Image Tampering Detection Using No-Reference Image Quality Metrics

Ying Li, Bo Wang, Xiang-Wei Kong, Yan-Qing Guo

(School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: In this paper, a new approach is proposed to determine whether the content of an image is authentic or modified with a focus on detecting complex image tampering. Detecting image tampering without any prior information of the original image is a challenging problem since unknown diverse manipulations may have different characteristics and so do various formats of images. Our principle is that image processing, no matter how complex, may affect image quality, so image quality metrics can be used to distinguish tampered images. In particular, based on the alteration of image quality in modified blocks, the proposed method can locate the tampered areas. Referring to four types of effective no-reference image quality metrics, we obtain 13 features to present an image. The experimental results show that the proposed method is very promising on detecting image tampering and locating the locally tampered areas especially in realistic scenarios.

Keywords: image forensics; tampering detection; no-reference; image quality metrics; tampering localization

CLC number: TP391.7 **Document code**: A **Article ID**: 1005-9113(2014)06-0051-06

1 Introduction

Digital images are important media for human communication, but it is difficult to ensure the primitiveness of digital images since they may be tampered intentionally or not, like uploading or saving images, embellishing photographs or even changing content of images, which may have deleterious effects when forged images are used as judicial evidences, news media and so on. As a consequence, the study of digital image tampering detection is significant for exploring the authenticity and primitiveness of digital images.

The passive image tampering detection, which is also called blind detection, is one of the most challenging and useful research fields of digital image forensics, because this passive method needs no knowledge of prior information of the original image. A number of approaches have been advocated to tackle this task. Fridrich et al. first present a block matching procedure [1] for copy-move detection. All studies following this approach are similar except for features selected to represent each block [2]. Dalgaard et al. focus on resampling detection [3]. Dong et al. take the GLCM (Gray Level Co-occurrence Matrix) of the DCT coefficients as the evidence of double JPEG compression [4]. Some methods detect splicing basing on shadows [5] or inconsistencies in perspective [6].

Different from the above methods, Battisti et al. made use of image quality assessment [7] to detect image forgery. Prior methods are limited to detecting specific image manipulations, while image quality assessment (IQA) may have an advantage over them as a methodology for detecting complex image modification. The principle of image tampering detection based on IQA is the presence of differences in quality degradations impairing the images, and IQA can just evaluate image quality with a quantitative metric. In fact, the similar theory has been used on image forensics before. Image Quality Metric (IQM) is considered to help to identify source cameras [8]. However, it is difficult to choose effective IQM for tampering detection because tampered images just appear tiny visible changes. In order to promote the tampering detection accuracy, more efficient IQMs are to be studied.

In this paper, a new method is proposed based on Combination of No-Reference Image Quality Metrics (CNR-IQM) consisting of 13 features for detecting image tampering, without any priori information of the initial image. The experimental results show that the CNR-IQM is more accurate than image quality features mentioned above [7-8], and the proposed method is effective in locating the locally tampered areas not only in the simulated tampering scenarios, but also in realistic scenarios with more complex modifications.

2 Proposed CNR-IQM for Image Tampering Detection Method

Our studies have led to the assumption that the

Received 2014-01-06.

Sponsored by the National Natural Science Foundation of China (Grant No. 60971095 and No. 61172109), Artificial Intelligence Key Laboratory of Sichuan Province (Grant No. 2012RZJ01) and also the Fundamental Research Funds for the Central Universities (Grant No. DUT13RC201).

Corresponding author: Xiang-Wei Kong. E-mail: kongxw@dlut.edu.cn

method based on the combination of different types of features usually performs better across all types of modifications than individual algorithms. So a hybrid method is presented as CNR-IQM in this paper. As is mentioned above, suitable IQM is a key point for our tampering detection approach. We select effective No-Reference Image Quality Metrics (NRIQMs) by testing the performance of several latest NRIQMs which can evaluate multiple modifications or mainly aim at blur and compression which are common operations performed on forged digital images and may cause the image quality degradation.

