify camera model. Hongmei Gou *et al.*, [1] introduced a novel approach for tampering detection and steganalysis on digital images using three sets of noise features. They explored the denoising algorithms to obtain image noise estimation. The second set of features was obtained by wavelet analysis and the third was obtained by utilizing prediction errors of neighborhood. Using these features a classifier was built to distinguish direct camera output from their tampered or stego versions. However, the supervised learning method did not provide the exact extent and location of the tampered regions. Another limitation is that only several specific camera models were examined by the learning algorithm.

Using image noise variance estimation at local image blocks to locate suspicious regions has been introduced in [2-4]. Popescu and Farid [4] proposed a noise inconsistencies detection method based on estimating the noise variances of overlapping blocks by which they tiled the entire investigated image. Their method used the second and fourth moments of analyzed block to estimate noise variance. Their method assumes white Gaussian noise and non Gaussian uncorrupted image. Their method also assumes that kurtosis of the original signal is known, which is mostly not true in practice. Another recently developed method capable of detecting image noise inconsistencies was proposed in [2] by Babak Mahdian and Stanislav Saic. They introduced a segmentation method detecting changes at noise level. The local noise estimation was based on tiling the high pass diagonal wavelet coefficients at the highest resolution with non-overlapping blocks. The noise standard deviation in each block was estimated using a widely used median-based method. Once the noise standard deviation of each block was estimated, it was used as the homogeneity condition to segment the investigated image into several homogenous sub-regions. This was carried out by a simple regions-merging segmentation technique. Although this method could locate tampered regions, the threshold must be carefully selected. Otherwise, it was difficult to separate the tampered region from rest of the image. Xunyu Pan *et al.*, [3] proposed a new image forgery detection method to effectively locate image forgeries based on the clustering of image blocks with different noise variances. The suspicious image was first segmented into image blocks for initial noise estimation by the technique introduced in [9], and clustered into clean and tampered blocks. The detected suspicious regions were further segmented into smaller blocks for refined noise estimation and classification in the second phase to obtain final detection results. Their method is based on the Kurtosis concentration property of natural image in the band pass filtered domains. Our work is similar to [3] but different in three aspects. First, our

work focuses on the saturation component of HSV color space but RGB color space. Second, we adopt the latest image noise estimation method [5] that not only has a good compromise between speed and accuracy but also can successfully process images containing only textures. Third, we propose a hybrid clustering method combined with k-means and SVM for better classification performance.

Because image manipulations like brightness and contrast adjustment can affect these noise features and lead to enlarged numerical difference between each actual and its estimated EXIF feature from the noise features, Jiayuan Fan *et al.*, [10] proposed a novel technique to correlate statistical image noise features with three Exchangeable Image File format (EXIF) header features for manipulation detection. They observe that the numerical differences between the actual EXIF features and their corresponding estimated EXIF features from the noise features can serve as an excellent indicator to determine whether an image represents the original camera recording or it has gone through the brightness and contrast adjustment. By using the numerical difference as a manipulation indicator, they achieved excellent performance in detecting common brightness and contrast adjustment. Ahmet Emir Dirik and Nasir Memon [11] proposed a detection method that did not target any specific operation but was applicable to the detection of a variety of operations such as splicing, retouching, re-compression, resizing, blurring etc. The proposed method differ from known universal tamper detection techniques [1, 12] in the sense that they did not require a complex classifier, instead they used only one threshold to make a decision about the image in question.

3. The Proposed Method

The proposed method for image forgeries detection based on noise estimation on HSV color space and hybrid clustering method combined with unsupervised clustering and supervised clustering is illustrated in Figure 1. The suspicious image is first converted to HSV color space from RGB color space. Then it is segmented into non-overlapping image blocks. The noise variance at each local image block is estimated using an effective noise estimation method [5]. A simple unsupervised clustering algorithm (k-means) is first used to classify the estimated noise variance. All image blocks are initially grouped into two clusters. Only labeled samples close to its learned cluster centroids are used to train a SVM classifier for better classification results.