2.1 No-Reference Image Quality Metrics from Spatial Domain

Battisti et al. prove a no-reference quality metric designed by Wang et al. [9] to be effective in detecting image forgery [7]. In this paper, we use the three features, i.e. Blocking, Activity and Zero-Crossing, as NRIQMs directly instead of the features fusion approach proposed in the prior works, considering that features fusion may introduce errors, and on the contrary, raising the dimensions of features may improve the performance of the SVM classifier. Given an image of size $[M,N]$, the Blocking feature along horizontal 8×8 blocks is calculated as:

$$B_h = \frac{1}{M\left(\lfloor \frac{N}{8} \rfloor - 1\right) \sum_{i=1}^{M} \sum_{j=1}^{\lfloor \frac{N}{8} \rfloor - 1} |\boldsymbol{d}_h(i,8j)|} \qquad (1)$$

where $\boldsymbol{d}_h$ is the horizontal difference between neighbor pixels:

$$\boldsymbol{d}_h(m,n) = x(m,n+1) - x(m,n),$$
$$n \in [1,N-1], \quad m \in [1,M-1] \qquad (2)$$

$\boldsymbol{x}(m,n)$ represents the value of the pixel at (m,n), $m \in [1,M]$, and $n \in [1,N]$. The activity feature along horizontal 8×8 blocks is calculated as:

$$A_h = \frac{1}{7}\left[\frac{8}{M(N-1)} \sum_{i=1}^{M} \sum_{j=1}^{N-1} |\boldsymbol{d}_h(i,j)| - B_h\right] \qquad (3)$$

and Zero-crossing is another aspect of activity. Zero-crossing feature along horizontal 8×8 blocks is:

$$Z_h = \frac{1}{M(N-2)} \sum_{i=1}^{M} \sum_{j=1}^{N-2} z_h(m,n) \qquad (4)$$

where $m \in [1,M]$, $n \in [1,N-2]$, and z_h is computed as:

$$z_h(m,n) = \begin{cases} 1, & \text{horizontal } Z \text{ at } d_h(m,n) \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

Similar to the horizontal features achieved above, vertical ones can be obtained. Then the three features can be computed as follows:

$$B = \frac{B_h + B_v}{2}, A = \frac{A_h + A_v}{2}, Z = \frac{Z_h + Z_v}{2} \qquad (6)$$

The $\boldsymbol{x}(m,n)$ in Eq.(2) represents the pixel's value of the matrix which is one of the three color components of an image in YC_bC_r color space. In this paper, we calculate the 3 features ($B\&A\&Z$) corresponding to each color component according to Eq.(6), resulting in 9 features (i.e. $3B\&3A\&3Z$) in total.

Another research of NRIQMs that we take into consideration is on the natural scene statistics model in the spatial domain [10], for the method performs better than other well-sited no reference algorithms in evaluating several common distortions. $\boldsymbol{I}$ is the image's luminance component, and its size is $M \times N$. The mean subtracted contrast normalized (MSCN) coefficients $\hat{\boldsymbol{I}}$ is computed as:

$$\hat{\boldsymbol{I}}(i,j) = \frac{\boldsymbol{I}(i,j) - \mu(i,j)}{\boldsymbol{\sigma}(i,j) + C}$$

where $i \in [1,M]$, and $j \in [1,N]$ are the spatial indices, $C = 1$ preventing the occurrence of the denominator to be zero,

$$\boldsymbol{\mu}(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} \boldsymbol{I}_{k,l}(i,j)$$

$$\boldsymbol{\sigma}(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} \left(\boldsymbol{I}_{k,l}(i,j) - \boldsymbol{\mu}(i,j)\right)^2}$$

where $\boldsymbol{I}_{k,l}(i,j)$ is the value of the luminance of the pixel $(i+k, j+l)$, $k \in \{-3,\cdots,3\}$, $l \in \{-3,\cdots,3\}$ and $w_{k,l}$ is a Circularly-Symmetric Gaussian filter sampled out to 3 standard deviations in Ref.[10]. However, our experimental results show that the Circularly-Symmetric Gaussian filter above does not generate Gaussian datasets. So in this paper we calculate $\boldsymbol{\mu}(i,j)$ as the average of 7 × 7 neighborhood of $\boldsymbol{I}(i,j)$ by setting $w_{k,l} = 1/(7 \times 7)$, making the distribution of MSCN coefficients more like Gauss according to our experiments. Besides, we replace the AGGD (Asymmetric Generalized Gaussian Model) with the Gauss model to fit the MSCN coefficients since the former is more complicated and takes longer time, and the latter one is enough for fitting the MSCN coefficients. In addition, to fix the fitting curve whose peak point is too high to keep the curve Gaussian, we expel the peak point during the fitting procedure. As a consequence, the absolute value of the fitting curve is far from that of the MSCN curve, but they still have similar shapes, so that the variance of fitting curve is used as the 10th feature.

NRIQMs from spatial domain are visualized and effective. To find out deeper relationships, statistical metrics in frequency domain are studied for more information.

2.2 No-Reference Image Quality Metrics from Frequency Domain

Since DCT (Discrete Cosine Transformation) is the basic step of JPEG compression, it is reasonable to associate DCT coefficients with the compression quality. Shen et al. took the indices of the second peak from the left of the LPMDC (Logarithmic Probability distribution function of the Magnitude of DCT Coefficients) as NRIQMs [11]. The LPMDC is calculated as:

$$h(x) = pdf(\log_{10}(|C|))$$

where C is the set of DCT coefficients, and pdf is the probability distribution function. However, in our experiments, the third peak from the left of the LPMDC is more sensitive than the second one, so the indices (the abscissa and the ordinate) of the third peak are taken as two NRIQMs instead, i.e. the 11th and 12th features.

Local phase coherence (LPC) has demonstrated good potentials in a number of image processing and computer vision applications, including image registration, fusion and sharpness evaluation. From previous results [12-13], we find that the proposed LPC feature is better than others in representing blur. $\boldsymbol{W}^{D_S}$ represents the S$^{\text{th}}$ scale of the wavelet decomposition of an image in the direction D (horizontal (H) or vertical (V)), the coefficients of which are classified into coherent and incoherent groups as:

$$\boldsymbol{W}^{D_S} = \boldsymbol{C}^{D_S} + \boldsymbol{I}^{D_S}$$

where $\boldsymbol{C}^{D_S}$ represents the coherent coefficients and $\boldsymbol{I}^{D_S}$ corresponds to the incoherent ones. An adaptive threshold is computed according to the steps below:

Algorithm 1 Adaptive threshold

```
Input: matrix $\boldsymbol{W}^{D_1}$, $\boldsymbol{W}^{D_2}$, $D = H$ or $V$
Set the threshold: $T \leftarrow 0$
loop:
    if $[\boldsymbol{W}^{D_1}(i,j)\ \boldsymbol{W}^{D_2}(i,j)] > T$ then
        $\boldsymbol{C}^{D_1}(i,j) \leftarrow \boldsymbol{W}^{D_1}(i,j)$, $\boldsymbol{I}^{D_1}(i,j) \leftarrow 0$
else
        $\boldsymbol{C}^{D_1}(i,j) \leftarrow 0$, $\boldsymbol{I}^{D_1}(i,j) \leftarrow \boldsymbol{W}^{D_1}(i,j)$
        $T_{\text{new}} \leftarrow$ compute the variance of $\boldsymbol{I}^{D_1}$
if $|T - T_{\text{new}}| > 1/(255^2)$ then
        $T \leftarrow T_{\text{new}}$
else
        end loop
```

The mean of standard deviations of $\boldsymbol{I}^{H_1}$ and $\boldsymbol{I}^{V_1}$ is taken as the 13th feature.

2.3 Image Tampering Detection by CNR-IQM

The CNR-IQM is a vector composed of the 13 NRIQMs above. The first 9 features are pixel-level ones obtained by comparing the differences between pixels like the discontinuity between neighbor pixels on the two sides of block boundaries (B) and the smoothness of pixels within blocks ($A\&Z$). The 10th feature is a statistic of the spatial domain, and the last 3 features are statistics from different transformation domains (Discrete Cosine Transformation and Wavelet Transformation).