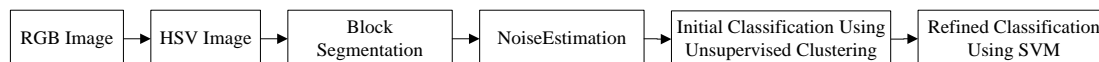


Figure 1. The proposed image forgery detection method based on noise estimation

3.1. Image Preprocessing

The Hue-Saturation-Value (HSV) color space is a different color system that is believed to be more natural than the RGB system for human perception. The three variables are: hue(H), saturation(S) and value(V). The S and V are linear features and take values in the interval [0, 1]. S encodes the “purity” of color and V stores the intensity information. The hue component H, of which the range is from 0 to 360 degree, denotes the spectral composition of color. We first convert RGB color space to HSV color space and select the saturation component of HSV color space for the next block segmentation.

3.2. Block Segmentation

To locate forged image regions, the suspicious image is segmented into non-overlapping image blocks B_i of $R \times R$ pixels for local noise estimation. Blocks are assumed to be smaller than the size of the corrupted regions to be detected. The total number of non-overlapping blocks for an image of $M \times N$ pixels is $r = \left\lfloor \frac{M}{R} \right\rfloor \times \left\lfloor \frac{N}{R} \right\rfloor$.

Generally, the accuracy of noise estimation relies on the size of image block. To obtain an appropriate block size for the image segmentation, we evaluate the noise estimation performance by applying the noise estimation algorithm [5] on randomly selected image blocks with various sizes. Twenty sample images in KODAK dataset [13] are randomly chosen to produce the image blocks for size evaluation. The image blocks generated are of size 16×16 , 32×32 and 64×64 pixels respectively. We produce 100 forged images which are randomly cropped at different locations from the images for each size. The cropped image blocks are processed by adding white Gaussian noise with standard deviation $\sigma = 5, 10, 15$ respectively. The evaluation results demonstrate that the noise estimation for image blocks with size of 32×32 achieve the best results.

3.3. Image Noise Estimation

Numerous methods have been proposed so far to perform the noise level estimation in digital images. Generally, these methods can be divided into following groups: block-based, smoothing-based and gradient-based. An effective block-based noise estimation method [5] is used in our method, which has ability to process images with textures, even if there are no homogeneous areas.

Let x be a noise-free image of size $S_1 \times S_2$, where S_1 is the number of columns and S_2 is the number of rows, $y = x + n$ be an image corrupted with signal-independent additive white Gaussian noise n with zero mean. Noise variance σ^2 is unknown and should be estimated.

Each of images x , n , y contains $N = (S_1 - M_1 + 1)(S_2 - M_2 + 1)$ blocks of size $M_1 \times M_2$, whose left-top corner positions are taken from set $\{1, \dots, S_1 - M_1 + 1\} \times \{1, \dots, S_2 - M_2 + 1\}$. These blocks can be rearranged into vectors with $M = M_1 M_2$ elements and considered as realizations x_i , n_i , y_i , $i = 1, \dots, N$ of random vectors X , N , and Y respectively. As n is signal-independent additive white Gaussian noise, $N \sim N_M(0, \sigma^2 I)$ and $\text{cov}(X, N) = 0$.

Let S_X , S_Y be the sample covariance matrices of X and Y respectively, $\lambda_{X,1} \geq \lambda_{X,2} \geq \dots \geq \lambda_{X,M}$ be the eigenvalues of S_X with the corresponding normalized eigenvectors $V_{X,1}, \dots, V_{X,M}$, and $\lambda_{Y,1} \geq \lambda_{Y,2} \geq \dots \geq \lambda_{Y,M}$ be the eigenvalues of S_Y with the corresponding normalized eigenvectors $V_{Y,1}, \dots, V_{Y,M}$. Then, $V_{Y,1}^T Y, \dots, V_{Y,M}^T Y$ represent the

sample principal components of Y , which have the property $s^2(V_{Y,k}^T Y) = \lambda_{Y,k}, k = 1, 2, \dots, M$ where s^2 denotes the sample variance.

To apply PCA to estimate noise variance, a class of noise-free images is defined. Such noise-free images satisfy the following assumption:

Assumption 1: Let m be a predefined positive integer number. The information in noise-free image x is redundant in the sense that all x_i lie in subspace $V_{M-m} \subset \mathbb{R}^M$, whose dimension $M-m$ is smaller than the number of coordinates M .

This assumption means the existence of a linear dependence between components of X , i.e. a linear dependence between pixels of x in the image blocks. This assumption also implies that X has zero variance along any direction orthogonal to V_{M-m} .

The following theorem provides a way to apply PCA for noise variance estimation:

Theorem 1: If Assumption 1 is satisfied then $E(|\lambda_{Y,i} - \sigma^2|)$ is bounded above by σ^2 / \sqrt{N} asymptotically:

$$E\left(|\tilde{\lambda}_{Y,i} - \sigma^2|\right) = O\left(\sigma^2 / \sqrt{N}\right), N \rightarrow \infty \quad (1)$$

for all $i = M - m + 1, \dots, M$.

Theorem 1 gives an asymptotic bound for the convergence speed. It states that the expected value of $|\lambda_{Y,i} - \sigma^2|$ converges to zero as σ^2 / \sqrt{N} or faster. According to Theorem 1, when assumption 1 is satisfied,

$$\lim_{N \rightarrow \infty} E\left(|\tilde{\lambda}_{Y,M} - \sigma^2|\right) = 0 \quad (2)$$

i.e. $\lambda_{Y,M}$ converges in mean to σ^2 . Therefore, the noise variance can be estimated as $\lambda_{Y,M}$. Since convergence in mean implies convergence in probability, $\lambda_{Y,M}$ is a consistent estimator of the noise variance.

3.4. Image Block Classification

Once the noise variance of each block is estimated, a simple unsupervised clustering algorithm (k-means) is first used to classify the estimated noise variance. All image blocks are grouped into two clusters. Generally, the area of the tampered region is usually smaller than their authentic counterparts. The cluster with fewer blocks is treated as tampered region. While image noise estimation method based on PCA [5] works well on the images with textures, even if there are no homogeneous areas, there is still a little false detection using k-means. Figure 2(a) shows that there exists false classification.

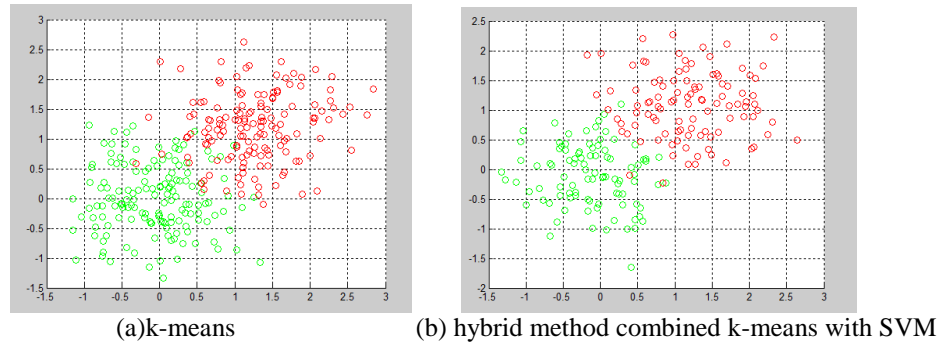


Figure 2. The result of classification

To improve the classification performance, we propose a hybrid classification method combined k-means with SVM for more accurate detection, as shown in Figure 3.

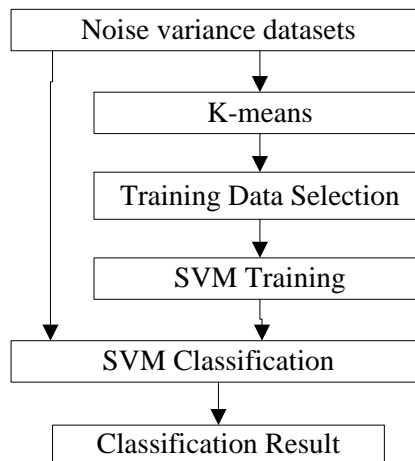


Figure 3. The framework of hybrid classification method combined k-means with SVM

After applying k-means to noise variance for obtaining labels, a training data selection step is performed for SVM training. Let the distance between two learned clusters centroids be r . We only select labeled sample that the distance to its learned cluster centroids is less than one third of r , as input of SVM training. Last, we apply trained SVM classifier to noise variance for achieving better classification results. Figure 2(b) shows the results using hybrid classification method. Compared with Figure 2(a) and Figure 2(b), we can find that improved clustering method achieve better classification performance than k-means.