By training SVM classifier with CNR-IQM, the training model can be obtained and then be used to detect tampered images. In the case of local tampering detection, we test the given image on slide blocks, and the test result is presented as a binary image.

The SVM classifier used is based on a library for support vector machines proposed byChih-Chung Chang and Chih-Jen Lin [14]. A radial basis function kernel is used, and the parameters C and γ both range from 2^{-5} to 2^5.

3 Experimental Results

In this section, we first verify that the CNR-IQM feature is effective for Gaussian blur and JPEG compression detection in both global images and local regions on three databases consisting of different formats of images (BMP, JPG, and TIFF) in simulated scenarios. And then we tamper some images by Photoshop for detection experiments to show the practicability of the proposed method on detecting complex modifications in realistic scenarios. The images of the databases are all 24-bits true color images:

Database 1 703 BMP images in total. The images are acquired by Canon Powershot Pro1 in JPG format originally and resized to 83% of the original images and converted to BMP images at last. The resolution of the images is 1328×996 pixels.

Database 2 523 JPG images in total, acquired by Nikon S210. The resolution of the images is 2592×1944 pixels. The images are originally saved as JPG images.

Database 3 530 TIFF images in total, acquired by Kodak DC290. The resolution of the images is 1440×960 pixels. The images are originally saved as TIFF images.

3.1 Image Tampering Detection Simulation

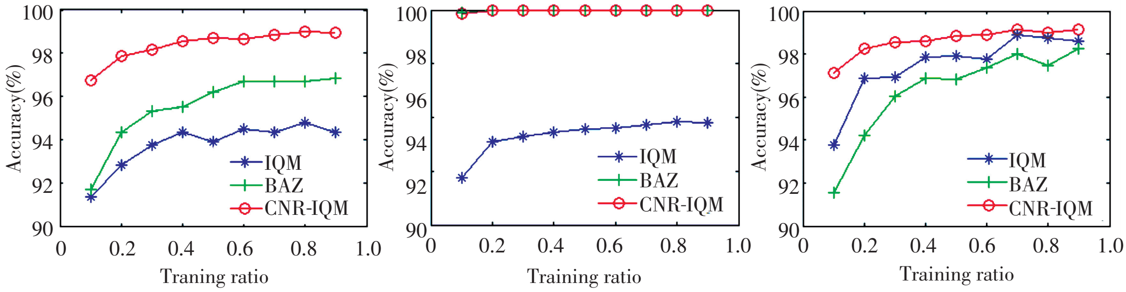
First, all of the images in databases are tampered globally by Gaussian blur and JPEG compression separately with MATLAB. In order not to affect the visual quality of images, in the tampering procedure, the size of Gaussian window is chosen as small as 5×5 pixels, and the quality factor of JPEG compression is as high as 90 in MATLAB. In the classifying procedure, both Gaussian blurred and JPEG compressed images are taken as tampered images. The experimental results are shown in Table 1. Training Ratio is the ratio of images taken for training to all images in the database, which is set from 0.1 to 0.9. The experiments are taken ten times on each Training Ratio, and Accuracy is the average accuracy of the ten testing results. The features of IQM [8] and BAZ [7] are also abstracted for comparison.

Table 1 and Fig.1 show that the CNR-IQM gets an observably higher accuracy in Database 1 and Database 3, and almost the same accuracy with BAZ in Database 2. So the proposed method is efficient for identifying whether images are tampered or not.