4. Experimental Results and Discussion

In the following results of experiments of the proposed detection method are described. In the experiment, we generated a set of forged images based on one uncompressed true color image in KODAK dataset [13]. A randomly chosen tampered region was processed by adding zero mean white Gaussian noise with standard deviation $\sigma = 5, 10, 15$ respectively. For each

of the noise levels, we generated 150 forged images. We segmented each forged image into image blocks with size of 32×32 pixel because the blocks of this size could reach the best performance. A support vector machine (SVM) classifier with RBF kernel in the LSSVM tools was used in our experiments. We used the “grid-search” method to find the optimal parameters σ and γ of RBF kernel.

Figure 4 shows the result of our proposed method on one of these forged images. It can be observed that our proposed method can effectively locate the tampered region.

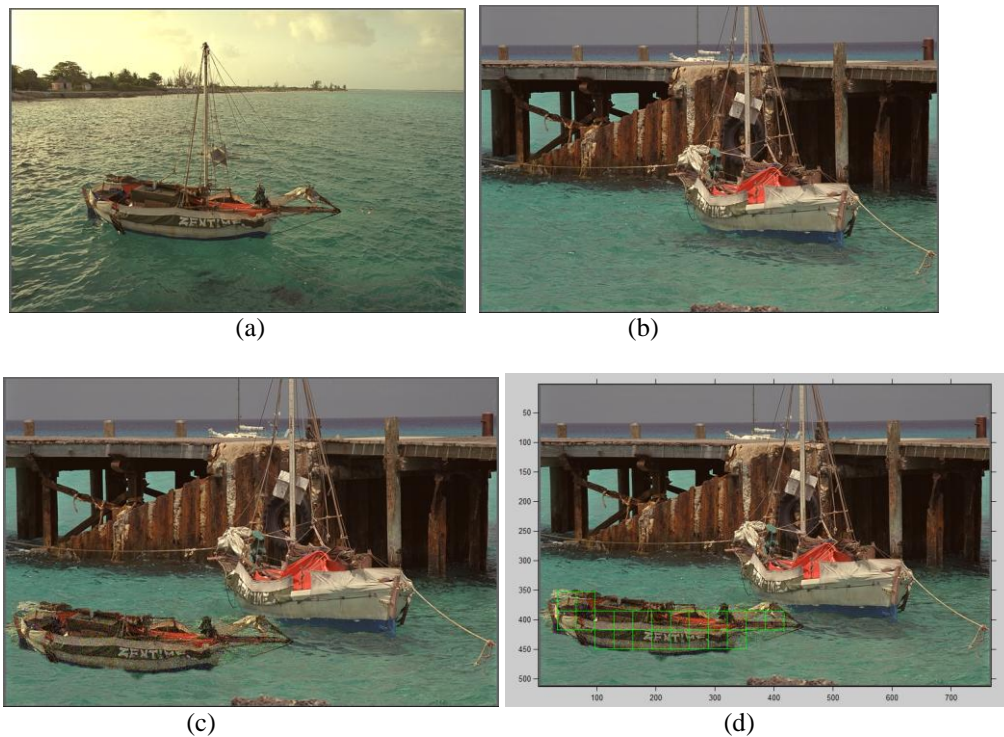


Figure 4. (a) and (b) are the original images. (c) is the tampered image and (d) is the detection result of our proposed method

Furthermore, we made comparisons with previous work [3]. The results shown in Figure 5 demonstrate that our proposed method outperforms Xunyu Pan’s method [3].