Table 1 Accuracy of image tampering detection simulation %

Training Ratio	Database 1			Database 2			Database 3		
	IQM	BAZ	CNR-IQM	IQM	BAZ	CNR-IQM	IQM	BAZ	CNR-IQM
0.1	91.33	91.68	96.74	68.83	99.59	99.36	96.89	95.77	98.57
0.2	92.81	94.33	97.87	75.42	99.92	99.86	98.44	97.11	99.13
0.3	93.77	95.33	98.17	76.55	99.93	99.90	98.45	98.01	99.27
0.4	94.32	95.50	98.54	77.17	99.94	99.86	98.94	98.44	99.30
0.5	93.88	96.21	98.71	77.89	99.95	99.90	98.96	98.40	99.41
0.6	94.49	96.67	98.64	77.94	100.00	99.97	98.88	98.68	99.45
0.7	94.34	96.70	98.86	78.66	100.00	99.96	99.45	99.01	99.56
0.8	94.80	96.71	98.98	79.21	100.00	100.00	99.37	98.74	99.50
0.9	94.32	96.85	98.97	78.93	100.00	100.00	99.31	99.12	99.56

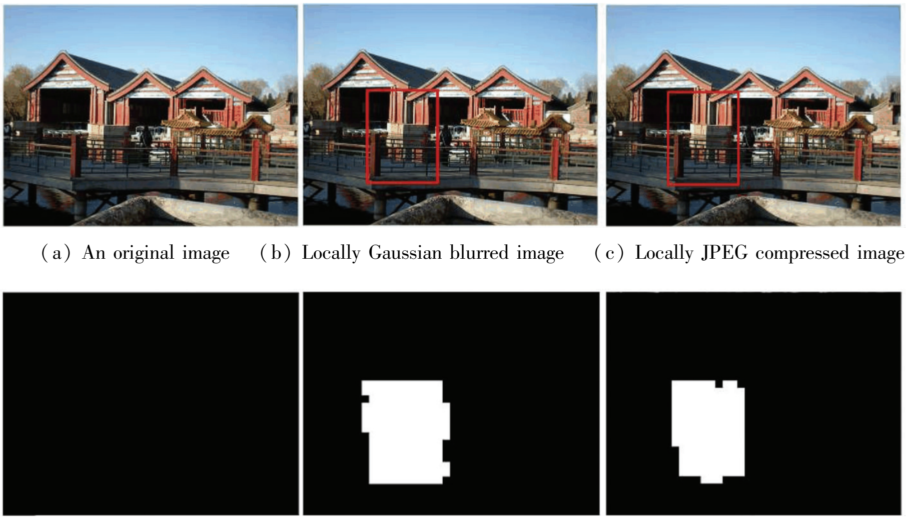
To further validate the proposed method on locating the locally tampered regions, we processimages locally by Gaussian blur and JPEG compression separately by MATLAB then. Only a rectangle block about 1/9 the area of original image is tampered in each image, which is 400×300 pixels in Database 1 and Database 3, 800 × 600 pixels in Database 2 considering the size of images in corresponding databases. An example of an image before and after locally tampering is shown in Fig.1, with the red rectangles marking the tampered regions.



(a) Database 1 (b) Database 2 (c) Database 3

Fig.1 Comparison of the accuracy results of image tampering detection simulation

In the test procedure, each slide block is 128×128 pixels and overlapped with the neighbor block of half of the area. The detection results are shown as binary images as in Fig.2, in which black regions correspond to clean districts, while white regions correspond to tampered districts. The proposed method can be viewed as effective in locating the tampered areas since the manipulations are so slight.



(a) An original image (b) Locally Gaussian blurred image (c) Locally JPEG compressed image

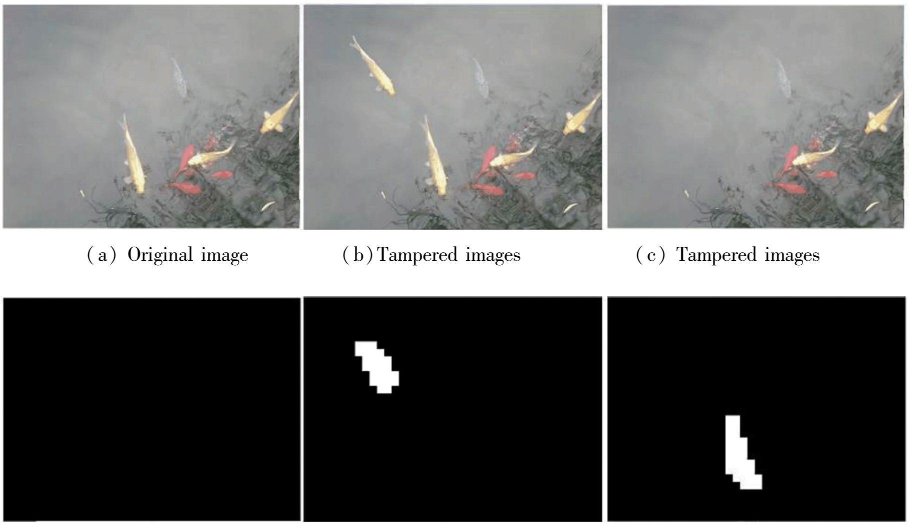
(d) Result of (a) (e) Result of (b) (f) Result of (c)

Fig.2 Local image tampering detection

3.2 Detection of Images Tampered by Photoshop

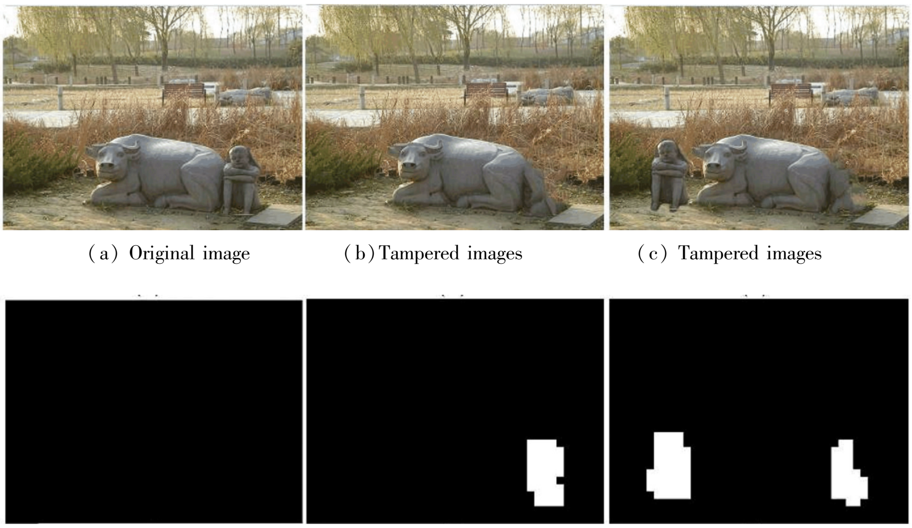
Photoshop is popular and powerful image processing software. We use some images tampered by Adobe Photoshop CS6 for experiments to verify the practicability of the proposed method. Common digital image tampering can be divided into two categories: the same image copy-move tampering, and splicing tampering with different images.

Examples for the test of former tampering are shown in Figs.3 and 4. Fig.3 (a) is an original image taken from Database 2, and we duplicate a fish in Fig.3(b) and resize and rotate the fish before moving it to the left corner. Besides, we use the Clone Stamp tool in Photoshop CS6 to hide a fish in Fig.3(c). Fig.4 is an example of the detection of Content-Aware Fill shown in Fig. 4 (b) and Content-Aware Move in Fig.4(c) which are new tools in Photoshop CS6 and can be viewed as automatically copy-move processes.


(a) Original image (b) Tampered images (c) Tampered images

(d) Detection results (e) Detection results (f) Detection results

Fig.3 The same image copy-move tampering

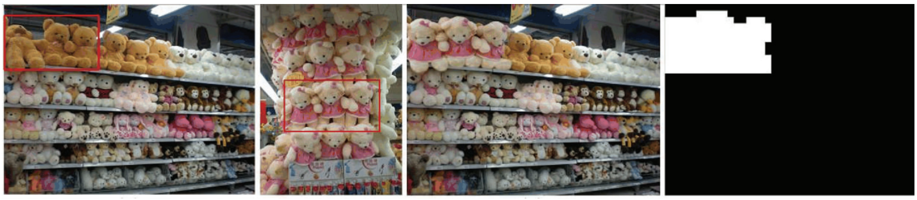

(a) Original image (b) Tampered images (c) Tampered images

(d) Detection results (e) Detection results (f) Detection results

Fig.4 The same image copy-move tampering (processed by content-aware tools in Photoshop CS6)

Another example for the test of latter tampering is shown in Fig.5. The three pink bear toys in the top left corner of Fig.5(c) are cut from Fig.5(b) and pasted on Fig.5(a) covering up the original brown bear toys. Resizing, rotation and blur are applied to make the tampered area more natural.