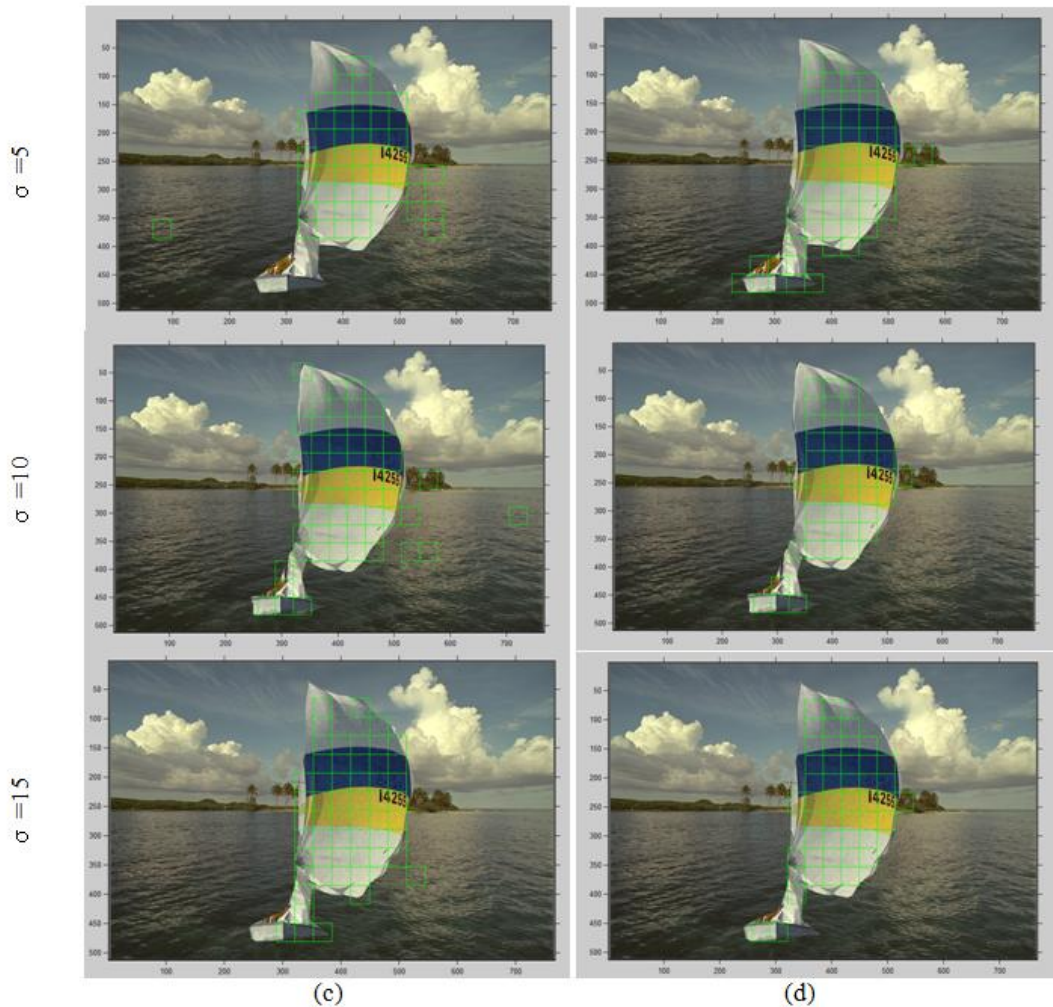


Figure 5. (a) and (b) are the original images. (c) is the detection results of Xunyu Pan's method on various levels of added noise, and (d) is our results.

One of latest tampering techniques using Photoshop can create interesting forgery in a single image. Figure 6 shows an example that a local retouch is created as falling rain on a living room window. Another example is shown in Figure 7, where waterfall is created over the statue. The image noise is artificially added to the original image during the manipulation process. The detection results demonstrate that our proposed method can also locate the tampered regions in these images with reasonable accuracy.