From the detection results, we can see clearly that for copy-move and splicing tampering accompanied with several complex processes in Photoshop, the tampering localization is accurate, justifying the effectiveness of our method on detecting complex modifications in realistic scenarios.



(a) Original images (b) Original images (c) Tampered image (d) Detection result

Fig.5 Splicing tampering with different images

4 Conclusion

In this paper, four types of efficient NRIQMs are combined as the CNR-IQM, and then result in 13 features to fulfill the detection of not only the global tampered images, but also the locally tampered districts of the images. The experimental results show that the proposed method is more effective than several existing methods, and verify the practicability of the proposed method on detecting complex modifications in realistic scenarios. But the accuracy of local image tampering detection is limited by the size of slide blocks divided for each time of detection. Although the proposed method achieved good results considering current research, the accuracy and certainty of blind tampering detection is not enough for legal purposes. Besides, the abstraction of more efficient NRIQMs will be another important component of our future work. With the prosperity of no-reference quality assessment technology, our method will achieve even broader development prospects.

References

[1] Fridrich A J, Soukal B D, Lukas A J. Detection of copy-move forgery in digital images. Proceedings of the Digital Forensic Research Workshop. Cleveland, Ohio. 2003.

[2] Piva A. An overview on image forensics. ISRN Signal Processing, 2013, 2013: 1-22.

[3] Dalgaard N, Mosquera C, Perez-Gonzalez F. On the role of differentiation for resampling detection. Proceedings of the International Conference on Image Processing. Piscataway: IEEE, 2010. 1753-1756.

[4] Dong Lisha, Kong Xiangwei, Wang Bo, et al. A robust JPEG image tampering detection method using GLCM features. Advances in Information Sciences and Service Sciences, 2011, 3(10): 384-391.

[5] Liu Q, Cao X, Deng C, et al. Identifying image composites through shadow matte consistency. IEEE Transactions on Information Forensics and Security, 2011, 6: 1111-1122.

[6] Kakar P, Sudha N, Ser W. Exposing digital image forgeries by detecting discrepancies in motion blur. IEEE Transactions on Multimedia, 2011, 13(3): 443-452.

[7] Battisti F, Carli M, Neri A. Image forgery detection by means of no-reference quality metrics. SPIE Conference on Media Watermarking, Security, and Forensics. Burlingame. California: SPIE, 2012.

[8] Mehdi K L, Sencar H T, Memon N. Blind source camera identification. Proceedings of the International Conference on Image Processing. Piscataway: IEEE, 2004. 709- 712.

[9] Wang Zhou, Sheikh H R, Bovik A C. No-reference perceptual quality assessment of JPEG compressed images. Proceedings of the International Conference on Image Processing. Piscataway: IEEE, 2002. 477-480.

[10] Mittal A, Moorthy A K, Bovik A C. Blind/referenceless image spatial quality evaluator. Proceedings of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR). Piscataway: IEEE, 2011. 723-727.

[11] Shen Ji, Li Qin, Erlebacher G. Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images. IEEE Transactions on Image Processing, 2011, 20(8): 2089-2098.

[12] Ciancio A, da Costa A L, da Silva E A B, et al. No-reference blur assessment of digital pictures based on multifeature classifiers. IEEE Transactions on Image Processing, 2011, 20(1): 64-75.

[13] Ciancio A, Targino A N, da Silva E A B, et al. Objective no-reference image quality metric based on local phase coherence. IET Electron. Lett., 2009, 45(23): 1162-1163.

[14] Chang C C, Lin C J. LIBSVM: A library for support vector machines. http://www. csie. ntu. edu. tw/~ cjlin/libsvm/. 2013-04-01.