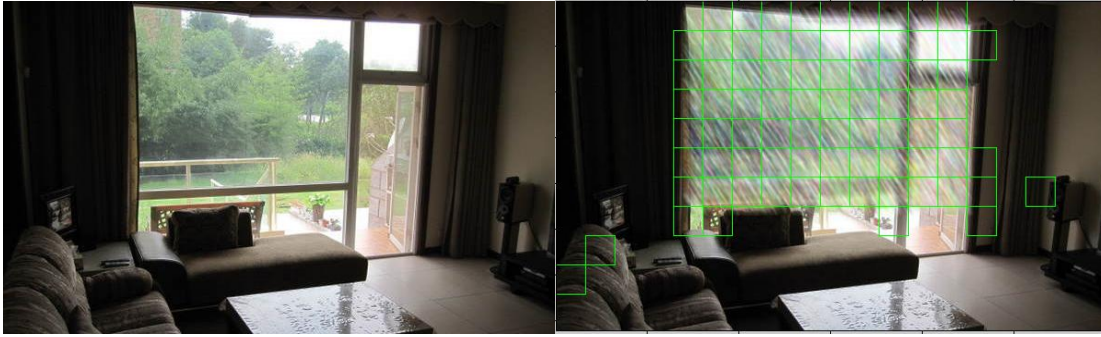


Figure 6. The detection result of our method on the window that falling is created



Figure 7. The detection result of local forgery that is created as waterfall

Blurring is a very common process in digital image manipulation. It could be used to reduce the degree of discontinuity or to remove unwanted defects, ultimately, it is used to generate plausible digital image forensics. We did another experiment of applying our proposed method on a manually blurred image. The experimental result shown in Figure 8 demonstrates that our method can also detect the tampered region of blur operation with reasonable accuracy.

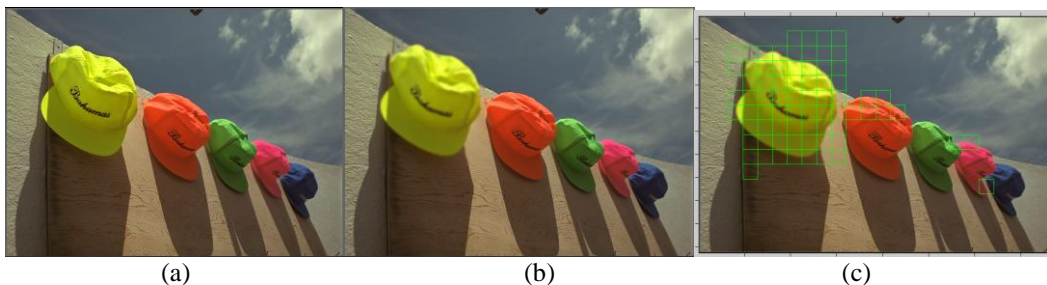


Figure 8. (a) is the original image. (b) is the local blurring on the yellow cap, and (c) is the detection result of our proposed method

5. Conclusions

In this paper, we proposed a novel approach based on noise level estimation using principal component analysis on saturation component of HSV color space for tampering detection and local retouching detection on digital images. We apply noise estimation method based on PCA to obtain the image noise variance, and classify all image blocks into two clusters using k-means. To improve the classification performance, a robust hybrid classifier combined k-means with SVM is used for more accurate detection. Experimental results show that our method can locate tampered regions such as image noise or forgeries created using artificially added noise and blurring for special visual effects. In the future work, we are interested in combining the current detection method with other blur estimation based methods to increase the detection performance and the range of the forensic applications.

Acknowledgements

This research was partially supported by the Natural Science Foundation of Tianjin, China (Grant No. 13JCYBJC15500).

References

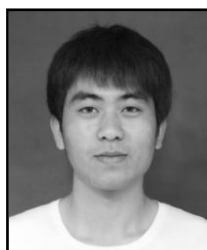
- [1] H. Gou, A. Swaminathan and M. Wu, "Noise Features for Image Tampering Detection and Steganalysis", Proceedings of the 2007 IEEE International Conference on Image Processing, (2007), pp. 2893-2896.
- [2] B. Mahdian and S. Saic, "Using Noise Inconsistencies for Blind Image Forensics", Image and Vision Computing, vol. 27, no. 10, (2009), pp. 1497-1503.
- [3] X. Pan, X. Zhang and S. Lyu, "Exposing Image Forgery with Blind Noise Estimation", Proceedings of the the Thirteenth ACM Multimedia Workshop on Multimedia and Security, (2011), pp. 15-20.
- [4] A. C. Popescu, "Statistical Tools for Digital Image Forensics", Proceedings of the 6th International Workshop on Information Hiding & LNCS, (2004), pp. 128-147.
- [5] S. Pyatykh, J. Hesser and L. Zheng, "Image Noise Level Estimation by Principal Component Analysis", IEEE Transactions on Image Processing, vol. 22, no. 2, (2013), pp. 687-699.
- [6] J. Lukáš, J. Fridrich and M. Goljan, "Digital Camera Identification from Sensor Pattern Noise", IEEE Transactions on Information Forensics and Security, vol. 1, no. 2, (2006), pp. 205-214.
- [7] M. Chen, J. Fridrich, M. Goljan and J. Lukáš, "Determining Image Origin and Integrity Using Sensor Noise", IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, (2008), pp. 74-90.
- [8] Y. Sutcu, S. Bayram, H. T. Sencar and N. Memon, "Improvements on Sensor Noise Based Source Camera Identification", Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, (2007), pp. 24-27.
- [9] D. Zoran and Y. Weiss, "Scale Invariance and Noise in Natural Images", Proceedings of the IEEE 12th International Conference on Computer Vision, (2009), pp. 2209-2216.
- [10] J. Fan, H. Cao and A. C. Kot, "Estimating EXIF Parameters Based on Noise Features for Image Manipulation Detection", IEEE Transactions on Information Forensics and Security, vol. 8, no. 4, (2013), pp. 608 – 618.
- [11] D. A. Emir and N. Memon, "Image Tamper Detection Based on Demosaicing Artifacts", Proceedings of the 16th IEEE International Conference on Image Processing (ICIP), (2009), pp. 1497-1500.
- [12] S. Bayram, I. Avcibas, B. Sankur and N. Memon, "Image Manipulation Detection", Journal of Electronic Imaging, vol. 15, no. 4, (2006), pp. 041102-041102-17.
- [13] Kodak Lossless True Color Image Suite, <http://r0k.us/graphics/kodak/>.

Authors



Yongzhen Ke

He obtained his BE of computer application at Tianjin Polytechnic University in China in 1997, and received his ME and PhD of computer application at Tianjin University in China in 2000 and 2008, respectively, on the research subjects of information security, image processing and data visualization. From 1997 to the present he worked in Tianjin Polytechnic University, China. Now he is an associate professor at School of Computer Science & Software Engineering, Tianjin Polytechnic University, China. His current research interests include image processing, digital image forensic, network security and application of IoT.



Qiang Zhang

He obtained her BE of computer application at Tianjin Polytechnic University in China in 2011, and she is a postgraduate at Tianjin University in China now, on the research subjects of image processing and digital image forensic.



Weidong Min

He obtained his BE, ME and PhD of computer application at Tsinghua University in China in 1989, 1991 and 1995, respectively, on the research subjects of computer graphics, image processing and computer aided geometric design. He was an assistant professor of Tsinghua University from 1994 to 1995. From 1995 to 1997 he was a postdoctoral researcher at University of Alberta, Canada. From 1998 to 2011 he worked as a senior research project manager at Corel, March Networks and other companies in Canada. Now he is an associate professor at School of Computer Science & Software Engineering, Tianjin Polytechnic University, China. His current research interests include computer graphics, image and video processing, software engineering, distributed system, network resources management.



Shuguang Zhang

He received his BE of computer application at Nanjing Artillery Academy in China in 2002, on the research subject of education informatization, SOA and ERP. From 2002 to the present, he worked in Jinzhi Education Information Technology Co., Ltd., Jiangsu, china. Now he is the general manager of Chinese Middle West Region. His current research interests include networking, wisdom campus, image and video processing, network security and application.

